

Лекция 13. Технические и психологические аспекты, связанные с голосом и слухом

Цель лекции: в лекции рассмотрены вопросы, касающиеся технических и психологических аспектов, связанных с голосом и слухом. Описана классификация систем распознавания речи. Перечислены преимущества и недостатки голосового взаимодействия с помощью звука. Наибольшее внимание уделено архитектуре распознавания речи и звука.

Вопросы для рассмотрения: Система восприятия человека. Слух. Голос. Звук без голоса. Использование звука в проектировании взаимодействия. Преимущества и недостатки голосового взаимодействия с помощью звука. Преимущества и недостатки неголосового взаимодействия с помощью звука. Технические вопросы, связанные со звуком. Звуковые волны. Машинно-сгенерированный звук. Распознавание речи.

Основные термины: распознавание речи, языковая модель, Септрон, звуковые волны, телетайп, интерфейс безмолвного доступа

Введение

Распознавание речи — автоматический процесс преобразования речевого сигнала в цифровую информацию (например, текстовые данные). Обратной задачей является синтез речи.

Первое устройство для распознавания речи появилось в 1952 году, оно могло распознавать произнесённые человеком цифры.^[1] В 1962 году на ярмарке компьютерных технологий в Нью-Йорке было представлено устройство IBM Shoebox.

В 1963 году в США были презентованы разработанные инженерами корпорации «Сперри» миниатюрные распознающие устройства с волоконно-оптическим запоминающим устройством под названием «Септрон» (Sceptron, но произносится [ˈseptraːn] без «к»),^[2] выполняющие ту или иную последовательность действий на произнесённые человеком-оператором определённые фразы. «Септроны» годились для применения в сфере фиксированной (проводной) связи для автоматизации набора номеров голосом и автоматической записи надиктовываемого текста телетайпом, могли применяться в военной сфере (для голосового управления сложными образцами военной техники), авиации (для создания «умной авионики», реагирующей на команды пилота и членов экипажа), автоматизированных системах управления и др.^{[2][3][4]} В 1983 году был презентован интерактивный комплекс «умной авионики» для ударных вертолётов «Апач», распознающий команды и запросы пилота, преобразующий их в сигналы управления на бортовое оборудование и односложно отвечающий ему голосом относительно возможности реализации поставленной им задачи^[5].

Коммерческие программы по распознаванию речи появились в начале девяностых годов. Обычно их используют люди, которые из-за травмы руки не в состоянии набирать большое количество текста. Эти программы (например, Dragon NaturallySpeaking (англ.)русск., VoiceNavigator (англ.)русск.) переводят голос пользователя в текст, таким образом, разгружая его руки. Надёжность перевода у таких программ не очень высока, но с годами она постепенно улучшается.

Увеличение вычислительных мощностей мобильных устройств позволило и для них создать программы с функцией распознавания речи. Среди таких программ стоит отметить

приложение Microsoft Voice Command, которое позволяет работать со многими приложениями при помощи голоса. Например, можно включить воспроизведение музыки в плеере или создать новый документ.

Все большую популярность применение распознавания речи находит в различных сферах бизнеса, например, врач в поликлинике может проговаривать диагнозы, которые тут же будут внесены в электронную карточку. Или другой пример. Наверняка каждый хоть раз в жизни мечтал с помощью голоса выключить свет или открыть окно. В последнее время в телефонных интерактивных приложениях все чаще стали использоваться системы автоматического распознавания и синтеза речи. В этом случае общение с голосовым порталом становится более естественным, так как выбор в нём может быть осуществлен не только с помощью тонового набора, но и с помощью голосовых команд. При этом системы распознавания являются независимыми от дикторов, то есть распознают голос любого человека.

Следующим шагом технологий распознавания речи можно считать развитие так называемых интерфейсов безмолвного доступа (silent speech interfaces, SSI). Эти системы обработки речи базируются на получении и обработке речевых сигналов на ранней стадии артикулирования. Данный этап развития распознавания речи вызван двумя существенными недостатками современных систем распознавания: чрезмерная чувствительность к шумам, а также необходимость четкой и ясной речи при обращении к системе распознавания. Подход, основанный на SSI, заключается в том, чтобы использовать новые сенсоры, не подверженные влиянию шумов в качестве дополнения к обработанным акустическим сигналам.

Классификация систем распознавания речи[\[править\]](#) [\[править код\]](#)

Системы распознавания речи классифицируются:^[6]

- 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).

Для систем автоматического распознавания речи, помехозащищённость обеспечивается, прежде всего, использованием двух механизмов:^[7]

- Использование нескольких, параллельно работающих, способов выделения одних и тех же элементов речевого сигнала на базе анализа акустического сигнала;
- Параллельное независимое использование сегментного (фонемного) и целостного восприятия слов в потоке речи.

Архитектура систем распознавания[\[править\]](#) [\[править код\]](#)

Одна из архитектур систем автоматической обработки речи, основанной на статистических данных, может быть следующей.^{[12][13]}

- Модуль шумоочистки и отделение полезного сигнала.
- Акустическая модель — позволяет оценить распознавание речевого сегмента с точки зрения схожести на звуковом уровне. Для каждого звука изначально строится сложная

статистическая модель, которая описывает произнесение этого звука в речи.

- Языковая модель — позволяют определить наиболее вероятные последовательности слов. Сложность построения языковой модели во многом зависит от конкретного языка. Так, для английского языка, достаточно использовать статистические модели (так называемые N-граммы). Для высокофлективных языков (языков, в которых существует много форм одного и того же слова), к которым относится и русский, языковые модели, построенные только с использованием статистики, уже не дают такого эффекта — слишком много нужно данных, чтобы достоверно оценить статистические связи между словами. Поэтому применяют гибридные языковые модели, использующие правила

русского языка, информацию о части речи и форме слова и классическую статистическую модель.

