

SVM 기법에 기초한 청각장애인의 영어모음 발음을 위한 음성 인식 및 입술형태 특징 추출

Speech Recognition and Lip Shape Feature Extraction for English Vowel Pronunciation of the Hearing - Impaired Based on SVM Technique

이근민*, 한경임, 박혜정

K. M. Lee, K. I. Han, H. J. Park

요 약

이 연구의 목적은 SVM 기술에 근거한 시각 보조기에 주로 의존하는 청각 장애인을 위한 영어 모음 발음에 대한 시각 교습 방법을 제안하는데 있다. 귀로 듣기 어려운 소리로부터 SVM 기술을 사용하여 소리 특징을 추출함으로써, 각 모음의 입술 모양이 추출되었다. 모음에 대한 입술 모양의 세련미는 언어 학습자가 발음기의 움직임을 눈으로 쉽게 볼 수 있다는 점에서 유리하며, 청각 장애인을 위한 영어 모음을 학습하고 가르치는 데 유용할 것이다.

ABSTRACT

The purpose of this study is to suggest the visual teaching method for the English vowel pronunciation, especially for the hearing-impaired who mostly rely on the visual aids, based on the SVM technique. By extracting phonetic features using the SVM technique from the sounds that are hard to hear by ear, the lip shapes for each vowel were refined. The lip shape refinement for vowels is advantageous in that language learners can easily see the movement of articulators by eye, and it is helpful for learning and teaching English vowels for the hearing-impaired.

Keyword : English vowel, formant frequency, lip shape, hearing-impaired, gender, SVM, speech recognition

1. 서 론

디지털 혁신과 소셜 미디어의 등장으로 데이터 수집 및 저장 용량이 증가하고 있다. 빅 데이터는 이러한 데이터를 분석하고 교육, 에너지 산업, 의료 산업, 정치, 스포츠 등 다양한 분야에서 미래의 변화를 예측함으로써 사회 및 인류에게 귀중한 정보

를 제공 할 수 있는 가능성을 제시한다. 미래에는 대용량 데이터의 적절한 사용이 대부분의 분야에서 성공의 열쇠가 될 것이다. 이 논문에서는 영어 모음 인식에 얼마나 큰 데이터가 사용되고 있는지를 보여줄 것이다.

오늘날 어쿠스틱 언어학에서 영어 모음 발음과 관련된 많은 연구 프로젝트는 모국어 및 비원어민 및 정상 청각 및 청각 장애인 간의 차이점을 찾아 내는 데 초점을 맞추고 있다[1-3]. 일반적으로 모음은 세 가지 요소로 설명 할 수 있다. (1) 혀의 몸높이 (2) 혀의 앞뒤 위치; (3) 립 라운딩의 정도 [4]. 모음에 대한 이러한 전통적인 설명은 포먼트 주파수와 관련이 있다. 포먼트 주파수들 중에서, 제 1 포먼트 (F1) 및 제 2 포먼트 주파수 (F2)는 모음을 구별하는데 결정적인 역할을 한다. 즉, F1과 F2는 모음/ 혀 높이와 모음/ 혀의 뒤쪽을 각각 나타낸다.

접 수 일 : 2017.08.12

심사완료일 : 2017.08.30

게재확정일 : 2017.08.31

* 이근민 : 대구대학교 재활공학과 교수

kunminrhee@hanmail.net (주저자)

한경임 : 계명대학교 교양교육대학 교수

kyungim@gw.kmu.ac.kr (공동저자)

박혜정 : 대구대학교 기초교육대학 교수

hye3255@daum.net (교신저자)

※ 이 논문 또는 저서는 2013년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2013S1A5A2A03045751)

그러나 모국어가 영어가 아닌 사람이 혀 위치에 의존하여 영어 모음을 배울 때 학습자가 각 모음에 대해 혀를 어디에 두어야 하는지를 알아내는 것은 매우 어렵다. 입은 육안으로는 보이지 않는다. 특히, 청각 장애인의 언어 시스템은 보통 청각 능력이 정상 청력보다 현저히 낮기 때문에 수화라는 시각적 언어로 구성된다. 언어로 청각 장애인에게 지시 할 때, 구두 능력뿐만 아니라 그들의 언어 능력이 정상 청력과 다르다는 것을 인정하는 것이 필요하다 [5-6]. 이러한 차이 때문에 청각 장애인이 시각적으로 정보를 받을 필요성을 고려하여 언어 교육을 위한 새로운 방법을 제안해야 한다.

