

한국어 음성의 스펙트럼 변화에 관한 연구

A Study on the Spectrum Variation of Korean Speech

이 수 길*
Sou-Kil Lee

송 정 영**
Jeong-Young Song

요 약

음성학에서 음성이 가지고 있는 주파수 특성을 이용하여 스펙트럼을 추출할 수 있고 이를 이용하여 음성을 분석할 수 있다. 그러나 음성의 스펙트럼은 단모음의 경우 어느 정도 일정한 형태를 유지하지만 음절, 단어 등과 같이 자음과 모음이 서로 결합되었을 때는 상당한 변화가 발생된다. 이는 음소단위 음성인식에 있어서 가장 큰 장애가 되고 있다.

본 논문에서는 주파수 영역과 청각적 인상을 고려한 멜 대역 그리고 멜 첵스트럼을 이용하여 각 자음과 모음이 가지고 있는 스펙트럼을 분석하고, 청각적 특성을 반영한 음성의 변화를 체계화하여 음성을 음소단위로 분할할 수 있는 기반을 제공한다.

Abstract

We can extract spectrum of the voices and analyze those, after employing features of frequency that voices have. In the spectrum of the voices monophthongs are thought to be stable, but when a consonant(s) meet a vowel(s) in a syllable or a word, there is a lot of changes. This becomes the biggest obstacle to phoneme speech recognition.

In this study, using Mel Cepstrum and Mel Band that count Frequency Band and auditory information, we analyze the spectrums that each and every consonant and vowel has and the changes in the voices reflects auditory features and make it a system. Finally we are going to present the basis that can segment the voices by an unit of phoneme.

↳ Keyword : speech, voice, spectrum, segmentation

1. 서 론

음성은 푸리에 변환을 통하여 주파수 영역으로 변환되고 이는 스펙트럼으로 해석된다[1]. 그리고 그 스펙트럼을 분석하여 음성을 음소단위로 분리할 수 있다. 그런데 이러한 음향학적 분석에는 한계가 있다. 즉 음성의 변화에 적절히 대응하지 못하고 있기 때문이다. 음성의 스펙트럼은 단모음의 경우 어느 정도 일정한 형태를 유지하지만 자음과 모음이 결합하여 음절, 단어 등으로 발음될 때는 상당한 변화를 일으키기 때문이다[2,3].

일반적으로 음성인식을 위한 음소분리는 단어 또는 연속 음성인식을 위해 정확한 Segmentation이 필수적으로 알려져 있으며, 현재 그 연구는 여러 곳에서 활발히 진행 중으로 보고되고 있다[2,4-6]. 또한 유성음/무성음을 분리하는 기법과 파열음/마찰음을 식별하는 연구도 현재 진행 중인 것으로 알려져 있다. 그러나 기존의 논문에서 시간영역의 파형 에너지와 영교차율 등을 주로 사용하여 환경에 따른 변화에 대한 적용이 어려웠고 일부 주파수 영역의 스펙트럼을 연구한 논문에서도 구체적인 연구의 필요성을 제시 하였다[7].

따라서 본 연구에서는 주파수 영역과 청각적 인상을 고려한 멜 대역 그리고 멜 첵스트럼을 이용하여 음성의 스펙트럼을 분석하고, 멜 대역으로 기존의 모음 사각도를 재정립한다. 그리고 모음과 모음의 연결과, 자음과 모음의 연결시

* 정 회 원 : 오류동대장

sklee2@mail.pcu.ac.kr(제1저자)

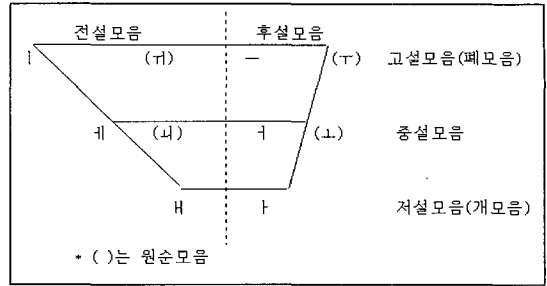
** 종신회원 : 배재대학교 컴퓨터공학과 부교수

jysong@mail.pcu.ac.kr

[2005/10/05 투고 - 2005/10/15 심사 - 2005/11/5 심사완료]

자음이 모음에 주는 영향 등에서 나타나는 음성의 변화를 체계화하여 음성을 음소단위로 분할 할 수 있는 기반을 제공하고자 한다.

