

СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ РАСПОЗНАВАНИЯ РЕЧИ

Гельдымурадов С.С.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Петрова Н.Е. – к.филол.н., доцент

В статье рассматриваются основные технологии и роль распознавания речи в современном мире. Также описываются классификации технологий распознавания речи. Делается акцент на использовании спектрального анализа, применимого к звуковой речи, основываясь на особенности звучания некоторых языков.

Информационные технологии являются неотъемлемой частью любой профессиональной деятельности. Развитие технологического процесса позволяет решать всё больше задач, связанных почти с любой деятельностью. Не обошел стороной прогресс и лингвистическую отрасль, которая выполняет всё больше исследований, связанных с информационными технологиями, такие как машинный перевод, синтез человеческой речи, распознавание речи и другие.

По мере развития компьютерных систем становится все более очевидным, что использование этих систем намного расширится, если станет возможным использование человеческой речи при работе непосредственно с компьютером, и в частности станет возможным управление машиной обычным голосом в реальном времени, а также ввод и вывод информации в виде обычной человеческой речи.

Автоматический анализ печатного текста, так же, как и анализ звучащей речи, начинается с его ввода в компьютер. Результат большинством OCR-программ весьма точен, хотя некоторые трудности в распознавании текста приводят к ошибкам, которые впоследствии приходится исправлять вручную [1].

Существующие технологии распознавания речи не имеют пока достаточных возможностей для их широкого использования, но на данном этапе исследований проводится интенсивный поиск возможностей употребления коротких многозначных слов (процедур) для облегчения понимания. Распознавание речи в настоящее время нашло реальное применение в жизни, пожалуй, только в тех случаях, когда используемый словарь сокращен до 10 знаков, например при обработке номеров кредитных карт и прочих кодов доступа в базирующихся на компьютерах системах, обрабатывающих передаваемые по телефону данные [2].

Далее рассмотрим классификацию систем распознавания речи. Различают системы:

- по размеру словаря (ограниченный набор слов, словарь большого размера);
- по зависимости от диктора (дикторозависимые и дикторонезависимые системы);
- по типу речи (слитная или раздельная речь);
- по назначению (системы диктовки, командные системы);
- по используемому алгоритму (нейронные сети, скрытые Марковские модели, динамическое программирование);
- по типу структурной единицы (фразы, слова, фонемы, дифоны, аллофоны);
- по принципу выделения структурных единиц (распознавание по шаблону, выделение лексических элементов) [3].

Для устранения негативного эффекта влияния варьирования голосового тракта на процесс распознавания речи было использовано множество методов. Первым делом рассматривалась характеристика пространства траектории артикуляторных органов, включая гласные, используемые говорящим. Наиболее удачные формы трансформации, использованной для сокращения различий, были впервые представлены Сакоя & Чибо и назывались динамическими искажениями (dynamic time warping). Техника динамического искажения используется для временного вытягивания и сокращения расстояния между искаженным спектральным представлением и шаблоном для говорящего. Использование данной техники дало улучшение точного распознавания (~20-30%). Метод динамического искажения используют практически все коммерчески доступные системы распознавания, показывающие высокую точность сообщения при использовании [4].

Очевидно, что спектральное представление речи позволяет характеризовать особенности голоса человека и способ использования его говорящим. Самый обычный способ моделирования специфических эффектов "модель-источник" – использование фильтров. Речевой аппарат моделируется с использованием источников, вызывающих резонанс, ведущий к пиковым точкам интенсивности звука в соседстве с отдельными частотами, называемыми формантами. При произнесении звуков вибрация голосовых связок является источником возбуждения, и эти короткие импульсы вызывают резонанс между голосовыми связками и губами. Так как язык, челюсть, губы, зубы и альвеолярный аппарат двигаются, размер и место этих резонансов меняются, давая возможность воспроизведения особых параметров звуков.

