國立臺灣大學電機資訊學院資訊網路與多媒體研究所

碩士論文

Department of Graduate Institute of Networking and Multimedia

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

利用零參數濾波器重構自己的聲音

Reconstruction of One's Own Voice with a Zero-Parameter Filter

林睦航

Mu-Hang Lin

指導教授: 洪一平 博士

Advisor: Yi-Ping Hung Ph.D.

中華民國 111 年 9 月

September, 2022

doi:10.6342/NTU202203643

摘要

眾所周知，人們在聽到自己的錄音時常常會有不熟悉甚至是不舒服的感覺。這種對自己錄製聲音的怪異感其實是聲音的傳輸機制所導致。當我們在錄製聲音時，麥克風只有錄製到透過空氣傳導的聲音；但我們自己所聽到的聲音，是透過空氣傳導和骨頭傳導這兩者聲音一起組合而成的。有鑒於透過骨頭傳導的聲音無法被錄製，我們希望能找到一個快速且簡單的轉換方法，將自己錄音的聲音轉換成自己熟悉的聲音。

在本篇論文中，我們設計並實現了一個不需要調整參數的零參數濾波器，能將使用者的錄音轉換成使用者自己的聲音。我們將此濾波器和其他的轉換方法（雙參數模型和八參數等化器）一起做盲測驗證實驗。實驗結果顯示，零參數濾波器在使用上最為方便快速，受試者們對轉換結果都能給出正向回饋，並且在整體表現上也不遜於雙參數模型及八參數等化器。最後，我們進一步構思出獨特的潛在應用。

關鍵字：濾波器、錄音、自己的聲音

doi:10.6342/NTU202203643

Abstract

As we know, when people hear their own voices, they often feel unfamiliar or even uncomfortable. This strange feeling about recorded voice is actually caused by the transmission mechanism of the voice. When we record a voice, the microphone only records the sound transmitted through the air; however, the sound we hear is a combination of air-conducted and bone-conducted sounds. Since bone-conducted sounds cannot be recorded, we wanted to find a quick and easy way to convert our own recorded voices into our own familiar voices.

In this paper, we design and implement a zero-parameter-filter approach that does not require parameter adjustment and converts one's recording into one's own voice. This filter was tested together with other conversion methods (two-parameter model and eight-parameter equalizer) in a blind validation experiment. The results showed that the zero-parameter filter was the easiest and fastest to use, and the subjects gave positive feedback on the conversion results, and the overall performance was no worse than that of the two-

doi:10.6342/NTU202203643

parameter model and the eight-parameter equalizer. Finally, we further conceive unique potential applications.

Keywords: filter, recording, own voice

doi:10.6342/NTU202203643

Contents

doi:10.6342/NTU202203643

doi:10.6342/NTU202203643

List of Figures

doi:10.6342/NTU202203643

doi:10.6342/NTU202203643

Chapter 1 Introduction

We all know how strange it often feels when we hear a recording of our own voice. The difference in perception is due to the fact that our body takes multiple pathways to transmit sound from our vocal cords to our ears when vocalizing, rather than a single air-conduction pathway when playing a recording.

The transmission pathway through the body is known as ”bone-conduction” and has been studied in various ways since the 19th century. However, because of the complex shape of the skull, it is difficult to elucidate the mechanism of bone-conduction in humans. Figure 1.1 shows the auditory pathway; number 1 indicates the bone-conduction (BC) pathway and number 2 indicates the air-conduction (AC) pathway.

However, the BC sound cannot be captured during the vocalization. Therefore, in order to obtain one’s own voice, Won et al. [11] try to find a transfer function H that converts one’s AC voice to one’s own voice.

They first propose an eight-parameter equalizer, which allows users to convert their own sounds by adjusting eight sliders. After collecting experimental data, they then used singular value decomposition (SVD) to reduce the eight-parameter equalizer to a two-parameter model. In other words, the user only needs to adjust two sliders to convert his or her own voice. The subjects also showed positive feedback on the two-parameter

doi:10.6342/NTU202203643

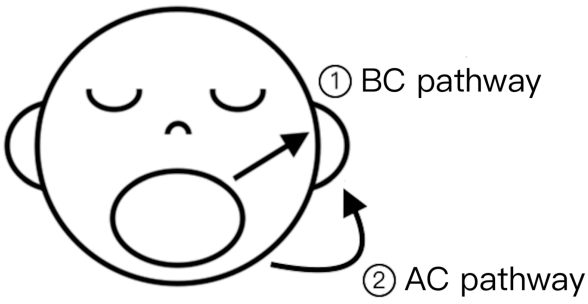


Figure 1.1: Auditory Pathway

model.

After carefully reading and studying the above papers, we found that there is an opportunity to simplify the two-parameter model even further. We propose a method using a zero-parameter filter, which does not require the user to adjust the sliders to convert their own voice. In this paper, we present information about the zero-parameter-filter approach, the experimental procedure, the analysis of the results, and its applications, as follows.

- **Filter Design:** We describe how we found the possibility of simplification from the two-parameter model. We also designed a graphical user interface. It allows users to quickly listen to the voice converted by the zero-parameter filter and compare the differences before and after the conversion.

- **Experiment:** We then performed a blinded validation experiment of the zero-parameter, two-parameter, and eight-parameter methods on subjects to fairly compare the results of the three methods. We analyzed the experimental results and showed the feasibility of using the zero-parameter filter with a bar chart. We also found the correlation between the subjects' background and the experimental results.

- **Conclusion and Future Works:** We combined the filter with a text-to-speech model to implement a text reader of one's own voice, and also conceived interest-

doi:10.6342/NTU202203643

ing applications. Finally, we conclude the whole paper and suggest how to further improve our zero-parameter-filter approach in the future.

doi:10.6342/NTU202203643

Chapter 2 Related Works

2.1 Air-Conduction and Bone-Conduction Sound

While listening to yourself, the voice is generated by your vocal cords and transmitted through the air and bone to your auditory organs. Although the transmission pathways of AC and BC sound are different, both sounds linearly combine at inner ear and excite the basilar membrane similarly [2, 8]. Figure 2.1 shows the overview of one's own voice [3, 9].

Other people can only hear the AC part of the sound, while you are the only person who can hear the BC sound. When you listen to a recording of yourself, the BC sound is lost. This is exactly why you think your own recordings sound weird.

