УДК 621.372:519.72

ИССЛЕДОВАНИЕ ЗВУКОВОГО СТРОЯ НАЦИОНАЛЬНОГО ЯЗЫКА НА ОСНОВЕ ИНФОРМАЦИОННОЙ ТЕОРИИ ВОСПРИЯТИЯ РЕЧИ

© 2010 г. ***В.В. Савченко, Д.Ю. Акатьев, И.В. Губочкин***

Нижегородский государственный лингвистический университет им. Н.А. Добролюбова

svv@lunn.ru

Поступила в редакцию 26.05.2009

Ставится и решается задача автоматического выделения и распознавания элементарных речевых едеництипа фонем из потока устной речи диктора по критерию их минимального информационного рассогласования в смысле Кульбака–Лейблера. Предложен новый алгоритм, основанный на решении последовательности задач R-альтернативной статистической классификации без учителя при переменном и нарастающем значении R = 2, 3, … Рассматривается пример его практического применения в задаче исследования звукового (фонетического) строя речи жителей Нижегородской области.

Ключевые слова: речевой сигнал, информационный центр фонемы, фонетическая база данных, сопоставление диалектов.

Введение

В различных приложениях лингвистики, включая современные речевые базы данных для систем автоматической обработки речи, большие трудности представляет квалификация степени самостоятельности отдельных языковых разновидностей: говоров, диалектов и т.п. относительно друг друга и окружающих их языков [1]. Во многих случаях в практике трактовки и классификации таких разновидностей наблюдается неоднозначное решение вопроса, являются ли они: а) самостоятельными языками; б) диалектами какого-либо единого языка; в) диалектами, составляющими «переходную зону», и т. д. Обычно при классификации языковых разновидностей, составляющих такого типа ареалы, возникает целый ряд проблем, которые решаются различными способами, в зависимости от конкретных установок. При этом часто определяющую роль играет субъективный подход – с позиций той лингвистической школы, к которой принадлежит исследователь. Следует подчеркнуть при этом, что проблема квалификации той или иной языковой разновидности в качестве самостоятельного языка или подчиненного какой-либо языковой общности диалекта возникла сравнительно недавно и связана с качественно новым этапом современной лингвистики, знаменующимся развитием ареальных исследований, функционального подхода к языку, обобщающими работами энциклопедического типа. Указанная проблема актуальна, например, для социолингвистической или функциональной характеристики данной локальной единицы, связанной с ее ролью в коммуникативном процессе в масштабе того или иного сообщества. Проблема возникает и в обобщающих трудах энциклопедического характера, таких, например, каким является многотомное научное издание «Языки мира». Она особенно обостряется в отсутствие для данной речевой единицы единой наддиалектной нормы для сопоставлений. *В представленной статье проблему квалификации языковых единиц из одного ареала предлагается решать на базовом, фонетическом уровне, т.е. на основе выявления и учета их различий по звуковому строю*.

Существование общих фонологических звуковых особенностей во многих соседствующих, но не родственных друг другу языках или диалектах констатировалось не раз [2]. Однако для объяснения этих фактов проявляли поспешность и обращались с этой целью к теории субстрата или к гипотезе о влиянии «ведущего» языка. Такие объяснения ничего не стоят, пока они объясняют лишь единичные случаи. Лучше вообще временно отказаться от всяких гипотез, пока не будет собран достаточный «полевой» материал. По мере увеличения такого материала в последние годы особенно актуальной становится задача его адекватного описания, с целью выяснения фактического состояния вещей в конкретных местностях. *Предлагаемая работа направлена на решение именно такой задачи в рамках нижегородского языкового ареала России.* Причем в ее математической основе используется аппарат информационной теории восприятия речи [3].

