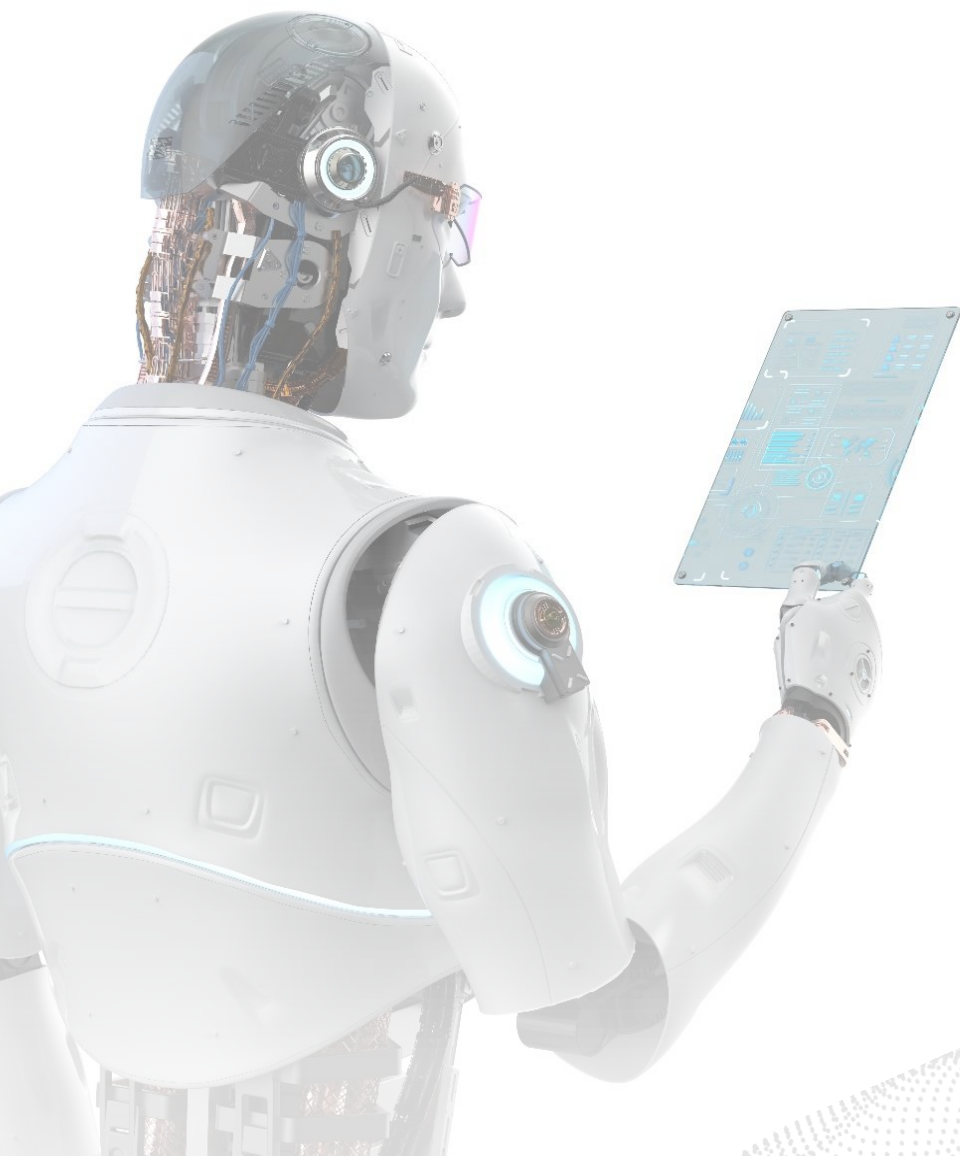


ТЕХНОЛОГИЯ СИНТЕЗА РЕЧИ

Формирование речевого
сигнала по печатному тексту

фцприи.рф



Также как предложение состоит из слов, а слова из букв, аудиофайл состоит из звуковых фрагментов, которые можно отобразить в виде дорожки аудиозаписей.

Например, имея запись из 100 слов, вы можете в нужном порядке собрать звуковое предложение.

В этом случае вы столкнетесь с тем, что у вас все слова в предложении звучат по разному:

- нет единой интонации;
- некорректные паузы;
- разная громкость слов;
- разный тон.

Задача Искусственного интеллекта/ нейросети:

- 1** Правильно переводить голос в текст для того, что бы впоследствии текст переводить в синтезированный голос (вы пишете текст, а в результате воспроизводится голос);
- 2** Обучить модели для правильного расставления пауз, правильных смысловых интонаций, тона и громкости в зависимости от смысла предложения.

Дополнительные возможности:

- Настройка синтеза для специализированных терминов, выражений, в том числе для иностранных слов;
- Эмоциональная окраска;
- Выбор синтезированного голоса и качества его произношения;
- Изменение тона и скорости речи;
- Использование предзаписанного голоса человека;
- Клонирование голоса (например, голос оператора КЦ);
- Добавление специальных или брендированных фильтров и звуковых эффектов.



Нажмите [«Сюда»](#) чтобы прослушать демо-запись синтеза речи.

Мы обучаем акустическую модель на речи людей, используя для этого нейронные сети

Для повышения качества синтеза мы дополнительно используем:



Открытые датасеты
(для экспериментов с моделями)



Платные датасеты и услуги диктора (для продакшена)



Дополнительные нейросетевые модули, повышающие качество и стабильность звучания



Шаг №1

NLP-препроцессор отвечает за подготовку данных и используется в ситуациях когда, например, необходимо расставить ударения, «е/ё» и так далее. Этот процесс осуществляется автоматически с помощью словарей и нейронных сетей

Шаг №2

Движок переводит текст в мел-спектограммы

Шаг №3

Вокодер переводит мелспектограммы в голос (для каждого диктора обучается своя модель)

Шаг №4

Постобработка — корректируется скорость, тон и громкость синтезируемого аудио

ОСНОВНЫЕ ХАРАКТЕРИСТИКИ



Технология синтеза речи

Формирует речевой сигнал по печатному тексту

40 часов

русскоязычной речи было использовано для обучения нейросетей



Плавная
речь



Естественная
интонация



Нейросетевая
обработка е/э



Нейросетевая
обработка е/ё



Расстановка
вопросов
и восклицаний в
интонации



Нейросетевая
нормализация
текста



Возможность
управлять
скоростью и
тоном



Нейросетевая
обработка
омографов



Возможность наложения
комфортного шума



Расстановка
пауз и ударений





- Эксперимент: скопировать голос известного радио- и телеведущего
- 10-12 часов речи
- Результат – голос с интонациями и привычными нотками

[Прослушать пример №1](#)



- Открытыми инструментами (бесплатными, в опенсорс) можно сделать очень качественные подделки
- Примеры ниже получены по 2 часам

[Прослушать пример №2](#)

[Прослушать пример №3](#)

**БЛАГОДАРИМ
ЗА ВНИМАНИЕ!**



fcprii.rf