ISSN 2005-8063
2017. 12. 31.
Vol.9 No.4
pp. 27-34

말소리와 음성과학

Phonetics and Speech Sciences

한국음성학회지

http://dx.doi.org/10.13064/KSSS.2017.9.4.027

영어와 한국어 자연발화 코퍼스에서의 무성 폐쇄음 개방 파열 스펙트럼 연구

A study on the release burst spectra of the voiceless plosives from the English and Korean spontaneous speech corpus

황 선 미 · 윤 규 철*

Hwang, Sunmi · Yoon, Kyuchul

Abstract

The purpose of this work is to examine the English and Korean voiceless plosives from the Buckeye[15] and Seoul[16] corpus in terms of their static spectral characteristics. The plosives were automatically extracted by a Praat script. In order to estimate the percent correctness in the classification of the plosives, discriminant analyses were performed whose trainings were based on four spectral moments, i.e. the center of gravity, variance, skewness and kurtosis as suggested in [6]. Another set of discriminant analyses were performed based on the spectral tilts. In the last set of analyeses, the spectral moments and tilts were both used in the training. Results showed that the correct classification rate did not exceed around 65% in the best case, which suggested that phonetic cues other than the release burst would be necessary including the dynamic spectral aspects and vowel-onset cues.

Keywords: Seoul corpus, Buckeye corpus, Korean, English, plosives, moments, spectral tilts, discriminant analysis

1. 서론

무성 폐쇄음의 개방 파열(release burst) 스펙트럼이 조음 위치에 따라 다를 수 있다는 것은 널리 알려진 사실이다. 이는 개방 파열 부분이 각 파열음의 조음 상태에 따른 공명 특성에 영향을 받기 때문이다.

개방 파열 스펙트럼과 관련된 주요한 관심사 중의 하나는 이 스펙트럼 차이만으로 각 파열음의 음성학적 차이를 나타낼 수 있느냐 하는 점이다. 가장 오래된 연구 중 하나는 패턴 재생기(pattern playback)를 활용한 연구[1]인데, 다양한 개방 파열 주파수와 모음 포먼트를 조합하여 시행한 청취 실험 결과, 모음의 제 2 포먼트의 영향을 받긴 했으나 개방 파열만으로도 파열음의 구분이 가능하다는 것이 밝혀졌다. 또 다른 연구[2]에 의하면 에너지가 집중되어 있는 영역이, 양순 파열음의 경우 대체로 500 ~ 1500Hz 정도였고, 치경 파열음의 경우 4000Hz 이상이며, 연구개 파열음의 경우는 그 중간인 1500 ~ 4000Hz 정도였다고 한다.

[3]과 [4]에서는 각 파열음의 조음 위치와 스펙트럼 템플릿 사이의 연관성에 주목하였다. 즉, 양순 파열음은 평평하거나 하강하는 스펙트럼, 치경 파열음은 상승하는 스펙트럼, 연구개 파열음은 중간이 솟은 스펙트럼과 연관되어 있다고 주장하고, 이 템플릿을 기반으로 한 후속 연구[5]에서 여섯 화자가 발화한 1800개 자극에 대하여 85%의 정확도로 각 파열음을 분류할 수 있었다고 한다. 이러한 스펙트럼 차이는 개방 파열의 광대역 스펙트로그램에서도 어느 정도 볼 수 있다. 즉, 양순 파열음의 경우에는 에너지가 주로 저주파 영역에 모여 있고, 치경 파열음의 경

* 영남대학교, kyoon@ynu.ac.kr, 교신저자

Received 11 October 2017; Revised 4 December 2017; Accepted 7 December 2017

우에는 고주파 영역에 모여 있으며, 연구개 파열음은 중간 정도의 주파수 영역에 모여 있다고 볼 수 있다.

어두 파열음의 통계적 분류를 시행한 한 연구[6]에서는 스펙트럼을 확률 분포로 가정하고 이들의 운동량(moment) 즉, 중력중심값(center of gravity), 분산(variance), 비대칭도(skewness), 첨예도(kurtosis) 등 네 가지 값을 구하였다. 무성 파열음의 개방 파열 부분 중 초기 40 ms에서 구한 이들 값들을 이용하여 92% 정도의 정확도로 조음 위치를 구분해낼 수 있었다고 한다. 더욱이 남성에게서 구한 값을 기준으로 여성을 분류할 때 94%의 정확도를 보여 성별에 무관함을 주장하였다. 하지만, 개방 파열을 이용한 미국 영어에 대한 조음 위치 분류는 연구 절차에 따라 정확도가 매우 달라 58%[7], 100%[8], 97%[9], 88%[10], [11], 92-98[6] 등 매우 차이가 난다.

영어 이외에는 연구가 매우 적은 편인데, 프랑스어 연구[12]에 따르면 개방 파열 정보를 이용해 조음 위치를 분류하는데 87%의 정확도를 보였다고 한다. 네덜란드어 연구[13]에서는 개방 파열 정보만으로 [k]의 구분은 잘 되었지만, [p]와 [t]는 정확도가 떨어졌다고 한다. 이들은 또한 유성 파열음보다는 무성 파열음의 개방 파열 정보가 네덜란드어의 조음 위치 구별에 효과적이었다고 주장했다. 한국어에 대한 연구는 특히 드문데, 그 중 한 연구[14]는 세 명의 화자를 대상으로 한국어 세 종류의 발성 유형(경음, 격음, 평음)에 따른 치조 파열음과 마찰음, 경구개 파찰음을 틀문장에 넣어 발화한 다음, 개방 파열의 강도와 중력중심값 및 비대칭도를 살펴보았다. 같은 조음 위치에서 발성 유형에 따른 차이를 살펴봤기 때문에 각 요인에 있어서 통계적인 차이는 없었고, 화자별로 약간의 차이가 있었다고 한다. 이처럼 무성 파열음의 개방 파열에 대한 연구는 영어에 비해 상대적으로 다른 언어의 연구가 매우 적은 편이다. 특히 한국어의 연구는 거의 전무한 상황이다.

