АНАЛИЗ ДАННЫХ И ЗНАНИЙ В ЭКСПЕРТНЫХ СИСТЕМАХ
(Вычислительные системы)
1990 год Выпуск 134

УДК 621.31:534.4

ПЕРВИЧНАЯ ОБРАБОТКА СИГНАЛОВ В СИСТЕМАХ РАСПОЗНАВАНИЯ РЕЧИ

А.В. Кельманов

Проблема первичной обработки сигналов является одной из ключевых проблем при построении систем распознавания речи и речевых интерфейсов к экспертным системам. К настоящему времени эту проблему можно считать решенной лишь частично.Цель данной работы состоит в том, чтобы, с одной стороны, дать обзор наиболее часто применяющихся методов первичного описания сигналов, с другой - попытаться сравнить эти методы, с третьей - обобщить исследования автора в данной области и, наконец, - очертить уровень решения данной проблемы и указать нерешенные задачи.

1. Два подхода к выделению признаков

Любая система распознавания речи содержит два основных блока: блок предварительной (первичной) обработки и блок принятия решения. Блок предварительной обработки выполняет функции измерения характеристик помех, обнаружения полезного сигнала, выделения признаков и компенсации помех. В блоке принятия решения осуществляется сравнение контрольной реализации с набором заранее подготовленных эталонов и отождествление этой реализации с ближайшим эталоном.

При построении блока предварительной обработки необходимо решить две основные задачи: что измерять и как измерять? С пер-

вым вопросом связана разработка математической модели сигнала, а со вторым - выбор алгоритма оценивания параметров этой модели.

Модели сигнала, как правило, базируются на данных теории речеобразования [1-6] и некоторых математических допущениях. Например, считается, что речевой сигнал стационарен на участках длительностью 15-40 мс и полностью определяется на этих участках своими моментами до второго порядка включительно. Эти допущения проверены и подтверждены в многочисленных экспериментах по распознаванию и синтезу речи. В связи с этим круг моделей сигнала ограничивается спектральными и корреляционными (временными) моделями и поэтому при выделении признаков стали уже традиционными спектральный и временной подходы.

Выбор алгоритма оценивания является многокритериальной задачей и диктуется, с одной стороны, желанием сократить трудоемкость обработки, а с другой - получить оценки с требуемыми свойствами (состоятельность, несмещенность, эффективность, устойчивость и т.д.). В некоторых случаях в алгоритмы оценивания включаются процедуры обработки сигнала, свойственные слуховому анализатору (выбор сетки частот для оценивания спектральной плотности, эффект маскировки и т.п.) [7-14].

Точно так же, как устройства распознавания речи классифицируются на универсальные (т.е. рассчитанные на произвольный по составу и объему словарь) и проблемно-ориентированные (ограниченный проблемно-ориентированный словарь), системы первичного описания можно разделить на универсальные и усеченные или ограниченные. В последнем случае за счет отказа от универсальности можно использовать упрощенные модели сигнала и сократить до минимума алфавит объектов распознавания [15], что, в свою очередь, позволяет сократить размерность пространства признаков (минимизировать описание), трудоемкость алгоритмов обработки и упростить устройство распознавания в целом. К настоя -

щему времени разработаны сотни подобных систем первичного описания [16-18]. Перечислить их здесь не представляется возможным. Следует лишь отметить, что область применения этих систем описания и основанных на них устройствах распознавания весьма ограничена.

Универсальные же системы первичного описания должны обеспечивать хорошую "разделимость" (по крайней мере, не хуже, чем в слуховом анализаторе человека) всего объективно существующего алфавита элементарных объектов распознавания, т.е. звукотипов, псевдофонем или фонем.

Кроме перечисленных требований, блок предварительной обработки должен обеспечивать минимальную размерность пространства признаков и трудоемкость оценивания. При этом желательными свойствами первичного описания являются: инвариантность к диктору и инвариантность к таким нелинейным преобразованиям (искажениям) сигнала, которые сохраняют разборчивость (например, нелинейные изменения уровня громкости, логарифмическое квантование, клиппирование и т.п.). Легко убедиться, что требование инвариантности приводит к изменению шкалы измерений признаков в сторону ее огрубления. Таким образом, если к разрабатываемой системе распознавания предъявляется требование инвариантности, то в алгоритмах формирования признаков должны быть предусмотрены процедуры перехода к более грубым шкалам (процедуры вторичной обработки).

Предварительная обработка сигнала производится в рамках принятой модели сигнала и может осуществляться во временной и спектральной областях.

Во временной области речь обычно аппроксимируется моделями авторегрессии или авторегрессии скользящего среднего (линейного предсказания). На квазистационарных участках производится оценивание параметров моделей сигнала и помехи. Эти же параметры используются при решении задачи поиска полезного сигнала

(начала и конца команды) при помощи статистических методов обнаружения разладки случайных процессов, описываемых разностными уравнениями [19-21]. Задача компенсации фоновых помех решается методами обратной фильтрации [4,22,23]. Параметры модели (коэффициенты уравнений) являются признаками, которые (иногда подвергаются вторичной обработке) и поступают в блок принятия решения.

При обработке речи в спектральной области основной характеристикой является энергетический спектр сигнала (точнее, спектральная плотность), который также измеряется на квазистационарных участках сигнала. По оценкам спектральной плотности на заранее выбранной сетке частот определяются характеристики помехи и границы полезного сигнала. Для принятия решения о появлении полезного сигнала применяются методы обнаружения момента изменения свойств случайных процессов, имеющих непрерывную спектральную плотность [20,21]. Оценки спектральной плотности используются как при компенсации помех известным методом вычитания энергетических спектров (предполагается, что сигнал и помеха аддитивны), так и в блоке распознавания или принятия решения в качестве признаков. Как и в случае обработки сигнала во временной области, в спектральной зачастую проводится вторичная обработка.

2. Дискретизация, квантование, сегментация

Под речевым сигналом $s(t)$ обычно понимается напряжение, снимаемое с микрофонного усилителя. Этот сигнал является непрерывной функцией непрерывного аргумента. При цифровой обработке значения сигнала $s(t)$ отсчитываются в дискретные моменты времени с периодом (интервалом дискретизации) T. В результате получается сигнал $s(nT)$ с дискретным аргументом (n - целая переменная). В дальнейшем, с целью упрощения, вместо $s(nT)$ будем писать s_n. Последовательность s_n, $n = 0, \pm 1, \pm 2, \ldots$, называют также временным рядом, частота от-

счетов которого определяется формулой $F_s = 1/T$. Для цифрового представления речи необходимо учитывать теорему отсчетов Шеннона: если сигнал занимает полосу частот Ω Гц, то частота дискретизации должна выбираться из условия $F_s \geq 2\Omega$ (2Ω - частота Найквиста). Значения сигнала s_n в дискретные моменты времени имеют непрерывный диапазон изменения. Поэтому перед цифровой обработкой необходимо выполнить операцию квантования, т.е. представить сигнал в виде двоичных слов длиной $\mathcal{N}$ бит. Типичными для речевых исследований являются величины F_s = 6-20 кГц и $\mathcal{N}$ = 8-14 бит.

Выше уже упоминалось, что основным (оправданным практически) допущением при обработке речи является то, что характеристики сигнала не изменяются на интервалах (сегментах или окнах) анализа длительностью T_a от 15 до 40 мс. Поэтому цифровая обработка сигнала ведется последовательно, как правило,с фиксированной шириной окна анализа, содержащего N отсчетов. При этом соседние интервалы анализа перекрываются или примыкают друг к другу. По совокупности выборочных значений $\{s_n\}$, $n = \overline{1,N}$, для каждого интервала анализа оценивается набор, содержащий $p < N$ параметров (признаков). Таким образом, речевой сигнал представляется в виде последовательности векторов, компонентами которых являются параметры выбранной модели.

3. Параметризация во временной области

В основе временного описания сигнала лежит модель авторегрессии проинтегрированного скользящего среднего [24-26], которая в операторной форме может быть записана в виде:

$$A(z^{-1}) \nabla s_n = B(z^{-1}) \xi_n , \qquad (1)$$

где $A(z^{-1}) = 1 - \sum_{i=1}^{p} a_i z^{-i}$ - стационарный оператор авторег-

рессии p-го порядка, $B(z^{-1})=1-\sum_{i=1}^{q} b_i z^{-i}$ - оператор скользящего среднего q-го порядка, z^{-i} - оператор сдвига назад такой, что $z^{-i}s_n = s_{n-i}$, $\nabla = 1-z^{-1}$ - разностный оператор со сдвигом назад, ξ_n - последовательность независимых, одинаково распределенных случайных величин с нулевым матожиданием и дисперсией $\sigma_\xi(0)$. Можно заметить, что модель (1) является линейным фильтром с передаточной функцией $A^{-1}(z^{-1})\,\nabla^{-1}B(z^{-1})$, на вход которой подается белый шум ξ_n, а на выходе наблюдается сигнал s_n.

Вопросам аппроксимации речевого сигнала моделью (1) посвящено достаточно много работ [1,4,5,22,27-34]. Однако широкое распространение в речевых исследованиях получила упрощенная модель проинтегрированной авторегрессии:

$$A(z^{-1})\,\nabla\, s_n = \xi_n . \qquad (2)$$

Основной причиной отказа от введения в модель оператора скользящего среднего $B(z^{-1})$ является то, что, с одной стороны, его использование приводит к существенному увеличению трудоемкости алгоритмов обработки сигнала, а с другой - его отсутствие не приводит к заметному снижению качества и разборчивости речи, восстановленной по модели (2), по сравнению с моделью (1).

