



Production and perception of Korean word-initial stops from a sound change perspective

Jin-Woo Kim*

Department of Korean Language Education, Korea University, Seoul, Korea

Abstract

Based on spontaneous speech data collected in 2020, this study examined the production and perception of Korean lenis, aspirated, and fortis stops. Unlike the controlled experiments of previous studies, lenis and aspirated stops of males in their 30s were not distinguished by voice onset time (VOT) in spontaneous speech. Perceptual experiments were conducted on young females, the leaders of language change. F0 was found to serve as the primary cue for the perception of lenis stops, and then VOT distinguished the aspirated and fortis stops. The fact that the sounds were always perceived as lenis stops when F0 was low, irrespective of whether VOT was short or long, showed that F0 plays an absolute role in the perception of lenis stops. However, in some cases the aspirated and lenis stops were distinguished only by VOT, which does not happen in production. In terms of sound change, disagreement between production and perception systems occurs when sound change is in progress. In particular, when production change precedes perception change, it indicates that the sound change is in its latter stages. Young females still maintain the previous system in perception because the distinction of lenis and aspirated stops by VOT was valid in their parents' generation. In other words, VOT is still used for perception to communicate with other groups.

Keywords: sound production and perception, Korean stops, sound change, VOT (voice onset time) merger

1. 서론

본 연구는 발성 유형에 따른 어두 폐쇄음의 발화 및 지각 양상을 살펴보고, 둘 사이에 어떠한 차이가 있는지 그리고 그 차이를 음 변화의 관점에서 어떻게 설명할 수 있을지 알아보는 것을 목적으로 한다.

지금까지 많은 연구들에서는 어두의 평음과 격음이 발화되는 방식이 변화하고 있음에 주목하였다(Bang et al., 2018; Byun,

2016; Jeong, 2010; Kang, 2014; Oh, 2011; Silva, 2006). 예전에는 평음과 격음이 VOT(voice onset time)에 의해 변별되었다면, 지금은 젊은 여성 세대를 중심으로 하여 평음과 격음의 VOT가 합류되어 F0에 의해 발화가 변별된다는 것이다. 이러한 양상을 통시적으로 살펴본 Bang et al.(2018), Kang(2014)에서는 젊은 여성일수록 격음의 VOT가 줄어들어 VOT 합류가 일어나고 있다고 보았다. Bang et al.(2018)은 나이가 더 다양한 요인을 고려하여 평음과 격음의 VOT 합류는 후행 모음이 고모음이 아닌 경우,

* iks2297@korea.ac.kr, Corresponding author

Received 29 July 2021; Revised 2 September 2021; Accepted 5 September 2021

© Copyright 2021 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

단어의 빈도가 높은 경우에 더 잘 관찰된다고 하였다. 한편 Byun(2016)은 격음의 VOT가 줄어들 뿐만 아니라 평음의 VOT가 늘어났고, 따라서 1980-1990년대 출생 여성에서 VOT가 평음과 격음 변별의 일차적 자질로 사용되지 못하고 그 기능을 F0이 이어받았다고 주장하였다.

위에서 언급된 연구들을 포함하여 지금까지의 선행 연구에서는 대부분 통제된 음성 자료만을 대상으로 하였고, 자유 발화를 대상으로 한 연구는 드물었다. 이에 Kang & Guion (2008)은 1977년 이후 출생 그룹과 1966년 이전 출생 그룹을 둘로 나누어 폐쇄음의 발화 양상이, 주의 깊은 발화(clear speech), 단독 발화(citation), 담화(conversation)에 따라 어떻게 다른지 살펴보았다는 점에서 의의를 찾을 수 있다. Kang & Guion (2008)의 연구 결과 1977년 이후 출생 그룹에서는 단독 발화와 주의 깊은 발화에서만 평음과 격음의 VOT에서 차이가 나타났지만 1966년 이전 출생 그룹에서는 모든 상황에서 VOT에 차이가 있었다. 또한 젊은 세대에서는 평음과 격음의 주의 깊은 발화 시 VOT보다는 F0을 주요한 자질로 이용한다는 것이 밝혀졌다.

이러한 연구에 더하여 본 연구에서는 2020년의 10-60대 남녀의 자유 발화를 대상으로 발성 유형에 따른 어두 폐쇄음의 발화 양상을 살펴보고자 한다. 점진적 음 변화를 알 수 있는 방법은 경계선의 변화를 살펴보는 것인데(Janson, 1983), 한국어의 어두 폐쇄음은 VOT와 F0을 축으로 하는 평면을 그리고 여기에 각 분절음에 대한 산점도와 경계선을 그릴 수 있으므로 이에 따른 경계선 변화를 확인할 것이다.

어두 폐쇄음의 발화 양상이 세대에 따라 달라지고 있는 것이 확인되었을 때, 함께 다뤄야 할 것은 지각적인 측면이다. 예를 들어 음 변화가 일어나고 있는 젊은 세대에서, 발화 시 측정되는 VOT 및 F0이 지각 시에도 그대로 그 역할을 하는지 혹은 그렇지 않은지 등의 연구가 필요하다. 이러한 점에서 어두 폐쇄음의 지각 양상을 살펴본 Jang & Shin(2010), Kang(2010), Kim(2001), Kim et al.(2020) 등은 주목할 만한 연구이다. 이들 연구에서는 공통적으로 서울 방언의 젊은 화자들이 평음과 격음의 지각에서 VOT보다는 F0을 이용한다고 밝혔다. 구체적으로 Kim(2001)에서는 20-30대 화자를 대상으로 지각 실험을 한 결과 낮은 F0이 평음의 지각에 핵심적인 역할을 한다고 보았다. 그리고 Jang & Shin(2010)에서는 대구 화자에 비해 서울 화자가 F0에 민감하다고 하였으며, Kang(2010)에서는 평음과 격음의 지각 변별에 있어서 40세 이상의 집단보다 20대의 집단에서 F0의 단서 역할이 더 크다고 하였다. 또한 Kim et al.(2020)에서는 청년층과 노년층 모두 격음과 평음을 지각할 때 VOT보다 F0이 더 큰 역할을 하지만, 청년 여성 집단일수록 F0에 더 민감하다는 것을 보여주었다. 하지만 아쉬운 점은 이들 연구에서 발화와 지각에 있어 어떠한 차이가 있는지는 구체적으로 다루지 않았다는 것이다. 즉 지금까지 대부분의 연구는 발화와 지각에서

평행하게 F0의 역할이 커지고 있다는 것에만 주목하였으며, 발화와 지각 양상이 구체적으로 어떻게 다른지를 살펴본 연구는 드물다고 할 수 있겠다.

발화와 지각의 차이는 음 변화가 완료되었는지 그렇지 않은지의 중요한 판단 기준이 된다. 음 변화가 진행 중일 때에는 발화와 지각의 불일치가 나타나며, 따라서 사람들은 자신이 발화하는 대로 지각하지 않는다(Janson, 1983; Ohala, 1993; Pinget et al., 2020). 한편 음 변화에서 발화가 선행하는지 지각이 선행하는지에 대해서는 이견이 있기도 하다. Ohala(1993)에서는 오지각(misperception)으로부터 음 변화가 발생한다고 보았고 반대로 Janson(1983)에서는 발화의 변화가 지각의 변화보다 선행한다고 보았다. 그런데 최근의 Pinget et al.(2020)에서는 음 변화의 단계에 따라 선행하는 것이 다를 수 있다고 주장하기도 하였는데, 음 변화의 시작은 지각에서 일어나긴 하지만 음 변화의 완성은 발화에서 먼저 완성된다고 하였다.

이에 본 연구는 통제된 음성 자료가 아닌 자유 발화의 평음·격음·경음의 VOT와 F0을 측정하여 서로 어떠한 경계를 형성하고 있는지 확인할 것이며, 이때 머신러닝(machine learning) 중 하나인 SVM(support vector machine) 알고리즘을 활용할 것이다. 그다음으로는 언어 변화의 주도 계층인 젊은 여성을 피험자로 하여 청취 실험을 실시한 후 발화와 지각의 차이를 살펴볼 것이다. 끝으로 이러한 차이를 음 변화의 관점에서 어떻게 이해할 수 있는지를 논의함으로써 마무리하고자 한다.

2. 어두 폐쇄음의 발화

2.1. 연구 방법

2.1.1. 발화 자료 수집

이 연구에서는 국립국어원에서 구축한 ‘일상 대화 음성 말뭉치 2020(National Institute of Korean Language, 2021)’을 사용하였다.¹ 이 음성 말뭉치는 2021년 3월에 공개된 최신 자료로, 두 화자가 어떤 주제에 대하여 15분 정도 자유롭게 대화하는 것이 녹음되어 있다. 예컨대 두 화자는 ‘스포츠/레저’, ‘여행지’, ‘계절/날씨’ 등의 주제를 받고 자유롭게 대화한다. 이 말뭉치 구축에 참여한 화자는 전국 방언의 총 2,739명(10-60대의 남성 및 여성)이며 말뭉치에는 약 500시간 정도의 분량이 담겨 있다. 음성 파일은 이미 역양구 단위로 분할되어 있고 JSON 파일에 발음 전사 및 철자 전사가 되어 있어 연구자가 사용하기에 매우 편리하다. 또한 각 화자별로 출생지, 주요 거주지, 현재 거주지 역시 JSON 파일로 제공되고 있어 화자의 지역까지 파악할 수 있다. 다만 음성 파일이 16 kHz의 표본 추출률, 16 bit 양자화의 선형 PCM 파일 형식으로 제공되기 때문에 PCM 파일을 wave 파일로 변환하는 과정이 필요한데, 본 연구에서는 Python (version 3.8)

1 이 말뭉치는 ‘모두의 말뭉치’ 홈페이지(<https://corpus.korean.go.kr/main.do>)에서 누구나 무료로 이용할 수 있으나, 사전에 말뭉치 이용을 신청해야 하며 내려받기는 최대 3회까지 가능하다.

