

대용량 데이터베이스를 이용한 한국어 운율 특성에 관한 연구

A Prosodic Study of Korean Using a Large Database

이 숙 향*, 김 종 진**

(Sook-Hyang Lee*, Jong-Jin Kim**)

*원광대학교 영중어문학부, **한국전자통신연구원

(접수일자: 2005년 2월 1일; 채택일자: 2005년 2월 17일)

본 연구는 대용량 데이터베이스 분석을 통한 한국어의 운율 특성 규명을 시도하였다. 남녀 화자 1명씩 각각 낭독한 650개의 뉴스기사 문장 (총 1,300 개)은 분절을 레이블링과 운율 레이블링을 한 후, 각 운율단위의 성조유형과 크기, 상위 운율단위의 크기와 하위 운율단위 수간의 상관관계, 그리고 강세구의 하강곡선과 상승곡선의 기울기와 F0값에 대한 비교분석을 수행하였다. 운율단위의 지속시간, 어절수, 음절수와 하위 운율단위의 수는 화자뿐만 아니라 운율단위 위치에 따라서 차이를 보였으며 하위 운율단위의 수는 상위 운율단위의 음절수와 지속시간과 높은 상관관계를 보였다. 강세구 내의 하강곡선의 기울기는 강세구 음절수와 반비례 관계를 보였으며 강세구 첫 성조의 유형에 따라 기울기값에 차이가 있었다. 이 차이는 고성조로 시작한 강세구 성조의 F0 상승과 성조 간 상승 폭의 차이로 설명할 수 있었다. 강세구 경계의 하강곡선은 일정하면서 강세구 내의 하강곡선에 비해 다소 급격한 기울기를 보였다. 강세구 내의 두 상승곡선은 유사한 기울기값을 보였으나 강세구 시작 상승곡선의 F0 변화량이 끝 상승곡선의 F0 변화량보다 크게 나타났다. 그리고 하나의 상승곡선이 강세구를 이루는 L Ha의 경우에는 강세구의 음절수에 반비례 관계를 보였다.

핵심용어: 억양구, 강세구, 상승곡선, 하강곡선, 기울기

투고분야: 말소리 생성 및 인지 분야 (12.1)

This study investigates the prosodic characteristics of Korean through the analysis of a large database. One female and one male speakers each read 650 sentences and they were segmentally and prosodically labeled. Statistical analyses were done on these utterances regarding the tonal pattern and the size of prosodic units, correlation between the size of higher level prosodic units and the number of lower level prosodic units, and the slope and F0 of the falling and rising contours of an accentual phrase. The results showed that the duration and the number of words and syllables of a prosodic unit were significantly different not only between speakers but also between its positions within a higher level prosodic unit. The number of a prosodic unit showed a high correlation with the duration and the number of syllables of its higher level units. The slope of the falling contour within an accentual phrase was inversely proportional to the number of its syllables. The slope was different depending on the first tone type of an accentual phrase, which could be explained with the F0 rising and the different amount of rising between tones when an accentual phrase starts with an H tone. The slope of the falling contour across an accentual phrase boundary showed a constant and larger value compared to one within an accentual phrase. The rising contours in the beginning and end of an accentual phrase were similar in their slopes but they differ in the amount of F0 change: the former showed a larger amount of change. The slope of the rising contour which forms an accentual phrase on its own was inversely proportional to the number of its syllables.

Keywords: Intonation phrase, Accentual phrase, Rising contour, Falling contour, slope

ASK subject classification: Speech Production and Perception (12.1)

I. 서론

한국어 운율구조 연구는 오랫동안 음성 음운연구자들에 의해 다양한 접근방법에 의해 이루어졌다[1-8]. Beckman & Jun[9]은 이러한 연구 결과들을 Pierrehumbert[10]의 억양음운론의 이론적 토대에 접목시켜 우리말 음성데이터에 포함된 운율 정보를 체계적으로 기술하는 운율기술체계의 하나인 K-ToBI 레이블링 규약 (Korean Tone and Break Indices Labelling Conventions)을 고안하였다). 이 레이블링 규약은 한국어 발화문장의 운율구조가 계층구조로 이루어져 있으며 하단의 음절층에서 시작하여, 운율단어층 (prosodic word), 강세구 (accentual phrase)층, 그리고 최상단에 억양구 (intonation phrase) 층으로 구성된다고 가정한다. 강세구는 하나 이상의 운율단어로 이루어지며 억양구는 강세구의 상위층으로서 하나 이상의 강세구로 이루어진다. 강세구의 기본 성조는 LHLH이며 강세구의 첫 소리가 유기음이나 경음이면 L 대신 H로 시작하고 강세구 음절수가 4 개 미만일 경우에는 중간 H나 L 성조, 또는 둘 모두 실현이 되지 않기도 한다. 그리고 강세구 끝이 H로 끝나는 경우가 대부분이나 L로 끝나기도 한다. 강세구의 음절수와 첫 소리 그리고 발화속도 등에 따라 결과적으로 14가지의 다양한 피치곡선의 형태로 강세구가 실현된다.

이 이론체계를 바탕으로 최근 10년간 우리말 운율구조에 대한 연구가 활발히 이루어져왔다. 예를 들면 강세구와 억양구 내 각 성조 이벤트의 음성적 특성에 관한 연구 [12]나 강세구 피치곡선상의 성조 간의 기울기에 관한 연구 [13,14], 우리말 억양곡선의 downstep과 declination에 관한 연구 [15], 또는 강세구 내에 나타나는 H 성조의 종류와 환경에 따른 피치 차이 등을 관찰 기술한 연구 [16] 등의 운율단위의 음성적 특성에 대한 정량적 연구를 비롯하여, 조사와 운율단위와의 관계를 살펴 본 연구 [17,18]나 프레이징을 달리 함으로써 분절음상으로 동일한 문장의 의미의 차별화에 관한 연구 [19], 전남, 경상 방언 및 제주방언 등 여러 방언의 운율구조에 대한 연구 [20-22], 성별에 따른 강세구 성조 실현 유형에 관한 연구 [23], 또는 제 2언어로서의 우리말 운율습득 양상에 대한 연구 [24]까지 다양한 연구가 수행되었다. 또한 공학적인 접근으로는 K-ToBI 레이블링 자료를 근거로 문장