Иногда модель еще более упрощают, отказываясь от использования оператора ∇ - оператора взятия первой разности (дискретный аналог дифференцирования) и описывают сигнал в виде простой авторегрессионной модели:

$$A(z^{-1})s_n = \xi_n , \qquad (3)$$

которая во временной области имеет вид:

$$s_n = \sum_{i=1}^{p} a_i s_{n-i} + \xi_n \ . \qquad (4)$$

В этой модели оцениванию подлежат p параметров авторегрессии (линейного предсказания) $\{a_i\}$ и дисперсия $\sigma_\xi(0)$. Имеется большое количество разнообразных алгоритмов оценивания параметров авторегрессии, вытекающих из различных постановок задачи оценивания: оценки максимального и условного правдоподобия, среднеквадратичные оценки, оценки по методу ковариаций и корреляций, рекуррентные оценки и т.п. [1,4,22,24-29,34]. Не вдаваясь в детальное обсуждение перечисленных подходов, отметим, что отличия в алгоритмах можно легко увидеть из следующих соображений. Если обозначить через

$$\mathrm{cov}[s_{n-i}, s_{n-j}] = \sigma_s(i-j) = M[s_{n-i}, s_{n-j}] \qquad (5)$$

ковариации процесса s_n (здесь и далее полагается, что сигнал s_n имеет нулевое среднее), то из (5) и (4) можно получить систему нормальных уравнений:

$$\sum_{i=1}^{p} a_i \sigma_s(i-j) = \sigma_s(j), \quad 1 \le j \le p, \qquad (6)$$

в которую входят теоретические ковариации. Многообразие алгоритмов оценивания можно теперь объяснить разнообразием оценок ковариаций или корреляций (после нормировки (6)), которые подставляются в (6) для нахождения оценок коэффициентов авторегрессии. Эти алгоритмы при небольших объемах выборки N дают, вообще говоря, различные результаты. Однако все они асимптотически эквивалентны [24] в том смысле, что для любого из них среднеквадратичная ошибка оценивания есть величина $O(1/N)$. При традиционных окнах анализа отличия в оценках становятся практически малозаметными. Поэтому при выборе способа оценивания на первый план выходят вычислительные аспекты обработки.

Одними из наиболее часто используемых являются оценки Юла-Уокера [24,25], которые находятся из системы уравнений:

$$\sum_{i=1}^{p} \hat{a}_i r_{j-i} = r_j, \quad 1 \le j \le p, \tag{7}$$

где

$$r_j = \frac{\hat{\sigma}_s(j)}{\hat{\sigma}_s(0)} = \frac{\sum_{n=1}^{N-j} s_n s_{n+j}}{\sum_{n=1}^{N} s_n^2}, \quad 0 \le j \le p, \tag{8}$$

оценки корреляций сигнала s_n. Эффективным методом решения системы уравнений (7) является метод Дурбина [24,25]. При использовании этого метода требуется $2p$ ячеек памяти и $p^2 + O(p)$ арифметических операций. Рекуррентные формулы имеют вид:

$$\left.\begin{aligned} \hat{a}_{j+1,i} &= \hat{a}_{ji} - \hat{a}_{j+1,j+1} \times \hat{a}_{j,j-i+1}, \\ \hat{a}_{j+1,j+1} &= \frac{r_{j+1} - \sum_{i=1}^{j} a_{ji} r_{j-i+1}}{1 - \sum_{i=1}^{j} a_{ji} r_i}, \\ i &= 1,2,\ldots,p; \quad \hat{a}_{11} = r_1 . \end{aligned}\right\} \tag{9}$$

Метод особенно удобен тем, что обеспечивает возможность простой проверки устойчивости по критерию Лемера-Шура [4,5] (проверка попадания нулей характеристического полинома в единичную окружность), который сводится к проверке неравенства:

$$|\hat{a}_{j+1,j+1}| < 1, \quad 0 \le j \le p-1. \tag{10}$$

Следует отметить, что теоретически для отличного от нуля сигнала s_n матрица системы (7) всегда положительно определена

[35]. Поэтому решение системы всегда так определяет характеристический полином $A(z)$, что его корни лежат внутри единичной окружности [24,35], т.е. теоретически устойчивость гарантируется. Однако проверка устойчивости необходима в том случае, если при оценивании используются вычислители с малым числом разрядов, т.е. в том случае, когда ошибки округления могут привести к плохой обусловленности матрицы уравнений (7).

Величины $\hat{a}_{j+1,j+1}$, $j=0,1,\ldots,p-1$, являются оценками частных корреляций между s_n и $s_{n-(j+1)}$ при фиксированных (оцененных) значениях $s_{n-1},\ldots,s_{n-j}$ [24]:

$$a_{j+1,j+1}=\frac{M[(s_n-\hat{s}_n)(s_{n-(j+1)}-\hat{s}_{n-(j+1)})]}{\{M(s_n-\hat{s}_n)^2 M[s_{n-(j+1)}-\hat{s}_{n-(j+1)}]^2\}^{\frac{1}{2}}}. \quad (11)$$

Частные корреляции в отличие от однозначно связанных с ними коэффициентов авторегрессии обладают целым рядом важных с практической точки зрения свойств. Нетрудно заметить, что они получаются путем ортогонализации множества параметров авторегрессии $\{a_{jj}\}$, $j=\overline{1,p}$. Для процесса авторегрессии p-го порядка $a_{jj}\neq 0$ при $j\leq p$ и $a_{jj}=0$ для всех $j>p$. При этом оценки $\hat{a}_{jj}$ для $j>p$ имеют дисперсию $O(1/N)$ и распределены асимптотически нормально с нулевым средним [24]. Эти свойства можно применять при автоматическом поиске числа независимых переменных в модели авторегрессии, т.е. для выявления порядка модели или размерности пространства признаков, используя методы статистической проверки гипотез. Перечисленные свойства сделали частные корреляции одними из наиболее популярных параметрических представлений речевого сигнала в системах анализа и синтеза речи.

Свойство (10) делает частные корреляции весьма удобным способом описания и в вычислительном аспекте, при реализации

обработки на процессорах с фиксированной арифметикой. Следует отметить, что с уменьшением отношения сигнал/помеха точность вычислений частных корреляций и коэффициентов авторегрессии начинает падать особенно с ростом порядкового номера параметра. Происходит это по той причине, что при оценивании применяется процедура обращения матриц, обусловленность которых ухудшается - ся, а также потому, что используется рекуррентная процедура вычисления очередного параметра по всем предыдущим.Поэтому, од - нажды появившись, ошибка вычислений растет с ростом порядка рекурсии. Распространение ошибки в зависимости от порядка модели p можно записать в виде [36]: $\epsilon(p) = 0.00005^{0.2p}$.

С частными корреляциями связан набор параметров,описывающих поперечные сечения однородной акустической трубы без потерь [1,4]. Можно установить эквивалентность между процессами фильтрации в авторегрессионном фильтре (лестничной структуры) и процессом прохождения звуковой волны через такую трубу [1,4]. При этом частные корреляции интерпретируются как коэффициенты отражения в стыках секций трубки, а отношения площадей между смежными секциями могут быть записаны в виде:

$$\frac{Q_{j+1}}{Q_j} = \frac{1-a_{jj}}{1+a_{jj}}, \quad 1 \le j \le p. \qquad (12)$$

Логарифм этого отношения является оптимальным преобразованием для проведения равномерного квантования параметров [1,4]. Данное описание часто применяется в вокодерных системах передачи речи, а в системах распознавания широкого применения не нашло.

Кроме рассмотренных выше параметров,для описания речи во временной области используют кепстральные коэффициенты, кото - рые однозначно связаны с параметрами авторегрессии [1,37]. Но эти коэффициенты на практике оказываются весьма чувствительными к диктору. Поэтому их чаще применяют в системах автоматической идентификации и верификации личности по голосу.

Наконец, отметим, что оператор $A(z^{-1})$ в (3) как полином может быть представлен в виде суммы двух полиномов одной и той же степени p. Один из способов подобного разложения предложен в [38,39]. Коэффициенты этих полиномов также могут быть использованы в качестве признаков. Однако число этих коэффициентов оказывается в 2 раза большим. Поэтому на практике вместо них используют пары корней указанных полиномов, т.е. фактически переходят к спектральному представлению или представлению в Z-плоскости. Последнее представление известно как описание в виде "линейных спектральных пар" [38,39] и применяется в системах синтеза речи. Возможность распознавания по этим параметрам находится в стадии исследования.

4. Обработка речи в спектральной области

Как было показано выше, обработка речи во временной области обычно связана с аппроксимацией сигнала какой-либо математической моделью. При этом процедуры восстановления сигнала по параметрам этой модели фактически становятся прототипами процессов речеобразования. Поэтому параметризация во временной области в бóльшей степени отражает процесс генерации речи и позволяет интерпретировать процесс речеобразования, а также использовать физиологические данные при построении модели.

Целью обработки речи в спектральной области является получение оценок спектральной плотности сигнала. При этом спектральная обработка, наоборот, в бóльшей степени связана с процессами восприятия речи, что позволяет интерпретировать процесс анализа речи ухом и использовать свойства слухового аппарата человека в качестве различного рода эвристик в формальных методах спектрального анализа.

