피치 변경 발성에 따른 모음의 음향적 특성

조 창 수, 홍 광 석
성균관대학교 정보통신공학부 휴먼컴퓨터연구실
전화 : 031-290-7196 / 휴대폰 : 011-892-1725

Acoustic characteristics of Korean vowels on pitch alteration utterance

Chang-Su Cho, Kwang-Seok Hong
HCI Lab, School of Information and Communication Engineering, Sungkyunkwan University
wildcho@netian.com, kshong@yurim.skku.ac.kr

Abstract

In this paper, we examine the acoustic characteristics of Korean vowels on pitch alteration utterance. The prosody is known as an indicator of acoustic characteristics of emotions. Also, speech is acoustically differenced according to the emotional variation and environmental variation, although speaker utters the same speech.

We analyzed the spectral envelopes and formants from the voiced regions as data points on the speech waveform.

I. 서론

사람의 발성 음은 서로 다른 크기의 성대와 성도 모양 때문에 동일한 음이더라도 기본 주파수(pitch)나 포만트(formant)와 같은 음향적 분석 값이 달라진다[1]. 실제 생활에서 음성은 화자 내에서나 화자 사이에서 상당히 다양하게 나타난다.

예를 들어, 동일한 화자가 동일한 단어를 발성하더라도 결코 물리적으로 동일한 발음을 두 번 이상 하지 못한다(화자내 변이). 즉 동일한 화자가 동일한 음을 발성했다 하더라도 정서적인 변화나 주변 환경 변화에 따라서 물리 음향적으로 달라진다[2].

이런 화자 변이는 주로, 방언적, 사회 언어학적 차이와 같은 언어적 요소와, 신체구조, 나이, 성별, 발화자의 정서 상태 등과 같은 비언어적 요소로 구분된다. 기본 주파수는 화자의 성대 전체길이에 반비례하며, 나이와 성별에 따라 성대 길이가 달라서 기본 주파수 값도 달라진다. 또한 인강과 구강의 길이 비율도 화자간 변이의 요소가 된다.

이러한 발화의 특성은 음성 인식에 있어서 인식률을 감소시키는 요인이 되고 음성 합성에 있어서는 개인감정을 표현하는데 필요한 요소가 된다. 또한 화자간 차이점의 하나인 발화속도가 빨라질수록 모음의 지속시간은 짧아진다는 것에는 대부분의 연구들이 유사한 결과들을 보여줬다[3][4].

또한 최근 음성 인식 기술의 향상으로 인해 음성 인식 시스템의 실용화가 진행되면서 인식할 화자 및 그 주변 환경에 적응화 할 수 있는 음성 인식 시스템 개발이 진행되고 있는데 그 적응화 방법으로 음성 신호의 음향적 특징을 화자에 따라 적응화하는 방법과 기존의 학습된 파라메타를 새로 적응하고자 하는 환경 및 화자에 맞는 적절한 파라메타를 찾아 인식하는 방법 등이 있다.

본 논문은 기본 주파수의 높낮이에 따른 한국어 단모음들의 음향적 특성을 기본 주파수와 제 3 포만트와의 관계로 밝히고 그 관계를 어떻게 규정할 수 있는지 연구하였으며 본 연구 결과로 화자에 맞는 적절한 특징

값을 기본 주파수와 제 3 포만트와의 관계로 설정할 수 있다고 본다.

Ⅱ. 실험 환경

본 실험에 쓰인 한국어 모음은 단모음 7개(ㅏ, ㅓ, ㅗ, ㅜ, ㅡ, ㅣ, ㅐ)를 사용하였다. 우리말이 몇 개의 단모음으로 구성되어 있는가에 대해서 국어 학자들은 일치된 견해를 보이지 않고 있다. 학자에 따라서 많게는 10모음 체계에서 적게는 7모음 체계로 현대 국어의 모음 체계를 설정한다[5].

본 논문에서는 'ㅔ'와 'ㅐ'는 따로 구분하지 않고 'ㅐ'를 'ㅔ'와 'ㅐ'들의 대표음으로 간주하였다. 실제로 서울 방언 화자의 경우는 거의 성별이나 연령에 무관하게 'ㅔ'와 'ㅐ'철자를 의식하여 부자연스러운 발화를 하지 않은 이상 거의 하나의 음소로 통합되어 발성하기 때문에 음성적 차이가 없다[5].

