

ОТЗЫВ

официального оппонента на диссертацию Ермилова Алексея Валерьевича «Методы, алгоритмы и программы решения задач идентификации языка и диктора», представленную на соискание ученой степени кандидата физико-математических наук по специальности 05.13.11 – «Математическое обеспечение вычислительных машин, комплексов и компьютерных сетей»»

Актуальность работы. В диссертации А. В. Ермилова был исследован ряд вопросов, касающихся идентификации диктора и языка. В настоящее время в условиях постоянно усиливающейся угрозы терроризма идентификация ключевых слов имеет очень большое значение. С другой стороны, разработка голосового интерфейса с авторизацией голоса имеет не менее важное значение для людей с ограниченными физическими способностями, например для управления автомобилем или другими домашними устройствами. Таким образом, актуальность данной работы несомненна.

Целью диссертации А. В. Ермилова является исследование акустических свойств речевого сигнала и на их основе разработка и теоретическое обоснование методов, алгоритмов и программ идентификации диктора и языка.

Научная новизна работы обусловлена новым подходом к выделению системы характерных признаков для распознавания языка с применением распределения Грам-Шарлье, модификации метода опорных векторов для повышения точности распознавания диктора на основе введения в базовый алгоритм Фишеровских ядер, модификации алгоритма симуляции отжига для повышения быстродействия и точности получения признаков, применяемых для распознавания языка, а также алгоритмическими и программными решениями, которые позволили достичь поставленной цели.

Научная и практическая ценность диссертации подтверждается тем, что диссертант не только разработал методы, алгоритмы и написал программы, но и практически реализовал их, о чем имеется акт о внедрении.

Диссертация состоит из введения, четырех глав, заключения и приложения.

В первой главе автор изложил физические аспекты акустического сигнала, его характеристики и особенности. Приведена модель речеобразования и введено понятие речевого тракта. Рассмотрены общие принципы генерации и восприятия звукового сигнала. Также даётся характеристика некоторых подходов к использованию методов распознавания речи. Кроме того, исследованы источники вариативности речевого сигнала, затрудняющие процесс идентификации, например физиологические или эмоциональные факторы.

Результатом этой главы можно считать выбранные характеристики речевого сигнала, которые далее используются в диссертации для распознавания языка и диктора.

Во второй главе диссертант исследует различные математические модели, используемые для построения систем распознавания языка и диктора, базирующиеся на аппарате Скрытых Марковских Моделей, при этом особое внимание уделено методам, применяемым для разработки системы распознавания языка, точность работы которой не зависит от диктора. Приводится способ построения дикторонезависимых признаков для описания речевого сигнала, опирающийся на психоакустическую модель восприятия человеком речевого сообщения. Для построения системы дикторонезависимых признаков речевого сигнала А.В. Ермилов использовал т.н. Auditory Image Model, разработанную Р. Петерсоном, выходом которой является нормализованный спектр сигнала. Для моделирования огибающей этого спектра диссертантом впервые предложено использование расширения Грам-Шарлье вместо обычно используемой гауссовской смеси.

Третья глава содержит результаты различных аспектов реализации системы идентификации языка и диктора с применением предложенных методов. Изложена структура такой системы и детально рассмотрены этапы идентификации языка и диктора. Приведена модель такой системы с использованием языка UML (Unified Modeling Language). Рассмотрены составляющие эту модель классы и процедуры. Рассмотрены аспекты использования кон-

вейерной обработки данных для повышения быстродействия работы всей системы.

В четвертой главе представлены результаты экспериментов с реальными и симулированными данными с применением моделей, описанных во второй главе. Приводятся результаты экспериментов по распознаванию в различных каналах: микрофонном, телефонном, а также GSM-канале. Изложено объяснение полученных результатов с различных точек зрения и с точки зрения практических ограничений. Результаты анализа проведенных экспериментов подтверждают практическую применимость разработанных алгоритмов и их эффективность. Далее автором проводятся исследования по сравнению качества подгонки распределений при различных аппроксимациях спектров.

В Заключении подводятся итоги проделанной работы, перечисляются основные результаты и следующие из них выводы.

Несмотря на хорошее качество представленной работы, она не лишена недостатков. Отметим некоторые из них.

1. В работе замечены грамматические и орфографические ошибки.
2. В качестве альтернативы распараллеливанию можно было бы рассмотреть возможности применения графических плат для повышения быстродействия.
3. В литературе имеются упоминания о системах распознавания дикторов на основе использования нейронных сетей, не рассмотренные в диссертации.
4. Кроме рассмотренных в диссертации методов имеется еще и универсальная фоновая модель (Universal Background Model - UBM), которую желательно было бы исследовать тоже.

Отмеченные недостатки не снижают высокой оценки диссертационной работы. В целом диссертацию А. В. Ермилова можно квалифицировать как законченное решение научно-практической задачи. Достоверность результатов подтверждается работающим прототипом и актом о внедрении.

Разработанные модели распознавания могут быть эффективно использо-

ваны для различных практических приложений, связанных с распознаванием дикторов. Дальнейшие исследования в этой области могут быть направлены на разработку текстонезависимых алгоритмов распознавания, в том числе, на создание более эффективных систем распознавания дикторов.

По результатам исследований диссертантом были опубликованы 5 статей (3 из которых в журналах из списка ВАК, одна в международном реферируемом журнале), 6 тезисов на международных конференциях. Автореферат правильно и полно отражает содержание диссертации.

Таким образом, диссертация А. В. Ермилова на соискание ученой степени кандидата физико-математических наук является научно-квалификационной работой, в которой содержится решение задачи по разработке методов, алгоритмов и программ по идентификации диктора и языка, имеющей существенное значение для разработки акустических интерфейсов и развития биометрии, что соответствует требованиям п. 9 «Положения о порядке присуждения ученых степеней» от 24.09.2013 г. № 842, а ее автор Ермилов Алексей Валерьевич заслуживает присуждения искомой ученой степени.

Д. т. н., старший научный сотрудник
Федерального государственного бюджетного
учреждения науки «Институт Высшей Нервной
Деятельности и Нейрофизиологии РАН»



А.А. Харламов



Подпись т. Харламова А.А.
УДОСТОВЕРЯЮ
Зав. кани. ИВНД и НФ Харламов