論文97-34S-4-4

조음 특성과 음소 대표 구간을 이용한 우리말 파열음의 인식

(Plosive Consonants Recognition Using Acoustic Properties with the Frames Representing Each Phoneme)

朴 贊 應 **, 李 夬 熙 *

(Chan Eung Park and Kwae Hi Lee)

요 약

우리말 음소들 중 모음계열의 음소들과 비음, 유음계열의 자음들은 비교적 긴 시간 동안 안정된 특성을 보이는 데 반하여 무성 자음들은 음소 구간에서의 특성이 안정적이지 못하다. 또한 조음점은 같으나 조음 방법에 따라 다른 음소로 나타나는 음소들은 유사한 특성을 갖고 있어 이들의 분류를 어렵게 한다. 따라서 이들 구간 내에서의 조음 특성들을 이용한 특징을 찾아 음소를 구분하므로서 인식율을 향상시킬 수 있다. 이들 무성자음들은 일부가 어중에서 유성자음화 할 경우를 제외하고는 이러한 특징이 뚜렷이 나타나므로 비교적 용이하게 인식될 수 있다. 이들의 인식을 위한 입력으로는 음소 대표구간 추출을 통한 음소의 대표 프레임들이 사용되었고, 이들 대표 프레임에서 특징을 추출하여 신경망에 적용하였고, 각 음소들의 조음 특성들을 이용한 후처리 과정을 통하여 음소를 분류하였다. 조음점이 양순, 치경, 연구개인 9개의 무성 파열음 자음들에 대한 인식을 통하여 유사 조음점을 갖는 음소들 내의 분류에서는 89.4%의 인식율을 얻었고, 조음계열 분류까지를 포함할 경우에는 85.6%의 인식율을 기록하였다.

Abstract

Korean unvoiced phonemes consist of nonstationary parts comparing that the vowels and nasal consonants consist of quasi-stationary part. And some phonemes, which have same point of articulation but different manner of articulation, has similar characteristics, so it makes to be hard to distinguish each other. A new method using changes and characteristics of acoustic properties of these phonemes to improve recognition rate are proposed. And because these changes and characteristics evidently occur in continuous speech except some unvoiced consonants are articulated as voiced phoneme in case to be used as an midial between voiced phonemes, this method can be applied easily. The features of the frames extracted to represent each phonemes are used as inputs to the hierarchical neural network. And with these results final decision for phoneme recognition is made through post processing which the new method is applied to. Through the experimental recognition results for 9 unvoiced consonants which belong to bilabial, alveolar, and velar phoneme series, 89.4% recognition rate to distinguish in same phoneme series is obtained, and 85.6% recognition rate is obtained in case of including distinguishing phoneme series.

* 正會員, 西江大學校 電子工學科

(Electronic Engineering Dept., Sogang Univ.)

** 正會員, 仁德專門大學 放送通信科

(Broadcasting & Communication Dept., Induk Junior College)

接受日字: 1996年11月22日, 수정완료일: 1997年1月8日

I. 서 론

음성 인식에의 접근방법에는 음소 등의 음성학적 단위들의 특성을 이용하여 음성을 음성학적 단위들로 구분하는 음성, 음향학적 접근방법, 통계적인 모델링을 이용하는 DTW(Dynamic Time Warping), HMM 등의 통계적 패턴인식 접근 방법, 인간의 인식과정을 모방

한 인공지능 접근 방법들이 있으며 신경망은 인공지능 접근 방법의 범주에 넣고 있다. 또한 인식단위(예: 음소, 음절, 어휘 등)의 구분과 인식이라는 관점에서의 접근방법도 2가지로 구분하여 볼 수 있는데, 첫째는 인식단위의 구분과 인식을 동시에 수행하는 접근 방법이 있고, 둘째는 음성 신호를 우선 인식단위들로 구분한 후에 그 인식단위들을 인식하는 접근방법이 있다. 근래에는 주로 HMM 혹은 HMM과 ANN의 결합된 형태의 알고리듬[1][2][3]을 이용한, 인식단위의 구분과 인식을 동시에 수행하는, 인식 시스템들이 연구의 주류를 이루고 있다.

본 논문에서는 음소를 인식단위로 사용하고, 음성신호를 음소로 구분한 후에 그 음소들을 인식하는 접근 방법을 사용하는 시스템에서 모음이나 비음 및 유음에 해당하는 자음들은 비교적 긴 시간 동안 특성을 지속적으로 변하지 않고 있어 음소를 대표하는 프레임을 추출하여도 높은 인식율을 얻을 수 있는 데 반하여, 무성 자음들은 음소 구간에서의 특성의 변화가 심하며, 음소 구간의 길이도 음소, 혹은 사용 위치 등에 따라 많은 차이가 있어, 음소단위의 인식에서 음소의 분류를 어렵게 하고 있다. 특히 파열음에 해당하는 양순, 치경, 연구개음 계열의 음소들은 각 계열 내에서 이완, 긴장, 대기의 조음 방법에 따라 구분되는 음소들 사이의 거리가 매우 근접해 있어 구분이 매우 어렵다[4]. 따라서 이들 음소 구간에서는 조음의 특성이 변화하는 특징을 찾아 이들 특징이 이완, 긴장, 대기의 어느 조음 방법에 해당하는 가를 비교해 봄으로서 음소를 인식하게 된다. 이들 조음상의 특징을 살펴보면, 이완 조음 방법에 해당하는 음소들은 짧은 파열음 뒤에 비교적 약한 대기음(aspiration)구간이 뒤에 나타나며, 긴장 조음 방법에 해당하는 음소는 짧은 파열음만이 존재하며 대기음 구간은 존재하지 않는다. 또한 대기 조음 방법에 해당하는 음소들은 마찬가지로 짧은 파열음 뒤에 대기음 구간이 존재하나 비교적 강한 대기음이 나타난다. 그리고 이러한 파열음들은 전설 모음 앞에서는 아주 약한 대기음을 띠므로서 다른 경우와 구분된다. 이와 같은 파열음들이 갖는 조음 특성들을 분류하여 이들 특징을 음소 결정 과정에서 적용하므로서 이들에 대해서도 높은 인식율을 얻을 수 있다.