4.1. Непараметрические оценки. Оценивание спектральной плотности по наблюдаемому речевому сигналу является, в сущности, непараметрическим методом, поскольку при этом по конечно-

му множеству наблюдений временного ряда на конечном отрезке оценивается функция, которая не определяется конечным числом параметров. Как известно, корреляционная функция и спектральная плотность стационарного процесса связаны по Фурье [24]:

$$\left.\begin{aligned} \sigma(k) &= \int_{-\pi}^{\pi} E(\omega)\cos k\omega d\omega = \int_{-\pi}^{\pi} E(\omega) e^{j\omega k} d\omega , \\ & \qquad k = 0, \pm 1, \pm 2, \ldots , \\ E(\omega) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \sigma(k)\cos k\omega = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \sigma(k) e^{j\omega k} , \\ & \qquad -\pi \le \omega \le \pi . \end{aligned}\right\} \qquad (13)$$

Если положить

$$\left.\begin{aligned} C(\omega) &= \frac{2}{N} \sum_{n=1}^{N} s_n \cos n\omega, \quad D(\omega) = \frac{2}{N} \sum_{n=1}^{N} s_n \sin n\omega, \\ R^2(\omega) &= C^2(\omega) + D^2(\omega) , \quad -\pi \le \omega \le \pi, \end{aligned}\right\} \qquad (14)$$

то выборочную спектральную плоскость можно записать в виде:

$$\hat{E}(\omega) = \frac{N}{8\pi} R^2(\omega) = \frac{1}{2\pi N} \left| \sum_{n=1}^{N} s_n \exp(j\omega n) \right|^2 , -\pi \le \omega \le \pi. \qquad (15)$$

При этом выборочные ковариации

$$c_k = \frac{1}{N} \sum_{n=1}^{N-k} s_n s_{n+k} , \quad k = 0, 1, \ldots, N-1 ,$$

и выборочные спектральные плотности оказываются связанными точно так же, как и в исходном процессе:

$$\hat{E}(\omega) = \frac{1}{2\pi} \sum_{k=-(N-1)}^{N-1} c_k \cos k\omega, \quad -\pi \le \omega \le \pi, \quad \left.\begin{matrix} \\ \\ \vdots \end{matrix}\right.$$

$$c_k = \int_{-\pi}^{\pi} \hat{E}(\omega)\cos k\omega d\omega, \quad k = 0, \pm 1, \ldots, \pm(N-1). \qquad (16)$$

При цифровой обработке вычисления $\hat{E}(\omega)$ обычно производятся для $\omega = 2\pi k/N$, $k=0,1,\ldots,N/2$, как правило, путем быстрого преобразования Фурье (БПФ).

Выборочные спектральные плотности подвержены значительной выборочной вариабельности, которая возникает в силу того, что $\hat{E}(\omega)$ не является состоятельной оценкой для $E(\omega)$: несмотря на то, что при $N \to \infty$ математическое ожидание оценки сходится к теоретической спектральной плотности, дисперсия оценки не стремится при этом к нулю. Поэтому для оценивания спектральной плотности применяют специальные приемы, заключающиеся в сглаживании с применением корреляционных и спектральных окон. Такие оценки можно записать в виде:

$$\hat{E}_c(\omega) = \frac{1}{2\pi} \sum_{k=-(N-1)}^{N-1} w_k c_k \cos k\omega =$$

$$= \int_{-\pi}^{\pi} W(\omega - \omega^*)\hat{E}(\omega^*)d\omega^*, \qquad (17)$$

где w_k и $W(\omega)$ - корреляционные и спектральные окна соответственно. Наиболее распространенными оценками являются (подробнее см.[24,26]): усеченная, Бартлетта, Даниэля, обобщенная Блэкмена-Тьюки, Парзена, а также оценки Блэкмена-Тьюки с использованием окна Хэмминга и Хэннинга. Перечисленные виды оценок асимптотически эквивалентны в том плане, что все они позволяют состоятельно оценить спектральную плотность.

Все существующие методы обработки речи в спектральной области (спектрально-полосные, формантные, гармонические и т.п.) основаны на получении оценок вида (17). Отличия заключаются в способе оценивания $E(\omega)$ или в том, как оно производится: непосредственно по выборке из (15) или по оценкам ковариаций из

(16), а также в типе окна анализа для вычисления (17). Здесь уместно отметить, что и выделение признаков (энергий или мощностей) при помощи полосовых (цифровых или аналоговых) фильтров является разновидностью оценок (17).

С другой стороны, отличия в методах спектральной обработки объясняются разнообразием сеток частот или точек, в которых производится оценивание спектральной плотности. Выбор этой сетки, как будет показано ниже, влияет на информативность и размерность пространства признаков, а также надежность распознавания. Известно по крайней мере три способа задания сетки частот. Это равномерный способ, когда оценивание производится в эквидистантных точках; формантный, когда оценивается от 3-х до 5-ти значений спектральной плотности в точках,которые соответствуют формантным частотам (этот способ опирается на данные теории речеобразования) и неравномерный, когда сетка частот устанавливается в соответствии с некоторым законом (этот способ опирается на данные теории восприятия речи или на статистические свойства самих оценок).

Следует отметить, что с вычислительной точки зрения вместо оценок спектральной плотности по ковариациям, т.е. по формуле (16), гораздо более выгодно применять быстрое преобразование Фурье к формуле (15). Поскольку признаки не должны зависеть от уровня громкости сигнала,необходимо оценивать нормированную выборочную спектральную плотность $\widetilde{E}(\omega) = E(\omega)/\sigma(0)$ и вычислять нормированные оценки спектральной плотности.

В большинстве случаев оценки спектральных плотностей представляются в логарифмическом или близком к нему масштабе.В речевых исследованиях стало уже традицией применение такого масштаба объяснять свойствами слухового анализатора человека. Точное же математическое обоснование этого масштаба состоит в том, что асимптотическая дисперсия логарифмов выборочных спектральных плотностей (16) не зависит от значений самих плотностей

[24], что позволяет применять простые в вычислительном плане метрики в блоке принятия решения.

4.2. Параметрическое сглаживание оценок. В первом приближении речь образуется в результате свертки функций возбуждения и импульсной реакции речевого тракта (без учета характеристики излучения) [2,3]. Поэтому спектральную плотность сигнала можно представить в виде:

$$E(\omega) = G(\omega)H(\omega) = V(\omega)Q(\omega), \qquad (18)$$

где $G(\omega)$ - спектральная плотность источника возбуждения; $H(\omega)$ - амплитудно-частотная характеристика речевого тракта; $V(\omega)$ - спектральная плотность идеализированного источника возбуждения (периодический или линейчатый спектр для голосового источника возбуждения и равномерный - для неголосового); $Q(\omega)$ - так называемая огибающая спектра сигнала. Фонетическая информация, за исключением характеристик вокализованный/невокализованный, содержится главным образом в огибающей спектра. Поэтому при спектральной обработке речи характеристики источника возбуждения становятся мешающими. Мешающий характер источника сказывается в том, что результаты анализа речи зависят от величины периода основного тона или от расстояния между линейками энергетического спектра.

Для устранения влияния источника возбуждения на результаты анализа можно применять различные методы оценивания огибающей энергетического спектра или аппроксимацию этого спектра спектром модели. Примером такой обработки является аппроксимация спектральной плотности речевого сигнала спектральной плотностью модели авторегрессии [1,4-6]. В основе этой аппроксимации лежит тот факт [24], что непрерывную спектральную плотность можно сколь угодно точно (путем увеличения порядка модели) аппроксимировать спектральной плотностью процесса авторегрессии. Последняя из (3) и (13) может быть записана в виде:

$$E_M(\omega) = \frac{\sigma_\xi(0)}{2\pi\left|1 - \sum_{k=1}^{p} a_k e^{-j\omega k}\right|^2}, \quad -\pi \le \omega \le \pi \ . \qquad (19)$$

Здесь $\sigma_\xi(0)/2\pi$ - спектральная плотность белого шума.

Порядок модели выбирается обычно из тех соображений, что на каждый пик спектральной плотности речевого сигнала приходится по крайней мере 2 параметра авторегрессии, т.е. модель авторегрессии второго порядка (резонатор) [4]. Поэтому порядок аппроксимирующей модели должен быть не меньше удвоенного числа пиков в спектральной плотности речевого сигнала. Иногда порядок модели оценивается при обработке адаптивно при помощи статистических критериев, основанных на свойствах оценок параметров авторегрессии [5,40].

По существу, данный метод обработки является параметрическим, поскольку для оценивания спектральной плотности необходимо предварительно вычислить параметры уравнения авторегрессии (4) и подставить их в (19). Вычисления $E_M(\omega)$ можно проводить двумя путями.

Непосредственно из (19) видно, что оценку $\hat{E}_M(\omega)$ можно получить путем деления $\hat{\sigma}_\xi(0)/2\pi$ на квадрат величины, полученной из последовательности $1, -a_1, -a_2, \ldots, -a_p$ с помощью быстрого преобразования Фурье. При визуальном анализе для получения бо́льшей разрешающей способности по частоте эту последовательность следует дополнить необходимым числом нулей. Если предполагается, что размерность пространства (спектральных) признаков может быть больше p, то длину последовательности путем добавления нулей следует увеличить до N, т.е. до размера исходной выборки. Дальнейшая обработка строится в соответствии с формулой (17) путем подстановки в нее вычисленных оце-

нок $\hat{E}_M(\omega)$. При этом, как правило, используется прямоугольное окно, соответствующее идеальному полосовому фильтру.

Другой способ получения сглаженных оценок можно получить, если переписать (19) в виде:

$$E_M(\omega)= \frac{\sigma_\xi(0)}{2\pi A(\omega)} , \quad -\pi \le \omega \le \pi, \tag{20}$$

где $A(\omega)$ - амплитудно-частотная характеристика обратного (к авторегрессионному) фильтра, т.е. такого фильтра, у которого вход и выход поменяны местами. Из (4) путем обращения сигнал на выходе такого фильтра можно представить в виде:

$$h_n = \xi_n - \sum_{i=1}^{p} a_i \xi_{n-i} . \tag{21}$$

Если на вход такого фильтра поступает белый шум ξ_n с единичной дисперсией, то спектральная плотность выходного процесса будет равна его амплитудно-частотной характеристике. Поэтому из (13) имеем:

$$A(\omega)= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \sigma_h(k)\cos k\omega, \tag{22}$$

где $\sigma_h(k)$ - ковариации выхода обратного фильтра, когда на входе белый шум с единичной дисперсией. Из определения ковариаций, свойств белого шума и формулы (21) можно получить:

$$\sigma_h(k)= \sum_{i=0}^{p-k} a_i a_{i+k} , \quad 0 \le k \le p , \quad a_0 = 1 . \tag{23}$$

Поэтому

$$E_M(\omega)= \frac{\sigma_\xi(0)}{\sum_{k=-\infty}^{\infty} \sigma_h(k)\cos k\omega} = \frac{\sigma_\xi(0)}{\sigma_h(0)-2\sum_{k=1}^{\infty} \sigma_h(k)\cos k\omega} =$$

$$= \frac{\sigma_\xi(0)}{\sigma_h(0) - 2\sum_{k=1}^{p-k} \sigma_h(k)\cos k\omega}. \qquad (24)$$

А выборочные оценки получаются из (23) и (24) после подстановки в них оценок $\hat{\sigma}_\xi(0)$ и $\{\hat{a}_i\}$, $i = \overline{1,p}$.