의 F0 곡선을 생성하려는 시도 [25] 등이 있다. 이러한 선행 연구들은 우리말 운율구조의 개괄적인 특성을 기술하는 것이 주목적이기 때문에 길이가 짧고 통사적으로 단순한 문장들만을 분석 대상으로 하였으며 분석 문장의 수 또한 많지 않았다. 본 연구는 문장의 길이와 통사적 구조가 다양한 비교적 많은 수의 문장을 대상으로 하여 우리말 운율구조의 특성을 정량적으로 기술하고자 한다. 특히 강세구 내와 강세구 경계에서 성조 간에 나타나는 피치 변화, 즉 피치곡선의 상승과 하강의 양에 대한 비교분석을 주목적으로 한다. 즉 강세구 내에 최대 2 개의 상승 (첫 두개 음절에 나타나는 상승 (L +H)과 마지막 두개 음절에 나타나는 상승 (L+ Ha))과 2개의 하강 (예, 둘째음절과 끝 둘째음절 사이에 나타나는 하강 (+H L+)과 마지막 음절에서 다음 강세구 첫음절까지의 전이구간 (Ha L))이 나타날 수 있다.²⁾ 또는 강세구의 음절수가 적어서 중간 성조가 실현되지 않을 경우 상승곡선 하나 (L Ha)만으로 실현되는 강세구도 있을 수 있다. 이들은 각각 다른 종류의 상승 또는 하강이기 때문에 대부분 청각적으로 분간이 가능하나 때로는 청각혼란이 될 경우에는 구분이 쉽지 않을 수도 있다. 즉, 예를 들어 하나의 상승곡선이 과연 하나의 독립된 강세구를 이루는지 (L Ha) 또는 강세구 시작 부분의 상승곡선 (L +H)인지 등 판단이 어려울 수 있으며, 하나의 하강곡선도 강세구 내의 하강곡선 (+H L+)인지 아니면 강세구 경계에 있는 하강곡선 (Ha L)인지 구분이 어려울 수도 있다. 따라서 이들에 대한 정량적 분석을 통하여 객관적 기준을 제공함으로써 청각적 판단에 보조적 수단이 될 수 있을 것이며 더 나아가서 우리말 강세구의 피치곡선에 대한 보다 정확한 기술을 통하여 운율 합성이나 인식에서 강세구 생성과 강세구 경계에 대한 판단에 중요한 자료를 제공할 수 있을 것이다.

II. 연구방법

실험에 참여한 화자는 남녀 각각 1명씩 2명이며 (앞으로 여자화자는 KSH로, 남자화자는 KSM으로 칭할 것임) 한국 내 학부 재학생으로서 서울 출신 20대이며 녹음 당

1) 수정본원을 거쳐 Sun-Ah Jun(11)이 2000년에 3차 버전을 썼으며 본 연구는 이 버전이 제시하는 이론적 근거와 기술방법을 따른다.

2) K-ToBI에서 강세구의 첫째 음절에 실현되는 저성조 또는 고성조는 'L'이나 'H'로, 보통 둘째음절에 실현되는 고성조는 '+H'로, 끝에서 두 번째 음절에 실현되는 저성조는 'L+'로 기술하며, 마지막 고성조 또는 저성조는 'Ha' 또는 'La'로 표기한다. 그리고 억양구 경계성조는 성조에 %를 더하여 기술한다 (예, H%, L% 등).

표 1. 발화문장의 평균 지속시간 및 억양구 수

Table 1. Utterance mean durations and the number of intonation phrases in an utterance.

운율특성	화자	평균 (표준편차)	동일문장 간 상관관계	화자 간 t-test 결과
지속시간	KSH	9445.78 (4403.73)	.981	.000
	KSM	8519.91 (3822.70)		
억양구 수	KSH	7.04 (3.12)	.867	.000
	KSM	5.89 (2.58)		

시 대학 방송반에서 아나운서로 활동하고 있었다. 두 화자 각각 1,000 문장을 한 달에 걸쳐 녹음을 하였으며 문장의 길이와 통사적 구조는 매우 다양하며 모두 평서문이다. 문장 내 음절수는 문장에 따라 16개에서 172개에 이르며 어절수는 5개에서 57개에 이르고 있다. 문장의 내용은 방송 뉴스에서 따온 것으로서 두 화자에게 방송 아나운서가 읽듯이 낭독해달라고 주문하였다. 녹음은 방음처리된 녹음실에서 TASCAM DAT 녹음기와 Sennheiser 다이나믹 마이크를 이용해 44.1kHz로 녹음한 후, Ariel DAT-Link+를 이용해 16kHz, 16bits로 다운샘플링하여 사용하였다. 분절음 및 운율 레이블링은 본 연구의 제 1 저자가 수동으로 레이블링했으며 레이블링된 자료에서 운율특징을 자동으로 추출하는 프로그램은 제 2저자가 썼다. 운율 레이블링은 K-ToBI 레이블링 규약에 따랐다. 본 연구의 분석대상 자료로 사용된 문장 수는 남녀 각각 1,000문장 중 650 문장으로서 총 1,300 문장이다. 통계분석은 SPSS 10.0을 이용하였으며 유의수준은 $P < .05$ 로 하였다.

III. 결과 및 논의

3.1. 운율단위의 특성

3.1.1. 발화문장

682 개의 발화문장의 지속시간과 발화문장 내 억양구 수의 분포를 화자별로 분석 비교하였다. 표 1에서 볼 수 있듯이 paired-samples *t*-test 결과 두 화자의 동일한 발화문장들의 지속시간 간에는 .981이라는 높은 pearson 상관계수를 보였으며 남녀 화자 간에 유의한 차이를 보였다. 남자화자 KSM의 지속시간 (8519.91ms)이 여자화자 KSH (9445.78ms)에 비해 짧게 나타났는데 이는

표 2. 억양구 경계성조의 화자별 출현빈도

Table 2. Occurrence frequency of intonation phrase boundary tones for each speaker.

IP경계성조	빈도수(백분율)	
	화자 KSH	화자 KSM
HL%	1818 (40.80)	2307 (61.67)
L%	1191 (26.73)	1051 (28.09)
H%	814 (18.27)	368 (9.84)
LH%	630 (14.14)	11 (.29)
LHL%	2 (0.04)	0 (0)
HLH%	1 (0.02)	4 (0.11)
불확실	120	90
합계	4576	3831

KSH의 발화속도가 KSM에 비해 느리거나 화자 KSH가 화자 KSM에 비해 한 문장을 더 많은 억양구로 나누어 발화하거나 또는 두 요인이 한꺼번에 작용한 결과일 수 있겠다³⁾. 두 화자의 동일한 발화문장들 내의 억양구 수 간에도 역시 .867이라는 높은 상관관계를 보였으며 두 화자 간에 유의한 문장 당 평균 억양구 수 차이를 보였다. 화자 KSH는 7.04개, 화자 KSM은 5.89개로 두 화자 간에 1개 정도의 차이를 보이고 있다.

