International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 3 (2019), pp. 378-384
© International Research Publication House. http://www.irphouse.com

# A Study on the Principles of Human Whistling and its Analysis by Using the Parameters of Speech Production Models

**Sang-Hyu Yoon[1], Kwang-Bock You[2] and Myung-Jin Bae[3*]**
*[1]Soong-sil University, Department of Information and telecommunication Engineering, Seoul, 06978, Korea*
*[2]Soong-sil University, Department of Information and telecommunication Engineering, Seoul, 06978, Korea*
*[3] Soong-sil University, Department of Information and telecommunication Engineering, Seoul, 06978, Korea*
*Orcid Id : 0000-0002-7585-0400*
**Corresponding author*

**Abstract:**

In this paper, the production principles of human whistling and its analysis were carried out utilizing the parameters of the speech production models. The speech production models have used in this paper are the vocal tract model and the source synthesis filter model. In order to ensure the validity of this study, the musical scales are utilized as the experimental data since it can be predicted its results. In this paper, four types of musical scale data were used for the purpose of comparison and analysis. The first is an artificial musical scale composed of sinusoidal signals, the second is a musical scale created by human whistle, the third is generated by human voice, and the last is a musical scale created by real piano. Here, the data by human whistling, voice, and piano are those of noisy environment.

It is shown that the fundamental frequency of human whistling has almost the same properties as that of the Helmholtz resonator. The results of the analysis of the vocal tract speech production model, composed of tubes with non-uniform cross-sectional areas, suggested the possibility of this model as a production model of human whistling. The formant frequency of the speech signal is described by using this modelling. The relationship among the fundamental frequency of human whistling, the formant frequency of the human speech and the resonant frequency of the Helmholtz resonator is explained. The source synthesis filter model, which modeled the speech production with impulse and random signals, and the linear Prediction Coding (LPC) equations which are the solutions of this model have been studied. Using the four types of data aforementioned, we performed computer simulation for these models.

In this paper, the applications using not only the Helmholtz resonator but also the properties of the fundamental frequency of human whistling are presented.

**Keyword:** Vocal Tract Model, Fundamental frequency, Helmholtz resonator, Formant frequency, Resonant frequency

## 1. INTRODUCTION

Human vocal organs can produce a considerable number of sounds such as whistling, singing, and voice. People in the small town of Kuskoy in northern Turkey use whistling like their mother tongue. An ancient Greek historian, Herodotus described that there were humans who are "talking like bats" in Ethiopia. It is known that people in the Canary Islands have spoken the whistling language for at least 600 years [1]. It is, also, known that whistling is widely used in the northwestern areas of Africa, northern part of Laos, and Amazon region of Brazil when the long-distance communication is needed. Recently, in Japan, the whistle blowing contests are held regularly, the winners of such contests teach musical whistling in places like music schools [2].

A whistling is a series of processes that produce a constant flow of air from the lungs and then collect it into the lips and control it using the tongue. While these processes are going on, by using tongue, lips, teeth, or fingers change the flow of air and operate the mouth as a kind of resonance device to get the desired sound. The whistle analyzed in this paper is the labial whistling that produces sound by making small circles with both lips and works well when exhale or inhale. These whistles have some harmonic content, but they are clearly signals with a single frequency or a single tone. Therefore, the fundamental frequency of human whistling is likely to be the same as the resonant frequency of the Helmholtz resonator [2], [3], [4].

Researches for the speech signal production models such as the vocal tract model and source filter model have been consistently studied for more than 40 years, but there are relatively few studies on the principle of whistling production. In this paper, not only the speech signal and whistle signal are analyzed but also the piano signal for the purpose of comparison is analyzed. For comparison purposes, the waveform of each data, the frequency component analysis of the signal, and the spectrum are obtained.

