ЭМПИРИЧЕСКОЕ ПРЕДСКАЗАНИЕ И РАСПОЗНАВАНИЕ ОБРАЗОВ
(Вычислительные системы)
1980 год Выпуск 83

УДК 519.226:534.78:681.327

СРАВНЕНИЕ СИСТЕМ ПРИЗНАКОВ, ОСНОВАННЫХ НА ЧАСТНОЙ АВТОКОРРЕЛЯЦИОННОЙ ФУНКЦИИ, ПРИ РЕШЕНИИ ЗАДАЧИ РАСПОЗНАВАНИЯ ИЗОЛИРОВАННЫХ СЛОВ

А.В. Кельманов

Исследованию описания речевых сигналов моделью авторегрессии, авторегрессии со стационарной разностью первого порядка и авторегрессии с адаптивной разностью первого порядка при решении задач анализа и синтеза речевых сигналов посвящено достаточно много работ [1-4, 8,9,13-19]. При решении этих задач используются как параметры моделей авторегрессии, так и частные автокорреляции соответствующих моделей.

Иначе обстоит дело с решением задачи распознавания речи. В работах [5,7,21-27], посвященных этой проблеме, речевые сигналы описываются только при помощи моделей авторегрессии, хотя взятие разности первого порядка в сочетании с другими системами описания ранее уже применялось [8] для анализа речи.

Отметим, что в упомянутых работах, как правило, используются непосредственно параметры авторегрессии и основное внимание уделяется исследованию меры сходства (различия) между сегментами речевого сигнала.

Частные автокорреляции, в отличие от параметров авторегрессии, в работах по распознаванию почти не используются за исключением работ [10,25]. Так, в [25] для описания применялась модель авторегрессии, а в [10] - модель авторегрессии с адаптивной разностью первого порядка. Интересно сравнить указанные способы описания.

Хотя в обеих работах получены высокие по надежности распознавания результаты, окончательный вывод в пользу какой-либо системы описания сделать нельзя по следующим причинам:

I) в [10] применялся алгоритм динамического программирования, предложенный в [6], а в [25] - его модификация;

2) использовались различные меры сходства (различия) между сегментами речевого сигнала;

3) исследования проводились на различном речевом материале (словарь, язык, дикторы);

4) различались условия ввода сигнала в ЭВМ (число разрядов аналого-цифрового преобразователя, частота квантования, уровень шума, микрофон);

5) использовались различные способы анализа по методу разбиения с перекрытием (длительности сегмента и сдвига).

Поскольку частная автокорреляционная функция модели авторегрессии со стационарной разностью первого порядка ранее не использовалась в качестве описания при построении систем распознавания, представляется целесообразным исследовать такой способ описания в сравнении с двумя ранее упомянутыми. Так как при распознавании речи признаки типа тон/шум (голос/не голос) и частота основного тона во многих случаях оказываются полезными, то также целесообразно исследовать те системы описания, в которые, кроме частных автокорреляций 3-х указанных моделей, включены (по отдельности) эти признаки. Наконец, представляет интерес вопрос об инвариантности систем описания к диктору.

Итак, цель данной работы заключается в экспериментальном сравнении при одинаковых условиях 9-ти систем описания (признаков) речевого сигнала. Критериями для сравнения будут: надежность распознавания, порядок моделей авторегрессии и число признаков, а также затраты времени на оценивание параметров (признаков) и принятие решения.

I. Системы признаков

Пусть речевой сигнал $s(t)$ проквантован с частотой $F_s = 1/T$ (T - интервал квантования) так, что $s_n = s(nT)$. Обозначим через K объем словаря системы распознавания изолированных слов. Каждое k-е слово словаря разобьем на L_k сегментов $\{s_n\}_k^{(l)}$, $n = \overline{1,N}$, $l = \overline{1,L_k}$, длительностью T_a со сдвигом от сегмента к сегменту на время $\Delta T < T_a$. Каждый l-й сегмент $\{s_n\}^{(l)}$ будем описывать моделью вида:

$$A^{(l)}(z^{-1})w_n^{(l)} = \alpha_n^{(l)}, \qquad (1)$$

$$w_n^{(1)} = \nabla_a^{(1)} s_n^{(1)} , \qquad (2)$$

$$A^{(1)}(z^{-1}) = 1 - \sum_{i=1}^{p} a_i^{(1)} z^{-i} , \qquad (3)$$

$$\nabla_a^{(1)} = 1 - a^{(1)} z^{-1}, \qquad (4)$$

где z^{-i} – оператор сдвига назад такой, что $z^{-i} w_n = w_{n-i}$; α_n – белый шум с дисперсией $[\sigma_\alpha^{(1)}]^2$; p – порядок модели; n – текущее равномерно дискретное время; $\{a_i\}^{(1)}$ – параметры авторегрессии.

Исследуемые модели можно получить из (I)–(4), изменяя в (4) значение коэффициента $a^{(1)}$. Если положить $a^{(1)}=0$, то уравнения (I)–(4) будут описывать модель авторегрессии p-го порядка (в дальнейшем именуемую как модель I). При $a^{(1)} = 1$ получим модель авторегрессии p-го порядка со стационарной разностью первого порядка (в дальнейшем – модель 2), а при $a^{(1)} = r_1^{(1)}(s)$, где $r_1^{(1)}(s)$ – автокорреляция сигнала $\{s_n\}^{(1)}$ для первой задержки, уравнения (I)–(4) будут описывать модель авторегрессии p-го порядка с адаптивной разностью первого порядка (модель 3).

Оценки параметров оператора $A^{(1)}(z^{-1})$ находятся из решения системы нормальных уравнений Юла–Уокера рекуррентным методом Дурбина [I2]. Соотношения имеют следующий вид:

$$\hat{a}_{j+1,i}^{(1)} = \hat{a}_{ji}^{(1)} - \hat{a}_{j+1,j+1}^{(1)} \times \hat{a}_{j,j-i+1}^{(1)}, \qquad i = 1,2,\ldots, j,$$

$$\hat{a}_{j+1\ j+1}^{(1)} = \frac{\hat{r}_{j+1}^{(1)} - \sum_{i=1}^{j} \hat{a}_{ji}^{(1)} \hat{r}_{j-i+1}^{(1)}}{1 - \sum_{i=1}^{j} \hat{a}_{ji}^{(1)} \hat{r}_i^{(1)}} , \qquad (5)$$

где первый индекс соответствует очередному шагу при рекуррентном оценивании, второй – номер параметра оператора $A_{j+1}^{(1)}(z^{-1})$, а $\hat{r}_i^{(1)}$ – оценка автокорреляционной функции сигнала $\{w_n\}^{(1)}$, $n = \overline{1, N-1}$:

$$\hat{r}_i^{(1)} = \frac{\sum_{n=1}^{N-1-i} w_n^{(1)} w_{n+i}^{(1)}}{\sum_{n=1}^{N-1} [w_n^{(1)}]^2} .$$

Величина a_{ii}, рассматриваемая как функция задержки i, называется частной автокорреляционной функцией.