본 연구에서는 최근에 구축된 자연발화 음성 코퍼스인 벅아이 코퍼스[15]와 서울 코퍼스[16]에서 발견되는 무성 파열음의 개방 파열 부분을 연구 대상으로 설정하여 이전 연구들의 절차를 따라 분석을 수행할 것이다. 특히, 한국어에 대한 연구는 관련 연구가 희박한 현 시점에서 의미 있는 자료가 될 것으로 생각된다. 녹음실에서 인위적으로 단어나 분절음을 틀문장에 넣어서 발화한 것을 대상으로 하는 이전의 연구 결과와 자연발화 음성 코퍼스를 대상으로 한 연구 결과는 정성적으로는 같은 경향을 보일 수도 있지만, 정량적인 면에서 다소 차이가 날 수도 있을 것이다. 구축 시점에서 연구의 목적이 알려져 있지 않고, 틀문장 등 보고 읽는 자료가 없는 상태에서 즉흥적으로 대화하듯이 녹음을 하기 때문에, 코퍼스 연구는 기존의 녹음 대상 연구에 비하여 피실험자의 발화에 있어서 그 자연스러움이 극대화되어 있다. 따라서 자연발화 음성 분석은 다른 음성 분석 연구에 비해 해당 언어의 실체에 보다 가깝게 접근할 수 있는 장점이 있다고 생각된다.

2. 연구 방법

2.1. 연구 대상 및 추출 방법

본 연구에서 사용한 자연발화 음성 코퍼스는 영어의 경우 벅아이 코퍼스, 한국어의 경우는 서울 코퍼스이다. 이 두 코퍼스는 모두 주어진 주제에 대하여 두 사람이 인터뷰 방식으로 자유롭게 발화한 것을 녹음한 것으로 단어 및 변이음별로 레이블링이 되어 있다.

이 두 코퍼스로부터 프랏[17] 스크립트를 이용하여 변이음 층으로부터 영어 [p, t, k], 한국어 격음 [p^h, t^h, k^h]에 해당하는 자료와 관련 정보를 추출하였다. 한국어의 경우 평음과 경음도 무성음으로 알려져 있지만, 본 연구에서는 모음 사이에서 유성음화가 이루어지지 않는 격음 분석을 통해 시범적으로 영어의 무성음의 경우와 얼마나 유사한 결과를 얻을 수 있는지 알아보고자 한다.

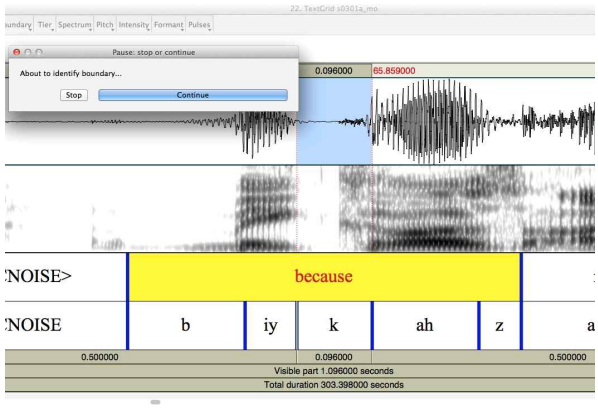

그림 1. 스크립트가 찾아낸 무성 파열음의 예
Figure 1. A sample voiceless plosive identified by a Praat script

특히, 해당 무성 파열음의 개방 파열 시작점을 수작업으로 지정하는 대신, 프랏 스크립트를 작성하여 코퍼스에서 자동으로 시작점을 찾아내도록 하였다. <그림 1>에서 보듯이, 본 연구의 연구 대상인 무성 파열음들은 개방 파열 시작 직전이 폐쇄구간이다.

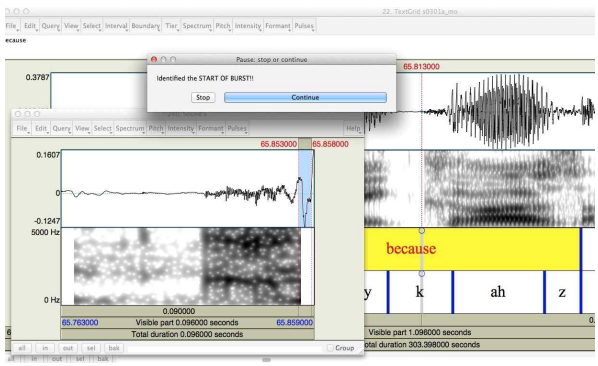

그림 2. 스크립트가 찾아낸 개방 파열의 시작 지점
Figure 2. The start of a release burst identified by a Praat script

이 점을 활용하여, 스크립트로 하여금 파열음으로 레이블링되어 있는 부분을 자동으로 감지하게 하였으며, 파열음 부분 전

체를 폐쇄구간 시작부터 후속 모음 시작 직전까지 5 ms 간격으로 스펙트럼을 추출하였다. 그 다음 스펙트럼의 에너지를 별도의 창에서 순차적으로 비교하여 갑자기 에너지가 증가하여 지정한 값을 초과하는 부분을 개방 파열의 시작점으로 지정하도록 하였다. 예를 들면 스크립트가 찾아낸 [k]의 개방 파열 시작점은 <그림 2>에서 해당 사운드를 관통하는 수직의 커서 위치로 볼 수 있다. 스펙트럼 추출의 간격이 5 ms이므로 사람의 측정치와는 최대 5 ms의 오차가 생길 수 있음을 의미한다.

