공 학 석 사 학 위 청 구 논 문

확산스펙트럼에 따른 음소사전 재분류와 신경망에 의한 모음인식에 관한 연구

2 0 0 1년 12월

부 경 대 학 교 대 학 원

전 자 공 학 과

김 수 화

공 학 석 사 학 위 청 구 논 문

확산스펙트럼에 따른 음소사전 재분류와 신경망에 의한 모음인식에 관한 연구

지도교수 허 만 탁

이 논문을 공학석사 학위청구논문으로 제출함

2001년 2월

부 경 대 학 교 대 학 원

전 자 공 학 과

김 수 화

김수화의 공학석사 학위청구논문을 인준함

2001년 12월 26일

주 심 공학박사 김 문 수 (인)

위 원 공학박사 정 수 태 (인)

위 원 공학박사 허 만 탁 (인)

목　　차

A Study of the Recognition of Vowels Using Neural-Network and Reclassification of the Phonemes with Spectrum by the Diffusion Filter

Kim-Suhwa

Department of Electronics Engineering, Graduate School.
Pukyung National University.

Abstract

In this paper, the recognition of vowels using the diffusion filter bank and a neural network is presented. Each vowels are reclassified data that have a same spectrum characteristic which are obtained by the diffusion filter bank pre-processor.

Conventional BP algorithm of the neural network have the disadvantage that will not complete because of local minima, and takes long time for training. So, the second neural network is presented. It takes the initial condition from the first neural network. It has completed the training, and is faster than the first neural network as a result of that. Used vowels in this paper are /ㅏ/, /ㅓ/, /ㅗ/, /ㅜ/, /ㅡ/, /ㅣ/, /ㅔ/, and /ㅐ/. Because /ㅔ/ and /ㅐ/ are have a similar characteristic, vowel /ㅔ/ is used as representative of them. Recognition rate of reclassified vowels by spectrum of diffusion filter bank is compared with those of not reclassified data. The results of experimentation are that the recognition rate of /ㅏ/, /ㅔ/, /ㅗ/, /ㅓ/, /ㅣ/, and /ㅡ/ are increased by 15%, 15%, 21%, 17%, 36%, and 18%, respectively, but that of /ㅜ/ is decreased by 18%. Total recognition rate is increased by 16.85%. The case of reclassification of vowels by spectrum of the diffusion filter bank have more teaching pattern and training time than the case of no reclassification. The result of that, the recognition rate is increased.

1. 서 론

음성은 인간의 정보 전달의 수단 중에 가장 편리하고 빠르다. 음성을 기계와 인간 사이의 직접적인 정보 전달 매개 수단으로 사용하는 것이 음성인식이다. 음성인식의 단위는 음소(phoneme), 음절(syllable), 단어(word)의 세 가지로 나눌 수 있다. 기존의 단어, 음절단위의 한국어 음성 인식 시스템은 제한된 인식대상에는 유용하나 많은 양을 인식할 경우는 하드웨어 구현 자체가 어렵다. 한국어 음성은 대부분 초성, 중성, 종성으로 음소가 구성되어 구분이 쉽고 각각의 단위음소로 나누기 쉽다. 따라서 한국어 음성인식의 경우 수 십여 개의 음소로 이루어진 음소 단위로 음성을 인식할 수 있다[1] [4].

음성 인식을 위한 과정을 전처리 과정과 인식 과정으로 나눌 수 있다. 먼저 전처리과정에서는 음성을 분석하여 특징 파라미터를 추출해야 하며 이 특징 파라미터를 이용하여 인식과정을 수행한다[5] [6]. 전처리 방법은 여러 가지가 있으며 대부분 주파수 영역에서의 파라미터를 추출하는 것이다. 음성 인식과정에는 두뇌의 학습능력을 모델화한 방법이 많이 사용되고 있다[7] [11]. 이중 신경망은 입력과 출력사이의 비선형 함수 관계를 학습하는 특성을 가지고 있어 음성의 비선형적 특성을 잘 수용하기 때문에 현재 많이 쓰이고 있는 HMM의 단점을 보강하는 수단으로 사용되고 있다[12] [17].

