IMPERIAL COLLEGE of SCIENCE, TECHNOLOGY & MEDICINE

DEPARTMENT of ELECTRICAL & ELECTRONIC ENGINEERING

MSc in Communications and Signal Processing

Formal Report No. 2

Name	Yibing Liu
Experiment Code	PN
Title of Experiment	Speech Signal Processing
Date of Submission	13/01/2021
Supervisor of Experiment	Dr. Patrick Naylor
Grade	

Communications and Signal Processing Laboratory

Signal Speech Processing

Yibing Liu

School of Electrical and Electronic Engineering

Imperial College London

yibing.liu20@imperial.ac.uk

Aims—The aim of this experiment is to implement a speech coder based on Linear Predictive Coding (LPC) and to investigate LPC analysis through a series of tests on some speech data. The experiment mainly learn the Winner-Hopf equations which are a classical technique to solve for the optimum linear-prediction coefficients for a second-order stationary stochastic process. Linear prediction will then be applied to speech signals. A number of exercises will be carried out to see the effects of the model order, the window, the reemphasise filter etc.

I. Background

The MATHEMATICAL analysis of the behavior of general dynamic systems has been an area of concern since the beginning of the last century. [1]. The analysis of the outputs of dynamic systems was for the most part the concern of "time series analysis," which was developed mainly within the fields of statistics, econometric, and communications. [2] Linear prediction is a mathematical operation where future values of a discrete-time signal are estimated as a linear function of previous samples. [3] In digital signal processing, linear prediction is often called linear predictive coding (LPC) and can thus be viewed as a subset of filter theory.

II. Introduction

A. Linear Prediction

Given a signal $x(n)$ we wish to predict future values of $x(n)$ from the previous samples. We call this process linear prediction. There are many models which are currently used in the processing of the linear prediction is that where a signal s_n is considered as the output of some system with some unknown input u_n such that the the following relation [1]:

$$s_n = -\sum_{k=1}^{p} a_k \cdot s_{n-k} + G \cdot \sum_{i=0}^{q} b_i \cdot u_{n-i}\ , \quad b_0 = 1 \quad (1)$$

the gain G are the parameters of the hypothesized system.

Equation 1 is can be specified in the frequency domain by taking the z transform on both sides. Let $H_{(z)}$ be the transfer function of the system, we can get :

$$H_{(z)} = \frac{s_{(z)}}{U_{(z)}} = \frac{1 + \sum_{l=1}^{q} b_l \cdot z^{-l}}{1 + \sum_{k=1}^{p} a_k \cdot z^{-k}} \quad (2)$$

where

$$S_{(z)} = \sum_{n=-\infty}^{\infty} s_n z^{-n} \tag{3}$$

is the z transform of s_n. For the $H_{(z)}$ in 2, there two special cases that are of interest.

- all-zero model: $a_k = 0$, $1 <= k <= p$
- all-pole model: $b_l = 0$, $1 <= l <= q$

In this experiment, we mainly study the all-polo model. In all-pole model, we assume the signal s_n is given as a linear combination of past values and some input u_n

$$s_n = -\sum_{k=1}^{p} a_k \cdot s_{n-k} + G \cdot u_n \tag{4}$$

where G is gain factor. The transfer function $H_{(z)}$ in 4 is:

$$H_{(z)} = \frac{G}{1 + \sum_{k=1}^{p} a_k \cdot z^{-k}} \tag{5}$$

We assume that the input u_n is totally unknown. Therefore, the signal s_n can be predicted only approximately from a linearly weighted summation of past samples. Let $\hat{s}_n$ be this approximation s_n , where

$$\hat{x}_{(n)=} - \sum_{k=1}^{p} a_k \cdot x_{(n-k)} \tag{6}$$

what we do is to find the coefficients a_k that minimize the mean-squared prediction error between the signal x_n and a predicted signal $\hat{s}_n$. For a p th order prediction, there are $(p+1)$ coefficients, $a_0 = 1$. And a_k can be designed by minimising , over the samples n ,the function of the sum of the squared difference between the true samples of $x_{(n)}$ and our estimate $\hat{x}_{(n)}$ can be written as:

$$E = \sum_{n} \left(x_{(n+r)} - \hat{x}_{(n)} \right)^2 \tag{7}$$

in the the equation (7) we can use auto-correlation method to choose the range n. From the Fig. 1, if $r = 0$ then the predictor predicts the current sample, if $r > 0$, then the predictor predicts a future sample.

