

# MULTISENSORY INTEGRATION: BRAIN, BODY AND THE WORLD

EDITED BY: Magda L. Dumitru, Achille Pasqualotto and Andriy Myachykov  
PUBLISHED IN: Frontiers in Psychology





# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2016 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88919-792-7

DOI 10.3389/978-2-88919-792-7

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)



# MULTISENSORY INTEGRATION: BRAIN, BODY AND THE WORLD

Topic Editors:

**Magda L. Dumitru**, Macquarie University, Australia & Middle East Technical University, Turkey

**Achille Pasqualotto**, Sabanci University, Turkey

**Andriy Myachykov**, Northumbria University, UK & National Research University Higher School of Economics, Russia



Cover image: “Confident psychologist is solving the client’s problems”, taken from <http://www.dreamstime.com/stock-photo-confident-psychologist-solving-client-problems-image62306097>

Behaviour, language, and reasoning are expressions of neural functions par excellence, as the brain must draw on sensory modalities to gather information on the rest of the body and on the outer world. Cortical areas processing the identity and location of sensory inputs were once thought to be organised hierarchically, with some branches dedicated to basic features and other branches dedicated to complex features. Yet current studies have uncovered synergistic effects at early sensory cortices as well as at higher-level association areas. A less hierarchical functional architecture of the brain has emerged such that, irrespective of sensory modality, inputs would be allocated to the best suited cortical substrate. It is our hope that the articles included in this special issue will offer novel insights into recent developments relating to multisensory integration and brain functioning.

**Citation:** Dumitru, M. L., Pasqualotto, A., Myachykov, A., eds. (2016). *Multisensory Integration: Brain, Body and the World*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-792-7

# Table of Contents

- 05 Editorial: Multisensory Integration: Brain, Body, and World**  
Achille Pasqualotto, Magda L. Dumitru and Andriy Myachykov
- 07 Is the auditory system cognitively penetrable?**  
Berit Brogaard and Dimitria Electra Gatzia
- 11 Auditory scene analysis and sonified visual images. Does consonance negatively impact on object formation when using complex sonified stimuli?**  
David J. Brown, Andrew J. R. Simpson and Michael J. Proulx
- 20 Semantic-based crossmodal processing during visual suppression**  
Dustin Cox and Sang Wook Hong
- 29 Modality use in joint attention between hearing parents and deaf children**  
Nicole Depowski, Homer Abaya, John Oghalai and Heather Bortfeld
- 37 Effects of word-evoked object size on covert numerosity estimations**  
Magda L. Dumitru and Gitte H. Joergensen
- 43 Encoding audio motion: spatial impairment in early blind individuals**  
Sara Finocchietti, Giulia Cappagli and Monica Gori
- 49 Voluntary initiation of movement: multifunctional integration of subjective agency**  
Patrick Grüneberg, Hideki Kadone and Kenji Suzuki
- 63 How our body influences our perception of the world**  
Laurence R. Harris, Michael J. Carnevale, Sarah D'Amour, Lindsey E. Fraser, Vanessa Harrar, Adria E. N. Hoover, Charles Mander and Lisa M. Pritchett
- 73 Nonlinear response speedup in bimodal visual-olfactory object identification**  
Richard Höchenberger, Niko A. Busch and Kathrin Ohla
- 84 Illusory visual motion stimulus elicits postural sway in migraine patients**  
Shu Imaizumi, Motoyasu Honma, Haruo Hibino and Shinichi Koyama
- 94 Amplitude-modulated stimuli reveal auditory-visual interactions in brain activity and brain connectivity**  
Mark Laing, Adrian Rees and Quoc C. Vuong
- 105 Feature activation during word recognition: action, visual, and associative-semantic priming effects**  
Kevin J. Y. Lam, Ton Dijkstra and Shirley-Ann Rueschemeyer
- 113 Aesthetic Emotions Across Arts: A Comparison Between Painting and Music**  
Andrei C. Miu, Simina Pițur and Aurora Szentágotai-Tătar
- 122 The multisensory body revealed through its cast shadows**  
Francesco Pavani and Giovanni Galfano
- 133 Audiovisual integration of emotional signals from others' social interactions**  
Lukasz Piwek, Frank Pollick and Karin Petrini



- 143 Sources of Confusion in Infant Audiovisual Speech Perception Research**  
Kathleen E. Shaw and Heather Bortfeld
- 149 Action sounds update the mental representation of arm dimension: contributions of kinaesthesia and agency**  
Ana Tajadura-Jiménez, Manos Tsakiris, Torsten Marquardt and Nadia Bianchi-Berthouze
- 167 What makes space-time interactions in human vision asymmetrical?**  
Chizuru T. Homma and Hiroshi Ashida
- 175 Tactile input and empathy modulate the perception of ambiguous biological motion**  
Hörmetjan Yiltiz and Lihan Chen
- 186 Corrigendum: Tactile input and empathy modulate the perception of ambiguous biological motion**  
Hörmetjan Yiltiz and Lihan Chen



# Editorial: Multisensory Integration: Brain, Body, and World

Achille Pasqualotto<sup>1</sup>, Magda L. Dumitru<sup>2,3\*</sup> and Andriy Myachykov<sup>4,5</sup>

<sup>1</sup> Faculty of Arts and Social Sciences, Sabanci University, Istanbul, Turkey, <sup>2</sup> Department of Cognitive Science, Macquarie University, Sydney, NSW, Australia, <sup>3</sup> Cognitive Science Department, Graduate School of Informatics, Middle East Technical University, Ankara, Turkey, <sup>4</sup> Department of Psychology, Northumbria University Newcastle, Newcastle-upon-Tyne, UK, <sup>5</sup> School of Psychology, Centre for Cognition and Decision Making, National Research University Higher School of Economics, Moscow, Russia

**Keywords:** multisensory integration, body representation, attentional deployment, emotional processing, numerical cognition, language, embodied reasoning, time processing

## The Editorial on the Research Topic

### Multisensory Integration: Brain, Body, and World

The brain is safely sealed inside the cranium, with virtually no direct interaction with other parts of the body and the outside world. Nevertheless, it constantly processes the information conveyed by several sensory modalities in order to create representations of both body and outer world and to generate appropriate motor responses (Ehrsson et al., 2005; Farnè et al., 2005; Green and Angelaki, 2010). For example, vision can convey information about dangerous stimuli to trigger the generation of appropriate motor response (e.g., escape, avoidance, fight, etc.). Rather than processing sensory inputs in isolation, the brain *integrates* sensory information (Stein and Meredith, 1993; Fetsch et al., 2012) by forming reliable and robust representation of the external world and body. For example, when both visual and auditory input inform about the same danger, an appropriate motor response is more rapid and efficient (Serenio and Huang, 2006; Laing et al.).

Until a few decades ago, it was strongly believed that sensory (or multisensory) integration occurred only in high-level/associative areas or the cortex (Ghazanfar and Schroeder, 2006; Pavani and Galfano). Recently, several “new” multisensory areas have been discovered (Gobbelé et al., 2003; Pietrini et al., 2004), suggesting that a larger portion of the cortex is engaged in multisensory processing. Additional evidence suggests that multisensory integration also occurs in sub-cortical areas (Kuraoka and Nakamura, 2007; Amad et al., 2014). Finally, and perhaps surprisingly, some studies have demonstrated that multisensory processing occurs in primary sensory areas that were traditionally considered to be *uni-sensory* (Zangaladze et al., 1999; Murray et al., 2005).

Theories such as “neural reuse” (Anderson, 2010) and “metamodal” organization of the brain (Pascual-Leone and Hamilton, 2001) attempt to provide new paradigms for brain functioning taking into account widespread multisensory integration. The evolutionary advantage of multisensory integration might be the resulting availability of more reliable representations of the external world and body (Elliott et al., 2010; Grüneberg et al.) based on multiple sensory inputs and the resilience to brain injuries and sensory loss (Sarno et al., 2003; Pasqualotto and Proulx, 2012; Brown et al.; Finocchietti et al.). Indeed, multisensory integration has been reported in various experimental tasks including spatial representation (Pasqualotto et al., 2005), object recognition (Woods and Newell, 2004; Harris et al.; Höchenberger et al.; Laing et al.), movement perception (Grüneberg et al.; Imaizumi et al.; Uesaki and Ashida), body representation (Pasqualotto and Proulx, 2015; Pavani and Galfano; Tajadura-Jiménez et al.; Yiltiz and Chen), emotional processing (Miu et al.; Piwek et al.), attentional deployment (Spence, 2002; Depowski et al.), language

## OPEN ACCESS

### Edited and reviewed by:

Bernhard Hommel,  
Leiden University, Netherlands

### \*Correspondence:

Magda L. Dumitru  
magda.dumitru@gmail.com

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 17 December 2015

**Accepted:** 23 December 2015

**Published:** 12 January 2016

### Citation:

Pasqualotto A, Dumitru ML and  
Myachykov A (2016) Editorial:  
Multisensory Integration: Brain, Body,  
and World. *Front. Psychol.* 6:2046.  
doi: 10.3389/fpsyg.2015.02046

(Gallese, 2008; Myachykov and Tomlin, 2008; Myachykov et al., 2012; Lam et al.; Shaw and Bortfeld), embodied reasoning (Dumitru, 2014), sensory awareness (Cox and Hong), numerical cognition (Dumitru and Joergensen), auditory perception (Brogaard and Gatzia), and time perception (Homma and Ashida).

The articles included in this special issue offer novel insights about recent developments within the field of multisensory integration, and we believe that they will help understanding the multisensory nature of brain functioning.

## REFERENCES

- Amad, A., Cachia, A., Gorwood, P., Pins, D., Delmaire, C., Rolland, B., et al. (2014). The multimodal connectivity of the hippocampal complex in auditory and visual hallucinations. *Mol. Psychiatry* 19, 184–191. doi: 10.1038/mp.2012.181
- Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* 33, 245–266. doi: 10.1017/S0140525X1000853
- Dumitru, M. L. (2014). Moving stimuli guide retrieval and (in)validation of coordination simulations. *Cogn. Process.* 15, 397–403. doi: 10.1007/s10339-014-0604-6
- Elliott, M. T., Wing, A. M., and Welchman, A. E. (2010). Multisensory cues improve sensorimotor synchronisation. *Eur. J. Neurosci.* 31, 1828–1835. doi: 10.1111/j.1460-9568.2010.07205.x
- Ehrsson, H. H., Holmes, N. P., and Passingham, R. E. (2005). Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas. *J. Neurosci.* 25, 10564–10573. doi: 10.1523/JNEUROSCI.0800-05.2005
- Farnè, A., Iriki, A., and Làdavas, E. (2005). Shaping multisensory action-space with tools: evidence from patients with cross-modal extinction. *Neuropsychologia* 43, 238–248. doi: 10.1016/j.neuropsychologia.2004.11.010
- Fetsch, C. R., Pouget, A., DeAngelis, G. C., and Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nat. Neurosci.* 15, 146–154. doi: 10.1038/nn.2983
- Gallese, V. (2008). Mirror neurons and the social nature of language: the neural exploitation hypothesis. *Soc. Neurosci.* 3, 317–333. doi: 10.1080/17470910701563608
- Ghazanfar, A. A., and Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci.* 10, 278–285. doi: 10.1016/j.tics.2006.04.008
- Gobbelé, R., Schürmann, M., Forss, N., Juottonen, K., Buchner, H., and Hari, R. (2003). Activation of the human posterior parietal and temporoparietal cortices during audiotactile interaction. *Neuroimage* 20, 503–511. doi: 10.1016/S1053-8119(03)00312-4
- Green, A. M., and Angelaki, D. E. (2010). Multisensory integration: resolving sensory ambiguities to build novel representations. *Curr. Opin. Neurobiol.* 20, 353–360. doi: 10.1016/j.conb.2010.04.009
- Kuraoka, K., and Nakamura, K. (2007). Responses of single neurons in monkey amygdala to facial and vocal emotions. *J. Neurophysiol.* 97, 1379–1387. doi: 10.1152/jn.00464.2006
- Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., et al. (2005). Grabbing your ear: Rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb. Cortex* 15, 963–974. doi: 10.1093/cercor/bbh197
- Myachykov, A., and Tomlin, R. (2008). Perceptual priming and syntactic choice in Russian sentence production. *J. Cogn. Sci.* 9, 31–48. doi: 10.17791/jcs.2008.9.1.31
- Myachykov, A., Garrod, S., and Scheepers, C. (2012). Determinants of structural choice in visually-situated sentence production. *Acta Psychol.* 141, 304–315. doi: 10.1016/j.actpsy.2012.09.006
- Pascual-Leone, A., and Hamilton, R. (2001). The metamodal organization of the brain. *Prog. Brain Res.* 134, 427–445. doi: 10.1016/S0079-6123(01)34028-1
- Pasqualotto, A., Finucane, C. M., and Newell, F. N. (2005). Visual and haptic representations of scenes are updated with observer movement. *Exp. Brain Res.* 166, 481–488. doi: 10.1007/s00221-005-2388-5
- Pasqualotto, A., and Proulx, M. J. (2012). The role of visual experience for the neural basis of spatial cognition. *Neurosci. Biobehav. Rev.* 36, 1179–1187. doi: 10.1016/j.neubiorev.2012.01.008
- Pasqualotto, A., and Proulx, M. J. (2015). Two-dimensional rubber-hand illusion: the Dorian Gray hand illusion. *Multisens. Res.* 28, 101–110. doi: 10.1163/22134808-00002473
- Pietrini, P., Furey, M. L., Ricciardi, E., Gobbi, M. I., Wu, W. H. C., Cohen, L. G., et al. (2004). Beyond sensory images: object-based representation in the human ventral pathway. *Proc. Natl. Acad. Sci. U.S.A.* 101, 5658–5663. doi: 10.1073/pnas.0400707101
- Sarno, S., Erasmus, L. P., Lipp, B., and Schlaegel, W. (2003). Multisensory integration after traumatic brain injury: a reaction time study between pairings of vision, touch and audition. *Brain Inj.* 17, 413–426. doi: 10.1080/0269905031000070161
- Sereno, M. I., and Huang, R. S. (2006). A human parietal face area contains aligned head-centered visual and tactile maps. *Nat. Neurosci.* 9, 1337–1343. doi: 10.1038/nn1777
- Spence, C. (2002). Multisensory attention and tactile information-processing. *Behav. Brain Res.* 135, 57–64. doi: 10.1016/S0166-4328(02)00155-9
- Stein, B. E., and Meredith, M. A. (1993). *The Merging of the Senses*. Boston, MA: The MIT Press.
- Woods, A. T., and Newell, F. N. (2004). Visual, haptic and cross-modal recognition of objects and scenes. *J. Physiol.* 98, 147–159. doi: 10.1016/j.jphysparis.2004.03.006
- Zangaladze, A., Epstein, C. M., Grafton, S. T., and Sathian, K. (1999). Involvement of visual cortex in tactile discrimination of orientation. *Nature* 401, 587–590. doi: 10.1038/44139

## AUTHOR CONTRIBUTIONS

AP wrote the first draft of the manuscript. MD and AM provided comments, additions, and further improvements. All authors have approved the final version of the manuscript.

## ACKNOWLEDGMENTS

MD was supported by a Marie Curie FP7-PEOPLE-IAPP fellowship (grant number 610986).

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Pasqualotto, Dumitru and Myachykov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Is the auditory system cognitively penetrable?

Berit Brogaard<sup>1,2\*</sup> and Dimitria Electra Gatzia<sup>3</sup>

<sup>1</sup> The Brogaard Lab for Multisensory Research, University of Miami, Miami, FL, USA, <sup>2</sup> Department of Philosophy, University of Oslo, Oslo, Norway, <sup>3</sup> Department of Philosophy, The University of Akron Wayne College, Akron, OH, USA

**Keywords:** auditory perception, cognitive penetration, McGurk illusion, semantic-coherence, top-down influences, top-down modulation, tritone illusion, perceptual learning

According to the hierarchical model of sensory information processing, sensory inputs are transmitted to cortical areas, which are crucial for complex auditory and speech processing, only after being processed in subcortical areas (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009). However, studies using electroencephalography (EEG) indicate that distinguishing simultaneous auditory inputs involves a widely distributed neural network, including the medial temporal lobe, which is essential for declarative memory, and posterior association cortices (Alain et al., 2001; Squire et al., 2004). More recent studies have even demonstrated plasticity of auditory signals as low as the brainstem (Suga, 2008). Collectively, studies suggest that the functional architecture of perceptual processing involves primarily top-down modulation (Suga et al., 2002; Gilbert and Li, 2013; Chandrasekaran et al., 2014). Top-down influences exerted throughout the auditory systems (Lotto and Holt, 2011) include: memory (Goldinger, 1998)<sup>1</sup>, attention (Choi et al., 2014), which has been found to modulate auditory encoding in the cochlea, a subcortical area (Maison et al., 2001), (prior) knowledge of syntax or words (Ganong, 1980; Warren, 1984)<sup>2</sup>, and experience-based expectations pertaining to the speaker's accent (Deutsch, 1996; Deutsch et al., 2004; Irino and Patterson, 2006), gender (Johnson et al., 1999), and vocal folds or tract (Irino and Patterson, 2002; Patterson and Johnsrude, 2008).

While a great deal has been written about the issue of cognitive penetrability in the case of vision, audition has received almost no attention. For example, a corresponding body of evidence for top-down modulation in vision has been used to undermine the Cognitive Impenetrability Thesis (CIT) (see Macpherson, 2012; Siegel, 2012; Wu, 2013; Cecchi, 2014). Brogaard and Gatzia (in press) have argued that top-down modulation on visual processes involving prior-knowledge, experience based expectation, or memory do not threaten the CIT, even after acknowledging that such influences are cognitive in nature (see also Pylyshyn, 1999; Raftopoulos, 2001). The reason is that such top-down influences, although cognitive in nature, are distinct from discursive thoughts that stand in a semantically-coherent relation to the phenomenology or content of experience, for instance, thoughts proceeding by argumentation or reasoning rather than by intuition or implicit hypothesis internal to the visual system<sup>3</sup>. If we insisted that instances of top-down modulation be counted as instances of cognitive penetration, the debate about cognitive penetrability would be trivial and, hence, unmotivated since studies clearly indicate that such top-down modulation in visual (or auditory) perception is extensive. A similar argument can be made in the case of audition.

## OPEN ACCESS

### Edited by:

Andriy Myachykov,  
Northumbria University, UK

### Reviewed by:

Andrew J. Lotto,  
University of Arizona, USA

### \*Correspondence:

Berit Brogaard,  
brit@miami.edu

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 10 April 2015

**Accepted:** 24 July 2015

**Published:** 11 August 2015

### Citation:

Brogaard B and Gatzia DE (2015) Is  
the auditory system cognitively  
penetrable? *Front. Psychol.* 6:1166.  
doi: 10.3389/fpsyg.2015.01166

<sup>1</sup> It has been suggested that the mechanism underlying auditory restoration (the auditory system's ability to compensate for expected missing sounds, see Warren, 1984) involves episodic memory, which involves memory traces left by an experience that are activated, according to the similarity with the stimulus, when a new stimulus such as a word is heard (see Goldinger, 1998).

<sup>2</sup> As the Ganong effect illustrates, phonemes such as /t/ or /d/ tend to be heard as /t/ when followed by "ask" to form "task" but as /d/ when followed by "usk" to form "dusk."

<sup>3</sup> Constancy computations, for example, are not obligatorily linked to experiencing sensibles and may precede it (Kentridge et al., 2014).

The CIT has traditionally been understood as a semantic thesis. Accordingly, the information a system computes is not sensitive (in a semantically-coherent way) to one's cognitive states and cannot be altered in a way that bears a logical relation to one's knowledge or reasons (Pylyshyn, 1984, 1999; Raftopoulos, 2009). For example, suppose that you experience a sound as /da-da/ and that causes you to form the belief that the sound is /da-da/. In this case, your belief and your auditory experience are semantically coherent: they have roughly the same content. Suppose now that you acquire the belief that the sound is in fact /ba-ba/ (say, because you have now come to believe that the Cartesian evil genius has made you hear it as /da-da/ when it is in fact a /ba-ba/ sound). According to the semantic thesis, your newly acquired belief, for which you may have ample justification, cannot alter the content computed by your auditory system; you will continue to experience the sound as /da-da/ despite that you have come to believe that it is /ba-ba/. Some proponents of the semantic thesis have argued that changes to the information a system computes are attributed to intra-perceptual principles that do not conform to standard tenets of rationality, such as standard rules of logic, probability theory and statistics, or rational choice theory (Brogaard and Gatzia, in press).

Undermining the CIT requires demonstrating that changes in the phenomenology of one's auditory perception are due to the listener's discursive or rational thoughts that stand in the right sort of semantic relation to her experience. So it is not enough that discursive thoughts influence experience; they must do so in a semantically-coherent way. Consider ventriloquism, for example. Suppose that I believe that the puppet is not actually producing the sounds (the person holding the puppet is) but I nevertheless hear the speech as coming from the puppet's mouth. In this case, the content of my belief differs from the content as my auditory experience. Now suppose that my discursive thoughts about what *really* goes on in the case of ventriloquism gives rise to a stress reaction in me (for some reason) and that this mood (the stress) changes the content of my experience: I no longer *hear* the speech as coming from the puppet. In this case, it may appear that my discursive thoughts have changed my auditory experience in a semantically-coherent way: my belief and my experience now have the same content. However, by hypothesis, it is the mood, not my beliefs, that changed my auditory experience. Since moods, unlike beliefs, have no contents, the stress (a mood) cannot have the same content as either my belief or my auditory experience. The content of my experience has thus changed but not in a semantically-coherent way. This semantic-coherence has to be involved in every step of the process for changes in phenomenology to threaten the CIT. For example, if my belief that the puppet is not actually producing the sounds were to cause me to no longer experience the speech as coming from the puppet via a chain of logically related processes, then the content of my belief would have changed the content of my experience in a semantically coherent-way. Such a case would indeed threaten the CIT.

Additionally, cases that involve the *indirect* influencing of auditory experience by beliefs (or discursive thoughts) need not threaten the CIT. For example, Fodor (1988) jokingly said that his heart is cognitively "penetrated" by his intention to do

calisthenics since it results in doing calisthenics, resulting in his heart rate increasing. What this joke illustrates is that the locution "receives input from" is not transitive, meaning that it is not the case that if a process B receives input from A, and C received input from B that C receives input from A since it is possible that none of B's outputs that were responses to inputs from A affected C (Lyons, 2015).

Cases of perceptual learning involve such indirect influencing of auditory perception. Typically, perceptual learning refers to the brain's plasticity, i.e., the gradual structural or functional changes in the connectivity of sensory systems resulting from training consisting of repeated exposure to particular stimuli (Roelfsema et al., 2010). However, the competition between verbal and implicit systems (COVIS) model suggests a dual-system framework, according to which learners, in information-integration tasks, initially use the reflective (rule-based) system, but switch to the reflexive (information-integration) system with practice (Maddox et al., 2013; Valentin et al., 2014)<sup>4</sup>. The fact that the reflective system is mediated by the prefrontal cortex and involves hypothesis testing by the learner seems to suggest that at least some cases of perceptual learning may constitute cases of cognitive penetration. This conclusion, however, is too hasty. The reflexive system is viewed as indirect and procedural: trial feedbacks reinforce associations of stimuli located in different regions of perceptual space with specific motor outputs (Maddox et al., 2013). It follows that the changes in auditory phenomenology associated with the reflexive system result indirectly from the brain's plasticity, not directly from the listener's discursive thoughts (in a semantically-coherent way). Perceptual learning, therefore, need not threaten the CIT, provided that the changes in phenomenology result indirectly from changes in the brain's plasticity, which cannot be attributed to the listener's discursive thoughts.

Auditory illusions are useful tools to illustrate the inability of our discursive thoughts to alter the phenomenology of our auditory experience in a semantically-coherent way. One example is the tritone illusion. Deutsch (2007) presented listeners with two tones in succession that are opposite in the positions along the pitch class space such as G# followed by D or C followed by F#, which comprised an interval of six semitones (known as tritone). When one of the pairs was played (say, G# followed by D) some of the listeners heard a descending pattern while others heard an ascending pattern. However, when another pair was played (say, C followed by F#) listeners who had previously heard a descending pattern now heard an ascending one and vice versa. The tritone illusion varies in correlation with the accent of the speaker. For example, while Californians tended to hear the pattern as ascending, Britons tended to hear it as descending (Deutsch, 1991). A considerable difference was also observed between mothers who had grown up in widely different geographical regions. Perhaps not surprisingly, significant similarities were observed among these mothers and their children, even though the children had not grown up in the same geographic regions as their mothers (Deutsch, 1996).

<sup>4</sup>We thank an anonymous reviewer for helpful comments on the issue of perceptual learning.

The tritone illusion persists even after listeners are informed that the two tones in succession are opposite in the positions along the pitch class space, indicating that their discursive thoughts cannot alter the phenomenology of their auditory experiences. What one hears depends on the configuration of one's auditory system, which is, among other things, subject to developmental influences (Deutsch et al., 2004). However, top-down modulation caused by adaptation- or development-based knowledge, experience-based expectation, memory, or attention are consistent with the claim that auditory perception is not cognitively penetrable, at least not in any interesting sense, as the changes in phenomenology cannot plausibly be attributed to the listener's discursive thoughts.

Another example is the McGurk illusion, which arises when auditory speech cues are presented in synchrony with incongruent visual speech cues (McGurk and MacDonald, 1976). For example, when the auditory syllable "ba" is presented in synchrony with a speaker mouthing "ga," subjects typically report hearing "da." However, when the auditory syllable "ga" is presented in synchrony with a speaker mouthing "ba," subjects typically report hearing "bga"<sup>5</sup>. As with the tritone illusion, the McGurk illusion persists even after subjects are informed that the auditory syllable is "ba" in the first case and "ga" in the second. Windmann (2004) found that the clarity and, to some extent, the probability of the illusion was significantly influenced by the listener's experience-based expectations, which do not threaten the CIT for the same reason: the information the system computes is not altered by the listener's discursive thoughts.

It may nevertheless be objected that other cases such as sine wave speech appear to threaten the CIT since they seem to involve changes in phenomenology which can be attributed to

subject's discursive thoughts<sup>6</sup>. For example, naive listeners tend to hear sine wave speech as tones or whistles, rather than speech. After being familiarized with the linguistic message, however, many listeners readily hear sine wave as speech (Sheffert et al., 2002). However, it is not clear, in this case, whether it is the listener's beliefs that cause a change in her experience. For example, it could be that such cases involve cognitive penetration if the listener's belief about the content of the linguistic message were to alter (in a semantically-coherent way) the phenomenology of the listener's experience. Or, it could be that the listener is still hearing the same tones or whistles but interprets them on the basis of the newly acquired knowledge of the linguistic message. The more likely explanation is that it is a case of normalization based on experience-based expectation given that the listener comes to understand sine wave speech only *after* learning its linguistic message. So it seems that the expectation that the sound has the linguistic message the listener expects it to have is what is doing all the work. Indeed, studies suggest that listeners use a range of information regarding the speaker, including the speaker's supposed nationality (Niedzielski, 1999), to create a frame of reference to be used during perception in order to normalize what is heard. In other words, listeners utilize adaptation- or development-based knowledge, experience-based expectation, memory, or attention to make sense of speech. However, as we have argued, such changes in phenomenology cannot plausibly be attributed to the listeners' discursive thoughts (at least not in a semantically-coherent way) and, thus, do not threaten the CIT.

## Acknowledgments

We would like to thank an anonymous referee for invaluable comments.

<sup>6</sup>We thank an anonymous reviewer for posing this question.

<sup>5</sup>Here too it is due to the non-transitivity of the locution "receives input from" that we cannot say that auditory processing is cognitively penetrated by visual processing (see Lyons, 2015).

## References

- Alain, C., Arnott, S. R., and Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 1072–1089. doi: 10.1037/0096-1523.27.5.1072
- Brogaard, B., and Gatzia, D. E. (in press). Color and cognitive penetrability. *Top. Cogn. Sci.*
- Cecchi, A. S. (2014). Cognitive penetration, perceptual learning, and neural plasticity. *Dialectica* 68, 63–95. doi: 10.1111/1746-8361.12051
- Chandrasekaran, B., Skoe, E., and Kraus, N. (2014). An integrative model of subcortical auditory plasticity. *Brain Topogr.* 27, 539–552. doi: 10.1007/s10548-013-0323-9
- Choi, I., Wang, L., Bharadwaj, H., and Shinn-Cunningham, B. (2014). Individual differences in attentional modulation of cortical responses correlate with selective attention performance. *Hear. Res.* 314, 10–19. doi: 10.1016/j.heares.2014.04.008
- Deutsch, D. (1991). The tritone paradox: an influence of language on music perception. *Music Percept.* 8, 335–347. doi: 10.2307/40285517
- Deutsch, D. (1996). Mothers and their children hear a musical illusion in strikingly similar ways. *J. Acoust. Soc. Am.* 99, 2482. doi: 10.1121/1.415579
- Deutsch, D. (2007). Music perception. *Front. Biosci.* 12, 4473–4482. doi: 10.2741/2402
- Deutsch, D., Henthorn, T., and Dolson, M. (2004). Speech patterns heard early in life influence later perception of the tritone paradox. *Music Percept.* 21, 357–372. doi: 10.1525/mp.2004.21.3.357
- Fodor, J. A. (1988). A reply to Churchland's "Perceptual plasticity and theory neutrality." *Philos. Sci.* 55, 188–194. doi: 10.1086/289426
- Ganong, W. F. III. (1980). Phonetic categorization in auditory word perception. *J. Exp. Psychol. Hum. Percept. Perform.* 6, 110–125. doi: 10.1037/0096-1523.6.1.110
- Gilbert, C. D., and Li, W. (2013). Top-down influences on visual processing. *Nat. Rev. Neurosci.* 14, 351. doi: 10.1038/nrn3476
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 105, 251–279. doi: 10.1037/0033-295X.105.2.251
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Irino, T., and Patterson, D. (2002). Segregation information about the size and shape of the vocal tract using a time-domain auditory model: the stabilised wavelet-Mellin transform. *Speech Commun.* 36, 181–203. doi: 10.1016/S0167-6393(00)00085-6



- Irino, T., and Patterson, R. D. (2006). A dynamic, compressive gammachirp auditory filterbank. *IEEE Audio Speech Lang. Process* 14, 2222–2232. doi: 10.1109/TASL.2006.874669
- Johnson, K., Strand, E. A., and D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *J. Phon.* 27, 359–384. doi: 10.1006/jpho.1999.0100
- Kentridge, R., Norman, L., Akins, K., and Heywood, C. (2014). “Colour constancy without consciousness,” in *Towards a Science of Consciousness Conference* (Tucson).
- Lotto, A., and Holt, L. L. (2011). Psychology of auditory perception. *Wiley Interdiscipl. Rev. Cogn. Sci.* 2, 479–489. doi: 10.1002/wcs.123
- Lyons, J. C. (2015). “Unencapsulated modules and perceptual judgment,” in *The Cognitive Penetrability of Perception: New Philosophical Perspectives*, eds J. Zeimekis and A. Raftopoulos (Oxford: Oxford University Press), 103–122.
- Macpherson, F. (2012). Cognitive penetration of colour experience: rethinking the issue in light of an indirect mechanism. *Philos. Phenomenol. Res.* 74, 24–62. doi: 10.1111/j.1933-1592.2010.00481.x
- Maddox, W. T., Chandrasekaran, B., Smayda, K., and Yi, H. G. (2013). Dual systems of speech category learning across the lifespan. *Psychol. Aging* 28, 1042–1056. doi: 10.1037/a0034969
- Maison, S., Micheyl, C., and Collet, L. (2001). Influence of focused auditory attention on cochlear activity in humans. *Psychophysiology* 38, 35–40. doi: 10.1111/1469-8986.3810035
- McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0
- Niedzielski, N. A. (1999). The effect of social information on the perception of sociolinguistic variables. *J. Lang. Soc. Psychol.* 18, 62–85. doi: 10.1177/0261927X99018001005
- Patterson, R. D., and Johnsrude, I. S. (2008). Functional imaging of the auditory processing applied to speech sounds. *Philos. Trans. R. Soc. B Biol. Sci.* 363, 1023–1035. doi: 10.1098/rstb/2007.2157
- Pylyshyn, Z. (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav. Brain Sci.* 22, 341–423. doi: 10.1017/S0140525X99002022
- Raftopoulos, A. (2001). Is perception informationally encapsulated? The issue of the theory-ladenness of perception. *Cogn. Sci.* 25, 423–451. doi: 10.1207/s15516709cog2503\_4
- Raftopoulos, A. (2009). *Perception and Cognition: How do Psychology and the Cognitive Sciences Inform Philosophy*. Cambridge, MA: MIT Press.
- Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331
- Roelofsma, P. R., van Ooyen, A., and Watanabe, T. (2010). Perceptual learning rules based on reinforcers and attention. *Trends Cogn. Sci.* 14, 64–71. doi: 10.1016/j.tics.2009.11.005
- Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., and Remez, R. E. (2002). Learning to recognize talkers from natural, sinewave, and reversed speech samples. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 1447–1469.
- Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Noûs* 46, 201–222. doi: 10.1111/j.1468-0068.2010.00786.x
- Squire, L. R., Stark, C. E., and Clark, R. E. (2004). The medial temporal lobe. *Annu. Rev. Neurosci.* 27, 279–306. doi: 10.1146/annurev.neuro.27.070203.144130
- Suga, N. (2008). Role of corticofugal feedback in hearing. *J. Comp. Physiol. A* 194, 169–183. doi: 10.1007/s00359-007-0274-2
- Suga, N., Xiao, Z., Ma, X., and Ji, W. (2002). Plasticity and corticofugal modulation for hearing in adult animals. *Neuron* 36, 9–18. doi: 10.1016/S0896-6273(02)00933-9
- Valentin, V. V., Maddox, W. T., and Ashby, F. (2014). A computational model of the temporal dynamics of plasticity in procedural learning: sensitivity to feedback timing. *Front. Psychol.* 5:643. doi: 10.3389/fpsyg.2014.00643
- Warren, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychol. Bull.* 96, 371–383. doi: 10.1037/0033-2909.96.2.371
- Windmann, S. (2004). Effects of sentence context and expectation on the McGurk illusion. *J. Mem. Lang.* 50, 212–230. doi: 10.1016/j.jml.2003.10.001
- Wu, W. (2013). Visual spatial constancy and modularity: does intention penetrate vision? *Philos. Stud.* 165, 647–669. doi: 10.1007/s11098-012-9971-y

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Brogaard and Gatzia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Auditory scene analysis and sonified visual images. Does consonance negatively impact on object formation when using complex sonified stimuli?

David J. Brown<sup>1,2\*</sup>, Andrew J. R. Simpson<sup>3</sup> and Michael J. Proulx<sup>1\*</sup>

<sup>1</sup> Crossmodal Cognition Lab, Department of Psychology, University of Bath, Bath, UK, <sup>2</sup> Biological and Experimental Psychology Group, School of Biological and Chemical Sciences, Queen Mary University of London, London, UK, <sup>3</sup> Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK

## OPEN ACCESS

### Edited by:

Achille Pasqualotto,  
Sabanci University, Turkey

### Reviewed by:

Pietro Pietrini,  
Azienda Ospedaliero-Universitaria  
Pisana, Italy  
Tina Iachini,  
Second University of Naples, Italy

### \*Correspondence:

David J. Brown and  
Michael J. Proulx,  
Crossmodal Cognition Lab,  
Department of Psychology, University  
of Bath, 2 South, Bath, BA2 7AY, UK  
djbrownmsp@gmail.com;  
m.j.proulx@bath.ac.uk

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 13 April 2015

**Accepted:** 22 September 2015

**Published:** 13 October 2015

### Citation:

Brown DJ, Simpson AJR  
and Proulx MJ (2015) Auditory scene  
analysis and sonified visual images.  
Does consonance negatively impact  
on object formation when using  
complex sonified stimuli?  
Front. Psychol. 6:1522.  
doi: 10.3389/fpsyg.2015.01522

A critical task for the brain is the sensory representation and identification of perceptual objects in the world. When the visual sense is impaired, hearing and touch must take primary roles and in recent times compensatory techniques have been developed that employ the tactile or auditory system as a substitute for the visual system. Visual-to-auditory sonifications provide a complex, feature-based auditory representation that must be decoded and integrated into an object-based representation by the listener. However, we don't yet know what role the auditory system plays in the object integration stage and whether the principles of auditory scene analysis apply. Here we used coarse sonified images in a two-tone discrimination task to test whether auditory feature-based representations of visual objects would be confounded when their features conflicted with the principles of auditory consonance. We found that listeners ( $N = 36$ ) performed worse in an object recognition task when the auditory feature-based representation was harmonically consonant. We also found that this conflict was not negated with the provision of congruent audio-visual information. The findings suggest that early auditory processes of harmonic grouping dominate the object formation process and that the complexity of the signal, and additional sensory information have limited effect on this.

**Keywords:** auditory scene analysis, consonance, signal complexity, blindness, cross-modal, sensory substitution

## Introduction

Our sensory systems provide a rich coherent representation of the world through the integration and discrimination of input from multiple sensory modalities (Spence, 2011). These low-level processes are modulated by high-order processing to selectively attend to task relevant stimuli. For example to attend to a speaker at a cocktail party we must select the low-level acoustic features that are relevant to the target, that is the person you are speaking with, from the environmental noise (Cherry, 1953). To accomplish this, feature-based sensory representations must be recombined into object-based representations in a rule based manner. In visual perception this is through scene analysis. Visual input is grouped into distinct objects based on Gestalt grouping rules such as feature proximity, similarity, continuity, closure, figure ground, and common fate

(Driver and Baylis, 1989; Ben-Av et al., 1992). Similarly, there are rules that govern the arrangement of low-level stimuli into haptic and auditory objects. For the latter the process is called auditory scene analysis (ASA). Contrary to the spatial principles that guide visual categorization, grouping in ASA is at either a temporal or melodic level governed by proximity or similarity over time, pitch or loudness continuation, or at spectral levels including common fate, coherent changes in loudness, frequency, or harmony (Bregman, 1994).

While principles of ASA, such as frequency and harmony, may seem relatively unimportant to visual perception they hold relevance for rehabilitation techniques for the substitution of vision for the visually impaired (Proulx et al., 2008; Brown et al., 2011). Researchers have long strived to provide crucial visual information with compensatory techniques via alternate modalities such as touch – Braille, embossed maps, tactile sensory substitution – (Bach-y-Rita and Kercel, 2003; Rowell and Ungar, 2003; Jiménez et al., 2009) or more recently sound – auditory sensory substitution and auditory workspaces – (Frauenberger and Stockman, 2009; Abboud et al., 2014; MacDonald and Stockman, 2014). The conversion principles of sonification algorithms are not arbitrary but instead based on natural cross-modal correspondences and cross-modal plasticity (Frasnelli et al., 2011; Spence, 2011) which allow the coding of visual features (brightness, spatial location) into auditory ones (pitch, loudness, stereo pan). Sensory substitution devices go beyond simple feature detection, and are also effective in ‘visual’ tasks such as object recognition and localisation, and navigation (Auvray et al., 2007; Brown et al., 2011; Maidenbaum et al., 2013). Given that the substitution of vision by other sensory modalities can evoke activity in visual cortex (Renier et al., 2005; Amedi et al., 2007; Collignon et al., 2007), it is unclear whether the mechanisms of scene analysis are processed as visual objects or auditory objects. Is the grouping of feature-based sensory representations into auditory objects based on visual grouping principles or those of ASA?

It seems natural that if the signal is a sonification it would be processed as an auditory feature and therefore be subjected to grouping principles of ASA. However, with extensive research showing activation of ‘visual’ areas in response to ‘auditory’ stimulation (Amedi et al., 2007; Striem-Amit and Amedi, 2014) and visually impaired users defining information from sonifications as ‘visual’ (Ward and Meijer, 2010) it is important to ascertain whether or not the auditory characteristics are more salient to the final perception using sonifications rather than a straight extrapolation from the unimodal literature. There are certainly valid comparisons between the two modalities. For example, shape and contour are crucial for the organization and recognition of visual objects. In parallel the spectral and temporal contour of a sound, the envelope, is critical in recognizing and organizing auditory objects (Sharpee et al., 2011).

However, there are also critical differences. The output signal of the sonification algorithm is dependent of the visual properties of the stimulus and therefore can be a coarse representation relative to a controlled audio-only presentation. For example, the sonification of equal-width visual lines will have different frequency bandwidths dependent on the stimulus baseline on

an exponential frequency scale – higher frequency baselines sonify to broader bandwidths, comprise of more sine waves, and are thus more complex than the sonification of an identical line lower down in the visual image. Thus, while the two pieces of visual information are perceived as having equivalent levels of complexity, there is variance between the complexities of the subsequent sonifications. Considering the purpose of sonifications is to convey visual information can we directly apply the principles of ASA, tested using auditory objects, to this?

If using the analog of two visual lines, equal in length (x-axis) but differing in elevation (y-axis), as two sonifications equal in duration (x-axis) but varying in baseline frequency (y-axis), we can apply ASA to make predictions on the mechanisms of feature segregation. Presented sequentially, with no requirement of identification (the two tones are separated in time), just noticeable differences (JND) in pitch should demonstrate low discrimination thresholds, typically between 1 and 190 Hz dependent on baseline frequency (Shower and Biddulph, 1931; Wever and Wedell, 1941). Presented concurrently, discrimination requires the identification of each tone based on the relative frequency components of each object. Considering this is one of the fundamental properties of the ear, the literature on this is scant. Thurlow and Bernstein (1957) reported two-tone discrimination at around 5% of the baseline frequency (at 4 kHz), while Plomp (1967), when assessing the ability to hear a harmonic in a harmonic complex, showed harmonic resolvability for five to seven lower harmonics. Plomp and Levelt (1965) evaluated explanations of consonance, that is the sensory experience of tonal fusion associated with isolated pairs of tones sharing simple frequency ratios, based on; frequency ratio, harmonic relationships, beats between harmonics, difference tones, and fusion. They concluded that the difference between consonant and dissonant intervals was related to the beats of adjacent partials, and that the transition range between these types of intervals were related to a critical bandwidth.

While this literature provides a solid grounding to predict results based on ASA it is important to note that in all these experiments the stimuli are generated as auditory objects, often with pure tones. This allows precision of the stimuli based on the exact auditory features you wish to test. For example, pure tones at specific frequencies can be used, or if testing the resolvability of harmonics complexes, tones with exact partials. Within the literature there appear to be no studies that contrast two-tone discrimination in which the precision of the stimuli is not controlled by auditory theory, as would be found when the signal is derived from visual features in a visual-to-auditory sonification. For example, with reference to the two line example above, would interval markers with varying complexity elicit similar results to what is found using controlled auditory stimuli? With this in mind we evaluated the segregation of two ‘auditory’ signals sonified from two equal length parallel lines at varying intervals. In a simple 2AFC paradigm the listener was required to indicate their perception of ‘one-ness’ or ‘two-ness’ in presented tonal complexes (Thurlow and Bernstein, 1957; Kleczkowski and Pluta, 2012). Based on the auditory literature we hypothesized that segregation of the two lines into separate objects would



be problematic when the sonifications had consonant harmonic relations.

In a second part of the experiment we used a multisensory paradigm to evaluate whether any influence in discrimination, due to ASA rules, could be negated by the provision of additional information in another modality. Our rationale and methodology were simple. Extensive research has demonstrated the efficacy of using multisensory, rather than uni-modal stimuli, with audio–visual information shown to enhance visual perception (Frassinetti et al., 2002) visual search (Iordanescu et al., 2008) and increase performance in spatial and temporal tasks. In speeded classification (SC) paradigms (Evans and Treisman, 2010) in which participants have to rapidly discriminate visual targets while presented with task irrelevant auditory stimuli, response times increase and accuracy decreases if the auditory stimulus is incongruous, i.e., high visual elevation paired with low pitch tone (Bernstein and Edelman, 1971; Marks, 1974; Ben-Artzi and Marks, 1995).

Crucial in multisensory integration is the binding of the unimodal stimuli into one perceived event based on: low-level spatial and temporal synchrony (Spence, 2011), temporal correlation (Radeau and Bertelson, 1987; Recanzone, 2003), or top down cognitive factors such as semantic congruency (Laurienti et al., 2004). For example, incongruent audio–visual spatial information shows a localisation bias toward visual information, in the ventriloquist effect, even when cued to the auditory stimulus (Bermant and Welch, 1976; Bertelson and Radeau, 1981) while separation of asynchronous audio–visual stimuli was perceived as shorter if presented in congruent rather than incongruent spatial locations (Soto-Faraco et al., 2002; Vroomen and de Gelder, 2003) with the auditory information appearing to dominate (Fendrich and Corballis, 2001; Soto-Faraco et al., 2004).

Considering this we manipulated the first task by providing either congruent multisensory stimuli, in which the sonification and visual presentation were associated (e.g., two-tone sonification and two visual lines) or incongruent (e.g., two-tone sonification and one visual line) to the listener. The task requirements were as before with the listener instructed to indicate how many visual lines had been sonified to create the stimulus. Based on the multisensory literature, we hypothesized that congruent audio–visual stimuli would facilitate superior performance in contrast to performance with both incongruent audio–visual and audio only stimuli.

## Materials and Methods

### Participants

We recruited 36 participants (28 female) via an Undergraduate Research Assistant module. Participant age ranged from 18 to 25 years old ( $M = 20.17$ ,  $SD = 1.30$ ). All participants provided informed written consent, and had normal or corrected eyesight, normal hearing and educated to undergraduate level. Four participants self-reported as left handed and all were naïve to the principles of sonification. 12 participants didn't return for the second part of the study and this is reflected in the analysis. The

study was approved by the University of Bath Psychology Ethics Committee (#13-204).

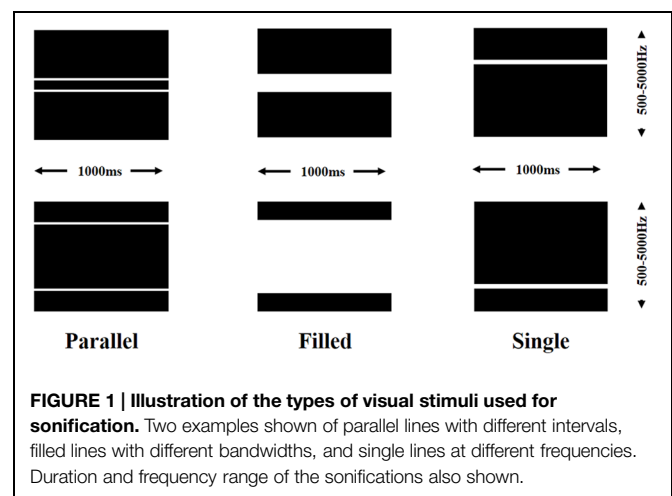
### Materials and Stimulus Design

Visual stimuli were created in Adobe Photoshop 3.0 with the sonifications using the principles of The vOICE (Meijer, 1992) algorithm. Frequency analysis of the sonifications was conducted in Cool Edit Pro 2.0 with all visual stimuli and sonifications presented in E-Prime 2.0 running on a Windows 7 PC. Sonifications were transmitted to the listener via Sennheiser HD 585 headphones. All statistical analysis was conducted using SPSS version 21.0.

### Stimulus Design

In Photoshop a grid of 48 pixel  $\times$  1.5 pixel rows was overlaid on a black background. Solid white lines were drawn over the full x-axis of the background with width and interval dependent on the stimulus type. Example of each type of line can be seen in **Figure 1**. For the parallel line stimuli two one-row lines, separated by the designated interval were created. The interval was varied from a two-row interval to a 42 row interval, with each interval gap increasing by two rows. The initial starting point was the center of the y-axis with each interval involving moving the top line up one row and the bottom line down 1 row from baseline or the previous stimulus. There were two types of single line stimuli. Filled stimuli took the parallel line stimuli and filled the gap between the two lines with white pixels. Thus the top and bottom lines were the same as the parallel line counterparts but with no interval between. The single line stimuli consisted of a line 2 rows thick (giving the same amount of white pixels as the parallel line). In total there were 23 parallel line, 24 single, and 24 filled stimulus images (two lines together at the central point of the y-axis was classified as a single line).

The lines were sonified using the following principles: the duration of each sonification, represented on the x-axis, was consistent for all stimuli (1000 ms), pitch was mapped to the y-axis with a range of 500 Hz (bottom) to 5000 Hz (top). White pixels were sonified at maximum volume ( $-65$  dB) with black pixels silent. Each sonification therefore comprised of two



complex tones at varying frequencies playing concurrently for 1000 ms (parallel lines), or one complex tone with the same top and bottom frequencies as the parallel line counterpart playing for 1000 ms (filled lines), or one complex tone at a consistent 'visual' width playing for 1000 ms (single line). Parallel line sonifications were categorized as consonant or dissonant based on the frequency range of the interval between the two lines.

## Procedure

Participants watched a PowerPoint presentation with audio-visual examples of the sonification process with a brief introduction to its applications. Example parallel lines, plus the two types of single lines with their sonifications were included as well as an example of the task procedure. For each trial of the main task the listener was presented with a soundscape which had been sonified from either 1 or 2 visual lines. Their task was to indicate on the PC keyboard whether the sonification was of 1 or 2 lines. Participants were explicitly told in both the instructions and PowerPoint that a filled line was classed as a single line. There was no visual information or post-trial feedback given. Each experimental block consisted of 96 trials ( $48 (2 \times 24) \times$  parallel,  $24 \times$  filled,  $24 \times$  single) with trial order fully randomized within block and no repeated trials. There were four blocks in total, randomized across participants, to give 386 trials in total.

The audio-visual task had the same listener requirements as the audio-only task, that is, to indicate how many lines were used to create the sonification. For each trial the listener heard a soundscape sonified from one or two lines. At the same time an image of one or two white lines appeared on the PC monitor. The audio-visual presentation could either be congruent, where the number of lines matched over both modalities, or incongruent where there was a mismatch. The participants were informed that while it was a requisite to look at the screen for timing purposes they were not required to indicate how many visual lines they perceived, just the number of 'lines' in the soundscape. As with the audio-only task there was no feedback. Again there were 4 blocks of 96 randomized trials. Examples of the example trials in both conditions are shown in **Figure 2**.

## Results

Consider accuracy for the parallel line condition first. **Figure 3** displays accuracy for individual parallel line frequencies, and clearly illustrates that the size of the interval between lines affects accurate recognition [ $F(8.52, 298.04) = 21.937, p < 0.0005, \eta_p^2 = 0.385$ ]. It is also clear that this cannot be solely due to proximity as some proximal lines (e.g., 498 Hz) are discriminated better than more distal lines (e.g., 3111 Hz), indicating that the predicted harmonic grouping is the relevant factor. **Figure 3** also displays the pattern for consonant (<50%) and dissonant (>50%) stimuli which matches the predictions from the categorization based on consonance and dissonance. Analysis of variance on these seven groups, as shown in **Figure 4**, again showed a main omnibus effect [ $F(3.19, 111.52) = 42.182, p < 0.0001, \eta_p^2 = 0.547$ ].

With harmonicity appearing the main factor in parallel line discrimination all relevant conditions were analyzed together:

audio-only consonant, audio-only dissonant, audio-visual consonant congruent, audio-visual consonant incongruent, audio-visual dissonant congruent, and audio-visual dissonant incongruent. Results are shown in **Figure 5** and **Table 1**. With accuracy as the D.V., an ANOVA, Greenhouse-Geisser corrected for violation of sphericity ( $\epsilon = 0.588$ ), showed an omnibus main effect [ $F(2.94, 64.69) = 19.162, p < 0.000, \eta_p^2 = 0.466$ ] again displaying that, when factoring in audio-visual conditions, the size of the interval between parallel lines is influential in line discrimination. To assess where these differences lay planned contrasts, Bonferroni corrected for multiple comparisons, were conducted.

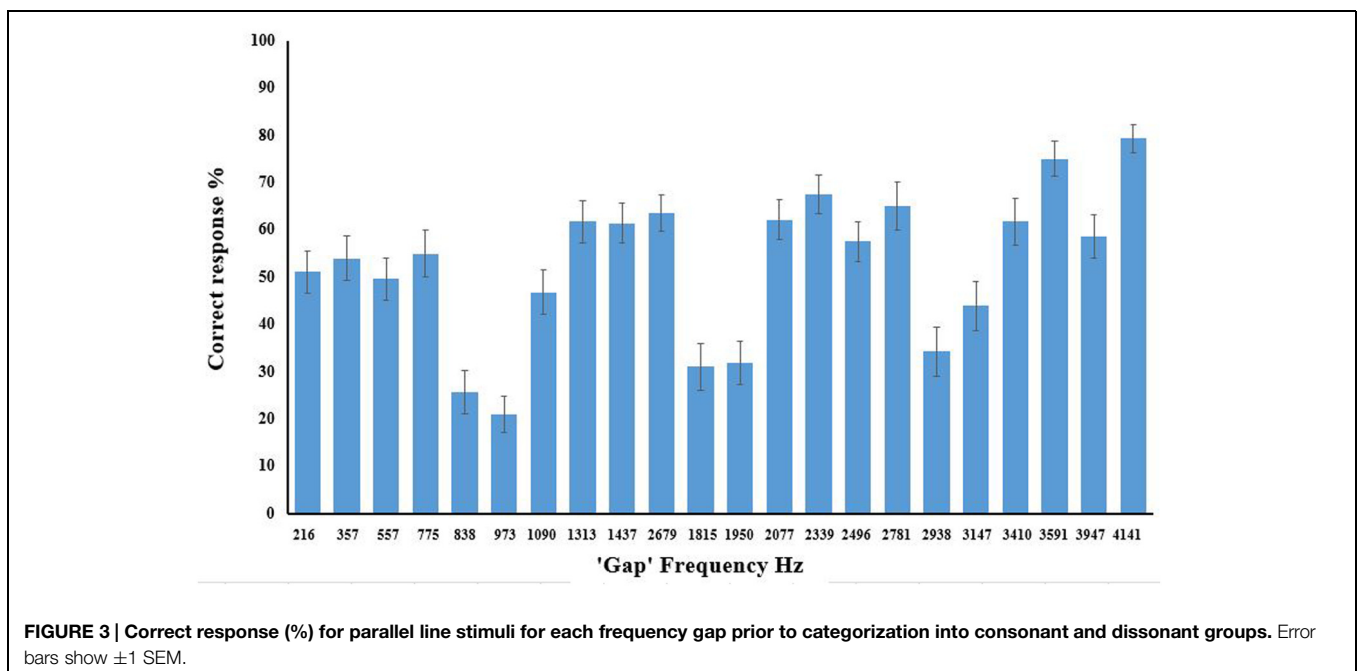
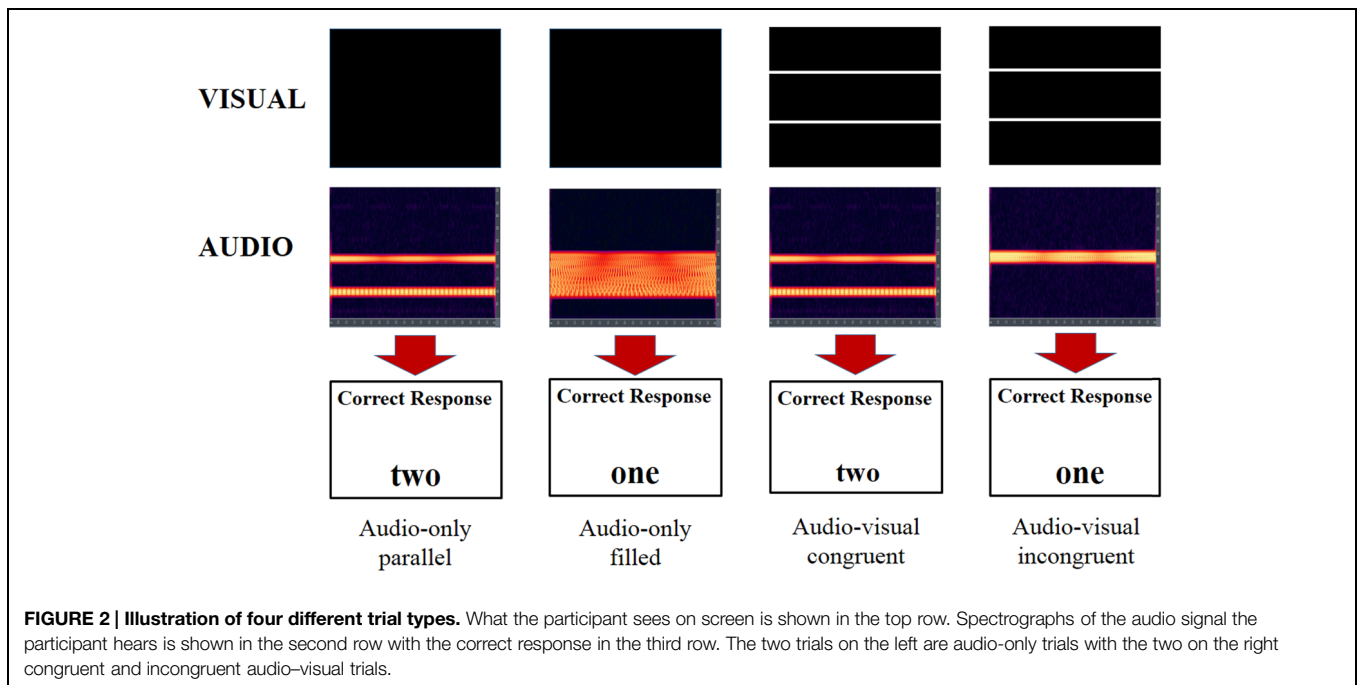
For trials where the stimuli were audio-only harmonicity had a large impact. Dissonant stimuli ( $M = 59.48$ ), where the interval should not elicit any tonal confusion, were discriminated more successfully than consonant stimuli ( $M = 30.73$ ) where harmonic relations should impact on performance [ $MD = 27.525, 95\% CI(15.84, 39.21), p < 0.0005$ ]. The latter were also significantly below what would be expected by chance [ $t(35) = -5.058, p < 0.0005, d = 1.34$ ] illustrating the magnitude of the 'confusion' caused by these harmonic relations.

Could this effect be anyway negated by using multisensory stimuli providing additional visual information? With the literature implying that multisensory binding requires some form of synchronicity we would only expect improved performance for audio-visual trials that were congruent, that is, provide the same line information via different modalities. The contrasts for the consonant stimuli showed no evidence of increased performance due to either congruent ( $M = 42.75$ ) or incongruent ( $M = 32.79$ ) audio-visual stimuli with significance levels of  $p = 0.797$  and  $p = 0.984$ , respectively.

For dissonant stimuli, where performance in the audio-only condition was already significantly above chance [ $t(35) = 2.912, p = 0.006, d = 3.04$ ] with no issues of harmonic relations we would expect an improvement in performance congruent trials in the audio-visual conditions. While the contrasts showed higher mean accuracy for the congruent condition ( $M = 70.58$ ) and a lower one for the incongruent ( $M = 54.55$ ), compared to the audio-only ( $M = 59.95$ ) neither differences were significant with  $p$ -values of 0.445 and 0.984, respectively.

Secondly we considered whether proximity was an influence on discrimination of parallel lines, that is, would sonified lines closer together be less likely to be segregated into separate objects? Looking at the seven groups categorized by the frequency ranges shown in **Figure 4**, we only contrasted within groups, that is, consonant versus consonant and dissonant versus dissonant. With the harmonicity effect having such a profound effect on performance comparisons between consonant and dissonant groups would naturally show a significant effect with the variance explained by these harmonic relations.

With accuracy as the dependent variable an ANOVA factoring in all consonant groups (audio-only, audio-visual congruent and audio-visual incongruent) showed an omnibus main effect for proximity [ $F(8, 176) = 3.528, p = 0.001, \eta_p^2 = 0.138$ ] with a separate ANOVA for dissonant groups showing similar [ $F(11, 242) = 5.335, p = 0.001, \eta_p^2 = 0.195$ ]. The Bonferroni corrected planned contrasts for both analyses tell a similar

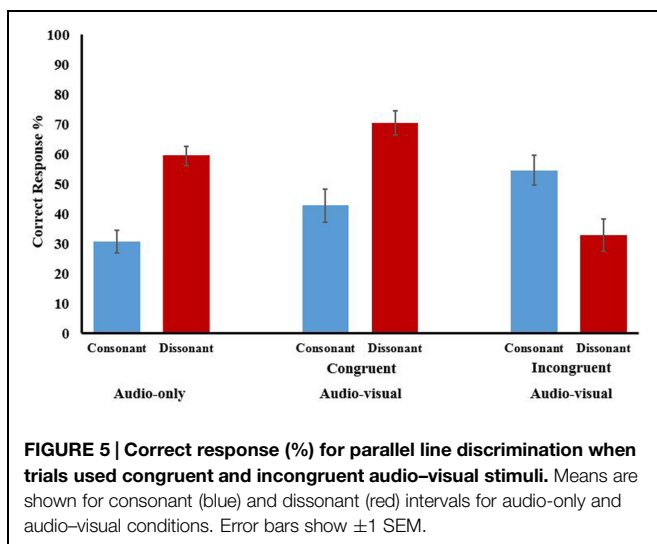
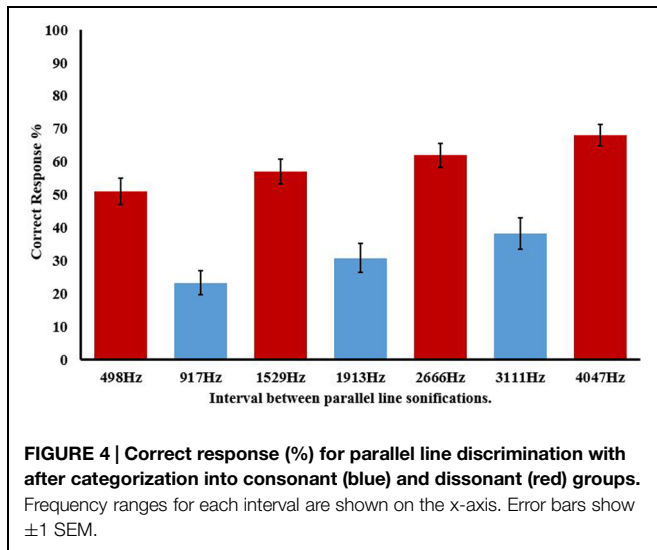


story. The only significant planned contrasts were between the congruent and incongruent audio-visual categories. For example, for consonant trials disregarding harmonicity, discrimination in the largest congruent category was better than for the two smallest incongruent categories ( $p = 0.008$ ) and ( $p = 0.013$ ), respectively. Dissonant trials in the smallest congruent group were better than for the smallest ( $p = 0.002$ ) and second smallest ( $p = 0.018$ ) incongruent groups. The second largest congruent elicited better scores than all four incongruent groups (smallest-to-largest,  $p = 0.026$ ,  $p = 0.001$ ,  $p = 0.008$ ,  $p = 0.001$ ), with

the largest congruent group better than the smallest ( $p = 0.009$ ) and largest ( $p = 0.009$ ) incongruent. There were no significant contrasts within groups or involving the audio-only trials.

Analysis of the filled line data corroborates the lack of any effect of proximity. These lines retained the same top and bottom frequencies as the parallel lines but with the interval filled with white pixels/sonified noise. Without the intervals there can be no effect of harmonicity and therefore any differences are due to proximity or signal bandwidth. With all groups (7  $\times$  audio-only, 7  $\times$  audio-visual congruent,





**TABLE 1 | Correct response (%), for parallel line discrimination for consonant and dissonant stimuli in; audio-only, congruent audio-visual, and incongruent audio-visual conditions.**

	Consonant		Dissonant	
	Mean	SD	Mean	SD
Audio-only	30.73	22.86	59.48	19.54
<b>Audio-visual</b>				
Congruent	42.75	26.33	70.58	19.33
Incongruent	32.79	25.85	54.55	24.16

$7 \times$  audio-visual incongruent) entered into an ANOVA there was a significant omnibus main effect [ $F(20,360) = 3.401$ ,  $p < 0.0005$ ,  $\eta_p^2 = 0.159$ ]. However, while there were 17 significant contrasts at an alpha of  $< 0.05$  these were all between audio-visual congruent (good) and incongruent groups (poor) with no differences within groups or involving the audio-only condition.

In summary. When presented with audio-only stimuli where the interval had no harmonic relations the task was relatively easy with participants scoring above chance. However, when the interval does have harmonic relations, signified by tonal fusion, the negative impact of this makes the task difficult with participants below chance levels. The use of audio-visual stimuli has little impact on lessening the effect of harmonicity and even when this effect is discounted, i.e., dissonant stimuli only, the congruent trials show a trend of better discrimination, but not reaching significance. Secondly, there is little evidence that proximity influences the discrimination of the sonifications with the only effects in this analysis being down to the use of congruent and incongruent audio-visual stimuli.

## Discussion

In this study we evaluated whether feature segregation of sonified horizontal lines would be influenced by rules of ASA. Unlike simple stimuli used in auditory research, the sonifications here were complex, with wider interval marker bandwidths dictated by the visual features of the stimulus interacting with the principles of the sonification algorithm. However, even with this coarse representation, sonifications with consonant intervals demonstrated poor segregation as predicted by ASA. Secondly we assessed whether the provision of additional multisensory information would negate the effects of harmonicity. While congruent audio-visual information displayed a trend for superior feature segregation, relative to incongruent audio-visual and audio-only, this only reached significance for the former contrast.

The results fall broadly in line with what is predicted in the auditory literature (Plomp and Levelt, 1965; Bregman et al., 1990; Bregman, 1994) demonstrating the negative impact of consonance on feature segregation. Even when visual lines were almost the full height (y-axis) of the workspace apart, with associated sonifications separated by  $> 3100$  hz, harmonic relations elicited the perception of one object. While these findings are not too surprising they do emphasize the robustness of the effect to interval markers of varying complexity. The logarithmic frequency conversion of the algorithm renders visual lines of equal width as sonifications whose bandwidths are dependent on their elevation in the visual field. For example, in our study the frequency bandwidth of a two-pixel wide line at the top of the screen was over 800 hz greater than the equivalent line at the bottom of the screen. Within the somewhat sparse simultaneous two-tone discrimination literature in the auditory domain, in which visual factors are not applicable, this interval marker bandwidth variability is not assessed as stimuli parameters can be more controlled. Of course it would be interesting to evaluate how much variance between the two markers, in bandwidth and other features, would be required to reduce the consonance effect. There is certainly evidence that two-tone complexes are more easily resolved if the amplitude of one of the tones is more intense (Arehart and Rosengard, 1999) and this could have been evaluated in the present

experiment by manipulating the shading of one of the visual lines.

Using The vOICE algorithm for the visual-to-auditory conversion necessitates a signal that is not static in the stereo field over time, that is, the signal initiates in the left headphone and pans across the stereo field to the right headphone over the duration of the scan. In a simultaneous two-tone pitch discrimination task Thurlow and Bernstein (1957) compared conditions where either the two tones were presented to the same ear (analogous to the present study), or presented to separate ears. Results showed little difference in discrimination for the five tested frequency levels when led to separate ears, however, when led to the same ear equivalent performance was only for stimuli where masking effects were minimized. If The vOICE signal was led to separate ears with the low frequency line scanning right-to-left and the high frequency line left-to-right, would this negate the masking effects demonstrated in the study? It is certainly a consideration for future research.

Simultaneous two-tone discrimination has been evaluated in different users to assess individual and group differences. An obvious group to test is trained musicians as successful pitch discrimination is an essential tool in their skillset. Kleczkowski and Pluta (2012) demonstrated that trained musicians were able to discriminate pitches at narrower levels than non-musicians, with similar results for musicians resolving harmonics in inharmonic complexes (Plomp, 1976). Musicians have also shown higher levels of performance using sensory substitution devices with Haigh et al. (2013) reporting musical ability correlating with higher acuity in a task using the vOICE and the Snellen Tumbling 'E'. All participants in the study were sighted and naïve to sensory substitution and yet demonstrated acuity approaching the legal blindness definition of 20/200. In a similar acuity test with blind participants trained to use the device even lower acuity was reported (Striem-Amit et al., 2012) illustrating not only the effect of training but also potentialities due to superior auditory abilities, such as frequency discrimination (Roder et al., 1999; Wan et al., 2010), posited to be found in these populations. It would therefore be of great interest to test whether highly trained blind users of The vOICE could overcome the effect of consonance found in the present study. If so, this psychophysical test will provide solid evidence whether, through perceptual learning, the user is truly 'seeing' the sound or just hearing it. Considering the strength of consonance reported, it is highly doubtful that the effect would be negated in auditory domain and thus any difference in performance in these populations would imply a percept beyond audition.

The strength of the consonance effect is further exemplified by the limited influence of congruent and incongruent visual information. In speeded classification tasks evaluating cross-modal congruency, classification of visual stimuli as 'high' or 'low' has been shown to be more rapid if accompanied by tones that were congruent rather than incongruent (Bernstein and Edelstein, 1971; Ben-Artzi and Marks, 1995) with Evans and Treisman (2010) showing that cross-modal mappings between audio and visual stimuli are automatic and affect performance even when irrelevant to the task. This integration

of temporally synchronous multisensory information is weighted to specific modalities as a function of the task (Spence, 2011), drawing support from a metamodal theory of the brain organization (Pascual-Leone and Hamilton, 2001). Here the brain is viewed as a task based machine with brain areas that are functionally optimal for particular computations; auditory areas for temporal tasks and visual for spatial (Proulx et al., 2014). In the present study the discrimination task can be considered spatial as the temporal features of the stimuli were identical. True to the metamodal theory, this adds weight to the visual information. If the audio-visual stimuli were congruent this should elicit better performance, and while the data showed a trend for this, it was not strong enough to bring discrimination of consonant stimuli above chance levels. Conversely, the incongruent visual information should reduce performance as there is extra weight attributed to the irrelevant distractor but again this trend was non-significant. Naturally with no access to visual information the blind users would not experience this audio-visual congruence, however, this could be tested using congruent and incongruent tactile stimuli. Simple methods such as embossed print outs of the visual workspace, or more technological based techniques involving haptic displays could be utilized to give multisensory information.

The results of our experiment show that the influence of consonance on object segregation is applicable to the sonification of coarse visual objects, but how can this information be suitably utilized? One approach to sonify a visual computer workspace is to evaluate the original visual stimulus and a spectrograph of it. Comparing these to the auditory representation would allow an evaluation of any potential auditory masking that might arise. This could include the direct mapping of spectrographs over the visual workspace in the development stage. Secondly, it would be interesting to evaluate how much consonance impacts on the use of sensory substitution devices when used in real-time. In such scenarios the sonified visual field updates at the device scan rate (1000 ms at default) to provide a continuous stream of 'static' frames. Thus, two parallel line sonifications masked in the first frame would only remain masked in the following frame if the device sensor, and background, remained static. For example, if the sensor was closer to the object in the second frame the parallel lines would be more disparate on the y-axis, the auditory interval increased, and the consonance negated.

A second consideration is variability and density of information provided in real-time device use. The present study utilized relatively simple stimuli, equal in all properties aside from auditory frequency, on a silent background. Objects encountered in everyday use are likely to be considerably more complex and therefore, even with masking, there should be sufficient unmasked signal to facilitate recognition. Indeed in a simple object recognition task using The vOICE, Brown et al. (2014) demonstrated equitable performance for degraded signals with limited information in contrast to more detailed stimuli.

Considering the above it seems unlikely that the negative effects of consonance would impact on real-time use of sensory substitution devices, although it should be considered

if using static objects in early training paradigms. Interestingly, however, reducing dissonance has already been applied to visual-to-auditory sensory substitution. The EyeMusic uses similar conversion principles to the vOICE as well as coding basic colors to musical instruments (Abboud et al., 2014). In an attempt to make device use less uncomfortable, a pentatonic scale, alongside a reduced frequency range, is used to reduce dissonance. This is logical considering dissonance in audition is associated with a harsh perceptual experience. However, as we have demonstrated in our simple object discrimination task, dissonance appears

important in feature segregation and it may be worth evaluating if there would be a comfort-function trade off in such tasks using EyeMusic.

## Acknowledgment

This work was supported in part by a grant from the EPSRC to MP (EP/J017205/1) and the EPSRC Doctoral Training Account studentship at Queen Mary University of London to AS.

## References

- Abboud, S., Hanassy, S., Levy-Tzedek, S., Maidenbaum, S., and Amedi, A. (2014). EyeMusic: Introducing a “visual” colorful experience for the blind using auditory sensory substitution. *Restor. Neurol. Neurosci.* 32, 247–257. doi: 10.3233/RNN-130338
- Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat. Neurosci.* 10, 687–689. doi: 10.1038/nn1912
- Arehart, K. H., and Rosengard, P. F. (1999). Psychometric functions for discrimination of two-component complex tones in listeners with normal hearing and listeners with hearing loss. *J. Acoust. Soc. Am.* 105, 2454–2465. doi: 10.1121/1.426850
- Auvray, M., Hanneton, S., and O'Regan, J. K. (2007). Learning to perceive with a visuo-auditory substitution system: localisation and object recognition with ‘the vOICE.’ *Perception* 36, 416–430. doi: 10.1068/p5631
- Bach-y-Rita, P., and Kercel, S. W. (2003). Sensory substitution and the human-machine interface. *Trends Cogn. Sci.* 7, 541–546. doi: 10.1016/j.tics.2003.10.013
- Ben-Artzi, E., and Marks, L. E. (1995). Visual-auditory interaction in speeded classification: role of stimulus difference. *Percept. Psychophys.* 57, 1151–1162. doi: 10.3758/BF03208371
- Ben-Av, M. B., Sagi, D., and Braun, J. (1992). Visual attention and perceptual grouping. *Percept. Psychophys.* 52, 277–294. doi: 10.3758/BF03209145
- Bermant, R. I., and Welch, R. B. (1976). Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition. *Percept. Mot. Skills* 42, 487–493. doi: 10.2466/pms.1976.42.2.487
- Bernstein, I. H., and Edelman, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *J. Exp. Psychol.* 87, 241–247. doi: 10.1037/h0030524
- Bertelson, P., and Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept. Psychophys.* 29, 578–584. doi: 10.3758/BF03207374
- Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT press.
- Bregman, A. S., Liao, C., and Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Can. J. Psychol.* 44, 400–413. doi: 10.1037/h0084255
- Brown, D. J., Macpherson, T., and Ward, J. (2011). Seeing with sound? Exploring different characteristics of a visual-to-auditory sensory substitution device. *Perception* 40, 1120–1135. doi: 10.1068/p6952
- Brown, D. J., Simpson, A. J. R., and Proulx, M. J. (2014). Visual objects in the auditory system in sensory substitution: how much information do we need? *Multisens. Res.* 27, 337–357. doi: 10.1163/22134808-00002462
- Cherry, C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229
- Collignon, O., Lassonde, M., Lepore, F., Bastien, D., and Veraart, C. (2007). Functional cerebral reorganization for auditory spatial processing and auditory substitution of vision in early blind subjects. *Cereb. Cortex* 17, 457–465. doi: 10.1093/cercor/bhj162
- Driver, J., and Baylis, G. C. (1989). Movement and visual attention: the spotlight metaphor breaks down. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 448–456. doi: 10.1037/h0090403
- Evans, K. K., and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *J. Vis.* 10, 1–12. doi: 10.1167/10.1.6
- Fendrich, R., and Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Percept. Psychophys.* 63, 719–725. doi: 10.3758/BF03194432
- Frasnelli, J., Collignon, O., Voss, P., and Lepore, F. (2011). “Crossmodal plasticity in sensory loss,” in *Progress in Brain Research, Part 1 Enhancing Performance for Action and Perception: Multisensory Integration, Neuroplasticity and Neuroprosthetics*, eds A. M. Green, C. E. Chapman, J. F. Kalaska, and F. Lepore (Amsterdam: Elsevier), 233.
- Frasnelli, F., Bolognini, N., and Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp. Brain Res.* 147, 332–343. doi: 10.1007/s00221-002-1262-y
- Frauenberger, C., and Stockman, T. (2009). Auditory display design—an investigation of a design pattern approach. *Int. J. Hum. Comput. Stud.* 67, 907–922. doi: 10.1016/j.ijhcs.2009.05.008
- Haigh, A., Brown, D. J., Meijer, P., and Proulx, M. J. (2013). How well do you see what you hear? The acuity of visual-to-auditory sensory substitution. *Front. Psychol.* 4:330. doi: 10.3389/fpsyg.2013.00330
- Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., and Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychon. Bull. Rev.* 15, 548–554. doi: 10.3758/PBR.15.3.548
- Jiménez, J., Olea, J., Torres, J., Alonso, I., Harder, D., and Fischer, K. (2009). Biography of louis braille and invention of the braille alphabet. *Surv. Ophthalmol.* 54, 142–149. doi: 10.1016/j.survophthal.2008.10.006
- Kleczkowski, P., and Pluta, M. (2012). Frequency discrimination in a simultaneous two-tone signal. *Acta Phys. Pol. Ser. A Gen. Phys.* 121, A120.
- Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., and Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Exp. Brain Res.* 158, 405–414. doi: 10.1007/s00221-004-1913-2
- MacDonald, D., and Stockman, T. (2014). *Towards a Method, Techniques and Tools to Support the Development of Auditory Displays. HCI Engineering Charting the Way towards Methods and Tools for Advanced Interactive Systems*, 29. Available at: [http://hci-engineering.di.uminho.pt/Program\\_files/MacDonald.pdf](http://hci-engineering.di.uminho.pt/Program_files/MacDonald.pdf)
- Maidenbaum, S., Levy-Tzedek, S., Chebat, D. R., and Amedi, A. (2013). Increasing accessibility to the blind of virtual environments, using a virtual mobility aid based on the “EyeCane”: feasibility study. *PLoS ONE* 8:e72555. doi: 10.1371/journal.pone.0072555
- Marks, L. E. (1974). On associations of light and sound: the mediation of brightness, pitch, and loudness. *Am. J. Psychol.* 87, 173–188. doi: 10.2307/1422011
- Meijer, P. (1992). An experimental system for auditory image representations. *IEEE Trans. Biomed. Eng.* 39, 112–121. doi: 10.1109/10.121642
- Pascual-Leone, A., and Hamilton, R. (2001). The metamodal organization of the brain. *Prog. Brain Res.* 134, 427–445. doi: 10.1016/S0079-6123(01)34028-1
- Plomp, R. (1967). Pitch of complex tones. *J. Acoust. Soc. Am.* 41, 1526–1533. doi: 10.1121/1.1910515
- Plomp, R. (1976). *Aspects of Tone Perception: A Psychophysical Study*. New York: Academic Press.
- Plomp, R., and Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* 38, 548–560. doi: 10.1121/1.1909741
- Proulx, M. J., Brown, D. J., Pasqualotto, A., and Meijer, P. (2014). Multisensory perceptual learning and sensory substitution. *Neurosci. Biobehav. Rev.* 41, 16–25. doi: 10.1016/j.neubiorev.2012.11.017

- Proulx, M. J., Stoerig, P., Ludowig, E., and Knoll, I. (2008). Seeing 'Where' through the Ears: effects of learning-by-doing and long-term sensory deprivation on localization based on image-to-sound substitution. *PLoS ONE* 3:e1840. doi: 10.1371/journal.pone.0001840
- Radeau, M., and Bertelson, P. (1987). Auditory-visual interaction and the timing of inputs. Thomas (1941) revisited. *Psychol. Res.* 49, 17–22.
- Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *J. Neurophysiol.* 89, 1078–1093. doi: 10.1152/jn.00706.2002
- Renier, L., Collignon, O., Poirier, C., Tranduy, D., Vanlierde, A., Bol, A., et al. (2005). Cross-modal activation of visual cortex during depth perception using auditory substitution of vision. *Neuroimage* 26, 573–580. doi: 10.1016/j.neuroimage.2005.01.047
- Roder, B., Teder-Salejarvi, W., Sterr, A., Rosler, F., Hillyard, S. A., and Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature* 400, 162–166. doi: 10.1038/22106
- Rowell, J., and Ungar, S. (2003). The world of touch: an international survey of tactile maps. Part 1: production. *Br. J. Vis. Impair.* 21, 98–104.
- Sharpee, T. O., Atencio, C. A., and Schreiner, C. E. (2011). Hierarchical representations in the auditory cortex. *Curr. Opin. Neurobiol.* 21, 761–767. doi: 10.1016/j.conb.2011.05.027
- Showers, E., and Biddulph, R. (1931). Differential pitch sensitivity of the ear. *J. Acoust. Soc. Am.* 3, 7–7. doi: 10.1121/1.1915561
- Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., and Kingstone, A. (2002). The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. *Brain Res. Cogn. Brain Res.* 14, 139–146. doi: 10.1016/S0926-6410(02)00068-X
- Soto-Faraco, S., Spence, C., and Kingstone, A. (2004). Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 330–345. doi: 10.1037/0096-1523.30.2.330
- Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Atten. Percept. Psychophys.* 73, 971–995. doi: 10.3758/s13414-010-0073-7
- Striem-Amit, E., and Amedi, A. (2014). Visual cortex extrastriate body-selective area activation in congenitally blind people “seeing” by using sounds. *Curr. Biol.* 24, 687–692. doi: 10.1016/j.cub.2014.02.010
- Striem-Amit, E., Guendelman, M., and Amedi, A. (2012). ‘Visual’ acuity of the congenitally blind using visual-to-auditory sensory substitution. *PLoS ONE* 7:e33136. doi: 10.1371/journal.pone.0033136
- Thurlow, W. R., and Bernstein, S. (1957). Simultaneous two-tone pitch discrimination. *J. Acoust. Soc. Am.* 29, 515–519. doi: 10.1121/1.1908946
- Vroomen, J., and de Gelder, B. (2003). Visual motion influences the contingent auditory motion aftereffect. *Psychol. Sci.* 14, 357–361. doi: 10.1111/1467-9280.24431
- Wan, C. Y., Wood, A. G., Reutens, D. C., and Wilson, S. J. (2010). Early but not late-blindness leads to enhanced auditory perception. *Neuropsychologia* 48, 344–348. doi: 10.1016/j.neuropsychologia.2009.08.016
- Ward, J., and Meijer, P. (2010). Visual experiences in the blind induced by an auditory sensory substitution device. *Conscious. Cogn.* 19, 492–500. doi: 10.1016/j.concog.2009.10.006
- Wever, E., and Wedell, C. (1941). Pitch discrimination at high frequencies. *Psychol. Bull.* 38, 727.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Brown, Simpson and Proulx. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Semantic-based crossmodal processing during visual suppression

Dustin Cox and Sang Wook Hong\*

Department of Psychology, College of Science, Florida Atlantic University, Boca Raton, FL, USA

To reveal the mechanisms underpinning the influence of auditory input on visual awareness, we examine, (1) whether purely semantic-based multisensory integration facilitates the access to visual awareness for familiar visual events, and (2) whether crossmodal semantic priming is the mechanism responsible for the semantic auditory influence on visual awareness. Using continuous flash suppression, we rendered dynamic and familiar visual events (e.g., a video clip of an approaching train) inaccessible to visual awareness. We manipulated the semantic auditory context of the videos by concurrently pairing them with a semantically matching soundtrack (congruent audiovisual condition), a semantically non-matching soundtrack (incongruent audiovisual condition), or with no soundtrack (neutral video-only condition). We found that participants identified the suppressed visual events significantly faster (an earlier breakup of suppression) in the congruent audiovisual condition compared to the incongruent audiovisual condition and video-only condition. However, this facilitatory influence of semantic auditory input was only observed when audiovisual stimulation co-occurred. Our results suggest that the enhanced visual processing with a semantically congruent auditory input occurs due to audiovisual crossmodal processing rather than semantic priming, which may occur even when visual information is not available to visual awareness.

**Keywords:** multisensory integration, semantic processing, continuous flash suppression, visual awareness, semantic priming

## OPEN ACCESS

### Edited by:

Magda L. Dumitru,  
Macquarie University, Australia

### Reviewed by:

Paula Goolkasian,  
University of North Carolina  
at Charlotte, USA  
Emily J. Ward,  
Yale University, USA

### \*Correspondence:

Sang Wook Hong,  
Department of Psychology, College  
of Science, Florida Atlantic University,  
777 Glades Road, BS 12, Room 101,  
Boca Raton, FL 33431, USA  
shong6@fau.edu

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 19 March 2015

**Accepted:** 14 May 2015

**Published:** 02 June 2015

### Citation:

Cox D and Hong SW (2015)  
Semantic-based crossmodal  
processing during visual suppression.  
*Front. Psychol.* 6:722.  
doi: 10.3389/fpsyg.2015.00722

## Introduction

The objects and events we encounter in everyday life are often experienced in multiple sensory modalities. Multisensory integration can enrich perceptual experience of objects and events by enhancing the saliency of stimuli (Stein and Stanford, 2008; Evans and Treisman, 2010). The advantages of multisensory integration have been evidenced by faster response times (RTs) in speeded classification tasks when auditory pitch and visual elevation are congruent with each other (Bernstein and Edelman, 1971; Ben-Artzi and Marks, 1995), improved visual motion perception with congruent auditory information (Cappé et al., 2009; Lewis and Noppeney, 2010), and enhanced speech perception with synchronous audiovisual inputs (Pandey et al., 1986; Plass et al., 2014).

Multisensory congruency generally indicates that multiple unimodal stimuli are present closely in space or time. Multisensory spatiotemporal congruency often results in enhancement of behavioral and perceptual performances (Stein et al., 1988). Auditory and visual stimuli that have spatial alignment can generate more efficient saccadic eye movements to the target in non-human primates (Bell et al., 2005). Human saccadic eye movements are also faster toward visual targets when auditory and visual stimuli have closer spatiotemporal proximity (Frens et al., 1995). The detection of unimodal objects and events can be enhanced by a spatially and/or temporally co-occurring stimulus

in another modality (Vroomen and de Gelder, 2000; Lovelace et al., 2003; Bolognini et al., 2005; Noesselt et al., 2008).

Semantic congruency is also considered to be an important factor that determines multisensory integration (Doehrmann and Naumer, 2008; Spence, 2011). Audiovisual crossmodal semantic congruency effects have been examined by testing whether behavioral performance is enhanced by pairing an auditory stimulus and a visual stimulus that match or mismatch in meaning, such as pairing the sound of a dog barking with an image of a dog or cat (Laurienti et al., 2003; Hein et al., 2007). Participants tend to be faster and/or more accurate when identifying visual stimuli paired with auditory stimuli that have a semantically congruent than incongruent relationship (Laurienti et al., 2003, 2004; Iordanescu et al., 2008; Schneider et al., 2008; Chen and Spence, 2010).

The semantic content of auditory information can also affect visual awareness. The auditory semantic context of sounds heard during viewing of bistable figures can influence the predominance of a given percept (Hsiao et al., 2012). When viewing different dichoptic images during binocular rivalry, the dominance duration of a visual stimulus paired with a semantically congruent sound is significantly longer than when the same stimulus is paired with a semantically incongruent sound (Chen et al., 2011). Considering that perceptual dominance during binocular rivalry is dependent on the relative strength of dichoptically presented stimuli (Levelt, 1965), a longer period of dominance for an audiovisually congruent stimulus suggests that multisensory integration can strengthen a visual stimulus, resulting in the prolonged predominance of the stimulus. The longer predominance of a visual stimulus paired with a semantically congruent sound during binocular rivalry, however, cannot indicate whether the congruent sound influences the strength of the visual stimulus while it is suppressed from visual awareness. Multisensory integration may only occur when the congruent visual stimulus is dominantly perceived, and thus congruent auditory input might only exert an influence on dominance durations when visual stimuli are consciously perceived.

The possibility of multisensory integration based on semantic congruency when visual stimuli are suppressed from visual awareness has been supported by recent studies using continuous flash suppression (CFS; see Tsuchiya and Koch, 2005). CFS is a modification of binocular rivalry, in that, dynamically changing, highly salient “noise” patterns presented to one eye can suppress a stimulus presented to the other eye from visual awareness for extended periods of time. The measurement of the time of the breakup of CFS can indicate the relative strength of visual stimuli to gain access to the visual awareness of observers (Stein et al., 2011).

The results of two recent studies demonstrate that congruent semantic auditory information in addition to temporal congruency can enhance the processing of *dynamic* visual stimuli, which are suppressed from visual awareness (Alsius and Munhall, 2013; Plass et al., 2014). A dynamic talking face suppressed from visual awareness by CFS can break suppression and reach visual awareness quicker when the original (matched) soundtrack accompanies the lip movements of the face compared to a mismatched soundtrack pair (Alsius and Munhall, 2013). In another study, a dynamic talking face presented during CFS can speed up the identification of a spoken target word if the lip-movements of the face

correspond synchronously (Plass et al., 2014). However, it is not clear whether this congruency effect on visual speech processing is mediated by purely semantic-based multisensory integration since the influence of audiovisual semantic congruency could not be separated from speech stimuli while fully controlling for audiovisual temporal synchrony during CFS.

In the current study, we examined whether *purely* semantic-based multisensory integration influenced access to visual awareness for familiar dynamic visual events while limiting spatiotemporal congruency. Using CFS, we measured participants' RTs to identify suppressed visual events when participants simultaneously heard soundtracks that were either semantically congruent or incongruent with the visual events. The audiovisual events, such as a moving racecar and an approaching train, were chosen because there is a lesser amount of specific congruent timing between their constituent auditory and visual event components. We specifically hypothesize that audiovisual crossmodal integration occurs even when visual stimuli are suppressed from visual awareness, and thus, semantically congruent audiovisual events will break up suppression and will be perceived earlier than incongruent events. In a control experiment, we tested whether the semantic congruency effect occurs due to crossmodal semantic priming by presenting the soundtracks prior to the visual events. In an additional control experiment, we tested our hypothesis further using static images with which any residual spatiotemporal crossmodal correspondences were removed.

## Experiment 1

To determine whether auditory semantic information can influence visual awareness of events, we measured the latencies for participants to identify one of three (3AFC task) familiar visual events with concurrent soundtracks that were initially suppressed by CFS. The soundtracks varied in their semantic relationships to the videos so that they matched (congruent audiovisual soundtrack condition), mismatched (incongruent audiovisual soundtrack condition), or were silent (neutral video-only condition). If semantic auditory contexts affect visual processing of dynamic events, which are suppressed from visual awareness, there should be a difference in the RT for participants to become aware of event videos as they break CFS across the different soundtrack conditions. We expected that visual event videos that were semantically congruent with a concurrently heard soundtrack would break up suppression relatively sooner than when soundtracks were incongruent or neutral as indicated by faster RTs in the congruent audiovisual soundtrack condition.

## Method

### Participants

Thirty-three (nine males) undergraduate students participated in Experiment 1 for course credit. The participants were naïve to the purpose of this study. All participants had normal or corrected-to-normal vision and normal hearing as indicated by self-report. All participants signed an informed consent form approved by the Florida Atlantic University Institutional Review Board before participating in this experiment.

## Apparatus and Stimuli

The visual stimuli were presented on a Sony CPD-G520, 21" CRT display (100 Hz frame rate). The presentation of stimuli and collection of response data was manipulated by the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) in Matlab (MathWorks). Visual stimuli were presented in a dark room to observers positioned 90 cm from the CRT monitor whose R, G, B guns were calibrated using a light meter (IL-1700) and a luminance meter (Minolta LS100), creating a linearized look-up-table (eight-bit for each R, G, and B guns). A four-mirror stereoscope was used to achieve dichoptic presentation of the visual stimuli characteristic of a binocular rivalry experiment. Auditory stimuli were presented using Acoustic Noise-Canceling headphones (Bose QuietComfort).

The visual stimuli used in Experiment 1 were three dynamic and familiar event video clips, one of which was presented to one eye of a participant in each trial. The three brief video clips (7 s in duration) were black and white and depicted an approaching train, a man playing guitar, or racecars circling a racetrack. The video clip stimuli were edited using iMovie. The video clips were presented within rectangular apertures ( $3.91^\circ \times 3.2^\circ$ ) created by black rectangular fusion contours ( $4.36^\circ \times 3.42^\circ$ ). Three audiovisual soundtrack conditions were tested. In the congruent audiovisual soundtrack condition, each video clip was presented with its original soundtrack (e.g., an approaching train with a train soundtrack). In the incongruent audiovisual soundtrack condition, each video clip was overdubbed with a soundtrack from one of the other two videos (e.g., an approaching train with a racecar soundtrack). In the neutral video-only condition, video clips were presented without any sound. The suppressors, dynamically changing Mondrian-like patterns, were presented to the other eye. Each suppressor was composed of 200 rectangular patches with random sizes. The luminance of each patch was randomly assigned, but within a predetermined range whose maximum and minimum values were used to compute the contrast of the suppressors. The mean luminance of the suppressors was fixed at  $55 \text{ cd/m}^2$ , which was identical to the luminance of the background. Sixty Mondrian-like patterns were created and presented every 100 msec (10 Hz).

Calibration of the stereoscope was achieved by participants' self-report of the vertical alignment of small nonius lines ( $0.04^\circ \times 0.22^\circ$ ) that extended from the center of the top and bottom of the inner edge of the rectangular image apertures presented to the left and right eye, respectively. The stereoscope was calibrated prior to the practice trial, and the calibration was checked again prior to the beginning of the experimental trial for each participant. Participants were also instructed to monitor the alignment of the nonius lines in between trials throughout the experiment and to inform the experimenter if they became misaligned.

## Procedure

Participants viewed a dichoptic presentation consisting of a dynamic Mondrian stimulus that was presented in one eye while the other eye was simultaneously presented with one of nine target event videos (three video conditions by three soundtrack conditions). The Mondrian stimuli served to initially suppress the

target video that was concurrently presented to the opposite eye from visual awareness. The eye that viewed the target video in each condition was considered the target eye. Each target event video condition (train, guitar, racecar), soundtrack condition (congruent audiovisual, incongruent audiovisual, neutral video-only), and target eye condition (left, right) was counterbalanced and randomly presented eight times to each participant for a total of 144 trials.

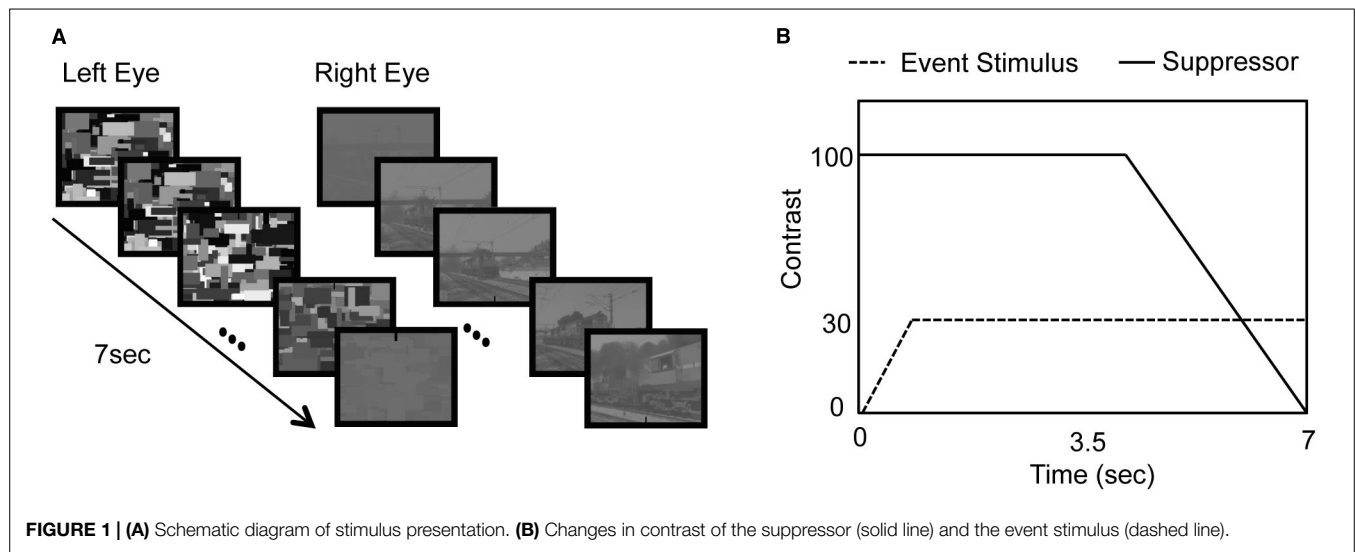
The relative luminance contrast in relation to the background for the Mondrian suppressors and target event videos was manipulated to ensure that the Mondrian stimuli achieved initial perceptual dominance followed by the breaking of suppression of the target video into perceptual dominance in each experimental trial (Yang et al., 2007). The target event videos were initially presented to one eye at 0% contrast before gradually increasing in contrast at equal increments over the first second of each experimental trial until reaching 30% contrast. In each trial, the Mondrian suppressor was initially presented at full (100%) contrast for the first 4 s before decreasing in contrast at equal decrements over the course of the remaining 3 s, so the Mondrian stimulus decreased in contrast to 0% by the end of the last 3 s of each 7-s trial duration (see **Figure 1**).

The participants' task was to report which target event video was viewed in each trial. Three response keys located at the numeric keypad portion on the right side of a computer keyboard were designated (participants were instructed to press the "1" key if they saw the train video, the "2" key if they saw the guitar video, or the "3" key if they saw the racecar video) prior to beginning the practice trials that were completed before the actual experiment. Participants were reminded again of the response key assignments as needed throughout the practice trials and once more prior to beginning the experimental trials of Experiment 1. The elapsed time from the moment of pressing the spacebar on the computer keyboard, which initiated each trial, until the moment of pressing the "1," "2," or "3" key on the keyboard was recorded as a RT. Participants were instructed to respond only when confident about identification of the video. Participants were also encouraged to not guess or respond based on the soundtracks they heard since the soundtracks would not always be informative for determining the correct response in the trials.

Participants were familiarized with the task during a run of practice trials that were identical to the experimental trials of Experiment 1 but consisted of only 12 repetitions. Successful practice trial performance was indicated once each participant demonstrated correct memorization of the response keys and was based on consistently correct responding as determined by the experimenter.

## Results and Discussion

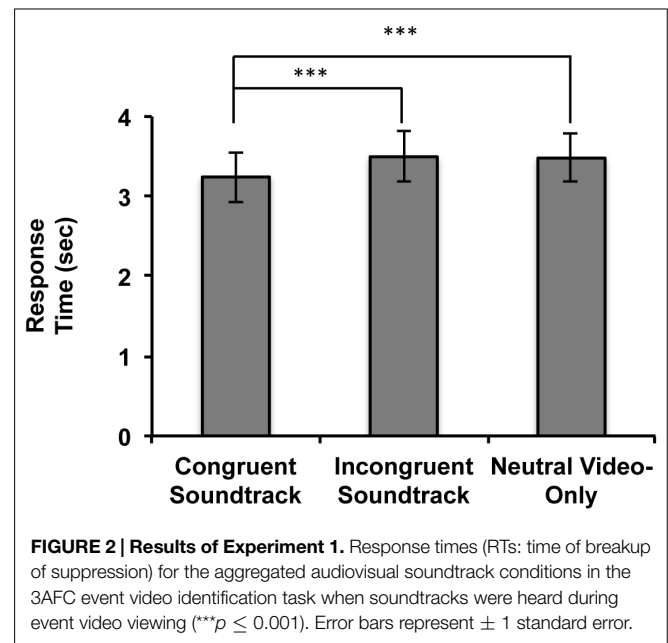
The data from 28 participants (seven males) were analyzed. We excluded five participants' data that had overall average error rates greater than or equal to chance level responding (i.e., chance responding rate on a 3AFC = 0.33). A trial in which the response did not correspond with the actual video presented was considered to be an error. Error rates greater than chance are potentially indicative of a lack of motivation and/or understanding of the task, or a tendency to guess when responding. To ensure that the mean



RT measurements were based only on correct responses, the RTs from incorrect response trials were excluded from the analysis. Trials where a video did not break up the suppression occurred when a key press response was not made during the 7-s duration of stimulus presentation. Since participants were encouraged to not guess the event video that was viewed, trials where no response was made were not considered to be incorrect, but they were also removed from the analysis.

A three by three (three event video conditions by three soundtrack conditions), two-way repeated measures ANOVA was conducted to examine the effect of the event viewed, the type of soundtrack heard, and the interaction between the event and soundtrack conditions on the mean RTs to discriminate the visual event videos. The analysis revealed a significant main effect of the event video viewed [ $F(2,26) = 10.058, p = 0.001, \eta^2 = 0.271$ ] and the type of soundtrack heard [ $F(2,26) = 10.263, p = 0.000, \eta^2 = 0.275$ ]. There was no significant interaction effect between the event and soundtrack conditions [ $F(4,24) = 0.808, p = 0.480, \eta^2 = 0.029$ ]. The significant main effect of the event video factor was not surprising since there were different amounts of luminance and motion information contained in the three videos. Differences in visual stimulus saliency may differentiate the time of the breakup of suppression. The lack of a significant interaction between sound and event conditions indicates a consistent effect of sound among the different events.

Since no significant interaction between sound and event conditions was found, we aggregated data based on the sound conditions from the three event conditions. We were more interested in examining the semantic influence of sound on visual event discrimination rather than the influence of differences in visual saliency of the three event videos. A one-way, repeated measures ANOVA with the aggregated data (**Figure 2**) revealed a significant main effect of audiovisual soundtrack condition [ $F(2,26) = 10.263, p = 0.001, \eta^2 = 0.275$ ]. Planned contrast tests revealed that the RTs were significantly faster when participants concurrently heard a semantically congruent soundtrack in comparison to hearing a semantically incongruent soundtrack [ $F(1,27) = 13.273, p = 0.001, \eta^2 = 0.330$ ] or no soundtrack



[ $F(1,27) = 12.710, p = 0.001, \eta^2 = 0.320$ ]. There was no significant difference between the RTs to discriminate the visual events when participants concurrently heard a semantically incongruent soundtrack in comparison to when no soundtrack was heard [ $F(1,27) = 0.106, p = 0.747, \eta^2 = 0.004$ ].

The results of Experiment 1 indicate that congruent auditory semantic information affects the time for dynamic visual events to gain access to visual awareness, and thus, suggest that semantic congruency-based audiovisual multisensory integration occurs while visual information is suppressed from visual awareness. The present results are consistent with a previous study showing that congruent semantic information contained within auditory soundtracks can enhance the perceptual dominance of dichoptically viewed images during binocular rivalry (Chen et al., 2011). These results suggest that the longer pre-dominance due to semantic congruency during binocular rivalry



(Chen et al., 2011) can result from a shortened suppression period due to multisensory information processing. The modulatory influence of auditory semantic context on unconscious visual processing further supports that purely semantic-based multisensory integration can happen regularly in everyday life situations.

## Experiment 2

What are the mechanisms that caused the semantic-based congruency effect observed in Experiment 1? Semantic priming is a plausible mechanism that can explain the early breakup of suppression for the congruent audiovisual events. Semantic priming can be observed when an enhancement of accuracy or reaction time in response to a target stimulus is due to the presentation of a semantically associated priming stimulus that precedes the presentation of a target stimulus (Dehaene et al., 1998; Costello et al., 2009). A target word suppressed from visual awareness by CFS breaks up suppression and is perceived earlier when a semantically congruent prime word is viewed prior to presentation of a target word with CFS, compared to when the prime word and target words are semantically incongruent (Costello et al., 2009). These results indicate that semantic congruency enhances the strength of a target stimulus and consequently the target breaks up suppression sooner. Recent studies suggest that crossmodal semantic priming of congruent naturalistic sounds presented prior to visual stimulus presentation can enhance visual sensitivity (Chen and Spence, 2011) and result in shorter reaction times to identify natural objects (Schneider et al., 2008).

Close temporal proximity of multiple unimodal sensory components has been shown to be important for multisensory integration (Meredith et al., 1987; van Atteveldt et al., 2007). We hypothesized that by presenting soundtracks prior to the discrimination of silent event videos, the potential influence of crossmodal semantic priming on participants' visual awareness of the events in Experiment 1 can be assessed while limiting the influence of concurrent multisensory integration. If the semantic congruency effect is abolished by the prior presentation of sound, this result indicates that the facilitatory effect of semantic congruency observed in Experiment 1 may be caused by a different mechanism than crossmodal semantic priming.

## Methods

Fifty-one undergraduate students participated in Experiment 2 that did not participate in Experiment 1. All apparatuses and stimuli were identical to those used in Experiment 1, except that the onset and offset of auditory soundtrack presentation immediately preceded the onset of dichoptic Mondrian and target video presentation. Auditory soundtrack presentation in Experiment 2 always lasted for 3 s to allow adequate time for semantic information to be accessed prior to performance of a 3AFC video discrimination task that was identical to that done by participants in Experiment 1. Following the initial soundtrack presentation, the event videos were always presented silently, so all discrimination trials of Experiment 2 resembled the silent audiovisual condition trials of Experiment 1.

## Results and Discussion

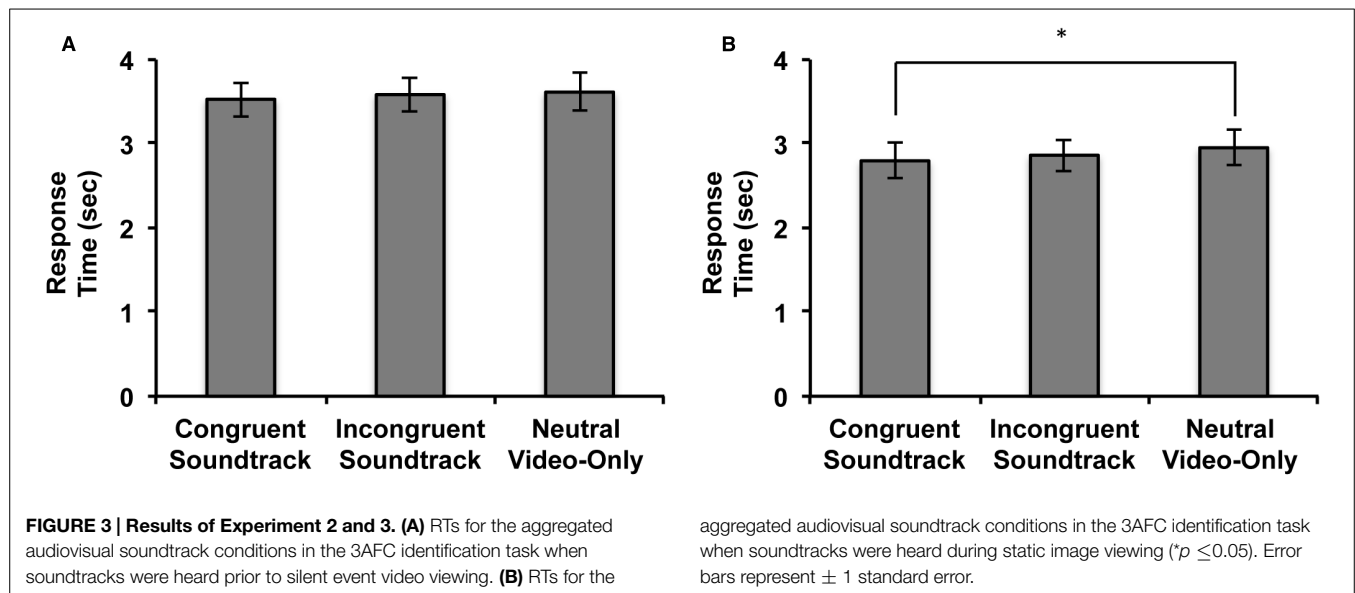
The data screening procedure based on individuals' error rate were identical to that used in Experiment 1. We excluded fourteen participants with greater than chance error rates leaving the data of 37 participants for analysis. A one-way repeated measures ANOVA conducted on the factor of audiovisual soundtrack condition, aggregated over three events, did not reveal a significant main effect on participants' overall RTs to discriminate the visual events [ $F(2,35) = 1.319, p = 0.274, \eta^2 = 0.035$ ]. This result indicates that when a soundtrack is played prior to the visual event, auditory semantic congruency has no significant influence on interocular suppression durations. However, despite the lack of significant differences, the overall average RTs in Experiment 2 when comparing the congruent, incongruent, and the neutral video-only audiovisual soundtrack conditions does resemble the one observed in Experiment 1 (Figure 3A). This tendency indicates that crossmodal semantic priming may partially contribute to the audiovisual semantic congruency effect observed in Experiment 1, but the temporal concurrence of auditory and visual stimulus presentation may be the factor that determines whether the multi-sensory integration of semantic information can occur.

To further assess the possibility that the results of Experiment 1 can be explained by semantic priming, a mixed design ANOVA was conducted on the aggregate data from Experiment 1 and Experiment 2 with the audiovisual soundtrack condition as a within-subjects factor and the temporal relationship between auditory and visual presentation (concurrent audiovisual presentation for Experiment 1, and auditory prior to visual presentation for Experiment 2) as a between-subjects factor. The mixed design ANOVA revealed a significant interaction between the audiovisual soundtrack condition and the temporal relationship of audiovisual presentation [ $F(2,63) = 3.200, p = 0.044, \eta^2 = 0.048$ ]. This result further supports that crossmodal semantic priming cannot completely account for the facilitatory effect of audiovisual semantic congruency observed in Experiment 1.

## Experiment 3

It is possible that spatiotemporal crossmodal correspondences could have influenced the results observed in the congruent audiovisual soundtrack condition of Experiment 1. For example, there is a close temporal alignment of the finger movements of the guitar player seen in the guitar event video that occurred synchronously with the sounds of the guitar being played. As mentioned before, audiovisual temporal synchrony can shorten interocular suppression durations for dynamic talking faces (Alsius and Munhall, 2013; Plass et al., 2014). Thus, observers could have been influenced by temporal synchrony cues when discriminating the guitar video in the congruent audiovisual soundtrack condition instead of being influenced only by semantically congruent multisensory information.

Looming or receding auditory signals, which respectively refer to increases or decreases in sound intensity (Ghazanfar and Maier, 2009), could have corresponded with the movement of objects seen in the event videos and influenced the results observed in Experiment 1. Multisensory integration of auditory and visual stimuli can enhance behavioral performance in humans



(Cappe et al., 2009). Looming and receding audiovisual correspondences could have been particularly relevant to the congruent audiovisual soundtrack conditions of the train and racecar events, because both events featured objects (an approaching train or circling racecars) that moved toward the perspective of the camera and then away in the case of the ending portion of the racecar event video. Additionally, a spatiotemporal correspondence related to the Doppler illusion may have influenced the results of Experiment 1. The Doppler illusion refers to an observer's changing perception of pitch as a sound-emitting object in motion approaches and recedes relative to the location of an observer despite the unchanging frequency of the auditory signal emitted by a moving object (Neuhoff and McBeath, 1996). Specifically, the experience of the Doppler illusion includes a perceived gradual decrease in the pitch of the auditory signal emitted by a moving object as it approaches an observer followed by another quick decrease in perceived pitch as the moving object then passes the spatial location of the observer (Rosenblum et al., 1987). Thus, it is possible that audiovisual Doppler cues could have also served as a spatiotemporal audiovisual cue when discriminating the train and racecar event videos in Experiment 1.

To address the possibility that spatiotemporal crossmodal correspondences, rather than semantic congruency, may cause the facilitatory congruency effect observed in Experiment 1, we conducted an additional control experiment using static image event stimuli that eliminated the potential influence of residual spatiotemporal crossmodal correspondences on visual awareness. If participants discriminate static visual event images faster when hearing semantically congruent soundtracks in comparison to when hearing incongruent or no soundtracks, this would provide further support for the facilitatory effect of congruent audiovisual semantic information.

## Methods

Thirty-four undergraduate students who did not participate in Experiment 1 or 2 participated in Experiment 3. All apparatuses

and stimuli were identical to those used in Experiment 1, except that the target stimuli used in Experiment 3 were static images that were selected from a single representative frame of each of the three target event videos used in Experiment 1 and 2.

## Results and Discussion

Identical data screening and aggregation procedures done prior to the analysis of data in Experiment 1 and 2 were done in Experiment 3. The data of seven subjects were excluded from analysis due to high error rates and the data from 27 participants were analyzed. A one-way repeated measures ANOVA conducted on the factor of audiovisual soundtrack condition revealed that there was a significant main effect on participants' discrimination RTs [ $F(2,25) = 3.377, p = 0.042, \eta^2 = 0.115$ ]. Planned contrast tests between the soundtrack conditions revealed that RTs were significantly faster when participants heard soundtracks that were congruent with the suppressed event image viewed in comparison to when no sound was heard (Figure 3B) [ $F(1,26) = 6.500, p = 0.017, \eta^2 = 0.200$ ]. Unlike in Experiment 1, there was no significant difference between the reaction times when participants concurrently heard a semantically congruent soundtrack in comparison to when they heard incongruent soundtracks [ $F(1,26) = 1.091, p = 0.306, \eta^2 = 0.040$ ]. Consistent with Experiment 1, there was also no significant difference between participants' RTs when they heard soundtracks that were incongruent in comparison to when nothing was heard [ $F(1,26) = 2.587, p = 0.120, \eta^2 = 0.090$ ].

The results of Experiment 3 further support that auditory semantic contexts can significantly influence the latency for suppressed static visual images to gain access to visual awareness. This result is consistent with Experiment 1, confirming that the beneficial effect of multisensory integration observed in Experiment 1 can be induced by a purely semantic congruency between auditory and visual stimuli. When considering the results of Experiments 1–3 together, our findings suggest that the multisensory integration of semantic information can occur even

when static and dynamic visual events are suppressed from visual awareness, but temporal concurrence of auditory and visual stimulation is required for audiovisual semantic congruency effects to occur.

## General Discussion

In the current study, we demonstrated that semantically congruent auditory information accelerated the time for visually suppressed familiar and dynamic events to gain access to visual awareness, indicating enhanced visual processing due to semantic congruency. In a control experiment, no significant audiovisual semantic congruency effect was observed when the soundtracks were presented prior to the onset of visual event presentation, which indicates that crossmodal priming cannot completely explain the congruency effect. We also replicated the crossmodal semantic congruency effect with static images, in which any residual spatiotemporal correspondences between the auditory and visual stimuli were removed. These results suggest that crossmodal integration of congruent semantic information occurs even when visual stimuli are not consciously perceived.

## Unconscious Semantic Processing?

Unconscious processing of emotional information has been consistently supported by behavioral (Adams et al., 2010; Yang et al., 2010) and functional imaging (Morris et al., 1999; Pasley et al., 2004; Jiang and He, 2006) studies. However, results are mixed for other types of unconscious semantic processing, such as that involving the semantics of written words and category-specific object information. Some behavioral studies show that interocularly suppressed words cannot induce semantic priming effects (Blake, 1988; Cave et al., 1998) and that high-level object adaptation is abolished if visual stimuli are rendered invisible during binocular suppression (Moradi et al., 2005). These results indicate that high-level semantic processing does not occur when visual stimuli are suppressed from visual awareness. Supporting this notion, human brain imaging studies show that object representation is eliminated during binocular rivalry suppression in inferior temporal cortex (Tong et al., 1998; Pasley et al., 2004). A recent ERP study also reveals that the N400 component, an index of semantic information processing, is missing when participants are completely unaware of the meaning of dichoptically presented words (Kang et al., 2011).

There is, on the other hand, accumulating evidence supporting unconscious processing of semantic information. Chinese (Hebrew) words suppressed by CFS break up suppression faster than Hebrew (Chinese) words for Chinese (Hebrew) readers, indicating that the meaning of words are processed unconsciously and can influence access to visual awareness (Jiang et al., 2007). Priming of associated visual words can result in a faster breakup of suppression for visually presented words suppressed by CFS (Costello et al., 2009, but see also Lupyan and Ward, 2013). It is also shown that suppressed words can affect behavioral performance in a problem-solving task (Zabelina et al., 2013). Human brain imaging studies demonstrate that multi-voxel pattern analysis can extract category-specific object information even when objects are suppressed from visual awareness during CFS

(rendering BOLD signals reduced close to baseline) in category-specific areas such as FFA and PPA (Sterzer et al., 2008) and other visual areas such as the lateral occipital area and the intra-parietal sulcus (Hesselmann and Malach, 2011). These results suggest that semantic information conveyed by visual objects can survive strong interocular suppression.

The current study demonstrates that interocularly suppressed dynamic events gain access to visual awareness faster when they are semantically congruent with sounds. Although indicating that audiovisual crossmodal integration occurs during visual suppression, our results do not necessarily indicate the unconscious processing of semantic information. The current study cannot determine whether crossmodal integration with invisible visual stimuli requires semantic processing of both auditory and visual information. It is possible that semantic processing of sound, which was clearly heard in the current study, may enhance visual processing of the suppressed event without unconscious visual semantic analysis. Further studies are required to clearly answer this question.

## Potential Mechanisms of the Crossmodal Semantic Congruency Effect

The semantic crossmodal congruency effect observed in the current study may not be caused by semantic priming as observed in an aforementioned study (Costello et al., 2009). Whereas semantic priming can occur when a prime precedes the presentation of target stimuli, we did not observe a significant semantic congruency effect when the sound was presented before the presentation of the target stimuli. This result indicates that multisensory integration based on semantic congruency may also require temporal proximity (Meredith et al., 1987; van Atteveldt et al., 2007). However, we observed a weak tendency toward a congruency effect, which suggests that semantic priming may be partially involved in the present crossmodal semantic congruency effect. Differences between the current study and previously mentioned studies (Costello et al., 2009; Lupyan and Ward, 2013) may explain why a congruency effect with temporal displacement was not presently observed. First, since our stimuli depicted dynamic events (e.g., an approaching train or circling racecars), concurrent presentation of information might be more important than previous studies using static images (written words or objects). However, we excluded this possible explanation with a control experiment using static images. Second, previous studies used lexical stimuli to induce a priming effect, but we used naturalistic non-speech sounds from the events. Word primes may activate greater amounts of information in a more extensive semantic network compared to the natural sounds of the individual events.

Hard-wired connections between primary sensory areas and multisensory areas, such as superior colliculus (SC) and posterior superior temporal sulcus (pSTS; Stein and Meredith, 1993; Stein, 1998; Noesselt et al., 2010) as well as between primary sensory cortices (Driver and Noesselt, 2008) have been suggested as underlying mechanisms for the beneficial effect of multisensory interaction based on spatial and temporal congruency. However, the neural mechanisms of purely semantic-based multisensory interaction are still not clear. A few recent

brain-imaging studies suggest that inferior frontal cortex (IFC) and pSTS areas are activated differentially between semantically congruent and incongruent audiovisual stimuli (Belardinelli et al., 2004; Hein et al., 2007; Plank et al., 2012). However, enhancement and reduction in BOLD responses to semantically congruent vs. incongruent audiovisual stimuli vary depending on the brain areas, and the interpretations for the changes in these BOLD responses are still under debate. Although further studies are required to reveal underlying neural mechanisms for semantic-based multisensory integration, we speculate that multisensory cortical areas contribute to the semantic congruency effect observed in the current study.

## Conclusion

In the current study, we examined whether an audiovisual semantic congruency effect can occur even when visual stimuli are suppressed from visual awareness. In a series of experiments, we show that visual events suppressed by CFS gain preferential access to visual awareness only when semantically congruent sound is concurrently heard but not when the same sound is heard before visual event presentation. Our results suggest that first, semantic-based audiovisual integration can occur when visual stimuli are rendered invisible, and second, multisensory integration based on semantic congruency also requires temporal proximity.