이 특별한 상황을 염두에 두고, 현재의 연구는 사람들이 쉽게 보고 객관화 할 수 있는 5 개의 느슨한 영어모음에 대한 입술 모양의 차이점을 보여준다. 입술 모양은 아래쪽 항아리와 혀의 움직임과 밀접하게 관련되어 입모양을 보고 혀의 위치를 판단할 수 있다. 이를 수행 할 때 첫 번째 포먼트 주파수(F1)와 두 번째 주파수(F2)를 모음 분류 기준으로 사용한다. 첫 두 포먼트는 혀 위치뿐 아니라 입술 라운딩과도 관련되기 때문이다. F1은 혀 높이에 반비례한다. F2의 경우 입술이 튀어 나와 구강이 길어지고 F2 값이 낮아진다. 우리의 연구는 김(Kim), 강(Kang), 코(Ko)[7]에 의해 제안 된, 포먼트 주파수와 입술형태 사이의 관계에 대한 가정을 취한다. 즉, F1은 입술의 높이와 관련이 있고 F2는 입술의 넓이와 상관관계가 있다.

이 연구의 목적은 두 가지이다. 첫째, 우리는 모음을 만들 때 입술 모양의 특징을 정확하게 식별하여 청각 장애인을 위한 영어 모음 발음을 시각적으로 가르치는 방법을 제안 한다. 둘째, SVM (Support Vector Machine) 기술을 사용하여 청각 장애인 데이터에서 추출한 음성 특징을 적용하여 입술 모양과 번역 시스템을 재생성 한다.

2. 연구 방법

2.1 연구 대상

이 연구를 위해 영어 원어민, 한국인 정상 청력 및 한국어 청력이 약화 된 세 그룹의 연사 그룹이 포함되었다. 각 그룹은 남녀 스피커로 구성된다. 영어 원어민 그룹의 경우 힐 브랜드 등의 포먼트 값을 사용했다[8]. 비교를 위해서, 한국의 정상 청각장애인과 관련하여, 20대 남녀가 동원되어 대구에 살던 10명의 피험자가 실험에 참여했다. 그들은 정상

적인 말하기와 듣기를 보고했으며 언어 장애의 병력이 없었다. 한국인 청력 장애 참가자 그룹은 대구에 거주하는 10 대 후반의 남성 5 명, 여성 5 명으로 구성되었다. 이 연구를 위해 선택된 참가자는 영어 단어를 인식하고 읽는 능력을 가지고 있다.

2.2 음성 자료 및 절차

오디오 녹음은 5 명의 느슨한 모음 /i, ε, æ, ɔ, ʌ/ 즉, "hit, head, hat, hood, hut"을 포함하는 단어를 읽는 한국어 사용자들로 이루어져 있다. 녹음하기 전에 참가자들은 단어를 올바르게 발음 할 때까지 주어진 영어 단어를 연습하도록 요청 받았다. 각 단어는 고립 된 맥락에서 모든 주제에 의해 세 번 발음되었다. 녹음 준비가 되었을 때, 참가자들은 책상 앞에 앉아 가능한 한 자연스럽게 단어를 읽으라는 메시지를 받았습니니다. 레코딩은 SONY 헤드 마운트 마이크와 디지털 레코더를 사용하여 조용한 방에서 이루어졌다. 각 주파수에 대해 얻어진 값을 확인하기 위해 'PRAAT' 음성 분석 소프트웨어 프로그램이 사용되었다[9]. 5개의 모음에 대한 포먼트 주파수가 스펙트로그램에서 측정되었다. 모음 섹션을 추출한 후, 스펙트로그램의 모음의 정상 상태 부분을 선택하여 측정했다.

