

На правах рукописи

**Симончик Константин Константинович**

**МЕТОД И АЛГОРИТМЫ АВТОМАТИЧЕСКОЙ  
ТЕКСТОНЕЗАВИСИМОЙ ВЕРИФИКАЦИИ ДИКТОРОВ  
И ИХ ПРОГРАММНАЯ РЕАЛИЗАЦИЯ**

Специальность: 05.13.18 - Математическое моделирование, численные  
методы и комплексы программ

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени  
кандидата технических наук

Санкт-Петербург – 2010

Работа выполнена в Санкт-Петербургском государственном электротехническом университете «ЛЭТИ» им. В.И. Ульянова (Ленина)

Научный руководитель – доктор технических наук, профессор Геппенер Владимир Владимирович

Официальные оппоненты:

доктор технических наук,  
профессор

Шульгин Евгений Александрович

кандидат технических наук,  
старший научный сотрудник

Егоров Владимир Викторович

Ведущая организация – Учреждение Российской академии наук "Санкт-Петербургский институт информатики и автоматизации Российской академии наук" (СПИИРАН)

Защита диссертации состоится « \_\_\_\_ » \_\_\_\_\_ 2010 г. в \_\_\_\_\_ часов на заседании совета по защите докторских и кандидатских диссертаций Д212.238.01 Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» им. В.И. Ульянова (Ленина) по адресу: 197376, Санкт-Петербург, ул. Проф. Попова, 5.

С диссертацией можно ознакомиться в библиотеке университета.

Автореферат разослан « \_\_\_\_ »

2010г.

Ученый секретарь совета по защите докторских и кандидатских диссертаций

Н. Л. Щеголева

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы.** В настоящий момент область речевых технологий является одной из наиболее динамично развивающихся областей прикладной науки. Это связано с наличием большого числа актуальных задач, связанных с обработкой речи человека. В последние два десятилетия, в связи с бурным ростом компьютерных технологий, произошел прорыв в области автоматической обработки речевого сигнала. Как оказалось, многие задачи из области речевых технологий представляют интерес как для бизнес-структур, так и для правительственных организаций. Действительно, практически все основные прикладные задачи этой области имеют выход как на чисто коммерческое, так и на специальное (связанное с обеспечением безопасности граждан) применение. Например, система автоматического выделения ключевых слов может использоваться как для индексирования речевых баз данных, так и для поиска в телефонном канале определенных слов и фраз; система идентификации диктора применяется как для контроля доступа пользователей к сервисам частных компаний, так и для выявления преступника по голосу. Поддержка речевых технологий из разных источников привела к развитию собственно автоматических методов, развивающихся и уже активно используемых настоящее время. Тем не менее, в виду сложности и комплексного характера задач, возникающих в процессе обработки естественного речевого сигнала, эти задачи далеки от того, чтобы можно было бы считать их решенными как в практическом, так и в научном плане.

Системы верификации дикторов по голосу относятся к множеству биометрических систем и, как следствие, области их применения пересекаются. Достоинством систем распознавания диктора по голосу является то, что такие биометрические системы чаще всего не требуют дополнительного оборудования и могут быть реализованы с использованием телефонных сетей или устройств ввода-вывода разных типов (микрофонов). Области применения таких систем различны:

– Автоматическая идентификация подозреваемого в телефонном канале. Подобная система позволяет проводить постоянный мониторинг телефонных переговоров (как в телефонном канале, так и для IP-телефонии) без присутствия оператора и сигнализировать об обнаружении диктора с голосом, похожим на голос подозреваемого в одном из каналов для дальнейших оперативных мероприятий (запись разговора, определение местоположения говорящих т.п.).

– Обработка речевых баз данных. Система идентификации диктора может выделить все вхождения голоса диктора в произвольной речевой базе данных. В таком случае система освобождает оператора от прослушивания всего звукового массива и предоставляет ему набор наиболее вероятных речевых сегментов, что в значительной степени удешевляет и ускоряет процесс поиска по речевым базам данных.

– Криминалистические исследования. Технология автоматического распознавания диктора по голосу уже сейчас используется в современных лабораториях криминалистических исследований как средство анализа фонограмм подозреваемых.

– Системы контроля доступа. Такие системы могут обеспечить дополнительный контроль доступа как к физическим объектам, так и к удаленным ресурсам в компьютерных сетях.

Задачей автоматической верификации дикторов можно считать создание модели, алгоритмов и, наконец, программного модуля, основанных на такой математической модели, которая позволила бы приблизиться к результатам, достигаемым человеком и, возможно, улучшить их. Данная диссертация посвящена описанию тех подходов, которые обеспечивают наилучшие результаты автоматической текстонезависимой идентификации дикторов и имеют хорошо просматриваемые перспективы дальнейшего развития. Так, ближайшие несколько лет модели с использованием смесей гауссовых распределений (СГР) и метод опорных векторов (МОВ) остаются доминирующими при моделировании голоса диктора в задаче текстонезависимой верификации. На основании этого анализа разрабатывается и исследуется система верификации дикторов с использованием СГР и МОВ подхода.