이렇게 설정한 단모음 7개를 20대 중반의 남성 화자 5명과 여성 화자 5명으로부터 음의 높이가 다른 단모음 7개의 음성을 받아 기본 주파수와 포만트를 측정하여 분석하였다.

발성할 화자들에게 음높이에 대한 어떠한 기준을 제시해 그 높이에 맞는 정확한 음을 얻기에는 어려운 일이었기 때문에 녹음하기 전에 한 옥타브(octave)의 음을 발성하게 한 뒤 실제 녹음 시에도 연습 때 발성한 음 높이만큼 발성하게 하였다.

이렇게 해서 각 모음별, 음높이별로 3번 반복하여 저장한 다음 음 높이별로 하나씩 추출하였다. 녹음은 16KHz의 표본화 율로 표본화하고 16비트 양자화 하여 녹음하였다. 포만트 측정시 LPC를 사용하였고 LPC 차수는 12차로 설정하였다.

분석은 각 화자별로 발성한 모음들의 피치를 각각 구하였고 포만트는 제 1포만트부터 제 4포만트까지 측정하였다. 잘못 발성한 음성은 분석에서 제외했고 포만트 추출이 제대로 이루어지지 않은 음성도 제외했다. 또한 녹음시 음높이가 올라감에 따라 amplitude가 커져 amplitude의 저장범위를 초과하게 된 음성들도 분석에서 제외했다.

Ⅲ. 피치 변경 발성 모음의 특징

표 1과 2에서는 남성과 여성이 발성한 단모음 'ㅏ'에 대해 화자별 기본 주파수와 제 3포만트의 측정값을 나타낸 것으로써 남성 4의 경우 피치가 증가함에 따라 제 3포만트의 값도 일괄적으로 증가했다. 여러 화자들이 발성한 단모음의 포만트를 측정한 결과 대부분 많은 단모음들이 기본 주파수가 증가함에 따라 포만트의 위치가 증가했는데 그 중에서 제 3포만트의 위치가 다른 포만트보다 일괄적으로 증가했다(그림 1, 그림 2) [6].

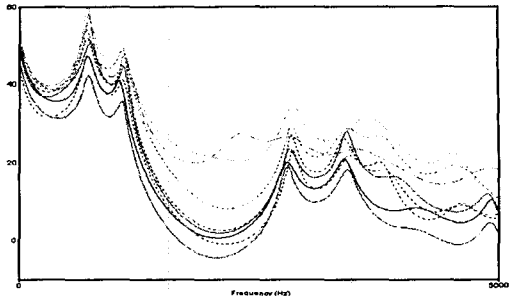

그림 1 남성 발성음 'ㅏ'의 LPC 12차 포락선 비교 (기본 주파수 값 : 89.3~187.6)

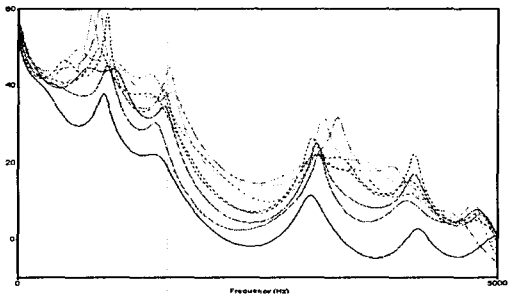

그림 2 여성 발성음 'ㅏ'의 LPC 12차 포락선 비교(기본 주파수 값 : 218.8~425.7)

표 1. 화자별 단모음 'ㅏ'의 기본 주파수와 3 formant 분석결과

남성1		남성2		남성3		남성4		남성5	
pitch	F3	pitch	F3	pitch	F3	pitch	F3	pitch	F3
113.8	2830	113.4	2950	89.31	2795	128.5	2545	95.08	2585
133.5	2875	135.4	3000	100.1	2795	143.9	2570	106.3	2595
151.6	2850	150.5	2930	113.3	2820	158	2600	118.	2585
163.1	2870	159.3	2880	123.7	2840	169.3	2655	132.9	2655
183.5	2865	182.1	2985	135.6	2850	192.8	2625	150.4	2650
211	2895	203.9	2985	155.4	2840	216.9	2705	162.2	2605
236.2	2855	230.2	3025	179.5	2855	248.3	2750	183.2	2585
251.3	2925	327.3	2930	187.6	2800	259.7	2815	192.7	2590