4.3. Эвристические приемы сглаживания. Существует еще целый ряд эвристических способов получения сглаженных оценок спектральной плотности [12-14,41,42]. Суть этих способов состоит в том, что перед вычислением сглаженной спектральной оценки выборочная спектральная плотность речевого сигнала преобразуется или корректируется по некоторому разумному закону или правилу. Обычно в выборочном спектре речевого сигнала визуально выделяется от трех до пяти пиков. Возрастающе-убывающий характер выборочной спектральной плотности около этих пиков заменяют "стандартными" прямыми линиями или монотонными кривыми, параметры которых не изменяются на протяжении всего речевого сигнала. Эта процедура напоминает "аппроксимацию", но таковой в строгом смысле не является. Параметры кривых выбираются из эвристических соображений так, чтобы, с одной стороны, наклон кривых не был слишком крутым и не привел к образованию глубоких впадин и нулевых значений в скорректированной выборочной спектральной плотности (в огибающей спектра), а с другой - чтобы убывание этих кривых в окрестности максимумов не было слишком медленным и не привело к поглощению соседних пиков. Иногда параметры этих кривых получают путем усреднения по группе дикторов.

В работе [41] для подобной коррекции применялась кусочно-линейная "аппроксимация", в [42] - кусочно-полиномиальная, а в [14] - "аппроксимация" гауссовоподобными кривыми. Причем в последнем случае параметры этих кривых подбирались так, чтобы эти кривые моделировали эффект маскировки, присущий слуховому анализатору человека. После проведения коррекции выборочной

спектральной плотности для получения сглаженных оценок используется формула (17), т.е. традиционный метод.

Цель, которую преследуют перечисленные методы корректировки, состоит в устранении линейчатости энергетического спектра и в построении системы признаков, инвариантных к диктору. Эта цель была бы достигнута, если бы вариации спектральной плотности от диктора к диктору заключались только в наклонах спектральной плотности. Многочисленные же экспериментальные данные по восприятию синтезированной речи свидетельствуют об обратном, т.е. о том, что крутизна наклонов спектральной плотности в окрестности ее пиков не имеет принципиального значения для индивидуальных особенностей речи. В гораздо большей степени эти особенности заключены (не считая частоты основного тона) в положении спектральных максимумов на частотной оси. Как следствие применение сглаживающих процедур подобного рода хотя и позволяет повысить надежность распознавания речи (с 60% при распознавании без подстройки под диктора), но не в той степени, которая приемлема для практики: для фонетически сбалансированных словарей получена надежность распознавания около 90% [14]. Следует также учесть, что процедуры корректировки заметно увеличивают трудоемкость алгоритмов обработки речи в целом. Поэтому при построении систем распознавания описанные методы получения оценок спектральной плотности не находят своего применения.

4.4. Выбор числа и границ спектральных полос. При спектрально-полосном распознавании речи возникает задача выбора числа и границ спектральных полос, в которых формируются признаки (спектральные энергии). Эта задача эквивалентна задаче выбора сетки частот, на которых будет проводиться оценивание спектральной плотности в соответствии с (17), и поиску оптимальной ширины спектрального окна $W(\omega)$ для сглаживания. Указанную задачу можно решать разными путями. Первый состоит в моделиро -

вании слухового анализатора и формировании такого разбиения частотного диапазона сигнала, каким пользуется человек, второй - в формулировании и оптимизации разумного формального критерия, учитывающего статистические свойства оценок спектральной плотности, и, наконец, третий - в задании сетки частот, исходя из данных спектральной теории речеобразования, т.е. учете неравнозначности (в информационном смысле) значений спектральной плотности речевого сигнала в зависимости от частоты.

Каждый из способов задания сетки частот имеет свои преимущества и недостатки. Выбор того или иного способа зачастую диктуется вкусами разработчика. Однако необходимо всегда помнить, что от задания сетки частот и границ спектральных окон зависят свойства оценок спектральной плотности и, в конечном счете, надежность распознавания. В свою очередь, от свойств этих оценок зависит вид меры сходства между векторами в спектрально-полосном пространстве признаков, которая будет задействована в блоке принятия решения для вычисления расстояния. Иными словами, мера сходства должна быть согласована с признаками. Вид же меры сходства влияет на время принятия решения: чем проще вычисляется мера сходства или расстояние, тем быстрее будет реакция системы. Поэтому при формальной эквивалентности всех способов задания сетки частот и границ спектральных окон для фиксированной надежности распознавания предпочтение следует отдавать тому способу, который приводит к наиболее простой в вычислительном плане метрике.

Если нет никаких априорных сведений о спектральных характеристиках сигнала, то оценивать спектральную плотность можно на равномерной сетке частот. При этом для обеспечения априорной равной информативности всех значений спектральной плотности спектральные окна или полосы должны примыкать друг к другу и, по возможности, обеспечивать минимальное просачивание энергий в соседние полосы для того, чтобы результаты оценивания на соседних частотах были независимы. Таким образом, в данном слу-

чае границы частотных полос легко определяются по верхней и нижней частоте частотного диапазона и числу полос.

С технической точки зрения следует придерживаться равномерных в логарифмическом частотном масштабе полос или окон, примыкающих друг к другу. Эти полосы принято называть октавными полосами. Обоснованием к подобному разбиению может служить тот факт, что при этом (по сравнению с обычной равномерной шкалой частот) улучшается качество звучания и разборчивость синтезированной полосными методами речи [43]. Границы частотных полос в данном случае определяются по нижней граничной частоте анализируемого диапазона частот, ширине первого спектрального окна и числу таких окон или числу точек для оценивания спектральной плотности.

Специалисты по восприятию речи рекомендуют придерживаться равномерной шкалы полос в масштабе Барк, поскольку такая шкала соответствует шкале слухового анализатора [44,45]. Экспериментальная зависимость Барк-Гц может быть аппроксимирована функцией:

$$\mathcal{B} = 6{,}7\,\mathrm{Arsh}\left[\frac{\Omega - \Omega_н}{600}\right], \qquad (25)$$

где $\mathcal{B}$ - тональность в Барк, Ω - частота в Гц, $\Omega_н$ - нижняя граница частотного диапазона в Гц; погрешность аппроксимации в диапазоне от 0.15 до 5 кГц не превосходит 3% по тональности и 5% по частоте. Исходя из приведенного выражения, по заданному числу полос и нижней и верхней граничных частот нетрудно вычислить границы частотных полос для формирования спектрально-полосных признаков. Для $\Omega_н$ = 110 Гц и верхней частоте 5000 Гц границы частотных полос приведены в табл.1.

С точки зрения теории речеобразования спектральную плотность следует оценивать в точках, соответствующих резонансным характеристикам речевого тракта. Считается, что бóльшая часть

Таблица 1

Число полос	Границы частотных полос в Гц
1	110-1249-5000
3	110-753-1997-5000
4	110-564-1249-2514-5000
5	110-463-930-1659-2884-5000
6	110-400-753-1249-1997-3161-5000
7	110-356-642-1013-1531-2278-3375-5000
8	110-324-564-860-1249-1779-2514-3545-5000
9	110-299-507-753-1062-1464-1997-2713-3683-5000
10	110-280-463-674-930-1249-1659-2190-2884-3797-5000
11	и т.д.

информации о речевом сообщении содержится именно в таких точках, называемых формантными частотами. Обычно оперируют с тремя-пятью значениями формантных частот. Выбор границ спектральных окон около формантных частот производится из эвристических соображений. Эти окна или полосы называют формантными. Поскольку резонансные характеристики речевого тракта изменяются с течением времени от звука к звуку, определение сетки частот в данном случае должно производиться для каждого участка анализа сигнала или сегмента, что сильно увеличивает трудоемкость алгоритмов обработки. К тому же значения формантных частот зависят от диктора. Поэтому при построении устройств распознавания речи формантная сетка частот практически не используется.

Перечисленные способы задания сетки частот носят эвристический характер. Можно определить эту сетку и формальными методами, исходя из статистических свойств оценок спектральной плотности. Как известно [24], эти оценки асимптотически некоррелированы и нормально распределены и, следовательно, асимптотически независимы. Поэтому в пределе совместная плотность распределения оценок-признаков может быть представлена в виде произведения одномерных плотностей, что при байесовском подходе к распознаванию для признаков, имеющих одинаковые диспер -

сии, приводит к евклидовой метрике. Однако на самом деле дисперсия оценок спектральной плотности зависит от самих значений спектральной плотности и имеет порядок $1/N$. Поэтому перед вычислением евклидовой метрики для произвольной сетки частот признаки должны нормироваться на дисперсию. Можно, однако, так выбрать сетку частот, чтобы дисперсия оценок была одинаковой. Обработка речевого сигнала показала [11], что примерно равные дисперсии оценок получаются в том случае, если сетку частот выбирать так, чтобы расстояния между точками увеличивались с ростом частоты нелинейно. При этом оказалось, что масштаб сетки частот будет приближенно соответствовать масштабу, рекомендуемому техниками и специалистами по теории речевосприятия, т.е. сетка частот должна быть равномерной в логарифмическом масштабе или масштабе Барк.

5. Исследование методов первичного описания

Первые успехи в области распознавания речи были связаны со спектрально-полосным подходом к выделению признаков. Затем широко применялись модели речевого сигнала во временной области, а именно модели авторегрессии (линейного предсказания) и авторегрессии скользящего среднего. В последние годы для выделения признаков все чаще снова применяются спектрально полосные методы. Причины возврата к спектральным методам станут ясны из дальнейшего изложения.

