

ОБЪЕДИНЕННЫЙ ИНСТИТУТ ЯДЕРНЫХ ИССЛЕДОВАНИЙ
JOINT INSTITUTE FOR NUCLEAR RESEARCH
ЛАБОРАТОРИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ
LABORATORY OF INFORMATION TECHNOLOGIES

141980 г. Дубна, Московская область
Российская Федерация, ОИЯИ ЛИТ
Тел. (496-21) 6-40-19
Факс (496-21) 6-51-45



141980 Dubna, Moscow reg
Russia, JINR LIT
Tel. (496-21) 6-40-19
Fax (496-21) 6-51-45



«УТВЕРЖДАЮ»

 Директор ЛИТ
ОИЯИ, д.т.н. Кореньков В.В.

«9» сентября 2014 г.

ОТЗЫВ

ведущей организации на диссертационную работу 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 темы

В настоящее время существуют значительное количество методов идентификации речи, диктора и языка, однако большинство из них работает только при большом отношении сигнал/шум. При попытках идентификации при меньшем отношении сигнал/шум происходит возрастание количества ошибок и/или полный отказ в работе систем распознавания. Наиболее распространенные массовые каналы передачи акустической информации в настоящее время - это телефон и сотовая связь, но если операторы сотовой связи еще как-то гарантируют отношение сигнал/шум на уровне 40дБ, то обычная телефонная линия не может обеспечить даже этого. В таких условиях применение известных методов идентификации диктора и распознавание языка становится малоэффективным. Тем не менее, необходимость в мониторинге сообщений по каналам связи в настоящее время возрастает по разным причинам, в том числе из-за угрозы террористических актов. В связи с этим диссертационная работа А.В. Ермилова, посвященная разработке методов повышения качества распознавания языка и диктора, представляется весьма актуальной. Кроме того, решение данной задачи полезно для организации акустических интерфейсов в тех случаях, когда другие методы просто не работают.

Научная новизна исследования

Автором диссертации проведен анализ существующих информационных признаков, используемых для идентификации языка и диктора на основе физиологических особенностей человеческого языка и дикции с учетом механизма восприятия звука человеком при распознавании речи. Для повышения качества идентификации им впервые в практике распознавания акустических сигналов предложена и теоретически обоснована система характерных признаков для распознавания языка с применением 4-х параметрического семейства распределений (расширение Грам-Шарлье). Разработана и теоретически обоснована модификация метода опорных векторов, основанная на применении фишеровских ядер, позволившая существенно повысить вероятности правильного распознавания диктора в каналах с низким отношением сигнал/шум. Кроме того, разработана модификация алгоритма симуляции отжига,

увеличивающая быстродействие системы при получении дикторонезависимых признаков за счет введения в алгоритм параллельно выполняющихся циклов. Предложенный автором диссертации подход реализован в виде алгоритмов и программ на языке MatLab. Им проведены экспериментальные исследования по оценке точности распознавания и быстродействию системы идентификации языка и диктора, которые показали преимущества разработанных методов по сравнению с применяемыми ранее.

Хотя автор использует уже хорошо зарекомендовавшие себя вычислительные методы, однако представленную их комбинацию, совмещающих алгоритм опорных векторов, симуляцию отжига и Фишеровские ядра, можно считать принципиально новым подходом, позволяющим усовершенствовать уже существующие методы, алгоритмы и основанные на них расчетные схемы. Данный подход позволяет нивелировать недостатки известных подходов и, тем самым, существенно повысить эффективность исследования. В данной работе впервые проведен сравнительный анализ алгоритмов оптимизации для вычисления акустических дикторонезависимых признаков по скорости и точности.

Практическая значимость работы

Разработанные принципы, алгоритмы и программное обеспечение могут найти широкое применение при создании человеко-машинных интерфейсов в медицине, акустическом управлении сложными системами в различных отраслях науки и техники, а также для идентификации личности и языка в работе различных государственных служб и органов.

Разработанный автором подход был успешно применен в отделе речевых технологий ООО "Стэл КС", показав хорошую предсказательную способность.

Достоверность и обоснованность научных положений и выводов диссертации

Достоверность научных положений полученных автором обосновано доказательствами соответствующих теорем, строгим формальным применением теории распознавания образов, теории параллельных вычислений, методов цифровой обработки и анализа сигналов а также положительными результатами полученными в ходе экспериментов по идентификации языков и дикторов. Основные результаты диссертации достаточно полно отражены в 11 печатных работах, в т.ч. в 4 статьях в журналах, входящих в перечень ВАК РФ.

Оценка содержания диссертации

Диссертационная работа А.В. Ермилова состоит из введения, четырех глав, заключения и списка литературы.

Во введении обоснована актуальность темы диссертации, сформулированы цель и задачи исследований, отмечены научная новизна, практическая ценность и результаты реализации работы.

В первой главе рассмотрена модель речеобразования и статистические особенности речевого сигнала. Проведен анализ существующих методов распознавания языка и диктора. Изложены методы выделения акустических информативных признаков речевого сигнала.

