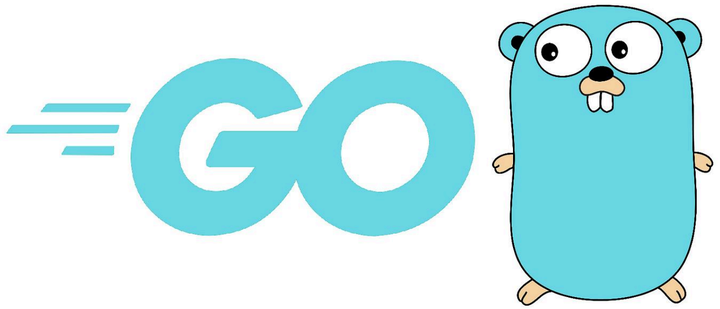
Использования модели распознавания речи Yandex SpeechKit для реализации субтитров с переводом занятия в реальном времени

Красиков Иван ПИ-21-1

Цель работы

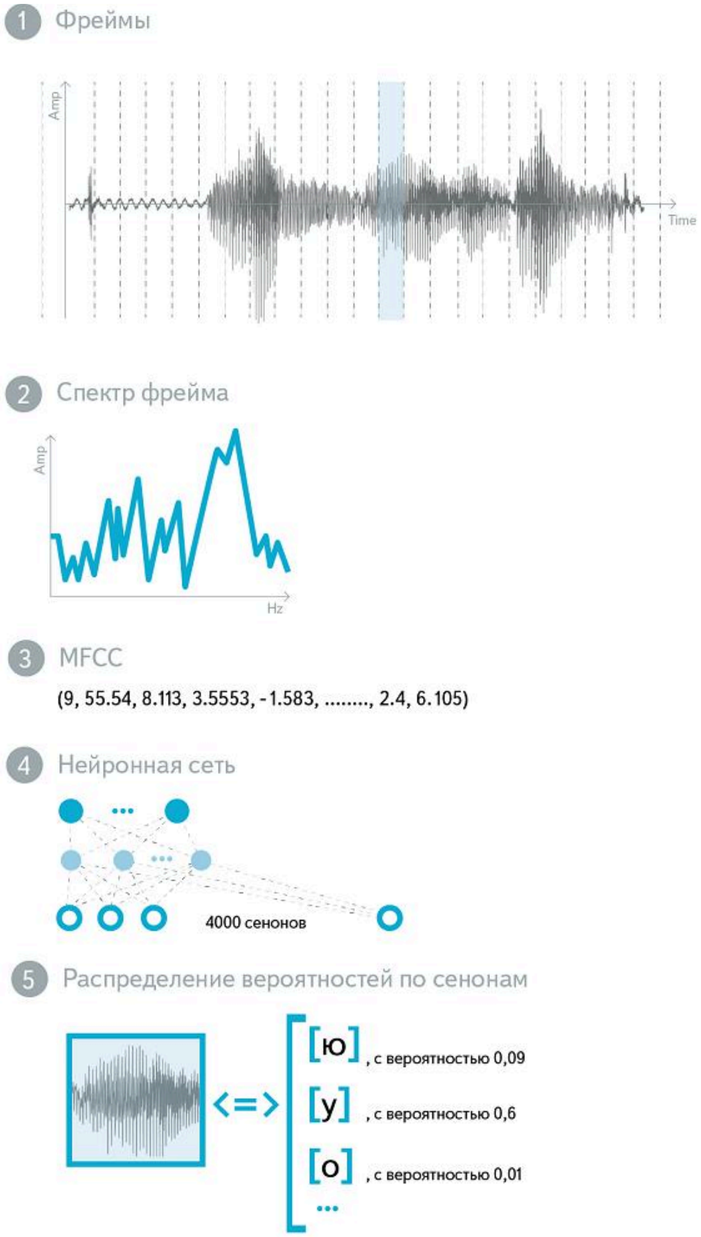
Ознакомится с процессом потокового распознавания речи и моделью Yandex SpeechKit. Реализовать веб-приложения для тестирования распознавания речи и вывода субтитров речи пользователя.