5.1. Обзор исследованных методов. Для того чтобы получить общую картину о наиболее популярных методах выделения признаков, были проведены сравнительные эксперименты по распознаванию изолированных команд (по возможности в одних и тех же условиях и на одном и том же речевом материале). В процессе проведения работ было исследовано около трех десятков различных систем первичного описания сигнала. Все результаты получены путем моделирования на универсальных ЭВМ. При этом речевой сиг-

нал вводился в ЭВМ через АЦП, имеющие 7 или 8 разрядов с использованием конденсаторного микрофона МД-59. Отношение сигнал/шум составляло величину, равную примерно 30 дБ. Получены данные по трудоемкости выделения признаков и надежности распознавания для каждой из систем. Кроме того, моделировались алгоритмы первичной обработки ряда отечественных и зарубежных систем. Основу алгоритмов распознавания составляла процедура динамического программирования. Ниже перечислены исследованные системы описания. Для отельных систем в скобках дополнительно указаны номера. Это сделано с целью пояснения результатов экспериментов, описанных в пп. 5.2-5.4.

1. Частные корреляции модели авторегрессии [46-48]:

A) без использования первых разностей,

B) со стационарной разностью первого порядка,

C) с адаптивной разностью первого порядка.

2. Те же 3 группы признаков, что и в п.1, дополненные бинарным признаком тон/шум [47,48]:

3. Те же 3 группы признаков, что и в п.1, дополненные признаком нормированная частота основного тона [47,48].

4. Частные корреляции модели авторегрессии со стационарной разностью [47-48]:

A) без использования первых разностей,

B) со стационарной разностью первого порядка,

C) с адаптивной разностью первого порядка.

5. Те же 3 группы признаков, что и в п.4, дополненные бинарным признаком тон/шум [47,48].

6. Те же 3 группы признаков, что и в п.4, дополненные признаком нормированная частота основного тона [47,48].

7. Спектрально-полосные признаки [10,11,40,49,52]:

A) энергия сигнала в полосах БПФ-спектра [10,11,49,52]:

AA) без использования первых разностей,

ВВ) с использованием стационарной разности первого порядка (далее система № 1),

В) энергия сигнала в полосах авторегрессионного спектра [10,11,40,52]:

АА) с фиксированным порядком модели авторегрессии:

ААА) без использования первых разностей,

ВВВ) с использованием стационарной разности первого порядка (далее система № 2),

ССС) с использованием адаптивной разности первого порядка,

ВВ) с адаптивным выбором порядка модели авторегрессии:

ААА) без использования первых разностей,

ВВВ) с использованием стационарной разности первого порядка,

ССС) с использованием адаптивной разности первого порядка.

8. Бинарные спектрально-полосные признаки, основанные на [50-52]:

А) дихотомическом разбиении частотного диапазона с использованием:

АА) БПФ-спектра,

ВВ) сглаженного спектра (два варианта: системы №3 и № 4) ,

В) знаке производной или 'наклоне":

АА) БПФ-спектра,

ВВ) сглаженного спектра (далее система № 5).

(Подробные характеристики этих систем приведены в работах, на которые даны ссылки.)

Сравнительный анализ экспериментальных результатов позволил упорядочить различные системы описания по надежности распознавания и трудоемкости. При этом оказалось, что, как правило, лучшие результаты по надежности дают системы признаков,имею-

щие бо́льшую трудоемкость. Забегая вперед, следует отметить,что опыт работы с различными первичными описаниями позволил уста - новить, что по надежности распознавания и спектральный, и временной подходы дают практически равноценные результаты.

5.2. Выбор оптимального числа полос.Минимальность первичного описания всегда желательна. Известно [48], что при временном описании для получения приемлемых результатов необходимо использовать не менее 10-16 авторегрессионных параметров. Для спектрального подхода подобные цифры ранее не были известны.Поэтому последующие работы были связаны со сравнительными исследованиями различных спектральных подходов (системы №1-5) и оптимизацией спектрально-полюсного описания. Преследовалась цель получения зависимости надежности от числа полос и выбора минимального числа полос для приемлемой надежности распознавания.

Для получения результатов распознавания речевой сигнал вводился в ЭВМ через 8-разрядный аналого-цифровой преобразователь с частотой квантования 10 кГц. Выделение признаков производилось каждый 16 мс при хэмминговом кадре анализа длитель - ностью 25.6 мс. Каждый кадр подвергался предварительной обра - ботке при помощи операции взятия первой разности (дифференци - рование) и процедуры быстрого преобразования Фурье. На основе полученного спектрального представления формировался вектор (система №1) $X = (x_1, x_2, \ldots, x_i, \ldots, x_p)$, где $x_i = \sqrt{\frac{E_i}{E_0}}$, E_0 - энергия сигнала в кадре анализа, E_i - энергия сигнала в i-й полосе, p - число полос. Речевой сигнал представляется последовательностью векторов.

Обучение системы заключалось в однократном произнесении словаря и побайтной упаковке признаков в память ЭВМ. Распоз - навание контрольных реализаций состояло в выборе одной из K

Таблица 2

p	%
2	80.5
3	90.5
4	96.
5	97.5
6	97.
7	97.5
8	97.
9	98.
10	99.
11	98.5
12	98.
13	98.5
14	98.5
15	99.
16	99.
17	98.5
18	98.5
19	98.
20	98.5

Таблица 3

p	%
2	81.5
3	93.5
4	97.
5	98.
6	98.5
7	98.5
8	99.
9	99.
10	99.5
11	99.5
12	99.5
13	99.5
14	99.5
15	99.5
16	99.5
17	99.5
18	99.5
19	99.5
20	99.5

гипотез (K - объем словаря). Для сравнения слов использовался модифицированный алгоритм динамического программирования, описанный в [46].

Результаты распознавания словаря из 200 слов для одного диктора приведены в табл.2. В левом столбце таблицы приведено число частотных полос, в правом указана надежность распознавания в [%].

В табл.3 приведены результаты распознавания того же речевого материала по авторегрессионному спектру при "оптимальном" порядке модели авторегрессии, равном 16 (система №2).

Из табл.2 видно, что с ростом числа полос надежность распознавания в среднем увеличивается. Можно заметить, что, начиная с 9-ти полос и выше, значения надежности незначительно колеблются около некоторого уровня. Эти колебания объясняются гармоничностью несглаженного БПФ-спектра речевого сигнала на озвученных сегментах речи, из-за которой при изменении числа полос происходит скачкообразная перекачка энергий в соседние полосы. Авторегрессионное (см. табл.3) сглаживание позволяет повысить надежность и "устранить" колебания, так что надежность распознавания становится монотонной функцией числа спектральных полос.

5.3. Исследование бинарных спектрально-полосных признаков.

В некоторых системах распознавания используется бинарное кодирование спектрально-полосных признаков. Например, в системах типа "Речь" [53] и "Икар" [54] для кодирования применяется знак производной спектра, а в системах типа "Марс" [55] - сравнение энергии в полосе с фиксированным порогом. Подобное кодирование позволяет сократить объем памяти, резервируемый под эталоны, и для сравнения элементарных участков речи использовать простое в вычислительном плане хэммингово расстояние.

В процессе работы по теме были разработаны еще 2 более совершенных способа бинарного представления спектрально-полосных признаков, основанных на дихотомическом разбиении частотного диапазона [50-52]. Представляют интерес результаты сравнительных испытаний этих систем на одном и том же речевом материале.

Первый из этих способов (система №3) кодирования состоит в следующем [50]. Сначала спектральный диапазон разбивается на 2 полосы равной ширины и производится сравнение энергий в этих полосах. При этом если разность энергий в левой и правой полосах выше некоторого порога и значение энергии в правой полосе также выше заданного порога, то первый разряд кода принимает значение, равное 1; в противном случае этот разряд принимает значение, равное 0. Далее каждая из полос разбивается на две полосы равной ширины и производится сравнение энергий в этих полосах способом, описанным выше. В результате сравнения получим еще 2 бита. Продолжая подобное разбиение до заданного числа полос, получим бинарное спектрально-полосное представление сигнала.

Второй способ [51] (система №4) кодирования, так же как и первый, базируется на последовательном разбиении частотного диапазона на полосы равной ширины. Первый шаг кодирования такой же, как и в предыдущем способе. На втором шаге (после получения четырех полос) и последующих сравнение энергий произ -

водится во всех соседних полосах. Поэтому на втором шаге получается 3 бита вместо двух.

Иными словами, в обоих случаях спектральный диапазон "просматривается" через "окно", уменьшающееся от шага к шагу в 2 раза по ширине, но содержащее на любом шаге ровно 2 полосы. При этом в первом случае "просмотр" спектра идет без перекрытия "окна", а во втором спектральные "окна" перекрываются ровно на одну полосу.

Способы кодирования, принятые в системах "Икар", "Речь", "Марс" (система №5), имеют недостаток, состоящий в том, что при кодировании различные спектральные функции могут иметь одинаковые коды, т.е. кодирование спектра неоднозначно. Описанные способы кодирования, вообще говоря, тоже неоднозначны, однако степень неоднозначности у второго предложенного способа ниже, чем у всех остальных (см. следующий пункт).

5.4. Сравнительные экспериментальные результаты. В экспериментах по распознаванию использовался словарь из 100 слов. Три реализации словаря одним диктором были введены в ЭВМ через 8-разрядный АЦП при частоте квантования сигнала 10 кГц. В качестве эталонов поочередно бралась одна из реализаций словаря, а остальные предъявлялись на контроль (т.е. по 6 контрольных реализаций на слово).