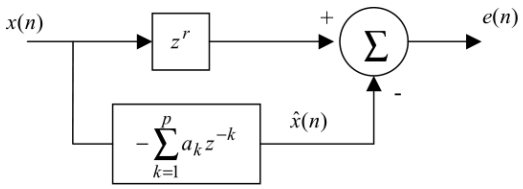

Fig. 1: Linear Predictor

B. Voiced sound Unvoiced sound

For the sound, there is an important concept: the voiced sound and the unvoiced sound [4]. Voiced sound is produced by forcing airflow through the glottis and adjusting the vocal cords to a proper tension. At this time, the vocal cords are relaxed and tense alternately, producing quasi-periodic air pulses to stimulate the vocal tract, and finally resulting in a quasi-periodic waveform. Unvoiced sound is produced by contraction in certain positions of the vocal tract, forcing air to be disturbed by the contraction point at a sufficiently high speed. This creates a wide-band noise source to excite the channel. When we plot the speech signal in time domain, we can see that the voiced sound has higher energy and often has a certain periodicity, but the unvoiced sound often doesn't has any periodicity and with lower energy.

C. Frequency Response

Frequency response is the quantitative measure of the output spectrum of a system or device in response to a stimulus, and is used to characterize the dynamics of the system. It is a measure of magnitude and phase of the output as a function of frequency [5]. When we have the poles and zeros for a system, we can write the frequency like this:

$$H_{(Z)} = G \cdot \frac{(z-z_1)(z-z_2)\dots(z-z_N)}{(z-p_1)(z-p_2)\dots(z-p_N)} \quad (8)$$

If we want to get the magnitude:

$$\left|H_{(Z)}\right| = G \cdot \frac{\prod|(z-z_i)|}{\prod|(z-p_i)|}, i \in [\mathbf{0}, N] \quad (9)$$

In Eq.9, it can be seen as the product of the distance. The distance from each point on the unit circle to the zero point is the numerator, and the distance to the pole is the denominator. Then

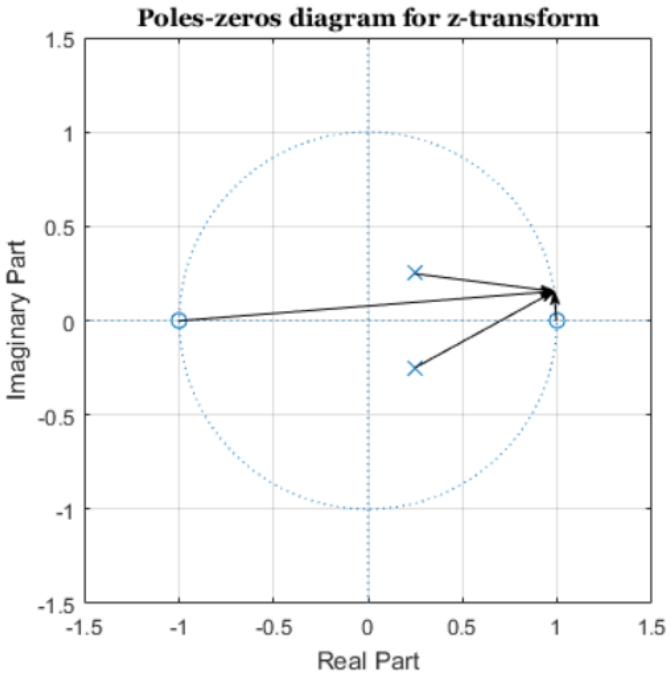

Fig. 2: Unit circle with zeros and poles

$$\left|H_{(Z)}\right| = G \cdot \frac{d_1 d_2 \dots d_N}{q_1 q_2 \dots q_N} \quad (10)$$