Постановка задачи

При анализе устного текста на русском языке мы опираемся на наши точные знания в отношении его фонетического строя, количественного и качественного состава используемой фонетической системы, а также закономерностей ее функционирования в разговорной речи. Этими знаниями мы пользуемся, например, при транскрибировании потока речи. Однако если мы анализируем звучащий текст на неизвестном языке и нам недоступна информация, относящаяся к его тонкой структуре, то мы можем либо, опираясь на наш лингвистический опыт, давать участкам речевого потока приблизительную интерпретацию в рамках международного фонетического алфавита, либо, обратившись к акустическим понятиям, членить речь на некие повторяющиеся минимальные звуковые единицы (МЗЕ) и давать им определенные метки. Очевидно, что второй подход со всех точек зрения наиболее информативен и универсален. Множество меток всех МЗЕ и составит в таком случае звуковой строй данного диалекта или языка.

Проблема состоит в том, что разговорная речь по своим акустическим характеристикам широко варьируется, причем не регулярным образом, не только от одного языка к другому, но и от одного носителя к другому носителю одного и того же языка. В указанных условиях становится проблематичной сама идея выделения *повторяющегося* набора МЗЕ из разговорного потока. Кроме того, длительность отдельных МЗЕ не превышает нескольких миллисекунд, и это главное препятствие для применения традиционных методов теоретической лингвистики к разговорной (устной) речи. С другой стороны, до настоящего времени проблема не была преодолена и методами экспериментальной фонетики. И главная причина здесь – отсутствие *адекватной системы описания отдельных фонем*.

В поисках путей решения указанной проблемы в работах [3, 4] само понятие «фонема» впервые было строго определено в теоретико-информационном смысле как «*множество однородных МЗЕ, объединенных в кластер по критерию минимального информационного рассогласования в метрике Кульбака–Лейблера».* Условно говоря, человеческий мозг объединяет и запоминает как нечто целое (в виде абстрактного образа) разные образцы (произношения) каждой отдельной фонемы в соответствующей «сфере» своей памяти вокруг абстрактного «центра» с заданным «радиусом». Нетрудно понять, что таким определением одновременно решается множество актуальнейших проблем в области фонологического анализа: и вариативности разговорной речи, и априорной неопределенности, и адекватного описания звукового строя языка с кардинальным сжатием данных и, наконец, проблема обновления речевых баз данных без разрушения их структуры.

Элементы ИТВР

Несмотря на существующие различия в реализациях некоторой *r*-й фонемы, все они воспринимаются человеком как нечто общее, иначе речь утратила бы свою информативность. Можно поэтому утверждать, что одноименные реализации $\mathbf{x}_{r,j}$, $j=\overline{1,J_r}, J_r >> 1$, в сознании человека группируются в соответствующие классы или речевые образы фонем $X_r = \{\mathbf{x}_{r,j}\}, r=\overline{1,R}$, вокруг некоторого центра – эталонной метки данного образа. В информационной теории восприятия речи указанные эталоны определяются в строгом, теоретико-информационном смысле [3]: речевая метка $\mathbf{x}_r^* \subset X_r$ образует *информационный центр-эталон r*-го речевого образа, если в пределах множества X_r она характеризуется минимальной суммой информационных рассогласований (ИР) по Кульбаку–Лейблеру относительно всех других его меток-реализаций $\mathbf{x}_{r,j}$, $j=\overline{1,J_r}$.

Нетрудно увидеть, что именно в понятии информационного центра (ИЦ) *r*-го множества реализаций X_r дается наиболее информативное описание свойств соответствующей фонемы. Одновременно становится очевидным и механизм формирования самого этого множества. Анализируемый (входной) речевой сигнал $X(t)$ в дискретном времени $t=0,1,\ldots$ сначала разбивается на ряд последовательных сегментов данных $\mathbf{x}(t)$ длиной в одну МЗЕ: примерно 10 – 15 мс. После этого каждый такой парциальный сигнал рассматривается в пределах конечного списка фонем $\{X_r\}$ и отождествляется с той X_ν из них, которой отвечает минимум информационного рассогласования (МИР) между вектором $\mathbf{x}(t)$ и соответствующим эталоном $\mathbf{x}_\nu^*, \nu \le R$. Это известная [4] формулировка критерия МИР в задачах автоматического распознавания речи.