2.2. 무성 파열음과 관련 정보 추출

스크립트를 이용하여 측정 대상이 된 무성 파열음들로부터 추출된 정보는 화자 번호, 성별, 연령대, 파열음 종류, 파열음 길이, 파열음 소속 단어, 단어/발화 내 위치(어두, 어중, 어말이나 발화초, 발화 중), 이전/이후 단어, 이전/이후 변이음, 개방 파열 시작/끝 시간, 중력중심값, 분산, 비대칭도, 첨예도 등이다. 특히, 중력중심값을 비롯한 네 가지의 값은 개방 파열 시작부터 (후속 모음 시작 지점을 침범하지 않는 범위 내에서) 10, 20, 30, 40 msec 지점에서 해밍 윈도우를 씌워 추출한 다음 계산한 값을 기록하도록 하였다. 추출된 무성 파열음 개수는 영어의 경우는 [p, t, k] 각각 6,548개, 13,780개, 11,775개이고, 한국어의 경우는 [p^h, t^h, k^h] 각각 3,862개, 8,647개, 6,375개이며, 세부적인 사항은 <표 1>에 나타낸 바와 같다.

표 1. 분석에 사용된 무성 파열음의 개수
Table 1. Number of voiceless plosives analyzed

벅아이 코퍼스				
무성 파열음		p	t	k
성별	남성	3,567	7,337	5,786
	여성	2,981	6,443	5,989
	소계	6,548	13,780	11,775
연령	낮음	2,973	5,743	5,469
	높음	3,575	8,037	6,306
	소계	6,548	13,780	11,775
위치	어두	3,962	7,987	6,701
	어중	2,157	4,495	3,110
	어미	429	1,298	1,964
	소계	6,548	13,780	11,775
합계		32,103		

서울 코퍼스			
격음	p^h	t^h	k^h
남성	2,050	4,592	3,171
여성	1,812	4,055	3,204
소계	3,862	8,647	6,375
10 대	720	1,792	1,186
20 대	893	2,262	1,402
30 대	1,237	2,464	2,031
40 대	1,012	2,129	1,756
소계	3,862	8,647	6,375
어두	1,668	1,617	1,560
어중	2,194	7,030	4,815
소계	3,862	8,647	6,375
총합	18,884		

2.3. 분석 방법

중력중심값 등 수치로 표시할 수 있는 값들은 표를 통하여 평균 및 표준편차 등의 기술통계량을 제시하였고 경우에 따라 히스토그램 등을 사용하여 분포의 경향을 나타내었다. 추론 통계 분석에는 통계 프로그램인 RStudio[23]를 사용하였고 유의성은 95% 신뢰구간을 기준으로 하였다.

3. 결과

3.1. 개방 파열 시작 탐지 스크립트의 검증

영어와 한국어 자연발화 코퍼스에 존재하는 방대한 양의 무성 파열음들의 개방 파열 시작 지점을 자동으로 알아내기 위해 프랏 스크립트를 활용하였음은 전술한 바와 같다. 하지만 중요한 것은 이 스크립트가 음성학자(저자)와 얼마나 비슷한 능력을 발휘하느냐 하는 것이다.

표 2. 개방 파열 측정 스크립트 성능
Table 2. Performance of the script used for identifying burst onset

화자	벅아이 코퍼스			서울 코퍼스		
	인식률 (사람)	인식률 (script)	정확도 (script)	인식률 (사람)	인식률 (script)	정확도 (script)
1	100	80	95	100	76	95
2	96	84	100	96	92	91
3	96	96	96	96	88	86
4	100	84	100	100	92	78
5	88	84	100	100	96	96
6	92	92	96	100	96	83
7	96	92	96	100	96	100
8	84	72	94	96	92	100
9	76	60	100	100	92	91
10	92	72	89	100	96	100
11	100	96	100	100	96	92
12	92	68	100	100	96	88
13	88	84	90	92	80	100
14	96	92	96	80	64	100
15	92	64	94	96	92	96
16	92	84	95	96	96	92
17	100	100	96	100	100	80
18	100	100	100	100	96	79
19	100	100	100	100	92	100
20	92	88	95	96	92	96
21	84	80	100	100	96	92
22	96	64	88	100	100	84
23	92	80	90	96	96	88
24	92	88	91	96	92	96
25	100	100	92	92	84	90
26	96	92	83	100	100	64
27	80	72	89	96	92	87
28	80	76	100	96	96	96
29	84	80	80	100	88	100
30	80	64	94	100	100	84
31	92	76	95	96	84	86
32	80	72	100	100	96	83
33	96	92	91	92	84	95
34	72	68	100	100	96	79
35	80	80	90	96	96	88
36	92	92	96	100	100	76
37	80	80	90	96	92	91
38	72	72	89	88	84	86
39	88	76	100	100	100	84
40	84	76	95	100	100	84
평균	**90%**	**82%**	**95%**	**97%**	**92%**	**89%**

이를 알아내기 위하여 사람과 스크립트의 능력을 두 가지 관점에서 비교 판단하였다. 첫째는 해당 무성 파열음에 측정 가능한 개방 파열이 존재하는지를 사람과 스크립트 모두 올바르게 인식해내는가를 판단하는 것이고, 다른 하나는 개방 파열이 존재한다는 가정 하에 그 시작 지점을 사람과 스크립트가 얼마나

정확하게 알아내는가 하는 점이다. 성능 평가의 기준은 사람을 기준으로 삼았다. 만일 스크립트가 사람만큼 100%의 능력을 발휘한다면, 스크립트로 알아낸 개방 파열의 시작 지점은 사람이 수행한 것과 동일하다고 볼 수 있을 것이다.

우선 영어 벅아이 코퍼스의 40 명 화자에 대하여 한 명당 무작위로 25 개의 무성 파열음, 총 1,000 개를 선정하였다. 그런 다음 각각의 분석 대상에 대하여 스크립트를 실행하여 개방 파열의 존재 유무를 인식하는지와 존재하는 개방 파열의 시작 지점을 정확하게 알아내는지를 조사하였다. 사람이 판단하기에 개방 파열이 존재한다고 보는 많은 경우에 스크립트도 같은 판단을 하였다. 물론 사람은 있다고 보는데 스크립트는 찾아내지 못하는 경우도 간혹 있었다. 사람과 스크립트 모두 개방 파열이 존재한다고 판단한 경우, 그 시작점을 스크립트로 하여금 측정하게 하였다. 측정된 지점에 사람도 동의하는 경우가 많았지만, 간혹 스크립트가 측정한 지점이 틀린 경우도 있었다. 위와 같은 방식으로 한국어 서울 코퍼스의 40 명 화자에 대해서도 무작위로 추출된 1,000 개의 무성 파열 격음에 대하여 스크립트의 성능을 가늠해 보았다. 그 결과는 <표 2>에 제시하였다.