우선 파열음들의 인식을 위하여 음소 대표구간 추출을 통하여 각 음소를 대표하는 프레임들을 추출하고, 이들 프레임에서 특징을 추출하여 계층 신경망에 적용하였다. 계층 신경망은 우선 파열음을 양순, 치경, 연구개 계열로 구분하는 신경망을 상위 계층에 두고, 그 결과를 이용하여 각 계열별 신경망에서 인식이 이루어지도록 하였다. 각 신경망들은 2개의 은닉층을 갖는 다층 퍼셉트론 신경망으로 구성되었다. 그리고 이들의 분류 결과를 앞에서 설명한 조음 특성들을 후처리 과정에 적용하여 음소를 결정하게 된다. 일반적으로 거리가 근접해 있어 분류에 어려움이 있는 것으로 알려진 조음점이 양순, 치경, 연구개인 9개의 파열음들에 대하여 인식 실험을 수행하였다.

Ⅱ. 우리말에서의 음소

1. 음운론적인 음소의 구성

우리말의 음운은 자음과 모음으로 설정되며 이를 분절 음운 또는 기본 음운이라고 한다. 자음의 음소 수는 반모음이 쓰이는 위치에서 보면 자음과 같으므로 자음으로 간주하면 21개가 되고, 모음의 음소 수는 단모음 10개(혹은 8개), 복모음 10개(혹은 12개)로 구성되어 총 41개가 된다[5].

우리말 자음에 속하는 음운들은 조음점과 조음 방법에 따라서 조음되며 총 19개의 자음들이 있다. 조음점으로는 양순, 치경, 치경구개, 연구개, 성문들이 있다. 조음체의 작용에 따른 조음 방법에는 단순 또는 이완, 긴장, 대기, 비강, 설측들이 있다. 이들 자음 음운을 조음할 때에는 조음점과 조음방법은 동시에 작용한다. 표 1은 조음 계열에 속한 자음들을 보인 것이다.

표 1. 자음 계열 음운들의 조음[6]
Table 1. Articulation of Korean consonants.

조음점 / 조음방법	양순	치경	치경구개 (파찰)	치경구개 (마찰)	연구개	성문
단 순	p(ㅂ)	t(ㄷ)	c(ㅈ)	s(ㅅ)	k(ㄱ)	
긴 장	pp(ㅃ)	tt(ㄸ)	cc(ㅉ)	ss(ㅆ)	kk(ㄲ)	
대 기	ph(ㅍ)	th(ㅌ)	ch(ㅊ)		kh(ㅋ)	h(ㅎ)
비 강	m(ㅁ)	n(ㄴ)			ŋ(ㅇ)	
설 측		l(ㄹ)				

모음에 속하는 음운들은 두 입술의 모양, 혀의 높낮이 및 혀의 앞과 뒤에 따라서 조음된다. 즉 입술의 모양에는 평순과 원순의 두 가지가 있고, 혀의 높낮이는 고, 중, 저의 세 가지가 있다. 또한 혀의 앞과 뒤는 높낮이와 동시에 작용되는 것으로, 혀의 앞을 높이느냐, 뒤를

높이느냐에 따라서, 높이가 같은 음운들일지라도, 서로 대립의 관계를 이룬다. 즉 전설과 후설 모음 음운이다. 모음의 구조적 계열을 표 2에 보였다.

표 2. 모음 계열 음운들의 조음
Table 2. Articulation of Korean vowels.

혀의 앞뒤 / 입술모양 / 혀의 높낮이	전 설		후 설	
	평 순	원 순	평 순	원 순
고	ㅣ	ㅟ	ㅡ	ㅜ
중	ㅔ	ㅚ	ㅓ	ㅗ
저	ㅐ		ㅏ	

이들 모음 중에서 'ㅡ'의 경우는 어두, 어중에 흔히 쓰이고, 어말에는 외래어에서 많이 쓰인다. 또한 반모음 'w'와 'y'들은 'w'의 경우는 조음은 'ㅜ'와 거의 같고, 'y'의 경우는 조음이 'ㅣ'와 거의 같으나, 이들은 독립적으로 쓰이는 것이 아니라 항상 다른 모음 앞에서만 쓰이므로 반모음(semi-vowel)이라 부른다. 'y'는 'ㅖ', 'ㅒ', 'ㅕ', 'ㅛ', 'ㅠ', 'ㅑ'들 앞에서만 조음되어 'ㅖ', 'ㅒ', 'ㅕ', 'ㅛ', 'ㅠ', 'ㅑ'들의 복모음이 되고, 'w'는 'ㅣ', 'ㅖ', 'ㅒ', 'ㅕ', 'ㅑ'들 앞에서만 조음되어 'ㅟ', 'ㅞ', 'ㅙ', 'ㅝ', 'ㅘ'들의 복모음이 된다. 'y'의 경우는 'ㅡ'의 뒤에서 조음됨으로서 'ㅢ'의 복모음으로 조음되기도 한다.