2.3 SVM을 이용한 스피커 및 모음 분류를 위한 모델링 및 추출 기능 세트

그림 1은 SVM을 이용하여 화자 집단과 모음 분류를 위한 모델링 및 추출 특징 집합을 보여준다. 그림 1의 ㉑은 분류를 위한 지원 벡터가 사용되는 화자의 특징을 추출하는 과정이다. 그림 1의 ㉒는 ㉑의 과정에서 화자 그룹의 특징을 반영하여 각 화자 그룹으로부터 입술 모양의 특징을 추출하는 과정이다.

지원 벡터 분류 후, 다중 지원 벡터 분류는 그림 1에서 ㉑ 및 ㉒에 대한 특징 추출에 사용된다. 우리는 3개의 2클래스 '1 대 모두' 분류자를 사용하여 모든 스피커 및 모음 분류에 대해 3가지 클래스로 기능을 도표화했다. SVM을 다음과 같이 두 번 사용했다.

3. Schmatized 입술 모양

모두가 다른 립 형태와 다른 입 크기를 가지므로, 모음을 위한 입술 열림의 정도를 디지털화하는 것은 어렵다. 또한, 사람들은 입 열기의 정도가 다른 모음을 발음한다. 입술 모양을 개선하기 위해 Kim,

Kang & Ko[6]가 제안한 간단한 입술 모델을 사용하였다.

3.1 모음의 평균 빈도수

그림 2는 5개의 느슨한 모음의 평균 F1을 화자 그룹과 성별로 나타낸 것이다. /ɪ/의 경우 남성을 위한 영어 원어민 강사의 F1은 한국의 정상 청력 및 청력 장애자보다 유의하게 높다. /ε 및 /æ/의 경우 청각 장애자는 남녀 그룹 모두 F1을 유의하게 낮다. /ʊ/의 경우, 발화 군과 성별 간에 유의한 차이가 없었다. /ʌ/의 경우, 청력 장애가 있는 여성 그룹은 F1 값이 낮았지만 남성 그룹에게는 유의한 차이가 없었다.

그림 3. 화자 그룹과 성별로 5개의 느슨한 모음의 평균 F2를 나타냅니다. /ɪ/의 경우, 청각 장애인은 두 성별 모두 다른 두 그룹보다 약간 높은 값을 보였다. /ε/의 경우 청각 장애인이 여성에 대해 유의하게 낮은 수치를 보였다. /æ/의 경우, 청각 장애인은 두 성별에 대한 낮은 F2를 보여 주었다. /ʊ/의 경우, 발화 집단 및 성별 측면에서 유의한 차이가 없었습니다. /ʌ/의 경우, 남성 그룹은 발화 그룹 중 F2에 큰 차이를 보이지 않았지만, 여성 그룹에서는 청력이 약한 여성에게 F2가 더 낮았다.

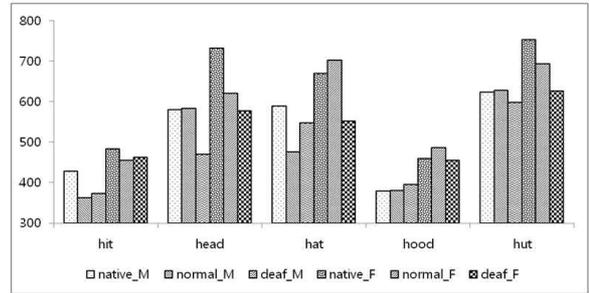


그림 2. 원어민, 한국인 정상 청력 및 한국인 청각 장애자에 대한 성별에 의한 F1 평균 빈도
Fig. 2. Mean F1 frequency for Native speaker, Korean Normal-Hearing, and Korean Hearing-Impaired by Gender

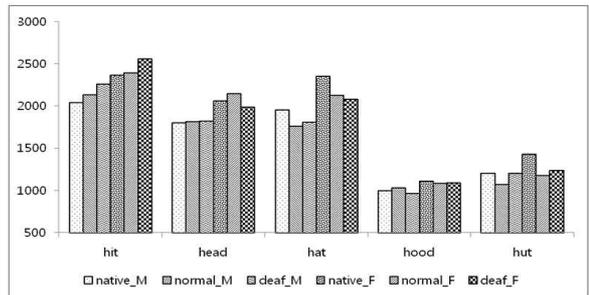


그림 3. 네이티브 스피커, 한국인 정상 청력 및 한국인 청각 장애인을 위한 F2 평균 주파수
Fig. 3. Mean F2 frequency for Native speaker, Korean Normal-Hearing, and Korean Hearing-Impaired by Gender