However, the BC sound cannot be recorded, we cannot synthesize our own voice

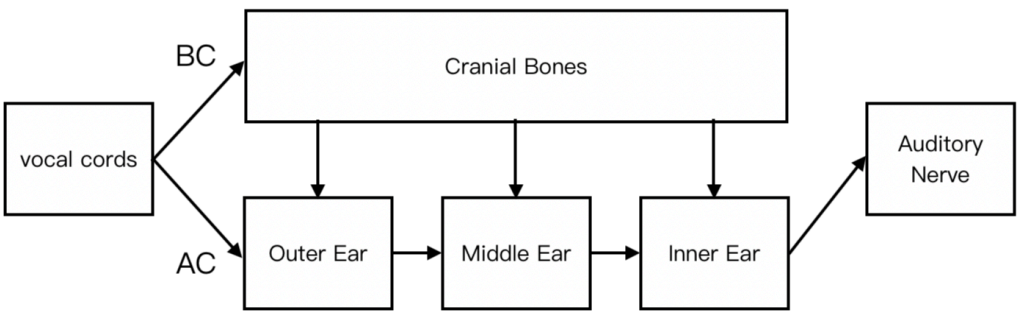


Figure 2.1: The Overview of One's Own Voice.

doi:10.6342/NTU202203643

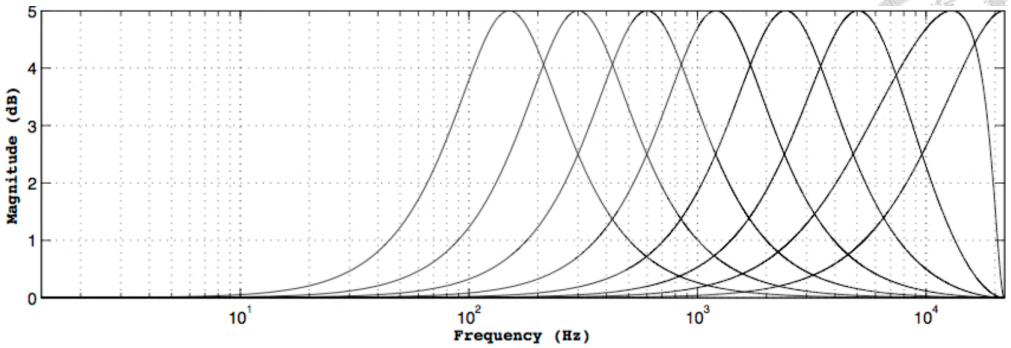


Figure 2.2: Seven Peak Filters and One High-Shelf Filter (Won, 2013).

in a linear equation. So won et al. [10] used multi-band filtering to experiment with a equalizer, trying to find a transfer function to convert our own voice. They do this by extending the equalizer method of [7].

2.2 Eight-Parameter Equalizer

In order to estimate the transfer function of one's own voice, They designed an experiment in which the human voice could be manipulated, so the experimental equalizer should effectively cover the range of the human voice. According to [4, 6], the frequency spectrum of speech extends from 100 Hz to 7 kHz.

Therefore, they positioned seven peak filters and one high-shelf filter in equal parts on the frequency axis of the logarithmic scale, as shown in Figure 2.2 [1]. Thus, they have eight frequency bands, corresponding to the eight sliders in the equalizer.

They designed a graphical user interface that allows users to easily convert voices. There were two subject pools in their study –a group of 8 amateurs and 13 singers with professional training. Participants were asked to sing an octave and to say four sentences.

As shown in Figure 2.3 and Figure 2.4, they observed that all participants had self-similar transmission functions except for two anomalies in the amateur group. The transfer

doi:10.6342/NTU202203643

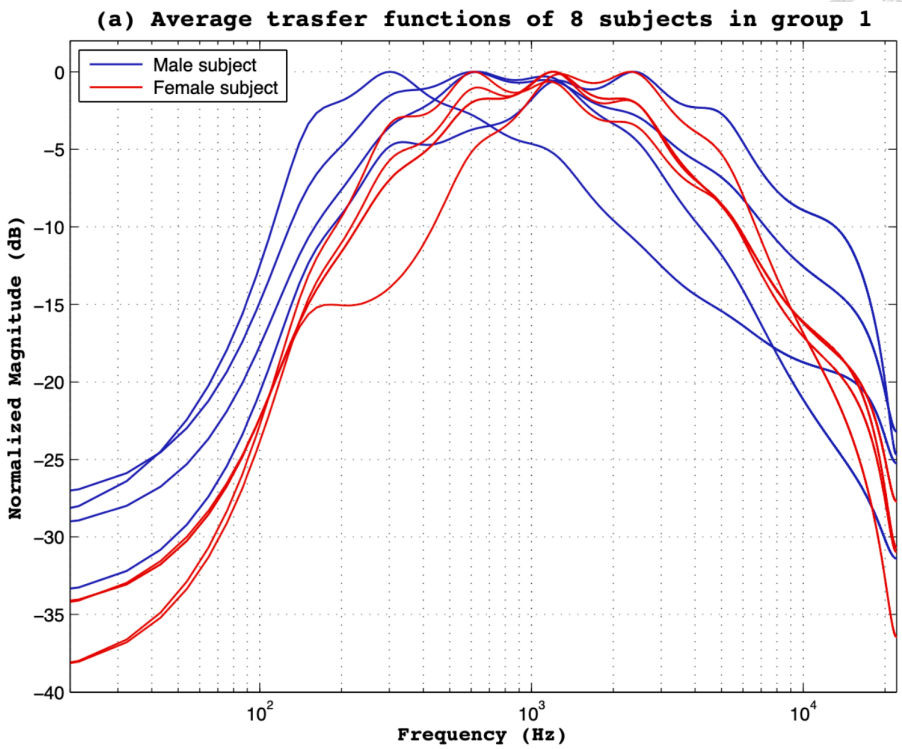


Figure 2.3: Average Transfer Functions of 8 Subjects in the Amateur Group (Won, 2013).

functions were emphasized from 300 Hz to 1200 Hz. Also, all mean transfer functions were very similar.

2.3 Two-Parameter Model

On the basis of the above results, Won et al. [10] used SVD to estimate the transfer function. After applying SVD in Matlab, they got three matrices which represent the filter shape, filter weight and individual difference, respectively.

The top two values of the filter weight (blue and green circles in Figure 2.5 (b)) [11] are distinctly higher than the others. They hope to approximate the transfer function by using these two filters (blue and green waveform in Figure 2.5 (a)) [11].

After several pilot tests, a model M with a constant filter and two variable filters is obtained, as shown in Equation 2.1. Filter1 is the constant blue waveform, Filter2 is

doi:10.6342/NTU202203643

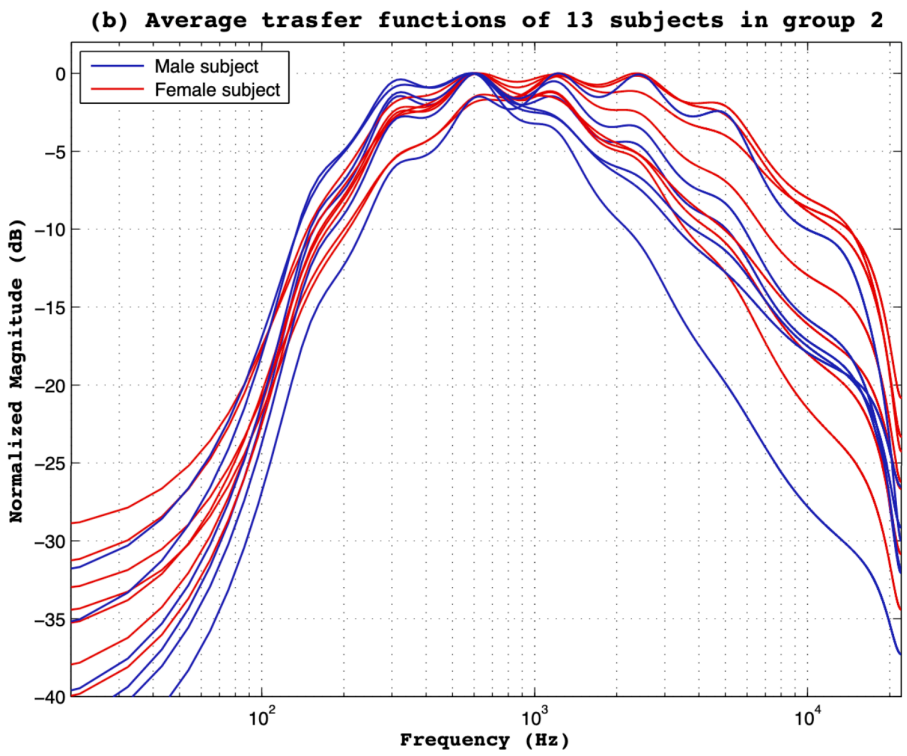


Figure 2.4: Average Transfer Functions of 13 Subjects in the Professional Group (Won, 2013).

