УДК 004.522

РАСПОЗНАВАНИЯ РЕЧЕВЫХ СИГНАЛОВ

С. А. Запрягаев, А. Ю. Коновалов

Воронежский государственный университет

Поступила в редакцию 10.10.2009

Аннотация. В работе представлено описание расширений программной оболочки для анализа и распознавания речевых сигналов с использованием вейвлет-преобразования, представленной в работе [1]. Данная оболочка имеет целью создание инструмента для изучения различных моделей, алгоритмов и методов обработки данных, содержащихся в речевых сигналах, а также методов распознавания речи, для выявления применимости методов и алгоритмов к анализу и распознаванию речевых образов. В настоящей работе на основе системы распознавания, реализованной в программной оболочке, исследуются условия применимости дискретного вейвлет-преобразования в качестве инструмента при выделении акустических характеристик речевых сигналов.

Ключевые слова: Распознавания речи, вейвлет-преобразование, анализ речевых сигналов, сопоставление с образцом (DTW), векторное квантование.

Annotation. In the work the description of addition to the program complex for a speech signals analysis and recognition using the wavelet transform method, represented in [1], is considered. The goal of this complex is making of a tool for research of techniques of analysis and processing of a speech data, and also for research of speech recognition techniques, to conclude about applicability of this methods and algorithms to speech analysis and recognition In the work applicability of discrete wavelet transform to speech acoustic features extraction in speech recognition systems is investigated using a recognition system, constructed by the realized program complex.

Keywords: Speech recognition, wavelet transform, speech signals analysis, template matching (DTW), vector quantization.

ВВЕДЕНИЕ

Распознавание речи является задачей классификации образов акустических характеристик речевых сигналов. В системах распознавания речи выделяются два основных блока: — блок акустического анализа, предназначенный для выделения информативных акустических характеристик речевого сигнала и формирования акустического образа, сигнала как набора характеристик и — блок классификации путем сравнения с обученными акустическими моделями — эталонами.

В настоящей работе представлено описание дополнений программного комплекса (оболочки) для анализа и распознавания речи с использованием вейвлет-преобразования, представленного в [1]. Данная оболочка имеет целью формирование инструмента для изучения различных методов и алгоритмов анализа данных, содержащихся в речевых сигналах, а также методов распознавания речи. В работе используется дискретное вейвлет-преобразование для реализации блока акустического анализа речевых сигналов.

Вейвлет-преобразование является наиболее подходящим методом для анализа нестационарных сигналов, позволяющий получать многокомпонентный частотно-временной образ сигнала [4]. Поэтому для речевых сигналов, являющихся нестационарными, вейвлет-преобразование обеспечивает более точное представление, чем методы анализа, предполагающие стационарность сигнала.

Для общности подхода и для реализации сравнительного анализа в оболочке реализовано несколько стандартных методов спектрального анализа, включая методы Фурье преобразования. При этом построение энергетического спектра сигнала основано на дискретном преобразовании Фурье:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi nk/N}, \quad k = \overline{0, N-1}, \qquad (1)$$

© Запрягаев С. А., Коновалов А. Ю., 2009

где x_n, $n = \overline{0, N-1}$, — дискретный сигнал, N - период преобразования (или количество преобразуемых отсчетов сигнала). Коэффициенты дискретного преобразования Фурье вычисляются с помощью алгоритма быстрого преобразования Фурье (БПФ). Энергетический спектр Фурье отображает для k-й анализируемой частоты значение $|X_k|^2$ — квадрат модуля комплексного значения коэффициента X_k. При этом чаще используется логарифм этого значения $\log|X_k|^2$ — лог-спектр. В оболочке реализованы оба варианта отображения. Реальное значение анализируемой частоты вычисляется в соответствии с выражением

$$f_k = k \times F / N, \; k = \overline{0, N-1}. \qquad (2)$$

Для устранения явления просачивания спектральных составляющих в оболочке реализована возможность свертки анализируемого участка сигнала с оконной функцией Хэмминга [5].

На рис. 1 показан результат применения инструмента спектрального анализа Фурье к сегменту речевого сигнала, на который наложена оконная функция Хэмминга с коэффициентом 0.5 — лог-спектр Фурье.

В окне данного инструмента настраивается отображение графика анализируемого сигнала (рис. 1, блок 1, 2). При настройке можно задать оконную функцию Хэмминга с коэффициентом, определяющим влияние функции на исходный сигнал, и параметры быстрого преобразования Фурье: используемый алгоритм, наличие или отсутствие нулевого дополнения, а также необходимость выполнения расширения размера преобразования (рис. 1, блок 3). После вычисления коэффициентов спектр Фурье отображается на экране (с ранее установленными вариантами его построения) (рис. 1, блок 4).

Спектральный анализ на основе преобразования Фурье введен в настоящее программное приложение для проверки результатов применения методов вейвлет-анализа и сравнения анализа с фиксированным и изменяющимся окном. Кроме того, энергетический спектр используется в оболочке при вычислении кепстральных коэффициентов.

