УДК 004.934.1'1

МЕТОДИКА ФОРМИРОВАНИЯ ЭТАЛОНОВ ФОНЕМ, БАЗИРУЮЩАЯСЯ НА ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИИ МОРЛЕ

Ермоленко Т.В.

Донецкий государственный институт искусственного интеллекта,
ул.Артема, 118-в, Донецк, Украина, 340048
e-mail: etv@iai.donetsk.ua

Abstract. At creation of systems of speech recognition the important role is played a choice of features, for references of phoneme generation. For the decision of this task in the article the technique of formation of the references of phoneme is developed. It is based on wavelet transformation. For increase of probability of recognition for each pair phoneme classes from an proposed set of features the optimal feature select, on which the division of this classes is made.

Постановка проблемы. На сегодняшний день актуальной является разработка образного компьютера, одним из основных элементов которого является речевой интерфейс. Для этого интерфейса важную роль играет процедура выделения признаков, входящая в состав подсистем обучения и распознавания речи. Для построения эффективных систем распознавания речи, необходимо решить задачу преобразования входного речевого сигнала в набор акустических параметров и формирования из них вектора признаков, который в дальнейшем будет использован для распознавания.

Анализ исследований. Анализ последних достижений и публикаций, посвященных этой проблеме, позволяет сделать вывод, что большинство методов обработки речевых сигналов, используемых в системах распознавания, базируются на анализе спектральных составляющих звуков речи. В настоящее время, кроме методов, основанных на дискретном преобразовании Фурье, цифровой фильтрации и кодировании с линейным предсказанием, разработаны подходы, опирающиеся на вейвлет-преобразование [1], [2], [3].

Нерешенным является вопрос о выборе оптимального признака для точного разделения двух классов фонем.

Постановка задачи. *Целью настоящей работы является* разработка методики формирования эталонов фонем, базирующаяся на вейвлет-преобразовании.

Решение задачи. В статье предлагается методика формирования эталонов фонем, для которой разрабатываются правила:

- вычисления вейвлет-преобразования речевого сигнала;
- построения векторов признаков фонем на основе меры контрастности;
- выделения среди этих признаков оптимальных для каждой пары классов фонем.

Непрерывное вейвлет-преобразование функции $f(t) \in L^2(R)$ [4],[5],[6]:

$$CWT(a,b) = |a|^{-1/2} \int\limits_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right) dt = \int\limits_{-\infty}^{\infty} f(t)\psi_{ab}(t)dt,$$

где $\psi_{ab} = |a|^{-1/2}\psi\left(\frac{t-b}{a}\right)$

На практике сигнал f имеет конечную длину N и квантован по времени ($f = f(n), \quad n = 0, \ldots, N-1$). Для получения вейвлет-коэффициентов сигнала f необходимо применить численное интегрирование, т. е. заменить интегралы суммами, для чего вводят правила квантования масштабирующей переменной a и сдвига b: $a = a_0^j, \quad b = kb_0,\ j,k \in \mathbb{Z},\ a > 1,\ b_0 \neq 0$

Тогда вейвлет-преобразование сигнала $f(n)$ может быть трансформировано к виду

$$d_{jk} = \sum_{n=0}^{N-1} f(n)\psi_{jk}(n)\Delta t \quad , \tag{1}$$

где $\psi_{jk}(n) = |a_0^j|^{-1/2}\psi\left(\frac{n-kb_0}{a_0^j}\right),\ j_{\min} \leq j \leq j_{\max},\ 0 \leq k \leq N-1,$

Δt — шаг квантования по времени, обратный частоте дискретизации, j — уровень вейвлет-разложения, k — параметр сдвига.

Вейвлет-функции ψ_{jk} представляют собой полосовые нерекурсивные фильтры (КИХ-фильтры), ширина полосы пропускания которых зависит от масштабирующей переменной a_0^j, а центральная частота — от параметра сдвига k [6].

В качестве материнского вейвлета был использован вейвлет Морле

$$\psi(n) = \cos(\omega_0 n) \cdot \exp\left(-\frac{n^2}{2}\right)$$

Все вейвлет-функции $\psi\left(\frac{t}{a_0^j}\right)$ локализованы, т. е. определены на интервале $[\tau_j^1; \tau_j^2]$ (временном окне), за пределами которого их можно положить равными 0. Формулы для вычисления центра и ширины временного окна приведены в [4]. С целью ускорения расчетов значения функций $\psi_{jk}(n)$ табулируются, а при вычислении коэффициентов d_{jk} по формуле (1 меняются индексы суммирования:

$$d_{jk} = \sum_{n=n1}^{n2} f(n)\psi_{jk}(n)\Delta t \quad , \tag{2}$$

где $n1 = [a_0^j\tau_j^1 + kb_0],\ n2 = [a_0^j\tau_j^2 + kb_0]$ (оператор [ ] выполняет выделение целой части числа).