the low-frequency part of the green waveform with a variable α, and Filter3 is the high-frequency part of the green waveform with a variable β, as shown in the Figure 2.6 [11].

$$M = Filter\ 1 + \alpha \cdot Filter\ 2 + \beta \cdot Filter\ 3. \tag{2.1}$$

They used this two-parameter model M to conduct their experiment. The experiment recruited 14 participants, including 7 males and 7 females. Similar to the experiment with the eight parameters, but this time only two sliders were adjusted. In addition, they were asked to provide a score of 0 for the AC voice and 100 for their own voice.

The final data showed that most of the participants scored between 70 and 80. Although the scores are slightly lower than those of the eight-parameter equalizer experiment, it still shows that the two-parameter model gives satisfactory results for the goal of

doi:10.6342/NTU202203643

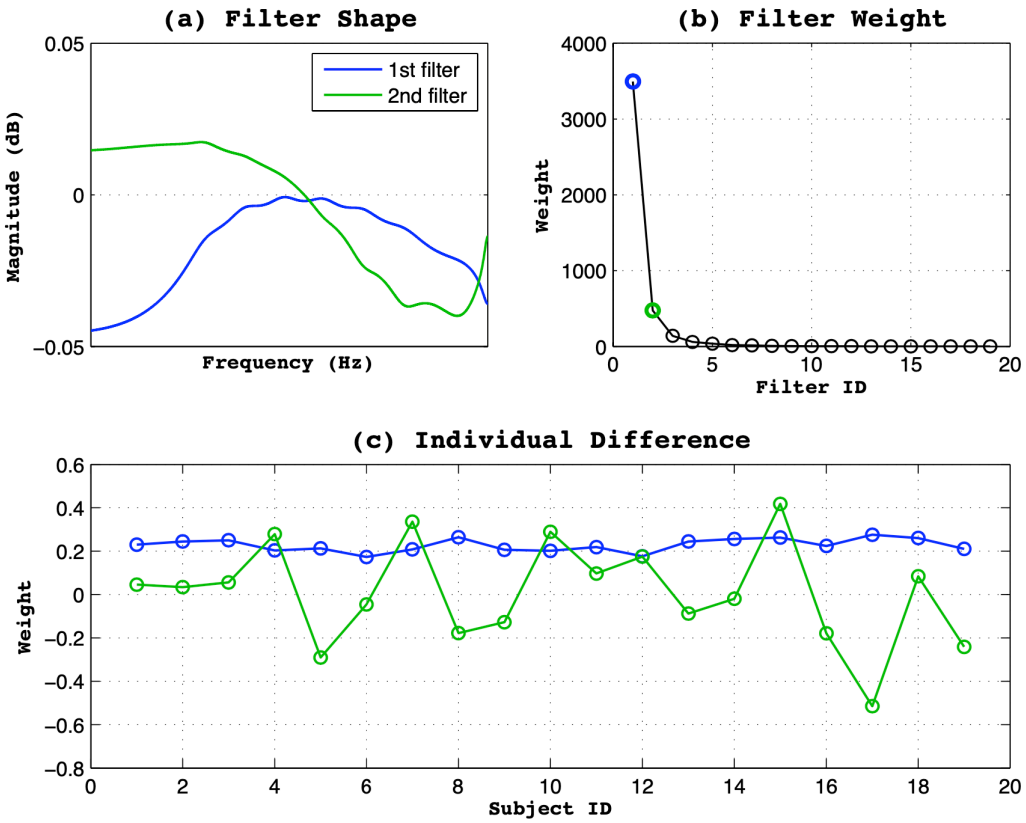


Figure 2.5: Decomposed Three Matrices by SVD. (a) Filter Shape (b) Filter Weight (c) Individual Differences (Won, 2013).

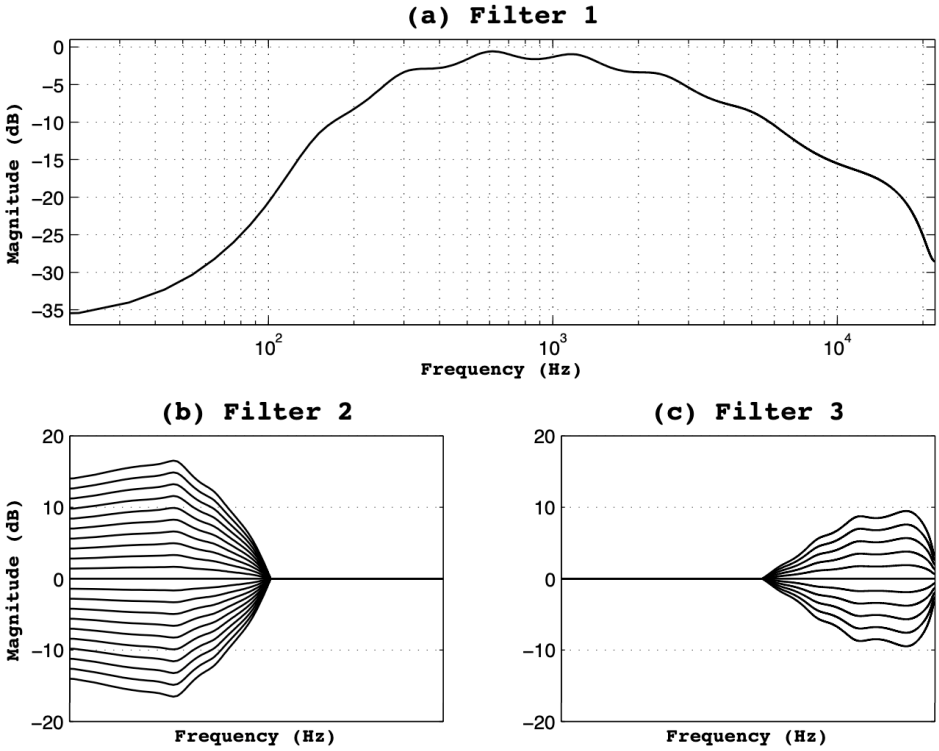


Figure 2.6: Three Filters Used in Equation 2.1. (a) Filter 1 (b) Filter 2 (c) Filter 3 (Won, 2013).

doi:10.6342/NTU202203643

simulating one’s own voice.

doi:10.6342/NTU202203643

Chapter 3 Filter Design

3.1 Zero-Parameter Filter

After carefully reading the paper by Won et al. [10], we found that Filter 1 in their two-parameter model may be the key to the results, as shown in Figure 2.4 (a). The transfer function they derived is a constant filter 1 with two variables, filter 2 and filter 3, to allow users to adjust their voice. However, the weight of filter 1 is 6 times more than the weight of filter 2 and filter 3. Then we wondered if there was a way to make the voice conversion easier.

$$One's\ own\ voice = Filter\ 1 \cdot AC\ voice. \tag{3.1}$$

Therefore, we only use the constant Filter 1 as our transfer function, without adding any variables such as equation 3.1. In this way, users do not have to spend time adjusting parameters, and this zero-parameter filter can achieve the easiest conversion. We hope to see if this zero-parameter-filter approach can give satisfactory results to most users.

doi:10.6342/NTU202203643

3.2 Graphical User Interface

We implement a graphical user interface as shown in Figure 3.1. The green waveform in the middle is the Filter 1 we used, the x-axis is the sound frequency, and the y-axis is the decibel. There are 6 buttons with the following functions.

