

**Министерство науки и высшего  
образования  
Российской Федерации**  
Федеральное государственное бюджетное  
образовательное учреждение  
высшего образования  
**«Ковровская государственная  
технологическая академия  
имени В.А. Дегтярева»**

Маяковского ул., д.19, г. Ковров,  
Владимирская область, 601910  
тел.(49232) 3-21-60 факс (49232) 3-21-60  
E-mail: ksta@dksta.ru

ОКПО 02069786, ОГРН 1023301953223

ИНН/КПП 3305007006/330501001

25. 11. 2019 № 10-Н-1-30/22-73

на № \_\_\_\_\_ от \_\_\_\_\_



**УТВЕРЖДАЮ:**  
Проректор по научной работе и  
международным связям  
Н.А. Кузнецов

2019

### **ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ**

на диссертацию Аль-Дайбани Абдулгани Мохаммеда Салех «Исследование методов и разработка алгоритмов обработки сигналов для автоматического распознавания телефонной речи в республике Йемен», представленную на соискание ученой степени кандидата технических наук по специальности 05.12.13 – «Системы, сети и устройства телекоммуникаций»

#### **Актуальность темы**

В настоящее время системы автоматического распознавания речи (САРР) начинают все шире применяться в автоматических справочных телефонных системах. Однако большое количество ошибок распознавания, вызванных наличием помех в речевом сигнале и вариациями частотной характеристики канала связи значительно сдерживают распространение САРР. Количество ошибок значительно возрастает, когда "общение" пользователей с системой происходит с использованием множества диалектов. Именно такая ситуация характерна для стран арабского мира, что не позволяет активно использовать САРР.

Точность распознавания диалектного арабского языка повышается, если использовать автоматическую идентификацию диалектов в составе

САРР. Исследования в области идентификации арабских диалектов распознавания очень малочисленны. Отсутствуют какие-либо данные о расчете вероятности ошибки идентификации диалекта. Представлены лишь результаты экспериментов по идентификации конкретных диалектов.

При обращении абонентов телефонной сети к автоматической справочной системе используются каналы связи с различными частотными характеристиками (ЧХ), что меняет состав спектра речевых сигналов (РС), поступающих на вход системы распознавания, точность распознавания речи снижается. Для подавления влияния ЧХ канала связи на параметры РС используется их нормализация по среднему значению. Но отсутствуют данные по анализу по оценке влияния способа обработки РС на степень стабилизации нормализованных параметров.

Как правило, речевые сигналы сопровождаются различными помехами, обусловленными акустическим окружением диктора. Точность распознавания при этом снижается. Для подавления помех в РС используются спектральное вычитание и фильтр Винера. Указанные методы не только подавляют помехи, но вносят некоторые искажения в РС. Степень искажений зависит не только от вида помех, но и от формы самих сигналов. Оценку степени искажений можно провести только по результатам экспериментов. Такие исследования в Йемене не проводились.

На основе вышеизложенного можно сделать вывод об **актуальности** темы диссертации, посвященной исследованию методов предварительной обработки РС, используемых при распознавании и разработке алгоритмов обработки речевых сигналов с целью использования их в составе системы распознавания диалектной речи в арабской республике Йемен.

### **Научная новизна исследований и полученных результатов**

1. Получены выражения для оценки вероятности ошибки оперативной идентификации диалекта в течение сеанса связи, использующей акустические модели произнесений контрольных слов.

2. Получены результаты эксперимента по оценке точности идентификации йеменских диалектов, основанной на использовании акустических моделей произнесений контрольных слов.
3. Получены выражения, определяющие зависимость нормализованных по среднему значению мел-частотных кепстральных коэффициентов от неравномерности амплитудно-частотной характеристики канала связи и вида оконной функции, используемой при дискретном преобразовании Фурье.
4. Получены результаты эксперимента по исследованию влияния различных оконных функций на значения параметров речевых сигналов - нормализованных мел-частотных кепстральных коэффициентов.
5. Получены результаты эксперимента по исследованию возможностей применения спектрального вычитания и фильтра Винера для подавления помех при автоматическом распознавании речи в Йемене.

