

한국어 비음의 음향학적 세부 기술을 위한 장구간 스펙트럼(LTAS) 분석

최순애 • 성철재

충남대학교 언어병리학과 • 충남대학교 언어학과

Long Term Average Spectral Analysis for Acoustical Description of Korean Nasal Consonants

Soonai Choi * • Cheol-Jae Seong * *

Department of Speech Pathology *, Department of Linguistics * *,
Chungnam National University

Abstract

The purpose of this study is to find the acoustic parameters on frequency domain to distinguish the Korean nasals, /m, n, ng/ from each other. Since it is not easy to characterize the antiformant on frequency domain, we suggest the new parameters that are calculated by LTAS(Long term average spectrum). Maximum energy value and its frequency and minimum energy and its frequency of zero are obtained from the spectrum respectively. In addition, slope1, slope2, total energy value, centroid, skewness, and kurtosis are suggested as new parameters as well. The parameters that are revealed as to be statistically significant difference are roughly peak1_a, zero_f, slope_1, slope_2, highENG, zero_ENG, and centroid.

Key words : parameter, LTAS, slope, zero, centroid, skewness, kurtosis

I. 머리말

말소리 중 비음은 구강의 한 부분을 막고 연구개를 내려 폐로부터 나오는 기류를 비강을 통해 내보내면서 조음하는 소리이다.[1][2]. 음향학적으로 비음은 본래 공명과 대치되는 반공명 (antiresonance)이 발생되는데, 성도 응답(vocal tract response)에 관한 측면

에서는 공명(formant)을 pole로 반공명을 zero라 한다.[3][4][5] 한국어에서 실현되는 비음 /m/, /n/, /ng/은 각각이 고유의 음향적 특성을 가지고 있을 뿐 아니라 환경에 따라 실현되는 음의 음향적인 특징이 다르다[3].

일반적으로 비음은 안면 두개골 구조에 있는 여러 개의 부비동과 비공의 좁힘 때문에 제1포먼트의 에너지가 낮은 주파수대에서 형성된다.[6] 제1포먼트 에너지는 스펙트로그램상에서 약 200Hz ~ 300 Hz 주파수대에 강한 에너지띠가 나타난다.[5][6]. 또한 비음에 앞뒤로 오는 모음에도 비음성이 영향을 주어 제1포먼트의 에너지를 약화시키고, 800Hz 아래에서 포먼트의 사이가 가깝게 형성된다.[4][5][6]

또 다른 비음의 음향적 특성은 포먼트 대역 (bandwidth)이 넓은 것이다.[7] 성대에서 생성된 성문음은 구강을 거쳐 비강을 통과하면서 감쇠된다. 성도의 면적이 넓어지면서 공명강의 면적과 체적이 커지고 비강의 복잡한 구조 때문이다.

반공명주파수는 제1포먼트 이상에서 에너지가 감소되며 반공명 주파수 근처의 주파수대역 에너지도 강하게 감소시킨다. 구강과 비강의 coupling의 정도에 따라 스펙트럼상에서 에너지 증폭이 전혀 안되는 경우도 있다.[3] 또한, 구강이 부가적인 공명강으로 역할을 하여 1.6dB 정도의 에너지 감소 효과를 준다. [m]은 1,100Hz 이상에서 포먼트 에너지가 약해지며[1][5], 반공명은 1,100Hz와 3,300Hz에서 생긴다.[8][9] [n]은 1,200Hz~ 1,300Hz 사이에서 포먼트 에너지가 약해지고, 1,600Hz와 4,800Hz에서 강하게 에너지가 감소되어 반공명이 생긴다. [ng]는 1,700Hz~ 2,000Hz에서 포

먼트 에너지가 약해지고 1,100 Hz, 3,200Hz에서 반공명이 있다. 또 다른 비음 특성은 스펙트럼상에서 인접 분절음과의 경계부분에 불연속적인 전이가 있고, 각 포먼트 사이의 거리가 감소한다.[6][7]

본 연구는 한국어 비음을 초성과 종성 환경에서 각각 구분할 수 있는 음향 파라미터를 주파수 영역에서 찾아보고자하는 시도의 일환이다. 한국어 비음 /ㄱ/, /ㄴ/, /ㅇ/을 각 분절음과 초성과 종성의 환경에 따라 구분하기 위한 음향변수를 찾고자 하였다. 반공명과 모음에 비해 상대적으로 넓은 포먼트 대역 등이 비음의 주요 특성이라는 것은 음성학적으로 이미 밝혀져 있는 사실이지만 주파수 영역에서 이를 구분하는 작업이 그리 쉽지 않은 관계로 기계적인 분류작업이 가능할 수 있는 새로운 파라미터를 제안한다. 새로운 파라미터는 장구간평균 스펙트럼(LTAS; Long term average spectrum) 값을 다양하게 이용하며, 이는 산술적으로 특정주파수 구간의 평균 파워 값으로 정의될 수 있다(그림 1 참고).

