

A Study on the Clarity and Cognition of Songs by Soprano Singers - Focused on the Relationship with the Formant Frequencies

Kwang-Bock You¹, Bong-Young Kim² and Myung-Jin Bae^{3*}

¹Soong-sil University, Department of Information and telecommunication Engineering, Seoul, 06978, Korea.

²Soong-sil University, Department of Information and telecommunication Engineering, Seoul, 06978, Korea.
ORCID: 0000-0002-3553-039X

³ Soong-sil University, Department of Information and telecommunication Engineering, Seoul, 06978, Korea.
ORCID: 0000-0002-7585-0400

Abstract:

This paper presents an engineering analysis is performed on whether the sound is audible and perceptual based on the clarity and intelligibility of the signal generated by the soprano singers. In this paper, the signal generated by the soprano singers is analyzed in three organs of the vocal cords, the vocal tract, and the lips that may have a great influence on the articulation of the phonation of soprano singers in the high vocal range. With parameters for comparing the clarity and cognition was utilized the formant frequency and the spectrum. In this study, voice signal algorithm is applied to soprano singer's song in wide frequency range, and pronunciation clarity and cognition is verified. The selected data were analyzed by measuring several formant frequencies and the spectrum of the signal. The analysis includes Korean songs by eight soprano singers "Missing Mt. Geumgangsan" sound sources were used. In this paper, the LPC spectra and the Fourier spectrum analysis are shown to be applicable to Soprano's song, which has a wide range from 250 Hz (C4) to 1300 Hz (E6) in frequency and the severe variation in its amplitude, by comparing with their clarity and intelligibility.

Keyword: Soprano Singer, Clarity, Formant Frequency, LPC spectra, Fourier spectrum

1. INTRODUCTION

Well-trained classic male singers have the spectral peak, is called "the singers formant" between 2 and 4 kHz. In this frequency range, the auditory system of the human being is very sensitive. This has a significant advantage in performances with orchestras with relatively weak power in these frequency ranges. However, for female singers like soprano, this "the singers formant" appears weak or rarely. They have a wide harmonic spacing so that it is known to be difficult to define a formant in the spectrum. The harmonic components make the fruitful musical tones produce richer sounds, but it is not the pitch or the integer multiples of the fundamental frequency [1] [2] [3].

The soprano singer's musical range spans two to three octaves. A singer who can produce notes from A3 or B3 to C6 or D6 is called soprano, and moreover, the soprano singers who can sing beyond C7 or F7 are called coloratura. In this soprano's vocal

range, it is difficult to recognize the exact words of the song, which can be anatomically related to both the larynx and the vocal tract. The lyrics of the soprano's song, which is sung while maintaining a high pitch, are still difficult to understand [1] [2] [3] [4]. This paper presents an engineering analysis is performed on whether the sound is audible and perceptual based on the clarity and intelligibility of the signal generated by the soprano singers. The production of voice signals including songs is the articulation, which is the interaction of each other including various organs such as vocal cords, vocal tract, lips, nasal cavity, oral cavity, etc. [1] [4] [5]. In this paper, the signal generated by the soprano singers is analyzed in three organs of the vocal cords, the vocal tract, and the lips that may have a great influence on the articulation of the phonation of soprano singers in the high vocal range.

