

論文94-31B-4-20

음소 특징파라미터를 이용한 무성자음 인식

(The Recognition of Unvoiced Consonants Using Characteristic Parameters of the Phonemes)

許萬鐸*, 李鐘赫**, 南基坤***, 尹台焘***, 金在昌***, 李梁成***

(Man Tak Huh, Jong Hyeok Lee, Ki Gon Nam, Tae Hoon Yoon, Jae Chang Kim and Yang Sung Lee)

要約

본 연구에서는 음절을 구성하고 있는 음소의 특징파라미터를 이용하여 무성자음을 인식하는 시스템을 제안하였다. 인식을 위해 자음구간의 영교차 율, 전구간 에너지, 반구간 에너지와 같은 시간영역 상의 특징과 천이구간의 주파수 스펙트럼인 주파수영역상의 특징을 특징파라미터로 사용하였다. 인식 대상의 무성자음은 /ㄱ/, /ㄷ/, /ㅈ/, /ㅊ/, /ㅋ/, /ㆁ/, /ㅍ/, /ㅌ/ 이며 이들의 인식을 위해 분할된 각 음소를 분석하여 구한 두 구간의 특징파라미터를 각 구간의 인식 시스템에서 독립적으로 처리하고 두 시스템의 출력을 상호 보완하여 최종적인 인식 결과를 출력하도록 하였다. 인식 시스템은 학습능력을 가진 다층신경망을 사용하여 구성하였다. 제안된 인식 시스템을 이용하여 112 가지 무성자음 데이터를 처리한 결과, 평균인식율은 학습율 60 %에서 93.7 %, 학습율 80 %에서 96.4 %이었다.

Abstract

In this study, we present unvoiced consonant recognition system using characteristic parameters of the phoneme of the each syllable. For the recognition, the characteristic parameters on the time domain such as ZCR, total energy of the consonant region and half region energy of the consonant region, and those on the frequency domain such as the frequency spectrum of the transition region are used. The objective unvoiced consonants in this study are /ㄱ/, /ㄷ/, /ㅈ/, /ㅊ/, /ㅋ/, /ㆁ/, /ㅍ/ and /ㅌ/. Each characteristic parameter of two regions extracted from these segmented unvoiced consonants are used for each recognition system of the region, independently. And, complementing two outputs of each other system, the final output is to be produced. The recognition system is implemented using MLP which has learning ability. The recognition simulation results for 112 unvoiced consonant samples are that average recognition rates are 96.4 % under 80 % learning rates and 93.7 % under 60 % learning rates.

* 正會員, 釜山工業大學校 電子工學科
(Dept. of Elec., Pusan Nat'l Ins. of Tech.)

** 正會員, 慶星大學校 컴퓨터工學科
(Dept. of Computer., Kyungsung Univ.)

*** 正會員, 釜山大學校 電子工學科
(Dept. of Elec., Pusan Nat'l Univ.)

接受日字: 1993年 9月 7日

1. 서론

음성인식 및 음성합성 분야에서는 인간과 기계사이의 정보전달 수단으로써 음성을 사용하려는 노력이 계속되어 왔다. 음성인식은 특징파라미터를 추출하는 단위에 따라 음소단위, 음절단위 및 단어단위로 처리된다. 음절이나 단어단위에 의한 음성인식에서는 비교적 높은 인식율을 보이고 있지만 인식대상이 증가함에 따라 많은 기억용량이 요구되어, 연속음성을 처리하기에는 많은 어려움이 있다. 음소단위 인식에서는 인식해야 할 음소의 수가 적고 연속음성 인식에 적합하지만 음소패턴의 변형이 심하여 인식율이 낮기 때문에 실용화하기 위해서는 지속적인 연구가 요구된다.

음소단위 인식에서 모음의 경우에는 포만트주파수, 선형예측계수, 켈스트럼계수등을 주로 이용하며¹ 최근 스펙트럼의 포락선을 이용하여 모음인식을 함으로써 높은 인식율을 얻은 보고도 있으며² 모음인식 신경회로망을 하드웨어로 구성한 보고도 있다.³ 자음의 경우에서 유성자음은 피치를 가지므로 모음인식의 기법을 사용하여 높은 인식율을 얻을 수 있으나 무성자음은 유색잡음(colored noise)의 특성을 가지며 후속 모음의 영향을 많이 받기 때문에 주파수 영역에서의 특징 추출 방법만으로는 모음에서와 같은 높은 인식율을 얻기가 어렵다.^{4,5}

