ISSN 2005-8063
2017. 12. 31.
Vol.9 No.4
pp. 43-50

말소리와 음성과학

Phonetics and Speech Sciences

한국음성학회지

http://dx.doi.org/10.13064/KSSS.2017.9.4.043

대학생들이 또렷한 음성과 대화체로 발화한 영어문단의 구글음성인식

Google speech recognition of an English paragraph produced by college students in clear or casual speech styles

양 병 곤*
Yang, Byunggon

Abstract

These days voice models of speech recognition software are sophisticated enough to process the natural speech of people without any previous training. However, not much research has reported on the use of speech recognition tools in the field of pronunciation education. This paper examined Google speech recognition of a short English paragraph produced by Korean college students in clear and casual speech styles in order to diagnose and resolve students' pronunciation problems. Thirty three Korean college students participated in the recording of the English paragraph. The Google soundwriter was employed to collect data on the word recognition rates of the paragraph. Results showed that the total word recognition rate was 73% with a standard deviation of 11.5%. The word recognition rate of clear speech was around 77.3% while that of casual speech amounted to 68.7%. The reasons for the low recognition rate of casual speech were attributed to both individual pronunciation errors and the software itself as shown in its fricative recognition. Various distributions of unrecognized words were observed depending on each participant and proficiency groups. From the results, the author concludes that the speech recognition software is useful to diagnose each individual or group's pronunciation problems. Further studies on progressive improvements of learners' erroneous pronunciations would be desirable.

Keywords: speech, recognition, English pronunciation, diagnosis, fricative, clear, casual, Google, soundwriter

1. 서론

오늘날 음성인식 기술은 다양한 국적을 가진 모국어 화자의 발음을 빠르고 정확하게 인식할 정도로 발전했다. 예전에는 바이보이스와 같이 일정 시간에 걸쳐 화자 개인의 목소리를 입력하여 발화 특징을 추출한 다음 그 화자에서만 작동하는 음성인식이 사용되었고, 단어인식률도 기존의 학습된 표현이나 어휘에 대해서는 높지만, 새로운 입력에 대해서는 오류가 나는 경우가 많았다. 최근의 구글음성인식기를 비롯한 새로운 소프트웨어의 성능은 많은 연구자들이 말로서 문서를 작성하고 편집할 수 있을 정도로 개선되었다. 영어를 외국어로 학습하는 한국인에게는 이러한 음성인식 기술을 이용하여 개인의 영어표현이 얼마나 잘 인식되고, 또 인식이 잘 되지 않는 단어의 발음상의 문제점을 찾는데 널리 활용할 수 있을 것으로 기대된다. 하지만 이러한 최근의 음성인식 기술의 발전에도 불구하고 인문학 분야에서는 컴퓨터 활용이 어려워서인지 아직까지 국내의 논문에서 실험을 통해 직접 조사하거나, 도구를 활용하는 방법을 구체적으로 제시한 연구가 드물다. 국외의 논문들은 주로 모국

* 부산대학교, bgyang@pusan.ac.kr
Received 1 November 2017; Revised 18 December 2017; Accepted 18 December 2017

어의 제한된 숫자나 단어에 대한 인식률에 대한 평가가 일부 있긴 하지만 본 연구에서 제안한 외국어로서 영어발음 인식률을 조사하고 발음의 문제점을 진단하는데 활용한 연구는 부족하다.

영어학습도우미 로봇을 만드는 데에 활용할 수 있는 음성인식기의 기초자료를 마련하기 위해 윤정희(2014)는 초등학생 16명을 대상으로 219개의 영어 단어를 읽게 하여 녹음한 음성을 안드로이드 모바일 기기에 설치된 Google Voice Actions를 이용해 음성인식률을 분석했다. 그의 연구 결과에 따르면 다섯 번의 발음에 대해 20점 단위로 매긴 평균 인식감도는 73.18점으로 다소 높게 나타났고, 모국어 발음으로 대체하거나 겹자음과 이중모음, 파열음, 한국어에 없는 모음 등의 음운배열과 관련된 정보들이 인식률에 영향을 준다고 보고했다. 덧붙여, 특정한 음소의 배열이 얼마의 확률로 결합되는지를 나타내는 음소배열확률(Phonotactic probability, Vitevitch & Luce, 2004)과, 비슷한 크기의 음소배열을 가진 실제 단어가 얼마나 많은지를 나타내는 어휘근접밀도(Lexical neighborhood density, Luce & Pisoni, 1998)로 오인식의 문제점들을 설명하려고 했다. 음소배열론은 Crystal(1992)에 따르면 어떤 한 언어에 나타나는 음소들의 배열유형을 나타낸다. 예를 들어, 영어에서는 초성자음군이 주로 spr, str, skr 등으로 규칙적으로 나타나며 smr과 같은 경우는 나오지 않는다. 이런 음절이 실제 나타나는 단어에서의 분포도 다르기 때문에 많이 나타날수록 어휘근접밀도도 높아지게 되고 인식기의 성능도 떨어질 수 있다. 그래서 윤정희(2013)는 학습하기 어려운 음운요소가 들어가 있는 단어들이 오히려 인식률이 높게 나타난 현상에 대해, 길이가 짧은 단어는 어휘근접밀도가 높고 단어수가 많은데 비해 길이가 긴 단어들은 잘못 인식이 될 만한 단어가 적어서 일부 음소를 부정확하게 발음하더라도 더 정확히 인식한 것으로 추정했다.

