На правах рукописи

Тупицин Геннадий Сергеевич

ПРЕДОБРАБОТКА РЕЧЕВЫХ СИГНАЛОВ В СИСТЕМАХ АВТОМАТИЧЕСКОЙ ИДЕНТИФИКАЦИИ ДИКТОРА

Специальность: 05.12.04

Радиотехника, в том числе системы и устройства телевидения

АВТОРЕФЕРАТ

диссертации на соискание учёной степени

кандидата технических наук

Владимир – 2015

Работа выполнена на кафедре динамики электронных систем
ФГБОУ ВПО «Ярославский государственный университет им. П.Г. Демидова».

Научный руководитель: **Брюханов Юрий Александрович**
доктор технических наук, профессор,
заведующий кафедрой динамики электронных систем ФГБОУ ВПО «Ярославский государственный университет им. П.Г. Демидова», г. Ярославль.

Официальные оппоненты: **Левин Евгений Калманович**
доктор технических наук, доцент кафедры радиотехники и радиосистем ФГБОУ ВПО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых», г. Владимир.

Савватин Алексей Иванович
кандидат технических наук, руководитель группы ООО «А-ВИЖН», г. Ярославль.

Ведущая организация: ОАО «Ярославский радиозавод», г. Ярославль

Защита диссертации состоится «23» декабря 2015 г. в 16 часов на заседании диссертационного совета Д 212.025.04 при Владимирском государственном университете имени Александра Григорьевича и Николая Григорьевича Столетовых по адресу: 600000, г. Владимир, ул. Горького, д. 87, ВлГУ, корп. 3, ФРЭМТ, ауд. 301.

С диссертацией можно ознакомиться в библиотеке Владимирского государственного университета имени Александра Григорьевича и Николая Григорьевича Столетовых и на сайте http://diss.vlsu.ru.

Автореферат разослан «21» октября 2015 г.

Отзывы на автореферат, заверенные печатью, просим направлять по адресу: 600000, г. Владимир, ул. Горького, д. 87, ВлГУ, корп. 3, ФРЭМТ.

Ученый секретарь диссертационного совета
доктор технических наук, профессор  А.Г. Самойлов

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы и состояние вопроса. Речь – существенный элемент человеческой деятельности, позволяющий человеку познавать окружающий мир, передавать свои знания и опыт другим людям, аккумулировать их для передачи последующим поколениям. Начиная с древних времен и по сей день она является основным способом обмена информацией между людьми.

Люди в процессе общения способны подсознательно различать голоса других людей. Это связано с тем, что характеристики голоса определяются анатомическими особенностями строения речевого аппарата, которые индивидуальны для каждого человека. Уникальность голоса послужила предпосылкой создания биометрических систем, использующих голос человека в качестве параметра.

Задача распознавания диктора по его голосу была поставлена более 40 лет назад, но исследования в этой области еще продолжаются. Ниже приведены лишь некоторые практические задачи, решение которых требует применения системы распознавания диктора.

- **Биометрический контроль доступа**. Системы биометрического контроля доступа предназначены для обеспечения безопасности доступа к физическим объектам, информационным и финансовым ресурсам.
- **Сопоставление голоса подозреваемого и некоторой фонограммы.** Технология автоматического распознавания диктора по голосу уже сейчас используется в современных лабораториях криминалистических исследований как средство анализа фонограмм подозреваемых.
- **Автоматическое управление тангентой в радиостанциях.** Полудуплексный режим работы широко используется в радиосвязи, однако в условиях занятости рук он может создавать неудобства для пользователя. В качестве решения обычно применяются детекторы речевой активности для автоматического управления тангентой. Однако при этом система может ошибочно активировать тангенту от голоса постороннего человека. Одним из перспективных способов избежать этого является добавление модуля распознавания диктора в радиостанцию.
- **Идентификация говорящего в радиостанциях**. Если радиостанция предназначена для использования несколькими людьми, то одной из возможностей, повышающей удобство эксплуатации устройства, которое принимает в данный момент сигнал с этой радиостанции, является отображение идентификатора говорящего. Определение идентификатора можно осуществлять с помощью системы распознавания диктора.
- **Голосовое управление роботом**. Управление с помощью голосовых команд является одним из важнейших естественных способов взаимодействия с роботом. Однако существуют приложения, в которых необходимо ограничить круг лиц, имеющих доступ к управлению. В этом случае перед распознаванием речевой команды можно выполнять верификацию диктора.

– **Голосовое управление подсистемами автомобиля**. Голосовое управление такими подсистемами, как кондиционер, навигатор, медиаплеер уже сейчас внедряется во многие модели автомобилей. Актуальным является создание индивидуальных профилей голосового управления для разных людей, что может быть реализовано с помощью системы распознавания диктора.

Уже сегодня системы распознавания диктора показывают достаточно высокую точность работы, однако присутствие фонового шума способно существенно ее снизить. Одним из наиболее эффективных способов повышения устойчивости систем распознавания диктора к шумам является применение алгоритмов шумоподавления.

Проблема восстановления речевого сигнала, искаженного аддитивным некоррелированным шумом, в случае, когда доступен только зашумленный сигнал, широко изучалась в прошлом и актуальна сейчас. Предложены методы подавления шума в частотной области, использующие различные функции коррекции спектра (ФКС), зависящие от апостериорного отношения сигнал/шум (ОСШ) и/или оценки априорного ОСШ. Для ФКС спектрального вычитания оценка априорного ОСШ не требуется. В ином случае она может осуществляться с помощью подхода прямого принятия решения (decision-directed), его модификации на основе двухступенчатого алгоритма (two step noise reduction, TSNR), а также других методов. Для коррекции спектра на практике используются различные ФКС: Винера, минимальной среднеквадратичной ошибки кратковременной амплитуды спектра (minimum mean square error short-time spectral amplitude, MMSE-STSA) и др. Помимо методов подавления шума в частотной области существуют и другие подходы.

Отметим, что алгоритмы шумоподавления, максимизирующие показатели качества и разборчивости речевых сигналов, не всегда столь эффективны для предобработки сигналов в задаче распознавания диктора. Сложность также представляет подбор параметров таких алгоритмов, т. к. вычислительная сложность существующих методик оценки систем распознавания диктора, как правило, намного выше вычислительной сложности алгоритмов оценки показателей качества и разборчивости речи.

