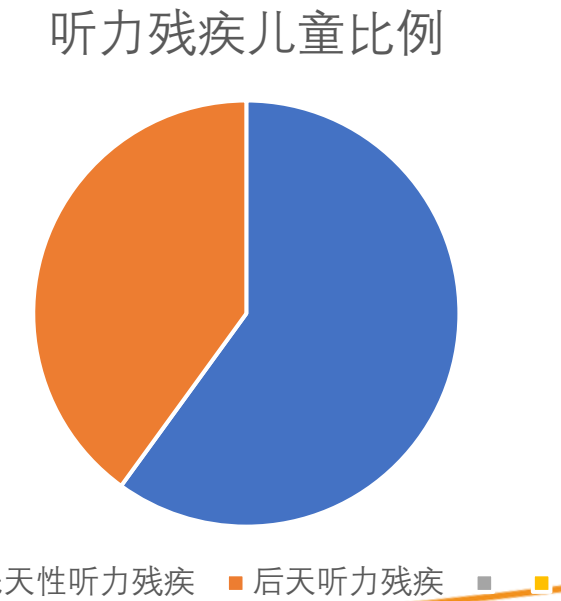
信号与系统Project 1
Speech synthesis and perception with envelope cue

12012305张子尚 12010514刘子羽
12010508华羽霄 12112807徐建辉

background

- 中国残疾人联合会2019 年资料显示，我国现有听力障碍的残疾人2,780 万人，其中0-6 岁的听力残疾儿童约有13.7 万人，每年新生听障儿童2-3 万人。在我国听力残疾人群中，约有60%是因为遗传基因缺陷而引发的耳聋



background

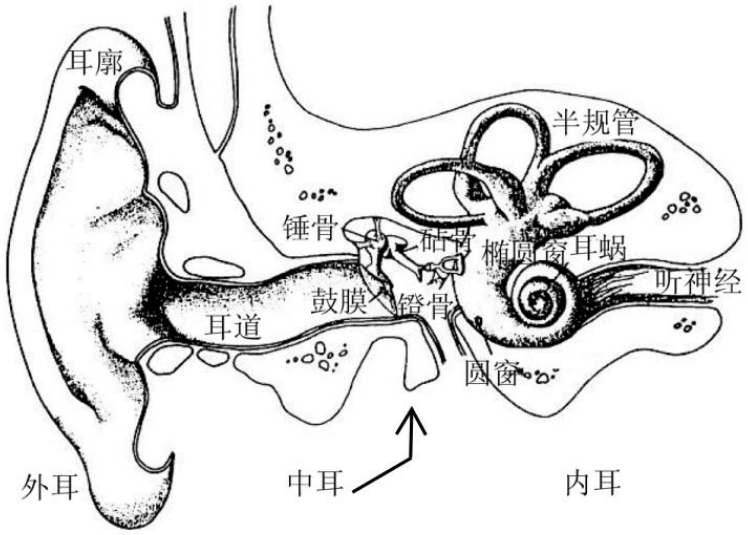


图 1　外周听觉系统示意图

Fig.1　Illustration of the structure of the peripheral auditory system. Redrawn from Moore(2007)

图 3　人工耳蜗总体结构框图

Fig.3　A block diagram of the overall structure of the prosthesis. Redrawn from Clark 1978.

background

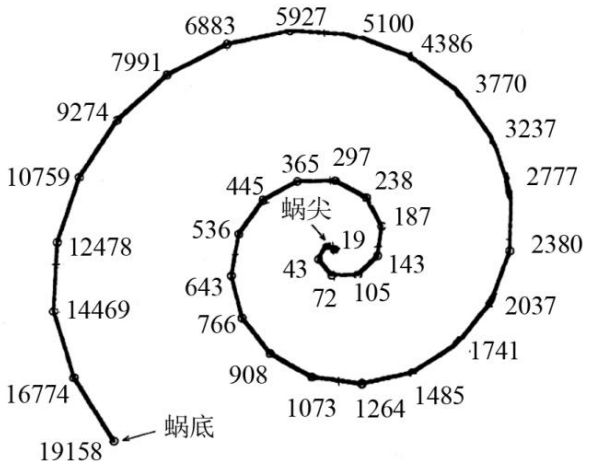
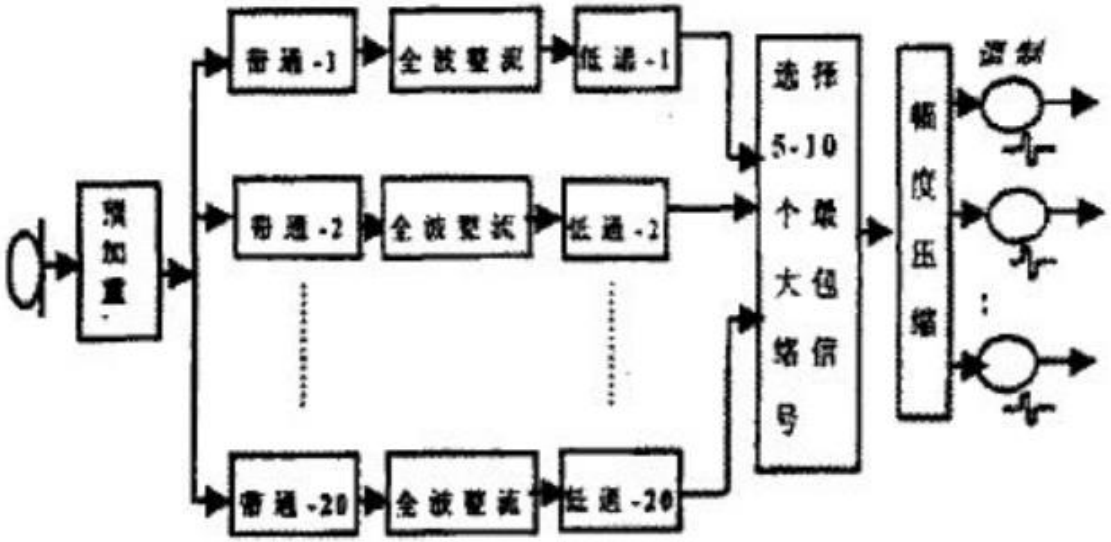


图 2　耳蜗感音位置理论示意图(单位为 Hz)

Fig.2　Diagram of the basilar membrane showing the base and apex. The position of maximum displacement in response to sinusoids of different frequency (in Hz) is indicated.



SPEAK方案

项目简介

- 利用matlab软件进行语音信号的低通滤波仿真处理，并保存处理后得到的音频。
- 除完成给定的参数条件以外，调试其他的阶数和截止频率，寻找滤波效果与这两者之间的关系。
- 寻找一个合适的阶数和截止频率的范围，使得处理后的信号与原始信号最为接近。
- 一些其他的。

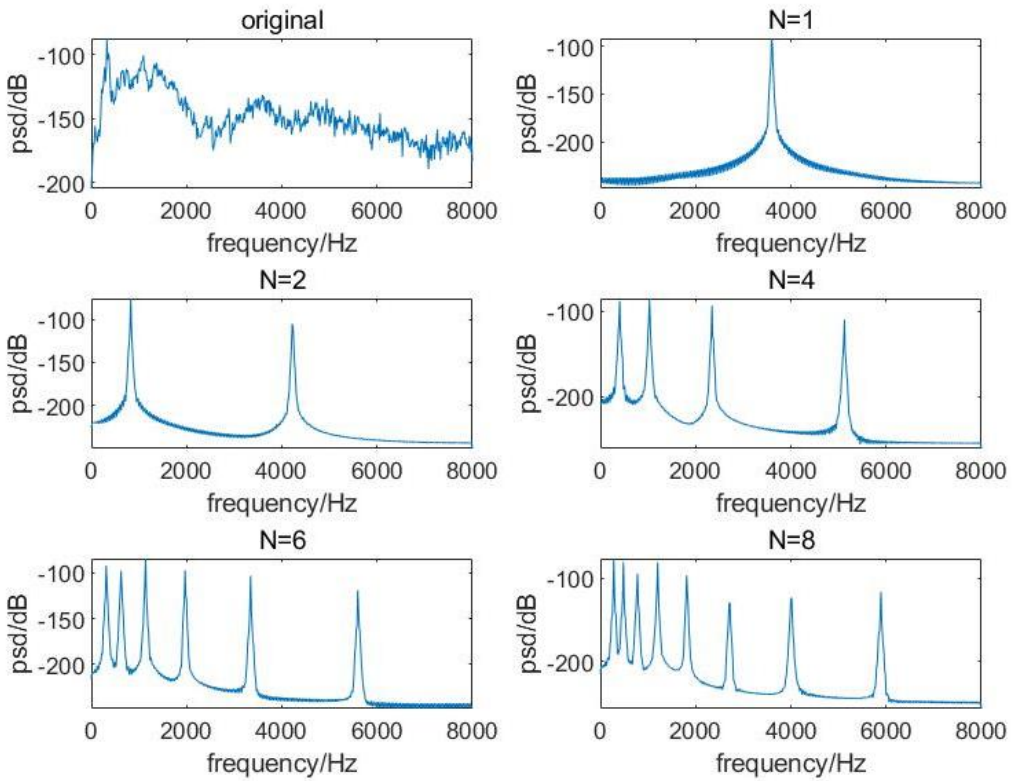
小组分工

- 张子尚:Task1+N频段的项目外的拓展和critical thinking
- 华羽霄:Task2+introduction，experience，团队组成和贡献
- 刘子羽:Task3+查项目背景
- 徐建辉:Task4+cutoff频率的项目外的拓展和critical thinking

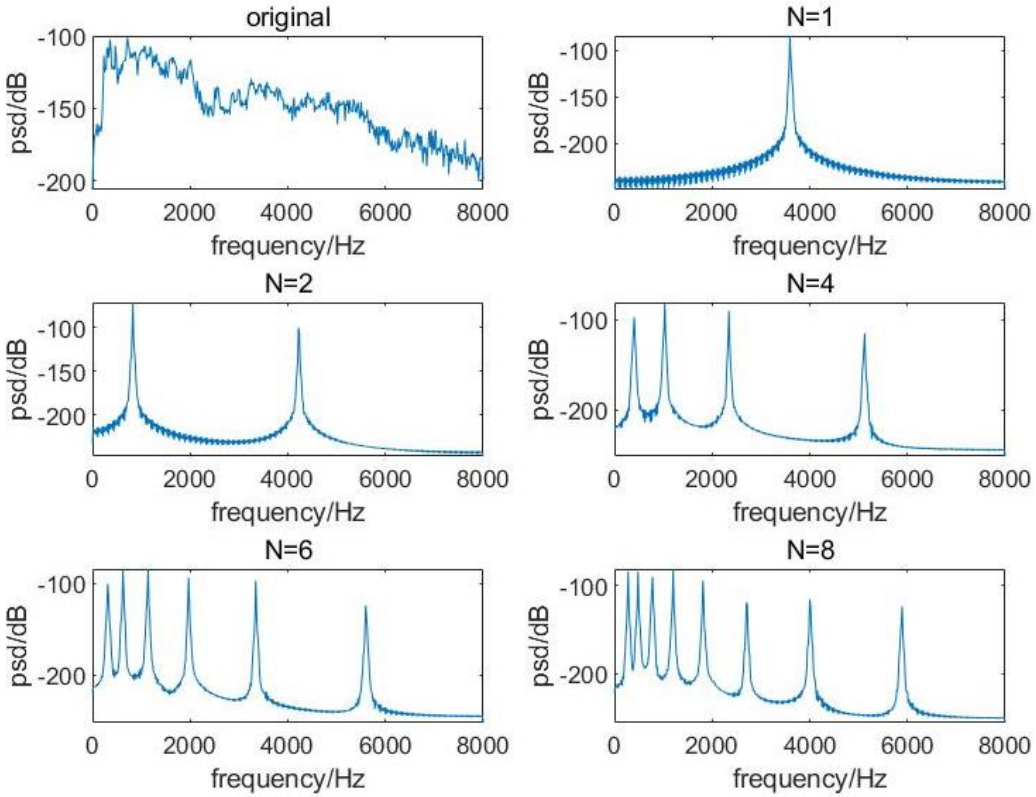
task1

- 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.