Для первых 3-х систем описания каждый 1-й сегмент слова будем характеризовать вектором $\hat{\vec{a}}_{pp}^{(1)} = (\hat{a}_{11}^{(1)}, \hat{a}_{22}^{(1)}, \ldots, \hat{a}_{pp}^{(1)})$, компонентами которого являются частные автокорреляции рассматриваемых моделей, а k-е слово словаря - последовательностью векторов $\{\hat{\vec{a}}_{pp}\}_k^{(1)}$, $1 = \overline{1, L_k}$.

Четвертая, пятая и шестая системы описания получаются из первых 3-х добавлением к ним признака тон/шум IV, т.е. каждый 1-й сегмент описывается вектором $\hat{\vec{b}}_{pp}^{(1)} = (IV^{(1)}, \hat{a}_{11}^{(1)}, \hat{a}_{22}^{(1)}, \ldots, \hat{a}_{pp}^{(1)})$. Признак $IV^{(1)}$ может принимать два значения: I - для голосовых сегментов и 0 - для неголосовых. Слово описывается последовательностью векторов $\{\hat{\vec{b}}_{pp}\}^{(1)}$.

Последние 3 системы признаков образуются аналогично предыдущим 3-м, только вместо признака тон/шум в них задействован признак частота основного тона $F_{от}^{(1)}$. Каждый 1-й сегмент слова в данном случае описывается вектором $\hat{\vec{c}}_{pp}^{(1)} = (F_0^{(1)}, \hat{a}_{11}^{(1)}, \hat{a}_{22}^{(1)}, \ldots, \hat{a}_{pp}^{(1)})$, первая компонента которого задается по правилу: $F_0^{(1)}$ равно 0 для неголосовых сегментов, а для голосовых $F_0^{(1)}$ равно частоте основного тона, нормированной к 1 по максимальному значению на всей длине слова, т.е.

$$F_0^{(1)} = \begin{cases} \dfrac{F_{от}^{(1)}}{F_{max}}, & F_{max} = \max\limits_{1} F_{от}^{(1)}, \text{ при } IV = 1, \\ 0, & IV = 0. \end{cases} \qquad (6)$$

Слово описывается последовательностью векторов $\{\hat{\vec{c}}_{pp}\}^{(1)}$.

Нормировка (6) частоты основного тона к диапазону от 0 до I и выбор значений признака $IV^{(1)}$ равным 0 или I сделаны для того, чтобы априори обеспечить "равномерный" вклад всех компонент векторов $\hat{\vec{c}}_{pp}^{(1)}$ и $\hat{\vec{b}}_{pp}^{(1)}$ в меру сходства, поскольку известно [12], что для стационарных моделей величина $|a_{ii}| < 1$, $i = \overline{1, p}$.

Вычисление значений признаков тон/шум и частоты основного тона производится специально разработанными алгоритмами. В настоящей работе потребуется только трудоемкость указанных алгоритмов.

2. Мера сходства и алгоритмы распознавания

2.I. А л г о р и т м I (см. [6]). По аналогии с [6] введем меру сходства между сегментами i и j :

$$d_{ij} = \frac{\alpha^2}{\alpha^2 + \rho_{ij}^2}, \qquad (7)$$

где $\rho_{ij}^2 = \sum_{q=1}^{p} [\hat{a}_{qq}^{(i)} - \hat{a}_{qq}^{(j)}]^2$ - квадрат евклидова расстояния между векторами $\hat{\vec{a}}_{pp}^{(i)}$ и $\hat{\vec{a}}_{pp}^{(j)}$; α - экспериментально подбираемая константа.

Мера сходства между сегментами, в описание которых входят признаки тон/шум и частота основного тона, определяется аналогично.

Меру сходства между двумя словами, содержащими I и J сегментов соответственно, будем вычислять последовательно вдоль ломаной максимального сходства i(j), начиная с концов слов (т.е. сегментов, номера которых I и J), при помощи метода динамического программирования из следующих соотношений:

$$D(i,j) = \max \begin{cases} D(i,j+1), \\ D(i+1,j+1) + d(i,j), \\ D(i+1,j), \end{cases} \qquad (8)$$

$$i = I, I-1, \ldots, 1; \quad j = J, J-1, \ldots, 1,$$

при нулевых начальных условиях:

$$D(I,J+1) = D(I+1,J) = D(I+1,J+1) = 0, \qquad (9)$$

граничных условиях:

$$\forall i,j (i < I \;\&\; j > J \Rightarrow D(i,j) = 0),$$
$$\forall i,j (i > I \;\&\; j < J \Rightarrow D(i,j) = 0) \qquad (10)$$

и условии монотонности

$$\forall i,j \; (i(j) - i(j-1) \geq 0). \qquad (11)$$

На последнем шаге, т.е. при $i=1$ и $j=1$, получаем меру сходства между двумя словами $D(1,1)$. Окончательная мера сходства определяется после нормировки по длине более длинного слова: $D = D(1,1)/L_{max}$, $L_{max} = \max(I,J)$.

Каждое m-е контрольное слово сравнивается со всеми эталонами так, что в результате формируется последовательность мер сходства $D_1^{(m)},\ldots,D_K^{(m)}$. Решение принимается по правилу:

$$h = \underset{k = \overline{1,K}}{\operatorname{argmax}}\, D_k^{(m)}. \qquad (12)$$

2.2. А л г о р и т м II. В этом алгоритме, как и в предыдущем, используется мера сходства между сегментами (7), мера сходства между словами $D(1,1)$ вычисляется из соотношений (8) с соблюдением условия монотонности (11) и при тех же начальных условиях (9). Но в отличие от алгоритма I введены ограничения на область допустимых значений ломаной максимального сходства. Если в алгоритме I ломаная была определена на всем множестве пар (i,j), $i = \overline{1,I}$, $j = \overline{1,J}$, то в алгоритме II эта область ограничена коридором:

$$|i - j\theta| \le \phi, \quad \theta = I/J, \qquad (13)$$

где ϕ – целая положительная константа, $0 < \phi < \min(I,J)$.

В соответствии с (13) граничные условия (10) для данного алгоритма принимают вид:

$$\forall i,j(|i - j\theta| > \phi \Rightarrow D(i,j) = 0),$$
$$\forall j\ ((I - \phi)/\theta \le j \le J-1 \Rightarrow D(I+1,j) = 0),$$
$$\forall i\ (I - \phi \le i \le I-1 \Rightarrow D(i,J+1) = 0).$$

Как и в предыдущем алгоритме, мера сходства между двумя словами $D(1,1)$ нормируется по длине более длинного слова, а решение принимается по правилу (12).