코딩을 통해 wave 파일로 변환하였다.

본 연구에서는 출생지와 주요 거주지가 모두 서울 혹은 경기인 사람만을 대상으로 하였다. 성별(남·여)과 연령대(10~60대, 만 나이)별의 12개 그룹에서 각각 15명의 화자를 분석 대상으로 삼고자 하였으나, 10대 남성과 60대 남성은 각각 9명, 13명만이 분석 대상이 되었다.² 이에 분석 대상이 된 화자의 수는 남성 82명, 여성 90명으로 총 172명이다.

어두 폐쇄음은 억양구 초의 어두 폐쇄음을 분석 대상으로 삼았다. 이는 평음의 경우 억양구 내부에서 유성음화나 경음화가 일어날 수 있기 때문이다. 그리고 어두 폐쇄음은 연구개음을 대표로 삼아 분석하였다. 자유 발화에서는 통제된 발화보다 다양한 변이형이 나타날 수 있다는 점을 고려하여 토큰 빈도가 높은 /ㄱ/은 각 화자 별로 최대 5까지 분석하였으며, /ㅋ/와 /ㄲ/는 등장하는 전부를 분석 대상으로 삼았다. 분석 대상이 된 어두 폐쇄음은 총 1,510 개이며 구체적으로는 다음 표 1과 같다.

표 1. 어두 폐쇄음별 표본의 크기(단위: 개)
Table 1. Sample size of word-initial stops

남성							
음소	10대	20대	30대	40대	50대	60대	합계
ㄱ	39	77	71	72	71	63	393
ㅋ	17	40	41	33	27	25	183
ㄲ	14	21	21	24	24	19	123
합계	70	138	133	129	122	107	699
여성							
음소	10대	20대	30대	40대	50대	60대	합계
ㄱ	74	75	74	74	80	67	444
ㅋ	42	38	45	34	31	27	217
ㄲ	25	26	25	26	26	22	150
합계	141	139	144	134	137	116	811

2.1.2. 발화 음성 분석

어두 폐쇄음의 VOT와 F0을 측정하는 데에는 Praat(Boersma, P., & Weenink, D., 2021, version 6.1.41)을 이용하였다. VOT는 음성 파형을 참고하여 파열 순간부터 성대 진동으로 인해 주기적 파형이 나타나는 모음 파형의 개시 순간까지를 측정하였다. 후행하는 모음의 F0은 F2 이상이 나타나는 지점부터 모음의 중간 지점까지의 F0을 평균 내어 계산하였다. 모음 시작 부분의 F0만을 측정하지 않은 이유는 한국어 어두 폐쇄음의 특징을 고려하기 위함이었다. 자음의 종류가 후행하는 모음 시작 부분의 F0에 영향을 미치는 것은 범언어적으로 존재하는 현상이지만, 한국어는 자음의 종류가 모음 시작 부분의 F0뿐만 아니라 전반적인 F0에 영향을 미치며 심지어 강세구의 두 번째 모라까지 영향을 미칠 수 있다(Jun, 1993).

이렇게 측정된 VOT와 F0은 12개의 그룹(남·여×10~60대)별로 머신러닝을 통해 평음, 경음, 격음이 어떤 분포를 보이는지

살펴보았다. 지금까지 어두 폐쇄음 분포 연구에서 머신러닝을 활용한 연구는 거의 없었는데, 최근 대두되고 있는 다양한 머신러닝들은 데이터를 정확하고 빠르게 분석할 뿐만 아니라 입력된 데이터를 기반으로 새로운 데이터를 예측하기도 한다. 본 연구에서는 어두 폐쇄음의 발화 양상의 분석에서 나아가 이를 토대로 새로운 데이터를 예측하여 후술할 청취 실험에서 비교할 것이므로 머신러닝 알고리즘을 적극 활용할 것이다.

2.2. 발화 음성 분석 결과

발성 유형별 VOT 측정 결과를 제시하면 다음 표 2 및 그림 1과 같다. 40~60대 그룹에서 VOT의 평균은 격음>평음>경음 순으로 나타나고 있다. 그런데 놀랍게도 10~30대 그룹에서는 격음과 평음의 VOT 평균에 유의미한 차이가 없었다($p>.05$). 그림 1에서 보이듯, 경음의 VOT는 연령대별로 서로 큰 차이가 나타나지는 않지만, 격음의 VOT는 젊은 세대로 오면서 줄어들어 평음의 VOT와 차이가 없어지고 있다. 또한 40~60대의 평음과 격음의 VOT 차이를 보면 여성의 경우 이미 남성에 비해 그 차이가 적었음을 알 수 있는데, 이는 이전의 연구들에서 보고된 바와 같이 한국어 VOT 변화에서 여성이 선두에 있었음을 알려주는 것이다. 그리고 여성 10~20대의 격음 VOT는 55~58 ms 정도로 다른 그룹에 비해 확연하게 낮는데, 이는 60대 남성의 평음 VOT와 비슷한 수준이다.

표 2. 자유 발화에서의 VOT 평균 및 표준편차(단위: ms)
Table 2. VOT mean and standard deviation of spontaneous speech

남성						
발성 유형	10대	20대	30대	40대	50대	60대
평음	66.7 (16.0)	66.2 (19.4)	59.1 (17.4)	61.5 (20.5)	60.1 (17.6)	56.3 (19.2)
격음	64.5 (21.6)	65.4 (22.7)	62.7 (19.7)	71.7 (18.5)	91.0 (25.8)	79.7 (21.3)
경음	16.4 (4.7)	15.6 (8.4)	16.9 (6.4)	20.9 (17.1)	23.4 (10.1)	18.0 (8.6)
여성						
발성 유형	10대	20대	30대	40대	50대	60대
평음	57.7 (19.6)	55.5 (21.1)	63.3 (19.6)	57.1 (21.0)	54.4 (18.9)	52.4 (16.1)
격음	58.3 (19.5)	55.2 (19.0)	62.3 (20.0)	67.3 (26.3)	69.8 (23.8)	68.4 (28.2)
경음	14.3 (9.1)	16.2 (8.8)	12.0 (5.9)	16.7 (9.6)	13.5 (5.1)	14.4 (7.0)

2 본 연구에서 활용한 음성 말뭉치에서 10대 남성의 경우 서울/경기 출신 화자의 수가 9명밖에 되지 않았으며, 60대 남성의 경우 15명이 있었지만 2명의 화자는 발화 상태가 좋지 않아 배제하였다.

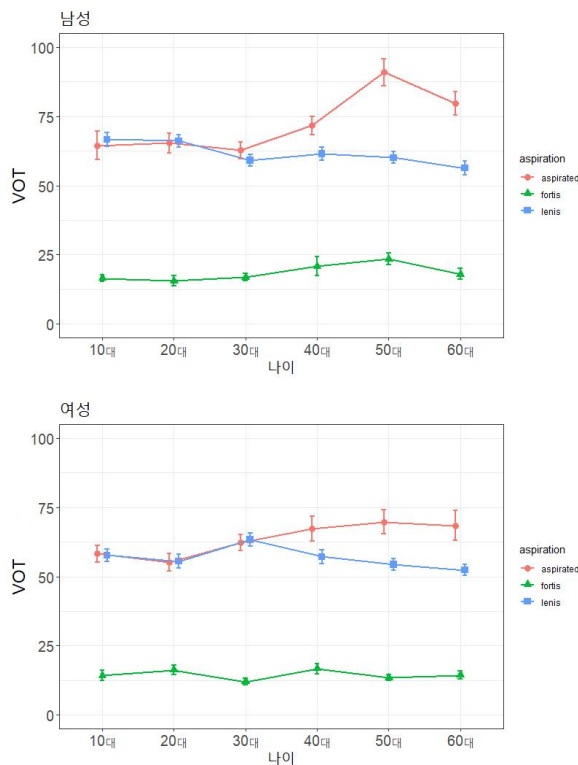


그림 1. 자유 발화에서의 VOT(ms) 평균 및 표준 오차
Figure 1. VOT(ms) mean and standard error of spontaneous speech

한편 발성 유형별 F0 측정 결과를 제시하면 다음 표 3 및 그림 2와 같다. F0은 Byun(2016)을 참고하여 개인마다 F0 값을 표준화한 z-score를 사용하였다. 이는 F0이 성별뿐만 아니라 연령에 따라서, 나이가 같은 연령이라도 개인마다 큰 차이가 있기 때문이다.

F0은 12개의 그룹에서 모두 격음>경음>평음 순으로 나타났다. 또한, 앞에서 살펴본 VOT와 비교했을 때 F0은 연령에 따라 크게 다르지 않은 모습을 보여준다. 다만 60대와 10대의 평음 F0 차이를 성별로 비교했을 때, 남성에게 비해 여성의 평음 F0이 더 많이 낮아진 것을 확인할 수 있다. 이는 Byun(2016)의 결과와 일치하는 것으로, Byun(2016)에서는 여성이 자음 구별을 위해 평음에서 F0을 낮추는 식으로 F0을 적극적으로 이용하는 것으로 해석하였다.