**Целью** диссертационной работы является разработка и исследование метода и алгоритмов текстонезависимой верификации дикторов по голосу, разработка программных средств текстонезависимой верификации дикторов, устойчиво работающей в различных каналах связи.

В соответствии с указанной целью в работе сформулированы и решены следующие **задачи**:

1. Выполнен обзор методов, алгоритмов и систем текстонезависимой верификации дикторов по голосу.
2. Разработаны метод и алгоритмы построения систем верификации с использованием СГР-МОВ подхода.
3. Разработано и введено в эксплуатацию программное обеспечение текстонезависимой системы верификации по голосу с использованием СГР-МОВ подхода.
4. Проведено экспериментальное исследование разработанной системы верификации дикторов и выполнен подбор параметров программных модулей системы с целью минимизации равновероятной ошибки пропуска/отклонения (EER).

**Предмет и методы исследования.** Предметом исследования является текстонезависимая система верификации по голосу. При решении поставленных задач использовался аппарат математического моделирования, численных методов, математической статистики, теории цифровой обработки сигналов, теории случайных процессов, теории распознавания образов.

**Научная новизна** работы представлена следующими положениями:

1. Предложен метод текстонезависимой верификации по голосу, использующий компенсацию канальных искажений (метод «собственных каналов») и классификатор на основе метода опорных векторов.
2. Предложен комплекс алгоритмов предобработки речевого сигнала, позволяющий повысить качество работы системы верификации в условиях различного рода помех: импульсных, мультитональных, музыкальных, а также при нелинейных амплитудных искажениях.

3. Разработана методика обучения системы текстонезависимой верификации по голосу на основе использования СГР и МОВ.

**Научные положения, выносимые на защиту:**

1. Метод текстонезависимой верификации по голосу, использующий компенсацию канальных искажений (метод «собственных каналов») и классификатор на основе метода опорных векторов.

2. Комплекс алгоритмов предобработки речевого сигнала, позволяющий повысить качество работы системы верификации в условиях различного рода помех: импульсных, мультитональных, музыкальных, а также при нелинейных амплитудных искажениях.

3. Методика обучения системы текстонезависимой верификации по голосу на основе использования СГР и МОВ.

**Практическая значимость работы** заключается в реализации предложенных методов и алгоритмов в виде комплекса программ системы текстонезависимой верификации, устойчиво работающего в различных каналах связи, включающего программные модули предобработки речевого сигнала. Также предложены программные реализации методики обучения системы, в том числе и для адаптации системы для работы с фонограммами, записанными в специфических каналах связи.

**Достоверность результатов исследования** подтверждается корректным использованием математического аппарата, результатами экспериментальных исследований на программных моделях и результатами испытаний реальной системы, при создании которой использовались предложенные методы и алгоритмы.

**Внедрение результатов.** Результаты работы использовались при проведении в ООО «Центр Речевых Технологий» 2 ОКР и производстве 3 коммерческих продуктов. Работа поддержана программой фонда содействия развитию малых форм предприятий в научно-технической сфере «Участник молодежного движения Научно-Инновационного Конкурса» 2007 года.

**Апробация результатов исследования.** Основные положения и результаты работы изложены в следующих документах:

1. Отчет по выполнению опытно-конструкторской работы «Разработка и создание автоматизированной фоноскопической системы поиска и учета лиц, представляющих оперативный интерес для правоохранительных органов, в том числе с учетом состояний наркотического опьянения», шифр «Невод-07», № гос. контракта 9/1/2818, 2009 г.

2. Отчет по выполнению опытно-конструкторской работы «Разработка аппаратно-программного комплекса по идентификации лиц, говорящих на разных языках (цыганском, таджикском и др.) и представляющих оперативный интерес для правоохранительных органов», шифр «Этнос-06», 2008 г.

Основные положения и результаты работы докладывались и обсуждались на ряде конференций, в число которых входят следующие:

1. 7-я Международная конференция "Распознавание образов и анализ изображений: новые информационные технологии". Санкт-Петербург, 18-23 октября 2004.

2. Технологии Microsoft в теории и практике программирования. Санкт-Петербург, 1-2 марта 2005.

3. VIII Международная конференция по мягким вычислениям и измерениям. Санкт-Петербург, 27-29 июня 2005.

4. 59-я научно-техническая конференция профессорско-преподавательского состава СПбГЭТУ «ЛЭТИ». Санкт-Петербург 24 января-31 февраля 2006.

5. IX Международная конференция по мягким вычислениям и измерениям. Санкт-Петербург, 27-29 июня 2006.

6. 60 научно-техническая конференция профессорско-преподавательского состава СПбГЭТУ «ЛЭТИ». Санкт-Петербург 30 января-7 февраля 2007.

