УДК 004.934

А.В. САВЧЕНКО

РЕЗУЛЬТАТЫ НАТУРНЫХ ИСПЫТАНИЙ МЕТОДА ФОНЕТИЧЕСКОГО ДЕКОДИРОВАНИЯ СЛОВ В ЗАДАЧАХ РАСПОЗНАВАНИЯ И ДИАРИЗАЦИИ РАЗГОВОРНОЙ РУССКОЙ РЕЧИ

Представлены результаты натурных испытаний экспериментального образца программного комплекса фонетического декодирования слов на основе принципа минимума информационного рассогласования Кульбака-Лейблера в режимах распознавания и диаризации разговорной русской речи. Показано, что предлагаемая система характеризуется повышенными надежностью и быстродействием при распознавании как отдельных слов, так и целых фраз. Даны рекомендации по ее практическому применению в задачах голосового управления.

Ключевые слова: *автоматическое распознавание русской речи; диаризация речи; метод фонетического декодирования; принцип минимума информационного рассогласования.*

ВВЕДЕНИЕ

Задача построения надежных систем голосового управления [1] является одной из центральных по всем направлениям распространения автоматического распознавания речи (АРР) [2]. К сожалению, использование здесь канонического подхода [2, 3], основанного на скрытых марковских моделях речевых единиц и применяемого, например, в системах диктовки текста, не всегда позволяет получить удовлетворительное качество распознавания, особенно если требуется за короткое время перенастраивать систему на новое множество команд. При этом наиболее естественный способ повышения надежности [4] – настройка системы АРР под конкретного диктора – наталкивается на проблему вычислительной сложности.

Одним из перспективных с точки зрения вычислительных затрат на реализацию может служить техническое решение [5], предложенное в устройстве для фонетического анализа и распознавания речи в информационной метрике Кульбака-Лейблера [6] по методу фонетического декодирования слов (ФДС) [7]. К сожалению, данный метод пока недостаточно был апробирован в реальных условиях. Указанный пробел в какой-то степени устраняет настоящая статья, посвященная организации и результатам натурных испытаний метода ФДС. Для этого был разработан экспериментальный образец программного комплекса фонетического декодирования слов (ЭО ПК ФДС). Причем наряду с функцией АРР в нем осуществлена и автоматическая диаризация речи (АДР) [8], используемая для идентификации диктора по входному речевому сигналу.

МЕТОД ФОНЕТИЧЕСКОГО ДЕКОДИРОВАНИЯ СЛОВ

Пусть задано множество из $L>1$ эталонных команд $\{X_l\}$, где $l=\overline{1,L}$ – номер слова-эталона. Согласно общепринятому фонетическому подходу [2], каждая эталонная команда разбивается на последовательность фонем (транскрипцию) $X_l=\{c_{l,1},c_{l,2},...,c_{l,L_l}\}$. Здесь L_l – длительность команды (в фонемах), а числа $c_{l,j}\in\{1.,,,.R\}$ – номера фонем из некоторого фонетического алфавита $\{\mathbf{x}_r^*\}, r=\overline{1,R}$, где R – количество фонем в алфавите. Задача состоит в том, чтобы поступившему на вход речевому сигналу X с частотой дискретизации F (в герцах) поставить в соответствие наиболее близкое к нему слово-эталон.

Для решения задачи на первом этапе сигнал X разбивается на непересекающиеся сегменты $\{\mathbf{x}(t)\}, t=\overline{1,T}$ длиной $\tau=0{,}01\text{-}0{,}015$ сек, где T – общее число сегментов. Далее каждый парциальный сигнал $\mathbf{x}(t)=\|x_1(t)...x_M(t)\|$ (здесь $M=\tau\cdot F$) рассматривается в

пределах конечного списка гласных фонем $\left\{\mathbf{x}_r^*\right\}$ (т.е. ФБД состоит только из гласных звуков) и отождествляется с той из них, которая отвечает принципу минимума величины заданной исследователем меры близости между сигналом $\mathbf{x}(t)$ и эталоном $\mathbf{x}_r^*$.