Вторая глава посвящена математическим методам используемым при идентификации языка и диктора. Изложены методы и алгоритмы описания признаков акустических сигналов на основе скрытых марковских цепей. Приведен и теоретически обоснован модифицированный метод распознавания диктора на основе метода опорных векторов и рассмотрены его вариации с различными ядрами и функциями штрафов. Приведено описание известной модели "Auditory Image Model", на базе которой автором диссертации были разработаны методы идентификации диктора и языка. Изложен механизм построения описания признаков языка с использованием четырех параметрового распределения, называемого "расширение Грамм-Шарлье".

В третьей главе автором рассмотрен общий подход к построению систем для распознавания языка и диктора. Представлена общая архитектура программной реализации такой системы. Приведены результаты исследования некоторых аспектов применения параллельных алгоритмов к задаче идентификации языка и диктора. Изложены принципы конвейерной обработки речевого сигнала.

В четвертой главе приводятся результаты экспериментов по распознаванию с использованием различных механизмов. Полученные автором результаты свидетельствуют о

повышении качества распознавания языка и диктора для разработанных А.В. Ермиловым методов.

В заключении автором изложены основные выводы и результаты диссертационной работы.

Выводы по главам диссертации сформулированы корректно и соответствуют содержанию проведенного научного исследования, а полученные в работе результаты обоснованы теоретически и экспериментально и достаточно полно отражены в публикациях, в том числе, входящих в перечень ВАК.

Автореферат соответствует содержанию и структуре диссертации и не искажает ее смысла и полученных результатов.

Объем и структура диссертации

Диссертация А.В. Ермилова изложена на 135 страницах машинописного текста и состоит из 4 глав, в которых отражены основные данные литературы и собственные результаты, выводов и списка литературы, включающего 81 источник.

Замечания

Работа выполнена и представлена на достаточно высоком научном уровне, с привлечением различных современных методов. Из недостатков работы, имеющихся в работе и в оформлении содержания диссертации и автореферата, можно отметить следующие:

- В диссертационной работе и автореферате замечены орфографические и синтаксические ошибки.
- На стр. 44 диссертации на рисунке 1.8 (Пример преобразования мел фильтров) не расшифрованы обозначения $\alpha_{\min}$, $\alpha_{\max}$, f_L , f_U .
- Как уже отмечалось выше, в работе были использованы некоторые методы моделирования вероятностных распределений речевых признаков. На наш взгляд следовало бы провести исследование практической применимости других методов моделирования вероятностных распределений, например, такие как Deep Belief Networks или обобщенного t-распределения.
- Ни в литературном обзоре, ни в работе не рассмотрены другие методы построения нелинейной классификации, в частности, широко используемые нейросетевые методы. Представляет научный интерес провести их сравнение с предложенным в работе модифицированным методом на основе метода опорных векторов.

Однако отмеченные недостатки носят частный или рекомендательный характер и не влияют на общую высокую оценку диссертационной работы.

Рекомендации по практическому использованию

Представленные в диссертации результаты рекомендуется использовать в дальнейших исследованиях в научных и образовательных организациях Минобрнауки, а также в работе правоохранительных органов, программах защиты свидетелей, в системах слежения и сбора улик.

Результаты исследований А. В. Ермилова могут быть также использованы в учебном процессе при чтении спецкурсов на факультетах Бизнес-Информатики и Компьютерных наук НИУ ВШЭ, МВТУ им. Баумана, МАИ и в ряде других профильных высших учебных заведений.

Заключение

В целом диссертационная работа А. В. Ермилова представляет собой завершенную научно-исследовательскую работу на актуальную тему, выполненную на высоком научном и методическом уровнях. Достоверность данных и полученных результатов не вызывает сомнений. Все выводы в достаточной мере обоснованы. Основные результаты опубликованы в российских и международных рецензируемых журналах, рекомендуемых Минобрнауки для публикации результатов кандидатских диссертаций и обсуждались в виде докладов на международных научных конференциях и достаточно полно отражены в автореферате.

Диссертационная работа А. В. Ермилова «Методы, алгоритмы и программы решения задач идентификации языка и диктора» по актуальности, объёму выполненных исследований и

практической значимости полученных результатов соответствует требованиям «Положения о присуждении ученых степеней», утвержденного Постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842. предъявляемым к диссертациям на соискание учёной степени кандидата наук, а её автор А. В. Ермилов заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 05.13.11 - «Математическое обеспечение вычислительных машин комплексов и компьютерных сетей»

Отзыв обсуждён и одобрен на заседании научного семинара по вычислительной и прикладной математике Лаборатории Информационных Технологий Объединенного Института Ядерных Исследований от 23 мая 2014 года.

Главный научный сотрудник ЛИТ ОИЯИ,
доктор физ.-мат.наук, профессор
Г. А. Ососков



9 сентября 2014 г.