

한국어 대용량발화말뭉치의 단모음분석 Monophthong Analysis on a Large-scale Speech Corpus of Read-Style Korean

윤 태 진¹⁾ · 강 윤 정²⁾
Yoon, Tae-Jin · Kang, Yoonjung

ABSTRACT

The paper describes methods of conducting vowel analysis from a large-scale corpus with the aids of forced alignment and optimal formant ceiling methods. ‘Read Style Corpus of Standard Korean’ is used for building the forced alignment system and a subset of the corpus for the processing and extraction of features for vowel analysis based on optimal formant ceiling. The results of the vowel analysis are reliable and comparable to the results obtained from using traditional analytic methods. The findings indicate that the methods adopted for the analysis can be extended and used for more fine-grained analysis without time-consuming manual labeling and without losing accuracy and reliability.

Keywords: large-scale spoken corpus, forced alignment, optimal formant ceiling, monophthongs

1. 서론

근래에 들어 음성말뭉치를 이용한 음성학 연구가 급속하게 팽창하는 연구분야로 떠오르고 있다. 하지만, 자동 음향 분석을 할 수 있는 적절한 도구가 널리 이용되고 있지 않다는 점에서, 말뭉치 기반의 음성학 연구가 널리 이용되고 있지는 않다. 본 연구는 강제음성정렬장치와 최적의 상한선(optimal ceiling)을 이용한 포먼트 추출방식을 이용한 연구 방법을 이용하여 대용량 음성자료에서의 모음 연구를 수행하는 방법을 제시하고자 한다. 음성 자료는 ‘서울말 낭독체 발화말뭉치’를 이용하였다. 말소리 중 모음의 포먼트(Formant)는 턱과 혀의 움직임에 따른 성도 모양(vocal tract configuration)의 변화를 반영하며, 모음을 발화하고 인지하는 데 중요한 역할을 한다

(cf. Adank, 2003; Yang, 2008). 이렇듯 모음의 연구는 F1과 F2의 포먼트를 중심으로 많은 연구자들에 의해 지속적으로 연구되어 오고 있다. *Language*에 실린 Labov, Rosendfelder, and Fruehwald (2013: 34-35)는 모음 분석과 관련한 방법론에 대해 다음과 같은 논평을 하고 있다: “1972년에 수행된 필라델피아 지역의 언어 변화와 변이 프로젝트(Linguistic Variation and Change; LVC)에서 모음 연구를 위해 포먼트를 측정할 적인 있는데, 당시의 모음 측정은 개별 화자에게서 각각의 모음 범주들에서 수작업을 통해 5개에서 10개 정도씩 토큰을 선택한 후 LPC (Linear Predictive Coding) 분석을 했다. 모음 체계를 처리하는 데 한 모음당 약 40시간이 걸렸고, 모음 토큰의 수는 각 모음 유형에 따라 약 300에서 350개 정도 남겼다. 모음 측정을 자동으로 할 수 있는 소프트웨어 기술은 그 후로 많은 발전을 이루어, PRAAT 또는 ESPS (Entropic Signal Processing System)같은 소프트웨어를 사용하면 된다. 하지만 특정한 모음을 포함하고 있는 단어를 선택하고 포먼트 측정을 하기 위해 모음의 측정 지점을 고르는 것은 자동음성정렬을 위한 프로그램(e.g. Yuan & Liberman 2008)과 자동으로 포먼트 측정을 할 수 있는 프로그램 (Evanini 2009)이 개발되기 전까지는 여전히 상당히 많은 시간을 소모하는 작업으로 남아 있었다.”

이처럼 영어는 말뭉치기반의 음성학(corpus phonetics) 혹은

1) 청주대학교, tyoon@cju.ac.kr, 교신저자

2) University of Toronto at Scarborough, kang@utsc.utoronto.ca
이 논문은 제 1저자가 2013년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받고(NRF-2013S1A5A8022540), 제2저자가 캐나다의 SSHRC로부터 지원을 받아 연구되었음 (SSHRC-Partnership Grant #890-2012-25).

접수일자: 2014년 8월 15일
수정일자: 2014년 9월 23일
게재결정: 2014년 9월 30일

말뭉치 기반의 음운론(corpus phonology)라는 학제 간 연구에서 강제음성정렬(forced alignment)과 자동 포먼트 측정을 토대로 한 음성말뭉치 연구가 이루어지고 있다. 본 논문의 연구 목적은 한국어를 대상으로 강제음성정렬장치와 LPC 기반의 포먼트 추출에서 필요한 최적의 포먼트 상한선(optimal ceiling) 설정을 자동화하여 모음연구를 하는데 필요한 방법론적인 연구를 하는 것이다.