In this way, we can get the frequency response from zeros and poles. And for the phase can be calculated by:

$$< \left|H_{(z)}\right| =< G + < (z-z_1) + < (z-z_2) + \dots + < (z-z_N) - < (z-p_1) - < (z-p_2) - \dots - < (z-p_N) \quad (11)$$

III. Procedure and Result

In this section, we conduct experiments under the guidance of experimental requirements. We will process such a speech signal:'Oak is strong and also gives shade', we can plot the magnitude of this signal:

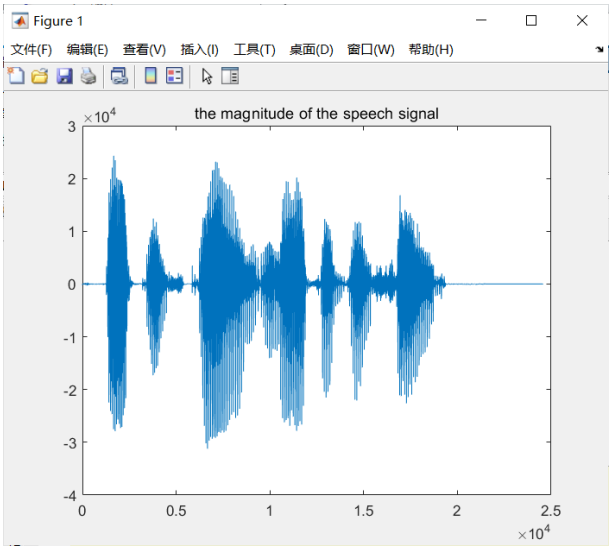

Fig. 3: The magnitude of the speech signal

A. Exercise 1

Write a MATLAB function to perform autocorrelation LPC Fig.4 :

```
function [A, G, r, a] = autolpc(x, p)
%AUTOLPC Autocorrelation Method for LPC
% Usage: [A, G, r, a] = autolpc(x, p)
% x : input samples
% p : order of LPC
% A : prediction error filter, (A = [1; -a])
% G : rms prediction error
% r : autocorrelation coefficients
% a : predictor coefficients
x = x(:);
L = length(x);
r = zeros(p+1,1);
for i=0:p
   r(i+1) = x(1:L-i)' * x(1+i:L);
end
R = toeplitz(r(1:p));
a = R\r(2:p+1);
A = [1; -a];
G = sqrt(sum(A.*r));
end
```

Fig. 4: LPC function

B. Exercise 2

We use a speech data file $s5.mat$, apply the LPC analysis on the data file using 12^{th} order predication using a Hamming window for two phonemes. We use the LPC function to get the Prediction error filter and the rms prediction error, and we can use the function $freqz$ in Matlab to get the frequency response vector.we can use

```
freqz(A,1,8000,'whole')
```

to get the prediction error filter's frequency response, and

```
freqz(G,a,8000,'whole')
```

to get the vocal tract model filter's frequency response. After that we can use $plot(20 * log10(abs(x)))$; to get the log magnitude.