7. Технологии Microsoft в теории и практике программирования. Санкт-Петербург, 13-14 марта 2007.

8. X Международная конференция по мягким вычислениям и измерениям. Санкт-Петербург, 25-28 июня 2007.

9. 10-я Международная научно-техническая конференция и выставка «Цифровая обработка сигналов и ее применение». Москва, 28-30 марта 2008.

10. 12-я Международная научно-техническая конференция и выставка «Цифровая обработка сигналов и ее применение». Москва, 31 марта – 2 апреля 2010.

**Публикации.** По теме диссертации опубликовано 14 работ, из них 5 статей опубликованы в журналах, рекомендованных Высшей аттестационной комиссией, и 9 работ в материалах научно-технических конференций.

#### **Структура и объем работы.**

Диссертация состоит из введения, пяти глав, заключения, списка литературы, включающего 85 наименований. Основная часть работы изложена на 175 страницах машинописного текста. Работа содержит 50 рисунков и 20 таблиц.

## **КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ**

**Во введении** обосновывается актуальность диссертационной работы, определяются цель и задачи исследования, формулируются научная новизна и практическая ценность результатов.

**В первой главе** вводятся основные определения задачи сравнения голосов дикторов, приводятся принципы построения систем верификации по голосу, дается краткое описание их основных модулей. Приводятся критерии эффективности систем голосового поиска. Дан обзор современных существующих систем идентификации/верификации по голосу как отечественных, так и зарубежных производителей.

По постановке задачи и методу работы системы поиска голосов и верификации диктора делятся на текстозависимые и текстонезависимые, а также на системы, работающие на открытом или закрытом множестве дикторов.

Система называется текстозависимой, если она заранее обучена на распознавание определенной фразы, которую должен произнести диктор. В реальных системах диктор произносит либо заранее определенную «парольную»

фразу, либо система сама определяет и запрашивает фразу, которую должен произнести диктор. В любом из этих случаев распознаваемая фраза должна быть произнесена каждым из возможных дикторов в процессе обучения системы. Система распознавания диктора называется текстонезависимой, если она не содержит информации о том, что именно диктор будет произносить. Система обучается и тестируется на произвольных речевых данных. Как следствие, эффективность текстонезависимых систем распознавания обычно ниже, чем текстозависимых.

Одним из факторов, по которым определяют эффективность биометрических систем, является частота появления ошибок. Существует два типа ошибок: ошибка отказа FRR (False Reject Rate), пропорциональная числу попыток клиентов, которые отвергаются и ошибка пропуска FAR (False Accept Rate), пропорциональная числу попыток злоумышленников, которые допускаются. Очень часто исследователи используют критерий равенства частоты появления ошибок FRR и FAR - EER (Equal Error Rate) для описания своих исследовательских систем.

На текущий момент на российском рынке представлены такие системы распознавания дикторов, как система автоматизации фоноучетов "Трал М" ООО "Центр Речевых Технологий" с единым уровнем ошибки EER 8%, система "PhonoBase" компании ООО "Целевые Технологии" с показателями FRR = 10% и FAR = 1%.

Во **второй главе** рассматриваются методы первичной обработки и выбора признаков речевых сигналов в задаче распознавания дикторов. Анализируются методы измерения и выбор совокупности параметров речевого сигнала, пригодных охарактеризовать индивидуальность диктора (система признаковых описаний речи), при этом необходимо выделить информативные признаки речевого сигнала, описывающие сигнал в наиболее краткой форме без существенной потери информации.

Особое внимание уделено выделению индивидуальных характеристик диктора из его речи. Важнейшими из них являются коэффициенты линейного предсказания, кепстральные признаки, а также спектрально-формантные.

При передаче речевого сигнала по тракту, включающему в себя помещение, канал записи, канал передачи и канал воспроизведения, на него могут накладываться шумы и помехи, например:

- шумы помещения,
- шумы электронных компонентов каналов записи, передачи и воспроизведения,
- шумы реверберации,
- помехи со стороны других дикторов и других каналов.

Кроме того, в речевой сигнал могут вноситься амплитудно-частотные искажения, определяемые характеристиками помещения, микрофона, каналов записи и передачи. Все эти помехи, шумы и искажения, накладываемые на речевой сигнал при его передаче по каналам связи, могут негативно повлиять на процесс верификации диктора по голосу.

В данной работе предлагается комплекс алгоритмов предобработки речевого сигнала, позволяющий повысить качество работы системы верификации в

условиях различного рода помех: импульсных, мультитональных, музыкальных, а также при нелинейных амплитудных искажениях.

Комплекс представлен пятью алгоритмами:

- 1) алгоритм детектирования щелчков;
- 2) алгоритм детектирования перегрузов;
- 3) алгоритм детектирования музыкальных сигналов;
- 4) алгоритм детектирования гудков;
- 5) алгоритм детектирования речевого сигнала.

На рис. 1 приведена схема взаимодействия компонент (программных модулей) модуля предобработки.