본 논문은 2장에서 모음의 스펙트럼 변화를, 3장에서는 자음과 모음간의 스펙트럼 변화를, 4장에서 분석 및 평가를, 마지막으로 5장에서 결론을 맺는다.



〈그림 1〉 모음 사각도

2. 모음의 스펙트럼 변화

2.1 실험 조건

본 연구는 탁상용 마이크를 사용하여 음성을 16kHz, 16bit, Mono로 샘플링 하였으며 푸리에 변환을 통하여 주파수를 구하고[8]-[10] 청각인상을 고려하기 위하여 이를 멜 대역으로 전환시킨 후 멜 켈프스트럼을 구하였다. 앞으로 사용될 각 그림들은 상단부터 푸리에 변환을 통한 8 kHz범위의 주파수별 스펙트럼(F/kHz), 발음 시간(Second), 멜 대역(MF), 멜 켈프스트럼(MC) 그리고 하단에는 멜 대역의 주파수가 강한 부분(MFC)을 나타낸다.

2.2 모음의 체계

한국어의 모음은 단모음과 이중모음으로 분류되며 단모음의 모음체계는 그림 1과 같이 발음시 혀의 높낮이와 혀의 앞뒤 위치에 따른 모음사각도에 의해 표현된다. 또한 단모음은 소리를 길게 발음해도 입술이나 혀가 고정되어 발음 도중 처음부터 끝까지 움직이지 않는다[11]-[14][16].

따라서 단모음은 발음시 혀의 앞뒤 위치에 따라서 전설모음과 중설모음, 후설모음으로 분류되고 혀의 높이에 따라서 고설모음, 중설모음, 저설모음으로 또는 고모음, 반고모음, 반저모음, 저모음으로 분류되고 있다[11].

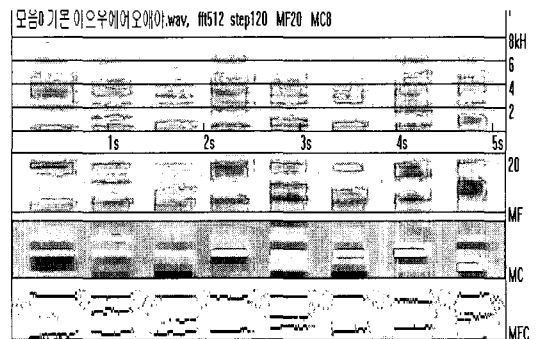
본 연구에서는 단모음과 이중모음 중 일부

내용을 주로 분석한다. 단모음 중 ‘ㅣ’ 와 ‘기’는 이중모음으로 취급될 수 있으므로 “ㅣ, ㅡ, ㅓ, 기, ㅕ, ㅗ, ㅛ, ㅜ”의 8개를 분석하고[12], 이중모음은 “유, 여, 요, 야”의 4개를 분석한다.

2.3 기본 단모음의 스펙트럼

청각인상을 고려한 멜 대역 위주로 분석하기 때문에 본 연구에서는 멜 대역(MF)의 가장 높은 스펙트럼을 H로, 높은 중간을 HL로 가장 낮은 스펙트럼을 L로, 중간을 M으로, 낮은 중간을 ML로 정의한다.

그림 2는 5초간 발음한 “이, 으, 우, 에, 어, 오, 애, 아”에 대한 단모음으로 스펙트럼과 멜 켈프스트럼이 발음 시작부터 끝까지 모두 안정되어 있음을 알 수 있다. F, MF, MC, MCF 모두 일관성이 있다. 특히 MF 대역에서의 H, MH, ML, L, ML 모두가 거의 변화가 없다.

