Стр.87

1.Наиболее актуальными задачами речевой обработки являются распознавание речи, синтез речи, анализ тональности, извлечение информации из аудиофайлов, идентификация говорящего и многие другие.

2.Решение этих задач находит применение в различных областях, таких как технологии голосового управления устройствами, системы распознавания речи для людей с ограниченными возможностями, системы мониторинга и анализа звучаний в медицине и безопасности, системы автоматизации и обработки голосовых уведомлений и т.д.

3.Процесс верификации говорящего основан на сравнении биометрических характеристик голоса человека с данными, зарегистрированными заранее в базе данных. Это позволяет проверить, является ли говорящий именно тем человеком, за которого себя выдаёт.

4.Отличие верификации от идентификации говорящего по голосу заключается в том, что при верификации проверяется подлинность голоса конкретного человека, который уже зарегистрирован в системе, в то время как при идентификации осуществляется поиск совпадений среди всех возможных голосов и определение личности говорящего.

5.Актуальность преобразования речи в текст обусловлена несколькими факторами. Во-первых, это делает информацию, передаваемую устно, доступной для людей с нарушениями слуха. Во-вторых, это упрощает процесс записи и хранения информации, так как текст может быть легко оцифрован и архивирован. Кроме того, преобразование речи в текст позволяет использовать автоматизацию и искусственный интеллект для обработки и анализа больших объемов аудиозаписей, что может быть полезно в различных областях, от медицины до маркетинга.

6.Отличие первых систем Speech-to-Text от современных заключается в их точности, скорости работы, возможности работы с различными диалектами и акцентами, а также способности адаптироваться к сленгу, профессиональной терминологии и другим специализированным жаргонам. Современные системы также часто используют нейронные сети и глубокое обучение для улучшения качества распознавания речи.

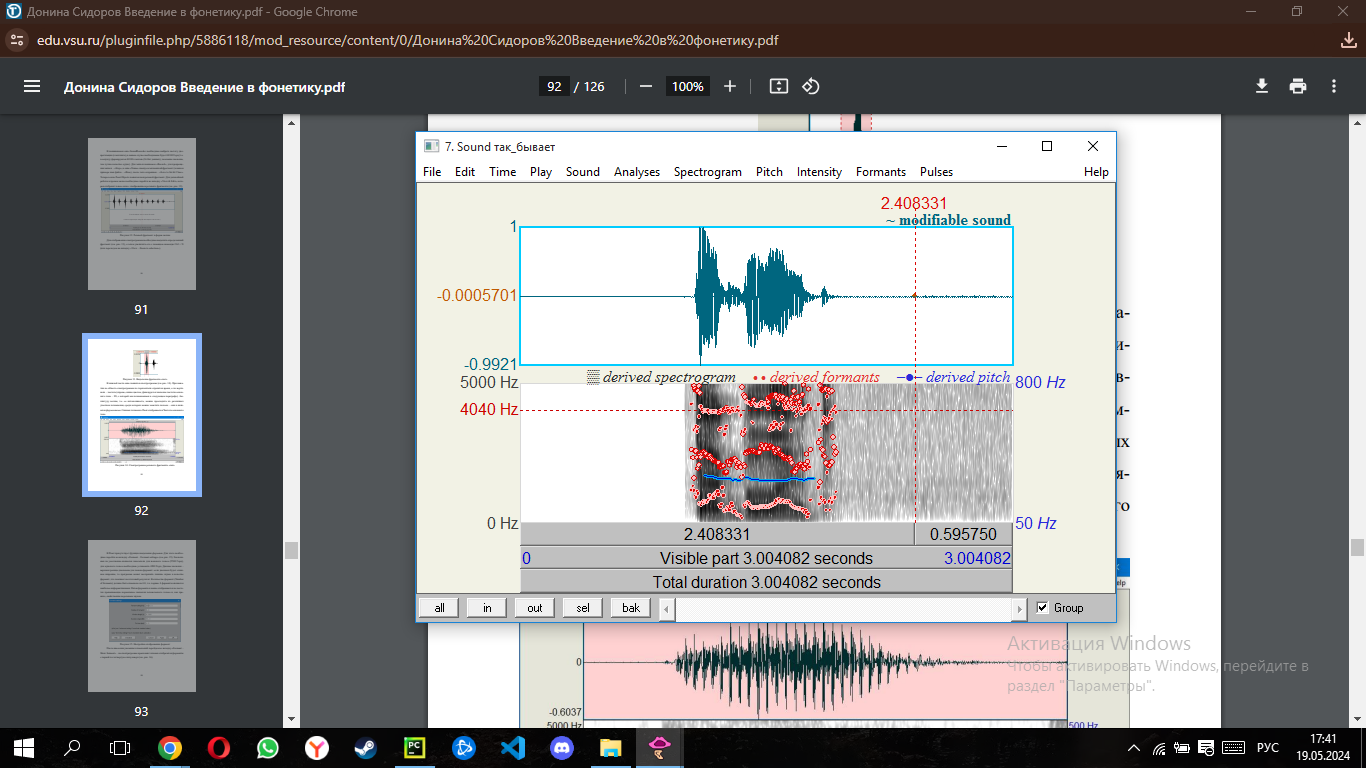
7. Системы распознавания речи стали допускать меньше ошибок за счет использования более сложных алгоритмов и моделей машинного обучения, увеличения объема и разнообразия данных для обучения, а также развития технологий передачи и обработки звука. Также улучшение качества микрофонов и акустической техники позволяет системам точнее и надежнее записывать речь для последующего преобразования в текст.

