Санкт-Петербургский государственный университет

Кочаров Даниил Александрович

АВТОМАТИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ ЗВУКОВ РЕЧИ

10 02 19 – Теория языка

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата филологических наук

Санкт-Петербург – 2008

Диссертация выполнена на кафедре фонетики и методики преподавания иностранных языков факультета филологии и искусств Санкт-Петербургского государственного университета

Научный руководитель – доктор филологических наук, профессор Скрелин Павел Анатольевич

Официальные оппоненты – доктор филологических наук, профессор Мартыненко Григорий Яковлевич,

кандидат филологических наук, Егорова Ольга Борисовна

Ведущая организация – Санкт-Петербургский институт информатики и автоматизации Российской академии наук

Защита состоится «27» июня 2008 года в «15[00]» часов на заседании совета Д 212.232 23 по защите докторских и кандидатских диссертаций при Санкт-Петербургском государственном университете по адресу: 199034, Санкт-Петербург, Университетская наб 11, факультет филологии и искусств, ауд 215

С диссертацией можно ознакомиться в Научной библиотеке им. М Горького Санкт-Петербургского государственного университета (Санкт-Петербург, Университетская набережная, 7/9)

Автореферат разослан «19» мая 2008 года.

Ученый секретарь
диссертационного совета Д 212 232.23,
доктор филологических наук, профессор К. А. Филиппов

Область применения речевых технологий постоянно расширяется Особенно это относится к автоматическому распознаванию и транскрипции речи Почти все современные системы автоматической обработки речи полностью основаны на статистических моделях, и в них используется довольно примитивное, с лингвистической точки зрения, моделирование речевого сигнала Современный статистический подход к моделированию речи сталкивается с существенными проблемами при переходе от лабораторных данных к реальному речевому материалу Это происходит в силу специфики статистического подхода система эффективно работает на том материале, на котором она обучалась

В настоящее время в мире немного исследований, задачей которых является разработка лингвистических методов решения существующих проблем Лингвистический подход может быть очень эффективен для создания антропоморфных моделей речевого сигнала, т е таких, которые учитывают то, как человек порождает и воспринимает речь

Целью настоящего диссертационного исследования является автоматическая интерпретация звуков речи, основанная на лингвистических принципах Интерпретация звуков речи осуществляется при помощи акустических моделей речевого сигнала, построенных на основе фонологических и фонетических принципов.

Предметом исследования являются устойчивые акустические характеристики звуков речи, основанные на фонетических свойствах, которые могут быть использованы для разработки автоматической процедуры выделения и интерпретации звуков

В ***задачи исследования*** входит

1 определение акустических характеристик звуков речи в разных типах речи (спонтанной речи и чтении),
2 создание процедур автоматического выделения полезных акустических признаков звуков в речевом сигнале,
3. создание процедур автоматической фонемной интерпретации акустических характеристик звуков речи,
4 проверка точности интерпретации звуков речи в разных видах речевого материала и условиях предварительной обработки, а также разных методов представления речевого сигнала.

Научная новизна заключается в применении современных достижений цифровой обработки речевого сигнала для сегментации речевого потока и выделения акустических коррелятов традиционных фонетических (артикуляторных и перцептивных) признаков

звуков речи с их последующей автоматической классификацией и интерпретацией на основе современных методов статистической обработки данных

Теоретическая ценность исследования заключается в выявлении и формально-акустическом представлении устойчивых свойств звуков русской речи, которые необходимы для их автоматической классификации и интерпретации и сохраняются в разных типах речи Данная работа опирается на традиционный подход Щербовской фонологической школы к классификации элементов звуковой системы языка на основе артикуляторных и акустических дифференциальных признаков

Практическая значимость работы определяется возможностью использования фонетических характеристик звуков речи для их автоматической классификации и интерпретации в системах автоматической транскрипции русской устной речи. Использование результатов работы в системах автоматического распознавания речи позволит увеличить их эффективность в применении к спонтанной речи, а также возможность адаптации подобных систем к обработке других языков

В результате исследования сформулированы и ***выносятся на защиту следующие положения***:

- Фонетические характеристики могут быть успешно использованы для моделирования звуков речи в разных условиях реализации для их автоматической интерпретации В ходе экспериментов была доказана эффективность применения акустических коррелятов звонкости, сонорности, местоположения формант, а также места и способа образования согласных Все фонетические характеристики были проверены в экспериментах на точность интерпретации звуков речи в разных видах речевого материала и условий предварительной обработки, а также методов представления речевого сигнала
- Для успешного моделирования речи необходимо учитывать особенности реализации звуков в спонтанной речи Модели, получаемые на спонтанной речи, не соответствуют вероятностным распределениям, выведенным на основе исследования «идеальной» речи.
- Комбинация математического и фонетического подходов к моделированию звуков речи более эффективна, чем наиболее распространенный в настоящее время математический подход Фонетический подход к акустическому моделированию звуков речи учитывает то, каким образом образуются звуки речи, как они противопоставляются друг другу в рамках фонологической системы отдельно взятого языка и как они влияют друг на друга в речевом сигнале.

Апробация работы. Достоверность диссертационного исследования была проверена при помощи экспериментов на материале новейших тестовых корпусов на русском, немецком и английском языках, общим объемом около 135 часов Результаты исследований были представлены в докладах на заседаниях кафедры фонетики и методики преподавания иностранных языков Санкт-Петербургского государственного университета, на семинарах, посвященных вопросам речевых технологий, на межвузовских конференциях преподавателей и аспирантов в СПбГУ (2003, 2006, 2007), на международных конференциях (SPECOM 2004, 2006 и Interspeech 2005) Результаты исследований опубликованы в 10-ти изданиях, в том числе в двух рецензируемых изданиях из списка ВАКа «Вестник СПбГУ» и «Speech Communication» (на основании системы цитирования «Web of Science», см перечень рецензируемых научных журналов и изданий ВАК от 21 апреля 2008 г)

Структура работы. Данная диссертационная работа содержит 169 страниц машинописного текста и состоит из введения, трёх глав, заключения, списка использованной литературы (138 наименований) и четырёх приложений Работа иллюстрирована рисунками и таблицами

В первой главе изложены основы анализа и акустического моделирования речевого сигнала, применяемые в современных работах в области речевых технологий В том числе описаны методы, применённые в диссертационном исследовании Кроме того, в первой главе рассмотрены особенности спонтанной речи Особое внимание уделяется акустическим свойствам реализованных в спонтанной речи аллофонов

При моделировании спонтанной речи следует учитывать особенности реализации звуков в спонтанной речи, так как почти все они являются потенциальными источниками проблем и множества исключений из правил или вероятностных распределений, выведенных на основе исследования лабораторного материала

Фонетический подход к акустическому моделированию звуков речи учитывает то, каким образом образуются звуки речи, как они противопоставляются друг другу в рамках фонологической системы языка, и как они влияют друг на друга в речевом сигнале

Построение акустической модели речевого сигнала и ее использование в качестве основы для принятия решений является основным подходом к автоматической интерпретации речи Акустическая модель звука речи определяется набором описывающих его акустических характеристик, представленных в виде одного или комбинации акустических векторов

Акустические реализации фонем значительно меняются в зависимости от скорости речи Стандартным математическим подходом к моделированию временной вариативности речи в системах распознавания речи является использование скрытых Марковских моделей

Во второй главе рассмотрены фонетические характеристики речевого сигнала, разработанные и реализованные в рамках данного диссертационного исследования Под фонетическими характеристиками понимаются те, которые различают фонологические или фонетические классы Для каждой характеристики подробно описан алгоритм ее получения из речевого сигнала В их числе были акустические корреляты звонкости, сонорности, местоположения формант и центр тяжести спектра для определения места и способа образования согласных

Акустические характеристики являются основными структурными элементами акустических моделей. От правильного выбора характеристик во многом зависит то, насколько полученная в итоге модель будет удовлетворять накладываемым на нее требованиям Все остальные процедуры, используемые для акустического моделирования, направлены в основном на компенсацию классификационных минусов и неточного определения акустических характеристик

При выборе акустических характеристик необходимо считаться с двумя противоречивыми требованиями. С одной стороны, для осуществления надежного распознавания необходимо сохранить исходную информацию С другой, для простоты технической реализации количество измеряемых параметров и точность их измерения должны быть по возможности сравнительно небольшими Из-за сильной вариативности речи и сложности речевых сигналов абсолютное выполнение этих условий невозможно