II. 실험 개요

2.1 실험 재료

실험에 사용된 음성샘플은 원광대학교에서 제작한 K-Kwon PBS data를 사용하였으며, 제작당시의 세부 사항은 다음과 같다: 음성은 방음부스에서 Senheizer HMD 224X 마이크를 사용하여 녹음하였으며, 발생된 데이터는 디지털 오디오 테이프에 저장하였다. 디지털 변환은 Sun Workstation Sparc20 환경에서 실시하였으며, DAT-Link+를 사용하여 입력하였다. 16kHz로 샘플링하고 16Bit로 양자화하였다.

피실험자의 구성은 남자 9명, 여자 11명이며, 연령별로는 40대가 4명, 20대가 16명이다. 출신지역은 서울과 경기, 인천, 충남, 경남, 경북, 제주 등으로 다양하며 현 거주 지역은 전북, 대전, 충남 지역이다. 직업은 학생과 교수이다.

2.2 데이터 추출 및 파라미터 값 측정

샘플은 각 화자별 동일한 단어에서 비음부분이 선명히 구분되는 구간을 선택하였고(대략 30ms~60ms), 초성과 종성, 남녀로 구분하여 개별 음성파일로 저장하였다. 실험에 사용된 분절음은 /ㄴ/(초성) 101개, /ㄴ/(종성) 100개, /ㅇ/(초성) 100개, /ㅇ/(종성) 99개, /ㅇ/(종성) 100개이다. 초성 비음은 NV(C)(\$의 환경에

서, 종성 비음은(\$)(C)VN(\$)(\$는음절경계)의 환경에서 선택하였다. 각 분절음의 파라미터 값은 음성 분석 프로그램 Praat(4.4.10)을 이용하여 파일을 추출하였고, Praat의 open source code를 이용하여 C++로 필요한 내용을 프로그램하여 측정하였다. 비음의 음향특징은 장구간평균 스펙트럼(LTAS)으로부터 저주파 대역과 고주파 대역에서의 스펙트럼 최고치, 스펙트럼의 기울기, 100~5000Hz사이의 최소진폭과 해당 주파수(zero_f, zero_a로 명칭), 대역별 에너지총량, 스펙트럼 무게중심, 왜도, 첨도를 선정하였고 이들의 통계치를 비음 데이터로부터 구하여 분석하였다.

