

## ОТЗЫВ

### официального оппонента

доцента Волго-Вятского филиала (г. Нижний Новгород) Московского технического университета связи и информатики (ВВФ МТУСИ), доктора технических наук, доцента Ю. М. Тулякова о диссертации «Исследование методов и разработка алгоритмов обработки сигналов для систем автоматического распознавания телефонной речи в республике Йемен» Аль-Дайбани Абдулгани Мохаммеда Салеха, представленной на соискание ученой степени кандидата технических наук по специальности 05.12.13 – «Системы, сети и устройства телекоммуникаций»

Одной из важных задач, стоящей перед разработчиками телефонных систем автоматического распознавания речи (САРР), является снижение числа ошибок распознавания. Ошибки распознавания, во многом обусловлены внешними воздействиями на речевой сигнал: аддитивными помехами и отличиями частотных характеристик телефонных каналов связи, которыми пользуются абоненты телефонной сети, от частотных характеристик, которые были использованы при создании выборки звуковых файлов, предназначенной для обучения САРР.

Особенностью арабской речи является наличие множества диалектов, что увеличивает изменчивость речевых сигналов, приводящую к дополнительному росту количества ошибок распознавания. Сложность создания САРР для арабского языка обусловила малое количество разработок в этой области. Для республики Йемен подобные разработки вообще отсутствуют.

Для решения задачи по реализации САРР необходимо на стадии предварительной обработки сигналов подавление влияний помех и частотной характеристики канала связи на параметры речевого сигнала, которые затем используются на стадии собственно распознавания речи. Поэтому решаемая в диссертации задача повышения устойчивости САРР к воздействию

внешних факторов на сигнал в условиях распознавания диалектной арабской речи является **актуальной**.

**Научную новизну данной работы можно** характеризовать следующими результатами.

1. Исследована работа САРР с использованием идентификатора диалектов. Получены выражения, позволяющие определить вероятность ошибки идентификации при произнесении контрольных слов. При идентификации используются акустические модели произнесений контрольных слов, созданные для каждого диалекта. Такой подход обеспечивает малую вероятность ошибки идентификации. Получены выражения, определяющие вероятность повторного произнесения контрольных слов, что создает дискомфорт при использовании системы.

2. По результатам проведения эксперимента получены значения достоверности распознавания при использовании идентификации диалекта. Показано, что при использовании акустических моделей распознаваемых слов отдельных для каждого диалекта достоверность распознавания значительно повышается.

3. Получены выражения, определяющие зависимость нормализованных по среднему значению параметров речевых сигналов от вида оконной функции, используемой при быстром преобразовании Фурье. Из приведенных выражений следует, что при использовании оконных функций, частотные характеристики которых имеют малый уровень боковых лепестков, достаточно удаленных от основного лепестка, можно снизить влияние частотной характеристики микрофона на нормализованные значения мел-частотных кепстральных коэффициентов (МЧКК), которые часто используются в качестве параметров речевых сигналов при распознавании речи.

4. На основе проведенных экспериментов определены (выделены) виды оконных функций, обеспечивающих снижение влияния частотной



характеристики канала связи на значения нормализованных параметров речевого сигнала.

5. С помощью экспериментальных исследований оценены возможности спектрального вычитания и фильтра Винера по снижению влияния аддитивных помех на достоверность распознавания. Показано, что использование фильтра Винера является более предпочтительным по сравнению со спектральным вычитанием, так как достигаемое повышение достоверности распознавания подтверждается и меньшим по сравнению со спектральным вычитанием изменением параметров речевого сигнала при воздействии помех.

**Практическая значимость** диссертации определяется следующими результатами.

1. Алгоритм идентификации диалектов, предложенный в диссертации, обеспечивает относительную ошибку идентификации равную 0,24%, и повышает достоверность распознавания арабских названий цифр, как минимум, на 7%.

2. Предложены методика и соответствующий алгоритм оценки влияния нормализации параметров речевого сигнала по среднему значению на достоверность распознавания речи. Предложены также методика и алгоритм определения степени стабильности нормализованных параметров при изменении частотной характеристики канала связи.

3. Предложены методика и алгоритм оценки возможной спектрального вычитания и фильтра Винера по снижению влияния помех на значения параметров РС и точность распознавания речи.

4. Разработаны компьютерные программы, которые прошли государственную регистрацию. Программы имеют развитый графический интерфейс и позволяют настроить САРР на максимальную достоверность распознавания с помощью рассмотренных средств подавления помех и применения нормализованных параметров сигнала.

5. Подготовлены выборки звукозаписей для настройки САРР.

6. Дополнительно следует отметить большую трудоемкость данной работы. В частности, трудоемкость обусловлена особенностью арабского письма, которая заключается в отсутствии букв, отражающих гласные звуки. Поэтому при анализе звукового состава слова требуется проводить так называемую огласовку согласных букв.

**Достоверность и обоснованность** научных положений, выводов сформулированных в диссертации, подтверждается использованием математического аппарата теории вероятностей, 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.

К **недостаткам** диссертации можно отнести следующее.

1. В диссертации проведены исследования по диалектам республики Йемен, однако нет исследований при использовании стандартного арабского языка в республике.

2. Неясно, почему при исследовании нормализации тестовый сигнал, является суммой амплитудно-модулированного гармонического колебания и белого гауссова шума. Такой сигнал может значительно отличаться от речевого сигнала.

3. Не дано сравнение разработанного программного обеспечения с существующими системами, которые используются при разработке САРР (например, ПО "НТК - Hidden Markov Model Toolkit", предназначенный для САРР, построенных с использованием скрытых марковских моделей)

4. В диссертации отсутствует подробное описание исследований, проведенных с аудиозаписями помех, присутствующих в реальности при эксплуатации САРР.

Отмеченные недостатки не являются определяющими 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 в республике 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».

Доцент Волго-Вятского филиала  
(г. Нижний Новгород) Московского  
технического университета связи  
и информатики (ВВФ МТУСИ),  
доктор технических наук, доцент



Ю. М. Туляков

«01» декабря 2019г

603011, Нижегородская область, г. Нижний Новгород, ул. Менделеева,  
д.15, телефон: +7 910 790 1111, Email: [yu.m.tulyakov@rambler.ru](mailto:yu.m.tulyakov@rambler.ru)

Подпись Тулякова Юрия Михайловича заверяю,  
зам. директора Волго-Вятского  
филиала Московского технического  
университета связи и информатики



С.В. Гвоздиков.