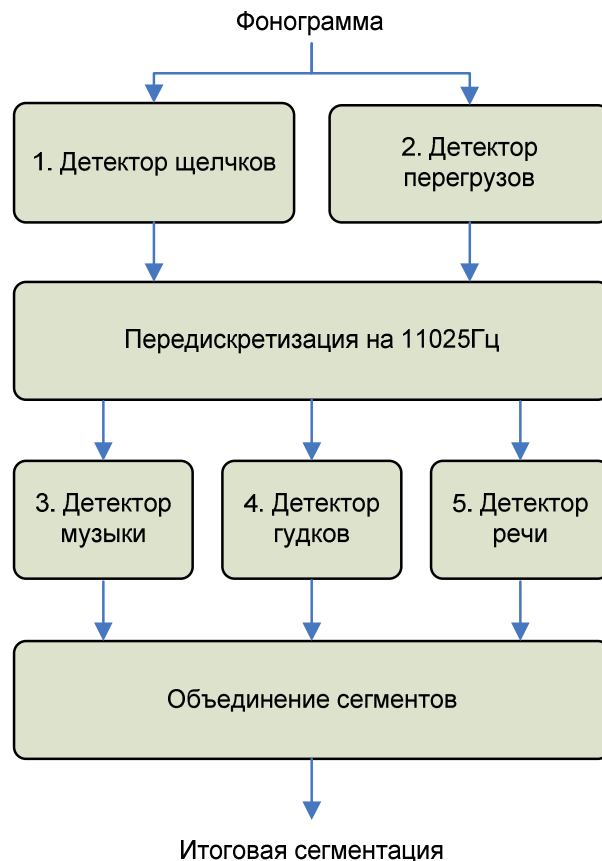


Рисунок 1. Схема модуля предобработки речевого сигнала.

В качестве входных данных для каждого детектора является фонограмма с исходным сигналом (детекторы щелчков и перегрузов) или передискретизированным на 11025 Гц (остальные детекторы). Выходными параметрами является список сегментов целевого для данного детектора сигнала.

**Третья глава** диссертации посвящена описанию математических методов, используемых при моделировании голоса диктора. Рассматривается использование моделей на основе использования смеси гауссовых распределений (СГР) и метода опорных векторов (МОВ).

В данной работе предлагается система текстонезависимой верификации дикторов по голосу, сочетающая метод компенсации канальных искажений



моделей СГР (метод «собственных каналов») и метод опорных векторов для классификации результатов сравнения моделей СГР.

Выбор сочетания использования этих методов обусловлен тем, что метод опорных векторов широко зарекомендовал себя в последнее время для эффективного сравнения СГР моделей голосов дикторов. С другой стороны, одним из основных факторов, влияющих на качество системы верификации, является искажения спектра речевого сигнала, вносимые АЧХ канала связи. Метод «собственных каналов», позволяет компенсировать эти искажения ещё на этапе построения СГР модели голоса, таким образом, облегчая задачу классификатору.

Модель гауссовой смеси голоса диктора представляется набором параметров определенных для каждой  $i$ -ой компоненты смеси: средними векторами  $\bar{\mu}_i$ , матрицами ковариации  $\Sigma_i$  и весами  $\alpha_i$ . Эта совокупность параметров называется моделью СГР  $\lambda$ :

$$\lambda = \{\alpha_i, \bar{\mu}_i, \Sigma_i\}, i = 1, \dots, M, \quad (1)$$

Для наблюдаемого входного вектора  $\bar{x}$  плотность Гауссовой смеси является взвешенной суммой  $M$  компонент смеси, которая задается выражением:

$$p(\bar{x}|\lambda) = \sum_{i=1}^M \alpha_i p_i(\bar{x}), \quad \sum_{i=1}^M \alpha_i = 1, \quad \alpha_i > 0, \quad i = 1, \dots, M \quad (2)$$

где  $\bar{x}$  –  $D$ -мерный случайный вектор,  $p_i(\bar{x})$ ,  $i = 1, \dots, M$ , являются компонентами смеси и  $\alpha_i$ ,  $\{i = 1, \dots, M\}$  являются весами смеси. Каждая компонента плотности – это функция Гаусса  $n$ -мерной случайной величины со средним вектором  $\mu_i$  и  $D \times D$  ковариационной матрицей  $\Sigma_i$ :

$$p_i(\bar{x}) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_i|}} \exp\left\{-\frac{1}{2}(\bar{x} - \bar{\mu}_i)^T \Sigma_i^{-1} (\bar{x} - \bar{\mu}_i)\right\} \quad (3)$$