일상생활에서 화자는 환경에 따라 발화방식을 달리하는데 Lindblom(1990)은 아주 시끄러운 환경에서는 좀 더 크고 또렷하게 천천히 발화하고(Hyper-dimension), 조용한 환경에서는 약간 낮은 목소리로 빠르게 발화하는(Hypo-dimension) 방식으로 청자를 의식해서 자신의 목소리를 조절한다고 했다. 실제 이런 구분은 이분법으로 나눠지기 보다는 양극단 사이에 여러 가지 연속된 단계의 한 부분에 해당한다. 이런 Hypo-dimension 내에서도 추가로 변화를 보이는데, Fowler & Housum(1987)은 자연발화에서 이미 한번 발화한 단어는 두 번째 이후에는 더 짧게 발음한다는 점을 보고했으며, Wright(2003)도 일상생활에서 빈도가 많이 나타나는 단어일수록 모음공간이 더 좁아지며 조음동작을 적게 한다고 했다. 음향적으로는 양병곤(2012, 2014)이 미국인 남성 9명이 또렷한 발음과 대화체 발음을 조사해본 결과 피치값과 스펙트럼에서 차이를 보임을 보고했다. 그는 또렷한 발음은 대화체에 비해 피치가 높으며, 스펙트럼에서도 고주파 영역으로 갈수록 높아짐을 보였다. 음성인식기는 음향적 자료를 근거로 이뤄지기 때문에 발화방식에 따라 인식률이 달라질 것으로 예상된다.

이 연구의 목적은 영어전공 대학생들이 서로 다른 발화 양식으로 영어문단을 읽었을 때 음성인식기가 어떤 표현이나 단어를 잘 인식하고, 잘 인식되지 않는 표현이나 단어의 음성언어학적 특징은 무엇인지 등을 살펴봄으로써 이들의 영어발음을 진단하고 교정하는데 근본적인 도움을 주는 것이다. 연구문제를 구체적으로 서술하면 다음과 같다.

1. 대학생들이 발화한 영어문단에 대한 단어인식률은 어떠한가?
2. 대학생들이 또렷한 음성과 대화체로 발화한 영어문단의 어구별 단어인식률은 각각 어떠한가?
3. 또렷한 음성의 단어인식률 평균값으로 구분한 상・하위 집단별 단어인식 유형은 어떠한가?

이러한 연구결과는 대학생들의 영어발음학습에 필요한 기초자료를 제공하고, 새로운 연구와 분석방법은 다양한 수준의 학습자와 한국어를 비롯한 외국인의 발음학습 성취도에 대한 평가를 통해 개인별 맞춤형 학습방향을 제시하거나, 궁극적으로는 구글음성인식기와 같은 장치의 개선에 도움이 될 것으로 기대된다.

2. 연구 방법

2.1. 참여자

연구의 참여자는 영어음성학 과목을 수강한 대학생 중 33명이다. 이들은 영어교육을 전공하는 2-3학년생으로 원어민의 회화수업을 비롯한 영어전공과목을 이수하였고, 본인이 평가한 영어의사소통능력에서 발음부분을 중급이상으로 표시한 학생들이다. 이들은 한 학기 동안 영어자음과 모음발음에 대해 기본적인 교육을 받았으며, 교수자는 학기말에 주어진 영어문단에 대한 말하기 평가를 한다고 공지했고, 이들이 원어민의 발음을 들으며 연습을 하도록 했다. 이렇게 영어전공 대상자를 선택한 이유는 구글음성인식기가 어느 정도 구별되는 영어발음을 하는 경우에 인식이 바르게 되기도 하고, 실제 교육현장에서 적극적으로 참여한 대학생들을 대상으로 함으로써 이들에게 공통으로 나타나는 영어발음 학습의 문제점을 찾아 차후의 교육과정에 반영할 수 있기 때문이다. 앞으로 보다 낮은 등급의 학습자들을 대상으로 단어인식률은 낮더라도 실험해 보거나, 일정 기간에 걸쳐 다양한 방식의 발음교육을 실시한 학습자들을 대상으로 발음학습의 성취도를 추정하는 데 활용하는 연구도 필요하다.

2.2. 녹음 자료 수집

실험에 사용한 영어문단은 The speech accent archive(2017)의 원문을 이용했다. 이 사이트에는 영어원어민과 비원어민화자들이 구어의 전달문 형태의 문단을 자연스럽게 읽은 녹음 표본을 제공하고 있다. 이 문단 자체는 영어에서 나타나는 자모음을 모두 넣으려고 구성하는 과정에서 담화적인 맥락의 전개로서는 부자연스러우나 영어학습자의 단어인식률을 점검하는

데는 도움이 될 것으로 보아 사용하게 되었다. 녹음에서는 구두점이 들어간 원문을 읽게 했지만, 이 논문에서는 69개의 연속된 단어로 된 텍스트를 아래와 같이 11개의 문장이나 어구로 나누어 설명하기로 한다. 이 가운데 일부는 완전한 문장도 있지만 이 논문에서는 어구로 표현한다.

