УДК 62-5:007:621.391

АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ ОГРАНИЧЕННОГО НАБОРА УСТНЫХ КОМАНД

В.И.Величко, Н.Г.Загоруйко

Задача автоматического распознавания ограниченного набора устных команд привлекает внимание исследователей речевого сигнала, главным образом, с точки зрения получения практически приемлемых решений таких частных задач, как управление голосом автоматическими устройствами, устный ввод исходных данных в программ в ЭВМ и т.д. (см., например, [1]).

В данной статье описываются эксперименты по распознаванию словаря из 168 слов (таблица 1). В качестве исходного материала взяты все термины языка "Алгол-60", дополненные некоторыми терминами входного языка α -транслятора и названиями некоторых элементарных функций. Большинство названий букв русского и латинского алфавитов заменено собственными именами (как это принято в системах связи); кроме того, вместо "мю" используется "ми" и вместо "пси" - "пси-штрих".

Для описания речевого сигнала в данном эксперименте была принята следующая система признаков. Акустический сигнал пропускался через систему из 5 фильтров второго порядка (резонансные контуры LCR). Центральные частоты фильтров были выбраны равными 112,5 гц; 450 гц; 900 гц; 1800 гц; 7200 гц [2] . Полоса пропускания каждого фильтра по уровню 0,5 равнялась центральной частоте, что соответствует добротности фильтров $\sim$ 2,45.

Слово разбивалось на сегменты фиксированной длительности 14 мсек. Длительность сегмента выбрана превышающей максимально возможный период основного тона. В каждом сегменте подсчитывалась энергия сигнала в общей полосе E_0 и на выходе каждого из 5 фильтров E_i (i = 1,... 5). В качестве параметров,ха-

Таблица I

I. Один
2. Два
3. Три
4. Четыре
5. Пять
6. Шесть
7. Семь
8. Восемь
9. Девять
I0. Ноль
II. Плюс
I2. Минус
I3. Разделить
I4. Запятая
I5. Точка
I6. Пробел
I7. Десять
I8. Степень
I9. Скобка
20. Открыть
2I. Закрыть
22. Круглая
23. Умножить
24. Равно
25. Точка с запятой
26. Квадратная
27. Звездочка
28. Кавычки
29. Не
30. Больше
3I. Меньше
32. Двоеточие
33. Надчеркнуть
34. Подчеркнуть
35. Восклицательный
36. Твердый
37. Градус
38. Минута
39. Секунда
40. Влево
4I. Вправо
42. Вопросительный
43. Вниз
44. Плюс-минус
45. Набла
46. Альфа
47. Бэта
48. Гамма
49. Дельта
50. Эпсилон
5I. Дзэта
52. Эта
53. Тэта
54. Каппа
55. Лямбда
56. Ми
57. Ню
58. Кси
59. Пи
60. Ро
6I. Сигма
62. Тау
63. Фи
64. Хи
65. Пси-штрих
66. Омега
67. Аш
68. Исправление и выделение
69. Примечание
70. Целый
7I. Вещественный
72. Комплексный
73. Логический
74. Собственный
75. Начало
76. Конец
77. Масс..в
78. Стоп

79. Для
80. Шаг
8I. До
82. Цикл
83. Раз
84. Пока
85. Если
86. То
87. Иначе
88. На
89. Процедура
90. Результат
9I. Значение
92. Метка
93. Барабан
94. Лента
95. Скаляр
96. Вектор
97. Матрица
98. Строка
99. Функция
I00. Истина
I0I. Ложь
I02. После
I03. Печать
I04. Заменить
I05. Число повторений
I06. Идентификатор
I07. Конец отладки
I08. Переключатель
I09. Присвоить
II0. Дифференциал
III. Интеграл
II2. Троеточие
II3. Фигурная
II4. Синус
II5. Косинус
II6. Арксинус
II7. Арктангенс
II8. Логарифм
II9. Экспонента
I20. Корень
I2I. Русская
I22. Латинская
I23. Греческая
I24. Большая
I25. Малая
I26. Аля
I27. Борис
I28. Владимир
I29. Геннадий
I30. Дмитрий
I3I. Елена
I32. Женя
I33. Зинаида
I34. Иван
I35. И краткая
I36. Клим
I37. Лидия
I38. Михаил
I39. Николай
I40. Ольга
I4I. Павел
I42. Роман
I43. Сергей
I44. Татьяна
I45. Ульяна
I46. Фёдор
I47. Харитон
I48. Цецилия
I49. Чарли
I50. Шура
I5I. Щукарь
I52. Ы
I53. Мягкий
I54. Знак