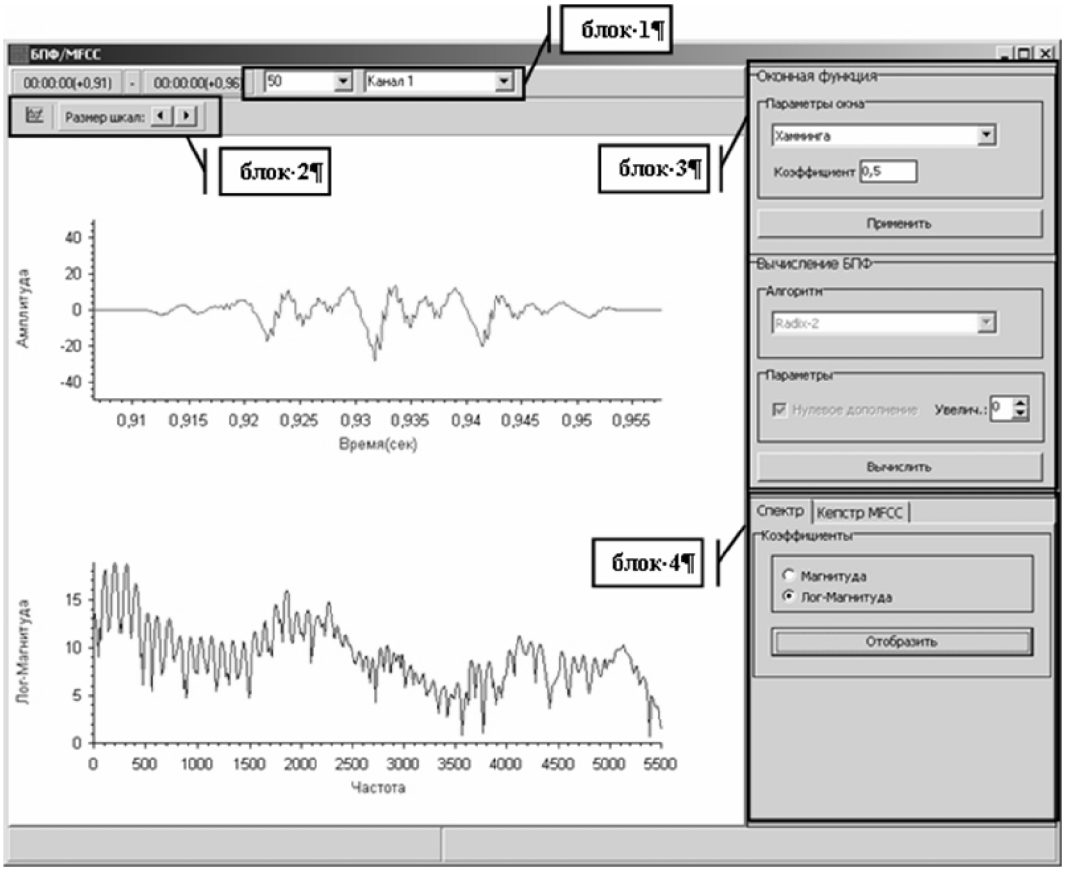

Рис. 1. Применение инструмента спектрального анализа Фурье к сегменту сигнала. Получен спектр Фурье

1. МЕЛ-ЧАСТОТНЫЙ КЕПСТРАЛЬНЫЙ АНАЛИЗ

Кепстральные коэффициенты являются результатом применения обратного преобразования Фурье к логарифмированному энергетическому спектру. В оболочке реализована возможность анализа сегментов сигнала через выделение кепстральных коэффициентов на мел-шкале (Mel Frequency Cepstral Coefficient — MFCC). Этот метод анализа основан на модели функционирования органов слуха человека и использует частотную шкалу мел, которая моделирует частотную чувствительность человеческого уха [5]. Мел-шкала линейная до 1 кГц и логарифмическая выше 1 кГц. MFCC-представление сигнала реализовано в оболочке как вещественный кепстр сегмента сигнала, выделенный с использованием быстрого преобразования Фурье, с отображением энергетического спектра на мел-шкалу. Отображение на мел-шкалу осуществляется с помощью блока треугольных фильтров (полосно-пропускающие фильтры), линейно распложенных на мел-шкале. Количество MFCC-коэффициентов определяется количеством фильтров в блоке треугольных фильтров.

На рис. 2 показан результат применения инструмента мел-частотного кепстрального анализа к сегменту речевого сигнала. График, отображающий значения MFCC-коэффициентов содержит по оси абсцисс — номер коэффициента, по оси ординат — значение. Ппри вычислении коэффициентов преобразования Фурье для последующего вычисления вещественного спектра использовалось наложение оконной функция Хэмминга с коэффициентом 0.5. Данный инструмент реализован совместно с инструментом спектрального анализа Фурье, поскольку также требует вычисления БПФ для расчета MFCC-коэффициентов. К описанному выше интерфейсу добавлена возможность вычисления и отображения MFCC-коэффициентов с указанием количества треугольных мел-фильтров и диапазона частот анализа (рис. 2, блок 1).

