ЭМПИРИЧЕСКОЕ ПРЕДСКАЗАНИЕ И РАСПОЗНАВАНИЕ ОБРАЗОВ
(Вычислительные системы)
1976 год Выпуск 67

УДК 621.391:681.3.06

АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ РЕЧИ ПО МАСКПРИЗНАКАМ

В.Г. Лебедев

В работе [1] в качестве механизма для выбора признаков из полного спектрального описания речевого сигнала использовалась машинная модель известного в психоакустике эффекта маскировки, присущего слуховой системе человека.

Исходные акустические сигналы, в качестве которых были взяты изолированные слова, через аналого-цифровой преобразователь с частотой квантования 20 кгц вводились в ЭВМ БЭСМ-6 и записывались на магнитную ленту. Анализируемое слово разбивалось на сегменты длительностью 16 мсек, и для каждого сегмента вычислялись значения модулей спектра по Фурье, а затем с помощью алгоритма, моделирующего эффект маскировки, выделялись маскпризнаки этого сегмента речи. Каждый маскпризнак записывался парой чисел: $F_i^{(j)}$, $E_i^{(j)}$, где $F_i^{(j)}$ – значение частоты i-го маскпризнака j-го сегмента, $E_i^{(j)}$ – значение амплитуды i-го маскпризнака j-го сегмента.

Таким образом, каждая акустическая реализация слова, хранящаяся на магнитной ленте ЭВМ БЭСМ-6, перерабатывалась в последовательность из L массивов, соответствующих числу сегментов длительности 16 мсек в обрабатываемом слове. Каждый массив с номером j содержит значения частот (F) и амплитуд (E) маскпризнаков j-го сегмента слова.

Настоящая работа посвящена описанию экспериментов по автоматическому распознаванию речевых сигналов, которые были проведены с целью исследования эффективности полученной системы маскпризнаков.

Во-первых, была введена мера сходства между двумя сегментами речевого сигнала. При этом исследовались два случая:

1. В первом случае использовалась только информация о значениях частот (F) маскпризнаков сегментов.

Пусть n_1 – число маскпризнаков в k-м сегменте, а n_2 – число маскпризнаков в m-м сегменте. При $n_2 \geq n_1$ расстояние ρ_{km} между сегментами k и m находится по формуле:

$$\rho_{km} = \sum_j [\ln F_i^{(k)} - \ln F_j^{(m)}]^2, \qquad (1)$$

где $F_i^{(k)}$ – значение частоты i-го маскпризнака k-го сегмента, i – номер ближайшего по частоте маскпризнака k-го сегмента к j-му признаку m-го сегмента.

Мера сходства сегментов k и m определялась следующим образом:

$$a_{km} = \frac{\alpha^2}{\alpha^2 + \rho_{km}^2}, \qquad (2)$$

где $\alpha^2 = \text{const}$. (Экспериментально выбрано значение $\alpha^2 = 1000$.)

2. Во втором случае наряду с информацией о значениях частот маскпризнаков использовалась информация о значениях их амплитуд.

Расстояние ρ_{km} в этом случае представляло собой сумму

$$\rho_{km} = \rho_{km}^{(1)} + \rho_{km}^{(2)}, \qquad (3)$$

где $\rho_{km}^{(1)}$ находится по формуле (1), а ρ_{km}^{2} вычисляется аналогично:

$$\rho_{km}^{(2)} = \sum_j [\ln E_i^{(k)} - \ln E_j^{(m)}]^2. \qquad (4)$$

Мера сходства определялась так же, как и в первом случае, по формуле (2).

Переход от меры сходства a для сегментов к мере сходства A для слов осуществляется по методу, описанному в [2]. Максимальная мера сходства между двумя словами (A_{max}) определяется с помощью метода динамического программирования. Мера сходства A между двумя словами находится путем нормировки по длине слова, т.е. делением A_{max} на длину более длинного слова:

$$A = \frac{A_{max}}{L_{max}}, \qquad (5)$$

$$L_{max} = \max (L_1, L_2),$$

где L_1 – длина I-го слова, L_2 – длина 2-го слова.

Эксперименты проводились на последовательности из 63-х изолированных слов, произнесенных двумя дикторами (мужчинами).