Для выбора меры близости в (1) воспользуемся широко использующейся в АРР авторегрессионной (АР) моделью речевого сигнала на интервалах его квазистационарности $\tau \approx \text{const}$. Известно [11, 18], что в этом случае критерий, основанный на принципе минимума информационного рассогласования Кульбака-Лейблера [6] с решающей статистикой вида

$$\nu\left(t; \left\{\mathbf{x}_r^*\right\}\Big|_{r=\overline{1,R}}\right) = \underset{r \in \{1,\ldots,R\}}{\arg\min} \frac{1}{F} \sum_{f=1}^{F} \left(\frac{G_X(f)}{G_{\mathbf{x}_r^*}(f)} - \ln \frac{G_X(f)}{G_{\mathbf{x}_r^*}(f)} - 1 \right) \quad (1)$$

эквивалентен оптимальному методу максимального правдоподобия. Здесь $G_X(f)$ – выборочная оценка спектральной плотности мощности (СПМ) входного сигнала $\mathbf{x}(t)$ в функции дискретной частоты f, $G_{\mathbf{x}_r^*}(f)$ – СПМ эталона r-ой фонемы $\mathbf{x}_r^*$, F – верхняя граница частотного диапазона речевого сигнала или используемого канала связи. Оценка СПМ чаше всего производится на основе АР-модели [10] речевого сигнала, главное достоинство в задаче АРР которой [7] состоит в возможности предварительной нормировки речевых сигналов по дисперсиям их порождающих процессов. Такая нормировка обусловлена физическими особенностями голосового механизма человека: воздушный поток на входе его модели «акустической трубы» имеет приблизительно одну и ту же интенсивность на интервалах, длительностью в целое слово. Тогда отношение СПМ в (1) приобретает вид [7]

$$\frac{G_X(f)}{G_{\mathbf{x}_r^*}(f)} = \frac{\left|1 + \sum_{m=1}^{p} a_r(m) \exp\left(-j\pi m f / F\right)\right|^2}{\left|1 + \sum_{m=1}^{p} a_X(m) \exp\left(-j\pi m f / F\right)\right|^2},$$

где p – порядок АР-модели, $j = \sqrt{-1}$, а $a_r(m)$ и $a_X(r)$ – оценки АР-коэффициентов эталона $\mathbf{x}_r^*$ и входного сигнала $\mathbf{x}(t)$, получаемые на основе алгоритма Левинсона-Дурбина и метода Берга [10].

Важнейшее достоинство АР-модели в задачах АРР – это возможность нормировки речевых сигналов по дисперсии порождающих процессов: $\sigma_0^2 = \sigma_X^2$, где σ_X^2 – дисперсия порождающего процесса. Известно [7], что при учете этого асимптотически оптимальное решение дает основанный на принципе МИР критерий

$$\nu\left(t; \left\{\mathbf{x}_r^*\right\}\Big|_{r=\overline{1,R}}\right) = \underset{r \in \{1,\ldots,R\}}{\arg\min} \frac{1}{2} \left[\frac{\sigma_r^2(\mathbf{x})}{\sigma_0^2} - 1 \right], \quad (2)$$

где $\sigma_r^2(\mathbf{x})$ – выборочная оценка дисперсии отклика r-го обеляющего фильтра (ОФ) $\mathbf{y}_r(t) = \left\| y_{r;1}(t) \ldots y_{r;M-p}(t) \right\|$, где p – порядок АР-модели, а

$$y_{r;\,j}(t) = x_{j+p}(t) - \sum_{m=1}^{p} a_r(m) x_{j+p-m}(t), \quad j = \overline{1, M-p}. \quad (3)$$

На втором этапе полученная согласно (2) транскрипция сигнала X обычно выравнивается по темпу речи с транскрипцией каждого слова-эталона для установления временного соответствия между звуками сопоставляемых речевых образов. Для этого можно воспользоваться, например, алгоритмом Dynamic Time Warping, основанным на принципах динамического программирования, или вероятностным аппаратом скрытых Марковских моделей [2, 3]. Наиболее близкое в смысле среднего рассогласования вида (2) после временного выравнивания слово и будет являться решением задачи АРР.

АЛГОРИТМ РАСПОЗНАВАНИЯ И ДИАРИЗАЦИИ РЕЧИ

Мы предполагаем, что входное слово X разбито на N слогов, причем границы каждого n-го слога ($n=\overline{1,N}$) определены с точностью до номера квазистационарного сегмента $\left(t_n^{(1)}, t_n^{(2)}\right)$. Будем проводить распознавание только среди гласных звуков. Для этого выполним настройку системы голосового управления для всех U потенциальных пользователей (где U – количество различных пользователей). Такая настройка потребует лишь произнесения пользователями каждого гласного звука. В результате получим набор фонетических алфавитов $\left\{\mathbf{x}^*_{r;u}\right\}, r=\overline{1,R}, u=\overline{1,U}$.

