

모방발화의 모음 포먼트 연구

안병섭* 신지영** 강선미***

* 고려대학교 민족문화연구원 음성언어정보연구실

** 고려대학교 국어국문학과

***서경대학교 컴퓨터학과

A Study On Formants of Voice Imitation

Byoung-Seob Ahn*, Jiyoung Shin** and Sunmee Kang***

* Spoken Language Information Lab, Institute of Korean Culture, Korea Univ.

** Department of Korean Language and Literature, Korea Univ.

*** Department of Computer Science, Seokyeong Univ.

abs001@empal.com, shinjy@korea.ac.kr, smkang@skuniv.ac.kr

Abstract

The aim of this paper is to analyze vowel in voice imitation, and to find the invariable phonetic features of the speaker. In this paper we examined the formants of vowel /a, u, i/. The results of the present are as follows : ① Speakers change their vocal tract cavity features. ② F1 changes easily compared to F2·F3·F4. ③ F3-F2 appears to be constituent for a speakers identification in vowel /a/ and F4-F2 in vowel /i/.

I. 서론

모방발화(voice imitation)란 특정 대상인의 음성적 특징을 포착하여 그 사람의 음성과 유사하도록 음성을 흉내낸 것을 말한다. 모방발화에 대한 선행 연구에 의하면 어떤 화자가 특정 음성을 모방한다는 것은 특정인의 음성을 정확하게 따라하는 것이 아니라 특징적인 몇 가지 음성 특성을 정형화하는 과정이라고 하였다 [2][3][4][5]. 다시 말해서 모방자는 모방 대상자의 모든 음성적 요소를 모방하는 것이 아니며, 모방자가 인지한 모방 대상자의 음성적 특징을 정확히 모방하는 것도 아니라는 것이다.

따라서 본 연구는 자신의 목소리를 변화시키더라도

변화시키기 힘든 음성적 요소가 있을 것이라는 가정에서 출발하였으며, 변화시키지 않는 음성적 특성이 있다면 그것이 무엇인지를 찾아내는 데 목적을 두었다. 모방자가 인위적으로 변화시키지 못한다면 그것은 화자 특징적 요인이 될 수 있기 때문이다. 이와 더불어 화자 정보(speaker information)를 상대적으로 정확히 알아 낼 수 있는 분절음(특히 모음)은 무엇이며, 이를 화자 인식(speaker recognition)에 이용할 수 있는지에 대한 검토도 병행될 것이다.²⁾

모방발화에 대한 이러한 연구는 인간의 음성적 특징을 좀더 잘 이해하는 과정이 될 것이며, 한 걸음 더 나아가서는 음성 인식 및 화자 인식과 같은 음성공학 분야나 범음성학 분야의 발전에 기여할 수 있는 정보를 제공할 수 있을 것으로 기대된다

II. 연구 방법

본 연구에서는 본인발화와 모방발화를 비교하기 위해서 여섯 명의 남성 모방자를 섭외하여 녹음하였고, 한 명(남성)의 라디오 음성 자료를 녹취하였다. 여섯 명의 피험자 가운데 한 명의 자료는 강한 기식성으로 인해 모음의 포먼트가 잘 나타나지 않았으므로 측정 대상에서 제외하였다. 피험자 정보는 다음과 같다.

1) 본 논문은 한국과학재단 목적기초연구

(R01-1999-000-00229-0) 지원으로 수행되었습니다.

자료 성격	화자	나이	키(cm)	경력
설의 녹음	화자A	19	174	인터넷 동호회에서 활동
	화자B	25	177	개그맨
	화자C	26	176	일반인
	화자D	23	178	"
	화자E	24	176	"
라디오 자료	화자F	35	?	개그맨

피험자들은 두 가지 방법으로 모방발화를 녹음하였다. 하나는 특정인(김대중)의 음성을 듣고 일정 기간 연습한 후에 모방하는 것이며(화자A·B·C·D·E), 다른 하나는 주어진 대본을 피험자들이 자신 있어 하는 대상자의 음성으로 모방하는 것이다.(화자A·B) 피험자들이 모방한 대상자들은 다음과 같다.

