Формирование решающих правил с использованием теории мультимножеств в задачах распознавания речи

Е. Е. Федоров[1], Т. В. Ермоленко[2]

В статье рассматривается подход формирования решающего правила при распознавании фонем и идентификации диктора, использующий аппарат теории мультимножеств. При этом также учитываются результаты проведенного сравнительного анализа эффективности наборов признаков для распознавания речи и идентификации диктора. Для выделения акустических характеристик речевого сигнала использовались методы, базирующиеся на психофизических особенностях восприятия речи, и методы, основанные на акустической теории речеобразования.

Введение

В настоящее время актуальной является разработка систем речевого управления, используемых для технического контроля и управления качемирования для станков с ЧПУ и др., которые облегчают работу оператора-технолога, при этом для выбора его словаря эталонов может использоваться процедура идентификации диктора. При разработке таких систем важную роль играет выбор признаков. Эффективность распознавания и идентификации существенно зависит от выбора характеристик сигнала, поступающих на вход классификатора.

Целью настоящей работы является формирование решающего правила при распознавании фонем и идентификации диктора, использующее аппарат теории мультимножеств, а также с учетом результатов анализа эф-

[1] *83086, Донецк, пр. Дзержинского, 7, Донецкий институт автомобильного транспорта, fee75@mail.ru*

[2] *83050, Донецк, пр. Б. Хмельницкого, 84, Государственный университет информатики и искусственного интеллекта, etv@iai.donetsk.ua*

фективности признаков, полученных различными методами выделения акустических параметров речи.

Методы, применяемые для параметризации речевого сигнала, можно разделить на две группы: одна базируется на физиологических и психофизических особенностях слушателя, другая основана на акустической теории образования речи.

Применение нейросетей в качестве средства распознавания имеет преимущества перед распознаванием с использованием статистического подхода: обучение сети настраивает ее на выделение существенных для распознавания признаков [Gupta et al, 2003]; нейросеть позволяет разделять классы, образующие в пространстве признаков области сложной формы, в т. ч. линейно неразделимые и неодносвязные. В связи с этим для распознавания звуков в работе был использован многослойный персептрон [Rabiner et al, 1993], а для идентификации диктора — алгоритм динамического искажения времени (DTW), хорошо зарекомендовавший себя при решении таких задач [Винцюк и др, 1987].

Для формального представления результатов распознавания фонем и идентификации, а также формирования решающего правила использовался аппарат теории мультимножеств.

1. Наборы признаков для распознавания фонем и идентификации дикторов

Банк фильтров можно рассматривать как упрощенную модель человеческой слуховой системы. Согласно психоакустическим принципам восприятия речи, целесообразно в качестве полос пропускания фильтров использовать критические полосы слуха [Секунов, 2001].

В настоящее время вместо банка цифровых фильтров используют быстрое преобразование Фурье, позволяющее ускорить процесс преобразования сигнала с целью формирования эталона. Для дискретного речевого $s(n)$ длиной N его спектр Фурье представлен в виде:

$$S(k) = \sum_{n=0}^{N-1} s(n)\ e^{-j\frac{2\pi nk}{N}},\ 0 \le k \le N/2-1 \qquad (1)$$

К недостаткам преобразования Фурье, традиционно применяемого для спектрального анализа и обработки речевых сигналов, можно отнести невозможность точного восстановления сигнала из-за эффекта Гиббса и от-

сутствие хорошей частотно-временной локализации. Ввиду чего в последнее время получили широкое распространение методы обработки сигнала, базирующиеся на вейвлет-преобразовании [Малла, 2005].

