# ПРОБЛЕМА РАСПОЗНАВАНИЯ РЕЧЕВЫХ СИГНАЛОВ КАК СЛОЖНАЯ СИСТЕМА

Н. Г. ЗАГОРУЙКО*

Речевой сигнал, переходя в процессе восприятия с одного иерархического уровня на другой, претепревает многократную перекодировку. Теория сложных систем (1) позволяет найти ответ на следующий важный вопрос: *В чем смысл иерархической структуры восприятия речи?*

Ответ состоит в том, что: 1. в результате перекодировки удаётся согласовать характеристики сигнала на выходе одной ступени с характеристиками простого классификатора следующей ступени; 2. при многоуровневой системе требуется небольшая оперативная память; 3. на каждом уровне появляется возможность использовать дополнительную информацию о речи и языке, которая может храниться в долговременной памяти.

В результате, с помощью простых структур при малых затратах оперативной памяти удается достигать надежного распознавания устного сообщения.

*За счет чего можно упростить структуру каждого уровня?*

Нами обнаружено (2), что в процессе принятия решения в многомерном пространстве признаков человек использует самый простой тип решающих границ-гиперплоскости, перпендикулярные осям координат. Это накладывает ограничение на характер распределения распознаваемых элементов-распределения должны обладать малой дисперсией. Уменьшение дисперсии может достигаться путем различных видов нормализации — по громкости, темпу, особенностям дикции и т. д. Экспериментальное исследование алгоритма подстройки эталонов под диктора (3) показала явную целесообразность этого вида адаптации.

В очень сильной степени сложность процедуры распознавания зависит от количества распознаваемых объектов. Предварительное сокращение промежуточных алфавитов может быть получено с помощью алгоритмов таксономии (4, 5). Большой выигрыш во времени дает сокращение альтер-

* Институт математики Сибирского отделения АН СССР г. Новосибирск.

натив за счет априорной информации, а так же за счет информации, поступающий в процессе распознавания. Алгоритмы, реализующие такую возможность, иммитируют хорошо известный в психологии феномен „установки“ и предложенный Л. А. Чистович метод „вычеркивания“ (6).

Рассмотрение известных фактов о воспроятии речи человеком с позиций теории сложных систем приводит нас к варианту структуры многоуровневого распознающего устройства, представленного на рис. 1. Основные элементы этой схемы обсуждались с Бондарко Л. В., Молчановым А. П., Кожевниковым В. А. и особенно часто с Чистович Л. А. Поэтому, заключительную часть доклада можно считать результатом нашей совместной работы.

Нам представляется, что речевой сигнал $f(t)$ преобразуется в самом начале в некоторое довольно полное описание в пространстве $(X_1)$ частота-время. На участке длительностью в несколько сотен миллисекунд определяются некоторые признаки $(S_1)$ типа статических и динамических характеристик формант, характеристик шумовой части спектра и т. д. Разумеется, надежность разпознавания $(P_1)$ этих признаков может оказаться недостаточно высокой, даже после использования всех возможностей классификатора $D_1$. Повысить эту надежность можно за счет использования блоком $H_1$ информации о физических законах речеобразования такого типа: частота основного тона не может меняться быстрее некоторой величины; одновременное существование таких-то признаков невозможно; вслед за такой комбинации признаков наиболее вероятно появление таких-то признаков и т. д. Использование информации подобного рода продолжается до тех пор, пока вероятность некоторого варианта признаков не станет больше вероятности других вариантов на некоторую пороговую величину $\Delta P_1$. Этот вариант поступает на вход следующей ступени преобразования, где последовательность таких признаков образует пространство описания $X_2$.

Если разница вероятностей меньше $\Delta P_1$, то запоминается несколько вариантов признаков $S_1$.

На втором уровне происходит распознавание фонем $(S_2)$, вернее их контекстуальных вариантов (звукотипов). С этой целью классификатор $D_2$ использует информацию на участке типа открытого слога, границы которого определяются сегментатором $C_2$. Для сокращения числа возможных вариантов некоторой фонемы используется информация, содержащаяся в описании $X_2$, а затем, если необходимо, и информация о структуре последовательности фонем. С этой целью блоком $H_2$ формируются последовательности наиболее вероятных вариантов фонем и, с учётом всех этих постериорных и априорных сведений, выбирается наиболее вероятная последовательность (7). Если разность вероятности этой избранной последовательности и любой другой превышает некоторый порог $\Delta P_2$, то фонемный код слога передается на вход следующего блока. В противном случае категорического решения не принимается и запоминаются коды фонем $S_2$ несколь-