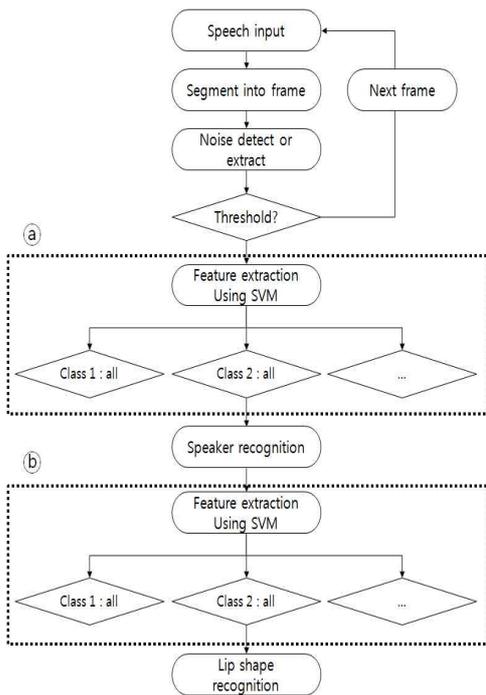


그림 1. 화자 그룹 및 모음 분류 절차 블록 다이어그램

Fig. 1. Block diagram of speaker group and vowel classification procedure

3.2 성형 주파수를 이용한 입술 형태

각 모음에 대한 입술의 모양의 차이를 조사하기 위해, 포먼트가 흉부의 공진 주파수를 나타내므로, 포먼트 주파수(formant frequency)와 입술모양(lip shape)사이의 관계를 보여주는 간단한 립 모델이 연구에 사용되었다. 모음을 설명 할 때 매우 유용한 기능이다. 표 I은 입술 반올림과 혀의 위치에 대한 다섯 가지 영어 모음의 분류를 보여준다.

3표 1. 입술 반올림과 혀의 위치에 대한 영어 다섯 모음의 분류

Table 1. Classification of the English five vowels in terms of lip rounding and tongue position

	Front		Central		Back	
	Flat	Round	Flat	Round	Flat	Round
High	/i/					/u/
Mid	/e/					
Low	/æ/		/ʌ/			

그림 4는 Kim, Kang & Ko [6]가 만든 간단한 입술 모델이다. 포먼트(formant) 빈도와 입술 모양 사이의 선형 관계는 입술의 높이와 너비로 매핑된다.

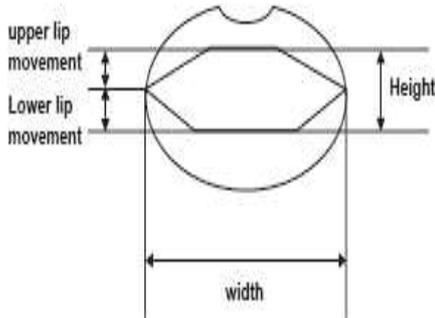


그림 4. 간단한 립 모델 [6]
Fig. 4. Simple Lip Model[6]

입술의 높이와 너비에 대한 공식 [6]

$$h = \alpha f_1 \quad (1)$$

$$w = \beta f_2 \quad (2)$$

α , β 는 상수, h 와 w 는 각각 입술의 높이를 의미

$$h_u = 0.3(h_b + \alpha(f_{bh} - f_1)) \quad (3)$$

$$h_l = 0.7(h_b + \alpha(f_{bh} - f_1)) \quad (4)$$

$$w = w_b + \beta(f_{bw} - f_2) \quad (5)$$

여기서 α , β 는 상수, h_u 는 상 립 동작, h_l 은 입술의 움직임, w 는 입술의 폭, h_b , w_b 는 입술의 기본 높이와 폭 크기, f_{bh} , f_{bw} 는 기본 주파수이며 f_1 , f_2 는 제 1 및 제 2 포먼트 주파수이다.

