**Лабораторная работа 16**

**Уровень 1**

**Стр87, ответы на вопросы + практическое задание**

1.Какие задачи речевой обработки являются наиболее актуальными?

**Распознавание речи (ASR):** Преобразование устной речи в текстовый формат.

**Генерация речи (TTS):** Преобразование текстового формата в устную речь.

**Анализ настроений**: Определение настроения или эмоциональной окраски текста или речи.

**Перевод речи:** Перевод устной речи с одного языка на другой.

**Диалоговые системы**: Разработка систем, которые могут вести естественные и информативные диалоги с людьми.

**Обработка естественного языка (NLP):** Задачи, связанные с пониманием и генерацией человеческого языка, такие как извлечение информации, обобщение и машинный перевод.

**Распознавание голоса:** Идентификация говорящего по его голосу.

**Поиск по речи**: Поиск устной речи или аудиозаписей по определенным словам или фразам.

**Голосовое управление**: Разработка систем, которые позволяют пользователям управлять устройствами или приложениями с помощью голосовых команд.

2. В каких областях решение этих задач находит свое применение?

**Обслуживание клиентов: Виртуальные помощники, системы** автоматического ответа и чат-боты.

**Здравоохранение:** Транскрипция медицинских записей, анализ настроений пациентов и диагностика заболеваний.

**Образование:** Персонализированное обучение, помощь в изучении языков и автоматическая оценка.

**Финансы**: Анализ финансовых данных, обнаружение мошенничества и обработка транзакций голосом.

**Медиа и развлечения**: Создание субтитров, поиск по видео и персонализированные рекомендации.

**Производство**: Управление складом, контроль качества и техническое обслуживание с помощью голосовых команд.

**Транспорт**: Голосовое управление навигационными системами, диагностика транспортных средств и управление флотом.

**Умный дом**: Управление устройствами умного дома, такими как освещение, отопление и безопасность, с помощью голосовых команд.

**Безопасность**: Распознавание голоса для доступа к зданиям и системам, а также анализ настроений для выявления угроз.

3. Как происходит процесс верификации говорящего?

**Верификация говорящего** - это процесс определения того, является ли конкретный человек тем, за кого он себя выдает, на основе его голоса.

1. **Регистрация**: Говорящий записывает образец своего голоса, который используется для создания голосового профиля.

2. **Извлечение характеристик**: Из голосового образца извлекаются уникальные акустические характеристики, такие как частота основного тона и особенности произношения.

3. **Создание модели**: Эти характеристики используются для создания математической модели голосового профиля говорящего.

4. **Верификация**: Когда говорящий пытается получить доступ к системе или выполнить транзакцию, его голос сравнивается с сохраненной моделью.

5. **Оценка сходства**: Алгоритм оценивает степень сходства между голосом говорящего и сохраненной моделью и принимает решение о том, является ли говорящий тем, за кого он себя выдает.

Существуют различные методы верификации говорящего, в том числе:

**Подход на основе текста**: Говорящего просят произнести определенный текст, а затем его голос сравнивается с сохраненной моделью.

**Подход на основе фразы-пароля**: Говорящего просят произнести секретную фразу-пароль, которую знает только он сам.

**Подход, не зависящий от текста**: Говорящему не нужно произносить никаких конкретных слов или фраз. Вместо этого сравниваются акустические характеристики его голоса.

4. В чем заключается отличие верификации от идентификации говорящего по голосу?

**Верификация** говорящего определяет, **является ли конкретный человек тем, за кого он себя выдает**, на основе его голоса. Это процесс **проверки личности.**

**Идентификация** говорящего определяет, **кто говорит, на основе его голоса**. Это процесс **определения личности**.

Верификация более строгий процесс, чем идентификация. При верификации цель состоит в том, чтобы убедиться, что говорящий действительно является тем человеком, за которого он себя выдает. При идентификации цель состоит в том, чтобы определить личность говорящего, даже если он не заявляет о своей личности.

5. Чем обусловлена актуальность решения такой задачи как преобразование речи в текст?