3. Трудоемкости выделения признаков и принятия решения

Оценим трудоемкости выделения признаков для 1,2 и 3-й систем описания (группа I) τ_{01}, τ_{02}, τ_{03}. Системы описания этой группы включают в себя только частные автокорреляции. Поэтому трудоемкости выделения признаков совпадают с трудоемкостями оценивания частных автокорреляций рассматриваемых моделей. Для модели авторегрессии число операций, необходимых для оценивания, складывается из числа операций на вычисление автокорреляций для p задержек, равного примерно $N(p+1)$ операциям (сложения и умножения) плюс число операций на решение системы уравнений Юла–Уокера, равного примерно p^2 операциям, т.е. $\tau_{01} = p^2 + N(p+1)$. Для модели авторегрес-

сии со стационарной разностью первого порядка необходимо дополнительно N операций на взятие разностей, поэтому $\tau_{02} = p^2+N(p+2)$. Наконец, для модели авторегрессии с адаптивной разностью по сравнению с предыдущей моделью требуется еще N операций на вычисление коэффициента $a^{(1)}$. Следовательно, $\tau_{03} = p^2 + N(p+3)$.

Трудоемкость выделения признаков 4,5 и 6-й систем описания (группа II) τ_{04}, τ_{05} и τ_{06} складывается из трудоемкостей оценивания частных автокорреляций τ_{01}, τ_{02} и τ_{03} соответственно плюс трудоемкость классификации тон/шум τ_{V}. Аналогично вычисляются трудоемкости выделения признаков для 7,8 и 9-й систем описания (группа III) τ_{07}, τ_{08} и τ_{09}, только вместо трудоемкости классификации тон/шум к трудоемкостям оценивания частных автокорреляций прибавляется трудоемкость оценивания частоты основного тона τ_{0T}.

Распознавание одного слова, содержащего L сегментов из словаря, состоящего из K слов, для алгоритма I требует выполнения $CKL\bar{L}p$ операций, где C = const, $\bar{L}$ – средняя длина эталонного слова, а p – число признаков. Трудоемкость алгоритма II равна $2CKL\phi p$ операциям.

Полное время на принятие решения состоит из времени на обработку (выделение признаков) слова и времени на распознавание. В табл.I приведены трудоемкости принятия решения для всех систем описания по алгоритму I (для алгоритма II символ $\bar{L}$ следует заменить на 2ϕ).

Таблица I

Трудоемкость принятия решения для различных систем описания

Система описания i	Группа	Трудоемкость принятия решения τ_i
1	I	$CKL\bar{L}p + L[p^2 + N(p+1)]$
2		$CKL\bar{L}p + L[p^2 + N(p+2)]$
3		$CKL\bar{L}p + L[p^2 + N(p+3)]$
4	II	$CKL\bar{L}(p+1) + L[p^2 + N(p+1) + \tau_V]$
5		$CKL\bar{L}(p+1) + L[p^2 + N(p+2) + \tau_V]$
6		$CKL\bar{L}(p+1) + L[p^2 + N(p+3) + \tau_V]$
7	III	$CKL\bar{L}(p+1) + L[p^2 + N(p+1) + \tau_{0T}]$
8		$CKL\bar{L}(p+1) + L[p^2 + N(p+2) + \tau_{0T}]$
9		$CKL\bar{L}(p+1) + L[p^2 + N(p+3) + \tau_{0T}]$

4. Эксперименты и их результаты

4.I. Р е ч е в о й м а т е р и а л. При проведении эксперимента использовались 2 словаря. Первый состоял из I00 русских слов (см.[I0, табл.I]), второй - из 45 слов (табл. 2).

Т а б л и ц а 2

Словарь №2

I. Один	I6. Отключение	3I. Запятая
2. Два	I7. Схема	32. Предельные
3. Три	I8. Защита	33. Справка
4. Четыре	I9. Автоматика	34. План
5. Пять	20. Авария	35. Станция
6. Шесть	2I. Архив	36. Подстанция
7. Семь	22. Заявки	37. Линия
8. Восемь	23. Оборудование	38. Шины
9. Девять	24. Сдача	39. Выключатель
I0. Ноль	25. Переток	40. Обходной
II. Нагрузка	26. Ремонт	4I. Смена
I2. Пэ	27. Задание	42. Час
I3. Ку	28. Рапорт	43. Минута
I4. Нарушение	29. Точка	44. Нормально
I5. Отметки	30. Тире	45. Отклонение

По словарю №I одним диктором-мужчиной (диктор I) были сформированы 2 акустические последовательности с интервалом произнесения полгода, т.е. всего 200 реализаций по 2 на слово. По словарю №2 каждым из дикторов-мужчин (дикторы I,2,3) были также сформированы 2 акустические последовательности, т.е. всего 45 х 2 х 3 = 270 реализаций, по 6 на слово.

4.2. У с л о в и я э к с п е р и м е н т о в. Все эксперименты проводились на ЭВМ "Минск-32". Ввод слов в ЭВМ осуществлялся непосредственно в машинном зале (уровень шума ~60 дб) через микрофон типа МД-59 и семиразрядный аналого-цифровой преобразователь с частотой квантования 20 кгц. После ввода производилось автоматическое определение границ слов и запись введенных слов на магнитную ленту.

4.3. У с л о в и я а н а л и з а. Акустические реализации слов обрабатывались при помощи процедур, описанных выше. Анализ слов проводился по методу разбиения с перекрытием при длительно -

Таблица 3

Надежность распознавания (%) по частным автокорреляциям; диктор I, словарь I00 слов; ошибка ±0,5; группа I

Порядок модели	Система описания		
	I	2	3
	Модель		
	I	2	3
2	32	48	34
4	6I	69	66
6	78	80	79
8	84	89	88
I0	9I	95	93
I2	96	96	97
I4	97	98	99
I6	98	99	I00
I8	98	99	I00
20	98	99	I00
22	99	I00	I00
24	99	I00	I00
26	99	I00	I00
28	99	I00	I00
30	99	I00	I00

сти сегмента T_a = 25,6 мсек и сдвиге от сегмента к сегменту на время ΔT = I6 мсек. Каждый сегмент взвешивался окном Хэмминга и описывался одной из 9 систем признаков в зависимости от типа эксперимента. Порядок моделей при обработке слов был равен 30.

4.4. Результаты распознавания. Алгоритмы распознавания, описанные выше, были опробованы для всех систем описания при изменении порядка моделей p от 2 до 30. Значение констант α и ϕ было выбрано равным I и 5 соответственно. Ниже приведены результаты экспериментов.

Эксперимент I. В качестве эталонов по очереди служили первые из двух акустических последовательностей, произнесенных каждым из трех дикторов. На контроль предъявлялись вторые последовательности того диктора, первая последовательность которого использовалась в качестве эталонной.

I. В табл. 3 приведены результаты распознавания по частным автокорреляциям моделей I,2 и 3 по алгоритму I для словаря №I и диктора I.

2. В табл.4 приведены результаты распознавания для всех девяти систем описания по алгоритму II для словаря № 2, усредненные по данным, полученным для каждого из 3-х дикторов.

Эксперимент 2. Тестирование проводилось на словаре №2. Эталонами служили 45 слов в произнесении диктора I (I последовательность). На контроль были предъявлены: вторая последовательность диктора I и первая и вторая последовательности каждого из дикторов 2 и 3, т.е. всего 225 слов. Результаты распознавания для всех систем описания сведены в табл. 5.

