

한국어 롬바드 음성의 지속시간 연구

김 선 희

한국과학기술원 전자전산학과

A Study on the Durational Characteristics of Korean Lombard Speech

Sunhee Kim

Department of Electrical Engineering and Computer Science

shkim@ee.kaist.ac.kr

Abstract

This paper presents durational characteristics of Korean Lombard speech using speech data, which consist of 500 Lombard utterances and 500 normal utterances of 10 speakers (5 males and 5 females). Each file was segmented and labeled manually and the duration of each segment and each word was extracted. The durational change of Lombard effect in comparison with normal speech was analyzed using a statistical method. The results show that the duration of words with Lombard effect is increased in comparison with normal style, and that the average unvoiced consonantal duration is reduced while the average vocalic duration is increased. Female speakers show a stronger tendency towards lengthening the duration in Lombard speech, but without statistical significance. Finally, this study also shows that the speakers of Lombard speech could be classified according to their different duration rate.

I. 서론

롬바드 효과란 잡음 환경에서 화자가 음성을 보다 명료하게 전달하고자 하여 목소리를 높여 말할 때 일어나는 조음상의 변화를 의미하는 것으로[1][2], 이는 음성인식 시스템의 한 성능 저하 요인으로 지적되어 왔다. 롬바드 음성이 효과적으로 음성인식 시스템에서 처리되기 위한 연구로는 크게 그 특징의 규명에 관한 연구와 보상 방법에 관한 연구로 구분될 수 있다. 먼저 롬바드 효과의 특징에 관한 대표적인 연구로는 [1][2][3][4]를 들 수 있는데, 이러한 연구들에 따른 롬바드 음성의 공통적인 음향학적 특징으로는 (i) 모음 지속시간의 증가, (ii) 에너지 분포의 이동, (iii) F1의 증가 등을 들 수 있다[2]. 음성 인식에서 롬바드 음성의 보상에 관한 연구로는 [2][3][5][6]을 들 수 있는데, 보상 방법으로는 (i) 학습 방법, (ii) 전처리 방법, (iii) 후처리 방법의 세 가지로 요약된다[2]. 한국어 데이터를 이용한 연구로는 [7][8]이 있는데, 이들은 모두 보상 방법에 관한 연구로서 한국어 롬바드 음성의 특징에 관한 음성학적 연구는 아직까지는 시도된 적이 없었다.

본 논문은 롬바드 음성의 특성을 규명하는 것을 목적으로 하는 연구의 한 부분으로서, 여러 가지 특성 가운데 운율적 특성인 지속시간을 중심으로 분석하고

는 모두 평이한 음성과 롬바드 음성의 길이 차이가 유의미함을 알 수 있었다.

다음 <표 3>은 단어 지속시간 증감과 그 비율이 남녀간 성별 차이를 보이는지를 비교한 것이다. 전체적으로는 <표 2>와 같이 1음절어에서 차이가 가장 크고 2, 3 음절어로 갈수록 차이가 줄어드는 경향은 보인데, 남성 화자들의 지속시간 증가 비율에 비하여 여성 화자들의 지속시간 증가비율이 두드러진 것을 볼 수 있었다. 지속시간의 증가 정도에 있어서 [6]의 지적과 같이 여성 화자들에게서 좀 더 그 차이가 크게 나타나는 것으로 보인다. 이 결과를 T검증을 하였는데, 예상과는 달리 남녀 간의 차이는 없다는 결과를 얻었다.

	Male		Female	
	Diff (SD)	Pct	Diff (SD)	Pct
1-syll word	51.6 (35.84)	19.9	96.9 (56.97)	29.53
2-syll word	44.33 (56.62)	8.7	112.88 (136.77)	18.16
3-syll word	35.73 (49.08)	5.91	123.53 (122.37)	17.73
4-syll word	-4.15 (126.23)	-0.44	-2.0 (79.47)	-0.18
Ave	31.87 (64.00)	5.52	82.82 (88.61)	12.07

<표 3> 성별에 따른 단어의 지속시간 차이 및 그 비율 (Diff: Average difference of duration, SD: Standard Deviation, Pct: Percentage of difference)

다음은 롬바드 음성에 있어서 각 분절음의 지속시간을 그룹별로 나누어 살펴본 것이다. 아래 <표 4>와 같이 분절음은 크게 자음과 모음으로 분류되는데, 자음의 경우는 유음을 제외하고는 평이한 음성의 지속시간보다 롬바드 음성의 지속시간이 감소한 것을 관찰할 수 있었다.