표 3. 자유 발화에서의 F0(z-score) 평균 및 표준편차(단위: z-score)
Table 3. F0(z-score) mean and standard deviation of spontaneous speech

남성						
발성 유형	10대	20대	30대	40대	50대	60대
평음	-0.92	-0.86	-0.90	-0.96	-0.88	-0.87
	(0.26)	(0.47)	(0.39)	(0.31)	(0.45)	(0.59)
격음	0.79	0.80	0.91	0.88	0.72	0.70
	(0.54)	(0.62)	(0.78)	(0.50)	(0.61)	(0.85)
경음	-0.03	0.14	-0.11	0.10	0.12	0.17
	(0.52)	(0.65)	(0.54)	(0.57)	(0.76)	(0.69)
여성						
발성 유형	10대	20대	30대	40대	50대	60대
평음	-0.99	-0.98	-0.89	-1.00	-0.85	-0.79
	(0.33)	(0.33)	(0.39)	(0.35)	(0.42)	(0.38)
격음	0.84	0.94	0.93	0.85	0.93	0.75
	(0.49)	(0.51)	(0.60)	(0.56)	(0.47)	(0.59)
경음	0.13	0.10	-0.05	0.13	-0.02	-0.16
	(0.61)	(0.54)	(0.63)	(0.72)	(0.54)	(0.69)

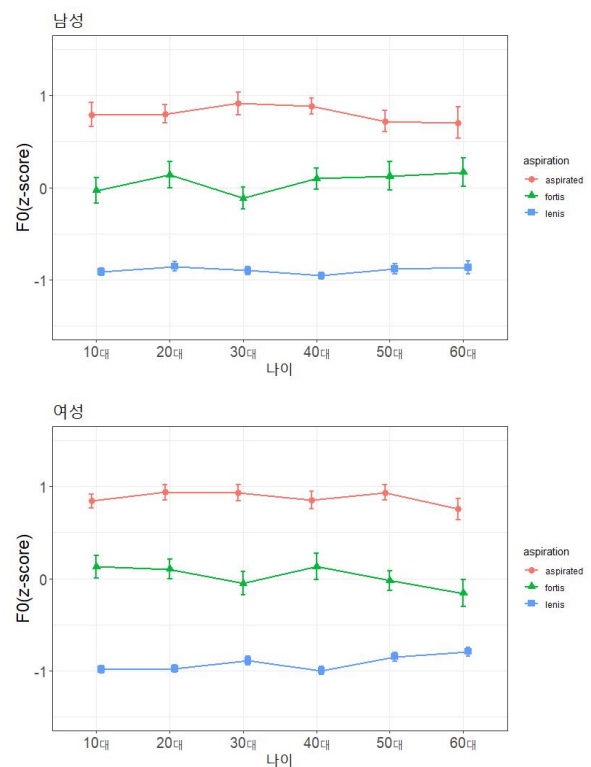


그림 2. 자유 발화에서의 F0(z-score) 평균 및 표준 오차
Figure 2. F0(z-score) mean and standard error of spontaneous speech

VOT와 F0의 데이터를 가지고 머신러닝 중 하나인 선형 SVM을 이용하면 평음, 격음, 경음을 최대한 잘 분류하는 경계선을 특정할 수 있다. 이때 독립 변수는 VOT, F0이며 종속 변수는 발성 유형이 된다. SVM은 주로 분류를 위해 사용되는 머신러닝으로, 중심선과 마진(margin) 사이가 최대화가 되는 경계선(decision boundary)을 찾는 알고리즘이다. 따라서 본 연구의 관심 대상인 평음, 격음, 경음의 분포를 알아보기 위한 최적의 알고리즘이라고 할 수 있다. 본 연구에서는 scikit-learn(Pedregosa et al., 2011)에서 제공하는 SVM 알고리즘을 Python에서, Package

e1071(Meyer et al., 2021)에서 제공하는 SVM 알고리즘을 R(R Core Team, 2021, version 4.0.5)에서 사용하였다. Python은 그림 3의 시각화를 위해서만 사용하였으며 나머지는 모두 R을 사용하였다.³

그림 3에서 SVM 알고리즘으로 나뉜 모습을 보면 평음·경음의 경계선이 대체로 양(+)의 기울기를 가진 것을 볼 수 있다. 즉 같은 F0이라면 VOT가 비교적 긴 것이 평음, 짧은 것이 경음으로 분류되고, 같은 VOT라면 비교적 높은 F0이 경음, 낮은 F0이 평음으로 분류된다는 것을 의미한다.⁴ 그러나 전체적인 산점도를 보면 평음과 경음의 VOT와 F0은 많이 겹치지 않는다. 경음의 F0은 평음의 F0보다 대개 높은 곳에 위치하며, 평음의 VOT는 20 ms 미만에는 거의 분포하지 않는다.

한편 평음과 격음의 경계는 남·여 50-60대를 제외하고는 대체로 기울기가 0, 즉 x축에 평행한 것을 볼 수 있다. 이는 평음과 격음이 F0만으로도 잘 분류된다는 것을 보여준다.⁵ 한편 남성 50-60대, 여성 60대의 평음·격음의 경계선은 음(-)의 기울기를 가지고 있는데, 이는 평음과 격음 일부가 같은 F0을 공유하기 때문이며, VOT에 의해 구분될 수 있다는 것을 의미한다. 즉 이

그룹들에서는 같은 F0일 때, VOT가 비교적 긴 것이 격음, 짧은 것이 평음으로 분류된다. 또한, VOT가 길수록 격음으로 분류될 가능성이 크다는 것도 동시에 보여준다.

R에서는 만들어진 SVM 모델의 독립 변수 중요도를 ROC(receiver operating characteristic) 곡선을 통한 AUC(area under the curve)로 제시해준다. AUC는 값에 따라 $0.5 < AUC \leq 0.7$ 일 때 낮은(low) 정확도, $0.7 < AUC \leq 0.9$ 일 때 보통의(moderate) 정확도, $0.9 < AUC < 1$ 일 때 높은(high) 정확도, $AUC = 1$ 일 때 완벽한(perfect) 정확도를 나타낸다(Fischer et al., 2003). 표 4의 각 셀은 이러한 기준에 따라 VOT와 F0의 변수 중요도를 4가지 단계의 진하기로 표현하였다.

표 4를 보면 평음과 격음의 분류에서 F0은 모든 그룹에서 0.9를 넘는 AUC를 보여 높거나 완벽한 정확도를 나타냈다. 반면 VOT는 남성 50-60대에서만 보통의 정확도를 보이고 나머지 그룹에서는 낮은 정확도를 보였다. 평음과 경음의 분류에서는 VOT가 모든 그룹에서 높은 정확도를 나타냈고, F0은 높은 정확도 혹은 보통의 정확도를 나타냈다. 그리고 모든 그룹에서 VOT의 AUC가 F0의 AUC보다 높았는데, 이는 평음과 경음의 분류

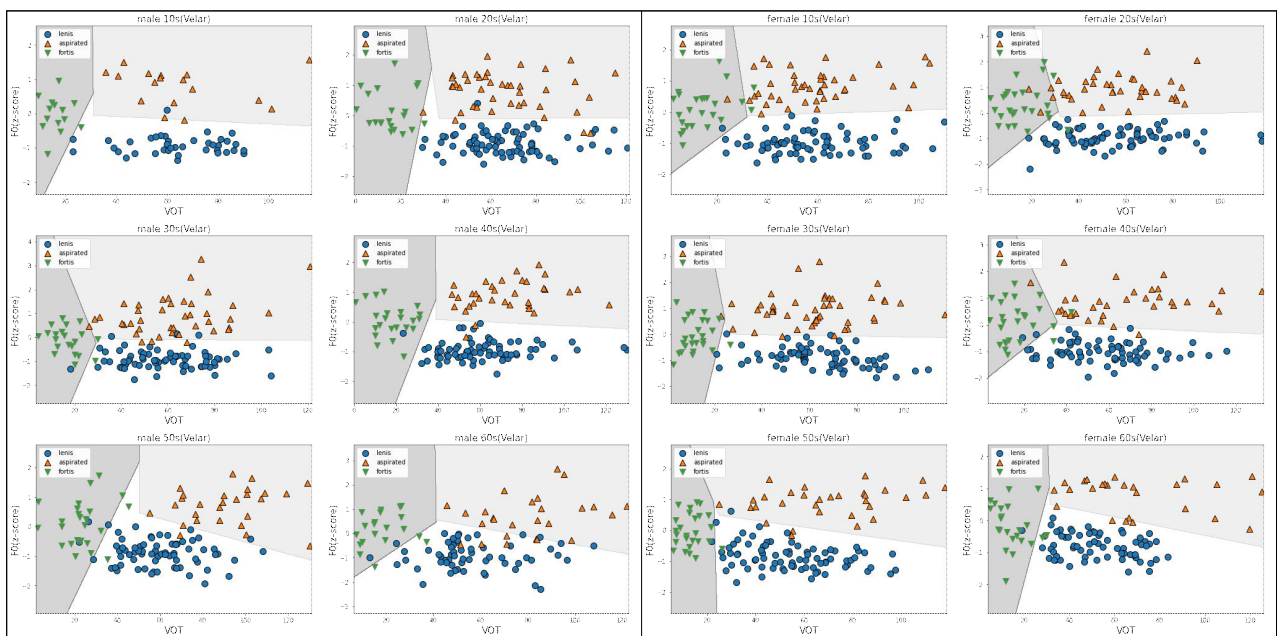


그림 3. SVM 알고리즘을 이용한 어두 평음, 격음, 경음의 경계선

Figure 3. Decision boundary of word-initial lenis, aspirated, fortis stops using SVM (support vector machine) algorithm.

³ Package e1071(Meyer et al., 2021)을 이용한 R에서 선형 SVM을 실행할 때에는 기본 설정을 유지하였다(type: C-classification, kernel: linear, cost: 1). 또한 트레이닝 데이터와 테스트 데이터는 수집된 데이터에서 7:3의 비율로 하여 테스트 데이터에서 학습 효과를 유지하는지 관찰하였다. 이에 따라 R에서 학습한 SVM 모델의 트레이닝 데이터에서는 ‘남성 10대: 94.0%, 20대: 91.5%, 30대: 89.9%, 40대: 94.4%, 50대: 89.5%, 60대: 84.9%, 여성 10대: 92.5%, 20대: 94.5%, 30대: 94.5%, 40대: 90.6%, 50대: 93.7%, 60대: 92.7%’의 정확성을, 테스트 데이터에서는 ‘남성 10대: 95.2%, 20대: 100%, 30대: 97.5%, 40대: 100%, 50대: 97.3%, 60대: 90.9%, 여성 10대: 97.6%, 20대: 100%, 30대: 93.1%, 40대: 92.6%, 50대: 95.2%, 60대: 91.4%’의 정확성을 보였다.

