УДК 621.372:519.72

Метод направленного перебора словаря в задаче автоматического распознавания речи на основе принципа минимума информационного рассогласования

Автоматическое распознавание речи, распознавание образов, распознавание с обучением, критерий минимума информационного рассогласования, генетический алгоритм.

А. В. Савченко

Государственный университет «Высшая школа экономики» - Нижегородский филиал

Ставится и решается задача автоматического распознавания речевых сигналов на основе принципа минимума информационного рассогласования. Предложен метод направленного перебора словаря эталонов как альтернатива методу полного перебора. Представлены результаты экспериментального исследования предложенного метода.

The Direct Search in a Dictionary's Method in a Problem of Automatic Speech Recognition Based on Minimum Information-Mismatch Principle

Automatic speech recognition, pattern recognition, recognition with training, criterion of the minimum of information mismatch, genetic algorithm**.**

A.V. Savchenko.

State University "High School of Economics", Nizhny Novgorod

The article deals with a problem of speech signals recognition based on the minimum of the informative divergent criterion. The new algorithm of direct search in a dictionary has been developed as an alternative for checking all patterns. The program and experimental results of this method have been produced.

Введение. Принцип минимума информационного рассогласования (МИР) является эффективным инструментом для решения разнообразных задач в области распознавания образов [1]. Задача автоматического распознавания речи (АРР) – одна из наиболее актуальных разновидностей такого рода задач. Существует множество подходов к ее решению. Среди них очевидный интерес представляет теоретико-информационный подход, разработанный в рамках информационной теории восприятия речи (ИТВР) [2] и основанный на упомянутом выше принципе МИР и методе обеляющего фильтра (МОФ). Его эффективность и преимущества по сравнению с другими подходами показаны в работах [3, 4] на ряде примеров из практики АРР. Между тем, далеко не все преимущества и возможности ИТВР на данный момент получили необходимое освещение и развитие. В частности, до настоящего времени практически не исследовались преимущества принципа МИР перед традиционными методами и подходами в задачах автоматического распознавания сложных

речевых единиц типа отдельных (изолированных) слов или целых фраз. Исследованиям в этом актуальнейшем направлении и посвящена предлагаемая статья. В ней при учете метрических свойств решающей статистики МИР предложен метод направленного перебора (МНП) словаря эталонов как альтернатива традиционному методу сплошного перебора при проверке конкурирующих гипотез. Полученные результаты и сделанные по ним выводы рассчитаны на широкий круг специалистов в области современной теории и практики автоматической обработки речевых сигналов и распознавания образов.

Элементы ИТВР. Центральным элементом ИТВР является понятие «фонема». Под фонемой обычно понимают минимальную единицу звукового (фонетического) строя национального языка, или «элементарную речевую единицу» (ЭРЕ) [4]. Разным языкам соответствуют разные списки фонем: и по составу, и по количеству R их элементов. Это базовый уровень описания каждого языка. В подтверждение можно привести пример: большинство современных речевых баз данных сопровождается транскрипцией речевых сигналов, т.е. их описанием через последовательность фонем. С другой стороны, фонетический строй языка предъявляет определенные требования к его носителям, посредством которых (и только так) этот строй и реализуется в коммуникациях. Сколько носителей – столько и разных реализаций фонетического списка национального языка. В этом проявляется краеугольная проблема вариативности устной речи. Однако, несмотря на существующие различия в реализациях каждой отдельной (r-ой) фонемы, все они воспринимаются человеком как нечто общее, иначе речь утратила бы свою информативность. Можно поэтому утверждать, что одноименные (однофонемные) реализации $\mathbf{x}_{r,\,j}$, $j=\overline{1,J_r}, J_r>>1$, в сознании человека группируются в соответствующие классы или речевые образы $X_r=\left\{\mathbf{x}_{r,\,j}\right\}$, $r=\overline{1,R}$, вокруг некоторого центра – эталонной метки данного образа [2]. В ИТВР указанные эталоны определяются в строгом, теоретико-информационном смысле: речевая метка $\mathbf{x}_r^*\in X_r$ образуется *информационный центр-эталон* r-го речевого образа, если в пределах множества X_r она характеризуется минимальной суммой информационных рассогласований (ИР) по Кульбаку-Лейблеру [5] относительно всех других его меток-реализаций $\mathbf{x}_{r,\,j}$, $j=\overline{1,J_r}$. По своей сути это статистический аналог понятия «центр массы» физического тела.

Нетрудно увидеть, что именно в понятии информационного центра (ИЦ) r-го множества реализаций $\mathbf{x}_r^*$ дается наиболее информативное

определение соответствующей фонемы. А множество всех ИЦ $\{\mathbf{x}_r^*\}$ определяет исчерпывающим образом фонетический состав речевого сигнала. Одновременно становится очевидным и механизм формирования самого такого множества. Анализируемый (входной) речевой сигнал $X(t)$ в дискретном времени $t = 0,1,...$ сначала разбивается на ряд последовательных сегментов данных $\mathbf{x}(t)$ длиной в одну ЭРЕ $\tau \approx (10 – 15)\ mc$ [4]. После этого каждый полученный парциальный сигнал рассматривается в пределах конечного списка фонем $\{X_r\}$ и отождествляется с той X_ν из них, которая отвечает принципу минимума величины ИР между вектором $\mathbf{x}(t)$ и соответствующим эталоном $\mathbf{x}_\nu^*$, $\nu \le R$. Это стандартная [1, 2] формулировка критерия МИР в задачах автоматического распознавания образов.