Используемые инструменты

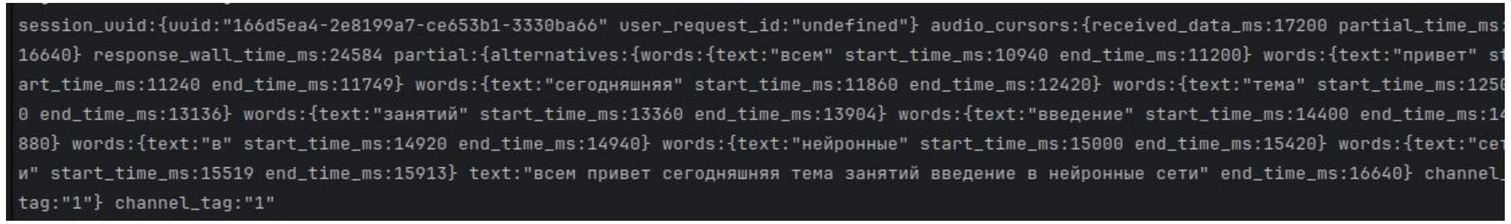


Краткое описание работы модели Yandex SpeechKit

1. Фреймы
   - Звуковой сигнал разбивается на короткие участки длительностью около 25 мс, называемые фреймами.
   - Между соседними фреймами есть перекрытие (обычно 10 мс). Это позволяет учитывать плавность изменений звука.
   - Каждый фрейм анализируется отдельно, что упрощает обработку длинных сигналов.
2. Спектр фрейма
   - Для каждого фрейма вычисляется его частотный спектр, который показывает распределение энергии звукового сигнала по частотам.
   - Это преобразование позволяет выделить ключевые особенности звука, которые невозможно увидеть в сыром виде (по амплитуде).
3. MFCC (Mel Frequency Cepstral Coefficients)
   - Спектральные данные преобразуются в компактное представление — MFCC.
   - Это вектор из нескольких чисел (обычно 13), который включает важную информацию о звучании фрейма, учитывая особенности человеческого слуха (Mel-преобразование).
   - Пример такого вектора показан: (9,55.54,8.113,3.5553,−1.583,...,2.4,6.105)(9, 55.54, 8.113, 3.5553, -1.583, ..., 2.4, 6.105)(9,55.54,8.113,3.5553, −1.583,...,2.4,6.105).
4. Нейронная сеть
   - На вход нейронной сети поступают несколько фреймов MFCC (учитывается текущий фрейм и его контекст).
   - Нейросеть анализирует входные данные и определяет, какой звук наиболее вероятен. На выходе сети — около 4000 сенонов.
   - Сеноны — это минимальные единицы звука, которые учитывают как фонему, так и её контекст (например, произношение звука в разных словах).
5. Распределение вероятностей по сенонам
   - Нейросеть выдаёт вероятностное распределение по сенонам:
     - Например, звук имеет 60% вероятность быть [у], 9% вероятность быть [ю] и 1% вероятность быть [о].
   - Эти вероятности помогают выбрать наиболее подходящий сенон, чтобы правильно распознать звук и интерпретировать его в текст.