- Load: Select the sound to be converted.
- Original: Play the unconverted sound and repeat it.
- Play: Play the converted sound and repeat it.
- Stop: Stop the sound being played.
- Save: Save the converted sound.
- Filtered All: Converts multiple audio files at once.

With the above button, users can easily listen to their converted voice to see if it is close to their own. Moreover, users can always listen to the original unconverted sound for comparison.

For the validation experiment, we also implemented a two-parameter model and an eight-parameter equalizer, for which the user interfaces are shown in the Figure 3.2 and the Figure 3.3, respectively.

As you can see, the two-parameter model has two gray sliders on the waveform interface that control the corresponding Filter 2 and Filter 3. As the user adjusts the sliders, the voice is converted in real time. And the bottom of the interface shows the gains corresponding to Filter 2 and Filter 3, while the functions of the six buttons remain the same.

doi:10.6342/NTU202203643

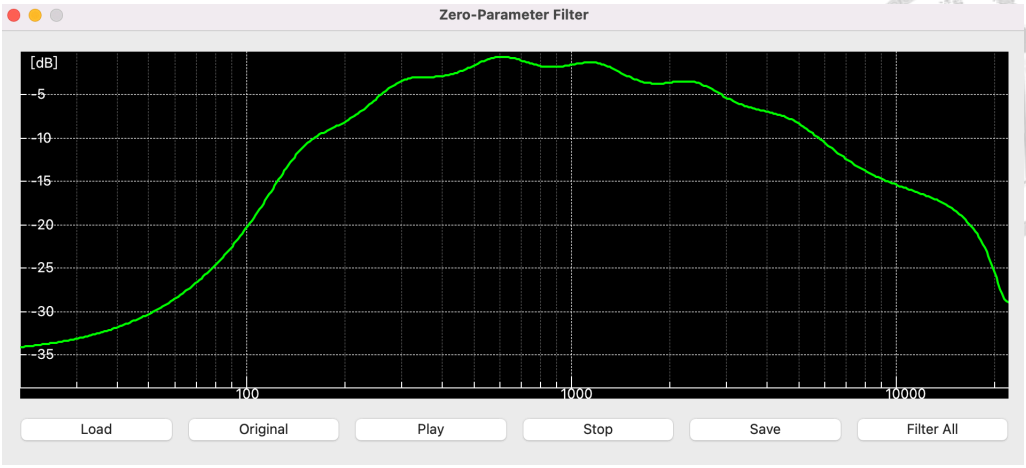


Figure 3.1: Graphical User Interface of Zero-Parameter Filter.

The eight-parameter equalizer has eight gray sliders on the waveform interface, corresponding to the center frequencies of 150, 300, 600, 1200, 2400, 4800, and 9600 Hz for the peak filters and 9600 Hz for the high-shelf filter, and the voice is also converted in real time. And the bottom of the interface shows the gains corresponding to each filter, while the functions of the six buttons also remain the same.

doi:10.6342/NTU202203643

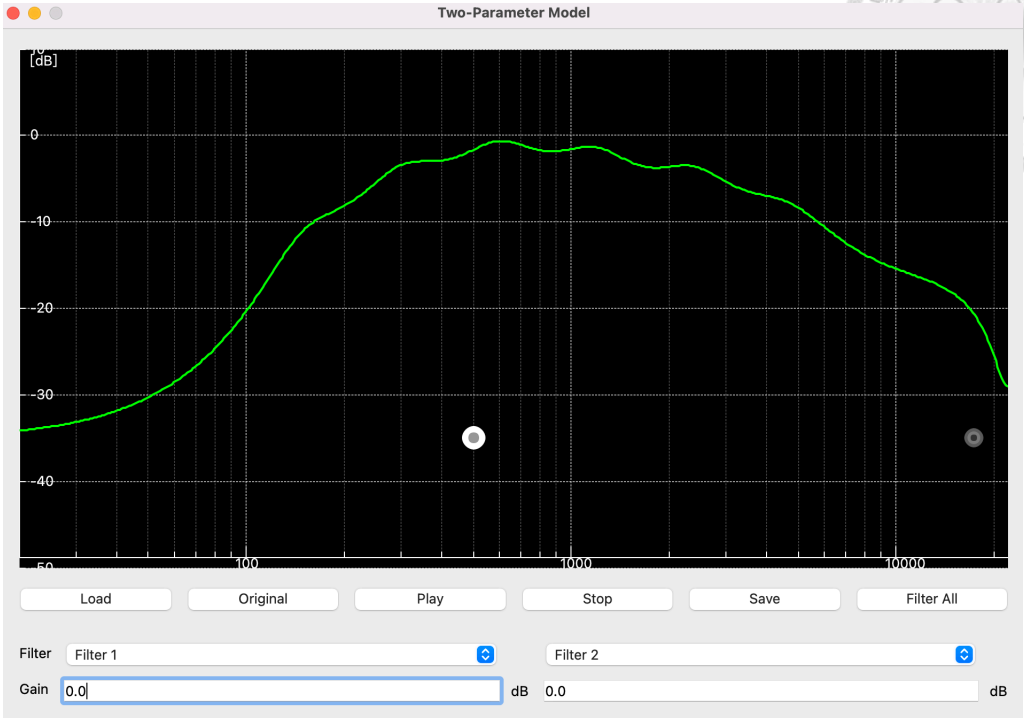


Figure 3.2: Graphical User Interface of Two-Parameter Model.