Задача существенно упрощается, если воспользоваться гауссовой (нормальной) аппроксимацией закона распределения каждой фонемы вида $\mathbf{P}_r = N(\mathbf{K}_r)$, где $\mathbf{K}_r$ – автокорреляци-

онная матрица (АКМ) размера $n \times n$, $n \geq 1$. Задача формулируется в этом случае как проверка простых гипотез о законе распределения МЗЕ. А соответствующий набор оптимальных решающих статистик может быть записан следующим образом:

$$\rho_r(\mathbf{x}) = \frac{1}{2n}\left[\mathrm{tr}\left(\hat{\mathbf{K}} \cdot \mathbf{K}_r^{-1}\right) - \log\left|\hat{\mathbf{K}} \cdot \mathbf{K}_r^{-1}\right| - n\right], \quad r = \overline{1,R}. \qquad (1)$$

Здесь $\hat{\mathbf{K}}$ – это выборочная оценка АКМ анализируемого сигнала $\mathbf{x} = \mathbf{x}(t)$, $t = 0, 1, 2, \ldots$. Решение принимается в пользу некоторой гипотезы $\mathbf{P}_\nu, \nu \leq R$, по признаку минимума ν-й решающей статистики. Ее значение $\rho_\nu(\mathbf{x})$ определяет удельную (на один отсчет данных) величину ИР между выборочным распределением МЗЕ на входе и $\mathcal{V}$-м из гипотетических гауссовых распределений. Причем в задачах с априорной неопределенностью в выражение (1) подставляют статистические оценки альтернативных АКМ по R (по числу фонем в списке) классифицированным выборкам наблюдений. Это стандартная формулировка критерия МИР с обучением.

Адаптивный алгоритм

Предположим, что речевой образ каждой фонемы $X_r = \{\mathbf{x}_{r,j}\}$ представлен по-прежнему конечным (объема $J_r > 1$) множеством своих различных векторов-реализаций $\mathbf{x}_{r,j}$, $j = \overline{1,J_r}$, составленных из L последовательных во времени отсчетов одноименных МЗЕ $\{x_{r,j}(t)\}$ с периодом $T = 1/(2F) = \mathrm{const}$. Здесь F – верхняя граница частотного диапазона речевого тракта. Рассматривая каждую такую реализацию в режиме «скользящего окна» длиной n отсчетов ($n << L$), будем иметь ($L - n$) векторов (столбцов) данных $\{\mathbf{x}_{r,j,i}\}$ размерностью $n = \mathrm{const}$ каждый. Используя после этого формулу среднего арифметического, определим по ним выборочную оценку для АКМ гипотетического гауссова распределения:

$$\hat{\mathbf{K}}_{r,j} \overset{\Delta}{=} (L-n)^{-1} \sum_{i=1}^{L-n} \mathbf{x}_{r,j,i} \cdot \mathbf{x}_{r,j,i}^T, \quad j = \overline{1,J_r}. \qquad (2)$$

Здесь T – символ транспонирования векторов. При этом учитывается естественная центрированность сигналов на выходе стандартного речевого тракта [4].

Подставляя систему оценок (2) в выражение (1), получим в пределах образа X_r матрицу ($J_r \times J_r$) величин информационных рассогласований (ВИР) между одноименными метками:

$$\rho_{j,k}^{(r)} \overset{\Delta}{=} \frac{1}{2n}\left[\mathrm{tr}\left(\hat{\mathbf{K}}_{r,j}\hat{\mathbf{K}}_{r,k}^{-1}\right) - \ln\left|\hat{\mathbf{K}}_{r,j}\hat{\mathbf{K}}_{r,k}^{-1}\right| - n\right], \quad j,k = \overline{1,J_r}. \qquad (1)$$

После суммирования ее элементов по столбцам $\sum_{j=l}^{J_r} \rho_{j,k}^{(r)} \overset{\Delta}{=} \rho_k^{(r)}$, $k = \overline{1,J_r}$, находим информационный центр-эталон речевого образа *r*-й фонемы:

$$\mathbf{x}_r^* \overset{\Delta}{=} \mathbf{x}_{r,\theta} = \underset{k}{\mathrm{Argmin}}\, \rho_k^{(r)}, \quad r = \overline{1,R}. \qquad (3)$$

Его выборочная АКМ $\hat{\mathbf{K}}_{r,\theta}$ из (2) при $j = \theta$ и определяет, в конечном итоге, оптимальную *r*-ю решающую статистику в выражении (1). Отметим, что в зависимости от состава множества X_r положение (значение) эталона (3) будет, разумеется, меняться. Чем больше объем J_r множества, тем устойчивее и, следовательно, точнее определяется соответствующий эталон.