Таким образом, проблема поиска новых алгоритмов предобработки речевых сигналов в задаче распознавания диктора, а также методик оценки их работы является актуальной.

Основополагающие работы по обработке и анализу речевых сигналов связаны с именами таких известных зарубежных и отечественных ученых, как Рабинер Л., Шафер Р., Фланаган Дж. Л., Римский-Корсаков А.В., Сапожков М.А., Михайлов В.Г. и др.

Интерес к задаче распознавания диктора нашел свое отражение в исследовательских работах Фуруи С., Атала Б., Бейджи Х., Рейнольдса Д., Кэмпбелла В., Ортега-Гарсия Дж., Матвеева Ю.Н., Новоселова С.А. и др.

В области подавления шума в речевых сигналах наибольшую известность получили работы Болла С., Лима Дж., Ефрайма Я., Малла Д., Маколлея Р.,

Малпасса М., Скалара П., Плапоса С., Коэна И., Лойзо Ф., Филхо Дж., Ванга Д., Петровского А.А.

Целью работы является разработка и анализ алгоритмов шумоподавления для повышения точности идентификации дикторов в условиях воздействия аддитивных шумов различных типов.

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

- Анализ существующих методов идентификации диктора, алгоритмов подавления шума в частотной области и способов объективной оценки качества речи с целью выбора прототипов для собственных решений.
- Разработка методики быстрой оценки точности идентификации дикторов и создание нового объективного показателя качества на основе нее для возможности быстрого подбора параметров алгоритмов шумоподавления в задаче идентификации диктора.
- Разработка новых алгоритмов подавления шума в речевых сигналах для повышения точности идентификации дикторов по сравнению с существующими решениями.
- Разработка программы для ЭВМ и исследование разработанных алгоритмов с ее помощью.

Методы исследования. При решении поставленных задач применялись методы математического анализа, линейной алгебры и аналитической геометрии, теории вероятности и математической статистики, цифровой обработки сигналов, спектрального анализа. Для исследования разработанных алгоритмов применялись методы математического и компьютерного моделирования.

Объектом исследований являются системы автоматической идентификации диктора с модулем предварительной обработки входных сигналов.

Предметом исследования являются методы и алгоритмы идентификации диктора, шумоподавления в частотной области, оценки качества речевых сигналов.

Научная новизна. Впервые получены следующие научные результаты:

- Произведена оценка тесноты статистической связи между точностью идентификации дикторов для двух баз речевых сигналов и показателями качества речи: PESQ, отношение сигнал/шум, сегментное отношение сигнал/шум, LLR, WSS.
- Разработан объективный показатель качества речевых сигналов, позволяющий оценить эффективность работы алгоритма шумоподавления в задаче идентификации диктора.
- Разработана методика быстрой оценки точности идентификации дикторов.
- Предложен новый подход к оценке мягкой маски, который может стать прототипом для широкого класса алгоритмов шумоподавления.

– Разработан новый двухступенчатый алгоритм на основе мягкой маски и функции коррекции спектра минимальной среднеквадратичной ошибки кратковременной амплитуды спектра.

Практическая значимость

– Методика быстрой оценки точности идентификации дикторов позволяет подбирать параметры алгоритмов шумоподавления быстрее, чем при использовании прямой оценки с помощью системы идентификации диктора. В частном случае достигнуто ускорение приблизительно в 88 раз.
– Предложенный двухступенчатый алгоритм на основе мягкой маски и функции коррекции спектра минимальной среднеквадратичной ошибки кратковременной амплитуды спектра позволяет повысить точность идентификации дикторов в среднем (среди ОСШ 5 дБ, 10 дБ, 15 дБ) для АБГШ на 13,4 процентных пункта по сравнению с алгоритмом на основе подхода прямого принятия решения и функции коррекции спектра Винера.
– Разработана программа «Speaker Recognition Test Framework – программа для исследования алгоритмов распознавания диктора» (свидетельство о государственной регистрации программы для ЭВМ № 2015660245), предназначенная для исследования алгоритмов распознавания диктора (идентификации и верификации) в условиях шумов.
– Разработана программа «NN-SCG speech recognition – научно-исследовательская программа по изучению алгоритмов нейросетевого дикторонезависимого распознавания речевых команд» (свидетельство о государственной регистрации программы для ЭВМ № 2015616920), с помощью которой может быть проведен анализ предложенных алгоритмов шумоподавления в задаче дикторонезависимого распознавания речевых команд.

Результаты работы внедрены в соответствующие разработки ООО «Оскар» (г. Ярославль) и ООО «Эймс Софтвэр» (г. Ярославль). Отдельные результаты диссертационной работы внедрены в учебный процесс Ярославского государственного университета им. П. Г. Демидова в рамках дисциплины «Цифровая обработка речевых сигналов». Все результаты внедрения подтверждены соответствующими актами.

Достоверность материалов диссертационной работы подтверждена согласованностью результатов математического моделирования разработанных алгоритмов и экспериментальной проверки в условиях компьютерного моделирования с использованием реальных речевых сигналов, апробацией в печати и на научно-практических конференциях различного уровня.

Апробация работы. Результаты работы докладывались и обсуждались на следующих конференциях:

– 14-й и 15-й Международной конференции «Цифровая обработка сигналов и её применение», Москва, 2012–2013;

- Международной конференции «Системы синхронизации, формирования и обработки сигналов в инфокоммуникациях», Ярославль, 2013;
- 11-й и 12-й Международных научно-технических конференциях «Оптико-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации», Курск, 2013, 2015;
- Международной конференции «Перспективные технологии в средствах передачи информации», Владимир, 2013;
- Международной научно-практической молодежной конференции «Путь в науку», Ярославль, 2013–2015;
- 66-й Всероссийской НТК студентов, магистрантов и аспирантов с международным участием, Ярославль, 2013;
- 69-й Международной конференции «Радиоэлектронные устройства и системы для инфокоммуникационных технологий», Москва, 2014;
- 15-й Всероссийской научно-практической конференции «Проблемы развития и применения средств противовоздушной обороны на современном этапе», Ярославль, 2014.

Публикации. По теме диссертации опубликовано 19 научных работ, из них 3 статьи в журналах, рекомендованных ВАК для публикации результатов кандидатских и докторских диссертаций, 16 докладов на научных конференциях; получено 2 свидетельства о регистрации программы для ЭВМ.