Results are like Fig.5 and Fig. 6 For the

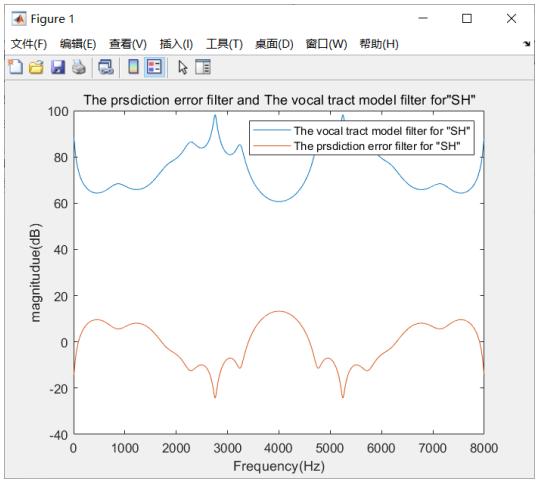

Fig. 5: The prediction error filter and The vocal tract model filter for"SH"

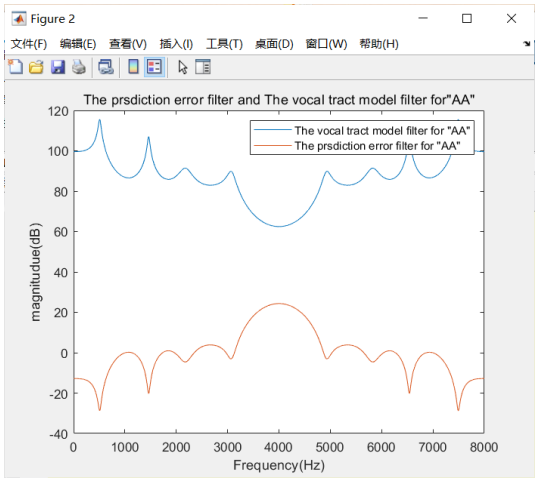

Fig. 6: The prediction error filter and The vocal tract model filter for"AA"

zeros of the prediction error filter for both cases, like Fig.7 and Fig.8

C. Exercise 3

We can just plot the signal of SH and AA to get the time domain figure Fig.9: In the experiment, all the signals we processed are windowed signals, which means we use a Hamming windows. Multiply the signal segment and the Hamming window to get the windowed segment, for the signal AA is like Fig.10

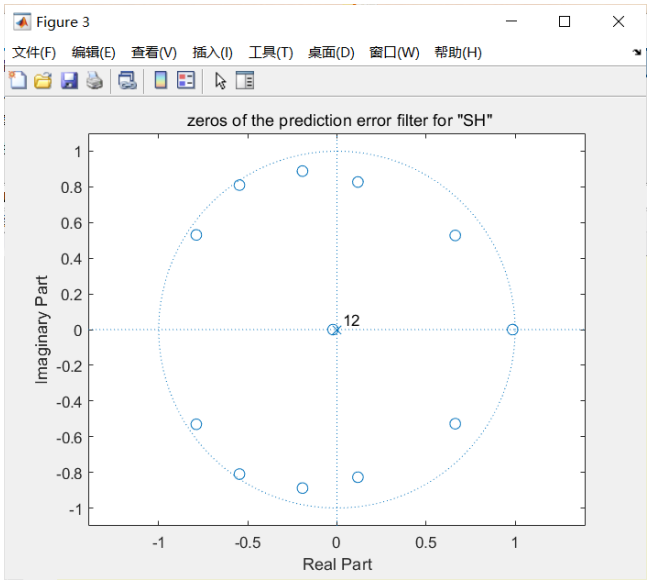

Fig. 7: Zeros for"SH"

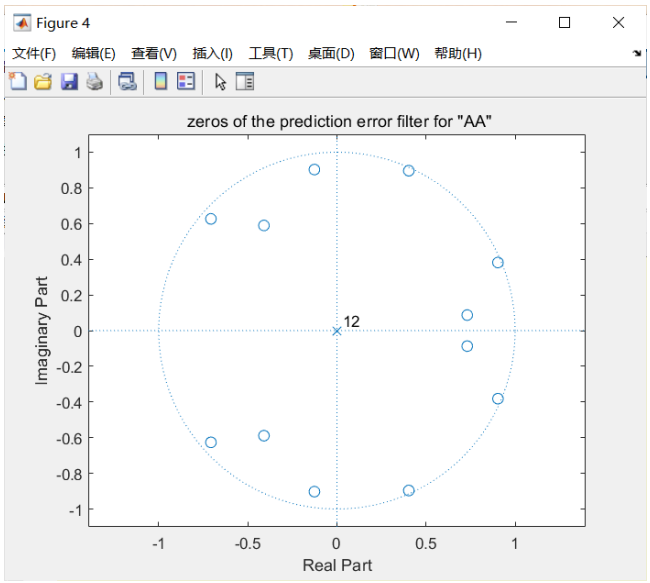

Fig. 8: Zeros for"AA"

