

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Локтионова Оксана Геннадьевна
Должность: проректор по учебной работе
Дата подписания: 31.05.2022 13:03:49
Уникальный программный ключ:
0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4851fda56d089

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра программной инженерии

УТВЕРЖДАЮ
Проректор по учебной работе
О.Г. Локтионова
«15» 12 2017 г.



МЕТОДЫ МАШИННОГО АНАЛИЗА И СИНТЕЗА РЕЧИ, АВТОМАТИЧЕСКАЯ ОБРАБОТКА ЗВУКА

Методические указания и задания к выполнению
лабораторных и практических работ
для студентов вузов

Курск 2017

УДК 004.2

Составители: В.В. Ефремов, И.Н. Ефремова

Рецензент

Кандидат технических наук, доцент *Т.М.Белова*

МЕТОДЫ МАШИННОГО АНАЛИЗА И СИНТЕЗА РЕЧИ,
АВТОМАТИЧЕСКАЯ ОБРАБОТКА ЗВУКА: методические указания к
выполнению лабораторных работ / Юго-Зап. гос. ун-т; сост.:
В.В.Ефремов, И.Н. Ефремова. Курск, 2017. - 14 с.

Содержат описание лабораторных и практических работ по дисципли-
нам.

Предназначено для студентов вузов

Текст печатается в авторской редакции

Подписано в печать

Формат 60x84 1/16.

Усл. печ. л. 1,8. Уч.-изд. л. 1,7. Тираж 100 экз. Заказ 48. Бесплатно

Юго-Западный государственный университет.

305040, г. Курск, ул. 50 лет Октября, 94.

Содержание

1 ПО компьютеров в лингвистике.....	4
2 Машинный анализ речи	7
3 Машинный синтез речи	10
Список используемых источников.....	14

Программное обеспечение (ПО) компьютеров в лингвистике.

Цель работы: Ознакомление с назначением и составом основных информационных и программных средств дисциплин. Получение практических навыков.

Задачи: Научиться работать с основными информационными и программными средствами изучаемых дисциплин.

Содержание отчёта

1. Титульный лист.
2. Формулировка задачи.
3. Ход работы
4. Ответы на контрольные вопросы.

Контрольные вопросы

1. Назначение ПО?
2. Роль корпуса речи?
3. ПО для работы со звуком?

Задание

Исследовать возможности речевого подкорпуса национально-го корпуса русского языка.

Научиться работать в программах Elan, PRAAT. Создать аннотацию звукового файла.

Теоретические и справочные сведения

ELAN – это очень удобная программа для аннотации аудио- или видео-записей. Программа ELAN – разработана в НИИ Психолингвистики им. Макса Планка в Неймегене, Голландия, специально для архивирования записей языков, находящихся под угрозой исчезновения, в рамках проектов Фонда Фольксвагена ELAN можно бесплатно скачать со следующего сайта, где можно скачать также и другие программы, полезные в работе лингвиста:

<http://www.mpi.nl/research/research-projects/language-archiving-technology/tools>

В ELANе можно работать либо с аудиофайлом, либо с аудио- и видеофайлом вместе.

Если Вы хотите работать с аудио- и видеофайлом вместе, нужно сначала выбрать аудиофайл (.wav), потом выбрать видеофайл (.mpg). Открывается главное окно ELANa, где видно либо только звуковые волны (если выбрали только аудиофайл), либо звуковые волны и видео (если выбрали аудио- и видеофайл).

Аннотация звукового файла — это его текстовое сопровождение. При создании аннотации участку звуковой волны ставится в соответствие текст. При этом можно использовать несколько текстовых подстрок, например, на разных языках.

Программа PRAAT также является программой свободного доступа.

Интерфейс программы состоит из двух окон. Главное окно программы содержит главное меню и все функциональные кнопки. Вспомогательное окно программы используется для графического отображения звука и работы с ним.

Praat позволяет пользователю работать как со звуковыми файлами, так и с записями собственного голоса. Любой только что загруженный файл или новая запись добавляются в список на главном окне программы. Если выделить любой объект из списка, то появятся кнопки, отображающие все возможные действия применимые к нему.

