

# 한국어 원거리 음성의 모음의 음향적 특성

이숙향(원광대), 김선희(KAIST)

## <차 례>

- |                  |                              |
|------------------|------------------------------|
| 1. 서론            | 3. 분석 결과 및 논의                |
| 2. 분석 대상 및 분석 방법 | 3.1. 발화 상황과 성별에 따른<br>음향적 특성 |
| 2.1. 음성 데이터      | 3.2. 화자 간 음향적 특성 차           |
| 2.2. 분석 방법       | 3.3. 모음 공간                   |
|                  | 4. 결론                        |

## <Abstract>

### Acoustic Characteristics of Vowels in Korean Distant-Talking Speech

Sook-hyang Lee, Sunhee Kim

This paper aims to analyze the acoustic effects of vowels produced in a distant-talking environment. The analysis was performed using a statistical method. The influence of gender and speakers on the variation was also examined. The speech data used in this study consist of 500 distant-talking words and 500 normal words of 10 speakers (5 males and 5 females). Acoustic features selected for the analysis were the duration, the formants (F1 and F2), the fundamental frequency and the total energy. The results showed that the duration, F0, F1 and the total energy increased in the distant-talking speech compared to normal speech; female speakers showed higher increase in all features except for the total energy and the fundamental frequency. In addition, speaker differences were observed.

\* Keywords: Distant-talking speech, Vowels, Acoustic characteristics, Gender differences, Speaker differences.

## 1. 서 론

본 연구는 일정한 거리에 있는 PC를 음성으로 조작하기 위한 경우에 관찰되는 한국어의 분절음의 음향적 변화 중 모음의 특징을 F0와 포먼트, 에너지 및 지속시간을 중심으로 평이한 음성과 비교하여 살펴보고자 한다. 이러한 연구는 원거리 음성의 보상 문제를 위한 기초 연구로서, 궁극적으로 음성인식 시스템 성능 개선에 기여하기를 기대한다.

본 연구에서 원거리 음성이란 일정한 거리에 있는 음성 시스템을 조작하기 위하여 화자가 음성을 보다 명료하게 전달하고자 하여 목소리를 높여 말하는 음성을 의미하는 것으로, 일반 음성과는 다른 여러 가지 음향적 효과가 발생하게 된다. 이러한 음향적 변화는 대상이 사람인 경우에 있어서는 일반적으로 그 명료성(intelligibility)으로 인하여 전달의 효과를 높이는데 반하여[1][2], 음성인식 시스템의 경우에 있어서는 오히려 그 성능을 저하시키는 것으로 알려져 있다[3][4]<sup>1)</sup>.

상대방에게 명료하게 들리도록 하기 위하여 목소리를 높이는 경우에 발생하는 효과는 여러 환경에서 관찰될 수 있는데, 잡음환경에서 의사소통을 시도하는 경우 [5][6][7]나 청각 장애인과 대화하는 경우[8][9], 거리가 떨어진 곳에 있는 사람과 대화하는 경우[10], 잘 알아듣지 못하는 외국인에게 말하듯이 말하는 경우[11], 일정한 거리에 있는 PC를 음성으로 조작하기 위한 경우[12], 또는 언어적 강세 혹은 포커스를 실현하기 위하여 명료하게 발음하는 등 다양하다[13][14].

기존의 연구들에 의하면 각각의 경우에 그 대상 언어 및 언어 수집 환경이나 수집 방법의 상이성으로 인하여 엄밀한 의미에서의 비교는 용이하지는 않지만, 대체로 (i) F0와 에너지 및 모음 지속시간의 증가, (ii) 에너지 분포의 이동, (iii) F1의 증가 등이 공통적인 음향학적 특징으로 보고 되었다. 연구 결과들을 환경별로 고찰하여 보면, 잡음환경에서의 말소리(롬바드 음성)의 음향적 특성을 연구한 [5]는 F0 증가, 저주파수대에서 중, 고주파수대로의 에너지 이동, 모음지속시간 증가, F1과 F2의 이동 등을 잡음환경의 말소리의 특징으로 기술하고 있다. [15]는 스페인어 롬바드 음성의 음향적 특징을 음소별 및 음소군별로 분석하였는데, 모음의 경우에 지속시간 증가, F0 증가, F1과 F2의 증가가 관찰되었고, 유성음에서 고주파수대(3-8 kHz)의 스펙트럼 기울기(spectral tilt) 변화율이 감소하고, 무성 마찰음에서는 고주파수대에서의 스펙트럼 기울기 변화율이 감소하며, 무성 마찰음을 제외한 모든 음소에서 저주파수대(0-3 kHz)의 스펙트럼 기울기 변화율이 증가한다고 보고하였다. 또한, F0의 경우는 남성 화자의 증가율이 두드러진데 반하여, F1의 증가와 유성음에서 고주파수대에서의 스펙트럼 기울기 변화율의 감소는 여성 화자에서 그 비율이 높게 나타나 성별에 따른 차이도 보고되었다. 프랑스어의 일음절 단모

1) 음성인식 시스템의 이러한 성능 저하는 학습 환경과 실제 사용 환경의 차이에 기인한 것으로 알려져 있다[3][4].

음으로 구성된 단어들을 0.5m(가까운 거리), 1.5m(보통 거리), 6m(먼 거리)에서 발음하도록 하여 녹음한 후 거리에 따른 말소리의 음향적 특성을 비교한 [10]은 원거리 음성의 특징으로 F0와 F1의 증가, F1, F2, F3의 에너지 증가와 함께 저주파수대보다는 고주파수대의 에너지 증가의 두드러짐(즉, 스펙트럼 기울기의 변화)을 보고하고 있다.

