

ОТЗЫВ

официального оппонента на диссертацию Ермилова Алексея Валерьевича «Методы, алгоритмы и программы решения задач идентификации языка и диктора», представленной 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 рассмотрена проблема идентификации диктора и языка по аудио данным. Теоретическая и практическая разработка акустических моделей и методов для аудио идентификации является очень актуальной задачей в связи с увеличением спроса на услуги голосовой биометрии на предприятиях и государственных организациях, на автоматизацию дистанционных услуг call-центров в различных сферах, на голосовые интерфейсы для управления мобильными и бытовыми устройствами, в связи с большим интересом в использовании идентификации и распознавания речи в судебно-медицинских и военных организациях, в оперативной работе правоохранительных органов с целью анализа угроз и предотвращения террористических актов.

Целью диссертации А. В. Ермилова является разработка моделей, методов, алгоритмов и программ идентификации языка и диктора путем выделения из акустического сигнала характерных признаков, построение голосовых моделей в признаковом пространстве, удобном для классификации и дальнейшей идентификации.

Научная новизна работы заключается в исследовании и в использовании ряда методов для повышения качества идентификации языка и диктора, основанных на ряде интересных алгоритмических решений. В работе предложено использовать четырехпараметрическое распределение, полученное на основе разложения

Грам-Шарлье, для представления характерных признаков в акустической модели, которая исследована и использована для распознавания языка. Для улучшения разделимости мел-частотных кепстральных акустических признаков (MFCC) в задаче идентификации диктора предложено использование спрямляющего пространства ядра Фишера для классификации методом опорных векторов с экспоненциальным ядром. Для задач оптимизации акустической модели предложена модификация алгоритма симуляции отжига с использованием параллельных вычислений.

Практическая ценность диссертации определяется разработанными алгоритмами, которые были реализованы в программном обеспечении, показали преимущество перед известными решениями и были успешно внедрены.

Диссертация состоит из введения, четырех глав, заключения и приложения

В первой главе изложены принципы и методы обработки голосовых сигналов, приведен обзор основных акустических моделей речевого и слухового тракта, признаков описаний речевых сигналов с учетом психофизиологических моделей. В главе подробно описаны модели речевой акустической системы на основе линейного фильтра, реализуемого кодированием с линейным предсказанием (Linear Prediction Coding, LPC), и мел-частотных кепстральных коэффициентов (Mel Frequency Cepstral Coefficients, MFCC), определены основные методы выделения акустических признаков, которые могут использоваться для построения голосовых моделей с целью распознавания языка и диктора.

Во второй главе А.В. Ермаковым проанализированы различные математические модели, используемые для построения систем распознавания языка и диктора. Описан подход использования скрытых марковских моделей для задачи распознавания последовательности фонем, которые могут использоваться для идентификации языка, разобраны сопутствующие этой модели задачи. Изложен способ выделения дикторонезависимых признаков для описания речевого сигнала, базирующийся на психоакустической модели восприятия человеком речевого сообщения. В качестве

базовой признаковой модели диссертантом была использована Auditory Image Model (AIM), предложенная Р. Петерсоном, выходом которой является нормализованный спектр речевого сигнала, моделирующий сигнал слухового нерва, поступающего в мозг. В подобных моделях для параметризации признаков нормализованного спектра используется аппроксимация огибающей спектра с помощью модели гауссовых смесей (GMM). Вместо использования многопараметрической GMM в диссертации А.В. Ермиловым предложена аппроксимация спектра с помощью функции четырехпараметрического распределения, полученного на основе разложения Грам-Шарлье. Аппроксимация достигается путем оптимизации методом симуляции отжига по параметрам распределения. Представлены две модификации метода, пригодные для параллельных вычислений. Первая – одновременный запуск процессов отжига из множества различных начальных точек, чем более надежно достигается глобальный оптимум. Вторая модификация, менее вычислительно емкая, – это запуск процесса отжига из оптимальной начальной точки, выбранной из множества претендентов путем подстановки в целевую функцию. Найденные оптимальные параметры аппроксимирующего распределения характеризуют признаковую реализацию дикторонезависимой фонемы, последовательность которых в диссертации идентифицируется с помощью скрытых марковских моделей для дальнейшей их классификации по принадлежности языкам с помощью метода опорных векторов с гауссовским ядром. Для задачи распознавания диктора используются мел-частотные кепстральные признаки (MFCC) и их классификация с помощью метода опорных векторов. В главе приводится описание метода опорных векторов, анализируются алгоритмы его обучения и проблема линейной неразделимости по дикторам исходного признакового пространства. Для улучшения разделимости речевых признаков А.В. Ермиловым предложено применение ядра Фишера, вычисленного на основе распределения вектора признаков моделей, и использование его в методе опорных векторов с экспоненциальным ядром. В главе приводится алгоритм оценки наилучших параметров ядра в процессе обучения метода на обучающей выборке с процедурой кросс-валидации, указывается на возможность параллельных вычислений.

