The Journal of the Convergence on Culture Technology (JCCT)
Vol. 8, No. 6, pp.637-646, November 30, 2022. pISSN 2384-0358, eISSN 2384-0366

http://dx.doi.org/10.17703/JCCT.2022.8.6.637

JCCT 2022-11-78

베트남인 한국어 학습자와 한국인의 한국어 겹받침 발음 비교 연구

A Comparative Study on the Pronunciations of Korean and Vietnamese on Korean Syllable Final Double Consonants

장경남*, 유광복**

Kyungnam Jang*, Kwang-Bock You**

요 약 본 논문은 한국어의 겹받침 발음에 대하여 베트남인 한국어 학습자와 한국인을 비교 연구하였다. 언어학적인 연구를 통하여 조사하고 분석한 겹받침 발음에 관한 여러 오류와 제시한 교육 방법에 대하여 공학적 특히 음성 신호처리의 분석 방법을 활용하여서 이런 연구 결과를 확인하였고 이에 우리는 본 논문에서 새로운 교육 방법을 제시하였다. 인공지능의 기계 학습에 많이 활용되고 있는 서포팅 벡터 머신 (supporting vector machine, SVM)을 사용하여서 베트남인 학습자의 발음과 한국인의 발음을 비교하였다. SVM의 초결정 평면을 구할 수 있다는 것은 베트남인 학습자의 겹받침 발음이 한국인의 발음과 차이를 보인다는 것이고, 그 반대라면 발음을 잘하고 있다는 것이다. 본 논문에서 우리가 제시한 새로운 교육 방법은 쓰기와 듣기로만 구성하는 것이 아닌 음성 신호의 시간 영역에서 파형과 그것에 대응하는 신호의 에너지 등과 같은 피교육자에게 보일 수 있는 것들을 포함하는 효율적인 발음 교육 방법이다.

주요어 : 초결정 평면, 한국어 겹받침, 음성신호처리, 서포팅 벡터 머신 (SVM), 파형과 에너지

Abstract In this paper the comparative study on the pronunciation of Vietnamese learners and Koreans for the Korean syllable final double consonants was performed. For many errors and the suggested teaching methods related to the pronunciation of the Korean syllable final double consonants that were investigated and analyzed through linguistic research the results of this study by using the analysis tools of speech signal processing were confirmed. Thus, we suggest the new educational method in this paper. Using SVM, which is widely used in machine learning of artificial intelligence the pronunciation of Vietnamese learners and that of Koreans were compared. Being able to obtain the decision hyperplane of the SVM means that Vietnamese learners' pronunciation of the Korean syllable final double consonants is quite different from that of Koreans. Otherwise their pronunciation are pretty similar each other. The new teaching method presented in this paper is not only composed of writing and listening but is included things such as the speech signal waveform in the time domain and its corresponding energy that can be visualized to the learners.

Key words : Decision Hyperplane, Korean Syllable Final Double Consonants, Speech Signal Processing, Supporting Vector Machine (SVM), Waveform and Energy

*정회원, 숭실대학교 국어국문학과 교수 (제1저자)
**정회원, 숭실대학교 전자정보공학부 교수 (교신저자)
접수일: 2022년 8월 30일, 수정완료일: 2022년 9월 25일
게재확정일: 2022년 10월 10일

Received: August 30, 2022 / Revised: September 25, 2022
Accepted: October 10, 2022
**Corresponding Author: kwangbockyou@ssu.ac.kr
Soongsil University School of Electronic Engineering

Ⅰ. 서 론

한국어의 겹받침은 모두 11종류 - ‘ㄱㅅ’ (/ㄱ/계), ‘ㄴㅈ’, ‘ㄴㅎ’ (/ㄴ/계), ‘ㄹㄱ’, ‘ㄹㅁ’, ‘ㄹㅂ’, ‘ㄹㅅ’, ‘ㄹㅌ’, ‘ㄹㅍ’, ‘ㄹㅎ’ (/ㄹ/계) 그리고 ‘ㅂㅅ’ (/ㅂ/계) - 가 있다. 표준어 겹받침 발음법에서 그것들의 발음을 어떻게 해야 하는지 설명을 한다 [1]. 한국어 겹받침의 발음에서 중요한 것은 자음 앞에서는 겹받침을 구성하고 있는 자음 중 하나를 선택해서 발음한다는 것이다. 또, 겹받침 뒤에 모음이 오면 첫 자음은 앞 음절에 남아서 소리를 내고, 두 번째 자음은 뒤 음절의 첫소리가 된다 [2]. 이런 중요한 두 가지 원칙 이외에도 여러 다른 상황에서의 겹받침의 발음은 달라진다. 이렇게 다양한 겹받침의 발음을 외국인 한국어 학습자는 어떻게 하는지 그리고 한국인의 발음과는 어떤 차이를 보이는지를 알아보고, 발음 교육 방법들을 제시하는 많은 연구가 있었다 [3, 4].

본 논문은 언어학 (국어학)에 기반을 둔 연구들이 조사하고 분석한 겹받침 발음의 여러 오류와 제시한 발음 교육 방법에 대하여 음성학적 분석과 방법들을 활용하여 연구들의 결과를 확인하고 새로운 겹받침 발음 교육 방법을 제시하였다. 본 논문에서는 음성을 분석하는 파라미터로 첫째로 화자의 발화 속도를 알 수 있는 음성 파형을 보았다. 두 번째로는 음성 신호의 에너지 분포를 볼 수 있는 Short-Time Fourier Transform (STFT)으로 분석하였다. 그리고 음성 신호의 주요한 파라미터들인 피치 주기와 포먼트 주파수를 계산하였다. 마지막으로는 Supporting Vector Machine (SVM)이라고 하는 이진 분류기를 사용하여서 외국인 화자와 한국인 화자의 발음을 비교하였다. 본 논문에서 활용한 외국인 화자의 겹받침 발음 데이터는 베트남인 한국어 학습자로 하였다. 2019년의 통계에 의하면 베트남인 학습자는 한국의 전체 유학생의 약 28%를 차지하고 어학 연수생은 전체의 약 60%를 점유하는 것으로 알려져 있다 [5, 6].