⁴ 다만 50대 여성의 자료에서는 평음의 VOT가 경음의 VOT와 겹치는 경우가 드물어 경계선이 거의 수직으로 나타났다. 이는 표본의 부족일 수 있다고 생각된다.

⁵ 이는 VOT에 의해 평음과 격음이 전혀 분류될 수 없다는 것을 의미하지는 않는다. 예를 들어 남·여 40대에서 평음·격음 경계선의 기울기는 0에 가깝지만, 전반적인 격음의 VOT 분포가 평음의 VOT에 비해 다소 오른쪽에 있음을 볼 수 있다. 이는 표 2에서도 확인한 바 있다.

에서 F0과 VOT가 모두 중요한 역할을 하지만 VOT가 F0보다 더 큰 역할을 한다는 것을 의미한다. 그리고 경음과 격음의 분류에서는 VOT가 모든 그룹에서 높거나 완벽한 정확도를 나타냈고, F0은 60대 남성과 50대 여성을 제외하고는 보통의 정확도를 나타냈다. 경음과 격음의 분류에서도 F0보다 VOT가 중요한 역할을 하는 것을 알 수 있다.

표 4. SVM 모델의 변수 중요도(AUC)
Table 4. Variable importance(AUC) in SVM model

남성(평음 vs 격음)						
독립 변수	10대	20대	30대	40대	50대	60대
VOT	0.570	0.505	0.548	0.663	0.882	0.814
F0	0.997	0.971	0.994	0.997	0.981	0.942
여성(평음 vs 격음)						
독립 변수	10대	20대	30대	40대	50대	60대
VOT	0.504	0.507	0.525	0.601	0.687	0.656
F0	1.000	1.000	0.998	0.998	0.993	0.979
남성(평음 vs 경음)						
독립 변수	10대	20대	30대	40대	50대	60대
VOT	0.996	0.999	0.993	0.957	0.971	0.973
F0	0.923	0.942	0.873	0.940	0.863	0.872
여성(평음 vs 경음)						
독립 변수	10대	20대	30대	40대	50대	60대
VOT	0.988	0.977	0.992	0.969	0.998	0.995
F0	0.952	0.963	0.877	0.918	0.891	0.807
남성(경음 vs 격음)						
독립 변수	10대	20대	30대	40대	50대	60대
VOT	1.000	0.998	0.995	0.963	1.000	1.000
F0	0.870	0.779	0.875	0.853	0.736	0.671
여성(경음 vs 격음)						
독립 변수	10대	20대	30대	40대	50대	60대
VOT	0.995	0.977	0.999	0.986	0.999	0.995
F0	0.825	0.874	0.871	0.777	0.906	0.851

SVM, support vector machine; AUC, area under the curve.

지금까지의 결과를 통해 어두 폐쇄음 발화 양상의 변화를 정리해보면 다음과 같다. 우선 경음의 경우 연령에 따른 VOT 및 F0의 변화가 크지 않아서 발화 양상이 현재 60대부터 10대까지는 거의 변화하지 않았음을 알 수 있다. 즉 경음은 짧은 VOT 및 중간 정도의 F0으로 자신의 경계를 유지하고 있다. 그러나 평음과 격음의 경계는 시간이 지남에 따라 변화했는데, 이는 격음의 VOT가 줄어들었기 때문이다. 다시 말해 평음과 격음에서 VOT와 F0이 모두 다르게 발화되던 시기(50-60대)를 지나 과도기를 거쳐(40대) F0만 다르게 발화되는 시기(30대 이하)로 이행한 것이다.⁶ 이에 남·여 50-60대에서 평음과 격음의 F0이 일부 겹치기도 하는 모습은 30대 이하에서는 발견되지 않는다. 따라서 자

유 발화 양상만을 고려했을 때, 30대 이하의 젊은 세대에서는 평음과 격음의 VOT 융합이 거의 완료된 상태로 보인다.

3. 어두 폐쇄음의 지각

3.1. 연구 방법

3.1.1. 연구 대상 및 실험 참가자

청자는 40명의 정상 청력을 가진 서울·경기 출신 20대 여성(만 20-29세, 평균 24.5세, 표준편차 2.84)으로 하였다. 20대 여성을 청자로 설정한 이유는 VOT 변화가 여성을 주도로 이루어졌으며 20대 여성은 발화의 변화가 가장 완료된 것으로 보이는 그룹 중 하나이기 때문이다. 앞서 언급하였듯 어두 폐쇄음의 발화 양상(특히 평음과 격음)의 변화는 현재 30대 이하 연령층에서 거의 완료된 것으로 보이며 격음의 VOT는 10-20대 여성 그룹에서 가장 짧게 나타났다. 따라서 본 연구에서는 20대 여성을 대상으로 한 지각 실험 결과를 20대 여성의 발화 자료의 분석 결과와 비교하여 발화와 지각에 어떠한 차이가 있는지 살펴보고 나아가 음 변화에 대해 논의하고자 한다.⁷

3.1.2. 청취 자료 제작

평음, 경음, 격음의 지각 분포가 VOT×F0의 2차원 평면에 어떻게 펼쳐지는지 살펴보기 위해, VOT와 F0의 연속체를 음성 자극으로 제작하였다. 음성 자극의 어두 폐쇄음은 과일 시점이 명확하고 VOT가 비교적 길어서 음성 조작이 용이한 연구개음으로 설정하였고, 후행 모음은 Winn(2020)의 VOT 조작 스크립트에서 추천하는 모음에 따라 ‘이’로 설정하였다.

음성 자극을 제작하기 위해서는 NAVER社에서 제공하는 CSS(clova speech synthesis)를 이용하였으며, 이때 CSS에서 제공하는 여러 목소리 중 여성 청년층인 ‘나눔 드림’의 목소리를 선택하였다. 보통의 음성 지각 실험에서는 실제 사람의 목소리를 녹음하여 이를 조작하곤 하는데, 본 연구에서 CSS를 이용한 이유는 다음과 같다. 첫째, 연구자가 듣기에 실제 사람의 목소리와 거의 차이가 없었으며 오히려 더 깔끔한 소리를 제공하여 음성을 조작하기에 용이하다. 한 예로 CSS로 만들어진 ‘기’의 스펙트로그램을 보이면 그림 4와 같다. VOT 구간, 모음 시작 부분, 포먼트가 명확히 보이며 F1과 F2는 각각 381.6 Hz, 2,986.3 Hz로, Moon(2007)에서 제시한 서울·경기 20대 여성 화자의 평균 ‘이’ 모음 F1(414 Hz, 표준편차 81), F2(2,953 Hz, 표준편차 153)와 다르지 않다.

6 물론 더 이전의 시기로 올라가면 또 다른데, Byun(2016), Kim(2015)에 따르면 1950년대 이전 출생자들의 평음과 격음은 F0으로 전혀 분류가 안 되었고, VOT로만 분류되었다.

7 가장 정확한 실험은 음성 자료를 제공한 발화자를 그대로 지각 실험에 참여시키는 것이다. 그러나 본 연구에서 활용한 말뭉치의 발화자가 누구인지 알 수 없으므로 새로 청자를 모집하였다. 20대 여성의 발화 자료를 Kruskal-Wallis 검정을 통해 분석한 결과 15명의 화자의 평음, 경음 VOT가 다르지 않았으며(평음: $df=14$, $p=0.2078$ / 경음: $df=14$, $p=0.1963$), 1명을 제외한 14명의 화자의 격음 VOT가 다르지 않았다($df=13$, $p=0.103$). 따라서 이들 화자 15명의 어두 폐쇄음 VOT는 서울·경기 20대 여성의 발화를 대표한다고 볼 수 있을 것이다. 그리고 청자는 40명이라는 충분한 크기의 표본이므로 역시 서울·경기 20대 여성을 대표한다고 할 만하다.

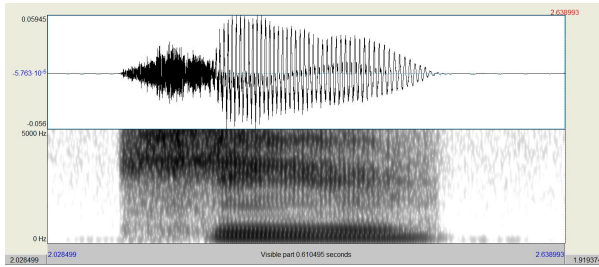


그림 4. CSS로 합성된 ‘기’의 스펙트로그램
Figure 4. Spectrogram of ‘gi’ synthesized by CSS

둘째, CSS와 같은 딥러닝 방식의 음성 합성은 수많은 데이터를 기반으로 하므로 제공되는 음성은 이미 여러 화자들로부터 표준화된 것이라고 할 수 있다. 이러한 이유들로 본 연구에서는 CSS에서 제공되는 ‘기’ 음성을 44.1 kHz의 표본 추출률, 16 bit 양자화로 컴퓨터 내부에서 녹음하였으며, 이를 조작하였다.

VOT 값은 20대 여성 발화 자료에서 /ㄱ/의 VOT(평균 16.22 ms, 표준편차 8.77)와 /ㅋ/의 VOT(평균 55.16 ms, 표준편차 18.99)를 고려하여 10–90 ms의 9단계로 조작하였다. F0 값은 CSS ‘기’의 F0(227.7 Hz)에서 반음(semitone) 단위로 조작하여 20대 여성 /ㅋ/의 F0 평균 306.97보다 높은 321.4 Hz까지 도달하게 하였다. 이런 방식으로 F0은 7단계로 조작되었다. VOT 조작은 기본적으로 Winn(2020)에 따른 Praat 스크립트를 이용하고 미세한 부분은 Manipulation 기능으로 수정하였으며, F0의 반음 조작은 Praat의 Manipulation 기능을 이용하였다. 이렇게 만들어진 구체적인 음성 자극의 목록은 표 5의 63개이다.