외국인에게 말하듯이 말하도록 하여 명료한 모음의 발음의 undershoot 특성을 연구한 [11]은 명료한 발음의 음향적 특성은 지속시간의 증가와 보통 발음에 비해 적은 양의 undershoot이며, undershoot은 포만트의 모음공간에서의 중앙화(centralization)가 아니라 인접자음의 포만트 주파수에 근접하는 것이고(즉, 동화) 그 정도는 모음자신의 지속시간과 포만트 전이의 속도에 따라 달라진다고 하였다.

실제로 사람이 명료한 발음을 인식하는 경우, 명료도가 증가한다는 연구보고가 주를 이루지만(예, 잠음환경에서 17-20% 포인트 명료도 증가 효과[1]), [8]은 청각 장애인을 대상으로 지각실험을 실시한 결과 모음공간의 확대와 포만트 전이의 벡터 증가 및 모음의 지속시간 증가를 보고하면서, 명료한 발음이 정상인에게는 명료도 증가효과가 있었으나 오히려 청각 장애인들에게는 명료도가 저하되었다고 보고했다. 이는 전설모음의 F2의 상승이 명료도 저하의 원인으로 F2의 주파수대(2000-2500 Hz)가 청각 장애인들의 점증적인 청력 상실이 일어나는 대역인 것으로 설명하고 있다. [9]는 화자가 한 말을 청자가 알아듣지 못한 척하여 화자로 하여금 다시 말하도록 유도하여 명료한 음성을 녹음하고 분석한 결과 F1의 증가와 F2의 감소, 모음의 지속시간과 F0의 증가를 관찰하였으며 더 나아가서 명료한 음성의 스펙트럼 기울기가 분절음의 종류에 따라 다르게 나타났다고 보고했다. 즉, 저해음의 스펙트럼 기울기는 완만한 반면, 모음을 비롯한 다른 음소의 스펙트럼 기울기는 급격하였다.

네덜란드어의 강세의 음향학적 특성을 분석하고 지각실험을 수행한 [13]과 [14]는 강세에 대한 지각적 단서로서 모음의 음가나 전체 에너지는 거의 역할이 없으며 모음의 지속시간과 함께 주파수대별 에너지의 분포(spectral balance)가 중요한 역할을 한다고 보고하고 있다. 여기서, 주파수대별 에너지의 분포는 강세 단어는 비강세 단어에 비해 상대적으로 저주파수대보다는 고주파수대의 에너지가 증가한다는 것을 의미한다.

그러나, 여러 다양한 환경에서 상대방에게 명료하게 들리도록 하기 위하여 목소리를 높여(increasing vocal effort, hyperarticulated speech) 명료하게 발음한다(speaking clearly)고 하는데 문제는 실제로 ‘목소리를 높인다’거나 ‘명료하게 발음한다’에 대한 명확한 정의를 내리지 않거나 또는 연구자들 각각 나름대로의 정의를 내리고 연구를 수행해오고 있다. 더욱이 연구에 따라 결과가 일치하지 않거나 그 결과가 서로 상반되는 경우도 종종 발견되기도 한다. 예를 들면, 앞에서 명료한 발음의 특성으로 느린 발화속도(즉, 긴 지속시간)와 F1의 증가 등을 들고 있는데,

F1의 증가는 오히려 빠른 발화(fast speech)의 특징으로 보고 되기도 할 뿐 아니라 [23], 영어 /i/와 /v/ 모음의 비강세 환경에서의 특징이기도 하다[16]. F1\*F2 모음공간에 관한 선행연구결과도 모음공간의 확대(즉, 모든 모음의 F1의 증가와 전설모음의 F2증가와 후설모음의 F2감소)를 보고하는 연구가 있는가 하면[8], F1만 증가하고 F2는 변하지 않아 모음공간의 면적의 변화가 없이 이동만 하는 경우도 관찰되었으며[5], F1은 증가하나 F2가 감소하는 경우도 관찰되었다[9]. 에너지 분포의 이동 또한 명료한 음성의 특성으로서 모음의 중, 고주파수대의 에너지 증가가 두드러져 완만한 스펙트럼 기울기를 보인다는 보고가 있는가 하면[10][13][14], 연구에 따라서는 앞에서 언급하였듯이 저해음에서는 고주파수대의 에너지 증가가 두드러져서 완만한 스펙트럼 기울기를 보이거나 다른 음소, 특히 모음에서는 오히려 저주파수대의 에너지 증가가 두드러져 급격한 스펙트럼 기울기를 보인다는 보고도 있다[9]. 많은 연구에서 전체 화자들의 명료한 음성의 특징으로 또는 성별에 따른 음향적 특성들의 차이를 보고하기도 하지만, 화자에 따라 명료한 음성을 실현시키는 책략이 다를 수 있다는 연구 결과들도 다수 있다([17][1][18] 등).

그 동안 한국어의 원거리 음성의 음향적 특성에 관한 연구는 많지 않으며 최근의 연구로 [19]는 일정한 거리에 있는 PC를 음성으로 조작하기 위한 경우의 한국어 원거리 음성의 지속시간 연구에서 저해음의 지속시간은 감소하는 반면 모음의 지속시간은 증가하는 것으로 보고하였다. [20]에 의하면 한국어 원거리 음성은 화자가 음성을 보다 명료하게 전달하고자 하여 목소리를 높여 발성하는 조음적 변화와 함께 거리에 의한 음성의 왜곡이 있게 되는데, 인식 실험을 통하여 조음적 변화에 의한 인식을 저하 정도와 거리에 의한 인식을 저하 정도를 살펴보고, 조음적 변화에 기인한 음향적인 특성을 피치, 지속시간, 포먼트, 에너지 및 대역별 에너지 분포를 중심으로 보고하였다. HTK를 이용한 실험 결과, 동일한 거리에서 수집된 평이한 음성과 원거리 음성의 차이는 거리에 의한 왜곡보다는 조음현상에 의한 인식을 저하로 볼 수 있는데, 조음변화에 의하여 인식이 9.19% 저하되는 것을 알 수 있었다. 그리고 원거리 음성의 음향적 특징으로는 모음의 지속 시간과 피치 및 에너지의 증가, 포먼트의 변화, 그리고 대역별 에너지가 저주파수 대역에서 중간 혹은 고주파수 대역으로 이동함을 보였다.