본 논문에서는 확산 필터 뱅크를 전처리기로 사용하여, 원시 음성데이터에서 모음을 분리하여 얻어진 음향데이터의 스펙트럼을 얻었다. 그리고 인식률을 높이기 위해 스펙트럼의 특성을 관찰하여 각 데이터를 재분류(소분류)하는 방법을 제안했다. 그리고 인식기로써 신경망을 이용하여 학습 및 인식 실험을 하였다. 이때 사용하는 신경망이 학습실패와 학

습에 많은 시간이 걸리는 점을 보완하기 위해 제 2 신경망을 새로이 구현하여 학습이 더 빠르고 효율적으로 완료되게 하였다. 실험 대상 모음은 /ㅏ/, /ㅔ/, /ㅐ/, /ㅗ/, /ㅓ/, /ㅣ/, /ㅡ/, /ㅜ/이다. 각각의 모음의 스펙트럼을 확산망 필터뱅크를 통해 얻는다. 이 때에 재분류하지 않은 데이터를 사용한 신경망의 시험(test) 결과를 재분류하여 학습해서 얻은 결과와 비교하였다. 후자의 경우에 음소 사전의 수가 늘어나고, 학습 시간이 늘어나는 반면, 인식률이 평균 16.85% 향상되어진 결과를 확인하였다.

2. 한국어 음성과 음성인식이론

음성에 포함된 언어정보를 자동으로 추출하여 그 정보를 기계가 이해하게 하는 작업을 음성인식이라 하며 시간패턴이나 음소환경, 주위환경 등의 조건에 따른 변동을 최소화하고 정규화하여 음성인식을 효율적으로 실현해야 한다. 일반적인 음성인식 시스템은 음성패턴의 특징을 기준패턴과 비교하여 가장 유사한 패턴을 선택하며 음성의 형태, 화자수, 음성단위 및 인식방법에 따라 여러 가지로 분류된다.

2.1 한국어 음성의 분류

음성의 분석단위로는 단어, 음절, 음소 단위로 구별되며 음성은 연속적인 아날로그 신호이다. 그러나 언어적 특성에서 보면 이산적인 음성기호로 나타낼 수 있다[18]. 음성적으로는 다르나 의미의 대립이 생기지 않는 음은 하나의 기호로 묶어 기호를 최소한으로 줄인 추상적인 기호를 음소(phoneme)라고 한다. 음소를 분류하면 크게 유성음(voiced phoneme)과 무성음(unvoiced phoneme), 또는 모음(vowel)과 자음(consonant)으로 나눌 수 있다. 또한 조음방법이나 조음의 위치에 따라서 파열음(plosive), 마찰음(fricative), 비음(nasals), 파찰음(affricate)과 전설(front tongue)모음, 후설모음 및 양순음(bilabial), 치음(dental), 설음(retroflex) 그리고 연구개음(velar)등으로 구별한다. 또 모음은 복합도에 따라서 단모음 및 복모음으로 나눈다.[19],[20]

말의 흐름 속에서 지각이 가능한 최소의 소리를 단음(phone)이라고 하며 단음은 제각기 소리 값이 다르지만 유사성과 공통성에 따라 분류된다. 분류되는 구분은 크게 자음과 모음으로 분류된다. 우리나라의 전통적인 음성의 분석은 음절구조의 관점에서 초성, 중성, 종성의 삼분법으로

나누며 초성과 종성은 자음에 중성은 모음에 해당한다. 자음은 구강 내에서 어떤 장애나 차단이 수반되면서 나는 소리이며 모음은 공기가 아무런 장애를 받지 않고 빠져나가면서 나는 소리이다. 단모음은 길게 발음하더라도 그 음이 변하지 않는 모음으로 /아/, /에, /애/, /이/, /어/, /으/, /오/, /우/ 등이 있다. 복모음은 단모음의 중첩이므로 단모음 두 개를 조합하면 복모음이 구성된다. 단모음의 분석에는 모음의 조음형태와 대응하여 특정 주파수 성분이 강조되며 모음을 특정 지우는 주파수 스펙트럼에 의한 방법을 주로 사용한다[19],[20].

