한국어 원거리 음성의 지속시간 연구

김선희(KAIST)

<차 례>

<Abstract>

A Study on the Durational Characteristics of Korean Distant-Talking Speech

Sunhee Kim

This paper presents durational characteristics of Korean distant-talking speech using speech data, which consist of 500 distant-talking 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. Using a statistical method, the durational change of distant-talking speech in comparison with normal speech was analyzed. The results show that the duration of words with distant-talking speech 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 distant-talking speech. Finally, this study also shows that the speakers of distant-talking speech could be classified according to their different duration rate.

* Keywords: Distant-talking speech, Normal speech, Duration, Word, Segment

1. 서 론

음성 인식 분야에서 원거리 음성 인식은 일반적으로 근거리 마이크를 사용할 수 없는 경우에 보통 화자로부터 일정한 거리에 있는 여러 개의 마이크를 통하여 음성이 입력될 때를 의미한다[1][2]. 대표적인 예로는 움직이는 로봇을 음성으로 조작하는 경우를 들 수 있는데, 이러한 경우에는 음성뿐만 아니라 다른 요인에 의해 유발되는 잡음과 음향적 환경에 따른 반향(echo) 등의 문제가 발생한다. 이와 같이 일정한 거리에 있는 음성 인식 시스템을 조작하는 경우에 화자는 음성을 보다 명료하게 전달하고자 목소리를 높여 발성을 하는 경우가 있는데, 이러한 화자의 조음상의 변화는 가산 잡음보다 음성인식 성능을 더 크게 저해하는 요인으로 지적되기도 하였다[3].

원거리 음성과 유사한 경우로 롬바드 음성을 들 수 있는데, 롬바드 음성이란 잡음 환경에서 화자가 음성을 보다 명료하게 전달하고자 목소리를 높여 발성을 하는 경우를 의미한다[4][5][6][7]. 잡음 환경의 성격이 다르기는 하지만 두 경우 모두 화자가 음성을 보다 명료하게 전달하고자 목소리를 높여 말할 때 발생하는 조음상의 변화가 야기되는 유사한 경우이다. 본 논문에서는 원거리 음성을 일정한 거리에 있는 음성 인식 시스템을 조작하기 위하여 화자가 음성을 보다 명료하게 전달하고자 목소리를 높여 발성을 하는 경우에 발생하는 조음상의 변화를 동반하는 음성이라 할 때, 평이한 음성과 비교하여 어떠한 특징이 관찰되는 지를 롬바드 음성 연구와 비교하여 고찰하고자 한다.

원거리 음성에 관한 연구는 주로 잡음 문제 해결에 집중되어 그 조음적 효과에 대한 음성적인 연구는 시도되지 않은 반면에[1][2], 롬바드 음성에 관한 연구로는 음성의 특징 규명에 관한 연구와 보상 방법에 관한 연구로 구분될 수 있다. 롬바드 효과의 특징에 관한 대표적인 연구로는 [4][5][6][8][9]를 들 수 있는데, 이러한 연구들에 의하면 평이한 음성과 롬바드 음성의 차이로는 (i) 모음 지속시간의 증가, (ii) 에너지 분포의 이동, (iii) F1의 증가 등을 들 수 있다[5]. 음성 인식에서 롬바드 음성의 보상에 관한 연구로는 [5][6][9][10]을 들 수 있는데, 보상 방법으로는 (i) 학습 방법, (ii) 전처리 방법, (iii) 후처리 방법의 세 가지로 요약된다[5]. 한국어 데이터를 이용한 연구로는 [11][12]가 있는데, 이들은 모두 잡은 보상 방법에 관한 연구로서, 한국어의 경우에 있어서 롬바드 음성이나 원거리 음성의 특징에 관한 음성학적 연구는 아직까지는 시도된 적이 없었다.

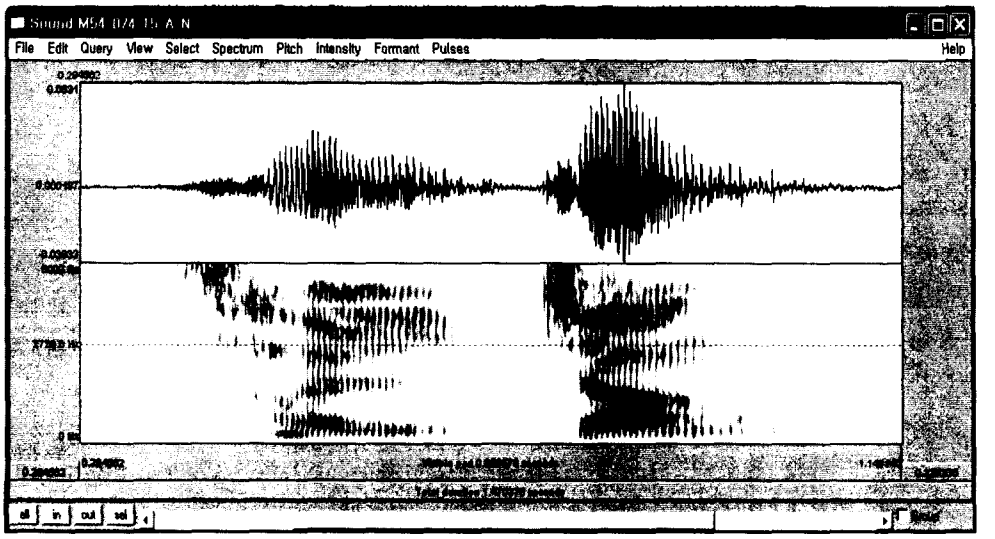
본 논문은 한국어 원거리 음성의 특성을 규명하는 것을 목적으로 하는 연구의 한 부분으로서, 여러 가지 특성 가운데 운율적 특성인 지속시간을 중심으로 분석하고자 한다. 이를 위하여 10명의 화자가 발성한 원거리 음성과 평이한 음성으로 발성한 데이터를 분절음 및 단어 단위로 지속 시간을 측정하여, 그 전체적인 변화 양상과 성별에 따른 변화 양상, 그리고 이러한 지속 시간의 화자별 변화 유형을