С точки зрения анализа звучащей речи лингвисту полезна кнопка «View end Edit». При ее нажатии, появится новое окно, содержащее визуализации звука. Программа может представить звук

как спектрограмму, графическое отображение формы волны, диаграмму изменения частоты тона, диаграмму интенсивности, длительности и формантной составляющей. Для примера в программу будет загружен звук, содержащийся в *.wav файле: это слово «forecast», произнесенное диктором с английского канала BBC. Ниже отображены результаты визуализаций.

Программа Praat отображает все визуализации в одном окне. Пользователь может включать и выключать отображение визуализаций, проставляя или убирая галочки в пунктах меню, соответствующих визуализациям. В данном примере включены все пункты.

Черный цвет: графическое отображение формы волны.

Синий: пульсации (сегменты с резко возрастающей интенсивностью)

Серый: спектрограмма.

Красный: формантная составляющая.

Голубой: изменение частоты основного тона.

Желтый: интенсивность.

Для загрузки больших звуковых файлов необходимо воспользоваться пунктом меню Open -> Open long sound file. Данное действие позволит программе хранить временные файлы на жестком диске.

Предусмотрена в программе и возможность создания панели, где можно прописать транскрипцию под или над соответствующими секторами визуализаций. Какой либо системы для удобного поиска и записи фонетических знаков не обнаружено.

Для добавления панели необходимо нажать кнопку «Annotate», после выделения звука в списке загруженных объектов. В появившемся окне в первом поле прописываем название панели и жмем кнопку «ОК».

В главном окне появится еще один объект с именем «TextGridX», где X – порядковый номер. Далее пользователю необходимо выделить сразу два объекта из списка на главном окне: звук и соответствующий ему «TextGridX». Затем нажать кнопку «View and Edit»

Далее используется раздел меню Boundary в новом окне для деления панели на интервалы. Пользователь выделяет мышью часть диаграммы и нажимает пункт меню Boundary -> Add on se-

lected tier, чтобы создать границу интервала. Далее, выделяя мышью разные интервалы, пользователь может вводить фонетические символы в них.

Также программа может производить простейшие виды редактирования: вырезание, копирование и вставка отрезков звука, которые вызываются соответствующими пунктами главного меню Edit -> Cut | Copy | Paste. Пользователь может выбрать интервалы на любой из визуализаций. [1]

Машинный анализ речи.

Цель работы: Ознакомление со средствами машинного анализа речи. Получение практических навыков.

Задачи: Провести анализ звучащей речи.

Содержание отчёта

1. Титульный лист.
2. Формулировка задачи.
3. Ход работы
4. Ответы на контрольные вопросы.

Контрольные вопросы

1. Общий принципы анализа речи?
2. Средства анализа речи?
3. Задачи анализа речи?

Задание

Провести анализ звукового файла, содержащего речевую информацию с помощью программы PRAAT, изучив фонетические и просодические закономерности звука.

Теоретические и справочные сведения

Основная часть голосовых звуков - в этом случае ее обычно называют частотой основного тона - соответствует частоте колебания голосовых связок. Исходящий из голосового аппарата звук кроме основного тона включает также многочисленные сопутствующие ему тоны. Основной тон и эти дополнительные тоны и составляют сложный звук. Если частота сопутствующих тонов превосходит частоту основного тона в целое число раз, то такой звук называют гармоническим. Сами же сопутствующие тоны и соответствующие им спектральные составляющие в амплитудно-частотном спектре звука называют гармониками. Расстояния по шкале частот между соседними гармониками соответствуют частоте основного тона, то есть частоте колебания голосовых связок.

