



pISSN 2005-8063  
eISSN 2586-5854  
2018. 06. 30.  
Vol.10 No.2  
pp. 7-14

# 말소리와 음성과학

Phonetics and Speech Sciences

한국음성학회지

<https://doi.org/10.13064/KSSS.2018.10.2.007>



## 성대진동 및 성별이 미국영어 마찰음에 미치는 효과에 관한 코퍼스 기반 연구\*

A corpus-based study on the effects of voicing and gender on American English Fricatives

윤 태 진\*\*

Yoon, Tae-Jin

### Abstract

The paper investigates the acoustic characteristics of English fricatives in the TIMIT corpus, with a special focus on the role of voicing in rendering fricatives in American English. The TIMIT database includes 630 talkers and 2,342 different sentences, and comprises more than five hours of speech. Acoustic analyses are conducted in the domain of spectral and temporal properties by treating gender, voicing, and place of articulation as independent factors. The results of the acoustic analyses revealed that acoustic signals interact in a complex way to signal the gender, place, and voicing of fricatives. Classification experiments using a multiclass support vector machine (SVM) revealed that 78.7% of fricatives are correctly classified. The majority of errors stem from the misclassification of /θ/ as [f] and /ʒ/ as [z]. The average accuracy of gender classification is 78.7%. Most errors result from the classification of female speakers as male speakers. The paper contributes to the understanding of the effects of voicing and gender on fricatives in a large-scale speech corpus.

**Keywords:** acoustic analysis, fricatives, gender, SVM, TIMIT corpus, voicing

### 1. 서론

마찰음은 구강 내에서 아주 좁은 협착으로 인한 마찰 소음을 수반하여 만들어지는 분절음이다. 구강 내의 좁은 틈을 통해 빠르게 기류가 흘러나옴으로 인해 난기류가 생기는데, 이러한 난기류가 음원(source of sound)의 역할을 한다(Stevens, 1971; Jongman *et al.*, 2000; Wilde, 2005; Johnson, 2012). 영어의 마찰음은 일반적으로 조음 위치에 따라 순치음 /f, v/, 치간음 /θ, ð/, 치경음 /s, z/, 그리고 후치경음 /ʃ, ʒ/의 4종류로 분류된다. /h/는 인접한 모음과 조음 위치를 공유하는 무성음으로 분류되어 음향 분석에

서 종종 제외된다(Ladefoged, 1982).

마찰음은 조음 위치뿐만 아니라, 성대 진동을 동시에 수반하여 발화될 수 있다. 성대 진동을 수반한 유성 마찰음이 가지는 효과는 마찰음을 발화하는 동안 기류의 속도가 낮아지고, 주기적인 성대의 진동으로 인한 기류의 변조로 인해, 마찰 소음의 전반적인 강도를 약화되는 효과를 가지고 온다고 알려져 있다(Jongman *et al.*, 2000).

이외에도 유성 마찰음과 무성 마찰음 사이에는 입술에서 협착 지점까지의 거리에도 차이가 있다는 실험 결과가 있다. MRI를 이용한 마찰음 연구를 Narayanan *et al.*(1994)에서 4명의 화자

\* 이 논문은 2017년도 성신여자대학교 학술연구조성비 지원에 의하여 연구되었음.

\*\* 성신여자대학교, tyoon@sungshin.ac.kr

Received 24 April 2018; Revised 24 May 2018; Accepted 24 May 2018

© Copyright 2018 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.

를 이용하여 하였는데, 이 MRI 연구에서 혀의 모양뿐만 아니라 성도의 길이와 넓이를 측정하였다. 전반적으로 화자별 차이는 있었지만, 유성 마찰음이 무성 마찰음보다 입술에서 협착지점까지의 거리가 좀 더 긴 것으로 관찰되었다. 예를 들어, 한 화자의 경우 그 거리가 /f/는 0.55 cm, /v/는 0.66 cm, /θ/는 0.89 cm, /ð/는 1.01 cm, /s/는 1.4 cm, /z/는 1.75 cm, /f/는 2.08 cm, 그리고 /z/는 2.24 cm로 측정되었다. 또한 인두강의 부피도 유성 마찰음이 무성 마찰음에 비해 더 큰 것으로 관찰되었는데, 이는 아마도 설근(tongue root)이 전진되어 배치되었기 때문으로 여겨진다(Narayanan *et al.*, 1994, 1995).

이러한 조음 위치와 성대 진동(voicing)에서 관찰되는 생리학적이고 공기역학적인 특징들이 마찰음에 대한 음향 연구에서도 반영된다는 선행연구들이 있다. 선행연구들은 무성 마찰음이 대응하는 유성 마찰음에 비해 스펙트럼의 평균 및 꼭지 값(spectral mean and peak values)이 높고, 스펙트럼 상의 주파수 분산(variance)이 덜하며 꼭지도 명확히 분간할 수 있으며, 음수의 왜도(negative skewness)와 보다 긴 길이(duration), 그리고 전반적으로 더 큰 강도(amplitude)를 가지고 있다고 한다(Maniwa *et al.*, 2009).

유성 마찰음의 성대 진동 타이밍은 다소 복잡한 것으로 알려져 있다(Docherty, 1992). 어두와 어말의 유성 마찰음을 청각에 의지해 분석을 해 보면, 이 유성 마찰음들이 협착이 진행되는 동안 내내 성대 진동을 수반하는 것은 아니다. 그리고 몇몇 경우에 있어서는 전혀 성대진동이 없는 경우도 발견된다(Docherty, 1992; Haggard, 1978). 영어 유성 마찰음이 문맥에 따라 영향을 받는다는 또 다른 청각에 기반한 연구는 Haggard(1978)에서 찾아 볼 수 있다. Haggard(1978)는 어두, 어말, 그리고 어중에서 관찰되는 유성 마찰음을 조사했는데, 이 마찰음들이 강세 모음과 비강세 모음 앞에서, 그리고 유성 혹은 무성 폐쇄음과 인접해서 나타나는 환경에서 살펴보았다. 연구 결과는 어떠한 환경에서도 마찰음의 협착 동안 성대 진동이 지속되는 토큰이 없다는 것이었다. 무성 폐쇄음 앞에서 유성 마찰음이 나타나는 경우는 이 유성 마찰음은 언제나 무성으로 실현되었으며, 비강세 모음 앞의 모음 간에 /v/가 나타날 경우에도 8% 정도는 무성으로 실현된다는 것을 밝혔다.