Linear Predictive Coding(선형예측코딩; LPC)을 이용한 포먼트 추출방식은 현재 가장 많이 이용되고 있는 방식이다. 그런데, 이 LPC를 이용한 포먼트 추출에는 근본적인 문제점이 있다. LPC를 이용해 포먼트를 추출할 때, LPC 차수(order)는 보통 포먼트 개수의 두 배가 되도록 지정한다. 일반적으로 남성 화자들의 포먼트를 신뢰성있게 추출하기 위해서는 여성화자들보다 더 많은 LPC 차수(order)를 지정해야 한다. Praat의 매뉴얼에 LPC를 구하기 위해서, 평균 성인 여성의 경우 5500Hz를 그리고 평균 성인 남성의 경우 5000Hz를 최고의 포먼트 값으로 정하는 것이 적절하다고 지적되어 왔다 (Boersma & Weenink, 2014). LPC 차수는 성별뿐만 아니라, 모음의 종류에 따라 영향을 받는다. 일반적으로 후설모음은 높은 LPC차수가 필요하고, 전설모음은 낮은 LPC차수가 필요하다 (Yao et al., 2010). 물론 이외에도 개인의 생리학적인 특징도 LPC 차수에 영향을 미칠 것이다. 이러한 복잡한 상호작용으로 인해, 보다 정확한 모음 분석을 원하는 음성학자들은 수작업을 통해 LPC 차수를 기입하고 결과를 확인하면서 포먼트를 추출하였다.

본 논문에서는 정확성을 유지하면서 모음분석작업을 자동화시킬 수 있는 방법을 사용하는 것을 제안한다. 일반적으로 각각의 모음은 성별에 따라 4000Hz에서 6000Hz사이의 주파수를 LPC를 기반으로 한 최적의 포먼트를 추출할 수 있는 최고값이라고 할 수 있으므로, 4000Hz에서 6000Hz중 성별과 모음에 따라 포먼트 측정치의 변이(variation)를 최소화하는 주파수값을 설정할 수 있다면, 그 최소의 변이를 가져오는 주파수값을 최적의 포먼트 상한선(optimal formant ceiling)이라고 할 수 있을 것이다 (Escudero et al., 2009). 본 논문에서는 자동 추출한 최적의 포먼트 상한선을 토대로 강제음성정렬된 음성신호에서 포먼트를 추출하여 대용량 말뭉치의 방대한 자료를 이용해 모음 정보를 추출하고자 한다.

2. 연구 방법

2.1 말뭉치

본 논문에서 자료로 이용할 말뭉치는 “서울말 낭독체 발화 말뭉치”이다. 이 ‘서울말 낭독체 발화 말뭉치’는 국립국어원에서 2002년에 개발하고 2005년도 경에 공개 및 분배한 말뭉치이다. 이 말뭉치는 20대에서 70대 사이의 서울 경기 지역 화자로, 부모 역시 서울 경기 지역 화자인 120명의 낭독체 자료

이다. 이 120명의 화자들은 남자 60명과 여자 60명으로 구성되어 있다. <표> 1은 연령대별 남녀 화자의 수를 보여준다.

표 1. 연령대별 남녀 화자 수

Table 1. Number of speakers by gender and age

연령	20대	30대	40대	50대	60대~70대
남성	20	20		11	9
여성	20		20	18	2

“서울말 낭독체 발화 말뭉치”는 120명의 남녀 화자들이 각각 초등학교에서부터 고등학교에 이르기까지 국어시간에 배운 동화, 수필, 소설 등 다양한 형태의 글들을 방음처리된 음성실험실에서 편안하게 읽은 음성자료들을 모아 놓은 말뭉치이다. 낭독한 텍스트와 한사람이 각 텍스트를 읽는 데 소요된 시간을 <표> 2에서 보여준다 (Yoon, 2013). 참고로 20대부터 40대까지는 19개의 작품을 약 1시간에서 1시간 20분 정도의 시간을 들여 모두 녹음하였고, 50대 이상은 9번부터 19번까지의 작품만 녹음하였다.

표 2. 서울말 낭독체 발화 말뭉치의 내용.

Table 2. Contents of the Reading Style Corpus of Standard Korean

	낭독체 내용	문장수	발화시간
1.	요람기	51	5분37초
2.	소나기	87	4분
3.	아름다움에 대하여	69	5분20초
4.	방망이를 깎던 노인	62	4분59초
5.	문학의 세계와 삶의 세계	47	5분35초
6.	독서와 인생	54	4분92초
7.	수난 시대	62	5분29초
8.	메밀꽃 필 무렵	94	7분10초
9.	토끼와 자라	60	3분09초
10.	선녀와 나무꾼	73	4분92초
11.	호랑이와 꾀감	42	2분67초
12.	해님 달님	28	2분07초
13.	그리운 시대가	36	4분61초
14.	광화문 지하도 아주머니	27	2분47초
15.	막 지은 밥	17	1분77초
16.	눈 오던 날	35	2분46초
17.	승냥의 지혜	19	1분58초
18.	까만 눈동자들 앞에서	27	3분43초
19.	내 고향 개울 물에서	40	2분72초
	합계	930	73분61초

