

АВТОМАТИЧЕСКОЕ ОПОЗНАВАНИЕ ЗВУКОВЫХ ОБРАЗОВ
(ОБЗОР ЛИТЕРАТУРЫ)

Н.Г. Загоруйко, Г.Я. Волошин, В.Н. Елькина

1. ВВЕДЕНИЕ

Количество публикуемых работ по автоматическому анализу и синтезу звуковых образов ежегодно возрастает.

Обзоры литературы [1, 2, 3] облегчают ознакомление с состоянием исследований в этом направлении. Однако даже наиболее поздние обзоры [1] не охватывают работ, опубликованных в 1963 - 1964 гг. Кроме того, в связи с повышением интереса к использованию ЭВМ было бы желательно систематизировать накопленный опыт применения ЭВМ для анализа и синтеза звуков.

Наконец, в имеющихся работах мало внимания уделяется теоретическим методам, лежащим в основе тех или иных исследований.

В данной статье содержится попытка систематизировать используемые методы и опыт применения ЭВМ в этих исследованиях и рассматривается содержание работ по анализу и синтезу звуковых образов, опубликованных за последние 1,5 - 2 года.

§ 1. Определения и классификация

В геометрической интерпретации под образом понимается локальное множество точек в n -мерном выборочном пространстве, где вдоль осей этого пространства отложены значения параметров (или признаков), характеризующих образ. Отдельная точка этого множества называется реализацией образа. Если n - конечное число, то локальное множество точек в n -мерном пространстве, соответствующее образу, называется эталоном данного образа. Набор эталонов образов представляет собой алфавит объектов опознавания. Параметры задаются распределениями их значений по осям n -мерного пространства.

В зависимости от характера распределений параметров, условий их измерения и способов использования информации о параметрах для опознавания образа, в литературе рассматривается несколько вариантов проблемы опознавания [4, 5].

1. Алфавит опознавания и значения их параметров определены однозначно. По этим значениям можно безошибочно относить предъявленную реализацию к тому или иному образу. Такое опознавание называется параметрическим. Устройства, реализующие эту процедуру опознавания, просты. Алгоритм их работы большого теоретического интереса не представляет.

2. Алфавит объектов опознавания устанавливается в процессе обучения, предшествующем процедуре опознавания. Распределения параметров накапливаются и уточняются как в режиме обучения, так и в режиме непосредственно опознавания. Этот тип процедуры носит название "адаптивное опознавание образов". Реализация данного типа процедуры осуществляется устройствами типа "Персептрон" [6, 7, 8].

3. Алфавит задан однозначно, но области выборочного пространства, соответствующие различным образам, перекрываются. Решение об отнесении реализации к тому или иному образу принимается на основании сравнения апостериорных вероятностей существования образов в данной точке пространства. Такой вид процедуры опознавания можно назвать вероятностным. Наибольшее число разрабатываемых в настоящее время опознающих устройств относится к этому виду. Подробнее рассмотрение вероятностной процедуры опознавания удобно вести на языке теории статистических решений [9, 10].

§ 2. СТАТИСТИЧЕСКАЯ ТРАКТОВКА ВЕРОЯТНОСТНОЙ ПРОЦЕДУРЫ ОПОЗНАВАНИЯ

Согласно [11], всякое опознающее устройство состоит из блока измерения параметров (рецептора), блока принятия решений (классификатора) и блока исполнительных устройств (эффектора).

Пусть S_i - алфавит K объектов опознавания, априорная вероятность появления которых равна ω_i , причем $1 \leq i \leq K$, $\sum_{i=1}^K \omega_i = 1$. Если рецептору предъявлена реализация i -го образа, то условная вероятность того, что значения параметров, измеренных рецептором, будут равны n , есть величина $\beta(n/S_i)$. Пространство величин $\beta(n/S)$ характеризует свойства рецептора.

По измеренному значению параметров классификатор определяет, к какому образу относится реализация, предъявленная для опознавания. Условная вероятность того, что по n значениям параметров будет опознаваться j -ый образ, равна $\delta(d_j/n)$. Множество величин $\delta(d_j/n)$ представляет собой пространство решений, или решающих правил.

Если стоимость потерь, обусловленных опознаванием j -го образа, когда действительно был предъявлен i -ый образ, задана матрицей потерь C_{ij} , то условные потери при S_i на входе и при использовании решающего правила $\delta(d_j/n)$ будут равны

$$R(\delta, S_i) = \sum_{j=1}^K C_{ij} \int_N \beta(n/S_i) \delta(d_j/n) dn, \quad (1)$$

или с учетом ω_i :

$$R(\delta) = \sum_{i=1}^K \sum_{j=1}^K C_{ij} \omega_i \beta(n/S_i) \delta(d_j/n) dn. \quad (2)$$

Оценка потерь будет полной, если учесть, что в результате помех в некоторых случаях будет приниматься "нулевое" решение - предъявленную реализацию опознают как образ, не содержащий в алфавите S_i . Стоимость потерь при отнесении i -го образа к нулевому есть C_{i0} , причем обычно

$$C_{ij} > C_{i0} > C_{ii} = 0$$

С учетом нулевых решений суммарные потери (ожидаемый риск) станут равными

$$R^* = \sum_{i=0}^K \omega_i C_{i0} + \sum_{i=1}^K \sum_{j=1}^K C_{ij} \omega_i \beta(n/S_i) \delta(d_j/n) dn. \quad (3)$$

Основное содержание исследований проблемы опознавания образов при заданном алфавите объектов опознавания S_i и заданной матрице потерь C_{ij} сводится к поиску оптимальных способов измерения образов (минимизация пространства N измерений) и оптимальных решающих правил δ , при которых достигается минимум ожидаемого риска R^* .

§ 3. Решающие правила

Минимизация ожидаемого риска потерь достигается в том случае, если в качестве решающего правила используется критерий Байеса:

$$\delta_i(n) = \frac{\omega_i \cdot \beta(n/S_i)}{\sum_{i=1}^K \omega_i \cdot \beta(n/S_i)}$$

и выбирается та гипотеза, которая соответствует максимальному значению $\delta_i(n)$. Процедура решения сводится к вычислению величин $\delta_i(n)$ по заданным значениям параметров n для всех $1 \leq i \leq K$ образов и к выбору максимальной из этих величин. Часто в решающее правило включается условие, по которому осуществляется выбор гипотезы в том случае, когда для нескольких гипотез получается одинаковая или близкая величина $\delta_i(n)$. Это условие состоит в том, что устанавливается некая пороговая величина разности $\Delta = \delta_i(n) - \delta_j(n)$. Решение считается принятым, если разность между максимальным значением $\delta_i(n)$ и ближайшим к нему по величине $\delta_j(n)$ превышает заданный порог Δ . В противном случае либо принимается нулевое решение (отказ), либо указываются конкурирующие гипотезы. Доказано, что класс методов Байеса является минимальным полным классом [9].

