Task 3: Fourier Transform

Our goal is to increase amplitude, improve speech quality, and ensure that the audio is free of clipping or distortion.

In order to find the region of harmonic speech frequency from the spectrum, we intercepted multiple vowel and consonant sounds and drew the spectrum diagram as follows.

Below is a spectrum of the sound of vowel “a”.

图表

描述已自动生成

Below is a spectrum of the sound of vowel “e”.

图表

描述已自动生成

Below is a spectrum of the sound of vowel “i”.

图表, 折线图

描述已自动生成

Below is a spectrum of the sound of vowel “o”.

图表, 折线图

描述已自动生成

Below is a spectrum of the sound of consonant “b”.

图表, 折线图

描述已自动生成

Below is a spectrum of the sound of consonant “d”.

图表, 折线图

描述已自动生成

As can be seen from the figures, the audio will drop significantly from 1000Hz or 2000Hz, and the higher frequency component is noise, so we can use the range of 50-2000Hz as the harmonic speech frequency region.

By the way, everyone with the same vowel will have difference, but the overall feature is about:

1. In the frequency domain, vowels have distinct peaks, which are sharp in shape, while consonants have smooth peaks.

Vowel [Consonant](javascript:;)

图表, 折线图

描述已自动生成 图表, 折线图

描述已自动生成 图表, 折线图

描述已自动生成 图表, 折线图

描述已自动生成 图表, 折线图

描述已自动生成

1. The highest vowel peak is further away from the second highest peak, while the consonant peak is closer together
2. Vowels have about three high peaks, and consonants have about four

So I think one way to distinguish vowels is to segment the audio at 0.1 second intervals and determine if each segment has a similar characteristic waveform, and if it does, it's a vowel.