

АЛГОРИТМ ПОСТРОЕНИЯ ВЕКТОРОВ ПРИЗНАКОВ НА ОСНОВЕ ВЕЙВЛЕТ ПРЕОБРАЗОВАНИЯ ДЛЯ КЛАССИФИКАЦИИ ФОНЕМ РУССКОГО ЯЗЫКА

А. М. Сорока

Белорусский государственный университет, Минск, Беларусь

ВВЕДЕНИЕ

В настоящее время для построения векторов признаков акустических сигналов в системах распознавания речи широко используются мел-частотные кепстральные коэффициенты (МЧКК) [1]. Однако, как показывают практические исследования [2], использование данного подхода не обеспечивает достаточной точности классификации акустических сигналов, что может быть обусловлено близостью векторов признаков в признаковом пространстве. В данной статье предложены два алгоритма извлечения векторов признаков на основе вейвлет преобразования, обладающего более высокой способностью к выделению локальных частотно-временных особенностей сигнала в сравнении с традиционным кратковременным Фурье-преобразованием.

ПОСТРОЕНИЕ ВЕКТОРОВ ПРИЗНАКОВ НА ОСНОВЕ ВЕЙВЛЕТ ПРЕОБРАЗОВАНИЯ

В данной статье рассматриваются два алгоритма извлечения векторов признаков для речевых сигналов на основе вейвлет-анализа [2].

Первый алгоритм фундируется возможностью провести сегментацию и распознавание фонемы посредством визуального анализа графического представления результатов вейвлет-преобразования. Данный способ построения векторов признаков (ВП1) основан на методах смежной дисциплины - распознавания графических образов и может быть описан следующей последовательностью действий. Графический вейвлет образ сегментируется на участки, соответствующие одному периоду в квазипериодической трактовке вейвлет образа, далее в каждом сегменте детектируются резкие характерные изменения с использованием алгоритма детектора Харриса. Следующим шагом является нормализация координат полученных характерных точек. Для формирования вектора признаков характерные точки представляются в виде смеси двумерных Гауссовых распределений [4]:

$$p^{(X)} = \prod_{j=1}^K w_j P(x|C_j)$$

где w_j - весовой коэффициент, $P(x|C_j)$ - вероятность того, что тестируемый

вектор, C_j - предполагаемый кластер, K - количество компонент в смеси,

~~$\mathbf{Z}$ - матрица вида $\mathbf{Z} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$~~ . Вектор признаков для

заданного образа может быть описан следующим выражением:

$$\mathbf{x} = (A^1, M^1, \dots, M^K, \dots, M^K)$$

Данный метод продемонстрирован на рис. 1

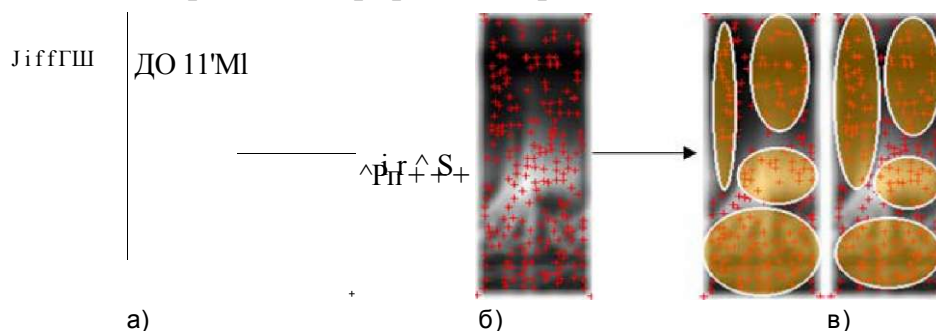


Рис. 1. Формирование вектора признаков с использованием методов анализа изображений.

Сегментация исходного изображения (а), нахождение характерных точек (б), аппроксимация распределения ключевых точек с использованием смеси Гауссовских распределений

Во втором случае (ВП2), для формирования вектора признаков вейвлет образ акустического сигнала разбивается на $3N$ прямоугольных окон, в каждом из которых находится усредненная энергия S_j , $i = 1 \dots N$, $j = 1 \dots 3$. В данном случае вектор признаков описывается следующим выражением:

$$\mathbf{x} = (S_{12}, S_{13}, \dots, S_{N2}, S_{N3}, \dots, S_{NX})$$

параметры $L_i = S_{i3} - S_{i1}$ введены для учета динамических процессов в начале и конце фонемы, обусловленных эффектами редукции и коартикуляции. Данный алгоритм представлен на рис. 2.