Критерий МИР. Задача существенно упрощается, если воспользоваться гауссовой (нормальной) аппроксимацией закона распределения речевого сигнала на интервалах его квазистационарности $\tau \approx \text{const}$ вида $\mathbf{P}_r = N(\mathbf{K}_r)$, где $\mathbf{K}_r$ - автокорреляционная матрица (АКМ) размера $n \times n$, $n \ge 1$. Известно [6], что в этом случае критерий МИР является оптимальным в байесовском смысле [1]. Задача формулируется как проверка простых гипотез о законе распределения ЭРЕ. А соответствующий набор оптимальных решающих статистик может быть записан следующим образом [2]:

$$\rho(\mathbf{x}/\mathbf{x}_r) = \frac{1}{2n}[tr\left(\hat{\mathbf{K}} \cdot \mathbf{K}_r^{-1}\right) - \log\left|\hat{\mathbf{K}} \cdot \mathbf{K}_r^{-1}\right| - n] \quad (1)$$

где $\hat{\mathbf{K}}$ - это выборочная оценка АКМ анализируемого сигнала $\mathbf{x} = \mathbf{x}(t)$, $t = 0,1,2,...$. Решение принимается в пользу гипотезы $\mathbf{P}_\nu$, $\nu \le R$, по признаку минимума ν-ой решающей статистики (1), т.е.

$$W_\nu(X): \rho(\mathbf{x}/\mathbf{x}_\nu) = \min_r \rho(\mathbf{x}/\mathbf{x}_r) \quad (2)$$

Причем, в задачах с априорной неопределенностью вместо неизвестных, в общем случае, фонемных АКМ $\mathbf{K}_r$, $r = \overline{1,R}$, в выражение (1) подставляют их статистические оценки, которые предварительно получают по R (число фонем в списке) классифицированным выборкам речевого сигнала. Это стандартная формулировка критерия МИР *с обучением*.

В работе [6] также показано, что в асимптотике, когда $n \to \infty$, и при распределении сигнала $\mathbf{P}_r = N(\mathbf{K}_r)$ с обратной АКМ $\mathbf{K}_r^{-1}$ ленточной структуры оптимальный алгоритм (2) сводится к минимизации выражения

$$\rho(\mathbf{x} / \mathbf{x}_r) = \frac{1}{F} \sum_{f=1}^{F} G(f) - 1 \quad (3)$$

Это известная формулировка критерия МИР на основе авторегрессионной (АР) модели речевого сигнала. Здесь введено обозначение $G(f) = \left(\frac{G_x(f)}{G_r(f)} + \ln \frac{G_r(f)}{G_x(f)} \right)$, где $G_x(f)$ – выборочная оценка спектральной плотности мощности (СПМ) входного сигнала $\mathbf{x}(t)$ в функции дискретной частоты f, а $G_r(f)$ – СПМ эталона r-ой фонемы $\mathbf{x}_r^* \in X_r$; F – верхняя граница частотного диапазона речевого сигнала или используемого канала связи.

Главное достоинство АР-модели, как известно [2, 3], состоит в возможности предварительной нормировки речевых сигналов по дисперсиям их порождающих процессов. Применительно к сигналам типа ЭРЕ 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 или даже фразу. При учете этого свойства выражение (3) приобретает предельно простой вид [4]

$$G(f) = \frac{\left| 1 + \sum_{m=1}^{p} a_r(m) e^{-\frac{j\pi mf}{F}} \right|^2}{\left| 1 + \sum_{m=1}^{p} a_x(m) e^{-\frac{j\pi mf}{F}} \right|^2} \quad (4)$$

Выражения (3) и (4) представляю собой стандартную формулировку МОФ в частотной области. Здесь выражение в числителе определяет квадрат амплитудно-частотной характеристики r-го обеляющего фильтра, настроенного на r-ю фонему $\mathbf{x}_r^*$, $r = \overline{1, R}$. Преимуществом такой интерпретации принципа МИР является, прежде всего, возможность его практической реализации в адаптивном варианте на основе быстрых вычислительных процедур авторегрессионного анализа, таких как метод Берга и др. [7]. Задача в общем случае сводится к двухэтапной проверке статистических гипотез. На первом этапе распознаются ЭРЕ типа отдельных

фонем. На втором – слова, фразы и целые тексты как соответствующим образом структурированные последовательности разных фонем.