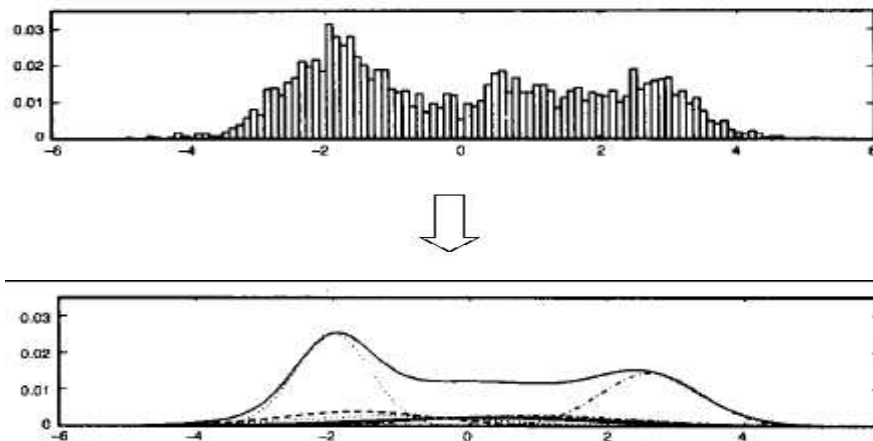


Рисунок 2. Приведение к линейным комбинация Гауссовых функций.

Основной проблемой при решении задачи верификации для систем с использованием СГР является проблема рассогласования, вызванная помехами, вносимыми каналом записи. Под понятием «канал» подразумевается совокупный эффект влияния микрофона мобильного или стационарного телефона, влияние АЧХ канала соединения и другие искажения, вносимые записывающей аппаратурой. Для компенсации такого рода помех для СГР в данной работе

используется метод «собственных каналов», который предполагает разложение супервектора средних модели СГР на составляющие:

$$\vec{M} = \vec{S} + \vec{C} \quad (4)$$

где супервектор  $\vec{S}$  зависит только от диктора, а супервектор  $\vec{C}$  зависит только от эффектов канала (рис. 4).

Вектор  $\vec{S}$  выступает в качестве опорного канала, тогда как для определения супервектора  $\vec{C}$  исходят из разложения:

$$\vec{C} = \hat{U} \cdot \vec{x}, \quad (5)$$

где  $\vec{x}$  – скрытый низко-размерный вектор каналов,  
 $\hat{U}$  – фиксированная матрица собственных каналов.

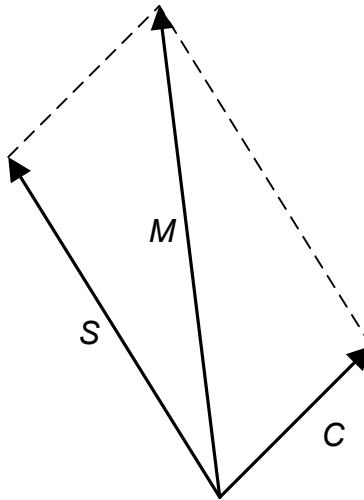


Рисунок 3. Разложение дикторо- и каналозависимого вектора  $M$ .

Матрица  $\hat{U}$  рассчитывается методом с помощью MLES-алгоритма (Maximum Likelihood Eigen Space). Скрытый вектор  $\vec{x}$  вычисляется методом MLED (Maximum Likelihood Eigen-Decomposition), и представляет собой точку в пространстве собственных каналов для модели СГР, построенной по заданному фрагменту речи.

В качестве классификатора в данной работе используется метод опорных векторов. Современное развитие МОВ в распознавании диктора выявило, что самое эффективное его применение – это использование МОВ в комбинации с методом СГР. В случае такой гибридной системы, МОВ использует в качестве входных данных для классификации супервектора средних  $\mu_i$  СГР.

Для гипотезы Н1 (диктор спорной и эталонной фонограмм один и тот же) и гипотезы Н2 (дикторы на фонограммах различаются), функция, разделяющая классы в пространстве акустических признаков, будет:

$$f(\mu) = \sum_{i=1}^N w_i t_i K(\mu, \mu_i) + b \quad (6)$$

где вектор  $w = (w_1, \dots, w_n) \in R^n$ ,  $b \in R^n$  - скалярный порог,  $t_i \in \{0,1\}$  - идеальные выходы классификатора,  $K(\mu, \mu_i)$  - ядро классификатора.

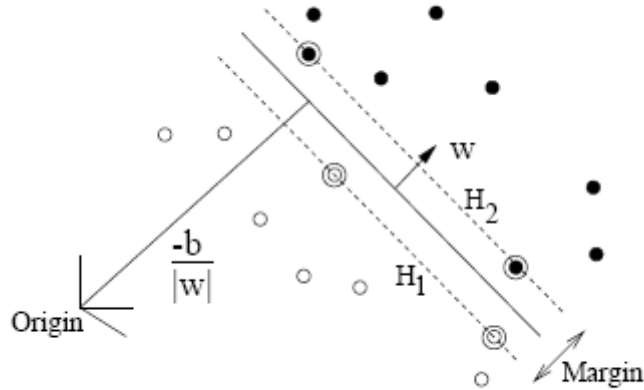


Рисунок 4. Построение оптимальной разделяющей гиперплоскости.