이 말뭉치를 연구의 대상으로 삼은 이유는 비록 수년 동안 국립국어원에서 수집한 공개 음성자료이긴 하지만, 이 음성 자료를 활용한 음성 혹은 언어학적인 연구가 많지 않다는 것과 기존에 수행된 세대별 서울말 연구에 대한 보완책으로서

활용될 수 있다는 것이다. 또한 통제된 환경에서 추출된 발화에서는 경험하지 못한 대용량 음성말뭉치 특유의 분석상의 어려움을 파악하고 극복하는 데 도움이 되기 때문이다.

2.2 강제음성정렬장치

강제음성정렬(forced alignment)은 음성인식장치(Automatic Speech Recognition System)에 제약을 가하여 주어진 발화체들의 단어 및 단어를 구성하는 음소들의 연쇄만을 찾으려 하는 장치이다 (Hosom, 2000). 본 연구는 대용량 말뭉치에서 모음을 추출하고 분석하는 작업을 수행하기 위해 발화체들을 구성하는 단어와 음소들을 강제음성정렬(forced aligned)시키는 작업을 우선적으로 수행하였다. 이 작업을 위해 은닉마르코프모형(Hidden Markov Model; HMM) 기반의 자동강제음성정렬장치를 개발하였다.

일반적으로 HMM은 음성인식시스템을 구현하는 데 사용된다. 음성신호에 담긴 정보는 짧은 구간의 파워스펙트럼(power spectrum)으로 표현된다. 이 정보를 해석하기 위해서는 파워스펙트럼과 시간에 따른 변화를 추적해야하는 기술이 필요한데, 여러 기술 중 통계적 모델링 방법인 Hidden Markov Model(HMM)이 가장 성공적인 학습기술로 알려져 있다. 음성인식구축을 위한 공개소프트웨어인 HTK(HMM Tool Kit, Version 3.4.1, Young et al., 2006)과 HTK을 운용하기 위해 스크립팅 언어인 Python (version 2.7)를 사용하여 강제음성정렬시스템을 구축하였다.

강제음성정렬장치를 구축하기 위해서는 음향모형과 발음사전이 반드시 필요하다. 이 두 모형을 만들기 위해 음성 파일로는 100시간이 넘고, 어절 타입수로는 5500어절이 넘는 ‘서울말 낭독체 발화 말뭉치’를 사용하였다.

우선 음향모형을 만든 방법은 다음과 같다: 42개의 음성단위(phone unit)를 설정하고, HMM의 상태수 (number of states) 3의 연속음성훈련을 수행하도록 설계되었다. 사용된 음향매개변수는 12개의 MFCC(Mel Frequency Cepstral Coefficients)와 1개의 Energy를 기본 매개변수로 하여, 이들의 일차도함수 13개 및 이차도함수 13개를 도출하여, 총 39개의 음향매개변수를 사용하였다. 기본적인 음성모델링을 진행한 후, 문장 발화에서 단어(word)와 단어사이, 구(phrase)나 구사이, 혹은 발화문(utterance)의 시작점과 종결점에서 발생할 수도 있는 짧은 휴지(short pause)를 인식할 수 있는 휴지모형(silence model)을 설정하였다. 그리고 정형화된 발음열을 가진 발음사전을 사용하여 각각의 음성단위에 대한 정규분포의 결합(Gaussian Mixture)을 증가시키면서 음향모형을 재훈련시키는 방법을 사용하여 강제음성정렬시스템의 성능을 향상시켰다 (Yoon, 2013).

발음사전을 만들기 위해서 ‘서울말 낭독체 발화 말뭉치’에

서 관찰된 어절을 추출한 후, 어절 당 하나씩의 정형화된 발음열을 가진 기본 사전을 구축하였다. 발화 말뭉치에는 약 5,500개의 변별적인 어절 유형이 19개의 낭독체 텍스트에 분포되어 있다. 한글자모를 토대로 음절화시킨 후, 음절화된 자모를 유니코드의 기준에 맞추어 로마자 표기(Romanization)를 하였다. 국어에서는 음절사이의 자음들이 위치동화(place assimilation), 비음화(nasalization), 설측음화(lateralization) 등과 같은 음운변화를 겪는다는 점을 반영하여 (Ahn, 1996), 어절내의 음절 간에 일어나는 음운규칙을 일부 적용시켰다 (Yoon, 2013). <그림> 1은 강제음성정렬시킨 발화의 예를 보여준다.