2.2 음성의 특징 추출 방법

음성분석에 있어서 가장 중요한 것 중에 하나는 특징을 추출하는 것이다. 특징 파라미터를 추출하는 방법으로는 음성의 발성 과정을 모델링하여 추출하는 방법과 음성 신호자체를 주파수 영역이나 시간 영역에서 해석하여 추출하는 방법 그리고 음성의 인지 과정을 모델링하여 추출하는 방법 등이 있다[5],[17]. 발성과정을 모델링하여 추출하는 방법은 발음이 성대로부터 성도를 거쳐 입술까지 이르는 과정을 모델링하고 이 모델로부터 특징 파라메터를 구하는 것으로서 전극모델(all-pole model)로 모델링한 LPC(linear predictive coding)방법, LPC계수로부터 켑스트럼(Cepstrum)계수를 구하고 이 계수에 적절한 가중치를 가하여 이를 특징 벡터로 표현하는 방법과 성대의 개방과 폐쇄 조건을 고려한 LSP(line spectral-pair frequency)방법 등이 있다[4],[19].

음성 신호 자체를 주파수 영역이나 시간 영역에서 해석하여 추출하는 방법은 각 영역에서 음성신호의 변화를 관찰하여 음성 특징을 추출해 내는 것이며 특징 파라미터로서는 에너지, 포만트(formant), 지속시간(time duration), 영교차율(zero crossing rate)등이 있다[6],[22]. 또 인지과정을

모델링하여 특징 파라미터를 추출하는 방법으로 달팽이관(cochlear)필터 뱅크 출력을 이용하는 방법을 들 수 있다[23]. 최근에는 시간영역의 동적 특성(dynamic feature)을 주파수 영역의 특징(spectral feature)들과 함께 사용하기도 한다[19].

본 논문에서는 인지과정을 모델링한 확산필터뱅크를 구현하여 신호의 스펙트럼 포락선을 구하는 방법을 사용하였다. 이 방법은 시간영역에서 음소(모음구간)를 추출하고 신경세포가 외부로부터 자극을 받아 그 자극의 세기가 인접세포로 확산되어 가는데 이 확산되어 가는 자극의 분포가 가우시안(gaussian)이 된다는 특성을 이용한 것이다. 확산필터뱅크는 확산망이 공간주파수를 추출해 내는 대역통과 특성이 있다는 점과 신경세포간의 연결이 매우 적어 하드웨어 구현이 용이하고 적은 용량으로 많은 채널의 필터를 구성할 수 있다[7] [9]. 확산 필터의 중심 주파수는 확산회수와 차분회수에 의해 결정되어지며 선택도는 차분회수만에 의해 결정되며 선택도 조정에 의해 적당히 스무드한 스펙트럼을 구할 수 있다[10] [11].

2.3 음성인식 방법의 분류

음성인식 시스템을 기술적 요인으로 분류하면 음성형태에 따라 격리단어, 연결단어 그리고 연속음성으로 분류되고, 인식단위에 따라 단어단위, 음절단위 그리고 음소단위로 분류된다[4],[5]. 그리고 어휘수의 양에 따라서도 구별되어지고 화자 수에 따른 화자종속과 화자독립으로 구별된다. 또 일반 자연 언어에 가까우나 정해진 문법이나 어순에 따르는 말과 소음의 정도, 사용목적이 정해져 있는 경우 등의 언어종류에 따라서도 구별된다. 기존의 한국어 음성인식 연구의 접근방식은 주로 단어단위로 음성인식 시스템을 구현해 왔다[1],[2],[4],[19]. 그러나 한국어 음성의 경우에

2000여 개의 음절과 40여 개의 음소로 나눌 수 있으며 이를 사용하여 모든 단어들을 구성할 수 있다. 단어 또는 음절단위의 인식은 인식률이 좋지만 단어수가 많기 때문에 비례적으로 많은 메모리가 필요하고 시스템의 복잡도가 커지는 등의 난점이 발생한다. 따라서 한국어 음성인식 시스템의 경우, 연속음성인식으로의 확장이 용이한 음소의 단위로 음성을 인식하는 접근이 필요하다.

최근의 음성인식 연구는 선형예측부호화(LPC: linear predictive coding)법과 동적계획법을 이용한 시간정합(DTW: dynamic time wrapping)에서 벡터양자화(VQ: vector quantization)법, 은닉마르코프모델(HMM: hidden markov model), 신경회로망(neural network), 퍼지이론 쪽으로 많은 연구가 진행되어 지고 있다[24] [26].

