

Модуль I. Лекция 4

Тема: Области применения информационных технологий в лингвистике

1. Автоматический анализ и синтез звучащей речи
2. Автоматическое распознавание текста

1. Автоматический анализ и синтез звучащей речи

Одним из первых важных шагов использования информационных технологий в лингвистике является цифровизация текстов — переводение языкового материала, существующего в печатном или устном виде, в цифровую форму. Именно в этом случае появляется возможность привлечения компьютеров для выполнения определенных операций над текстами на естественном языке: их преобразования, выделения из них отдельных элементов и создания (синтеза) аналогичных текстов.

В связи с принципиальными различиями в способах цифровизации и обработки звучащей речи и печатных текстов в нашей работе эти явления рассматриваются в разных параграфах. Первый параграф посвящен вопросам автоматической обработки и синтеза звучащей речи, а во всех последующих рассматриваются автоматические операции, производимые над печатными текстами. При автоматическом анализе звучащей речи она преобразуется в печатный текст, над которым можно производить дальнейшие операции. Автоматический синтез звучащей речи представляет собой обратный процесс преобразования печатного текста, существующего в цифровой форме, в звучащий текст на естественном человеческом языке.

Процесс автоматического анализа речи включает следующие этапы:

- 1) ввод звучащей речи в компьютер с помощью микрофона.
- 2) выделение компьютерной программой в звуковом потоке отдельных знаков.
- 3) идентификация выделенных знаков звучащей речи со знаками языка.

Минимальными знаками звучащей речи являются звуки, производимые артикуляторным аппаратом человека. Каждый звук имеет свои акустические характеристики (высота, частота колебаний звуковых волн и т.д.), которые можно измерить специальными приборами (например, осциллографом).

Параметры звукового сигнала непрерывно меняются, и такой(непрерывный) тип сигнала называется аналоговым. В отличие от аналогового, цифровой сигнал представляет собой набор дискретных(отдельных) числовых значений, фиксирующих разные уровни звуковой волны. При использовании микрофона аналоговый звуковой сигнал преобразуется в аналоговый электрический, который с помощью аналогово-цифровых преобразователей, встроенных в звуковые карты современных компьютеров, переводится в дискретный цифровой сигнал.

Первые устройства автоматического распознавания устной речи, которых на сегодняшний день большинство, в качестве выделяемых в речевом потоке знаков использовали не звуки, а слова. Слова вводимой в компьютер речи идентифицировались со словами, заранее записанными диктором, читающим слова. Но такой тип распознавания речи связан с определенными ограничениями:

- личность говорящего: автомат распознает речь только определенного говорящего,

- запас слов: автомат распознает только ограниченное количество слов,

- подготовленность речи: автомат распознает речь, лишь если она подготовлена [3].

Для преодоления этих ограничений требуется, чтобы компьютерная программа распознавала не слова, а звуки, т.е. работала не с дискретной речью (которая содержит паузы между словами), а со слитной естественной человеческой речью.

В основе по фонемного распознавания звуков речи лежит анализ

1) длительности и динамики звучания, 2) чередования акустического сигнала и пауз. При этом на основе универсальной классификации звуков Гуннара Фанта, Морриса Халле и Романа Якобсона акустические признаки звуков выводятся из артикуляционных. Правда, акустические признаки в отношении к артикуляционным оказываются недостаточно универсальными. Кроме того, в этой теории недостаточно учитывается слоговоеделение, акцентуация и ритм.

В настоящее время наиболее доступной формой точной фиксации звучащей речи становится спектрограмма — фотографическое изображение звуков. Результаты наблюдений показывают, что в произнесении звуков активно используются четыре частоты называемые формантами. Так, на рис.3 изображены форманты русских звуков и и у. При переходе от звука и к звуку j наиболее заметно изменение форманты F2 (рис.3).

Задачей автоматического анализа звучащей речи при использовании спектрограмм становится перевод спектрограмм в фонологическую транскрипцию.

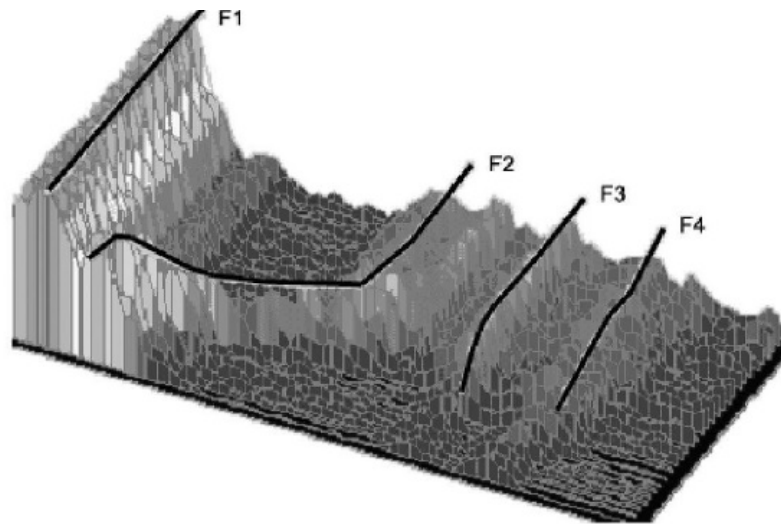


Рис. 3. Спектрограмма русских звуков и и у [Фролов, Фролов]