이상 HTK와 Python를 기반으로 하여 ‘서울말 낭독체 발화 말뭉치’를 훈련시켜 개발한 강제음성정렬장치는 Korean Phonetic Aligner라는 이름으로 웹기반에서 이용할 수 있도록 하였다³⁾. 이 온라인 강제음성정렬장치에는 서울말 낭독체 발화말뭉치에서 나타나는 5500여개의 어절만이 발음사전으로 탑재되어 있어서, 새로운 단어들을 처리하는 데 한계가 있다. 하지만, 새로운 어절이나 단어도 발음사전을 생성하는 장치를 통해 처리할 수 있도록 개발되었다.

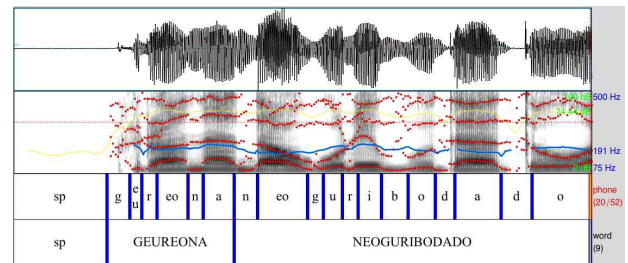


그림1. 강제음성정렬시킨 예.

Figure 1. Example of the output of forced-aligned utterance

2.3 포먼트 추출

현재 널리 사용되고 있는 음성분석프로그램인 Praat에서 일반적으로 권장되는 포먼트 측정방식은 To Formant(burg)이다. 물론 Burg 방식을 사용한 측정방식 외에도 Split-Levinson 알고리즘이나 자기상관방식(autocorrelation)을 이용해 포먼트를 측정할 수 있는 방법들이 있다. 하지만, Praat manual에서 Burg 방식을 권장한다. 양병근(2008: 50)에 의하면, 합성음의 포먼트 분석에서 Praat의 Burg 방식에서 가장 적은 측정오류가 나온다. 따라서 본 논문에서는 Burg 방식에 의한 포먼트 추출을 따르기로 한다.

우선 Burg 방식에 따른 포먼트 방식에 대해 살펴보고, 대용량의 음성자료에 자동처리 방식을 적용하여 결과를 추출하였을 때 어떠한 이론적 및 실용적 문제점을 지니는 지 살펴볼도록 하겠다. 모음 분석을 위해서는 말뭉치의 자료 중 '호랑

3) korean.utsc.utoronto.ca/kpa/

이와 콧감'에서 나타난 단모음만을 대상으로 하였다. 분석 화자의 수는 음성파일에 이상이 없고 강제음성정렬에서도 오류를 나타내지 않은 88명을 대상으로 하였다. <표> 3은 '호랑이와 콧감'에 나오는 모음의 분포를 화자 1명이 발화한 것이다.

표 3. '호랑이와 콧감'에 나타난 단모음의 분포

Table 3. Number of monophthongal vowel tokens in the story 'The Tiger and the Dried Persimmon'

모음	한글	IPA	화자당 토큰수
a	(ㅏ)	[a]	2540
ae	(ㅓ)	[ɛ]	421
e	(ㅕ)	[e]	427
eo	(ㅛ)	[ʌ]	1224
eu	(ㅡ)	[i]	1555
i	(ㅣ)	[i]	1683
o	(ㅜ)	[o]	1165
u	(ㅜ)	[u]	724
			9739

<그림> 2는 한 여성 화자의 음성자료에서 F1과 F2를 Burg 방식을 이용하여 추출한 모음도(vowel chart)이다. 일반적인 관행을 따라, 여성 화자이므로 포먼트 상한선(Formant ceiling)을 5500Hz으로 한 후, 모음 측정의 오류를 줄이기 위해 모음의 중앙 20% 구간의 평균 F1과 F2 값을 측정한 결과임에도 불구하고, 측정상의 에러(measurement error)를 보여주는 토큰들이 다수 있다는 것을 알 수 있다.

Formant values with the ceiling of 5500Hz

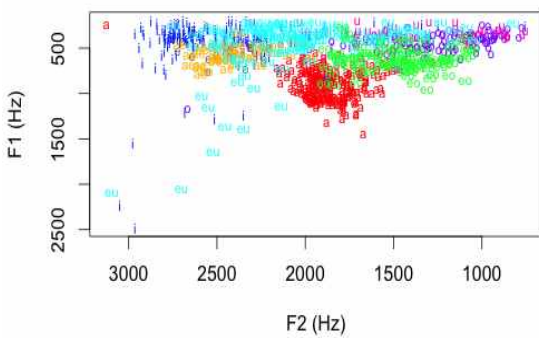


그림 2. 5500Hz의 포먼트 상한선(formant ceiling)을 토대로 구성된 여성화자의 모음 분포도

Figure 2. Vowel formant chart of a female speaker drawn from the formant values with the 5500Hz formant ceiling

