

## Слайд 1

**Здравствуйте, уважаемые члены аттестационной комиссии!**

Тема моей выпускной квалификационной работы - «...».

**Актуальность темы обусловлена** тем, что кластерная модель позволяет существенно снизить требования к объему обучающей выборки в задаче автоматического распознавания речи. Если скрытой марковской модели для уверенного обучения требуется не менее 20 реализаций каждого слова из распознаваемого словаря, то при использовании кластерной модели для обучения достаточно всего нескольких (в идеальном случае одного) эталонных реализаций каждого слова. Это позволяет не только существенно снизить требования к объему обучающей выборки, но и сократить временные затраты на обучение системы распознавания речи.

## Слайд 2

**Целью работы** является разработка эффективных (по точности и скорости вычислений) алгоритмов автоматического анализа, распознавания и обучения речи на основе кластерных моделей элементарных речевых единиц.

Для реализации цели был поставлен и решен ряд **задач**.

**Объектом** исследования является устная речь, ее математические модели и методы обработки.

**Предметом** исследования являются кластерная модель элементарных речевых единиц (ЭРЕ) и критерий минимального информационного рассогласования.

## Слайд 3

В ходе работы были рассмотрены известные в мировой литературе методы автоматического распознавания речи, в том числе на основе аппарата скрытых марковских моделей.

Для достижения цели, во второй главе работы производится разработка адаптивной кластерной модели элементарных речевых единиц.

Известно, что задача формирования ЭРЕ (которые могут представлять собой, к примеру, фонемы) является неотъемлемой частью фонетического анализа речи (ФАР). Соответственно работа всей системы автоматической обработки во многом зависит от эффективности ФАР. При этом состав и качество базового словаря, или фонетической базы данных, некоторого фиксированного объема  $R < \infty$  во многом определяют достигаемое качество фонетического анализа речи. Поэтому формирование или настройка фонетической базы данных является ключевой для большинства современных методов обработки речевого сигнала. В рамках ИТВР для решения указанной задачи предложена модель переменного дерева и информационный  $(R+1)$ -элемент.

## Слайд 4

Для проведения математического моделирования метода переменного дерева был выбран текст объемом около одной стандартной машинописной страницы, взятый из первой главы романа А.С. Пушкина "Капитанская дочка". Этот текст был проговорен в среднем темпе группой дикторов. Полученные сигналы сначала записывались в память ПК в виде соответствующих звуковых файлов. Для этого применялись специальные программные и аппаратные средства: динамический микрофон AKG D77 S и ламповый микрофонный предусилитель ART TUBE MP Project Series USB.

На слайде показана зависимость от значений  $p_0$  и  $V_0$  числа выявленных ЭРЕ для первого диктора. Видно, что при увеличении обоих порогов сначала происходит резкое уменьшение количества выделенных ЭРЕ.

### **Слайд 5**

Хорошей иллюстрацией к сказанному может служить временная диаграмма речевого сигнала на слайде, которая отображает короткий (2 с) фрагмент выбранного текста в произношении нашего диктора. Здесь для наглядности синхронно с текстом отмечены буквами алфавита соответствующие ЭРЕ русского языка.

### **Слайд 6**

Сделанные выводы остаются справедливыми для всех, без исключения, дикторов. В их подтверждение на слайде представлена зависимость для второго диктора.

Хорошо видно, что она качественно сохраняет прежний вид, хотя суммарное число выделенных ЭРЕ в этом случае оказалось существенно меньшим. Из сказанного можно сделать важный вывод об устойчивости оптимальных значений всех основных параметров обработки речи по методу переменного дерева по отношению к разным текстам, разным дикторам и их фонетическим особенностям речи.

### **Слайд 7**

В ходе работы была рассмотрена эффективность работы предложенного алгоритма. Для этого проведен аналитический и экспериментальный анализ вероятностей ошибок обнаружения фонем во входном сигнале.

Практическая реализация предложенного алгоритма фонетического анализа слитной речи была положена в основу создания информационной системы фонетического анализа слитной речи.

Нами предложен алгоритм формирования адаптивной кластерной модели элементарных речевых единиц для создания новых и обновления существующих ФБД. Разработанный алгоритм является мощным инструментом для фонетического анализа слитной речи: варьируя значения порогов по величине информационного рассогласования и длине ЭРЕ можно в широких пределах изменять его чувствительность.

Анализ эффективности работы алгоритма показал, что он позволяет достичь, по крайней мере, в 1,4 раза меньшей вероятности перепутывания фонем, чем, например, метод на основе известного критерия Пирсона. Следует отметить, что такой анализ проводился по паре проблемных фонем, для которых возникает наибольшее количество ошибок.

### **Слайд 8**

Для реализации предложенного алгоритма фонетического анализа слитной речи был разработан лабораторный образец информационной системы фонетического анализа слитной речи (ИС ФАР). Данная система представляет собой фонетический анализатор.

Блок-схема ИС ФАР показана на слайде. БВР осуществляет извлечение данных из БД, соответствующих требуемому критерию, и отображение их в виде, удобном для пользователя. БОД предназначен для организации возможности работы с группами дикторов и формирования входных данных для БВР.

### **Слайд 9**

Информация, содержащаяся в БД, может быть, как непосредственно считана БВР, так и может поступать в БВР через БОД.

Разработанная информационная система использует базу данных формата Microsoft Access. На слайде показана схема этой базы данных.