본 연구는 이러한 연구결과들을 바탕으로 한국어의 원거리 음성에 관한 연구를 시도한 것으로, 일정한 거리에 있는 PC를 음성으로 조작하기 위한 경우에 예상되는 음향적 변화들 가운데 모음의 특징을 F0와 포먼트, 에너지 및 지속시간을 중심으로 평이한 음성과 비교하여 살펴보고자 한다. 이러한 연구는 원거리 음성의 보상 문제를 위한 기초 연구로서, 궁극적으로 음성인식 시스템 성능 개선에의 기여를 기대할 뿐만 아니라, 한국어의 ‘명료한 발음’에 대한 음향적 특성을 살펴봄으로써 그 간의 연구결과들 간의 차이 또는 상반되는 결과들에 대한 원인 규명에 조그마한 기여를 할 수 있기를 기대한다.

## 2. 분석 대상 및 분석 방법

### 2.1. 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. 50개의 단어는 1음절어 4개, 2음절어 36개, 3음절어 6개, 4음절어 4개로 구성되었다. 발성 단어를 구성하는 모음으로는 7개의 단모음 /i, ɛ(ɨ), ɐ, ʌ, ɔ, ʌ, ɯ/, 그리고 6개의 이중 모음 /ɪɪ, ɛɛ, ɔɔ, ɛɪ, ɔɪ, ɔɯ/이 사용되었는데, 본 연구에서는 단모음만을 연구의 대상으로 삼았다.

[12]의 음성 수집에 참여한 화자로는 남성 화자 56명, 여성 화자 34명으로 총 90명으로 지역별 연령별로 구분하여 수집하였으나, 본 연구는 이 가운데 경기도 지역에 거주하는 남성 화자, 여성 화자 각 5명의 음성을 그 분석 대상으로 하였다.

평이한 음성의 수집 방법은 화자로부터 0.5m 지점에 설치한 마이크로 일상적인 어조로 발성된 음성을 수집한 것이다. 원거리 음성은 음성인식 시스템의 사용자가 3m 거리에 있는 시스템을 조작하기 위하여 목소리를 높여서 발성하도록 한 다음, 평이한 음성과 마찬가지로 화자로부터 0.5m, 1m, 2m와 3m의 지점에 설치된 마이크로 수집하였다. 본 연구에서는 화자와 0.5m 떨어진 지점에 설치된 마이크를 통해 녹음한 원거리 음성자료만을 이용하여 평이한 음성과 비교하였다. 따라서 각 화자 당 평이한 음성으로 발성한 단어 50개와 원거리 음성으로 발성한 단어 50개를 수집하여 총 평이한 음성의 500개 단어와 원거리 음성의 500개 단어가 분석 대상이 되었다.

### 2.2. 분석 방법

모음의 F0와 포먼트, 에너지 및 지속 시간을 측정하기 위하여 먼저 Praat 4.2.21을 이용하여 각 음성 파일을 수작업으로 하여 분절음 단위로 레이블링(labeling)하였다. 이 때 레이블링의 기준으로는 기본적으로 [21]을 따른 것으로, 분절음 레이블링에 있어서 음소 단위 이상의 세부적인 레이블링, 즉, 기식구간, 유성음화, 공명음화, 파찰음의 마찰구간, 탄설음의 공명음화 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 /r/의 경우는 /l/과 /r/을 구분하였다.

분절음 단위로 레이블링한 자료들 가운데 모음을 추출하여 Praat script를 이용

하여 F0와 F1, F2, 그리고 에너지 및 지속 시간을 측정하고 원거리 음성에서의 변화율을 계산하였다. F0와 에너지는 모음 전 구간의 평균값을 구하였으며 포만트는 모음의 50% 지점의 값을 구하였다. 그 결과를 SPSS 10.0을 이용하여 통계 분석을 하였으며 통계적 유의 수준은  $P < .05$ 이다.

### 3. 분석 결과 및 논의

#### 3.1. 발화상황과 성별에 따른 음향적 특성

발화상황과 성별이 원거리 음성의 음향적 특성, 즉 F0, 지속시간, 에너지, F1, F2에 미치는 영향을 보기 위하여 발화상황과 성별을 독립변수로 하여 이원분산분석을 실시하였다. 표 1에서 볼 수 있듯이 발화상황과 성별의 주요효과(main effect)는 F2의 발화상황을 제외하고 모두 통계적으로 유의하였으며 발화상황과 성별 간의 상호작용(interaction)은 F0와 F2를 제외한 지속시간, 에너지, 그리고 F1에서 통계적으로 유의하였다. 원거리 음성은 평이 음성에 비해 F0와 지속시간, 그리고 에너지 및 F1의 증가를 보였으며, 에너지를 제외한 나머지 특성에서 여성화자 그룹이 남성화자 그룹에 비해 큰 값을 보였다. 발화상황과 성별간의 상호작용을 보인 지속시간과 F1에서는 그림 1에서 명확히 볼 수 있듯이 여성화자그룹이 남성화자 그룹에 비해 평이 음성과 원거리 음성 간의 변화폭이 크게 나타난 반면 에너지는 반대로 남성화자그룹에서 큰 변화폭을 보이고 있다.