최적의 포먼트 상한선을 찾기 위해서는 다음과 같은 방식이 사용되었다. 4000Hz에서 시작하여 100Hz씩 증가시켜 6500Hz까지 최적의 포먼트 상한선의 후보 값이 될 수 있도록 하였다. 이 26 개의 후보값 중에서, F1과 F2 측정치의 변이(variance)가 가장 적게 나타나는 포먼트 상한선 값을 최적형

으로 선택하였다 (Escudero et al. 2009). <그림> 3은 각 모음별로, 포먼트 상한선값(x-축)에 따라 F1, F2의 평균값과 변이(y-축)가 어떻게 달라지는 지를 나타낸다. 수직선으로 표시된 부분은 최소의 변이를 보이는 포먼트 상한선값을 보여주는 부분이다.

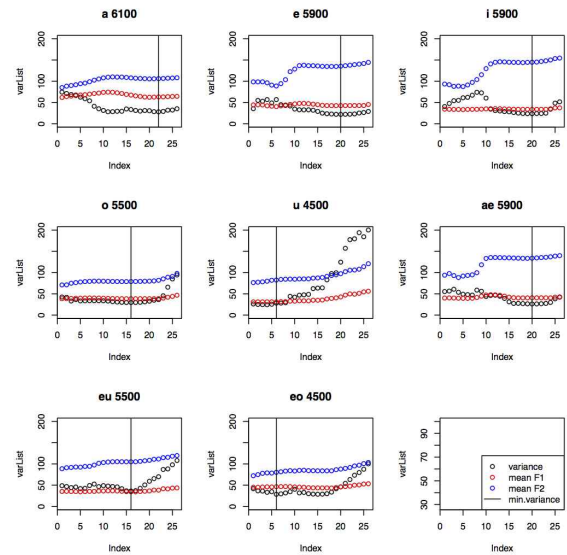


그림 3. 최소의 변이형을 보여주는 최적의 최고값

Figure 3. Optimal ceiling that reveals minimal variation

이상과 같은 방식을 통하여, 88명의 화자들에게서 각 단모음에 대한 최적의 포먼트 상한선을 결정하였다. <그림> 4는 20대 여성 화자 중 한 명 (FV03)에게서 추출한 최적의 포먼트 상한선을 모음별로 도식화한 것을 보여주고 있다.

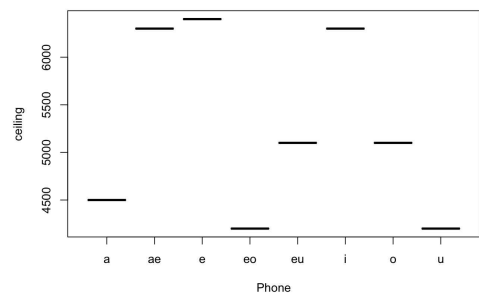


그림 4. 한 여성화자(FV03)의 모음별 최적의 포먼트 상한선

Figure 4. Optimal formant ceilings of vowels for a female speaker (FV03)

성별에 따라 최적의 포먼트 상한선(formant ceiling)이 달리 나타나는 이유는 여성과 남성의 성도의 길이가 차이가 나기 때문이다. 그리고 모음에 따라 포먼트 상한선이 다르게 나타나는 이유는 후설모음 [o]와 [u] 같은 경우 LPC 차수가 낮으

면 두 개의 포먼트가 가까이 접근하여 하나의 포먼트로 인식 되는 오류가 발생할 수 있기 때문이다. <그림> 5는 20대 여성 화자들의 단모음에 따른 optimal ceiling의 분포를 보여주고 있다. <그림> 5를 통해 후설모음일 경우의 최적의 포먼트 상한선 값이 전설모음일 경우의 최적의 포먼트 상한선 값보다 다소 낮아지는 경향이 있다는 것을 관찰할 수 있다.

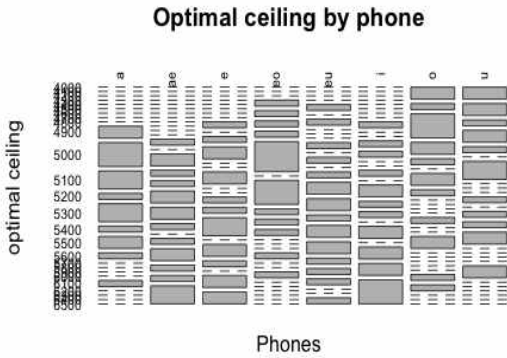


그림 5. 20대 여성화자들의 모음별 최적의 포먼트 상한선 분포

Figure 5. Distribution of optimal ceiling values for all female speakers in their 20's.