Тогда на первом этапе для каждого сегмента $\mathbf{x}(t)$ вычисляются ближайшие к нему коды эталонов для каждого пользователя

$$\nu_u(t)=\nu\left(t;\left\{\mathbf{x}^*_{r;u}\right\}\Big|_{r=\overline{1,R}}\right), u=\overline{1,U}. \tag{4}$$

На втором этапе АРР на основе всех $\nu_u(t), t=\overline{t_n^{(1)}, t_n^{(2)}}, u=\overline{1,U}$ будем принимать решение в пользу принадлежности распознаваемого слога к одной из R гласных. Воспользуемся простым агрегированием – n-му слогу ставится в соответствие последовательность частот $\mu_n(r), r=\overline{1,R}$, где

$$\mu_n(r)=\frac{1}{t_n^{(2)}-t_n^{(1)}+1}\cdot\sum_{t=t_n^{(1)}}^{t_n^{(2)}}\delta\left(r-\arg\min_{\nu_u(t);u=\overline{1,U}}\rho\left(\mathbf{x}(t);\mathbf{x}^*_{\nu_u(t)}\right)\right), \tag{5}$$

где $\delta(x)$ – дискретная дельта-функция, $\rho\left(\mathbf{x}(t);\mathbf{x}^*_{\nu_u(t);u}\right)$ – рассогласование между $\mathbf{x}(t)$ и ближайшей к нему фонемой-эталоном u-го пользователя вида (1) или (2). Далее для каждого слова-эталона X_l оценивается его корреляция с распознаваемым речевым сигналом:

$$\mu(l)=\begin{cases}\sum_{n=1}^{N}\mu_n(c_{l,n}), & L_l=N\\ 0 & L_l\neq N\end{cases}. \tag{6}$$

Тогда решение задачи АРР принимается в пользу слова X^* по критерию максимума величины μ_l. Аналогично решается и задача идентификации диктора [9], являющаяся основой алгоритмов АДР [8] – вначале n-му слогу ставится в соответствие последовательность частот $\lambda_n(u), u=\overline{1,U}$:

$$\lambda_n(u)=\frac{1}{t_n^{(2)}-t_n^{(1)}+1}\cdot\sum_{t=t_n^{(1)}}^{t_n^{(2)}}\delta\left(u-\arg\min_{w=\overline{1,U}}\rho\left(\mathbf{x}(t);\mathbf{x}^*_{\nu_w(t)}\right)\right). \tag{7}$$

Решение задачи идентификации принимается в пользу диктора u^* с максимальной корреляцией по всему входному сигналу:

$$\lambda(u) = \begin{cases} \sum_{n=1}^{N} \lambda_n(u), & L_l = N \\ 0 & L_l \neq N \end{cases}. \quad (8)$$

Таким образом, система выражений (2-8) и определяет используемый нами подход к решению задач АРР и АДР.

ПРОГРАММА ЭКСПЕРИМЕНТАЛЬНОГО ИССЛЕДОВАНИЯ

Для тестирования описанных выше алгоритмов АРР и АДР был разработан ЭО ПК ФДС. В тестировании принимают участие два пользователя программы, которые не должны иметь явно выраженных дефектов речи, а также оператор-контролер. Его задача – контроль действий диктора; самостоятельного задания на тестирование он не имеет. Предварительным условием для проведения эксперимента является загрузка в ЭО ПК ФДС рабочего словаря из текстового файла [11], содержащего не более 10000 слов/словосочетаний и настройка ФБД [12] для всех дикторов, участвующих в тестировании. Диктор, произносящий слова в момент тестирования, выбирается в ЭО ПК ФДС как текущий пользователь.

Программа испытания включает в себя следующие этапы:

1. Испытание быстродействия алгоритма АРР;
2. Испытание качества АРР (2-6);
3. Испытание качества АДР (2-4), (7), (8).

Для тестирования первого пункта программы последовательно в течение 100 раз средствами ЭО ПК ФДС выполняется запись речевого сигнала с микрофона длительностью не более 10 сек. При этом время распознавания отображается справа в нижней части главного окна программы (рис. 1). Комплекс считается выполнившим проверку, если время распознавания на испытательном стенде не превысило 0,1 сек.