Compute the DFT of the windowed segment of speech and plot is magnitude (in dB) on the same graph as the vocal tract model. Like Fig.11 and Fig. 12

D. Exercise 4

Repeat the above tests on other phonemes. We use the o in *strong* and i in *give* and use the method in Exercise 3 to find the time

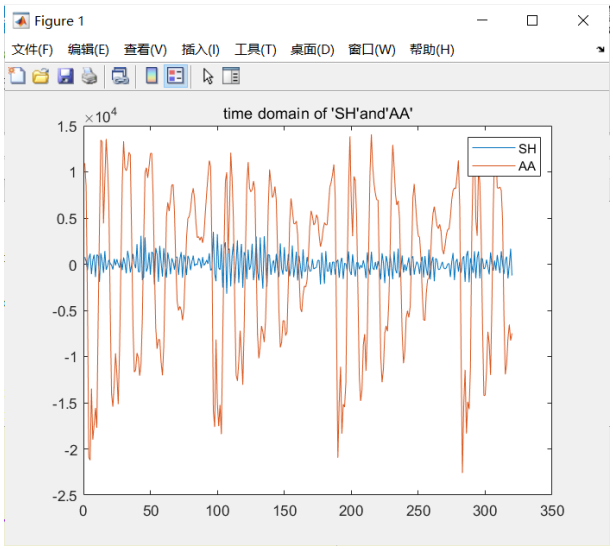

Fig. 9: Time domain of SH and AA

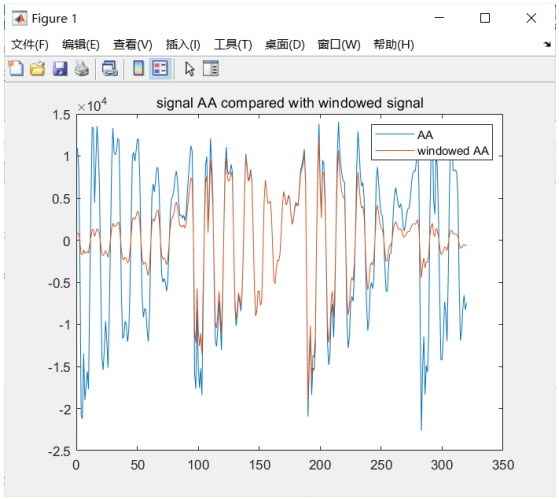

Fig. 10: The signal AA and the windowed AA

domain figure like Fig.13 and DFT like Fig. 14 and 15

E. Exercise 5

Now, we change the order of LPC to investigate the effect of varying the model order on the magnitude spectral plots. Like Fig. 16

F. Exercise 6

Apply a preemphasis filter to speech before using linear prediction analysis. the filter is:

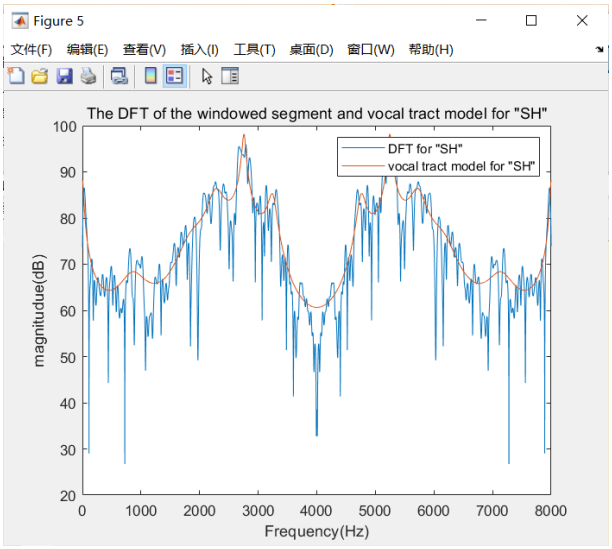

Fig. 11: The DFT of the windowed segment and vocal tract model for "SH"