Для использования Байесова критерия выборочное пространство должно быть разбито на K подпространств. Граничная поверхность между i и j подпространством будет проходить через точки, в которых $\omega_i \cdot \beta(n/S_i) = \omega_j \cdot \beta(n/S_j)$. Такая поверхность называется оптимальной решающей границей. В зависимости от вида распределения величин $\omega_i \cdot \beta(n/S_i)$ решающая граница представляет собой гиперповерхность той или иной степени сложности.

Естественное стремление к упрощению схемы классификатора приводит к попыткам упростить форму граничных поверхностей. Де-

лается это, главным образом, путем замены оптимальных решающих гиперповерхностей гиперплоскостями [12].

Линейные решающие функции (л.р.ф.), используемые в этом случае вместо оптимальных решающих функций, легко реализуются простыми схемами из пассивных сопротивлений. Применение л.р.ф. сводится к решению уравнения гиперплоскости

$$\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \dots + \alpha_n x_n + \alpha_0 = a, \quad (4)$$

где $x_1, \dots, x_n$ - измеренные рецепторами значения n -параметров, а $\alpha_1, \dots, \alpha_n$ - постоянные коэффициенты однозначно задающие положения гиперплоскости в n -мерном пространстве. Знак величины a показывает, на какой стороне каждой гиперплоскости лежит опознаваемая точка. Положение гиперплоскости определяется видом распределения величин $\omega_i \cdot \beta(n/S_i)$. Если вид этих распределений неизвестен (что часто имеет место в практических задачах), а известны лишь их математические ожидания, то решающей гиперплоскостью является гиперплоскость, перпендикулярная линии, соединяющей математические ожидания, и делящая эту линию пополам. На рис. 1 показана оптимальная ре-

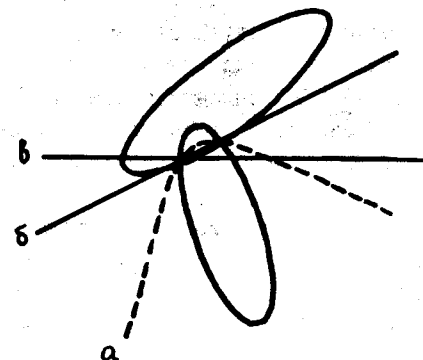


Рис. 1

шающая граница (линия a) и аппроксимирующая ее л.р.ф. (линия b). Линия b представляет собой л.р.ф., заданную с использованием информации лишь о математических ожиданиях значения параметров образа i и j . При определенном типе распределений величин $\omega_i \cdot \beta(n/S_i)$ [13] л.р.ф. является оптимальной решающей функцией, в частности, если поверхности равной вероятности представляют собой эллипсоиды и плотность вероятности монотонно уменьшается в направлении от математических ожиданий. Более того, имеются такие виды распре-

делений $\omega_i \propto (1/S_i)$, для которых оптимальной решающей функцией является плоскость, перпендикулярная линии, соединяющей математические ожидания. Вот некоторые из этих распределений:

1. Распределения независимы вдоль направлений гиперпространства, ортогональной линии, соединяющей математические ожидания, то есть линии $[M]$. Вдоль линии $[M]$ распределения монотонно убывают с удалением от математического ожидания.

2. Распределения монотонно убывают вдоль всех направлений с удалением от математических ожиданий, причем убывает и крутизна их изменения.

3. Распределения для всех образов попарно зеркально симметричны относительно какой-нибудь гиперплоскости.

В любом случае использование в качестве решающего правила л.р.ф. дает результат не хуже, чем использование ряда других решающих правил, в частности, таких, как минимум евклидова расстояния, максимум нормализованной взаимной корреляционной функции и т.д. [12].

В связи с этим попытки [14] улучшения метода гиперплоскостей за счет гибридизации его с корреляционным методом вряд ли будут успешными. При разбиении выборочного пространства положение каждой гиперплоскости следует задавать с учетом положения всех остальных. Такая "параллельная" процедура задания гиперплоскостей сложна. В [12] описана более простая "последовательная" процедура. Один из способов задания положения линейных решающих границ, основанный на "гипотезе компактности", описан в [15].

Интересным упрощением корреляционного метода оценки правдоподобия гипотез является метод Browning - Bledsoe (В-В схема) [16]. Значения каждого из n параметров ℓ -ой реализации записываются в виде столбца с единственным ненулевым элементом (см. табл. 1).

Таблица 1

	$\mathcal{L}$			M			G
	$\ell=1$	$\ell=2 \dots \ell=L$	S_1	$S_2 \dots S_K$			
$n=1$	0	0	0	0	0	0	I
	I	0	I	I	0	0	0
	0	I	0	I	0	0	0
	0	0	0	0	0	I	0
$n=2$	0	0	0	0	0	0	0
	0	0	0	0	I	0	I
	I	I	I	I	I	I	0
	0	0	0	0	0	I	0
$n=3$	0	I	I	I	0	I	0
	I	0	0	I	I	0	I
	0	0	0	0	0	0	0
	0	0	0	0	0	0	0

Эталон i -го образа (S_i) получен как результат логического построчного сложения матрицы, образованной из $\mathcal{L}$ векторов - реализаций. Таким путем получают эталоны всех остальных $K-1$ образов. В результате этого "процесса обучения" в памяти опознающего устройства создается матрица M эталонов.

Процесс опознавания состоит в следующем. Параметры реализации (G), предъявленной для опознавания, измеряются и представляются в виде таких же столбцов с одним ненулевым элементом. Затем вектор G логически умножается на матрицу M . В результате строки матрицы, которые совпадают с нулевыми элементами вектора G , вычеркиваются. То, что осталось от матрицы M представлено в табл. 2.

Таблица 2

	S_1	$S_2 \dots S_K$
1	0	I
2	0	I
n	I	I
	I	3
		0

Столбец, сумма единиц в котором максимальна, указывает образ, представителем которого является опознаваемая реализация.

Ясно, что чем меньше единиц содержит матрица M , тем меньше вероятность ошибок классификации. Стохастическая модель В-В схемы, позволяющая оценить ожидаемый риск потерь, описана в [17].

Вид решающей функции может быть существенно упрощен, если предварительно преобразовать координаты выборочного пространства, подвергнув их, например, инверсии, переносу, сжатию или растяжению, повороту и т.д. Если бы выборочное пространство было упругим, то преобразование координат выражало бы способ, с помощью которого различные области пространства должны быть растянуты или сжаты для того, чтобы приблизить друг к другу точки, принадлежащие одному образу, наиболее плотно. Интересны в этом плане работы [18, 19].

Устройства, реализующие преобразования координат, могут являться составной частью либо классификатора, либо рецептора. Простейшим представителем устройств этого типа можно считать, например, нормализатор уровня сигнала, подаваемого на вход рецептора.