Таким образом, в форме системы выражений (1)…(3) представлен оптимальный по критерию МИР алгоритм автоматического распознавания МЗЕ из слитной речи с предварительным его обучением на заданном R-множестве (списке) фонем $\{X_r\}$.

Проблемы возникают, однако, в случае отсутствия априори классифицированных выборок $\{X_r\}$, т.е. при распознавании образов «без учителя». Автоматический анализ фонетического состава речи чаще всего относится именно к такому кругу задач. И статистические характеристики фонем, и их используемое каждым диктором число R из общего списка зависят от особенностей его речевого аппарата. Здесь требуется алгоритм с самообучением, или адаптивный алгоритм фонетического анализа речи (ФАР). Аналогичная задача рассматривалась в работе [3], в которой для ее решения был предложен информационный $(R+1)$-элемент.

Информационный $(R+1)$-элемент – это условный термин, обозначающий устройство или алгоритм для автоматической классификации или распознавания сигнала $\mathbf{x}$ в пределах некоторого множества классов-альтернатив $\mathbf{P}_r, r = \overline{1,R}$. В основе его функционирования применяется статистический подход и критерий

МИР. В отличие от других аналогичных алгоритмов с R выходами $(R+1)$-элемент имеет дополнительный, $(R+1)$-й выход, который сигнализирует об отказе при распознавании образов одновременно от всех R заданных альтернатив. Указанная особенность и служит основой для построения эффективного алгоритма распознавания образов в условиях априорной неопределенности. Задача сводится к последовательности задач статистической классификации «с учителем» при переменном (нарастающем) числе альтернатив $R = 1, 2,\ldots$

Выделим в анализируемом речевом сигнале $X(t)$ от некоторого диктора первые L отсчетов из соображений сохранения в них свойства приблизительной стационарности или однородности распределения $\mathbf{P}_r$. Например, при стандартной частоте дискретизации телефонного канала связи в 8 кГц обычно полагают [3] L = = 100…200 (это те же 10–15 мс). Используем полученный минимальный сегмент данных $\mathbf{x}_1 = \{x_1,...,x_L\}$ в качестве обучающей выборки X_1 для оценивания по формуле (2) АКМ первой МЗЕ из сигнала. Соответствующий закон распределения $\mathbf{P}_1 = N(\mathbf{C}_1)$ – это первый из элементов нашего будущего списка. После этого приравниваем $R = 1$ и берем второй сегмент выборки для анализа: $\mathbf{x}_2 = \{x_{L+1},...,x_{2L}\}$. Следуя выражению для решающей статистики (1), определим для него удельную величину ИР

$$\rho(X_2, X_r) = \rho_r(\mathbf{x})\big|_{\mathbf{x}=\mathbf{x}_2} \qquad (4)$$

относительно первой МЗЕ (при $r = 1$). Полученный результат сопоставляется с порогом сегментации в роли допустимой величины рассогласований между разными реализациями одних и тех же фонем устной речи:

$$\rho(X_2, X_r) \le \rho_0. \qquad (5)$$

При нарушении данного неравенства в нашем начальном списке фонем появится второй элемент, и вслед за этим приравниваем число выявленных фонем $R = 2$. В противном случае принимается решение об объединении выборок X_1 и X_2 в один речевой образ $\mathbf{P}_1$: в качестве или одной МЗЕ удвоенной длительности $L_r = 2L$, если выборки смежные, или двух разных реализаций первой фонемы, если выборки не стыкуются. Равенство $R = 1$ в обоих случаях сохраняется.

Нетрудно понять, что в форме условия (5) реализуется проверка гипотез об однородности выборок, а понятие фонемы определяется здесь как кластер однородных МЗЕ по критерию МИР. Это типичная формулировка информационного $(R+1)$-элемента.