- 화자A 김대중, 권영길, 노무현, 이회창, 정몽준, 멀더(극중 인물), 앙드레김, 전도연, 엄앵란
화자B: 김대중, 김제동, 박준형, 신문선
화자C·D·E: 김대중
화자F: 김대중, 김영삼, 노무현, 정준하

녹음 환경은 화자A와 화자B가 다른데, 화자A는 PC에서 GoldWave v.4.26 프로그램, SHURE 58SM 마이크를 사용하여 조용한 방에서 녹음하였고, 화자B·C·D·E는 TASCAM DA-20MKII DAT(일본), SHURE 58SM 마이크를 사용하여 고려대 음성언어정보연구실의 녹음실에서 녹음하였다. 모든 화자의 음성 자료는 Cool Edit Pro를 이용하여 16000Hz와 16bit로 디지털화하였으며, wavesurfer 15.2로 분석하였다.

녹음 문장은 연결문 문장 22개와 단어 30개이며, 피험자들은 녹음 문장을 각 발화마다 2회 반복하였다.

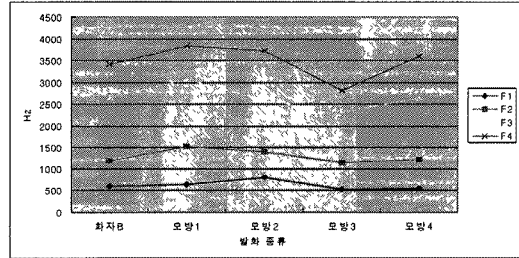
본 연구에서는 강세구 및 억양구 초에 오는 단음절 모음 /아, 우, 이/를 비교·분석의 대상으로 삼았다. 모음의 포만트 값들은 wavesurfer의 formant plot 기능을 이용하여 측정하였으며, 각 측정 지점의 FFT값과 대조하여 해당 모음의 가운데를 중심으로 전후 20ms(총 40ms)를 10ms 단위로 다섯 곳의 값들을 뽑아 평균을 내었다.

III. 결과 및 논의

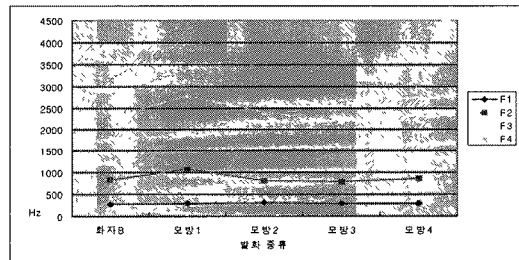
1. 모음의 포만트

모음 포만트를 측정된 결과, 피험자 모두의 경우에서 모음과 모방발화의 종류에 관계 없이 F1은 일정한 값을 보였다. 반면에 F2·F3·F4는 발화 종류에 따라

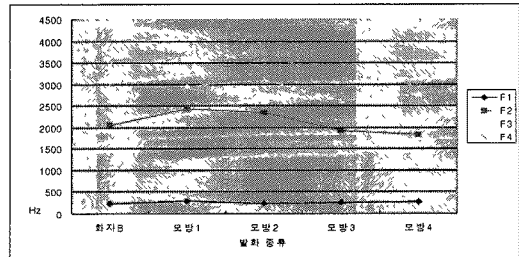
큰 변화를 보여주었다. 그 변화의 폭은 단어 발화보다 문장 발화에서 컸는데, 그것은 짧은 단어를 발화할 때보다 긴 문장을 발화할 때 안정된 조음 방법을 유지하기가 힘들었기 때문인 것으로 보인다.