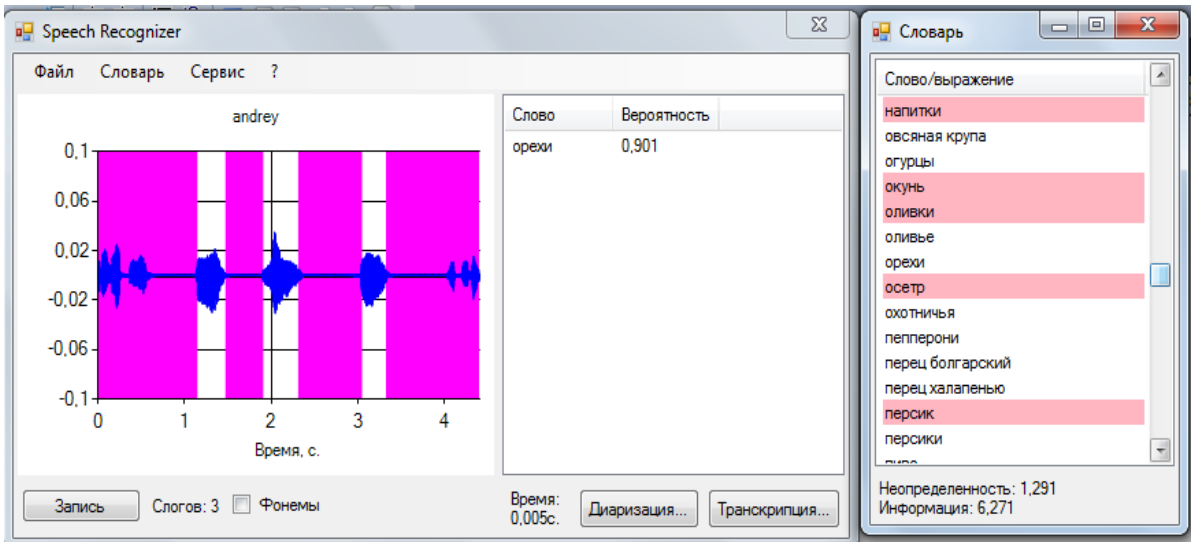

Рисунок 1 – Главное окно ЭО ПК ФДС

До окончания тестирования средствами ЭО ПК ФДС не менее 10 раз (для каждого слова) производится ввод с микрофона любого слова/словосочетания из рабочего словаря длительностью не менее трех слогов. Фрагменты акустического (речевого) сигнала (отдельные слова или целые фразы) проговорены диктором в относительно медленном

темпе, по слогам, с четкими паузами между слогами. Речевой сигнал, подаваемый на вход ЭО ПК ФДС, должен удовлетворять следующим требованиям:

а) отношение сигнал/шум не менее 20 дБ;

б) речевой сигнал не должен иметь выраженных нелинейных искажений (АРУ, клиппирование).

Программной компонентой испытательного стенда учитываются только те слова-попытки, которые отвечают требованиям к эталонному диктору: все слова/словосочетания из рабочего словаря должны проговариваться диктором в относительно медленном (не менее 200 мс на один слог) темпе, по слогам, с паузами между слогами не короче 70 мс., при автоматически детектированной системе количество слогов во входном акустическом сигнале должно совпадать с выбранным оператором-контроллером. Если попытка засчитывается, но произнесенное в микрофон слово среди наиболее вероятных эталонов отображаемых в главном окне системы, отсутствует, делается вывод об ошибке распознавания. В противном случае распознавание считается безошибочным. Далее, если идентифицированный системой пользователь не совпадает с текущим пользователем, делается вывод об ошибке АДР. Комплекс считается выдержавшим проверку, если по завершении всех операций средняя вероятность верного распознавания оказалась не ниже 90%, а средняя вероятность диаризации превысила 75%.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИСПЫТАНИЙ

Рассмотрим применение ЭО ПК ФДС в задаче распознавания голосовых команд. В качестве словаря эталонов возьмем список названий лекарств, продаваемых в одной аптеке города Н. Новгорода, состоящий из 1913 слов/словосочетаний на русском языке. Запись речевого сигнала осуществлялась через внешний микрофон с функцией шумоподавления из гарнитуры A4Tech HS-12. Записанный речевой сигнал (отдельные звуки и слова/словосочетания из словаря эталонов) сохранялся в виде отдельного звукового wav-файла (моно, частота дискретизации $F=8000$ Гц, 16 бит на отсчет). В качестве предварительной обработки из сигнала удалялись начальные и конечные паузы. Порядок АР-модели $p=20$. Выделенные слоги членились на последовательность пересекающихся сегментов длительностью τ=0,015 сек ($M=120$ отсчетов).