Фонетический анализ речи

Вычисления по схеме (4), (5) повторяются циклически для всех последующих сегментов данных из речевого сигнала $X(t)$, причем повторятся «нарастающим итогом» для переменного значения R = 2, 3,… Каждый очередной сегмент данных сопоставляется по правилу (5) одновременно со всеми R множествами $\{X_r\}$ из текущего списка фонем. При этом не исключается возможность объединения одного и того же сегмента данных с элементами одновременно нескольких разных множеств. В результате будем иметь список фонем с некоторым фиксированным числом элементов R^*. Это важная характеристика как речевого сигнала, так и самого диктора. Чем больше значение R^* для конкретного диктора, тем богаче с фундаментальной, фонетической точки зрения его речь.

В данном выводе и состоит, по-видимому, главный смысл и назначение фонетического анализа речи (ФАР). Однако здесь же присутствует и очевидная проблема: чрезмерно большое число фонем в речи диктора – это признак ее нечеткости, или неинформативности. С точки зрения качества устной речи первостепенный интерес, безусловно, представляет собой множество четких МЗЕ. Его в таком случае и следует считать основным итогом ФАР. Поэтому логика подсказывает: после выполнения всех перечисленных выше вычислений некоторые «фонемы» из окончательного списка можно исключить как маргинальные.

Следуя такой логике, рассортируем полученные множества $\{X_r\}$ в два вида фонем: четкие и сомнительные. Критерии здесь очевидны: это либо ограничение на минимальную длительность реализаций фонем вида

$$L_r \ge L_0, \qquad (6)$$

либо ограничение на минимальный объем каждого множества

$$J_r \ge J_0. \qquad (7)$$

Здесь L_0, J_0 – некоторые пороговые уровни в каждом случае.

Выражения (4)…(7) в совокупности определяют адаптивный алгоритм формирования R множеств реализаций каждой фонемы для последующего определения по ним соответствующего списка их информационных центров-эталонов (3). В результате создается адаптив-

ная, самонастраиваемая на одного или группу разных дикторов, фонетическая база данных (ФБД) для ее дальнейшего применения в задачах ФАР согласно критерию МИР (1).

Добавим к сказанному, что предложенный алгоритм имеет множество разнообразных модификаций за счет, главным образом, применения рекуррентных вычислительных процедур корреляционно-спектрального анализа [3]. Среди них наибольший интерес представляет метод обеляющего фильтра (МОФ) [4], основанный на авторегрессионной модели МЗЕ.

В работе [5] было показано, что в асимптотике, когда $n \to \infty$, и при гауссовом распределении речевого сигнала $\mathbf{P}_r = N(\mathbf{K}_r)$ с обратной АКМ ленточной структуры выражение для оптимальной решающей статистики из выражения (1) сводится к виду

$$\rho_r(\mathbf{x}) \overset{\Delta}{=} \frac{1}{F+1} \times$$

$$\times \sum_{f=0}^{F} \left(\frac{G_x(f)}{G_r(f)} + \ln \frac{G_r(f)}{G_x(f)} \right) - 1 \to \min.$$

Это известная формулировка критерия МИР на основе авторегрессионной (АР) модели речевого сигнала. Здесь $G_x(f)$ – выборочная оценка спектральной плотности мощности (СПМ) входного сигнала $\mathbf{x}$ в функции дискретной частоты f; а $G_r(f)$ – СПМ эталона r-й фонемы $\mathbf{x}_r^* \subset X_r$.

Главное достоинство АР-модели – это возможность предварительной нормировки речевых сигналов по дисперсиям их порождающих процессов. Применительно к сигналам типа МЗЕ такая нормировка обусловлена физическими особенностями голосового механизма человека: воздушный поток на входе его модели «акустической трубы» [5] имеет приблизительно одну и ту же интенсивность на интервалах длительностью в целое слово или даже фразу. При учете этого свойства последнее выражение приобретает предельно простой вид:

$$\rho_{x,r} = \frac{1}{F+1} \times$$

$$\times \sum_{f=0}^{F} \frac{\left| 1 + \sum_{m=1}^{p} a_r(m) e^{-j\pi mf/F} \right|^2}{\left| 1 + \sum_{m=1}^{p} a_x(m) e^{-j\pi mf/F} \right|^2} - 1 \geq 0. \qquad (8)$$