자음이 포함된 음성의 특징을 이용한 음성인식방법의 연구로는 HMM(Hidden Markov Model)⁶, DTW(Dynamic Time Warping)⁷, TDNN(Time Delay Neural Network)⁴, MLP(Multi Layer Perceptron)⁵ 등을 이용한 것이 있다. HMM은 음성의 발생확률을 이용하는 통계적인 방법이고 DTW는 기준패턴과 시퀀스패턴과의 거리를 계산하여 최소거리를 가지는 패턴을 선택하는 패턴정합 방법으로서, 음절내지 단어단위 인식에서는 좋은 인식율을 보이고 있으나 음소단위 인식에는 적합하지 않은 것으로 보고되고 있다.¹⁴ TDNN과 MLP는 인간의 신경망구조를 모델링한 신경회로망으로서 학습특성과 적응특성으로 인하여 음성인식 분야에서 현재 많이 연구하고 있는 방법이다. 하지만 TDNN은 시간의 추이에 따른 음성의 특징을 인식할 수 있는 장점을 가지지만 그 구조가 복잡하여 학습시간이 오래 걸릴 뿐만 아니라 하드웨어로 구성할 경우 많은 소자가 요구되는 단점을 갖고 있다. 최근까지 음소단위로 자음소를 인식한 연구보고는 제한된 조건하에서 자음소 인식율이 80% 내지 90% 정도의 결과를 보여주고 있으나⁸⁻¹⁰ 실용화하기에는 부족하므로 많은 연구가 요

구된다.

본 연구에서는 음소단위 무성자음 인식을 위해, 음성의 시간영역상의 특징과 주파수영역상의 특징을 둘 다 고려한 인식시스템을 제안한다. 무성자음 인식을 위해 자음구간의 시간영역상의 특징파라미터를 주로 사용하며, 또한 천이구간이 선행 자음의 특징을 가지고 있을 뿐만 아니라¹⁰ 후속 모음의 특징도 함께 가지고 있으므로 천이구간의 주파수영역의 특징파라미터도 함께 이용한다. 인식시스템은 먼저 인식대상의 단음절을 자음구간, 천이구간 및 모음구간으로 분할하고 각 구간별로 특징파라미터를 추출한다. 자음 및 천이구간의 특징파라미터는 모음구간의 특징파라미터를 처리한 모음인식망¹¹의 결과와 함께 자음인식망에 입력되어 인식과정에 이용된다. 자음구간의 특징파라미터는 분할된 자음구간의 전구간에너지, 반구간에너지¹² 및 영교차율을 이용하며 천이구간의 특징파라미터는 주파수스펙트럼을 이용한다. 천이구간의 주파수스펙트럼은 신경세포의 특성과 비슷한 ∇^2G 연산자를 이용한 interconnection 모델¹³을 이용하여 추출하며, 자음구간의 특징파라미터는 신호처리 기법을 사용하기로 한다. 끝으로, 제안한 무성자음 인식시스템에 112 개의 음성시료에 대하여 시뮬레이션하여 이를 토대로 인식결과를 검토하기로 한다.

II. 무성 자음의 특징파라미터 추출

각 음소구간의 특징파라미터를 추출하기 위해서 우선, /자음/ + /모음/ 형태의 단음절 음성에서 자음구간, 천이구간 및 모음구간의 결정이 필요하다. 본 연구에서 이들 구간을 결정하기 위하여 사용한 특징파라미터는 자음구간의 에너지, 피치의 존재유무 및 묵음구간의 평균에너지이다. 자음 중에서 무성자음은 그 파형이 잡음과 유사하며 후속모음에 비해 에너지가 아주 작지만 유성자음은 피치를 가지고 있을 뿐만 아니라 에너지도 크다. 그러므로 유성자음은 모음의 인식 기법을 이용하여 인식할 수 있으므로 본 연구에서는 표 1에 나타난 한국어 자음의 분류¹⁴ 상에서 평음과 격음에 속하는 8 개 무성자음(/ㄱ/, /ㄷ/, /ㅂ/, /ㅅ/, /ㅋ/, /ㅌ/, /ㅍ/, /ㅈ/)을 연구의 대상으로 한다.

