International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 21 (2017) pp. 11624-11626

A Study on a Synthesized Synthesis using a Pitch Change Compensation in Speech Signal

Sang-Bum Park[1]
[1]*Information and Telecommunication of Department, Soongsil University, Seoul, Korea.*
369, Sangdo-ro Dongjak-gu, Seoul

Myung-Jin Bae [*2]
[2]*Information and Telecommunication of Department, Soongsil University, Seoul, Korea.*
369, Sangdo-ro Dongjak-gu, Seou

[2]*Orcid Id: 0000-0002-7585-0400*
*(*Correspondence author)*

Seong-Geon Bae [3]
[3]*School of Softwre Application, Kangnam University, Gyunggido, Korea.*

[3]*Orcid Id: 0000-0003-3252-0062*

Abstract

In this paper, as mankind develops, there are people who have won great men and cheers every century. Great people and popular characters can only meet with visual elements, but they are difficult to reach audibly. If already leaving the world to restore his voice no longer able to hear the actual voices could accept more vivid accept their achievements and emotions. In this thesis analyzed for speech reconstruction by comparing the statements of similar sound and the voice of singer Kim Gwang-seok untimely death in 1990 in order to restore the speech signal. The similarity was confirmed by extracting the information about the common formants of the sound source, and the similarity was emphasized by making additional energy. Pitch is applicable to various fields as a structure for synthesizing speech which can be detected in a number of ways, but is detected by using a general method to predict the synthesis here four formant and pitch.

Keywords: Denture, Spectrum, Pronunciation, Formant, Attachment

INTRODUCTION

As mankind develops, there are people who have won great men and cheers every century. In the present generation, there is also a statue built to honor these great men and to engrave them in the heart, and memorial memorials and memorials for great men. But these things have a lot of visual elements. Human beings live enriched by five senses: sight, hearing, smell, taste, touch. The deceased great men can be seen visually, but the public wants to remember the voices of their lives. [1] If they are already out of the world and are restoring their voices that can no longer hear real voice, they will be able to take their achievements and emotions more vividly. [1] They will also be able to fill their voices with memorials or memorial streets that have been left with only visual elements.

In this paper, we analyzed the voice based on the vocal similar to the voice of the singer Kim Kwang Suk, The purpose of the study is to reproduce the voices of great men and popular figures who leave the world, so that their emotions and wills can be grasped clearly because they can be remembered visually as well as audibly. Chapter 2 explains the previous research on restoration of speech signal, and Chapter 3 concludes the experiment and the results finally.

TIME-FREQUENCY DOMAIN PITCH DETECTION METHODS

A voice is the most basic communication method used by humans and has superior characteristics in terms of convenience and economy compared with other methods. In terms of information transmission of speech signals to human beings, promptness and ease make voice a medium of important information exchange, and speech is more important in MMI (Man-Machine Interface) for information exchange between humans and machines. [2] [3] [4] The pitch period is usually known to be in the range of 2.5 to 25ms, and the time domain detection method has a high resolution because the range is 20 to 200 samples even when the speech is sampled at 8 kHz. In the time domain detection method, the pitch is detected by the decision logic after emphasizing the periodicity of the waveform. Periodic emphasis is aimed at

eliminating the effects of formants in the resonance of saints and emphasizing only the fundamental frequencies of excitons. [2] [3] [4] The most widely used method for pitch detection is pitch detection by autocorrelation function. Equation 1 shows the autocorrelation function.

$$\phi(k) = \sum_{n=0}^{N-1} x(n)x(n+k) \qquad k = 0, 1, 2, \ldots \tag{1}$$

The pitch detection in the frequency domain is a method of detecting the fundamental frequency of the voiced sound by measuring the harmonic interval of the voice spectrum. In general, the spectrum is obtained in units of one frame (20 to 40ms), so that the noise is less influenced because noise is fluctuated or background noise is averaged even in this interval. However, the computation is complicated because it requires the conversion process to the frequency domain in the process. Increasing the number of FFT pointers to increase the precision of the fundamental frequency increases the processing time. The spectral harmonics of voiced sound are characterized by the form of peak and valley each time an integral multiple of the existing harmonics is generated, and a slope from the fundamental frequency to the first formant increases. The characteristics of the peak are different in width and shape as the speaker's influence or voice changes. [2] [3] [4]

When the fundamental harmonic is known, the formant spectrum can be passed through Equation 3. Where K is the spectral frequency range 0 -size. The flattened harmonic spectrum obtained from this formant spectrum can be expressed by Equation 4. [2] [3]

$$K_0 = \frac{size}{pitch} \tag{2}$$

$$H(k) = \frac{1}{K_0} \sum_{L+1}^{K_0} S(K-L) \qquad (K = 1, 2, 3, \ldots, size) \tag{3}$$

$$E(K - \frac{K_0}{2}) = S(K - \frac{K_0}{2}) - H(K) \tag{4}$$

We propose a block diagram using the autocorrelation function used in pitch detection and the pitch detection method using harmonic flattening in the frequency domain in time domain. This block diagram can be regarded as a preprocessing process for voice reconstruction after pitch detection.[2] [4][5][6][7][8]