В третьей главе рассмотрены практические алгоритмические и программные решения системы идентификации языка и диктора с применением вышеизложенных методов. В главе представлена общая программная архитектура системы, приведена ее модель с использованием языка UML, описана её детальная структура, включая составляющие классы и функции, рассмотрены вопросы ее реализации. Изложены методы повышения быстродействия алгоритмов выделения признаков из акустического сигнала и методов идентификации на основе параллельных вычислений и с использованием конвейерной обработки.

Четвертая глава содержит результаты экспериментов на основе предлагающихся моделей распознавания. Приводятся данные экспериментов распознавания диктора по различным аудио записям, сделанных с микрофона, стационарного телефона и GSM телефона. Результаты показывают увеличение точности распознавания диктора с помощью предлагаемых методов по сравнению с распространенными и известными алгоритмами. Автором проведен анализ полученных результатов с учетом различных особенностей методов и условий получения аудиоданных. В главе описаны результаты экспериментов системы распознавания языка, приводятся данные и анализ по скорости сходимости предлагаемого метода симуляции отжига по сравнению с другими методами для аппроксимации спектральной огибающей, анализируется оценка расхождения полученного приближения и делается вывод об преимуществе предлагаемого подхода. В главе описана процедура обучения классификатора по множеству фонов 6 языков и приведены данные результатов распознавания, проанализирован вопрос устойчивости классификатора, путем моделирования ошибок в фоновой последовательности. Результаты, полученные диссертантом, подтверждают теоретические выкладки и свидетельствуют о практической применимости разработанных алгоритмов.

В Заключении диссертантом подведены итоги исследовательской работы и перечислены основные полученные результаты.

Несмотря на хорошее качество представленной работы, она не лишена

недостатков. Отметим некоторые из них.

1. В работе отсутствуют оценки области применимости предложенных методов.
2. Недостаточно подробно описана процедура обучения классификатора в методе опорных векторов и структура обучающих данных для распознавания диктора.
3. Неполно описана методика получения результатов экспериментов. Не приведены численные данные по ошибкам первого и второго рода распознавания диктора.
4. Непроанализированная возможность использования GPU для параллельных вычислений в предлагаемых алгоритмах.
5. Некоторые погрешности в связности изложения обсуждаемых методов и алгоритмов.

Отмеченные недостатки и стилистические погрешности не снижают качества диссертационной работы и её можно считать законченным решением научно-практической задачи. Достоверность результатов подтверждена внедренным и эксплуатируемым прототипом. Предложенные А. В. Ермаковым модели и решения могут быть использованы для дальнейших разработок и практического использования.

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

На основании вышеизложенного, диссертацию А. В. Ермакова на соискание ученой степени кандидата физико-математических наук можно считать научно-квалификационной работой, в которой содержится решение задачи по разработке методов, алгоритмов и программ по идентификации диктора и языка, имеющей существенное значение для применения при разработке акустических интерфейсов, что соответствует требованиям п. 9 «Положения о порядке присуждения ученых степеней» от 24.09.2013 г. № 842, а ее автор Ермаков Алексей Валерьевич

заслуживает присуждения ученой степени кандидата физико-математических наук.

к.ф.-м.н., научный сотрудник

Федерального государственного бюджетного

учреждения науки Вычислительный

центр им. А. А. Дородницына РАН,

119333, Москва, ул. Вавилова, 40,

тел.: 8-499-930-75-89

email: gneushev@ccas.ru

 /А.Н. Гнеушев/