Тестирование модели YandexSpeechKit
Пример ответа модели

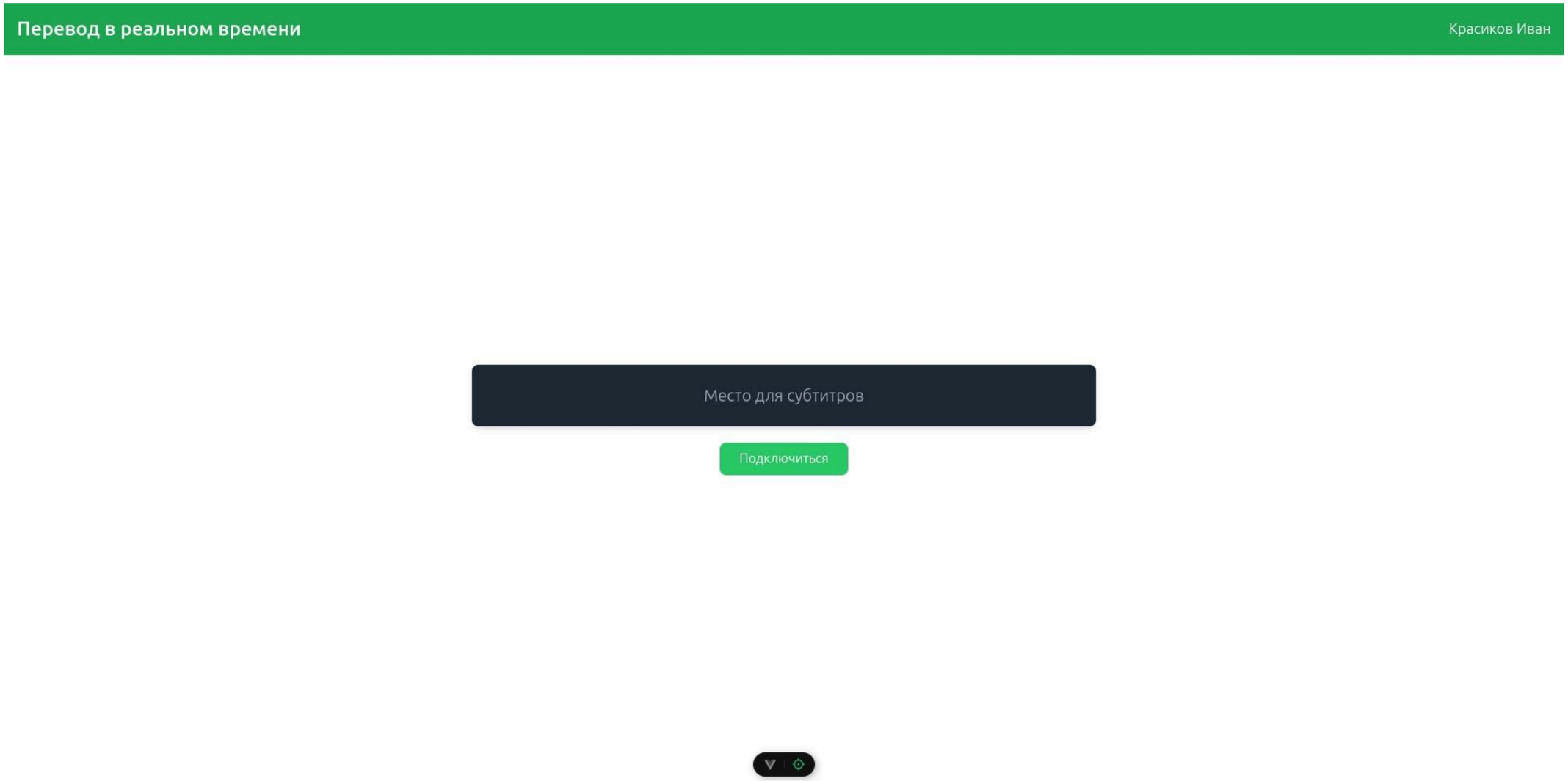


session_uuid:{uuid:"166d5ea4-2e8199a7-ce653b1-3330ba66" user_request_id:"undefined"}
audio_cursors:{received_data_ms:17200 partial_time_ms:16960} response_wall_time_ms:24654
partial:{alternatives:{words:{text:"всем" start_time_ms:10940 end_time_ms:11200} words:
{text:"привет" start_time_ms:11240 end_time_ms:11750} words:{text:"сегодняшняя"
start_time_ms:11860 end_time_ms:12420} words:{text:"тема" start_time_ms:12500
end_time_ms:13144} words:{text:"занятий" start_time_ms:13360 end_time_ms:13907} words:
{text:"введение" start_time_ms:14400 end_time_ms:14880} words:{text:"в" start_time_ms:14920
end_time_ms:14944} words:{text:"нейронные" start_time_ms:15000 end_time_ms:15420} words:
{text:"сети" start_time_ms:15519 end_time_ms:15920} text:"всем привет сегодняшняя тема
занятий введение в нейронные сети" end_time_ms:16960} channel_tag:"1"} channel_tag:"1"