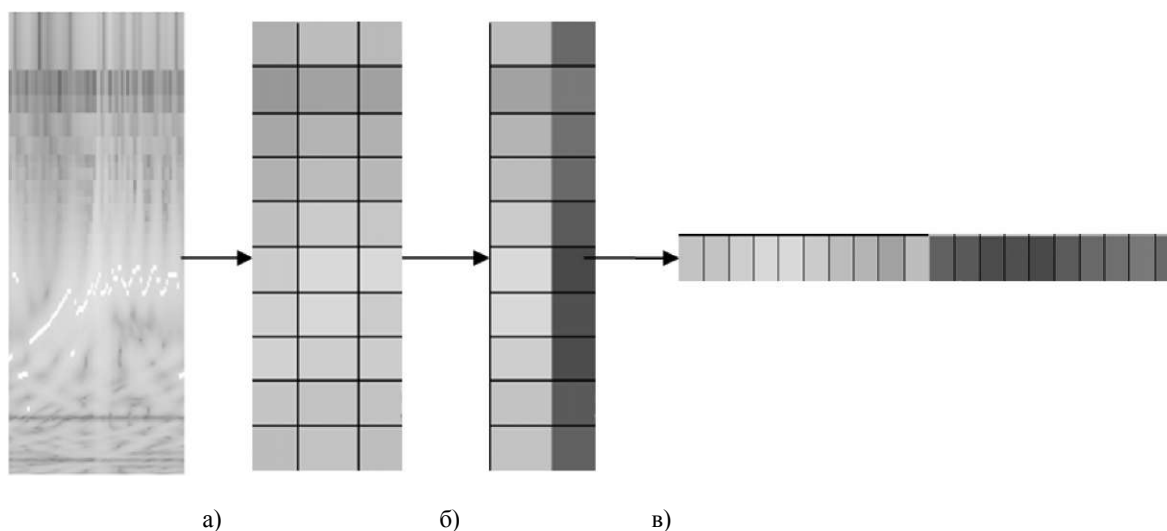


Рис. 2. Формирование вектора признаков с использованием вейвлет-преобразования. Разделение вейвлет образа на $3 \times N$ окон (а), учет динамических процессов (б), формирование численных признаков (в)

ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ

В рамках данной статьи проведена серия экспериментов по классификации фонем русского языка. Для проведения экспериментов создана база данных с использованием акустического материала, собранного на кафедре радиофизики Белорусского Государственного Университета, а так же свободной базы VoxForge [3]. В качестве классификатора выбран метод опорных векторов (МОВ) с радиальной базисной ядерной функцией (RBF). Оптимальные параметры классификатора и ядерной функции найдены с использованием методов поиска по сетке и кросспроверки. Результаты экспериментов по определению оптимальных параметров предложенных алгоритмов представлены на рисунках 3 и 4.

При проведении экспериментов по определению оптимальных параметров в качестве тестового набора данных использовались 100 реализаций фонемы [а], при этом в обучающей выборке объемом 1000 реализаций не были представлены другие гласные фонемы, что обеспечило возможность достижения полностью верной классификации.

Для определения характеристик разработанных методов проведено сравнительное тестирование метода построения векторов признаков на основе мел-частотных кепстральных коэффициентов (МЧКК) и предложенных методов.

Для проведения эксперимента сформирована обучающая выборка из 4000 звуков различных фонем русского языка, из которых 700

соответствуют фонеме [а] и тестовая выборка из 300 звуковых реализаций фонемы [а].

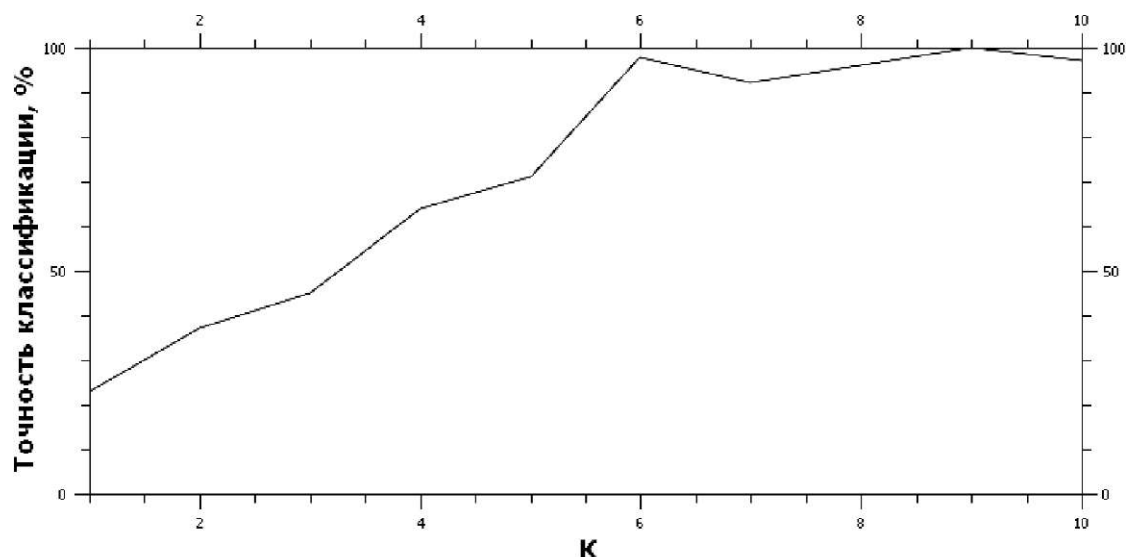


Рис. 3. Зависимость точности классификации от количества компонент в смеси Гауссовских распределений для метода ВП1