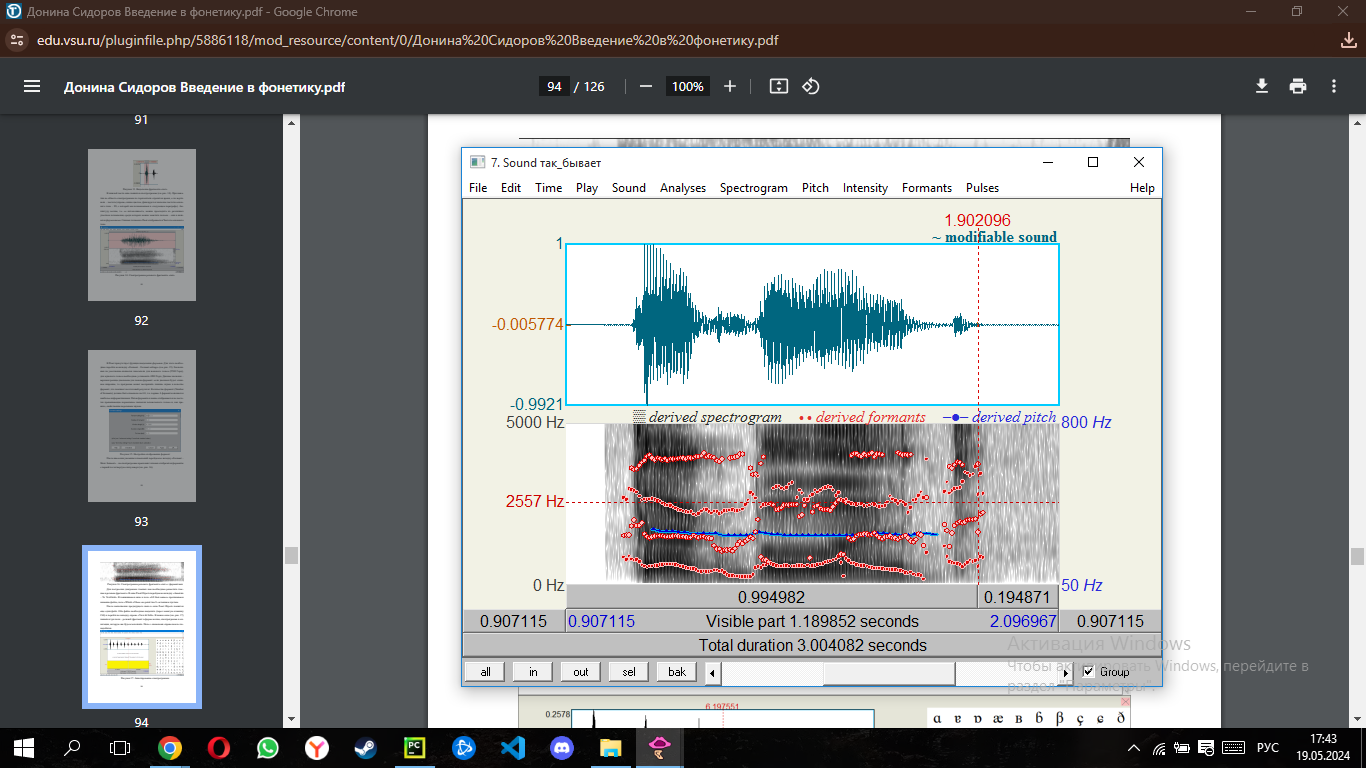
8. Захват аудиосигнала - система записывает аудиосигнал, который содержит речь говорящего. Предварительная обработка аудиосигнала - включает в себя фильтрацию шума, усиление сигнала, преобразование аналогового сигнала в цифровой формат. Извлечение признаков - процесс выделения ключевых особенностей аудиосигнала, которые будут использоваться для распознавания речи. Моделирование речевых единиц - создание моделей для каждого звука, фонемы или слова, чтобы система могла сопоставить их с признаками извлеченных из аудиосигнала. Распознавание речи - процесс сопоставления извлеченных признаков с заранее созданными моделями речевых единиц для определения текста, произнесенного говорящим. Преобразование речи в текст - окончательный этап, в котором текстовый результат распознавания речи выводится на экран или сохраняется в файле. Корректировка результатов и улучшение качества распознавания - включает в себя проверку и исправление ошибок распознавания, а также обновление моделей речевых единиц на основе обратной связи для улучшения точности системы.