Well-trained soprano singers in the musical range between C5 and E5 can modify the vocalization method, which is mainly moving around the lips to transfer the notes from mid-range to high range. There are researchers to study this topic. Also, the soprano singers increase their fundamental frequency (or pitch) to a range of up to 1 kHz by making their mouth wide open. This is a very important vocal technique, which is the process of making the first resonance frequency, R1 (or F1), close to the first harmonic (fundamental frequency), f0. It is R1 (F1): f0 tuning [5] [6]. Here, R1 is the resonant frequency that the sounds (including voice and music) scientists called and F1 is the resonant frequency of the vocal tract which is called by the speech engineers. In this paper, R1 is used as a resonant frequency having constant areas, while F1 is the first formant frequency of the signals. In many songs of soprano singers with a high fundamental frequency (f0), there rises an intelligibility problem. This greatly reduces the clarity and perception of pronunciation (especially vowels) because it tries to shift vocalization with the mouth wide open to make the first resonance frequency region R1 higher than the fundamental frequency f0. Since R1: f0 tuning is an important characteristic of the soprano singers, the problem of the articulation of phonation is raised. As a feature of the speech signal, the formant frequencies F1 and F2 are mainly based on the shape of the vocal tract and affect the cognition of vowels. The problem of the articulation of the soprano singer's phonation inherent in these properties must be the interrelated phenomenon of various organs such as vocal cords, vocal tract,

lips, and etc. In this paper, this phenomenon was analyzed by using the formant frequency, mainly. There were studies that analyzed the phenomena by the resonance frequency of soprano singers with high fundamental frequencies and tremors in the larynx during transition to high notes [6] [7] [8]. The most of preceding studies were based on musical, scientific, or medical (anatomical) analysis. There were not many analysis studies based on signal processing. This study started with measuring the vibration depth of soprano singers' songs on how they touch to people as shown in [9].

This study confirmed whether the algorithms for the analysis of voice signals can also be applied to songs by soprano singers with the wide frequency range and much variation in its amplitude to achieve significant results. The song of soprano singer is very good data of this research because it spans over the wide range of frequency of 2 octaves or more. For the articulation of the phonation of soprano singers based on the clarity and cognition, several formant frequencies were measured and the spectra were analyzed. Comparative studies on the articulation of the phonation may help train soprano singers. Adding scientific and analytical data to the training methods based on empirical, intuitive and innate abilities seems to be expected to provide a systematic new training method.

This paper is organized as follows: In section 2, the algorithms and measuring parameters with respect to the speech production model and analysis methods are explained. Section 3 presents the configuration of the data used in this study and the simulation results. Finally, in section 4, the conclusion of this paper was described and the future works were discussed.

2. SPEECH PRODUCTION MODEL AND ITS ANALYSIS

A telephone transmits the narrowband information of voice from 300 Hz to 3400 Hz. Since the fundamental frequency of human beings, from 130 Hz (male) to 220 Hz (female) is lower than the narrowband of voice, the quality of telephone sound's perception is reduced. It means that the telephone does not transmit the fundamental frequency (f_0) of human beings. In addition, the formant frequencies, as shown in Fig. 1, affect the transmission of voice information depending on the location of each occurrence. F_0 is the fundamental frequency (pitch) that is created primarily in the vocal cord. The first formant frequency, F_1 , occupies generally a constant range of frequencies that change mainly based on the shape of the vocal tract (trachea). The second formant frequency, F_2 , is a frequency that varies with the shape of the oral cavity, most of which is in the higher frequency range than F_1 . The third formant frequency, F_3 , which varies based on how it radiates from the lips and the position of the teeth, is positioned in a higher range than F_1 and F_2 [5] [10]. Since formant frequencies, F_1 , F_2 , and F_3 are not clearly distinct in their roles and positions, so that each frequency does not do their own work, but rather, in most cases, their combinations are able to produce the corresponding unique voice signals.

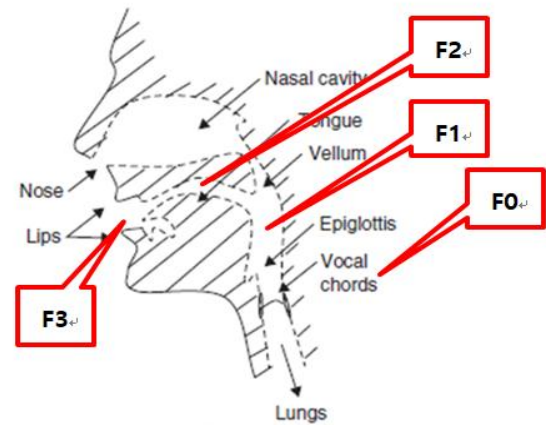


Fig 1. Sectional diagram of human vocal apparatus [5] [10]