1. 자음구간의 특징파라미터

8개 자음을 하나의 군으로 하여 인식처리하려면 인식대상의 중복성(redundancy)이 많기 때문에 인식율이 좋지 못하다. 그러므로 표 1의 자음의 분류표에

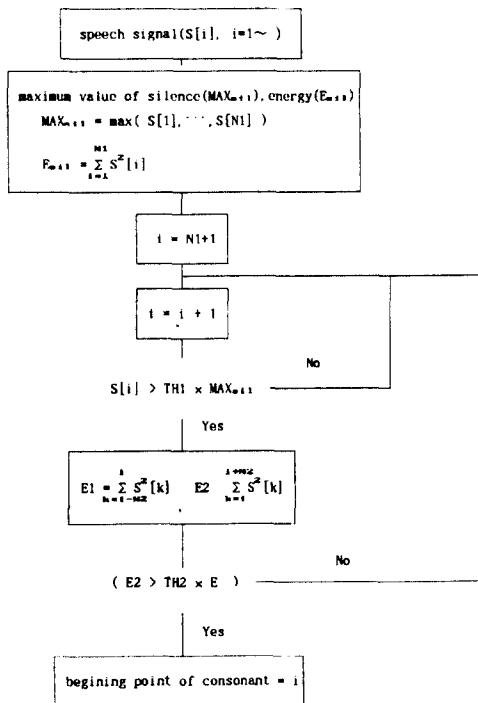
표 1. 한국어 자음의 분류

Table 1. Classification of the Korean consonants.

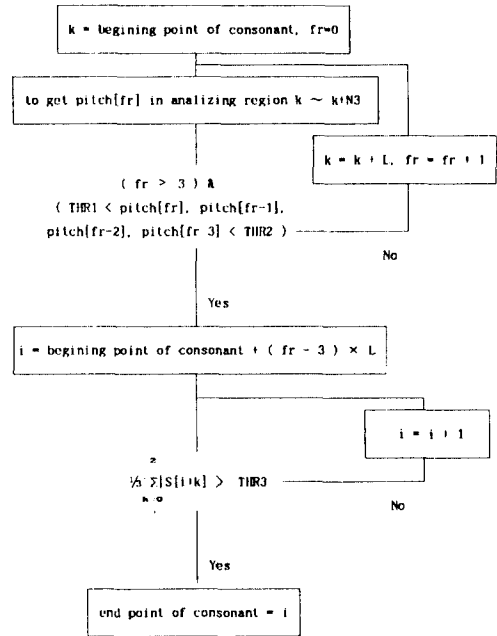
구별요인	양순성	전 설 성			후설성	후두성
		정지성	파찰성	마찰성		
병 음	ㅂ	ㄷ	ㅌ	ㅍ	ㄱ	
경 음	ㅃ	ㄸ	ㅊ	ㅍ	ㅋ	
격 음	ㅍ	ㅌ	ㅊ		ㅋ	ㅎ
비 음	ㅁ	ㄴ			ㅇ	
유 음		ㄹ				

나타난 것과 같이 평음과 격음의 군으로 나누어 인식할 필요가 있다. 피치패턴은 초성자음에서 후속모음으로 천이할 때 피치 주파수가 시간에 대해 어떤 기울기로 변화하는 것인가를 나타내는 특징파라미터로서 일반적으로 격음에서의 기울기가 평음에서의 기울기보다 크다. 본 연구에서는 피치패턴에 의해 평음군과 격음군으로 분류하고 분류된 자음을 인식하기로 한다.

신호대 잡음비가 그렇게 낮지 않은 환경에서는 목



(a)



(b)

그림 1. 자음구간 결정 알고리즘

(a) 자음의 시작점 결정

(b) 자음의 끝점 결정

Fig. 1. Algorithm of consonant region decision.

(a) Decision of beginning point of consonant.

(b) Decision of end point of consonant.

음구간과 자음구간의 구별은 신호의 크기에 의해 할 수 있다. 먼저 자음구간은 시간의 추이에 따라 음성의 각 샘플의 크기를 비교하여 임계값 이상이 되는 시점을 시작점으로 간주한다. 이때의 임계값은 묵음구간의 최대 신호값보다 큰 값으로 정한다. 그리고 자음구간의 시작점을 중심으로 일정한 구간의 앞쪽 프레임과 뒤쪽 프레임의 에너지를 비교하여 뒤쪽 프레임의 에너지가 클 경우 자음의 시작점으로 최종 결정한다. 음성신호는 자음에서 모음으로 변화하는 천이구간부터 피치를 가지기 시작하므로 자음구간의 시작점부터 시간의 추이에 따라 분석구간을 설정하여 피치를 구한다. 만약 연속적으로 피치가 존재할 경우, 피치가 존재하는 첫 분석구간의 시작점을 자음구간의 끝점으로 한다. 이때 분석구간만큼의 분할(segmentation)오차가 생길 수 있으므로 이 시점부터 시간의 추이에 따라 신호의 크기를 비교하여 임계

값 이상이 되는 점을 자음구간의 끝점으로 결정한다. 이상에서 설명한 자음구간 결정 알고리즘을 그림 1에 나타내었다.