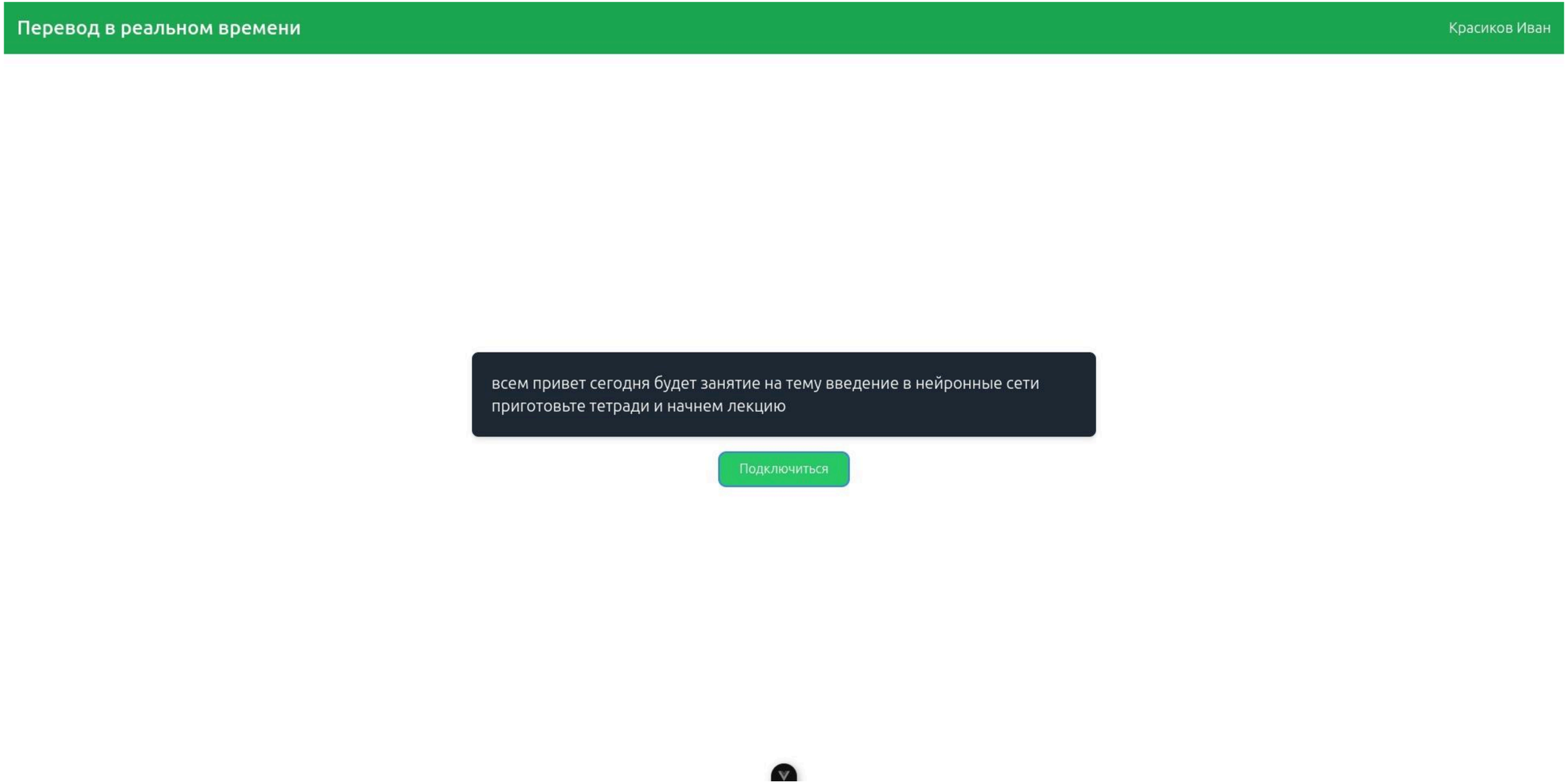
Тестирование модели YandexSpeechKit
Распознанный текст из ответа модели

```
всем привет
всем привет
всем привет
всем привет
всем привет сегодня
всем привет сегодня
всем привет сегодня занятие
всем привет сегодня занятие
всем привет сегодня занятие будет на канале
всем привет сегодня занятие будет на канале
всем привет сегодня занятие будет на тему
всем привет сегодня занятие будет на тему
всем привет сегодня занятие будет на тему введение
всем привет сегодня занятие будет на тему введение
всем привет сегодня занятие будет на тему введение в
всем привет сегодня занятие будет на тему введение в
всем привет сегодня занятие будет на тему введение в нейронные сети
всем привет сегодня занятие будет на тему введение в нейронные сети
всем привет сегодня занятие будет на тему введение в нейронные сети
всем привет сегодня занятие будет на тему введение в нейронные сети
всем привет сегодня занятие будет на тему введение в нейронные сети
всем привет сегодня занятие будет на тему введение в нейронные сети
всем привет сегодня занятие будет на тему введение в нейронные сети
всем привет сегодня занятие будет на тему введение в нейронные сети
всем привет сегодня занятие будет на тему введение в нейронные сети
всем привет сегодня занятие будет на тему введение в нейронные сети
```