C_01_01.wav & C_01_02.wav Power Spectral Density

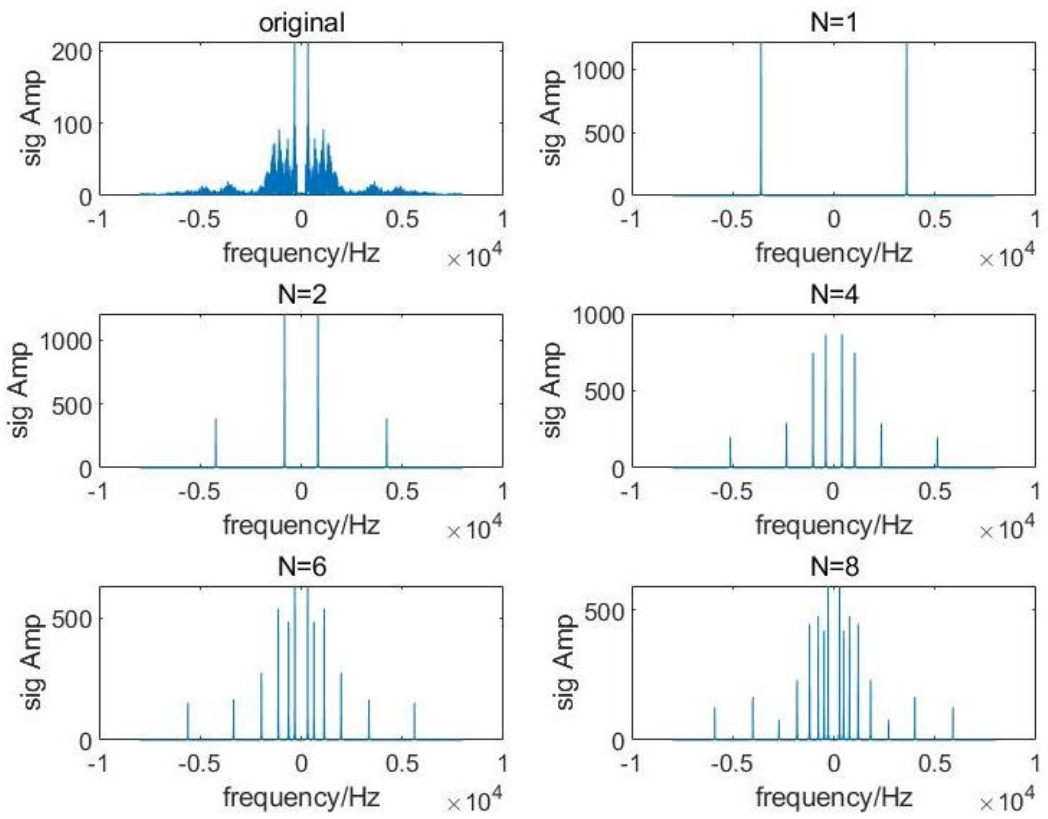


C_01_01.wav

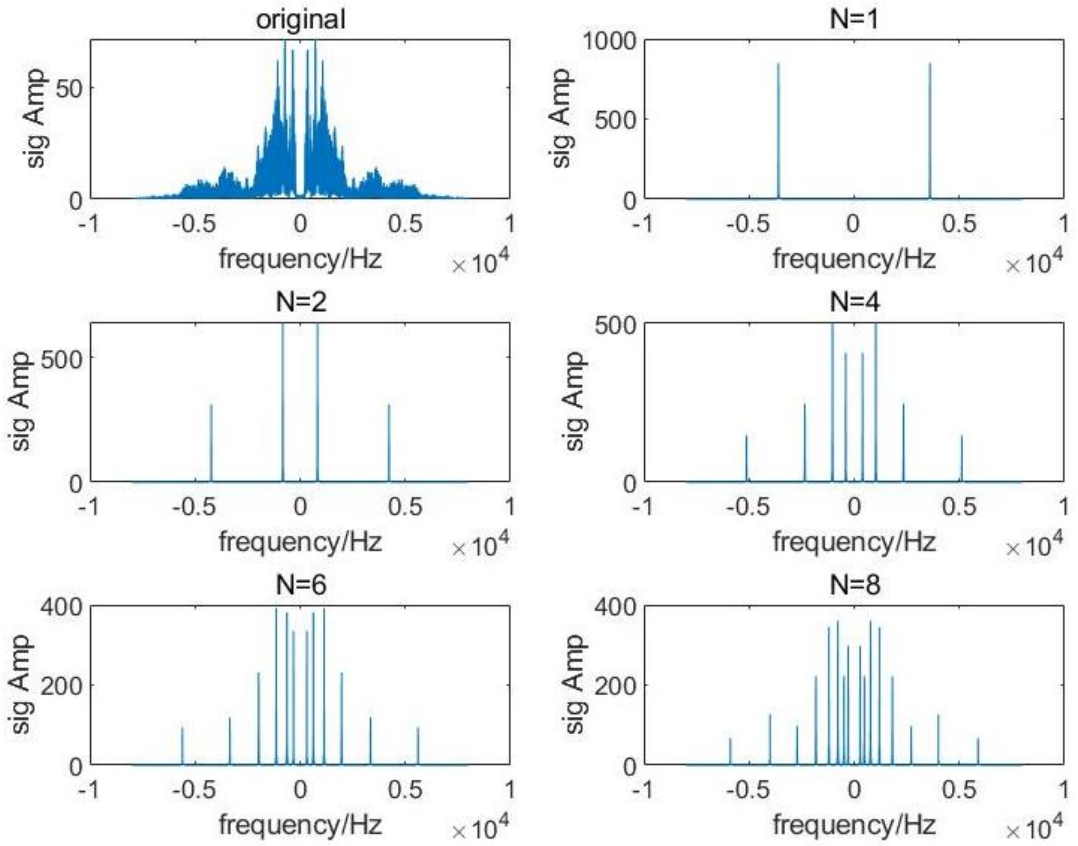


C_01_02.wav

C_01_01.wav & C_01_02.wav Graphs

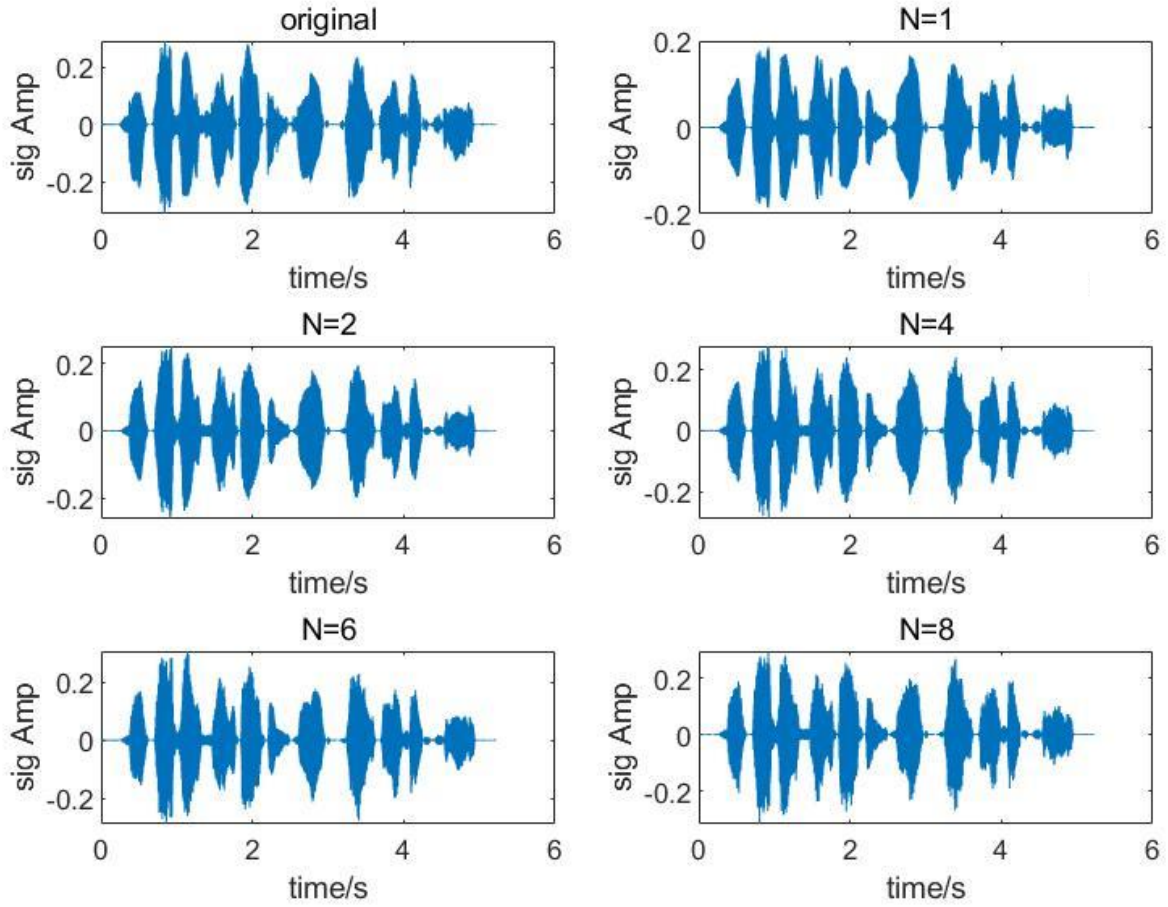


C_01_01.wav



C_01_02.wav

C_01_01.wav Graph （cut-off frequency to 50 Hz）



Original

N=1

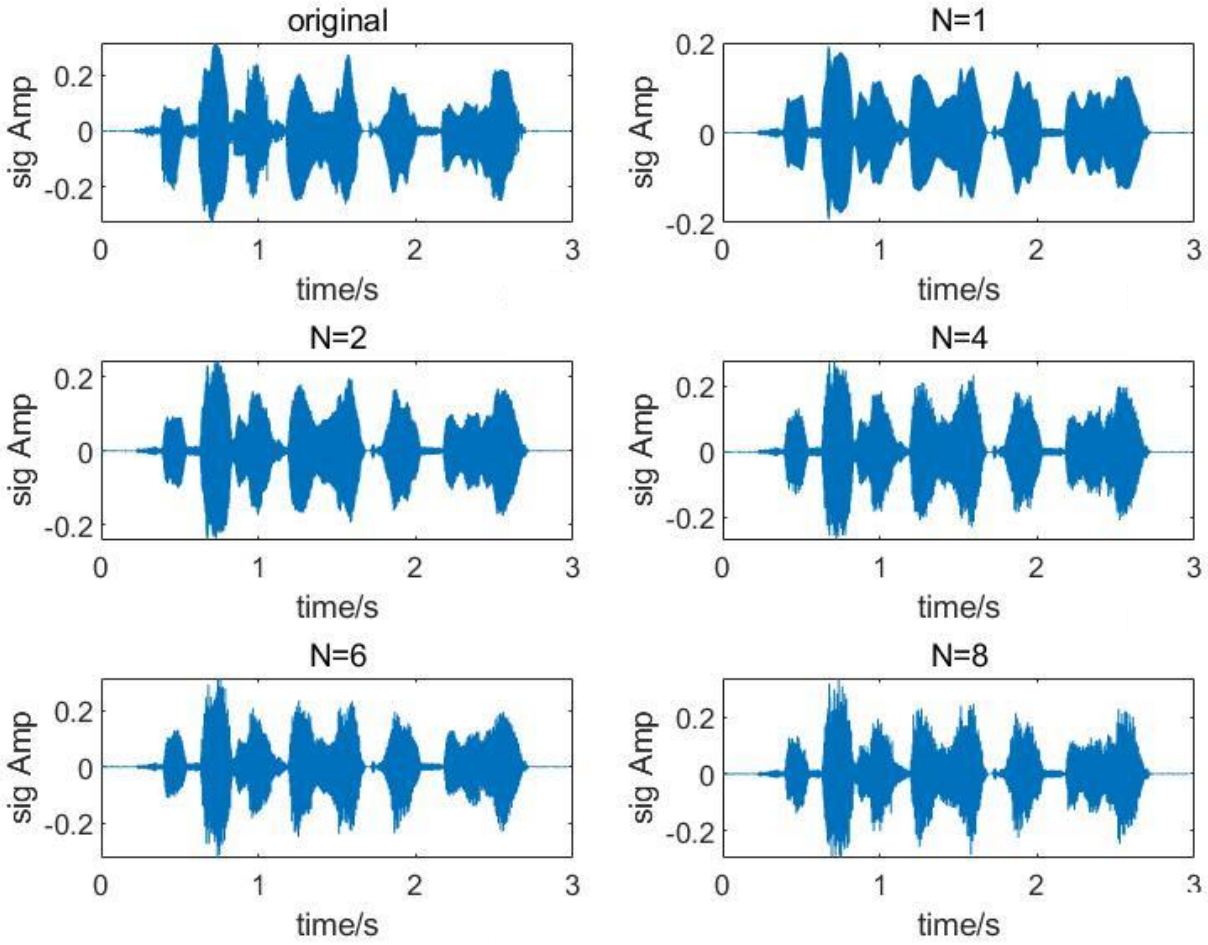
N=2

N=4

N=6

N=8

C_01_02.wav Graph （cut-off frequency to 50 Hz）



