

Учреждение образования
**«БЕЛОУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИНФОРМАТИКИ И
РАДИОЭЛЕКТРОНИКИ»**

УДК 681.327.12

РАКУШ
Валентин Владимирович

**Методы и алгоритмы стохастической обработки речи в системах
текстонезависимой верификации диктора**

Специальность 05.13.15 – Вычислительные машины и системы

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Минск – 2004

Работа выполнена в учреждении образования «Белорусский государственный университет информатики и радиоэлектроники»

Научный руководитель –
доктор технических наук, профессор Садыхов Р.Х.,
кафедра Электронных вычислительных машин, БГУИР

Официальные оппоненты:
доктор технических наук, профессор Петровский А.А.,
кафедра Электронных вычислительных средств, БГУИР

кандидат технических наук, старший научный сотрудник Левковская Т.В.,
Объединенный институт проблем информатики НАНБ

Оппонирующая организация –
Научно-исследовательское учреждение «Институт прикладных физических проблем
им А.М. Севченко» Белгосуниверситета.

Защита состоится 22 декабря 2004 г. в 14 часов на заседании совета по защите
диссертаций Д 02.15.04 при учреждении образования «Белорусский государственный
университет информатики и радиоэлектроники» по адресу: 220013, Минск, ул.
П. Бровки, 6, корп. 1, ауд. 232, тел. 2398989.

С диссертацией можно ознакомиться в библиотеке Учреждения образования
«Белорусский государственный университет информатики и радиоэлектроники».

Автореферат разослан «19» ноября 2004 г.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследований. Успехи в развитии вычислительной техники, в особенности аппаратного обеспечения процессоров цифровой обработки сигналов, развитии мультимедиа-приложений в быту и в коммерческих структурах создает экономические предпосылки для построения быстродействующих систем верификации диктора. Круг пользователей персональных компьютеров, сотовых телефонов, сетевых приложений расширяется. Появляются новые области применения, такие, как верификация и идентификация диктора для защиты от террористов и обеспечения безопасности. Для решения задачи верификации личности человека по голосу до недавнего времени требовалось приложить немало усилий специалистов из области экспертизы, что тормозило широкое распространение методов верификации. Математическое и алгоритмическое обеспечение, разработанное в последние десятилетия, позволяет реализовать полностью автоматические верификаторы. Причем число областей применения верификаторов становится с каждым годом все больше.

В настоящее время разработан большой набор методов и алгоритмов для распознавания диктора по голосу. Так существует уже ряд систем, показывающих высокие результаты в распознавании диктора. Однако, также существует и широкий круг задач, которые невозможно решать, применяя одни и те же алгоритмы. Каждая специфическая задача требует уникального подхода. Область текстонезависимой верификации диктора пересекается с другими областями распознавания диктора по голосу и в настоящее время имеет некоторые слабоисследованные аспекты. Более качественное исследование первичного анализа речи именно для задачи распознавания диктора (в частности, верификации) может принести значительное увеличение точности распознавания системы в целом. Хотя первичный анализ хорошо исследован для построения систем распознавания речи, задача распознавания голоса говорящего имеет свои особенности и требует отдельного подхода.

Речь человека представляет собой квази-стационарный стохастический процесс, который сложно описать математической моделью и решить аналитически. Построение стохастических моделей позволяет более точно описать речь на основании статистики, собранной на продолжительной записи голоса диктора. Стохастическую модель можно рассматривать как способ исследования речеобразования и различий между голосами дикторов. К настоящему времени достаточно полно исследованы методы стохастического моделирования речи в целом и диктора в частности. Однако существуют возможности повышения эффективности и производительности систем верификации. Выбор и построение моделей определяет текстозависимость или текстонезависимость системы верификации. Текстонезависимая система имеет более широкий круг применения, использует общие подходы и тем самым является более интересной для исследователей. Основными методами моделирования речи диктора для текстонезависимой верификации являются методы скрытых Марковских моделей, модели гауссовых смесей, векторное квантование, нейронные сети. Из этого списка векторное квантование является самым малоисследованным методом. Однако, применение алгоритмов векторного квантования дает возможность значительно

повысить скорость работы систем верификации. Также алгоритмы векторного квантования демонстрируют хорошие результаты совместно с другими стохастическими методами моделирования.

В связи с развитием научных исследований в области обработки и распознавания речи и диктора существенным является использование универсальных методов оценки качества работы систем, создание единых общедоступных речевых баз, на которых можно проводить эксперименты. Такой подход позволит сравнивать системы, разработанные различными авторами. Однако существует необходимость в исследовании уже существующих методов оценки производительности систем верификации и выборе наиболее универсальных, подходящих как можно для большего числа систем.

Связь работы с крупными научными программами и темами. Диссертационная работа выполнена на кафедре ЭВМ учреждения образования «Белорусский государственный университет информатики и радиоэлектроники» в рамках государственной научно-технической программы «Ахова» по теме «Разработать экспериментальные стенды для распознавания личности по голосу» (1997-1998 гг.); в рамках Государственной научно-технической программы «Информатика» по теме «Разработать систему аналитического и имитационного моделирования управляющих вычислительных комплексов реального времени» (1998-1999 гг.); а также Государственной программы ориентированных фундаментальных исследований на 2003-2005 года по теме ИНФОТЕХ 08 «Разработка методов и алгоритмов обработки изображений и идентификации объектов в системах технического зрения».

Цель и задачи исследования – разработать стохастические методы и алгоритмы обработки речевого сигнала и распознавания образов применительно к верификации диктора по произвольной речи.