이 결과는 표 2의 평이음성과 원거리 음성 간의 변화율에서 더욱 확연히 드러난다. 화자 10명 전체의 평균변화율을 보면 모든 음향적 특성에서 양수의 변화율을 보이고 있으며, 성별 간 차이를 보면 남성화자 그룹이 F0와 에너지, 그리고 F2에서 여성화자 그룹에 비해 큰 변화율을 보이는 반면 지속시간과 F1은 여성화자 그룹에 비해 적은 변화율을 보이고 있다. 여성화자 그룹의 F0 증가 폭이 남성화자 그룹에 비해 적은 것은 여성은 평이 음성에서 이미 비교적 높은 F0값을 지니고 있어서 원거리 음성에서 증가 폭이 남성에 비해 클 수가 없을 것이며, 따라서 지속시간이나 포만트 변화율 등에 더 의존하는 것이 아닐까 추정해본다. 그러나 이는 단지 하나의 가정일 뿐 이에 대한 실험적 연구가 있어야 할 것이다.

#### 3.2. 화자 간 음향적 특성 차

화자별로 각 음향적 특성의 변화율을 구한 결과 표 3에서 볼 수 있는 바와 같이 화자 간에 변화율의 차이를 보이고 있다. 화자 간 변화율 차이를 보기 위하여 화자를 독립변수로 하여 일원분산분석을 실시하였다. 표 4에서 볼 수 있는 바와

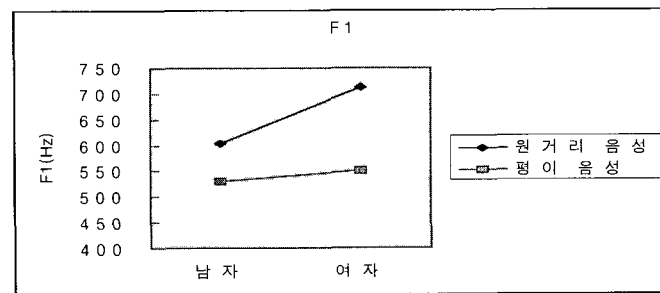
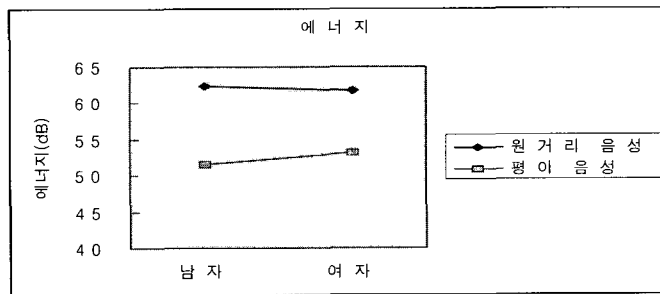
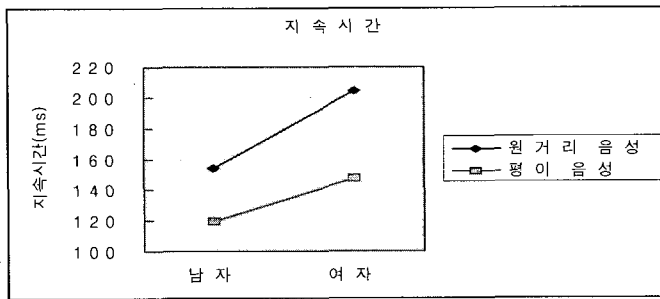
같이 모든 음향적 특성에서 화자의 영향이 통계적으로 유의하였다. Tukey 사후검정 결과(P<.05) F0는 여2, 여5<여3, 여1, 남1< 남1, 남17, 여79< 남74, 남40, 남48 변화율 순으로 4개 그룹으로 나뉘었다. 지속시간은 여1, 남5, 남1, 남4, 여5, 남3, 여2< 남4, 여5, 남3, 여2, 여4 <남3, 여2, 여4, 남2<여4, 남2, 여3 변화율 순으로 4개 그룹으로 나뉘었다. 에너지는 여1< 여2 < 여3, 여5, 남1 < 여5, 남1, 남5, 남2 < 남1, 남5, 남2, 남4 < 남5, 남2, 남4, 남3 < 여4 변화율 순으로 7개 그룹으로 나뉘었다. F1값은 남1, 남4, 여2, 남2, 남3, 여5 < 남4, 여2, 남2, 남3, 여5, 여3, 여4 < 여5, 여3, 여4, 남5, 여1 변화율 순으로 3개 그룹으로 나뉘었으며 F2값은 여1, 남2, 여4, 남3, 남1, 여5, 여3, 여2, 남5 < 여5, 여3, 여2, 남5, 남4 변화율 순으로 2개 그룹으로 나뉘었다.

<표 1> 발화상황과 성별에 따른 원거리 음성과 평이 음성의 F0, 지속시간, 에너지, F1, F2의 평균값과 표준편차 및 t-test 결과. 음영으로 표시한 경우는 통계적으로 유의한 차이가 있음을 나타냄.

음향적 특성 발화상황	F0 (Hz)	지속시간 (ms)	에너지 (dB)	F1 (Hz)	F2 (Hz)
원거리 음성	<u>256.33</u> (63.54)	<u>178.91</u> (104.26)	<u>63.51</u> (6.10)	<u>658.35</u> (244.11)	1795.64 (616.58)
평이 음성	<u>170.79</u> (52.12)	<u>133.37</u> (73.567)	<u>52.32</u> (7.05)	<u>540.76</u> (235.80)	1805.78 (583.37)
음향적 특성 성별	F0 (Hz)	지속시간 (ms)	에너지 (dB)	F1 (Hz)	F2 (Hz)
남자	<u>170.63</u> (55.44)	<u>136.56</u> (74.67)	<u>58.36</u> (9.07)	<u>570.48</u> (239.80)	<u>1729.47</u> (573.86)
여자	<u>256.60</u> (60.37)	<u>175.76</u> (104.80)	<u>57.44</u> (8.18)	<u>637.17</u> (238.75)	<u>1831.55</u> (648.91)
음향적 특성 (남자)발화상황	F0 (Hz)	지속시간 (ms)	에너지 (dB)	F1 (Hz)	F2 (Hz)
원거리 음성	<u>212.62</u> (43.09)	<u>153.96</u> (79.85)	<u>62.26</u> (6.63)	<u>602.51</u> (246.77)	1726.88 (587.69)
평이 음성	<u>128.64</u> (27.66)	<u>119.16</u> (64.64)	<u>51.47</u> (5.01)	<u>530.14</u> (219.96)	1748.18 (522.50)
음향적 특성 (여자)발화상황	F0 (Hz)	지속시간 (ms)	에너지 (dB)	F1 (Hz)	F2 (Hz)
원거리 음성	<u>300.15</u> (48.85)	<u>203.92</u> (118.94)	<u>61.73</u> (4.91)	<u>711.38</u> (229.54)	1860.94 (636.61)
평이 음성	<u>213.04</u> (33.26)	<u>147.61</u> (79.08)	<u>53.18</u> (8.55)	<u>550.84</u> (249.73)	1860.48 (631.54)

