

A Study on Comparison of Pronunciation Accuracy of Soprano Singers

Uk-Jin Song¹, Hyungwoo Park^{1†}, Myung-Jin Bae^{1†}

¹Dept of Information and Telecommunication Engineering, Soongsil University, Seoul, Korea
{imduj, †pphw, †mjbae}@ssu.ac.kr

Abstract

There are three sorts of voices of female vocalists: soprano, mezzo-soprano, and contralto according to the transliteration. Among them, the soprano has the highest vocal range. Since the voice is generated through the human vocal tract based on the voice generation model, it is greatly influenced by the vocal tract. The structure of vocal organs differs from person to person, and the formants characteristic of vocalization differ accordingly. The formant characteristic refers to a characteristic in which a specific frequency band appears distinctly due to resonance occurring in each vocal tract in the vocal process. Formant characteristics include personality that occurs in the throat, jaw, lips, and teeth, as well as phonological properties of phonemes. The first formant is the throat, the second formant is the jaw, the third formant and the fourth formant are caused by the resonance phenomenon in the lips and the teeth. Among them, pronunciation is influenced not only by phonological information but also by jaws, lips and teeth. When the mouth is small or the jaw is stiff when pronouncing, pronunciation becomes unclear. Therefore, the higher the accuracy of the pronunciation characteristics, the more clearly the formant characteristics appear in the grammar spectrum. However, many soprano singers can not open their mouths because their jaws, lips, teeth, and facial muscles are rigid to maintain high tones when singing, which makes the pronunciation unclear and thus the formant characteristics become unclear. In this paper, in order to confirm the accuracy of the pronunciation characteristics of soprano singers, the experimental group was selected as the soprano singers A, B, C, D, E of Korea and analyzed the grammar spectrum and conducted the MOS test for pronunciation recognition. As a result, soprano singer B showed a clear recognition from F1 to F5 and MOS test result showed the highest recognition rate with 4.6 points. Soprano singers A, C, and D appear from F1 to F3, but it was difficult to find formants above 2kHz. Finally, the soprano singer E had difficulty in finding the formant as a whole, and MOS test showed the lowest recognition rate at 2.1 points. Therefore, we confirmed that the soprano singer B, which exhibits the most distinct formant characteristics in the grammar spectrum, has the best pronunciation accuracy.

Key words: Use the word “Keywords” as the title, in 9-point Times New Roman, boldface, initially capitalized, justified, and indented for 2 characters. 4-6 keywords for a subject index should be included in this section. Keywords must be 9-point Times New Roman, multiple-spacing at 1.1. Leave 11-point one single space after keywords, then begin Introduction.

1. Introduction

There are three sorts of voices of female vocalists: soprano, mezzo-soprano, and contralto according to the transliteration. Contralto speaks the lowest vowel voice of female vocalists, mezzo soprano speaks the midrange voice, and soprano speaks the highest vowel voice. Since the voice is generated through the human vocal tract based on the voice generation model, it is greatly influenced by the vocal tract. The structure of the vocal organs differs from person to person, and the voices at the time of vocalization are also different from each other. The spectral spectrum has individual resonance information that occurs in the throat, jaw, lips, and teeth, as well as in the phonological properties of the phoneme. Among them, pronunciation is influenced not only by phonological information but also by jaws, lips and teeth. When the mouth is small or the jaw is stiff when pronouncing, pronunciation becomes unclear. Therefore, the higher the accuracy of the pronunciation characteristic, the more distinctive the characteristic of each vocal tract is. However, many soprano singers do not open their mouths because their jaws, lips, teeth, and facial muscles are rigid to maintain high tones when singing, which makes the pronunciations unclear and thus the character of each vocal organ is also unclear. In this paper, five soprano singers A, B, C, D, and E were selected as experimental groups to confirm the accuracy of the pronunciation characteristics of soprano singers. The analysis was conducted in two ways for the experimental group. First, the gated spectrum was compared and analyzed. Finally, the MOS test was conducted for the population. In Chapter 2, the basic theory of sound engineering, Chapter 3, experimental analysis and results, and Chapter 4 are concluded.