## 2. HUMAN WHISTLING, HELMHOLTZ RESONATOR, AND SPEECH PRODUCTION MODELS

### 2.1 The Principles of Human Whistling Production and Helmholtz Resonator

Unlike voice and song, whistling is not involved with the Larynx or Voice box, which is located at the center of the neck to carry out the function of breathing and vocalization. The whistling is accomplished by a constant flow of air, which is produced starting from the lungs across to the mouth cavity, to the outside by almost periodic inhale and exhale airflow at the lips [4].

It is said that the frequency of the whistle is determined by the frequency of the Helmholtz resonator in the mouth cavity, and that the fundamental frequency of the whistle is very close to the resonant frequency of the Helmholtz resonator. Experiments to demonstrate these have been done using the vocal tract model [2], [3]. However, some whistles are similar to sounds with very high pitch produced by wind instrument using air column response frequency [5].

Helmholtz resonance is a resonance phenomenon that occurs in the cavity or chamber of a bottle when air is injected from the bottleneck. In this resonant phenomenon, the loudest sound is produced at or near the resonant frequency. This resonant frequency is determined by the length of the neck of and the volume of the capacity of the Helmholtz resonator. Musical instruments such as guitar and ocarina are major examples of use. Often it is used to increase the low frequency response of the speakers [6], [7], [8].

$$2\pi F_h = c\sqrt{\frac{A}{VL}} \tag{1}$$

Using equation (1) Helmholtz resonance frequency $(F_h)$ is calculated. Here, A, V, and L are the cross-sectional area of the neck, the volume of cavity, and the length of the neck of the Helmholtz resonator, respectively, and c is the speed of sound in air. It is noticed from this equation that, during whistling, the neck of the resonator formed by the lips is affected, although its geometric shape does not change. The volume of cavity changes by the movement of the tongue so that the whistling can make different tones [9].

Figure 1 below shows a typical frequency response of a Helmholtz resonator. As it is seen the Helmholtz resonant frequency is appeared at 68 Hz.

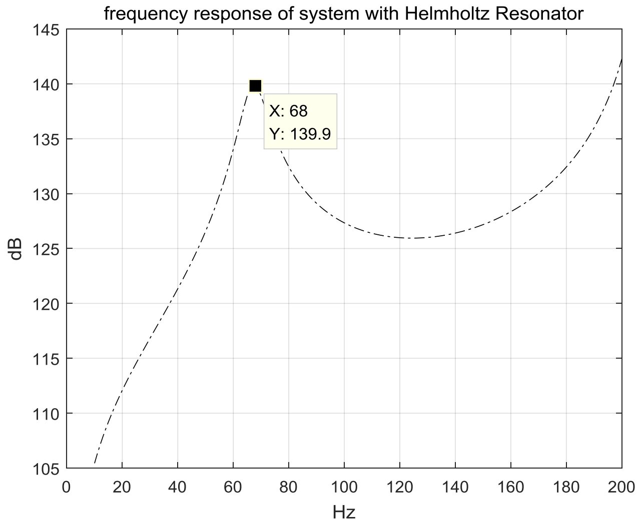

**Fig 1.** Typical frequency response for Helmholtz Resonator

### 2.2 Speech Signal Production Principles and Formant Frequencies

The source synthesis filter model and the vocal tract model are typical examples for modeling of the speech signal production. In fact, the vocal tract between the larynx and pharynx is almost bent at a right angle. However, the vocal tract model is modeled by connecting multiple tubes with non-uniform cross sectional areas. Such non-uniform cross sectional areas modeling allows precise mathematical simulations of sound transmission in the vocal tract. When a signal is transmitted through these tubes, the frequency spectrum is shaped by the characteristics of the tube. This phenomenon is very similar to the resonance effect that appears in the pipe of the wind instrument or organ. These resonance frequencies are called Formant frequencies in the speech signal analysis [5]. Formant frequencies are depending upon the dimension of the vocal tract and its shape such as length and form. This means that the characteristics of the voice signal passing through the tube are varying depending on the length of the tubes and the cross sectional areas of the vocal tract. The vocal tract can be analyzed with a lossless tube model which is connected with both lossless and one-sided-closed acoustic tubes. Therefore, the reflection coefficient at tube of between the $k^{th}$ and $(k+1)^{th}$ can be expressed by the following equation (2).

