



A study on the voiceless plosives from the English and Korean spontaneous speech corpus*

Kyuchul Yoon**

Department of English Language & Literature, Yeungnam University, Gyeongsan, Korea

Abstract

The purpose of this work was to examine the factors affecting the identities of the voiceless plosives, i.e. English [p, t, k] and Korean [p^h, t^h, k^h], from the spontaneous speech corpora. The factors were automatically extracted by a Praat script and the percent correctness of the discriminant analyses was incrementally assessed by increasing the number of factors used in predicting the identities of the plosives. The factors included the spectral moments and tilts of the plosive release bursts, the post-burst aspirations and the vowel onsets, the durations such as the closure durations and the voice onset times (VOTs), the locations within words and utterances and the identities of the following vowels. The results showed that as the number of factors increased up to five, so did the percent correctness of the analyses, resulting in 74.6% for English and 66.4% for Korean. However, the optimal number of factors for the maximum percent correctness was four, i.e. the spectral moments and tilts of the release bursts and the following vowels, the closure durations and the VOTs. This suggests that the identities of the voiceless plosives are mostly determined by their internal and vowel onset cues.

Keywords: Seoul corpus, Buckeye corpus, Korean, English, plosives, moment, spectral tilt, discriminant analysis

1. 서론

파열음은 조음에 있어서 대체로 폐쇄구간(closure) 후에 개방파열(release burst)이 일어나고 성대진동개시시간(voice onset time, 이하 VOT)을 거쳐 후속 모음으로 진행된다고 알려져 있다. 각기 다른 파열음이 만들어지는 경우, 어떠한 요인들이 이에 기여하는지에 대한 연구는 다양한 각도에서 이루어져 왔다.

영어에 있어 선행 모음이 존재하는 경우, 파열음은 대체로 50~150 ms 정도의 폐쇄구간을 나타내고, 조음 기관이 개방될

때 5~40 ms 정도의 짧은 개방파열을 보이며, 유무성 여부에 따라 다양한 길이를 보이는 성대진동개시시간을 나타내게 된다 (Kent & Read, 2002). 파열음을 구성하는 이러한 요인들에 관한 주요 관심사 중 하나는, 이들이 각 파열음의 음성학적 차이를 얼마나 잘 나타낼 수 있는나 하는 점이다.

요인들을 하나씩 살펴보면, 우선 폐쇄구간의 경우, Zue(1976)와 Byrd(1993)는 영어의 [p]가 [t, k]보다 길다고 보고하였으나, Crystal & House(1988)는 [p, k]가 [t]보다 다소 길다고 주장하였다. 영어 백아이 코퍼스 일부를 대상으로 한 Yao(2007)의 연구

* This work was supported by the 2019 Yeungnam University Research Grant.

** kyoony@ynu.ac.kr, Corresponding author

Received 16 October 2019; Revised 29 November 2019; Accepted 30 November 2019

© Copyright 2019 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.

에서는 [p]의 길이가 60~90 ms로 제일 길었고, [k, t]에 있어서는 40~70 ms 정도의 분포를 보였으나, [k]가 10 ms 내외의 차이로 다소 긴 양상을 보였다고 하여 이전의 두 연구를 합쳐놓은 듯한 결과를 얻었다. 한국어의 경우는 경음이 제일 길고, 격음과 평음의 순서로 짧아지고, 조음위치로 보면 연구개음이 더 짧다고 보고되었다(Pae et al., 1999; Shin, 1997).

개방과열을 연구한 Halle et al.(1957)에서는 에너지가 집중되어 있는 부분이 양순, 치경, 연구개 과열음에 대하여 각각 500~1,500 Hz, 4,000 Hz 이상, 1,500~4,000 Hz 정도였다고 보고하고 있다. 이러한 양상을 스펙트럼 템플릿으로 살펴본 Stevens & Blumstein(1975, 1978)를 기반으로 Blumstein & Stevens(1979)는 화자들의 무성 과열음 자극을 85%의 정확도로 조음 위치에 따라 분류할 수 있었다고 한다. 스펙트럼 양상을 확률 분포로 간주하여 네 가지의 운동량(moment), 즉 중력중심값(center of gravity), 분산(variance), 비대칭도(skewness), 첨예도(kurtosis)의 관점에서 살펴본 Forrest et al.(1988)의 연구에서는 무성 과열음의 개방과열 시작 부분 40 ms를 분석하였다. 그 결과 개방과열의 운동량을 이용하여 92%의 정확도로 조음 위치를 구분할 수 있었다고 한다. 하지만, 대화체 영어 자료에서 뽑아낸 Winitz et al.(1972)의 연구에 있어서는 겨우 58% 정도로 정확도가 떨어져 연구마다 편차가 매우 심한 양상을 보였다.

영어 이외의 언어에 있어서는, Bonneau et al.(1996)의 프랑스어 연구에서 87%의 정확도를 보였고, 네덜란드어의 [k]는 개방과열만으로도 구분이 잘 되었으나 [p, t]는 정확도가 낮다고 하였다(Smits et al., 1996). 한국어의 경우는 서울 코퍼스의 격음 과열음을 연구한 Hwang & Yoon(2017)에서 63%의 정확도를 보였다고 한다. 같은 연구에서 영어 벽아이 코퍼스는 66%의 정확도를 보였다.

VOT의 경우 비교적 많은 연구가 이루어져왔다. 과열음의 유무성을 구분하는 기준으로서뿐 아니라 조음위치를 알려주는 역할로서도 잘 알려져 있다. Azou et al.(2000)은 여러 연구 자료를 바탕으로 영어의 경우 [p, t, k]의 VOT가 각각 -46~85 ms, -65~95 ms, -70~110 ms의 분포를 이룬다고 하였다. 중첩되기는 하지만, 대체로 양순<치경<연구개 과열음의 순서로 값이 증가한다(Kent & Read, 2002). 한국어는 Lee & Yoon(2016)에서 서울 코퍼스 20대 남성의 격음 과열음의 경우, 치경<양순<연구개 과열음의 순서로 값이 증가한다고 보고하고 있다.