이상의 LPC 차수의 최적의 포먼트 상한선(optimal formant ceiling)을 이용하여 추출한 각 단모음에 해당하는 F1과 F2를 가지고, 평균과 표준편차를 구하였다. 평균에서 2x표준편차 내에 95%이상의 토큰들이 분포를 이루고 있으므로, 2x표준편차 이상으로 벗어나는 토큰들의 포먼트값은 이상치(outliers)로 간주하고 분석에서 제외하였다. <그림> 6은 이와 같은 방식을 통해 얻은 한 남성 화자의 모음도이다.

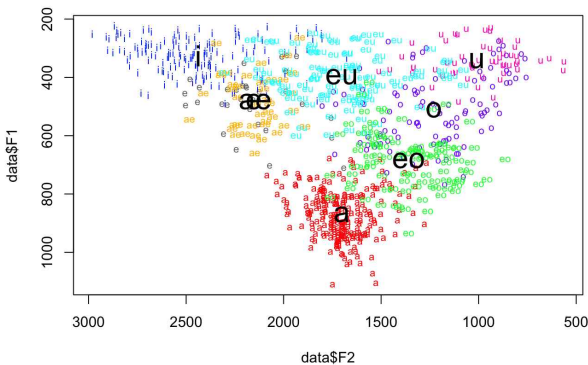


그림 6: 한 남성(mz08)의 최적의 포먼트 상한선에 의한 모음도

Figure 6. Vowel chart of a male speaker (mz08) obtained by the optimal ceiling method.

<표> 4는 <그림> 6에 표시된 포먼트들의 평균값과 표준편차

를 수치화한 것이다.

표 4. 한 남성화자(mz08)의 F1과 F2 평균 및 표준편차

Table 4. Mean and standard deviation of F1 and F2 of the male speaker (mz08)

mz08	F1 (Hertz)		F2 (Hertz)	
	mean	std. dev.	mean	std. dev.
i	329	46	1911	207
e	485	48	1725	96
ae	475	37	1689	128
a	667	75	1264	109
eo	500	42	1050	167
o	445	47	943	173
u	331	37	1017	208
eu	407	46	1911	207

이상 88명의 화자들이 ‘호랑이와 꽃감’라는 작품을 낭독한 음성자료를 가지고 자동음성정렬장치와 최적의 포먼트 상한선을 선택하는 방식을 통해 대용량의 음성자료를 처리할 수 있는 방법을 실증하였다. 다음의 <표> 5과 <표> 6은 모든 남성과 여성의 F1과 F2의 평균과 표준편차를 각각 보여준다.

표 5. 남성 화자들의 F1과 F2 평균 및 표준편차

Table 5. Mean and standard deviation of F1 and F2 by all male speakers

남성	F1 (Hertz)		F2 (Hertz)	
	mean	std. dev.	mean.	std. dev.
i	325	29	1960	159
e	448	42	1739	124
ae	444	46	1717	113
a	638	58	1333	105
eo	486	42	1100	64
o	412	38	984	60
u	333	34	1066	70
i (eu)	374	40	1381	98

표 6. 여성 화자들의 F1과 F2 평균 및 표준편차

Table 6. Mean and standard deviation of F1 and F2 by all female speakers

여성	F1 (Hertz)		F2 (Hertz)	
	mean	std. dev.	mean	std. dev.
i	360	35	2442	147
e	500	38	2168	124
ae	480	47	2173	138
a	825	65	1688	80
eo	595	56	1286	84
o	450	44	1103	98
u	357	36	1203	120
i (eu)	410	42	1706	141

<그림> 7은 모든 화자들의 F1과 F2의 평균값을 성별로 나누

어 그려놓은 모음도이다. 파란색은 여성 화자, 그리고 빨간색은 남성 화자들의 평균값을 나타낸다.

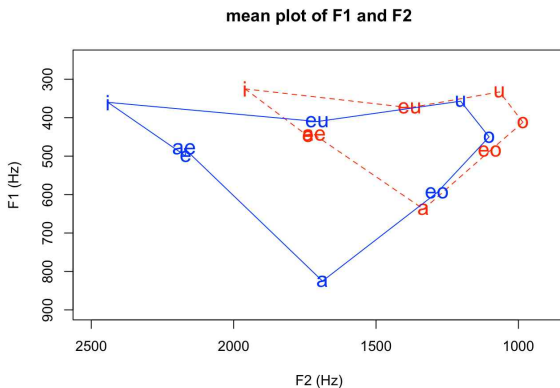


그림 7. 남녀화자들의 평균 F1 및 F2 값을 나타내는 모음도

Figure 7. Vowel chart with mean F1 and F2 of male and female speakers, respectively