Задача первого этапа. На пути к практическому осуществлению решающего правила (2) сначала требуется определить множество всех ЭРЕ $\{X_r\}$ как результат линейного расчленения речевого сигнала на квазистационарные последовательности отсчетов $\mathbf{x} = \{x_1,...,x_n\}$ конечного объема *n*. Подробно указанная процедура описана в работе [8]. Разработанный в ней алгоритм сводится к последовательной проверке условия

$$\rho\left(\mathbf{x}/\mathbf{x}_r^*\right) < \rho_0, r \le R \quad (5)$$

об однородности распределений вектора отсчетов **x** сигнала анализируемой (текущей) ЭРЕ и вектора отсчетов ИЦ каждой фонемы из текущего списка $\{\mathbf{x}_r^*\}$. Здесь ρ_0 - допустимый уровень ИР в пределах однородного множества X_r. При нарушении условия (5) в списке $\{\mathbf{x}_r^*\}$ появляется еще одна, (*R*+1)-я фонема $\mathbf{x}_{R+1}^*$.

Вычисления по схеме (3…5) повторяются циклически для всех последующих сегментов данных из речевого сигнала X, причем повторятся «нарастающим итогом» для переменного значения *R*=1, 2,… В результате получим множество из *R** выявленных на первом этапе обработки фонем. При этом понятно, что с точки зрения качества полученного результата первостепенный интерес представляет собой множество четких фонем. Поэтому в работе [8] была предложена дополнительная процедура отбраковки сомнительных по своей четкости фонем путем установления к ним требования по минимальной длительности эталонных ЭРЕ вида

$$V_r \ge V_0 \quad (6)$$

Здесь V_r - число отсчетов в векторе *r*-й фонемы $\mathbf{x}_r^*$; V_0 – пороговый уровень для минимального числа отсчетов в результирующем списке фонем. Множество $R<R^*$ четких фонем $\{\mathbf{x}_r^*\}$ и следует считать, в общем случае, основным результатом фонетического анализа речи на первом этапе обработки речевого сигнала.

Его наиболее полная информационная характеристика – это (*R*×*R*)-матрица $\left\|\rho_{r,\nu}\right\|$ величин ИР $\rho_{r,\nu} = \rho\left(\mathbf{x}_\nu^*/\mathbf{x}_r^*\right)$ между всеми парами выявленных фонем. В качестве примера в табл. 1 представлен фрагмент такой матрицы для случая *R* =20, полученной экспериментальным путем в условиях и с применением программных средств из упомянутой выше

работы [8] для одного диктора-мужчины при установленных в алгоритме (3…5) порогах $\rho_0 = 1{,}0;\ \ \mathrm{V}_0 = 800$.

Таблица 1

	$\mathbf{x}_1^*$	$\mathbf{x}_2^*$	$\mathbf{x}_3^*$	$\mathbf{x}_4^*$	**…**	$\mathbf{x}_{20}^*$
$\mathbf{x}_1^*$	0	0.98286	4.4857	0.6369	…	1.8222
$\mathbf{x}_2^*$	0.6611	0	2.6279	0.8887	…	2.9414
$\mathbf{x}_3^*$	10.221	7.369	0	10.845	…	2.8603
$\mathbf{x}_4^*$	0.4823	0.60722	4.9626	0	…	1.4164
…	…	…	…	…	…	…
$\mathbf{x}_{20}^*$	0.8513	0.96772	3.2227	0.93173	…	0

При этом применялись специальные программные и аппаратные средства: динамический микрофон AKG D77 S и ламповый микрофонный предусилитель ART TUBE MP Project Series USB. Частота дискретизации встроенного АЦП была установлена равной 8 кГц – общепринятое значение при обработке устной речи. Продолжительность записи составила около полутора минут**.** Это был художественный текст, взятый из первой главы романа А.С. Пушкина "Капитанская дочка". Длина одного сегмента данных составляла *n* = 80 отсчетов, или 10 *мс* по времени. А для расчета коэффициентов авторегрессии $\{a_r(m)\}$ в (4) использовалась рекуррентная процедура Берга-Левинсона [7] с высокой скорость сходимости. Полученное множество $\{\mathbf{x}_r^*\}$ совместно с выражениями (3), (4) и определяло в конечном итоге принятое по критерию МИР (2) решение в пользу одной из фонем $\mathbf{x}_r^*$ по каждой очередной ЭРЕ $\mathbf{x}_i$ в составе анализируемого речевого сигнала $X(t)$.

Распознавание изолированных слов. Мысленно разобьем анализируемый составной сигнал $X(t)$, теперь длиной в одно слово, на некоторую последовательность фонем $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_L\}, \mathbf{x}_i \in \{\mathbf{x}_\nu^*\}$ длиной в одну ЭРЕ τ каждая. Здесь *L* – длина изолированного слова, выраженная в числе входящих в него фонем. При этом некоторые фонемы в нем могут повторяться. В качестве иллюстрации на рис.1 показаны временные диаграммы двух слов разной длины.