마찰음에 관한 여러 음향 연구에도 불구하고, 성대의 진동과 관련된 음향적 자질이 사용된 선행 연구는 찾기 힘들다. 비록 유성 마찰음이 무성음화되는 경향이 강하다는 연구도 실험자의 청각에 의존하는 경향이 강하며, 유성 마찰음과 무성 마찰음의 음향적 자질은 길이, 강도 혹은 스펙트럼 상의 평균값 혹은 꼭지 값에 의존하고 있다. 또한 선행 연구 중 다양한 문장 내에서 나타나는 마찰음에 대한 연구를 한 것은 극히 드물다. 예를 들어, 마찰음과 관련된 선행 연구에서 많이 인용되는 연구는 Jongman *et al.*(2000)을 들 수 있다. 이 논문에서는 20명의 코넬 대학 학부생들을 대상으로 8개의 유·무성 마찰음이 V+p (V는 /i, e, æ, ʌ, o, u/ 중 하나)의 연쇄 앞에 위치한 단음절 단어를 “Say \_\_\_ again”라는 틀로 된 문장에서 3번씩 반복하게 하였으며, 이렇게 해서 수집된 토큰들을 가지고 마찰음들의 음향적인 특징을 연

구하였다.

본 논문은 600명 이상의 화자가 발화한 2,000개 이상의 문장을 담고 있는 TIMIT 코퍼스를 사용하여 마찰음에 대한 음향적 연구를 진행하고자 한다. 본 연구에는 특히 성대의 진동에 대한 정량적인 측정 방법을 사용하여 마찰 소음 구간 내에서의 성대 진동 여부 및 정도를 측정하는 방법을 사용하여 모음 앞에서 나타나는 마찰음의 음향 분석을 시도하고자 한다. 이를 통해 마찰음의 조음 위치와 성대 진동에 대한 음향 연구를 확대하여, 이러한 성대 진동 자질이 조음 위치와 성별과 관련하여 어떠한 관계를 살펴보고자 한다. 또한 여러 음향 자질들이 마찰음의 분류에 어떻게 기여하는지를 기계 학습 중 널리 사용되는 SVM(support vector machine)을 사용하여 시험해 보고자 한다.

## 2. 연구 방법

### 2.1. TIMIT 코퍼스

본 연구에서는 TIMIT 코퍼스를 사용하여, 마찰음의 음향 자질을 분석한다. TIMIT 코퍼스는 MIT 대학과 텍사스 인스트루먼트(Texas Instruments), 그리고 Stanford 대학교의 SRI International 이 자동 음성 인식 시스템 및 음향 음성에 대한 기초 연구를 위해 1980년대 후반에 개발한 코퍼스이다(Byrd, 1994). 이 코퍼스는 발화자의 수가 많고, 이미 단어 및 분절음 단위의 레이블링이 되어 있는 관계로, 음향 음성 연구를 하는데 상당한 이점을 가지고 있으며, 자동 음성 인식의 개발에는 거의 표준적인 코퍼스로 자리 잡고 있다. 반면 개발된 지 30년이 되었지만 이 코퍼스를 사용하여 음향 음성을 분석한 음성학 논문은 상대적으로 많지 않은 실정이다. TIMIT 코퍼스에는 630명 이상의 화자가 참여하여 2,342개의 문장들을 발화한 총 5시간 이상의 음성파일과 분절음 및 단어 정보들이 포함되어 있다. 문장의 종류에 대한 보다 상세한 설명은 Byrd(1994), Yoon(2015), Zue & Seneff(1988)에 기술되어 있으므로, 여기서는 그 설명을 생략하기로 한다. <그림 1>은 TIMIT 코퍼스의 한 문장을 Praat으로 시각화한 것을 예로 보여주고 있다.

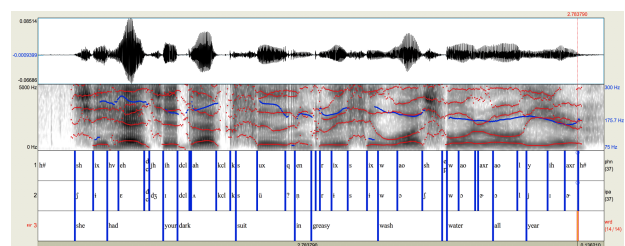


그림 1. 예시문 'She had your dark suit in greasy wash water all year'  
Figure 1. An example of a sentence 'She had your dark suit in greasy wash water all year'

TIMIT은 총 8개의 방언 지역에 있는 남녀 화자들의 발화를 담고 있다. 8개의 방언 지역은 1) New England, 2) Northern, 3) North Midland, 4) South Midland, 5) Southern, 6) New York City, 7)

Western, 8) Army Brat (moved around)을 지칭한다. 참고로 8) Army Brat은 지역적 특성이 일정하지 않고 여러 곳을 돌아다니는 군인들이 참여하였다는 점과 단지 남녀 총 33명만 참여하였다는 점으로 인해 본 논문에서는 이들 화자들의 음성파일에 대한 분석을 제외하였다. <표 1>은 유·무성 마찰음의 음향 분석을 위하여, TIMIT 코퍼스에서 무성 마찰음 /f/, /θ/(=th), /s/, /ʃ/(=sh)와 유성 마찰음 /v/, /ð/(=dh), /z/, /ʒ/(=zh)를 추출한 토큰 수를 남성과 여성으로 구분하여 제시하였다.

표 1. 성별에 따른 무성 마찰음의 분포  
Table 1. Distribution of fricative phones by gender

Variables	/f/	/v/	/θ/	/ð/	/s/	/z/	/ʃ/	/ʒ/
Male	912	716	285	1,850	2,014	952	839	63
Female	370	287	101	746	822	377	357	24
Total	1,282	1,003	386	2,596	2,836	1,329	1,196	87

<표 1>에서는 TIMIT 코퍼스의 단점 중 하나를 간접적으로 보여주고 있는데, 이는 성별에 따른 분포의 불균형과 음소에 따른 토큰 수의 불균형이라고 할 수 있다. 예를 들어, 본 논문에서 분석할 대상이 되는 남성 화자는 438명 중 8) Army Brats의 화자 22명을 제외한 416명이 있는 반면, 여성 화자는 192명 중 역시 8) Army Brats의 화자 11명을 제외한 181명만 있다. 또한 /s/가 2,836개로 가장 많은 토큰 수를 가지는 반면, /ʒ/는 남성 63개와 여성 24개 총 87개의 토큰 밖에 되지 않는다. 그 외 TIMIT 코퍼스의 특징으로 들 수 있는 것은 대부분 20대와 30대의 백인들이 녹음한 낭독형(read-style) 코퍼스라고 할 수 있다.