$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k} \tag{2}$$

And the transfer function between these connections is expressed by the equation (3) in below.

$$V(z) = \frac{\frac{1}{2}(1+r_G)\prod_{k=1}^{N}(1+r_k)z^{-\frac{N}{2}}}{1-\sum_{k=1}^{N}\alpha_k z^{-k}} \tag{3}$$

In the equations, $A_k$, $r_G$, and $\alpha_k$ are the cross sectional areas at the $k^{th}$ tube, the reflection coefficient at glottal, and the prediction coefficients, respectively [10], [11], [12].

The source synthesis filter model is a method to implement speech signal production by using two different signal models, which are the voiced signal (modeling with impulse signal) and the unvoiced signal (modeling with random signal). It can be expressed as a time-varying filter that characterizes the glottal, vocal tract, and lips effects that contribute to the production of speech signal. The transfer function of the time-varying filter is shown in the following equation.

$$H(z) = \frac{S(z)}{X(z)} = \frac{G}{1+\sum_{k=1}^{p}\alpha_k z^{-k}} \tag{4}$$

where $X(z)$, $S(z)$, and $\alpha_k$ are the input signal, the output of the filter, and the prediction coefficients, respectively. The equation (4) can be converted into the time domain as following:

$$s(n) = Gx(n) - \sum_{k=1}^{p}\alpha_k\, s(n-k) \tag{5}$$

The equation (5), which is the difference equation of the Linear Prediction Coding (LPC), shows that the current output $s(n)$ of the synthesis filter is determined by a linear combination of the current input $Gx(n)$ and the past output samples. Therefore, this model becomes a problem of finding the prediction parameters of its linear combination for a given speech signal. The ways to find these parameters are Autocorrelation Function and AMDF (Average Magnitude Difference Function) in time domain, and Cepstrum method in frequency domain. It can be said that speaker and language are not important factors in these analyzes. Nevertheless, it is more important to apply the appropriate parameters to the model using the key features of the speaker or language.

