International Journal of Internet, Broadcasting and Communication Vol.16 No.4 495-501 (2024)
http://dx.doi.org/10.7236/IJIBC.2024.16.4.495

IJIBC 24-4-55

Study on Improving Machine Learning Discriminators using Vocal Parameter of Korean Learners

Kyungnam Jang[1], Kwang-Bock You[2], and Hyungwoo Park[3]

[1] Department Of Korean Language & Literature, Soongsil University, Korea
[2] Electronic Information Engineering IT Convergence, Soongsil University, Korea
[3] Department Of IT-Convergence, Dong-Seoul University, Korea
michael.park@du.ac.kr

Abstract

South Korea has transformed from one of the world's poorest countries into one of its wealthiest. Since the Korean War, the nation has not only elevated its standard of living through technological innovations but has also become a prolific producer of globally popular cultural content. This rise in the popularity of K-culture has attracted learners from various countries to the Korean language. Located strategically between China and Japan, Korea draws numerous foreign language learners, including international students and industrial trainees from countries such as Vietnam and Uzbekistan. Pronouncing Korean accurately poses challenges due to the pronunciation habits rooted in the learners' native languages. Previous research focused on analyzing the pronunciation characteristics of Chinese or Vietnamese speakers and proposed the use of a Support Vector Machine (SVM) discriminator. This study aims to refine the parameters of the SVM's hyperplane to better distinguish pronunciation variations. It introduced research that leverages this discriminator to facilitate more precise Korean pronunciation among non-native speakers.

Keywords: *Korean Learning, Machine Learning, Support Vector Machine, Korean Pronunciation*

1. Introduction

Since the 2020s, Korea has emerged as one of the world's advanced nations. Rapid advancements in science and technology since the 1960s have spurred significant developments in promoting unique cultural content. Known as Hallyu or K-culture, Korean dramas, movies, music, food, broadcasting technology, and entertainment programs have gained global popularity. As Korea spreads its culture and shares its scientific and technological advances with various countries, including those in Asia, the Korean language plays a pivotal role in this dissemination. Mastery of Korean is crucial for a deep understanding of Korean culture.

Furthermore, to facilitate the transmission of cultural and scientific knowledge, industrial trainees, language learners, and international students are pursuing education in various fields within Korea. According to the "Education Statistics Yearbook" by KESS, which provides educational statistics, approximately 153,000 international students, including those enrolled in degree courses and industrial training, were studying Korean in Korea in 2020 [1]. The demographic breakdown shows that Chinese nationals constitute the largest group at 42%, followed by Vietnamese at 24%, and Uzbeks at 6%. For these growing numbers of international students, learning Korean is vital for academic success and effective communication in professional environments [2].

Manuscript received: October. 30, 2024 / Revised: November. 4, 2024 / Accepted: November. 9, 2024
Corresponding Author: michael.park@du.ac.kr
Tel: +82-31-720-2085, Fax: +82-31-720-2085
Assistant Professor, Department Of IT-Convergence, Dong-Seoul University, Korea

Effective learning methods are essential to reduce pronunciation errors in Korean language education. Prior studies on the pronunciation characteristics of foreign learners have primarily focused on aspects that lead to unnatural pronunciation. Given that Vietnamese and Chinese are tonal languages, research has also been conducted on the dynamics of energy changes and rapid formant shifts over brief durations. Specifically, one study [3] examined the characteristics and variations related to the stress of consonants and vowels over short intervals, along with practices to enhance naturalness in pronunciation. It was found that Vietnamese learners often perceive Korean tones through the lens of Vietnamese tonality, leading to unclear Korean pronunciation due to habitual speech patterns from their native language [3].

Further research has been suggested to address stress and compare finality. Another study [4] investigated the pronunciation habits of Vietnamese speakers, particularly focusing on plosive consonants. This research compared the final plosive consonants in Korean and Vietnamese, presenting findings on pronunciation difficulties and unnaturalness among Vietnamese learners. Additionally, a study [5] proposed optimal strategies for identifying and correcting habitual pronunciation errors in Korean learners, specifically targeting the phonemes /ㄴ/and /ㄹ/[5].

These investigations collectively highlight the need for precise linguistic analysis to clarify unclear pronunciations. Research in speech began earnestly in the mid-20th century and by 2020, the capability of computers to listen, respond, inquire, and deliver results through machine learning has significantly evolved. The ultimate goal is to enhance communication among people of various countries and ethnic backgrounds worldwide through natural language processing and automatic translation.