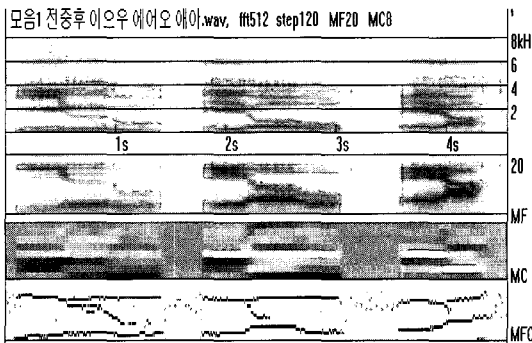


〈그림 2〉 기본 단모음 “이으우에어오애아”

2.4 혀의 앞뒤 위치에 따른 변화

그림 3은 발음시 혀의 앞(앞니)과 뒤(목구멍)의 위치에 따른 변화를 관찰하기 위하여 고모음 중 “이으우”, 중모음 중 “에어오”, 저모음 중 “애아”를 나타내고 있다.

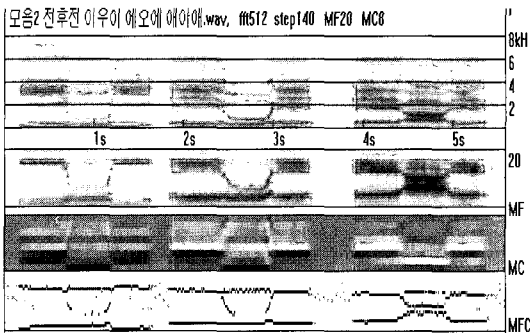
그림에서 혀의 위치가 뒤로 갈수록 H가 약화되고 M이 강화된다. 즉 MF의 스펙트럼이 H→MH→ML로 이동하는 것을 볼 수 있다.



〈그림 3〉 모음 전중후 “이으우 에어오 애아”

그림 4는 혀의 앞뒤 위치에 따른 스펙트럼의 변화를 자세히 관찰하기 위하여 고(이,으,우), 중(에,오,에), 저모음(애,이,에)에 따라 가장 앞의 모음(이,에,애)→가장 뒤의 모음(우,오,아)→다시 가장 앞의 모음(이,에,애)을 각각 비교하였다.

발음시 혀가 뒤쪽에 있을 때 H가 약화되고 M이 강화되는 것이 더욱 명확하다.

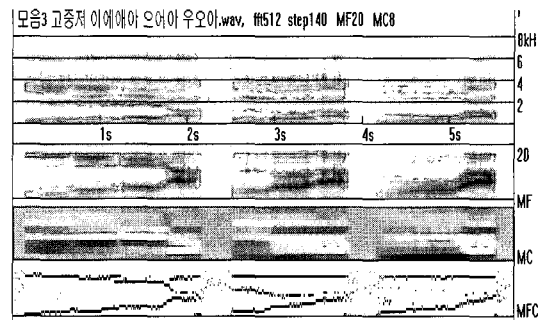


〈그림 4〉 모음 전후전 “이우이 에오에 애아애”

2.5 혀의 높이에 따른 변화

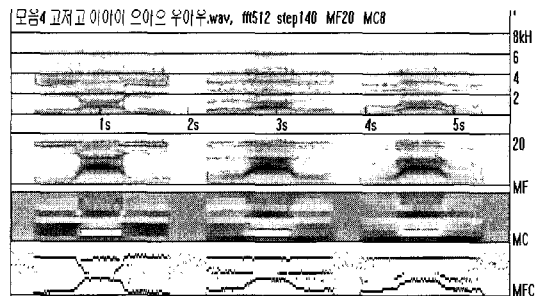
그림 5는 발음시 혀의 높이에 따른 변화를 관찰하기 위하여 혀의 높이가 높은 음과 낮은 음을 결합하여 “이, 에, 애, 아”, “으, 어, 아”, “우, 오, 아”의 그룹으로 분류하여 비교하였다.

그림 5에서 혀의 위치가 아래로 갈수록 L이 M으로 이동하는 것을 볼 수 있다.



