

한국어 음성 파형의 편집에 의한 한국어 음운 변화에 관한 연구

A study on the phonemic feature changes according to Korean speech waveform edition

김 선 일*, 홍 기 원*, 이 행 세*
Seonil Kim*, Kiwon Hong*, Haing Sei Lee*

要 約

한국어의 음성 파형의 일부분을 제거하거나, 일부분을 교체하므로써 얻어진 파형을 인간의 청각에 의한 음운 변별하는 실험을 수행하였다. 이 실험을 통해서 위치에 따라 급격한 음운 변화가 발생하는 천이구간, 가감해도 음운 변화가 극히 미세하여 음운 변화가 없는 구간, 그리고 상호 치환해도 변화를 가져오지 않는 등가 음운 구간 및 음가에 결정적 영향을 미치는 구간등을 확인하였다.

Abstract

A study on phonemic feature changes is accomplished by human perception of the discrimination of the phonemic feature of Korean edited speech waveform which is partially elimination or exchange. We found that speech waveforms has transitional, stationary, equivalent and critical phonemic parts.

I. 서 론

언어는 사람과 사람 사이의 가장 기본적인 통신 수단으로서 언어의 발달이 문명을 발전시키고 그 사회를 변모시켜 왔다고 해도 과언이 아니다. 이 언어를 구현하는 수단으로서 인간이 의미있는 소리를 낼 때 이것을 음성이라 한다.

지금까지 여러가지 통신 수단이 발전되었지만 그러나 음성은 여전히 강력한 통신 수단이며 단지 이를

멀리 전달하기 위하여 전화, 위성, 통신, 변조, 복조 등 여러가지 기술들이 발전하여 왔다. 기계 문명이 발전한 오늘날에는 사람과 사람 사이의 통신 뿐만 아니라 사람과 기계 사이의 통신 및 의사 전달도 중요한 상황이 되었다. 따라서 사람의 음성을 통하여 기계에 의사를 전달하기 위한 연구가 꾸준히 이루어져 왔다.

그러나 음성은 사람에 따라 특유한 성질을 갖고 있으며 같은 단어를 발음하더라도 동일한 형태로 발음하지 않는다는 문제가 있다⁽¹⁾⁽²⁾.

한국어는 한글과의 관계에서 특이한 과학적 구조로 인하여 자모를 분리하기 쉬운 형태로 구성되어 있

*아주대학교 전자공학과

*Dept. of Elec. Eng., Ajou Univ.

접수일자: 1994년 7월 27일

다. 따라서 초성, 중성, 종성으로 구별되는 한국어는 한글에서도 그 관계가 그대로 유지되고 있다. 물론 일부 예외도 있지만 발음되고 있는 음소가 그대로 대응되고 있는 것이다.

한국어 인식에서도 이러한 특성을 살려 초성, 중성, 종성을 음소 분리를 통하여 인식하려는 연구가 시도되고 있다³⁾⁽⁴⁾. 그러나 각 음소들은 연속해서 발음될 때 서로가 서로에게 영향을 끼쳐서 기대와는 달리 발음되고 있는 경우가 종종 있다. 앞의 음소가 뒤의 음소에 영향을 주어 변형되고 뒤의 음소가 앞의 음소에 영향을 주어 변형된 음소까지 포함할 경우 상당한 양의 음소가 존재하게 되어 점점 복잡한 문제로 발전하게 된다. 그러나 사람이 음성을 인지할 때는 이 변형된 음소는 무시되는 경향이 있다. 기계로 음성을 인식할 때에는 프레임(frame)별로 구간을 나누어서 프레임 인식을 하는 경우가 많다⁵⁾. 그러나 이 프레임 내의 데이터가 변형된 음소가 나타나는 천이 구간에서는 우리가 기대하는 바와 다르게 나타나는 데 보통은 이 사실을 외면하고 자음과 모음, 모음과 자음 등으로 칼로 두부 자르듯 나누는 경우가 많다. 이런 경우에 변형된 음이 어떤 음으로 변화하는가 하는 것은 상당히 흥미로운 과제로서 본 연구에서는 변형된 음소, 즉 천이구간에서의 음소가 어떤 음소로 변형되는지를 음성 편집과 편집된 음성의 청음을 통하여 연구해 본다.