## 2.2. 음향 자질 추출

마찰음은 많은 음향 단서들(cues)을 가지고 있다(McMurray & Jongman, 2011). 본 연구에서는 선행 연구에서 밝힌 결과들을 토대로 하여 TIMIT 코퍼스의 음성 파일 중 모음에 선행하는 마찰음 토큰에서 다음의 음향 자질들을 추출하였다. 즉, (1) 마찰소음의 길이(duration), 스펙트럼상의 모멘트(Spectral moment) 중 (2) 무게 중심(Center of gravity), (3) 분산(Dispersion) (4) 왜도(Skewness), (5) 소음 강도(Noise intensity)를 추출하였다. 조음 위치는 네 가지의 스펙트럼 모멘트(스펙트럼 상의 주파수 값들의 평균, 변이, 왜도, 첨도)(Forrest *et al.*, 1988; Jongman *et al.*, 2000)에 의해서 구별될 수 있으며, 이 음향자질들은 유성 마찰음과 무성 마찰음을 구별하는데도 이용된다(Maniwa *et al.*, 2009). 마찰음의 길이와 강도는 조음 위치 및 성대 진동과 관련이 있다고 알려져 있다. 주로 치찰음을 비치찰음과 구별하는데 관련이 있다(Baum & Blumstein 1987; Behrens & Blumstein 1988a, 1988b; Crystal & House 1988; Jongman *et al.*, 2000). 또한 선행논문에서는 많이 활용되지 않지만, 성대 진동과 관련이 있다고 판단한 (6) 국부 무성 프레임 비율(fraction of locally unvoiced frames)과 (7) 고주파와 저주파수 간의 에너지 차(energy difference between high-frequency and low-frequency)와 같은 음향 자질을 추출하여 분석에 사용하였다.

Praat(Boersma & Weenink, 2015)의 스크립팅 기능을 이용해 위에서 언급한 자질들을 추출하여 결과물 텍스트 파일로 저장하였으며, 자질 값들이 저장된 텍스트 파일은 R(R Core Team, 2015)과 이 프로그램의 여러 패키지들(예, pastecs, ggplot2, arm, e1071, kernlab 등)을 이용하여 통계 분석과 시각화에 이용하였다. 그 외 선행연구에서는 언급이 되었지만, 본 논문에서는 다루지 않은 음향적 특성들도 있다. 예를 들면, 후행하는 모음 시작 지점이 F2 값도 마찰음의 유형에 따라 영향을 받는다는 연구결과(Wagner *et al.*, 2006; Jongman *et al.*, 2000)가 있지만, 본 논문에서는 다루지 않았다.

## 3. 결과

### 3.1. 마찰 소음의 길이(duration)

Jongman *et al.*(2000: 1255)에 따르면 마찰 소음의 길이는 치찰음과 비치찰음을 구분하는데 쓰일 수 있다고 한다. 이는 일반적으로 /s, ʃ/가 /f, θ/보다는 더 긴 마찰 소음구간을 가지고 있기 때문이다. Behrens and Blumstein(1988a)은 /s/와 /ʃ/ 사이에는 길이의 차이가 없는 반면, /θ/와 /f/를 비교하였을 때, /θ/가 /f/보다 더 짧은 경향이 있음을 보고한다. 또한 무성 마찰음이 대응하는 유성 마찰음에 비해 길이(duration)가 길다는 Maniwa *et al.*(2009)의 연구 결과도 있다. 다음의 <그림 2>는 모음 앞에서 관찰된 유·무성 마찰음의 평균 길이를 성별을 구별하여 보여주고 있다.

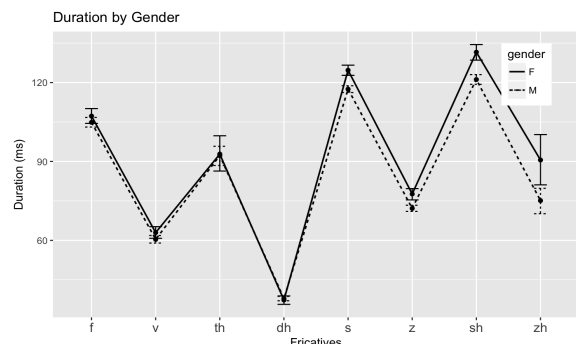


그림 2. 마찰 구간의 길이  
Figure 2. The duration of the fricative portion

Maniwa *et al.*(2009)의 연구 결과와 마찬가지로, 본 연구에서도 일반적으로 무성 마찰음에 해당하는 /f/, /θ/ (th), /s/, /ʃ/ (sh)가 유성 마찰음에 해당하는 /v/, /ð/ (=dh), /z/, /ʒ/ (=zh)보다 더 길게 발음되는 것을 <그림 2>를 통해 관찰할 수 있다. <그림 2>에서 유성 마찰음 /ð/가 다른 마찰음보다도 훨씬 더 짧게 발음되는 것을 관찰할 수 있는데, 이는 이 마찰음을 포함하는 단어들의 대다수가 *the, they, their* 등과 같은 빈도수가 높은 기능어(function words)로 구성되어 있기 때문으로 판단된다.

마찰 소음의 길이에 대한 이원분산(two-way ANOVA) 분석의 결과는 마찰음 유형[F(7, 10,699)=2,740.7,  $p<.001$ ,  $\eta^2=0.64$ ]과 성별[F(1, 10,699)=63.1,  $p<.001$ ,  $\eta^2=0.005$ ]에 따른 주효과(main effects)

가 있으며, 성별과 마찰음 유형 간의 상호작용효과[interaction effect;  $F(7, 10,699)=7.08, p<.001, \eta^2=0.004$ ]도 관찰된다. Tukey 사후 검증 시험에서 마찰음의 유형과 관련하여서는 /z/와 /ʒ/ 사이에서만 유의미한 차이를 보이지 않았으며, 다른 마찰음 쌍들은 성대 진동 유무 및 조음 위치 모두에서 유의미한 차이를 보였다. 성별을 고려한 경우는 비치찰음인 /f/, /v/, /θ/, /ð/의 경우에는 유의미한 차이를 보이지 않았지만, 치찰음인 경우에는 성별에 따른 유의미한 차이를 보이고 있다.