자음구간을 결정한 후 이를 분석하여 자음구간의 특징파라미터로서 전구간에너지, 반구간에너지 및 영교차율을 구한다. 여기서 자음구간의 반구간에너지는 자음구간을 반으로 나누었을 때의 앞쪽 반구간동안의 에너지를 구한 것이다. 이것은 자음파형에서 특정한 자음들의 에너지가 앞쪽에 많이 몰려있는 것을 이용하고자 하는 것이다. 자음구간의 전구간에너지 E_{total} 과 반구간에너지 E_{half} 는 다음과 같다.

$$E_{total} = \sum_{n=CI}^{CE} S[n]^2 \quad (1)$$

$$E_{half} = \sum_{n=CI}^{(CI+CE)/2} S[n]^2 \quad (2)$$

여기서 CI와 CE는 자음구간의 시작점과 끝점의 표본수이다.

영교차율은 다음 식과 같이 정의한다.

$$ZCR = \#(S[n] \cdot S[n+1] < 0), \quad n = CI, CI+1, \dots, CE \quad (3)$$

2. 천이구간의 특징파라미터 추출

천이구간은 음성신호가 자음에서 모음으로 변화해 나가는 구간에서 후속모음과 비슷한 피치가 나타나기 시작하는 점부터 시간의 추이에 따라 안정된 후속모음의 피치형태를 가지게 되는 점까지의 구간을 말한다. 천이구간의 주파수스펙트럼은 후속모음의 영향을 받지만 선행자음의 주파수특징도 포함하고 있기 때문에 후속모음의 인식 결과에 따른 자음인식의 특징파라미터로 유용하다.

천이구간의 스펙트럼은 ∇^2G 연산자를 이용하여 시간영역에서 추출할 수 있다. 먼저, 신경회로망으로 구성하기에 편리한 ∇^2G 연산자의 특성에 대하여 기술하고 이를 이용한 천이구간의 주파수스펙트럼 추출에 대해서 논한다. 음성신호에서 특정한 주파수의 신호만을 추출하기 위하여는 연산자가 대역통과 특성을 가져야 한다. 그러나 첨예한 대역통과특성을 가지기 위해서는 시간영역에서의 연산자가 커지게 된다. 가우스 함수에서는 시간영역에서의 연산자 크기와 주파수 영역에서의 대역폭과의 곱이 작기 때문에¹² 이 두가지 조건이 적절히 만족될 수 있다. 1차원 가우스 함수의 2차 미분은

$$-\nabla^2 G(t) = G''(t) = \frac{1}{\sqrt{2\pi}\sigma^3} \left(1 - \frac{t^2}{\sigma^2} \right) \exp\left(-\frac{t^2}{2\sigma^2} \right) \quad (4)$$

과 같이 표시되며, 이의 푸리에 변환을 $LoG(\omega)$ 라 하면

$$LoG(\omega) = \omega^2 \exp\left(-\frac{\sigma^2 \omega^2}{2} \right) \quad (5)$$

로 표시되며, 중심주파수 ω_0 와 진폭 $A(\omega_0)$ 는

$$\omega_0 = \frac{\sqrt{2}}{\sigma} \quad (6)$$

$$A(\omega_0) = LoG(\omega_0) = \frac{2}{\sigma^2} \exp(-1) \quad (7)$$

이 된다. 여기서 중심주파수가 ω_0 인 신호 성분을 찾기 위한 ∇^2G 연산자의 표준편차 σ 는 식 (6)에서 다음과 같이 유도된다.

$$\sigma = \frac{\sqrt{2}}{2\pi \omega_0} \quad (8)$$

음성 신호 $s(t)$ 의 푸리에 변환을 $S(\omega)$ 라고 하면, 이 신호에서 특정한 주파수 신호는 대역통과 특성을 가지는 ∇^2G 연산자를 통과시킴으로써 추출되고 이들 관계는 다음과 같다.

$$S(\omega) \cdot LoG(\omega) \Leftrightarrow s(t) * \nabla^2G(t) \quad (9)$$

여기서 '*'은 콘볼루션 연산기호이다. 식 (9)에 의하여, 주파수 영역에서의 특정한 주파수의 검출은 시간영역에서 입력신호 $s(t)$ 와 ∇^2G 연산자와의 콘볼루션 결과로써 구할 수 있다. 그리고 식 (9)에서 $LoG(\omega)$ 를 정규화하면, 입력 신호의 각 주파수에 대한 진폭특성은 식 (9)의 우변에서 콘볼루션의 최대값으로 대체할 수 있다.