В итоге процесс автоматического анализа речи включает ввод слов в компьютер через микрофон, начитанных разными дикторами, их спектральную обработку и создание набора признаков, своеобразного образца слова, который выступает знаком языка. При распознавании звучащей речи реальные признаки составляющие ее единицы сравниваются с признаками и образцами слов, существующими в памяти машины. Результатом сравнения является транскрипция или орфографическая запись слова.

Но при автоматическом анализе слитной речи дополнительную трудность составляет отсутствие четких границ между словами. Человек для преодоления этой трудности кроме акустических сигналов обычно использует самые разные другие источники информации: ситуацию, контекст, структуру языкового высказывания, прошлый опыт в данной области и т.п. Аналогичные

правила ученые пытаются применить и к машинам и стремятся задействовать в современных системах анализа речи кроме акустического другие уровни системы языка: лексический, синтаксический, семантический, прагматический.

Включение семантического уровня в автоматический анализ речи приводит, в частности, к следующим последствиям:

1) машина устанавливает, что введенные предложения многозначны и правдоподобны;

2) машина прогнозирует, что в определенных речевых контекстах могут возникать определенные типы общения; в зависимости от такого прогнозируемого типа общения машина интерпретирует предложение.

Очевидно, что создание систем анализа речи такого сложного уровня предусматривает сотрудничество представителей самых разных специальностей. Для экономии времени и усилий ученых и практиков различные компании, в том числе Microsoft, выпускают средства анализа и синтеза речи в виде программных модулей и интерфейсов.

Программисты, не обладающие познаниями в области лингвистики, математики и биологии, могут использовать готовые интерфейсы и программные модули в собственных разработках.

Правда, в этом случае речевые возможности программ будут ограничены использованными средствами и технологиями. Например, многие средства анализа и синтеза речи неспособны работать с русским языком, что ограничивает их использование в России [9].

Можно назвать следующие примеры программ, в которых применяются средства автоматического анализа речи:

- программы голосового управления компьютером и бытовой техникой VoiceNavigatori Truffaldino (компания «Центр речевых технологий», С.-Петербург);

- комплекс голосового управления мобильным телефоном DiVo («Центр речевых технологий»);

- программный модуль Voice Key для идентификации личности по парольной фразе длительностью 3—5 секунд («Центр речевых технологий»);

- программы диктовки текста на английском языке: VoiceType Dictation(IBM), DragonDictate («DragonSystems»); на русском языке: Комбат («БайтГруп») к Диктограф («Voice Member Technology»);

- система распознавания речи, встроенная в Microsoft OfficeXP (работает только с английским языком);

- голосовой поиск (например, в поисковой системе Google).

Так, программа VoiceNavigator позволяет запускать компьютерные приложения и выполнять заданные команды голосом без использования клавиатуры. Перед применением программы ее необходимо обучить, произнеся в микрофон слова команд (команды можно произносить на любом языке и любым голосом). Чтобы программа начала распознавать голосовые команды, ее необходимо «разбудить», произнеся ключевое слово.

Использование модулей распознавания речи весьма перспективно в различных областях деятельности: в обслуживании клиентов, проведении судебных экспертиз, биометрии, обучении, научных исследованиях и т.д. Но массовое внедрение речевых технологий тормозится высокой стоимостью разработок и предлагаемых технологий, а также их пока еще низким качеством.

В целом задача автоматического анализа речи является весьма сложной и решена лишь отчасти. В сравнении с ней задача

автоматического синтеза речи оказывается более простой, и с примерами ее массового использования в обиходной жизни мы сталкиваемся постоянно. В частности, автоматически синтезируется речь в следующих ситуациях:

- называние текущего времени по телефону,
- объявление остановок в метро,
- называние остатка средств на счету и другие услуги мобильных операторов,
- оповещение систем гражданской безопасности и т.д.

Автоматический синтез (генерация) речи в настоящее время осуществляется путем составления слови фраз и заранее записанных диктором образцов отдельных звуков (метод компилятивного синтеза) или путем моделирования речевого тракта человека (формантно-голосовой метод).

Первый метод используется главным образом для синтеза относительно небольшого и заранее известного набора фраз. При этом обеспечивается довольно высокое качество звучания, поскольку синтезируемая речь базируется на элементах естественной человеческой речи. Тем не менее на стыке составляемых звуковых фрагментов возможны интонационные искажения и разрывы, заметные на слух.