II. 시작점 이동에 따른 음운의 변화

남녀 각각 1인의 단음을 12bits A/D 변환기를 통하여 10kHz로 샘플링(sampling) 하였다. 샘플링된 음성을 화일에 저장한 후 컴퓨터를 이용하여 편집하였다. 편집된 음성은 D/A 변환기를 통하여 아날로그 신호로 바뀌고 이 신호를 증폭기로 증폭한 후 음성을 청취하였다.

먼저 편집되지 않은 음성을 청취한 후 음성의 시작점을 조금씩 이동시켜 이를 청취하고 원래의 음성과 다른 음성이 들릴 경우 이를 기록하였다. 그림 1은 여성의 '사' 음성 파형을 나타내고 있다. A의 앞부분을 떼어 내었을 때 즉 청음의 시작점을 A로 옮겼을 때의 음성 파형은 처음 조사에서 마치 '타'처럼 들리고 있다. 시작점을 B로 이동시켰을 때 음성 파형은 '카'처럼 들린다. 시작점을 C로 옮겼을 때 음성 파형은 '가'처럼 들리고, 시작점을 D로 옮겼을 때 음성 파형은

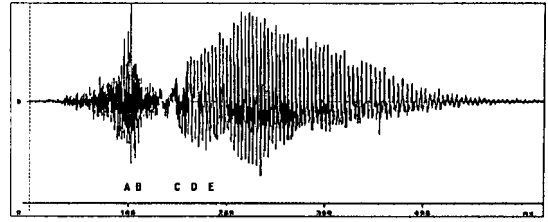


그림 1. 여성 음성 파형 '사'의 시작점 이동 파형
Fig 1. Various starting point waveforms for the female speech waveform /sa/

'바'처럼 들리며, 시작점을 E로 옮겼을 때 음성 파형은 '아'로 들린다. 그러나 파형의 끝부분에 가서는 심한 변음 현상이 나타났다.

표 1은 전술한 바와 같은 방식으로 청음 조사를 한 결과이다. 남녀 발성 구조의 차이로 인해 조금씩 다른 결과를 보여주고 있으나 대체로 잡음 성분이 강한 자음으로 시작되는 '사', '자', '차', '카', '타', '짜' 등에서 다른 음운이 많이 나타나고 있으며 따라서 이들이 백색 잡음성 자음에 가까우며 그중에서도 여성의 '사'가 백색 잡음성에 가장 가깝지 않나 추정된다. 표 2에는 '스'와 단모음이 결합된 음의 청음 결과가 나와 있다. 여기서도 여성의 음성이 남성보다 다양한 음운으로 들리고 있음을 알 수 있다. 물론 청음이란 것이 사람의 감각에 의존하기 때문에 기계적으로 이것은 '라'이고 저것은 '타'이다 라고 명확하게 구분하는 것이 아니고 감각적으로 자기가 알고 있는 음운 중의 하나에 사상되므로 그 정확도는 의문시 될 수도 있다. 그러나 중요한 것은 원래의 음성 파형에서 시작점을 이동시켜 청음 조사하였을 때 다른 음운으로 들리게 된다는 것이다. 이것은 기존의 음성 인식 프로그램들이 프레임이라는 세그먼트(segment)로 음성을 잘게 나누어서 프레임 단위로 인식을 시도하는 경우가 많은데 이 경우에 왜 오인식 되느냐 하는 것을 명확히 설명해 주고 있는 것이다. 음성 인식에 있어 요즈음의 추세는 인간의 인식 시스템을 모방하는 것이라 할 수 있다. 따라서 시작점을 바꾸어 들었을 때 다른 음운으로 들리거나 때로는 모호한 음운으로 들린다는 것은 음성 파형을 잘게 나누어서 행하는 인식의 근원적 모순점을 지적하는 것으로 생각할 수 있다.

