project1：Speech synthesis and perception with envelope cue

Author	Name：张子尚	Student ID:12012305
	Name：华羽霄	Student ID:12010508
	Name：刘子羽	Student ID:12010514
	Name：徐建辉	Student ID:12112807

Introduction

Use Matlab software to simulate the low-pass filtering of the speech signal and save the audio obtained after processing. In addition to completing the given parametric conditions, test other orders and cutoff frequencies to find the relationship between the filtering effect and the two. We are expected to find a suitable order and cutoff frequency range so that the processed signal is closest to the original signal.

Lab results & Analysis:

Task 1(张子尚)

Sentences for pro 1: 'C_01_01.wav' & 'C_01_02.wav'

– Set LPF cut-off frequency to 50 Hz.

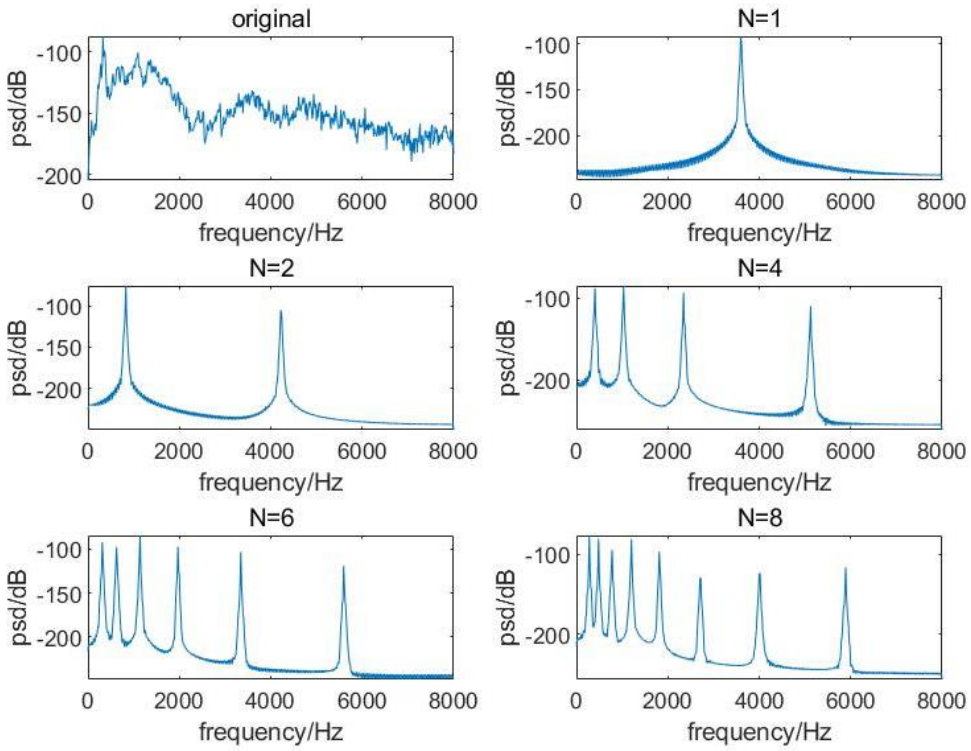
– Implement tone-vocoder by changing the number of bands to N=1, N=2, N=4, N=6, and N=8.

– Save the wave files for these conditions, and describe how the number of bands affects the intelligibility (i.e., how many words can be understood) of synthesized sentence.

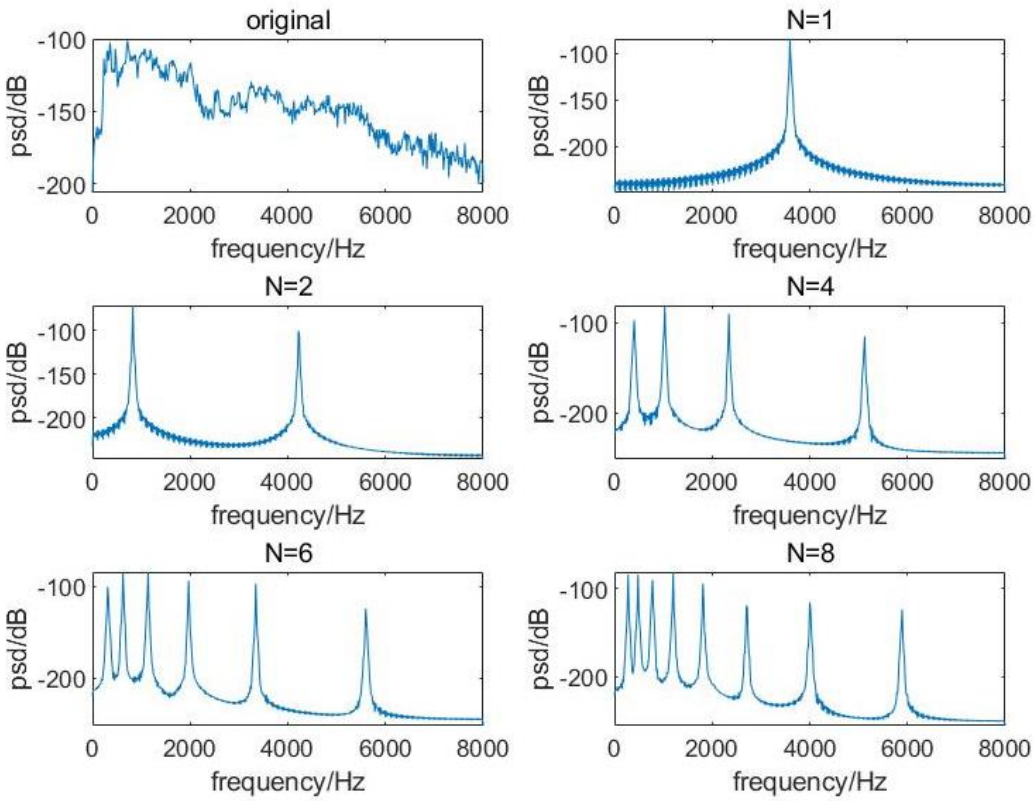
Result:

Power Spectral Density figures of C_01_01.wav & C_01_02.wav

C_01_01.wav：

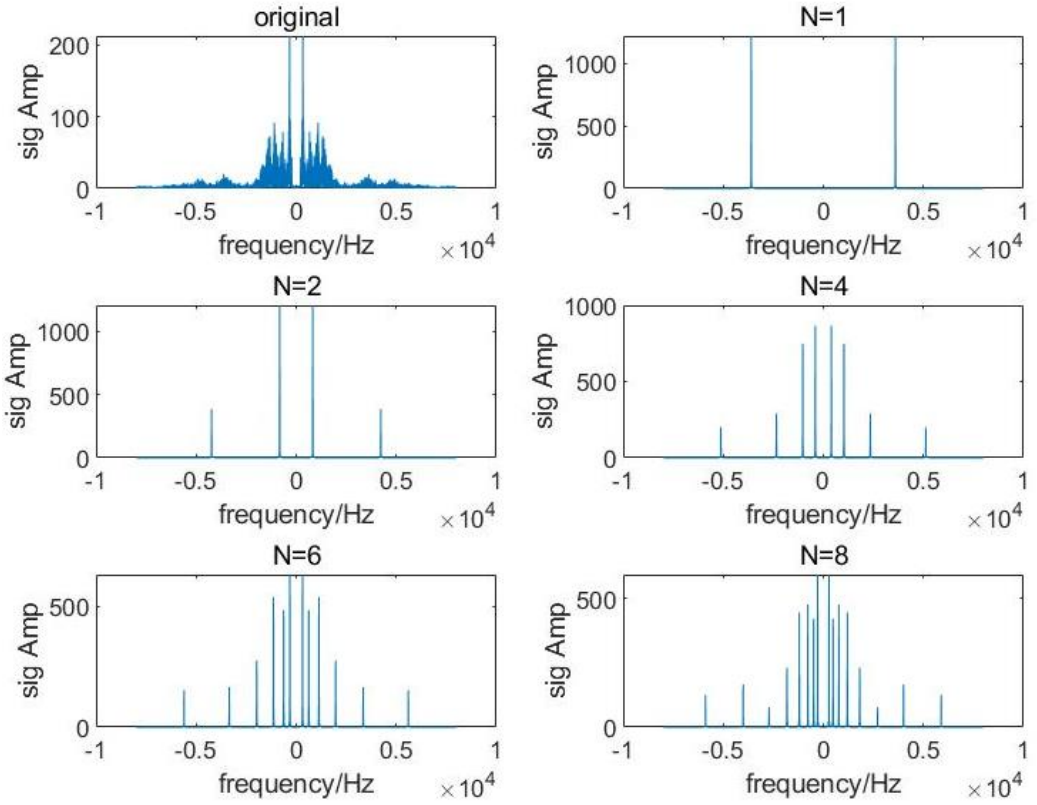


C_01_02.wav：

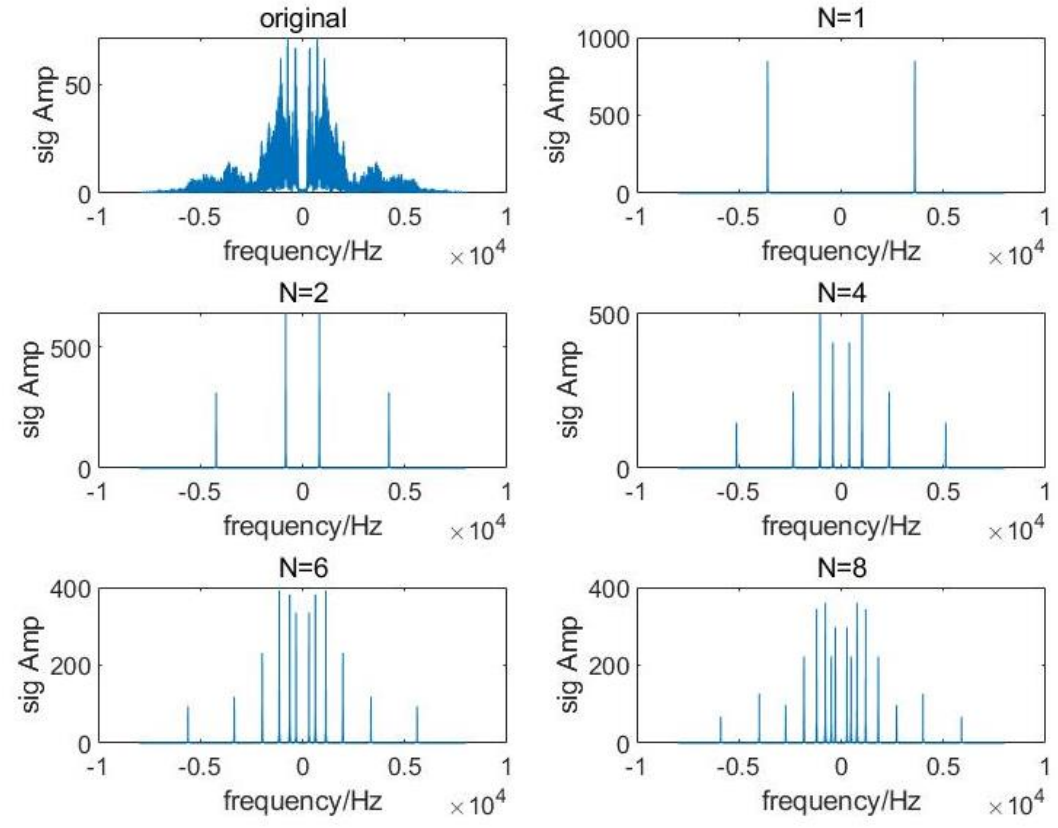


Graph of C_01_01.wav & C_01_02.wav in frequency domain:

C_01_01.wav:

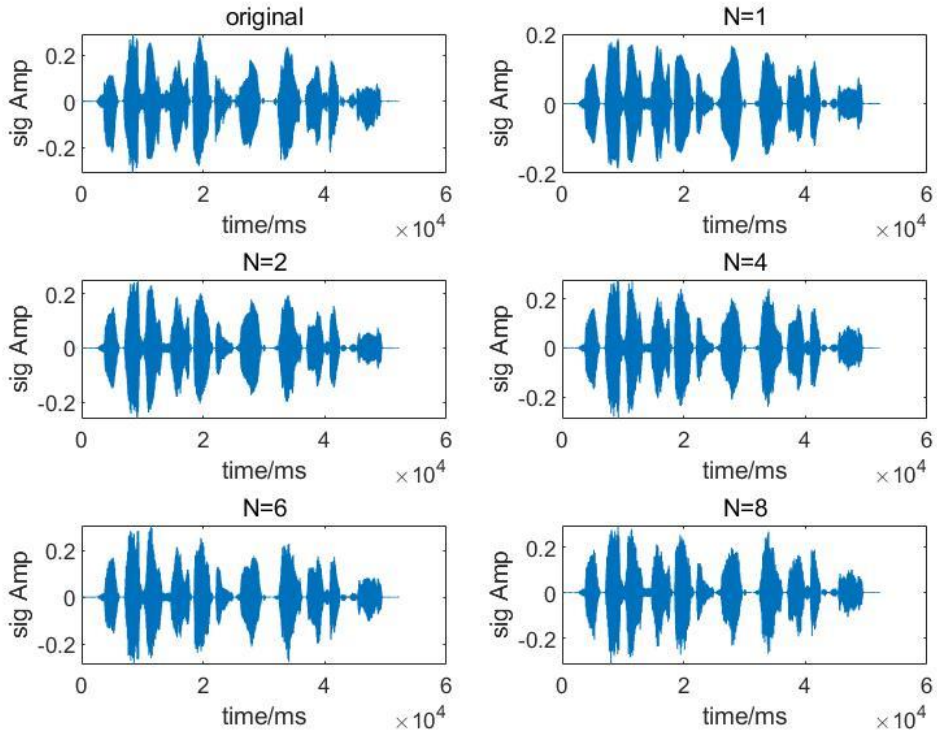

original
N=1
N=2
N=4
N=6
N=8
sig Amp
frequency/Hz
×10^4

C_01_02.wav:

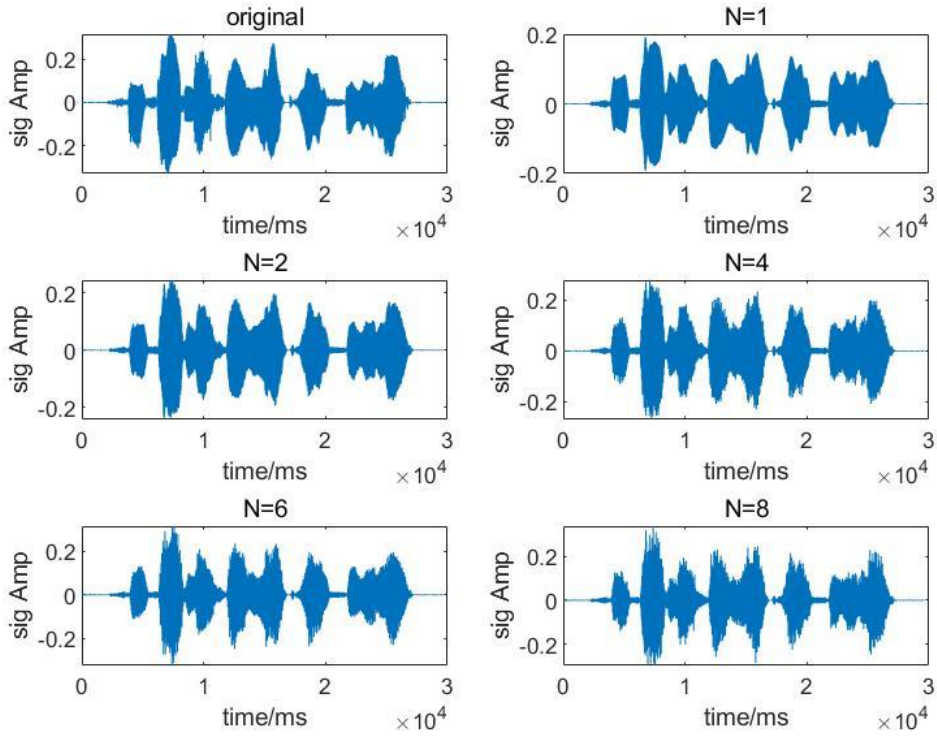

original
N=1
N=2
N=4
N=6
N=8
sig Amp
frequency/Hz
×10^4

Graph of C_01_01.wav & C_01_02.wav in time domain:

C_01_01.wav:



C_01_02.wav:

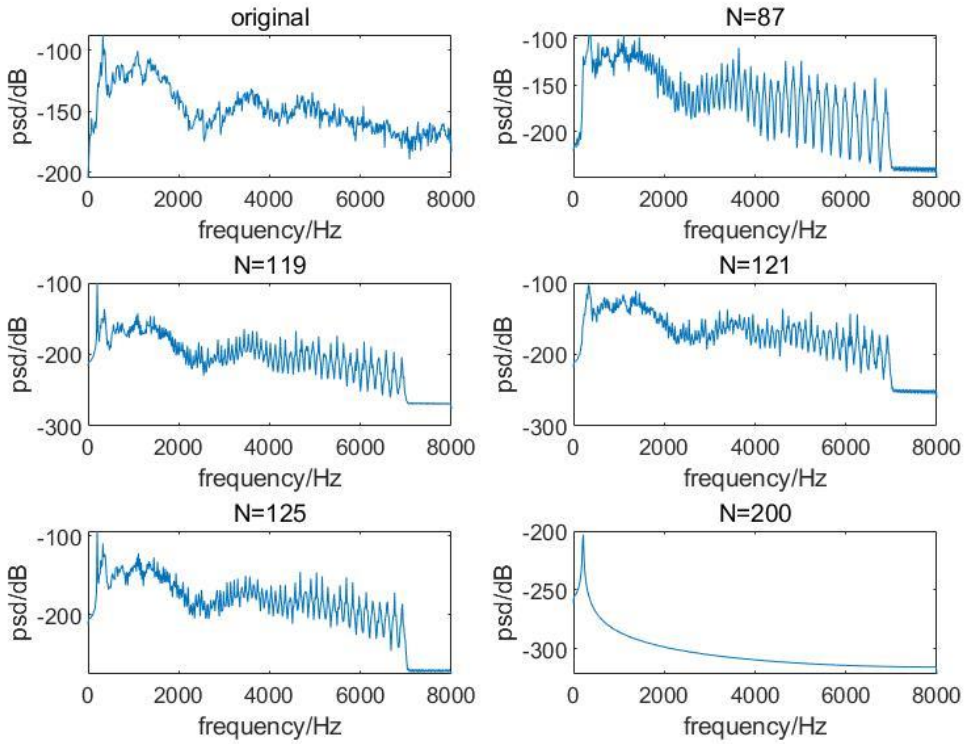


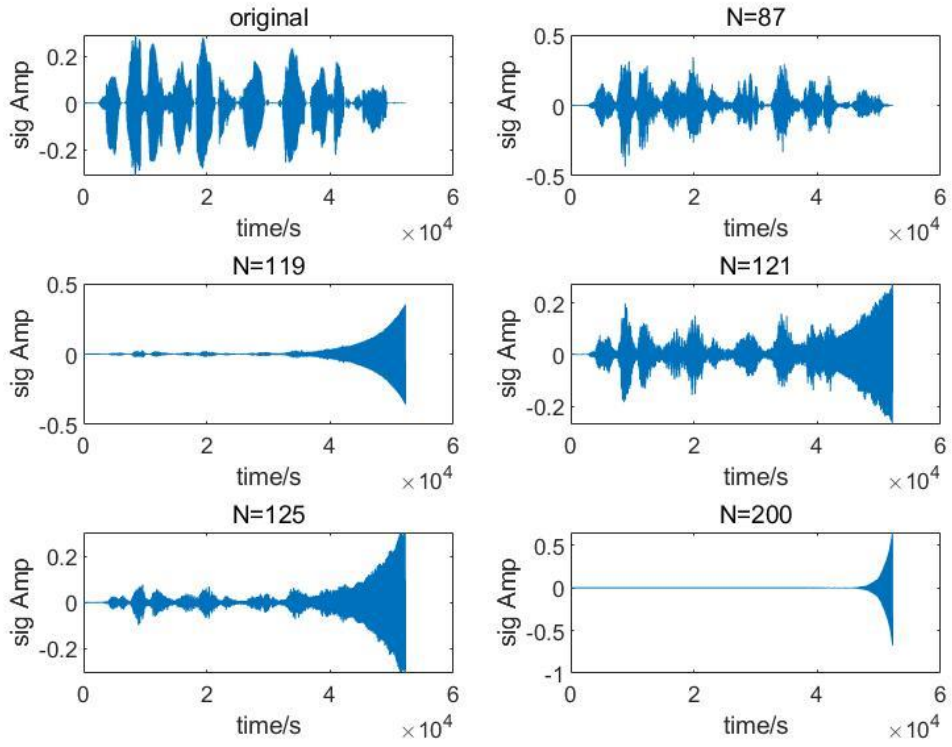
Analysis:

1. Analysis of PSD, frequency domain response and waveform when frequency band N is of different values shows that the processed image is closer to the original signal image with the increase of frequency band number. Meanwhile, with the increase of frequency band N, the readability of audio gradually increases
2. In task1, the N we picked varied from 1,2,4,6 to 8, and even though we gradually understood the audio better, we still heard only a vague audio, with no specific information.

Guess: In Task1, the readability of the audio improves significantly as N increases. But the maximum value of N, 8, is still small. So the audio may be unclear because the value of N is too small. We suspect that increasing frequency band N can improve the readability of processed audio within a certain range.

Critical thinking (about N):





1. As can be seen from the PSD figure, in the case of high N, the energy density of high frequency band gradually decreases with the increase of frequency band N
2. From the sound waveform diagram, when N is high, with the increase of frequency band N, the waveform amplitude of the sound part bearing useful information decreases and its features become less obvious. However, when N=121 and 125, the amplitude of the sound wave carrying useful information increases in reverse. It is supposed that the frequency band N is exactly divided into the discontinuity of speech frequency. When N is very large, the amplitude of the sound waveform is mainly at the end of the sound, which is the original audio clip without speech.
3. In the filter, we set the order of Butter as 4, which will rise and fall in low frequency and high frequency parts, which is not an ideal filter and will affect the experimental results.
4. After repeated attempts, the speech is most clear when the cutoff frequency is 50Hz and the value of N is 87.

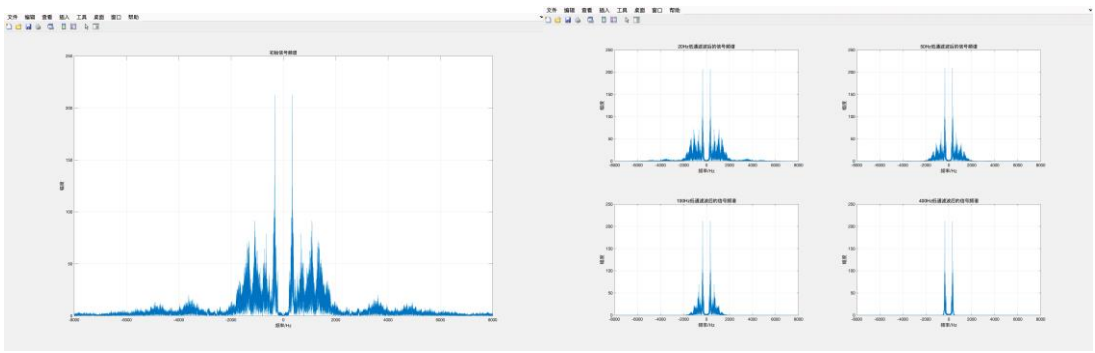
Task2(华羽霄)

– Set the number of bands N=4.

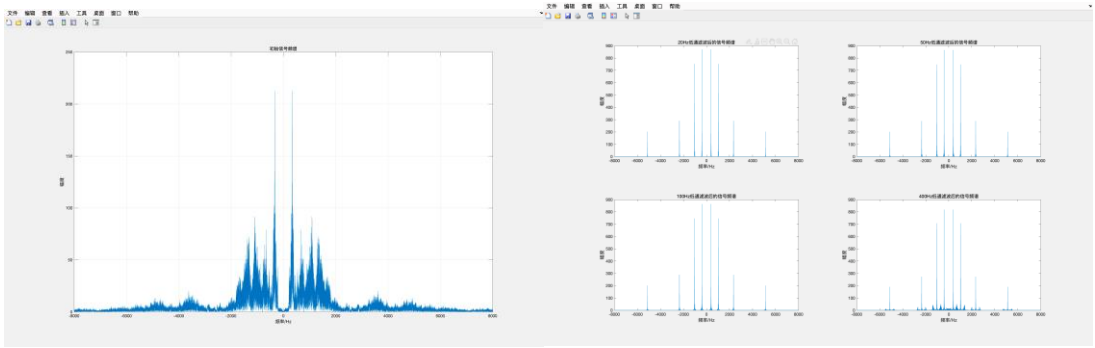
– Implement tone-vocoder by changing the LPF cut-off frequency to 20 Hz, 50 Hz, 100 Hz, and 400 Hz.

– Describe how the LPF cut-off frequency affects the intelligibility of synthesized sentence.