The radiation at the lips can be the one of major reason to produce the third formant frequency, so it may have the certain relationships with the articulation of soprano singers, have studied in this paper. In particular, F_3 shows several properties unlike other formant frequencies. First, it does not change from vowel to vowel. Rather, it varies with gender, age, and length of vocal tract. Second, unlike other formant frequencies, there does not change much with the fundamental frequency (f_0) of the voice signal. Third, it appears clearly in the sound of whistling. These properties of F_3 seem to be used for a speaker identifier and to be a good parameter for measuring and comparing variations due to the speaker and the vowel [5] [11] [12].

In this paper, the F_3 of each soprano singer and its associated bandwidths are measured and compared. The Discrete Fourier Transform (DFT) was used to measure this parameter. The N-point DFT can be expressed as equation 1.

$$S(k) = \sum_{n=0}^{N-1} x(n) e^{-\frac{2\pi jkn}{N}} \quad \text{for } k = 0, 1, \dots, N-1 \quad (1)$$

Where $x(n)$ is the time sequence of N-point and $S(k)$ is the corresponding coefficient of the DFT [12].

The Linear Prediction Coding (LPC) spectrum was calculated and compared. The LPC model describes the production of speech signal by a time varying linear filter that characterized the voice tract excited by voiced and unvoiced signals. The system function of this time varying linear filter is represented by equation 2.

$$H(z) = \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (2)$$

Where the order of this filter is p and a_k , which are the coefficients of the filter, are parameters representing the characteristics of the vocal tract. Therefore, the solution of LPC model is to find filter coefficient a_k , in equation 2.

3. SIMULATION RESULTS

3.1 The Sample Data

In this paper, the Korean song "Missing Mt. Geumgangsán" sung by eight professional soprano singers was chosen as the data for analysis. The eight sopranos are famous singers in their 20s and 60s. The data were collected in public spaces so that it is accompanied by the sound of musical instruments. As aforementioned, female singers do not have strong "singing formant" unlike male singers. Therefore, in this paper, the data were sampled at 16 kHz to minimize the sound components of musical instrument. In other words, since only the components up to 8 kHz are analyzed, the influence of the above frequency components, which are mainly produced by the musical instrument, can be minimized. Then, the data were divided into high notes (or relatively loud ones, or high frequencies), mid notes, and low notes and the simulation was performed.

3.2 Simulation Results

The songs of soprano singers, that is, the data $x(n)$, were normalized in the following manner to indicate $\widehat{x(n)}$.

$$\widehat{x(n)} = \frac{x(n) - \mu_x}{\sigma_x} \quad (2)$$

Where μ_x and σ_x represent sample mean and sample variance, respectively [13] [14]. The normalization of the signal by equation 3 is relatively high in reliability of the results of the analysis because it preserves the characteristics of the signal

well compared to the normalization of the signal by the maximum value of the signal analyzed.

Table 1. Measured values of formant frequencies

Data	F1	F2	F3		F4
"Mid_산.wav"	Peak	Peak	Peak	BW	Peak
Soprano A	529	1344	3054	120	6257
Soprano B	521	1179	3206	210	3706
Soprano C	79	1193	3431	160	6179
Soprano D	1180	2993	3515	260	5933
Soprano E	500	1204	3086	30	6473
Soprano F	393	1171	3214	130	5675
Soprano G	806	2996	3880	240	6995
Soprano H	628	1242	3312	360	6360

For the data which "san" is separated from "malg-go gounsán", the formant frequencies and the bandwidth of F3 of each soprano singer were measured and listed in Table 1. Figures 2 and 3 show the frequency spectrum of Soprano G and Soprano H used in the analysis.