1 please call stella
2 ask her to bring these things
3 with her from the store
4 six spoons of fresh snow peas
5 five thick slabs of blue cheese
6 and maybe a snack for her brother bob
7 we also need a small plastic snake
8 and a big toy frog for the kids
9 she can scoop these things into three red bags
10 and we will go meet her Wednesday
11 at the train station

문단을 어구로 나눈 이유는 구글음성인식기에서 구두점 없이 대소문자를 가리지 않고 단어 인식을 했고, 인식 자료 분석 과정에서 사용한 R(v.3.4.1, 2017)의 문자열 처리 함수(str_count)에서 the가 두 번 나오는 행이나, these에 들어 있는 the가 함께 단어로 처리하는 것을 피하기 위해서 분리했다. 하지만, 6행의 her brother는 빈칸 다음에 오는 her와 다음 단어의 두 번째 음절에 쓰인 빈칸 없는 her를 다르게 분석해 주어서 같은 행에 두었다. 덧붙여, 아래한글에서 행을 바꿀 때 자동으로 첫 글자가 대문자로 바뀌는 Wednesday를 그대로 두고 나머지는 소문자를 사용했다.

녹음하는 과정은 대학생들이 조용한 연구실에서 PC컴퓨터에 연결된 젠하이져 헤드셋 마이크를 입에서 10 cm 정도로 띄운 채 발음하게 하고 이를 GoldWave(v5.70)버전을 이용해서 PCM signed 16 bit mono로 컴퓨터에 녹음했다. 이들은 영어문단을 모두 두 번씩 발음했는데 처음에는 또렷한 음성으로 약간 느린 속도로 모든 단어를 명확하게 발음했고, 이어서 대화체로 평소의 대화속도에 맞춰 약간 빠르게 발음했다. 서로 다른 양식으로 녹음하는 음성 사이에는 약간 쉬게 했고, 잘못 발음한 경우에는 그 문장 부분만 다시 되풀이 발음하게 하고 인식결과 텍스트에서는 되풀이된 앞부분을 삭제했다.

2.3. 음성인식과정

음성인식 과정은 먼저 구글문서작성(docs.google.com)에 로그인하여 Add-ons에 EFV-Solution이 딥러닝 뉴럴네트워크 알고리즘을 이용하여 개발한 Speech Recognition Soundwriter를 설치한다. 이어서 음성인식을 켜고, 컴퓨터 본체에서 재생되는 대학생들의 녹음된 음성이 윈도우즈의 스테레오믹스기능을 통해 인식기에 입력되고 실시간으로 인식된 단어가 연이어 텍스트파일로 저장됐다. 사전에 녹음한 음성의 음량은 헤드셋을 사용해서 양호한 편이나 일부 대학생의 원래 녹음한 목소리가 다소 약한 경우가 있어서 프랏의 폴더읽기와 scale to peak 함수를 이용해서 각 음성의 최대값을 기준으로 한꺼번에 증폭한 다음 재생했다. 인식기의 성능을 사전에 확인하기 위해 The speech accent archive(2017)에 제공되어 있는 피츠버그 태생의 42세 미국인 남성(speakerid=61)과 브룩클린 태생의 45세 여성(speakerid=121)의 음성을 컴퓨터에서 바로 재생함과 동시에 스테레오믹스기능으로 GoldWave에 녹음한 뒤, Praat(Boersma & Weenink, 2017)에서 최대값으로 증폭하여 인식시켜 보았다. 그 결과 두 사람의 음성에 대해 spoons를 moons로 잘못 인식한 것을 제외하고는 영어문단의 나머지 모든 단어를 바르게 인식했다.

2.4. 인식문단 분석

인식된 영어문단은 참여자별로 연속된 텍스트 파일 형태로 저장하여 3.2절에서 제시한 것처럼 3~9개의 단어로 된 11개의 문장이나 어구로 나누었다. 아래한글의 찾아 바꾸기를 이용하여 첫머리에 오는 단어를 찾아 한꺼번에 줄바꾸기를 통해 행별로 분할했다. 이어서 잘못된 단어로 인식되거나 누락된 단어 때문에 분리가 안 된 어구는 연구자가 원문 문단과 대조하면서 수작업으로 하나씩 분할한 다음, 참여자별로 인식된 단어목록으로 된 11개의 행을 복사하여 엑셀의 열마다 붙여서 하나의 통합문서로 저장했다.

각 행에 대한 분석은 이 통합문서를 R로 불러와 참여자와 11개의 어구별로 되풀이하며 인식된 단어수와 목록, 인식되지 않는 단어수와 목록을 매트릭스자료에 모아 하드디스크의 결과파일에 덧붙여 쓰는 스크립트를 만들어 실행했다. R에서의 처리과정을 간략하게 서술해보면, 먼저 문자열 처리 라이브러리인 stringr패키지를 설치하고 11개 행의 원래의 단어목록을 r1, r2...r11 등의 변수 리스트로 불러들여 메모리에 저장했다. 앞에서도 언급했듯이 문자열 처리 함수(str_count)에서 빈칸이 있는 경우와 없는 경우가 다르게 처리되기 때문에 행의 시작과 끝에 있는 단어는 각각 뒤쪽과 앞쪽에 빈칸을 넣어 저장했다. 이어서, 결과를 저장할 빈 매트릭스를 13개 행과 7개의 열로 구성한 다음, 첫 줄에 화자의 번호와 속도에 따른 구분기호를 넣고, 두 번째 행부터는 원래의 단어목록에 들어있는 단어 순서대로 하나씩 해당 단어가 화자별로 인식된 행의 단어목록에 들어 있는지를 str_count 함수를 이용해 확인하고, 있으면 인식된 단어수를 증가시키고 동시에 인식된 단어를 문자열에 덧붙여 변수로 저장하고, 없으면 잘못 인식된 단어수를 증가시키고 동시에 문자열을 덧붙여 변수로 처리한 뒤, 결과로 구한 변수값을 매트릭스의 열에 차례로 입력했다. 이 매트릭스의 뒤쪽 열에는 화자마다 인식된 문자열을 입력하여 나중에 잘못 인식된 단어를 확인하기 편하도록 했다. 각각의 매트릭스는 최종적으로 하드디스크의 결과파일에 자동으로 덧붙여 쓰며 저장했다.