한편으로는 프레임별 인식에 있어서 다른 음운으로 청음되는 부분을 천이구간으로 정의한다면 이 천이구간에 있어서의 오인식은 오인식 자체를 타당한

표 1. 시작점이 이동된 음성 파형의 청음 결과(초성 자음과 'ㅏ'가 결합된 음절).

Table 1. The listening results for the variable starting point speech waveforms(monosyllables composed of consonant and /a/).

음성 성별	가	나	다	라	마	바	사	자	차
남		빠	바		바, 마	바, 아	타	다	타
								라	
여	다	나, 라	바, 다				타	다	타
							카	바	파
							가		
						바			

음성 성별	카	타	파	하	까	따	빠	싸	짜
남	라	빠	타	타				따, 짜	따
	타, 파		다	파, 다				빠	
	파			바					
여		바, 파		카		따, 라		짜	까
		파		파				따	

표 2. 시작점이 이동된 음성 파형의 청음 결과('ㅏ'과 단모음이 결합된 음절).

Table 2. The listening results for the variable starting point speech waveforms(monosyllables composed of /s/ and monophthong).

음성 성별	사	서	소	수	스	시	새	세
남	타	터	토	투	트	티	태	테
	다	더	도	두	드	디	대	데
				뚜			배	
여	타	터	토	투	트	키	태	테
	카	더	도	두	드	치	대	케
	가			뚜	뜨	티	배	페
	바					따, 끼		베

인식으로 생각할 수 있으므로 인식을 향상에도 도움을 줄 수 있다.

Ⅲ. 자음의 치환에 따른 청음 효과

'사'와 '싸'는 자모 결합상 같은 모음에 자음이 바뀌는 형태이다. 각각의 파형에서 모음 부분은 그대로 두고 자음 부분만 치환한 후 이를 청음 조사하였다. 여기서 '사'의 'ㅏ'를 '싸'의 'ㅏ'로 바꾸었음에도 불구하고 '사'로 들린다는 사실을 발견했다. 또한 '싸'의 'ㅏ'를 '차'의 'ㅏ'로 대체해도 여전히 '싸'로 들리었다. 이것은 모음 'ㅏ'의 앞에 자음 'ㅏ'이 있느냐 또는 자음 'ㅏ'이 있느냐에 따라 모음 'ㅏ'의 음가가 바뀐다는 것을 의미하고 청음 조사에서 에너지가 적은 자음 부분은 청음에 거의 영향을 끼치지 못하고 에너지가 높은 부분이 크게 영향을 주므로 일반적인 기대와는 다른 청음조사 결과를 나타내고 있다.

이런 현상은 '사'와 '싸' 뿐만이 아니고 '자'와 '짜', '가'와 '까', '바'와 '빠'에서도 관찰되고 있다. 또한 '소', '수', '시', '세'와 '쏘', '쑤', '씨', '세'에서 동종의 모음에 대해 자음을 치환한후 청음 조사하면 같은 결과를 관찰할 수 있다. 또한 '쓰'의 'ㅏ'를 '사', '소', '수', '시', '세'의 모음과 결합시켰을 때 '스'의 'ㅏ'를 '싸', '쏘', '쑤', '씨', '세'와 결합시킨 후 청음 조사했을때도 같은 결과를 나타낸다. 다만 이중 모음과의 결합, 예를 들어 '시'의 'ㅏ'과 '쏘'의 'ㅏ'와의 결합에서는 부자연스러운 음으로 들리게 된다. 그림 2에 '사'와 '싸'에 대한 원 음성 파형과 합성된 음성 파형이 나타나 있다. 이런 결과를 종합해보면 자음보다는 모음에 정보가 집중되어 있음을 알 수 있으며 다른 모음과의 결합에서도 이런 현상이 나타날 것으로 유추할 수 있다. 따라서 이런 현상을 나타내는 음성에서는 프레임별 인식 방법으로 자음과 모음을 각각 인식한다는 것이 무의미할 수도 있다. 우리가 같은 모음이라고 인식하고 있는 모음조차도 서로 다른 모음이기 때문이다.

