

# 한국어 원거리 음성의 운율적 특성<sup>1)</sup>

이숙향\* 김선희\*\* 김종진\*\*\*

\* 원광대학교 영중어문학부, \*\* 서울대학교 언어정보 연구소, \*\*\* 한국전자통신연구원

## Prosodic Characteristics of Korean Distance Speech

Sook-hyang Lee\*, Sunhee Kim\*\*

\* Division of English and Chinese, Wonkwang Univ.

\*\* Humanities Information Research Center, Seoul National Univ.

\*\*\* Electronics and Telecommunication Research Institute

[shlee@wonkwang.ac.kr](mailto:shlee@wonkwang.ac.kr), [snhkim@snu.ac.kr](mailto:snhkim@snu.ac.kr), [kimji@etri.re.kr](mailto:kimji@etri.re.kr)

### Abstract

The aim of this paper is to investigate the prosodic characteristics of Korean distant speech. 36 2-syllable words of 4 speakers (2 males and 2 females) produced in both distant-talking and normal environments were used. The results showed that ratios of second syllable to first syllable in vowel duration and vowel energy were significantly larger in the distant-talking environment compared to the normal environment and f0 range also bigger in the distant-talking environment. In addition, 'HL%' contour boundary tone in the second syllable and/or 'L +H' contour tone in the first syllable were used in the distant-talking environment.

### I. 서론

본 연구는 지속시간, 에너지, F0와 억양 패턴을 중심으로 원거리 음성의 운율적 특성을 살펴보고자 한다. 본 연구에서 원거리 음성은 일정한 거리에 있는 음성 시스템을 조작하기 위하여 화자가 음성을 보다 명료하게 전달하고자 발화한 음성을 의미한다.

1) 본 논문에서 사용된 음성 데이터는 ETRI의 지원으로 KAIST에서 공동 제작되었습니다. 또한, 본 연구는 제 1저자의 원광대학교 2005학년도 교내 일반과제 지원으로 수행되었습니다.

### 2. 분석 대상 및 분석 방법

본 연구에서 사용하는 음성데이터는 PC 명령어를 녹음한 것으로서 이전 연구[1,2]에서 이미 사용했던 데이터 가운데, 남, 녀 각 2명이 50개 단어를 평이한 음성으로 발성한 것 중 2음절어 단어 36개의 음성파일을 그 분석 대상으로 한다. 피험자는 모두 서울 경기지역에서 성장하였으며 남녀 각 2명으로, 여성 화자는 30대, 남성 화자는 20대와 40대였다. 피치의 최소값, 최고값 및 그 대역을 추출하기 위하여서는 Praat 4.2.21을 이용하였다

### III. 분석 결과 및 논의

#### 1. 지속시간

표 1에서 볼 수 있듯이 모음 지속시간은 발화상황에 상관없이 첫째 음절에 비해 둘째 음절에서 길며 4명의 화자 모두 음절 간 유의한 차이를 보이고 있다. 표 2에서 볼 수 있는 바와 같이 음절 간 모음의 지속시간 비율((둘째음절/첫째음절)\*100)은 대체적으로 평이 음성에 비해 원거리 음성에서 큰 값을 보이고 있다. 발화상황 간 차이를 화자별로 보면 화자 F26와 M48만이 유의한 차이를 보이고 있으며 이는 뒤에서 논의될 발화상황 간

경계성조 억양 패턴 차이에 기인하는 것 같다. 즉, 이 두 화자는 다른 두 화자와는 달리 대부분의 단어에서 둘째 음절의 경계성조를 평이 음성에서는 simple tone인 'L%'에서 원거리 음성에서 contour tone인 'HL%'로 변화시킴으로써 원거리 음성의 둘째 음절이 평이음성의 둘째 음절에 비해 상당히 길어지는 결과를 초래한다고 할 수 있다.

화자	발화상황	평균 (표준편차)		샘플수	t-test 유의도
		첫째 음절 (ms.)	둘째 음절 (ms.)		
전체	원거리	120.36 (68.70)	222.15 (87.44)	139	.000
	평이	112.05 (61.10)	153.48 (69.52)	139	.000
M1	원거리	97.24 (43.19)	171.67 (68.42)	34	.000
	평이	73.38 (32.76)	122.44 (52.23)	34	.000
F26	원거리	102.71 (54.16)	244.09 (103.33)	35	.000
	평이	130.54 (55.95)	176.17 (71.05)	35	.007
M48	원거리	105.43 (57.61)	215.80 (69.41)	35	.000
	평이	114.11 (67.35)	139.69 (60.37)	35	.100
F79	원거리	175.40 (83.01)	255.60 (82.20)	35	.000
	평이	129.06 (65.52)	174.74 (78.24)	35	.018

표 1. 음절별 발화상황별 모음 지속시간 평균, 표준편차 및 음절 간 차에 대한 Paired Samples t-test 결과.