표 2. 화자별 단모음 'ㅏ'의 기본 주파수와 3 formant 분석결과

여성1		여성2		여성3		여성4		여성5	
pitch	F3	pitch	F3	pitch	F3	pitch	F3	pitch	F3
194.5	3030	204.8	208.9	208.9	3225	218.4	3280	218.8	3055
229	3025	224	214.4	214.4	3085	251.1	3270	243.5	3140
247.3	3180	232.1	237.3	237.3	3255	267.2	3250	263.8	3115
270.1	3195	257.1	260.8	260.8	3250	283.2	3305	281.1	3070
298.4	3255	285.1	276.8	276.8	3210	302.9	3310	310.9	3105
330	3280	302.9	302.4	302.9	3315	330.5	3330	350.3	3135
339.7	3360	344.5	329.1	329.1	3520	361.7	3280	396.8	3185
342.9	3385	378.9	361.8	361.8	3215	384.9	3465	425.7	3340

위의 그림 1과 2처럼 기본 주파수 변경에 따른 제 3 포만트의 위치가 다른 포만트보다 좀 더 일관되게 변화

하는 결과를 바탕으로 각 발성한 모음별 피치와 포만트를 측정한 뒤에 피치에 따른 r값을 구할 수 있는데 r은 다음과 같이 정의했다.

$$r = \frac{P}{F3} \times 100$$

(P = 기본 주파수, F3 = 제 3포만트)

이렇게 정의한 r을 각 화자별, 모음별, 피치별로 구하여 그래프로 그려보면 그림 3~9, 그림 10~16과 같이 나타난다. 그림에서 X축은 발성한 각각의 모음에 대해 음높이를 높여서 발성한 모음수로 1은 제일 낮은 피치 주파수를 갖는 음이고 7은 제일 높은 피치 주파수를 갖는 음이다. Y축은 r값을 나타낸다.