MFCC-коэффициенты (в основном младшего порядка [5]) широко используются в системах распознавания речи, так как этот метод

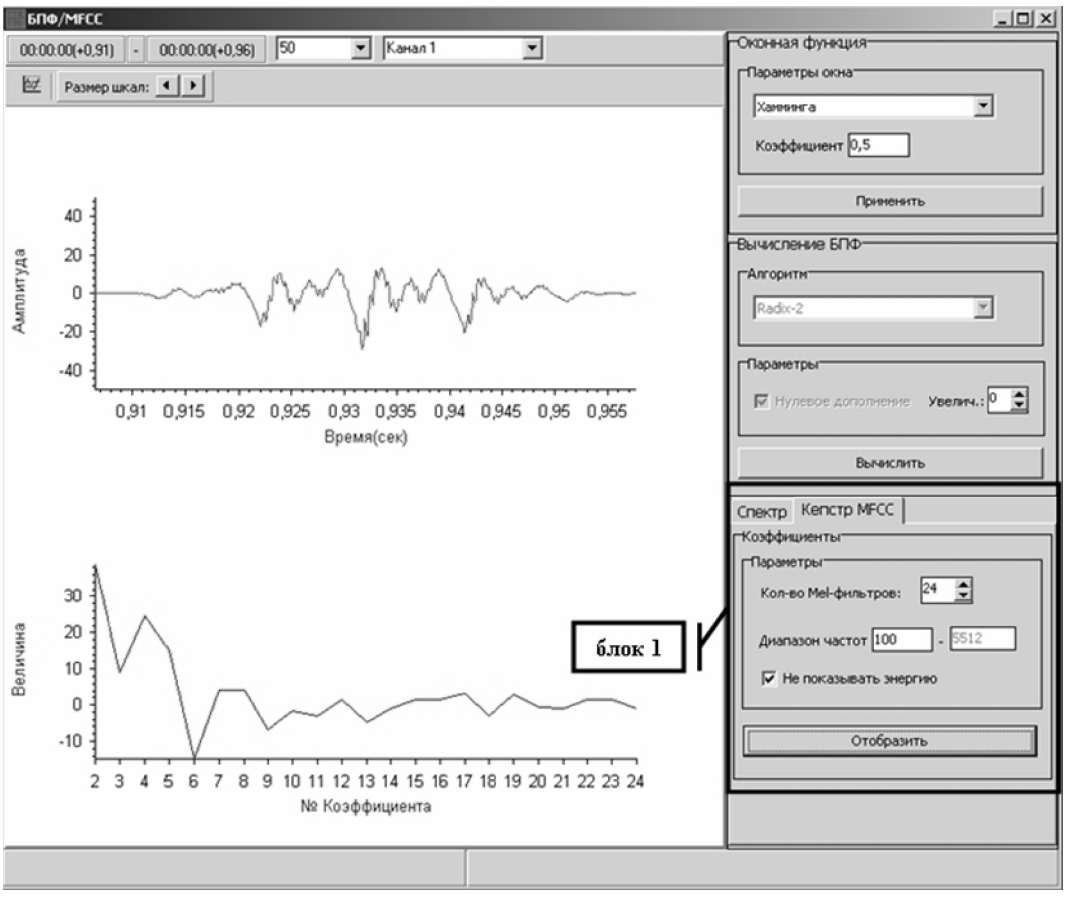

Рис. 2. Применение инструмента мел-частотного кепстрального анализа к сегменту сигнала. Получен график значений MFCC-коэффициентов

выделения акустических характеристик позволяет получить информацию об общем виде спектра и дает наиболее подходящие для классификации образы. Нулевой коэффициент содержит информацию о средней энергии спектра.

2. КРАТКОВРЕМЕННЫЙ АНАЛИЗ

В оболочке реализована обработка сегмента речи с помощью кратковременного (или оконного) анализа. Идея кратковременного анализа в следующем: окно анализа перемещается во времени с некоторым смещением, в каждом положении рассчитывается вектор характеристик для участка сигнала, попавшего в окно, последовательный набор векторов рассматривается как образ сигнала. Полученный образ сигнала отображается в оболочке с помощью цветовой поверхности, окраска которого вычисляется аналогично вейвлет-спектру в спектральном вейвлет-анализе, при этом возможно изменять границы максимума и минимума шкалы серого. Все точки со значениями характеристик выше установленного максимального значения окрашиваются в черный, меньше установленного минимального значение — в белый. Это повышает качество визуальной оценки образов сигналов. В настоящей оболочке для расчета вектора характеристик используются спектральный анализ на основе преобразования Фурье, кратномасштабный анализ на основе дискретного вейвлет-преобразования и мел-частотный кепстральный анализ описанные выше. Заметим, что в случае использования преобразования Фурье кратковременный анализ является оконным преобразованием Фурье. При использовании кратномасштабного анализа в качестве метода выделения вектора характеристик берутся максимальные по абсолютному значению вейвлет-коэффициенты на каждом уровне вейвлет-спектра для данного участка сигнала. На рис. 3 показан результат применения инструмента кратковременного анализа к сегменту сигнала — цветовая поверхность —