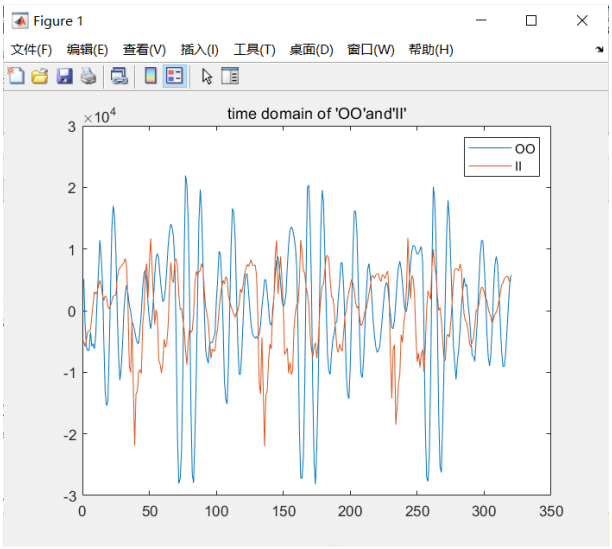

Fig. 13: Time domain of O and I

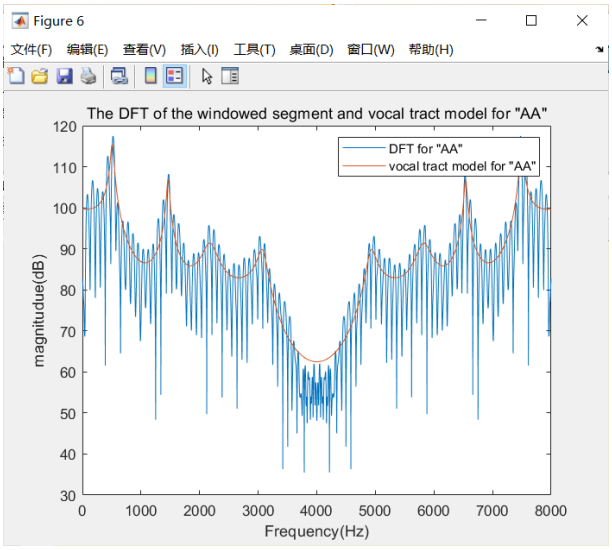

Fig. 12: The DFT of the windowed segment and vocal tract model for "AA"

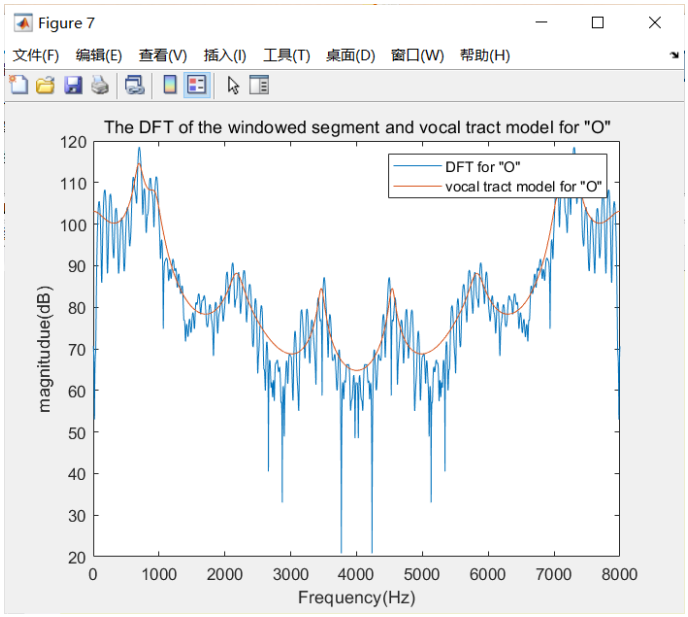

Fig. 14: The DFT of the windowed segment and vocal tract model for "O"

```
y=filter([1,-0.98], 1, s5);
```