<표 2>의 결과에서 보듯이, 인식률과 정확도에 있어서 스크립트는 물론 사람의 능력을 따라잡지는 못했다. 하지만, 영어의 경우 개방 파열이 존재한다고 사람이 판단한 경우가 90%라면 스크립트도 82% 정도는 동일하게 판단하였다. 한국어의 경우는 사람이 97%이고 스크립트가 92% 정도로 영어보다는 인식률이 조금 나았다. 개방 파열이 있는데도 놓친 경우를 제외해 놓고 보면, 일단 존재한다고 인식한 경우에 대하여 정확도를 살펴보면 사람을 100%로 봤을 경우 스크립트는 영어의 경우 95%, 한국어의 경우 89% 정도의 정확도를 보인 것으로 나타났다.

영어의 경우 정확도가 다소 높지만, 사람을 100 으로 봤을 때 스크립트가 대략 90% 혹은 이상의 능력을 보인다는 것은 주목할 만하다. 이 말은 <표 1>에서 분석 대상이 된 영어 무성 파열음 32,103 개와 한국어 격음 18,884 개의 파열음이 인식률 때문에 실제로 존재하는 수보다 다소 적기는 해도 측정의 정확도에 있어서는 사람 능력의 90% ~ 95%정도가 된다는 의미이다. 물론 전문가가 일일이 수작업으로 개방 파열 시작 지점을 파악하는 것이 이상적이겠지만, 코퍼스 연구의 특성상 대량의 자료를 신뢰할만한 방법으로 자동 추출해내는 것도 의미 있는 작업이라고 생각된다. 따라서 추후의 분석 결과를 해석함에 있어서 이러한 사항들을 유의하기 바란다.

3.2. 분석 대상 파열음의 중력중심값 분포

표 3. 개방 파열 중력중심값의 평균 및 표준편차
Table 3. means and standard deviations for the center of gravity values of the release bursts

단위: Hz	벅아이 코퍼스			서울 코퍼스		
파열음	p	t	k	p^h	t^h	k^h
평균	**527**	**2150**	**1526**	**537**	**1284**	**1390**
표준편차	488	1511	750	740	1263	1017

분석 대상인 영어와 한국어의 파열음에 대하여 [6]에서와 같이 개방 파열 40 ms에 대한 중력중심값을 구해보면 <표 3>과 같고 히스토그램을 그려보면 <그림 3>과 같은 양상을 나타낸다.

<표 3>의 평균값으로만 보면, 영어와 한국어 두 언어에 있어서 양순 및 연구개 파열음의 경우는 중력중심값이 서로 유사하다고 할 수 있지만, 치경 파열음의 경우는 서로 상당히 다르다는 것을 알 수 있다. 이러한 양상은 <그림 3>에서도 볼 수 있는데, 왼쪽 패널의 영어에서는 세 파열음의 중력중심값 분포 양상이 서로 다소의 차이를 보이는데 반해, 오른쪽 패널의 한국어의 경우는 값들이 주로 저주파에 몰려 있고, 치경 및 연구개 파열음의 경우는 분포하는 영역이 상당이 많이 겹쳐있음을 볼 수 있다.

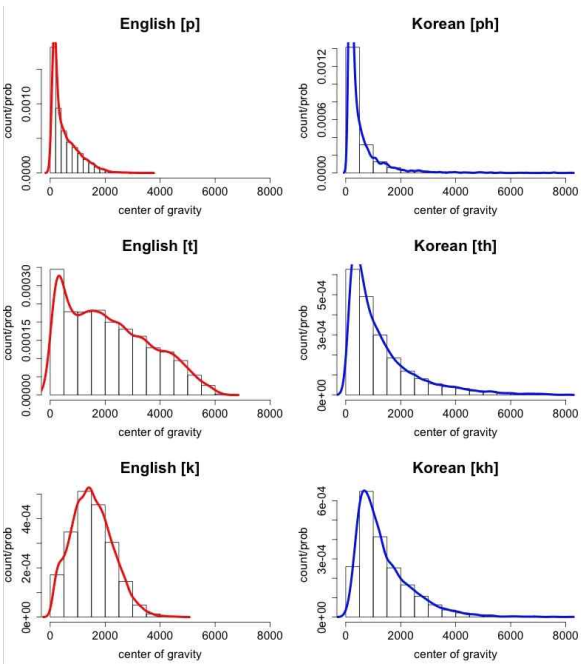

그림 3. 영어와 한국어 개방파열의 중력중심값 히스토그램
Figure 3. Historgrams for the center of gravity values of English and Korean release bursts

한국어의 경우 요인을 어두로 제한하여 어두 위치에 있는 파열음의 개방 파열에 대하여만 히스토그램을 다시 그려보았고 그 결과는 <그림 4>에 나타내었다.

<그림 4>에서 볼 수 있듯이, 양순음의 경우는 값들의 분포 범위가 치경음에 비해 상대적으로 좁았으나, 최빈값의 분포 위치는 유사했고, 치경음과 연구개음의 경우는 분포 범위는 양순음에 비해 상대적으로 넓었고 최빈값의 위치도 다른 양상을 보였다.