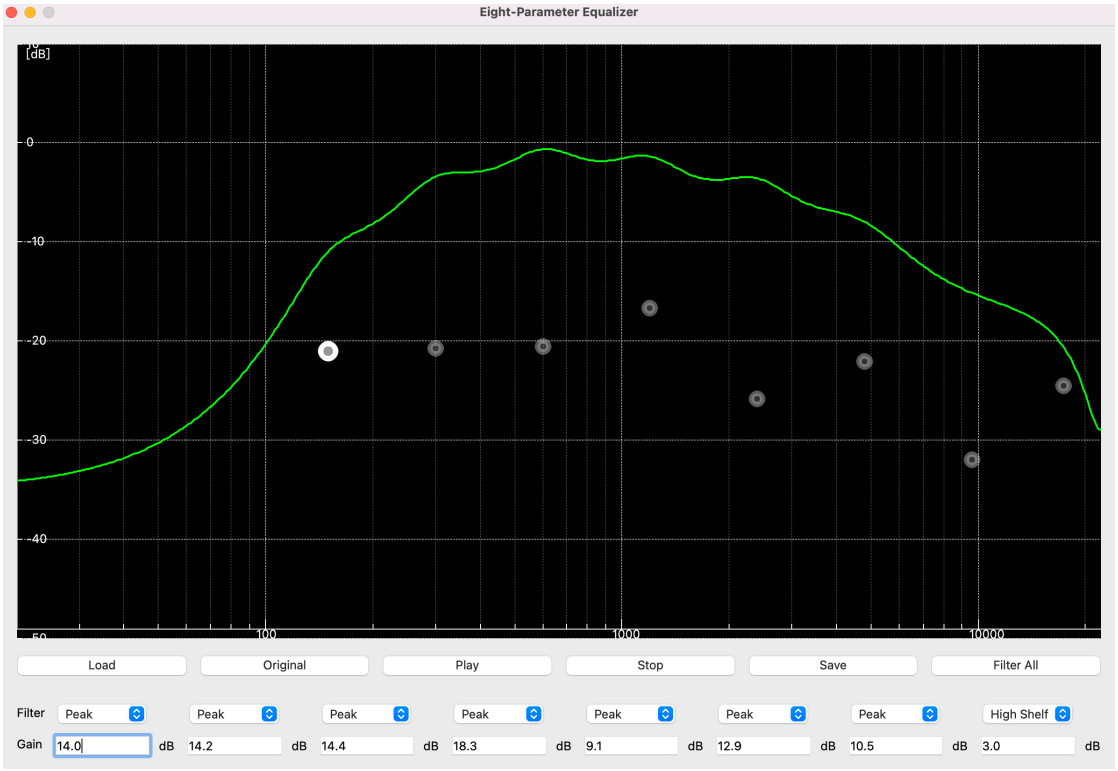


Figure 3.3: Graphical User Interface of Eight-Parameter Equalizer.

doi:10.6342/NTU202203643

Chapter 4 Experiment

To demonstrate the feasibility of the zero-parameter filter, we design a validation experiment that fairly compares the zero-parameter filter, the two-parameter model, and the eight-parameter equalizer.

4.1 Experiment Setup

A total of 16 subjects were recruited, 10 males and 6 females. The whole experiment took 45 minutes to 1 hour. We used Blue Yeti USB microphone to record and TaoTronics TT-BH046 Hybrid ANC Headphone to listen. The subjects were all using the same set of equipment and the experiment was conducted in a quiet environment.

4.2 Procedure

4.2.1 Questions and Recording

Before the experiment, the subjects were asked several questions about their musical background. We hope to find out the possible correlations by understanding the subject's musical background.

doi:10.6342/NTU202203643

- Are you a professional musician?
- Do you play any musical instruments?
- How many years in total?

Then, we asked the subjects to record one octave and ten sentences as follows (in Chinese).

1) Do-Re-Mi-Fa-So-La-Ti-Do

2) 我靜靜的躺在沙丘的椅子上

3) 感覺非常安全，非常舒適

4) 微風輕輕吹著樹梢

5) 一切是那麼輕鬆，那麼愉快

6) 深深的吸氣，深深的吐氣

7) 愛是恆久忍耐又有恩慈

8) 空即是色，色即是空

9) 一花一世界，一葉一如來

10) 人生天地之間，若白駒過隙

11) 上善若水，水善利萬物而不爭

doi:10.6342/NTU202203643

4.2.2 Adjusting and Blind Test

After recording, we asked the user to adjust the two-parameter model and the eight-parameter equalizer using the octave recording, so that the user could find the waveform setting that was closest to his or her own voice, respectively. The zero-parameter filter did not need to be adjusted by the subject. On average, it took about 2 minutes to adjust the two-parameter model and 5 minutes to adjust the eight-parameter equalizer.

Next, we conducted a blind test for each subject. Based on the subject's adjusted waveform settings, we converted 10 sentences (number 2 to number 11) recorded by the subject. Each sentence was converted by the zero-, two-, and eight-parameter methods, which resulted in a total of 30 sentences. The 30 sentences were randomly ordered and the subject listened to them one by one and scored each one. The score was given on the same scale as for Won et al. [10], where the air-conduction voice was 0 and their own voice (AC + BC) was 100. By doing this, we were able to obtain fair scores for the analysis of the results.

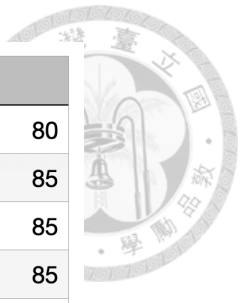
4.3 Result and Analysis

We present here the tables and bar charts of the subjects and the whole group so that we can analyze the experimental results in detail.

4.3.1 Result

The scores obtained from the experiment are displayed in a table for each participant. The 30 individual scores for the three methods (0-param, 2-param, and 8-param)

doi:10.6342/NTU202203643



	0-param	2-param	8-param
1	75	80	80
2	65	75	85
3	75	85	85
4	80	70	85
5	65	80	70
6	85	90	85
7	85	80	75
8	80	85	85
9	70	75	85
10	65	75	80
AVERAGE	74.5	79.5	81.5
STDEV	7.9756574093369	5.9860949986893	5.2967495273569

Figure 4.1: Score Table of Subject 1.

are shown, along with the average scores and standard deviations. Figure 4.1 shows the scores of subject 1.

In order to analyze the data more effectively, we exported the score table for each subject into a bar chart. We see the average scores and standard deviations of the three methods, as shown in Figure 4.2. In this way, we can easily know which method performs best, second best, or worst.

After obtaining the average score for each participant, we then calculated the average score for all subjects, as shown in Figure 4.3. Each row represents the average score of one subject, and there are 16 rows in total, and we finally calculate the overall average score and standard deviation of 16 subjects.

We also exported the overall average score table as a bar chart, as shown in Figure 4.4. With this chart, we can analyze the overall performance of the zero-parameter filter, the two-parameter model and the eight-parameter equalizer.

doi:10.6342/NTU202203643

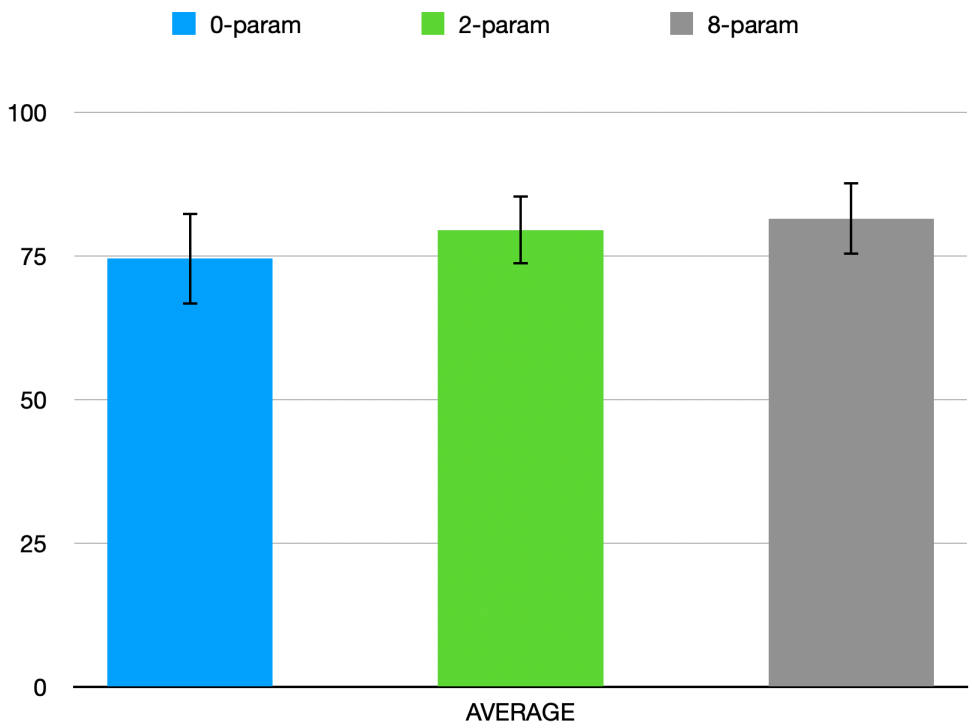


Figure 4.2: Score Bar Chart of Subject 1.