Таблица 4

Результаты распознавания (с точностью до ±1%) с подстройкой под диктора, усредненные по трем дикторам (%); словарь 45 слов

Порядок модели	Группа I			Группа II			Группа III		
	Система описания			Система описания			Система описания		
	1	2	3	4	5	6	7	8	9
	Модель			Модель			Модель		
	1	2	3	1	2	3	1	2	3
4	97	94	94	97	98	97	97	97	97
6	97	98	97	97	98	97	98	98	97
8	98	98	97	98	98	98	98	98	98
10	99	99	99	98	99	99	98	99	99
12	99	99	99	99	99	99	99	99	99
14	99	99	99	99	99	99	99	99	99
16	99	100	100	99	99	99	99	99	99
18	99	100	100	99	99	99	99	99	99
20	99	100	100	99	99	99	99	99	99
22	99	100	100	99	99	99	99	99	99
24	99	100	100	99	99	99	99	99	99
26	99	100	100	99	99	99	99	100	100
28	100	100	100	99	99	99	100	100	100
30	100	100	100	99	99	99	100	100	100

5. Обсуждение результатов

Переходя к сравнению систем описания, напомним, что внутри каждой из групп системы описания отличаются лишь типом модели авторегрессии (модели 1,2,3). Поэтому, сравнивая модели внутри групп, можно сделать выводы о предпочтительности той или иной модели, а сравнивая группы помодельно (т.е. по одинаковому типу моделей), - выводы о предпочтительности той или иной группы.

Во введении были выписаны критерии, по которым будет проводиться сравнение. Выбор этих критериев не случаен, поскольку любая система распознавания должна удовлетворять двум основным тре-

Таблица 5

Результаты распознавания (с точностью до ± 0,2%) без подстройки под диктора (%), словарь 45 слов

Порядок модели	Группа I			Группа II			Группа III		
	Система описания			Система описания			Система описания		
	1	2	3	4	5	6	7	8	9
	Модель			Модель			Модель		
	I	2	3	I	2	3	I	2	3
4	80,9	81,8	78,2	84,9	84,0	82,7	84,0	83,I	82,2
6	84,4	85,8	82,7	88,0	88,4	87,6	88,4	88,4	86,7
8	87,6	90,2	88,4	90,7	92,9	92,0	92,9	93,3	9I,6
I0	88,9	93,3	9I,6	92,9	94,2	94,2	94,2	95,I	94,7
I2	90,2	93,8	92,9	94,7	95,I	95,I	95,I	96,0	95,6
I4	92,4	94,2	93,8	94,7	95,I	95,I	95,I	96,0	95,6
I6	93,8	95,6	94,7	95,6	96,0	95,6	96,0	96,9	96,0
I8	94,7	96,4	95,I	95,6	96,0	95,6	96,0	96,9	96,0
20	95,I	96,4	96,0	95,6	96,0	95,6	96,0	96,9	96,0
22	95,6	96,4	96,0	95,6	96,0	95,6	96,0	96,9	96,0
24	96,0	96,4	96,0	95,6	96,0	95,6	96,0	96,9	96,0
26	96,0	96,4	96,0	95,6	96,0	95,6	96,0	96,9	96,0
28	96,0	96,4	96,0	95,6	96,0	95,6	96,0	96,9	96,0
30	96,0	96,4	96,0	95,6	96,0	95,6	96,0	96,9	96,0
Порядок предпочтения	9	7	8	6	4	5	3	I	2

бованиям: I) работать с высокой надежностью, 2) иметь минимальное время реакции. В связи с этим при анализе табличных данных сравнение следует проводить по значениям надежности распознавания при фиксированном порядке моделей, по максимальной надежности распознавания, по порядкам моделей, при которых надежности распознавания равны или близки, и по порядкам, при которых достигается максимальная надежность.

Приступая к анализу экспериментальных данных, следует отметить, что данные в табл. 3,4,5 приведены с точностью до ±I, ±0,5

и ±0,2% соответственно. Делая выводы, это необходимо учитывать. Поэтому в качестве основной таблицы, по которой будут сделаны выводы, в дальнейшем принята табл. 5. Кроме того, отметим общую характерную для всех таблиц черту, которая облегчит проведение сравнения: зависимость надежности распознавания от порядка моделей носит монотонный характер.

5.I. С р а в н е н и е с и с т е м о п и с а н и я п о н а д е ж н о с т и.

5.I.I. Сравнение моделей внутри групп.

Г р у п п а I. Сравнивая модели I,2 и 3, нетрудно заметить (см. табл. 5), что при каждом фиксированном значении порядка надежность распознавания по параметрам модели 2 больше либо равна надежности распознавания по параметрам моделей I и 3, а разность по надежности относительно модели I лежит в диапазоне 0-4,4% и в диапазоне 0,4-3,6% относительно модели 3. Максимальная надежность распознавания по параметрам модели 2 (96,4%) больше либо равна (с точностью до ошибки) максимальным надежностям распознавания по параметрам моделей I и 3 (96,0%). Следовательно, наилучшей по надежности распознавания является модель 2.

Из сопоставления моделей I и 3 по надежности распознавания при фиксированном порядке вытекает, что при $p \leq 6$ лучше (на 2,2-2,7%) модель I, а при $p > 6$ лучше (на 0-2,7%) модель 3. Максимальные надежности для обеих моделей совпадают. Если при $p \leq 6$ надежность распознавания для обеих моделей так низка, что нет смысла использовать их в системах распознавания, то предпочтение в данном случае следует отдать модели 3.

Сравним модели по значениям порядка при равных или близких к равным надежностях. Из табл.5 видно, что при этих условиях для модели 2 требуется порядок меньший, чем для моделей I и 3. При этом разность в значениях порядка относительно модели I лежит в диапазоне от 0 до 4-6 и в диапазоне 0-2 относительно модели 3. Значение порядка модели 2 ($p = 18$), при котором достигается максимальная надежность, меньше аналогичных значений порядка для моделей I ($p = 22-24$) и 3 ($p = 20$). Следовательно, в смысле минимума числа параметров, необходимых для описания, наилучшей является модель 2.

Из сравнения моделей I и 3 вытекает, что при $p > 6$ для модели 3 требуется на 0-4 параметров меньше, а разность между минимальными значениями порядка, при которых надежность распознавания максимальна, равна 2-4. Поэтому модель 3 предпочтительнее модели I.

Аналогичные выводы относительно предпочтительности моделей можно сделать и по табл.4. Так, для моделей 2 и 3 наблюдается 100%-ная надежность при p = 16, в то время как для модели 1 только при p = 28. При p < 16 из двух моделей 2 и 3 лучше модель 2. Однако по табл.4 нельзя сделать столь же категоричные выводы относительно максимальной надежности и порядка моделей, как это было сделано по табл.5, поскольку данные табл.4 получены с точностью до ±1%, а надежность распознавания, отличающаяся от максимальной как раз на 1%, достигается всеми моделями при одном и том же значении порядка p = 10.