[그림1] 화자B의 단어 발화 /아/의 포만트 평균값



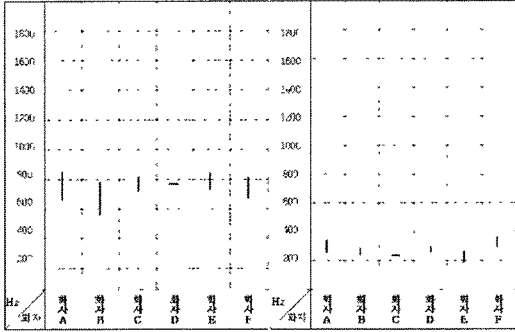
[그림2] 화자B의 단어 발화 /우/의 포만트 평균값



[그림3] 화자B의 단어 발화 /이/의 포만트 평균값

이러한 결과는 피험자들이 모방발화에 따라 성도 공명장의 모양을 다양하게 변화시키더라도 인두강의 길이는 대체로 변화시키지 못한다는 것을 의미한다. 이 사실로 보아 F1 값이 화자 특징으로 이용될 수 있음을 시사한다고 볼 수 있다.

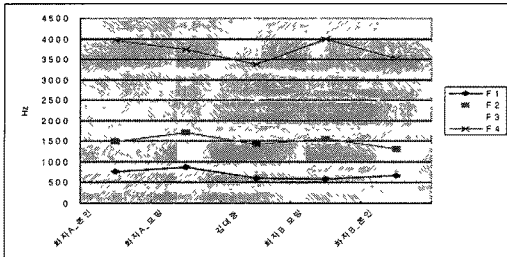
그러나 F1 값이 화자별로 모방 여부와 관계 없이 일정하여 화자 특징적 요인으로 보이기에는 하지만 다음 그림에서와 같이 화자 간에서 F1 값은 변별적이라고 하기 어려울 만큼 화자 간 유사한 범위를 보여주고 있다. 화자간 F1 값의 비교는 발화 종류를 구분하지 않고 화자별 모든 발화에서 나온 F1 평균값 가운데 최저값과 최고값의 범위를 보인 것이다.



[그림4] 화자 간 /아/(좌)와 /이/(우)의 F1값 비교

따라서 F1이 화자 특징적 요인이기는 하지만 화자 간 F1 값의 범위가 변별적이지 못하고 중복되어 나타났기 때문에 F1 값만으로는 화자 인식에 결정적 단서가 될 수 없다고 판단하였다.

다음은 특징인을 두 화자가 모방한 경우를 비교한 것이다.



[그림5] 특징인 모방발화의 포먼트 비교

그림을 보아서도 알 수 있듯이 모방자들은 특징인의 포먼트를 정확히 모방하지 못하였다. 오히려 포먼트 값이 모방발화보다 본인발화에서 모방 대상자와 더 비슷한 곳도 관찰되었다.

결국 모방자는 성도 공명강들의 모양을 변화시켰으나 그것은 정확한 모방이 아니었으며, 발화에 따른 포먼트 변화의 정도를 보았을 때도 제1포먼트를 제외한 각 포먼트를 화자의 특징적 요소로 보기가 어려울 수 있었다.

2. 포먼트 간의 차이값

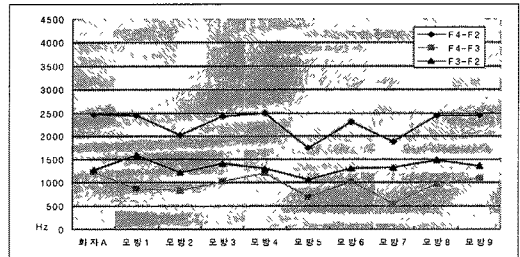
앞의 실험에서 우리는 화자가 발화에 따라서 모음의 포먼트를 자유롭게 변형시킬 수 있음을 알았다. 그러나 모방자들이 모음이나 발화 종류에 따라 성도의 다양한 공명강들을 변화시키더라도 공명강의 모양을 조절할 때는 상관되는 공명강들이 있을 것이라는 가정하에 포먼트들의 차(F4-F2, F4-F3, F3-F2)를 비교함으로써 그러한 특징을 찾고자 하였다. 앞의 실험을 통해

서 알 수 있듯이 F1 값은 일정하게 유지되는 반면 다른 포먼트 값들은 큰 변화를 보였으므로, F1 값과 다른 포먼트와의 관련성은 적을 것이라고 판단하여 F1과 다른 포먼트들과의 차는 비교에서 제외하였다.