Продолжение таблицы I

I55. Эдуард	I62. Дизъюнкция
I56. Юрий	I63. Конъюнкция
I57. Яков	I64. Импликация
I58. Ки	I65. Целое деление
I59. Ку	I66. От
I60. Дубльвэ	I67. Тождественно
I6I. Зэт	I68. Отрицание

рактеризующих сегмент, брались величины $\ell_n \frac{E_0}{E_i}$, т.е. сегмент описывался точкой в пятимерном пространстве. Слово описывалось как последовательность сегментов. Частичное обоснование выбора логарифмических параметров состоит в том, что слух подчиняется известному в физиологии закону Вебера-Фехнера [3], согласно которому минимально заметный прирост ΔJ внешнего воздействия на органы чувств пропорционален величине воздействия J

$$\Delta J \sim \Delta \beta J ,$$

откуда $\beta \sim \ell_n J$ - сила восприятия, пропорциональная логарифму внешнего воздействия.

Для сравнения различных сегментов выбрана мера сходства

$$\alpha = \frac{\alpha^2}{\alpha^2 + \rho^2} . \qquad (I)$$

Здесь

$$\rho^2 = \sum_{i=1}^{5} \left[\ell_n \frac{E_0^{(1)}}{E_i^{(1)}} - \ell_n \frac{E_0^{(2)}}{E_i^{(2)}} \right]^2$$

-квадрат евклидова расстояния между сегментами в пятимерном пространстве, α^2 принято равным 2 на основе предыдущих экспериментов [4]. Мера сходства α может принимать значения от I (для одинаковых сегментов) до 0 (для бесконечно удаленных).

Чтобы перейти от меры сходства α для сегментов к мере сходства A для слов, использована идея, изложенная в работе [5]. Пусть сравниваются две реализации одного и того же слова "восемь" (рис.I). По координатным осям отложим время от начала слова и условно отметим границы фонем. Слова из-за различия в произношении имеют, как правило, разную временную структуру - одни фонемы удлиняются, другие укорачиваются, причем в разной степени. Интуитивно очевидно, что для получения максималь-

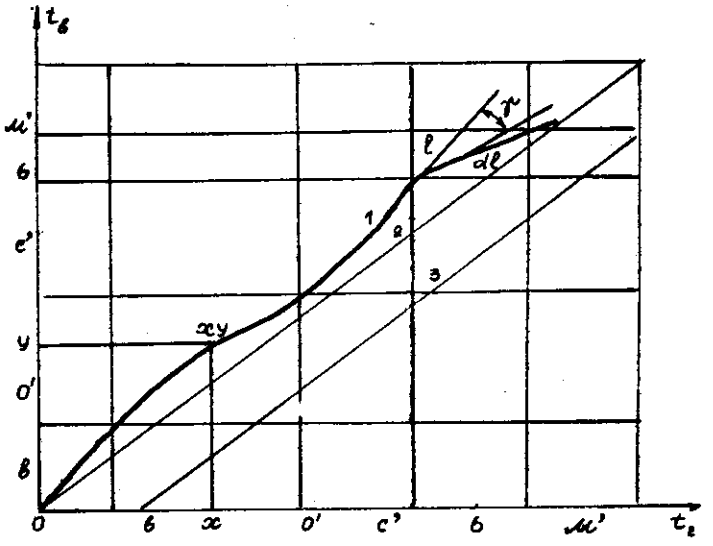

Рис.I.

ной меры сходства между одинаковыми словами (не определяя пока понятия меры сходства) желательно сравнивать похожие участки слов: "$б$" - "$б$", " О " - " О " и т.д. Пусть момент времени x первой ("горизонтальной") реализации слова соответствует моменту времени y второй ("вертикальной") реализации того же слова. Тогда сходство между параметрами "горизонтального" и "вертикального " слов в этих точках будет велико (например, в смысле меры сходства (I)). Построим в плоскости $t_г\ t_б$ геометрическое место точек, подобных точке $x\,y$ кривую I. Каждая точка этой кривой проецируется в "похожие" точки "горизонтального" и "вертикального" слов, а вся кривая является как бы "кривой максимального сходства". Пусть $\beta(\ell)$ - мера сходства между параметрами слов в момент, соответствующий точке ℓ кривой I. Тогда мера сходства B между словами можно определить как сумму мер сходства для всех точек кривой I (в непрерывном приближении сумма заменяется интегралом):

$$B = \int_{(I)} \beta(\ell)\, d\ell . \qquad (2)$$

Выражение (2) может быть скорректировано с учетом общей длины кривой I и наклона кривой I в различных точках. Ясно, что чем длинее слова, тем больше получается B. Такое влияние длины слова на величину сходства нежелательно. Далее, чем больше на-