2. 인식 단위로서의 음소

앞에서 살펴본 음운론적인 음소의 구분에 의하면 우리말에서의 음소의 수는 총 41 개였다. 그러나 이들 음소는 복모음 10개(혹은 12개)를 포함한 것으로 복모음을 반모음+단모음('ㅖ', 'ㅒ', 'ㅕ', 'ㅑ', 'ㅛ', 'ㅠ', 'ㅞ', 'ㅙ', 'ㅝ', 'ㅘ', 'ㅚ', 'ㅟ') 혹은 단모음+반모음('ㅢ')이라는 관점에서 분리하여 인식할 경우, 이들 복모음은 음소의 수에서 제외될 수 있다. 또한 단모음 중 'ㅐ'와 'ㅔ'의 경우는 다차원 척도상의 거리가 매우 근접하여 있어 단모음의 인식에서는 동일 모음으로 취급하고, 상위 레벨의 처리에서 구문론적인 분석에 의하여 구분하는 것이 더 효율적일 것이다.

표 3. 인식 단위로서의 모음
Table 3. Korean vowels as phonetic-like unit.

	단 모 음	반 모 음
인식 단위로의 음소	ㅏ, ㅓ, ㅗ, ㅜ, ㅡ, ㅣ	y, w

자음들의 경우에는 더욱 복잡한 양성을 띤다. 즉 조음점과 조음 방법은 일치하나, 쓰이는 위치에 따라 다른 특성을 갖도록 조음된다. 예를 들면, 'ㄹ'은 어두(語頭)에서는 'r'과 같이 조음되나, 어말(語末)에서는 'l'로 조음되며, 어중(語中)에서는 인접 음소에 따라 'r' 또는 'l'로 조음되기도 한다. 따라서 이러한 음소는 인식의 관점에서는 'r'과 'l'로 구분하여 인식하는 것이 인식율을 높일 수 있다. 또한 조음 방법들 중에서 파열음 중 'ㅂ', 'ㄷ', 'ㄱ'들은 구강 내부의 어느 부분이 완전히 폐쇄되거나 파열되면서 나는 소리로, 어두에서는 파열음으로, 어말에서는 폐쇄음으로 조음되며 무성음으로 분류된다. 그러나 이러한 음운들은 어중에서는 모음과 모음사이 혹은 모음과 'ㅁ', 'ㄴ', 'ㅇ', 'ㄹ'들의 사이에서는 유성음으로 조음된다. 또한 'ㅈ'의 경우는 어말에는 사용되지 아니하나, 어중에서는 모음과 모음사이 혹은 모음과 'ㅁ', 'ㄴ', 'ㅇ', 'ㄹ'들의 사이에서는 역시 유성음 화한다. 따라서 이들 음소들은 쓰이는 위치 즉 어두, 어중, 어말에 따라 다른 특성을 갖도록 조음되므로 별도의 음소, 즉 변이음으로 간주하여 인식하는 것이 바람직하다. 나머지 자음들의 사용 위치에 따른 특성의 차이는 매우 적다. 이러한 경우까지를 모두 감안하면 인식 단위로서의 자음 음소는 표 4.6에서와 같이 총 26개가 된다.

표 4. 인식 단위로서의 자음
Table 4. Korean consonants as phonetic-like unit.

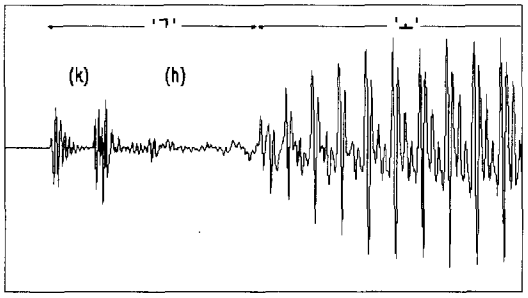
조음점 / 조음방법	양순	치경	치경구개 (파찰)	치경구개 (마찰)	연구개	성문
단 순	$ㅂ^{i}, ㅂ^{t}, ㅂ^{m}$	$ㄷ^{i}, ㄷ^{t}, ㄷ^{m}$	$ㅈ^{i}, ㅈ^{m}$	ㅅ	$ㄱ^{i}, ㄱ^{t}, ㄱ^{m}$	
긴 장	ㅃ	ㄸ	ㅉ	ㅆ	ㄲ	
대 기	ㅍ	ㅌ	ㅊ		ㅋ	ㅎ
비 강	ㅁ	ㄴ			ㅇ	
설 측		$ㄹ^{r}, ㄹ^{l}$				

($()^{i}$: 사용 위치 어두, $()^{t}$: 사용 위치 어말, $()^{m}$: 사용위치 어중, $ㄹ^{r}$: 'r' 발음되는 'ㄹ', $ㄹ^{l}$: 'l' 발음되는 'ㄹ')