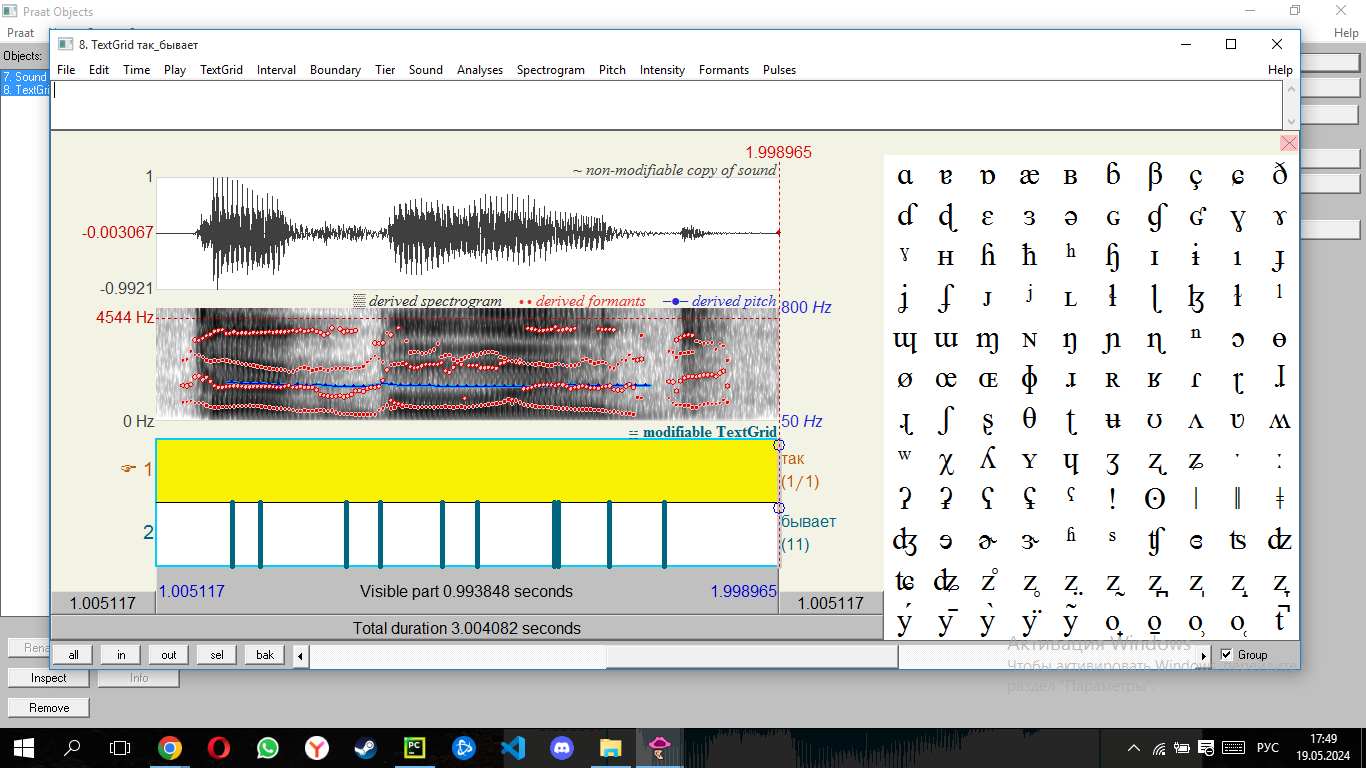
Практическое задание:

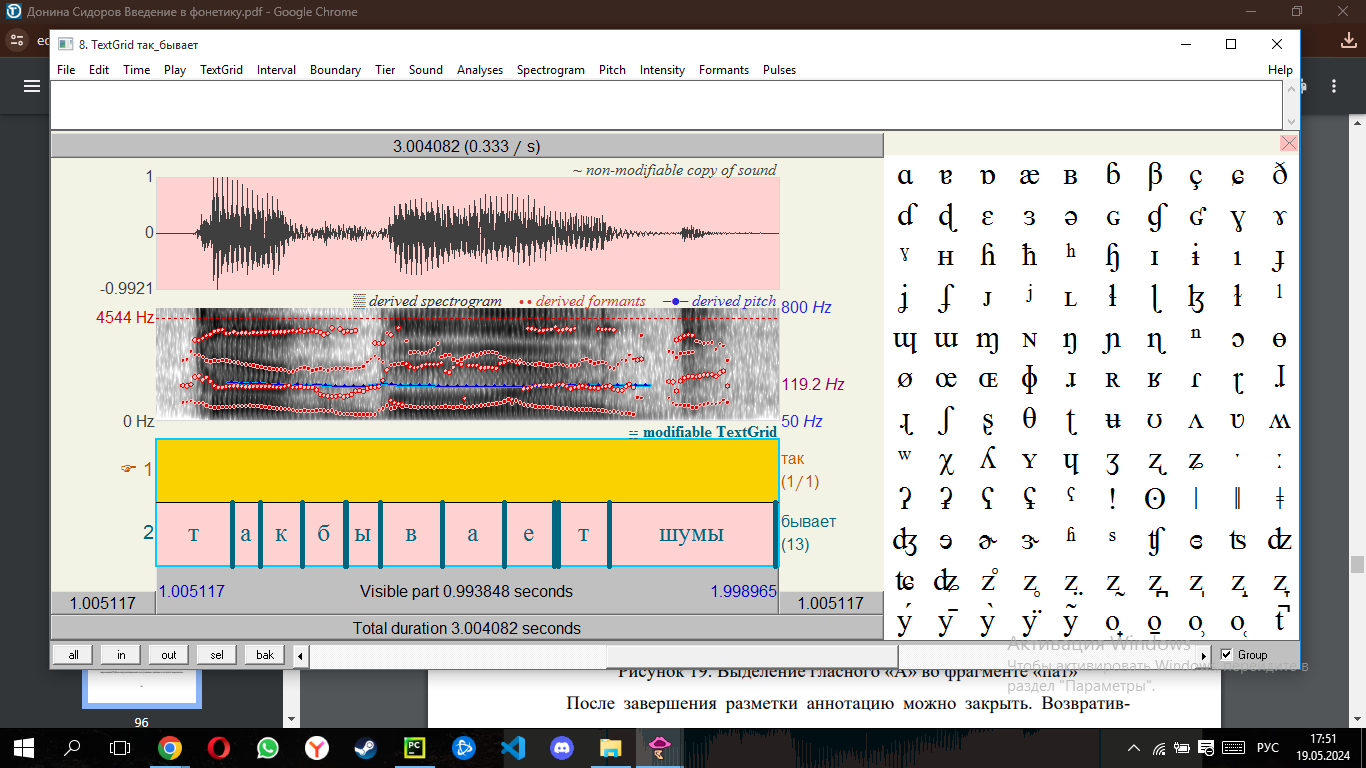
Я взял Алису и Google Translate. При постороннем разговоре значительных ошибок не было,а вот при несвязанных и несогласованных словах алиса не распознала практически ничего, а гугл более менее справился.

