

한국어 유성 파열음 / ㄱ, ㄷ, ㅂ /의 분류에 대한 연구

A Study on the Discrimination of
the Korean Voiced Stops / ㄱ, ㄷ, ㅂ /金 基 奭* · 黃 熙 隆**
(Ki-Seok Kim · Hee-Yeung Hwang)

요 약

본 논문은 한국어 파열음 중 유성 파열음인 / ㄱ, ㄷ, ㅂ /의 분류에 관한 연구이다. 유성 파열음의 조음점의 추론을 위해 유용한 특징 변수의 추출을 위한 알고리즘을 개발하고 이를 실험하여 개선하였으며 이중 가장 유효한 것으로 판단되는, 자음부와 후속 모음의 포르مان트의 추이 정도, 준 근원점 및 시간에 해당하는 다섯 개의 특징 변수를 사용하여 결정 프로그램을 설계하였다. 이들 특징 변수의 정확한 추출을 위해 후속 모음의 중앙부로부터 파열의 시작점까지의 포르مان트를 추적하는 포르مان트 추적 알고리즘을 구현하여 실험하였으며 결정 프로그램의 실세를 위해 후속 모음의 여러 환경에 따르는 특징 변수의 유용성을 확률로 계산하여 이를 지식으로 하여 정확한 결정을 할 수 있도록 하였다. 음성 데이터로는 각 자음당 64개의 모음 조합에 해당하는 모음·자음·모음 연쇄 음성 데이터를 사용하였다. 모두 192개의 자음에 대하여 70-80%의 인식율을 나타내었다. 이 실험의 방법과 연구 결과는 음성 발생 모델에 의한 한국어 음성 인식 시스템을 위한 여러 특징 변수의 추출 방법 및 패턴 인식 전략에 중요한 지침을 제공할 것이다.

Abstract - This paper is the study on the discrimination of Korean voiced stops / ㄱ, ㄷ, ㅂ /. We developed the algorithms for extraction of characteristic parameters for the inference of the place of articulation of Korean voiced stops and experimented their accuracy and improved them. And we selected five parameters among the experimented parameters, which are transition time of formants, formant pseudoloci and formant slopes, and then designed the decision program using these parameters. For the accurate extraction of these parameters, we designed the formant tracking algorithm that tracks the formant values from the central point of following vowel to the burst point. We calculated the weight of five parameters in according to various contexts of following vowels and made decision program for accurate recognition based on the knowledge of these weight. The speech data are VCV speech chain, V is the eight Korean simple vowels and C is the three Korean voiced stops. The recognition rate is resulted in about 70 or 80%. The method and results of the experiments will provide a guidepost for the method of extraction of the feature parameters and the pattern recognition strategy for the Korean phoneme recognition system following the speech production model.

*正 會 員 : 서울대 大學院 컴퓨터工學科 博士課程

**正 會 員 : 서울대 工大 컴퓨터工學科 教授·工博

接受日字 : 1988年 5月 12日

1次修正 : 1988年 9月 8日

1. 서 론

한국어 음성 인식 시스템의 구현에 있어서 무엇

보다도 선행되어야 할 문제가 한국어 음성의 음성학적 구조 및 특성의 연구라 하겠다. 한국어 음성의 음향 특성, 음소 분류에 필요한 여러 변별적 요소의 발견, 이음들의 집합에 대한 상세한 분류와 나타나는 예, 빈도수들에 대한 기초조사 등이 선행되어야 하며 이에 근거한 한 여러 음성학 및 음운학적 규칙들을 이용하여야 효율적인 음성 인식 시스템이 구현될 수 있다.

이러한 음성 인식 시스템의 구현시 고려해야 할 사항 중의 하나로, 음성의 최소 인식 단위를 무엇으로 하느냐 하는 것을 들 수 있다. 음성 인식의 단위로서는 이음(allophones), 음소(phonemes)이중음(diphones), 음절(syllables), 단어(words) 등이 있으며, 이러한 음성인식 단위와 음성 신호에서 추출한 특징(features)과의 관계를 기술하는 음운론적 규칙(phonological rule)의 발견은 음성 인식 시스템의 효율을 결정짓는 중요한 작업이라 하겠다. 이러한 음운론적 단위 중에서 음성 인식 시스템의 최소 인식 단위로 가장 많이 쓰이는 것중의 하나가 음소(phonemes)이다. 음소란 한 언어에서 서로의 사이에 의미적인 구분을 이루지 않고 같은 음으로 받아들여지는 이음(allophone)들의 집합을 말한다.¹⁾

음소를 음성 인식의 단위로 할 경우의 잇점은 다음과 같다. 첫째로 특정 언어에 있어서 존재하는 음소의 종류는 기타의 음운 단위보다 적다. 즉 일반적인 언어에 있어서 음소의 종류는 20-60여개이다. 그러나 이음 또는 음절의 종류는 수천가지가 존재한다. 따라서 찾아야할 탐색 공간이 적어지거나 표준 패턴의 저장 용량이 적어지는 등의 잇점을 가질 수 있다. 둘째로는 일반적인 개념으로 음소는 어휘 레벨의 지식과 곧바로 일치가 된다는 것이다. 즉 어휘 지식을 포함한 의미, 문법 지식이 음소 단위로 표현되어 있기 때문에 해석된 음소의 추론 작업이 용이할 수 있다. 따라서 상위 레벨의 지식을 이용하여 음성을 이해하는 시스템의 구성에 있어 해석된 음소의 또다른 변형이 필요 없는 효율적인 시스템을 구성할 수 있다.