Основной применяемой фонетической характеристикой является наличие основного тона (ОТ) В рамках данной работы был реализован алгоритм определения наличия ОТ на основе автокорреляционного метода

Перспективным подходом представляется выделение акустических характеристик синхронно периодам частоты основного тона (ЧОТ) там, где это возможно. Поэтому одним из направлений исследований была выбрана разработка алгоритма максимально точного автоматического определения ЧОТ. Можно допустить, что в рамках одного периода ЧОТ частоты формант не изменяются, т к они кратны ЧОТ, которая, естественно, не изменяется в таком окне Для того чтобы сделать правильную обработку синхронно периодам ЧОТ, помимо самой ЧОТ необходимо еще и знать точные границы ее периодов Ошибка даже в один отсчет может дать нежелательные искажения. В случае

отсутствия ОТ речевой сигнал обрабатывается при помощи стандартных окон постоянной длины

Ни один алгоритм автоматического определения ЧОТ не работает абсолютно правильно. каждый из них имеет свои плюсы и минусы, но при использовании комбинации различных методов можно достичь достаточно высокой эффективности В ходе исследований и экспериментов, целью которых были разработка, реализация и сравнение эффективности различных алгоритмов определения ЧОТ, были выбраны наиболее перспективные, и на основе их комбинации было реализовано автоматическое определение ЧОТ Это следующие алгоритмы

1 Вычисление автокорреляционной функции,

2 Анализ через синтез,

3 Вычисление отношения длины текущего периода с длинами предыдущих периодов ЧОТ

Для окончательного определения периода ЧОТ используется линейная комбинация логарифмов значений четырех характеристик. значения автокорреляционной функции, дистанции до синтезированного сигнала и отношения длины текущего периода к длинам двух предыдущих периодов. Каждый из методов имеет свои плюсы и минусы, а совместное применение позволяет акцентировать их сильные стороны и, соответственно, уменьшить количество неправильно определенных периодов ЧОТ Наиболее вероятная последовательность периодов ЧОТ определялась с помощью алгоритма Витерби Определенные таким образом периоды ЧОТ использовались для назначения окон обработки речевого сигнала при выделении фонетических характеристик, в частности местонахождения формант

Для идентификации гласных фонем в рамках данного исследования был разработан алгоритм определения местонахождения формант на основе обработки речевого сигнала синхронно периодам основного тона Была выбрана процедура, когда длина окна обработки была равна трем периодам частоты основного тона, а его шаг был равен одному периоду

Форманты определялись на основе спектра речевого сигнала Для этого определяются все гармоники частоты основного тона до 4000 Гц В качестве гармоник берутся частоты кратные частоте основного тона Далее строится гребенка непересекающихся треугольных фильтров таким образом, что центральной частотой фильтров являются гармоники, а ширина полосы фильтрации равна ЧОТ. Каждой гармонике соответствует отдельный фильтр гребенки Значения спектра в каждой полосе

суммируются, и полученные значения суммы сравниваются с взвешенным значением, соответствующим частоте основного тона Если оно больше порога, то выход фильтра равен «1», если меньше, то «0» Таким образом, на выходе гребенки полосных фильтров получается бинарный вектор Значение «1» в векторе обозначает присутствие форманты на месте соответствующей гармоники, а «0» - отсутствие форманты Размерность характеристического вектора является переменной и зависит от текущей ЧОТ. Из-за того, что ЧОТ постоянно меняется, даже соседние векторы могут быть разной длины, что делает их сравнение крайне неудобным По этой причине была введена еще одна, «статическая», гребенка прямоугольных фильтров, состоящая из 12 фильтров На этот раз параметры фильтров уже не зависят от частоты основного тона и заданы заранее с учетом информации о формантной структуре гласных русского языка Полученный 12-ти мерный вектор является значением акустической характеристики звука речи, отражающей его формантную структуру

Кроме местоположения формант и наличия основного тона была разработана акустическая характеристика, отражающая сонорность звука речи. Сонорность фонемы можно определить как степень ее звучности или как, наоборот, степень участия шумовых составляющих

