

Отзыв

официального оппонента на диссертационную работу Тупицина Геннадия Сергеевича на тему «Предобработка речевых сигналов в системах автоматической идентификации диктора», представленную на соискание ученой степени кандидата технических наук по специальности 05.12.04 Радиотехника, в том числе системы и устройства телевидения

Актуальность работы

Разработка алгоритмов автоматического распознавания диктора по его голосу является актуальным направлением цифровой обработки речевых сигналов. Данные алгоритмы уже успешно применяются в составе мультимодальных биометрических систем контроля и управления доступом и повышают точность работы таких систем. Алгоритмы распознавания диктора постепенно внедряются в системы дистанционного обслуживания населения для возможности проведения простой и быстрой процедуры аутентификации. Востребованы они и при обеспечении контроля доступа к различным радиотехническим устройствам и медицинскому оборудованию.

При разработке систем распознавания диктора важно учитывать и возможность воздействия фонового шума. Это проблема довольно актуальна, т. к. даже при относительно высоких отношениях сигнал/шум порядка 20 дБ точность работы существенно падает. Одним из эффективных способов улучшения качества работы систем распознавания в этих случаях является применение алгоритмов шумоподавления. При этом нужно отметить, что не всегда алгоритмы шумоподавления, предназначенные для повышения качества речевых сигналов, хорошо работают с системами распознавания диктора.

Разработке алгоритмов шумоподавления для системы идентификации диктора, а также оценке качества их работы и посвящена представленная диссертационная работа.

Структура и содержание работы

Работа состоит из введения, трех глав и заключения.

Во введении обосновывается актуальность темы исследования, определяются предмет, цель, задачи и методы исследования, научная новизна, практическая ценность работы, достоверность полученных результатов, приводятся сведения об апробации работы и внедрении результатов исследования, личном вкладе соискателя, указываются положения, выносимые на защиту.

В первой главе произведен обзор современных способов идентификации диктора и алгоритмов подавления шума в речевых сигналах. Подробно описаны мел-частотные кепстральные коэффициенты как эффективный способ представления речевого сигнала с помощью набора векторов относительно небольшой размерности. Приведено описание EM-алгоритма для представления модели диктора с помощью гауссовых смесей, а также алгоритма MAP-адаптации универсальной фоновой модели. В разделе, посвященном описанию алгоритмов шумоподавления, приведена подробная информация о широко используемых способах подавления шума в частотной области. Также в данной главе описаны методы оценки качества речевых сигналов, таких как PESQ, ОСШ, SegОСШ, WSS, LLR.

Вторая глава посвящена поиску альтернативных способов оценки точности работы системы идентификации диктора. Оценена возможность использования для этих целей популярных показателей качества речи, разработан комбинированный показатель качества (так называемый «АОТИД»), позволяющий аппроксимировать точность идентификации дикторов, полученную с использованием реальной системы идентификации. Рассматривается способ уменьшения числа тестовых сигналов, необходимых для ускорения процедуры оценки точности идентификации. В результате получена методика быстрой оценки точности идентификации дикторов. При тестировании ее для практической задачи подбора параметра алгоритма

шумоподавления получено значительное ускорение по сравнению с традиционным способом, использующим систему идентификации диктора.

В третьей главе производится попытка обобщить новый и еще малоизученный подход подавления шума с помощью мягких масок. Разработана новая методика оценки мягкой маски и двухступенчатый алгоритм на ее основе. Проведено исследование полученных алгоритмов с использованием двух баз речевых сигналов, в результате чего выявлено преимущество предлагаемого алгоритма при некоторых типах шума и отношениях сигнал/шум.

В заключении сформулированы основные выводы диссертационной работы.

По характеру поставленной цели, перечню решаемых задач и методов исследования представленная работа соответствует заявленной специальности 05.12.04 Радиотехника, в том числе системы и устройства телевидения.

Научная новизна и практическая значимость результатов работы

Впервые получены следующие научные результаты:

1. Определена теснота статистической связи точности идентификации дикторов с показателями качества речи PESQ, ОСШ, SegОСШ, WSS, LLR.
2. Разработана альтернативная оценка точности идентификации дикторов.
3. Предложена методика быстрой оценки точности идентификации дикторов.
4. Разработана новая методика оценки мягкой маски.
5. Предложен двухступенчатый алгоритм на основе мягкой маски.

Данные результаты имеют и практическую значимость. Так методика быстрой оценки точности идентификации дикторов позволяет быстрее

подбирать параметры алгоритмов шумоподавления. В частном случае получено ускорение в 88 раз. Предложенный двухступенчатый алгоритм шумоподавления на основе мягкой маски позволяет повысить точность идентификации дикторов в среднем (среди ОСШ 5, 10, 15 дБ) для аддитивного белого гауссовского шума на 13,4 процентных пункта по сравнению с алгоритмом на основе подхода прямого принятия решения и функции коррекции спектра Винера.

Практическая значимость также подтверждается соответствующими актами внедрения и двумя свидетельствами о государственной регистрации программы для ЭВМ.

Автореферат и публикации

Автореферат достаточно полно отражает содержание диссертации. Основные результаты диссертационной работы отражены в многочисленных публикациях, среди которых имеются 3 публикации в изданиях, рекомендованных ВАК.

Замечания по диссертационной работе

По представленной диссертационной работе имеются следующие замечания:

1. Рассмотрение алгоритмов шумоподавления ограничивается только алгоритмами, работающими в частотной области.
2. При получении комбинированного показателя качества используется относительно небольшая по объему выборка.
3. Задача шумоподавления в речевых сигналах ограничивается только рассмотрением воздействия аддитивных шумов.
4. Статистические параметры шума оцениваются только по имеющемуся фрагменту, что не всегда доступно.
5. База речевых сигналов «РУС-31-5» недостаточно полно описана.

Указанные замечания не носят принципиального характера и существенным образом не снижают научной и практической ценности диссертации.

Заключение

Диссертация «Предобработка речевых сигналов в системах автоматической идентификации диктора» является законченной научно-квалификационной работой, выполненной на актуальную тему и посвященной решению задачи разработки помехозащищенных систем идентификации диктора по его голосу. Она удовлетворяет требованиям ВАК, предъявляемым к кандидатским диссертациям, а ее автор Тупицин Геннадий Сергеевич заслуживает присуждения ученой степени кандидата технических наук по специальности 05.12.04 Радиотехника, в том числе системы и устройства телевидения.

Официальный оппонент
кандидат технических наук,
руководитель группы
связи и передачи данных
проектного отдела,
ООО «А-ВИЖН», г. Ярославль.



Савватин Алексей
Иванович

150000, г. Ярославль, ул. Большая Октябрьская, д. 45, оф. 212
ООО «А-ВИЖН»
Тел.: +7 (4852) 26-50-10, E-mail: savvatin@a-vsn.ru

Подпись к.т.н. А.И. Савватина заверяю:

Секретарь

1 декабря 2015 г.



А.А. Катаева