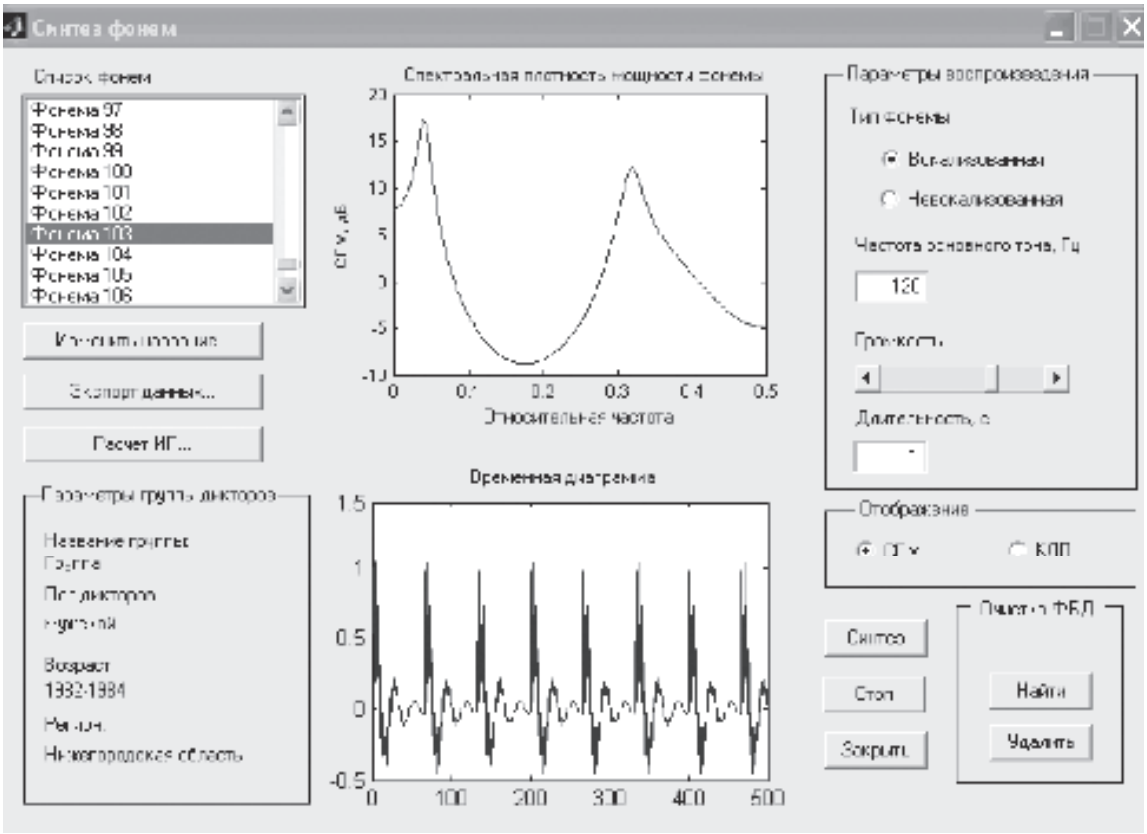

Рис. Интерфейс ИС ФАР

Таблица

Зоны ареала	Юг 18-ж	Юг 40-ж	Север 18-ж	Юг 18-м	Север 18-м	Центр 18-м
Юг 18-ж	0	0.7541	0.5738	1.1303	0.8784	0.8120
Юг 40-ж	0.7239	0	0.6783	1.3232	0.9749	0.9917
Север 18-ж	0.6549	0.8699	0	1.1758	0.7883	0.8397
Юг 18-м	1.0494	1.7584	0.9648	0	0.6544	0.7157
Север 18-м	0.9717	1.5158	0.8647	0.6865	0	0.6787
Центр 18-м	0.8468	1.2635	0.7973	0.7383	0.6942	0

Здесь $\{a_x(m)\}, \{a_r(m)\}$ – два вектора АР-коэффициентов: входного сигнала и *r*-го эталона, оба одного порядка $p > 1$. Это стандартная формулировка МОФ в частотной области. Преимуществом данной интерпретации критерия МИР является, прежде всего, возможность его эффективной реализации в адаптивном варианте на основе быстрых вычислительных процедур АР-анализа, таких как метод Берга и др. [4]. Именно такой вариант МОФ был реализован в дальнейшем для проведения его экспериментальных исследований в типовой задаче ФАР.