2. Sound engineering basic theory

2.1 Voice generation model

The sound is a sound wave emitted by the air at the stenosis point as the air released from the lungs escapes through the soul. [1] [2] The saints are from the vocal cords to the end of the lips. Figure 1 shows a speech generation model that illustrates the speech generation process mentioned above. Because speech contains both phonological information and personality, it has been studied in many fields such as speech recognition research and analysis of the same person's voice, all of which are based on the voice generation model.

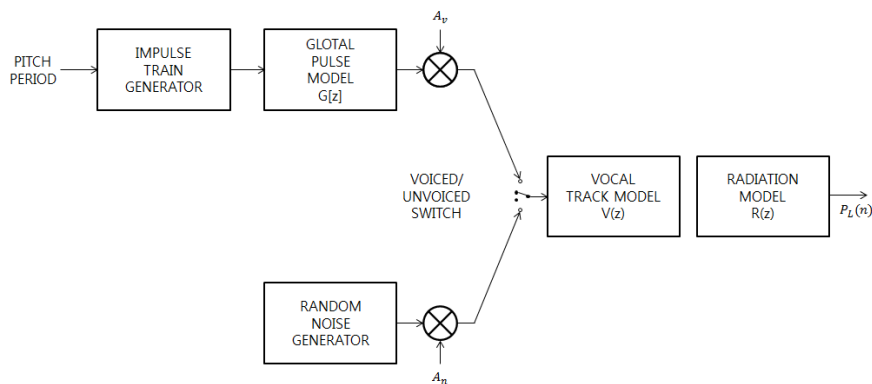


Figure 1. Voice generation model

2.2 Formant

The voices that are generated based on the voice generation model have different frequencies at which resonance occurs depending on different body structures and are referred to as the 1formant, the 2formant, from a lower frequency. The 1formant is the throat, the 2formant is the jaw, the 3formant and the 4formant are

caused by resonance in the lips and teeth. Phonological information of phonemes appears in the 1formant and the 2formant, and the 3formant, the 4formant, and the 5formant appear individually in individual vocal organs. A good pronunciation characteristic for a person's hearing means a vocalization that appears clearly from the 4formant to the 5formant.

3. Experiment and Analysis

For the experiment of this paper, the sound source of ‘A nostalgic Geum-gangmountain’ called by Korean soprano singers A, B, C, D, E was sampled at 8000Hz and quantized by 16bit. Based on the sound source, we analyzed the spectrum of the soprano singer's gospel and conducted a MOS test for Korean populations. In order to investigate the accuracy of pronunciation characteristics, we analyzed mainly the /~ri-un/ vocal part where there is a double vowel section. The dual vowel means that the vowel and the vowel form a syllable and the pronunciation accuracy is low during the Korean vowel. Therefore, the formant characteristic of the vowel is analyzed and the pronunciation accuracy is derived. Figure 2 below shows the spectrum of /~ri-un/ of soprano singer A, Figure 3 shows the spectrum of /~ri-un/ of soprano singer B, Figure 4 shows the spectrum of /~ri-un/ of soprano singer C, Figure 5 shows the spectrum of /~ri-un/ of soprano singer D, Figure 6 shows the spectrum of /~ri-un/ of soprano singer E.