과열음에서 후속 모음으로의 공명 패턴의 급격한 변화를 반영하는 모음의 포먼트 전이구간(formant transition)도 조음위치에 따른 과열음 구분에 도움을 준다고 알려져 있다. 영어 유성 과열음 [b, d, g]를 스펙트럼 패턴재생기를 통해 연구한 Delattre et al.(1955)에서는 포먼트 전이구간에서의 각 포먼트 주파수가 과열음의 조음위치 구분에 도움이 된다고 하였고, 아랍어 과열음을 연구한 Alwan(1989)은 전이구간의 제1 포먼트가 조음위치 구분에 핵심적이라고 주장하였다. 또한 Kent & Read(2002)는 후속 모음 시작 부분의 스펙트럼 양상도 역할을 한다고 주장하였다.

위에서 살펴본 연구들을 종합해 보면, 과열음의 각 구성 요소와 후속 모음의 시작 부위는, 조음위치에 따른 과열음 결정에

있어서 정도의 차이는 있으나 매우 중요한 역할을 하고 있다고 말할 수 있다. 따라서 본 연구에서는 각 과열음의 정체를 파악할 수 있게 해주는 여러 요인들을 통합적으로 살펴보고, 이들이 과열음의 조음위치 결정에 어느 정도로 기여하는지를 다변량 통계분석 기법 중의 하나인 판별분석(discriminant analysis)을 통해 살펴보고자 한다.

많은 선행 연구들이 주어진 단어나 문장을 녹음실에서 그대로 읽어 녹음한 자료를 바탕으로 하였기 때문에 실제로 자연스럽게 대화하는 상태에서 발화한 자료와는 많은 차이가 있을 것으로 예상할 수 있다. 그래서 본 연구에서는 자연발화 음성 코퍼스인 영어 벽아이 코퍼스(Pitt et al., 2007)와 한국어 서울 코퍼스(Yun et al., 2015)를 대상으로 하여 영어의 무성 과열음 [p, t, k]와 한국어 격음 과열음 [p^h, t^h, k^h]를 살펴보고자 한다.

2. 연구 방법

2.1. 연구 대상 및 추출 방법

무성 과열음을 추출하는 대상으로 사용된 코퍼스는 영어 벽아이 코퍼스와 한국어 서울 코퍼스이다. 두 코퍼스 모두 인터뷰 방식으로 두 사람이 여러 주제에 대하여 자유롭게 발화한 것을 녹음한 후, 단어와 변이음별로 레이블링한 자연발화 음성 코퍼스이다.

영어의 경우와는 달리, 한국어 과열음은 격음 이외에도 평음과 경음도 무성음으로 알려져 있으나, 본 연구에서는 모음 사이에서 유성음화가 이루어지지 않는 격음 분석을 통해, 영어 무성 과열음의 경우와 상호 비교해 보고자 한다.

두 코퍼스로부터 자료를 추출하는 방법은 프랏(Boersma, 2002) 스크립트를 작성하여 사용하였다. 영어 무성 과열음 [p, t, k]와 한국어 격음 [p^h, t^h, k^h]을 코퍼스의 변이음 층에서 모두 찾아 분석할 요인들에 대한 정보를 자동으로 추출하도록 하였다.

특히 개방과열의 시작 부분은 코퍼스에서 따로 표시가 되어 있지 않으므로 Hwang & Yoon(2017)에서 사용한 방법을 이용하여 스크립트를 통해 자동으로 찾도록 하였다. 과열음의 시작이 코퍼스에 경계로 표기되어 있으므로 개방과열 위치를 파악하면 폐쇄구간은 계산이 가능하고, 후속 모음의 시작도 코퍼스에 경계로 표시되어 있으므로 개방과열의 시작을 통해 자동으로 VOT 계산이 가능하다(그림 1 참조).

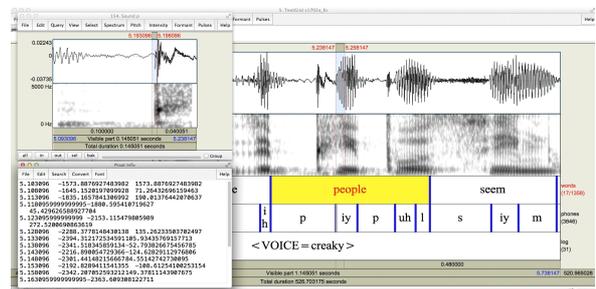


그림 1. 스크립트 실행 화면
Figure 1. Praat script running

2.2. 무성 파열음의 조음위치에 영향을 미치는 요인

프랏 스크립트를 통해 무성 파열음으로부터 자동으로 추출된 정보는 다음과 같다. 영어와 한국어 모두, 화자 번호, 성별, 연령대, 무성 파열음 종류, 파열음 길이, 파열음 소속 단어, 단어/발화 내 위치(어두, 어중, 어말 혹은 발화초, 발화중, 발화말), 직전/직후 단어, 직전/직후 변이음, 개방파열 시작/끝 시간, VOT, 폐쇄구간 길이, 개방파열 40 ms에 대한 중력중심값, 분산, 비대칭도, 첨예도 등 네 가지 운동량과 스펙트럼 기울기(10 ms 윈도우를 5 ms 간격으로 이동하면서 최대 7개의 값), 개방파열 후방 부위의 네 가지 운동량과 스펙트럼 기울기(VOT가 40 ms보다 긴 경우에만 후방 부위에 대하여 추출), 마지막으로 후속 모음 시작 부분의 운동량과 스펙트럼 기울기(대체로 포먼트 전이구간에 해당하는 것으로 보여지는 20 ms 부분)이다.

개방파열에 대한 운동량을 정적으로 추출하지 않고 짧은 윈도우를 단계적으로 이동시키면서 추출한 이유는 정적으로 추출한 Hwang & Yoon(2017)와 Kent & Read(2002)의 제안대로 기존 연구의 부족함을 극복하기 위해서이다.