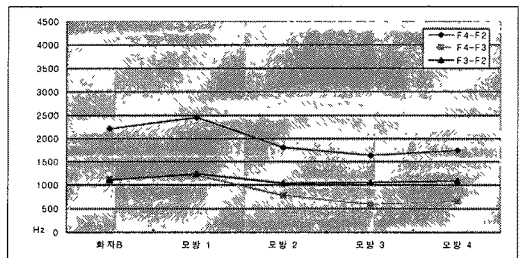
1) 모음 /아/

모음 /아/의 경우, 화자B·C·D·E·F의 경우에서 F3-F2의 값이 일정하게 나타났다. 그러나 화자A의 경우에는 다른 화자들에 비해 F3-F2 값이 상대적으로 일관적이지 않았다. 그렇다 하더라도 다른 값들보다는 일관성을 보이고 있음이 관찰되었다

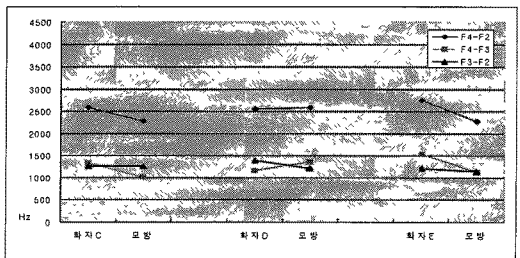
화자별로는 F3-F2와 더불어 F4-F2, F4-F3의 값들이 일정하게 나타나는 경우도 있었으나(화자D), F3-F2 값이 화자별로 큰 변화를 보이지 않은 반면에 F4-F2, F4-F3 값은 화자 간 변이가 커서 화자 특징적 요인으로 보기에는 어려웠다.



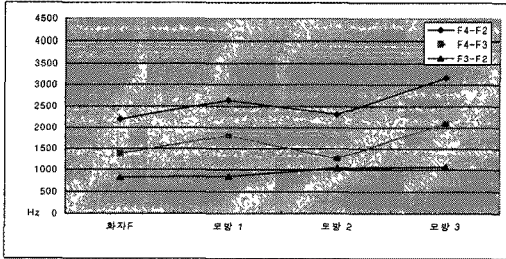
[그림6] 화자A의 포먼트 변화값(단어_아)



[그림7] 화자B의 포먼트 차이값(단어_아)



[그림8] 화자C·D·E의 포먼트 차이값(단어_아)

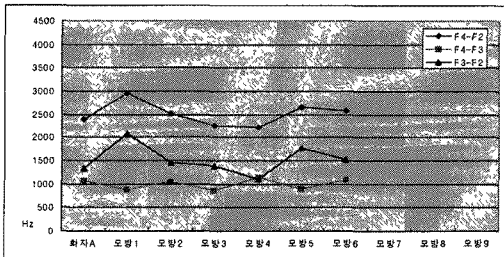


[그림9] 화자F의 포먼트 차이값(문장_아)