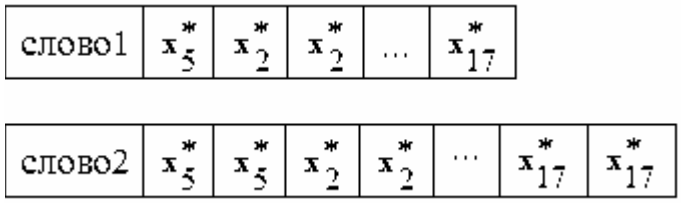

Рис. 1. Временные диаграммы слов

В задаче АРР общего вида (2) сигналу с входа $X(t)$ противопоставляется R альтернатив из множества эталонных слов $\{X_r\}$. Каждое из таких слов-эталонов предварительно разбивается на соответствующую последовательность фонем $X_r = \left\{\mathbf{x}^*_{r,1}, \mathbf{x}^*_{r,2}, ..., \mathbf{x}^*_{r,L_r}\right\}$, $\mathbf{x}^*_{r,i} \in \left\{\mathbf{x}^*_{\nu}\right\}$. Предположим сначала, что все они согласованы по своей длине с длиной L анализируемого слова. При учете статистической независимости выделенных фонем в слове его АКМ размера $(L \cdot n) \times (L \cdot n)$ имеет блочно-диагональную структуру и состоит из L квадратных $n \times n$ АКМ составляющих слова фонем: $\mathbf{K}_r = diag\left\{\mathbf{K}_{r,1}, \mathbf{K}_{r,2}, \ldots, \mathbf{K}_{r,Lr}\right\}$. В таком случае, как это следует из (1), рассогласование Кульбака-Лейблера между двумя рассматриваемыми словами

$$\rho(X / X_r) = L^{-1} \sum_{i=1}^{L} \rho(\mathbf{x}_i / \mathbf{x}_{r,i}) \quad (7)$$

определяется суммой ИР $\rho(\mathbf{x}_i / \mathbf{x}_{r,i})$ между парами их фонем на одноименных позициях в слове.

Все выше сказанное сохраняет свою справедливость и в общем случае разной длины $L_r \neq const$ каждого отдельного слова, например, из-за разного темпа речи одного или разных дикторов. На нашем рис. 1 отображена именно эта ситуация. Во всех таких случаях перед применением выражения (7) входной сигнал $\mathbf{X}$ и каждый сигнал-эталон $\mathbf{X}_r$ должны быть предварительно выровнены по темпу речи. Для этого на практике применяются специальные алгоритмы, основанные, как правило, на методе динамического программирования [9].

Таким образом, распознавание слов по МОФ в общем случае реализуется на основе многоканальной обработки, в которой число каналов R определяется количеством слов-эталонов. При этом в каждом r-м канале используется набор из L_r обеляющих фильтров, настроенных на последовательные стационарные участки (фонемы) соответствующего эталонного слова. Решение (2) принимается по критерию минимума суммы решающих статистик (7) по всем L сегментам анализируемого слова.

Метрические свойства решающей статистики МИР. Рассмотрим случай $R>>1$, когда решается задача АРР с объемом эталонного словаря (ЭС) в сотни и даже тысячи слов. В указанных условиях практическая реализация МОФ по схеме R-канальной обработки (3)…(7) наталкивается на очевидную проблему вычислительной сложности. Выполнение проверки гипотез по схеме (7), включающей в себя предварительное выравнивание анализируемых слов по длительности или темпу речи, становится в данном случае непреодолимым препятствием для работы системы АРР в режиме реального времени. В поиске путей решения указанной проблемы за счет отказа от сплошного перебора ЭС и состоит центральная задача настоящей работы.

Прежде всего, отметим метрические свойства решающей статистики МИР $\rho(X/X_r)\geq 0$ с равенством ее нулю лишь в идеальном случае совпадения входного и эталонного сигналов. Поэтому вначале преобразуем критерий МИР (2) к упрощенному (в его практической реализации) виду

$$W_\nu(X):\ \rho(X/X_\nu)<\rho_1. \quad (8)$$

Здесь $\rho_1<\rho_0$ - это порог для допустимой величины ИР по Кульбаку-Лейблеру на множестве одноименных слов за счет известной вариативности устной речи. Значение такого порога нетрудно установить опытным путем [9]. Например, в условиях предыдущего эксперимента (табл. 1) на уровне значимости критерия (8) $\alpha=0{,}1...0{,}15$ было получено приближенное равенство $\rho_1\approx 0{,}3...0{,}5$. Это примерно в два раза меньше значения «информационного радиуса» ρ_0 в (5) при определении множества реализаций каждой отдельной фонемы. По своей сути выражение (8) определяет условие «останова» при переборе альтернатив в рамках проверочной процедуры по МОФ (3)…(7).

Таким образом, при принятии решения на основе принципа МИР (2) требуется просматривать не весь словарь целиком, а вычислять величину ИР лишь до тех пор, пока оно не будет меньше некоторого порогового уровня. Нетрудно понять, что само по себе указанное обстоятельство позволит сократить объем перебора в среднем на 50%. Иными словами, благодаря использованию правила останова (8) удается в два раза сократить объем выполняемых вычислений и этим существенно ослабить проблему практической реализуемости АРР в режиме реального времени. В этом состоит принципиальное преимущество МОФ по сравнению со всеми его наиболее известными аналогами, такими как группа методов на основе скрытых марковских моделей (СММ-методы) и другие, в которых применяются классические (байесовские) критерии: минимума среднего риска, максимума апостериорной вероятности и др. Между тем, как это

выясняется ниже, рассмотренный выигрыш в производительности далеко не исчерпывает всех преимуществ МОФ и принципа МИР в задаче распознавания образов.