## References

- Adams, W. J., Gray, K. L. H., Garner, M., and Graf, E. W. (2010). High-level face adaptation without awareness. *Psychol. Sci.* 21, 205–210. doi: 10.1177/0956797609359508
- Alsius, A., and Munhall, K. G. (2013). Detection of audiovisual speech correspondences without visual awareness. *Psychol. Sci.* 24, 423–431. doi: 10.1177/0956797612457378
- Belardinelli, M. O., Sestieri, C., Di Matteo, R., Delogu, F., Del Gratta, C., Ferretti, A., et al. (2004). Audio-visual crossmodal interactions in environmental perception: an fMRI investigation. *Cogn. Process.* 5, 167–174. doi: 10.1007/s10339-004-0024-0
- Bell, A. H., Meredith, M. A., Van Opstal, A. J., and Munoz, D. P. (2005). Cross-modal integration in the primate superior colliculus underlying the preparation and initiation of saccadic eye movements. *J. Neurophysiol.* 93, 3659–3673. doi: 10.1152/jn.01214.2004
- Ben-Artzi, E., and Marks, L. E. (1995). Visual-auditory interaction in speeded classification: role of stimulus difference. *Percept. Psychophys.* 8, 1151–1162. doi: 10.3758/BF03208371
- Bernstein, I. H., and Edelman, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *J. Exp. Psychol.* 87, 241–247. doi: 10.1037/h0030524
- Blake, R. (1988). Dichoptic reading: the role of meaning in binocular rivalry. *Percept. Psychophys.* 44, 133–141.
- Bolognini, N., Frassinetti, F., Serino, A., and Ladavas, E. (2005). “Acoustical vision” of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Exp. Brain Res.* 160, 273–282. doi: 10.1007/s00221-004-2005-z
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357
- Cappe, C., Thut, G., Romei, V., and Murray, M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia* 47, 1045–1052. doi: 10.1016/j.neuropsychologia.2008.11.003
- Cave, C. B., Blake, R., and McNamara, T. P. (1998). Binocular rivalry disrupts visual priming. *Psychol. Sci.* 9, 299–302. doi: 10.1111/1467-9280.00059
- Chen, Y.-C., and Spence, C. (2010). When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures. *Cognition* 114, 389–404. doi: 10.1016/j.cognition.2009.10.012
- Chen, Y.-C., and Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1554–1568. doi: 10.1037/a0024329
- Chen, Y.-C., Yeh, S.-L., and Spence, C. (2011). Crossmodal constraints on human perceptual awareness: auditory semantic modulations of binocular rivalry. *Front. Psychol.* 2:212. doi: 10.3389/fpsyg.2011.00212
- Costello, P., Jiang, Y., Baartman, B., McClennen, K., and He, S. (2009). Semantic and subword priming during binocular suppression. *Conscious. Cogn.* 18, 375–382. doi: 10.1016/j.concog.2009.02.003
- Dehaene, S., Naccache, L., Le Clec'H, G., Koehlin, E., Mueller, M., Dehaene-Lambertz, G., et al. (1998). Imaging unconscious semantic priming. *Nature* 395, 597–600. doi: 10.1038/26967
- Doehrmann, O., and Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain Res.* 1242, 136–150. doi: 10.1016/j.brainres.2008.03.071
- Driver, J., and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on ‘sensory-specific’ brain regions, neural responses, and judgements. *Neuron* 57, 11–23. doi: 10.1016/j.neuron.2007.12.013
- Evans, K. K., and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *J. Vis.* 10, 1–12. doi: 10.1167/10.1.6
- Frens, M. A., Van Opstal, A. J., and Van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept. Psychophys.* 57, 802–816. doi: 10.3758/BF03206796
- Ghazanfar, A. A., and Maier, J. X. (2009). Rhesus monkeys (*Macaca mulatta*) hear rising frequency sounds as looming. *Behav. Neurosci.* 4, 822–827. doi: 10.1037/a0016391
- Hein, G., Doehrmann, O., Muller, N. G., Kaiser, J., Muckli, L., and Naumer, M. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J. Neurosci.* 27, 7881–7887. doi: 10.1523/jneurosci.1740-07.2007
- Hesselmann, G., and Malach, R. (2011). The link between fMRI-BOLD activation and perceptual awareness in “stream-invariant” in the human visual system. *Cereb. Cortex* 21, 2829–2837.
- Hsiao, J.-Y., Chen, Y.-C., Spence, C., and Yeh, S.-L. (2012). Assessing the effects of audiovisual semantic congruency on the perception of a bistable figure. *Conscious. Cogn.* 21, 775–787. doi: 10.1016/j.concog.2012.02.001
- Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., and Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychon. Bull. Rev.* 15, 548–554. doi: 10.3758/PBR.15.3.548
- Jiang, Y., Costello, P., and He, S. (2007). Processing of invisible stimuli. *Psychol. Sci.* 18, 349–355. doi: 10.1111/j.1467-9280.2007.01902.x
- Jiang, Y., and He, S. (2006). Cortical responses to invisible faces: dissociating subsystems for facial-information processing. *Curr. Biol.* 16, 2023–2029. doi: 10.1016/j.cub.2006.08.084
- Kang, M.-S., Blake, R., and Woodman, G. F. (2011). Semantic analysis does not occur in the absence of awareness induced by interocular suppression. *J. Neurosci.* 31, 13535–13545. doi: 10.1523/jneurosci.1691-11.2011
- Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., and Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Exp. Brain Res.* 158, 405–414. doi: 10.1007/s00221-004-1913-2
- Laurienti, P. J., Wallace, M. T., Maldjian, J. A., Susi, C. M., Stein, B. E., and Burdette, J. H. (2003). Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. *Hum. Brain Mapp.* 19, 213–223. doi: 10.1002/hbm.10112
- Levelt, W. J. M. (1965). *On Binocular Rivalry*. Soesterberg: Institute for Perception RVO-TNO.
- Lewis, R., and Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *J. Neurosci.* 30, 12329–12339. doi: 10.1523/jneurosci.5745-09.2010
- Lovelace, C. T., Stein, B. E., and Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Cogn. Brain Res.* 17, 447–453. doi: 10.1016/S0926-6410(03)00160-5
- Lupyan, G., and Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14196–14201. doi: 10.1073/pnas.1303312110



- Meredith, M. A., Nemitz, J. W., and Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J. Neurosci.* 7, 3215–3229.
- Moradi, F., Koch, C., and Shimojo, S. (2005). Face adaptation depends on seeing the face. *Neuron* 45, 169–175. doi: 10.1016/j.neuron.2004.12.018
- Morris, J. S., Ohman, A., and Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating “unseen” fear. *Proc. Natl. Acad. Sci. U.S.A.* 96, 1680–1685. doi: 10.1073/pnas.96.4.1680
- Neuhoff, J. G., and McBeath, M. K. (1996). The Doppler illusion: the influence of dynamic intensity change on perceived pitch. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 970–985. doi: 10.1037/0096-1523.22.4.970
- Noesselt, T., Bergmann, D., Hake, M., Heinze, H. J., and Fendrich, R. (2008). Sound increases the saliency of visual events. *Brain Res.* 1220, 157–163. doi: 10.1016/j.brainres.2007.12.060
- Noesselt, T., Tyll, S., Boehler, C. N., Budinger, E., Heinze, H. J., and Driver, J. (2010). Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. *J. Neurosci.* 30, 13609–13623. doi: 10.1523/jneurosci.4524-09.2010
- Pandey, P. C., Kunov, H., and Abel, S. M. (1986). Disruptive effects of auditory signal delay on speech perception with lipreading. *J. Audit. Res.* 26, 27–41.
- Pasley, B. N., Mayes, L. C., and Schultz, R. T. (2004). Subcortical discrimination of unperceived objects during binocular rivalry. *Neuron* 42, 163–172. doi: 10.1016/S0896-6273(04)00155-2
- Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366
- Plank, T., Rosengarth, K., Song, W., Ellermeier, W., and Greenlee, M. W. (2012). Neural correlates of audio-visual object recognition: effects of implicit spatial congruency. *Hum. Brain Mapp.* 33, 797–811. doi: 10.1002/hbm.21254
- Plass, J., Guzman-Martinez, E., Ortega, L., Grabowecy, M., and Suzuki, S. (2014). Lipreading without awareness. *Psychol. Sci.* 25, 1835–1837. doi: 10.1177/0956797614542132
- Rosenblum, L. D., Carello, C., and Pastore, R. E. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception* 16, 175–186. doi: 10.1068/p160175
- Schneider, T. R., Engel, A. K., and Debener, S. (2008). Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Exp. Psychol.* 55, 121–132. doi: 10.1027/1618-3169.55.2.121
- Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Attent. Percept. Psychophys.* 73, 971–995. doi: 10.3758/s13414-010-0073-7
- Stein, B. E. (1998). Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. *Exp. Brain Res.* 123, 124–135. doi: 10.1007/s002210050553
- Stein, B. E., Huneycutt, W. S., and Meredith, M. A. (1988). Neurons and behavior: the same rules of multisensory integration apply. *Brain Res.* 448, 355–358. doi: 10.1016/0006-8993(88)91276-0
- Stein, B. E., and Meredith, M. A. (1993). *The Merging of the Senses*. Cambridge, MA: MIT Press.
- Stein, B. E., and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.* 9, 255–266. doi: 10.1038/nrn2377
- Stein, T., Hebart, M. N., and Sterzer, P. (2011). Breaking continuous flash suppression: a new measure of unconscious processing during interocular suppression? *Front. Hum. Neurosci.* 5:167. doi: 10.3389/fnhum.2011.00167
- Sterzer, P., Haynes, J.-D., and Rees, G. (2008). Fine-scale activity patterns in high-level visual areas encode the category of invisible objects. *J. Vis.* 8, 1–12. doi: 10.1167/8.15.10
- Tong, F., Nakayama, K., Vaughan, J. T., and Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 21, 753–759. doi: 10.1016/S0896-6273(00)80592-9
- Tsuchiya, N., and Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nat. Neurosci.* 8, 1096–1101. doi: 10.1038/nn1500
- van Atteveldt, N. M., Formisano, E., Blomert, L., and Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb. Cortex* 17, 962–974. doi: 10.1093/cercor/bhl007
- Vroomen, J., and de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *J. Exp. Psychol.* 26, 5, 1583–1590. doi: 10.1037/0096-1523.26.5.1583
- Yang, E., Hong, S.-W., and Blake, R. (2010). Adaptation aftereffects to facial expressions suppressed from visual awareness. *J. Vis.* 10, 1–13. doi: 10.1167/10.12.24
- Yang, E., Zald, D. H., and Blake, R. (2007). Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion* 7, 882–886. doi: 10.1037/1528-3542.7.4.882
- Zabelina, D. L., Guzman-Martinez, E., Ortega, L., Grabowecy, M., Suzuki, S., and Beeman, M. (2013). Suppressed semantic information accelerates analytic problem solving. *Psychon. Bull. Rev.* 20, 581–585. doi: 10.3758/s13423-012-0364-1

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Cox and Hong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Modality use in joint attention between hearing parents and deaf children

Nicole Depowski<sup>1</sup>, Homer Abaya<sup>2</sup>, John Oghalai<sup>2</sup> and Heather Bortfeld<sup>3\*</sup>

<sup>1</sup> Department of Psychology, University of Connecticut, Storrs, CT, USA, <sup>2</sup> Head and Neck Surgery, Department of Otolaryngology, Stanford University School of Medicine, Stanford, CA, USA, <sup>3</sup> Psychological Sciences, University of California, Merced, Merced, CA, USA

The present study examined differences in modality use during episodes of joint attention between hearing parent-hearing child dyads and hearing parent-deaf child dyads. Hearing children were age-matched to deaf children. Dyads were video recorded in a free play session with analyses focused on uni- and multimodality use during joint attention episodes. Results revealed that adults in hearing parent-deaf child dyads spent a significantly greater proportion of time interacting with their children using multiple communicative modalities than adults in hearing parent-hearing child dyads, who tended to use the auditory modality (e.g., oral language) most often. While these findings demonstrate that hearing parents accommodate their children's hearing status, we observed greater overall time spent in joint attention in hearing parent-hearing child dyads than hearing parent-deaf child dyads. Our results point to important avenues for future research on how parents can better accommodate their child's hearing status through the use of multimodal communication strategies.

## OPEN ACCESS

### Edited by:

Magda L. Dumitru,  
Macquarie University, Australia

### Reviewed by:

Louis A. Schmidt,  
McMaster University, Canada  
Amy M. Lieberman,  
University of California, San Diego,  
USA

### \*Correspondence:

Heather Bortfeld  
heather.bortfeld@uconn.edu

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 16 February 2015

**Accepted:** 25 September 2015

**Published:** 12 October 2015

### Citation:

Depowski N, Abaya H, Oghalai J  
and Bortfeld H (2015) Modality use  
in joint attention between hearing  
parents and deaf children.  
Front. Psychol. 6:1556.  
doi: 10.3389/fpsyg.2015.01556

**Keywords:** joint attention, multimodal communication, Parent-child communication, ELAN, cochlear implants, deaf

## INTRODUCTION

Imagine a world in which the way that people communicate is inherently different from how you communicate: You use visual information and they insist on using auditory information. The result is confusion and miscommunication. For many children who are born deaf, this is the reality they initially face. This is because 90 percent of deaf children are born to hearing parents (Mitchell and Karchmer, 2004), meaning that there is an inherent mismatch between parent and child in the dominant modality used for communication. Here we examine how hearing parents accommodate their deaf children's hearing status by documenting the modality or modalities used in communication between parents and children. To what extent do parents use changes in modality to accommodate their child's hearing loss and how do children adapt to those changes?

Language development is often delayed in deaf children of hearing parents (Lederberg and Everhart, 1998) because the majority of hearing parents of deaf children have no prior experience using sign language to communicate (DeMarco et al., 2007) and must adjust to their child's hearing status. Parents may choose to learn sign language, they may choose to have their deaf child evaluated for cochlear implant candidacy, or they may do both of these things. Regardless, early communication between hearing parents and deaf children presents a significant obstacle, as suggested by evidence that deaf children of hearing parents have an increased rate of behavioral issues and that these issues are related to communication difficulties (Barker et al., 2009).

This mismatch may pose difficulties for parents as well. Hearing parents of deaf children may experience stress specifically with regard to their child's deafness (Lederberg and Golbach, 2002), and those with the highest levels of stress also tend to have deaf children with more social and emotional development problems (Hitermair, 2006). Given the clear impact of maternal stress on children's development and the importance of communication between parents and their children to mitigate sources of stress, the present study was designed to compare communication in hearing parent-deaf child dyads (in which parents were using a predominantly auditory-oral approach) and hearing parent-hearing child dyads. One way that hearing mothers appear to mitigate the difficulties in communication with their deaf children is by changing their own behavior to accommodate the children's limited access to the auditory modality. For example, during free play sessions, hearing mothers of deaf infants have been shown to use exaggerated gestures relative to deaf mothers of deaf children (Koester et al., 1998a), suggesting that they are trying to use a non-auditory modality to communicate even if they are not learning sign themselves. In another study, hearing mothers of deaf infants were found to move objects into a child's visual field and tap on or point to objects to get the child to attend to them (Waxman and Spencer, 1997). Our goal in the current study was to characterize how hearing parents of deaf children who use an auditory-oral approach accommodate their children's communicative needs.

Research that is informative on the issue of communicative accommodation involves use of the Still Face Paradigm, in which a mother is instructed to maintain a neutral, unemotive face at prescribed intervals during normal interaction with her infant (Cohn and Tronick, 1983). Although this paradigm was initially developed for the study of effects of depressive mothers on young children (Cohn and Tronick, 1983), it has since been used to probe other areas of early development (see Mesman et al., 2009). Typically, when the mother tries to re-engage with the infant after a period of maintaining a blank face, she must work harder than usual to successfully re-engage with her infant. When used with deaf children, the paradigm has revealed that hearing mothers use spoken language to engage their 9-month-old infants more than deaf mother-deaf child dyads (Koester et al., 1998b), despite the child's lack of access to the auditory modality. While this is not entirely surprising, no difference between dyad types was found in use of the visual or tactile modalities to re-engage the infants (Koester et al., 1998b; Koester, 2001), showing that hearing parents were accommodating their deaf infants communicatively.

Another way to examine whether and how parents accommodate their children communicatively is through their efforts to establish joint attention. Joint attention is the ability to focus simultaneously on an object or event and another person, sometimes described as "shared intentionality" (Tomasello, 1995; Tomasello and Carpenter, 2007). Joint attention can be further divided into the acts of both initiating a bid for attention (e.g., pointing at a balloon) and responding to a bid for attention (e.g., commenting that the balloon is red; Mundy et al., 2007). This is an act of reciprocating communication within a dyad, and is essential to basic human communication. The act of successfully

initiating joint attention is more sophisticated than responding to it; thus, the act of a child successfully initiating joint attention can be considered to be the start of formalized and intentional communication (Brinck, 2001). Although joint attention is commonly the focus of research on language development, it is also relevant to more general social and emotional development (Mundy et al., 1990; Corkum and Moore, 1998; Mundy and Gomes, 1998; Mundy and Neal, 2000). Given these broad developmental implications, joint attention provides a way to characterize interactions between hearing parents and their deaf children.

Gale and Schick (2009) focused on symbol-infused joint attention or joint attention during symbolic communication, between 24-month-old deaf children and their hearing parents. Although deaf children of hearing parents did not differ from deaf children of deaf parents or hearing children of hearing parents on most language measures, they did engage in significantly fewer sustained interactions (Gale and Schick, 2009). This difference is notable given that much of the cognitive benefit derived from joint attention originates from the sustained interaction between parent and child, suggesting that this may be the source of some of the negative developmental outcomes seen in this population. In other research, hearing parents of hearing children (between 18 and 36 months of age) rated their children as having higher adaptive social behavior than hearing parents of deaf children of the same age range. Moreover, these researchers found that higher rates of successful joint attention were associated with higher ratings of the children's adaptive social behavior, regardless of hearing status (Nowakowski et al., 2009). This highlights the substantial role that joint attention plays in development in general. Importantly, hearing mothers of deaf 36-month-olds have been shown to use more modalities of communication to gain their child's attention during interaction than hearing parents of hearing children do (Lederberg and Everhart, 1998).

While the comparison of hearing parent-deaf child dyads to hearing parent-hearing child dyads is helpful, it is just as important to compare modality-matched dyads (e.g., hearing parent-hearing child dyads and deaf parent-deaf child dyads). Lieberman et al. (2014) did just this, focusing on the specific types of gaze used by these dyads during joint attention. Their results demonstrate that the way in which partners in these different dyads engage one another is qualitatively different. Deaf children switched gaze between the parent and the object of interest much more often than hearing parents of hearing children, suggesting that deaf children who are exposed to sign [in this case, American Sign Language (ASL)] are able to meet the attention-switching requirements of joint attention (Lieberman et al., 2014). Compared to hearing parent-hearing child dyads, hearing parent-deaf child dyads spent less time overall in joint attention (Prezbindowski et al., 1998). Considering that much of the benefit of joint attention derives from the interaction inherent in it, and much of what is learned in joint attention can be symbolic (including language), this difference is of concern. The authors hypothesized that the reason for this is that hearing parents of deaf children try to engage their children in symbol-infused joint attention by using oral language (i.e., the auditory modality; Prezbindowski et al., 1998).

A hearing parent's use of the auditory modality with a deaf child highlights one of the primary difficulties in hearing parent-deaf child communication. Hearing parents rely on oral language in the rest of their lives but cannot use it to communicate with their children effectively. While this may seem obvious, the instinctive use of oral communication by hearing parents affects important basic interactions, such as when parents direct their children's attention to objects and events in their surroundings. In a study on children's visual perception of Manual Coded English (MCE), a communication method which involves a hearing mother speaking while signing to her deaf child, mothers who used more deaf-friendly means of communication had children who saw more complete versions of the mothers' signed utterances (Swisher, 1991). For example, the most successful mother, as measured by the percentage of complete utterances seen by her child, tapped the child to ensure that the child was paying attention to before the mother began to sign and did so more frequently than any other mother in the study (Swisher, 1991). This study is relevant to the current study as it highlights the link between language and joint attention. For deaf children, attention established in the visual modality is necessary for subsequent access to visual language. Even if a hearing parent makes the decision to have a deaf child implanted, there will be a period of time—that is, the preimplantation period—during which the dominant modality of communication is mismatched between parents and child.

Clearly, which modalities are used in communication between hearing parents and deaf children is a topic that merits further research. While previous research (e.g., Trautman, 2009) has examined modality differences in communication between hearing parents and deaf children in broad terms, in the present study we sought to establish more precise coding of modality use during establishment of joint attention between hearing parent and deaf children and compare that to similarly precise coding of hearing parent-hearing child dyads. We were particularly interested in seeing whether differences emerged between the two dyad types in terms of how the parent worked to establish a child's attention. More generally, this study represents a first step toward documentation of the communicative modalities that non-signing hearing parents use to establish joint attention with their deaf children.

## MATERIALS AND METHODS

### Participants

Four severely to profoundly deaf children ( $n = 4$  females) aged 18.2–36.7 months ( $M = 26.83$ ,  $SD = 7.78$ ; specifically, ages 18.2, 24.1, 28.3, 36.7) and their hearing parents ( $n = 4$  females), participated in the study. While all children were candidates for cochlear implantation, none of them had received an implant; the children were being instructed predominantly using the oral method. None of the children produced any spoken or signed language during videotaping in our sample. Each child was receiving at least 1 h per week of speech therapy, as well as some basic instruction in ASL. In addition, four hearing children ( $n = 4$  females) ages 18.3–36.7 months ( $M = 26.85$ ,  $SD = 7.72$ ;

specifically, 18.3, 24.1, 28.3, 36.7) and their hearing parents ( $n = 4$  females) took part in the study. Participants were aged-matched, and were from the Southwestern and Northeastern United States. Each was recruited via the National Institute of Health website or local recruitment. The sample was primarily Caucasian (two of the deaf children identified as Caucasian, Hispanic/Latino), and all but one parent had completed at least high school. This study was carried out in accordance with the recommendations of the University of Connecticut Institutional Review Board and the Stanford University School of Medicine Institutional Review Board with written informed consent from all subjects. For participants who were young children, parents provided written informed consent. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

### Materials

Age appropriate toys (a ball, a set of large blocks, a set of stacking cups, tableware, a tower of stacking rings, and toy cars) were used during a free-play session between the child and his/her primary caregiver, which occurred as part of a visit with a speech language pathologist (deaf children) or to the Husky Pup Language Lab at the University of Connecticut (hearing children). The speech language pathologist or experimenter instructed the caregiver to play with the child as she would at home; play sessions were video recorded for approximately five minutes ( $M = 464.23$ ,  $SD = 154.35$ ; see **Table 1**). Videos of hearing parent-deaf child dyads were then transmitted from collaborators at Stanford University to researchers at University of Connecticut using Research Electronic Data Capture (REDCap) electronic data capture tools hosted at Stanford University (Harris et al., 2009). REDCap is a secure, web-based application designed to support data capture for research studies. It provided the two labs with a vehicle for validated data entry with audit trails for tracking data entry and export, as well as procedures for importing data from external sources. For the current study, REDCap was used solely as a means of secure transfer of videos between collaborators, and was not used for any analytical/coding purposes.

### Procedure

The videos were coded for joint attention using ELAN (Wittenburg et al., 2006), language annotation software created at the Max Planck Institute for Psycholinguistics, (The Language Archive, Nijmegen, The Netherlands). ELAN allows for multimodal analyses of language and other behavior (<http://tla.mpi.nl/tools/tla-tools/elan/>), and is available free of charge. We use coding criteria for joint attention based on the work of Tek (2010), which was a modified version of the Early Social Communication Scales, a measure of early development that can be used on typically developing populations (Mundy et al., 1996.) Coded variables were analyzed using ELAN, Microsoft Excel, VassarStats, and SPSS.

### Video Processing

Videos were reviewed for visual clarity and Adobe Premiere Pro (CS6) was used to cut the video to the start and end time of the play session. The start time of the play session was at the first frame in which the testing room's door was closed, leaving the



**TABLE 1 | Lengths of play sessions (in seconds), proportion of time spent in joint attention episodes, and length of joint attention episodes (in seconds).**

Dyad/age (in months)	Dyad 1 (18 m)	Dyad 2 (24 m)	Dyad 3 (28 m)	Dyad 4 (36 m)
Hearing parent-hearing child length of play session <i>Total time (in seconds)</i>	385.44	540.76	679.84	408.58
Hearing parent-deaf child length of play session <i>Total time (in seconds)</i>	375.98	247.77	427.15	655.97
Hearing parent-hearing child joint attention <i>Proportion/length of time (in seconds)</i>	0.12/48.02	0.34/184.76	0.19/129.34	0.55/225.28
Hearing parent-deaf child joint attention <i>Proportion/length of time (in seconds)</i>	0.11/39.79	0.19/43.25	0.38/158.29	0.02/12.34

child and parent alone. The end of the play session was at the first frame in which the experimenter opened the door to end the play session. These two values were subtracted to give a baseline length of time for the play session. Next, intervals in which the video was uncodeable were marked. An uncodeable interval was defined as an interval of at least 5 s in which at least one participant's face was not visible. The amount of uncodeable time was subtracted from the baseline length of time to yield a total length of play session for each participant.

### Joint Attention Coding

Very few instances of child-initiated joint attention were observed; thus, this construct was not included for analysis in the paper. Moreover, in the present study, only successful bids for joint attention (i.e., joint attention episodes) were coded and quantified. A successful joint attention episode involved the adult making a bid for the child's attention using pointing, gaze switching between the object and the child, tapping or touching the child, deliberate waving in the child's visual field, changing affect, and/or language; this bid was then responded to by the child using pointing, gaze switching between the object and the parent, tapping or touching the parent, grasping the object of interest, deliberate waving in the parent's visual field, changing affect, and/or language. This type of episode could also occur if a parent shifted the child's attention from one object to another using the previously mentioned techniques. Any indication of the auditory modality being used is *during* its use within a joint attention episode. Most instances of use of the auditory modality

with deaf children were brief and the hearing parent would proceed to a different modality of engagement.

To record joint attention in ELAN, a 5 s "rule of engagement" was followed (i.e., after interacting with an object, a member of the dyad had 5 s to begin to engage with the other member of the dyad and vice versa for interactions beginning with a member of the dyad). Similarly, there was a 5 s rule of disengagement, i.e., a joint attention episode was deemed to be terminated after neither participant engaged in joint attention behavior for 5 s. If either participant re-engaged within the 5-s window, the length of the episode was extended; the episode ended at the start of the first period of 5 s that displayed no joint attention behaviors.

### Coding for Modality

All successful, adult-initiated joint attention episodes were then coded separately for both the parent's and the child's uses of the following modalities: auditory, visual, tactile, auditory-visual, auditory-tactile, visual-tactile, and auditory-visual tactile. The criteria are as follows. One episode could have multiple modalities used within it, as specified by the following categories:

#### Auditory

Behaviors in the auditory modality involved using sound to gain the attention of the other member of the dyad. These included language, humming, other vocal sounds (e.g., "psst!"), hitting an object to make noise, clapping (if the other member of the dyad was unable to see the clap), and causing a toy to produce noise (i.e., squeaking a small toy or pressing a button on a toy to cause the toy to produce noise such as music or animal sounds.) This modality was coded for when there was no possible way for the other dyad member to have received visual input with the auditory input.

#### Visual

The visual modality included behaviors that somehow incorporated the visual field in getting the other member's attention. These included waving, gesturing, pointing, making eye contact, holding an object directly in the other member's visual field, causing a toy to light up (but not produce sound), demonstrating play with toys, offering a toy to the other partner (without using any of the behaviors described in the auditory section), making faces, and changing affect. As no ASL was produced in any of the dyads, it was subsequently excluded from the coding criteria.

#### Tactile

The tactile modality involved using touch, either direct or indirect. Examples included tapping/touching the other person, tickling, hugging, holding, grabbing on to the other person's clothing, tapping the ground to create vibrations, and touching the other person with a toy (out of their visual field).

#### Auditory-visual

This multimodal classification involved criteria for both the auditory and visual modalities occurring simultaneously. Examples included gesturing while talking, presenting a toy while describing it, reacting to a visual event (e.g., saying "uh oh" when a toy rolls under a table), and demonstrating affect while producing any sort of sound.

**TABLE 2 | Mean ranks of adult modality use.**

	Auditory	Visual	Auditory-visual	Visual-tactile	Auditory-visual-tactile
Hearing parent-deaf child	3.3	3	5.5	4.3	6.3
Hearing parent-hearing child	5.8	6	3.5	4.8	2.8

A higher mean rank indicates that adults in that dyad type spent significantly more time in that particular modality. All results were significant at  $p < 0.05$ .

### Auditory-tactile

This multimodal classification included criteria for both the auditory and tactile modalities. This included running a toy over the other partner while making appropriate noises (e.g., running a toy car over the other member's back while saying "vroom" or making other vehicular noises), holding/grasping hands while signing (e.g., the parent grabs the child's hands to help them do the motions for "Patty Cake"), and touching the other person with a toy that made noise.

### Visual-tactile

This multimodal classification included criteria for visual and tactile modalities. It included behaviors such as taking a toy and making it "hop" up the other person's arm (without making noise), making eye contact with the other person while also touching them, grabbing the other person's arm while pointing, and touching the other person with a toy within their visual field while not producing any auditory output.

### Auditory-visual-tactile

This multimodal classification included simultaneously occurring behaviors encompassed by the criteria for the auditory, visual, and tactile modalities. It included holding a child while pointing and talking to them, making eye contact while singing and touching the other person, and both people playing a clapping game that involves auditory output of some sort while making eye contact.

### Coding Modality in ELAN

To record modality use in ELAN, the start of the production of a modality was coded in real time (i.e., there was no rule of engagement). However, there was a two second rule of discontinuing the modality, i.e., a participant could pause in production of the modality for up to 2 s and have the subsequent production be part of the same episode. Modality episodes were deemed to be terminated after neither participant engaged in any of the modality criteria behaviors for over 2 s, with the end time of the episode being the end time of the last modality production. Abrupt changes in modality type (e.g., the parent switches from speaking to speaking and pointing) were coded in real time, with no rule of engagement or disengagement.

### Extracting Data for Analyses

Data were extracted from individual videos using the "View Annotation Statistics" function in ELAN. Total times were

extracted for length of time spent in joint attention. In addition, modality times were extracted, after having been coded as a controlled vocabulary in ELAN (and a dependent tier of joint attention). These data were then analyzed as described in Section "Analyses." Inter-observer reliability ( $n = 3$ ) for these measures was calculated at >90% agreement.

### Analyses

In order to account for differences in the lengths of free play sessions, the metric of proportion of total session length spent in joint attention was computed. To compute this metric, the total amount of time spent in this episode type was extracted from ELAN for each participant. These times were divided by the total session length (excluding uncodeable time) for each participant, i.e., total time spent in adult-initiated, successful bids for joint attention was divided by total session length (see **Table 1** for proportions and lengths of time spent in joint attention for each dyad). With regards to modality, seven modality metrics were computed for both parents and children. This was done by extracting the total amount of time spent in each of the seven modalities, and dividing each in turn by the total amount of time spent in joint attention in the free play session. Mann-Whitney  $U$  analyses were conducted not only to compare joint attention behavior between the hearing parent-hearing child and hearing parent-deaf child groups, but also to compare modality use by both parents and children in the two dyad types.

### RESULTS

We first compared the overall proportion of time spent in joint attention between parents and children in the two dyad types. The results of a Mann-Whitney  $U$  analysis indicated that hearing parent-hearing child dyads spent a significantly higher proportion of time in joint attention than hearing parent-deaf child dyads,  $U = 15$ ,  $p < 0.05$ .

We then evaluated modality use by adults across dyad types during periods of joint attention. Because no instances of tactile-only or auditory-tactile modality combinations were produced by adults in either dyad type, these modalities were excluded from further analysis. First, a comparison of the differences in proportion of time spent in the auditory modality reveals that adult in hearing parent-hearing child dyads spent significantly more time in the auditory modality than adults in hearing parent-deaf child dyads,  $U = 13$ ,  $p < 0.05$ . Moreover, adults in hearing parent-hearing child dyads spent a greater proportion of time in the visual modality than hearing parent-deaf child dyads,  $U = 14$ ,  $p < 0.05$ . Thus, hearing parents of hearing children were more likely to use unimodal forms of communication than hearing parents of deaf children (see **Table 2** for a summary of results; see **Table 3** for descriptive statistics).

What about instances in which two modalities were used during joint attention episodes? A comparison of the proportion of time adults in the two dyad types spent communicating in the auditory-visual modality revealed a significant difference,  $U = 4$ ,  $p < 0.05$ , such that adults in hearing parent-deaf

**TABLE 3 | Descriptive statistics of proportion of time spent by adults in each modality type during joint attention.**

	Auditory	Visual	Auditory-visual	Visual-tactile	Auditory-visual-tactile
Hearing parent-deaf child <i>M (SD)</i>	0.03 (0.06)	0.24 (0.12)	0.46 (0.11)	0.02 (0.04)	0.18 (0.13)
Hearing parent-hearing child <i>M (SD)</i>	0.11 (0.12)	0.45 (0.14)	0.36 (0.10)	0.01 (0.01)	0.02 (0.04)

**TABLE 4 | Mean ranks of child modality use.**

	Auditory	Visual	Auditory-visual	Auditory-tactile	Visual-tactile	Auditory-visual-tactile
Hearing parent-deaf child	3	6	3.1	4	3.3	3.9
Hearing parent-hearing child	6	3	5.9	5	5.8	5.1

A higher mean rank indicates that children in that dyad type spent significantly more time in that particular modality. All results were significant at  $p < 0.05$ .

**TABLE 5 | Descriptive statistics of proportion of time spent by children in each modality type during joint attention.**

	Auditory	Visual	Auditory-visual	Auditory-tactile	Visual-tactile	Auditory-visual-tactile
Hearing parent-deaf child <i>M (SD)</i>	0.00 (0.00)	0.62 (0.12)	0.06 (0.11)	0.00 (0.00)	0.12 (0.14)	0.06 (0.11)
Hearing parent-hearing child <i>M (SD)</i>	0.07 (0.06)	0.32 (0.17)	0.15 (0.11)	0.003 (0.005)	0.35 (0.35)	0.06 (0.05)

child dyads spent a greater proportion of time using this combination than hearing parents of hearing children. In contrast, analysis of the visual-tactile modality demonstrated that adults in hearing parent-hearing child dyads spent a significantly greater proportion of time using this combination than adults in hearing parent-deaf child dyads,  $U = 9$ ,  $p < 0.05$ . Finally, the only case in which adults used three modalities simultaneously involved auditory-visual-tactile communication. In this case, adults in hearing parent-deaf child dyads spent significantly more time in the auditory-visual-tactile modality than adults in hearing parent-hearing child dyads,  $U = 1$ ,  $p < 0.05$ .

We now turn to analyses of children's use of different modalities during joint attention. No instances of the tactile modality were observed, so this was excluded from further analysis. Beginning with the auditory modality, results demonstrated—not surprisingly—that children in hearing parent-hearing child dyads (that is, hearing children) spent a significantly greater proportion of time using the auditory modality than children in hearing parent-deaf child dyads (that is, deaf children),  $U = 14$ ,  $p < 0.05$ . Likewise, deaf children spent a significantly higher proportion of time using the visual modality than hearing children,  $U = 2$ ,  $p < 0.05$ . These differences in unimodal communication channel make sense and, we would argue, validate our measurement system (see **Table 4** for a summary of results; see **Table 5** for descriptive statistics).

Turning to multimodal comparisons, analyses revealed that hearing children of hearing parents spent a greater proportion of time in the auditory-visual modality than deaf children of hearing parents,  $U = 13.5$ ,  $p < 0.05$ . A comparison of use of the auditory-tactile combination of modalities revealed that hearing children spent significantly more time using it than deaf children,  $U = 10$ ,  $p < 0.05$ , as was the case for the visual-tactile combination as well,  $U = 13$ ,  $p < 0.05$ . Finally, we observed that deaf children spent significantly less time using

the auditory-visual-tactile combination than hearing children,  $U = 10.5$ ,  $p < 0.05$ .

## DISCUSSION

Our results highlight interesting differences in both unimodal and multimodal communication used during episodes of joint attention by parents and children in hearing-hearing and hearing-deaf dyads. Some of the results make sense; others are more surprising and, perhaps, concerning. At the very least, these data demonstrate the variability in accommodation made by parents across different parent-child dyads.

First, we found that hearing parents of hearing children spent a significantly greater proportion of time communicating with their children in both the auditory-only modality and the visual-only modality than hearing parents of deaf children. In other words, hearing parents of hearing children used more unimodal communication during joint attention episodes than hearing parents of deaf children. Of course, the shared use of oral language in hearing parent-hearing child dyads produced a richer body of linguistic interactions overall and, because rich linguistic interactions beget rich attentional interactions, joint attention is no doubt easier for these dyads to establish. The lack of complex, language-based interactions between hearing parents and their deaf children could explain some of the discrepancies in modality use between the two dyad types.

Second, in contrast to previous research showing that hearing parents tend to use the auditory modality most often when trying to engage their deaf children (Koester et al., 1998a), our findings revealed that hearing parents accommodate their deaf children's hearing status at least somewhat by engaging them via multiple modalities. In particular, adults in hearing parent-deaf child dyads spent a higher proportion of time using the audio-visual modality combination than those in hearing parent-hearing child dyads. However, the reverse pattern was observed for the

visual-tactile combination. Why are hearing parents of hearing children spending more time using this combination than hearing parents of deaf children? One possibility is that the hearing children in this study were simply engaged in more physical play, which elicited more tactile interaction with the parent. However, in instances during which three modality combinations were observed, they were produced by hearing parents of deaf children, a finding that is not consistent with such an interpretation. Regardless, the fact that parents in the mismatched dyads were more likely to use multiple modalities during communication than those in matched dyads demonstrates these parents' effort to accommodate their children's hearing status.

With regard to children's use of uni- and multimodal communication, we observed that deaf children spent a greater proportion of time than hearing children using only the visual modality. While this is not surprising given that the visual modality is accessible to a deaf child while the auditory is not, this raises the question of whether children are aware that their parents communicate differently than they do. Another item of note is that hearing children produced the only instances of the auditory-tactile combinations that we observed. When considering the parent and the child data, the overall pattern suggests that hearing parent-hearing child dyads were communicating more in general, an interpretation that is consistent with our finding that hearing parent-hearing child dyads spent a greater proportion of time *in* joint attention relative to hearing parent-deaf child dyads. Of course, the overall amounts of joint attention were small and so we do not wish to make too much of this difference. However, while the present study extends the body of research on this topic by further detailing modality use between the two dyad types, it raises several questions about the nature of communication

between hearing parents and their deaf children that merit further investigation. Thus, the preliminary findings of the present study should serve to motivate future research on this issue.

There are several additional factors that necessarily constrain interpretation of our results. First, it is important to note that the sample size is quite small. More observations are needed from more dyads of both types. Another shortcoming is that, although the deaf and hearing children were age-matched, the children are quite varied in age across the dyads. Free play with an 18-month-old is quite different from that with a 36-month-old. Thus, this variability undoubtedly influenced the findings of the present study. Moreover, the hearing parent-deaf child dyad with the oldest child produced the least amount of joint attention. Why? We can only speculate that this older child found the new toys provided in the study of great interest and willfully chose to focus on the toys rather than the parent. An examination of how parent-child interaction changes over time and relates across time would help clarify some of these questions, as well as facilitate more sophisticated and detailed understanding of the dynamics of age and interaction.

Nonetheless, while the present study has raised more questions than it has answered with regard to modality use in joint attention between parents and their children, it demonstrates that detailed coding of modality use in parent-child communication can provide important insights into how parents accommodate their children's particular communicative needs, whether they are hearing or deaf. This should motivate additional research of this type. Future studies will be needed to address not only *how* communication is facilitated in joint attention in the two types of dyads, but *what* is going on during these different types of engagement and how it affects the children's subsequent development.

## REFERENCES

- Barker, D. H., Quittner, A. L., Fink, N. E., Eisenberg, L. S., Tobey, E. A., Niparko, J. K., et al. (2009). Predicting behavior problems in deaf and hearing children: the influences of language, attention, & parent-child communication. *Dev. Psychopathol.* 21, 373–392. doi: 10.1017/S0954579409000212
- Brinck, I. (2001). Attention and the evolution of intentional communication. *Pragmat. Cogn.* 9, 259–277. doi: 10.1075/pc.9.2.05bri
- Cohn, J. F., and Tronick, E. Z. (1983). Three-month-old infants' reaction to simulated maternal depression. *Child Dev.* 54, 185–193. doi: 10.2307/1129876
- Corkum, V., and Moore, C. (1998). The origins of joint visual attention in infants. *Dev. Psychol.* 34, 28–38. doi: 10.1037/0012-1649.34.1.28
- DeMarco, I., Colle, L., and Bucciarelli, M. (2007). Linguistic and extralinguistic communication in deaf children. *J. Pragmat.* 39, 134–156. doi: 10.1016/j.cognition.2014.03.014
- Gale, E., and Schick, B. (2009). Symbol-infused joint attention and language use in mothers with deaf and hearing toddlers. *Am. Ann. Deaf* 153, 484–503. doi: 10.1353/aad.0.0066
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J. G. (2009). Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* 42, 377–381. doi: 10.1016/j.jbi.2008.08.010
- Hitermair, M. (2006). Parental resources, parental stress, and socioemotional development of deaf and hard of hearing children. *J. Deaf Stud. Deaf Educ.* 11, 493–513. doi: 10.1093/deafed/enl005
- Koester, L. S. (2001). Nonverbal communication between deaf and hearing infants and their parents: a decade of research. *Hrvatska Revija Rehabil. Istraž.* 37, 61–76.
- Koester, L. S., Brooks, L. R., and Karkowski, A. M. (1998a). A comparison of the vocal patterns of deaf and hearing mother-infant dyads during face-to-face interactions. *J. Deaf Stud. Deaf Educ.* 3, 290–301. doi: 10.1093/oxfordjournals.deafed.a014357
- Koester, L. S., Karkowski, A. M., and Traci, M. A. (1998b). How do deaf and hearing mothers regain eye contact when their infants look away? *Am. Ann. Deaf* 143, 5–13. doi: 10.1353/aad.2012.0071
- Lederberg, A. R., and Everhart, V. S. (1998). Communication between deaf children and their hearing mothers: the role of language, gesture, and vocalizations. *J. Speech Lang. Hear Res.* 41, 887–899. doi: 10.1044/jslhr.4104.887
- Lederberg, A. R., and Golbach, T. (2002). Parenting stress and social support in hearing mothers of deaf and hearing children: a longitudinal study. *J. Deaf Stud. Deaf Educ.* 7, 330–345. doi: 10.1093/deafed/7.4.330
- Lieberman, A. M., Hatrak, M., and Mayberry, R. I. (2014). Learning to look for language: development of joint attention in young deaf children. *Lang. Learn. Dev.* 10, 9–35. doi: 10.1080/15475441.2012.760381
- Mesman, J., van IJzendoorn, M. H., and Bakermans-Kranenburg, M. J. (2009). The many faces of the Still-Face Paradigm: a review and meta-analysis. *Dev. Rev.* 29, 120–162. doi: 10.1016/j.dr.2009.02.001
- Mitchell, R. E., and Karchmer, M. A. (2004). Chasing the mythical ten percent: parental hearing status of deaf and hard of hearing students in the United States. *Sign Lang. Stud.* 4, 138–163. doi: 10.1353/sls.2004.0005



- Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., and Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Dev.* 78, 938–954. doi: 10.1111/j.1467-8624.2007.01042.x
- Mundy, P., and Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Inf. Behav. Dev.* 21, 469–482. doi: 10.1016/S0163-6383(98)90020-0
- Mundy, P., Hogan, A., and Doehring, P. (1996). *A Preliminary Manual for the Abridged Early Social Communication Scales (ESCS)*. Available at: [www.ekgp.ugent.be/pages/nl/vragenlijsten/ESCS\\_Preliminary\\_Manual.doc](http://www.ekgp.ugent.be/pages/nl/vragenlijsten/ESCS_Preliminary_Manual.doc)
- Mundy, P., and Neal, A. R. (2000). Neural plasticity, joint attention, and a transactional social-orienting model of autism. *Int. Rev. Res. Ment. Retard.* 23, 139–168. doi: 10.1016/S0074-7750(00)80009-9
- Mundy, P., Sigman, M., and Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *J. Autism. Dev. Disord.* 20, 115–128. doi: 10.1007/BF02206861
- Nowakowski, M. E., Tasker, S. L., and Schmidt, L. A. (2009). Establishment of joint attention in dyads involving hearing mothers of deaf and hearing children, and its relation to adaptive social behavior. *Am. Ann. Deaf* 154, 15–29. doi: 10.1353/aad.0.0071
- Prezbindowski, A. K., Adamson, L. B., and Lederberg, A. R. (1998). Joint attention in deaf and hearing 22-month-old children and their hearing mothers. *J. Appl. Dev. Psychol.* 19, 377–387. doi: 10.1016/S0193-3973(99)80046-X
- Swisher, V. (1991). “Conversational interaction between deaf children and their hearing mothers: the role of visual attention,” in *Theoretical Issues in Sign Language Research*, eds S. D. Fisher and P. Siple (Chicago, IL: University of Chicago Press), 111–134.
- Tek, S. (2010). *A Longitudinal Analysis of Joint Attention and Language Development in Young Children with Autism Spectrum Disorders*. Doctoral Dissertations, University of Connecticut, Storrs, CT.
- Tomasello, M. (1995). “Joint attention as social cognition,” in *Joint Attention: Its Origins and Role in Development*, eds C. Moore and P. J. Dunham (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc), 103–130.
- Tomasello, M., and Carpenter, M. (2007). Shared intentionality. *Dev. Sci.* 10, 121–125. doi: 10.1111/j.1467-7687.2007.00573.x
- Trautman, C. H. (2009). *Early Lexical Acquisition in the Real World: Benefits of Child-Centered and Multimodal Input in the Absence of Coordinated Joint Attention*, doctoral dissertation, Retrieved from ERIC, The University of Texas at Dallas, Dallas, TX.
- Waxman, R. P., and Spencer, P. E. (1997). What mothers do to support infant visual attention: sensitivities to age and hearing status. *J. Deaf Stud. Deaf Educ.* 2, 104–114. doi: 10.1093/oxfordjournals.deafed.a014311
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). “ELAN: a professional framework for multimodality research,” in *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*, Genoa.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Depowski, Abaya, Oghalai and Bortfeld. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Effects of word-evoked object size on covert numerosity estimations

Magda L. Dumitru<sup>1\*</sup> and Gitte H. Joergensen<sup>2,3</sup>

<sup>1</sup> Department of Cognitive Science, Macquarie University, Sydney, NSW, Australia, <sup>2</sup> University of York, York, UK, <sup>3</sup> University of Connecticut, Storrs, CT, USA

We investigated whether the size and number of objects mentioned in digit-word expressions influenced participants' performance in covert numerosity estimations (i.e., property probability ratings). Participants read descriptions of big or small animals standing in short, medium, and long rows (e.g., *There are 8 elephants/ants in a row*) and subsequently estimated the probability that a health statement about them was true (e.g., *All elephants/ants are healthy*). Statements about large animals scored lower than statements about small animals, confirming classical findings that humans perceive groups of large objects as being more numerous than groups of small objects (Binet, 1890) and suggesting that object size effects in covert numerosity estimations are particularly robust. Also, statements about longer rows scored lower than statements about shorter rows (cf. Sears, 1983) but no interaction between factors obtained, suggesting that quantity information is not fully retrieved in digit–word expressions or that their values are processed separately.

**Keywords:** numerosity estimation, digit–word expression, numerical cognition, embodied cognition

## Introduction

People usually count concrete objects and living things and would rather speak of “8 baskets” or of “8 elephants” than simply of “8,” for instance. Despite their frequency, these complex numerical expressions composed of a digit followed by a word referring to a concrete object have been overlooked in current research on numerical cognition. The present study is the first to investigate whether word representations impact digit values to yield combined numerosity estimations. We are particularly interested in how robust these effects are as well as in whether the two magnitudes, for digits and for words, have distinct or shared conceptual and cortical representations when processed together.

Current behavioral and neural evidence suggests that numerical abilities are flexible and depend on context, habit, and cortical development (e.g., Dehaene, 1992; Lipton and Spelke, 2003; Siegler and Opfer, 2003; Cantlon et al., 2006). Moreover, numerical abilities are common across sensory modalities (cf. Walsh, 2003) and generate interaction and interference effects with space, time, size, and luminance (e.g., Pinel et al., 2004; Cohen Kadosh et al., 2007; Conson et al., 2008). Recent research investigating the common basis of numerical values and object size (i.e., Gabay et al., 2013) has confirmed that size-congruency effects are distinct from response-initiation effects triggered by the primary motor cortex (cf. Cohen Kadosh et al., 2007) and are truly conceptual in nature. Gabay et al. (2013) used equally-sized images of small and large animals in a parity-judgment task and reported that, in conditions where response conflict effects were controlled for, images of small animals primed small numbers whereas images of large animals primed large numbers.

## OPEN ACCESS

### Edited by:

Roberta Sellaro,  
Leiden University, Netherlands

### Reviewed by:

Anna M. Borghi,  
University of Bologna and Institute  
of Cognitive Sciences and  
Technologies, Italy  
Shai Gabay,  
University of Haifa, Israel

### \*Correspondence:

Magda L. Dumitru,  
Department of Cognitive Science,  
Macquarie University, 16 University  
Avenue, Sydney, NSW 2109, Australia  
magda.dumitru@gmail.com

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 13 April 2015

**Accepted:** 13 June 2015

**Published:** 03 July 2015

### Citation:

Dumitru ML and Joergensen GH  
(2015) Effects of word-evoked object  
size on covert numerosity estimations.  
*Front. Psychol.* 6:876.  
doi: 10.3389/fpsyg.2015.00876

These findings led us to hypothesize that, when processing complex numerical expressions such as “8 elephants” or “8 ants,” the size of the objects to which the nouns refer would exert a certain influence on people’s combined numerosity estimations that is, on their concluding that both expressions refer to eight objects. So, for example, we expect that even numerate adults might unduly estimate that rows of large objects (e.g., 8 elephants) contain more members than rows of small objects (e.g., 8 ants), whereby they would be tempted to combine the two magnitude types, for digits and for objects. Indeed, numerosity estimations of concrete objects vary with object size such that groups of small objects are judged to be less numerous than groups of large objects (cf. Binet, 1890). We therefore anticipate that merely mentioning a group of objects would evoke their combined size, which in turn would affect the overall numerosity estimations of digit-word expressions.

Language instantly evokes object properties including object size (Rubinsten and Henik, 2002; Setti et al., 2009; Sellaro et al., 2015), as predicted by theories of embodied and grounded cognition (Barsalou, 2008). These theories hold that people evoke multimodal perceptual simulations during online language processing based on their experience with concrete situations. Therefore, since language expressions are grounded in situations where people routinely use them, merely reading about an object is likely to evoke a full array of related experiences that gives instant access to associated perceptual and cognitive processes. Furthermore, results from brain imaging studies indicate that the same regions become active when objects are presented in pictorial form or when they are mentioned by language (e.g., Chao et al., 1999; Just et al., 2010). Research has also found that the retrieval of number magnitude is a spontaneous process similar to automatic language processing (Paivio, 1971; Barsalou, 2008) such that numbers are rapidly assigned approximate representations prior to further refinement in specific cortical areas (e.g., Tzelgov et al., 1992).

Among the studies devoted to investigating language-evoked object size, we recall the evidence reported in Rubinsten and Henik (2002), who used a Stroop-like paradigm to show that, in physical-comparison tasks (i.e., estimating which font size is larger) as well as in conceptual-comparison tasks (i.e., estimating which real-life animal is larger), judgments were faster for congruent animal names (e.g., “lion” written in large font or “ant” written in small font) than for incongruent names (e.g., “lion” written in small font or “ant” written in large font). Similar evidence was provided by Setti et al. (2009) who used an indirect task (i.e., category decision) asking participants to decide whether two objects evoked by a prime word and by a target word belonged to the same category. People responded faster to targets following same-size primes (e.g., “elephant” following “giraffe”) than to targets following different-size primes (e.g., “hare” following “giraffe”).

In our study, we used an indirect task (i.e., property probability ratings) to explore the hypothesis that object size affects numerosity estimations in digit-word expressions. We relied on a well-established finding that people tend to evaluate single entities more positively than groups (i.e., the “person-positivity

bias hypothesis” cf. Sears, 1983), which results in lower probability ratings for a particular property as groups grow larger. For example, when participants are presented with the information “There are 8 elephants in a row” or “There are 156 elephants in a row” and subsequently rate the probability that the statement “All elephants are healthy” is true, their scores should be lower for the statement about 156 elephants than for the about 8 elephants. We further predict that participants will rate small animals’ health higher than large animals’ health (e.g., “All ants are healthy” following “There are 8 ants in a row” would score higher than “All elephants are healthy” following “There are 8 elephants in a row”). In other words, adults might consider rows composed of large animals as being more numerous than rows composed of the same number of small animals and thus think of animals in “long” rows as being less healthy than animals in “short” rows. Object size effects may occur despite people’s ability to instantly recover the representation of the digit “8” in “8 ants” and “8 elephants,” for instance, because they are also able to rapidly evoke the size of the animals mentioned.

Our covert task (i.e., object-property probability ratings) taps into the later stages of combined magnitude processing hence obtaining significant effects of word-evoked object size on numerosity estimations would indicate that object-size effects are particularly robust. To further preclude confounds relating to whether size affects digit magnitude in virtue of the form of the statement rather than in virtue of the way that sentence fragments combine (i.e., jointly or independently), we varied the quantifier type to suggest aggregate (i.e., “All elephants are healthy”) as well as discrete numerosities (i.e., “Each elephant is healthy”).

## Materials and Methods

### Subjects

Fifty-two native English speakers volunteered for an online study in return for course credit.

### Stimuli

Stimuli were 36 sentences of which half included small animals (e.g., bats, mice, crabs) and half included large animals (e.g., tigers, bears, wolves), as determined from a previous rating study summarized in **Table 1**. Average ratings were calculated based on individual size ratings ( $N = 22$ ) of 100 items from two categories (animals and vegetables). Participants rated the size of each item presented individually in a scale from “0” (“not very big”) to “10” (“very big”). We then selected 36 items (i.e., names of small and large animals) from the rating study such that large animals received ratings at least twice as high as small animals and were also matched for frequency and length. Each sentence was followed by a statement about the health of the animals mentioned, as explained below. We constructed two lists (Latin square design) such that all participants saw each number once, paired with a small animal in the first list and with a large animal in the second list. In each list, half of the animals were small and the other half were large. Numbers ran from 3 to 8 in short rows, from 43 and 95 in medium rows, and from 1269 to 8421 in long rows. Both the numerosity study and the preliminary rating study

**TABLE 1 | Animal names in all stimuli statements used in the experimental study<sup>a</sup>.**

	Big Animals				Small Animals		
	Length	Frequency	Size		Length	Frequency	Size
Bears	5	3.17	7.77	Bats	4	2.74	2.95
Bisons	6	1.04	7	Bees	4	2.39	0.91
Camels	6	2.14	7.45	Beetles	7	1.76	1.14
Chimpanzees	11	1.6	5.86	Crabs	5	2.27	2.95
Crocodiles	10	1.85	6.59	Crickets	8	1.97	1.14
Deer	4	2.39	6.45	Doves	5	2.12	3
Donkeys	7	2.21	6.36	Finches	7	1.8	2.45
Foxes	5	2.61	4.59	Flies	5	3.32	0.55
Giraffes	8	1.76	8.32	Goldfish	8	2	1.77
Goats	5	2.53	4.82	Hamsters	8	1.75	2.32
Gorillas	8	2.19	7.05	Magpies	7	1.11	2.77
Hippos	6	1.6	7.77	Mice	4	2.6	2.68
Horses	6	3.19	7.05	Pigeons	7	2.26	2.95
Panthers	8	1.67	6.68	Rats	4	2.97	2.73
Reindeer	8	1.93	6.77	Robins	6	2.44	2.45
Tigers	6	2.64	7.09	Sparrows	8	1.77	2.36
Wolves	6	2.57	5.82	Spiders	6	2.36	1.45
Zebras	6	1.77	6.45	Squirrels	9	2.22	3.23
<b>Average Scores</b>	<b>6.72</b>	<b>2.16</b>	<b>6.66</b>	<b>Average Scores</b>	<b>6.22</b>	<b>2.21</b>	<b>2.21</b>

<sup>a</sup>20 participants rated the size of 36 animals on a scale from 1 (“not very big”) to 10 (“very big”). Names of big and small animals were matched in length and frequency.

were conducted in accordance with the ethics requirements of the University of York and followed relevant regulatory standards.

## Design and Procedure

The experiment followed a 2 (Size: Small vs. Large animals) × 3 (Row-length: Short vs. Medium vs. Long) fully factorial design. We also introduced “quantifier” as a between-subjects factor such that half of the participants read statements containing the quantifier *all* and the other half read statements containing the quantifier *each*. On a typical trial, participants read a description (e.g., *There are 3 crocodiles in a row*) followed by a statement (e.g., *All crocodiles are healthy* or *Each crocodile is healthy*), which they rated on a scale from 0 (“not very likely”) to 10 (“very likely”), as seen in **Figure 1**.

## Results

**Figure 2** summarizes the average likelihood scores across conditions. A 2 (Size: Small vs. Large animal) × 3 (Row-length: Short vs. Medium vs. Long) ANOVA revealed a main effect of size,  $F(1, 50) = 6.62$ ,  $p = 0.013$ ,  $\eta_p^2 = 0.117$ , and a main effect of row-length,  $F(2, 100) = 173.76$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.777$ , but no interaction between factors,  $F(2, 100) = 2.06$ ,  $p = 0.132$ , suggesting that group size as well as word-evoked object size influence property-probability ratings and thereby covert numerosity estimations.

We also calculated Cohen’s *d* for each row-length condition separately and found a sizeable difference between the effect size in the long-row condition and the effect size in the short- and medium-row conditions, namely a value of 0.311 for long rows, a value of 0.183 for short rows, a value of 0.123 for medium rows, suggesting the existence of a qualitative distinction between small and medium groups comprising at most tens of individuals on

the one hand, and very large groups comprising thousands of individuals on the other hand.

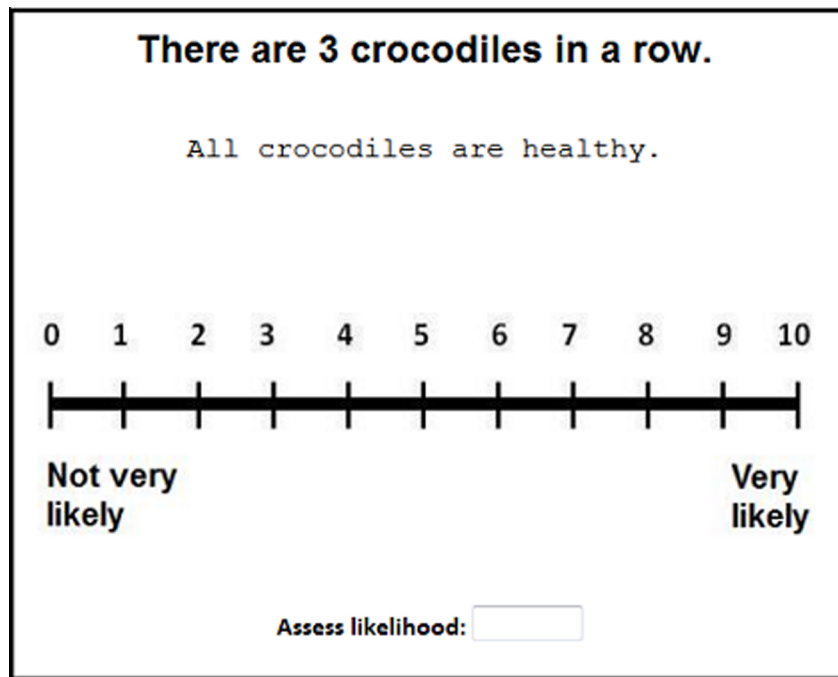
Importantly, we found no effect of quantifier type,  $F(1, 50) = 0.189$ ,  $p = 0.665$ , suggesting that magnitude estimations were not dependent on whether the quantifiers accompanying animal names prompted participants to view the groups (i.e., rows of animals) as aggregates (i.e., the quantifier “all”) or as discrete sums of individuals (i.e., the quantifier “each”).

## Discussion

We provided evidence that word-evoked object size impacts numerosity estimations in a covert task where participants rated the probability that several objects (i.e., 8 elephants) mentioned in a previous statement are healthy. We obtained a main effect of object size such that participants rated health statements about large animals lower than health statements about small animals, thereby confirming previous findings that language evokes object size, which in turn impacts number processing (Rubinsten and Henik, 2002; Setti et al., 2009; Gabay et al., 2013; Sellaro et al., 2015). Unsurprisingly (cf. Sears, 1983), we also obtained a main effect of group size such that health statements about long rows of animals scored lower than statements about medium rows, which in turn scored lower than statements about short rows.

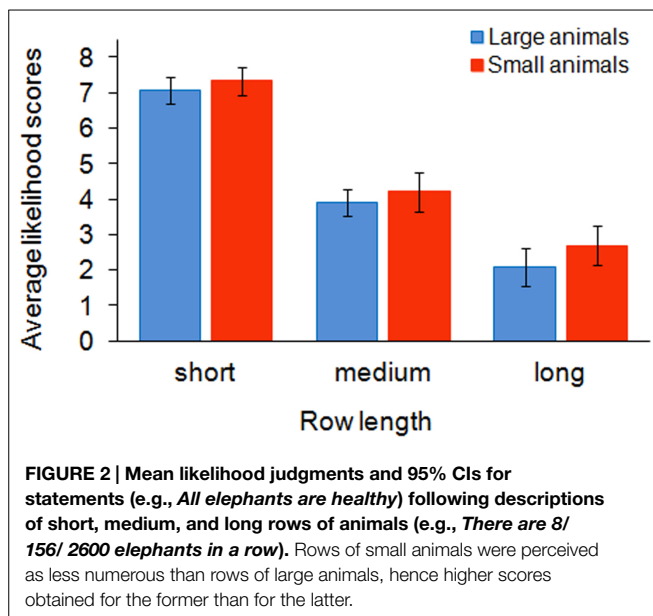
Interestingly, we observed no interaction between factors, which might suggest that quantity information is not fully retrieved in digit–word combinations or that digit and word magnitudes are processed separately at some level. Indeed, current evidence suggests that de-composition may occur for expressions containing same-type magnitude values, in particular for two-digit combinations (e.g., Nuerk et al., 2001) such that each digit is processed separately. Unfortunately, a decomposition account of same or different magnitude types runs counter previous evidence





**FIGURE 1 | Example stimulus in our study.** In each trial, participants read a description (e.g., *There are 3 crocodiles in a row*) followed by a statement (e.g., *All crocodiles are healthy*), whose likelihood they rated

on a scale from 0 (not very likely) to 10 (very likely). In half of the trials, the statement contained a different quantifier (e.g., *Each crocodile is healthy*).



**FIGURE 2 | Mean likelihood judgments and 95% CIs for statements (e.g., *All elephants are healthy*) following descriptions of short, medium, and long rows of animals (e.g., *There are 8/156/2600 elephants in a row*). Rows of small animals were perceived as less numerous than rows of large animals, hence higher scores obtained for the former than for the latter.**

(i.e., size congruency effects in reaction-time studies) supporting a shared magnitude code across quantity dimensions. Nevertheless, the predictions of the decomposition account and of the size-congruency principle could be reconciled if we examined more closely the particularities of our task and associated cognitive processes.

Most notably, the effect of object size is robust but small that is, numerical estimations of digit–word expressions are

largely determined by digit values, which are subsequently modulated by the size of a single object rather than by the combined size of a group whose cardinality matches the digit value. In other words, the plural form on the noun in “8 ants” does nothing to influence overall numerosity estimations, which suggests that language processing constraints might be responsible for the lack of interaction between object number and object size. In particular, linearity requires that items in a string be processed one by one in the order in which they are mentioned and is thus compatible with the so-called “anchoring bias” (cf. Tversky and Kahneman, 1974), which is a general-cognitive tendency toward grounding upcoming information into information already acquired. In digit-word expressions, the information provided by digit representations serves to anchor subsequent information provided by word representations, with lasting effects. In particular, our covert numerosity task (i.e., property probability ratings) explored the late combination stages of word-evoked object size and overall numerosity estimations rather than early behavioral reactions in item-by-item processing, as was the case in previous studies. The linearity constraint is likely to be responsible for the incomplete retrieval of quantity information. It is a matter for further research to confirm this hypothesis as well as whether full magnitude retrieval might be obtained for languages with a different word order, namely for languages where digits follow object names.

Let us now briefly consider the score differences between the short and medium row conditions on the one hand and the large row condition on the other hand, which were rather sizeable in the absence of a significant interaction between object

number and object size. We believe that these findings too are amenable to task properties, in particular to the stimuli used (i.e., digit magnitudes). Unlike previous studies where small numbers ran from 1 to 10 and large numbers would not surpass 100, our study included extremely large values (i.e., thousands) in the long-row category, which people might find less familiar or more difficult to grasp. The qualitative properties of very large magnitudes are likely to result from the comprehension effort they require, which might help explain why score differences between small and large animals were greatest in effortful trials (i.e., in the “long row” condition). By comparison, the tasks used in previous behavioral and neuro-cognitive studies reporting significant object size effects strongly evoked motor control and were thus inherently effortful. Importantly, effortful processing depends on participants’ goals hence specific cortical areas are recruited for handling the response types required. These findings suggest that the mapping between number magnitude and action representation is rather flexible (Koch and Prinz, 2005; Koch and Rumiati, 2006; Wenke and Frensch, 2005). Indeed, as shown in Fias et al. (2001) and in Lammertyn et al. (2002), effects of Spatial-Numerical Association Response Code (SNARC – e.g., Dehaene et al., 1993) were obtained only when participants judged the orientation of a digit, but not when they judged the color of the digit, arguably because the processing of numbers as well as orientation relies on regions of the parietal cortex, which belongs to the dorsal stream, while color processing relies mainly on regions of the inferior temporal cortex, which belongs to the ventral stream (Zeki et al., 1999). Since particular tasks involve different magnitude representations in the ventral and dorsal pathways, the extent of their neural overlap determines the interaction between numbers and action as well as between numbers and space (e.g., Badets et al., 2007).

In the present study, the object size effect as well as the qualitative difference between small and medium groups on the one hand and large groups on the other hand might stem from

a basic tendency toward translating different magnitude types onto each other as well as from an instant appraisal of the effort required for manipulating the objects, as predicted by theories of embodied cognition (Barsalou, 2008), thus engaging specific cortical pathways. It remains an issue for future research to carefully determine the relevance of the manipulability hypothesis (e.g., Moretto and Di Pellegrino, 2008; Badets and Pesenti, 2010, 2011; Ranzini et al., 2011) for the processing of digit-word expressions by varying response type and/or object affordability (e.g., manipulable vs. non-manipulable).

Numerate adults’ susceptibility to object-size biases also remains to be investigated in future research. Whereas it is widely acknowledged that the number sense is influenced by maturation levels, which generate differences in cortical activity between children and adults (Dehaene et al., 2003; Cantlon et al., 2006; Hyde et al., 2010), the extent to which maturation levels reflect expertise levels is largely unknown. The existence of correlations between maturation and expertise levels might help explain why children’s ability to discriminate numerosities and their capacity to map numbers onto distinct numerosities are not perfected before adolescence, once they have been exposed to a full range of numerical information (e.g., Lipton and Spelke, 2003). We believe that, in our study, adults’ numeracy expertise has prevented them from unduly concluding that the result of counting 8 elephants would be very different from the result of counting 8 ants, thus yielding only small effects of object size and no interaction between number and size. In other words, though object size exerted only a limited influence on adults’ numerosity estimations, it might have a greater impact on children and adults who lack extensive expertise with numerical calculations (e.g., tribal populations). The results of our study suggest that words can readily evoke object properties, which numerate adults factor in when making overt property likelihood judgments and thereby covert numerosity estimations.

## References

- Badets, A., Andres, M., Di Luca, S., and Pesenti, M. (2007). Number magnitude potentiates action judgments. *Exp. Brain Res.* 180, 525–534. doi: 10.1007/s00221-007-0870-y
- Badets, A., and Pesenti, M. (2010). Creating number semantics through finger movement perception. *Cognition* 115, 46–53. doi: 10.1016/j.cognition.2009.11.007
- Badets, A., and Pesenti, M. (2011). Finger-number interaction: an ideomotor account. *Exp. Psychol.* 58, 287–292. doi: 10.1027/1618-3169/a000095
- Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639
- Binet, A. (1890). La perception des longueurs et des nombres chez quelques petits enfants. *Revue Philosophique France L’Étranger* 30, 68–81.
- Cantlon, J. F., Brannon, E. M., Carter, E. J., and Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biol.* 4:e125. doi: 10.1371/journal.pbio.0040125
- Chao, L. L., Haxby, J. V., and Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat. Neurosci.* 2, 913–919. doi: 10.1038/13217
- Cohen Kadosh, R., Cohen Kadosh, K., Linden, D. E. J., Gevers, W., Berger, A., and Henik, A. (2007). The brain locus of interaction between number and size: a combined functional magnetic resonance imaging and event related potential study. *J. Cogn. Neurosci.* 19, 957–970. doi: 10.1162/jocn.2007.19.6.957
- Conson, M., Cinque, F., Barbarulo, A. M., and Trojano, L. (2008). A common processing system for duration, order and spatial information: evidence from a time estimation task. *Exp. Brain Res.* 187, 267–274. doi: 10.1007/s00221-008-1300-5
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N
- Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol. Gen.* 122, 371–393. doi: 10.1037/0096-3445.122.3.371
- Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239
- Fias, W., Lauwereyns, J., and Lammertyn, J. (2001). Irrelevant digits affect feature-based attention depending on the overlap of neural circuits. *Cogn. Brain Res.* 12, 415–423. doi: 10.1016/S0926-6410(01)00078-7
- Gabay, S., Leibovich, T., Henik, A., and Gronau, N. (2013). Size before numbers: conceptual size primes numerical value. *Cognition* 129, 18–23. doi: 10.1016/j.cognition.2013.06.001
- Hyde, D. C., Boas, D. A., Blair, C., and Carey, S. (2010). Near-infrared spectroscopy shows right parietal specialization for number in pre-verbal infants. *Neuroimage* 53, 647–652. doi: 10.1016/j.neuroimage.2010.06.030
- Just, M. A., Cherkassky, V. L., Aryal, S., and Mitchell, T. M. (2010). A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoS ONE* 5:e8622. doi: 10.1371/journal.pone.0008622

- Koch, I., and Prinz, W. (2005). Response preparation and code overlap in dual tasks. *Mem. Cogn.* 33, 1085–1095. doi: 10.3758/BF03193215
- Koch, I., and Rumiati, R. I. (2006). Task-set inertia and memory-consolidation bottleneck in dual tasks. *Psychol. Res.* 70, 448–458. doi: 10.1007/s00426-005-0020-8
- Lammertyn, J., Fias, W., and Lauwereyns, J. (2002). Semantic influences on feature based attention due to overlap of neural circuits. *Cortex* 38, 878–882. doi: 10.1016/S0010-9452(08)70061-3
- Lipton, J., and Spelke, E. (2003). Origins of number sense: large-number discrimination in human infants. *Psychol. Sci.* 14, 396–401. doi: 10.1111/1467-9280.01453
- Moretto, G., and Di Pellegrino, G. (2008). Grasping numbers. *Exp. Psychol.* 188, 505–515. doi: 10.1007/s00221-008-1386-9
- Nuerk, H. C., Weger, U., and Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition* 82, B25–B33. doi: 10.1016/S0010-0277(01)00142-1
- Paivio, A. (1971). *Imagery and Verbal Processes*. New York, NY: Holt, Rinehart & Winston.
- Pinel, P., Piazza, M., Le Bihan, D., and Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron* 41, 983–993. doi: 10.1016/S0896-6273(04)00107-2
- Ranzini, M., Lugli, L., Anelli, F., Carbone, R., Nicoletti, R., and Borghi, A. M. (2011). Graspable objects shape number processing. *Front. Hum. Neurosci.* 5:147. doi: 10.3389/fnhum.2011.00147
- Rubinsten, O., and Henik, A. (2002). Is an ant larger than a lion? *Acta Psychol.* 111, 141–154.
- Sears, D. O. (1983). The person-positivity bias. *J. Pers. Soc. Psychol.* 44, 233–250. doi: 10.1037/0022-3514.44.2.233
- Sellaro, R., Treccani, B., Job, R., and Cubelli, R. (2015). Spatial coding of object typical size: evidence for a SNARC-like effect. *Psychol. Res.* [Epub ahead of print].
- Setti, A., Caramelli, N., and Borghi, A. M. (2009). Conceptual information about size of objects in nouns. *Eur. J. Cogn. Psychol.* 21, 1022–1044. doi: 10.1080/09541440802469499
- Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–243. doi: 10.1111/1467-9280.02438
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124
- Tzelgov, J., Meyer, J., and Henik, A. (1992). Automatic and intentional processing of numerical information. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 166–179. doi: 10.1037/0278-7393.18.1.166
- Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space, and quantity. *Trends Cogn. Sci.* 7, 483–488. doi: 10.1016/j.tics.2003.09.002
- Wenke, D., and Frensch, P. A. (2005). The influence of task instruction on action coding: constraint setting or direct coding. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 803–819. doi: 10.1037/0096-1523.31.4.803
- Zeki, S., Aglioti, S., McKeefry, D., and Berlucchi, G. (1999). The neurological basis of conscious color perception in a blind patient. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14124–14129. doi: 10.1073/pnas.96.24.14124

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Dumitru and Joergensen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Encoding audio motion: spatial impairment in early blind individuals

Sara Finocchietti\*, Giulia Cappagli and Monica Gori

Science and Technology for Visually Impaired Children and Adults Group, Istituto Italiano di Tecnologia, Genoa, Italy

The consequence of blindness on auditory spatial localization has been an interesting issue of research in the last decade providing mixed results. Enhanced auditory spatial skills in individuals with visual impairment have been reported by multiple studies, while some aspects of spatial hearing seem to be impaired in the absence of vision. In this study, the ability to encode the trajectory of a 2-dimensional sound motion, reproducing the complete movement, and reaching the correct end-point sound position, is evaluated in 12 early blind (EB) individuals, 8 late blind (LB) individuals, and 20 age-matched sighted blindfolded controls. EB individuals correctly determine the direction of the sound motion on the horizontal axis, but show a clear deficit in encoding the sound motion in the lower side of the plane. On the contrary, LB individuals and blindfolded controls perform much better with no deficit in the lower side of the plane. In fact the mean localization error resulted  $271 \pm 10$  mm for EB individuals,  $65 \pm 4$  mm for LB individuals, and  $68 \pm 2$  mm for sighted blindfolded controls. These results support the hypothesis that (i) it exists a trade-off between the development of enhanced perceptual abilities and role of vision in the sound localization abilities of EB individuals, and (ii) the visual information is fundamental in calibrating some aspects of the representation of auditory space in the brain.

**Keywords:** auditory perception, blindness, spatial cognition, movement, early blind

## OPEN ACCESS

### Edited by:

Achille Pasqualotto,  
Sabanci University, Turkey

### Reviewed by:

Tina Iachini,  
Second University of Naples, Italy  
Maria J. S. Guerreiro,  
University of Hamburg, Germany

### \*Correspondence:

Sara Finocchietti,  
Science and Technology for Visually  
Impaired Children and Adults Group,  
Istituto Italiano di Tecnologia,  
Via Morego 30, 16163 Genoa, Italy  
sara.finocchietti@iit.it

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 18 June 2015

**Accepted:** 24 August 2015

**Published:** 07 September 2015

### Citation:

Finocchietti S, Cappagli G and Gori M  
(2015) Encoding audio motion: spatial  
impairment in early blind individuals.  
*Front. Psychol.* 6:1357.  
doi: 10.3389/fpsyg.2015.01357

## Introduction

Together with the visual information, audition provides important cues for the perception of object localization and movement. Visual and auditory spatial cues are usually associated. It has been demonstrated that our brain can increase spatial localization precision by integrating these two cues (Stein and Stanford, 2008). However, which is the role of vision on the development of auditory spatial skills is still unclear. Auditory space representation in visually deprived individuals has been extensively studied. The loss of vision results in changes in auditory perceptual abilities and in the way sounds are processed within the brain. An enhancement of certain aspects of spatial hearing and an impairment of some others have been observed in visually impaired individuals (Thinus-Blanc and Gaunet, 1997). The enhanced performance seems to be related to the recruitment of occipital areas deprived of their normal visual inputs (Gougoux et al., 2005; Collignon et al., 2009). Early-blind subjects are properly able to form spatial topographical maps (Tinti et al., 2006; Fortin et al., 2008), express superior auditory pitch discrimination (Gougoux et al., 2004), and can map the auditory environment with superior accuracy (Lessard et al., 1998; Voss et al., 2004). However, neurophysiological studies support the hypothesis of auditory impairment in absence of vision, showing that vision drives the maturation of auditory spatial properties of superior colliculus neurons (King et al., 1988; King, 2009). The superior sound localization accuracy has usually been reported only for peripheral rather than for central regions of space (Röder et al., 1999;



Voss et al., 2004) and for monaural testing conditions (Lessard et al., 1998; Gougoux et al., 2005). Furthermore, the localization in the mid-sagittal plane (Zwiers et al., 2001; Lewald, 2002) and the performance of more complex tasks requiring a metric representation of the auditory space (Gori et al., 2010, 2013) tend to be worse in these subjects than in sighted controls.

Importantly, most of these studies investigated spatial skills of blind using static stimuli. In contrast, the dynamic localization of sounds – which requires a continuous encoding in time and space of a moving sound source – has been largely neglected in the literature, with only a few studies investigating it. Poirier et al. (2006) showed that blind individuals can determine both the nature of a sound stimulus (pure tone or complex sound) and the presence or absence of its movement. Lewald (2013) showed that visually deprived individuals were superior in judging the direction of a sound motion on the horizontal direction. These two studies investigate simple aspects of dynamic sound evaluation like its presence and its direction. Both these tasks do not require the presence of a metric representation of space. Since it has been shown that blind individuals results impaired in performing tasks that require a metric representation of the auditory space (Gori et al., 2013), one may expect to find an impairment when a more complex auditory dynamic task, like the capability of blind individuals to completely reproduce a continuous dynamic sound and to determine its end point, is evaluated. While the discrimination of sound direction can be evaluated by comparing the position of the two sounds in a relative way, the reproduction and definition of a sound end point requires the creation of a complex Euclidian map which considers the relationship between sounds positions in space and time [as well as it occurs in the space bisection task, see (Gori et al., 2013)].

For this reason, we studied the ability of early and late blind (EB and LB) individuals and of sighted blindfolded controls, to encode the trajectory of a 2-dimensional sound motion, reproducing the complete movement, and reaching the correct end-point sound position.

We advance the hypothesis that (i) EB, LB, and sighted blindfolded individuals are able to correctly determine the direction of sound motion on the horizontal direction (as previously shown by (Lewald, 2013), but not in the vertical direction; (ii) contrary to LB and sighted blindfolded individuals, EB individuals are impaired in encoding the complete trajectory and in correctly localizing the end-point sound position.

## Materials and Methods

### Subjects

Forty participants have been enrolled in the study: EB ( $N = 12$ , 7 females; average age  $34 \pm 11$  years old), LB ( $N = 8$ , 3 females; average age:  $33 \pm 13$  years old), and sighted blindfolded controls ( $N = 20$ , 11 females; average age:  $32 \pm 13$  years old). All the participants had similar education (at least an Italian high school diploma, indicating 13 years of school). Clinical details regarding the blind participants are presented in **Table 1**. All the EB participants were blind at birth. All the

**TABLE 1 | Clinical details of the early blind (EB) and late blind (LB) participants.**

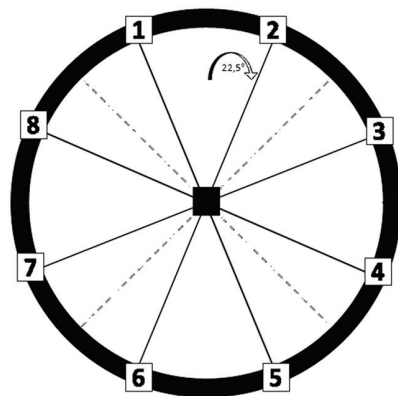
Participant	Age at test	Gender	Pathology	Age complete blindness
<b>EB</b>				
#EB1	21	Female	Congenital glaucoma and retinal detachment	Birth
#EB2	25	Female	Retinopathy of prematurity	Birth
#EB3	26	Female	Retinopathy of prematurity	Birth
#EB4	20	Female	Congenital Glaucoma	Birth
#EB5	36	Male	Retinopathy of prematurity	Birth
#EB6	37	Female	Retinopathy of prematurity	Birth
#EB7	49	Male	Retinopathy of prematurity	Birth
#EB8	32	Female	Congenital cataract	Birth
#EB9	26	Female	Retinopathy of prematurity	Birth
#EB10	56	Male	Retinopathy of prematurity	Birth
#EB11	56	Male	Congenital Glaucoma	Birth
#EB12	42	Male	Leber amaurosis	Birth
<b>LB</b>				
#LB1	27	Male	Corneal opacity	17
#LB2	45	Female	Leber amaurosis	40
#LB3	65	Male	Glaucoma	14
#LB4	25	Male	Retinal detachment	22
#LB5	22	Female	Bilateral uveitis	12
#LB6	27	Male	Damage to the optic nerve	10
#LB7	24	Male	Retinal detachment	13
#LB8	37	Female	Exudative retinopathy	14

*The table shows the age at test, the gender, the pathology, and the age since they became completely blind.*

participants had no history of hearing impairment and were right handed. The handedness was defined by the Edinburgh handedness inventory (Oldfield, 1971). The participants provided written informed consent in accordance with the Declaration of Helsinki. The study was approved by the ethics committee of the local health service (*Comitato Etico, ASL3 Genovese, Italy*).

### Set-Up and Protocol

The experiment was performed in a dark room. The apparatus consisted of a graduated circular perimeter (radius = 45 cm) mounted on a wooden panel positioned in front of the participant on the frontal plane. Eight different positions were marked on the perimeter, starting at  $22.5^\circ$  and increasing of  $45^\circ$  (**Figure 1**). Sighted participants were blindfolded before entering the experimental room. Each participant was seated, the center of the circle corresponding to the tip of his nose, and was able to comfortably reach and explore with their hand the graduated circular perimeter. Two experimenters instructed the participant and performed all the experiments (SF, GC). The two experimenters were previously trained to perform the task as similar as possible, so that the movement's velocity was consistent across trials, positions, and groups. The experimenter was seated opposite to the participant, holding a sound source. The sound source was a digital metronome (single pulse at



**FIGURE 1 | Experimental set-up.** The graduated circular perimeter (radius = 45 cm) is mounted on a wooden panel positioned in front of the participant on the frontal plane. The eight different positions are marked on the perimeter starting at 22.5° and increasing of 45°.

500 Hz, intermittent sound at 180 bpm) and was clearly audible by every participant. A spherical marker was mounted on the distal phalanges of the index finger on both the participant and experimenter for motion tracking (Vicon Motion Systems Ltd., UK). The experimenter moved the sound source from the center of the plane toward one of the possible positions highlighted on the circular perimeter in a randomized order. The participant was instructed to keep his index finger pointed to the center until the end of the audio motion. He then had to reproduce the complete trajectory, reach the estimated sound end-point position, and return to the original central position. The movement was performed at participant's own pace. All the eight positions were reached five times, for a total of 40 trials per participant.

### Data Analysis

Kinematic data were post-processed and analyzed using Matlab (R2013a, The MathWorks, USA). The spatial accuracy, indicated by localization bias and localization error, was computed for each participant and for each spatial position. Each end-point position was computed as the average of the last 10 samples and normalized on the origin position (the center of the circumference), in order to avoid movement's errors. The localization bias represents the average position in the space of the end-point reached by the participant. The localization error was calculated as the Euclidean distance (in mm) between the end-point position reached by the participant and the one reached by the experimenter. This error was averaged on the number of trials per position and on the number of participants. In order to evaluate top-bottom and left-right judgments, the end-point positions of the experimenter and the participants were categorized as follows: 1 = top, related to position 1–2, and with ordinate value higher than 0; 2 = right, related to position 3–4, and with abscissae value higher than 0; 3 = bottom, related to position 5–6, and with ordinate value less than 0; 4 = left, related to position 7–8, and with abscissae value less than 0. The correct

direction of judgment was defined as the difference between the experimenter and participant categorization was used for further analysis.

### Statistics

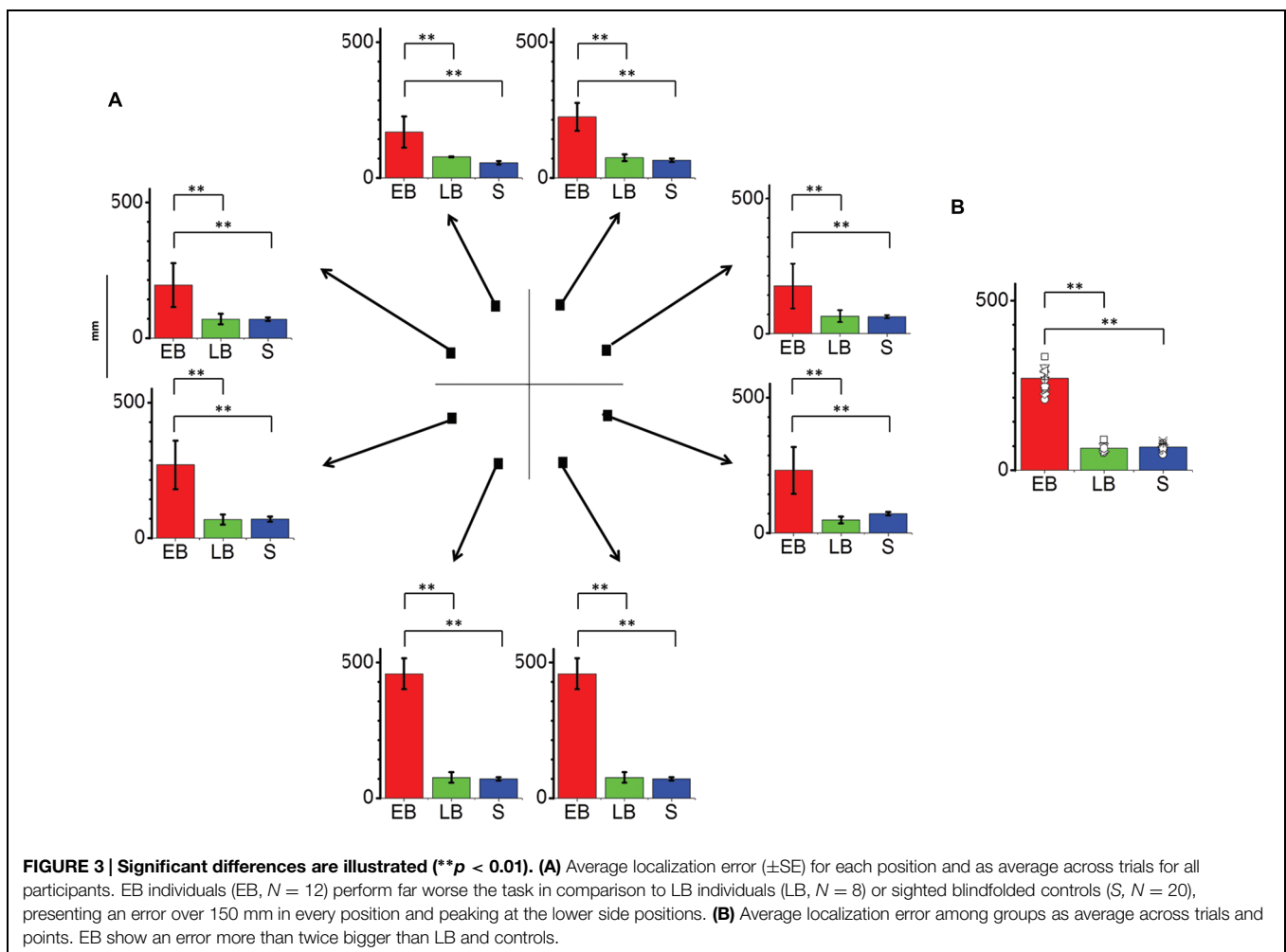
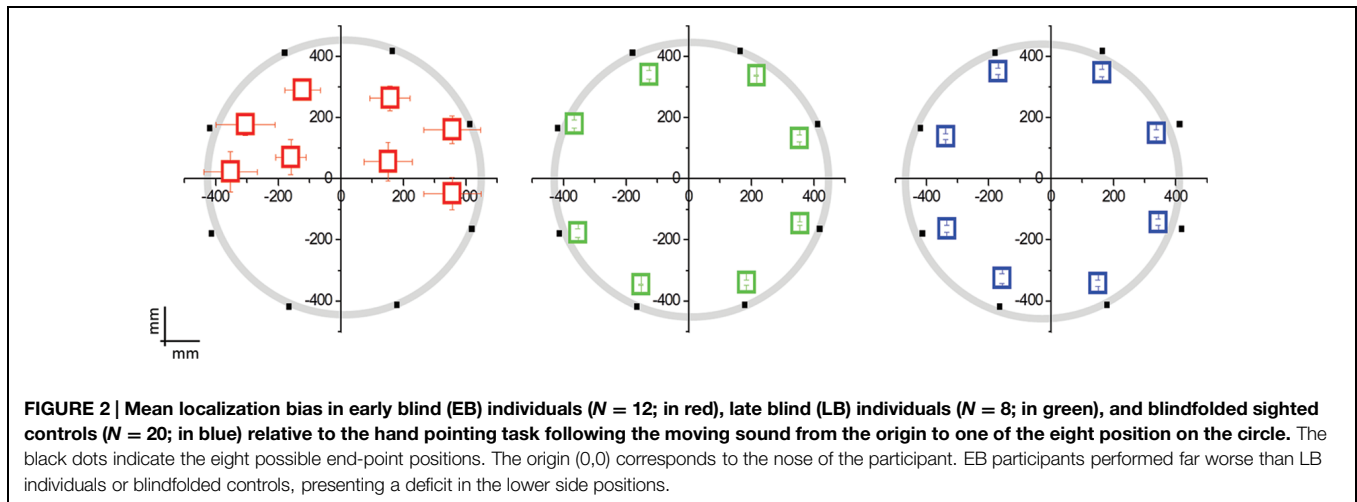
Data were normally distributed, confirmed by visual inspection of Q–Q plots. Data are presented as mean and SE. Localization bias was analyzed by two separate factorial ANOVA (one for the abscissae value, one for the ordinate) with factors participant group (EB, LB, controls), and trajectory (experimenter, participant). The Levene's test for homogeneity of variance was used to compare EB and LB. In order to evaluate the left–right and top–bottom judgments, a factorial ANOVA of the correct direction judgment, with factor participant group (EB, LB, controls), and panel area (top, left, right, bottom) was performed. The localization error was analyzed by a factorial ANOVAs, with between factors participant group (EB, LB, controls), and point (1–8). The mean velocity was analyzed by a one-way ANOVA, with between factor participant group (EB, LB, controls). The Bonferroni *post hoc* test was used in the case of significant factors.  $P < 0.05$  was considered significant.

### Results

As can be observed in **Figures 2 and 3**, the pattern of results for the EB is completely different with respect to the ones for the other two groups.

#### Localization Bias

The interaction Group  $\times$  Point resulted significant for both the abscissae ( $F_{14,296} = 5.14$ ;  $P = 0.001$ ) and the ordinate value ( $F_{14,296} = 33.76$ ;  $P = 0.001$ ). LB individuals and sighted individuals do not show any localization bias for both the abscissae and ordinate value (Levene's:  $F_{1,222} = 0.005$ ;  $P = 0.94$ ): their responses are superimposed with the physical



endpoint position ( $P > 0.1$ ; back dots in **Figure 2**). On the contrary, EB individuals showed a strong localization bias and a general compression of the targets toward the upper part of the space (Point 5 and 6, ordinate value,  $P < 0.003$ ).

The interaction Group  $\times$  Panel area of the correct motion judgment resulted significant ( $F_{6,308} = 24.27$ ;  $P = 0.001$ ). In agreement with previous results (Lewald, 2013), the left/right motion judgment did not show any statistical difference among the three groups ( $P > 0.05$ ), EB individuals, LB individuals

and controls were able to correctly judge the stimulus direction in the horizontal axis. On the opposite, the top/bottom motion judgment show statistical difference among the three groups, as EB individuals were not able to correctly judge the stimulus direction in the vertical axis for the bottom positions ( $P < 0.001$ ).

### Localization Error

The interaction between group and point resulted significant ( $F_{14,296} = 17.10$ ,  $P = 0.01$ ). In fact the average localization error (Figure 3) on lower side positions was more than 400 mm compared to less than 100 mm for both LB individuals and blindfolded controls, respectively. On the opposite, LB individuals performed equal to blindfolded healthy participants, as no statistical difference was present in both localization bias and error ( $P > 0.1$ ).

### Velocity

Every participant was free to perform the movement at his own pace, but no difference in mean velocity between groups was observed ( $F_{2,317} = 0.51$ ;  $P > 0.1$ ).

## Discussion

We present the first study whose aim is to evaluate the dynamic audio localization in visually impaired individuals with a task requiring a continuous encoding in time and space of a sound source in the sagittal plane. This is a complex task that requires the ability to distinguish the spatio-temporal change imposed on moving sounds in space by the dynamic filtering mechanism of the two external ears from the intrinsic spectral structure of the sound (Wightman and Kistler, 1989; Hofman et al., 1998).

Early blind individuals result impaired in performing this task, which results more complex than a static localization task, and they show a clear deficit in encoding the sound motion in the lower side of the plane. On the contrary, LB individuals and blindfolded controls perform much better with no deficit in the lower side of the plane. In agreement with previous results (Lewald, 2013), no deficit was observed in EB subjects for the identification of sound direction.

Some studies suggest that the absence of vision does not impact audio perception in visually impaired humans (Lessard et al., 1998; Röder et al., 1999; Lewald, 2013) and animals (Rauschecker et al., 1995; King and Parsons, 1999). These auditory spatial abilities are more remarkable in peripheral than in central regions of space and in the horizontal plane (King and Parsons, 1999; Röder et al., 1999; Voss et al., 2004). In contrast, localization in the mid-sagittal plane tends to be worse in blind individuals than in sighted controls (Zwiers et al., 2001; Lewald, 2002). A possible explanation about the different static localization ability in the horizontal vs. mid-sagittal plane is that the vertical localization is based primarily on spectral cues that are mainly guided by vision (Tollin et al., 2013). In addition visually impaired individuals, especially EB individuals, show impairments in performing more

complex tasks that require a metric representation of space (Gori et al., 2013).

In the first years of life, the brain continuously needs to calibrate the developing system (Gori et al., 2008, 2010; Nardini et al., 2008; Gori, 2015). In case of a sensory loss, such as vision in EB individuals, the important communication between sensory modalities cannot occur (Warren and Pick, 1970) and this can directly affect the development of the audio spatial maps in the superior colliculus (King et al., 1988). While the development of a complex Euclidian representation of space is compromised in absence of vision from birth (Gori et al., 2013), results obtained in LB individuals suggest that even a short early visual experience can guarantee this representation (Fine et al., 2003). Some EB individuals can partly build a representation of space in the case of simple audio spatial tasks, like monaural static sound localization (Röder et al., 1999; Voss et al., 2004), with changes within the auditory pathway and the recruitment of the visual cortex (Merabet and Pascual-Leone, 2009; Bavelier and Hirshorn, 2010). In our case, both early and LB individuals show, in agreement with previous studies (Poirier et al., 2006; Lewald, 2013), an equal auditory motion perception on the horizontal axis. Our task resulted more complex as it required the ability to relate sound positions in a two dimensional space and time; in this case other brain areas cannot intervene, and EB individuals clearly result impaired. What is the reason for this? When the visual calibration is not possible, audio spatial information may be self-calibrated by the auditory system. This audio self-calibration is limited by: (i) the physiology of the auditory system and associated processing of the audio signal; and (ii) the audio environmental statistics.

First, like in the case of the elevation-related spectral cues (Zwiers et al., 2001; Lewald, 2002), the auditory system is not equally good in perceiving sounds coming from the frontal or from the peripheral plane (King, 2009). This suggests a trade-off in the localization proficiency between the two auditory spatial planes that has recently proposed for a static auditory localization task (Voss et al., 2015). The ability to perform such a complex task may then require a full development of the audio spatial maps in superior colliculus, where signals from the different senses are combined and used to guide adaptive motor responses.

Second, in the peri-personal space, the most frequent dynamic sounds we face with are the ones related to individuals speaking around us, sounds that generally are at our height. Recent findings show that the natural auditory scene statistics shapes human spatial hearing, suggesting that both sound localization behavior and ear anatomy are fine-tuned to the statistics of natural auditory scenes (Parise et al., 2014). This statistical environmental cue may then affect the way blind individuals built their spatial representation.

## Conclusion

The absence of spatial references from the visual inputs has widespread consequences on the brain; the important



communication between sensory modalities cannot be created, therefore auditory space perception can only rely on the physiological and statistical information heterogeneity. This information results insufficient in dynamic localization tasks, as the one presented here, producing direct impairments on auditory space cognition in blind individuals.

## References

- Bavelier, D., and Hirshorn, E. A. (2010). I see where you're hearing: how cross-modal plasticity may exploit homologous brain structures. *Nat. Neurosci.* 13, 1309–1311. doi: 10.1038/nn1110-1309
- Collignon, O., Voss, P., Lassonde, M., and Lepore, F. (2009). Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. *Exp. Brain Res.* 192, 343–358. doi: 10.1007/s00221-008-1553-z
- Fine, I., Wade, A. R., Brewer, A. A., May, M. G., Goodman, D. F., Boynton, G. M., et al. (2003). Long-term deprivation affects visual perception and cortex. *Nat. Neurosci.* 6, 915–916. doi: 10.1038/nn1102
- Fortin, M., Voss, P., Lord, C., Lassonde, M., Pruessner, J., Saint-Amour, D., et al. (2008). Wayfinding in the blind: larger hippocampal volume and supranormal spatial navigation. *Brain* 131, 2995–3005. doi: 10.1093/brain/awn250
- Gori, M. (2015). Multisensory integration and calibration in children and adults with and without sensory and motor disabilities. *Multisens. Res.* 28, 71–99. doi: 10.1163/22134808-00002478
- Gori, M., Del Viva, M., Sandini, G., and Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Curr. Biol.* 18, 694–698. doi: 10.1016/j.cub.2008.04.036
- Gori, M., Sandini, G., Martinoli, C., and Burr, D. (2010). Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Curr. Biol.* 20, 223–225. doi: 10.1016/j.cub.2009.11.069
- Gori, M., Sandini, G., Martinoli, C., and Burr, D. C. (2013). Impairment of auditory spatial localization in congenitally blind human subjects. *Brain* 137, 288–293. doi: 10.1093/brain/awt311
- Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R. J., and Belin, P. (2004). Neuropsychology: pitch discrimination in the early blind. *Nature* 430, 309. doi: 10.1038/430309a
- Gougoux, F., Zatorre, R. J., Lassonde, M., Voss, P., and Lepore, F. (2005). A functional neuroimaging study of sound localization: visual cortex activity predicts performance in early-blind individuals. *PLoS Biol.* 3:e27. doi: 10.1371/journal.pbio.0030027
- Hofman, P. M., Van Riswick, J. G., and Van Opstal, A. J. (1998). Relearning sound localization with new ears. *Nat. Neurosci.* 1, 417–421. doi: 10.1038/1633
- King, A. J. (2009). Visual influences on auditory spatial learning. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 331–339. doi: 10.1098/rstb.2008.0230
- King, A. J., Hutchings, M. E., Moore, D. R., and Blakemore, C. (1988). Developmental plasticity in the visual and auditory representations in the mammalian superior colliculus. *Nature* 332, 73–76. doi: 10.1038/332073a0
- King, A. J., and Parsons, C. H. (1999). Improved auditory spatial acuity in visually deprived ferrets. *Euro. J. Neurosci.* 11, 3945–3956. doi: 10.1046/j.1460-9568.1999.00821.x
- Lessard, N., Pare, M., Lepore, F., and Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature* 395, 278–280. doi: 10.1038/26228
- Lewald, J. (2002). Vertical sound localization in blind humans. *Neuropsychologia* 40, 1868–1872. doi: 10.1016/S0028-3932(02)00071-4
- Lewald, J. (2013). Exceptional ability of blind humans to hear sound motion: implications for the emergence of auditory space. *Neuropsychologia* 51, 181–186. doi: 10.1016/j.neuropsychologia.2012.11.017
- Merabet, L. B., and Pascual-Leone, A. (2009). Neural reorganization following sensory loss: the opportunity of change. *Nat. Rev. Neurosci.* 11, 44–52. doi: 10.1038/nrn2758

## Acknowledgments

This research was supported by the EU Project ABBI (FP7-ICT 611452). The authors have no conflict of interest to report. The authors would like to thank Prof. Giulio Sandini and Ing. Marco Jacono for their support during the development of the study, and Prof. Flavio Parmiggiani for proofreading the manuscript.

- Nardini, M., Jones, P., Bedford, R., and Braddick, O. (2008). Development of cue integration in human navigation. *Curr. Biol.* 18, 689–693. doi: 10.1016/j.cub.2008.04.021
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Parise, C. V., Knorre, K., and Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6104–6108. doi: 10.1073/pnas.1322705111
- Poirier, C., Collignon, O., Scheiber, C., Renier, L., Vanlierde, A., Tranduy, D., et al. (2006). Auditory motion perception activates visual motion areas in early blind subjects. *Neuroimage* 31, 279–285. doi: 10.1016/j.neuroimage.2005.11.036
- Rauschecker, J. P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268, 111–114. doi: 10.1126/science.7701330
- Röder, B., Teder-Sälejärvi, W., Sterr, A., Rösler, F., Hillyard, S. A., and Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature* 400, 162–166. doi: 10.1038/22106
- Stein, B. E., and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.* 9, 255–266. doi: 10.1038/nrn2331
- Thinus-Blanc, C., and Gaunet, F. (1997). Representation of space in blind persons: vision as a spatial sense? *Psychol. Bull.* 121, 20–42. doi: 10.1037/0033-2909.121.1.20
- Tinti, C., Adenzato, M., Tamietto, M., and Cornoldi, C. (2006). Visual experience is not necessary for efficient survey spatial cognition: evidence from blindness. *Q. J. Exp. Psychol.* 59, 1306–1328. doi: 10.1080/17470210500214275
- Tollin, D. J., Ruhland, J. L., and Yin, T. C. (2013). The role of spectral composition of sounds on the localization of sound sources by cats. *J. Neurophysiol.* 109, 1658–1668. doi: 10.1152/jn.00358.2012
- Voss, P., Lassonde, M., Gougoux, F., Fortin, M., Guillemot, J.-P., and Lepore, F. (2004). Early- and late-onset blind individuals show supra-normal auditory abilities in far-space. *Curr. Biol.* 14, 1734–1738. doi: 10.1016/j.cub.2004.09.051
- Voss, P., Tabry, V., and Zatorre, R. J. (2015). Trade-off in the sound localization abilities of early blind individuals between the horizontal and vertical planes. *J. Neurosci.* 35, 6051–6056. doi: 10.1523/JNEUROSCI.4544-14.2015
- Warren, D. H., and Pick, H. L. (1970). Intermodality relations in localization in blind and sighted people. *Percept. Psychophys.* 8, 430–432. doi: 10.3758/BF03207040
- Wightman, F. L., and Kistler, D. J. (1989). Headphone simulation of free-field listening. I: stimulus synthesis. *J. Acous. Soc. Am.* 85, 858–867. doi: 10.1121/1.397557
- Zwiers, M., Van Opstal, A., and Cruysberg, J. (2001). A spatial hearing deficit in early-blind humans. *J. Neurosci.* 21, 141–145.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Finocchietti, Cappagli and Gori. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Voluntary initiation of movement: multifunctional integration of subjective agency

Patrick Grüneberg<sup>1,2\*</sup>, Hideki Kadone<sup>3</sup> and Kenji Suzuki<sup>4</sup>

<sup>1</sup> School of Global Japanese Studies, Meiji University, Tokyo, Japan, <sup>2</sup> Artificial Intelligence Laboratory, University of Tsukuba, Tsukuba, Japan, <sup>3</sup> Center for Innovative Medicine and Engineering, University of Tsukuba Hospital, Tsukuba, Japan, <sup>4</sup> Center for Cybernetics Research, University of Tsukuba, Tsukuba, Japan

## OPEN ACCESS

### Edited by:

Achille Pasqualotto,  
Sabanci University, Turkey

### Reviewed by:

Paul Van Schaik,  
Teesside University, UK  
Derrick L. Hassert,  
Trinity Christian College, USA

### \*Correspondence:

Patrick Grüneberg,  
Artificial Intelligence Laboratory,  
University of Tsukuba, 1-1-1 Tennodai,  
Tsukuba 305-8573, Japan  
patrick@ai.iit.tsukuba.ac.jp

### Specialty section:

This article was submitted to  
Cognition, a section of the journal  
Frontiers in Psychology

**Received:** 15 March 2015

**Accepted:** 10 May 2015

**Published:** 22 May 2015

### Citation:

Grüneberg P, Kadone H and Suzuki K  
(2015) Voluntary initiation of  
movement: multifunctional integration  
of subjective agency.  
Front. Psychol. 6:688.  
doi: 10.3389/fpsyg.2015.00688

This paper investigates subjective agency (SA) as a special type of efficacious action consciousness. Our central claims are, firstly, that SA is a conscious act of voluntarily initiating bodily motion. Secondly, we argue that SA is a case of multifunctional integration of behavioral functions being analogous to multisensory integration of sensory modalities. This is based on new perspectives on the initiation of action opened up by recent advancements in robot assisted neuro-rehabilitation which depends on the active participation of the patient and yields experimental evidence that there is SA in terms of a conscious act of voluntarily initiating bodily motion (phenomenal performance). Conventionally, action consciousness has been considered as a sense of agency (SoA). According to this view, the conscious subject merely echoes motor performance and does not cause bodily motion. Depending on sensory input, SoA is implemented by means of unifunctional integration (binding) and inevitably results in non-efficacious action consciousness. In contrast, SA comes as a phenomenal performance which causes motion and builds on multifunctional integration. Therefore, the common conception of the brain should be shifted toward multifunctional integration in order to allow for efficacious action consciousness. For this purpose, we suggest the heterarchic principle of asymmetric reciprocity and neural operators underlying SA. The general idea is that multifunctional integration allows conscious acts to be simultaneously implemented with motor behavior so that the resulting behavior (SA) comes as efficacious action consciousness. Regarding the neural implementation, multifunctional integration rather relies on operators than on modular functions. A robotic case study and possible experimental setups with testable hypotheses building on SA are presented.

**Keywords:** subjectivity, agency (psychology), motion, multimodality, multifunctionality, neurorehabilitation, assistive robotics

## 1. Introduction

The concept of multisensory (multimodal) integration emerged as an alternative approach to the problem of sensory integration, i.e., how different sensory modalities interact in order to form coherent representations of objects or processes underlying sensory input. According to the standard view, sensory modalities are processed independently in their respective brain areas and later on integrated by means of binding. As this kind of integration depends on single

modalities the standard approach can be referred to as *unimodal integration* or just *binding*. By postulating multisensory neurons, this standard view has been challenged (Calvert and Thesen, 2004, Alais et al., 2010). Multimodality implies that there is no one-to-one mapping of sensory input to a certain brain area. Instead, different sensory modalities can be processed by one and the same area so that integration already takes place at the primary level of sensory processing. This new perspective, referred to as *multimodal (multisensory) integration*, contains far-reaching implications for the functional organization of the brain as well as for the cognitive and phenomenal (conscious) aspects of sensory processing and action control (Musseler et al., 2014).

While unimodal and multimodal integration are usually concerned with the processing of sensory input and perceptual consciousness, an analogous point for shifting from a unimodal to a multimodal setup can be made in the case of action consciousness. As action consciousness mainly builds on behavioral functions of action initiation and control, the uni-/multimodal-distinction turns into the distinction between *unifunctional* and *multifunctional integration* of behavioral functions. Unifunctional integration (binding) is mainly applied in order to explain the phenomenal experience of action consciousness in terms of a sense of agency (SoA), authorship, or control (Gallagher, 2000, Gallagher, 2012)<sup>1</sup>. According to SoA as an experiential concept, current phenomenology of action as well as psychological and neurocognitive research on action do not leave any space for bodily motion being initiated by the conscious agent. Even if voluntary initiation is well-known in neuroscientific research on motion, the question “Does Consciousness Cause Behavior?” (Pockett et al., 2006) tends to be negatively answered. While initiation is usually left to the locomotor system, the conscious agent is limited to experiencing action *post-hoc*. Therefore, the conscious subject merely echoes motor performance (Haggard and Johnson, 2003) and is not regarded to be an efficacious agent who causes bodily motion (Bayne and Levy, 2009). In this view, action consciousness is an epiphenomenal addition to sub-personal processes of the locomotor system.

Opposed to this common view, robot-assisted rehabilitation (Tejima, 2001, Feil-Seifer and Mataric, 2005) opens up a new perspective on the phenomenology of action. The rehabilitative application of robotic devices which crucially depend on the active participation of the patient (Hogan et al., 2006, Duschau-Wicke et al., 2010), yields experimental evidence that there is action consciousness prior to conducted motion. There is subjective agency (SA) in terms of a conscious and therefore subjective act of voluntarily initiating bodily motion (Zhu, 2004, Kawamoto et al., 2010). “Subjective” concerns the individual conception of reality. An agent is subjective if her behavior is not completely predefined in terms of its task-orientation and (functionally defined) course of action (Grüneberg and Suzuki, 2014). In this view, action consciousness is a particular instance of subjective in terms of autonomous behavior. Accordingly,

“voluntary” here means that the human agent initiated the motion of her body based on her (spontaneous) decision and regardless whether there has been a previous external stimulus provoking a reflex or any internal constraint like Libet’s urge. It is also irrelevant whether motion actually occurs as SA’s efficacy concerns the release of a controlling neural signal (motor program) which may or may not result in bodily motion.

Conceding that robotic research serves as a source for investigating human cognition and behavior (Oudeyer, 2010, Morse et al., 2011), we use robotic experiments for identifying SA<sup>2</sup>. While action in general is a long-known candidate for integrating different modalities (Gallese, 2000), SA—compared to SoA—suggests a basically different type of action consciousness which in turn asks for a different explanation. In the same way as the multimodal approach suggests intersensory integration at the basic neuronal level of sensory processing, we suggest that *interfunctional* integration already occurs at the basic neuronal level of action initiation. Accordingly, our focus lies on the *functional organization* of action consciousness which allows SA as efficacious action consciousness. Following this approach, this paper aims at revealing a substantial constituent of action consciousness and at suggesting an explanation for SA as a case of multifunctional integration.

The remainder of this paper is divided into two parts. Section 2 and 3 identify SA as a distinct type of efficacious action consciousness. Section 4 and 5 investigate SA as a case of multifunctional integration and present experimental evidence as well as hypotheses based on SA. Because SA as an efficacious capacity is usually not regarded as a feature of action consciousness, we will first identify SA. For this purpose, we introduce robotic neurorehabilitation and in particular focus on the patient’s role in the therapeutic process (Section 2.1). By means of analyzing the implementation and effects of robotic neurorehabilitation, we argue that SA is efficacious in terms of voluntary initiation of motor programs (Section 2.2). Then we show that SA does not fall under common action consciousness (SoA). Due to the experiential stance of SoA (Section 3.1) and the corresponding functional organization (unifunctional integration), SoA does not capture SA as a conscious and at the same time efficacious capacity (Section 3.2). The identification of SA and its exclusion from common action consciousness lead to the conclusion that SA comes as a distinct type of the phenomenology of action and is classified as a phenomenal performance (Section 3.3). Based on this finding, we argue that the brain must be able to implement SA (Section 4.1) and present a functional organization (multifunctional integration) which allows for SA as efficacious action consciousness (Section 4.2). Finally, we illustrate SA by means of a case study of a robotic device for lower limb rehabilitation (Section 5.1). Hypotheses regarding neurorehabilitation and athletic sport are suggested which promise to gain insight into the link between SA and its implementation in motor behavior together with the detection of effects of neurorehabilitation (Section 5.2).

<sup>2</sup>Legrand also uses the concept of “subjective agency” for the pre-reflective condition of phenomenal experience (Legrand, 2007) and thereby remains within the common understanding of agentive as experiential consciousness. In turn, we relate SA to the agent’s capacity to initiate action voluntarily and consciously.

<sup>1</sup>If not otherwise specified, we will subsume the different senses related to action (sense of agency, ownership, intention, control etc.) to the sense of agency (SoA).

## 2. Subjective Agency in the Course of Robotic Neurorehabilitation

In recent years, exoskeleton robots have been developed for the rehabilitation of impairments of upper and lower limbs. Traditional physiotherapy follows a bottom-up approach in terms of acting on the (distal) physical level in order to influence the neural system. In comparison, robotic devices build on therapeutic top-down control for the purpose of neurorehabilitation (Belda-Lois et al., 2011). Hereby, neurorehabilitation depends on the state of the brain after a stroke or other damage and not on the physical level of the impaired limbs. In order to exploit neuroplasticity for rehabilitative purposes and motor learning, the patient's voluntary involvement in the therapeutic process is essential (Hogan et al., 2006) similar to cognitive-behavioral therapy where therapeutic effects also depend on the conscious modification of thought or behavior by the patient herself (Brewin, 1996; McKay et al., 2015). Therefore, robotic rehabilitation devices for upper (Maciejasz et al., 2014 for an overview) and lower limbs (Daz et al., 2011 for an overview) enable the patient to move her impaired limbs voluntarily despite the impairment.

### 2.1. From Being Moved to Voluntary Initiation: Device Control by Biosignals

The standard electromechanical approach to exoskeleton robots consists mainly of replacing motion support delivered by a physiotherapist. It is the task of an exoskeleton robot to move a limb according to a predetermined kinematic trajectory; thereby the patient makes use of the autonomous motion generated by the robot (Belda-Lois et al., 2011). Thus, the patient is being moved. Accordingly, purely mechanically based exoskeleton robots do not fully utilize a top-down approach by making use of remaining brain capacities and increasing the patient's involvement in motion generation because they do not consider the patient's intention to move her limbs voluntarily. Motion support remains passive as a purely mechanical and automated process closer to common physiotherapeutic support.

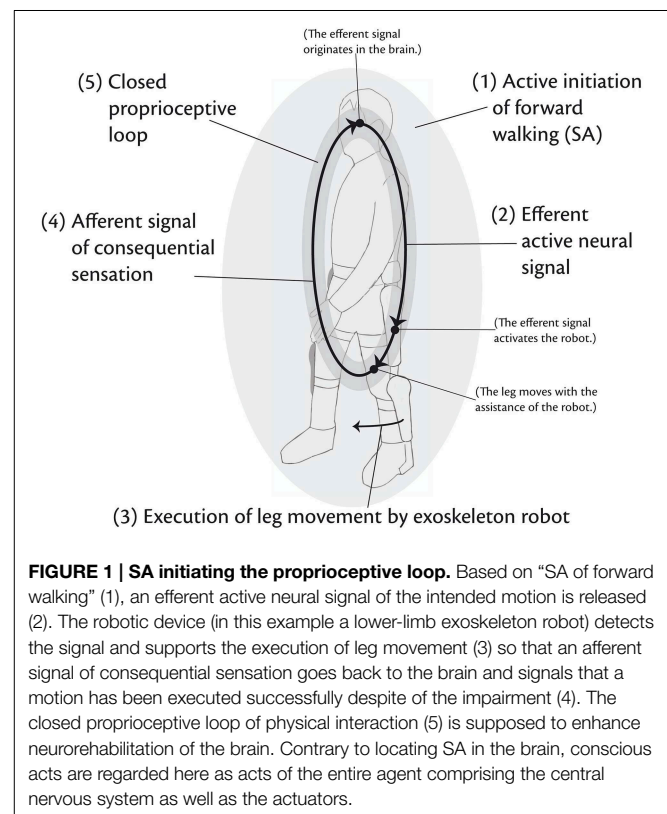
In order to increase the patient's participation, the control strategy of the robotic device has to be extended for implementing active support. Instead of letting the robot execute predetermined kinematic patterns, biosignals of patients can be exploited for the control of the robotic device (de Almeida Ribeiro et al., 2013). As especially EMG signals of neural muscle activity can be detected even in patients with severe impairments, these signals can be used to interpret the patient's intention to move, i.e., for human intention estimation (Suzuki et al., 2007). Thus, the patient is no longer being passively moved by the robotic device, but is enabled to control the robot directly by her capacity to voluntarily initiate bodily motion. Motion support is delivered according to the patient's needs<sup>3</sup>.

<sup>3</sup>Numerous examples for upper limb devices depending on biosignals or mechanical control can be found in Maciejasz et al. (2014); for an example of a lower limb device building on biosignal processing see Section 5.1.

### 2.2. Closing the Proprioceptive Loop

The obvious reason for arguing for SA lies in the therapeutic effects achieved by devices using biosignals (Kawamoto et al., 2013; Maciejasz et al., 2014). Lacking a decisive neuroscientific explanation for these effects, the following hypothesis might serve as a starting point to understand the implementation and effects of robotic neurorehabilitation: *Depending on SA, robotic devices allow for the closing of the proprioceptive loop (Kawamoto et al., 2013) of physical interaction between the efferent active neural signal and the afferent signal of consequential sensation of the intended motion and thereby enhance neurorehabilitation in that the brain detects successful initiation and execution of motion despite of the impairment.* According to this hypothesis, the therapeutic effects of a recovery of motivity and the underlying recovery of the corresponding brain regions are derived as follows (cf. **Figure 1**; for the sake of simplicity, we will illustrate the hypothesis by lower limb rehabilitation of forward gait which could be replaced by any other limb):

- 0 In case of locomotively impaired patients, there is no automatic (sub-personal) initiation of forward walking as the neural signals are not sufficient in order to activate the leg muscles. The patient remains in a resting position when no therapeutic actions are taken.
- 1 After being equipped with an exoskeleton robot, the patient voluntarily initiates forward walking (SA), i.e., consciously issues the command to move.





- 2 The neural activity of motor commands issued by the patient can be detected in the leg muscles: there is an efferent active neural signal.
- 3 The exoskeleton detects this signal by its EMG sensors and launches its motion support: actual walking motion is executed.
- 4 Due to the execution of a walking motion, an afferent signal of consequential sensation goes back to the brain and signals that a motion has been executed successfully.
- 5 The proprioceptive loop is closed. The brain regions responsible for motion control can chalk up a successful motion and remain active or become (partially) restored<sup>4</sup>.

The key assumption in this hypothetical sequence of the therapeutic process concerns the recovery of the brain regions responsible for motion control by means of closing the proprioceptive loop of efferent and afferent neural signals. Even if current research does not yield a final neuroscientific explanation for this effect, two findings are nevertheless obvious: Following (0), there is no automatic (sub-personal) initiation of motion in the therapeutic setting. This means that the sub-personal mechanisms of the locomotor system, which are usually held responsible for motion initiation (Haggard, 2005, Frith, 2013), no longer provide sufficient resources to initiate bodily motion automatically. The result is the obvious impairment of the patient and her corresponding inability to move her body. *Thus, firstly, in the case of patients with locomotive impairment there is no sub-personal (automatic) initiation of motion as the brain and/or the spinal cord are impaired to the extent that control and initiation of motion are no longer available automatically.*

Limited knowledge about neurorehabilitation does not challenge the clinical evidence (cf. Section 5.1) that there are significant therapeutic effects by robotic rehabilitation. So following (1) and (2), robotic therapy shows that there is a certain conscious and efficacious capacity of SA in order to voluntarily initiate motion. This finding can be directly concluded from the fact that the rehabilitation robot is only operated if the patient voluntarily (consciously) seeks to walk forward (Hogan et al., 2006, Eitam et al., 2013)<sup>5</sup>. If the patient does not voluntarily engage in the therapeutic process, nothing will happen (as stated above) and the patient's condition might even deteriorate. Accordingly, the *initiation* of the proprioceptive loop by means of the efferent active neural signal depends on the voluntary initiation by the patient. Regarding the motor-related objective of the initiation, behavioral research suggests that cognitive action control concerns the synergetic level of bodily motion (Latash et al., 2007). Voluntary motor programs identified by Ivanenko et al. are possible candidates to implement initiated motion physically (Ivanenko et al., 2004, Lacquaniti et al., 2012). They suggest five basic locomotion motor programs for gait which are possibly superimposed by voluntary motor programs depending

<sup>4</sup>For current purposes, the fact is crucial that the patient's ability to move is (partially) restored whereas it exceeds the scope of this paper to specify which brain regions in particular are affected. For further considerations of the neural implementation of SA cf. Section 4.2.

<sup>5</sup>The patient's capacity of SA can be generalized to healthy subjects as healthy subjects are also capable of a particularly active engagement in physical motion which exceeds automatic motion (Haggard and Johnson, 2003).

on the subject's control<sup>6</sup>. These motor programs are released even if no actual motion occurs. *Thus, secondly, SA is efficacious in terms of voluntary initiation of motor programs.*

### 3. Unifunctional Integration and the Sense of Agency

In general, conscious experience is supposed to form a particular and rather problematic case of unimodal integration as the coherence of objects and the phenomenal homogeneity of experience ask for a relatively high degree of integration. The same counts for motor behavior and is here referred to as *unifunctional integration*: Basic functions of motor behavior are integrated (bound) in order to make certain aspects of motor behavior contents of phenomenal experience. Action consciousness is usually spelled out in terms of SoA which forms a result of unifunctional integration. After a brief look at the common phenomenology of action and the underlying experiential stance toward agency (Section 3.1), we will turn to Pacherie (2008) as she links a strong phenomenology of action with its neural implementation. In particular, we examine how Pacherie draws on binding in order to explain SoA. It can be shown by means of the functional organization of SoA that unifunctional approaches to action consciousness such as SoA inevitably lead to *post-hoc* (experiential) action consciousness and therefore do not capture SA as an efficacious capacity (Section 3.2). Finally, the identification of SA (cf. Section 2) and its exclusion from common action consciousness (Section 3.2) lead to the conclusion that SA comes as a distinct type of the phenomenology of action which will be classified as a phenomenal performance (Section 3.3).