그림 3의 스펙트로그램을 보면 '사'의 'ㅏ'보다는 '싸'의 'ㅏ'가 발음 초기에 고주파 성분을 많이 가지고 있는 것으로 나타나 있다. 따라서 '싸'를 발음할 때 성대가 더 긴장하여 고주파 성분이 나타난다고 볼 수 있으며 결국 이것이 '사'와 '싸'를 구별짓는 요인이 된다.

음성 인식 시스템에 있어서 '사'와 '싸' 및 위에서 예로 든 음성들을 서로 구별하는 것은 쉬운 일이 아니다. 이런 음성에서는 자모의 개별적 인식보다는 모

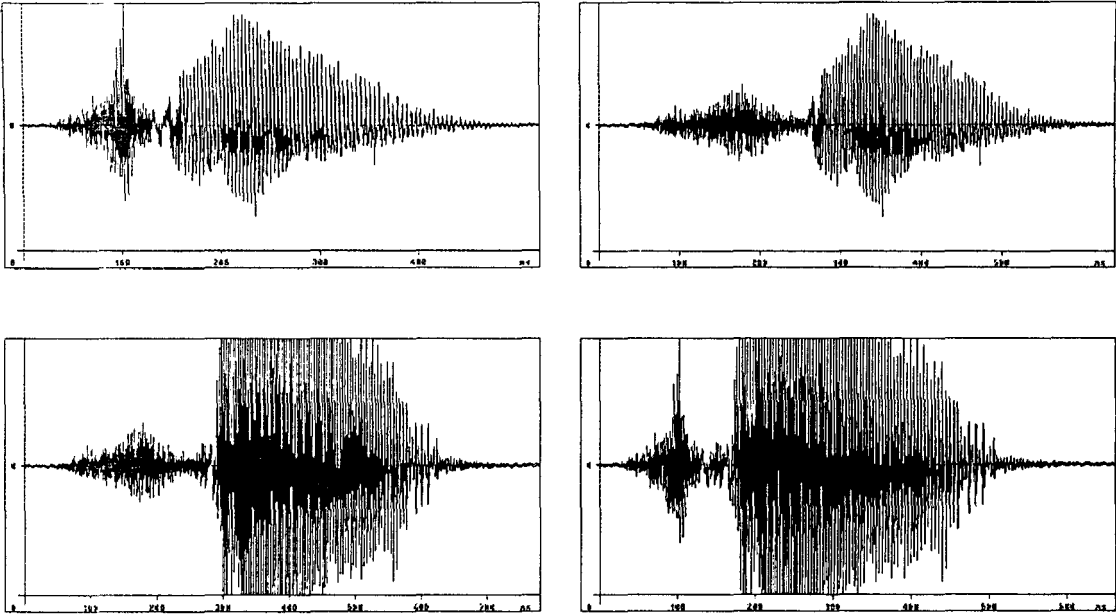


그림 2. '사'와 '싸' 음성 파형
 왼쪽 위 : '사'의 원 파형
 왼쪽 아래 : '싸'의 원 파형
 오른쪽 위 : '싸'의 'ㅅ'과 '사'의 'ㅏ'의 합성 파형('사'로 들림)
 오른쪽 아래 : '사'의 'ㅏ'과 '싸'의 'ㅏ'의 합성 파형('싸'로 들림)