Личный вклад автора. Выносимые на защиту положения предложены и реализованы автором самостоятельно в ходе выполнения научно-исследовательских работ на кафедре динамики электронных систем Ярославского государственного университета им. П. Г. Демидова.

Структура и объем работы. Диссертация состоит из введения, трех глав, заключения, списка литературы и двух приложений. Содержание работы изложено на 133 страницах. Список литературы включает 102 наименования. В работе представлено 24 рисунка и 37 таблиц.

Основные научные положения и результаты, выносимые на защиту

- Методика быстрой оценки точности идентификации дикторов, позволяющая подбирать параметры алгоритмов шумоподавления быстрее, чем при использовании системы идентификации диктора.
- Новый подход к оценке мягкой маски, который может стать прототипом для широкого класса алгоритмов шумоподавления.
- Двухступенчатый алгоритм на основе мягкой маски и функции коррекции спектра минимальной среднеквадратичной ошибки кратковременной амплитуды спектра.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность выбранной темы, сформулированы цель и задачи исследования, изложены основные положения, выносимые на защиту, показана научная новизна и практическая значимость работы.

В первой главе проведен обзор современных способов идентификации диктора и подробно описана система на основе мел-частотных кепстральных коэффициентов и моделей гауссовых смесей с применением универсальной фоновой модели. Данная система наиболее широко применяется для задачи идентификации диктора и выбрана автором в качестве объекта исследований.

Рассмотрена проблема идентификации диктора в условиях шумов. Для повышения устойчивости системы идентификации диктора к шумам выбран такой известный способ, как предварительная обработка входных сигналов с помощью алгоритмов шумоподавления. Отмечено, что на текущий момент применение данного способа ограничивается использованием довольно простых алгоритмов шумоподавления: на основе ФКС спектрального вычитания; на основе подхода прямого принятия решения и ФКС Винера.

Проведен обзор способов подавления шума в частотной области. Рассмотрены такие проблемы, как моделирование речи и шума, выбор функции коррекции спектра, оценка априорного отношения сигнал/шум. Рассмотрено использование бинарных масок для подавления шума в речевых сигналах, а также новое направление в шумоподавлении – мягкие маски.

Поставлена проблема оценки качества речевых сигналов. Рассмотрены наиболее известные показатели качества речи – ОСШ, СегОСШ, LLR, WSS, PESQ.

Во второй главе поставлена задача подбора параметров алгоритмов шумоподавления для использования в системе идентификации диктора. Если решать эту задачу прямо и рассчитывать так называемую точность идентификации дикторов (ТИД), то эксперимент займет много времени из-за большого числа обрабатываемых тестовых сигналов.

Предложены новые показатели качества речи на основе расстояния между мел-частотными кепстральными коэффициентами (МЧКК) незашумленного сигнала и зашумленного:

$$Q^{МЧКК} = \frac{1}{W}\sum_{w=1}^{W} d\left(\vec{x}_w^R, \vec{x}_w^A\right),$$

где $\vec{x}_w^R$ – вектор МЧКК исследуемого сигнала для окна *w*, $\vec{x}_w^A$ – вектор МЧКК незашумленного сигнала для окна *w*, $d\left(\vec{x}_w^R, \vec{x}_w^A\right)$ – расстояние между векторами $\vec{x}_w^R$ и $\vec{x}_w^A$. На основе евклидова расстояния получен показатель качества речи МЧКК-Э, на основе расстояния городских кварталов – МЧКК-L1, на основе расстояния Махаланобиса – МЧКК-М.

Исследована теснота статистической связи между ТИД и показателями качества речи PESQ, ОСШ, СегОСШ, WSS, LLR, МЧКК-Э, МЧКК-L1, МЧКК-М. Для исследования использовался тип шума АБГШ, а также 2 шума из фонотеки NOISEX-92: «Speech babble» (далее – SB) и Vehicle interior noise (далее – VIN). Значение ОСШ изменялось от 6 до 15 дБ с шагом 1 дБ. Использовалось 10 алгоритмов предобработки. Измерения производились для двух баз речевых сигналов – «РУС-31-5» и «АНГЛ-20-5».

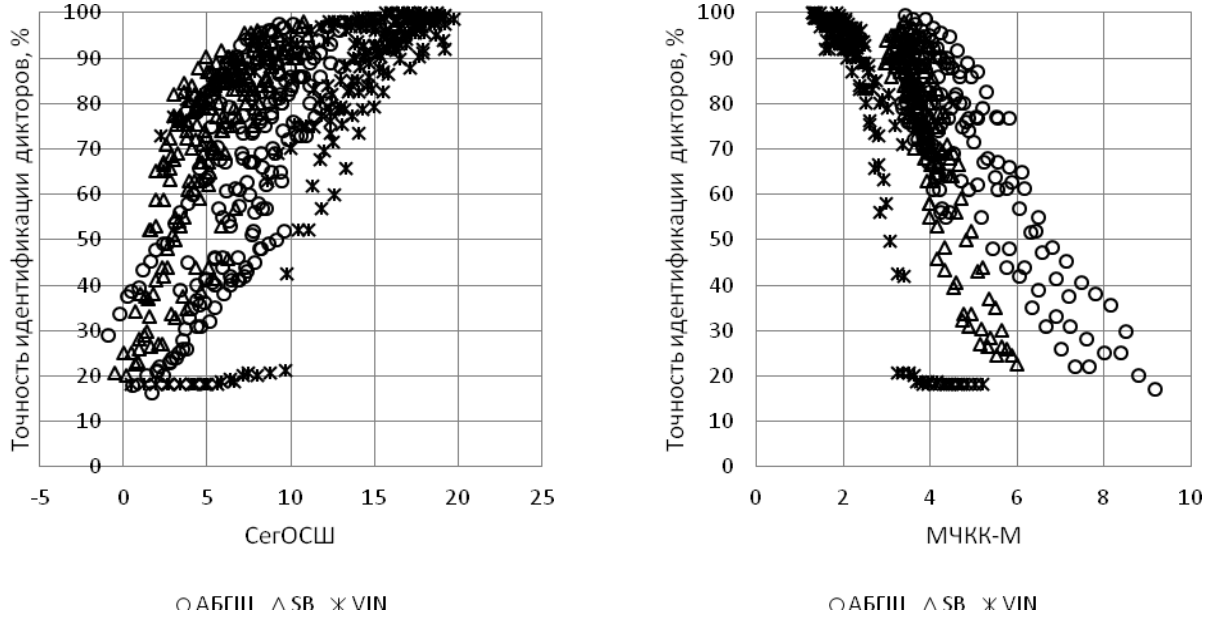
Рассчитан линейный коэффициент корреляции между каждым из показателей качества речи и ТИД для двух случаев – с использованием алгоритмов, избыточно подавляющих шум, и без использования. Результаты исследования приведены в табл. 1.