Для достижения поставленной цели необходимо решить следующие задачи:

- выполнить аналитический обзор существующих методов и систем распознавания диктора по произвольной речи и выделить наиболее перспективные направления;
- разработать алгоритмы первичной обработки речи в системах верификации диктора;
- исследовать и создать алгоритмы выделения признаков для построения систем верификации диктора на основе стохастических методов моделирования речевого сигнала;
- разработать надежные методы и алгоритмы построения стохастических моделей голосов дикторов для систем верификации;
- исследовать и практически реализовать указанные выше алгоритмы в конкретных системах верификации диктора по произвольной речи.

Объект и предмет исследования. В данной диссертационной работе объектом исследования является речь дикторов, оцифрованная и записанная с телефонного канала. Предметом исследования являются методы обработки речевого сигнала и выделения первичных признаков, а также стохастические методы моделирования диктора по голосу.

Методология и методы проведенного исследования. В качестве аппарата исследования применены психоакустические принципы восприятия речи человеком, стохастические методы моделирования речевого сигнала, векторное квантование и модели гауссовых смесей. Для практической реализации разработанных алгоритмов использовались технология создания программных систем, методы оптимизации программных комплексов, структурный и модульный подход к технологии проектирования программного обеспечения.

Научная новизна и значимость полученных результатов.

В качестве основных научных результатов, полученных в диссертационной работе, можно отметить следующее:

- Разработан новый способ формирования векторов параметров речевого сигнала, позволяющий увеличить эффективность верификации всей системы; представлен метод моделирования параметров голоса диктора при помощи весовых функций.
- Создан метод выделения субречевых единиц на основе алгоритмов векторного квантования, ориентированный на повышение скорости вычислений и возможность нормализации двух различных кодовых книг.
- Разработан способ инициализации модели гауссовых смесей на основе алгоритма векторного квантования, позволяющий повысить точность верификации и скорость сходимости алгоритма обучения моделей гауссовых смесей.
- Разработаны алгоритмы нормализации оценки интегрального шума квантования кодовой книги, позволяющие объединять несколько методов классификации для повышения точности верификации.
- Разработана процедура оценки качества алгоритмов верификации диктора, позволяющая оптимально подобрать набор указанных алгоритмов, первичных признаков и конфигурацию всей системы в целом.

Практическая значимость полученных результатов. Рассмотренные в диссертационной работе методы предварительной обработки речевого сигнала и методы стохастического моделирования векторов параметров используются в созданной системе текстонезависимой верификации и идентификации диктора. Результаты диссертационной работы внедрены в НИИ ЭВМ, в малом предприятии ООО «СЕНСОР-М», а также в учебном процессе в части, касающейся курса лекций по дисциплине «Цифровая обработка сигналов и изображений» на кафедре ЭВМ БГУИР.

Основные положения, выносимые на защиту

1. Способ формирования вектора первичных признаков речевого сигнала с использованием функций весов, аргументом которых является число элементов вектора параметров, предназначенные для нормализации и психоакустической корректировки параметров.
2. Метод выделения субречевых единиц с использованием алгоритмов векторного квантования для определения порога значения интегрального шума кодовой книги и применение этого порога для нормализации значений шумов тестовой и эталонной книг.

3. Способ инициализации модели гауссовых смесей на основе алгоритма векторного квантования для увеличения эффективности распознавания систем верификации, построенных на базе модели гауссовых смесей (МГС).
4. Алгоритмы нормализации значений интегральных шумов двух кодовых книг, построенных при помощи различных статистических данных.
5. Система автоматической верификации личности по произвольной речи, записанной на фонограмме, позволяющая объединить стохастическое моделирование речи при помощи модели гауссовых смесей и алгоритмов векторного квантования.

Личный вклад соискателя. Все предложенные методы и алгоритмы были разработаны и программно реализованы лично автором. Научный руководитель принимал участие в постановке задач, определении возможных путей решения.

Апробация результатов диссертации. Результаты диссертационной работы докладывались и обсуждались на X научно-технической конференции «Новые технологии в машиностроении и вычислительной технике» (Брест, 1998), II Международной научно-технической конференции «Современные методы цифровой обработки сигналов в системах измерения, контроля, диагностики и управления» (Минск, 1998), The 5th International Conference “Pattern Recognition and Information Processing” (Minsk, 1999), The 3rd International Conference on Neural Networks and Artificial Intelligence (Minsk, 2003), 1st International Conference on Informatics in Control, Automation and Robotics (Setubal – Portugal, 2004).

Опубликованность результатов. По материалам выполненных исследований опубликовано 10 научных работ, в том числе 5 статей, 5 докладов на международных и республиканских конференциях. Общее количество страниц публикаций: 54.

Структура и объем диссертации. Диссертация изложена на 136 страницах машинописного текста, содержит 32 рисунка (22 страницы), 11 таблиц (5 страниц), и состоит из введения, общей характеристики работы, четырех глав, заключения, списка использованных источников на 8 страницах, включающего 113 названий, и 2 приложений на 26 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении показана актуальность темы диссертационной работы, формируются цель и задачи исследования, приведена краткая характеристика работы.

В главе 1 исследуется задача распознавания диктора по произвольной речи, которая делится на две смежных подзадачи: идентификация и верификация диктора. Приводится общая структура системы верификации диктора, которая состоит из блока формирования первичных признаков, блока оценки параметров стохастических моделей, блоков памяти моделей и классификации. Описывается функционирование отдельных блоков системы.