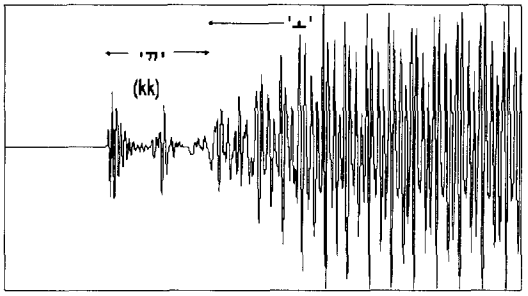
3. 파열음 인식을 위한 음소 인식단위

파열음은 조음점이 양순('ㅂ', 'ㅃ', 'ㅍ'), 치경('ㄷ', 'ㄸ', 'ㅌ'), 연구개('ㄱ', 'ㄲ', 'ㅋ')이고, 조음방법으로는 이완('ㅂ', 'ㄷ', 'ㄱ'), 긴장('ㅃ', 'ㄸ', 'ㄲ'), 대기('ㅍ', 'ㅌ', 'ㅋ')인 자음 음소들을 말한다. 이들 음소들이 어두에서

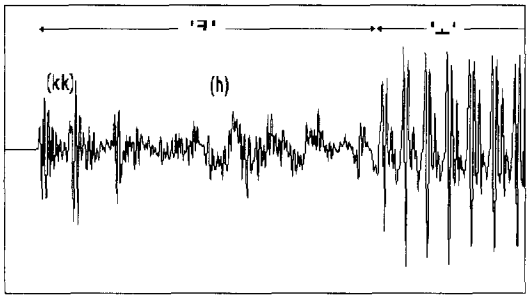
사용될 경우에 이들 조음상의 특징을 살펴보면, 이완 조음 방법에 해당하는 음소들은 짧은 파열음 뒤에 비교적 약한 대기음(aspiration)구간이 뒤에 나타나며, 긴장 조음 방법에 해당하는 음소는 짧은 파열음만이 존재하며 대기음 구간은 존재하지 않는다. 또한 대기 조음 방법에 해당하는 음소들은 마찬가지로 긴장 조음에 가까운 짧은 파열음 뒤에 대기음 구간이 존재하며 비교적 강한 대기음의 형태로 나타난다.



(a)



(b)



(c)

그림 1. 연구개 조음의 예 - (a) 이완계열 (b) 긴장계열 (c) 대기계열

Fig. 1. Examples velar articulation. - (a) lax (b) tension (c) aspiration

그림 1에서 파열음 중 연구개에 해당하는 'ㄱ', 'ㄲ', 'ㅋ'의 음소의 예를 보였다. 즉 파열음의 조음 특성은 그림 1의 예에서와 같이 이완 조음의 경우는 이완파열음 + 대기음, 긴장 조음의 경우는 긴장 파열음, 대기 조음의 경우는 유사 긴장 파열음(긴장도가 긴장 조음에 가까움) + 강한 대기음 들로 조음되는 특성을 갖는다.

즉 파열음의 조음 특성은 그림 1의 예에서와 같이 이완 조음의 경우는 이완파열음 + 대기음, 긴장 조음의 경우는 대기음이 따르지 않는 긴장 파열음만으로 조음되며, 대기 조음의 경우는 유사 긴장 파열음(긴장도가 긴장 조음에 가까움) + 강한 대기음 들로 조음되는 특성을 갖는다.

또한 이들 파열음들은 전설 평순 모음('ㅣ', 'ㅔ')들의 앞에서는 그림 2에서와 같이 조음되며 이것은 그림 1에서의 원순 후설 모음('ㅏ', 'ㅓ', 'ㅗ', 'ㅜ', 'ㅡ')들 앞에서 조음될 때와 뚜렷한 차이를 갖는다.

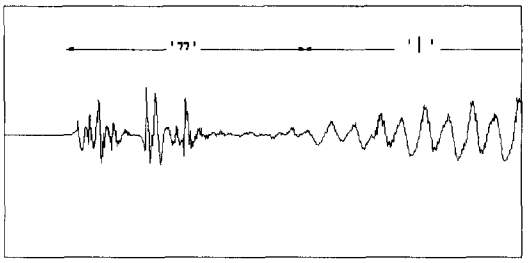


그림 2. 전설 평순 모음 앞에서의 파열음('ㄱ')의 예

Fig. 2. Example of plosive('ㄱ') prior to front-spread articulated vowels.

즉 전설 평순 모음 앞에서 조음 될 때는 원순 후설 모음 앞에서 조음될 때와는 조음 위치가 조금 달라지기 때문이다[6]. 따라서 이들 두 경우는 별도의 인식 단위로서 취급하는 것이 인식율 향상에 도움이 된다. 따라서 이상의 사항들을 정리해 보면 표 5와 같은 결과를 얻는다.

표 5. 파열음에 대한 인식 단위로서의 음소 구성

Table 5. The composition of phonetic-like unit for explosive phonemes.

조음점 / 조음 방법	양 순		치 경		연 구 개	
	원순 모음 앞	전설 평순 모음 앞	원순 모음 앞	전설평순 모음 앞	원순 모음 앞	전설 평순 모음 앞
단 순	p+h(ㅂ)	p*+h(ㅂ)	t+h(ㄷ)	t*+h(ㄷ)	k+h(ㄱ)	k*+h(ㄱ)
긴 장	pp(ㅃ)	pp* (ㅃ)	tt(ㄸ)	tt*(ㄸ)	kk(ㄲ)	kk*(ㄲ)
대 기	pp+h(ㅍ)	pp*+h(ㅍ)	tt+h(ㅌ)	tt*+h(ㅌ)	kk+h(ㅋ)	kk*+h(ㅋ)

표 5에서 ( )*는 전설 평순 모음 앞에서 조음되는 경우를 말하며 따라서 한 계열의 경우, 예를 들면 양순 계열의 파열음에서는 p, pp, p*, pp*, h 들의 5가지 형태의 인식 단위로서의 음소를 인식하여 조음 방법에 따른 음소의 조음 특성을 감안하면, 표 5에서의 조합에 의하여 음소를 분류해 낼 수 있다.