	0-param	2-param	8-param
1	74.5	79.5	81.5
2	80.7	73	88.4
3	78	77.5	77
4	71.3	63.4	72.4
5	72.5	80	84
6	72.3	84.1	73.1
7	57.5	61.5	59.5
8	81.5	80	78.5
9	62	58	59
10	72	70.7	82.7
11	63	69	73
12	64.5	72	75.5
13	61	65.5	71.5
14	68.5	74.5	77
15	83.5	86.5	83.5
16	83	81.5	83
AVERAGE	71.6125	73.54375	76.225
STDEV	8.3291356094135	8.4209238408462	8.2369492734466

Figure 4.3: Mean Score Table of All Subjects.

doi:10.6342/NTU202203643

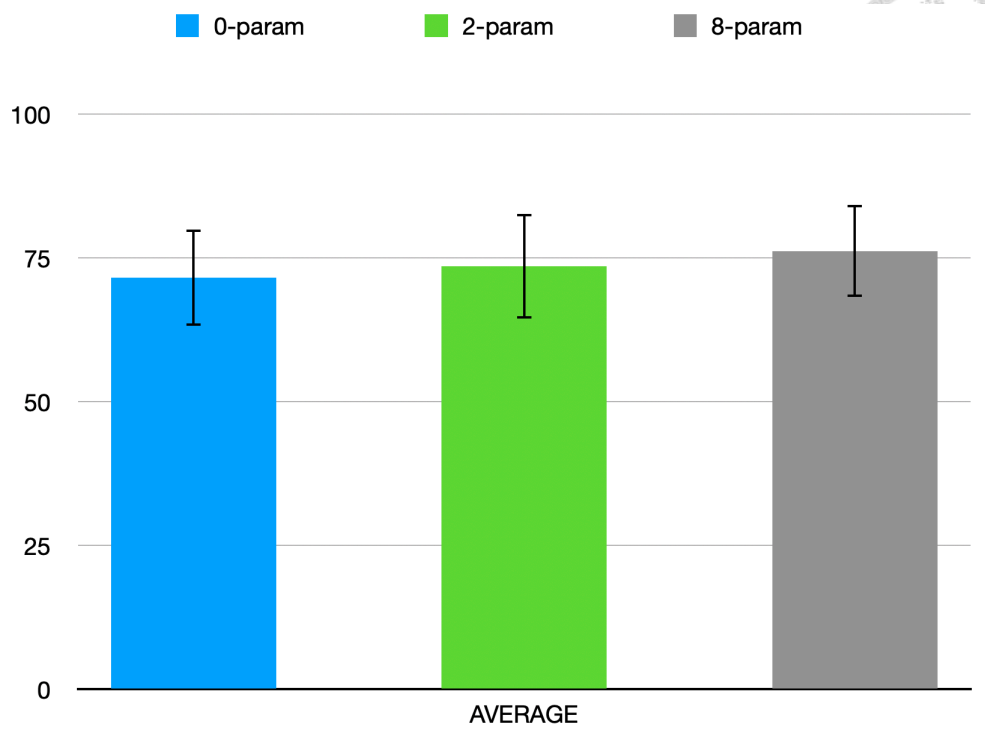


Figure 4.4: Mean Score Bar Chart of All Subjects.

4.3.2 Analysis

Theoretically, the zero-parameter filter, two-parameter model and eight-parameter equalizer should show the incremental form of lowest, middle and highest in terms of scores. Because the number of parameters directly affects the range of waveform that a subject can adjust, the incremental form should be the most theoretical result. Figure 4.4 indeed confirms our prediction, with the overall average scores of 71.61, 73.54, and 76.23 for the three methods, showing an increment.

However, in a closer analysis, not every participant's result showed a theoretical form. Only 6 of them showed a incremental form in the bar chart, while the other 10 did not. For example, the Figure 4.5 shows a hill shape. And the Figure 4.6 shows a valley shape. There is also a decreasing, as shown in the Figure 4.7.

Therefore, we started with the musical background of the subjects, hoping to clarify

doi:10.6342/NTU202203643

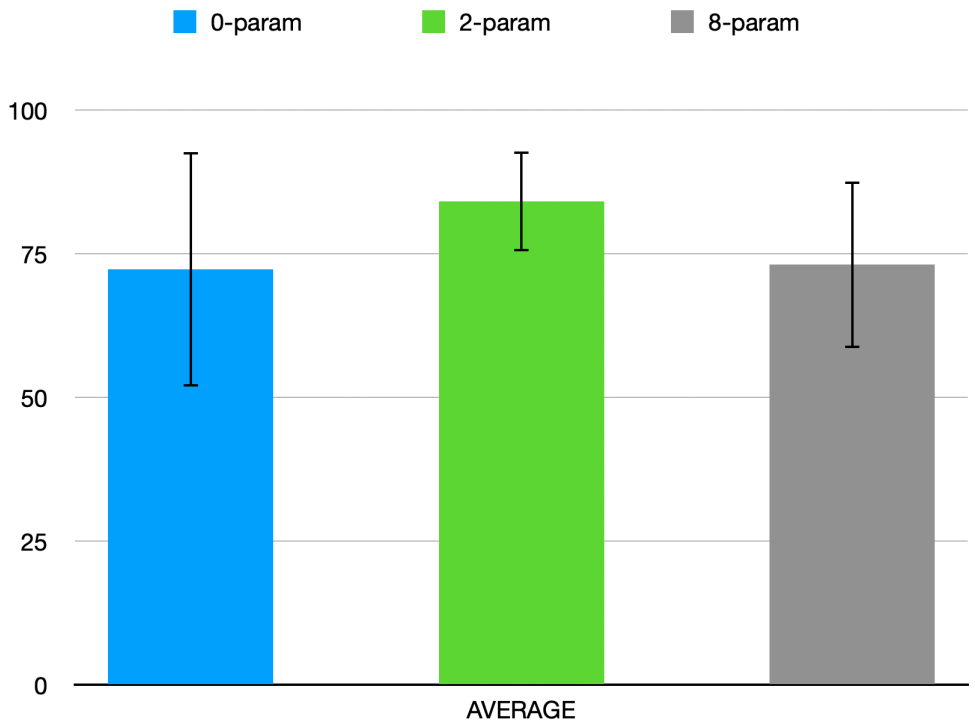


Figure 4.5: Score Bar Chart of Subject 6.