В качестве акустического коррелята и показателя фонетической характеристики сонорности предлагается использовать сумму производных спектра в каждой точке по шкале частот С математической точки зрения применение производной спектра мотивируется тем, что производная функции выражает скорость ее изменения Производная функции обладает таким свойством, что чем выше скорость изменения функции, тем выше значение модуля производной Таким образом, производная спектра в частотной области должна отражать скорость изменения спектра в частотной области, и, соответственно, его «изломленности» Это коррелирует с количеством и качеством пиков в спектре и, следовательно, может выразить сонорность звука речи

Получение величины акустического коррелята сонорности основано на вычислении производных амплитудного спектра во всех точках частотной области и суммировании модулей полученных производных. В дискретном случае производная функции равна разности значений функции в последовательных точках. Значение акустической характеристики вычисляется как логарифм суммы модулей производных спектра. Сумма логарифмируется для того, чтобы уменьшить диапазон значений акустической характеристики

Акустическая характеристика сонорности напрямую зависит от качества спектра речевого сигнала, поэтому на ее производительность влияют различные преобразования спектра и фильтрация речевого сигнала Было решено рассмотреть спектр, представленный в различных нелинейных шкалах, и отфильтрованным низкочастотным фильтром с различными частотами среза, чтобы посмотреть, как это будет влиять на эффективность применения производной спектра в качестве акустической характеристики сонорности. Было опробовано несколько преобразований спектра, таких как преобразование из шкалы герц в шкалы мелов и барков, а также преобразование спектра билинейной функцией Этап, заключавшийся в модификации спектра, осуществлялся сразу после нормализации амплитудного спектра перед его дифференцированием Эксперименты показали, что преобразование спектра перед его дифференцированием может значительно увеличить эффективность акустической характеристики сонорности Наилучшие результаты были получены при использовании преобразования при помощи билинейной функции с коэффициентом преломления равным 0,8. Для низкочастотной фильтрации применялся идеальный FFT-фильтр нижних частот Были проведены несколько экспериментов с постепенным увеличением частоты среза низкочастотного фильтра от 500 Гц до 6000 Гц Результаты экспериментов показали, что использование фильтра нижних частот с частотой среза равной 1000 Гц дало наилучшие результаты Они почти совпадают с результатами, полученными при предварительном преобразовании спектра при помощи билинейной функции с коэффициентом преломления равным 0,8.

Для интерпретации согласных применялось вычисление центра тяжести спектра Идея этого подхода состоит в том, что на фоне достаточно плоского спектра на определенных частотах согласные имеют усиление спектральных составляющих Значения этих частот связанно с местом образования согласных. Так, губные согласные имеют увеличение амплитуды спектра на частотах, находящихся в области 800 Гц, заднеязычные – около 1400 Гц, а переднеязычные – выше 2000 Гц Соответственно, вычисляя центр тяжести спектра согласных можно оценивать место их образования Для этого была осуществлена полосная фильтрация, для вычисления центра тяжести спектра использовались значения, полученные на выходах полосных фильтров Гребенка полосных фильтров состоит из трех фильтров, которые были выбраны с учетом корреляции между местом образования согласного и частотными областями, доминирующими в спектре этих согласных

В ходе исследований акустические характеристики комбинировались на уровне акустических векторов при помощи ЛДА

В третьей главе представлены результаты, полученные в ходе экспериментов как по интерпретации отдельных звуков речи при помощи фонетических характеристик, так и по использованию таких характеристик в системах автоматического распознавания слитной речи Результаты приводятся на материале русского, немецкого и английского языков

Для оценки эффективности разработанных и реализованных акустических характеристик, основанных на фонетических принципах, были проведены несколько экспериментов В ходе экспериментов были опробованы следующие методы автоматической интерпретации звуков речи

1 определение гласных по местоположению их формант,
2 определение согласных по акустическим характеристикам, связанным с местом и способом образования согласных,
3 использование акустической характеристики наличия основного тона в комплексной статистической системе автоматического распознавания слитной речи,
4 использование акустической характеристики сонорности звуков речи в комплексной статистической системе автоматического распознавания слитной речи

Первые два эксперимента проводились на материале выделенных вручную звуков русской речи, классификация звуков речи основывалась на вычислении Евклидова расстояния В последних двух экспериментах использовалась система автоматического распознавания речи, основанная на статистических принципах и включающая в себя самые современные процедуры математического анализа данных, где фонетические характеристики использовались в качестве дополнительных к общепринятым акустическим характеристикам