## 3. SIMULATION AND RESULTS

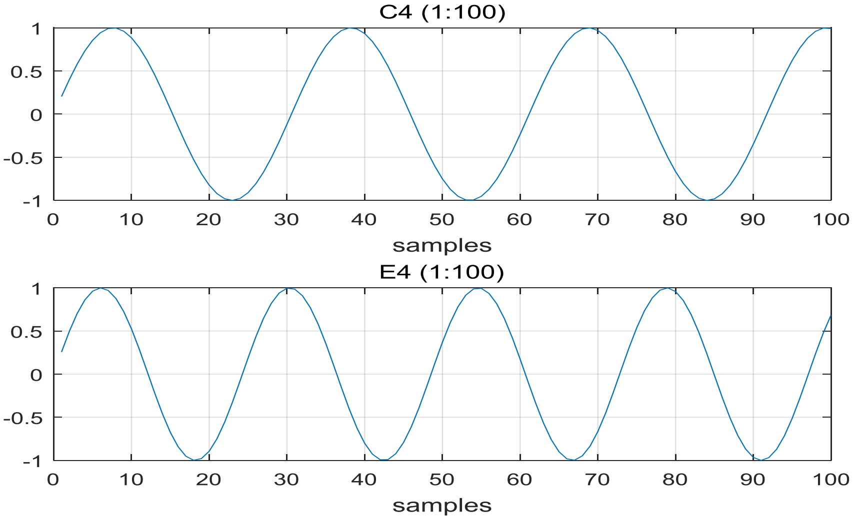

**Fig 2.** Time-Domain waveforms of C4 (Do) and E4 (Mi)

A musical note has its own pitch which is determined by its perceived fundamental frequency of composing sinusoidal signal. There can be a combination of different frequencies existed in a musical note, which affect the timbre of the note. The musical scale is tied to a frequency reference where A4 is set to 440 Hz. This scale defines that the frequency relationship of one semitone is the twelfth root of two. Therefore, since A4 is 440 Hz, one semitone higher -A#4- should be $440 \times 2^{\left(\frac{1}{12}\right)} = 466.16\ Hz$. In this way, based upon the semitone relationship between musical tones, we can calculate the frequency, in Hz, of any note on the musical scale. In figures 2 and 3, the waveforms and frequency components of C4 (261.63 Hz) and E4 (329.63 Hz), which the artificial musical scales composed of sinusoid signal, are shown. It was compared with data – human whistling, piano, and voice in noise environment.

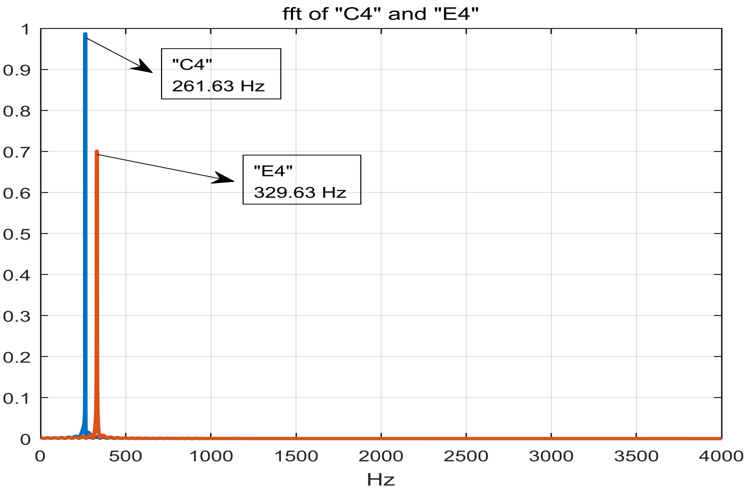

**Fig 3.** The Fourier Transform of "C4 (Do) and E4 (Mi)"

The data used in this paper are “C (Do)/E (Mi)” of the musical scales, and they are compared and analyzed with the signals implemented by whistle, piano, and voice. First, these signals are preprocessed. The first step of the preprocessing is the normalization, which is exploited the mean and variance of the input random signal. In the second, in order to remove the DC component of the signal the high pass filter (HPF) of Equation (6) was used [12].

$$H_{hpf}(z) = \frac{0.946-1.892z^{-1}+0.946z^{-2}}{1-1.889033z^{-1}+0.89487z^{-2}} \qquad (6)$$

Figure 4 is the waveforms in the time-domain that passed the HPF of equation (6). Figures 5, 6, and 7 show the waveforms of the time-domain of the data and its results of the Fourier transform. Figures 7, 8, and 9 below show the waveforms in the time-domain (top) and its corresponding spectra (bottom) for the musical scale "E". Of course, these signals also performed the preprocessing steps in the same way.