베트남어의 음절 구조는 한국어와 같아서 여러 자음이 종성으로 쓰인다. 또한, 한국어와 베트남어 모두 종성에 두 개의 자음을 쓸 수 있다. 하지만 한국어의 겹받침은 베트남어와 다르게 두 개의 받침 중에서 하나를 선택하여 발음한다. 따라서 베트남인 화자들은, 물론 중국인 화자들도 마찬가지지만, 겹받침을 발음할 때 뒤의 자음을 발음할 때에도 앞 자음을 발음하는 오류를 보이기도 하고 종성 발음을 생략하기도 한다 [1, 2].

일반적으로 베트남인 학습자들이 한국어를 학습하는 데 몇 가지 중대한 오류를 보인다. 첫 번째로는 초성에서는 폐쇄음, 마찰음, 그리고 파찰음에서 오류를 보이고, 종성에서는 /ㄹ/을 /ㄴ/로 대체하여 발음하는 오류가 있다. 두 번째는 모음과 /ㄴ/ 사이에 약한 유음을 삽입하여 발음하는 오류를 보인다. 세 번째는 베트남어는 표기로 두 개의 자음이 있는데 그 두 자음이 초성에 있을 때와 같은 한 음소로 나타난다. 그래서 베트남어가 모국어인 화자가 한국어를 배울 때 종성의 구조 때문에 어려움을 겪을 수 있다. 그리고 한국어의 겹받침에서 뒤 자음을 발음하는 ‘삶’과 ‘읊다’에서 많은 베트남 화자들은 앞 자음을 발음하는 오류를 보인다고 한다 [1-4].

한국어 표준어 겹받침 발음법에 따르면, ‘ㄱㅅ’, ‘ㄴㅈ’, ‘ㄹㅂ’, ‘ㄹㅅ’, ‘ㅂㅅ’의 겹받침은 어말 또는 자음 앞에서 각 [ㄱ, ㄴ, ㄹ, ㅂ]으로 발음한다. 그러나 ‘밟-’은 자음 앞에서 [밥]으로 발음하고, ‘넓-’은 ‘넓죽하다 [넙쭈카다]’와 같은 경우 [넙]으로 발음한다. 겹받침 ‘ㄹㄱ’, ‘ㄹㅁ’, ‘ㄹㅍ’은 어말 혹은 자음 앞에서 [ㄱ, ㅁ, ㅂ]으로 발음한다. ‘ㄹㄱ’은 ‘ㄱ’ 앞에서 같은 ‘ㄱ’을 탈락시키고 [ㄹ]로 발음한다 (예, 맑게 [말께], 묽고 [묵꼬], 읽거나 [일꺼나]). 겹받침 뒤에 모음이 오면 첫 번째 자음은 앞 음절에 남아서 소리를 내고, 뒤 자음은 뒤 음절의 첫소리로 발음된다. 표기상 11개의 겹자음이 받침에 올 수 있지만, ‘ㄴㅎ’과 ‘ㄹㅎ’을 제외하면 받침에서 두 자음은 모두 발음될 수 없다. 겹받침 ‘ㄱㅅ’, ‘ㄴㅈ’, ‘ㄴㅎ’, ‘ㄹㅂ’, ‘ㄹㅅ’, ‘ㄹㅌ’, ‘ㄹㅎ’ 그리고 ‘ㅂㅅ’ 인 경우에는 앞 자음이 발음되고, 겹받침이 ‘ㄹㄱ’, ‘ㄹㅁ’, ‘ㄹㅍ’ 인 경우에는 뒤 자음으로 발음된다. 다만, ‘ㄹㅍ’의 경우에는 뒤 자음이 ‘ㅍ’이 중화 현상으로 [ㅂ]으로 발음된다 [1, 2].

베트남 학습자의 겹받침 발음의 오류를 분석하기 위해서 겹받침의 발음을 쓰기를 하여 분석하였다 [1]. 이 분석의 결과를 보면, 겹받침을 발음할 때 앞 자음을 발음하려는 경향이 높아서 뒤 자음을 발음하는 ‘ㄹㅂ’, ‘ㄹㄱ’, ‘ㄹㅁ’, ‘ㄹㅍ’ 계열의 단어들에서 많은 오류를 보였다고 한다. 상대적으로 겹받침 ‘ㄱㅅ’, ‘ㄴㅈ’에서는 오류가 적었다 [1].

겹받침 발음 교육을 위한 몇 가지 구체적인 제안을 하였다 [2]. 그 첫째는 한국어의 겹받침은 음절 말이나 자음 앞에서는 두 자음 중 하나만 발음한다는 사실을 인식하게 하는 것이다. 두 번째는 ‘ㄹㅂ’에 대한 발음 교육이다. ‘ㄹㅂ’은 원래 앞 자음으로 발음되지만 /밟다/, /넓

둥글다/, 넓적하다/ 등 몇 가지 단어에서는 뒤 자음이 발음돼 베트남 학습자들이 혼란스러워하므로 이 발음에 대한 반복 연습이 필요하다. 세 번째는 'ㄹㄱ'의 발음 교육이다. [ㄱ]으로 발음하는 것이 원칙이지만 겹받침 뒤에 'ㄱ'이 오면 [ㄹ]로 발음한다. 이 발음은 한국인 화자들도 자주 범하는 오류이므로 연습이 필요하다 [2].

이상과 같이 한국어 겹받침의 발음에 대한 베트남 학습자들이 많이 범하는 오류들과 그에 대한 교육 방법들에 대하여 주로 [1-4]에서 검토하였다. 이 논문들에서 주장하는 것은 모두 언어학적 조사와 분석에 근거를 둔다. 본 논문은 이러한 오류들과 분석들을 신호처리의 방법을 활용하여 음성학적 실험을 수행하여 확인하고 제안하는 발음 교육 방법에 대한 근거를 제시하고자 한다.

본 논문의 구성은 다음과 같다. 다음 장에서는 본 논문에서 활용한 음성 신호의 다양한 파라미터들에 대한 그 의미와 수학적 표현에 관하여 기술한다. 특별히, 본 논문에서는 인공지능 연구에 많이 활용하고 있는 SVM을 결과를 제안하는 분류기로 사용하였다. 3장에서는 본 논문에서 사용한 음성 데이터와 시뮬레이션의 과정을 설명하였다. 그리고 4장에서는 시뮬레이션의 결과들을 보였다. 마지막으로 5장에서 본 논문의결론을 의논한다.