그러나 음향 분석 단계에 있어서 각각의 음소들을 분할 및 분류 인식하는 것은 무척 어려운 작업이며, 또한 어떤 음소는 두 가지 이상의 음소군으로 인식될 수 있으며 음소에서의 상호 동화 현상(Coarticulation)과 인접 규칙(Junctural rule)에 의한 변화의 보정이 필요하다. 그러나 한국어에 있어서는 음소의 갯수가 적을 뿐 아니라 음소인식이 곧 음절 및 단어의 인식과 연결이 되므로 이러한 음운

론적 규칙들을 많은 실험에 의하여 발견한다면 보다 효율적인 음성 인식 시스템을 구성할 수 있으리라 생각된다.²⁾

한국어 음소의 특징중 하나는 18개의 자음 음소 중 파열음이 절반인 9개가 있다는 것이며 이는 보통 외국어의 경우 파열음의 종류가 6개인 경우와 비교가 된다. 9개의 파열음은 조음의 방식에 따라 크게 3가지 형태의 파열음으로 분류되며 이들은 각각 (a) 약간의 대기 현상과 약한 긴장성의부류(b) 대기 현상이 나타나지 않으며 강한 긴장성을 갖는 부류(c) 많은 대기 현상과 강한 긴장성을 갖는 부류등으로 특징지어진다. (a)에 해당되는 파열음은/ㄱ, ㄷ, ㅂ/이 있으며 이는 평음 파열음이라 칭하고(b)에 해당하는 파열음은/ㅋ, ㅌ, ㅃ/이 있으며 이는 무기 경음 파열음이라 칭하고(c)에 해당하는 파열음은/ㅋ, ㅌ, ㅍ/이 있으며 이는 유기 경음 파열음이라 칭한다.³⁾

조음 위치에 따른 한국어 파열음의 분류는 양순음(bilabials), 치경음(alveolars), 연구개음(velars)등으로 구분한다. 양순 파열음에는/ㅂ, , ㅍ/이 있으며 치경파열음은/ㄷ, ㅌ, ㅌ/, 연구개 파열음은/ㄱ, ㅋ, ㅋ/이 있다. 이러한 조음 방식과 조음 위치에 따라 9개의 파열음이 구분된다.

이러한 파열음은 초성의 위치에서 발음될때는 모두 무성으로 발음된다. 또한 유성음 사이의 중간음으로 발음될 때는(a)에 해당하는 파열음은 유성으로 발음된다. 그러나(b), (c)에 해당하는 파열음은 자기 고유의 특성을 변이시키지 않는다. 종성으로 발음될 때는 모든 파열음은 불파음으로 발음되며 이때의 조음 위치는 그대로이다.⁴⁾

이러한 한국어 파열음의 집합은 외국의 음성에서는 찾아 볼 수 없는 것이며 따라서 외국의 음성학에서 미처 연구되지 못하거나 고려되지 못한 여러 자질들에 대한 연구가 한국어 파열음에대상으로조시되고 있다. 이는 한국어 음성의 자동 인식을 위하여서도 한국어 파열음에 대한 새로운 연구가 필요함을 말한다. 본 연구는 이러한 한국어 파열음의 음성학적 분류를 위한 여러 자질들을 제안하고 이를 구현한 실험 결과들을 분석하여 이들 중 일부를 기지고 한국어 유성 파열음을 자동 분류해 보려는 실험이다. 이러한 연구의 궁극적인 목적은 이들 특성 중에서 컴퓨터에 의한 자동 추출이 가능한 여러 자질들을 골라 실험해 보기 위한 것이며, 결과적으로 자동 인식을 위해 유용한 자질이라고 실험적으로 판단된 것들은 한국어 음성 인식 시스템의 음소

인식 모듈에서 사용될 것이다.

이러한 파열음 음소의 자동 인식에 대한 연구는 외국의 경우에 있어서 음성 발생 모델, 음성 인지 모델 등에 근거한 여러 방법으로 수행되어 왔으며 대체적으로 70-95% 정도의 인식율을 나타내고 있다. 그러나 이들의 경우에는 파열음의 종류가 대략 6 개 정도로 한국어 음성과는 다른 환경에서 이루어졌으므로 이들의 결과를 한국어에 그대로 적용할 수 없다. 따라서 이 연구는 한국어의 음성 특징에 근거하여 한국어 파열음의 자동 분류를 행하는데 그 의의가 있는 것이다.

본 연구는 이를 위하여 한국어의 단모음중 가장 많이 쓰이는 8 개의 단모음을 대상으로 이 단모음과 한국어 파열음 중 유성 파열음에 해당하는 /ㄱ, ㄷ, ㅂ /의 조합 결합인 모음·자음·모음 연쇄 음성(VCV speech) 데이터를 사용하여 각 모음의 조합 하에서 /ㄱ, ㄷ, ㅂ /이 구분되는 특징들을 조사하고 이들 특징들을 이용하여 /ㄱ, ㄷ, ㅂ /의 분류를 행하였다. 제 2 장에서는 한국어 파열음을 스펙트로그램을 통하여 그 특성을 관찰한 결과들을 정리하였고, 3 장에서는 발견된 여러 특징들을 컴퓨터로 구현하기 위해 구성된 음향 분석 프로그램과 이들 프로그램에 의해 출력된 여러 그래프에 의한 음향 분석 과정을 기록하였다. 4 장에서는 인식 모듈에서 사용될 특징 변수의 정의 및 추출 방법을, 5 장에서는 특징 변수 추출 결과 및 분석을 행하였으며 6 장에서는 이들 특징 변수를 이용하여 구성된 인식 모듈의 실험 결과 및 분석을 행하였다.

2. 한국어 파열음의 스펙트로그램을 통한 음향 특성의 분석

파열음은 혀에서 올라오는 공기를 혀와 이와의 어떤 지점에서 일단 완전히 막아 압력이 높아진 다음에 갑자기 파열시켜 얻는 소리를 말하며 음향 분석에 있어서 어느 일정 기간 동안의 묵음과 폐쇄 해제에 따르는 진폭의 급격한 상승으로 특징지어진다. 그 폐쇄 해제시에는 약간의 마찰 잡음(friction noise)과 파열 잡음(burst noise)이 동반된다. 그림 1에는 한국어 파열음 /어타 /의 발성시 나타나는 파열음의 조음 단계를 스펙트로그램 상에 나타내었다.

