New chat

Model: GPT-4

T

I'm working to develop an ASR Model.

I have a dataset of .wav recordings, each recording is a long conversation between people. Each .wav recording is link to another transcript file that contains information of the marking of each utterance's in the conversation.

I want to break down the conversation .wav file into utterance .wav file with the provided marked timing. Are there any python libraries to do this?

Yes, there are Python libraries that can help you accomplish this task. One such library is `pydub`, which is a simple and easy-to-use library for manipulating audio files. You can use it to slice the .wav recordings into utterance .wav files based on the marked timings from the transcript.

First, you need to install `pydub`:

bash Copy code

```bash
pip install pydub
```

Then you can use the following code to split the .wav file into separate utterance .wav files:

python Copy code

```python
from pydub import AudioSegment

def split_wav_file(input_file, start_time, end_time, output_file):
    # Load the input .wav file
    audio = AudioSegment.from_wav(input_file)

    # Convert start and end time to milliseconds
    start_time_ms = start_time * 1000
    end_time_ms = end_time * 1000

    # Slice the audio
    sliced_audio = audio[start_time_ms:end_time_ms]

    # Export the sliced audio to a new .wav file
    sliced_audio.export(output_file, format="wav")

# Example usage:
input_wav = "path/to/your/conversation.wav"

# Assuming you have a list of tuples with start and end times for each utterance
utterance_timings = [(0, 5), (5, 12), (12, 20)]  # Replace this with actual timings from the transcript
```