#### **Практическая значимость проведенных исследований.**

1. Разработаны методика и алгоритм оценки целесообразности применения спектрального вычитания и фильтра Винера для снижения влияния помех на значения параметров РС и точность распознавания речи.
2. Разработаны методика и алгоритм оценки влияния нормализации параметров речевого сигнала на достоверность распознавания речи и определения степени стабильности нормализованных параметров при изменении ЧХ канала связи.
3. Разработано прошедшее государственную регистрацию программное обеспечение для настройки средств подавления влияния помех и ЧХ канала связи на работу SAPP, реализующее разработанные алгоритмы.
4. Предложенный алгоритм оперативной идентификации диалектов во время сеанса связи обеспечивает относительную ошибку

идентификации равную 0,24%, и повышает достоверность распознавания названий цифр, как минимум, на 7%.

5. Составлены и обработаны выборки аудиозаписей арабской речи для обучения и тестирования системы распознавания.

### **Структура и объем работы**

Диссертация состоит из Введения, четырех глав, Заключения, библиографического списка, включающего 100 наименований источников, и Приложений.

**Во Введении** обоснована актуальность темы диссертационной работы, сформулированы цель исследования и перечислены решаемые задачи, показаны научная новизна и практическая значимость полученных результатов, определены положения, выносимые на защиту.

**В первой главе** рассмотрены особенности и проблемы построения систем автоматического распознавания арабской речи. Арабский язык характеризуется большим разнообразием диалектов, что увеличивает число ошибок распознавания. Поэтому возникает необходимость включения идентификатора диалекта в состав системы распознавания речи. Проведенный анализ существующих результатов и подходов к идентификации диалектов арабского языка показал, что достаточно высокая точность идентификации достигается при использовании акустических моделей на основе скрытых марковских моделей и при использовании мел-частотных кепстральных коэффициентов (МЧКК) в качестве параметров речевых сигналов.

Во многих системах распознавания речи процесс распознавания основан на использовании акустических моделей звуков речи в виде скрытых марковских моделей. В составе звуков арабской речи есть такие звуки, которые отсутствуют в других национальных языках. Поэтому акустические модели звуков, используемые при распознавании арабской речи, значительно

отличаются от моделей, которые используются в составе существующих систем.

Системы автоматического распознавания речи включают в себя модуль предварительной обработки речевых сигналов, где подавляются помехи и определяются параметры сигналов – мел-частотные кепстральные коэффициенты (МЧКК). Для повышения устойчивости МЧКК к изменениям частотной характеристики канала связи используют нормализацию коэффициентов по их среднему значению.

Для очистки зашумленной речи применяют спектральное вычитание (СВ) и фильтр Винера (ФВ). В диссертации исследованы возможности подавления помех с помощью СВ и ФВ на примере автоматического распознавания названий цифр на диалектном языке Йемена. Проведены экспериментальные исследования изменений параметров РС (МЧКК) при воздействии помехи в виде модели белого шума и использовании указанных методов подавления помех.

Результаты экспериментов показывают, что при отношении сигнал-помеха меньше 35 дБ применение СВ и ФВ повышает достоверность распознавания.

**Во второй главе** проанализировано использование оперативной (в начале сеанса связи) идентификации диалекта для повышения достоверности распознавания путем произнесения контрольных слов. Рассмотрен случай, когда распознаются произнесения арабских названий отдельных цифр. Целесообразность идентификации диалектов оценена экспериментально путем сравнения результатов автоматического распознавания с учетом идентификации, когда для каждого названия цифры на каждом диалекте создается своя акустическая модель, и без ее учета, когда акустическая модель названия каждой цифры является общей для всех диалектов.

Йеменский диалектный язык можно разделить на три основные диалектные группы: северный (СД), южный (ЮД) и западный (ЗД) диалекты. Результаты проведенного эксперимента показывают, что при отсутствии

идентификации, относительная частота правильного распознавания, усредненная по всем названиям цифр, равна 90,8 %. Если же для каждого диалекта используется своя совокупность моделей для распознаваемых произнесений (случай произнесения двух контрольных слов), то такая же относительная частота для ЮД составляет 98,4%, СД - 97,8% и для ЗД - 97,5%.