기술 통계량으로 살펴본 분포 양상을 추론 통계를 통해 자세히 알아보기 위하여 언어별 파열음의 조음위치를 요인으로, 중력중심값을 결과값으로 보고 반복측정 분산분산을 실시하였다. 영어의 경우에는 각 길이의 개방 파열에 대하여 [p, t, k]가 중력중심값에 대하여 통계적으로 의미 있는 영향을 미치고 있는 것으로 나타났다(40 ms의 경우, $F(2,37) = 11.24$, $p < 0.05$; 30 ms의

경우, F(2,37) = 11.42, p <0.05; 20 ms의 경우, F(2,37) = 8.792, p < 0.05; 10 ms의 경우, F(2,37) = 4.008, p < 0.05). 그러나, 한국어의 경우에는 [p^h, t^h, k^h]가 통계적으로 의미 있는 영향을 미치지 않는 것으로 나타났다(40 ms의 경우, F(2,37) = 2.191, p > 0.05; 30 ms의 경우, F(2,37) = 1.691, p > 0.05; 20 ms의 경우, F(2,37) = 0.786, p > 0.05; 10 ms의 경우, F(2,37) = 0.15, p > 0.05).

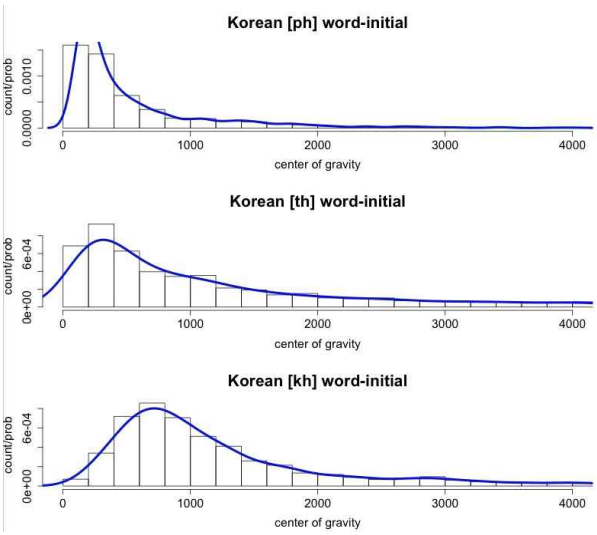

그림 4. 한국어 어두 파열음의 개방파열에 대한 중력중심값 히스토그램

Figure 4. Historgrams for the center of gravity values of the release bursts for Korean word-initial plosives

3.3. 운동량 판별분석을 통한 파열음 자동 분류

파열음 개방파열의 스펙트럼에서 얻은 네 가지 운동량 즉, 중력중심값, 분산, 비대칭도, 첨예도를 바탕으로 해당 파열음의 조음위치를 예측할 수 있는지를 알아보기 위하여 프랏에 구현되어 있는 다변량 통계 분석 기능 중의 하나인 판별분석(discriminant analysis)을 활용하였다.

자료를 트레이닝 자료와 테스트 자료로 나누는 방법은 프랏 매뉴얼에서 추천하는 방식 중 하나인 Jacknife (leave-one-out) 분류 방식을 이용하여 프랏 스크립트를 작성한 다음 시행하였다. 이 방식은 테스트 자료로 하나의 자료만 제외하고 나머지 모든 자료를 트레이닝 자료로 이용하여 예측 모델을 구축하는 것인데, 이때 구축된 예측 모델로 테스트 자료를 예측하는 것이다. 이 절차가 모든 자료에 대하여 반복되고 예측의 정확도가 혼동행렬표(confusion matrix)의 형태로 출력된다.

영어와 한국어의 자료인 32,103 개와 18,884 개를 대상으로 각각 판별분석을 시행하였는데, 개방 파열의 시작 지점에서 10, 20, 30, 40 msec 등 네 경우를 측정하였으므로 각 언어별로 네 번의 판별분석을 시행하여 어떤 경우에 가장 예측 성능이 좋은지를 알아보았다. 시행한 총 판별분석의 수는 203,948 회이다 (영어: 32,103 × 4 = 128,412 번, 한국어: 18,884 × 4 = 75,536 번). 판별분석 결과와 가장 예측 정확도가 높은 경우에 대한 혼동 행렬표를 <표 4>에 나타내었다. 혼동 행렬표에서 보듯이 영어의 경우에는 [p, t, k]를 각각 [k, k, t]로 예측하는 실수가 많았고, 한국어의 경우에는 [p^h, t^h, k^h]를 각각 [t^h, k^h, t^h]로 잘못 예측하는 경우가 많았다.

표 4. 판별분석 후 % 정확도와 혼동 행렬표

Table 4. Results of the discriminant analyses

판별 분석	개방 파열 측정 길이			
% 정확도	10 msec	20 msec	30 msec	40 msec
영어	51.8	63.1	**63.2**	60.9
한국어	50.2	60.9	**61.4**	58.5

영어 30 msec		예측값		
		p	t	k
관측값	p	**2,822**	312	3,414
	t	1,076	**8,601**	4,103
	k	530	2,366	**8,879**

한국어 30 msec		예측값		
		p^h	t^h	k^h
관측값	p^h	**1,609**	2,040	213
	t^h	547	**7,300**	800
	k^h	436	3,247	**2,692**

<표 4>의 % 정확도 부분에서 보듯이 영어와 한국어 모두, 개방 파열 시작부터 30 msec까지의 스펙트럼에서 얻은 중력중심값, 분산, 비대칭도, 첨예도가 가장 예측력이 강하였다. 하지만 두 경우 모두 60% 초반 대에 그친 예측 정확도는 기존의 연구결과와는 차이를 보인다. 영어의 경우 90% 이상의 높은 수치를 보이는 [6]의 연구는 어두 무성 파열음만을 분석한 것을 감안하더라도 본 연구의 결과와는 매우 차이가 크다. 이 차이는 틀문장에 넣은 단어를 읽은 녹음과 자연 발화 음성 코퍼스의 차이에서 오는 것일 수도 있다.