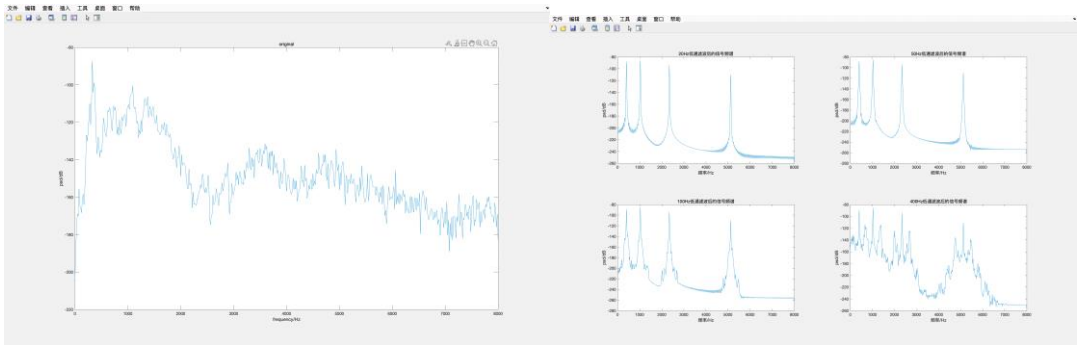
Result:



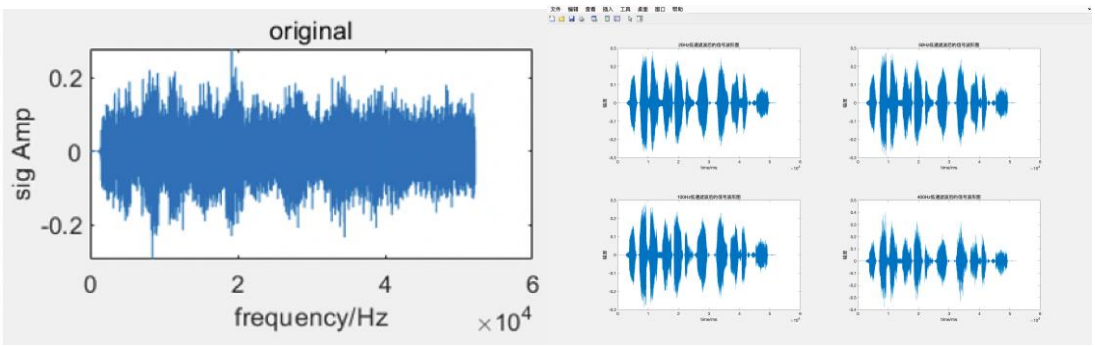
Signal spectrogram



fft plot



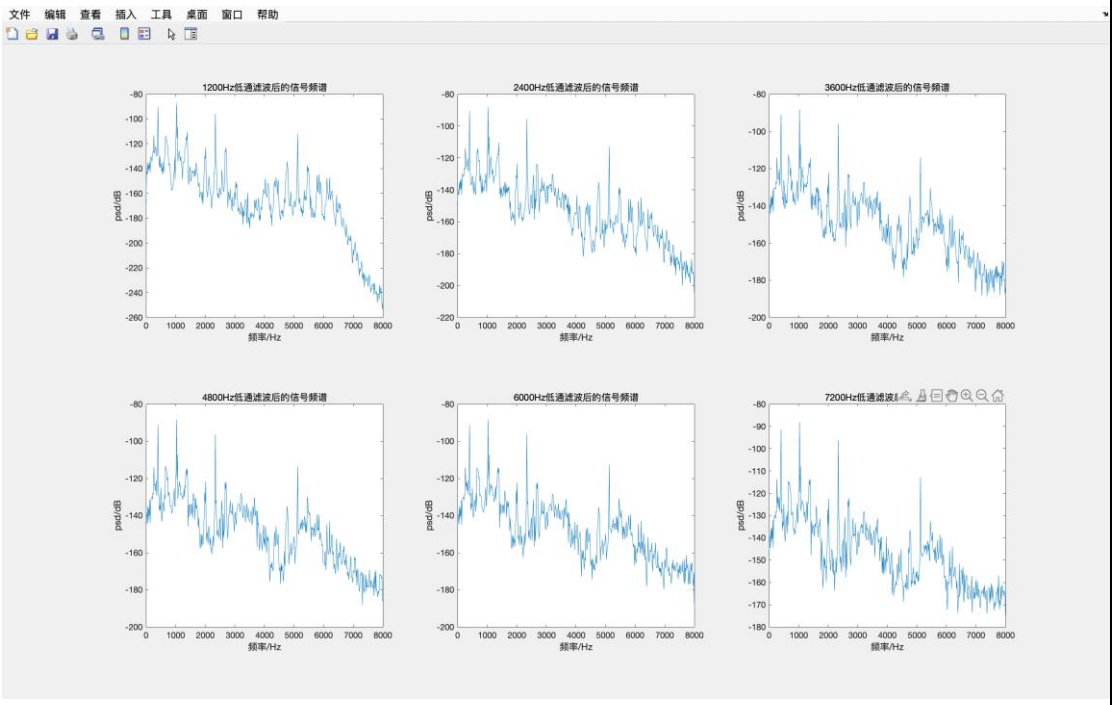
envelop plot

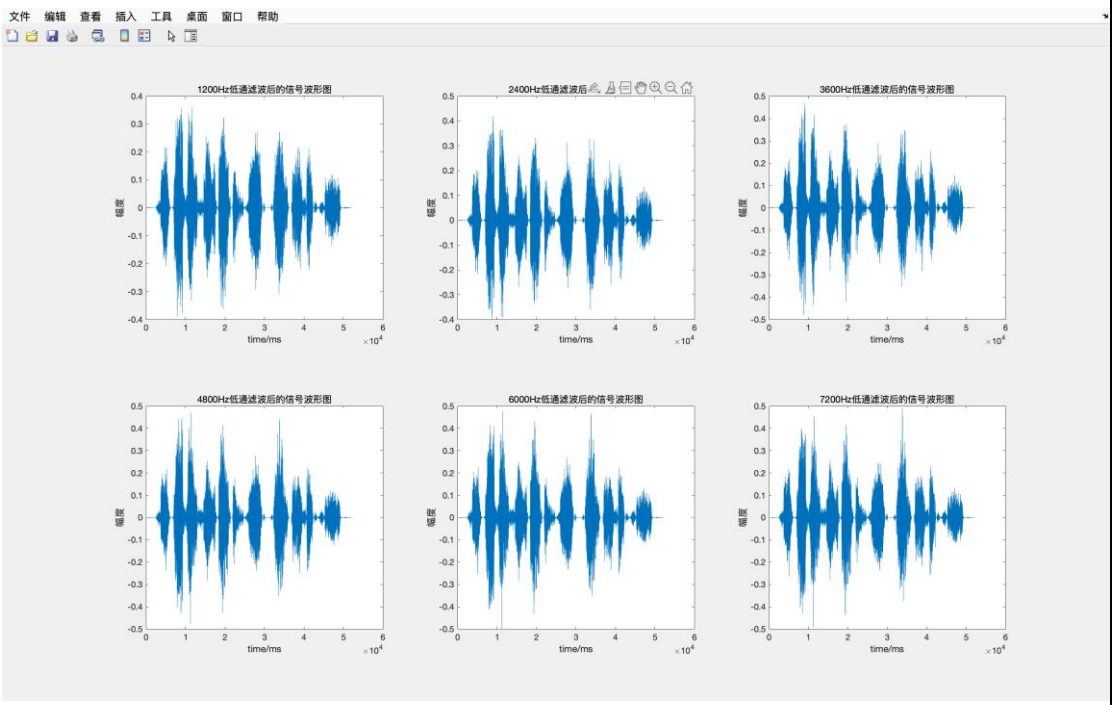


Waveform plot

Analysis:

It is not difficult to find that when the cutoff frequency is between 20Hz and 400Hz, the voice signal processing effect is very poor. So, I expanded the upper limit of the cutoff frequency to 7200Hz, and tested every 1200Hz, and the results are as follows:





Analysis:

After setting the number of bands to N=4, the result sound signal can hardly be recognized when the cutoff frequency is below 400Hz, and we can not determine whether the sound is from a

human or machine. As a result, we gradually increase the value of cutoff frequency in order to determine the possible relationship between the effect of low pass filter and the cutoff frequency. We found that with the cutoff frequency increasing, the result sound signal is getting more and more recognizable. However, since the limit of the number of bands is set to a fixed number 4, the result sound signal is hard to be recognized from beginning to end. And the higher limit of the low pass filter should be lower than 8400Hz, which means that we cannot continue our work to higher frequency tests.

Conclusion:

Under the condition of a fixed number of bands, we can easily find that, with higher cutoff frequency, the result sound signal is easier to be recognized. This is because with higher cutoff frequency, more information of the original signal will be concluded into the result signal, and thus the result signal is more similar to the original one.

In future tests, we are trying to find the relationship between the effect of the filter and the choice of different number of bands and cutoff frequency. Finally, we will be able to find the most suitable value of bands and frequency, under which circumstance the result sound signal is the most similar to the original one.

Task 3(刘子羽)

Project tasks3

-Generate a noisy signal (summing clean sentence and SSN) at SNR = -5 dB.

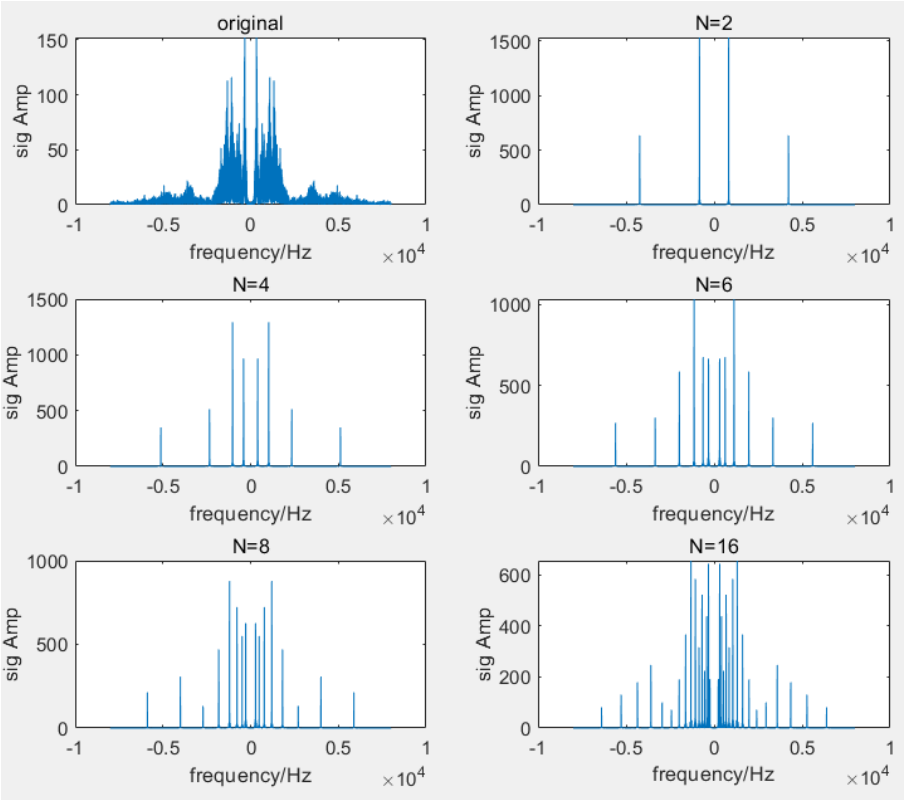
-Set LPF cut off frequency to 50 Hz.

-Implement tone vocoder by changing the number of bands to N=2, N=4, N=6, N=8, and N=16.

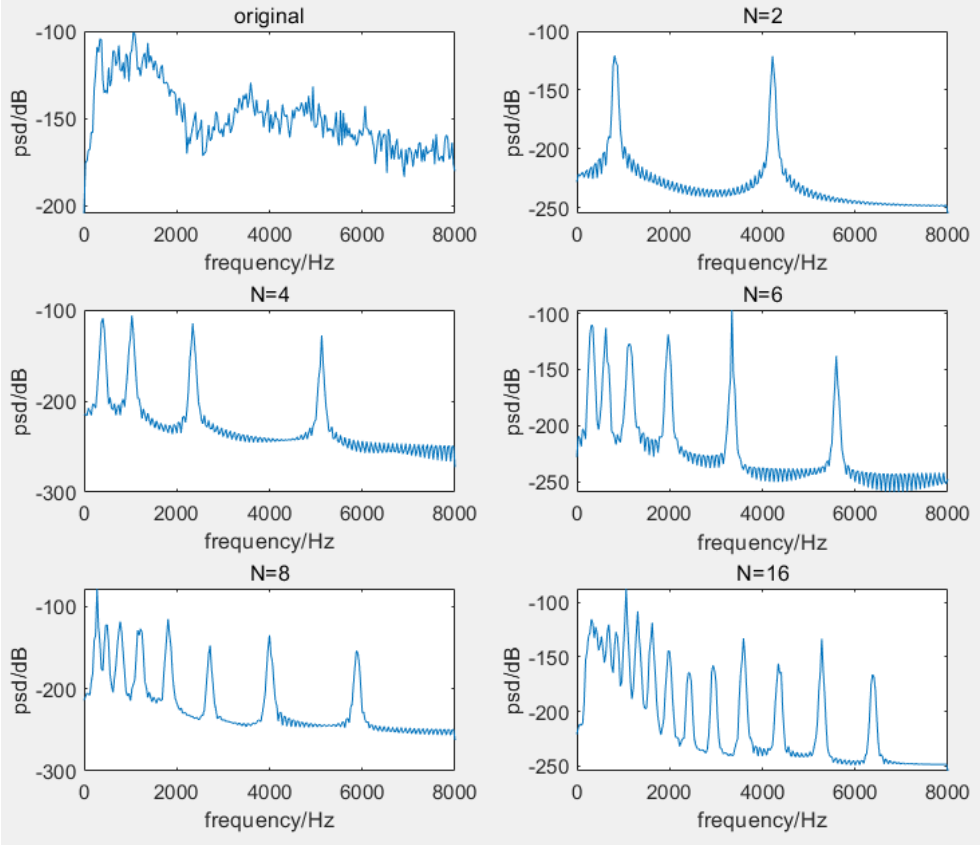
-Describe how the number of bands affects the intelligibility of synthesized sentence, and compare findings with those obtained