Таблица 1

Значения линейного коэффициента корреляции с использованием/без использования алгоритмов, избыточно подавляющих шум

Показатель качества	АБГШ	SB	VIN	Среднее
PESQ	0,88 / 0,83	0,72 / 0,65	0,60 / 0,66	0,73 / 0,71
ОСШ	0,87 / 0,87	0,66 / 0,75	0,64 / 0,73	0,72 / 0,79
СегОСШ	0,72 / 0,70	0,80 / 0,78	0,76 / 0,77	0,76 / 0,75
WSS	-0,36 / -0,08	-0,15 / 0,16	-0,60 / -0,71	-0,37 / -0,21
LLR	-0,89 / -0,92	-0,47 / -0,27	-0,29 / -0,55	-0,55 / -0,58
МЧКК-Э	-0,87 / -0,87	-0,50 / -0,33	-0,87 / -0,92	-0,75 / -0,71
МЧКК-L1	-0,78 / -0,72	-0,32 / -0,11	-0,87 / -0,91	-0,66 / -0,58
МЧКК-М	-0,45 / -0,85	-0,56 / -0,91	-0,89 / -0,92	-0,64 / -0,89

Таким образом, наибольшую тесноту связи с ТИД без использования алгоритмов, избыточно подавляющих шум, имеет предлагаемый показатель качества речи МЧКК-М. При использовании полного набора алгоритмов предобработки наибольшую тесноту статистической связи с ТИД имеет показатель качества речи СегОСШ. Однако значение линейного коэффициента корреляции в обоих случаях может оказаться недостаточно большим для точного подбора параметров алгоритмов шумоподавления. Диаграммы рассеяния значений СегОСШ и МЧЧК-М приведены на рис. 1.



а) б)

Рис. 1. Диаграмма рассеяния значений СегОСШ (а) и МЧКК-М (б) относительно ТИД

Разработан новый показатель качества речи, аппроксимирующий оценку ТИД на основе линейной комбинации PESQ, ОСШ, СегОСШ, WSS, LRR, МЧКК-Э, МЧКК-L1, МЧКК-М, которая получена с помощью линейной регрессии для каждого типа шума. Предлагается для нового показателя качества использовать название «альтернативная оценка точности идентификации дикторов» (АОТИД).

Формулы расчета АОТИД для используемых типов шума приведены ниже:

$$Q^{АОТИД-АБГШ} = \left[-0{,}352\ \ 0{,}03\ \ -0{,}016\ \ -0{,}01\ \ -0{,}15\ \ -0{,}336\ \ 0{,}161\ \ 0{,}002\ \ 0{,}224\right] * A;$$

$$Q^{АОТИД-SB} = \left[-0{,}405\ \ -0{,}027\ \ 0{,}019\ \ -0{,}002\ \ 0{,}624\ \ -0{,}473\ \ 0{,}145\ \ 0{,}002\ \ 0{,}555\right] * A;$$

$$Q^{АОТИД-VIN} = \left[1{,}106\ \ -0{,}036\ \ 0{,}049\ \ 0{,}027\ \ -0{,}27\ \ -0{,}25\ \ 0{,}042\ \ -0{,}132\ \ 0{,}034\right] * A;$$

$$A = \left[Q^{PESQ}\ \ Q^{ОСШ}\ \ Q^{СегОСШ}\ \ Q^{WSS}\ \ Q^{LLR}\ \ Q^{МЧКК-Э}\ \ Q^{МЧКК-L1}\ \ Q^{МЧКК-М}\ \ 1\right]^T.$$

Значение линейного коэффициента корреляции для АБГШ составляет 0,96, для шума SB – 0,94, для шума VIN – 0,97. Диаграммы рассеяния значений новых показателей качества речи для соответствующих типов шума приведены на рис. 2.

Весьма высокая (по шкале Чеддока) теснота статистической связи между АОТИД и ТИД доказана при использовании всего набора тестовых сигналов баз «РУС-31-5» и «АНГЛ-20-5». В то же время общее число тестовых сигналов оказывается достаточным большим (155 сигналов для базы «РУС-31-5» и 100 для базы «АНГЛ-20-5»), что влечет за собой увеличение вычислительной сложности предлагаемой методики оценки ТИД.

