**РОССИЙСКАЯ АКАДЕМИЯ НАУК**

**Федеральное государственное бюджетное учреждение науки**

**Российской академии наук**

**Вычислительный центр им. А.А.Дородницына РАН**

«УТВЕРЖДАЮ»

Директор ВЦ РАН

академик РАН,

д.ф.-м.н., профессор

 \_\_\_\_\_\_\_\_\_\_\_\_\_\_ Ю.Г.Евтушенко

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2012 г.

**РАБОЧАЯ ПРОГРАММА**

УЧЕБНОЙ ДИСЦИПЛИНЫ

«Основы речевых технологий»

для подготовки аспирантов по специальности
05.13.11 - Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

Москва 2012

**1. ЦЕЛИ И ЗАДАЧИ**

Цель курса - освоение аспирантами фундаментальных знаний в области обработки и анализа речевой информации, изучение основных проблем компьютерной обработки речи и современных подходов к их решению.

Задачами данного курса являются:

* формирование базовых знаний в области компьютерной обработки речевой и аудио информации как дисциплины, обеспечивающей технологические основы современных инновационных сфер деятельности;
* обучение аспирантов принципам решения задач обработки (цифровой обработки, анализа, распознавания) естественной речи на основе методов машинного обучения и распознавания образов;
* подготовка аспирантов к выполнению самостоятельных исследований в области речевой технологии.

**2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ПОСЛЕВУЗОВСКОГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ (АСПИРАНТУРА)**

Дисциплина «Основы речевых технологий» относится к дисциплинам по выбору учебного плана подготовки аспирантов по научной специальности 05.13.11 «Математическое обеспечение вычислительных машин, комплексов и компьютерных сетей».

Изучение данной дисциплины базируется на следующих дисциплинах подготовки бакалавров или специалистов:

* «Линейная алгебра»;
* «Теория вероятности и математическая статистика»;
* «Программирование и основы алгоритмизации»;
* «Базы данных»;
* «Искусственный интеллект»;
* «Методы оптимизации»;

а также на дисциплинах подготовки магистра:

* «Современные проблемы информатики и вычислительной техники»;
* «История и методология информатики и вычислительной техники»;
* «Компьютерные технологии в науке и образовании».

Для успешного изучения курса аспиранту необходимо знать общесистемное программное и техническое обеспечения автоматизированных систем, а также уметь работать с персональной ЭВМ.

Основные положения дисциплины будут использованы при подготовке к кандидатскому экзамену по научной специальности 05.13.11 «Математическое обеспечение вычислительных машин, комплексов и компьютерных сетей», в научно-исследовательской работе и при выполнении диссертации на соискание ученой степени кандидата физико-математических или технических наук.

**3. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ СОДЕРЖАНИЯ ДИСЦИПЛИНЫ**

В результате изучения дисциплины «Основы обработки текстовой информации» аспирант должен:

* иметь представление: о месте и роли дисциплины «Основы обработки текстовой информации» в своей будущей научной и практической деятельности, о взаимосвязи дисциплины с другими дисциплинами, наукой и техникой; о современных автоматизированных системах, используемых для обработки текстов;
* знать: модели и алгоритмы, применяемые для обработки текстовой информации; современные проблемы обработки текстовой информации; подходы к экспериментальному исследованию качества решения задач обработки текстовой информации;
* уметь: решать задачи из области обработки текстов; проводить самостоятельные научные исследования по теме дисциплины; применять изученные модели и алгоритмы для решения поставленных задач.

**4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ**

**Объем дисциплины и виды учебной работы**

|  |  |  |  |
| --- | --- | --- | --- |
| Вид учебной работы | Всего часов | Семестры | Семестры |
| 1 | 2 |
| Общая трудоемкость дисциплины | 90 | 46 | 44 |
| Аудиторные занятия, в том числе: | 28 | 14 | 14 |
| Лекции | 28 | 14 | 14 |
| Семинары (С) |   |   |   |
| Лабораторные работы (ЛР) |   |   |   |
| Другие виды аудиторных занятий |   |   |   |
| Самостоятельная работа | 62 | 32 | 30 |
| Вид итогового контроля (зачет, экзамен) |   |   | экзамен |