스펙트로그램은 음성을 받아들여 이를 주파수 성분으로 분석하여 시각적으로 관찰할 수 있도록 한 그림을 말하며, 수평축은 시간을, 수직축은 진동수를

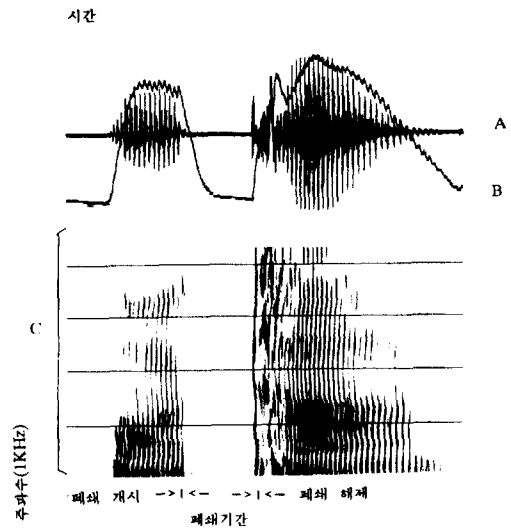


그림 1 스펙트로그램을 통해서 본 파열음의 조음 단계 (/어타 /)

Fig. 1 The articulation steps of the stops through the spectrogram.

말하고, 진동수의 강도는 명암으로 나타나 있다. 그림 1을 보면 A는 입력된 음성 파형 /어타 /의 전체 음성 신호를 나타낸 것이고 B는 이 음성의 소리 강도(amplitude)를 나타낸 것이며 C는 음성을 주파수별로 분석한 그림이다. C의 각 칸은 1KHz에 해당한다. 이 스펙트로그램에서 /어타 /의 발성시 파열음의 특징인 폐쇄 기간이 존재하며 이 기간은 소리 강도의 하강으로 나타나고 있음을 볼 수 있다. 그리고 스펙트로그램상의 주파수 대역에서도 명암이 공백으로 나타나는 부분이다. 그리고 /어타 /의 발성중 /ㅌ /에 해당하는 부분은 폐쇄 해제 직후 부분에도 고주파 대역에 있어서의 잡음과 500Hz 이하의 짙은 강도의 음성으로 나타남을 관찰할 수 있다.

이러한 조음 방식으로 발음되는 파열음의 분류를 위한 음향 특성(Acoustic features)으로는 파열 시작점과 후속 모음의 정상 발성 지점과의 시간 간격인 상대 진동 개시 시간(VOT; voicing onset time)이 있으며 이는 무성 파열음과 유성 파열음을 구분하는 중요한 특징 변수가 된다. VOT는 대기(Aspiration)을 위해 이완된 상태가 후속 모음의 발성을 위해 다시 수축되는 시간으로 볼 수 있다. 그림 2는 한국어 파열음 /ㅂ, ㅃ, ㅍ /의 발성시 나타나

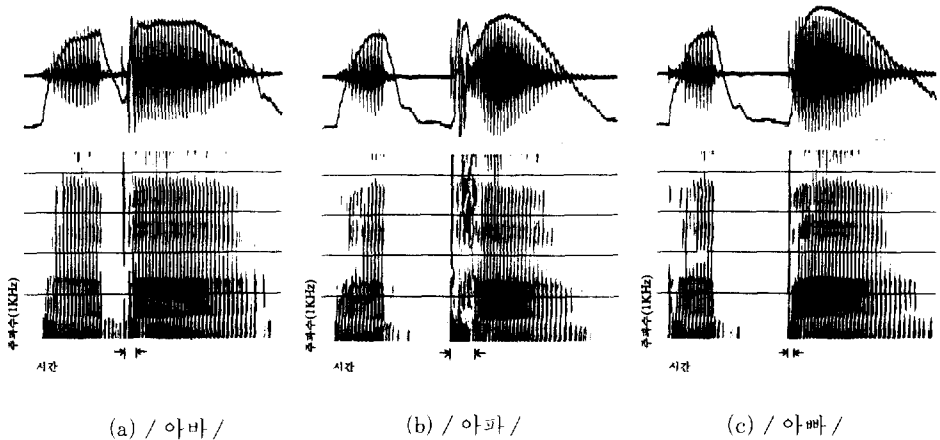


그림 2 조음 방식에 따른 한국어 파열음의 VOT의 차이

Fig. 2 The difference of VOT in Korean stops by the manner of articulation.

는 VOT의 차이를 스펙트로그램에서 관찰한 것이다. 그림에서 보듯 /ㅂ/에 해당하는 성대 진동 개시 시간은 /ㅃ/의 것보다 길며 /ㅍ/의 것보다는 짧다. 따라서 VOT는 파열음의 조음 방식을 구별하는 데 좋은 단서가 된다. 한편 VOT는 조음점에 따라서도 규칙적으로 변화된다는 것이 밝혀져 있다. 따라서 VOT는 유성 파열음 사이의 구분을 위해서도 좋은 특징 변수가 된다. 또한 제 1 포르مان트의 추이 및 기본 주파수 형상 등은 파열음의 유무성을 파악하는 주요 단서이다.⁴

파열음의 조음점의 위치를 파악하기 위한 특성으로는 파열시의 주파수 특성과 인접 모음의 포르مان트의 추이, 포르مان트의 준 근원점 등을 들 수 있다. 포르مان트란 성대에서 발생된 음성이 성도의 형상에 따라 공명하는 공명 주파수를 말하며 모음의 음질을 구분하는 척도이기도 하다. 자음의 음질은 포르مان트로 구분되지는 않지만 자음의 조음이 미치는 성도의 형상에 따르는 변화로 말미암아 인접 모음의 포르مان트의 형상에 변화를 주게 된다.

그림 3은 한국어 파열음 /ㄱ, ㄷ, ㅂ/의 발생시 나타나는 포르مان트의 추이를 스펙트로그램을 통해 관찰한 것이다. 그림에서 횡으로 검은 띠를 형성하고 있는 것이 포르مان트를 나타낸다. 모음부의 포르مان트의 값이 자음부에 가면서 서로 다른 모양으로 나타나는 것을 관찰할 수 있다. 각 파열음의 전후 모음을 모두 /아/로 하였을 때 각 자음에 따라 포르مان트의 추이 정도가 달라짐을 볼 수 있다. 즉 /

아/에 해당하는 부분의 제 1, 2, 3 포르مان트의 값은 각각 700, 1200, 2500Hz 정도이다. 그러나 /아가/의 경우 폐쇄 시작점에 있어서의 제 1, 2, 3 포르مان트의 값은 약 500, 1400, 2100Hz 정도로 나타나며 폐쇄 해제점에서는 명확히 나타나지는 않지만 거의 비슷한 값으로 나타난다고 볼 수 있다. 그리고 /아다/의 경우는 뚜렷한 변화가 없으며 /아바/의 경우는 제 3 포르مان트의 값이 폐쇄 시작점에서 낮아지고, 제 2 포르مان트의 값이 폐쇄 해제를 시점으로 각각 상승함을 보여준다. 이와 같이 포르مان트의 추이 정도 및 형상이 각 파열음마다 다르게 나타나며 따라서 이 특징으로 말미암아 파열음을 구분할 수 있다.