〈그림 5〉 모음 고중저 “이에애아 으어아 우오아”

그림 6은 혀의 높이에 따른 스펙트럼의 변화를 분명하게 관찰하기 위하여 가장 높은 모음(이,으,우)→가장 낮은 모음(아)→다시 가장 높은 모음(이,으,우)을 비교하였다. 혀의 위치가 낮을 때 L이 약화되고 M이 새롭게 형성되는 것이 더욱 명확하다.



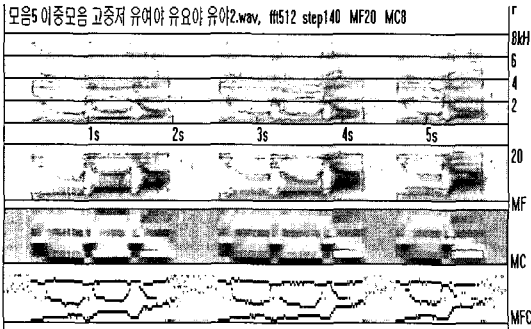
〈그림 6〉 모음 고저고 “이아이 으어으 우아우”

2.6 이중모음

이중모음은 소리를 내는 도중에 입술 모양이

나 혀의 위치가 처음과 나중에 달라지는 모음이다. 그림 7은 이중모음을 혀의 높이가 높은음(고)과 중간음(중), 낮은음(저)을 결합하여 발음한 ‘유여야’, ‘유요야’, ‘유야’ 이다.

그림 7에서 스펙트럼이 ‘유’는 “이→우”, ‘여’는 “이→어”, ‘야’는 “이→아”, ‘요’는 “이→오”의 방향으로 움직이고 있는 것을 볼 수 있다. 즉 그림의 이중모음은 ‘이’ 모음에서 “우, 어, 아”의 방향으로 스펙트럼이 이동하고 있다.



〈그림 7〉 모음 이중모음 고중저 “유여야 유요야 유아”

3. 자음과 모음 간의 스펙트럼 변화

3.1 실험 조건

실험은 모음과 동일한 조건에서 자음이 모음에 미치는 M의 변화를 살펴보기 위해 스펙트럼이 M부근에 집중되어 있는 ‘아’와 결합하여 발음하고. HL의 변화를 살펴보기 위해서 M이 거의 없는 ‘이’모음과 결합하여 실험하였다.

3.1 자음의 실험 체계

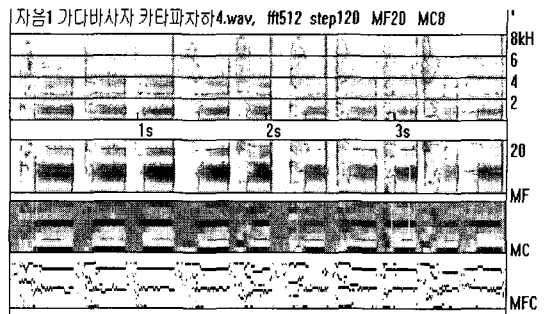
자음은 훈민정음 초성의 체계인 전청(ㄱㄷㅂㅅㅈ), 차청(ㅋㅌㅍㅊㅎ), 전탁(ㄱㄷㅂㅅㅈㅊ), 불청불탁(ㅇㄴㄹㄴ)의 순서로 하였으며[15] 이는 현대 음성학의 평음, 유기음, 경음, 공명음과 같다. 이어서 모음사이의 울림소리와 중성을 분석한다.

3.2 자음과 모음(‘ㅏ, ㅣ’)을 결합한 실험

그림 8은 자음 중 전청(ㄱㄷㅂㅅㅈ)과 차청(ㅋㅌㅍㅊㅎ)을 ‘ㅏ’ 모음과 결합한 발음이다.

주파수 대역(F)에서는 고주파의 확인이 용이하고, 멜 대역(MF)에서는 고주파를 약화시키고 저주파를 강화시켰기 때문에 동시에 비교하면 고주파가 많고 저주파가 적은 자음 분석이 용이하다.