Быстрое вейвлет-преобразование сигнала $s(n)$ на P уровней представляет собой свертку на текущем i-м уровне сигнала с полосовыми фильтрами с коэффициентами lg_n, h_n для получения высоко- (d_{im}) и низкочастотных (c_{im}) составляющих:

$$d_{im} = 2^{1/2} \sum_{n=0}^{N/2^{i-1}-1} c_{i-1,n} g_{n+2m}, \quad c_{im} = 2^{1/2} \sum_{n=0}^{N/2^{i-1}-1} c_{i-1,n} h_{n+2m}, \tag{2}$$

$$c_{0n} = s(n), 0 \le m \le N/2^{i-1} - 1, \quad 1 \le i \le P.$$

Быстрое вейвлет-преобразование формирует компактный результирующий набор коэффициентов, но при этом накладывает серьезные ограничения на выбор масштабов преобразования. Масштаб, на котором проводится анализ сигнала, может быть выбран только из фиксированного ряда значений. Непрерывное вейвлет-преобразование (3) сигнала $s(t)$ является наиболее информативным представлением частотно-временных и масштабно-временных свойств сигнала:

$$CWT_s(a,b) = |a|^{-1/2} \int_{-\infty}^{\infty} s(t) \psi\left(\frac{t-b}{a}\right) dt, \tag{3}$$

где $\psi(t)$ — вейвлет, a — масштабный коэффициент, b — сдвиг.

Чтобы получить вейвлет-коэффициенты дискретного сигнала $s(n)$ длиной N на уровнях разложения от $j_{\min}$ до $j_{\max}$, необходимо применить численное интегрирование и заменить интегралы в (3) суммами. В результате чего получим формулу, представляющую собой аппроксимацию непрерывного вейвлет-преобразования:

$$d_{ml} = \sum_{n=0}^{N-1} s(n) \psi_{ml}(n) \Delta t, \quad j_{\min} \le m \le j_{\max}, \quad 0 \le l \le N-1, \tag{4}$$

где Δt — величина, обратная частоте дискретизации;

$$\psi_{ml}(t) = a_0^{-m/2} \psi\left(a_0^{-m} t - b\, l\right), \quad a_0 > 1, \quad b \ne 0.$$

Другим подходом выделения акустических параметров, основанным на теории образования речи, является метод кодирования с линейным предсказанием (КЛП) [Chu, 2003].

Линейный предсказатель порядка p с коэффициентами a_k определяется как система:

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) \quad 0 \le n \le N-1.$$

Коэффициенты линейного предсказания a_j вычисляются рекуррентно согласно алгоритму Дарбина с помощью автокорреляционной функции (6) и коэффициентов отражения (7):

$$R(i) = \sum_{m=0}^{N-1-i} s(m)s(m+i), \quad 1 \le i \le p, \tag{6}$$

$$k_i := \frac{R(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} R(i-j)}{R(0)\prod_{l=1}^{i-1}\left(1-k_l^2\right)}, \quad 1 \le i \le p, \tag{7}$$

где $a_j^{(i)}$ — j-й коэффициент для линейного предсказателя порядка i.

Коэффициенты k_i однозначно определяют форму кусочно-постоянной акустической трубы, содержащей $(p + 1)$ цилиндрическую секцию фиксированной длины. Процессы в этой трубе — распространение плоской акустической волны, а площади A_i поперечных сечений соседних секций вычисляются согласно:

$$A_{i+1} = \frac{1-k_i}{1+k_i} A_i, \quad A_1 = 1, \quad 2 \le i \le p+1. \tag{8}$$

С помощью коэффициентов КЛП, автокорреляционной функции (6) и автокорреляционной функции коэффициентов КЛП:

$$r(0) = \sum_{j=0}^{p} a_j^2, \quad r(n) = 2\sum_{j=0}^{p-n} a_j a_{j+n}, \quad a_0 = 1 \tag{9}$$

вычисляется сглаженный энергетический спектр:

$$W(k)=\frac{R_n(0)-\sum_{k=1}^{p}a_k R_n(k)}{r(0)-\sum_{n=1}^{p}r(n)\cos\left(\frac{2\pi}{N}nk\right)},\quad 0\le k\le N/2-1 \tag{10}$$