Кроме того, создание крупной базы данных звуковых фрагментов, учитывающей все особенности произношения фонем с разными интонациями, представляет собой сложную и кропотливую работу.

Второй метод оказывается более сложным, поскольку здесь необходимо точное моделирование особенностей речевого тракта человека, а также учет интонационной модуляции речи. В силу

названных особенностей формантно-голосовая модель обладает относительно низкой точностью синтезируемых звуков речи.

В качестве примера программы, синтезирующей речь, можно назвать программу Govorka. Основные особенности данной программы состоят в следующем:

- программа читает текст разными голосами и на разных языках, в том числе на русском;
- исходный текст для чтения может быть загружен из текстового файла или набран в окне программы при помощи клавиатуры;
- можно сохранить результаты синтеза речи, записав файл формата WAV или MP3.

Таким образом, несмотря на мощность современных компьютеров, проблема оснащения компьютера полноценным речевым интерфейсом еще далека от своего завершения. Главной проблемой при создании программ автоматического распознавания речи является то, что компьютер не умеет работать со смыслом. В синтезе речи уже имеются определенные достижения, которые внедрены в массовую практику.

2. Автоматическое распознавание текста

Автоматический анализ печатного текста, также как и анализ звучащей речи, начинается с его ввода в компьютер. Поскольку современный человек имеет дело главным образом с информацией, размещенной на печатных носителях, остановимся на процессах автоматической обработки печатных текстов подробнее и посвятим им несколько отдельных параграфов. В первом из этих параграфов рассмотрим ввод печатного текста в компьютер и связанный с этим процесс распознавания печатного текста.

Для ввода информации в компьютер используются специальные устройства — клавиатура, мышьи др., но наиболее удобным инструментом для ввода большого количества печатных текстов является сканер.

Сканер — это устройство ввода, работающее по принципу фотоаппарата, т.е. позволяющее компьютеру «увидеть» текст в виде-фотографии. Чтобы компьютер смог «понять» этот текст, т.е. перевести графическое (растровое) изображение символов в текстовую форму, при которой у каждого символа имеется свой двоичный код (например, в системе кодировок ASCII), требуется программа автоматического распознавания символов (англ. OCR = Optical Character Recognition).

Символами являются любые буквы, знаки препинания и другие знаки текста (апостроф, кавычки, тире, скобки и т.д.). Слово понимается как последовательность символов между двумя соседними пробелами.

Программа автоматического распознавания текста (OCR-программа)— это компьютерная программа, позволяющая преобразовать текст с бумажного носителя в электронный текстовый файл, который в дальнейшем может обрабатываться человеком в любом текстовом редакторе. Такие программы обычно предлагаются с каждым приобретаемым сканером, но наиболее известными и полифункциональными являются OCR-программы FineReader (компании Abbyy) и CuneiForm (фирмы Cognitive Technologies).

Результат распознавания большинством OCR-программ весьма точен, хотя некоторые трудности в распознавании текста приводят к ошибкам, которые впоследствии приходится исправлять вручную.

Трудности распознавания могут быть вызваны следующими особенностями печатного текста:

- использование шрифта разной гарнитуры и размера,
- использование в тексте нескольких языков,
- размещение текста в несколько колонок,
- включение в текст таблицы рисунков,
- искажения символов (разрывы, слипания букв и т.п.),
- посторонние включения в изображение и т.д.

Названные трудности решаются, если дополнить системы автоматического распознавания текстов возможностью работы машины со смыслом документа, т.е. вывести OCR-программы на наиболее высокий уровень искусственного интеллекта.

В целом точность распознавания OCR-программ на текстах хорошего и среднего качества достигает 99%, что позволяет считать проблему массового ввода печатных текстов в компьютер практически решенной.

Вопросы для обсуждения

1. Что такое автоматический анализ речи?
2. Из каких этапов состоит процесс автоматического анализа речи?
3. Как проводится анализ по фонемного распознавания звуков?
4. Что такое спектограмма? Назовите виды спектограмм
5. Когда и в каких случаях применяется автоматический анализ речи?
6. Назовите компьютерные программы используемые для распознавания автоматического анализа речи?

Рекомендуемая литература

1. Башмаков И.А, Башмаков А.И. Интеллектуальные информационные системы. М.: МГТУ им. Н.Э.Баумана, 2005. С.32—40.

2. Всеволодова А.В. Компьютерная обработка лингвистических данных: учеб. пособие. 2-е изд., испр. М.: Флинта; Наука, 2007. С.47.

3. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: учеб. пособие. М.: Академия, 2004. С.53—55.

4. Сканирование и распознавание: профессиональный подход: материалы курса дистанционного обучения, www.online-academy.ru/scan.htm (дата обращения: 02.02.3012).