This study will explore elements from the perspective of speech signal processing to clarify Korean pronunciation. By analyzing short-term pitch, formant changes, energy envelopes, waveform alterations, and spectrogram analysis, we aim to identify characteristic differences and develop parameters that will enhance future machine learning performance.

2. Korean Vocalization and Speech Signal Processing Parameters

2.1 During Korean Vocalization /ㄹ/Characteristics of Vocalization

The Korean sound /ㄹ/[liul] is realized with two variants. It is pronounced as a lingual sound between the vocabulary and vowels, and as lingual sidetones. On the other hand, the Vietnamese sound is usually pronounced as a lingual sound [l]. Vietnamese speakers tend to pronounce [liul] between Korean vowels as a lingual sound rather than a sonant sound, which can be seen as a pronunciation error from the phonetic perspective. In addition, in Vietnamese, the sound /l/ is allowed in the initial position of a syllable, but Vietnamese speakers often pronounce it as /n/ when it appears in the first syllable of Korean. This seems to be the effect of the restrictions that do not allow /l/ at the end of syllables in Vietnamese [2].

In this paper, we compared and analyzed how the energy change of Korean /ㄹ/pronunciation differs between Korean speakers and Chinese and Vietnamese learners. In addition, the impact of this change was examined. The measurement of energy in a speech signal is mainly used as the main parameter to distinguish consonants and vowels in the fact that voiced sounds (in many cases, vowels) have more energy than unvoiced sounds. Energy can be calculated by the following (1) equation [6-8].

$$E(n) = \sum_{n=1}^{N} |x(n)|^2 \qquad (1)$$

2.2 speech processing

Voice is the easiest way to communicate. With the development of information and communication

technology, voice can transmit information not only from person to person, but also from machine to person, and between machine and machine. For processing, it is necessary to analyze the voice to find out what components are present and to judge the linguistic meaning of the characteristic. Speech analysis is widely using the linear predictive speech generation model proposed in the mid-20th century. In the LPC model, speech is divided into short quasi-periodic signal intervals. And it is created through the convolutional synthesis in the time domain of the amplitude of the excitation signal in that section and the resonance parameter of the vocal track. And finally, the synthesized voice is radiated into the air through characteristic parameters radiated from the mouth and nose. The linguistic shape and meaning are determined according to the characteristics of the excitation signal in the short section, the characteristic in the vocal tract, and the change in the transfer function of the vocal track in the adjacent section [9].

The pitch of speech signal processing is called the fundamental frequency. Pitch is a quasi-periodic parameter in the short-term analysis of a speech signal, which means the vibration characteristic of the excitation signal, that is, the vibration characteristic for a unit time of the vocal cords. It is a characteristic of the sound produced when the air in the lungs rises and the vibrations of the vocal cords pass through the vocal cords. By accurately detecting these pitches, you can reduce the influence of the speaker who recognizes the speech and change the naturalness and personality of speech synthesis. Also, each person has a different vocal cord and different voices. Also, the range of the pitch frequency varies from person to person, and the characteristics of the change appear differently. Pitch is the part where the energy of the voice signal is prominent, and it is a parameter that can find the change of energy, change of accent, position or change of accent, etc. [10][11].

There is a frequency at which the voice signal generated by the excitation source is amplified along the vocal tract. This is called a formant. Starting with a lower frequency, the names are sequentially increased in the order of f1, f2, f3, and so on. The sound produced by the source here is similar, but the amplitude of the vocal track depends on the thickness, length of the vocal tract, and the shape and location (tongue) of the vocal organs. Among the vocal organs, a lot of oral transfer characteristics appear. And if you look at the overall energy distribution, you can get information such as stress, intonation, and speed of speech, and be evaluated the gradient of energy, which can be analyzed information such as psychology and health. Through the detection and comparison of these analytical parameters, it is possible to determine which vocalization was made and how [12].

2.3 Measuring Method of Voice Signal Pitch and Formant

In the previous chapter, that the pitch and formant of the voice signal are important parameters. The pitch is a part in which quasi-periodic vibration of the voice signal appears a lot and is a part that expresses the characteristics of the excitation signal well. The formant is a part that shows the transmission characteristics of a vocal track well and shows a lot of change in the analysis of short and long sections, and the type of pronunciation is determined according to the location and size of [6][13].