**СТРУКТУРА ПРЕПОДАВАНИЯ ДИСЦИПЛИНЫ**

**Перечень разделов дисциплины и распределение времени по темам**

|  |  |
| --- | --- |
| № темы и название | Количество часов |
| 1. Предмет речевых технологий | 4 |
| 2. Элементарные сведения о речеобразовании и восприянии речи человеком.  | 4 |
| 3. Элементы фонетики. | 4 |
| 4. Обработка сигналов в частотной области. Кратковременный анализ. БПФ. Параметрическое и признаковое описание речевых образов в частотной области. | 8 |
| 5. Параметрическое описание речевых сигналов во временной области. Модель линейного предсказания речи | 8 |
| 6. Кодирование речевых сигналов. Векторное квантование. Примеры современных речевых кодеков | 4 |
| 7. Меры сходства речевых сигналов.Оценка меры схожести для образов различной длительности. Распознавание изолированных слов. | 6 |
| 8. Структура систем распознавания речи. Статистический подход к распознаванию речи. Критерии эффективности работы системы распознавания речи. | 4 |
| 9. Статистические методы моделирования последовательностей образов. Скрытая марковская модель (СММ). Модель смеси нормальных распределений (СГР). Оценка параметров СММ и СГР. | 12 |
| 10. Акустико-фонетическое моделирование в системах распознавания речи. | 8 |
| 11. Статистические модели языка для систем распознавания речи | 8 |
| 12. Алгортмическая реализация процедур распознавания слитной речи. | 8 |
| 13. Методы подстройки параметров системы распознавания речи на голос говорящего | 4 |
| 14. Методы автоматического синтеза речи по тексту | 8 |
| ВСЕГО( зач. ед.(часов)) | 90  |

**ВИД ЗАНЯТИЙ**

**Аудиторные занятия**

|  |  |
| --- | --- |
| № темы и название | Количество часов |
| 1. Предмет речевых технологий | 2 |
| 2. Элементарные сведения о речеобразовании и восприятии речи человеком.  | 2 |
| 3. Элементы фонетики. | 2 |
| 4. Обработка сигналов в частотной области. Кратковременный анализ. БПФ. Параметрическое и признаковое описание речевых образов в частотной области. | 2 |
| 5. Параметрическое описание речевых сигналов во временной области. Модель линейного предсказания речи | 2 |
| 6. Кодирование речевых сигналов. Векторное квантование. Примеры современных речевых кодеков | 2 |
| 7. Меры сходства речевых сигналов. Оценка меры схожести для образов различной длительности. Распознавание изолированных слов. | 2 |
| 8. Архитектура систем распознавания речи. Статистический подход к распознаванию речи. Критерии эффективности работы системы распознавания речи. | 2 |
| 9. Статистические методы моделирования последовательностей образов. Скрытая марковская модель (СММ). Модель смеси нормальных распределений (СГР). Оценка параметров СММ и СГР. | 2 |
| 10. Акустико-фонетическое моделирование в системах распознавания речи. | 2 |
| 11. Статистические модели языка для систем распознавания речи. | 2 |
| 12. Алгортмическая реализация процедур распознавания слитной речи. | 2 |
| 13. Методы подстройки параметров системы распознавания речи на голос говорящего | 2 |
| 14. Методы автоматического синтеза речи по тексту | 2 |
| ВСЕГО( зач. ед.(часов)) | 28 |

**ВИДЫ САМОСТОЯТЕЛЬНОЙ РАБОТЫ**

|  |  |  |
| --- | --- | --- |
|  | Темы | Трудоёмкость в зач. ед.(количество часов) |
| 1 | Проработка и повторение лекционного материала и материала рекомендованной литературы – выполняется самостоятельно каждым аспирантом по итогам каждой из лекций, результаты контролируются преподавателем на лекционных занятиях, используются конспект лекций, учебники, рекомендуемые данной программой | 40 |
| 2 | Самостоятельное изучение отдельных подразделов программы – выполняется каждым аспирантом по заданию преподавателя, результаты контролируются преподавателем на лекционных занятиях, используются материалы, рекомендуемые данной программой | 22 |
|  | ВСЕГО ( зач. ед.(часов)) | 62 часа |