본 논문에서 제시한 화자들의 포먼트 평균값은 통제된 음성 실험을 통해 추출한 모음의 포먼트 값과 비슷한 결과를 보여주고 있다 (cf. Yang, 1996)⁴⁾. 이는 본 논문에서 시도한 포먼트 자동 측정 방식이 신뢰할 만한 정도의 정확성을 가지고 있음을 나타내는 것이다. <그림> 7을 보면, /e/와 /æ/의 경우, F1-F2 모음 공간상에서 이미 합병(merger)이 되었다는 것을 알 수 있다. 현대 한국어의 표준말 단모음 체계는 연구자에 따라 7모음체계로 보는 경우와 10모음 체계로 보는 경우도 있다. 7모음 체계를 주장하는 분석은 /i, ε, ʌ, a, i, u, o/와 같은 단모음들을 설정하고 있다. 10모음 체계는 7모음 체계에 /y, ø, æ/를 음소 목록에 첨가한 경우이다. /y/와 /ø/ 모음의 경우는 현재 대부분의 한국어 화자들이 두 모음을 각각 /wi/와 /we/와 같이 이중모음으로 조음하고 있다 (Han and Kim, 2014). /e/와 /æ/의 경우 합병(merger)이 되었다는 것을 <그림> 7을 통해서 알 수 있다. /e/와 /æ/의 합병(merger)은 Hong (1988) 등의 여러 연구에서 검증된 바 있다. 한국어 모음 /o/와 /u/가 F1과 F2의 평균값에서 서로 큰 차이를 보이지 않고 하나의 집단으로 분류되는 경향이 지속적으로 보고되고 있다 (cf. Chae, 2005; Yun and Seung, 2013; Kang, 2013). <그림> 7에서도 다른 모음들 간의 거리보다 모음 /o/와 /u/의 거리가 가깝다는 것을 보여 준다.

4) 본 논문에서의 모음 분석의 결과가 기존의 통제된 음성 실험에서 보고된 결과와 유사하다는 것은 넓은 의미에서 해석할 필요가 있다. 유사성이 각 모음의 평균 값이 오차 범위 내에 있다는 의미보다는 조음과 포먼트의 관계에 비추어 보았을 때 우리가 예측할 수 있는 기대치에 근접한 모음도 혹은 기존의 통제된 음성 실험에서 보고된 모음도(vowel chart)에 유사한 형태로 구해졌다는 의미로 해석할 수 있다.

4. 결론

음성말뭉치(spoken corpus)를 이용한 음성분석을 위해서는 음성말뭉치에서 음성정보를 추출할 수 있는 도구가 있어야 한다. 음성말뭉치를 이용하는 데 있어서 우리가 넘어야 할 산은 크게 두 가지가 있다. 첫째는 음성신호의 변이성(variability)이다. 문자로 된 말뭉치(written corpus)와 달리 음성말뭉치는 여러 환경의 영향으로 인해 보다 많은 변이를 보인다. 또한, 음성말뭉치는 일반적으로 다양한 화자들에게서 녹음한 음성자료를 포함하고 있다. 이러한 이유로 인해 화자간의 생리학적 차이, 사회적 요소 및 개별 화자의 특성 등 화자간의 변이(inter-speaker variation)도 분석에 있어서 커다란 난제라고 할 수 있다 (Adank, 2003).

음성말뭉치를 사용할 때 주의해야 할 또 다른 문제점은 “측정상의 오류”(measurement error)이다 (Yao et al., 2010). 최근 사용되는 자동음성분석방식 중 많은 방식들이 자동음성인식이나 음성합성을 위해 개발된 것들이다. 그런데 음성인식이나 음성합성은 일반적으로 음성분석에 필요한 정도의 정확성보다는 전반적인 인식 혹은 합성의 정확성에 신경을 쓰는 경우가 있다.

본 논문은 이러한 측정상의 오류를 극복하고 대용량의 음성자료를 처리할 수 있는 방법을 실증해 보여주었다. 다시 말해, 본 논문은 대용량의 음성 말뭉치를 사용하여 강제음성정렬장치와 최적의 포먼트 측정을 통한 한국어 단모음의 분석방법을 제시하였다. 우선, 강제음성정렬장치를 통해 음성신호에서 단어 및 유사음소단위 정렬을 하였다. 강제음성정렬된 정보를 이용하여 단모음들을 파악한 후, F1과 F2를 각각의 단모음들이 가지고 있는 음향음성학적 특성과 각각의 화자들이 가지고 있는 특성에 맞추어 추출하기 위해 최적의 포먼트 상한선을 설정하는 작업을 수행하였다. 그리고 마지막으로 각 화자들이 가지고 있는 평균 포먼트 값에서 2x표준편차 이상으로 벗어나서 추출된 포먼트 값은 이상치(outlier)라고 판단하여 분석대상에서 제외하였다. 강제음성정렬장치에서도 오류가 발생할 수 있고, 고정된 매개변수를 상정하여 포먼트를 추출하는 과정에서도 오류가 발생할 수 있다. 포먼트의 분포에 제약을 가함으로써 분석상의 오류들을 최소화 할 수 있으며, 이는 대용량의 음성자료를 생태학적으로 유효한 결과를 도출하고 기존의 통제된 환경에서 도출한 모음 분석 결과와 비교할 수 있는 유효성을 가져다주는 효과를 가지고 있다.