또한 선행 연구에서는 주로 개방 파열 시작 부분 40 ms 정도를 살펴보았지만, 파열음의 종류나 환경에 따라 후속 모음이 시작되기 전까지 개방파열 이후에 기식음이 존재하는 경우가 있다. 본 연구에서는 이 부분이 파열음 조음위치 결정에 미치는 영향을 알아보기 위하여 개방파열 후방 부위에 대한 정보도 추출하였다. 발화초이면서 어두에 위치하여 폐쇄구간의 길이를 정할 수 없는 경우는 0값으로 처리하였다.

전술한 방법으로 추출된 자료는, 영어의 경우 [p, t, k] 각각 6,488개, 13,597개, 11,541개이고, 한국어의 경우 [p^h, t^h, k^h] 각각 3,862개, 8,647개, 6,375개이다. 성별, 연령별 위치별 세부 사항은 표 1에 나타내었다.

표 1. 추출된 무성 파열음의 개수
Table 1. Number of voiceless plosives extracted

		벽아이 코퍼스			서울 코퍼스			
무성 파열음		p	t	k	격음 파열음	p ^h	t ^h	k ^h
성별	남성	3,530	7,227	5,663	남성	2,050	4,592	3,171
	여성	2,958	6,370	5,878	여성	1,812	4,055	3,204
	소계	6,488	13,597	11,541	소계	3,862	8,647	6,375
연령	낮음	2,945	5,683	5,351	10대	720	1,792	1,186
					20대	893	2,262	1,402
	높음	3,543	7,914	6,190	30대	1,237	2,464	2,031
					40대	1,012	2,129	1,756
	소계	6,488	13,597	11,541	소계	3,862	8,647	6,375
위치	어두	3,551	7,154	6,080	어두	1,668	1,617	1,560
	어중	2,565	5,316	3,731	어중	2,194	7,030	4,815
	어미	372	1,127	1,730	어미	2,194	7,030	4,815
	소계	6,116	12,470	9,811	소계	1,668	1,617	1,560
	발화초	459	762	776	발화초	387	365	452
발화중	6,029	12,835	10,765	발화중	3,475	8,282	5,923	
소계	6,575	13,232	10,587	소계	2,055	1,982	2,012	
합계		31,626			총합	18,884		

2.3. 분석 방법

운동량 값 등 수치로 표시 가능한 것들은 먼저 표를 통하여 기술통계량을 제시하였고 필요한 경우 히스토그램 등의 그래프를 이용하여 분포 경향을 나타내었다. 추론 통계 분석에는 통계 프로그램인 RStudio(RStudio Team, 2015)를 사용하였고, 유의성은 95% 신뢰구간을 기본으로 하였다. 무성 파열음의 조음위치 결정에 영향을 미칠 것으로 예상되는 요인들을 대상으로 한 판별분석의 경우 프랏에 구현되어 있는 기능을 이용하였다.

3. 결과

3.1. 무성 파열음의 폐쇄구간 분석

한국어와 영어 무성 파열음의 폐쇄구간 길이를 나타내면 표 2와 같고, 분포를 히스토그램으로 나타내면 그림 2와 같다. 영어의 경우를 살펴보면 선행 연구에서 살펴보았듯이, [p]가 제일 길고, 그 다음으로 [k]와 [t]의 순서로 짧아지는 것을 볼 수 있다 (Byrd, 1993; Crystal & House, 1988; Yao, 2007). 히스토그램에서도 [p]는 100 ms 정도까지 분포하는 반면, [t, k]는 70~80 ms 정도까지 분포 범위가 다소 짧은 것을 볼 수 있다.

표 2. 폐쇄구간 길이의 평균 및 표준편차
Table 2. Means and standard deviations for the closure durations

단위: ms	벽아이 코퍼스			서울 코퍼스		
파열음	p	t	k	p ^h	t ^h	k ^h
평균	53	36	39	57	59	49
표준편차	29	24	22	37	35	31

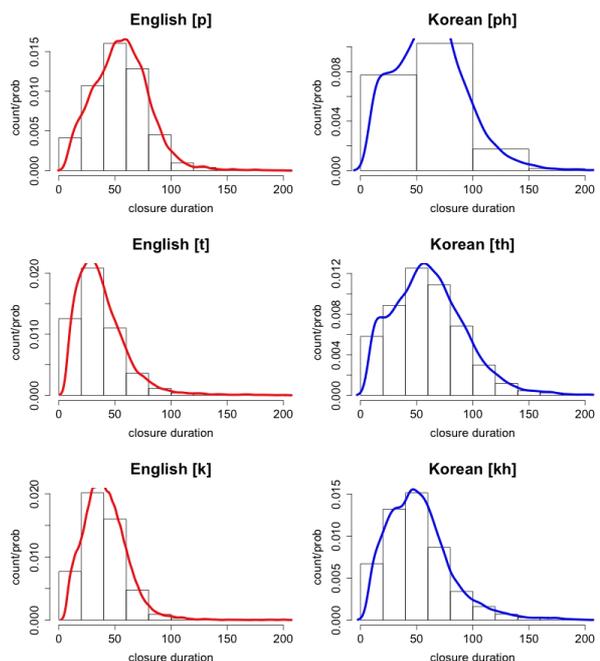


그림 2. 영어와 한국어의 폐쇄구간 히스토그램
Figure 2. Histograms for the closure durations (ms) of English and Korean

한국어의 경우도 선행 연구에서 보고한대로 연구개음의 폐쇄구간이 평균적으로 제일 짧았고(Pae et al., 1999; Shin, 1997), 양순음과 치경음은 매우 유사한 평균값을 나타내었다. 영어의 경우와는 달리 치경음의 폐쇄구간이 양순음과 거의 유사한 경향을 나타내었다. 히스토그램에서도 연구개음은 다른 파열음에 비해 다소 좁은 분포를 나타내었다.

폐쇄구간 길이가 파열음의 조음위치에 미치는 영향을 알아보기 위하여 일원분산분석을 영어와 한국어에 대하여 실시하였고, 그 결과 두 언어 모두 유의미한 영향을 나타내었다(영어: $F(2, 29626)=1,323, p<0.05$; 한국어: $F(2,17677)=182.2, p<0.05$). 사후분석 결과, 영어는 세 파열음 모두 서로 차이를 보였으나, 한국어는 [p^h, t^h] 사이에만 차이를 보이지 않고 나머지 경우 모두에는 차이를 보였다.