клон кривой в какой-то точке отличается от $\pi/4$, тем больше разница в темпе произнесения слов, что должно быть учтено при вычислении меры сходства. Поэтому может быть принято уточненное определение меры сходства между словами:

$$B = \frac{1}{L}\int_{(l)} f(\gamma)B(l)dl \quad , \qquad (3)$$

где L – длина кривой I;

$f(\gamma)$ – убывающая функция угла $|\gamma|$ между биссектрисой координатного угла и кривой I, например, $\cos 2\gamma$.

Мера сходства B существенно зависит от кривой, вдоль которой она вычисляется. Нам заранее не известна "кривая максимального сходства", поэтому в соответствии с [5] определим B как максимум функционала (3) вдоль всех возможных кривых l ,соединяющих начала и концы сравниваемых слов. На кривые наложено единственное ограничение – они должны монотонно возрастать, т.е. $-\frac{\pi}{4} \le \gamma \le \frac{\pi}{4}$ (4)

Мера сходства (4) введена для непрерывного случая.Её вычисление сводится к решению вариационной задачи нахождения максимума функционала B . Нами принято дискретное описание слова в виде последовательности точек шестимерного пространства параметров. Переформулируем задачу для дискретного случая. Построим матрицу $\{a_{ik}\}$, элементами которой являются меры сходства (в смысле формулы (I)) между сегментами сравниваемых слов:

$$a_{ik} = \frac{\alpha^2}{\alpha^2 + \rho_{ik}^2} \quad . \qquad (5)$$

Здесь ρ_{ik}^2 – квадрат евклидова расстояния между k – м сегментом первого ("горизонтального") и i – м сегментом второго ("вертикального") слова. Нумерация строк матрицы совпадает с нумерацией сегментов "вертикального" слова, нумерация столбцов – с нумерацией сегментов "горизонтального" слова (см. рис.2).Матрица $\{a_{ik}\}$ является дискретным аналогом непрерывного поля мер сходства $B(l)$ (рис.I). Точка A на рис.2 соответствует началу координат на рис.I. Введем правило вычисления функционала сходства F между словами по матрице $\{a_{ik}\}$: функционал F вдоль пути ABCD... $PQST$ есть сумма тех элементов матрицы $\{a_{ik}\}$, через которые проходит выбранный путь, деленная на длину более длинного слова. В приведенном примере

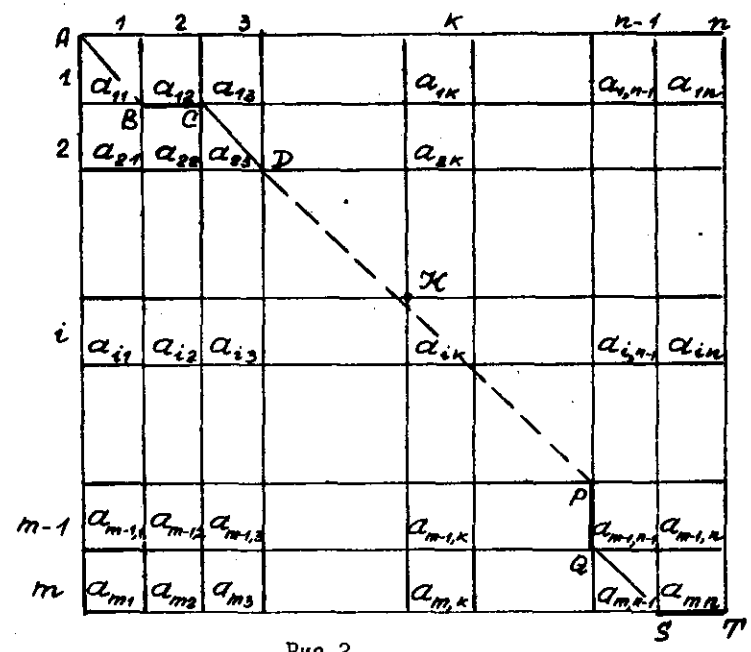

Рис.2.