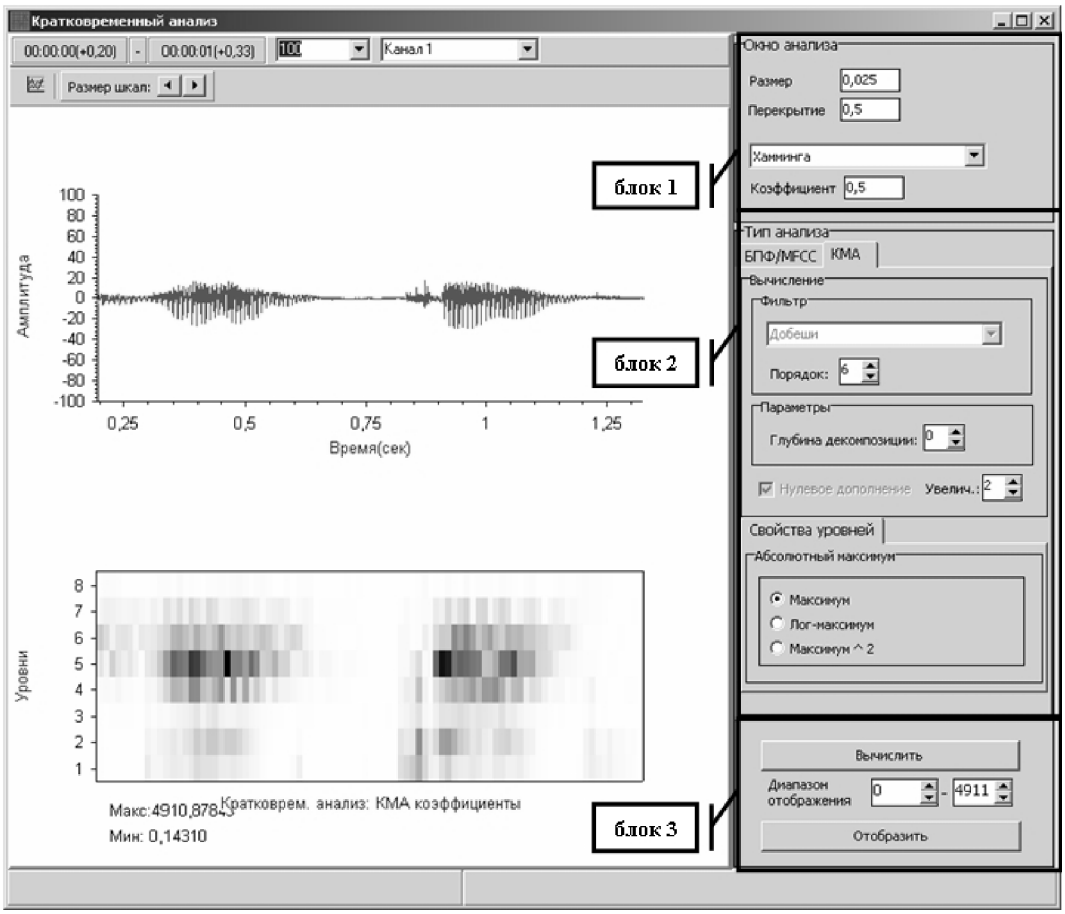

Рис. 3. Применение инструмента кратковременного анализа к сегменту сигнала. В качестве метода выделения характеристик используется кратномасштабный анализ. Получена цветовая поверхность — акустический образ сегмента сигнала

акустический образ сигнала. В окне данного инструмента можно осуществлять настройку окна анализа (рис. 3, блок 1), выбор метода выделения характеристик (рис. 3, блок 2), и установку его параметров (рис. 3, блок 3).

3. ИСПОЛЬЗУЕМЫЕ МЕТОДЫ КЛАССИФИКАЦИИ АКУСТИЧЕСКИХ ОБРАЗЦОВ РЕЧИ

Для классификации акустических образов речи в настоящей оболочке реализованы два метода: — нелинейное сопоставление с образцом и — векторное квантование.

НЕЛИНЕЙНОЕ СОПОСТАВЛЕНИЕ С ОБРАЗЦОМ

Речь является нелинейным во времени процессом. На формирование речи диктора, и в частности ее темпа, влияет множество факторов. Одно слово, произнесенное в разные моменты времени, обычно отличается по суммарной длительности и по длительности отдельных участков. Нелинейности речевых сигналов отражаются в соответствующих акустических образах при выделении характеристик. Для корректного сопоставления речевых образов в данной программной оболочке производится их выравнивание по длине. При этом линейное выравнивание не удовлетворительно из-за нелинейности самого процесса речи. Поэтому при построении систем распознавания, базирующихся на сопоставления с образцом, используется алгоритм динамического искажения временной шкалы (Dynamic Time Warping. DTW) [5]. Данный алгоритм обеспечивает нелинейное выравнивание сопоставляемых образов с оценкой затрат на это выравнивание, которое выражается в длине глобального пути выравнивания. Классификатор на основе DTW оценивает меру совпадение образа распознаваемого участка речи с эталонными образами по длине пути выравнивания. Эталонный образ, требующий меньших затрат на выравнивание с образом распознаваемого участка, выбирается в качестве образца для распознаваемого участка и определяет решение системы распознавания. Алгоритм DTW вычисляет матрицу размерности (M,N), где M — количество векторов характеристик образа распознаваемого сигнала и N — количество векторов эталона. Матричный элемент $D(i,j)$ — является оценкой глобального пути выравнивания до точки (i,j). Процесс вычисления описывается формулами:

$$\begin{gathered} D(i,j) = \min[D(i-1,j), D(i- \\ -1,j), D(i,j-1)] + d(i,j), \\ D(1,1) = d(1,1), \end{gathered} \qquad (3)$$

где $d(i,j)$ — локальная оценка в точке (i,j) — расстояние между i-м и j-м векторами образов. В оболочке реализовано DTW-сопоставление образов, использующее евклидово расстояние в качестве локальной оценки. Алгоритм DTW построен на принципе динамического программирования. Решение задачи в точке (i,j), то есть оценка глобального пути, определяется минимальным расстоянием из точек $(i-1,j)$, $(i-1,j-1)$ и $(i,j-1)$. После завершения вычисления матрицы элемент $D(M,N)$ содержит оценку глобального пути выравнивания сегментов.