주파수 분석을 위한 filter bank의 중심주파수를 결정하는 방법으로는 FFT분석 방법과 같이 중심주파수를 균등 간격으로 설정하는 방법과 비균등 간격으로 설정하는 방법이 있지만 음성신호처리에서는 귀가 높은 주파수에서는 감도가 떨어지는 점을 감안하여 후자의 방법을 많이 택한다. 비균등 간격으로 중심주파수를 설정하는 방법으로 여러가지가 제안되었다. 그러나 주파수 간격당 청각세포의 분포는 800-1000Hz 범위에서 가장 많이 분포되어 있으며 이 범위를 중심으로 저주파 부분보다는 고주파부분에 적게 분포되어 있다. 본 연구에서는 이와 같은 청각계의 특성을 감안하여 각 중심주파수를 결정하였으며 filter bank의 채널 수는 40으로 하였다. 이상과 같은 방법을 ∇^2G 연산자를 이용한 interconnection

[13] 모델로서 스펙트럼을 구하고 천이구간의 특징과 라미터로 취한다.

Ⅲ. 자음인식 시스템의 구현

전체적인 자음인식을 위하여 학습특성과 적응특성을 가지고 있는 다층 신경망을 인식기로 사용한다. 본 연구에서 구성한 자음인식 시스템은 그림 2와 같으며 그 동작은 다음과 같다. 먼저, 입력음성에 대해 유성음, 무성음 결정 블록에서 자음, 모음 및 천이구

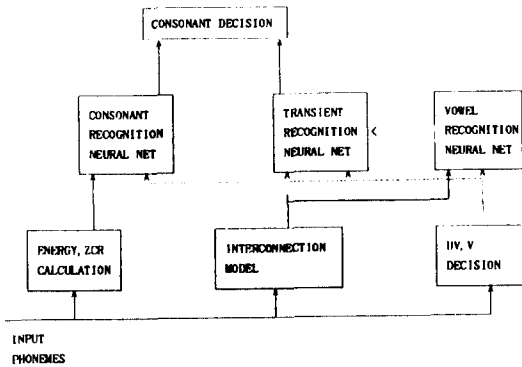


그림 2. 자음인식 시스템
Fig. 2. Consonant recognition system.

간의 구분을 짓는다. 자음구간의 경우에는 신호처리 기법에 의해 시간영역의 특징을 구하여 자음구간 신경망의 입력으로 사용한다. 천이구간이나 모음구간일 경우에는 interconnection 모델을 이용하여 주파수를 분석하고 그 분석 결과는 천이구간 신경망과 모음 인식 신경망의 입력으로 사용된다. 모음인식 신경망

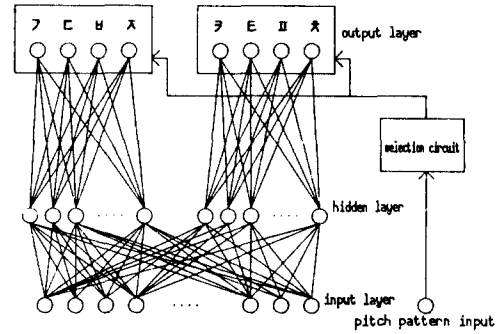


그림 3. 자음구간에 대한 인식 신경망
Fig. 3. The recognition neural network for the consonant region.

의 출력에 의해 입력음성의 모음이 결정되고 이 모음에 해당되는 천이구간 신경망이 활성화된다. 자음구간 신경망의 활성화값과 천이구간 신경망의 활성화값을 자음결정 블록에 입력하여 정해진 알고리즘에 의해 입력된 자음을 인식하도록 한다. 자음결정 블록에서는 각 자음에 대응하는 자음구간 신경망의 활성화값과 천이구간 신경망의 활성화값을 더하여 가장 큰 값을 갖는 뉴런에 해당되는 자음을 최종적인 인식 자음으로 결정한다. 자음인식 시스템에서 시간영역상의 특징과 라미터인 자음구간의 전구간 에너지, 반구간 에너지, 영교차율을 이용하여 자음을 인식하는 자음구간 신경망은 그림 3과 같다. 자음구간의 특징과라미터가 입력층에 입력되고 피치패턴이 선택회로에 입력되면 먼저 평음군과 격음군으로 구분된다. 선택된 군 중에서 입력된 임의의 자음에 가장 가까운 출력뉴런의 활성화값이 가장 크게 된다. 천이구간의 파형은 선