### 3.2. 스펙트럼상의 모멘트들(Moments)

스펙트럼상의 모멘트들은 마찰음들의 조음 위치를 구분하는데 유용하게 사용될 수 있다고 알려져 있다(Gordon *et al.*, 2002). 스펙트럼상 모멘트 값들 중 (1) 무게중심, (2) 분산 및 (3) 왜도의 분석 결과를 아래에 기술하였다.

#### 3.2.1. 무게 중심(Center of Gravity)

무게 중심은 스펙트럼에서 낮은 주파수대의 에너지 합이 높은 주파수대의 에너지 합과 같게 되는 주파수 값에 대응하는 값을 지칭한다. 따라서, 고주파 에너지를 많이 포함한 마찰음들은 무게 중심 값이 높게 설정될 것이다(Forrest *et al.*, 1998). 이러한 이유로 인해, 치찰음이 비치찰음보다 무게 중심의 값이 크게 나타난다. Jongman *et al.*(2000)은 스펙트럼상의 꼭지 값(Spectral peak)이 비치찰음의 경우에는 유무성의 차이를 밝히는데 통계적으로 유의미한 차이를 보이지만, 치찰음의 경우에는 스펙트럼상의 꼭지 값이 유의미한 차이를 보이지 않는다고 한다. 이러한 결과를 토대로 무게 중심도 비슷한 차이를 보일 것이라고 유추할 수 있다. 하지만, 본 연구에서는 치찰음과 비치찰음 모두 무게중심의 값이 유성 마찰음과 무성 마찰음을 구별하는데 사용되는 것으로 관찰된다. 다시 말해, 모든 조음 위치에서 마찰음이 유성일 경우가 무성일 때보다 무게 중심의 값이 낮아진다는 것을 <그림 3>을 통해서 관찰할 수 있다. 한편 성별에 대한 차이는 치찰음의 경우가 비치찰음의 경우보다 크다는 것을 <그림 3>을 통해서 관찰할 수 있다.

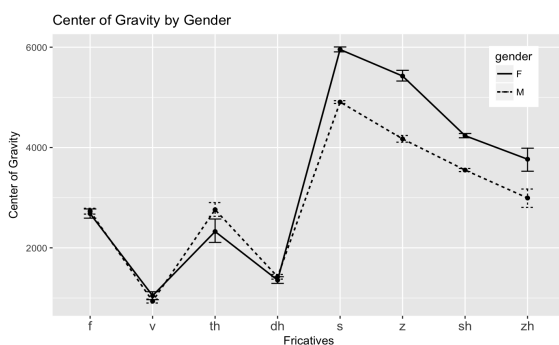


그림 3. 무게 중심  
Figure 3. Center of gravity

무게 중심에 대한 이원분산(two-way ANOVA) 분석의 결과는 마찰음 유형[ $F(7, 10,699)=5,812.7, p<.001, \eta^2=0.79$ ]과 성별[ $F(1,$

$10,699)=769.4, p<.001, \eta^2=0.06$ ]에 따른 주효과(main effects)가 있으며, 성별과 조음 위치간의 상호작용효과[interaction effect;  $F(7, 10,699)=142.4, p<.001, \eta^2=0.08$ ]도 관찰된다. Tukey 사후 검증 시험에서 마찰음의 유형과 관련하여서는 /θ/와 /f/ 사이에서만 유의미한 차이를 보이지 않았으며, 다른 마찰음 쌍들은 성대 진동 유무 및 조음 위치 모두에서 유의미한 차이를 보였다. /f/와 /v/, 그리고 /ð/의 경우 성별에 따른 유의미한 차이를 찾을 수 없었고, 다른 마찰음들은 성별에 따른 유의미한 차이를 보였다.

#### 3.2.2. 분산(Dispersion)

분산은 에너지가 무게 중심을 중심으로 좁은 주파수 대역에 몰려있는지, 아니면 넓은 범위의 주파수대에 퍼져 있는지를 밝혀주는 측정치이다. <그림 4>는 분산의 경우 치찰음과 비치찰음 사이에서 비대칭성을 관찰할 수 있다. 일반적으로 비치찰음의 경우 무성 마찰음이 유성 마찰음보다 분산의 정도가 큰 반면, 치찰음의 경우는 그 반대로 유성 마찰음이 무성 마찰음보다 분산의 정도가 다소 크다는 것을 관찰할 수 있다. 이러한 유·무성에 따른 비대칭성이 어떤 생리학적 혹은 공기역학적 이유로 나타나는지는 현재로서는 밝혀지지 않았다. 비치찰음의 경우는 채널난기류(channel turbulence)에 의해 만들어져서, 치찰음의 경우 장애난기류(obstacle turbulence)에 의해 만들어진다는 점을 고려해 볼 수 있을 것이다.

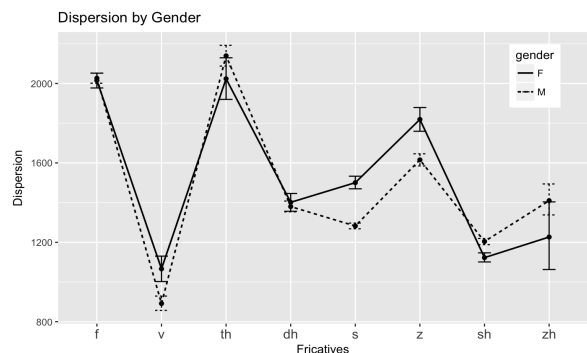


그림 4. 분산  
Figure 4. Dispersion

분산에 대한 이원분산(two-way ANOVA) 분석의 결과는 마찰음 유형[ $F(7, 10,699)=639.07, p<.001, \eta^2=0.29$ ]과 성별[ $F(1, 10,699)=73.7, p<.001, \eta^2=0.006$ ]에 따른 주효과(main effects)가 있으며, 성별과 조음 위치 간의 상호작용효과[interaction effect;  $F(7, 10,699)=19.3, p<.001, \eta^2=0.01$ ]도 관찰된다. Tukey 사후 검증 시험에서 마찰음의 유형과 관련하여서는 /ð/와 /ʒ/ 사이, 그리고 /ʒ/와 /s/ 사이에서만 유의미한 차이를 보이지 않았다. 다른 마찰음 쌍들은 성대 진동 유무 및 조음 위치 모두에서 유의미한 차이를 보였다. 성별에 따른 경우 /f/, /θ/, /ð/, /ʃ/, /ʒ/의 마찰음은 유의미한 차이를 보이지 않았고, /v/, /s/, /z/의 경우에만 성별이 유의미한 차이를 가져왔다.