<표 2> 화자 10명 전체의 변화율과 성별에 따른 변화율. 변화율 공식은 ((원거리 음성 - 평이 음성) / 평이 음성) \* 100임.

음향적 특성 화자그룹	F0 (%)	지속시간 (%)	에너지 (%)	F1 (%)	F2 (%)
전체	55.12 (29.16)	40.75 (60.48)	22.84 (15.06)	31.82 (62.10)	3.67 (39.01)
남자	67.19 (27.11)	37.00 (49.71)	27.01 (8.64)	25.27 (78.43)	5.59 (52.73)
여자	43.03 (25.97)	45.41 (69.37)	18.65 (18.58)	38.05 (40.09)	1.85 (18.00)



<그림 1> 발화상황과 성별 간에 상호작용을 보이는 지속시간, 에너지, F1.



이를 표 5에서와 같이 각 음향적 특성별로 작은 값을 보이는 화자에서 큰 값을 보이는 화자 순으로 열거하면 화자 간 특성을 보다 일목요연하게 볼 수 있다. F0 변화율과 에너지 변화율, 그리고 F1 변화율에서 성별 간 차이가 뚜렷이 보이거나, F2 변화율에서는 성별 간에 일관적인 차이를 볼 수 없다. 지속시간 변화율에서는 F0 변화율과 에너지 변화율만큼 아주 뚜렷하지는 않으나, 화자 여1을 제외하면 대체적으로 큰 변화율을 보이고 있다. F0 변화율과 에너지 변화율에서는 화자 여4를 제외한 네 명의 여성화자들이 남성화자들에 비해 작은 값을 보이고 있으며, 예외적으로 화자 여4는 F2를 제외한 네 가지 음향적 특성의 변화율에서 모두 큰 값을 보이고 있다. 반면에 F1 변화율에서는 화자 여2를 제외한 여성화자들이 남성화자들에 비해 큰 값을 보이고 있다. 여기에서 주목할 만한 것은 원거리 음성에서 F0 변화율과 에너지 변화율이 크지 않은 여성화자들 중 화자 여1은 F1 변화율에서 비교적 큰 값을 보이고 있는 반면, F0와 에너지 변화율에서뿐만 아니라 F1 변화율에서도 큰 값을 보이지 않은 화자 여2는 지속시간과 F2 변화율에서 비교적 큰 값을 보이고 있어, 원거리 음성 전달에서의 음향적 특성이 성별 간 차이만이 아니라 화자 간 차이 또한 보이고 있다는 점이다. 이는 발화속도에 따른 한국어 단모음의 음향학적 특성을 연구한 [22]에서 화자마다 다른 책략을 쓸 수 있음을 보여준 연구 결과와 함께 화자 간 차이에 대한 연구의 필요성을 보여주는 연구 결과라고 할 수 있겠다.

<표 3> F0, 지속시간, 에너지, F1, F2의 화자별 변화율. 변화율 공식은  

$$((\text{원거리 음성} - \text{평이 음성}) / \text{평이 음성}) * 100\text{임.}$$

음향적 특성 화자	F0 (%)	지속시간 (%)	에너지 (%)	F1 (%)	F2 (%)
남1	48.82	30.85	23.17	4.08	0.65
남2	58.87	62.09	27.08	26.35	-0.72
남3	76.40	38.76	30.23	26.78	-0.10
남4	79.68	31.74	27.97	15.78	19.87
남5	71.50	17.66	26.88	50.18	7.67
여1	47.03	13.40	-0.84	59.15	-0.95
여2	30.53	39.61	11.99	22.56	6.64
여3	44.84	81.17	19.26	35.90	2.56
여4	59.70	57.50	40.38	39.43	-0.4625
여5	32.38	35.12	22.26	32.74	1.55

&lt;표 4&gt; 화자를 독립변수로 한 F0, 지속시간, 에너지, F1, F2에 대한 일원분산분석 결과.

음향적 특성	통계분석 F 비율	Tukey 사후검정 (P<.05)
F0	F(9,911)=46.999, P=.000	여2, 여5 < 여3, 여1, 남1 < 남1, 남17, 여79 < 남74, 남40, 남48
지속시간	F(9,925)=12.274, P=.000	여1, 남5, 남1, 남4, 여5, 남3, 여2 < 남4, 여5, 남3, 여2, 여4 < 남3, 여2, 여4, 남2 < 여4, 남2, 여3
에너지	F(9,925)=103.403, P=.000	여1 < 여2 < 여3, 여5, 남1 < 여5, 남1, 남5, 남2 < 남1, 남5, 남2, 남4 < 남5, 남2, 남4, 남3 < 여4
F1	F(9,881)=6.20, P=.000	남1, 남4, 여2, 남2, 남3, 여5 < 남4, 여2, 남2, 남3, 여5, 여3, 여4 < 여5, 여3, 여4, 남5, 여1
F2	F(9,881)=2.431, P=.010	여1, 남2, 여4, 남3, 남1, 여5, 여3, 여2, 남5 < 여5, 여3, 여2, 남5, 남4

&lt;표 5&gt; F0, 지속시간, 에너지, F1, F2 값의 화자 간 순서로서 작은 값을 보이는 화자에서 큰 값을 보이는 화자 순으로 열거함. 여성과 남성화자 간에 뚜렷한 대조를 보이는 음향적 특성에는 시각적 효과를 위하여 그룹핑을 보이는 여성 화자들을 음영으로 표시하였음.