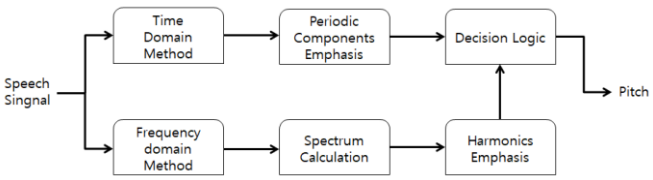

Figure 1: time-frequency domain pitch detection block diagram

EXPERIMENTS

In this study, we restored the speech signal by using similar voice to the existing voice signal. Even if the oral structure is similar, it can be difficult to follow the rhythm, vocal state, and breathing if you adjust to the tone of voice. Conversely, adjusting rhythm, vocal state, and breathing can interfere with maintaining the basic voice tone. The following figures are the result of comparing 5 sentences each of sound sources similar to existing sound sources. As shown in the figure, the fundamental frequency and the first formant are almost similar but the third and fourth formants have different slopes. As a result, as shown in all five sentences, we can see that the rhythm, vocal state, and breathing cannot be uttered in the same way as the conventional sound source by adjusting the voice tone to the existing sound source.

Figure 2: Sound source comparison spectrogram similar to existing sound source

Figure 3: Sound source comparison spectrogram similar to existing sound source

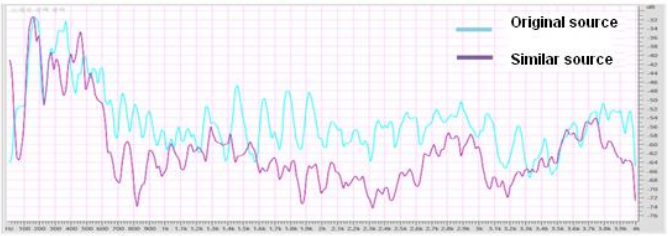

Figure 4: Sound source comparison spectrogram similar to existing sound source

Figure 5: Sound source comparison spectrogram similar to existing sound source

Figure 6: Sound source comparison spectrogram similar to existing sound source

CONCLUSION

As the development of civilization and the quality of life increase in the present generation, cultural facilities and memorials are performed to honor those who have received the greatest or most popular love of the past. The voice has different characteristics and emotions for each person, and he wants to hear the voices of great people and popular characters and to feel warmth. Therefore, in this study, we want to restore the voice of a great person or a popular person. Because the past people cannot say the words that the celadon wants to listen to, he tried to restore the voice through similar sound sources based on the voices left in his lifetime. Even if the oral structure is similar, it can be difficult to follow the rhythm, vocal state, and breathing if you adjust to the tone of voice. Conversely, adjusting rhythm, vocal state, and breathing can interfere with maintaining the basic voice tone. As a result of analysis, it can be seen that the fundamental frequency and the first formant are almost similar to each other but the slopes of the second formant, the third formant, and the fourth formant are different from each other. As a result, we can see that the rhythm, vocal state, and breathing cannot be uttered in the same way as the original sound source. Based on this research, we aim to help voice restoration by similar voice such as pitch control and background noise to restore voice signal.

REFERENCES

[1] M.J. Bae, M.S. Kim, Professor Bae's Sound Story, Gimm-Young Publishers, Inc, 2013

[2] S.G. Bae, M.S. Kim, and M.J. Bae, "A New Hybrid Non-Uniform Coding with Low Bit Rates for Sound Signal in Near Field Communication," 2013, IEEE ICTC, pp. 1039-1040, ICTC 2013.

[3] S.G. Bae, M.S. Kim, and M.J. Bae, "On Enhancement Signal Using Non-uniform Sampling in Clipped Signals for LTE Smart Phones," 2013, IEEE ICCE-berlin, pp.125-126, ICCE-berlin 2013.

[4] Seong-Geon Bae, Won-Hee Lee, and Myung-Jin Bae, " A Judgment of Intoxication using Hybrid Analysis with Pitch Coutour Compare in Speech Signal Processing," IJAER, Vol. 12(2017), pp.2342-2346, 2017.

[5] Seong-Geon Bae and Myung-Jin Bae, " A Study on Recovery in Voice Analysis through Vocal Changes before and After Specch Using Speech Signal Processing," IJAER, Vol. 12(2017), pp.5299-5303, 2017.

[6] Seong-Geon Bae, Myung-Sook Kim, and Myung-Jin Bae, "Using High Frequency Accentuation in Speech Signals as a New Parameter in Intoxication Judgment. Information," An International Interdisciplinary Journal, Vol. 17 (2014), pp.6531-6536, 2014.

[7] Seong-Geon Bae, Myung-Jin Bae, "A New Speech Coding using Harmonics Emphasis Filter," ISAAC 2013, AACL Vol. 1(2013), pp.43-44, 2013.

[8] Won-Hee Lee, Seong-Geon Bae, and Myung-Jin Bae, "A Study on Improving the Overloaded Speech Waveform to Distinguish Alcohol Intoxication using Spectral Compensation," IJET Vol.7, No.5, pp.235-237.