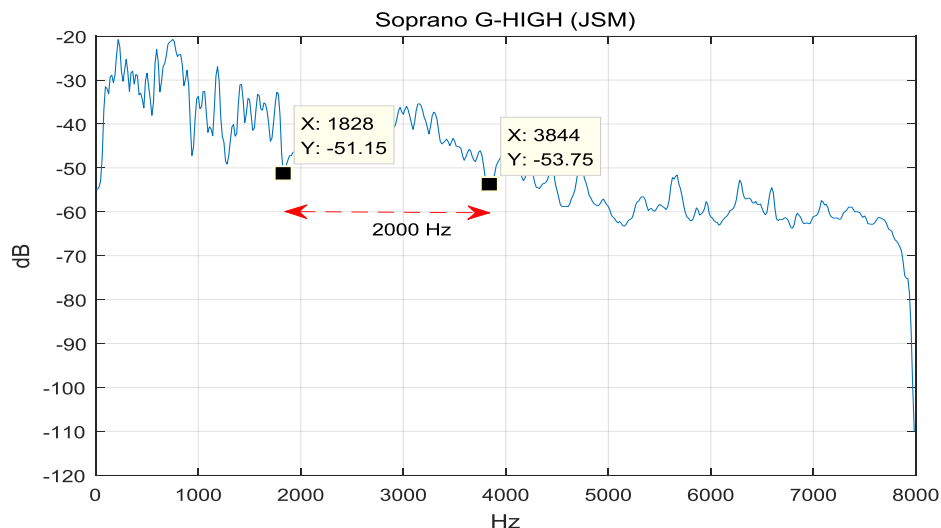


Fig 2. Frequency spectrum of soprano G

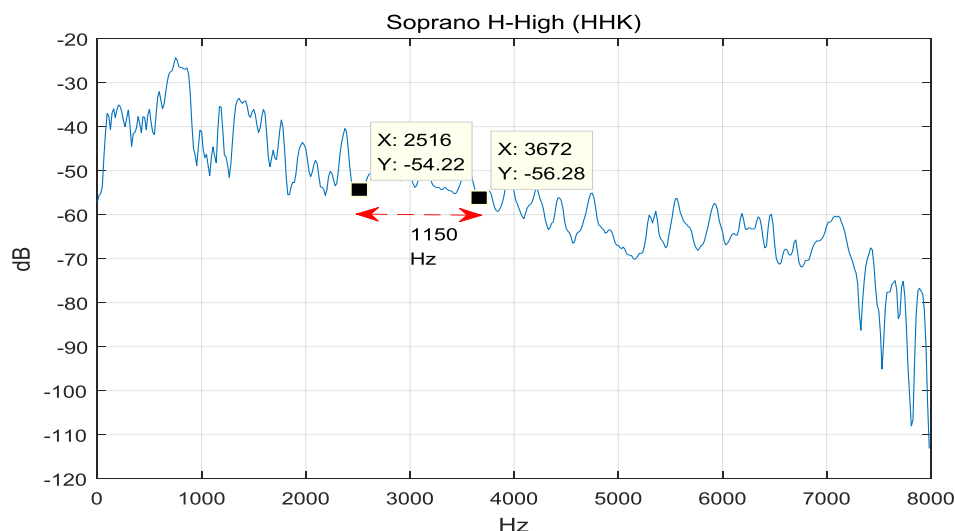


Fig 3. Frequency spectrum of soprano H

The LPC spectrum was extracted to identify the formant from the frequency spectrum of each sound source. From the fact that all F3 values of 8 soprano singers are located between 3 kHz and 4 kHz, it can be explained the relationship between the clarity and cognition, which are described before, and this formant frequency. The LPC spectra of the top three soprano singers with the wide bandwidths and one soprano singer with the narrow one were shown in Figure 4 – Figure 7. As it is seen in the spectrum of Figure 4, F3 is very peaky and its bandwidth

is very narrow. The foreign soprano singer of 20s with the Korean song had the lowest clarity in phonation. In the spectrum of Figures 5 and 6, soprano D and soprano G used the method of transitioning the notes at F3 to widen the bandwidth, whereas soprano H in Figure 7 had the largest bandwidth without changing the notes. Their sounds did not make much difference, but the articulation of phonation of soprano H was relatively clear.