- Декодер — программный компонент системы распознавания, который совмещает данные, получаемые в ходе распознавания от акустических и языковых моделей, и на основании их объединения, определяет наиболее вероятную последовательность слов, которая и является конечным результатом распознавания слитной речи.

Этапы распознавания^[12]

1. Обработка речи начинается с оценки качества речевого сигнала. На этом этапе определяется уровень помех и искажений.
2. Результат оценки поступает в модуль акустической адаптации, который управляет модулем расчета параметров речи, необходимых для распознавания.
3. В сигнале выделяются участки, содержащие речь, и происходит оценка параметров речи. Происходит выделение фонетических и просодических вероятностных характеристик для синтаксического, семантического и прагматического анализа. (Оценка информации о части речи, форме слова и статистические связи между словами.)
4. Далее параметры речи поступают в основной блок системы распознавания — декодер. Это компонент, который сопоставляет входной речевой поток с информацией, хранящейся в акустических и языковых моделях, и определяет наиболее вероятную последовательность слов, которая и является конечным результатом распознавания.

Автоматическое распознавание речи на естественном языке является самым актуальным направлением развития искусственного интеллекта

Результаты в этом направлении будут способствовать установлению эффективного взаимодействия между человеком и компьютером.

Внедрение речевых технологий в любую сферу производства осуществляется исследовательскими центрами и научными центрами.

Исследованием распознавания речи занимаются специалисты нескольких научных областей уже более пятидесяти лет.

Он широко используется в любых системах голосового управления.

Звуковой ввод

природность, оперативность, смысловая точность введения, освобождение рук и глаз пользователя, экстремальная

Автоматическое распознавание речи-текст или соответствующий акустический речевой сигнал компьютера.

Более сложная задача заключается в автоматическом понимании речи: он включает в себя семантический анализ признанного текста, включая распознавание речи. Системы распознавания речи можно классифицировать по следующим признакам::

- зависимость от диктора;
- объем словаря;

- характер опознаваемого речевого потока.

Система, зависящая от диктора, создается отдельным пользователем.

С помощью функции распознавания речи можно вводить данные, указывать действия или команды и выполнять задачи.

Важные API-интерфейсы: [Windows.Media.SpeechRecognition](#)

Для распознавания речи используется специальная среда выполнения, API распознавания для программирования среды выполнения, готовые грамматики для диктовки и веб-поиска, а также системный пользовательский интерфейс по умолчанию, который помогает пользователям обнаруживать и использовать функции распознавания речи.

Настройка распознавания речи

Для поддержки распознавания речи в приложении пользователь должен подключить и включить микрофон на устройстве, а также принять политику конфиденциальности Майкрософт, предоставив приложению разрешение на его использование.

Чтобы автоматически запрашивать у пользователя диалоговое окно, запрашивающее разрешение на доступ и использование канала звука микрофона (пример из [примера распознавания речи и синтеза речи](#)), просто установите [функцию микрофона](#) в [манифесте пакета приложения](#). Дополнительные сведения см. в разделе [объявления возможностей приложений](#).

Если пользователь нажмет кнопку Да для предоставления доступа к микрофону, приложение будет добавлено в список утвержденных приложений на странице Параметры > Конфиденциальность — > микрофон. Однако, так как пользователь может отключить этот параметр в любое время, прежде чем пытаться использовать его, убедитесь, что приложение имеет доступ к микрофону.

Контрольные вопросы

1. Что можно отнести к технологическим и психологическим аспектам распознавания речи, связанным с голосом и слухом?
2. Какие существуют основные аспекты оценивания проекта?
3. Какова архитектура распознавания речи и слуха?
4. Какую роль в распознавании речи играют звуковые волны?
5. В каком году появилось первое устройство распознавания речи?

Литература:

1. Купер А. т.б. Об интерфейсе: основы проектирования взаимодействия. 4-е изд. — СПб: Питер, 2018. — 720 б.
2. Норман Д. Дизайн привычных вещей. — М: Манн, Иванов и Фербер, 2019. — 384 б.
3. The Encyclopedia of Human-Computer Interaction, 2nd Ed. [электрондық басылым] — URL: <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed> (соңғы алынған 8/1/2020)
4. Sharp H. Interaction Design: Beyond Human-Computer Interaction. 5th Ed. — Wiley, 2019. — 656 р.

5. Shneiderman B. et.al. Designing the user interface: strategies for effective human-computer interaction. 6th Ed. – Pearson, 2016. – 616 p.
6. Тидвелл Д. т.б. Разработка пользовательских интерфейсов. 2-е изд. – М: Питер, 2011 – 480 б.
7. Круг С. т.б. Веб-дизайн или «не заставляйте меня думать». – СПб.: Символ Плюс, 2008. – 224 .
8. Нильсен Я. т.б. Веб дизайн. СПб: Символ Плюс, 2006. – 512 б.
9. Уильямс Р. т.б. Не дизайнерская книга о дизайне. – СПб: Весь, 2004. – 128 б.
10. <http://appscamp.io/> – Онлайн-курс который дает начальное понимание разработки на HTML и мобильных платформах.
11. <http://phonegap.com/book/> – Список книг по разработке HTML и мобильных приложений с помощью фреймворкаPhoneGap.
12. <http://creator.ionic.io/> – HTML фреймворк пользовательского интерфейса для мобильных приложений.