#### 3.1. Senses and Experiences: the Experiential Stance Toward Agency

The phenomenology of action is usually regarded as thin and evasive to the extent that its phenomenal content cannot be identified clearly (Metzinger, 2006). However, there is a degree of consensus about how to capture the phenomenology of action in terms of SoA. According to the basic definition provided by Gallagher, SoA is the sense that I am the one who is causing or generating an action, and comprises multiple aspects ranging from first-order experience linked to intentional aspects and bodily movements to second-order reflective attribution (Gallagher, 2007). Further conceptual refinement led to the distinction of a feeling of agency and a judgment of agency as different levels of SoA (Synofzik et al., 2008) and the framework of optimal cue integration building on prediction and postdiction (Synofzik et al., 2013)<sup>7</sup>. Based on the phenomenal

<sup>6</sup>Here we skip the question whether voluntary initiation refers to one motor program or possibly a complex of motor programs in order to initiate bodily motion.

<sup>7</sup>See David et al. (2008) for an overview of different accounts on SoA, distinguishing between the comparator model, simulation theory, intentional binding, and the multifactorial two-step account. Regarding gait, Kannape and Blanke argue that the central monitoring framework (comparator model) suffices for gait agency despite certain necessary refinements in case of gait as a whole body motion (Kannape and Blanke, 2012).

states associated with agency, experience of a movement can take different shapes, such as an action (of one's own), an action that one is of control of, an action that one is performing with a certain degree of effort, or an action that one is performing freely (Bayne, 2008).

All these approaches share the common ground that SoA is a *phenomenal echo* of sub-personal motor processes, i.e., "Awareness is a delayed and attenuated version of motor performance." (Haggard and Johnson, 2003, p. 81). Even if Haggard and Johnson stress the aspect that the phenomenology of action becomes more accessible during tasks which ask for active engagement of the agent, such as rehabilitation or motor learning and recreational activity, their far-reaching observation is not further elaborated. Common approaches to SoA and the phenomenology of action in general still take an *experiential stance toward agency* (Horgan et al., 2003, Bayne, 2008) which binds any phenomenally present agency to the dimension of perceptual experience. This experiential stance toward agency depends on the common triadic structure of (1) a subject of experience, i.e., the agent, (2) the experiential content, i.e., the phenomenal state of an action, which depends on (3) the object of experience, i.e., (aspects of) the physical movement. As will be shown by means of analyzing the functional organization of SoA, this experiential stance and the dependency of the state (2) on the object (3) inevitably renders agency a *post-hoc* phenomenon of sub-personal motor processes so that agentic experience in terms of SoA falls under an experiential caveat and is not supposed to play any efficacious role (Bayne and Levy, 2009).

### 3.2. Binding: Making Subjective Agency Impossible

In the following, we will analyze the *functional organization of SoA* within Pacherie's model (cf. **Table 1**). Beginning with a brief presentation of Pacherie's framework (Section 3.2.1), we will then focus on the organizational principle, the implementation of behavioral functions and the resulting type of consciousness (Section 3.2.2; cf. **Table 1**). This analysis will clarify why efficacious action consciousness is generally made impossible by unifunctional integration.

#### 3.2.1. Pacherie's Approach to the Sense of Agency

Despite the fact that SA forms a cornerstone of folk psychology (Malle, 2004) and the organization of societal life in general, Libet set off the latest avalanche which seeks to explain any action consciousness as an epiphenomenal consequence of locomotor processing (Libet et al., 1983). Following corresponding accounts of the brain, action consciousness is captured by experiential (*post-hoc*) SoA and epiphenomenally attached to sub-personal processes in the brain (Flohr, 1991, Metzinger, 2003).

In light of this development, Pacherie's approach is insofar of interest as she generally argues for *conscious* agency and considers processes of action initiation and control that not only represent goals or executed actions, but more actively organize and structure motor processes (Pacherie, 2014). This position possibly leads a way to SA. Yet on the other hand, she proposes a complex model which attributes SoA to a cybernetic model of action specification (Pacherie, 2008, also Kumar and Srinivasan, 2012 drawing on Clark, 2013). It is exactly this latter model which renders conscious agency, as Pacherie proposes, impossible. Regarding the aetiology of agentic experiences, Pacherie goes for a comparator-based approach according to which bodily action is initiated and controlled by inverse and forward models in a central monitoring framework (Frith et al., 2000) which basically follows a cybernetic setup of control mechanisms (Wolpert, 1997).

The comparator model serves to instantiate a dynamical model of intentions consisting of three hierarchical levels from distal D-intentions down to proximal P-intentions and motoric M-intentions (Pacherie, 2008). *D-intentions* consist of beliefs and desires and therefore concern the overall decision-making and rational control of bodily actions. *P-intentions* form a link between the rational level of D-intentions and motoric implementation. They integrate D-intentions and the situational constraints of a particular action (situational anchoring) and control the execution of an action. Finally, *M-intentions* concern the selection of motor-programs and serve for basic motor control. Based on this model, Pacherie breaks SoA down into the *sense of intentional causation*, the *sense of initiation* and the *sense of control*. These phenomenal agentic experiences of SoA are explained by attributing them to the neural processes underlying D-, P- and M-intentions.

#### 3.2.2. Hierarchical Binding and Phenomenal Counterparts

The standard account of unimodal integration (binding) builds on temporal synchrony. According to von der Malsburg and his correlation theory (von der Malsburg, 1994), neurons which process input of the same sensory object, are supposed to fire in temporal synchrony<sup>8</sup>. By means of roughly simultaneous fire rates, (populations of) neurons which relate to one and the same stimulus, synchronize even if the neurons are located in rather distant areas of the brain. Every external object can therefore evoke a certain representational pattern in the brain, a so-called assembly. These assemblies are supposed to bring about the homogeneous phenomenal consciousness of objects as single entities in that unimodal representations of an

<sup>8</sup>While temporal synchrony can be regarded as the standard account, also other accounts have been proposed; cf. Cleeremans (2003).

**TABLE 1 | Functional organization of unifunctional (SoA) and multifunctional (SA) integration.**

Functional organization	Organizational principle	Implementation of behavior	Type of consciousness
Unifunctional integration	Hierarchy: binding	Neuronal modules	SoA: experiential, non-efficacious
Multifunctional integration	Heterarchy: asymmetric reciprocity	Neuronal operators	SA: performative, efficacious

external object are bound together into one coherent neural state (Metzinger, 1995). In this view, conscious experience depends on the integration of basic sensory modalities and therefore emerges as an epiphenomenal higher-order process.

While the standard account of binding aims at binding features derived from sensory perception, Pacherie uses the underlying mechanism also for the integration of behavioral functions, referring to it as *efferent binding* (Pacherie, 2008). Based on the comparator model, a number of behavioral functions can be identified in her framework:<sup>9</sup>

- Comparator function
- Prediction of a movement
- Feedback of a movement
- Awareness of an intention to move
- Awareness of movement onset
- Motion initiation
- Motion supervision
- Motion execution

Pacherie explains the *sense of intentional causation* as the result of a comparison between the prediction and the feedback of a movement and the subsequent binding of movement and consequence. This type of efferent binding is also discussed as intentional binding (Haggard, 2005, Moore and Obhi, 2012). The *sense of initiation* results from binding the awareness of an intention to move and an awareness of movement onset. The *sense of control* depends on the comparison between desired, predicted, and actual states of a motion.

The behavioral functions underlying these senses which constitute SoA are mapped onto neural modules (modularization) so that there are specific brain areas, so-called neural correlates (of consciousness) (NCC) (Chalmers, 2000, Kühn et al., 2013), which instantiate behavioral functions. Thus, according to Pacherie's model SoA is the result of the integration of independent neural modules which implement the corresponding behavioral functions. One example is the supposed implementation of the comparator model by the posterior parietal cortex (PPC) which concerns the comparison of self-produced actions and their visual consequences, the cerebellum which concerns discrepancies between predicted and sensory consequences of actions and possibly the extrastriate body area (EBA) of the visual association cortex regarding visuo-motor incongruence (David et al., 2008). Accordingly, the sense of intentional causation is supposed to result from the bound (synchronized) activity of neural modules such as PPC, the cerebellum and EBA for comparison and modules for prediction and feedback processing which are possibly implemented by the supplementary motor area (SMA) (Eccles, 1982, Pfurtscheller et al., 2014). Resulting from the bound activities of neuronal modules at the basic level of processing, SoA emerges at higher levels of processing. The hierarchical binding of the behavioral functions constitutes SoA as "phenomenal counterpart[s]"

<sup>9</sup>For the sake of simplicity and as the general mechanism of binding remains the same, we do not further distinguish between non-sensory behavioral functions which control motion or which entail sensory input related to motion such as proprioceptive feedback. For the same reason, we also ignore the distinction between awareness and experience of action.

(Pacherie, 2008, p. 193) which are epiphenomenally attached to cybernetic control mechanisms<sup>10</sup>.

Considering the hierarchy of unifunctional integration (with locomotory modules at the bottom and SoA at the top), it is the *temporal organization* which renders SoA inefficacious. As the neural modules work independently at the basic neuronal level, SoA follows on their independent activities. SoA occurs only *after* the proprioceptive loop has been closed as the comparator model depends on the efferent neural signal as well as on the afferent signal of consequential sensation of the intended motion. Accordingly, the *sense of intentional causation* is not efficacious as it relies on the afferent feedback of an actual motion<sup>11</sup>. The same limitation holds for the *sense of control* which also relies on actual states of a motion and therefore depends on the closed proprioceptive loop. The remaining *sense of initiation* does not rely on any afferent signal and therefore conveys the impression to be a suitable candidate for efficacious action consciousness. Yet, a patient can try to initiate motion even if no movement onset occurs so that also the sense of initiation presupposes an already initiated motion.

The temporal dependency on the closed proprioceptive loop and therefore on the integration of independent neural modules renders SoA a mere phenomenal counterpart of sub-personal motor processes. As a purely experiential consciousness, a phenomenal counterpart cannot play any efficacious role because it merely follows on locomotory events instead of effecting the latter. Moreover, SoA immediately vanishes once the corresponding locomotory mechanisms are out of order as in the case of patients with locomotive impairments. These findings show that *SA as efficacious action consciousness does not fall under common experiential action consciousness such as SoA*.

### 3.3. Subjective Agency as a Phenomenal Performance

Regarding the results of robotic neurorehabilitation which gave rise to identify SA (Section 2) and the exclusion of SA from experiential action consciousness (Section 3.2), we suggest a preliminary working definition of SA as *phenomenal performance*. Accounts such as Chisholm (1966), O'Connor (2000) argue for something like SA on a conceptual level. But besides a certain conceptual plausibility, it is also important to fix the conscious phenomena of action initiation in an empirically verifiable manner<sup>12</sup>.

<sup>10</sup>The explanation of SoA in terms of phenomenal counterparts of cybernetic processes could also be extended to the general explanatory conflict that the explanandum (SoA) becomes superfluous in face of the explanans as cybernetic processes do not necessarily imply any phenomenal experience of agency, cf. Grüneberg (2013), chapter 5.

<sup>11</sup>Research on processes of action selection suggests that there is a prospective generation of SoA which does not rely on afferent signals (feedback) (Chambon et al., 2014). On the one hand, action selection should be regarded as preceding SA. However, the authors suggest that this prospective generation merges fully in the *post-hoc* experience of SoA, so that it is not obvious how the prospective generation should be efficacious.

<sup>12</sup>The proposed account of SA might at a first glance be similar to O'Connors' concept of agent causation (O'Connor, 2000) as in both cases the agent is supposed to be the cause of her action. The most important difference is that the phenomenality of SA refers to the real (embodied) agent and does not imply any metaphysical foundation in terms of O'Connors' agent.

SA is consciousness of an action *during* its initiation and therefore occurs previous to visible motor behavior. On the one hand, SA, just as SoA, bears a certain qualitative state of consciousness and *phenomenal* content (Nagel, 1974). The agent brings to mind that she is about to move (e.g., to move to another place by forward walking). In healthy agents, the volition just passes by as the intended motion is immediately implemented. If the motion requires efforts (e.g., walking uphill), the volition is phenomenally stronger and includes exertion. And in case no bodily motion occurs, the volition might even be stronger in terms of futile attempts of initiation. On the other hand, SA is a prospect of the intended motion. The phenomenal content of SA is present in the very moment of initiation and not given after its initiation. In the moment of initiation, one acts voluntarily (e.g., starts to move to another place by forward walking) so that the conscious content of SA is equal to the voluntary initiation of that action and therefore comes as a “performance.” Taking together the phenomenal (qualitative) presence of SA and its performative content, we suggest the working definition of *phenomenal performance* in order to describe SA as a distinct type of efficacious action consciousness. In contrast, SoA is bound to intentional objects of experience (here aspects of motor behavior) and therefore relates to already executed acts. It can be characterized as a *phenomenal representation* of motor behavior.

In sum, the rehabilitation scenario yields particular evidence for SA in that the patient can make efforts to move consciously comparably with the conscious modification of thought or behavior during cognitive-behavioral therapy. Even if the patient’s efforts to move do not result in any motion, SA still bears a phenomenally present performative act, and the corresponding neural signal occurs. Thus, even in the case of locomotory impairment, SA is still efficacious in releasing an efferent neural signal. But SA does not necessarily imply an awareness that one acts in terms of the action as an intentional object of experience as spelled out by SoA. Regarding the robotic rehabilitation scenario, SoA also plays an important role *after* motion has been initiated and implemented with the help of the robot. The patient receives different kinds of feedback, such as proprioceptive and visual feedback of her own motion. This information is also supposed to play an important role in the process of rehabilitation (Kawamoto et al., 2013). Thus, there are different types of experientially based consciousness of one’s action, as SoA shows. But this phenomenal representation has to be distinguished from SA as a phenomenal performance.

## 4. Multifunctional Integration and Subjective Agency

Hitherto, SA has been, firstly, identified as efficacious action consciousness (Section 2) which does, secondly, not fall under experiential action consciousness and comes as phenomenal performance (Section 3). As unfunctional integration or binding is not sufficient to explain SA’s efficacy, the question arises as to what is needed in order to explain SA as efficacious and therefore immediate (instead of epiphenomenally attached) action consciousness. In the following, we will present a

multifunctional approach to SA which could also be adapted for voluntary control of thought or combinations of thought and behavior as in cognitive-behavioral therapy. For this purpose, we will argue that the brain should be conceived in a way that allows the neural implementation of SA (Section 4.1). Then we suggest a functional organization of SA in terms of multifunctional integration (Section 4.2) and some general hypothesis on neurorehabilitation following SA (Section 4.3).

### 4.1. Not Underestimating the Brain

Whereas it should be the task of any scientific research about consciousness to explain what actually occurs in our conscious life, the current situation literally seems to have reversed. Instead of finding a conception of the brain which suffices for obvious phenomena such as SA, the latter are generally refuted by the prevailing conception of the brain as a representational device (cf. also Section 3.1). Hence, the situation arises that an obvious phenomenon such as SA is not allowed to be a conscious *and* efficacious phenomenon at the same time. This problem of recognizing SA stems from the underlying assumption of what the brain is capable of. If consciousness and cognition, as shown in Section 3.2, are supposed to result from neurocomputational brain processes, then the former can only achieve what the latter allow for. This bias excludes conscious processes from being efficacious regarding bodily action.

From a biological perspective Latash is making the same point when he explains that a biological system as the brain is explained in cybernetic terms which have originally been developed for much less complex systems such as the control of missiles (Latash, 2008, p. 323). In face of fundamental limitations of representational and information-theoretic explanations of consciousness (Eimer, 1990, Grüneberg, 2013), an analog point can be made here. Information-processing, which is mainly inspired by computational approaches and lies at the ground of neuroscientific approaches to cognition and behavior, does not capture complex intelligent behavior such as SA. Accordingly, from the viewpoint of SA, the fundamental questions arises why consciousness should *necessarily* and *exclusively* be experiential (*post-hoc*) and, subsequently, how to extend our understanding of the brain in order to include SA. As well as SA as multifunctional integration and in general the idea of multimodality, the concept of plasticity can be seen as another striking example that sticking to a certain conception of the brain avoids the recognition of its capabilities (Rubin, 2011). So it is important to continuously question the explanatory framework underlying the brain (Perruchet and Poulin-Charronnat, 2012).

### 4.2. Functional Organization of Subjective Agency

Analogous to the functional analysis of SoA in Section 3.2, the functional organization of SA will be clarified in terms of the organizational principle, the implementation of behavioral functions (Section 4.2.1) and the resulting type of consciousness (Section 4.2.2; cf. Table 1).

#### 4.2.1. Heterarchy: Asymmetric Reciprocity

SA comprises the behavioral functions of voluntary initiation and the respective motor programs. Both can be distinguished as both



can be performed independently of each other. While initiation can refer to other behavioral patterns such as cognitive behavior (Bayne and Montague, 2011), motor programs can also be initiated automatically without any contribution by the conscious agent. However, in case of SA both are integrated in a way that makes SA an *efficacious* action consciousness so that hierarchical binding with motor behavior at the basic level (as in case of SoA) is not any more feasible. Instead, we draw on the *heterarchic principle of asymmetric reciprocity* (Grüneberg, 2013, ch. 7, 8, Grüneberg and Suzuki, 2014) in order to explain the integration underlying SA. The general idea of asymmetric reciprocity is that action consciousness depends on a bidirectional relation of voluntary and automatic behavior with the former prevailing the latter. Such a bidirectional and asymmetric relation is what McCulloch (1945) and Günther (1971) call a heterarchy. A heterarchic relation allows for the *simultaneous* and therefore reciprocal activity of independent elements in a network so that behavioral functions are implemented reciprocally and at the same level of neuronal processing. At the same time, the heterarchic relation allows for one element governing other elements in that it includes a hierarchic and therefore *asymmetric* moment. In contrast to a strictly hierarchic setup where the governing element is predetermined by the hierarchy, the governing element in a heterarchy can change depending on the situation.

From this viewpoint, initiation as voluntary behavior and motor programs as automatic behavior asymmetrically depend on each other for the sake of SA. On the one hand, SA depends in two respects on motor programs. Firstly, if the agent wants to initiate a movement, the agent must be able to access her actuators. This job is done by motor programs (Ivanenko et al., 2004, Lacquaniti et al., 2012) which activate the locomotor system on a synergetic level (Latash et al., 2007). Voluntary behavior is enabled in that a voluntarily initiated motion is automatically executed after its initiation so that, for example, the agent can turn her attention to other tasks (Gallagher, 2006). Thus, automatic motion is not a contradiction to voluntary initiation, but the latter builds on automatic motor resources which comprise learned and habituated motor behavior and allow for new motor behavior. Secondly, if the agent selected a certain motor program, she is constrained to the respective motion and will move correspondingly. Even if she immediately modifies her motion by selecting a different motor program, every act of initiation is bound to its previous selection. Thus, any selection depends on the currently running motor program. Regarding the dependency of motor programs, a selection out of the pool of available motor programs is necessary in order to allow for coordinated (goal-directed) motion. Without a selection, no movement would occur. Thus, motor programs ask for a controlling instance. While this selection is often done by automatic selection, SA shows that this selection can also be done by the agent's voluntary initiation. According to this mutual dependency, initiation and motor programs are organized reciprocally<sup>13</sup>. At the same time, the selection of a

specific motor program, i.e., the efficacy of initiation, implies an asymmetric relation in that initiation releases one specific motor program. In case of SA, the prevalence is in favor of the voluntary initiation with the motor program being selected so that initiation and motor programs are organized by *asymmetric reciprocity*.

Regarding SA as action consciousness, we suggest that its conscious appearance depends on asymmetric reciprocity. Generally, the content of phenomenal consciousness comprises particular objects. The main feature of phenomenal consciousness is the persistence and homogeneity of those objects—may these be physical objects externally perceived or cognitive contents such as thoughts, intentions or inner images. All these objects are characterized by the fundamental feature that they form homogeneous entities which can be distinguished from other entities and therefore identified as single entities (Metzinger, 2003). Analogous to the problem of experiential consciousness how objects composed of different features and mediated by different sensory modalities can appear phenomenally as homogeneous and therefore distinguishable objects, action consciousness faces the problem how the performing agent can distinguish between different behaviors so that these can become identifiable contents of phenomenal consciousness. Regarding SA, the question is how the agent can distinguish between her voluntary initiation and the initiated automatic motor program so that both become identifiable phenomenal contents.

We suggest that this can be done by means of asymmetric reciprocity. (It has to be noted that we are here in the first place concerned with asymmetric reciprocity as the basic organizational principle for the implementation of subjectivity (Grüneberg and Suzuki, 2014). Phenomenal consciousness (whether experiential or performative) as a particular instance of subjectivity asks for further relational processing which is figured out in more detail in Grüneberg, 2013, ch. 8). Take again the case of SA of forward gait. According to reciprocity, voluntary initiation and the motor program for forward gait mutually depend on each other and are implemented simultaneously so that they are contents of the same phenomenal state. At the same time, the voluntary behavior (that the agent seeks to walk forward) and the selection of the corresponding movement depends on the agent's self-determination (it is up to the agent how to behave). In turn, the content of the automatic behavior itself is pre-determined because a certain motor-program implies one particular motion (here forward gait). According to this asymmetry, both behaviors can be distinguished from each other in that the voluntary behavior (initiation) becomes distinguished *as voluntary* from automatic behavior (forward gait) *as automatic*. Voluntary initiation and the automatic motor program for forward gait can therefore be identified as particular phenomenal contents of one and the same state, i.e., SA of forward gait. It is this *mutual distinction* between voluntary and

<sup>13</sup>Analogously, Chalmers et al. argue that so-called higher-order (conceptual) and lower-order (perceptual) processes necessarily depend on each other and thereby

show that alleged high-level or subjective cognition is already at play in so-called low-level cognition (Chalmers et al., 1992). The same counts here in that locomotor processes involve subjective selection processes.

automatic behavior that distinguishes both behaviors from each other and allows for SA becoming conscious.

SA is also efficacious as the phenomenal content of SA is no other than voluntary initiation of a motor program. The conscious act does not refer to any higher-order or epiphenomenal level as in the case of SoA where the content of agentic consciousness (the phenomenal state) is different from the underlying behavioral functions (the object of that state) and therefore cannot bear any efficacy. For example while the sense of control, the object of experience, comprises the comparison between desired, predicted and actual states, it appears phenomenally as the feeling that one is in control of an action (Pacherie, 2008). In case of SA, the phenomenal performance can be directly identified with voluntary initiation of a motor program so that SA can be efficacious and conscious at the same time.

In the therapeutic scenario of neurorehabilitation (or cognitive-behavioral therapy), SA clearly prevails motor behavior. However, the same behavioral functions could also be arranged differently. Another scenario might include the ongoing walking motion while the agent is having a conversation. In this latter scenario, the motor behavior is not being prevailed by SA but performs automatically without being consciously initiated compared to the rehabilitation scenario. The automatic execution allows an agent to focus on other tasks such as motion related aspects (e.g., navigation) or tasks completely distinct from motion (e.g., conversation or observation of the environment during walking). Therefore, if motor behavior is not initiated voluntarily but performs automatically or is not performed by the agent at all, this behavior is not conscious as there is no mutual distinction with any voluntary behavior. It depends on a particular situation which kind of functional behaviors are implemented reciprocally so that a phenomenal performance such as SA might arise.

#### 4.2.2. Multifunctional Integration: Operators Sharing Functions

According to modularization, behavioral functions are implemented by independent neural modules so that the integration of several functions follows *after* each independent function has been activated. Therefore, unifunctional integration depending on binding comes as a *secondary integration*. In contrast, SA asks for a *primary integration* of behavioral functions, i.e., the behavioral functions have to be immediately activated as integrated functions. Such a heterarchy cannot be facilitated by unimodal (secondary) integration. For this reason, we argue that SA requires *multifunctional integration*.

In the following, we refer to the concept of the *operator* in order to neurally implement SA as multifunctional behavior. This means that both voluntary initiation and the motor programs have to be implemented at the same basic neuronal level. After identifying what Bassin et al. (based on the works of Bernstein) called “neuronal polysensority” (Latash et al., 2000, p. 136<sup>14</sup>), they proposed the concept of an *operator* in

order to describe the modular (basic functional) units of the brain. Derived from control theory, an operator designates the particular design of a neuronal net which fulfills a specific operation in the neurodynamic processes of a brain region (Isomura et al., 2009). These operators can implement different behavioral functions and therefore come as the independent units of neural processing. For example, there are operators (neural circuits) that perform mathematical or action-related operations which can be shared by different functions (Latash, 2008) such as action planning, action initiation or learning. It is beyond the scope of this paper to identify particular neuronal operators. But, regarding SA, it can be suggested that there should be operators for the decision for, selection and release of a motor program which implement voluntary initiation. Neural circuits in the SMA and the insula might be possible candidates for implementing these operators (Eccles, 1982, Pfurtscheller et al., 2014). Other operators would comprise synergistic components which implement motor programs (Latash, 2008)<sup>15</sup>. In contrast to unimodal integration, multifunctional integration implies that behavioral functions are not directly (one-to-one) implemented by neural modules so that each single function has to be activated independently and then integrated. Instead, multifunctional operators implement behavioral functions *simultaneously as integrated functions* in that single functions are only realized reciprocally and in the context of a comprehensive multifunctional behavior such as SA. Due to their multimodal/-functional operability, operators allow for a primary and therefore multiple integration of behavioral functions.

Multifunctionality also implies that SA is a non-localizable function. There is no rigid modularization on the neural level according to which SA could be attributed to a NCC. Building on operators, there are not only several brain areas involved in SA but also the spinal cord<sup>16</sup> so that SA as a behavioral function is attributed to the entire agent as an embodied and conscious entity.

In sum, the functional organization of SA as a multifunctional setup resolves shortcomings of unifunctional integration of action consciousness. As SoA merely covers *post-hoc* experience and therefore neglects the efficacious nature of SA, the organization of the brain should be modified to that extent that phenomenal performance as an efficacious capacity can be implemented. For this purpose, we suggest the heterarchic relation of asymmetric reciprocity as the organizational principle and neural operators as the implementation of the functional organization of SA.

<sup>15</sup>Downward causation might serve as a comprehensive framework of the neural implementation of SA (Murphy et al., 2009).

<sup>16</sup>Control of movement roughly involves the spinal cord and brainstem circuits (lower motor neurons), the motor cortex and brainstem centers (upper motor neurons), the cerebellum and the basal ganglia (Purves et al., 2011). Depending on lower motor neurons and the generation of synergies in the spinal cord by central pattern generators (Grillner and Wallen, 1985), the neural control of movement encompasses the entire central nervous system.

<sup>14</sup>The cited paper is a translated reprint (Latash et al., 1999, Latash et al., 2000) of Bassin et al. (1966) which was originally published in Russian language.

### 4.3. Improving Neurorehabilitation by Utilizing Subjective Agency

Currently, discussions on neurorehabilitation center around whether an active or passive approach is more effective (Belda-Lois et al., 2011). This issue concerns the degree to which a patient's active participation is required in order to activate and control the therapeutic device (cf. Section 2). A related issue concerns neurorehabilitation as a form of motor learning (Huang and Krakauer, 2009, Kitago and Krakauer, 2013). Whereas motor learning approaches also consider the effect of active participation in terms of initiation of a movement by the patient, they mainly focus on the ongoing execution of a movement and the subsequent learning effects.

From the viewpoint of SA, an active approach which stresses the importance of voluntary *initiation* compared to the execution of a movement is advocated. This leads to the following hypothesis: (1.) *Effects of neurorehabilitation are significantly increased by voluntary initiation which* (2.) *enables motor learning.* Regarding the neuronal dynamics, SA initiates the proprioceptive loop so that the patient executes motor programs successfully (cf. Section 2). This effect builds on the multifunctional integration of SA according to which voluntary initiation directly activates motor programs. Accordingly, a patient can initiate movement comparable to a healthy condition (Section 4) so that an active approach to neurorehabilitation is supposed to be more effective than a passive approach because the active rehabilitation entails activation of the entire processes related to the intended movement whereas the passive rehabilitation incorporates solely local processes that are directly related to the treated joints. Furthermore, utilizing SA in supervised and unsupervised learning scenarios with robotic devices, a patient will receive proprioceptive feedback regardless whether the trained movement was successful or asks for further improvement. This allows a patient to enter into a learning process even if execution of movements is limited. Thus, SA also comprises enabling conditions for motor learning so that voluntary initiation should be emphasized compared to motor learning which performs often automatically once a motion has been initiated. Both parts of the hypothesis can be tested within the robotic framework presented in Section 5 as there is also behavioral evidence for the efficacy of neurorehabilitation initiated by SA (Section 5.2).

## 5. Experimental Evidence for Subjective Agency

### 5.1. Robotic Case Study: Exoskeleton Robot HAL

For the purpose of illustrating SA, we will present the exoskeleton robot HAL (hybrid assistive limb) (Sankai, 2006, Sankai, 2011) which is used for gait rehabilitation of spinal cord injury and stroke patients who suffer from severe impairments of motion (cf. Figures 2, 3). Currently HAL supports straightforward walking, standing up and sitting down. As different clinical studies show, HAL has successfully supported rehabilitation of 16 stroke patients (Kawamoto et al., 2013), 32 patients with stroke, SCI, musculoskeletal and other diseases (Kubota



FIGURE 2 | Patient wearing HAL in a walking device (front view).



FIGURE 3 | Patient wearing HAL in a walking device (side view).



et al., 2013), and one patient with ossification of the posterior longitudinal ligament (OPLL) (Sakakima et al., 2013). Compared to mechanically based exoskeleton robots which facilitate passive support<sup>17</sup>, HAL makes use of biosignals and facilitates active support.

Drawing on the proprioceptive loop (cf. **Figure 1** and Section 2.2), HAL's functionality can be described as follows: After the patient has been equipped with HAL, she voluntarily initiates a motor program for forward walking. HAL's crucial feature consists of EMG sensors attached to the flexor and extensor muscles of hip and knee. By means of this sensors, HAL detects the efferent active neural signal released by SA. In case there remains enough neural activity in the leg muscles, HAL interprets the neural impulse from the brain as a command to support walking motion and generates torque so that leg movement is facilitated. An afferent signal of consequential sensation is reported back to the brain and closes the proprioceptive loop and thereby supports neurorehabilitation. Thus, the patient initiates HAL's online gait support so that HAL is able to close the proprioceptive loop by estimating the patient's intention to move (Suzuki et al., 2007). Without HAL these patients are not able to initiate the physical gait motion *efficaciously*. The motor program is indeed issued, but not actually implemented. The fact that with HAL they are able to move implies that patients are able to initiate sub-personal motor-processes consciously by means of their SA.

In sum, the HAL scenario illustrates how SA is implemented as multifunctional behavior depending on asymmetric reciprocity. Voluntary initiation is directly bound to motor programs for forward gait in that the patient seeks to walk forward. Reversely, motor programs for forward gait are only initiated due to the agents conscious efforts to walk forward. Thus, in that both behavioral functions are activated simultaneously with voluntary initiation governing the selection of motor programs, SA is multifunctionally integrated and comes as efficacious action consciousness.

## 5.2. Testable Hypotheses Building on Subjective Agency

There are two possible areas where SA leads to testable hypotheses. One concerns neurorehabilitation by means of robotic devices. For the purpose of robotic neurorehabilitation, two different approaches are pursued as described in Section 2. On the one hand, patients use robots which build on the physiological signals of the patients motion. As these signals directly represent the intended motion, patients with locomotor impairments are enabled to initiate motion voluntarily (by themselves) while using a robot device. On the other hand, exercise is done by passive motion in that a therapist or a robot moves the patients limbs or body irrespective of motion initiation by the patient. In case that the human locomotor system would not allow for SA but only for SoA, therapeutic outcome of these two kinds of therapy would make no significant difference.

<sup>17</sup>Examples include Lokomat (Colombo et al., 2000), ALEX (Banala et al., 2009) or AutoAmbulator (Fisher et al., 2011).

There are some reports on the importance of participants efforts to initiate motion (Hogan et al., 2006; Eitam et al., 2013) during motor learning (Lotze, 2003) or hand rehabilitation (Takahashi et al., 2008) as well as the examples of the lower-limb exoskeleton robot that we discussed in the previous sections. Future analysis of the outcome of robotic rehabilitation could investigate the differences between the two approaches in a more evidence based manner. *A testable hypothesis concerns the extent of rehabilitative effects.* In case of SA, reflecting its characteristics as whole body phenomenal performance, whole body coordination including stability, efficiency in multiple muscle coordination, limb synergies and head/posture control during locomotion is improved while in the case of SoA only limb joint motion might be improved. This difference can be physically evaluated by means of motion measurement and analysis technology using 3D motion tracker and EMG sensors in addition to the conventional 10m walking speed test and by applying gait analysis methods which are commonly used in the field of behavioral science.

The other area concerns conscious initiation of motion and online control. Based on the functional organization of SA, experiments should focus on the link between voluntary initiation and motor programs as SA plays a major role in the selection of a single motion out of a pool of available motions. Of particular interest is the question how phenomenal performance controls motor programs, i.e., how an agent can shape her motor behavior by means of initiation and online control. In case of athletes, motion in competitive contexts entails a variety of extraordinarily rapid movements so that feed-forward control of motion is widely exploited whereas feedback control might be too slow to be included. Here, it should be considered to test conscious self-recognition of motion. In case of SoA, self-recognition reflects the conducted motion since SoA depends on the perception of represented motion. In case of SA, self-recognition might be rather different from the actually conducted motion. Considering that an athlete by means of SA might have learned an appropriate way of tricking sub-personal locomotor processes through training, she might in some situations be able to manipulate sub-personal processes much more effectively for better performance than by sending naive straight forward commands. *Thus, the subsequent hypothesis states that there are subjective motoric behaviors which allow for a goal-directed manipulation of motion.*

To test this hypothesis, motion measurement technology can be used again. First athletes are interviewed how they control motion and what is the key variable to control for example the height of a jump and the angle of rotation during turning in their specialized sports motion. Then we can compare their self-recognition of the motion to the physically measured motion. Differences between these two measurements can support the existence and efficacy of SA. Predictions include that SA concerns the global synergetic level of motion and rather not kinematic and kinetic details of motion. Moreover, the conscious access to or initiation of motion is supposed to contain highly subjective motoric behaviors which are not necessarily observed in objective kinetic and kinematic measurements.



## 6. Conclusion

Robotic rehabilitation yields evidence that there is action consciousness prior to conducted motion. A similar finding can also be derived from cognitive-behavioral therapy where the voluntary involvement of the patient does also form an essential part of the therapeutic process. Based on this evidence, we argued for SA in terms of voluntary initiation of motor programs for movement. By analyzing robotic neurorehabilitation and introducing the proprioceptive loop, it could be concluded that, firstly, SA as an efficacious conscious act does exist. Secondly, we distinguished SA from common action consciousness by means of an analysis of the functional organization of SoA that showed that SoA depends on unifunctional binding which inevitably leads to *post-hoc* and therefore inefficacious action consciousness. Because SoA is implemented by independent neural modules corresponding to the behavioral functions, consciousness emerges not until the functions are integrated (bound) and therefore beyond functional efficacy. Therefore, SA implies a different type of action consciousness and has been identified as a phenomenal performance: a conscious act which consists of voluntarily initiating motor behavior.

For the sake of implementing SA, we suggested multifunctional integration of the behavioral functions underlying SA. Drawing on the heterarchic principle of asymmetric reciprocity, voluntary initiation and motor programs can be integrated at the same neuronal level simultaneously with the prevalence of initiation. We argued that it is the *mutual distinction* between voluntary and automatic behavior that allows for SA becoming conscious. Regarding the neural

implementation of SA, we referred to the concept of the multifunctional operator which forms the basic neuronal module and is shared by different functions so that the activation of behavioral functions goes hand in hand with their integration. This means that the behavioral functions are not implemented independently as modules and then possibly integrated, but immediately integrated at the time of their activation. The multifunctional integration makes SA conscious with functional efficacy. Finally, we presented a robotic case study as experimental evidence for SA and sketched experimental setups of neurorehabilitation and athletic motion control in order to gain behavioral evidence for SA.

In sum, we propose that there is the phenomenal performance of SA as a type of efficacious action consciousness. Our analysis showed that a unifunctional approach to the brain is too narrow in order to capture the complexity of human behavior. Future research should seek to integrate multimodal input and multifunctional behavior. For this purpose, research in bodily motion forms an instructive starting point as movement implies a broad range of sensory and behavioral processing which are inherently integrated.

## Funding

Work by Patrick Grüneberg was funded by a JSPS Postdoctoral Fellowship (P14706).

## Acknowledgments

We thank Hassan Modar for valuable support and advice.

## References

- Alais, D., Newell, F. N., and Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. *Seeing Perceiving* 23, 3–38. doi: 10.1163/187847510X488603
- Banala, S. K., Kim, S. H., Agrawal, S. K., and Scholz, J. P. (2009). Robot assisted gait training with active leg exoskeleton (ALEX). *IEEE Trans. Neural Syst. Rehabil. Eng.* 17, 2–8. doi: 10.1109/TNSRE.2008.2008280
- Bassin, P. V., Bernstein, N. A., and Latash, L. P. (1966). “On the problem of the relation between structure and function in the brain from a contemporary point of view,” in *Physiology in Clinical Practice*, ed N. I. Grastschenkov (Moscow: Nauka), 38–71.
- Bayne, T. (2008). The phenomenology of agency. *Philos. Compass* 3, 182–202. doi: 10.1111/j.1747-9991.2007.00122.x
- Bayne, T., and Levy, N. (2009). “The feeling of doing: deconstructing the phenomenology of gency,” in *Disorders of Volition*, eds N. Sebanz and W. Prinz (Cambridge, MA: MIT Press), 49–68.
- Bayne, T., and Montague, M. (2011). *Cognitive Phenomenology, 1st Edn.* Oxford, NY: Oxford University Press.
- Belda-Lois, J.-M., Horno, S. M.-D., Bermejo-Bosch, I., Moreno, J. C., Pons, J. L., Farina, D., et al. (2011). Rehabilitation of gait after stroke: a review towards a top-down approach. *J. Neuroengineering Rehabil.* 8:66. doi: 10.1186/1743-0003-8-66
- Brewin, C. R. (1996). Theoretical foundations of cognitive-behavior therapy for anxiety and depression. *Annu. Rev. Psychol.* 47, 33–57. doi: 10.1146/annurev.psych.47.1.33
- Calvert, G. A., and Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *J. Physiol. Paris* 98, 191–205. doi: 10.1016/j.jphysparis.2004.03.018
- Chalmers, D. (2000). “What is a neural correlate of consciousness?” in *Neural Correlates of Consciousness. Empirical and Conceptual Questions*, ed T. Metzinger (Cambridge; London: MIT Press), 17–39.
- Chalmers, D. J., French, R. M., and Hofstadter, D. R. (1992). High-level perception, representation, and analogy: a critique of artificial intelligence methodology. *J. Exp. Theor. Artif. Intell.* 4, 185–211. doi: 10.1080/09528139208953747
- Chambon, V., Sidarus, N., and Haggard, P. (2014). From action intentions to action effects: how does the sense of agency come about? *Front. Hum. Neurosci.* 8:320. doi: 10.3389/fnhum.2014.00320
- Chisholm, R. M. (1966). “Freedom and action,” in *Freedom and Determinism*, ed K. Lehrer (New York, NY: Random House), 11–44.
- Clark, A. (2013). Are we predictive engines? Perils, prospects, and the puzzle of the porous perceiver. *Behav. Brain Sci.* 36, 233–253. doi: 10.1017/S0140525X12002440
- Cleeremans, A. (ed.). (2003). *The Unity of Consciousness. Binding, Integration, and Dissociation.* Oxford: Oxford University Press.
- Colombo, G., Joerg, M., Schreier, R., and Dietz, V. (2000). Treadmill training of paraplegic patients using a robotic orthosis. *J. Rehabil. Res. Dev.* 37, 693–700.
- David, N., Newen, A., and Vogeley, K. (2008). The sense of agency and its underlying cognitive and neural mechanisms. *Conscious. Cogn.* 17, 523–534. doi: 10.1016/j.concog.2008.03.004
- Daz, I., Gil, J. J., and Snchez, E. (2011). Lower-limb robotic rehabilitation: literature review and challenges. *J. Rob.* 2011:e759764. doi: 10.1155/2011/759764
- de Almeida Ribeiro, P. R., Brasil, F. L., Witkowski, M., Shiman, F., Cipriani, C., Vitiello, N., et al. (2013). Controlling assistive machines in paralysis using brain waves and other biosignals. *Adv. Hum. Comput. Int.* 2013, 3:3–3:3. doi: 10.1155/2013/369425

- Duschau-Wicke, A., Caprez, A., and Riener, R. (2010). Patient-cooperative control increases active participation of individuals with SCI during robot-aided gait training. *J. Neuroeng. Rehabil.* 7:43. doi: 10.1186/1743-0003-7-43
- Eccles, J. C. (1982). The initiation of voluntary movements by the supplementary motor area. *Arch. Psychiatr. Nervenkr.* 231, 423–441. doi: 10.1007/BF00342722
- Eimer, M. (1990). *Informationsverarbeitung und Mentale Representation. Die Analyse Menschlicher Kognitiver Fähigkeiten am Beispiel der Visuellen Wahrnehmung*. Berlin; Heidelberg: Springer.
- Eitam, B., Kennedy, P. M., and Higgins, E. T. (2013). Motivation from control. *Exp. Brain Res.* 229, 475–484. doi: 10.1007/s00221-012-3370-7
- Feil-Seifer, D., and Mataric, M. (2005). “Defining socially assistive robotics,” in *9th IEEE International Conference on Rehabilitation Robotics, ICORR2005* (Chicago, IL), 465–468.
- Fisher, S., Lucas, L., and Thrasher, T. A. (2011). Robot-assisted gait training for patients with hemiparesis due to stroke. *Top. Stroke Rehabil.* 18, 269–276. doi: 10.1310/tsr1803-269
- Flohr, H. (1991). Brain processes and phenomenal consciousness a new and specific hypothesis. *Theory Psychol.* 1, 245–262. doi: 10.1177/0959354391012006
- Frith, C. (2013). The psychology of volition. *Exp. Brain Res.* 229, 289–299. doi: 10.1007/s00221-013-3407-6
- Frith, C. D., Blakemore, and Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 355, 1771–1788. doi: 10.1098/rstb.2000.0734
- Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn. Sci.* 4, 14–21. doi: 10.1016/S1364-6613(99)01417-5
- Gallagher, S. (2006). “Where’s the action? Epiphenomenalism and the problem of free will,” in *Does Consciousness Cause Behavior?* eds S. Pockett, W. P. Banks, and S. Gallagher (Cambridge, MA: MIT Press), 109–124.
- Gallagher, S. (2007). The natural philosophy of agency. *Philos. Compass* 2, 347–357. doi: 10.1111/j.1747-9991.2007.00067.x
- Gallagher, S. (2012). Multiple aspects in the sense of agency. *New Ideas Psychol.* 30, 15–31. doi: 10.1016/j.newideapsych.2010.03.003
- Gallese, V. (2000). The inner sense of action. Agency and motor representations. *J. Conscious. Stud.* 7, 23–40.
- Grillner, S., and Wallen, P. (1985). Central pattern generators for locomotion, with special reference to vertebrates. *Annu. Rev. Neurosci.* 8, 233–261. doi: 10.1146/annurev.ne.08.030185.001313
- Grüneberg, P. (2013). *Projektives Bewusstsein. Th. Metzingers Selbstmodelltheorie und J.G. Fichtes Wissenschaftslehre*. Münster: Mentis.
- Grüneberg, P., and Suzuki, K. (2014). An approach to subjective computing: a robot that learns from interaction with humans. *IEEE Trans. Auton. Ment. Dev.* 6, 5–18. doi: 10.1109/TAMD.2013.2271739
- Günther, G. (1971). “Cognition and volition. A contribution to a cybernetic theory of subjectivity,” in *Cybernetics Technique in Brain Research and the Educational Process*, ed V. Alexandria (Washington, DC: American Society for Cybernetics), 119–135.
- Haggard, P. (2005). Conscious intention and motor cognition. *Trends Cogn. Sci.* 9, 290–295. doi: 10.1016/j.tics.2005.04.012
- Haggard, P., and Johnson, H. (2003). Experiences of voluntary action. *J. Conscious. Stud.* 10, 72–84.
- Hogan, N., Krebs, H. I., Rohrer, B., Palazzolo, J. J., Dipietro, L., Fasoli, S. E., et al. (2006). Motions or muscles? Some behavioral factors underlying robotic assistance of motor recovery. *J. Rehabil. Res. Dev.* 43, 605. doi: 10.1682/JRRD.2005.06.0103
- Horgan, T., Tienson, J., and Graham, G. (2003). “The phenomenology of first-person agency,” in *Physicalism and Mental Causation: The Metaphysics of Mind and Action*, eds S. Walter and H.-D. Heckmann (Exeter: Imprint Academic), 323–340.
- Huang, V. S., and Krakauer, J. W. (2009). Robotic neurorehabilitation: a computational motor learning perspective. *J. Neuroeng. Rehabil.* 6:13. doi: 10.1186/1743-0003-6-5
- Isomura, Y., Harukuni, R., Takekawa, T., Aizawa, H., and Fukai, T. (2009). Microcircuitry coordination of cortical motor information in self-initiation of voluntary movements. *Nat. Neurosci.* 12, 1586–1593. doi: 10.1038/nn.2431
- Ivanenko, Y. P., Poppele, R. E., and Lacquaniti, F. (2004). Five basic muscle activation patterns account for muscle activity during human locomotion. *J. Physiol.* 556, 267–282. doi: 10.1113/jphysiol.2003.057174
- Kannape, O., and Blanke, O. (2012). Agency, gait and self-consciousness. *Int. J. Psychophysiol.* 83, 191–199. doi: 10.1016/j.ijpsycho.2011.12.006
- Kawamoto, H., Kamibayashi, K., Nakata, Y., Yamawaki, K., Ariyasu, R., Sankai, Y., et al. (2013). Pilot study of locomotion improvement using hybrid assistive limb in chronic stroke patients. *BMC Neurol.* 13:141. doi: 10.1186/1471-2377-13-141
- Kawamoto, H., Taal, S., Niniss, H., Hayashi, T., Kamibayashi, K., Eguchi, K., et al. (2010). Voluntary motion support control of Robot Suit HAL triggered by bioelectrical signal for hemiplegia. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2010, 462–466. doi: 10.1109/IEMBS.2010.5626191
- Kitago, T., and Krakauer, J. W. (2013). Motor learning principles for neurorehabilitation. *Handb. Clin. Neurol.* 110, 93–103. doi: 10.1016/b978-0-444-52901-5.00008-3
- Kubota, S., Nakata, Y., Eguchi, K., Kawamoto, H., Kamibayashi, K., Sakane, M., et al. (2013). Feasibility of rehabilitation training with a newly developed wearable robot for patients with limited mobility. *Arch. Phys. Med. Rehabil.* 94, 1080–1087. doi: 10.1016/j.apmr.2012.12.020
- Kühn, S., Brass, M., and Haggard, P. (2013). Feeling in control: neural correlates of experience of agency. *Cortex* 49, 1935–1942. doi: 10.1016/j.cortex.2012.09.002
- Kumar, D., and Srinivasan, N. (2012). Hierarchical event-control and subjective experience of agency. *Theor. Philos. Psychol.* 3:410. doi: 10.3389/fpsyg.2012.00410
- Lacquaniti, F., Ivanenko, Y. P., and Zago, M. (2012). Patterned control of human locomotion. *J. Physiol.* 590, 2189–2199. doi: 10.1113/jphysiol.2011.215137
- Latash, L. P., Latash, M. L., and Meijer, O. G. (1999). 30 Years later: the relation between structure and function in the brain from a contemporary point of view (1966), part I. *Motor Control* 3, 329–345.
- Latash, L. P., Latash, M. L., and Meijer, O. G. (2000). 30 Years later: the relation between structure and function in the brain from a contemporary point of view (1966), part II. *Motor Control* 4, 125–149.
- Latash, M. L. (2008). *Synergy*. Oxford: Oxford University Press.
- Latash, M. L., Scholz, J. P., and Schoner, G. (2007). Toward a new theory of motor synergies. *Motor Control* 11:276.
- Legrand, D. (2007). Naturalizing the acting self: subjective vs. anonymous agency. *Philos. Psychol.* 20, 457–478. doi: 10.1080/09515080701441736
- Libet, B., Gleason, C. A., Wright, E. W., and Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential) the unconscious initiation of a freely voluntary act. *Brain* 106, 623–642. doi: 10.1093/brain/106.3.623
- Lotze, M. (2003). Motor learning elicited by voluntary drive. *Brain* 126, 866–872. doi: 10.1093/brain/awg079
- Maciejasz, P., Eschweiler, J., Gerlach-Hahn, K., Jansen-Troy, A., and Leonhardt, S. (2014). A survey on robotic devices for upper limb rehabilitation. *J. Neuroengineering Rehabil.* 11:3. doi: 10.1186/1743-0003-11-3
- Malle, B. F. (2004). *How the Mind Explains Behavior*. Cambridge, MA: MIT Press.
- McCulloch, W. S. (1945). A heterarchy of values determined by the topology of nervous nets. *Bull. Math. Biophys.* 7, 89–93. doi: 10.1007/BF02478457
- McKay, D., Sookman, D., Neziroglu, F., Wilhelm, S., Stein, D. J., Kyrios, M., et al. (2015). Efficacy of cognitive-behavioral therapy for obsessive-compulsive disorder. *Psychiatry Res.* 225, 236–246. doi: 10.1016/j.psychres.2014.11.058
- Metzinger, T. (1995). “Faster than thought: holism, homogeneity, and temporal coding,” in *Conscious Experience*, ed T. Metzinger (Paderborn: Ferdinand Schöningh), 425–462.
- Metzinger, T. (2003). *Being no One. The Self-model Theory of Subjectivity*. Cambridge, MA: MIT Press.
- Metzinger, T. (2006). “Conscious volition and mental representation: toward a more fine-grained analysis,” in *Disorders of Volition*, eds N. Sebanz and W. Prinz (Cambridge, MA: MIT Press).
- Moore, J. W., and Obhi, S. S. (2012). Intentional binding and the sense of agency: a review. *Conscious. Cogn.* 21, 546–561. doi: 10.1016/j.concog.2011.12.002
- Morse, A. F., Herrera, C., Clowes, R., Montebelli, A., and Ziemke, T. (2011). The role of robotic modelling in cognitive science. *New Ideas Psychol.* 29, 312–324. doi: 10.1016/j.newideapsych.2011.02.001
- Murphy, N., Ellis, G. F. R., and O’Connor, T. (eds.). (2009). *Downward Causation and the Neurobiology of Free Will*. Berlin; Heidelberg: Springer.

- Musseler, J., Drewing, K., and Sutter, C. (eds.). (2014). *Multisensory Integration in Action Control, Frontiers Research Topics*. Lausanne: Frontiers SA Media.
- Nagel, T. (1974). What is it like to be a bat? *Philos. Rev.* 83, 435–450. doi: 10.2307/2183914
- O'Connor, T. (2000). *Persons and Causes: The Metaphysics of Free Will*. Oxford: Oxford University Press.
- Oudeyer, P.-Y. (2010). On the impact of robotics in behavioral and cognitive sciences: from insect navigation to human cognitive development. *IEEE Trans. Auton. Ment. Dev.* 2, 2–16. doi: 10.1109/TAMD.2009.2039057
- Pacherie, E. (2008). The phenomenology of action: a conceptual framework. *Cognition* 107, 179–217. doi: 10.1016/j.cognition.2007.09.003
- Pacherie, E. (2014). Can conscious agency be saved? *Topoi* 33, 33–45. doi: 10.1007/s11245-013-9187-6
- Perruchet, P., and Poulin-Charronnat, B. (2012). Is consciousness a passive recipient of the end-product of sophisticated unconscious computations? *Theor. Philos. Psychol.* 3:413. doi: 10.3389/fpsyg.2012.00413
- Pfurtscheller, G., Andrade, A., Koschutnig, K., Brunner, C., and da Silva, F. L. (2014). Initiation of voluntary movements at free will and ongoing 0.1-Hz BOLD oscillations in the insulae pilot study. *Front. Integr. Neurosci.* 8:93. doi: 10.3389/fnint.2014.00093
- Pockett, S., Banks, W. P., and Gallagher, S., (eds.). (2006). *Does Consciousness Cause Behavior?* Cambridge, MA: MIT Press.
- Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., and White, L. E. (2011). *Neuroscience, 5th Edn.* Sunderland, MA: Sinauer Associates, Inc.
- Rubin, B. (2011). “Das menschliche Gehirn im Wandel,” in *Modell Mensch. Konturierungen des Menschlichen in den Wissenschaften*, eds R. Egloff, P. Gisler, and B. Rubin (Zürich: Chronos), 123–140.
- Sakakima, H., Ijiri, K., Matsuda, F., Tominaga, H., Biwa, T., Yone, K., et al. (2013). A newly developed robot suit hybrid assistive limb facilitated walking rehabilitation after spinal surgery for thoracic ossification of the posterior longitudinal ligament: a case report. *Case Rep. Orthop.* 2013:e621405. doi: 10.1155/2013/621405
- Sankai, Y. (2006). “Leading edge of cybernics: robot suit HAL,” in *SICE-ICASE, 2006. International Joint Conference (Busan)*, 1–2.
- Sankai, Y. (2011). “HAL: hybrid assistive limb based on cybernics,” in *Robotics Research The 13th International Symposium ISRR, Springer Tracts in Advanced Robotics*, eds M. Kaneko and Y. Nakamura (Berlin: Springer), 25–34.
- Suzuki, K., Mito, G., Kawamoto, H., Hasegawa, Y., and Sankai, Y. (2007). “Intention-based walking support for paraplegia patients with robot suit HAL,” in *Climbing and Walking Robots*, Vol. 21, ed B. Miripour (Rijeka; Shanghai: InTech), 1441–1469.
- Synofzik, M., Vosgerau, G., and Newen, A. (2008). Beyond the comparator model: a multifactorial two-step account of agency. *Conscious. Cogn.* 17, 219–239. doi: 10.1016/j.concog.2007.03.010
- Synofzik, M., Vosgerau, G., and Voss, M. (2013). The experience of agency: an interplay between prediction and postdiction. *Conscious. Res.* 4:127. doi: 10.3389/fpsyg.2013.00127
- Takahashi, C. D., Der-Yeghiaian, L., Le, V., Motiwala, R. R., and Cramer, S. C. (2008). Robot-based hand motor therapy after stroke. *Brain* 131, 425–437. doi: 10.1093/brain/awm311
- Tejima, N. (2001). Rehabilitation robotics: a review. *Adv. Robo.* 14, 551–564. doi: 10.1163/156855301742003
- von der Malsburg, C. (1994). “The correlation theory of brain function,” in *Models of Neural Networks II*, eds E. Domany and J. L. K. van Hemmen (Berlin: Springer), 94–119.
- Wolpert, D. M. (1997). Computational approaches to motor control. *Trends Cogn. Sci.* 1, 209–216. doi: 10.1016/S1364-6613(97)01070-X
- Zhu, J. (2004). Understanding volition. *Philos. Psychol.* 17, 247–273. doi: 10.1080/0951508042000239066

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Grüneberg, Kadone and Suzuki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# How our body influences our perception of the world

Laurence R. Harris<sup>1,2\*</sup>, Michael J. Carnevale<sup>1,2</sup>, Sarah D'Amour<sup>1,2</sup>, Lindsey E. Fraser<sup>1,2</sup>, Vanessa Harrar<sup>3</sup>, Adria E. N. Hoover<sup>1,2</sup>, Charles Mander<sup>1,2</sup> and Lisa M. Pritchett<sup>1,2</sup>

<sup>1</sup> Multisensory Integration Laboratory, The Centre for Vision Research, York University, Toronto, ON, Canada, <sup>2</sup> Department of Psychology, York University, Toronto, ON, Canada, <sup>3</sup> School of Optometry, University of Montreal, Montreal, QC, Canada

## OPEN ACCESS

### Edited by:

Achille Pasqualotto,  
Sabanci University, Turkey

### Reviewed by:

Michiel M. Spapé,  
Helsinki Institute for Information  
Technology HIIT/Aalto University,  
Finland

Manuela Ruzzoli,  
Pompeu Fabra University, Spain

### \*Correspondence:

Laurence R. Harris,  
Department of Psychology, York  
University, 4700 Keele Street,  
Toronto, ON M3J 1P3, Canada  
harris@yorku.ca;  
Vanessa Harrar,  
School of Optometry,  
University of Montreal, 3744  
Jean-Brillant, Montreal,  
QC H3T 1P1, Canada  
vanessa.harrar@umontreal.ca

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 26 March 2015

**Accepted:** 29 May 2015

**Published:** 12 June 2015

### Citation:

Harris LR, Carnevale MJ, D'Amour S,  
Fraser LE, Harrar V, Hoover AEN,  
Mander C and Pritchett LM (2015)  
How our body influences our  
perception of the world.  
Front. Psychol. 6:819.  
doi: 10.3389/fpsyg.2015.00819

Incorporating the fact that the senses are embodied is necessary for an organism to interpret sensory information. Before a unified perception of the world can be formed, sensory signals must be processed with reference to body representation. The various attributes of the body such as shape, proportion, posture, and movement can be both derived from the various sensory systems and can affect perception of the world (including the body itself). In this review we examine the relationships between sensory and motor information, body representations, and perceptions of the world and the body. We provide several examples of how the body affects perception (including but not limited to body perception). First we show that body orientation effects visual distance perception and object orientation. Also, visual-auditory crossmodal-correspondences depend on the orientation of the body: audio “high” frequencies correspond to a visual “up” defined by both gravity and body coordinates. Next, we show that perceived locations of touch is affected by the orientation of the head and eyes on the body, suggesting a visual component to coding body locations. Additionally, the reference-frame used for coding touch locations seems to depend on whether gaze is static or moved relative to the body during the tactile task. The perceived attributes of the body such as body size, affect tactile perception even at the level of detection thresholds and two-point discrimination. Next, long-range tactile masking provides clues to the posture of the body in a canonical body schema. Finally, ownership of seen body parts depends on the orientation and perspective of the body part in view. Together, all of these findings demonstrate how sensory and motor information, body representations, and perceptions (of the body and the world) are interdependent.

**Keywords:** body representation, distance, gravity, auditory, crossmodal, tactile, self-perception

## Introduction

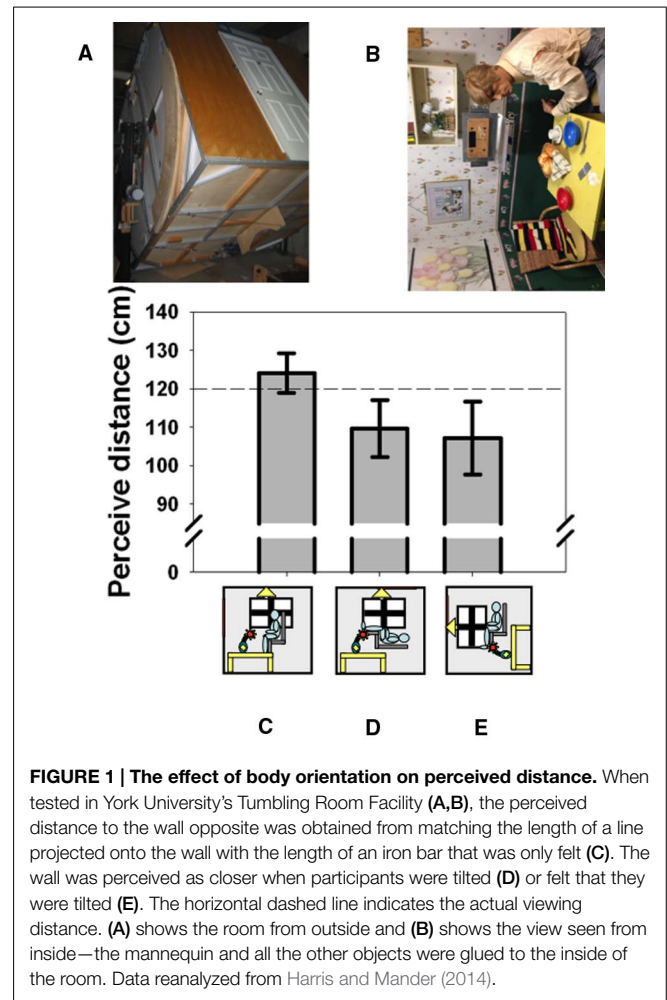
Since the pioneering philosophical approach of Merleau-Ponty (1945), it has been acknowledged that the senses are embodied. The implication of this approach is that the senses can only be understood by acknowledging the attributes of the body in which they are necessarily situated. In vision, it is obvious that the eyes are in the head and that their viewpoints will be affected by the head's position and orientation. What is perhaps less obvious is that these properties of the eyes' vehicle contribute to processing such “visual” judgments as the orientation of the ground plane (Schreiber et al., 2008) and, as we will see, perceived distance. Head position influences the three-dimensional position of the eyes by means of static and dynamic three-dimensional vestibulo-ocular reflexes and through eye height. Information concerning head position is therefore



critical to “externalize” the information in the retinal images: that is in creating a representation of the external world. Similar arguments apply to the ears, which are also passengers on the head. Head motion can even help to correctly scale the representation of external space, e.g., the distance between objects, which is notoriously hard to extract from static auditory or visual information alone (Gogel, 1963; Philbeck and Loomis, 1997). Information about the body is also needed to interpret tactile information about the world. When the hands explore and interact with objects in the world, it is necessary to take into account the arrangements of the hands and fingers in order to interpret the patterns of pressure sensed by the fingertips. The representation of the body is also needed to interpret the pressure and location of even simple touches on the skin in order to take into account the uneven density of tactile receptors over the surface of the body in the same way as the visual system must take into account the uneven density of photoreceptors in the retina. In this review we will outline some of the interesting, unexpected and fundamental roles that the body plays in determining our perception of the world.

### The Effect of Body Orientation on Perceived Distance

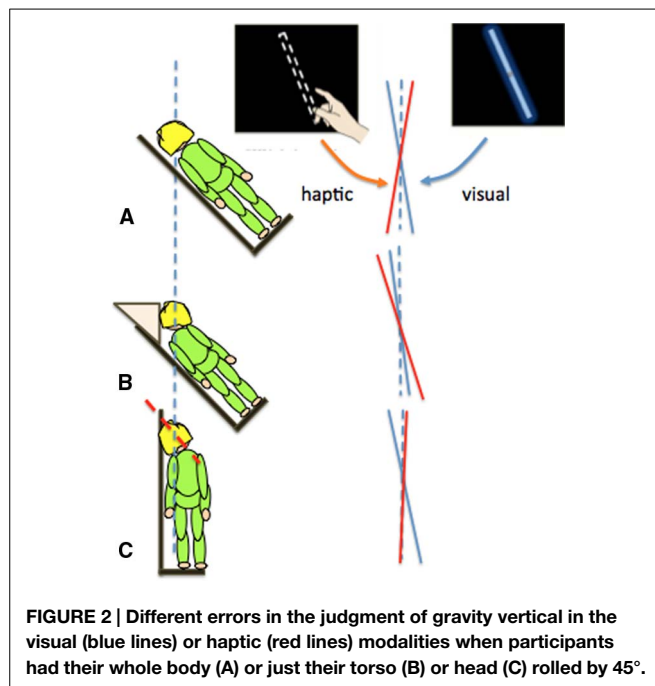
Things look different when viewed with the head in an unusual orientation. It is amusing, for example, to look out of a tall building and watch people walking on the street below. Their legs seem to move in a strange way and they often look too small, “like ants,” suggesting a failure in size-distance constancy when looking straight downward which also extends to the perception of speed (Owens et al., 1990). It has long been suspected that body orientation or perceived body orientation may be connected to perhaps the most famous distance-related illusion in psychology: the moon illusion (Rock and Kaufman, 1962). Casual observation shows that the moon appears smaller when it is in the zenith and viewed by looking straight up than when it is close to the horizon and viewed straight ahead. Although the illusion continues to defy complete explanation (Hershenson, 1989; Ross and Plug, 2002; Weidner et al., 2014), it is usually explained with reference to changes in the moon’s perceived distance. We (Harris and Mander, 2014) were the first to measure the effect of posture (and perceived posture) on the perceived distance of objects at biologically significant distances (Cutting and Vishton, 1995), as opposed to the unknowable distance of celestial bodies. We used the York University Tumbling Room facility (Howard and Hu, 2001) in which the orientation of an observer and the surrounding room can be independently varied (Figures 1A,B). We showed that lying supine causes the opposite wall of the room to appear closer than when viewed from an upright position (Figures 1C,D). Rotating the room around an upright observer (Figure 1E) produces an illusion of lying supine (Howard and Hu, 2001). Just *feeling* supine due to this illusion turned out to be sufficient to create this shortening of perceived distance (Harris and Mander, 2014). Thus, it is the perceived orientation of the body that is important in interpreting visual cues to distance. This may be related to the geometrical requirement of taking eye orientation—itsself dependent on head orientation—into account in order to interpret binocular cues correctly (Blohm et al., 2008).



This unexpected involvement of the body in visual distance perception underscores the importance of the body in interpreting sensory information—it is not a raw sensory signal that leads to perception, but rather the representation in the brain of world features (including the body itself) that is modified in response to sensory input and that determines perception.

### The Effect of Body Orientation on the Perceived Orientation of Objects

In the previous section we showed that the body’s orientation in pitch (head over heels) affected distance and size perception. Other changes in self-orientation can also lead to errors in perceptual judgments. When the body is rolled to one side (Figure 2), individuals systematically misperceive the orientation of an object relative to gravity. For example, when judging the orientation of a visual line with gravity vertical, estimates are biased toward the body midline (Aubert, 1886; Mittelstaedt, 1983). In contrast, when setting a bar to gravity vertical using only touch, there is a bias in the opposite direction, away from the body midline (Bauermeister et al., 1964). We (Fraser et al., 2014; Harris et al., 2014) compared visual and manual, touch-based estimates of gravity vertical while the body and head were tilted relative to each other (Figure 2).



Touch-based orientation judgments were affected more by the orientation of the body (**Figure 2B**), whereas visual errors were largely driven by head tilt (**Figure 2C**; Guerraz et al., 1998; Tarnutzer et al., 2009). Together, these results show that it is oversimplistic to refer to the representation of the body as a single unit when considering the effect of self-orientation on perception. Changes in orientation of the head and body can have different effects on different sensory inputs and so they should be taken into account separately: posture is an important factor.

### The Effect of Body Orientation on Auditory Localization

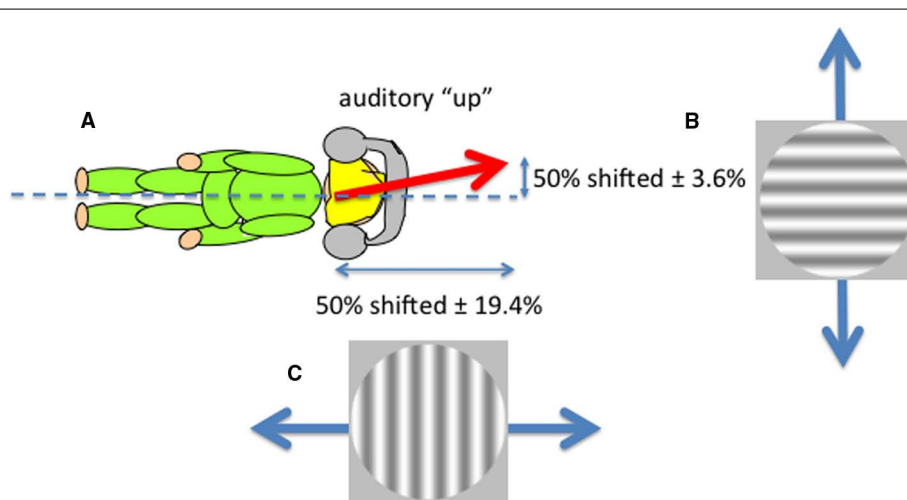
The ability to localize sound in elevation is tricky. What ability we have depends largely on reflections within the external pinna (Batteau, 1967; Fisher and Freedman, 1968; Makous and Middlebrooks, 1990; Blauert, 1996) and is thus bound to the head. Deducing where sounds are in the external world therefore requires taking into account the position and orientation of the head. Errors in sound localization when the head and body are tilted show that head orientation is only partially taken into account (Goossens and van Opstal, 1999; Parise et al., 2014). In fact, the perceived elevation of a sound, like the perceived orientation of a line we described above, depends on the perceived orientation of the head, which is determined by several factors.

Sounds that are played through headphones with no intrinsic location at all can nevertheless be perceived as having an elevation by virtue of their frequency content. This is an example of a cross-modal correspondence (Spence, 2011), in this case between pitch and perceived elevation, in which “higher” frequencies are perceived as coming from “higher” in space. But is this elevation defined in head or space coordinates? We showed that such sounds were perceived as lying on an axis defined neither by the head nor gravity but rather that lined up with the perceptual

upright (Carnevale and Harris, 2013). Non-spatial sounds (tones played through headphones) that differed only in their frequency content (either rising or falling frequencies) were presented while observers viewed ambiguous visual motion in either the horizontal or vertical directions created by superimposing two gratings moving in opposite directions (left and right or up and down; **Figure 3**). Observers were tested lying on their sides to separate body and gravitational uprights. A disambiguating effect of sound was found in both directions (up relative to the head and up relative to gravity), suggesting that an auditory upright exists in between the head and gravitational reference frames—a direction very similar to the perceptual upright. The perceptual upright is the orientation in which objects are best identified and represents the brain’s best guess of the direction of up derived from a combination of visual and gravity cues (Dyde et al., 2006) and a tendency to revert to the body midline (Mittelstaedt, 1983). As we showed above for the influence of the body in determining perceived distance and orientation of objects, the perceived orientation of the body also determines the layout of auditory space (see also Parise et al., 2014, who used external sounds). So both visual and auditory perceptions depend on body orientation. What about the perception of touch?

### Tactile Responses Depend on the Direction of Gaze

The orientation of the eyes and head are also involved in determining the perceived location of a touch such that the perceived location of a touch is shifted depending on gaze position (Harrar and Harris, 2009; Pritchett and Harris, 2011). Of course the direction of gaze is usually also the point to which attention is directed and attention is known to affect some aspects of tactile perception (Michie et al., 1987) in a way that depends on eye position (Gherri and Forster, 2014). However, Harrar and Harris (2009) found, by overtly orienting attention away from eye position, that attention could account for only about 17% of the effect. Even actions toward a touch are directed toward the shifted perceived position (Harrar and Harris, 2010). The effect appears to be equally affected by either eye or head displacement and is therefore best described as relating to gaze, the sum of eye and head position (Pritchett and Harris, 2011). The perceived location of touch also depends on whether a participant moves their gaze between the presentation of the touch and reporting its perceived location (Pritchett et al., 2012; Mueller and Fiehler, 2014). The perceived location shifts in the same direction as gaze if a gaze change occurs before the report (Harrar and Harris, 2009; Pritchett and Harris, 2011; Harrar et al., 2013), but in the opposite direction if the person does not move before making their report (Ho and Spence, 2007; Pritchett et al., 2012). What do these strange reversals tell us about the involvement of the body in the coding of touch? We can partially explain these gaze-related shifts in terms of the frame of reference in which touch location is coded. The direction of gaze and the direction in which the body is facing are misperceived toward one another when gaze is held eccentrically: the perceived straight ahead of the body is shifted in the direction of gaze, and the perceived direction of gaze is underestimated and perceived as closer to the body’s “straight ahead” (Hill, 1972; Morgan, 1978; Yamaguchi and Kaneko, 2007;



**FIGURE 3 | The perceived direction of auditory “up” (A)** corresponds to the perceptual upright which is determined by a combination of the body idiotropic vector and gravity with more emphasis on the body (Dyde et al., 2006). Non-spatial sounds differing only in their frequency content were used to disambiguate ambiguous visual motion created by superimposing

two gratings moving in opposite directions either vertically (B) or horizontally (C) viewed by an observer lying on their side. The effect of sound in shifting the contrast balance for “ambiguous motion” from 50:50 provided the horizontal and vertical components of the perceived direction of sound.

Harris and Smith, 2008). **Figure 4** shows how the direction in which the perceived location of touch shifts, may depend on whether it is coded relative to one or other of these misperceived reference directions. Displacements in a gaze-centered frame might also be evoked if the location of touch were attracted toward the direction of gaze. We can therefore conclude that touch is initially coded relative to the body midline but, if the location needs to be remembered during a gaze movement, it is switched to a gaze-based reference frame. Touch localization therefore depends on the orientation of the body and gaze. In next section we consider the effect of body size on the perception of touch.

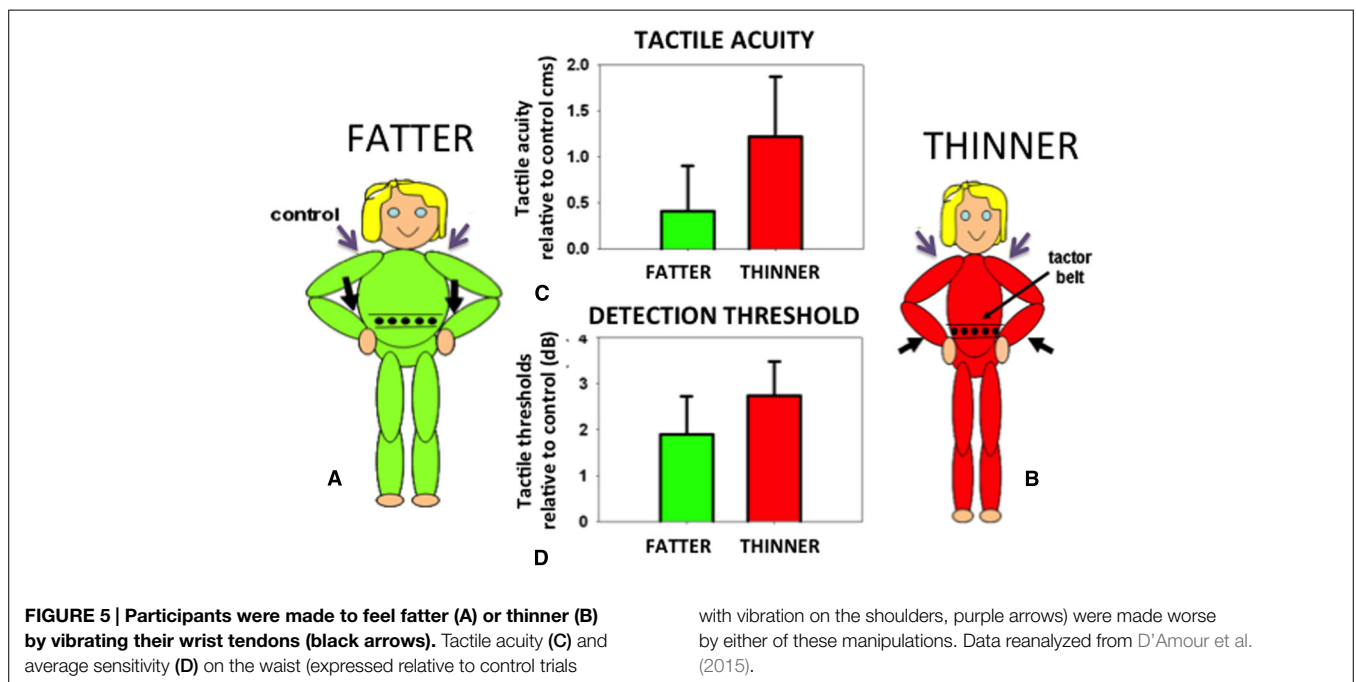
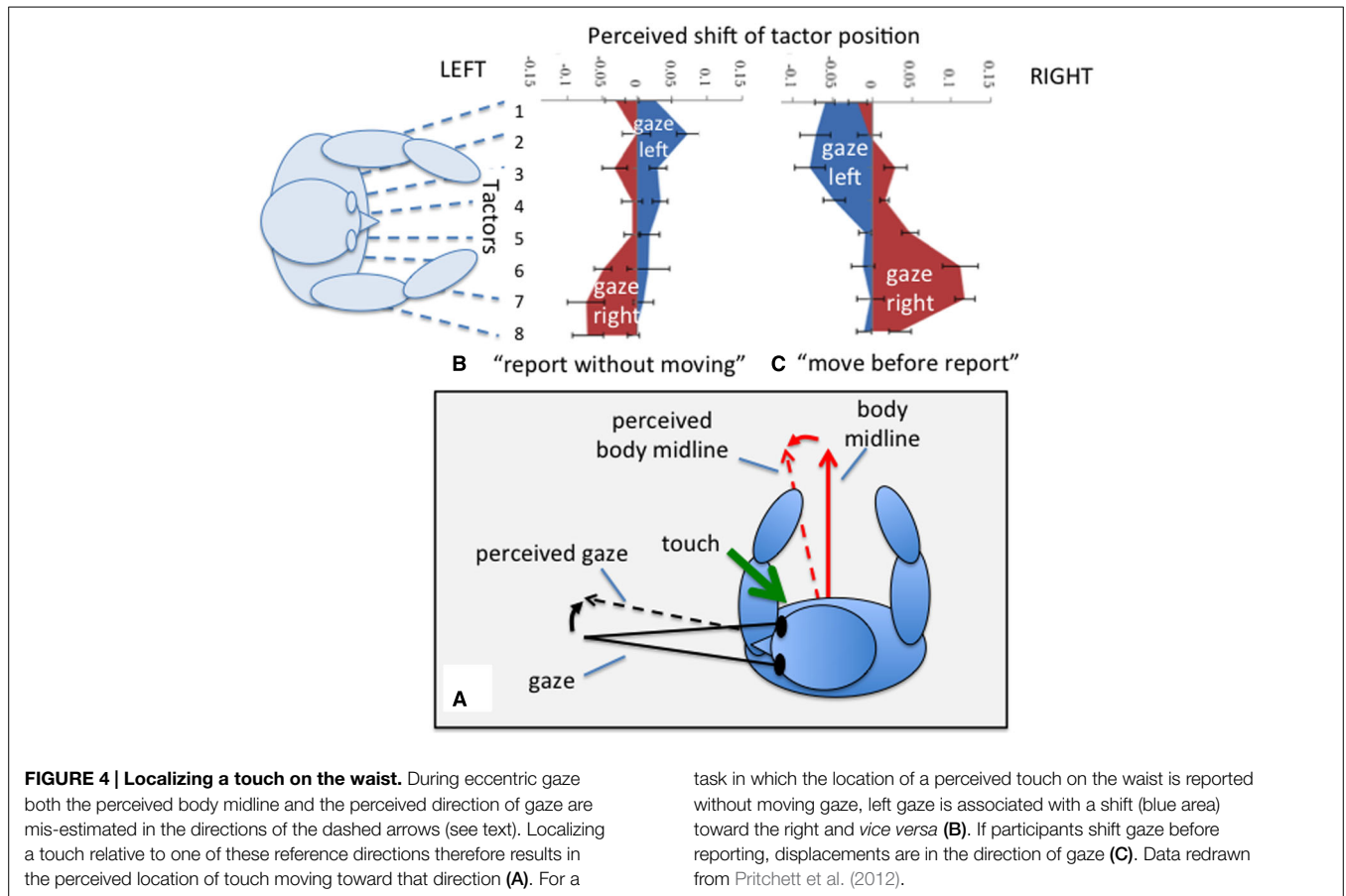
### Tactile Responses Depend on the Perceived Size of the Body

In order to identify the size of an object held against the skin it is necessary to correct for the variation in the density of tactile receptors on that part of the body surface. The object will stretch over an array of receptors on the body. The same size object will extend over a different number of receptors depending on the density of receptors in that area of skin. Receptor density must therefore be taken into account if an object's felt size and proportions are to be accurately determined. In fact, small errors in the perceived size of felt objects are found in which an object felt on an area with a high density of receptors (e.g., the hand) is judged as slightly larger than when the same object is felt on an area with a low density of receptors (e.g., the back). This phenomenon, known as the Weber Illusion (Longo and Haggard, 2011), suggests incomplete compensation for the variation in receptor density and the associated distortions of the homunculus found in the primary somatosensory cortex (Penfield and Boldrey, 1937). The perceived size of the body even in adults is rather plastic and can be altered not only in response to normal growth but also in response to altered feedback concerning body size. For example, the perceived position of a limb can be manipulated by applying vibration to the

associated tendon organs. If the affected limb is in contact with another body part, for example the tip of the nose, its perceived location in space will be inferred from the distorted position of the limb. Thus, the nose can appear lengthened: the aptly-named Pinocchio Illusion (Lackner, 1988). Such distortions in the size of a body part are passed on to objects felt on the skin (de Vignemont et al., 2005). If the body part is extended, an object pressed onto the skin is felt as correspondingly longer. Curiously, when perceived body size is distorted in either direction (made either larger or smaller) tactile sensitivity and acuity are both reduced compared to control conditions with non-tendon vibration and attention maintained constant throughout (Figure 5; D'Amour et al., 2015; but cf. Volcic et al., 2013). Distorting the perceived size of the body represents a major change in the critical, universal reference system of the brain: the body. Disrupting the body reference system has multiple fundamental consequences. But what might a reliable body reference look like?

### The Body Reference System

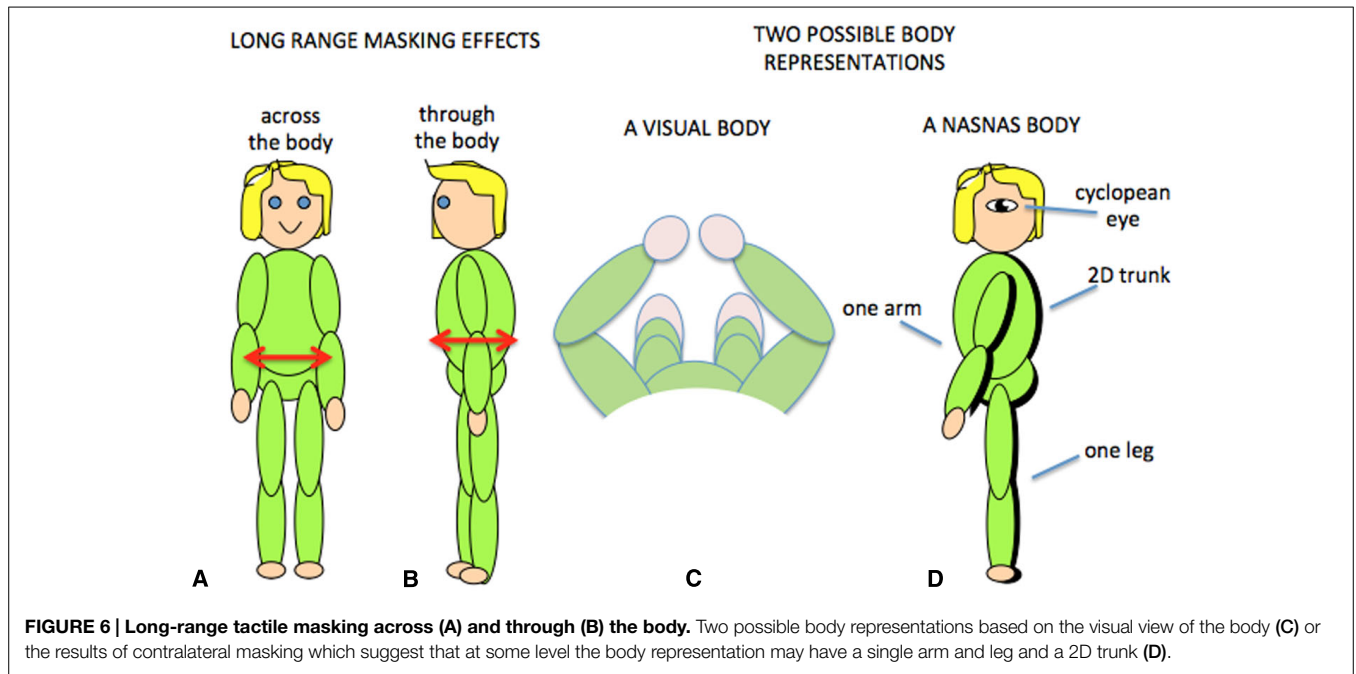
Tactile sensitivity depends on many things. We have demonstrated that it depends on the body representation (Figure 5; D'Amour et al., 2015), and it is very likely that cognitive factors such as attention are also involved (Michie et al., 1987; Spence, 2002; Gherri and Forster, 2012, 2014). An additional factor is that tactile sensitivity can be influenced by simultaneous tactile stimulation on remote areas of the body. This is known as long-range tactile masking (Sherrick, 1964; Braun et al., 2005; Tamè et al., 2011) and seems to indicate a precise connection between the representations of certain patches of skin. For example, the sensitivity to touch on one arm can be influenced by long-range masking only by touch on the corresponding point on the other arm (Figure 6A; D'Amour and Harris, 2014a). Likewise touches on the stomach can be affected by simultaneous touch on the corresponding part of the back (Figure 6B; D'Amour and Harris,



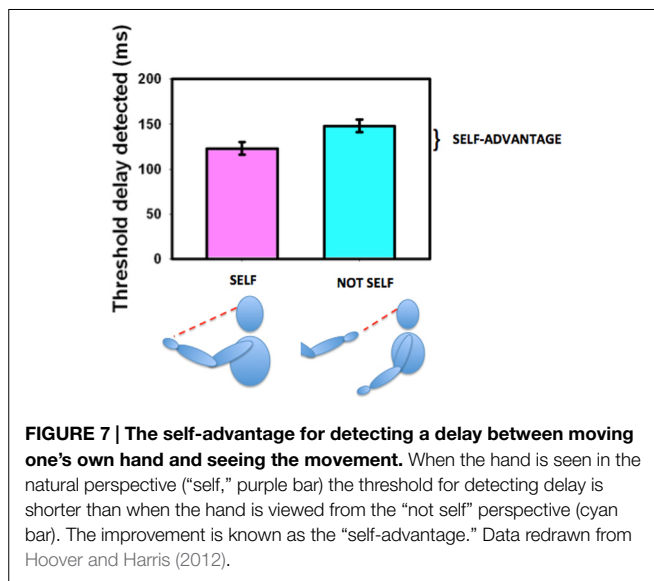
2014b). These effects are quantified relative to when the masking stimulus is positioned at another point on the body so that any attentional effects caused by the presence of a second tactile

stimulus were controlled. The question then becomes, how are the “corresponding points” defined and what can they tell us about the nature of the brain’s body schema?





**FIGURE 6 | Long-range tactile masking across (A) and through (B) the body.** Two possible body representations based on the visual view of the body (C) or the results of contralateral masking which suggest that at some level the body representation may have a single arm and leg and a 2D trunk (D).



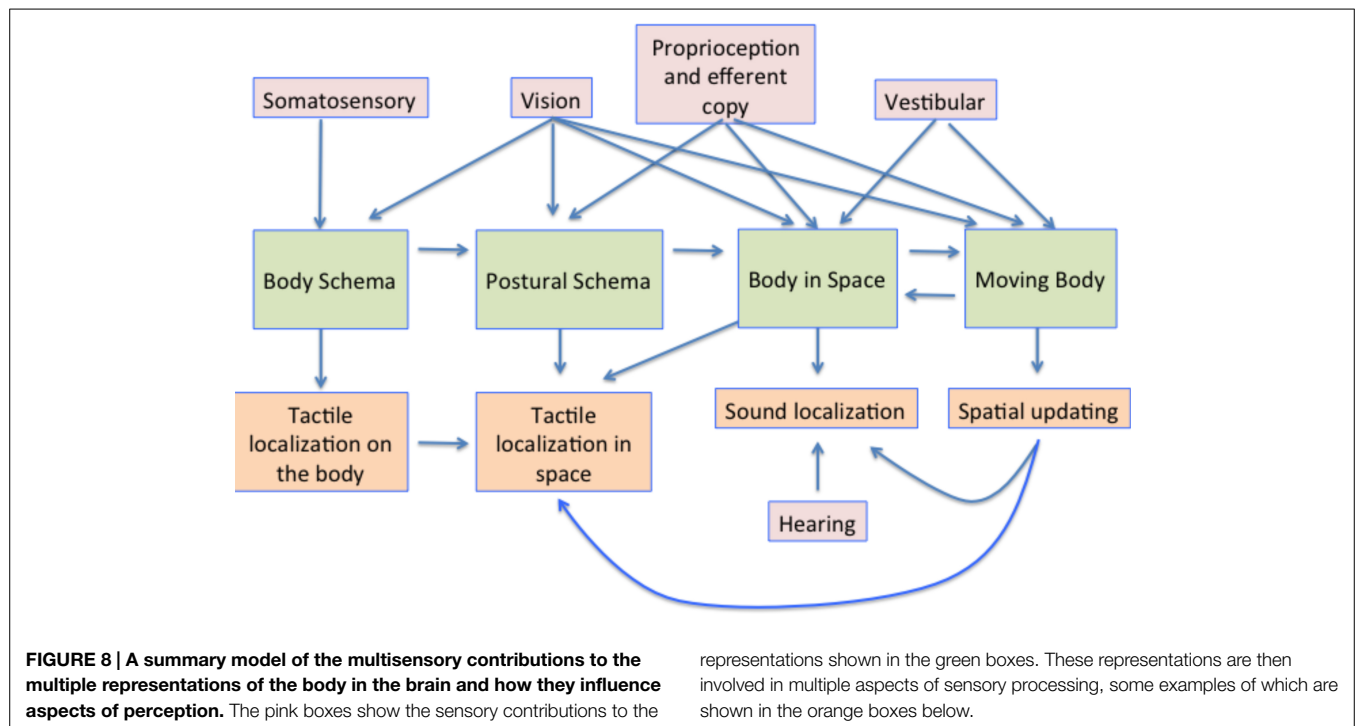
**FIGURE 7 | The self-advantage for detecting a delay between moving one's own hand and seeing the movement.** When the hand is seen in the natural perspective ("self," purple bar) the threshold for detecting delay is shorter than when the hand is viewed from the "not self" perspective (cyan bar). The improvement is known as the "self-advantage." Data redrawn from Hoover and Harris (2012).

In Head and Holmes' (1911) original description of the representation of the body in the brain, they postulate a body schema in a "canonical posture" to which the actual posture is later added. The nature of this canonical representation can only be inferred but is presumably based on statistical probabilities of where the various body parts are likely to be (Bremner et al., 2012), that is a prior with the left arm and leg on the left and *vice versa*. This might correspond to the "position of orthopedic rest" (Bromage and Melzack, 1974), the position that astronauts adopt when relaxed in zero gravity<sup>1</sup> although the detailed layout is hard to access. The prior is likely to rely on visual information about the body (Röder et al., 2004), which might provide a

<sup>1</sup>JSC-09551, Skylab Experience Bulletin No. 17—Neutral Body Posture in Zero G, NASA-JSC, 7–75 cited in <http://msis.jsc.nasa.gov/sections/section03.htm>

representation of the type shown in Figure 6C although the existence of phantom limbs in people born without arms or legs (Ramachandran and Hirstein, 1998; Brugger et al., 2000) indicates a genetic component to the body schema. Positioning the limbs in a non-canonical position (e.g., crossed) can provide hints about the canonical arrangement. If a touch is applied to the left hand while it is positioned on the right side of the body, saccades toward the touch will often start off directed toward its expected position on the left side (Groh and Sparks, 1996) and reaction times to the touch will be speeded by a visual cue on the left side (Azañón and Soto-Faraco, 2008). More detailed work is required testing many parts of the body (such as the hands, and the upper and lower sections of the limbs) to obtain a more precise impression of the canonical representation. Further, there are likely to be multiple schemas each adapted to a particular aspect of perception (de Vignemont, 2007; Longo et al., 2010).

Obviously the relationship between the front and back of the torso is fixed in all frames of reference, but for the limbs this is not the case. By varying the position of the limbs relative to each other, we have demonstrated that long-range tactile masking also depends on the position of the limbs in space (D'Amour and Harris, 2014a). Such modulation by posture suggests that long-range tactile masking is a phenomenon at or beyond the point at which the postural body schema is derived rather than at or before the level of the primary somatosensory cortex. The connections between the sides of the body has a neurophysiological correlate in which many somatosensory cells with receptive fields on the arms and hands are responsive to stimuli from either side of the body (Iwamura et al., 1994, 1993; Taoka et al., 1998). Such cells thus provide a signal that an arm was touched but do not distinguish which arm: at some level the postural schema seems to have only one arm! There is some indication that cross-body connections might also occur between the legs (Gilson, 1969; Iwamura et al., 2002, 2001) which suggests that this bilateral representation may



include the whole body (**Figure 6D**). We refer to this as a Nasnas body after the monster in *the Book of 1001 Nights*. The Nasnas body may be a somatosensory equivalent to the way that vision is referred to a single cyclopean eye (Mapp and Ono, 1999).

### The Representation of the Body in Defining the Self

The ability to move one's own body and see that it behaves in the expected way is an important aspect of determining agency (Gallagher, 2000; Tsakiris et al., 2007a,b) and thus in deriving, establishing, and maintaining a sense of ownership of our own body. We established that sensitivity for detecting delay between initiating and seeing a movement was enhanced if the moving body part were seen in its natural orientation (the first-person perspective) as opposed to if it were seen as if it were someone else's hand (from a third-person perspective, Hoover and Harris, 2012, see **Figure 7**). This variation with perspective gives us an objective measure of what the brain regards as the body's first person view. Hand and head movements that are seen from a natural first-person perspective (looking down at the hand or seeing the hand or head in a mirror) are associated with a strong self-advantage, but views of the body from behind or of an arm stretched out toward us in a third-person perspective are not (Hoover and Harris, 2015). This suggests that body parts that can be seen in a first-person perspective are preferentially treated as belonging to us (Petkova et al., 2011b). Parts of the body that cannot be seen directly and thus have no representation from a first-person perspective, such as the back of the body, may not be regarded as parts of the self in the same way as parts of the body that can be seen directly. However, this can be altered by providing an unusual first-person visual view of the back (Ehrsson, 2007; Lenggenhager et al., 2007; Spapé et al., 2015) demonstrating the

role of learning and experience in forming our perception of our "self." The suggestion that vision determines what is regarded as self either directly or from the view in a mirror, is compatible with our representation of the body in the brain as having only a two-dimensional representation of the torso as shown in **Figure 6D**.

### Discussion

This review emphasizes the reciprocal nature of the perception of our bodies in the world and the world that we perceive. Multisensory integration operates not only at the level of integrating redundant cues about object properties—such as when auditory and visual cues signal the location of an event (Alais and Burr, 2004; Burr and Alais, 2006) or when cues about the size of an object are conveyed by both vision and touch (Ernst and Banks, 2002). Multisensory integration also determines the representation of the body in the brain (Maravita et al., 2003; Petkova et al., 2011a), and this representation in turn is fundamental in interpreting all sensory information.

### The Body in the Brain

What is the nature of the body's representation in the brain? Here we are not considering the consciously accessible representation of the body which may be divided into parts known as body mereology (de Vignemont et al., 2006) with their various cultural associations and accessible to consciousness (Longo, 2014). That is better referred to as a body image. Instead we are attempting to access the internal, possibly monstrous, representation(s) to which all sensory information is related at a neurophysiological level. This representation may be fragmented (Coslett and Lie, 2004; Kammers et al., 2009; Mancini et al., 2011) and apparently

illogical in its arrangement. Many converging studies (e.g., Driver and Grossenbacher, 1996; Röder et al., 2004; Soto-Faraco et al., 2004; Longo et al., 2008) suggest that, counterintuitive to the idea of a fragmented, distorted representation, there might be a strong visual component to this representation, at least in normally sighted individuals (see **Figure 6C**). However, the view of the body is limited in the sense that only some parts can be seen at all and mostly from what we might paradoxically think of as an “odd angle” (see **Figure 6C**). In which case it is not surprising that there is reduced ownership of the back, which is not directly visible (Hoover and Harris, 2015), and that perception of the back may be closely linked to the more visible front (D’Amour and Harris, 2014b). Representing the three-dimensional body using the two-dimensional flat mapping process that seems to be so common in the brain (Chklovskii and Koulakov, 2004) clearly requires some transformations. It is necessary that sensory inputs are connected appropriately so that for example, a stimulus drawn across the body’s surface is perceived as moving continuously at a constant speed and without discontinuities as it moves from one side to the other or between regions of high and low acuity. That is, it is necessary that the unconscious, distorted body schema be related to the consciously accessible, three-dimensional body image in some.

The processes involved in creating and using a representation of the body in the brain are summarized in **Figure 8**. The body

schema, in some canonical posture, has posture added to it, using information from proprioception and vision. This representation is then situated in space using proprioceptive vision (vision about the body and its relationship with space) and vestibular cues concerning the direction of up (Harris, 2009). The movement of the body, obtained also from visual and vestibular cues also needs to be taken into account, so that the position of earth-fixed features can be appropriately updated to register their new positions relative to the body both during the movement itself and following repositioning in space.

To consider sensory functioning in isolation of the multisensory context provided by the other senses and without regard to the body of which they are a part has to be regarded as being artificial. It is now the turn of our own bodies to take central stage if we are to understand how we are able to construct our perception of the external world and interact with it.

## Acknowledgments

The core funding for the experiments described in this paper was provided by the Natural Sciences and Engineering Council of Canada (NSERC) to LH. MC, SD, LF, and LP were partly supported from the NSERC CREATE program. LF, VH, and AH were supported by NSERC post-graduate scholarships. MC, SD, and LF received Ontario Graduate Scholarships.

## References

- Alais, D., and Burr, D. C. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262. doi: 10.1016/j.cub.2004.01.029
- Aubert, H. (1886). Die bewegungsempfindung. *Pflugers Arch. J. Physiol.* 39, 347–370. doi: 10.1007/BF01612166
- Azañón, E., and Soto-Faraco, S. (2008). Changing reference frames during the encoding of tactile events. *Curr. Biol.* 18, 1044–1049. doi: 10.1016/j.cub.2008.06.045
- Batteau, D. W. (1967). The role of pinna in human localization. *Proc. R. Soc. Lond. B Biol. Sci.* 168, 158–180. doi: 10.1098/rspb.1967.0058
- Bauermeister, M., Werner, H., and Wapner, S. (1964). The effect of body tilt on tactual-kinesthetic perception of verticality. *Am. J. Psychol.* 77, 451–456. doi: 10.2307/1421016
- Blauert, J. (1996). *Spatial Hearing. The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press.
- Blohm, G., Khan, A. Z., Ren, L., Schreiber, K. M., and Crawford, J. D. (2008). Depth estimation from retinal disparity requires eye and head orientation signals. *J. Vis.* 8, 3.1–3.23. doi: 10.1167/8.16.3
- Braun, C., Hess, H., Burkhardt, M., Wuhle, A., and Preissl, H. (2005). The right hand knows what the left hand is feeling. *Exp. Brain Res.* 162, 366–373. doi: 10.1007/s00221-004-2187-4
- Bremner, A. J., Holmes, N. P., and Spence, C. (2012). “The development of multisensory representations of the body and of the space around the body,” in *Multisensory Development*, eds A. J. Bremner, D. J. Lewkowicz, and C. Spence (Oxford: Oxford University Press), 113–136.
- Bromage, P. R., and Melzack, R. (1974). Phantom limbs and the body schema. *Can. Anaesth. Soc. J.* 21, 267–274. doi: 10.1007/BF03005731
- Brugger, P., Kollias, S. S., Müri, R. M., Crelier, G., Hepp-Reymond, M. C., and Regard, M. (2000). Beyond re-membering: phantom sensations of congenitally absent limbs. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6167–6172. doi: 10.1073/pnas.100510697
- Burr, D., and Alais, D. (2006). Combining visual and auditory information. *Prog. Brain Res.* 155, 243–258. doi: 10.1016/S0079-6123(06)55014-9
- Carnevale, M. J., and Harris, L. R. (2013). The contribution of sound in determining the perceptual upright. *Multisens. Res.* 26, 125. doi: 10.1163/22134808-000S0091
- Chklovskii, D. B., and Koulakov, A. A. (2004). Maps in the brain: what can we learn from them? *Annu. Rev. Neurosci.* 27, 369–392. doi: 10.1146/annurev.neuro.27.070203.144226
- Coslett, H. B., and Lie, E. (2004). Bare hands and attention: evidence for a tactile representation of the human body. *Neuropsychologia* 42, 1865–1876. doi: 10.1016/j.neuropsychologia.2004.06.002
- Cutting, J. E., and Vishton, P. M. (1995). “Perceiving layout and knowing distances: the interaction, relative potency, and contextual use of different information about depth,” in *Perception of Space and Motion*, eds W. Epstein and S. Rogers (San Diego, CA: Academic Press), 66–117.
- D’Amour, S., and Harris, L. R. (2014a). Contralateral tactile masking between forearms. *Exp. Brain Res.* 232, 821–826. doi: 10.1007/s00221-013-3791-y
- D’Amour, S., and Harris, L. R. (2014b). Vibrotactile masking through the body. *Exp. Brain Res.* 232, 2859–2863. doi: 10.1007/s00221-014-3955-4
- D’Amour, S. D., Pritchett, L. M., and Harris, L. R. (2015). Bodily illusions disrupt tactile sensations. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 42–49. doi: 10.1037/a0038514
- de Vignemont, F. (2007). How many representations of the body? *Behav. Brain Sci.* 30, 204–205. doi: 10.1017/S0140525X07001434
- de Vignemont, F., Ehrsson, H. H., and Haggard, P. (2005). Bodily illusions modulate tactile perception. *Curr. Biol.* 15, 1286–1290. doi: 10.1016/j.cub.2005.06.067
- de Vignemont, F., Tsakiris, M., and Haggard, P. (2006). “Body mereology,” in *Human Body Perception from the Inside Out*, eds G. Knoblich, I. M. Thornton, M. Grosjean, and M. Shiffrar (Oxford: Oxford University Press), 147–170.
- Driver, J., and Grossenbacher, P. G. (1996). “Multimodal spatial constraints on tactile selective attention,” in *Attention and Performance XVI: Information Integration in Perception and Communication*, eds I. Toshio and J. L. McClelland (Cambridge, MA: MIT Press), 209–236.
- Dyde, R. T., Jenkin, M., and Harris, L. R. (2006). The subjective visual vertical and the perceptual upright. *Exp. Brain Res.* 173, 612–622. doi: 10.1007/s00221-006-0405-y
- Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science* 317, 1048. doi: 10.1126/science.1142175
- Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433. doi: 10.1038/415429a

- Fisher, H. G., and Freedman, S. J. (1968). The role of the pinna in auditory localization. *J. Aud. Res.* 8, 15–26.
- Fraser, L., Makoöie, B., and Harris, L. R. (2014). Contributions of the body and head to perceived vertical: cross-modal differences. *J. Vis.* 14, 1097. doi: 10.1167/14.10.1097
- Gallagher, I. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn. Sci.* 4, 14–21. doi: 10.1016/S1364-6613(99)01417-5
- Gherri, E., and Forster, B. (2012). The orienting of attention during eye and hand movements: ERP evidence for similar frame of reference but different spatially specific modulations of tactile processing. *Biol. Psychol.* 91, 172–184. doi: 10.1016/j.biopsycho.2012.06.007
- Gherri, E., and Forster, B. (2014). Attention to the body depends on eye-in-orbit position. *Front. Psychol.* 5:683. doi: 10.3389/fpsyg.2014.00683
- Gilson, R. D. (1969). Vibrotactile masking: some spatial and temporal aspects. *Percept. Psychophys.* 5, 176–180. doi: 10.3758/BF03209553
- Gogel, W. (1963). The visual perception of size and distance. *Vision Res.* 3, 101–120. doi: 10.1016/0042-6989(63)90033-6
- Goossens, H. H., and van Opstal, A. J. (1999). Influence of head position on the spatial representation of acoustic targets. *J. Neurophysiol.* 81, 2720–2736.
- Groh, J. M., and Sparks, D. L. (1996). Saccades to somatosensory targets. 1. behavioral characteristics. *J. Neurophysiol.* 75, 412–427.
- Guerraz, M., Poquin, D., Luyat, M., and Ohlmann, T. (1998). Head orientation involvement in assessment of the subjective vertical during whole body tilt. *Percept. Mot. Skills* 87, 643–648.
- Harrar, V., and Harris, L. R. (2009). Eye position affects the perceived location of touch. *Exp. Brain Res.* 198, 403–410. doi: 10.1007/s00221-009-1884-4
- Harrar, V., and Harris, L. R. (2010). Touch used to guide action is partially coded in a visual reference frame. *Exp. Brain Res.* 203, 615–620. doi: 10.1007/s00221-010-2252-0
- Harrar, V., Pritchett, L. M., and Harris, L. R. (2013). Segmented space: measuring tactile localisation in body coordinates. *Multisens. Res.* 26, 3–18. doi: 10.1163/22134808-00002399
- Harris, L. R., and Mander, C. (2014). Perceived distance depends on the orientation of both the body and the visual environment. *J. Vis.* 14, 1–8. doi: 10.1167/14.12.17
- Harris, L. R., and Smith, A. (2008). The coding of perceived eye position. *Exp. Brain Res.* 187, 429–437. doi: 10.1007/s00221-008-1313-0
- Harris, L. R. (2009). “Visual-vestibular interactions,” in *The International Encyclopedia of Neuroscience*, ed. L. R. Squire (North Holland: Elsevier), 381–387.
- Harris, L. R., Fraser, L., and Makoöie, B. (2014). Vestibular and neck proprioceptive contributions to the perceived vertical. *Iperception* 5, 441.
- Head, H., and Holmes, G. (1911). Sensory disturbances from cerebral lesions. *Brain* 34, 102–254. doi: 10.1093/brain/34.2-3.102
- Hershenson, M. (1989). *The Moon Illusion*. New Jersey, NJ: Lawrence Erlbaum Associates.
- Hill, A. L. (1972). Directional constancy. *Percept. Psychophys.* 11, 175–178. doi: 10.3758/BF03210370
- Ho, C., and Spence, C. (2007). Head orientation biases tactile localization. *Brain Res.* 1144, 136–141. doi: 10.1016/j.brainres.2007.01.091
- Hoover, A. E. N., and Harris, L. R. (2012). Detecting delay in visual feedback of an action as a monitor of self recognition. *Exp. Brain Res.* 222, 389–97. doi: 10.1007/s00221-012-3224-3
- Hoover, A. E. N., and Harris, L. R. (2015). The role of the viewpoint on body ownership. *Exp. Brain Res.* 233, 1053–1060. doi: 10.1007/s00221-014-4181-9
- Howard, I. P., and Hu, G. (2001). Visually induced reorientation illusions. *Perception* 30, 583–600. doi: 10.1068/p3106
- Iwamura, Y., Iriki, A., and Tanaka, M. (1994). Bilateral hand representation in the postcentral somatosensory cortex. *Nature* 369, 554–556. doi: 10.1038/369554a0
- Iwamura, Y., Tanaka, M., Iriki, A., Taoka, M., and Toda, T. (2002). Processing of tactile and kinesthetic signals from bilateral sides of the body in the postcentral gyrus of awake monkeys. *Behav. Brain Res.* 135, 185–190. doi: 10.1016/S0166-4328(02)00164-X
- Iwamura, Y., Tanaka, M., Sakamoto, M., and Hikosaka, O. (1993). Rostrocaudal gradients in the neuronal receptive field complexity in the finger region of the alert monkey's postcentral gyrus. *Exp. Brain Res.* 92, 360–368. doi: 10.1007/BF00229023
- Iwamura, Y., Taoka, M., and Iriki, A. (2001). Bilateral activity and callosal connections in the somatosensory cortex. *Neuroscientist* 7, 419–429. doi: 10.1177/107385840100700511
- Kammers, M. P., Longo, M. R., Tsakiris, M., Dijkerman, H. C., and Haggard, P. (2009). Specificity and coherence of body representations. *Perception* 38, 1804–1820. doi: 10.1068/p6389
- Lackner, J. R. (1988). Some Proprioceptive influences on the perceptual representation of body shape and orientation. *Brain* 111, 281–297. doi: 10.1093/brain/111.2.281
- Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science* 317, 1096–1099. doi: 10.1126/science.1143439
- Longo, M. R. (2014). Three-dimensional coherence of the conscious body image. *Q. J. Exp. Psychol.* 68, 1116–1123. doi: 10.1080/17470218.2014.975731
- Longo, M. R., Azañón, E., and Haggard, P. (2010). More than skin deep: body representation beyond primary somatosensory cortex. *Neuropsychologia* 48, 655–668. doi: 10.1016/j.neuropsychologia.2009.08.022
- Longo, M. R., Cardozo, S., and Haggard, P. (2008). Visual enhancement of touch and the bodily self. *Conscious. Cogn.* 17, 1181–1191. doi: 10.1016/j.concog.2008.01.001
- Longo, M. R., and Haggard, P. (2011). Weber's illusion and body shape: anisotropy of tactile size perception on the hand. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 720–726. doi: 10.1037/a0021921
- Makous, J. C., and Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *J. Acoust. Soc. Am.* 87, 2188–2200. doi: 10.1121/1.399186
- Mancini, F., Longo, M. R., Iannetti, G. D., and Haggard, P. (2011). A supramodal representation of the body surface. *Neuropsychologia* 49, 1194–1201. doi: 10.1016/j.neuropsychologia.2010.12.040
- Mapp, A., and Ono, H. (1999). Wondering about the wandering cyclopean eye. *Vision Res.* 39, 2381–2386. doi: 10.1016/S0042-6989(98)00278-8
- Maravita, A., Spence, C., and Driver, J. (2003). Multisensory integration and the body schema: close to hand and within reach. *Curr. Biol.* 13, R531–R539. doi: 10.1016/S0960-9822(03)00449-4
- Merleau-Ponty, M. (1945). *Phenomenology of Perception*. Paris: Gallimard.
- Michie, P. T., Bearpark, H. M., Crawford, J. M., and Glue, L. C. (1987). The effects of spatial selective attention on the somatosensory event-related potential. *Psychophysiology* 24, 449–463. doi: 10.1111/j.1469-8986.1987.tb00316.x
- Mittelstaedt, H. (1983). A new solution to the problem of the subjective vertical. *Naturwissenschaften* 70, 272–281. doi: 10.1007/BF00404833
- Morgan, C. L. (1978). Constancy of egocentric visual direction. *Percept. Psychophys.* 23, 61–63. doi: 10.3758/BF03214296
- Mueller, S., and Fiehler, K. (2014). Effector movement triggers gaze-dependent spatial coding of tactile and proprioceptive-tactile reach targets. *Neuropsychologia* 62, 184–193. doi: 10.1016/j.neuropsychologia.2014.07.025
- Owens, D. A., Hahn, J. P., Francis, E. L., and Wist, E. R. (1990). Binocular vergence and perceived object velocity: a new illusion. *Investig. Ophthalmol. Vis. Sci.* 31(Suppl.), 93.
- Parise, C. V., Knorre, K., and Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proc. Natl. Acad. Sci. U. S. A.* 111, 6104–6108. doi: 10.1073/pnas.1322705111
- Penfield, W., and Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain* 60, 389–443. doi: 10.1093/brain/60.4.389
- Petkova, V. I., Bjornsdotter, M., Gentile, G., Jonsson, T., Li, T. Q., and Ehrsson, H. H. (2011a). From part- to whole-body ownership in the multisensory brain. *Curr. Biol.* 21, 1118–1122. doi: 10.1016/j.cub.2011.05.022
- Petkova, V. I., Khoshnevis, M., and Ehrsson, H. H. (2011b). The perspective matters! Multisensory integration in ego-centric reference frames determines full-body ownership. *Front. Psychol.* 2:35. doi: 10.3389/fpsyg.2011.00035
- Philbeck, J. W., and Loomis, J. M. (1997). Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 72–85. doi: 10.1037/0096-1523.23.1.72
- Pritchett, L. M., Carnevale, M. J., and Harris, L. R. (2012). Reference frames for coding touch location depend on the task. *Exp. Brain Res.* 222, 437–445. doi: 10.1007/s00221-012-3231-4
- Pritchett, L. M., and Harris, L. R. (2011). Perceived touch location is coded using a gaze signal. *Exp. Brain Res.* 213, 229–234. doi: 10.1007/s00221-011-2713-0
- Ramachandran, V. S., and Hirstein, W. (1998). The perception of phantom limbs. The D. O. Hebb lecture. *Brain* 121, 1603–1630. doi: 10.1093/brain/121.9.1603
- Rock, I., and Kaufman, L. (1962). The Moon Illusion II: the moon's apparent size is a function of the presence or absence of terrain. *Science* 136, 1023–1031. doi: 10.1126/science.136.3521.1023



- Röder, B., Rosler, F., and Spence, C. (2004). Early vision impairs tactile perception in the blind. *Curr. Biol.* 14, 121–124. doi: 10.1016/j.cub.2003.12.054
- Ross, H. E., and Plug, C. (2002). *The Mystery of the Moon Illusion: Exploring Size Perception*. Oxford: Oxford University Press.
- Schreiber, K. M., Hillis, J. M., Filippini, H. R., Schor, C. M., and Banks, M. S. (2008). The surface of the empirical horopter. *J. Vis.* 8, 7.1–7.20. doi: 10.1167/8.3.7
- Sherrick, C. (1964). Effects of double simultaneous stimulation of the skin. *Am. J. Psychol.* 77, 42–53. doi: 10.2307/1419270
- Soto-Faraco, S., Ronald, A., and Spence, C. (2004). Tactile selective attention and body posture: assessing the multisensory contributions of vision and proprioception. *Percept. Psychophys.* 66, 1077–1094. doi: 10.3758/BF03196837
- Spapé, M. M., Ahmed, I., Jacucci, G., and Ravaja, N. (2015). The self in conflict: actors and agency in the mediated sequential Simon task. *Front. Psychol.* 6:304. doi: 10.3389/fpsyg.2015.00304
- Spence, C. (2002). Multisensory attention and tactile information-processing. *Behav. Brain Res.* 135, 57–64. doi: 10.1016/S0166-4328(02)00155-9
- Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Atten. Percept. Psychophys.* 73, 971–995. doi: 10.3758/s13414-010-0073-7
- Tamè, L., Farnè, A., and Pavani, F. (2011). Spatial coding of touch at the fingers: insights from double simultaneous stimulation within and between hands. *Neurosci. Lett.* 487, 78–82. doi: 10.1016/j.neulet.2010.09.078
- Taoka, M., Toda, T., and Iwamura, Y. (1998). Representation of the midline trunk, bilateral arms, and shoulders in the monkey postcentral somatosensory cortex. *Exp. Brain Res.* 123, 315–322. doi: 10.1007/s002210050574
- Tarnutzer, A. A., Bockisch, C. J., and Straumann, D. (2009). Head roll dependent variability of subjective visual vertical and ocular counterroll. *Exp. Brain Res.* 195, 621–626. doi: 10.1007/s00221-009-1823-4
- Tsakiris, M., Hesse, M. D., Boy, C., Haggard, P., and Fink, G. R. (2007a). Neural signatures of body ownership: a sensory network for bodily self-consciousness. *Cereb. Cortex* 17, 2235–2244. doi: 10.1093/cercor/bhl131
- Tsakiris, M., Schutz-Bosbach, S., and Gallagher, S. (2007b). On agency and body-ownership: phenomenological and neurocognitive reflections. *Conscious. Cogn.* 16, 645–660. doi: 10.1016/j.concog.2007.05.012
- Volcic, R., Fantoni, C., Caudek, C., Assad, J. A., and Domini, F. (2013). Visuomotor adaptation changes stereoscopic depth perception and tactile discrimination. *J. Neurosci.* 33, 17081–17088. doi: 10.1523/JNEUROSCI.2936-13.2013
- Weidner, R., Plewan, T., Chen, Q., Buchner, A., Weiss, P. H., and Fink, G. R. (2014). The moon illusion and size-distance scaling—evidence for shared neural patterns. *J. Cogn. Neurosci.* 26, 1871–1882. doi: 10.1162/jocn\_a\_00590
- Yamaguchi, M., and Kaneko, H. (2007). Integration system of head, eye, and retinal position signals for perceptual direction. *Opt. Rev.* 14, 411–415. doi: 10.1007/s10043-007-0411-8

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Harris, Carnevale, D'Amour, Fraser, Harrar, Hoover, Mander and Pritchett. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Nonlinear response speedup in bimodal visual-olfactory object identification

Richard Höchenberger<sup>1\*</sup>, Niko A. Busch<sup>2,3</sup> and Kathrin Ohla<sup>1</sup>

<sup>1</sup> Psychophysiology of Food Perception, German Institute of Human Nutrition Potsdam-Rehbrücke (DIFE), Nuthetal, Germany, <sup>2</sup> Institute of Medical Psychology, Charité - Universitätsmedizin Berlin, Berlin, Germany, <sup>3</sup> Berlin School of Mind and Brain, Humboldt University, Berlin, Germany

## OPEN ACCESS

### Edited by:

Andriy Myachykov,  
Northumbria University, UK

### Reviewed by:

Wen Li,  
Florida State University, USA  
Ashley James Chapman,  
Northumbria University, UK

### \*Correspondence:

Richard Höchenberger,  
German Institute of Human Nutrition  
Potsdam-Rehbrücke (DIFE),  
Arthur-Scheunert-Allee 114–116,  
14558 Nuthetal, Germany  
hoechenberger@dife.de

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 10 April 2015

**Accepted:** 14 September 2015

**Published:** 30 September 2015

### Citation:

Höchenberger R, Busch NA and  
Ohla K (2015) Nonlinear response  
speedup in bimodal visual-olfactory  
object identification.  
Front. Psychol. 6:1477.  
doi: 10.3389/fpsyg.2015.01477

Multisensory processes are vital in the perception of our environment. In the evaluation of foodstuff, redundant sensory inputs not only assist the identification of edible and nutritious substances, but also help avoiding the ingestion of possibly hazardous substances. While it is known that the non-chemical senses interact already at early processing levels, it remains unclear whether the visual and olfactory senses exhibit comparable interaction effects. To address this question, we tested whether the perception of congruent bimodal visual-olfactory objects is facilitated compared to unimodal stimulation. We measured response times (RT) and accuracy during speeded object identification. The onset of the visual and olfactory constituents in bimodal trials was physically aligned in the first and perceptually aligned in the second experiment. We tested whether the data favored coactivation or parallel processing consistent with race models. A redundant-signals effect was observed for perceptually aligned redundant stimuli only, i.e., bimodal stimuli were identified faster than either of the unimodal components. Analysis of the RT distributions and accuracy data revealed that these observations could be explained by a race model. More specifically, visual and olfactory channels appeared to be operating in a parallel, positively dependent manner. While these results suggest the absence of early sensory interactions, future studies are needed to substantiate this interpretation.

**Keywords:** multisensory integration, olfaction, visual-olfactory, race model, response time

## 1. Introduction

Olfactory and visual sensory information are continuously flooding the brain and are, therefore, often experienced with a marked temporal overlap or even simultaneously. Both the smell and visual appearance serve a vital function in the localization of food, the assessment of edibility, as well as the identification of potential environmental hazards, thereby allowing for fast and appropriate behavior not only limited to food-choice. The integration of redundant sensory information by the neural system has been proven beneficial for perception and subsequent behavior: it speeds up processing and improves accuracy. However, it is unclear whether this holds true for the combination of olfaction and vision.

Recent studies have shown that odors modulate visual perception and performance, particularly by directing attention to and influencing the saliency of a congruent visual object, e.g., during attentional blink (Robinson et al., 2013), binocular rivalry (Zhou et al., 2010, 2012), spatial

attention and visual search (Chen et al., 2013), and eye movements (Seo et al., 2010). These effects occur even when odors are task-irrelevant and suggest spontaneous binding between visual and olfactory inputs (Zhou et al., 2012). In contrast, odor perception is not only influenced by vision (and the other senses), but odor identification also critically depends on additional information because odors in isolation are notoriously ambiguous (Cain, 1979). Observations that humans have difficulties identifying and discriminating odors in the absence of additional information (Davis, 1981) and that color cues (Zellner et al., 1991), verbal labels (Herz and von Clef, 2001) and images (Gottfried and Dolan, 2003) assist odor perception corroborate this notion. Most previous studies investigated modulatory effects of visual cues on olfaction and their interaction at cognitive levels, when semantic representations were available. It remains unknown whether sensory information from the olfactory and visual modalities is in fact pooled at early perceptual stages, that is, integrated.