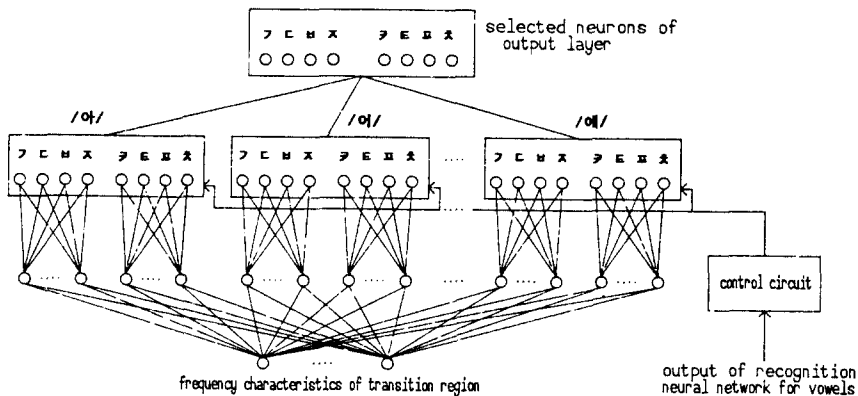


그림 4. 천이구간에 대한 인식 신경망
Fig. 4. The recognition neural network for the transition region.

행자음 뿐만 아니라 후속 모음의 영향도 받기 때문에, 자음 뒤의 후속 모음 7 개(/ㅏ/, /ㅑ/, /ㅓ/, /ㅕ/, /ㅗ/, /ㅛ/, /ㅜ/, /ㅠ/)에 대해서 각 모음에 따른 천이 구간의 주파수 분석결과가 그 모음에 해당하는 천이 구간 신경망에 입력되도록 하였고 따라서 천이구간 신경망을 그림 4에 나타낸 바와 같이 그림 3과 같은 구조가 7 개인 신경망으로 구성한다.

IV. 인식 시뮬레이션

인식 시뮬레이션에 사용된 데이터는 20대 남자 2 명이 발성한 자음 8 개(/ㄱ/, /ㄷ/, /ㅌ/, /ㅈ/, /ㅊ/, /ㅋ/, /ㅍ/, /ㅍ/)와 후속모음 7 개 (/ㅏ/, /ㅑ/, /ㅓ/, /ㅕ/, /ㅗ/, /ㅛ/, /ㅜ/, /ㅠ/)로 조합된 112 개이다. 앞에서 기술한 방법으로 음성신호를 분할하였으며 그 결과 목적과 잘 일치하였다. 음성신호 /가/를 분할한 결과를 그림 5에 나타내었다.

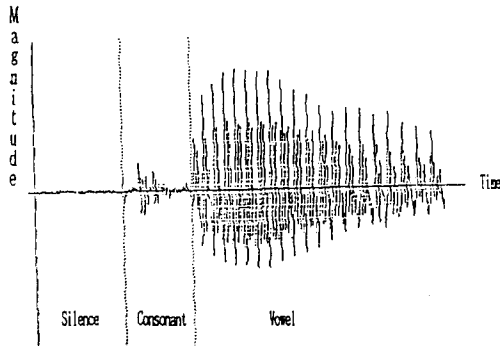


그림 5. 음절 /가/의 음소분할 결과
Fig. 5. The result of phoneme segmentation for syllable /ga/.

분할된 자음구간에 대하여 추출한 특징파라미터인 자음구간의 전구간 에너지, 반구간 에너지 및 영교차율을 인식 신경망의 입력으로 사용하기 위해서 이들 특징파라미터 값을 각각 0~99사이의 십진 정수값으로 하였으며 이때 시도한 코드변환 방법은 binary 코드, 5 bit shift 코드, 6 bit shift 코드, Gray 코드, BCD 코드등이다. 그리고 신경망에 천이구간의 정보를 학습시키기 위해 interconnection 모델에서 구한주파수스펙트럼을 코딩하여 입력패턴으로 만들었다. 이때 시도한 코딩방법은 interconnection 모델의 출력신호에서 평균을 구한 후 평균값보다 큰 경우면 "1"로 하는 평균치 방법과, 포락선을 구한 후 증