참고문헌

- [1] Adank, P. (2003). *Vowel normalization: A perceptual-acoustic study of Dutch vowels*. Ph.D. thesis, University of Nijmegen.
- [2] Ahn, S.-C. (1996). *An Introduction to Korean Phonology*.

- Seoul: Hanshin Publishing Co.
- [3] Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.3.85, retrieved 19 September 2014 from <http://www.praat.org>.
- [4] Chae, S.-Y. (2005). *External Constraints on Sound Change: The Raising of /o/ in Seoul Korean*. Doctoral Dissertation, University of Pennsylvania.
- [5] Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. A. H. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *Journal of Acoustical Society of America*, 126, 1379-1393.
- [6] Evanini, K. (2009). *The permeability of dialect boundaries: a case study of the region surrounding Erie, Pennsylvania*. Doctoral Dissertation, University of Pennsylvania.
- [7] Han, J.-I. & Kim, J.-Y., (2014). A phonetic investigation of Korean monophthongs in the early twentieth century. *Phonetics and Speech Sciences*, 6(1), 31-38.
(한정임 & 김주연. (2014). 20세기 초 한국어 단모음의 음향 음성학적 연구. *말소리와 음성과학* 6(1), 31-38.)
- [8] Hong, Y. (1988) *A sociolinguistic study of Seoul Korean*. Unpublished PhD thesis, University of Pennsylvania.
- [9] Hosom, J.-P. (2000). *Automatic Time Alignment of Phonemes Using Acoustic-Phonetic Information*. Ph.D. thesis, Oregon Graduate Institute.
- [10] Kang, Y. (2013). A corpus-based study of positional variation in Seoul Korean vowels. Paper presented at *Japanese Korean Linguistics Conference 23*. October 11-13.
- [11] Labov, W., I. Rosenfelder, & J. Fruehwald. (2013) One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, 89(1), 30-65.
- [12] Yang, B. (2008). Formant measurement of complex waves and vowels produced by students. *Korean Journal of Phonetic Sciences* 15(3), 39-52.
(양병곤 (2008). 복합음과 대학생이 발음한 모음 포먼트 측정. *음성과학* 15(3), 39-52)
- [13] Yang, B. (1996) A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics*, 24, 245-261.
- [14] Yao, Y., S. Tilsen, R. L. Sprouse, & K. Johnson. (2010). Automatic Measurement of Vowel Formant in the Buckeye Corpus. *UC Berkeley Phonology Lab Annual Report*.
- [15] Yoon, T.-J. (2013). Large-scale Studies of Phonetic Data using Forced Alignment System. *Studies in Linguistics*, 29, 207-227.
(윤태진 (2013). 강제음성정렬장치를 이용한 대용량음성자료 연구. *언어학연구*, 29, 207-227.)
- [16] Young, S. J., G. Evermann, M. J. F. Gales, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, & P. C. Woodland. (2006). *The HTK Book* (version 3.4), Cambridge University Engineering Department
- [17] Yuan, J. & M. Liberman (2008). "Speaker identification on the SCOTUS corpus," *Proceedings of Acoustics 2008*, 5687-5690.
- [18] Yun, J. & Seung, C. (2013). Effects of F1/F2 manipulation on the perception of Korean vowels /o/ and /u/. *Speech Sciences*, 5(3), 39-45.
(윤지현, 성철재. (2013). F1/F2의 변화가 한국어 /오/, /우/ 모음의 지각관별에 미치는 영향. *말소리와 음성과학*, 5(3), 39-45.)

• 윤태진 (Yoon, Tae-Jin)

청주대학교 영어영문학과
충북 청주시 청원구 대성로 298, 360-764
Tel: 043-299-7573 Fax: 043-229-8310
Email: tyoon@cju.ac.kr
관심분야: 음성학, 음운론

• 강윤정 (Kang, Yoonjung)

Centre for French and Linguistics,
University of Toronto Scarborough
1265 Military Trail, HW314, Toronto ON MIC 1A4, Canada
Tel: 416-287-7172 Fax: 416-208-2669
Email: kang@utsc.utoronto.ca
관심분야: 음성학, 음운론