Original

N=1

N=2

N=4

N=6

N=8

Analysis

- 对频段N为不同值时的PSD图、频域响应图以及波形图分析可以看出，随着频段数目的增加，处理后的图像与原信号图像更加接近。同时随着频段N的增加，音频的可读性也逐步增加

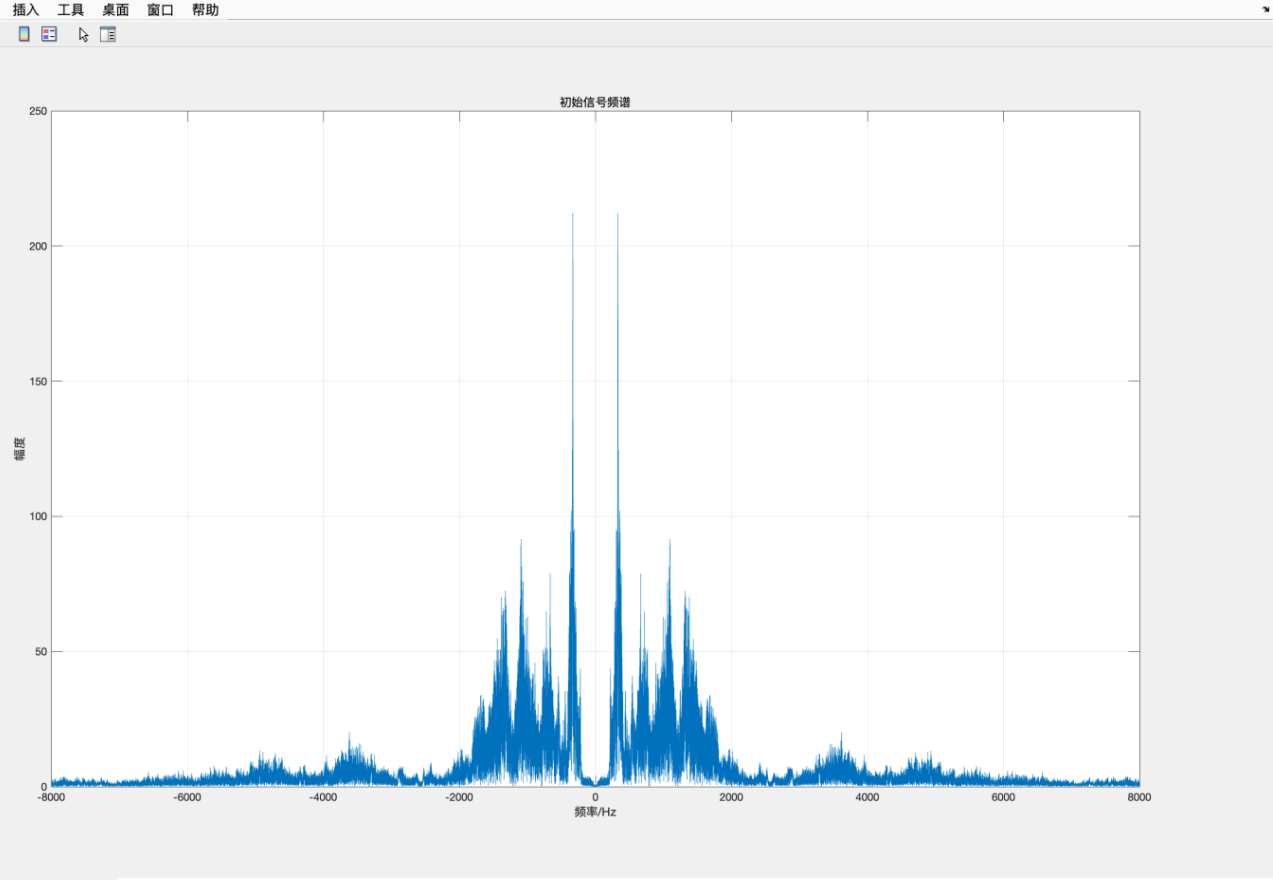
- 在task1中，我们选取的N从1，2，4，6一直变化到8时，尽管我们能够逐渐更加理解音频的内容，但仍然只能听到一个模糊的音频，无法具体听清楚其准确的信息。

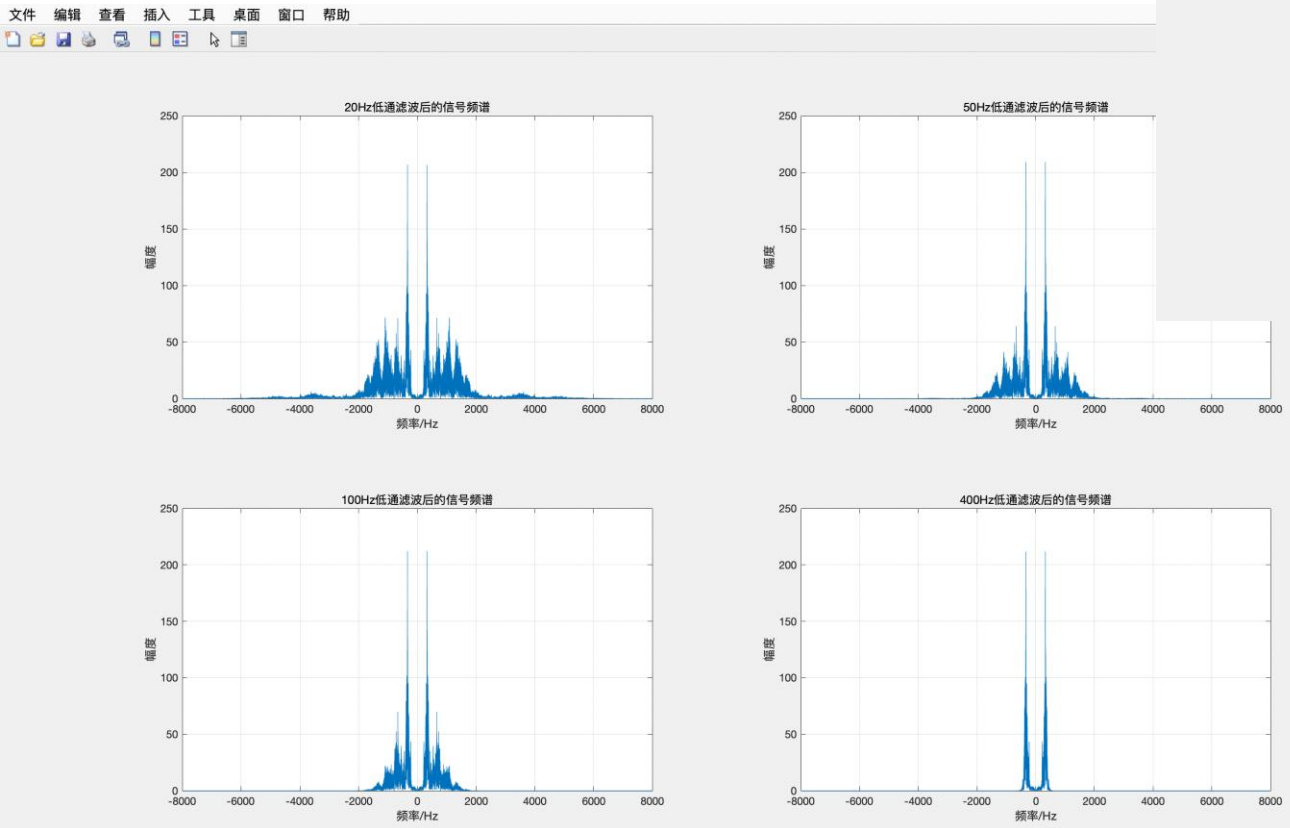
- 猜想：在task1中，随着N的增大，音频的可读性有了明显提升。但N所选取的最大值为8仍较小。所以音频的不清晰可能是由于N的值太小所致。我们猜想：在一定范围内，增大频段N可以提升处理后音频的可读性。

