Урок 20. Распознавание и синтез речи

9 класс · Урок 20 из 34 · 45 минут

Вся программа 9 класса

Распознавание и синтез речи соединяются в голосовом помощнике. Проследим, в каком звене может появиться ошибка.

Содержание урока

- 01. Цель урока
- 02. Ход урока
- 03. Этап 1. Введение: Команда с числом · 7 мин
- 04. Этап 2. Путь команды и ответа · 12 мин
- 05. Этап 3. Расследование и контрольный повтор · 20 мин
- 06. Этап 4. Итоги урока · 6 мин

01. Цель урока

Объяснять, как распознавание и синтез речи работают в голосовом помощнике. Определять, на каком шаге возникла ошибка и как её проверить.

02. Ход урока

Этап	Мин	Учитель	Ученики
Этап 1. Введение: Команда с числом	7	Предлагает несколько причин сбоя.	Записывают, что надо посмотреть кроме звука ответа.
Этап 2. Путь команды и ответа	12	Объясняет вход, распознавание, действие и озвучку.	Сопоставляют звенья с доступными следами.
Этап 3. Расследование и контрольный повтор	20	Раздаёт журналы, затем проводит короткую пробу произнесения.	Определяют шаг ошибки и подбирают контроль.
Этап 4. Итоги урока	6	Подводит итог и сверяет ответы.	Письменно отвечают и сверяют решение.

Этап	Мин	Учитель	Ученики
Итого	45		

03. Этап 1. Введение: Команда с числом 7 мин

ГОВОРИМ УЧЕНИКАМ

В макете планетария светом управляет голосовой помощник. Ему сказали включить третий режим, а на панели загорелся индикатор пятого. В комнате было шумно. Мог ли помощник правильно распознать слова и всё равно включить не тот режим?

Распознавание и синтез речи отвечают за разные стороны разговора. Сначала можно проверить, какие слова записала система. Если в тексте уже стоит «пятый», искать нужно на входе. Если стоит «третий», смотрим, какое действие программа выбрала после распознавания.

04. Этап 2. Путь команды и ответа 12 мин

ГОВОРИМ УЧЕНИКАМ

Микрофон превращает звук в сигнал. Программа выделяет короткие участки и оценивает, какие частоты в них слышны. Такое представление помогает модели заметить, как звук меняется во времени. На входе могут быть и голос, и посторонний шум. Если часть слога перекрыта, модели труднее выбрать подходящее слово. Модели распознавания часто учитывают окружающие слова, чтобы предпочесть более вероятную запись.

После распознавания остаётся текст. Помощник должен понять, какое действие выполнить, или составить ответ. Например, в фразе «включи тёплый свет на втором столе» ему надо выделить действие, режим и номер стола. Даже точная запись ничего не включает сама по себе. Если программа сопоставит слово «тёплый» с настройкой холодного света, ошибка возникнет на шаге выбора действия.

Синтезатор, наоборот, получает текст и создаёт звук. Ему нужно решить, как прочитать сокращения и числа, где поставить паузу, куда падает ударение. Слово «замок» можно прочитать по-разному. В объявлении для выставки о средневековой крепости это «за́мок», а в инструкции к двери — «замо́к». Синтезатор сначала преобразует запись в произносимую фразу, выбирает звуки, ударения и паузы, а затем строит звуковой сигнал. Контекст помогает выбрать чтение, но результат лучше послушать.

Эти технологии встречаются отдельно. Распознавание полезно для субтитров или поиска по записи. Синтез речи позволяет прослушать инструкцию, например когда читать её с экрана неудобно. В помощнике они соединяются с программой, которая отвечает или выполняет действие.

Что можно проверить	К какому шагу относится
Сигнал с микрофона и расшифровка	Распознавание: звук стал текстом.
Расшифровка и выбранная команда	Интерпретация: текст стал решением.

Что можно проверить	К какому шагу относится
Текст ответа и звучание	Синтез: текст стал речью.

