

# 모방발화의 기본주파수 연구

박미영\* 신지영\*\* 강선미\*\*\*

\*고려대학교 민족문화연구원 음성언어정보연구실

\*\*고려대학교 국어국문학과, \*\*\*서경대학교 컴퓨터학과

## A Study of Fundamental Frequency about Voice Imitation

Park, Miyoung\* Shin, Jiyoung\*\* Kang, Sunmee\*\*\*

\*Spoken Language Information Lab, Institute of Korean Culture, Korea University

\*\*Department of Korean Language and Literature, Korea University

\*\*\*Department of Computer Science, Seokyeong University

pikel@unitel.co.kr, shinjy@korea.ac.kr, smkang@skuniv.ac.kr

### Abstract

The purpose of this paper is to find prosodic characteristics in voice imitation. Speakers change various phonetic features in voice imitation. Speakers change their pitch ranges in the most cases. Especially, the pitch range is important for word conditions. And, as imitators change the voice, the average value of  $f_0$  is close to high frequency than low frequency or middle level.

### I. 서론

본 연구<sup>1)</sup>는 모방 발화 시에 변화는 음성 요소 및 변하지 않는 음성 요소를 살펴, 화자 인식에 활용할 수 있는 화자 고유의 음성을 찾는 것을 연구의 목표로 한다. 이를 위해 본 연구는 문장 및 단어 수준에서의 기본주파수 변화 양상을 대상으로 모방 발화의 특징을 살펴보았다.

### II. 연구 방법

본 연구는 두 가지 방법으로 실험 자료를 수집하였다. 우선은 단일 목표 음성을 모방하는 다수 모방 화자의 음성 모방 자료를 대상으로 수집하였고, 또한 일부 모방 화자에게서 음성 모방이 가능한 목표 음성을 모두 수행토록 하여 단일 모방 화자에 대한 복수 목표 모방 음성을 자료로 구축하였다.

본 연구를 위해 음성 모방이 가능하며 모방 수준이 적절한 피실험자 6명을 선정하였다. 본 연구의 실험에

참여한 피실험자들의 명단은 다음과 같다

LJH	남	25세	LKR	남	19세
KHJ	남	27세	CSE	남	25세
CSB	남	24세	CWI	남	26세

피실험자 LKR의 녹음은 조용한 방의 PC상에서 이루어졌으며, 녹음에 사용된 프로그램은 GoldWav v.4 26, 마이크는 SHURE 58SM이다. 이 외의 피실험자들의 녹음은 일방향 마이크 C414 B-ULS와 TASCAM DA-20MK DAT 녹음기기를 이용하여 고려대학교 음성언어정보연구실의 녹음실에서 이루어졌다. 음성 파일은 Cool Edit Pro를 이용하여 16bit, 1로 디지털화하였다. WaveSurfer 1.5 2를 사용하여 기본주파수값들을 분석하였다. 기본주파수값들의 계산을 위해서는 pitch\_chal(2004 3 자체 제작)이라는 프로그램을 이용하였는데, 예측되지 않은 기본주파수값의 오류는 모두 개별적으로 검토 후 수정하였다.

피실험자들은 연구의 목적에 맞게 설계된 연설문 텍스트와 단어 목록<sup>2)</sup>을 단일 목표 음성인 '김대중 전 대통령'의 음성으로 동일하게 발화하였다. 이 중 LJH와 LKR은 복수 음성 모방이 가능하여 추가로 복수 목표 발화를 각 4명, 8명씩 수행하였다.