Task Result:

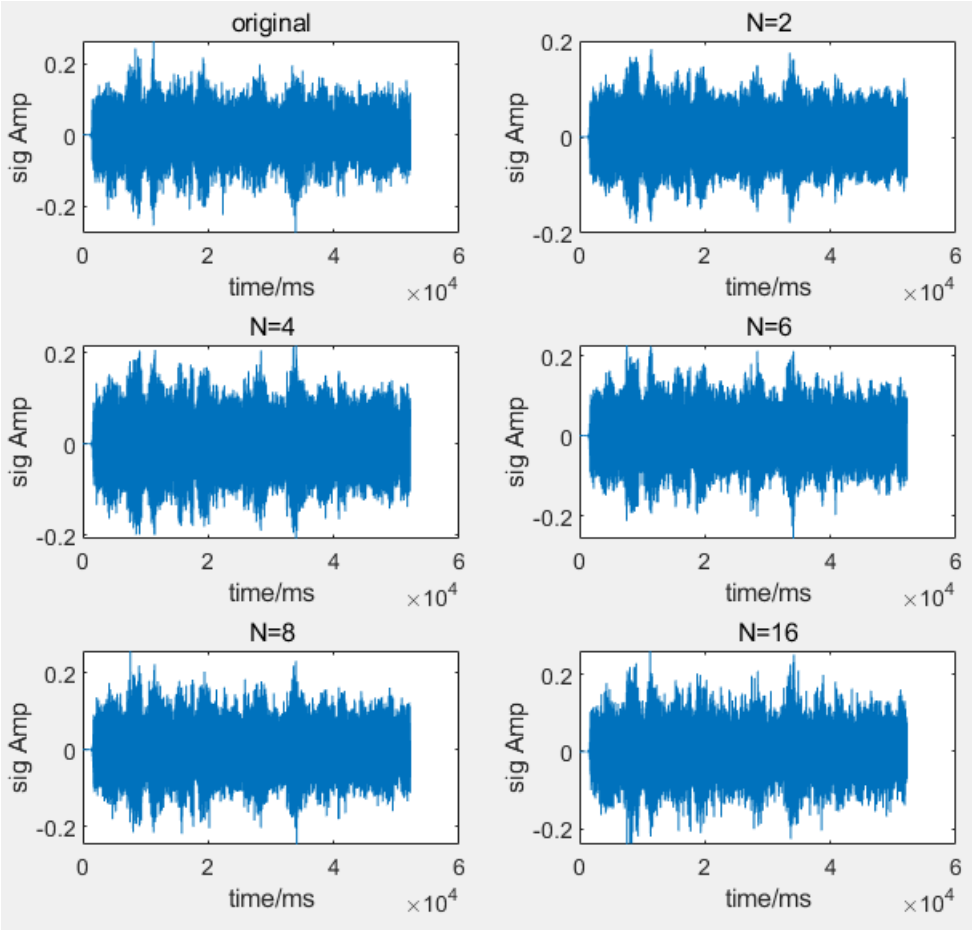
Figure(1): C_01_01.wav in frequency domain:



Figure(2): Power Spectral Density of C_01_01.wav:0



Figure(3):



Analysis:

1. When listening to the output audio, we found that when N = 2, 4, 6, we could hardly hear any human voice; when N = 8, we could vaguely hear someone talking; when N = 16, we could already hear what people were saying

2. It can also be seen from FFT and PSD images that FFT and PSD are closer to the original value with the increase of N

3. According to the experimental results, cut-off frequency = 50Hz, 2<=N<=16, the overall readability of the sound gradually increases with the increase of N, and gradually tends to the original signal

4. Due to the addition of speech spectrum noise, the final signal is not as readable as task1's

Task 4 (徐建辉)

Generate a noisy signal (summing clean sentence and SSN) at SNR -5dB.

Set the number of bands to N = 6

Implement tone-vocoder by changing the LPF cut-off frequency to 20Hz, 50Hz, 100Hz and 400Hz.

Describe how the LPF cut-off frequency affects the intelligibility of synthesized sentence.

Answer:

1. First, we generate SSN. We get the PSD of the original signal and then we use fir2 function to get a filter of it. Using this filter, we process a white noise signal to obtain the SSN. The

PSD of these signals are shown in Fig. 4(a) and the wave of SSN is plotted in Fig. 4(b)

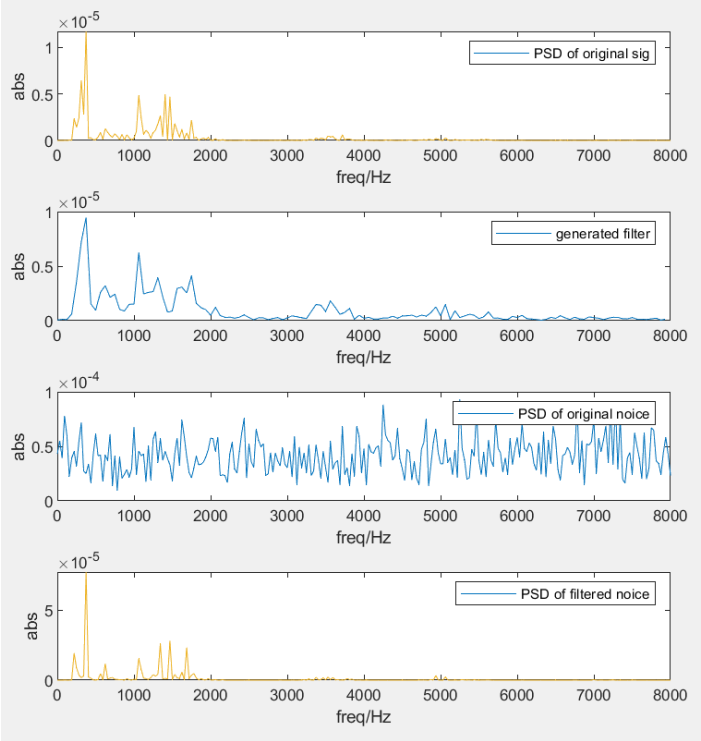


Fig. 4(a)