Предложенный метод распознавания гласных по формантам синхронно периодам основного тона был проверен на материале русского языка Материал состоял из двух частей· отдельных гласных и отдельных слов

Первая часть представляла собой корпус гласных, которые были вручную выделены из фонетически представительного текста, прочитанного диктором-мужчиной, нормативным носителем русского языка Всего был использован 3771 гласный В этот набор входили реализации всех комбинаторных и позиционных аллофонов гласных В то же самое время сохранялась относительная частотность встречаемости каждого из гласных в реальных текстах

Вторая часть материала состояла из корпуса изолированно произнесенных слов Были использованы 33 слова (команды голосового меню мобильного телефона) Список слов предоставлен в приложении 3 Эти слова были произнесены 20-ю дикторами 10-ю мужчинами и 10-ю женщинами. Каждое слово было записано три раза, с разной скоростью чтения Таким образом, тестовая часть состояла из 1880 слов Гласные автоматически выделялись из слов и затем подвергались процедуре распознавания

Данные об эффективности распознавания изолированных гласных приведены в таблице 1

Таблица 1. Результаты распознавания изолированных гласных

Гласный	Кол-во ошибок (%)
a	6,10
e	4,54
ı	5,20
o	5,15
u	4,50
ɨ	5,60
В среднем	4,92

В ходе экспериментов по распознаванию гласных в составе слов речевой сигнал автоматически сегментировался на звуки, и которых затем анализировались и интерпретировались гласные звуки Отличие от предыдущего эксперимента в том, что звуки выделялись из речевого сигнала не так точно как во время ручной сегментации

Сначала выделяются озвонченные участки речевого сигнала при помощи описанного алгоритма Затем выделенные участки речевого сигнала сегментируются на основе изменения значений корреляционных функций средней энергии сигнала, спектральной интенсивности и огибающей сигнала на отдельных периодах Там, где все эти три функции имеют локальный максимум, ставится граница Вычисляются три функции, так как каждая из них в отдельности дает максимумы не только на границах звуков, но и внутри них Комбинируя три функции корреляции, мы убираем лишние потенциальные границы Это очень грубый метод, но он обеспечил достаточную точность для выделения гласных из отдельно произнесенных слов

Затем выделенных гласных был сформирован тестовый корпус. База эталонов использовалась та же, что и для распознавания отдельных гласных Полученные результаты были ниже, чем при распознавании отдельных гласных, что и предполагалось априори Для этого есть две причины. междикторская вариативность речи и более низкая точность автоматической сегментации по сравнению с ручной Результаты автоматической сегментации примерно соответствовали результатам, получающимся в

итоге при автоматическом распознавании слитной речи В таблице 2 приведена результативность автоматической интерпретации гласных в составе слов

Таблица 2 Результаты распознавания гласных в словах

Гласный	Кол-во ошибок (%)
a	15,28
e	15,00
i	15,51
o	15,28
u	18,10
ɨ	16,95
В среднем	16,02

В целом, результаты показывают эффективность применения предложенных акустических характеристик, основанных на выделении формант гласных синхронно периодам основного тона Результаты распознавания примерно одинаковы для всех гласных

Алгоритм не является зависимым от русского языка и может быть применен для распознавания гласных на другом языке, с единственным изменением, касающимся характеристик полосных фильтров, которые зависят от фонологической системы рассматриваемого языка

Для эксперимента по автоматической интерпретации согласных в качестве материала был использована часть корпуса русской спонтанной речи Всего использовалось около 6800 согласных Использовались записи 10-ти дикторов: 5-ти мужчин и 5-ти женщин В рамках эксперимента сначала определялся способ образования согласного по присутствию или отсутствию смычки Затем определялось место образования согласного при помощи широкополосных фильтров