### 3.2.3. 왜도(Skewness)

왜도는 스펙트럼상의 주파수대가 가지는 비대칭성의 정도를 일컫는 척도이다. 스펙트럼상의 평균을 중심으로 대칭적인 경우 왜도는 0의 값을 가지며, 왜도가 음수일 경우에는 주파수대의 에너지가 왼쪽 부분에 긴 꼬리를 가지며, 주파수대의 에너지가 평균값의 오른쪽 부분에 더 많이 분포해 있다. <그림 5>는 비치찰음의 유·무성 마찰음과 치찰음 중 후치경음은 왜도가 양수인 경우를 보이고 있으며, 이는 스펙트럼상의 주파수대의 에너지가 평균값보다 왼쪽(즉 낮은 주파수대)에 더 많이 분포해 있다는 것을 나타낸다. 치찰음 중 치경음의 경우는 반대로 음의 왜도를 가지는데, 이는 높은 주파수대인 평균값보다 왼쪽에 에너지가 더 많이 분포되어 있다는 것을 나타낸다. <그림 5>를 통해 관찰할 수 있듯이, 왜도는 유·무성에 따라서도 차이를 보이는데, 비치찰음의 경우 유성 마찰음들의 왜도 값이 더 높으며, 치찰음의 경우는 유무성에 따른 차이가 그렇게 크지 않다.

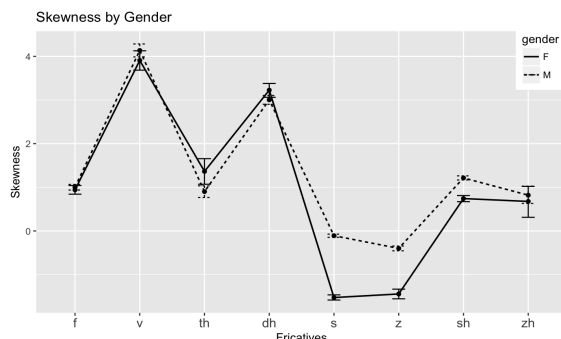


그림 5. 왜도  
Figure 5. Skewness

왜도에 대한 이원분산(two-way ANOVA) 분석의 결과는 마찰음 유형 [ $F(7, 10,699)=2,027.12, p<.001, \eta^2=0.57$ ]과 성별 [ $F(1, 10,699)=275.07, p<.001, \eta^2=0.02$ ]에 따른 주효과(main effects)가 있으며, 성별과 조음 위치 간의 상호작용효과[interaction effect;  $F(7, 10,699)=65.2, p<.001, \eta^2=0.04$ ]도 관찰된다. Tukey 사후 검증 시험에서 마찰음의 유형과 관련하여서는 /θ/와 /f/ 사이에서만 유의미한 차이를 보이지 않았으며, 다른 마찰음 쌍들은 성대 진동 유무 및 조음 위치 모두에서 유의미한 차이를 보였다. 성별의 경우 /f/, /v/, /θ/의 경우는 유의미한 차이를 보이지 않았다.

### 3.3. 마찰 구간의 강도(Intensity)

마찰음의 강도와 관련된 연구들은 무성 마찰음이 유성 마찰음보다 일반적으로 강도가 더 세다고 밝히고 있다. 무성 마찰음 사이에서도 강도의 차이는 있는데, 일반적으로 치찰음인 /s/와 /ʃ/가 비치찰음인 /f/와 /θ/보다 강도가 훨씬 높다는 것으로 알려져 있다(Behrens & Blumstein, 1988a, 1988b).

<그림 6>은 성대 진동의 유무와 상관없이 일반적으로 치찰음이 비치찰음보다 강도가 높다는 것을 보여주고 있다. 성대 진동의 유무를 고려하면, 치찰음과 비치찰음 사이에 비대칭적인 차이가 관찰된다. 치찰음의 경우 유성 마찰음보다는 무성 마찰음

이 강도가 더 강한 반면, 비치찰음의 경우 그 반대로 무성 마찰음이 유성 마찰음보다 강도가 더 센 것으로 보인다.

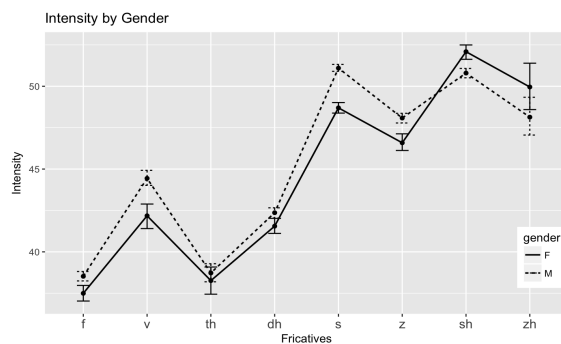


그림 6. 마찰 소음의 강도  
Figure 6. Intensity of the fricative noise

비치찰음과 비치찰음 내에서는 같은 그룹에 속하는 마찰음들이 서로 다르지는 않다고 하는데(Jongman et al., 2000: 1254), 동일한 유형이 TIMIT 코퍼스에서도 관찰되었다. 다시 말해, 치찰음은 비치찰음보다는 높은 강도를 가지며, 비치찰음 내의 두 마찰음 유형들은 강도를 통해서 더 이상 세분화되지는 않았다.

마찰음의 강도에 대한 이원분산(two-way ANOVA) 분석의 결과는 마찰음 유형 [ $F(7, 10,699)=1,168.43, p<.001, \eta^2=0.43$ ]과 성별 [ $F(1, 10,699)=112.5, p<.001, \eta^2=0.01$ ]에 따른 주효과(main effects)가 있으며, 성별과 조음 위치 간의 상호작용효과[interaction effect;  $F(7, 10,699)=14.7, p<.001, \eta^2=0.009$ ]도 관찰된다. Tukey 사후 검증 시험에서 마찰음의 유형과 관련하여서는 /θ/와 /f/, 그리고 /z/와 /ʒ/ 사이에서만 유의미한 차이를 보이지 않았으며, 다른 마찰음 쌍들은 성대 진동 유무 및 조음 위치 모두에서 유의미한 차이를 보였다. 설별의 경우 /f/, /θ/, /ʒ/의 경우는 유의미한 차이를 보이지 않았다.