분석 대상은 복수 목표 발화를 수행한 피실험자 LJH가 '박준형, 신문선, 김제동, 김제동'의 자료, 피실험자 LKR이 모방한 '김대중, 권영길, 노무현, 이규화(성우, X파일 멀더역), 이회창, 정몽준 -이상 남성 6명

1) 본 논문은 한국과학재단 목적기초연구(R01-1999-000-00229-0) 지원으로 수행되었습니다

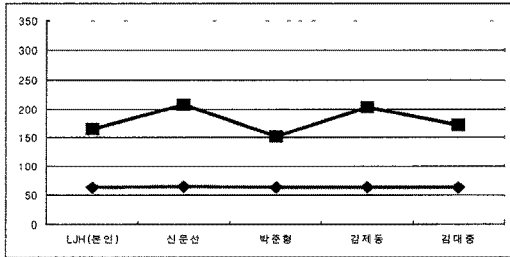
2) 연설문 텍스트는 총 22문장으로 구성되어 있으며, 추가로 단어 30개와 한국어의 단모음 7개를 각 2회씩 발화하도록 하였다

엄앵란, 전도연 -이상 여성 2명'의 자료와 다른 4명의 피실험자가 모방한 '김대중 전 대통령'의 자료이다.

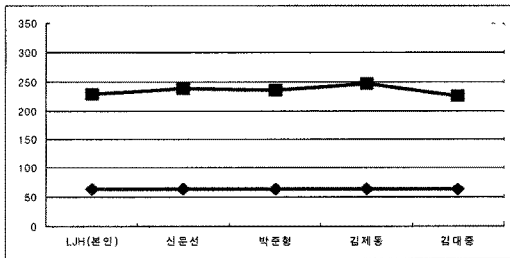
### Ⅲ. 분석 결과 및 논의

#### 3.1 기본주파수 대역

각 피험자의 문장 녹음과 단어 녹음 자료를 대상으로 본인 발화, 모방발화의 기본주파수 대역폭을 측정하였다. 이때에 기본주파수 대역폭의 값은 절대값을 의미하는데, 분석 결과는 다음과 같다

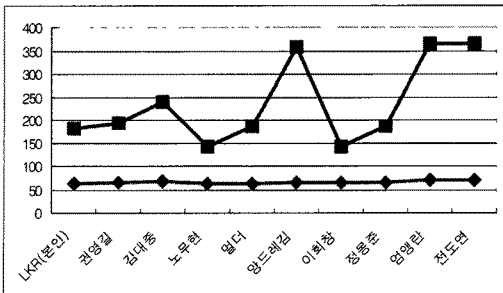


〈그림 1 화자 LKH의 기본주파수 대역폭 단어〉

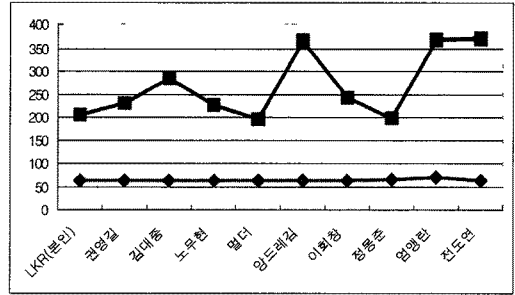


〈그림 2 화자 LKH의 기본주파수 대역폭 문장〉

위의 〈그림 1〉과 〈그림 2〉는 피실험자 LKH의 단어와 문장 발화 시 기본주파수 대역폭이며, 아래의 〈그림 3〉과 〈그림 4〉는 피실험자 LKR의 것이다 모두 본인 발화의 기본주파수 대역폭이 모방 시 변화한 것을 알 수 있다.



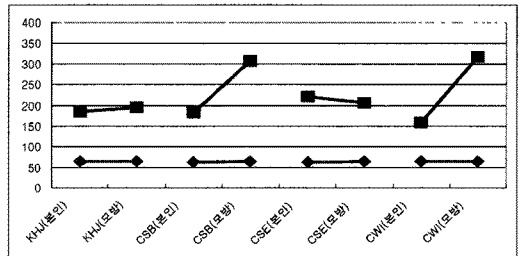
〈그림 3 화자 LKR의 기본주파수 대역폭 단어〉



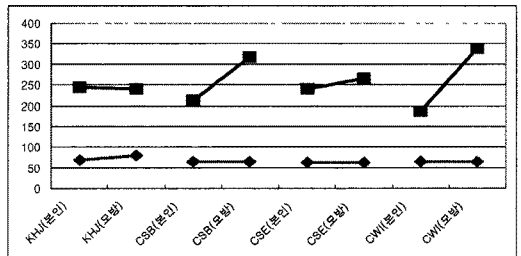
〈그림 4 화자 LKR의 기본주파수 대역폭 문장〉

이들 자료를 통하여 단어 수준에서 기본주파수의 대역폭이 확대되거나 축소된 경우를 모두 확인할 수 있다. 특히 LKH, LKR의 복수목표발화의 경우는 LKH '박준형' 모방과 LKR의 '노무현, 이회창' 모방에서 기본주파수 대역폭의 축소를 확인할 수 있다.