**СОДЕРЖАНИЕ ДИСЦИПЛИНЫ**

**Развёрнутые темы и вопросы по разделам**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|   | Разделы и темы | Содержание | Объем (зачетные единицы - часы) | Общее количество часов |
|   |   | Аудиторная работа  | Самостоятельная работа  |   |
| 1 | Предмет речевых технологий | Основные задачи речевой тех-нологии: распознавание речи; распознавание диктора; компрессия речи;фильтрация речи; синтез речи; распознава-ние языка;понимание речи; обучение произношению; диагностика патологий. Вариативность (интер- и интра) дикторная. Речевой сигнал. Основные этапы генерации и восприятия речи. Визуализация речевого сигнала.  | 2 | 2 | 4 |
| 2 | Элементарные сведения о речеобразовании восприятии речи человеком.  | Восприятие речи. Схема слухового тракта. Естественная частотная шкала слуха Частота колебаний и высота тона. Час-тотные группы слуха. Эффект маскировки. Восприятие громкости звуков. Качество речи. Разборчивость и комфортность. Схема речеобразующего тракта. Частота основного тона. Формантные частоты. | 2 | 2 | 4 |
| 3 | Элементы фонетики. | Фонемы. Классы фонем: гласные, согласные, твердые, мягкие, взрывные, звонкие, глухие. Фонетические алфавиты. Система Аванесова и машинные алфавиты. Алфавит IPA. Алфавит SAMPA. Произносительная транскрипция. Запись произношения в виде фонематической транскрипции. | 2 | 2 | 4 |
| 4 | Обработка сигналов в частотной области. Кратковре-менный анализ. БПФ. Параметри-ческое и признаковое описание речевых образов в частотной области. | Временн**ы**е и частотные параметры. Проблема анализа протяженных сигналов и идея кратковременного анализа. Преобразование Фурье. ДФП. Алгоритм БПФ. Свойства ПФ. Амплитудный спектр. Кратковременное преобра-зование Фурье. Использование оконных функций. Кратко-временный спектр. Цифровая фильтрация: Гребенка ЦФ. Представление ДПФ как гребенки ЦФ. Кепстр. Мел-спектральные и мел-кепстральные коэффициенты. | 2 | 6 | 8 |
| 5 | Параметрическое описание речевых сигналов во временной области. Модель линейного предсказания речи | Модель авторегресии. Постановка задачи предсказания для временных последовательностей. Оценка параметров АР-модели в общем виде. Модель линейного предсказания речи. Физический смысл модели ЛПР. Коэффициенты линейного предсказания (КЛП). Авторегрессионный и автокорреляционный методы вычисления КЛП. Алгоритм Дарбина. | 2 | 6 | 8 |
| 6 | Кодирование речевых сигналов. Векторное квантование. Примеры современных речевых кодеков | Постановка задачи сжатия речевого сигнала. Измерение скорости передачи. A и U законы кодирования. Кодек G 711. Векторное квантование парамтеров речевых сигналов. Алгоритм k-средних. Кодек G723. Кодек MPEG | 2 | 2 | 4 |
| 7 | Меры сходства речевых сигналов. Переход от кратковре-менных параметров к протяженным во времени образам. Меры схожести для образов различной длительности.  | Локальные меры сходства речевых сигналов. Метрика L2.Кепстральные расстояния. Связь со спектральными. Мера Итакуры-Саито (это не расстояние!). Методы нелинейного сравнения (деформации) протяженных образов на основе динамического программирования. Алгоритм Т.К.Винцюка. | 2 | 4 | 6 |
| 8 | Структура систем распознавания речи. Статистичес-кий подход к распознаванию речи. Критерии эффективности работы системы распознавания речи. | Статистическая формулировка проблемы распознавания речи (уравнение для распознавания слитной последовательности слов). Основные компоненты системы распознавания. Количественная оценка эффективности работы системы распознавания речи. Ошибки первого и второго рода. Примеры значений оценок эффективности для современных систем распознавания речи: WER, DER, LER, CER. Оценка эффективости работы систем распознавания ключевых слов – FAHR Оценка эффективости работы систем речевого диалога. | 2 | 2 | 4 |
| 9 | Статистичес-кие методы моделирования последовательностей образов. Скрытые марковские модели. Модель смеси нормальных (гауссовых) распределений (СГР). Оценка параметров СММ и СГР по обучающей выборке. | Скрытые марковские модели (СММ или HMM). Представление речевого сигнала в виде цепи Маркова. Три основные проблемы, связанные с практическим применением СММ. Алгоритм прямого и возвратного хода. Алгоритм Витерби (Viterbi). Непрерывные и дискретные СММ. Оценка параметров дискретной СММ. Основные недостатки СММ. Модель смеси нормальных гауссовых распределений (СГР - GMM). Оценивание параметров СММ-СГР с помощью ЕМ-алгоритма. Алгоритм Баума-Уэлча (Baum-Welch). | 2 | 10 | 12 |
| 10 | Акустико-фонетическое моделирование в системах распознавания речи. | Выбор множества акустико-фонетических моделей. Монофоны, бифоны и трифоны. Дискретные, непрерывные и полунепрерывные СММ. Автоматический выбор алфавита моделей. Использование деревьев бинарных решений для вычисления оптимального множества марковских моделей. Моделирование вариативности произношения. | 2 | 6 | 8 |
| 11 | Статистические модели языка | Что такое модель языка. Грамматики как модели языка. Стандарт SRGS. Вероятностная модель языка. Способы оценки качества модели языка. Перплексия. N-граммные модели языка.Дисконтные методы оценки параметров N граммных моделей языка.  | 2 | 6 | 8 |
| 12 | Алгоритмы поиска - декодирования для слитной речи. Алгоритм перемещения маркера (фишки). | Представление произноси-тельного лексикона в виде графа. Интерпретация распознавания речи как поиска на графе. Алгоритм А\*. Алгоритм перемещения фишки. Интеграция модели языка в процедуру поиска на лексической сети. Сохранение вычислительной эффектив-ности алгоритма. Генерация списка N лучших гипотез. Проблема кроссвордов | 2 | 6 | 8 |
| 13 | Методы настройки системы распознавания на голос диктора | Адаптация признаков речевого сигнала в частотной области по алгоритму VTLN. Настройка на голос путем адаптации параметров GMM: алгоритм MLLR | 2 | 2 | 4 |
| 14 | Синтез речи | Синтез речи. Компилятивный, формантный и артикуляторный синтез.Методы OLA и PSOLA. Метод Unit Selection. Оценки качества речевого сигнала: субьективные и обьективные. | 2 | 6 | 8 |
| ВСЕГО | 28 | 62 | 90 |