고찰하고자 한다. 이러한 연구는 먼저, 한국어 원거리 음성의 특성을 규명하고, 나아가 음성인식 시스템을 비롯한 음성 처리 분야에서 그 시스템의 성능 향상에 기여할 수 있을 것으로 본다.

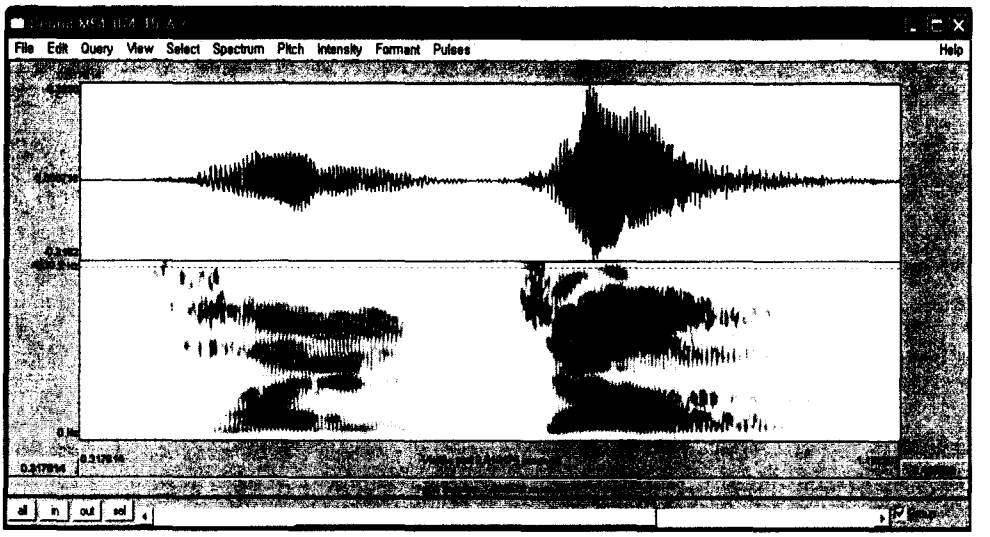
본 논문은 먼저 다음 2장에서 원거리 음성의 일반적인 특징을 살펴보고, 3장에서는 본 연구의 분석 대상 및 분석 방법을 소개한 다음, 4장에서 연구 결과와 해석을 제시하고, 5장의 결론으로 마무리할 것이다.

2. 원거리 음성의 특징

위에서 정의한 대로, 원거리 음성이란 일정한 거리에 있는 음성 인식 시스템을 조작하기 위하여 화자가 음성을 보다 명료하게 전달하고자 목소리를 높여 발성을 하는 경우에 발생하는 조음상의 변화를 동반하는 음성으로, 일반 음성과는 다른 조음에 기인한 여러 가지 음향적 효과가 발생하게 된다. 다음의 <그림 1>과 <그림 2>는 각각 단어 '설정'을 평이한 음성으로 발성한 경우와 원거리 음성으로 발성한 경우의 파형과 스펙트로그램이다. 이 그림들을 비교해 보면, 먼저 파형 상으로는 발성 에너지가 현저히 차이가 나는 것을 관찰할 수 있고, 스펙트로그램의 경우에는 롬바드 음성의 경우에 에너지 증가와 함께 평이한 음성과는 다른 곡선적인 형상이 관찰되는 것을 볼 수 있다.



<그림 1> 단어 '설정'의 일반 음성의 파형 및 스펙트로그램



<그림 2> 단어 '설정'의 원거리 음성의 파형 및 스펙트로그램

롬바드 음성의 음향학적 특징에 관한 연구로는 위에서 이미 언급한 바와 같이 [4][5][6][8][9]와 각 논문에 언급된 여러 선행 연구들이 있어 왔는데, [8]에 의한 롬바드 음성에 관여하는 음향적 특징으로는 (i) 음소 지속 시간, (ii) 전체 에너지, (iii) 저대역 스펙트럼 기울기, (iv) 고대역 스펙트럼 기울기, (v) 피치, (vi) F1, F2, (vii) 마찰음 포먼트, (viii) 마찰 정도, (ix) 대역별 에너지 분포 등이다. 이러한 특징이 롬바드 효과인 화자가 음성을 보다 명료하게 전달하고자 목소리를 높여 발성을 하는 경우에 발생하는 조음상의 변화와 관련이 있다면, 동일한 특징이 원거리 음성에도 관여한다고 추정할 수 있다.