**Удобство**: STT позволяет пользователям взаимодействовать с устройствами и приложениями с помощью голосовых команд. Это особенно удобно для людей, у которых заняты руки или которые не могут печатать.

**Доступность**: STT делает контент более доступным для людей с нарушениями зрения или другими ограниченными возможностями.

**Повышение эффективности**: STT может помочь повысить эффективность в различных областях, таких как обслуживание клиентов и транскрипция.

**Расширение возможностей искусственного интеллекта**: STT является ключевой технологией для многих приложений искусственного интеллекта, таких как виртуальные помощники и чат-боты.

6. В чем заключается отличие первых систем Speech-to-Text от тех, которые разрабатываются в настоящее время?

**Первые системы** распознавания речи (STT) были разработаны в 1950-х годах и **использовали простые акустические модели** и небольшое количество обучающих данных. Они могли распознавать только ограниченный набор слов и часто допускали ошибки.

**Современные системы** STT используют более сложные модели глубокого обучения и огромные наборы обучающих данных. Они **способны распознавать широкий спектр слов** и языков с высокой точностью. Кроме того, современные системы STT более устойчивы к шуму и другим помехам.

7. За счет чего системы распознавания речи стали допускать меньше ошибок?

Системы распознавания речи стали допускать меньше ошибок за счет следующих факторов:

**Улучшенные акустические модели**: Современные системы STT используют более сложные акустические модели, которые лучше улавливают особенности человеческой речи.

**Большие наборы обучающих данных**: Современные системы STT обучаются на огромных наборах данных, которые содержат миллионы часов речи. Это позволяет моделям изучать тонкие нюансы человеческой речи.

**Более мощные алгоритмы**: Современные системы STT используют более мощные алгоритмы глубокого обучения, которые могут обрабатывать большие объемы данных и распознавать сложные речевые паттерны.

**Использование дополнительных данных**: Некоторые системы STT используют дополнительные данные, такие как контекст и знания о мире, чтобы улучшить распознавание речи.

8. Перечислите этапы работы систем распознавания говорящего и преобразования речи в текст.

Этапы работы систем распознавания говорящего и преобразования речи в текст:

**Распознавание говорящего:**

1)Запись голосового образца.

2)Извлечение акустических характеристик из голосового образца.

3)Сравнение акустических характеристик с сохраненными моделями голоса.

4)Определение того, является ли говорящий тем, за кого он себя выдает.

**Преобразование речи в текст:**

1)Запись речевого сигнала.

2)Преобразование речевого сигнала в последовательность акустических признаков.

3)Использование акустической модели для распознавания последовательности фонем.

4)Использование языковой модели для определения наиболее вероятной последовательности слов.

5)Вывод текстовой транскрипции.

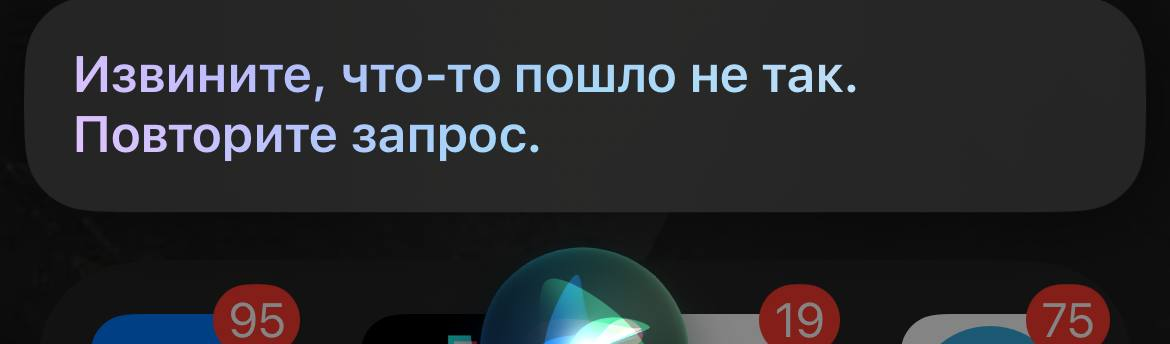
**Практическое задание**

Оцените качество работы систем распознавания речи на русском и английском языках. Проверьте, ухудшится ли результат, если на вход системы STT сообщить: несвязный речевой сигнал (случайные наборы слов), несогласованные друг с другом слова (например, по роду) или речевой сигнал с параллельной речью нескольких людей (или посторонним шумом). Можете использовать системы распознавания речи, внедренные в онлайн переводчики (Яндекс Переводчик, Google Translation) или в голосовые помощники (Google Assistant, Яндекс Алиса, Siri, Alexa)