Другим представлением сигнала является комплексный кепстр импульсной характеристики системы линейного предсказания, который получается с помощью рекурсивных соотношений:

$$\hat{h}(n)=a_n+\sum_{k=1}^{n-1}\frac{k}{n}\hat{h}(k)a_{n-k},\quad \hat{h}(0)=a_0,\quad 1\le n\le p. \tag{11}$$

Модификация кепстра применяется для вычисления мел-частотных кепстральных коэффициентов (MFCC), широко используемых в современных системах распознавания речи в качестве набора признаков речевого сигнала. Признаки, построенные на основе MFCC, учитывают психоакустические принципы восприятия речи, поскольку используют мел-шкалу, связанную с критическими полосами слуха, и вычисляются следующим образом [Shannon et al, 2003]:

$$MFCC_k=\sum_{l=1}^{L}E_l\cos\left(k\left(l-0.5\right)\pi/L\right),\quad 1\le k\le L, \tag{12}$$

где $E_l=\lg\left(\sum_{k=k1_l}^{k2_l}\left(S(k)\right)^2 w\left(k-\left(k1_l+\Delta K_l/2\right)\right)\right)$, $S(k)$ — спектр Фурье (1), $k1_l, k2_l$ — границы частотных диапазонов l-й мел-частотной полосы, $\Delta K_l=k2_l-k1_l$ — четное число, $w(x)$ — оконная функция, L — количество мел-полос.

Таким образом, в работе исследуются следующие наборы признаков. $\left\{X_k=\frac{S^2(k)}{\sum_{i=0}^{N/2-1}S^2(i)}\right\}_{k=0}^{N/2-1}$ — нормированный энергетический спектр, вычисленный на основе спектра Фурье (1).

$$\left\{ X_k = \frac{\sum_{i=0}^{N/2-1} S^2(i)}{\sum_{i=0}^{N/2-1} S^2(i) - \sum_{i=0}^{k} S^2(i)} \right\}_{k=1}^{N/2-1}$$ — кумулятивное отношение [Старушко, 2002], построенное на основе спектра Фурье (1) и характеризующее изменение энергии сигнала в зависимости от частоты.

A. $\left\{ X_k = \lg\left(\frac{\sum_{n=N1_k}^{N2_k} S^2(n)}{\sum_{i=0}^{k} \sum_{n=N1_i}^{N2_i} S^2(n)} \right) \right\}_{k=1}^{L}$ — мера контрастности, построенная на основе спектра Фурье (1) и характеризующая изменение энергии в зависимости от полосы частот, где L — количество спектральных полос, $N1_k$ и $N2_k$ — границы k-й полосы.

B. $\left\{ X_k = MFCC_k \right\}_{k=1}^{L}$ — MFCC, вычисляемые согласно (12).

C. $\left\{ X_k = \lg\left(\frac{\sum_{n=0}^{N-1} d_{k+1n}^2}{\sum_{j=1}^{k+1} \sum_{n=0}^{N-1} d_{jn}^2} \right) \right\}_{k=1}^{P-1}$ — мера контрастности, построенная на основе вейвлет-спектра (2) быстрого преобразования по P уровням разложения [Ермоленко, 2005].

D. $\left\{ X_k = \lg\left(\frac{\sum_{n=0}^{N-1} d_{k+j_{\min} n}^2}{\sum_{j=j_{\min}}^{k+j_{\min}} \sum_{n=0}^{N-1} d_{jn}^2} \right) \right\}_{k=1}^{j_{\max}-j_{\min}}$ — мера контрастности, построенная на основе вейвлет-спектра (4) непрерывного преобразования по уровням разложения от $j_{\min}$ до $j_{\max}$ [Ермоленко, 2005].

E. $\left\{X_k = a_k\right\}_{k=1}^{p}$ — коэффициенты КЛП.

F. $\left\{X_i = k_i\right\}_{i=1}^{p}$ — коэффициенты отражения КЛП, полученные по (7).

G. $\left\{X_k = \dfrac{r(k)}{r(0)}\right\}_{k=1}^{p}$ — нормированная автокорреляция КЛП, где $r(k)$ автокорреляция КЛП, получаемая из (9).