그러나, 문장 수준에서는 피실험자 모두 기본주파수 대역이 확대의 양상으로만 변화한 것을 알 수 있다 다음의 자료를 더 살펴보도록 하자



〈그림 5 KHJ, CSB, CSE, CWI 기본주파수 대역폭 단어〉



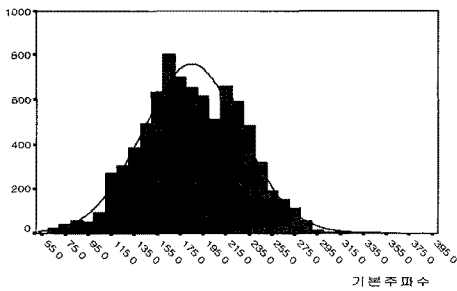
〈그림 6 KHJ, CSB, CSE, CWI 기본주파수 대역폭 문장〉

〈그림 5〉와 〈그림 6〉은 단일 목표인 '김대중'의 음성을 모방한 나머지 4명의 피실험자의 기본주파수 대역폭이다 앞서 살핀 2명의 피실험자의 경우와 같이 문장 발화에서는 기본주파수 대역폭이 모두 확대되었지만, 피실험자 CSE가 '김대중' 모방에서 단어 수준에서는 본인 발화보다도 모방 시 좁은 대역폭이 실현된 것을 확인할 수 있다.

이런 결과로 문장 발화 자체에서 비롯되는 역동적인 (Dynamic) 기본주파수값의 자료보다는 단어 수준에서의 음성 모방 자료를 통하여 기본주파수의 변화를 더 객관적으로 확인할 수 있다. 그러나 기본주파수의 최대, 최소값의 절대값만으로 모방 정보를 파악하는 데에는 미흡하므로 다음에서 기본주파수의 평균값을 살펴보았다.

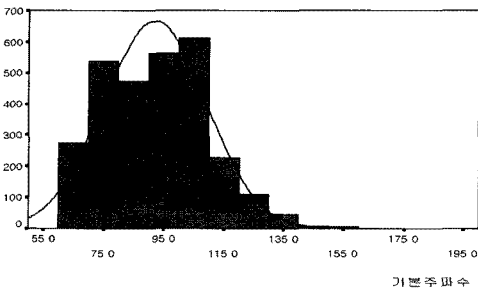
### 3.2 기본주파수 평균값의 양상

본 절에서는 전체 기본파수 값의 양상을 도출하여 비교하였다 아래의 그래프들은 각 화자별로 추출된 전체 기본주파수값들의 빈도를 히스토그램으로 나타낸 것이다. 다음의 <그림 7>은 전체 6명의 피실험자들이 대상으로 한 목표화자 '김대중'의 기본주파수값의 분포를 그래프로 나타낸 것이다.

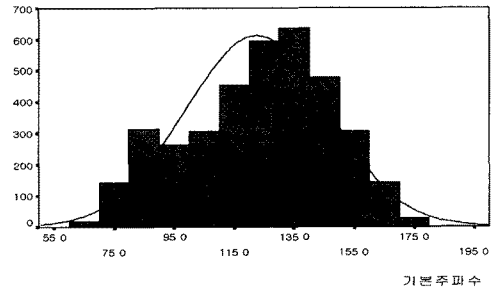


<그림 7 목표화자 김대중의 기본주파수 양상>

다음은 피실험자들이 '김대중'을 목표 발화로 하였을 때의 단어 목록 전체 자료의 기본주파수값 분포를 그래프로 나타내 보았다. <그림 8>은 피실험자 LJH의 본인 발화 기본주파수의 전체적인 실현 양상을 보인 것이고, <그림 9>는 '김대중'의 음성을 모방한 것이다.