3.1.2. 억양구

억양구의 구말 성조 유형을 살펴보면 표 2에 나타나 있듯이 2 가지 단순경계성조 H%, L%와 4 가지 복합경계성조 HL%, LH%, LHL%, HLH%이 사용되고 있다. 이 중 HL% 경계성조가 두 화자 모두에서 가장 높은 출현빈도를 보이고 있으며 L%와 H%가 그 다음 순으로 많이 사용되고 있다. 본 연구의 분석 대상 발화문장이 모두 평서문 형태라서 내림조로 끝날 것을 감안하면 HL% 또는 L%가 빈번히 나타나는 것은 예측 가능한 일이나 H%와 LH%의 출현빈도 또한 높다는 것은 (LH%의 높은 출현빈도는 여자화자에만 해당됨) 문장 중간의 억양구에서 이들 경계성조가 종종 사용되었음을 말해주고 있다. 억양구말 경계성조의 유형과 빈도에서 두 화자 간에 전반적으로 같은 유형을 보이고 있으나 몇 가지 차이 또한 관찰할 수 있다. 화자 KSH는 HL%, L%, H% 외에 LH%도 많이 사용하고 있는 반면 (14.14%) 화자 KSM은 거의 사용하지 않고 있다 (.29%). 또한 출현빈도가 가장 높은 HL%인 경우에도 화자 간에 큰 차이를 보이고 있으며 화

3) 억양구는 구말 장음화에 의해 마지막 음절이 길어질 뿐만 아니라 억양구 사이에 무음간격이 올 수도 있기 때문에 같은 문장을 보다 많은 억양구를 두어 읽게 되면 문장 전체 지속시간이 길어지는 한 요인이 될 수 있다. 참조로 우리말 강세구에서는 구말 장음화 현상이 일어나지 않는다.

표 3. 억양구의 평균 지속시간과 억양구 내 평균 강세구 수, 어절수, 음절수
Table 3. Mean durations of intonation phrases and the number of accentual phrases, words, and syllables within an intonation phrase.

운율특성	화자	평균 (표준편차)	화자 간 t-test 결과
지속시간	KSH	1173.64 (522.37)	.000
	KSM	1273.39 (539.02)	
강세구 수	KSH	1.84 (.89)	.000
	KSM	2.07 (.95)	
어절수	KSH	3.31 (1.36)	.000
	KSM	3.76 (1.50)	
음절수	KSH	7.74 (3.85)	.000
	KSM	9.23 (4.23)	

표 4. 강세구 마지막 성조 유형 및 출현빈도
Table 4. Tonal types of the last tone of an accentual phrase and their occurrence frequency.

강세구 경계성조	출현빈도 (백분율%)	
	KSH	KSM
H	3683 (95.94)	3866 (94.36)
L	164 (4.07)	231 (5.64)
합계	3847	4097

자 KSM (61.67%)이 화자 KSH (40.80%)에 비해 훨씬 자주 사용하는 것으로 나타났다.

억양구 내의 운율적 특성으로서는 표 3에서 볼 수 있듯이 억양구의 지속시간과 억양구 내 강세구 수, 어절수, 음절수의 성별 간 차이에 대한 independent samples *t*-test 결과 화자 KSM의 억양구에서 강세구 수와 어절수, 그리고 음절수가 화자 KSH의 억양구에 비해 많으며 화자 KSM의 억양구의 지속시간 또한 화자 KSH에 비해 길게 나타났다. 주목할 만한 것은 화자 KSM의 한 문장 내의 억양구 수는 화자 KSH에 비해 적은 대신 한 억양구 내의 강세구 수는 많다는 점이다.

3.1.3. 강세구

표 4에 제시된 바와 같이 대부분의 강세구 마지막 성조는 Ha가 약 95%, La가 약 5%의 출현빈도를 보이고 있다.

표 5와 6은 억양구 중간에 오는 강세구와 억양구 끝에 오는 강세구의 성조 유형과 출현빈도를 보여주고 있다. 두 화자 간에 유사한 분포를 보이고 있는데 억양구 중간의 강세구의 경우에는 L Ha, L+H L+Ha, H L+Ha가 가장 높은 출현빈도를 보이고 있다. 반면 억양구 말 강세구의 경우에는 H L+ HL%, L+H L+ HL%, H L%, L+H L%의 유형이 많이 나타나고 있다.

강세구의 지속시간과 음절수 및 어절수에 대한 분석은

표 5. 억양구 중간 강세구 성조 유형 및 출현빈도
Table 5. Tonal types of accentual phrases in the intonation phrase medial positions and their occurrence frequency.

강세구 성조유형	출현빈도 (백분율%)	
	화자 KSH	화자 KSM
L Ha	1177 (30.60)	1379 (33.66)
L+H L+Ha	928 (24.12)	1016 (24.80)
H L+Ha	1152 (29.95)	824 (20.11)
L L+Ha	139 (3.61)	272 (6.64)
H+H L+Ha	277 (7.20)	356 (8.69)
H La	91 (2.37)	100 (2.44)
L+H La	41 (1.07)	73 (1.78)
H+H La	20 (.52)	50 (1.20)
H Ha	8 (.21)	8 (.20)
L La	4 (1.10)	5 (1.2)
L+H Ha	2 (.05)	10 (2.4)
H L+La	5 (1.3)	2 (.05)
H+H L+La	2 (.05)	2 (.05)
L+H L+La	1 (.03)	0 (0)
총합	3847	4097

표 6. 억양구말 강세구 성조 유형 및 출현빈도
Table 5. Tonal types of accentual phrases in the intonation phrase final positions and their occurrence frequency.

강세구 성조유형 ¹⁾	출현빈도 (백분율%)	
	화자 KSH	화자 KSM
H L+ HL%	730 (16.38)	815 (21.79)
L+H L+ HL%	607 (13.62)	873 (23.33)
H L%	607 (13.62)	427 (11.41)
L+H L%	476 (10.68)	444 (11.87)
H L+ H%	266 (5.97)	101 (2.70)
L+H L+ LH%	289 (6.49)	2 (.05)
H L+ LH%	267 (5.99)	6 (.16)
H+H L+ HL%	232 (5.20)	358 (9.57)
L+H L+ H%	214 (4.80)	65 (1.74)
L HL%	209 (4.69)	130 (3.48)
L+H H%	197 (4.42)	106 (2.83)
H+H L%	93 (2.09)	121 (3.23)
H H L+ LH%	63 (1.41)	3 (.08)
H+H L+ H%	49 (1.10)	40 (1.07)
H H%	40 (0.90)	21 (.56)
L H%	36 (.81)	20 (.53)
L L+ HL%	34 (.76)	116 (3.1)
L L%	15 (.34)	42 (1.12)
L L+ H%	11 (.25)	15 (.40)
기타 ¹⁾	21 (.47)	36 (.96)
합계	4456	3741

억양구 말의 장음화 현상 등 억양구 내의 위치에 따라 강세구의 특성이 다르게 나타날 가능성을 고려하여 억양구 내의 위치에 따라 억양구 중간에 오는 강세구와 억양구 말에 오는 강세구 두 그룹으로 나누어 특성을 살펴보았다. 표 6에서 볼 수 있듯이 억양구 중간에 오는 강세구의 지속시간은 두 화자 간에 유의한 차이를 보이지 않은 반면 음절수와 어절수는 화자 간에 유의한 차이를 보이며 화자 KSM에서 더 많은 것으로 나타났다. 이는 화자 KSH의 발화속도가 화자 KSM에 비해 느리다는 것을 보

표 7. 강세구의 지속시간과 강세구 내 어절수와 음절수 평균
Table 7. Mean durations of accentual phrases and the number of words and syllables within an accentual phrase.