롬바드 음성의 특징에 관한 개별 언어 연구로는 영어, 불어, 스페인어, 일본어 등에 관한 연구가 있었는데, 화자 종속적인 요소가 많은 롬바드 음성을 위한 데이터가 각각 독자적인 방식으로 수집되고 화자 수도 제한적인 경우가 많아서, 각 언어에 따른 결과들을 비교하는 것이 용이하지 않은 것으로 알려져 있다. [10]에 따르면, 영어와 불어의 경우에는 거의 비슷한 특성을 보이는데, 스페인어의 경우에는 롬바드 음성의 경우에 F2 변화의 차이가 두드러지고 성별 차이가 나타나며, 일본어의 경우는 캡스트럴 계수가 감소하는 경향을 보인다고 한다.

롬바드 음성의 특징 가운데 지속시간에 관한 연구 결과들은 모두 평이한 음성에 비하여 롬바드 음성의 경우에 모음의 지속시간이 증가하는 것으로 알려져 있고, 자음의 경우는 감소하는 경향을 보인다고 한다[4][5][6][10]. 본 논문에서는 한국어 원거리 음성에 있어서 지속시간의 변화 양상과 화자들의 지속시간 변이 유형을 파악하고, 이를 이러한 롬바드 음성에 관한 선행 연구의 결과와 비교해 보고자 한다.

3. 분석 대상 및 분석 방법

3.1. 음성 데이터

본 연구에서 사용한 음성 데이터는 [12]에서 제작하여 사용한 음성 데이터 가운데 그 일부를 대상으로 하였다. 전체 음성 데이터는 고립 단어 인식 시스템 구축을 위한 것으로, 평이한 음성과 원거리 음성으로 구분하여 50개의 PC 명령어를 발성한 것이다. 발성 단어의 음소 구성은 /ㅃ, ㅋ/을 제외한 17개의 자음, /ㅂ, ㅍ, ㄷ, ㄸ, ㅌ, ㅅ, ㅆ, ㅈ, ㅉ, ㅊ, ㄱ, ㄲ, ㅎ, ㅁ, ㄴ, ㅇ, ㄹ/과, 모음은 7개의 단모음 /ㅣ, ㅔ(ㅐ), ㅡ, ㅓ, ㅏ, ㅜ, ㅗ/, 그리고 7개의 이중 모음 /ㅠ, ㅕ, ㅛ, ㅖ, ㅚ, ㅟ, ㅘ/으로 이루어져 있다. 여기에서 한국어의 가능한 자음과 이중 모음의 일부가 제외된 것은, 실제 음성인식 시스템에서 일정한 영역에서의 사용을 위하여 제작되었기 때문이다. 사용된 분절음 수는 조음 방법에 따라 분류하면, 폐쇄음 41개, 마찰음 17개, 파찰음 28개, 비음 41개, 유음 19개, 단모음 95개, 그리고 15개의 이중 모음으로 구성되어, 총 256개의 분절음이 발성되었다. 발성 단어를 음절수에 따라 구분하면 1음절어 4개, 2음절어 36개, 3음절어 6개, 4음절어 4개로 구성되었다. 음성 데이터에 사용된 발성 목록은 아래와 같다.

<표 1> 발성 목록

1음절어	끝, 예, 온, 집
2음절어	검색, 날짜, 다음, 닫기, 뒤로, 메뉴, 메일, 반복, 보기, 볼륨, 설정, 스톱, 시작, 암호, 연결, 열기, 오전, 오프, 오후, 이전, 일정, 잠금, 재생, 저장, 전송, 정지, 종료, 주소, 중지, 처음, 추가, 축소, 통화, 해제, 확대, 확인
3음절어	끝내기, 되감기, 아니오, 앞으로, 재발신, 플레이
4음절어	볼륨낮춤, 볼륨높임, 일시정지, 일시중지