We can get the figure like Fig. 17

IV. Conclusion

In Exercise 1, when we input the samples and the order of the LPC, we can get the prediction error filter A , which can be gotten from predictor coefficients a by $A = [1; -a]$, and the rms prediction error G. And using G/A we can get the vocal tract model filter.

In Exercise 2, we can see that in both Fig.5 and Fig. 6, the figure of vocal tract model filter and the prediction error filter is exactly symmetrical. Because in the linear prediction processing, the vocal tract model filter is just

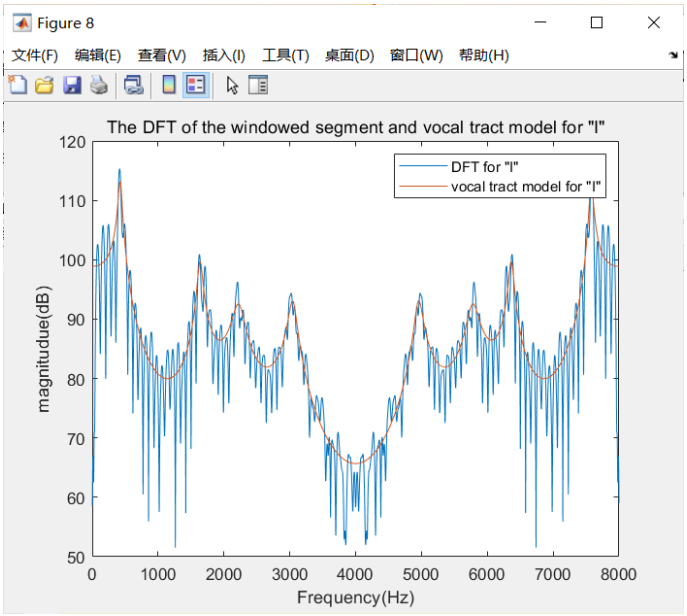

Fig. 15: The DFT of the windowed segment and vocal tract model for "I"

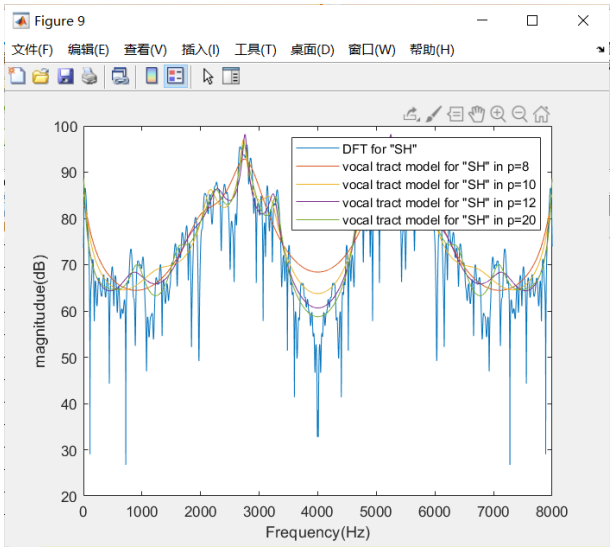

Fig. 16: The DFT of the windowed segment and vocal tract model in different p for "SH"

the reciprocal of the prediction error for vocal tract model filter $H = G/A$.

- zeros of the prediction error filter are just corresponding to the poles of vocal tract filter and the number are equal to the filter order.
- number of dips in the frequency response of the prediction error model is equal to the number of zeros of the prediction error filter which close to the unit circle.