В данной работе оценивались акустические характеристики самих звуков речи без учета контекста, поскольку анализ влияния контекста выходит за рамки диссертационного исследования Как следствие, было решено не разделять мягкие и твердые согласные, так как в речи часто этот дифференциальный признак реализуется за счет движения формант окружающих гласных На основе только собственных акустических данных отдельного согласного практически невозможно определить его мягкость, может быть за исключением сибилянтов /s'/ и /z'/ При построении системы, основанной полностью на антропоморфных принципах, которая в том числе учитывала бы контекст при анализе и акустическом моделировании отдельных звуков речи так, как это делает человек, эта проблема могла бы быть решена В таблице 3 приведены результаты распознавания согласных

Таблица 3 Результаты распознавания отдельных согласных

	f	v	s	z	š	ž	š'	h	p	b	t	d	k	g
f	87,4	2,2	1,4	0	0	0	0	5,2	3,0	0	1,3	0	0	0
v	4,1	88,6	0	0	0	1,9	0	2,4	0	3,0	0	0	0	2,3
s	6,9	0	74,7	4,2	3,7	0	2,4	9,3	2,2	0	8,5	0	0,5	0
z	0	2,4	3,1	77,6	0,4	5,9	0	2,1	0	0	0	3,1	0	0
š	0	0	10,9	3,2	89,4	2,8	7,1	0	0	0	4,0	0	0	0
ž	0	3,7	0	13,5	1,1	88,0	1,1	0	0	0	0	0	0	0
š'	0	0	8,3	1,5	5,4	0,4	89,4	0	0	0	2,5	0	0	0
h	3,1	0	1,6	0	0	0	0	73,4	3,7	0	0	0	5,4	0
p	0,5	0	0	0	0	0	0	1,8	81,8	4,1	4,1	1,9	4,0	2,1
b	0	3,1	0	0	0	0	0	0	7 8	83,4	0	1,5	0	8,4
t	0	0	0	0	0	0	0	0,9	0,5	0	72,8	2,8	3,0	0
d	0	0	0	0	0	0	0	2,3	1,0	6,3	1,5	74,0	0	7,2
k	0	0	0	0	0	0	0	1,6	0	0	5,3	2,2	84,6	4,0
g	0	0	0	0	0	0	0	0	0	3,2	0	10,5	2,5	76,0

В таблице 4 приведены результаты распознавания, полученные при использовании акустической характеристики звонкости совместно с различными акустическими характеристиками такими, как MFCC, PLP Кроме того, исследуемая фонетическая характеристика была опробована в комбинации с MFCC характеристикой, адаптированной к диктору при помощи алгоритма нормализации длины речевого тракта (VTLN) Для комбинации акустических характеристик использовался алгоритм ЛДА Эксперименты проводились на корпусе с малым объемом словаря, SieTill, и на корпусах с большим объемом словаря, VerbMobil II и EPPS.

Таблица 4 Результаты распознавания при использовании акустической характеристики звонкости (Зв)

Корпус	Акуст хар-ки	Кол-во ошибок (%)
SieTill	MFCC	1,8
	MFCC + Зв	1,6
VerbMobil II	MFCC	21,0
	MFCC + Зв	20,3
	VTLN	19,1

	VTLN + Зв	18,7
	PLP	21,4
	PLP + Зв	20,6
EPPS	MFCC	14,7 / 15,3
	MFCC + Зв	14,3 / 14,8
	VTLN	14,2 / 14,1
	VTLN + Зв	13,8 / 14,0
	PLP	15,4 / 15,8
	PLP + Зв	15,1 / 15,4

Для корпуса EPPS ошибки указаны в следующем виде: корпус разработки / тестовый корпус В таблице количество ошибок указано на уровне слов, т. е приведено относительное количество неправильно распознанных слов – это стандартная мера эффективности систем автоматического распознавания слитной речи Результаты показывают, что использование акустической характеристики звонкости в качестве дополнительной во всех случаях увеличивает эффективность системы вне зависимости от корпуса и используемых базовых акустических характеристик Во время экспериментов было получено увеличение результативности системы автоматического распознавания на ≈11 % для корпуса с малым объемом словаря и на ≈3 % для корпуса с большим объемом словаря

Акустическая характеристика сонорности (С) тестировались совместно с MFCC, VTLN и характеристикой звонкости (Зв) Для объединения акустических характеристик использовался алгоритм ЛДА Эксперименты проводились на корпусе с малым объемом словаря, SieTill, и на корпусах с большим объемом словаря, VerbMobil II и EPPS В таблице 5 представлены результаты использования акустической характеристики сонорности совместно с акустическими характеристиками MFCC и VTLN на разных корпусах