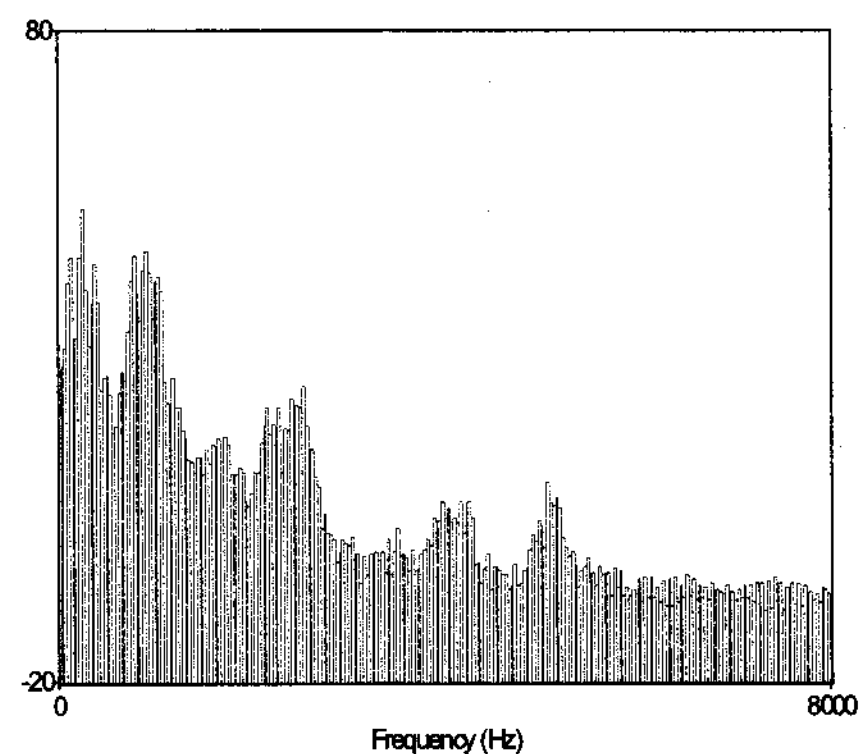


그림 1. [ㄴ]의 LTAS /
bin bandwidth 35.12 Hz(8000/256)

스펙트럼 최고치와 zero는 저주파 대역(0Hz-4000Hz)과 고주파 대역(4000Hz-8000Hz)에서 각각 주파수와 진폭을 구하였다(peak1_f, peak1_a; peak2_f, peak2_a). 스펙트럼의 기울기는 100Hz에서 6000Hz 사이의 주파수 대역에서 스펙트럼 최고치를 선정하고, 이를 중심으로 좌우 주파수 대역에서의 스펙트럼의 기울기를 선형추정(linear estimation) 하였을 때에 최소자승오차(minimum square error)를 나타내는 값으로 정하였다(slope1, slope2). 또한, 에너지총량을 0Hz~4,000Hz, 4,000Hz~ 8,000Hz의 주파수 도메인에서 구하고(lowENG, highENG), zero의 에너지 총량을 위한 주파수 도메인은 [n]의 특징 주파수 영역인 1,500Hz~2,200Hz에서 구하였다(zeroENG).

자음의 음소별 특징을 나타내는 또 다른 특징으로 스펙트럼의 무게중심을 사용하였다. 스펙트럼의 무게중심(centroid)은 전체 주파수 대역에서 에너지 스펙트럼이나 전력 스펙트럼의 분포상의 무게중심에 해당하는 주파수를 의미한다. 스펙트럼의 왜도(skewness)는 스펙트럼의 무게중심으로부터의 기울어진 정도, 즉, 비대칭성을 나타내는 척도이다. 스펙트럼의 첨도(kurtosis)는 스펙트럼의 뾰족 정도를 나타내는 척도로써 해당주파수의 공명정도를 말해준다.

환경적 변수를 최소화하여 초성 비음과 종성 비음으로 나누어 각기 음향 특징을 살펴보고 최종적으로는 하나의 범주 집단으로 간주하여 $p < 0.05$ 수준에서 독립 t-검정과, 분산분석(ANOVA)을 실시하였다.

III. 실험결과

남녀간에는 성대의 모양, 성도의 길이와 용적 등 해부학적 차이가 있으므로 산출된 비음의 음향학적 특징이 차이나는 것은 당연할 것이다. 그러나, 성대 진동의 강약에서 비롯된 진폭의 남녀간, 개인간 편차를 없애기 위하여 데이터에 대한 정규화 작업이 이루어졌으므로, 남녀 간에는 구강과 비강으로 열린 성도 특성의 차이만 남게 된다. 측정결과에 대한 기술통계와 t-검정을 통하여 통계적으로 유의미한 차이가 나는 파라미터를 정리하면 다음과 같다.

표1. 남성 비음 /ㄱ/과 /ㄴ/ 기술통계표

파라미터	토큰수	평균	표준 편차	평균의 표준오차
peak1_f	ㄱ_100	264.3	49.64	4.964
	ㄴ_101	277.6	58.45	5.816
peak1_a	ㄱ_100	1.281	.0430	.0043
	ㄴ_101	1.279	.0338	.0033
peak2_f	ㄱ_100	4915	1014	101.4
	ㄴ_101	5149	1120	111.4
peak2_a	ㄱ_100	1.062	.0319	.0031
	ㄴ_101	1.062	.0337	.0033
zero_f	ㄱ_100	4032	987.6	98.76
	ㄴ_101	3978	987.7	98.28
zero_a	ㄱ_100	.862	.0426	.0042
	ㄴ_101	.8566	.0421	.0041
slope1	ㄱ_100	.5305	.0074	.0007
	ㄴ_101	.5305	.0085	.0008
slope2	ㄱ_100	.4734	.0073	.0007
	ㄴ_101	.4733	.0085	.0008
lowENG	ㄱ_100	.0933	.0021	.0002
	ㄴ_101	.0924	.0024	.0002
highENG	ㄱ_100	-35.7	8.027	.8027
	ㄴ_101	-34.8	9.864	.9815
zeroENG	ㄱ_100	-19.4	34.63	3.463
	ㄴ_101	-3.97	34.63	3.446
centroid	ㄱ_100	3838	31.41	3.141
	ㄴ_101	3841	35.61	3.543
skewness	ㄱ_100	.0713	.0216	.0021
	ㄴ_101	.0708	.0249	.0024
kurtosis	ㄱ_100	-1.20	.0166	.0016
	ㄴ_101	-1.20	.0150	.0015