## Слайд 10

Повышение точности распознавания речи на сегодняшний день является одной из ключевых задач АРР. В рамках теоретико-информационного подхода на основе кластерной модели ЭРЕ ее решение заключается в учете того факта, что различные сегменты анализируемого слова  $X$  могут вносить различный вклад в сумму предложенного выше алгоритма.

В таблице на слайде показаны оценки вероятности правильного распознавания десяти слов из подготовленной экспериментальной базы для двух различных алгоритмов распознавания: по сегментам с решающей статистикой и по сегментам с решающей статистикой. Оценка вероятности правильного распознавания слов вычислялась по результатам серии испытаний и формуле относительной частоты случайных событий.

Представленные результаты указывают на существенное уменьшение ошибок перепутывания слов, за счет амплитудного ограничения слагаемых в решающей статистике.

## Слайд 11

В таблице на слайде показаны вероятности правильного распознавания слов для решающей статистики в зависимости от значения порога  $U$ . Видно, что максимальная вероятность правильного распознавания достигается при  $U=0,5$ . Таким образом, данное значение ограничения можно выбрать в качестве оптимального.

Повышение точности АРР при использовании статистики можно объяснить, рассматривая эффект накопления в ней величины ИР. Зависимость суммарного информационного рассогласования между сигналом  $X$  и истинным словом-эталоном от количества анализируемых сегментов при применении двух статистик (сплошная линия) и (пунктирная линия) для оптимального порога  $U=0,5$  для одной из реализаций слова "ноль".

В предложенном алгоритме распознавания речи могут также применяться и другие пути повышения быстродействия. Наиболее перспективным подходом в данном случае можно считать использование метода направленного перебора словаря и его модификаций

## Слайд 12

В настоящее время проблема обучения речи людей с нарушениями слуха является весьма актуальной. Общеизвестно, что страдающие нарушениями слуха не могут четко и правильно произносить слова и даже отдельные звуки, поскольку не слышат того, что говорят, и потому не могут контролировать свое произношение. Современная дефектология имеет широкий арсенал методов для постановки произношения у глухих и слабослышащих детей [101 — 103]. Однако все они требуют постоянного вмешательства логопеда и практически не предполагают индивидуальной работы ребенка над собой. Вместе с тем, для сохранения разборчивости речи слабослышащим, а особенно глухим требуется постоянная тренировка собственного произношения. Значительную помощь в данной ситуации могут оказать автоматизированные системы контроля.

Компьютеры и другие технические устройства предоставляют большое число возможностей по улучшению произношения через визуализацию артикуляционного процесса при помощи различных средств отображения. Здесь применяется большое число методов и подходов.

Главный принцип, реализуемый в представленных системах постановки произношения - это сравнение исходного сигнала с некоторым эталоном. Основной

недостаток такого подхода заключается в том, что диктор (обучаемый) в силу ряда причин, может быть в принципе не способным точно повторить эталон.

### **Слайд 13**

Благодаря применению кластерной модели РЕ была разработана новая технология информационного обеспечения в задаче обучения речи людей с дефектами в органах слуха. На слайде представлена схема учебного процесса на ее базе.

Особенностью данной технологии является новый метод визуализации данных в информационной метрике Кульбака-Лейблера, основанный на двух разновидностях кластерных моделей речи: фонетической и морфологической.

Было показано, что обучение глухих детей произношению по разработанной технологии позволяет им улучшить собственно произношение. Об этом наглядно свидетельствуют показатели качества речи каждого из них. У всех обучаемых уровень нечеткости речи (процент ошибок) к концу обучения снизился на 10 — 20%.

Рассмотренная технология может использоваться не только при обучении речи глухих и слабослышащих, но и при решении задачи обучения иностранным языкам и диалектам. Одна из основных проблем при изучении иностранного языка заключается во влиянии звуков национального языка на произношение звуков изучаемого языка. Все это оказывает негативное влияние, как на восприятие, так и на артикуляцию. Разработанная система может быть успешно использована при решении данной проблемы. Кроме того, с помощью режима тестирования качества речи обучаемый может контролировать произношение не только отдельных звуков, но так же и общее качество своей речи.

### **Слайд 14**

На основании полученных результатов можно сделать следующие выводы представленные на слайде.

Из сделанных выводов следует, что предложенные в диссертационной работе алгоритмы на основе кластерной модели элементарных речевых единиц могут иметь практическое применение при решении широкого круга актуальных задач в области анализа, распознавания и обработки речи. Например, это задача разработки разнообразных речевых корпусов в пределах национального языка - как от отдельных его носителей, так и их групп, объединенных по признакам пола, года рождения, места жительства и т.п. с автоматическим наполнением и непрерывным развитием (модернизацией) во времени. На этом примере проявляются все основные преимущества кластерной модели элементарных речевых единиц: высокая чувствительность к рассогласованию данных, способность к эффективной выборочной адаптации, минимум вычислительных затрат и требований к объему запоминающих устройств и другие. Сама идея информационного центра-эталона речевого образа здесь приобретает очевидный практический смысл и дополнительное обоснование.

К числу приоритетных направлений практического применения принадлежит также область современной прикладной лингвистики в части теории и практики преподавания иностранных языков. Разработанная информационная система обучения речи позволяет существенно расширить арсенал используемых здесь технических и программных средств, особенно для самостоятельной работы учащихся.

**Таким образом, Цель работы** - разработка эффективных (по точности и скорости вычислений) алгоритмов автоматического анализа, распознавания и обучения речи на основе кластерных моделей элементарных речевых единиц – **достигнута.**

Спасибо за внимание! Доклад окончен.