В данной работе в качестве ядра использовалось линейное ядро Кампбелла:

$$K_{lin}(s^a, s^b) = \sum_{i=1}^N w_i \mu_i^a \Sigma_i^{-1} (\mu_i^a)^t = \sum_{i=1}^N \left( \sqrt{w_i} \Sigma_i^{-\frac{1}{2}} \mu_i^a \right) \left( \sqrt{w_i} \Sigma_i^{-\frac{1}{2}} \mu_i^b \right)^t \quad (7)$$

На рис. 6 показаны основные компоненты СГР-МОВ системы верификации.

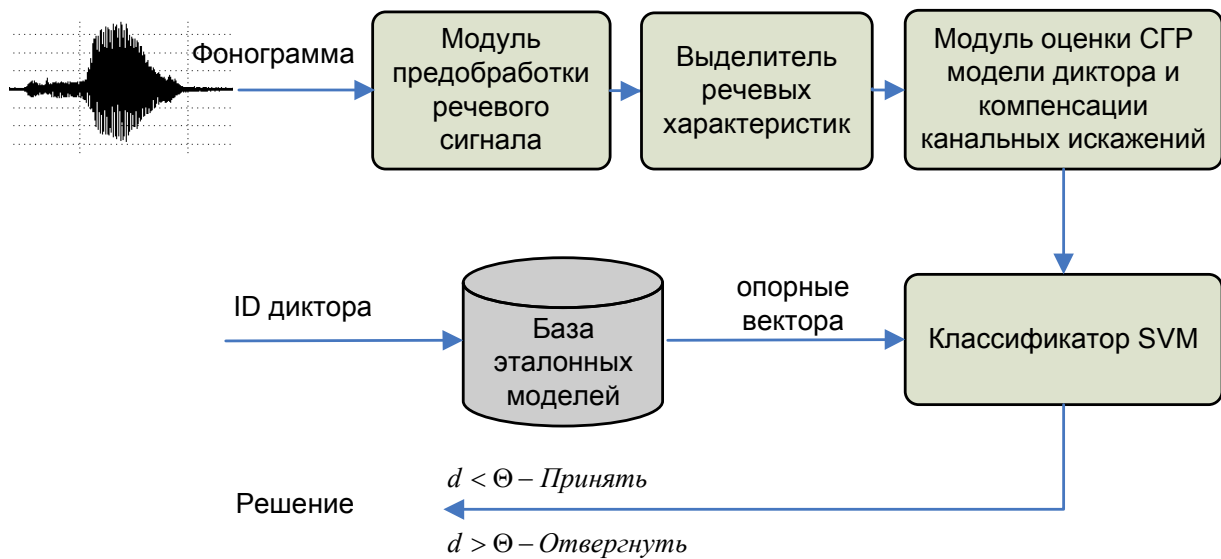


Рисунок 5. Система верификации диктора на основе СГР-SVM.

Роль блока формирования акустических векторов-признаков сводится к предобработке входного речевого сигнала для определения параметров, которые передают дикторозависимую информацию. На выходе этого блока получается последовательность акустических векторов-признаков, представляющих тестовый или тренировочный сегмент:  $X = \{\bar{x}_1, \dots, \bar{x}_T\}$ , где  $x_t$  - вектор,  $t = \{1, \dots, T\}$ .

Блок «оценки модели диктора» выполняет алгоритм обучения СГР модели диктора. Модель диктора строится в рабочем цикле системы, когда требуется верифицировать спорную фонограмму, т.е. определить, относится ли эта

фонограмма к вероятному диктору  $S$  или нет. Модель диктора СГР, построенная с учетом компенсации канальных искажений методом "собственных каналов", представляется далее в качестве входных данных для классификатора на базе МОВ. С использованием специальной базы СГР моделей, выступающих в качестве опорных векторов, строится гиперплоскость разделяющая пространство речевых признаков на 2 искомого класса:  $H_1$  и  $H_2$ . Количество опорных векторов было выбрано равным 1200 как оптимальное по соотношению скорость верификации/качество.

Качество работы системы верификации на основе использования СГР и МОВ напрямую зависит от того, каким образом была обучена универсальная фоновая модель (УФМ) и матрица компенсации канальных искажений  $U$ . В данной работе предложена методика обучения системы текстонезависимой верификации по голосу с использованием СГР и МОВ, которая представляет собой последовательность из трех этапов:

1) *Создание универсальной фоновой модели.* Выдвигается требование того, чтобы тренировочная и тестовая речь дикторов для УФМ была собрана на различных микрофонах или каналах, что является принципиальным для достижения хорошей эффективности системы распознавания. Кроме того, при создании УФМ очень важно разделение дикторов по полу при формировании базы данных. Предлагается использовать способ параллельного обучения СГР моделей каждого диктора тренировочного множества с последующим объединением этих моделей в УФМ.