또한 그림 3에서 발음 /아가/에서 보듯 자음 /ㄱ/ 과 후속 모음 /아/ 사이에 나타나는 포르مان트의 추이가 한 곳으로 모아져 있음을 볼 수 있다. 이렇게 파열음의 폐쇄 지점에 따라 포르مان트의 추이의 방향이 일정한 지점을 가르키고 있으며 이 지점을 근원점(loci)이라 부른다. 물론 이 근원점은 정확하게 추정되지 않으며, 인접 모음에 따라서도 많이 달라지기 때문에 준근원점(Pseudoloc)이라 부르며 이 준근원점은 파열음의 조음점을 추론하는데 중요한 단서가 되기도 한다.

한편 이러한 포르مان트의 추이의 정도 및 형상은 인접한 모음에 따라서도 다르게 나타난다. 그림 4는 같은 파열음 /ㅂ/이 인접한 모음이 달라짐에 따라 포르مان트의 추이의 형상이 달라짐을 보여주고

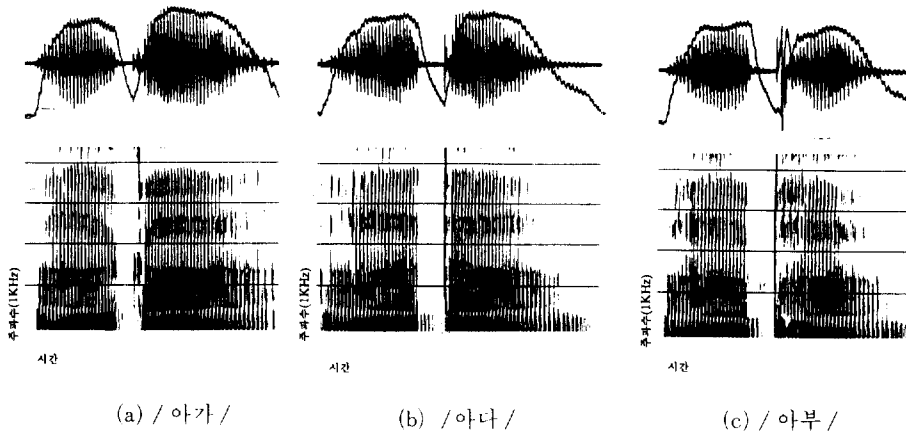


그림 3 조음점에 따른 한국어 파열음의 포르만트의 추이 형상

Fig.3 The pattern of formant transition of the Korean stops by the place of articulation.

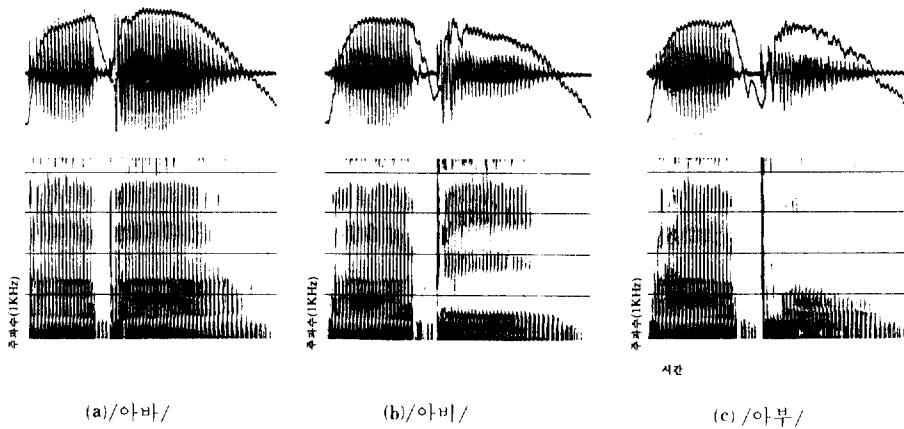


그림 4 후속 모음에 따른 한국어 파열음의 포르만트의 추이 형상

Fig.4 The pattern of formant transition of the Korean stops by the following vowel.*

있다. 즉 후속 모음만 달리 발음한 세 경우의 스펙트로그램을 보면 /아바/인 경우 전위 모음의 제2, 제3 포르만트의 값이 거의 수평으로 변화가 적지만 /아비/인 경우는 제2 포르만트의 값이 상승하는 것이 눈에 띄이며, 제3 포르만트는 급격한 하

락에서 밝혀진 여러 특징 변수들을 조사한 후 이들 중에서 자동 인식에 기여할 수 있는 특징 변수를 골라 이들을 구현하는 알고리즘을 실험하였으며, 그 중에서 한국어 유성 파열음을 자동 인식하기 위해 유용하다고 판단되는 포르만트의 추이 속도 및 추이

강을 나타낸다. 이와같이 파열음의 자동 인식을 위해서는 전후의 모음의 문맥에 따른 분석이 필요하다.

본 연구에서는 이러한 음향 특성을 가진 한국어 파열음의 자동 분류를 위해, 그동안의 실험 음성

시간을 전후 모음에 따라 조사하여 이에 따르는 다섯 가지의 특징 변수를 설정, 이를 가지고 한국어 유성 파열음을 자동 인식하는 실험을 행하였다.