II. 음성신호 파라미터

1. 자기 상관 함수 (Autocorrelation Function-ACF)

준주기적 (Quasi-periodic)이고 시간에 따라 변하는 시변 신호인 음성 신호에서 주기적인 특성을 나타내는 유성음 (Voiced)의 피치 주기 (F0)를 추정하는 것은 일반적으로 매우 어려운 일이라고 알려져 있다. 피치 주기를 추정하는 대표적인 방법으로는 본 논문에서 사용한 ACF 방법이 있고 이외에 Average Magnitude Difference Function (AMDF)과 Cepstrum FUnction (CF) 알고리즘이 있다. ACF와 AMDF는 시간 영역에서 알고리즘을 수행한다는 점에서 같은 방식이다. 반면에CF는 주파수 영역에서 피치 주기를 계산한다. 아래의 식은 ACF의 정의식이다.

$$R(m)=\frac{1}{N}\sum_{n=0}^{N-1-m}x(n)x(n+m),\ (0\le m\le M_0) \qquad (1)$$

(1) 식에서 m은 지연 (lag 혹은 delay)이라 하며 이 함수는 $m=0$에서 최대값을 갖는 $R(m)=R(-m)$인 우함수이다 [7, 8].

음성 신호 생성 모델의 관점에서 보면 피치 주기는 성문 (vocal cords)에서 생성되는 기본 주파수 (Fundamental Frequency)로 각 개인의 특징을 나타내는 파라미터이다.

2. 포먼트 주파수 (Formant Frequencies)

선형예측 코딩 (Linear Prediction Coding, LPC)는 현재의 시간 n에서의 음성 신호는 과거 $(n-1)$까지의 음성 신호들의 선형 조합 (Linear Combination)으로 근사할 수 있다는 것이다. 선형예측 코딩은 음성 신호의 생성 모델에서 중요한 기관인 성도 (vocal tract)를 선형화하여 음성 신호 생성을 모델링 했다고 할 수 있다. 이 선형화를 통해서 발생 되는 주파수들을 포먼트 주파수라고 하며 이론적으로는 LPC 알고리즘에서 사용하는 방정식의 차수에 의존한다. 차수가 10차인 LPC 방정식에서 실제적으로 3-4개의 포먼트 주파수가 보인다. 첫 번째 포먼트 주파수 F1은 기관지 (trachea)의 모양에 따라서 변동하며, 두 번째 주파수 F2의 변화는 구개 (oral cavity)의 모양 (shape)에 의존한다. 세 번째 포먼트 주파수인 F3는 입술의 방사로 인해 생성된다고 한다 [7, 8].

시변 디지털 필터의 출력 $s[n]$은 입력이 유성음이면, 여기 신호 (excitation signal) 가 성대의 진동 (피치)를 만들고 이 떨림 (진동)은 성도를 지나면서 소리를 만든다. 반면에 무성음일 경우에는 랜덤 여기 신호 (random excitation signal) 가 성도를 따라서 소리를 생성한다. 성도의 반응을 시변 시스템 (필터)로 설정하면 필터의 계수들은 음성생성의 주요한 파라미터가 된다. 이 시스템에서 계수를 구하는 것은 성도의 특징을 나타내는 필터 계수를 추출 해내는 과정이 된다. 정상상태 시스템 (steady state system) 함수로 이루어진 디지털 필터의 전달함수 (transfer function)은 식 (2)와 같이 pole-zero 시스템으로 나타낼 수 있다. 이를 차분방정식 (difference equation)으로 표현하면 식 (3)이 된다.

$$H(z)=\frac{S(z)}{X(z)}=\frac{G(1-\sum_{j=1}^{M}b_j z^{-1})}{1-\sum_{i=1}^{N}a_i z^{-1}} \qquad (2)$$

$$s(n)=Gu(n)+\sum_{j=1}^{p}a_j s(n-j) \qquad (3)$$

LPC 분석 문제, 즉 식 (3)을 푸는 것은 신호의 측정값이 주어지면 필터의 계수 a_j를 구하는 것이다. 본 논문에서는 LPC 방정식의 차수를 10차로 하였다. 그러므로 이론적으로는 포먼트 주파수가 5개까지 나올 수 있지만, 분석하는 데이터에 따라 보통 2-4개의 포먼트 주파수가 검출된다. 본 논문에서는 포먼트 주파수는 3개까지 검출하였다.

모음식별에서 첫 번째 포먼트와 두 번째 포먼트가 가장 크게 관련되며, 첫 번째 포먼트는 혀의 구강 내에서 높낮이에 따른 인두강 부피와 관련이 있고 고모음보다 저모음에서 더 높다. 두 번째 포먼트는 혀의 앞뒤 위치에 따른 구강 길이에 영향을 받고 후설모음보다 전설 모음에서 더 높다. 세 번째 포먼트 주파수는 비음의 특성을 나타낸다 [8].

LPC 계수를 구하는 과정에서 10차의 방정식의 근들이 (poles) 서로 영향을 주는 상호간섭 (interaction)이 일어나기에 5개의 포먼트 주파수가 2-4개로 보이는 것으로 예측할 수 있다. 그러므로 이런 상호간섭을 줄여주면 이론적인 포먼트 주파수가 나타날 것이기에 Split-LPC 방법을 사용한다 [9].

먼저 10차의 LPC 방정식을 5개의 2차 방정식으로 분리하고 이들을 직렬로 연결하였다. 이제 2차 방정식을 다루는 문제가 되므로 10차 방정식에서 일어나는 poles 간의 상호간섭 문제를 줄일 수 있다. 아래의 식 (4)로 이 과정을 보였다.

$$H(z)=\frac{1}{A(z)}=\frac{1}{1+\sum_{k=1}^{10}a_k z^{-k}}=\frac{1}{\prod_{k=1}^{5}(1-p_k z^{-1})(1-p_k^* z^{-1})} \quad (4)$$

여기서 a_k 는 예측 계수이고 p_k 와 p_k^* 는 $A(z)$의 근 (pole)으로 켤레 복소근 이다.

본 논문에서는 Split-LPC 방법으로 구한 포먼트 주파수들 (F1, F2, F3)에 대한 피치 주기 (F0)로 SVM을 적용하였다 [10].