2) *Расчет матрицы компенсации канальных искажений.* Для расчета матрицы компенсации каналов  $\vec{U}$  предлагается алгоритм, основанный на работах Куна и Нгуена. В качестве обучающей речевой базы должна быть выбрана ровно тот набор фонограмм, что и для обучения УФМ. Обучение матрицы  $\vec{U}$  проходит итерационно. Каждая итерация состоит из 2-х этапов:

- MLED-оценка положения вектора  $\vec{x}(s, h)$  подпространства каналов;
- MLES-оценка значений элементов матрицы  $\vec{U}$ .

Насыщение значений элементов матрицы  $\vec{U}$  происходит при 20 итерациях.

3) *Выбор множества опорных векторов.* Выбор опорных векторов (ОВ) происходит подобно тому, как производится подбор фонограмм для создания УФМ. Требование того, чтобы СГР модели ОВ были представлены различными микрофонами или каналами, является принципиальным для достижения хорошей эффективности системы верификации. Приводится зависимость ошибки EER верификации от количества используемых ОВ. Так, показано, что при достижении количества ОВ равного 1200 наступает насыщение ошибки EER – дальнейшее её падение даже для фонограмм при малых длительностях речевого сигнала прекращается. Поскольку время работы МОВ на этапе классификации пропорционально кубу от мощности множества ОВ, то выбор их количества является компромиссным решением относительно качества и скорости работы системы. Таким образом, рекомендуется использовать порядка нескольких десятков – сотню опорных векторов.

В четвертой главе приводятся исследования разработанной системы верификации на различных речевых базах с целью оптимизации её параметров. В качестве критерия качества использовался показатель EER. В таблице 1 приведены значения EER на 7-ми собранных базах для системы с оптимальными параметрами.

Таблица 1. Качество работы системы на разных речевых базах.

Канал записи спорной и эталонной фонограмм	Кол-во сравнений "свой-свой"	Кол-во сравнений "свой-чужой"	EER, %
GSM	2505	61739	4.5
IP	1340	49850	7.0
E1	1204	20505	4.7
Радиоканал	568	10409	14.0
Аналоговый	541	10234	7.5
Микрофон	1950	50395	1.5
NIST SRE 2004	5032	12850	2.4

Оценка надежности сравнения фонограмм голосов дикторов в зависимости от длительности фонограмм приведена в таблице 2. В данном исследовании спорная и эталонная фонограммы принадлежали различным каналам связи.

Таблица 2. Значение EER в зависимости от сочетания длительностей фонограмм, канал связи GSM-микрофон.

Длительность, сек	4	7	11	18	29	47	77
4	20.2%	19.5%	17.1%	16.6%	15.4%	13.3%	11.9%
7		18.2%	15.4%	13.4%	11.6%	9.5%	8.0%
11			13.0%	11.0%	9.2%	6.7%	5.7%
18				8.0%	6.3%	4.5%	3.8%
29					4.5%	3.1%	2.7%
47						2.5%	1.8%
77							1.8%

Проведены экспериментальные исследования влияния алгоритмов предобработки сигнала на показатель EER качества верификации. В таблице 3 приведена зависимость для речевой базы канала записи GSM

Таблица 3. Влияние алгоритмов предобработки на качество верификации для речевой базы GSM.

Детектор щелчков	Детектор перегрузов	Детектор музыкальных сигналов	Детектор гудков	Детектор речи	EER, %
-	-	-	-	-	14.5
-	-	-	-	+	6.5
-	-	-	+	+	5.9
-	-	+	+	+	5.0
-	+	+	+	+	4.7
+	+	+	+	+	4.5

### **Основные результаты работы**

Представленная диссертационная работа содержит результаты исследований по разработке системы текстонезависимой верификации дикторов по голосу.

Основные научные и практические результаты работы можно сформулировать следующим образом:

- Проведен обзор современных существующих систем идентификации/верификации по голосу как отечественных, так и зарубежных производителей.
- Предложен комплекс алгоритмов предобработки речевого сигнала, позволяющий повысить качество работы системы верификации в условиях различного рода помех: импульсных, мультитональных, музыкальных, а также при нелинейных амплитудных искажениях.
- Предложен метод текстонезависимой верификации по голосу, использующий компенсацию канальных искажений (метод «собственных каналов») и классификатор на основе метода опорных векторов.
- Разработана методика обучения системы текстонезависимой верификации по голосу на основе использования СГР и МОВ.
- Разработано и введено в эксплуатацию программное обеспечение текстонезависимой системы верификации по голосу с использованием СГР-МОВ подхода.
- Проведено экспериментальное исследование разработанной системы верификации дикторов и выполнен подбор параметров программных модулей системы с целью минимизации равновероятной ошибки пропуска/отклонения (EER).

Разработанная модель системы распознавания дикторов может быть эффективно использована для различных практических приложений. Реализованная программная архитектура позволяет сделать данную систему расширяемой для работы на фонограммах, записанных в условиях специфических канальных помех и искажений, а также для верификации дикторов, говорящих на различных языках.