In the case of pitch, a quasi-periodic signal is detected, and analysis is mainly performed in a short section. The short section means a section in which about one phoneme is included for about 30ms per week, and it appears prominently during this section. The pitch is detected in both the time domain and the frequency domain and is seen as a rapidly changing parameter in both domains [11]. In the past, the computational power of computers was not good, so values were extracted by emphasizing periodicity in the time domain. However, even today, even a microcomputer has a module that calculates FFT(Fast Fourier Transform) exclusively, so it is sometimes detected using the position in the frequency domain [12]. In addition, it converts to the cepstrum area to obtain the pitch and format at the same time [13][14].

In the case of formants, it is not easy to detect in the time domain, so detection is mainly performed in the frequency domain. At this time, there is a method of converting a short section voice signal into a frequency

domain and detecting an envelope in the frequency domain by removing the pitch [10][15]. Alternatively, it is also indicated by the position of the line spectrum pair that finds the part of the peak in the frequency domain and expresses the area together [12][16]. Finally, the goal is to record the location and size of the peaks and display them in the form of an overall envelope.

3. Voice characteristics Data and Simulation Results

3.1 Voice Data used for Analysis

In this study, two types of speech data were utilized. The first dataset comprised recordings of Korean native speakers and foreign learners delivering a text that included vocabulary with resonant sounds typical of Korean emotional expression. The foreign participant was a student in his twenties studying Korean at a local university, possessing an intermediate proficiency level in the language. The second dataset involved recordings from learners who had undergone extensive speech training and correction, specifically focusing on the pronunciation of /ㄹ/. Notably, the participant who experienced this learning and correction was a Vietnamese learner from a mixed background of Hanoi and Ho Chi Minh City, highlighting the significant regional pronunciation differences within Vietnam. The voice analyses focused on the pronunciation of /ㄹ/at the beginning and middle of words, assessing their naturalness in sound. Audio signals were recorded at a sampling rate of 16 kHz and quantized at 16 bits. Analysis of these samples was conducted with a 30% overlap at intervals of 16 milliseconds. A short moving window function employing a Hamming window was used to smooth the overlaps across the entire cross-section. The time-domain waveform, energy slope, and spectrogram were analyzed to characterize the frame variations, facilitating a nuanced interpretation of the pronunciation adjustments. This revised version enhances clarity and specificity, detailing the methodology and objectives of the study while maintaining a formal academic tone appropriate for publication or presentation in scholarly contexts. It underscores the technical aspects of the speech data analysis and the educational context of the research, providing a clear overview of the experimental setup and the phonetic focus of the study.

4. Experiments and Result

To verify the accuracy of the speech, pitch detection was performed in the transitional sections using the pitch auto-correlation method. Figures 1 and 2 display the results of pitch detection for short-term speech from Korean and foreign learner speakers, respectively. The upper part of each figure shows the waveform of the short speech section, and the lower part illustrates the results of peak detection using autocorrelation. A red dot marks detected peaks, indicating the pitch at quasi-periodic positions. In the Korean vocalizations, peaks are detected at very regular intervals, confirming periodicity. However, for the Vietnamese speaker's vocalizations, peak detection proves to be problematic. This indicates that the pronunciation of the voiced components within the consonant sounds is inaccurately rendered. Specifically, the Vietnamese speaker's production of /ㄹ/diverges from the Korean pronunciation, as it is articulated as a silent syllable, reflecting the habitual phonetic patterns of the speaker's native language. This revised paragraph refines the explanation of the methodology and findings, clarifying the technical descriptions and outcomes while maintaining an academic tone suitable for a scholarly article. The contrast between the Korean and Vietnamese speakers' pronunciations is highlighted to emphasize the study's focus on phonetic differences and their implications.

