

포르만트를 이용한 한국어 단모음 분류 시스템의 개발에 관한 연구

최 윤 석                      김 기 식                      황 희 용  
서 울 대 학 교

A Study on The Development of the Korean Vowel Discrimination System Using Foxmant

(요약)

한국어 음소 인식 시스템의 구현을 위한 한 단계로서 한국어 모음의 분류를 위한 시스템을 설계하였다. 특징 파라미터로는 에너지, 영고차율 그리고 제 1, 제 2, 제 3 포르만트를 사용하였다. [아, 어, 오, 우, 으, 이, 예, 애]의 8개의 단모음에 해당하는 총 132개의 음성 데이터에 대하여 실험하였다. 각각의 모음에 대한 포르만트의 평균치를 구한 뒤 이들의 값을 표준 편차로도 갖는 인식 시스템을 구현한다. 구현된 인식 시스템에 의하여 새로운 음성의 모음부가 어떤 모음인지도 인식하며 결정할 때의 거리 측정 방식으로는 선형 분류 함수를 사용하였다. 132개의 모음에 대하여 62.9%의 인식율을 나타내었다.

1. 서론 : 모음 분류의 필요성

모음은 자음보다는 정보 전달량은 작다. 그러나 음성 인식 시스템의 구현에 있어 자음부의 분류보다는 모음부의 분류가 더 쉽고 안정적이다. 따라서 자음부의 분류 이전에 먼저 모음부를 찾아내어 안정정점을 찾은뒤 이 정보를 중심으로 자음부의 분류가 이루어지는 것이 일반적인 음성 인식의 절차이다. 따라서 한국어 음소 인식 시스템의 구현을 위해서도 음성에서의 모음부를 찾아내기 위한 알고리즘의 개발이 필요하며 모음부의 분류를 위한 실험 및 이론을 위한

정확한 지식을 찾아내는 연구가 필요하다. 또한 한국어는 영어나 일어에 비해 모음의 존재가 뚜렷하고 모음의 구조가 복잡하며 무성음 뒤에는 반드시 모음이 온다는 특징이 있다. 따라서 한국어 음소 인식 시스템의 구현시 모음의 정확한 분류 인식은 중요한 비중을 차지하게 될 것이다. [1]

2. 한국어 모음의 종류 및 특성

음성은 크게 모음과 자음으로 나뉘어진다. 모음은 발음할 때 성대에서 진동된 공기가 입 안에서 아무런 장애도 받지 않고 자유로이 입 밖으로 흘러나오는 소리이다. 이와 반해 자음은 입 안에서 여러가지로 방해됨 받아 발음되는 소리이다.

모음은 단순모음과 이중모음으로 나뉘 수 있다. 단순모음은 치음부터 갈카지 소리값에 변화가 없이 내는 /이/, /아/, /어/같은 모음을 말한다. 이중모음은 치음과 나중의 소리값이 달라지는 모음이며 /외/, /야/, /의/와 같은 모음이 여기에 속한다. 다음은 한국어의 모음을 단모음, 이중 모음, 또는 반모음으로 구분해 놓은 것이다. [3]

다음 그림 1은 한국어 모음의 음가들 모음 사각도에 표시해 놓은 것이다.

3. 포르만트의 의미

모음의 자동 분류 인식을 위해 가장 많이 사용되는 특징 변수로는 포르만트가

있다. 포르만트(formant)란 어떤 한 음성의 음가가 다른 단 음성과 식별되게 하여 하나의 독립된 음가를 지니게 하는 공명강의 진동 주파수 성분을 말한다. 이 포르만트는 '음성 기관의 주요한 공명강인 인강과 구강의 공명에서 나오는 것으로 알려져 있으며 이 같은 혀와 입술의 움직임에도 관계한다. 모음의 포르만트는 우두소리와 성도 진체가 가져오는 공명상의 영향, 그리고 조음점의 움직임 등이 모두어저서 되는 것이다. [1]