표2. 각 비음별 남성과 여성의 비교에서 유의미한 파라미터

비음구분	유의미한 파라미터
초성 /ㄱ/	peak1_f, peak1_a, zero_f, zero_a, slope_1, slope_2, highENG, zeroENG, centroid, skewness
초성 /ㄴ/	peak1_a, peak2_a, zero_f, zero_a, slope_1, slope_2, highENG, zeroENG, centroid, skewness
종성 /ㄱ/	peak1_a, peak2_a, zero_f, slope_1, slope_2, lowENG, highENG, zeroENG, centroid, skewness
종성 /ㄴ/	peak1_a, peak2_a, zero_f, slope_1, slope_2, highENG, zeroENG, centroid, skewness
종성 /ㅇ/	peak1_a, peak2_a, zero_f, slope_1, slope_2, lowENG, highENG, zeroENG, centroid, skewness

남성과 여성의 발화샘플에서 초성과 종성 비음을 각각 비교했을 때 전 비음에 걸쳐 남녀 간 차이를 가장 잘 드러내는 공통 파라미터는 peak1_a, zero_f, slope_1, slope_2, highENG, zeroENG, centroid 이다.

표3. 남성 초성과 종성의 유의미한 파라미터

비음구분	유의미한 파라미터
초성_ㄱ/ㄴ	zeroENG: $t = -2.277, p < 0.03; df = 91$
종성_ㄱ/ㄴ/ㅇ	peak1_a: $F(2,132) = 3.857, p < 0.03$ peak2_f: $F(2,132) = 4.182, p < 0.03$ lowENG: $F(2,132) = 4.458, p < 0.03$

남성 비음의 비교에서 사후검정으로 Dunnett의 t-검정 결과 종성 /ㅇ/과 /ㄱ/($p = 0.014$), 종성 /ㄴ/과 /ㅇ/($p = 0.001$) 사이에서 통계적으로 유의미한 차이를 보였다. 종성의 유의미한 파라미터는 모두 /ㅇ/과 /ㄱ/, /ㅇ/과 /ㄴ/ 사이를 구별하는 파라미터이다. 즉, 현 파라미터 세트에서 초성 /ㄱ/과 /ㄴ/의 구별에 관여하는 파라미터는 있으나 종성 /ㄱ/과 /ㄴ/을 구별하는 파라미터는 찾을 수 없다.

표4. 여성 초성과 종성의 유의미한 파라미터

비음구분	유의미한 파라미터
초성_ㄱ/ㄴ	lowENG: $t = 3.18, p < 0.003; df = 106$ zeroENG: $t = -2.497, p < 0.02; df = 106$
종성_ㄱ/ㄴ/ㅇ	peak1_a : $F(2,161) = 5.023, p < 0.01$

여성 비음의 비교에서 초성 /ㄱ/과 /ㄴ/ 사이에서는 lowENG, zeroENG 유의미한 파라미터이며, 종성 /ㄱ/, /ㄴ/, /ㅇ/ 사이에서는 peak1_a가 유의미한 파라미터이다. 사후검정에서 /ㄴ/과 /ㅇ/은 유의미한 차이가 ($p < 0.005$) 있지만, /ㄱ/과 /ㄴ/ 사이는 유의미한 차이가 없는 것으로 나타났다.