Рассматриваются типы и способы формирования первичных признаков речевого сигнала в системах верификации диктора. Определяется, что параметры, используемые для распознавания диктора, должны удовлетворять следующим требованиям:

- быть различными для голосов разных дикторов и в то же время быть сходными для голоса одного и того же диктора;
- быть легко выделяемыми из речевого сигнала;
- не изменяться во времени;
- быть устойчивыми к подражанию.

Описываются достоинства и недостатки различных способов моделирования речевого сигнала, которые включают в себя метод представления речевого сигнала при помощи последовательности векторов признаков, метод моделирования при помощи скрытых марковских моделей, алгоритмы векторного квантования, модели гауссовых смесей.

Дается описание существующих речевых баз, используемых для исследования эффективности систем верификации, приводятся их технические характеристики. Также приводится методология оценки эффективности.

Приводится краткая характеристика систем распознавания и верификации диктора по голосу, их характеристики, проводится анализ используемых методов первичной обработки и моделирования речевого сигнала.

Во второй главе определяются основные источники искажения речевого сигнала и рассматриваются методы и алгоритмы первичного анализа речевого сигнала. На рис. 1 изображена структурная схема подсистемы выделения первичных признаков. Эта подсистема является общей для систем верификации диктора и может состоять из различных блоков. В большинстве систем, проанализированных в первой главе, используются кепстральные коэффициенты, вычисленные на основании коэффициентов линейного предсказания.

Для повышения эффективности системы верификации очень важно на этапе выделения первичных признаков не потерять информации, которую невозможно восстановить на последующих этапах моделирования. Поэтому в блок формирования вектора параметров включены шаги вычисления не только кепстральных коэффициентов, но и их временных производных, (рис. 2).

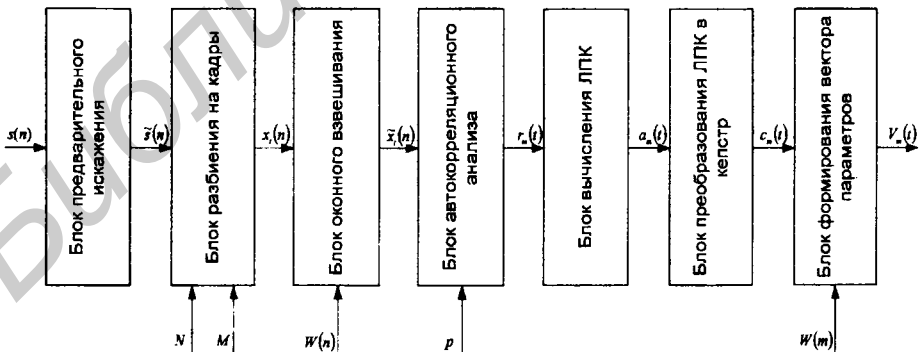


Рис. 1. Структура подсистемы формирования вектора первичных признаков

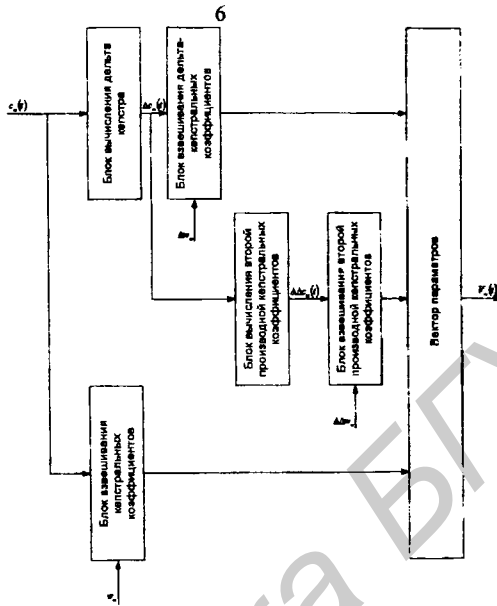


Рис. 2. Структура блока формирования вектора параметров

Показано, что при формировании вектора первичных параметров, который включает кепстральные коэффициенты, а также первую и вторую производные кепстральных коэффициентов, очень важно учитывать тот факт, что не все коэффициенты имеют одинаковый «вес» при стохастическом моделировании. Для «выравнивания» значимости всех коэффициентов предложено использовать следующие весовые функции отдельно для кепстральных, дельта-кепстральных и дельта-дельта-кепстральных коэффициентов:

$$w_m = 1 + 9 \sin\left(\frac{\pi m}{72}\right), \quad 1 \leq m \leq 36 \quad (1)$$

$$w_m = 2,5 \left(1 + \sin\left(\frac{\pi m}{72}\right) \right), \quad 1 \leq m \leq 36 \quad (2)$$

$$w_m = 12,5 \left(1 + 13 \sin\left(\frac{\pi m}{72}\right) \right), \quad 1 \leq m \leq 36 \quad (3)$$

При определении весовой функции для параметров временных производных необходимо учитывать выражения дифференцирования кепстральных коэффициентов

$$\Delta \hat{c}_m(n) \approx \frac{d}{dt} \hat{c}_m(n) \approx \hat{c}_m(n) - \hat{c}_m(n-1) \quad (4)$$

$$\Delta \hat{c}_m(n) \approx \frac{d}{dt} \hat{c}_m(n) \approx \hat{c}_m(n+1) - \hat{c}_m(n) \quad (5)$$

$$\Delta \hat{c}_m(n) \approx \frac{d}{dt} \hat{c}_m(n) \approx \sum_{p=-N}^N p \hat{c}_m(n+p), \quad (6)$$

из которых следует, что кепстральные весовые коэффициенты являются весовыми коэффициентами и для производных параметров

$$\sum_{p=-N}^N p \hat{x}_m^*(n+p) = \sum_{p=-N}^N p w_m x_m(n+p) = w_m \sum_{p=-N}^N p x_m(n+p) = w_m \Delta x_m \quad (7)$$

Выражение (7) учитывается в блоке формирования вектора параметров, как изображено на рис. 2.