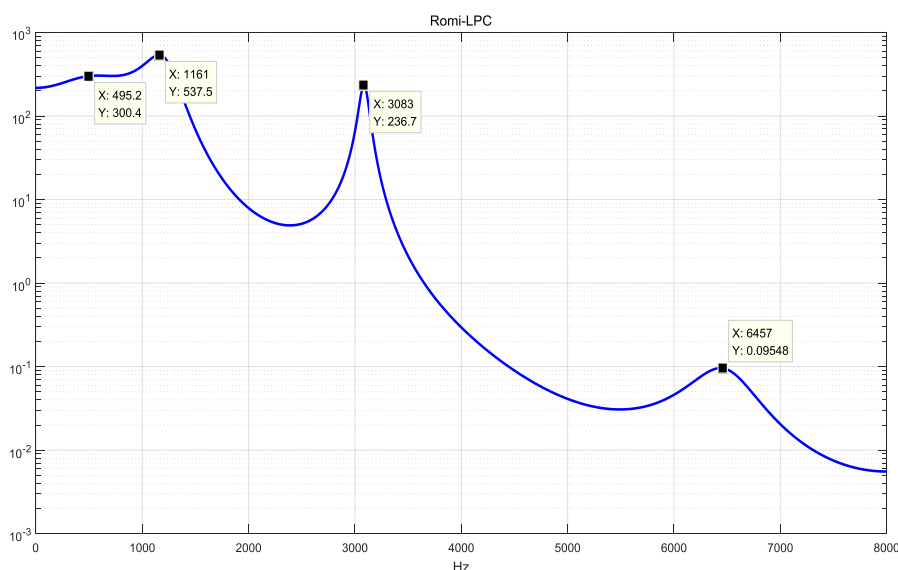


Fig 4. LPC Spectrum of soprano E

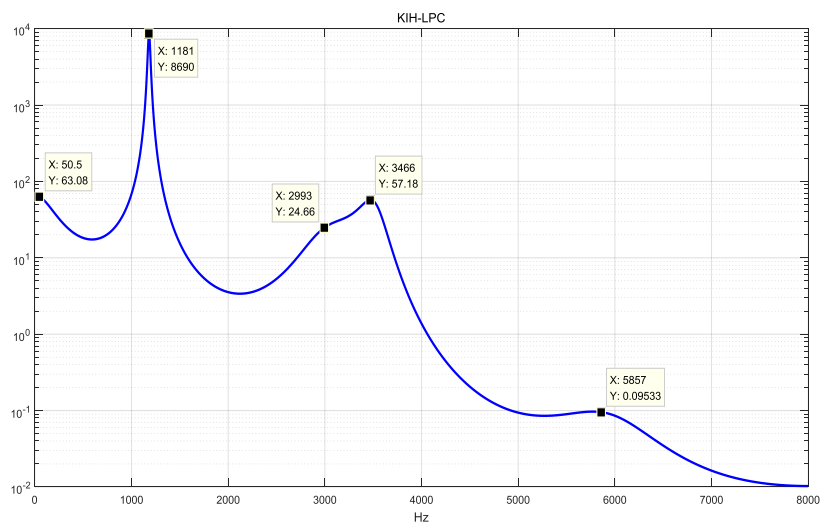


Fig 5. LPC Spectrum of soprano D

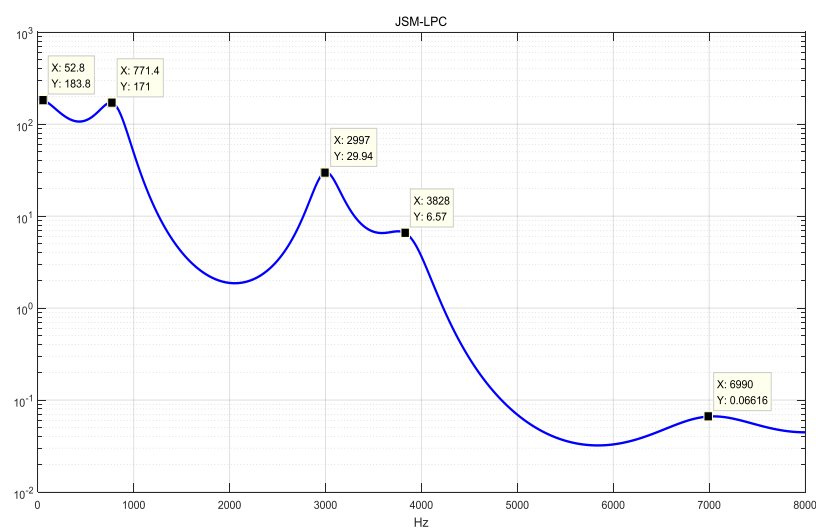


Fig 6. LPC Spectrum of soprano G

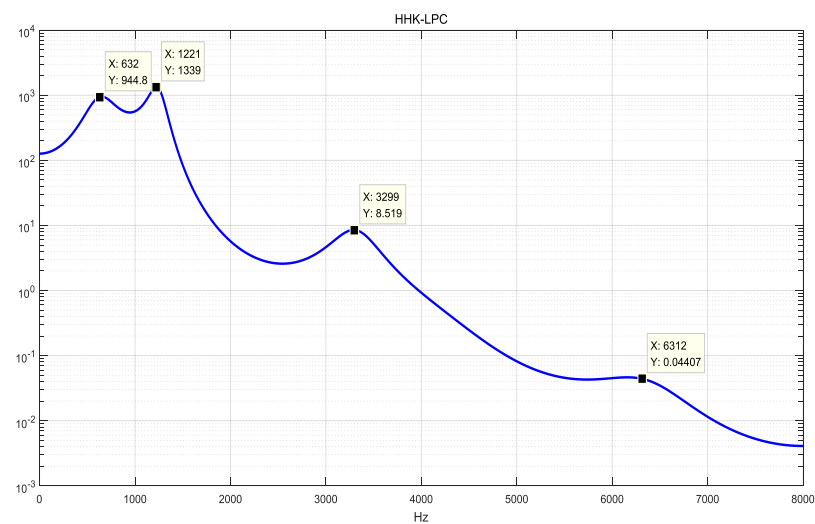


Fig 7. LPC Spectrum of soprano H

The results of Table 1 and from Figure 4 to Figure 7 show that soprano H has relatively good clarity and cognition in the mid-range notes, the group of soprano B, D, and G are good, and the foreigners soprano A and E is the worst as expected. However, even though soprano H has good clarity in pronunciation, the amplitude of the formant F4 is less pronounced than in other singers. Soprano E has a distinct formant F4 compared to other singers. Looking at the formant F4, soprano E can be said to clarify the pronunciation of high notes, such as friction and rupture, even though the clarity of pronunciation is generally poor.

4. CONCLUSION

The results of the measurements in Table 1 show that the clarity and cognition of the phonation of the 8 soprano singers is comparatively related to the bandwidth of the third formant frequency, even though it cannot be said that there are many differences. These results are also shown in the LPC spectra from Figure 4 to Figure 7. The pronunciation clarity of the soprano singer was confirmed by formants F3 and F4 in the LPC spectrum.

The analysis methods used in this paper are shown to be applicable to the songs of soprano, which the frequency ranges span from 250 Hz (C4) to 1300 Hz (E6) and its amplitude varies severely, by comparing the clarity and cognition of their articulation.