F0	지속시간	에너지	F1	F2
<b>여2</b>	여1	<b>여1</b>	남1	여1
<b>여5</b>	남5	<b>여2</b>	남4	남2
<b>여3</b>	남1	<b>여3</b>	여2	여4
<b>여1</b>	남4	<b>여5</b>	남2	남3
남1	<b>여5</b>	남1	남3	남1
남2	남3	남2	<b>여5</b>	여5
여4	<b>여2</b>	남4	<b>여3</b>	여3
남5	<b>여4</b>	남5	<b>여4</b>	여2
남3	남2	남2	남5	남5
남4	<b>여3</b>	여4	<b>여1</b>	남4

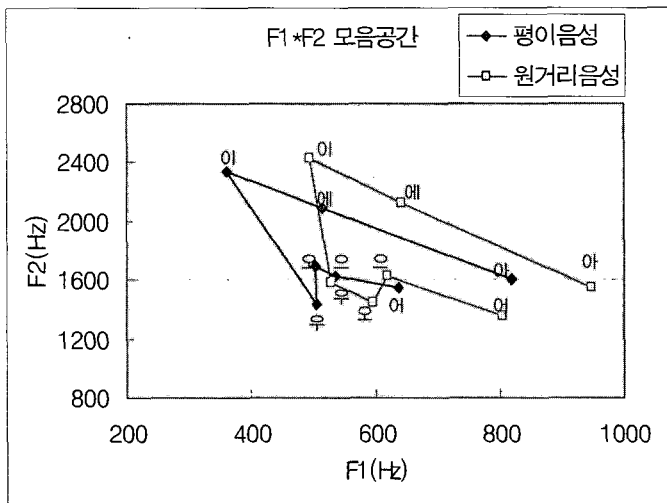
### 3.3. 모음 공간

원거리 음성 전달에서의 여러 음향적 특성이 성별 간 차이만이 아니라 화자 간 차이 또한 보이고 있는데 F1\*F2 모음공간에서는 어떠한 양상을 보일 것인가? 서론에서 언급한 바와 같이 모음공간에 관한 선행연구를 보면 모음공간의 확대에서 축소까지 다양한 결과를 보여주고 있다. 본 연구에서는 평이 음성과 원거리 음성의 F1\*F2 모음공간의 차이를 보기 위하여 우선 모음별 paired samples t-test를 실시하였다. 표 6에서와 같이 F1은 /우/를 제외한 나머지 모음에서 유의한 차이를 보

였으며, F2는 /아, 이, 어/ 모음만이 유의한 차이를 보였고, 원거리 음성에서 /아/와 /어/ 모음의 F2값은 감소한 반면 /이/ 모음에서는 증가하였다. 이 결과를 이용하여 그림 2에서와 같이 모음공간을 그려본 결과, F1의 상승이동이 뚜렷이 관찰되었다.

<표 6> 발화상황에 따른 각 모음의 F1, F2 평균값과 표준편차 및 t-test 결과. 음영으로 표시한 경우는 통계적으로 유의한 차이가 있음을 나타냄.

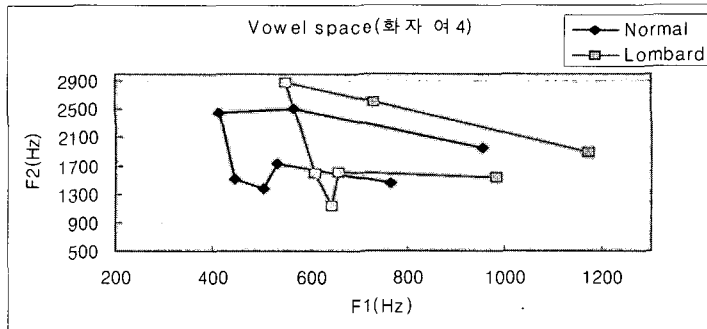
	F1		F2	
	평균	원거리	평균	원거리
아	<b>818.17</b>	<b>946.50</b>	<b>1602.76</b>	<b>1549.88</b>
	(195.72)	(169.00)	(343.90)	(267.95)
에/애	<b>515.02</b>	<b>641.72</b>	2088.43	2118.69
	(89.761)	(101.57)	(355.92)	(328.075)
이	<b>361.53</b>	<b>494.42</b>	<b>2338.14</b>	<b>2422.66</b>
	(184.32)	(278.70)	(331.57)	(352.45)
오	<b>503.38</b>	<b>596.26</b>	1428.76	1439.42
	(125.39)	(116.61)	(670.69)	(713.93)
우	501.83	526.66	1689.49	1578.95
	(387.47)	(116.76)	(516.06)	(385.55)
으	<b>535.71</b>	<b>618.49</b>	1620.99	1622.73
	(206.87)	(183.67)	(498.19)	(482.21)
어	<b>638.76</b>	<b>805.68</b>	<b>1546.90</b>	<b>1353.44</b>
	(183.88)	(167.49)	(414.05)	(347.96)



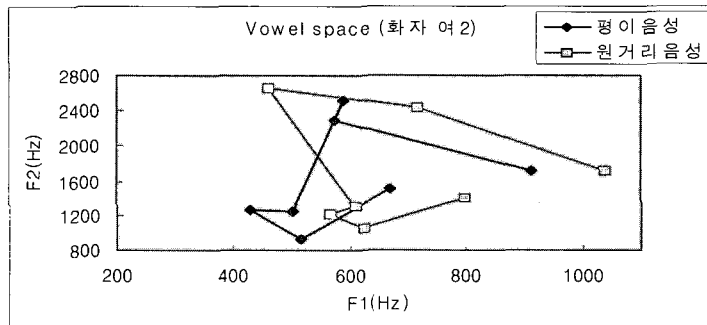
<그림 2> 화자 10명 전체의 평균 F1\*F2 모음공간.