Была выдвинута гипотеза, что при увеличении числа коэффициентов эффективность системы должна расти. Эта гипотеза была подтверждена экспериментально. Так, результаты экспериментов для различного числа коэффициентов вектора параметров приведены в таблице 1. Использование большого числа коэффициентов в векторе параметров приводит к уменьшению производительности системы, поэтому требуется находить компромисс между числом параметров, и требуемым временем реакции системы.

Таблица 1.

	Число кепстральных параметров					
	22			36		
	%	EER	C _{det}	%	EER	C _{det}
Взвешивание кепстральных коэффициентов	96,6	0,433	0,214	97,7	0,433	0,098
Отдельное взвешивание производных параметров	98,7	0,217	0,19	100	0,252	0,095

EER – частота ожидания ошибки, которая представляет собой сумму величин ошибок первого и второго рода;

C_{det} – коэффициент выявленной ошибки детектирования.

Полагая, что речевой сигнал и спектральная характеристика канала могут быть представлены при помощи модели линейного предсказания с достаточно малой погрешностью, можно утверждать, что влияние канала на речевой сигнал приводит к аддитивной составляющей в кепстре результирующего речевого сигнала. Далее, полагая, что математическое ожидание кепстральных коэффициентов неискаженной речи равно нулю, получаем, что оценка спектральной характеристики канала равна математическому ожиданию речи, искаженной каналом. Таким образом, компенсация влияния канала на речевой сигнал производится по формуле

$$c_{\text{гов}}(n) = c_p(n) - E[c_p(n)], \quad (8)$$

где $E[\bullet]$ – символ математического ожидания.

В таблицах 2, 3 приведены результаты экспериментов, демонстрирующих зависимость эффективности системы верификации диктора по произвольной речи от типа и числа использованных параметров.

Таблица 2.

	Число параметров					
	22			36		
	%	EER	C_{det}	%	EER	C_{det}
LPCC	87,6	0,357	0,223	91,0	0,489	0,214
LPCC+ Δ	88,7	0,357	0,218	91,0	0,496	0,362
LPCC+ Δ + Δ	88,7	0,362	0,198	92,1	0,511	0,406
LPCC+ Δ + Δ +MCN	98,9	0,154	0,102	100	0,252	0,095

Проведенные эксперименты показывают, что эффективность распознавания возрастает с ростом числа кепстральных параметров, участвующих в распознавании. Также добавление первой и второй производной кепстральных параметров улучшает качество системы. Для систем верификации, функционирующих с сигналом, зашумленным аддитивным шумом, использование техники вычитания среднего кепстра повышает достоверность распознавания.

Таблица 3.

	Число параметров, 48		
	%	EER	C_{det}
LPCC	98,9	0,334	0,111
LPCC+ Δ	100	0,219	0,070
LPCC+ Δ + Δ	100	0,207	0,046
LPCC+ Δ + Δ +MCN	100	0,171	0,037

LPCC – кепстральные коэффициенты, вычисленные из коэффициентов линейного предсказания;

MCN – нормализация среднего кепстра;

Δ - первая временная производная параметров;

$\Delta + \Delta$ - первая и вторая временные производные параметров.

В третьей главе рассматриваются стохастические методы моделирования речевого сигнала в системах текстонезависимой верификации диктора по произвольной речи. Рассматривается ряд алгоритмов и методов выделения признаков, в том числе алгоритмы векторного квантования и моделей гауссовых смесей, применявшиеся в экспериментах. Рассматриваются различные классификаторы, являющиеся модификациями алгоритма векторного квантования.

В рамках рассмотрения алгоритмов векторного квантования и моделей гауссовых смесей особое внимание уделяется моделированию голоса диктора при

помощи выделения субречевых единиц. Основным является предположение, что кластера кодовой книги представляют собой некоторые звуки, составляющие произвольную фразу, которую произносит диктор. Таким образом, кодовая книга, построенная на основании векторов, полученных при первичном анализе речи, является моделью голоса диктора и в дальнейшем может использоваться для верификации в качестве классификатора. Критерием принятия решения является интегральный шум квантования тестовой фразы.

В случае применения моделей гауссовых смесей можно сделать аналогичное предположение о моделировании субречевых единиц речи диктора. Так как модель гауссовых смесей состоит из нескольких гауссовых распределений, то можно считать, что одно гауссово распределение моделирует один звук. Таким образом, построенная на речи диктора модель является классификатором голоса этого диктора, а критерием принятия решения является вероятность, вычисленная для тестовой фразы по заданной эталонной модели.

Для нормализации тестовой и эталонной кодовых книг предложено использовать модифицированный алгоритм векторного квантования:

Шаг 1. Начальное разбиение на M кластеров.

$m=0$. Для $m=0$ разбиение осуществляется с помощью обычного дихотомического алгоритма, результатом работы которого является последовательность векторов-центроидов $\bar{y}_i(0), 1 \leq i \leq L$.

Шаг 2. Классификация.

Входные векторы $\{\bar{x}_i, 1 \leq i \leq M\}$ разбиваются по кластерам C_i с помощью правила ближайшего соседа $\bar{x} \in c_i(m)$ тогда и только тогда, когда $d(\bar{x}, \bar{y}_i(m)) \leq d(\bar{x}, \bar{y}_j(m))$ для всех $j \neq i$.