3.2. 무성 파열음의 VOT 분석

두 언어에 대한 무성 파열음 VOT의 평균값과 표준편차를 나타내면 표 3과 같고, 분포를 히스토그램으로 나타내면 그림 3과 같다. 영어의 경우를 살펴보면 선행 연구에서 제시한 바와 같이 양순<치경<연구개 파열음의 순서로 값이 증가하는 것을 알 수 있다. 히스토그램에서도 유사한 분포를 보이면서 약간씩 길어지는 양상을 보인다.

표 3. VOT의 평균 및 표준편차

Table 3. Means and standard deviations for the closure durations

단위: ms	백아이 코퍼스			서울 코퍼스		
파열음	p	t	k	p ^h	t ^h	k ^h
평균	46	53	57	57	52	65
표준편차	26	29	29	26	27	32

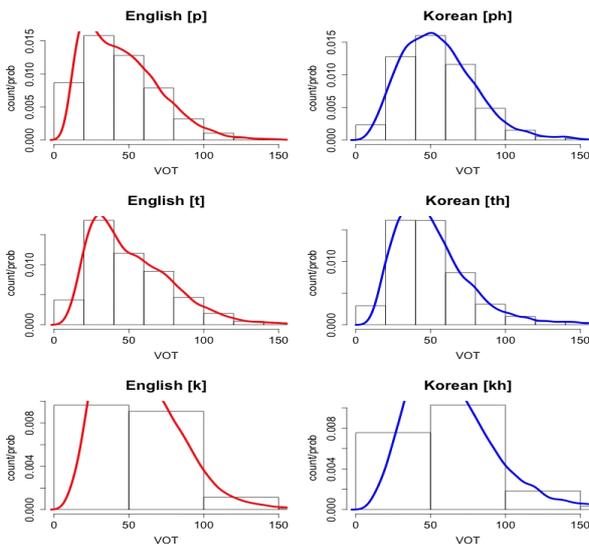


그림 3. 영어와 한국어의 VOT 히스토그램

Figure 3. Histograms for VOT (ms) of English and Korean

한국어의 경우는 Lee & Yoon(2016)에서 20대 남성에게 대하여 보고한 바와 같이 치경<양순<연구개 파열음의 순서로 평균값

이 증가하고 있다. 본 연구에서는 서울 코퍼스의 화자 전체를 대상으로 측정하였으므로 이러한 증가 추세는 모든 연령층에서 나타나는 것으로 볼 수 있다.

VOT 길이가 파열음의 조음위치에 미치는 영향을 알아보기 위하여 역시 일원분산분석을 두 언어에 대하여 실시하였고, 그 결과 두 언어 모두 유의미한 영향을 나타내었다(영어: $F(2, 29626)=287, p<0.05$; 한국어: $F(2,17677)=343.7, p<0.05$). 사후분석 결과, 영어와 한국어 모두 세 파열음 사이에 서로 차이를 보이는 것으로 나타났다.

3.3. 판별분석을 통한 파열음 자동 분류

다음으로 두 언어의 무성 파열음에서 얻은 여러 요인들이 해당 파열음의 조음위치 결정에 미치는 영향을 알아보기 위하여 다변량 통계 분석 기법인 판별분석(discrimant analysis)을 이용하여 알아보려고 한다. 판별분석은 주어진 자료들을 바탕으로 통계적 모델을 구축하고 이 모델을 이용하여 미지의 자료를 예측하는 방법으로 프랏에도 구현되어 있다. 기존 자료와 미지의 자료를 구분하여 모델의 성능을 검증하는 방법에는 여러 가지가 있지만 본 연구에서는 프랏에 구현되어 있는 방법 중에서 Jackknife(=leave-one-out) 방식을 이용했다. 이 방식은 검증 자료로 쓸 하나의 데이터만 제외하고 나머지 모든 데이터를 모델 구축에 이용한다. 이 절차가 단계적으로 모든 데이터에 대하여 반복 적용되어 모델의 측정된 예측 정확도가 혼동행렬표(confusion matrix)와 정확도 백분율(fraction correct)로 출력된다.

두 언어의 코퍼스로부터 추출된 정보들 중에서 판별분석에 이용가능한 요인 항목들은 (1) 10 ms 크기의 윈도우로 5 ms씩 이동하면서 측정된 개방파열 시작 40 ms 부분의 네 가지 운동량과 스펙트럼 기울기, (2) 개방파열 나머지 후방 부분의 네 가지 운동량과 스펙트럼 기울기, (3) 직후 모음 포먼트 전이 구간(20 ms)의 네 가지 운동량과 스펙트럼 기울기, (4) 폐쇄구간 길이, (5) VOT, (6) 발화 내 위치, (7) 단어 내 위치, (8) 직후 모음의 종류 등 총 여덟 가지이다. 발화초 위치이면서 어두인 경우의 폐쇄구간 값처럼 측정이 불가능한 경우는 모두 0으로 처리되었다. 특히 개방파열 시작 40 ms 부분은 짧은 윈도우가 5 ms씩 이동하므로 최대 7 세트의 값이 추출될 수 있다.

백아이 코퍼스에서 추출된 세 종류의 무성 파열음이 표 1에서 보듯 총 31,626개이고, 서울 코퍼스에서 추출된 파열음은 18,884개이므로, 모든 데이터에 대하여 시행된 판별분석의 횟수는 이들의 개수와 동일하여, 영어와 한국어 각각 31,626회, 18,884회 실시되었다. 어떠한 항목 혹은 항목의 조합들이 최상의 예측 정확도를 나타내는지 알아보기 위하여 이미 알려진 적은 수의 요인들로부터 단계적으로 요인들을 늘려가는 방식으로 판별분석을 실시하였다.

모델 구축에 사용되는 요인의 수뿐만 아니라 훈련에 사용되는 데이터 세트의 수도 중요한 역할을 수행한다. 표 4에서 보듯이 영어에 있어서 모델 구축에 사용되는 데이터 세트 수가 증가하면 예측 정확도가 증가하지만, 어느 정도 이상되면 증가의 폭이 큰 변화가 없어보인다.