화자	평균 (표준편차)		샘플수	t-test 유의도
	원거리 (%)	평이 (%)		
전체	253.48 (189.94)	181.33 (128.49)	139	.000
M1	220.51 (140.94)	209.30 (143.71)	34	.478
F26	313.9562 (204.84)	162.02 (99.34)	35	.000
M48	284.92 (236.30)	170.73 (127.05)	35	.000
F79	193.60 (140.69)	184.06 (140.24)	35	.677

표 2. 발화상황별 둘째 음절 대 첫째 음절의 모음 지속시간 비율 평균 및 표준편차와 발화 상황 간 차이에 대한 Paired Samples t-test 결과.

## 2. 에너지

표 3에서 볼 수 있는 바와 같이 모음 에너지는 발화상황에 따라 다른 양상을 보이고 있다. 대체적으로 원거리 음성에서는 첫째 음절에 비해 둘째 음절에서 큰 값을 보이며 반대로 평이 음성에서는 둘째 음절에 비해 첫째 음절이 큰 값을 보이고 있다. 화자별 유의도를 보면 평이 음성 화자 M1과 원거리 음성 F79을 제외하고 모두 유의한 차이를 보이고 있다. 표 4에서 볼 수 있듯이 음절 간 모음의 에너지 비율 ((둘째음절/첫째음절)\*100)은 4명의 화자 모두 발화 상황 간에 유의한 차이를 보이고 있으며 지속시간에서 유의한 차이를 보이지 않은 화자 M1과 F79도 에너지에서는 유의한 차이를 보이고 있으며 이는 둘째 음절, 즉 마지막 음절의 에너지를 크게 발음하는 것이 원거리 음성의 전달의 효율성을 높이는 전략의 하나일 수 있지 않을까 한다.

화자	발화상황	평균 (표준편차)		샘플수	t-test 유의도
		첫째 음절 (dB)	둘째 음절 (dB.)		
전체	원거리	62.92 (6.65)	65.50 (5.59)	139	.000
	평이	54.32 (9.50)	52.64 (9.31)	139	.000
M1	원거리	58.24 (3.90)	60.94 (2.03)	34	.001
	평이	47.68 (2.98)	48.76 (2.19)	34	.118
F26	원거리	61.49 (7.20)	66.54 (4.37)	35	.002
	평이	66.94 (6.37)	64.63 (5.89)	35	.030
M48	원거리	69.71 (3.75)	72.46 (2.15)	35	.002
	평이	56.80 (2.70)	54.80 (4.04)	35	.019
F79	원거리	62.11 (5.13)	61.91 (3.74)	35	.838
	평이	45.66 (4.19)	42.26 (4.19)	35	.000

표 3. 음절별 발화상황별 모음 에너지 평균, 표준편차 및 음절 간 차에 대한 Paired Samples t-test 결과.

화자	평균 (표준편차)		샘플수	t-test 유의도
	원거리 (%)	평이 (%)		
전체	104.98 (12.38)	97.35 (10.43)	139	.000

M1	105.20 (9.47)	102.81 (9.99)	34	.010
F26	110.03 (17.56)	97.09 (10.98)	35	.000
M48	104.30 (7.78)	96.72 (9.05)	35	.000
F79	100.37 (10.86)	92.95 (9.59)	35	.000

표 4. 발화상황별 둘째 음절 대 첫째 음절의 모음 에너지 비율 평균 및 표준편차와 발화 상황 간 차이에 대한 Paired Samples t-test 결과.

### 3. F0

표 5에서 볼 수 있듯이 두 발화상황 모두에서 f0의 최고값, 최저값, 그리고 대역폭에 있어서 성별 간 유의한 차이를 보이고 있다.

분석대상	원거리 음성	t-test 유의도	평이 음성	t-test 유의도
	평균 (표준편차)		평균 (표준편차)	
대역폭	98.21/148.53 (43.87/57.65)	.000	46.61/86.40 (18.24/35.14)	.000
최고값	224.48/349.24 (32.32/31.11)	.000	131.49/238.44 (17.07/37.33)	.000
최저값	126.27/200.71 (33.115/46.32)	.000	84.89/152.04 (10.85/11.67)	.000

표 5. F0대역폭과 최고값, 최저값의 성별 간 차이. '/' 왼쪽은 남자화자, 오른쪽은 여자화자의 값을 나타냄.

f0 대역폭은 표 6에서와 같이 모든 화자가 두 발화상황 간에 유의한 차이를 보이고 있으며 원거리 음성에서 훨씬 큰 값을 보이고 있다.