Шаг 3. Коррекция кодового вектора.

Для $m = m+1$ производится пересчет центроидов всех кластеров $\bar{y}_i(m) = \text{cent}(c_i(m)) \quad 1 \leq i \leq L$.

Шаг 4. Проверка на окончание процедуры.

При формировании кодовой книги-эталона значение M заранее известно. В этом случае квантование завершается, если интегральная мера близости

$$D(m) = \frac{1}{L \cdot N} \sum_{i=1}^L \sum_{j=1}^N d_{\min}(\bar{x}_j, \bar{y}_i) > P \quad (9)$$

на итерации m не уменьшилась по отношению к $D(m-1)$. В противном случае осуществляется переход к шагу 2.

Шаг 5. Окончание

Результатом работы алгоритма являются значения параметров векторов центроидов кодовой книги и интегральная мера близости, вычисленная на шаге 4.

Таким образом, одинаковое значение интегрального шума тестовых и эталонных книг гарантирует, что интегральная ошибка квантования для различных эталонных и тестовых кодовых книг может сравниваться.

Вторым способом нормализации значения интегрального шума тестового высказывания является нормализация при помощи значения шума кодовой книги тестового высказывания D_{norm} . В этом случае при построении кодовой и тестовой кодовых книг задаются размеры этих книг, соответственно K и M . Однако для тестирования высказывания используются не только центры кодовой книги, но и значение шума тестовой фразы для тестовой кодовой книги

$$D_{final} = \frac{D_{compare}}{D_{norm}} \quad (10)$$

Третий способ нормализации значения шума квантования тестового высказывания осуществляется при помощи весов кластеров эталонной кодовой книги. Весом кластера в этом случае называется число векторов параметров эталонного высказывания, попавших в кластер. Эмпирическим обоснованием такого метода нормализации может служить то утверждение, что чем больше векторов попадает в кластер, тем выше значимость этого кластера для последующего процесса верификации. Этот метод нормализации дает наибольшее увеличение эффективности верификации, так как использует больше информации об эталонной фразе, чем предыдущие методы.

Модель гауссовых смесей представляет собой взвешенную сумму M компонент и может быть записана выражением

$$p(\bar{x}|\lambda) = \sum_{i=1}^M p_i b_i(\bar{x}), \quad (11)$$

где \bar{x} – это D - мерный вектор случайных величин, $b_i(\bar{x}), i=1, \dots, M$ функции плотности распределения составляющих модели и $p_i, i=1, \dots, M$ – веса компонентов смеси. Каждый компонент является D -мерной гауссовой функцией распределения вида

$$b_i(\bar{x}) = \frac{1}{(2\pi)^{D/2} |\delta_i|^{1/2}} \exp\left\{-\frac{1}{2}(\bar{x} - \bar{\mu}_i)' \delta_i^{-1} (\bar{x} - \bar{\mu}_i)\right\}, \quad (12)$$

где $\bar{\mu}_i$ – вектор математического ожидания, δ_i – ковариационная матрица. Веса смеси удовлетворяют выражению $\sum_{i=1}^M p_i = 1$.

Полностью модель гауссовой смеси определяется векторами математического ожидания, ковариационными матрицами и весами смесей для каждого компонента модели. Эти параметры все вместе записываются в виде $\lambda = \{p_i, \bar{\mu}_i, \delta_i\}, i=1, \dots, M$.

При обучении модели гауссовых смесей существует проблема инициализации параметров модели перед началом процесса обучения. Алгоритмы обучения не гарантируют нахождение глобального максимума в пространстве векторов обучения и поэтому результат обучения системы существенно зависит от начальных значений параметров системы.

Предполагается, что число компонент в модели гауссовых смесей совпадает с размером кодовой книги, т.е. $M=L$. Тогда математическое ожидание $\bar{\mu}_i$ каждого компонента инициализируется значениями элементов кодового вектора \bar{y}_i . Ковариационная матрица δ_i вычисляется на основании векторов, принадлежащих i -му

кластеру кодовой книги. Веса компонентов p_i модели инициализируются числом векторов, попавших в кластер C_i . Так как веса должны удовлетворять выражению $\sum_{i=1}^M p_i = 1$, то каждый вес нормализуется общим числом векторов \bar{x} , участвующих в обучении модели.

В четвертой главе рассматриваются практические применения предложенных методов и алгоритмов первичного анализа, а также стохастического моделирования голоса диктора в текстонезависимых системах верификации диктора по голосу.

При построении системы верификации диктора и в течение разработки алгоритмов первичного анализа и статистического моделирования использовалась библиотека обработки и распознавания речи НТК (Hidden Markov Models Toolkit) версии 3.0, созданная в CUED (Cambridge University Engineering Department). Библиотека состоит из модулей, построенных таким образом, что они обеспечивают одинаковый интерфейс с внешними программами и операционной системой, что позволяет легко конфигурировать последовательность обработки речевых файлов. Также библиотека предоставляет централизованный доступ к наиболее часто используемым ресурсам и функциям.