표 4. 데이터 세트 수에 따른 판별분석 예측 정확도
Table 4. Results of the discriminant analyses

데이터 세트 수	1,000개	5,000개	10,000개	15,000개	20,000개
예측 정확도 %	68.2	75.3	74.6	75.5	75.0

3.3.1. 운동량과 스펙트럼 요인

먼저 기존 연구에서 살펴본 개방과열의 운동량과 스펙트럼 기울기를 중심으로 무성 과열음 종류의 예측 정확도를 살펴보자. 무성 과열음 개방과열 시작 40 ms 부분의 네 가지 운동량, 즉 중력중심값, 분산, 비대칭도, 첨예도만을 가지고 판별분석을 수행한 결과 얻은 조음위치에 대한 예측 정확도와 혼동행렬표는 표 5와 같다.

표 5. 개방과열의 운동량에 대한 판별분석 결과
Table 5. Discriminant analyses on the burst moments

영어: 68.4%			예측값		
			p	t	k
관측값	61.5%	p	3,993	581	1,919
	66.0%	t	1,409	8,992	3,216
	75.1%	k	1,011	1,875	8,690
한국어: 62.8%			예측값		
			p ^h	t ^h	k ^h
관측값	38.9%	p ^h	1,503	1,483	876
	78.5%	t ^h	408	6,787	1,452
	56.1%	k ^h	271	2,528	3,576

영어 전체의 예측 정확도는 68.4%로 한국어의 62.8%보다는 다소 높은 결과를 보였다. 이는 대화체 영어 자료를 바탕으로 수행한 연구(Winitz et al., 1972)에서의 58%보다는 10% 이상 높은 정확도를 보이고 있다. 영어의 경우 [k>t>p]의 순서로 예측 정확도가 높았으나 한국어의 경우는 [t^h>k^h>p^h]의 순서로 연구 개와 치경음의 순서가 바뀌는 것을 알 수 있다.

표 6. (1) 개방과열의 운동량과 스펙트럼에 대한 판별분석
Table 6. (1) Discriminant analyses on the burst moments & spectra

영어: 72.5%			예측값		
			p	t	k
관측값	64.0%	p	4,152	639	1,697
	73.5%	t	1,148	9,999	2,450
	76.0%	k	867	1,907	8,767
한국어: 65.1%			예측값		
			p ^h	t ^h	k ^h
관측값	41.9%	p ^h	1,619	1,210	1,033
	77.5%	t ^h	409	6,704	1,534
	62.4%	k ^h	252	2,146	3,977

운동량에 추가적으로 개방과열의 스펙트럼 양상을 기울기와 절편의 형태로 추가하여 판별분석을 시행하면 표 6과 같은 예측 정확도와 혼동행렬표를 얻게 된다. 이는 전술한 요인 (1)번에 해당된다.

이번에도 영어의 예측 정확도가 72.5%로 한국어의 65.1%보

다는 다소 높은 값을 보였다. 개방과열의 운동량만을 이용한 판별분석에 비해서 스펙트럼을 추가한 경우 영어의 예측 정확도는 68.4%에서 72.5%로 증가하는데 특히 [t]의 정확도 증가가 두드러진다. Blumstein & Stevens(1979)의 85%에 비해서는 낮지만, 그들의 연구는 주어진 음절만을 인위적으로 녹음한 자료이므로, 본 연구의 결과와 단순 비교는 불가하다. 한국어의 경우에도 스펙트럼 정보가 추가된 경우 예측 정확도가 62.8%에서 65.1%로 증가하는데, 영어와는 달리 [k]의 정확도 증가가 상대적으로 크다. 영어는 [k>t>p]의 순서로, 한국어도 [t^h>k^h>p^h]의 순서로 이전의 경우와 마찬가지로 예측 정확도가 높았다.

긴 개방과열, 즉, 개방과열에서 모음으로 이어지는 길이가 40 ms 이상 되는 경우, 개방과열 후속 기식음이 과열음의 조음위치 결정에 도움이 될 수 있으므로, 이들의 기여도를 알아보기 위하여 추가적으로 개방과열 후속 기식음의 운동량과 스펙트럼 양상을 데이터에 추가하여 판별분석을 시행하였고, 그 결과를 표 7에 나타내었다. 전술한 요인으로 보면 (1)번과 (2)번을 합친 것이다.

표 7. (1-2) 긴 개방과열의 운동량과 스펙트럼에 대한 판별분석
Table 7. (1-2) Discriminant analyses on the moments & spectra of long bursts

영어: 72.6%			예측값		
			p	t	k
관측값	64.2%	p	4,168	629	1,691
	73.7%	t	1,108	10,025	2,464
	75.9%	k	868	1,911	8,762
한국어: 65.0%			예측값		
			p ^h	t ^h	k ^h
관측값	42.0%	p ^h	1,621	1,234	1,007
	77.4%	t ^h	400	6,697	1,550
	62.0%	k ^h	298	2,127	3,950

표 6과 표 7을 비교해 보면, 개방과열 후속 기식음의 여부는 예측 정확도에 거의 영향을 미치지 못한 것을 알 수 있다. 정확도 차이가 영어와 한국어 모두 0.1% 포인트에 불과한 것을 알 수 있다. 따라서 개방과열 후 기식음이 긴 경우 여기에 담겨있는 정보는 무성과열음의 조음위치 결정에 영향을 거의 미치지 못하는 것으로 볼 수 있다.

과열음 직후 존재하는 모음의 시작 부분 운동량과 스펙트럼이 무성과열음의 예측 정확도에 미치는 영향을 알아보기 위하여 모음 시작 20 ms 부분의 운동량 및 스펙트럼 양상을 데이터에 추가하여 판별분석을 시행하였고, 그 결과를 표 8에 나타내었다. 전술한 요인 (1), (2), (3)번 모두를 합친 것에 해당된다.