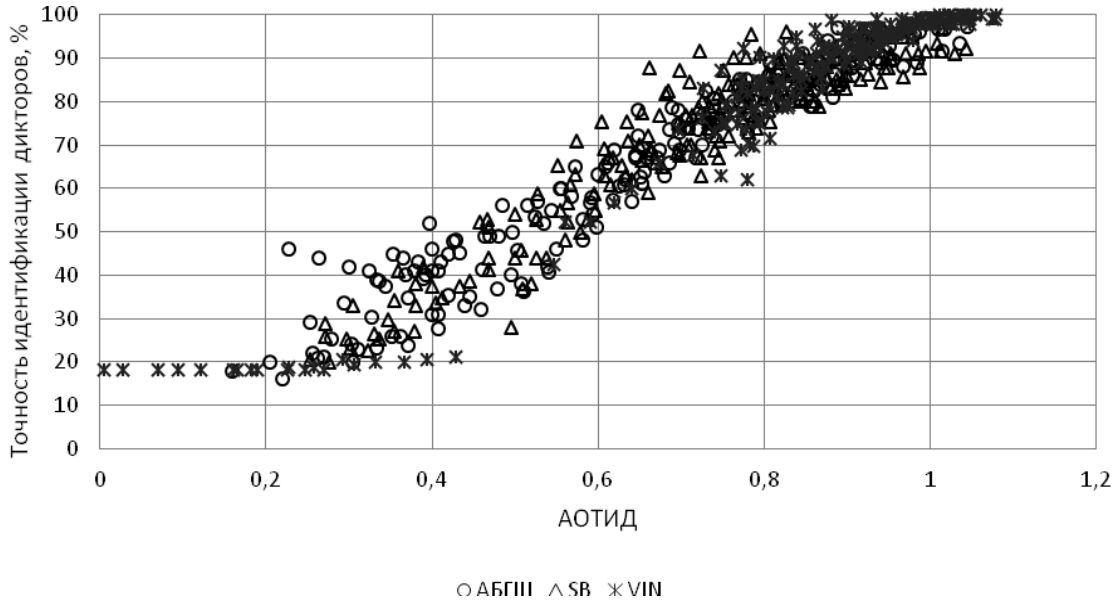


Рис. 2. Диаграммы рассеяния значений АОТИД относительно значений ТИД

Поставлена задача модифицировать разработанный показатель качества для быстрой аппроксимации среднего значения ТИД для двух используемых баз речевых сигналов. Модифицированный показатель качества речи не будет

использовать PESQ в линейной комбинации, т. к. последний обладает относительно высокой вычислительной сложностью.

Первым этапом предлагается для каждого тестового сигнала составить показатель качества, являющийся линейной комбинацией ОСШ, СегОСШ, WSS, LLR, МЧКК-Э, МЧКК-L1, МЧКК-М, и рассчитать линейные коэффициенты корреляции между полученными показателями качества и ТИД. Линейные коэффициенты корреляции рассчитываются отдельно для двух используемых баз речевых сигналов. Для каждого типа шума также проводится отдельное исследование.

Далее в каждой из речевых баз выбирается 30 тестовых сигналов, при которых получены наибольшие коэффициенты корреляции для выбранного типа шума. Сигналы сортируются в порядке убывания линейного коэффициента корреляции. Полученный массив сигналов назовем «релевантные тестовые сигналы».

Берутся несколько первых сигналов из наборов релевантных тестовых сигналов каждой речевой базы и рассчитываются значения показателей качества ОСШ, СегОСШ, WSS, LLR, МЧКК-Э, МЧКК-L1, МЧКК-М. Значения каждого показателя качества усредняются среди выбранных тестовых сигналов. Далее с помощью линейной регрессии рассчитываются коэффициенты нового показателя качества для аппроксимации среднего значения ТИД среди двух используемых баз речевых сигналов. Для каждой речевой базы берется равное число сигналов из набора релевантных, поэтому всего получено 30 новых показателей качества для каждого типа шума. Каждый новый показатель качества обозначается по числу используемых для его получения релевантных тестовых сигналов из каждой базы. Далее рассчитывается значения линейного коэффициента корреляции между новыми показателями качества и средним значением ТИД среди двух используемых речевых баз. Из всех полученных показателей качества выбирается по одному для каждого типа шума. Предлагается для нового показателя качества использовать название «быстрая оценка точности идентификации дикторов» (БОТИД). В работе для расчета БОТИД используется по 4 тестовых сигнала из каждой базы (набор сигналов различен для каждого типа шума).

Диаграммы рассеяния для БОТИД и соответствующих типов шума приведены на рис. 3. Предложенные показатели качества имеют весьма высокую тесноту связи с усредненным ТИД для двух используемых баз речевых сигналов. Значение линейного коэффициента корреляции для каждого типа шума составляет 0,99.

Методика быстрой оценки точности идентификации дикторов использована для подбора параметра α двухступенчатого алгоритма шумоподавления на основе ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра. Для сравнения данный параметр также был подобран стандартным способом с использованием системы идентификации диктора. Значение параметра с максимальной усредненной ТИД среди значений ОСШ 5 дБ, 10 дБ, 15 дБ и типов шума АБГШ, SB, VIN составило 0,99. Аналогичный

результат получен при использовании методики на основе БОТИД. Но с помощью нее удалось подобрать параметр приблизительно в 88 раз быстрее, чем при оценке ТИД напрямую.

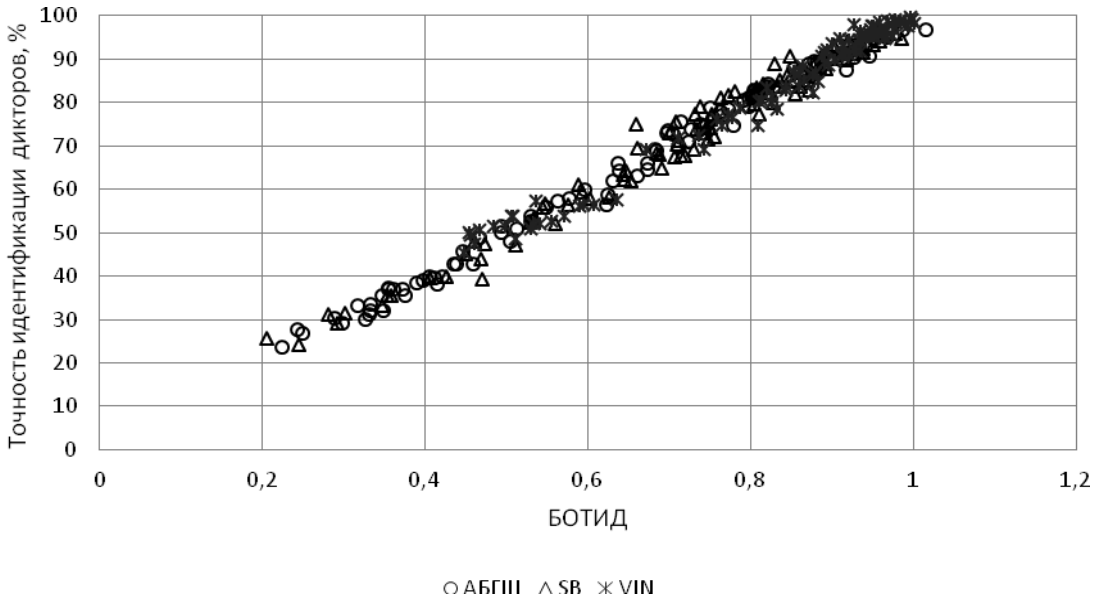


Рис. 3. Диаграмма рассеяния значений БОТИД относительно значений ТИД

<u>В третьей главе</u> предложен новый подход к расчету мягкой маски, основанный на определении вероятности присутствия речи в точках частотно-временного представления сигнала и модифицированном подходе прямого принятия решения. Рассчитать мягкую маску можно, последовательно применяя следующие формулы:

$$a = 2 - \alpha\,,\; b = -3(1-\alpha)R_{k,w},$$

$$c = \alpha\hat{A}^2_{k,w-1} + 3(1-\alpha)R^2_{k,w},\; d = -R_{k,w}\left(\alpha\hat{A}^2_{k,w-1} + (1-\alpha)R^2_{k,w}\right),$$

$$p = \frac{c}{a} - \frac{b^2}{3a^2}\,,\; q = \frac{2b^3}{27a^3} - \frac{bc}{3a^2} + \frac{d}{a},$$

$$Q = \left(\frac{p}{3}\right)^3 + \left(\frac{q}{2}\right)^2,\; \varphi = \sqrt[3]{\sqrt{Q} - \frac{q}{2}},$$

$$S^{\theta}_{k,w} = F^{Рэлей}_{k}\left(\varphi - \frac{p}{3\varphi} - \frac{b}{3a}\right)^{\theta},$$

где $S^{\theta}_{k,w}$ – обобщенная мягкая маска (обобщение предложено в работе); $R_{k,w}$ – амплитуда спектра зашумленного сигнала для частотной полосы k и текущего окна w; $\hat{A}_{k,w-1}$ – оценка амплитуды спектра незашумленного сигнала для частотной полосы k и предыдущего окна w-1; α, θ – параметры алгоритма; $F^{Рэлей}_{k}$ – функция распределения значений амплитуды спектра шума для частотной

полосы *k*. Амплитуда спектра шума моделируется как случайная величина с рэлеевским законом распределения.

С помощью методики БОТИД подобраны параметры α и θ данного алгоритма, для которых среднее значение БОТИД максимально среди АБГШ, шума SB, шума VIN, – 0,99 и 1 соответственно.

Полученный алгоритм шумоподавления на основе мягкой маски можно использовать как первый этап в модифицированном двухступенчатом алгоритме шумоподавления. В качестве функции коррекции спектра для второго этапа предлагается выбрать ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра.

Для второго этапа двухступенчатого алгоритма предлагается также, выполнить сглаживание априорного ОСШ с помощью экспоненциального скользящего среднего с ограничением сверху значения в текущем окне:

$$\hat{\xi}_{k,w}^{TSNR} = \varepsilon \cdot \min(\delta, \frac{\hat{A}_{k,w}^2}{\lambda_{k,w}^D}) + (1-\varepsilon) \cdot \hat{\xi}_{k,w-1}^{TSNR},$$

$$0 < \varepsilon \le 1,\ \delta >> 1,$$

где ε – сглаживающий параметр, который подбирается исходя из задачи; δ – ограничивающий параметр, предотвращающий переоценку априорного ОСШ; $\lambda_{k,w}^D$ – спектральная плотность шума. В работе принимается $\delta = \infty$. С помощью методики на основе БОТИД подобрано значение сглаживающего параметра ε для данного алгоритма – 0,75.

Проведено сравнение предложенных алгоритмов шумоподавления с существующими в задаче предобработки речевых сигналов для системы автоматической идентификации диктора. Используются следующие алгоритмы шумоподавления: 1) алгоритм на основе подхода прямого принятия решения ($\alpha = 0{,}98$) и ФКС Винера; 2) двухступенчатый алгоритм шумоподавления ($\alpha = 0{,}99$) на основе ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра; 3) предлагаемый алгоритм на основе мягкой маски ($\alpha = 0{,}99; \theta = 1$); 4) предлагаемый двухступенчатый алгоритм шумоподавления ($\alpha = 0{,}98$) на основе мягкой маски и ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра ($\alpha = 0{,}99; \theta = 1; \varepsilon = 0{,}75$).

Точность идентификации дикторов определяется для баз речевых сигналов «РУС-31-5» и «АНГЛ-20-5», значений ОСШ 5 дБ, 10 дБ, 15 дБ. Расчет ТИД повторяется 10 раз, результаты усредняются, и определяется доверительный интервал с доверительной вероятностью 0,95. Результаты для АБГШ приведены в табл. 2. Значения ТИД округляются до десятых долей процента. Лучшие результаты среди каждой речевой базы и значения ОСШ выделены жирным. Методика их определения следующая. 1) Выбирается алгоритм шумоподавления, для которого среднее значение ТИД наибольшее. Этот алгоритм пометим, как «наилучший». 2) Выбираются алгоритмы, для которых

доверительный интервал ТИД пересекается с доверительным интервалом ТИД «наилучшего» алгоритма.

Таблица 2

Оценка ТИД (%) для АБГШ

Алг.	ОСШ, дБ					
	5		10		15	
	РУС-31-5	АНГЛ-20-5	РУС-31-5	АНГЛ-20-5	РУС-31-5	АНГЛ-20-5
Нет	26,1 ± 0,3	15,6 ± 0,9	46,1 ± 0,6	25,7 ± 1,1	67,5 ± 0,5	70,5 ± 0,9
1	60,9 ± 1,1	56,3 ± 1,4	77,2 ± 0,7	71,9 ± 1,0	87,2 ± 1,1	80,1 ± 1,4
2	74,7 ± 1,0	63,1 ± 1,7	**90,6 ± 0,9**	82,7 ± 1,2	**97,9 ± 0,6**	93,2 ± 0,9
3	73,2 ± 0,7	63,7 ± 1,6	87,4 ± 1,3	**84,6 ± 1,3**	95,5 ± 0,6	92,2 ± 0,9
4	**77,0 ± 1,2**	**66,8 ± 1,0**	**91,6 ± 1,0**	**86,0 ± 1,2**	**97,8 ± 0,5**	**95,1 ± 0,5**

Также выполнялась оценка работы алгоритмов шумоподавления с помощью методики БОТИД. Значение БОТИД измерялось 10 раз, результаты усреднялись и округлялись до тысячных. В табл. 3 приведены усредненные среди двух баз речевых сигналов значения ТИД, а также значения БОТИД для АБГШ.