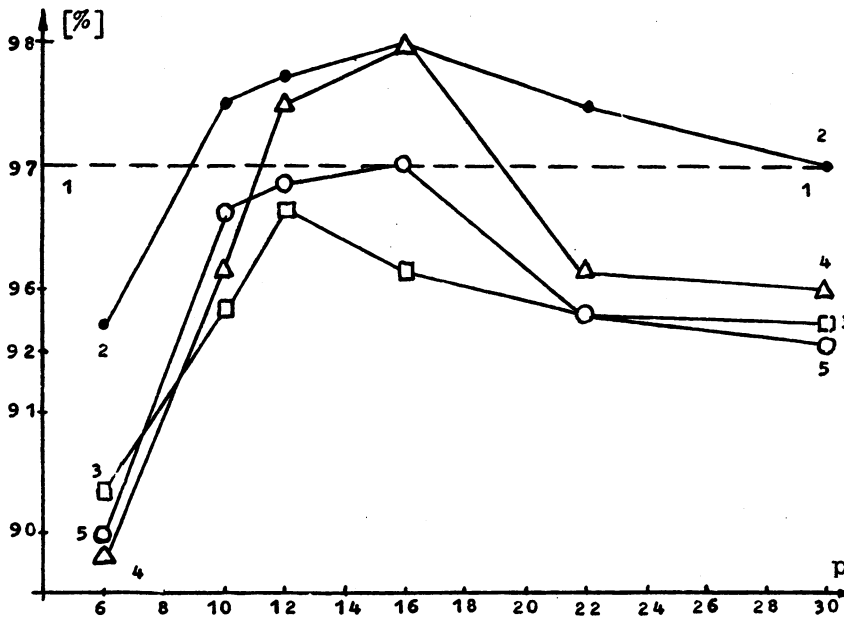
Для системы №1 был выбран диапазон частот от 230 Гц до 5 кГц, так что для 10 полос в соответствии с результатами п.4.4 гребенка спектроанализатора выглядела следующим образом: 230-398-580-788-1040-1354-1754-2273-2949-3835-5000 Гц. Число полос было выбрано на основе результатов предыдущих экспериментов (см. табл.1-2). Длина интервала анализа составила 25.6 мс. Перекрытия между интервалами не было. Результат распознавания - 97%.

Для системы №2 параметры гребенки фильтров были такие же, как и для системы №1.

В системах описания №3-5 частотный диапазон разбивался на полосы равной в шкале Гц ширины. Максимальное число полос равно 32. Для сглаживания применялась модель авторегрессии. В табл.4 приведены результаты распознавания при различном порядке модели (различной степени сглаживания).

Таблица 4

Порядок модели АР	2	10	12	16	22	30
Система №2	93.33	97.33	97.50	97.67	97.33	97.00
Система №3	90.33	95.50	96.67	96.17	94.67	93.17
Система №4	89.83	96.17	97.33	97.67	96.17	96.00
Система №5	88.50	96.67	96.83	97.00	94.67	92.50



Зависимость надежности распознавания от порядка модели авторегрессии для различных систем описания.

Таблица 5

Системы				%
1	A			97.7
	B			97.8
	C			97.8
2	A			97.5
	B			97.5
	C			97.5
3	A			97.7
	B			97.8
	C			97.8
4	A			97.1
	B			97.1
	C			97.1
5	A			97.1
	B			97.1
	C			97.1
6	A			97.1
	B			97.1
	C			97.1
7	A	AA		97.9
		BB		97.9
	B	AA	AAA	98.1
			BBB	98.1
			CCC	98.1
		BB	AAA	97.0
			BBB	97.0
			CCC	97.0
8	A	AA		96.0
		BB		97.0
	B	AA		94.0
		BB		97.0

Для наглядности результаты распознавания приведены на рисунке (с.125).

На рисунке система №1 - это энергии дифференцированного сигнала в полосах БПФ-спектра; система №2 - энергии дифференцированного сигнала в полосах авторегрессионного спектра; система №3 - дихотомически перекодированные в бинарные авторегрессионные спектрально-полосные признаки; система №4 - бинарно перекодированные дихотомические авторегрессионные спектрально-полосные признаки; система №5 - двоичные признаки, основанные на знаке производной авторегрессионного спектра.

Наилучшие результаты по надежности дает система №2. Из систем, использующих бинарные признаки, предпочтительнее система №4.

В табл.5 сведены значения надежности распознавания для исследованных систем описания (нумерация систем такая же, как и в п.5.1). Значения надежности в этой таблице усреднены по результатам распознавания пяти словарей объемом 45, 100, 125, 200 и 300 слов. Среди этих словарей были проблемно-ориентированные, фонетически сбалансированный и частотный словари.

Проведенные исследования позволили аппаратно реализовать близкий к оптимальному блок первичной обработки сигналов, который уже около 10 лет используется в различных вариантах мало -

габаритной системы распознавания речи "Сибирь". Один из таких вариантов имеет 6 полос с нижней граничной частотой 300 Гц, верхней - 5000 Гц и следующими спектральными полосами: 300-500-800-1250-1900-3100-5000 Гц (расчетные оптимальные значения: 300-585-932-1414-2136-3251-5000 Гц). Шесть оценок спектральной плотности позволяют работать со словарями до 200 слов с надежностью 98%. Другой вариант системы включает двоичные спектральные дихотомические признаки, построенные по этим шести полосам. При этом на один сегмент речи традиционной длительности используется всего 1 байт. Эта система при весьма сильной компрессии работает с надежностью около 95% на словарях из 100 слов.

5.5. Обсуждение результатов. В целом результаты экспери - ментов сводятся к следующему.

1. При построении универсальных систем распознавания для обработки речи как во временной, так и в спектральной областях необходимо оценивать по 10-16 величин при частоте квантования сигнала 10 кГц. Во временной области этими величинами могут быть: частные корреляции, коэффициенты авторегрессии, кепстральные коэффициенты, логарифмы отношения площадей однородной акустической трубы и др. параметры, однозначно связанные с перечисленными. В спектральной области в качестве признаков следует использовать сглаженные оценки спектральной плотности.При выборе сетки частот для оценивания желательно придерживаться логарифмического масштаба или масштаба Барк.

Все перечисленные способы первичного описания позволяют строить универсальные системы распознавания речи, работающие с надежностью 97-99% при отношении сигнал/шум на входе блока выделения признаков не менее 24 дБ.

2. Метрика в пространстве выбранного первичного описания должна быть согласована с правилом принятия решения. В против-

ном случае (как это имеет место в некоторых работах) высокая надежность не будет гарантирована.

6. Анализ эффективности спектрального и временного подходов

Теоретически подходы к первичной обработке сигналов во временной и спектральной областях эквивалентны в том плане, что между временными и спектральными характеристиками имеется взаимно-однозначное Фурье-соответствие. При этом, приняв статистический подход к распознаванию и, например, байесовскую стратегию принятия решения, метрике во временной области можно всегда однозначно сопоставить метрику в спектральной (поскольку максимизация апостериорной плотности во "временном" и "спектральном" пространстве признаков эквивалентны ввиду их взаимно-однозначной связи). Поэтому при фиксированном правиле принятия решения и способе нелинейной нормализации по темпу при помощи алгоритма динамического программирования временной и спектральный подходы к обработке речи при построении системы распознавания равносильны.

Однако при практической реализации схем или подходов к обработке они оказываются не совсем равнозначными, что выражается в отличиях по надежности распознавания речи, лежащих в пределах 1-3%. Примерно в этих же пределах наблюдаются колебания в надежности распознавания при использовании различных алгоритмов первичной обработки для оценивания признаков в рамках фиксированного подхода (временного или спектрального). В чем же причина этих колебаний?

Независимо от принятого подхода к обработке основная причина состоит в том, что алгоритмы оценивания признаков отличаются по точности. В свою очередь, различия по точности объясняются отличиями в свойствах оценок (смещение, дисперсия и т.п.), которые используются в этих алгоритмах. Здесь следует

отметить, что в речевых исследованиях на свойствах оценок уже традиционно почти не обращается внимания, что, конечно же, не совсем правомерно. Между тем, зная, например, дисперсии оценок для одних и тех же величин (признаков), полученных при помощи различных алгоритмов предобработки, можно заранее установить, какой из способов предобработки лучше. При этом отпадает необходимость в "дорогостоящих" экспериментах по распознаванию. На практике же, как правило, основным показателем считается высокая (97-99%) надежность распознавания, за которой остаются фактически скрытыми методы предобработки (их трудоемкость, устойчивость, точность). Последнее обстоятельство весьма затрудняет сравнительный анализ разработанных систем распознавания. К тому же результаты по надежности систем получают в различных условиях и на отличном речевом материале.

Отсутствие сравнительных данных (по свойствам оценок) и ориентация только на надежность распознавания как основную характеристику системы были присущи для исследований, проводимых примерно до конца 80-х годов. Подобная ориентация породила серию работ, которую в целом можно охарактеризовать как гонку за процентами и долями процента в повышении надежности распознавания. Это соревнование сыграло свою положительную роль в том смысле, что к настоящему времени все существующие универсальные системы распознавания работают примерно с одинаковой надежностью 97-99% (в отсутствие помех) при программной реализации алгоритмов с использованием "плавающей" арифметики. Вместе с тем эти системы отличаются по своей сложности в первую очередь из-за алгоритмов предобработки или из-за свойств оценок признаков. Невысокая устойчивость, большая дисперсия оценок или их низкая точность компенсируются в этих системах распознавания усложнением самих алгоритмов предобработки путем введения специальных эвристических приемов.

Другая причина колебаний в надежности распознавания заключается в переводе признаков в целочисленный диапазон, т.е. в эффектах квантования, которые неравнозначно сказываются на оценках признаков. Так, например, экспериментальное исследование эффектов квантования показало, что при спектрально-полосном подходе к распознаванию, при равномерной в масштабе Барк сетке частот для логарифмов оценок спектральной плотности под признаки необходимо отводить не менее 4-5 разрядов. При этом надежность распознавания остается на том же уровне, как и в случае большего числа разрядов (от 6 до 16). Однако стоит лишь изменить сетку частот или масштаб измерения спектральных плотностей, как сразу же изменятся дисперсии оценок, что потребует перераспределения в числе бит, отводимых под признаки, и сделает квантование неравномерным. При этом евклидова метрика будет рассогласована с признаками, что, в свою очередь, приведет к снижению надежности распознавания.