표 2. 모음을 위한 혀의 전진과 높이 [10]

Table 2. Tongue Advancement and Height for vowels[10]

		←Tongue Advancement					
		6	5	4	3	2	1
Tongue Height ↓	i				o		u
			i	ə		o	
	æ		ε	Λ	ɔ		
			α				

그림 2는 입의 개도에 대한 것이며, 각 모음에 대한 입의 길이와 폭을 나타낸다. x 축에서 숫자가 증가할수록 립이 넓어짐을 나타낸다. Y 축의 경우 숫자가 높을수록 더 낮은 립이 윗입술에서 멀어진다. 예를 들어, / i /의 경우 입을 6 번 수평으로, 1 번을 수직으로 떨어뜨린다. 숫자는 각 모음의 입술 높이와 입술 폭의 비율로 사용된다.

그림 4의 간단한 입술 모델과 표에서 포먼트 (formant) 빈도를 가진 입술의 높이와 너비에 대한 공식을 사용한다.

그림 2를 참조하면, 영어 원어민, 한국인 정상 청력 및 한국 청각 장애인에 의해 생성된 각각의 모음에 대한 입술 모양이 계산된다. 그림 5와 그림 6은 영어 원어민, 한국인 정상 청력 및 남학생과 여성의 청각 장애인이 각각 제작한 5 개의 영어 모음에 대한 입술 모양을 나타낸다.

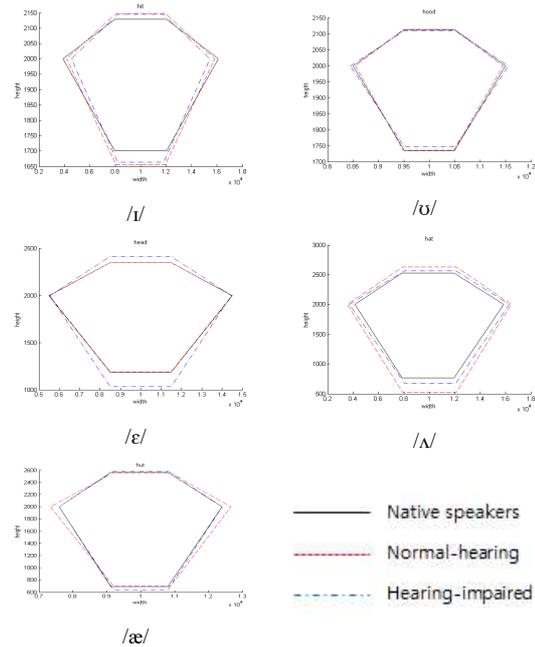


그림 5. 영어 원어민, 한국의 정상 청각인과 청각 장애인(남성)에 의한 5 개의 영어 모음에 대한 입술 모양(참고 : 각 모음에 대한 x 축과 y 축의 비율은 다르다.)

Fig. 5. Lip shape for the five English vowels produced by English native speakers, the Korean normal-hearing and the Korean hearing-impaired (male)(Note: scales of x-axis and y-axis for each vowel are differ.)

/ i /의 경우 청각 장애인과 정상 청취자는 아래쪽 립을 아래쪽으로 움직여 이 소리를 낸다. / ε /의 경우 청각 장애자는 윗입술을 내리고 윗입술을 올리면서 발음했다. 여성용 스피커는 입술이 넓어지면서 발음했다. / æ /의 경우 청각 장애인의 남성 연설자는 입술을 퍼지면서 소리를 냈지만 청각 장애가 있는 여성 연설자의 경우에는 입술을 넓게 펼치고 아래 립을 내리고 감피를 들어 올리는 것으로 발음한다. / ʊ /의 경우 여러 연사 그룹의 남성은 큰 차이를 보이지 않지만 여성 청력 장애자는 입술 퍼기가 더 많이 사용된다. / Λ /의 경우, 정상 청력 및 청력 장애 그룹은 윗입술을 낮추고 윗입술을 올리고 입술을 넓혀서 발음했다.

높은 모음 / i / 및 / ʊ /와 비교하면, 중간 모음 / ε / 및 낮은 모음 / æ / 및 / Λ /은 특히 청력에

있어 연결자 그룹 사이의 입술의 높이와 너비에 분명한 차이가 있다 - 손상. 이 연구에서 영어 모음 발음에 대한 청력 장애자의 일반적 경향은 입술의 높이와 입술의 폭을 더 많이 사용함으로써 청력이 약한 발음으로 발음된다는 것이다.