Вектор признаков $P = (C_{j\min+1}, \ldots, C_{j\max})$ имеет размерность $j\max - j\min - 1$ и строится на основе энергетической характеристики, которая в [4] называется мерой

контрастности:

$$C_j = \frac{E_j}{\sum\limits_{i=j\,\min}^{j} E_i} \quad , \tag{3}$$

где $E_j = \sum\limits_k d_{jk}$

Набор фильтров ψ_{jk} с различной центральной частотой и шириной полосы пропускания позволяет анализировать спектральный состав исследуемого сигнала во всем диапазоне частот, а мера контрастности определяет изменение энергии сигнала в различных частотных полосах. Спектры звуков, несмотря на их вариативность в разных реализациях и на разных уровнях вейвлет-разложения одной фонемы, имеют большое сходство для данной фонемы и значительно отличаются для разных фонем. Это отличие имеет стабильный характер и сохраняется от реализации к реализации, значит, по вектору признаков, построенному таким образом, можно будет адекватно провести квалификацию.

Выбор оптимального признака распознавания пары фонем. Все фонемы русского языка по способу их образования делятся на несколько широких фонетических классов [7]: гласные, сонорные, шумные звонкие и шумные глухие. В свою очередь, сонорные и шумные подразделяются на щелевые и смычные, а в шумных глухих выделяют еще и аффрикаты (смычно-щелевые). Таким образом, всё пространство фонем $W = \{w\}$ можно разбить на непересекающиеся классы: $W_1, W_2, \ldots, W_L$. Из предложенной системы признаков распознавания для каждой пары классов фонем W_l и W_m выберем один, оптимальный, обеспечивающий максимальную компактность этих классов и их разделимость.

Каждый из классов W_l и W_m описан в пространстве признаков множеством векторов $W_l = \{\vec{X}_1^l, \vec{X}_2^l, \ldots, \vec{X}_{N_l}^l\}$ и $W_m = \{\vec{X}_1^m, \vec{X}_2^m, \ldots, \vec{X}_{N_m}^m\}$, где N_l, N_m — число реализаций фонем из классов W_l и W_m соответственно. Несмещенная оценка математического ожидания (МО) значения j-того признака в классе W_l имеет вид:

$$M_j^l = \frac{1}{N_l} \sum_{k=1}^{N_l} X_{kj}^l \tag{4}$$

Центроидом $\overrightarrow{Avg}^l$ класса W_l будем считать вектор $\overrightarrow{Avg}^l = (M_{j\,\min}^l, \ldots, M_{j\,\max}^l)$. Вектора-центроиды каждого класса сохраняются в базе данных и являются эталонами фонем соответствующих классов.

Мерой компактности класса W_l по j-тому признаку назовем величину, являющуюся несмещенной оценкой дисперсии значений этого признака

$$D_j^l = \frac{1}{N_l - 1} \sum_{k=1}^{N_l} (X_{kj}^l - M_j^l)^2 \tag{5}$$

Сравнение признаков по обеспечению компактности пары классов W_l и W_m может быть выполнено на основе оценки средней дисперсии:

$$D_j^{lm} = D_j^l \frac{N_l}{N_l + N_m} + D_j^{lm} + D_j^m \frac{N_m}{N_l + N_m}, \tag{6}$$

Оценкой МО средних значений j-того признака классов W_l и W_m является:

$$M_j^{lm} = M_j^l \frac{N_l}{N_l + N_m} + D_j^{lm} + M_j^m \frac{N_m}{N_l + N_m}, \tag{7}$$

Величиной, характеризующей разделительные свойства пары классов W_l и W_m по j-тому признаку является дисперсия МО j-того признака:

$$D_{Mj}^{lm} = (M_j^l - M_j^{lm})^2 \frac{N_l}{N_l + N_m} + (M_j^m - M_j^{lm}) \frac{N_m}{N_l + N_m} \tag{8}$$

Признак, реализующий минимум комбинированного критерия, построенного на основе средней дисперсии D_j^{lm} и дисперсии МО D_{Mj}^{lm}, будем считать оптимальным для распознавания классов W_l и W_m:

$$j* = \arg\min_j \frac{D_j^{lm}}{D_{Mj}^{lm}} \tag{9}$$

В результате численного исследования для каждой пары широких фонетических классов рассчитывается номер признака оптимального для их разделимости. Таким образом, для классов L строится матрица $[A_{ij}]_{i,j=1}^L$ смежности классов, элемент которой A_{lm} —номер признака, обеспечивающего максимальную разделимость классов W_l и W_m. Вместе со сформированными ранее эталонами фонем полученная матрица включается в базу данных, которая в дальнейшем будет использоваться в процедуре распознавания.