운율 특성	화자	억양구 중간 강세구		억양구말 강세구		억양구 내 위치 간 t-test 결과
		평균 (표준 편차)	화자 간 t-test 결과	평균 (표준 편차)	화자 간 t-test 결과	
지속 시간 (ms)	KSH	496.47 (193.69)	.386	761.82 (227.44)	.099	.000
	KSM	492.70 (193.98)		753.34 (235.21)		
음절 수	KSH	3.60 (1.41)	.000	4.75 (1.65)	.000	.000
	KSM	3.80 (1.51)		5.22 (1.86)		
어절 수	KSH	2.13 (.58)	.006	2.37 (.71)	.000	.000
	KSM	2.16 (.64)		2.52 (.76)		

여주고 있다. 억양구 말에서 또한 억양구 중간에서와 마찬가지로 강세구의 지속시간은 유의한 차이를 보이지 않았으나 음절수와 어절수는 두 화자 간에 유의한 차이를 보였으며 KSM에서 어절수와 음절수가 더 많은 것으로 나타났다. 이 결과는 앞에서 화자 KSH의 발화문장 전체 지속시간이 화자 KSH에 비해 긴 이유 중 하나를 제시하고 있다고 할 수 있다. 즉, 앞에서 보았듯이 화자 KSH의 한 문장 내의 억양구 수가 화자 KSM보다 더 많은 것이 한 가지 이유이며 또 한 가지 이유로는 화자 KSH의 발화속도가 화자 KSM에 비해 상대적으로 느리기 때문이다. 다음으로 강세구의 억양구 내 위치에 따라 강세구의 지속시간, 음절수 및 어절수가 달리 나타나는지를 보기 위하여 t-test를 실시하였다. 두 화자 모두 두 위치 간에 유의한 차이를 보였으며 억양구 말 강세구에서 억양구 중간 강세구에 비해 음절수와 어절수가 더 많고 지속시간 또한 더 길었다. 특히 지속시간에서 상당한 차이를 보이고 있는데 이는 억양구말 강세구는 억양구 중간 강세구에 비해 크기가 클 뿐만 아니라(음절수와 어절수가 많음) 억양구말에서 구말 장음화가 일어남으로 인한 결과로 볼 수 있겠다.

3.2. 운율단위 길이와 운율단위 수 간의 상관관계

각 운율단위 안에 나타나는 하위 단위의 수와 각 단위 내의 어절수와 음절수, 그리고 지속시간 간에 어떤 상관관계를 보이는가를 보기 위하여 bivariate 상관분석을 실시하였다. 표 8에서 볼 수 있듯이 발화문장 내 억양

표 8. 발화문장 내 억양구 수와 어절수, 음절수, 지속시간 간의 상관관계
Table 8. Correlation of the number of intonation phrases within an utterance with its number of words and syllables and duration.

화자	어절수	음절수	지속시간
KSH	.904	.916	.938
KSM	.853	.869	.910

표 9. 억양구 내 강세구 수와 어절수, 음절수, 지속시간 간의 상관관계
Table 9. Correlation of the number of accentual phrases within an intonation phrase with its number of words and syllables and duration.

화자	어절수	음절수	지속시간
KSH	.904	.916	.938
KSM	.853	.869	.910

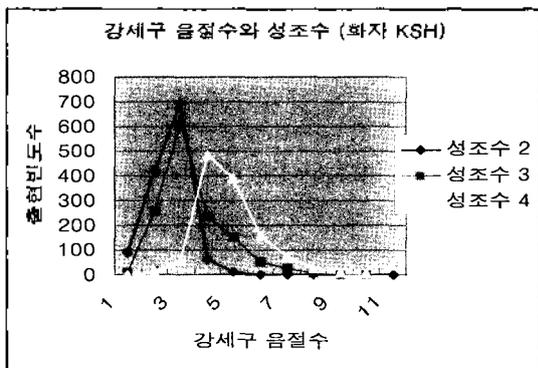
화자	어절수	음절수	지속시간
KSH	.761	.840	.838
KSM	.745	.827	.828

구 수는 두 화자 모두에서 세 가지 특성 모두와 높은 상관관계를 보이며 그 중 지속시간과의 상관관계가 가장 높은 것으로 나타났다. 한편 억양구 내의 강세구 수의 경우는 발화문장 내의 억양구 수의 경우에 비해 비교적 상관관계가 낮으며 세 특성 중 음절수와 지속시간이 강세구 수와 상관관계가 높은 것으로 나타났다 (표 9 참조). 세 가지 특성 중 어절수와 음절수는 한 문장이나 억양구를 어떻게 읽느냐와는 상관없이 그 수가 이미 주어진 변수이지만 지속시간은 발화속도와 몇 개의 하위단위를 두느냐에 따라 다르게 나타나는 변수이다. 따라서 이미 읽은 문장의 운율단위 수 예측에는 지속시간과 음절수가 중요한 역할을 할 수 있지만 텍스트만 주어진 상태에서의 예측에는 음절수가 중요한 역할을 한다고 할 수 있겠다. 강세구 내의 음절수와 성조 수와의 관계는 표 10에서 볼 수 있듯이 음절수가 넷 이상이면 4개 성조 모두 실현되는 경향이 뚜렷해지며 음절 수가 하나나 둘인 강세구는 2개의 성조만 실현되는 경우가 주를 이루는 반면, 음절수가 셋인 강세구는 화자 KSH에서는 2개 또는 3개의 성조가 실현되는 경우가 주를 이루고 있고 화자 KSM에서는 3개 성조가 실현되는 경우보다는 2개 성조가 실현되는 경우수가 더 크게 나타나고 있다. 즉, 그림 1에서 볼 수 있듯이 성조가 2개 실현된 강세구와 3개 실현된 강세구는 음절수 3에서 출현 빈도의 최고치를 보이고 있는 반면, 성조가 4개인 강세구는 음절수 4에서 출현 빈도의 최고치를 보이고 있다. 특기할만한 사실은 음절수가 3인 경우의 강세구 내에 실현되는 성조의 수가 2개인 경우수와 3개인 경우수가 유사하거나 (화자 KSH) 2개

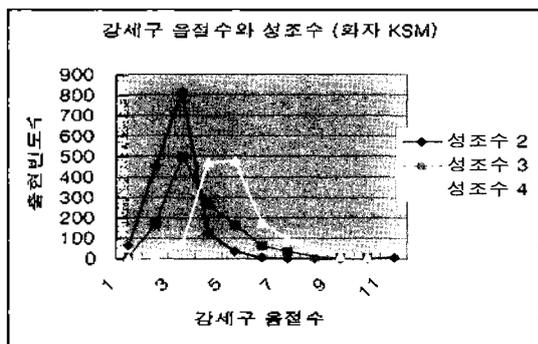
표 10. 강세구 내 음절수와 성조 수와의 관계
Table 10. The relation between the number of tones and syllable within an accentual phrase.