H. $\left\{X_k = \hat{h}_k\right\}_{k=1}^{p}$ — кепстр импульсной характеристики системы линейного предсказания, вычисляемый по (11).

I. $\left\{X_k = A_k\right\}_{k=2}^{p+1}$ — площади поперечных сечений кусочно-постоянной акустической трубы, вычисляемые с помощью (8).

J. $\left\{X_k = \dfrac{R(k)}{R(0)}\right\}_{k=1}^{p}$ — нормированная автокорреляция, где $R(k)$ — автокорреляция, вычисляемая по (6).

K. $\left\{X_k = \dfrac{W^2(k)}{\sum_{i=0}^{N/2-1} W^2(i)}\right\}_{k=0}^{N/2-1}$ — нормированный сглаженный энергетический спектр, где $W(k)$ — энергетический спектр, определяемый из (10).

L. $\left\{X_k = \lg\left(\dfrac{W^2(k)}{\sum_{i=0}^{k} W^2(i)}\right)\right\}_{k=1}^{N/2-1}$ — меры контрастности сглаженного энергетического спектра, где $W(k)$ — энергетический спектр, вычисляемый с помощью (10).

2. Численное исследование эффективности наборов признаков для распознавания фонем с помощью нейросети

Предлагаемая методика распознавания фонем состоит из двух этапов: обобщенной и детальной классификации. В ходе работы первого этапа распознаваемый звук относится к одному из широких фонетических классов (ШФК), полученных согласно классификации звуков русской речи по их образованию:

$$\Omega = \left\{ \Omega_i \right\}_{i=1}^{6} = \left\{ Sh, P, Cons1, Cons2, Son, Vow \right\}, \tag{13}$$

где *Sh* — шумные глухие щелевые ([ф], [с], [х], [ш]) и шумные глухие смычно-щелевые ([ц], [ч]); *P* — шумные глухие смычные ([к], [т], [п]); *Cons*1 — шумные звонкие щелевые ([в], [з], [ж]); *Cons*2 — шумные звонкие смычные ([б], [д], [г]); *Son* — сонанты ([й], [л], [м], [н], [р]); *Vow* — гласные ([а], [и], [о], [у], [ы], [э]). Обобщенная классификация проводится с помощью дерева принятия решений, узлами которого являются ШФК, а условиями перехода — близость к эталонным значениям признаков распознавания. Для каждого узла набор этих признаков индивидуален, в качестве признаков, по которым осуществляется классификация, выступают длина квазипериода сигнала и энергии спектра на мел-частотных полосах.

Детальная классификация распознаваемого звука осуществляется в пределах введенных ШФК с помощью нейросети. При этом для каждого из классов строится и обучается соответствующая сеть. Исключение составляет класс *P*, поскольку распознавать фонемы, входящие в него, не представляется возможным в силу малой амплитуды шумных глухих смычных звуков и практически идентичного их спектрального состава.

Для проведения численного исследования был программно реализован алгоритм распознавания фонем внутри ШФК с помощью трехслойного персептрона (входной слой — нулевой). Для обучения нейросети использовался алгоритм обратного распространения ошибки. Количество входных нейронов соответствует размерности вектора признаков, количество выходных нейронов *fnum* соответствует количеству фонем в ШФК, количество нейронов в скрытых слоях равно 3*fnum* и 2*fnum* соответственно. В качестве функции активации первого слоя была выбрана сигмоида (14), второго скрытого и выходного слоев — гиперболический тангенс (15):

$$F(x) = \frac{1}{1+e^{-x}}, \tag{14}$$

$$F(x) = th(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}. \qquad (15)$$

В экспериментах участвовало 100 дикторов, каждый из них по 10 раз произносил слова, из которых выделялись фонемы. В качестве наборов признаков использовались признаки (**A**)-(**N**). Признаки (**G**)-(**N**) получены методом КЛП при $p = 30$, признаки (**E**) были построены на основе вейвлета Добеши 4-го порядка, количество уровней разложения P=8, признаки (**F**) были получены на основе вейвлета Морле при $a_0 = 1.1$; $j_{\min} = 10$; $j_{\max} = 50$.