표 5. 지각 실험에 사용된 음성 자극
Table 5. Speech stimuli used for perception experiment

F0 (Hz)	VOT (ms)
227.7 (0 semitone)	10, 20, 30, 40, 50, 60, 70, 80, 90
241.4 (+1 semitone)	10, 20, 30, 40, 50, 60, 70, 80, 90
255.8 (+2 semitone)	10, 20, 30, 40, 50, 60, 70, 80, 90
270.8 (+3 semitone)	10, 20, 30, 40, 50, 60, 70, 80, 90
286.7 (+4 semitone)	10, 20, 30, 40, 50, 60, 70, 80, 90
303.5 (+5 semitone)	10, 20, 30, 40, 50, 60, 70, 80, 90
321.4 (+6 semitone)	10, 20, 30, 40, 50, 60, 70, 80, 90

3.1.3. 지각 실험

지각 실험으로는 식별 과제가 사용되었다. 이 과제에서 피험자는 헤드셋을 통해 제시되는 음성을 듣고 ‘기’, ‘키’, ‘끼’ 중 어느 것으로 들리는지 판단하였다. 이때 음성 자극들은 “이것은 ‘___’라고 들립니다.”라는 문장 안에서 함께 제시되었다.

지각 실험에는 Praat에서 제공하는 청지각 실험용 스크립트

인 ExperimentMFC를 이용하였으며 음성은 2번씩 랜덤화되어 총 126번(표 5의 63개 자극×2번씩) 제시되었다. 또한, 126개의 음성이 끊임없이 제시되면 피험자가 피로할 수 있기 때문에 10 개씩 끊어갈 수 있도록 스크립트를 작성하였다. 다시 듣기는 한 번까지만 가능하도록 설정하였다. 이 실험은 대략 7–8분이 소요되었으며 대면 및 비대면으로 실험되었다. 대면 실험에서는 Britz 社の BR-H850 헤드셋을 이용하였다.⁸

3.2. 지각 양상 분석 결과

어두 폐쇄음 /ㄱ/, /ㅋ/, /ㄱ/의 지각 양상을 시각화하여 표현하면 다음 그림 5와 같다. 원의 크기가 클수록, 색이 짙을수록 응답 비율이 100%에 가까운 것이다. 그리고 50%가 넘는 부분을 선으로 묶어 표시하였다.

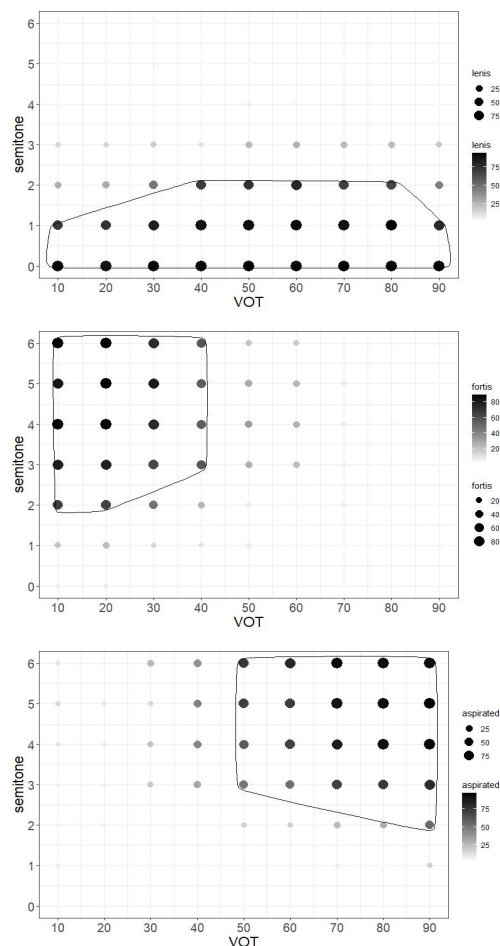


그림 5. 어두 폐쇄음 /ㄱ/ (위), /ㄱ/ (중간), /ㄱ/ (아래)의 지각 양상
Figure 5. Perception aspect of word-initial stops /ㄱ/ (above), /ㄱ/ (middle), /ㄱ/ (below)

평음과 경음을 먼저 비교해 보자. 경음으로 지각되는 음성 자

⁸ 비대면 실험은 zoom을 통해 연구자의 컴퓨터를 원격 제어할 수 있는 기능을 활용하여 대면 실험과 동일하게 Praat ExperimentMFC로 시행되었다. 또한 피험자는 본인이 가지고 있는 헤드셋을 꼭 착용하여 실험에 참여할 것을 요구받았으며, 헤드셋의 소리 크기는 중간에 맞추도록 하였다. 실험은 연구자의 컴퓨터를 제어하여 진행되었으므로 컴퓨터에서 출력되는 소리는 연구자가 제어하였다.

극은 주로 VOT 40 ms 이하에 분포되어 있다. 그런데 VOT가 짧다고 하여 항상 경음으로 지각되는 것은 아닌데, 0+1 semitone에서는 경음이 아닌 평음으로 지각되기 때문이다. 이는 경음의 지각에 있어서 VOT뿐만 아니라 F0 역시 영향을 미친다는 것을 보여준다. 즉 아무리 VOT가 짧아도 F0이 높지 않으면 경음으로 지각될 수 없다.

다음으로 경음과 격음을 비교해 보자. 경음과 격음은 대체로 +2+3 semitone 이상에 위치한다. F0에서는 차이를 보이지 않고, VOT 경계가 40-50 ms 사이에서 형성되는 것을 확인할 수 있다. 이는 경음과 격음의 지각 변별에서 F0은 영향을 미치지 못하며, F0이 일정 값 이상만 되면 VOT에 따라 경음과 격음이 변별된다는 것을 보여준다.

마지막으로 평음과 격음의 경계이다. 격음은 주로 50 ms 이상, +3 semitone 이상에 분포한다. VOT 50 ms 이상에서는 같은 VOT라고 하더라도 F0이 낮으면 평음, 높으면 격음으로 지각된다. 이러한 점은 20대 여성의 발화 양상과도 크게 다르지 않다. 그러나 주목할 점은 VOT 90 ms의 +2 semitone 부분이다. 이 부분에서는 평음(43.75%)보다는 격음(53.75%)으로 지각되는 비율이 높았는데, 같은 F0인 +2 semitone의 VOT 60-80 ms에서는 평음의 응답 비율이 격음의 응답 비율보다 높았다는 점에서 차이를 보인다. 즉 이러한 모습은 F0이 변화하지 않아도 VOT가 특정한 값 이상으로 길어지면 평음이 아닌 격음으로 지각될 수 있다는 것을 보여준다.

4. 어두 폐쇄음 발화와 지각의 비교

여기에서는 언어 변화 주도 계층인 20대 여성의 발화와 지각을 구체적으로 비교하고자 한다. 다음 그림 6의 왼쪽은 F0과 VOT에 따른 평음, 격음, 경음의 실제 응답 비율이며 오른쪽은 20대 여성의 발화 /ㄱ/, /ㅋ/, /ㄲ/를 트레이닝 데이터로 하여 머신러닝 SVM으로 학습한 후 표 5의 음성 자극을 예측한 비율이다⁹. 그래프의 자세한 정보는 부록에 제시하였다.

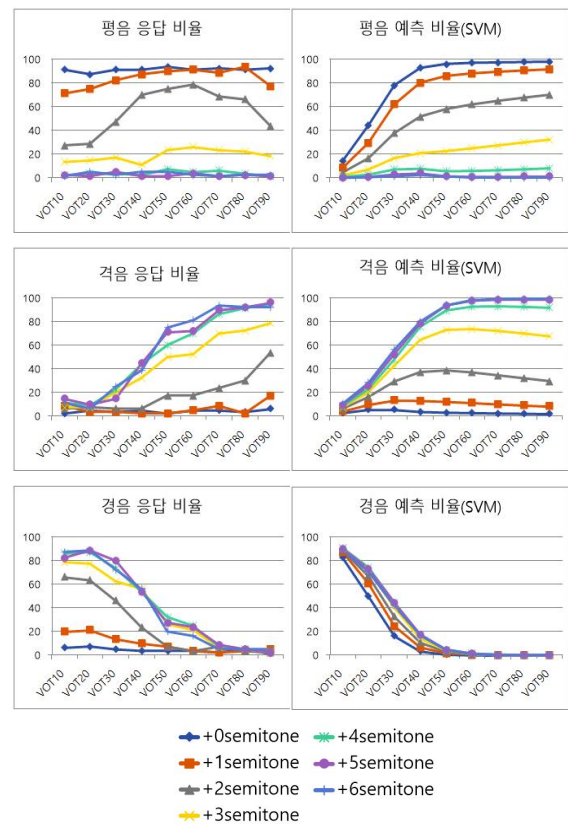


그림 6. 실험을 통한 실제 응답 비율과 SVM으로 예측한 비율
Figure 6. Response percentage based on perception experiment and prediction percentage based on SVM(support vector machine)

그림 6을 보면, 실제 응답과 발화를 기반으로 한 예측에서 차이가 가장 크게 나타나는 경우는 0+1 semitone의 그래프들이다. 0+1 semitone은 VOT와 관계없이 최소 71.25% 이상 평음으로 지각되었으며 경음으로 지각되는 경우는 최고 21.25%에 불과하였다. 그러나 발화를 기반으로 예측한 SVM 예측 비율을 보면 0+1 semitone이 VOT가 짧을 때는 경음으로 분류되고 길어질수록 평음으로 분류될 확률이 높은 모습을 보여준다. 사실 VOT 20 ms 이하의 0 semitone과 +1 semitone의 조합은 실제 발화될 일이 거의 없는 영역이기도 한데, 특히나 연구개 폐쇄음은 치조 폐쇄음이나 양순 폐쇄음보다 VOT가 길기 때문에 어두에서 VOT 20 ms 이하의 0+1 semitone으로 발화될 일은 흔하지 않다. 그럼에도 청자들은 VOT 20 ms 이하의 0+1 semitone을 /ㄲ/가 아닌 /ㄱ/로 지각하였고, 이는 평음의 지각에 낮은 F0이 무엇보다 중요하며 1차적 역할을 한다는 것을 보여준다. 경음뿐만 아니라 격음의 응답 비율을 보면 0+1 semitone은 VOT가 길어져도 격음으로 거의 지각되지 않는데, 이 역시 낮은 F0이 평음으로 지각되는 데에 결정적인 역할을 한다는 것을 보여준다.