### 3.4. 국부 무성 프레임 비율

성대 진동의 정량적인 자질 추출 방식으로 Praat의 Voice Report의 기능 중 하나인 “국부 무성 프레임 비율(fraction of locally unvoiced frames)”을 활용하였는데, 이를 통해 마찰음 구간에서의 무성 프레임 비율을 구하였다. <그림 7>은 마찰 구간 내에서 무성 프레임의 비율 값을 0과 1로 나타낸 그림이다. 마찰 구간이 무성이면 1을, 완전한 무성이면 0을 갖도록 계산하였다. <그림 7>에서 무성 마찰음들은 0.8에서 1 사이에 값들이 분포되어 있으며, 유성 마찰음들은 0.8 이하의 값을 가지고 있다. 또한 유성 마찰음 중 /v/와 /zh/의 경우, 성별에 따른 차이가 있다는 것을 관찰할 수 있다.

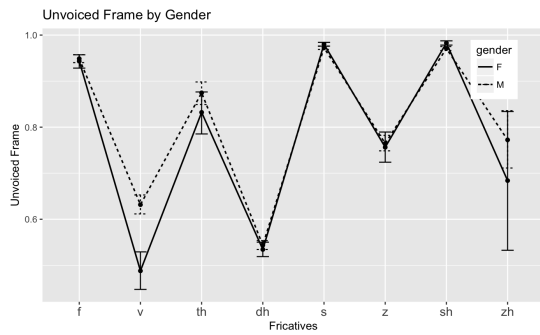


그림 7. 마찰 구간의 국부 무성 프레임 비율  
Figure 7. fraction of locally unvoiced frames for the fricative portion

무성 프레임 비율에 대한 이원분산(two-way ANOVA) 분석의 결과는 마찰음 유형 [ $F(7, 10,699)=1,444.89, p<.001, \eta^2=0.48$ ]과 성별 [ $F(1, 10,699)=15.39, p<.001, \eta^2=0.001$ ]에 따른 주효과(main effects)가 있으며, 성별과 마찰음 유형 간의 상호작용효과[interaction effect;  $F(7, 10,699)=15.82, p<.001, \eta^2=0.01$ ]도 관찰된다. Tukey 사후 검증 시험에서 마찰음의 유형과 관련하여서는 /s/와 /ʃ/ 사이, 그리고 /z/와 /ʒ/ 사이에서만 유의미한 차이를 보이지 않았으며, 다른 마찰음 쌍들은 성대 진동 유무 및 조음 위치 모두에서 유의미한 차이를 보였다. 성별까지 고려한 경우는 /v/는 유의미한 차이를 보인 반면, 다른 마찰음들은(/f/, /θ/, /ð/, /s/, /z/, /ʃ/, /ʒ/) 유의미한 차이를 보이지는 않았다.

### 3.5. 고주파수와 저주파수 간의 에너지 차이

본 논문에서 고주파수와 저주파수 간의 에너지 차이를 구하는 방식은 4,000 Hz를 기준으로 4,000 Hz부터 8,000 Hz까지의 주파수대의 에너지 총합에서 4,000 Hz 이하의 주파수대 에너지의 총합을 뺀 값을 지칭한다. 참고로 TIMIT 코퍼스의 표본 추출 주파수(sampling frequency)는 16,000 Hz이다. <그림 8>에서 관찰되듯이, 고주파수대 에너지에서 저주파수대 에너지를 뺀 에너지 차이(Energy Difference)의 값은 대체적으로 비치찰음의 경우 음수가, 치찰음의 경우 양수의 값을 가진다.

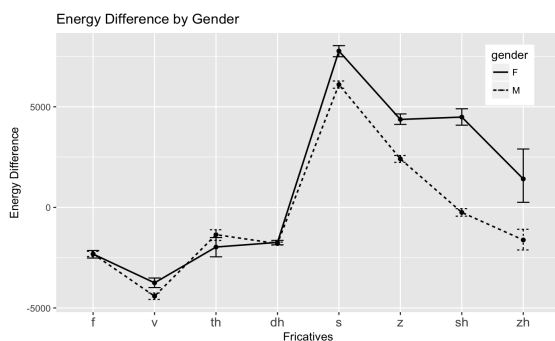


그림 8. 에너지 차  
Figure 8. The energy difference

이는 비치찰음의 경우, 저주파수대의 에너지 합이 고주파수대의 에너지합보다 더 크다는 것을 나타내며, 치찰음의 경우 특

히 /s/와 /z/의 대체적으로 고주파수대에 에너지가 많이 분포되어 있다는 것으로 해석된다. 남성의 경우, 후치경 마찰음의 경우 에너지 차이가 거의 없거나 음의 값을 보이는데, 이는 후치경 마찰음의 에너지 다소 낮은 주파수대에 집중해 있다는 것을 나타낸다.

에너지 차이에 대한 이원분산(two-way ANOVA) 분석의 결과는 마찰음 유형 [ $F(7, 10,699)=2,668.17, p<.001, \eta^2=0.63$ ]과 성별 [ $F(1, 10,699)=441.27, p<.001, \eta^2=0.03$ ]에 따른 주효과(main effects)가 있으며, 성별과 조음 위치 간의 상호작용효과[interaction effect;  $F(7, 10,699)=81.62, p<.001, \eta^2=0.05$ ]도 관찰된다. Tukey 사후 검증 시험에서 마찰음의 유형과 관련하여서는 /θ/와 /ð/ 사이, 그리고 /θ/와 /ʒ/ 사이에서만 유의미한 차이를 보이지 않은 반면, 다른 마찰음 쌍들은 성대 진동 유무 및 조음 위치 모두에서 유의미한 차이를 보였다. 성별과 관련하여 비치찰음들인 /f/, /v/, /θ/, /ð/는 유의미한 차이를 보이지 않은 반면, 치찰음들은 성별에 따른 유의미한 차이를 나타내고 있다.

## 4. 논의 및 결론

유·무성 마찰음의 음향적 특성에 관한 선행연구와 마찬가지로, 본 논문은 음향 자질에 초점을 맞추어 마찰음의 성별, 조음 위치 및 성대 진동의 유무에 따른 특징을 살펴보았다. 성별을 고려하지 않은 상태에서, 조음 위치와 성대 진동의 유무에 따라 대략적인 특징을 살펴보면 (1) 국부 무성 프레임 비율의 경우 유성 마찰음과 무성 마찰음 사이의 구분, (2) 고주파와 저주파수 간의 에너지 차의 경우는 치찰음과 비치찰음 사이의 구분, (3) 마찰 소음의 길이는 유성 마찰음과 무성 마찰음 사이의 구분, (4) 무게 중심은 유성 마찰음과 무성 마찰음의 구분 및 치찰음과 비치찰음 사이의 구분, (5) 분산의 경우는 유성과 무성 마찰음 사이의 구분을, (6) 왜도의 경우 치찰음과 비치찰음 사이의 구분 및 비치찰음의 경우 유성 마찰음과 무성 마찰음 사이의 구분, (7) 소음 강도의 경우 치찰음과 비치찰음 사이의 구분 및 유성 마찰음과 무성 마찰음 사이의 구분을 가져오는데 사용될 수 있을 것으로 보인다.