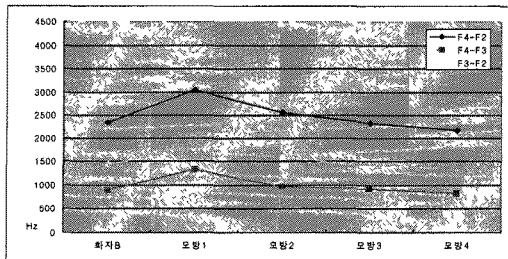
2) 모음 /우/³⁾

모음 /우/의 경우, 화자A에서는 세 변수 모두 일정한 값들을 보이지 않았다. 반면에 화자B의 경우에는 모음 /아/와 마찬가지로 F3-F2가 일정한 값을 유지하는 것으로 관찰되었다. 하지만 모음 /우/는 내재적 지속시간이 짧은 데다가 전이 구간이 다른 모음에 비해 상대적으로 일찍 시작된다는 점에서 측정에 어려움이 있다. 그러므로 모음 /우/를 화자 인식의 측정 대상으로 삼기에는 부적절할 것으로 판단된다.

다만 상대적으로 긴 안정구간을 보여주었던 화자B의 경우에서 F3-F2 값이 일정한 값을 보여주고 있어, 모음 /우/에서도 F3-F2 값이 화자 특징적 요소일 가능성을 배제할 수는 없을 것으로 보인다.



[그림10] 화자A의 포먼트 차이값(단어_우)

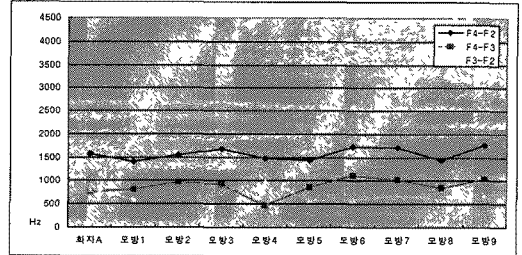


[그림11] 화자B의 포먼트 차이값(단어_우)

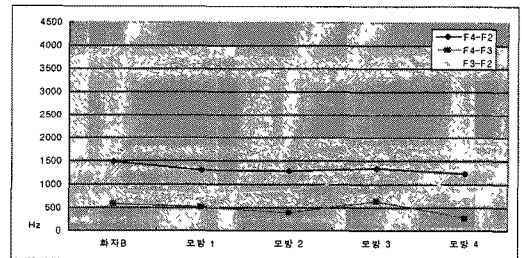
3) 화자C·D·E·F의 모음 /우/는 분석 대상에서 제외하였다 이 화자들은 사후 검증 과정에서 모음 /우/의 측정상의 문제점을 고려하여 제외하였다

3) 모음 /이/

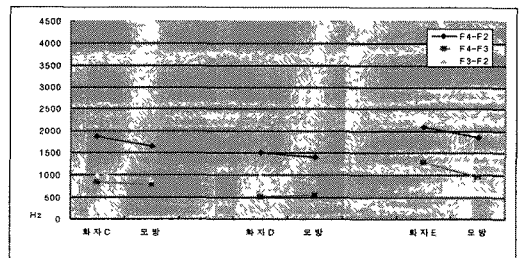
모음 /이/에서는 F4-F2 값이 일정하게 나타났다. 이로써 발화 시 공명강의 모양을 조절하는 과정에서 개별 모음마다 관련된 공명강의 종류가 다르다는 사실을 알 수 있었다.



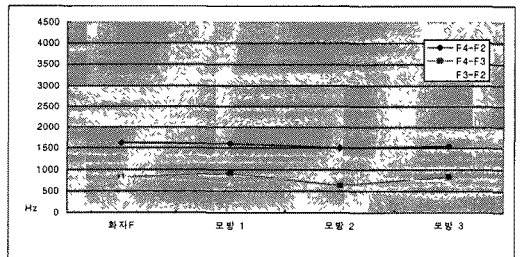
[그림12] 화자A의 포먼트 차이값(단어_이)



[그림13] 화자B의 포먼트 차이값(단어_이)