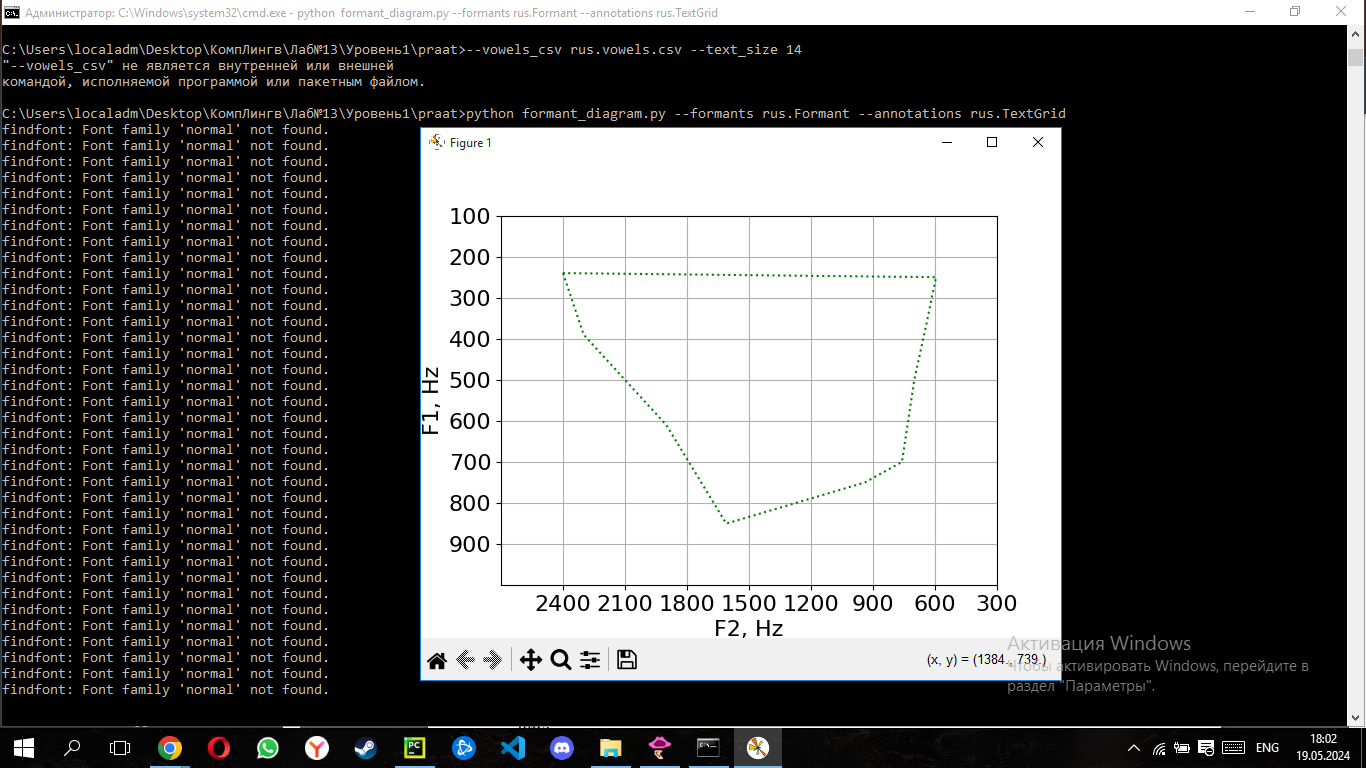
ЛАБА:

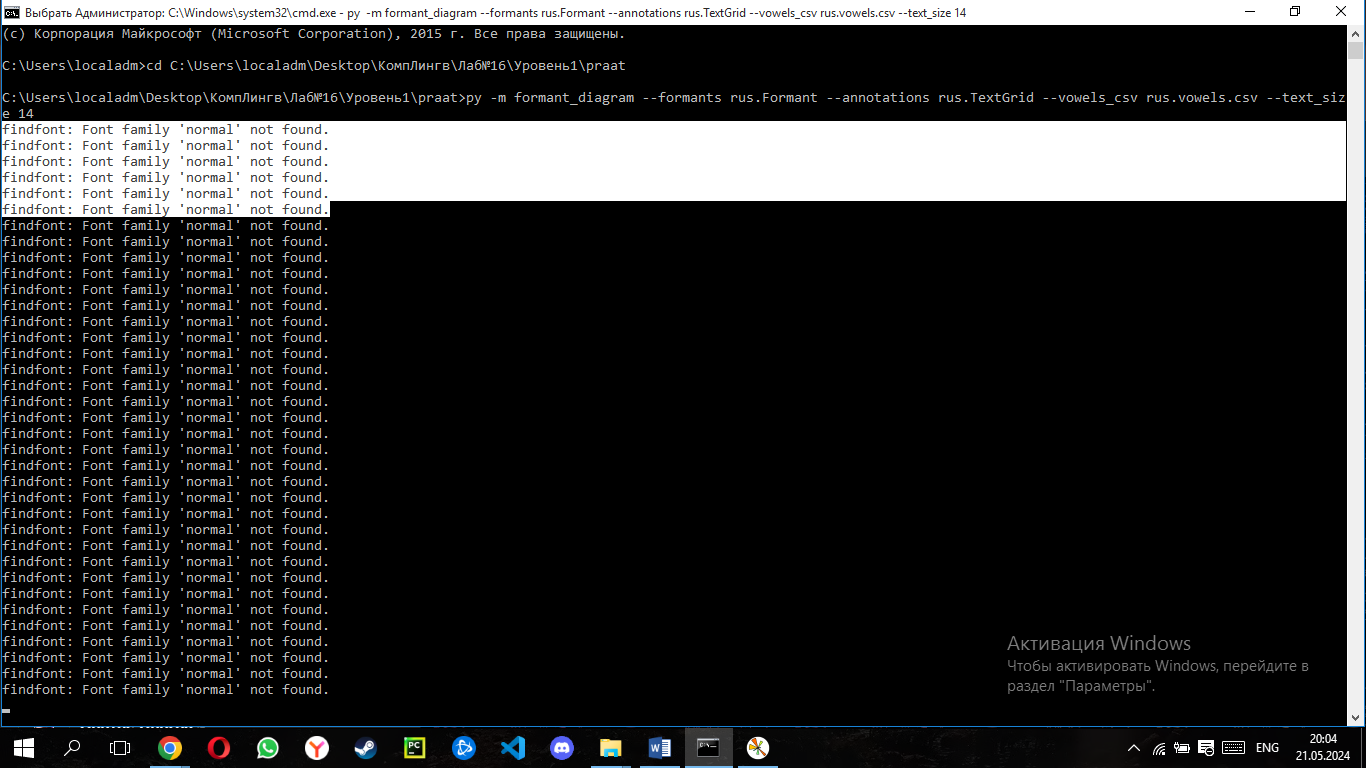












Стр.118

1. Основная задача систем преобразования текста в речь (TTS – Text-ToSpeech) – создание голосового сообщения без участия человека напрямую.

2. Основные критерии для оценки качества синтеза речи: разборчивость, т.е. возможность человеком декодировать речевой сигнал, и естественность – автоматически созданное высказывание не должно значительно отличаться от речи человека.

Под естественностью речи подразумевается наличие следующих свойств: соответствие синтезированных звуков речи входному тексту; определенная длительность и скорость произнесения каждого звука; наличие интонации; наличие тона соответствующей высоты; наличие акцента, что особенно важно для языков с несколькими вариантами (например, английский или итальянский).

3. В основе компилятивного синтеза речи заложены заранее записанные сегментные единицы (от звуков речи до целых высказываний), т.е. на вход системы подаются подготовленные человеком образцы речевых сигналов, которые система объединяет и озвучивает в требуемой последовательности

4. Формантный синтез осуществляется посредством искусственного создания акустических свойств звуков. Для создания качественного речевого сигнала необходимо около 60 акустических параметров. Данные параметры подразумевают всю информацию о частоте, амплитуде и ширине формант отдельных звуков. Процесс создания такого синтезатора требует много времени и подготовки.

Результатом работы подобного синтезатора является имитация речевых колебаний воздуха. Получаемый сигнал весьма сложно перепутать с человеческой речью, он имеет характерный «электронный» голос. Несмотря на то, что генерируемая формантными синтезаторами речь не естественна, она весьма удачно распознается человеком.

Статистический синтез речи подразумевает использование скрытых Марковских моделей, т.е. систем нейронных сетей. Их работу можно описать двумя этапами: во-первых, на вход системе подается достаточно большой объем фрагментов человеческой речи (минимум 5-10 часов), что позволяет нейронным сетям «обучиться» на речи диктора; во-вторых, система пытается самостоятельно воспроизвести речь по акустическим параметрам.

5.Артикуляционный синтез основывается на создании модели человеческого речевого аппарата и ее реализации с помощью некоторого механизма – «говорящей» машины, которая порождает речевой сигнал путем артикуляции. Акустические свойства речи на выходе производятся с учетом особенностей речевого аппарата человека, которые мы подробно рассмотрели в параграфе “Артикуляционная фонетика”. Создание аналога системы человеческого речевого аппарата все еще сложная задача и требует множества исследований в сферах нейрофизиологии, механики, аэродинамики и биоакустики. Имеющиеся на данный момент разработки способны порождать разборчивую речь, но назвать ее естественной пока нельзя.