task2

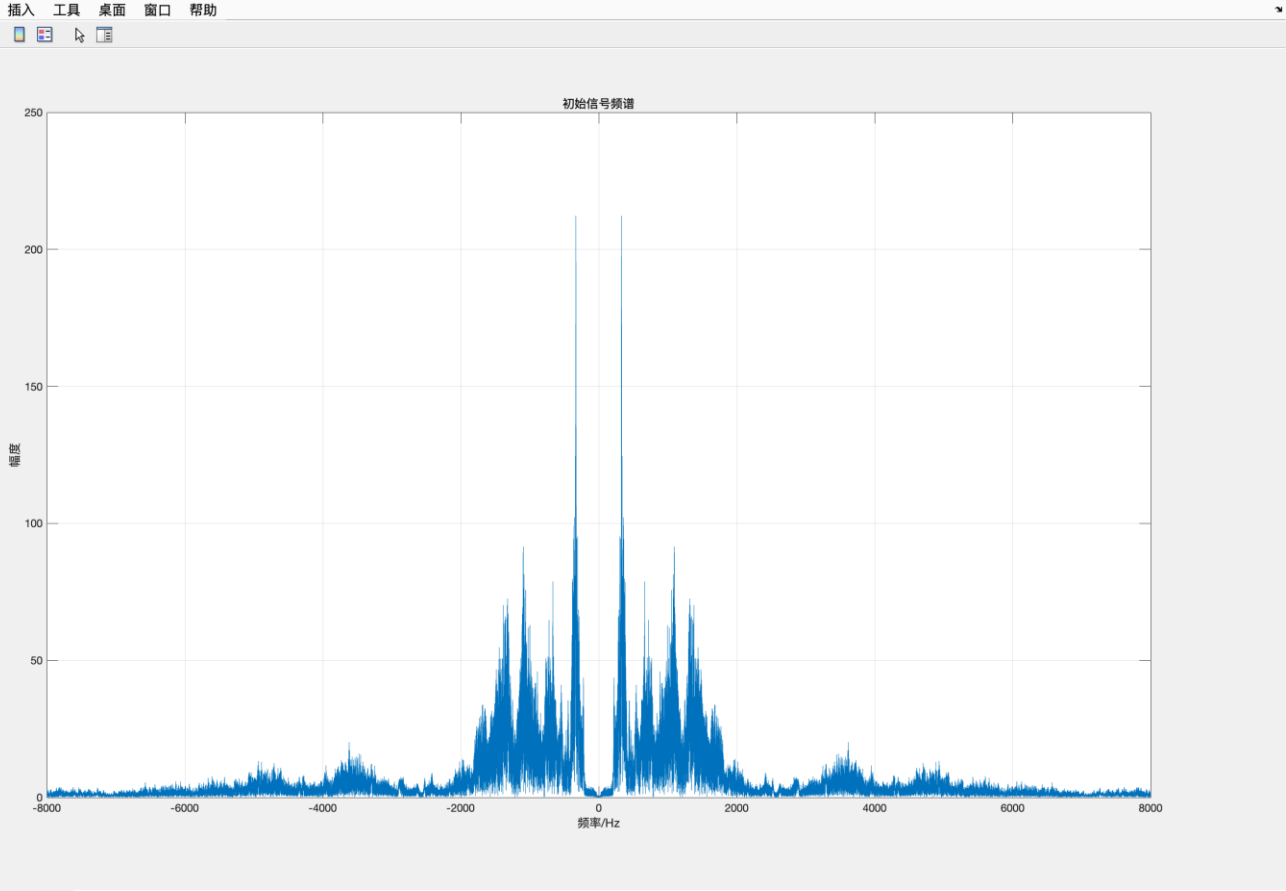
– Set the number of bands N=4.
– Implement tone-vocoder by changing the LPF cut-off frequency to 20 Hz, 50 Hz, 100 Hz, and 400 Hz.
– Describe how the LPF cut-off frequency affects the intelligibility of synthesized sentence.