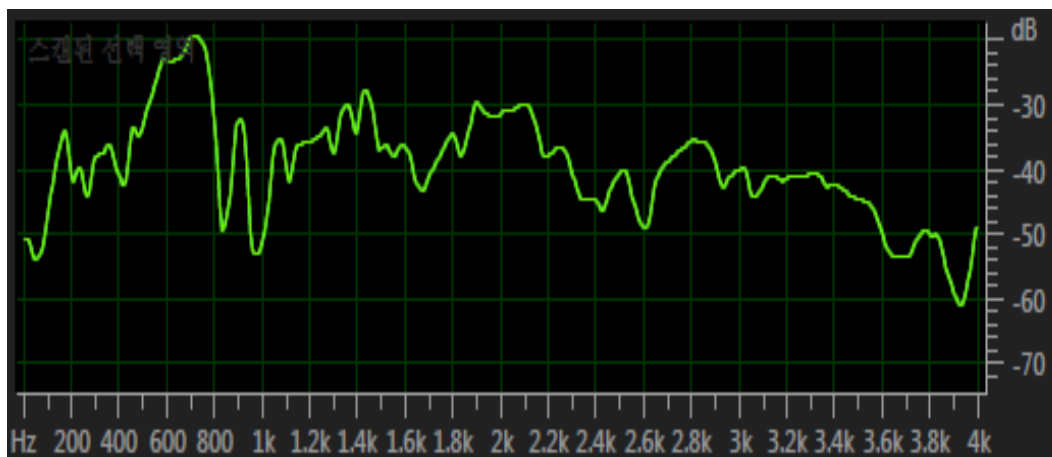


Figure 2. spectrum of soprano singer A

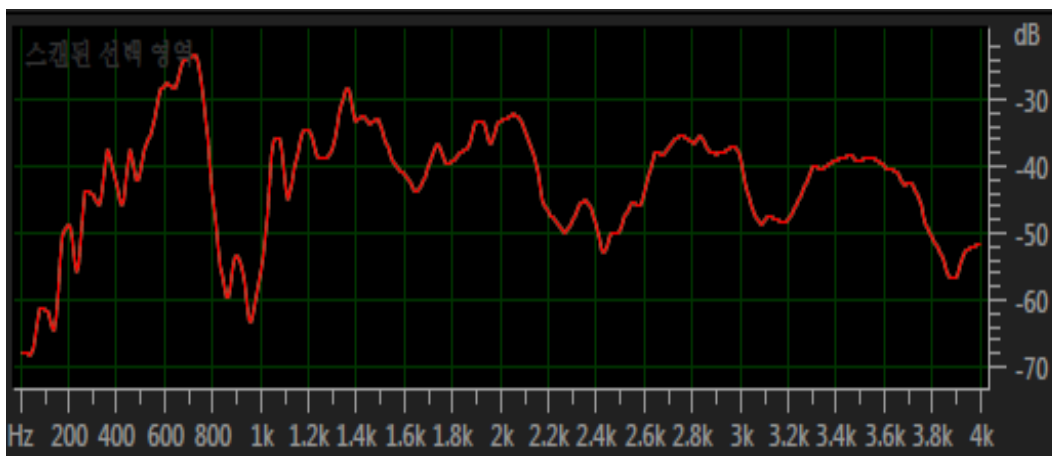


Figure 3. spectrum of soprano singer B

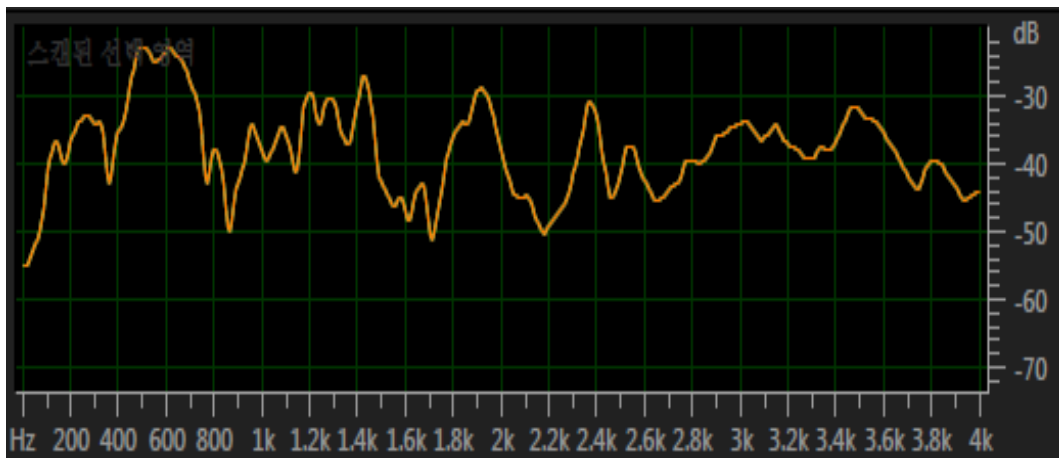


Figure 4. spectrum of soprano singer C

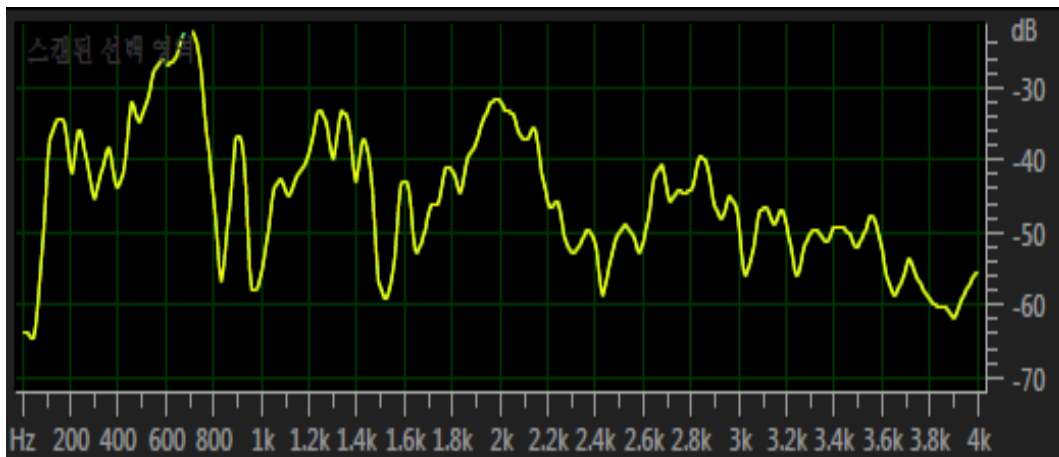


Figure 5. spectrum of soprano singer D

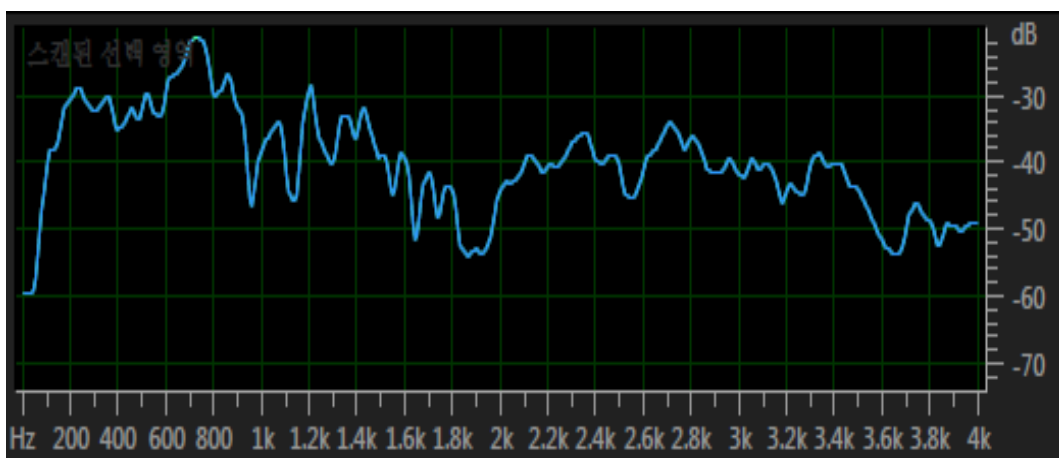


Figure 6. spectrum of soprano singer E

Comparing the formants of the soprano singer A, B, C, D, and E, the soprano singer B's gospel spectrum clearly showed the first to fifth formants. On the other hand, the soprano singer A, C, and D can be seen from the 1st formant to the 3rd formant in the sentence spectrum but it was difficult to find the formants above 2 kHz.

Finally, in the gospel spectrum of soprano singer E, it was difficult to distinguish the formants as a whole.