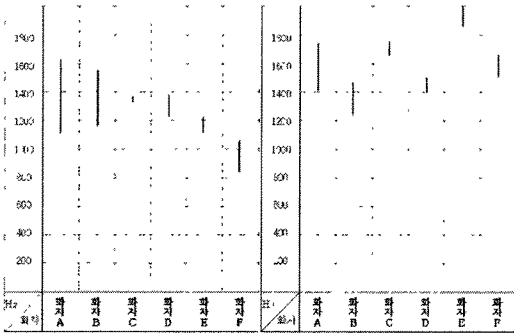
[그림14] 화자C·D·E의 포먼트 차이값(단어_이)



[그림15] 화자F의 포먼트 차이값(문장_이)

지금까지 살펴본 바에 의하면 모음 /아/는 F3-F2 값이, 모음 /이/는 F4-F2 값이 일정하게 나타나서 화자 특징적 변수임을 알 수 있었다. 이러한 결과가 화자간 F1 값을 비교했을 때보다 변별적인지를 다음 그림에서

비교해 보았다.



[그림16] 화자 간 /아/의 F3-F2(좌), /이/(우)의 F4-F2 값 비교

그 결과 두 변수를 비교했을 때가, 화자 간 F1 값을 비교했을 때보다 화자 간 변별성이 상대적으로 컸음을 알 수 있다. 따라서 화자 인식에서 포먼트 값을 모음 특징으로 이용할 것이 아니라 포먼트 간 차이 값을 보완적으로 이용한다면 화자 인식 성공률이 더 높아질 것으로 기대된다.

IV. 결론

본인발화와 모방발화를 대상으로 한 이상의 실험을 통하여 우리는 다음과 같은 사실을 알 수 있었다.

모음의 각 포먼트와 관련해서는,

- 1) 모음의 포먼트 가운데 F1은 모방 여부에 무관하게 일정한 값을 유지하였다 이로써 F1은 화자 특징적인 요소라 할 수 있다.
- 2) F2·F3·F4는 발화에 따라 많은 변화를 보였다. 따라서 이와 관련된 공명강의 모양을 모방자들이 쉽게 변화시켰음을 알 수 있었다

그러나 F1만으로는 화자를 식별하는 데 어려움이 있었다 그것은 F1이 화자 간 큰 차이를 보이지 않기 때문이다. 따라서 각 포먼트 간의 관계를 통해서 그러한 문제점을 보강하고자 하였다. 이와 관련해서는,

- 3) 발화 시 공명강의 모양을 조절하는 과정에서 개별 모음마다 관련된 공명강의 종류가 다르다는 사실을 알 수 있었다
- 4) 모음 /아/에서는 F3-F2 값이, 모음 /이/에서는 F4-F2 값이 화자별로 모방 여부와 무관하게 일관된 값을 나타내었다.
- 5) 4)의 두 변수는 화자 간 F1 값을 비교한 것보다 더욱 변별적이었다.

따라서 화자 인식에는 모음 /아/나 /이/를 사용하여 비밀 단어를 생성하는 것이 좋을 것으로 보이며, 이럴 경우 3), 4)의 결과들이 1), 2)의 문제점들을 보강할 수 있을 것으로 기대된다. 향후 연구에서는 포먼트 간 차이 값들이 과연 성공적인 화자 변별력을 가지는가에 대해서 다양한 방법의 검증이 이루어져야 할 것이다.

참고문헌

- [1] 신지영(2000). 말소리의 이해 한국문화사.
- [2] 안병섭·이영배·신지영·강선미(2003). 모방발화의 음향음성학적 연구(1), 분절음의 특징과 발화속도 및 지속시간을 중심으로. 한국음성과학회 발표논문집 73~80쪽.
- [3] Anders Eriksson and Par Wretling(1997) How Flexible is the Human Voice?-A Case study of Mimicry, In Proc. *Eurospeech* Vol 2, 1043~1046.
- [4] Elisabeth Wetterholm(2000) The Significance of Phonetics in Voice imitation, In Proc. *Australian International Conference on Speech Science and Technology*, Canberra, 342~347
- [5] Laver, J.(1994) *Principles of Phonetics*, Cambridge: Cambridge University Press.