마지막으로 이렇게 처리한 결과파일을 엑셀에서 다시 불러와 공백을 기준으로 텍스트나누기를 시행하여 엑셀의 평균과 표준편차 등을 구하는 함수나 정렬기능을 이용하여 전체적인 단어인식률을 계산하거나 화자, 어구, 발화양식별로 잘못 인식된 단어를 분석하는데 활용했다. 인식 단어의 빈도 분포는 정렬하여 텍스트파일로 저장한 다음 R의 table 함수를 이용하여

구했다. 구글음성인식기에서 5와 6 등은 숫자로 표기되는 경우가 많아 분석의 편의상 모두 찾아서 five와 six로 바꾸어 처리했고 things나 peas는 복수형이 제대로 인식된 경우만 옳게 인식된 단어수로 포함했음을 밝힌다. 여기서는 편의상 단어인식률을 중심으로 음성인식결과를 다루지만, 복수형은 쉽게 주변 단어를 통해 추정할 수 있기 때문에 앞으로 원어민에게 문단 텍스트를 들려주어서 필요한 정보를 제대로 전달했는지를 평가해보는 연구가 필요하다.

3. 분석 결과와 논의

3.1. 음성인식 결과

대학생들이 또렷한 음성과 대화체로 발화한 총 단어수는 4,554개이고, 이중에 3,325개를 바르게 인식하여 평균 73%(표준편차=11.5%)의 전체 단어인식률을 기록했다. 발화 방식별 음성인식 결과를 보면 먼저 또렷하게 발화한 총 단어수 2,277개 중 1,761개를 바르게 인식하여 77.3%(표준편차=9.6%)의 단어인식률을 기록했다. 대화체로 발화한 경우에는 바르게 인식한 단어가 1,564개이고 전체에서 차지하는 비율은 68.7%(표준편차=11.5%)를 차지해서 또렷한 발화에 비해 인식률이 8.6% 하락했다. <표 1>은 이러한 결과를 요약하여 보여준다.

Styles	Number of correct words	Number of incorrect words	Sum	Word recognition rates (%)
Clear	1,761	516	2,277	77.3
Casual	1,564	713	2,277	68.7
Total	3,325	1,229	4,554	73.0

표 1. 대학생들이 발음한 영어문단의 음성인식
Table 1. Word recognition of the English paragraph produced by Korean college students

구체적으로 조사해본 결과 또렷한 음성에서 가장 높은 단어인식률을 보인 대학생은 95.7%를 기록했고, 가장 낮은 단어인식률을 보인 대학생은 52.2%를 보였다. 대화체에서는 92.8%에서 44.9%까지 범위에 걸쳐 다양하게 분포되어 있다. 이렇게 대화체의 단어인식률과 범위가 상대적으로 낮은 이유는, 대화체로 발음할 때 조금 빠른 속도로 연음이나 음운변동을 일으켜 발음했기 때문으로 여겨진다. 음성인식이 되지 않은 구체적인 단어들과 빈도수 분포에 대해서는 다음 절에서 자세히 살펴보기로 한다. 한편 일부 대학생들은 발음평가를 의식해서 대화체라고 해도 또렷한 음성보다 약간 더 빠른 대화체로 발음했기 때문에 이 차이가 적었다고 여겨진다.

이번에는 어구별로 또렷한 음성과 대화체에 대한 단어인식률을 <표 2>를 통해 살펴본다.

Row No.	Clear		Word recognition rates	Casual		Word recognition rates
	Correct	Wrong		Correct	Wrong	
1	95	4	96.0	78	21	78.8
2	182	16	91.9	149	49	75.3
3	146	19	88.5	136	29	82.4
4	84	114	42.4	59	139	29.8
5	92	106	46.5	78	120	39.4
6	238	26	90.2	237	27	89.8
7	223	8	96.5	211	20	91.3
8	224	40	84.8	197	67	74.6
9	154	143	51.9	124	173	41.8
10	195	36	84.4	166	65	71.9
11	128	4	97.0	129	3	97.7

표 2. 대학생들이 발화한 영어문단의 어구와 발화양식별 단어인식
Table 2. Word recognition of the English paragraph produced by Korean college students according to the phrase number and speech styles