Классификатор на основе DTW обеспечивает построение дикторозависмых систем распознавания речи с небольшими базами слов. Точность распознавания может быть повышена, если использовать составной эталонный образ, сформированный из нескольких одиночных эталонов слова, что позволяет в некоторой степени повысить устойчивость метода к изменениям в произношении слова, не связанным только лишь с нелинейностью речи. В качестве оценки составного эталона в таком случае берется минимальная оценка по всем включенным эталонам. Однако такой подход трудно реализуем для систем с большими словарями, так как приводит к необходимости обучать и хранить большое количество эталонов или искать методы усреднения эталонов, количество которых сказывается и на производительности метода.

В оболочке реализован инструмент для DTW-сопоставления сегментов сигналов. Для вычисления образов сигналов используется кратковременный анализ с ранее перечисленными возможностями. На рис. 4 представлено окно данного инструмента. Программа позволяет загрузить 2 сигнала из файлов WAV-формата (рис. 4, блок 1) и рассчитать для них DTW-матрицу с отслеживанием глобального пути выравнивания. В результате сопоставления в окне отображаются образы сигналов (цветовые поверхности), полученные выбранным методом (рис. 4, блок 2), глобальный путь выравнивания (ломаная линия) и его оценка.

Данный инструмент предназначен для визуального изучения результатов работы алго-

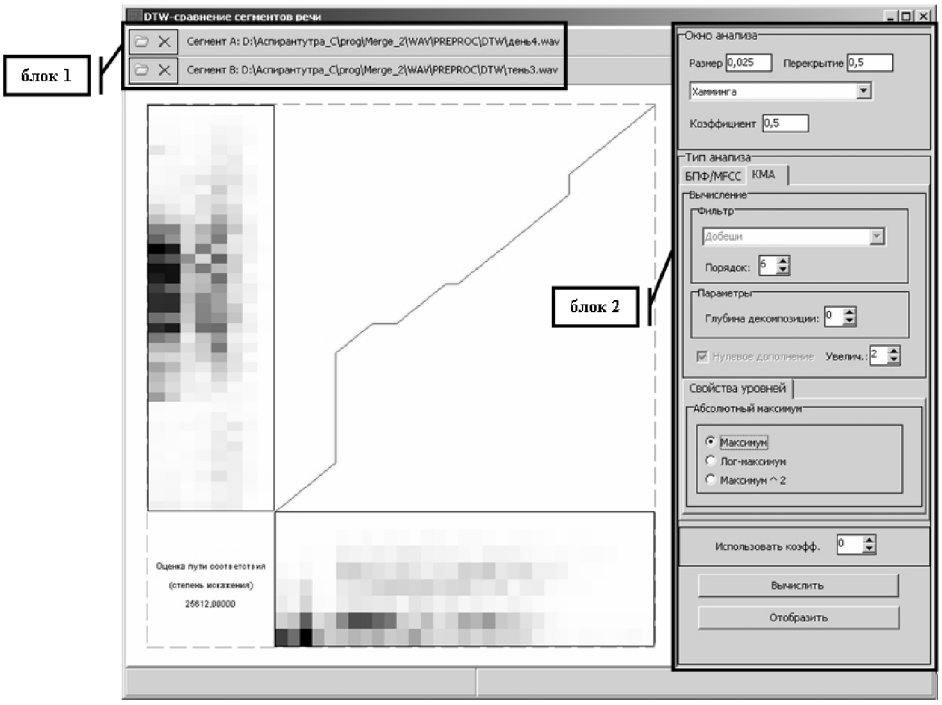

Рис. 4. Применение инструмента DTW-сопоставления сегментов речи. В качестве метода выделения характеристик используется кратномасштабный анализ. Получены образы сигналов, глобальный путь выравнивания и оценка

ритма DTW на образах, полученных разными методами выделения акустических характеристик.

DTW-сопоставление используется в реализованном модуле распознавания речи как один из вариантов классификатора для блока классификации.