**5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ**

 **В учебном процессе используются следующие образовательные технологии:**

|  |  |  |  |
| --- | --- | --- | --- |
| № | Вид занятия | Форма проведения занятий | Цель |
| 1 | Лекция | Изложение теоретического ма териала | Получение теоретических знаний по дисциплине |
| 2 | Лекция | Изложение теоретического ма териала с помощью презентаций | Повышение степени понимания материала |
| 3 | Лекция | Разбор конкретных примеров применения современных технологий обработки текстов | Осознание связей между теорией и практикой, а также взаимозависимостей разных дисциплин |
| 4 | Самостоятельная работа аспирантов | Самостоятельное изучение отдельных подразделов программы.Самоподготовка (проработка и повторение лекционного материала и материала рекомендованной литературы) | Повышение степени понимания материала  |

**6. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ И УЧЕБНО\_МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ АСПИРАНТОВ.**

**Форма контроля знаний:**

**- кандидатский экзамен по специальности.**

**Контрольно-измерительные материалы**

На кандидатском экзамене аспирант должен продемонстрировать знания в объеме основной программы кандидатского экзамена по специальности 05.13.11 «Математическое обеспечение вычислительных машин, комплексов и компьютерных сетей», а также дополнительной программы, в которую, в зависимости от выбранной аспирантом специализации, могут входить вопросы, рассматриваемые в данном курсе.

**Перечень контрольных вопросов для дополнительной программы:**

|  |  |
| --- | --- |
| 1 | Назовите и опишите основные типы задач, которые решает речевая технология |
| 2 | Особенности слухового восприятия. Эффект маскировки, критические полоски слуха. Как особенности восприятия используются в речевой техноогии, в частности в алгоритмах компрессии и подавления помех в речи.  |
| 3 | Фонемы, фоны. Как соотносятся буква и фонемы. Что такое аллофон. Пример алфавита фонем для русского языка. Произносительная транскрипция слова. Каноническая транскрипция. Произношение, вариативность произношения в разговорной речи, способы моделирования вариативности произношения. |
| 4 | Статистическая формулировка проблемы распознавания речи. Что понимается под моделью языка, акустико-фонетической моделью, моделью произношения. |
| 5 | Дискретное преобразование Фурье последовательностей. Определение и основные свойства. Быстрое преобразование Фурье (БПФ).  |
| 6 | Кратковременный анализ речевых сигналов. Кратковременный спектр. Оконные функции. Амплитудный спектр. Мел-спектр. |
| 7 | Что такое кепстр сигналов и Мел-кепстр. Алгоритм вычисления мел-кепстральных коэффициентов речевого сигнала. |
| 8 | Модель линейного предсказания речи (ЛПР). Коэффициенты линейного предсказания. Автокорреляционный метод оценки коэффициентов ЛПР. Алгоритм Дарбина. |
| 9 | Локальные меры схожести для речевых сигналов: Метрика L2. Cпектральное расстояние. Кепстральное расстояние, усеченное кепстральное расстояние. Информационные меры, мера Итакуры-Саито. |
| 10 | Скрытая Марковская модель речевых сигналов. Основные параметры СММ и их физический смысл. Три основные проблемы, связанные с использованием СММ. |
| 11 | Вычисление полной вероятности для наблюдений с помощью процедуры прямого и обратного хода.  |
| 12 | Вычисление вероятности наилучшей последовательности состояний с помощью процедуры Витерби |
| 13 | Основные виды СММ. Вычисление вероятностей наблюдений для дискретных и непрерывных СММ. |
| 14 | Оценка параметров СММ. Процедура Баума – Уэлча (Baum-Welch). |
| 15 | Количественная оценка эффективности работы системы распознавания речи: точность распознавания в терминах пословная и др. ошибок распознавания (WER, DER, LER), операционная характеристики приемника (ROC). Показатели эффективности работы систем обнаружения ключевых слов (FOM, FAHR).  |
| 16 | Методы акустико-фонетического моделирования в системах распознавания речи. Контексто-зависимые модели звуков: монофоны, бифоны, трифоны. Использование деревьев бинарных решений для нахождения оптимального множества марковских моделей. |
| 17 | Представление языка в диалоговых системах распознавания речи. Стандарт SRGS. |
| 18 | Статистическая модель языка. Энтропия и перплексия языка. Вероятностные n-граммные модели, основные методы оценки параметров моделей языка. Дисконтные модели. Модель классов слов. Триггерные модели. |
| 19 | Синтез речи. Компилятивный, формантный и артикуляторный синтез. Алгоритмы OLA и PSOLA. Метод Unit Selection. |
| 20 | Методы быстрой настройки параметров моделей на голос говорящего. Алгоритмы VTLN и MMLR. |
| 21 | Представление произносительного лексикона в виде префиксного графа. Алгоритм А\*. Алгоритм перемещения фишки. Рекомбинация путей в вершинах графа. Интеграция модели языка в процедуру поиска на лексической сети. Пререживание путей и друнгие способы поддержания вычислительной эффективности процедур распознавания речи.  |