Таблица 3

БОТИД и усредненное значение ТИД (%) для АБГШ

Алг.	ОСШ, дБ							
	5		10		15		Среднее	
	БОТИД	ТИД	БОТИД	ТИД	БОТИД	ТИД	БОТИД	ТИД
	0,190	20,8	0,365	35,9	0,578	69,0	0,378	41,9
1	0,594	58,6	0,719	74,5	0,825	83,7	0,713	72,3
2	0,717	68,9	0,862	86,7	0,976	95,5	0,852	83,7
3	0,668	68,5	0,829	86,0	0,964	93,8	0,821	82,8
4	**0,720**	**71,9**	**0,869**	**88,8**	**0,993**	**96,5**	**0,861**	**85,7**

В условиях отсутствия шума (за исключением естественного шумового фона записи) в обеих базах речевых сигналов правильно идентифицированы все дикторы. Для АБГШ наиболее предпочтительным оказался предложенный двухступенчатый алгоритм на основе мягкой маски (алгоритм № 4). Отметим, что с помощью методики БОТИД удалось точно предсказать выбор предпочтительного алгоритма.

В табл. 4 приведены усредненные среди используемых баз речевых сигналов и значений ОСШ результаты для АБГШ, шума SB, шума VIN.

Таблица 4

Среднее значение ТИД (%) и БОТИД среди используемых ОСШ

Алг.	Тип шума							
	АБГШ		SB		VIN		Среднее	
	БОТИД	ТИД	БОТИД	ТИД	БОТИД	ТИД	БОТИД	ТИД
Нет	0,378	41,9	0,441	43,9	0,486	47,0	0,435	44,2
1	0,713	72,3	0,822	78,8	**0,960**	**96,3**	0,831	82,5
2	0,852	83,7	**0,837**	**80,4**	0,933	94,1	0,874	86,1
3	0,821	82,8	0,817	78,6	0,955	94,8	0,864	85,4
4	**0,861**	**85,7**	0,834	79,1	0,939	95,9	**0,878**	**86,9**

Таким образом, наибольшее среднее значение ТИД среди ОСШ 5 дБ, 10дБ, 15 дБ, типов шумов АБГШ, SB, VIN и баз речевых сигналов «РУС-31-5» и «АНГЛ-20-5» получено при использовании предложенного двухступенчатого алгоритма на основе мягкой маски и ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра.

В заключении подводятся итоги выполненной работы.

В приложениях приведены копии актов о внедрении результатов работы.

ОСНОВЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

На основании проведенных исследований в области цифровой обработки речевых сигналов в работе получены следующие результаты:

1. Предложены новые объективные показатели качества речи на основе расстояния между МЧКК незашумленного сигнала и зашумленного. На основе евклидова расстояния получен показатель качества речи МЧКК-Э, на основе расстояния городских кварталов – МЧКК-L1, на основе расстояния Махаланобиса – МЧКК-М.

2. Исследована теснота статистической связи между ТИД и показателями качества речи PESQ, ОСШ, СегОСШ, WSS, LLR, МЧКК-Э, МЧКК-L1, МЧКК-М. При использовании алгоритмов, избыточно подавляющих шум, наибольшей теснотой статистической связи с ТИД для АБГШ, шума SB, шума VIN обладает показатель качества речи СегОСШ. Для АБГШ значение линейного коэффициента корреляции составляет 0,72; для шума SB – 0,8; для шума VIN – 0,76. Без использования алгоритмов, избыточно подавляющих шум, наибольшей теснотой статистической связи с ТИД обладает предложенный показатель качества речи МЧКК-М. Для АБГШ значение линейного коэффициента корреляции составляет -0,85; для шума SB – -0,91; для шума VIN – -0,92.

3. Предложен новый показатель качества речи АОТИД (альтернативная оценка точности идентификации дикторов) на основе линейной комбинации PESQ, ОСШ, СегОСШ, WSS, LLR, МЧКК-Э, МЧКК-L1, МЧКК-М. Весовые коэффициенты для каждого используемого показателя качества речи в линейной комбинации подобраны индивидуально для АБГШ, шума SB, шума VIN. Для АБГШ значение линейного коэффициента корреляции между АОТИД для соответствующего типа шума и ТИД составляет 0,96; для шума SB – 0,94; для шума VIN – 0,97.

4. Предложена методика быстрой оценки усредненной среди двух используемых баз речевых сигналов точности идентификации дикторов – БОТИД. Значение линейного коэффициента корреляции между БОТИД и усредненной ТИД составило 0,99 для АБГШ, шума SB, шума VIN.

5. Методика на основе БОТИД использована для подбора параметра двухступенчатого алгоритма шумоподавления на основе ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра. Для сравнения данный параметр также подобран стандартным способом с

использованием системы идентификации диктора. Значение параметра с максимальной усредненной ТИД среди значений ОСШ 5 дБ, 10 дБ, 15 дБ и типов шума АБГШ, SB, VIN совпало со значением, полученным при использовании методики на основе БОТИД. Но с помощью нее удалось подобрать параметр приблизительно в 88 раз быстрее, чем при оценке ТИД напрямую.

6. Обобщено понятие мягкой маски, математически показана правомерность такого обобщения, пояснен физический смысл параметра θ – показателя степени обобщенной мягкой маски.

7. Предложен новый подход к расчету мягкой маски, основанный на определении вероятности присутствия речи в точках частотно-временного представления сигнала и модифицированном подходе прямого принятия решения. Подобраны параметры α и θ данного алгоритма, для которых среднее значение БОТИД максимально среди АБГШ, шума SB, шума VIN, – 0,99 и 1 соответственно. Новый подход к расчету мягкой маски может стать прототипом для широкого класса алгоритмов шумоподавления.

8. Предложена модификация двухступенчатого алгоритма шумоподавления, которая использует сглаживание априорного ОСШ, полученного на втором этапе алгоритма, с помощью экспоненциального скользящего среднего с ограничением сверху значения в текущем окне. Для двухступенчатого алгоритма на основе ФКС среднеквадратичной ошибки кратковременной амплитуды спектра данная модификация не способна серьезно повысить точность идентификации дикторов. Однако предложен двухступенчатый алгоритм на основе мягкой маски и ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра, для которого данная модификация повышает ТИД. Подобрано значение сглаживающего параметра ε для данного алгоритма – 0,75.