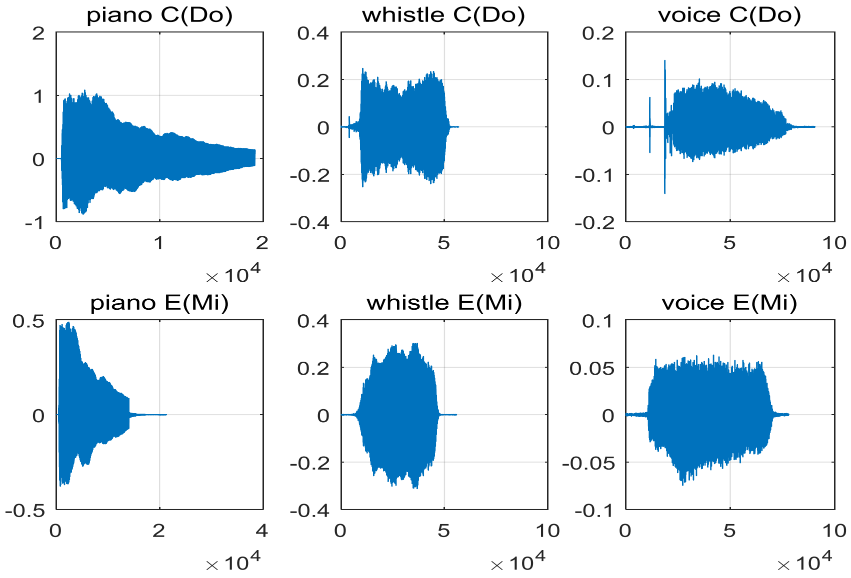

**Fig 4.** Time-Domain waveforms after HPF

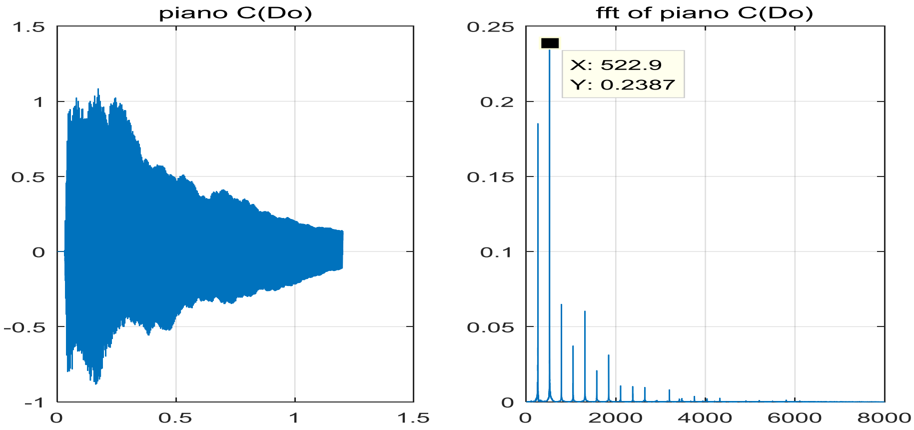

**Fig 5.** The waveform and its Fourier Transform of piano's "Do (C)"

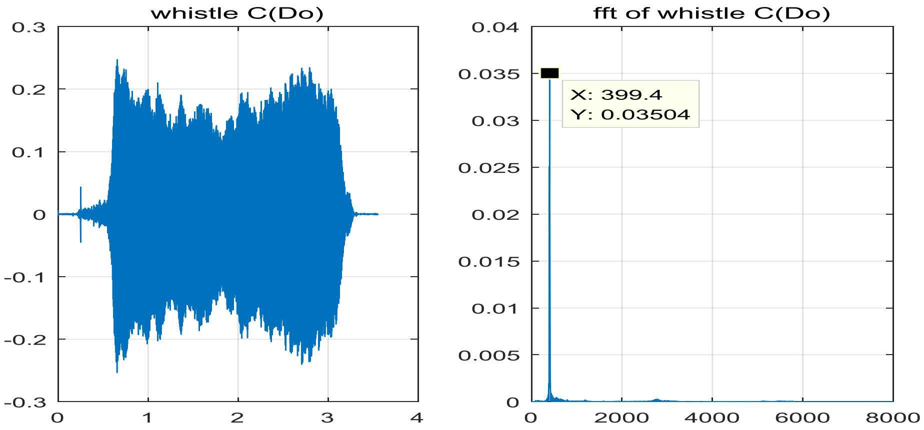

**Fig 6.** The waveforms and its Fourier Transform of Human Whistling "Do (C)"

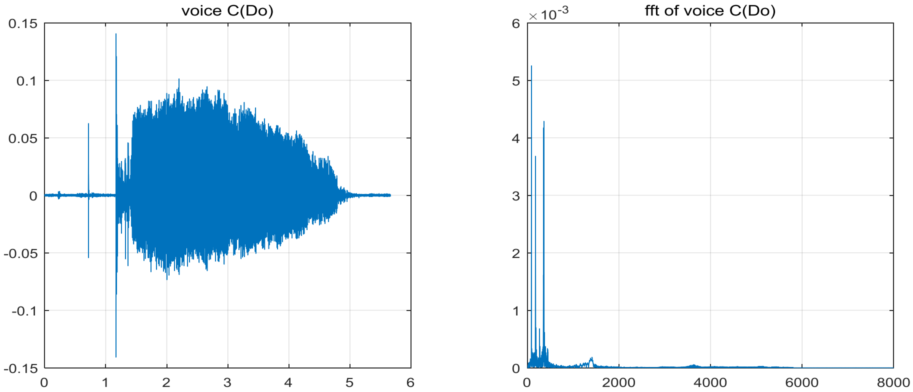

**Fig 7.** The waveform and its Fourier Transform of Voicing "Do (C)"

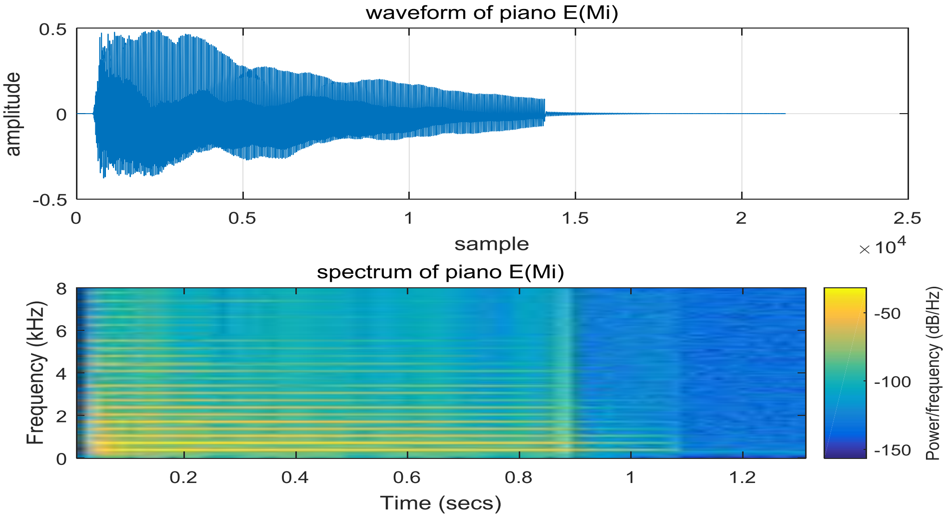

**Fig 8.** The waveform of piano "Mi (E)" and corresponding spectrogram

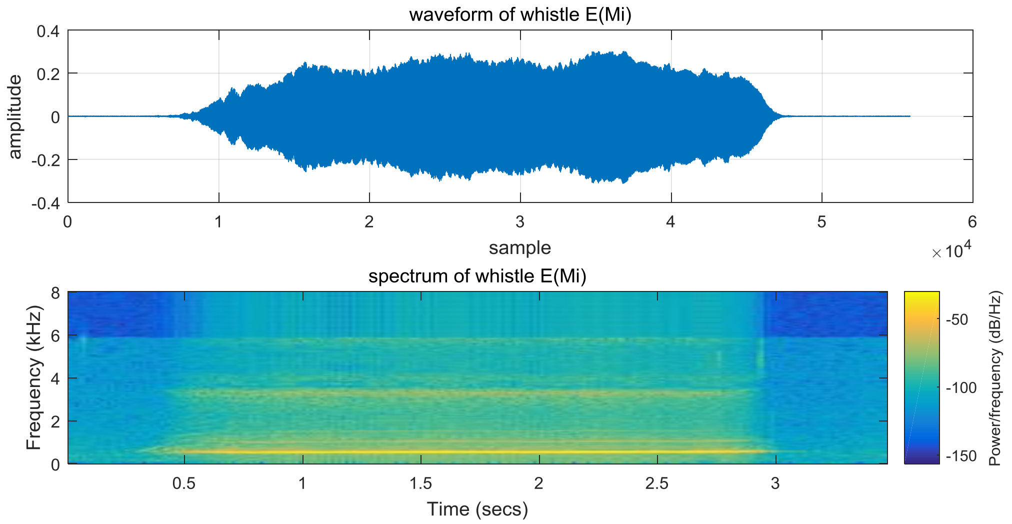