하지만 단편적인 음향 단서들이 통합되었을 때 마찰음 유형 간의 구분을 얼마나 잘 만들어낼 수 있는 지는 실험을 통해서 살펴볼 필요가 있다. 예를 들면 치찰음과 같은 경우 강도가 비치찰음에 비해 높게 나오는 반면(<그림 6>), 왜도와 같은 경우는 그 반대의 현상을 보여 비치찰음들이 상대적으로 치찰음들보다 높은 값을 가진다(<그림 5>).

본 논문에서는 SVM(support vector machine)을 사용하여 위의 7가지의 음향 자질들을 설명 변수로 하고, 마찰음의 종류를 독립 변수로 설정하여 다층 클래스 분류 시험(multi-class classification tests)을 했다. 이를 위해서 kernlab(Karatzoglou et al., 2004)이라는 R 패키지를 사용하였으며, 자료를 90%의 훈련 데이터와 10%의 실험 데이터로 분리하여 진행하였다. 본 논문에서 추출한 음향자질들은 정량화되어 있으나, 단위가 큰 자질도 있으며, 그렇지 않은 자질들도 있어서 R의 arm 패키지를 사용하여 표준화

(standardized)하여 사용하였다. arm 패키지에서 제공되는 표준화 공식은 정량 변수에서 평균을 뺀 후, 이를 2x 표준 편차로 나눈 것이다.

클래스가 여러 개인 문제를 해결할 때는 다층 클래스 SVM 알고리즘을 사용할 수 있다. 일반적으로 n개의 클래스가 있다면 1개와 나머지 (n-1)개의 클래스의 합집합을 비교하는 OVA(one versus All) 방식과 두 클래스의 짝의 경우의 수를 모두 구하는 OVO(one versus one) 방식이 있다. 본 논문에서는 kernlab 패키지에서 제공하는 OVO 방식에 따라 마찰음의 종류를 분류하는 실험을 하였다. 또한 다층 클래스를 설정하는 과정에서 (1) 모든 마찰음의 종류를 분류하는 시험과 (2) 성별을 분류하는 시험을 수행하였다.

본 연구에서는 우선 모든 마찰음의 유형을 예측하는 실험을 하였는데, 90%에 해당되는 9,644개의 토큰은 훈련 데이터로 사용되었고, 10%에 해당되는 1,071개의 토큰은 시험 데이터로 사용되었다. Table 2에 제시되어 있듯이, 이러한 방식을 통하여 78.7%의 평균적인 정확성(accuracy)을 달성하였다.

표 2. 마찰음 분류 결과(단위: %)  
Table 2. Classification results of fricative phones reported in percentage

Pred.	True classes						
	/f/	/v/	/θ/	/ð/	/s/	/z/	/ʃ/
/f/	<b>91.1</b>	2.0	22.2	0.0	1.4	0.0	0.7
/v/	5.6	<b>96.9</b>	8.3	4.8	0.0	1.3	0.0
/θ/	0.8	0.0	<b>47.2</b>	0.4	0.0	0.6	0.0
/ð/	0.0	1.03	16.6	<b>93.8</b>	0.0	1.3	0.0
/s/	0.8	0.0	0.0	0.0	<b>93.7</b>	13.6	13.1
/z/	1.6	0.0	2.7	0.8	4.0	<b>82.8</b>	0.0
/ʃ/	0.0	0.0	2.7	0.0	0.7	0.0	<b>86.1</b>
/ʒ/	0.0	0.0	0.0	0.0	0.0	0.0	<b>8.3</b>

<표 2>의 분류 결과를 보면 /θ/와 /ʒ/에서 대다수의 오류가 발행한다는 것을 알 수 있다. /θ/의 경우 정확하게 예측되는 비율은 47.2%이고, /f/로 22.2%, 그리고 유성음인 /ð/로 16.6%로 부정확하게 예측되었다. 이와 관련하여 McGuire & Babel(2012)에 따르면, 두 비치찰 마찰음 /θ/와 /f/의 청각적인 혼동성은 잘 알려진 사실이다(예, Miller & Nicely, 1955). 그리고 이 두 소리가 서로 대체되기도 하는 원인이기도 하다고 주장되었다(Labov et al., 1968: 93). /θ/는 공식적으로 그리고 통시적으로도 불안정한 소리이며, 이 소리와 매우 혼동되는 소리인 /f/와 자주 대체되고 했다. McGuire & Babel(2012)은 /θ/가 미국영어에서 /f/보다 더 변이가 심하며, 이러한 변이의 정도로 인해 미국영어에서 안정이 결여되어 있다고 주장한다.

/ʒ/의 경우는 단지 8.3%만 정확하게 예측되었고, /z/로 50.5%, 그리고 무성음인 /ʃ/로 25%가 잘못 예측되었다. 이와 관련하여서는 TIMIT 데이터에서 /ʒ/의 토큰 수가 다른 마찰음에 비해 절대적으로 부족한 것이 음성적으로 유사한 다른 마찰음으로 분류되는데 기여하였을 것이라고 판단한다. 마찰 구간의 길이를 보면 /z/와 /ʒ/ 사이에서만 유의미한 차이를 보이지 않았는데, 이러한 음향 자질의 유사성을 지니고 있음을 보여준다고 할 수

있다. 또한 TIMIT에서 /ʒ/를 포함한 단어를 보면 pleasure (9), measured (9), composure (7), visual (7), usually (4) 등이 빈도수가 많은 단어들로 나타나고 있다(여기서 괄호 안의 숫자는 빈도수를 나타낸다). 본 분류 시험에서는 음향 변수만을 설명 변수(explanatory variable)에 설정하고, 단어는 설명 변수에 포함하지 않았다. 단어를 설명변수에 포함하면 정확성은 88% 이상으로 높아지는데, 이는 /ʒ/를 직각하는데 순전한 음향적인 자질들로 한계가 있음을 보여주는 반증이라고 할 수 있다.