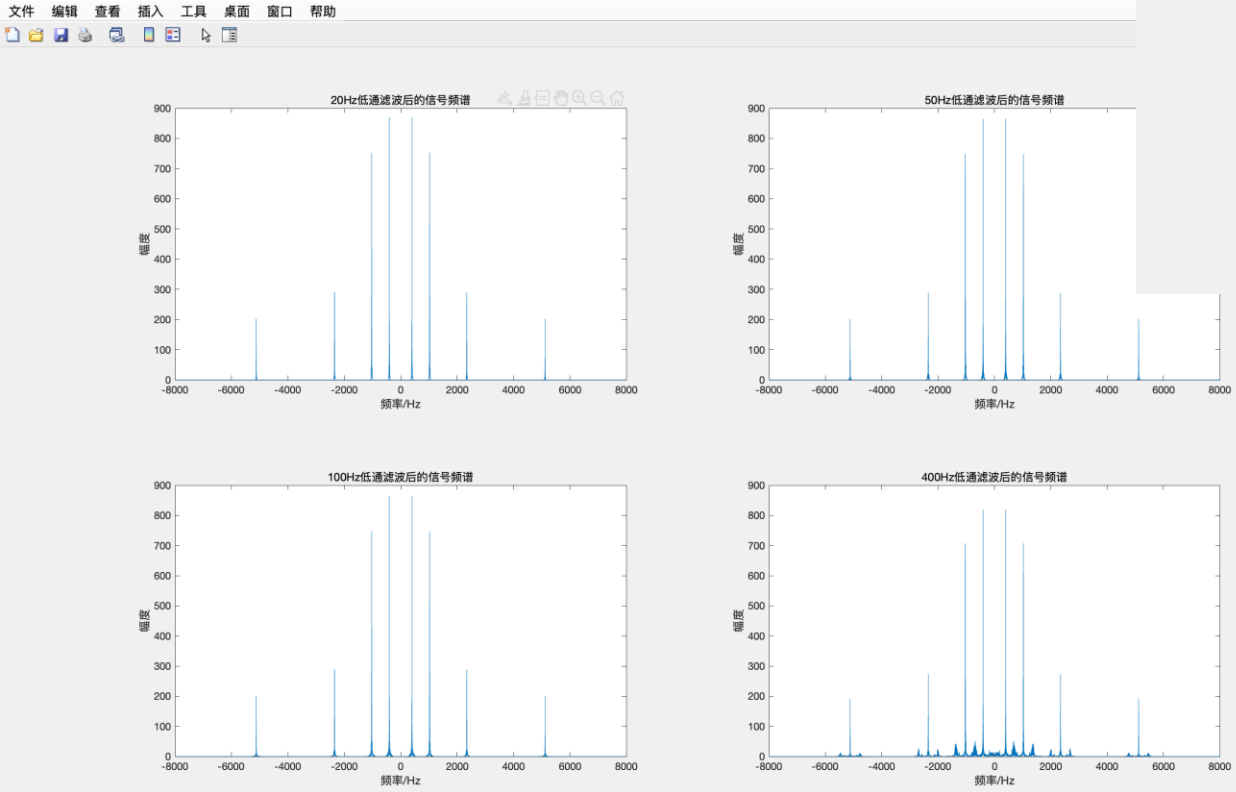
原始信号与低通滤波后的信号频谱



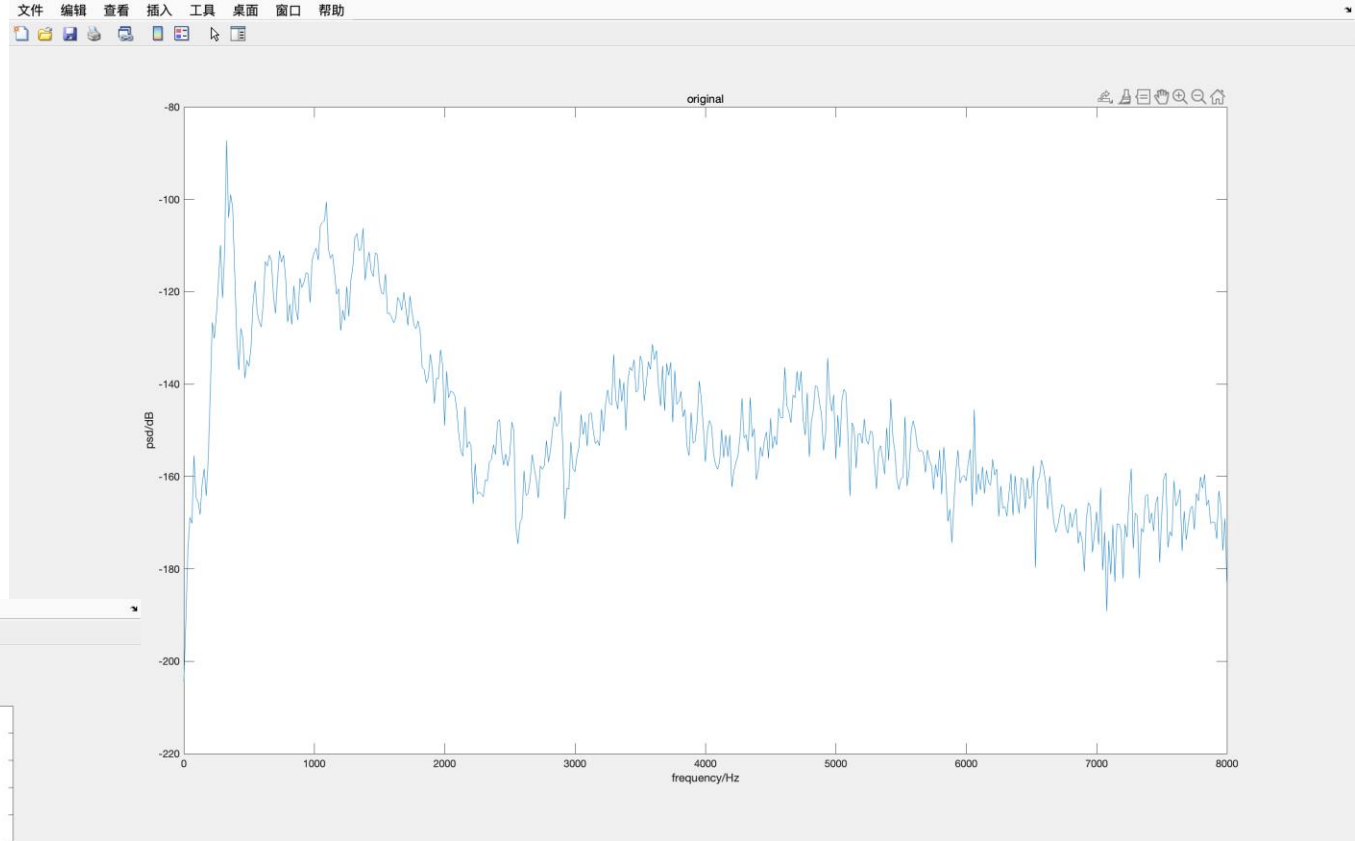


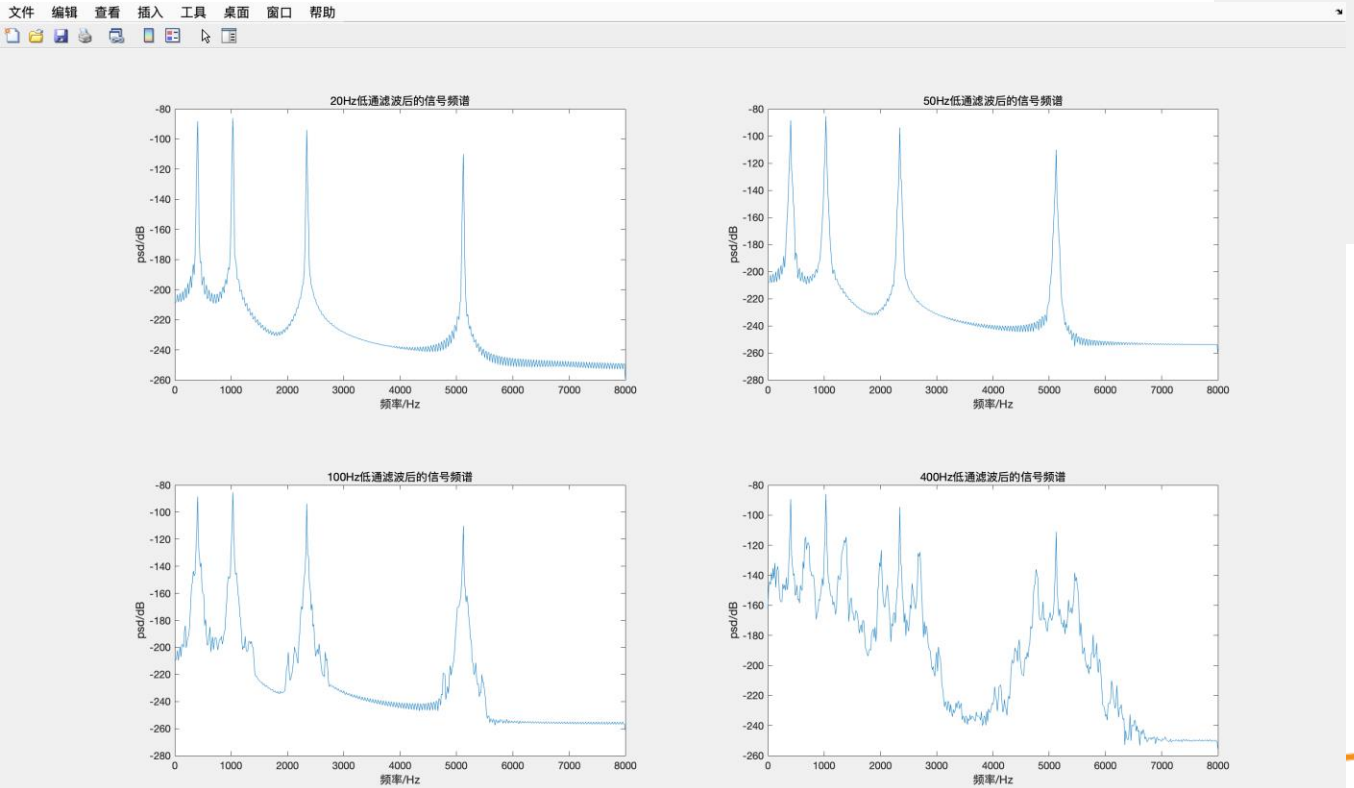
原始信号与低通滤波后的信号的fft



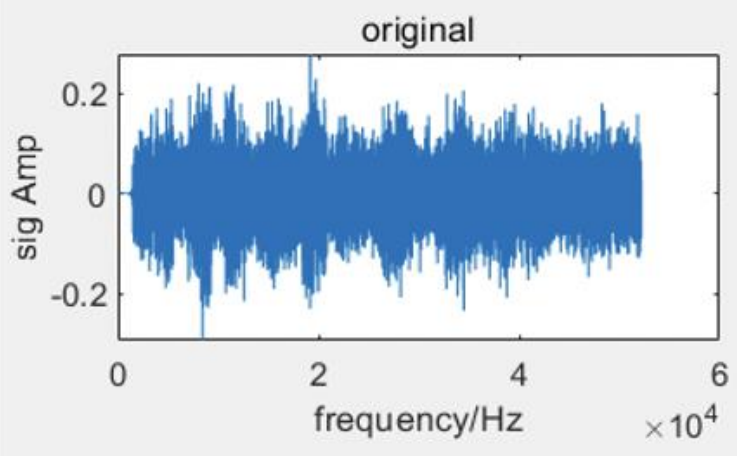


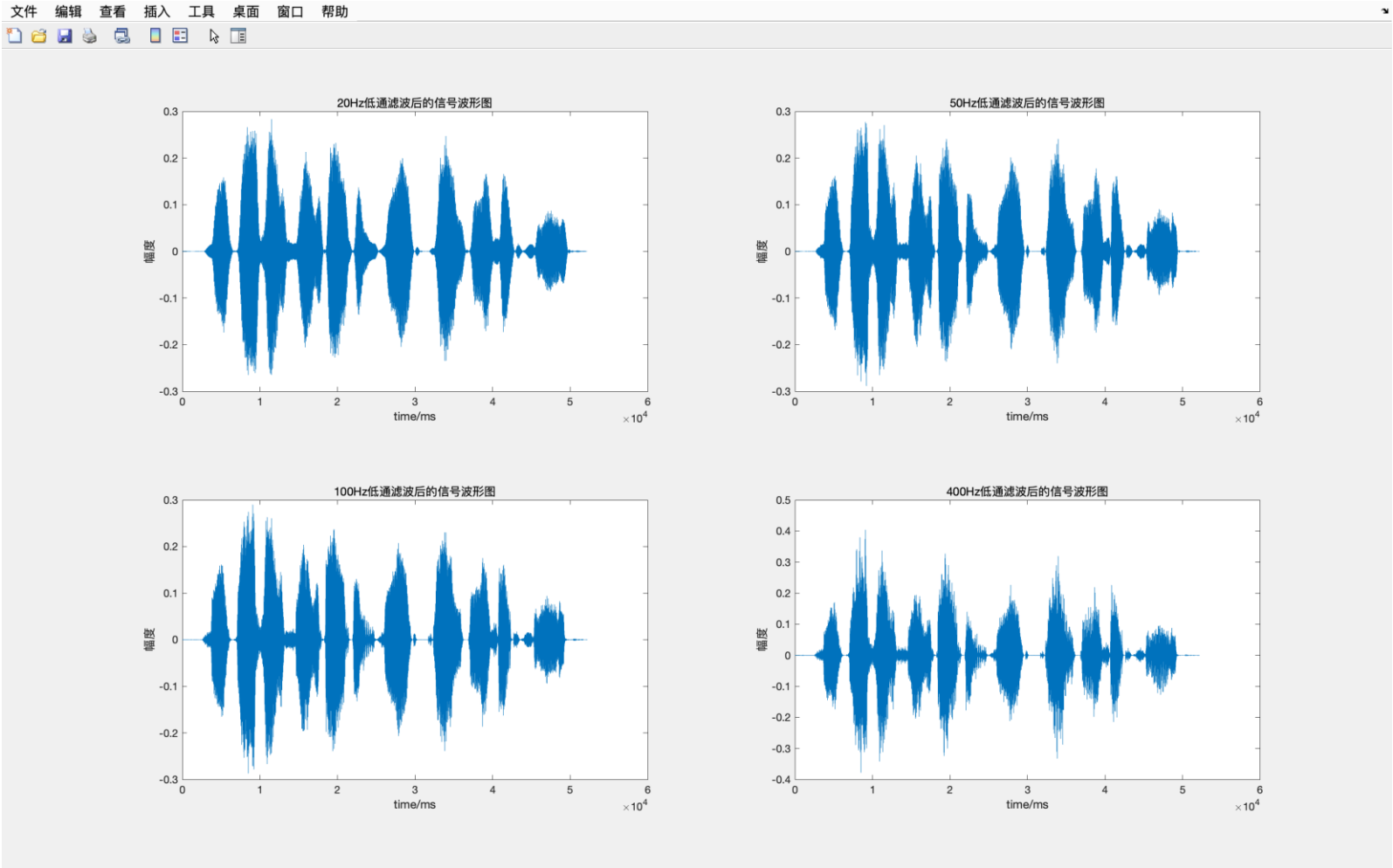
原始信号与低通滤波后的信号包络





波形图对比


original
sig Amp
0.2
0
-0.2
0
2
4
6
frequency/Hz
×10^4


文件 编辑 查看 插入 工具 桌面 窗口 帮助
20Hz低通滤波后的信号波形图
50Hz低通滤波后的信号波形图
100Hz低通滤波后的信号波形图
400Hz低通滤波后的信号波形图
幅度
time/ms
×10^4

语音信号处理效果试听

原始信号

20Hz低通滤波后的信号

50Hz低通滤波后的信号

100Hz低通滤波后的信号

400Hz低通滤波后的信号

注：以上结果均在N=4的条件下测得

Critical thinking

不难发现，当截止频率处于20Hz到400Hz之间时，语音信号处理效果非常差。于是，我扩展了截止频率的上限到7200Hz，每隔1200Hz进行一次检测，得到的结果如下：