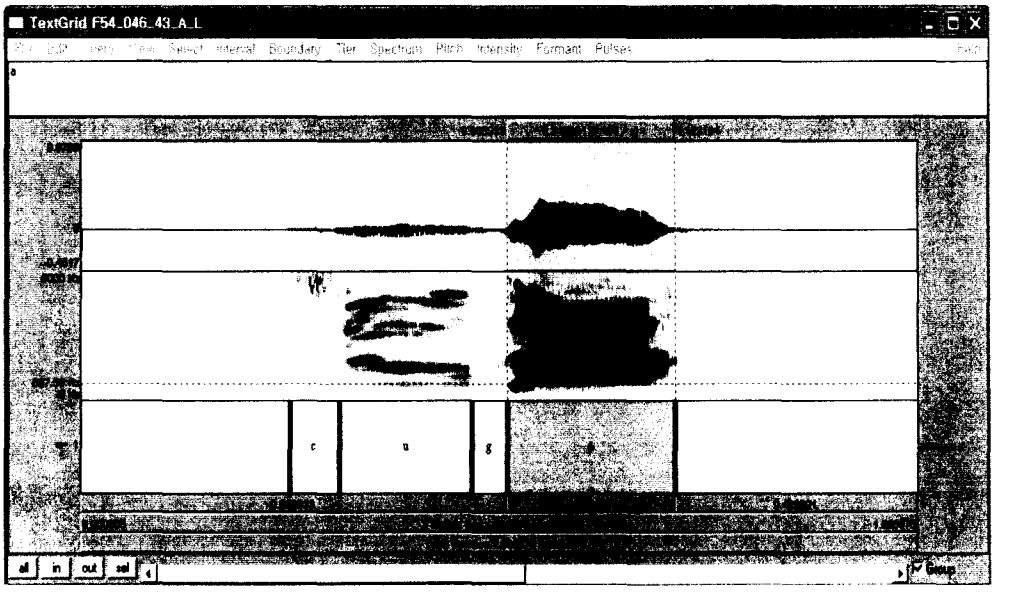
음성 수집에 참여한 화자로는 남성 화자 56명, 여성 화자 34명으로 총 90명으로 지역별 연령별로 구분하여 수집하였는데, 본 연구는 이 가운데 경기도 지역에 거주하는 남성 화자, 여성 화자 각 5명의 음성을 그 분석 대상으로 하였다. 음성 데이터는 각각 한 명의 화자 당 평이한 음성을 2회, 원거리 음성으로 3회를 발성하게 한 다음, 검증 과정을 거쳐 발성 단어 당 하나의 평이한 음성 파일과 원거리 음성 파일을 저장한 것이다. 따라서 본 논문에서 분석 대상이 되는 음성 파일은 화자 10명 당 평이한 음성 50개와 원거리 음성 50개로 전체 총 1,000개의 음성 파일이 사용되었다.

평이한 음성의 수집 방법은 화자로부터 0.5 m 지점에 설치한 마이크로 일상적인 어조로 발성된 음성을 수집한 것이다. 원거리 음성은 음성인식 시스템의 사용

자가 3 m 거리에 있는 시스템을 조작하기 위하여 목소리를 높여서 발성하도록 한 다음, 평이한 음성과 마찬가지로 화자로부터 0.5 m 의 지점에 설치된 마이크로 수집한 것이다.

3.2. 분석 방법

지속 시간은 모든 분절음의 지속 시간과 단어의 지속 시간을 측정하였는데, 이를 위하여 먼저 Praat 4.2.21을 이용하여 각 음성 파일을 수작업으로 하여 분절음 단위로 레이블링(labeling)하였다. 이 때 레이블링의 기준으로는 기본적으로 [13]을 따른 것으로, 분절음 레이블링에 있어서 음소 단위 이상의 세부적인 레이블링, 즉, 기식구간(H), 유성음화, 공명음화, 파찰음의 마찰구간, 탄설음 등의 표기는 하지 않았다. 모음의 경우는 해당 모음의 F1과 F2가 관찰되기 시작하는 곳을 시작점으로, 모음의 끝은 파형에서 모음의 파형(여러 성분들이 복합되어 나타나는)이 관찰되는 곳의 끝을 잡았고, 유음 /ㄹ/의 경우는 l/r을 구분하였다. 다음 <그림 3>은 이와 같은 기준에 의하여 레이블링한 예이다. 레이블링된 분절음에 대한 통계 분석은 SPSS 12.0을 이용하여 수행하였다.



<그림 3> Praat 4.2.21을 이용한 단어 '추가'의 분절음 레이블링

4. 결과 및 해석

4.1. 지속시간 분석

지속시간은 실험 음성학에서 음성적 환경이나 운율, 혹은 발화 양식(speaking style)이나 발화 속도 등과 관련하여 많이 다루어진 주제이고[14][15][16], 음성 처리 분야에서 지속시간의 모델링 문제는 중요한 주제로 다루어져 왔다. 음성 처리 분야에서는 세부적으로 음성 합성 분야[17][18]과 음성 인식 분야[19][20], 그리고 화자 인식 분야[21] 등에서 그 모델링 연구가 수행되어 오고 있다. 여기에서는 이런 다른 분야의 연구와 연계하여 연구될 수 있다는 가능성만을 언급하고, 연구의 범위를 원거리 음성 데이터의 지속시간을 분석하는 것만으로 한정하기로 한다.

4.1.1. 단어의 지속시간

다음의 <표 2>는 분석 대상이 되는 음성 데이터의 50개 단어를 음절수에 따라 분류한 다음, 각 단어들에서 원거리 음성과 평이한 음성의 단어 지속시간의 차이를 평균과 표준편차로 나타내고, 그 차이를 비율로 나타내었다. 여기에서 차이의 비율이란 평이한 음성과 비교하여 원거리 음성의 지속시간이 변화한 비율을 나타낸 것으로 (원거리 음성 지속시간 - 일반 음성 지속시간)/일반 음성 지속시간 * 100에 의하여 산출하였다.