화자	측정 대상	최저 (Hz)	최고 (Hz)	평균(Hz)	대역폭 (Hz)	t-test 유의도
M1	최고	180/102	222/177	200.17/125.91 (10.6/18.75)	71.66/ 39.88	.000
	최저	84/73	182/140	128.51/86.03 (30.50/14.08)		
F26	최고	288/186	390/294	333.64/227.03 (30.84/34.17)	129.64/ 81.59	.000
	최저	143/105	302/159	204/145.44 (46.65/10.00)		
M48	최고	183/115	29/168	248.11/136.22 (28.55/13.41)	124.03/ 52.24	.000
	최저	84/76	214/105	124.08/83.78 (35.77/6.35)		
F79	최고	300/201	400/353	364.83/249.86 (22.62/37.29)	167.66/ 91.22	.000

	최저	59/135	339/187	197.17/158.64 (46.37/37.29)		
--	----	--------	---------	--------------------------------	--	--

표 6. 화자별 F0대역 및 발화상황 간 대역폭 차에 대한 t-test결과. ('/' 왼쪽은 원거리 음성의 값이며 오른쪽은 평이 음성의 값임).

### 4. 억양 패턴

평이 음성에서의 2음절 단어의 억양 패턴은 화자에 따라 정도는 다르지만 'H L%'가 주를 이루고 있다(표 6 참조). 이에 반해 원거리 음성에서는 화자 M48과 F26

단어	F26	M48	F79	M1
검색	L HL%	H L%	H L%	H L%
날짜	H L%	L+H L%	H L%	L HL%
다음	H L%	L HL%	H L%	H L%
달기	H L%	L HL%	H L%	H L%
뒤로	H L%	H L%	H L%	H L%
메뉴	L+H L%	H L%	H L%	L HL%
메일	H L%	H L%	H L%	L HL%
반복	H L%	H L%	H L%	H L%
보기	H L%	H L%	H L%	L HL%
분류	H L%	H L%	H L%	H L%
설정	H L%	H L%	H L%	H L%
스톱	H L%	H L%	H L%	H L%
시작	H L%	H L%	H L%	H L%
암호	H L%	L HL%	H L%	L HL%
연결	H L%	H L%	H L%	L+H L%
열기	H L%	H L%	H L%	L+H L%
오전	H L%	H L%	H L%	L+H L%
오프	H L%	H L%	H L%	L HL%
오후	H L%	H L%	H L%	L+H L%
이전	H L%	L+H L%	H L%	H L%
일정	H L%	H L%	H L%	L+H L%
잠금	H L%	L+H L%	H L%	L+H L%
재생	H L%	L+H L%	H L%	H L%
저장	H L%	L+H L%	H L%	H L%
전송	H L%	?	H L%	L+H L%
정지	L+H L%	L+H L%	H L%	H L%
종료	H L%	?	H L%	H L%
주소	H L%	L+H L%	H L%	H L%
중지	H L%	H L%	H L%	?
처음	H L%	H L%	H L%	H L%
추가	H L%	H L%	H L%	H L%
축소	H L%	H L%	H L%	H L%
통화	H L%	H L%	H L%	H L%
해제	H L%	H L%	H L%	H L%
확대	H L%	H L%	H L%	H L%
확인	H L%	H L%	H L%	H L%

표 7. 평이 음성의 화자별 억양 패턴. 음영은 'H L%'의 높은 사용빈도를 보기 쉽게 하기 위하여 사용하였음. '?'은 분석이 어려운 경우임.

은 'L HL%'가 주를 이루고 있다(표 7 참조). 이 두 화자의 큰 차이점은 첫 음절이 평이 음성에서는 첫음절이 어떤 분절음으로 시작하든 대부분 'H'로 시작하는 데 반해 원거리 음성에서는 화자 F26는 첫 음절의 분절음 유형에 상관없이 'L'로 시작하는 반면 화자 M48은 강

자음으로 시작하는 경우에는 'H'로 시작하고 그 외에는 'L'로 시작하고 있다[3]. 이러한 차이에도 불구하고 두 화자는 공통적으로 원거리 음성에서는 'HL%'라는 contour tone을 사용하여 원거리 음성의 전달 효과를 높이려는 것으로 나타났다. 이에 반해 화자 F79과 M1은 첫 음절 성조 'L'을 'L+H'로 바꿈으로써 첫 음절을 길게 contour tone으로 발음하여 원거리 음성의 전달 효과를 높이려는 것 같다. 화자 76의 특기할 만한 사실은 위 두 가지 전략('HL%' 경계성조 사용과 첫 음절의 'L+H' contour tone 사용)을 다 사용하고 있다는 점이다.