§ 4. МИНИМИЗАЦИЯ ПРОСТРАНСТВА ИЗМЕРЕНИЙ

Минимизация описания образов, то есть выбора того наименьшего числа параметров, которые несут информацию, достаточную для опознавания образов с заданной степенью надежности, составляет предмет многих работ по автоматическому опознаванию. Если прежде выбор параметров основывался на интуитивных представлениях, то в последние годы все больше уделяется внимания выработке способов объективного выбора наиболее эффективных параметров [20]. Краткое описание этой проблемы на языке теории статистических решений содержится в [13]. Более плодотворными оказались попытки решить эту проблему в терминах теории информации [21]. Стратегия поиска минимального маршрута измерения опознаваемого объекта как для случая зависимых, так и независимых параметров, описана в работе [4]. Более детальное изложение методики оценки информационной эффективности независимых параметров речевых сигналов содержится в [22]. Под информационной эффективностью параметра понимается величина $R = \frac{J}{m}$,

где J - количество информации, содержащееся в параметре, а m - характеристики затрат (аппаратуры или машинных операций), необходимых для измерения этого параметра. J определяется разностью между исходной энтропией пространства сигнала H_0 и энтропией H_1 , оставшейся после того, как стало известно значение данного параметра. При независимых параметрах поиск минимального описания заключается в выборе тех параметров, информационная эффективность которых максимальна, а суммарное количество информации, несомой ими, достаточно для опознавания образов с заданной достоверностью.

Если параметры связаны между собой некоторой зависимостью, то выбор минимального маршрута, вообще говоря, может быть сделан путем полного перебора всех возможных сочетаний параметров и оценки количества информации, которую несут различные сочетания параметров.

В литературе имеются работы [4, 23], посвященные решению этой задачи методом направленного перебора сочетаний. В основе этих методов лежат следующие приемы. В качестве первого параметра используется тот, который несет максимум информации, если его считать независимым. Вторым выбирается тот параметр, который несет наибольшую информацию при условии, что значение первого параметра известно. Третьим - тот параметр, который несет наибольшее количество информации, при заданных значениях первых двух и т.д.

Другой прием состоит в том, что находят тот параметр, исключение которого из рассмотрения, увеличивает ошибки опознавания на минимальную величину. Затем находят наименее информативный из оставшихся параметров и исключают его и т.д.

По существу, 1 и 2 приемы - это один и тот же процесс. Разница заключается в направлении перебора. В отличие от полного перебора, эти методы направленного перебора не оптимальны. Полученные с их помощью маршруты обследования реализации могут оказаться не оптимальными, а минимальными лишь для локальной области решений.

Дальнейшая возможность минимизации выборочного пространства достигается уменьшением числа уровней (градаций), на которые разбиваются выбранные параметры. Но простым заглублением точности измерения цели достичь невозможно. Нужно решить вопрос о том, как при заданном числе градаций расположить границы между ними, чтобы извлечь из параметра максимальное количество информации. Метод оптимального размещения границ между за-

данным числом градаций для одномерного случая описан в [22]. Полное решение этой задачи следует искать в распространении этого способа на многомерный случай с использованием метода динамического программирования [24].

Как указывалось раньше, энтропию выборочного пространства можно уменьшить, если воспользоваться определенными преобразованиями его координат. Эти преобразования реализуются, например, с помощью устройств, осуществляющих "нормализацию" измеряемого сигнала. В [22] описан метод оценки информационной эффективности нормализации.

§ 5. Рациональный выбор алфавита объектов опознавания

Составляя алфавит объектов опознавания, иногда исследователь имеет свободу выбора. При опознавании ограниченного набора устных команд соответствующим подбором слов-синонимов можно существенно повысить надежность опознавания. Более интересен случай, когда требуется опознавать устные сообщения, составленные из сравнительно большого числа слов (несколько сот и больше). Тогда можно ставить вопрос о выборе между элементами различных иерархических уровней такими, как фонемы, слоги, слова, синтагмы и т.д. Вероятно, что наиболее рациональным алфавитом в этом случае будет определенное сочетание элементов всех иерархических уровней речи [20]. Для конкретного решения этого вопроса необходимо знать статистику всех этих элементов. Уже имеются работы, посвященные исследованию статистики некоторых элементов языка [25, 26, 27, 29].

Результаты исследований зрительного анализатора дают основание полагать, что и при опознавании изображений, алфавит опознаваемых объектов, вероятно, целесообразно составлять из элементов различного иерархического уровня (например, образ целого слова наряду с образами отдельных букв). Так, в [28] сделан вывод о том, что в зрительной системе кодирование осуществляется крупными блоками. Чем крупнее блок, по которому закодирован образ, тем меньше вероятность ошибки и тем меньше количество клеток (или групп клеток), необходимых для наиболее экономной записи образа.

Ш. ИСПОЛЬЗОВАНИЕ ЭВМ В РАБОТАХ ПО АВТОМАТИЧЕСКОМУ ОПЗНАВАНИЮ ОБРАЗОВ (ЗВУКОВ РЕЧИ)

В течение последних двух-трех лет в печати не появлялось сообщений о крупных успехах в решении проблемы автоматического распознавания речи в том плане, в каком она была поставлена исследователями 20-30 лет назад - выделить аппаратными (объективными) методами из непрерывного потока речи ограниченное число элементарных отрезков (фонем) и опознать их, решив тем самым проблему автоматического распознавания речи в целом (то есть в объеме неограниченного словаря).

К настоящему времени стало очевидным, что вряд ли удастся надежно выделить и опознать фонемы аппаратными методами без привлечения сведений о более крупных элементах речи и без применения сложных алгоритмов информационной и логической обработки параметров при помощи современных электронных вычислительных машин (ЭВМ). Применение ЭВМ в решении задач автоматического распознавания речи позволяет значительно удешевить и ускорить исследования. Сейчас ни одна сколько-нибудь значительная работа по анализу и распознаванию речевых сигналов не проводится без использования вычислительной техники.

В рассматриваемой области исследований ЭВМ сравнительно широко стали применяться совсем недавно, 3-4 года назад. Естественно, что прежде всего разрабатывались алгоритмы и программы, позволявшие моделировать принципы действия уже существовавшей аппаратуры.

Показательным в этом отношении является результат, полученный в электротехническом отделении Колумбийского университета (США) [36, 37]. Сотрудниками отделения разработаны программы выделения параметров речевого сигнала с помощью машины IBM 7090 в реальном масштабе времени. Анализируется диапазон частот до 8 кГц. Предварительно речевой сигнал проходит через волновой анализатор "Simotamic". Отсчеты спектра производятся через каждые 24 мсек. Показания волнового анализатора при помощи преобразователя аналог-цифра вводятся в ЭВМ. Шаг квантования по оси частот равен 30 Гц. Детальное исследование спектра сигнала позволяет определять частоту основного тона. Преобразователь работает с частотой 330 000 проб в сек. Такое высокое быстродействие потребовалось для того, чтобы из 24 мсек (интервал между соседними анализируемыми спектральными разрезами) как можно меньше времени потратить на ввод данных в ЭВМ.

При быстросействии 330 000 отсчетов в сек на ввод затрачивается меньше 1 мсек, а остальных 23 мсек оказывается достаточно для выделения комплекса параметров.