### 1.1. How can we Investigate Whether Multisensory Integration is Taking Place?

Multisensory integration has been mostly studied between the non-chemosensory modalities vision, hearing, and somatosensation; these senses have been shown to interact already at the level of the superior colliculi (Stein and Meredith, 1990). Classically, super-additive responses, i.e., more than the sum of the parts, are considered an indication of multisensory integration. Key aspects governing multisensory integration are the so-called principles of spatial and temporal proximity: stimuli presented at the same location and at the same time, respectively, most likely belong to the same object and are therefore more likely to be bound together to a unitary percept (Stein and Meredith, 1993). Additionally, Meredith and Stein (1986) found that cells in the superior colliculus produced the strongest response amplification for the weakest stimuli, a principal phenomenon called *inverse effectiveness*. While these observations could be replicated on the behavioral level in numerous studies, it has been suggested that these findings might largely be statistical artifacts (Holmes, 2007). While imaging studies have mostly focused on superadditive effects when trying to identify functional correlates of multisensory integration, it is unclear whether the results from single-neuron recordings can be readily transferred to the cortical level (Laurienti et al., 2005) and behavior.

### 1.2. Response Facilitation Can Serve as a Possible Measure of Multisensory Integration

Stimulus detection, on average, is faster and more accurate in situations where the target is presented redundantly, i.e., on several sensory channels. In the multisensory literature, this facilitation is commonly called *redundant-targets effect* (RTE) or *redundant-signals effect* (RSE); both terms are largely used synonymously. In the remainder of this paper, we will refer to these effects of multisensory processing exclusively as redundant-signals effects. The response speedup is commonly explained by assuming that an internal decision criterion is reached faster when multiple targets are presented simultaneously, compared

to the single-target situation. Similarly, redundant information reduces stimulus ambiguity, hence allowing for a higher accuracy of responses.

However, RSEs can result from *statistical facilitation* merely due to probability summation alone. A popular probability summation model was introduced by Raab (1962) with the idea of a *race* between parallel single-target detection processes during a multiple-target situation. The process finishing first “wins the race,” elicits a response, and, therefore, determines the behavioral response time. These so-called race models operate according to a *separate-activation model* with a *first-terminating stopping rule* (see e.g., Colonius and Vorberg, 1994). They implicitly assume unlimited-capacity processing (Colonius, 1990), meaning that the speed of one detection process is not influenced by other, simultaneous, detection processes. For example, detection of a unimodal target should happen at the same speed as detection of the same target in a multimodal situation. Therefore, if RT distributions of the single-target detection processes overlap, the observed RTs in redundant-target trials will, on average, be faster than the unimodal RTs. “Slow” responses of one single-target detection process can be replaced by “faster” responses of another, simultaneous detection process. The observed RT speedup would thus be a statistical artifact only. In sum, an RSE that can be fully accounted for by a race model does not provide strong evidence for multisensory integration.

Nevertheless, integration can be inferred if RTs are faster than predicted by race models. Specifically, Miller (1982) derived an upper bound to the bimodal RT speedup possible in *any* race model, the so-called *race model inequality* (RMI) or *Miller bound*. It is based on the assumption of maximum negative dependence between the channel processing speeds (Colonius, 1990), that is, if the participant detects a signal on one channel at a given bimodal trial, the other channel will fail to detect the target. Violations of this criterion, i.e., faster responses than predicted by the RMI, support *coactivation models*. They demand that processing of different sensory channels be pooled prior to the decision stage and therefore refute all race models in favor of “true” multisensory integration. Satisfaction of the RMI, on the contrary, does not necessarily exclude coactive processing.

Numerous studies have investigated response facilitation to bimodal stimuli in the visual, auditory, and somatosensory modalities (see e.g., Gielen et al., 1983; Miller, 1986; Forster et al., 2002; Diederich and Colonius, 2004). Whether the combined presentation of congruent (that is, redundant) visual-olfactory information can likewise facilitate object perception remains unclear and was investigated with the present study. Specifically, we tested the hypothesis that bimodal visual-olfactory object identification is facilitated compared to identification of either of the unimodal constituents alone.

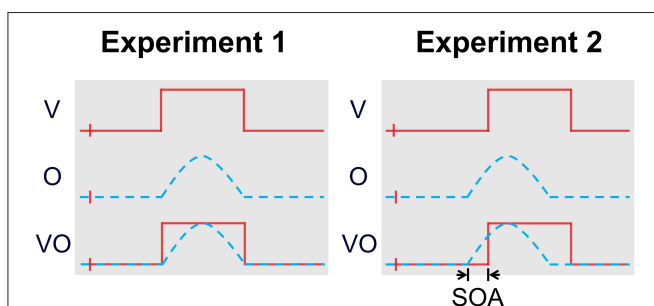
Furthermore, we examined whether facilitation is more pronounced for perceptually aligned compared to physically aligned stimuli. For this, we conducted two experiments in which the bimodal constituents were either presented physically (Experiment 1) or perceptually (Experiment 2) simultaneously. We compared the observed RTs and response accuracies to the predictions of different models of probability summation.

## 2. Results

Seven participants smelled and viewed different food objects presented either alone as unimodal visual (V) or olfactory (O) stimuli or as congruent bimodal combinations (OV) and performed a speeded two-alternative forced-choice (2-AFC) object identification task. Stimulus strength was adjusted to achieve approximately 75% accuracy. The biggest RSE for response time (RT) can be observed when the RT distributions of the unimodal constituents overlap largely (Hilgard, 1933; Hershenson, 1962; Raab, 1962; Miller, 1986; Colonius, 1990; Diederich and Colonius, 2004; Gondan, 2009). Therefore, we conducted two separate experiments, in which OV stimuli consisted of *physically* (Experiment 1; **Figure 1**, left) or *perceptually* (Experiment 2; **Figure 1**, right) aligned unimodal constituents. Perceptual alignment was achieved by introduction of a stimulus-onset asynchrony (SOA) equal to the RT differences between the unimodal stimuli.

For perceptually aligned OV stimuli (Experiment 2) we observed a significant RSE for RTs [RSE = 54 ms,  $t_{(6)} = 3.05$ ,  $p = 0.02$ ], that is a speedup of median RTs to bimodal OV compared to the fastest unimodal stimuli (**Figure 2A**). No RSE was found for physically aligned OV stimuli during Experiment 1 [RSE = 0 ms,  $t_{(6)} = 0.02$ ,  $p = 0.98$ ]. Accuracy showed no RSE in either Experiment [Experiment 1: RSE = 1.8 %-points,  $t_{(6)} = 0.73$ ,  $p = 0.49$ ; Experiment 2: RSE = -0.6 %-points,  $t_{(6)} = -0.29$ ,  $p = 0.78$ ; **Figure 2B**].

Experiment 1 yielded a significant difference between response times (RT) to unimodal V and O stimuli, indicating only little overlap between the unimodal RT distributions; V stimuli were perceived 362 ms faster than O stimuli [ $t_{(6)} = -4.72$ ,  $p < 0.01$ ; **Figure 3A**]. By contrast, we could observe a markedly reduced difference between the SOA-corrected unimodal RTs in Experiment 2 of only 74 ms [ $t_{(6)} = -2.48$ ,  $p < 0.05$ ; **Figure 3B**], indicating a strong overlap of unimodal RT distributions.

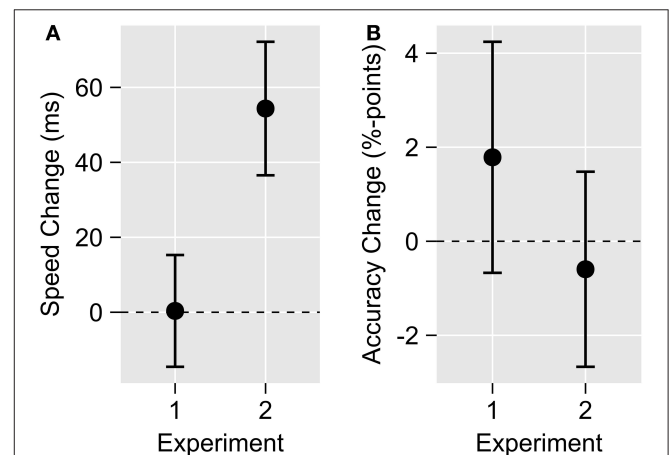


**FIGURE 1 | Stimulus timing in unimodal and bimodal trials of Experiments 1 and 2 (schematic).** The start of a trial is marked with a small vertical line and identifies the onset of the fixation cross. The visual and olfactory constituents were presented physically simultaneously in the bimodal trials in Experiment 1, and perceptually simultaneously in Experiment 2. The SOAs employed in Experiment 2 were individually estimated for every participant and object (banana and lemon, respectively) during Experiment 1. Note that the depicted delayed presentation of the visual stimulus in Experiment 2 caused the fixation cross to be displayed for a longer duration before stimulus onset in the unimodal vision-only condition in order to ensure context-invariance.

These performance differences were clearly reflected in the cumulative RT distributions: While the bimodal distribution mostly followed the visual distribution in Experiment 1, it was shifted toward faster responses throughout its whole range in Experiment 2 (**Figures 3C,D**).

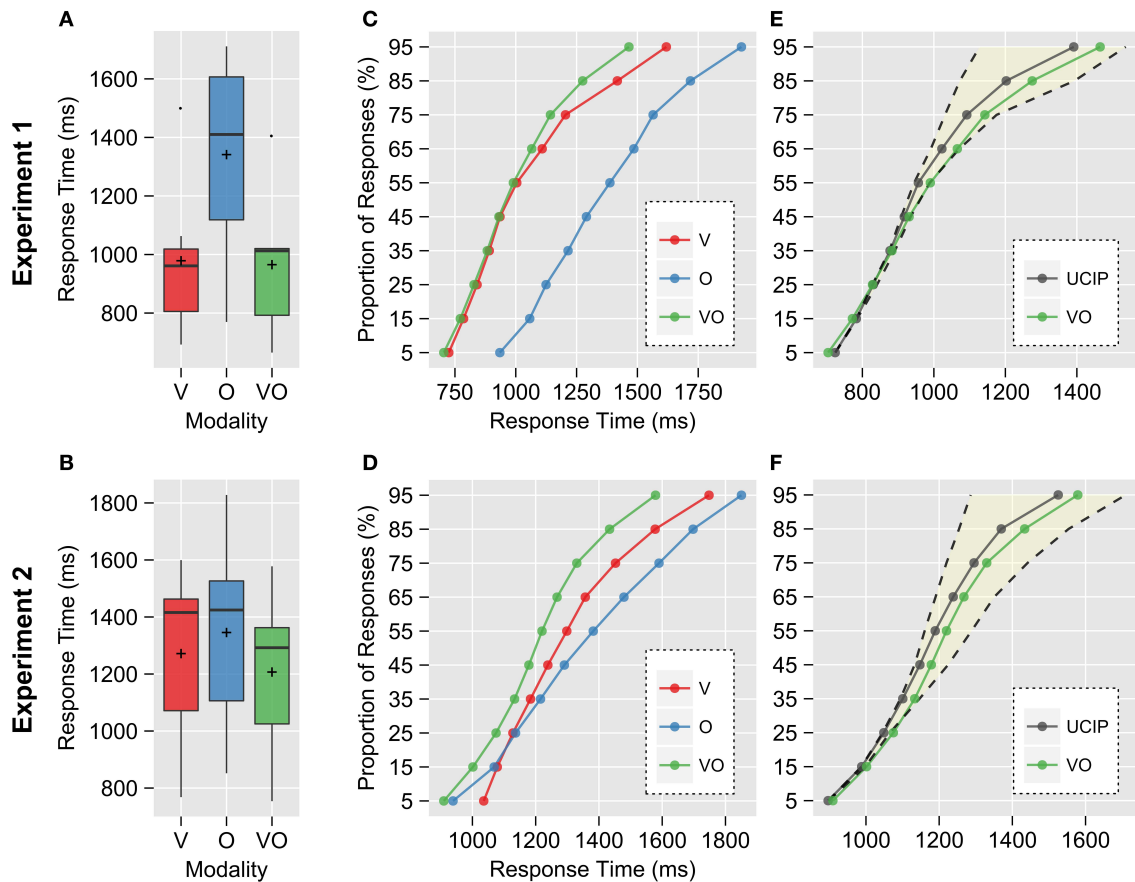
We next tested whether the bimodal RT speedup could be explained by statistical facilitation in a separate-activation model with unlimited capacity (race model). The theoretical upper performance limit was given by the *Miller bound*: If observed responses were *faster* than this boundary at any time, all race models could be ruled out at once, and the system would be assumed to be super-capacity at this time (Townsend and Nozawa, 1995; Townsend and Wenger, 2004). Additionally, we compared our data to a lower performance bound proposed by Grice et al. (1984b), referred to as *Grice bound*. It assumes that responses in the bimodal situation should be at least as fast as responses to the fastest unimodal constituents. If responses were *slower* than this boundary, the system would be assumed to be limited-capacity (Townsend and Wenger, 2004) at this time. Both super-capacity and limited-capacity processing violate the assumption of context-invariance, invalidating an essential requirement of race models (Colonius, 1990). We found that the observed bimodal RTs did not exceed the *Miller* or the *Grice bounds* significantly in both experiments, indicating parallel processing of the visual and olfactory channels (**Figures 3E,F**, and **Table 1**). Our data could thus be attributed for by a race model.

Colonius (1990) pointed out that the *Miller* and *Grice bounds* can only be reached under the implicit assumption of perfect negative and positive, respectively, dependence between channel processing speeds. To gain further insight into the



**FIGURE 2 | Mean redundant-signals effects in Experiments 1 and 2.** Positive values indicate a bimodal facilitation, negative values a bimodal impairment relative to the unimodal constituents. **(A)** No redundancy gain in response speed was observed for physically simultaneous bimodal stimulation (Experiment 1), but it was clearly evident for perceptually simultaneous stimulation (Experiment 2). **(B)** For both physically simultaneous (Experiment 1) and perceptually simultaneous bimodal stimulation (Experiment 2), no significant accuracy improvement could be observed. Data were calculated individually for each participant and object, and subsequently averaged. Error bars show standard error of the mean.





**FIGURE 3 | Comparison of observed response times (RTs) in the visual (V), olfactory (O), and bimodal (VO) trials.** Visual RTs were corrected for SOA. **(A)** RT distributions for V and VO appeared to be almost identical, with the O distribution shifted to much slower RTs and very little overlap with V, indicating that responses in VO trials were mostly driven by the visual constituent. The crosses depict group means. **(B)** Individual timing adjustments by introduction of an SOA aligned the unimodal distributions, suggesting perceptual simultaneity. **(C,D)** Empirical cumulative RT distributions. Quantile values were averaged across participants. In Experiment 1, the VO distribution seemed to follow V up to the 55% quantile. Approximately at the same time, the fastest olfactory responses could be observed, i.e., the unimodal constituents were starting to perceptually overlap. Coincidentally, the VO distribution started to diverge from V, and shifted to faster responses. This bimodal speedup is not reflected in the global (mean) RSE. In Experiment 2, the VO distribution was shifted to faster responses relative to both unimodal distributions across its whole range. **(E,F)** Comparison of the unlimited capacity, independent, parallel (UCIP) model prediction with the observed bimodal data. The highlighted area depicts the possible phase space under the assumption of separate-activation models with unlimited capacity and a first-terminating time rule, but possibly dependent processing (i.e., possible race models would have to lie within this area); accordingly, the dashed line to the left shows the Miller and the right the Grice bound (upper and lower performance limits, respectively).

underlying processing mechanisms, we compared our data to a model assuming uncorrelated processing between the visual and olfactory channels, the so-called *unlimited-capacity, independent, parallel* (UCIP) model. The bimodal RTs were slower than predicted by this model in the 75 % quantile in Experiment 1 and from the 45 % to the 85 % quantiles in Experiment 2 (all  $p < 0.05$ ). However, only the deviation in the 75 % quantile in Experiment 2 survived Holm-Bonferroni correction for multiple testing ( $p_{\text{corr}} = 0.03$ ). All comparisons are summarized in **Table 1**. Because the deviations of the observed data from the model predictions are shifted in direction of the *Grice bound*, i.e., toward perfect positive dependence, the results suggest a race model with positively correlated channel processing speed between the visual and olfactory channels (see Grice et al., 1984a).

Next, the accuracy data (grand means shown in **Figures 4A,B** for Experiments 1 and 2, respectively) were compared to models of probability summation. We first adopted equivalents of the *Miller* and *Grice bounds* to derive upper and lower performance limits, respectively (Colonius, 2015). The upper bound was at 100 % accuracy in both experiments and therefore never violated; observed accuracies were significantly below this bound [Experiment 1:  $\Delta = -13.5\%$ -points,  $t_{(6)} = -8.95$ ,  $p < 0.001$ ; Experiment 2:  $\Delta = -13.9\%$ -points,  $t_{(6)} = -6.76$ ,  $p < 0.001$ ]. The lower bound was never significantly violated [Experiment 1:  $\Delta = 1.8\%$ -points,  $t_{(6)} = 0.73$ ,  $p = 0.49$ ; Experiment 2:  $\Delta = -0.6\%$ -points,  $t_{(6)} = -0.29$ ,  $p = 0.78$ ]. Note that the lower bound was identical to the baseline used earlier to identify an RSE for accuracy. These results suggest that probability summation could in fact explain the observed bimodal accuracies.

**TABLE 1 | Each participant contributed one quantile value, following the procedure from Ulrich et al. (2007).**

Experiment	Quantile (%)	VO–Miller bound			VO–Grice bound			VO–UCIP		
		$\Delta$ (ms)	$t$	$p$	$\Delta$ (ms)	$t$	$p$	$\Delta$ (ms)	$t$	$p$
1	5	–20	–1.68	0.144	–20	–1.68	0.145	–20	–1.68	0.144
	15	–11	–1.15	0.295	–13	–1.52	0.180	–11	–1.18	0.281
	25	0	0.04	0.971	–13	–1.99	0.094	–2	–0.27	0.800
	35	8	0.82	0.443	–8	–1.43	0.204	5	0.58	0.583
	45	22	1.30	0.242	–4	–0.36	0.732	14	0.92	0.391
	55	45	2.30	0.061	2	0.11	0.915	33	1.85	0.114
	65	74	2.65	0.038	–6	–0.21	0.844	43	1.79	0.123
	75	111	4.26	0.005	–33	–1.05	0.333	50	2.62	<b>0.039</b>
	85	203	4.11	0.006	–118	–1.84	0.116	73	1.77	0.128
95	337	4.41	0.004	–72	–0.91	0.400	74	0.94	0.384	
2	5	13	0.77	0.472	11	0.65	0.543	13	0.77	0.473
	15	14	0.92	0.395	4	0.24	0.822	13	0.86	0.422
	25	29	2.01	0.091	9	0.52	0.622	26	1.71	0.138
	35	39	2.20	0.070	–13	–1.26	0.255	33	1.99	0.093
	45	46	3.98	0.007	–46	–3.05	0.022	31	3.49	<b>0.013</b>
	55	60	4.08	0.007	–69	–2.54	0.044	31	3.01	<b>0.024</b>
	65	77	5.08	0.002	–83	–2.92	0.027	29	2.92	<b>0.027</b>
	75	110	5.92	0.001	–112	–3.39	0.015	35	4.82	<b>*0.003</b>
	85	179	6.16	0.001	–120	–2.39	0.054	64	3.47	<b>0.013</b>
95	293	4.89	0.003	–131	–2.09	0.082	54	1.37	0.218	

Negative  $t$ -values for the Miller and positive  $t$ -values for Grice bound comparisons indicate violations of the respective bounds. Significant model violations in bold; the asterisk marks a significant violation after Holm-Bonferroni correction for multiple testing.

The data were then compared to a model predicting stochastic independence, equivalently to the UCIP model employed for RTs (Stevenson et al., 2014). Bimodal accuracy was lower than predicted by the model in both Experiment 1 [ $\Delta = -9.0\%$ -points,  $t_{(6)} = -5.97$ ,  $p < 0.001$ ; **Figure 4C**] and Experiment 2 [ $\Delta = -10.6\%$ -points,  $t_{(6)} = -4.99$ ,  $p < 0.01$ ; **Figure 4D**]. In line with the RT data, the accuracy data indicate that the visual and olfactory channels are stochastically dependent.

### 3. Discussion

The present study found a bimodal response facilitation for perceptually, but not for physically aligned bimodal visual-olfactory stimuli. The facilitation could be accounted for by race models assuming probability summation across positively dependent processing channels. Thus, the results yielded no proof of coactivation.

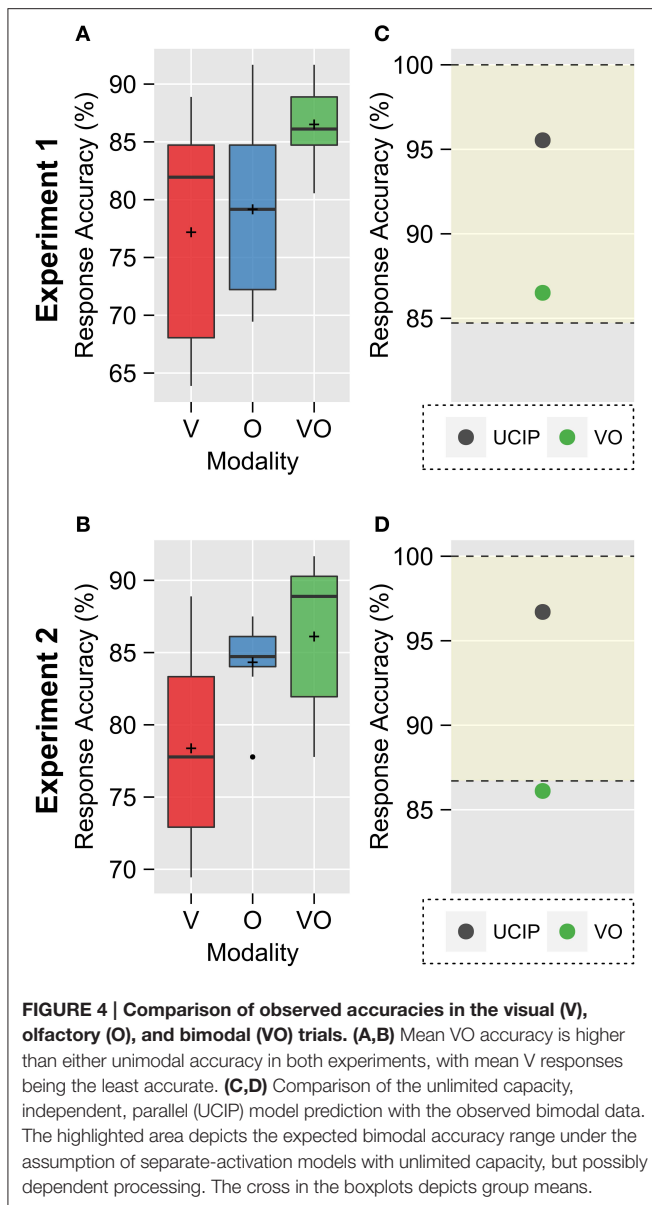
The observation of a significant bimodal visual-olfactory response speedup indicated by an RSE for perceptually, but not for physically aligned unimodal constituents suggests that temporal proximity subserves visual-olfactory response facilitation. Increasing temporal parity amplifies multisensory interactions in other sensory modalities, e.g., for visual-tactile (Forster et al., 2002) or visual-auditory (Lovelace et al., 2003) stimuli albeit the temporal binding window, i.e., the range of inter-stimulus intervals over which multisensory stimuli are

integrated, is not universal. While multisensory binding windows as large as several hundred milliseconds exist for example in audio-visual speech perception (see e.g., van Wassenhove et al., 2007), the effects of stimulus timing on visual-olfactory perception are unknown. To our knowledge, this is the first study demonstrating that perceptual, rather than physical, simultaneity is vital to elicit an RSE for bimodal visual-olfactory objects.

However, the response speedup could be the result of statistical facilitation alone and is not necessarily proof of neural integration processes. Therefore, we examined whether the present data could be explained by race models, or if we could find evidence for coactivation.

Response time distributions never significantly exceeded the *Miller bound*. We can therefore exclude coactivation as a possible explanation of the observed RSEs. Further we can exclude strictly limited-capacity processing over an extended period of time because the *Grice bound* was never violated (Townsend and Wenger, 2004). Taken together, the observed bimodal response times are consistent with separate-activation models with a first-terminating time rule and unlimited-capacity processing, i.e., race models (Miller, 1982; Grice et al., 1984a; Colonius, 1990).

Classically, it has been shown that violations of the *Miller bound* are more easily produced in go/no-go tasks due to the absence of “response competition” (Grice and Canham, 1990; Grice and Reed, 1992). Yet, race models can successfully be rejected in choice response time studies as well (see e.g., Miller, 1982; Hecht et al., 2008; Girard et al., 2011).



No change in response accuracy was observed in the bimodal conditions, compared to unimodal stimulation. This finding is in contrast to previous reports of improved accuracy for multisensory stimuli. A possible reason for this discrepancy might be that olfaction and vision do not integrate in the same way as other senses. However, it is also possible that we were not able to observe improved accuracy simply for statistical reasons due to the low number of trials (owing to the long inter-trial intervals, ITIs, necessary for olfactory stimuli) and high inter-subject variability.

Comparison of the observed bimodal response time distributions to a more restrictive race model assuming stochastic independence of channel processing speed (UCIP model) revealed significantly slower responses than predicted in both experiments, suggesting positively dependent channel

processing speeds between the visual and olfactory channels (Grice et al., 1984a). Although only the deviation in one quantile in Experiment 2 was significant after correction for multiple testing, the additional finding of lower bimodal response accuracies than predicted further corroborates the assumption of a possibly positive stochastic dependence of visual and olfactory processing.

In contrast, the bimodal combination of odor and taste stimuli yielded faster responses than predicted by a UCIP model in a recent study (Veldhuizen et al., 2010). Notably, odor and taste perception are closely intertwined; evidence exists for direct and indirect anatomical connections between the primary gustatory and olfactory cortices (Rolls and Baylis, 1994; Shepherd, 2006) as well as for convergence areas responding to both smell and taste, for example in the orbitofrontal cortex (OFC) (O'Doherty et al., 2001; de Araujo et al., 2003; Small and Green, 2012), the anterior insula, and frontal and parietal opercula (Small et al., 1999; Cerf-Ducastel and Murphy, 2001; Poellinger et al., 2001). Perceptually, the combined odor-taste experience typically exceeds the sum of the two chemosensory modalities, being perceived as more *Gestalt*-like, intense and rewarding, and yields superadditive activation in the frontal operculum (Seubert et al., 2015). Although no monosynaptic connection between the primary visual and olfactory cortices has been found, the perirhinal cortex is a prime candidate as a processing hub between the visual and olfactory modalities due to its numerous reciprocal connections, particularly with the inferior temporal cortex. The inferior temporal cortex is involved in object perception (Grill-Spector and Weiner, 2014) and associations of sensory representations, and a subdivision, the rhinal cortex, has been proven critical for the association of flavor with visual food objects in monkeys (Parker and Gaffan, 1998).

### 3.1. Conclusion

The present data are consistent with models of parallel processing with unlimited-capacity and positive dependence between the visual and olfactory channels. Notably, these models do not refute the possibility of coactive processing. Although odor perception is highly ambiguous and susceptible to other sensory information (Herz and von Clef, 2001), the olfactory stimuli may in fact have contributed to the bimodal object identification by generating further perceptual evidence, allowing an internal decision criterion to be reached faster. This assumption is supported by the observation of positive channel dependence, indicating that the identification of the visual and olfactory constituents in bimodal trials co-occurs. The objects used in the present study carried a semantic meaning, which had to be decoded before mapping it to the appropriate response button. Semantic representations emerge only at later stages in the perception process (Olofsson, 2014). Further, no direct connections between the visual and olfactory cortices have been discovered yet, questioning the plausibility of early bimodal visual-olfactory interactions. Future studies will have to show whether the present findings are transferable to other stimulus objects, SOAs, and experimental tasks.

## 4. Materials and Methods

### 4.1. Participants

Eight participants completed the study; one participant was excluded because his accuracy was far below chance level for the unimodal olfactory lemon stimulus in both experiments (mean accuracy was approx. 33 %); data of seven participants (4 female; age in years:  $29.9 \pm 2.4$  SD, range: 26–32; all right-handed) are reported here. Participants were recruited from the German Institute of Human Nutrition and local universities; they gave written informed consent and received compensatory payment. They reported no neurological disorders or chronic diseases, in particular no smell impairment, and normal or corrected-to-normal vision. The study was conducted in accordance with the requirements of the revised Declaration of Helsinki and had been approved by the ethics committee of the German Society for Psychology (DPGs).

### 4.2. Stimuli

#### 4.2.1. Visual Stimuli

Six images (three different images of bananas and lemons, respectively) with different complexities were selected from the Food-pics database (no. 276, 282, 341, 379, and 415; Blechert et al., 2014) or purchased online. Images displayed a food object centered on a white background. They were resized to  $1024 \times 1024$  pixels and converted to grayscale. A Gaussian blur (order 0,  $\sigma = 3$ ) was applied to remove sharp edges. The fast Fourier transform (FFT) of all images was calculated and the phase space was randomly scrambled. The inverse FFT of the image with the scrambled phase yielded blurry images of the food objects with superimposed cloud-like noise patterns. Noise-only images were also derived for every object using the same method, yielding  $2 \times 3$  target and  $2 \times 3$  noise-only stimuli in total. The spatial frequency of those noise patterns was similar to the spatial frequency of the original object. Images were presented on a TFT monitor with a resolution of  $1680 \times 1050$  pixels. The refresh rate was set to 60 Hz. Participants viewed the images at an eye distance of approx. 60 cm, corresponding to an object size of approximately  $12^\circ$  of visual angle, embedded in visual noise of approximately  $27^\circ$  of visual angle.

#### 4.2.2. Olfactory Stimuli

Odorants were 10 mL aliquots of isoamyl acetate (banana; Sigma-Aldrich Chemie GmbH, Steinheim, Germany, CAS 123-92-2) and lemon oil (lemon; same vendor, CAS 8008-56-8) diluted with mineral oil (Acros Organics, Geel, Belgium, CAS 8042-47-5) to produce solutions of 0.1 % v/v concentration. The solvent, pure mineral oil, served as neutral control. The odors were congruent to the visual objects banana and lemon; odor intensity was chosen to yield identifiable, yet weak stimuli based on a pilot study ( $n = 7$ ). Odorants were presented birhinally using a custom-built 16-channel air-dilution olfactometer (Lundström et al., 2010). Teflon tubes with an inner diameter of  $1/16''$  delivered the odorous air via custom-made anatomically shaped nose pieces into the participants' nostrils. A constant flow of clean air (approximately  $0.5 \text{ L min}^{-1}$ ) was present at all times to rinse the tubing system and the nose. Stimuli were delivered with a flow

rate of approximately  $3.0 \text{ L min}^{-1}$ , totaling to a flow of about  $3.5 \text{ L min}^{-1}$  during stimulation. Stimulus timing was measured using a photo-ionization detector (PID; 200B miniPID, Aurora Scientific Inc., Aurora/ON, Canada) and defined as the time point 254 ms after sending the trigger to the olfactometer. To ensure a constant odor concentration and to reduce depletion of head space in the odor jars in the course of the experiment, one of three identical odor jars was used in sequential order from trial to trial.

### 4.3. Procedure

Participants completed two experimental sessions on separate days. In the first session, a visual identification threshold assessment was conducted, followed by a choice response time (CRT) Experiment in which bimodal stimulus components were presented physically simultaneous. A second CRT experiment with perceptually aligned bimodal stimuli was conducted during the next session. The experiments were carried out in a sound-attenuated experimental booth. Participants were seated centered in front of the screen. Responses were collected using a button box (Serial Response Box, Psychology Software Tools, Sharpsburg/PA, USA) connected to a USB port of the stimulation computer via a serial-to-USB adapter. Timing accuracy was verified to be better than 2 ms. In-ear headphones delivered Brownian noise during the CRT experiments at a volume chosen such that the change in air flow at stimulation on- and offset was inaudible. The stimulation was controlled using PsychoPy 1.79.01 (Peirce, 2009) running on a personal computer.

#### 4.3.1. Visual Threshold Estimation

We adjusted the strength of the noise so that objects could be perceived approximately on every second trial using a QUEST staircase procedure (Watson and Pelli, 1983). The Experiment started with a short practice block, in which all target and noise-only images were presented once. Then, images of objects + noise were presented interleaved with noise-only images (equal proportions) for 900 ms with a randomly varied ITI between 1.5 and 2.0 s during which a white screen was presented. Participants indicated by button press the detection of an object within the noise. The staircase adjusted the strength of the noise to yield a performance level of 50 % correct object detection when stimuli were present (false alarms on noise-only trials were very rare, ranging from 0 to approx. 3 %, with a grand mean of 1.3 %). Separate staircases were run for each of the six different object images. Overall, the threshold procedure entailed 240 trials, 20 repetitions of each of the six images and their respective noise-only images ( $2 \times 6 \text{ images} \times 20 \text{ repetitions}$ ). Participants were allowed a short break; the procedure lasted about 12 m. Note that stimuli yielding 50 % accuracy in this detection task are expected to yield approximately 75 % performance in the 2-AFC task as used in the main experiment.

#### 4.3.2. Bimodal CRT Experiments

During the CRT experiments, participants were to identify the presented object (banana or lemon) as quickly as possible (while avoiding anticipatory responses) by pressing either of two buttons on the button box. Stimuli were either unimodal visual (V) objects presented at individual 50% identification threshold,



unimodal olfactory (O) objects, or bimodal visual-olfactory (VO) objects. V stimuli were always paired with the neutral control odorant. O stimuli were paired with a randomly assigned noise-only image derived from a visual stimulus of the same object. OV stimuli consisted of the combined presentation of congruent V and O stimuli.

VO stimulus pairs were presented simultaneously in Experiment 1. In Experiment 2, bimodal stimulus timing was adjusted to achieve perceptual simultaneity by introducing an SOA equal to the difference of unimodal median RTs individually for each object and participant. The mean SOAs were  $330 \pm 295$  ms SD for banana, and  $395 \pm 205$  ms SD for lemon. Note that all SOAs were positive, i.e., delaying visual presentation, except for banana in one participant, where the odor had to be presented 182 ms prior to the visual stimulus to achieve perceptual simultaneity. To ensure context-invariance in Experiment 2, we also adjusted the timing of the unimodal stimulus presentations. Specifically, if the estimated SOA indicated a delayed presentation of the visual constituent in bimodal conditions, we also delayed the visual stimulation in the unimodal conditions for the same amount of time (meaning the fixation cross was visible for a longer duration before the stimulus appeared; note that this was also true for the unimodal olfactory stimulation, where the visual stimulus was noise-only). The stimulus timing is illustrated in **Figure 1**.

Each trial started with a fixation cross centered on the screen, which informed participants to prepare and to slowly inhale. At the same time, the air flow through the neutral jar was initiated to remove the tactile cue from the later stimulus presentation. After a random period of 1–2 s, a stimulus (O, V, or VO) was presented for 900 ms. After stimulation, the neutral control odorant was presented for 4.1 s to remove residual odor molecules. The ITI was randomly varied between 20 and 21 s.

The experiments started after a short practice block in which each stimulus combination was presented once. Each Experiment consisted of six blocks during which all stimulus combinations were presented twice and in pseudo-random order, totaling to 216 stimuli (6 blocks  $\times$  2 repetitions  $\times$  (6 V + 6 O + 6 VO)), and lasted 95–120 min. Participants were allowed self-paced breaks in the middle of each block and between blocks.

RT measurement started with the onset of the image in V trials and the physical onset of the odorant as determined by PID measurements in O trials. In bimodal trials, RT measurement started with the physical onset of the stimuli (Experiment 1, physically simultaneous presentation), or with the onset of the earlier stimulus (Experiment 2, perceptually simultaneous presentation).

#### 4.4. Data Analysis

Only trials with positive and correct identification responses were analyzed. RT medians and standard deviations (SDs) of the aggregated data were calculated for each of the six conditions (O, V, VO for banana and lemon objects, respectively). All trials with a reaction time deviating more than two SDs from the median were discarded as outliers. In Experiment 1 and 2, 6.0 % and 5.5 % of the trials were removed, respectively.

A short summary of the analyses will be given in the next paragraph, followed by a detailed method description in the remaining section.

Faster responses to bimodal, compared to unimodal, stimuli indicate an RSE. Therefore, we first compared bimodal to unimodal RTs by calculating the difference between the bimodal and the faster of the two unimodal RTs (visual or olfactory). Because this global RSE is relatively insensitive to effects that are not present across the whole response time range, we next estimated cumulative distribution functions (CDFs) from the RTs. Analyses based on these CDFs can take into account the whole RT distribution. We evaluated the CDFs at 10 quantiles. Since an observed response speedup can be caused by statistical facilitation alone, in a next step we calculated theoretical model boundaries based on the unimodal CDFs under the assumption of parallel processing of the visual and olfactory channels (race model), that is the data range that could be explained by statistical facilitation. Any observation exceeding these limits would support the hypothesis of true integrative processing. To examine whether the channels operated in a stochastically independent manner, we additionally compared our data to a very specific race model assuming stochastic independence of the channel processing speeds (UCIP model). A very similar approach was chosen in the analysis of the accuracy data, although it was naturally based on mean accuracies and not single-trial responses, i.e., no equivalent of a CDF could be estimated.

##### 4.4.1. Response Times

The RT distributions were heavily positively skewed; we therefore used the median as measure of central tendency. This measure is not without criticism (cf. Miller, 1988), but alternatives like the commonly applied log-transformations are not universally applicable approaches either (Feng et al., 2014).

Response times to unimodal V and O stimuli and their bimodal VO combination are defined as non-negative random vectors  $RT_V$ ,  $RT_O$  and  $RT_{VO}$ . Their respective expected values shall be labeled  $E(RT_V)$ ,  $E(RT_O)$  and  $E(RT_{VO})$ , and their distribution functions as  $F(RT_V)$ ,  $F(RT_O)$ , and  $F(RT_{VO})$ . An RSE can be observed if

$$E(RT_{VO}) < \min[E(RT_V), E(RT_O)], \quad (1)$$

i.e., if mean RTs for bimodal VO stimuli are faster than for either unimodal component.

We calculated the difference between the medians of the fastest unimodal and the bimodal RTs, i.e.,  $\min[E(RT_V), E(RT_O)] - E(RT_{VO})$ . Positive values indicate a bimodal speedup, i.e., a facilitation in processing of bimodal as compared to unimodal stimuli. Note that RSEs were calculated separately for each object (banana and lemon) before collapsing and submission to one-sample *t*-tests against zero to identify bimodal facilitation.

To quantify the effect of in perceptually aligning the unimodal constituents of bimodal trials in Experiment 2, we compared the median RTs of the unimodal V and O conditions (collapsed across objects) using paired *t*-tests for each experiment.

Next, we tested whether the RT distributions fit probability summation models. In bimodal trials, only the marginal distribution  $F(RT_{VO})$ , but not the distributions of the unimodal constituents  $F(RT_V)$  and  $F(RT_O)$  can be observed.

Probability summation models critically rely on the assumption of *context-invariance* (Colonius, 1990), which states that the processing speed of a channel is identical in unimodal and bimodal stimulations, that is additional work load on one channel does not influence processing speed in another channel, suggesting *unlimited capacity*.

The unlimited capacity, independent, parallel (UCIP) model makes the additional assumption that the processing speeds of individual channels are uncorrelated and hence *stochastically independent* (Raab, 1962; Meijers and Eijkman, 1977). According to a UCIP model, the cumulative distribution function for the bimodal stimulation is:

$$F(RT_{VO})(t) = F(RT_V)(t) + F(RT_O)(t) - F(RT_V)(t) \times F(RT_O)(t). \quad (2)$$

The last term is always equal to or greater than zero, i.e.,  $F(RT_V)(t) \times F(RT_O)(t) \geq 0$ .

Miller (1982) discarded the assumption of stochastic independence and instead assumed a maximally negative dependence between the channel processing speeds (Colonius, 1990). This allowed him derive an upper bound for the maximum achievable performance gain under *any* parallel processing model called *Miller bound* or *race model inequality* (RMI), commonly expressed as:

$$F(RT_{VO})(t) \leq F(RT_V)(t) + F(RT_O)(t) \quad (3)$$

All parallel processing models have to satisfy inequality (3). If the inequality is violated, the assumption of parallel processing must be dropped, i.e., all race models are ruled out immediately, and the results can only be accounted for by what Miller called *coactivation models* (Miller, 1982)<sup>1</sup>. Similarly, a lower performance bound was defined by Grice et al. (1984b), implying perfect positive dependence (Colonius, 1990) between the channels' processing speeds:

$$F(RT_{VO})(t) \geq \max[F(RT_V)(t), F(RT_O)(t)] \quad (4)$$

That is, performance in the bimodal conditions should be equal to or faster than in the fastest unimodal condition.

In the case of asynchronous stimulation, i.e., by delaying the presentation of the visual stimulus by the time  $\tau$ , Equations (1), (2), respectively, become (Miller, 1986):

$$E(RT_{VO(\tau)}) < \min[E(RT_V + \tau), E(RT_O)], \text{ and} \quad (5)$$

$$F(RT_{VO(\tau)})(t) = F(RT_V)(t - \tau) + F(RT_O)(t) - F(RT_V)(t - \tau) \times F(RT_O)(t). \quad (6)$$

Note that the visual RT distribution  $F(RT_V)(t - \tau)$  is shifted to the right, which is the correct adjustment for the SOA. The adjusted

<sup>1</sup>However, it should be noted that the reverse is not true: Showing that the observations can be described using a parallel processing model does not necessarily exclude coactivation models.

*Miller and Grice bounds* from Equations (3), (4) can then be expressed as:

$$F(RT_{VO(\tau)})(t) \leq F(RT_V)(t - \tau) + F(RT_O)(t), \text{ and} \quad (7)$$

$$F(RT_{VO(\tau)})(t) \geq \max[F(RT_V)(t - \tau), F(RT_O)(t)]. \quad (8)$$

We estimated empirical cumulative distribution functions (CDFs) of the RTs using a Python implementation of the algorithm suggested by Ulrich et al. (2007). The CDFs predicted by the UCIP model denoted in Equation (6), as well as the theoretical race model boundaries from Equations (7), (8) were calculated based on the unimodal CDFs, resulting in six CDFs per participant (unimodal O and V, bimodal VO, UCIP model, upper and lower bound). All CDFs were then evaluated at ten evenly spaced quantile points (0.05, 0.15, . . . , 0.95), which were subsequently collapsed across both objects. The resulting values were submitted to separate paired *t*-tests for every quantile to test for deviations from the model predictions.

#### 4.4.2. Accuracy

Similar to Equation (1), an RSE in accuracy can be observed if

$$E(ACC_{VO}) > \max[E(ACC_V), E(ACC_O)], \quad (9)$$

i.e., if mean accuracy for bimodal VO stimuli is higher than for the most accurate of the unimodal components.

We calculated the difference between the means of the most accurate unimodal and the bimodal responses, i.e.,  $E(ACC_{VO}) - \max[E(ACC_V), E(ACC_O)]$ . Positive values indicate a bimodal accuracy enhancement. Note that RSEs were calculated separately for each object before collapsing to one-sample *t*-tests against zero.

Following the assumption of the UCIP model, Equation (2) can be applied to accuracy data and becomes (Stevenson et al., 2014):

$$p(ACC_{VO}) = p(ACC_V) + p(ACC_O) - p(RT_V) \times p(ACC_O). \quad (10)$$

Equivalents of the *Miller and Grice bounds* for bimodal accuracy were proposed by Colonius (2015). Formulas (3) and (4), respectively, then become

$$p(ACC_{VO}) \leq p(ACC_V) + p(ACC_O), \text{ and} \quad (11)$$

$$p(ACC_{VO}) \geq \max[p(ACC_V), p(ACC_O)]. \quad (12)$$

The model predictions and boundaries were calculated for each Experiment and object separately. The results were then collapsed across objects. The resulting values were then submitted to paired *t*-tests to test for deviations from the model predictions.

## Author Contributions

RH, NB, and KO designed the experiments; RH acquired the data and performed the analyses with input from KO and NB. RH and KO wrote the paper.

## Acknowledgments

The authors thank Andrea Katschak for help with data collection and stimulus preparation. Data exploration and aggregation was carried out with IPython (Pérez and Granger, 2007) and pandas (McKinney, 2010). Statistical analyses were conducted with R

(Wickham, 2011; R Core Team, 2014). Plots were generated by ggplot2 (Wickham, 2009) and arranged with Inkscape (<https://inkscape.org/>). The GIMP (<http://www.gimp.org/>) was used for image preparation; the Gaussian blur was generated with SciPy (<http://scipy.org/>). FFT transforms were calculated using FFTW3 (Frigo and Johnson, 2005).

## References

- Blechert, J., Meule, A., Busch, N. A., and Ohla, K. (2014). Food-pics: an image database for experimental research on eating and appetite. *Front. Psychol.* 5:617. doi: 10.3389/fpsyg.2014.00617
- Cain, W. S. (1979). To know with the nose: keys to odor identification. *Science* 203, 467–470. doi: 10.1126/science.760202
- Cerf-Ducastel, B., and Murphy, C. (2001). fMRI activation in response to odorants orally delivered in aqueous solutions. *Chem. Sens.* 26, 625–637. doi: 10.1093/chemse/26.6.625
- Chen, K., Zhou, B., Chen, S., He, S., and Zhou, W. (2013). Olfaction spontaneously highlights visual saliency map. *Proc. R. Soc. Lond. B Biol. Sci.* 280:20131729. doi: 10.1098/rspb.2013.1729
- Colonius, H. (1990). Possibly dependent probability summation of reaction time. *J. Math. Psychol.* 34, 253–275. doi: 10.1016/0022-2496(90)90032-5
- Colonius, H. (2015). Behavioral measures of multisensory integration: bounds on bimodal detection probability. *Brain Topogr.* 28, 1–4. doi: 10.1007/s10548-014-0421-3
- Colonius, H., and Vorberg, D. (1994). Distribution inequalities for parallel models with unlimited capacity. *J. Math. Psychol.* 38, 35–58. doi: 10.1006/jmps.1994.1002
- Davis, R. G. (1981). The role of nonolfactory context cues in odor identification. *Percept. Psychophys.* 30, 83–89. doi: 10.3758/BF03206139
- de Araujo, I. E., Rolls, E. T., Kringelbach, M. L., McGlone, F., and Phillips, N. (2003). Taste-olfactory convergence, and the representation of the pleasantness of flavour, in the human brain. *Eur. J. Neurosci.* 18, 2059–2068. doi: 10.1046/j.1460-9568.2003.02915.x
- Diederich, A., and Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Percept. Psychophys.* 66, 1388–1404. doi: 10.3758/BF03195006
- Feng, C., Wand, H., Lu, N., Chen, T., He, H., Lu, Y., et al. (2014). Log-transformation and its implications for data analysis. *Shanghai Arch. Psychiatry.* 26, 105–109. doi: 10.3969/j.issn.1002-0829.2014.02.009
- Forster, B., Cavina-Pratesi, C., Aglioti, S. M., and Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Exp. Brain Res.* 143, 480–487. doi: 10.1007/s00221-002-1017-9
- Frigo, M., and Johnson, S. (2005). The design and implementation of FFTW3. *Proc. IEEE* 93, 216–231. doi: 10.1109/JPROC.2004.840301
- Gielen, S. C., Schmidt, R. A., and Van den Heuvel, P. J. (1983). On the nature of intersensory facilitation of reaction time. *Percept. Psychophys.* 34, 161–168. doi: 10.3758/BF03211343
- Girard, S., Collignon, O., and Lepore, F. (2011). Multisensory gain within and across hemispaces in simple and choice reaction time paradigms. *Exp. Brain Res.* 214, 1–8. doi: 10.1007/s00221-010-2515-9
- Gondan, M. (2009). Testing the race model inequality in redundant stimuli with variable onset asynchrony. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 575. doi: 10.1037/a0013620
- Gottfried, J. A., and Dolan, R. J. (2003). The nose smells what the eye sees: crossmodal visual facilitation of human olfactory perception. *Neuron* 39, 375–386. doi: 10.1016/S0896-6273(03)00392-1
- Grice, G. R., and Reed, J. M. (1992). What makes targets redundant? *Percept. Psychophys.* 51, 437–442. doi: 10.3758/BF03211639
- Grice, G. R., and Canham, L. (1990). Redundancy phenomena are affected by response requirements. *Percept. Psychophys.* 48, 209–213. doi: 10.3758/BF03211520
- Grice, G. R., Canham, L., and Boroughs, J. M. (1984a). Combination rule for redundant information in reaction time tasks with divided attention. *Percept. Psychophys.* 35, 451–463. doi: 10.3758/BF03203922
- Grice, G. R., Canham, L., and Gwynne, J. W. (1984b). Absence of a redundant-signals effect in a reaction time task with divided attention. *Percept. Psychophys.* 36, 565–570. doi: 10.3758/BF03207517
- Grill-Spector, K., and Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* 15, 536–548. doi: 10.1038/nrn3747
- Hecht, D., Reiner, M., and Karni, A. (2008). Multisensory enhancement: gains in choice and in simple response times. *Exp. Brain Res.* 189, 133–143. doi: 10.1007/s00221-008-1410-0
- Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *J. Exp. Psychol.* 63, 289. doi: 10.1037/h0039516
- Herz, R. S., and von Clef, J. (2001). The influence of verbal labeling on the perception of odors: evidence for olfactory illusions? *Perception* 30, 381–391. doi: 10.1068/p3179
- Hilgard, E. R. (1933). Reinforcement and inhibition of eyelid reflexes. *J. General Psychol.* 8, 85–113. doi: 10.1080/00221309.1933.9713177
- Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behaviour: multisensory integration versus normal variability. *Neuropsychologia* 45, 3340–3345. doi: 10.1016/j.neuropsychologia.2007.05.025
- Laurienti, P. J., Perrault, T. J., Stanford, T. R., Wallace, M. T., and Stein, B. E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Exp. Brain Res.* 166, 289–297. doi: 10.1007/s00221-005-2370-2
- Lovelace, C. T., Stein, B. E., and Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Brain Res. Cogn. Brain Res.* 17, 447–453. doi: 10.1016/S0926-6410(03)00160-5
- Lundström, J., Gordon, A. R., Alden, E. C., Boesveldt, S., and Albrecht, J. (2010). Methods for building an inexpensive computer-controlled olfactometer for temporally precise experiments. *Int. J. Psychol.* 78, 179–189. doi: 10.1016/j.ijpsycho.2010.07.007
- McKinney, W. (2010). “Data structures for statistical computing in Python,” in *Proceedings of the 9th Python in Science Conference*, eds S. van der Walt and J. Millman (Austin, TX), 51–56.
- Meijers, L., and Eijkman, E. (1977). Distributions of simple rt with single and double stimuli. *Percept. Psychophys.* 22, 41–48. doi: 10.3758/BF03206078
- Meredith, M. A., and Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J. Neurophysiol.* 56, 640–662.
- Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cogn. Psychol.* 14, 247–279. doi: 10.1016/0010-0285(82)90010-X
- Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Percept. Psychophys.* 40, 331–343. doi: 10.3758/BF03203025
- Miller, J. (1988). A warning about median reaction time. *J. Exp. Psychol. Hum. Percept. Perform.* 14, 539–543. doi: 10.1037/0096-1523.14.3.539
- O’Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., and Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat. Neurosci.* 4, 95–102. doi: 10.1038/82959
- Olofsson, J. K. (2014). Time to smell: a cascade model of human olfactory perception based on response-time (rt) measurement. *Front. Psychol.* 5:33. doi: 10.3389/fpsyg.2014.00033
- Parker, A., and Gaffan, D. (1998). Memory after frontal/temporal disconnection in monkeys: conditional and non-conditional tasks, unilateral and bilateral frontal lesions. *Neuropsychologia* 36, 259–271. doi: 10.1016/S0028-3932(97)00112-7

- Peirce, J. W. (2009). Generating stimuli for neuroscience using psychopy. *Front. Neuroinformatics* 2:10. doi: 10.3389/neuro.11.010.2008
- Pérez, F., and Granger, B. E. (2007). IPython: a system for interactive scientific computing. *Comput. Sci. Eng.* 9, 21–29. doi: 10.1109/MCSE.2007.53
- Poellinger, A., Thomas, R., Lio, P., Lee, A., Makris, N., Rosen, B. R., et al. (2001). Activation and habituation in olfaction fMRI study. *Neuroimage* 13, 547–560. doi: 10.1006/nimg.2000.0713
- R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Trans. N.Y. Acad. Sci.* 24, 574–590. doi: 10.1111/j.2164-0947.1962.tb01433.x
- Robinson, A. K., Mattingly, J. B., and Reinhard, J. (2013). Odors enhance the saliency of matching images during the attentional blink. *Front. Integr. Neurosci.* 7:77. doi: 10.3389/fnint.2013.00077
- Rolls, E. T., and Baylis, L. L. (1994). Gustatory, olfactory, and visual convergence within the primate orbitofrontal cortex. *J. Neurosci.* 14, 5437–5452.
- Seo, H.-S., Arshamian, A., Schemmer, K., Scheer, I., Sander, T., Ritter, G., et al. (2010). Cross-modal integration between odors and abstract symbols. *Neurosci. Lett.* 478, 175–178. doi: 10.1016/j.neulet.2010.05.011
- Seubert, J., Ohla, K., Yokomukai, Y., Kellermann, T., and Lundström, J. N. (2015). Superadditive opercular activation to food flavor is mediated by enhanced temporal and limbic coupling. *Hum. Brain Mapp.* 36, 1662–1676. doi: 10.1002/hbm.22728
- Shepherd, G. M. (2006). Smell images and the flavour system in the human brain. *Nature* 444, 316–321. doi: 10.1038/nature05405
- Small, D. M., and Green, B. (2012). “A Proposed Model of a Flavor Modality,” in *The Neural Bases of Multisensory Processes*, eds M. M. Murray and M. T. Wallace (Boca Raton, FL: CRC Press), 717–738. doi: 10.1201/9781439812174-46
- Small, D. M., Zald, D. H., Jones-Gotman, M., Zatorre, R. J., Pardo, J. V., Frey, S., et al. (1999). Human cortical gustatory areas: a review of functional neuroimaging data. *Neuroreport* 10, 7–13. doi: 10.1097/00001756-199901180-00002
- Stein, B., and Meredith, M. (1993). *The Merging of the Senses*. Cambridge, MA: MIT Press.
- Stein, B. E., and Meredith, M. A. (1990). Multisensory integration. neural and behavioral solutions for dealing with stimuli from different sensory modalities. *Ann. N.Y. Acad. Sci.* 608, 51–65; discussion: 65–70. doi: 10.1111/j.1749-6632.1990.tb48891.x
- Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., et al. (2014). Identifying and quantifying multisensory integration: a tutorial review. *Brain Topogr.* 27, 707–730. doi: 10.1007/s10548-014-0365-7
- Townsend, J. T., and Nozawa, G. (1995). Spatio-temporal properties of elementary perception: an investigation of parallel, serial, and coactive theories. *J. Math. Psychol.* 39, 321–359. doi: 10.1006/jmps.1995.1033
- Townsend, J. T., and Wenger, M. J. (2004). A theory of interactive parallel processing: new capacity measures and predictions for a response time inequality series. *Psychol. Rev.* 111:1003. doi: 10.1037/0033-295X.111.4.1003
- Ulrich, R., Miller, J., and Schröter, H. (2007). Testing the race model inequality: an algorithm and computer programs. *Behav. Res. Methods* 39, 291–302. doi: 10.3758/BF03193160
- van Wassenhove, V., Grant, K. W., and Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* 45, 598–607. doi: 10.1016/j.neuropsychologia.2006.01.001
- Veldhuizen, M. G., Shepard, T. G., Wang, M.-F., and Marks, L. E. (2010). Coactivation of gustatory and olfactory signals in flavor perception. *Chem. Senses* 35, 121–133. doi: 10.1093/chemse/bjp089
- Watson, A. B., and Pelli, D. G. (1983). QUEST: A bayesian adaptive psychometric method. *Percept. Psychophys.* 33, 113–120. doi: 10.3758/BF03202828
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer.
- Wickham, H. (2011). The split-apply-combine strategy for data analysis. *J. Statist. Softw.* 40, 1–29. doi: 10.18637/jss.v040.i01
- Zellner, D. A., Bartoli, A. M., and Eckard, R. (1991). Influence of color on odor identification and liking ratings. *Am. J. Psychol.* 104, 547–561. doi: 10.2307/1422940
- Zhou, W., Jiang, Y., He, S., and Chen, D. (2010). Olfaction modulates visual perception in binocular rivalry. *Curr. Biol.* 20, 1356–1358. doi: 10.1016/j.cub.2010.05.059
- Zhou, W., Zhang, X., Chen, J., Wang, L., and Chen, D. (2012). Nostril-specific olfactory modulation of visual perception in binocular rivalry. *J. Neurosci.* 32, 17225–17229. doi: 10.1523/JNEUROSCI.2649-12.2012

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Höchenberger, Busch and Ohla. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Illusory visual motion stimulus elicits postural sway in migraine patients

Shu Imaizumi<sup>1,2</sup>, Motoyasu Honma<sup>3</sup>, Haruo Hibino<sup>1</sup> and Shinichi Koyama<sup>1,3,4\*</sup>

<sup>1</sup> Graduate School of Engineering, Chiba University, Chiba, Japan, <sup>2</sup> Japan Society for the Promotion of Science, Tokyo, Japan, <sup>3</sup> School of Medicine, Showa University, Tokyo, Japan, <sup>4</sup> School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore, Singapore

Although the perception of visual motion modulates postural control, it is unknown whether illusory visual motion elicits postural sway. The present study examined the effect of illusory motion on postural sway in patients with migraine, who tend to be sensitive to it. We measured postural sway for both migraine patients and controls while they viewed static visual stimuli with and without illusory motion. The participants' postural sway was measured when they closed their eyes either immediately after (Experiment 1), or 30 s after (Experiment 2), viewing the stimuli. The patients swayed more than the controls when they closed their eyes immediately after viewing the illusory motion (Experiment 1), and they swayed less than the controls when they closed their eyes 30 s after viewing it (Experiment 2). These results suggest that static visual stimuli with illusory motion can induce postural sway that may last for at least 30 s in patients with migraine.

## OPEN ACCESS

### Edited by:

Magda L. Dumitru,  
Macquarie University, Australia

### Reviewed by:

Thierry Lelard,  
Université de Picardie Jules Verne,  
France  
Arnold Jonathan Wilkins,  
University of Essex, UK

### \*Correspondence:

Shinichi Koyama,  
Graduate School of Engineering,  
Chiba University, 1-33 Yayoi, Inage,  
Chiba 263-8522, Japan  
skoyama@faculty.chiba-u.jp

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 02 February 2015

**Accepted:** 15 April 2015

**Published:** 28 April 2015

### Citation:

Imaizumi S, Honma M, Hibino H  
and Koyama S (2015) Illusory visual  
motion stimulus elicits postural sway  
in migraine patients.  
*Front. Psychol.* 6:542.  
doi: 10.3389/fpsyg.2015.00542

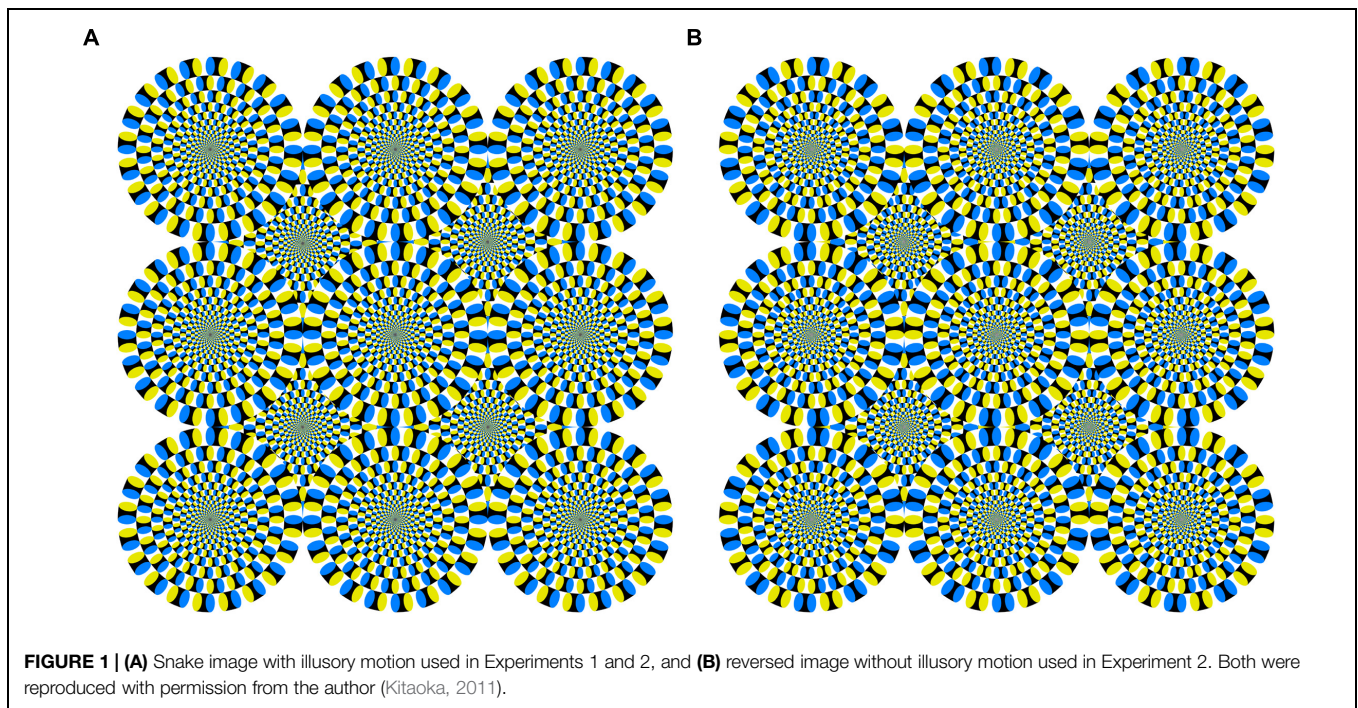
**Keywords:** migraine, vision, optical illusion, postural control, visuo-vestibular interaction, multisensory integration

## Introduction

Postural control is modulated not only by vestibular functioning (Birren, 1945) but also by visual stimulation. For example, visual input simulating forward or backward self-motion, such as expanding, or contracting optic flow, elicits postural sway in observers (Lee and Lishman, 1975; van Asten et al., 1988). This visually induced postural modulation occurs even in infants (Lee and Aronson, 1974). These and other recent studies (Guerraz and Bronstein, 2008; Meyer et al., 2013) suggested that postural sway was induced by the visual stimulus with motion energy (i.e., a physically moving stimulus).

However, human observers do not necessarily need motion energy to perceive motion in a visual stimulus. Illusory motion perception is one type of optical illusion in which observers perceive physically static images as moving. In the Fraser–Wilcox illusion (Fraser and Wilcox, 1979), a static figure consisting of repeating patterns with saw-tooth luminance profiles induces illusory motion. The Rotating Snakes (Kitaoka, 2003) is an optimized Fraser–Wilcox illusion that has patterns with stepwise luminance profiles, which induces stronger illusory motion (Kitaoka and Ashida, 2003; See **Figure 1A** for an example). One explanation for the Rotating Snakes is that each component of the stepwise luminance profiles in this figure elicits motion energy caused by differences in the latency of neural activity for each luminance component (Backus and Oruc, 2005; Conway et al., 2005). Recent studies have suggested that the neural basis for the illusory motion induced by Rotating Snakes is found in the human cortical pathway from primary visual cortex to the middle temporal area (Kuriki et al., 2008; Ashida et al., 2012).

**Abbreviations:** HMD, head-mounted display; RMANOVA, repeated measures analysis of variance.



The effects of illusory motion on human body movements are not well documented. One previous study reported that illusory expanding motion can induce the perception of body movement in physically stationary observers (i.e., forward vection; Seno et al., 2013). Another study showed that, following adaptation to a leftward or rightward 20-s random-pixel array motion, the aftereffect resulting from the static random pixels increased postural sway in the direction opposite to that of the adapted motion (Holten et al., 2014). The researchers argued that the neural motion signal itself influences postural control, even after moving stimulus observation. Indeed, physically moving visual stimulation has been demonstrated to activate the visual cortex's middle temporal area (Zeki et al., 1991; Morrone et al., 2000), which can also be activated by illusory motion (Kuriki et al., 2008; Ashida et al., 2012) and motion aftereffect (He et al., 1998). On the other hand, an optical flow stimulus that is congruent with self-motion can activate not only the middle temporal area (Slobounov et al., 2006), but also the cingulate sulcus visual area, which receives vestibular inputs (Smith et al., 2012) and represents self-motion (Wall and Smith, 2008; Fischer et al., 2012). Since physical motion perception shares common neural bases with illusory motion and motion aftereffect and is represented in the self-motion sensitive cortex, illusory motion may influence postural control, as well as physical motion (e.g., Lee and Lishman, 1975) and motion aftereffect (Holten et al., 2014). However, it remains unclear whether postural sway increases *during* the illusory motion inducing static visual stimulus observation.

Illusory motion and/or visual distortion in static geometrical stimuli (e.g., striped patterns) are more likely to be perceived by individuals with chronic migraine headaches than non-chronic headache sufferers (Wilkins et al., 1984; Marcus and Soso, 1989;

Huang et al., 2003; Imaizumi et al., 2011). This effect is perhaps caused by altered cortical processing in the primary visual cortex (Aurora et al., 1998; Huang et al., 2003) and middle temporal areas (Granziera et al., 2006). On the other hand, migraine patients are known to be susceptible to motion sickness (Cutrer and Baloh, 1992; Drummond, 2005; Marcus et al., 2005), which is caused by the conflict between visual and vestibular input (Reason and Brand, 1975; Yates et al., 1998). Especially in patients with migraine, motion sickness can be evoked solely by visual stimulation when it conflicts with vestibular signals. For instance, the stationary observation of horizontally moving vertical stripes can induce motion sickness more in patients than in normal controls (Drummond, 2002; Drummond and Granston, 2004). Thus, it can be assumed that patients with migraine, who are susceptible to visually induced motion sickness, might be more dependent on visual input when their posture is controlled. Although postural sway increases in both patients and normal controls when they close their eyes because of the lack of visual control (Travis, 1945; Edwards, 1946; Honma et al., 2012), a previous study demonstrated that postural sway increases by a greater amount in patients with migraine while they have their eyes closed (Ishizaki et al., 2002). Taken together, we hypothesize that the patients' postural control should be more influenced by visual stimuli than that of normal individuals, especially when the stimuli are capable of inducing illusory motion.

The present study aimed to examine whether illusory motion can influence postural sway and whether there are any distinguishing characteristics in patients with migraine in terms of postural control. We attempted to measure the postural sway of both patients and normal controls during observations of static visual stimuli with and without illusory motion with a stabilometer to track the displacement of centers of gravity.

## Experiment 1

We measured postural sway during migraine patients' and normal controls' viewing of static stimulus with and without illusory motion (Rotating Snakes and a gray plane, respectively).

## Materials and Methods

### Participants

This experiment included 11 patients with migraine (six female; mean age  $22.18 \pm 0.30$  years) and nine controls without chronic headaches (two female; mean age  $22.22 \pm 0.40$  years). One of the patients had visual aura symptoms. We separated the patients from the controls and determined the presence of visual aura using a questionnaire based on the second edition of the International Classification of Headache Disorders (Headache Classification Subcommittee of the International Headache Society, 2004), which includes 18 questions about chronic headache occurrence, as well as their characteristics, duration, frequency, and accompanying symptoms. All participants had normal or corrected-to-normal visual acuity with no visual deficits, such as color blindness. The experiment was conducted during headache-free periods. Written informed consent was obtained from each participant. This study was approved by the ethical committee of the Graduate School of Engineering, Chiba University, and was conducted in accordance with the principles of the Declaration of Helsinki.

### Apparatus

Figure 2 shows an example of the apparatus. Stimuli were presented on a HMD, (HMZ-T1, Sony Corporation). The luminance output from the HMD ranged from 0.40 to  $28.36 \text{ cd/m}^2$ . A stabilometer (UM-BAR2, Unimec Corporation), which was placed on the floor 60 cm away from the wall, tracked participants' centers of gravity displacements and sampled their fluctuations at 60 Hz.

### Stimuli

We used two static visual stimuli: a homogeneous gray plane and an illusory motion image (Rotating Snakes Kitaoka, 2003, 2011). The illusory motion image (the "snake image"), as shown

in Figure 1A, has been used in many motion perception studies (Conway et al., 2005; Kuriki et al., 2008; Ashida et al., 2012). The smallest unit of the snake image composition was an arrangement of "black-blue-white-yellow" patches. This color patch order was arranged in the same direction throughout, thus inducing illusory rotational motion. Each stimulus included a fixation cross at its center. All had the same mean luminance of  $13.56 \text{ cd/m}^2$  and were subtended at approximately  $29$  by  $29^\circ$  on the HMD's black background.

### Procedure

Our procedure followed a standard stabilometric protocol based on Kapteyn et al. (1983) and Ishizaki et al. (2002), who investigated postural control in patients with migraine. The participants removed their shoes and stood erect, with their knees straight and hands down at their sides, on the stabilometer. First, they stood on the stabilometer without HMD and viewed an eye-level fixation point on the wall for 30 s (eyes open condition). Immediately afterward, they closed their eyes and kept standing for 30 s (eyes closed condition). Next, they stood on the stabilometer with the HMD on their heads and fixated on the center cross on one of the two stimuli for 30 s. Then, they closed their eyes and kept standing on the stabilometer for 30 more seconds. The stimuli were presented in a random order. These procedures were the same across three trials (one per condition). The number of trials was limited in order to prevent excessive visual stress (Wilkins, 1995), such as eye strain and visual discomfort, and to reduce the risk of migraine attacks (Harle et al., 2006).

We recorded the stabilometric parameters of postural sway, total path length (total length of center-of-gravity displacement), rectangular area (area of the maximum amplitude of center-of-gravity displacement), and Romberg ratio (postural sway parameter ratio of measurement under the eyes closed condition to that of the eyes open condition). The Romberg ratio assesses the stabilizing effect of vision in postural control (Diener et al., 1984) and typically measures more than 1 because one's postural sway tends to increase when one's eyes are closed (Travis, 1945; Edwards, 1946; Honma et al., 2012).

### Data Analysis

Total path length (eyes open and closed condition, and its Romberg ratio) and rectangular area (eyes open and closed condition, and its Romberg ratio) were independently analyzed using RMANOVA with a between-participants factor (migraine: patients, controls) and a within-participants factor (stimulus type: without HMD, gray plane, snake image). Because of our relatively small sample size, we did not analyze the effect of the presence of visual aura, although migraine with aura has been suggested to be associated with strong perceptual disturbances (Chronicle and Mulleners, 1994; Shepherd, 2000; Cucchiara et al., 2014). When the sphericity assumption of the RMANOVA was violated, Greenhouse-Geisser correction was applied to the degrees of freedom. Bonferroni correction was used for multiple comparisons. The significance level was set at  $p < 0.05$ . The effect size was reported as eta squared ( $\eta^2$ ).

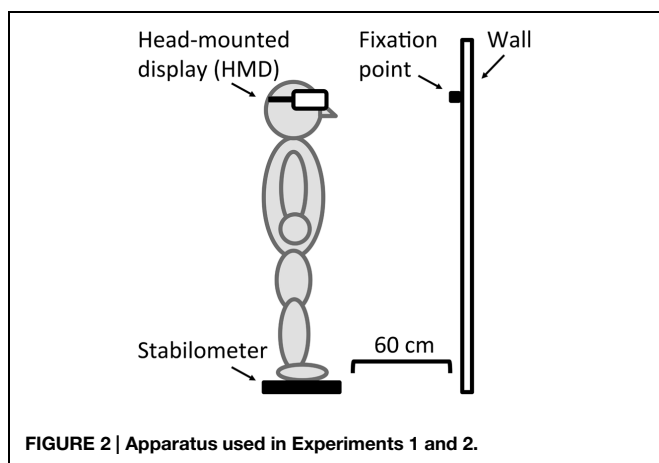


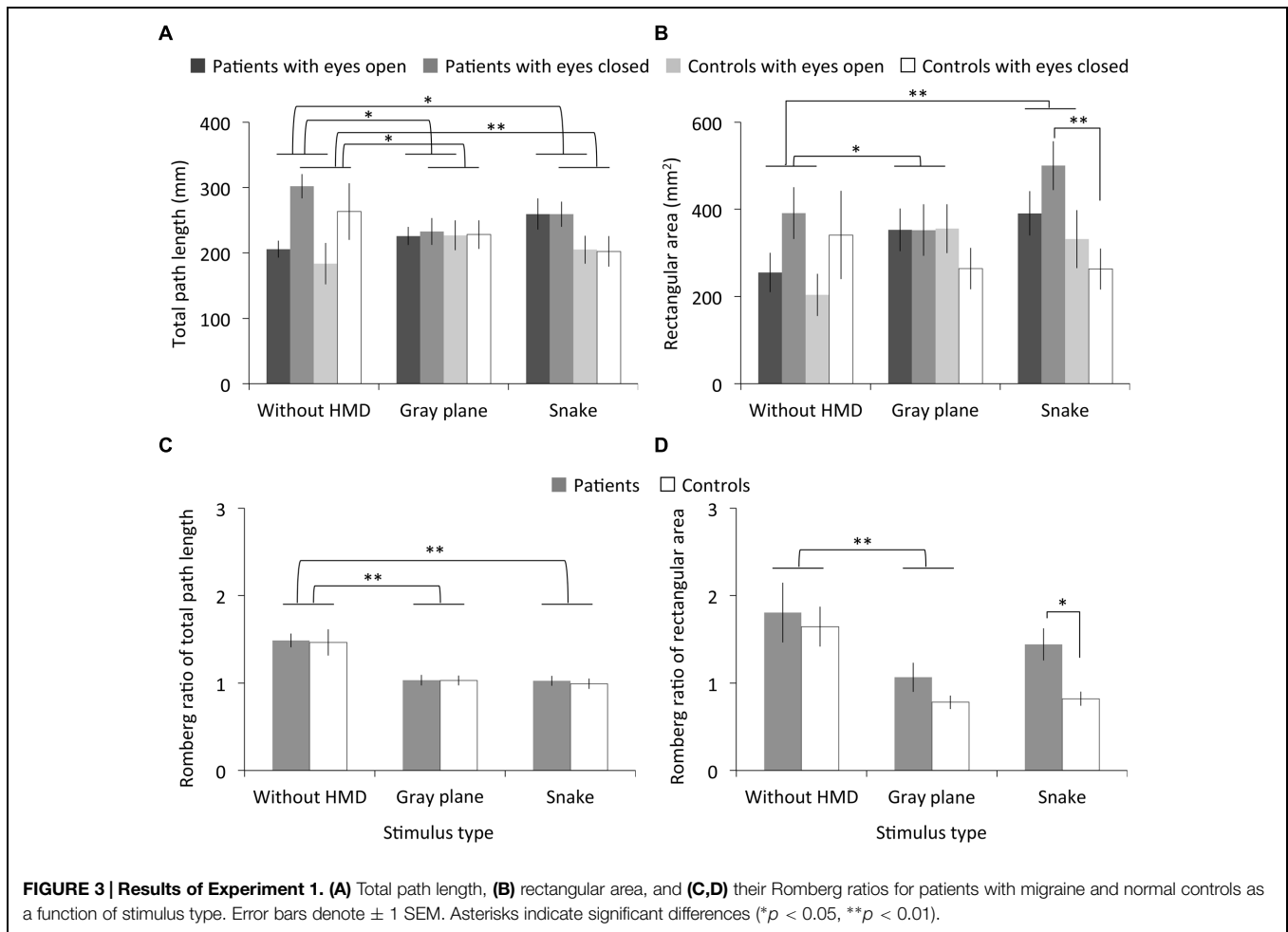
FIGURE 2 | Apparatus used in Experiments 1 and 2.



**Results**

**Figure 3** shows the measured total path length, rectangular area, and their Romberg ratio of both the migraine patients and controls. The RMANOVA revealed significant main effects of stimulus type on total path length under the eyes open and closed conditions and on the Romberg ratio of the total path length [eyes open:  $F_{(2,36)} = 4.48, p < 0.05, \eta^2 = 0.20$ ; eyes closed:  $F_{(2,36)} = 7.16, p < 0.01, \eta^2 = 0.29$ ; Romberg ratio:  $F_{(1.50,27.08)} = 19.69, p < 0.01, \eta^2 = 0.52$ ]. Multiple comparisons revealed significantly larger total path length in the eyes open condition with the gray plane and snake image than in the without HMD condition ( $ps < 0.05$ ) and smaller total path length in the eyes closed condition and its Romberg ratio with the gray plane and snake image ( $ps < 0.01$ ; except for the eyes closed with gray plane condition:  $p < 0.05$ ). We found no significant main effects of migraine or interaction between migraine and stimulus type on total path length and the Romberg ratio of total path length [migraine on eyes open condition:  $F_{(1,18)} = 0.95, p = 0.34, \eta^2 = 0.05$ ; eyes closed:  $F_{(1,18)} = 1.23, p = 0.28, \eta^2 = 0.06$ ; Romberg ratio:  $F_{(1,18)} = 0.10, p = 0.76, \eta^2 = 0.01$ ; interaction on eyes open condition:  $F_{(2,36)} = 2.14, p = 0.13, \eta^2 = 0.11$ ; eyes closed:  $F_{(2,36)} = 1.39, p = 0.26, \eta^2 = 0.07$ ; Romberg ratio:  $F_{(1.50,27.08)} = 0.01, p = 0.97, \eta^2 = 0.00$ ].

On the other hand, we found significant main effects of stimulus type on rectangular area under the eyes open condition and for the Romberg ratio of the rectangular area [eyes open:  $F_{(2,36)} = 8.52, p < 0.01, \eta^2 = 0.32$ ; Romberg ratio:  $F_{(2,36)} = 7.65, p < 0.01, \eta^2 = 0.30$ ] but not on rectangular area under the eyes closed condition [ $F_{(2,36)} = 1.26, p = 0.30, \eta^2 = 0.07$ ]. Multiple comparisons revealed significantly larger rectangular area in the eyes open condition with the gray plane ( $p < 0.05$ ) and snake image ( $p < 0.01$ ) than in the without HMD condition and a smaller Romberg ratio of total path length with the gray plane, although there was no main effect of stimulus type on Romberg ratio ( $p < 0.01$ ). We found no significant main effects of migraine or interaction between migraine and stimulus type on the rectangular area and the Romberg ratio of the rectangular area [migraine on eyes open condition:  $F_{(1,18)} = 0.34, p = 0.57, \eta^2 = 0.02$ ; eyes closed:  $F_{(1,18)} = 3.15, p = 0.09, \eta^2 = 0.15$ ; Romberg ratio:  $F_{(1,18)} = 4.17, p = 0.06, \eta^2 = 0.19$ ; interaction on eyes open condition:  $F_{(2,36)} = 0.44, p = 0.65, \eta^2 = 0.02$ ; eyes closed:  $F_{(2,36)} = 2.05, p = 0.14, \eta^2 = 0.10$ ; Romberg ratio:  $F_{(2,36)} = 0.63, p = 0.54, \eta^2 = 0.03$ ]. However, multiple comparisons revealed a significantly larger rectangular area in patients compared to controls in the eyes closed condition after the observation of the snake image ( $p < 0.01$ ). Consequently, the patients'





Romberg ratio of the rectangular area was significantly higher ( $p < 0.05$ ).

## Discussion

No differences in total path length were found between the gray plane and the snake image observations, while the total path length under the without HMD condition increased more than that under the gray plane and snake image eyes open conditions and their Romberg ratios. This is the case concerning the rectangular area, except for the Romberg ratio in the without HMD and snake image conditions. Participants likely increased their postural sway during the observation of both the gray plane and the snake image. Postural sway may be elicited by visual stimulation with HMD, regardless of illusory motion (Hakkinen et al., 2002).

Concerning the differences between participants, there were no total path length differences between the migraine patients and controls. However, the patients showed larger rectangular area while closing their eyes after viewing the illusory rotating snake image, whereas such differences were not found during the actual observation. There are three possible explanations for these results. First, since migraine patients perceive stronger motion aftereffects than controls (Shepherd, 2006), the illusory motion aftereffect may have increased the patients' postural sway. Indeed, postural sway can be elicited by the motion aftereffect following continuous observations of a horizontally moving visual stimulus (Holten et al., 2014). An alternative hypothesis is that visual stress *per se* induced postural sway. Migraine patients are known to be particularly susceptible to striped patterns with unnatural characteristics (Fernandez and Wilkins, 2008; Juricevic et al., 2010; Penacchio and Wilkins, 2015), and such visual patterns are likely to evoke excess visual cortex excitation (Huang et al., 2003, 2011). Because our snake image contained visual patterns similar to high-contrast stripes, they might have induced the non-specific visual disturbance and the visual pattern cortical response, which would induce postural sway even after the eyes were closed. Finally, migraine patients may simply be more susceptible to sway with closed eyes (Ishizaki et al., 2002). To test these hypotheses, we carried out another experiment including a 30-s interval between the eyes open and closed conditions. If the patients' sway during the eyes closed condition is induced by motion aftereffect or visual stress, the effect will be reduced after the 30-s interval.

## Experiment 2

To examine whether the illusory motion-generated aftereffect can increase postural sway, we inserted an interval between the eyes open and closed conditions to decay the aftereffect. The aftereffect decay should decrease postural sway. Furthermore, we used a snake image without illusory motion as a control stimulus (i.e., one that looked like the Rotating Snakes without the rotating effect; **Figure 1B**). If illusory motion is enough to modulate postural sway, then the control stimulus should not have the same effect.

## Materials and Methods

The material and methods were identical to those used in Experiment 1, except as noted below.

### Participants

This experiment included eight patients with migraine (four female; mean age  $21.29 \pm 3.09$  years) and 14 controls without chronic headaches (seven female; age  $22.36 \pm 2.24$  years) who did not participate in Experiment 1. Two of the patients had visual aura symptoms.

In this experiment, we attempted to investigate migraine patients' motion sickness susceptibility, since this is a common complaint among this population (e.g., Cutrer and Baloh, 1992) and is associated with visually induced postural instability in individuals highly susceptible to motion sickness (Smart et al., 2002; Yokota et al., 2005). According to a standardized questionnaire (Golding, 1998), patients and controls had compatible motion sickness susceptibility (patients: mean = 54.88, SD = 38.15; controls: mean = 53.06, SD = 30.20;  $t_{(20)} = 0.12$ ,  $p = 0.90$ , Cohen's  $d = 0.05$ ). The patients showed slightly low, and controls showed high, scores in comparison with Jeong et al. (2010), who investigated migraine patients' abnormal vestibular functions of migraine patients (patients: approximately 59; controls: approximately 38. Note they reported only graphs without detailed values).

### Stimuli

We used three stimuli: the gray plane and snake image used in Experiment 1 and a reversed image without illusory motion (Kitaoka, 2011) as a control stimulus (**Figure 1B**). The color patch order in the reversed image was reversed between adjacent units to nullify the illusory motion signal. Each stimulus included a fixation cross at its center. All had the same mean luminance of  $13.56 \text{ cd/m}^2$  and were subtended at approximately  $29$  by  $29^\circ$  on the HMD's black background.

### Procedure

To prevent the illusory motion-generated aftereffect from modulating postural sway in the eyes closed condition, we added intervals of 30 s between the eyes open and closed conditions for each measurement. During this interval, the participants who had their eyes open kept standing on the stabilometer while being exposed to a blank display for 30 s. They then closed their eyes, and their postural-sway indices were measured under the eyes closed condition. Directly after the stabilometric measurements, the participants orally rated the magnitude of illusory motion for each stimulus using an 11-point Likert scale, where 0 meant "the image did not appear to move at all," and 10 meant "the image appeared to move most strongly." These procedures were the same across four trials (one per condition).

As in Experiment 1, we conducted only a few trials in order to prevent excessive visual stress and reduce migraine attack risk. For the same reason, we decided not to conduct another trial for measuring the magnitude of illusory motion. Instead, we asked participants to report the perceived illusory motion retrospectively.

**Data Analysis**

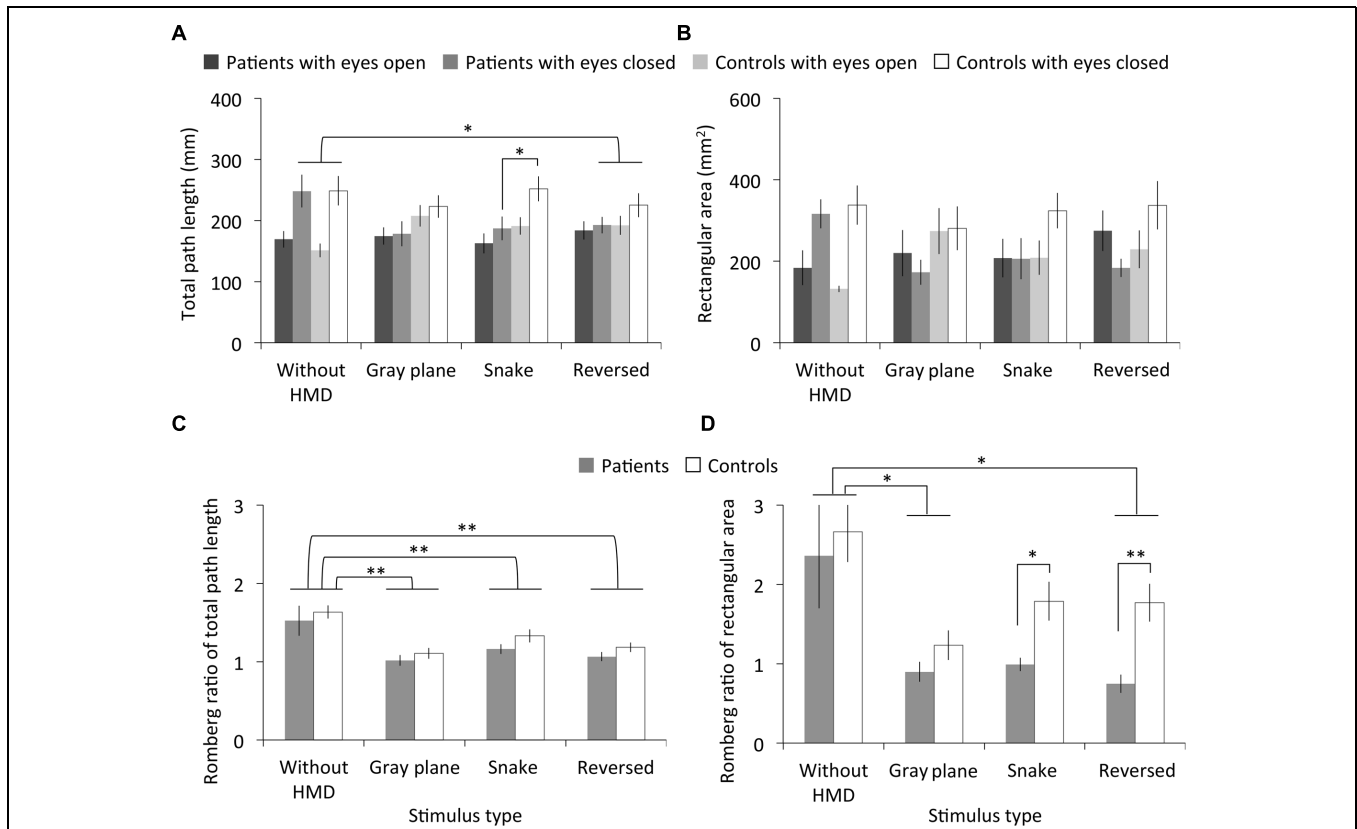
Along with total path length and rectangular area, the illusory motion ratings were analyzed using RMANOVA with a between-participants factor (migraine) and a within-participants factor (stimulus type: without HMD, gray plane, snake image, reversed image).

**Results**

**Figure 4** shows the measured total path length, rectangular area, and Romberg ratio of both the patients and controls. The RMANOVA revealed significant main effects of stimulus type on total path length under the eyes open and closed conditions and on the Romberg ratio of total path length [eyes open:  $F_{(2.24,44.70)} = 4.16, p < 0.05, \eta^2 = 0.17$ ; eyes closed:  $F_{(1.99,39.87)} = 4.68, p < 0.05, \eta^2 = 0.19$ ; Romberg ratio:  $F_{(3,60)} = 15.43, p < 0.01, \eta^2 = 0.44$ ]. Multiple comparisons revealed significantly smaller total path length in the eyes closed condition after the observation of the reversed image than in the without HMD condition ( $p < 0.05$ ), and a smaller Romberg ratio of total path length with the gray plane, snake, and reversed images than was observed in the without HMD conditions ( $ps < 0.01$ ). We found no significant main effects of migraine or interaction between migraine and stimulus type on total path length and the Romberg ratio of total path length [migraine on eyes open condition:  $F_{(1,20)} = 0.49, p = 0.49, \eta^2 = 0.02$ ;

eyes closed:  $F_{(1,20)} = 1.78, p = 0.20, \eta^2 = 0.08$ ; Romberg ratio:  $F_{(1,20)} = 2.33, p = 0.14, \eta^2 = 0.10$ ; interaction on eyes open condition:  $F_{(2.24,44.70)} = 2.61, p = 0.08, \eta^2 = 0.12$ ; eyes closed:  $F_{(1.99,39.87)} = 1.96, p = 0.15, \eta^2 = 0.09$ ; Romberg ratio:  $F_{(3,60)} = 0.08, p = 0.97, \eta^2 = 0.00$ ]. However, multiple comparisons revealed a significantly smaller total path length in patients compared to controls in the eyes closed condition after the snake image observation ( $p < 0.05$ ).

There were no significant main effects of stimulus type and migraine or interaction between migraine and stimulus type on rectangular area under the eyes open and closed conditions [stimulus type on eyes open condition:  $F_{(2.27,45.43)} = 2.54, p = 0.08, \eta^2 = 0.11$ ; eyes closed:  $F_{(2.34,46.74)} = 2.35, p = 0.10, \eta^2 = 0.11$ ; migraine on eyes open condition:  $F_{(1,20)} = 0.05, p = 0.83, \eta^2 = 0.00$ ; eyes closed:  $F_{(1,20)} = 3.10, p = 0.09, \eta^2 = 0.13$ ; interaction on eyes open condition:  $F_{(2.27,45.43)} = 0.81, p = 0.49, \eta^2 = 0.04$ ; eyes closed:  $F_{(2.34,46.74)} = 1.07, p = 0.37, \eta^2 = 0.05$ ]. However, although no main effect of stimulus type was found, multiple comparisons revealed a significantly smaller Romberg ratio for the rectangular area with the gray plane and reversed image than was observed in the without HMD condition ( $ps < 0.05$ ). On the other hand, we found significant main effects of stimulus type and migraine on the Romberg ratio of rectangular area [stimulus type:  $F_{(1.48,29.57)} = 8.57, p < 0.01, \eta^2 = 0.30$ ; migraine:  $F_{(1,20)} = 7.56, p < 0.05, \eta^2 = 0.27$ ], but no significant



**FIGURE 4 | Results of Experiment 2. (A)** Total path length, **(B)** rectangular area, and **(C,D)** their Romberg ratios for patients with migraine and normal controls as a function of stimulus type. Error bars denote  $\pm 1$  SEM. Asterisks indicate significant differences (\* $p < 0.05$ , \*\* $p < 0.01$ ).

interactions between these factors [ $F_{(1.48,29.57)} = 0.63, p = 0.60, \eta^2 = 0.03$ ]. Contrary to Experiment 1's results, multiple comparisons revealed that the Romberg ratio of the rectangular area significantly decreased in patients relative to controls following both the snake ( $p < 0.05$ ) and reversed image observations ( $p < 0.01$ ).

**Figure 5** depicts the subjective magnitude of illusory motion for both the patients and controls. The RMANOVA revealed significant main effects of stimulus type on magnitude [ $F_{(2,40)} = 24.53, p < 0.01, \eta^2 = 0.55$ ]. No significant main effects of migraine or interaction between migraine and stimulus type were found [migraine:  $F_{(1,20)} = 0.53, p = 0.48, \eta^2 = 0.03$ ; interaction:  $F_{(2,40)} = 0.88, p = 0.42, \eta^2 = 0.04$ ]. Multiple comparisons revealed that illusory motion significantly increased for the snake image relative to both the gray plane and the reversed image, and for the reversed image relative to the gray plane ( $p < 0.01$ ).

## Discussion

The results showed differences in total path length and rectangular area Romberg ratios between the without HMD condition and each of the three visual stimuli conditions, except for the Romberg ratio of rectangular area with the snake image, while no differences were found among the stimuli. Similar to Experiment 1's findings, postural sway in both patients and controls was apparently elicited by visual stimulation with HMD, regardless of illusory motion (Hakkinen et al., 2002).

There were no total path length differences between migraine patients and controls, except for longer total path length among the controls under the eyes closed condition after snake image observation. However, contrary to Experiment 1's results, a smaller Romberg ratio for the migraine patients suggested they showed decreased postural sway in the eyes closed condition after observing both the snake and reversed images following a 30-s interval. Therefore, an illusory motion-generated aftereffect can increase postural sway in migraine patients.

Visual stress and discomfort due to stimulus spatial properties (e.g., Fernandez and Wilkins, 2008) can also explain the increased postural sway following observation. Although the snake image

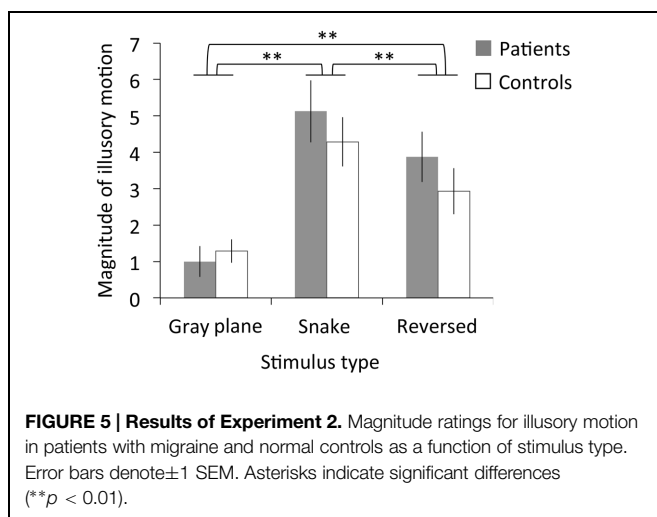
created stronger illusory motion than did the reversed image for both patients and controls, there were no significant differences between the patients' postural sway for either image. In addition, the reversed image also induced more illusory motion than did the gray plane, suggesting that the geometric repetitive patterns of the reversed image may have induced perceptual distortions in migraine patients and controls as a consequence of neural overload (Wilkins, 1995; Imaizumi et al., 2011).

## General Discussion

The present study investigated how migraine patients' postural sway can be modulated by static visual stimuli, especially stimuli with illusory motion perception. In Experiment 1, patients showed larger sway while closing their eyes after viewing the illusory motion. In Experiment 2, they showed decreased sway while closing their eyes after a 30-s interval following their viewing of the illusory motion. Thus, static visual stimuli can induce illusory motion and postural sway, and this effect may last for at least 30 s among the patients.

We hypothesized two mechanisms underlying the increased sway in patients with migraine who closed their eyes after viewing the illusory motion. First, due to their sensitivity to the illusory motion (Huang et al., 2003; Imaizumi et al., 2011) and/or motion aftereffect (Shepherd, 2006), the motion aftereffect continued even after the patients closed their eyes, and this induced postural sway. Although this finding is speculative due to a lack of evidence for the occurrence of aftereffects in the patients, recent findings suggesting that the motion aftereffect itself can induce postural sway (Holten et al., 2014) may support this hypothesis. Second, visual stress in the patients with migraine, which was caused by the stimuli (Wilkins, 1995; Huang et al., 2003, 2011), resulted in the propagation of the visual activities to the more anterior motion- and vestibular-related areas. Consequently, these abnormal neural responses may have induced postural sway due to perceptual disturbances that last for 30 s after the stimulus observation. Given that high-contrast stripes with unnatural spatial characteristics, in terms of the Fourier amplitude spectrum of images (Fernandez and Wilkins, 2008; Juricevic et al., 2010; O'Hare and Hibbard, 2011; Penacchio and Wilkins, 2015), can evoke visual stress (Huang et al., 2003, 2011), our snake and reversed images with patterns similar to high-contrast stripes might have induced the visual stress-induced sway. Such postural sway could be found in both patients and controls, because visual stress is not limited to migraine patients. Normal individuals also find some images uncomfortable to view (Conlon et al., 1999; Fernandez and Wilkins, 2008). However, no studies have reported how long, and to what extent, visual stress can influence postural control when one's eyes are closed. Future studies testing these hypotheses should be beneficial in understanding vision, postural control, and their interactions, especially in migraine patients.

Moreover, we found differences between migraine patients and controls, mostly in the rectangular area. Generally, rectangular area reflects how widely, whereas total path length reflects how frequently the centers of pressure fluctuate. Therefore, patients' greater postural sway as induced by the visual stimuli with



illusory motion can appear widely and slowly after their eyes closed. This characteristic of sway is consistent with Ishizaki et al. (2002), who reported that patients with eyes closed showed larger rectangular area than normal controls but no total path length differences between them, although they did not examine the effect of visual stimulation.

However, it is unclear why our participants did not show more postural sway during their illusory motion observations. There are three possible explanations. First, although visual stimuli with illusory motion may elicit perceptions of body movement (Seno et al., 2013), such stimuli may not lead to actual body movement (i.e., postural sway), which suggests postural sway can be modulated only by direct visual-motion stimulation. Second, HMD weight (~420 g) itself may have caused posture-controlling difficulties, thus attenuating the conditions' effects on postural sway. Indeed, postural instability during an observation with HMD may occur more strongly than that occurring during television viewing (Hakkinen et al., 2002). Finally, negative emotional processes may have influenced postural control. Postural sway can be decreased by visually evoked negative emotions such as disgust (Azevedo et al., 2005; Stins and Beek, 2007), and by imagined painful situations (Lelard et al., 2013), suggesting the activation of a defensive "freezing" posture. Our results showing no increased postural sway during the snake image observation may indicate that visual discomfort cancels out postural sway during observation of the illusory motion stimuli, even though we did not measure perceived visual discomfort. Further investigations should overcome the abovementioned methodological issues by manipulating emotional components in illusory motion stimuli to clarify the effects of illusory motion and visual discomfort on postural sway in light of migraine patients' perceptual characteristics.

Although the two experimental procedures were identical except for the trial number and the 30-s interval between the eyes open and eyes closed measurements, the results obtained from the two experiments seem to differ in several ways besides the illusory motion aftereffect, as noted above. Decreased sway *during* the stimulus observation was found in Experiment 2, although the presence of the 30-s intervals should affect postural sway *after* the observation. We speculate that inter-individual variability in visually induced postural sway (Akiduki et al., 2003), in addition to the migraine effect, may have led to such inter-experiment differences, given that all participated in either Experiment 1 or 2. Besides, motion sickness susceptibility might be the potential factor in increasing postural sway, since visually induced postural instability can be found in highly susceptible individuals (Smart et al., 2002; Yokota et al., 2005); however, there is lack of susceptibility evidence from Experiment 1's participants.

The present study has several limitations. First, the illusory rotating motion parallel to the coronal plane induced by the snake image did not allow us to examine how illusory motion direction and magnitude were associated with those of postural sway, although the perceived motion direction will be consistent

with the direction of increased sway (Lee and Lishman, 1975; Bronstein, 1986). Furthermore, the illusory rotation of one part of the snake image might be counterbalanced by the opposite rotation of another part. If this is the case, we can speculate that overall rotation decreased and, consequently, did not elicit postural sway in the specific direction. Indeed, a follow-up analysis revealed that the ratio of medio-lateral to antero-posterior path length did not differ among stimuli for patients and controls in either Experiment (no main effects of stimulus type:  $F_s < 3.91$ ,  $ps > 0.06$ ,  $\eta^2_s < 0.17$ ; no main effects of migraine:  $F_s < 0.11$ ,  $ps > 0.74$ ,  $\eta^2_s < 0.01$ ). This suggests that our stimuli that included the illusory motion stimulus influenced the amount of postural sway but did not bias the direction of the sway. As Holten et al. (2014) used the horizontally moving stimuli in the coronal plane, further investigation is needed to clarify the direction and magnitude of sway induced by illusory motion in the antero-posterior and medio-lateral dimensions. Second, we measured only one trial for each experimental condition in order to avoid excessive visual stress and the risk of migraine attacks being triggered by visual stimuli (Wilkins, 1995; Harle et al., 2006), resulting from long-term exposure to the illusory motion stimuli, in particular. Finally, we did not measure the perceived illusory motion *during* the stimulus presentation. Instead, we measured this *after* the presentation and limited the number of trials for the abovementioned ethical reason. However, given that there is large inter-individual variability in postural sway (Akiduki et al., 2003) and probable inaccuracy of retrospective perceptual judgment, future studies with larger sample sizes and adequate inter-trial intervals will allow for the repeated measurement of postural sway and separate sessions with which to measure illusory motion more accurately.

In conclusion, the present study examined how illusory motion influenced postural sway in migraine patients and normal controls. We proposed the possibility that illusory motion and visual stress may induce postural sway in migraine patients after illusory motion stimulus observation, although we could not dissociate their effects. Future studies are required to confirm this possibility, considering the multiple factors associated with vision and postural control in migraine patients, such as motion sickness susceptibility and visual discomfort.

## Acknowledgments

The authors would like to thank Haruka Lee for her data collection assistance and two reviewers for their helpful comments. This work was supported by Grant-in-Aids for JSPS Fellows to SI (13J00943), for Challenging Exploratory Research to MH (24650142), and for Scientific Research (B) to MH and SK (23330218) from the Japan Society for the Promotion of Science, and a grant for supporting the recovery efforts following the Great East Japan Earthquake to MH from the Japanese Psychological Association.



## References

- Akiduki, H., Nishiike, S., Watanabe, H., Matsuoka, K., Kubo, T., and Takeda, N. (2003). Visual-vestibular conflict induced by virtual reality in humans. *Neurosci. Lett.* 340, 197–200. doi: 10.1016/S0304-3940(03)00098-3
- Ashida, H., Kuriki, I., Murakami, I., Hisakata, R., and Kitaoka, A. (2012). Direction-specific fMRI adaptation reveals the visual cortical network underlying the “Rotating Snakes” illusion. *Neuroimage* 61, 1143–1152. doi: 10.1016/j.neuroimage.2012.03.033
- Aurora, S. K., Ahmad, B. K., Welch, K. M., Bhardhwaj, P., and Ramadan, N. M. (1998). Transcranial magnetic stimulation confirms hyperexcitability of occipital cortex in migraine. *Neurology* 50, 1111–1114. doi: 10.1212/wnl.50.4.1111
- Azevedo, T. M., Volchan, E., Imbiriba, L. A., Rodrigues, E. C., Oliveira, J. M., Oliveira, L. F., et al. (2005). A freezing-like posture to pictures of mutilation. *Psychophysiology* 42, 255–260. doi: 10.1111/j.1469-8986.2005.00287.x
- Backus, B. T., and Oruc, I. (2005). Illusory motion from change over time in the response to contrast and luminance. *J. Vis.* 5, 1055–1069. doi: 10.1167/5.11.10
- Birren, J. E. (1945). Static equilibrium and vestibular function. *J. Exp. Psychol.* 35, 127–133. doi: 10.1037/h0058240
- Bronstein, A. M. (1986). Suppression of visually evoked postural responses. *Exp. Brain Res.* 63, 655–658. doi: 10.1007/BF00237488
- Chronicle, E., and Mulleners, W. (1994). Might migraine damage the brain? *Cephalalgia* 14, 415–418. doi: 10.1046/j.1468-2982.1994.1406415.x
- Conlon, E. G., Lovegrove, W. J., Chekaluk, E., and Pattison, P. E. (1999). Measuring visual discomfort. *Visual Cogn.* 6, 637–663. doi: 10.1080/135062899394885
- Conway, B. R., Kitaoka, A., Yazdanbakhsh, A., Pack, C. C., and Livingstone, M. S. (2005). Neural basis for a powerful static motion illusion. *J. Neurosci.* 25, 5651–5656. doi: 10.1523/JNEUROSCI.1084-05.2005
- Cucchiara, B., Datta, R., Aguirre, G. K., Idoko, K. E., and Detre, J. (2014). Measurement of visual sensitivity in migraine: validation of two scales and correlation with visual cortex activation. *Cephalalgia*. doi: 10.1177/0333102414547782 [Epub ahead of print].
- Cutrer, F. M., and Baloh, R. W. (1992). Migraine-associated dizziness. *Headache* 32, 300–304. doi: 10.1111/j.1526-4610.1992.hed3206300.x
- Diener, H. C., Dichgans, J., Bacher, M., and Gompf, B. (1984). Quantification of postural sway in normals and patients with cerebellar diseases. *Electroencephalogr. Clin. Neurophysiol.* 57, 134–142. doi: 10.1016/0013-4694(84)90172-X
- Drummond, P. D. (2002). Motion sickness and migraine: optokinetic stimulation increases scalp tenderness, pain sensitivity in the fingers and photophobia. *Cephalalgia* 22, 117–124. doi: 10.1046/j.1468-2982.2002.00332.x
- Drummond, P. D. (2005). Triggers of motion sickness in migraine sufferers. *Headache* 45, 653–656. doi: 10.1111/j.1526-4610.2005.05132.x
- Drummond, P. D., and Granston, A. (2004). Facial pain increases nausea and headache during motion sickness in migraine sufferers. *Brain* 127, 526–534. doi: 10.1093/brain/awh061
- Edwards, A. S. (1946). Body sway and vision. *J. Exp. Psychol.* 36, 526–535. doi: 10.1037/h0059909
- Fernandez, D., and Wilkins, A. J. (2008). Uncomfortable images in art and nature. *Perception* 37, 1098–1113. doi: 10.1068/p5814
- Fischer, E., Bulthoff, H. H., Loughtholm, N. K., and Bartels, A. (2012). Visual motion responses in the posterior cingulate sulcus: a comparison to V5/MT and MST. *Cereb. Cortex* 22, 865–876. doi: 10.1093/cercor/bhr154
- Fraser, A., and Wilcox, K. J. (1979). Perception of illusory movement. *Nature* 281, 565–566. doi: 10.1038/281565a0
- Golding, J. F. (1998). Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. *Brain Res. Bull.* 47, 507–516. doi: 10.1016/S0361-9230(98)00091-4
- Granziera, C., DaSilva, A. F., Snyder, J., Tuch, D. S., and Hadjikhani, N. (2006). Anatomical alterations of the visual motion processing network in migraine with and without aura. *PLoS Med.* 3:e402. doi: 10.1371/journal.pmed.0030402
- Guerraz, M., and Bronstein, A. M. (2008). Mechanisms underlying visually induced body sway. *Neurosci. Lett.* 443, 12–16. doi: 10.1016/j.neulet.2008.07.053
- Hakkinen, J., Vuori, T., and Paikka, M. (2002). “Postural stability and sickness symptoms after HMD use,” in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* (Yasmine Hammamet: IEEE), 147–152. doi: 10.1109/icsmc.2002.1167964
- Harle, D. E., Shepherd, A. J., and Evans, B. J. (2006). Visual stimuli are common triggers of migraine and are associated with pattern glare. *Headache* 46, 1431–1440. doi: 10.1111/j.1526-4610.2006.00585.x
- He, S., Cohen, E. R., and Hu, X. (1998). Close correlation between activity in brain area MT/V5 and the perception of a visual motion aftereffect. *Curr. Biol.* 8, 1215–1218. doi: 10.1016/S0960-9822(07)00512-x
- Headache Classification Subcommittee of the International Headache Society. (2004). The international classification of headache disorders: 2nd edition. *Cephalalgia* 24, 9–160. doi: 10.1111/j.1468-2982.2003.00824.x
- Holten, V., van der Smagt, M. J., Donker, S. F., and Verstraten, F. A. (2014). Illusory motion of the motion aftereffect induces postural sway. *Psychol. Sci.* 25, 1831–1834. doi: 10.1177/0956797614540177
- Honma, M., Endo, N., Osada, Y., Kim, Y., and Kuriyama, K. (2012). Disturbances in equilibrium function after major earthquake. *Sci. Rep.* 2:749. doi: 10.1038/srep00749
- Huang, J., Cooper, T. G., Satana, B., Kaufman, D. I., and Cao, Y. (2003). Visual distortion provoked by a stimulus in migraine associated with hyper-neuronal activity. *Headache* 43, 664–671. doi: 10.1046/j.1526-4610.2003.03110.x
- Huang, J., Zong, X., Wilkins, A., Jenkins, B., Bozoki, A., and Cao, Y. (2011). fMRI evidence that precision ophthalmic tints reduce cortical hyperactivation in migraine. *Cephalalgia* 31, 925–936. doi: 10.1177/0333102411409076
- Imaizumi, S., Koyama, S., and Hibino, H. (2011). A quantitative evaluation of stripe and grid patterns by migraine patients and normal controls. *Bull. Jpn. Soc. Sci. Des.* 58, 69–74.
- Ishizaki, K., Mori, N., Takeshima, T., Fukuhara, Y., Ijiri, T., Kusumi, M., et al. (2002). Static stabilometry in patients with migraine and tension-type headache during a headache-free period. *Psychiatry Clin. Neurosci.* 56, 85–90. doi: 10.1046/j.1440-1819.2002.00933.x
- Jeong, S. H., Oh, S. Y., Kim, H. J., Koo, J. W., and Kim, J. S. (2010). Vestibular dysfunction in migraine: effects of associated vertigo and motion sickness. *J. Neurol.* 257, 905–912. doi: 10.1007/s00415-009-5435-5
- Juricevic, I., Land, L., Wilkins, A., and Webster, M. A. (2010). Visual discomfort and natural image statistics. *Perception* 39, 884–899. doi: 10.1068/p6656
- Kapteyn, T. S., Bles, W., Njiokiktjien, C. J., Kodde, L., Massen, C. H., and Mol, J. M. (1983). Standardization in platform stabilometry being a part of posturography. *Agressologie* 24, 321–326.
- Kitaoka, A. (2003). *Rotating Snakes*. Available at: <http://www.ritsumei.ac.jp/~akitaoka/index-e.html> [accessed January 20, 2015].
- Kitaoka, A. (2011). *A Test Stimulus and a Control Stimulus*. Available at: <http://www.psy.ritsumei.ac.jp/~akitaoka/rotsnakes13e.html> [accessed January 20, 2015].
- Kitaoka, A., and Ashida, H. (2003). Phenomenal characteristics of the peripheral drift illusion. *Vision* 15, 261–262.
- Kuriki, I., Ashida, H., Murakami, I., and Kitaoka, A. (2008). Functional brain imaging of the Rotating Snakes illusion by fMRI. *J. Vis.* 8, 1–10. doi: 10.1167/8.10.16
- Lee, D. N., and Aronson, E. (1974). Visual proprioceptive control of standing in human infants. *Percept. Psychophys.* 15, 529–532. doi: 10.3758/bf03199297
- Lee, D. N., and Lishman, J. R. (1975). Visual proprioceptive control of stance. *J. Hum. Mov. Stud.* 1, 87–95.
- Lelard, T., Montalan, B., Morel, M. F., Krystkowiak, P., Ahmaidi, S., Godefroy, O., et al. (2013). Postural correlates with painful situations. *Front. Hum. Neurosci.* 7:4. doi: 10.3389/fnhum.2013.00004
- Marcus, D. A., Furman, J. M., and Balaban, C. D. (2005). Motion sickness in migraine sufferers. *Expert Opin. Pharmacother.* 6, 2691–2697. doi: 10.1517/14656566.6.15.2691
- Marcus, D. A., and Soso, M. J. (1989). Migraine and stripe-induced visual discomfort. *Arch. Neurol.* 46, 1129–1132. doi: 10.1001/archneur.1989.00520460125024
- Meyer, G. F., Shao, F., White, M. D., Hopkins, C., and Robotham, A. J. (2013). Modulation of visually evoked postural responses by contextual visual, haptic and auditory information: a ‘virtual reality check.’ *PLoS ONE* 8:e67651. doi: 10.1371/journal.pone.0067651
- Morrone, M. C., Tosetti, M., Montanaro, D., Fiorentini, A., Cioni, G., and Burr, D. C. (2000). A cortical area that responds specifically to optic flow, revealed by fMRI. *Nat. Neurosci.* 3, 1322–1328. doi: 10.1038/81860
- O’Hare, L., and Hibbard, P. B. (2011). Spatial frequency and visual discomfort. *Vision Res.* 51, 1767–1777. doi: 10.1016/j.visres.2011.06.002

- Penacchio, O., and Wilkins, A. J. (2015). Visual discomfort and the spatial distribution of Fourier energy. *Vision Res.* 108, 1–7. doi: 10.1016/j.visres.2014.12.013
- Reason, J. T., and Brand, J. J. (1975). *Motion Sickness*. Oxford: Academic Press.
- Seno, T., Kitaoka, A., and Palmisano, S. (2013). Vection induced by illusory motion in a stationary image. *Perception* 42, 1001–1005. doi: 10.1068/p7511
- Shepherd, A. J. (2000). Visual contrast processing in migraine. *Cephalalgia* 20, 865–880. doi: 10.1046/j.1468-2982.2000.00119.x
- Shepherd, A. J. (2006). Local and global motion after-effects are both enhanced in migraine, and the underlying mechanisms differ across cortical areas. *Brain* 129, 1833–1843. doi: 10.1093/brain/awl124
- Slobounov, S., Wu, T., Hallett, M., Shibasaki, H., Slobounov, E., and Newell, K. (2006). Neural underpinning of postural responses to visual field motion. *Biol. Psychol.* 72, 188–197. doi: 10.1016/j.biopsycho.2005.10.005
- Smart, L. J. Jr. Stoffregen, T. A., and Bardy, B. G. (2002). Visually induced motion sickness predicted by postural instability. *Hum. Factors* 44, 451–465. doi: 10.1518/0018720024497745
- Smith, A. T., Wall, M. B., and Thilo, K. V. (2012). Vestibular inputs to human motion-sensitive visual cortex. *Cereb. Cortex* 22, 1068–1077. doi: 10.1093/cercor/bhr179
- Stins, J. F., and Beek, P. J. (2007). Effects of affective picture viewing on postural control. *BMC Neurosci.* 8:1–7. doi: 10.1186/1471-2202-8-83
- Travis, R. C. (1945). An experimental analysis of dynamic and static equilibrium. *J. Exp. Psychol.* 35, 216–234. doi: 10.1037/h0059788
- van Asten, W. N., Gielen, C. C., and van der Gon, J. J. D. (1988). Postural adjustments induced by simulated motion of differently structured environments. *Exp. Brain Res.* 73, 371–383. doi: 10.1007/BF00248230
- Wall, M. B., and Smith, A. T. (2008). The representation of egomotion in the human brain. *Curr. Biol.* 18, 191–194. doi: 10.1016/j.cub.2007.12.053
- Wilkins, A., Nimmo-Smith, I., Tait, A., McManus, C., Della Sala, S., Tilley, A., et al. (1984). A neurological basis for visual discomfort. *Brain* 107, 989–1017. doi: 10.1093/brain/107.4.989
- Wilkins, A. J. (1995). *Visual Stress*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198521747.001.0001
- Yates, B. J., Miller, A. D., and Lucot, J. B. (1998). Physiological basis and pharmacology of motion sickness: an update. *Brain Res. Bull.* 47, 395–406. doi: 10.1016/S0361-9230(98)00092-6
- Yokota, Y., Aoki, M., Mizuta, K., Ito, Y., and Isu, N. (2005). Motion sickness susceptibility associated with visually induced postural instability and cardiac autonomic responses in healthy subjects. *Acta Otolaryngol.* 125, 280–285. doi: 10.1080/00016480510003192
- Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., and Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. *J. Neurosci.* 11, 641–649.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Imaizumi, Honma, Hibino and Koyama. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Amplitude-modulated stimuli reveal auditory-visual interactions in brain activity and brain connectivity

Mark Laing, Adrian Rees\* and Quoc C. Vuong\*

*Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK*

## OPEN ACCESS

### Edited by:

Achille Pasqualotto,  
Sabanci University, Turkey

### Reviewed by:

Anton Ludwig Beer,  
Universität Regensburg, Germany

Deborah Aphorpe,  
Australian National University,  
Australia

### \*Correspondence:

Adrian Rees and  
Quoc C. Vuong,  
Institute of Neuroscience, Newcastle  
University, Framlington Place,  
Newcastle upon Tyne NE2 4HH, UK  
quoc.vuong@newcastle.ac.uk;  
adrian.rees@ncl.ac.uk

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 10 April 2015

**Accepted:** 09 September 2015

**Published:** 02 October 2015

### Citation:

Laing M, Rees A and Vuong QC  
(2015) Amplitude-modulated stimuli  
reveal auditory-visual interactions  
in brain activity and brain connectivity.  
*Front. Psychol.* 6:1440.  
doi: 10.3389/fpsyg.2015.01440

The temporal congruence between auditory and visual signals coming from the same source can be a powerful means by which the brain integrates information from different senses. To investigate how the brain uses temporal information to integrate auditory and visual information from continuous yet unfamiliar stimuli, we used amplitude-modulated tones and size-modulated shapes with which we could manipulate the temporal congruence between the sensory signals. These signals were independently modulated at a slow or a fast rate. Participants were presented with auditory-only, visual-only, or auditory-visual (AV) trials in the fMRI scanner. On AV trials, the auditory and visual signal could have the same (AV congruent) or different modulation rates (AV incongruent). Using psychophysiological interaction analyses, we found that auditory regions showed increased functional connectivity predominantly with frontal regions for AV incongruent relative to AV congruent stimuli. We further found that superior temporal regions, shown previously to integrate auditory and visual signals, showed increased connectivity with frontal and parietal regions for the same contrast. Our findings provide evidence that both activity in a network of brain regions and their connectivity are important for AV integration, and help to bridge the gap between transient and familiar AV stimuli used in previous studies.

**Keywords:** auditory-visual integration, temporal congruence, brain network, psychophysiological interaction, amplitude modulation

## Introduction

Everyday events and objects concurrently stimulate multiple senses, and an important task for the brain is to determine whether signals received by different modalities belong to the same or different sources. Perceptually combining different sensory signals from the same source can enhance performance, particularly when environmental conditions are not ideal. For example, visual information about a speaker's lips can enhance the intelligibility of her spoken speech in a noisy room (Sumbly and Pollack, 1954; Grant and Seitz, 2000). Combining information from different sources can lead to multi-sensory illusions; most notably, when the syllable conveyed by a speaker's voice does not match the one conveyed by her lips, observers perceive a syllable that is neither the auditory syllable nor the visual syllable (McGurk and MacDonald, 1976). There is accumulating behavioral and neural evidence that the strength of multi-sensory integration depends on the congruence between sensory signals. This congruence can be defined by spatial or temporal information, such as sensory signals originating from the same spatial location or occurring in close temporal proximity (e.g., Frassinetti et al., 2002). Congruence can also be defined

by semantic information, such as a dog's bark matching a picture of a dog rather than a picture of a cat (e.g., Naumer et al., 2011).

In the current study, we focused on how temporal congruence facilitates auditory-visual (AV) integration at the neural level. Events in the environment are dynamic and present multi-sensory information continuously over a range of time scales. With many events occurring at similar locations, the temporal congruence of multi-sensory information may be a powerful cue for combining sensory signals: congruence will generally be higher for sensory signals originating from the same source than from different sources. Indeed, temporal congruence can lead to behavioral advantages across various stimuli and tasks. Following the example above, focusing on the speaker's lips would enhance the intelligibility of her speech despite other simultaneous conversations and events. In this case, the temporal congruence is produced by the synchrony between the continuously changing shape of the lips and the changing amplitude of the speech envelope over an extended period (Grant and Seitz, 2000; Vander Wyk et al., 2010). Not only will the synchrony between the speaker's lips and speech be higher than between the lips and other environmental sounds, there may also be congruent semantic information derived from lip reading and the speech itself (Calvert et al., 2000). For non-meaningful stimuli (e.g., simple tones and visual shapes), temporal congruence can lead to higher target detection (e.g., Frassinetti et al., 2002; Lovelace et al., 2003; Maddox et al., 2015), better motion discrimination (e.g., Lewis and Noppeney, 2010; Ogawa and Macaluso, 2013) and faster responses (e.g., Nozawa et al., 1994; Diedrich and Colonius, 2004) when the auditory and visual signals are congruent.