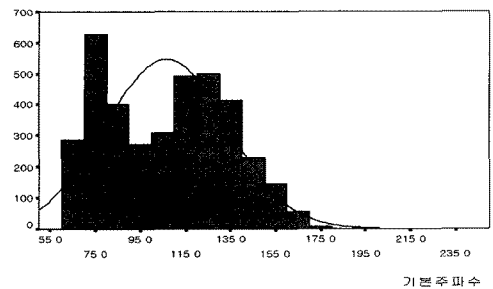


<그림 8 화자 LJH 본인 발화 기본주파수 양상>

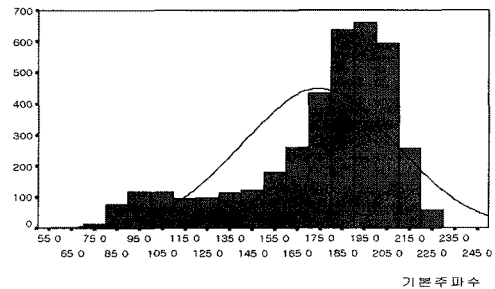


<그림 9 화자 LJH의 김대중 모방 시 기본주파수 양상>

<그림 10>과 <그림 11>은 피실험자 LKR의 본인 발화와 모방 발화 시의 기본주파수의 분포이다



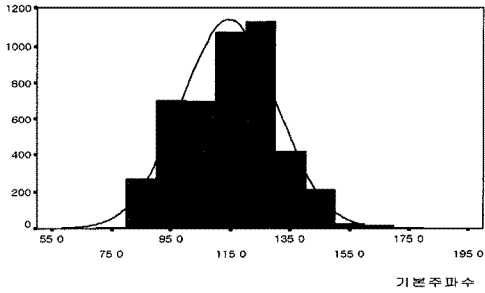
<그림 10 화자 LKR의 본인 발화 기본주파수 양상>



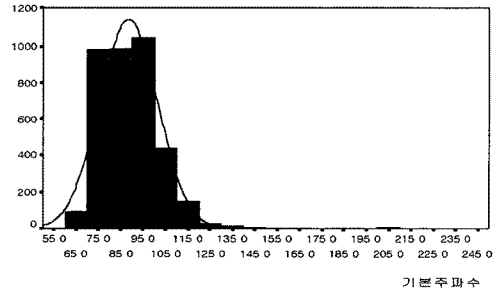
<그림 11 화자 LKR의 김대중 모방 시 기본주파수 양상>

목표 발화 '김대중'의 기본주파수 분포도는 평균 주파수대가 확연하게 중앙에 위치하는 전형적인 분포 곡선의 모습을 보이고 있다.

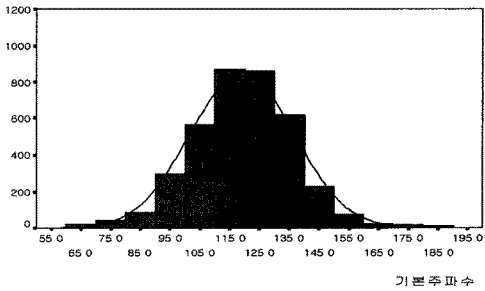
뒤 이어 나오는 화자 LJH와 LKR의 각 자료의 그래프에서 자세하게 볼 수 있듯이, 피실험자 본인 발화의 기본주파수값은 앞서 살펴본 기본주파수 대역 내에서 주로 저주파수대나 중간 수준에서 분포하고 있음을 알 수 있다 이와 반대로 모방 발화 시에는 전체적인 기본주파수값이 고주파수대로 편중되어 분포하고 있는 것을 확인할 수 있다. 아래의 <그림 12, 13, 14, 15, 16, 17, 18, 19>은 피실험자 KHJ, CSB, CSE, CWI의 기본주파수 분포도이다. 같은 경향성을 확인할 수 있다.