В табл. 1 приведены результаты распознавания фонем внутри ШФК. По этим классам фонем в табл.1 занесены: N_ε — количество циклов обучения; p_{nw} — вероятность распознавания фонем внутри исследуемых ШФК.

Как видно из табл. 1, наилучшие результаты распознавания для всех классов фонем по рассмотренным наборам признаков дают признаки, построенные на основе меры контрастности вейвлета Морле. Фонемы, входящие в класс *Cons*2, в своей стационарной части имеют незначительные

Таблица 1

Результаты численного исследования наборов признаков, используемых при классификации фонем

Набор признаков	*Sh*		*Cons*1		*Cons*2		*Son*		*Vow*	
	N_ε	p_{nw}	N_ε	p_{nw}	N_ε	p_{nw}	N_ε	p_{nw}	N_ε	p_{nw}
A	3653	0.73	198633	0.97	58421	0.25	14205	0.84	2490	0.79
B	263	0.78	1292	0.97	13722	0.29	12500	0.89	237	0.81
C	233	0.86	2870	0.97	12850	0.33	13783	0.89	4091	0.79
D	1285	0.79	22	0.97	65	0.38	87	0.77	31	0.81
E	300	0.67	1656	0.95	33530	0.31	12143	0.52	780	0.82
F	336	0.93	110	0.99	1287	0.52	1150	0.96	60	0.99
G	115	0.61	3619	0.91	15341	0.32	2562	0.80	633	0.78
H	382	0.91	7132	0.92	10821	0.35	1234	0.86	358	0.84
I	5843	0.71	1014	0.91	45418	0.28	56368	0.72	53758	0.68
J	327	0.70	239	0.93	54274	0.30	2470	0.80	110	0.84
K	5971	0.72	13727	0.92	534	0.29	213405	0.75	2698	0.72
L	5812	0.71	1201	0.93	51021	0.33	55842	0.71	52734	0.69
M	416	0.87	567	0.94	2350	0.32	165	0.81	579	0.64
N	4376	0.70	1067	0.93	3475	0.31	62597	0.68	57246	0.65

отличия, в связи с этим вероятность распознавания внутри этого класса крайне низка по всем предложенным наборам признаков, поэтому распознавать их между собой в пределах этого ШФК нецелесообразно.

3. Численное исследование эффективности наборов признаков для идентификации диктора с помощью DTW

Для проведения численного исследования был программно реализован алгоритм DTW, при этом в качестве меры близости была выбрана евклидова метрика. В экспериментах участвовало 100 дикторов. Каждый диктор 5 раз произносил ключевое слово. В качестве наборов признаков использовались признаки (**A**)-(**N**). Признаки (**G**)-(**N**) получены методом КЛП при $p = 30$, признаки (**E**) были построены на основе вейвлета Добеши 4-го порядка, количество уровней разложения P=8, признаки (**F**) были получены на основе вейвлета Морле при $a_0 = 1.1$; $j_{\min} = 10$; $j_{\max} = 50$.

Результаты проведенного исследования сведены в табл. 2, куда занесены вероятности идентификации диктора по ключевому слову «Саша», фонемам [ш] и [а].