예측 정확도는 영어 73.8%, 한국어 66.1%로 다소 증가한 양상을 보였고, 개별 과열음의 정확도도 이전과 동일하게 영어는 [k]가 한국어는 [t^h]가 제일 높았다.

표 8. (1-3) 개방과열과 직후 모음의 운동량 및 스펙트럼에 대한 판별분석

Table 8. (1-3) Discriminant analyses on the moments & spectra of the bursts & following vowels

영어: 73.8%			예측값		
			p	t	k
관측값	65.5%	p	4,249	631	1,608
	75.4%	t	1,090	10,246	2,261
	76.6%	k	882	1,815	8,844
한국어: 66.1%			예측값		
			p ^h	t ^h	k ^h
관측값	41.9%	p ^h	1,617	1,260	985
	78.8%	t ^h	403	6,812	1,432
	63.5%	k ^h	305	2,019	4,051

3.3.2. 폐쇄구간 길이와 VOT 요인 추가

이번에는 폐쇄구간 길이와 VOT 요인을 데이터에 추가하여 판별분석을 시행하였고 그 결과를 표 9에 나타내었다. 전술한 요인 중 (4)번과 (5)번이 추가되어 (1)~(5)번까지의 요인이 분석 대상이 된 것이다.

표 9. (1-5) 폐쇄구간과 VOT 추가 후 판별분석

Table 9. (1-5) Discriminant analyses after adding closure durations and VOTs

영어: 74.6%			예측값		
			p	t	k
관측값	67.7%	p	4,390	625	1,473
	75.4%	t	1,016	10,254	2,327
	77.4%	k	797	1,809	8,935
한국어: 66.4%			예측값		
			p ^h	t ^h	k ^h
관측값	41.9%	p ^h	1,618	1,278	966
	78.7%	t ^h	408	6,805	1,434
	64.4%	k ^h	304	1,963	4,108

표 8과 표 9를 비교해 보면 영어의 경우 이전에 비해 0.8% 포인트의 정확도 증가를 보였으나, 한국어는 0.3% 포인트의 증가에 그쳐, 각각 74.6%와 66.4%의 예측 정확도를 보였다. 개별 파열음의 정확도 순서도 변함이 없어서, 폐쇄구간 길이와 VOT 정보는 전체적인 예측 정확도 증가에 있어 미미하지만 영향을 미치는 것으로 나타났다.

3.3.3. 발화/단어 내 위치, 직후 모음 종류 요인 추가

마지막으로 발화 혹은 단어 내 위치와 파열음 직후 모음의 종류가 예측 정확도에 미치는 영향을 알아보기 위하여 이들을 데이터에 추가한 후 판별분석을 시행하였고 결과를 표 10에 나타내었다. 전술한 요인 (6), (7), (8)번이 모두 합쳐져 (1)~(8)번의 여덟 개 요인 모두가 판별분석에 사용한 셈이다.

표 9와 표 10을 비교해 보면, 요인이 추가되었음에도 불구하고 예측 정확도는 영어의 경우 오히려 74.6%에서 70.2%로 감소하였다. 한국어의 경우는 변화 없이 66.4%를 유지했다. 영어 개별 파열음의 정확도를 살펴보면, [p, k]의 정확도가 비교적 크게 증가하였으나 [t]의 정확도는 거의 20% 포인트 이상 감소한 것

을 볼 수 있다.

표 10. (1-8) 발화/단어 내 위치와 직후 모음 종류 추가 후 판별분석

Table 10. (1-8) Discriminant analyses after adding locations in utterance/word and following vowel identities

영어: 70.2%			예측값		
			p	t	k
관측값	75.5%	p	4,898	134	1,456
	54.6%	t	1,646	7,419	4,532
	85.8%	k	1,172	469	9,900
한국어: 66.4%			예측값		
			p ^h	t ^h	k ^h
관측값	44.8%	p ^h	1,729	1,172	961
	77.4%	t ^h	480	6,693	1,474
	64.6%	k ^h	351	1,905	4,119

영어와 한국어의 무성 파열음 조음위치에 영향을 미치는 요인을 점진적으로 추가하여 살펴본 결과, 요인에 따라 증감의 정도가 다른 것을 알 수 있었다. 따라서, 증가폭에 그다지 영향을 미치지 않거나 감소를 야기시키는 개방과열 후 기식음과 직후 모음 종류, 발화/단어 내 위치를 빼고, (1) 개방과열과 (4) 폐쇄구간, (5) VOT만을 가지고 추가적으로 판별분석을 시행하였고, 그 결과를 표 11에 나타내었다.

표 11. (1, 4, 5) 개방과열과 폐쇄구간, VOT 요인의 판별분석

Table 11. (1, 4, 5) Discriminant analyses on bursts, closure durations and VOTs

영어: 73.3%			예측값		
			p	t	k
관측값	66.5%	p	4,317	613	1,558
	73.7%	t	1,056	10,023	2,518
	76.6%	k	786	1,912	8,843
한국어: 65.4%			예측값		
			p ^h	t ^h	k ^h
관측값	41.7%	p ^h	1,609	1,244	1,009
	77.5%	t ^h	403	6,701	1,543
	63.4%	k ^h	261	2,072	4,042

모든 요인을 다 포함했을 경우(표 10 참조)의 예측 정확도와 비교해 보면, 영어의 경우는 정확도가 70.2%에서 73.3%로 크게 증가한 반면, 한국어는 66.4%에서 65.4%로 다소 감소한 것을 볼 수 있다. 개방과열 후 기식음에 관해서는(표 7 참조) 영어와 한국어 모두 해당 요인이 없어도 정확도는 영어는 0.7% 포인트, 한국어는 0.4% 포인트 가량 다소 증가하는 것으로 나타났다. 직후 모음의 운동량 및 스펙트럼 요인(표 8 참고)을 비교해 보면, 이 요인이 제거됨으로 인해서 정확도는 영어의 경우 0.5% 포인트, 한국어는 0.7% 포인트 감소된 것을 볼 수 있다. 따라서 이 요인은 정확도를 다소 증가시키는 것으로 보인다. 폐쇄구간과 VOT 요인이 추가된 표 9와 비교해 보면, 여기에는 정확도를 다소 증가시키는 요인인 직후 모음 운동량 및 스펙트럼 요인이 포함되어 있으므로 표 11의 경우보다 영어는 1.3% 포인트, 한국어는 1.0% 포인트 더 정확도가 높다.