<표 2>에서 보면 또렷한 음성에서 가장 높은 단어인식률을 보인 어구는 11번으로 단어수도 4개이고 전치사구라는 하나의 의미단위로 되어 있어서 그런지 97%의 단어인식률을 보였다. 이어서 7번과 1, 2, 6번 어구가 90%이상의 단어인식률을 보였다. 그런데 4번과 5번 어구는 각각 42.4%와 46.5%로 50%가 채 안 되는 아주 낮은 단어인식률을 보인다. 이렇게 낮은 인식률을 보인 이유는 다른 어구에 비해 마찰음 [s]가 많이 들어간 것이 한 가지 원인으로 보이는데, 다음 단락에서 구체적으로 분석해 보기로 한다. 9번 어구도 51.9%를 기록해서 이들 세 어구는 우연의 확률에 못 미치거나 가깝다. 대화체에서도 11번과 7번 어구는 각각 97.7%와 91.3%를 기록했고, 6번 어구는 또렷한 음성에 비해서 한 단어만 차이가 나고 세 번째로 잘 인식된 어구다. 대화체에서 아주 낮은 단어인식률을 보인 어구는 또렷한 음성에 대한 인식률의 순서와 같이 4번과 5번 어구로 각각 29.8%와 39.4%를 기록했고, 9번 어구는 41.8%로 또렷한 음성의 발음에 비해 단어인식률이 10%나 떨어졌다. 약간 느린 속도로 발음한 또렷한 음성에 비해 평소의 대화 속도로 다소 빠르게 발음한 대화체 음성으로 갈수록 단어인식률이 떨어지는 경향을 보이고 있는데, 1번과 2번 어구에서 각각 17.2%와 16.7%로 가장 많이 떨어졌고, 3번, 5번, 6번, 7번 어구에서 안정적으로 가다가 8번, 9번, 10번 어구부터 10%이상을 기록했고 마지막 어구에서는 0.8%상승했다. 마지막 어구를 뺀 나머지 어구의 평균을 내면 9.8%의 하락률을 보였다. 이러한 경향은 연구자가 녹음된 음성을 여러 번 들으며 음성인식의 문제점이 무엇인지 파악하는 과정에서 일부 대학생들이 대화체의 발음 요구에 서둘러 속도를 내어 발음하다가 중간 부분에서는 약간 늦추었다가 다시 빠르게 발음하면서 오류가 늘어났거나, 어구에 들어있는 단어들의 조음 특성상 발화양식에 관계없이 에러를 보인 것 등을 원인으로 추정하는데 앞으로 더 세밀한 조사가 필요하다.

이번에는 구체적으로 어떤 단어들이 이런 발화양식에 따른 단어인식률의 차이를 보였는지를 <표 3, 4>를 통해 살펴보기로 한다. 지면상 오인식 단어수가 7개 이상이란 임의 기준을 적

용해서 표로 나타내었다.

Words	Freq	Words	Freq	Words	Freq	Words	Freq
slabs	33	she	20	bob	15	we	10
thick	33	snow	20	of	15	for	8
spoons	31	six	19	kids	14	the	8
bags	28	fresh	18	blue	11	things	8
red	26	these	18	frog	11	will	7
scoop	26	can	17	cheese	10		
peas	20	her	16	five	10		

표 3. 대학생들이 또렷한 음성으로 발화한 영어문단의 오인식 단어와 빈도

Table 3. Frequency distribution of unrecognized words of the English paragraph produced by Korean college students in clear speech

Words	Freq	Words	Freq	Words	Freq	Words	Freq
these	37	red	26	six	14	stella	8
her	35	of	24	a	13	Wednesday	8
slabs	33	she	24	bob	12	call	7
spoons	33	can	20	will	12	for	7
thick	33	things	19	five	11	go	7
bags	28	frog	18	meet	11	into	7
fresh	27	kids	16	three	11		
peas	27	we	16	toy	11		
scoop	27	blue	15	with	11		
snow	27	cheese	15	ask	8		

표 4. 대학생들이 대화체로 발화한 영어문단의 오인식 단어와 빈도

Table 4. Frequency distribution of unrecognized words of the English paragraph produced by Korean college students in casual speech