<표 2> 단어의 지속시간 차이 및 그 비율

	Average difference of duration (단위: ms)	Percentage of difference (단위: %)
1-syllable word	74.25 (50.83)	25.28
2-syllable word	78.6 (105.09)	13.9
3-syllable word	79.63 (99.33)	12.24
4-syllable word	-3.07 (99.45)	-0.3
Average	57.35 (77.66)	9.08

(괄호 안의 숫자는 표준 편차를 나타냄)

<표 2>에서 보는 바와 같이, 원거리 음성의 단어 지속시간은 평이한 음성의 지속시간과 비교할 때 평균 9.08% 증가하였다. 1음절어의 경우에 그 증가 정도가 가장 두드러지고, 음절수가 증가함에 따라 그 차이 비율이 줄어든다. 4음절어의 경우는 전체 단어의 지속시간이 조금 감소하는데, 그 정도는 1%에 못 미치는 미미한 정도였다. 원거리 음성과 평이한 음성의 지속 시간 차이의 유의미성을 판별하기 위하여 T검증을 한 결과, T 값의 유의도가 4음절어의 경우에는 유의미하지 않

으나, 1음절어인 경우는 t=4.61, p<.001, 2음절어인 경우는 t=2.63, p<.05, 3음절어인 경우는 t=2.81, p<.05로 나타나서, 1음절어, 2음절어, 3음절어의 경우에는 모두 평이한 음성과 원거리 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 보이는데, 남성 화자들의 지속시간 증가 비율에 비하여 여성 화자들의 지속시간 증가비율이 두드러진 것을 볼 수 있었다. 지속시간의 증가 정도에 있어서 [8]의 지적과 같이 여성 화자에게서 좀 더 그 차이가 크게 나타나는 것으로 보인다. 이 결과를 통계적으로 분석하기 위하여 T검증을 하였는데, 예상과는 달리 남녀 간의 차이가 유의미하지 않다는 결과를 얻었다.

<표 3> 성별에 따른 단어의 지속시간 차이 및 그 비율

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

(Diff: Average difference of duration, SD: Standard Deviation, Pct: Percentage of difference)

4.1.2. 분절음의 지속시간

다음은 원거리 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 원거리 음성의 지속시간이 감소한 것을 관찰할 수 있었다. 전체 자음의 지속시간의 변화율은 평균 -9.10%로 원거리 음성의 지속시간이 감소함을 보인다. 반면에 모음의 경우는 단모음인 경우에 34.28%, 이중모음의 경우는 24.71%로 증가하였다.

그룹별 분절음에서 평이한 음성과 원거리 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<.01로 무성 자음과 모음의 경우에는 유의한 차이가 있었으나, 비음과 유음은 차이가 없는 것으로 나타났다. 즉, 한국어에 있어서 원거리 음성의 지속시간 차이는 무성 자음인 경우에는 감소하고 모음인 경우에는 증가하며 유성 자음은 그 차이가 없다고 볼 수 있다.

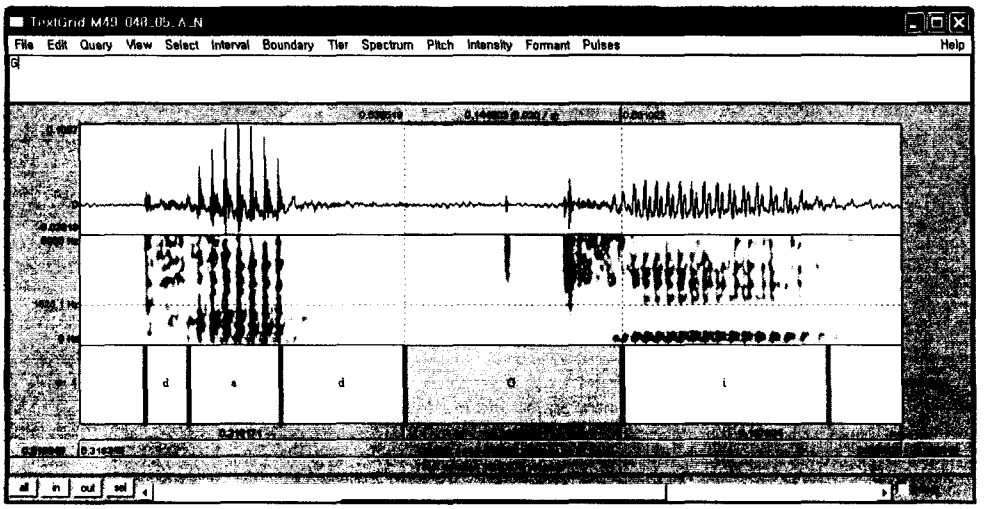
<표 4> 분절음의 지속시간 차이

	Consonants					Vowels	
	Plosives	Fricatives	Affricates	Nasals	Liquids	Mono-V	Diph-V
Diff (SD)	-14.54 (14.41)	-16.16 (17.87)	-8.53 (23.96)	-10.11 (23.17)	.58 (7.71)	44.09 (30.48)	45.21 (39.09)
Pct	-15.59	-15.65	-7.42	-7.77	.94	34.28	24.71