VQ는 음성신호 데이터를 압축시키기 위해 기준 패턴을 생성할 때 이용하며 신호 파형 중 대표적인 몇 가지의 모양을 미리 준비해 놓고 입력파형에 대하여 가장 비슷한 모양을 정합 시키는 것이다. MS(multi section) VQ는 한 단어를 발성순서에 따라 몇 개의 구간으로 나누어 독립된 codebook을 작성하므로 시간적 정보를 포함하는 알고리즘이다[25].

DTW는 기준패턴과 시험패턴과의 거리를 계산하여 최소거리를 가지는 패턴을 선택하는 패턴 정합방법으로 음절이나 단어단위 음성의 인식에 있어 음성의 시간적 변화를 모델화 할 수 있는 반면에 인식할 데이터가 많아질수록 인식시간이 길어져 음소단위 인식에는 적합하지 않다[14],[25].

HMM(hidden markov model) 알고리즘은 음성이 Markov model로 모델링 될 수 있다는 가정 하에 학습과정에서 Markov model의 파라미터를 얻어 기준 Markov model을 만들고 인식과정에서는 입력음성과 가장 유사한 기준 Markov model을 찾아냄으로써 인식한다. HMM은 음성

인식에 널리 이용되고 있지만 모든 확률이 현재상태에만 의존한다는 가정 하에 처리되므로 음성의 조음효과를 충분히 모델링하기에 부족하며 근접 입력 프레임의 상관관계를 무시해가며 독립적 처리를 하는 단점이 있다. 이는 인식할 데이터수가 많아질수록 복잡도가 커지며 방대한 변수를 다루어야 하는 문제가 있다[1],[2],[25].

신경망은 입력과 출력사이의 비선형 함수관계를 학습하는 특성이 있어 HMM의 단점을 보강하는 수단으로 많이 연구되고 있다[12]. 신경망은 생물 신경조직의 기본단위인 뉴런을 모델링의 대상으로 하고 있다. 뉴런은 중심의 체세포(soma)와 이로부터 뻗어 나온 수상돌기(dendrite)와 축색돌기 방향으로 전달하며 수상돌기와 체세포 사이에는 공간적 합산과 시간적 합산이 일어나 그로 인한 결과가 시냅스(synapse)를 통하여 다른 뉴런의 수상돌기에 전달된다. 신경망은 이러한 과정의 반복으로 신경전달이 이루어지는 것을 본떠서 만든 것이고, 다층신경망의 각 층간 가중치를 정해주는 효율적 학습방법인 BP(backpropagation)학습법에 의해 학습이 가능하게 되었고 신경망에 중간층을 둘 수 있게되어 2층 퍼셉트론에서 불가능했던 XOR의 문제도 학습 가능하게 되었다.[21] 신경망은 자율적으로 판정 경계(decision boundary)를 정하며, 스스로 학습할 수 있는 기능을 가지고 있으며 새로운 입력패턴에 대해서도 가장 비슷한 범주를 찾을 수 있고 학습 받은 패턴과 형태가 조금 다르더라도 인식할 수 있으며 병렬 구조로 이루어져 있어서 처리 속도가 빠르다[12],[13].

3. 음성 인식 도구

음성인식을 위한 과정을 전처리 과정과 인식 과정으로 나누는데 전처리 과정에서 먼저 음성을 분석하여 특징 파라미터를 추출하여야 한다. 본 논문에서는 확산필터뱅크를 사용하여 특징 파라미터를 추출하고, 인식 과정에서는 다층신경망을 이용해 음성 정보를 학습 및 인식한다.

3.1 확산필터뱅크

동물의 신경망은 신경세포가 자극을 받게 되면 이웃하는 신경세포 쪽으로 그 자극이 전달되어 가는데 이를 모델화하여 확산망이 제안되었으며 확산망의 전달함수는 가우스함수의 특성을 이용한 신호처리에 아주 유용하게 적용될 수 있다. 본 절에서는 가우스함수를 이용한 확산필터의 원리와 이를 이용한 확산필터뱅크의 이론 및 확산필터뱅크에 의한 스펙트럼 추출원리를 기술한다.