화자	성조 수	음절수					
		1	2	3	4	5	6
KSH	2	90	419	696	65	9	1
	3	8	259	612	238	154	53
	4	0	6	56	486	391	157
	합계	98	684	1364	789	554	211
KSM	2	62	449	810	126	34	7
	3	5	172	499	265	163	67
	4	0	9	74	473	477	179
	합계	67	130	1383	864	674	253

화자	성조 수	음절수					
		7	8	9	10	11	합계
KSH	2	0	0	0	0	1	1281
	3	26	7	2	0		1359
	4	72	26	9	4		1207
	합계	98	33	11	4	1	3847
KSM	2	1	2	0	1	4	1496
	3	34	13	5	8		1231
	4	97	43	13	5		1370
	합계	132	58	18	14	4	4097



(a)



(b)

그림 1. 강세구 내 음절수와 성조 수와의 관계. (a): 화자 KSH, (b): 화자 KSM

Figure 1. The relation between the number of tones and syllable within an accentual phrase. (a): speaker KSH, (b): speaker KSM.

성조가 실현되는 경우수가 더 크게 나타났다는 점이다. (화자 KSM) 또한 음절수가 2인 강세구에서 성조 3개가 실현된 경우수와 음절수가 4인 강세구에서 3개 성조만 실현된 경우수도 비교적 크다는 점 또한 주목할 만하다. 음절수가 2에서 4인 강세구들은 음절수와 성조수가 1:1 관계를 보이는 경우가 일반적이긴 하나 그렇지 않은 경우도 종종 나타나고 있어 이를 결정하는 요인에 어떤 것이 있는지는 추후 연구에서 밝혀져야 할 것이다.

3.3. 강세구의 상승곡선과 하강곡선의 기울기⁴⁾

3.3.1. 하강곡선 +H L+ 과 Ha L 기울기 비교⁵⁾

강세구 내와 강세구 경계에서 나타나는 하강곡선 (앞으로 각각 +H L+, Ha L로 칭할 것임)은 단순히 청각적으로나 시각적으로 구분하기가 쉽지 않은 경우가 종종 있어 강세구 경계에 어려움을 주기도 한다. 보다 정확한 강세구 경계 판별을 위해 두 성조 간의 F0 (Hz) 차이를 두 성조 간의 지속시간 (ms) 차이로 나누어 두 가지 하강곡선의 기울기를 비교분석하였다. 그림 2와 그림 3에서 볼 수 있듯이 두 화자 모두 +H L+ 하강곡선은 강세구의 음절수가 많아질수록 기울기가 작아져 완만한 하강곡선을 그리는 반면 Ha L 하강곡선은 강세구 음절수와 상관없이 일정한 기울기를 보였다. 두 하강 곡선 모두에서 두 화자 간에 기울기의 차이를 보이고 있는데 대체적으로 화자 KSH의 기울기가 KSM의 기울기보다 큰 값을 보여 화자 KSH가 화자 KSM에 비해 큰 피치 레인지를 쓰는 것을 간접적으로 보여주고 있다고 할 수 있겠다. 기울기 분석은 +H L+ 하강곡선의 경우, 강세구가 저성조 L로 시작하는 경우와 고성조 H로 시작하는 경우 둘로 나누어 분석하였다. 평균기울기는 화자 KSH에서 저성조로 시작한 경우와 고성조로 시작한 경우 각각 -0.19와 -0.21이고 최소-최대값 범위는 각각 -0.10~-0.31과 -0.13~-0.39이었다. 화자 KSM은 저성조로 시작하는 경우 평균기울기는 -0.10이고 최소-최대값 범위는 -0.09~-0.24였으며 고성조로 시작하는 경우의 평균은 -0.15이고 범위는 -0.09~-0.30였다. 이는 Jun(13)의 결과 (-0.08~-0.2)와 거의 유사하다. 한 가지 특기할 만한 사실은 대체적으로 고성조로 시작하는 강세구에서의 기울기가 저성조로 시작하는 강세구에서의 기울기보

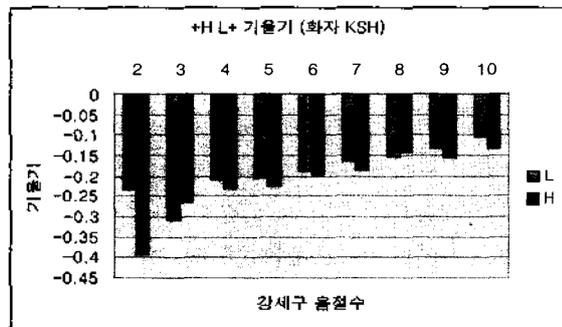
4) 강세구가 저성조로 끝나는 La형은 출현빈도가 아주 낮기 때문에 논의대상에서 제외하였다.
5) 기울기 논의에서는 출현빈도가 아주 낮은 음절수는 논의대상에서 제외하였다.

다 값이 크게 나타나고 있다. 이런 차이를 초래하는 가능한 요인으로 몇 가지를 가정해볼 수 있겠다. 그 중 한 가지는 +H의 F0값 차이로서 고성조로 시작하는 강세구의 +H의 F0값이 저성조로 시작하는 강세구의 +H의 F0값에 비해 크기 때문일 가능성을 들 수 있겠고, 또 다른 가능성으로는 강세구가 고성조로 시작할 때 +H의 F0값은 일정하나 +L의 F0값이 많이 떨어져 이런 결과를 초래할 가능성이 있을 수 있겠다. 또는 강세구가 고성조로 시작하는 경우 두 성조의 F0 모두 상승이 있으나 상승의 폭이 달라서 오는 결과일 수도 있겠다. 이를 위한 F0 비교분석은 다음 절에서 다룬다. Ha L 하강곡선의 평균기울기는 화자 KSH에서 -0.32, KSM에서 -0.18이었으며 +H L+의 기울기에 비해 대체적으로 큰 값을 보인다. 이 값은 Jun[13]의 연구결과가 제시하는 값 (-0.5~-0.7)과는 큰 차이를 보이고 있는데 이는 단순히 화자 간 차이일 수도 있으나 실험 환경의 차이에서 오는 결과일 가능성 또한 있겠다. 즉, 두 연구 모두 실험실에서 이루어진 낭독체 문장이긴 하나 Jun[13]의 연구에서의 분석 대상 문장은 길이가 무척 짧으며 목표단어만 달라하는 틀문장을 낭독하는 반면, 본 연구의 분석 대상

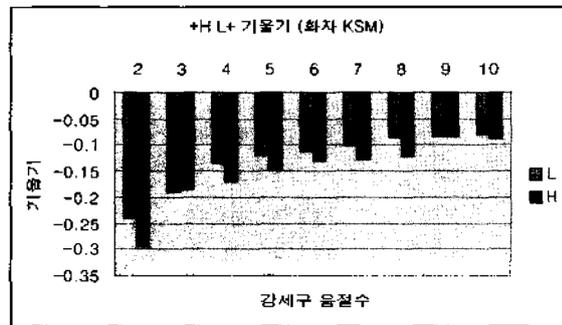
대부분의 문장은 길이가 무척 길고 통사적으로도 복잡하며 아나운서가 뉴스기사처럼 읽는 상황이었다. 강세구간의 피치레인지의 downstep과 문장 전체에 걸쳐 나타나는 declination을 고려할 때, 한 문장 내에 나타나는 강세구 수가 많을 경우 화자의 피치 레인지 내에서 많은 수의 강세구 경계의 하강곡선이 다 실현되기 위해서는 피치의 변화가 그리 클 수가 없을 것이며 따라서 기울기 값이 작게 나타날 것이다. 이 가능성에 대한 연구는 추후 연구에서 보다 자세히 이루어져야 하나, 실험 환경 또는 화자 등 어떤 요인에 의한 것인지는 불분명하나, 여러 요인에 의해 분석 결과에 상당한 차이를 보일 수 있음을 보았다.