Таблица 5 Результаты распознавания речи при использовании акустической характеристики сонорности

Корпус	Акуст хар-ки	Кол-во ошибок (%)
SieTill	MFCC	1,8
	MFCC + C	1,6

VerbMobil II	MFCC	21,0
	MFCC + C	20,3
	VTLN	19,1
	VTLN + C	18,6
EPPS	MFCC	14,7 / 15,3
	MFCC + C	14,7 / 15,1
	VTLN	14,2 / 14,1
	VTLN + C	14,2 / 14,1

Результаты экспериментов показали эффективность применения предложенной акустической характеристики сонорности звуков речи Использование производной спектра в качестве дополнительной акустической характеристики увеличило эффективность системы распознавания речи по сравнению с использованием MFCC или VTLN

Улучшение равно ≈13 % на корпусе SieTill и ≈3 % на корпусе VerbMobil II По неизвестным причинам акустическая характеристика сонорности не смогла существенно улучшить распознавание речи на корпусе EPPS С другой стороны она и не вызвала ухудшения полученных результатов распознавания Поэтому эксперимент по ее применению в системе автоматического распознавания слитной речи можно считать успешным.

Среди исследованных вопросов, связанных с производной спектра, наиболее важными были число производных спектра, используемых для распознавания, и влияние фильтрации и представления спектра в нелинейных шкалах на результативность системы распознавания речи. Результаты экспериментов отражают влияние, которое оказывает представление спектра речевого сигнала в нелинейных шкалах на эффективность производной спектра Эксперименты проводились только на материале корпуса VerbMobil II Наилучшие результаты были получены при представлении спектра речевого сигнала в шкале мелов, и при его преобразовании с помощью билинейной функции со значением коэффициента преломления, равного 0,8 Было достигнуто относительное увеличение эффективности на 4,5 %. Результаты экспериментов отражают влияние, которое оказывает предварительная фильтрация речевого сигнала при помощи низкочастотного спектра с разной частотой среза на эффективность производной спектра Эксперименты проводились только на материале корпуса VerbMobil II Наилучшие результаты, относительное увеличение эффективности системы автоматического

распознавания речи на 3,5 %, были получены при фильтрации с частотой среза равной 1000 Гц

Результаты экспериментов показали достаточно высокую эффективность фонетического подхода к интерпретации звуков речи, что говорит о перспективности применения представленных фонетических характеристик в автоматическом распознавании и транскрипции речи, а также о перспективности антропоморфных моделей, опирающихся не только на собственные акустические характеристики звука речи, но и на окружающий контекст Применяя фонетический подход к интерпретации звуков, были достигнуты результаты соизмеримые, а в некоторых случаях лучше тех, что были заявлены исследователями при использовании акустических характеристик Развитием этого подхода был бы учет динамических процессов, происходящих внутри звуков речи, заключающихся в изменении и движении формант в зависимости от контекста

В заключительной части работы приведены основные выводы по результатам диссертации

В настоящее время в речевых технологиях преобладает статистический подход к моделированию речевого сигнала Несмотря на постоянное развитие математических алгоритмов обработки и классификации данных, системы автоматической обработки речи не достигают результативности, показываемой людьми в сходных условиях

Статистический подход к акустическому моделированию речи сталкивается с существенными проблемами при переходе от лабораторных данных к реальному материалу, представленному в разных типах слитной речи Практически все системы, основанные на чисто статистических методах, не используют знания о том, как человек порождает и воспринимает речь, а также знания о фонологических системах языков и фонетических процессах, происходящих со звуками речи под влиянием тех или иных условий

В ходе исследования были определены фонетические характеристики, которые возможно успешно использовать для анализа звуков речи в разных условиях реализации Были созданы процедуры выделения и идентификации выбранных акустических характеристик из речевого сигнала Все акустические характеристики были проверены в экспериментах на точность интерпретации звуков речи в разных видах речевого материала и условий предварительной обработки, а также методов представления речевого сигнала Эксперименты были проведены на материале самых последних тестовых корпусов на разных языках, общим объемом около 135 часов