Fig 2. Speech waveforms for the /sa/ and /ssa/
 Upper left : Original waveform of /sa/
 Lower left : Original waveform of /ssa/
 Upper right : Mixed waveform of /ss/ of /ssa/ and /a/ of /sa/ (It sounds /sa/)
 Lower right : Mixed waveform of /s/ of /sa/ and /a/ of /ssa/ (It sounds /ssa/)

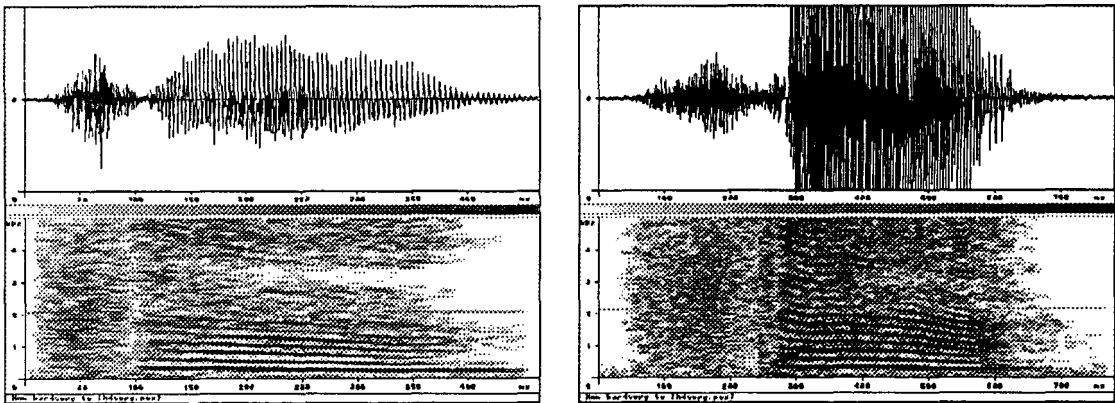


그림 3. 여자 음성 파형 '사'와 '싸'에 대한 스펙트로그램
 Fig 3. Spectrogram for the female speech waveform 'Isal' and 'Issal'

음의 인식에 더 비중을 두어야 할 것으로 생각한다.

IV. 결 론

음성 인식 기술은 나날이 발전하지만 그 기본적인 틀을 두가지이다. 즉 단어를 통째로 인식하는 방식이 그 하나고 다른 하나는 음소 단위 인식을 하여 전체를 인식하는 방법이다. 여러가지 방식의 기술이 쓰이고 있지만 가장 많이 쓰이는 기술은 프레임별 인식일 것이다. 본 연구는 프레임별 인식에서 앞 프레임이 없는 나머지 음성은 어떻게 들리겠느냐 하는 것에서 출발했다. 그래서 남녀 각각 1인이 '가', '나', '다', '라', '마', '바', '사', '자', '차', '카', '타', '하', '까', '따', '따', '빠', '싸', '짜'를 발음하게 하고 이 음성 데이터로 음성을 편집한 다음에 청음 조사를 실시하였다. 여기서 선택된 단어들은 우리나라에서 쓰이는 초성 자음에 모음 'ㅏ'를 붙여서 발음한 것이다. 잡음성 데이터로 분류되는 자음에 대해 파형의 시작점이 바뀔 때 큰 청음 조사를 실시하였다. 그 결과 남녀의 차이는 대부분의 음성이 다른 음성으로 들린다는 것이었다. 따라서 음성을 프레임별로 분할한다는 것 자체가 음성 인식에 상당히 부정적인 영향을 끼칠 수 있는 요인을 안고 있음을 증명하였다. 또한 프레임별 인식에서 오인식이 발생하는 근거를 명확히 제시하였다. 'ㅏ'가 아닌 다른 모음에 대해서도 이러한 현상을 관찰할 수 있는데 이것은 'ㅏ'와 단모음이 결합된 '샤', '서', '소', '수', '스', '시', '새', '세'에서 확인할 수 있었다. 자모 분리 인식의 경우에도 '사'와 '싸', '소'와 '쏘', '수'와 '쑤', '사'와 '씨', '세'와 '세', '자'와 '짜', '가'와 '까', '바'와 '빠'의 자모 치환 청음 조사를 통하여 모음은 그대로 두고 자음을 서로 치환하여도 치환 전의 음가와 같은 음가를 내는 것을 확인하였으며 '스'의 자음을 '싸', '쏘', '쑤', '씨', '세'의 자음에 대치시켰을 때 '쓰'의 자음을 '사', '소', '수', '시', '새'의 자음에 대치시켰을 때도 치환 전의 음가를 갖는 것으로 확인할 수 있었다. 따라서 자음보다는 모음의 인식이 더 중요한 것으로 결론 지을 수 있다.