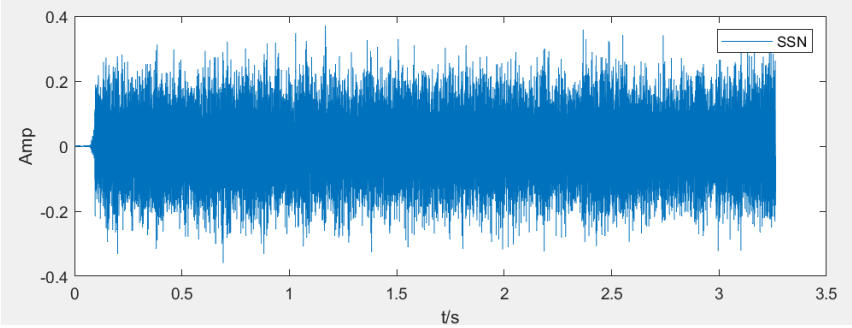


Fig.4(b)

2. Then we set N = 6. We visualize the filters as Fig.4(c)

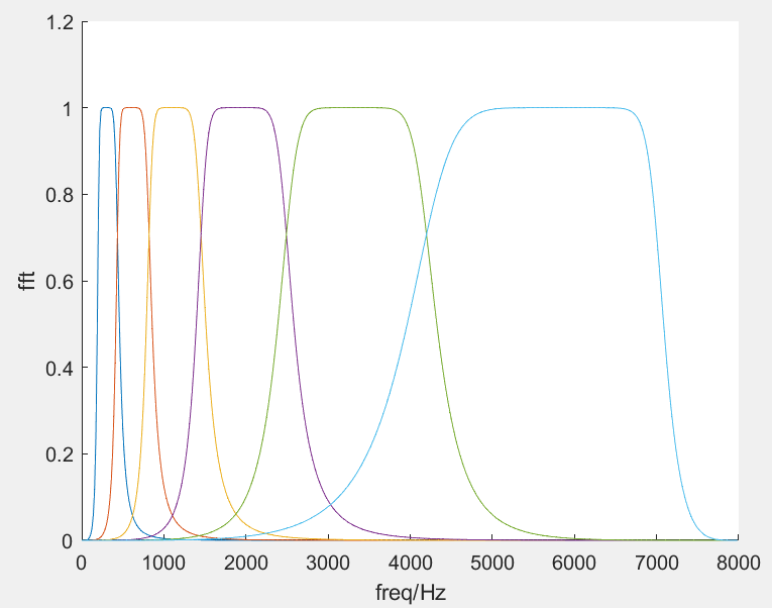


Fig.4(c)

3. The fft of each signal with different frequency is shown as Fig.4(d). We can see that more noise is introduced with the cut-off freq increases.

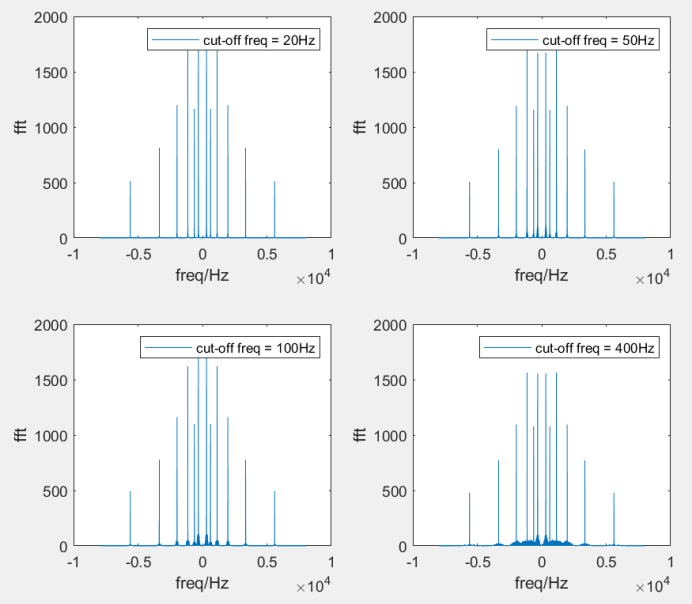


Fig.4(d)

Wave function is plotted in Fig.4(e). It's hard to tell difference between them using human eyes.

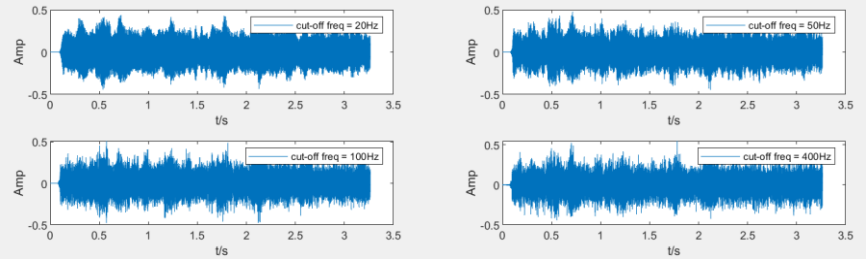


Fig.4(e)

4. After listening to the sound. We find that the sound the nearly the same with little difference

that the sound will be clearer when the cut-off freq increases from 20Hz to 100Hz and be noisier when the cut-off freq increases from 100Hz to 400Hz.

We analyze the phenomena and conclude that:

1. Too high cut-off frequency passed too much detailed high-frequency information, making the result somewhat noisy.
2. While too slow cut-off frequency only passes quite low-frequency signals, losing much of the detailed information.

Finally, we choose cut-off freq = 100Hz as our proper setting.

Experience

We learn to use Matlab to simulate low-pass filters, and test different parameters to determine the effect of different filters.

Through self-study, we have mastered a lot of grammar, functions, etc. other than the content of the class, and have a deeper understanding of the structure of the loop.

The project is a 4-person group project that helps to increase the team spirit of the group members.