인식된 단어들을 기능어와 내용어로 나누어 살펴보면 기능어가 오인식률을 높인 주된 요인으로 여겨진다. 영어학에서는 단어를 내용어와 기능어로 구분하는데, 명사, 형용사, 동사 등은 의사전달의 핵심이 담긴 내용어이고 관사, 전치사, 대명사 등은 문법적인 기능을 담당하고 있어서 기능어로 분류한다 (Fromkin & Rodman, 2013). 특히, 또렷한 음성에서 대화체로 갈수록 her, of, she, can, we, a, will, with, for, into등의 기능어가 눈에 띄게 높은 오인식 빈도수를 보인다. 구체적으로 <표 3>에서 her의 오인식 빈도수는 16번인데 비해 <표 4>에서는 35번으로 두 배 이상이나 증가했고, of도 또렷한 음성에서는 15번이었는데 대화체에서 24번으로 증가했음을 알 수 있다. 한편 많은 단어에서 마찰음이 오인식의 원인으로 보인다. 앞서 원어민 2명의 발화에서도 유일하게 spoons가 moons로 잘못 인식되었는데, 대학생이 발음한 음성의 인식률을 <표 3, 4>에서 보면, slabs, spoons, scoop, snow 등이 높은 오인식을 보였고, 앞에서 분석한 어구별 오인식률에서도 4, 5, 7, 9번 어구는 주로 [s]로 시작하는 단어가 눈에 띄게 많은 것을 알 수 있다. 이와 마찬가지로 복수형접미사로 쓰이는 마찰음 때문에 단어인식률이 낮은 단어들로는 spoons, bags, peas, kids, things 등의 순서로 오인식률이 낮아진다. peas는 piece로 많이 인식되었다. 이런 경향을 뒷받침하듯이 또렷한 음성에서 초성과 종성 양쪽에 이 마찰음이 들어간 slabs는 두 가지 발음양식에서 모두 인식이 되지 않았다. 이러한 낮은 단어인식률은 개인별 발음오류에도 문제가 있지만, 원어민의 발음(spoons)에서도 오류가 난 점을 생각해볼 때, 음향적으로 공명도(sonority scale)가 다른 자음에 비해 상대적으로 낮고, 이 마찰음의 스펙트럼에서는 4000~8000 Hz 사이에 절단주파수(cutoff frequency)가 나타나(Kent & Read, 2002; Pickett, 1987), 이 부분이 적절히 입력이 되지 않거나 활용되지 않아 인식에 오류를 가져왔을 것으로 여겨진다. 또 다른 마찰음이 포함된 these, thick, things, three의 오인식도 fresh, frog, five와 함께 높게 나타났는데 대화체에서 유성마찰음이 들어간 these의 오인식 빈도가 가장 높은 37번을 보였다. 실제 이 단어는 1번과 9번 어구에 두 번 쓰였기 때문에 반 이상이 오인식되었다고 볼 수 있다. her도 4번 쓰였기 때문에 빈도가 높게 나타났다. 이렇게 빈도가 높은 단어의 인식오류는 서론에서 살펴본 어휘근접밀도에 따른 원인도 포함되어 있을 것이다 (Vitevitch & Luce, 2004; Luce & Pisoni, 1998). 이 외에도 초성에 유성음이 없는 한국어의 특성상 bags, bob, blue가 오인식되는 비율이 높았고, red, frog, three발음에서는 대학생들의 영어조음이 불안정한 [r] 발음이 원인의 일부로 보인다. 또렷한 음성에서 bags는 5명의 발음에서만 바르게 인식되었고, 21명의 발음에서는 back으로 잘못 인식되었고, 대화체에서 8명의 발음이 바르게 인식되고 back으로 잘못 인식된 경우가 13명이나 되었다. 종성의 유성자음도 발음에 문제가 있음을 알 수 있다. kids는 또렷한 음성에서 19명의 음성이 바르게 인식되었고, 또렷한 음성과 대화체 음성 모두에서 10명의 발음이 key(s)로 인식되었는데 이완모음 [ɪ]를 너무 강하게 발음한 것이 원인으로 보인다. 어절로 된 fresh snow는 대화체에서 6명의 발음이 바르게 인식되었고, 19명의 발음은 fresno로 인식되었다. these things는 대화체에서 15명의 발음이 바로 인식되었고, 13명의 발음이 things로 인식되었는데 앞 단어를 this로 인식한 경우가 9번이 나왔고, 나머지는 his나, a, the로 인식된 예도 있었다. 구글인식기의 기능에 이런 수의 일치 부분을 처리해 주거나 문맥에 따라 인식단어를 결정하는 상위수준의 후처리가 필요할 것으로 여겨진다. 여기에서는 조사할 수 없었지만, 서론에서 보았던 어휘근접밀도(Luce & Pisoni, 1998)도 영향을 주었을 것으로 보여진다. 결국 이런 단어들을 중심으로 개인별 발음의 문제점을 진단하고 이를 바르게 교정시켜서 인식률을 높이거나 음성인식기의 근본적인 문제점을 찾아 성능을 개선하는 방향의 실험이 필요하다.

요약하면, 영어문단의 단어인식률은 대학생들마다 독특한 개인별 발화특성과 발화양식에 따라 달라지고, [s]와 같은 특정한 음성이 들어간 어구에서 보았듯이 구글음성인식기 자체의 음성처리방식에서 오는 요인들이 결합되어 단어인식률이 낮아졌음을 알 수 있다.

3.2. 상·하위 집단별 음성인식 결과

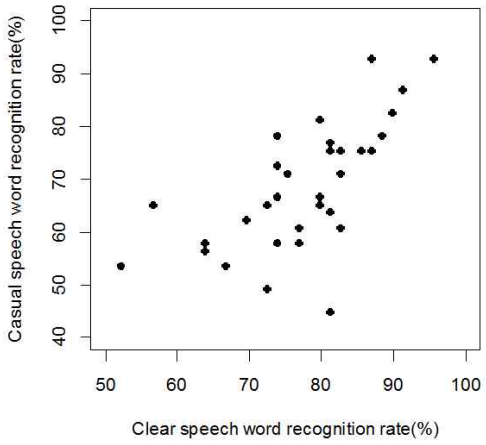
이번에는 또렷한 발음에 대한 단어인식률의 전체평균인 77.3%

를 기준으로 상위집단 17명과 하위집단 16명으로 나누어 집단별 단어인식 경향을 조사해 보았다. 이렇게 집단별로 나누면 전체 결과에서 볼 수 없는 특징을 파악할 수 있고 집단별 또는 수준별 영어발음 교육의 방안이나 개인별 맞춤형 지도방안을 찾는데 활용할 수 있을 것이다. 덧붙여, 이러한 집단별 구분은 대학생들의 발음을 원어민에게 들려주어 전체적인 발음평가 점수를 기준으로 나눌 수도 있고, 단어인식률과 평가점수가 상관관계를 보인다면 발음평가를 대체할 수 있을 것으로 기대되는데 앞으로 더 연구가 필요하다. 우선 상·하위집단별 단어인식률의 평균과 표준편차를 살펴보면 <표 5>와 같다.