Complementing behavioral evidence, human brain imaging studies have identified regions that respond more to AV stimuli than to auditory or visual stimuli alone (e.g., Calvert et al., 1999, 2000, 2001; Beauchamp et al., 2004; Stevenson et al., 2007; Sadaghiani et al., 2009; Stevenson and James, 2009; Vander Wyk et al., 2010; Naumer et al., 2011; for a review see Stein and Stanford, 2008). These putative multi-sensory regions include those within the temporal [e.g., superior temporal sulcus (STS)], the parietal [e.g., intraparietal sulcus (IPS)] and the frontal lobes [e.g., inferior frontal gyrus (IFG)], as well as subcortical structures such as the superior colliculus (Meredith and Stein, 1983). Several of these studies show the importance of temporal congruence in increasing regional activity for congruent AV stimuli and decreasing regional activity for incongruent AV stimuli. In an early human-imaging paper, Calvert et al. (2001) presented auditory white noise bursts in parallel with a visual checkerboard pattern with reversing black and white squares. Each sensory stimulus type had a different duration (auditory: 39 s on, 39 s off; visual: 30 s on, 30 s off) giving rise over time to auditory, visual, and AV periods. In separate blocks, Calvert et al. (2001) also manipulated whether the onset of the sound and onset of the checkerboard occurred at the same time (congruent) or whether the onsets were randomly out of temporal phase with respect to each other (incongruent). Observers listened passively to all stimuli. Importantly, their study showed that temporal congruence led to response enhancement when the auditory and visual signals were congruent and to response suppression when they were incongruent, emphasizing the importance of

temporal information for modulating brain activations. Using a similar paradigm, but with speech stimuli, Calvert et al. (2000) found that the temporal congruence of meaningful stimuli also elicited similar response enhancement and suppression, with the strongest response in the left posterior STS. In this study, they paired visual lip movements with either the correct sound track (congruent) or another sound track (incongruent). On incongruent blocks, the mis-match between the lip movements and sound track gave rise to different temporal patterns of the auditory and visual signals (as well as semantic incongruency due to lip reading). These overall patterns of results have been replicated with different types of auditory and visual stimuli such as non-meaningful transient tone-bursts (i.e., "beeps") and flashes (Noesselt et al., 2007), speech-like stimuli (circles and ellipses animated with speakers' speech envelopes; Vander Wyk et al., 2010) and meaningful non-speech stimuli (e.g., videos of tool use; Stevenson et al., 2007; Stevenson and James, 2009; Werner and Noppeney, 2010). These studies suggest that congruent AV stimuli typically lead to stronger responses than incongruent AV stimuli but this is not always the case (e.g., Noesselt et al., 2012). For instance, when congruency is defined along a semantic dimension, semantically incongruent AV stimuli can lead to larger responses than semantically congruent AV stimuli (e.g., Hocking and Price, 2008; Meyer et al., 2011; Beer et al., 2013).

The regional responses to AV stimuli are important but they do not necessarily provide a complete picture of multi-sensory integration at the neural level for at least two complementary reasons. First, there are anatomical connections between brain regions, allowing information to be transmitted quickly between them (Felleman and Van Essen, 1991; Beer et al., 2011, 2013; van den Brink et al., 2014). Second, brain regions can show functional connectivity with each other; that is, activity in different regions can co-vary over time (Hagmann et al., 2008). These anatomical and functional connections may, for instance, allow regions to pool information from other regions (e.g., Noppeney et al., 2010; Beer et al., 2013). Several human studies have investigated brain connectivity patterns for AV integration (e.g., Noesselt et al., 2007, 2012; Lewis and Noppeney, 2010; Noppeney et al., 2010; Werner and Noppeney, 2010; Lee and Noppeney, 2011; Ogawa and Macaluso, 2013; Kim et al., 2015). For example, Werner and Noppeney (2010) found interactions between auditory and visual regions (see also Lewis and Noppeney, 2010, and Ogawa and Macaluso, 2013, for motion discrimination). They had observers categorize videos of everyday actions as tools or instruments, and varied both the presence of a sensory signal and (if present) how informative it was about the action. Auditory and visual signals were degraded by adding visual or auditory noise. This manipulation reduced the reliability of the sensory signal, which is known to increase the strength of multi-sensory integration. The concurrent presentation of a visual signal automatically increased responses in auditory cortex via direct connectivity with the visual cortex or indirectly through the STS. Interestingly, Noesselt et al. (2012) found that perceived temporal congruence could also modulate functional connectivity. They presented observers with AV speech streams in which the auditory



stream was physically leading, the visual stream was physically leading, or the streams were physically synchronous. The authors further manipulated the stimulus onset asynchrony between the auditory and visual streams to create bistable percepts. That is, observers would perceive physically asynchronous AV streams (visual leading or auditory leading) sometimes as asynchronous and sometimes as synchronous. Noesselt et al. (2012) found that despite the same physical stimuli (e.g., visual leading), there was an increased functional connectivity between the STS and right prefrontal regions when observers correctly perceived the AV stimulus as asynchronous relative to when they incorrectly perceived the AV stimulus as synchronous. For transient auditory tone and visual flash stimuli, Noesselt et al. (2007) found increased functional connectivity between the STS and primary visual and auditory regions, rather than frontal regions, when the tones and flashes were temporally coincident (synchronous) relative to when they were temporally non-coincident (asynchronous).

Most human imaging studies have focused on speech, music and other meaningful (e.g., animals or tools) stimuli that carry high-level cognitive and/or semantic information. We do not know if the same brain regions are activated by simpler AV constructs. Furthermore observers may have differential experiences with familiar stimuli, which can shape how the brain responds to them. For example, Lee and Noppeney's (2011) data showed that connectivity could change with expertise. On the other hand, previous studies of AV interactions using non-meaningful AV stimuli often use transient sounds and visual patterns that rarely occur in nature (Sekuler et al., 1997; Shams et al., 2000; Calvert et al., 2001; Noesselt et al., 2007). Here we used continuous sounds and shapes which are nonetheless unfamiliar AV stimuli. These consisted of a three-dimensional object that was sinusoidally modulated in size and combined with a tone that was sinusoidally modulated in amplitude. Both the auditory and visual signals were thus continuous and were modulated at modulation rates commonly experienced in familiar stimuli such as speech (e.g., Plomp, 1983; Rosen, 1992; Shannon et al., 1995). Using these AV stimuli, we reported that observers' sensitivity to amplitude differences between two sequentially presented AV stimuli were affected if the auditory and visual signals were modulated at the same rate (congruent) but not when they were modulated at different rates (incongruent; Vuong et al., 2014). This temporal manipulation allowed us to test how combining auditory and visual information changes brain activation and/or brain connectivity, without the confound of speech, language, and semantic information. We found that temporally congruent AV stimuli led to increased activation in putative multi-sensory areas in temporal and parietal lobes, consistent with previous reports (e.g., Calvert et al., 2000, 2001; Noesselt et al., 2007, 2012), but temporally incongruent AV stimuli led to increased functional connectivity between auditory/visual regions and predominantly frontal regions (see also Noesselt et al., 2012). Overall, the results suggest that both brain activation and connectivity changes support AV integration. Our results provide an important link between transient, unfamiliar stimuli and continuous real-world objects, speech and music.

## Materials and Methods

### Participants

Nine right-handed adults (seven males, two females; age in years:  $M = 24$ ,  $SD = 1.6$ ; range: 21–26 years) participated in the study. All participants reported normal hearing and normal or corrected-to-normal vision. All participants provided informed consent. The study was performed in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Newcastle University.

### Apparatus

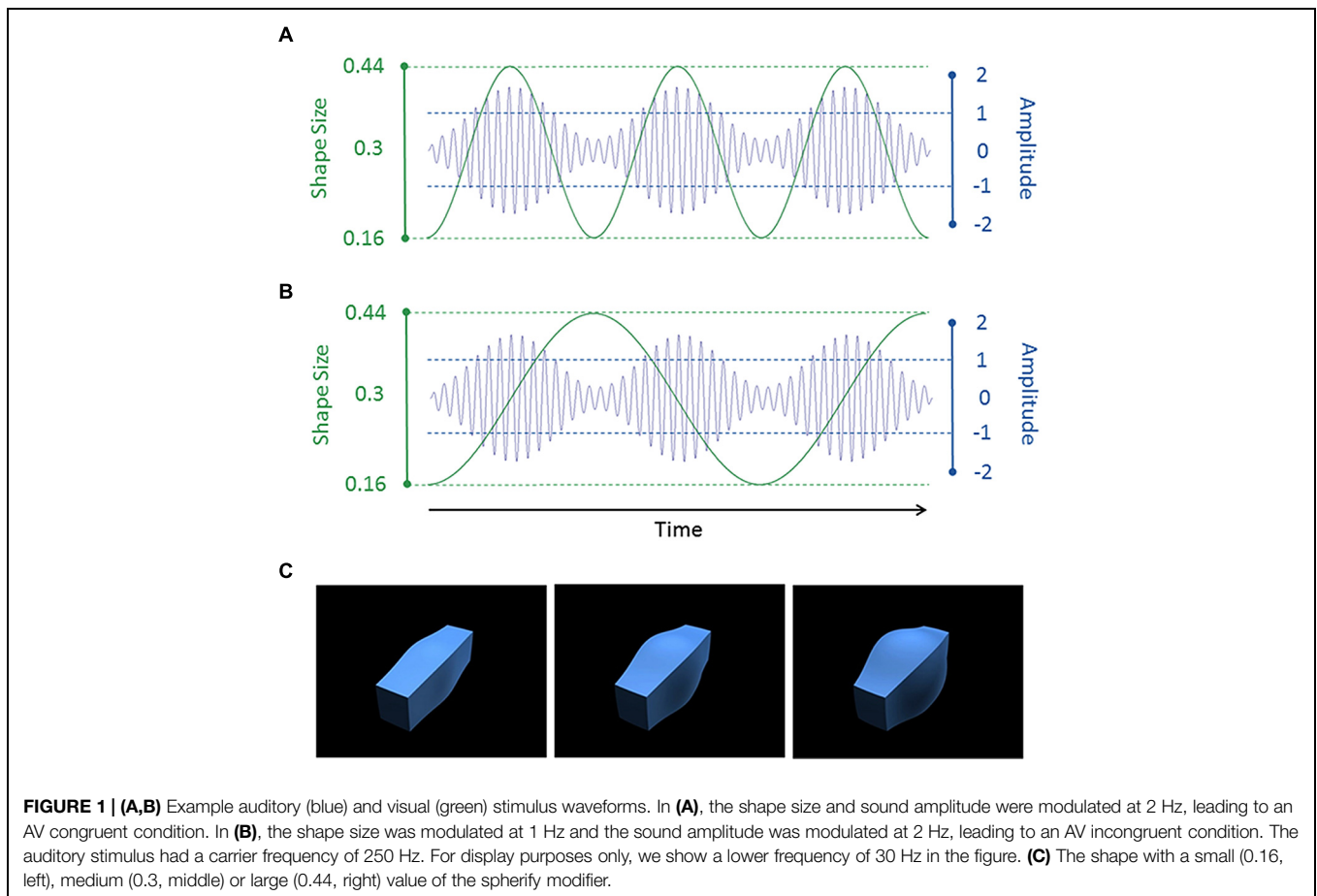
The visual stimuli were back-projected onto a screen at the foot end of the scanner using a canon XEED LCD projector (1280 × 1024 pixels, 60 Hz). Participants viewed the projection through an angled mirror attached to the head coil ~10 cm above their eyes. The sounds were presented using an MR-compatible audio system and delivered with electrostatic transducer headphones (NordicNeuroLab). Participants wore earplugs to further protect against scanner noise. Head motion was restricted by placing foam pads between the head and the head coil. The experiment was run on a Windows 7 PC using the Psychophysics Toolbox version 3<sup>1</sup> (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007; run on 32-bit MATLAB 2012, Mathworks, Inc.) to control the experiment, present the stimuli and record behavioral responses. Participants responded via a MR-compatible response pad (LumiTouch).

### Stimuli

**Figure 1** illustrates the auditory and visual stimuli used in the study. The auditory stimuli consisted of amplitude-modulated tones (see **Figures 1A,B**), with a 250 Hz carrier frequency sinusoidally amplitude-modulated at 1 or 2 Hz with a modulation depth of 70%. They were created in MATLAB 2012 and saved as stereo wav files with a 44.1 kHz sampling rate. We were unable to measure the volume within the scanner. We therefore set the sound level of our stimuli to 75 dB SPL in a sound-attenuated room. The sounds were presented via headphones on a high fidelity MR-compatible audio system (NordicNeuroLab). We used a fixed setting for the audio system (volume level = 4) for all participants in the scanner but they could all clearly hear the tones with our sparse imaging protocol.

The visual stimuli consisted of size-modulated three-dimensional (3D) cuboids (see **Figures 1A–C**). The cuboid was created using 3D Studio Max version 7 (Autodesk, Inc.). The “spherify” modifier was applied to a blue rectangular box (1.0 × 1.2 × 4.0 units [width × height × length]) to vary the size of the central portion of the cuboid. This modifier can vary from 0 (rectangle) to 1.0 (sphere). As with the tones, a 1 or 2 Hz sinusoid waveform was used to modulate the modifier between 0.16 and 0.44 (oscillating around a mean of 0.3). The cuboid was rendered against a uniform black background from an oblique camera viewpoint. The bounding box of the cuboid subtended a visual angle of 13.7° × 13.7° (300 pixels × 300 pixels). The videos

<sup>1</sup>[www.psychtoolbox.org](http://www.psychtoolbox.org)



were saved as Quicktime movie files (240 frames; 60 frames per second; H.264 compression).

The auditory and visual stimuli were 4.0 s in duration. There were thus four cycles with the 1 Hz modulation rate and eight cycles with the 2 Hz rate. The two modalities and two modulation rates were factorially combined to produce four stimuli. Importantly, there were two congruency conditions which reflected whether the auditory and visual stimuli had the same (congruent) or different (incongruent) modulation rates. The 1 Hz modulation rate was considered to be “slow” and the 2 Hz modulation rate was considered to be “fast.”

## Design and Procedure

There were six experimental conditions in the current study. Participants were instructed to attend to either the auditory or visual stimulus. For each attended stimulus, they were presented with the audio- or video-only stimulus (A or V), the AV congruent stimulus (AVC) and AV incongruent stimulus (AVI). Each experimental condition was presented twice in each functional run in a random order giving a total of 12 experimental blocks. Before each experimental block, there was an instruction block to inform participants to attend to the auditory or visual stimulus. Each functional run was ~10 min in duration. There were three functional runs for eight of the participants and two runs for one participant.

A 10.0 s instruction screen appeared before each experimental block in which the label “AUDITION” or “VISION” was presented at the center of the screen (Courier, 64 font size, white text). There were four trials in each 40.0 s experimental block. Participants judged whether the attended stimulus (audio or video) was “slow” (1 Hz) or “fast” (2 Hz) while ignoring the modulation rate of the unattended stimulus (if present). They used a response pad to make their response (with the response mapping counterbalanced across participants). In each 10.0 s trial, a fixation cross was presented for 2.0 s, followed by the stimuli for 4.0 s, and by a blank screen for 2.5 s. Participants could only respond during a 1.5 s period in which the word “respond” was displayed (Courier, 24 font size, white text). If they responded before this period or did not respond within this period, the next trial continued and the response was counted as an error. The fMRI image acquisition occurred at the beginning of each trial whilst the fixation cross was displayed and recorded the brain response to the preceding trial. Thus there was no interference from the scanner noise during the presentation of the auditory stimuli. Outside the scanner, participants were given a practice block for each experimental condition to familiarize them with the trial sequence and enable them to appreciate the difference between “slow” and “fast” auditory and visual stimuli. The modulation-rate judgment task ensured that participants remained alert in the scanner

but was designed to be an easy task, and was not used to assess the extent to which participants integrated the AV stimuli.

## Image Acquisition

All participants were scanned at the Newcastle Magnetic Resonance Centre. Anatomical T1-weighted images and functional T2\*-weighted echo planar images (EPIs) were acquired from a 3 T Philips Intera Achieva MR scanner using a Philips 8-channel receive-only head coil. The high resolution T1-weighted scan consisted of 150 slices and took approximately 5 min to acquire. The parameters of the structural scan were: repetition time (TR) = 9.6 ms, echo time (TE) = 4.6 ms, flip angle = 8°. The field of view (FOV) was 240 mm × 240 mm × 180 mm with a matrix size of 208 × 208 pixels. Each voxel was 0.94 mm × 0.94 mm × 1.2 mm in size. The T2\*-weighted EPIs consisted of 28 axial slices acquired from the bottom to the top of the head. The parameters of the EPIs were: acquisition time (TA) = 1.3 s, TR = 10 s, TE = 30 ms, flip angle = 90°. The FOV was 192 mm × 192 mm × 125.5 mm with a matrix size of 64 × 64 pixels. Each voxel was 3 mm × 3 mm × 4 mm in size, with a 0.5 mm gap between slices. We used sensitivity encoding (SENSE) with factor = 2 to increase the signal-to-noise ratio of the functional images. For each participant, a total of 62 functional images were acquired in each run (~10 min per run). Due to some technical problems, 64 functional images were acquired in each run for one participant. Before each functional run, four “dummy” scans were acquired to allow for equilibration of the T1 signal.

## fMRI Pre-processing

Functional images were realigned to the first image across all runs for each participant and re-sliced to correct for head motion. These images were normalized to a standard Montreal Neurological Institute (MNI) EPI T2\*-weighted template with a resampled voxel size of 3 mm × 3 mm × 3 mm. They were then spatially smoothed with a 6 mm full-width-at-half-maximum Gaussian kernel to improve the signal-to-noise ratio and to allow for comparisons across participants. To remove low-frequency drifts in the signal, we applied a high-pass filter with a cutoff of 180 s.

## fMRI Whole-brain Analysis

The preprocessed data were analyzed using SPM8<sup>2</sup> (Friston et al., 1994). We used the general linear model (GLM) with a two-step mixed-effects approach. First, a fixed-effects model was used to analyze each participant's data set. Second, a random-effects model was used to analyze the individual datasets at the group level. No additional smoothing of the images was used at the group level.

The design matrix for each participant was constructed as follows. The onset and duration for each of the six experimental blocks and the instruction (baseline) block were modeled as boxcar functions (40.0 s for experimental blocks, 10.0 s for

the instruction block). These boxcar functions were convolved with a finite impulse response function (Order 1) implemented in SPM8. In addition to these regressors of interest, the six movement parameters (roll, yaw, pitch, and three translation terms) and a constant term for each session were included in the design matrix as regressors of no interest. A linear combination of the regressors was fitted to the BOLD signal to estimate the beta weight for each regressor.

For the first-level analysis, contrast images were computed from the beta-weight images. We used the contrasts  $A > \text{instruction}$  and  $V > \text{instruction}$  to localize uni-sensory auditory and visual regions. There are several statistical criteria for localizing multi-sensory regions (Beauchamp, 2005). Given our temporal congruency manipulation, we focused on the contrast  $AVC > AVI$  (averaging across the attention conditions) to localize multi-sensory regions (e.g., Calvert et al., 2000, 2001; Beauchamp et al., 2004; Noesselt et al., 2007, 2012). For the second-level group analysis, one-sample *t*-tests of participants' contrast images were conducted at each voxel.

The goal of the whole-brain analyses was to functionally localize well-established uni- and multi-sensory regions. These regions served as seeds for the functional connectivity analyses described below. We therefore used a liberal statistical threshold (uncorrected  $p < 0.001$  at the voxel level) and we focused on those clusters that were within cortical regions reported in previous studies (e.g., Calvert et al., 2000, 2001; Beauchamp et al., 2004; Noesselt et al., 2007, 2012). For all other statistical tests, we used  $\alpha = 0.05$  and considered  $0.05 < p < 0.10$  as marginal effects.

## fMRI Psychophysiological Interaction Analysis

We used the generalized form of context-dependent psychophysiological interaction (PPI) analyses<sup>3</sup> (McLaren et al., 2012; see also Friston et al., 1997; Gitelman et al., 2003) to identify regions which show changes in functional connectivity as a function of audio-visual congruency. For the PPI analyses, we derived three regressors from the BOLD time series. First, a regressor representing the physiological activity in a seed area was computed by deconvolving the first eigenvariate of the BOLD time series from all voxels in that area to estimate changes in neural activity in that area. Second, a regressor representing the psychological context was computed by convolving a boxcar time series for the two congruency conditions with the canonical hemodynamic response function implemented in SPM8. To test for increased connectivity on AV congruent trials, AVC blocks were coded as +1 and AVI blocks were coded as -1. Conversely to test for increased connectivity on AV incongruent trials, AVC blocks were coded as -1 and AVI blocks were coded as +1. Lastly and importantly, a regressor representing a PPI was computed by multiplying the first two regressors. These three regressors were used to augment each participant's design matrix from the whole-brain analyses (see above). In this augmented design matrix, the experimental conditions and head-movement parameters were treated as regressors of no interest to factor out the contribution of the

<sup>2</sup><http://www.fil.ion.ucl.ac.uk/spm>

<sup>3</sup>[www.nitrc.org/projects/gppi](http://www.nitrc.org/projects/gppi)

experimental conditions on the PPI analyses (McLaren et al., 2012).

We used the functionally localized uni-sensory and multi-sensory regions (see analysis above) as the bases of our seeds. To generate seed areas, we defined a 6 mm sphere centered on the peak voxel of a given region (i.e., the voxel with the largest response). Only significant voxels within this sphere were included in the seed. Although our multi-sensory regions were based on contrasting AVC and AVI conditions, it is important to note that the PPI regressor combined with factoring out the contribution of the experimental conditions meant that we did not bias our sampling for the multi-sensory seeds. As with the whole-brain analyses, we first estimated regressor beta weights for each participant (first-level analysis). We then submitted the participants' beta-weight image for the PPI regressor to a one-sample *t*-test against zero for the contrasts AVC > AVI or AVI > AVC (second-level analysis).

## Results

### Behavioral Results

**Table 1** presents the behavioral results in the scanner. As expected, participants had no difficulty distinguishing the fast and slow rates in the modulation-rate judgment task (accuracy > 90%). The proportion correct data and response times from correct trials were submitted to a 2 attended stimulus (audio, video)  $\times$  3 AV congruency (audio/video-only, AV congruent, AV incongruent) repeated measures analysis of variance (ANOVA). For accuracy, there was only a marginally significant main effect of attended stimulus,  $F(1,8) = 5.3$ ,  $p = 0.051$ ,  $\eta_p^2 = 0.40$ . Participants were marginally more accurate when attending to the visual compared to the

auditory stimulus (vision:  $M = 0.97$ ,  $SEM = 0.01$ ; audition:  $M = 0.93$ ,  $SEM = 0.02$ ). The effect of congruency and the interaction between the two factors were not significant,  $F_s < 1.0$ . For correct response times, there was no main effect of attended stimulus or congruency, and there was no interaction between the two factors, all  $F_s < 1.4$  and  $p_s > 0.28$ .

### fMRI Whole-brain Results Uni-Sensory Regions

We localized auditory and visual regions using the contrasts A > instruction and V > instruction, respectively. For the auditory contrast, we used an initial threshold of  $p = 0.01$  and  $k = 20$ . For the visual contrast, we used an initial threshold  $p = 0.001$  and  $k = 20$ . **Tables 2** and **3** present the auditory and visual results, respectively. For these and subsequent tables, we also present regions which had uncorrected  $p < 0.001$  peak voxels and we used the WFU Pickatlas toolbox to label the reported regions (with exceptions as noted). The labels are based on the peak voxel (Maldjian et al., 2003). For the auditory contrast, we found activations in the area of the posterior right STG corresponding to Heschl's gyrus, and activations in the left posterior and right anterior STG. These auditory regions were used as seeds in the PPI analyses below. There were further activations in a white-matter region of the temporal lobe, in frontal regions and in the cerebellum. These clusters are not known to process auditory information. We therefore did not use them as seeds. For the visual contrast, we found activations in the visual cortex (three clusters in the right MOG and one in the left FG). These visual regions were used as seeds in the PPI analyses below. There was a further activation in the medial frontal gyrus, which is not known to process visual information. We therefore did not use this cluster as a seed. **Figure 2** illustrates the auditory

**TABLE 1 | Behavioral results in the scanner.**

	Attend audio			Attend video		
	Audio	Cong	Incong	Video	Cong	Incong
Proportion Correct (sem)	0.92 (0.05)	0.96 (0.01)	0.92 (0.03)	0.97 (0.02)	0.98 (0.01)	0.97 (0.01)
Correct response time in msec (sem)	548 (34)	554 (32)	570 (26)	578 (31)	568 (25)	564 (23)

**TABLE 2 | Audio-only > instruction results.**

Structure	Hem	<i>k</i>	<i>Z</i>	MNI coordinate			<i>p</i> <sub>unc</sub>	<i>p</i> <sub>corr</sub>
				<i>x</i>	<i>y</i>	<i>z</i>		
Superior temporal gyrus†	R	80	3.59	51	-31	10	0.0002	0.050
Superior temporal gyrus†	R	21	3.86	54	2	-8	0.0001	0.550
Superior temporal gyrus†	L	25	4.27	-60	-16	7	<10 <sup>-5</sup>	0.503
Temporal lobe (sub-gyral)	R	121	4.18	36	-55	-2	<10 <sup>-5</sup>	0.010
Medial frontal gyrus	L	258	4.31	-3	5	49	<10 <sup>-5</sup>	<10 <sup>-5</sup>
Postcentral gyrus	L	429	3.75	-39	-31	61	0.0001	<10 <sup>-5</sup>
Cerebellum	R	65	3.45	9	-55	-17	0.0003	0.083

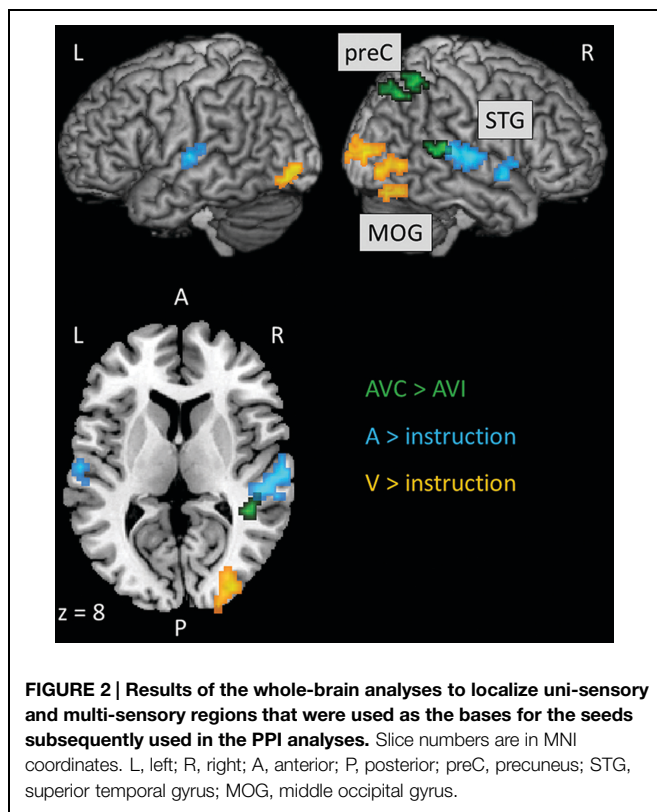
The † indicates regions used as seeds for the psychophysiological interaction analyses. *k*, cluster size; *Z*, Z-score; *p*<sub>unc</sub>, uncorrected *p*-value; *p*<sub>corr</sub>, cluster-corrected *p*-value.



**TABLE 3 | Video-only > instruction results.**

Structure	Hem	k	Z	MNI coordinate			$p_{unc}$	$p_{corr}$
				x	y	z		
Middle occipital gyrus†	R	58	4.62	33	-88	7	<10 <sup>-5</sup>	<10 <sup>-5</sup>
Middle occipital gyrus†	R	50	4.32	45	-82	-5	<10 <sup>-5</sup>	<10 <sup>-5</sup>
Inferior occipital gyrus†	R	22	3.98	42	-76	-20	<10 <sup>-5</sup>	0.017
Fusiform gyrus†	L	29	3.84	-42	-82	-8	0.0001	0.006
Medial frontal gyrus	L	20	4.59	-3	8	49	<10 <sup>-5</sup>	0.020

The † indicates regions used as seeds for the psychophysiological interaction analyses. k, cluster size; Z, Z-score;  $p_{unc}$ , uncorrected p-value;  $p_{corr}$ , cluster-corrected p-value.



and visual regions from the whole-brain analysis that were used as the bases for the seeds used in the PPI analyses.

### Multi-sensory Regions

Following previous studies (Calvert et al., 2000, 2001; Beauchamp et al., 2004; Noesselt et al., 2007, 2012), we used congruency contrasts to localize multi-sensory regions. For these contrasts, we used an initial threshold of  $p = 0.001$  and  $k = 20$ . **Table 4** presents the activations from the AVC > AVI contrast. We found one cluster in the right posterior STG and two clusters in the right parietal lobe that showed activations, which have previously been established as AV regions (e.g., Calvert et al., 2000, 2001; Beauchamp et al., 2004; Noesselt et al., 2007, 2012). We therefore used these clusters as seeds in the PPI analyses below. There was also activation in the left cingulum but this region was not used as a seed as no previous studies reported this region to be involved in

processing AV stimuli. **Figure 2** also illustrates the multi-sensory regions from the whole-brain analysis used as the bases for the seeds in the PPI analyses. There were no activations with the AVI > AVC contrast.

### fMRI PPI Results

We ran PPI analyses to test whether the functional connectivity between regions depended on whether the AV stimuli were temporally congruent (same modulation rate) or incongruent (different modulation rates). We used uni-sensory and multi-sensory regions identified in the whole-brain analyses to derive our seeds (see regions with † in **Tables 2–4**). For these contrasts, we used an initial threshold  $p = 0.005$  and  $k = 20$ . As shown in **Table 5**, the analyses identified several target regions that showed a positive change in functional connectivity with the different seeds on incongruent relative to congruent AV blocks (i.e., for the contrast AVI > AVC). **Figure 3** illustrates those target regions that were significant at the cluster-corrected level. These regions clustered in frontal and parietal cortices. There was one marginally significant target region in the STG that showed a marginally significant positive change in the functional connectivity with the right auditory seed on congruent relative to incongruent AV blocks. None of the visually localized seeds and none of the significant regions outside of the temporal lobe from the AVC > AVI contrast showed changes in functional connectivity as a function of the temporal congruence between the auditory and visual signals.

## Discussion

We used unfamiliar stimuli to investigate the role of temporal congruence in AV integration and to reveal the underlying neural mechanisms supporting integration. We manipulated temporal congruence by modulating the amplitude of a tone and the size of a 3D cuboid either at the same (congruent) or different (incongruent) amplitude-modulation rate. Here we show that both regional activations in the temporal lobe and functional connectivity between temporal, parietal and frontal regions support AV integration of continuous and unfamiliar stimuli independently of their semantic content.

Using whole-brain analyses, we localized a significant auditory region in the right temporal lobe and significant visual regions in the occipital-temporal lobe. We further found increased

**TABLE 4 | Auditory-visual congruent > Auditory-visual incongruent results (pooling over attention conditions).**

Structure	Hem	k	Z	MNI coordinate			p <sub>unc</sub>	p <sub>corr</sub>
				x	y	z		
Superior temporal gyrus†	R	20	4.09	42	-37	13	<10 <sup>-5</sup>	0.068
Precuneus†	R	20	4.21	9	-67	43	<10 <sup>-5</sup>	0.068
Intraparietal sulcus† <sup>1</sup>	R	26	4.21	27	-58	52	<10 <sup>-5</sup>	0.068
Cingulum	L	21	3.89	-15	-40	46	0.0001	0.068

The † indicates regions used as seeds for the psychophysiological interaction analyses. k, cluster size; Z, Z-score; p<sub>unc</sub>, uncorrected p-value; p<sub>corr</sub>, cluster-corrected p-value. <sup>1</sup>Although the peak voxel was located in the precuneus, most of the region was in the parietal lobe (see Figure 3).

**TABLE 5 | Psychophysiological interaction results.**

Structure	Hem	k	Z	MNI coordinate			p <sub>unc</sub>	p <sub>corr</sub>
				x	y	z		
<b>AVI &gt; AVC</b>								
<i>Auditory seed (left superior temporal gyrus, -60 -16 7, k<sub>seed</sub> = 15)</i>								
Inferior frontal gyrus	R	80	3.87	36	20	28	0.0001	0.014
Inferior frontal gyrus	R	29	3.45	51	26	10	0.0003	0.215
Middle frontal gyrus	R	65	3.47	36	44	-8	0.0003	0.022
Middle frontal gyrus	R	26	3.38	27	32	43	0.0004	0.220
Precentral gyrus	R	29	3.47	27	-19	73	0.0003	0.215
Precuneus	R	28	3.36	12	-49	37	0.0004	0.215
Cerebellum	L	22	4.08	-45	-49	-29	<10 <sup>-5</sup>	0.290
<i>Auditory seed (right superior temporal gyrus, 51 -31 10, k<sub>seed</sub> = 18)</i>								
Middle frontal gyrus	R	105	3.86	39	26	40	0.0001	0.001
Middle frontal gyrus	R	28	3.04	39	14	49	0.001	0.222
Superior frontal gyrus	R	28	3.37	24	44	34	0.0004	0.222
<i>Congruency seed (right superior temporal gyrus, 42 -37 13, k<sub>seed</sub> = 7)</i>								
Inferior frontal gyrus	R	46	3.83	51	20	10	0.0001	0.063
Medial frontal gyrus	R	361	3.92	3	32	43	<10 <sup>-5</sup>	<10 <sup>-5</sup>
Superior frontal gyrus	R	99	3.64	24	62	4	0.0001	0.002
Supramarginal gyrus	R	54	3.79	48	-49	37	0.0001	0.043
<b>AVC &gt; AVI</b>								
<i>Auditory seed (right superior temporal gyrus, 51 -31 10, k<sub>seed</sub> = 18)</i>								
Superior temporal gyrus	L	52	3.87	-39	-58	16	0.0001	0.08

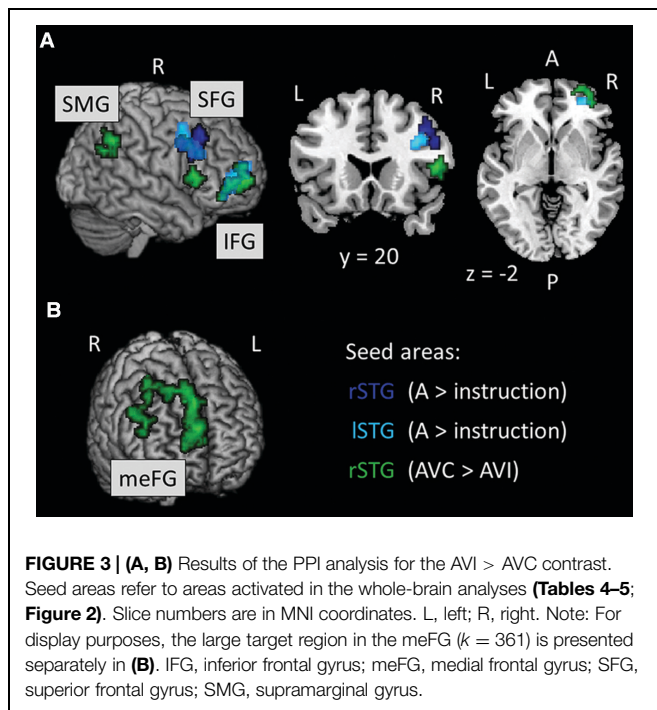
k, Cluster size; Z, Z-score; p<sub>unc</sub>, uncorrected p-value; p<sub>corr</sub>, cluster-corrected p-value; k<sub>seed</sub>, number of voxels in the seed.

activation in the right STG and the right parietal cortex when the modulation rates of the auditory and visual stimuli were temporally congruent (e.g., both modulated at 2 Hz) relative to when they were incongruent (e.g., amplitude modulation at 1 Hz and size modulation at 2 Hz). Although these multi-sensory regions are marginally significant at the cluster level ( $p = 0.068$ ), they are consistent with a large number of previous human imaging studies (e.g., Calvert et al., 2000, 2001; Beauchamp et al., 2004; Noesselt et al., 2007; Stevenson et al., 2007; Vander Wyk et al., 2010).

Importantly, we found that temporal congruence significantly modulated the functional connectivity between regions within the temporal, parietal and frontal lobes. We showed that there was an increase in functional connectivity between functionally localized auditory seed regions in the temporal lobe and frontal target regions when the auditory and visual signals had incongruent relative to congruent modulation rates. We also

found that a functionally localized multi-sensory region in the right posterior STS showed increased functional connectivity with both parietal and frontal target regions for temporally incongruent as opposed to congruent AV stimuli. Lastly, we found a marginally significant increase in functional connectivity between the auditory seed region within the right STG and a target region within the left STG with congruent compared to incongruent AV stimuli. Our connectivity results are consistent with previous work showing inter-regional interactions during AV integration across a variety of stimuli and tasks (e.g., Noesselt et al., 2007, 2012; Lewis and Noppeney, 2010; Noppeney et al., 2010; Werner and Noppeney, 2010; Lee and Noppeney, 2011; Ogawa and Macaluso, 2013; Kim et al., 2015).

We found regional interactions predominantly between bilateral regions within the anterior STS and regions within the right frontal gyrus including inferior, middle, superior and medial regions for temporally incongruent AV stimuli (see



**Table 5;** Bushara et al., 2001; Dhamala et al., 2007; Noesselt et al., 2012). Noesselt et al. (2012) recently reported greater functional connectivity between the STS and frontal regions when observers perceived AV stimuli to be asynchronous (i.e., temporally incongruent) relative to when they perceived the AV stimuli to be synchronous even though the stimuli were always physically asynchronous. In their study, Noesselt et al. (2012) used dynamic faces and voices and adjusted the stimulus onset asynchrony of facial movements and voices to produce temporally bistable percepts. They suggested that asynchronous perception is more demanding than synchronous perception as it requires the maintenance of two separate working memory representations (i.e., the auditory and visual percepts); hence the increased functional connectivity with the prefrontal cortex. In Noesselt et al.'s (2012) study, the functional connectivity was between multi-sensory regions within more posterior STS and prefrontal regions. We found that auditory regions in more anterior STS and a multi-sensory region in the posterior STS both showed increased functional connectivity with frontal regions, thereby demonstrating a large network of temporal and frontal regions (among others) in supporting AV integration. Our results further help generalize Noesselt et al.'s (2012) findings to non-ambiguous perception. The non-ambiguous nature of our stimuli may have led to the increased functional connectivity between auditory regions in the STS and frontal regions.

Noppeney et al. (2010) proposed another role for regional interactions between the STS and frontal regions. In their study, Noppeney et al. manipulated the reliability of auditory and visual information. Participants judged whether a stimulus was a tool or a musical instrument in eight different conditions derived by manipulating whether the auditory signal was intact or degraded (thereby reducing its reliability), whether the visual

signal was intact or degraded, and whether the auditory and visual signals were congruent (i.e., same category) or incongruent (i.e., different categories). The authors found that the inferior frontal sulcus (IFS) inhibited superior temporal activations for unreliable auditory input, and suggested that the IFS accumulates AV evidence by weighting its connectivity to auditory or visual cortex according to the stimulus reliability and the salience of each modality for a perceptual decision. Other researchers have proposed that the STS and frontal regions may form a network that combines sensory and semantic information and that premotor cortex in the frontal lobe may be particularly important for integrating auditory and visual information for speech and other body movements (e.g., Meyer et al., 2011; Wuerger et al., 2012). However, these latter studies did not measure connectivity between these regions.

Lastly, we found that temporal congruence did not modulate the functional connectivity between visual seed regions and any other brain regions. This modulation may not have occurred for visual regions because vision tends to be a more reliable source of sensory information than audition (Witten and Knudsen, 2005). However, in future work, it would be interesting to systematically degrade the reliability of the auditory or visual signal. With our stimuli, we can reduce the magnitude of the modulations which may be a more naturalistic method of degradation than adding noise (e.g., Stevenson et al., 2007; Stevenson and James, 2009; Noppeney et al., 2010).

Interestingly, there is evidence that frontal regions may be more involved in integrating AV communication signals (e.g., Sugihara et al., 2006) or semantic categorization (e.g., Meyer et al., 2011; Wuerger et al., 2012). Vander Wyk et al. (2010) also showed that an ellipse combined with congruent speech led to activations in frontal regions whereas a circle combined with congruent speech did not. The authors argued that the ellipse was mouth-like and therefore resembled lips more than the circle did. Further work is needed to investigate the extent to which activation in frontal regions to AV stimuli and their functional connectivity with other regions are driven by stimulus properties (e.g., familiarity or duration) as opposed to task demands and attention. Our stimuli and paradigm could be systematically manipulated (e.g., reducing the stimulus duration) to address this question (see also Vander Wyk et al., 2010).

There are two outstanding issues that we did not address in the current study. First, PPI analyses do not indicate the direction of connectivity. Future work is needed to determine whether auditory and visual information is transmitted in a bottom-up stimulus-driven manner from uni-sensory to multi-sensory and frontal regions or whether there is top-down feedback from higher to lower regions, for example, using dynamic causal modeling (e.g., Lewis and Noppeney, 2010; Werner and Noppeney, 2010; Lee and Noppeney, 2011; Ogawa and Macaluso, 2013). Second, the functional connectivity between regions within the STS and the frontal lobe may reflect neural inhibition rather than AV integration. That is, the frontal regions may help to reduce responses to the incongruent signal in the unattended modality. However, the results of Noppeney et al. (2010) and Noesselt et al. (2012) suggest that our findings are

due to AV integration (although we cannot completely rule out neural inhibition).

One advantage of our stimuli is that they capture key aspects of naturalistic stimuli such as speech yet do not carry any semantic content (see also Vander Wyk et al., 2010). We are also able to manipulate the auditory and visual signals in comparable ways (i.e., modulation of the amplitude or size). With our current stimuli, there is some degree of correlation even when the auditory and visual signals have different modulation rates because the “fast” modulation rate (2 Hz) is a harmonic of the “slow” one (1 Hz) and close in value (see **Figure 1**). However, in a separate study using these stimuli, we found that the AV congruent stimulus affected performance on an amplitude-modulation discrimination task, but not the AVI stimuli (Vuong et al., 2014). This finding suggests that observers’ were sensitive to the difference in temporal congruence between the two types of AV stimuli. It would be interesting in future work to more systematically manipulate the frequency difference and the harmonicity between the modulation rates.

## Conclusion

In summary, using amplitude-modulated tones and size-modulated shapes, our functional imaging study revealed the

importance of both regional activation and inter-regional connectivity in AV integration across a network of temporal, parietal, and frontal regions. Supporting our findings, diffusion imaging data in humans suggest that there are anatomical connections between some of these regions (Beer et al., 2011, 2013; van den Brink et al., 2014). Moreover, recent studies in non-human primates suggest that there are also effective functional (Petkov et al., 2015) and anatomical (Yeterian et al., 2012) connections between the STS and frontal regions. Compared to congruent stimuli, temporally incongruent stimuli elicited increased functional connectivity between auditory and multi-sensory regions in the STS and prefrontal regions. Importantly, these physiological changes were obtained using continuously varying non-meaningful stimuli. The AV interactions observed in this study are not confounded by semantic content, and therefore they provide an important link between transient, non-meaningful stimuli and continuous real-world objects, speech and music.

## Acknowledgments

ML was supported by a Wellcome Trust studentship [102558/Z/13/Z]. We would like to thank Chris Petkov for the use of the NordicNeuroLab audio system, and Tim Hodgson, Louise Ward and Dorothy Wallace for help with scanning.

## References

- Beauchamp, M. (2005). Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics* 3, 93–113.
- Beauchamp, M. S., Lee, K. E., Argall, B. D., and Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41, 809–823. doi: 10.1016/S0896-6273(04)00070-4
- Beer, A. L., Plank, T., and Greenlee, M. W. (2011). Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex. *Exp. Brain Res.* 213, 299–308. doi: 10.1007/s00221-011-2715-y
- Beer, A. L., Plank, T., Meyer, G., and Greenlee, M. W. (2013). Combined diffusion-weighted and functional magnetic resonance imaging reveals a temporal-occipital network involved in auditory-visual object processing. *Front. Integr. Neurosci.* 7:5. doi: 10.3389/fnint.2013.00005
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357
- Bushara, K. O., Grafman, J., and Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *J. Neurosci.* 21, 300–304.
- Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., and David, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport* 10, 2619–2623. doi: 10.1097/00001756-199908200-00033
- Calvert, G. A., Campbell, R., and Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657. doi: 10.1016/S0960-9822(00)00513-3
- Calvert, G. A., Hansen, P. C., Iversen, S. D., and Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage* 14, 427–438. doi: 10.1006/nimg.2001.0812
- Dhamala, M., Assisi, C. G., Jirsa, V. K., Steinberg, F. L., and Kelso, J. A. (2007). Multisensory integration for timing engages different brain networks. *Neuroimage* 34, 764–773. doi: 10.1016/j.neuroimage.2006.07.044
- Diedrich, A., and Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Attent. Percept. Psychophys.* 66, 1388–1404. doi: 10.3758/BF03195006
- Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1
- Frassinetti, F., Bolognini, N., and Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp. Brain Res.* 147, 332–343. doi: 10.1007/s00221-002-1262-y
- Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., and Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6, 218–229. doi: 10.1006/nimg.1997.0291
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., and Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210. doi: 10.1002/hbm.460020402
- Gitelman, D. R., Penny, W. D., Ashburner, J., and Friston, K. J. (2003). Modeling regional and psychophysiological interactions in fMRI: the importance of hemodynamic deconvolution. *Neuroimage* 19, 200–207. doi: 10.1016/S1053-8119(03)00058-2
- Grant, K. W., and Seitz, P. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *J. Acoust. Soc. Am.* 108, 1197–1208. doi: 10.1121/1.1288668
- Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., et al. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol.* 6:e159. doi: 10.1371/journal.pbio.0060159
- Hocking, J., and Price, C. J. (2008). The role of posterior superior temporal sulcus in audiovisual processing. *Cereb. Cortex* 18, 2439–2449. doi: 10.1093/cercor/bhn007
- Kim, H., Hahn, J., Lee, H., Kang, E., Kang, H., and Lee, D. S. (2015). Brain networks engaged in audiovisual integration during speech perception revealed by persistent homology-based network filtration. *Brain Connect.* 5, 245–258. doi: 10.1089/brain.2013.0218
- Kleiner, M., Brainard, D., and Pelli, D. (2007). What’s new in Psychtoolbox-3? *Perception (ECVP Abstr. Suppl.)* 36, 1.
- Lee, H., and Noppeney, U. (2011). Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proc. Natl. Acad. Sci. U.S.A.* 108, E1441–E1450. doi: 10.1073/pnas.1115267108



- Lewis, R., and Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *J. Neurosci.* 30, 12329–12339. doi: 10.1523/JNEUROSCI.5745-09.2010
- Lovelace, C. T., Stein, B. E., and Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Cogn. Brain Res.* 17, 447–453. doi: 10.1016/S0926-6410(03)00160-5
- Maddox, R. K., Atilgan, H., Bizley, J. K., and Lee, A. K. (2015). Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners. *eLife* 4, 1–11. doi: 10.7554/eLife.04995
- Maldjian, J. A., Laurienti, P. J., Kraft, R. A., and Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19, 1233–1239. doi: 10.1016/S1053-8119(03)00169-1
- McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0
- McLaren, D. G., Ries, M. L., Xu, G., and Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *Neuroimage* 61, 1277–1286. doi: 10.1016/j.neuroimage.2012.03.068
- Meredith, M. A., and Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science* 221, 389–391. doi: 10.1126/science.6867718
- Meyer, G. F., Greenlee, M., and Wuerger, S. (2011). Interaction between auditory and visual semantic stimulus classes: evidence for common processing networks for speech and body actions. *J. Cogn. Neurosci.* 23, 2291–2308. doi: 10.1162/jocn.2010.21593
- Naumer, M. J., van den Bosch, J. J. F., Wibrall, M., Kohler, A., Singer, W., Kaiser, J., et al. (2011). Investigating human audio-visual object perception with a combination of hypothesis-generating and hypothesis-testing fMRI analysis tools. *Exp. Brain Res.* 213, 309–320. doi: 10.1007/s00221-011-2669-0
- Noesselt, T., Bergmann, D., Heinze, H.-J., Münte, T., and Spence, C. (2012). Coding of multisensory temporal patterns in human superior temporal sulcus. *Front. Integr. Neurosci.* 6:64. doi: 10.3389/fnint.2012.00064
- Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H.-J., et al. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J. Neurosci.* 27, 11431–11441. doi: 10.1523/JNEUROSCI.2252-07.2007
- Noppeney, U., Ostwald, D., and Werner, S. (2010). Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. *J. Neurosci.* 30, 7434–7446. doi: 10.1523/JNEUROSCI.0455-10.2010
- Nozawa, G., Reuter-Lorenz, P., and Hughes, H. (1994). Parallel and serial processes in the human oculomotor system: bimodal integration and express saccades. *Biol. Cybern.* 72, 19–34. doi: 10.1007/BF00206235
- Ogawa, A., and Macaluso, E. (2013). Audio-visual interactions for motion perception in depth modulate activity in visual area V3A. *Neuroimage* 71, 158–167. doi: 10.1016/j.neuroimage.2013.01.012
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366
- Petkov, C. I., Kikuchi, Y., Milne, A. E., Mishkin, M., Rauschecker, J. P., and Logothetis, N. (2015). Different forms of effective connectivity in primate frontotemporal pathways. *Nat. Commun.* 6, 6000. doi: 10.1038/ncomms7000
- Plomp, R. (1983). “The role of modulation in hearing,” in *Hearing: Physiological Bases and Psychophysics*, eds R. Klinke and R. Hartmann (Berlin: Springer-Verlag), 270–276.
- Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 336, 367–373. doi: 10.1098/rstb.1992.0070
- Sadaghiani, S., Maier, J. X., and Noppeney, U. (2009). Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. *J. Neurosci.* 29, 6490–6499. doi: 10.1523/JNEUROSCI.5437-08.2009
- Sekuler, R., Sekuler, A. B., and Lau, R. (1997). Sound alters visual motion perception. *Nature* 385, 308. doi: 10.1038/385308a0
- Shams, L., Kamitani, Y., and Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature* 408, 788. doi: 10.1038/35048669
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science (N.Y.)* 270, 303–304. doi: 10.1126/science.270.5234.303
- Stein, B. E., and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.* 9, 255–266. doi: 10.1038/nrn2331
- Stevenson, R. A., Geoghegan, M. L., and James, T. W. (2007). Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Exp. Brain Res.* 179, 85–95. doi: 10.1007/s00221-006-0770-6
- Stevenson, R. A., and James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage* 44, 1210–1223. doi: 10.1016/j.neuroimage.2008.09.034
- Sugihara, T., Diltz, M. D., Averbeck, B. B., and Romanski, L. M. (2006). Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J. Neurosci.* 26, 11138–11147. doi: 10.1523/JNEUROSCI.3550-06.2006
- Sumbly, W., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215. doi: 10.1121/1.1907309
- van den Brink, R. L., Cohen, M. X., Van Der Burg, E., Talsma, D., Vissers, M. E., and Slagter, H. A. (2014). Subcortical, modality-specific pathways contribute to multisensory processing in humans. *Cereb. Cortex* 24, 2169–2177. doi: 10.1093/cercor/bht069
- Vander Wyk, B. C., Ramsay, G. J., Hudac, C. M., Jones, W., Lin, D., Klin, A., et al. (2010). Cortical integration of audio-visual speech and non-speech stimuli. *Brain Cogn.* 74, 97–106. doi: 10.1016/j.bandc.2010.07.002
- Vuong, Q. C., Laing, M., and Rees, A. (2014). Auditory-visual interactions in brain and behaviour using modulated stimuli. *Abstr. Present. Soc. Neurosci.*
- Werner, S., and Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *J. Neurosci.* 30, 2662–2675. doi: 10.1523/JNEUROSCI.5091-09.2010
- Witten, I. B., and Knudsen, E. I. (2005). Why seeing is believing: merging auditory and visual worlds. *Neuron* 48, 489–496. doi: 10.1016/j.neuron.2005.10.020
- Wuerger, S. M., Parkes, L., Lewis, P. A., Crocker-Buque, A., Rutschmann, R., and Meyer, G. F. (2012). Premotor cortex is sensitive to auditory-visual congruence for biological motion. *J. Cogn. Neurosci.* 24, 575–587. doi: 10.1162/jocn\_a\_00173
- Yeterian, E. H., Pandya, D. N., Tomaiuolo, F., and Petrides, M. (2012). The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex* 48, 68–81. doi: 10.1016/j.cortex.2011.03.004

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Laing, Rees and Vuong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Feature activation during word recognition: action, visual, and associative-semantic priming effects

Kevin J. Y. Lam<sup>1,2\*</sup>, Ton Dijkstra<sup>1</sup> and Shirley-Ann Rueschemeyer<sup>3</sup>

<sup>1</sup> Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands, <sup>2</sup> International Max Planck Research School for Language Sciences, Nijmegen, Netherlands, <sup>3</sup> Department of Psychology, University of York, York, UK

## OPEN ACCESS

### Edited by:

Andriy Myachykov,  
Northumbria University, UK

### Reviewed by:

Olaf Hauk,  
Medical Research Council, Cognition  
and Brain Sciences Unit, UK  
Bo Yao,  
University of Manchester, UK

### \*Correspondence:

Kevin J. Y. Lam,  
Postbus 9104, 6500 HE, Nijmegen,  
Netherlands  
K.Lam@donders.ru.nl

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 28 November 2014

**Accepted:** 05 May 2015

**Published:** 27 May 2015

### Citation:

Lam KJY, Dijkstra T and  
Rueschemeyer S-A (2015) Feature  
activation during word recognition:  
action, visual, and  
associative-semantic priming effects.  
*Front. Psychol.* 6:659.  
doi: 10.3389/fpsyg.2015.00659

Embodied theories of language postulate that language meaning is stored in modality-specific brain areas generally involved in perception and action in the real world. However, the temporal dynamics of the interaction between modality-specific information and lexical-semantic processing remain unclear. We investigated the relative timing at which two types of modality-specific information (action-based and visual-form information) contribute to lexical-semantic comprehension. To this end, we applied a behavioral priming paradigm in which prime and target words were related with respect to (1) action features, (2) visual features, or (3) semantically associative information. Using a Go/No-Go lexical decision task, priming effects were measured across four different inter-stimulus intervals (ISI = 100, 250, 400, and 1000 ms) to determine the relative time course of the different features. Notably, action priming effects were found in ISIs of 100, 250, and 1000 ms whereas a visual priming effect was seen only in the ISI of 1000 ms. Importantly, our data suggest that features follow different time courses of activation during word recognition. In this regard, feature activation is dynamic, measurable in specific time windows but not in others. Thus the current study (1) demonstrates how multiple ISIs can be used within an experiment to help chart the time course of feature activation and (2) provides new evidence for embodied theories of language.

**Keywords:** embodied language comprehension, feature activation, semantic priming, action priming, visual priming

## Introduction

One of the oldest issues in cognitive psychology concerns the mental representation of meaning. In the past decade, embodied theories of language, postulating that language meaning is stored in modality-specific brain areas, have gained in popularity and empirical support. For example, the meaning of the word “grasp” activates some of the neural areas involved in planning and performing everyday grasping actions (e.g., Hauk et al., 2004; Rueschemeyer et al., 2007), while comprehension of the word “red” entails activation of parts of the neural visual pathway (e.g., Simmons et al., 2007; van Dam et al., 2012). Nevertheless, despite much research important questions remain unanswered. One of these is when, and to what end, modality-specific information becomes activated during language comprehension.

In line with a general embodied framework, a number of behavioral studies have demonstrated that words with shared perceptual features prime each other. This indicates that the physical properties of an object in the real world influence how the word denoting the object is processed. For example, words referring to objects with similar shapes, such as *pizza* and *coin*, prime each other (Schreuder et al., 1984; Pecher et al., 1998), as do words referring to objects with shared manipulation features such as *typewriter* and *piano* (Myung et al., 2006). Note that in both of these examples, participants showed priming (which is interpreted as facilitation of processing) for words with shared perceptual or action features, even in the absence of any obvious conceptual or semantic relationship (see McNamara, 2005 for an in-depth treatment).

Neuroimaging evidence such as those reported by Kiefer et al. (2011) using EEG and fMRI further substantiate the results above. Participants saw a prime presented either as a word or as a picture followed by a target picture and were asked to name both stimuli when cued. Pairs were either congruent (*pliers–nutcracker*) or incongruent (*pliers–horseshoe*) with regard to the implied action. Most notably, pictures primes elicited early (N1) and late (N400) priming effects; word primes, by contrast, showed effects only later in the N400 component. The authors interpreted the finding as evidence of two stages of priming effects: fast and slow activation of action features with pictures, but slow activation with words. Specifically, the authors argued that pictures make certain features more salient, therefore activating more detailed representations which may also lead to earlier activation. Word stimuli appear less suitable to generate early action priming effects, at least when manipulations of congruency are employed to induce priming effects.

Other experimental methods have also been used to test for the activation of visual information (i.e., information about the visual form of an object in the real world) in conjunction with word processing. In an eye-tracking study (Yee et al., 2011), participants heard a spoken word and saw pictures of four objects on a screen in a given trial. In this visual world paradigm, participants identified the picture that best matched the spoken word. Notably, participants spent significantly longer looking at distractor items with a visual form matching that of the object denoted by the spoken word. For example, when participants heard “*frisbee*”, they looked significantly longer at a picture of a pizza (both objects are round) than to a linguistically matched control with no shared perceptual features (e.g., a thimble). Interestingly, this effect appeared only if participants had a relatively short amount of time to explore the visual scene (1000 ms); the effect was not present when the visual scene was presented for a longer time period (2000 ms). The authors argue that effects of visual form are seen early in word and object identification but decay over time.

Altogether, the reviewed studies suggest that different aspects of a word’s referent (i.e., the object’s features) can be independently activated, as evidenced by action and visual priming effects. Nonetheless, studies make the implicit assumption that feature activation is constant and stable over time, as is evident from their use of a single inter-stimulus interval (ISI) or stimulus-onset asynchrony (SOA) value in most

(priming) experiments. Interestingly, in Yee et al. (2011), the authors reported a possible decay of visual features over time which was determined by comparing two ISIs. In Kiefer et al. (2011), the authors proposed fast and slow feature activation as a function of stimuli type, but it is unclear if the same conclusion will hold if at least another ISI was tested for a comparison. Still, unlike the commonly held assumption, the authors of both studies assume that the time course of feature activation is dynamic.

This focus on the timing aspect is non-trivial, especially when considered alongside the discussion of embodied theories of language. Some authors have demonstrated very fast (160–250 ms after word onset) and automatic (activation even when attention is diverted away) activation of sensorimotor information, taking this as evidence for the integral role of such information in the word’s representation (Pulvermüller et al., 2001; Hauk and Pulvermüller, 2004; for a review, see Pulvermüller, 2005). However, other researchers have come to slightly different positions with respect to timing. In the Language as Situated Simulation (LASS) model, Barsalou and colleagues have claimed that perceptual and action information, while being an integral part of conceptual knowledge, are activated relatively late during language comprehension (Solomon and Barsalou, 2004; Barsalou, 2008; Barsalou et al., 2008). The slower reaction times for property verification judgements reflect the need for participants to activate deep, perceptually-based conceptual knowledge, because quickly accessed language-based relationships will not suffice to perform the task.

The Symbol Interdependency Hypothesis (SIH; Louwerse, 2011), too, claims that perceptual simulations play a greater role later on and reflect more detailed representations but, unlike the LASS model (Barsalou, 2008), it emphasizes the symbolic (linguistic) rather than the embodied (modality-specific) aspect of linguistic processing. For example, Louwerse showed that the results of a previous iconicity study (Zwaan and Yaxley, 2003) which was interpreted as support for embodied representations, could actually be accounted for predominantly by linguistic frequency. In other words, SIH argues that perceptual simulations can be traced back to language itself. The symbolic aspect serves to create underspecified representations quickly for good-enough comprehension whereas the embodied aspect goes further when full and deep comprehension is needed by relying on embodied relations already encoded into language. In sum, LASS and SIH make similar proposals; only the relative importance of each component differs whereby the task dictates which component is more or less relevant. Both theories claim that perceptual features are time-consuming and resource-hungry.

In the current behavioral priming experiment, we aim to disentangle some of the issues surrounding the time course along which different types of perceptual and motor information are activated during word comprehension. Firstly, we focus on action and visual features because previous research has shown the two features to be highly relevant in word processing. More importantly, action and visual features have not been directly compared within one study to determine whether they each have unique time courses. Furthermore, other studies

(e.g., Wheatley et al., 2005) that suggest relative importance of different features for the conceptual representations of objects are another source for our hypothesis. Based on previous literature (e.g., Schreuder et al., 1984; Pecher et al., 1998; Myung et al., 2006), we hypothesize that both feature-based action and visual priming should show effects similar in direction to those seen for associative semantic priming. That is, word pairs related along action and visual features should show facilitation of reaction times.

Secondly, we use the Go/No-Go task in combination with lexical decision. In the current experiment, participants are instructed to respond with a button press when a stimulus pair consists of words (“Go”); otherwise, they did not need to respond (“No-Go”). Notably, some authors have used the task to examine time-course related issues in language processing (e.g., van Turennout et al., 1997, 1998, 1999). Nevertheless, we included associative-semantic related stimuli to elicit standard semantic priming effects as a verification measure (see Gomez et al., 2007 for a comparison of the Go/No-Go and standard two-choice tasks).

Thirdly, we systematically vary the time between presentation of the prime word and the target word; that is, the ISI, which is the interval between the offset of the prime and the onset of the target. Participants in the pilot phase reported that they could not always identify the prime and target stimuli if a presentation duration of less than 400 ms was used (mean word length no less than 9.5 letters; see Supplementary Material for complete listing). Consequently, we used a fixed prime and target duration of 400 ms and varied the ISIs accordingly. The ISI factor is thus a manipulation of preview time between prime and target word presentation to determine the relative timing of and processing differences between different features. We assume that activation is a dynamic process, thus there is likely no single ISI value that can capture all features; the use of multiple ISI values therefore is intended to sample feature activation over time (see Moss et al., 1995; Hauk et al., 2012 for similar arguments). Previous relevant priming studies (Myung et al., 2006; Kiefer et al., 2011) have used ISIs of 50 and 70 ms, with SOAs ranging between 370 and 1250 ms. In line with those earlier studies, we employed three ISIs in 150 ms-increments: 100, 250, and 400 ms (corresponding to SOAs of 500, 650, and 800 ms, respectively). The fourth ISI of 1000 ms (equal to an SOA of 1400 ms) serves as a long interval in which we expect the greatest modulation of effects to occur.

In summary, we investigate priming in three distinct conditions: (1) associative semantic priming (e.g., *bolt-screwdriver*), (2) feature-based action priming (e.g., *housekey-screwdriver*), and (3) feature-based visual form priming (e.g., *soldering iron-screwdriver*). By including three priming conditions within one experimental design, we investigate whether feature-based action and visual priming produce effects directly comparable with associative semantic priming. More importantly, by looking at priming at four ISIs we assess how long after presentation of a prime word, specific types of information become available in order to affect comprehension of the target word. In this manner, we can draw conclusions about the relative timing of different types of feature-based semantic knowledge. Different embodied theories

of language predict such effects at different time intervals: strong embodied theories (e.g., Pulvermüller, 2005; Pulvermüller and Fadiga, 2010) predict effects in the early phase, but moderate and disembodied theories (e.g., Mahon and Caramazza, 2008) in the late phase. Hybrid theories such as LASS (e.g., Barsalou, 2008) and SIH (Louwerse, 2011) allow the involvement of both language-based and perceptual-based information, with more or less emphasis on either depending on the task. The current study will provide detailed timing information to help adjudicate between the competing theories.

## Materials and Methods

### Participants

One hundred and seventy-six right-handed native German speakers aged 18–25 years (136 females; mean age = 21 years) with normal or corrected-to-normal vision were recruited within the Radboud University Nijmegen. Participants were assigned to one of the four inter-stimulus interval (ISI) groups, consisting of 44 participants each. Participants gave informed consent and were offered course credit or monetary compensation. This study was approved by the local Nijmegen Ethical Committee of the Faculty of Social Sciences (ECG2012-2711-05).

### Stimulus Materials

German words denoting familiar tools or manipulable objects were used either as prime or target words. Each of the 24 target words was paired with four prime words corresponding to the four prime conditions (see sample stimuli in **Table 1**; full stimulus materials in Supplementary Material): (1) semantically related, (2) action-related, (3) visual-related, and (4) unrelated. In the semantically related condition, the prime and target pair denoted related objects by association (e.g., *bolt-screwdriver*) and had no action and visual relatedness. In the action-related condition, the prime and target pair denoted objects that are used in a similar manner but do not have any semantic or visual relatedness (e.g., *housekey-screwdriver*). Also, all actions implied by these objects are restricted to the hands or arms. In the visual-related condition, the prime and target pair denoted objects similar in form or appearance but did not share any semantic or action relatedness (e.g., *soldering iron-screwdriver*). Finally, the prime and target pair in the unrelated condition denoted objects that shared none of the above relationships (e.g., *charger-screwdriver*).

**TABLE 1 | Sample primes from the four conditions paired with the same target in German with their corresponding English translations.**

	Prime word	Target word
Unrelated	Ladegerät “charger”	Schraubenzieher “screwdriver”
Semantic	Dübel “bolt”	
Action	Haustürschlüssel “housekey”	
Visual	LötKolben “soldering iron”	

See supplementary materials for full set.



A norming study using a new selection of participants ( $n = 10$ ) confirmed our manipulations (see Supplementary Material). For all comparisons of interest, words were matched for length and frequency (see Supplementary Material) using the SUBTLEX-DE database (Brysbaert et al., 2011). Also, 24 pseudowords were added from a pseudoword generator (Keuleers and Brysbaert, 2010) to serve as catch trials in the Go/No-Go lexical decision task, described below.

## Design

Participants were presented with a total of 140 trials: 96 critical trials containing 24 target words paired with four different prime words, 24 catch trials containing one or two pseudowords, and another 20 filler trials similar to critical and catch trials. The trials were divided into four blocks of 35 trials each, with five dummy trials at the beginning of each block. Crucially, target words appeared only once per block and lists were pseudo-randomized to ensure that no more than three consecutive trials were from the same condition. In result, four lists were generated and one version was randomly assigned to each participant.

## Procedure

Participants sat approximately 80 cm in front of the computer screen. Button presses were recorded from a response box. The start of a trial was indicated by an asterisk positioned at the center for 2000 ms. Next, prime and target words were each presented for 400 ms; the interval of the intervening blank screen—the inter-stimulus interval (ISI)—was 100, 250, 400, or 1000 ms. A black blank screen was presented for an inter-trial interval of 2000 ms.

Participants were instructed to press the response button with their right index finger whenever a trial consisted only of German words (i.e., both prime and target words). Otherwise, they were instructed to withhold their response—thus, catch trials (containing pseudowords) did not require a button press. A short break was given between blocks of trials. Participants were first presented with a practice block of 12 trials that did not contain any critical stimuli but reflected the experimental conditions. In total, each version of the experiment lasted about 20 min.

## Results

Participants were excluded if (1) their overall mean reaction times (RTs) exceeded 800 ms, and if (2) the  $d$ -prime scores of at least three conditions were less than 2.9 out of a maximal possible score of 4.7. Of the remaining data, we excluded incorrect trials and trials containing RTs faster than 250 ms and slower than 1800 ms, as well as those slower than 2.5 standard deviations of a participant's mean. This resulted in the removal of 3% trials. Priming scores were calculated by subtracting each of the three conditions (Semantic, Action, Visual) from the Unrelated condition.

For the  $F_1$  analyses, subject-based means were then submitted to a Two-Way Condition (Semantic, Action, Visual)  $\times$  ISI (100, 250, 400, 1000-ms) ANOVA with Condition as a within-subject variable and ISI as a between-subject variable. For the  $F_2$  analyses,

item-based means were submitted to a Two-Way Condition  $\times$  ISI ANOVA with Condition and ISI both as within-subject variables. We also report complementary  $F_1$  and  $F_2$  analyses using only Action and Visual for the Condition factor to verify that the two main effects of interest indeed differ in time course. We report Greenhouse-Geisser corrected  $p$ -values whenever the sphericity assumption is violated.

Within each ISI group, paired samples  $t$ -tests were conducted for the three critical pairwise comparisons. All  $p$ -values resulting from the  $t$ -tests have been controlled for multiple comparisons using the Benjamini and Hochberg False Discovery Rate (FDR) procedure (Benjamini and Hochberg, 1995). Effect sizes reported reflect Cohen's  $d$  using pooled variance. See Table 2 for an overview of mean RTs.

## Interactions and Main Effects

A summary of the ANOVA analyses is shown in Table 3. Table 4 lists the priming scores of the three conditions; asterisks denote significant effects at FDR-corrected  $p$ -values  $< 0.05$ . The presence of a (nearly) significant interaction between Condition and ISI allowed us to consider the effects in each of the ISIs separately.

### ISI = 100 ms

A statistically reliable semantic priming effect was present at this ISI: Mean RTs were faster to semantically related target words (525 ms) than to unrelated target words (547 ms),  $t_{(33)} = 5.33$ ,  $p < 0.01$ ,  $d = 0.253$ . An action priming effect was also statistically reliable: Mean RTs were faster to action-related target words (536 ms) relative to unrelated target words,  $t_{(33)} = 2.00$ ,  $p < 0.05$ ,  $d = 0.143$ . There was no statistically reliable visual priming effect, however. Mean RTs to visual-related target words (552 ms) were not distinguishable from those to unrelated target words,  $t_{(33)} = -0.91$ ,  $p = 0.19$ ,  $d = 0.060$ .

**TABLE 2 | The sample size, mean reaction times, and standard deviation values (within parentheses) of the four conditions across the four inter-stimulus interval manipulations.**

	100 ms $n = 34$	250 ms $n = 35$	400 ms $n = 41$	1000 ms $n = 37$
Unrelated	547 (87)	560 (71)	558 (87)	553 (67)
Semantic	525 (84)	543 (68)	542 (87)	534 (67)
Action	536 (69)	550 (70)	554 (84)	542 (70)
Visual	552 (95)	554 (72)	553 (83)	543 (70)

**TABLE 3 | The ANOVA summary of  $F_1$  and  $F_2$  results using all three priming conditions and only the two main conditions of interest.**

	Condition (semantic, action, visual)	Condition (action, visual)
Condition	$F_{1(5, 691, 271, 255)} = 2.079$ ; $p = 0.059$	$F_{1(3, 143)} = 2.309$ ; $p = 0.079$
$\times$ ISI	$F_{2(6, 138)} = 2.667$ ; $p = 0.018$	$F_{2(3, 69)} = 3.321$ ; $p = 0.025$
Condition	$F_{1(2, 286)} = 19.261$ ; $p = 0.000$	$F_{1(1, 143)} = 3.794$ ; $p = 0.053$
	$F_{2(2, 46)} = 5.927$ ; $p = 0.005$	$F_{2(1, 23)} = 0.776$ ; $p = 0.388$
ISI	$F_{1(3, 143)} = 0.366$ ; $p = 0.778$	$F_{1(3, 143)} = 0.684$ ; $p = 0.563$
	$F_{2(3, 69)} = 0.187$ ; $p = 0.905$	$F_{2(3, 69)} = 0.515$ ; $p = 0.673$

**TABLE 4 | Priming scores (in ms) of the three conditions and standard error of differences values (within parentheses) across the four inter-stimulus interval manipulations.**

	Semantic	Action	Visual
100 ms	22 (4.1)*	11 (5.6)*	-5 (5.9)
250 ms	17 (4.5)*	9 (4.4)*	6 (4.8)
400 ms	16 (4.4)*	4 (4.7)	5 (4.2)
1000 ms	19 (4.2)*	12 (4.8)*	10 (4.7)*

Asterisks denote significant effects at FDR-corrected  $p$ -values  $< 0.05$ .

### ISI = 250 ms

A similar pattern of results as for ISI = 100 ms was found. A statistically reliable semantic priming effect was present: Mean RTs were significantly faster to semantically related target words (543 ms) than to unrelated target words (560 ms),  $t_{(34)} = 3.75$ ,  $p < 0.01$ ,  $d = 0.242$ . A statistically significant action priming effect indicated that mean RTs were significantly faster to action-related target words (550 ms) than to unrelated target words,  $t_{(34)} = 2.09$ ,  $p < 0.05$ ,  $d = 0.132$ . However, the visual priming effect was not statistically reliable: Mean RTs were not significantly faster to visual-related target words (554 ms) than to unrelated target words,  $t_{(34)} = 1.25$ ,  $p = 0.12$ ,  $d = 0.086$ .

### ISI = 400 ms

Only a semantic priming effect was obtained: Mean RTs were significantly faster to semantically related target words (542 ms) than to unrelated target words (558 ms),  $t_{(40)} = 3.51$ ,  $p < 0.05$ ,  $d = 0.178$ . Action and visual priming effects, however, were not statistically reliable. Mean RTs were not significantly faster to action-related target words (554 ms) than to unrelated target words,  $t_{(40)} = 0.81$ ,  $p = 0.22$ ,  $d = 0.045$ . Similarly, mean RTs were not significantly faster to visual-related target words (553 ms) than to unrelated target words,  $t_{(40)} = 1.23$ ,  $p = 0.17$ ,  $d = 0.061$ .

### ISI = 1000 ms

All three priming effects were statistically significant. Mean RTs were significantly faster to semantically related target words (534 ms) than to unrelated target words (553 ms),  $t_{(36)} = 4.55$ ,  $p < 0.01$ ,  $d = 0.283$ . Mean RTs were faster to action-related target words (542 ms) than to unrelated target words,  $t_{(36)} = 2.39$ ,  $p < 0.05$ ,  $d = 0.169$ . Finally, mean RTs were faster to visual-related target words (543 ms) than to unrelated target words,  $t_{(36)} = 2.14$ ,  $p < 0.05$ ,  $d = 0.146$ .

## Discussion

In this study, we investigated the time course of activation for different modality-specific features using a Go/No-Go priming paradigm with varying inter-stimulus intervals (ISIs). Four groups of participants performed lexical decisions to word pairs from three priming conditions: (1) associative semantic priming (e.g., *bolt-screwdriver*), (2) feature-based action priming (e.g., *housekey-screwdriver*), (3) feature-based visual priming (e.g., *soldering iron-screwdriver*), and we compared these to a fourth

unrelated condition (e.g., *charger-screwdriver*). By varying the amount of time between presentation of the prime word and of the target word (i.e., ISI), we assessed how soon the activation of semantically relevant (i.e., feature-based) information became effectively available after prime word presentation.

Our results show that feature-based information present in the prime word facilitates recognition of subsequent target words (i.e., priming takes place). Importantly, the relative timing at which feature-based information becomes activated varies between modalities. Feature-based action relationships elicited priming effects at ISIs of 100, 250, and 1000 ms. Feature-based visual relationships, by contrast, elicited priming effects only at ISI of 1000 ms. Unlike both feature-based relationships, associative semantic relationships elicited consistent priming effects across all four ISIs.

In the following, we will first discuss the time course of activation of semantic, action, and visual features individually. As noted in the introduction, by varying the ISI (preview time), we can determine the relative timing of and processing differences between different features. We will argue that the finding of different time course of activation for different modality-specific features requires a reassessment of current opposing views on embodied representations, moving to views that highlight the flexible recruitment of feature activations (e.g., Hoening et al., 2008; Kiefer and Pulvermüller, 2012) and a combination of amodal and embodied representations (e.g., Barsalou, 2008; Louwerse, 2011).

### Associative Semantic Priming Effects are Activated at all ISIs

We observed associative semantic priming effects at all four ISIs. These effects show that the experiment is sensitive to our manipulations and able to elicit priming effects at all four intervals tested. The findings agree with the literature on semantic priming wherein reports of semantic priming effects have been shown using very short and very long ISIs (e.g., Perea and Gotor, 1997; Hutchison et al., 2001; Perea and Rosa, 2002; Chiarello et al., 2003; see Hutchison, 2003 for a review).

### Different Time Courses of Activation: Action Precedes Visual Feature Activation

The results show that words referring to manipulable objects can indeed elicit action priming effects, as reported in the object representation literature (e.g., Ellis and Tucker, 2000; for a review, see Martin, 2007). In a similar action priming study (Myung et al., 2006 Experiment 1), participant made lexical decisions to primes and targets (e.g., piano-typewriter) presented over headphones. Another study (Kiefer et al., 2011) showed that picture targets preceded by word primes elicited effects relatively late in processing, namely in the N400 time window. By contrast, picture targets preceded by picture primes showed effects sooner in the N1 time window. Kiefer and colleagues argue that pictorial stimuli make certain features more salient, thus generating more detailed representations. However, retrieving more detailed representations does not necessarily lead to activation of a feature earlier in time, because such retrieval may be more time-consuming and effortful. Regardless, our results demonstrate that

visually presented word pairs can elicit action priming effects in time windows subsequent to the N400.

We also observed priming effects of visually related word pairs in the longest ISI of 1000 ms. Unlike action features, visual features do not appear to be activated as quickly as action features. Seen alongside the semantic and action priming results, this suggests that different features may have different activation profiles.

Certain visual features may be particularly difficult to elicit using word stimuli. Using pairs of perceptually related stimuli which shared shape or color features, Schreuder and colleagues (Schreuder et al., 1984; Flores d'Arcais et al., 1985) reported priming effects using the lexical decision task. Subsequent studies, however, failed to replicate these effects unless these features were made explicit for the task, such as the use of a preceding activation task (Pecher et al., 1998; stimulus-onset asynchrony, SOA = 350 ms, ISI = 50 ms).

A possible clarifying factor is that the perceptual priming effect in Schreuder et al. (1984; also see Flores d'Arcais et al., 1985) is not strictly visual priming in the sense used here and elsewhere (e.g., Kellenbach et al., 2000). Their perceptual condition was composed of visually–(primarily) and color-related stimuli. Though color-related items made up a small part of the stimuli, the effects may have largely originated from these items. Color has been shown to be a prominent component of an object's representation, more so than action features for certain classes of object nouns (e.g., van Dam et al., 2012). Similarly, the perceptual stimuli used in Pecher et al. (1998) differ from our stimuli in that they consisted of nouns referring to a range of categories like food, body part, animals, etc., and could thus have confounded the results.

Using pictorial stimuli as targets, a recent study has indeed reported early visual effects (Yee et al., 2011) but, as is the case in the Kiefer et al. (2011) study with action features and pictorial targets, these early effects may appear sooner when pictorial stimuli are used. There is suggestive evidence that pictures are processed faster and yield larger effects than words across a range of tasks (e.g., Glaser, 1992). Future studies are needed to explicitly test different stimulus types using multiple ISIs, or even a combination of different experimental methods (e.g., RT and EEG as in Kellenbach et al., 2000; ISI = 150 ms).

### Implications for Embodied Theories of Language

In the Language and Situated Simulation (LASS) theory, Barsalou et al. (2008) proposed that linguistic and situated simulation systems interact continuously. The fast linguistic system processes linguistic forms, not meaning, and thus allows for quick and effective performance in many cases. Meaning is derived by the slower and more central simulation system when the task at hand requires the retrieval of detailed representations. Similarly, the Symbol Interdependency Hypothesis (SIH; Louwerse, 2011) makes explicit predictions in terms of early and late contributions of symbolic (linguistic) and embodied (simulation) representations. Unlike LASS, SIH places greater emphasis on linguistic representations because “language encodes perceptual information” (Louwerse, 2011,

p. 279); thus meaning can be derived already from linguistic representations.

The current findings very broadly support the distinction between early and late stages of feature activation described by both LASS and SIH. Although both theories attribute early and late effects to different systems (linguistic and simulation, respectively), our results suggest that both systems are in play already at an ISI of 100 ms (equal to an SOA of 500 ms). Associative semantic priming effects across all ISIs show that the linguistic system is continuously activated, whereas action and visual priming effects at different ISIs show differential involvement of the simulation system. We attribute action and visual priming effects to the simulation system because it is unclear how statistical interdependencies which drive the linguistic system can pick up, for example, shared manipulation features between “housekey” and “screwdriver” that do not co-occur to any regularity. In our view, both associative semantic and action priming effects demonstrate the parallel activation of the linguistic and simulation systems (but see Louwerse and Hutchinson, 2012; Hutchinson et al., 2014), thus demonstrating the fast and dynamic nature of the overall conceptual system.

From a theoretical standpoint, the current findings can be interpreted as support for both LASS and SIH. Whether meaning is derived from (or, “resides” in) either the linguistic or simulation system requires further experimentation, but we suspect that both systems are involved and interdependent through flexible recruitment of feature activations (e.g., Hoenig et al., 2008; Kiefer and Pulvermüller, 2012) and a combination of amodal and embodied representations (e.g., Barsalou, 2008; Louwerse, 2011). Indeed, we argue that a more beneficial pursuit for embodied theories of language is to describe how the time course of feature activation relates to the way knowledge is acquired, represented, and retrieved given that these theories emphasize how conceptual representations are deeply rooted in interactions of the body and the world. Furthermore, future studies should chart changes in time courses as a function of task and context to clarify how the brain makes available different kinds of information according to present needs (e.g., Hoenig et al., 2008).

### Conclusions

Our results support the following account of the time course of visual word recognition. Feature activation is both fast and slow (e.g., Zwaan, 2003; Pulvermüller et al., 2005; Barsalou et al., 2008; Louwerse, 2011), and once a feature is activated, it can affect relatively early aspects of target word recognition (i.e., priming effects *do* occur). Different features have different time courses, and the relative timing of each feature is informative about the role the feature plays in the word representation of the object. Much empirical support has been offered in support of either the early or late activation of embodied representations (e.g., Glenberg and Kaschak, 2002; Louwerse and Jeuniaux, 2010; for a review, see Meteyard et al., 2010), but by comparing different ISIs within one study, we were able to determine that different modality-specific information

is activated at different time points during visual word recognition.

## Acknowledgments

This work was supported by a Donders Graduate School for Cognitive Neuroscience TopTalent grant (NWO 022.001.026) awarded to KL. The authors thank S. Bultena and T. Uhlmann for additional help with data collection. For all technical assistance,

the authors thank the Technical Support Group of the Faculty of Social Sciences Nijmegen. We also thank both reviewers for their invaluable feedback.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00659/abstract>

## References

- Barsalou, L. W., Santos, A., Simmons, W. K., and Wilson, C. D. (2008). "Language and simulation in conceptual processing," in *Symbols, Embodiment, and Meaning*, eds M. De Vega, A. M. Glenberg and A. C. Graesser (Oxford: Oxford University Press), 245–283.
- Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300.
- Brysaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., and Böhl, A. (2011). The word frequency effect: a review of recent developments and implications for the choice of frequency estimates in German. *Exp. Psychol.* 58, 412–424. doi: 10.1027/1618-3169/a000123
- Chiarello, C., Liu, S., Shears, C., Quan, N., and Kacirik, N. (2003). Priming of strong semantic relations in the left and right visual fields: a time-course investigation. *Neuropsychologia* 41, 721–732. doi: 10.1016/S0028-3932(02)00220-8
- Ellis, R., and Tucker, M. (2000). Micro-affordance: the potentiation of components of action by seen objects. *Br. J. Psychol.* 91, 451–471. doi: 10.1348/000712600161934
- Flores d'Arcais, G., Schreuder, R., and Glazenborg, G. (1985). Semantic activation during recognition of referential words. *Psychol. Res.* 47, 39–49. doi: 10.1007/BF00309217
- Glaser, W. R. (1992). Picture naming. *Cognition* 42, 61–105. doi: 10.1016/0010-0277(92)90040-O
- Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565. doi: 10.3758/BF03196313
- Gomez, P., Ratcliff, R., and Perea, M. (2007). A model of the go/no-go task. *J. Exp. Psychol. Gen.* 136, 389–413. doi: 10.1037/0096-3445.136.3.389
- Hauk, O., Coutout, C., Holden, A., and Chen, Y. (2012). The time-course of single-word reading: evidence from fast behavioral and brain responses. *Neuroimage* 60, 1462–1477. doi: 10.1016/j.neuroimage.2012.01.061
- Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307. doi: 10.1016/S0896-6273(03)00838-9
- Hauk, O., and Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Hum. Brain Mapp.* 21, 191–201. doi: 10.1002/hbm.10157
- Hoenig, K., Sim, E. J., Bochev, V., Herrnberger, B., and Kiefer, M. (2008). Conceptual flexibility in the human brain: dynamic recruitment of semantic maps from visual, motor, and motion-related areas. *J. Cogn. Neurosci.* 20, 1799–1814. doi: 10.1162/jocn.2008.20123
- Hutchinson, S., Tillman, R. N., and Louwerse, M. M. (2014). "Quick linguistic representations and precise perceptual representations: language statistics and perceptual simulations under time constraints," in *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (Austin, TX: The Cognitive Science Society) 2399–2404.
- Hutchison, K. A., Neely, J. H., and Johnson, J. D. (2001). With great expectations, can two "wrongs" prime a "right"? *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 1451. doi: 10.1037/0278-7393.27.6.1451
- Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychon. Bull. Rev.* 10, 785–813. doi: 10.3758/BF03196544
- Kellenbach, M. L., Wijers, A. A., and Mulder, G. (2000). Visual semantic features are activated during the processing of concrete words: event-related potential evidence for perceptual semantic priming. *Cogn. Brain Res.* 10, 67–75. doi: 10.1016/S0926-6410(00)00023-9
- Keuleers, E., and Brysaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behav. Res. Methods* 42, 627–633. doi: 10.3758/BRM.42.3.627
- Kiefer, M., and Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex* 48, 805–825. doi: 10.1016/j.cortex.2011.04.006
- Kiefer, M., Sim, E. J., Helbig, H., and Graf, M. (2011). Tracking the time course of action priming on object recognition: evidence for fast and slow influences of action on perception. *J. Cogn. Neurosci.* 23, 1864–1874. doi: 10.1162/jocn.2010.21543
- Louwerse, M., and Hutchinson, S. (2012). Neurological evidence linguistic processes precede perceptual simulation in conceptual processing. *Front. Psychol.* 3:385. doi: 10.3389/fpsyg.2012.00385
- Louwerse, M. M., and Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition* 114, 96–104. doi: 10.1016/j.cognition.2009.09.002
- Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics Cogn. Sci.* 3, 273–302. doi: 10.1111/j.1756-8765.2010.01106.x
- Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol.* 102, 59–70. doi: 10.1016/j.jphysparis.2008.03.004
- Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45. doi: 10.1146/annurev.psych.57.102904.190143
- McNamara, T. P. (2005). *Semantic Priming: Perspectives from Memory and Word Recognition*. New York, NY: Psychology Press.
- Meteyard, L., Rodriguez Cuadrado, S., Bahrami, B., and Vigliocco, G. (2010). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex* 48, 788–804. doi: 10.1016/j.cortex.2010.11.002
- Moss, H. E., Ostrin, R. K., Tyler, L. K., and Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: evidence from priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 863–883. doi: 10.1037/0278-7393.21.4.863
- Myung, J. Y., Blumstein, S. E., and Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: manipulation knowledge of objects. *Cognition* 98, 223–243. doi: 10.1016/j.cognition.2004.11.010
- Pecheur, D., Zeelenberg, R., and Raaijmakers, J. G. (1998). Does pizza prime coin? Perceptual priming in lexical decision and pronunciation. *J. Mem. Lang.* 38, 401–418. doi: 10.1006/jmla.1997.2557
- Perea, M., and Gotor, A. (1997). Associative and semantic priming effects occur at very short stimulus-onset asynchronies in lexical decision and naming. *Cognition* 62, 223–240. doi: 10.1016/S0010-0277(96)00782-2
- Perea, M., and Rosa, E. (2002). Does the proportion of associatively related pairs modulate the associative priming effect at very brief stimulus-onset asynchronies? *Acta Psychol.* 110, 103–124. doi: 10.1016/S0001-6918(01)00074-9
- Pulvermüller, F., and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360. doi: 10.1038/nrn2811



- Pulvermüller, F., Härle, M., and Hummel, F. (2001). Walking or talking?: behavioral and neurophysiological correlates of action verb processing. *Brain Lang.* 78, 143–168. doi: 10.1006/brln.2000.2390
- Pulvermüller, F., Shtyrov, Y., and Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *J. Cogn. Neurosci.* 17, 884–892. doi: 10.1162/0898929054021111
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582. doi: 10.1038/nrn1706
- Rueschemeyer, S. A., Brass, M., and Friederici, A. D. (2007). Comprehending prehending: neural correlates of processing verbs with motor stems. *J. Cogn. Neurosci.* 19, 855–865. doi: 10.1162/jocn.2007.19.5.855
- Schreuder, R., Flores D'Arcais, G. B., and Glazenborg, G. (1984). Effects of perceptual and conceptual similarity in semantic priming. *Psychol. Res.* 45, 339–354. doi: 10.1007/BF00309710
- Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., and Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia* 45, 2802–2810. doi: 10.1016/j.neuropsychologia.2007.05.002
- Solomon, K. O., and Barsalou, L. W. (2004). Perceptual simulation in property verification. *Mem. Cogn.* 32, 244–259. doi: 10.3758/BF03196856
- van Dam, W. O., van Dijk, M., Bekkering, H., and Rueschemeyer, S. A. (2012). Flexibility in embodied lexical–semantic representations. *Hum. Brain Mapp.* 33, 2322–2333. doi: 10.1002/hbm.21365
- van Turenout, M., Hagoort, P., and Brown, C. (1999). The time course of grammatical and phonological processing during speaking: Evidence from event-related brain potentials. *J. Psycholinguist. Res.* 28, 649–676. doi: 10.1023/A:1023221028150
- van Turenout, M., Hagoort, P., and Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 787.
- van Turenout, M., Hagoort, P., and Brown, C. M. (1998). Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science* 280, 572–574.
- Wheatley, T., Weisberg, J., Beauchamp, M. S., and Martin, A. (2005). Automatic priming of semantically related words reduces activity in the fusiform gyrus. *J. Cogn. Neurosci.* 17, 1871–1885. doi: 10.1162/089892905775008689
- Yee, E., Huffstetler, S., and Thompson-Schill, S. L. (2011). Function follows form: activation of shape and function features during object identification. *J. Exp. Psychol. Gen.* 140, 348–363.
- Zwaan, R. A., and Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychon. Bull. Rev.* 10, 954–958. doi: 10.3758/BF03196557
- Zwaan, R. A. (2003). The immersed experienter: toward an embodied theory of language comprehension. *Psychol. Learn. Motiv.* 44, 35–62. doi: 10.1016/S0079-7421(03)44002-4

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Lam, Dijkstra and Rueschemeyer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Aesthetic Emotions Across Arts: A Comparison Between Painting and Music

Andrei C. Miu<sup>1</sup>, Simina Pițur<sup>1</sup> and Aurora Szentágotai-Tătar<sup>2\*</sup>

<sup>1</sup> Cognitive Neuroscience Laboratory, Department of Psychology, Babeș-Bolyai University, Cluj-Napoca, Romania,

<sup>2</sup> Department of Clinical Psychology and Psychotherapy, Babeș-Bolyai University, Cluj-Napoca, Romania

## OPEN ACCESS

### Edited by:

Magda L. Dumitru,  
Macquarie University, Australia

### Reviewed by:

Sascha Topolinski,  
University of Cologne, Germany  
Elena Alessandri,  
Lucerne University of Applied  
Sciences and Arts, Switzerland

### \*Correspondence:

Aurora Szentágotai-Tătar  
auraszentagotai@psychology.ro

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

Received: 19 June 2015

Accepted: 04 December 2015

Published: 05 January 2016

### Citation:

Miu AC, Pițur S and  
Szentágotai-Tătar A (2016) Aesthetic  
Emotions Across Arts: A Comparison  
Between Painting and Music.  
*Front. Psychol.* 6:1951.  
doi: 10.3389/fpsyg.2015.01951

Emotional responses to art have long been subject of debate, but only recently have they started to be investigated in affective science. The aim of this study was to compare perceptions regarding frequency of aesthetic emotions, contributing factors, and motivation which characterize the experiences of looking at painting and listening to music. Parallel surveys were filled in online by participants ( $N = 971$ ) interested in music and painting. By comparing self-reported characteristics of these experiences, this study found that compared to listening to music, looking at painting was associated with increased frequency of wonder and decreased frequencies of joyful activation and power. In addition to increased vitality, as reflected by the latter two emotions, listening to music was also more frequently associated with emotions such as tenderness, nostalgia, peacefulness, and sadness. Compared to painting-related emotions, music-related emotions were perceived as more similar to emotions in other everyday life situations. Participants reported that stimulus features and previous knowledge made more important contributions to emotional responses to painting, whereas prior mood, physical context and the presence of other people were considered more important in relation to emotional responses to music. Self-education motivation was more frequently associated with looking at painting, whereas mood repair and keeping company motivations were reported more frequently in relation to listening to music. Participants with visual arts education reported increased vitality-related emotions in their experience of looking at painting. In contrast, no relation was found between music education and emotional responses to music. These findings offer a more general perspective on aesthetic emotions and encourage integrative research linking different types of aesthetic experience.

**Keywords:** aesthetic emotions, painting, music, art education

## INTRODUCTION

Emotional responses to art (i.e., aesthetic emotions) have long interested philosophers, psychologists, and art critics (Robinson, 2004). Theories in psychology and aesthetics (James, 1890/1950; Bell, 1914; Berlyne, 1974) initially focused on positive emotional responses that arise from the appreciation of the form of expression as beautiful, harmonious, or powerful (Robinson, 2004). Recent studies have found that indeed, emotions (i.e., brief affective states triggered by the appraisal of an event in relation to current goals; Scherer and Zentner, 2001) such as awe

(Shiota et al., 2007) and wonder (Zentner et al., 2008) are frequently reported in relation to the contemplation of artworks. These emotions typically occur when an object or event is appraised as highly complex and novel, and creates a sense of being in the presence of something greater than oneself (Keltner and Haidt, 2003).

However, it has also been recently emphasized that affective responses to art are more diverse (Silvia, 2011) and often include emotions such as sadness (Vuoskoski and Eerola, 2012) and nostalgia (Barrett et al., 2010), which are also experienced in other everyday situations that do not involve contemplation of artworks. These emotions may be related to the content and personal interpretation of an artwork, rather than its form (Robinson, 2004; Silvia, 2011). For instance, one may admire Caravaggio's skill in *David with the Head of Goliath*, but also feel disgust at the sight of dripping blood, and sadness at the thought that this artwork may express the painter's remorse. Similarly, someone listening to the *Adagietto* from Mahler's 5th Symphony may feel blends of awe, tenderness and nostalgia related to the skillful orchestration, on the one hand, and knowing that this piece captures the composer's love for his wife and worries for his deteriorating health, on the other hand. Therefore, art contemplation can trigger multiple emotions, which include aesthetic emotions driven by positive appraisals of the form of expression, and other positive or negative emotions, driven by appraisals of the content or meaning of artworks (Silvia, 2011). Given the increasing interest in affective science (Gross and Barrett, 2013), recent studies have focused on describing emotions associated with aesthetic experiences such as looking at painting and listening to music, and on examining their underlying mechanisms and motivation (for review see Silvia, 2011; Swaminathan and Schellenberg, 2015).

Influential theoretical frameworks, which have guided research on preferences for painting (Leder et al., 2004; Lindell and Mueller, 2011) and emotional responses to music (Scherer and Zentner, 2001), argue that one's reactions to artworks involve an interplay of multiple factors related to stimulus, person, and situation. The contribution of perceptual features and formal characteristics conveying style has been pointed out by observations that aesthetic preferences form very rapidly (i.e., in less than 1 s), whether in the form of beauty judgments of graphic patterns (Jacobsen and Höfel, 2003) or emotional categorization of music excerpts (Bigand et al., 2005). Indeed, these rapid responses may involve automatic mechanisms such as visual disambiguation (Topolinski et al., 2015) and premotor simulation (Leder et al., 2012; Ticini et al., 2014), although recent studies also report their interaction with consciously controlled processes such as expectations (McLean et al., 2015). The relations between the structural characteristics of music (e.g., mode, tempo) and emotional responses have been systematically investigated (Gabrielsson and Lindstrom, 2001; Gomez and Danuser, 2007). Taking a more general approach, research relevant to painting has mostly focused on non-aesthetic stimuli (e.g., geometrical shapes) and broad aesthetic preferences instead of specific emotions (Jacobsen and Höfel, 2002). Nonetheless, theory in both fields (Scherer and

Zentner, 2001; Lindell and Mueller, 2011) has acknowledged that stimulus-driven or "bottom-up" processing interacts with education and psychological characteristics that can influence emotional responses to art through knowledge-driven or "top-down" processing.

Many studies have therefore examined whether art education facilitates art-related emotions through a better understanding of the formal means of expression in painting or music. Indeed, students in art history compared to students in other fields categorize paintings using more criteria and favor style-related rather than affective criteria (Augustin and Leder, 2006). Similarly, musicians perceive the links between a musical theme and its variations better than non-musicians (Bigand and Poulin-Charronnat, 2006), and describe music using adjectives related to novelty and originality rather than emotional characteristics (Istok et al., 2009). However, despite these differences in processing styles, music-related emotions are not markedly dissimilar in musicians and non-musicians (Bigand and Poulin-Charronnat, 2006; Baltes et al., 2012) and the same may be true for painting-related emotions. While no study investigated the influence of visual arts expertise on emotional responses to painting, experimental evidence suggests that providing additional information that facilitates understanding of paintings does not influence preference for paintings (Leder et al., 2006). In addition to art education, other individual differences such as prior mood may also influence emotional responses to artworks (Hunter et al., 2011; Vuoskoski and Eerola, 2011a,b; Baltes and Miu, 2014).

Situational factors may also modulate art-related emotions. For instance, the presence of other people such as in the attendance of live music performance or during a visit to an art gallery may influence emotional responses to artworks. Field studies (Juslin et al., 2008) and experimental studies (Liljestrom et al., 2013) showed that the presence of the romantic partner or a close friend during music listening increases the frequency of affective states such as happiness-elation, pleasure-enjoyment and admiration-awe. These findings highlight social facilitation as one of the factors that may contribute to the increased enjoyment of music during live music performance (Lamont, 2011). The influence of context has also been acknowledged in theories of painting-related emotions (Leder et al., 2004) and one study (Pelowski et al., 2014) suggested that social encounters in art galleries may be detrimental to aesthetic experience by inducing competition between social awareness and self-focused enjoyment of paintings. However, the influence of social factors and other contextual variables (e.g., location; Scherer and Zentner, 2001) needs further research, particularly in the case of painting-related emotions.

In addition to mechanisms, recent studies have also focused on motivation for exposure to art. The most commonly reported reason for music listening is "mood repair" or emotion regulation, but social reasons (e.g., alleviate loneliness; keep up with art trends) and self-actualization needs (e.g., explore and express identity) are also frequently reported (Lonsdale and North, 2011). People use music to manage their mood to a greater extent than they use other leisure activities such as reading or exercising (Lonsdale and North, 2011). However, the tendency to

use music for mood repair may be influenced by music training considering that musicians use music for cognitive (e.g., attention to structural complexity or performing technique) rather than emotional reasons (Getz et al., 2014). To our knowledge, no study has yet investigated motivation for aesthetic experience with painting.

In summary, painting and music-related emotions seem to involve a similar interplay of factors related to stimulus, person and context. However, any attempt to generalize across experience with these arts is currently hampered by the lack of empirical evidence on certain issues, particularly in the case of painting (e.g., frequency of specific emotions; influence of visual arts expertise, prior mood and social context; motivation), as well as the absence of integrative studies systematically comparing the characteristics of aesthetic experience in relation to painting and music (but see Rawlings et al., 2000; Cleridou and Furnham, 2014). In this study, parallel surveys on the experience of looking at painting and listening to music were filled in online by two samples of volunteers. Self-reported frequency of emotions, evaluation of contributing factors, and motivation in aesthetic experience with painting and music were compared between samples. In addition, the influence of art education on the characteristics of aesthetic experience was also investigated.

## MATERIAL AND METHODS

### Participants

The surveys on looking at painting and listening to music were separately advertised online, mainly through social media (e.g., Facebook), as part of a psychological study on aesthetic experience. The survey on looking at painting was filled in by 260 participants, and the survey on listening to music was filled in by 711 participants. The surveys were in Romanian and all participants reported Romanian as their first language. **Table 1** shows the distributions of age, sex, general education, and occupational status, which were not significantly different between the samples. Participants were informed that they would answer questions about their experience of looking at painting or listening to music, and signed a consent form before accessing the survey. The study followed the recommendations of the Declaration of Helsinki regarding participant safety and was approved by the Ethics Committee of Babeş-Bolyai University.

### Surveys

The questions and answer options were equivalent in the two surveys. Other than the reference to painting or music, the phrasing was identical.

The surveys were divided into three sections. The first section focused on socio-demographic characteristics: age, sex, education level, and occupational status.

The second section surveyed art education, asking participants whether they had graduated from a high school or college in the field of visual arts or music. Participants who filled in the survey on painting-related experiences were also asked to report whether they had knowledge related to painting or drawing, sculpture, and/or art history. Those who filled in the survey on music-related experiences were asked to report whether they had knowledge related to sight reading of musical scores, instrument playing and/or musicology. They were also asked to assess how experienced they thought they were in looking at painting or listening to music (five-point scale: 1, beginner; 5, experienced), as well as the personal importance of these art-related activities (five-point scale: 1, not at all important; 5, very important).

The third section included questions about frequency of art-related emotions, perception of contributing factors, and motivation for aesthetic experience. Emotional experience was assessed by asking participants to rate the frequency of several emotions in relation to looking at painting or listening to music, using a five-point scale (1, never; 2, rarely; 3, sometimes; 4, frequently; 5, very frequently). The emotion labels were taken from the 25-item version of the Geneva Emotional Music Scale (Zentner et al., 2008), representing nine emotion categories: wonder, transcendence, tenderness, nostalgia and peacefulness (facets of the more general dimension of “sublimity”); power and joyful activation (facets of “vitality”); and tension and sadness (facets of “unease”). To our knowledge, GEMS is the only standardized instrument covering the whole spectrum of emotional responses to artworks, including both positive aesthetic emotions (e.g., wonder, transcendence), and other positive (e.g., joyful activation, power) and negative emotions (e.g., nostalgia, sadness) that occur in various situations in everyday life. There is no equivalent standardized assessment of emotional responses to painting and developing such an instrument was beyond the purpose of this study. However, we thought GEMS was suitable for this exploratory study considering the potential similarities between emotional responses to music (Zentner and Eerola, 2010) and painting (Silvia, 2011). The Romanian translation of GEMS was used in several previous studies (e.g., Miu and Baltes, 2012; Baltes and Miu, 2014).

In addition to assessing the frequency of emotions using GEMS, another item asked participants to rate the similarity between everyday emotions and emotional experience with

**TABLE 1 | Socio-demographic characteristics of the survey samples.**

	Age (M ± SD)	Sex (%)		General education (%)			Occupational status (%)			
		Women	Men	Primary	Secondary	Higher	Student	Employed	Unemployed	Retired
Looking at painting	30.27 ± 11.53	82.31	17.69	5	30.38	64.62	40.77	52.31	4.23	2.69
Listening to music	28.10 ± 10.17	78.62	21.38	3.94	36.71	59.35	45.15	49.37	3.52	1.96

Abbreviations: M, mean; SD, standard deviation.



painting or music using a five-point scale (1, not at all; 5, very much).

Participants also rated, on a scale from 1 (not at all) to 5 (very much), the extent to which painting or music-related emotions involved one of the following factors: (1) structural features of the aesthetic stimulus, such as form, color, contrasts and composition for painting, and mode and tempo for music; (2) physical context (e.g., location); (3) prior mood, immediately before exposure to artworks; (4) previous knowledge about artwork and artist (i.e., painter or composer); and (5) presence of other people, when aesthetic experience occurs in social contexts. These factors were inspired by previous studies (Scherer and Zentner, 2001).

Another item focused on motivation, and participants were asked to rate the importance of five potential reasons in their aesthetic experience with painting or music: (1) mood management or relaxation; (2) experiencing new emotions, which are not typical of everyday life; (3) self-education; (4) sharing emotions with others; and (5) keeping company when one feels lonely. These types of motivation were also derived from previous literature (Lonsdale and North, 2011).

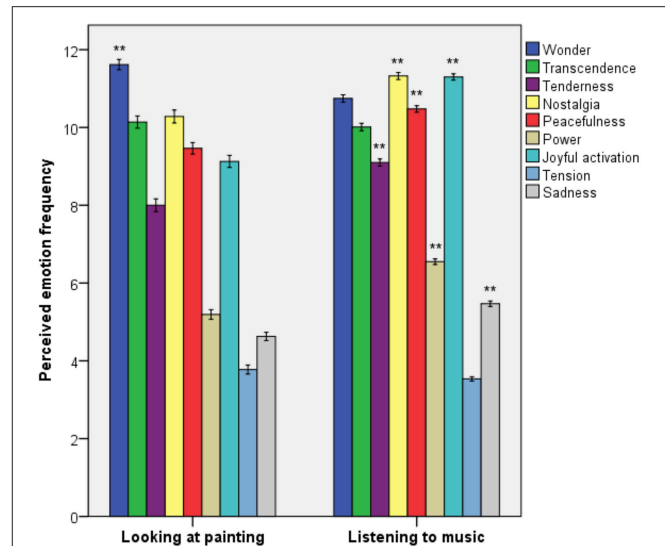
## Statistical Analyses

The main analyses compared self-reported frequency of emotions, contributing factors and motivation for the two types of aesthetic experience: looking at painting and listening to music. Other analyses compared between participants with and without art education. Considering the unequal sizes of the two samples, as well as of the groups with and without art education, we used analysis of variance (ANOVA) with Welch's correction for unequal variance, which is a robust method to protect against type I errors while conserving power (Kohr and Games, 1974). In addition, we used the Bonferroni method to correct the threshold of statistical significance for each set of analyses, as follows:  $p \leq 0.005$  (0.05/9) for self-reported frequency of emotions;  $p \leq 0.01$  (0.05/5) for perceived contributing factors; and  $p \leq 0.01$  (0.05/5) for self-reported motivation. Effect sizes are reported as  $\eta_p^2$ , where an effect of 0.01 is small, one of 0.06 is medium, and one of 0.14 is large (Cohen, 1988). All analyses were run in SPSS.

## RESULTS

### Painting and Music-Related Emotions

By comparing self-reported frequency of each emotion between samples (Figure 1), we found that those who described their experience of looking at painting reported higher frequencies of wonder compared to those who described their experience of listening to music [ $F_{(1, 525.29)} = 28.49, p < 0.001, \eta_p^2 = 0.03$ ]. In contrast, the frequencies of tenderness [ $F_{(1, 434.56)} = 33.86, p < 0.001, \eta_p^2 = 0.04$ ], nostalgia [ $F_{(1, 419.57)} = 30.09, p < 0.001, \eta_p^2 = 0.03$ ], peacefulness [ $F_{(1, 438.95)} = 35.83, p < 0.001, \eta_p^2 = 0.04$ ], power [ $F_{(1, 447.32)} = 89.75, p < 0.001, \eta_p^2 = 0.09$ ], joyful activation [ $F_{(1, 410.84)} = 151.69, p < 0.001, \eta_p^2 = 0.15$ ], and sadness [ $F_{(1, 501.01)} = 43.55, p < 0.001, \eta_p^2 = 0.04$ ] were higher in relation to listening to music compared to looking at painting.



**FIGURE 1 | Perceived frequency of emotions in the experience of looking at painting and listening to music.** Error bars indicate standard error of the mean. \*\* $p < 0.01$ .

The frequency of transcendence and tension were not different in the two samples.

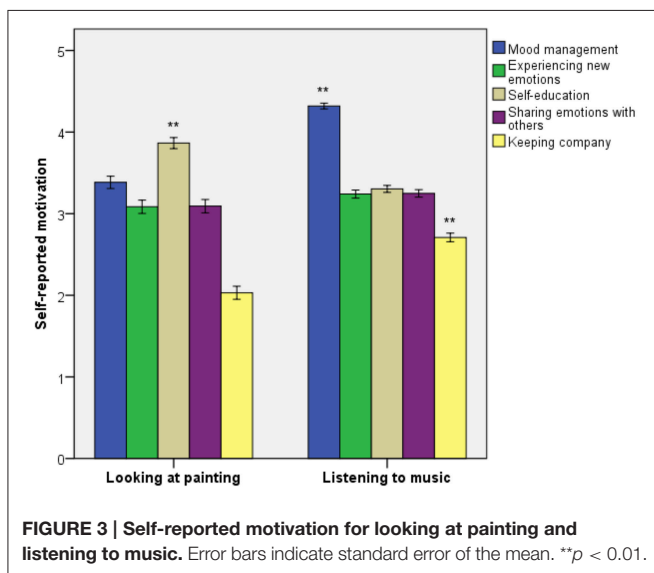
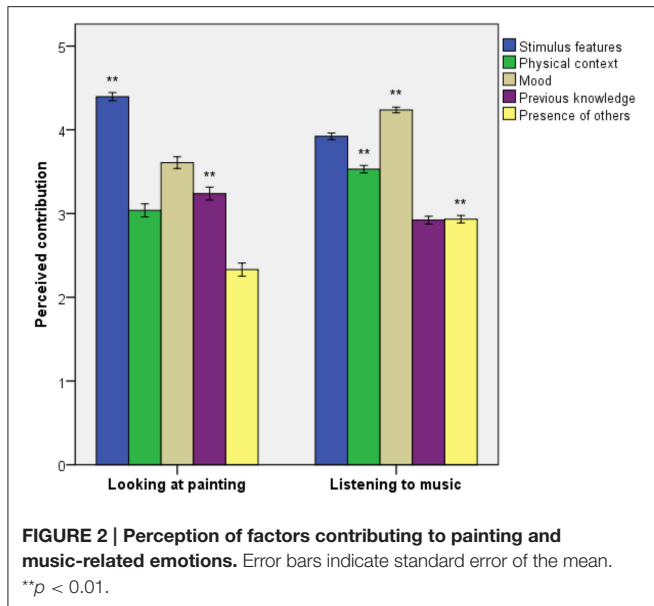
The perceived similarity between art-related emotions and everyday emotions was also analyzed. Painting-related emotions ( $M = 3.25; SD = 0.99$ ) were rated as significantly less similar to emotions in other everyday situations, compared to music-related emotions ( $M = 3.53; SD = 0.97$ ):  $F_{(1, 450.29)} = 15.89, p < 0.001, \eta_p^2 = 0.02$ .

### Perception of Contributing Factors

Figure 2 shows the perceived contributions of several factors to art-related emotions. The contributions of stimulus features [ $F_{(1, 624.81)} = 56.85, p < 0.001, \eta_p^2 = 0.04$ ] and previous knowledge [ $F_{(1, 461.09)} = 12.48, p < 0.001, \eta_p^2 = 0.01$ ] were rated at higher levels for painting-related emotions, whereas the contributions of prior mood [ $F_{(1, 384.60)} = 65.93, p < 0.001, \eta_p^2 = 0.08$ ], physical context [ $F_{(1, 437.99)} = 30.29, p < 0.001, \eta_p^2 = 0.03$ ], and the presence of others [ $F_{(1, 433.12)} = 44.99, p < 0.001, \eta_p^2 = 0.05$ ] were rated at higher levels for music-related emotions.

### Self-Reported Motivation

Self-reported motivation was also compared between participants who described their experience of looking at painting and listening to music (Figure 3). Self-education was rated as more important for looking at painting [ $F_{(1, 481.05)} = 48.48, p < 0.001, \eta_p^2 = 0.05$ ], whereas mood management [ $F_{(1, 375.83)} = 125.61, p < 0.001, \eta_p^2 = 0.14$ ] and keeping company [ $F_{(1, 506.15)} = 50.21, p < 0.001, \eta_p^2 = 0.05$ ] were rated as more important for music listening. Experiencing new emotions and sharing emotions with others were rated at comparable levels for looking at painting and music listening.



## Art Education

There were 69 visual arts graduates in the sample that answered the painting survey, and 42 music graduates in the sample that answered the music survey. The majority of visual arts graduates reported knowledge about painting (99.65%), sculpture (55.07%), and art history (95.65%). The self-reported level of experience with painting [ $t_{(258)} = 7.53, p < 0.001, \text{Cohen's } d = 1.05$ ], and the personal importance of painting [ $t_{(258)} = 5.04, p < 0.001, \text{Cohen's } d = 0.72$ ] were significantly higher for visual art graduates compared to the other participants who filled in the painting survey. Similarly, most music graduates reported knowledge related to sight reading of music scores (92.86%), instrument playing (95.23%), and musicology (85.71%). The self-reported levels of experience with music [ $t_{(51.74)} = 7.56,$

$p < 0.001, \text{Cohen's } d = 1.02$ ] and the personal importance of music [ $t_{(64.99)} = 5.95, p < 0.001, \text{Cohen's } d = 0.65$ ] were significantly higher for music graduates compared to the other participants who filled in the survey on listening to music.

Next, self-reported frequency of emotions, perception of contributing factors, and self-reported motivation for looking at painting and listening to music were compared between participants with and without art education in each sample (Table 2).

Participants with visual arts education reported significantly higher frequencies of power [ $F_{(1, 106.04)} = 10.18, p = 0.002, \eta_p^2 = 0.04$ ] and joyful activation [ $F_{(1, 120.54)} = 17.32, p < 0.001, \eta_p^2 = 0.06$ ] in their experience with painting, in comparison to participants without visual arts education. Frequencies of the other painting-related emotions were not significantly different between those with and without visual arts education. Self-reported frequencies of all music-related emotions were similar in participants with and without music education.

Perceived similarity between art-related (i.e., painting or music) and everyday emotions was not significantly different in participants with and without art education (i.e., visual arts education or music education).

Both participants with visual arts education [ $F_{(1, 122.63)} = 6.81, p = 0.010, \eta_p^2 = 0.03$ ] and those with music education [ $F_{(1, 46.04)} = 23.91, p < 0.001, \eta_p^2 = 0.03$ ] rated the contribution of previous knowledge to painting-related emotions and music-related emotions, respectively, as more important, in comparison to participants without art education (Table 3).

There were no significant differences related to art education in self-reported motivation for looking at painting or listening to music (Table 4).

## DISCUSSION

In this study, participants answered surveys on their experience of looking at painting and listening to music. The main aims were to compare between perceptions regarding frequency of emotions, contribution of several factors to art-related emotions, and motivation for these two types of aesthetic experience. In addition, we examined the influence of art education on these dimensions.

Previous studies identified emotions that are commonly experienced by music listeners (Zentner et al., 2008). Aesthetic emotions such as awe (Shiota et al., 2007) and other positive and negative emotions that occur in various everyday situations (Silvia, 2011) have also been described in the experience of looking at painting. These studies suggested that looking at painting and listening to music are associated with blends of different types of emotions. However, no study has yet compared the relative frequency of different emotions in these two types of aesthetic experience. The present results indicate that wonder may be more frequently experienced while looking at painting rather than while listening to music. In addition, the experience of looking at painting may be associated with relatively lower frequency of vitality-related emotions (Zentner et al., 2008) such as joyful activation and power. These two emotions were much

**TABLE 2 | Perceived frequency of emotions in participants with and without arts education.**

		Perceived frequency of emotions									Similarity to emotions in other situations
		Wonder	Transcendence	Tenderness	Nostalgia	Peacefulness	Power	Joyful activation	Sadness	Tension	
Looking at painting	Visual arts graduates	11.46 ± 2.24	10.36 ± 2.56	8.3 ± 2.84	10.79 ± 2.57	9.81 ± 2.55	5.88 ± 2.17	10.17 ± 2.43	4.78 ± 2.01	4.26 ± 2.18	3.29 ± 0.98
	No formal visual arts education	11.67 ± 2.12	10.05 ± 2.52	7.89 ± 2.57	10.09 ± 2.72	9.34 ± 2.28	4.94 ± 1.86	8.74 ± 2.44	4.57 ± 1.59	3.6 ± 1.72	3.23 ± 1
Listening to music	Music graduates	10.9 ± 2.1	10.5 ± 2.57	8.78 ± 2.29	11.54 ± 2.59	10.09 ± 2.03	6.59 ± 1.86	10.85 ± 2.24	5.42 ± 1.66	3.57 ± 1.43	3.36 ± 1.2
	No formal music education	10.73 ± 2.50	9.98 ± 2.53	9.11 ± 2.48	11.3 ± 2.4	10.49 ± 2.24	6.54 ± 1.93	11.32 ± 2.18	5.47 ± 1.89	3.53 ± 1.51	3.54 ± 0.95

Values in cells are means and standard deviations.

**TABLE 3 | Perception of factors contributing to art-related emotions in participants with and without arts education.**

		Perception of contributing factors				
		Stimulus features	Physical context	Mood	Previous knowledge	Presence of other people
Looking at painting	Visual arts graduates	4.51 ± 0.76	3.35 ± 1.27	3.55 ± 1.15	3.57 ± 1.2	2.57 ± 1.30
	No formal visual arts education	4.36 ± 0.78	2.93 ± 1.22	3.63 ± 1.11	3.12 ± 1.23	2.25 ± 1.23
Listening to music	Music graduates	4.29 ± 0.94	3.4 ± 1.14	4.21 ± 0.84	3.83 ± 1.24	2.74 ± 1.08
	No formal music education	3.9 ± 1.06	3.54 ± 1.18	4.24 ± 0.89	2.86 ± 1.21	2.94 ± 1.17

Values in cells are means and standard deviations.

**TABLE 4 | Self-reported motivation for looking at painting and listening to music in participants with and without arts education.**

		Motivation				
		Mood management	Experiencing new emotions	Self-education	Sharing emotions with others	Keeping company
Looking at painting	Visual arts graduates	3.43 ± 1.27	3.22 ± 1.32	4.14 ± 1.01	3.23 ± 1.33	2.13 ± 1.32
	No formal visual arts education	3.37 ± 1.2	3.04 ± 1.31	3.76 ± 1.11	3.04 ± 1.31	1.99 ± 1.26
Listening to music	Music graduates	3.86 ± 1.29	3.62 ± 1.39	3.64 ± 1.24	3.38 ± 1.36	2.6 ± 1.43
	No formal music education	4.35 ± 0.9	3.22 ± 1.31	3.28 ± 1.14	3.24 ± 1.22	2.72 ± 1.42

Values in cells are means and standard deviations.

more frequently (i.e., large or medium effect size) reported in relation to listening to music, which suggests that “vitality” may best distinguish emotional responses to music and painting. Other emotions (i.e., tenderness, nostalgia, peacefulness, sadness) were also more frequently reported in the experience of listening to music compared to looking at painting, but to a lesser degree, that is, with small effect sizes.

Painting-related emotions were perceived as less similar to emotions experienced in other everyday life situations compared to music-related emotions. This perception may be connected to the relatively higher frequency of wonder associated with looking at painting, considering that this emotion is experienced

in limited contexts (e.g., contemplation of artworks or nature scenes; Shiota et al., 2007) that create the sensation of being in the presence of something greater than oneself (Keltner and Haidt, 2003). The reduced vitality of emotions associated with looking at painting may also contribute to the impression that they are different from emotional experience in general.

These results also indicate differences in the perception of factors that may contribute to art-related emotions. Participants rated stimulus features and previous knowledge as making more important contributions to emotional responses to painting than to music. These impressions are in line with theories (Berlyne, 1974) and experimental evidence (Jacobsen and Höfel,

2002; Leder et al., 2012; Ticini et al., 2014; McLean et al., 2015; Topolinski et al., 2015) that support the relation between perceptual features of paintings and their emotional impact. The present observations do not exclude the contribution of these factors to music-induced emotions, which is well documented in the literature (Gabrielsson and Lindstrom, 2001; Gomez and Danuser, 2007), but merely suggest that people perceive them as weighing more in the experience of looking at painting. In addition, the perception that previous knowledge plays an important role in painting-related emotions was corroborated by another observation in this study (see below), namely that the frequency of certain painting-related emotions was higher in visual art graduates, who reported higher levels of art knowledge. In a complementary way, the influence of prior mood, physical context, and the presence of other people were rated as more important in relation to music-induced emotions. These subjective evaluations are also in line with previous evidence showing that indeed, both mood prior to music exposure, whether in laboratory (Hunter et al., 2011; Vuoskoski and Eerola, 2011b) or concert hall (Vuoskoski and Eerola, 2011b; Baltes and Miu, 2014), and the presence of others, particularly close persons (Juslin et al., 2008; Liljestrom et al., 2013), influence emotional responses to music.