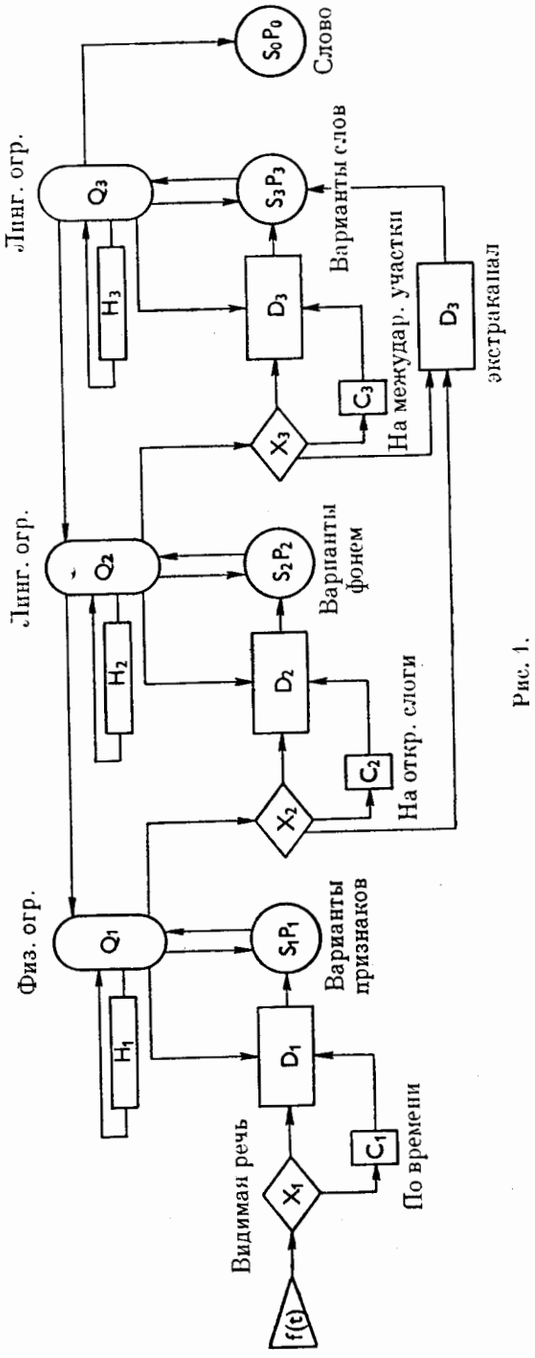


Рис. 1.

ких (наиболее вероятных) слогов. Если этих вариантов слишком много, то можно повторить процедуру, вызвав на вход блока другой вариант признаков $X_2$.

Для распознавания слов из словаря $S_0$ пространство $X_3$ кроме кодов фонем должно содержать информацию и об ударениях. Сегментатор $C_3$ осуществляет членение речевого потока на участки от ударения до ударения. В двух таких соседних участках содержится, как минимум, одно слово словаря $S_0$. Поиск нужного слова и одновременное определение его границ может осуществляться с помощью алгоритма Лисенко (8). На этом этапе, как и раньше, при выборе решения могут использоваться (блоком $H_3$) дополнительные априорные сведения об элементах словаря и, если окажется необходимым, осуществляться вызов (по линии $Q_3$—$Q_2$) на вход ($X_3$) нескольких вариантов фонемных последовательностей.

Наряду с этим должен существовать механизм ($D_3$) распознавания слов по неполной информации — например, только по началу или только по некоторому ограниченному набору признаков на определенном участке слова.

Рассмотрение такой модели позволяет наметить цели конкретных исследований, необходимых для конструирования искусственных систем распознавания речи.

## *ЛИТЕРАТУРА*


1. Гуд, Г. Х., Макол, Р. Э.: Системотехника. Введение в проектирование больших систем. Перевод с англ. Изд-во „Советское Радио", Москва, 1962.
2. Загоруйко, Н. Г.: Какими решающими функциями пользуется человек? Сб. Тр. ИМ СО АН СССР „*Вычислительные системы*" вып. 28, Новосибирск, 1967.
3. Загоруйко, Н. Г., Лозовский, В. С.: Подстройка, под диктора при распознавании ограниченного набора устных команд. Сб. Тр. ИМ СО АН СССР „*Вычислительные системы*" вып. 28, Новосибирск, 1967.
4. Елкина, В. Н., Загоруйко, Н. Г.: Об алфавите объектов распознавания. Сб. Тр. ИМ СО АН СССР „*Вычислительные системы*" вып. 22, Новосибирск, 1966.
5. Загоруйко, Н. Г., Елкина, В. Н.: Алфавит с минимальной избыточностью. Сб. Тр. ИМ СО АН СССР „*Вычислительные системы*" вып. 28, Новосибирск, 1967.
6. Загоруйко, Н. Г.: Комбинированный метод принятия решений. Сб. Тр. ИМ СО АН СССР „*Вычислительные системы*" вып. 22, Новосибирск, 1966.
7. Волошин, Г. Я.: Об использовании языковой избыточности для повышения надежности автоматического распознавания речевых сигналов. Сб. тр. ИМ СО АН СССР „*Вычислительные системы*", вып. 28, Новосибирск, 1967.
8. Лисенко, D. М.: Выделение и морфологический анализ слов в речевом потоке. Диссертация. Ленинград, 1966.




## DISCUSSION

*Райнов:*

Вопрос: Каким образом осуществляется вычеркивание элементов, которые, как вы показали, выполняют функцию избыточности?

И второй вопрос: Какой процент надежности обеспечивает это устройство для распознавания?

*Загоруйко:*

К Райнову: 1. Алгоритмы сокращения избыточности алфавита описаны в работах Загоруйко Н. Г. и Елкиной В. Н. в трудах Института математики СО АН СССР (г. Новосибирск) „Вычислительные системы" № 22, 1967 г.

2. Языковая избыточность сильно повышает надежность распознавания фонем. Так, в опыте со 125 словами информация о правилах сочетания двух фонем повысила процент правильно распознанных фонем с 15 до 94. Этот алгоритм описан в работе Волошина Г. Я. в трудах Института математики СО АН СССР (г. Новосибирск) „Вычислительные системы", № 28, 1967 г.

К Фанту: Мы считаем, что элементами распознавания на первом уровне должны быть статические и динамические характеристики формант—частота, ширина, скорость движения. Кроме этого должны распознаваться признаки источника-шума, голоса и т. д.