	High level group		Low level group	
Styles	mean(%)	s.d.(%)	mean(%)	s.d.(%)
clear	84.5	4.7	69.7	7.3
casual	74.4	12.0	62.6	8.1

표 5. 상·하위 집단별 단어인식률 평균과 표준편차
Table 5. Means and standard deviations of word recognition rates by high and low level groups

개인별 특징을 자세히 살펴보기 위해서 상·하위집단별 단어인식률 분포를 그래프로 나타내면 <그림 1>과 같다.



각 화자별로 발화양식별 대응상관계수는 0.66($p<.05$)으로 약한 상관을 보이고 있다. 이런 관계는 단어인식률에서 영어전공대학생을 대상으로 했고, 특정단어의 분포나 기능어 등을 화자마다 다르게 발음한 결과가 영향을 미친 것으로 생각된다. 특히 상위집단에서 또렷한 음성의 단어인식률과 대화체의 단어인식률은 최대 36.2%의 차이를 보였고, 이어서 21.7%, 17.4%, 14.5%, 13%, 11.6%로 이어지고, 나머지는 모두 10%이하로 떨어지는 경향을 보였다. 하위집단에서는 최대 23.2%, 18.8%, 15.9%, 13% 등의 순서로 단어인식률의 차이를 보였고, 나머지는 모두 10%이하를 보였다. 대다수 화자들이 대화체에서 단어인식률이 대체로 하락하는 경향을 보였는데, 이와 반대로 상위집단에서는 두 명의 단어인식률이 2%~6%상승하였고, 하위집단에서는 세 명의 단어인식률이 2%~9%로 상승한 결과를 보였다. <그림 1>의 맨 아래에 위치한 한 화자는 또렷한 음성에서는 81.2%의 단어인식률을 보였으나 대화체에서는 10번과 11번 어구가 완전히 인식되지 않아 44.9%로 낮아졌다. 이러한 참여자별 특성을 살펴보기 위해 <그림 2>에 개인별로 또렷한 음성과 대화체의 발화양식에 따라 인식률의 변화를 나타내어 보았다.

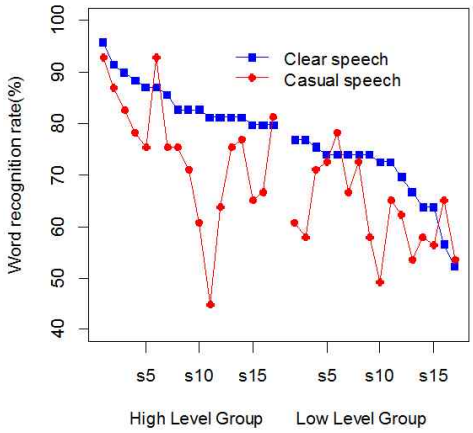


그림 2. 또렷한 음성과 대화체의 상위집단(High Level)과 하위집단(Low Level)의 개인별 단어인식률 분포
Figure 2. Distribution of word recognition rates of high and low level groups in clear and casual speech

<그림 2>에서 보면 또렷한 음성보다는 대화체에서 개인별 차이가 크게 나타나고 있는데, 이는 앞의 <표 3, 4>에 대한 논의에서도 보았듯이 개인별로 기능어에 대한 발음이나 마찰음에 대한 오인식에 덧붙여, 보다 빠른 속도로 발음하는 과정에서 생략이나 연음을 자연스럽지 발음하지 못해서, 구글인식기가 기능어와 내용어가 합쳐진 새로운 영어단어로 인식했기 때문으로 여겨진다. <그림 1>에서 대화체에서 인식률이 가장 많이 떨어진 대학생은 상위집단의 s11번임을 알 수 있고, 또한 하위집단에서는 s10 대학생도 대화체에서 매우 낮아진 단어인식률을 보이고 있음을 알 수 있다. 이러한 화자 개인별 특성을 그림으로 나타낸다면, 개인별 발음의 문제점을 진단하고 맞춤형으로 개선하는 방법을 찾는데 활용할 수 있을 것으로 기대된다. 덧붙여, 일정 기간에 걸쳐 영어발음을 지도하고 학습자의 발음 습득여부를 알아보기 위해 음성인식기로 단어인식률을 구해 학습효과를 비교분석할 때는 발화양식과 수준별 특성을 고려해서 측정해야 할 것으로 여겨진다.

지금까지 발화실험에 참여한 대학생들의 단어인식률의 평균을 기준으로 상·하위집단으로 나누어 살펴본 결과 전체적으로는 수준별로 약한 상관관계를 보이지만, 일부 개인의 발화양식별 차이가 있으므로 이런 요인을 연구방법에 반영할 필요가 있음을 알 수 있다.