앞 절에서 원거리 음성 전달에 있어 화자간의 전략이 다를 수 있음을 보았는데 각각의 화자별 F1\*F2 모음공간에 대한 분석 결과, 모음공간에 있어서도 화자에 따라 각기 다른 전략을 쓰고 있음을 볼 수 있었다. 예를 들어 원거리 음성에서 화자 여4와 여5는 F1값을 증가시키고 F2값은 후설모음에서는 변화를 주지 않고 비후설 모음에서만 증가시켜 모음 전체의 개구성을 증가시키고 비후설 모음의 전설성을 증가시킴으로써 어느 정도의 모음공간의 확대를 꾀한 반면(그림 3 참조), 화자 여2는 F1값과 F2값을 모두 증가시켜 개구성과 전설성을 증가시킴으로써 모음공간의 크기의 변화는 없이 전체적인 이동만을 보이고 있다(그림 4 참조). 화자 남3과 남2는 F1값을 증가시키고 F2값은 비후설 모음에서는 증가시킨 반면 후설모음에서는 감소시켜 모음 전체의 개구성을 증가시키고 비후설 모음에서는 전설성을 증가시킴과 동시에 후설모음에서는 후설성을 증가시킴으로써 모음공간의 확대를 가져왔다(그림 5 참조). 화자 여1, 여3과 남5는 F2값의 변화 없이 F1값의 증가, 즉 개구성의 증가만 있었다(그림 6 참조). 예외적으로 화자 남1은 F1값과 F2값 모두에서 감소를 보였다. 즉, 화자 남1을 제외한 9명의 화자들은 모두 F1의 증가를 보였으며 F2는 화자에 따라, 모음에 따라 감소에서 증가까지 다양한 변화 유형을 보였다.

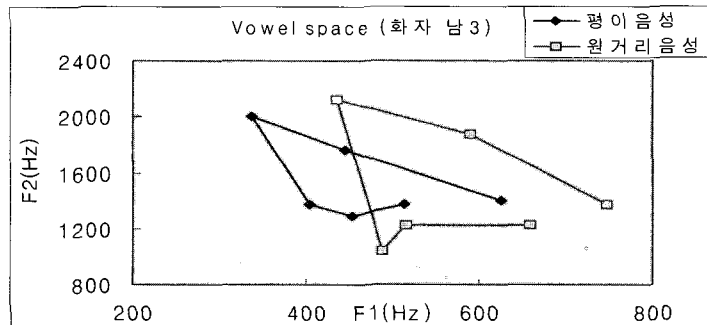
즉, 원거리 음성에서는 평이 음성에 비해 화자 공통적으로 입을 더 많이 벌리고 있으며 화자에 따라 모음의 전/후설성에 따라 그 특성을 그대로 유지하거나 또는 증가시킴으로써 모음공간의 크기를 그대로 유지하거나 또는 확대시키고 있다. 이렇게 화자에 따라 원거리 음성 전달 책략에 있어 다양한 형태를 보이는데 과연 각각의 책략이 모두 같은 전달 효율성을 보일지는 의문이다. 이 화자들의 발화를 대상으로 지각실험을 할 경우 명료도에 있어 화자 간에 어떤 차이를 보일지, 그리고 여러 음향적 특성 간에 전달의 효율성에 있어 어떤 차이를 보일지 살펴보는 것이 흥미로운 것이다. 또 한 가지 흥미로운 사실은 발화속도가 느린 원거리 음성의 음향적 특성 중 F1의 증가는 빠른 발화속도의 특성이기도 할 뿐만 아니라 비강세 모음의 특성이기도 하다는 것이다. 이는 보통 속도로 평이하게 말할 때가 해당 분절음의 목표점에 가장 가깝게 도달할 가능성을 간접적으로 시사한다고 볼 수 있지 않을까 한다.



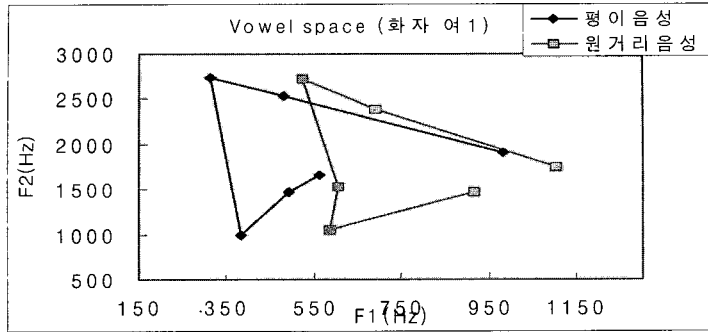
<그림 3> F1값 증가, 비후설 모음 F2값 증가 예 (화자 여4).



<그림 4> F1과 F2 모두 증가 예 (화자 여2).



<그림 5> F1 증가, 비후설 모음 F2 증가, 후설모음 F2 감소 예 (화자 남3).



<그림 6> F1 증가 예 (화자 여1).

#### 4. 결 론

본 연구에서는 한국어 원거리 음성의 모음 특징을 F0와 F1, F2 포먼트, 에너지 및 지속 시간을 중심으로 살펴보았다. 분석 대상으로는 5명의 남성 화자와 5명의 여성 화자가 50개 단어를 원거리에서 발성했을 경우와 평이한 음성으로 산출한 녹음 자료이며, 이에 대한 음향 분석을 통하여 원거리 음성의 음향적 특성을 살펴보았다. 그 결과, 원거리 음성에서 일반적으로 F0, 지속시간, 에너지와 F1 평균값이 평이 음성에 비해 큰 값을 보였다. 성별 간 차이 또한 보였는데 에너지 평균값을 제외한 F0, 지속시간, F1, F2 평균값이 여성화자 그룹이 남성화자 그룹에 비해 큰 값을 보였다.