Действительно, общая формулировка задачи (2) позволяет рассматривать ее как задачу оптимизации и применять алгоритмы поиска оптимального решения с заданным условием останова (8). В такой задаче на множестве эталонных слов $\{X_r\}$ требуется найти такое слово X_ν, которое будет минимизировать статистику МИР. В таком случае метод, сводящийся к полному перебору ЭС, является одним из множества известных методов оптимизации систем. Главным препятствием для применения в нашей задаче более эффективного оптимизационного метода является то, что, во-первых, задача относится к области дискретной математики и, во-вторых, в ней требуется найти глобальный минимум решающей статистики (2). По-видимому, наиболее обоснованным способом поиска глобального экстремума в указанных условиях можно считать метод случайного поиска и, в частности, его наиболее распространенную разновидность – генетический алгоритм (ГА) [10]. В нашем случае генотипом для ГА будет номер *r* слова из ЭС, а фенотипом – выражение (2). Возможности применения ГА будут подробно описаны ниже в результатах экспериментальных исследований.

Предварительно же можно отметить следующий главный недостаток ГА в рамках решения поставленной задачи. Перебор организуется только по *номеру* эталонного слова в ЭС. То есть информация о самих словах, рассогласованиях между ними, в стандартном ГА никак не учитывается. На помощь снова приходит принцип МИР. Действительно, на основе того же выражения (8) мы можем сформулировать критерий останова работы ГА. В этом случае появляется гарантия того, что решение задачи, если оно существует (то есть если эталон входного слова присутствует в словаре), с помощью ГА будет найдено. Естественным развитием этой идеи может служить предложенный ниже метод направленного перебора (МНП) словаря, в котором метрические свойства решающей статистики МИР (8) используются в наиболее полной степени.

Идея МНП. Следуя общей схеме вычислений по МОФ (3)…(7), сведем задачу автоматического распознавания слова $X(t)$ к проверке N первых вариантов $X_1,...,X_N$ из R альтернатив $\{X_r\}$ при условии $N<<R$. Если, по крайней мере, одна из них $X_\nu, \nu \le N$, отвечает требованию останова (8), процесс поиска оптимального решения по критерию МИР (2) на ней и завершается. Однако в общем случае можно предположить, что ни одна из первых N альтернатив проверку (8) не проходит. В таком случае можно проверить вторые N эталонных слов из множества $\{X_r\}$, потом

третьи и т.д. – до момента выполнения условия (8). Но есть и иной, предпочтительный, вариант поведения.

Расставим слова из нашей контрольной выборки $X_1,...,X_N$ в порядке убывания их ИР $\rho(X/X_n), n=\overline{1,N}$. В результате будем иметь упорядоченную (ранжированную) последовательность эталонных слов вида $\left\{X_{i_j}\right\}=\left\{X_{i_1},X_{i_2}\dots X_{i_N}\right\}, i_j \le N$. Соответствующая последовательность $\left\{\rho_j\right\}$ их ИР $\rho_j=\rho\left(X/X_{i_j}\right), i_j \le N,$ будет иметь характер монотонно убывающей зависимости. Ее вид, в частности, скорость убывания, будет зависеть как от состава контрольной выборки эталонов $X_1,...,X_N$, так и ее теоретико-информационных характеристик, в частности, от величины взаимных информационных рассогласований (ВИР) $\rho_{(j+1)/j}=\rho\left(X_{i_{j+1}}/X_{i_j}\right)$ между парами соседних элементов в упорядоченной контрольной выборке $\left\{X_{i_j}\right\}$. Проиллюстрируем указанную зависимость с помощью специально для этого нами созданного экспериментального ЭС.

При его формировании в качестве базы эталонных данных использовался тот же набор из 20 фонем $\left\{\mathbf{x}_r^*\right\}$, который был положен в основу построения табл.1. На его основе путем простого перебора разных сочетаний фонем было сформировано множество эталонных слов $\left\{X_r\right\}$ большого объема *R=10000*. В нем каждое слово отображалось определенной последовательностью фонем длиной в одну ЭРЕ, как это было показано на рис.1. Их сочетания в экспериментальном варианте словаря устанавливались случайным образом. Тем самым достигались наиболее жесткие условия для последующего автоматического распознавания слов *X(t).* При этом длина каждого слова варьировалась также случайным образом в рабочем интервале ее значений: от 10 до 20 ЭРЕ. А для их выравнивания по динамике применялась стандартная процедура [9].

При этом на первом этапе вычислений состав контрольной выборки $X_1,...,X_N$ был выбран нами наугад, ее объем был установлен равным *N=100*. Для вычисления ИР (7) использовались данные предварительного фонетического анализа из табл.1. По результатам вычислений была

получена упорядоченная контрольная выборка $\left\{X_{i_j}\right\}$. Ее фрагмент из 30 последних элементов $X_{i_{71}},...,X_{i_{100}}$ отображен на рис. 2,а в виде графика зависимости их ИР $\left\{\rho_j\right\}$ относительно определенного слова на входе $X(t)\in\{X_r\}$. Здесь же для сравнения на рис.2,б (сплошная линия) представлен график соответствующей зависимости величины ВИР $\left\{\rho_{(j+1)/j}\right\}$. Хорошо, в частности, видно, что данная последовательность вслед за последовательностью ИР $\left\{\rho_j\right\}$ для элементов упорядоченной контрольной выборки $\left\{X_{i_j}\right\}$ имеет характер затухающих колебаний.