Таблица 2

Результаты численного исследования наборов признаков, используемых при идентификации диктора

Набор признаков	слово «Саша»	фонема [а]	фонема [ш]
A	0.86	0.78	0.22
B	0.94	0.78	0.22
C	0.94	0.66	0.38
D	0.98	0.9	0.86
E	0.78	0.54	0.38
F	0.82	0.58	0.6
G	0.86	0.36	0.68
H	0.98	0.72	0.86
I	0.86	0.36	0.68
J	0.82	0.4	0.58
K	0.7	0.52	0.32
L	0.76	0.54	0.34
M	0.6	0.24	0.3
N	0.6	0.2	0.22

Численное исследование позволяет сделать вывод, что из исследуемых наборов признаков при идентификации диктора по одной фонеме наиболее эффективными являются MFCC. Для тональных фонем (в данном случае [а]) вероятность идентификации составила 0.9. Для глухих щелевых (или смычно щелевых) фонем (в данном случае [ш]) вероятность идентификации — 0.86. Эффективность признаков при идентификации диктора по слову (в данном случае «Саша») выше, чем по фонемам. При этом лучшие результаты показывают MFCC и коэффициенты отражения КЛП (вероятность идентификации — 0.98).

По результатам распознавания фонем и идентификации диктора формируется решающее правило, которое можно конструировать как на основе количества наборов признаков, подтверждающих диктора, так и на основе расстояний между ключевым словом и эталоном. В статье рассматриваются оба варианта с использованием аппарата мультимножеств [Петровский, 2003].

4. Формирование решающих правил распознавания фонем и идентификации диктора на основе аппарата теории мультимножеств

Результаты распознавания *i*-го звука в слове представлены следующим образом:

$$A_{is} = \left\{ \left(k^{1}_{i1s}, p^{1}_{s}, x^{1}_{1} \right), \ldots, \left(k^{1}_{iN1s}, p^{1}_{s}, x^{1}_{N1} \right), \ldots, \left(k^{j}_{i1s}, p^{j}_{s}, x^{j}_{1} \right), \ldots, \right.$$

$$\left. \left(k^{j}_{iNjs}, p^{j}_{s}, x^{j}_{Nj} \right), \ldots, \left(k^{6}_{i1s}, p^{6}_{s}, x^{6}_{1} \right), \ldots, \left(k^{6}_{iN6s}, p^{6}_{s}, x^{6}_{N6} \right) \right\}, \qquad (16)$$

$$k^{j}_{ils} \in \{0,1\}, s \in \overline{1, M},$$

где k^{j}_{ils} — логический признак, который показывает в случае *s*-го набора признаков, относится ли *i*-й распознаваемый звук к фонеме класса Ω_j (13), имеющей код x^{j}_{l}, Nj — количество фонем класса Ω_j, p^{j}_{s} — вероятность

распознавания фонем внутри класса Ω_j с помощью нейросети, использующей s-й набор признаков (табл. 1). При чем, поскольку в пределах классов Ω_2 и Ω_4 распознавание с помощью нейросети не проводится, то

$$N2 = N4 = 1, \quad p_s^2 = p_s^4 = 1 \forall s.$$

Результаты идентификации дикторов по i-й реализации кодового слова или фонемы по каждому s-му набору признаков представлены соответствующим множеством:

$$A_{is} = \left\{ p_s, \left(k_{is1}, x_1 \right), \ldots, \left(k_{isj}, x_j \right), \ldots, \left(k_{isN}, x_N \right) \right\},$$
$$k_{isj} \in \{0,1\}, \quad s \in \overline{1,M} k_{isj}, \tag{17}$$

где p_s — вероятность идентификации диктора с помощью DTW на базе s-ого набора признаков (табл. 2), k_{isj} — логический признак, показывающий относится ли i-я реализация ключевого слова (фонемы) к эталону диктора с кодом x_j в случае s-го набора признаков.

Однако представление результатов в виде (16), (17) приводит к следующим недостаткам:

а) необходимость хранения M множеств для каждого звука распознаваемого слова в случае пофонемного распознавания, каждой реализации слова (фонемы) в случае идентификации диктора;

б) затруднено построение решающего правила (разные наборы признаков могут давать разные результаты).