Автоматическая интерпретация звуков речи подразумевает предварительную обработку речевого сигнала и акустическое моделирование звуков речи на основе выделенных из сигнала акустических характеристик. Сама интерпретация производится посредством сравнения акустической модели опознаваемого звука речи с эталонными моделями Поэтому в данной диссертационной работе последовательно был описан процесс разработки, выделения и применения акустических моделей звуков речи для автоматической интерпретации звуков речи. В исследовании подробно описаны алгоритмы получения всех акустических характеристик

При моделировании спонтанной речи следует учитывать особенности реализации звуков в спонтанной речи, т к все они являются потенциальными источниками проблем, а также большого количества несоответствий вероятностным распределениям, выведенным на основе исследования «идеальной» речи.

Использование антропоморфных моделей решает эту проблему, так как человек может намного эффективнее понимать спонтанную речь и является своего рода идеальной системой распознавания Фонетический подход к акустическому моделированию звуков речи учитывает то, каким образом образуются звуки речи, как они противопоставляются друг другу в рамках фонологической системы отдельно взятого языка и как они влияют друг на друга в речевом сигнале.

В качестве фонетических характеристик предложены акустические корреляты звонкости, сонорности, местоположения формант, а также места и способа образования согласных

Для оценки эффективности разработанных в ходе диссертационного исследования фонетических характеристик был проведен ряд экспериментов Часть экспериментов проводились на материале вручную выделенных звуков речи В других экспериментах использовалась полноценная система автоматического распознавания речи, где фонетические характеристики использовались в качестве дополнительных к общепринятым акустическим характеристикам Во всех экспериментах были получены успешные результаты

Результаты, представленные в диссертационной работе показывают эффективность разработанных фонетических характеристик, а также общую перспективность применения фонетических характеристик в системах автоматического распознавания речи

Основные положения диссертации отражены в следующих публикациях

1 Кочаров, Д А Автоматическая обработка и распознавание гласных (на материале русского языка) / Д А Кочаров // Материалы XXXII международной филологической конференции, секция фонетики и методики преподавания иностранных языков, часть 1 . изд-во СПбГУ, 2003 – стр 35–38,

2 Кочаров, Д А Моделирование системы автоматического распознавания гласных в шуме (на материале русского языка) / Д А Кочаров // Ученые записки молодых филологов, вып 2 · изд-во СПбГУ, 2004 – стр 214–227,

3 Кочаров, Д А Автоматическое распознавание гласных в потоке речи (на материале русского языка) / Д А. Кочаров // Фонетический лицей, вып. 1 : изд-во СПбГУ, 2004 – стр 43–49,

4 Kocharov, D Automatic Vowel Recognition in Fluent Speech (on the Material of the Russian Language) / D Kocharov // Proc. of SPECOM 2004 Saint-Petersburg, 2004 – pp 308–309,

5 Kocharov, D Articulatory Motivated Acoustic Features for Speech Recognition / D Kocharov, A. Zolnay, R Schluter, H Ney // Proc. of European Conf on Speech Communication and Technology 2005, vol 2 Portugal 2005 – pp 1101–1104;

6 Кочаров, Д А Использование акустической характеристики сонорности для автоматического распознавания речи / Д А Кочаров // Материалы XXXV международной филологической конференции, секция фонетики и методики преподавания иностранных языков : СПбГУ, 2006 – стр 23–27;

7 Kocharov, D. Sonority Measure for Automatic Speech Recognition / D Kocharov // Proc of SPECOM 2006 Saint-Petersburg, 2006 – pp 359–362,

8 Zolnay, A Using Multiple Acoustic Feature Sets for Speech Recognition / A Zolnay, D. Kocharov, R. Schluter, H Ney // Speech Communication, Volume 49, №6, 2007 – pp 514–525;

9 Кочаров, Д А Использование фонетических характеристик для автоматического распознавания речи / Д А Кочаров // Вестник Санкт-Петербургского государственного университета, вып. 2, часть. 2, серия 9 . изд-во СПбГУ, 2007 – стр 45–54.

10 Кочаров, Д А Автоматическое определение частоты основного тона методом анализа через синтез / Д А Кочаров // Материалы XXXVI международной филологической конференции, Секция формальных методов анализа русской речи, вып. 6 СПбГУ, 2007 – стр 70–74;