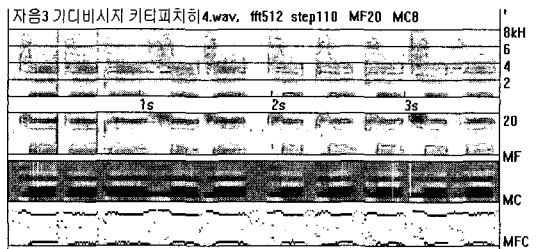
그림 8에서 자음은 고주파를 형성하며 멜 대역의 저주파 스펙트럼이 아주 적다. 단, 표준 숨결 잡음 형성으로 저주파가 발생했다. ‘ㅈ’과 ‘ㅊ’은 고주파에서 MH로 스펙트럼이 변화하고 있다. 즉 모음의 MH가 위로 향한다.



〈그림 8〉 자음 “가다바사자 카타파차하”

그림 9는 자음 중 전청(ㄱㄷㅂㅅㅈ)과 차청(ㅋㅌㅍㅊㅎ)을 ‘ㅣ’ 모음과 결합한 발음이다.

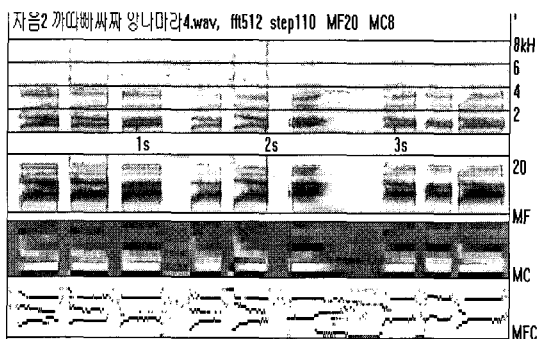
그림 9에서 전청과 차청은 H와 L에는 거의 영향을 주지 않으며 고주파를 형성하고 있는 것을 볼 수 있다.



〈그림 9〉 자음 “기디비시지 키티피치히”

그림 10은 전탁(ㄱㅈㅊㅅㅆ)과 불청불탁(ㅇㄹㄴ)을 ‘ㅓ’ 모음과 결합한 발음이다

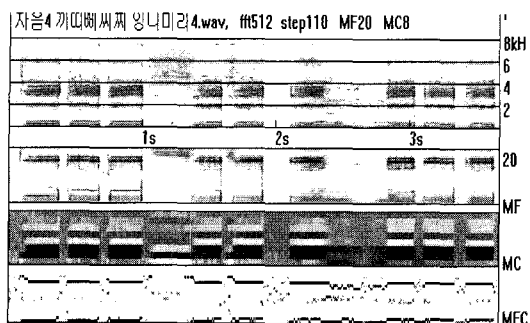
그림 10에서 전탁과 불청불탁의 자음은 M을 상(ㅈㄴㄹ)과 하(ㄱㅈㅊㅅㅆㅈㄴㄹ)로 변화시켰으며 저주파가 거의 없다. 즉 전탁은 ML을 아래로 변화시키며 고주파를 형성하고 있다. 불청불탁 중 ‘ㅇ’은 L부근에 약한 저주파를 새로 만들고 기존 모음의 여음을 형성하며, ‘ㄴ’은 ML을 아래로, ‘ㄹ’은 MH와 ML을 아래로 ‘ㄹ’은 MH와 ML을 위아래로 변화시키고 있다.



〈그림 10〉 자음 “까따빠싸짜 양나마라”

그림 11은 전탁(ㄱㅈㅊㅅㅆ)과 불청불탁(ㅇㄹㄴ)을 ‘ㅣ’ 모음과 결합한 발음이다

그림 11에서 전탁과 불청불탁은 H, L에는 큰 영향을 주지 않으며 전탁은 고주파를 형성하고 불청불탁(ㄴㅇㄹ)은 약한 저주파와 L을 만들어 낸다.