인간의 성대 (Vocal bands)는 글라리넷의 리드와 마찬가지로 특납은 주파대의 소리를 낸다. 이 소리에 대해 구강, 비강, 인강은 공명체 구성을 하게 되며 이들 공명체의 여러 다른 모양의 조합이 여러 가지의 공명 곡선을 만들어 성대에서 나오는 소리 가운데서 이 공명 곡선에 맞는 소리만이 공명하게 한다. 즉 여러 다른 모음들은 구강과 인강의 모양을 여러가지로 달리 함으로써 만들어지게된다. 이 공명은 성도의 공명 곡선의

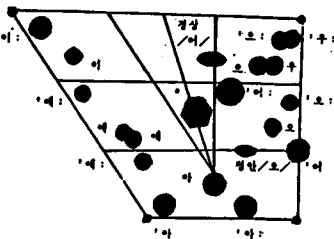


그림 1. 모음 사각도상의 한국어 음가.

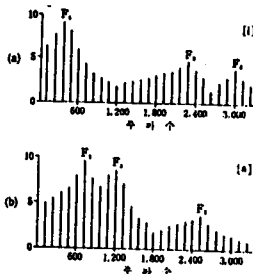


그림 2. [i]와 [a]의 스펙트럼과 안면도.

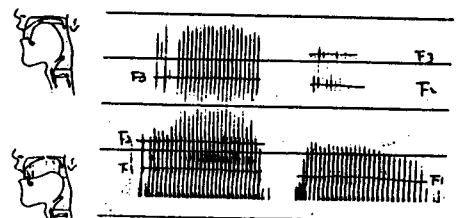


그림 3. 한국어 /아비/의 스펙트로그램을 통한 포르만트의 관찰.

특성에 따라 이루어지는데 대략 기본 주파수 500c/s, 1500c/s, 2500c/s가 봉우리를 이루고 있으며 이들을 중심으로 소리의 에너지가 집중해져 있다. 이때의 에너지가 집중된 부분의 주파수를 포르만트라 하며 차례로 제 1 포르만트, 제 2 포르만트, 제 3 포르만트라 한다. 그림 2에는 [i]와 [a]의 제 1, 제 2, 제 3 포르만트를 나타내는 스펙트럼이 나타나있다.

그림 4는 영어의 여덟개 모음의 F1, F2, F3의 세 포르만트를 정확한 주파수와 함께 표시한 것이다. 이 그림에서 보면 제 1 포르만트는 고모음에서 저모음으로 갈수록 높아지고 반대로

저모음에서 고모음으로 갈수록 낮아진다. 즉 제 1 포르만트는 혀의 높이에 반비례한다. 그리고 제 2 포르만트는 전심모음(front vowel)에서 후심모음(back vowel)으로 갈수록 낮아진다. 그리고 전심모음일수록 F1과 F2의 상호 간격은 벌어지고 후심 모음일수록 그 상호 간격은 좁아진다.

그림 5는 모음의 제1, 제2 포르만트의 좌표를 구해본 것이다. 이 때 원순 모음(round vowel)의 경우 입술을 둥글게 함으로써 입술이 앞으로 나오게 되고 이것은 결국 성도의 앞부분을 길게 해서 제 2 포르만트의 주파수를 낮게 한다. 이것이 원순 모음인 [ɔ]의 제 2 포르만트의 값이 낮은 이유이다. 그림 6은 한국어의 기본 모음의 제 1 포르만트와 제 2 포르만트의 모음 사각도를 그린 것이다. 이 그림을 통하여 F1과 F2가 한국어의 모음 분류의 중요한 단서가 됨을 알 수 있다.

	2,890	2,560	2,470	2,490
	2,250	1,920	1,770	1,660
	280	400	550	690
[i]	[ɪ]	[e]	[ɛ]	
	2,540	2,540	2,380	2,250
	1,100	880	1,030	870
	710	590	450	310
[a]	[ɔ]	[o]	[u]	

그림 4. 영어의 여덟개 모음의 포르만트의 값.