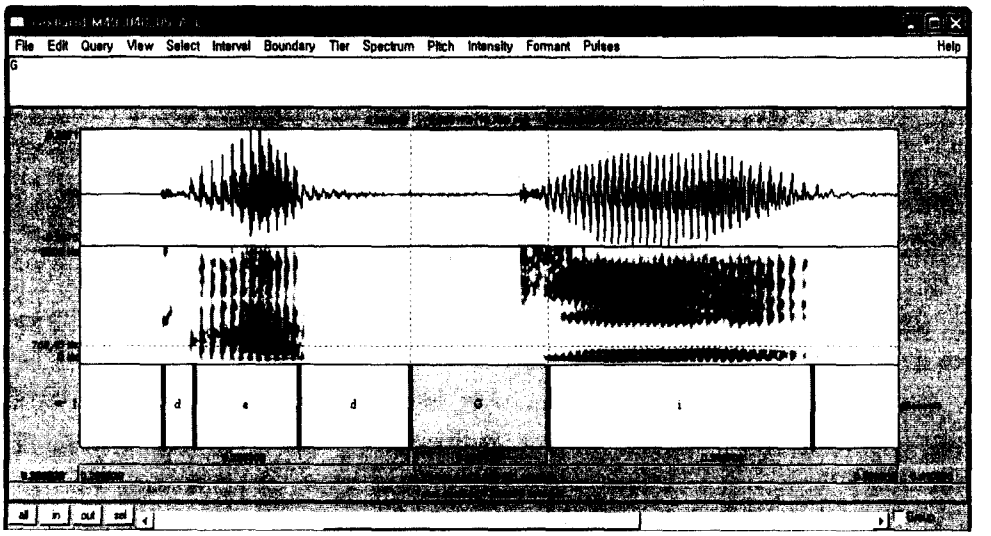
(Diff: Average difference of duration, SD: Standard Deviation, Pct: Percentage of difference)

원거리 음성을 평이한 음성과 비교하면 무성 자음이 짧아지고 모음이 길어지는데, 모음이 길어지는 비율이 자음이 짧아지는 비율보다 2배 이상 크므로, 결국 단어 전체의 지속시간은 위에서 이미 살펴본 바와 같이 증가하게 되었다고 볼 수 있다. 이러한 자음 감소 현상은 영어[5]와 스페인어[8]에서도 관찰된 현상으로, 한국어의 경우에는 원거리 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.

아래 <그림 4>와 <그림 5>는 단어 '닫기'의 평이한 음성과 원거리 음성을 각각 레이블링 한 것으로, 어두의 /ㄷ/과 어중의 자음 연쇄인 /ㄷ-ㄱ/의 길이가 평이한 음성 보다 원거리 음성에서 감소한 것을 볼 수 있다.



<그림 4> 평이한 음성 '닫기'의 레이블링 결과



<그림 5> 원거리 음성 '닫기'의 레이블링 결과

분절음 지속시간의 남녀간 성별 차이를 살펴보기 위한 표가 다음의 <표 5>이다. 자음에 있어서는 전체적인 경향과 같이 지속시간이 감소하는 것을 볼 수 있으나 그 감소율에 있어서는 차이를 보이는 것을 볼 수 있다. 남자의 경우에 자음의 지속시간 감소율은 평균 13.12%이고 여자는 7.99%로, 남자에 있어서 자음의 지속시간 감소율이 훨씬 높았다. 여성 화자의 경우에는 자음 가운데 지속시간이 감소하지 않고 증가하는 그룹으로는 유음이 있었다. 모음의 경우에도 지속 시간 증가율은 남녀 간의 차이를 보이는데, 남자의 경우 지속시간 증가율이 단모음과 이중모음에 있어서 각각 27.37%, 19.68%인데, 여자의 경우는 10% 내외의 차이를 보이는 39.9%와 28.81%로 높게 나타났다. 그러나 이미 위 <표 3>에서 살펴본 바와 같이 원거리 음성과 평이한 음성 간에 있어서 전체 단어의 지속 시간의 차이가 유의미하지 않음에 따라서 분절음 그룹별 성별 차이도 유의미하지 않다고 볼 수 있지만, 그 차이는 대략 볼 수 있었다.

<표 5> 성별에 따른 분절음의 지속시간 차이

		Consonants					Vowels	
		Plosives	Fricatives	Affricates	Nasals	Liquids	Mono-V	Diph-V
M	Diff	-17.45	-24.74	-18.64	-3.23	-1.86	31.59	32.34
	(SD)	(8.32)	(21.54)	(17.17)	(25.1)	(8.75)	(11.68)	(31.23)
	Pct	-19.95	-22.99	-16.41	-2.96	-3.3	27.37	19.68
F	Diff	-11.63	-7.57	-11.61	-16.98	3.03	56.58	58.07
	(SD)	(19.4)	(8.38)	(17.25)	(21.44)	(6.49)	(39.55)	(45.26)
	Pct	-15.74	-7.68	-9.76	-11.24	4.48	39.9	28.81

4.2. 지속시간 변이의 화자 유형 분석

지금까지는 평이한 음성과 원거리 음성의 지속시간 차이의 전체적인 경향을 단어와 분절음으로 나누어 살펴보았고, 그러한 차이가 다시 남녀 성별에 따라 어떻게 나타나는 지를 보았다. 이번에는 이러한 경향이 여러 화자 간에 어떻게 나타나는 지을 알아보고, 이러한 결과에 의하여 원거리 음성 발화시에 지속시간의 변이에 따라 화자들을 유형화하여 본다.