본 연구의 결과는 50% 후반대의 수치를 보이는 [7]의 연구와 결과값이 유사함을 알 수 있는데 해당 연구에서 분석에 사용된 자료는 대화체 녹음에서 추출한 것이다. 따라서 본 연구에 사용된 코퍼스가 인터뷰 형식의 대화체 음성 발화임을 감안하면, 틀문장을 통해서 만들어진 파열음과 대화체 녹음에서 추출된 파열음의 성질이 적어도 개방 파열 부분에 있어서 상이하다는 것과 개방 파열 부분에서 구한 네 가지 운동량값들만으로는 파열음 조음 위치의 구분이 만족스럽지 않다는 것을 알 수 있다.

성별과 단어 내 위치 요인으로 인해 판별분석의 예측력이 영향을 받았을 가능성이 있으므로 이번에는 성별을 남성으로, 단어 내 위치를 어두로 하여 자료를 추출한 다음 판별분석을 다시 실시하였고 그 결과를 <표 5>에 제시하였다. 표에서 보듯이 영어와 한국어 모두 예측력이 다소 하락한 것을 볼 수 있다. 따라서 두 요인을 통제한 것이 예측력에 긍정적으로 작용하지 않았음을 알 수 있다.

운동량을 네 개의 수치로 변환시켜 스펙트럼의 경향을 포착하려는 위의 시도와는 다소 다르지만, [4]나 [5]의 연구에서는 스펙트럼의 2 차원 분포 양상을 조음 위치별로 특정한 기울기(spectral tilt) 혹은 오르내리는 모양을 지니고 있다고 가정을 하여 스펙트럼 템플릿을 설정하였다. 운동량을 바탕으로 한 판별

분석과는 어떤 차이를 보일지 분석을 해 보았다.

표 5. 추가 판별분석 후 % 정확도와 혼동 행렬표
Table 5. Results of the second discriminant analyses

판별 분석	개방 파열 측정 길이			
% 정확도	10 msec	20 msec	30 msec	40 msec
영어	49.4	**61.0**	60.5	57.7
한국어	52.4	**56.8**	55.6	51.6

영어 20 msec		예측값		
		p	t	k
관측값	p	**1,096**	587	540
	t	436	**2,592**	1,141
	k	230	873	**2,261**

한국어 20 msec		예측값		
		p^h	t^h	k^h
관측값	p^h	**642**	213	72
	t^h	218	**526**	151
	k^h	234	252	**328**

3.4. 스펙트럼 템플릿 판별분석을 통한 파열음 자동 분류

선행 연구들([2], [3], [4], [5])에 의하면 파열음 개방 파열의 스펙트럼 템플릿은 대략적으로 <그림 5>와 같이 볼 수 있다. 또한 대략 4,000 Hz를 중심으로 치경 파열음의 경우는 그 위 주파수에, 양순 파열음은 그 아래 주파수에, 연구개 파열음은 그 부분에 에너지가 집중되어 있는 양상을 보인다고 한다.

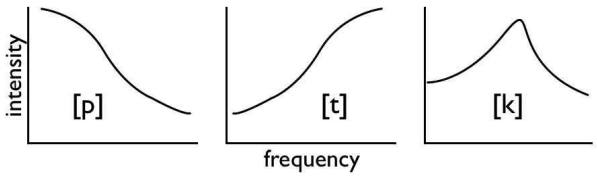

그림 5. 양순, 치경, 연구개 파열음의 스펙트럼 템플릿 [4]
Figure 5. Spectral templates for bilabial, alveolar and velar plosives [4]

따라서 스펙트럼 템플릿을 판별분석의 자료로 이용하기 위하여 4,000 Hz를 중심으로 그 아래 부분과 위 부분의 스펙트럼 기울기를 각각 구하였다. 이렇게 하면, 양순 파열음의 스펙트럼은 두 기울기 모두 하강하는 양상을, 치경 파열음은 모두 상승하는 양상을, 연구개 파열음의 경우는 상승하다가 하강하는 양상을 보이게 될 것이므로 세 파열음을 구분할 수 있어 판별분석에 긍정적인 효과를 보일 것으로 예상할 수 있다.

10 에서 40 ms까지의 네 가지 개방 파열 길이에 대하여 LPC 처리를 거친 스펙트럼 템플릿만을 자료로 사용해 판별분석을 시행한 결과는 <표 6>에 나타내었다. 60% 초반대의 정확도를 나타내는 운동량 기반의 판별분석 결과(<표 4> 참조)와는 달리, 스펙트럼 기울기를 기반으로 하는 판별분석 결과는 대략 50% 중반대의 정확도를 나타내고 있다. 특히 혼동 행렬표에서 볼 수 있듯이 영어와 한국어 모두 양순 파열음의 예측 정확도가 극히 좋지 않다. 운동량[6]과 스펙트럼 기울기[5]를 중심으로 파열음 분류의 정확도를 살펴본 선행 연구에서는 각각 90%, 80% 이상의 높은 정확도를 보고하고 있으나, 이들 연구 모두 미리 주어진 단어나 음절을 보거나 듣고 따라 읽는 형태로 녹음된 음성 자료를 분석한 것이기 때문에, 주어진 목록이나 자료 없이 즉흥적인 대화 방식의 자연스런 인터뷰를 녹음한 코퍼스와는 발음의 명료성이나 발화 속도 등의 여러 음성적 특질에 있어서 차이를 보일 수 있다.

표 6. 스펙트럼 템플릿을 이용한 판별분석 결과
Table 6. Results of the discriminant analyses from spectral templates

판별 분석	개방 파열 측정 길이			
% 정확도	10 msec	20 msec	30 msec	40 msec
영어	53.7	**55.4**	55.4	53.3
한국어	53.3	**54.7**	53.5	53.0

영어 20 msec		예측값		
		p	t	k
관측값	p	**2**	2,153	4,393
	t	8	**10,750**	3,022
	k	9	4,719	**7,047**

한국어 20 msec		예측값		
		p^h	t^h	k^h
관측값	p^h	**0**	3,382	480
	t^h	0	**8,092**	555
	k^h	0	4,132	**2,243**

이번에는 운동량과 스펙트럼 템플릿 모두를 자료로 사용해 추가적으로 판별분석을 시행하였고, 그 결과를 <표 7>에 나타내었다.