정확도를 다소 증가시키는 것으로 나타난 (3) 직후 모음 포면

트 전이구간의 운동량 및 스펙트럼 요인을 추가하여 (1) 개방과열과 (4) 폐쇄구간, (5) VOT의 네 가지 요인만을 가지고 판별분석을 시행하여 그 결과를 표 12에 나타내었다. 결과를 보면 표 11의 경우에 비해 영어의 정확도는 74.5%로 1.2% 포인트, 한국어는 66.3%로 0.9% 포인트 증가한 것을 알 수 있다. 결국, 사용 가능한 여덟 가지 요인들 중에서 예측 정확도의 증가에 기여하는 네 가지 요인만을 사용하여도 최고의 경우와 단지 0.1% 포인트 차이가 나는 예측 정확도를 얻을 수 있는 것으로 나타난 것이다.

표 12. (1, 3, 4, 5) 개방과열, 폐쇄구간, VOT, 직후 모음 운동량 및 스펙트럼 요인의 판별분석

Table 12. (1, 3, 4, 5) Discriminant analyses on bursts, closure durations, VOTs and following vowel moments

영어: 74.5%			예측값		
			p	t	k
관측값	67.7%	p	4,393	629	1,466
	75.2%	t	1,028	10,230	2,339
	77.4%	k	806	1,800	8,935
한국어: 66.3%			예측값		
			p ^h	t ^h	k ^h
관측값	41.7%	p ^h	1,609	1,298	955
	78.5%	t ^h	408	6,789	1,450
	64.5%	k ^h	292	1,970	4,113

폐쇄구간과 VOT는 합하게 되면 과열음 자체의 길이를 의미하기도 한다. 폐쇄구간을 빼고 판별분석을 시행해보면 예측 정확도는 영어의 경우 73.9%, 한국어의 경우 66.0%로 다소 감소하는 것을 볼 수 있다. 이는 VOT 정보는 남아있지만, 폐쇄구간 정보의 제거로 인해 과열음 길이 정보가 사라져서 정확도가 감소하는 것으로 추측할 수 있다. 따라서 정도는 크지 않으나 폐쇄구간 길이 요인도 정확도 증가에 기여하는 것으로 볼 수 있을 것이다.

4. 결론

본 논문에서는 영어와 한국어 자연발화 음성 코퍼스에서 각각 무성 과열음 [p, t, k]와 [p^h, t^h, k^h]를 추출하고, 이들의 조음위치 결정에 영향을 미칠 것으로 예상되는 여덟 가지의 요인들 즉, (1) 개방과열의 운동량/스펙트럼, (2) 개방과열 나머지 부분의 운동량/스펙트럼, (3) 직후 모음의 운동량/스펙트럼, (4) 폐쇄구간 길이, (5) VOT, (6) 발화 내 위치, (7) 단어 내 위치, (8) 직후 모음의 종류 등을 대상으로 판별분석을 시행하여 예측 정확도를 분석하였다. 분석 결과를 나타낸 표들을 요인 번호를 중심으로 막대그래프로 나타내면 그림 4와 같다.

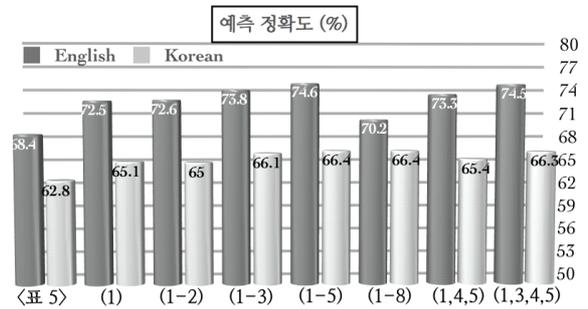


그림 4. 영어와 한국어 무성 과열음의 예측 정확도
Figure 4. Predicted percent correctness of English and Korean voiceless plosives

요인의 수가 (1)번에서 (5)번으로 다섯 개로 증가함에 따라 영어는 74.6%, 한국어는 66.4%까지 두 언어 모두 예측 정확도도 증가하는 양상을 보였으나, (6), (7), (8)번 요인 등 세 개의 요인이 추가되면 한국어는 변화가 없었으나 영어의 예측 정확도는 오히려 70.2%로 감소하는 것으로 나타났다. 개별 요인 분석을 통해 다섯 개의 요인이 아닌 네 개의 요인만을 이용해도 영어는 74.5%, 한국어는 66.3%로 최고 예측 정확도보다 단지 0.1% 포인트 적은 값을 달성할 수 있음도 확인하였다.

결과적으로 두 언어 모두 (1) 개방과열, (3) 직후 모음, (4) 폐쇄구간 길이 (5) VOT가 무성 과열음의 예측 정확도에 주요한 기여를 하는 것으로 나타났고, 여기에 (2) 개방과열 나머지 부분이 추가되면 최고 예측 정확도인 74.6%(영어), 66.4%(한국어)를 달성할 수 있음을 알게 되었다. 이는 같은 코퍼스를 이용한 Hwang & Yoon(2017)에서 보고한 영어의 66%, 한국어의 63% 예측 정확도에 비해 증가한 값으로, 개방과열의 정보 이외에 추가적으로 사용한 요인들이 이러한 증가에 기여한 것으로 볼 수 있다. 결국 영어와 한국어의 무성 과열음 결정에 주요한 영향을 미치는 정보는 폐쇄구간과 개방과열 주변에 존재하는 무성 과열음 자체의 요인들뿐 아니라 직후 모음의 시작 부위에도 존재한다는 것을 암시한다.