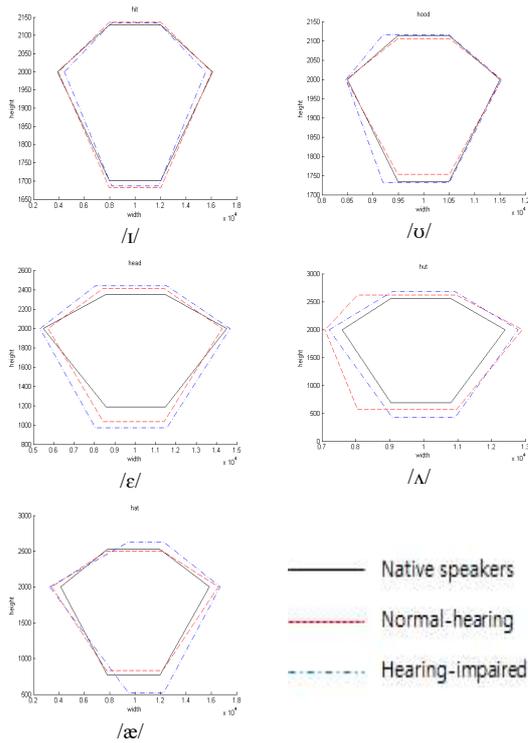


그림 6. 영어 원어민, 한국의 정상 청각인과 청각 장애인(여성)에 의한 5 개의 영어 모음에 대한 입술 모양(참고 : 각 모음에 대한 x 축과 y 축의 비율은 다르다.)

Fig. 6. Lip shape for the five English vowels produced by English native speakers, the Korean normal-hearing and the Korean hearing-impaired (female)(Note: scales of x-axis and y-axis for each vowel are differ.)

표 3은 발표자 그룹과 성별에 따른 다섯 가지 영문 모음의 분류비율 결과를 보여준다. 각 분류 백분율은 SVM을 사용하여 계산되었다. 표 3에서 알 수 있듯이, 화자 그룹이 분리되지 않은 경우, 모음 분류에 대한 정확한 비율은 63 %에 불과했다. 그러나 3 그룹의 연사와 성별을 따로 고려할 때 정확한 분류 비율은 더 높았다. 원어민 그룹의 경우 올바른 분류비율은 79%였고 모든 스피커를 조합 한 경우는 63%였다. 또한, 남성과 여성 그룹을 따로 비교했을 때 정확도는 더 높았다(남성과 여성 모두 89%). 한국의 정상 청력 그룹과 한국의 청력 장애 그룹에서 동일한 결과가 발견되었다. 한국의 정상청력 그룹에

서 올바른 비율은 69%였다. 성별로는 남성과 여성이 각각 78%와 74%였다. 한국인 청력 장애자의 정확한 비율은 남성과 여성이 함께 64 %였지만 그룹을 따로 고려할 때 각각 70%와 75%였다. 화자 집단과 성별을 따로 따로 다루었을 때 올바른 비율이 훨씬 높았기 때문에 SVM은 그림 2의 화자 분류와 성별 분류에 대해 두 번 사용되었고 그림 2의 입술형상 인식에서는 모음 분류에 사용되었다.

표 3. 화자 집단 및 성별에 따른 모음 분류의 분류비율

Table 3. Classification rate for vowel classification in terms of speaker group and gender

Speaker group	Classification	Total	Male	Female
All speaker groups	Correct classification rate	63%	-	-
	Misclassification rate	38%	-	-
Native speaker group	Correct classification rate	79%	89%	89%
	Misclassification rate	21%	11%	11%
Korean Normal-Hearing group	Correct classification rate	69%	78%	74%
	Misclassification rate	31%	23%	26%
Korean Hearing-Impaired group	Correct classification rate	64%	70%	75%
	Misclassification rate	36%	30%	25%

4. 결론

이 논문에서는 SVM 기법을 사용하여 5 개의 영어 모음에 대한 입술 모양을 개선하였다. 모음 입술 모양의 미세화는 영어 원어민, 한국인 정상 청력 및 한국어 청각 장애인이라는 세 가지 다른 연결자 그룹에서 모음 발음이 입술의 움직임과 어떻게 다른지 명확하게 보여준다. 이 연구의 결과에서, 청각 장애자가 영어 모음을 배울 때, 립 모양을 조정하여 원초 같은 소리를 얻는 것이 매우 중요하다.