Программная реализация

Для реализации предложенного алгоритма (3)...(8) была разработана информационная система фонетического анализа слитной речи (ИС ФАР)[6].

Интерфейс ИС ФАР состоит (см. рис.) из главной формы, на которой показано главное меню программы и список дикторов, внесенных в БД, а также форм для настроек информационной системы, проведения расчетов и отображения результатов.

При выборе любого диктора из списка в правой части окна появляется его краткая характеристика. С помощью имеющегося меню можно выбирать различные режимы работы, загрузки, сохранения и отображения данных.

Процесс создания фонетической базы данных на базе данной ИС выполняется в несколько этапов. На первом этапе формируется группа дикторов и каждый из них проговаривает в среднем темпе некоторый лингвистически сбалансированный текст, например отрывок из художественного произведения. Объем текста составляет минимум (1 … 1.5) тыс. печатных знаков. Каждая такая запись с помощью звукового редактора сохраняется в виде соответствующего звукового файла. На втором этапе в ИС ФАР производится обработка полученных данных по адаптивному алгоритму (3)...(8). В результате формируется множество персональных ФБД $\{X_r\}$, учитывающих особенности разных дикторов. Это первый результат автоматической обработки речевых сигналов. На третьем, заключительном, этапе обработки речевых сигналов для исследуемой группы дикторов формируется отдельная подгруппа по признаку подобия, или похожести, их произношения. В результате отбирается множество соответствующих персональных ФБД. В пределах полученного множества осуществляется объединение выбранных нескольких списков фонем в один общий список $\{X_r\}$ – согласно тому же понятию информационного центра-эталона речевого образа (3) и при учете того же ограничения (5), что и на этапе формирования персональных ФБД.

Программа полевых испытаний

Разработанная методика в период с 1 по 14 декабря 2008 г. была апробирована в полевых (натурных) условиях нижегородского языкового ареала. Указанный ареал был предварительно разбит на три зоны: «юг», «центр» и «север». В зону «юг» вошли Арзамасский и Павловский районы, в «центр» – город Нижний Новгород и в зону «север» – Семёновский район. В пределах каждого района были получены записи разговорной речи от представительной группы дикторов, разделенной в две подгруппы по признаку пола и три – возраста. Всего таким образом было сформировано 18 групп из общего числа 150 дикторов.

Запись одного образца речи по каждому из них имела продолжительность порядка 2 мин. В пересчете к периоду стационарности речевого сигнала (10–15 мс) это составило корпус МЗЕ суммарным объемом около 2.5 миллионов зву-

ковых единиц. Все они были в дальнейшем обработаны по методике ИТВР на персональном компьютере с применением информационной системы [3]. При этом были зафиксированы следующие её основные параметры: порядок АР-модели $p = 20$; порог сегментации $\rho_0 = 1.0$; порог по минимальному объему множества реализаций каждой фонемы $J_0 = 4$; длина сегмента L задается в отсчетах и равна 80. В результате по каждому диктору была сформирована индивидуальная ФБД. Все они были в дальнейшем подвергнуты взаимному информационному анализу в формате матрицы ИР. При этом показателем степени информационных различий (k,l)-пары персональных ФБД служила величина

$$\rho_{kl} = (n+m)^{-1} \times \times \left[\sum_{i=1}^{m} \min_{j=1,n} \rho_{kl}(i,j) + \sum_{j=1}^{n} \min_{i=1,m} \rho_{kl}(i,j) \right] \quad (9)$$

суммы минимальных значений их $(m \times n)$-матрицы ИР $\|\rho_{kl}(i,j)\|$ по строкам и столбцам в расчете на каждое ее слагаемое. По результатам таких вычислений делались выводы относительно фонетических различий в речи жителей как в пределах каждой географической зоны рассматриваемого региона, так и между дикторами разных зон и разных возрастных категорий. Одновременно – по критерию минимума среднего рассогласования (9) в группе – определялся и оптимальный (наиболее информативный) диктор каждой отдельной зоны и региона.