На предварительном этапе осуществлялась настройка системы под конкретного диктора. В режиме настройки диктор четко проговаривал каждый из 10 гласных звуков русского языка: «а», «е», «ё», «и», «о», «у», «ы», «э», «ю», «я». Для каждой фонемы-эталона АР-коэффициенты оценивались по всему сигналу целиком без его разделения на сегменты. Запись звука повторялась до тех пор, пока синтезированный АР-процесс по звучанию не становился близок к произнесенному диктором звуком. Близость оценивалась самим диктором «на слух». В среднем для каждого звука потребовалось 2-3 итерации. Среднее время настройки всей фонетической базы данных в расчете на одного диктора составило 2,75 минут (минимальное время настройки – 1,5 мин, максимальное – 5,1 мин).

Для сравнения быстродействия используемого алгоритма сопоставления минимальных звуковых единиц (2) с традиционными алгоритмами, основанными на сопоставлении спектральных плотностей мощности (СПМ) воспользуемся известной [6] формулировкой принципа МИР в частотной области (1). Для ускорения процедуры распознавания в (1) сопоставлялись не все значения СПМ для $f=\overline{1,F}$, а только частоты с шагом $\Delta f=10$ Гц. Для этого значения параметра скорость АРР в 10 раз превышает скорость распознавания (1), при этом качество распознавания остается практически неизменным и не отличается от точности критерия (1). Среднее время распознавания для критериев (1) и (2) в зависимости от числа слогов n во входном словосочетании показано на рисунке 2. Хорошо видно, что вычислительная эффективность адаптивного критерия (2) на порядок превышает аналогичный показатель для традиционного сопоставления отсчетов СПМ (1).

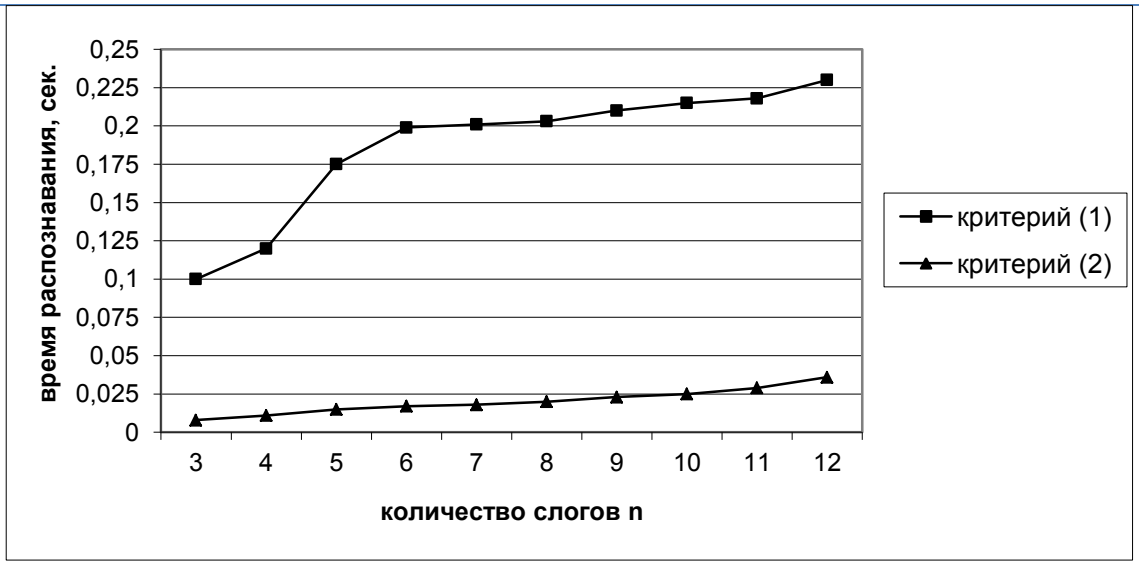

Рисунок 2 – Зависимость времени распознавания (сек) для ЭО ПК ФДС от рассогласования и числа слогов n во входном словосочетании

Для тестирования качества сформулированного алгоритма АРР (2-6) каждым диктором были произнесены по 10 реализаций каждого слова из словаря эталонов. Основным требованием к произнесению команд было разделение слов на открытые слоги с четкой паузой между слогами. В процессе распознавания слоги выделялись простейшим амплитудным детектором паузы, определенной как сигнал с малой амплитудой длительностью не менее 70 мс. Качество распознавания оценивалось по двум показателям: 1) вероятность ошибки (отсутствии произнесенной команды в списке альтернатив); 2) среднее количество *K* альтернатив, которые после упорядочивания находятся ближе к входному слову, чем произнесенная команда.

Результаты в виде зависимости усредненных по дикторам вероятности ошибки и величины *K* от количества слогов *n* во входном словосочетании показаны на рисунках 3 и 4 в виде диаграмм типа «ящик с усами». Заметим, что большая часть ошибок при распознавании связана с неверным определением количества слогов во входной команде из-за присутствия постороннего шума или из-за малой паузы между слогами (менее 70 мс).