Обращаясь к табл.3, замечаем, что и в ней данные для моделей 2 и 3 лучше, чем для модели 1. Надежность распознавания 99-100% достигается при p = 20 для модели 1, при p = 16-20 для модели 2 и при p = 14-16 для модели 3. Сравнивая модели 2 и 3, видим, что при p ≤ 10 лучше модель 2, а при p > 10 лучше модель 3, т.е. данные табл.3 несколько отличаются по характеру поведения надежности распознавания от данных, приведенных в табл. 4,5, в которых при всех значениях порядка модель 2 была лучше модели 3. Однако эти незначительные отклонения носят скорее всего случайный характер, поскольку данные табл.3 получены по одной контрольной последовательности, в то время как данные табл.4,5 получены соответственно по 3 и 5 контрольным последовательностям.

На основании вышеизложенного получается, что в группе 1 и по надежности распознавания и по порядку наилучшей является модель 2, затем модель 3 и, наконец, модель 1.

Г р у п п а II. Сравнивая модели 1,2 и 3 точно так же, как это было сделано для группы 1, можно обнаружить, что между данными внутри группы II существуют такие же зависимости, как и между данными внутри группы 1, только при других значениях надежности распознавания и порядка моделей, которые без труда могут быть выписаны по табл. 4,5 и поэтому в тексте не приводятся.

Г р у п п а III. Анализируя данные группы III по табл. 4,5, приходим к выводу, что все заключения относительно предпочтительности моделей внутри групп 1 и II справедливы и для данной группы.

Таким образом, во всех группах описания имеем следующий порядок предпочтения: 2,3,1.

5.1.2. Сравнение групп по однотипным моделям. Переходя к сравнению данных из разных групп по однотипным моделям, напомним, что в отличие от группы 1, в которой размерность пространства призна-

ков и порядок моделей совпадают, в группах II и III число признаков на I больше порядка задействованных в них моделей за счет добавления к параметрам моделей группы I признака тон/шум или признака частоты основного тона.

М о д е л ь I (I,4,7 системы описания). Сопоставим группы I, II и III по модели I. Из табл.5 видно, что, начиная с p = 6, при каждом фиксированном значении порядка надежность распознавания в группе III больше либо равна надежности распознавания в группах I и II, а разность по надежности относительно группы I лежит в диапазоне 0-5,3% и относительно группы II в диапазоне 0,4-I,3%. Максимальные надежности распознавания для групп I,II и III равны 96,0%, 95,6%, 96,0% соответственно и могут рассматриваться как совпадающие с точностью до ошибки округления данных. Следовательно, наилучшей по надежности распознавания является группа III.

Из сравнения групп I и II по надежности при фиксированном порядке вытекает, что при $p < 24$ лучше (на 0-4,5%) группа II. При $p \geq 24$ надежности распознавания практически (с точностью до $\pm 0,2$) одинаковы. Поэтому предпочтение следует отдать группе II.

Сравним группы по значениям порядка при равных или близких к равным надежностям. Из табл.5 видно, что при этих условиях для группы III требуется порядок меньший, чем для групп I и II. При этом разность в значениях порядка относительно группы I лежит в диапазоне от 0 до 8 и в диапазоне 0-4 относительно группы II. Значение порядка в группе III (p = I6), при котором достигается максимальная надежность, меньше, чем в группе I (p = 22-24) и равно той же величине в группе II (p = I6). Поэтому в смысле минимума порядка наилучшей является группа III.

Из сопоставления групп I и II находим, что для группы II требуется порядок на 0-6 меньший, чем для группы I, а разность между минимальными значениями порядка, при которых надежность распознавания максимальная, равна 6-8. Поэтому группа II предпочтительнее группы I.

Сравнивая группы I,II,III по модели I с помощью табл.4, нетрудно заметить, что все 3 группы практически равноценны.

Следовательно, по надежности распознавания и порядку наилучшей является группа III, затем группа II и, наконец, группа I.

М о д е л ь 2 (2,6,8 системы описания). Сравнивая группы I, II,III так же, как это было сделано для модели I, можно обнаружить, что для модели 2 справедливы такие же зависимости, как и для мо-

дели I,с той лишь разницей,что проявляются эти зависимости при других значениях порядка и надежности распознавания. Эти значения легко находятся по табл.4,5.

М о д е л ь 3 (3,7,9 системы описания). Анализируя данные для модели 3, приходим к выводу, что все заключения о предпочтительности групп по моделям I и 2 справедливы и для данной модели.

Таким образом, по каждой из моделей наиболее предпочтительной является группа Ш, а наименее предпочтительной - группа I.

Остается сравнить все системы описания между собой. Для этого по тем же критериям, что и ранее, сопоставим следующие системы описания: наихудшую в группе Ш (7) с наилучшей в группе П (5) и наихудшую в группе П (4) с наилучшей в группе I (2). Из этого сопоставления вытекает, что все системы описания группы Ш лучше (или, по крайней мере, не хуже) систем описания группы П, а все системы признаков из группы П лучше (или,по крайней мере,не хуже) каждой системы признаков из группы I. Таким образом, все системы описания упорядочены по предпочтению (в последней строке табл.5 приведены номера систем признаков в порядке ухудшения).

5.2. С р а в н е н и е с и с т е м о п и с а н и я п о т р у д н о с т и проведем так же, как и сравнение по надежности распознавания и порядку, сопоставляя трудоемкости систем описания внутри групп и трудоемкости для однотипных моделей между группами. Обращаясь к табл.I,прежде всего отметим,что при фиксированном порядке трудоемкость принятия решения возрастает с ростом порядкового номера системы описания. Однако результаты проведенных экспериментов показывают, что по мере возрастания порядкового номера системы описания порядок модели имеет тенденцию уменьшаться. Уменьшение порядка, а следовательно, и размерности пространства признаков, как видно из табл.I, приводит к сокращению трудоемкости принятия решения. Поэтому принятие решения по более сложной системе описания может проходить за меньшее время. В связи с этим представляется целесообразным выяснить, при какой разности в значениях порядка моделей для различных систем признаков возможно сокращение трудоемкости и происходит ли сокращение трудоемкости на практике.

5.2.I. Сравнение моделей внутри групп.

Г р у п п а I. Сравним модели I и 2. Из табл.I находим: $\tau_1 - \tau_2 = CKL(p_1 - p_2) + L[(p_1 - p_2) \times (p_1 + p_2) + N(p_1 - p_2 - 1)$. Отсюда получаем, что $\tau_1 > \tau_2$ тогда и только тогда, когда $p_1 - p_2 > 0$.

Это означает, что для сокращения трудоемкости принятия решения по модели 2 по сравнению с моделью I необходимо и достаточно пони - зить порядок модели 2 хотя бы на I. Поскольку эксперименталь - ные результаты показывают возможность понижения порядка на 4-6 (см. табл. 5), принятие решения по параметрам модели 2 происходит существенно раньше, чем по параметрам модели I.