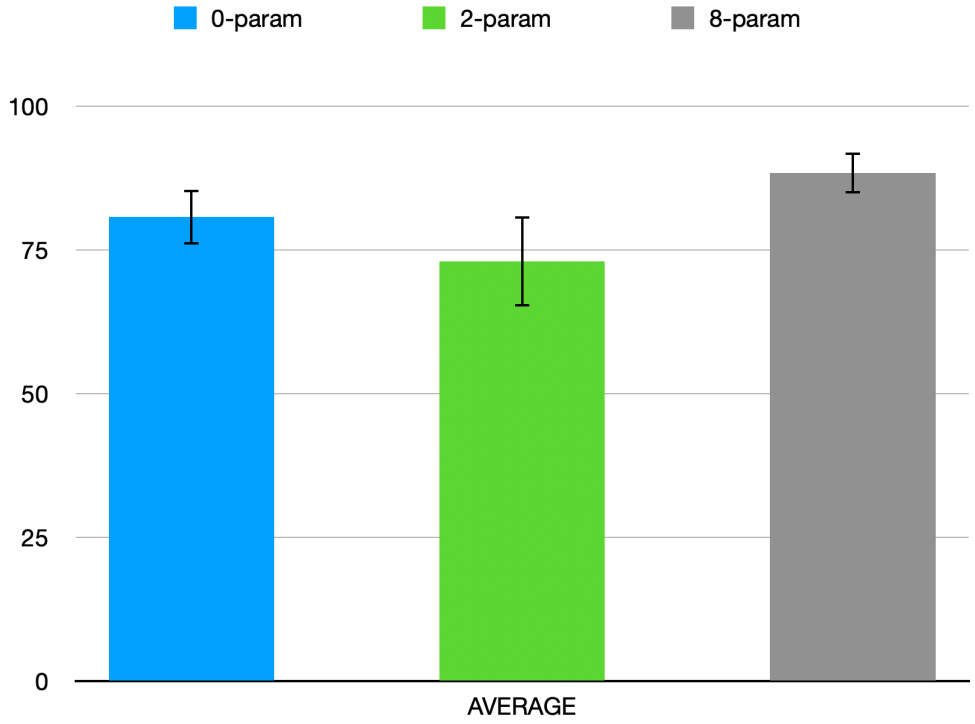


Figure 4.6: Score Bar Chart of Subject 2.

doi:10.6342/NTU202203643

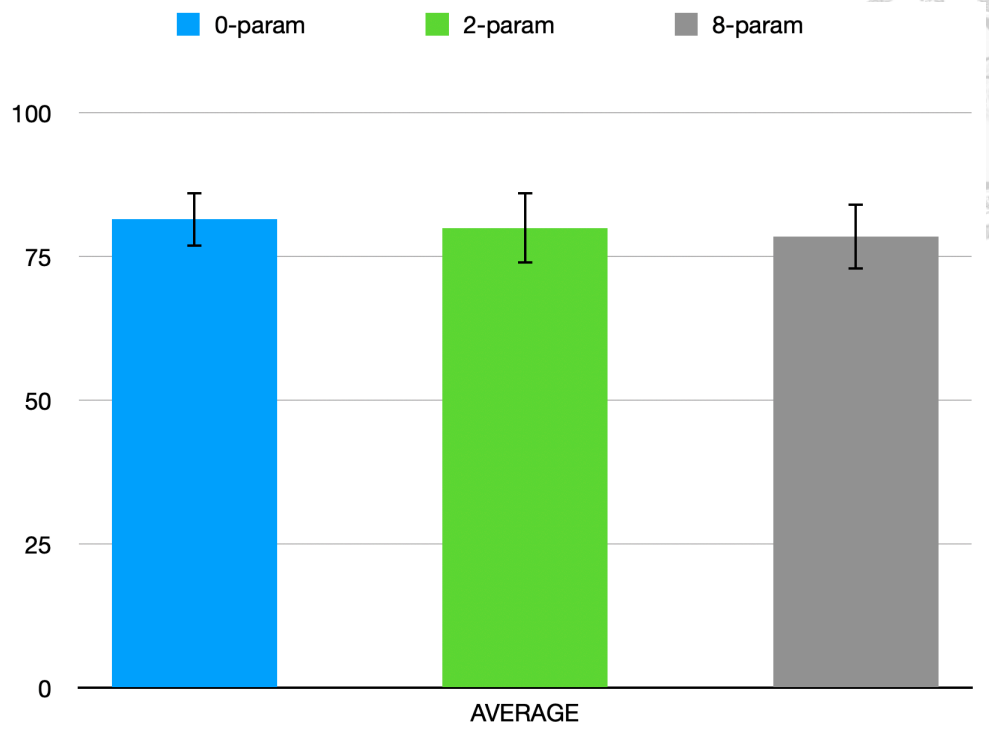


Figure 4.7: Score Bar Chart of Subject 8.

the reasons for these non-theoretical forms.

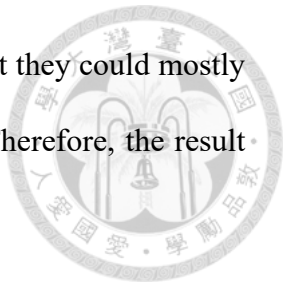
4.4 Discussion

We hope to identify potential correlations through the experimental results and the musical background of the subjects.

4.4.1 Sensitivity to Sound

First, we discuss the 6 subjects who showed incremental form. We found that they all had a strong background of playing musical instruments or producing music. Two of them had more than 10 years of experience playing musical instruments, two had 5-10 years, one had played band, and the last one is a professional musician. In other words, they all have an amazingly great sensitivity to sound. It is not very difficult for them to

doi:10.6342/NTU202203643



listen and adjust the voice. After the blind test, they also reported that they could mostly distinguish the differences between the three conversion methods. Therefore, the result can be in line with the theory.

We then looked at the other 10 subjects who were unable to show incremental form. In contrast to the 6 above, none of the 10 subjects had a strong musical background. Four of them had played musical instruments but had less than 5 years of experience; the others had never touched an instrument and had only attended music lessons at the elementary or secondary level. They may not be able to listen and tune the sound precisely.

Of these 10 subjects, several provided feedback that they were unsure of the results even after a long time of adjustment, and that they would inadvertently make the voice worse and worse. On the contrary, they felt that the results of the zero-parameter filter, which did not need to be adjusted, were already close to their own voices. Moreover, they reported that they were unable to clearly distinguish the differences between the three conversion methods in the blind test. From the above, we know that they deviated from the theory mainly because they have only normal sensitivity to sound.

4.4.2 Statistical Significance

First, we sorted out those 6 people who were very sensitive to sound. We did an analysis of these people using the statistical program SPSS to see if there was a statistical significance between their zero- and two-parameter data, and between their zero- and eight-parameter data.

The results of the analyses showed p-values of 0.044 and 0.004, both of which were below the threshold of 0.05, as shown in Figure 4.8 and Figure 4.9. From the above results,

doi:10.6342/NTU202203643

群組統計量

	參數	N	平均值	標準差	標準誤平均值
分數	0-param	6	67.3333	5.40987	2.20857
	2-param	6	73.4167	5.75688	2.35024

獨立樣本檢定

		變異數等式的 Levene 檢定		平均值等式的 t 檢定							
						顯著性				差異的 95% 信賴區間	
		F	顯著性	t	df	單面 p	雙面 p	平均值差異	標準誤差異	下限	上限
分數	採用相等變異數	.003	.956	-1.886	10	.044	.089	-6.08333	3.22512	-13.26935	1.10268
	不採用相等變異數			-1.886	9.962	.044	.089	-6.08333	3.22512	-13.27310	1.10644

Figure 4.8: Results of the Analysis Between Zero-parameter and Two-parameter for 6 Subjects.