Словарь из 63-х слов "ПОМПА" – "ХИХИКАНЬЕ"

I) ПОМПА	22) ЗЫБЬ	43) ГНЕЗДОВЬЕ
2) СВИСТ	23) ЧЕШУЯ	44) ХВОЩ
3) ЛОПАСТЬ	24) ШАБАШ	45) ХИЛОСТЬ
4) ЖЕМЧУЖИНА	25) ШИХТА	46) ПИКЕТ
5) ФЕЛЬДФЕБЕЛЬ	26) ЧЕЧЕВИЦА	47) ХИМИК
6) ДИФФЕРЕНЦИАЦИЯ	27) РЕЗЕРВ	48) ПИТОМНИК
7) СМЕХ	28) БЕГЕМОТИКИ	49) ГРЯЗИЩА
8) СПЕЦИФИКА	29) ШКИВ	50) ПИГМЕЙ
9) УСТРОЙСТВО	30) ФИЗИК	5I) СДОБА
I0) ВЗРЫВ	3I) ШАВЕЛЬ	52) ПИЧУГА
II) ГЕГЕМОН	32) ЩЕБЕНЬ	53) ПЕВИЦА
I2) ГРУЗДЬ	33) ХОДЬБА	54) ВОЗДУХ
I3) ГИМН	34) СДВИГ	55) ДВЕРЬ
I4) КИЗИЛ	35) ЩУПЛЫЙ	56) ГНЕВ
I5) МЕДЯНКА	36) БУДДИЗМ	57) ЩЕЛИН
I6) ЛЕМЕХ	37) ЖУЖЕЛИЦА	58) БИРЮЗА
I7) ЛЫЖИ	38) БУБЕН	59) ШКВАЛ
I8) ДВУХВОСТКА	39) ДУГИ	60) ЩУПАЛЬЦА
I9) СКИПЕТР	40) БУРНУС	6I) КЕФАЛЬ
20) ХИЩНИК	4I) БЕЗДЕНЕЖЬЕ	62) ГИСТОХИМИя
2I) ЖИЛЫ	42) КИМОГРАФ	63) ХИХИКАНЬЕ

В качестве эталонов использовалась последовательность, произнесенная одним диктором, а в качестве контрольной – та же последовательность, произнесенная другим диктором. Для каждого слова, предъявленного для распознавания, выдавалось I0 слов претендентов из эталонной последовательности, имеющих наиболь-

шие меры сходства (в смысле (5)) с предъявленным словом. Слово считалось распознанным верно, если соответствующее ему слово эталонной последовательности имело наибольшую меру сходства, т.е. стояло на первом месте.

Были получены следующие результаты:

I. При использовании информации только о значениях частот маскпризнаков надежность распознавания составила 60,3%.

2. При совместном использовании информации о значениях частот и амплитуд маскпризнаков надежность распознавания повысилась до 80,9%.

Для дальнейшего повышения надежности распознавания была привлечена информация о словесном ударении [3]. Для каждого слова как эталонной, так и контрольной последовательности указывался вручную сегмент начала ударной гласной. Таким образом, слово делилось точкой ударения на две части. При распознавании сравнение двух слов велось раздельно по начальным (до точки ударения) и конечным (после точки ударения) участкам.

Мера сходства в этом случае вычисляется следующим образом:

$$A = \frac{A_1 L_{1\,max} + A_2 L_{2\,max}}{L_{1\,max} + L_{2\,max}},$$

где A_1 – мера сходства начальных участков 2-х слов,

A_2 – мера сходства конечных участков 2-х слов,

$L_{1\,max} = \max (L_1^{(1)}, L_1^{(2)})$, $L_1^{(1)}, L_2^{(2)}$ – соответственно длины начальных участков I-го и 2-го слов;

$L_{2\,max} = \max (L_2^{(1)}, L_2^{(2)})$, $L_2^{(1)}, L_2^{(2)}$ – длины конечных участков I-го и 2-го слов;

A_1 и A_2 вычислялись с использованием информации о значениях частот и амплитуд маскпризнаков.

Надежность распознавания 63-х слов составила 85,7%.

Для сравнения полученного результата с другими отметим, что при распознавании того же словаря для тех же 2-х дикторов, но в случае использования в качестве признаков значения энергий на выходах пяти октавных фильтров, была получена надежность распознавания около 50%.

В настоящее время проводится работа по оптимизации маскпризнаков с целью получения более высокой надежности распознавания.

Литература

I. ЗАГОРУЙКО Н.Г., ЛЕБЕДЕВ В.Г. Эффект маскировки и автоматический анализ речевых сигналов. -В кн.: Вычислительные системы. Вып. 61. Новосибирск, 1975, с. 103-111.

2. ВЕЛИЧКО В.М., ЗАГОРУЙКО Н.Г. Автоматическое распознавание ограниченного набора устных команд. -В кн.: Вычислительные системы. Вып. 36. Новосибирск, 1969, с. 101-110.

3. ХАЙРЕТДИНОВА А.Г. Автоматическое выделение словесного ударения с использованием спектрально-временного описания гласных. -Труды УП Всесоюзной школы семинара "Автоматическое распознавание слуховых образов". Алма-Ата, 1973, с. 45-48.

Поступила в ред.-изд.отд.
15 января 1976 года