Разработанные программы позволяют выделять:

- 1) частоты 1-й, 2-й и 3-й формант (F_1 , F_2 , F_3) и их амплитуды;
- 2) распределение энергии по спектру;
- 3) наличие или отсутствие сигнала основного тона (по отношению энергии в диапазоне до 1 кГц к полной энергии);
- 4) частоту основного тона (среднее расстояние между гармониками по всему исследуемому диапазону речевого сигнала);
- 5) среднюю спектральную плотность шумовой энергии, наличие которой определяется по отношению энергии в диапазоне выше 5 кГц к полной энергии сигнала;
- 6) мгновенную мощность сигнала (суммирование квадратов амплитудных значений отсчетов за каждый временной интервал);
- 7) начало и конец слова (начало слова определяется наличием 5 и более отсчетов выше определенного порога, конец слова — по отсутствию энергии в течение 4-х отсчетов).

В этом же университете проведена работа [53] по моделированию на ЭВМ IBM 7090 фильтров с различными характеристиками и по исследованию реакций этих фильтров на различные сигналы.

Ряд программ по моделированию на ЭВМ аппаратных методов выделения параметров речевых сигналов разработан в Институте математики СО АН СССР: спектральный и корреляционный анализ, определение мгновенной мощности, подсчет частоты перехода через нуль клипированной речи (клипирование подвергается как исходный, так и предварительно продифференцированный сигнал) [20].

Применение ЭВМ позволяет не только моделировать существующую аппаратуру, но и применять новые методы исследований, трудно реализуемые аппаратными методами.

Так, в Массачусетском технологическом институте (США) разработан интересный метод анализа через синтез, или просто "анализ — синтез". Этот метод появился в результате поиска путей сокращения количества информации о звуках речи.

В методе анализ-синтез имеется две модели: артикуляционная и акустическая.

а) Артикуляционная модель [39]

В ЭВМ имеется ограниченный набор параметров, характеризующих определенную совокупность фиксированных положений артикуляционных органов. По этим параметрам на ЭВМ спектры синтезируются и сравниваются со спектром анализируемого звука. Таким образом, находится спектр, наиболее похожий на спектр анализируемого сигнала. Это позволяет описывать звуки в терминах параметров артикуляционных органов.

В работе [40] изложена методика применения артикуляционной модели для исследования возможности сегментации потока речи на фонемы. Использовались следующие артикуляционные параметры:

- 1) расстояние между губами,
- 2) координаты x, y высшей точки языка,
- 3) расстояние в этой точке между языком и небом.

Синтез спектра по этим данным производился на ЭВМ IBM 7090.

б) Акустическая модель

В памяти ЭВМ имеется набор параметров, характеризующих ограниченную совокупность фиксированных акустических картин. В качестве параметров используются частоты и амплитуды формант (обычно трех) и частота антиформанты.

По этим параметрам на ЭВМ синтезируются спектры и сравниваются со спектрами исследуемого звука.

Метод анализ-синтез в настоящее время используется для исследований. Он не разработан в такой степени, чтобы быть применимым непосредственно для автоматического отождествления звуков речи. Однако использование этого метода может привести к созданию удовлетворительных алгоритмов опознавания. Так, в работе [41] сообщается, что изучение фрикативных и гласных на акустической модели метода анализ-синтез привело к установлению схемы распознавания, которая в дальнейшем была успешно реализована (использовались звуко сочетания типа C-T-C).

При помощи метода анализ-синтез удается исследовать не только фонемы, но и сллабические элементы [42]. В настоящее время продолжается совершенствование рассматриваемого метода в Массачусетском технологическом институте и лабораториях Бедла [43].

Применение ЭВМ позволяет устранить трудности, связанные с

аппаратурными методами выделения инфразвуковых параметров. В Институте математики СО АН СССР разработана программа инфразвукового спектрального анализа, позволяющая устранить влияние постоянной составляющей, точнее прямоугольного импульса постоянного по направлению тока, возникающего в результате детектирования, предшествующего любому инфразвуковому анализу речевых сигналов [48].

Большая работа по анализу звуков речи с помощью ЭВМ ведется в лаборатории биологических вычислительных машин (БВМ) Иллинойского университета (США) [49, 50]. Разрабатывается методика распознавания фонем с использованием "анализатора динамических сигналов", изготовленного в той же лаборатории в 1962 г. по заказу ВВС США для целей акустического анализа (рис. 2).

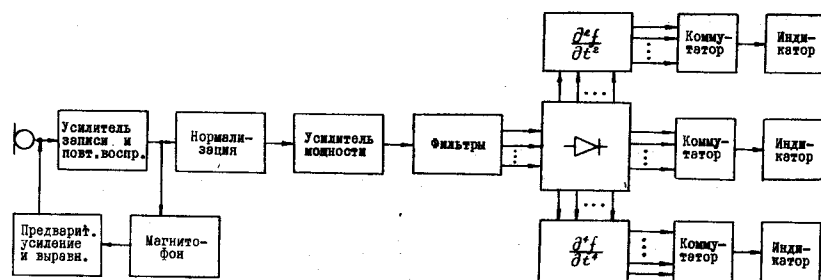


Рис. 2

Исследуемый диапазон частот перекрывает 8 октав (30–8000 гц). На каждую октаву приходится 12 фильтров. Итого весь диапазон перекрывается 96 фильтрами.

Нормализация звуков речи по громкости производится регулировкой усиления при повторном воспроизведении. После детектирования отфильтрованные сигналы подвергаются дифференцированию. Руководитель исследований д-р Форстер считает, что использование 2-й и 4-й производной приводит к уменьшению числа констант, описывающих звук речи в частотно-временном пространстве.

Индикация работы анализатора осуществляется при помощи

трехлучевого осциллоскопа. Коммутационные переключатели позволяют делать 1200 проб в сек на каждую полосу. Опыт показал, что по осциллоскопическим картинкам легко опознаются фонемы. В настоящее время ведутся работы по автоматизации процесса опознавания. Логическое устройство, выполняющее операции опознавания, строится на основе изготовленной в лаборатории БВМ аппаратуры расчета параметров зрительных образов с произвольными размерами, формой и местоположением. Эта аппаратура отличается высоким быстродействием: параметры 40 объектов рассчитываются в течение 200 мсек. Такое быстродействие обеспечивается применением большого числа аналоговых вычислительных устройств, работающих параллельно.

Предусмотрена возможность применения цифровой вычислительной техники, для чего выход 2-й производной квантуется на 8 уровней, а 4-й — на 32.

Для опознавания используются статистические методы. О конкретных результатах этой лаборатории сообщений в литературе нет. Форстер считает, что в конечном счете им удастся создать аппаратуру для опознавания фонем в реальном масштабе времени размером с портативную пишущую машинку.

Широко применяются ЭВМ для исследований по автоматическому опознаванию звуков речи в лаборатории распознавания образов фирмы "General Dynamics Electronics" (США) [50]. Речевой сигнал предварительно проходит через 30-канальный полосный анализатор. С помощью ЭВМ IBM 1401 рассчитываются частоты F_1 , F_2 и F_3 через каждые 10 мсек. Участки речи, на которых частоты формант не меняются, из рассмотрения выпадают.