For more detailed evaluation of the articulation of phonation, it seems necessary to measure the fundamental frequency (F0) and compare it with other resonant frequencies by using the precise algorithms such as Autocorrelation Function (ACF) and Average Magnitude Differentiation Function (AMDF), in the future.

REFERENCE

- [1] Johan Sundberg, "Articulatory Interpretation of the singing formant", J. Acoustic. Soc. Am., Vol. 55, No. 4, April 1974.
- [2] www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/singer.html
- [3] E. Joliveau, J. Smith, and J. Wolfe, "The tuning of vocal tract resonances by sopranos", Nature, Vol. 427, 116, Jan. 2004.
- [4] M. Garnier, Nathalie Henrich, J. Smith, and Joe Wolfe, "Vocal tract adjustments in the high soprano range", J. Acoustic. Soc. Am., Vol. 127, No. 6, June 2010
- [5] <https://www.animations.physics.unsw.edu.au/jw/voice.html>
- [6] E. Joliveau, J. Smith, and J. Wolfe, "Vocal tract resonance in singing: The soprano voice", J. Acoustic. Soc. Am., Vol. 116, No. 4, Pt. 1, October 2004
- [7] M. Echternach, P. Birkholz, L. Traser, T. Flugge, R. Kamberger, F. Burk, M. Burdumy, and B. Richter, "Articulation and vocal tract acoustics at soprano subject's high fundamental frequencies", J. Acoustic. Soc. Am., Vol. 137, No. 5, May 2015
- [8] M. Garnier, Nathalie Henrich, L. Crevier-Buchman, C. Vincent, J. Smith, and J. Wolfe, "Glottal behavior in the high soprano range and the transition to the whistle register", J. Acoustic. Soc. Am., Vol. 131, No. 1, Pt. 2, January 2012
- [9] Uk-Jin Song, Kwang-Bock You, and Myung-Jin Bae, "A Study on the Measurement of Vibration Depth of Soprano Singers", AJMAHS – HSST, Vol. 8, No. 4, pp. 277-284, April 2018
- [10] L. Rabiner and R. Schafer, Theory and Applications of Digital Speech Processing, 1st Edition, Prentice Hall, 2011
- [11] P. Monahan and W. Idsardi, "Auditory sensitivity to formant ratio: Toward an account of vowel normalization", Language and Cognitive Processes, Vol.25, No. 6, pp. 808-839, 2010
- [12] I. McLoughlin, Speech and Audio Processing, Cambridge University Press, 2016
- [13] Shinae So, Kang Hee Lee, Kwang-Bock You, Ha-Young Lim, Jisu Park, "A Study of Peak Finding Algorithms for the Autocorrelation Function of Speech Signal," of The Korea Society of Computer and Information, Vol. 21, No. 12, pp. 131-137, Dec. 2016
- [14] Shinae So, Kang Hee Lee, Kwang-Bock You, Ha-Young Lim, Jisu Park, "A Study of the Pitch Estimation Algorithms of Speech Signal by Using Average Magnitude Difference Function(AMDF)", AJMAHS – HSST, Vol.7, No. 4, pp. 235-242, April 2017
- [15] Song, U.-J., Ahn, I.-S., Kim, M.-S., Bae, M.-J., "A Study on the Stability of Ultra-High Frequency Vocalization of Soprano Singers", Studies in Computational Intelligence 789, pp. 245-253, 2019
- [16] Lee, W.-H., You, K.-B., Bae, M.-J., "On a study on drinking judgement using phonation characteristics in speech", Journal of Engineering and Applied Sciences 13(5), pp. 1093-1097, 2018
- [17] Park, S.-B., Kim, M.-S., Bae, M.-J., "A study on the high-pitched vocal voice of Lee, Nan-Young", Information (Japan) 20(6), pp. 4063-4070, 2017