ВЕКТОРНОЕ КВАНТОВАНИЕ

Векторное квантование является одним из наиболее эффективных методов кодирования исходной информации для передачи по каналам связи, обеспечивающей минимальное представление исходной информации перед передачей и наиболее точное (с минимальным искажением) восстановление исходной информации. Квантование — процесс аппроксимации сигналов дискретными символами. Если входными параметрами являются векторы значений, то квантование векторное или многопараметрическое. В аспекте построения классификатора для системы распознавания метод векторного квантования применяется для конструирования аппроксимирующих акустических прототипов, используемых в качестве акустических моделей-эталонов. Метод основывается на построения кодовой книги — наборе эталонных векторов — кодовых слов. В нашем случае эти кодовые слова — векторы акустических характеристик, аппроксимированных для лучшего представления классифицируемого элемента речи. В реализуемой оболочке обучение системы распознавания, построенной на кодовых книгах, базируется на алгоритмах LBG [5] (для получения начальной книги) и K-means (K-средних) (для уточнения на последующих обучающих данных) [5]. Процесс квантизации состоит в следующем. Входящий вектор сравнивается с каждым кодовым словом с помощью некоторой меры искажения (в настоящей оболочке применяется евклидово расстояние). Кодовое слово с наименьшим искажением выбирается как эталон для входящего вектора и заменяется уникальным символом данного кодового слова. Следовательно, прошедший такую обработку акустический образ, представляется как последовательность символов, при этом может быть рассчитано общее искажение. При распознавании кодовая книга, дающая наименьшее общее искажение для распознаваемого образа, определяет распознаваемый элемент речи. Метод не учитывает последовательность векторов признаков в образе и временное выравнивание не требуется. Поскольку в самом этом методе заложено усреднение акустических

характеристик, он более подходит для построения систем распознавания с большими словарями для построения дикторонезависимых систем, но это в свою очередь требует большего объема обучающих данных.

Данный метод используется в реализованном модуле распознавания речи как один из вариантов классификатора для блока классификации.

4. МОДУЛЬ РАСПОЗНАВАНИЯ РЕЧИ

В реализованной оболочке модуль распознавания речи определен как система блоков, описанных выше. Работа блока акустического анализа базируется на кратковременном анализе. Возможна настройка окна анализа (размер, перекрытие, использование функции окна), метода выделения характеристик сигнала и количества коэффициентов вектора характеристик, используемых при построении образа. Перед подачей на блок классификации (как при распознавании так и при обучении, тестировании) сигналы требуют предварительной обработки с помощью редактора звуковых файлов, реализованного в оболочке [1]: акцентируются высокие частоты, определяются границы распознаваемого участка. Для блока классификации может быть выбран любой из выше описанных классификаторов. При этом возможна настройка специфичных параметров классификатора. В целом, модуль построен как дикторозависимая система распознавания изолированных слов, работающая в off-line режиме. Речевые сигналы для распознавания подаются в файлах формата WAV. После построения образа сигнала и классификации модуль сообщает индекс класса (специальная метка элемента речи), показавшего лучшую оценку для данного сигнала. В модуле реализованы инструменты обучения и тестирования систем. Модуль позволяет производить пакетный (массовый) процесс обучения и тестирования, загружая пакеты обучающих или тестирующих данных — набор файлов формата WAV, размещенных в каталоге особой структуры на жестком диске компьютера. Инструмент обучения информирует о ходе процесса обучения и по завершении сообщает количество обученных эталонов для системы. После обучения проект системы распознавания можно сохранить на жесткий диск компьютера и впоследствии загружать в модуль для тестирования и распознавания. В проекте системы помимо акустических моделей сохраняется конфигурация системы: параметры блока акустического анализа и блока классификации. После загрузки системы для последующего использования модуль настраивает систему распознавания по этим параметрам, чтобы обеспечить соответствие входящих образов и обученных данных. Инструмент тестирования информирует о ходе тестирования и по завершении сообщает о точности распознавания системы на тестовых данных. Кроме того, доступна информация об ошибках распознавания: файл, на котором произошла некорректная классификация, некорректный индекс классификации, общая ошибка для текущего тестирующего набора элементов речи одного класса.

На рис. 5, 6, 7 представлены окна соответствующие различным режимам использования модуля распознавания, блок классификации которого реализован через DTW; обучение, тестирование и распознавание соответственно; система конфигурируется для распознавания командных слов.

5. ПОСТРОЕНИЕ СИСТЕМ РАСПОЗНАВАНИЯ РЕЧИ

С помощью реализованной оболочки и модуля распознавания в его составе осуществлен эксперимент по построению двух дикторозависимых систем распознавания изолированных слов на основе дискретного вейвлет-преобразования с использованием метода нелинейного сопоставления с образцом и метода векторного квантования.

БАЗА СЛОВ

В эксперименте использовалась следующие наборы слов:

– набор букв английского алфавита {b, c, d, e, g, p, t, v, z};

– цифры {один, два, три, четыре, пять, шесть, семь, восемь, девять};

– набор общеупотребимых командных слов {открыть, закрыть, пуск, выполнить, да, нет, отмена}.

Во время тестирования 6 различных вариантов произношения слова использовалось для обучения модели-эталона для слова. 14 различных вариантов произношения слова, независимых от обучающих вариантов, использовались для тестирования системы и определения точности распознавания. Все варианты записывались и редактировались с помощью редактора