3.3.2. 상승곡선 L +H, L+ Ha, L Ha 기울기 비교

강세구에 나타날 수 있는 상승곡선은 첫 두개 음절에 실현되는 L +H, 마지막 두개 음절에 실현되는 L+ Ha, 그리고 두 개의 중간 성조가 실현되지 않고 나머지 두 개의 성조로만 강세구를 이루는 L Ha가 있다. 이들의 기울기에 대한 비교 분석 결과, 그림 4-6에 나타나 있듯이 L +H와 L+ Ha는 음절수에 상관없이 일정한 기울기 값을 보이는 반면 L Ha의 기울기값은 음절수가 증가함에 따라 감소하여 완만한 상승곡선을 그린다. 하강곡선에서와 마찬가지로 화자 KSH의 기울기가 화자 KSM에 비해 큰 값을 보인다. L +H 평균 기울기값은 화자 KSH에서 0.27, 화자 KSM에서 0.17이었으며 L+ Ha의 평균 기울기값은 화자 KSH에서 0.30, 화자 KSM에서 0.15이었다. 즉, L +H와 L+ Ha는 모두 보통 2개의 음절 사이에 실현되는 상승곡선이며 평균 기울기값 또한 거의 유사하여 기울기만으로는 서로 구분이 되지 않는 것으로 나타났다. L Ha 평균 기울기값은 화자 KSH에서 0.11, 화자 KSM에서 0.08이었으며 최소-최대값 범위는 각각 0.009 ~0.32, 0.02 ~0.22이었다.



(a)



(b)

그림 2. 하강곡선 +H L+의 기울기. 범례 L은 저성조로 시작하는 강세구를 의미하며 범례 H는 고성조로 시작하는 강세구를 의미함. (a): 화자 KSH, (b): 화자 KSM

Figure 2. The slope of +HL+. L means an accentual phrase whose first tone is H and H means an accentual phrase whose first tone is L. (a): speaker KSH, (b): speaker KSM.

3.4. 강세구 성조의 F0값 비교

3.4.1. 강세구 첫 성조의 유형에 따른 강세구 내 성조의 F0 값 변화

앞 절에서 강세구의 첫 성조의 유형에 따라 +H L+의 기울기값이 차이를 보였다. 즉, 고성조로 시작할 때의 기울기값이 저성조로 시작할 때의 기울기값보다 크게 나타났고 이에 대한 한 가지 가능한 요인으로 강세구 첫 성조의 유형에 따른 H+의 F0값, 또는 L+의 F0값 차이

또는 두 성조 F0 변화의 크기 차이를 제시했다. 이 차이는 H.J. Lee[16]에서 논의된 바 있으며 강세구 첫 성조의 유형이 보통 강세구 둘째 음절에 실현되는 +H의 F0만이 아니라 강세구 끝에서 둘째음절에 실현되는 +L의

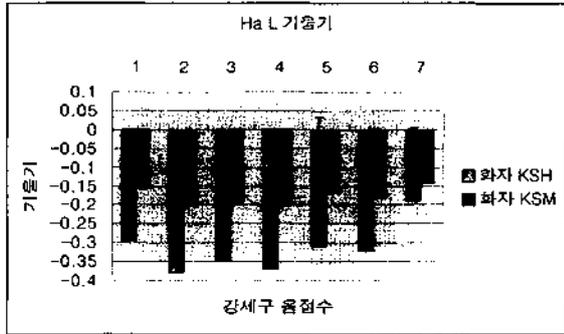


그림 3. 하강곡선 Ha L의 기울기
Figure 3. The slope of Ha L.

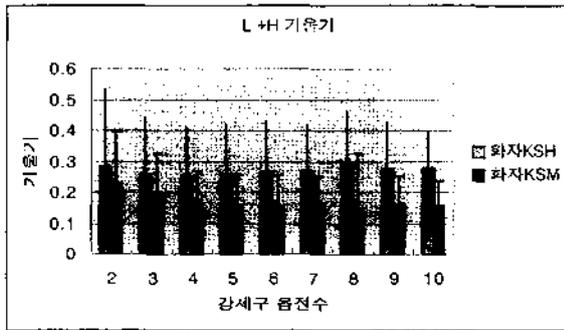


그림 4. 상승곡선 L + H의 기울기
Figure 4. The slope of L + H.

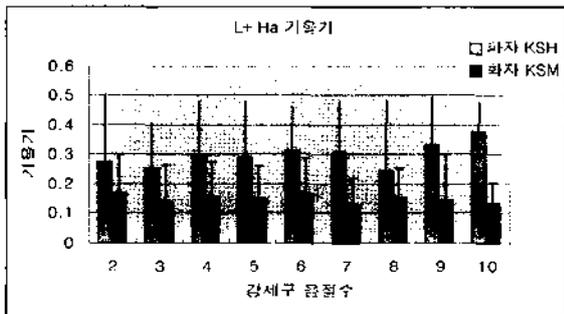


그림 5. 하강곡선 L + Ha의 기울기
Figure 5. The slope of L + Ha.

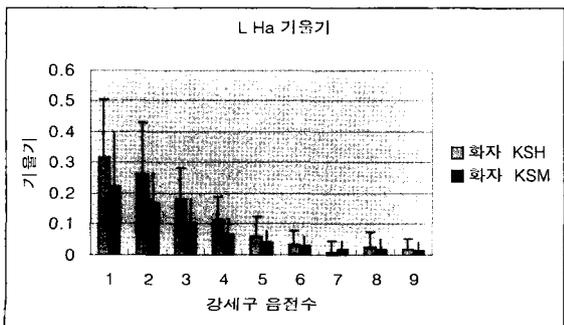


그림 6. 하강곡선 L Ha의 기울기
Figure 6. The slope of L Ha.

표 11. 강세구 첫 성조 유형에 따른 F0값 비교 분석 결과 (분석대상 강세구 수는 첫 성조가 L일 때 화자 KSH 1033, 화자 KSM 1096, H일 때 화자 KSH 299, 화자 KSM 376개임)

Table 11. F0 values of tones of accentual phrases depending on its first tone types.