Для проведения численного исследования предложенные методики были реализованы в программном комплексе. Вейвлет-преобразование проводилось при $a_0 = 1.1$, $b_0 = 1$, $j_{\min} = 10$, $j_{\max} = 50$. Количество реализаций фонем из каждого класса было не менее 40. Было проведено сравнение и исследование щелевых и гласных звуков, глухих согласных и гласных, звонких и сонатных звуков. Эти классы звуков наилучшим образом разделяются на начальных уровнях разложения. При анализе по разделению гласных звуков между собой можно сделать вывод, что звуки |а, о, э, и, у, ы| различаются при всех сочетаниях в окрестностях уровня 25.

На рис. 1 приведены графики, на которых показаны центроиды и разброс вокруг них в пределах трехсигмовой зоны для классов звонких и сонатных звуков. Оптимальный признак для распознавания этих классов соответствует уровню $j = 2$, показанному непрерывной прямой линией.

Распознавание между парами групп звуков по одному оптимальному признаку из предложенной системы признаков позволило получить точность распознавания 90 процентов.

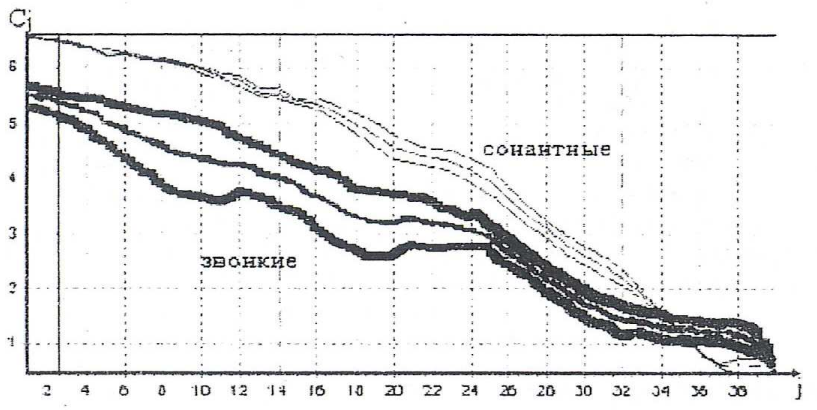

Рис.1: Разброс значений признаков вокруг центроидов для классов звонких и сонантных звуков

Рис. 1. Разброс значений признаков вокруг центроидов для классов звонких и сонатных звуков

Новизна. *Основным результатом данной статьи является* разработанная методика формирования эталонов фонем. В соответствии с чем были предложены правила вычисления вейвлет-преобразования речевого сигнала, построение векторов признаков фонем и выделение среди этих признаков оптимальных для каждой пары классов фонем. Использование порогов в виде оптимальных признаков вместо векторов признаков позволяет уменьшить пространство поиска и время распознавания.

Практическое значение. *Представляется перспективным дальнейшее изучение и разработка* правил построения векторов признаков фонем, базирующихся на вейвлет-преобразовании. Основные положения данной работы предназначены для реализации в интеллектуальных системах управления, в которых команды поступают на естественном языке.

Список литературы

1. Бойков Ф.Г., Старожилова Т.К. Применение вейвлет-анализа сигнала в системе распознавания речи. Сб. докладов 12-й Всероссийской конференции «Математические методы распознавания образов». Москва, 2005.
2. Юрков П.Ю., Федоров В.М., Бабенко Л.К. «Распознавание фонем русского языка с помощью нейронных сетей на основе вейвлет преобразования». Нейрокомпьютеры: разработка, применение. - 2001. - № 8. - С. 81-185.
3. Ермоленко Т.В. Использование непрерывного вейвлет-преобразования при распознавании вокализованных участков речевого сигнала. Искусственный интеллект. - 2004. - № 4. -С. 499-503.
4. Астафьева Н.М. Вейвлет-анализ: основы теории и примеры применения. Успехи физических наук. - 1998,Т.166. - № 11. -С. 1145-1170.

5. Воробьев В.И., Грибунин В.Г. Теория и практика вейвлет-преобразования. - СПб.: Изд-во ВУС, 1999. - 208 с.
6. Добеши И. Десять лекций по вейвлетам. - Москва-Ижевск: НИЦ «Регулярная и хаотическая динамика», 2004. - 464 с.
7. Современный русский язык: Учеб. для филол. спец. высших учебных заведений. Под ред. В.А. Белошапковой. - М.: Азбуковник, 1997. - 928 с.