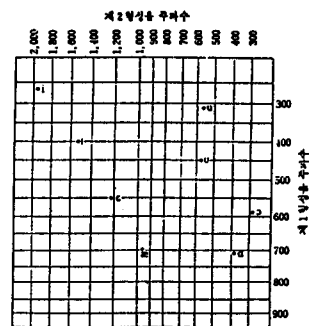


그림 5. 모음의 제 1, 제 2 포르만트의 좌표.

#### 4. 실험 방법

##### 4.1. 음성 시료

본 연구시 사용된 음성 시료로는 이미 구축되어 있는 한국어 VCV 음성 데이터의 일부를



5. 실험 결과 및 분석

실제된 모음 인식 모듈에 의하여 구분된

모음의 성공률이 다음 그림 11에 나타나있다.

분석 결과로는 [아]와 [오]의 성공률이 다른

모음에 비해 극히 저조함을 알 수 있다. 그 이유로는 첫째 표준 패턴과 매칭을 시도하는 테스트 데이터, [아]와 [오]의 모음의 중심부에서 추출한 포르만트의 값의 분포가 매우 넓어 (각각 다른 모음의 27.9배, 22.4배, 7.5배) 그 분포 구간 내의 표준 패턴을 갖는 다른 모음들과 올바른 구분이 되지 않기 때문이다. 둘째로 데이터의 잡음을 들 수 있다. 데이터가 좀 더 좋은 상황에서 녹음될 필요가 있으며 표준화시의 잡음도 많은 영향을 끼쳤으리라 생각된다. 그리고 마지막으로 포르만트 추적 알고리즘을 사용하지 않음으로 인하여 극의 추가 및 탈락 현상을 회복하지 못한 이유도 있다.

6. 결론

본 논문에서는 포르만트에 의한 모음의 분류인식을 행하는 시스템의 설계 방법을 제안하고 실험하였다. 음성 데이터가 필요 충분하고 또 적은 수의 데이터만 가지고 실험하였으나 [아], [오]를 제외하고는 비교적 좋은 결과가 나왔다. 좀더 빠르고 정확한 포르만트의 추출 방법의 연구가 필요하며 특히 모음부의 구분을 위한 알고리즘의 개발은 시급한 연구 분야일 것이다. 본 연구가 한국어 음소 인식 시스템의 구현에 좋은 기어가 될 것이다.

7. 참고 문헌

- [1]. 전상범, 영어 음성학, 을서 문화사, 1983
- [2]. 김재하, 음소 분류를 이용한 한국어 고립 단어 인식에 관한 연구, 한국 과학원 석사 논문, 1986.
- [3]. 이현복, 한국어의 표준 발음, 대한 음성학회, 1987.
- [4]. J. D. Markel, A. H. Gray, Linear Prediction of Speech, 1976.
- [5]. Stephanie S. Mccandless, "An Algorithm for Automatic Formant Extraction Using Linear Prediction Spectra", IEEE Trans. on ASSP, Vol. ASSP-22, No. 2, April, 1974, pp 135-141
- [6]. Clifford J. Weinstein, Stephanies. Mccandless, "A System For Acoustic Phonetic Analysis of Continuous Speech", IEEE TRANS. On ASSP, Vol. ASSP-23, No. 1, Feb. 1975, pp 54-72.
- [7]. Patrik F. Castelaz, Russel J. Niederjohn, "A Comparison of Linear Prediction, FFT, Zero-crossing Analysis Technique for vowel Recognition", ICASSP, 1978, pp 541-545.
- [8]. Renato De Mori, Pietro Laface, "Use of Fuzzy Algorithms for Phonetic and Labeling of Continuous Speech", IEEE TRANS ON PAMI, Vol. PAMI-2, No 2, March 1980, pp136-148.