Основные результаты

Исследования были проведены по каждой из 18 контрольных групп дикторов. В итоге были отобраны 18 оптимальных дикторов для дальнейшего анализа, а именно сравнения зон и возрастных категорий между собой по фонетическим различиям в речи их оптимальных дикторов. Для этого использовалась матрица информационных рассогласований (ИР) и критерий средней минимальной величины ИР. Полученные результаты отображаются в виде квадратной (18×18)-матрицы ИР. Ее конечный (6×6)-фрагмент показан в таблице. В ее строках и столбцах перечислены 6 категорий (из 18 рассмотренных) дикторов.

Из таблицы можно сделать множество интересных выводов. Например, это вывод в отношении наибольших фонетических различий между жителями центра и юга Нижегородской области, особенно из разных возрастных групп. Интуитивно понятный вывод о том, что произношение дикторов разного пола существенно различается, получил не только дополнительное подтверждение, но и количественную характеристику по результатам полевых испытаний.

Заключение

К числу приоритетных направлений применения ИТВР наряду с автоматической обработкой речи относятся прежде всего проблемы современной диалектологии. Как сопоставить разные диалекты между собой по степени их объективной близости или различий на базовом, фонетическом, уровне? Какова количественная мера таких различий? Какие тенденции: сближения или удаления по фонетическому строю – доминируют в настоящий момент в процессе исторического развития тех или иных диалектов? И, наконец, как можно лучше обучиться данному диалекту или, напротив, максимально ослабить его? Благодаря методологии ИТВР впервые в мировой науке открываются возможности дать четкие ответы на все перечисленные выше и подобные им вопросы. В их изучении и состоит главная цель предлагаемого исследовательского подхода. Ожидаемые по результатам исследований выводы и обобщения должны составить материал для подготовки к изданию первой фонологической карты России с многоуровневым членением языкового ареала на родственные диалекты при учете степени их звуковых различий, т.е. впервые в практике лингвистического картографирования – с указанием переходных диалектных зон. Осуществление предлагаемого проекта стимулирует, в свою очередь, научные исследования в области не только современной лингвистики, но и информатики в целом, прежде всего прикладной информатики.

Работа выполнена при поддержке РГНФ, проект № 09-06-12125в.

Список литературы

1. Трубецкой Н.С. Фонология и лингвистическая география // В кн.: Избранные труды по филологии. М., 1987. С. 31–36.

2. Яковлев Н.Ф. // Вопросы языкознания. 1983. № 6. С. 128–134.

3. Савченко В.В. // Изв. высших учебных заведений России. Радиоэлектроника. 2007. Вып. 6. С. 3–9.

4. Савченко В.В., Акатьев Д.Ю. // Системы управления и информационные технологии. 2008. № 1 (31). С. 60–66.

5. Савченко В.В. // Радиотехника и электроника. 1997. Т. 42. № 4. С. 426–431.

6. Программа для ЭВМ. Информационная система фонетического анализа слитной речи / В.В. Савченко и др. Роспатент: Свидетельство о гос. регистрации № 2008615442 по заявке № 2008614233 от 15.09.2008.

INVESTIGATION OF THE NATIONAL LANGUAGE SOUND STRUCTURE BASED ON THE SPEECH PERCEPTION INFORMATION THEORY

V.V. Savchenko, D.Yu. Akatiev, I.V. Gubochkin

Among the priority application trends of the speech perception information theory (SPIT) are, above all, the problems of modern dialectology. How to compare different dialects by their proximity or differences on a basic phonetic level? And what is a quantitative measure of such differences? SPIT methodology for the first time in the world science has opened up opportunities for giving rigorous answers to both of these and many other similar questions.

Keywords: speech signal, phoneme information centre, phonetic database, dialect comparison.