Experiences of looking at painting and listening to music were also differentiated by self-reported motivation. Relatively more participants reported that self-education motivated them to look at painting. In addition, relatively more participants reported that mood repair and keeping company drove their experience of listening to music. These motivational differences may be supported by many factors, including the wider accessibility of music on portable devices, which may increase its use for everyday life needs such as mood repair (Lonsdale and North, 2011), and the relatively higher vitality of emotional responses to music, which may contribute to increasing function in everyday life. Pending on replication of these results, future research could examine why people use the experience of looking at painting and listening to music for relatively different reasons.

Visual arts graduates reported higher frequencies of power and joyful activation in their experience of looking at painting. Considering that these emotions had the lowest frequencies in the overall sample that answered the painting survey, this indicates that visual arts formal training has a significant impact on emotional responses to painting and may specifically enhance vitality-related emotions. In contrast, music formal training had no significant effect on the frequency of music-related emotions, which is in line with previous evidence (Bigand and Poulin-Charronnat, 2006; Baltes et al., 2012). These findings suggest that painting-related emotions may involve knowledge-driven or top-down information processing to a larger extent than music-related emotions. However, both visual arts and music graduates rated the contribution of previous knowledge (e.g., information about artwork and artist) to emotional responses at higher levels than participants without formal art training. No differences in motivation for looking at painting and listening to music were linked to formal art education. Given that art

graduates reported increased levels of art-related knowledge—although note that this type of knowledge was not limited to those with formal training—, as well as increased experience with and personal importance of art, these differences may have driven the present observations on the influence of formal art training.

This study has at least two main limitations. First, being based on surveys, these findings describe how art-related experience is perceived by people, and may thus be subjectively biased. For instance, all art graduates reported that increased levels of art knowledge would enhance art-related emotions, but only visual arts education seemed to influence emotional responses to painting. Second, we assessed emotional experience using a scale that focuses on emotions which are common in the experience of listening to music. There is no similar scale for painting-related emotions, so the only available options for this study were measures focused on music-induced emotions such as GEMS (Shiota et al., 2007) and general measures such as PANAS (Watson and Clark, 1994). We chose the former option considering that GEMS, which was developed through a factorial approach based on self-reported experience of music listeners (Zentner et al., 2008), may offer a more specific assessment of aesthetic emotions, leaving out emotions that are not representative for the experience of music listening and may be equally unrepresentative for the experience of looking at painting. Previous studies suggested some similarities between emotional responses to painting and music (Shiota et al., 2007). In addition, GEMS and PANAS partially overlap, with emotions like wonder, power, joyful activation, tension, and sadness from the former scale paralleling emotions like serenity, self-assurance, joviality, hostility, and sadness from the latter scale. Notwithstanding these reasons in favor of our approach, it is possible that we did not assess emotions that are more specific to looking at painting and are not covered by GEMS. For instance, recent studies identified so-called “knowledge emotions” such as surprise, interest and confusion in the experience of looking at painting (Silvia, 2011). Therefore, the specificity of painting-related emotions may have been underestimated in this study. Future research may identify other specific aspects of emotional responses to painting.

In conclusion, our results highlighted multiple differences in the perceived qualities of looking at painting and listening to music: emotional responses to painting may be characterized by higher levels of wonder and lower vitality, and are perceived as less similar to emotions in other everyday life situations, compared to music-induced emotions; people outweigh the contributions of stimulus features and previous knowledge in relation to emotional responses to painting, and the contributions of prior mood, physical context, and the presence of others in relation to emotional responses to music; looking at painting is driven by self-education motivation, whereas listening to music is associated with emotional and social motivation; and formal art training influences emotional responses to painting (e.g., by increasing vitality), but not to music, which suggests that the former may depend more on knowledge-driven information processing. We hope this study will encourage the integration



of theories and approaches in research on painting and music, which have largely developed in parallel until now, and stimulate future research that could give a more detailed perspective on common and specific aspects of aesthetic experiences with different forms of art.

## REFERENCES

- Augustin, M. D., and Leder, H. (2006). Art expertise: a study of concepts and conceptual spaces. *Psychol. Sci.* 48, 135–156.
- Baltes, F. R., Miclea, M., and Miu, A. C. (2012). Does everybody like Vivaldi's Four Seasons? Affective space and a comparison of music-induced emotions between musicians and non-musicians. *Cogn. Brain Behav. Interdiscip.* 16, 107–119.
- Baltes, F. R., and Miu, A. C. (2014). Emotions during live music performance: links with individual differences in empathy, visual imagery, and mood. *Psychomusicol. Music Mind Brain* 24, 58–65. doi: 10.1037/pmu0000030
- Barrett, F. S., Grimm, K. J., Robins, R. W., Wildschut, T., Sedikides, C., and Janata, P. (2010). Music-evoked nostalgia: affect, memory, and personality. *Emotion* 10, 390–403. doi: 10.1037/a0019006
- Bell, C. (1914). *Art*. London: Chatoo and Windus.
- Berlyne, D. E. (1974). *Studies in the New Experimental Aesthetics*. New York, NY: Wiley.
- BZigand, E., Filipic, S., and Lalitte, P. (2005). The time course of emotional responses to music. *Ann. N.Y. Acad. Sci.* 1060, 429–437. doi: 10.1196/annals.1360.036
- Bigand, E., and Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition* 100, 100–130. doi: 10.1016/j.cognition.2005.11.007
- Cleridou, K., and Furnham, A. (2014). Personality correlates of aesthetic preferences for art, architecture, and music. *Empir. Stud. Arts* 32, 231–255. doi: 10.2190/EM.32.2.f
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Gabrielsson, A., and Lindstrom, E. (2001). “The influence of musical structure on emotional expression,” in *Music and Emotion: Theory and Research*, eds P. N. Juslin and J. A. Sloboda (New York, NY: Oxford University Press), 223–248.
- Getz, L. M., Marks, S., and Roy, M. (2014). The influence of stress, optimism, and music training on music uses and preferences. *Psychol. Music* 42, 71–85. doi: 10.1177/0305735612456727
- Gomez, P., and Danuser, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion* 7, 377–387. doi: 10.1037/1528-3542.7.2.377
- Gross, J. J., and Barrett, L. F. (2013). The emerging field of affective science. *Emotion* 13, 997–998. doi: 10.1037/a0034512
- Hunter, P. G., Schellenberg, E. G., and Griffith, A. T. (2011). Misery loves company: mood-congruent emotional responding to music. *Emotion* 11, 1068–1072. doi: 10.1037/a0023749
- Istok, E., Brattico, E., Jacobsen, T., Krohn, K., Muller, M., and Tervaniemi, M. (2009). Aesthetic responses to music: a questionnaire study. *Musicae Sci.* 13, 183–206. doi: 10.1177/102986490901300201
- Jacobsen, T., and Höfel, L. (2002). Aesthetic judgments of novel graphic patterns: analyses of individual judgments. *Percept. Mot. Skills* 95, 755–766. doi: 10.2466/pms.2002.95.3.755
- Jacobsen, T., and Höfel, L. (2003). Descriptive and evaluative judgment processes: behavioral and electrophysiological indices of processing symmetry and aesthetics. *Cogn. Affect. Behav. Neurosci.* 3, 289–299. doi: 10.3758/CABN.3.4.289
- James, W. (1890/1950). *The Principles of Psychology*. New York, NY: Dover.
- Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G., and Silva, A. (2008). An experience sampling study of emotional reactions to music: listener, music, and situation. *Emotion* 8, 668–683. doi: 10.1037/a0013505
- Keltner, D., and Haidt, J. (2003). Approaching awe, a moral, spiritual, and aesthetic emotion. *Cogn. Emot.* 17, 297–314. doi: 10.1080/026999303022297
- Kohr, R. L., and Games, P. A. (1974). Robustness of the analysis of variance, the Welch procedure and a box procedure to heterogeneous variances. *J. Exp. Educ.* 43, 61–69. doi: 10.1080/00220973.1974.10806305
- Lamont, A. (2011). University students' strong experiences of music: pleasure, engagement, and meaning. *Musicae Sci.* 15, 229–249. doi: 10.1177/1029864911403368
- Leder, H., Bär, S., and Topolinski, S. (2012). Covert painting simulations influence aesthetic appreciation of artworks. *Psychol. Sci.* 23, 1479–1481. doi: 10.1177/0956797612452866
- Leder, H., Belke, B., Oeberst, A., and Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *Br. J. Psychol.* 95, 489–508. doi: 10.1348/0007126042369811
- Leder, H., Carbon, C. C., and Ripsas, A. L. (2006). Entitling art: influence of title information on understanding and appreciation of paintings. *Acta Psychol.* 121, 176–198. doi: 10.1016/j.actpsy.2005.08.005
- Liljeström, S., Juslin, P. N., and Västfjäll, D. (2013). Experimental evidence of the roles of music choice, social context, and listener personality in emotional reactions to music. *Psychol. Music* 41, 579–599. doi: 10.1177/0305735612440615
- Lindell, A. K., and Mueller, J. (2011). Can science account for taste? Psychological insights into art appreciation. *J. Cogn. Psychol.* 23, 453–475. doi: 10.1080/20445911.2011.539556
- Lonsdale, A. J., and North, A. C. (2011). Why do we listen to music? A uses and gratifications analysis. *Br. J. Psychol.* 102, 108–134. doi: 10.1348/000712610X506831
- McLean, C., Want, S. C., and Dyson, B. J. (2015). The role of similarity, sound and awareness in the appreciation of visual artwork via motor simulation. *Cognition* 137, 174–181. doi: 10.1016/j.cognition.2015.01.002
- Miu, A. C., and Baltes, F. R. (2012). Empathy manipulation impacts music-induced emotions: a psychophysiological study on opera. *PLoS ONE* 7:e30618. doi: 10.1371/journal.pone.0030618
- Pelowski, M., Liu, T., Palacios, V., and Akiba, F. (2014). When a body meets a body: an exploration of the negative impact of social interactions on museum experiences of art. *Int. J. Educ. Arts* 15, 1–47.
- Rawlings, D., Vidal, N. B. I., and Furnham, A. (2000). Personality and aesthetic preference in Spain and England: two studies relating sensation seeking and openness to experience to liking for paintings and music. *Eur. J. Pers.* 14, 553–576. doi: 10.1002/1099-0984(200011/12)14:6<553::AID-PER384>3.0.CO;2-H
- Robinson, J. (2004). “The emotions in art,” in *The Blackwell Guide to Aesthetics*, ed P. Kivy (Oxford: Blackwell Publishing), 174–192.
- Scherer, K. R., and Zentner, M. R. (2001). “Emotional effects of music: production rules,” in *Music and Emotion: Theory and Research*, eds P. N. Juslin and J. A. Sloboda (Oxford: Oxford University Press), 361–392.
- Shiota, M. N., Keltner, D., and Mossman, A. (2007). The nature of awe: elicitors, appraisals, and effects on self-concept. *Cogn. Emot.* 21, 944–963. doi: 10.1080/02699930600923668
- Silvia, P. J. (2011). “Human emotions and aesthetic experience: an overview of empirical aesthetics,” in *Aesthetic Science: Connecting Minds, Brains, and Experience*, eds A. P. Shimamura and S. E. Palmer (New York, NY: Oxford University Press), 250–275.
- Swaminathan, S., and Schellenberg, E. G. (2015). Current emotion research in music psychology. *Emot. Rev.* 7, 189–197. doi: 10.1177/1754073914558282
- Ticini, L. F., Rachman, L., Pelletier, J., and Dubal, S. (2014). Enhancing aesthetic appreciation by priming canvases with actions that match the artist's painting style. *Front. Hum. Neurosci.* 8:391. doi: 10.3389/fnhum.2014.00391
- Topolinski, S., Erle, T. M., and Reber, R. (2015). Necker's smile: immediate affective consequences of early perceptual processes. *Cognition* 140, 1–13. doi: 10.1016/j.cognition.2015.03.004
- Vuoskoski, J. K., and Eerola, T. (2011a). Measuring music-induced emotion: a comparison of emotion models, personality biases, and intensity of experiences. *Musicae Sci.* 15, 159–173. doi: 10.1177/1029864911403367

## ACKNOWLEDGMENTS

We thank the Editor and the two reviewers for helping us to improve this manuscript. We are particularly grateful to Reviewer 1 for detailed and constructive comments.

- Vuoskoski, J. K., and Eerola, T. (2011b). The role of mood and personality in the perception of emotions represented by music. *Cortex* 47, 1099–1106. doi: 10.1016/j.cortex.2011.04.011
- Vuoskoski, J. K., and Eerola, T. (2012). Can sad music really make you sad? Indirect measures of affective states induced by music and autobiographical memories. *Psychol. Aesthetics Creativity Arts* 6, 204–213. doi: 10.1037/a0026937
- Watson, D., and Clark, L. A. (1994). *The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form*. Iowa: The University of Iowa.
- Zentner, M., and Eerola, T. (2010). “Self-report measures and models of musical emotions,” in *Handbook of Music and Emotion: Theory, Research, Applications*, eds P. N. Juslin and J. A. Sloboda (New York, NY: Oxford University Press), 187–223.
- Zentner, M., Grandjean, D., and Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion* 8, 494–521. doi: 10.1037/1528-3542.8.4.494
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Miu, Pițur and Szentágotai-Tătar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# The multisensory body revealed through its cast shadows

Francesco Pavani<sup>1,2\*</sup> and Giovanni Galfano<sup>3,4\*</sup>

<sup>1</sup> Center for Mind/Brain Sciences, University of Trento, Rovereto, Italy, <sup>2</sup> Department of Psychology and Cognitive Science, University of Trento, Rovereto, Italy, <sup>3</sup> Department of Developmental and Social Psychology, University of Padua, Padua, Italy, <sup>4</sup> Center for Cognitive Neuroscience, University of Padua, Padua, Italy

## OPEN ACCESS

### Edited by:

Achille Pasqualotto,  
Sabanci University, Turkey

### Reviewed by:

Matthew R. Longo,  
Birkbeck, University of London, UK  
Francesca Frassinetti,  
University of Bologna, Italy

### \*Correspondence:

Francesco Pavani,  
Center for Mind/Brain Sciences,  
University of Trento, Corso Bettini 31,  
I-38068 Rovereto, Italy  
francesco.pavani@unitn.it;  
Giovanni Galfano,  
Department of Developmental  
and Social Psychology, University  
of Padua, Via Venezia 8,  
I-35131 Padua, Italy  
giovanni.galfano@unipd.it

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 20 February 2015

**Accepted:** 07 May 2015

**Published:** 19 May 2015

### Citation:

Pavani F and Galfano G (2015)  
The multisensory body revealed  
through its cast shadows.  
Front. Psychol. 6:666.  
doi: 10.3389/fpsyg.2015.00666

One key issue when conceiving the body as a multisensory object is how the cognitive system integrates visible instances of the self and other bodies with one's own somatosensory processing, to achieve self-recognition and body ownership. Recent research has strongly suggested that shadows cast by our own body have a special status for cognitive processing, directing attention to the body in a fast and highly specific manner. The aim of the present article is to review the most recent scientific contributions addressing how body shadows affect both sensory/perceptual and attentional processes. The review examines three main points: (1) body shadows as a special window to investigate the construction of multisensory body perception; (2) experimental paradigms and related findings; (3) open questions and future trajectories. The reviewed literature suggests that shadows cast by one's own body promote binding between personal and extrapersonal space and elicit automatic orienting of attention toward the body-part casting the shadow. Future research should address whether the effects exerted by body shadows are similar to those observed when observers are exposed to other visual instances of their body. The results will further clarify the processes underlying the merging of vision and somatosensation when creating body representations.

**Keywords:** shadow, spatial attention, multisensory, body perception, self-recognition, touch, vision

## Introduction

The processing of shadows has been the target of an increasing number of studies in recent years. The results stemming from this line of investigation have demonstrated that information conveyed by shadows can support several tasks performed in everyday life. It is now well established that our visual system can process shadows very rapidly (e.g., Elder et al., 2004; Rensink and Cavanagh, 2004) and use shadows for several visual functions (see Mamassian et al., 1998; Dee and Santos, 2011; for reviews). For instance, it has been shown that shadows can foster object recognition (Norman et al., 2000; Mascialzoni et al., 2009). Moreover, several studies have shown that cast shadows of objects can play a critical role in defining the spatial arrangement of objects within a scene, in both dynamic and static contexts (e.g., Kersten et al., 1997; Yonas and Granrud, 2006; Imura and Tomonaga, 2009). Furthermore, reaching movement kinematics can also be affected by the shadow casted by the target object (Bonfiglioli et al., 2004).

One very special class of objects casting shadows in the environment is represented by human bodies. Others that we perceive in the visual scene often cast shadows of their body or body parts. Moreover, our own body is frequently a source of shadows, projecting images of our bodily self in the environment. It is now widely acknowledged that full bodies or body parts represent special

stimuli for the brain and they are processed by specialized neural pathways (e.g., Downing et al., 2001; Arzy et al., 2006; Pourtois et al., 2007; Calvo-Merino et al., 2010; Cazzato et al., 2015). This special salience of body-related stimuli is also well reflected in behavioral effects, which suggest that body parts undergo prioritized processing compared to other objects (e.g., Ro et al., 2007; Igarashi et al., 2008, 2010), especially when they belong to one's own body (e.g., Frassinetti et al., 2009; Ferri et al., 2011).

Recently, researchers have asked whether shadows cast by body parts may represent a unique class of stimuli for the visuomotor system. More specifically, the focus of research has covered two related, yet distinct issues, i.e., the generic effects of someone else's body shadow vs. the specific effects of one's own body shadows on cognitive processing. One first relevant question in this literature is whether shadows cast by bodies can also undergo prioritized processing compared to other objects—similar to what has been documented for visible instances of bodies. A second important question is whether body shadows may trigger reflexive orienting of attention toward the body that casts them. It has been proposed that, when seeing a cast shadow, our visual system is somehow forced to find an association between the visible shadow and the object that most likely casts it, thus solving the so-called “shadow correspondence problem” (Mamassian, 2004). While for generic objects this could serve the main purpose of reducing the perceptual complexity of the visual scene by promoting perceptual bindings between segmented elements, in the case of body shadow it could serve a different yet fundamental function: deciding which visible instances of bodies in the scene belong to the self and which belong to others. When applied to body shadows, the shadow-correspondence problem may thus be central to a perceptual decision that ultimately promotes self-identification and self-recognition.

The primary aim of the present review is to provide a comprehensive perspective of the studies that examined how cast shadows of bodies affect our cognitive processes. We will first discuss the limited literature on the influence on behavior of shadows cast by the body of others, and then we will turn to the issue of shadows cast by our own body. This organization has been adopted with the goal of introducing the more specific topic of the review (the influence of one's own shadows when creating body representations) starting from a more general perspective. In particular, we will examine (1) how cast shadows of our own body can change a sense of bodily space, by promoting binding between personal space and the space occupied by one's own shadow; (2) whether cast body shadow of our own body can “push” attention toward the body itself; (3) the extent to which this orienting effect may occur automatically. We will conclude by discussing the implications of this literature for the study of body perception in general and outlining some possible development of this research field, which is still in its infancy.

## The Effects of Someone Else's Body Shadows

Research in this subtopic has primarily converged on the attempt to address the basic question of whether someone else's body

shadow can affect one's motor behavior. Tentative evidence supporting a positive answer was provided by Liden and Herberholz (2008), who investigated whether fake shadows resembling the body of a predator might influence movement in crayfishes. To this purpose, they used an experimental setting in which an object moving at different velocities effectively mimicked the shadow of an attacking predator. Crayfishes exhibited two different types of escape responses whose prevalence critically depended of the velocity of the moving shadow.

As concerns humans, Alaerts et al. (2009) conducted a study in which participants were required to watch video clips in which either a hand of a stranger or its cast shadow were shown executing abduction/adduction movements of the index finger while transcranial magnetic stimulation was administered over the hand-related area of the primary motor cortex and electromyographical activity was recorded from the muscle of the participants' index finger. Motor-evoked potentials showed an increased amplitude for both the real hand and the hand shadow conditions as compared to when movements were performed by an unrecognizable object (control condition). This pattern of results has been taken to support the idea that visible body parts and body shadows alike are sufficient to activate motor areas, as long as a biological movement is implied. In a similar study which combined electromyography and transcranial magnetic stimulation, Sartori and Castiello (2013) addressed the mirror neuron system's ability (for a review, see Rizzolatti and Craighero, 2004) to resonate with movements shown in full illumination vs. shadowed movements, in which the hand performing a reach-and-grasp sequence was shown with the little finger in shadow. Note that in this study the manipulation involved attached rather than cast shadows (i.e., shadows falling on the body, rather than the shadow projected by the body). Motor-evoked potentials for shadowed movements exhibited a decrease in amplitude as compared to the full illumination condition. Sartori and Castiello (2013) interpreted this finding as suggesting that body shadow processing can be reflected at the level of the human mirror neuron system, even when shadows are not relevant for the task at hand.

Turning to behavioral studies, recent evidence has been reported indicating that observing a cast shadow of one hand can affect imitative behaviors in humans. Badets et al. (2013) presented their participants with two superimposed visual stimuli (one on the foreground and the other on the background). One of the two stimuli depicted a hand and the other depicted its cast shadow. The participants were required to imitate the movement (opening vs. closing the fingers) of one stimulus (the target) while ignoring the other (the distractor). Crucially, there were congruent trials (in which the hand and the shadow performed the same movement) and incongruent trials (in which the hand and the shadow executed opposite movements). In addition, there was a real shadow condition (in which the shadow always appeared on the background and the hand appeared in the foreground), and a no-shadow condition (in which the shadow appeared in the foreground and the hand appeared on the background, i.e., a situation which is known to break one of the shadow priors, see, e.g., Casati, 2003). A response time distributional analysis demonstrated that participants suffered from an interference effect (i.e., they were



slower in initiating movements on incongruent trials as compared to congruent trials). Crucially, this effect vanished for the slowest responses in the real shadow condition only. Badets et al. (2013) have argued that imitation abilities can be deeply influenced by body shadows. They interpreted the fact that interference was present (also for slowest responses) in the no-shadow condition as suggesting that participants likely treated these stimuli as real hands.

Recently, the role of body shadows cast by others has been investigated also in the context of computer vision and robotics (Dee and Santos, 2011). A particularly interesting applied research domain in this regard is related to person identification for vision-based surveillance systems. Aerial search and surveillance systems typically rely on a top view of the human body, with much less details than in side views. Iwashita et al. (2012) have demonstrated that shadows provide additional information regarding body biometrics that enhance person identification and gait recognition both inside a building (using artificial light) and outside (under the natural sunlight). It would be interesting to extend this line of research to animal species that use aerial view (e.g., birds), to explore to what extent cast shadow can also constitute a cue for object recognition. Furthermore, although humans typically do not see other humans from an aerial perspective, it would be interesting to examine to what extent adding shadow stimuli could promote recognition of people in natural scenes (e.g., Reeder and Peelen, 2013).

## One's Own Body Shadows Bind Personal and Extrapersonal Space

The data reviewed in the previous section indicate that body shadows (of others) can have a strong impact on the visuomotor system, in both humans and other animal species. One's own body shadows, however, may be even more salient. Each shadow cast by our own body broadly refers to a location (the body part casting it) for which we have exteroceptive, proprioceptive and interoceptive experience. This feature makes body shadows potentially capable to contribute to the construction of the internal representation of body shape and its extension in space.

A pivotal role in starting this line of investigation has been played by the work of Pavani and Castiello (2004). In their experiments, Pavani and Castiello used a very popular experimental setting in multisensory research, that is the visuo-tactile interference paradigm (e.g., Pavani et al., 2000; Spence et al., 2004a,b). The participants performed a tactile elevation discrimination task (with thumb and index finger arranged one below the other, judge which of the two fingers was stimulated) while ignoring a simultaneous task-irrelevant visual stimulus. The typical finding observed with this setting is that tactile localization performance is worse when tactile and visual stimuli occur at different elevations (e.g., touch at the index, vision at the thumb) compared to when they occur at the same elevation (e.g., touch and vision both at the index finger). Crucially, this visuo-tactile interference is greater when the visual distractors are presented near the stimulated hand, compared to when they are presented further away from the body (Spence et al., 2004a).

Interestingly, Pavani and Castiello (2004) observed that task-irrelevant visual stimuli presented far and equidistant from both hands but in close proximity to the shadow cast by one of the two hands produced a much stronger interference effect when tactile targets were delivered to the hand casting the shadow as compared to when they were presented at the other hand. Such modulation was genuinely related to body shadows, as it vanished when participants wore a shaped glove projecting an unnatural polygonal shadow or viewed a line drawing silhouette of a hand. Pavani and Castiello (2004) argued that participants reacted to the visual stimuli near the shadow of the hand as if the stimuli were affecting the hand itself. Also in consideration of previous reports that visuo-tactile interference can be observed also when visual distractors are presented to fake hands aligned to the real hands (see Pavani et al., 2000) and that it can be influenced by active tool-use (e.g., Maravita et al., 2002a), Pavani and Castiello (2004) have interpreted the magnification of visuo-tactile interference as evidence that body shadows may create some sort of binding between personal and extrapersonal space (i.e., the space occupied by the body and the space occupied by the shadow, respectively).

The notion that our own body shadows can be incorporated into our personal multisensory space of the self (see Cardinali et al., 2009; de Vignemont, 2011; for reviews), has recently been supported also by findings reported by Kuylen et al. (2014), who used a perceptual matching task. Based on the idea that the ability to interact with an object at any distance shrinks the perceived distance between object and observer (e.g., Witt et al., 2005), Kuylen et al. (2014) tested whether viewing the shadow of one's own body extending toward a target object may result in the subsequent underestimation of the distance between the body and the same target object. The results confirmed that, compared to a baseline condition in which no body shadow was visible, the participants exhibited an estimation bias to report a shorter distance when the body shadow was present. Interestingly, this phenomenon, was also reliable when participants interacted with the target object by means of a tool (a laser pointer), but it did not emerge when the body shadow was replaced by the shadow projected by a different object (a large file cabinet placed behind the participant which covered the shadow cast by the body). This latter finding clearly indicates that cast shadows of our own body are different from other types of shadows and suggests that they may indeed act as extensions of the body, as originally proposed by Pavani and Castiello (2004).

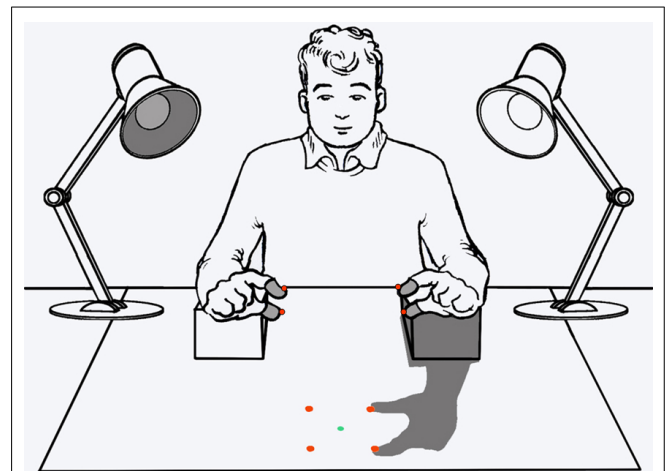
Before exploring the effects of one's own body shadows for body perception further, it is worth noting that owned body shadows have also been studied in applied cognitive science, especially in the context of user interface research. Devices exploiting shadows cast by the body of users have been implemented for operating graphical information on large displays (e.g., Xu et al., 2006). These shadow-based interfaces enable users to interact with a computer by simply using the shadows cast on the screen by the upper limbs (and more specifically by the fingers). Takeuchi et al. (2014) have demonstrated that body shadows can be very effective as pointing cursors. This may be due to the fact that users do not have particular difficulties in understanding the correspondence

between the movement of the fingertips and the movement of the related cast shadow. Specifically, the cognitive ergonomics validity of using one's own body shadows for the interaction with distal surfaces may relate to the natural tendency of our cognitive system to bind personal and extra-personal spaces through one's own body shadows.

The research reviewed so far, stemming from different disciplines and perspectives, highlights that body shadows are highly peculiar stimuli. Interestingly, unlike tools or other objects such as rubber hands, they are immaterial and can only provide visual information (they are not multisensory stimuli). Another important point is that, unlike other objects that are capable of shaping the subjective extension of the body in space, the type of visual information they convey is quite coarse, being only two-dimensional. Although the two-dimensional nature of cast shadows does not prevent extracting useful three-dimensional information about the object casting it (Norman et al., 2009), the correspondence between the 2D cast body-shadow and the 3D body part remains underspecified. There cannot be a 1:1 mapping between points on the shadow and points on the body. The shadow of one's head, for instance, could relate to either the front or the back of the head (we thank one of the reviewers for this interesting remark).

## One's Own Body Shadows Shift Attention to the Body

Pavani and Galfano (2007, Galfano and Pavani, 2005; Pavani et al., 2014) have addressed another critical possibility concerning the role of body shadows, namely, the possibility that they can serve as important cues to the multimodal sense of body. Galfano and Pavani (2005) hypothesized that body shadows may indeed represent a high-priority class of stimuli that act by "pushing" attention toward the body itself. To this purpose, they modified the paradigm used by Pavani and Castiello (2004) to implement an exogenous or reflexive spatial cueing paradigm (e.g., Jonides, 1981; see Spence and Santangelo, 2009; for a review in the context of multisensory research), in which hand shadows served as spatially uninformative visual cues (**Figure 1**). The participants were delivered tactile targets unpredictably to the thumb or index finger of either hands and were asked to localize them irrespective of the stimulated hand. At the same time, they viewed the shadow of either the touched or untouched hand cast in front of them by a lateral light source. In the first experiment, the hand casting the shadow remained fixed within a block of trials, but the participants were explicitly told that the tactile target had the same probability to be delivered on the hand casting shadow and in the other hand. This, in turn, made the shadow entirely irrelevant for the task at hand. Nevertheless, localization performance was better when targets touched the hand casting the shadow (spatially congruent trials) than the other hand (spatially incongruent trials). This pattern was very robust, suggesting that body shadows somehow cued attention back the body part casting it. Although the body shadow conveyed no predictive information about the target location, in a second experiment the hand casting the shadow varied unpredictably from trial to trial. This manipulation had



**FIGURE 1 | Schematic illustration of the experimental setting in the experiments investigating orienting of attention mediated by body shadows adapted from Pavani et al. (2014).** A trial with a left-hand shadow is shown. Tactile stimulators are embedded in the gray sheaths around the index and thumb of each hand. Red LEDs were used only in experiments addressing visual modality in personal (at the hand) and extrapersonal (near the shadow cast by the hand) space and were illuminated one at a time. The green LED served as fixation point. See the text for details.

the purpose of discouraging participants from adopting implicit strategies to deliberately attend to the hand casting the shadow. The results, again, showed that tactile localization performance was significantly better at the hand casting shadow than at the other hand. This finding was taken as evidence that the attentional cueing effect toward the body part casting the shadow was indeed genuinely reflexive rather than the consequence of some top-down strategy.

Galfano and Pavani (2005, Experiment 4) also conducted an experiment in which participants were prevented from seeing their own hands. This manipulation had the purpose of ruling out the possibility that the observed spatial cueing effect resulted from the fact that the visible hand casting the shadow was illuminated more strongly than the other hand. Orienting of attention mediated by body shadows was still present, suggesting that the alternative account could be dismissed. For another experimental condition, in which the cast shadow of an object (a piece of cardboard) overlapped and completely masked any shadow cast by the hand, the data showed no reliable effects. This latter pattern rules out yet a further alternative account which would attribute the better performance on spatially congruent trials over spatially incongruent trials to the fact that the lateralized light source that was turned on to create the shadow might also potentially convey somatosensory (thermal) stimulation to the hand casting the shadow. Such account can be rejected because the asymmetrical thermal stimulation (if any) was present also in the object-shadow condition and no differences in performance emerged. The observed pattern clearly demonstrated that shadow-driven orienting was specific to body shadows. However, it is worth noting that the object shadow condition did not differ from the body shadow condition only for the shape of the shadow. Indeed, the object shadow was stationary throughout each block

of trials, whereas body shadow was obviously spatio-temporally correlated to the movements, if any, of the hand (likely inducing a sense of agency).

The possible role of the sense of agency (for a review, see Tsakiris et al., 2007) as a key factor for accounting for the orienting of attention mediated by body shadows reported by Galfano and Pavani (2005) has been explored by Pavani and Galfano (2007). They specifically addressed whether shadow-induced benefits on tactile localization performance were dependent on the correspondence between the seen shadow and the object casting it, that is self attribution of the visible image (i.e., shadows) of the body. To this aim, they implemented three different cue conditions in their experimental set up. Beside the standard hand-shadow condition, similar to Pavani and Castiello (2004), Pavani and Galfano (2007) also included a condition in which participants wore a shaped glove casting an unnatural polygonal shadow (real shadow with unnatural shape), and a condition in which participants were presented with photographs consisting of shadow-like images projected from above (fake shadow with natural shape). This allowed dissociation of two different factors that may be at work for endorsing self attribution of shadows and to estimate their impact in isolation. In the real shadow with unnatural shape condition, self attribution, if any, was promoted by spatio-temporal movement correlation between hands and their shadows alone. In contrast, in the fake shadow with natural shape condition, the only factor at work was represented by the visual similarity between the hands and the (static) shadow-like images. Overall, participants exhibited a significantly faster tactile localization performance for cued over uncued hands only for the real shadow condition. In a more in-depth analysis aimed to uncovering possible fluctuations of orienting of attention mediated by body shadows within each block of trials, Pavani and Galfano (2007) observed an interesting pattern of data also for the other experimental conditions. The analysis revealed that for the fake-shadow with natural shape condition, a reliable shadow-mediated orienting effect was present in the first part of each experimental block. In sharp contrast, in the real shadow with unnatural shape condition this effect was significant in the last portions of each block only. The overall findings were taken as strong evidence that orienting of attention mediated by body shadows is critically bound to self attribution of shadows. The temporally diverging trend for the fake-shadow with natural shape condition and the real shadow with unnatural shape condition was interpreted as evidence that the sense of ownership of shadows is strongly mediated by both spatio-temporal correlation between hands and shadows (i.e., a sense of agency) and visual similarity, although these two factors operate in a different fashion (e.g., van den Bos and Jeannerod, 2002; Whiteley et al., 2004; Tsakiris et al., 2005, 2006).

Another critical question addressed by Pavani and Galfano (2007) is whether attention shifts induced by body shadows comprise the whole portion of visual space they occupy or the body part referred to by shadows exclusively. In so doing, they modified the basic paradigm by adding visual targets located at either the external boundaries of the hand shadow (i.e., close to fixation and far from the hand), or at the index finger and thumb of both hands. The results showed that, overall, shadow-mediated orienting was reliable for tactile targets only, strongly suggesting

that body shadows push attention to the body part they refer to, rather than cueing the portion of space they cover.

## The Attentional Link between One's Own Shadows and the Body is Fast and Mandatory

One important question that arises from the studies that showed that one's own body shadows can orient attention to the body—and specifically to touches on the body—is whether this effect is mandatory. Recently, Pavani et al. (2014) have addressed the automaticity of attention shifts elicited by body shadows by focusing on two different features that are widely assumed to characterize exogenous orienting of spatial attention: the speed of attention orienting and its sensitivity to contextual modulations. It is important to reiterate that in all the experiments reported by both Galfano and Pavani (2005) and Pavani and Galfano (2007), the body shadow effectively cued attention to the body part casting the shadow despite shadow being spatially non-predictive of the target location. While this is considered a critical feature of reflexive orienting (e.g., Galfano et al., 2011, 2012), another feature that is often deemed as a hallmark for automatic processing is that this type of orienting typically results in a very early rising effect (e.g., Müller and Rabbitt, 1989; Cheal et al., 1994). In behavioral studies, this latter feature is reflected in the observation of a significant benefit in performance for spatially congruent over spatially incongruent trials with very short (below 200 ms) cue-target stimulus onset asynchrony (SOA). Because both Galfano and Pavani (2005) and Pavani and Galfano (2007) invariably used a fixed 2750-ms SOA between cue (the cast shadow of the hand) and target (tactile or visual), Pavani et al. (2014) manipulated SOA and included also a 100-ms SOA. This very short SOA is known to reveal reliable spatial orienting effects with other types of attentional cues, such as eye gaze (e.g., Driver et al., 1999; Galfano et al., 2012). The results showed a robust orienting of attention mediated by body shadows early in processing and sustained over time, as it was not modulated as a function of SOA for both tactile targets (Pavani et al., 2014; Experiment 1) and visual targets delivered near the shadow and far from the hands, i.e., in extrapersonal space (Pavani et al., 2014; Experiment 2).

The second feature addressed by Pavani et al. (2014) was whether shadow-driven orienting is resistant to contextual modulations. It is a widely shared assumption that strongly automatic processing should be impervious to changes in the experimental setting and task demands (e.g., Zbrodoff and Logan, 1986; Ristic and Kingstone, 2005; Pavan et al., 2011). Pavani et al. (2014) addressed this criterion of automaticity by intermixing target modality in the same experiment. Unlike previous experiments, in which target modality remained fixed, their participants responded to unpredictable tactile and visual targets. These latter targets were delivered near the shadow and far from the hands (i.e., in extrapersonal space; Pavani et al., 2014, Experiment 3) or directly at the hands (i.e., in personal space; Pavani et al., 2014, Experiment 4). The results showed a reliable orienting of attention mediated by body shadows for tactile targets in agreement with the previous studies in which touch was the only target modality

(Galfano and Pavani, 2005; Pavani and Galfano, 2007). However, the effect for targets in the visual modality became inconsistent, irrespective of whether they appeared in personal or extrapersonal space (i.e., near or far from the hand). Overall, these findings provide support for the notion that orienting of attention mediated by body shadows for tactile targets is a strongly automatic phenomenon, as it appears early in processing and is unaffected by contextual changes (e.g., Zbrodoff and Logan, 1986). In sharp contrast, orienting of attention by body shadows was visible for visual targets (in extrapersonal space) only to the extent that sensory modality was fixed. Hence, orienting to visual targets cannot be said to be strongly automatic as it is clearly sensitive to contextual manipulations (also see Pavani and Galfano, 2007).

Taken together, the studies on orienting of attention triggered by our own body shadows indicate that cast shadows of body parts may indeed represent a high-priority class of stimuli. They act by “pushing” attention toward the body itself and this effect has the characteristics of a mandatory process, at least for the tactile modality. Seeing our own body shadow is a powerful cue toward tactile sensations at the body part casting the shadow. The effect is also influenced by self-attribution of the cast shadow: its presence is tightly linked to perceived ownership of the cast shadows. When this attribution fails, cast shadows can quickly become ineffective as a cue for attention. In the next paragraphs, we examine the extent to which the effects observed for body shadows may extend to other types of visual instances of the body in the environment.

## Are Body Shadows Special?

One important issue in relation to the observations reviewed here for body shadows is to the extent to which they imply mechanisms specific to shadows only, or instead constitute examples of more general processes, such as those involved in multisensory body perception. Consider, for instance, the shadow correspondence problem briefly illustrated in the Introduction section. The problem for the cognitive system is to find the correct correspondence between the seen cast shadow and the object in the environment to which it belongs (Mamassian, 2004). The findings reviewed above, showing that vision of task-irrelevant shadows of one’s own body automatically triggers attention orienting to touches on the body, might stem from the solution of the body-shadow correspondence problem. This interpretation would link the observed findings to a process which has been proposed specifically for cast shadows. An alternative possibility, however, is that a somewhat similar process exists also whenever we experience visible body parts in the environment. During our waking life, images of our own body are almost always present and available in first person perspective. Furthermore, we have third-person views of ourselves through mirrors, photos, videos and nowadays also virtual-reality setups and avatars. Because the body of others is also a frequent stimulus in the environment, occasionally in first-person view and most often in third-person view, choosing which of these visual instances of bodies correspond to our own corporeal awareness is a fundamental task that our cognitive system is constantly asked to solve.

By analogy with the shadow correspondence problem, one could argue for a more general “visible-body correspondence

problem,” and posit the existence of a cognitive process whose aim is to correctly match the seen bodies with our own corporeal awareness. This process would involve binding instances of the body across sensory modalities (vision and somatosensation) and, sometimes, across different spatial locations (extra-personal and personal)—just like it occurs with body shadows. Solving the visible-body correspondence problem could be at the roots of the discrimination between body images that belong to oneself and body images that belong to others, strengthening self-other distinction, bodily self-recognition and, ultimately, the psychological experience of the self.

Thus, the key question is whether we can generalize from the body-shadow correspondence problem to a more general “visible-body correspondence problem.” If this is the case, it should be possible to find parallels between the results that emerged from the literature on body shadows and the more general literature on multisensory body perception. Specifically, there should be evidence (1) that a seen body part in the environment (i.e., a photograph or video of one’s own hand) “pushes” attention to the corresponding body part; (2) that this process occurs particularly for touch (or somatosensation); and (3) that this process is largely automatic. As we shall see in the next paragraphs, although several studies in the literature do suggest that visible body parts can affect somatosensory processing, parallels between the findings reviewed here for own body shadows and the studies on own pictorial images of the body are still limited.

When searching for effects of seen body parts on tactile perception one key phenomenon described in the literature is the so-called “Visual enhancement of touch” (VET). VET emerges as improved tactile detection and discrimination at a specific body part (typically a hand), when the body part is either seen directly (Kennett et al., 2001; Taylor-Clarke et al., 2002, 2004; Press et al., 2004; Whiteley et al., 2004) or through a pictorial representation (either video or photograph; Tipper et al., 1998, 2001). Critically, VET emerges despite the fact that vision of the body part is completely task-irrelevant and uninformative about somatosensation. This multisensory effect has been reported in neurologically healthy participants, but there is also evidence that vision of body parts can ameliorate the somatosensory deficits in brain-damaged patients (Serino et al., 2007; see also Rorden et al., 1999; for related findings with vision of a rubber hand).

In many VET studies, the importance of self-attribution of the seen hand remains unclear. This is because tactile enhancements were measured as the difference in performance between a condition in which participants observed an owned body part vs. an object. This contrast does not allow to determine whether the crucial factor is seeing “a” hand, or seeing the “owned” hand (for discussion see Longo et al., 2008). There have been two attempts to address this issue. Haggard (2006) asked participants to discriminate the orientation of gratings delivered to the index finger tip, under three different viewing conditions. Participants either viewed their own hand, or viewed a neutral object, or viewed the hand of a third person aligned with the tactually stimulated hand. Compared to the viewing of a neutral object, both viewing one’s own body part and viewing the body part of another person produced orientation discrimination enhancements. This finding seems to suggest that VET can generalize also to the viewing of



body parts that belong to others. In a subsequent study, however, VET emerged specifically for the self-attributed visible body parts. Longo et al. (2008) asked participants to perform a similar orientation discrimination task, while viewing a rubber hand that appeared in the felt location of the real hand through a mirror. To manipulate the perceived ownership of the visible rubber hand, they stimulated the real and fake hands in synchrony (leading to an illusion of ownership) or out of synchrony (no illusion of ownership) across blocks (also see Botvinick and Cohen, 1998). The results showed that VET boosted performance particularly for those participants who performed the tactile discrimination task near threshold. Importantly, they also showed that among participants performing near threshold, VET was larger when the rubber hand was self-attributed compared to when it was considered an extraneous body part. Thus, it appears that under certain circumstances modulations of tactile performance in the presence of visible body parts can be strengthened by self-attribution, similar to the case of body shadows.

In the typical VET experiment, the viewing condition is continuous during the entire block of trials. Tipper et al. (2001, Experiment 2) tested VET using an experimental setup that allowed timed presentation of the visible body part and control over the temporal interval between the onset of the viewing condition and the tactile stimulation. They used three cameras to project displays of different body parts of the participant: face, neck or hand. On each experimental trial, one of these visual displays was shown, either 200 or 700 ms before the tactile target. Participants were instructed to detect touches at a specific body part (e.g., the face), while ignoring distractors at another body site (e.g., the neck). The results showed that response speed advantages emerged regardless of the onset asynchronies between the visible body part and the tactile target. A recent EEG study by Cardini et al. (2012) has provided consistent evidence that VET reflects a phasic effect and can be elicited by very brief exposure to one's body part. Overall, these findings on VET are reminiscent of the early-rising effect of body shadows on tactile targets documented by Pavani et al. (2014). In that study, orienting of attention mediated by body shadows occurred even at the 100 ms SOA, and this effect was particularly stable and robust for tactile targets.

A different, yet related, line of research worth mentioning is the one that explored the interpretation of mirror reflection of body parts. To correctly interpret mirror-reflections, our brain needs to understand that the object that appears in the mirror (e.g., our face) occupies in fact a different location in space. This process is clearly similar to the shadow-correspondence problem, and it is probably the closest match to the cognitive mechanism at play when interpreting shadows in the environment. Furthermore, it is classically considered evidence of self-awareness in human development and ethology (Gallup, 1982). Interestingly, there is evidence that human infants typically succeed in interpreting mirror-reflections of themselves by their 2 year of life (Gallup et al., 2002). As for body shadows, shadow self-recognition appears to emerge at age 3 (Cameron and Gallup, 1988).

Using the visuo-tactile interference paradigm later used also by Pavani and Castiello (2004) for body shadows, Maravita et al. (2002b) explored the interpretation of mirror reflections of body

parts. They asked participants to perform a speeded spatial discrimination for touches at the hands, while ignoring concurrent visual distractors. Critically, in one condition the visual distractors were physically close to the participant's hands, but were seen only as distant mirror reflections; in another condition they were physically in far space, and appeared near a dummy hand or the hand of another person. The results showed that the strongest visuo-tactile interference emerged for the mirror condition, suggesting that participants recoded the true source of the visual distractors near the body. Similar to the body-shadow studies, vision of the hands (mirror reflected, dummy, or someone else's) was completely task irrelevant. One interpretation of this finding is that participants mandatorily remapped the self-attributed hand to the actual space the hand occupied, hence coding visual distractors close to the mirror-reflection of the hands from far to near space.

## Future Directions for a Novel Research Field

Research on body shadows is still in its infancy. However, it has the potential to provide a window onto the cognitive and neural mechanisms that regulate the multisensory construction of body representation and the bodily-grounded sense of the self. More generally, it can provide useful insights on the multisensory representation of space, on shadow perception in general, or even on the principles that make shadows a useful and ergonomic tool for human-computer interfaces. While these multiple directions are all worth exploring, we suggest here four possible future developments for this new research domain.

The first one, builds on the considerations offered in the previous section, and is concerned with the possibility of exploring the effects of body shadows on somatosensory perception and self-processing further, with the goal of finding parallels between processing of body shadows and processing of other seen instances of the body in the environment. For instance, it would be very interesting to understand the extent to which body shadows and other seen instances of the body could trigger attention to somatosensation in general. At the moment, all the studies conducted on body shadows examined their effects for spatial touch. Whether similar cueing effect could also exist for other aspects of somatosensation, such as pain perception, proprioception or interoception is unknown. Exploring this aspect would help understanding the extent to which seeing body shadows may be a cue for all bodily sensations—i.e., a cue for the body in general. Interestingly, indications that seeing one's own body, through direct vision or mirrors, can affect somatosensation in general and not just touch, are already available in the literature. For instance, looking at an image of ourselves in the mirror has been shown to improve the perception of heart-beat signals, and specifically heart-beat counts which are considered a proxy of the person's ability to pay attention to interoceptive signals (Ainley et al., 2012). There is also evidence that vision of one's own body parts can modulate pain perception (Longo et al., 2009, 2012; Romano et al., 2014). Another line of investigation within this aim of finding parallels between perception of body-shadow and perception of other visible instances of the body,

is related to the validation of the existing findings obtained for shadows of body-parts, to shadows of the whole-body. In recent years, seminal works using virtual reality approaches have already taken the study of multisensory body perception in this direction using whole-body illusions (e.g., Ehrsson, 2007; Lenggenhager et al., 2007; Slater et al., 2010) and the study of the interactions between whole-body and body-part perception is also very promising (Liang et al., 2015). At present, however, there has been no attempt to explore the effects of whole-body shadows on body perception.

A second direction worth exploring concerns the neural correlates of body shadow perception. A number of studies in the last decade have examined the neural correlates of visible bodies or body-parts (e.g., Pourtois et al., 2007; Cazzato et al., 2015). In addition, studies have documented the influences of visible body parts on somatosensory processing, primarily exploring the neural correlates of the VET effect described above (Macaluso et al., 2000; Taylor-Clarke et al., 2002; Sambo et al., 2009; Gillmeister and Forster, 2010). These studies have revealed that task-irrelevant vision of body parts can modulate somatosensory processing, including the earliest stages involving the primary somatosensory cortex (Longo et al., 2011). It would be informative to unravel whether the same neural mechanisms described for the visible bodies are also recruited during vision of body shadows. Also, given the importance of self-attribution of body shadows in cueing attention to the body (Pavani and Galfano, 2007), it would be interesting to explore the role of possible right-hemispheric specializations for self-processing (Keenan et al., 2000; Sugiura et al., 2005) using a neuropsychological approach. Behavioral evidence has suggested that implicit self-attribution of seen body parts can enhance performance on match-to-sample body discrimination tasks—a phenomenon which has been labeled “self-advantage” (Frassinetti et al., 2008). The use of this paradigm in brain-damaged patients has revealed an interesting dissociation, with left-brain damaged patients retaining self-advantage in body discrimination tasks, whereas right-brain damage patients performing equally regardless of whether the seen body part belong to themselves or not (Frassinetti et al., 2008, 2010; see also Frassinetti et al., 2012; for similar results in brain-damaged children from the age of 4 years). If right-hemispheric lesions undermine implicit self-recognition, then they should also impair the mechanisms of orienting of attention toward the body triggered by self-attributed body-shadows.

A third direction concerns the effects of body shadows of others. As reviewed above, the literature on this topic is currently very limited and it has primarily explored the consequences on motor behavior of participants observing images of others or images of their shadow acting in the environment. As already anticipated, it would be interesting to examine to what extent body shadows of others could promote person recognition in complex natural scene (Reeder and Peelen, 2013). Furthermore, building on the literature on one's own body shadows, it might be expected that body shadows of others could also trigger attention to the individuals that cast them—a process that could in itself also foster detection of conspecifics in the environment.

Finally, moving from mechanisms of body perception to more general mechanisms of visual processing, it would be interesting

to understand whether some of the principles that have emerged from the literature on body-shadow could apply also to processing of shadows cast by non-bodily objects. For instance, there is evidence that shadows can be treated as objects in the scene (albeit at a coarse spatial scale; see Lovell et al., 2009) and as such can favor within-object advantages for attention orienting (de-Wit et al., 2012). It is unknown, however, whether the cast shadow and the object to which it belongs are bound together at some stage of visual processing, into a unique perceptual entity. The literature on body shadows would suggest that this is the case and that cueing the shadow could result in attention being directed to the object, but this is currently an open empirical question.

## Conclusion

In the present review we pursued two aims. First, we attempted to provide the first systematic account of the effects of body shadows on behavior, considering both the studies that examined the effects of body shadows cast by other people and the relatively larger literature on the effects of body shadows cast by one's own body. The latter literature, in particular, revealed that shadows cast by one's own body can promote binding between personal and extrapersonal space and can orient attention toward the body-part casting the shadow. These effects emerge despite body shadows being completely task-irrelevant and they conform to several of the features that characterize automatic processes.

The second aim of the present review was to examine to what extent the effects documented for body shadows may be specific to shadows only or may also extend to other multisensory processes involving the body perception and attention. Although we delineated possible parallels between the effects of cast shadows of one's own body and the effect of viewing other visual instances of one's own body, it is clear that this remains an open empirical question. We believe that addressing this issue in future studies will be highly informative. If processing of body shadows is somewhat unique, then this would imply the existence of a cognitive and neural mechanism that developed (perhaps through phylogenesis) to quickly resolve and exploit the redundant information provided by cast shadows in the environment. In this scenario, it would be important to assess whether such a process is selective for body shadows or generalizes to the processing of shadows cast by any of the objects in the environment. By contrast, if processing of body shadows is similar to that involved in the processing of other visual instances of the body, then the studies reviewed here could offer insights into the more general mechanisms that subtend the complex but necessary task of merging vision and somatosensation when constructing body representations.

## Author Contributions

All authors provided substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; and contributed in drafting the work or revising it critically for important intellectual content; and approved the final version for publication; and agreed to be accountable

for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The authors are grateful to Paola Rigo and Tommaso Sega for the artwork in **Figure 1**.

## References

- Ainley, V., Tajadura-Jiménez, A., Fotopoulou, A., and Tsakiris, M. (2012). Looking into myself: changes in interoceptive sensitivity during mirror self-observation. *Psychophysiology* 49, 1504–1508. doi: 10.1111/j.1469-8986.2012.01468.x
- Alaerts, K., Van Aggelpoel, T., Swinnen, S. P., and Wenderoth, N. (2009). Observing shadow motions: resonant activity within the observer's motor system? *Neurosci. Lett.* 461, 240–244. doi: 10.1016/j.neulet.2009.06.055
- Arzy, S., Thut, G., Mohr, C., Michel, C. M., and Blanke, O. (2006). Neural correlates of embodiment: distinct contributions of temporo-parietal junction and extrastriate body area. *J. Neurosci.* 26, 8074–8081. doi: 10.1523/JNEUROSCI.0745-06.2006
- Badets, A., Toussaint, L., Blandin, Y., and Bidet-Ildei, C. (2013). Interference effect of body shadow in action control. *Perception* 42, 873–883. doi: 10.1068/p7502
- Bonfiglioli, C., Pavani, F., and Castiello, U. (2004). Differential effects of cast shadows on perception and action. *Perception* 33, 1291–1304. doi: 10.1068/p5325
- Botvinick, M., and Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature* 391, 756. doi: 10.1038/35784
- Calvo-Merino, B., Urgesi, C., Orgs, G., Aglioti, S. M., and Haggard, P. (2010). Extrastriate body area underlies aesthetic evaluation of body stimuli. *Exp. Brain Res.* 204, 447–456. doi: 10.1007/s00221-010-2283-6
- Cameron, P. A., and Gallup, G. G. (1988). Shadow recognition in human infants. *Infant Behav. Dev.* 11, 465–471. doi: 10.1016/0163-6383(88)90006-9
- Cardinali, L., Brozzoli, C., and Farnè, A. (2009). Peripersonal space and body schema: two labels for the same concept? *Brain Topogr.* 21, 252–260. doi: 10.1007/s10548-009-0092-7
- Cardini, F., Longo, M. R., Driver, J., and Haggard, P. (2012). Rapid enhancement of touch from non-informative vision of the hand. *Neuropsychologia* 50, 1954–1960. doi: 10.1016/j.neuropsychologia.2012.04.020
- Casati, R. (2003). *The Shadow Club: The Greatest Mystery in the Universe—Shadows—and the Thinkers who Unlocked their Secrets*. New York: Knopf.
- Cazzato, V., Mian, E., Serino, A., Mele, S., and Urgesi, C. (2015). Distinct contributions of extrastriate body area and temporoparietal junction in perceiving one's own and others' body. *Cogn. Affect. Behav. Neurosci.* 15, 211–228. doi: 10.3758/s13415-014-0312-9
- Cheal, M.-L., Lyon, D. R., and Gottlob, L. R. (1994). A framework for understanding the allocation of attention in location-precued discrimination. *Q. J. Exp. Psychol.* 47A, 699–739. doi: 10.1080/14640749408401134
- Dee, H. M., and Santos, P. E. (2011). The perception and content of cast shadows: an interdisciplinary review. *Spat. Cogn. Comput.* 11, 226–253. doi: 10.1080/13875868.2011.565396
- de Vignemont, F. (2011). Embodiment, ownership and disownership. *Conscious. Cogn.* 20, 82–93. doi: 10.1016/j.concog.2010.09.004
- de Wit, L., Milner, D., and Kentridge, R. (2012). Shadow remain segmented as selectable regions in object-based attention paradigms. *Iperception* 3, 150–158. doi: 10.1068/i7164
- Downing, P. E., Jiang, J., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science* 293, 2470–2473. doi: 10.1126/science.1063414
- Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., and Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuo-spatial orienting. *Vis. Cogn.* 6, 509–540. doi: 10.1080/135062899394920
- Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science* 317, 1048. doi: 10.1126/science.1142175
- Elder, J. H., Trithart, S., Pintilie, G., and MacLean, D. (2004). Rapid processing of cast and attached shadows. *Perception* 33, 1319–1338. doi: 10.1068/p5323
- Ferri, F., Frassinetti, F., Costantini, M., and Gallese, V. (2011). Motor simulation and the bodily self. *PLoS ONE* 6:e17927. doi: 10.1371/journal.pone.0017927
- Frassinetti, F., Fiori, S., D'Angelo, V., Magnani, B., Guzzetta, A., Brizzolara, D., et al. (2012). Body knowledge in brain-damaged children: a double-dissociation in self and other's body processing. *Neuropsychologia* 50, 181–188. doi: 10.1016/j.neuropsychologia.2011.11.016
- Frassinetti, F., Maini, M., Benassi, M., Avanzi, S., Cantagallo, A., and Farnè, A. (2010). Selective impairment of self body-parts processing in right brain-damaged patients. *Cortex* 46, 322–328. doi: 10.1016/j.cortex.2009.03.015
- Frassinetti, F., Maini, M., Romualdi, S., Galante, E., and Avanzi, S. (2008). Is it mine? Hemispheric asymmetries in corporeal self recognition. *J. Cogn. Neurosci.* 20, 1507–1516. doi: 10.1162/jocn.2008.20067
- Frassinetti, F., Pavani, F., Zamagni, E., Fusaroli, G., Vescovi, M., Benassi, M., et al. (2009). Visual processing of moving and static self body-parts. *Neuropsychologia* 47, 1988–1993. doi: 10.1016/j.neuropsychologia.2009.03.012
- Gallup, G. G. (1982). Self-awareness and the emergence of mind in primates. *Am. J. Primatol.* 2, 237–248. doi: 10.1002/ajp.1350020302
- Gallup, G. G., Anderson, J. R., and Shillito, D. J. (2002). "The mirror test," in *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*, eds M. Bekoff, C. Allen, and G. Burghardt (Cambridge, MA: The MIT Press), 325–334.
- Galfano, G., Dalmaso, M., Marzoli, D., Pavan, G., Coricelli, C., and Castelli, L. (2012). Eye gaze cannot be ignored (but neither can arrows). *Q. J. Exp. Psychol.* 65, 1895–1910. doi: 10.1080/17470218.2012.663765
- Galfano, G., and Pavani, F. (2005). Long-lasting capture of tactile attention by body shadows. *Exp. Brain Res.* 166, 518–527. doi: 10.1007/s00221-005-2392-9
- Galfano, G., Sarlo, M., Sassi, F., Munafò, M., Fuentes, L. J., and Umiltà, C. (2011). Reorienting of spatial attention in gaze cuing is reflected in N2pc. *Soc. Neurosci.* 6, 257–269. doi: 10.1080/17470919.2010.515722
- Gillmeister, H., and Forster, B. (2010). Vision enhances selective attention to body-related information. *Neurosci. Lett.* 483, 184–188. doi: 10.1016/j.neulet.2010.08.004
- Haggard, P. (2006). Just seeing you makes me feel better: interpersonal enhancement of touch. *Soc. Neurosci.* 1, 104–110. doi: 10.1080/17470910600976596
- Igarashi, Y., Kimura, Y., Spence, C., and Ichihara, S. (2008). The selective effect of the image of a hand on visuotactile interactions as assessed by performance on the crossmodal congruency task. *Exp. Brain Res.* 184, 31–38. doi: 10.1007/s00221-007-1076-z
- Igarashi, Y., Kitagawa, N., and Ichihara, S. (2010). Influence of the body on cross-modal interference effects between tactile and two-dimensional visual stimuli. *Exp. Brain Res.* 204, 419–430. doi: 10.1007/s00221-010-2267-6
- Imura, T., and Tomonaga, M. (2009). Moving shadows contribute to the corridor illusion in a chimpanzee (*Pan troglodytes*). *J. Comp. Psychol.* 123, 280–286. doi: 10.1037/a0015839
- Iwashita, Y., Stoica, A., and Kurazume, R. (2012). Gait identification using shadow biometrics. *Pattern Recognit. Lett.* 33, 2148–2155. doi: 10.1016/j.patrec.2012.07.022
- Jonides, J. (1981). "Voluntary versus automatic control over the mind's eye movement," in *Attention and Performance*, Vol. 9, eds J. B. Long and A. D. Baddeley (Hillsdale, NJ: Erlbaum), 187–203.
- Keenan, J. P., Wheeler, M. A., Gallup, G. G., and Pascual-Leone, A. (2000). Self-recognition and the right prefrontal cortex. *Trends Cogn. Sci.* 4, 338–344. doi: 10.1016/S1364-6613(00)01521-7
- Kennett, S., Taylor-Clarke, M., and Haggard, P. (2001). Noninformative vision improves the spatial resolution of touch in humans. *Curr. Biol.* 11, 1188–1191. doi: 10.1016/S0960-9822(01)00327-X
- Kersten, D., Mamassian, P., and Knill, D. C. (1997). Moving cast shadows induce apparent motion in depth. *Perception* 26, 171–192. doi: 10.1068/p260171
- Kuylen, C., Balas, B., and Thomas, L. E. (2014). My shadow, myself: cast-body shadows are embodied. *Psychon. Bull. Rev.* 21, 676–681. doi: 10.3758/s13423-013-0545-6
- Lenggenhager, B., Tadi, T., Metzinger, M., and Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science* 317, 1096–1099. doi: 10.1126/science.1143439

## Acknowledgment

This research was financially supported by the University of Padua.

- Liang, C., Chang, S.-Y., Chen, W.-Y., Huang, H.-C., and Lee, Y.-T. (2015). Body ownership and experiential ownership in the self-touching illusion. *Front. Psychol.* 5:1591. doi: 10.3389/fpsyg.2014.01591
- Liden, W. H., and Herberholz, J. (2008). Behavioral and neural responses of juvenile crayfish to moving shadows. *J. Exp. Biol.* 211, 1355–1361. doi: 10.1242/jeb.010165
- Longo, M. R., Betti, V., Aglioti, S. M., and Haggard, P. (2009). Visually induced analgesia: seeing the body reduces pain. *J. Neurosci.* 29, 12125–12130. doi: 10.1523/JNEUROSCI.3072-09.2009
- Longo, M. R., Cardozo, S., and Haggard, P. (2008). Visual enhancement of touch and the bodily self. *Conscious Cogn.* 17, 1181–1191. doi: 10.1016/j.concog.2008.01.001
- Longo, M. R., Iannetti, G. D., Mancini, F., Driver, J., and Haggard, P. (2012). Linking pain and the body: neural correlates of visually induced analgesia. *J. Neurosci.* 32, 2601–2607. doi: 10.1523/JNEUROSCI.4031-11.2012
- Longo, M. R., Pernigo, S., and Haggard, P. (2011). Vision of the body modulates processing in primary somatosensory cortex. *Neurosci. Lett.* 489, 159–163. doi: 10.1016/j.neulet.2010.12.007
- Lovell, P. G., Gilchrist, I. D., Tolhurst, D. J., and Troscianko, T. (2009). Search for gross illumination discrepancies in images of natural objects. *J. Vis.* 9, 1–14. doi: 10.1167/9.1.37
- Macaluso, E., Frith, C., and Driver, J. (2000). Selective spatial attention in vision and touch: unimodal and multimodal mechanisms revealed by PET. *J. Neurophysiol.* 83, 3062–3075.
- Mamassian, P. (2004). Impossible shadows and the shadow correspondence problem. *Perception* 33, 1279–1290. doi: 10.1068/p5280
- Mamassian, P., Knill, D. C., and Kersten, D. (1998). The perception of cast shadows. *Trends Cogn. Sci.* 2, 288–295. doi: 10.1016/S1364-6613(98)01204-2
- Maravita, A., Spence, C., Kennett, S., and Driver, J. (2002a). Tool-use changes multimodal spatial interactions between vision and touch in normal humans. *Cognition* 83, B25–B34. doi: 10.1016/S0010-0277(02)00003-3
- Maravita, A., Spence, C., Sergeant, C., and Driver, J. (2002b). Seeing your own touched hands in a mirror modulates cross-modal interactions. *Psychol. Sci.* 13, 350–355. doi: 10.1111/j.0956-7976.2002.00463.x
- Mascalzoni, E., Regolin, L., and Vallortigara, G. (2009). Mom's shadow: structure-from-motion in newly-hatched chicks as revealed by an imprinting procedure. *Anim. Cogn.* 12, 389–400. doi: 10.1007/s10071-008-0198-4
- Müller, H. J., and Rabbitt, P. M. A. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to suppression. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 315–330. doi: 10.1037/0096-1523.15.2.315
- Norman, J. F., Dawson, T. E., and Raines, S. R. (2000). The perception and recognition of natural object shape from deforming and static shadows. *Perception* 29, 135–148. doi: 10.1068/p2994
- Norman, J. F., Lee, Y., Phillips, F., Norman, H. F., Jenning, L. R., and McBride, T. R. (2009). The perception of 3-D shape from shadows cast onto curved surfaces. *Acta Psychol.* 131, 1–11. doi: 10.1016/j.actpsy.2009.01.007
- Pavan, G., Dalmaso, M., Galfano, G., and Castelli, L. (2011). Racial group membership is associated to gaze-mediated orienting in Italy. *PLoS ONE* 6:e25608. doi: 10.1371/journal.pone.0025608
- Pavani, F., and Castiello, U. (2004). Binding personal and extrapersonal space through body shadows. *Nat. Neurosci.* 7, 13–14. doi: 10.1038/nn1167
- Pavani, F., and Galfano, G. (2007). Self-attributed body-shadows modulate tactile attention. *Cognition* 104, 73–88. doi: 10.1016/j.cognition.2006.05.007
- Pavani, F., Rigo, P., and Galfano, G. (2014). From body shadows to bodily attention: automatic orienting of tactile attention driven by cast shadows. *Conscious Cogn.* 29, 56–67. doi: 10.1016/j.concog.2014.07.006
- Pavani, F., Spence, C., and Driver, J. (2000). Visual capture of touch: out-of-the-body experiences with rubber gloves. *Psychol. Sci.* 11, 353–359. doi: 10.1111/1467-9280.00270
- Pourtois, G., Peelen, M. V., Spinelli, L., Seeck, M., and Vuilleumier, P. (2007). Direct intracranial recording of body-selective responses in human extrastriate visual cortex. *Neuropsychologia* 45, 2621–2625. doi: 10.1016/j.neuropsychologia.2007.04.005
- Press, C., Taylor-Clarke, M., Kennett, S., and Haggard, P. (2004). Visual enhancement of touch in spatial body representation. *Exp. Brain Res.* 154, 238–245. doi: 10.1007/s00221-003-1651-x
- Reeder, R. R., and Peelen, M. V. (2013). The contents of the search template for category-level in natural scenes. *J. Vis.* 13, 1–13. doi: 10.1167/13.3.13
- Rensink, R. A., and Cavanagh, P. (2004). The influence of cast shadows on visual search. *Perception* 33, 1339–1358. doi: 10.1068/p5322
- Ristic, J., and Kingstone, A. (2005). Taking control of reflexive social attention. *Cognition* 94, B55–B65. doi: 10.1016/j.cognition.2004.04.005
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230
- Ro, T., Friggel, A., and Lavie, N. (2007). Attentional biases for faces and body parts. *Vis. Cogn.* 15, 322–348. doi: 10.1080/13506280600590434
- Romano, D., Pfeiffer, C., Maravita, A., and Blanke, O. (2014). Illusory self-identification with an avatar reduces arousal responses to painful stimuli. *Behav. Brain Res.* 261, 275–281. doi: 10.1016/j.bbr.2013.12.049
- Rorden, C., Heutink, J., Greenfield, E., and Robertson, I. H. (1999). When a rubber hand 'feels' what the real hand cannot. *Neuroreport* 10, 135–138. doi: 10.1097/00001756-199901180-00025
- Sambo, C. F., Gillmeister, H., and Forster, B. (2009). Viewing the body modulates neural mechanisms underlying sustained spatial attention in touch. *Eur. J. Neurosci.* 30, 143–150. doi: 10.1111/j.1460-9568.2009.06791.x
- Sartori, L., and Castiello, U. (2013). Shadows in the mirror. *Neuroreport* 24, 63–67. doi: 10.1097/WNR.0b013e32835c6e6a
- Serino, A., Farnè, A., Rinaldesi, M. L., Haggard, P., and Làdavas, E. (2007). Can vision of the body ameliorate impaired somatosensory function? *Neuropsychologia* 45, 1101–1107. doi: 10.1016/j.neuropsychologia.2006.09.013
- Slater, M., Spanlang, B., Sanchez-Vives, M. V., and Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLoS ONE* 5:e10564. doi: 10.1371/journal.pone.0010564
- Spence, C., Pavani, F., and Driver, J. (2004a). Spatial constraints on visual-tactile cross-modal distractor congruency effects. *Cogn. Affect. Behav. Neurosci.* 4, 148–169. doi: 10.3758/CABN.4.2.148
- Spence, C., Pavani, F., Maravita, A., and Holmes, N. (2004b). Multisensory contributions to the 3-D representation of visuotactile peripersonal space in humans: evidence from the crossmodal congruency task. *J. Physiol. Paris* 98, 171–189. doi: 10.1016/j.jphysparis.2004.03.008
- Spence, C., and Santangelo, V. (2009). Capturing spatial attention with multisensory cues: a review. *Hear. Res.* 258, 134–142. doi: 10.1016/j.heares.2009.04.015
- Sugiura, M., Watanabe, J., Maeda, Y., Matsue, Y., Fukuda, H., and Kawashima, R. (2005). Cortical mechanisms of visual self-recognition. *Neuroimage* 24, 143–149. doi: 10.1016/j.neuroimage.2004.07.063
- Takeuchi, M., Isogawa, M., Iwai, D., and Sato, K. (2014). “Weak perspective shadow interface for seated user's pointing on large wall display,” in *Proceedings of the 2014 IEEE/SICE International Symposium on System Integration (SII)*, (Tokyo: Chuo University), 316–321.
- Taylor-Clarke, M., Kennett, S., and Haggard, P. (2002). Vision modulates somatosensory cortical processing. *Curr. Biol.* 12, 233–236. doi: 10.1016/S0960-9822(01)00681-9
- Taylor-Clarke, M., Kennett, S., and Haggard, P. (2004). Persistence of visual-tactile enhancement in humans. *Neurosci. Lett.* 354, 22–25. doi: 10.1016/j.neulet.2003.09.068
- Tipper, S. P., Lloyd, D., Shorland, B., Dancer, C., Howard, L. A., and McGlone, F. (1998). Vision influences tactile perception without proprioceptive orienting. *Neuroreport* 9, 1741–1744. doi: 10.1097/00001756-19980610-00013
- Tipper, S. P., Phillips, N., Dancer, C., Lloyd, D., Howard, L. A., and McGlone, F. (2001). Vision influences tactile perception at body sites that cannot be viewed directly. *Exp. Brain Res.* 139, 160–167. doi: 10.1007/s002210100743
- Tsakiris, M., Haggard, P., Franck, N., Mainy, N., and Sirigu, A. (2005). A specific role for efferent information in self-recognition. *Cognition* 96, 215–231. doi: 10.1016/j.cognition.2004.08.002
- Tsakiris, M., Prabhu, G., and Haggard, P. (2006). Having a body versus moving your body: how agency structures body-ownership. *Conscious Cogn.* 15, 423–432. doi: 10.1016/j.concog.2005.09.004
- Tsakiris, M., Schütz-Bosbach, S., and Gallagher, S. (2007). On agency and body-ownership: phenomenological and neurocognitive reflections. *Conscious Cogn.* 16, 645–660. doi: 10.1016/j.concog.2007.05.012
- van den Bos, E., and Jeannerod, M. (2002). Sense of body and sense of action both contribute to self-recognition. *Cognition* 85, 177–187. doi: 10.1016/S0010-0277(02)00100-2
- Whiteley, L., Kennett, S., Taylor-Clarke, M., and Haggard, P. (2004). Facilitated processing of visual stimuli associated with the body. *Perception* 33, 307–314. doi: 10.1068/p5053



- Witt, J. K., Proffitt, D. R., and Epstein, W. (2005). Tool use affects perceived distance, but only when you intend to use it. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 880–888. doi: 10.1037/0096-1523.31.5.880
- Xu, H., Iwai, D., Hiura, S., and Sato, K. (2006). “User interface by virtual shadow projection,” in *Proceedings of the 2006 SICE-ICASE International Joint Conference SA14-1, A0897*, Busan.
- Yonas, A., and Granrud, C. E. (2006). Infants’ perception of depth from cast shadows. *Percept. Psychophys.* 68, 154–160. doi: 10.3758/BF03193665
- Zbrodoff, N. J., and Logan, G. D. (1986). On the autonomy of mental processes: a case study of arithmetic. *J. Exp. Psychol. Gen.* 115, 118–130. doi: 10.1037/0096-3445.115.2.118

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Pavani and Galfano. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Audiovisual integration of emotional signals from others' social interactions

Lukasz Piwek<sup>1\*</sup>, Frank Pollick<sup>2</sup> and Karin Petrini<sup>3</sup>

<sup>1</sup> Behaviour Research Lab, Bristol Business School, University of the West of England, Bristol, UK, <sup>2</sup> School of Psychology, College of Science and Engineering, University of Glasgow, Glasgow, UK, <sup>3</sup> Department of Psychology, Faculty of Humanities & Social Sciences, University of Bath, Bath, UK

## OPEN ACCESS

### Edited by:

Achille Pasqualotto,  
Sabanci University, Turkey

### Reviewed by:

Beatrice De Gelder,  
Maastricht University, Belgium  
Katja Koelkebeck,  
University of Muenster, Germany

### \*Correspondence:

Lukasz Piwek,  
Behaviour Research Lab,  
Bristol Business School, University of  
the West of England, Coldharbour  
Lane, 4D16, Bristol, BS16 1QY, UK  
lpiwek@gmail.com

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 28 February 2015

**Accepted:** 23 April 2015

**Published:** 08 May 2015

### Citation:

Piwek L, Pollick F and Petrini K (2015)  
Audiovisual integration of emotional  
signals from others' social  
interactions. *Front. Psychol.* 6:611.  
doi: 10.3389/fpsyg.2015.00611

Audiovisual perception of emotions has been typically examined using displays of a solitary character (e.g., the face-voice and/or body-sound of one actor). However, in real life humans often face more complex multisensory social situations, involving more than one person. Here we ask if the audiovisual facilitation in emotion recognition previously found in simpler social situations extends to more complex and ecological situations. Stimuli consisting of the biological motion and voice of two interacting agents were used in two experiments. In Experiment 1, participants were presented with visual, auditory, auditory filtered/noisy, and audiovisual congruent and incongruent clips. We asked participants to judge whether the two agents were interacting happily or angrily. In Experiment 2, another group of participants repeated the same task, as in Experiment 1, while trying to ignore either the visual or the auditory information. The findings from both experiments indicate that when the reliability of the auditory cue was decreased participants weighted more the visual cue in their emotional judgments. This in turn translated in increased emotion recognition accuracy for the multisensory condition. Our findings thus point to a common mechanism of multisensory integration of emotional signals irrespective of social stimulus complexity.

**Keywords:** multisensory integration, social interactions, point-light displays, voice, happiness, anger

## 1. Introduction

Perception of emotions is a multimodal event; by integrating signals from facial expressions, body movements, vocal prosody and other cues, we make emotional judgments about others. This multisensory integration of emotional expressions has been studied with faces and voices (de Gelder and Vroomen, 2000; Kreifelts et al., 2007; Collignon et al., 2008), body expression and faces (Meeren et al., 2005; Van den Stock et al., 2007), body expression with sound stimuli (Vines et al., 2006; Petrini et al., 2010, 2011), and body expressions and voices (Pichon et al., 2008; Stienen et al., 2011). A number of studies investigating the perception of emotions from facial expression and voices suggested strong bidirectional links between emotion detection processes in vision and audition (Massaro and Egan, 1996; de Gelder and Vroomen, 2000; Collignon et al., 2008; Jessen et al., 2012). For instance, de Gelder and Vroomen (2000) presented participants with static photographs of emotional faces combined with short vocal verbalizations, and found that participants emotional judgments reflected multisensory integration. When asked to identify the expression of a face, while ignoring a simultaneously heard voice, their judgments were nevertheless influenced by

the tone of the voice, and vice versa. Similarly, Collignon et al. (2008) showed that participants were faster and more accurate to identify fearful and disgust expressions when they observed faces combined with voices than either faces or voices alone. This multisensory behavioral facilitation became particularly evident when the most reliable visual information was degraded, thus changing the participants weighting strategy (i.e., they weighted the auditory cue more when judging the expressed emotion). Only a small number of studies have examined how observers integrate signals from emotional body movement and voice, and results so far follow a similar pattern to studies of emotional faces and voices (Van den Stock et al., 2007).

These studies have examined perception of emotions involving faces, voices or body movement using single agent displays. However, a growing number of studies point to substantial differences between the social situations involving a single person compared to the situations involving two people interacting. Social interaction has been shown to change fundamental aspects of visual perception and recognition (Scherer, 2003; Shiffrar, 2011). For example, Neri et al. (2006) and Manera et al. (2011) demonstrated that observers can use information detected from one of the agents in the observed social interaction to predict the action or response from the other agent. Besides behavioral studies also neuroimaging studies (Centelles et al., 2011; Petrini et al., 2014) have examined which brain regions were recruited during the observation of two interacting agents. While the “mirror neuron” system and “mentalizing networks” are rarely concurrently active (Van Overwalle and Baetens, 2009), these studies found that both of these networks were needed to process the social intentions carried by the biological motion of the two humans interacting. This adds to the argument that observation and understanding of multiagent social interactions may involve a wider network of brain regions than that of a single agent social action.

We do not know however if these differences in behavioral and neural processing between multiagent and single agent social situations extend to multisensory recognition of emotions. Here we ask whether the multisensory facilitation in emotion recognition, reported by previous studies using single agent social displays (e.g., de Gelder and Vroomen, 2000; Collignon et al., 2008; Petrini et al., 2010, 2011), extends to multiagent social interactions. To this end we carried out two experiments, utilizing a paradigm frequently employed in studies of multisensory integration of emotional signals (e.g., de Gelder and Vroomen, 2000; Collignon et al., 2008; Petrini et al., 2010). In the both experiments we asked participants to recognize the emotion expressed (happiness or anger) in audiovisual, audio, and video clips of two agents interacting. In Experiment 1, we varied the reliability of the auditory information by using two different degrading methods (low-pass filtering and addition of brown noise), and the emotional congruency between visual and auditory cues. In Experiment 2, we also varied the level of relevance attributed to the two signals by asking participants to ignore one of the information while performing the task (e.g., to judge the visual emotion while ignoring the auditory emotion).

## 2. Materials and Methods

### 2.1. Motion and Voice Capture of Stimuli Set

Motion capture took place at the University of Glasgow in the School of Psychology using 12 Vicon MXF40 cameras (Vicon, 2010) that offer online monitoring of 3D motion signals. The audio capture was done simultaneously using a custom-upgraded Vicon Analogue Card (Vicon, 2010) connected to amplifier with AKG D7S Supercardioid Dynamic Microphone, recording at 44.1 kHz and 24-bit sampling rate. Twelve repetitions of happy and angry interactions were recorded between eight pairs of actors (mean age of 26.12 years, ranging from 17 to 43 years). Actors were asked to interact exchanging one of two simple, single-sentence dialogues in each capture trial (e.g., Actor 1: “Where have you been?,” Actor 2: “I’ve just met with John”). A single capture trial lasted between 3–5 s. During the capture trial actors were positioned, one facing the other, at a distance specified by a marked position on the floor, approximately 1.3 m. This interpersonal distance varied between 1 and 1.6 m and it flexibly changed during the capture trials, depending on how much actors moved when interacting. However, at the beginning of each single capture trial actors were asked to come back to the start position marked on the floor.

To help actors convey angry and happy interactions they were given short and simple scenarios of the emotional situations and asked to imagine themselves in those situations. Actors were also instructed to recall their own past situations associated with the relevant emotional scenario to help them induce the emotion. The hypothetical scenarios were based on simple common situations (Scherer, 1986). Actors were given relative freedom in expressing the emotions during interactions (Clarke et al., 2005). They were encouraged to act naturally, but they were instructed to avoid touching each other and we were careful to give them only verbal instructions rather than performing actions ourselves (Clarke et al., 2005; Ma et al., 2006; Roether et al., 2009).

MATLAB 2010 (Mathworks, 2010) was used to convert captured movement into format useful for animation—as point-light displays. Point-light display (see **Figure 1** for an example) is a method of representing movement separately from other cues like clothing or body shape and is one of the most common approaches in the study of human motion (Johansson, 1973). Point-light display contains little or no static spatial information and enables complex manipulation of different features such as temporal coordination (Bertenthal and Pinto, 1994) or position of points (Cutting, 1981; Verfaillie, 1993). We chose point-lights over full-body displays to avoid any emotional bias that could be associated with cues such as identity, clothing or body shape, and to make sure we are primarily looking at the effects of body movement with visual displays (Hill et al., 2003). Point-light displays also enable us to easily manipulate various parameters of displays (e.g., viewpoint, number of points), and therefore help us to “future proof” our stimuli set for other studies without the need to re-capture a new interactions.

To convert motion capture coordinates to point-light displays we used an approach similar to Dekeyser et al. (2002), Troje (2002), and Ma et al. (2006). Specifically, we computed the location of 15 virtual markers positioned at major joints of

the body. The algorithm converted those 15 virtual markers from each actor into point-light displays (Pollick et al., 2001), generated as white dots on a black background from the side view. The algorithm exported point-light displays in the Audio Video Interleave (AVI) format with the frame rate of 60 fps.

Adobe Audition 3 (Adobe Systems, 2007) was used to post-process the dialogues. Every audio dialogue was first amplified by 10 dB and then a noise reduction was applied. All audio dialogues were then normalized to create a consistent level of amplitude, and to obtain the average volume of around 65 dB. Finally, each audio dialogue was exported as a Waveform Audio File Format (WAV) with a resolution of 44.1 kHz and 24-bit sampling rate.

The final stimulus was created using Adobe Premiere 1.5 (Adobe Systems, 2004) and consisted of 192 unique audio-visual clips (each clip was between 2500 and 3500 ms long) including 8 actor couples, 2 emotions (happy and angry) and 12 repetitions. Each unique clip was created in three modality formats: as point-lights (visual display), dialogue (auditory display) and a combination of point-lights with dialogues (audio-visual display). An example of angry and happy audio-visual clips can be viewed in Supplementary Movie.

## 2.2. Stimuli Validation Study

To examine whether observers could identify emotions conveyed in point-light displays and voice dialogues from created stimuli set, we conducted a stimuli validation study. Participants were presented with the displays as point-lights (visual group with 7 male and 8 female participants), voice dialogues (auditory group with 6 male and 7 female participants) or a combination of point-lights and dialogues (audio-visual group with 8 male and 7 female participants). Each group was presented with 192 displays described above. The reason for using a between-subject design was to avoid audio-visual facilitation, or carry-over effects, that could impact emotional identification when visual, auditory, and audio-visual displays are presented together in one set (Vines et al., 2006; Collignon et al., 2008). We also wanted to restrict presentation of every display to a single occasion to avoid effects of practice that can occur when participants see a repetition of a specific display (Heiman, 2002). The task was exactly the same for each group: after being presented with the display, participants were asked to identify whether interaction was happy or angry. Each display was presented only once and the order of all displays was randomized. The results provided us with average accuracy scores for each display we created in the stimuli set. Based on those results we selected a subset of eight angry and eight happy displays that were identified with 75% accuracy. However, by averaging across displays, we found that identification accuracy was higher in audio-visual (82%) and auditory-only (78%) groups than visual group (62%), indicating that the auditory information was more reliable than the visual. Hence, we decreased the reliability of the auditory stimuli to a level similar to the visual stimuli, as a greater increase in multisensory precision is obtained in situations for which the two sensory cues have a similar level of reliability (e.g., Ernst and Banks, 2002; Alais and Burr, 2004). To this end we used two methods frequently utilized in the literature: addition of brown noise to dialogues (Barnes and Allan, 1966; You et al., 2006; Hammerschmidt and Jürgens, 2007; Gardiner,

2009) and application of low-pass filter (Rogers et al., 1971; Frick, 1985; Scherer, 2003; Knoll et al., 2009). The use of both low-pass filtering and brown noise was guided by the principles of ecological validity—to choose a method of audio distortion that emulates real-life conditions. In this context, low-pass filtering made the voice dialogues sound like neighbors arguing behind a thick wall, or like the sounds heard when submerged in water; the words are unintelligible but the emotion behind the words is detectable. Accordingly, brown noise emulated real-life conditions such as listening to other people's conversation during heavy rainfall. Examples of those filtering methods applied to happy and angry audio can be heard in Supplementary Movie.

## 3. Experiment 1

Experiment 1 examined whether participants were more accurate in recognizing the expressed emotions when presented with both visual and auditory signals than only visual or auditory. We used a similar procedure to the one applied by Collignon et al. (2008) and Petrini et al. (2010). Participants were asked to recognize angry and happy expressions either displayed aurally, visually or audio-visually, in a congruent (the same expressions in the two modalities) or incongruent way (different expressions in the two modalities).

### 3.1. Participants

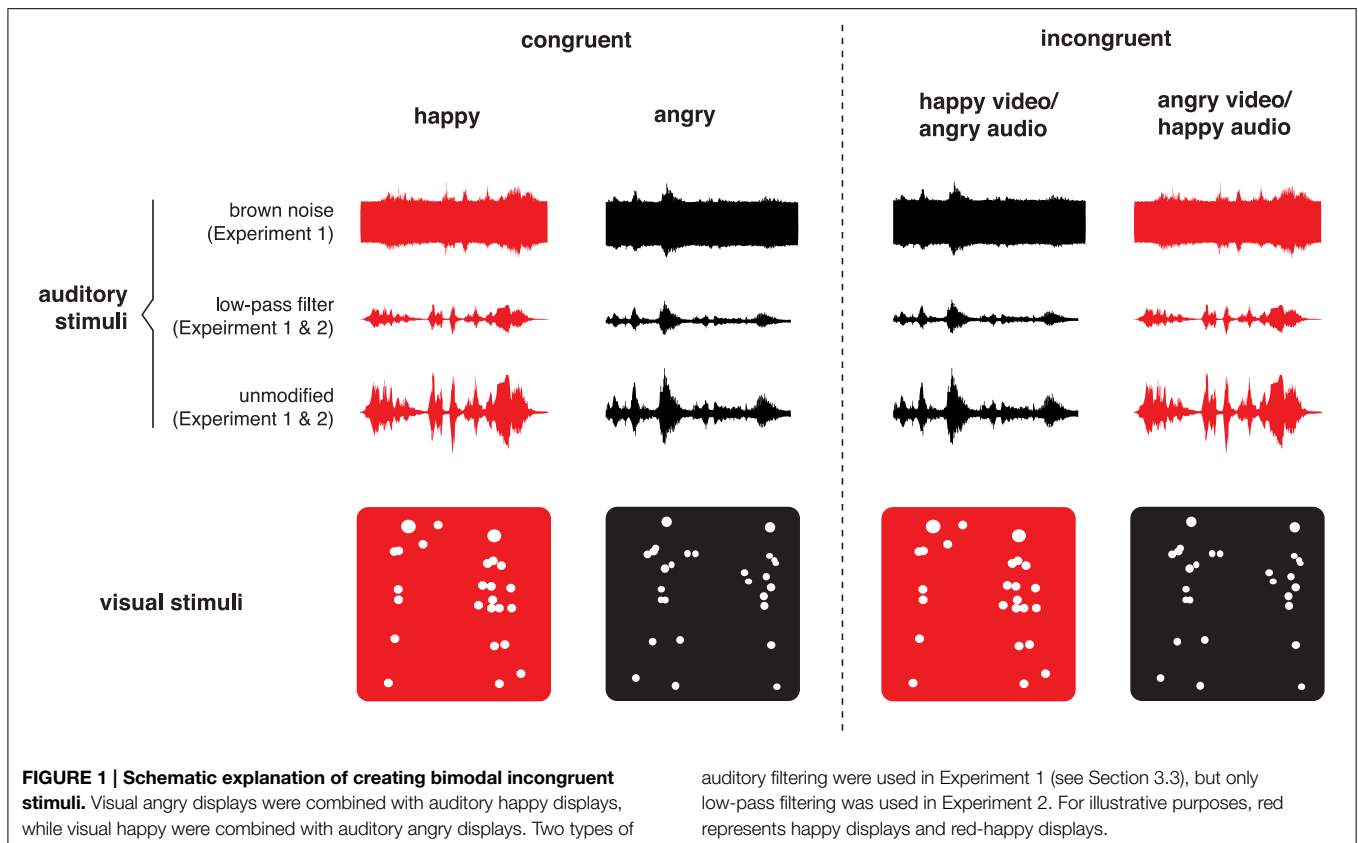
A total of 31 participants were recruited for Experiment 1: 15 female and 16 male, with a mean age of 22 years, ranging from 17 to 34 years. All participants were English speakers and UK born. All reported normal hearing and normal or corrected-to-normal vision. All participants were naive to the purpose of the study and had no prior experience with point-light display movies or images. The study received ethical approval from the University of Glasgow's Faculty of Information and Mathematical Sciences Ethics Review Board and every participant signed a consent form.

### 3.2. Stimuli

The auditory stimuli consisted of unmodified voice dialogues, low-pass filtered (LPF) dialogues, and dialogues with brown noise applied to them. All dialogues were processed using Adobe Audition 3 (Adobe Systems, 2008). To create LPF versions of the dialogues, a filter with a 400 Hz cut-off was applied to the unmodified dialogues attenuating signals with frequencies higher than the cut-off frequency. It is sometimes called a high-cut filter, or treble cut filter in audio applications (MacCallum et al., 2011). To create noisy dialogues, brown noise was added to the unmodified clip. All clips were normalized to the same amplitude level of around 65 dB.

The visual stimuli were a side view, unmodified dyadic point-light displays, an example of which can be seen on **Figure 1**. The bimodal stimuli were obtained by combining corresponding point-light displays with voice dialogues. The matching could either be “congruent,” with the use of point-light displays and voice dialogues expressing the same emotion (e.g., angry point-lights/angry voices), or “incongruent,” with point-light displays and voice dialogues expressing different emotions (e.g., happy point-lights/angry voices). We created two incongruent





versions of bimodal stimuli: point-light displays combined with unmodified voice dialogues, and point-light displays combined with dialogues filtered with brown noise or LPF. A schematic explanation of how bimodal incongruent stimuli were created is shown on **Figure 1**.

To summarize, the final stimuli set used in Experiment 1 consist of 112 stimuli with: 2 emotions (happy, angry), 7 stimulus types (visual, auditory unmodified, auditory filtered, bimodal congruent with unmodified dialogue, bimodal congruent with filtered dialogue, bimodal incongruent with unmodified dialogue, bimodal incongruent with filtered dialogue), and 8 actor pairs.

### 3.3. Design and Procedure

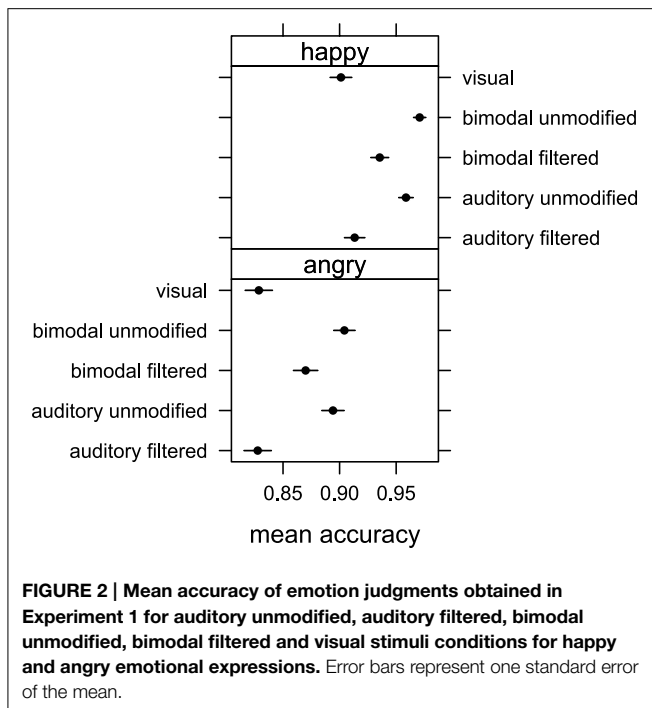
Participants were tested in a dark room, with only a small lamp to illuminate the keyboard. They were seated approximately 65 cm from a 21" Cathode Ray Tube (CRT) monitor with resolution of 1024 by 768 pixels, and 60 Hz refresh rate. Point-light displays subtended a maximum visual angle of approximately  $8.5^\circ$  in height and  $6^\circ$  in width. Voice dialogues were presented simultaneously with a white fixation cross shown during each display. Participants wore headphones (Beyer Dynamic DT Headphones), with an intensity at the sound source of 60 dB. We used Neurobehavioral Presentation 13.1 software (Neurobehavioral Systems, 2008) to present the displays and collect the responses. After each display, participants were asked to identify whether the presented interaction was happy or angry. They did so by pressing "H" for happy, or "A" for angry on the

keyboard. Each display lasted between 2500–3500 ms and the next display was presented immediately after participants pressed the response key. Overall, participants were presented with a total of 336 displays that included three repetitions of all conditions randomly interleaved in 3 separate blocks of 112 stimuli.

It's important to note that in Experiment 1 auditory filtered stimuli were presented either with addition of brown noise (15 participants) or filtered with LPF (16 participants). We wanted to compare whether either of these two filtering methods was particularly better in filtering and decreasing reliability of auditory signal. We conducted two-sample *t*-tests on the averaged accuracy scores to establish whether there was a difference in correct discriminations when participants were presented with the auditory condition filtered with a low-pass filter rather than brown noise. Results showed that there was no significant difference in participants' performance between the two filtering methods ( $t = -0.42$ ,  $df = 29$ ,  $p = 0.68$ ). Therefore, Experiment 1 included responses collated across two filtering methods as we found no differences between them.

### 3.4. Results

The averaged proportion of correct responses were submitted to a repeated measure ANOVA with "emotion" (happy and angry) and "stimuli" (visual, auditory unmodified, auditory filtered, bimodal congruent unmodified, and bimodal congruent filtered) as within factors. The ANOVA returned a main effect of "emotion" [ $F_{(1, 29)} = 13.81$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.15$ ]. **Figure 2** clearly shows that participants were overall more accurate when

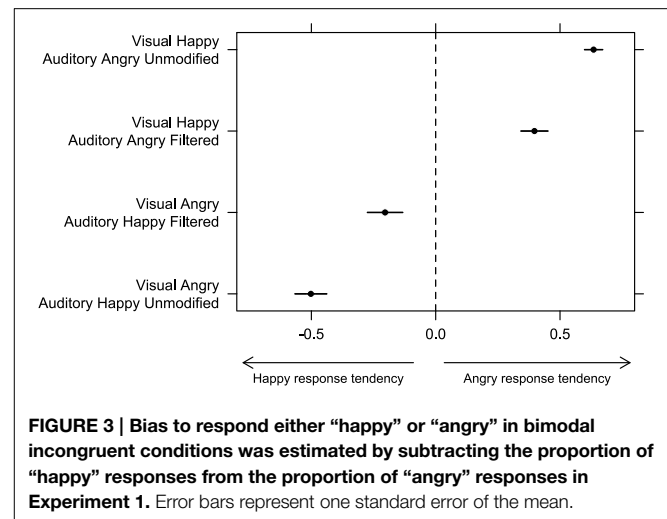


**FIGURE 2 | Mean accuracy of emotion judgments obtained in Experiment 1 for auditory unmodified, auditory filtered, bimodal unmodified, bimodal filtered and visual stimuli conditions for happy and angry emotional expressions.** Error bars represent one standard error of the mean.

judging happy rather than angry displays though the average recognition accuracy for the emotion expressed in the clips was far above the level of chance (50%). We also found a main effect of the factor ‘stimuli’ [ $F_{(4, 116)} = 20.46, p < 0.001, \eta_G^2 = 0.11$ ] indicating that some stimuli conditions were judged more correctly than others. No interaction between factors ‘emotion’ and ‘stimuli’ [ $F_{(4, 116)} = 0.24, p = 0.91, \eta_G^2 = 0$ ] was found, indicating that differences observed between various stimuli conditions were not influenced by emotional valence.

Pairwise comparison with correction for multiple testing showed that the emotion expressed in the visual displays was recognized less accurately than that expressed in the auditory unmodified ( $p < 0.001$ ), bimodal unmodified ( $p < 0.001$ ) and bimodal filtered ( $p < 0.001$ ) displays. No difference in accuracy was found between visual and auditory filtered conditions ( $p = 0.56$ ), and bimodal unmodified and auditory unmodified ( $p = 0.48$ ). Finally, participants were more accurate in recognizing the correct emotion in the bimodal filtered condition than in either the auditory filtered condition ( $p < 0.001$ ), or the visual condition ( $p < 0.001$ ).

To analyze responses for incongruent bimodal stimuli we had to use a different approach, as there were no ‘correct’ responses for this stimulus. We used the same approach of Collignon et al. (2008) and Petrini et al. (2010). We calculated a tendency to respond either ‘angry’ or ‘happy’ by subtracting the proportion of ‘happy’ judgments from the proportion of ‘angry’ judgments in the four incongruent stimulus conditions (happy point-light display/angry unmodified voice; happy point-light display/angry filtered voice; angry point-light display/happy unmodified voice; and angry point-light display/happy filtered voice). The index, which varied between -1 (subject always responded ‘happy’) and 1 (subject always responded ‘angry’) was then submitted to



**FIGURE 3 | Bias to respond either ‘happy’ or ‘angry’ in bimodal incongruent conditions was estimated by subtracting the proportion of ‘happy’ responses from the proportion of ‘angry’ responses in Experiment 1.** Error bars represent one standard error of the mean.

ANOVA with ‘auditory emotion’ (happy or angry) and ‘auditory filtering’ (filtered or unmodified) as within-subject factors.

There was no significant effect of factor ‘auditory filtering’ [ $F_{(1, 30)} = 1.49, p = 0.23, \eta_G^2 = 0$ ], but we found a significant effect of factor ‘auditory emotion’ [ $F_{(1, 30)} = 163.10, p < 0.001, \eta_G^2 = 0.65$ ] as well as a significant interaction between factors ‘auditory emotion’ and ‘auditory filtering’ [ $F_{(1, 30)} = 86.07, p < 0.001, \eta_G^2 = 0.15$ ]. Pairwise comparison with correction for multiple testing revealed that the index was significantly more positive with ‘visual happy/auditory angry unmodified’ stimuli than with ‘visual happy/auditory angry filtered’ ( $p < 0.01$ ), and that the index was significantly more negative with ‘visual angry/auditory happy unmodified’ stimuli than with ‘visual angry/auditory happy filtered’ stimuli ( $p < 0.001$ ).

Figure 3 shows that for all bimodal incongruent combinations, participants’ response were biased toward the auditory modality, but this tendency was weaker when filtering was present in the auditory signal. These results are consistent with the previous findings in showing a clear auditory dominance when no filtering or noise was applied, and a clear change in weighting strategy toward the visual information when the auditory reliability was decreased.

## 4. Experiment 2

In Experiment 2, we requested the participants to pay attention to only one modality at a time to ascertain whether any multimodal effects found in Experiments 1 were due to automatic processes and would not disappear when participants were asked to ignore one of the two modalities. The underlying idea was that if audiovisual integration operates in an automatic fashion, multisensory influence should occur even if the participants only focus their attention toward one single modality (de Gelder and Vroomen, 2000; Vroomen and de Gelder, 2000).

### 4.1. Participants

Sixteen participants were recruited for Experiment 2: 6 female and 10 male, with a mean age of 22.7 years, ranging from 18 to

36 years. All participants were English speakers and UK born. All reported normal hearing and normal or corrected-to-normal vision. All participants were naive to the purpose of the study and had no prior experience with point-light display movies or images. The study received ethical approval from the University of Glasgow's Faculty of Information and Mathematical Sciences Ethics Review Board and every participant signed a consent form.

## 4.2. Stimuli

The stimulus set used in Experiment 2 was exactly the same as in Experiments 1 (Section 3.2). As we didn't find a difference between two methods of auditory filtering in Experiment 1, we only used low-pass filter for audio filtering in Experiment 2 (see end of Section 3.3 for details).

## 4.3. Design and Procedure

In Experiment 2 participants also performed an emotion recognition task but were explicitly asked to focus their attention on one sensory modality at a time, ignoring the other modality. As a result we introduced two separate blocks in Experiment 2: a visual and an auditory block. The visual block included 2 emotions (happy, angry), 5 stimulus types (visual, bimodal congruent with unmodified audio, bimodal congruent with filtered audio, bimodal incongruent with unmodified audio, bimodal incongruent with filtered audio), and 8 actor pairs. The auditory block included the same conditions of the visual blocks with only one difference; the auditory unimodal condition replaced the visual unimodal condition. Participants were presented with a total of 480 stimuli. Each block (i.e., auditory and visual) consisted of 240 stimuli, which included three repetitions of 80 stimulus conditions randomly interleaved within three separate blocks. Before starting to the visual block, participants were instructed to focus their attention on the visual displays and ignore the audio. In contrast, before starting to the auditory block, participants were instructed to focus their attention on the audio and ignore the visual displays. The order of visual and auditory blocks was counterbalanced across participants.

## 4.4. Results

The averaged proportion of correct responses were submitted to a repeated measure ANOVA with "emotion" (happy and angry), "attention" (attend vi-sual, attend auditory), and "stimuli" (unimodal, bimodal unmodified, and bimodal filtered) as within factors. We found a main effect of "emotion" [ $F_{(1, 15)} = 5.27, p < 0.05, \eta_G^2 = 0.10$ ] and **Figure 4** shows that participants were again more accurate when judging happy rather than angry displays. We also found a main effect of "stimuli" [ $F_{(2, 30)} = 6.35, p < 0.05, \eta_G^2 = 0.02$ ] indicating that some stimulus conditions were judged with more accuracy than others. No interaction between 'emotion' and "attention" [ $F_{(1, 15)} = 0.16, p = 0.7, \eta_G^2 = 0$ ]; "emotion" and "stimuli" [ $F_{(2, 30)} = 0.47, p = 0.63, \eta_G^2 = 0$ ]; "attention" and "stimuli" [ $F_{(2, 30)} = 2.12, p = 0.14, \eta_G^2 = 0.01$ ]; and "emotion," "attention," and "stimuli" [ $F_{(2, 30)} = 1.57, p = 0.23, \eta_G^2 = 0.01$ ] was found. Pairwise comparison with correction for multiple testing showed that bimodal unmodified

condition was judged more accurately than unimodal ( $p < 0.05$ ) and bimodal filtered ( $p < 0.05$ ) conditions. There was no difference between unimodal and bimodal filtered ( $p = 0.95$ ). We found no significant effect of factor "attention" [ $F_{(1, 15)} = 0.11, p = 0.74, \eta_G^2 = 0$ ] indicating that the level of accuracy for emotion recognition did not depend on the specific modality attended.

In Experiment 2, we again looked at the tendency to choose happy or angry emotion when observers were presented with incongruent displays. The index calculated for incongruent displays, which varied between  $-1$  (subject always responded "happy") and  $1$  (subject always responded "angry"), was analyzed by means of a Three-Way ANOVA with "auditory emotion" (happy or angry), "auditory filtering" (unmodified or filtered), and "attention" (visual or auditory) as within-subject factors. No significant effect of "attention" [ $F_{(1, 15)} = 1.93, p = 0.19, \eta_G^2 = 0.01$ ] was found in line with the previous findings. Overall, **Figure 5** shows that participants were biased toward the modality they attended—regardless of whether they attended auditory or visual signal. We found a significant effect of "auditory emotion" [ $F_{(1, 15)} = 7.11, p < 0.05, \eta_G^2 = 0.06$ ] as well as a significant interaction between "auditory emotion" and "auditory filtering" [ $F_{(1, 15)} = 22.54, p < 0.001, \eta_G^2 = 0.07$ ]. **Figure 5** shows that the presence of auditory filtering weakened participants tendency to use auditory signal in their responses, but this effect was stronger with happy than angry audio ( $p < 0.05$ ).

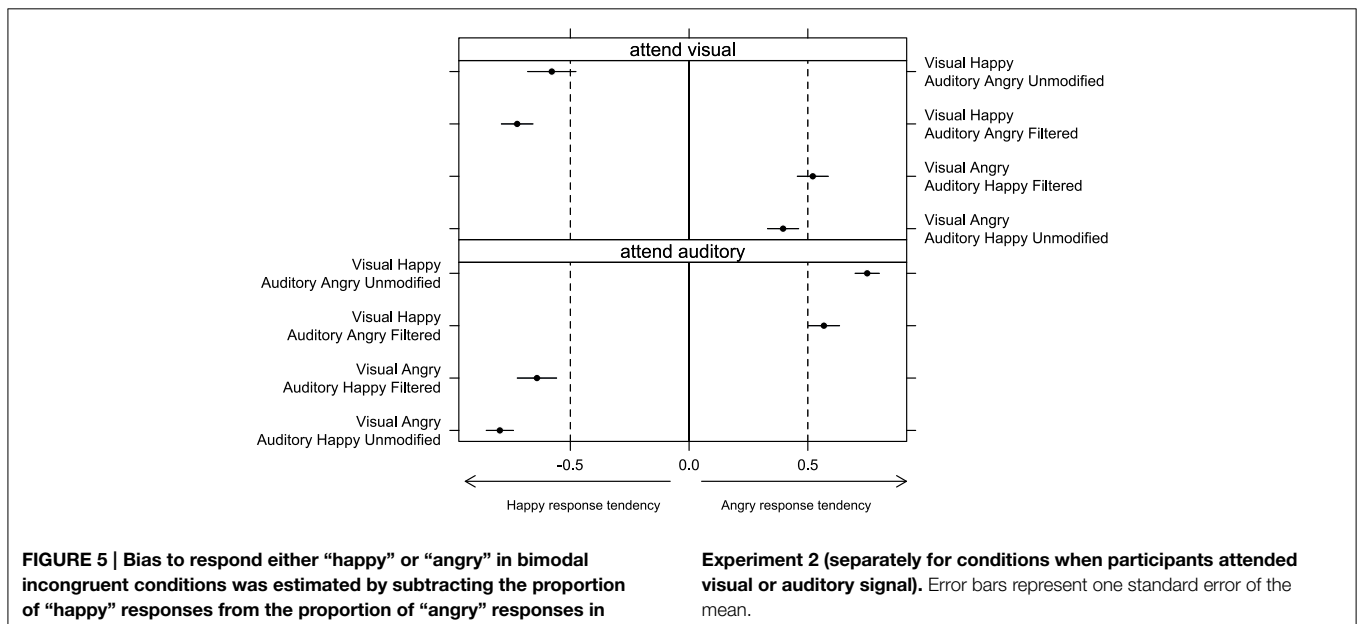
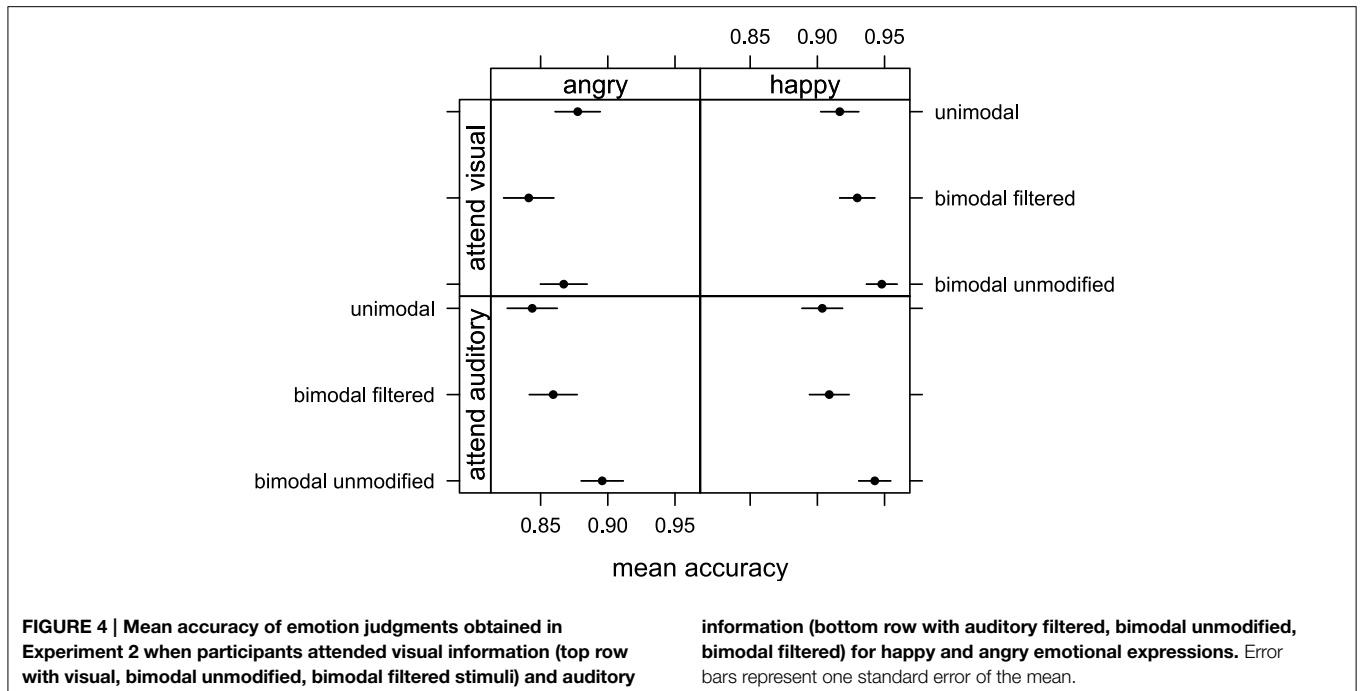
We also observed a significant interaction between "auditory emotion" and "attention" [ $F_{(1, 15)} = 245.45, p < 0.001, \eta_G^2 = 0.84$ ]. **Figure 5** shows that participants were biased toward the auditory information with the same extent for both happy and angry audio when they attended auditory rather than visual information. However, response tendency shifted more toward zero with happy audio than angry audio when visual signal was attended, which was not the case when auditory signal was attended.

Displays with happy auditory signal were shifted more toward the zero than with angry auditory signal when the visual was attended, but this is not the case when auditory was attended.

No other significant interaction was found.

## 5. Discussion

In the present study we ask whether the multisensory facilitation in emotion recognition, reported by previous studies using single agent social displays (e.g., de Gelder and Vroomen, 2000; Kreifelts et al., 2007; Collignon et al., 2008; Petrini et al., 2010), extends to multiagent social interactions. The results of both experiments consistently indicate that the auditory signal dominated the visual signal in the perception of emotions from social interactions. Participants were less accurate in discriminating emotions when making judgments on visual stimuli than on auditory stimuli. This result is in line with previous findings demonstrating that the auditory emotional information dominates the visual information in multisensory integration of emotional signals from body movements and sound (e.g., Vines et al., 2006; Petrini et al., 2010). However,



degrading the auditory information so to match its level of reliability to that of the visual information changed the participants weighting of the two cues. The level of accuracy with which participants could recognize the emotion portrayed in the audio clips (when the auditory reliability was lower) was no better than that for the video clips. Integrating the two cues when the auditory was less reliable resulted in multisensory facilitation (i.e., participants were more accurate in recognizing the correct emotion when using both cues) as described by single agent studies (e.g., Collignon et al.,

2008). Similarly, in both experiments we found that when participants judged the emotion in incongruent displays (e.g., happy visual information and angry auditory information), they shifted their responses toward the emotion represented by the visual signal if the auditory signal was less reliable. This supports earlier results by de Gelder and Vroomen (2000) and Collignon et al. (2008) that an incongruent combination of two signals would cause some disruption in the emotion interpretation of those signals, and a shift toward perceiving the emotion expressed by the most reliable information. The



similarity between our findings and those using a single agent provides evidence for a common mechanism of multisensory integration of emotional signals irrespective of social stimulus complexity.

Our results also show an interesting difference in the way we interpret emotional signals from body movement and voice as compared to face and voice. Specifically, studies on the perception of emotions from face and voice show that observers make their judgments based mainly on faces rather than voices, although such dominance can shift depending on the visual and auditory reliability of the stimuli (Massaro and Egan, 1996; de Gelder and Vroomen, 2000; Collignon et al., 2008; Jessen et al., 2012). In contrast, our results suggest that auditory stimuli (voice) rather than visual stimuli (body movement) plays a particularly important role in the perception of emotional social interactions. Vines et al. (2006) and Petrini et al. (2010) show a similar pattern of results but with the musical sound dominating body expression when observers judged musical performance from those two cues. Petrini et al. (2010) highlight that making of music requires specific coupling between the performer and instrument, but the complexity of information in music sound is difficult to achieve with body expression. In short, body expression plays a “secondary” role as an accenting factor in the observation of musical performance. However, music is a special case since not only the majority of movements are constrained by the instrument, but those movements are also produced by a tool (the instrument) rather than coming from the body action *per se* (Petrini et al., 2010). Another possible explanation for the strong effect of voice found in our study is that we used point-light displays rather than full body displays. Reduced cue point-light expressions could render visual signal less “informationally rich” when comparing to unmodified voice. Such argument is particularly valid when looking at the studies that used combination of static full body displays and voices (Stienen et al., 2011; Van den Stock et al., 2011). Specifically, those studies indicate that recognition performance for bodies and voices is on the similar level (i.e., visual signal is as reliable as auditory signal, as long as they are both congruent and unmodified).

Finally, it is possible that the source of the sound from the dyadic point-light displays in our study is uncertain due to lack of conversational cues such as the mouth or face movements. A potential solution to those issues would be to increase the reliability of the visual signal by introducing full body displays (but with a blurred faces like in studies by Van den Stock et al. 2011 or Stienen et al. 2011), or to introduce conditions with only a single actor at the time so to specify the source of sound production.

In a separate argument, a broad literature on deception and non-verbal communication show a strong interrelation between body movement and voice. Ekman et al. (1976) found that measures of hand movements and voice were interrelated but changed incongruently when a person shifted from honest to deceptive expressions. Specifically, the amount of symbolic hand movements decreased in deception, while pitch variance into high tones increased with deception, making the voice more accessible as cue as well as creating a discrepancy between voice

and body movement. Moreover, studies on body movement and speech rhythm in social conversation clearly show that speakers tend to use their body movement to highlight specific aspects of their spoken messages (Dittmann and Llewellyn, 1969). Movement output and speech output were found to be quite closely correlated (Boomer, 1963). Renneker (1963, p. 155) described what he called “speech-accompanying gestures,” which “seek to complement, modify, and dramatize the meanings of words,” Freedman and Hoffman (1967) separated what they called “punctuating movements” from other speech-related movements. It is possible that, in a conversational context, body movements play an accenting function to the voice—a claim also supported and suggested by Ekman (1965) regarding non-verbal behavior in general. This claim is further supported by brain imaging studies. For instance, Hubbard et al. (2009) found that non-primary auditory cortex showed greater activity when speech was accompanied by “beat” gesture than when speech was presented alone. Hubbard et al. (2009) results pointed toward a common neural substrate for processing speech and gesture, likely reflecting their joint communicative role in social interactions.

Considering our results on the emotional identification, we found that happy interactions were repeatedly identified more accurately than angry interactions in both experiments. The accuracy of recognition between angry and happy affect has long been a point of debate between researchers. A number of studies have shown that observers were better at identifying angry rather than happy emotional expressions when listening to voices (Scherer, 1986), viewing faces (Massaro and Egan, 1996; Fox et al., 2000; Knyazev et al., 2009), watching the actions of a single actor (Pollick et al., 2001) or watching interactions between two actors (Clarke et al., 2005). Several studies also argue that detection of anger serves as an evolutionary indicator of threat (Pichon et al., 2008), and specific brain areas such as the amygdala are tuned to detect angry actions from body movement (de Gelder, 2006). However, others found similar results to ours highlighting that happy expression is a highly salient social signal. For example, Dittrich et al. (1996) showed that happy displays of point-light dancers were identified more accurately compared to angry displays. Belin et al. (2008) created and experimentally validated a dataset of non-verbal affect bursts showing that vocal expressions of happiness were better recognized than anger. Johnstone et al. (2006) found that greater activation to happy vs. angry vocal expressions in amygdala and insula regions when explicitly attending to these expressions. In such context, our study adds further evidence that happy expressions from movement and voice are potentially more salient social signals when compared to anger.

In conclusion, we found that the auditory signal dominated the visual signal in the perception of emotions from social interactions, but only to the extent of auditory signals’ reliability. When reliability of auditory signal was degraded, participants weighted visual cues more in their judgments, which followed pattern of results similar to de Gelder and Vroomen (2000), Collignon et al. (2008), and Petrini et al. (2010). Similarly, when participants watched emotionally mismatched bimodal displays, filtering auditory signal increased

the weighting of visual cue. Our results suggest that when identifying emotions from complex social stimuli, we use similar mechanism of multimodal integration as with simple social stimuli.

## Author Contributions

Conception and design of the work: LP, KP, FP. Stimuli capture, processing and production: LP, FP. Acquisition of the data: LP. Analysis of the data: LP, KP. Wrote the paper: LP, KP, FP.