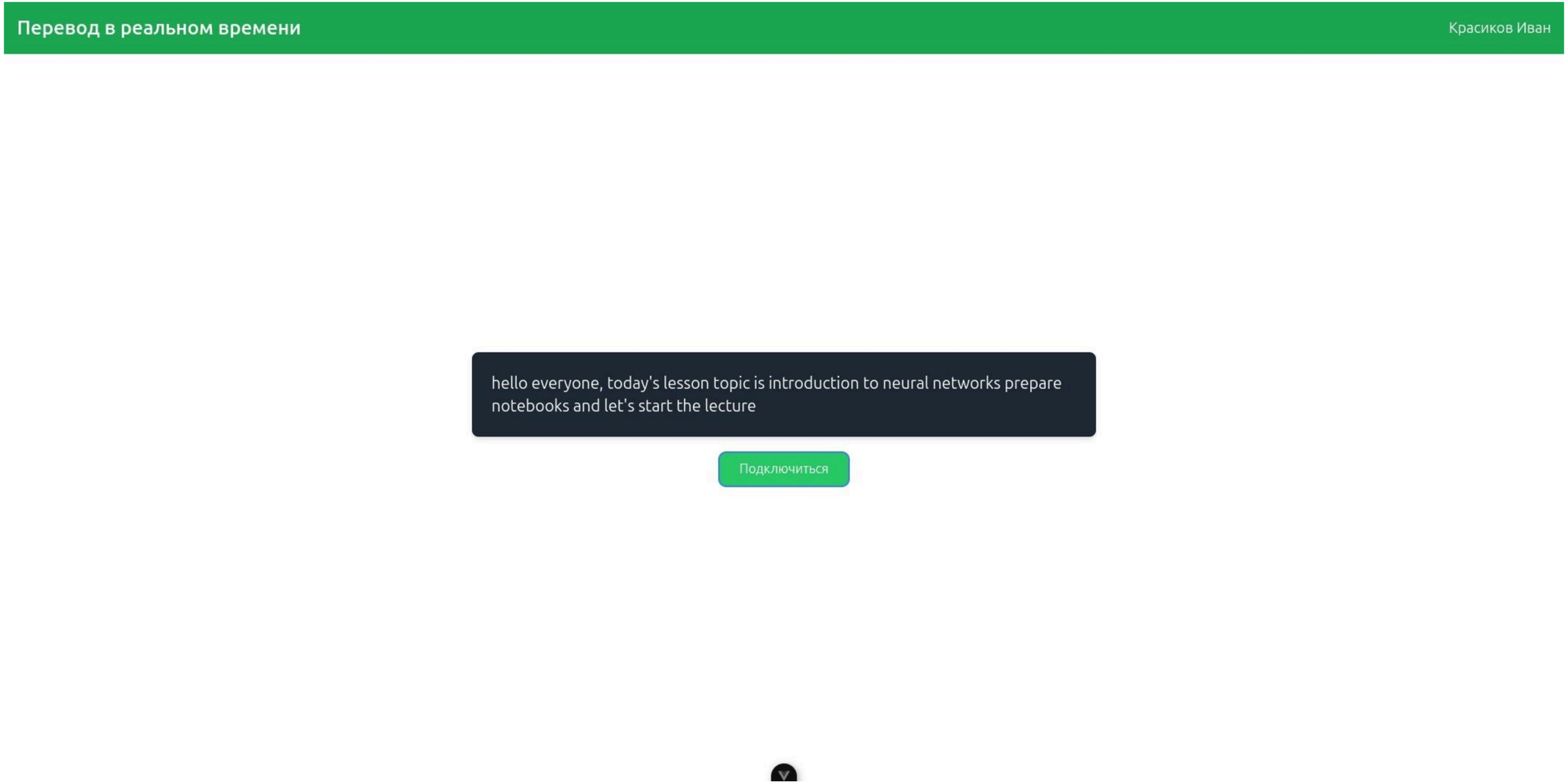
Внешний вид клиентской части



Пример субтитров в клиентской части



Пример субтитров с переводом в клиентской части



Спасибо за внимание