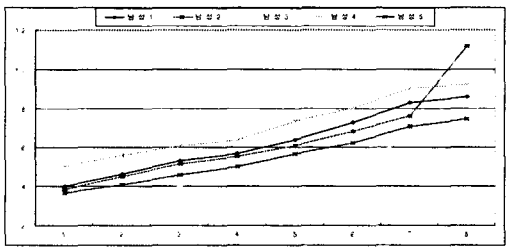

그림 3. 남성 화자에 따른 단모음 'ㅏ'의 피치별 r값

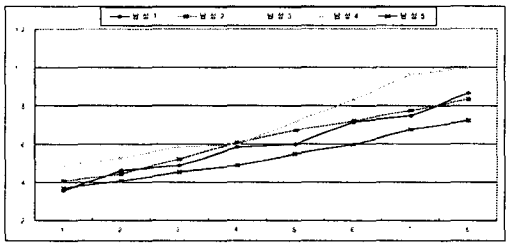

그림 4. 남성 화자에 따른 단모음 'ㅓ'의 피치별 r값

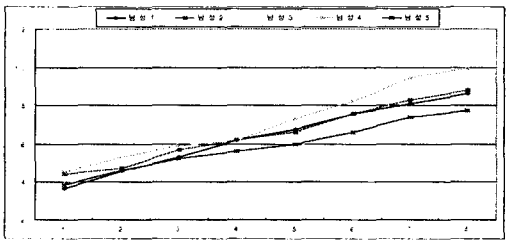

그림 5. 남성 화자에 따른 단모음 'ㅗ'의 피치별 r값

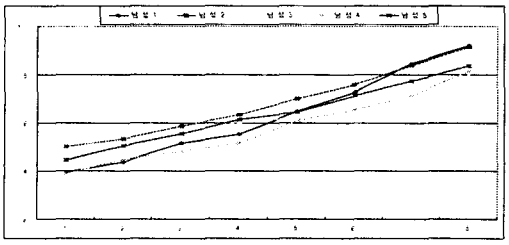

그림 6. 남성 화자에 따른 단모음 'ㅜ'의 피치별 r값

그림 3, 남성 1에서 첫 번째 음의 피치는 113.8Hz r값은 4.021201이며 두 번째 음의 피치는 133.5Hz r값은 4.643478이었다. 여기서 피치 1Hz에 대한 r값을 구할 수 있다. 이렇게 해서 구한 r값들을 더해 평균을 구했더니 1Hz에 대한 r값은 0.0345694가 나왔다. 이 값을 이용하여 남성 1이 발성한 'ㅏ'음에서 하나를 선택하여 피치 값과 1Hz에 대한 r값 변화율을 곱해주어 발성한 음의 r값을 구하였다. 이러한 방법을 적용해 각 피치별 발성한 음의 r값과 비교한 결과 화자별로 99%, 102%, 97%, 92%, 90%로 비슷하게 나왔다. 이러한 r값은 포만트의 위치를 포함하고 있다.