3.1.1 확산필터

동물신경망의 수용장(receptive field)에 자극이 인가되면 그 자극은 인접세포로 확산되고 그 분포는 가우스함수의 형태를 형성하며 이 수용장에 인가된 신호는 가우스함수의 콘볼루션(convolution)이 되고 또 결과는 흥분중심(ON center)세포들의 세기와 억제중심(OFF center)세포들의 세기가 결합되어 두 가우스함수의 차인 DOG가 형성되도록 한다[8],[9]. 이러한 일련의 과정은 특정한 주파수를 검출해낼 수 있는 대역통과 특성을 가지게 된다.

한 신경세포에 단위 임펄스 자극이 인가되면 인접한 다른 신경세포로

자극이 전달되면서 확산이 계속 발생하면 결과적으로 자극의 확산은 가우스 분포에 가깝게 근접한다. 따라서 이 모델은 신경세포간의 연결이 매우 적으면서 가우스함수를 실현할 수 있는 효율적인 시스템이 된다. 이러한 원리는 식 (3.1)과 같이 표현된다[9],[10].

$$V_i[k+1] = (1-2\alpha)V_i[k] + \alpha V_{i-1}[k] + \alpha V_{i-1}[k] \tag{3.1}$$

단, (i= ⋯ -3, -2, -1, 0, 1, 2, 3, ⋯)

여기서 i는 신경세포의 일련번호이고 k는 확산회수이며 가우스 함수를 구성한다. 또, 흥분중심세포와 억제중심세포간의 출력차는 다음 식과 같이 둘 수 있다[9],[10].

$$U_i[n+1] = 2U_i[n] - U_{i-1}[n] - U_{I+1}[n] \tag{3.2}$$

식 (3.1)와(3.2)으로부터 n이 1씩 증가함에 따라 확산 1회와 동시에 2차 차분한 것과 같이 된다[16]. U_i를 차분한다는 것은 연속신호를 미분한 것과 근사함으로 식 (3.2)은 $g^{(2)}(t)$를 구현하는 것이며 이를 n번 수행하면 n차 $g^{(2)}(t)$ 즉 $g^{(2n)}(t)$를 구현하게 된다. 확산망을 k번 수행하고 차분망을 n번 수행했을 때 $\sigma = \sqrt{2\alpha(k+n)}$이므로 α, k, n의 함수가 되어 아래와 같이 된다[9],[10].

$$\omega_0 = \frac{\sqrt{2n}}{\sigma} = \sqrt{\frac{n}{\alpha(k+n)}} = \sqrt{\frac{1}{\alpha(1+\frac{k}{n})}} \tag{3.3}$$

$$Q = \sqrt{\frac{n}{\ln 2}} \tag{3.4}$$

위 식에서 ω_0는 필터의 중심주파수이며 총 확산회수(k+n)의 자승근에 반비례하고 차분회수의 제곱근에 비례하며 n은 2차 차분회수 이고 선택도 Q는 차분회수 n의 제곱근에 비례하며 확산회수 k와는 무관하다는 것을 알 수 있다.

3.1.2 확산필터뱅크

확산망의 수행회수와 차분망의 수행회수는 시간에 따라 증가시킬 수 있으므로 추출 채널의 순서는 그 중심주파수에 해당하는 k와 n이 증가하는 방향으로만 설정할 수 있다. 또한 확산과 차분의 연산은 순서에 상관없이 회수에만 의존한다. 식 (3.4)에서 선택도 Q는 확산회수 k가 비교적 클 때 근사적으로 차분회수 n의 제곱근에 비례하고 확산회수 k와는 관계가 없다. 그림 3.1은 차분회수 n을 고정시켜 선택도를 변화하지 않고 중심주파수만 변화시킬 수 있음을 보여 준다. 그림 3.2는 식 (3.3)에서 나타난 것과 같이 확산회수와 차분회수가 같은 비율로 변화함에 따라 중심주파수는 변화하지 않고 선택도만 변화함을 보여준다[9] [11].