Реальный случай. Как Whisper представляет звук

В сентябре 2022 года разработчики Whisper описали, как модель получает спектрограмму, то есть картину изменений частот во времени. На таком представлении система учится распознавать речь. Это помогает понять, зачем для диагностики сравнивают исходный звук и расшифровку.

05. Этап 3. Расследование и контрольный повтор 20 мин

ГОВОРИМ УЧЕНИКАМ

В журнале видно несколько следов одной команды. Чтобы найти причину, начните с первой пары записей, в которой появилось различие. После исправления попробуйте поменять одно условие, иначе будет трудно понять, что сработало.

Лаборатория голосовой команды

Команда в учебном журнале — «Включи тёплый свет на втором столе». Четыре записи показывают независимые попытки.

Попытка	Расшифровка	Действие	Текст ответа и произнесённое
А	Включи тёплый свет на втором столе	Тёплый, стол 2	Написано и произнесено: «Свет включён на втором столе».
Б	Включи тёплый свет на третьем столе	Тёплый, стол 3	Написано и произнесено: «Свет включён на третьем столе».
В	Включи тёплый свет на втором столе	Холодный, стол 2	Написано и произнесено: «Холодный свет включён на втором столе».
Г	Включи тёплый свет на втором столе	Тёплый, стол 2	Написано «Свет включён на втором столе», произнесено «Свет включён на втором сто́ле».

1. Для каждой записи определите, что сработало, и укажите первый шаг, на котором появилась ошибка.
2. Придумайте контрольный повтор для одной строки. Скажите, что измените и какой след посмотрите.
3. Проверьте другую задачу. Озвучьте запись «На северную лестницу идут участники мастерской ждут экскурсанты у гардероба». Сосед на слух распределяет две группы по местам. Уточните текст и паузы так, чтобы «участники мастерской» остались одной группой, и повторите пробу.

Для учителя

А — ошибки нет, в Б первым меняется расшифровка, в В — действие, в Г — произношение. Контроль для Б: та же фраза в тишине, сравнить расшифровку. Для В: повтор с другим режимом света и проверка выбранной команды. Для Г: сравнить озвучку текста, затем попробовать явное ударение там, где сервис его поддерживает. Для объявления подойдут две самостоятельные фразы: «Участники мастерской, идите на северную лестницу. Экскурсанты, ждите у гардероба».

Для показа в SpeechKit Playground заранее запишите голос учителя с одной нейтральной фразой в тишине и на фоне хлопков. Загрузите два файла и сравните расшифровки. Затем воспроизведите два варианта текста объявления. Используйте голос учителя и учебный текст. В офлайн-версии учитель озвучивает варианты, а класс отделяет модельное чтение по журналу от собственной пробы.

06. Этап 4. Итоги урока 6 мин

ГОВОРИМ УЧЕНИКАМ

Важные команды полезно подтверждать до действия. Для объявления, которое нельзя перечитать, лучше заранее проверить на слух, кого касается каждое указание и где слышна граница фразы.

Главное за урок

- Распознавание превращает речь в текст, синтез превращает текст в звук.
- Точный текст ещё не гарантирует правильно выбранное действие.
- Следы каждого шага позволяют найти причину сбоя.

Итоговые вопросы

1. Текст расшифровки и действие верны, но фамилия в ответе звучит неверно. Какой шаг проверить первым?
2. Какая проверка поможет, если устройство понимает команду в тишине и путает её на перемене?

Для учителя

В первом случае проверяем синтез и выбранное произношение фамилии. Во втором повторяем одну и ту же фразу при разных условиях шума и сравниваем расшифровки.

Источники

1. OpenAI. Introducing Whisper · 21.09.2022 · Материал старше 2024 года. «converted into a log-Mel spectrogram» Источник
2. Яндекс. Проверяем распознавание и синтез речи в SpeechKit Playground · 19.06.2025. «загрузите файл или введите текст и послушайте результат прямо в браузере» Источник