성조	첫 성조 유형	화자 KSH		화자 KSM	
		평균 (표준편차)	t-test 결과	평균 (표준편차)	t-test 결과
L/H	L	241.56 (26.93)	.000	122.72 (17.56)	.000
	H	286.70 (34.49)		154.27 (24.09)	
+H	L	287.10 (25.69)	.000	149.74 (16.26)	.000
	H	301.88 (31.61)		165.16 (21.65)	
L+	L	240.52 (30.43)	.003	127.02 (17.06)	.000
	H	246.58 (31.16)		133.02 (19.94)	
Ha	L	276.76 (29.83)	.004	144.85 (17.25)	.000
	H	282.34 (29.13)		150.58 (18.70)	

F0값에까지 영향을 미친다고 하였다. 즉 강세구가 고성조로 시작하면 전반적으로 강세구 전체의 F0값의 상승을 초래하여 강세구 셋째 성조에까지 영향을 미친다고 하였다. 본 연구는 강세구 첫 성조에 따른 +H L+하강곡선의 기울기의 차이를 설명하기 위해 4개의 성조가 모두 실현되면서 Ha로 끝나는 강세구들만을 대상으로 F0값을 분석하였다. 그 결과, 표 11에 제시된 바와 같이 화자 간에 큰 F0값 차이를 보였으며 두 화자 모두 강세구 첫 성조 L과 H의 F0값 간에 유의한 차이가 있었다. 뿐만 아니라 둘째 성조에서 마지막 성조까지 모든 성조에서 유의한 차이를 보였다. 즉 강세구가 고성조로 시작된 경우 +H와 L+상승만이 아니라 마지막 성조 Ha의 상승 또한 있었다. 상승의 폭은 +H에서 14~15Hz, L+에서 6Hz, Ha값에서 6Hz 정도로 나타났다. 즉 강세구가 고성조로 시작할 때 +H L+ 하강곡선의 기울기값이 큰 것은 +H의 F0 상승 폭과 L+의 F0 상승 폭이 다르기 때문이라고 할 수 있겠다.

3.4.2. L vs. L+, +H vs. Ha

앞 절에서 L + H 와 L+ Ha는 보통 2개의 음절 사이에 실현되는 상승곡선이며 평균 기울기값 또한 거의 유사하여 (L + H: 화자 KSH 0.27, 화자 KSM 0.17, L+ Ha: 화자 KSH 0.30, 화자 KSM 0.15) 기울기만으로는 서로 구분이 되지 않는 것으로 나타났다. 이 두 곡선이 F0값에서 차이를 보이는지 보기 위하여 L과 L+, +H

표 12. 강세구 내 두 상승곡선의 성조간 F0값 비교 분석 결과 (분석대상 강세구 수는 화자 KSH 1033, 화자 KSM 1096개임)

Table 12. F0 values of tones of the two rising contours within an accentual phrase.

성조	화자 KSH		화자 KSM	
	평균 (표준편차)	t-test 결과	평균 (표준편차)	t-test 결과
L	241.5644 (26.93)	.33	122.719 (17.56)	.000
L+	240.5227 (30.43)		127.0155 (17.06)	
+H	287.1045 (25.69)	.000	149.7363 (16.26)	.000
Ha	276.7638 (29.83)		144.8476 (17.25)	

와 Ha 간의 F0값을 paired samples t-test를 통하여 비교분석하였다. 표 12에서 볼 수 있듯이 화자 KSH의 L과 L+를 제외하고 유의한 차이를 보였다. 화자 KSM의 L은 L+보다 낮으며 두 화자 모두 +H의 F0값이 Ha의 F0값보다 높게 나타났다. 즉, 강세구 시작 상승곡선은 끝 상승곡선에 비해 큰 F0 차이를 보이면서 저성조는 끝 상승곡선의 저성조와 같거나 더 낮고 고성조는 끝 상승곡선의 고성조에 비해 더 높게 실현되었다. 따라서 두 상승곡선은 기울기는 비슷하나 F0값에서 차이를 보였다.

IV. 결론

본 연구는 대용량 데이터베이스 분석을 통하여 한국어의 운율 특성을 밝혀보고자 하였다. 첫째, 각 운율 단위의 성조 유형과 크기, 즉 지속시간, 어절수, 음절수 등에 대한 비교분석 결과, 운율단위의 지속시간, 어절수, 음절수 등에서 화자 간 차이가 있었으나 성조유형 및 출현 빈도에는 화자 간 유사한 결과를 보였다. 일반적으로 여성화자 KSH의 발화속도가 남성화자 KSM에 비해 느리게 나타났으며 화자 KSH가 화자 KSM에 비해 한 문장 내의 억양구 수는 많은 반면 한 억양구 내의 강세구 수는 적었다. 강세구의 억양구 내 위치에 따른 차이 또한 관찰되었는데 억양구말 강세구의 지속시간이 억양구 중간의 강세구에 비해 훨씬 큰 값을 보였다. 이는 화자들이 억양구 중간의 강세구에 비해 억양구말의 강세구에 보다 많은 음절과 어절을 할당할 뿐만 아니라 억양구말에서 구말 장음화가 일어나기 때문에 일어나는 현상으로 설명할 수 있었다. 운율단위의 성조 유형을 보면, 억양구의 경계성조의 경우, HL%의 출현빈도가 가장 높았고

L%와 H%도 비교적 높은 출현빈도를 보였으며 화자 KSH의 경우에는 LH%도 종종 사용되었다. 강세구는 L Ha, L +H L+ Ha, H L+ Ha 유형이 주를 이루었다.

둘째, 상위 운율단위의 운율 특성과 하위 운율단위의 개수와의 관계를 살펴본 결과, 하위 운율단위의 개수는 상위 운율단위의 음절수와 지속시간과 높은 상관관계를 보임으로써 이 두 변수가 하위 운율단위의 개수를 예측하는 데 중요한 변수임을 알 수 있었다.

셋째, 강세구에 나타나는 여러 상승곡선과 하강곡선의 기울기값과 F0값에 대한 비교분석 결과, 강세구 내의 하강곡선의 기울기는 강세구의 음절수와 반비례 관계를 보인 반면 강세구 경계의 하강곡선의 기울기는 일정하면서 강세구 내 하강곡선의 기울기보다 다소 큰 값을 보였다. 강세구 내의 하강곡선의 기울기는 강세구 첫 성조의 유형에 따라 차이를 보였는데 고성조로 시작하는 강세구의 기울기가 저성조로 시작하는 강세구의 기울기에 비해 큰 값을 보였다. 이는 고성조로 시작한 강세구 성조의 F0 상승과 성조 간 상승 폭의 차이에서 기인하는 것임을 알 수 있었다. 즉, 고성조로 시작하는 강세구의 4개 성조 모두 F0값의 상승이 있었으며 +H의 F0값 상승 폭이 L+의 F0값 상승 폭에 비해 큰 것으로 나타났다. 강세구 내의 두 상승곡선은 유사한 기울기값을 보였으나 강세구 시작 상승곡선의 F0 변화량이 강세구 끝 상승곡선의 F0 변화량보다 크게 나타났다. 결과적으로 강세구의 피치곡선의 모습은 F0 변화량이 큰 첫 상승곡선에 이어 음절수에 반비례하여 떨어지는 하강곡선이 뒤를 이으며 이 하강곡선의 L+의 F0는 첫 상승곡선의 L만큼 낮아지거나(화자 KSH) 또는 그보다 다소 높았으며(화자 KSM) 비교적 F0 변화량이 적은 상승곡선으로 이어졌다. 이 상승곡선의 Ha의 F0는 첫 상승곡선의 H+만큼은 올라가지 않았다. 마지막으로 하나의 상승곡선이 하나의 강세구를 이루는 L Ha 하강곡선의 기울기는 강세구 음절수에 반비례 관계를 보였다. 이와 같이 강세구에 나타나는 여러 종류의 하강곡선과 상승곡선은 기울기의 차이를 보이는 경우도 있으나 F0값의 차이 또한 보임으로써 이러한 정량적 분석이 곡선 간 구분에 어느 정도의 도움을 줄 수 있는 것으로 보인다.