9 앞서 만들어진 SVM 모델의 트레이닝 데이터가 표준화된 데이터였으므로, 예측할 때에도 표준화된 데이터를 입력해야 한다. 따라서 발화 데이터의 표준화 방법과 똑같이 CSS의 /ㄱ/, /ㅋ/, /ㄲ/ 뒤 다양한 모음의 F0을 측정하여 표준화를 진행하였다. 그 결과 0 semitone부터 +6 semitone까지 각각 -0.749, -0.469, -0.175, 0.130, 0.455, 0.798, 1.164의 z-score가 도출되었으며 이를 모델에 입력하였다.

다음으로 +2 semitone 그래프 역시 지각과 발화의 중요한 차이를 나타낸다. 평음 응답 비율과 평음 예측 비율 그래프 모두에서 +2 semitone의 꺾은선을 보면 VOT 60 ms까지는 점차 상승하였지만, 70 ms부터 서로 다른 모습을 보여준다. 평음 응답 비율에서는 다시 하강하는 모습을 보여주며 심지어 90 ms에서는 평음의 응답 비율이 50% 미만으로 떨어졌다. 한편 평음 예측 비율 그래프를 보면 끝까지 상승하여 90 ms에서 70.2%의 비율로 평음으로 예측되었다. 또한 +2 semitone의 격음 응답 비율은 90 ms에서 80 ms에서보다 23.75% 포인트 상승하여 53.75%의 응답 비율을 보였다. 그러나 +2 semitone, 90 ms의 격음 예측 비율은 29.7%로 실제 응답 비율과 크게 차이난다. 이러한 모습이 나타나는 이유는 발화와 달리 지각에서 VOT가 평음과 격음 변별에 기능하였기 때문이다. 즉 발화와 달리 지각에서는 평음과 격음의 지각이 VOT만으로 달라질 수 있다는 것이다. 표 2와 그림 1에서 보았듯이 20대 여성 발화에서는 /ㄱ/와 /ㅋ/에서 VOT 차이가 나지 않았지만, VOT 30 ms 이상의 음성 자극만을 대상으로 t-검정한 결과 지각에서는 ‘기’로 응답한 음성 자극의 평균 VOT와 ‘키’로 응답한 음성 자극의 평균 VOT가 유의미하게 달랐음이 밝혀졌다($df=3,193, t=11.379, p<.001$).¹⁰

+3~+6 semitone은 평음으로는 거의 지각되지 않고, 경음 혹은 격음으로만 지각되는데, 특히 VOT 30~50 ms 사이에서 그래프의 기울기가 가파르게 나타나고 있다. 이는 SVM 알고리즘으로 예측한 그래프의 모양과도 크게 다르지 않다. 또한 +4~+6 semitone의 그래프들은 거의 겹쳐 있으므로, 발화에서 격음의 F0이 경음의 F0에 비해 유의미하게 높았던 것과는 대조적으로 (표 3, 그림 2), 지각에서는 F0이 격음과 경음의 변별에 거의 영향을 미치지 않는다는 것을 보여준다. 한편 +3 semitone의 경우 +4~+6 semitone과 달리 격음의 응답 비율이 다소 낮는데, 이는 +3 semitone의 F0이 +4~+6 semitone에 비해 낮아서 평음으로 지각되기도 했기 때문이다.

지금까지의 서울·경기 20대 여성의 발화와 지각 차이를 정리하면 다음과 같다. 발화의 경우 VOT는 ‘격음=평음>경음’, F0은 ‘격음>경음>평음’으로 나타나지만, 지각의 경우 VOT는 ‘격음>평음>경음’, F0은 ‘격음=경음>평음’으로 나타난다. 또한, 지각 변별에서 1차적 단서로 F0이 기능하여 낮으면 평음, 높으면 경음 혹은 격음으로 지각하고, 2차적 단서로 VOT가 기능하여 길면 격음 짧으면 경음으로 지각한다는 것을 알 수 있었다. 그러나 일부 F0에서는 여전히 VOT가 평음과 격음의 지각 변별에서 기능을 하기도 하는데, 이는 +2 semitone에서 VOT 70 ms부터 평음의 응답 비율이 낮아지고 격음의 응답 비율이 조금씩 높아지다가 VOT 90 ms에서는 격음의 응답 비율이 평음의 응답 비율을 넘어서는 것에서 확인할 수 있었다. 또한 +1 semitone에서도 VOT 90 ms에서는 VOT 80 ms에 비해 평음의 응답 비율이

16.25% 포인트 떨어졌으며, 격음의 응답 비율은 15% 포인트가 올랐다. 발화와 달리 지각에서는 VOT가 아직 평음과 격음을 변별하는 요소로 작게나마 기능하고 있는 것이다.

5. 논의 및 결론

지금까지 2020년에 수집된 자유 발화 자료를 바탕으로 하여 10~60대 남·여의 발성 유형에 따른 어두 폐쇄음 발화가 어떻게 다른지, 그리고 언어 변화의 주도 계층 중 하나인 20대 여성을 중심으로 하여 발화와 지각에서 어떠한 차이가 나타나는지를 음성 분석과 지각 실험을 통해 자세히 살펴보았다.

자유 발화 환경에서 VOT는 모든 집단에서 경음이 평음, 격음에 비해 훨씬 짧았다. 그러나 격음과 평음의 VOT 차이는 세대와 성별에 따라 달랐는데, 30대 이하의 남·여 모두에서 격음과 평음의 VOT는 유의미한 차이를 보이지 않았고, 40대 이상에서는 나이가 많을수록, 여성보다는 남성에서 평음과 격음의 VOT가 유의미한 차이를 보였다. 그리고 F0은 모든 집단에서 격음>경음>평음의 순으로 나타났다.

그런데 30대 이하 남성에서 평음과 격음의 VOT 차이가 유의미하지 않다는 결과는 지금까지의 연구와 차이를 보이기도 한다. Byun(2016)에서는 1950년대 출생 남성부터 1990년대 출생 남성(현재 20~60대 남성) 모두에서 격음>평음>경음의 순서로 VOT가 유의하게 길었으며, Oh(2011)에서는 18~27세 남성(현재 20~30대 남성)의 /p/와 /pʰ/, /k/와 /kʰ/의 VOT 겹침이 50% 미만으로 나타났다. 또한 Kang(2014)에서도 1930년대 출생 남성부터 1980년대 출생 남성(현재 30~80대 남성)에서 모두 격음의 VOT가 평음의 VOT보다 유의미하게 길었다. 그러나 이는 실험 환경에서 유발된 차이일 수 있는데, Byun(2016), Oh(2011)에서는 통제된 상황에서 주어진 단어 혹은 문장을 읽도록 요구하였으며, Kang(2014)에서는 서울말 낭독체 발화 말뭉치를 이용하였기 때문이다. 기존 연구들에서 젊은 남성 세대가 평음과 격음의 유의미한 VOT 차이를 보여준 것은 통제된 실험이라는 상황 속에서 평음과 격음의 VOT를 다소 보수적인 방향으로 발화했기 때문이었을 가능성이 있다고 생각된다.

그러나 남성과 달리 현재 30대 이하의 여성 발화(1980년대 이후 출생)에서 평음과 격음의 VOT 차이가 유의미하지 않다는 사실은 Byun(2016), Jeong(2010), Kang(2014) 등 여러 논의에서 언급된 것과 동일하다. Byun(2016)에서는 주의 깊은 발화의 VOT를 따로 살펴보기도 하였는데, 심지어 1990년대 출생 여성(현재 20대 여성)은 격음의 VOT 평균이 평음의 VOT 평균보다 5 ms 정도 짧은 모습을 보여주기도 하였다. 따라서 젊은 여성의 경우 발화 환경과 관계없이 평음과 격음의 VOT 합류가 보이며, 이에 젊은 여성 세대의 발화에서는 평음과 격음의 VOT 융합이 거의

10 +2 semitone의 경우 다른 F0들과 달리 유일하게 VOT에 따라 평음, 격음, 경음이 50% 이상으로 지각된 구간이 모두 존재하였는데, VOT 10~20 ms에서는 경음으로, 40~80 ms에서는 평음으로, 90 ms에서는 격음으로 지각된 비율이 모두 50% 이상이었다. 이를 통해 +2 semitone이 평음과 격음의 경계에 있는 F0이라는 사실과, 이처럼 애매한 F0에서는 청자들이 VOT를 사용하여 어두 폐쇄음을 변별한다는 것을 알 수 있다.

완전하게 이루어졌다고 할 수 있다. 젊은 남성 세대의 경우 자유 발화에서만 이러한 현상이 관찰되므로 여성에 비해 덜 완전한 VOT 합류가 일어났다고 보인다.