또 한 가지 특기할 만한 것은 본 연구에서와 같이 하나의 강세구 그 자체가 억양구를 이룰 때 강세구 내의 음절수가 적은 경우에 억양구의 경계성조 유형에 따라 강세구 내의 성조가 영향을 받을 수 있음을 보여주고 있다는 것이다. 그 예로 평이 음성에서 경계성조가 'L%'인 경우, 강세구가 강자음으로 시작하지 않더라도 첫 성조가 L'로 실현되지 못하고 'H'로 실현되고 있다. 이와는 반대로 원거리 음성에서 경계성조가 'HL%'인 경우, 강세구가 강자음으로 시작하더라도 음절수가 부족하면 첫 음절 성조 'H'가 'L'로 바뀌는 경우도 관찰되었다(화자 M48).

	F26	M48	F79	M1
검색	L HL%	L+H L%	L HL%	L+H L%
날짜	L HL%	L HL%	L+H L%	L HL%
다음	L HL%	L HL%	H HL%	L+H L%
달기	L HL%	L HL%	L HL%	L HL%
뒤로	L+H L%	L+H L%	HL HL%	L+H L%
메뉴	L+H L%	L HL%	L HL%	L HL%
메일	L HL%	L HL%	L HL%	L HL%
반복	L HL%	L HL%	L HL%	L+H L%
보기	L HL%	L HL%	HL HL%	L HL%
블록	L HL%	L HL%	L HL%	L HL%
설정	H HL%	L HL%	HL HL%	H L%
스톱	H HL%	H HL%	H L%	H L%
시작	?	L HL%	H HL%	H L%
암호	L HL%	L HL%	L+H HL%	L HL%
연결	L HL%	L HL%	L+H HL%	L+H L%
열기	L HL%	L HL%	L+H HL%	L+H L%
오전	L HL%	L HL%	L+H HL%	L+H L%
오프	L HL%	L HL%	L+H HL%	L HL%
오후	L HL%	L HL%	L HL%	L HL%
이전	L HL%	L HL%	L+H L%	L HL%
일정	L HL%	L HL%	L+H HL%	L+H L%
잠금	L HL%	L HL%	L+H L+HL%	L+H L%
재생	L HL%	L HL%	L HL%	L+H L%
저장	L HL%	L HL%	H L+HL%	L+H L%
전송	L HL%	L HL%	L HL%	L+H L%
정지	L HL%	L HL%	L+H L+HL%	L+H L%

종료	L HL%	L HL%	H HL%	L+H L%
주소	L+H L%	L+H HL%	L HL%	L+H L%
중지	L HL%	L HL%	L+H L+HL%	L+H L%
처음	H L%	H HL%	H L+HL%	H L%
추가	H L%	L HL%	H L%	H L%
축소	H L%	L HL%	H HL%	H L%
통화	H L%	L HL%	H L%	H L%
해제	H L%	L HL%	H HL%	L+H L%
확대	H L%	L+H HL%	H HL%	H L%
확인	H L%	L+H HL%	H L+HL%	L HL%

표 8. 원거리 음성의 화자별 억양 패턴. 음영은 'L HL%'의 높은 사용빈도(특히 화자 M48과 F26)를 보기 쉽게 하기 위하여 사용하였음.

#### IV. 결론

본 연구는 2음절어 고립어를 이용하여 원거리 음성의 운율적 특성을 살펴보았다. 지속시간은 모음 지속시간은 발화상황에 상관없이 첫째 음절에 비해 둘째 음절에서 길며 음절 간 모음의 지속시간 비율 ((둘째음절/첫째음절)\*100)은 평이 음성에 비해 원거리 음성에서 큰 값을 보였다. 에너지는 원거리 음성에서 첫째 음절에 비해 둘째 음절에서 큰 값을 보이며 평이 음성에서는 그 반대 양상을 보였다. 음절 간 모음의 에너지 비율 ((둘째음절/첫째음절)\*100)은 발화 상황 간에 유의한 차이를 보였다. f0 대역폭 또한 두 발화상황 간에 유의한 차이를 보였으며 억양 패턴은 원거리 음성에서 'HL%' contour 경계성조를 사용하거나 첫 음절에 'L+H' contour tone을 사용하였다.

요약하자면, 지속시간과 에너지의 음절 간 차이와 f0 대역폭을 원거리 음성에서 더 크게 함으로써, 그리고 simple tone 대신 contour tone 사용으로 억양 패턴을 달리 함으로써 원거리 음성의 전달 효과를 높이려는 것으로 나타났다.

#### 참고문헌

- [1] 김선희, "한국어 롬바드 음성의 지속시간," 말소리 54호, 1-14, 2005.
- [2] 이숙향, 김선희, "한국어 원거리 음성의 모음의 음향적 특성," 말소리 55호, 61-76, 2005
- [3] Sun-Ah Jun, "K-ToBI (Korean ToBI) labelling conventions: Version 3", *UCLA Working Papers in Phonetics* 99, pp.149-173, 2000.