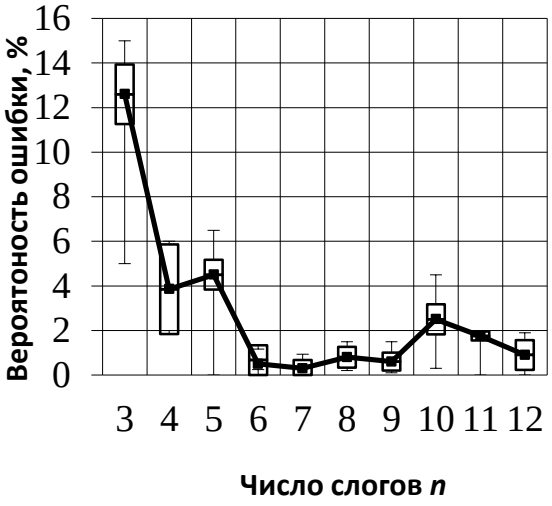

Рисунок 3 – Зависимость вероятности ошибки для ЭО ПК ФДС от количества слогов n во входном словосочетании

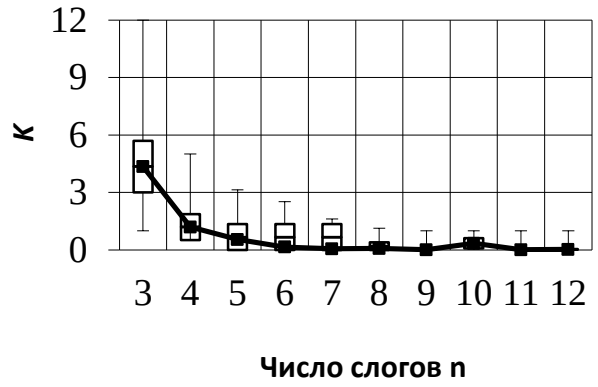

Рисунок 4 – Зависимость числа проверяемых эталонов K для ЭО ПК ФДС от числа слогов n во входном словосочетании

Для сравнения: на рисунке 5 показана вероятность ошибки для распознавания тех же словосочетаний в русскоязычной версии системы Google Voice Search [13]. Система выдает только наиболее близкое словосочетание, поэтому оценивать величину *K* здесь не нужно. Здесь качество дикторонезависимого распознавания из большого словаря очень высоко. Тем не менее, в среднем вероятность ошибки на 10% выше, чем для предложенного подхода. Кроме того, некоторые словосочетания отсутствуют в словаре системы, поэтому они остались нераспознанными.

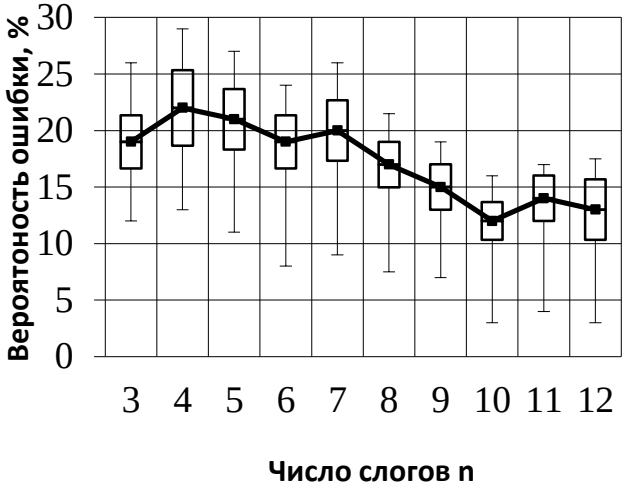

Рисунок 5 – Зависимость вероятности ошибки для Google Voice Search от количества слогов n во входном словосочетании

Для тестирования точности алгоритма АДР (2-4), (7), (8) (последний пункт программы эксперимента) проводилась идентификация диктора при условии сохранения ФБД всех потенциальных дикторов. Для построения фонетического кода каждого сегмента речевого сигнала используются эталоны из объединенных ФБД всех дикторов, а решение о идентификации диктора для слога принимается в пользу пользователя, звуки из ФБД которого наиболее близки (по частотному признаку – по аналогии с (3)) к сегментам этого слога. В эксперименте принимали участие два диктора – мужчина и женщина. Результаты в виде зависимости точности диаризации от количества слогов во входном словосочетании приведены на рисунке 6.