### **Статьи, опубликованные в изданиях, включенных в перечень ВАК**

1. Симончик К.К., Геппенер В.В. Разработка систем автоматической верификации дикторов с использованием нейронных сетей // Нейрокомпьютеры: разработка и применение. 2006. №7. Издательство "Радиотехника", С. 14-23.
2. Design of Speaker Verification Systems with the Use of an Algorithm of Dynamic Time Warping (DTW) (Разработки системы верификации дикторов по голосу с использованием Алгоритма Динамического Искращения Времени (ДИВ)), V. V. Geppener, K. K. Simonchik, and A. S. Haidar, Pattern Recognition and Image Analysis, Vol. 17, No. 4, 2007, pp. 470-479.
3. Симончик К.К., Головкин Д.Б. Исследование проблемы интерполяции дискретных сигналов в аудио приложениях, Известия Государственного электротехнического университета. №5/2005. Издательство СПбГЭТУ "ЛЭТИ", С. 3-12.

4. Козлов А.В., Лоханова А.И., Симончик К.К. Алгоритм детектирования музыкальных фрагментов в задачах речевой обработки, «Научно-технические ведомости СПбГПУ», Издательство Политехнического университета, 2010, №4(103), стр. 7-11.
5. Симончик К.К., Галинина О.С., Капустин А.И. Алгоритм обнаружения речевой активности на основе статистик основного тона в задаче распознавания диктора, «Научно-технические ведомости СПбГПУ», Издательство Политехнического университета, 2010, №4(103), стр. 18-23.

#### Другие работы

6. Development of an Automatic Speaker-Verification System Using the Dynamic Time-Warping Algorithm (Разработки системы верификации дикторов по голосу с использованием Алгоритма Динамического Искращения Времени (ДИВ)), V. V. Geppener, K. K. Simonchik, and A. S. Haidar, Pattern Recognition and Image Analysis, Vol. 15, No. 2, 2005, p. 397.
7. Разработка систем автоматической верификации дикторов с использованием алгоритма динамического искажения времени (ДИВ), В.В. Геппенер, К.К. Симончик, Технологии Microsoft в теории и практике программирования, Издательство Политехнического университета, 2005г, стр. 60-61.
8. Разработка систем автоматической верификации дикторов с использованием нейронных сетей, Геппенер В.В., Тихонов Д.В., Симончик К.К., сборник докладов VII Международной конференции по мягким вычислениям и измерениям, SCM'2005, СПбГЭТУ «ЛЭТИ», 2005.
9. Разработка библиотеки функций Matlab для использования в задачах верификации дикторов по голосу, В.В. Геппенер, К.К. Симончик, О.А. Черемушкина, сборник докладов IX Международной конференции по мягким вычислениям и измерениям, SCM'2006, СПбГЭТУ "ЛЭТИ", 2006, том 1, С. 124-127.
10. Разработка встроенной системы автоматической верификации дикторов по голосу, В.В. Геппенер, К.К. Симончик, Технологии Microsoft в теории и практике программирования (материалы межвузовского конкурса-конференции студентов, аспирантов и молодых ученых Северо-Запада), Издательство Политехнического университета, 2007г, стр. 68-69.
11. Разработка встроенной системы автоматической верификации дикторов по голосу, В.В. Геппенер, К.К. Симончик, сборник докладов X Международной конференции по мягким вычислениям и измерениям, SCM'2007, СПбГЭТУ «ЛЭТИ», 2007, том 2, стр. 55-58.
12. Разработка встроенной системы автоматической верификации дикторов по голосу, Симончик К.К., Доклады 10-й Международной конференции «Цифровая обработка сигналов и ее применение», Инсвязьиздат, Выпуск X-1, Москва, 2008, том 1, стр. 245-248.
13. Система верификации дикторов по голосу на основе использования СГР-SVM подхода, Капустин А.И., Симончик К.К., Доклады 12-й Международной конференции «Цифровая обработка сигналов и ее применение», Инсвязьиздат, Выпуск XII-1, Москва, 2010, том 1, стр. 207-210.
14. Алгоритм детектирования музыкальных фрагментов в задачах речевой обработки, Лоханова А.И., Симончик К.К., Козлов А.В., Доклады 12-й Международной конференции «Цифровая обработка сигналов и ее применение», Инсвязьиздат, Выпуск XII-1, Москва, 2010, том 1, стр. 210-213.

---

Подписано в печать .11.10. Формат 60\*84 1/16  
Бумага офсетная. Печать офсетная. Печ. л. 1.0  
Тираж 100 экз. Заказ

---

Отпечатано с готового оригинал-макета  
в типографии Издательства СПбГЭТУ "ЛЭТИ"

Издательство СПбГЭТУ "ЛЭТИ"  
197376, С.-Петербург, ул. Проф. Попова, 5