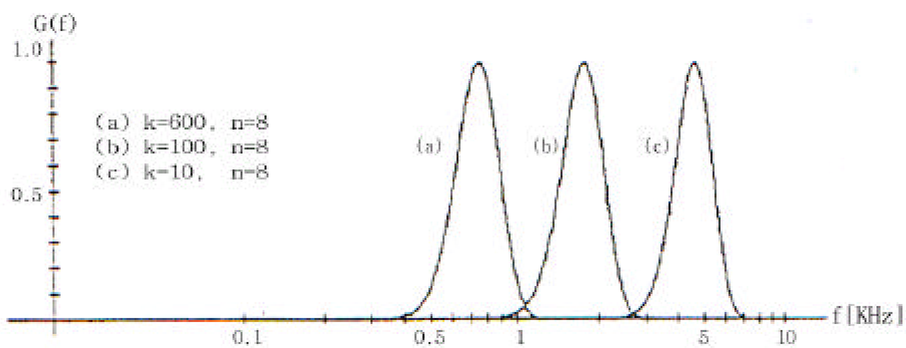


그림 3.1 중심주파수만 변화하는 경우

주어진 확산망이 추출할 수 있는 중심주파수의 범위는 확산회수 k와 차분회수 n의 가변범위 및 샘플링 주파수에 의하여 정해진다. 확산계수를 고정하면 입력된 신호는 k와 n의 값, 그리고 샘플링 주파수에 의하여 결정된 중심주파수에 따라 여파 처리된다.

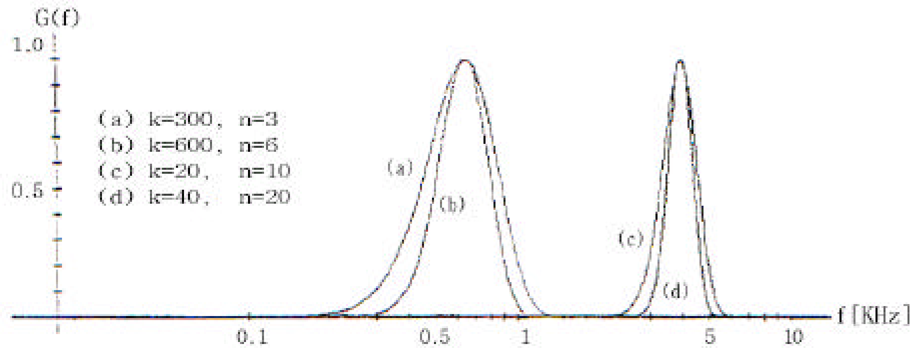


그림 3.2 선택도만 변화하는 경우

각 필터 채널을 조합하여 목적하는 필터뱅크 시스템을 구성하게 되면, 확산필터뱅크 시스템의 전체 구성은 그림 3.3과 같다. 여기서 입력신호의 스펙트럼만을 요구할 때 여파출력의 실효치만 검출하면 되므로 한 프레임의 길이를 t_f 라하고 샘플링 주파수를 f_s , 추출할 최저주파수를 f_l , 한 프레임에 입력되는 최저주파수 성분의 주기 수를 T_l 라 하면, 한 프레임에 필요한 신경 세포의 수 M 은

$$M = t_f f_s = \frac{T_l}{f_l} f_s \tag{3.5}$$

와 같이 된다. 식 (3.5)에 의해 각 확산망에서 출력망을 제외한 입력망, 확산망, 차분망 및 이득 보상 회로의 신경 세포의 수는 여파할 프레임의 샘플링수 M과 가장자리 효과[16]를 제거하기 위한 여분의 신경세포수 m을 합한 크기이며 출력망의 크기는 M이다. 여기서 한 프레임의 길이 t_f는 최저 주파수에 따라 결정되므로 M은 최저 주파수와 샘플링 주파수에 비례하여 정해진다. 그림 3.3에 있는 확산필터뱅크의 동작은 다음과 같다.