群組統計量

	參數	N	平均值	標準差	標準誤平均值
分數	0-param	6	67.3333	5.40987	2.20857
	8-param	6	77.0833	4.85198	1.98081

獨立樣本檢定

		變異數等式的 Levene 檢定		平均值等式的 t 檢定							
						顯著性				差異的 95% 信賴區間	
		F	顯著性	t	df	單面 p	雙面 p	平均值差異	標準誤差異	下限	上限
分數	採用相等變異數	.275	.611	-3.286	10	.004	.008	-9.75000	2.96671	-16.36025	-3.13975
	不採用相等變異數			-3.286	9.884	.004	.008	-9.75000	2.96671	-16.37080	-3.12920

Figure 4.9: Results of the Analysis Between Zero-parameter and Eight-parameter for 6 Subjects.

we did find a statistical significance between zero- and two-parameter, as well as between zero- and eight-parameter. In other words, for those subjects who were highly sensitive to sound, they were indeed able to effectively distinguish the differences between the zero-parameter filter and the other two methods.

Then, we used SPSS to do a statistical significance analysis on the data of 10 people who had normal sensitivity to sound. Figure 4.10 shows the results of the analysis between zero- and two-parameter, whereas Figure 4.11 shows between zero- and eight-parameter.

Not surprisingly, the p-values of both analyses were far above the threshold value of 0.05. We couldn't find a statistical significance in the data. This also suggests that they are unable to effectively distinguish the differences between the zero-parameter filter and the other two methods.

doi:10.6342/NTU202203643

群組統計量

	參數	N	平均值	標準差	標準誤平均值
分數	0-param	10	74.1800	8.93493	2.82547
	2-param	10	73.6200	9.98786	3.15844

獨立樣本檢定

		變異數等式的 Levene 檢定		平均值等式的 t 檢定							
						顯著性				差異的 95% 信賴區間	
		F	顯著性	t	df	單面 p	雙面 p	平均值差異	標準誤差異	下限	上限
分數	採用相等變異數	.282	.602	.132	18	.448	.896	.56000	4.23781	-8.34331	9.46331
	不採用相等變異數			.132	17.781	.448	.896	.56000	4.23781	-8.35117	9.47117

Figure 4.10: Results of the Analysis Between Zero-parameter and Two-parameter for 10 Subjects.

群組統計量

	參數	N	平均值	標準差	標準誤平均值
分數	0-param	10	74.1800	8.93493	2.82547
	8-param	10	75.7100	9.96064	3.14983

獨立樣本檢定

		變異數等式的 Levene 檢定		平均值等式的 t 檢定							
						顯著性				差異的 95% 信賴區間	
		F	顯著性	t	df	單面 p	雙面 p	平均值差異	標準誤差異	下限	上限
分數	採用相等變異數	.067	.798	-.362	18	.361	.722	-1.53000	4.23140	-10.41984	7.35984
	不採用相等變異數			-.362	17.792	.361	.722	-1.53000	4.23140	-10.42731	7.36731

Figure 4.11: Results of the Analysis Between Zero-parameter and Eight-parameter for 10 Subjects.

doi:10.6342/NTU202203643

Chapter 5 Conclusion and Future Works

5.1 Conclusion

We propose a zero-parameter-filter approach that can quickly convert one' s recorded voice to one' s own voice. To show the feasibility of this filter, we compared it with the two-parameter model and the eight-parameter equalizer in a blind validation experiment. The results showed that the overall average scores were in line with the theory we expected. While the zero-parameter filter is the easiest to use, and has only a slight decrease in overall performance compared to the two-parameter model and the eight-parameter equalizer. For normal people, the zero-parameter filter can give satisfactory results. However, for those who are highly sensitive to sound, they are able to notice the shortcoming of the zero-parameter filter. We will continue to make improvements in the future.

5.2 Future Works

As we mentioned in our conclusion, our proposed zero-parameter-filter approach is relatively ineffective for those who are very sensitive to sound. To improve this, first, we

doi:10.6342/NTU202203643

would like to find a single-parameter model that can outperform the zero-parameter filter and is simpler to use than the two-parameter model. Second, we hope to find multiple possible constant filters by collecting a large amount of user data, that users can hear their own voices by selecting these filters. Then, we expect to find a model that allows the user to input information such as age, gender, body type, and environment. The model could customize a waveform setting for the user's own voice based on the information entered. These are the goals we want to accomplish in the future.

5.3 Potential Applications

Here, we conceive two potential applications using the zero-parameter filter.

Text Reader

We conceived a text reader with one's own voice by combining the zero-parameter filter with a text-to-speech model (SV2TTS) developed by Google's team [5]. The user can record 30 to 60 minutes of their own voice through the reader to obtain a model of the user's recorded voice. After adding a zero-parameter filter to the model, a text reader with the user's own voice is completed. In this way, users can hear their own familiar voices reading various articles or long novels.

Relaxation and Meditation

We also believe that this filter can be used for relaxation and meditation. Listening to instructions is often an important part of relaxation and meditation, such as breathing instructions. We wanted to use this filter to allow users to listen to the breath instructions in the voice they are most familiar with. This may help the users focus and improve their

doi:10.6342/NTU202203643

performance in relaxation or meditation.

doi:10.6342/NTU202203643

References

[1] J. S. Abel and D. P. Berners. Filter design using second-order peaking and shelving sections. *In proceedings of the International Computer Music Conferences*, 2004.

[2] G. V. Békésy. Zur theorie des hörens bei der schallaufnahme durch knochenleitung. *Annalen der Physik*, page 111–136, 1932.

[3] G. V. Békésy. Note on the definition of the term: Hearing by bone conduction. *The Journal of the Acoustical Society of America*, pages 106–107, 1954.

[4] H. Fastl and E. Zwicker. *Psychoacoustics: Facts and Models.* chapter 9. Springer, London, 2007.

[5] Y. Jia, Y. Zhang, R. J. Weiss, Q. Wang, J. Shen, F. Ren, Z. Chen, P. Nguyen, R. Pang, I. L. Moreno, and Y. Wu. Transfer learning from speaker verification to multispeaker text-to-speech synthesis. *Advances in Neural Information Processing Systems*, pages 4485–4495, 2018.

[6] C. Pörschmann. Influences of bone conduction and air conduction on the sound of one’ s own voice. *Acta Acustica united with Acustica*, page 1038–1045, 2000.

[7] L. I. Shuster and J. D. Durrant. Toward a better understanding of the perception of self-produced speech. *Journal of Communication Disorders*, pages 1–11, 2003.

doi:10.6342/NTU202203643

[8] S. Stenfelt. Simultaneous cancellation of air and bone conduction tones at two frequencies: Extension of the famous experiment by von békésy. *Hearing Research*, page 105–116, 2007.

[9] J. Tonndorf. A new concept of bone conduction. *Arch Otolaryngol*, page 595–600, 1968.

[10] S. Y. Won. *Simulating How Humans Hear Themselves Vocalize: A Two-Parameter Spectral Model.* PhD thesis, Department of Music, Stanford University, USA, December 2014.

[11] S. Y. Won, J. Berger, and M. Slaney. Simulation of one's own voice in a two-parameter model. *International Conference on Music Perception and Cognition*, 2013.

doi:10.6342/NTU202203643