**вопросы:**

1. К наиболее важным задачам речевой обработки и анализа можно отнести распознавание говорящего по голосу и преобразование речи в текст.

2. Решение этих задач позволяет реализовать полученные знания в самых разных направлениях: автоматическое создание транскрипции, синхронное преобразование живой речи в текстовый формат, управление голосовыми помощниками, выявление особенностей речи говорящего, выделение наиболее значимых слов в речи, идентификация и верификация личности, создание и разработка технологий для людей с ограниченными возможностями, создание диалоговых систем и чат-ботов.

3. Под верификацией понимают определение системой говорящего как «своего» или «чужого». В качестве примера верификации можно рассмотреть следующую ситуацию: говорящий производит речевой сигнал – сообщает данные о своей личности (инициалы или специальный код), а система автоматического распознавания индивидуальных особенностей голоса и речи должна подтвердить личность говорящего, исходя из эталонной модели, т.е. имеющейся записи голоса говорящего

4. Верификация говорящего по голосу предполагает подтверждение личности человека на основе акустических характеристик его голоса. Это может использоваться, например, для аутентификации при доступе к информационным ресурсам.

С другой стороны, идентификация говорящего по голосу означает определение личности человека на основе его акустических особенностей.

5. Данная технология активно применяется: при автоматическом создании субтитров; при разработке приложений и технических средств для людей с ограниченными возможностями, у которых возникают сложности при

наборе текста вручную; для распознавания команд голосовыми помощниками (Google Assistant, Яндекс Алиса, Siri, Alexa).

6. Первые образцы систем STT работали исключительно на основе эталонных моделей, при этом речь членилась системой на отдельные крупные блоки, после чего происходил поиск наиболее близкого акустического сигнала из уже записанных в базе данных, что приводило к большому количеству ошибок.

В настоящее время, благодаря использованию баз данных коллокаций (устойчивых сочетаний) и N-gramm (последовательностей из N символов), системы STT способны предугадывать целые фразы, исходя из правильно распознанного фрагмента речи.

7. Немаловажную роль в сокращении количества ошибок при распознавании играет и внедрение машинного обучения, которое позволяет системам STT постоянно совершенствоваться. +н-граммы, коллокации и вот это все статистическое

8. Можно выделить три этапа преобразования речи в текст:

● **Анализ сигнала**. Полученный на вход речевой сигнал очищается от

лишних шумов и делится на минимальные фрагменты.

● **Распознавание сигнала**. Фрагменты проходят через эталонную модель, определяющую, какие звуки были произнесены; после чего система пытается объединить их в значимые единицы.

● **Преобразование в текст**. При помощи баз данных и модели языка

происходит подбор нераспознанных единиц, исходя из контекста, и

преобразование полученных результатов в графический вид.

Связный текст на русском и английском распознается хорошо, как я думаю, в силу факторов из 7 вопроса

Вчера кот летел на розовом велосипеде по облакам – сначала был терракот на облаках, английский также распознался после некоторых заминок

С фоновым шумом (треки разные) работает хорошо, биты не мешают (минусов нет у песен)

При несогласованном наборе слов Яндекс сработал после некоторых задержек, скорее всего из-за невозможности n-граммного предсказывания