3. 서포팅 벡터 머신 (Supporting Vector Machine)

SVM은 패턴인식, 지도 학습모델 분류와 회귀 분석에 주로 사용한다. SVM은 두 카테고리 중 어느 하나에 속한 데이터 집합이 주어졌을 때, 주어진 데이터 집합을 기반으로 새로운 데이터가 어느 카테고리에 놓이는지 판단하는 이진 선형분류 모델이다. 이 분류기는 두 카테고리 사이의 여백 (Margin)이 넓어지면 잘 분류하였다고 할 수 있다. 즉 margin이 주어진 조건에서 최대가 되면 좋은 분류인 것이다. data set의 최외곽에 위치하는 support vector를 통해서 margin이 최대가 되는 결정 초평면 (Decision Hyperplane)을 찾아야 한다 [10-12].

결정 초평면을 구하기 위해서 식 (5)에서 초평면에 수직 하는 가중치 벡터 $\vec{w}$와 상수 b를 구해야 한다.

$$d(x)=\vec{w}\cdot\vec{w}+b \quad (5)$$

각 supporting vector가 위치에서의 margin을 최대화를 norm을 이용하여 표현하면,

$$\max margin=\max\frac{2}{\|\vec{w}\|} \quad (6)$$

이제 라그랑주 승수법 (multiplier)을 이용해 풀자.

$$\max_{\alpha}\min_{w,b}L(w,b,\alpha)=\frac{1}{2}\|w\|^2-\sum_{i=1}^{n}\alpha_i[y_i(wx_i+b)-1] \quad (7)$$

여기서 $\alpha_i\geq 0, i=1,2,\cdots,n$ 이다. 그래서 식 (7)은

$$-\frac{\partial L(w,b,\alpha)}{\partial w}=0 \quad \therefore w=\sum_{i=1}^{n}\alpha_i y_i x_i \quad (8)$$

$$-\frac{\partial L(w,b,\alpha)}{\partial b}=0 \quad \therefore \sum_{i=1}^{n}\alpha_i y_i=0 \quad (9)\ (8)$$

과 (9)식이 되고 이를 정리하면 (10)과 (11)식을 얻는다.

$$L(\alpha)=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j \quad (10)$$

$$\sum_{i=1}^{n}\alpha_i y_i=0 \quad (11)$$

이 두 식으로 α, w, b를 구 할 수 있다. 이 해를 통해서 결정 초평면을 구할 수 있다. 여기서 α는 라그랑주

승수이며 y_i는 훈련 집합이고 n은 훈련 샘플의 개수이다 [10-12].

본 논문에서는 SVM으로 한국인 화자와 베트남인 한국어 학습자의 한국어 겹받침의 발음을 분류하였다. 분류에 사용하는 파라미터는 피치 주기와 포먼트 주파수로 하였다. 분류가 잘 된다는 것은 두 그룹의 발음의 차이가 있다는 것이 된다.

III. 데이터와 시뮬레이션

1. 데이터 (Data)

본 논문에서 사용한 데이터는 한국어의 겹받침을 포함하는 문장으로 구성하였다. 화자의 부자연스러움과 얼버무림을 배제하기 위해서 시작과 끝부분을 채우는 문장을 두었다. 스마트 폰과 디지털 녹음기 같은 일반적인 기기로 Noisy 환경에서 녹음하였다.

본 논문에서 음성 신호 분석이 수행된 겹받침은 /굶는다/, /앉니?/, /밝아/, /밟아야지/, /흙을/, /닭을/ 의 단어들에 포함된 'ㄹㅁ', 'ㄴㅈ', 'ㄹㄱ', 'ㄹㅂ' 이다. 특별히, 본 논문에서 분석한 겹받침은 /굶는다 (ㄹㅁ)/와 /앉니? (ㄴㅈ)/ 이다. 먼저, 남과 여 각 2명의 총 4명의 베트남인 학습자들과 3명의 한국인이 발성한 두 단어를 청음하고 Mean Opinion Score (MOS) 테스트로 5점의 비율로 점수를 부여하였다. 이 테스트는 훈련된 listeners (청자)가 발음된 음성 신호를 듣고 정확도, 투명도, 등으로 주관적인 (subject) 판단으로 점수를 부여하는 것이다. 본 논문에서의 MOS 테스트는 이러한 일반적인 기준에 표준어 발음 규칙에 맞게 발음을 하는지를 보았다. 이 결과를 표 1에 보였다. 표 1에는 MOS 테스트를 포함해서 이 두 단어의 피치 주기와 포먼트 주파수의 측정값을 보였다. 피치 주기와 포먼트 주파수는 겹받침의 모음, 즉 /굶/인 경우는 /우/를, /앉/일 때는 /아/를 측정한 것이다.

2. 모의실험 (Simulation)

먼저 noisy 환경에서 녹음한 데이터를 16kHz의 주파수로 샘플링을 하여 본 시뮬레이션을 진행하였다. 본 논문에서는 앞서 언급한 겹받침들이 포함된 단어들로 구성된 문장들을 사용하였다. 예를 들면, /먹기 싫어서 굶는다/와 /값도 모르고 앉니?/ 같은 문장이다. 시뮬레이션은 먼저 데이터를 512 샘플 단위로 프레이밍 한다. 이 512 샘플로 구성된 단위로 즉, 프레임별로 음성 신호의 파라미터들을 추출한다. 이 과정을 위해서 먼저 유성음 프레임과 무성음 프레임을 구분할 필요가 있다. 이를 위해서 영 교차율 (zero-crossing rate, ZCR)와 short-time energy (STE)를 활용하였다.