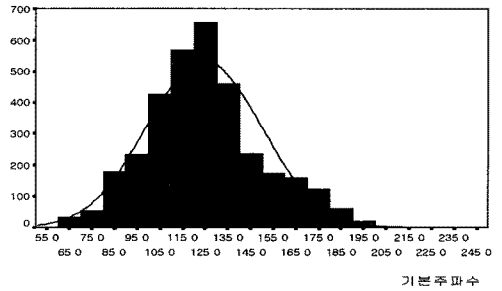
〈그림 12 화자 KHJ 본인 발화 기본주파수 양상〉



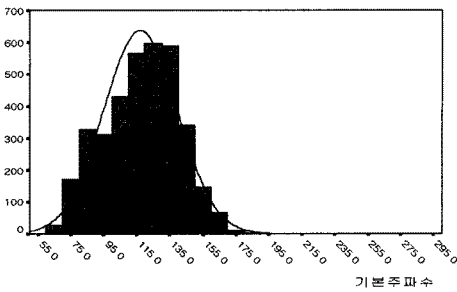
〈그림 16 화자 CSE 본인 발화 기본주파수 양상〉



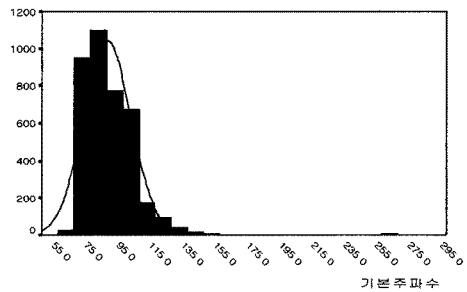
〈그림 13 화자 KHJ 김대중 모방 시 기본주파수 양상〉



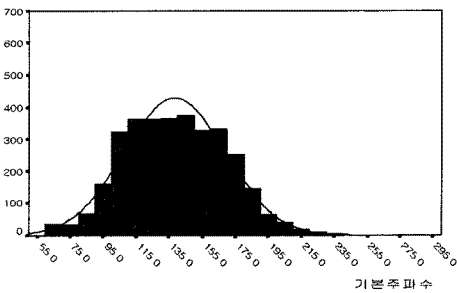
〈그림 17 화자 CSE 김대중 모방 시 기본주파수 양상〉



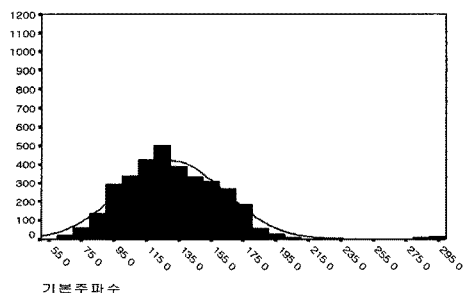
〈그림 14 화자 CSB 본인 발화 기본주파수 양상〉



〈그림 18 화자 CWI 본인 발화 기본주파수 양상〉



〈그림 15 화자 CSB 김대중 모방 시 기본주파수 양상〉

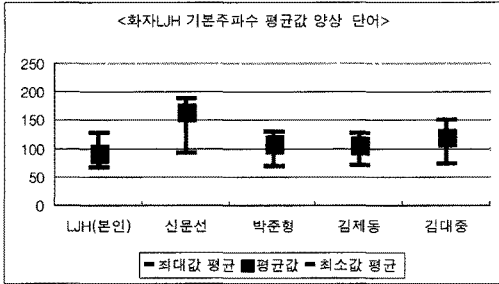


〈그림 19 화자 CWI 김대중 모방 시 기본주파수 양상〉

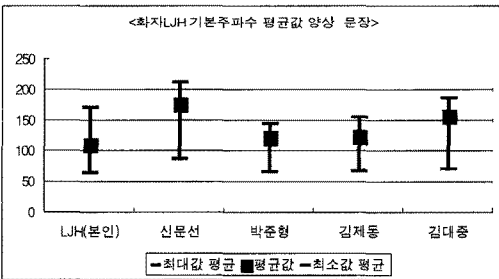
이상에서 살펴 본 것처럼 전체적인 기본주파수의 분포는 모방 발화 시에 본인의 발화에서보다 고주파수에 치우쳐 나타나고 있는 것을 알 수 있었다. 이러한 분석 과정을 거쳐 복수 목표 화자의 음성 모방과 전체