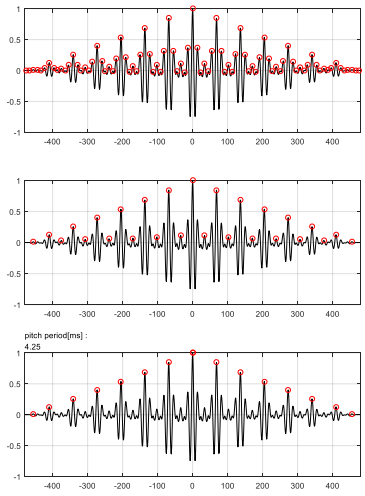

Figure1. Pitch detection results for Korean speakers

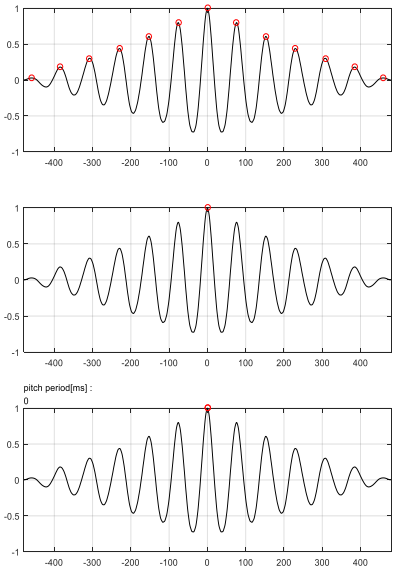

Figure 2. Pitch detection results for foreign learner speakers

As a result of analyzing the /ㄹ/vocalization of foreign Korean language learners, it was confirmed that the intonation, which is difficult to find in the standard Korean vocalization, and the change in formant according to the position of the initial and final voices were found. It can be seen that Korean pronunciation is different according to the habit of vocalizing in the mother tongue. In a previous study, a comparison of the characteristics of Chinese vocalization was performed [6]. This study evaluated how the /ㄹ/vocalization was different in terms of the influence of accents and interactions, such as changes in the overall energy, rather than being centered. To confirm the vocalization characteristics, prepare each word and perform the 10th LPC analysis. The 1st and 2nd formants frequencies of the results were displayed as coordinates and their positions were compared. By analyzing the main first and second formants as coordinates, it quickly and easily indicated the characteristics of your vocalization. Figure_e is the first and second formant

coordinates of a foreign learner speakers, and figure 3(a) and (b) is the first and second formant coordinates of a Korean speaker. The secondary formant of the Korean speaker in the Figure 3(b)was analyzed highly, and the position of the dot on the vertical axis of the picture was raised as a whole. In addition, it was confirmed that the second formant is concentrated in the 2kHz band in the vocalization of the Vietnamese speaker in figure 3(a), and the characteristics of the first and second formants appear close to each other.

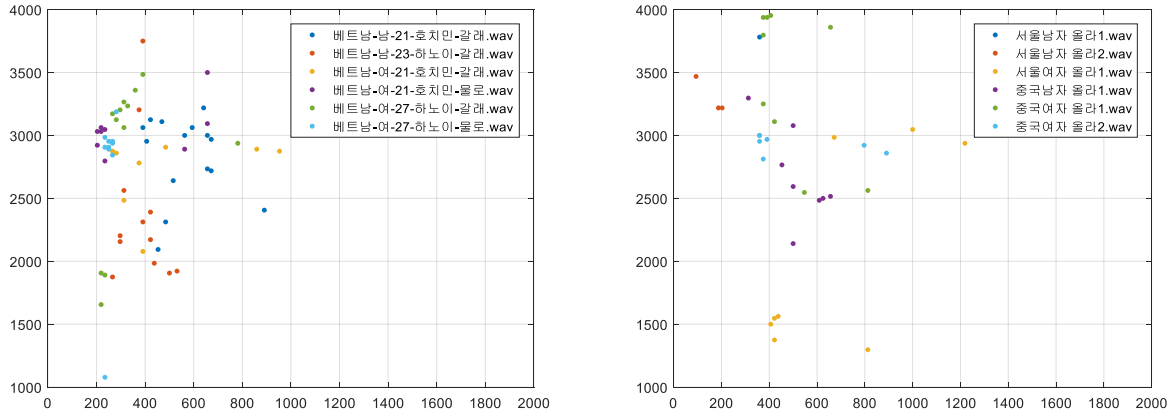

Figure 3 Graph of the change points of the 1st and 2nd formants according to pronunciation(a), (b)

When analyzing the vowel triangle methods through the first and second formants, it was observed that the pronunciation of Korean speakers exhibited variations along the horizontal axis, whereas the pronunciation of foreign speakers showed variations along the vertical axis. These observed changes in parameter orientation within the vowel triangle can be leveraged to enhance machine learning algorithms that are designed to correct pronunciation. This adjustment in the description focuses on clarifying the technical aspects of formant analysis and its application in improving pronunciation correction algorithms. By explicitly stating the directional differences in pronunciation variations between Korean and foreign speakers, the explanation aids in understanding the potential of machine learning applications in linguistic studies. This level of detail is crucial for academic discussions on speech processing and machine learning.