Аналогичным образом обстоит дело с временными признаками - параметрами авторегрессии (коэффициентами линейного предсказания) и частными корреляциями. Так как дисперсия оценок частных корреляций растет с порядковым номером коэффициента, квантование этих коэффициентов должно быть неравномерным (примерно от 9-10 разрядов для первого коэффициента до 3-4 для последнего при $p = 10$). При этом оценки частных корреляций согласованы с евклидовой метрикой. Переход же к равномерному квантованию - 9 бит под каждый коэффициент - снижает надежность с 97-99% до 88-90%. Точно так же эти эффекты проявляются при распознавании по коэффициентам авторегрессии, следует только учесть, что при байесовском подходе с параметрами авторегрессии асимптотически согласуется метрика Итакуры.

Следующий фактор, влияющий на разброс результатов по надежности, - устойчивость методов обработки, т.е. устойчивость к ошибкам измерения (наличию помех в измеряемом сигнале) и вычис-

лительная устойчивость (скорость нарастания ошибки вычислений). Традиционные временные методы в отличие от спектральных весьма чувствительны к аддитивным помехам (фоновому шуму) и обладают меньшей вычислительной устойчивостью. Это связано с тем, что для них требуется выполнение процедур обращения матриц, которые в условиях помех и ограниченности разрядной сетки вычислителя становятся неустойчивыми вследствие плохой обусловленности задачи.С увеличением мощности помехи точность оценок заметно снижается.Для получения удовлетворительных результатов требуется применение калмановской или обобщенной марковской фильтрации. При этом существенно возрастает трудоемкость алгоритмов обработки: они приобретают полиномиальную сложность. На практике это приводит к усложнению спецпроцессоров обработки речевых сигналов: вместо скорости обработки 1,5-2 млн.оп/с требуется производительность не менее 10 млн.оп./с. Поскольку имеются сотни вариантов обработки речи, вряд ли имеет смысл приводить здесь точные оценки сложности каждого из алгоритмов.

В целом же, как показывает практика, трудоемкость обработки речи во временной области примерно в 2 раза выше,чем трудоемкость обработки в спектральной. Сложность временного подхода оказывается большей и при вычислении расстояния между элементарными участками сигнала. Для сравнения сегментов, описываемых коэффициентами уравнения авторегрессии,требуется вычисление скалярного произведения,что по трудоемкости эквивалентно вычислению евклидова расстояния для коэффициентов частной корреляции или для оценок спектральной плотности. Однако, как показывают эксперименты, при сравнении элементарных участков сигнала в спектральной области евклидово расстояние можно заменить чебышевской метрикой и получить выигрыш в скорости вычислений в 2-5 раз. Более того, при переходе к бинарным спектральным признакам можно обойтись вычислением хеммингова расстояния, что наряду с уменьшением требуемой оперативной памяти

сокращает трудоемкость вычисления расстояния в десятки раз без заметного снижения надежности распознавания. Таким образом, спектральные признаки на практике оказываются эффективнее временных несмотря на их теоретическую равноценность.

З а к л ю ч е н и е

В данной статье рассмотрены два подхода к выделению признаков речевого сигнала - спектральный и временной. В рамках этих подходов описаны известные и разработаны новые способы выделения признаков. Проведенные исследования и анализ известных работ позволяют утверждать, что спектральный и временной подходы к выделению признаков эквивалентны в смысле результатов по надежности распознавания. Поэтому любой из рассмотренных способов предобработки может быть использован в универсальных системах распознавания речи. В то же время спектральных подход требует меньших вычислительных затрат и более устойчив к помехам при выделении признаков. При переходе от спектральный признаков к двоичным возможно существенное сжатие данных без снижения надежности распознавания.

Высокие по надежности распознавания результаты, получае - мые в последнее время,ни в коей мере не означают полного решения проблемы первичного описания речевого сигнала. Вот уже на протяжении около 40 лет остаются нерешенными близкие по своей сути главные задачи - задачи построения первичного описания инвариантного к диктору, а также инвариантного к нелинейным искажениям сигнала. Важность решения этих задач ни у кого не вызывает сомнений. Отсутствие инвариантных признаков сдерживает широкое практическое применение систем распознавания речи. В качестве примера можно привести такой факт. Универсальная система распознавания речи, работающая с надежностью 98% на словаре из 200 слов с подстройкой под диктора, снижает свою надеж - ность до 60% при наличии нелинейных искажений (типа клиппиро -

вания) в канале связи. Понижение надежности такого же порядка наблюдается у всех известных систем распознавания речи при работе с произвольным диктором (когда система обучена на конкретного диктора).

Если проблему инвариантности к диктору можно частично обойти путем настройки системы распознавания на диктора, то устойчивость или инвариантность к нелинейным искажениям обеспечить подобным образом путем компенсации этих искажений не всегда возможно. Основная причина состоит в том, что вид этих искажений в большинстве случаев неизвестен и зачастую случаен.

Снижение надежности распознавания из-за возникающих после обучения нелинейных искажений является следствием принятого во всех работах подхода к выделению признаков, который не отражает в полной мере свойств процессов речеобразования и восприятия речи человеком. Суть этого подхода состоит в том, что и аппроксимация сигнала моделью (при выделении признаков) и вычисление расстояния в пространстве признаков в конечном счете осуществляются по среднеквадратичному критерию близости. При этом в случае выделения признаков рассматривается близость исходного сигнала и сигнала, генерируемого аппроксимирующей моделью, а в случае сравнения эталонной реализации с контрольной вычисляется близость в среднеквадратичном между сигналами, генерируемыми двумя моделями - эталонной и контрольной. Наличие нелинейных искажений в контрольной реализации сигнала приводит к увеличению значения критерия и как следствие к увеличению расстояния между эталонным и контрольным представителями одного и того же образа в пространстве признаков, что в конце концов приводит к возрастанию ошибок классификации.

Между тем большинство нелинейных искажений не только сохраняют разборчивость сигнала, но и узнаваемость диктора. Этим собственно речевой сигнал и примечателен и существенно отли -

чен от сигналов другой природы. Однако именно эти свойства речи и не учитываются в существующих распознающих системах.

Таким образом, решение проблемы инвариантности первичного описания к нелинейным искажениям амплитуды сигнала лежит на пути полного учета указанных особенностей речи. Представляется, что один из подходов к решению проблемы может состоять в переходе к более грубым шкалам признаков: от традиционной относительной шкалы к шкале порядка или бинарной шкале. Другой способ, основанный на методе обратных оценок, по-видимому, найден в работах [56-58]. К настоящему времени получены обнадеживающие результаты. Однако требуется их всесторонняя проверка.

Литература

1. РАБИНЕР Л., ШАФЕР Р. Цифровая обработка речевых сигналов. - М.: Радио и связь, 1981. - 495 с.

2. ФАНТ Г. Акустическая теория речеобразования.-М.:Наука, 1964. - 283 с.

3. ФЛАНАГАН Дж. Анализ, синтез и восприятие речи: Пер.с англ. /Под ред. А.А.Пирогова.-М.: СВязь, 1968. - 396 с.

4. МАРКЕЛ Дж.Д, ГРЭЙ А.Х. Линейное предсказание речи:Пер. с англ./Под ред. Ю.Н.Прохорова, В.С.Звездина. -М.: Связь,1980. - 308 с.

5. МАКХОЛ Дж. Линейное предсказание речи. Обзор //ТИИЭР.- 1975. - Т. 63, № 4. - С. 20-44.

6. ШАФЕР Р., РАБИНЕР Л. Цифровое представление речевых сигналов //ТИИЭР. - 1975. -Т.63, №4. -С. 141-159.

7. САПОЖКОВ М.А. Речевой сигнал в кибернетике и связи. - М.: Связьиздат, 1963. - 450 с.

8. ВЕЛИЧКО В.М., ЗАГОРУЙКО Н.Г. Автоматическое распознавание ограниченного набора устных команд //Вычислительные системы. - Новосибирск, 1969. - Вып. 36. -С.101-110.

9. АКИНФИЕВ Н.Н. Об одном психоакустическом законе восприятия человеком речевого сигнала и об объективных параметрах сигнала, содержащих речевую информацию //Тез.докл. 15-го Все - союз.семинара по автоматическому распознаванию образов (АРСО-15), Таллин; март, 1989. - Таллин, 1989. -С. 179-181.

10. КЕЛЬМАНОВ А.В. О выборе числа и границ спектральных полос при распознавании речевых сигналов //Тез. докл. и сообщ 12-го Всесоюз. семинара по автоматическому распознаванию слу · ховых образов (АРСО-12), Киев-Одесса, сентябрь,1982. - Киев, 1989. -С. 313-315.

11. Его же. К вопросу выбора числа и границ спектральных полос при распознавании речевых сигналов //Советско-французск. симп. "Акустический диалог чел[illegible] с машиной", Москва, сент., 1984.: Тез докл. - М., 1984. С.70-73.

12. KELMANOV A.V., LEBE[illegible]/ V.G., VELICHKO V.M., ZAGORUI - KO N.G. A stady on speaker independent speech characteristics //Symposium Franco-Sovietique sur la parole, Grenoble, 20-22 oct. 1981. - Grenoble, 1981. -P. 130-141.

13. Idem. A stady on speaker independent speech characteristics //Proc. 6-th Intern. Conf. on Pattern Recognition, Munich, Germany, 19-22 oct. 1982. - Munich, 1982.

14. ЛЕБЕДЕВ В.Г. Автоматическое распознавание речи по маскпризнакам //Эмпирическое предсказание и распознавание образов. - Новосибирск, 1976. -Вып.67: Вычислительные системы. - С. 136-140.