Во время произнесения любого гласного колеблющиеся голосовые связки создают сложный звук, спектр которого состоит из ряда гармоник с постепенно уменьшающейся амплитудой. Для всех гласных спектр звука, создаваемого голосовыми связками, одинаков. Различие в звучании гласных достигается благодаря изменениям конфигурации и размеров воздушных полостей голосового тракта. Так, например, когда мы произносим звук "и", мягкое небо перекрывает доступ воздуха в носовую полость и к небу поднимается передняя часть спинки языка, в результате чего ротовая полость приобретает определенные резонансные свойства, видоизменяя первоначальный спектр созданного голосовыми связками звука. В этом спектре появляется ряд специфических для данного гласного звука пиков амплитуды спектральных составляющих, называемых спектральными максимумами. В таком случае говорят об изменении огибающей спектра звука. Энергетически наиболее выраженные спектральные максимумы, обусловленные работой голосового тракта как резонатора и фильтра, называют формантами. Форманты обозначают порядковыми номерами, причем первой формантой считают ту, которая следует сразу после частоты основного тона.

В настоящее время анализ звуков осуществляется с помощью устройств или программ, таких, как анализаторы частот и динамические спектрографы, или сонографы. В первом случае представ-

ляется возможность оценивать амплитудно-частотные характеристики звука, но без учета изменений свойств звука во времени. Применение динамических спектрографов дает возможность рассматривать звук в системе трех координат: частота и амплитуда колебаний, и время. Выраженность его спектральных составляющих, то есть интенсивность звучания, отражается на динамической спектрограмме степенью ее затемнения. Изображение сложных звуков в виде динамических спектров впервые было применено для анализа речи и получило название "видимая речь". Пока это является единственным более или менее удовлетворительным приемом преобразования звукового образа в зрительный.

Осциллограмма дает картину изменения во времени амплитуды звука. Таким образом, осциллограмма отражает амплитудно-временные характеристики звукового сигнала: его длительность и форму - то есть амплитудную модуляцию.

Спектрограмма существует для измерения и отображения спектра сигнала – распределения энергии сигнала по частотам. Для этого используется преобразование Фурье, современными анализаторами - FFT – быстрое преобразование Фурье.

Машинный синтез речи.

Цель работы: Ознакомление с назначением и принципами работы средств синтеза речи. Получение практических навыков.

Задачи: Изучить методы работы компилятивного синтеза.

Содержание отчёта

1. Титульный лист.
2. Формулировка задачи.
3. Ход работы
4. Ответы на контрольные вопросы.

Контрольные вопросы

1. Способы синтеза речи?
2. Назначение задачи?
3. Плюсы и минусы различных способов синтеза?

Задание

Наработать голосовую базу. Используя компилятивный способ, синтезировать речевую информацию.

Теоретические и справочные сведения

Автоматический синтез звучащей речи представляет собой процесс преобразования печатного текста, существующего в цифровой форме, в звучащий текст на естественном человеческом языке.

Все способы синтеза речи можно подразделить на группы:

- параметрический синтез;
- конкатенативный, или компиляционный (компилятивный) синтез;
- синтез по правилам;
- предметно-ориентированный синтез.

Параметрический синтез речи является конечной операцией в [вокодерных системах](#), где речевой сигнал представляется набором небольшого числа непрерывно изменяющихся параметров. Параметрический синтез целесообразно применять в тех случаях, когда набор сообщений ограничен и изменяется не слишком часто. Достоинством такого способа является возможность записать речь для любого языка и любого [диктора](#). Качество параметрического синтеза может быть очень высоким (в зависимости от степени сжатия информации в параметрическом представлении). Однако параметрический синтез не может применяться для произвольных, заранее не заданных сообщений.

Компиляционный синтез сводится к составлению сообщения из предварительно записанного словаря исходных элементов синтеза. Размер элементов синтеза не меньше слова. Очевидно, что содержание синтезируемых сообщений фиксируется объёмом словаря. Как правило, число единиц словаря не превышает нескольких сотен слов. Основная проблема в компилятивном синтезе — объёмы памяти для хранения словаря. В связи с этим используются разнообразные методы сжатия/кодирования речевого сигнала. Компилятивный синтез имеет широкое практическое применение. В западных странах разнообразные устройства (от военных самолётов до бытовых устройств) оснащаются системами речевого ответа. В

России системы речевого ответа до недавнего времени использовались в основном в области военной техники, сейчас они находят всё большее применение в повседневной жизни, например, в справочных службах операторов сотовой связи при получении информации о состоянии счета абонента.