영 교차율 (ZCR)은 음성 신호가 x축 (혹은 0)을 통과하는 횟수를 말한다. k번째의 샘플과 $k+1$번째를 곱해서 그 값이 음수이면 영 교차가 일어났다는 것으로 두 신호의 부호가 서로 다르다는 것을 의미한다. 이런 경우가 한 프레임에서 몇 번 일어났는지 그 횟수를 계산한다. 음성 신호에서 유성음은 준주기적 신호이므로 무성음 구간에 비해서 상대적으로 영 교차율이 적다. 이에 비해 무성음은 랜덤 신호로 표현되므로 유성음 구간보다 영 교차율이 높다. 이를 sign 함수로 표현하면 아래와 같다 [13, 14].

$$ZCR=\sum_{m=-\infty}^{\infty}|sgn[x(m)]-sgn[x(m-1)]|h(n-m) \quad (12)$$

여기서,

$$sgn[x(n)]=\begin{cases}1, & x(n)\geq 0\\ -1, & x(n)<0\end{cases} \quad (13)$$

$$h(n)=\begin{cases}\frac{1}{2N}, & 0\leq n\leq N-1\\ 0, & otherwise\end{cases} \quad (14)$$

STE는 시간의 변화에 따른 에너지의 값을 구하는 것이다. 에너지의 식에 윈도우 함수를 곱해줌으로 계산할 수 있다. 윈도우 함수를 곱해준다는 것은 아주 짧은 시간 동안에서의 에너지를 구한다는 것이다.

$$E(n)=\sum_{m=-\infty}^{\infty}[x(m)h(n-m)]^2 \quad (15)$$

ZCR과는 반대로 STE가 클수록 유성음 구간일 수 있고 랜덤 신호인 무성음은 ZCR이 높으므로 그 에너지는 상대적으로 작을 것이다. 그러므로 ZCR이 작고 에너지가 크면 유성음으로 구분할 수 있고 ZCR이 높고 에너지가 작으면 무성음으로 구분할 수 있다. 본 논문에서는 유성음/ 무성음 구분을 최대 ZCR의 60% 그리고 최대 STE의 20%를 기준으로 하였다. 즉, 입력 음성 신호에서 최대 ZCR의 60% 이하이고 최대 STE의 20% 이상인 프레임을 유성음으로 하였다 [13, 14].

IV. 모의실험 결과 (Simulation Results)

앞에서 언급했듯이 본 논문에서 자세히 살펴본 겹받침은 /굶는다 (ㄹㅁ)/와 /앉니? (ㄴㅈ)/ 이다. 첫 번째 단어 /굶는다/는 [굼는다]로 발음이 되어야 한다. 즉, 'ㄹㅁ'의 겹받침은 자음 앞에서 뒤 자음 'ㅁ'을 발음한다. 두 번째 단어 /앉니?/는 의문문에서 선택한 것으로 [안니?]로 앞 자음으로 발음해야 한다.

표 1의 MOS 테스트의 결과로 보인 것과 같이 베트남인 한국어 학습자들은 [1]과 [2]에서 예상한 것과 같이 뒤 자음을 발음하는 경우가 앞 자음을 발음하는 것보다 어려움을 겪는 것으로 보인다.

MOS 스코어가 뒤 자음 발음의 경우 한국인과 베트남 학습자 간의 차이가 확실히 보이지만 앞 자음을 발음하는 경우에는 의문문에서 선택된 겹받침임에도 차이가 나지 않음을 알 수 있다.

그림 1과 그림 2에서는 본 논문에서 분석한 두 가지 겹받침을 포함한 두 단어 /굶는다/ 와 /앉니?/의 시간 영역에서의 파형 (waveform)과 스펙트럼 (에너지)를 각 화자에 대해 보인다. 파형은 화자의 발화를 시간 영역에서 그 변화를 보였다. 스펙트럼은 Short-Time Fourier Transform (STFT)을 의미한다. 음성신호는 시간에 따라 그 성질이 변하는 랜덤 신호이기에 Fourier 변환을 매우 작은 시간 구간에서 수행하여 신호의 성질이 시간에 따라서 변하지 않는다는 가정을 한 것이다. 본 논문에서는 MatLab의 내장함수인 "spectrogram function"을 사용하여서 Fourier 변환의 값을 구했다. spectrogram function은 주파수 도메인에서 음성신호의 시변 특성들을 나타내는 Short-time Fourier transform에서 $x_k(e^{j\omega})$의 크기에 로그를 취해서 시간 축과 주파수 축으로 그린 이미지이다. 이는 음성신호가 시간으로 변해 갈 때 이에 따라서 변하는 주파수 성분이 어떻게 분포되어 있는지를 보여준다 [15, 16].

그림 1에서는 /굶는다 (ㄹㅁ)/의 파형 (waveform)과 스펙트럼을 보인다. 시간 영역에서의 파형을 보면 전체적으로 한국인이 베트남인 학습자들에 비해서 발화 속도가 빠른 것을 알 수 있다. 스펙트럼을 보면 베트남 여성 화자 (1)과 (2) 가 다른 화자들과 다른 것을 알 수 있고 이는 MOS 테스트에서도 다른 것으로 나타난다.

표 1. 한국어 겹받침 /굶는다/ 와 /앉니?/의 음성신호에 대한 파라미터
Table 1. Simulation parameters for Korean Final Double Consonants - /굶는다/ and /앉니?/

word	Participants			Vietnamese				Korean		
	Parameters			Male (1)	Male (2)	Female (1)	Female (2)	Male (1)	Female (1)	Female (2)
굶는다 ㅡ 굼는다 ㅡ	청음 (MOS)			4	4	1	2	5	4	5
	Pitch period (F0) [ms]			5.0625	5.74	4.95	4.12	9.875	5.25	5.96
	Formant Freq.	LPC [Hz]	F1	318.1	342.9	256.7	348	1502	380	279
			F2	2553.8	3393.8	2914.2	3308.2	3499	2782.9	3014.9
			F3	5152.5	4343.8	4464	5528.5	4846.7	4899	4922.7
		Split-LPC [Hz]	F1	256.2	338.3	194.8	287.7	164	285.1	322.2
			F2	2493.8	2421.1	2215.2	1141.7	2626	2556.4	2474
			F3	5415.6	5483.6	4422.8	4350	6901.4	5288.5	4815.5
앉니 ㅡ 안니 ㅡ	청음 (MOS)			5	4	4	5	5	5	5
	Pitch period (F0)			5.56	7.1	5.9	4.06	7.25	5.06	4.875
	Formant Freq.	LPC	F1	296.87	418.2	595.8	348.9	358.7	398.8	361.8
			F2	2679.2	3155.3	3204.5	3577.4	2807.1	3416.9	2567.3
			F3	3637.3	4611.3	4618.9	5563.2	4971.5	4526.3	5276
		Split-LPC	F1	292.7	399.8	146.7	244	234.4	174.3	311.3
			F2	2312.5	2288.6	2914.4	2372.8	2613.5	2549.3	2146.6
			F3	4392.7	5404.4	5163	5137.6	5546.9	5386	5039.7

그림 2는 앞 자음이 발음되는 /앉니?/의 음성 신호의 파형과 스펙트럼을 보인다. 그림 1의 신호 파형과 스펙트럼과 다르게 7명 화자가 모두 매우 비슷한 것을 알 수 있다.