선행 연구 중에서 대화체 영어 자료를 이용한 Winitz et al. (1972)에서 얻어낸 58%의 정확도보다는 이번 연구의 74.6%가 월등히 높은 편이지만 여전히 완벽한 예측 정확도를 보이지는 못하고 있다. 특히, 한국어 예측 정확도는 영어에 비해 매우 낮다. 이는 영어와 한국어의 차이에 기인할 뿐 아니라, 본 연구에서 분석한 요인들 이외에도 무성 과열음의 조음위치 결정에 영향을 미치는 다른 요인들이 있음을 암시하는 것이다. 예를 들어, 주어진 자모 분절음들의 완벽하지는 않지만 부분적인 정보들이 합쳐져 특정 단어를 형성할 확률이나 가능성의 정도가 화자의 두뇌에서 예측될 수도 있을 것이다. 이러한 가능성은 후속 과제의 일부로 남겨둔다.

References

Alwan, A. (1989). Perceptual cues for place of articulation for the voiced pharyngeal and uvular consonants. *The Journal of the*

- Acoustical Society of America*, 86(2), 549-556.
- Auzou, P., Ozsancak, C., Morris, R. J., Jan, M., Eustache, F., & Hannequin, D. (2000). Voice onset time in aphasia, apraxia of speech and dysarthria: A review. *Clinical Linguistics and Phonetics*, 14(2), 131-150.
- Blumstein, S. E., & Stevens, K. N. (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(4), 1001-1017.
- Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341-345.
- Byrd, D. (1993). 54,000 American stops. *UCLA Working Papers in Phonetics*, 83, 97-115.
- Bonneau, A., Djeddar, L., & Laprie, Y. (1996). Perception of the place of articulation of French stop bursts. *The Journal of the Acoustical Society of America*, 100(1), 555-564.
- Crystal T. H., & House, A. S. (1988). The duration of American-English stop consonants: An overview. *Journal of Phonetics*, 16(3), 285-294.
- Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *The Journal of the Acoustical Society of America*, 27(4), 769-773.
- Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America*, 84(1), 115-123.
- Halle, M., Hughes, G. W., & Radley, J.-P. A. (1957). Acoustic properties of stop consonants. *The Journal of the Acoustical Society of America*, 29, 107-116.
- Hwang, S., & Yoon, K. (2017). A study on the release burst spectra of the voiceless plosives from the English and Korean spontaneous speech corpus. *Phonetics and Speech Sciences*, 9(4), 27-34.
- Kent, R. D., & Read, C. (2002). *The acoustic analysis of speech* (2nd ed.). Albany, NY: Singular Thomson Learning.
- Lee, Y., & Yoon, K. (2016). A study on the voice onset times of the Seoul corpus males in their twenties. *Phonetics and Speech Sciences*, 8(4), 1-8.
- Pae, J., Shin, J., & Ko, D. H. (1999). Some acoustical aspects of Korean stops in various utterance positions: Focusing on their temporal characteristics. *Korean Journal of Speech Sciences*, 5(2), 139-159.
- Pitt, M. A., Dille, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). Buckeye Corpus of Conversational Speech (2nd release). Columbus, OH: Department of Psychology, Ohio State University. Retrieved from <http://www.buckeyecorpus.osu.edu>
- RStudio Team. (2015). RStudio: Integrated development for R [Computer software]. Boston, MA: RStudio. Retrieved from <http://www.rstudio.com/> on March 31, 2016.
- Shin, J. (1997). *Consonantal production and coarticulation in Korean* (Doctoral dissertation). University of London, London, UK.
- 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(6), 3582-3864.
- Stevens, K. N., & Blumstein, S. E. (1975). Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *Journal of Phonetics*, 3(4), 215-233.
- Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, 64(5), 1358-1368.
- Winitz, H., Scheib, M. E., & Reeds, J. A. (1972). Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech. *The Journal of the Acoustical Society of America*, 51, 1309-1317.
- Yao, Y. (2007). Closure duration and VOT of word-initial voiceless plosives in English spontaneous connected speech. *UC Berkeley Phonology Lab Annual Report*, 2007, 183-225.
- 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.
- Zue, V. W. (1976). *Acoustic characteristics of stop consonants: A controlled study* (Doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

• 윤규철 (Kyuchul Yoon) 교신저자

영남대학교 영어영문학과
경상북도 경산시 대학로 214
영남대학교
Tel: 053-810-2145
Email: kyoony@ynu.ac.kr
관심분야: 음성학

영어와 한국어 자연발화 음성 코퍼스에서의 무성 파열음 연구*

윤 규 철

영남대학교 영어영문학과

국문초록

본 논문의 목적은 자연발화 음성 코퍼스를 대상으로 영어 무성 파열음 [p, t, k]과 한국어 격음 파열음 [pʰ, tʰ, kʰ]의 조음위치 결정에 영향을 미치는 요인들을 살펴보는 것이다. 프랏 스크립트를 이용하여 요인들은 자동 추출하였고, 판별분석을 통해 요인의 수를 점차 증가시켜가면서 무성 파열음의 예측 정확도를 계산하였다. 분석에 사용된 요인들은 개방파열, 파열 후 기식음과 모음 시작 부분의 운동량과 스펙트럼 기울기, 폐쇄구간과 VOT, 단어와 발화 내 위치, 마지막으로 직후 모음의 종류 등이었다. 분석 결과에 따르면, 요인의 수가 다섯 개까지 증가하는 경우 예측 정확도가 최대로 증가하여 영어는 74.6%, 한국어는 66.4%를 나타내었다. 그러나 사실상의 최대값에 도달하는 데는 네 개의 요인으로도 충분하였고, 이들은 개방파열과 직후 모음의 운동량과 스펙트럼 기울기, 폐쇄구간과 VOT였다. 이는 무성파열음의 조음위치가 자신의 내부 요인들과 직후 모음의 영향을 동시에 받는다는 것을 의미한다고 볼 수 있다.

핵심어: 서울 코퍼스, 벅아이 코퍼스, 한국어, 영어, 파열음, 모먼트, 스펙트럼 기울기, 판별분석

참고문헌

배재연, 신지영, 고도홍 (1999). 음성 환경에 따른 한국어 폐쇄음의 음향적 특성: 시간적 특성을 중심으로. *음성과학*, 5(2), 139-159.

* 이 연구는 2019년도 영남대학교 학술연구조성비에 의한 것임.