Ⅲ. 음소 대표구간 추출[7]

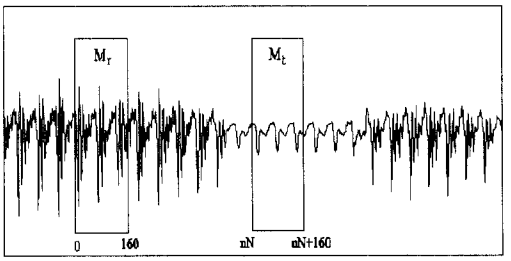
음소의 경계 추출은 Basseville, Andre-Obrecht들에 의하여 여러 가지 알고리듬들이 연구되었다. 주로 통계적인 접근방법을 사용한 이들 경계추출 방법은 Basseville와 Benveniste의 divergence test 방법[8]과 divergence test 방법의 문제점을 보완한 Andre-Obrecht의 forward-backward 방법[9]등이 대표적이라 할 수 있다. 이들 방법에서는 보다 정확한 음소의 경계점을 추출하기 위하여 음성신호의 블록단위 처리보다는 샘플단위의 처리를 하고있으나 계산량이 과다하고 일부 음소들에 대하여는 음소의 경계를 잘 찾지 못하는 단점들을 갖고 있다. 이들 방법 이외에도 조정호들이 제안한 방법[10]도 있다.

한 음소에서 다른 음소로 천이 되는 구간에서는 주파수 특성의 변화가 크게 두 가지 다른 형태로 나타난다. 즉 모음과 모음 혹은 모음과 유성자음 사이에서는 변화가 서서히 이루어지며, 무성자음과 모음 사이에서는 급격히 변화가 이루어진다. 경계부근의 천이구간에서의 음성샘플들은 음소인식을 위하여 별다른 의미를 갖지 못한다. 왜냐하면 이러한 천이구간의 음성샘플은 상호 조음현상에 의하여 인접음소의 영향을 받으므로 인접음소에 따라 특성이 달라지기 때문이다. 따라서 이러한 음소의 경계를 정확히 찾는 것보다는 특성의 변화가 적은 안정된 대표구간을 찾는 것이 더 효율적이라 할 수 있다.

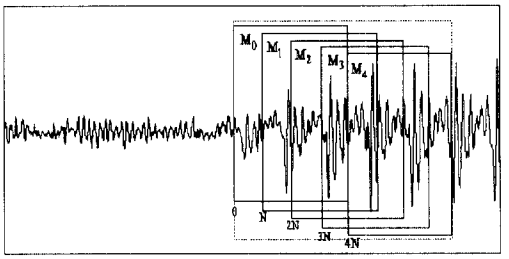
우선 단기간 음성특징 벡터의 변화 특성을 이용하여 하나의 대표 구간을 찾은 다음, 이를 기준구간으로 하여, 다음 음소의 대표구간을 찾는다. 이 과정과 병행하여 단기간 음성특징 벡터가 급격히 변화하는 구간을 추출하여 이를 음소간의 천이구간 혹은 짧은 음소구간으로 간주하게 된다. 음성 특징 벡터로는 비교적 계산이 간단한 켑스트랄 계수를 사용하였고, 특징의 변화는 기준구간과의 가중 켑스트랄 거리를 이용하여 구하였다.

그림 3-(a)에서와 같이 모델 M_r의 위치를 결정하고, 모델 M_r를 N 샘플만큼 씩 이동시켜 가면서 모델 M_r와 M_t의 거리를 측정한다. 모델 M_r과 M_t의 거리를 $d_{long\ term} = d(M_r, M_t)$ 라 하고, $d_{long\ term} > \lambda_0$이면 다른 음소로 바뀐 것으로 한다. 이 장기간시험은 주파수 특성의 변화가 매우 천천히 일어나는 모음과 모음 혹은 모음과 유성자음들 사이에서의 특성변화를 찾기 위한 것이다.

또한 그림 3-(b)와 같이 짧은 구간을 N 샘플씩 이동하면서 구간 내에서 모델 M_0와 nN 샘플 거리의 모델 M_n들과의 거리를 각각 구한다. 이들의 평균 거리와 최대거리를 $d_{mean(st)}$, $d_{\max(st)}$ 이라 하고 $d_{mean(st)} < \lambda_1$이면, 그 구간을 안정된 주파수 특성을 갖는 음소의 대표구간으로 간주하고, $d_{\max(st)} > \lambda_2$ 이면, 주파수 특성의 변화가 심한 음소의 천이구간 혹은 무성자음 구간으로 간주하게 된다. 이 단기간시험은 주파수 특성의 변화가 급격히 일어나는 음소 천이구간 혹은 초성, 종성에서의 폐쇄음구간들을 찾고, 긴 구간동안 비교적 동일한 주파수 특성이 계속되는 모음, 유성자음들에서의 대표구간을 찾기 위한 것이다.



(a) 장기간시험
(a) Long term test



(b) 단기간시험
(b) Short term test

그림 3. 음소 대표구간 추출을 위한 모델 위치
Fig. 3. Model position to extract frames representing phonemes.