4. 요약 및 결론

이 논문에서는 음성인식 기술을 발음진단과 개선에 활용할 목적으로 33명의 대학생들이 또렷한 음성과 대화체의 두 가지 양식으로 발음한 영어문단을 구글음성인식기를 이용해 인식시키고, 원문텍스트와 인식된 어구의 단어를 비교했다. R의 분석 스크립트와 table 함수를 이용해서 단어별로 바르게 인식된 빈도수와 오인식된 단어들을 조사했다. 이어서 또렷한 발음의 단어인식률 평균값을 기준으로 상·하위집단으로 나누었을 때 집단별 단어인식률의 분포와 개인별 음성인식 경향을 단어를 구성하는 음성의 유형별로 분석하여 대학생들의 영어발음의 문제점을 찾아보았다. 연구결과를 요약하면 다음과 같다.

첫째, 대학생들이 발화한 영어문단의 전체 단어인식률은 73%이고 표준편차가 11.5%로 나타났다. 영어교육을 전공하고 있는 대학생들이 영어발음에 대해 기본적인 교육을 받고 적극적으로 실험에 참여하게 되어 다소 높은 단어인식률을 보였다.

둘째, 발화 양식에 따라 분리하여 보면 또렷한 음성에서 77.3%를 기록했고, 대화체에서 68.7%를 보였는데, 다소 빠른 속도로 발화한 대화체의 단어인식률이 전체적으로 낮았다. 구체적으로 단어인식률은 개인별 발화의 특성과 내용어보다는 기능어의 인식이 낮았고, 마찰음이 들어간 단어들에서 대체로 낮은 인식을 보였는데, 그 원인으로 개인별 발화오류와 원어민의 음성인식에서도 드러난 점을 고려해볼 때 음성인식기 자체의 문제점도 있었다.

셋째, 실험에 참여한 대학생들의 또렷한 음성에서 구한 단어인식률의 평균을 기준으로 상·하위집단으로 나누어 살펴본 결과 전체집단에서 볼 수 없었던 개인별 특성이 드러났다. 음성인식기를 이용해서 발음의 문제점을 진단하고 이를 집단별 수준별 맞춤형 교육을 제공하는데 활용하려면 발화양식에 따른 변수도 고려할 필요가 있다.

이러한 결과를 보면 대학생들의 영어발음의 문제점을 진단하는데 음성인식기가 매우 유용하다고 결론을 내릴 수 있다.

앞으로 영어학습자들이 단기간이나 장기간에 걸쳐 발음학습을 했을 때 단어인식률이 얼마나 변하는지, 또 외국인들의 한국어학습에서도 이런 분석방식을 적용하여 연구해볼 계획이다.

참고문헌

Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer. Retrieved from http://www.fon.hum.uva.nl/praat/ on October 2, 2017.

Crystal, D. (1992). *An encyclopedic dictionary of language and languages.* Middlesex, U.K.: Blackwell.

Fowler, C., & Housum, J. (1987). Talkers' signalling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26, 489-504.

Fromkin, V., & Rodman, R. (2013). *An introduction to language.* Belmont, CA: Wadsworth.

Jusczyk, P., Luce, P., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory & Language*, 33, 630-645.

Kent, R., & Read, C. (2002). *Acoustic analysis of speech.* San Diego, CA: Singular Publishing Group.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H-H theory. In W. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403-439). London: Kluwer Academic Press.

Luce, P., & Pisoni, D. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, 19, 1-36.

Pickett, J. (1987). *The sounds of speech communication: A primer of acoustic phonetics and speech perception.* Austin, Texas: pro-ed.

R. Core Team. (2017). R: A language and environment for statistical computing. Retrieved from https://www.r-project.org/ [R Foundation for Statistical Computing, Vienna, Austria] on October 1, 2017.

Vitevitch, M., & Luce, P. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, 36(3), 481-487.

Wright, R. (2003). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI* (pp. 75-87). Cambridge: Cambridge University Press.

Yang, B. (2012). Pitch and formant trajectories of English vowels by American males with different speaking styles. *Phonetics and Speech Sciences*, 4(1), 21-28. (양병곤 (2012). 발화방식에 따른 미국인 남성 영어모음의 피치와 포먼트 궤적. *말소리와 음성과학*, 4(1), 21-28.)

Yang, B. (2014). Spectral characteristics and formant bandwidths of English vowels by American males with different speaking styles. *Phonetics and Speech Sciences*, 6(4), 91-99. (양병곤 (2014). 발화방식에 따른 미국인 남성 영어모음의 스펙트럼 특성과 포먼트 대역. *말소리와 음성과학*, 6(4), 91-99.)

Yun, J. (2014). *Analysis of Google Voice Actions' recognition of English word pronunciations by Korean young learners of English for the purpose of developing an English teaching assistant robot.* M.A. Thesis, Kyungpook National University. (윤정희 (2014). *Google 음성인식프로그램에 의한 한국 어린이 영어학습자의 영어단어 발음인식 실태분석: 영어학습도우미 로봇개발을 목적으로.* 경북대학교 석사학위논문.)

• **양병곤 (Yang, Byunggon)**
부산대학교 영어교육과
부산시 금정구 장전동 30
Tel: 051-510-2619
Email: bgyang@pusan.ac.kr
Homepage: http://fonetiks.info/bgyang