Далее, сравнивая модели I и 3 по табл.I, находим:

$$\tau_1 - \tau_3 = CK\bar{L}\bar{L}(p_1 - p_3) + L[(p_1-p_3)(p_1+p_3) + N(p_1-p_3-2)].$$

Поэтому для того чтобы $\tau_1 > \tau_3$, необходимо, чтобы $p_1 - p_3 > 2N/(CK\bar{L} + p_1 + p_3 + N)$, и достаточно, чтобы $p_1 - p_3 \geq 2$. При $p_1 - p_3 = 1$ выполняется $\tau_1 - \tau_3 = L(CK\bar{L} + p_1 + p_3 - N)$, что на практике больше нуля. Отсюда получаем, что для сокращения трудоемко - сти принятия решения по модели 3 по сравнению с моделью I доста - точно понизить порядок модели 3 хотя бы на I. Результаты экспери-ментов свидетельствуют о возможности понижения порядка на 2-4 (см. табл.5). Следовательно, на практике модель 3 позволяет принимать решение раньше, чем модель I.

Для сравнения моделей 2 и 3 достаточно заметить, что при фик-сированном порядке последняя модель требует выполнения большего числа операций и не позволяет понизить порядок по сравнению с мо-делью 2. Поэтому на практике модель 2 "быстрее" модели 3.

Г р у п п ы II и III. Проводя сравнение моделей I,2,3 по табл.I, замечаем, что для трудоемкостей внутри групп II и III спра - ведливы те же соотношения, что и для группы I. Обращаясь к экспе-риментальным данным, можно убедиться, что для групп II и III возможно понижение порядка на I для тех же моделей, что и для группы I. По-этому для данных групп, как и для группы I, наименее трудоемкая модель 2, а наиболее трудоемкая - модель I.

Таким образом, во всех группах принятие решения по парамет-рам модели 2 на практике осуществляется быстрее, чем по парамет-рам моделей I и 3, а принятие решения по параметрам модели 3 раньше, чем по параметрам модели I.

5.2.2. Сравнение групп по однотипным моделям. Для проведения сравнения необходимо знать трудоемкости выделения признаков часто-ты основного тона и тон/шума. Цифровые методы классификации тон/шум и оценивания частоты основного тона, как правило, предполагают низкочастотное ограничение речевого сигнала и понижение часто-ты квантования. Поэтому трудоемкость этих методов зависит не толь-

ко от длительности интервала анализа, но и от частоты квантования сигнала на выходе фильтра-ограничителя и от исходной частоты квантования сигнала. В данной работе для выделения указанных признаков исходная частота квантования 20 кгц понижалась до 2 кгц (после фильтрации). При этих значениях частот трудоемкости алгоритмов, использовавшихся в экспериментах для выделения признаков частоты основного тона и тон/шума, равны примерно 12N и 10N соответственно.

Сравним группы I и II. Из I и 4 строк табл.I находим:

$$\tau_1 - \tau_4 = CKL\bar{L}(p_1 - p_4 - 1) + L[(p_1 - p_4)(p_1 + p_4) + N(p_1 - p_4 - 10)].$$

Отсюда видно, что если $p_1 - p_4 \geq 10$, то $\tau_1 > \tau_4$. Из практических соображений последнее соотношение можно усилить: если $p_1 - p_4 \geq 5$, то $\tau_1 > \tau_4$. Для групп I и III по табл.I можно сделать вывод: если $p_1 - p_7 \geq 12$, то $\tau_1 > \tau_7$, который также может быть усилен из практических соображений: если $p_1 - p_7 \geq 7$, то $\tau_1 > \tau_7$.

Сравнивая группы II и III из табл.I, можно получить соотношение

$$\tau_4 - \tau_7 = CKL\bar{L}(p_4 - p_7) + L[(p_4 - p_7)(p_4 + p_7) + N(p_4 - p_7 - 2)],$$

которое совпадает с соотношением между трудоемкостями моделей I и 3 (см. п.5.2.I). Поэтому для сокращения трудоемкости принятия решения по модели из группы III по сравнению с той же моделью из группы II достаточно понизить порядок моделей группы III хотя бы на I.

Учитывая данные табл.5, нетрудно сравнить практическое время реакции по параметрам модели из какой-либо группы относительно времени реакции по параметрам той же модели из другой группы. Характеризуя группы признаков в целом, заключаем, что время принятия решения можно сократить с помощью III группы относительно I и II групп и с помощью II группы относительно группы I. В отдельных случаях указанные соотношения не выполняются, т.е. трудоемкость повышается или остается такой же из-за недостаточного понижения порядка моделей. Однако прирост трудоемкости незначителен.

5.3. Замечания об алгоритмах распознавания.

После того как несколько лет назад в работе [6] был изложен алгоритм распознавания изолированных слов на основе метода динамического программирования, появилось достаточно много его модификаций [22,23,25-28], основанных на эвристических допущениях, которые, как показали эксперименты, вместе с сокращением трудоемко-

сти позволяют повысить надежность распознавания. Суть всех модификаций - уменьшение области допустимых значений ломаной максимального сходства путем введения так называемого коридора и изменения граничных условий (см.п.2.2). Общим свойством упомянутых модификаций является то, что коридор в них задается в виде:

$$| i-j | \leq \phi, \qquad (I4)$$

т.е. вдоль прямой $i-j$, а отличаются модификации видом граничных условий, схемой вычислений, использованием в схеме вычислений меры сходства или различия и способом нормировки.

Задание коридора в виде (I4) предполагает, что эталонное и контрольное слова одного образа имеют равные или близкие к равным длительности при отличающихся темпах произнесения. На практике довольно часто представители одного образа существенно различаются по длительности. При этом использование коридора (I4) может привести к тому, что при малых ϕ ломаная максимального сходства будет лежать за пределами коридора, что может привести к ошибке. Этот недостаток можно устранить путем увеличения ϕ, но тогда возрастает трудоемкость алгоритма распознавания.

В данной работе используется иной способ устранения указанного недостатка: коридор задается вдоль прямой $i = \theta j$, где θ вычисляется адаптивно по длинам слов (по (I3)). Таким образом, упомянутые модификации [22,23,25-28] являются частным случаем (при $\theta = I$) модификации, используемой в данной работе. На θ наложено единственное ограничение: $0{,}5 \leq \theta \leq 2$, т.е. длины сравниваемых слов не должны отличаться больше, чем в 2 раза. Что касается времени распознавания, то для алгоритма П оно меньше, чем для алгоритма I примерно во столько раз, во сколько ширина коридора 2 меньше среднего числа сегментов в слове.

Постоянная ϕ эмпирически подбирается с учетом тех соображений, что, с одной стороны, уменьшение ϕ сокращает объем вычислений, но может привести к ошибкам распознавания из-за "выхода" из коридора истинной ломаной максимального сходства, а с другой,- с увеличением ϕ уменьшается вероятность "выхода" из коридора, но при этом трудоемкость алгоритма возрастает.