Был проведен эксперимент по опознаванию 10 моносyllабических слов, произнесенных 3 дикторами (мужчинами) по 3 раза. Правильно опознаны все 90 образов. Также успешно произведен эксперимент по опознаванию диктора на том же материале.

Ввод данных в вычислительную машину осуществляется специальным преобразователем, разработанным в Массачусетском технологическом институте. Одновременно с вводом данных в IBM 1401 осуществляется их запись в цифровой форме на магнитную ленту, что позволяет осуществлять обработку данных на других типах ЭВМ.

По заданию Научно-исследовательского центра ВВС в Реме лаборатория прикладных исследований фирмы "Sylvania Electric Products" (США) ведет поисковые работы по анализу и синтезу речевых сигналов с помощью ЭВМ [51]. Метод исследований заклю-

чается в следующем: речевой сигнал вводится в ЭВМ и подвергается тому или иному математическому анализу. Затем полученные символы используются для синтеза. По качеству звучания на выходе определяется эффективность принятого метода анализа. Лучший результат получается при аппроксимации речевого сигнала 30-ю ортогональными функциями в виде экспоненциально затухающего ряда Фурье. При изменении звука меняются коэффициенты, сами же функции остаются неизменными. Все эксперименты проводятся на ЭВМ, никакой аппаратуры не создается.

В Японии широкие исследования в рассматриваемой области сосредоточены в основном в 3-х организациях: Nippon Electric Company, Radio Research Laboratories и Университет в Киото. Ими используется предварительный спектральный анализ с помощью гребенки фильтров. Последующая обработка осуществляется в ЭВМ. Цель исследований - сегментация и опознавание фонем. В качестве параметров в основном используются формантный состав, различные моменты спектра и частота переходов через нуль клипированной речи [44, 51, 52].

Разработана программа определения частоты основного тона по автокорреляционной функции речевого сигнала, расчет которой осуществляется на ЭВМ MEAC II03 [45].

В связи с разработкой метода анализ-синтез стали применяться различные программы синтеза звуков речи при помощи ЭВМ. Кроме Массачусетского технологического института значительных результатов в этой области достигли исследователи Белловских лабораторий. Там разработаны программы синтеза речи с помощью машины IBM 7090 [47].

Имеются программы двух способов синтеза:

а) формантное кодирование (используется информация только о полюсах спектра);

б) кодирование в терминах размеров полостей речевого тракта.

На входе машины задается последовательность фонем, на выходе получается форма звуковой волны. Основная трудность программы - сглаживание переходов между фонемами.

При реализации артикуляционной модели синтеза речевой тракт представляется в виде последовательности 21 секции длиной по 0,8 см каждая. Хотя система кодирования в терминах размеров полостей речевого тракта представляется более эффективной, она еще не разработана в такой степени, чтобы сравниться с системой формантного кодирования по качеству получаемого звучания.

В связи с разработкой машинных методов анализа звуков речи появилось много работ, описывающих устройства ввода речевых сигналов в ЭВМ. Эти устройства можно разбить на два класса:

1) преобразователи типа аналог-цифра, обеспечивающие ввод в ЭВМ непосредственно речевых сигналов [51, 54, 55];

2) аппаратура предварительного выделения параметров речевых сигналов и последующего ввода их в ЭВМ [36, 37, 50].

Второй класс устройств используется более широко, чем первый, так как быстродействие современных ЭВМ не позволяет в реальном масштабе времени выделять достаточно большое число параметров непосредственно из речевых сигналов. Наибольшие затруднения в этом отношении представляет спектральный анализ на ЭВМ. В то же время использование аналоговых фильтров затрудняет оперативность в выборе числа полос, их ширины и расположения по звуковому диапазону, приводит к необходимости построения громоздкой аппаратуры предварительной обработки сигналов.

Весьма заманчивым выходом из создавшегося положения представляется предложение фирмы "Spectra Gyroscopu Co" (США) использовать скенетрон (волоконную оптику) в качестве спектрального анализатора [50, 56]. Стекловолокна (в количестве от нескольких сотен до нескольких тысяч) различной длины и диаметра ($d = 0,05$ мм) укрепляются одним концом на суппорте. На этот же конец стекловолокон направлен источник света. Суппорт приводится в движение усиленным звуковым сигналом. Лучи света, выходящие из тех волокон, собственные частоты которых близки к частоте звукового сигнала, описывают наиболее длинные траектории на специальной мишени, за которой расположены считывающие фотодиоды. Ток в фотодиодах пропорционален размаху колебаний соответствующих стекловолокон, а следовательно, и энергии звукового сигнала, приходящейся на данные частоты. Построенный скенетрон анализирует диапазон частот от 100 гц до 20 кГц. Объем этой модели $5,5 \cdot 10^{-2}$ см³, плотность упаковки 20 тысяч частотных сигналов на 1 см³.

В настоящее время на этой фирме проводятся работы по созданию устройства типа "Sceptron" для автоматического распознавания ограниченного набора команд.

Использование волоконной оптики в качестве спектрального анализатора дает возможность оперативно, с большой точностью и без дополнительных затрат менять количество полос, их шири-

ну, размещение по звуковому диапазону простым суммированием токов соответствующих фотодиодов.

IV. НАПРАВЛЕНИЯ РАБОТ ПО АВТОМАТИЧЕСКОМУ РАСПОЗНАВАНИЮ ЗВУКОВ РЕЧИ

До настоящего времени остается неясным вопрос о выборе параметров, по которым можно эффективно опознавать звуки речи. Исследователи выделяют и используют возможно большее число параметров. Это приводит к необходимости перерабатывать огромный поток информации. Аппаратура получается громоздкой, ненадежной и дорогостоящей, не выходящей за рамки лабораторных образцов. К сокращению потока данных, как отмечалось выше, направлены работы по оценке информационной эффективности исследуемых параметров.

Кроме того, уменьшение числа параметров и разработка оптимальных алгоритмов распознавания звуков речи по этим параметрам существенно зависит от успехов в изучении процессов речеобразования и слухового восприятия.

Один из экономически оправданных путей исследования, развитый в настоящее время, заключается в разработке методов и устройств автоматического распознавания ограниченного набора команд по небольшому числу параметров.

И, наконец, продолжает развиваться направление работ по поиску и определению значимости различных параметров для распознавания фонем. Рассмотрим состояние исследований в указанных направлениях.

а) Фундаментальные исследования по речеобразованию и восприятию звуков речи

В этом направлении наиболее успешные исследования проводятся в Институте физиологии им. И.П.Павлова АН СССР, лабораториях Хаскинса (США), Королевском технологическом институте (Швеция), Единбургском и Лондонском университетах и др.

Теория слухового восприятия, разрабатываемая в лабораториях Хаскинса [31], построена на том предположении, что восприятие звуков речи включает процесс декодирования на основании эквивалентных артикуляционных картин. Звуковая последовательность преобразуется в последовательность двигательных команд. Эта теория, по-видимому, еще далека до завершения, так как не предложено ни одной модели нервной сети.