Чтобы сократить занимаемое место, сделать формальную запись более компактной и сконструировать решающее правило, в работе предлагается использовать аппарат мультимножеств, что позволяет хранить только одно мультимножество вместо M множеств. Для распознавания фонем соответствующее мультимножество имеет вид (18), решающее правило — (19), для идентификации диктора мультимножество может быть записано в виде (20), решающее правило — (21):

$$B_i = \left\{ k_{Bi}\left(x_1^1 \right) \cdot x_1^1, \ldots, k_{Bi}\left(x_l^j \right) \cdot x_l^j, \ldots, k_{B_i}\left(x_{N6}^6 \right) \cdot x_{N6}^6 \right\},$$
$$k_{B_i}\left(x_l^j \right) = \sum_{s=1}^{M} p_s^j k_{ils}^j, \tag{18}$$

$$IF\left\langle \arg\max_{(m,n)} k_{B_i}\left(x_n^m\right) = x_l^j \right\rangle, THEN\left\langle ObjectB_i \leftrightarrow x_l^j \right\rangle, \quad (19)$$

$$B_i = \left\{ k_{Bi}\left(x_1\right)\cdot x_1, \ldots, k_{B_i}\left(x_N\right)\cdot x_N \right\}, \quad k_{B_i}\left(x_j\right) = \sum_{s=1}^{M} p_s k_{isj}, \quad (20)$$

$$IF\left\langle \arg\max_{m} k_{B_i}\left(x_m\right) = x_j \right\rangle, THEN\left\langle ObjectB_i \leftrightarrow x_j \right\rangle. \quad (21)$$

Заключение

В статье был предложен подход формирования решающих правил распознавания фонем и идентификации диктора на основе аппарата мультимножеств. При этом учитывались результаты проведенного численного исследования эффективности наборов признаков речевого сигнала, базирующихся на различных методах параметризации речевого сигнала. Для сравнительного анализа наборов признаков в качестве метода распознавания был выбран многослойный персептрон, а в качестве метода идентификации — DTW.

В результате исследования для распознавания фонем внутри ШФК наиболее эффективными для всех классов фонем являются признаки, построенные на основе меры контрастности непрерывного вейвлет-преобразования.

Для методики идентификации диктора по результатам проведенных численных исследований в качестве набора признаков было выбрано сочетание коэффициентов отражения и MFCC.

Основные положения работы могут быть использованы при разработке систем распознавания речи и идентификации диктора, которые могут применяться в естественно-языковых интерфейсах АСУ промышленного производства.

Литература

1. *Chu W. C.* Speech coding algorithms. New Jersey: WILEY-intercience, 2003. P. 558.
2. *Gupta M. M., Jin L., Homma N.* Static and Dynamic Neural Networks. New Jersey: John Wiley & Sons, 2003. P. 722.
3. *Rabiner L. R., Jang B. H.* Fundamentals of speech recognition. New Jersey: Prentice Hall PTR, Englewood Cliffs, 1993. P. 507.
4. *Shannon B. J., Paliwal K. K.* A comparative study of filter bank spacing for speech recognition // Microelectronic engineering research conference. 2003. P. 79–81.

5. *Винцюк Т. К.* Анализ, распознавание и интерпретация речевых сигналов. К.: Наук. думка, 1987. С. 261.
6. *Ермоленко Т. В.* Разработка системы распознавания изолированных слов русского языка на основе вейвлет-анализа // Искусственный интеллект. 2005. № 4. С. 595–601.
7. *Малла С.* Вейвлеты в обработке сигналов. М.: Мир, 2005. С. 671.
8. *Петровский А. Б.* Пространства множеств и мультимножеств. М.: URSS, 2003. С. 248.
9. *Секунов Н. Ю.* Обработка звука на PC. СПб.: БХВ — Санкт-Петербург, 2001. С. 1248.
10. *Старушко Д. Г.* Вычисление кумулятивного отношения на основе быстрого преобразования Хартли и нейросетевое распознавание речевых единиц с его использованием // Мат-лы Междунар. науч.-техн. конф. «Искусственный интеллект – 2002». Т. 2. Таганрог, 2002. С. 327–330.