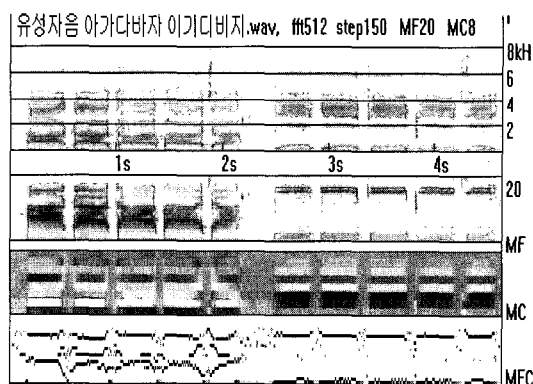
〈그림 11〉 자음 “끼띠빠씨찌 양니미리”

3.3 모음 사이의 유성자음

그림 12는 연속된 발음으로 울림소리 사이에서 울림소리로 발음되는 자음(ㄱ,ㄷ,ㅂ,ㅈ)이 모음에 미치는 영향을 나타내고 있다.

그림 12의 “아가다바”에서 “ㄱ,ㄷ”은 앞과 뒤의 모음의 ML을 아래로, ‘ㅂ’은 뒷 모음을 아래로, ‘ㅈ’은 MH와 ML을 각각 위와 아래로 변화시킨다.

결국 “ㄱ,ㄷ,ㅂ,ㅈ”은 앞과 뒤의 모음을 “ㄱ,ㄷ,ㅂ,ㅈ”와 같이 변화시킨다. 그리고 유성음화된 이러한 자음들은 H와 L에는 거의 영향을 주지 않음을 그림 12의 뒷부분인 “이기디비지”에서 확인되고 있다.

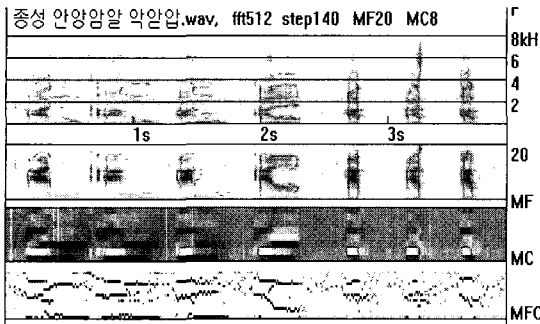


〈그림 12〉 모음 사이의 유성자음 “아가다바자 이기디비지”

3.4 종성

현재 사용되는 종성은 7개(ㄱ,ㅇ,ㄷ,ㄴ,ㅂ,ㄹ,ㄴ)가 있다. 그림 13은 ‘ㅓ’와 결합된 종성의 스펙트럼 변화를 나타내고 있다.

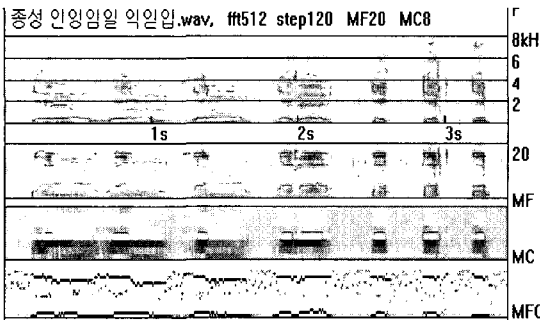
그림 13에서 공명음인 “ㄴ,ㅇ,ㄹ”은 H, M, L에서 약화된 여음을 형성하고 ‘ㄹ’은 H를 유지하고 MH와 ML을 상하로 벌려서 마치 모음 ‘ㅣ’와 같은 강한 음을 형성하고 있다. 그러나 “ㄱ,ㄷ,ㅂ”은 앞의 모음을 짧게 발음시키면서 ML을 L방향(하단)으로 내리고 있다.



〈그림 13〉 종성 “안양암알 악알입”

그림 14는 ‘ㅣ’와 결합된 종성의 스펙트럼 변화를 나타내고 있다.

그림 14에서 공명음인 “ㄴ, ㄹ, ㅁ”은 약화된 모음의 여운을 형성하지만 ‘ㄹ’은 H와 L부근에서 ‘ㄱ’과 같은 모음을 형성하고 있다. “ㄱ, ㄷ, ㅂ”은 앞의 모음을 급격히 단절시키며 아주 약한 여운을 형성한다.