3. 컴퓨터에 의한 음향 분석

3.1 실험 시료

본 연구에서 쓰인 음성 시료로는 한국어 연속 음성의 모든 문맥의 모음 조합인 모음·자음·모음연쇄 음성 데이터를 사용하였다. V=[아, 어, 오, 우, 으, 이, 예, 애], C=[ㄱ, ㄷ, ㅂ]을 사용하였으며 모두 8×3×8=192개의 데이터를 갖고 실험하였다. 조용한 방에서 메탈 테이프에 녹음한 뒤 4.95 KHz로 고주파 성분을 제거하였다. 10KHz로 표본화한 후 음성부만 플라 디스콰에 저장하였다. 화자는 1인의 남성 화자로 하였다.⁵⁾

3.2 음향 분석 방법

위의 2장에서 언급한 파열음의 특성을 자동 추출하기 위해서 표본화된 음성 데이터로부터 포르مان트를 비롯한 여러 음향 파라미터를 추출하여야 한다. 이를 위하여 음향 분석을 행하였다. 음향 분석을 위해 쓰인 분석 파라미터는 대수 에너지, 영교차율, 레지듀얼 에러, 포르مان트, 대역별 에너지 등이다.⁷⁾ 이들을 각각의 음성 데이터에 관하여 그 결과를 출력하여 유성 파열음 사이를 구별하는 음향 특성을 분석하는데 사용하였다. 즉 각 유성 파열음에서 후속 모음에 따라 포르مان트의 추이 정도 및 형상, V-OT의 차이 등을 관찰하기 위하여 사용되었다. 또한 여러 특징 변수 추출을 위한 기준점을 추출하기 위하여 사용되었다. 다음 그림 5, 6은 /어가/음성으로부터 계산된 각각의 음향 파라미터의 계산된 결과를 LOTUS 123에 의하여 그래프로 출력한 결과이다. 그림 5에서 에너지 곡선을 따라 급강하하는

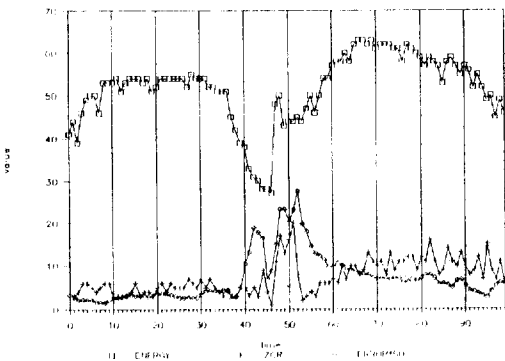


그림 5 대수 에너지, 영교차율, 레지듀얼 에러(/어가/)

Fig.5 Log energy, zero crossing rate and residual error. (/어가/)

시간축 약 40에 해당하는 점이 폐쇄 시작점이며, 시간축 약 45부근에서 에너지 곡선은 급상승하며 이는 폐쇄 해제점이다. 그리고 시간축 약 53부근에서 에너지 곡선이 약간의 굴곡을 나타내며 이 지점에서 영교차율의 급격한 하강을 관찰할 수 있다. 이 지점을 후속 모음의 성대 진동 개시점으로 볼 수 있다. 여기서 시간축 눈금 1은 시간으로 64msec 정도이다. 한편 그림 6에서 나타나는 5개의 궤적은 각각 아래로부터 제 1, 2, 3, 4, 5 포르مان트의 값이다.⁶⁾

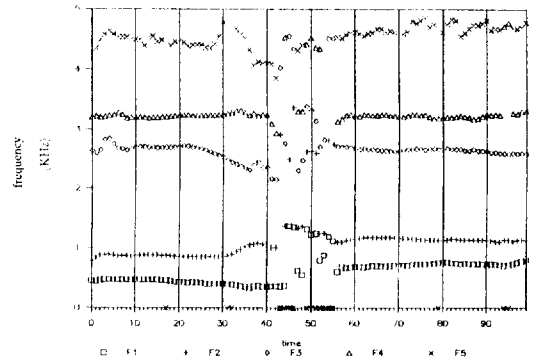


그림 6 포르مان트(/어가/)

Fig.6 Formants(/어가/)

4. 특징 변수의 정의 및 추출 방법

각각의 음성 데이터 192개에 대한 음향 분석 그래프를 모두 얻어 이를 분석하여 유성파열음의 분류를 위한 특징 변수를 정의하고 이를 분류한다. 유성 파열음의 분석을 위한 유용한 특징 변수라고 생각되는 다섯 가지의 특징 변수로서 FOSLF2, FOSLF3, FOPSF2, FOPSF3, DIST를 이용하였다. FOSLF2, FOSLF3란 후속 모음의 중앙부로부터 추적된 제 2, 제 3 포르مان트의 추이의 정도를 말하며 FOPSF2, FOPSF3란 파열시작시의 제 2, 제 3 포르مان트의 값을 말하며 DIST는 포르مان트의 추이가 일어나는 시간 간격을 말한다.^{8),9)} 이들을 구하는 알고리즘은 다음과 같다.

(step 1)

후속모음 중앙부 (IMDFV)를 추출한다.

파열 시작점 (ISTBURST)를 추출한다.

후속모음 안정 시작점 (IENDTR)을 추출한다.

(step 2)

포르مان트 추적 알고리즘을 이용하여 후속 모음안

정점에서부터 파열 시작점까지 포르مان트를 추적한다.

(step 3)

FOPSF2 = 파열 시작점의 F2의 값

= F2(ISTBURST)

FOPSF3 = 파열 시작점의 F3의 값

= F3(ISTBURST)

FOSLF2 = F2(ISTBURST) - F2(IMDFV)

FOSLF3 = F3(ISTBURST) - F3(IMDFV)

DIST = IENDTR - ISTBURST

첫번째 단계에서 후속모음중앙부(IMDFV), 파열시작점(ISTBURST), 후속모음안정시작점(IENDTR) 등은 각 특징 변수의 정확한 추출 및 해석을 위하여 그래프에서 수작업에 의하여 추출하였다. 여기에서 포르مان트 추적 알고리즘을 설명하면 다음과 같다.¹²⁾

전체 시간 프레임의 85-90%는 역 필터의 스펙트럼에서 추출한 피크 주파수가 곧바로 포르مان트 주파수가 되지만 약 10-15%는 피크의 추가 혹은 탈락 현상이 있게 된다. 따라서 보통 성인 남성의 발음시 제 1, 제 2, 제 3 포르مان트는 모두 3 KHz 보다 작다는 사실에 근거하여 피크 주파수로부터 정확한 포르مان트의 제적을 구할 필요가 있다.