**Fig 9.** The waveform of Human Whistling "Mi (E)" and corresponding spectrogram

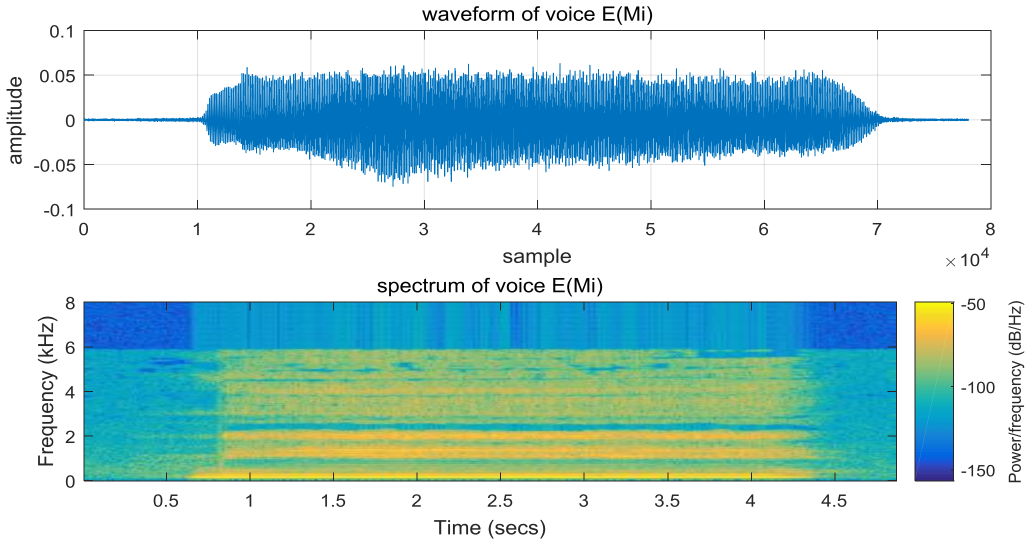

**Fig 10.** The waveform of Voicing "Mi (E)" and corresponding spectrogram

## 4. CONCLUSION

In figure 5, the Fourier Transform results show the dominant frequency at 522.9 Hz, which is equal to the frequency (523 Hz) of the "C (Do)" of the fifth octave in graph of frequency changes corresponding to musical scale and octave. This result ensures the legitimacy of the analytical programs that used in this paper. In figure 6, "C (Do)" in human whistling shows that the fundamental frequency is forming at approximately 400 Hz. "C (Do) of the human voice, shown in figure 7, displays at least three Formant frequencies, as like the general speech signals.

Comparing the spectra of "E (Mi)" in Figures 8, 9, and 10 shows that the signal energy of the piano is collected near the inherent frequencies of the musical scales, as be expected. It is known that the energy of the voicing “E (Mi)” as shown in Figure 10 is gathered at least three places so that the formant frequencies of the signal are formed. It is shown that the spectrum of whistling has the greatest amount of energy in the vicinity of 1330 Hz.