그림 3에서 그림 5까지에 뒤 자음이 발음되는 /굶는다/에 대해서 SVM으로 베트남인 학습자와 한국인을 분류하였다. SVM 분류기는 피치 주기에 대해 포먼트 주파수 F1, F2, F3를 사용하여 결정 초평면을 구하였고 이를 각 포먼트 주파수에 따라 보였다.

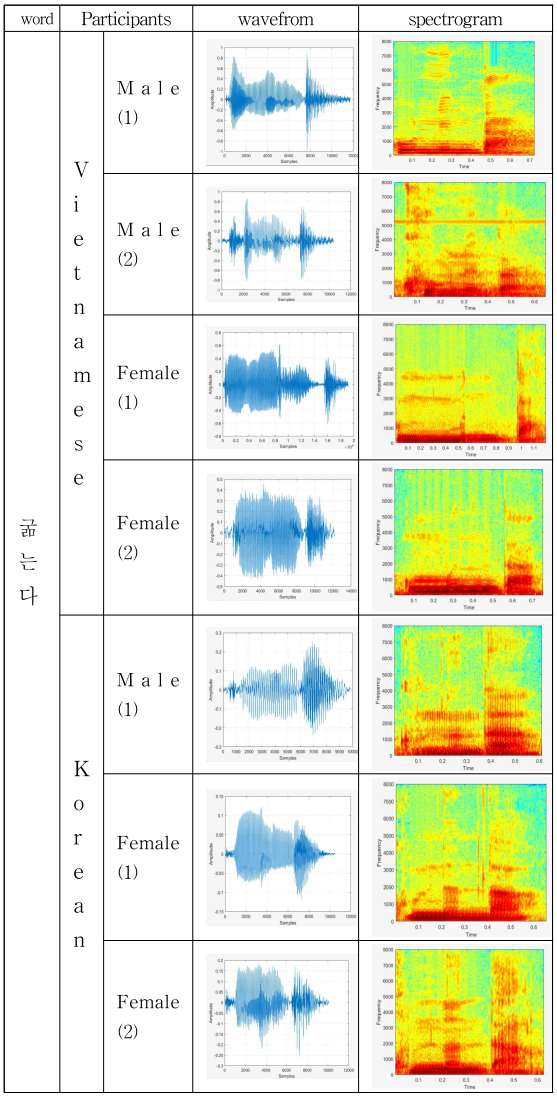

그림 1. /굶는다 (ㄹㅁ)/ 의 waveform과 spectrum
Figure 1. Waveform and its spectrum of /굶는다 (ㄹㅁ)/

그림 6에서 그림 8에서는 앞 자음을 발음하는 한국어 겹받침 /앉니?/에 대한 SVM 분류 결과를 보인다. 그림 6과 그림 8에서는 결정 초평면이 베트남인 학습자와 한국인을 분류하고 있으나, 그림 7에서는 결정 초평면을 구하지 못했다.

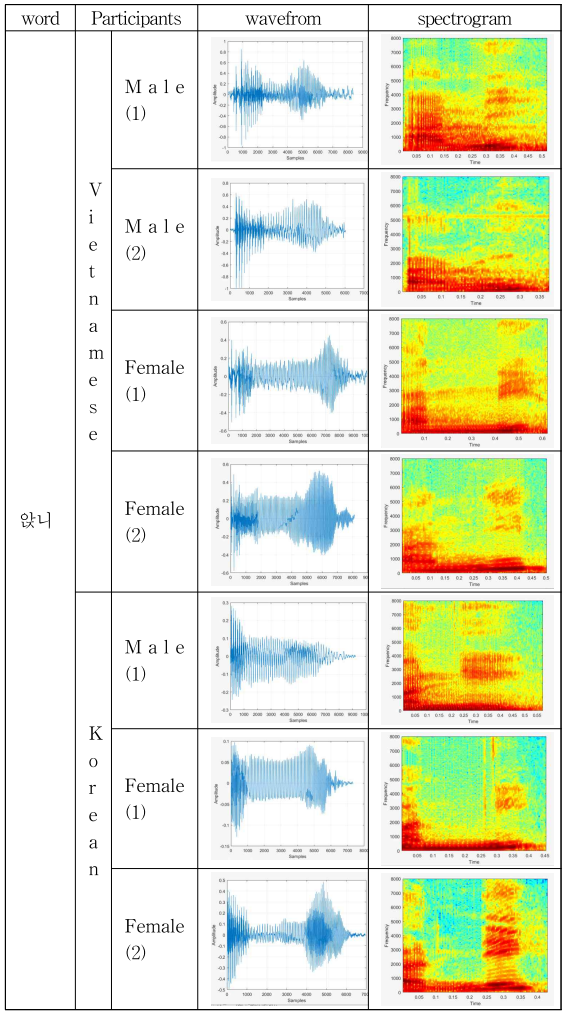

그림 2. /앉니? (ㄴㅈ)/ 의 waveform과 spectrum
Figure 2. Waveform and its spectrum of /앉니? (ㄴㅈ)/

V. 결 론

본 논문은 언어학적인 연구를 통하여 조사하고 분석한 겹받침 발음에 관한 여러 오류와 제시한 교육 방법에 대하여 공학적 특히 음성 신호처리의 분석 방법을 활용하여서 이런 연구 결과를 확인하였다.

겹받침 단어 /굶는다/에 대한 표 1의 MOS 데스트의 결과와 그림 1에서의 파형과 스펙트럼을 보면 베트남인 학습자와 한국인이 확실한 차이를 보인다.

