Towards a Customizable Text-to-speech Personal Assistant

Audio indexing and processing: Project proposal

Shupei Li, Chenyu Shi, Siwen Tu, Shuang Fan
Oct.10 2023

Problem Statement

- There are many mature language models related to

the text-to-speech generator:

  - Audiobooks reading (using Text-to-speech

  can save the cost and customize voice)

  - Siri in iPhone 'I could read this text for you'

  - Google translate 'The pronunciation of the word'

- Can we build a customizable text-to-speech personal assistant that can speak a specific language, even with a specific voice?
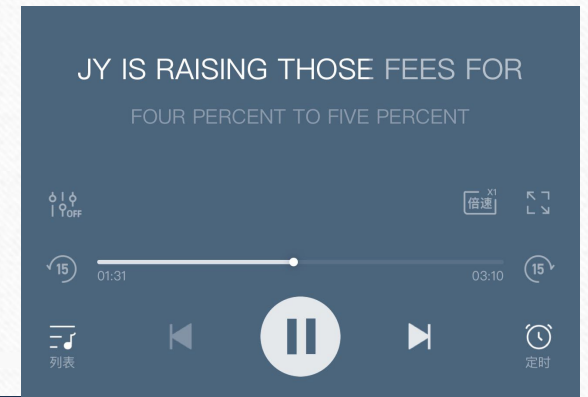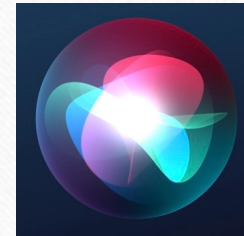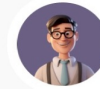
Challenges

1. Lack of available datasets for multi-language and multi-speaker.

2. Difficult to fine-tune the multi-language and multi-speaker models to a satisfied result.

Goals for Demo Day

Change voices

1. Accomplish the text-to-speech task based on YourTTS pre-trained model.

2. Multi-language transform. E.g. fine-tune the model on Chinese datasets.

3. Multi-speaker transform. E.g. fine-tune the model on multi-speaker datasets.

Combining 2 and 3, we can get a customize assistant with a specific language and voice.

Work Division

Text-to-speech model construction: all group members

Multi-language fine-tuning task: Shupei Li, Siwen Tu

Multi-speaker fine-tuning task: Chenyu Shi, Shuang Fan

Report writing and presentation preparation: all group members