피실험자의 자료를 대상으로 기본주파수의 최대값, 최소값의 평균값과 전체 기본주파수값의 평균값을 추출하여 다음과 같은 그래프들로 나타내보았다

〈그림 20〉과 〈그림 21〉은 피실험자 LJH의 모방 자료인데, 단어 발화와 문장 발화의 분석 결과이다



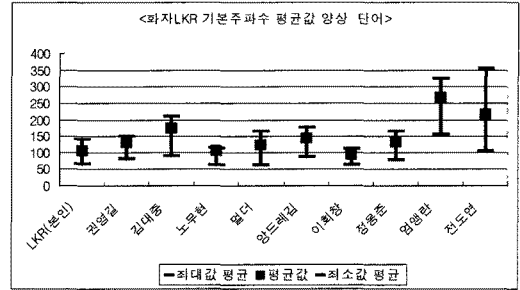
〈그림 20 화자 LJH의 기본주파수 평균값 양상 단어〉



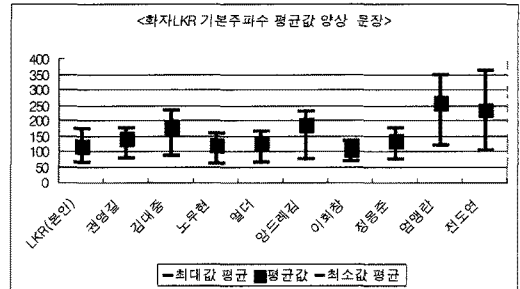
〈그림 21 화자 LJH의 기본주파수 평균값 양상 문장 수준〉

위의 분석 결과에서 피실험자는 본인 발화에서 전체 기본주파수의 평균값이 최고값과 최저값의 평균값의 중간 수준보다 약간 낮은 주파수대에 위치하고 있음을 알 수 있다 그러나 모방 발화 시에 전체 기본주파수의 평균값들은 최고값과 최저값의 중간 주파수대가 아니라 이보다 고주파수대로 이동되어 위치하고 있음을 확인할 수 있다.

반면에, 최저값은 모방 시에도 본인의 발화와 큰 차이가 없는데, 이것으로 우리는 음성 모방 시 자신의 최저기본주파수는 일정하게 유지되는 것을 알 수 있다

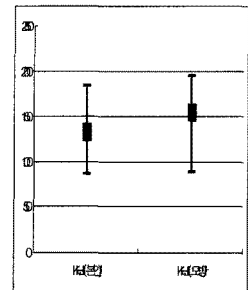
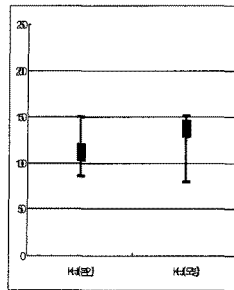


〈그림 22 화자 LKR의 기본주파수 평균값 양상 단어〉

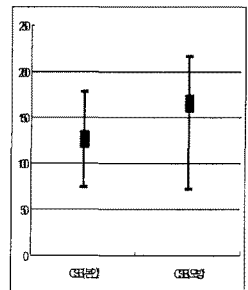
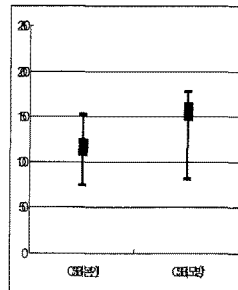