В Исследовательской лаборатории ВВС (Кембридж, США) [38] проведены работы по изучению параметров, несущих информации об эмоциональной окраске речи. Исследовались 8 стандартных фраз, произнесенных 3 дикторами (мужчинами) в разном эмоциональном состоянии (оттенки страха, счастья, вопроса и т.д.). Угадывание эмоциональной окраски при воспроизведении неисканной речи составило 85%. Затем из исходного сигнала была выделена частота основного тона, и в соответствии с ее изменением модулировалась частота специального генератора. Угадывание эмоциональной окраски при прослушивании сигналов этого генератора составило 44%. При добавлении информации об изменении амплитуд угадывание возросло до 47%. Наличие только амплитудной информации дает распознаваемость всего 14%. Таким образом, для выяснения эмоциональной окраски речевого сигнала недостаточно информации только об огибающей и основном тоне. Большое значение имеет и фонетическое содержание фразы. В ходе исследований выяснилось, что различные эмоции передаются разными параметрами. Это еще раз подчеркивает тот факт, что речевой сигнал описывается сложной системой взаимозависимых параметров и вряд ли удастся найти ключевые параметры, которые были бы одинаково эффективны во всех ситуациях.

Общая цель, характеризующая большинство современных работ по автоматическому распознаванию речи — это уменьшение числа параметров речевых сигналов. К этому же направлены и работы по изучению восприятия синтетической речи. Базой для подобных исследований служат различного рода синтезирующие устройства.

В лаборатории электроники Научно-исследовательского отделения фирмы "Melpar" (США) разработан прибор *EVA* — электронный голосовой аналог, синтезирующий звуки речи по кривым, нанесенным от руки [33]. В приборе имеется программируемый генератор функций и управляемый синтезатор речи формантного типа. На специальной ленте можно рисовать кривые изменения 12 параметров речи: высоты звука, частот формант, амплитуды носовых звуков, амплитуды фрикативных звуков, амплитуды возбуждения источника и частот полюсов переходной функции носового тракта. Прибор *EVA* предназначен для фундаментальных исследований в области синтеза речи и оценки различных принципов распознавания речи.

В Королевском технологическом институте (Стокгольм) раз-

работан синтезатор формантного типа *OVE II* [35]. Синтезатором управляет функциональный генератор, в котором проводящие чернильные линии используются для образования управляющих сигналов. Синтезатор содержит три фильтровые системы, каждая из которых состоит из последовательно включенных контуров. Эти три системы предназначены для формирования вокалических, назальных и фрикативных шумных звуков.

В Массачусетском технологическом институте создана динамическая аналоговая модель речевого тракта *DAVO*, управляемая вычислительной машиной *TX-0*, в памяти которой имеется большая совокупность конфигураций речевого тракта, позволяющая синтезировать отрезки непрерывной речи, длительностью до 1 сек [34].

Исследования по восприятию синтетической речи в Эдинбургском университете проводятся на базе синтезатора гласных звуков *PAT* [34]. В ходе этих исследований показано, что темп речи влияет на локус 2-й форманты, но не влияет на скорость ее перехода. Этот результат противоречит данным, полученным в лабораториях Хаскинса. Окончательное решение подобных вопросов весьма интересно, так как при этом выяснится роль и устойчивость переходных (динамических) характеристик речевых сигналов.

б) Разработка методов и устройств автоматического распознавания ограниченного набора команд

Это направление работ представляет большой практический интерес. Вот краткий перечень возможных применений подобных устройств:

- управление станками и механизмами;
- ввод устных команд в ЭВМ;
- работающий с голоса телефонный номеронабиратель;
- сужение ширины каналов специальной связи;
- управление различными системами в особых условиях, когда руки оператора не могут быть использованы для работы с рычажными или кнопочными пультами (например, в условиях сильных перегрузок) и т.д.

К настоящему времени разработана единственная серийная модель устройства, воспринимающего устные команды. Это "Шубокс" (фирма IBM, США) - портативный электронный арифмометр, реагирующий на произнесенные вслух 10 цифр (от 0 до 9) и 6 ко-

манд: плюс, минус, сумма, частная сумма, ошибка, сброс. После слова "ошибка" гасятся все произведенные операции. Устройство для распознавания слов собрано всего лишь на 31 транзисторе. В качестве параметров для распознавания используются сведения о групповой принадлежности первой фонемы, местоположении ударной гласной, групповой принадлежности последней фонемы и временной огибающей положительных и отрицательных пиков оциллограммы слова [57]. Помехоустойчивость устройства "Шубокс" низкая. Для нормальной работы прибора приходится менять голос, подлаживаясь к аппарату.

Фирмой "Bulova Electric" (США) разработана первая модель машинки "Аудитрэн", распознающей слова [58]. Произносимые слова записываются на магнитную ленту. Затем они нормализуются по длительности (регулировкой скорости протяжки ленты при воспроизведении) и по амплитуде (ручной регулировкой усиления). После этого слова вновь записываются на магнитную ленту. При повторном воспроизведении речевой сигнал усиливается, детектируется и сглаживается с помощью фильтра нижних частот (ФНЧ). Выделенная таким образом огибающая слова сравнивается с эталонными огибающими. Сравнение производится на аналоговом вычислительном устройстве. Оpoznание слов осуществляется по критерию

$$\Delta = \int_0^T |F_{\text{pear}}(t) - F_{\text{эт}}(t)| dt;$$

где $F_{\text{pear}}(t)$ - огибающая реализации;
 $F_{\text{эт}}(t)$ - эталонная огибающая.

По минимальному значению Δ срабатывает реле, зажимающее соответствующую лампочку световой индикации. Первая модель машинки "Аудитрэн" опознавала 10 слов (цифр), произнесенных всеми дикторами (количество их не указывается) с надежностью 99,96%. Сообщается, что ведется разработка 2-й модели машинки на 300 слов. Для расчета Δ используется несколько аналоговых вычислительных устройств, работающих параллельно, что позволит осуществлять анализ и распознавание в реальном масштабе времени. Разработчики считают, что используемый ими принцип позволит осуществить распознавание 1000 слов.

Японской фирмой "Nippon Electric Co" [59] изготовлено устройство различения произносимых вслух цифр. Это устройство можно использовать вместо телефонного номеронабирателя, если игнорировать экономический фактор. Эталоны цифр строились на основе 500 реализаций. В качестве параметров используются:

- количество голосовых звуков в слове;
- длительность слова;
- положение центра 1-й форманты на начальном участке слова;
- положение центра 2-й форманты на начальном участке слова.

Опознавание производится по Байесовской стратегии. Надежность опознавания цифр для 1-го диктора составила 99,7%, для 20 дикторов - 97,9%. В настоящее время ведется усовершенствование прибора. Разработчики считают, что без существенных переделок им удастся довести надежность опознавания для 20 дикторов до 99%.