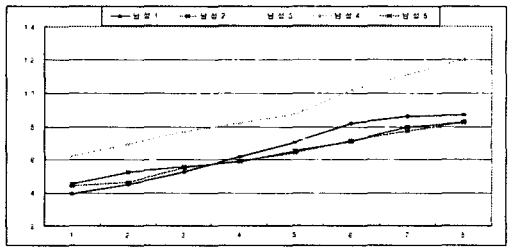

그림 7. 남성 화자에 따른 단모음 'ㅡ'의 피치별 r값

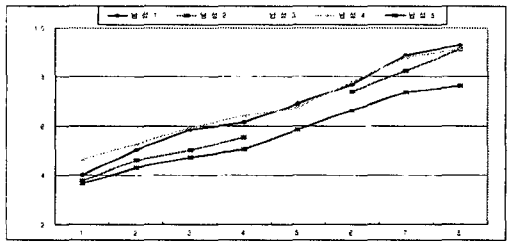

그림 8. 남성 화자에 따른 단모음 'ㅣ'의 피치별 r값

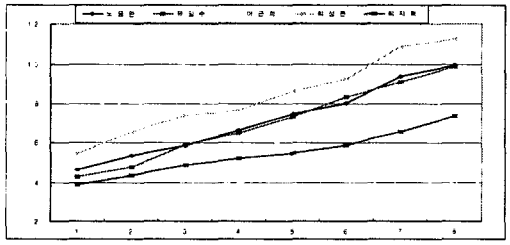

그림 9. 남성 화자에 따른 단모음 'ㅐ'의 피치별 r값

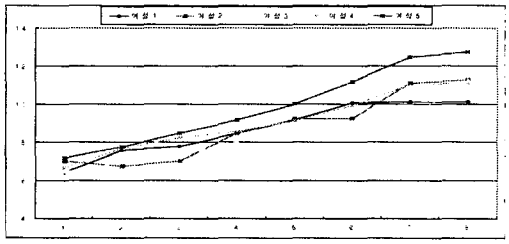

그림 10. 여성 화자에 따른 단모음 'ㅏ'의 피치별 r값

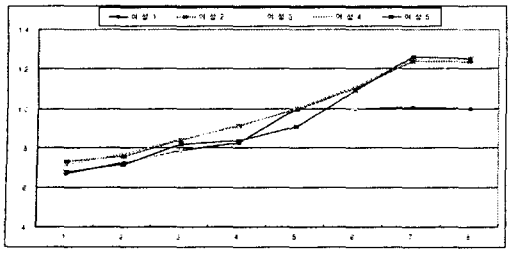

그림 11. 여성 화자에 따른 단모음 'ㅓ'의 피치별 r값

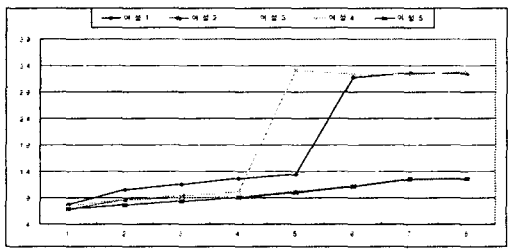

그림 12. 여성 화자에 따른 단모음 'ㅗ'의 피치별 r값

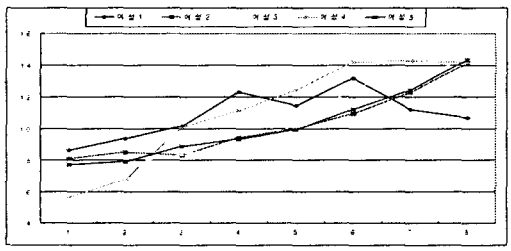

그림 13. 여성 화자에 따른 단모음 'ㅜ'의 피치별 r값

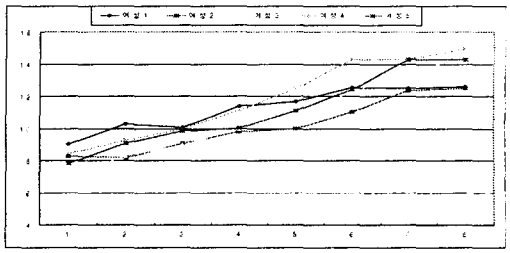

그림 14. 여성 화자에 따른 단모음 'ㅡ'의 피치별 r값

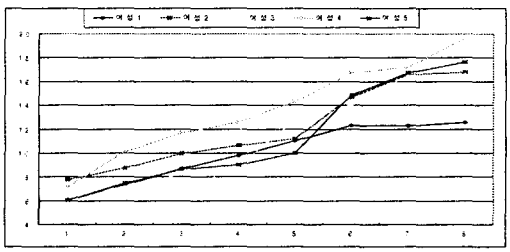

그림 15. 여성 화자에 따른 단모음 'ㅣ'의 피치별 r값

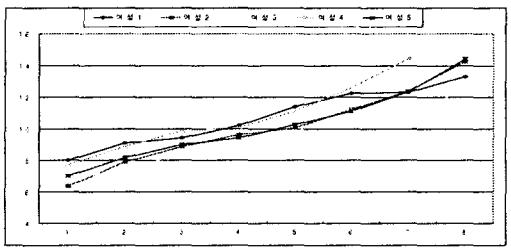

그림 16. 여성 화자에 따른 단모음 'ㅣ'의 피치별 r값

Ⅳ. 결론

본 연구에서는 기본 주파수인 피치를 변경하여 발성하였을 경우 한국어 단모음에 따른 기본 주파수와 제 3 포만트의 관계를 살펴보았다. 기본 주파수가 높아질수록 단모음의 제 3포만트도 높아져 갔다. 각 화자별 피치 측정값과 포만트 측정값을 분석한 결과 발성한 모음 중에 피치가 증가함에 따라 제 3포만트의 측정값이 일괄적으로 증가한 현상이 많이 나타났다. 이를 기반으로 여기서 정의한 r값을 구하였다. r값을 구하여 그래프로 그려본 결과 피치에 따른 r값이 비슷하게 나왔다. 본 연구 결과로 화자에 맞는 적절한 특징 값을 기본 주파수와 제 3 포만트와의 관계로 설정할 수 있다고 본다.

본연구는한국과학재단목적기초연구(R05-2002-000-01007-0)지원으로 수행되었음

참고문헌

[1] J. Junqua "Robust speech recognition in embedded system and PC applications" Kluwer Academic Publishers, 2000.

[2] J. Junqua, "The Lombard Reflex and Its Role on Human Listners and Automatics Speech recognizers," J.Acoustical Soc., 1993.

[3] T. Gay "Effect of speaking rate on vowel formant movement" J. Acoustical Soc., 1978

[4] J. L. Miller, "Effect of speaking rate on segmental distinction." Perspectives on the Study of Speech, Hillsdale, 1981.

[5] 신지영, "말소리의 이해", 한국문화사 2000

[6] K.S. Hong, "An Intra speaker Factor Estimation to Intra Speaker Normalization", ICCSA, 2003