Ⅳ. 신경망의 구성

신경망의 구성은 학습 시간을 줄이고 인식율을 높이기 위하여 여러 가지 제약 조건에 따라 각 신경망에서의 인식 대상을 구분하여 계층화 하였다. 즉 상층부는 조음점에 따라 양순음('ㅂ', 'ㅃ', 'ㅍ'), 치경음('ㄷ', 'ㄸ', 'ㅌ'), 연구개음('ㄱ', 'ㄲ', 'ㅋ') 및 성문음('ㅎ')들의 계열별로 구분하는 신경망을 구성한 뒤, 하층부에서 상층부 결과에 따라 각 조음 계열에 따른 음소를 구분하는 구조를 갖도록 하였다. 어두 음소를 위한 복합 신경망은 그림 4와 같다.

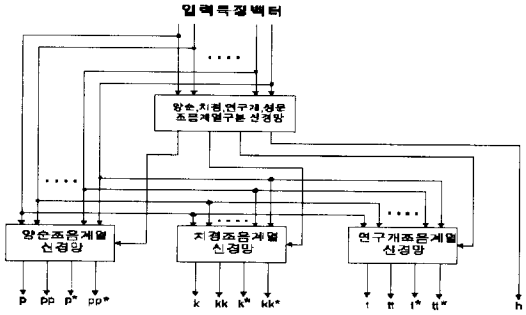


그림 4. 파열음 인식을 위한 복합 신경망의 구성도
Fig. 4. Block diagram of hierarchical neural network for recognition of plosive phonemes.

2. 신경망의 학습 및 인식

파열음 인식을 위하여 그림 4에서와 같이 복합 신경망을 구축하였고, 이 복합 신경망의 기본이 되는 단위 신경망은 그림 5와 같이 가장 기본적인 형태의 다층 피드포워드 신경망을 사용하였다.

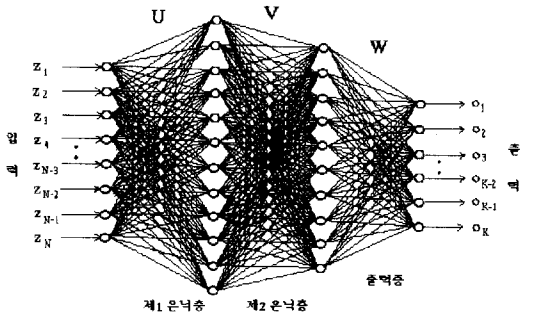


그림 5. 3계층 피드포워드 신경망
Fig. 5. 3-layered feed-forward network.

여기에서 입력은 18차의 켑스트랄 계수와 정규화 에너지, 정규화 영교차율을 위하여 입력 노드를 20개로 하였고, 출력 노드는 인식 대상 음소의 수에 따라 4개로 하였다. 제1 은닉층과 제2 은닉층의 노드 수는 40개와 8개로 하였다.

즉 두 개의 은닉층을 갖는 3계층 피드포워드 신경망으로 되어 있고 다음과 같은 과정의 오차 역전파 알고리듬에 의하여 학습이 이루어지도록 한다. 학습 과정을 통하여 수렴 속도를 빠르게 하기 위하여 모멘텀 방법(momentum method)을 사용하였다. 모멘텀 방법은 가중치의 조정시 바로 이전의 가중치 조정의 일부분을 다음 식과 같이 포함시키는 것이다.

$$\Delta w(t) = -\eta \nabla E(t) + \alpha \Delta w(t-1)$$

모멘텀율 α는 보통 0.1~0.8 사이에서 선택된다.

Ⅴ. 실험 결과 및 고찰

1. 학습 및 시험용 음성 데이터 베이스

학습에 사용된 음성 특징 벡터들은 실내 환경에서 성인 남자 3명에 의하여 발음된 한국어 문장들로부터 추출되었다. 사용된 문장은 국민교육헌장의 문장을 사용하였고, 사용 빈도가 적은 음소들의 경우수의 확보를 위하여 여러 가지 어휘들과 음절들을 추가하였다. 발음 속도는 초당 평균 약 3.3자(음절)의 속도를 갖도록 하였다. 음성 데이터의 취득은 PC에 장착된 사운드 카드를 이용하여 실내 환경에서 직접 마이크로 음성 데이터를 취득하였다. 샘플링 주파수는 8㎑, A/D변환은 16 bit로 선형 변환한 것을 하드 디스크에 저장하였다. 이들 발음된 문장과 어휘들 중에서 시험용 음성 특징 벡터의 추출에 사용된 450개의 음소를 제외하고 총 1420개의 음소들 중에서 비교적 특징이 뚜렷한 음소들을 선택하여 인식 대상 음소당 200~300개의 프레임을 선택하여 이들 프레임들로부터 특징벡터를 추출하였다. 따라서 인식 대상 음소당 200~300개의 특징 벡터들이 최종적으로 신경망의 학습에 사용되었다.

신경망의 학습은 모멘텀율 $\alpha=0.7$, 학습율 $\eta=0.3$을 선택하였고, 최대 오차는 0.001 이하가 되도록 하였다. 충분한 수렴을 위하여 학습 반복회수는 최대 10,000회까지로 하였다.