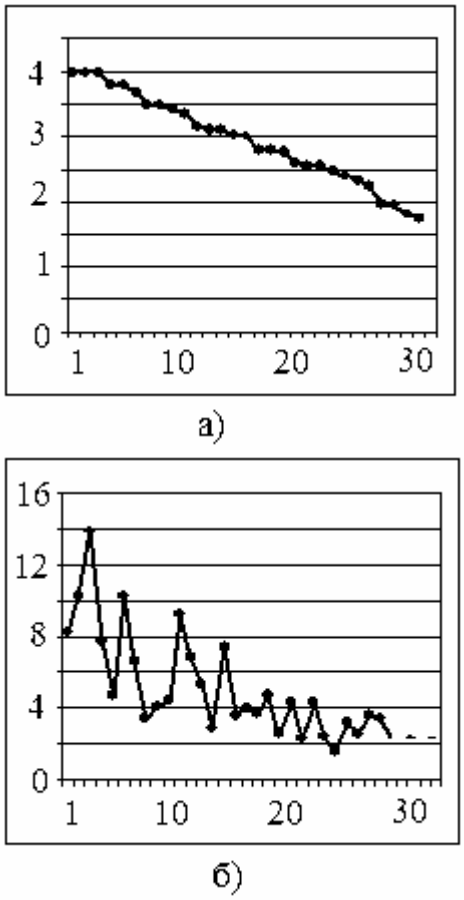

Рис.2. Последовательность ИР

Указанное наблюдение усиливается результатами экстраполяции последовательности значений ВИР на рис.2б (штриховая линия) за

границы контрольной выборки по формуле оценки линейного прогнозирования *P*-го порядка:

$$\rho_{(N+1)/N} = \sum_{i=1}^{P} a_i \rho_{(N-i+1)/(N-i)} \quad (9)$$

Здесь $\{a_i\}$ – вектор АР-коэффициентов. При этом, как и ранее, для АР-анализа использовалась рекурсивная вычислительная процедура Берга-Левинсона [7], а порядок авторегрессии был установлен равным P=5. Отсюда и вытекает главная идея МНП: использовать последний элемент X_{i_N} из упорядоченной контрольной выборки $\left\{ X_{i_j} \right\}$ как наилучшее приближение к искомому слову $X(t)$ в роли точки отсчета для поиска наиболее подходящих «кандидатов» $X_1,...,X_N$ в очередную контрольную выборку. При этом данные экстраполяции ВИР (9) могут служить ориентиром для определения максимально допустимой различий слов-эталонов из новой контрольной выборки по отношению к «точке отсчета» в теоретико-информационном смысле. Проиллюстрируем сказанное с помощью диаграммы поисковой процедуры МНП на рис.3.

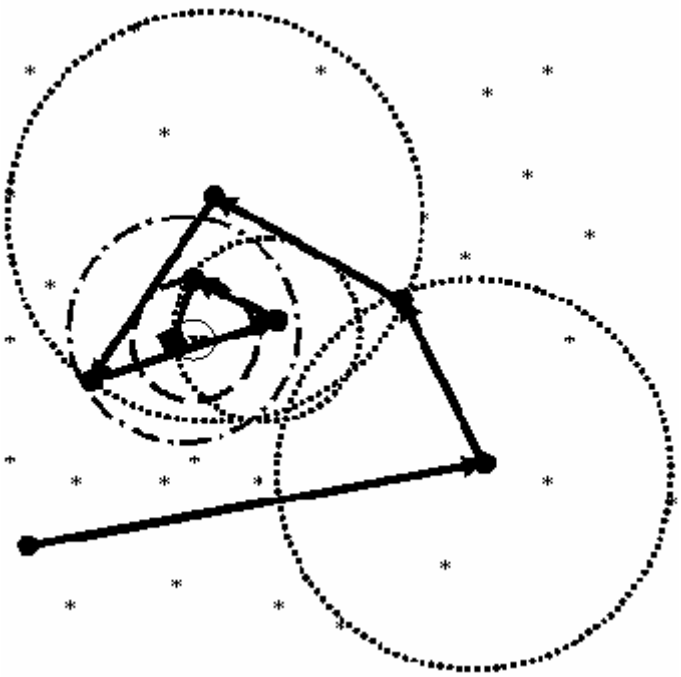

Рис. 3. Поисковая процедура МНП

Здесь звездочками обозначены все имеющиеся слова-эталоны, буквой X – входное слово, а ромбиком – наиболее близкий к $X(t)$ эталон. Он и определяет искомое оптимальное решение задачи. Траектория поиска отображается на рисунке ломаной направленной линией. Жирными точками на ней обозначена последовательность наиболее близких к

оптимуму слов X_{i_N} после нескольких подряд этапов вычислений. Окружностями здесь отмечены границы соответствующих контрольных точек-выборок $X_1,...,X_N$. Их радиусы определяются согласно выражению (9). Хорошо видно, что траектория поиска имеет вид скручивающейся спирали.