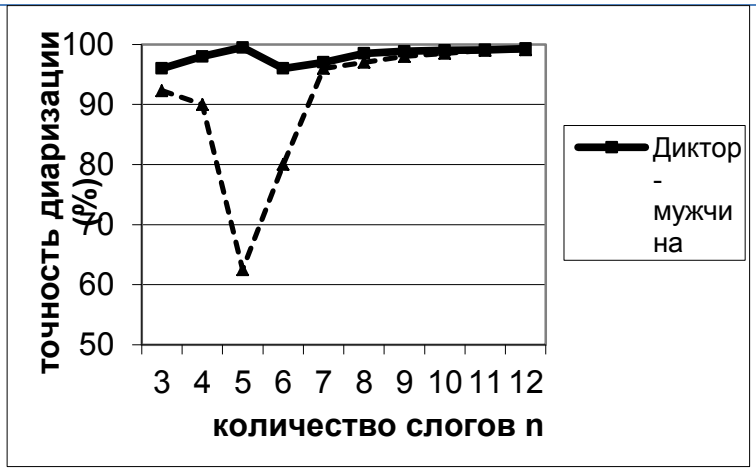

Рисунок 6 – Зависимость точности АДР (в %) для ЭО ПК ФДС от числа слогов n во входном словосочетании

ВЫВОДЫ

По результатам проведенного экспериментального исследования можно сделать следующие выводы. Во-первых, точность классификации для ЭО ПК ФДС выше, чем для универсального решения задачи АРР, такого, как Google Voice Search (рис. 5). Во-вторых, качество предложенной системы можно повысить, если использовать более совершенные алгоритмы выделения открытых слогов, основанные, например, на предварительном выделении однородных последовательных сегментов речевого сигнала. В-третьих, важное преимущество адаптивной реализации (1) метода ФДС состоит в существенном (на порядок и более) сокращении вычислительных затрат на реализацию, при этом и точность, и надежность АРР обеспечиваются на высоком уровне. И, наконец, можно сделать главный вывод – разработанный ЭО ПК ФДС, основанный на требовании четкого слогового произношения, позволяет построить надежную систему голосового управления, в которой преодолены такие проблемы канонического подхода к АРР, как длительность процедуры настройки на диктора и сложность адаптации к новому рабочему словарю.

Таким образом, в работе показано, что если искусственно ввести требование к четкому слоговому произношению команд пользователем [14], то можно построить достаточно надежную систему с автоматически перенастраиваемым словарем и с увеличением качества распознавания по мере накопления информации в процессе функционирования. Отметим, что предложенный подход может применяться не только в голосовом управлении, но в и других сферах, требующих быстрой адаптации рабочего словаря, например, таких, как заказы по телефону продуктов и услуг в пределах переменного ассортимента. В задачах подобного рода требование к слоговому произношению слов диктором является более чем приемлемой платой за достигаемые преимущества в точности и надежности их распознавания.

СПИСОК ЛИТЕРАТУРЫ

1. Tan B. A Distributed Speech Remote Control System Based on Web Service and Automatic Speech Recognition, Electrical Power Systems and Computers, Lecture Notes in Electrical Engineering, 2011. – P. 771-778, 99 p.
2. Benesty J., Sondh M., Huang Y. (eds.). Springer Handbook of Speech Recognition, Springer, New York, 2008. – 1159 p.
3. Rabiner L. A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of the IEEE, 1989. – Vol. 77. – № 2. – P. 257-285.

4. Anusuya M.A., Katti S.K. Speech recognition by Machine: A Review, International Journal of Computer Science and Information Security, 2009. – № 6(3).
5. Патент РФ №2011125526/08 21.06.2011. Савченко А.В., Савченко В.В., Акатьев Д.Ю. Устройство для фонетического анализа и распознавания речи. Патент России на полезную модель № 111944, 2011. – Бюл. № 36.
6. Kullback S. Information Theory and statistics, Dover Pub., 1997. – 399 p.
7. Савченко В.В. Метод фонетического декодирования слов в задаче автоматического распознавания речи на основе принципа минимума информационного рассогласования // Известия ВУЗов России. Радиоэлектроника, 2009. – № 5. – С. 31-41.
8. Fredouille C., Senay G. Technical improvements of the EHMM based speaker diarization system for meeting records, Lecture Notes in Computer Scienceб Proc. of Machine Learning for Multimodal Interaction (MLMI), 2007.
9. Савченко В.В. Разработка фонетических алгоритмов распознавания и диаризации речи с автоматически перенастраиваемым рабочим словарем // Системы управления и информационные технологии, 2012. – № 3(49). – С. 99-100.
10. Марпл С.Л.-мл. Цифровой спектральный анализ и его приложения. – М.: Мир, 1990. – 584 с.
11. Савченко В.В., Савченко А.В. Методика формирования рабочего словаря в системах автоматического распознавания речи по тематическому файлу в текстовом формате // Системы управления и информационные технологии, 2012. – № 2.2(48). – С. 284-289.
12. Савченко А.В. Автоматическое построение фонетической транскрипции речи на основе принципа минимума информационного рассогласования // Вестник компьютерных и информационных технологий, 2012. – № 8. – С. 14-19.
13. Schuster M. Speech Recognition for Mobile Devices at Google // Lecture Notes in Computer Science, 2010 (6230). – P. 8-10.