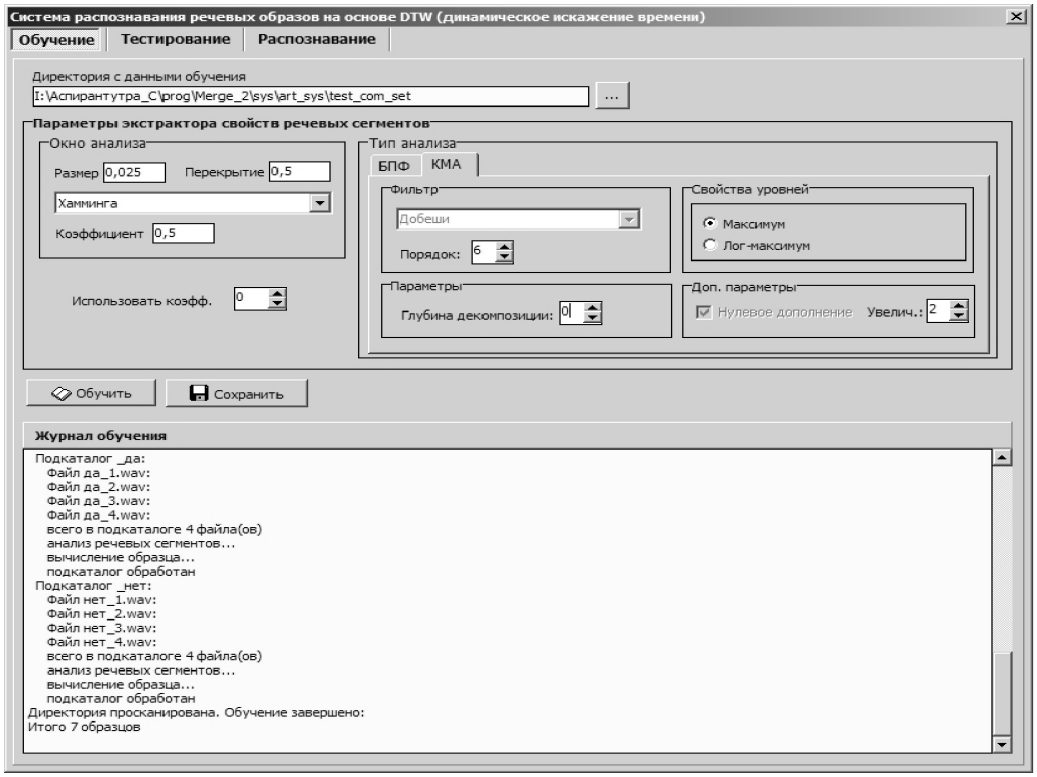

Рис. 5. Окно модуля распознавания речи: режим обучения, тип классификатора — DTW. В журнале обучения представлена информация о процессе обучения системы. Система обучается на наборе команд слов

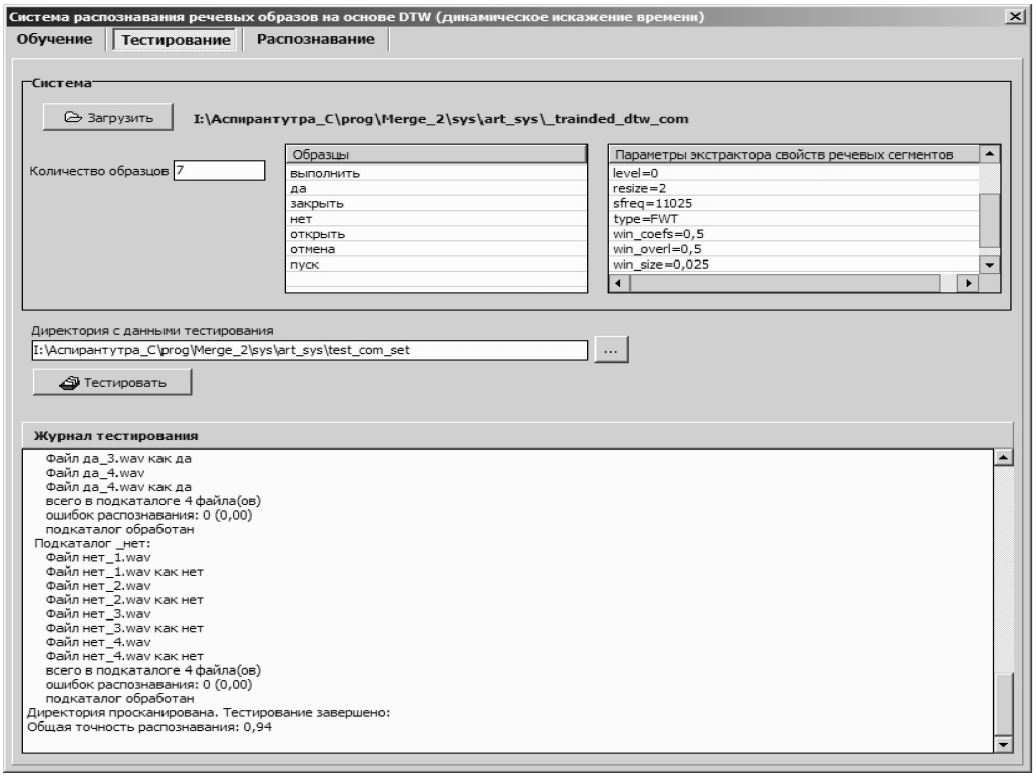

Рис. 6. Окно модуля распознавания речи: режим тестирования, тип классификатора — DTW. В журнале тестирования представлена информация о процессе тестирования системы. Система тестируется на независимом наборе командных слов. Точность распознавания 0,94