В Технологическом институте Кейса (США) разработано устройство перевода устных команд в цифровую форму для управления станками. Прибор опознает 14 слов: цифры от 0 до 9, x , y , "минус" и "точка". Сообщается, что достигнута высокая распознаваемость для всех дикторов. Она же зависит от интонации, акцента и пр. Сведения о методе опознавания и используемых параметрах не приводятся, так как подана заявка на патент.

Работа по созданию методики распознавания ограниченного набора команд проводится в отделе наземной связи фирмы "RCA" (по заказу ВВС США) [50]. К настоящему времени удалось добиться распознавания ограниченного набора моносиллабических слов (типа CIG) с надежностью 85%. Речевой сигнал проходит через фильтры, затем кодируется и вводится в ЭВМ RCA 501. Программа вычислений позволяет выделять и опознавать гласные независимо от диктора. Окончательные результаты неизвестны.

Интересная попытка использования визуальных данных о движении губ для распознавания команд сделана в Морской электронной лаборатории (Сан-Диего, США) [47]. Устройство считывания движения губ смонтировано следующим образом. С обеих сторон губ установлены осветители с направляющими рефлекторами. Прямо перед губами расположено фотосопротивление с собирающим рефлектором. Фотосопротивление включено в одно из плеч сбалансированного моста. Напряжение, снимаемое с моста, подается на дифференциальный усилитель и затем на самописец. Осветители и фотосопротивление с рефлекторами смонтированы на провололочной оправе, которая одевается как очки.

Проведены эксперименты по распознаванию 10 цифр, произнесенных полным стилем (вместо слова "zero" произносилось слово "oh"). Все данные по надежности опознавания получены

для 1-го диктора, так как в эксперименте участвовало всего 2 диктора, что явно недостаточно для набора статистических эталонов, относящихся ко многим дикторам.

Каждое из слов было произнесено обоими дикторами по 20 раз. В промежутках между словами губы были плотно сжаты.

Сначала распознавание осуществлялось вручную. Группе людей для обучения дали кривые 3-х реализаций каждого слова, произнесенных обоими дикторами. Остальные реализации должны были быть опознаны. Средняя надежность опознавания для 1-го диктора составила 91%, для 2-го - 78,3%.

Затем процесс опознавания был автоматизирован. Кривые, нормированные по амплитуде и длительности, были закодированы (6 разрядов) и введены в ЭВМ. Эталоны строились по 10 реализациям. Метод распознавания - статистический, каждая точка кривой считается независимым параметром. Надежность опознавания достигла 71%.

Для повышения надежности был привлечен еще один параметр: скорость движения потока воздуха около губ. Одна из осветительных лампочек была заменена датчиком анемометра (та же лампочка, но со снятым стеклянным колпаком; во избежание сгорания нити накала на воздухе применялось пониженное питающее напряжение). В этом случае автоматическое распознавание реализаций, участвовавших в создании эталона составило 100%, для остальных реализаций - 81%.

В настоящее время исследователи ведут работу по привлечению акустических параметров слова для опознавания.

в) Работы по поиску и определению значимости различных параметров для распознавания фонем

Это направление работ имеет самую длинную историю. Интерес к нему не уменьшается, так как решение задачи надежного автоматического распознавания фонем было бы решением задачи автоматического распознавания речи в целом.

В течение последних двух лет в печати не появлялось сообщений о принципиальных сдвигах в этой области, поэтому достаточно полное представление по этому поводу дают ранее опубликованные обзоры [1, 2].

На Стокгольмском семинаре [84] несколько докладов было посвящено вопросам распознавания фонем. В. Ясsem (лаборатория

передачи речи Королевского технологического института, Швеция) сообщил об исследовании фриктивных звуков шведского, английского и польского языков [45]. Он доказал, что признаки этих фонем общие для всех дикторов лежат в диапазоне частот от 5 до 13 кГц. Признаки же, характеризующие индивидуальные особенности дикторов, приходятся на область частот 2,3 и 4-й форманты.

В докладе Лилленкранца Н. (из той же лаборатории) сообщается о сегментации непрерывного потока речи на отрезки (по признаку наличия или отсутствия голосового источника) и о разбиении фонем на два класса: сонорный — несонорный (по соотношению энергий в нижней и верхней частях звукового диапазона) [46]. Несколько докладов (Козна А. и Харта Д.; Метьюза М.В., Миллера Д.Е. и Девида Е.Е.; Фанта Г. и др.) было посвящено исследованию формант [34]. В этих докладах показано, что само выделение формант связано с большими погрешностями, которые принципиально нельзя устранить. Кроме того, многие форманты, наблюдаемые в естественной речи, не всегда существенно важны. Таким образом, частоты формант, по-видимому, не являются решающими параметрами для распознавания фонем и без использования других параметров вряд ли приведут к решению задачи.

В Лаборатории акустики (Университет Миссисипи, США) проведены исследования значимости некоторых параметров, характеризующих согласные фонемы [30]. Установлено, что эффективность признаков убывает в следующей последовательности: 1) способ образования, 2) наличие голоса, 3) место образования.

Обещающее направление исследований по автоматическому распознаванию фонем принято в Иллинойском университете [49] и [50] (см. § 3). По мнению специалистов, это направление является в США наиболее перспективным.

ЛИТЕРАТУРА

1. Сапожков М.А. Речевой сигнал в кибернетике и связи, М., Связьиздат, 1963.
2. Цеммель Г.И. Автоматическое (объективное) распознавание звуков речи (обзор). Зарубежная радиоэлектроника, 1961, № 14, стр. 52-73.
3. Advances in Computers, New York-London, Acad. Press, 1960, vol. 1, p. 83.
4. Барабан Ю.Л. и др. Автоматическое распознавание образов, Киев, 1963 (КВАНУ).
5. Sackschewsky V.E., Olstreichner H.L. Pattern Recognition as a Problem in Decision Theory and Application to Speech Recognition, vol. Mil-7, Apr.-July 1963, N 2, 3, p. 186.
6. Перцептроны. Книжки и журнальные статьи 1958-1963 гг. 53 назв. яз., дек. 1963, ГБЛ. 13531/Н.
7. Electronics (p.n) 1962, vol. 35, N 6, p. 34-37.
8. Electronic Design, 1961, vol. 9, IX, N 19, p. 34-37.
9. Андерсон Т. Введение в многомерный статистический анализ. Пер. с англ. М., Физматгиз, 1963.
10. Мидлтон Д. Введение в теорию статистических решений. Пер. с англ. под ред. Левина Б.Р. М., "Сов. радио", I. I 1962, I. II 1961.
11. Sebestyen G.S. Decision-making Processes in Pattern Recognition, 1962, New York, Macmillan Company.
12. Highleyman W.H. Linear Decision Function, With Application on Pattern Recognition, Proc. IRE, 1963, vol. 50, pp. 1501-1514.
13. Cooper P.W. The Hyperplane in Pattern Recognition. Cybernetica, 1962, vol. 5, N 4, pp. 215-239.
14. Fischer Martin A. Hyperplane Techniques in Pattern Recognition. Proc. JRE, 1963, vol. 51, N 3, p. 497-498.
15. Браверман Э.М. Опыты по обучению машины распознаванию зрительных образов. Автоматика и телемеханика, 1962, ТХХI, I, № 3 стр. 349-364.
16. Bledsoe W.W. and Browning J. Pattern Recognition and Reading by Machine, Proc. of the Eastern Joint Computer Conf., 225, 1959.
17. Streck G.P. Stochastic Model for the Browning-Bledsoe Pattern Recognition Scheme. JRE Trans. on Electr. Comp., 1962, EC-II, pp. 274-282.