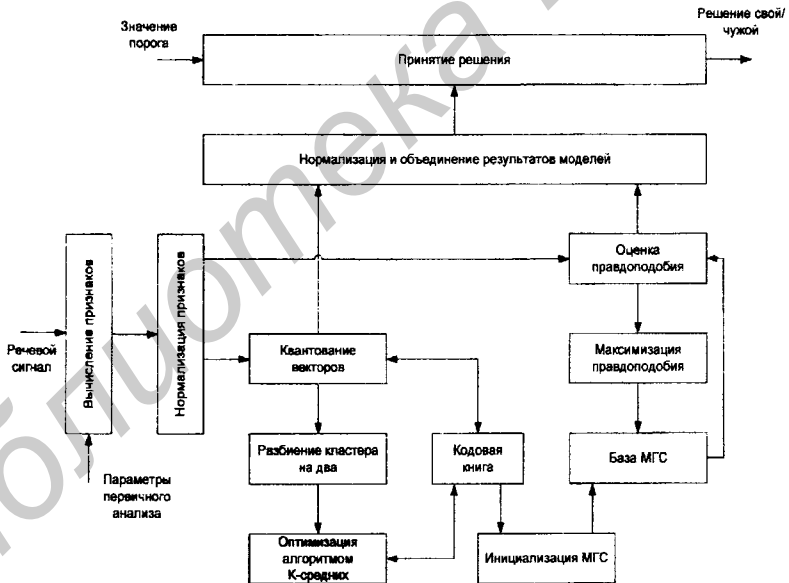


Рис. 3. Структурная схема системы верификации диктора по голосу

Главная цель программной оболочки – это объединить усилия разработчиков по реализации одной системы, обеспечить повторную используемость кода и предоставить унифицированные интерфейсы к функциональным модулям системы.

Программная оболочка позволяет легко конфигурировать систему для проведения различных экспериментов с целью определения оптимальных параметров.

Была разработана система верификации диктора по произвольной речи на основании алгоритмов векторного квантования и моделей гауссовых смесей, представленная на рис. 3.

Экспериментальная проверка системы текстонезависимой верификации диктора по голосу проводилась на двух различных речевых базах:

1. База предложенная Oregon Institute of Science and Technology, Centre of Spoken Language Understanding. Экспериментальная выборка состояла из 10 мужских и 10 женских голосов. Звуковой сигнал записывался с телефонной линии с частотой дискретизации 8 кГц и разрядностью АЦП 16 бит. Продолжительность обучающего высказывания составляла примерно 60 секунд. Продолжительность тестового высказывания составляла примерно 15 секунд. Эта база речевых высказываний использовалась для разработки и проверки алгоритмов верификации.

2. SWITCHBOARD'96, предложенная национальным институтом стандартов и технологии (NIST). Экспериментальная выборка из базы высказываний состояла из 20 женских голосов и 20 мужских голосов. Звуковой сигнал записывался с телефонной линии с частотой дискретизации 8 кГц и кодировался при помощи μ -Law кодера. Эта база использовалась для проведения экспериментов представленных в диссертационной работе.

Таблица 4

	Oregon Institute		SWITCHBOARD'96	
	%	C_{det}	%	C_{det}
VQ	97	0,07	92	0,189
GMM	100	0,05	95	0,115
VQ+GMM	100	0,04	97	0,07

VQ – алгоритм векторного квантования;

GMM – модели гауссовых смесей;

VQ+GMM – объединение алгоритма векторного квантования и моделей гауссовых смесей.

Проверка системы верификации проводилась на двух различных базах высказываний (таблица 4) с использованием вектора первичных параметров размером 36. Для алгоритма векторного квантования был выбран метод нормализации интегрального шума при помощи весов кластеров. Размер кодовой книги составлял 32 кластера. Также было использовано 32 гауссовых смеси. Для объединения результатов верификации от двух моделей использовалось выражение $H(z) = 1 - 0,97z^{-1}$ с коэффициентами $\alpha_1 = 0,545$ для гауссовых моделей и $\alpha_2 = 0,455$ для модели векторного квантования.

Результаты экспериментов демонстрируют повышение эффективности системы верификации диктора по произвольной речи на 1,5-2 % при применении предложенных алгоритмов стохастического моделирования.

Библиотека БГУИР

ЗАКЛЮЧЕНИЕ

В диссертационной работе представлены новые методы, алгоритмы и способы первичной обработки речевого сигнала, стохастического моделирования субречевых единиц, построения системы верификации диктора по произвольной речи с использованием нескольких моделей. Результаты, полученные в ходе диссертационного исследования, можно сформулировать следующим образом:

1. Было установлено, что на этапе выделения первичных признаков речевого сигнала важно обеспечить сохранность информации, которая используется на последующих этапах обработки речи. Потеря информации на первом этапе не может быть восполнена на последующих этапах. Поэтому очень важно использовать различные типы параметров, которые эффективно подчеркивают индивидуальность голоса диктора: кепстральные и дельта кепстральные, вторую производную кепстра [1]. Экспериментально доказано, что оптимальный выбор информативных признаков позволяет достичь максимального эффекта, недостижимого другими средствами. Необходимо усовершенствовать также и другие этапы обработки и моделирования речевого сигнала с целью повышения эффективности распознавания, скорости обучения и верификации. Надежность распознавания зависит не только от типа и способа объединения параметров в вектор, но и от числа параметров [6]. Так как информация об индивидуальности голоса говорящего содержится в 3 – 5 формантах, а точность вычисления зависит от дискретности выбора частот спектра голоса диктора, то увеличение числа кепстральных параметров ведет к значительному увеличению эффективности распознавания [5].

2. Исследованы алгоритмы стохастического моделирования в задачах распознавания речи и диктора. Для распознавания диктора используются в основном стохастические методы моделирования речи, такие как скрытые Марковские модели, алгоритмы векторного квантования, модели гауссовых смесей. Каждый из этих методов [6, 7] имеет свои достоинства, недостатки и ограничения в применении. Для построения систем верификации диктора по произвольной речи наиболее подходящими оказываются модели гауссовых смесей и алгоритмы векторного квантования, которые представляют собой алгоритмы моделирования многомерных случайных величин.