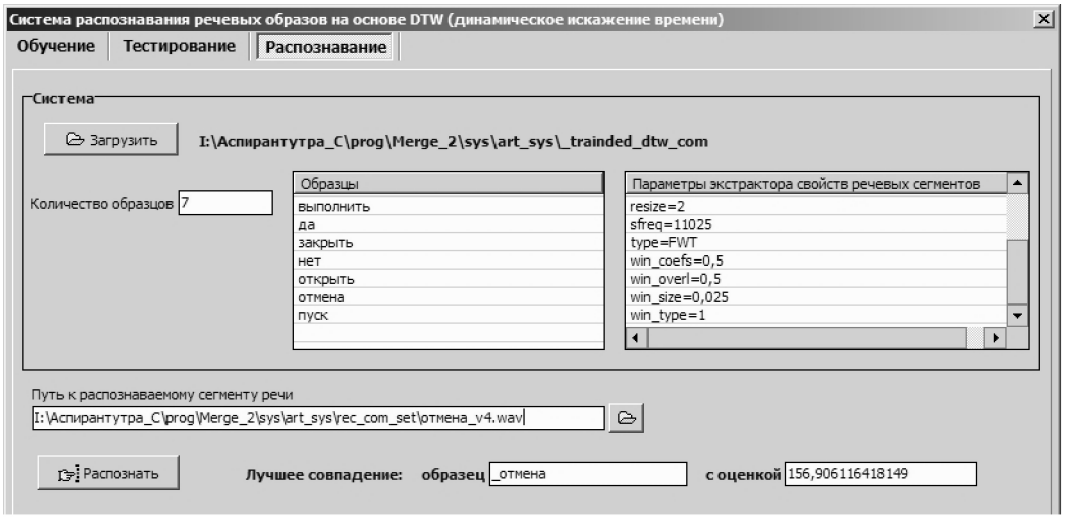

Рис. 7. Окно модуля распознавания речи: режим распознавания, тип классификатора — DTW. Система корректно распознала сегмент речи «Отмена» из набора командных слов

звуковых файлов. Редактирование состояло в акцентировании высоких частот и определении границ слов. Был выбран следующий формат сигналов: частота дискретизации 22050 Гц, 16 бит, что дает достаточное количество информации для акустического анализа [5].

КОНФИГУРАЦИЯ СИСТЕМ

Блок акустического анализа строился на кратковременном анализе сигнала, использующем дискретное вейвлет-преобразование для выделения акустических характеристик. В качестве анализирующих вейвлетов использовались вейвлеты Добеши 4-го порядка. Окно и метод анализа конфигурировались для получения 8 уровней вейвлет-декомпозиции сигнала. На каждом уровне выбирался один максимальный по абсолютному значению элемент как характеристика деталей сигнала на данном уровне. Таким образом, для каждого окна анализа вычисляется 8 коэффициентов. В блоке классификации применялись методы нелинейного сопоставления с образцом (DTW) и векторного квантования. Для обучения системы, основанной на DTW, был выбран следующий способ формирования эталонного образа. Попарно сопоставлялись (на основе алгоритма DTW) все варианты одного слова и 2 наиболее схожих образа выбирались в качестве эталона. При распознавании оценка для составного эталона считается как минимум оценки по всем включенным эталонам. Для метода векторного квантования размер кодовой книги был выбран 8 для построения моделей (для всех наборов слов).

РЕЗУЛЬТАТЫ

Оценки точности распознавания сконфигурированных систем на различных наборах представлены в табл.

Таблица

Оценка точности распознавания на трех наборах слов

Классификация / Набор	DTW	Векторное квантование
Набор букв латинского алфавита	76%	85%
Цифры	89%	70%
Командные слова	94%	75%

В соответствии с представленными в табл. результатами можно сделать заключение об удовлетворительном уровне распознавания методами реализованными в данной программной оболочке.

СПИСОК ЛИТЕРАТУРЫ

1. *Коновалов А. Ю.* Программный комплекс для анализа и распознавания речевых сигналов с применением вейвлет-преобразования / А. Ю. Коновалов, С. А. Запрягаев // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. — 2009. — № 1. — С. 199—107.

2. *Короновский А. А., Храмов А. Е.* Непрерывный вейвлетный-анализ и его приложения. — М.: ФИЗМАТЛИТ, 2003. — 176 с.

3. *Добеши И.* Десять лекций по вейвлетам / И. Добеши; пер. с английского. — Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001. — 464 с.

4. *Астафьева Н. М.* Вейвлет-анализ: Основы теории и примеры применения / Н. М. Астафьева. — М.: Успехи физических наук, 1996, Т. 166. — № 11. — С. 1145—1170.

5. *X. Huang, A. Acero, H. Hon.* Spoken language processing: a guide to theory, algorithm, and system development. — Prentice Hall PTR, 2001. — P. 936.

Запрягаев Сергей Александрович — д. ф.-м.н, проф. каф. цифровых технологий Воронежского государственного университета. Тел.(4732) 208-257. E-mail: zsa@main.vsu.ru

Коновалов Алексей Юрьевич — аспирант кафедры цифровых технологий Воронежского государственного университета E-mail: konovalov.alekse@mail.ru. Тел. 8 960 109 50 88

Zapryagaev S. A. — Doctor of Physics-math. Sciences, Professor of the dept. of digital technologies Voronezh State University. Tel.(4732) 208-257. E-mail:zsa@main.vsu.ru

Konovalov A. Yu. — Post-graduate student of the dept. of digital technologies Voronezh State University. E-mail: konovalov.alekse@mail.ru; Tel. 8 960 109 50 88