표 7. 운동량과 스펙트럼 템플릿 모두를 사용한 판별분석 결과
Table 7. Results of the discriminant analyses from momentum and spectral templates

판별 분석	개방 파열 측정 길이			
% 정확도	10 msec	20 msec	30 msec	40 msec
영어	57.6	**65.5**	64.5	62.3
한국어	53.6	63.0	**63.3**	61.0

영어 20 msec		예측값		
		p	t	k
관측값	p	**2,845**	1,023	2,680
	t	1,157	**10,133**	2,490
	k	663	3,060	**8,052**

한국어 30 msec		예측값		
		p^h	t^h	k^h
관측값	p^h	**1,641**	1,893	328
	t^h	587	**7,153**	907
	k^h	187	3,031	**3,157**

표에서 보듯이 판별분석의 정확도는 영어와 한국어 모두 60% 중반에 가까운 값을 보이고 있다. 이는 스펙트럼 템플릿만을 이용하였을 때보다 10% 정도 증가한 값이지만, <표 4>의 운동량 경우와 비교해보면 약 2% 정도의 상승에 불과하다. 혼동 행렬표를 보면 영어의 경우 [p], 한국어의 경우는 [p^h, k^h]의 예측

정확도가 그리 높지 않을 것을 볼 수 있다. 양순 파열음 개방 파열의 경우는 스펙트럼 기울기가 <그림 5>처럼 이상적이지 못하고 다른 파열음과 구분이 쉽지 않음을 짐작할 수 있다.

파열음 개방 파열 부분의 스펙트럼에서 얻은 중력중심값, 분산, 비대칭도, 첨예도 등의 운동량을 기반으로 판별분석을 시행한 경우와, 스펙트럼 기울기로 나타내어진 스펙트럼 템플릿을 기반으로 판별분석을 시행하였을 경우, 혹은 이들 두 가지 경우를 모두 합하여 판별분석을 시행하였을 경우 등 세 가지 경우에 대하여 영어와 한국어에 대한 비교를 <그림 6>에 나타내었다. 그림에서 볼 수 있듯이 어떤 경우에도 70% 정도의 예측 정확도를 넘기지 못하는 것을 알 수 있었다.

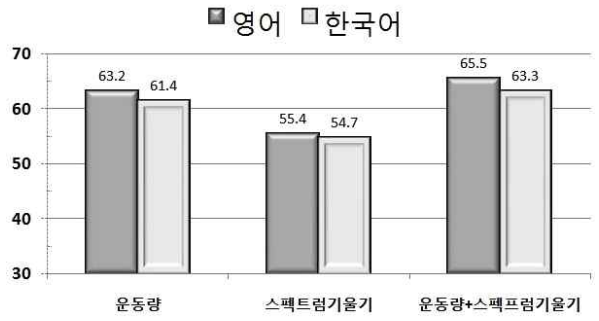

그림 6. 영어와 한국어의 판별분석 결과 비교
Figure 6. Comparison of the discriminant analyses of English and Korean

이러한 사실은 80%에서 90%를 상회하는 선행 연구들에서 쓰인 단어나 틀문장 낭독체의 음성 자료가 자연발화 음성 코퍼스와 개방 파열 부분의 음성 특징에 있어서 차이가 있음을 반영하는 것으로 볼 수 있으며, 파열음의 조음 위치 구분이 파열음의 개방 파열 단서만으로는 만족스럽게 구분할 수 없다는 것을 암시하는 것으로 볼 수 있을 것이다.

4. 결론

본 연구에서는 영어 벅아이 코퍼스와 한국어 서울 코퍼스에서 각각 무성 파열음과 격음 파열음을 자동 추출하여 개방 파열 부분에서의 스펙트럼의 양상을 살펴보았다. 특히, 중력중심값, 분산, 비대칭도와 첨예도 등으로 표현되는 운동량과 스펙트럼 기울기로 표현되는 스펙트럼 템플릿을 기반으로 다변량 통계 기법 중의 하나인 판별분석을 통하여 파열음의 조음 위치를 자동으로 분류하여 보았다.

개방 파열을 자동으로 인식하여 측정하는 프랏 스크립트의 경우 사람의 판단 지점과 5 ms 이내의 오차 범위 내에서 영어의 경우 인식률은 82%(사람은 90%), 측정 정확도는 95%(사람은 100%), 한국어의 경우 인식률은 92%(사람은 97%), 측정 정확도는 89%(사람은 100%)였다. 사람보다 인식률과 정확도는 다소 뒤지지만 일단 찾아낸 개방 파열의 시작 지점을 측정하는 데에는 89 ~ 95% 정도의 정확도를 보여 수작업이 대체 가능하다는 가능성을 보여주었다고 볼 수 있다.

영어의 무성 파열음 32,103 개의 경우, 개방 파열의 조음 위치가 중력중심값에 통계적으로 의미 있는 영향을 미치고 있었지만, 한국어의 격음 파열음 18,884 개의 경우, 그렇지 않았다. 네 가지 운동량을 기반으로 수행한 판별분석 결과를 보면, 두 언어 모두 개방 파열의 길이가 30 ms일 때 가장 높은 예측 정확도를 보여, 영어의 경우 63.2%, 한국어는 61.4%를 나타내었다. 스펙트럼 템플릿을 기반으로 수행한 판별분석 결과는 개방 파열의 길이가 20 ms일 때 가장 높은 정확도를 보였으나, 전체적인 예측 정확도는 운동량의 경우보다 낮아서 영어는 55.4%, 한국어는 54.7%를 나타내었다. 운동량과 템플릿 모두를 기반으로 판별분석을 수행했을 때는 예측력이 다소 높아져 영어는 65.5%, 한국어는 63.3%의 예측 정확도를 보였다.