Синтез алгоритма. Следуя определению ИР (7), (3), (4) и используя процедуру динамического выравнивания слов-эталонов разной длины, составим $R\times R$-матрицу $\mathrm{P}=\left\|\rho_{ij}\right\|$ значений ИР $\rho_{ij}=\rho(X_i/X_j),\ i,j\le R$. Эту весьма сложную в вычислительном отношении операцию требуется выполнить лишь раз: на предварительном этапе вычислений – для каждого конкретного ЭС.

Сначала, как это было описано выше, зададимся в пределах имеющегося R-словаря произвольной первой контрольной выборкой $X_1,...,X_N$ некоторого фиксированного объема N, по ней получим ранжированный по критерию МИР (2) ряд данных $\left\{X_{i_j}\right\}$ и, наконец, находим из него первый локальный оптимум X_{i_N}. На этом завершается первый этап вычислений. Переходя ко второму этапу, для выделенного слова-эталона X_{i_N} по матрице P найдем множество из $M<R$ слов $X^{(M)}=\left\{X_{i_{N+1}},\dots X_{i_{N+M}}\right\}, i_j\le R$, находящихся от слова X_{i_N} на «расстоянии» (7), не превышающем порогового значения $\rho_{(N+1)/N}$, или

$$\begin{gathered}\left(\forall X_i\notin X^{(M)}\right)\left(\forall X_j\in X^{(M)}\right)\\ \Delta\rho\left(X_i\right)\ge\Delta\rho\left(X_j\right)\end{gathered}\qquad(10)$$

Здесь $\Delta\rho(X)=\left|\rho\left(X/X_{i_N}\right)-\rho_{(N+1)/N}\right|$ - отклонение спрогнозированного рассогласования $\rho_{(N+1)/N}$ от ИР между словом X и словом X_{i_N}.

На рис.3 каждое такое множество ограничивается соответствующей окружностью с центром в точке X_{i_N}. Добавим к этому множеству еще

один, (*M*+1)-й элемент $X_{i_{N+M+1}}$ из числа не попавших в состав контрольной выборки по результатам предыдущего этапа вычислений. Этим мы вносим в поисковую процедуру определенный элемент случайности как способ достижения глобального оптимума за конечное число шагов оптимизации (этапов вычислений).

В результате получаем вторую контрольную выборку слов-эталонов $\left\{X_{i_{N+1}},\ldots X_{i_{N+M+1}}\right\}, i_j \le R$, для анализа. Далее все вычисления первого этапа циклически повторяются. Повторяются до тех пор, пока на некотором *K*-м этапе не будет выполнено условие останова

$$\rho\left(X / X_{i_N}\right) < \rho_1 . \quad (11)$$

На рис.3 в такой момент входное слово оказывается в пределах границ множества контрольных точек последнего этапа вычислений. Решение здесь принимается в пользу наиболее близкого образа X^*. Или, в худшем случае, после перебора всех альтернатив из словаря $\{X_r\}$ в отсутствие решения (11) делается вывод о том, что прозвучавшее (входное) слово в данном ЭС отсутствует и необходимо задействовать режим переспроса. В общем же случае, суммарное число $N + M \cdot K \le R$ выполняемых согласно (11) проверок может существенно выигрывать по сравнению с объемом используемого ЭС. В этом и состоит эффект направленного перебора.

Таким образом, система выражений (7)…(11) и определяет, в конечном итоге, предлагаемый метод в задаче АРР.

Результаты экспериментальных исследований. После составления описанного выше экспериментального ЭС достаточно большого объема *R=10000* был поставлен и 100 раз выполнен следующий эксперимент. На основе некоторого (каждый раз разного) слова-эталона $\mathbf{X}_r$ путем дублирования или, напротив, удаления части его ЭРЕ каждый раз создавалось некоторое анализируемое слово $X(t)$. На нашем предыдущем рис.1 представлены типичные диаграммы именно такой пары слов X_r и $X(t)$. В каждом случае решалась задача автоматического распознавания слова $X(t)$ из множества всех его допустимых альтернатив $\{X_r\}$. Сначала для этого применялся ГА. Его параметры были выбраны следующим образом: количество особей в начальной популяции *N=128*, количество потомков *M=80*. Порог $\rho_1 = 0{,}1$ для правила останова (8) был выбран, исходя рекомендаций работы [8]. В этом варианте ГА дал точное решение X^* в 98% случаев проверок гипотез. При этом на каждое такое решение в среднем потребовалось проверить (перебрать) 5700 слов, или

57% от объема всего ЭС. При учете вычислительной сложности обязательной в таких случаях процедуры выравнивания слов по динамике подобное ускорение обработки (почти в 2 раза) представляется, на первый взгляд, весьма существенным достижением. Однако данный вывод требует уточнения.

Для этого рассмотрим на рис.4 (график ГА) гистограмму количества осуществляемых в данном случае проверок слов (в процентах). Хорошо видно, что она имеет приблизительно прямоугольный вид во всей своей области определения от 0 до *R*. По сути это означает равновероятно случайный поиск. Примерно тот же результат мы будем иметь по результатам сплошного перебора ЭС с автоматическим остановом по правилу (8). Таким образом, применение ГА почти ничего не дает в решаемой задаче АРР. Совсем иное дело – МНП.