при $n > m$ это будет сумма $\frac{1}{n}(a_{11} + a_{23} + \ldots + a_{m,n-1})$,соответствующая участкам пути $AB, CD, \ldots, QS$. Участки пути BC, $\ldots, PQ$, ST проходят между элементами матрицы и на сумму не влияют. На возможные пути накладывается ограничение: разрешенными являются только участки пути по горизонтали слева направо, по вертикали сверху вниз и по диагонали слева-сверху направо-вниз.

Мерой сходства A между словами будем считать максимум функционала F , вычисленного вдоль всех возможных путей на матрице $\{a_{ik}\}$: $A = \max_{A \ldots T} F$. Нетрудно видеть, что матрица $\{a_{ik}\}$ представляет собой граф или сеть (см.,например, [6] , а задача максимизации функционала сходства есть задача нахождения максимального пути на графе. В данном эксперименте длина максимального пути вычислялась методом динамического программирования (см.,например, [7]). Ниже приведено краткое описание примененного алгоритма.

Пронумеруем узловые точки матрицы $\{a_{ik}\}$ (т.е. точки A, $B, \ldots T$) по следующему правилу. Точка,лежащая левее и выше элемента a_{ik} , имеет номер (i, k). На рис. 2 точка K имеет номер (i, k),точка A – (I,I),B– (2,2) и т.д.Точки, лежащие на правой границе матрицы,имеют номера ($i, n+1$),на нижней границе –

($m+1$, κ), где $i = 1,\ldots,m+1$; $\kappa = 1,\ldots,n+1$ Длина пути $A_{i\kappa}$ от точки (i, κ) до точки $T(m+1, n+1)$ есть максимальное из следующих трех чисел: $A_{i,\kappa+1}$, $A_{i+1,\kappa}$ и $a_{i\kappa} + A_{i+1,\kappa+1}$. Требуется найти длину пути от точки (1,1) до ($m+1$, $n+1$), т.е. A_{11}. Граничные условия задаются в следующем виде: $A_{i,n+1} = A_{m+1,\kappa} = 0$. Вычисления ведутся последовательно, начиная с точки (m, n):

1) $A_{m,n} = max[A_{m,n+1}, A_{m+1,n}, a_{mn} + A_{m+1,n+1}] = a_{mn}$,

2) $A_{m-1,n} = max[A_{m-1,n+1}, A_{m,n}, a_{m-1,n} + A_{m,n+1}] = max(a_{mn}, a_{m-1,n}$

. . . . . .

i) $A_{m-i+1,n} = max[A_{m-i+1,n+1}, A_{m-i+2,n}, a_{m-i+1,n} + A_{m-i+2,n+1}]$

. . . . . .

m) $A_{1,n} = max[A_{2,n}, A_{1,n+1}, a_{1n} + A_{2,n+1}]$

$m+1$) $A_{m,n-1} = max[A_{m+1,n}, A_{m,n}, a_{m,n-1} + A_{m+1,n}] = max(a_{mn}, a_{m,n-1})$

. . . . . .

$m.n$) $A_{11} = max[A_{2,1}, A_{1,2}, a_{11} + A_{22}]$.

На последнем шаге вычислений мы получаем требуемую длину пути A_{11}. После нормировки по длине слова, т.е. после деления A_{11} на длину более длинного слова, получаем A -меру сходства между словами.

Вышеописанный алгоритм вычисления точной меры сходства между словами довольно громоздок и требует большого расхода машинного времени - количество операций пропорционально квадрату длины слов n^2. Для значительного сокращения времени счета применён комбинированный метод принятия решения [8], в котором сначала вычисляются приближенные меры сходства между словами. Приближенная мера сходства определялась следующим образом: совмещаются начала и концы сравниваемых слов и суммируются (с последующей нормировкой) меры сходства между сегментами, одинаково удаленными от начала слов. Это соответствует нахождению функционала вдоль прямой 2 на рис. I. Для корректировки возможных ошибок при автоматическом определении границ слова эта процедура повторяется со словами, сдвинутыми друг относительно друга на некоторое число сегментов (см.прямую 3 на рис.I). В нашем эксперименте сдвиг начала одного слова относительно другого составлял -8,-4,0, +4, +8 сегментов.Из полученных мер сходства истинной считалась максимальная.