1200Hz低通滤波后的信号

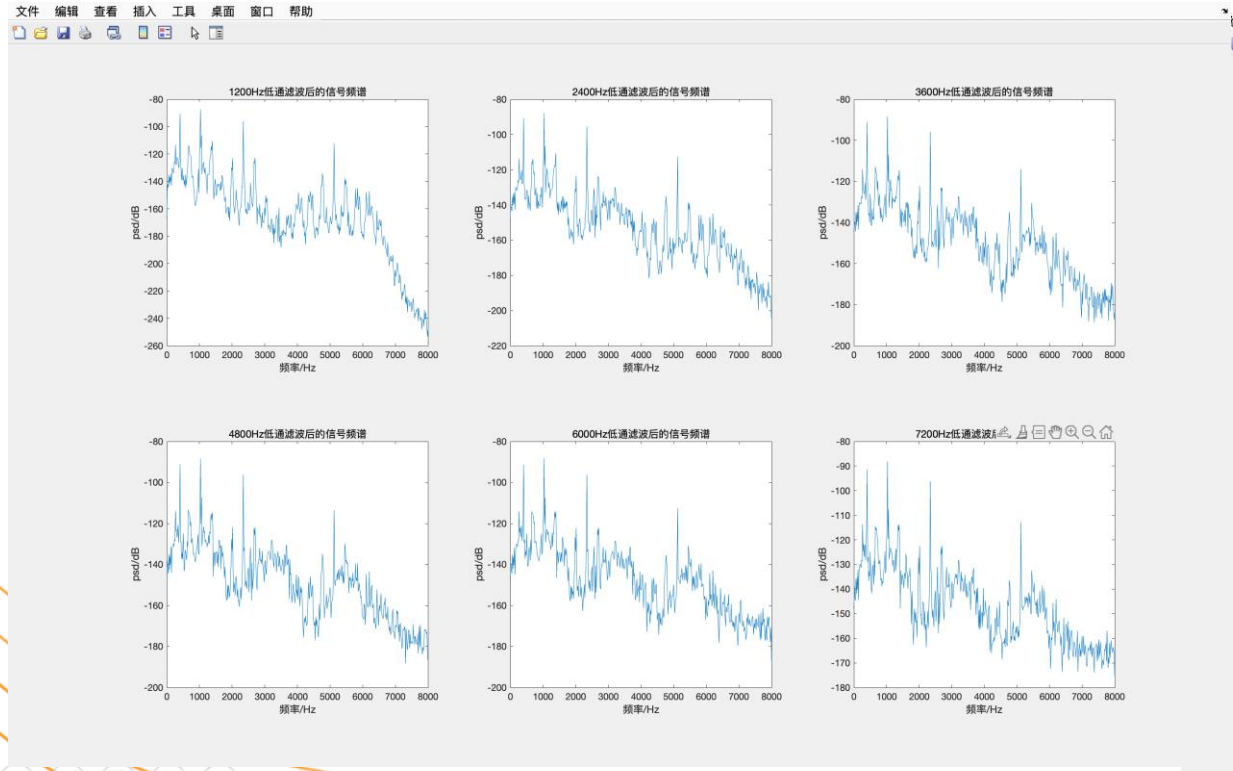
2400Hz低通滤波后的信号

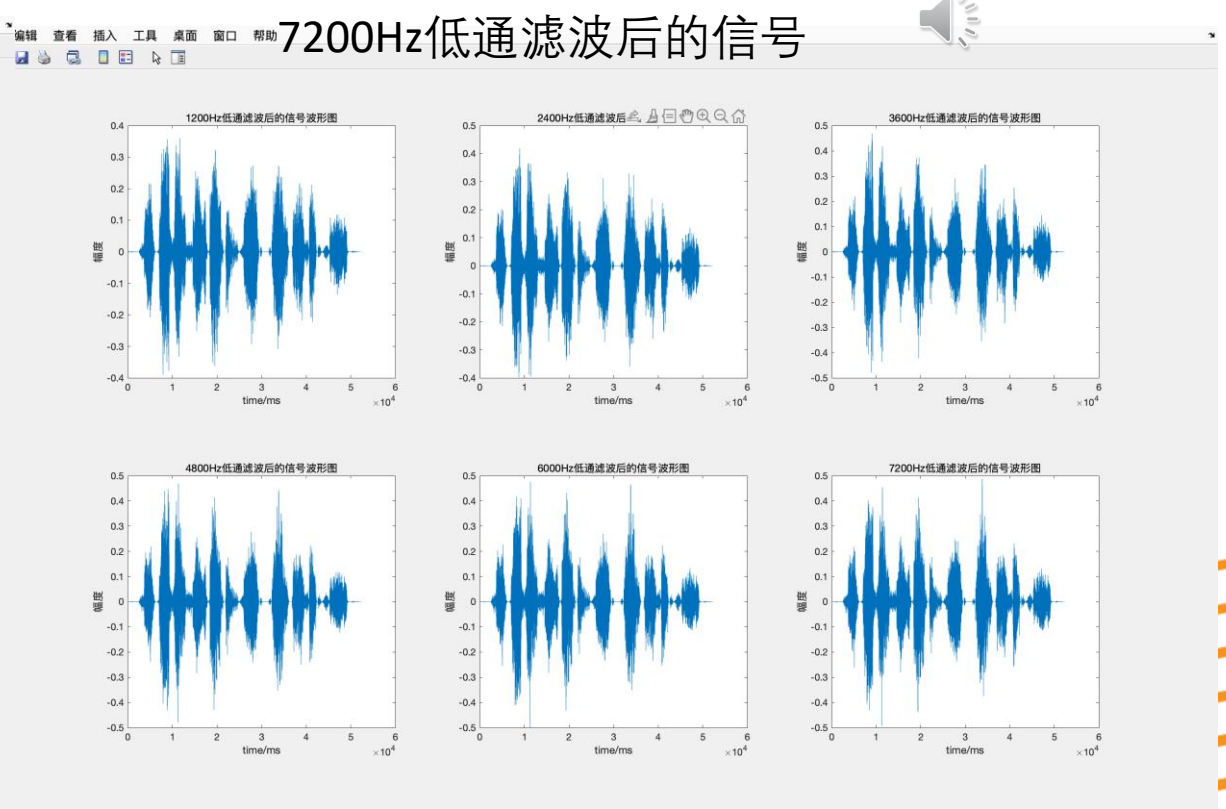
3600Hz低通滤波后的信号

4800Hz低通滤波后的信号

6000Hz低通滤波后的信号

7200Hz低通滤波后的信号





总结

- 当N=4时，随着截止频率的降低，滤波效果越显著，也就是说能够听到的频率段越少，语音中的信息越不容易被识别出。
- 当截止频率低于400Hz时，如果不是事先知道语音信息的内容，几乎无法识别出滤波后的语音信息。
- 随着截止频率的不断增加，滤波效果逐渐变弱，因此处理后的信号与原始信号的区别也逐渐减小。
- 由于滤波阶数的限制，无论截止频率再怎么增加，得到的结果始终无法与原始信号相媲美。

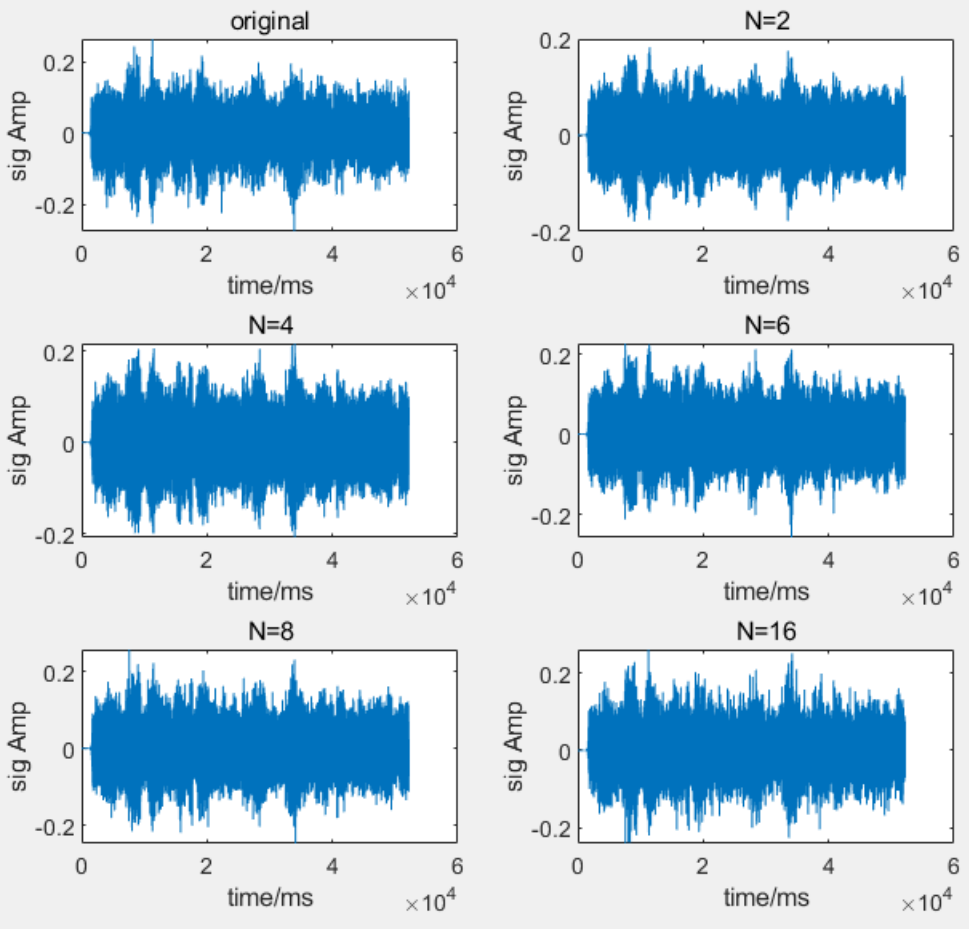
Task 3

- Problem:
  - 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.

Task 3

- Analysis:

Audio wave



Noisy signal

NS N=2

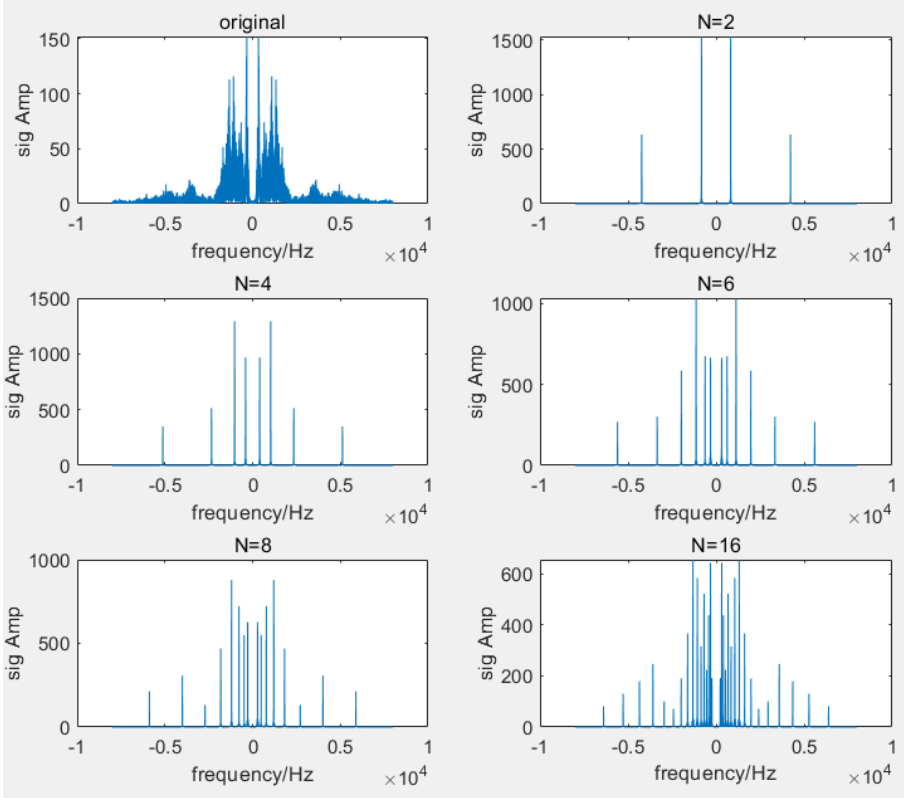
NS N=4

NS N=6

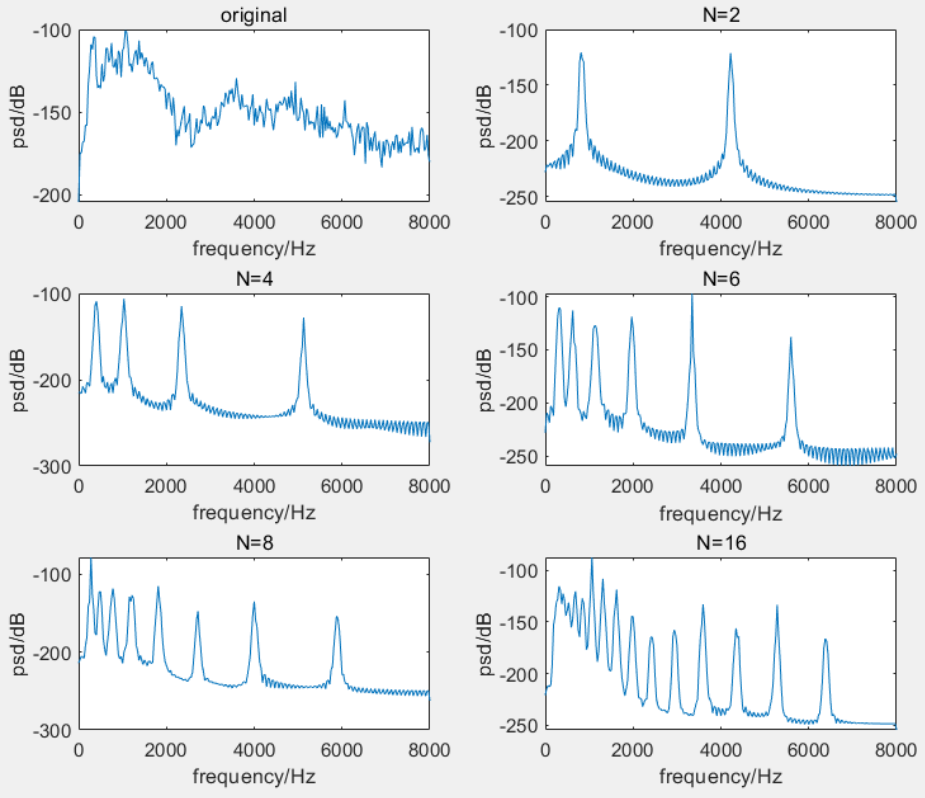
NS N=8

NS N=16

Task 3



FFT



PSD

Task 3

- Analysis:
  - 在试听输出音频时发现，N = 2，4，6时几乎听不出任何人声，N = 8 时隐约听到有人说话，N = 16 时已经可以听出人说的内容了
  - 在FFT 以及PSD图像中也可以看出随着N的增大， FFT 和PSD都更接近原始值
  - 由实验结果得出，在截止频率 cut-off frequency = 50Hz，波段 2 <=N<=16 的范围内，声音整体的可读性随着N的增大逐步提高，逐渐趋于原信号
  - 由于加了言语谱噪声，最终的信号可读性不如task1的好

Task 3 code

- **Generate a noisy signal at SNR -5 dB**

```
[x1, fs1] = audioread("C_01_01.wav");%读取音频
sig1 = x1';
[pxx1,w2]=pwelch(sig1,[],[],512,fs1);%估计功率密度
b1=fir2(3000,w2/(fs1/2),sqrt(pxx1/max(pxx1)));%生成滤波系数
noise1=1-2*rand(1,length(x1));%生成白噪音
SSN1= filter(b1,1,noise1);%生成言语普噪声
E1=norm(sig1);%求模
E2=norm(SSN1);
SSN1=sqrt(E1/E2)*SSN1;%调整信噪比
figure(3)
plot(1:length(SSN1),SSN1)
% SNR=20*log(norm(sig1)/(norm(SSN1)))
y1=sig1+SSN1;%混合声音
y1=y1/norm(y1)*norm(sig1);%调整声音大小
```

Task 3 code

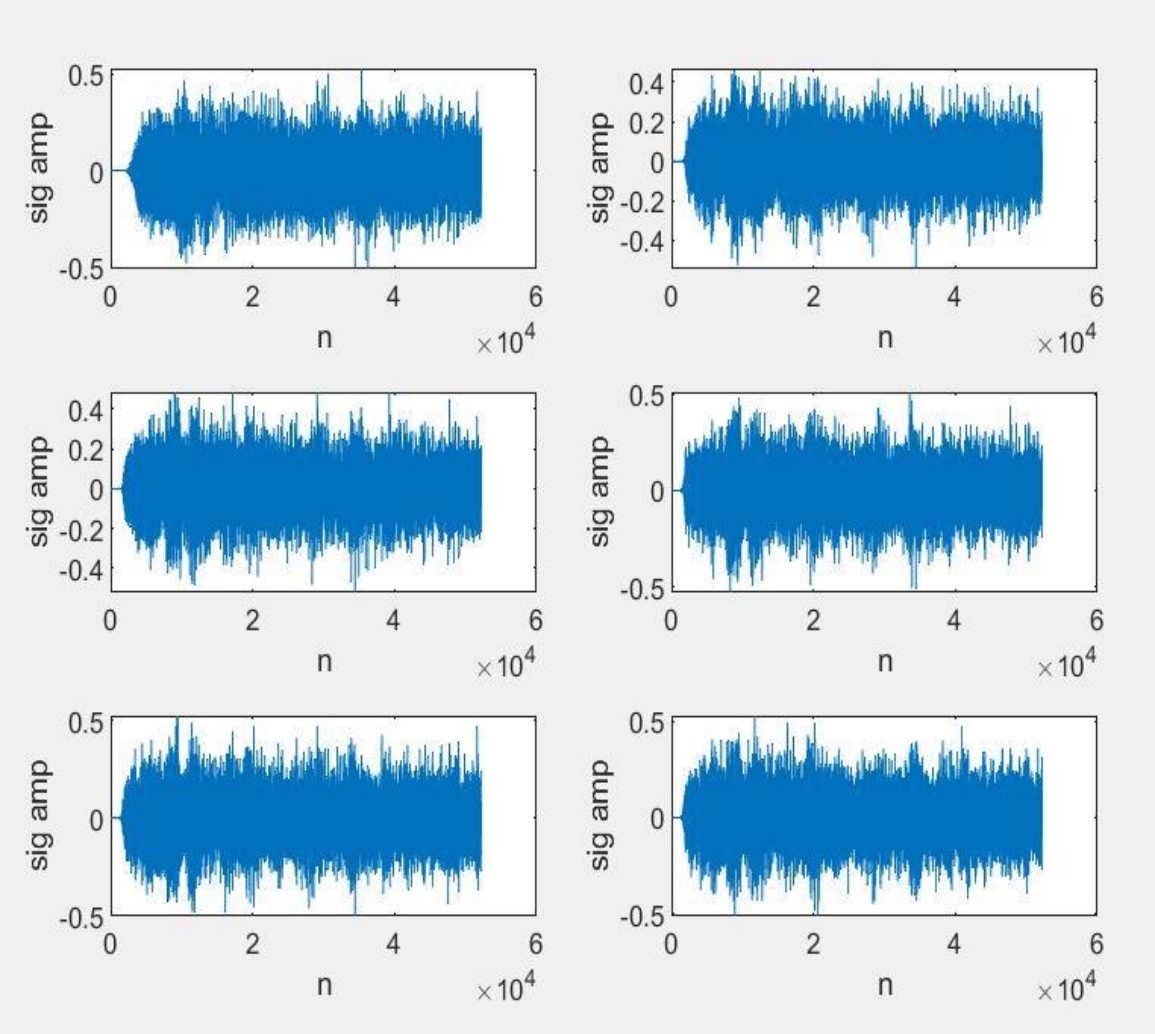
- **Graph & save**

```
% 以N = 2 为例
yg2 = tonevocoder(y1,fs1,50,2); %生成滤波信号
plot(1:length(y1),yg2) %画出波形图
f = fs1*(-length(y1)/2:length(y1)/2-1)/length(y1);
plot(f,fftshift(abs(fft(yg2)))) % 画出FFT图像
[Pxx2,w2] = pwelch(yg2,[],[],512,fs1); % 生成PSD
plot(w2,20*log10(Pxx2)) % 画出PSD图像
audiowrite('NS_N2_f50.wav',yg2,fs); % 将声波存储为.wav
```

task4

- 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.