2. 인식 실험 결과

신경망에서의 인식 시험은 학습용으로 사용되지 않은 음성에 속해 있는 음소들을 20~100 프레임 선택하여

행하였다. 그림 3에서와 같이 어두에 올 수 있는 파열음의 인식을 위하여 조음 계열 별로 구분하기 위한 신경망과 각 계열에 속해 있는 음소의 구분을 위한 신경망의 계층 구조로 구성되어 있다. 전체적으로 파열음에 대하여 85.6%의 인식율을 얻었고, 이 결과를 계층을 구성하고 있는 각 신경망에 별로 구분하여 보면, 양순, 긴장, 대기, 성문의 조음계열별로 분류하는 상층 신경망에서는 전체적으로 95.7%의 높은 인식 결과를 얻었고, 하층조음계열별로는 조음 특성을 고려하는 후처리 과정까지를 고려하면, 양순 조음 계열 음소들('ㅂ', 'ㅃ', 'ㅍ')의 인식을 위한 신경망의 경우는 89.5%의 인식 결과를 보였고, 치경 조음 계열 음소들('ㄷ', 'ㄸ', 'ㅌ')의 인식을 위한 신경망의 경우는 92.7%의 인식 결과를 보였으며, 연구개 조음 계열의 음소들('ㄱ', 'ㄲ', 'ㅋ')의 인식을 위한 신경망에서는 86.0%의 인식 결과를 보였다. 표 6에 전체적인 인식 결과를 정리하였고, 표 7, 8, 9에 각 조음 계열별 인식결과를 자세히 보였다.

표 6. 파열음의 전체적인 인식율
Table 6. Total recognition rate for explosive phonemes.

		시험수	인식 시험 결과			
			양순	치경	연구개	성문
입력음소	양순	105	97	3	1	4
	치경	110	7	98	5	
	연구개	107	1	3	103	
	성문	60	2			58

위의 결과에서 양순 조음 계열 음소들과 치경 조음 계열 음소들이 상대적으로 낮은 분류율을 보였다. 이는 양순 계열의 음소들과 치경 계열의 음소들이 거리 척도상에 서로 인접하기 때문에 생기는 현상으로 보인다.

표 7. 양순 조음 계열 음소들의 인식 결과
Table 7. Recognition result for bilabial series phonemes.

		시험수	인식 시험 결과			
			p	pp	p*	pp*
입력음소	p	30	27	2		1
	pp	30	4	25		1
	p*	20	1		18	1
	pp*	25		1	3	21

위의 결과를 살펴보면 상대적으로 pp, pp*의 인식율이 낮은 것을 관찰할 수 있다. 이것은 대기음 'ㅍ'의 초기 파열음이 'ㅂ'과 유사한 경우가 있기 때문에 이러한 결과가 얻어진 것이다. 전체적인 인식율은 86.7%이나 조음 특성을 고려한 후처리 과정에서 'p'로 인식된 것들 중 뒤에 대기음을 수반치 않는 경우는 긴장음 'ㅃ'으로 수정 인식하므로서 긴장음 'ㅃ'에 대하여 1번의 수정이 이루어졌고, 위에서 'pp'와 'pp*'의 경우와 'p'와 'p*'의 경우는 오인식이 최종 음소 결정에 오류를 가져오지 않으므로 전체적인 인식율은 89.5%를 보였다.

표 8. 치경 조음 계열 음소들의 인식 결과
Table 8. Recognition result for alveolar series phonemes.

		시험수	인식 시험 결과			
			t	tt	t*	tt*
입력음소	t	30	29	1		
	tt	30	2	26	1	1
	t*	25	1		23	1
	tt*	25		1	3	21

위의 결과를 살펴 보면 상대적으로 인식율이 높은 것을 관찰할 수 있다. 이것은 대기음 'ㅌ'의 초기 파열음이 'ㄷ'의 파열음과 특성의 차이를 보이고 있음을 보여주고 있다. 전체적인 인식율은 90.0%이나 조음 특성을 고려한 후처리 과정에서 't'로 인식된 것들 중 뒤에 대기음을 수반치 않는 경우는 긴장음 'ㄸ'으로 수정 인식되었고, 표에서 'tt'와 'tt*'의 경우와 't'와 't*'의 경우는 오인식이 최종 음소 결정에 오류를 가져오지 않으므로 전체적인 인식율은 92.7%를 보였다.

표 9. 연구개 조음 계열 음소들의 인식 결과
Table 9. Recognition result for velar series phonemes.

		시험수	인식 시험 결과			
			k	kk	k*	kk*
입력음소	k	30	25	4	1	
	kk	30	5	24		1
	k*	25	2		19	4
	kk*	22		1	3	18

위의 결과를 살펴보면 상대적으로 인식율이 앞의 3가지 계열들 중에 가장 낮은 것을 알 수 있다. 이것은 이완음 'ㄱ'의 초기 파열음과 대기음 'ㅋ'의 초기 파열음이

매우 유사하여 인식율이 저조한 것을 알 수 있었다. 전체적인 인식율은 80.4%이나 조음 특성을 고려한 후처리 과정에서 'k'로 인식된 것들 중 뒤에 대기음을 수반치 않는 2개는 긴장음 'ㄲ'으로 수정 인식되었고, 표에서 'kk'와 'kk*'의 경우와 'k'와 'k*'의 경우는 오인식이 최종 음소 결정에 오류를 가져오지 않으므로 이에 해당하는 4개가 오인식에서 제외되어 전체적인 인식율은 86.0%를 보였다.