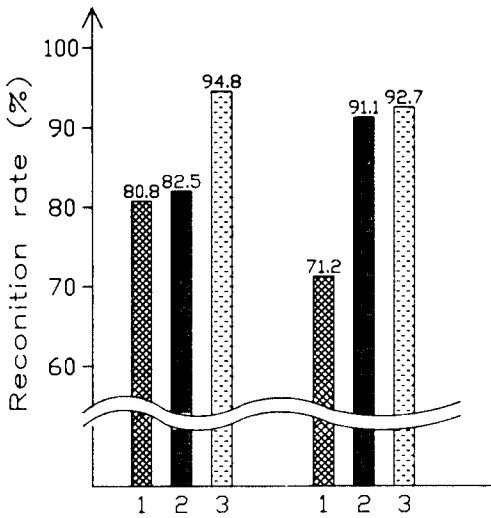
가하는 부분만 "1"로 하는 경사치 방법이다

표 2. 자음구간 상의 특징 파라미터를 사용한 경우의 인식율

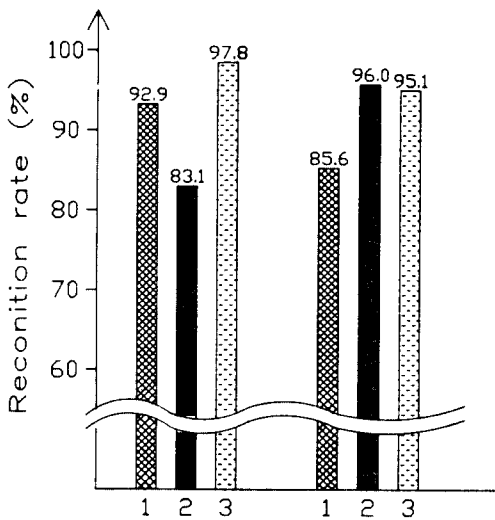
Table 2. Recognition rates using characteristic parameter of the consonant region.

히든 수	코딩방법	인식율(격음)		인식율(평음)	
		학습율 60%	학습율 80%	학습율 60%	학습율 80%
10	binary	67	83	76	90
	5bit shift	70	83	81	89
	6bit shift	67	85	83	92
	Gray	76	89	78	90
	BCD	70	81	80	94
20	binary	65	81	76	90
	5bit shift	74	83	87	90
	6bit shift	65	83	83	96
	Gray	76	89	76	89
	BCD	70	80	80	90
30	binary	67	81	74	89
	5bit shift	72	81	87	92
	6bit shift	67	83	85	92
	Gray	78	87	76	87
	BCD	70	81	80	90

표 2에 음성 데이터의 자음구간 특징파라미터를 자음구간 신경망에 입력하여 조건을 달리 하면서 인식 실험한 결과를 나타내었다. 전체적인 측면에서 고려해 볼 때 격음보다 평음에서의 인식율이 높았으며, 여러가지 코딩 방법 중 5bit shift 코딩 방법에 의한 것이 인식율이 좋았다. 신경망의 은닉층의 노드 수를 10개, 20개 및 30개로 하여 실험해 본 결과 인식율에 큰 차이는 없었지만 코스트와 성능면에서 비교해 볼 때 20개일 경우가 가장 적당하였다. 이들 선택된 조건을 취하여 인식 시뮬레이션한 결과, 학습율 60%에서 평균군은 87%, 격음군은 74%의 인식율을, 학습율 80%에서 평균군은 90%, 격음군은 83%의 인식율을 보여주었다. 천이구간의 주파수특징을 이용하여 인식율을 구하여 본 결과 학습율 60%에서 격음은 90%이상으로 좋은 인식율을 얻었다. 그러나 평음의 인식율은 평균치 방법이 경사치 방법보다 좋았지만 인식율이 82%정도로서 자음구간의 특징파라미터를 사용한 때보다 좋지 못하였다.



(a)



(b)

그림 6. 자음인식 시스템에서의 인식율
 1: 자음구간 이용 2: 천이구간 이용 3: 모두 이용 (a) 학습율 60%의 경우 (b) 학습율 80%의 경우

Fig. 6. Recognition rates of consonant recognition system.
 1: Using consonant region 2: Using transition region, 3: Using regions of both consonant and transition. (a) The case of 60% learning rates. (b) The case of 80% learning rates.

무성자음의 인식을 위해 자음구간 신경망만을 사용한 경우, 천이구간 신경망만을 사용한 경우 및 두 구간의 신경망을 같이 사용한 경우의 인식율을 전체적으로 그림 6에 나타내었다. 학습율 80%에서 천이구간 신경망만을 사용한 경우, 격음에 대한 인식율은 96.0%로서 두 구간의 신경망을 같이 사용한 방법보다 조금 높았지만, 그 이외의 경우는 두 구간의 신경망을 같이 사용한 방법이 각각의 신경망 하나만을 사용한 경우보다 인식율이 6.9% ~ 17.7% 높았다.