〈그림 14〉 종성 “인양임일 익일입”

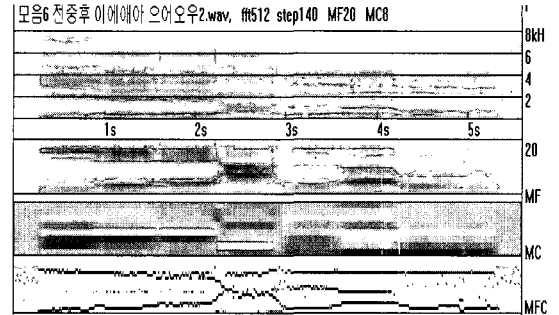
4. 분석 및 평가

4.1 모음

발음시 혀의 앞뒤 위치에 따른 그림 3과 그림 4는 H의 변화를 잘 나타내고 있다. 이를 H의 이동 정도에 따라서 다시 정리해 보면 그림 15와 같다.

그림 15에서 H에서 M으로 스펙트럼의 이동 정도에 따라 “이에애 으어오우”의 순서로 정리

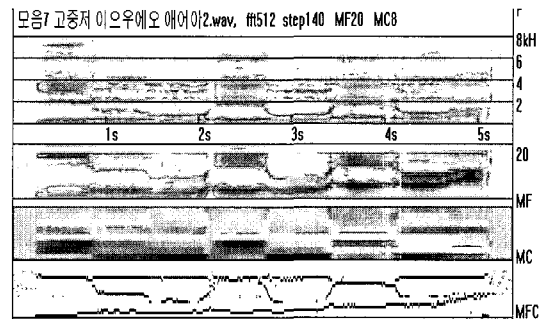
될 수 있다. 다만 ऐ애, 으아, 우오는 매우 비슷하여 같은 위치로 분류한다.



〈그림 15〉 모음 전중후 “이에애 으어오우2”

발음시 혀의 높이에 따른 그림 5, 6은 L의 변화를 나타내고 있다. 이를 L의 이동 정도에 따라 다시 정리해 보면 그림과 같다.

그림 16에서 L에서 M으로의 스펙트럼 이동 정도에 따라 “이으우에오 애어아”의 순서가 됨을 알 수 있다. 여기서 “으우, ऐ어, 애어”는 상당히 비슷하여 같은 위치로 분류한다.



〈그림 16〉 모음 고중저 “이으우에오 애어아2”

지금까지의 결과를 가지고 청각인상에 따라서 모음 음성도를 표 1과 같이 다시 구성할 수 있다.

표 1은 H에서 M으로 스펙트럼의 이동 정도에 따라 좌에서 우로 “이, ऐ애, 으아, 어, 우오”의 순서로 정리되었고 “ऐ애, 으아, 우오”는 같은 위치로 분류하였다. 또한 L에서 M으로

<표 1> 청각인상을 고려한 모음 위치

구분	전설← 혀의 앞뒤 위치 → 후설			
고 ↑ 혀 의 높이 ↓ 저				

스펙트럼 이동 정도에 따라 “이, 으우, 에오, 애어, 아”의 순서로 정리되었고 “으우, 에어, 애어”는 같은 위치로 분류하였다.

4.2 자음

자음은 ‘ㄱ, ㄴ’ 모음과 결합된 실험에서 H와 L에는 큰 영향을 주지 않으나 M을 변화시키고 고주파를 형성하며 불청불탁은 H, M, L의 여음을 형성한다.

또한 유성음화 된 자음은 전탁과 같이 앞뒵 모음에 영향을 주고 종성은 앞의 모음을 급격히 단절시키며 특히 ‘ㄹ’은 ‘ㄱ’모음에 가까운 강한 여음을 형성한다.

5. 결 론

본 연구에서는 음소 단위의 음성인식을 위해 발음시 음성의 변화를 보기 위해 모음과 음절, 단어 등과 같이 자음과 모음이 결합되어 발음될 때의 스펙트럼에 관하여 주파수 영역과 청각적 인상을 고려한 멜 대역, 그리고 멜 첩스트럼의 변화를 분석하였다.