3. Разработан метод выделения субречевых компонент на основании алгоритма K-средних. После выделения первичных признаков распознавания диктора, которые основаны на кепстральных коэффициентах, необходимо выделить вторичные признаки, которые будут использованы для моделирования различных индивидуальных особенностей голоса диктора. Речь различных дикторов может различаться лексическим составом и поэтому определение основных компонент лексики дикторов и является задачей стохастического моделирования речи. [4, 8]. Разработанный метод выделения субречевых единиц является продолжением развития алгоритма K-средних, в котором в качестве критерия окончания итераций предложено использовать значение шума квантования. Таким образом, величина шума

квантования является критерием точности выделения субречевых единиц. Разработанный метод позволил использовать алгоритмы векторного квантования для построения систем текстонезависимого распознавания диктора по голосу.

4. Предложен новый способ инициализации параметров моделей гауссовых смесей [9]. Этот алгоритм основан на алгоритмах векторного квантования. Проводится аналогия между параметрами модели гауссовых смесей и параметрами кодовой книги. Модели гауссовых смесей обучаются при помощи достаточно эффективного алгоритма оценки правдоподобия, который является итерационным алгоритмом. Применение алгоритма векторного квантования для инициализации моделей гауссовых смесей позволяет увеличить эффективность распознавания и скорость обучения моделей.

5. Предложены алгоритмы нормализации среднеквадратичной меры близости для использования классификаторов на основе алгоритмов векторного квантования совместно с моделями гауссовых смесей [10]. Использование комбинации двух различных моделей позволяет уменьшить время обучения системы верификации и повысить эффективность распознавания. Модели гауссовых смесей и алгоритмы векторного квантования могут быть совместно использованы для построения систем верификации диктора по произвольной речи. Однако, для построения такой системы необходимо найти меру сходства оценок двух различных методов. Так алгоритмы векторного квантования в качестве меры сходства двух высказываний используют среднеквадратичное расстояние в пространстве признаков, а модели гауссовых смесей определяют меру сходства величиной вероятности. Это приводит к тому, что необходимо нормализовать значение среднеквадратичного расстояния в диапазоне от 0 до 1, что в свою очередь, предоставляет возможность использовать две системы совместно.

6. Приведены описания различных систем и подсистем распознавания речи и диктора, которые используют методы и алгоритмы, исследуемые в работе [3, 2]. Приложения создавались в рамках концепции разработки программного обеспечения используемого в работе. Различные сферы практического применения указанных алгоритмов и систем позволяют сделать вывод о возможности использования полученных результатов в области распознавания речи, а также в других областях стохастического моделирования и классификации.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

Статьи в журналах:

1. Садыхов Р.Х., Ракуш В.В. Система идентификации диктора по голосу с использованием предварительного взвешивания дельта-кепстральных коэффициентов // Весці НАНБ сер. фізіка-тэхн. навук. – 2003. – №3. – С. 70-73.
2. Садыхов Р.Х., Ракуш В.В. Система распознавания изолированных слов на основе скрытых марковских моделей // Автоматика и вычислительная техника. – Рига, 1998, №4. – с.75-83.
3. Садыхов Р.Х., Ракуш В.В. Модели гауссовых смесей для верификации диктора по произвольной речи. – Доклады БГУИР. – Минск, 2003. – №4. – С. 95 – 103.

Сборники научных трудов

4. Ракуш В.В., Рылов А.С. Алгоритмы распознавания изолированных слов на основе скрытых Марковских моделей // Вопр. криминологии, криминалистики и судебной экспертизы. - Мн.: НИПККиСЭ, 1998. - №13. – С. 234 – 237.
5. Рылов А.С., Ракуш В.В. Алгоритм структуризации речи со среднеквадратичным критерием качества // Вопр. криминологии, криминалистики и судебной экспертизы. – Мн.: НИПККиСЭ, 1998. - №13. – С. 238 – 240.

Тезисы докладов и материалы конференций:

6. Садыхов Р.Х., Ракуш В.В. Моделирование вейвлетных коэффициентов при помощи скрытых Марковских цепей // Современные методы цифровой обработки сигналов в системах измерения, контроля, диагностики и управления: Материалы II междунар. науч.-техн. конф. – Минск, 1998 – С. 261-266.
7. Садыхов Р.Х., Ракуш В.В. Алгоритмы векторного квантования в системе идентификации речевых сигналов // Новые технологии в машиностроении и вычислительной технике: Тр. X науч.–техн. конф. – Брест, 1998.-С.83-88
8. Rakush V.V., Sadykhov R.H. Speaker identification system on arbitrary speech // Pattern Recognition and Information Processing: Proc. Of 5th Inter. Conf. – Minsk, 1999. - P. 52-57.
9. Sadykhov R. Kh., Rakush V.V. Training Gaussian models with vector quantization for speaker verification // Proc. of the 3rd International Conf. on Neural Networks and Artificial Intelligence. – 2003. – P. 175 - 183.
10. Rakush V.V., Sadykhov R.H. Speaker verification system based on the stochastic modeling // Proc. of Inter. Conf. on Informatics in Control, Automation and Robotics. – 2004. – Vol. 3. – P. 183-189.



РЭЗЮМЭ

РАКУШ Валянцін Уладзіміравіч

МЕТАДЫ І АЛГАРЫТМЫ СТАХАСТІЧНАЙ АПРАЦОУКІ МОВЫ У СЫСТЭМАХ
ТЭКСТАНЕЗАЛЕЖНАЙ ВЕРЫФІКАЦЫІ ДЫКТАРА

Ключавыя словы: верыфікацыя дыктара, стахастічнае мадэляванне, апрацоўка мовы, вектарнае квантаванне.