Полный синтез речи по правилам (или синтез по печатному тексту) обеспечивает управление всеми параметрами речевого сигнала и, таким образом, может генерировать речь по заранее неизвестному тексту. В этом случае параметры, полученные при анализе речевого сигнала, сохраняются в памяти так же, как и правила соединения звуков в [слова](#) и [фразы](#). Синтез реализуется путём моделирования речевого тракта, применения аналоговой или цифровой техники. Причём в процессе синтезирования значения параметров и правила соединения фонем вводят последовательно через определённый временной интервал, например 5—10 мс. Метод синтеза речи по печатному тексту (синтез по правилам) базируется на запрограммированном знании акустических и лингвистических ограничений и не использует непосредственно элементы человеческой речи. В системах, основанных на этом способе синтеза, выделяется два подхода. Первый подход направлен на построение модели рече-производящей системы человека, он известен под названием *артикуляторного синтеза*. Вторым подходом — [формантный синтез по правилам](#). Разборчивость и натуральность таких синтезаторов может быть доведена до величин, сравнимых с характеристиками естественной речи.

Синтез речи по правилам с использованием предварительно запомненных отрезков естественного языка — это разновидность синтеза речи по правилам, которая получила распространение в связи с появлением возможностей манипулирования речевым сигналом в оцифрованной форме. В зависимости от размера исходных элементов синтеза выделяются следующие виды синтеза:

- микросегментный (микроволновый);
- [аллофонический](#);
- дифонный;
- [полуслоговой](#);
- слоговой;

- синтез из единиц произвольного размера.

Обычно в качестве таких элементов используются полуслоги — сегменты, содержащие половину согласного и половину примыкающего к нему гласного. При этом можно синтезировать речь по заранее не заданному тексту, но трудно управлять интонационными характеристиками. Качество такого синтеза не соответствует качеству естественной речи, поскольку на границах сшивки дифонов часто возникают искажения. Компиляция речи из заранее записанных словоформ также не решает проблемы высококачественного синтеза произвольных сообщений, поскольку акустические и просодические (длительность и интонация) характеристики слов изменяются в зависимости от типа фразы и места слова во фразе. Это положение не меняется даже при использовании больших объёмов памяти для хранения словоформ.

Предметно-ориентированный синтез компилирует слова, записанные заранее, а также фразы для создания полных речевых сообщений. Он используется в приложениях, где многообразие текстов системы будет ограничено определенной темой/областью, например объявления об отправлении поездов и [прогнозы погоды](#). Эта технология проста в использовании и достаточно долго применялась в коммерческих целях: её так же применяли при изготовлении электронных приборов, таких как говорящие часы и [калькуляторы](#). Естественность звучания этих систем потенциально может быть высокой благодаря тому, что многообразие видов предложений ограничено и близко с соответствием интонацией исходных записей. А так как эти системы ограничены выбором слов и фраз в базе данных, они в дальнейшем не могут иметь широкое распространение в сферах деятельности человека, лишь потому, что способны синтезировать комбинации слов и фраз, на которые они были запрограммированы.

Синтезировать речь можно с помощью специального ПО. Сервисы синтеза предоставляют также такие ресурсы как <http://text-to-speech.imtranslator.net> и др.

Список используемых источников

- 1. Щипицина Л.Ю.** Информационные технологии в лингвистике [Текст] учебное пособие.- Москва: Флинта: Наука, 2015.- 128 с
- 2. Зубов А. В.** Информационные технологии в лингвистике [Текст] : учебник / А. В. Зубов, И. И. Зубова. - Москва: Академия, 2012. - 208 с.
- 3. Баранов А. Н.** Введение в прикладную лингвистику [Текст] : учебное пособие / А. Н. Баранов. - 2-е изд., испр. - М.: УРСС, 2003. - 360 с.
- 4. Елисеева О.Е.** Речевой интерфейс. Лабораторный практикум. Учебно-методическое пособие; под ред.проф. В.В.Голенкова: в 2-х Ч., Ч.1/ О.Е.Елисеева: - Минск: БГУИР, 2008.-44 с.