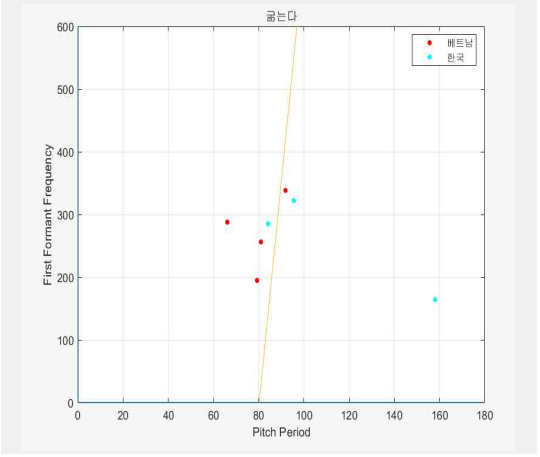

그림 3. /굶는다/ 한국인과 베트남 학습자의 분류 (피치주기 - 첫 번째 포먼트 주파수)
Figure 3. /굶는다/ Classification of Korean and Vietnamese Learners (Pitch period - 1st Formant Frequency)

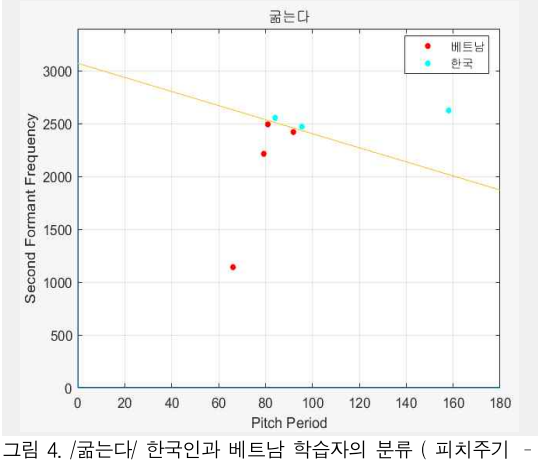

그림 4. /굶는다/ 한국인과 베트남 학습자의 분류 (피치주기 - 두 번째 포먼트 주파수)
Figure 4. /굶는다/ Classification of Korean and Vietnamese Learners (Pitch period - 2nd Formant Frequency)

그러나 겹받침 단어 /앉니?/는 모든 참여자의 파라미터들이 거의 같다고 할 수 있다. 이의 두 사실에서 알 수 있는 것은 언어학의 연구에서 예측한 것과 같이 뒤 자음을 발음하는 겹받침 단어들이 앞 자음을 발음하는 단어들보다 베트남인 학습자가 발음에 어려움이 있다는 것을 확인하였다.

표 1에서 LPC와 Split-LPC의 두 방법으로 포먼트 주파수를 구하였다. 본 논문에서 적용한 포먼트 주파수는 Split-LPC로 측정한 것이다. 이 결과가 LPC로 구한 주파수보다 그 동적 범위가 안정적이었다. 포먼트 주파수는 주파수 영역에서의 특징들이다. 이에 반해 피치 주기는 시간 영역에서 특징이다. 그러므로 이 두 가지 특징들은 서로에게 종속성이 없다고 할 수 있으니 측정값들을 믿을 수 있다. 이런 이유로 본 논문에서 사용한 SVM 분류기에는 피치 주기 (F0)와 포먼트 주파수 (Fi)를 활용하였다.

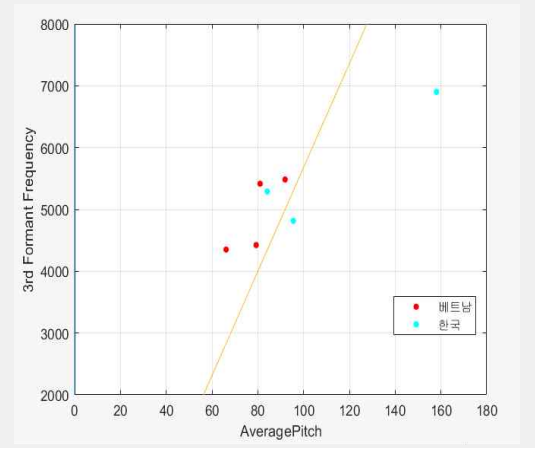

그림 5. /굶는다/ 한국인과 베트남 학습자의 분류 (피치주기 - 세 번째 포먼트 주파수)
Figure 5. /굶는다/ Classification of Korean and Vietnamese Learners (Pitch period - 3rd Formant Frequency)

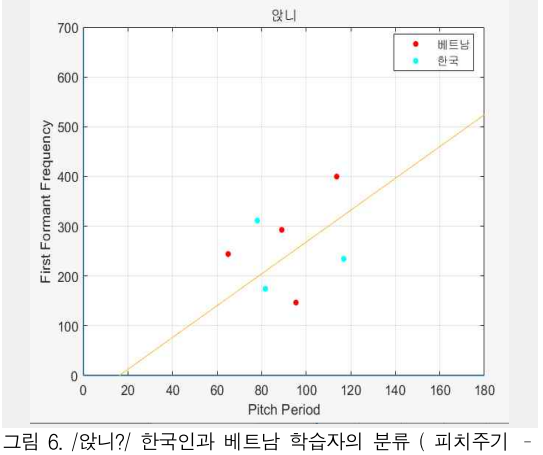

그림 6. /앉니?/ 한국인과 베트남 학습자의 분류 (피치주기 - 첫 번째 포먼트 주파수)
Figure 6. /앉니?/ Classification of Korean and Vietnamese Learners (Pitch period - 1st Formant Frequency)

그림 3-5 에서는 겹받침 단어 /굶는다/에 대한 SVM 분류기가 초결정 평면을 잘 보인다. 즉, 베트남인 학습자의 발음과 한국인의 발음에 차이가 있음을 알려준다. 그림 3에서 한국인 1명과 베트남인 학습자 1명이 잘 분류가 되지 않고 있는데 이들의 MOS 테스트가 나빴음을 알 수 있다.

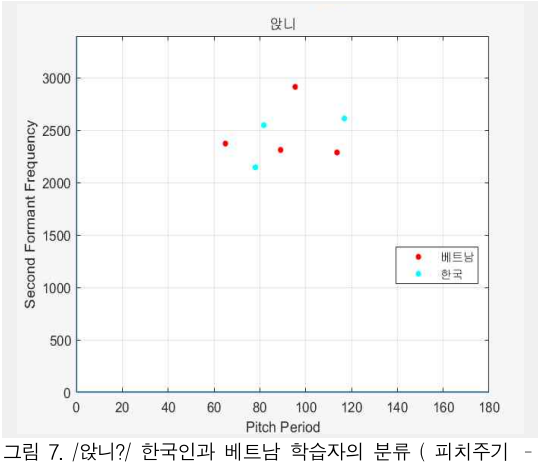

그림 7. /앉니?/ 한국인과 베트남 학습자의 분류 (피치주기 - 두 번째 포먼트 주파수)
Figure 7. /앉니?/ Classification of Korean and Vietnamese Learners (Pitch period - 2nd Formant Frequency)