VI. 결 론

본 논문에서는 인식 단위인 음소를 대표할 수 있는 프레임들을 추출하고 추출한 프레임들로부터 특징 벡터들을 구하여 파열음 계열의 음소들을 인식하기 위한 신경망의 입력으로 사용하였다. 파열음 계열의 음소들은 그 조음점에 따라 양순, 치경, 연구개 계열의 음소들로 구분되며, 각 계열은 다시 조음 방법에 따라 이완, 긴장, 대기들의 조음 계열로 분류된다. 이들 중 각 조음점에 따른 계열내의 음소들은 초기 파열음이 서로 매우 유사하여 이들의 분류가 매우 어려웠다. 본 논문에서는 이완, 대기 계열의 음소들이 초기 파열음 뒤에 대기음이 나타나는 조음 특성을 이용하여 파열음 인식 신경망에서 이들 음소 대신에 초기 파열음들을 인식케 하고 다음에 대기음의 존재 여부에 따라 음소를 결정하는 방법을 제안하였다. 또한 파열음들이 평순 전설 모음 앞에서는 원순 모음 앞에서의 경우와 달리 조음되는 것을 감안하여 인식율을 높이기 위하여 별도의 인식 단위의 음소로 취급하였다. 이렇게 하여 얻어진 신경망의 출력은 파열음 음소들이 가지는 조음 특성을 감안한 후처리 과정을 거침으로서 높은 인식율을 얻을 수 있었다.

인식 실험 과정을 통하여 파열음에 대하여 85.6%의 높은 인식율을 얻을 수 있었고, 이 인식율은 계열별 분류 신경망에서의 분류 결과와 계열 내에서의 인식 결과를 곱한 것으로서 계열별 분류 신경망에서의 계열 인식율은 95.7%의 아주 높은 인식율을 가지나 계열 내에서의 음소의 인식은 전체적으로는 89.7%의 인식율을 보이고 있다. 계열 내에서의 인식율이 더 낮은 이유는 계열 내에서의 이완, 긴장, 대기의 조음방법을 갖는 음소들이 서로 유사한 특성을 갖기 때문이다. 특히 연구개 계열의 음소들의 경우는 이러한 현상이 상대적으로 더 심해서 인식율이 가장 떨어지고 있다. 이러한 오류들은 상위 레벨에서의 전후 문맥 등을 이용한 처리과정을 통하여 상당 부분 수정될 수 있으리라 생각된다.

본 논문에서의 연구는 파열음이 어두에서 쓰이는 경우에 한정하고 있으나, 무성자음이 어말에서 쓰이는 경우('ㄱ', 'ㄷ', 'ㅂ')는 어두에서의 경우보다 상호 유사도가 훨씬 적어 보다 용이하게 인식이 가능한 것으로 알려져있고, 어중에서도 일부 유성음 사이에서 유성음 화하는 음소들을 제외하고는 어두에서와 같은 특성으로 조음되므로 인식에 크게 어려움이 없으리라 생각된다. 따라서 본 논문에서의 알고리듬에 무성자음 인식에 있어 또 하나의 어려움으로 알려져 있는 마찰, 파찰음 계열의 음소들에 대한 효율적인 인식 방법이 연구된다면, 우수한 성능의 음소단위의 연속 음성 인식 시스템의 구축이 가능해질 것이다.

참 고 문 헌

[1] N. Morgan and H. Bourlard, " Continuous Speech Recognition," *IEEE Signal Processing Magazine*, pp 25-42, May, 1995.

[2] G. Zavaliagkos, Y. Zhao, R. Schwartz, and J. Makhoul, " A Hybrid Segmental Neural Net / Hidden Markov Model System for Continuous Speech Recognition," *IEEE Trans. Speech and Audio Processing*, vol. 2, No. 1, Part II, pp 151-159, January, 1994.

[3] C. Dugast, L Devillers, and X. Aubert, " Combining TDNN and HMM in a Hybrid System for Improved Continuous-Speech Recognition," *IEEE Trans. Speech and Audio Processing*, vol. 2, No. 1, Part II, pp 217-223, January, 1994.

[4] 권영욱, 정현열, "다차원 척도 구성법을 이용한 한국어 음소의 분석," *전자공학회논문지*, 제 29권, B편 제 11호, pp 22-30, 1992년 11월

[5] 정연찬, *한국어 음운론*, 개문사, p127-151, 1980

[6] 박창배, *한국어 구조론 연구*, (주)탑출판사, p11-37, 1990

[7] 박찬응, 이쾌희,"연속음에서의 각 음소의 대표구간 추출에 관한 연구", *전자공학회논문지*, 제 33권 B편, 제 4호, 1996년 4월

[8] M. Basseville and A. Benveniste, spectral characteristics of digital signals," *IEEE Trans. Inform. Theory*, vol. IT-29, pp.709-723, Sept. 1983.

[9] R. Andre-Obrecht, "A new statistical approach for the automatic segmentation of continuous speech signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.36, No.1, Jan. 1988.

[10] 조정호, 홍재근, 김수중, "통계적인 방법에 의한 연결음의 음소분할 알고리듬," *전자공학회논문지*, 제 26권, 제 4호, pp.151-163, 1989년 4월

저 자 소 개

朴 贊 應(正會員)

1950년 8월 4일생. 1977년 8월 서강대학교 전자공학과(공학사). 1989년 8월 서강대학교 대학원 전자공학과(공학석사). 1997년 2월 서강대학교 대학원 전자공학과(공학박사). 1984년 9월 ~ 1992년 2월 대우통신(주) 종합연구소 수석연구원. 1995년 3월 ~ 현재 인덕전문대학 방송통신과 조교수. 주관심분야는 음성 및 영상신호처리, 통신시스템 등임

李 夫 熙(正會員) 第 33卷 B編 第 4號 參照