본 연구는 화자 당 분석 대상 문장 수가 비교적 많고 문장의 길이와 통사적 복잡도가 다양하기 때문에 화자별 운율 특성을 규명하는 데에 많은 강점을 가지고 있다. 그러나 화자 수가 둘 밖에 되지 않기 때문에 한국어 운율 특성에 관한 일반화에는 제한이 있을 수밖에 없다.

또한 같은 낭독체 발화문장이라도 문장의 내용과 종류 및 길이와 통사적 복잡성에 따라 운율적 특성들이 달리 나타날 수 있으며 실험실 환경에서의 낭독체가 아닌 자연스런 대화 상황에서의 발화문장 또한 운율적 특성을 달리 할 수 있을 것이다. 추후 연구에서는 보다 많은 화자를 대상으로, 뉴스문만이 아닌 다양한 스타일의 문장들에 대한 분석과 더 나아가서 대화체 발화문장에 대한 분석의 시도가 있어야 할 것이다.

감사의 글

본 연구는 2002학년도 원광대학교 교내 일반과제 지원으로 수행되었습니다. 본 연구를 위해 분절음 및 운율 레이블링 자료를 사용하도록 허락해준 한국전자통신연구원에 감사드립니다.

참고 문헌

1. 이현복, "서울말의 리듬과 억양," *어학연구*, 10 (2), pp. 15-25, 1974.
2. 김차균, *우리말의 성조*, 태학사, 1993.
3. H. Koo, *An Experimental Acoustic Study of the Phonetics of Intonation in Standard Korean*, Doctoral dissertation, University of Texas at Austin, 1986.
4. S.H. Lee, "Intonational domains of the Seoul dialect of Korean," a paper presented at the 117th meeting of the Acoustical Society of America, Syracuse, N.Y.; An abstract in *Journal of the Acoustical Society of America*, 85, suppl. 1, S99, 1989.
5. de-Jong, "Initial tones and prominence in Seoul Korea," *Working Papers in Linguistics*, 43, 1-14, Ohio State University, 1994.
6. S.-A. Jun, *The Phonetics and Phonology of Korean Prosody: Intonational Phonology and Prosodic Structure*, Doctoral dissertation, Ohio State University, 1993.
7. 이호영, *The Structure of Korean Prosody*, Doctoral dissertation of University College London, 1990.
8. 성철재, *한국어 리듬의 실험음성학적 연구—시간 구조와 관련하여*, 서울대학교 대학원 박사논문, 1995.
9. M. Beckman & S.-A. Jun, *K-ToBI annotation convention*, *Unpublished manuscript*, Ohio State University and UCLA, 1995.
10. J. B. Pierrehumbert, *The Phonology and Phonetics of English Intonation*, Doctoral dissertation, MIT, 1980.
11. S.-A. Jun, *K-ToBI(Korean ToBI) labeling conventions, version 3.0, unpublished manuscript*, UCLA, 2000.
12. J.J. Kim S.H. Lee, H.J. Ko, Y.J. Lee, S.H. Kim, J.C. Lee, "An analysis of some prosodic aspects of Korean utterances using K-ToBI labelling system," *Proceedings of ICSP '97*, pp. 87-92, 1997.
13. S.-A. Jun, "The accentual phrase in the Korean prosodic hierarchy," *Phonology*, 15, pp. 189-226, 1998.

14. 이숙향, 한국어 운율구조 기술 체계에 대한 연구: K-ToBI 기술 체계중 중심으로, *언어학* 20 (2), pp. 1-18, 2002.
15. 이호영, "국어 억양의 상승과 하강 현상," *한글*, 257, 5-26, 2002.
16. H.J. Lee, *Tonal Realization and Implementation of the Accentual Phrase in Seoul Korean*, MA thesis, UCLA, 1999.
17. S.-A. Jun & M. Oh, "A prosodic analysis of three types of Wh-phrases in Korean", *Language and Speech*, 39 (1), pp. 37-61, 1996.
18. 김기호, 김화영, 김민정, *담화상에 나타나는 목적격표지 {-를}의 의미고찰*, *음성과학* 6, pp. 55-82, 1999.
19. 전 은, 이숙향, "한국어 조사의 운율적 특성—낭독체 문장을 중심으로," *말소리* 37, pp. 73-85, 1999.
20. S.-A. Jun, "The accentual pattern and prosody of the Chonnam dialect of Korean," *Havard Studies in Korean Linguistics*, 3, pp. 89-100, 1989.
21. 문수미, *한국어 액센트에 관한 실험음성학적 연구*, 서울대학교 대학원 박사논문, 1999.
22. 이숙향, "제주방언 억양패턴의 실험음성학적 연구," *한국음향학회 학술발표대회 논문집* 18 (2(s)), pp. 369-372, 1999.
23. M. Oh, "An optimality theoretic analysis of tone realization in Korean," *Speech Sciences*, 10 (3), pp. 89-101, 2003.
24. Jun, S.-A and M. Oh, "Acquisition of second language intonation," *The 6th Proceedings of International Conference on Spoken Language Processing*, pp. 73-76, 2000.
25. Y.J. Lee, S.H. Lee, J.J. Kim, H.J. Ko, Y.I. Kim, S.H. Kim, J.C. Lee, "A computational algorithm for F0 contour generation in Korean developed with prosodically labeled databases using K-ToBI system," *Proceedings of 1998 International conference on Spoken Language Processing*, pp. 1995-1998, 1998.

저자 이력

● 이 숙 향 (Sook-Hyang Lee)

한국음향학회지 제23권 제2호 참조

● 김 종 진 (Jong-Jin Kim)



1969년 8월 24일생
 1995.2 원광대학교 컴퓨터공학과 박사
 1997.2 원광대학교 컴퓨터공학과 석사
 2000.2 원광대학교 컴퓨터공학과 박사수료
 2000.7 한국전자통신연구원 ~ 현재
 선임연구원