## References

- Alais, D., and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262. doi: 10.1016/j.cub.2004.01.029
- Barnes, J. A., and Allan, D. (1966). A statistical model of flicker noise. *Proc. IEEE* 54, 176–178.
- Belin, P., Fillion-Bilodeau, S., and Gosselin, F. (2008). The montreal affective voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behav. Res. Methods* 40, 531–539. doi: 10.3758/BRM.40.2.531
- Bertenthal, B. I., and Pinto, J. (1994). Global processing of biological motions. *Psychol. Sci.* 5, 221–224.
- Boomer, D. S. (1963). Speech disturbance and body movement in interviews. *J. Nerv. Ment. Dis.* 136, 263–266.
- Centelles, L., Assaiante, C., Nazarian, B., Anton, J.-L., and Schmitz, C. (2011). Recruitment of both the mirror and the mentalizing networks when observing social interactions depicted by point-lights: a neuroimaging study. *PLoS ONE* 6:e15749. doi: 10.1371/journal.pone.0015749
- Clarke, T. J., Bradshaw, M. F., Field, D. T., Hampson, S. E., and Rose, D. (2005). The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception* 34, 1171–1180. doi: 10.1068/p5203
- Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., et al. (2008). Audio-visual integration of emotion expression. *Brain Res.* 1242, 126–135. doi: 10.1016/j.brainres.2008.04.023
- Cutting, J. E. (1981). Coding theory adapted to gait perception. *J. Exp. Psychol* 7, 71–87. doi: 10.1037/0096-1523.7.1.71
- de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nat. Rev. Neurosci.* 7, 242–249. doi: 10.1038/nrn1872
- de Gelder, B., and Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cogn. Emot.* 14, 289–311. doi: 10.1080/026999300378824
- Dekeyser, M., Verfaillie, K., and Vanrie, J. (2002). Creating stimuli for the study of biological-motion perception. *Behav. Res. Methods* 34, 375–382. doi: 10.3758/BF03195465
- Dittmann, A. T., and Llewellyn, L. G. (1969). Body movement and speech rhythm in social conversation. *J. Pers. Soc. Psychol* 11, 98–106.
- Dittrich, W. H., Troscianko, T., Lea, S. E., and Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception* 25, 727–738.
- Ekman, P. (1965). “Communication through nonverbal behavior: a source of information about an interpersonal relationship,” in *Affect, Cognition and Personality*, eds S. Tomkins and C. Izard (New York, NY: Springer), 390–444.
- Ekman, P., Friesen, W. V., and Scherer, K. (1976). Body movement and voice pitch in deceptive interaction. *Semiotica* 16, 23–27.
- Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433. doi: 10.1038/415429a
- Fox, E., Lester, V., Russo, R., Bowles, R., Pichler, A., and Dutton, K. (2000). Facial expressions of emotion: are angry faces detected more efficiently? *Cogn. Emot.* 14, 61–92. doi: 10.1080/026999300378996
- Freedman, N., and Hoffman, S. P. (1967). Kinetic behavior in altered clinical states: approach to objective analysis of motor behavior during clinical interviews. *Perc. Mot. Skills* 24, 527–539.
- Frick, R. W. (1985). Communicating emotion: the role of prosodic features. *Psychol. Bull.* 97, 412–429.
- Gardiner, C. (2009). *Stochastic Methods: a Handbook for the Natural and Social Sciences, 4th Edn.* New York, NY: Springer.
- Hammerschmidt, K., and Jürgens, U. (2007). Acoustical correlates of affective prosody. *J. Voice* 21, 531–540. doi: 10.1016/j.jvoice.2006.03.002
- Heiman, G. W. (2002). *Research Methods in Psychology, 3rd Edn.* Boston; New York: Houghton Mifflin Company.
- Hill, H., Jinno, Y., and Johnston, A. (2003). Comparing solid-body with point-light animations. *Perception* 32, 561–566. doi: 10.1068/p3435
- Hubbard, A. L., Wilson, S. M., Callan, D. E., and Dapretto, M. (2009). Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. *Hum. Brain Mapp.* 30, 1028–1037. doi: 10.1002/hbm.20565
- Jessen, S., Obleser, J., and Kotz, S. A. (2012). How bodies and voices interact in early emotion perception. *PLoS ONE* 7:e36070. doi: 10.1371/journal.pone.0036070
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perc. Psychophys.* 14, 201–211.
- Johnstone, T., van Reekum, C. M., Oakes, T. R., and Davidson, R. J. (2006). The voice of emotion: an fMRI study of neural responses to angry and happy vocal expressions. *Soc. Cogn. Affect. Neurosci.* 1, 242–249. doi: 10.1093/scan/nsl027
- Knoll, M., Uther, M., and Costall, A. (2009). Effects of low-pass filtering on the judgment of vocal affect in speech directed to infants, adults and foreigners. *Speech Commun.* 51, 210–216. doi: 10.1016/j.specom.2008.08.001
- Knyazev, G. G., Bocharov, A. V., and Slobodskoj-Plusnin, J. Y. (2009). Hostility- and gender-related differences in oscillatory responses to emotional facial expressions. *Aggress. Behav.* 35, 502–513. doi: 10.1002/ab.20318
- Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., and Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage* 37, 1445–1456. doi: 10.1016/j.neuroimage.2007.06.020
- Ma, Y., Paterson, H. M., and Pollick, F. E. (2006). A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behav. Res. Methods* 38, 134–141. doi: 10.3758/BF03192758
- MacCallum, J. K., Olszewski, A. E., Zhang, Y., and Jiang, J. J. (2011). Effects of low-pass filtering on acoustic analysis of voice. *J. Voice* 25, 15–20. doi: 10.1016/j.jvoice.2009.08.004
- Manera, V., Becchio, C., Schouten, B., Bara, B. G., and Verfaillie, K. (2011). Communicative interactions improve visual detection of biological motion. *PLoS ONE* 6:e14594. doi: 10.1371/journal.pone.0014594
- Massaro, D. W., and Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychon. Bull. Rev.* 3, 215–221.
- Meeren, H. K. M., van Heijnsbergen, C. C. R. J., and de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proc. Natl. Acad. Sci. U.S.A.* 102, 16518–16523. doi: 10.1073/pnas.0507650102
- Neri, P., Luu, J. Y., and Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nat. Neurosci.* 9, 1186–1192. doi: 10.1038/nn1759
- Petrini, K., Crabbe, F., Sheridan, C., and Pollick, F. E. (2011). The music of your emotions: neural substrates involved in detection of emotional correspondence between auditory and visual music actions. *PLoS ONE* 6:e19165. doi: 10.1371/journal.pone.0019165
- Petrini, K., McAleer, P., and Pollick, F. E. (2010). Audiovisual integration of emotional signals from music improvisation does not depend on temporal correspondence. *Brain Res.* 1323, 139–148. doi: 10.1016/j.brainres.2010.02.012

## Acknowledgments

This study was supported by a grant from the ESRC (RES-060-25-0010).

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00611/abstract>

- Petrini, K., Piwek, L., Crabbe, F., Pollick, F. E., and Garrod, S. (2014). Look at those two!: the precuneus role in unattended third-person perspective of social interactions. *Hum. Brain Mapp.* 35, 5190–5203. doi: 10.1002/hbm.22543
- Pichon, S., de Gelder, B., and Grezes, J. (2008). Emotional modulation of visual and motor areas by dynamic body expressions of anger. *Soc. Neurosci.* 3, 199–212. doi: 10.1080/17470910701394368
- Pollick, F. E., Paterson, H. M., Bruderlin, A., and Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition* 82, B51–B61. doi: 10.1016/S0010-0277(01)00147-0
- Renneker, R. (1963). “Kinesic research and therapeutic processes: further discussion,” in *Expression of the Emotions in Man*, ed P. H. Knapp (New York, NY: International Universities Press).
- Roether, C. L., Omlor, L., Christensen, A., and Giese, M. A. (2009). Critical features for the perception of emotion from gait. *J. Vision* 9, 15.1–15.32. doi: 10.1167/9.6.15
- Rogers, P. L., Scherer, K. R., and Rosenthal, R. (1971). Content-filtering human speech: a simple electronic system. *Behav. Res. Methods* 3, 16–18.
- Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychol. Bull.* 99, 143–165.
- Scherer, K. R. (2003). Vocal communication of emotion: a review of research paradigms. *Speech Commun.* 40, 227–256. doi: 10.1016/S0167-6393(02)00084-5
- Shiffrar, M. (2011). People watching: visual, motor, and social processes in the perception of human movement. *Wiley Int. Rev. Cogn. Sci.* 2, 68–78. doi: 10.1002/wcs.88
- Stienen, B. M. C., Tanaka, A., and de Gelder, B. (2011). Emotional voice and emotional body postures influence each other independently of visual awareness. *PLoS ONE* 6:e25517. doi: 10.1371/journal.pone.0025517
- Troje, N. F. (2002). Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. *J. Vision* 2, 371–387. doi: 10.1167/2.5.2
- Van den Stock, J., de Jong, S. J., Hodiamont, P. P. G., and de Gelder, B. (2011). Perceiving emotions from bodily expressions and multisensory integration of emotion cues in schizophrenia. *Soc. Neurosci.* 6, 537–547. doi: 10.1080/17470919.2011.568790
- Van den Stock, J., Righart, R., and de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion* 7, 487–494. doi: 10.1037/1528-3542.7.3.487
- Van Overwalle, F., and Baetens, K. (2009). Understanding others’ actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage* 48, 564–584. doi: 10.1016/j.neuroimage.2009.06.009
- Verfaillie, K. (1993). Orientation-dependent priming effects in the perception of biological motion. *J. Exp. Psychol* 19:992. doi: 10.1037/0096-1523.19.5.992
- Vines, B. W., Krumhansl, C. L., Wanderley, M. M., and Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition* 101, 80–113. doi: 10.1016/j.cognition.2005.09.003
- Vroomen, J., and de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *J. Exp. Psychol* 26, 1583–1590. doi: 10.1037/0096-1523.26.5.1583
- You, M., Chen, C., Bu, J., Liu, J., and Tao, J. (2006). “Emotion recognition from noisy speech,” in *Proceedings of the 2006 IEEE International Conference on Multimedia and Expo* (Toronto, ON: IEEE), 1653–1656.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Piwek, Pollick and Petrini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Sources of Confusion in Infant Audiovisual Speech Perception Research

Kathleen E. Shaw<sup>1</sup> and Heather Bortfeld<sup>2,3\*</sup>

<sup>1</sup> Department of Psychology, University of Connecticut, Storrs, CT, USA, <sup>2</sup> Psychological Sciences, University of California, Merced, Merced, CA, USA, <sup>3</sup> Haskins Laboratories, New Haven, CT, USA

Speech is a multimodal stimulus, with information provided in both the auditory and visual modalities. The resulting audiovisual signal provides relatively stable, tightly correlated cues that support speech perception and processing in a range of contexts. Despite the clear relationship between spoken language and the moving mouth that produces it, there remains considerable disagreement over how sensitive early language learners—infants—are to whether and how sight and sound co-occur. Here we examine sources of this disagreement, with a focus on how comparisons of data obtained using different paradigms and different stimuli may serve to exacerbate misunderstanding.

**Keywords:** audiovisual perception, multimodal integration, infant perception, temporal binding window, sine wave speech, speech perception, speech disorders

## OPEN ACCESS

### Edited by:

Andriy Myachykov,  
Northumbria University, UK

### Reviewed by:

Jean Vroomen,  
University of Tilburg, Netherlands  
Clemens Wöllner,  
University of Hamburg, Germany

### \*Correspondence:

Heather Bortfeld  
hbortfeld@ucmerced.edu

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 15 February 2015

**Accepted:** 13 November 2015

**Published:** 15 December 2015

### Citation:

Shaw KE and Bortfeld H (2015)  
Sources of Confusion in Infant  
Audiovisual Speech Perception  
Research. *Front. Psychol.* 6:1844.  
doi: 10.3389/fpsyg.2015.01844

## INTRODUCTION

Although the development of early speech perception abilities is often framed as an auditory-only process, speech is a sensory-rich stimulus, with information provided across multiple modalities. Our focus here is on the auditory (i.e., spoken language) and visual (i.e., moving mouth) modalities, which together provide relatively stable, tightly correlated cues about the resulting speech. If we focus only on the articulators, both their visual form and the corresponding auditory stream they produce share onsets and offsets, intensity changes, amplitude contours, durational cues, and rhythmic patterning (Chandrasekaran et al., 2009). This reliable co-occurrence of cues serves to support speech comprehension (Sumby and Pollack, 1954), particularly in noisy environments (Massaro, 1984; Middelweerd and Plomp, 1987) and during language learning, whether first (Teinonen et al., 2008) or subsequent (Navarra and Soto-Faraco, 2007). Yet despite the clear relationship between spoken language and the moving mouth that produces it, there remains considerable disagreement about how sensitive early language learners—particularly infants—are to whether and how sight and sound co-occur. Here we examine the bases for this disagreement, with a particular focus on how data obtained using different methodologies and different stimuli may actually serve to exacerbate it.

One issue to consider is whether infants have initial biases toward attending to one or the other modality in the first place. On the one hand, infants have considerable prenatal experience with sound (DeCasper and Spence, 1986). Although the tissue and liquid barriers of the womb filter out frequencies greater than 5000 Hz, external acoustic stimuli are heard *in utero* beginning early in gestation (Jardri et al., 2008). Indeed, both behavioral data (Hepper and Shahidullah, 1994) and physiological data (Rubel and Ryals, 1983; Pujol et al., 1991) demonstrate that the fetal auditory system begins to process sounds between about 16 and 20 weeks. From that time forward, the cochlea matures anatomically during gestation such that its frequency response broadens



(Graven and Browne, 2008). Likewise, fetal abilities to discriminate among simultaneous frequencies, to separate rapid sequences of sounds (as in ordinary speech), and to perceive very quiet sounds all improve during the remaining gestational period (for reviews of empirical work see Busnel and Granier-Deferre, 1983; Lecanuet, 1996). As infants near term, their sensitivity to more complex auditory stimuli improves, allowing them to perceive details such as variations in music (Kisilevsky et al., 2004) and contrasting prosodic cues in familiar and novel rhymes (DeCasper et al., 1994). From this, one might conclude that development of auditory perceptual abilities has an initial advantage over vision, at least chronologically. On the other hand, and despite processing of visual stimuli beginning only postnatally (Turkewitz and Kenny, 1982; Slater, 2002), newborns' preference for faces (or face-like patterns) relative to any other visual stimulus is well documented (Goren et al., 1975; Morton and Johnson, 1991). This combination of early exposure in the auditory domain and precocious preference for faces—the source of spoken language—in the visual one would seem to position the newborn to easily recognize the relationship between spoken language and visual speech.

Not surprisingly, a talking face is more salient to a newborn than is a still face (Nagy, 2008), due at least in part to its inherent multimodality (Watson et al., 2014). But even when presented with a talking face with no accompanying sound (i.e., to visual speech alone), by the second half of the first year infants show greater sensitivity to the patterns of mouth movements found in their native language than in an unfamiliar language (Weikum et al., 2007). This suggests that they already recognize how specific movements of the visual articulators shape the speech signal, and a strong case has been made that the perception of the visual component of audiovisual speech facilitates the development of speech production abilities (Tenenbaum et al., 2015). Indeed, babbling infants tend to focus on the mouth of a speaker more than pre-babbling infants (Tenenbaum et al., 2013). Infants' own vocal productions interact with this as well, such that their real time attention to audiovisual speech changes as a function of their own articulatory modulations (Yeung and Werker, 2013); when presented with audiovisually produced vowels, infants imitate presentations more often when the audio and visual tokens are congruent than when they are incongruent (Legerstee, 1990). These and other findings inevitably lead to questions about what role, if any, the motor system plays in speech processing (e.g., Liberman and Mattingly, 1985). However, where perception of audiovisual speech clearly engages regions of sensorimotor cortex in both children and adults (Dick et al., 2010), other data indicate that motor activation is not necessary for audiovisual speech integration (Matchin et al., 2014). Therefore, we will set that debate aside to focus on the issue of integration itself.

Although a growing body of evidence demonstrates that substantial fine-tuning for various forms of audiovisual processing continues throughout childhood and well into adolescence (Baart et al., 2015; Tomalski, 2015), suffice it to say that at least some primitive form of multimodal perception emerges in early infancy (Bahrick et al., 2004). This can be characterized as guided by both *modal* cues (i.e., those that

are specific to a single modality, such as color information in the visual domain or the timbre of someone's voice in the auditory domain) and *amodal* ones (i.e., those that are available across modalities and are thus redundant; Bahrick, 1988). These amodal cues provide perceptual evidence that distinct sensory events can share a point of origin. By gaining experience with the correlated cues in audiovisual speech (or their intersensory redundancy, Lickliter and Bahrick, 2000), infants should come to identify information shared between them.

## ASSOCIATION IS NOT INTEGRATION

What remains unclear is when in the course of development association of these cues becomes actual integration of them. This is because, generally speaking, research techniques that are compatible with testing infants do not allow researchers to distinguish between these two processes. While this may seem like a subtle distinction, it is not a trivial one, in that it differentiates between those neural systems that evaluate cross-modal coincidence of physical stimuli (association) and those that actually mediate perceptual binding (integration; Miller and D'Esposito, 2005). Substantial animal research indicates that cumulative perceptual experience is critical to the development of the neural foundation for integration (Wallace and Stein, 2007; Yu et al., 2010), where presumably the cortical regions that contribute to such perceptual coding are fed by those regions engaged in initial associations between stimuli. It follows, then, that infants' perception of the relationship between the auditory and visual signals, as measured by looking procedures, contributes to the development of those neural underpinnings that will eventually support adult-like audiovisual integration. But implicit in that is the view that association precedes integration. The primary challenge to our understanding of the time course of this developmental process is that we have limited research methodologies for probing infants' perceptual experiences in a way that differentiates between behavioral evidence of association (e.g., looking behavior) and integration (e.g., some measure of perceptual fusion; c.f., Rosenblum et al., 1997). Although advances in infant-friendly neurophysiological testing techniques are allowing researchers new ways of tackling this issue (e.g., Kushnerenko et al., 2013), there remain many constraints on what can be reasonably asked of (and therefore concluded about) infant perception, whether with behavioral or neurophysiological techniques.

Nonetheless, infants clearly demonstrate sensitivity to audiovisual relations (see Shaw et al., 2015, for an example of how familiarity and coherence differentially influence infants' perception of audiovisual speech). Interest in the topic stemmed initially from a now classic study, in which 4-month-olds matched auditory vowels to videos of their corresponding articulation (Kuhl and Meltzoff, 1982). Follow-up studies replicated that original finding and extended it to male speakers (Patterson and Werker, 1999), as well as to infants of younger ages (Patterson and Werker, 2003). However, when the structured spectral elements of speech were replaced with simple tones, 5-month-olds struggled to recognize the

appropriate cross-modal match (Kuhl and Meltzoff, 1984; Kuhl et al., 1991). Because of this, much of the theoretical discussion of these early findings focused on whether and to what degree infants show privileged processing of speech and whether that indicates they have early access to phonetic representations. In the process, infants' ability to simply *match* auditory and visual streams was often mischaracterized as their ability to *integrate* audiovisual speech, leading to the loss of this important distinction. This formed the basis for much of the subsequent disagreement about early perceptual integration abilities. In more recent years, although this source of confusion has been recognized (see Stein et al., 2010, for a review), the broadly held view that infants integrate (rather than associate) has prevented the establishment of a more mechanistic account of how, for example, early association happens, and how it relates to the development of integration at a neural level.

## NON-COMPARABLE STIMULI

Another source of confusion stems from generalizations made based on findings obtained using stimuli that vary in complexity. For example, much of the early infant research employed the simplest form of audiovisual speech possible: single vowels or consonant-vowel combinations (e.g., Kuhl and Meltzoff, 1984). And, although these stimuli were characterized as audiovisual speech, it is well understood that the cues that support comprehension are both spatial and temporal in nature. For example, one of the strongest available cues is timing (i.e., temporal correlations between duration, onsets, offsets, and rate of the auditory and visual streams; Parise et al., 2012), so the truncated speech stimuli used in many of the early studies inadvertently limited infants' access to that class of cues. In other words, the infant data demonstrate their sensitivity to how visual spatial cues relate to auditory spectral cues (and vice versa) but say nothing about their ability to map articulator motion to the unfolding temporal information in continuous speech. Infants *are* sensitive to timing relationships in a variety of simple non-speech, multimodal events (Lewkowicz, 1992, 1994, 2003), but their ability to deal with timing relationships between streams of continuous auditory and visual speech has only recently become the focus of systematic research (e.g., Baart et al., 2014; Kubicek et al., 2014; Lewkowicz et al., 2015; Shaw et al., 2015).

Beyond inconsistencies in stimulus complexity, there are other sources of variability in infant audiovisual research, such as which dimension (spectral or temporal) is manipulated to create the non-matching (i.e., control) stimuli. Although these are not entirely orthogonal sources of information, spectral integration generally relies more on stimulus congruence and temporal integration generally relies more on stimulus timing. Much of the behavioral research with infants has been conducted using some form of a multimodal preferential looking technique in which one of two side-by-side visual displays matches the auditory stream while the other does not. The non-matching stimulus might differ in congruence (i.e., a different stimulus, such as visual /e/ and visual /a/ presented side-by-side with auditory /e/) or in timing (i.e., the identical stimulus but offset

in time relative to the audio). Congruence traditionally has been the more commonly manipulated dimension, as reflected by the matching/non-matching vowel stimuli used by Kuhl and colleagues in their early work. The McGurk effect (McGurk and MacDonald, 1976) also motivated a substantial line of research on perceptual fusion, typically with a single screen, and auditory and visual streams of single consonant-vowel pairs that are either congruent or non-congruent. In recent years, researchers have made substantial progress in using these sorts of stimuli in combination with electrophysiological measures with infants to identify neural indicators of perceptual fusion (e.g., Kushnerenko et al., 2008), but the former approach is far more commonly used.

Likewise, the synchrony of auditory and visual timing was manipulated early on (e.g., Dodd, 1979), revealing that older children (between 10 and 20 months of age) prefer synchronous over asynchronous running speech. More recently, questions have been raised about the extended developmental time course of such timing sensitivities and whether the temporal binding window continues to adjust further on in development. This refers to the period during which two sensory events can be separated in time yet still be perceptually bound into a unified event (see Wallace and Stevenson, 2014). Critically, testing this sensitivity requires temporally manipulating stimuli (i.e., comparing synchronous to non-synchronous audiovisual signals) rather than spatially manipulating them (i.e., comparing visual speech that matches the auditory speech to that which does not). If individuals have a temporal binding window that is too large, they may erroneously bind those events together (Van Wassenhove et al., 2007). In contrast, if the window is too narrow, individuals may be overly sensitive to whatever temporal discontinuity exists between two events and fail to recognize a cause-effect relationship between them (Dogge et al., 2012; Stevenson et al., 2012). Growing evidence of age-related differences in this form of temporal sensitivity is adding support to the view that data on infant association does not necessarily reflect integration of the sort that the temporal binding measures. For example, adolescents and pre-adolescents have larger temporal binding windows for audiovisual non-speech displays than older adolescents and adults (Hillock et al., 2011; Innes-Brown et al., 2011), and infants fail to indicate any sensitivity to temporal asynchrony unless the component signals are offset by over half a second (Lewkowicz, 2010; Pons et al., 2012).

While the research on timing sensitivities in typical development is still limited, there is even less data from atypical populations. Nevertheless, interest has grown recently in the role that temporal binding plays in a variety of developmental disorders such as autism (Bebko et al., 2006; Foss-Feig et al., 2010; de Boer-Schellekens et al., 2013) and dyslexia (Hairston et al., 2005), as well as with speech processing by cochlear implant users (Bergeson et al., 2005). Temporal-order-judgment tasks reveal that individuals with dyslexia, even when given non-linguistic audiovisual signals, tend to provide simultaneity judgments at longer lags than typical readers (Hairston et al., 2005). In this case, wider temporal binding windows may underlie reading deficits, reflecting poor temporal sensitivity to the auditory signal, visual signal, or both. By better understanding audiovisual

integration and the factors that lead to appropriate binding of events across senses, we will better understand the pathways leading to different developmental disorders and whether atypical perceptual integration may be at their base (Wallace and Stevenson, 2014).

## FURTHER ISOLATING SPECTRAL AND TEMPORAL INFLUENCES ON PROCESSING

While the correlation between the spectral and temporal information in the visual and auditory components of audiovisual speech makes it difficult to determine the influence of each, researchers have begun trying to isolate these components by degrading stimuli, for example, by using vocoded or sine wave speech (e.g., Tuomainen et al., 2005; Möttönen et al., 2006; Vroomen and Baart, 2009). Sine wave speech is natural speech that is synthetically reduced to three sinusoids replicating the frequency and amplitude of the first three formants (Remez et al., 1981). Unlike typical speech signals, sine wave speech is stripped of most extraneous spectral cues yet retains the temporal qualities of natural speech. Adults have difficulty recognizing the underlying phonetic content of sine wave speech unless they have been trained to hear it as language, or put into “speech-mode” (Vroomen and Baart, 2009). Because of this, sine wave speech is an ideal tool for examining the relative influence of top-down and bottom-up information on speech perception, and it is proving useful in isolating the relative influences of spectral and temporal information in infants’ processing of audiovisual speech (e.g., Baart et al., 2014).

In typical experiments, participants are first exposed to sine wave speech without prior knowledge of its relationship to natural speech. After a training phase in which participants are put into speech mode, they are tested again to ascertain whether phonetic knowledge provides a top-down processing advantage in speech perception. Differences between naïve and informed sine wave speech perception demonstrate that the top-down forces (e.g., phonetic representations) underlie a variety of perceptual phenomena, including phonetic recalibration (Vroomen and Baart, 2009), McGurk responses (Vroomen and Stekelenburg, 2011), and enhanced neural responsiveness (Stekelenburg and Vroomen, 2012). So what happens when participants do not have access to the phonetic representation corresponding to the sine-wave signal, as is the case with young infants?

There are clues from an early series of studies in which infants’ audiovisual perception was tested using stimuli that, though not sine wave speech, were quite similar to it. In an effort to assess which cues infants were relying on to cross-modally match audio and visual vowels in their initial study (Kuhl and Meltzoff, 1982), Kuhl and colleagues (Kuhl and Meltzoff, 1984; Kuhl et al., 1991) then asked whether modulating the spectral content of the acoustic signal impaired this ability. Four- to five-month-old infants were presented with audiovisual displays of a model silently articulating target vowels, but the auditory vowels were replaced by either pure tones, tones

that matched the fundamental frequencies of the vowels, or three-tone vowel analogs somewhat akin to sine wave speech (i.e., tones were matched to the first three formants of the naturally spoken vowels). As before, when given the natural acoustic speech signal, infants matched the auditory vowels to the appropriate articulating face. However, across all three spectral manipulations, they failed to attend to the matching face relative to the mismatching face.

Although not interpreted by the authors as such, these results suggest that temporal correlations between the auditory and visual signals did not provide enough information for infants to match stimuli across the auditory and visual modalities. Instead, Kuhl and colleagues suggested that the phonetic identity of the component signals served as the basis for early audiovisual sensitivity and that infants needed the natural speech stimulus (with its full phonetic realization) to process these cross-modal relationships. Moreover, they argued that audiovisual speech perception is a holistic process whereby infants are relatively insensitive to low-level cues. Therefore, when the phonetic content of the stimulus is reduced, any top-down processing advantages for infants are eliminated. In other words, their argument was that spectral information above and beyond the first three formants must be available for infants to combine heard and seen speech.

Critically, however, this study suffers from both of the stimulus problems we have outlined (i.e., very short stimuli; congruency manipulation rather than timing manipulation). Given a single vowel, it is not surprising that infants were unable to use the degraded spectral information to match the auditory to the visual vowel because there was virtually no corresponding temporal information to support them in the process. In recent research (Baart et al., 2014), we have addressed this problem by giving infants longer stimuli. In this study, we presented infants and adults with trisyllabic non-words in natural speech or the sine wave tokens of that speech, together with two visual displays of the same woman articulating each of the two non-words. In both the natural speech and sine wave speech conditions, only one display matched the auditory signal. Adults performed significantly worse with sine wave speech than natural speech across trials, suggesting that they were unable to match the articulatory information in the degraded auditory signal to the corresponding visual speech. In contrast, infants performed identically for both sine wave speech and natural speech, apparently able to access whatever cues existed across both signals to appropriately match the audio to the visual display. It is important to note, however, that infants performed significantly worse than adults did with natural speech; after all, adults have full access to the detailed phonetic representations that being a native speaker of a language entails. Not surprisingly, they performed near ceiling in this simple matching task when the full spectral and temporal information is made available. Without it, however, they were not able to use the temporal cues any more than the infants. Critically, there was no difference in infants’ performance in the natural speech and sine wave speech conditions, indicating that the temporal correlation between the auditory and visual signals was the basis for their performance rather than the spectral content of the speech itself. In other

words, infants' audiovisual association—at least in this case—was driven by relatively low level timing cues rather than by any form of phonetic representation. Importantly, this was only revealed by providing infants with the relevant temporal information in the form of sufficiently long stimuli, as well as by varying their access to the spectral information.

We are the first to admit that much remains unclear about how infants use spectral and temporal cues in audiovisual speech and how this contributes to their development of mature audiovisual integration. Nonetheless, we would argue that the factors we have identified here (i.e., lack of terminological

precision, paradigmatic differences, variable stimulus length, and inconsistent manipulation of spectral and temporal dimensions of test stimuli) underlie much of the disagreement about infants' audiovisual perceptual abilities. Attention to such factors will improve the quality of the research and the clarity of the discussion.

## FUNDING

This work was supported by NIH R01 DC10075 and National Science Foundation IGERT Training Grant 114399.

## REFERENCES

- Baart, M., Bortfeld, H., and Vroomen, J. (2015). Phonetic matching of auditory and visual speech develops during childhood: evidence from sine-wave speech. *J. Exp. Child Psychol.* 129, 157–164. doi: 10.1016/j.jecp.2014.08.002
- Baart, M., Vroomen, J., Shaw, K., and Bortfeld, H. (2014). Degrading phonetic information affects matching of audiovisual speech in adults, but not in infants. *Cognition* 130, 31–43. doi: 10.1016/j.cognition.2013.09.006
- Bahrack, L. E. (1988). Intermodal learning in infancy: learning on the basis of two kinds of invariant relations in audible and visible events. *Child Dev.* 59, 197–209. doi: 10.2307/1130402
- Bahrack, L. E., Lickliter, R., and Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Curr. Dir. Psychol. Sci.* 13, 99–102. doi: 10.1111/j.0963-7214.2004.00283.x
- Bebko, J. M., Weiss, J. A., Demark, J. L., and Gomez, P. (2006). Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *J. Child Psychol. Psychiatry* 47, 88–98. doi: 10.1111/j.1469-7610.2005.01443.x
- Bergeson, T. R., Pisoni, D. B., and Davis, R. A. O. (2005). Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. *Ear Hear.* 26, 149–164. doi: 10.1097/00003446-200504000-00004
- Busnel, M. C., and Granier-Deferre, C. (1983). "And what of fetal audition?" in *The Behavior of Human Infants*, eds A. Oliveira and M. Zappella (New York, NY: Plenum), 93–126.
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., and Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Comput. Biol.* 5:e1000436. doi: 10.1371/journal.pcbi.1000436
- de Boer-Schellekens, L., Eussen, M., and Vroomen, J. (2013). Diminished sensitivity of audiovisual temporal order in autism spectrum disorder. *Front. Integr. Neurosci.* 7:8. doi: 10.3389/fnint.2013.00008
- DeCasper, A. J., Lecanuet, J. P., Busnel, M. C., Granier-Deferre, C., and Maugeais, R. (1994). Fetal reactions to recurrent maternal speech. *Infant Behav. Dev.* 17, 159–164. doi: 10.1016/0163-6383(94)90051-5
- DeCasper, A. J., and Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behav. Dev.* 9, 133–150. doi: 10.1016/0163-6383(86)90025-1
- Dick, A. S., Solodkin, A., and Small, S. L. (2010). Neural development of networks for audiovisual speech comprehension. *Brain Lang.* 114, 101–114. doi: 10.1016/j.bandl.2009.08.005
- Dodd, B. (1979). Lip reading in infants: attention to speech presented in-and-out-of-synchrony. *Cogn. Psychol.* 11, 478–484. doi: 10.1016/0010-0285(79)90021-5
- Dogge, M., Schaap, M., Custers, R., Wegner, D. M., and Aarts, H. (2012). When moving without volition: implied self-causation enhances binding strength between involuntary actions and effects. *Conscious. Cogn.* 21, 501–506. doi: 10.1016/j.concog.2011.10.014
- Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., et al. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Exp. Brain Res.* 203, 381–389. doi: 10.1007/s00221-010-2240-4
- Goren, C. C., Sarty, M., and Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics* 56, 544–549.
- Graven, S. N., and Browne, J. V. (2008). Sensory development in the fetus, neonate, and infant: introduction and overview. *Newborn Infant Nurs. Rev.* 8, 169–172. doi: 10.1053/j.nainr.2008.10.007
- Hairston, W. D., Burdette, J. H., Flowers, D. L., Wood, F. B., and Wallace, M. (2005). Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Exp. Brain Res.* 166, 474–480. doi: 10.1007/s00221-005-2387-6
- Hepper, P. G., and Shahidullah, B. S. (1994). Development of fetal hearing. *Arch. Dis. Child.* 71, F81–F87. doi: 10.1136/fn.71.2.F81
- Hillock, A. R., Powers, A. R., and Wallace, M. (2011). Binding of sights and sounds: age-related changes in multisensory temporal processing. *Neuropsychologia* 49, 461–467. doi: 10.1016/j.neuropsychologia.2010.11.041
- Innes-Brown, H., Barutcha, A., Shivdasani, M. N., Crewther, D. P., Grayden, D. B., and Paolini, A. G. (2011). Susceptibility to the flash-beep illusion is increased in children compared to adults. *Dev. Sci.* 14, 1089–1099. doi: 10.1111/j.1467-7687.2011.01059.x
- Jardri, R., Pins, D., Houfflin-Debarge, V., Chaffiotte, C., Rocourt, N., Pruvo, J. P., et al. (2008). Fetal cortical activation to sound at 33 weeks of gestation: a functional MRI study. *Neuroimage* 42, 10–18. doi: 10.1016/j.neuroimage.2008.04.247
- Kisilevsky, B. S., Hains, S. M. J., Jacquet, A. Y., Granier-Deferre, C., and Lecanuet, J. P. (2004). Maturation of fetal response to music. *Dev. Sci.* 7, 550–559. doi: 10.1111/j.1467-7687.2004.00379.x
- Kubicek, C., Hillairet de Boisferon, A., Dupierriex, E., Pascalis, O., Lœvenbruck, H., Gervain, J., et al. (2014). Cross-modal matching of audio-visual German and French fluent speech in infancy. *PLoS ONE* 9:e89275. doi: 10.1371/journal.pone.0089275
- Kuhl, P. K., and Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science* 218, 1138–1141. doi: 10.1126/science.7146899
- Kuhl, P. K., and Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behav. Dev.* 7, 361–381. doi: 10.1016/S0163-6383(84)80050-8
- Kuhl, P. K., Williams, K. A., and Meltzoff, A. N. (1991). Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 829–840. doi: 10.1037/0096-1523.17.3.829
- Kushnerenko, E., Teinonen, T., Volein, A., and Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11442–11445. doi: 10.1073/pnas.0804275105
- Kushnerenko, E., Tomalski, P., Ballieux, H., Ribeiro, H., Potton, A., Axelsson, E., et al. (2013). Brain responses to audiovisual speech mismatch in infants are associated with individual differences in looking behaviour. *Eur. J. Neurosci.* 38, 3363–3369. doi: 10.1111/ejn.12317
- Lecanuet, J. P. (1996). "Prenatal auditory experience," in *Musical Beginnings: Origins and Development of Musical Competence*, eds I. Deliège and J. Sloboda (New York, NY: Oxford University Press), 3–34.
- Legerstee, M. (1990). Infants use multimodal information to imitate speech sounds. *Infant Behav. Dev.* 13, 343–354. doi: 10.1016/0163-6383(90)90039-B
- Lewkowicz, D. (1992). Infants' responsiveness to the auditory and visual attributes of a sounding/moving stimulus. *Percept. Psychophys.* 52, 519–528. doi: 10.3758/BF03206713
- Lewkowicz, D. J. (1994). Limitations on infants' response to rate-based auditory-visual relations. *Dev. Psychol.* 30, 880–892. doi: 10.1037/0012-1649.30.6.880



- Lewkowicz, D. J. (2003). Learning and discrimination of audiovisual events in human infants: the hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Dev. Psychol.* 39, 795–804. doi: 10.1037/0012-1649.39.5.795
- Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Dev. Psychol.* 46, 66–77. doi: 10.1037/a0015579
- Lewkowicz, D. J., Minar, N. J., Tift, A. H., and Brandon, M. (2015). Perception of the multisensory coherence of fluent audiovisual speech in infancy: its emergence and the role of experience. *J. Exp. Child Psychol.* 130, 147–162. doi: 10.1016/j.jecp.2014.10.006
- Liberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36. doi: 10.1016/0010-0277(85)90021-6
- Lickliter, R., and Bahrack, L. E. (2000). The development of infant intersensory perception: advantages of a comparative convergent-operations approach. *Psychol. Bull.* 126, 260–280. doi: 10.1037/0033-2909.126.2.260
- Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Dev.* 55, 1777–1788. doi: 10.2307/1129925
- Matchin, W., Groulx, K., and Hickok, G. (2014). Audiovisual speech integration does not rely on the motor system: evidence from articulatory suppression, the McGurk Effect, and fMRI. *J. Cogn. Neurosci.* 26, 606–620. doi: 10.1162/jocn\_a\_00515
- McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0
- Middelweerd, M. J., and Plomp, R. (1987). The effect of speechreading on the speech-reception threshold of sentences in noise. *J. Acoust. Soc. Am.* 82, 2145–2147. doi: 10.1121/1.395659
- Miller, L. M., and D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J. Neurosci.* 25, 164–181. doi: 10.1523/JNEUROSCI.0896-05.2005
- Morton, J., and Johnson, M. H. (1991). CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychol. Rev.* 98, 164–181. doi: 10.1037/0033-295X.98.2.164
- Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., et al. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage* 30, 563–569. doi: 10.1016/j.neuroimage.2005.10.002
- Nagy, E. (2008). Innate intersubjectivity: newborns' sensitivity to communication disturbance. *Dev. Psychol.* 44, 1779–1784. doi: 10.1037/a0012665
- Navarra, J., and Soto-Faraco, S. (2007). Hearing lips in a second language: visual articulatory information enables the perception of second language sounds. *Psychol. Res.* 71, 4–12. doi: 10.1007/s00426-005-0031-5
- Parise, C. V., Spence, C., and Ernst, M. O. (2012). When correlation implies causation in multisensory integration. *Curr. Biol.* 22, 46–49. doi: 10.1016/j.cub.2011.11.039
- Patterson, M. L., and Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Dev. Sci.* 6, 191–196. doi: 10.1111/1467-7687.00271
- Patterson, M. L., and Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behav. Dev.* 22, 237–247. doi: 10.1016/S0163-6383(99)00003-X
- Pons, F., Teixidó, M., Garcia-Morera, J., and Navarra, J. (2012). Short-term experience increases infants' sensitivity to audiovisual asynchrony. *Infant Behav. Dev.* 35, 815–818. doi: 10.1016/j.infbeh.2012.06.006
- Pujol, R., Lavigne-Rebillard, M., and Uziel, A. (1991). Development of the human cochlea. *Acta Otolaryngol.* 111, 7–13. doi: 10.3109/00016489109128023
- Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science* 212, 947–949. doi: 10.1126/science.7233191
- Rosenblum, L. D., Schmuckler, M. A., and Johnson, J. A. (1997). The McGurk effect in infants. *Percept. Psychophys.* 59, 347–357. doi: 10.3758/BF03211902
- Rubel, E. W., and Ryals, B. M. (1983). Development of the place principle: acoustic trauma. *Science* 219, 512–514. doi: 10.1126/science.6823549
- Shaw, K., Baart, M., Depowski, N., and Bortfeld, H. (2015). Infants' preference for native audiovisual speech dissociated from congruency preference. *PLoS ONE* 10:e0126059. doi: 10.1371/journal.pone.0126059
- Slater, A. (2002). Visual perception in the newborn infant: issues and debates. *Intellectica* 34, 57–76. Available online at: [http://intellectica.org/SiteArchives/archives/n34/n34\\_table.htm](http://intellectica.org/SiteArchives/archives/n34/n34_table.htm)
- Stein, B. E., Burr, D., Constantinidis, C., Laurienti, P. J., Alex Meredith, M., Perrault, T. J., et al. (2010). Semantic confusion regarding the development of multisensory integration: a practical solution. *Eur. J. Neurosci.* 31, 1713–1720. doi: 10.1111/j.1460-9568.2010.07206.x
- Stekelenburg, J. J., and Vroomen, J. (2012). Electrophysiological evidence for a multisensory speech-specific mode of perception. *Neuropsychologia* 50, 1425–1431. doi: 10.1016/j.neuropsychologia.2012.02.027
- Stevenson, R. A., Zemtsov, R. K., and Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 1517–1529. doi: 10.1037/a0027339
- Sumbly, W. H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215. doi: 10.1121/1.1907309
- Teinonen, T., Aslin, R. N., Alku, P., and Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition* 108, 850–855. doi: 10.1016/j.cognition.2008.05.009
- Tenenbaum, E. J., Shah, R. J., Sobel, D. M., Malle, B. F., and Morgan, J. L. (2013). Increased focus on the mouth among infants in the first year of life: a longitudinal eye-tracking study. *Infancy* 18, 534–553. doi: 10.1111/j.1532-7078.2012.00135.x
- Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., and Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *J. Child Lang.* 42, 1173–1190. doi: 10.1017/S0305000914000725
- Tomalski, P. (2015). Developmental trajectory of audiovisual speech integration in early infancy: a review of studies using the McGurk paradigm. *Psychol. Lang. Commun.* 19, 77–100. doi: 10.1515/plc-2015-0006
- Tuomainen, J., Andersen, T. S., Tiippana, K., and Sams, M. (2005). Audio-visual speech perception is special. *Cognition* 96, B13–B22. doi: 10.1016/j.cognition.2004.10.004
- Turkewitz, G., and Kenny, P. A. (1982). Limitations on input as a basis for neural organization and perceptual development: a preliminary theoretical statement. *Dev. Psychobiol.* 15, 357–368. doi: 10.1002/dev.420150408
- Van Wassenhove, V., Grant, K. W., and Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* 45, 598–607. doi: 10.1016/j.neuropsychologia.2006.01.001
- Vroomen, J., and Baart, M. (2009). Phonetic recalibration only occurs in speech mode. *Cognition* 110, 254–259. doi: 10.1016/j.cognition.2008.10.015
- Vroomen, J., and Stekelenburg, J. J. (2011). Perception of intersensory synchrony in audiovisual speech: not that special. *Cognition* 118, 75–83. doi: 10.1016/j.cognition.2010.10.002
- Wallace, M., and Stein, B. (2007). Early experience determines how the senses will interact. *J. Neurophysiol.* 97, 921–926. doi: 10.1152/jn.00497.2006
- Wallace, M. T., and Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia* 64, 105–123. doi: 10.1016/j.neuropsychologia.2014.08.005
- Watson, T. L., Robbins, R. A., and Best, C. T. (2014). Infant perceptual development for faces and spoken words: an integrated approach. *Dev. Psychobiol.* 56, 1454–1481. doi: 10.1002/dev.21243
- Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., and Werker, J. F. (2007). Visual language discrimination in infancy. *Science* 316, 1159–1159. doi: 10.1126/science.1137686
- Yeung, H. H., and Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychol. Sci.* 24, 603–612. doi: 10.1177/0956797612458802
- Yu, L., Rowland, B. A., and Stein, B. E. (2010). Initiating the development of multisensory integration by manipulating sensory experience. *J. Neurosci.* 30, 4904–4913. doi: 10.1523/JNEUROSCI.5575-09.2010

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Shaw and Bortfeld. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Action sounds update the mental representation of arm dimension: contributions of kinaesthesia and agency

Ana Tajadura-Jiménez<sup>1\*</sup>, Manos Tsakiris<sup>2</sup>, Torsten Marquardt<sup>3</sup> and Nadia Bianchi-Berthouze<sup>1</sup>

<sup>1</sup> UCL Interaction Centre, University College London, University of London, London, UK, <sup>2</sup> Lab of Action and Body, Department of Psychology, Royal Holloway, University of London, Egham, UK, <sup>3</sup> UCL Ear Institute, University College London, University of London, London, UK

## OPEN ACCESS

### Edited by:

Achille Pasqualotto,  
Sabanci University, Turkey

### Reviewed by:

Konstantina Kilteni,  
Karolinska Institute, Sweden  
Elisa Canzoneri,  
École Polytechnique Fédérale  
de Lausanne, Switzerland

### \*Correspondence:

Ana Tajadura-Jiménez,  
UCL Interaction Centre, University  
College London, University  
of London, Malet Place Engineering  
Building, 8th Floor, Gower Street,  
London WC1E 6BT, UK  
a.tajadura@ucl.ac.uk

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 13 March 2015

**Accepted:** 10 May 2015

**Published:** 29 May 2015

### Citation:

Tajadura-Jiménez A, Tsakiris M,  
Marquardt T and Bianchi-Berthouze N  
(2015) Action sounds update  
the mental representation of arm  
dimension: contributions  
of kinaesthesia and agency.  
*Front. Psychol.* 6:689.  
doi: 10.3389/fpsyg.2015.00689

Auditory feedback accompanies almost all our actions, but its contribution to body-representation is understudied. Recently it has been shown that the auditory distance of action sounds recalibrates perceived tactile distances on one's arm, suggesting that action sounds can change the mental representation of arm length. However, the question remains open of what factors play a role in this recalibration. In this study we investigate two of these factors, kinaesthesia, and sense of agency. Across two experiments, we asked participants to tap with their arm on a surface while extending their arm. We manipulated the tapping sounds to originate at double the distance to the tapping locations, as well as their synchrony to the action, which is known to affect feelings of agency over the sounds. Kinaesthetic cues were manipulated by having additional conditions in which participants did not displace their arm but kept tapping either close (Experiment 1) or far (Experiment 2) from their body torso. Results show that both the feelings of agency over the action sounds and kinaesthetic cues signaling arm displacement when displacement of the sound source occurs are necessary to observe changes in perceived tactile distance on the arm. In particular, these cues resulted in the perceived tactile distances on the arm being felt smaller, as compared to distances on a reference location. Moreover, our results provide the first evidence of consciously perceived changes in arm-representation evoked by action sounds and suggest that the observed changes in perceived tactile distance relate to experienced arm elongation. We discuss the observed effects in the context of forward internal models of sensorimotor integration. Our results add to these models by showing that predictions related to action sounds must fit with kinaesthetic cues in order for auditory inputs to change body-representation.

**Keywords:** auditory-dependent body-representation, kinaesthesia, agency, action sounds, body-related sensory inputs

## Introduction

Sounds accompany almost every bodily movement and action we produce. Think for instance about the sound of your footsteps, the impact sound of an object falling from your hand onto the floor, or the sound produced when typing on a keyboard. These sounds are highly rich in information about one's own body and its effects on the outside world; for instance, footstep sounds vary according to body weight and strength, as well as according to the emotional state of the walker (Li et al., 1991; Bresin et al., 2010; Tajadura-Jiménez et al., 2015a). But, to what extent are we making use of this "soundtrack" that accompanies most of our actions, for gathering information about one's actions and body? Here we focus on recent findings that sounds produced when tapping one's hand on a surface recalibrate the mental representation of arm length (Tajadura-Jiménez et al., 2012). We specifically seek to disambiguate the effects of kinaesthetic cues and feelings of agency in this recalibration.

Action- and body-awareness are critical for our interaction with the environment. For instance, according to our perceived body dimensions, we may ponder whether we can reach a particular object or whether there is enough space for us to get onto a crowded bus. Importantly, research has shown that the mental representation of our body (i.e., our body-representation) is not fixed, but it is continuously updated by the body-related multisensory cues received from the environment (de Vignemont, 2010; Longo et al., 2010; Serino and Haggard, 2010). For example, an artificial hand may feel like part of one's own body when one sees it being touched and in synchrony receives touch on one's own, unseen, hand. This is the result of the integration of information coming from different sensory channels – vision, touch, and proprioception (Botvinick and Cohen, 1998). Similarly, altering proprioception (de Vignemont et al., 2005; Ehrsson et al., 2005) or vestibular information (Lopez et al., 2012; although see also Ferrè et al., 2013) may result in perceived distortions in body size. Studies using virtual reality set-ups have shown that observing a very long arm (Kilteni et al., 2012; Preston and Newport, 2012) or a very large or very small body (van der Hoort et al., 2011) can result in the illusion of owning that arm or body, provided that visuo-tactile and visuo-motor temporal and spatial congruency is kept constant between the observed body and one's own felt body. Furthermore, using a tool to act with one's arm upon relatively distant objects can also result in an increase of represented arm length (Cardinali et al., 2009, 2012; Canzoneri et al., 2013b).

Despite this known link between body-related sensory cues and body-representation, the contribution of auditory cues to body-awareness has been addressed only in a few research studies. It has been shown that action sounds recruit motor areas of the brain that are involved in the planning, preparation, and observation of these actions (Aglioti and Pazzaglia, 2010; see Pazzaglia et al., 2008 for related findings in the visual domain). In addition, there are evidences that self-produced action sounds can also influence the way actions are subsequently performed. For example, altering in real-time the sound of someone's footsteps influences her walking style (Bresin et al., 2010; Menzer

et al., 2010) and altering cues related to applied strength on sounds generated by tapping one's hand on a surface influences the tapping behavior (Tajadura-Jiménez et al., 2015b). Regarding awareness of one's own body, on the one hand, it is known that blocking audition by wearing earplugs often results on people reporting an altered body-awareness, apart from a sensation of detachment from the surroundings (Murray et al., 2000). On the other hand, the provision of sound feedback on body movement of a person with reduced body awareness and mobility is known to increase physical self-efficacy (Singh et al., 2014).

In addition to these links between movement and self-produced sound, a few studies have started to show that self-produced sounds contribute to update body-representation. For instance, it has been shown that self-produced sounds update the representation of one's own entire body size and weight (Tajadura-Jiménez et al., 2015a) and even the experienced material of one's own body (Senna et al., 2014). The former was achieved by altering the frequency of the self-produced walking sounds, and the latter by altering the sound of the impact of an object on one's hand. The first demonstration of a link between audition and body-representation was actually provided by a study in which we showed that represented limb length updates by action sounds (Tajadura-Jiménez et al., 2012). In that study we asked participants to tap on a surface while progressively extending their arm sideways. Exposure to tapping sounds originating at double the distance at which participants actually tapped, and presented in synchrony with the taps of participants (Double distance – 2D – condition), changed the perception of tactile distance on the tapping arm, as compared to the perceived tactile distances before tapping. These changes in perceived tactile distances on the arm evidenced a change in represented arm length (e.g., Taylor-Clarke et al., 2004; de Vignemont et al., 2005; Canzoneri et al., 2013a,b). The effects were not observed when the tapping sounds originated at quadruple the distance (4D condition; see also Kilteni et al., 2012, for similar findings on plasticity of represented arm length when manipulating visual cues) or when the sounds were presented in asynchrony with participants' taps (Double distance asynchronous – 2DA – condition). Self-reports showed that in the 2D condition, as opposed to the 4D condition, participants felt that sound and tap originated at the same location. They also showed that in the 2D condition, as opposed to the 2DA condition, participants felt that the sound was caused by their own hand tapping and that they were in control of their arm. Indeed, temporal contingency is known to be crucial for correct action attribution (Moore et al., 2009). Tajadura-Jiménez et al. (2012) also ran a second experiment in which participants did not generate the taps and did not displace their arm, but they received externally-generated taps to their still arm. Results showed that simply hearing sounds in synchrony and at double the distance at which taps are felt, while keeping the arm stationary, does not elicit changes in perceived tactile distance.

Hence, several factors might be implicated in the auditory-induced changes in perceived tactile distance, namely (1) the magnitude of the manipulation of auditory distance, (2) the synchrony between the tapping sounds and the participants'

taps, (3) the feeling of being the agent of the tapping sounds, and (4) the displacement of the arm when tapping and when displacement of the sound occurs. Which of these factors are necessary and/or sufficient to observe an effect on perceived tactile distance remains unknown. Tajadura-Jiménez et al. (2012) addressed the two first factors and showed that hearing sounds in synchrony and at double the distance at which taps occur were necessary but not sufficient factors to elicit changes in perceived tactile distance. Hence, a remaining question is about the third and fourth factors described above. We hypothesize that a coherent representation of the motor command sent to the tapping arm and the sound feedback received from the tapping action needs to arise during the audio-tactile adaptation in order to observe changes in felt tactile distance. According to the 'forward internal models' of the motor system (e.g., Wolpert and Ghahramani, 2000), both temporal and spatial mismatches between motor and sensory representations reduce the likelihood that different sources of sensorimotor information merge to form a coherent and robust percept (Ernst and Bühlhoff, 2004), and they interfere with the sense of control or agency over one's action (Blakemore et al., 2002).

We sought to disambiguate the effect of kinaesthetic cues from the feelings of agency on the observed auditory-driven changes in the representation of arm dimension. For that reason, we opted for keeping the sound presentation equal to that in the study by Tajadura-Jiménez et al. (2012), and manipulated instead kinesthetic cues and the feelings of agency over the generated tapping sounds across different conditions. We asked participants to tap with their arm on a surface. In the "Displacement" conditions participants were required to tap while extending their arm, and they were presented with tapping sounds that originated at double the distance to the tapping locations. Feelings of agency over the tapping sounds were manipulated by presenting the tapping sounds either in synchrony or in asynchrony with the tapping actions (i.e., we expected agency to be preserved only in the synchronous conditions). Across two experiments kinaesthetic cues signaling arm displacement, and therefore change in hand position, were manipulated by having additional control conditions ("No Displacement" conditions) in which participants did not displace their arm but kept tapping at a fixed location, which was either close to their body torso (i.e., arm flexed, in Experiment 1) or far from it (i.e., arm stretched, in Experiment 2). Importantly, the tapping sounds were presented at the same locations across all experimental conditions, with the tapping sounds originating at double the distance to the points where participants tapped during the arm Displacement conditions. Having two posture positions for the No Displacement conditions (i.e., arm flexed in Experiment 1 and arm completely stretched in Experiment 2) allowed controlling for the effect of distance between hand and body torso (close or far) and for the effect of distance between hand and sound source, which was larger for the posture adopted in Experiment 1 than for the posture in Experiment 2. We quantified the effects on subjective feelings and on perceived tactile distance related to represented arm length.

## Experiment 1

### Materials and Methods

#### Participants

Twenty participants ( $M_{\text{age}} \pm \text{SD} = 22.85 \pm 2.5$  years; age range from 18 to 28 years; 16 females) took part individually in the experiment. The sample size was chosen by a power analysis calculation based on our previous work (Tajadura-Jiménez et al., 2012). This calculation showed that with a sample size of 14, there was 80% likelihood that the study will yield a statistically significant difference between the means of the Synchronous and Asynchronous Displacement conditions. All participants reported having normal hearing and normal tactile perception, and were naïve as to the purposes of the study. They were paid for their time and gave their informed consent prior to their inclusion in the studies. The experiment was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and approved by the ethics committee of University College London.

#### Apparatus and Materials

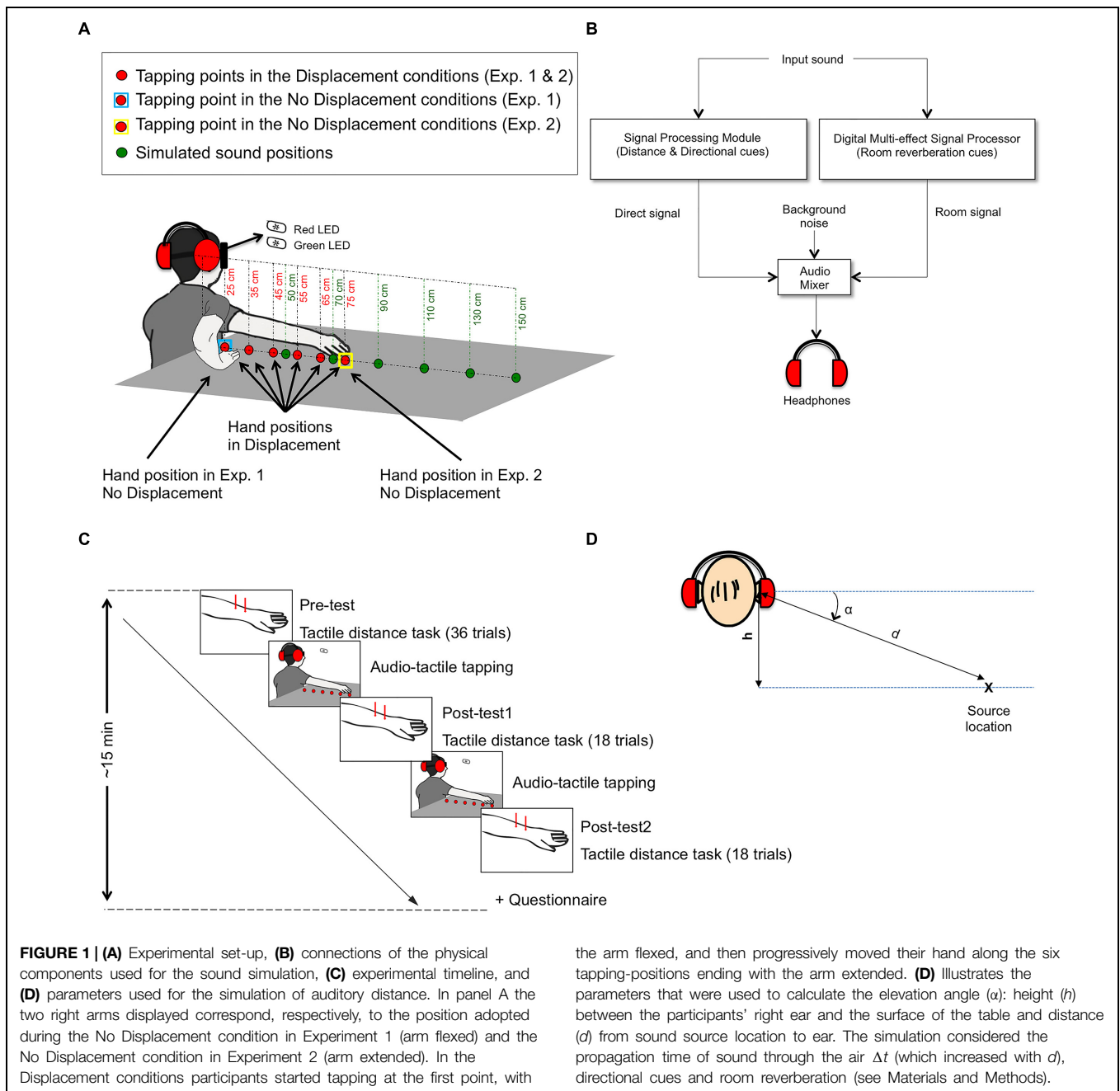
A schema of the experimental set-up is displayed in **Figure 1A**. Participants were seated in a chair, blindfolded, and wearing a pair of closed headphones with very high passive ambient noise attenuation (Sennheiser HDA 200). A pair of light-emitting diodes (LED), one green and one red, was positioned in front of the participants, at eye level and a distance of 50 cm. They were bright enough so that participants could see the light through the blindfold. The green LED served as the center fixation point, and the red LED was used by participants to perform the experimental task, as described in the next section. During the experimental blocks participants were instructed to refrain from turning their head sideways from the fixation point.

A table was placed to the right of the participants. The height ( $h$ ) between the participants' right ear and the surface of the table was approximately 40 cm. The participants were instructed to tap on the surface of the table, at six different positions ("tapping-positions"), which were located  $90^\circ$  to the right at 25 cm, 35 cm, 45 cm, 55 cm, 65 cm, and 75 cm from a vertical line traced between the participants' right ear and the table surface.

We simulated the auditory "source locations" by using virtual acoustic techniques (e.g., Begault, 1994). A "dry" recording of two fingers tapping on a cardboard box was made in an anechoic chamber. The recording lasted 125-ms and had a broad spectrum. This recording was later loaded onto a real-time signal processing module (RP2.1, Tucker-Davis Technologies) to manipulate the virtual location of the sound arriving directly from the sound-source. In a parallel processing path, room reverberation was added to the "dry" signal using a digital multi-effect signal processor (Digitech DSP 128 plus) to simulate a small room with low reflective surfaces ( $\text{RT60} = 0.36$  s). The direct and the room signal were then added (TEAC audio mixer model 2A) and presented to the right and left ear via stereo headphones (**Figure 1B**).

The "dry" pre-recorded sound was modified in the real-time-processor to provide the listener with distance cues





and directional cues using RPvdsEx software (Tucker–Davis Technologies). Increased source distance was simulated by increasing delay and decreasing intensity of the direct signal, thus decreasing the direct-to-reverberation ratio, which is one of the strongest distance cues in reverberant environments. The intensity  $I$  of the direct sound decreased with the square-power of the distance to ear  $d$  ( $I = \frac{1}{d^2}$ ; see **Figure 1D**). A delay  $\Delta t$ , of 3 ms per meter, was introduced to the direct sound to simulate the velocity of sound in air:

$$\Delta t = \frac{d}{\text{speed of sound}} = \frac{d[m]}{350[\frac{m}{s}]} = d \times 0.003 [s]$$

It should be noted that it is not this delay of the direct sound relative to the tapping sensation, but the consequently decreasing difference between the delay of the direct sound and fixed delay of the first reflection that provides a distance cue. The latency of the signal processing module (RP2) consisted largely of the A/D and D/A conversion time, and was in total less than 4 ms. Such short latency is unperceivable across sensory modalities, as it falls well within the intersensory temporal synchrony window (Lewkowicz, 1996, 1999).

Directional cues were then introduced to the direct sound by convolving the signal with the left and right sets of head-related transfer functions (HRTFs) that correspond to the desired spatial

direction of the source. Sets of generic HRTFs are provided by RPvdsEx software. We used the set for 90° azimuth, and an elevation angle  $\alpha$  (see **Figure 1D**):

$$\alpha = \arccos(h/d)$$

A piezoelectric transducer (Schaller Oyster 723 Piezo Transducer Pickup), attached to the table, was used to detect the participants' taps and trigger the auditory stimulation. In the Synchronous condition, the auditory stimulus was presented in synchrony with the participant's tap on the table. In the Asynchronous condition, the auditory stimulus was presented with a small delay with respect to the participant's tap. This delay varied randomly over a range of 300–800 ms. It should be noted that the minimum delay value (i.e., 300 ms) was chosen to fall outside of the multisensory integration window during which asynchronous stimuli in different modalities are perceived as simultaneous (Lewkowicz, 1996, 1999).

An array of six spatial “source locations” was simulated. The source locations were aligned with the tapping-positions but were at double the distance than those (i.e., the “source locations” were separated by 20 cm). An additional array (“practice array”) had “source locations” identical to the tapping-positions and was simulated for the practice block that participants performed to get familiar with the tasks.

In the arm Displacement conditions, the tapping sounds originated at double the distance to the tapping-positions (the Synchronous and Asynchronous conditions resemble, therefore, the 2D, and 2DA conditions in Tajadura-Jiménez et al., 2012). Hence, the last auditory stimulus of the experimental trial was delivered from the sixth source location in the array, 150 cm away from the vertical line traced from the participants' ear (while the last tapping-position was 75 cm away). In the No Displacement conditions, the tapping sounds originated at exactly the same locations than during the arm Displacement conditions. Thus, these conditions did not differ in the sounds presented, but rather in that participants did not displace their arm while tapping in the No Displacement conditions and they did in the Displacement conditions. The actual sound of the participants' taps on the table was attenuated by the high ambient-noise attenuation headphones, and masked by adding background noise (interaurally uncorrelated pink noise, 20–13000 Hz) to the headphone signals throughout the entire experimental session (see Procedure section).

The stimuli for the tactile distance task consisted of three pairs of wooden posts (diameter 3 mm) mounted in foam board, as in Longo and Haggard's (2011), study. The pairs of posts differed in the separation between the posts, which was fixed at 4, 5, or 6 cm. They were presented at two different body locations, the participant's right forearm (test stimuli) and the forehead (reference stimuli; see Canzoneri et al., 2013a,b, for similar procedure). The minimum distance of 4 cm was chosen to be clearly suprathreshold at both body locations (Nolan, 1982, 1985). Each tactile contact lasted for approximately 1 s.

MATLAB software was used to control stimulus delivery and record responses.

## Audio-Tactile “Tapping” Task

Participants were required to centrally fixate the green LED and to perform the simple action of tapping on the table using their right hand, while keeping the arm ventral side down. They tapped at the first tapping-position for ten times and the auditory stimulus was delivered at the first source location in the array, in synchrony or in asynchrony with the participant's tapping, depending on the condition. Participants were asked to pace their rhythm keeping a frequency of approximately one tap per second.

In the Displacement trials, after ten taps, a signal (red LED) indicated the participants to extend their arm rightward by 10 cm, and tap again for 10 times at the new tapping-position, with the auditory stimulus presented from the subsequent source location in the array (i.e., at double the distance to the tapping-position). This procedure was repeated six times, for a total of 60 taps on the table, in the Displacement trials 10 at each of the six tapping-positions, and in the No Displacement trials 60 taps at the same, first tapping-position. In the No Displacement trials, participants kept tapping at the first tapping-position while the auditory stimulus changed to the subsequent source location in the array every ten taps. After these 60 taps, participants were asked to repeat the procedure again, starting from the first tapping-position. Hence, an experimental trial included two sets of 60 taps. At the end of the Displacement trials, participants were instructed to keep the right hand open on the last tapping-position. At the end of the No Displacement trials, they were instructed to fully extend their arm and place their hand open on the last tapping-position, being assisted by the experimenter.

## “Tactile Distance” Task

This task, adapted from previous studies (de Vignemont et al., 2005; Lopez et al., 2012; Tajadura-Jiménez et al., 2012; Canzoneri et al., 2013a,b), serves as an indirect measure of the mentally represented body part size. Participants were required to adopt the same position that they had by the end of the audio-tactile “tapping” task, i.e., right arm extended laterally, ventral side up, with the hand open, and placed approximately 75 cm away. Dual tactile stimuli were delivered manually by the experimenter using pairs of wooden posts on two different body locations consequently (right forearm – test location and forehead – reference location), in a randomized order. The duration of each touch was approximately 1 s, with approximately 1 s between the touches to the two body locations. A sequence of 36 tactile trials, which constituted one “tactile distance” block, was generated beforehand and randomized. In one third of the trials the tactile distance on the test and reference locations was the same, in another third differed by  $\pm 1$  cm and in the last third differed by  $\pm 2$  cm. The task for participants was to indicate verbally whether the two points felt farther apart in the first or the second stimulated location (adapted from Longo and Haggard, 2011).

## Procedure

Participants sat on a chair and a sound test was performed to check listeners' perceived azimuth for the simulated sound sources. This test revealed that participants perceived the sound sources originating on average at 102° (SE = 5.26; range from 60 to 140°), which corresponds to locations on the right, slightly

back from participants' right ear. This test provided evidence that the simulated sound direction was perceived as expected. Then, participants were instructed and were asked to practice the tapping, paying attention to keep the required tapping rhythm (approximately one tap per second) and the location of the six tapping-positions (separated by 10 cm). Participants first practiced without wearing the blindfold, and then, once again wearing the blindfold, with the experimenter giving them feedback on their performance and correcting their movements if necessary. Next, they completed a full experimental practice block (Synchronous – Displacement, as described below) to familiarize themselves with the audio-tactile “tapping” and the tactile distance tasks. The audio-tactile “tapping” task in this practice block differed from the one in the experimental blocks in that the “practice array” of source locations was used. Given this extensive practice before the experiment start, during the experiment participants managed to tap approximately at the tapping-positions. The experimenter kept close to participants and visually monitored that the required pace and distances of movement were kept during the whole experiment, and when necessary, corrected participants by grabbing and leading their hand to the exact tapping-position.

Next, participants completed four experimental blocks, each containing five stages (See **Figure 1C**): pre-stimulation tactile distance task (Pre-test, 36 trials), audio-tactile tapping task, post-stimulation tactile distance task (Post-test, 18 trials), audio-tactile tapping task, post-stimulation tactile distance task (Post-test, 18 trials). The experimental blocks differed in the auditory condition (Synchronous or Asynchronous) for the audio-tactile tapping task and in the arm displacement condition (Displacement or No Displacement). The Pre-test values were taken as baseline measures to which refer the Post-test values. The Post-test was split in two parts (18 trials each), with participants performing a second round of the audio-tactile tapping task in between, to ensure that the effect of the audio-tactile tapping task was not lost due to the length of the procedure of the tactile distance task. Future studies may determine how long the effects of the audio-tactile tapping task last. Participants were blindfolded throughout the experimental block and were not allowed to see the tactile stimuli at any point during the experiment.

At the end of each block, the subjective experience of participants during the audio-tactile tapping task was assessed with a questionnaire containing eight statements, adapted from our previous study (Tajadura-Jiménez et al., 2012). The list of statements is presented in the Subjective Results section. Participants rated their level of agreement with the statements using a 7-item Likert scale, ranging from  $-3$  (strongly disagree) to  $+3$  (strongly agree), with 0 referring to “neither agree, nor disagree.”

The order of presentation of the blocks was randomized, with each experimental block lasting on average for 15 min. This resulted in a total duration of the experiment of about 90 min (four experimental blocks and one practice block, plus instructions and debriefing).

## Data Analyses

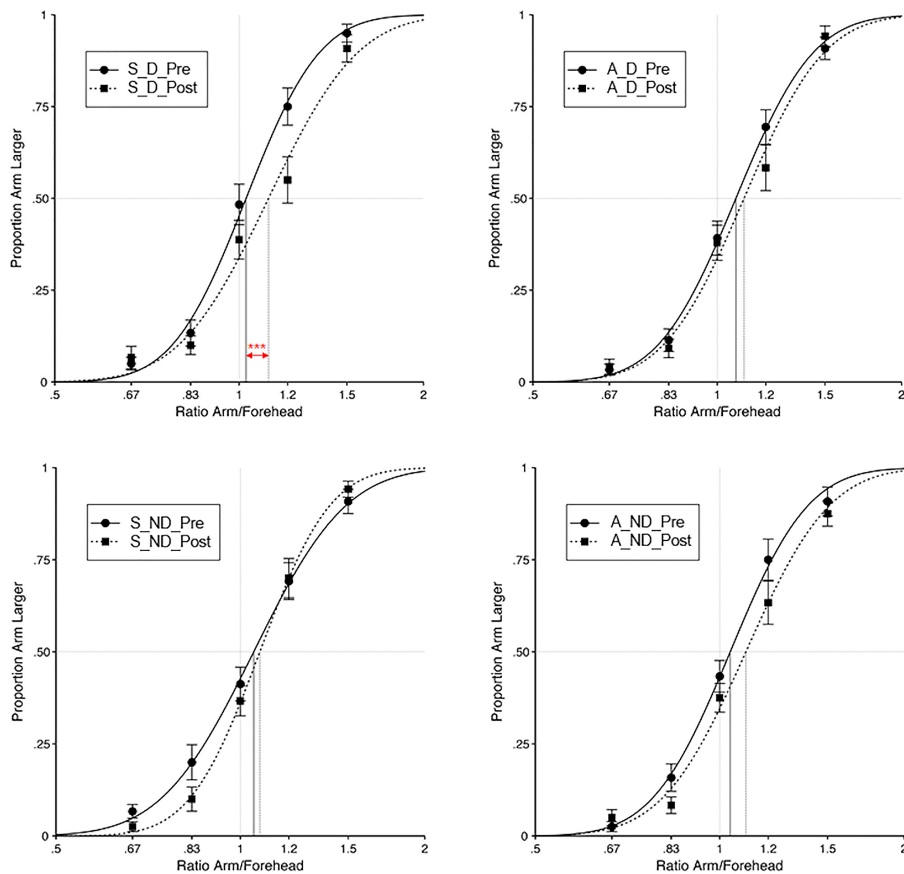
For data analyses, we followed the procedure described in Longo and Haggard (2011). For each experimental condition, and for both the Pre- and Post-test, the proportion of judgments that the distance between dual tactile stimuli on the right arm felt greater than on the forehead was analyzed as a function of the ratio of the length of the arm and forehead stimuli (i.e., 4/6, 5/6, 1, 6/5, or 6/4). The proportion of judgments was plotted logarithmically to produce a symmetrical distribution about the point of actual equality (i.e., the point at which the ratio equals 1; see **Figure 2**). Cumulative Gaussian functions were fit to each participant's data with least-squares regression using R 3.0.1. The point of subjective equality (PSE) was calculated as the point at which the fitted psychometric function crossed 50%. Thus, the PSE corresponds to the ratio of the length of the arm and forehead stimuli for which participants perceived the distance between dual stimuli on both locations to be the same. Given that tactile distance perception is directly related to tactile sensitivity (Taylor-Clarke et al., 2004; Longo and Haggard, 2011), on average a PSE greater than 1 is expected for the Pre-test, given the greater tactile sensitivity of the forehead with respect to the forearm. Given that tactile distance perception also links to the size of the represented body part (Taylor-Clarke et al., 2004; de Vignemont et al., 2005; Longo and Haggard, 2011; Tajadura-Jiménez et al., 2012; Canzoneri et al., 2013a,b), a change in PSE from Pre- to Post-test would provide evidence of the effect of the audio-tactile tapping task on the size of the represented forearm. For all statistical tests alpha level was set at 0.05, 2-tailed.

## Results

### Behavioral Results

The mean PSE values  $\pm$  SE are presented in **Table 1**. Initial analyses did not show any difference in the Pre-test PSE values across the different trial conditions ( $p > 0.05$ ), thus confirming the validity of the Pre-test values as baseline. In addition, another initial analysis was performed to investigate potential differences between the two Post-test sets of 18 trials in each condition. A  $4 \times 2 \times 5$  analysis of variance (ANOVA) with the factors ‘condition’ (Synchronous – Displacement, Asynchronous – Displacement, Synchronous – No Displacement, Asynchronous – No Displacement), ‘Post-test set’ (first, second), and ‘ratio of the length of the arm and forehead stimuli’ (4/6, 5/6, 1, 6/5, or 6/4) did not show any significant effect or interaction of the factor ‘Post-test set’ (all  $ps > 0.05$ ), thus justifying the treatment of both Post-tests as a single test.

Our main analysis focused on the effect of audio-tactile stimulation across conditions. A normality check of the residuals with Shapiro–Wilk tests and Q–Q plots showed moderate deviations from normality for three out of the eight variables (Synchronous – Displacement pre-test:  $W(20) = 0.87$ ;  $p = 0.014$ ; Asynchronous – No Displacement pre-test:  $W(20) = 0.73$ ;  $p < 0.001$ ; Asynchronous – No Displacement post-test:  $W(20) = 0.87$ ;  $p = 0.013$ ). Given that ANOVAs are quite robust to moderate deviations from normality (e.g., McDonald, 2014) we opted for the use of ANOVAs, which allow a factorial design and to explore the interaction between factors. Pre- and



**FIGURE 2 | Results from Experiment 1.** For each experimental condition (S-D, Synchronous – Displacement; S-ND, Synchronous – No Displacement; A-D, Asynchronous – Displacement; A-ND, Asynchronous – No Displacement) and for both the Pre- and Post-test, the proportion of judgements that the distance between dual tactile stimuli on the right arm felt greater than on the forehead was analyzed as a function of the ratio of the length of the arm and forehead stimuli (i.e., 4/6, 5/6, 1, 6/5, or 6/4). Curves are cumulative Gaussian

function fits to the group data, for each condition, with least-squared regression. Error bars indicate the SEM. Vertical lines indicate the interpolated points of subjective equality (PSE) between the perceived distance on the arm and on the forehead. Red asterisks denote a significant change in PSE from Pre- to Post-test (\*\*\*) denotes  $p < 0.001$ , corrected for multiple comparisons). Note that an increase in the PSE meant that perceived tactile distances on the arm were felt smaller, as compared to distances on a reference location.

**TABLE 1 | Results from Experiment 1.**

Time of test	Synchronous – Displacement	Asynchronous – Displacement	Synchronous – No Displacement	Asynchronous – No Displacement
Pre-test	1.035 (0.031)	1.085 (0.03)	1.068 (0.037)	1.067 (0.043)
Post-test	1.151 (0.044)	1.123 (0.032)	1.078 (0.026)	1.115 (0.032)

Mean point of subjective equality (PSE  $\pm$  SE) for each experimental condition and for both the Pre- and Post-test.

Post-test PSE values were submitted to a  $2 \times 2 \times 2$  within-subjects ANOVA with ‘audio-tactile synchronicity’ (Synchronous and Asynchronous), ‘arm displacement’ (Displacement and No Displacement) and ‘time of test’ (Pre-test and Post-test) as factors. The 3-way interaction ‘audio-tactile synchronicity’ by ‘arm displacement’ by ‘time of test’ was significant [ $F(1,19) = 4.51$ ;  $p = 0.047$ ], as well as the main effect of ‘time of test’ [ $F(1,19) = 18.39$ ;  $p < 0.001$ ], while the other main effects or interactions failed to reach significance (all  $ps > 0.05$ ).

In order to explore the 3-way interaction, we conducted two further  $2 \times 2$  within-subjects ANOVA, one for the Displacement and one for the No Displacement condition, with factors

‘audio-tactile synchronicity’ (Synchronous and Asynchronous) and ‘time of test’ (Pre-test and Post-test). The ANOVA for the Displacement condition revealed a significant main effect of ‘time of test’ [ $F(1,19) = 15.26$ ;  $p = 0.001$ ], as well as a significant 2-way interaction ‘audio-tactile synchronicity’ by ‘time of test’ [ $F(1,19) = 11.11$ ;  $p = 0.003$ ], while the main effect of ‘audio-tactile synchronicity’ was not significant ( $p > 0.05$ ). Independent-samples  $t$ -tests showed that the observed interaction was driven by a significant increase in the PSE from Pre- to Post-test in the Synchronous – Displacement condition [ $t(19) = -4.49$ ,  $p < 0.001$ ], which was not observed for the Asynchronous – Displacement condition ( $p > 0.05$ ). Such increase in the



PSE meant that exposure to the Synchronous – Displacement condition resulted in the perceived tactile distances on the arm being felt smaller, as compared to distances on a reference location. The ANOVA for the No Displacement condition did not yield any significant main effect or interaction (all  $p$ s > 0.05).

## Subjective Results

The full set of statements, mean responses and tests for significance are presented in **Table 2**. In order to investigate the effect of audio-tactile stimulation on the subjective experience of participants across the conditions, first, we tested whether

the distributions of the obtained data were normal using the Shapiro–Wilk test. None of the variables passed the normality test, and therefore we used non-parametrical statistical tests to analyze the data (Friedman and Wilcoxon Signed Ranks Test). We observed significant differences between the four conditions for all statements except S4 and S8.

In order to explore these significant differences and to validate our manipulations of synchronicity and kinaesthetic cues, we looked separately at the differences due to ‘audio-tactile synchronicity’ and due to ‘arm displacement’, by

**TABLE 2 | Mean ratings ( $\pm$ SE) and tests for significance for each questionnaire item across conditions in Experiment 1.**

During the audio-tactile stimulation it seemed like...	Mean ratings ( $\pm$ SE) and results of Friedman test	Synchronous vs. Asynchronous ( $\alpha = 0.025$ )	Displacement vs. No Displacement ( $\alpha = 0.025$ )	S–D vs. A–D ( $\alpha = 0.017$ )	S–D vs. S–ND ( $\alpha = 0.017$ )	S–D vs. A–ND ( $\alpha = 0.017$ )
S1: ...the sound I heard was caused by me	S–D: 1.9 (0.38) A–D: –0.5 (0.5) S–ND: 1.15 (0.41) A–ND: –1.05 (0.35) $\chi^2(3) = 31.35, p < 0.001$	<b><math>z = 3.93</math></b> <b><math>p &lt; 0.001</math></b>	<b><math>z = 2.30</math></b> <b><math>p = 0.021</math></b>	<b><math>z = 3.34</math></b> <b><math>p = 0.001</math></b>	$z = 1.38$ $p = 0.17$	<b><math>z = 3.64</math></b> <b><math>p = 0.000</math></b>
S2: ...my hand was at the same location as the sound	S–D: 1.25 (0.45) A–D: –0.15 (0.42) S–ND: –0.85 (0.45) A–ND: –1.5 (0.39) $\chi^2(3) = 10.83, p = 0.013$	$z = 2.08$ $p = 0.037$	<b><math>z = 2.88</math></b> <b><math>p = 0.004</math></b>	$z = 1.99$ $p = 0.047$	<b><math>z = 2.5</math></b> <b><math>p = 0.013</math></b>	<b><math>z = 3.09</math></b> <b><math>p = 0.002</math></b>
S3: ...my arm felt longer than usual	S–D: –0.25 (0.42) A–D: –0.9 (0.35) S–ND: –1.6 (0.29) A–ND: –1.2 (0.3) $\chi^2(3) = 11.89, p = 0.008$	$z = 0.36$ $p = 0.721$	<b><math>z = 2.35</math></b> <b><math>p = 0.019</math></b>	$z = 2.07$ $p = 0.038$	<b><math>z = 2.7</math></b> <b><math>p = 0.007</math></b>	$z = 1.94$ $p = 0.052$
S4: ...my arm felt shorter than usual	S–D: –0.85 (0.34) A–D: –0.7 (0.36) S–ND: –1.05 (0.35) A–ND: –1.0 (0.3) $\chi^2(3) = 2.02, p = 0.568$	$z = 0.45$ $p = 0.651$	$z = 1.12$ $p = 0.265$	$z = 0.5$ $p = 0.62$	$z = 0.84$ $p = 0.4$	$z = 0.81$ $p = 0.42$
S5: ...my own arm was out of my control	S–D: –1.5 (0.38) A–D: –0.4 (0.44) S–ND: –1.6 (0.26) A–ND: –0.25 (0.38) $\chi^2(3) = 10.07, p = 0.018$	<b><math>z = 2.55</math></b> <b><math>p = 0.011</math></b>	$z = 0.17$ $p = 0.862$	$z = 1.94$ $p = 0.052$	$z = 0.32$ $p = 0.75$	$z = 2.37$ $p = 0.018$
S6: ...I couldn't remember how long my arm was	S–D: –0.15 (0.39) A–D: –0.45 (0.44) S–ND: –1.45 (0.37) A–ND: –0.9 (0.31) $\chi^2(3) = 12.29, p = 0.006$	$z = 0.36$ $p = 0.722$	<b><math>z = 2.76</math></b> <b><math>p = 0.006</math></b>	$z = 0.89$ $p = 0.37$	<b><math>z = 2.97</math></b> <b><math>p = 0.003</math></b>	$z = 1.94$ $p = 0.052$
S7: ...I couldn't really tell where my hand was	S–D: 1 (0.38) A–D: 2 (0.42) S–ND: –1.4 (0.36) A–ND: –1.0 (0.4) $\chi^2(3) = 16.83, p = 0.001$	$z = 0.73$ $p = 0.467$	<b><math>z = 3.35</math></b> <b><math>p = 0.001</math></b>	$z = 0.32$ $p = 0.75$	<b><math>z = 2.09</math></b> <b><math>p = 0.002</math></b>	$z = 2.1$ $p = 0.036$
S8: ...the experience of my arm was less vivid than normal	S–D: –0.25 (0.38) A–D: 0.15 (0.42) S–ND: –0.2 (0.42) A–ND: 0.25 (0.4) $\chi^2(3) = 2.46, p = 0.482$	$z = 1.01$ $p = 0.315$	$z = 0.31$ $p = 0.759$	$z = 0.92$ $p = 0.36$	$z = 0.11$ $p = 0.91$	$z = 0.89$ $p = 0.37$

Participants rated their level of agreement with the statements using a 7-item Likert scale (i.e., –3 to +3). Results from Friedman tests comparing all conditions are presented in the second column. In addition, in the other columns planned pairwise comparisons with correction for multiple comparisons ( $\alpha$  is indicated in the column header) are presented. Significant differences between conditions are marked in **bold font**. Differences between conditions indicate changes as a result of the auditory manipulation. S–D, Synchronous – Displacement; S–ND, Synchronous – No Displacement; A–D, Asynchronous – Displacement; A–ND, Asynchronous – No Displacement.

running Wilcoxon Signed Ranks Tests (with correction for multiple comparisons  $\alpha = 0.025$ ). First, we compared the average of the two Synchronous conditions and of the two Asynchronous conditions. We found that while participants in the Synchronous audio-tactile conditions felt that the sound was caused by them (S1), they did not feel the same for the Asynchronous conditions, thus providing evidence that our manipulation of synchronicity had the expected effect on agency. In addition, participants significantly disagreed more when enquired about the loss of control over their arm (S5) during the Synchronous than the Asynchronous conditions. This less experienced control over the sounds and over hand movement during the Asynchronous conditions provides evidence that our manipulation of synchronicity had the expected effect on agency.

Second, we compared the average of the two Displacement conditions and of the two No Displacement conditions. We found a significant difference between the Displacement and No Displacement conditions in the felt sensation that the sound came from the same location where the hand was (S2). This provides evidence that our manipulation of kinaesthetic cues derived from arm Displacement had the expected effect on the feelings that sound and tap originate at the same location. Importantly, the felt sensation that the sound comes from the same location where the hand is has been previously identified as being fundamental to the auditory-induced changes in perceived tactile distance (Tajadura-Jiménez et al., 2012). We also found that in the Displacement conditions people felt more as the agents of the sounds (S1). Importantly, we found that in the Displacement conditions people reported not being able to tell where one's hand was (S7), as well as less disagreement with the statements "my arm felt longer than usual" (S3) and "I couldn't remember how long my arm was" (S6).

Furthermore, having identified the Synchronous – Displacement condition as the critical condition for which changes in the experience of one's arm as a result of the auditory manipulation were observed, *post hoc* analyses compared the mean responses to each statement of the questionnaire for Synchronous – Displacement to the responses given after exposure to the other three conditions (with correction for multiple comparisons  $\alpha = 0.017$ ; see **Table 2**). These analyses further confirmed a difference between our critical Synchronous – Displacement condition and the Asynchronous conditions in the felt sensation of being the agent of the sounds (S1), and between the Synchronous – Displacement condition and the No Displacement conditions, in the felt sensation that the sound came from the same location where the hand was (S2). Importantly, the Synchronous – Displacement condition significantly differed from the Synchronous – No Displacement condition in the sensation that one's arm felt longer than usual (S3), that one couldn't remember how long one's arm was (S6) and that one couldn't really tell where one's hand was (S7). This provides evidence that the combination of synchronicity, which resulted in the subjective experience of being the agent of the tapping sounds, and arm displacement, which involved additional kinaesthetic cues signaling a change in hand position, resulted in subjective

changes in the perceived length of the arm and in the perceived location of the hand.

### Summary Experiment 1

These results demonstrate that hearing the tapping sounds with double auditory distance under certain conditions results in a significant change in participants' perceived tactile distance on the test arm and in the subjective feelings of arm length. First of all, synchrony between the sounds and the actual taps is critical for this change to occur, because it preserves the subjective experience of being the agent of the sounds, as previously indicated in Tajadura-Jiménez et al. (2012). In addition, kinaesthetic cues signaling arm displacement when displacement of the sound source occurs are necessary in order to observe audio-tactile adaptation, as changes were observed for the Synchronous – Displacement but not for the Synchronous – No Displacement condition, while in both conditions the feelings of producing the sound were preserved.

The results on perceived tactile distance were further confirmed by the subjective reports, which show, for the first time, a significant effect on the subjective experience of arm length (S3). Importantly, we observed significant differences between the Synchronous – Displacement and the Synchronous – No Displacement conditions in the felt sensations that the sound came from the same location where the hand was (S2), and that one could not really tell where one's hand (S7) was. It exists the possibility that these differences derive from an effect of the posture adopted by participants in the No Displacement conditions. In particular, we identified two differences between Displacement and No Displacement conditions due to posture, which did not allow us to conclude that the difference in results between these conditions was only due to the presence/absence of kinaesthetic cues signaling arm displacement. First, the distance between sound source and hand was larger in the No Displacement conditions (it increased from 25 to 125 cm, as the hand is kept at 25 cm but the sound source moves from a position 50 cm away to a position 150 cm away) than in the Displacement conditions (where it increased from 25 to 75 cm, as the hand moves from a position 25 cm away to a position 75 cm away, and the sound source moves from a position 50 cm away to a position 150 cm away). Could this smaller hand-sound source distance in the Displacement conditions have accounted for the difference in felt sensation that the sound came from the same location where the hand was (S2)? Second, the distance between hand and body torso differed between conditions. While in the No Displacement conditions the hand was kept 25 cm away, in the Displacement condition the hand could be as far as 75 cm away. Could this larger hand-body torso distance in the Displacement conditions have accounted for the difference in felt sensation that one could not really tell where one's hand was (S7)?

Given these findings and the discussed possible confounds, a second experiment was run. Experiment 2 served to control for the possible confounding variables by having a modified version of the No Displacement conditions in which participants kept their arm stretched and their hand placed at the last tapping-position, thus far away from the participants' torso. With

this modified version of the No Displacement conditions, we made sure, first, that in both the Displacement and the No Displacement conditions the initial distance between the sound source and hand location was 25 cm, and that the maximum distance was 75 cm. Second, we made sure that the hand location in the No Displacement condition equalled the maximum hand-body torso distance in the Displacement condition, this is, 75 cm.

Hence, Experiment 2 was run with the hypothesis that if different results were found between the Displacement and No Displacement conditions, we could conclude that they were due to the presence of kinaesthetic cues signaling arm displacement, and not due to differences in the maximum hand-sound source, and hand-body torso distances.

## Experiment 2

### Materials and Methods

#### Participants

Seventeen participants took part in the experiment. We applied the same participant selection criteria as in Experiment 1, and the experiment was conducted in accordance with the same ethical standards. Three participants were removed from the analyses, given that two of them were unable to complete the audio-tactile “tapping” task as required, due to difficulties in remembering the instructions, and one was unable to complete the “tactile distance” task, due to lack of tactile sensitivity in the arm. Therefore, only results from fourteen participants are reported here ( $M_{\text{age}} \pm SD = 23.64 \pm 3.6$  years; age range from 18 to 30 years; seven females).

#### Apparatus, Materials, Procedure, and Data Analyses

Identical apparatus, materials, and data analyses to the ones in Experiment 1 were used. The procedure used was also identical to the one in Experiment 1, except that in this case, during the audio-tactile “tapping” task in the No Displacement trials, participants were required to keep their right arm stretched and tap always at the same, sixth tapping-position.

### Results

#### Behavioral Results

The mean PSE values  $\pm$  SE are presented in **Table 3**. Identical analyses to those in Experiment 1 were conducted and the behavioral results mirrored those in Experiment 1. After validating the Pre-test values as baseline, and the treatment of both post-tests as a single test, our main analysis focused, as before, on the effect of audio-tactile stimulation across conditions.

A normality check of the residuals with Shapiro–Wilk tests and Q–Q plots showed moderate deviations from normality for two out of the eight variables (Synchronous – No Displacement pre-test:  $W(14) = 0.83$ ;  $p = 0.015$ ; Synchronous – No Displacement post-test:  $W(14) = 0.83$ ;  $p = 0.015$ ), and hence we opted for the use of ANOVAs. As in Experiment 1, the 3-way interaction ‘audio-tactile synchronicity’ by ‘arm displacement’ by ‘time of test’ was significant [ $F(1,13) = 7.04$ ;  $p = 0.02$ ], as well as the main effect of ‘time of test’

[ $F(1,13) = 14.01$ ;  $p = 0.002$ ], while the other main effects or interactions failed to reach significance (all  $ps > 0.05$ ; see **Figure 3**).

In order to explore the 3-way interaction, we conducted two further  $2 \times 2$  within-subjects ANOVA, one for the Displacement and one for the No Displacement condition, with factors ‘audio-tactile synchronicity’ (Synchronous and Asynchronous) and ‘time of test’ (Pre-test and Post-test). The ANOVA for the Displacement condition revealed a significant main effect of ‘time of test’ [ $F(1,13) = 14.28$ ;  $p = 0.002$ ], as well as a significant 2-way interaction ‘audio-tactile synchronicity’ by ‘time of test’ [ $F(1,13) = 9.47$ ;  $p = 0.009$ ], while the main effect of ‘audio-tactile synchronicity’ was not significant ( $p > 0.05$ ). Independent-samples  $t$ -tests showed that the observed interaction was driven by a significant increase in the PSE (i.e., perceived tactile distances on the arm felt smaller) from Pre- to Post-test in the Synchronous – Displacement condition [ $t(13) = -4.40$ ,  $p < 0.001$ ], which was not observed for the Asynchronous – Displacement condition ( $p > 0.05$ ). The ANOVA for the No Displacement condition yielded a significant main effect of ‘time of test’ [ $F(1,13) = 4.65$ ;  $p = 0.05$ ], while the main effect of ‘audio-tactile synchronicity’ or its interaction with ‘time of test’ were not significant (all  $ps > 0.05$ ).

#### Subjective Results

The full set of statements, mean responses and test for significance are presented in **Table 4**. Identical analyses to those in Experiment 1 were conducted and the subjective results mostly mirrored those in Experiment 1. We observed significant differences between the four conditions for statements S1, S2, S3, and S5.

When comparing the average of the two Synchronous conditions and of the two Asynchronous conditions, apart from the effects reported in Experiment 1 on feelings of being the agent of the sound (S1) and of one’s arm being out of one’s control (S5), we also found that in the Synchronous conditions people felt more that the sound came from the same location where the hand was (S2), and disagreed less with the statement “*my arm felt longer than usual*” (S3). When comparing the average of the two Displacement conditions and of the two No Displacement conditions, we found a similar effect as that reported in Experiment 1 on feelings that the sound came from the same location where one’s hand was (S2). Finally, when comparing the Synchronous – Displacement responses to those responses given after exposure to the other three conditions we found similar effect as those reported in Experiment 1 on feelings of being the agent of the sound (S1) and that the sound came from the same location where the hand was (S2). We also observed a close to significant larger loss of control of one’s own arm (S5) in the Asynchronous conditions as compared to the Synchronous – Displacement condition.

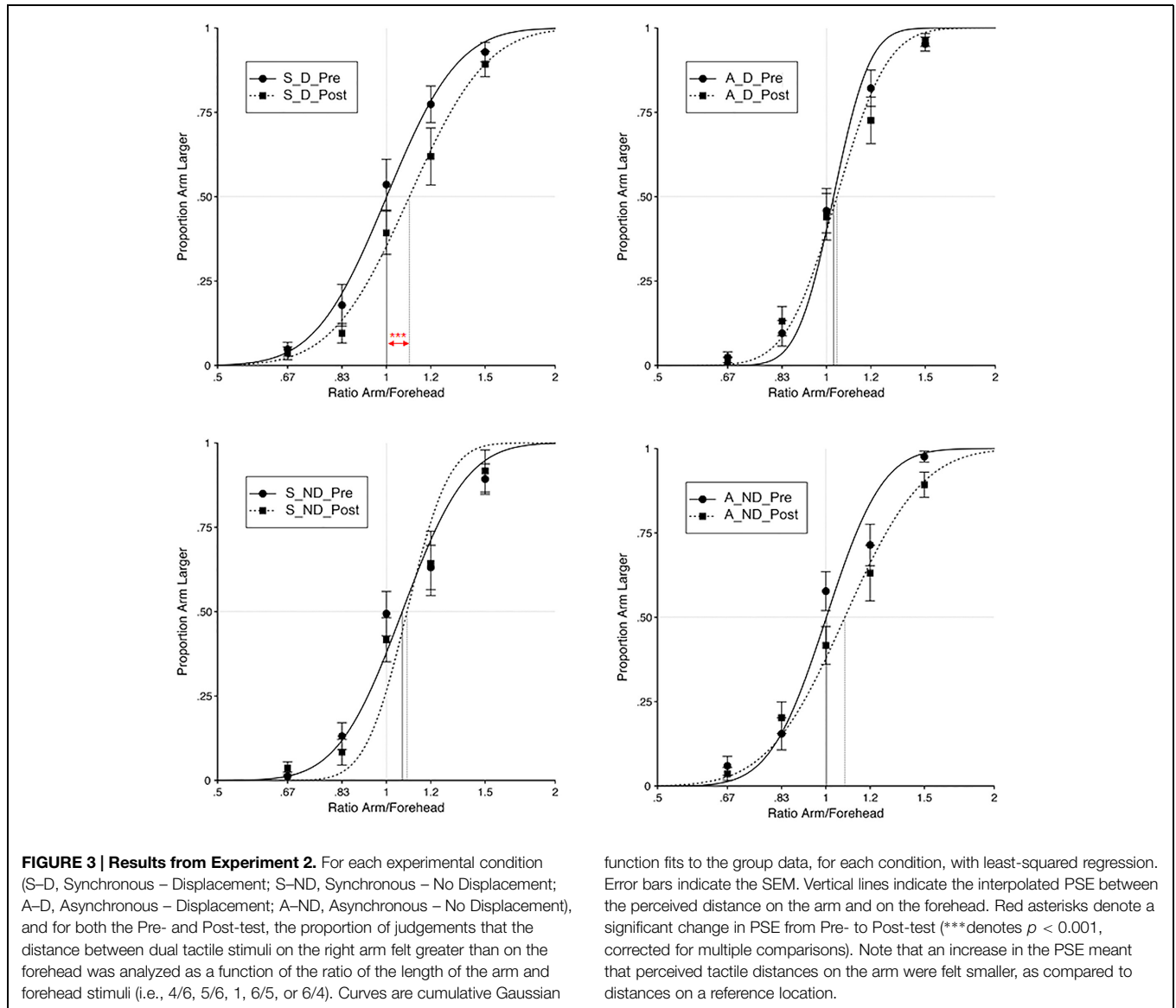
#### Direction of Changes in Represented Arm Length

We investigated how the observed tactile distance changes in the Synchronous – Displacement condition related to

**TABLE 3 | Results from Experiment 2.**

Time of test	Synchronous – Displacement	Asynchronous – Displacement	Synchronous – No Displacement	Asynchronous – No Displacement
Pre-test	1.008 (0.032)	1.033 (0.025)	1.079 (0.046)	1.009 (0.032)
Post-test	1.127 (0.044)	1.052 (0.036)	1.101 (0.05)	1.093 (0.045)

Mean PSE  $\pm$  SE for each experimental condition and for both the Pre- and Post-test.



participants' subjective experience of changes in arm length. Given that the Synchronous – Displacement condition was identical in Experiments 1 and 2, we pooled the results from the total 34 participants in both experiments and performed correlation analyses between behavioral and subjective data. In particular, we looked at Spearman's rho correlations between the change from Pre- to Post-test in PSE in the tactile distance task and the self-reported level of agreement for all statements (S1–S8) in the Synchronous – Displacement condition. Results showed that changes in PSE

correlated significantly with changes in level of agreement with the statement S3 “my arm felt longer than usual” [ $r_s(34) = 0.41$ ,  $p = 0.015$ ] and with the statement S7 “I couldn't really tell where my hand was” [ $r_s(34) = 0.36$ ,  $p = 0.038$ ], while correlations with data from the other statements were all not significant. In particular, linear regression analyses revealed that positive changes in PSE predicted increased feelings of one's arm being longer than usual and of not being able to tell where one's hand was (see **Figure 4**).



**TABLE 4 | Mean ratings ( $\pm$ SE) and tests for significance for each questionnaire item across conditions in Experiment 2.**

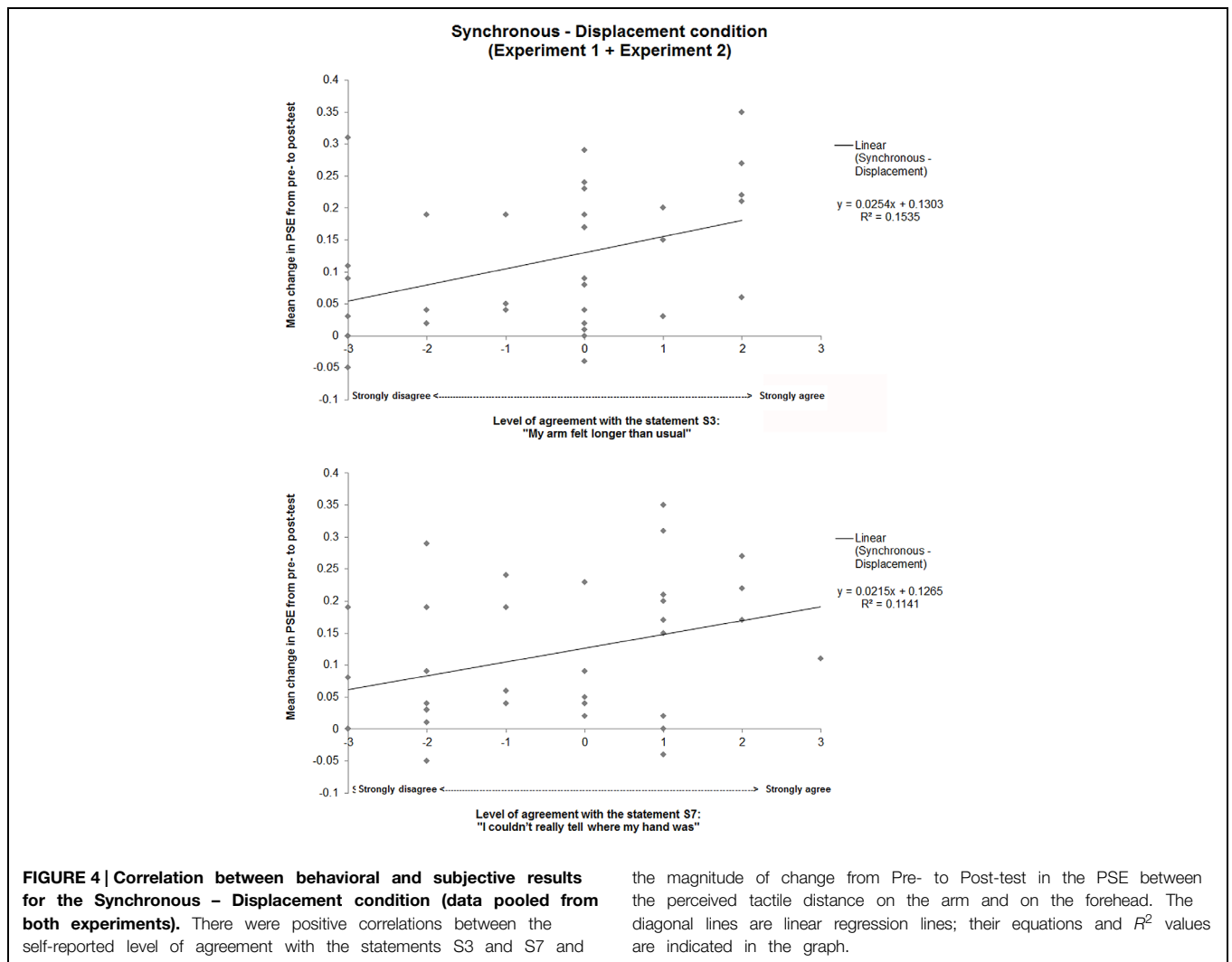
During the audio-tactile stimulation it seemed like...	Mean ratings ( $\pm$ SE) and results of Friedman test	Synchronous vs. Asynchronous ( $\alpha = 0.025$ )	Displacement vs. No Displacement ( $\alpha = 0.025$ )	S-D vs. A-D ( $\alpha = 0.017$ )	S-D vs. S-ND ( $\alpha = 0.017$ )	S-D vs. A-ND ( $\alpha = 0.017$ )
S1: ...the sound I heard was caused by me	S-D: 2.07 (0.37) A-D: 0.07 (0.67) S-ND: 2.21 (0.35) A-ND: -0.71 (0.59) $\chi^2(3) = 19.8, p < 0.001$	<b><math>z = 3.08</math></b> <b><math>p = 0.002</math></b>	$z = 1.44$ $p = 0.149$	<b><math>z = 2.46</math></b> <b><math>p = 0.014</math></b>	$z = 0.26$ $p = 0.79$	<b><math>z = 3.00</math></b> <b><math>p = 0.003</math></b>
S2: ...my hand was at the same location as the sound	S-D: 1.71 (0.41) A-D: 0.5 (0.51) S-ND: -0.29 (0.61) A-ND: -1.36 (0.44) $\chi^2(3) = 12.32, p = 0.006$	<b><math>z = 2.56</math></b> <b><math>p = 0.01</math></b>	$z = 2.77$ <b><math>p = 0.006</math></b>	$z = 1.7$ $p = 0.089$	$z = 2.25$ $p = 0.025$	<b><math>z = 2.96</math></b> <b><math>p = 0.003</math></b>
S3: ...my arm felt longer than usual	S-D: -0.86 (0.36) A-D: -1.29 (0.34) S-ND: 0.07 (0.42) A-ND: -0.93 (0.41) $\chi^2(3) = 10.84, p = 0.013$	<b><math>z = 2.39</math></b> <b><math>p = 0.017</math></b>	$z = 1.84$ $p = 0.065$	$z = 1.51$ $p = 0.13$	$z = 1.98$ $p = 0.048$	$z = 0.69$ $p = 0.49$
S4: ...my arm felt shorter than usual	S-D: -0.71 (0.38) A-D: -1.0 (0.38) S-ND: -1.36 (0.36) A-ND: -0.79 (0.38) $\chi^2(3) = 6.2, p = 0.102$	$z = 0.79$ $p = 0.429$	$z = 0.88$ $p = 0.38$	$z = 1.19$ $p = 0.23$	$z = 2.04$ $p = 0.041$	$z = 0.32$ $p = 0.75$
S5: ...my own arm was out of my control	S-D: -1.79 (0.39) A-D: -0.36 (0.44) S-ND: -1.29 (0.40) A-ND: -0.71 (0.50) $\chi^2(3) = 11.75, p = 0.008$	$z = 2.21$ $p = 0.027$	$z = 0.36$ $p = 0.720$	$z = 2.39$ $p = 0.017$	$z = 1.22$ $p = 0.22$	$z = 1.98$ $p = 0.048$
S6: ...I couldn't remember how long my arm was	S-D: -1.07 (0.43) A-D: -0.93 (0.43) S-ND: -0.36 (0.37) A-ND: -0.93 (0.46) $\chi^2(3) = 2.77, p = 0.428$	$z = 0.78$ $p = 0.436$	$z = 1.45$ $p = 0.146$	$z = 0.71$ $p = 0.48$	$z = 1.91$ $p = 0.056$	$z = 0.00$ $p = 1.00$
S7: ...I couldn't really tell where my hand was	S-D: -1.14 (0.40) A-D: -0.21 (0.42) S-ND: -0.64 (0.32) A-ND: -1.0 (0.44) $\chi^2(3) = 4.18, p = 0.243$	$z = 1.09$ $p = 0.278$	$z = 0.57$ $p = 0.57$	$z = 1.80$ $p = 0.072$	$z = 0.90$ $p = 0.365$	$z = 0.00$ $p = 1.00$
S8: ...the experience of my arm was less vivid than normal	S-D: -0.86 (0.40) A-D: -0.21 (0.41) S-ND: -0.43 (0.23) A-ND: -0.14 (0.49) $\chi^2(3) = 3.41, p = 0.332$	$z = 1.5$ $p = 0.134$	$z = 1.33$ $p = 0.185$	$z = 1.26$ $p = 0.21$	$z = 0.84$ $p = 0.40$	$z = 2.16$ $p = 0.031$

Participants rated their level of agreement with the statements using a 7-item Likert scale (i.e., -3 to +3). Results from Friedman tests comparing all conditions are presented in the second column. In addition, in the other columns planned pairwise comparisons with correction for multiple comparisons ( $\alpha$  is indicated in the column header) are presented. Significant differences between conditions are marked in **bold** font. Differences between conditions indicate changes as a result of the auditory manipulation. S-D, Synchronous - Displacement; S-ND, Synchronous - No Displacement; A-D, Asynchronous - Displacement; A-ND, Asynchronous - No Displacement.

## Summary Experiments 1 and 2

In Experiment 2 we controlled for the effect of arm posture adopted in the conditions lacking kinaesthetic cues signaling arm displacement (i.e., No Displacement conditions). In particular, we made sure that the lack of results for the No Displacement conditions in Experiment 1 was independent of the distance between body torso and hand (hand close to the body torso in Experiment 1 and far from the body torso in Experiment 2), and of the distance between hand and sound source, which was larger for the posture adopted in Experiment 1 than for the posture in Experiment 2.

The obtained results support the finding that exposure to the tapping sounds with double auditory distance significantly changes participants' perceived tactile distance on the test arm, in comparison to the reference location. Importantly, Experiment 1 and Experiment 2, together demonstrate that both synchrony between the tapping sounds and the actual taps of participants, and the update in kinaesthetic cues during the displacement of one's arm while tapping, are critical conditions for this change to occur. In other words, changes in perceived tactile distance on the arm do not occur in the absence of kinaesthetic cues signaling arm displacement (i.e., when the arm remained tapping at a fixed



location), even when the feelings of one being the agent of the sounds are preserved. In addition, results from both experiments suggest that changes in perceived tactile distance on the arm correlate with feelings of one's arm changing length.

## Discussion

Taken together our results elucidate necessary factors for auditory-induced recalibration of perceived tactile distances to occur. Extending previous results (Tajadura-Jiménez et al., 2012), we show that the manipulation of the auditory distance of the triggered tapping sounds can change the perceived tactile distance on the arm used for tapping. Importantly, we here show that the involvement of kinaesthetic cues signaling arm displacement during tapping when displacement of the sound sources occurs is a necessary condition for these sound sources to induce changes in the perceived tactile length of external objects. In particular, the feeling of being the agent of the tapping sounds is not sufficient to induce changes in the perceived tactile distance, but a coherent representation of the motor

command sent to the displacing and tapping arm and the sound feedback received from the tapping action needs to arise in order to observe such changes. Furthermore, we provide the first evidence that self-produced sounds can evoke consciously perceived changes in body-representation, specifically in the represented arm length, and that these changes correlate with changes in the perceived tactile length of external objects in contact with one's arm. In the following sections, we discuss the implications of the observed effects and the limitations of the study.

### Is the Feeling of Being the Agent of the Sound Sufficient to Change the Perceived Tactile Distance?

Our study shows that the feeling of being an agent of the sound, achieved by keeping temporal contingency between the action and its attributed sound, alone is not sufficient to observe auditory-induced changes in perceived tactile distance. Instead, the involvement of kinaesthetic cues signaling arm displacement when displacement of the sound sources occurs is also necessary

for these changes to occur, provided that the distance at which sounds originate is within certain limits from the tapping hand (i.e., we observed changes after exposure to tapping sounds originating at double but not at quadruple the distance at which participants actually tapped; Tajadura-Jiménez et al., 2012).

It should be noted that since people made voluntary movements when tapping throughout all experimental conditions, they did retain a basic sense of agency in so far they themselves were moving. However, our manipulation of temporal contingency impacted on the sense of agency, as it influenced the experience of being the agent of the sounds and of being in control over hand movement, as evidenced by the subjective results. It should also be noted that, while an overall increase in PSE from Pre- to Post-test occurred in all conditions, such systematic baseline shifts after adaptation are often reported in multisensory adaptation paradigms. For instance, exposure to fixed audiovisual time lags for several minutes results in shifts in subjective simultaneity responses in the direction of the exposure lag, indicating a perceptual temporal recalibration of multisensory perception (Fujisaki et al., 2004; Vroomen et al., 2004). While we cannot fully clarify here whether the baseline change observed in our experiments derives from some sort of perceptual temporal recalibration of multisensory perception or other processes, what is critical in our results is that this change in PSE from Pre- to Post-test significantly interacted with synchronicity (i.e., with the presence/absence of feelings of agency), in the Displacement conditions.

We suggest that these findings can be interpreted in the context of the proposed 'forward internal models' of motor-to-sensory transformations (Wolpert and Ghahramani, 2000). These models serve to predict the movement dynamics and the sensory outputs that derive from one's actions (i.e., reafference). Hence, when we move an arm, the central nervous system estimates the next state (e.g., the next position of the hand) by combining the current efferent motor outflow (the motor commands sent to the arm) with the predictions of arm's dynamics for the current state. The central nervous system also estimates the sensory reafference that will accompany the next state by combining the current reafferent multisensory inflow with the sensory predictions for the current state. The discrepancies between prediction and reafference are used to do adjustments in next state estimates (Wolpert et al., 1995), as well as to do fine adjustments in the subsequent motor commands (Blakemore et al., 2002). Studies introducing temporal and spatial discrepancies between movement and its visual consequences have shown that only discrepancies between prediction and reafference exceeding a certain threshold become available to awareness (Blakemore et al., 2002). Trespassing this threshold can result in delusions of control over produced actions, although the exact threshold for these discrepancies to reach awareness is debated. Indeed, this threshold can be relatively large, as long as our intentions are successfully achieved (Blakemore et al., 2002).

In our study, which involves motor-to-sensory transformations when moving an arm, we observed changes in perceived arm length for the Synchronous – Displacement condition, but not for the other conditions. We suggest that this condition provides a better temporal and spatial match between

reafference and sensory predictions than the other conditions. In the Synchronous – Displacement condition, as opposed to the Asynchronous conditions, there is a temporal agreement between the action and its attributed sound, which results in the feeling of being the agent of the sound (e.g., Moore et al., 2009). Nevertheless, we show that the conscious experience of agency alone is not enough to evoke changes in represented arm length, because these changes were not observed in the Synchronous – No Displacement condition. The Synchronous – Displacement condition, in addition to temporal synchrony, provides a better spatial match between reafference and sensory predictions made on the basis of the efferent motor outflow: the kinaesthetic cues in the motor outflow indicate a change in location of the hand and, similarly, the reafferent sensory input indicates a change in location of the sound source in the same direction. It is important to note that in the Synchronous – Displacement condition the feeling that "*the sound comes from the same location where the hand is*" (see S2) is preserved, even if the tapping sounds originated at double the distance at which participants actually tapped. This temporal and spatial mismatch reduction during the action-perception loop allows forming an association between action and sound (Ernst and Bühlhoff, 2004). It should also be noted that our design indirectly includes also the testing of a condition where neither kinaesthetic cues signal a change in location of the hand nor the reafferent sensory input indicates a change in location of the sound source. The last ten taps of the Synchronous – No Displacement condition in Experiment 2 correspond to a situation in which participants do not displace their arm and sound sources are at double the distance to the tapping location. While this exposure is short (~10 s), work on other sensory-driven bodily illusions have shown that such short periods may be enough to elicit the illusions (e.g., Ehrsson et al., 2004). However, we did not observe any significant behavioral or subjective changes for this condition. These results seem to suggest that kinaesthetic cues signaling arm displacement are needed for recalibrating arm length in this context, at least for short-term exposures. The testing of long-term exposure remains beyond the scope of this study, but it is nevertheless a topic interesting for further research.

Taken in this context of 'forward internal models,' our results add to the theories on these models. These theories have mainly considered that the reafferent sensory inflow used by forward models is constituted by visual and proprioceptive information (Wolpert et al., 1995). Here we propose, not only that action sounds also constitute part of this reafferent inflow, as suggested by recent neuroimaging studies demonstrating the link between action sounds and brain areas involved in the planning, preparation, and observation of actions involved in the production of those sounds (for a review see Aglioti and Pazzaglia, 2010), but also that predictions related to action sounds must fit with kinesthetic cues related to the performed actions in order to make use of the auditory inputs to update the model.

Furthermore, our study sheds light into the magnitude of the threshold for which the model can compensate for auditory-motor spatial discrepancies. We showed that action sounds may

be attributed to the outputs of the actions performed by one's hand even when the sounds originate at double the distance at which the hand is, provided that the feelings of agency and kinaesthetic cues signaling arm displacement are preserved. Previously we also showed that when the sound originates at quadruple the distance auditory-motor spatial discrepancies become too large (Tajadura-Jiménez et al., 2012). This threshold is similar to the one reported in a related study in the visual domain, in which the illusion of owning a very long arm, seen from a first-person perspective, starts breaking when the length of this arm exceeds three times the actual length of the participants' arm (Kilteni et al., 2012). While future work should further clarify the exact threshold for temporal and spatial discrepancies to disrupt these illusions, we hypothesize that this threshold may relate to sources falling inside the represented near space.

Finally, theories of 'forward internal models' have mainly discussed how these models are continuously updated by sensorimotor information in order to estimate, for instance, the position and velocity of a hand moving (Wolpert et al., 1995), as well as to do fine adjustments in the subsequent motor commands (Blakemore et al., 2002). Our study provides more insight into the updating of these models, by showing that the auditory feedback on one's hand actions is not only used to update the estimated position of the hand, but also to update the represented arm length that allows the hand to be in that position. We suggest that when engaged in limb actions, in order to estimate the current position of the hand, predictions must integrate, apart from multimodal information extracted from previous sensorimotor feedback, internal knowledge about the configuration and length of the limbs (i.e., mental representation of one's limbs). Hence, when moving the limb the sensory feedback is weighted against the predictions and, if potential discrepancies arise but are kept below a certain threshold, these discrepancies are used to do fine adjustments in the mental representation of one's body. In support of our suggestion, previous studies have shown that the representation of an action engages a mental representation of the general body structure that allows this action to be produced (Holmes and Spence, 2004; Maravita and Iriki, 2004). Indeed, studies on audio-motor mirroring of action sounds have shown that action representation engages both agency and mental representations of the body part involved in the action (see Aglioti and Pazzaglia, 2010). Because our body configuration can change, the internal body-representation is plastic to adapt to changing circumstances by tuning to the incoming sensory feedback. Importantly, as mentioned above, we showed that predictions related to action sounds must fit with kinesthetic cues related to the performed actions in order to make use of the auditory inputs to change body-representation.

### Do Action Sounds Change the Subjective Experience of Arm Length?

We here show for the first time that action sounds can indeed change the subjective experience of arm length. Our results demonstrate that the level of agreement with statement S3 "my arm felt longer than usual" significantly changed

across conditions. Moreover, for the critical Synchronous – Displacement condition, we observed that those participants showing larger levels of agreement with statement S3 showed larger audio-tactile driven increases in PSE for the tactile distance judgment task, thus suggesting that changes in perceived tactile distance relate to experienced arm elongation.

It should be noted that in our previous study, the observed behavioral changes in represented arm length were not accompanied by significant changes in the subjective experience of arm length, thus providing evidence that changes in body-representation can occur outside of awareness (Holmes and Spence, 2004; Maravita and Iriki, 2004). We argue that the listening experience provided by the headphone-based setup may have been a factor favoring that changes in represented arm length reached awareness. Note that in the current study we simulated the array of auditory spatial positions, which allowed using headphones to present the tapping sounds, instead of loudspeakers, which were used in our previous study (Tajadura-Jiménez et al., 2012). Although we did not directly measure immersion, headphone-based listening has previously been shown to provide more intense, immersive experiences, as compared to loudspeaker-based listening (Kallinen and Ravaja, 2007; Tajadura-Jiménez et al., 2008, 2011). Moreover, in the current setup participants were blindfolded. Therefore, visual cues, as well as auditory cues other than the experimental stimuli, were reduced. These differences might have favored immersion on the listening experience and positively impacted on the subjective experience of arm length.

### Direction of Changes in Perceived Tactile Distance

Recently, there has been some controversy on the direction of changes resulting from the tactile-distance task. Studies on the effects on perceived tactile distance of body-related inputs from sensory modalities other than sound have shown that an increase in the represented part of the body relates either to an increase (e.g., Taylor-Clarke et al., 2004; de Vignemont et al., 2005; Lopez et al., 2012) or to a decrease (Canzoneri et al., 2013a,b) in perceived tactile distance on that part of the body. Our present results add to this controversy. Here we demonstrate that exposure to manipulated auditory body-related inputs results in the perceived tactile distances on the arm being felt smaller, as compared to distances on a reference location. However, in our previous study (Tajadura-Jiménez et al., 2012) exposure to similar inputs resulted on tactile distances felt bigger.

It should be noted that the fact that tactile distances were felt smaller on the arm is not necessarily in contrast with an increase in the represented length of the arm. Indeed, we showed that larger feelings of arm elongation correlated with smaller felt tactile distances. In the studies by Canzoneri et al. (2013a,b), other additional behavioral measures supported the interpretation that the decrease in perceived tactile distance in the arm following tool-use results from an increase in the represented length of the arm. These authors related their findings to other studies showing that the larger one's body (or body part) is perceived, the smaller objects external to one's body are perceived (Linkenauger et al., 2011; van der Hoort et al., 2011). It has



been suggested that one's body is used as a "perceptual ruler" to measure object's size (Linkenauger et al., 2011; Canzoneri et al., 2013b). This controversy on the direction of changes resulting from the tactile-distance task has been recently discussed in a publication by Miller et al. (2014). They nicely summarize the two opposing views in previous studies, in favor of either an inverse or a proportional relationship between represented body size and perceived tactile size, and they suggest that none of the views is correct or incorrect, but rather that one needs to take into account possible factors that might influence how tactile information is used when providing the tactile distance judgements.

Tactile distance judgements for stimuli delivered on the arm are both dependent upon the mental representation of arm length, as well as upon the geometry of receptive fields (RFs) in primary somatosensory cortex (SI; Longo and Haggard, 2011). Miller et al. (2014) discuss that visual bodily feedback can result in an update in the stored visual body template and cause reorganization of SI RF geometry (Haggard et al., 2007). They suggest that top-down sensory signals can cause this reorganization of SI RF geometry leading to changes in tactile size perception. Similarly, we previously suggested that tactile perception is referenced to an implicit body-representation which is updated through auditory feedback, presumably by auditory-induced recalibration of SI RFs (Tajadura-Jiménez et al., 2012). However, both Miller et al. (2014) and us suggest that, in addition to reorganization of SI RF geometry, there might be other top-down factors (e.g., contextual/task demands) that might influence the direction of the tactile distance judgments: tactile information is used differently in distinct, but related tactile tasks, such as tactile distance perception and tactile localization, which are both affected by sensory information on body size, presumably following reorganization of SI RF geometry.

We suggest that task differences between our two studies may explain why this opposite direction of the results is observed. We introduced differences in the task in order to use a more sound methodology by addressing some potential biases affecting previous results. A first difference between the two studies is the body location used as reference: while previously we used as reference location the left arm, here we used the forehead. Previous studies have shown asymmetries in perceived arm length (i.e., participants may perceive their right arm to be longer than their left arm), which correlate with factors such as participants' handedness and hand strength (Linkenauger et al., 2011). It might be that these asymmetries are also affected by the experimental task, and in order to control for this, we chose as reference location the forehead, a location that has been previously used in a number of studies assessing changes in perceived body size (de Vignemont et al., 2005; Lopez et al., 2012; as well as the studies by Canzoneri et al., 2013a,b). Second, while previously the task for participants was to indicate whether the distance on the left or on the right arm felt greater, here they had to indicate whether the two points felt farther apart in the first or the second stimulated location. This change in task was introduced because our

previous task may suffer from a first-order response bias, while the current task does not (Longo and Haggard, 2011). The current task has been previously used in other studies assessing tactile size perception (Longo and Haggard, 2011). Third, while previously the minimum experienced tactile distance was 2 cm, here it was 4 cm. This change was introduced in order to make sure that tactile distances were clearly suprathreshold at both body locations (Nolan, 1982, 1985), following the suggestion made by Canzoneri et al. (2013a,b).