본 논문에서 멜 대역을 통하여 청각적 특성을 고려한 단모음의 위치를 혀의 높이와 앞뒤

위치에 따라 새로운 청각음성의 모음 음성도를 각각 5단계로 재정립하였다.

그리고 자음을 M의 변화를 살필 수 있는 모음 ‘ㅏ’와 결합하고 H와 L의 변화를 살필 수 있는 모음 ‘ㅣ’를 결합한 실험에서 자음은 H와 L에는 큰 영향을 주지 않으나 M을 각각 변화시키고 고주파를 형성하며 불청불탁은 H, M, L의 여음을 형성하는 것과 종성의 변화를 확인하였다.

향후 연구로는 다양한 발음에 따라 약화되거나 생략되는 음성의 스펙트럼에 관한 연구가 추가된다면 더욱 정확한 음소 분리가 가능하리라 라고 생각된다.

참 고 문 헌

- [1] L. R. Rabiner, “Fundamentals of Speech Recognition”, Prentice-Hall, 1993.
- [2] Lee Sou-Kil, Song Jeong-Young, “A Study on the Phonemic Analysis for Korean Speech Segmentation p.134 - p.139”, The Journal of the Acoustical Society of Korea, 2004.
- [3] L. R. Rabiner / R.W. Schafer, “Digital Processing of Speech Signal”, Prentice-Hall, 1993.
- [4] 이시우, 연속음성에서 모음, 무성자음, 천이구간의 신호처리에 관한 연구, 상명대학교, 2000.
- [5] 성철재, “한국어 단모음 8개에 대한 음향 분석”, 한국음향학회지 제23권 제6호 pp. 454-461, 2004.
- [6] 박정일·하동경·신옥근, “잡음환경에서 우리말 연속음성의 무성자음 구간 추출 방법”, 2003.
- [7] 이수길·송정영, “고립단어에 대한 음소분리 시스템”, 한국인터넷 정보학회 춘계학술 발표대회 논문집 제4권 1호 pp. 320-323, 2003.

- [8] Peter B. Denes, Elliot N. Pinson, "THE SPEECH CHAIN", 한신문화사, 1999.
- [9] 오영환, "음성정보처리", 홍릉과학 출판사, 1998.
- [10] James H. McCellan · Ronald W. Schafer · Mark A. Yoder, "DSP FIRST", 1999.
- [11] 배주채, "국어음운론 개설", 신구문화사, 1997.
- [12] 이승재, 모음의 발음, 특집 새국어생활 제3권3호, 1993.
- [13] John Clark · Colin Yallop, "음성학과 음운론, 한신문화사", 1998.
- [14] 양병곤, 프라트(Praat)를 이용한 음성분석의 이론과 실제, 민수출판사, 2003.
- [15] 유창균, "훈민정음역주", 형설출판사, 1998.
- [16] 이수정, 한국어의 자음과 모음 음성학적 고찰, 성균관 대학교, 2004.

● 저자 소개 ●



이 수 길 (Sou-Kil Lee)

1985년 한국방송대학교 행정학과 졸업(학사)
 2000년 한밭대학교 전자계산학과 졸업(석사)
 2002~현재 배재대학교 컴퓨터공학과 박사과정
 1996~현재 오류동대장
 관심분야 : 음성인식, 패턴인식
 E-mail : sklee2@mail.pcu.ac.kr



송 정 영 (Jeong-Young Song)

1984년 한남대학교 전자계산학과 졸업(이학사)
 1992년 일본 와세다대학 전기전자정보공학연구과 졸업(공학석사)
 1995년 와세다대학 전기전자정보공학연구과 졸업(공학박사)
 1995~1997 청운대학교 전자계산학과 전임강사
 1997~현재 배재대학교 컴퓨터공학과 부교수
 2004 미국 University of Florida 교환교수
 관심분야 : 패턴인식, 음성처리, Human Interface Technology
 E-mail : jysong@mail.pcu.ac.kr