현재의 분석 프레임 k에 대하여 nearest neighbor criteria에 의한 포르مان트 결정 방법은 다음과 같다.

각 i (i=1, 2, 3)에 대하여

$V_{ij} = | \bar{F}_i(k) - F_j(k-1) |$ 가 최소인 j를 구하여

$F_j(k) = \bar{F}_i(k)$ 로 한다.

이때 $\bar{F}_i(k)$: 측정하려는 k번째 프레임의 i번째 피크 주파수

$F_j(k-1)$: (k-1)번째 프레임의 j번째 포르مان트 주파수

L을 분석 프레임의 피크중 주파수 대역 임계치가 500Hz 미만인 피크의 갯수라고 하면

L=1 일때는 1% 미만이며

$F_j(k-1)$ 에 가장 가까운 값에 $\bar{F}_i(k)$ 으로 정해지고 나머지는 이전 프레임의 해당 포르مان트로 대체된다.

L=2 일때는 피크의 탈락이 생긴 경우로써 L개의 포르مان트 값은 nearest neighbor criteria에 의하여 구하고 나머지 (3-L)개의 값은 직전 프레임의 해당 포르مان트의 값으로 결정한다.

L=3 일때는 85-90%로 이 때는 피크 주파수가 곧 포르مان트 주파수가 된다.

즉, $F_i(k) = \bar{F}_i(k)$

L=4 일때는 피크의 추가가 생긴 경우로써 nearest neighbor criteria에 의해 3개의 포르مان트 값을 결정한다.

그림 7은 포르مان트 추적 알고리즘을 적용하기 이전의 제 2, 3 포르مان트이며 그림 8은 위의 방법에 의해 포르مان트의 피크의 추가와 탈락을 보정한 후의 포르مان트의 제적이다.

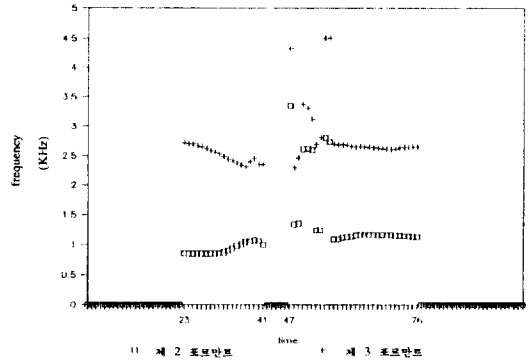


그림 7 포르مان트 추적 알고리즘 수행 이전 (/어가/)
 Fig.7 Before performing the formant tracking algorithm. (/어가/)

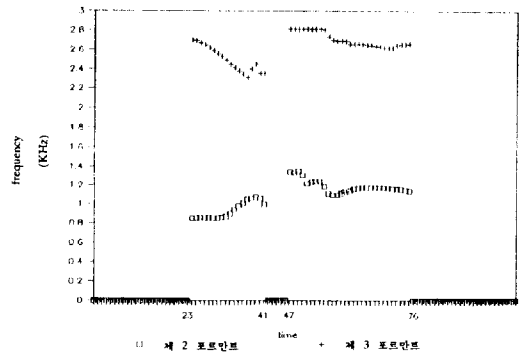


그림 8 포르مان트 추적 알고리즘 수행 이후 (/어가/)
 Fig.8 After performing the formant tracking algorithm. (/어가/)

5. 특징 변수 추출 결과의 분석

위의 알고리즘에 의하여 192개의 음성 데이터에 대하여 특징 변수를 모두 구한 뒤 각각을 후속 모음의 종류에 따라 분석한 뒤 이들 결과에 따라 표준 패턴 및 유효성을 출력한다.

5.1 표준 패턴 (TEMPLATE)의 작성

5개의 특징 변수 P_i 에 대해 다음과 같은 방법으로 표준 패턴을 얻어낸다. 먼저 n 개의 훈련 데이터 d_i 로부터 P_i 를 추출한 후 P_i 의 표준 편차가 최소화 될 것으로 예상할 수 있는 70%의 P_i 를 취해 그것의 평균을 표준 패턴의 값으로 갖는다.

표 1 표준패턴표(후속 모음을 구별하지 않을 때)

Table 1 Template table (when the following vowel not discriminated)

i j	FOPSF2	FOSLF2	FOPSF3	FOSLF3	DIST
/ㄱ/	1403	-54	2526	62	7.4
/ㄷ/	1427	-70	2509	88	4.6
/ㅂ/	1345	24	2543	65	5

단위 : FOxxFx (Hz)
DIST (frame)

표 1은 후속 모음을 구별하지 않을 때 각특징 변수의 /ㄱ, ㄷ, ㅂ/간의 평균치를 나타낸 것이다. 표 1을 보면 후속 모음을 구별하지 않을 때의 각각의 특징 변수의 값을 나타낸 것으로 FOPSF2는 폐쇄 해제점에 있어서의 제 2 포르مان트의 값으로 /ㄱ/이 1403Hz, /ㄷ/은 1427Hz, /ㅂ/은 1345Hz로써 /ㅂ/의 평균이 /ㄱ, ㄷ/의 평균과 큰 차이가 있음을 알 수 있다. 특히 FOSLF2 즉 포르مان트의 추이 정도는 /ㄱ/과 /ㄷ/에 있어서는 음수로 포르مان트가 후속 모음으로 갈수록 상승함을 알 수 있으며 그 기울기가 /ㄷ/의 경우가 더 큼을 볼 수 있다. 한편 /ㅂ/의 경우는 하강한다. FOPSF3 즉 폐쇄 해제점에 있어서의 제 3 포르مان트의 값은 /ㄱ, ㄷ, ㅂ/이 약간의 차이를 갖고 분포하고 있으며 제 3 포르مان트의 기울기는 모두 하강한다. DIST는 폐쇄 해제부터 후속 모음 안정 시작점까지 시간 간격을 말하며, 연구개음인 /ㄱ/이 다른 음

표 2 표준패턴표(후속 모음이 /아/일때)

Table 2 Template table. (when the following vowel is /아/)

i j	FOPSF2	FOSLF2	FOPSF3	FOSLF3	DIST
/ㄷ/	1233	-57	2506	89	7.7
/ㄷ/	1416	-190	2605	125	3
/ㅂ/	1106	59	2640	53	3.8

단위 : FOxxFx (Hz)
DIST (frame)

보다 대체로 늦다는 것을 확인할 수 있었다. 표 2는 후속 모음이 /아/일 경우에 있어서의 각각의 특징 변수의 값으로 표 1의 경우보다 더 뚜렷한 차이를 나타낸다.