У меня айфон, поэтому я буду использовать Siri

Сири первый голосовой помощник, внедренный в телефон. Эпл с каждым годом улучшает качество, поэтому речь как на русском, так и на английском она распознает без каких-либо ошибок, выполняя все запросы, которые я от нее требую

При несвязном речевом сигнале она говорит следующее:



Когда произношу несогласованные друг с другом слова Сири все равно понимает, что я от нее требую  
Также при посторонних шумах, голосовой помощник без труда распознает мою речь, видимо концентрируясь только на мой голос

Стр117-118

Ответы на вопросы

1. Что из себя представляют системы Text-To-Speech и что является их основной задачей?

**Системы Text-To-Speech (TTS)** - это системы, которые преобразуют письменный текст в устную речь. Их основная задача - создать синтетическую речь, которая звучит как естественная человеческая речь.

2. По каким критериям оценивают сгенерированную речь?

**Естественность**: Насколько синтетическая речь звучит как естественная человеческая речь.

**Разборчивость**: Насколько легко понять, что говорит синтетическая речь.

**Выразительность**: Насколько хорошо синтетическая речь передает эмоции и интонации человеческой речи.

3. Какой способ синтеза речи подразумевает использование записанных фрагментов человеческой речи? Опишите принцип его работы.

конкатенативный синтез.

**1)Записывается большое количество** фрагментов человеческой речи, охватывающих все возможные звуки и комбинации звуков.

2)**Эти фрагменты хранятся в базе данных.**

3)Когда система TTS получает текстовый ввод, она **разбивает текст на отдельные звуки или слоги.**

4)Затем система ищет соответствующие фрагменты речи в базе данных и соединяет их вместе, чтобы сформировать синтетическую речь.

4. Какие два способа синтеза речи основываются на акустических свойствах речи? В чем их отличие?

формантный синтез и синтез на основе синусоидальных моделей.

**Формантный синтез моделирует резонансные частоты** (форманты) человеческого вокального тракта. Он создает синтетическую речь путем генерации серии резонансных фильтров, которые накладываются друг на друга.

**Синтез на основе синусоидальных моделей** моделирует акустический сигнал речи как сумму синусоидальных волн. Он создает синтетическую речь путем генерации серии синусоидальных волн с соответствующими частотами и амплитудами.

Основное отличие между этими двумя способами заключается в том, что **формантный синтез моделирует резонансные свойства вокального тракта, а синтез на основе синусоидальных моделей моделирует акустический сигнал непосредственно.**

5. В чем заключается особенность артикуляционного способа синтеза речи?

он **моделирует работу человеческого речевого аппарата**. Он создает **синтетическую речь путем управления виртуальным вокальным** трактом, который имитирует движение губ, языка и других артикуляторов.

6. Перечислите основные сложности для синтеза речи каждым способом. Какие существуют пути их решения?

**1)Конкатенативный синтез:**

**Сложность**: Обеспечение плавных переходов между фрагментами речи.

**Пути решения**: Использование техник сшивания и перекрытия.

**2)Формантный синтез:**

**Сложность**: Точное моделирование резонансных свойств вокального тракта.

**Пути решения**: Использование сложных акустических моделей и техник адаптации.

**3)Синтез на основе синусоидальных моделей:**

**Сложность:** Синтез высококачественного акустического сигнала.

**Пути решения**: Использование сложных алгоритмов и больших наборов данных.

**4)Артикуляционный синтез:**

**Сложность**: Точное моделирование работы человеческого речевого аппарата.

**Пути решения**: Использование сложных биомеханических моделей и техник обучения с подкреплением.