To sum up, in our view, the nature of the task and the specific body parts used as test and reference locations seem to play a role in this relationship between tactile distance judgements and represented arm length changes, as different reference frames may be used for different body parts. Our study was not designed to directly tease these effects apart and therefore the exact relationship between tactile distance judgments and represented arm length remains open for further research. Having said this, while in our previous study we could not interpret the behavioral results in relation to the direction of changes in the represented arm length because the observed changes in perceived tactile distance were not accompanied by changes in the phenomenal experience of arm length (neither feeling that the arm elongated nor that it shrank was reported), in our present study we did observe such phenomenal changes. Changes in level of agreement with the statement "*my arm felt longer than usual*" significantly correlated with changes in perceived tactile distance on the arm, and in particular, larger increases in Pre- to Post-test PSE for the tactile distance task predicted larger feelings of one's arm being longer than usual. Given this subjective evidence, we suggest that the observed changes in perceived tactile distance relate to arm elongation.

## Conclusion

Our results show that self-produced tapping sounds can change perceived tactile distances but only when cues indicating both that one is the agent of the sounds and that when sound sources displace the tapping arm also displaces (i.e., kinaesthetic cues) are preserved. The present study adds to theories on forward internal models of motor-to-sensory transformations by showing that predictions related to action sounds must fit with kinesthetic cues related to the performed actions in order to make use of the auditory inputs to change body-representation. These cues reduce the spatial mismatch in the motor-to-sensory transformations, allowing a coherent, and robust sensory percept to emerge. Our results thus provide further insights on the necessary conditions (i.e., synchrony, agency, kinaesthesia) to observe audio-tactile influences on the coherence of body-representations. In addition, we showed, for the first time, that self-produced sounds can evoke consciously perceived changes in body-representation if a sufficiently immersing setup is provided. Finally, our results suggest that the nature of the tactile distance task and the specific body parts involved in the task may influence tactile distance judgements, as they may involve

different reference frames. With the task used in this study, a decrease in perceived tactile distance on the arm caused by audio-tactile adaptation seemed to predict feelings of one's arm elongating.

Future research should determine whether the kinaesthetic feedback should be resulting from active movements or whether passive movement is sufficient in order to observe audio-tactile influences on the coherence of body-representations, as well as determine whether the use of other sounds (i.e., non-action sounds) could have similar influences. Future studies may also test how long the effects of the audio-tactile adaptation last, as well as whether the observed effects would be enhanced or would diminish due to longer audio-tactile adaptation periods. Finally, it remains to be tested whether experienced changes in arm length scale with changes in the extent of the represented near space, as previous research has found a relation between arm lengths and represented near space (Longo and Lourenco, 2007).

This research on the dependency of body-representation upon auditory information complements previous research addressing the contribution of visual, tactile (Botvinick and Cohen, 1998), proprioceptive (de Vignemont et al., 2005; Ehrsson et al., 2005),

and vestibular (Lopez et al., 2012; Ferrè et al., 2013) channels to body-representation (for a recent review see Kilteni et al., 2015).

## Author Contributions

All authors contributed to the conception and design of the work, interpretation of data and revision of the drafts of the work. AT acquired and analyzed the data, and drafted the work. All authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved, and approved this final version of the manuscript.

## Acknowledgments

AT was supported by the ESRC grant ES/K001477/1 ("The hearing body"). MT was supported by the European Research Council (ERC-2010-StG-262853) under the FP7. The authors are grateful to Matthew R. Longo and Ophelia Deroy for useful comments on the experimental design and analyses.

## References

- Aglioti, S. M., and Pazzaglia, M. (2010). Representing actions through their sound. *Exp. Brain Res.* 206, 141–151.
- Begault, D. R. (1994). *3-D Sound for Virtual Reality and Multimedia*. San Diego, CA: Academic Press Professional, Inc.
- Blakemore, S.-J., Wolpert, D. M., and Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends Cogn. Sci.* 6, 237–242.
- Botvinick, M., and Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature* 391, 756.
- Bresin, R., de Witt, A., Papetti, S., Civolani, M., and Fontana, F. (2010). "Expressive sonification of footstep sounds," in *Proceedings of the Interaction Sonification Workshop (ISON) 2010*, eds R. Bresin, T. Hermann, and A. Hunt (Stockholm: KTH Royal Institute of Technology), 51–54.
- Canzoneri, E., Marzolla, M., Amoresano, A., Verni, G., and Serino, A. (2013a). Amputation and prosthesis implantation shape body and peripersonal space representations. *Sci. Rep.* 3, 2844. doi: 10.1038/srep02844
- Canzoneri, E., Ubaldi, S., Rastelli, V., Finisguerra, A., Bassolino, M., and Serino, A. (2013b). Tool-use reshapes the boundaries of body and peripersonal space representations. *Exp. Brain Res.* 228, 25–42.
- Cardinali, L., Frassinetti, F., Brozzoli, C., Urquizar, C., Roy, A. C., and Farne, A. (2009). Tool-use induces morphological updating of the body schema. *Curr. Biol.* 19, R478–R479.
- Cardinali, L., Jacobs, S., Brozzoli, C., Frassinetti, F., Roy, A. C., and Farne, A. (2012). Grab an object with a tool and change your body: tool-use-dependent changes of body representation for action. *Exp. Brain Res.* 218, 259–271.
- de Vignemont, F. (2010). Body schema and body image—pros and cons. *Neuropsychologia* 48, 669–680.
- de Vignemont, F., Ehrsson, H. H., and Haggard, P. (2005). Bodily illusions modulate tactile perception. *Curr. Biol.* 15, 1286–1290.
- Ehrsson, H. H., Kito, T., Sadato, N., Passingham, R. E., and Naito, E. (2005). Neural substrate of body size: illusory feeling of shrinking of the waist. *PLoS Biol.* 3:e412. doi: 10.1371/journal.pbio.0030412
- Ehrsson, H. H., Spence, C., and Passingham, R. E. (2004). "That's my hand!" activity in the premotor cortex reflects feeling of ownership of a limb. *Science* 305, 875–877.
- Ernst, M. O., and Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cogn. Sci.* 8, 162–169.
- Ferrè, E. R., Vagnoni, E., and Haggard, P. (2013). Vestibular contributions to bodily awareness. *Neuropsychologia* 51, 1445–1452.
- Fujisaki, W., Shimojo, S., Kashino, M., and Nishida, S. Y. (2004). Recalibration of audiovisual simultaneity. *Nat. Neurosci.* 7, 773–778.
- Haggard, P., Christakou, A., and Serino, A. (2007). Viewing the body modulates tactile receptive fields. *Exp. Brain Res.* 180, 187–193.
- Holmes, N., and Spence, C. (2004). The body schema and multisensory representation(s) of peripersonal space. *Cogn. Process.* 5, 94–105.
- Kallinen, K., and Ravaja, N. (2007). Comparing speakers versus headphones in listening to news from a computer: individual differences and psychophysiological responses. *Comput. Hum. Behav.* 23, 303–317.
- Kilteni, K., Maselli, A., Koerding, K. M. V., and Slater, M. (2015). Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception. *Front. Hum. Neurosci.* 9:141. doi: 10.3389/fnhum.2015.00141
- Kilteni, K., Normand, J. M., Sanchez-Vives, M. V., and Slater, M. (2012). Extending body space in immersive virtual reality: a very long arm illusion. *PLoS ONE* 7:e40867. doi: 10.1371/journal.pone.0040867
- Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 1094–1006.
- Lewkowicz, D. J. (1999). "The development of temporal and spatial intermodal perception," in *Cognitive Contributions to the Perception of Spatial and Temporal Events*, eds G. Aschersleben, T. Bachman, and J. Musseler (Amsterdam: Elsevier), 395–420.
- Li, X. F., Logan, R. J., and Pastore, R. E. (1991). Perception of acoustic source characteristics: walking sounds. *J. Acoust. Soc. Am.* 90, 3036–3049.
- Linkenauger, S. A., Witt, J. K., and Proffitt, D. R. (2011). Taking a hands-on approach: apparent grasping ability scales the perception of object size. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1432–1441.
- Longo, M. R., Azañón, E., and Haggard, P. (2010). More than skin deep: body representation beyond primary somatosensory cortex. *Neuropsychologia* 48, 655–668.
- Longo, M. R., and Haggard, P. (2011). Weber's illusion and body shape: anisotropy of tactile size perception on the hand. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 720–726.
- Longo, M. R., and Lourenco, S. F. (2007). Space perception and body morphology: extent of near space scales with arm length. *Exp. Brain Res.* 177, 285–290.
- Lopez, C., Schreyer, H. M., Preuss, N., and Mast, F. W. (2012). Vestibular stimulation modifies the body schema. *Neuropsychologia* 50, 1830–1837.

- Maravita, A., and Iriki, A. (2004). Tools for the body (schema). *Trends Cogn. Sci.* 8, 79–86.
- McDonald, J. H. (2014). *Handbook of Biological Statistics*, 3rd Edn. Baltimore, MD: Sparky House Publishing.
- Menzer, F., Brooks, A., Halje, P., Faller, C., Vetterli, M., and Blanke, O. (2010). Feeling in control of your footsteps: conscious gait monitoring and the auditory consequences of footsteps. *Cogn. Neurosci.* 1, 184–192.
- Miller, L. E., Longo, M. R., and Saygin, A. P. (2014). Tool morphology constrains the effects of tool use on body representations. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 2143–2153.
- Moore, J. W., Lagnado, D., Deal, D. C., and Haggard, P. (2009). Feelings of control: contingency determines experience of action. *Cognition* 110, 279–283.
- Murray, C. D., Arnold, P., and Thornton, B. (2000). Presence accompanying induced hearing loss: implications for immersive virtual environments. *Presence* 9, 137–148.
- Nolan, M. F. (1982). Two-point discrimination assessment in the upper limb in young adult men and women. *Phys. Ther.* 62, 965–969.
- Nolan, M. F. (1985). Quantitative measure of cutaneous sensation: two-point discrimination values for the face and trunk. *Phys. Ther.* 65, 181–185.
- Pazzaglia, M., Smania, N., Corato, E., and Aglioti, S. M. (2008). Neural underpinnings of gesture discrimination in patients with limb apraxia. *J. Neurosci.* 28, 3030–3041.
- Preston, C., and Newport, R. (2012). How long is your arm? Using multisensory illusions to modify body image from the third person perspective. *Perception* 41, 247–249.
- Senna, I., Maravita, A., Bolognini, N., and Parise, C. V. (2014). The marble-hand illusion. *PLoS ONE* 9:e91688. doi: 10.1371/journal.pone.0091688
- Serino, A., and Haggard, P. (2010). Touch and the body. *Neurosci. Biobehav. Rev.* 34, 224–236.
- Singh, A., Klapper, A., Jia, J., Fidalgo, A., Tajadura-Jiménez, A., Kanakam, N., et al. (2014). “Motivating people with chronic pain to do physical activity: opportunities for technology design,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto, ON: ACM Press, 2803–2812.
- Tajadura-Jiménez, A., Basia, M., Deroy, O., Fairhurst, M., Marquardt, N., and Berthouze, N. (2015a). “As light as your footsteps: altering walking sounds to change perceived body weight, emotional state and gait,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Seoul: ACM Press, 2943–2952.
- Tajadura-Jiménez, A., Furfaro, E., Bianchi-Berthouze, N., and Bevilacqua, F. (2015b). Sonification of virtual and real surface tapping: evaluation of behavior changes, surface perception and emotional indices. *MultiMedia IEEE* 22, 48–57.
- Tajadura-Jiménez, A., Pantelidou, G., Rebaz, P., Västfjäll, D., and Tsakiris, M. (2011). I-space: the effects of emotional valence and source of music on interpersonal distance. *PLoS ONE* 6:e26083. doi: 10.1371/journal.pone.0026083
- Tajadura-Jiménez, A., Våljamäe, A., Toshima, I., Kimura, T., Tsakiris, M., and Kitagawa, N. (2012). Action sounds recalibrate perceived tactile distance. *Curr. Biol.* 22, R516–R517.
- Tajadura-Jiménez, A., Våljamäe, A., and Västfjäll, D. (2008). Self-representation in mediated environments: the experience of emotions modulated by auditory-vibrotactile heartbeat. *Cyberpsychol. Behav.* 11, 33–38.
- Taylor-Clarke, M., Jacobsen, P., and Haggard, P. (2004). Keeping the world a constant size: object constancy in human touch. *Nat. Neurosci.* 7, 219–220.
- van der Hoort, B., Guterstam, A., and Ehrsson, H. H. (2011). Being barbie: the size of one’s own body determines the perceived size of the world. *PLoS ONE* 6:e20195. doi: 10.1371/journal.pone.0020195
- Vroomen, J., Keetels, M., de Gelder, B., and Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cogn. Brain Res.* 22, 32–35.
- Wolpert, D. M., and Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nat. Neurosci.* 3, 1212–1217.
- Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Tajadura-Jiménez, Tsakiris, Marquardt and Bianchi-Berthouze. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# What makes space-time interactions in human vision asymmetrical?

Chizuru T. Homma\* and Hiroshi Ashida

Graduate School of Letters, Kyoto University, Kyoto, Japan

## OPEN ACCESS

### Edited by:

Achille Pasqualotto,  
Sabanci University, Turkey

### Reviewed by:

Olesya Blazhenkova,  
Sabanci University, Turkey  
Shuichiro Taya,  
Taisho University, Japan

### \*Correspondence:

Chizuru T. Homma,  
Graduate School of Letters, Kyoto  
University, Yoshida-Honmachi,  
Sakyo-ku, Kyoto-shi, Kyoto 606-8501,  
Japan  
chizuru\_take@live.jp

### Specialty section:

This article was submitted to  
Cognition, a section of the journal  
Frontiers in Psychology

**Received:** 15 February 2015

**Accepted:** 21 May 2015

**Published:** 08 June 2015

### Citation:

Homma CT and Ashida H (2015)  
What makes space-time interactions  
in human vision asymmetrical?  
*Front. Psychol.* 6:756.  
doi: 10.3389/fpsyg.2015.00756

The interaction of space and time affects perception of extents: (1) the longer the exposure duration, the longer the line length is perceived and vice versa; (2) the shorter the line length is, the shorter the exposure duration is perceived. Previous studies have shown that space-time interactions in human vision are asymmetrical; spatial cognition has a larger effect on temporal cognition rather than vice versa (Merritt et al., 2010). What makes the interactions asymmetrical? In this study, participants were asked to judge exposure duration of lines that differed in length or to judge the lengths of the lines with different exposure time; to judge the task-relevant stimulus extents that also varied in the task-irrelevant stimulus extents. Paired spatial and temporal tasks in which the ranges of task-relevant and -irrelevant stimulus values were common, were conducted. In our hypothesis, the imbalance in saliency of spatial and temporal information would cause asymmetrical space-time interaction. To assess the saliency, task difficulty was rated. If saliency of relevant stimuli is high, the difficulty of discrimination task would be low, and vice versa. The saliency of irrelevant stimuli in one task would be reflected in the difficulty of the other task, in the pair of tasks. If saliency of irrelevant stimuli is high, the difficulty of paired task would be low, and vice versa. The result supports our hypothesis; spatial cognition asymmetrically affected on temporal cognition when the difficulty of temporal task was significantly higher than that of spatial task.

**Keywords:** space-time interaction, temporal cognition, spatial cognition, saliency, human vision, task difficulty

## Introduction

When people imagine that they are spending their time in a small room model, like a doll's house, they tends to feel the time shorter in a smaller room model compared to the estimated time in a larger room model (DeLong, 1981; Mitchell and Davis, 1987). Spatial extents of room model can alter subjective time. It is also known that more time was required to scan across mental images with greater distances, and to scan subjectively larger images (Mental Scanning; Kosslyn, 1973; Kosslyn et al., 1978). These are examples that show interactions between spatial and temporal cognition.

There are also cognitive interactions between number and space dimensions. In a numerosity discrimination task to compare two numbers, participants can react more rapidly when numerical and spatial extents are congruent, high digit with large size and low digit with small size, than when they are incongruent, high digit with small size and low digit with large size (Henik and Tzelgov, 1982). Many other cognitive interactions, like above, among three different dimensions (space, time and number), has been reported (e.g., Vicario, 2011; Javadi and Aichelburg, 2012). Accordingly, common mechanisms to process magnitude information of space, time and number has been suggested (a theory of magnitude; ATOM, Walsh, 2003).



The research topic of this study is on the cognitive interactions between space and time dimensions: (1) the longer the exposure duration is, the longer the line length tends to be judged and vice versa; (2) the shorter the line length is, the shorter the exposure duration tends to be judged and vice versa. Previous studies have repeatedly shown asymmetrical space-time interactions in vision of human adults; spatial cognition has a larger effect on temporal cognition rather than vice versa (Casasanto and Boroditsky, 2008; Merritt et al., 2010). However, such interactions in monkeys have been shown to be symmetrical (Merritt et al., 2010). In addition, space-time interactions in vision of human infants might be symmetrical. In 9-month-age infants, learning could be transferred among the three dimensions of time, space and number in vision, equally in each direction (Lourenco and Longo, 2010): learning of an arbitrary combination in one dimension, such as a stripe pattern of visual stimuli associated with a short exposure duration, can be transferred to the other dimension in every direction to a similar extent. Then, what is the difference between human adults and monkeys? How does the balance between space-time interactions in vision of human adults differ from that of infants? Trying to answer these questions is important for better understandings of spatial and temporal cognitions, the cognitive interaction and the development. The present study approaches to the specific question; what makes the interactions asymmetrical in vision of human adults?

As mentioned above, previous works have shown asymmetrical space-time interactions in vision of human adults; spatial cognition affects time cognition more than vice versa in discrimination tasks (Casasanto and Boroditsky, 2008; Merritt et al., 2010). One problem is that the balance of spatial and temporal information in the experimental stimuli has not been considered much. Therefore, in vision of human adults, the saliencies of spatial and temporal information might be one of the factors that make space-time interactions asymmetrical.

Many studies on cross-modal audiovisual interaction have shown the predominance of vision over audition (e.g., Thurlow and Jack, 1973; Kitagawa and Ichihara, 2002). However, when the ambiguity of visual information is high and the saliency of auditory information is high, auditory information could affect visual perception was presented later (Shimojo et al., 2001; Vroomen et al., 2004). A similar phenomenon has been found in the space-time cognitions, the saliency of stimuli should affect the balance between the space-time interactions.

Cai and Connell (2015) showed that time cognition can asymmetrically affect space cognition: spatial information from haptic perception can be affected by temporal information from audition but not vice versa. However, when spatial information from vision was added, space-time cognitions affect each other to a similar extent. According to them, the balance of space-time interaction could be affected by the perceptual acuity of the modality to perceive spatial information. The results indicated that the saliency of stimuli could affect the balance of space-time interaction in multi-modal perception.

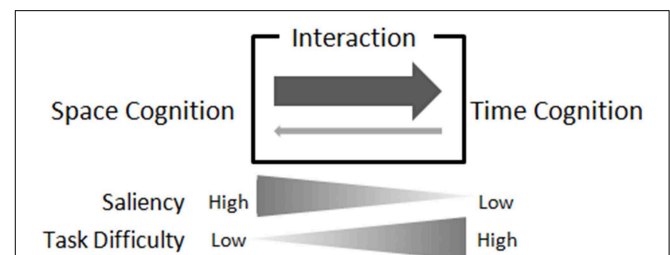
Thus, the present study aimed to investigate the effects of saliencies of spatial and temporal information in space-time interactions in vision. We conducted an experiment in which the tasks to discriminate exposure durations or line lengths. In

order to assess space-time interactions, we adapted the method of Merritt et al. (2010); line lengths were varied during duration judgments, and durations for presenting line stimuli were also varied during judgments of line lengths, and the method of Droit-Volet and Zélanti (2013); anchor stimuli, longest and shortest stimuli, were presented several times before an anchor training section. Task difficulty was rated to assess saliencies of spatial and temporal information. In a simple discrimination task, when the saliency of relevant stimuli is low, the automaticity in cognitive processes, such as discrimination, could be low, and thus the task difficulty could be high. Our hypothesis was that asymmetrical space-time interaction is caused by the imbalance in saliency of the spatial and temporal information, and the imbalance in difficulties in the spatial and temporal tasks (see **Figure 1**). In order to see the effect of saliency on the extent of interaction, two sets of paired spatial and temporal tasks that would be differed in the balance of difficulty (the pair that one task was more difficult than the other task, and the pair that two tasks were both difficult to the similar extent) were conducted. In each pair, the ranges of stimulus values were same; the shortest and the longest relevant stimuli in one task were same extents as the shortest and the longest irrelevant stimuli in the other task, in a pair. Thus, the difficulty of paired task would indicate the saliency of the irrelevant stimuli.

## Methods

### Participants

Twenty four adults (12 males, mean age: 23.13 years,  $SD = 3.13$ ) performed four tasks; two line length judgment tasks and two duration judgment tasks. All the participants had normal or corrected-to-normal vision. They were paid for the time by the standard of Kyoto University. The experiments were conducted in conformity to the standards of ethical review committee in Kyoto University. Through ethical considerations, before the



**FIGURE 1 | Balance of saliency, task difficulty, and spatial/temporal interaction.** In our hypothesis, asymmetrical space-time interaction is caused by the imbalance in saliency of the spatial and temporal information, and the imbalance in difficulties in the spatial and temporal tasks. When the saliency of relevant stimuli is high, the automaticity in cognitive processes, such as discrimination, could be high, and thus the task difficulty could be low; and vice versa. The saliency of irrelevant stimuli in one task would be reflected in the difficulty of the other task, in the pair of spatial/temporal tasks. If saliency of irrelevant stimuli is high, the difficulty of paired task would be low, and vice versa. When space cognition asymmetrically affects time cognition, saliency of spatial information would be high and/or saliency of temporal information would be low.

experiments, the content of the experiment and the rights of participants were explained, and the participants were asked to sign with agreement documents if they understood and agreed to participate in the experiment.

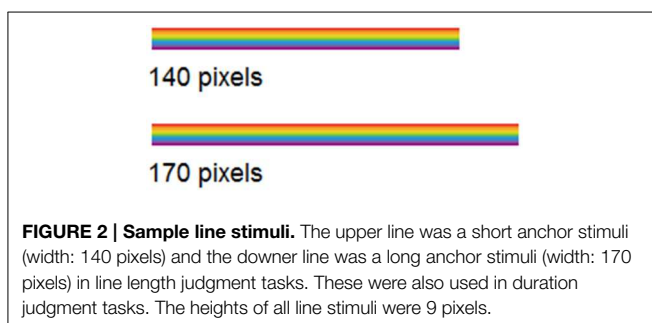
## Stimuli

Rainbow colored line stimuli were presented against a gray background (**Figure 2**). Similar experiments were planned for children, and the stimuli were colored in order to attract their attentions. The width of line stimuli was varied within the range of 140–170 pixels, and the exposure duration was varied within the range of 400–800 ms, or 1000–2000 ms. To make the balance of task difficulties different in two pairs of duration and line length judgment tasks, the exposure duration was varied within the above two ranges. In the line length judgment tasks, seven different widths of line stimuli were presented for three different durations. In the duration judgment task, three different widths of line stimuli were presented for seven different durations. The stimulus extents of relevant dimension had seven levels and that of irrelevant dimension had three levels. In each task, there were 21 stimulus presentation patterns (**Figure 3**). The line stimuli were presented on a 13-inch LCD display with a resolution of 1024 × 768 pixels.

## Procedure

The tasks varied in relevant dimensions for discrimination (line length or duration), and also varied in the range of exposure durations. There were therefore four conditions that were conducted in separate blocks: *Duration 400–800 ms, Line Length 400–800 ms, Duration 1000–2000 ms, Line Length 1000–2000 ms*. Half of the participants completed two line length judgment tasks ahead and two duration judgment tasks later, and the other half completed two duration judgment tasks ahead and two length judgment tasks later (**Figure 4**). The order of the two blocks for each task was counter-balanced. Task difficulty was rated after each two blocks ended, in five levels from one to five (1, easy; 2, a bit easy; 3, neither easy nor difficult; 4, a bit difficult; 5, difficult).

At the beginning of the experiments, participants were instructed to keep the same posture and the same position with a constant distance (varied across participants between 30 and 40 cm) from the monitor, to see all stimuli in the same way during tasks. Each task block consisted of three phases that were anchor training, bisection testing and cross-dimensional testing (**Figure 5**). For duration judgment tasks, the participants



were asked not to count, and for duration and line length judgment tasks, they were asked to think anything as much as possible, during the stimulus presentations. The experiments were controlled by a PC with E-prime software (Psychology Software Tools, Inc., USA).

## Duration Judgement Task

Participants were initially presented with 155 pixel width of line stimuli that were shortest and longest in exposure durations (anchor stimuli; 400 and 800 ms in *Duration 400–800 ms*; 1000 and 2000 ms in *Duration 1000–2000 ms*; see **Figure 3**) three times each in alternation, and were asked to remember them as the standard for later duration judgments. Before the presentation of the anchor stimuli started, the fixation cross was presented for 1000 ms. The interstimulus interval (ISI) of the anchor stimuli was fixed at 500 ms.

## Anchor Training

The participants were trained to judge short or long of exposure durations for the anchor stimuli that was presented once in each trial without any reference stimulus. The stimuli appeared following the fixation cross presented for 1000 ms. Immediately after the disappearance of the stimuli, the participants were asked to respond by pressing one of the two keys (“f” for short and “j” for long).

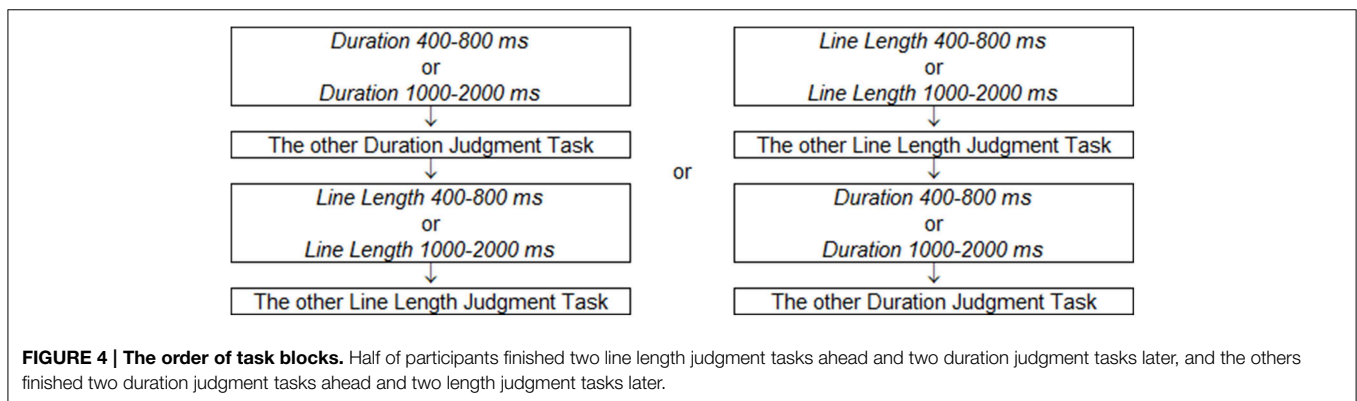
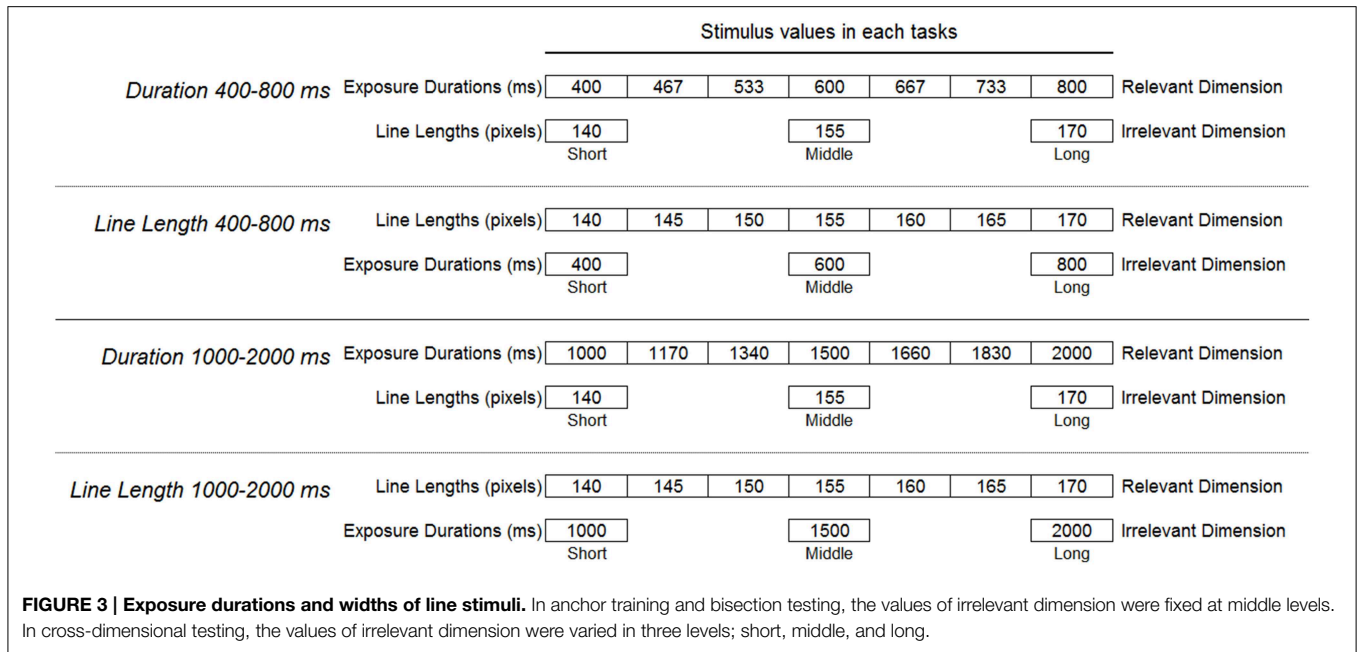
A visual feedback was given after each response: a red circle for a correct answer and a blue x-mark for a wrong answer. The feedback remained on the screen for 500 ms. In this phase, stimulus value of the irrelevant dimension, the length of line, was fixed at the Middle level (mean; 155 pixel width). There were 20 test trials that were separated into two trial blocks; the short and long anchor stimuli were randomly presented for five times, respectively in one block.

## Bisection Testing

The procedure was similar to the anchor training except that the exposure duration of stimuli was varied in seven levels (two anchor and five intermediate levels). The stimuli that have seven different exposure durations were presented in a random order. The number of presentations was differed depending on whether the stimulus was anchor or intermediate levels. In one trial block, the short and long anchor stimuli were presented for three times, and the five intermediate stimuli were presented for once. There were two blocks and 22 trials in total, and participants could take a rest between the blocks. The extent of the irrelevant dimension, line length, was fixed at the Middle level (155 pixel width). The flow of trials was the same as in the anchor training phase. There was a negative/positive feedback only for the anchor stimuli.

## Cross-dimensional Testing

The procedure was basically the same as the bisection testing. The exposure duration, the extent of the relevant dimension, was varied in seven levels. The line length, the extent of the irrelevant dimension, was varied in three levels: Short, Middle and Long. In one trial block, each anchor stimulus was presented for three times and each intermediate stimulus was presented once for



each level of irrelevant dimension in a random order. There were two blocks of 33 trials in the cross-dimensional testing, and rests were available between blocks. A negative/positive feedback was given only after the anchor stimuli as in bisection testing.

### Line Length Judgement Task

The procedure was basically the same as duration judgment task. The relevant and irrelevant dimensions were interchanged. The “short” and the “long” length of line stimuli (anchor stimuli; 140 pixel and 170 pixel widths; see **Figure 3**) were initially presented three times each in alternation, and participants were asked to remember them as the standard for later line length judgments.

### Anchor Training

The participants were trained to judge short or long of the width for anchor stimuli. A visual feedback was given after each response. In this phase, stimulus values of the irrelevant dimension, the exposure durations, were fixed at the Middle

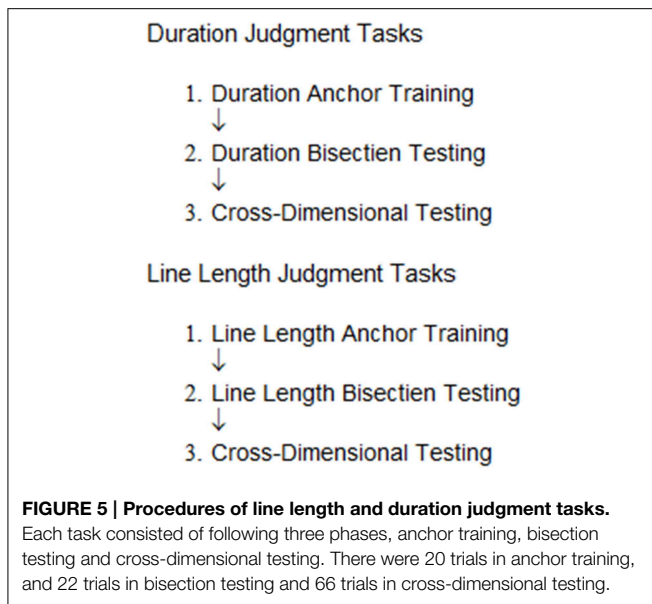
levels (mean; 600 ms in *Duration 400–800 ms*; 1500 ms in *Duration 1000–2000 ms*).

### Bisection Testing

The participants judged the width of stimuli varied in seven levels (two anchor and five intermediate levels). The seven different width of line stimuli were presented in a random order. In one trial block, the short and long anchor stimuli were presented for three times, and the five intermediate stimuli were presented for once. There were two blocks and 22 trials in total. The extents of the irrelevant dimension, the exposure durations, were fixed at the Middle levels (mean; 600 ms in *Duration 400–800 ms*; 1500 ms in *Duration 1000–2000 ms*). There was a negative/positive feedback only for the anchor stimuli.

### Cross-dimensional Testing

The line length, the extent of the relevant dimension, was varied in seven levels (two anchor and five intermediate levels). The exposure duration, the extent of the irrelevant dimension, was



varied in three levels: Short, Middle, and Long. In one trial block, each anchor stimulus was presented three times and each intermediate stimulus was presented once for each level of irrelevant dimension in a random order. There were two blocks of 33 trials in the cross-dimensional testing, and rests were available between blocks. A negative/positive feedback was given only after the anchor stimuli as in bisection testing.

## Results

Before the later analyses, the data of responses that took longer than 4000 ms were excluded as outliers. In trials which reaction time was longer than 4000 ms, the participants might not have concentrated on the stimuli or the task. The data of participants, who did not reach the criteria to judge the anchor stimuli correctly more than 80% in the last 10 trials of anchor training session, were also excluded. Two participants were excluded in *Duration 400–800 ms*, and one subject was excluded in each of the other three tasks.

To assess how saliency of irrelevant stimulus extents would affect on the discrimination of relevant stimuli, the results of combined two tasks (*Duration 400–800 ms* and *Line Length 400–800 ms*, *Duration 1000–2000 ms* and *Line Length 1000–2000 ms*) in which the ranges of stimulus extents were the same for both space and time, were separately analyzed. The 50% points of subjective equality (PSE) were estimated by the maximum likelihood method (Probit Analysis, Finney, 1971; Lieberman, 1983), in all conditions.

### **Duration 400–800 ms and Line Length 400–800 ms Bisection Testing**

The PSE in *Duration 400–800 ms* was 610.85 ms, and the PSE in *Line Length 400–800 ms* was 154.64 pixels. Reaction time was different between *Duration 400–800 ms* and *Line Length 400–800 ms*; the response in duration judgments took significantly

longer than line length judgments [ $t_{(20)} = 4.77, p = 0.000$ ] (see **Figure 7**).

### **Cross-dimensional Testing**

The PSE values were 608.44 ms, 593.85 ms, and 587.86 ms in *Duration 400–800 ms*; 154.48 pixels, 153.81 pixels, and 153.74 pixels in *Line Length 400–800 ms*, for Short, Middle, and Long extents of task irrelevant stimuli. The long line length stimuli were judged longer in duration, and the long exposure duration stimuli were judged longer in line length.

To assess cognitive interactions as effects from the extent of irrelevant dimension on judgments of relevant dimension, a mixed-effects logistic regression was conducted. In the analysis, seven values of the relevant dimension and three values of irrelevant dimension were used as predictors of “short” or “long” responses, and participants were considered as random effects. As a result, the main effect of the irrelevant dimension was significant for *Duration 400–800 ms* [ $\chi^2_{(1, N=1451)} = 4.54, p = 0.000$ ] but not for *Line Length 400–800 ms* [ $\chi^2_{(1, N=1513)} = 0.59, p = 0.44$ ] (see **Figure 6**). Reaction time was significantly longer in *Line Length 400–800 ms* than in *Duration 400–800 ms* and *Line Length 400–800 ms* [ $t_{(20)} = 5.73, p = 0.000$ ] (see **Figure 7**).

### **Task Difficulty**

The averaged task difficulty ratings for *Duration 400–800 ms* was the highest in all conditions, and significantly higher than that of *Line Length 400–800 ms*, according to t-test [ $t_{(20)} = 2.28, p = 0.03$ ] (see **Figure 7**).

### **Duration 1000–2000 ms and Line Length 1000–2000 ms Bisection Testing**

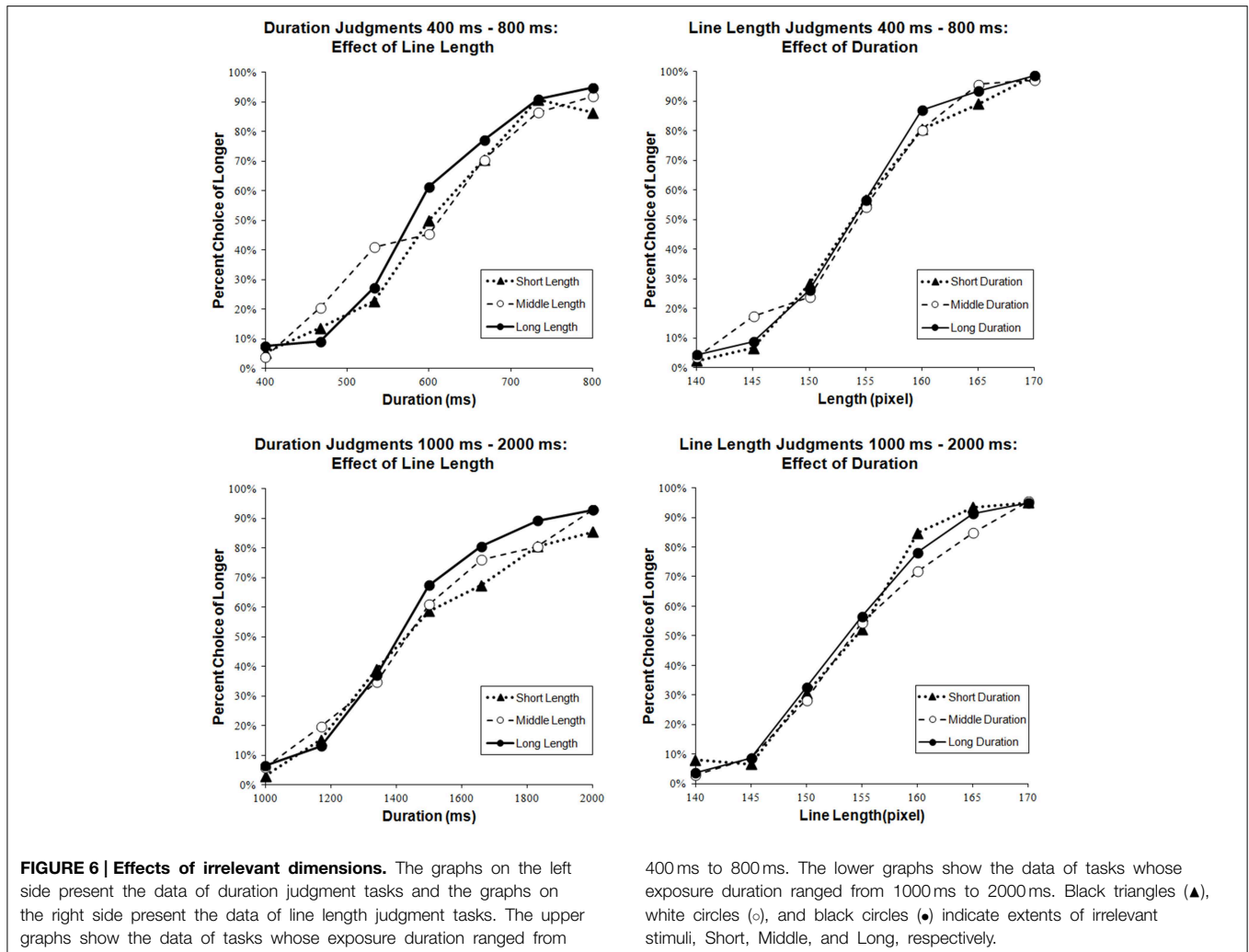
The PSE for *Duration 1000–2000 ms* was 1530.17 ms, and the PSE for *Line Length 1000–2000 ms* was 153.79 pixels. The response in *Duration 1000–2000 ms* took significantly longer than *Line Length 1000–2000 ms* [ $t_{(21)} = 2.84, p = 0.01$ ] (see **Figure 7**).

### **Cross-dimensional Testing**

The PSE values were 1507.06 ms, 1466.24 ms, and 1441.01 ms in *Duration 1000–2000 ms*; 153.96 pixels, 155.27 pixels, and 154.25 pixels in *Line Length 1000–2000 ms*, for Short, Middle, and Long extents of task irrelevant stimuli. The long line length stimuli were judged longer in duration, but such tendency could not be observed in line length; the long exposure duration stimuli were not always judged longer.

To assess cognitive interactions, a mixed-effects logistic regression was conducted in the same way as for *Duration 400–800 ms* and *Line Length 400–800 ms*. The main effects of the irrelevant dimension were not significant for *Duration 1000–2000 ms* [ $\chi^2_{(1, N=1501)} = 0.98, p = 0.32$ ] and for *Line Length 1000–2000 ms* [ $\chi^2_{(1, N=1505)} = 0.43, p = 0.51$ ] (see **Figure 6**). The response in *Duration 1000–2000 ms* took significantly longer than in *Line Length 1000–2000 ms* [ $t_{(21)} = 2.52, p = 0.02$ ] (see **Figure 7**).





**FIGURE 6 | Effects of irrelevant dimensions.** The graphs on the left side present the data of duration judgment tasks and the graphs on the right side present the data of line length judgment tasks. The upper graphs show the data of tasks whose exposure duration ranged from

400 ms to 800 ms. The lower graphs show the data of tasks whose exposure duration ranged from 1000 ms to 2000 ms. Black triangles ( $\blacktriangle$ ), white circles ( $\circ$ ), and black circles ( $\bullet$ ) indicate extents of irrelevant stimuli, Short, Middle, and Long, respectively.

### Task Difficulty

There was no significant difference between difficulty ratings of *Duration 1000–2000 ms* and *Line Length 1000–2000 ms*, according to *t*-test [ $t_{(21)} = -0.18, p = 0.86$ ; **Figure 7**]. The task difficulties were relatively high in both conditions.

### Discussion

The results of this study supported the hypothesis: asymmetrical space-time interaction is supposed to be caused by the imbalance in saliency of the spatial and temporal information, and difficulties in the spatial and temporal tasks, given the different pattern of results of combined two tasks (*Duration/Line Length 400–800 ms*, and *Duration/Line Length 1000–2000 ms*) in task difficulties and effects from the irrelevant dimension on relevant dimension.

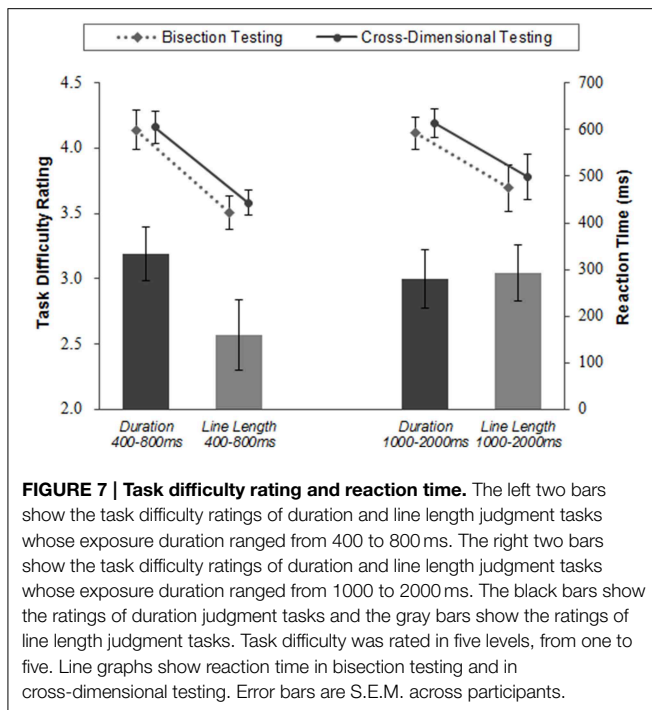
According to the results of a mixed-effects logistic regression, the effect of the irrelevant dimension was the largest in *Duration 400–800 ms* that was the most difficult, and the rating was significantly higher than that of *Line Length 400–800 ms*. On the other hand, the difficulties of *Duration 1000–2000 ms* and

*Line Length 1000–2000 ms* were similar. In this case, the effect of the irrelevant dimension on the judgment was not observed. These results can be interpreted as that the balance of difficulty between spatial-temporal cognitive tasks would affect the balance of cognitive interaction.

In discrimination tasks of this study, as already mentioned, when the task difficulty is high, the saliency of relevant stimuli would be low, and vice versa. In the sets of paired spatial and temporal tasks (*Duration/Line Length 400–800 ms*, *Duration/Line Length 1000–2000 ms*), the ranges of stimulus values were common, therefore the saliency of irrelevant stimuli would be high, when the difficulty of paired task is easy, and vice versa.

In *Duration 400–800 ms*, the task difficulty was high but the difficulty of paired task, *Line Length 400–800 ms* was low, thus the saliency of relevant stimuli was low but the saliency of irrelevant stimuli was high, so that the effect of irrelevant stimuli on the discrimination was statistically significant. In other tasks, the saliency of irrelevant stimuli would not be high enough to affect on the discrimination significantly.

There was a significant difference in reaction time between the spatial and the temporal cognitive tasks. The reaction time



of duration judgment was significantly longer than that of line length judgment. This difference in reaction time might reflect the imbalance of stimulus saliency between visual space-time cognitions, as discussed above for the asymmetrical interactions. The processes and/or representations of spatial and temporal information might be partially common or similar (ATOM: Walsh, 2003), although fundamental differences might exist between spatial and temporal cognitions with vision. Such differences may be reflected in reaction time; reaction time of temporal cognitive tasks was longer than that of spatial cognitive tasks, even in the bisection testing in which the irrelevant stimulus extent was fixed at the Middle level, and even though when the task difficulties were similar. The saliency of visual spatial information (the line length extents) would tend to be higher than the saliency of visual temporal information, therefore the automaticity of line length discrimination would tend to be higher than that of duration discrimination via visual perception.

Human adults have well-developed visual perception, and vision dominates over other modalities such as audition, in many cases, in the process to integrate information from several modalities especially in spatial cognition, such as the ventriloquism effect (Thurlow and Jack, 1973). However, in time perception, visual information can be affected by auditory information. It can be seen in temporal ventriloquism and visual illusions by audition, the phenomenon that the number or timing of flashes can be differently perceived from actual vision, which is

caused by hearing sounds (Shimojo et al., 2001; Vroomen et al., 2004). In human adults, spatial information via vision tends to be more precise compared to that via audition, and thus vision has predominance over audition in cross-modal spatial cognition. In contrast, the saliency of visual temporal information is low so that audition dominates over vision in cross-modal temporal cognition, in many cases.

In integration of cross-modal information, information with higher saliency would have the predominance over that with lower saliency. As well as in cross-modal perception, in space-time interaction, spatial information affects time information more than vice versa, due to the balance of saliency, in vision of human adults. Such a common hypothesis on integration in cross-modal and cross-dimensional cognitions is supposed to be plausible in terms of ecological validity. Such biases in integration would make the information integration more efficient. In addition, this hypothesis can approach a remaining question: what is the difference between human adults and monkeys? In monkeys, saliency of spatial information from vision might not be so high, or saliency of temporal information from vision might not be so low compared to human adults, or both. Therefore, spatial and temporal information in vision would be reliable to the same degree, which may have led to the symmetrical interaction between spatial and temporal cognition in vision of monkeys, as found in Merritt et al. (2010).

It remains unknown whether it is possible to make the balance between spatial-temporal cognitive interactions in vision reversed: the interaction from time to space is larger than vice versa. Cai and Connell (2015) have proved that the balance of space-time interaction in multi-modal perception can be reversed. So such a reversal in space-time interaction in vision could happen, if the saliency of temporal information becomes higher than that of spatial information. It is still open to the future studies to elaborate the way to assess saliency. In the present study, the task difficulty rating was considered as a related variable to the saliency of judged stimuli, but there could be other ways. It would be better to consider reaction time and other factors such as discrimination sensitivity comprehensively, as well as task difficulty.

## Acknowledgments

This work was supported by JSPS Grant-in-Aid for Scientific Research (S22220003). We would like to express our deep gratitude to Prof. Dr. Toshiyuki Yamashita whose enormous support and insightful comments were invaluable during the course of our study. Advice and comments given by reviewers has been a great help in improving this work. We would like to show our great appreciation to them and the assistance we received from the editor, Dr. Achille Pasqualotto.

## References

- Cai, Z. G., and Connell, L. (2015). Space-time interdependence: evidence against asymmetric mapping between time and space. *Cognition* 136, 268–281. doi: 10.1016/j.cognition.2014.11.039
- Casasanto, D., and Boroditsky, L. (2008). Time in the mind: using space to think about time. *Cognition* 106, 579–593. doi: 10.1016/j.cognition.2007.03.004
- DeLong, A. J. (1981). Phenomenological space-time: toward an experiential relativity. *Science* 213, 681–683. doi: 10.1126/science.7256273

- Droit-Volet, S., and Zélandi, P. (2013). Development of time sensitivity: duration ratios in time bisection. *Q. J. Exp. Psychol.* 66, 671–686. doi: 10.1080/17470218.2012.712148
- Finney, D. J. (1971). *Probit Analysis*. Cambridge: Cambridge University Press.
- Henik, A., and Tzelgov, J. (1982). Is three greater than five: the relation between physical and semantic size in comparison tasks. *Mem. Cogn.* 10, 389–395. doi: 10.3758/BF03202431
- Javadi, A. H., and Aichelburg, C. (2012). When time and numerosity interfere: the longer the more, and the more the longer. *PLoS ONE* 7:e41496. doi: 10.1371/journal.pone.0041496
- Kitagawa, N., and Ichihara, S. (2002). Hearing visual motion in depth. *Nature* 416, 172–174. doi: 10.1038/416172a
- Kosslyn, S. M. (1973). Scanning visual images: some structural implications. *Percept. Psychophys.* 14, 90–94. doi: 10.3758/BF03198621
- Kosslyn, S. M., Ball, T. M., and Reiser, B. J. (1978). Visual images preserve metric spatial information: evidence from studies of image scanning. *J. Exp. Psychol.* 4, 47–60. doi: 10.1037/0096-1523.4.1.47
- Lieberman, H. R. (1983). Computation of psychological thresholds using the probit technique. *Behavior Res. Meth. Instrum.* 15, 446–448. doi: 10.3758/BF03203681
- Lourenco, S. F., and Longo, M. R. (2010). General magnitude representation in human infants. *Psychol. Sci.* 21, 873–881. doi: 10.1177/0956797610370158
- Merritt, D. J., Casasanto, D., and Brannon, E. M. (2010). Do monkeys think in metaphors? Representations of space and time in monkeys and humans. *Cognition* 117, 191–202. doi: 10.1016/j.cognition.2010.08.011
- Mitchell, C. T., and Davis, R. (1987). The perception of time in scale model environments. *Perception* 16, 5–16. doi: 10.1068/p160005
- Shimojo, S., Scheier, C., Nijhawan, R., Shams, L., Kamitani, Y., and Watanabe, K. (2001). Beyond perceptual modality: auditory effects on visual perception. *Acoust. Sci. Technol.* 22, 61–67. doi: 10.1250/ast.22.61
- Thurlow, W. R., and Jack, C. E. (1973). Certain determinants of the “ventriloquism effect”. *Percept. Motor Skills* 36, 1171–1184.
- Vicario, C. M. (2011). Perceiving numbers affects the subjective temporal midpoint. *Perception* 40, 23–29. doi: 10.1068/p6800
- Vroomen, J., de Gelder, B., and Vroomen, J. (2004). Temporal ventriloquism: sound modulates the flash-lag effect. *J. Exp. Psychol.* 30, 513–518. doi: 10.1037/0096-1523.30.3.513
- Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* 7, 483–488. doi: 10.1016/j.tics.2003.09.002

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Homma and Ashida. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Tactile input and empathy modulate the perception of ambiguous biological motion

Hörmetjan Yiltiz<sup>1</sup> and Lihan Chen<sup>1,2\*</sup>

<sup>1</sup> Department of Psychology, Peking University, Beijing, China

<sup>2</sup> Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China

## Edited by:

Magda L. Dumitru, Macquarie University, Australia

## Reviewed by:

Quoc Vuong, Newcastle University, UK

Yves Philippe Rybarczyk, New University of Lisbon, Portugal

## \*Correspondence:

Lihan Chen, Department of Psychology, Peking University, 5 YiHeYuan Road, HaiDian District, Beijing 100871, China  
e-mail: clh@pku.edu.cn

Evidence has shown that task-irrelevant auditory cues can bias perceptual decisions regarding directional information associated with biological motion, as indicated in perceptual tasks using point-light walkers (PLWs) (Brooks et al., 2007). In the current study, we extended the investigation of cross-modal influences to the tactile domain by asking how tactile input resolves perceptual ambiguity in visual apparent motion, and how empathy plays a role in this cross-modal interaction. In Experiment 1, we simulated the tactile feedback on the observers' fingertips when the (upright or inverted) PLWs (comprised of either all red or all green dots) were walking (leftwards or rightwards). The temporal periods between tactile events and critical visual events (the PLW's feet hitting the ground) were manipulated so that the tap could lead, synchronize, or lag the visual foot-hitting-ground event. We found that the temporal structures between tactile (feedback) and visual (hitting) events systematically biases the directional perception for upright PLWs, making either leftwards or rightwards more dominant. However, this effect was absent for inverted PLWs. In Experiment 2, we examined how empathy modulates cross-modal capture. Instead of giving tactile feedback on participants' fingertips, we gave taps on their ankles and presented the PLWs with motion directions of approaching (facing toward observer)/receding (facing away from observer) to resemble normal walking postures. With the same temporal structure, we found that individuals with higher empathy were more subject to perceptual bias in the presence of tactile feedback. Taken together, our findings showed that task-irrelevant tactile input can resolve the otherwise ambiguous perception of the direction of biological motion, and this cross-modal bias was mediated by higher level social-cognitive factors, including empathy.

**Keywords:** tactile, point-light walker, temporal, empathy, apparent motion, binocular rivalry

## INTRODUCTION

Perceiving and recognizing biological motion patterns in a complex and cluttered environment is vital for human survival. Our understanding of the perception of biological motion has been increased by advancements in research methodology and paradigms (Cutting and Kozlowski, 1977; Cutting, 1978; Watson et al., 2004; Kim et al., 2010; van Boxtel and Lu, 2013). One development in methodology that has benefitted research in this domain is the use of point-light walkers. Johansson first introduced point-light walkers to examine how well human observers could extract motion and form information for a simulated walking person from the characteristic light dots rendering key parts of the human body (Johansson, 1973). This novel paradigm proved to be very successful and has been used extensively to investigate perceptual organization and visual attention in complex environments for more than two decades (Schmuckler and Fairhall, 2001; Servos et al., 2002; Beauchamp et al., 2003; Hirai and Hiraki, 2005; Troje et al., 2006; Brooks et al., 2007; Arrighi et al., 2009; Das et al., 2009; Hirai et al., 2009; Herrington et al., 2011; Pavlova et al., 2014). Researchers initially examined how observers could use visual cues to facilitate the detection of certain features (either

static or dynamic motion information) among the given PLWs (Das et al., 2009; de Lussanet and Lappe, 2012).

Studies have also addressed how people process social information that is embedded in the PLW, such as gender (Barclay et al., 1978; Pollick et al., 2005) and emotion (Ma et al., 2006; Johnson et al., 2011; Henry et al., 2012). Perception of PLWs has been shown to be modulated by individual differences and personality traits, such as age (Norman et al., 2004), identity (Barclay et al., 1978; Cutting, 1978; Troje et al., 2005), and the self-serving bias. Regarding the self-serving bias, a recent study revealed that in perceiving the receding/approaching directional information for PLWs, observers with high social anxiety are less likely to report the PLW as approaching, compared to observers with low social anxiety. This bias might reflect an assumption that mistaking approach for withdrawal is worse than the reverse (Van de Cruys et al., 2013; Weech et al., 2014).

In naturalistic settings and daily life however, it is often the case that biological motion involves information from more than one modality. Thus, research into the role of multi-modal information in biological motion is necessary for a more comprehensive understanding of biological motion. While studies utilizing PLWs



were originally confined to the visual modality, they have fortunately been extended to a multisensory context in recent years. In particular, several studies have targeted how auditory inputs resolve the otherwise ambiguous directional perception of PLWs. Brooks et al. (2007) investigated the effect of suprathreshold auditory motion on perceptions of visually-defined biological motion. Here, researchers manipulated the same (congruent) or opposite (incongruent) directions between auditory motion and visual motion, and found a direction-congruent effect between auditory events and visual PLWs. Relative to control auditory conditions, auditory motion in the same direction as the visually-defined biological motion target increased its detectability. However, it decreased detectability of the biological motion target when the directions of auditory motion and the visual PLW were incongruent (Brooks et al., 2007). In a similar vein, Kim et al. (2010) found a general improvement for the detection of a point-light *talking* face among point-light distractors, in the presence of congruent/matched auditory speech. This suggests that concomitant action-consistent sounds enhance visual sensitivity to the presence of coherent point-light displays of human movement. Thomas and Shiffrar (2010) examined further whether the visual detection sensitivity of PLWs is modulated by the meaningfulness of sounds that are concomitant with observed point-light actions. They revealed that detection sensitivity increased as a result of the veridical auditory cues (footfalls) but not as a result of pure tones. Taken together, the above studies suggest that the correspondence of auditory information to visual information, whether in lower perceptual features (direction) or higher cognitive factors (semantic relatedness), could to a large extent enhance visual sensitivity to the presence of coherent point-light displays of human movement.

The cross-modal influence of sensory inputs on perception of PLWs was driven mainly by temporal factors. For instance, performance on identifying upright PLWs was better when the visual “footsteps” were phase-locked with the auditory events. However, this advantage disappeared when the visual footsteps were out of phase with the auditory events (Saygin et al., 2008). The cross-modal influence on the temporal “capture” effect has been termed the “temporal ventriloquism effect.” In a typical dynamic ventriloquism effect, the perceived direction of the bistable visual motion (either leftwards or rightwards) is discerned by temporal alignments between distractor events (auditory events) and target (visual or tactile) events in the apparent motion (Slutsky and Recanzone, 2001; Bertelson and Aschersleben, 2003; Morein-Zamir et al., 2003; Vroomen et al., 2004; Shi et al., 2010; Chen and Vroomen, 2013). However, the distractor events provided no spatial cue (or motion direction) information and the temporal disparity between cross-modal events was beyond conscious perception (Freeman and Driver, 2008; Chen et al., 2011).

The current study aims to extend the research just discussed. Its purpose is two-fold. First, tactile events, like auditory signals, share the Gestalt principle of perceptual organization, so that paired tactile events could serve as temporal cues to influence the timing of visual/auditory events, and even cause a multisensory illusion-ventriloquism effect (Gallace and Spence, 2010, 2011). Therefore, events from a third modality, such as tactile input associated with veridical and ecologically meaningful feedback on the

visual footfalls of PLWs, could affect the perception of PLWs. This would be the case as long as there was appropriate temporal alignment between the onset times of the tactile inputs and the motion simulated by the PLW. Investigation along this line has not yet been documented. Therefore, we aimed to examine how the tactile temporal perceptual grouping (with visual frames of PLWs) influences the perception of the directional information of PLWs. The effect of the cross-modal temporal capture was measured by the variation in the perceived dominant durations of PLWs in one direction.

Second, as we described previously, perception of PLWs mobilizes not only low-level visual processing, but involves high-level cognitive inputs such as the cognitive states of the observers, due to the fact that PLWs can invoke social and emotional responses (Van de Cruys et al., 2013). Social neuroscience models have assumed that people tend to use the self as a reference point to perceive the world and gain information about other people’s mental states. Further, people rely mainly on their own cognitive states as a reference for empathy (Silani et al., 2013). Recent studies have also shown the neural basis for individual differences in empathy. Somatosensory response in the primary somatosensory cortex (SI) has been associated with the empathy subscale of perspective taking (Schaefer et al., 2012). This link demonstrates that vicarious somatosensory responses for simple touch are influenced by the observer’s personality traits. That is, people with higher empathic concern would be more sensitive to other individuals’ suffering (Banissy and Ward, 2007). We intend to apply tactile feedback to the participants as vicarious feedback from the PLWs. This essentially requires the participants to associate the experience of the first-person (the participant) and the third person (the PLWs) when they interpret the motion state of the PLWs with (dissociated) tactile feedback. From the above reasoning, we speculate that people with higher empathy will involve themselves more in the current cross-modal interaction task (Gallese et al., 2004; Cattaneo and Rizzolatti, 2009), and would therefore show a modulation effect of empathy upon the tactile temporal capture effect. Among the many operational techniques in PLWs, binocular rivalry remains a rigorous paradigm that induces potential perceptual bistability (Watson et al., 2004). This could however, be explained by different factors, including postures and cross-modal sensory inputs (Brooks et al., 2007; Kim et al., 2010).

Using the paradigm of binocular rivalry, we conducted two experiments to test the following hypotheses: (1) Tactile events as simulations of visual footsteps could help to organize the directional information of the otherwise ambiguous/bistable apparent motion of PLWs; (2) The tactile-visual dynamic temporal capture effect of the directional perception of PLWs is constrained by higher-level social-cognitive factors, including an individual’s empathy.

## EXPERIMENT 1

### METHOD

#### *Participants*

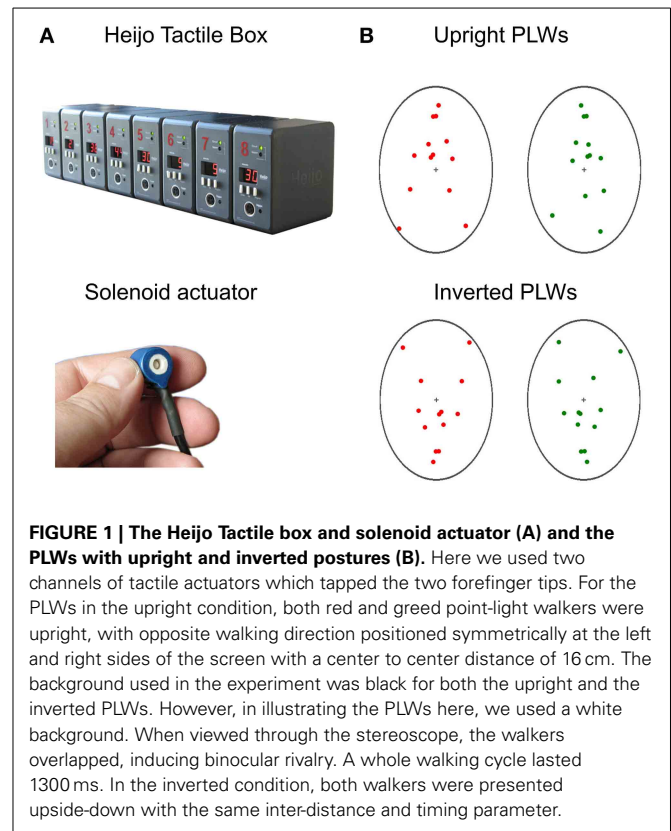
Sixteen undergraduate students (7 female) from Peking University, aged 19–23 years, with normal or corrected-to-normal vision participated in the experiment. None of them had color-blindness or partial colorblind symptoms, they reported

normal hearing, and normal somatosensory sensation. The experiment was conducted on each participant individually, in a dimly lit standard experimental booth. The experiment was performed in compliance with all institutional guidelines set by the Academic Affairs Committee of the Department of Psychology at Peking University. All participants provided written informed consent according to institutional guidelines and the Declaration of Helsinki. Participants were reimbursed after the experiment.

### Stimuli and apparatus

The raw data for composing the point-light walker's stimuli were obtained from CMU Graphics Lab Motion Capture Database (<http://mocap.cs.cmu.edu>). We presented two point-light walkers. Each PLW was either completely red or completely green, and was either upright or inverted. A point-light walker consisted of 13 dots, representing some of the key joints of the body, including the head, shoulders, elbows, hands, hips, knees, and feet (Ahlstrom et al., 1997). Each PLW extended approximately 6 (high)  $\times$  4 (wide) degrees of visual angle on screen, viewed from a distance of 60 cm to the eyes of the observer. The distance between the center of the two PLWS was kept at 16 cm, where the walking direction for each PLW was either leftwards or rightwards. However, the two PLWs were mirror-reflected in the stereoscope so that they converged and overlapped at the center of the screen. As a result, each eye of the observer only saw a single PLW at the corresponding side, which induces binocular rivalry (see the following procedure). The walking directions for the PLWs in each trial were randomized and counterbalanced. A full walking cycle for a PLW was 1300 ms, with 130 frames presented at a vertical refresh rate of 10 ms per frame. The visual display was a 19 inch CRT (ViewSonic) with a resolution of 1024  $\times$  768, at a vertical refresh rate of 100 Hz, which enabled the inter-frame time interval between visual stimuli to be set at 10 ms. Red and green stimuli were equiluminant at 14.88 and 10.49 cd/m<sup>2</sup> respectively, on a black screen background with a luminance of 0.17 cd/m<sup>2</sup>.

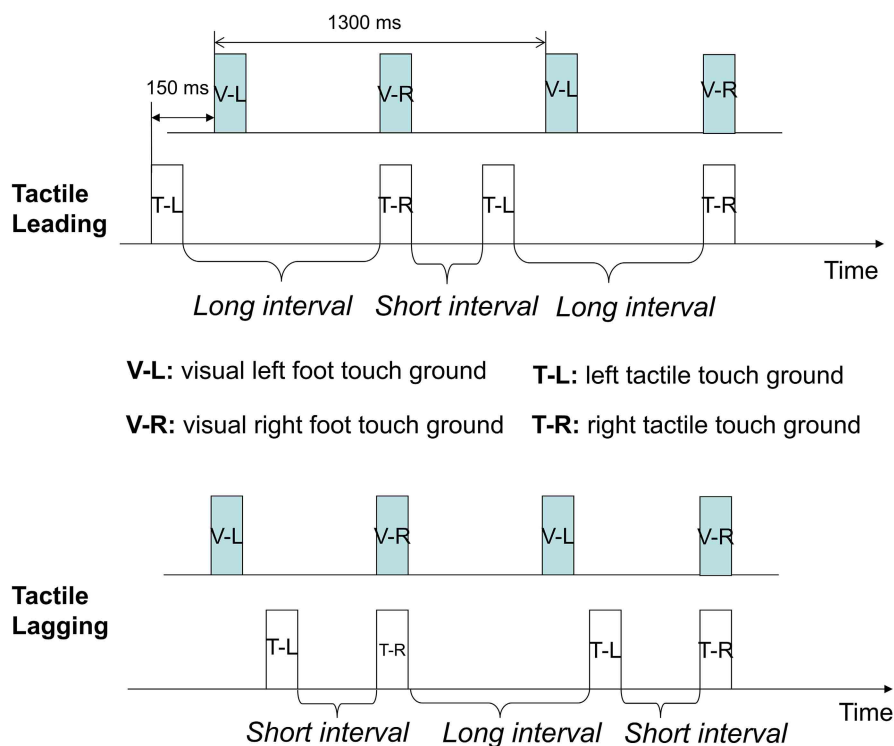
The tactile stimuli were produced using solenoid actuators with embedded cylinder metal tips, which would tap the fingertips to induce indentation taps when the solenoid coils were magnetized (Heijo Box, Heijo Research Electronics, UK, as shown in **Figure 1**). The maximum contact area is about 4 mm<sup>2</sup> and the maximum output is 3.06 W. Two tactile stimuli, simulating one of the (randomly chosen by trial) point-light walker's footsteps touching the ground, were presented on the index fingers. The temporal structures for the tactile stimuli and visual stimuli were as follows: the first tactile stimulus for each trial (e.g., the left tactile stimulus simulating the tactile feedback of a visual left footstep) was synchronized with the corresponding visual stimulus (e.g., the left visual footstep) for the whole trial. The second tactile stimulus either *preceded* 150 ms, *synchronized*, or *lagged* 150 ms to the corresponding visual frame of the PLW's footstep hitting the ground, as shown in **Figure 2**. The duration for a single tap lasted 10 ms. Each initial tap was assigned to either the left forefinger tip or the right forefinger tip. The order was randomized and counterbalanced across all experimental trials, also shown in **Figure 2**. To give more detail, in the "tactile leading" temporal condition, one tap was leading 150 ms to one visual



**FIGURE 1 | The Heijo Tactile box and solenoid actuator (A) and the PLWs with upright and inverted postures (B).** Here we used two channels of tactile actuators which tapped the two forefinger tips. For the PLWs in the upright condition, both red and green point-light walkers were upright, with opposite walking direction positioned symmetrically at the left and right sides of the screen with a center to center distance of 16 cm. The background used in the experiment was black for both the upright and the inverted PLWs. However, in illustrating the PLWs here, we used a white background. When viewed through the stereoscope, the walkers overlapped, inducing binocular rivalry. A whole walking cycle lasted 1300 ms. In the inverted condition, both walkers were presented upside-down with the same inter-distance and timing parameter.

footstep (visually touching the ground), while the other tap was synchronous with the second visual footstep. In contrast, the lower figure showed the "tactile lagging" condition, in which one tap was lagging 150 ms to one visual footstep while the other tap was synchronous with the onset of the second visual footstep. The pairing of visual and tactile stimuli could be organized into interleaved short intervals and long intervals along the whole presentation duration (70 s) of PLWs. There were another two conditions: "synchronous" and "baseline." In the synchronous condition, both taps were synchronous with the corresponding critical visual footsteps (hitting twice on the ground), while in the baseline condition, no taps were given. Participants' responses in the tactile leading or tactile lagging conditions were further recorded as either "congruent" or "incongruent." For the tactile leading condition, responses were recorded as congruent if they were in the opposite of the direction of the initial tactile motion (a "left" response for initial rightward motion was recorded as congruent). In the tactile lagging condition, responses were recorded as congruent if they were in accordance with the direction of the initial tactile motion (a "left" response for initial leftward tactile motion was recorded as congruent), this recoding method was based on the perceived direction of tactile motion from the above different temporal structures and was in accordance with previous studies (Freeman and Driver, 2008; Chen et al., 2011).

The computer programs used in Experiments 1 and 2 were developed with Matlab (Mathworks Inc.) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The test booth was semi-anechoic and dimly lit throughout the experiment, with ambient



**FIGURE 2 | Temporal structures of visual-tactile stimuli in PLWs.**

Here, two of the eight experimental conditions of Experiment 1 are shown. The upper figure shows the “tactile leading” temporal condition, in which one tap was leading 150 ms to one visual footstep (visually touching the ground), while the other tap was synchronous

with the second visual footstep. In contrast, the lower figure shows the tactile lagging condition, in which one tap was lagging 150 ms to one visual footstep while the other tap was synchronous with the onset of the second visual footstep. V, visual; T, tactile feedback (tap); L, left; R, right.

luminance of  $0.05 \text{ cd/m}^2$ . The viewing distance was fixed at 60 cm, which was maintained by using a chin-rest.

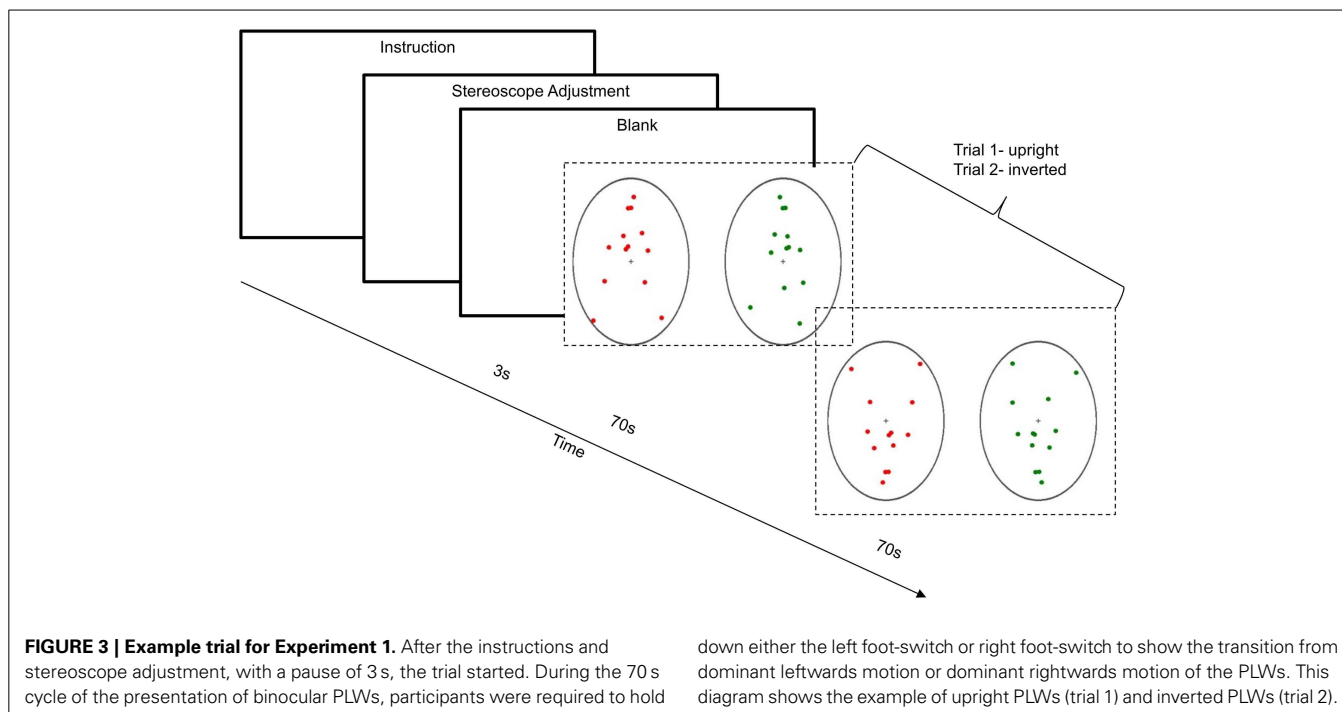
### Design and procedure

A 2 (posture: upright vs. inverted)  $\times$  4 (temporal structure: tactile leading, synchronous, lagging to the visual footstep, and baseline without taps) factorial design was adopted in this experiment. Participants were asked to report the perceived dominant walking direction of the point-light walker on the screen by pressing and holding the corresponding foot switch. The left switch was used to indicate leftward motion and the right switch was used to indicate rightward motion).

A complete cycle for the presentation of PLWs lasted 1300 ms. The total time duration for each single trial (i.e., the apparent motion of PLWs) was 70 s. Each condition was repeated and had five trials. The above tactile-visual temporal conditions were randomized and counterbalanced across all the trials. The inter-trial interval (ITI) between the two trials was 600–1000 ms. The onset of the first tactile stimulus was not started until 3000 ms (with a standard deviation of 500 ms) after the onset of the visual PLWs. The responses of the participants were not recorded for the first 10 s of each trial, beginning with the onset of the PLWs. This was done to prevent the initial bias of response arising from the first events (taps and visual PLWs), as shown in **Figure 2**.

Before taking part in the formal experiment, participants were asked to read the instructions and were provided with further detailed information related to the task when necessary. However, none of the participants knew the purpose of the experiment. The position of the stereoscope was adjusted in advance so that for each individual, the center of the point-light walkers could be perceived as overlapping before starting the experimental trials. A short video demonstration of the binocular PLWs was given before the formal experiment so that the participants would be familiar with the task. Then, they were trained in a pre-experiment with four trials containing each condition, to ensure they were capable of performing the required task. Each participant wore sponge earplugs and a headset to prevent any faint tactile noise during the experiment. During the experiment, they were required to focus on the central cross (fixation point) and report the perception of the dominant motion direction (leftwards vs. rightwards) of the perceived PLW projected through the stereoscope for 70 s by holding down the left foot-switch or right foot-switch, as shown in **Figure 3**. As explained earlier, the first 10 s of responses were not recorded.

After the formal experiment, we conducted a control test in which participants were asked to report the perceived dominant direction (leftwards or rightwards) of tactile apparent motion, based on the same temporal conditions as in the main experiment (tactile preceding 150 ms, synchronous, or lagging 150 ms to the

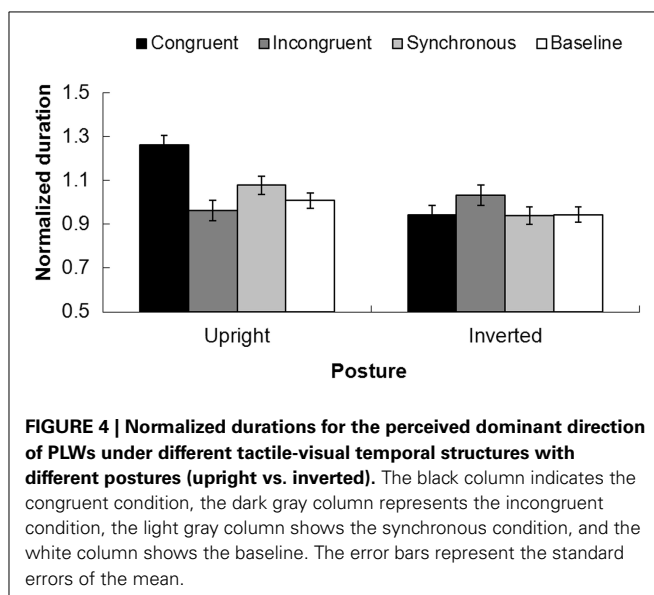


visual footstep of one PLW). We examined whether different temporal intervals between taps give rise to the dominant directional perception of the tactile motion, as in Chen et al. (2011), which contribute to capturing the dominant directional perception of the PLWs.

### Results

The durations for holding the left switch or right switch were sorted separately by each temporal structure in upright and inverted postures. Since there was a large amount of individual variance, we normalized the duration by dividing the holding time with the mean across the four temporal conditions. The averaged normalized duration for all the participants are shown in Figure 4.

An Analysis of Variance (ANOVA) with the postures of point-light walkers (upright or inverted) and the recoded temporal conditions (“congruent,” “incongruent,” “synchronous,” and “baseline”) as independent factors and dominant durations as a dependent factor showed a significant main effect of posture,  $F_{(1, 30)} = 15.050$ ,  $p < 0.01$ . The duration of the perceived normalized dominant direction for the upright point-light walker (Mean = 1.007, SEM = 0.185) was longer than the one in the inverted posture (Mean = 0.964 SEM = 0.174). The main effect for temporal conditions was also significant,  $F_{(3, 90)} = 3.558$ ,  $p < 0.05$ . Bonferroni-corrected pairwise analysis showed the dominant duration in the congruent condition (Mean = 1.102, SEM = 0.225) was significantly longer than the ones in the synchronous condition (Mean = 1.008, SEM = 0.188) and baseline (Mean = 0.976, SEM = 0.169) conditions, but no difference between synchronous and baseline conditions,  $p > 0.05$ . The interaction between the temporal structure between tactile stimuli and visual stimuli and the posture was significant,  $F_{(3, 90)} = 7.645$ ,  $p < 0.001$ .



A repeated measures ANOVA was implemented for upright and inverted postures separately. For the upright posture, normalized durations for congruent, incongruent, synchronous, and baseline conditions were 1.261 (0.044), 0.962 (0.047), 1.078 (0.041), and 1.008 (0.034), respectively. The main effect of the temporal structure was significant,  $F_{(3, 45)} = 14.448$ ,  $p < 0.001$ . Bonferroni adjusted pairwise analysis showed the duration in the congruent condition (1.261) was significantly longer than the ones in the synchronous (1.078) and baseline (1.008) conditions, while the normalized duration in the incongruent condition (0.962) was significantly lower than the ones in the synchronous and baseline conditions,  $p < 0.05$ . For the inverted



condition, the durations of the perceived dominant direction for congruent, incongruent, synchronous and baseline conditions were 0.942 (0.044), 1.033 (0.047), 0.939 (0.041), and 0.944 (0.034), respectively. In contrast to the results for the upright posture, however, the inputs for tactile stimuli imposed no noticeable influence upon the perceived dominant motion direction of PLWs,  $F_{(3, 45)} = 0.907$ ,  $p = 0.436$ . This is shown in **Figure 4**).

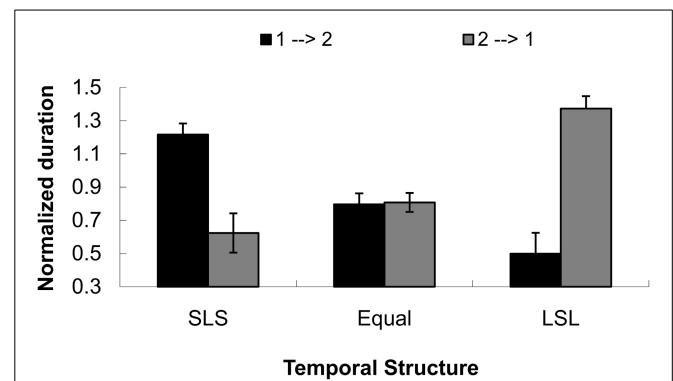
In light of these results, it appears that the temporal structure of tactile stimuli resolved the ambiguity of perceived dominant direction information for the binocular PLWs. However, to obtain the modulation effect from the tactile feedback, the PLWs should take on upright postures, which resemble the normal stance for walking people and suggest ecological constraints during cross-modal influence. This will be addressed in more detail in the Discussion section.

Sixteen additional subjects from the same population (undergraduate students, 8 female, from Peking University, aged 18–23 years) participated in a control experiment to judge the dominant direction of tactile apparent motion in the absence of visual stimuli. The mean normalized duration for the direction that went from the initial tap to the second tap (i.e., 1→2) was 0.837(0.048), and for the direction that went from the second tap to the initial tap was 0.935(0.051). The main effect of direction was not significant,  $F_{(1, 15)} = 1.634$ ,  $p = 0.221$ . The mean durations for SLS (short-long-short), equal (equal temporal intervals), and LSL (long-short-long) were 0.920(0.051), 0.802(0.032), and 0.936(0.042), respectively. The main effect of temporal condition was significant,  $F_{(2, 30)} = 4.336$ ,  $p < 0.05$ . Bonferroni-corrected pairwise comparison showed the mean duration for the equal condition (0.802) was shorter than for the mean duration for LSL (0.936). Importantly, the interaction between direction and temporal condition was significant,  $F_{(2, 30)} = 19.418$ ,  $p < 0.001$ . Further simple effects analysis with multivariate analysis of variance (MANOVA) indicated that the two perceived directions (1→2 and 2→1) were significantly different in the two SLS and LSL conditions,  $F_{(1, 15)} = 12.97$ ,  $p < 0.01$  and  $F_{(1, 15)} = 21.70$ ,  $p < 0.001$ . However, there was no difference in the equal condition,  $F < 1$ , as shown in **Figure 5**.

The results indicated that the capture of visual apparent motion in PLWs could mainly be based on the information of the perceived dominant direction of tactile apparent motion, which captures the directional perception of PLWs.

## EXPERIMENT 2

The walking direction (leftwards vs. rightwards) in Experiment 1 as a means of horizontal movement is seldom observed in real life situations. Therefore, in Experiment 2, we adopted receding/approaching walking postures to simulate the more common daily walking style. In addition, in order to better simulate the natural somatosensory perception related to walking, we moved the tactile stimuli from the fingertips to the ankles. In Experiment 2 we were interested in how the social-cognitive factor of empathy modulates the cross-modal (tactile-visual) temporal dynamic capture of the perceived direction of PLWs.



**FIGURE 5 | Normalized duration for dominant directional perception in three temporal structures (short-long-short, equal interval and long-short) for a control test to Experiment 1.** The directions were defined as from the initial tap to the second tap (1→2) or from the second tap to the initial tap (2→1). SLS indicates the temporal structure of short-long-short, equal means equal temporal intervals, and LSL shows the temporal structure of long-short-long intervals.

## METHOD

### Participants

Twenty-six undergraduates (ten female) from Peking University, aged 19–24 years, who met the same requirements of Experiment 1 participated in this experiment. The experiment was performed in compliance with all institutional guidelines set by the Academic Affairs Committee of the Department of Psychology at Peking University. All participants provided written informed consent according to institutional guidelines and the Declaration of Helsinki. Participants were reimbursed at a 20RMB/hour rate.

### Stimuli, apparatus, and procedure

The same apparatus and tactile stimuli of Experiment 1 were used in Experiment 2, except that the tactile actuators were attached to the front and back side of the ankle area, rather than on the fingertips. Two taps were put on the back of the two ankles while another two vibrators were put on the front of the ankles. All the PLWs took upright postures.

For the tactile stimuli, four stimuli were presented, with two attached to each ankle, either on the front or the back side of it. Tactile stimuli on the same side (e.g., front) were always presented at the same time, but the time interval between front and back side taps was manipulated with the same temporal structures as in Experiment 1. The tactile stimuli used in this study could simply be seen as the tactile stimuli used in Experiment 1, but rotated horizontally to the vertical motion, by attaching the tactile stimuli to each of the ankles. Participants were informed that while they could perceive the directional information of the tactile stimuli, the taps were irrelevant for determining the directions (receding vs. approaching) of the PLWs.

To render the binocular visual stimuli, two red and green PLWs were displayed on both the left and the right half of the screen and adjusted with a minor angular rotation (7° disparity) relative to its vertical location. Doing so ensured that the walking direction of the PLW on the left visual field was 97° while that of the PLW on the right visual field was 83° (in reference to

the right-hand X-axis for both). Note that the walking direction of the PLW appeared either facing away from (receding) or toward (approaching) the participants, as shown in **Figure 6**. These settings guaranteed the ambiguous nature of the apparent motion for the PLWs, and that for the given time period (70 s, with the same recording method as in Experiment 1), the participants could report their subjective dominant perception of the PLWs: either receding from or approaching themselves. The data was recorded by pressing and holding down two buttons of a custom-made response box (interfaced with a parallel port of the computer).

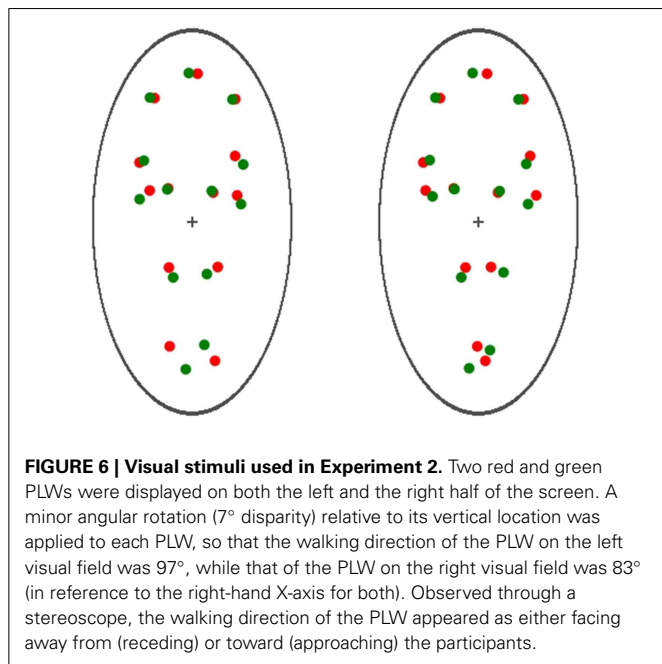
Similarly, we would expect that the temporal organization of tactile motion *per se* contributes to the observed cross-modal dynamic capture effect. A baseline task was implemented after the experiment to examine the effect of the temporal structure of the tactile stimuli upon the perceived dominant direction (receding vs. approaching) of the tactile apparent motion.

After the behavioral experiment, we asked the participants to fill in the Interpersonal Reactivity Index scale (Chinese version, IRI-C) (Rong et al., 2010), which includes four sub-scales of perspective-taking (PT), fantasy (FS), empathic concern (EC), and personal distress (PD); see the IRI-C is presented in the Supplementary Material. Based on the scores and according to common practice as described in above literature, we separated the individuals into two groups: a higher empathy group (with higher scores) and a lower empathy group (with lower scores), according to the above the median and below the median value of the scores (IRI  $\geq$  39, high empathy group; and IRI  $\leq$  38, low empathy group; 38 was the median).

## RESULTS

### CROSS-MODAL TEMPORAL CAPTURE EFFECT

The mean normalized durations for congruent, incongruent, synchronous, and baseline conditions were 1.402(0.076),



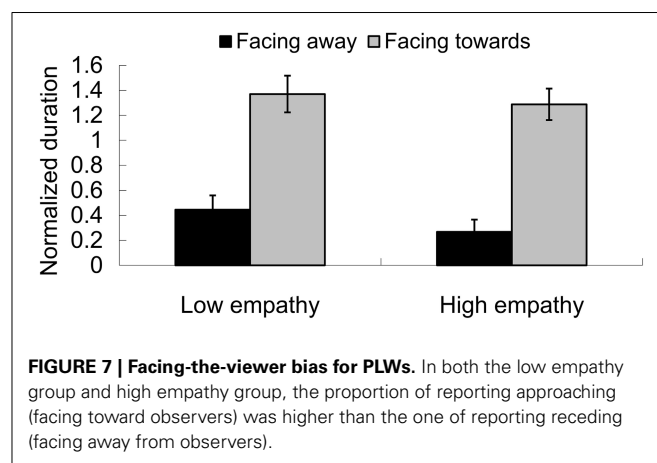
0.694(0.046), 0.942(0.049), and 1.067(0.038), respectively. A repeated measures ANOVA with temporal congruency as the independent variable showed a significant main effect of congruency,  $F_{(3, 75)} = 24.16$ ,  $p < 0.001$ . Bonferroni-corrected pairwise analysis showed that the duration for the congruent condition (1.402) was longest ( $p$ 's  $< 0.01$ ) and the duration for the incongruent condition (0.694) was shortest ( $p$ 's  $< 0.05$ ) among the four temporal structures. However, the duration for the synchronous condition (0.942) was statistically equal to the one in the baseline condition (1.067),  $p > 0.05$ . This result pattern suggests a significant impact of the cross-modal temporal structures on the perceived dominance of directional information for PLWs, just as we observed in Experiment 1.

### BASELINE TESTS: FACING-THE-VIEWER BIAS AND PERCEIVED DIRECTION FOR TACTILE APPARENT MOTION

In the visual-only condition, the normalized duration for a receding perception (facing away from the observer) was 0.356 (0.076) and for an approaching perception (facing toward the observer) was 1.329 (0.097),  $F_{(1, 24)} = 54.539$ ,  $p < 0.001$ . Therefore, a facing-the-viewer bias was manifested. This replicates several studies reported on in the literature (Vanrie et al., 2004; Brooks et al., 2008; Miller and Saygin, 2013; Van de Cruys et al., 2013; Heenan and Troje, 2014). However, there was no main effect of group. The mean duration for the low empathy group was 0.907 (0.086) and 0.778 (0.073),  $F_{(1, 24)} = 1.311$ ,  $p = 0.264$ . Also, there was no interaction effect between group and direction,  $F_{(1, 24)} = 0.129$ ,  $p = 0.722$ , as shown in **Figure 7**.

An additional control test (14 participants from Peking University, aged from 18 to 24 years old) discriminating the perceived direction of tactile apparent motion) showed that indeed, the temporal (interval) structure between tactile events caused a subjective bias of the perceived dominant direction of tactile apparent motion. The main effect of direction was not significant,  $F_{(1, 13)} = 3.476$ ,  $p = 0.085$ . The main effect of temporal condition was also not significant,  $F_{(2, 26)} = 1.463$ ,  $p = 0.250$ . The interaction between direction and temporal condition, however, was significant,  $F_{(2, 26)} = 13.952$ ,  $p < 0.001$ .

Further, simple effects analysis with MANOVA indicated that the two perceived directions (1 $\rightarrow$ 2 and 2 $\rightarrow$ 1) were significantly



different in the two SLS and LSL conditions,  $F_{(1, 13)} = 7.23$ ,  $p < 0.05$  and  $F_{(1, 13)} = 18.19$ ,  $p < 0.01$ , but not significantly different in the Equal condition,  $F < 1$ , as shown in **Figure 8**. This result pattern replicated the findings of the control test in Experiment 1, showing that the temporal structures between tactile events could lead to a dominant directional perception that gives rise to a capture effect in visual motion.

### THE INDIVIDUAL DIFFERENCE OF HIGH OR LOW EMPATHY

We compared the performance of two groups (high empathy vs. low empathy). In the incongruent condition, a group difference was observed. Individuals with high empathy had a shorter normalized dominant duration 0.604 (0.054) than those with low empathy, with a mean duration of 0.818 (0.063),  $F_{(1, 25)} = 6.595$ ,  $p < 0.05$ , as shown in **Figure 9**. This result pattern indicates that high empathy individuals were more readily captured by the tactile input. The tactile capture effect was shown mainly in the incongruent condition, in which the incongruent temporal structure between tactile events and biological motion somehow inhibited the perceived dominant directional information for PLWs.

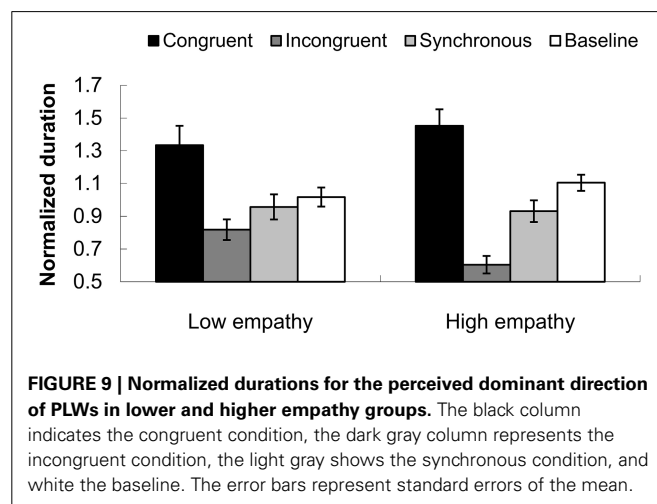
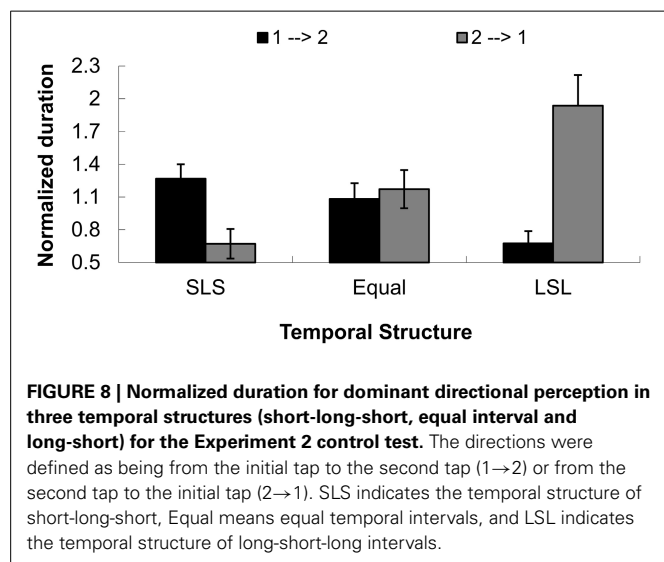
The variances of the mean durations could also be used to measure the tactile capture effect on visual perception. The mean standard deviations for congruent, incongruent, synchronous, and baseline conditions were 1.143(0.071), 1.936(0.096), 1.550(0.067), and 1.608(0.062), respectively. The main effect of condition was significant,  $F_{(3, 72)} = 21.175$ ,  $p < 0.001$ . Bonferroni-corrected pairwise comparisons showed that while there was no significant difference between synchronous (1.550) and baseline (1.608) conditions, the differences among the other cohorts were significant ( $p$ 's  $< 0.05$ ). The group effect was not significant,  $F_{(1, 24)} = 0.004$ ,  $p = 0.640$ . However, the interaction between temporal conditions and group was significant,  $F_{(3, 72)} = 21.175$ ,  $p < 0.001$ . Further analysis using a One-Way ANOVA indicated that on the dimension of congruency, the variance was lower for the higher empathy group (1.014) than the variance for the lower empathy group (1.319),

$F_{(1, 25)} = 5.196$ ,  $p < 0.05$ . This shows that for higher empathy individuals, the tactile capture effect was relatively stable in the congruency condition.

### DISCUSSION AND CONCLUSION

In this study, we revealed that the perception of directional information for PLWs under binocular rivalry conditions could be resolved by using tactile inputs, which simulate the tactile feedback of visual footsteps hitting the ground. By systematically manipulating the temporal intervals between tactile and visual events, we first extended the cross-modal dynamic capture effect from the visual-auditory domain to the visual-tactile domain, using PLWs. Specifically, when the walking pace signaled by the tactile stimuli were temporally congruent with the visual PLWs, the temporal structure facilitated the dominant directional perception—either dominant leftwards/rightwards movement (Experiment 1) or dominant receding/approaching movement (Experiment 2), with increased normalized durations. However, when the temporal structure of tactile feedback was incongruent with the visual footsteps, the perceived dominant directional information was inhibited with reduced normalized durations. *Post-hoc* observations and control tests indicated that the observers had on chance level to report the temporal synchronies with 150 ms between the tactile stimuli and visual footsteps, suggesting that the temporal dynamic capture effect was largely genuine perceptual processing.

The capture effect was larger for the congruent condition, rather than the temporally synchronous condition. This result pattern was in agreement with some previous studies on cross-modal temporal dynamic capture (Freeman and Driver, 2008; Shi et al., 2010). The results for the control test of discerning the dominant direction of tactile apparent motion in the absence of visual events indicate that the cross-modal dynamic capture effect was mainly driven by the perceived directional information of tactile events. In the unisensory modality (the tactile modality), the variation in temporal intervals between tactile inputs caused a potent directional perception of tactile motion (leftwards/rightwards in Experiment 1, and facing toward/away



in Experiment 2), which further captured the perceived dominant direction of the PLWs. During the visual-tactile interaction, the intra-modality perceptual grouping might precede the cross-modal (visual vs. tactile) binding process to produce the capture effect (Keetels et al., 2007; Cook and Van Valkenburg, 2009; Roseboom et al., 2013). The capture effect was not shown in the “synchronous” condition, which was seemingly contradictory to the findings that use other paradigm such as visual Ternus apparent motion (Shi et al., 2010). For example, in Shi et al. (2010) the two tones synchronously paired with two visual frames would change the observers’ categorization of motion percept (more “group motion” vs. “element motion”). Those differential findings are probably due to the differential tasks involved in different research paradigms. The current study used directional information of long-range apparent motion for probe, the capture effect stems from the build-up of the perceived temporal structure based on the varied temporal intervals (Freeman and Driver, 2008; Chen et al., 2011), which is absent in the “synchronous” condition. Therefore, we did not observe, if any, noticeable cross-modal capture effect when visual and tactile events were synchronous.

The cross-modal capture effect was observed in the upright visual configurations rather than in the inverted configurations, suggesting that cross-modal temporal capture is orientation specific (Pavlova and Sokolov, 2000), and that the sociobiological meaning (normal upright posture) of the biological motion is very important for detecting PLWs (Watson et al., 2004). This ecological constraint of perceiving PLWs was also shown in other studies (Cutting et al., 1988; Mather et al., 1992; Bertenthal and Pinto, 1994; Neri et al., 1998; Thornton, 1998). Pavlova and Sokolov (2000) reported an abrupt improvement in recognition of point-light walkers when the orientation changed from inverted to upright. These researchers used masking and priming procedures to investigate how display orientation affects recovery of a known point-light figure and found a high sensitivity to a camouflaged point-light walker with an upright orientation. A priming effect in biological motion was observed only if a prime corresponded to a range of deviations from the upright orientation within which the display was spontaneously recognizable. In their masking and priming paradigms, the recovery of a coherent structure is connected primarily with top-down processing of biological motion. However, their results indicated that orientation influences bottom-up processing of biological motion and influences top-down processing less. In Experiment 1 of our study, ecological constraints in perceiving PLWs were also shown. Here, the cross-modal capture effect on PLWs was observed with the upright posture, but not with the inverted posture.

We further showed that the capture pattern was modulated by empathy. Generally, high empathy individuals were more readily influenced by tactile inputs, with the characteristic capture effect in the incongruent condition. That is, high empathy group showed decreased normalized duration in the incongruent condition, compared to the low empathy group. High empathy individuals also demonstrated relatively stable performance with small variance (standard deviations) for the normalized duration in the congruent condition. These results suggest that multisensory

interaction can be modulated by an individual’s cognitive traits, and conform to an unwritten social norm. This effect might arise in people with high anxiety, as mistaking an approaching person for someone who is receding might have more severe consequences than the opposite mistake (Van de Cruys et al., 2013; Weech et al., 2014). People with higher empathic concern might be more sensitive to the direction of conflicting sensory cues (as in the incongruent condition), so as to avoid a potential mistake, like those in the high-anxiety group just mentioned. With the enhanced shared (mirror) touch experience of the first-person (the participant) and the third person (the PLWs), people with higher empathic concern could better exploit the vicarious somatosensory responses for simple touch and be more sensitive to others’ situations, including suffering (Banissy and Ward, 2007). In the current experimental scenario, the modulation arising from the factor of individual differences magnifies the difference of the temporal ventriloquism effect (tactile captures visual) between the high empathy group and low empathy group.

Other researchers have also recently found that individual differences in cognitive traits can influence the perception of PLWs. For example, with respect to ambiguous visual stimuli, more anxious individuals display a bias toward perceiving a more threatening image compared to those who are less anxious (Fox et al., 2002; Gray et al., 2009; Singer et al., 2012; Van de Cruys et al., 2013; Heenan and Troje, 2014). Heenan and Troje (2014) presented data to support that the facing-the-viewer bias is influenced at least in part by the social relevance of biological motion stimuli. Individuals with high anxiety level demonstrate a higher degree of facing-the-viewer bias than individuals with low anxiety. Evidence from the clinical field has shown that people with higher levels of Autism Spectrum Disorder have impaired global, but compensatory local, biological motion processing (van Boxtel and Lu, 2013). The studies cited have shown that personal cognitive/emotional states, whether in normally developing or atypically developing groups, could shape the perception of PLWs. Our study provides further evidence to support the idea that social-cognitive abilities can effectively modulate the otherwise ambiguous perception of point-light walkers. However, there might be individual differences in the ability to complete tasks that rely more heavily on the use of different cues in biological motion (form vs. motion and translational cues) (Rybarczyk and Santos, 2006; Wang et al., 2010; Miller and Saygin, 2013). Moreover, further study should aim to elucidate the intricate mechanisms underlying how individual differences modulate cross-modal interaction, as we have observed with the paradigm of PLWs.

Taken together, the above evidence suggests that tactile input helped to resolve the otherwise ambiguous perception of biological motion, and that this cross-modal effect is modulated by higher level social-cognitive factors, such as empathic concern.

## ACKNOWLEDGMENTS

We are grateful to Professor Denis Pelli (NYU) for valuable comments on an early draft. This study was supported by grants from the Natural Science Foundation of China (31200760), National



High Technology Research and Development Program of China (863 Program) (2012AA011602) and Fund for fostering talents in basic science (J1103602).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2015.00161/abstract>

## REFERENCES

- Ahlstrom, V., Blake, R., and Ahlstrom, U. (1997). Perception of biological motion. *Perception* 26, 1539–1548. doi: 10.1068/p261539
- Arrighi, R., Marini, F., and Burr, D. (2009). Meaningful auditory information enhances perception of visual biological motion. *J. Vis.* 9, 25–27. doi: 10.1167/9.4.25
- Banissy, M. J., and Ward, J. (2007). Mirror-touch synaesthesia is linked with empathy. *Nat. Neurosci.* 23, 145–152. doi: 10.1038/nn1926
- Barclay, C. D., Cutting, J. E., and Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Percept. Psychophys.* 23, 145–152. doi: 10.3758/BF03208295
- Beauchamp, M. S., Lee, K. E., Haxby, J. V., and Martin, A. (2003). fMRI responses to video and point-light displays of moving humans and manipulable objects. *J. Cogn. Neurosci.* 15, 991–1001. doi: 10.1162/089892903770007380
- Bertelson, P., and Aschersleben, G. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgment. *Int. J. Psychophysiol.* 50, 147–155. doi: 10.1016/S0167-8760(03)00130-2
- Bertenthal, B. I., and Pinto, J. (1994). Global processing of biological motions. *Psychol. Sci.* 5, 221–225. doi: 10.1111/j.1467-9280.1994.tb00504.x
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357
- Brooks, A., Schouten, B., Troje, N. F., Verfaillie, K., Blanke, O., and van der Zwan, R. (2008). Correlated changes in perceptions of the gender and orientation of ambiguous biological motion figures. *Curr. Biol.* 18, R728–R729. doi: 10.1016/j.cub.2008.06.054
- Brooks, A., van der Zwan, R., Billard, A., Petreska, B., Clarke, S., and Blanke, O. (2007). Auditory motion affects visual biological motion processing. *Neuropsychologia* 45, 523–530. doi: 10.1016/j.neuropsychologia.2005.12.012
- Cattaneo, L., and Rizzolatti, G. (2009). The mirror neuron system. *Arch. Neurol.* 66, 557–560. doi: 10.1001/archneurol.2009.41
- Chen, L., Shi, Z., and Muller, H. J. (2011). Interaction of perceptual grouping and crossmodal temporal capture in tactile apparent-motion. *PLoS ONE* 6:e17130. doi: 10.1371/journal.pone.0017130
- Chen, L., and Vroomen, J. (2013). Intersensory binding across space and time: a tutorial review. *Atten. Percept. Psychophys.* 75, 790–811. doi: 10.3758/s13414-013-0475-4
- Cook, L. A., and Van Valkenburg, D. L. (2009). Audio-visual organisation and the temporal ventriloquism effect between grouped sequences: evidence that unimodal grouping precedes cross-modal integration. *Perception* 38, 1220–1233. doi: 10.1068/p6344
- Cutting, J. E. (1978). A program to generate synthetic walkers as dynamic point-light displays. *Behav. Res. Methods Instrum.* 10, 91–94. doi: 10.3758/BF03205105
- Cutting, J. E., and Kozlowski, L. T. (1977). Recognition of friends by their walk: gait perception without familiarity cues. *Bull. Psychon. Soc.* 9, 353–356. doi: 10.3758/BF03337021
- Cutting, J. E., Moore, C., and Morrison, R. (1988). Masking the motions of human gait. *Percept. Psychophys.* 44, 339–347. doi: 10.3758/BF03210415
- Das, S. R., Lazarewicz, M. T., Wilson, R. C., and Finkel, L. H. (2009). Sensitivity to motion features in point light displays of biological motion. *Spat. Vis.* 22, 105–125. doi: 10.1163/156856809787465627
- de Lussanet, M. H., and Lappe, M. (2012). Depth perception from point-light biological motion displays. *J. Vis.* 12:14. doi: 10.1167/12.11.14
- Fox, E., Russo, R., and Dutton, K. (2002). Attentional bias for threat: evidence for delayed disengagement from emotional faces. *Cogn. Emot.* 16, 355–379. doi: 10.1080/02699930143000527
- Freeman, E., and Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. *Curr. Biol.* 18, 1262–1266. doi: 10.1016/j.cub.2008.07.066
- Gallace, A., and Spence, C. (2010). The science of interpersonal touch: an overview. *Neurosci. Biobehav. Rev.* 34, 246–259. doi: 10.1016/j.neubiorev.2008.10.004
- Gallace, A., and Spence, C. (2011). To what extent do Gestalt grouping principles influence tactile perception? *Psychol. Bull.* 137, 1174–1181. doi: 10.1037/a0022335
- Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn. Sci. (Regul. Ed.)* 8, 396–403. doi: 10.1016/j.tics.2004.07.002
- Gray, K. L., Adams, W. J., and Garner, M. (2009). The influence of anxiety on the initial selection of emotional faces presented in binocular rivalry. *Cognition* 113, 105–110. doi: 10.1016/j.cognition.2009.06.009
- Heenan, A., and Troje, N. F. (2014). Both physical exercise and progressive muscle relaxation reduce the facing-the-viewer bias in biological motion perception. *PLoS ONE* 9:e99902. doi: 10.1371/journal.pone.0099902
- Henry, J. D., Thompson, C., Rendell, P. G., Phillips, L. H., Carbert, J., Sachdev, P., et al. (2012). Perception of biological motion and emotion in mild cognitive impairment and dementia. *J. Int. Neuropsychol. Soc.* 18, 866–873. doi: 10.1017/S1355617712000665
- Herrington, J. D., Nymberg, C., and Schultz, R. T. (2011). Biological motion task performance predicts superior temporal sulcus activity. *Brain Cogn.* 77, 372–381. doi: 10.1016/j.bandc.2011.09.001
- Hirai, M., and Hiraki, K. (2005). An event-related potentials study of biological motion perception in human infants. *Brain Res. Cogn. Brain Res.* 22, 301–304. doi: 10.1016/j.cogbrainres.2004.08.008
- Hirai, M., Watanabe, S., Honda, Y., and Kakigi, R. (2009). Developmental changes in point-light walker processing during childhood and adolescence: an event-related potential study. *Neuroscience* 161, 311–325. doi: 10.1016/j.neuroscience.2009.03.026
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14, 195–204. doi: 10.3758/BF03212378
- Johnson, K. L., McKay, L. S., and Pollick, F. E. (2011). He throws like a girl (but only when he's sad): emotion affects sex-decoding of biological motion displays. *Cognition* 119, 265–280. doi: 10.1016/j.cognition.2011.01.016
- Keetels, M., Stekelenburg, J., and Vroomen, J. (2007). Auditory grouping occurs prior to intersensory pairing: evidence from temporal ventriloquism. *Exp. Brain Res.* 180, 449–456. doi: 10.1007/s00221-007-0881-8
- Kim, J., Kroos, C., and Davis, C. (2010). Hearing a point-light talker: an auditory influence on a visual motion detection task. *Perception* 39, 407–416. doi: 10.1068/p6483
- Ma, Y., Paterson, H. M., and Pollick, F. E. (2006). A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behav. Res. Methods* 38, 134–141. doi: 10.3758/BF03192758
- Mather, G., Radford, K., and West, S. (1992). Low-level visual processing of biological motion. *Proc. Biol. Sci.* 249, 149–155. doi: 10.1098/rspb.1992.0097
- Miller, L. E., and Saygin, A. P. (2013). Individual differences in the perception of biological motion: links to social cognition and motor imagery. *Cognition* 128, 140–148. doi: 10.1016/j.cognition.2013.03.013
- Morein-Zamir, S., Soto-Faraco, S., and Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Brain Res. Cogn. Brain Res.* 17, 154–163. doi: 10.1016/S0926-6410(03)00089-2
- Neri, P., Morrone, M. C., and Burr, D. C. (1998). Seeing biological motion. *Nature* 395, 894–896. doi: 10.1038/27661
- Norman, J. F., Payton, S. M., Long, J. R., and Hawkes, L. M. (2004). Aging and the perception of biological motion. *Psychol. Aging* 19, 219–225. doi: 10.1037/0882-7974.19.1.219
- Pavlova, M. A., and Sokolov, A. N. (2000). Orientation specificity in biological motion perception. *Percept. Psychophys.* 62, 889–899. doi: 10.3758/BF03212075
- Pavlova, M. A., Sokolov, A. N., and Bidet-Ildes, C. (2014). Sex differences in the neuromagnetic cortical response to biological motion. *Cereb. Cortex*. doi: 10.1093/cercor/bhu175. [Epub ahead of print].
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366
- Pollick, F. E., Kay, J. W., Heim, K., and Stringer, R. (2005). Gender recognition from point-light walkers. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 1247–1265. doi: 10.1037/0096-1523.31.6.1247

- Rong, X., Sun, B., Huang, X., Cai, M., and Li, W. (2010). Reliabilities and validities of chinese version of interpersonal reactivity index. *Chin. J. Clin. Psychol.* 2, 158–160.
- Roseboom, W., Kawabe, T., and Nishida, S. (2013). Direction of visual apparent motion driven by perceptual organization of cross-modal signals. *J. Vis.* 13:6. doi: 10.1167/13.1.6
- Rybarczyk, Y., and Santos, J. (2006). “Direction perception of a translational point-light-walker,” in *Encontro da Associação Portuguesa de Psicologia Experimental (Aveiro)*.
- Saygin, A. P., Driver, J., and de Sa, V. R. (2008). In the footsteps of biological motion and multisensory perception: judgments of audiovisual temporal relations are enhanced for upright walkers. *Psychol. Sci.* 19, 469–475. doi: 10.1111/j.1467-9280.2008.02111.x
- Schaefer, M., Heinze, H. J., and Rotte, M. (2012). Embodied empathy for tactile events: interindividual differences and vicarious somatosensory responses during touch observation. *Neuroimage* 60, 952–957. doi: 10.1016/j.neuroimage.2012.01.112
- Schmuckler, M. A., and Fairhall, J. L. (2001). Visual-proprioceptive intermodal perception using point light displays. *Child Dev.* 72, 949–962. doi: 10.1111/j.1467-8624.00327
- Servos, P., Osu, R., Santi, A., and Kawato, M. (2002). The neural substrates of biological motion perception: an fMRI study. *Cereb. Cortex* 12, 772–782. doi: 10.1093/cercor/12.7.772
- Shi, Z., Chen, L., and Muller, H. J. (2010). Auditory temporal modulation of the visual Ternus effect: the influence of time interval. *Exp. Brain Res.* 203, 723–735. doi: 10.1007/s00221-010-2286-3
- Silani, G., Lamm, C., Ruff, C. C., and Singer, T. (2013). Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. *J. Neurosci.* 33, 15466–15476. doi: 10.1523/JNEUROSCI.1488-13.2013
- Singer, N., Eapen, M., Grillon, C., Ungerleider, L. G., and Hendler, T. (2012). Through the eyes of anxiety: dissecting threat bias via emotional-binocular rivalry. *Emotion* 12, 960–969. doi: 10.1037/a0027070
- Slutsky, D. A., and Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport* 12, 7–10. doi: 10.1097/00001756-200101220-00009
- Thomas, J. P., and Shiffrar, M. (2010). I can see you better if I can here you coming: action-consistent sounds facilitate the visual detection of human gait. *J. Vis.* 10, 1–11. doi: 10.1167/10.12.14
- Thornton, I. M. (1998). The visual perception of human locomotion. *Cogn. Neuropsychol.* 15, 535–552. doi: 10.1080/026432998381014
- Troje, N. F., Sadr, J., Geyer, H., and Nakayama, K. (2006). Adaptation aftereffects in the perception of gender from biological motion. *J. Vis.* 6, 850–857. doi: 10.1167/6.8.7
- Troje, N. F., Westhoff, C., and Lavrov, M. (2005). Person identification from biological motion: effects of structural and kinematic cues. *Percept. Psychophys.* 67, 667–675. doi: 10.3758/BF03193523
- van Boxtel, J. J., and Lu, H. (2013). A biological motion toolbox for reading, displaying, and manipulating motion capture data in research settings. *J. Vis.* 13:7. doi: 10.1167/13.12.7
- Van de Cruys, S., Schouten, B., and Wagemans, J. (2013). An anxiety-induced bias in the perception of a bistable point-light walker. *Acta Psychol. (Amst)*. 144, 548–553. doi: 10.1016/j.actpsy.2013.09.010
- Vanrie, J., Dekeyser, M., and Verfaillie, K. (2004). Bistability and biasing effects in the perception of ambiguous point-light walkers. *Perception* 33, 547–560. doi: 10.1068/p5004
- Vroomen, J., de Gelder, B., and Vroomen, J. (2004). Temporal ventriloquism: sound modulates the flash-lag effect. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 513–518. doi: 10.1037/0096-1523.30.3.513
- Wang, L., Zhang, K., He, S., and Jiang, Y. (2010). Searching for life motion signals. Visual search asymmetry in local but not global biological-motion processing. *Psychol. Sci.* 21, 1083–1089. doi: 10.1177/0956797610376072
- Watson, T. L., Pearson, J., and Clifford, C. W. (2004). Perceptual grouping of biological motion promotes binocular rivalry. *Curr. Biol.* 14, 1670–1674. doi: 10.1016/j.cub.2004.08.064
- Weech, S., McAdam, M., Kenny, S., and Troje, N. F. (2014). What causes the facing-the-viewer bias in biological motion? *J. Vis.* 14, 1–15. doi: 10.1167/14.12.10

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 04 December 2014; accepted: 01 February 2015; published online: 20 February 2015.

Citation: Yiltiz H and Chen L (2015) Tactile input and empathy modulate the perception of ambiguous biological motion. *Front. Psychol.* 6:161. doi: 10.3389/fpsyg.2015.00161

This article was submitted to *Cognition*, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Yiltiz and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Corrigendum: Tactile input and empathy modulate the perception of ambiguous biological motion

Hörmetjan Yiltiz<sup>1</sup> and Lihan Chen<sup>1,2\*</sup>

<sup>1</sup> Department of Psychology, Peking University, Beijing, China, <sup>2</sup> Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China

**Keywords:** tactile, point-light walker, temporal, empathy, apparent motion, binocular rivalry

## A corrigendum on

**Tactile input and empathy modulate the perception of ambiguous biological motion** by Yiltiz, H., and Chen, L. (2015). *Front. Psychol.* 6:161. doi: 10.3389/fpsyg.2015.00161

The citation of:

Weech, S., McAdam, M., and Troje, N. F. (2014). What causes the facing-the-viewer bias in biological motion? *J. Vis.* 14, 1–15. doi: 10.1167/14.12.10

should be corrected as:

Weech, S., McAdam, M., Kenny, S., and Troje, N. F. (2014). What causes the facing-the-viewer bias in biological motion? *J. Vis.* 14, 1–15. doi: 10.1167/14.12.10

The original article has been updated.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Yiltiz and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## OPEN ACCESS

### Edited and reviewed by:

Richard A. Abrams,  
Washington University, USA

### \*Correspondence:

Lihan Chen,  
clh@pku.edu.cn

### Specialty section:

This article was submitted to  
Cognition, a section of the journal  
*Frontiers in Psychology*

**Received:** 18 August 2015

**Accepted:** 28 August 2015

**Published:** 08 September 2015

### Citation:

Yiltiz H and Chen L (2015)  
Corrigendum: Tactile input and  
empathy modulate the perception of  
ambiguous biological motion.  
*Front. Psychol.* 6:1384.  
doi: 10.3389/fpsyg.2015.01384

## ADVANTAGES OF PUBLISHING IN FRONTIERS



### FAST PUBLICATION

Average 90 days  
from submission  
to publication



### COLLABORATIVE PEER-REVIEW

Designed to be rigorous –  
yet also collaborative, fair and  
constructive



### RESEARCH NETWORK

Our network  
increases readership  
for your article



### OPEN ACCESS

Articles are free to read,  
for greatest visibility



### TRANSPARENT

Editors and reviewers  
acknowledged by name  
on published articles



### GLOBAL SPREAD

Six million monthly  
page views worldwide



### COPYRIGHT TO AUTHORS

No limit to  
article distribution  
and re-use



### IMPACT METRICS

Advanced metrics  
track your  
article's impact



### SUPPORT

By our Swiss-based  
editorial team