	Male		Female	
	D (SD)	PV	D (SD)	PV
Stops	-17.45 (8.32)	-19.95	-11.63 (19.4)	-15.74
Fricatives	-24.74 (21.54)	-22.99	-7.57 (8.38)	-7.68
Affricates	-18.64 (17.17)	-16.41	-11.61 (17.25)	-9.76
Nasals	-3.23 (25.1)	-2.96	-16.98 (21.44)	-11.24
Liquids	-1.86 (8.75)	-3.3	3.03 (6.49)	4.48
Vowels	31.59	27.37	56.58	39.9

	(11.68)		(39.55)	
Diphthongs	32.34 (31.23)	19.68	58.07 (45.26)	28.81

<표 4> 분절음의 지속시간 차이 (Diff: Average difference of duration, SD: Standard Deviation, Pct: Percentage of difference)

그룹별 분절음에서 평이한 음성과 롬바드 음성에 대한 T검증 결과 파열음은 $t=5.87$, $p<.001$, 마찰음은 $t=2.74$, $p<.05$, 파찰음은 $t=3.33$, $p<.01$, 모음은 $t=-4.57$, $p<.01$, 이중모음은 $t=-3.65$, $p<.01$ 로 무성 자음과 모음의 경우에는 유의한 차이가 있었으나, 비음과 유음은 차이가 없는 것으로 나타났다. 즉, 한국어에 있어서 롬바드 음성의 지속시간 차이는 무성 자음인 경우에는 감소하고 모음인 경우에는 증가하며 유성 자음은 그 차이가 없다고 볼 수 있다.

이러한 자음 감소 현상은 영어[2]와 스페인어[5]에서도 관찰된 현상으로, 한국어의 경우에는 롬바드 음성에서 유성 자음인 비음과 유음을 제외한 자음의 지속시간은 감소하고 모음의 지속시간은 증가하나, 모음의 지속시간 증가율이 자음의 지속시간 감소율에 비하여 현저하게 큰 값을 보여, 전체 단어의 지속시간 증가는 모음 지속시간의 증가에 기인한다고 할 수 있다.

2. 지속시간 변이에 의한 화자 유형 분석

다음의 <표 5>는 10명 화자의 단어 지속 증가율(%)을 음절수에 따라 분류하여 나타낸 것이다. 1음절어의 경우에는 모든 화자가 롬바드 음성을 더 길게 발화하는 것을 알 수 있다. 2음절 이상으로 음절이 길어질수록 3명의 화자(M4, M5, F1)를 제외한 다른 화자들은 변화하는 비율은 서로 다르나, 모두 롬바드 음성의 지속시간이 증가하였다.

	1 Syll-W	2 Syll-W	3 Syll-W
M1	12.02	20.4	14.37
M2	51.49	23.94	18.53
M3	22.98	9.82	7.74
M4	43.49	17.21	13.34
M5	10.92	-0.23	-3.39
F1	16.73	-13.29	-3.59
F2	9.35	-2.9	-0.83
F3	47.73	44.11	40.27
F4	25.76	25.77	24
F5	12.32	18.44	12.16

<표 5> 화자별 단어의 지속시간 차이 비율 (M1~M5: Male speakers, F1~F5: Female speakers)

다음 <표 6>에서는 분절음 지속시간 증가율을 기준으로 화자들의 발성 유형을 살펴보았다. 지속시간 증가율이 크게 모음과 자음으로 구분되어 자음인 경우에 감소하고 모음인 경우에 증가하는 경향을 보이는 화자들로서는 M3, M4, M5, F1, F5 등 5인이었다. 지속시간 증가율이 유성음과 무성음으로 구분되어 차이를 보여 유성음인 경우에 증가하고 무성음인 경우에 감소하는 경향을 보이는 화자로는 M1, M2, F2의 3인이 있었다. F3, F4의 경우는 유성음 가운데 비음은 감소하고 유음만이 증가하는 경향을 보였다. 이와 같이, 분절음 그룹에 따른 증가율에도 일정한 유형이 관찰된다고 할 수 있다.