마찰음과 관련한 음성 자질을 사용하여 성별을 분류하려고 시도하던 연구들이 있다. 예를 들면, Whiteside(1998)는 3명의 청각 실험자에게 3명의 남성과 3명의 여성이 발화한 문장에서 추출한 30-100 ms에 해당하는 무성 마찰음 구간을 듣고 화자의 성별을 식별하는 실험을 실행하였다. 이 실험을 통해 지각 실험자들이 64.4%의 정확도를 가지고 화자의 성별을 식별할 수 있었다는 실험 결과를 보고하였다. 본 논문에서는 무성마찰음뿐만 아니라 유성 마찰음까지 포함하여, 본 논문에서 보고한 음향자질들이 성별을 구분하는데 어느 정도의 정확성을 가지는지 살펴해보았다. 무성 마찰음뿐만 아니라, 유성 마찰음까지로 확장하여 SVM을 적용한 본 실험에서는 77.59%의 평균 정확도를 달성하였다. 남성은 97.3%로 정확하게 예측되는 반면, 여성은 30.3%만 정확하게 분류되었다. 그리고 대부분의 오류는 여성을 남성으로 잘못 분류한 것에서 기인하는데, 추출된 음향자질들 중 많은 자질들이 치찰음들의 경우는 성별에 따른 구별이 현저하게 높지만, 비치찰음들의 경우는 성별에 따른 차이가 그다지 많지 않은 것이 성별에 따른 분류상의 오류라고 파악된다.

이상 본 논문은 TIMIT 코퍼스를 사용하여 미국영어에서 유·무성 마찰음의 음향적인 특징을 성별, 조음 위치, 성대 진동의 유무에 따라 조사하였다. 본 논문은 마찰음의 음향연구를 다른 대부분의 선행연구와는 차별점이 있는데, 이는 선행연구들이 한정된 화자들을 대상으로 한정된 구나 환경에서 발화한 마찰음 혹은 인접한 자음의 음향적 특징을 추출하여 분석하였다. 본 논문은 모음을 선행하는 마찰음의 음향적 특징을 600명 이상의 화자가 발화한 2,000개 이상의 문장에서 추출하여 분석을 하여 선행하는 연구와의 유사성 및 차별성을 논의하였다는 점에서 차이점을 찾을 수 있다. 또한 성대 진동에 대한 정량적인 분석과 고주파수대와 저주파수대의 에너지의 총합의 차이를 정량적으로 구하여 해석을 시도하였다는 점에서도 선행하는 연구들과의 차별성이라고 할 수 있다. 본 논문에서는 모음 앞에 선행하는 마찰음들만을 고려하고 그 외의 문맥은 고려하지 않았다. 추후 문맥에 따른 마찰음들의 음향적인 효과들을 살펴보는 연구를 할 필요가 있을 것이다.

## 참고문헌

- Baum, S. R., & Blumstein, S. E. (1987). Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English. *Journal of the Acoustical Society of America*, 82(3), 1073-1077.



- Behrens, S. J., & Blumstein, S. E. (1988a). Acoustic characteristics of English voiceless fricatives: A descriptive analysis. *Journal of Phonetics*, 16(3), 295-298.
- Behrens, S. J., & Blumstein, S. E. (1988b). On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. *Journal of the Acoustical Society of America*, 84(3), 861-867.
- Boersma, P., & Weenink, D. (2015) *Praat: Doing phonetics by computer*. [Computer program], Version 5.4.05. Retrieved from <http://www.praat.org> on January 5, 2015.
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, 15, 39-54.
- Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America*, 83(4), 1553-1573.
- Docherty, G. (1992). *The timing of voicing in British English obstruents*. Berlin: Foris Publications.
- 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), 115-124.
- Gordon, M., Bathmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, 32(2), 141-174.
- Haggard, M. (1978). The devoicing of voiced fricatives. *Journal of Phonetics*, 6(2), 95-101.
- Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(3), 1252-1263.
- Johnson, K. (2012). *Acoustic and auditory phonetics* (2<sup>nd</sup> edition). Oxford: Wiley-Blackwell.
- Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). Kernlab – An S4 package for kernel methods in R. *Journal of Statistical Software*, 11(9), 1-20.
- Labov, W., Cohen, P., Robins, C., & Lewis, J. (1968). A study of the non-standard English of Negro and Puerto Rican speakers in New York City. *Final report, Cooperative Research Project 3288, Vols. I and II*, U.S. Regional Survey, Philadelphia.
- Ladefoged, P. (1982). *A course in phonetics*. New York: Harcourt Brace Jovanovich.
- Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristic of clearly spoken English fricatives. *Journal of the Acoustical Society of America*, 125(6), 3962-3973.
- McGuire, G., & Babel, M. (2012). A cross-modal account for synchronic and diachronic patterns of /f/ and /θ/ in English. *Laboratory Phonology*, 3(2), 215-271.
- McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, 118(2), 219-246.
- Miller, G., & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, 27(2), 338-352.
- Narayanan, S., Alwan, A., & Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 98(3), 1325-1347.
- Narayanan, S., Alwan, A., & Haker, K. (1994). An MRI study of fricative consonants. *Proceedings of the International Conference on Spoken Language Processing* (pp. 627-630), Yokohama, Japan.
- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org/> on April 4, 2018.
- Stevens, K. N. (1971). Airflow and turbulence noise for fricative and stop consonants: Static considerations. *The Journal of the Acoustical Society of America*, 50(4B), 1180-1192.
- Wagner, A., Ernestus, M., & Cutler, A. (2006). Formant transitions in fricative identification: The role of native fricative inventory. *The Journal of the Acoustical Society of America*, 120(4), 2267-2277.
- Whiteside, S. (1998). Identification of a speaker's sex: A fricative study. *Perceptual and Motor Skills*, 86(2), 587-591.
- Wilde, L. F. (2005). *Analysis and synthesis of fricative consonants*. Ph.D. Dissertation, MIT.
- Yoon, T. (2015). A corpus-based study on the effects of gender on voiceless fricatives in American English. *Phonetics and Speech Sciences*, 7(1), 117-124.
- Zue, V., & Seneff, S. (1988). Transcription and alignment of the TIMIT database. *Proceedings of the Second Meeting on Advanced Man-Machine Interface through Spoken Language*: 11.1-11.10.

• 윤태진 (Yoon, Tae-Jin)

성신여자 대학교 영어영문학과  
서울특별시 성북구 보문로 34다길2  
Tel: 02-920-7185  
Email: tyoon@sungshin.ac.kr  
관심분야: 음성학, 음운론, 언어공학