Что касается уровня артикуляторного контроля, первым уровнем является индивидуальный фонетический сегмент, иначе говоря, – фонема. Во многих естественных языках их примерно 40. Но их набор существенно различается. Поэтому, например, английские гласные могут быть носовыми, даже ненамеренно, в то время как во французском назализация гласных является фонетическим контрастом, и поэтому влияют на значение произносимого. Во французском языке носовая коартикуляция доминирует в гласных и существенно влияет на восприятие фонем и, следовательно, на главный смысл значения. Хотя все говорящие имеют одинаковый голосовой аппарат, использование его разное. Так, например, использование кончика языка или прищелкивание, как в некоторых африканских языках. Ясно, что природа артикуляционных движений имеет сильное влияние на метод воспроизведения речи. Эти ограничения всегда активно используются в практических системах. Говорящие на родном языке избегают этих ограничений или могут активно их использовать во время процесса восприятия [5].

Многие исследования обнаружили, что характеристика слов при введении разбиения на 5 жестких классов фонетических сегментов может быть сокращена до минимума, часто имея единственное в своем роде распознавание. Далее слишком усиливается эффект порядка двух букв и фонетических сегментов с тех пор, как в изучении английских и французских словарей было обнаружено, что более 90% слов имели единственное значение и только 0,5% имели 2 и больше альтернатив. На фонемном уровне было обнаружено, что все слова в английском словаре из 20 тысяч слов имели одно значение из-за беспорядочных фонемных пар. Этот пример помогает показать, что все еще существует ограничивающее влияние на лексическом уровне, которое еще не определено в современных системах распознавания речи. Естественно, что исследования в этой области продолжаются [6].

Несмотря на сложность описания характеристик источников различных ограничений, немаловажную роль играют современные системы влияния, которые представлены всеми возможными вариантами произнесения звуков. Например, система HARPI университета Carnegie-Mellon University является системой, в которой звуковоспроизведение описывается как путь через комплексную сеть. В этом способе ограничения структуры слога, слова и синтаксиса связаны одной структурой. Структура контроля, используемая для поиска, является адаптацией динамичной программной техники. Более сильный подход был предложен моделями использования цепей Маркова. Эти модели использовались как единая структура, где возможности могут быть точно изучены экспериментальным путем.

Чтобы иметь возможность сфокусированного исследования ограничений взаимодействия и интеграции в контексте, необходимо применять обе системы. Те системы, которые описывают ограничения взаимодействия, сфокусированы во многом на воспроизведении знаний, и они относительно слабо контролируются, а системам с математической поддержкой, которые в свою очередь имеют великолепную технику для установления параметров и оптимизации изучения, не хватает использования комплексной структуры данных, необходимых для характеристики ограничений высокого уровня, таких как синтаксис. Оба направления в настоящий момент находятся в процессе развития [7].

В заключение следует сделать акцент на влияние производственных технологий на эти системы. Технология интеграции не является большой проблемой для систем распознавания речи, наоборот, это является архитектурой этих систем, включая способ представления ограничений. Необходимо провести опыты и эксперименты и найти новые способы, которые необходимы для развития технологий распознавания речи и текстов. В таком случае это откроет новые возможности и в профессиональных отраслях науки и техники.

Список использованных источников:

1. Информационные технологии в лингвистике [Электронный ресурс]. – Режим доступа: <https://narfu.ru/university/library/books/1580.pdf>. – Дата доступа 20.12.2019.
2. Распознавание речи [Электронный ресурс]. – Режим доступа: https://ru.wikipedia.org/wiki/Распознавание_речи. – Дата доступа 20.12.2019.
3. Распознавание речи [Электронный ресурс]. – Режим доступа: https://ru.wikipedia.org/wiki/Распознавание_речи. – Дата доступа 20.12.2019.
4. Распознавание речи [Электронный ресурс]. – Режим доступа: <http://chernykh.net/content/view/1116/1200/>. – Дата доступа 20.12.2019.
5. Задача распознавания речи еще не решена [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/post/408017/>. – Дата доступа 20.12.2019.
6. Распознавание речи [Электронный ресурс]. – Режим доступа: https://ru.wikipedia.org/wiki/Распознавание_речи. – Дата доступа 20.12.2019.
7. Распознавание речи [Электронный ресурс]. – Режим доступа: <http://chernykh.net/content/view/1116/1200/>. – Дата доступа 20.12.2019.