5.2 각 특징 변수의 유효성

각 특징 변수가 자음의 특성을 나타내는 정도를 측정하기 위해 분석된 결과가 표 3과 4에 나타나 있다. 표에서 관찰하면 후속 모음을 구별하지 않을 때 FOSLF2는 /ㄱ/을 구분하는 데는 잘 기여하지 못함을 알 수 있으며 /ㄷ, ㅂ/을 구별하는 데는 대체로 기여함을 나타낸다. FOSLF2는 /ㅂ/을 잘 구분해내며 FOSLF3는 /ㄱ/을 구분해 내지 못한다. 표 4에서 DIST는 후속 모음이 /아/일 때 /ㄱ/을 확실히 구분해 내며, /ㄷ, ㅂ/인 경우에도 인식에 상당히 기여함을 볼 수 있다. 또한 5개의 특징 변수 모두 /ㄷ/에 대해서는 각각 45-50%의 인식율을 보이며 5개의 변수에 대해서 적절한 지식을 갖춘 결과 75%의 인식율을 나타내었다. 5개의 모든 경우에 대하여 위의 실험 결과로써 각 특징 변수가 어떤 자음의 분류에 유효한가를 조사하였다.

표 3 각 특징 변수의 인식율(후속 모음을 구별하지 않을 때)

Table 3 The recognition rates of each characteristic parameters. (when the following vowel not discriminated)

i j	FOPSF2	FOSLF2	FOPSF3	FOSLF3	DIST
/ㄱ/	4	23	0	50	70
/ㄷ/	54	48	45	53	54
/ㅂ/	53	82	46	9	48

표 4 각 특징 변수의 인식율(후속 모음이 /아/일 때)

Table 4 The recognition rates of each characteristic parameters. (when the following vowel is /아/)

i j	FOPSF2	FOSLF2	FOPSF3	FOSLF3	DIST
/ㄱ/	62	62	75	37	100
/ㄷ/	75	75	62	62	75
/ㅂ/	62	75	62	62	75

5.3 지식의 획득

각 자음 C_i 에 대한 특징 변수 P_i 의 유효성을 극

대화하기 위해 n개의 훈련 데이터 d_x 에 대해 다음과 같은 훈련을 시도하여 P_j 의 C_i 에 대한 지식 K_{ij} 를 얻어낸다. 즉 모든 d_x 에 대해서 P_j 에 의해 C_i 라고 인식된 횟수를 k라고 하고 k중 ($d_x=C_i$)인 횟수를 k' 라고 하면 $K_{ij}=(k'/k) \times 100$ 이다. 얻어진 K_{ij} 에 대해 인식 실험을 반복하여 최대의 인식율을 보이는 K_{ij} 값을 지식으로 갖게 한다. 지식에 대한 값이 표 5에 나타나있다.

표 5 지식표 (후속 모음을 고려하지 않는 경우)
Table 5 Knowledge table. (when the following vowel not discriminated)

i \ j	FOPSF2	FOSLF2	FOPSF3	FOSLF3	DIST
/ㄱ/	60	47	14	36	65
/ㄷ/	38	46	46	40	51
/ㅂ/	35	55	25	38	57

6. 인식 시스템의 구현과 실험 결과 분석

위에서 구한 특징 변수 값 및 지식을 근거로 하여 인식 시스템을 설계하였다. 인식 시스템은 크게 세 가지 방법으로 구성하여 실험하였다. 첫째, 후속 모음을 고려하지 않고 구해진 표준 패턴과 비교하는 법, 둘째로는 후속 모음을 음소 단위로 인식하지 않고 단지 전설, 중설, 후설의 여부만을 인식하여 그 결과에 따라 표준 패턴과 비교하는 법, 셋째로 후속 모음을 인식한 후 그 인식된 모음에 해당하는 표준 패턴과 비교하는 법 등으로 인식 실험을 해 보았다.

6.1 인식 결정 방법

5개의 특징 변수 P_j 중 테스트 데이터 d_x 에 대해서 자음 C_i 라고 인식하는 P_j 의 C_i 에 대한 지식 K_{ij} 의 합이 최대가 되는 자음을 최종 인식 자음으로 결정한다.

6.2 실험 결과

실험 결과가 표 6, 7, 8에 나타나있다. 표 6은 후속 모음을 구별하지 않은 채 구한 표준 패턴에 의해 인식한 인식율이다. 표 7은 후속 모음을 전설모음, 중설모음, 후설모음으로 구별하여 각각에 따른 표준 패턴을 따로 구하여 인식한 실험결과로 약 71%의 인식율을 보인다. 한편 표 8은 후속

모음을 정확히 인식한 후의 인식율로 약 80%의 인식율을 보였다.

표 6 후속 모음을 고려하지 않은 자음 인식 결과

Table 6 The recognition rates. (when the following vowel not discriminated)

대상결과	/ㄱ/	/ㄷ/	/ㅂ/	인식율
/ㄱ/	30	18	16	46(%)
/ㄷ/	3	48	13	75(%)
/ㅂ/	3	18	43	67(%)
			총	63(%)

표 7 전설, 중설, 후설 인식 후 자음 인식 결과

Table 7 The recognition rates. (when the following vowel discriminated into the front, middle, back vowels)

대상결과	/ㄱ/	/ㄷ/	/ㅂ/	인식율
/ㄱ/	43	10	11	67(%)
/ㄷ/	3	45	16	70(%)
/ㅂ/	9	5	50	78(%)
			총	71(%)

표 8 후속 모음 인식 후 자음 인식 결과

Table 8 The recognition rates. (when the following vowel discriminated)

대상결과	/ㄱ/	/ㄷ/	/ㅂ/	인식율
/ㄱ/	52	3	9	81(%)
/ㄷ/	3	54	7	84(%)
/ㅂ/	7	8	49	76(%)
			총	80(%)