Его гистограмма при тех же значениях параметров *N=128*, *M=80* и $\rho_1 = 0{,}1$ алгоритма (7)…(11) показана на том же рис.4 (график МНП). Здесь среднее количество проверок составило примерно 26% от объема ЭС. С вероятностью 80% МНП это количество не превышает 5000 или 50% от общего числа слов-эталонов для проверки. При этом в 100% случаев было получено абсолютно точное решение $X^* = X_r$.

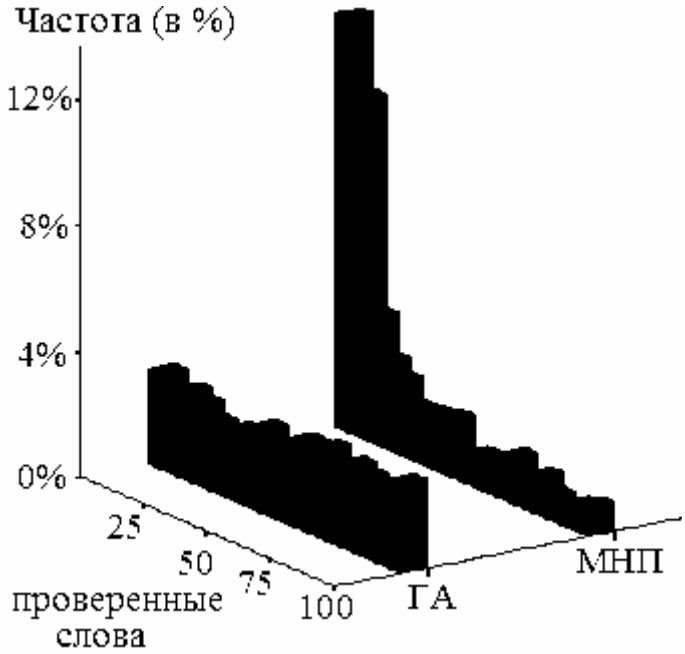

Рис. 4. Гистограммы количества проверок слов

Заключение. Вопрос о повышении скорости вычислений вызывает повышенный интерес среди специалистов как в области теории, так и практики АРР. Действительно, в тех случаях когда объем рабочего словаря начинает составляет сотни и тысячи единиц, большинство известных алгоритмов, работающих на основе сегментирования на отдельные

фонемы и их последующего выравнивания по динамике [9], не могут быть реализованы в режиме реального времени. Поэтому решению проблемы вычислительной сложности для больших словарей в последние годы и уделяется повышенное внимание. В представленной работе для этого предложен новый метод: направленного перебора, основанный на теоретико-информационном подходе и отталкивающийся от метрических свойств решающей статистики МИР [1]. При этом принципиальное значение при переборе альтернатив имеет правила автоматического останова (7). В самом невыгодном варианте своего применения оно сокращает объем вычислений в среднем в 2 раза. А при использовании предложенного метода в формулировке (11) суммарный выигрыш в вычислительной сложности алгоритма возрастает почти до 4 раз. При этом не утрачивается (по сравнению со сплошным перебором ЭС) и качество достигаемого по МНП решения X^*.

Таким образом, благодаря проведенному исследованию предложен новый метод АРР на основе принципа МИР, обладающий широкими функциональными возможностями и высокими эксплуатационными свойствами.

Библиографический список

1. Савченко В.В., Савченко А.В. Принцип минимального информационного рассогласования в задаче распознавания дискретных объектов // Известия вузов России. Радиоэлектроника. 2005. Вып.3. С.10-18.

2. Савченко В.В. Информационная теория восприятия речи. // Известия вузов России. Радиоэлектроника. 2007. Вып.6. С.3-9.

3. Савченко В.В., Акатьев Д.Ю. Теоретико-информационное обоснование метода обеляющего фильтра в задачах автоматического распознавания речи. // Системы управления и информационные технологии. 2008. №1 (31). С.21-30.

4. Савченко В. В., Акатьев Д. Ю., Карпов Н. В. Автоматическое распознавание элементарных речевых единиц методом обеляющего фильтра. // Известия вузов. Радиоэлектроника. 2007. Вып.4. С.11-19.

5. Кульбак С. Теория информации и статистика. М.: Наука, 1967, 408 с.

6. Савченко В.В. Различение случайных сигналов в частотной области // Радиотехника и электроника. 1997. Т.42, №4. С. 426–431.

7. Марпл С.Л.-мл. Цифровой спектральный анализ и его приложения. М.: Мир, 1990, 584 с.

8. В.В.Савченко, Карпов Н.В. Анализ фонетического состава речевого сигнала методом переопределенного дерева. // Системы управления и информационные технологии 2008. №2 (32), С. 297-303.

9. Д.Ю. Акатьев, И.В. Губочкин, В.В.Савченко. Автоматическое распознавание изолированных слов методом обеляющего фильтра с сегментированием и амплитудным ограничением сигналов переспросом // Изв. вузов России. Радиоэлектроника. 2007. Вып. 5. С. 11–18.

10. Goldberg, D.: Genetic Algorithms in search, optimization, and machine learning. Addison–Wesley, 1989.