5.4. О сокращении времени принятия решения. Обозначим через p^* минимальное значение порядка модели, при котором надежность распознавания достигает максимума. Пусть при распознавании k-го контрольного слова получена после -

довательность мер сходства $\{p_i^{(k)}\}$, $i = \overline{1,K}$. После упорядочения ее по убыванию получим некоторую последовательность мер сходства, в которой мера сходства для k-го контрольного слова находится на $\mu(p,k)$-м месте, $1 \le \mu(p,k) \le K$, а в случае правильного распознавания - на I месте. Для каждой модели при $0 < p \le 30$ определим функцию $K'(p)$ следующим образом:

$$K'(p) = \max_{1 \le k \le K} \mu(p,k) .$$

Экспериментальное исследование функции $K'(p)$ показало, что она является невозрастающей функцией порядка p и достигает минимума при $p \ge p^*$. В случае 100%-ной надежности минимум равен I. Используя эти свойства функции $K'(p)$, можно сократить время принятия решения за счет уменьшения числа претендентов. Для этого процедуру распознавания следует разбить на два этапа: сначала распознавать все K слов при $p < p^*$, а затем $K'(p) < K$ слов при $p = p^*$. Выбор оптимального p производится так, как это сделано в [10] для модели авторегрессии с адаптивной разностью первого порядка.

5.5. Об операциях взятия разностей. Результаты распознавания показывают, что из 3-х рассматриваемых моделей наилучшие результаты дают модели авторегрессии с разностями, а из моделей с разностями предпочтительнее модель авторегрессии со стационарной разностью первого порядка.

Для выявления причин, приведших к указанному порядку предпочтения между моделями, обратимся к работам [8,9,13-19], в которых операции взятия разностей исследовались с позиций анализа речи (выделение конфигурации речевого тракта, оценивание передаточной функций и формантных частот и т.п.). Из этих работ следует, что: I) при анализе большинства голосовых звуков речи каждая из операций взятия разностей позволяет уменьшить ошибку оценивания параметров речевого тракта [9,13-19] за счет частичного устранения влияния голосового источника [14,15], которое достигается подъемом верхних частот спектра речевого сигнала и завалом низких частот. Подъем верхних и завал низких частот являются свойствами операции взятия разности [17]. Поскольку для голосовых звуков значение коэффициента $a^{(1)}$ в (4) близко к I (см. [9]), результаты оценивания параметров речевого тракта при помощи обеих моделей практически совпадают [9]; 2) при анализе неголосовых звуков речи операция взятия разности приводит к увеличению ошибки оценивания параметров речевого тракта из-за подъема верхних частот, в

то время как операция адаптивной разности позволяет уменьшить ошибку оценивания за счет частичного устранения влияния шумового источника, которое достигается завалом верхних частот. Завал верхних частот обеспечивается тем, что для неголосовых звуков коэффициент $a^{(1)}$ в (4), как правило, отрицателен. А это приводит к замене операции взятия разности операцией суммирования.

Обращаясь к результатам данной работы, заключаем, что частичное устранение влияния голосового источника при помощи взятия разностей повышает надежность распознавания. Этот вывод еще раз подтверждается тем фактом, что основная информация в речевом сообщении передается с помощью речевого тракта, модулирующего колебания голосовых связок и шумового источника. Отличия в результатах распознавания по моделям, использующим разности, вызваны различными значениями параметров моделей, получающимися при анализе неголосовых звуков.

Хотя модель авторегрессии с адаптивной разностью позволяет точнее выделять параметры речевого тракта за счет более полного устранения влияния голосового источника, при распознавании она дает менее надежные результаты, чем модель авторегрессии со стационарной разностью. Это явление объясняется тем, что для многих голосовых и неголосовых звуков речи конфигурации речевого тракта практически совпадают. В связи с этим более полное устранение влияния источника возбуждения приводит к тому, что значения параметров модели для голосовых и неголосовых звуков становятся близкими и плохо различимыми. Поэтому указанные звуки, грубо говоря, отличаются по параметру, который фактически изымается из описания. Это не означает, что использование в системах описания речевого сигнала модели авторегрессии с адаптивной разностью хуже, чем использование для той же цели модели авторегрессии со стационарной разностью, а означает лишь то, что в системах описания, использующих модель авторегрессии с адаптивной разностью, необходим специальный признак для описания вида источника возбуждения.

Возникает вопрос: как лучше строить систему описания - на основе раздельного оценивания параметров речевого тракта и источника или без раздельного оценивания? Экспериментальные результаты говорят о том, что первый способ построения предпочтительней. В пользу раздельного оценивания говорит то, что устранение влияния голосового источника повышает надежность, а также то, что различия в надежностях распознавания между моделями с разностями менее ярко выражены в группах II и III, в которых наряду с параметрами мо-

делей были использованы признаки тон/шум и частота основного тона, сигнализирующие о виде источника возбуждения. Эти факты говорят о том, что необходимы специальные исследования для того, чтобы выяснить вопрос о том, как описывать источник возбуждения в случае раздельного оценивания.

5.6. О б и н в а р и а н т н о с т и с и с т е м о п и с а н и я. Несмотря на длительные и многочисленные речевые исследования, направленные на поиск системы описания, инвариантной к диктору, такая система еще не найдена. Поэтому представляют интерес способы повышения инвариантности описаний. Результаты проведенных экспериментов позволяют указать ряд таких способов.

Сравнивая данные табл.4, в которой приведены результаты для случая, когда контрольная и обучающая выборки принадлежали одному диктору, с данными табл.5, полученными для случая, когда обучающая выборка формировалась одним диктором, а контрольная другими, видим, что повысить инвариантность можно: 1) увеличением порядка моделей; 2) использованием моделей с разностями; 3) введением в систему описания признака тон/шум и 4) введением в систему описания признака нормированной частоты основного тона.

5.7. О в ы б о р е п о р я д к а м о д е л е й. Не выписывая конкретных значений порядка моделей, требуемых для построения систем распознавания (их можно найти по табл. 3,4,5), отметим, что для систем распознавания с подстройкой под диктора порядок моделей ниже, чем для систем без подстройки.

Кроме того, если при анализе речи, квантованной с частотой 20 кгц, порядок моделей авторегрессии должен быть не ниже 24–30 [9, 10], то, по результатам данной работы, при распознавании можно обойтись существенно меньшим порядком. Одна из причин этого явления, по-видимому, состоит в том, что минимальное значение порядка модели, при котором достигается максимальная надежность, зависит от объема словаря, в чем нетрудно убедиться, сравнивая данные табл.3 и 4: для меньших по объему словарей можно ограничиться меньшими значениями порядка. В связи с этим целесообразно выяснить, при каком объеме словаря его дальнейшее увеличение не потребует соответствующего повышения минимального значения порядка, обеспечивающего максимальную надежность. Найденное таким образом значение порядка представляет интерес для проведения автоматической сегментации речевого потока, а сам поиск явится одним из направлений дальнейших исследований.

6. Выводы и рекомендации

Проведено сравнение 9 систем описания речевого сигнала, основанных на частной автокорреляционной функции, из которых только одна ранее применялась для построения систем распознавания изолированных слов. Кроме того, предложена модификация алгоритма динамического программирования, ускоряющая процесс принятия решения. Проведенные эксперименты показали весьма высокую эффективность систем описания, построенных на основе частных автокорреляций моделей авторегрессии,и позволили сделать следующие рекомендации и выводы.