평이 음성 대비 원거리 음성의 변화율은 F0, 에너지, F2에서 남성화자 그룹이 더 큰 값을 보인 반면, 지속시간과 F1에서는 여성화자 그룹이 더 큰 값을 보였다. 화자 간 변화율의 차이 또한 관찰되었다.

F1\*F2 모음공간에서는 10명의 화자 중 한명의 화자를 제외하고는 공통적으로 모두 F1 증가를 보였으며 F2 증감에 있어서는 화자에 따라, 모음에 따라 감소에서 증가까지 다양한 양상을 보였다.

향후 연구에서는 보다 많은 화자를 대상으로 화자별 음향적 특성을 살펴보아야 할 것이며, 본 연구에서 다루지 못한 스펙트럼 기울기와 같은 음향적 특성에 대한 연구 또한 시도되어야 할 것이다. 더 나아가서 음향적 특성과 조음적 특성을 동시에 살펴볼 수 있는 연구도 이루어져야 할 것이다.

## 감사의 글

본 논문에서 사용된 음성 데이터는 ETRI의 지원으로 KAIST에서 공동 제작되었습니다. 또한, 본 연구는 제 1저자의 원광대학교 2004학년도 교내 일반과제 지원으로 수행되었습니다.

## 참고 문헌

- [1] M. A. Picheny, N. I. Durlach, and L. D. Braidia, "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," *J. Speech Hear. Res.* Vol. 28, pp.96-103, 1985.
- [2] K. L. Payton, R. M. Uchanski, and L. D. Braidia, "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," *J. Acoust. Soc. Am.*, Vol. 95, pp. 1581-1592, 1994.
- [3] S. Oviatt, M. MacEachern, and G. Levow, "Predicting hyperarticulated speech during human-computer error resolution," *Speech Communication* Vol. 24, No. 2, pp.1-23, 1998.
- [4] H. Soltau, and A. Waibel., "Specialized acoustic models for hyperarticulated speech," *Processings of the International Conference on Spoken Language Processing*, Istanbul, Turkey, 2000.
- [5] 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.
- [6] 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.
- [7] E. Lombard, "Le signe de l'lvation de la voix". *Annales des maladies de l'oreille, du larynx, du nez et du pharynx*, Vol. 37, 101-119, 1911.
- [8] S. H. Ferguson, and D. Kewley-Port, "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* Vol. 112, No. 1, pp.259-271, 2002.
- [9] S. Köster, "Acoustic Characteristics of Hyperarticulated speech for different speaking style," *Proc. ICASSP*, Vol. 2, pp.873-876, 2001.
- [10] J.-S. Linéard, and M.-G. Di Benedetto, "Effect of vocal effort on spectral properties of vowels," *J. Acoust. Soc. Am.* Vol.106, No.1, pp.411-422, 1999.
- [11] S. J. Moon & B. Lindblom, "Interaction between duration, context, and speaking style in English stressed vowels," *J. Acoust. Soc. Am.* Vol. 96, No. 1, pp.40-55, 1994.
- [12] 우수영, 롬바드 효과 보상 필터를 이용한 강인한 특징 추출 방법, 한국과학기술원 석사 논문, 2003.
- [13] A. M. C. Sluijter and V. J. van Heuven, "Spectral balance as an acoustic correlate of linguistic stress," *J. Acoust. Soc. Am.* 100(4), 2471-2485, 1996.
- [14] A. M. C. Sluijter, V.J. van Heuven, and J.J.A. Pacilly, "Spectral balance as a in the

- perception of linguistic stress,” *J. Acoust. Soc. Am.* Vol. 101, No. 1, pp.503-513, 1997.
- [15] A. Castellanos, J.-M. Benedi, and F. Casacuberta, “An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect”, *Speech Communication* 20, pp.23-35, 1999.
- [16] U. Stålhamma, I. Karlsson, and G. Fant, “Contextual effects on vowel nuclei,” *STL-QPSR* 4, 1-18, 1973.
- [17] J.-P. Gagné, V. M. Masterson, K. G. Munhall, N. Bilida, and C. Querengesser, “Across talker variability in auditory, visual, and audio-visual speech intelligibility for conversational and clear speech,” *J. Acad. Rehab. Audiol.* 27, 135-158, 1994.
- [18] S. H. Ferguson, “Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners, *J. Acoust. Soc. Am.* 116(4), 2365-2373, 2004.
- [19] 김선희, “한국어 롬바드 음성의 지속시간,” *말소리* 54호, pp.1-14, 2005.
- [20] S. Kim, S. Y. Park and C. D. Yoo, “An analysis of acoustic features caused by articulatory changes for Korean distant-talking speech,” *The Journal of the Acoustical Society of Korea*, Vol. 24-2E, pp.1-6, 2005.
- [21] 이숙향, 신지영, 김봉완, 이용주, “음성 코퍼스 구축을 위한 SiTEC 분절음/운율 레이블링 기준의 검토 및 제안,” *말소리* 46호, pp.127-143, 2003.
- [22] 이숙향, 고현주, “발화속도와 한국어 분절음의 음향학적 특성,” *한국음향학회지* 제23권 2호, pp.162-172, 2004.
- [23] R. J. J. H. Van Son and L. C. W. Pols, “Formant movements of Dutch vowels in a text, read at normal and fast rate”, *J. Acoust. Soc. Am.* Vol. 92, pp.121-127, 1992.

접수일자: 2005년 8월 21일

게재결정: 2005년 9월 12일

▶ 이숙향(Sook-Hyang Lee)

주소: 570-749 전북 익산시 신용동 344-2 원광대학교 인문대학 영중어문학부

소속: 원광대학교 인문대학 영중어문학부

전화: 063) 850-6913

E-mail: shlee@wonkwang.ac.kr

▶ 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