	S	F	A	N	L	V	D
M1							
M2							
M3							
M4							
M5							
F1							
F2							
F3							
F4							
F5							

<표 6> 화자별 분절음 지속시간 차이 비율 (M1~M5: Male speakers, F1~F5: Female speakers, S: Stops, F: Fricatives, A: Affricates, N: Nasals, L: Liquids, V: Vowels, D: Diphthongs)

IV. 결론

본 논문에서는 롬바드 음성 데이터를 분석하여 그 지속시간의 변화를 살펴보았다. 본 연구의 결과는 대체적으로 전체 단어의 지속 시간은 증가하고, 자음의 경우는 감소하는 반면에 모음의 경우는 증가하는 것으로 알려진 기존의 연구와 같은 경향을 보인다고 할 수 있다. 다만, 한국어의 경우에 세부적으로는 무성 자음은 감소하고 모음은 증가하며, 유성 자음은 비음과 유음의 경우는 그 차이가 유의미하지 않은 것으로 나타났다.

본 논문은 기존 연구들이 대부분 롬바드 음성에 관련하여 여러 음향적/음성적 특성을 일괄적으로 분석한 것과 달리, 지속시간이라는 하나의 특성을 좀 더 정밀하게 분석하여, 한국어 롬바드 음성의 특징을 구체적으로 규명하였다고 볼 수 있다. 또한 롬바드 음성이 화자 의존적인 특성이 많다고는 일반적으로 알려져 있

지만 그 구체적인 사례는 발표된 바가 없었는데, 본 논문에서 롬바드 음성의 지속시간이 일반적으로 증가하지만, 화자와 단어를 구성하는 음절 수에 따라서는 다른 양상이 관찰되는 것을 보임으로써, 롬바드 음성의 화자 의존성의 구체적인 사례를 제시했다는 의미를 부여할 수 있을 것이다.

참고문헌

- [1] Junqua, J.-C., "The Influence of Acoustics on Speech Production: A Noise-Induced Stress Phenomenon as the Lombard Reflex," *Speech Communication*, Vol. 20, pp.13-22, 1996.
- [2] Hansen, J., "Analysis and Compensation of Speech under Stress and Noise for Environmental Robustness in Speech Recognition," *Speech Communication*, Vol. 20, pp.151-173, 1996.
- [3] Bou-Ghazale, S. E., J. Hansen, "A Comparative Study of Traditional and Newly Proposed Features for Recognition of Speech Under Stress," *IEEE Transactions on Speech and Audio Processing*, Vol. 8-4, pp.429-442, 2000.
- [4] Castellanos, A., J.-M. Benedi, F. Casacuberta, "An Analysis of General Acoustic-Phonetic Features for Spanish Speech Produced with the Lombard Effect", *Speech Communication*, Vol. 20, pp.23-35, 1996.
- [5] Stanton, B., L. Jamieson, G. Allen, "Robust recognition of loud and Lombard speech in the fighter cockpit environment," *Proc. ICASSP*, Vol. 2, pp.675-678, 1989.
- [6] Junqua, J.-C., "The Lombard effect and its role on human listeners and automatic speech recognizers," *J. Acoust. Soc. Amer*, Vol. 93, pp.510-524, 1996.
- [7] Chi, S., Y.-H. Oh, "Lombard Effect compensation and noise suppression for Noisy Lombard Speech Recognition", *Proc. ICASSP*, pp.2013-2016, 1996
- [8] 우수영, 롬바드 효과 보상 필터를 이용한 강인한 특징 추출 방법, 한국과학기술원 석사논문, 2003
- [9] 이숙향, 신지영, 김봉완, 이용주, "음성 코퍼스 구축을 위한 SiTEC 분절음/음운 레이블링 기준의 검토 및 제안", *말소리*, 46호, pp 127-143, 2003.