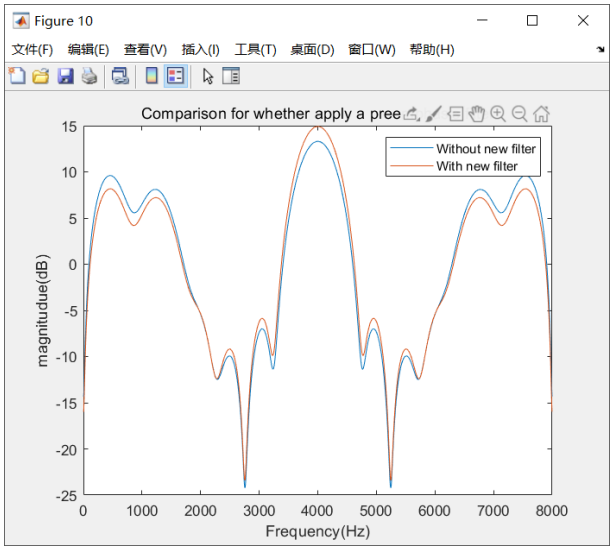

Fig. 17: Comparison for whether apply a preemphasis filter for prediction error filter of SH

In Exercise 3, according to the Fig.10, the windowed signal's center is strengthened by filtering out the edge part. And it is obvious that the frequency spectral density in Fig.12 for 'AA' is much higher than Fig.11 for 'SH'. Which means energy of AA is larger than SH. And according to the Fig.9 The waveform of AA has the property of quasi-period and can be found easily, while the waveform of SH is live noise and hard to be recognized. So the phoneme AA is voiced and SH is unvoiced.

In Exercise 4, from the Fig.13, both the O and the I all show a certain periodicityso both o and i belong to voiced phonemes. However, there are still some differences between them. The magnitude of the o is bigger than the i and the peaks of DFT of the two signals are

different. And o is so-called middle vowel, while i is front vowel. When the front vowel is uttered, the vocal tract is contracted by the tongue bulging forward, with a lower first resonance frequency and a higher second resonance frequency. The middle vowel has a higher first resonance frequency and a lower second resonance frequency

In Exercise 5, we take SH as example, we can see that, with the increasing of the order p, the vocal tract model filter fits better to the DFT of the windowed signal as shown in Fig16. Because with higher order, after the filter, the signal has more similarity to the original signals. However, with the increasing of the order p, the amount of calculation is also increasing, so if we use very high order, there are two obvious disadvantages: high computing and over-fitting modelling.

In Exercise 6, as shown in Fig.17 the new filter we used is a high pass filter.We take the prediction error filter of SH as an example. As we set $[1, -0.98]$, where the frequency components will be filtered out. Pre-emphasis on the input digital voice signal is to emphasize the high-frequency part of the voice, remove the influence of lip radiation, and increase the high-frequency resolution of the voice. The frequency spectrum after pre-emphasis has been improved in the high frequency part.

V. Acknowledgement

The author would like to express the special thanks to all people offering help and support to her during the experiment.

Thanks go to:

Dr. Patrick Naylor (Academic supervisor)

Mr. Neo Vincent(Lab demonstrator)

References

[1] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[2] L. R. Rabiner and R. W. Schafer, *Introduction to digital speech processing*. Now Publishers Inc, 2007.

[3] Wikipedia contributors, "Linear prediction — Wikipedia, the free encyclopedia," 2020, [Online; accessed 12-January-2021]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Linear_prediction&oldid=993353572

[4] ——, "Voice (phonetics) — Wikipedia, the free encyclopedia," 2020, [Online; accessed 15-January-2021]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Voice_(phonetics)&oldid=971311493

[5] ——, "Frequency response — Wikipedia, the free encyclopedia," 2020, [Online; accessed 15-January-2021]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Frequency_response&oldid=973947722