Technical requirements and specification

«Accent chat-bot»

Content

Description

Product Perspective

Most of the foreign language learning on-line applications today use a lot of different automatic features to examine language skills of their end-users. Although reading, writing, spelling and even listening are common to be tested, some functionalities that review speaking skills are rarely automatic and easy to use. Accent Chatbot, however, will try to address this issue and propose to its users an opportunity to score their English pronunciation quality.

There will be three main units to the system working together to provide better functionality:

- Conversational framework - capable of asks user required audio for a particular short text in order to score them and answer in a simple and natural form
- Audio enhancement - capable of noise removal to produce clearer speech
- Accent scoring - capable of providing level of similarity of user speech to the set of predefined English accents

Product Features

The major features for Accent chat-bot will be the following:

- Web API: An API call will include a sample audio provided by the client as a request attached file and the service will reply in JSON.
- Audio processing: The system will preprocess provided audio files in order remove some additional noise and enhance next steps of the model.
- English accent scorer: The answer will be written in standard and understandable English to provide simple pronunciation scoring results.

Constraints

Language: The system will only support English accents recognition.

Audio duration: The system will only support short audios due to computation costs

Functional Requirements

API calls

1. Client responsibilities

    1.1 The client will send an audio message (a voice file in wav/OGG format) with the sequence of English words proposed by the bot

2. Server responsibilities

    2.1 The server will send the sentence that the user is asked to read

    2.2 The server will send a score of closeness to the English accent and top of 3 accent recognized in speech

    2.3 The server will respond with a 400 Bad Request status code if an audio message does not specify requirements

Error Handling

1. If the system doesn't recognize any accent in the voice message. Example message, “Sorry, I didn't hear you. Let's try again”
2. If the audio message length exceeds 20 seconds. Example message, “Try to keep within 20 seconds”

Audio processing pipeline

1. Audio preprocessing

A model that removes noise will be used as a pre-processing to improve the quality of recognition.