이러한 결과는 분류의 정확도나 가능성 측면에서 80 ~ 90% 혹은 그 이상의 수치를 보이고 대부분의 선행 연구와는 큰 차이를 보이고 있다. 거의 모든 선행 연구들은 분석 대상인 단어나 음절만을 낭독체 발화로 읽어 녹음한 것들인 반면, 본 연구의 분석 대상은 대화체를 기반으로 하는 자연발화 녹음이다. 녹음 타겟 단어나 음절을 명료하게 혹은 또박또박 발음하는 경향이 두드러질 수밖에 없는 낭독체 발화 방식은 분절음의 특성에 영향을 미쳤을 가능성이 크다. 유일하게 대화에서 추출한 자료를 분석 대상으로 설정한 [7]의 연구 결과 수치들이 본 연구 결과에 나타난 수치들과 유사한 범위에 분포한다는 것은 이러한 추정을 뒷받침하는 것이라고 볼 수 있다.

본 연구에서는 개방 파열 부분만을 분석의 대상으로 삼아 파열음 조음 위치에 대한 분류 예측 정확도를 살펴보았지만, 발화의 일부로서의 파열음은 후속 모음과의 상호작용도 무시할 수 없을 것이다. 결국 특정한 인지적 단서는 여러 가지 음성학적 단서의 종합체(cue package)로 구성된다는 [19]의 주장처럼, 개방 파열에서만의 단서들뿐 아니라 후속 모음과 유기적으로 생성되는 성대진동개시시간(voice onset time)의 길이나 후속 모음 시작 부분의 스펙트럼 단서, 이러한 추가적인 단서들의 시간에 따른 변화 양상 등도 파열음 인지에 기여하고 있음이 분명하다. 물론 이번 연구에서 10 ~ 40 ms까지 네 단계로 개방 파열 길이를 변화시켜 시간적인 정보를 어느 정도 확보하기는 하였으나 판별분석 시행 결과, 분류 예측력의 정확도에서 있어서 크게 기여하는 점은 없었다고 생각된다.

개방 파열 스펙트럼의 분석만으로 대부분의 파열음 분류가 가능했던 낭독체 음성 분석 기반의 선행 연구들과는 달리, 즉흥적 대화체 음성을 기반으로 만들어진 영어 및 한국어의 자연발화 음성 코퍼스는 개방 파열 분석만으로는 무성 파열음 분류에 한계가 있으며, 정적인 스펙트럼 분석뿐 아니라 동적인 변화와 후속 모음과의 유기적 관계도 살펴봐야 할 것이라는 점을 암시하고 있다는 면에서, 또한 개방 파열 부분의 정적인 스펙트럼 분석 시에 스펙트럼 템플릿보다는 운동량 정보가 파열음 분류의 정확도에 있어서 예측력이 다소 높다는 것을 밝혀낸 것이 본 논문의 의의라고 할 수 있겠다.

참고문헌

[1] Cooper, F., Delattre, P., Liberman, A., Borst, J., & Gerstman, L. (1952). Some experiments on the perception of synthetic speech sounds. *The Journal of the Acoustical Society of America*, 24, 597-606.

[2] Halle, M., Hughes, G., & Radley, J. (1957). Acoustic properties of stop consonants. *The Journal of the Acoustical Society of America*, 29, 107-116.

[3] Stevens, K., & Blumstein, S. (1975). Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *Journal of Phonetics*, 3, 215-233.

[4] Stevens, K., & Blumstein, S. (1978). Invariant cues for the place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, 64, 1358-1368.

[5] Blumstein, S., & Stevens, K. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America*, 66, 1001-1017.

[6] Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *The Journal of the Acoustical Society of America*, 84, 115-123.

[7] Winitz, H., Scheib, M., & Reeds, J. (1972). Identification of stops and vowels for the burst portion of the /p, t, k/ isolated from conversational speech. *The Journal of the Acoustical Society of America*, 51, 1309-1317.

[8] Cole, R., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review*, 81, 348-374.

[9] Ohde, R., & Sharf, D. (1977). Order effect of acoustic segments of VC and CV syllables on stop and vowel identification. *Journal of Speech and Hearing Research*, 20, 543-554.

[10] Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, 73, 322-335.

[11] Kewley-Port, D. (1983). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. *The Journal of the Acoustical Society of America*, 72, 379-389.

[12] Bonneau, A., Djezzar, L., & Laprie, Y. (1996). Perception of the place of articulation of French stop bursts. *The Journal of the Acoustical Society of America*, 100, 555-564.

[13] Smits, R., ten Bosch, L., & Collier, R. (1996). Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment. *The Journal of the Acoustical Society of America*, 100, 3582-3864.

[14] Park, H. (2003). Spectral characteristics of release bursts. *Proceedings of the Korean Society of Speech Sciences* (pp. 159-162). (박한상 (2003). 개방 파열의 스펙트럼상의 특성. *대한음성학회 학술대회 논문집*, 159-162.)

[15] Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). Buckeye Corpus of Conversational Speech (2nd release). [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).

[16] Yun, W., Yoon, K., Park, S., Lee, J., Cho, S., Kang, D., Byun, K., Hahn, H., & Kim, J. (2015). The Korean corpus of spontaneous speech. *Phonetics and Speech Sciences*, 7(2), 103-109.

[17] Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5(9/10), 341-345.

[18] R Studio Team. (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. Retrieved from http://www.rstudio.com/ on March 31, 2016.

[19] Steriade, D. (1999). Phonetics in Phonology: The Case of Laryngeal Neutralization. *UCLA Working Papers in Linguistics*, 2, 25-146.

• **황선미 (Hwang, Sunmi)**
영남대학교 영어영문학과 영어학 박사과정
경상북도 경산시 대학로 280
Tel: 053-810-2130 Fax: 053-810-4607
Email: teasnake@naver.com
관심분야: 영어학
현재 영어영문학과 대학원 박사과정 수료

• **윤규철 (Yoon, Kyuchul)** 교신저자
영남대학교 영어영문학과
경상북도 경산시 대학로 280
Tel: 053-810-2145 Fax: 053-810-4607
Email: kyoon@ynu.ac.kr
관심분야: 음성학, 전산언어학