입술 모양의 세련미에는 두 가지 이점이 있다. 첫째, 시각적으로 보기에 의존하는 청각 장애인을 위한 영어 모음에 대한 시각적 인 교수법이 눈으로 보기 힘들 정도로 혀 높이와 혀의 뒤쪽으로 모음을 가르치기보다는 제안되었다. 둘째, 이 연구는 SVM 기술을 사용하여 청각 장애인의 귀에 구별 할 수없는 소리에서 소리 특징을 추출 할 수 있음을 보여준다.

이 연구에서는 청각 장애인으로부터 데이터 수집의 어려움으로 인하여 5개의 영어 느슨한 모음만을

조사했다. 앞으로의 연구를 위해 다른 언어로 된 더 많은 영어 모음과 모음이 원어민과 제 2 외국어 사용자를 비교하여 조사 될 필요가 있다. 이 연구가 외국어 교육을 위한 새로운 교수법 및 학습 방법으로 인도 할 수 있기를 바란다.

REFERENCES

[1] K. I. Han, and H. J. Park, "Acoustic Analysis of English Vowels Produced by Korean Speakers with Hearing Impairment: With Reference to F0, F1 and F2 Frequencies," J. Language Sciences, vol.21 , no. 3, pp. 233-252, 2014.

[2] M. A. Higgins, A. Carney, and L. Schulte, "Physiological Assessment of Speech and Voice Production of Adults with Hearing Loss," J. Speech Hearing Research, vol. 37, pp. 510-521, 1994.

[3] K. Nicoladis, and A. Sfakiannaki, "An Acoustic Analysis of Vowels Produced by Greek Speakers with Hearing Impairment," Icphs XVI, pp. 1969-1972.

[4] P. Ladefoged, A Course in Phonetics. Orland: Hartcourt, 2001.

[5] S. H. Yoon, "Analysis of English Class Operation Status in Hearing Impaired Schools," Master thesis, Daegu University, South Korea, 2001.

[6] S. R. Hong, "Information Communication of English Education for Hearing Impaired: Based on Holistic Education," Korean Society for Holistic Education, vol. 14, no. 2, pp. 133-146, 2011.

[7] T. Y. Kim, Y. S. Kang, and H. S. Ko, "Achieving real-time lip-synch via SVM-based phoneme classification and lip shape refinement,"

[8] J. Hillenbrand, L. Getty, M. Clark, and K. Wheeler. "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am, vol. 97, pp. 3099-3111, 1995.

[9] P. Boersma, and K. Weenink, Praat: Doing Phonetics by Computer, <http://www.praat.org>, 2014.

[10] Y. W. Jung, A Study on Teaching on English Vowel Pronunciation through the Shaping of the Mouth," Master thesis, Yeosu National University, South Korea, 2002.



이 근 민(Kun-Min Lee)

1997년 5월 Johns Hopkins University, Special Education & Rehabilitation Technology 졸업 (박사)
 1997년 9월 - 현재 대구대학교 재활공학과 정교수
 2007년 3월 - 현재 대구대학교 재활공학센터 소장
 2010년 6월 - 현재 대구광역시 보조기구센터 소장

Interest: Rehabilitation Technology, Assistive Technology, Computer Assistance, Service Delivery System



한 경 임(Kyung-Im Han)

2006년 12월 University of Colorado, Linguistics 졸업(박사)
 2007년 - 2009년 대구카톨릭대학교 외래교수
 2008년 - 2013년 대구대학교 초빙교수
 2017년 - 현재 계명대학교 조교수

Interest: Acoustic Analysis, Linguistic Comparison



박 혜 정(Hye-Jung Park)

2005년 2월 대구가톨릭대학교, 전산통계학과 졸업(박사)
 2008년 3월 - 현재 대구대학교 기초교육원 초빙교수
 2014년 3월 - 현재 대구대학교 외래교수
 2015년 9월 - 현재 계명대학교 외래교수

Interest: Data Mining, Machine Learning, Bigdata Analysis