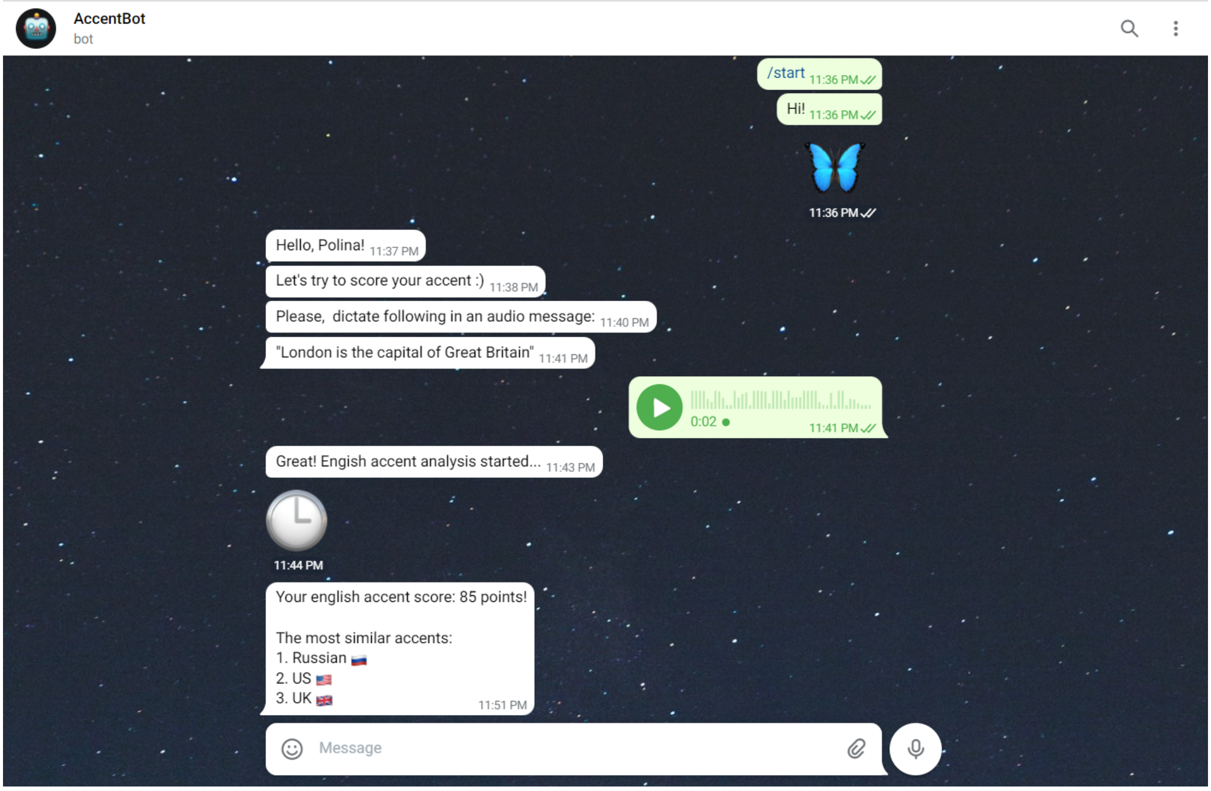
2. Accent recognition

Accent recognition model supports 10 accents (Chinese, American, England, Korean, Japanese, Russian, Indian, Portuguese, Spanish and Canadian). System uses 3 the model's greatest output probability to create the top of accents recognized in speech.

3. English accent scoring

System uses English accent output model probability to compute English accent score.

User Interfaces



AccentBot
bot

/start 11:36 PM

Hi! 11:36 PM

11:36 PM

Hello, Polina! 11:37 PM

Let's try to score your accent :) 11:38 PM

Please, dictate following in an audio message: 11:40 PM

"London is the capital of Great Britain" 11:41 PM

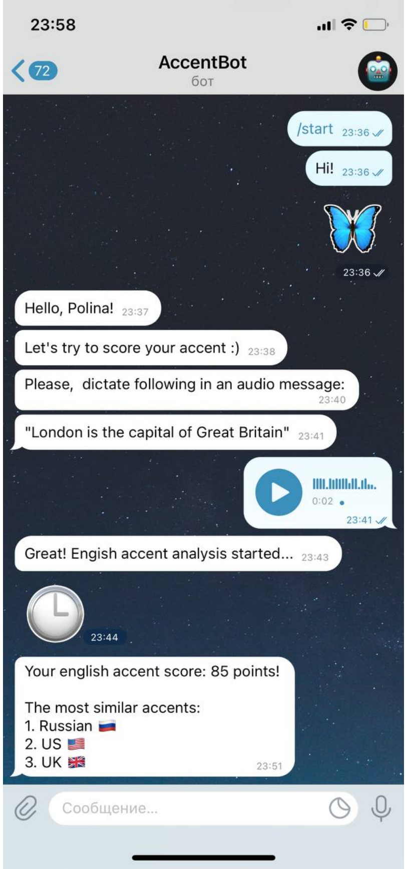
0:02 ● 11:41 PM

Great! Engish accent analysis started... 11:43 PM

11:44 PM

Your english accent score: 85 points!

The most similar accents:
1. Russian 🇷🇺
2. US 🇺🇸
3. UK 🇬🇧
11:51 PM



/start 23:36

Hi! 23:36

23:36

Hello, Polina! 23:37

Let's try to score your accent :) 23:38

Please, dictate following in an audio message: 23:40

"London is the capital of Great Britain" 23:41

0:02 • 23:41

Great! Engish accent analysis started... 23:43

23:44

Your english accent score: 85 points!

The most similar accents:
1. Russian 🇷🇺
2. US 🇺🇸
3. UK 🇬🇧
23:51

Non-functional Requirements

Modularity

1. The system will be designed in such a way that the algorithms for the noise removal and accent scorer units will be able to be easily swapped out.

Model Quality / Accuracy

1. The overall accuracy of the Web API’s response will be measured using a developer-made testing set.
2. The overall accuracy is measured by two metrics top-1 and top-3 accuracy.
3. Top-1 accuracy is calculated by dividing total number of correct answers by the number of questions asked.
4. Top-3 accuracy is calculated by dividing total number of hits of the correct class in the top-3 predictions by the number of questions asked.
5. Top-3 accuracy of the accent recognition part will be close to 80%.
6. Top-1 accuracy of the accent recognition part will be close to 70%.
7. Top-1 and top-3 accuracy will not decrease significantly if noise is present in the audio data.

Performance / Fast Response

1. The average time for the server to respond, over the question testing set, will be less than or equal to 2 seconds.
2. Audio preprocessing and accent scorer units will not require mandatory GPU inference

Use cases

Use case Flow

Web API

Precondition: The server that is running the API is online.

Main Flow: The user sends their audio as an attached file in POST request to the API

Postcondition: The user receives the answer to their request in JSON.

Telegram UI

Entering a question

Preconditions: The user starts the Telegram application.

Main Flow:

1. The service sends a sample text to the chat window.
2. User sends an audio message with the dictated text to the chat.
3. If the audio message does not align with the requirements, the service will send aт error message and go to 1.
4. After some time, some answer generated by the service appears in the chat window with the accent scoring results in a natural language format.

Postcondition: The service will send a message with a suggestion to end the dialog or to start scorning over.