Using the frequencies investigated in this paper - resonant frequency, fundamental frequency, and formant frequency - it can be possible to invent a novel musical instrument useful for everyday life [13], [14]. And, by using the Helmholtz frequency, new musical scales, which might be for a musical synthesizer, can be created.

If a very large sound can be generated in the Helmholtz resonance, a convenient fire extinguisher may be able to invent.

This can be known from the spectrum of whistles that energy is aggregated in one frequency band. Looking at the frequency characteristic of the whistling a very sharp notch at the peak is existed. By using this property it seems to be created a good noise cancelling filter.

## REFERENCE

[1] https://www.newyorker.com/tech/elements/the-whistled-language-of-northern-turkey, August 17, (2015).

[2] T. Shigetomi and Mikio Mori, "Principles of sound resonance in human whistling using physical models of human vocal tract", *Acoust. Sci. & Tech.*, Vol.37, No.2, 2016

[3] R.G. Busnel, *Whistled Languages*, Springer-Verlag, Berlin Heidelberg (1976)

[4] Oliver Keller, *Pitch Detection in Human Whistling Sounds on an Embedded System*, Project Report, Hochschule Karlsruhe Technik und Wirtschaft University (2013)

[5] Seiji Adachi, "Principles of Sound Production in Wind Instruments", *Acoust. Sci. & Tech.*, Vol.25, No.6, 2004

[6] https://en.wikipedia.org/wiki/Helmholtz_resonance

[7] Takayuki Arai, "Education system in acoustics of speech production using physical models of the human vocal tract", *Acoust. Sci. & Tech.*, Vol.28, No.3, 2007

[8] M. Nilsson, Josef Bartunek, J. Nordberg, and I. Claesson, "Human Whistle Detection and Frequency Estimation", *CISP, IEEE computer society*, (2008)

[9] Myonghyun Han, *Sound Reduction by a Helmholtz Resonator*, Thesis Master of Science, Lehigh University (2018)

[10] L.R. Rabiner and R.W. Schafer, *Digital Speech Processing*, Prentice Hall, Upper Saddle River, NJ (2011)

[11] Ian Vince McLoughlin, *Speech and Audio Processing*, Cambridge University Press, New York (2016)

[12] http://newt.phys.unsw.edu.au/jw/voice.html

[13] G.E.Poliner, Daniel Ellis, Andreas Ehmann, E. Gomez, S. Streich, and B. Ong, "Melody Transcription From Music Audio: Approaches and Evaluation", *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 15, No. 4, 2007, pp.1247-1256.

[14] https://en.wikipedia.org/wiki/Whistle_register

[15] You, K.-B., Bae, M.-J., "A study of the variations of energy of the speech signal when english speakers speak Korean", *Information (Japan)*, Vol.20, No.7, 2017, pp. 5035-5040.

[16] Kim, B.-Y., Bae, M.-J., "A study on the sound characteristics analysis of drum-sound rock", *Journal of Engineering and Applied Sciences*, Vol.13, No.12, 2018, pp. 4414-4418.

[17] Park, S.-B., Kim, M.-S., Bae, M.-J. "A study on the high-pitched vocal voice of Lee, Nan-Young", *Information (Japan)*, Vol.20, No.6, 2017, pp. 4063-4070.

[18] Kim, B.-Y., Kim, M.-S., Bae, M.-J., "Study on the Sound Transmission Characteristics of Rescue Signal 'Ya-Ho'", *International Journal of Engineering Research and Technology*, Vol.12, No.2, 2019, pp. 222-226.

[19] Ahn, I.-S., Kim, B.-Y., You, K.-B., Bae, M.-J., "A Study on the Characteristics of an EEG Based on a Singing Bowl's Sound Frequency", *Studies in Computational Intelligence*, Volume 789, 2019, pp. 233-243.

[20] You, K.-B., Bae, M.-J., "Study on energy changes during Korean vocalization of english natives", *Information (Japan)*, Vol.19, No.9A, September 2016, pp.3913-3918.