4.2.1. 단어의 지속시간

다음의 <표 6>은 10명 화자의 단어 지속 증가율(%)을 음절수에 따라 분류하여 나타낸 것이다. (여기에서는 이미 <표 2>에서 유의미하지 않은 결과를 보였던 4음절어는 제외하였다.) 1음절어의 경우에는 모든 화자가 원거리 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 원거리 음성의 지속시간이 증가하였다. 즉, 일반 음성을 발화할 때와 비교하여 원거리 음성을 발화하는 경우에 지속 시간 변이 양상은 음절수에 관계없이 지속시간이 증가하는 경우와, 증감이 동시에 관찰되는 경우로 나누어짐을 볼 수 있었다. 즉, 대부분의 화자가 원거리 음성을 발화하는 경우에 평이한 음성에 비하여 길게 발화하는 경향을 보이지만, 화자에 따라, 그리고, 단어를 구성하는 음절수에 따라, 그러지 않을 수도 있다는 것을 알 수 있다.

<표 6> 화자별 단어의 지속시간 차이 비율

	M1	M2	M3	M4	M5	F1	F2	F3	F4	F5
1syll W	12.02	51.49	22.98	43.49	10.92	16.73	9.35	47.73	25.76	12.32
2syll W	20.4	23.94	9.82	17.21	[illegible]			44.11	25.77	18.44
3syll W	14.37	18.53	7.74	13.34	[illegible]			40.27	24	12.16

4.2.2. 분절음의 지속시간

다음 <표 7>과 <표 8>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 증가율에

도 일정한 유형이 관찰된다고 할 수 있다.

<표 7> 화자별 분절음 지속시간 차이 비율(남성 화자의 경우)

	Plosives	Fricatives	Affricates	Nasals	Liquids	Mono-V	Diph-V
M1	-14.18	0.72	-6.29				
M2	-11.98	-23.94	0.34				
M3	-18.3	-16.78	-19.68	-3.07	-6.16		
M4	-30.39	-28.39	-33.3	-19.25	-14.5		
M5	-22.56	-16.6	-16.04	-18.89	-15.83		

<표 8> 화자별 분절음 지속시간 차이 비율(여성 화자의 경우)

	Plosives	Fricatives	Affricates	Nasals	Liquids	Mono-V	Diph-V
F1	-15.13	-17.8	-12.69	-31.77	-7.48		
F2	-19.61	-13.62	-17.92				
F3	-9.07	-2.86	-1.69	-9.43			
F4	-11.91	-3.73	-4.82	-11.38			
F5	-2.62	2.15	-5.7	-4.46	-2.24		

5. 결 론

본 논문에서는 원거리 음성 데이터를 분석하여 그 지속시간의 변화를 살펴보았다. 먼저, 원거리 음성을 일반 음성의 데이터를 비교하여 단어 전체의 지속시간과 분절음의 지속시간을 중심으로 분석한 결과, 1, 2, 3음절어의 경우에 단어 전체의 지속시간은 일반적으로 증가하고, 분절음의 지속 시간은 무성 자음의 경우에는 지속시간이 감소하는 반면에 모음의 경우는 증가한다. 이 때 모음의 증가율이 자음의 증가율보다 크기 때문에 전체 단어의 지속시간은 증가하게 된다고 할 수 있다. 원거리 음성에 있어서 성별에 의한 차이에 관해서는 이전의 연구에서 지적한 대로[8], 여성 화자의 경우에 지속시간의 특성과 관련하여 그 증가율이 남성 화자보다 높게 나타나는 경향을 보이지만, 통계적으로는 유의미하지 않았다.

다음으로는, 지속시간 증가율을 다시 화자별로 유형화하여 보았는데, 전체 단어 지속시간을 기준으로 분류할 때 음절수에 관계없이 모든 경우에 전체 지속시간이 증가하는 경우가 7명으로 대부분이었다고 할 수 있고, 나머지 3명의 경우는 2음절 이상으로 이루어진 단어의 경우에 감소하기도 하였다. 분절음의 지속시간으로 화자들을 분류했을 때는, 대부분의 경우 자음은 감소하고 모음은 증가하거나, 무성음은 감소하고 유성음은 증가하는 크게 두 가지의 양상을 보인다고 할 수 있었다.