C_01_02.wav Graph （N=114）



5Hz

20Hz

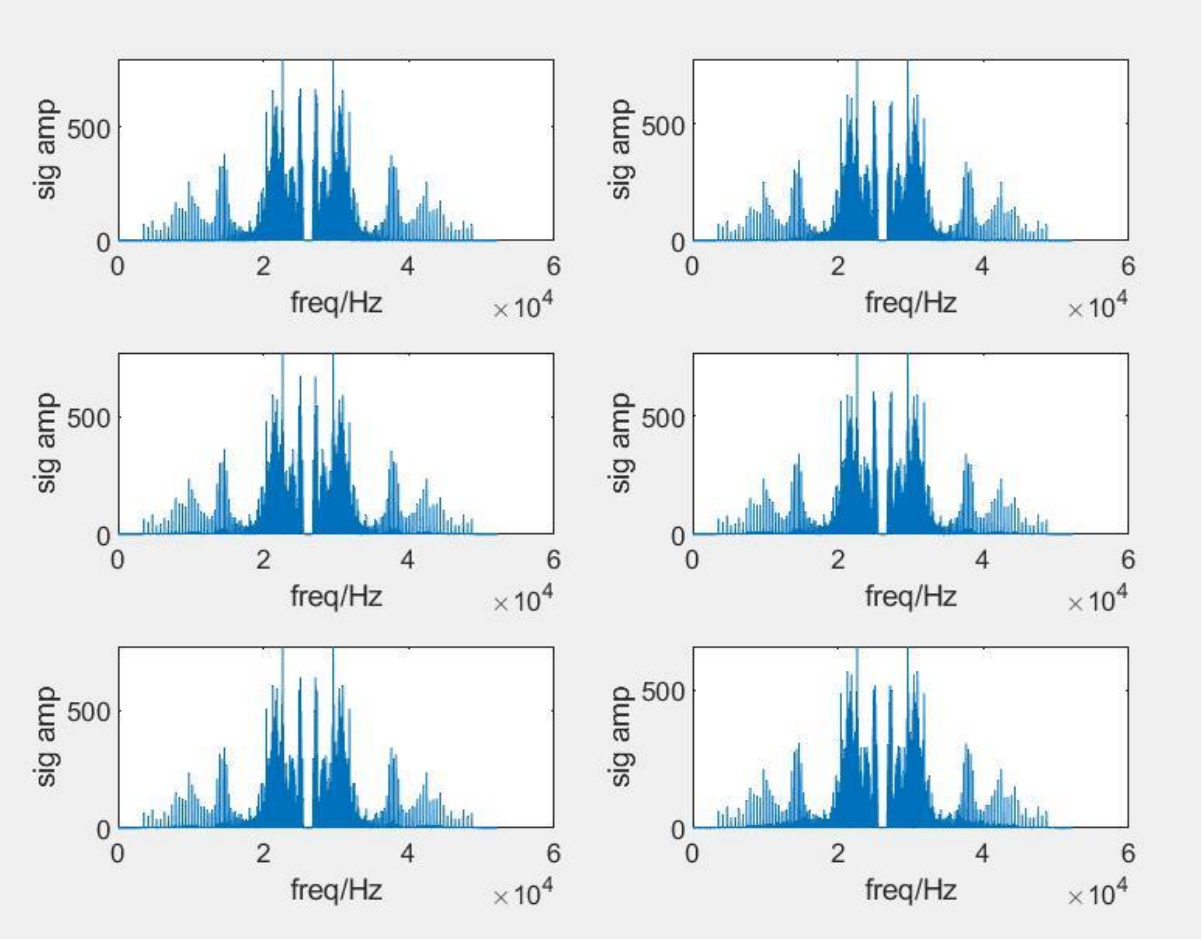
50Hz

100Hz

400Hz

2500Hz

C_01_02.wav Graph （N=114）



Conclusion

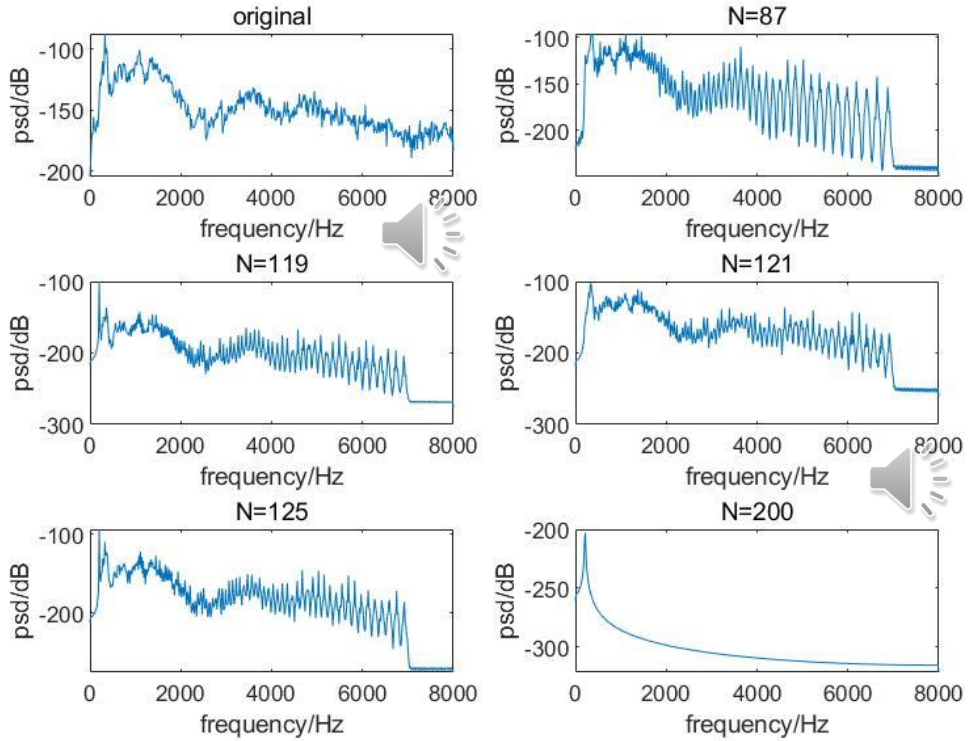
1. Too high cut-off frequency passed too much detailed high-frequency information, making the result somewhat noisy.

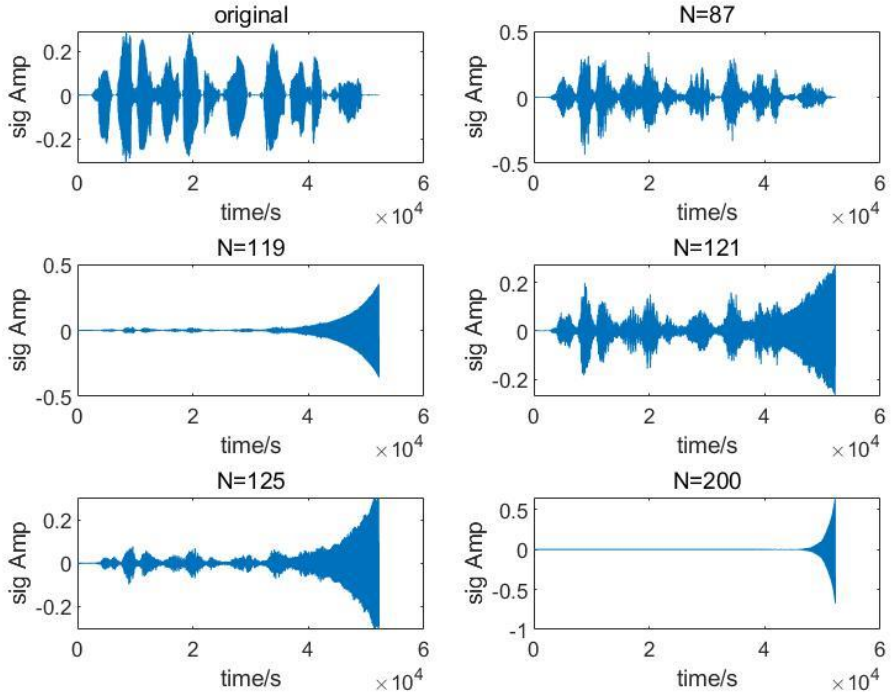
2. While too slow cut-off frequency only passes quite low-frequency signals, losing much of the detailed information.

After weighing the pros and cons, we choose 100Hz as the best cut-off frequency.

Critical Thinking

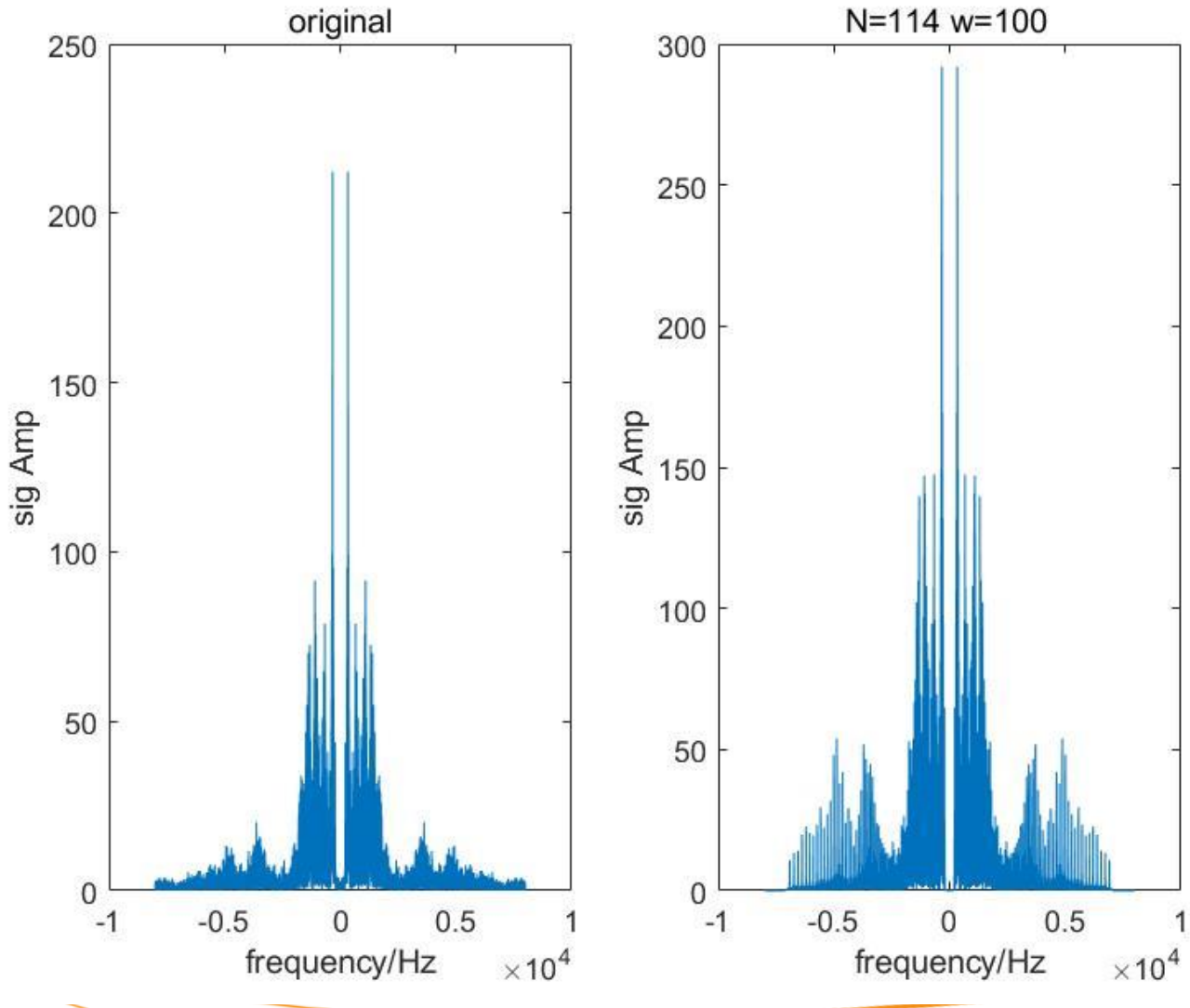
关于频段N





1. 从PSD图中可以看出，在N较高的情况下，随着频段N的增加，高频段的能量密度逐渐下降
2. 从声音的波形图来看，在N较高的情况下，随着频段N的增加，承载有用信息声音部分的波形幅值减小、特征变得不明显，但在N=121、125时出现反向升高的情况，猜想此时频段N恰好分在语音频率的间断处。在N很大时，声音波形幅值主要在声音末尾段，为原始音频无语音的片段。
3. 经过多次的尝试，在截止频率为50Hz时，语音最清晰的N为87。
4. Butter滤波器中阶数我们设置为4，会在低频率与高频率部分有上升与下降的部分，不是理想滤波器，对实验结果存在影响。

Best situation



经过寻找试听，我们得出：
频带数在114左右，LPF截止频率在100Hz时效果最好。
我们还绘制了最容易识别的结果图如下
($N = 114$ & LPF截止频率 = 100Hz)。

我们可以观察到结果非常接近原始信号。

经历收获

- 学会了使用matlab进行低通滤波器的仿真模拟，调试不同参数得到不同的滤波效果。
- 通过自学掌握了不少课堂内容以外的语法、函数等等，并且对循环结构有了更深刻的理解。
- 该项目是4人小组项目，有助于提高小组成员的团队意识。

SUSTech Southern University of Science and Technology

感谢观看！