표5. 남녀통합 초성과 종성의 유의미한 파라미터

비음구분	유의미한 파라미터
초성_ ㄱ/ㄴ	zeroENG($t = -3.177$, $p < 0.003$; $df = 199$)
종성_ ㄱ/ㄴ/ㅇ	peak1_a: $F(2,296) = 6.658$, $p < 0.003$ peak2_f: $F(2,296) = 3.812$, $p < 0.03$ peak2_a: $F(2,296) = 3.227$, $p < 0.05$ lowENG: $F(2,296) = 5.897$, $p < 0.005$

남녀의 양성의 데이터를 합쳐서 비교한 초성 /ㄱ/와 /ㄴ/의 t-검정에서는 zeroENG만이 초성 비음 사이에 유의미한 차이를 보이는 파라미터이다. 종성 /ㄱ/, /ㄴ/, /ㅇ/은 peak1_a, peak2_f, peak2_a가 유의미한 차이를 보였다. Tukey 사후검정에서 종성 /ㅇ/과 /ㄱ/ ($p = 0.020$), 종성 /ㅇ/과 /ㄴ/ ($p = 0.002$)에서 통계적으로 유의미한 차이를 보였고 종성 /ㄱ/과 /ㄴ/은 peak2_a에서만 유의미한 차이 ($p < 0.05$)를 보였다.

IV. 맺음말

본 연구에서는 한국어 비음 /ㄱ/, /ㄴ/, /ㅇ/을 초성과 종성의 환경과 성별로 나누어 음향적으로 구별 가능한 파라미터를 찾고자, 기계적인 분류작업이 가능할 수 있는 새로운 파라미터를 장구간평균 스펙트럼(LTAS; Long term average spectrum)으로 구한 값을 이용하였다. 저주파 대역과 고주파 대역에서의 스펙트럼의 최고치, 스펙트럼의 기울기, zero의 최소진폭과 에너지, 그리고 에너지총량을 구하였다. 또한 스펙트럼 무게중심, 왜도, 첨도를 선정하여 이들의 통계치를 비음 데이터로부터 구하여 분석하였다.

실험결과 각 성별 비음의 환경으로 구분하여 통계적으로 유의미한 파라미터를 요약하면,

첫째, 한국어의 비음 /ㄱ/, /ㄴ/, /ㅇ/에서 각 환경별로 남녀 간의 차이를 보이는 파라미터는 peak1_a, zero_f, slope_1, slope_2, highENG, zero_ENG, centroid가 공통적이다.

둘째, 남성의 경우 초성 /ㄱ/과 /ㄴ/의 비교에서

는 zeroENG가, 종성 /ㄱ/, /ㄴ/, /ㅇ/에서는 peak1_a, peak2_f, lowENG가 유의미한 파라미터이다.

셋째, 여성의 경우 초성 /ㄱ/과 /ㄴ/의 비교에서는 lowENG, zeroENG가, 종성 /ㄱ/, /ㄴ/, /ㅇ/에서는 peak1_a가 유의미한 파라미터이다.

넷째, 양성의 경우 초성비교에서 zeroENG가, 종성비교에서 peak1_a, peak2_f, peak2_a, lowENG가 해당된다.

현 실험의 미비점은 종성 비음을 완벽히 구별해 낼 수 있는 파라미터를 아직 찾지 못했다는 것이며, 토كن 수가 좀 부족하다는 데 있다. 후속 연구에서 종성 비음 모두를 구분할 수 있는 파라미터를 찾아보는 것이 과제라 하겠다.

참고문헌

- [1] Ladefoged, P., *A Course in Phonetics*, New York: Harcourt, Brace, & Jovanovich, 1975
- [2] Fant, G., *Acoustic Theory of Speech Production*, The Hague: Mouton, 1960
- [3] 성철재, "한국어 비음의 음향적 특성에 관한 실험 음성학적 연구", 말소리 31-32호, pp. 9-22, 1996.
- [4] Hayward, K., *Experimental Phonetics*, Pearson Education, 2000.
- [5] Pickett, J. M. *The Sounds of Speech Communication: A Primer of Acoustic Phonetics and Speech Perception*, Baltimore, University Park Press, 1985
- [6] Stevens, K. N., *Acoustic Phonetics*, Massachusetts, The MIT Press, 1998.
- [7] O'shaughnessy, "A study of French Spectral Patterns for Synthesis", *Journal of Phonetics* 10, 377-399, 1982.
- [8] Ladefoged, P., *Phonetic Data Analysis*, Blackwell, 2003
- [9] Johnson, K., *Acoustic & Auditory Phonetics*, Blackwell, 2003.