V. 결론

본 연구에서는 무성자음을 인식하기 위해 자음구간의 시간영역상의 특징과 천이구간의 주파수영역상의 특징을 모두 이용한 신경회로망 형태의 무성자음인식 시스템을 제안하였다. 자음구간의 시간영역상의 특징만을 이용한 자음구간 신경망을 이용하여 8개 자음을 인식 시뮬레이션한 결과, 격음보다 평음에서의 인식율이 높았으며, 여러가지 코딩 방법 중 5 bit shift 코딩 방법에 의한 것이 인식율이 좋았다. 자음구간과 천이구간의 특징파라미터를 따로 사용하였을 때보다 두 쪽 다 같이 사용하여 인식 시뮬레이션을 했을 때의 인식율이 6.9 ~ 17.7% 향상되었으며 그 결과, 112개의 무성자음 데이터에 대해 학습율 60%에서 평음이 94.8%, 격음이 92.7%이며 학습율 80%에서 평음이 97.8%, 격음이 95.1%였다. 이 결과는 본 연구에서 제안한 인식 시스템에 의한 음소단위 무성자음의 인식율이 개선된 것을 보여준다. 이 방법을 더욱 개선하여 연속음성 인식에 응용할 수 있도록 하여야 할 것이다.

參考文獻

[1] L.R. Rabiner, R.W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, 1978.
 [2] 이종혁, "▽²G 연산자에 의한 신호분석 및 음성인식을 위한 신경회로망의 구현," 박사학위논문, 부산대학교, 1991.
 [3] 채상훈, 전석보, 윤태훈, 김재창, "아날로그 입력방식에 의한 음성인식 신경회로망의 구현," 대한전자공학회 논문지-B, 제 29권, B편, 제 6호, pp. 80-85, 1992.
 [4] K. J. Lang and A. H. Waibel, "A Time-Delay Neural Network Architecture

- for Isolated Word Recognition, " *Neural Networks*, vol. 3, pp.23-43, 1990.
- [5] C. Wheddon and R.Linggard, *Speech and Language Processing*, Chapman and Hall, pp.291-304,1990.
- [6] A. Ljolje and S. E. Levinson, "Development of an Acoustic-Phonetic Hidden Markov Model for Continuous Speech Recognition," *IEEE Trans. Signal Processing*, vol. 39, no. 1, pp.29-39, 1991.
- [7] C. Myers, L.R. Labiner and A.E Rosenberg, "Performance trade offs in dynamic time warping algorithm for isolated word recognition," *IEEE Trans. Acoust., Speech Signal Processing*, vol. ASSP-28, no.6, pp. 623-635, 1980.
- [8] 김석동, 이행세, "신경망을 이용한 우리말 음성의 인식에 관한 연구," *한국음향학회지*, 11 권, 3 호, pp. 14-24, 1992.
- [9] 김영일, 이건기, 이문수, "음소에 의한 한국어 음성의 분석과 인식," *한국음향학회지*, 6 권, 2 호, pp.38-47, 1987.
- [10] 조명환, *언어심리학*, 민음사, pp.54-56, 1985.
- [11] 이종혁, 정용근, 남기곤, 윤태훈, 김재창, 박의열, 이양성, "▽²G 특성을 이용한 음성인식 신경회로망에 관한 연구," *대한전자공학회 논문지-B*, 제 29 권, B 편, 제 10 호, pp. 90-99, 1992.
- [12] 이상근, 허만탁, 이종혁, 윤태훈, 남기곤, 김재창, 이양성, "▽²G 연산자를 이용한 한국어 자음인식에 관한 연구," *대한전자공학회 추계종합학술대회 논문집*, 제 15권, 제 2호, pp. 670-673, 1992.
- [13] 이종혁, 정용근, 윤태훈, 김재창, 박의열, 이양성, "▽²G 연산자를 이용한 음성신호의 주파수 분석," *대한전자공학회 논문지-B*, 제 28 권, B편, 제 4호, pp. 24-32, 1991.
- [14] 허웅, *국어 음운학*, 샘 문화사, 1992.

 著者紹介

許萬鐸(正會員) 第30卷 B編 第9號 參照
 현재 부산대학교 전자공학과 교수

李鍾赫(正會員) 第29卷 B編 第10號 參照
 현재 경성대학교 컴퓨터공학과 교수

南基坤(正會員) 第30卷 B編 第1號 參照
 현재 부산대학교 전자공학과 교수

尹台燾(正會員) 第30卷 B編 第1號 參照
 현재 부산대학교 전자공학과 교수

金在昌(正會員) 第30卷 B編 第1號 參照
 현재 부산대학교 전자공학과 교수

李梁成(正會員) 第29卷 B編 第10號 參照