〈그림 23 화자 LKR의 기본주파수 평균값 양상 문장〉

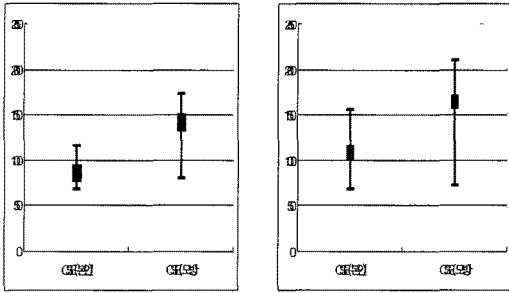
위의 〈그림 22〉와 〈그림 23〉은 피실험자 LKR의 것인데, 앞의 LJH에서와 마찬가지로 비슷한 경향성을 가지고 평균값들이 분포하고 있다. 아래의 자료들도 나머지 피실험자들의 기본주파수값 분포 양상이다 왼쪽은 단어 발화, 오른쪽은 문장 발화에서의 양상이다.



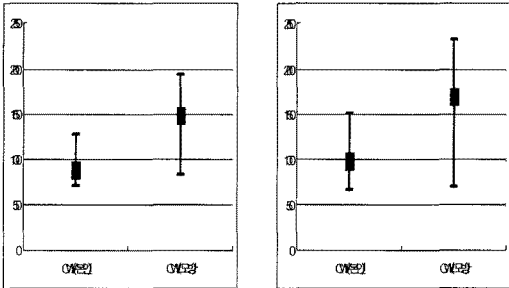
〈그림 24, 25 KHJ 김대중 모방 시 기본주파수 변화 양상〉



〈그림 26. 27 CSB 김대중 모방 시 기본주파수 변화 양상〉



〈그림 28. 29 CSE 김대중 모방 시 기본주파수 변화 양상〉



〈그림 30. 31 CWI 김대중 모방 시 기본주파수 변화 양상〉

지금까지의 결과에서 전체 기본주파수 평균값의 분포는 각 화자별 자료에서 최대값과는 10~60Hz, 최소값과는 50~110Hz의 편차를 보이는 곳에 위치한다는 것을 알 수 있었다. 결국 본인 발화 시에는 기본주파수의 평균값이 최저값과 최고값의 중간에 위치하는 반면에, 모방 발화 시에는 기본주파수의 평균값은 고주파수대에서 실현되고 있음을 알 수 있다.

#### IV. 결론

본 연구는 모방 발화 시 기본주파수 대역폭이 변화하였음을 확인하였고, 문장 수준에서는 전체적으로 기본주파수 대역이 확대되었지만, 단어 수준에서는 축소된 것을 관찰하였다.

또한 본인 발화 시 전체 기본주파수의 평균값은 모두 최대값과 최소값의 중간 수준이나 최소값에 가까운 영역에서 실현된 반면에 모방 발화 시 전체 기본주파수의 평균값은 최대값에 가까운 고주파수대에서 실현되어 아주 특징적인 모방 발화의 경향으로 나타났음을 알 수 있다.

전체적으로 기본주파수의 평균값은 최대값의 평균값과 단어 수준에서는 10~40Hz 정도, 문장 수준에서는 40Hz 이상의 편차가 나타났다 따라서, 문장 발화에서 보다는 단어 발화가 최대값과 평균값의 편차가 적은 고주파수대에서 실현하였음을 확인할 수 있었다.

본고의 이러한 결과들을 토대로 단어 수준의 발화에서 기본주파수의 최대값, 최저값, 평균값의 분포를 근거로 화자 식별이나 인식에 활용하는 방안을 기대할 수 있을 것이다.

#### 참고문헌

- [1] 박미영, 박지혜, 신지영, 강선미(2003) 모방발화의 음향음성학적 연구(2). 운율 특징을 중심으로 대한음성학회
- [2] 박지혜, 신지영, 강선미(2003) 모방발화에 사용되는 음성학적 특성에 관한 연구. 대한음성학회
- [3] 신지영(2000). 말소리의 이해. 한국문화사.
- [4] 안병섭, 이영배, 신지영, 강선미(2003). 모방발화의 음향음성학적 연구(1): 분절음의 특징과 발화속도 및 지속시간을 중심으로. 한국음성과학회