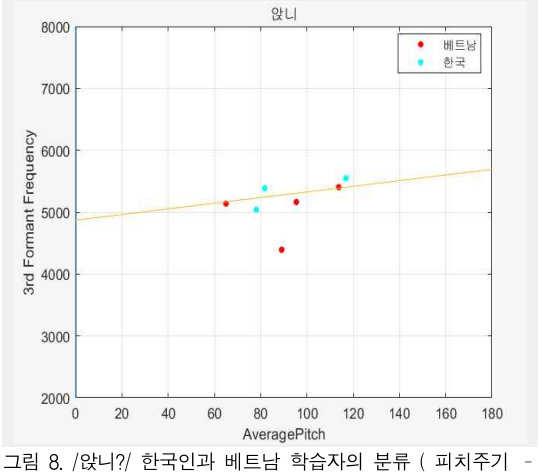

그림 8. /앉니?/ 한국인과 베트남 학습자의 분류 (피치주기 - 세 번째 포먼트 주파수)
Figure 8. /앉니?/ Classification of Korean and Vietnamese Learners (Pitch period - 3rd Formant Frequency)

특히 한국인 1인은 그림 5에서도 분류가 잘 안 되고 있다. 이 화자는 F1 과 F3에서 분류가 되지 않는다. 상대적으로 이 화자는 발음이 좋지 않았다는 것이다. 그림 6-8에서는 /앉니?/에 대한 SVM 분류기의 초결정 평면을 구하는 어려움을 보인다. 특히 그림 7에서는 초결정 평면을 구하지 못했다. 그런데 표 1에서 이 단어의 MOS 테스트 결과를 보면 참여자 모두가 4~5점을 받았다. 즉, 모든 참여 화자의 발음이 좋았다는 것이다. 따라서 본 논문에서 활용한 SVM 분류기는 외국인 학습자의 발음과 한국인의 발음을 비교에 활용할 수 있고, 이를 통해서 외국인 학습자의 발음 교육에 유용한 파라미터라고 할 수 있다.

언어학적 연구를 통해서 제시된 겹받침 발음 교육은 쓰기와 듣기 그리고 표준 발음법을 이해하고 인지 교육을 통한 반복 학습이다. 이에 본 논문이 제시하는 방법은 쓰기, 듣기, 그리고 반복 학습에 본 논문에서 보았던 신호의 파형과 에너지 같은 음성 신호처리의 파라미터를 visualize 하여 role-playing이 가능한 가상 현실 프로그램 (Metaverse 같은)을 만들어 훈련한다면 좀 더 효율적인 방법이 될 것이다.

본 논문에서 활용한 외국인 한국어 학습자는 베트남인으로 하였기에 데이터가 부족하다. 앞으로 여러 국가의 학습자와 더 많은 겹받침 발음 데이터로 하면 더욱 의미 있는 결과를 도출할 수 있다. SVM 분류기에 활용한 파라미터가 피치 주기와 포먼트 주파수이었다. 향후에 SVM 분류기의 활용 파라미터를 다양하게 조합하면 - 예를 들면, ZCR 과 STE - SVM 분류기의 기능과 성능이 향상될 것으로 기대된다.

References

[1] Choe Kyeong-bok and Song ji-young, “A Study on the Pronunciation Errors of Korean Double-final Consonants by Vietnamese Learners”, The Journal of Korean Language and Literature, vol. 110, pp. 175-203, Sept. 2019. DOI: https://dx.doi.org/10.21793/koreall.2019.110.175

[2] Lee Mi Young, A Comparative Study of the Actual Pronunciations od Korean, Chinese, Vietnamese and Filipino Groups on Korean Syllable Final Double Consonants and Education methods Master Thesis, Graduate School of Education Kunsan National University Kunsan, Korea, 2016.

[3] KangEun-jin, Tran Thi Huong, and Cho Chae-hyung, “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", The Journal of Korean Language and Literature, vol. 36, pp. 191-219, Feb. 2020. DOI: http://doi.org/10.24227/jkll.2020.02.36.191

[4] Liang Hai-sheng and Bao Juan, "A Study on Korean Final Double Consonants' Teaching Model in China - Focusing on Korean major students at lower grades in four-year undergra duate college", The Journal of Korean Language Education, vol. 12, pp. 71-103, Feb. 2017.

[5] 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, Sept. 2020. DOI: https://doi.org/10.17703/IJACT.2020.8.3.316

[6] Ministry of Education, Support Plan for University Innovation in Response to Demographic Changes and the Fourth Industrial Revolution, "Higher Education Policy Office, Ministry of Education, 2019.

[7] Shinae So, Kang-Hee Lee, and 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, 21 (12), 2016.

[8] A. M. Kondoz, Digital Speech(Coding for Low Bit Rate Communications Systems), 1st ed, WILEY, 1995.

[9] K.B You and K.H Lee "A study on splitting lpc synthesis filter", Information Technology Convergence, LNEE, volume 253, pp 1003-1009, Springer, July 2013.

[10]Jeong-seok Yeom, Kwang-Bock You, and Kyungnam Jang, "A Study on the Emotion Classification of the Speech Signal Using Support Vector Machine", The Journal of Korean Institute of Communications and Information Sciences '21-10 Vol.46 No.10, Oct. 2021. DOI: https://doi.org/10.7840/kics.2021.46.10.1741

[11]Durgesh K. Srivastava and Lekha Bhambhu, "Data classfication using support vector machine", Journal of theoretical and applied information technology", 2005.

[12]E. Osuna, R. Freund, and F. Girosi, "Support Vector Machines Training and Applications", AI No. 1602, Artificial Intelligence Laboratory, MIT, 1997.

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

[14]Bachu R.G, Kopparthi S, Adapa B, and Barkana B.D, "Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal", IEEE trans. On ASSP, vol. ASSP-24, pp. 201-212, 2014.

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

[16]Park, Hyungwoo, "Improvement of Sound Quality of Voice Transmission by Finger." International Journal of Advanced Culture Technolgy, vol. 7, no. 2, IPACT, pp. 218-226, June 2019. DOI: 10.17703/IJACT.2019.7.2.218

※ 본 연구는 2020년도 숭실대학교 교내연구비 지원에 의한 연구임.