이러한 결과는 대체적으로 전체 단어의 지속 시간은 증가하고, 자음의 경우는

감소하는 반면에 모음의 경우는 증가하는 것으로 알려진 기존의 롬바드 음성의 연구 결과와 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 여러 음향적/음성적 특성을 일괄적으로 분석한 것과 달리, 지속시간이라는 하나의 특성을 좀 더 정밀하게 분석하여, 한국어 원거리 음성의 특징을 구체적으로 규명하였다고 볼 수 있다. 또한 롬바드 음성이 화자 의존적인 특성이 많다고는 일반적으로 알려져 있지만 그 구체적인 사례는 발표된 바가 없었는데, 본 논문에서 원거리 음성의 경우에 지속시간이 일반적으로 증가하지만, 화자와 단어를 구성하는 음절수에 따라서는 다른 양상이 관찰되는 것을 보임으로써, 화자 의존성의 구체적인 사례를 제시했다는 의미를 부여할 수 있을 것이다.

감사의 글

본 논문에서 사용된 음성 데이터는 ETRI의 지원으로 KAIST에서 공동 제작되었습니다.

참 고 문 헌

[1] G. A. Fink and S. Hohenner, "Experiments in distant talking speech recognition using a standard database", *in Proc. of 31. Deutsche Jahrestagung für Akustik,* DAGA '05, München, 2005.

[2] M. Matassoni, M. Omologo, D. Giuliani, and P. Svaizer, "Hidden Markov model training with contaminated speech material for distant-talking speech recognition", *Computer Speech & Language*, 16(2), pp205-223, 2002.

[3] J.-C. Junqua, H. Wakita, "A comparative study of cepstral lifters and distance measures for all-pole modles of speech in noise", *in Proc. of ICASSP,* pp.841-844, 1989.

[4] J.-C. Junqua, "The Influence of Acoustics on Speech Production: A Noise-Induced Stress Phenomenon as the Lombard Reflex", *Speech Communication,* Vol. 20, pp.13-22, 1996.

[5] J. Hansen, "Analysis and Compensation of Speech under Stress and Noise for Environmental Robustness in Speech Recognition", *Speech Communication,* Vol. 20, pp.151-173, 1996.

[6] S. E. Bou-Ghazale, 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.

[7] E. Lombard, "Le signe de l'élévation de la voix", *Annales des maladies de l'oreille, du larynx, du nez et du pharynx,* Vol. 37, pp.101-119, 1911.

[8] A. Castellanos, 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.

[9] B. Stanton, L. Jamieson, G. Allen, "Acoustic-phonetic analysis of loud and Lombard speech in simulated cockpit conditions", *in Proc. of ICASSP,* Vol. 1, pp.331-334, 1988.

[10] J.-C. Junqua, "The Lombard Effect and its Role on Human Listeners and Automatic Speech Recognizers", *J. Acoust. Soc. Amer.,* Vol. 93-1, pp.510-524, 1993.

[11] S. Chi, Y-.H. Oh, "Lombard Effect compensation and noise suppression for Noisy Lombard Speech Recognition", *in Proc. of ICASSP,* pp.2013-2016, 1996.

[12] 우수영, *롬바드 효과 보상 필터를 이용한 강인한 특징 추출 방법,* 한국과학기술원 석사 논문, 2003.

[13] 이숙향, 신지영, 김봉완, 이용주, "음성 코퍼스 구축을 위한 SiTEC 분절음/운율 레이블링 기준의 검토 및 제안", *말소리,* 46호, pp.127-143, 2003.

[14] S.-J. Moon, B. Lindblom, J. Lame, "A perceptual study of reduced vowels in clear and casual speech", *in Proc. of XIIIth International Congress of Phonetic Sciences,* Vol. 2, pp.670-673, 1995.

[15] S. Köster, "Acoustic Characteristics of Hyperarticulated speech for Different Speaking Style", *in Proc. of ICASSP,* Vol. 2, pp.873-876, 2001.

[16] 이숙향, "발화속도와 한국어 분절음의 음향적 특성", *한국음향학회지,* Vol. 22-1, pp.14-22, 2003.

[17] J. van Santen, "Assignment of segmental duration in text-to-speech synthesis", *Computer Speech and Language,* Vol. 8, pp.95-128, 1994.

[18] H. Chung, "Duration Models and the Perceptual Evaluation of Spoken Korean", *Pro. Speech Prosody,* www.lpl.univ-aix.fr/sp2002/pdf/chung.pdf, 2002.

[19] J. E. Fosler-Lussier, *Dynamic Pronunciation Models for Automatic Speech Recognition,* Ph.D. thesis, University of California, Berkeley, 1999.

[20] J. Pylkkönen, M. Kurimo, "Duration Modeling Techniques for Continuous Speech Recognition", *in Proc. of ICSLP,* Vol. 1, pp.385-388, 2004.

[21] H. J. Künzel, "Some General Phonetic and Forensic Aspects of Speaking Tempo", *Forensic Linguistics,* Vol. 4-1 pp.48-83, 1997.

접수일자: 2005년 2월 18일
게재결정: 2005년 3월 24일

▶ 김선희(Sunhee Kim)
주소: 305-701 대전광역시 유성구 구성동 373-1번지 한국과학기술원 엘지홀 1108호
소속: 한국과학기술원(KAIST) 전자전산학과 전기및전자공학전공
전화: 042) 869-3493
E-mail: shkim@ee.kaist.ac.kr