음성은 사람마다 다르고 같은 사람도 상황에 따라 그 발성이 조금씩 다르다. 그리고 묘하게도 같은 단어에 대한 파형을 관찰하더라도 아주 판이하게 다른 모습을 관찰할 수 있다. 따라서 우리 말의 음가를 더 깊이 연구해 볼 필요가 있다. 그리고 음성은 연속적이고 통합적인 것이기 때문에 음성을 분할해서 연구

할 때에는 더 깊은 주의를 기울일 필요가 있다. 본 연구는 그런 면에서 문제 제기를 하고 있는 것이다.

V. 결 론

우리말 초성 자음과 단모음이 결합된 고립음을 사용하여 음성편집을 하고, 청음 실험을 통해서 음운의 변화를 추적하였다.

어떤 부분에서는 명확한 천이가 나타나기도 하며, 다른 부분에서는 약간 애매한 음가를 가지는 부분도 있고, 잡음부분을 교체하여도 음운에 변화를 초래하지 아니하는 경우 등, 편집에 따른 음운 변화에 대해서 기초적인 연구를 수행하였다. 이것은 앞으로 음성 부호화, 음성 합성 및 음성 인식은 물론 각종 음운 연구분야에서 활용이 가능할 것이다.

參 考 文 獻

1. T. W. Parsons, *Voice and Speech Processing*, McGraw-Hill Book Company, 1986.
2. R. D. Peacock and D. H. Graf, "An Introduction to Speech and Speaker Recognition," *Computer*, vol. 23, no. 8, August 1990.
3. 김석동, 이행세, "복합 신경망을 이용한 단음 인식에 관한 연구," 한국음향학회지 제11권 5호, 1992. 10.
4. 김석동, 신경회로망을 이용한 우리말 단음 인식에 관한 연구, 박사학위 논문, 1993.
5. J. R. Deller, J. G. Proakis and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Company, 1993.

▲이 행 세 1943年 8月 29日生.
 1966年 2月 : 전북대학교 전기공학과 졸업(공학사)
 1972年 2月 : 서울대학교 대학원 전자공학과 졸업(공학석사)
 1984年 8月 : 고려대학교 대학원 전자공학과 졸업(공학박사)
 1973年 2月 ~ 現在 : 아주대학교 전자공학과 교수

▲홍 기 원(Kiwon Hong) 1969년 1월 3일생
 1994年 2月 : 아주대학교 전자공학과 졸업(공학사)
 1994年 3月 ~ 現在 : 아주대학교 대학원 전자공학과 석사과정



▲김 선 일 1960年 3月 19日生
 1983年 2月 : 아주대학교 전자공학과 졸업(공학사)
 1985年 2月 : 아주대학교 대학원 전자공학과 졸업(공학석사)
 1994年 2月 : 아주대학교 박사과정 수료
 1985年 3月 ~ 1990年 2月 : 한국기계연구소 자동제어실 연구원
 1990年 3月 ~ 1990年 8月 : 한국기계연구소 자동제어실 선임연구원
 1990年 8月 ~ 現在 : 거제전문대 전자과 조교수