У дысертацыйнай рабоце даследавалася задача верыфікацыі дыктара па адвольнай мове. Мэтай было распрацаваць і ўдасканаліць метады і алгарытмы распазнавання голаса дыктара адносна да тэкстанезалежнай верыфікацыі.

Распрацаван метад фармавання першасных прызнакаў, дазваляючы павялічыць эфектыўнасьць працы сістэмы верыфікацыі ў выпадку запісу мовы ў неспрыяльных умовах.

Распрацаваны алгарытмы верыфікацыі, якія выкарыстоўваюць у якасці крытэрыя нармалізацыі значэнне інтэгральнага шума квантавання тэставай кодавай кнігі і веса кластэроў. Распрацаваныя алгарытмы дазваляюць скараціць аб'ём патрабуемай памяці і павялічыць хуткасьць верыфікацыі.

Распрацавана сістэма верыфікацыі дыктара па адвольнай мове з выкарыстаннем смесі гаўсавых мадэляў, якая ўжывае прапанаваны падыход субмоўных адзінак. У гэтай сістэме быў выкарыстан новы спосаб ініцыялізацыі параметраў смесі гаўсавых мадэляў, які прадэманстравваў значна большы працэнт верыфікацыі, чым спосаб выпадковай ініцыялізацыі. Вызначаны вылічальныя затраты алгарытма ацэнкі максімізацыі для абучэння смесей гаўсавых мадэляў.

Распрацавана сістэма верыфікацыі дыктара па адвольнай мове заснаваная на камбінацыі метадаў вектарнага квантавання і мадэляў гаўсавых смесяў.

РЕЗЮМЕ

РАКУШ Валентин Владимирович

МЕТОДЫ И АЛГОРИТМЫ СТОХАСТИЧЕСКОЙ ОБРАБОТКИ РЕЧИ В СИСТЕМАХ ТЕКСТОНЕЗАВИСИМОЙ ВЕРИФИКАЦИИ ДИКТОРА

Ключевые слова: верификация диктора, стохастическое моделирование, обработка речи, векторное квантование.

В диссертационной работе исследовалась задача верификации диктора по произвольной речи. Целью было разработать и усовершенствовать методы и алгоритмы распознавания голоса диктора применительно к текстонезависимой верификации.

Разработан метод формирования первичных признаков, позволяющий увеличить эффективность работы системы верификации в случае записи речи в неблагоприятных условиях.

Разработаны алгоритмы верификации, которые используют в качестве критерия нормализации значение интегрального шума квантования тестовой кодовой книги и веса кластеров. Разработанные алгоритмы позволяют сократить объем требуемой памяти и увеличить скорость верификации.

Разработана система верификации диктора по произвольной речи с использованием смеси гауссовых моделей, которая использует предложенный подход субречевых единиц. В этой системе был реализован новый способ инициализации параметров смеси гауссовых моделей, который продемонстрировал более высокий процент верификации, чем способ случайной инициализации. Определены вычислительные затраты алгоритма оценки максимизации для обучения смесей гауссовых моделей.

Разработана система верификации диктора по произвольной речи основанная на комбинации методов векторного квантования и моделей гауссовых моделей.

RESUME

Rakush Valentin Vladimirovich

**THE METHODS AND ALGORITHMS OF THE STOCHASTIC SPEECH
PROCESSING IN THE TEXT INDEPENDENT SPEAKER VERIFICATION SYSTEMS**

Keywords: speaker verification, stochastic modeling, speech processing, vector quantization

The task of the text independent speaker verification has been investigated in the PhD thesis. The main goal is to develop and improve methods and algorithms of the text independent speaker verification.

The primary properties extraction method has been developed that allows improve the verification system quality on the speech recorded in the adverse conditions. The method provides new approach in the cepstral coefficients heightening and formation.

The verification algorithms, which use the integral distortion value of the test code book and cluster's weights as a criteria for normalization has been developed. Those algorithms allow reduce the computation and memory usage and reduce verification time.

The text independent speaker verification system has been developed. It uses the Gaussian mixture models to implement the proposed approach of the sub-language units. This system implements new approach for Gaussian mixture model parameters initialization that shows higher verification percentage than random initialization approach. The expectation-maximization algorithm computation expenses for training Gaussian mixture models have been estimated.

The new speaker verification text independent system based on the fusion of the Gaussian mixture models and vector quantization code books has been developed.

Ракуш Валентин Владимирович

**МЕТОДЫ И АЛГОРИТМЫ СТОХАСТИЧЕСКОЙ ОБРАБОТКИ РЕЧИ В
СИСТЕМАХ ТЕКСТОНЕЗАВИСИМОЙ ВЕРИФИКАЦИИ ДИКТОРА**

Специальность 05.13.15 – Вычислительные машины и системы

**Автореферат диссертации
на соискание ученой степени кандидата технических наук**

Подписано в печать	16.11.2004.	Формат 60x84 1/16.	Бумага офсетная.
Гарнитура «Таймс».	Печать ризографическая.		Усл. печ. л. 1,4.
Уч.-изд. л. 1,2.	Тираж 80 экз.		Заказ 677.

Издатель и полиграфическое исполнение: Учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники»
Лицензия на осуществление издательской деятельности №02330/0056964 от 01.04.2004.
Лицензия на осуществление полиграфической деятельности №02330/0133108 от 30.04.2004.
220013, Минск, П. Бровка, 6.