지각 실험으로는 20대 여성을 대상으로 하였다. 그 결과 낮은 F0에서는 VOT와 관계없이 항상 평음으로 우세하게 지각되었으며, 높은 F0에서는 VOT가 짧으면 경음, VOT가 길면 격음으로 지각되었다. 이는 굉장히 흥미로운 결과로, ‘VOT가 짧다면 경음으로 지각된다.’라고 단순히 말할 수 없음을 보여준다. 20대 남성을 대상으로 지각 실험을 수행한 Kong & Lee(2017)에서는 VOT가 9-13 ms일 때 경음으로 지각되는 비율이 75%를 넘는다고 하면서 VOT가 경음의 지각에 있어 가장 중요한 요소라고 언급하였다. 그러나 본 연구에서는 F0이 낮으면 VOT가 아무리 짧아도 경음으로 지각되기는 힘들고 오히려 평음으로 지각된다는 것이 드러났다. 따라서 VOT보다 F0이 먼저 평음과 평음이 아닌 것을 변별하는 역할을 한다고 볼 수 있다. 이는 Jang & Shin(2010)의 결론과도 일치하는 것으로, 20대 서울 출신 여성은 1차적으로 F0을 가지고 평음과 나머지를 구별하고, 그 다음으로 VOT를 가지고 경음과 격음을 구별하였다.

그런데 VOT가 평음과 격음의 지각 변별에 아무런 단서로 작용하지 않는 것은 아니다. 같은 F0일 때 VOT에 따라 평음과 격음이 변별되기도 했기 때문이다. 이는 발화에서 평음과 격음의 VOT가 합류된 것과는 달리 지각에서는 여전히 VOT가 지각 요소로 사용되고 있다는 것을 보여준다. Kim et al.(2020)에서도 청년 여성 집단에서 평음과 격음의 지각 변별에 있어 F0의 변별력이 크긴 하지만 VOT 역시 낮은 변별력을 보여준다고 하였다. 즉 현재 여성 20대의 말하는 방식과 지각하는 방식에는 차이가 있는 것이다.

이러한 상황은 지각에서의 음 변화가 발화에서의 음 변화보다 뒤쳐지고 있는 것을 보여주는데, Pinget et al.(2020)에 따르면 이는 음 변화의 후반부에서 나타나는 현상이다. Pinget et al.(2020)에서는 네덜란드어의 /v/와 /b/가 각각 무성음화가 되어가는 변화를 관찰하였는데, 거의 완성되어가는 변화인 /v/의 무성음화는 발화의 변화가 지각의 변화에 선행하였고, 이제 막 변화의 시작 단계에 있는 /b/의 무성음화는 지각의 변화가 발화의 변화에 선행하였다. 지각의 변화가 발화의 변화에 뒤쳐지는 현상은 Janson(1983)에서도 발견된다. 스웨덴 스톡홀름의 모음 /a:/는 음 변화를 겪어 다른 방언에 비해 훨씬 후설화되어 /o:/와 가까운 위치에 있는데, Janson(1983)에서 스톡홀름의 젊은 세대를 대상으로 발화와 지각을 비교한 결과 발화가 지각에 비해 /a:/가 후설화되어 있는 것을 확인하였다. 즉 /a:/로 의도하여 발화한 음을 들려주면 /o:/로 지각한다는 것이다.

따라서 지각의 변화는 발화의 변화보다 늦게 완성되며 예전의 체계를 유지하고 있는 것인데, 이는 어떻게 보면 자연스러운 결과로 이해된다. 점진적으로 일어나는 음 변화에서 젊은 세대는 위 세대보다 한발 나아간 방식의 발화를 보이지만, 태어났을 때에는 위 세대로부터 언어를 학습했을 것이고 위 세대와 함께 소통해야 하기 때문에 지각은 예전의 체계를 함께 지니고 있어야 한다. Pinget et al.(2020)에서는 지각 체계가 예전의 방식을 유

지하고 있는 것은 이전의 방식으로 발화하는 집단이 아직 존재하기 때문이라고 해석하였으며, Janson(1983)에서도 자식 세대에서 지각의 변화가 발화의 변화에 후행하는 것은 위 세대의 소리를 잘 인지하기 위함이며 이러한 지각 체계는 이미 어린아이였을 때 형성된 것으로 보고 있다.

지각과 발화의 관계는 언어 형성과도 크게 관련되어 있다. 만약 아이들의 발화가 부모가 발화하는 것과 완전히 일치한다면 언어 변화는 일어날 수 없기 때문이다. 따라서 아이들이 형성하고 사용하는 언어는 부모의 언어와는 조금 다르며, 그렇기 때문에 언어 변화가 조금씩 생긴다. Labov(2001:447-448)에서는 아이들(특히 여자아이들)이 처음 부모로부터 습득한 언어를 재구성함으로써 언어 변화를 증가시킨다고 하였는데, 이러한 언어 변화의 증가는 부모의 언어적 지배를 벗어난 4-5세부터 언어 체계가 안정되는 시기인 17-20세 사이에 일어난다고 보았다.

따라서 본 연구에서 살펴본 자유 발화의 자식 세대라고 할 수 있는 10-30대에서 발화 시에는 평음과 격음의 변별에 VOT 정보를 거의 사용하지 않음에도, (20대 여성을 대표로 살펴봤을 때) 지각 시에는 VOT 정보에 따라 평음과 격음의 변별이 달라지는 이유는 부모 세대의 발화에 VOT에 의한 변별이 남아 있기 때문이다. 이렇게 발화와 지각이 일치하지 않는 사실은 음 변화가 완전히 종료된 것이 아니라, 여전히 진행 중인 것으로 해석되며, 발화에서의 음 변화가 지각에서의 음 변화보다 선행하는 상황이므로 그 진행 단계는 후반부라고 할 수 있겠다. 또한 가장 낮은 F0인 0 semitone에서는 VOT가 격음 지각에 전혀 기능하지 못하였으므로 순차적으로 발화와 지각의 일치, 즉 평음과 격음 VOT의 완전한 융합이 이루어지고 있다는 것을 알 수 있다.

남겨진 과제는 현재 30대인 부모의 자식 세대라고 할 수 있는 10세 이하 아동들의 지각 양상을 살펴보는 것이다. 2-4세 아동의 폐쇄음 지각에 관하여 살펴본 Son(2020)에서는 ‘평음-격음’의 최소대립쌍에서 VOT가 유지된 채 F0이 높아질수록 격음으로 지각되는 경향이 유의미하다고 밝혔으나, VOT에 따른 지각 양상에 관한 실험은 실행하지 않았다. 부모 세대인 30대 여성은 평음과 격음의 VOT를 같게 발화하고 있기 때문에 자식 세대에서는 VOT가 평음과 격음의 지각 변별에 개입하지 않을 가능성이 있으며, 이에 대한 연구가 추가적으로 필요하다. 만약 이것이 사실이라면 현대 서울 방언에서 평음과 격음의 VOT 융합은 현재 아동에서 완료되었다고 할 수 있을 것이다. 따라서 이에 관해서는 면밀한 연구가 필요하며 후고를 기약하기로 하겠다.

References

- Bang, H. Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T. J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics*, 66, 120-144.
- Boersma, P., & Weenink, D. (2021). Praat: Doing phonetics by computer (version 6.1.41) [Computer program]. Retrieved from <http://www.praat.org/>

- Byun, H. G. (2016). F0 as a primary cue for signaling word-initial stops of Seoul Korean. *Phonetics and Speech Sciences*, 8(1), 25-36.
- Fischer J. E., Bachmann L. M., & Jaeschke R. (2003). A readers' guide to the interpretation of diagnostic test properties: Clinical example of sepsis. *Intensive Care Med*, 29(7), 1043-1051.
- Jang, H., & Shin, J. (2010). A dialectal differences in the perception of phonation types in word-initial stops: Based on comparison between Seoul and Daegu dialect. *Korean Linguistics*, 49, 369-388.
- Janson, T. (1983). Sound change in perception and production. *Language*, 59(1), 18-34.
- Jeong, Y. (2010). Speech perception of Korean plain, aspirated, and tense considering pitch of the following vowel. *Urimol*, 27, 73-94.
- Jun, S. A. (1993). *The phonetics and phonology of Korean prosody* (Doctoral dissertation). Ohio State University, Columbus, OH.
- Kang, K. H. (2010). Generational differences in the perception of Korean stops. *Phonetics and Speech Sciences*, 2(3), 3-10.
- Kang, K. H., & Guion, S. G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America*, 124(6), 3909-3917.
- Kang, Y. (2014). Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics*, 45, 76-90.
- Kong, E. J., & Lee, H. (2017). Attentional modulation and individual differences in explaining the changing role of fundamental frequency in Korean laryngeal stop perception. *Language and Speech*, 61(3), 384-408.
- Kim, J., Kim, S., Ahn, J., Nam, K., & Choi, J. (2020). Perception of lenis and aspirated stops in Seoul Korean by younger and older male and female listeners. *Phonetics and Speech Sciences*, 12(4), 1-8.
- Kim, J. E. (2015). Reinterpretation of stop production in Korean elderly speakers. *Phonetics and Speech Sciences*, 7(2), 139-145.
- Kim, M. R. (2001). Consonantal and vocalic effects in Korean stop identification. *Speech Sciences*, 8(1), 93-111.
- Labov, W. (2001). *Principles of linguistic change* (Vol. 2). Malden, MA: Wiley-Blackwell.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C. C., Lin, C. C. (2021). e1071: Misc functions of the Department of Statistics, Probability Theory Group. R package version 1.7-7. Retrieved from <https://cran.r-project.org/web/packages/e1071/index.html>
- Moon, S. J. (2007). A fundamental phonetic investigation of Korean monophthongs. *Malsori*, 62, 1-17.
- National Institute of Korean Language. (2021). National Institute of Korean language daily conversation corpus 2020 (version 1.0). Retrieved from <https://corpus.korean.go.kr>
- Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics*, 39(1), 59-67.
- Ohala, J. J. (1993). Sound change as nature's speech perception experiment. *Speech Communication*, 13, 155-161.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., & Thirion, B. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Pinget, A. F., Kager, R., & Van de Velde, H. (2020). Linking variation in perception and production in sound change: Evidence from Dutch obstruent devoicing. *Language and Speech*, 63(3), 660-685.
- R Core Team (2021). R: A language and environment for statistical computing (version 4.0.5) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, 23(2), 287-308.
- Son, G. (2020). Korean-speaking children's perceptual development in multidimensional acoustic space. *Journal of Child Language*, 47, 579-599.
- Winn, M. B. (2020). Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *The Journal of the Acoustical Society of America*, 147, 852-866.