**7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

**Необходимое оборудование для лекций и практических занятий:** Компьютер и мультимедийное оборудование (проектор, звуковая система)

**Необходимое программное обеспечение:** ОС Microsoft Windows, Linux, MS Office, включая MS PowerPoint, любой браузер для доступа в Интернет

**Обеспечение самостоятельной работы** - базы данных по журналам Computational Linguistics, ACL Journal

**8. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

 **Основная литература**

* Ронжин А.Л., Карпов А.А., Ли И.В. Речевой и многомодальный интерфейсы, Наука, 2006 г.
* Потапова Р.К. Речевое управление роботом: лингвистика и современные автоматизированные системы / Р.К. Потапова. - М.:КомКнига, 2005. - 328 с.
* Винцюк Т.К. Анализ, распознавание и смысловая интерпретация речевых сигналов, Наукова думка, 1987
* Рабинер Л, Шафер Р. Цифровая обработка речевых сигналов, М., Радио и связь, 1981
* Маркел Дж., Грей. Линейное предсказание речи, М., Радио и связь, 1980
* Методы автоматического распознавания речи под ред. У.Ли, М.Мир 1983 г.
* Corpus-Based Methods in Language and speech processing, Kluwer Academic Publishers, 1997 [библиотека ВЦ РАН]
* X.Huang, Acero А. Spoken Language Processing: a Guide to Theory, Algorithm and System Development, 2001, Prentice Hall [ доступна в http://[www.ccas.ru](http://www.ccas.ru)/sites/speech]
* Rabiner L., Juang B.-H. Fundamentals of Speech Recognition. Prentice Hall, 1993 , Prentice Hall PTR [доступна в http://[www.ccas.ru](http://www.ccas.ru)/sites/speech]

**Информационные ресурсы:** **Журналы**

* Proceedings of IEEE Conference on Acoustic, Speech and Signal Processing (1980-2012)
* Proceedings of InterSpeech Conference (1995-2012 гг.)
* Proceedings of SPECOM (Speech and Computer Int.Conference) (1998- 2012 гг.)
* Speech Communication, Elsevier
* Computer Speech and Language, Academic Press, Elsevier
* IEEE Transactions on Speech and Audio Processing,
* IEEE Transactions on Audio, Speech and Language Processing ( IEEE Signal Processing Society)
* Речевые технологии, м., Нар.образование (2008-2012)

**Электронные ресурсы, включая доступ к базам данных и т.д.**

* htk.eng.cam.ac.uk/ - HTK Toolkit - сайт Кембриджский университета
* <http://www.spectrum.uni-bielefeld.de/~gibbon/gibbon_handbook_1997> - eaglebook
* <http://speech.cs.cmu.edu/> сайт университета Карнеги-Меллона, США

Программу составил к.ф.-м.н. Чучупал В.Я.

Принята на заседании ученого совета ВЦ РАН

Протокол № \_\_\_\_\_ от \_\_\_\_ декабря 2012г.