**В третьей главе** исследована возможность применения нормализации параметров сигнала по их среднему (во времени) значению для подавления влияния частотной характеристики канала связи на работу системы распознавания. Рассмотрено влияние вида оконной функции и неравномерности амплитудно-частотной характеристики (АЧХ) канала на степень стабилизации нормализованных параметров сигнала при изменении АЧХ. Получено выражение для разности нормированных логарифмов спектров сигнала для случаев обучения и тестирования системы распознавания, когда "звуковой" участок канала связи меняется.

Показано теоретически и экспериментом подтверждено, что использование оконных функций, частотная характеристика которых имеет малый уровень тех боковых лепестков, частоты которых достаточно далеко удалены от частоты главного лепестка (например, оконная функция Ханна) обеспечивает большую стабильность нормализованных параметров речевого сигнала и большую достоверность распознавания речи.

**В четвертой главе** описано программное обеспечение экспериментов, предназначенное для оценки помехоустойчивости систем распознавания речи в телефонии. Разработанный программный комплекс позволяет оценить эффективность подавления аддитивных помех с помощью спектрального вычитания и фильтра Винера. Комплекс позволяет также оценить возможности нормализации параметров речевого сигнала по повышению достоверности распознавания.

В приложении представлены акты внедрения результатов диссертации в ООО "Центр речевых технологий", а также - в учебный процесс

Владимирского государственного университета. Кроме того, приводятся сведения о государственной регистрации разработанного программного обеспечения. Приведены также дополнительные экспериментальные данные по классификации диалектов и результаты экспериментов по оценке влияния некоторых видов помех, встречающихся на практике, на достоверность распознавания.

**Достоверность** научных положений, выводов и рекомендаций, сформулированных в диссертации, подтверждается использованием математического аппарата теории вероятностей, математической статистики, цифровой обработки сигналов, а также использованием имитационного моделирования.

Основные положения диссертации отражены в публикациях автора, а также прошли апробацию на научных конференциях.

Текст автореферата полностью соответствует содержанию диссертации и позволяет оценить результаты работы.

Результаты диссертационной работы можно рекомендовать к применению в научно-исследовательских организациях, занимающихся исследованием и разработкой систем автоматического распознавания речи.

Собранная выборка звукозаписей названий цифр, отсортированная по диалектам и дикторам может использоваться для обучения и тестирования систем распознавания арабской речи.

К недостаткам диссертации относится следующее.

1. В диссертации указывается на наличие нескольких видов нормализации параметров речевого сигнала. Однако исследована лишь нормализация параметров по среднему (во времени) значению. Неясно, почему не рассмотрены иные виды нормализации.
2. Не обоснован выбор фильтра, который использовался для имитации изменений частотной характеристики канала связи при экспериментальном исследовании влияния нормализации на параметры речевого сигнала.

3. Нет сравнения рассмотренного метода идентификации диалекта на основе использования скрытых марковских моделей с идентификацией, основанной на применении искусственных нейронных сетей.

### Заключение

Отмеченные недостатки носят частный характер и не сказываются существенно на научной и практической значимости полученных в диссертации результатов. Диссертация является законченной научно-квалификационной работой, в которой решена актуальная научная задача по созданию средств исследования и повышения помехоустойчивости телефонных систем автоматического распознавания арабской речи (республика Йемен).

Диссертация удовлетворяет требованиям п.п. 9, 10, 11, 13, 14 Положения о присуждении ученых степеней, а ее автор **Аль-Дайбани Абдулгани Мохаммед Салех** заслуживает присуждения ученой степени кандидата технических наук по специальности 05.12.13 «Системы, сети и устройства телекоммуникаций».

Отзыв на диссертацию и автореферат обсужден и одобрен на заседании № 11 от 15 ноября 2019 года кафедры Робототехника и комплексная автоматизация ФГБОУ ВО «Ковровская государственная технологическая академия имени В.А. Дегтярева»

Карпенков А.С.  
кандидат технических наук