18. Sebestyen G.S. Recognition of Membership in Classes, JRE Trans. on Information Theory, 1961, JT-7, pp. 44-50, Jan.
19. Ming-Kuei Hu. Visual Pattern Recognition by Moment Invariants, JRE Trans. on Inf., 1962, vol. JT-8, N 2, pp. 179-187.
20. Загоруйко Н.Г. Об обмене устной информацией между человеком и вычислительными системами. Вычислительные системы (Сб. трудов) Новосибирск, 1964, вып. 10, стр. 3-12 (Ин. мат. СО АН).
21. Файнштейн А. Основы теории информации., М., Изд-во иностр. лит., 1960 г.
22. Загоруйко Н.Г. Методика оценки информационной эффективности независимых параметров речевого сигнала. Вычислительные системы (Сб. трудов), Новосибирск, 1964, вып. 10, стр. 13-17 (Институт математики СО АН СССР).
23. Marill T., Green D.M. On Effectiveness of Receptors in Recognition Systems, JRE Trans. on Inf. Theory, 1963, vol. JT-9, N 1, Jan., 1963.
24. Беллман Р. Динамическое программирование. Пер. с англ. М., ИЛ., 1960.
25. Сб. "Вопросы статистики речи", Л., Изд-во ЛГУ, 1958.
26. Ёлкина В.Н., Юдина Л.С. Статистика слогов русской речи. Сб. Вычислительные системы (Сб. трудов), Новосибирск, 1964, вып. 10, стр. 58-78 (Институт математики СО АН СССР).
27. Алексеев П.М., Калинин В.М., Чернядьева Е.А. О статистических закономерностях в современных русских и английских текстах по электронике. Вопросы радиоэлектроники, серия VIII, вып. 3, стр. 97-112.
28. Глезер В.Д. и др. Об опознавании образов в зрительной системе. Сб. "Биологические аспекты кибернетики". М., Изд-во АН СССР, 1962, стр. 164-173.
29. Ёлкина В.Н., Юдина Л.С. Статистика открытых слогов русской речи. (Данный сборник).
30. Peters R.W. Dimensions of Perception for Consonants, JASA, vol. 35, N 12, 1963.
31. Liberman A.M., Cooper F.S., Harris K.S., Macneilage P.F. A Motor Theory of Speech Perception. Speech Communication Seminar, Stockholm, 1962.
32. Kelly J., Anthony J.K. Tempo and Transition. Speech Communication Seminar, Stockholm, 1962.
33. Electronics, 1963, N 41, pp. 46-48.
34. Фант Г. Семинар по речевой связи (обзор). Акустический журнал, 1963, I. IX, № 2.
35. Fant G., Martony J., Rengman V., Risberg A. Synthesis strategy. Speech Communication Seminar, Stockholm, 1962.
36. Harris C.M. Pitch Extraction by Computer Processing of High-Resolution Fourier Analysis Data. JASA, 1963, vol. 35, N 3.
37. Weiss M.R., Harris C.M. 1963, JASA, vol. 35, N 2, pp. 207-214.
38. Lieberman P., Michaels S. JASA, 1962, vol. 34, N 7, pp. 922-927.
39. Heinz J.M. An Analysis of Speech Spectra in Terms of a model of Articulation. Speech Communication Seminar, Stockholm, 1962.
40. Inomata S. Program for Active Segmentation and Reduction of Phonetic Parameters. Speech Communication Seminar, Stockholm, 1962.
41. Hogan D.L. Some Studies of Fricative Using Automated Analysis by Synthesis Techniques. Speech Communication Seminar, Stockholm, 1962.
42. House A.S., Paul A.P., Stevens K.N., Arnold J.B. Acoustical Description of Syllabic Nuclei: Data Derived by Automatic Analysis Procedures. Speech Communication Seminar, Stockholm, 1962.
43. Mathews M.V., Miller J.E., David E.E. Strategies for Automatic Pole-Zero Analysis of Speech. Speech Communication Seminar, Stockholm, 1962.
44. Sakai T. et al. Phonetic Typewriter. Speech Communication Seminar, Stockholm, 1962.
45. Jassem W. Noise Spectra of Swedish, English and Polish Fricatives. Speech Communication Seminar, Stockholm, 1962.
46. Liljencrants J. A Few Experiments on Voiced-Voiceless Identification and the time segmentation of speech. Speech Communication Seminar, Stockholm, 1962.
47. Hillix W. Use of Two Nonacoustic Measures in Computer Recognition of Spoken Digits, JASA, 1963, vol. 35, N 12.
48. Волошин Г.Я. К вопросу об инфразвуковом анализе речевых сигналов. Проблемы автоматического распознавания образов. Киев, 1964 г. (КВИАУ).
49. AF Unit Designed to Study Basic Elements of Speech. Electronic News, 1962, N 19, p. 36.
50. Varied Approaches Used to Develop System for Reliable Recognition of Voice Commands. Electronic News, 1963, N 383, pp. 4-5.
51. Electronics, August 4, 1961, N 31.

52. Судзуки Даоудзи и др. Анализ речи с помощью вычислительной машины. Дэмпа кюкюсё кихо, Rev. Radio Res. Labs, 1962, 8, N 36, 277-285.
53. Harris C.M., Waite W.M. Response of Spektrum Analyzers of the Bank-of-Filters Type to Signals Generated by Vowel Sounds. JASA, 1963, vol. 35, N 12.
54. Волошин Г.Я. Преобразователь аналог-цифра для ввода речевых сигналов в ЭВМ. Вычислительные системы (Сб. трудов) Новосибирск, 1964, вып. 10, стр. 46-53. (Институт математики СО АН).
55. Лейбман Ю.А., Соболев В.Н. Преобразователь аналог-цифра для ввода речевого сигнала в вычислительную машину. Электросвязь, 1963, № 8.
56. Scepton - Fiber Optics Pattern Recognition. Electr. Inds., 1963, 22, N 1, 134, 235.
57. Machines Controlled by Spoken Commands. Data Processing, 5, N 2, 1963.
58. Машина для распознавания речи "Аудитрен", РЭЗР, 1961, вып. 39.
59. Japan Firm Builds Spoken Voice Digit Recognizer. Electronic News, 1963, vol. 8, N 390, p. 41.
60. Gamba A.J. oth. Preliminary Experimental Results with PAPA no. 2. Nuovo Cimento, 1962, 23, Suppl., N 2, 280-284.
61. Electronics, 1963, vol. 36, N 30, p. 7-8.
62. Aviation Week, 1963, vol. 78, N 9, p. 84-85.