Савченко Андрей Владимирович
Национальный исследовательский университет Высшая школа экономики, Нижний Новгород
Кандидат технических наук, доцент кафедры информационных систем и технологий
Тел.: 8 950 624 32 85
Email: avsavchenko@hse.ru

A.V. SAVCHENKO *(Candidate of Engineering Sciences, Associate Professor of the Department of Information Systems and Technologies)*
National Research University Higher School of Economics, Nizhny Novgorod

EXPERIMENTAL STUDY RESULTS OF THE PHONETIC WORDS DECODING METHOD IN RUSSIAN SPEECH RECOGNITION AND DIARIZATION PROBLEMS

The results of experimental study of software prototype of the words phonetic decoding method with the Kullback-Leibler information discrimination principle in Russian speech recognition and diarization are discussed. The proposed system is shown to be characterized by high reliability and computing efficiency of the isolated words recognition. The recommendations of its practical usage in a remote control applications are given.

Keywords: *automatic russian speech recognition; speech diarization; words phonetic decoding method; minimum information discrimination principle.*

BIBLIOGRAPHY (TRANSLITERATED)

1. Tan B. A Distributed Speech Remote Control System Based on Web Service and Automatic Speech Recognition, Electrical Power Systems and Computers, Lecture Notes in Electrical Engineering, 2011. – P. 771-778, 99 p.
2. Benesty J., Sondh M., Huang Y. (eds.). Springer Handbook of Speech Recognition, Springer, New York, 2008. – 1159 p.
3. Rabiner L. A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of the IEEE, 1989. – Vol. 77. – № 2. – P. 257-285.

4. Anusuya M.A., Katti S.K. Speech recognition by Machine: A Review, International Journal of Computer Science and Information Security, 2009. – № 6(3).
5. Patent RF №2011125526/08 21.06.2011. Savchenko A.V., Savchenko V.V., Akat'ev D.Yu. Ustrojstvo dlya foneticheskogo analiza i raspoznavaniya rechi. Patent Rossii na poleznuyu model' № 111944, 2011. – Byul. № 36.
6. Kullback S. Information Theory and statistics, Dover Pub., 1997. – 399 p.
7. Savchenko V.V. Metod foneticheskogo dekodirvaniya slov v zadache avtomaticheskogo raspoznavaniya rechi na osnove principa minimuma informacionnogo rassoglasovaniya // Izvestiya VUZov Rossii. Radioe'lektronika, 2009. – № 5. – S. 31-41.
8. Fredouille C., Senay G. Technical improvements of the EHMM based speaker diarization system for meeting records, Lecture Notes in Computer Scienceб Proc. of Machine Learning for Multimodal Interaction (MLMI), 2007.
9. Savchenko V.V. Razrabotka foneticheskix algoritmov raspoznavaniya i diarizacii rechi s avtomaticheski perenastraivaemy'm rabochim slovaryom // Sistemy' upravleniya i informacionny'e texnologii, 2012. – № 3(49). – S. 99-100.
10. Marpl S.L.-ml. Cifrovoj spektral'ny'j analiz i ego prilozheniya. – M.: Mir, 1990. – 584 s.
11. Savchenko V.V., Savchenko A.V. Metodika formirovaniya rabochego slovarya v sistemax avtomaticheskogo raspoznavaniya rechi po tematicheskomu fajlu v tekstovom formate // Sistemy' upravleniya i informacionny'e texnologii, 2012. – № 2.2(48). – S. 284-289.
12. Savchenko A.V. Avtomaticheskoe postroenie foneticheskoj transkripcii rechi na osnove principa minimuma informacionnogo rassoglasovaniya // Vestnik komp'yuterny'x i informacionny'x texnologij, 2012. – № 8. – S. 14-19.
13. Schuster M. Speech Recognition for Mobile Devices at Google // Lecture Notes in Computer Science, 2010 (6230). – P. 8-10.