Finally, we conducted MOS tests on pronunciation accuracies for experimental groups A, B, C, D, and E. Similar to the grammar spectrum analysis, the test was conducted on the /geu-ri-un-geum-gang-san / vocal part where the double vowel /~ri-un/ vowel exists. The soprano singers A, B, C, D, and E were singed to the mother group and 1 ~ 5 points were given according to the pronunciation recognition rate.

Table 1. Result of MOS test

Soprano singer A	3.0 point
Soprano singer B	4.6 point
Soprano singer C	2.9 point
Soprano singer D	3.5 point
Soprano singer E	2.1 point

Table 1 above shows MOS test results for pronunciation accuracy. Soprano B had the highest recognition rate of 4.6, soprano singer D 3.5, soprano singer A 3.0, soprano singer C 2.9 and soprano singer E 2.1. Soprano singers A, C, and D were similar scores. Finally, the soprano singer E, who could not confirm the formant characteristics in the grammar spectrum, showed the lowest recognition rate at 2.1 points.

4. Conclusions

The soprano singer is the highest vocalist among female vocalists. The structure of vocal organs differs from person to person, so it shows the spectral spectrum. Pronunciation characteristics are influenced not only by phonological information but also by chin, lips and teeth. The higher the accuracy of pronunciation characteristics, the more pronounced the characteristics of each vocal tract. The frequency at which the resonance occurs occurs differently depending on the vocal tract, and is referred to as a first formant, a second formant, a third formant, a fourth formant, a fifth formant from a low frequency. The first formant is the throat, the second formant is the jaw, the third formant and the fourth formant are caused by the resonance phenomenon in the lips and the teeth. However, many soprano singers do not open their mouths wide to maintain high treble when singing, resulting in less pronounced pronunciations, and thus the formants characteristic of each vocal organ are also unclear. In this paper, to verify the accuracy of the pronunciation characteristics of soprano singers, five groups of Korean soprano singers A, B, C, D, and E were selected and analyzed. In order to analyze the accuracy of phonetic characteristics, we analyzed the formant characteristics of the grammatical spectrum of the /~ri-un/ vocal part where there is a double vowel section where pronunciation accuracy is low. As a result of the analysis of the spectral spectrum, the soprano singer B 's spectral spectrum clearly showed the first to fifth formants. The soprano singer A, C, and D show the first formant to the third formant, but it was difficult to find the formant above 2kHz. In soprano singer's spectral spectrum, it was difficult to characterize formants as a whole. Finally, MOS test results showed that soprano singer B had the highest recognition rate of 4.6 and soprano singer E had the lowest pronunciation recognition rate of 2.1. Therefore, we found that the pronunciation accuracy of the soprano singer B, which has the most pronounced formant characteristics in the grammar spectrum and the highest pronunciation test result, is the best.

References

- [1] Jong-Guk Kim, Ki-Young Lee, Myung-Jin Bae, "Speech transformation technology", The Magazine of the IEEK Vol.31, No.6, pp.708-718, (2004)
- [2] Guem-Ran Baek, Jong-Jin Yun, Myung-Jin Bae, "A study on the documentary narration voice of famous entertainer", KSCSP Vol.27, No.1, (2010)
- [3] Lawrence Rabiner, Editor, "Theory and Applications of Digital Speech Processing, Pearson Education", (2010)
- [4] Bae, M,J, Lee, S,H., "Digital Voice Analysis"
- [5] Byung-Min Lim, Seong-Geon Bae, Myung-Sook Kim, Myung-Jin Bae, "A Study on Pronunciation Improvement with Denture Sticking", Vol.1, No.3, October (2015)
- [6] Sang-Hwi Jee, Hyung-Woo Park, Myung-Jin Bae, "A Study on Pronunciation Characteristics of Denture Intended for Dentures Users", KSCSP, Vol.33, No.1, (2016)
- [7] Byung-Min Lim, Myung-Sook Kim, Myung-Jin Bae, "A Study on Pronunciation Errors with Different Denture Positions", THE INSTITUTE OF ELECTRONICS ENGINEERS OF KOREA (2016)