15. ЗАГОРУЙКО Н.Г. Методы распознавания и их применение.- М.: Сов. радио, 1972. - 206 с.

16. МАРТИН Т.Б. Практические применения речевого ввода в вычислительную машину //ТИИЭР. - 1976. -Т.64, №4. -С.80-85.

17. Современные устройства распознавания речи //Радиоэлектроника за рубежом. - 1983. -Вып. 23. -С. 1-31.

18. Речевая связь с машинами. Тематический выпуск //ТИИЭР. - 1985. -Т. 73, №11.

19. НИКИФОРОВ И.В. Последовательное обнаружение изменения свойств временных рядов. -М.: Наука, 1983. - 199 с.

20. КЛИГЕНЕ Н., ТЕЛЬКСНИС Л. Методы обнаружения моментов изменения свойств случайных процессов //Автоматика и телемеханика. - 1983. - № 10. -С. 5-56.

21. ТОРГОВИЦКИЙ И.Ш. Методы определения момента изменения вероятностных характеристик случайных величин //Зарубежная радиоэлектроника. - 1976. - №1. -С. 3-52.

22. ПРОХОРОВ Ю.Н. Статистические модели и рекуррентное предсказание речевых сигналов. -М.: Радио и связь,1984.-238 с.

23. ITAKURA F. Minimum prediction residual principle applied to speech recognition //IEEE Trans. Acoust. Speech Signal Process. -1975.- Vol. ASSP-23, N 1. -P. 67-72.

24. АНДЕРСОН Т. Статистический анализ временных рядов : Пер. с англ. /Под ред. Ю.К.Беляева. -М.: Мир, 1976. - 755 с.

25. БОКС Дж., ДЖЕНКИНС Г. Анализ временных рядов. Прогноз и управление. Вып. 1,2: Пер. с англ. /Под ред. В.Ф.Писаренко.- М.: Мир, 1974. - 406 с., 197 с.

26. ДЖЕНКИНС Г., ВАТТС Д. Спектральный анализ и его приложения. Вып. 1,2: Пер. с англ. /Под ред. В.Ф.Писаренко , -М.: Мир, 1971. - 316 с., 287 с.

27. РАБИНЕР Л., ГОУЛД Б. Теория и применение цифровой обработки сигналов.: Пер. с англ. /Под ред. Ю.Н.Александрова. - М.: Мир, 1978. - 848 с.

28. ЛОЗОВСКИЙ В.С. Анализ и синтез речи на основе Z-описания //Вычислительные системы. - Новосибирск, 1971. -Вып. 44. - С. 136-140.

29. ЛЮДОВИК Е.К. Совместное оценивание нуль-полюсных параметров речевого тракта и характеристик источников возбуждения //Обработка и распознавание сигналов. - Киев, 1975.-С. 121-134.

30. КЕЛЬМАНОВ А.В. Алгоритм классификации тон/шум, основанный на критерии адекватности модели авторегрессии //Методы обработки информации. - Новосибирск, 1978. -Вып. 74: Вычислительные системы. -С. 129-148.

31. Его же. Алгоритм классификации тон/шум по частным автокорреляциям //Эмпирическое предсказание и распознавание образов. -Новосибирск, 1980. -Вып. 83: Вычислительные системы. - С. 67-73.

32. Его же. О некоторых алгоритмах классификации тон/шум и выделения траектории основного тона //Тез. докл. XI-й Всесоюз. школы-семинара по автоматическому распознаванию слуховых образов (АРСО-XI), Ереван, дек. 1980. - Ереван, 1980. - С. 88-90.

33. Его же. Алгоритм выделения основного тона по разностной функции ряда остаточных ошибок модели авторегрессии //Методы обнаружения закономерностей с помощью ЭВМ. - Новосибирск, 1981. - Вып. 91: Вычислительные системы. -С. 113-124.

34. ВИНЦЮК Т.К. Анализ, распознавание и интерпретация речевых сигналов. - Киев: Наукова думка, 1987. - 262 с.

35. ГРЕНАНДЕР У., СЕГЕ Г. Теплицевы формы и их приложения. - М.: ИЛ, 1961. - 308 с.

36. ICHIKAWA A., NAKANO V., NAKATA K. Evaluation of various parameter sets in spoken digits recognition //IEEE Trans on Audio and Electroacoustics. - 1973. -Vol. AU-21,N3.-P.202-209.

37. ОППЕНГЕЙМ А.В., ШАФЕР Р.В. Цифровая обработка сигналов: Пер. с англ. /Под ред. С.Я.Шаца. -М.: Связь, 1979.-416 с.

38. WAKITA H. Linear predictiom voice synthesizers: line spectrum pairs (LSP) is the newest of several techniques // Speech Tecnology. - 1981. -Vol. 1. -P. 17-22.

39. PALIWAL K.K. A study of the spectrum pair frequencies for vowel recognition //Speech Communication. - March 1989. - Vol. 8, N 1. -P. 27-33.

40. ВЕЛИЧКО В.М., КЕЛЬМАНОВ А.В.. Распознавание изолированных слов по авторегрессионному спектру с применением информационного критерия Акаика //Тез.докл. XI-й Всесоюз . школы-семинара по автоматическому распознаванию слуховых образов (АРСО-XI), Ереван, дек. 1980. - Ереван, 1980. - С. 263-264.

41. ГАЛУНОВ В.И., ОРЛОВА М.И., ЯГУНОВА Н.Н. Использование метода аппроксимации спектрального среза речевого сигнала в задачах распознавания речевых образов //Тез.докл. 13-й Всесоюз. школы-семинара по автоматическому распознаванию слуховых образов (АРСО-13), Новосибирск, июль, 1984.-Ч.2.-Новосибирск,1984. - С. 84-85.

42. РУДЖЕНИС А.И., ЧЮКШИС Э.А. Анализ признаков на основе кусочно-полиномиального приближения спектральной функции // Тез. докл. 13-й Всесоюз. школы-семинара по автоматическому распознаванию слуховых образов (АРСО-13), Новосибирск, июль,1984. - Ч.1. - Новосибирск, 1984. - С.66.

43. Вокодерная телефония. Методы и проблемы /Под ред. А.А.Пирогова. -М.: Связь, 1974. - 536 с.

44. Физиология речи. Восприятие речи человеком /Л.А.Чистович и др. - Л.: Наука, 1976. - 388 с.

45. ЦВИКЕР Э., ФЕЛЬДКЕЛЛЕР Р. Ухо как приемник информации.: Пер. с нем. /Под ред. Б.Г.Белкина. -М.: Связь, 1971. - 256 с.

46. КЕЛЬМАНОВ А.В. Сравнение систем признаков, основанных на частной автокорреляционной функции при решении задачи распознавания изолированных слов //Эмпирическое предсказание и распознавание образов. - Новосибирск, 1980: - Вып. 83: Вычислительные системы. -С. 74-97.

47. Его же. Экспериментальное исследование 18 систем первичного описания речевого сигнала //Тез.докл. XI-й Всесоюз. школы семинара по автоматическому распознаванию слуховых образов (АРСО-XI), Ереван, дек. 1980. - Ереван, 1980. -С.112-113.

48. Его же.Система распознавания изолированных слов по частной автокорреляционной функции //Эмпирическое предсказание и распознавание образов. -Новосибирск, 1978. - Вып.76: Вычислительные системы. - С. 132-143.

49. Его же.Сравнительное исследование двух алгоритмов динамического программирования //Тез. докл. и сообщ. 12-го Всесоюз. семинара по автоматическому распознаванию слуховых образов (АРСО-12), Киев-Одесса, сентябрь, 1982. - Киев, 1989. - С. 474-476.

50. ЗАГОРУЙКО Н.Г., КЕЛЬМАНОВ А.В. Распознавание речи по дихотомически перекодированным в бинарные спектрально-полосным признакам //Тез.докл. 13-й Всесоюз. школы-семинара по автоматическому распознаванию слуховых образов (АРСО-13), Новосибирск, июль 1984. - Ч.2. - Новосибирск, 1984. -С. 86-89.

51. КЕЛЬМАНОВ А.В. Распознавание речи по бинарно перекодированным дихотомическим спектрально-полосным признакам //Там же. - С. 89-90.

52. Его же. Экспериментальное исследование некоторых систем первичного описания речевого сигнала при распознавании изолированных слов //Там же. - С. 91-93.

53.ВИНЦЮК Т.К., ЛОБАНОВ Б.М., ШИНКАЖ А.Г. Система распознавания речи и система устного диалога СРД "Речь-1" на основе микро-ЭВМ// Тез. докл. и сообщ. 12-го Всесоюз. семинара по автоматическому распознаванию слуховых образов (АРСО-12), Киев-Одесса, сентябрь 1982. - Киев, 1989. -С. 516-521.

54. АВРИН С.Б. О характеристиках надежности распознавания устных команд устройством ИКАР //Тез.докл. 13-й Всесоюз.школы-семинара по автоматическому распознаванию слуховых образов (АРСО-13), Новосибирск, июль, 1984. -Ч.2. - Новосибирск, 1984. - С. 179-180.

55. АФАНАСЬЕВ В.П. и др. Архитектура речевого видеотерминала МАРС-1 //Там же. - Ч.1. -С. 31-33.

56. КЕЛЬМАНОВ А.В. Корректоры нелинейно искаженной речи, основанные на методе обратных оценок //Тез.докл. 15-го Всесоюз. семинара по автоматическому распознаванию слуховых образов (АРСО-15), Таллин, март 1989. -Таллин, 1989. -С. 158-159.

57. Его же.Алгоритмы анализа речевых сигналов по искаженным наблюдениям //Там же. -С. 206-207.

58. Его же. Методы обратных оценок в задачах первичной обработки речевых сигналов //Автоматическое распознавание и синтез речевых сигналов /Сб. Тр. Института кибернетики АН УССР. - Киев, 1989. - С. 20-22.

Поступила в ред.-изд.отд.

12 июня 1990 года