Полностью алгоритм распознавания выглядел следующим образом. Предъявленное для распознавания слово сравнивалось приближенным методом с каждым из 168 слов-эталонов обучающей последовательности. Выбирались 16 слов-эталонов, наиболее похожих на предъявленное слово. Затем методом динамического программирования проводилось точное определение меры сходства предъявленного слова с каждым из 16 эталонных слов и наиболее похожее слово-эталон выбиралось как результат распознавания.Комбинированная схема распознавания позволяет существенно сократить время принятия решения при незначительном снижении надежности распознавания.

Статистические испытания проводились на следующем материале: один диктор (мужчина), 4 последовательности слов, из которых 3 включают по 168 слов, а 4-я - 166 слов (при вводе 4-й последовательности в ЭВМ были допущены ошибки в двух словах). Весь материал был записан на бытовом магнитофоне "Астра-4" с микрофона типа МД-47 в тихой комнате с обычной акустикой.Все 4 последовательности с магнитофона той же марки с помощью 9-разрядного аналого-цифрового преобразователя (частота квантования 20 кгц) были введены в ЭВМ БЭСМ-6, где производилось автоматическое определение границ слов, фильтрация с помощью аппарата Z - преобразования (см.,например, [9]), выделение логарифмических параметров слов и запись этих параметров на магнитную ленту БЭСМ-6. Обработка слова продолжается — в 4 раза дольше длительности слова. Затем производилось распознавание, причем в качестве обучающих (эталонных) последовательностей использовались поочередно 1-я, 2-я и 3-я последовательности, а в качестве контрольных - все остальные. Результаты эксперимента приведены в таблице 2. Общая надежность распознавания, усредненная по 1506 реализациям, составляет свыше 95%.

Таблица 2

№оп \ №кп	число ошибок				надежность, %
	1	2	3	4	
1		3	9	12	95,6
2	3		8	5	96,8
3	16	13		4	93,4

Распознавание одного слова на ЭВМ БЭСМ-6 занимает в среднем около 8 сек без учета времени работы внешних устройств (печать, магнитные ленты, магнитные барабаны) и около 10 сек с учетом этих устройств.

Результаты экспериментов показывают, что с помощью данного метода можно получить практически приемлемую надежность распознавания 150-200 слов, если имеется возможность производить подстройку системы под диктора, состоящую в однократном произнесении всех слов словаря.

Дальнейшие исследования должны быть направлены, главным образом, на сокращение времени принятия решения.

Литература

1. Г.Я.ВЫСОЦКИЙ, Б.Н.РУДНЫЙ, В.Н.ТРУНИН-ДОНСКОЙ, Г.И.ЦЕМЕЛЬ. Алгоритм опознавания 40 слов на ЦВМ БЭСМ-3М. В сб. "Работы по технической кибернетике", М., вып.2, ВЦ АН СССР, 1968, стр.3-33.
2. Б.М. КУРИЛОВ, Б.Г.ГАВРИЛКО. Членение речевого потока на фонемы. Данный сборник, стр.
3. Физический энциклопедический словарь, т.1, 1960, стр.242.
4. Н.Г. ЗАГОРУЙКО, В.М.ВЕЛИЧКО, Г.Я.ВОЛОШИН, В.Д.ГУСЕВ, В.Н.ЕЛКИНА, И.В.БАХМУТОВА, А.Г.ХАЙРЕТДИНОВА, Л.С.ЮДИНА. Эксперименты по автоматическому распознаванию речевых сигналов.-Доклад на IУ школе-семинаре по автоматическому распознаванию слуховых образов (АРСО-IУ). Канев, 1968.
5. А.А.ПИРОГОВ, Г.С.СЛУЦКЕР. К фонетической теории речи. Доклад на У1 Всесоюзной акустической конференции, М., 1968.
6. А.КОФМАН, Г.ДЕБАЗЕЙ. Сетевые методы планирования и их применение. Изд. "Прогресс", 1968.
7. Е.С.ВЕНТЦЕЛЬ. Элементы динамического программирования. Изд. "Наука", 1964.
8. Н.Г.ЗАГОРУЙКО. Комбинированный метод принятия решения. Вычислительные системы. Новосибирск, 1966, вып.22, Изд."Наука", Сиб.отд., стр.
9. Приспосабливающиеся автоматические системы. Под ред.Э.Мишкина и Л.Брауна. ИЛ., 1963.

Поступила в редакцию
6.I.1969г.