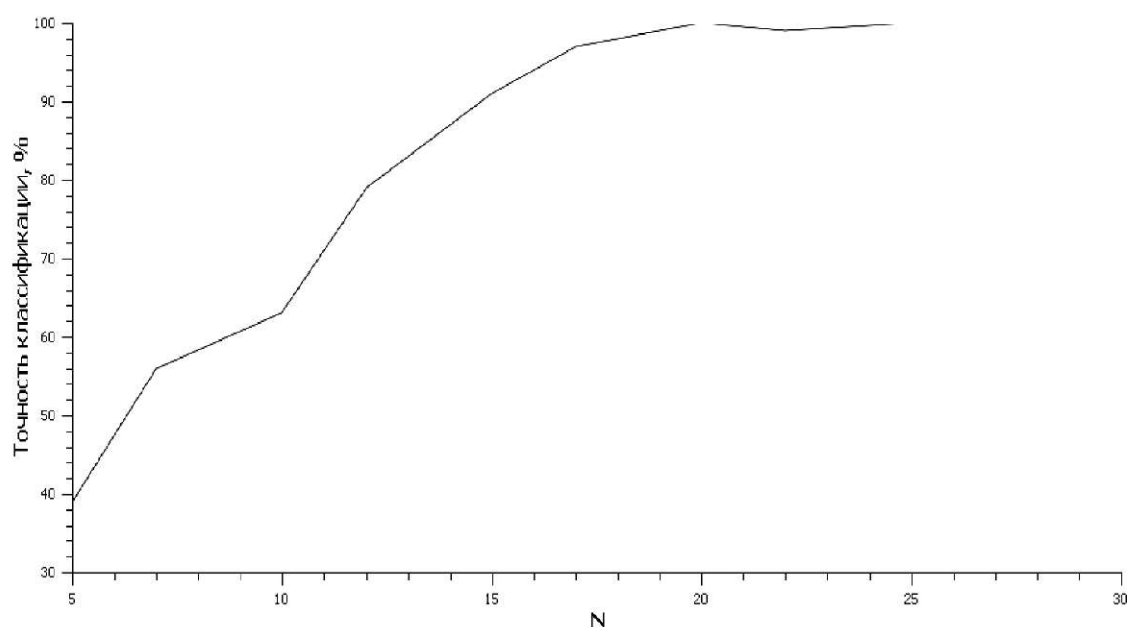


Рис. 4. Зависимость точности классификации от числа окон N для алгоритма ВП2

Точность классификации с использованием алгоритмов ВП1, ВП2 и МЧКК составила 60%, 82% и 80% соответственно. Для проведения эксперимента по классификации близкорасположенных в признаковом пространстве фонем сформирована обучающая выборка из 1000 звуков гласных фонем и тестовая выборка из 100 звуков фонемы [а]. В данном

эксперименте точность классификации с использованием алгоритмов ВП1, ВП2 и МЧКК составила 76%, 92% и 82% соответственно.

ЗАКЛЮЧЕНИЕ

В данной статье рассматриваются два алгоритма построения векторов признаков для акустических сигналов на основе вейвлет преобразования. Использование первого метода (ВП1) не показало практически значимых результатов, что может быть обусловлено некорректным моделированием распределения характерных точек на вейвлет образе. В тоже время, использование второго метода (ВП2) показало результаты, превосходящие результаты использования традиционно используемых методов формирования векторов признаков МЧКК на 2% при классификации фонов в общем случае и на 10% при классификации близкорасположенных в признаковом пространстве фонов.

Литература

1. *Huang, X.* Spoken Language Processing: a guide to theory, algorithm, and system development. - New Jersey: Prentice-Hall Inc. Upper Saddle River, 2001.
2. *Siafarikas, M.* Speech Recognition using Wavelet Packet Features / M. Siafarikas, I. Mporas, T. Ganchev, N. Fakotakis // Journal of Wavelet Theory and Applications. 2008. Vol. 2, N. 1. P. 41-59.
3. *Шмырев, Н. В.* Свободные речевые базы данных VoxForge.org // Сборник трудов международной конференции «Диалог 2008». 2008. С. 585-588.
4. *Rennie, J.* A short tutorial on using expectation-maximization with mixture models // The Internet [www.ai.mit.edu/people/jrennie/writing/mixtureEM.pdf], 2004.