6.3 실험 결과 분석

구성된 인식 시스템의 실험 결과에 의하면 후속 모음을 고려하지 않은 경우 인식율이 가장 낮으며, 후속 모음을 고려하는 경우에는 인식율이 높아짐을 볼수 있다. 이는 파열음의 인식에 있어서 후속 모음의 인식이 선행되어야 함을 입증한다. 또한 전위 모음의 인식도 후속되는 파열음의 분류에 영향을 미치리라 예측할 수 있다. 왜냐하면 인식을 위해 사용한 주요 특징 변수인 포르만트에 있어서 해당 파열음에서의 패턴 변화 형태가 양쪽 모음에 의하여 달라지기 때문이다.¹²⁾

또한 대체적으로 후속 모음을 고려하지 않거나, 전설, 중설, 후설 모음 정도로만 분류하는 경우에는 연구개 파열음인 / ㄱ / 의 분류 인식율이 제일 낮은 것으로 나타났다. 이는 / ㄱ / 의 분류에 있어서 폐쇄 해제시 공기의 파열이 다른 것 보다도 약하게 나타나며 따라서 정확한 폐쇄 해제점 및 상대 진동 개시점을 찾을 수 없기 때문이다. 이는 후속 모음을 구별했을 경우에는 비교적 높은 인식율을 회복할 수 있었다.

또한 본 실험에서는 5 개의 특징 변수를 이용하여 / ㄱ, ㄷ, ㅂ / 의 분류를 시도하였다. 실험결과 임의의 n개의 특징변수 (n=1, 2, 3, 4)를 선택했을 때 보다 5 개 모두를 선택했을 때 평균 인식율이 높게 나타났다. 따라서 유효한 특징 변수를 많이 찾으려는 연구가 필요하다고 본다. 그러나 / ㄱ, ㄷ, ㅂ / 에 대해서 공통적으로 유용한 특징 변수를 기대하기 어려우므로 잘 훈련된 지식이 필요하다.

7. 결 론

본 연구에서는 한국어 파열음중 유성 파열음인 / ㄱ, ㄷ, ㅂ / 에 관하여 후속 모음에 따라 이들을 구별할 수 있는 5 개의 특징 변수를 구해 이들의 값과, 그 유효성을 지식으로 유성 파열음을 분류하는 실험을 행하였다. 후속 모음의 구별 정도에 따라 인식율이 달라졌으며 후속 모음을 각각 구별한 후 얻은 특징 변수 값을 지식으로 했을 경우 약 80%까지의 인식율을 보였다.

이 연구는 음성 패턴의 결정시 사용되는 표준 패턴을 특징 변수들의 평균치 및 계산된 유효성값만을 사용함으로써, 표준 패턴의 저장 메모리를 기존의 고전적인 패턴 인식의 것보다 적게 했으며, 또한 입력된 음성을 기존의 패턴과 비교하는 매칭 시간도 적게 나타났다.

본 연구는 1인 남성 화자의 음성만을 이용하여 인식 실험을 행하였으므로 화자 독립적인 인식 시스템을 구성하기 위해서는 여성을 포함한 다수의 음성 데이터로 실험을 행하여야 하며, 또한 더 의미 있는 특징 변수를 발견하고 컴퓨터 프로그램으로 구현하는 실험을 행하여야 함과 아울러 기존의 특징 변수들도 더 정확히 추출하기 위한 실험도 행하여야 한다. 또한 이 실험 과정을 응용하여 파열음의 모든 이음들의 분류에 관한 연구 및 기타 자음의 분류에 관한 연구 등이 계속되어야 한다. 또한 패턴의 매칭 작업시 확률에 의한 거리 계산 방법 등도

개선되어야 할 것이다.

이 실험 결과는 한국어 자음 음소의 반을 차지하는 파열음의 분류에 이용되며, 또한 음성 발생 모델에 의한 한국어 음성 인식 시스템의 구성에 이용될 것이다.

본 논문은 1986년도 문교부 자유 과제 학술연구 조정에 의하여 수행된 연구 결과임

참 고 문 헌

- 1) 김기석, 최윤석, 황희용, "연속 음성 상에서의 한국어 자음 / ㄱ, ㄷ, ㅂ / 의 자동 분류에 관한 연구", 1987년도 한국 정보 과학회 가을 학술 발표 논문집, vol. 14, no.2, pp163-166.
- 2) Wayne, A. Lea, "Trends in speech recognition", 1980.
- 3) Yang-Soo Moon, "Some Aspect of Korean Stop Consonants", 언어학 3호, pp81-88.
- 4) Kim Chin Woo, "On the Autonomy of Tensity Feature in Stop Classification", Word 21.3, 1965, 339-359.
- 5) 오영환, 국어 음성의 모음·자음·모음 연쇄의 음향 분석, 한국 전자 통신 연구소 연구논문집, 1987.
- 6) 오영환, 숫자음 인식을 위한 패턴 매칭 알고리즘 개발 연구, 한국 전자 통신 연구소 연구논문집, 1986.
- 7) 김기석, 최윤석, 황희용, "포르만트를 이용한 한국어 난모음 분류 시스템의 개발에 관한 연구", 1987년도 전기·전자공학 학술대회 논문집, 87 / 7, pp1107-1110
- 8) Hiroya Fujisaki, Masahiko Tominaga, "Automatic Recognition of Voiced Stop Consonants In CV and VCV Utterances," IEEE ICASSP, 1982, pp. 1996-1999.
- 9) Piero D emichelis, Renato De Mori, Pietro Laface, Mary O, Kane, "Computer Recognition of Plosive Sounds Using Contextual Information," IEEE Trans. on ASSP., vol, ASSP-31, no. 2, April, 1983, pp. 59-377.
- 10) Renato De Mori, Pietro Laface, and Yu Mong, "Parallel Algorithm for Syllable Recongnition in Continuous Speech," IEEE Trans. on PAMI, vol. PAMI-7, no. 1, Jan., 1985, pp. 56-69.
- 11) V.W.Zue, Acoustic Characteristics of Stop Consonants: A Controlled Study, 1980.
- 12) J.D.Markel and A.H.Gray, Linear Prediction of Speech, 1976