I. Наилучшие по надежности результаты дает система описания, построенная по модели авторегрессии со стационарной разностью первого порядка, дополненная признаком нормированной частоты основного тона, а наихудшие - система описания, построенная по модели авторегрессии без разностей,т.е. имено та система, которая ранее уже применялась для построения систем распознавания.

2. Построение систем распознавания по параметрам моделей авторегрессии с разностями предпочтительнее построения по параметрам модели без разностей, так как позволяет повысить надежность распознавания, сжать описание и понизить время принятия решения. Модель авторегрессии со стационарной разностью первого порядка предпочтительнее модели авторегрессии с адаптивной разностью.Анализ причин такого предпочтения указывает на необходимость дальнейшего сравнительного исследования этих моделей.

3. Для повышения инвариантности описания в системы распознавания, работающие без подстройки под диктора, рекомендуется включать признаки тон/шум и нормированной частоты основного тона, а также применять модели авторегрессии с разностями. Эти способы повышения инвариантности на практике, как правило, позволяют сократить описание. При этом время принятия решения либо незначительно увеличивается, либо остается таким же, либо в большинстве случаев сокращается.

4. Одним из перспективных направлений дальнейших исследований является выявление зависимости от объема словаря минимального значения порядка моделей авторегрессии, при котором достигается максимальная надежность.

Литература

I. ЛОЗОВСКИЙ В.С. Аппроксимация отклика системы в Z -плоскости и формантный анализ речи. -В кн.: Вычислительные системы, вып. 37. Новосибирск, 1969, с.22-37.

2. МАКХОЛ Дж. Линейное предсказание. Обзор - Тр. ин-та инж. по электротех. и радиоэлектронике, 1975, т.63, №4, с.20-44.

3. ШАФЕР Р.В., РАБИНЕР Л.Р. Цифровое представление речевых сигналов.- Там же, с. 141-159.

4. КОРОТАЕВ Г.А. Система анализа и синтеза сигнала с линейным предсказанием. Обзор. - Зарубежная электроника, 1976, № 10, с. 3-14.

5. РЕДДИ Д.Р. Машинное распознавание речи. Обзор. - Тр.ин-та инж. по электротех. и радиоэлектронике, 1976, т.64, № 4, с.95-130.

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

7. ВИНЦЮК Т.К., ЛУДОВИК Е.К., ШИНКАЖ А.Г. Распознавание речи на основе параметров предсказания. -В кн.: Автоматическое распознавание слуховых образов. (Материалы всесоюзной школы-семинара.) Тбилиси, 1978, с.181-182.

8. КЕЛЬМАНОВ А.В. Алгоритм классификации тон/шум, основанный на критерии адекватности модели авторегрессии.-В кн.: Методы обработки информации.(Вычислительные системы, вып.74.) Новосибирск, 1978, с. 129-148.

9. КЕЛЬМАНОВ А.В. Оценивание параметров речевого тракта в классе моделей авторегрессии со стационарной сезонной разностью первого порядка и стационарной разностью первого порядка. -В кн.: Эмпирическое предсказание и распознавание образов. (Вычислительные системы, вып. 76.) Новосибирск, 1978, с. 110-131.

10. КЕЛЬМАНОВ А.В. Система распознавания изолированных слов по частной автокорреляционной функции. -Там же, с.132-143.

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

12. БОКС Дж., ДЖЕНКИНС Г. Анализ временных рядов, прогноз и управление. - М.: Мир, т.1, 1974. -206 с.

13. ATAL B.S., HANAUER S.L. Speech analysis and synthesis by linear prediction of speech waveform.- J.Acoust. Soc. Amer.,1971, v.50, N 2,pt 2, p.637-655.

14. MARKEL J.D. Applikation of digital invers filter for automatik formant and F_0 analysis. - IEEE Trans.Audio Electroacoust., 1973,v.AU-21, N 3,p.154-160.

15. WAKITA H. Direct estimation of the vocal tract shape by invers filtering of acoustic speech waveforms.- IEEE Trans. Audio Electroacoust.,1973,v.AU-21, N 5,p.417-427.

16. VISWANATHAN R., MAKHOUL J. Quantization propertes of transmission parameters in linear predictiv systems.-IEEE Trans.Acoust. Speech and Signal Processing,1975,v.ASSP-24,N 2,p.309-321.

17. MAKHOUL J. Spectral Linear prediction: properties and applications.- IEEE Trans.Acoust.,Speech and Signal Processing,1975, v.ASSP-23, N 3,p.283-296.

18. McCANDLESS S.S. An algorithm for automatic formant extraction using linear prediction spectra. - IEEE Trans.Acoust.,Speech and Signal Processing,1974,v.ASSP-22, N 2,p.135-141.

19. MARKEL J.D., GRAY A.M. A linear prediction vovoder simulation based upon the autocorrelation method. - IEEE Trans.Acoust., Speech and Signal Processing,1974,v.ASSP-22, N 2,p.124-134.

20. FLANAGAN J.L. Evaluation of two automatic formant extrae - tors.- J.Acoust.Soc.America,1956,v.28, N 1,p.118-125.

21. De SOUZA P.V. Statistical tests and distans mesure for LPC coefficients.-IEEE Trans.Acoust.,Speech and Signal Processing, 1977,v.ASSP-25, N 6,p.554-564.

22. WHITE G.M., NEELY R.B. Speech recognition experiments with linear prediction, bandpass filtering and dinamic programming. - IEEE Trans.Acoust., Speech and Signal Processing,1976,v.ASSP-24 , N 2,p.183-188.

23. RABINER L.R., SAMBUR M.R. Some preliminary experiments in the recognition of connected digits.- IEEE Trans.Acoust., Speech and Signal Processing,1976,v.ASSP-24, p.170-182.

24. GRAY A.M., MARKEL J.D. Distance mesures for speech processing.- IEEE Trans.Acoust., Speech and Signal Processing, 1976 , v.ASSP-24, N 5,p.380-391.

25. ICHIKAWA I., NAKANO Y., NAKATA K. Evaluation of variaus parameter sets in spoken digits recognition. - IEEE Trans.Audio Electroacoust.,1973,v.AU-21, N 3,p.202-209.

26. ITAKURA F. Minimum prediction residual principle applid to speech recognition. - IEEE Trans.Acoust.,Speech and Signal Processing,1975,v.ASSP-23, N 1,p.67-72.

27. COKER M.J. An improved isolation word recognition system based upon the linear prediction residual. - Proceedings of the IEEE International Joint Conference on ASSP, Philadelphia, April 1976, p.206-209.

28. SAKOE H., CHIBA S. Dynamic programming algorithm optimization for Spoken word recognition. - IEEE Trans.Acoust., Speech and Signal Processing,1978,v.ASSP-26, N 1,p.43-49.

Поступила в ред.-изд.отд.
21 октября 1979 года