9. Произведено сравнение предложенных алгоритмов в задаче идентификации диктора. Для АБГШ наиболее предпочтительным оказался предложенный двухступенчатый алгоритм шумоподавления на основе мягкой маски и ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра. Усредненное по используемым ОСШ и базам речевых сигналов значение ТИД для него на 13,4 процентных пункта (п. п.) выше, чем для алгоритма на основе подхода прямого принятия решения и ФКС Винера, и на 2,9 п. п. выше, чем для двухступенчатого алгоритма на основе ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра.

10. Для шума SB наибольшее значение ТИД обеспечивает двухступенчатый алгоритм шумоподавления на основе ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра. Усредненное по используемым ОСШ и базам речевых сигналов значение ТИД для него на 1,6 п. п. выше, чем для алгоритма на основе подхода прямого принятия решения и ФКС Винера, и на 0,5 п. п. выше, чем для предложенного двухступенчатого алгоритма на основе

мягкой маски и ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра.

11. Для шума VIN наиболее предпочтительным оказался алгоритм шумоподавления на основе подхода прямого принятия решения и ФКС Винера. Усредненное по используемым ОСШ и базам речевых сигналов значение ТИД для него на 1,5 п. п. выше, чем для двухступенчатого алгоритма на основе ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра, и на 0,4 п. п. выше, чем для предложенного двухступенчатого алгоритма на основе мягкой маски и ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра.

12. Наибольшее среднее значение ТИД среди используемых ОСШ, типов шумов АБГШ, SB, VIN и баз речевых сигналов обеспечивает предложенный двухступенчатый алгоритм на основе мягкой маски и ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра. Среднее значение ТИД для него на 4,4 п. п. выше, чем для алгоритма на основе подхода прямого принятия решения и ФКС Винера, и на 1,5 п. п. выше, чем для двухступенчатого алгоритма на основе ФКС минимальной среднеквадратичной ошибки кратковременной амплитуды спектра.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в журналах из перечня ВАК

1. Сагациян, М.В. Повышение эффективности коллективного нейросетевого алгоритма в задаче дикторонезависимого распознавания речевых команд в условиях шумов / М.В. Сагациян, Г.С. Тупицин, С.А. Кравцов, А.Л. Приоров // Информационные системы и технологии. – 2015. – № 4. – С. 39–46.
2. Сагациян, М.В. Разработка и исследование нейросетевого алгоритма дикторонезависимого распознавания речевых команд / М.В. Сагациян, А.В. Куликов, Г.С. Тупицин // Вестник Поволжского государственного технологического университета. – 2014. – Т. 20. – № 1. – С. 62–68.
3. Сагациян, М.В. Анализ эффективности нейросетевых алгоритмов в задаче дикторонезависимого распознавания речевых команд / М.В. Сагациян, Г.С. Тупицин // Информационные системы и технологии. – 2015. – № 3. – С. 16–26.

Материалы российских и международных конференций

4. Скопинцев, Я.М. Использование бинарных масок для повышения качества закрытой текстонезависимой идентификации диктора в условиях шумов / Я.М. Скопинцев, Г.С. Тупицин // Доклад 69-й Международной конференции «Радиоэлектронные устройства и системы для инфокоммуникационных технологий». – Москва, 2014. – С. 392–395.
5. Веселов, И.А. Использование априорного отношения сигнал/шум для построения бинарных масок в задаче подавления шума в речевых сигналах / И.А. Веселов, А.В. Куликов, Я.М. Скопинцев, Г.С. Тупицин // Доклад 15-й

международной конференции «Цифровая обработка сигналов и её применение». – Москва, 2013. – С. 246–249.
6. Тупицин, Г.С. Повышение качества идентификации диктора в условиях шумов с помощью бинарных масок / Г.С. Тупицин, А.В. Куликов, М.В. Сагациян // доклад международной конференции «Перспективные технологии в средствах передачи информации». – Владимир, 2013. – С. 180-182.
7. Кравцов, С.А. Алгоритм неэталонной оценки степени зашумлённости речевых сигналов / С.А. Кравцов, Г.С. Тупицин, М.В. Сагациян, А.В. Куликов // Доклад 14-й международной конференции «Цифровая обработка сигналов и её применение». – Москва, 2012. – С. 177–179.
8. Куликов, А.В. Использование априорного отношения сигнал/шум для построения бинарных масок в задаче идентификации диктора / А.В. Куликов, М.В. Сагациян, Г.С. Тупицин // Доклад международной конференции «Системы синхронизации, формирования и обработки сигналов в инфокоммуникациях». – Ярославль, 2013. – С. 168–170.
9. Тупицин, Г.С. Повышение качества закрытой текстонезависимой идентификации диктора в условиях шумов с помощью бинарных масок / Г.С. Тупицин, М.В. Сагациян // Доклад 12-й международной научно-технической конференции «Оптико-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации». – Курск, 2015. – С. 376–378.
10. Кравцов, С.А. Алгоритм обнаружения речевой активности на основе моделей гауссовых смесей / С.А. Кравцов, Г.С. Тупицин // Доклад 15 всероссийской научно-практической конференции «Проблемы развития и применения средств противовоздушной обороны на современном этапе». – Ярославль, 2014. – С. 39–44.
11. Куликов, А.В. Зависимость точности дикторонезависимого распознавания речевых команд базовым нейросетевым алгоритмом от количества обучающих дикторов / А.В. Куликов, М.В. Сагациян, Г.С. Тупицин // Доклад международной конференции «Системы синхронизации, формирования и обработки сигналов в инфокоммуникациях». – Ярославль, 2013. – С. 119-121.

Свидетельства о государственной регистрации программы для ЭВМ

12. Тупицин, Г.С. Speaker Recognition Test Framework – программа для исследования алгоритмов распознавания диктора / Г.С. Тупицин, А.И. Топников, А.Л. Приоров // Свидетельство о государственной регистрации программы для ЭВМ № 2015660245 от 25 сентября 2015 г.
13. Сагациян, М.В. NN-SCG speech recognition – научно-исследовательская программа по изучению алгоритмов нейросетевого дикторонезависимого распознавания речевых команд / М.В. Сагациян, Г.С. Тупицин // Свидетельство о государственной регистрации программы для ЭВМ № 2015616920 от 30 апреля 2015 г.

Подписано в печать «20» октября 2015 г.
Формат 60×84 1/16. Тираж 100 экз.

Ярославский государственный университет
150000, Ярославль, ул. Советская, 14.