• 김진우 (Jin-Woo Kim) 교신저자
고려대학교 국어교육학과 박사과정
서울시 중구 황학동 31
Tel: 010-2905-1269
Email: iks2297@korea.ac.kr
관심분야: 음성학, 음운론, 국어사

표 A1. 어두 폐쇄음의 응답 비율 및 SVM 예측 비율

Table A1. Response percentage of word-initial stops and prediction percentage from SVM model

(단위: %)

평음 응답 비율	VOT 10	VOT 20	VOT 30	VOT 40	VOT 50	VOT 60	VOT 70	VOT 80	VOT 90
0 semitone	91.25	87.5	91.25	91.25	93.75	91.25	92.5	91.25	92.5
+1 semitone	71.25	75	82.5	87.5	90	91.25	88.75	93.75	77.5
+2 semitone	27.5	28.75	47.5	70	75	78.75	68.75	66.25	43.75
+3 semitone	13.75	15	17.5	11.25	23.75	26.25	23.75	22.5	18.75
+4 semitone	2.5	3.75	3.75	2.5	7.5	5	6.25	3.75	1.25
+5 semitone	2.5	1.25	5	1.25	1.25	3.75	1.25	2.5	1.25
+6 semitone	1.25	5	2.5	5	5	2.5	1.25	2.5	2.5
평음 예측 비율	VOT 10	VOT 20	VOT 30	VOT 40	VOT 50	VOT 60	VOT 70	VOT 80	VOT 90
0 semitone	14.242	44.315	78.137	92.715	96.071	96.977	97.364	97.659	97.913
+1 semitone	8.477	29.687	62.112	80.430	86.084	88.277	89.627	90.704	91.646
+2 semitone	4.653	16.626	37.815	51.838	58.168	62.060	65.072	67.744	70.241
+3 semitone	1.890	6.961	16.581	20.826	22.506	24.937	27.359	29.784	32.277
+4 semitone	0.709	2.736	7.067	7.702	5.620	5.803	6.469	7.219	8.037
+5 semitone	0.288	1.141	3.174	4.110	1.737	1.086	1.144	1.280	1.455
+6 semitone	0.120	0.478	1.371	2.247	1.164	0.270	0.179	0.192	0.216
격음 응답 비율	VOT 10	VOT 20	VOT 30	VOT 40	VOT 50	VOT 60	VOT 70	VOT 80	VOT 90
0 semitone	2.5	5	3.75	5	2.5	5	5	3.75	6.25
+1 semitone	8.75	3.75	3.75	2.5	2.5	5	8.75	2.5	17.5
+2 semitone	6.25	7.5	6.25	6.25	17.5	17.5	23.75	30	53.75
+3 semitone	7.5	7.5	20	32.5	50	52.5	70	72.5	78.75
+4 semitone	11.25	8.75	22.5	43.75	60	70	86.25	91.25	95
+5 semitone	15	10	15	45	71.25	72.5	90	92.5	96.25
+6 semitone	11.25	6.25	25	38.75	75	81.25	93.75	92.5	92.5
격음 예측 비율	VOT 10	VOT 20	VOT 30	VOT 40	VOT 50	VOT 60	VOT 70	VOT 80	VOT 90
0 semitone	2.845	5.555	5.490	3.869	3.247	2.897	2.589	2.314	2.067
+1 semitone	4.157	9.773	13.422	13.239	12.437	11.369	10.286	9.274	8.348
+2 semitone	6.227	16.298	29.409	37.427	38.823	37.126	34.703	32.188	29.734
+3 semitone	7.476	20.391	42.802	64.691	73.156	73.810	72.281	70.113	67.692
+4 semitone	8.402	22.979	48.294	75.834	89.718	92.914	93.160	92.671	91.928
+5 semitone	9.467	25.589	52.436	78.360	93.659	97.684	98.501	98.615	98.528
+6 semitone	10.720	28.430	56.538	80.480	94.167	98.601	99.505	99.714	99.784
경음 예측 비율	VOT 10	VOT 20	VOT 30	VOT 40	VOT 50	VOT 60	VOT 70	VOT 80	VOT 90
0 semitone	82.913	50.129	16.373	3.417	0.682	0.127	0.047	0.027	0.020
+1 semitone	87.366	60.539	24.466	6.331	1.479	0.354	0.087	0.022	0.006
+2 semitone	89.119	67.076	32.776	10.735	3.009	0.814	0.224	0.067	0.025
+3 semitone	90.634	72.648	40.617	14.483	4.338	1.253	0.360	0.103	0.031
+4 semitone	90.889	74.285	44.640	16.463	4.662	1.283	0.371	0.110	0.035
+5 semitone	90.246	73.270	44.391	17.531	4.603	1.230	0.355	0.105	0.017
+6 semitone	89.160	71.091	42.092	17.273	4.668	1.129	0.316	0.093	0.000
경음 응답 비율	VOT 10	VOT 20	VOT 30	VOT 40	VOT 50	VOT 60	VOT 70	VOT 80	VOT 90
0 semitone	6.25	7.5	5	3.75	3.75	3.75	2.5	5	1.25
+1 semitone	20	21.25	13.75	10	7.5	3.75	2.5	3.75	5
+2 semitone	66.25	63.75	46.25	23.75	7.5	3.75	7.5	3.75	2.5
+3 semitone	78.75	77.5	62.5	56.25	26.25	21.25	6.25	5	2.5
+4 semitone	86.25	87.5	73.75	53.75	32.5	25	7.5	5	3.75
+5 semitone	82.5	88.75	80	53.75	27.5	23.75	8.75	5	2.5
+6 semitone	87.5	88.75	72.5	56.25	20	16.25	5	5	5

SVM, support vector machine; VOT, voice onset time.

음 변화 관점에서 바라본 한국어 어두 폐쇄음의 발화 및 지각

김 진 우

고려대학교 국어교육학과

국문초록

본고에서는 2020년에 수집된 자유 발화 자료를 바탕으로 어두 폐쇄음 평음, 격음, 경음의 발화가 어떻게 달라지고 있는지 그리고 지각은 어떻게 이루어지고 있는지에 대해 자세히 살펴보았다. 기존 연구의 통제된 실험에서와 다르게, 자유 발화에서는 30대 남성까지도 VOT(voice onset time)에 의해 평음과 격음의 변별이 이루어지지 않았다. 지각 실험은 언어 변화의 주도 계층인 젊은 여성 세대를 대상으로 하였다. 그 결과 어두 폐쇄음의 지각에서 F0이 평음과 나머지 둘을 변별하는 역할을 하였고, 그 다음 VOT가 격음과 경음을 변별하는 역할을 하였다. 특히 F0이 낮을 때에는 VOT의 길이가 짧은 길든 간에 항상 평음으로 지각되었다는 점은 평음의 지각에서 F0이 절대적인 역할을 한다는 것을 보여주었다. 그러나 어떤 F0의 경우에는 VOT에 따라 평음과 격음이 변별되기도 하였는데, 이는 20대 여성의 발화와 지각 체계가 일치하지 않기 때문이다. 음 변화의 관점에서 발화와 지각 체계가 일치하지 않는다는 것은 음 변화가 진행 중이라는 것을 의미하는데, 특히 20대 여성에서처럼 발화의 변화가 지각의 변화에 앞서는 경우는 음 변화가 후반 단계에 있을 때이다. 이처럼 언어 변화를 주도하는 계층에서 여전히 지각에서 이전의 체계를 유지하고 있는 이유는 그들의 부모 세대의 발화에서 VOT에 의한 평음과 격음의 변별이 이루어졌기 때문이며 이를 습득한 것이라고 볼 수 있다. 즉 다른 집단과의 소통을 위해 여전히 VOT를 지각에 사용하고 있는 것이다.

핵심어: 한국어 폐쇄음, VOT(voice onset time) 융합, 소리의 발화와 지각, 음 변화

참고문헌

- 국립국어원 (2021). *국립국어원 일상 대화 음성 말뭉치 2020*(버전 1.0). Retrieved from <https://corpus.korean.go.kr>
- 김제홍, 김소안, 안주희, 남기춘, 최지연 (2020). 한국어 서울 방언의 평음과 격음 변별 지각에서 연령과 성별에 따른 차이. *말소리와 음성과학*, 12(4), 1-8.
- 김지은 (2015). 노년층 파열음 발음의 재해석. *말소리와 음성과학*, 7(2), 139-145.
- 변희경 (2016). 서울 방언 어두 폐쇄음의 후속모음 F0. *말소리와 음성과학*, 8(1), 25-36.
- 오은진 (2009). 발화 속도에 따른 한국어 폐쇄음의 VOT 값 변화. *말소리와 음성과학*, 1(3), 39-48.
- 장혜진, 신지영 (2010). 어두 폐쇄음의 발성 유형 지각에서 나타나는 방언 간 차이: 서울 방언과 대구 방언의 비교를 바탕으로. *한국어학*, 49, 369-388.
- 정윤자 (2010). 후행모음의 피치를 고려한 한국어 평음, 경음, 격음의 인지 실험. *우리말연구*, 27, 73-94.