5. Conclusion

Korean cultural content is gaining international popularity, extending beyond Asia. Concurrently, there is a rising interest in learning the Korean language. This study focuses particularly on the phonetic characteristics of Chinese and Vietnamese students, who constitute a significant proportion of those learning Korean through industrial training or study abroad programs. The research involves analyzing and comparing their pronunciation with standard Korean pronunciation, identifying phonological discrepancies, and proposing criteria to mitigate these differences. Foreign speakers often struggle with the stress and speed of Korean pronunciation due to the phonetic habits from their native languages. Specifically, Vietnamese speakers face challenges in correctly articulating /ㄹ/ due to the lack of distinction between /ㄴ/ and /ㄹ/ in their native phonetic inventory. It has been observed that these pronunciation differences can be diminished through repetitive learning aimed at correcting these habitual pronunciations. Notably, many learners are unaware of the unique characteristics of Korean pronunciation when they begin learning as adults, which significantly prolongs the learning process.

Therefore, this study has analyzed the persistent influence of native phonetic characteristics on the Korean pronunciation of foreign speakers, assessed the variations in their pronunciation performance, and implemented guided practice sessions. This approach is anticipated to be an effective educational method for teaching Korean pronunciation to foreign learners. Furthermore, future research is planned to refine Korean pronunciation correction techniques through analogous phonological analysis and the development of machine learning parameters to enhance learning outcomes.

ACKNOWLEDGEMENT

This research was supported by the Soongsil University Research Fund in 2018.

References

[1] Korean Educational Statistics service, “International students by country and school,” Center for Educational Statistics, 2020.

[2] Tae-kyung Kim, "An Optimality-theoretic Approach to Interlanguage: Focused on /l/-substitution of Korean-learning Vietnamese", *Journal of The Korean Language and Literature*, Vol. 80, pp. 31-55, March 2019.

[3] E.J. Kang, Thi-Huong Tran, J.H Cho., " A Study on the Vietnamese Learner’s Korean Intonation phonetic research - Focused on University students Learning the Korean language as a foreign language in Vietnam ." *Korean collection of treatises*, Vol.36, pp.191-219, 2020.

[4] Duyong Lee, "The Production of Korean Stops by Vietnamese Beginner Learners." *Korean Linguistics*, Vol.82, pp.73-94, 2019.

[5] 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," *The Journal of KSCI*, Vol. 21, No.12, 2016.

[6] Savitha Upadhya, "Pitch Detection in Time and Frequency Domain," *ICCICT*, 2012.

[7] L. Rabiner and R. Schafer, *Theory and Applications of Digital Speech Processing*, 1st Edition, Prentice-Hall, 2011.

[8] M. J. Bae, and S. H. Lee, *Digital Voice Analysis*, Seould, Korea: Dongyoung publish, 1998.

[9] H. W. Park, M. S. Kim, and M. J. Bae, “Improving Pitch Detection through Emphasized Harmonics in Time-Domain,” *Communications in Computer and Information Science*, vol. 352, pp. 184-189, 2012.

[10] H. W. Park, S. G. Bae, and M. J. Bae, “Analysis of Confidence and Control through Voice of Kim Jung-un,” *International Information Institute*, vol. 19, no. 5, May 2016.

[11] L. Rabiner, and R. Schafer, *Digital Processing of Speech Signals*, NY, USA: Pearson, 1978.

[12] H.W. Park, A.R. Khil and M.J. Bae, “Pitch Detection based on Signal-to-Noise-Ratio Estimation and Compensation for Continuous Speech Signal,” *ICHIT, Volume 310 of the series Communications in Computer and Information Science*, pp. 767-774, 2012.

[13] Kyungnam Jang, Kwang-Bock You and Hyungwoo Park, "A Study on Correcting Korean Pronunciation Error of Foreign Learners by Using Supporting Vector Machine Algorithm. " *International Journal of Advanced Culture Technology*, Vol.8 No.3, pp.316-324, 2020.

[14] Hyungwoo Park, "Improvement of Sound Quality of Voice Transmission by Finger." *International Journal of Advanced Culture Technology*, Vol.7 No.2, pp.218-226, 2019.

[15] Won-Hee Lee, Hyungwoo Park, Seong-Geon Bae and Myung-Jin Bae, " A Study on the Possibility of Drinking through speech Waveform Compensation in Wireless Communication Environments. " *The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)*, Vol. 17, No. 3, pp.47-53, 2017.