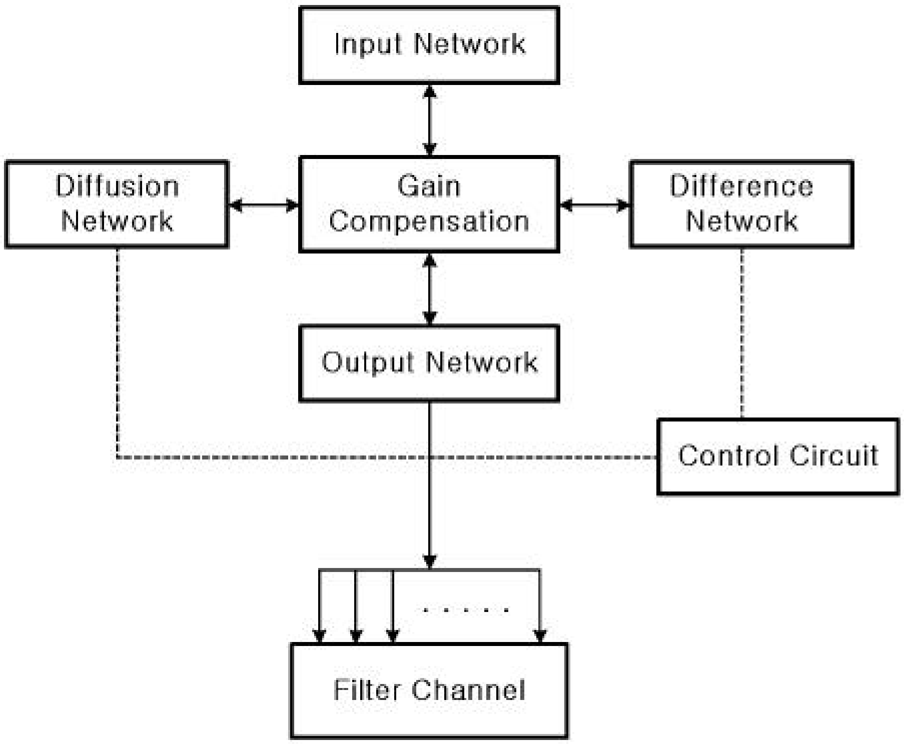


그림 3.3 확산필터뱅크의 블록도

입력망은 저역 필터를 통과한 후 샘플링된 외부신호를 직렬로 받아들여가면서 쉬프트 동작을 하는데, 샘플링된 외부 신호를 최초에는 $M+m$개를 받아들인 후 다음부터는 M개씩 쉬프트한 후 이득 보상 회로를 병렬로 내 보낸다. 출력망은 여파 특성을 결정하는 k와 n의 합이 작은 순으로 여파된 신호를 이득 보상 회로로부터 받아 k와 n의 합이 작은 필터 채널 순으로 차례대로 내 보낸다. 이득 보상 회로는 먼저 입력망으로부터 신호를 병렬로 받아 확산망 또는 차분망으로 내 보내면서 제어 시스템으로부터 전송된 이득 보상 계수에 의하여 스케일링한다. 확산망과 차분망은 이득 보상 회로로부터 받은 신호를 제어 시스템으로부터 보내

온 회수만큼 처리하여 다시 이득 보상 회로로 전송한다. 제어 시스템은 각 블록에서의 입력과 출력의 출입을 사전 계획된 대로 제어하고 이득 보상 계수를 계산하여 이득 보상 회로로 전송하며 사전 프로그램된 확산 회수와 차분 회수를 확산망과 차분망으로 각각 전송한다. 필터 뱅크의 최종 출력은 출력망으로부터 보내온 여파된 프레임별 신호의 실효치를 구하여 각 필터 채널로 전송된 값들이다.

이런 확산필터뱅크 시스템은 1개의 필터 시스템으로 하드웨어양의 증대 없이 많은 채널을 가진 필터뱅크를 구현할 수 있으며 하드웨어의 구조 변경 없이 필터뱅크의 파라미터를 바꾸어 중심 주파수 간격을 세분화할 수 있다[9] [11],[31].

3.2 다층 신경망(EBP)

신경망은 생체 두뇌의 처리 단위인 신경세포군의 병렬처리를 모사하여 기존의 순차 처리 컴퓨터에서 어려웠던 패턴 인식을 가능하도록 하였다. 신경망의 연구 목적은 크게 두뇌의 물리적 모사를 통한 지능의 연구와 실용적인 용도에의 응용이라 할 수 있다. 생체의 두뇌와 비교할 때, 인공신경망은 측정 오차의 영향이 적으며 변화상태에 대한 계측이 용이하고 학습 조건과 상태를 임의로 조정할 수 있다는 장점을 지니고 있다.[27][28]

3.2.1. 신경망

인공 신경망은 생체의 두뇌 구조에 대한 물리적 모사를 통하여 패턴 분류기로서의 역할을 수행한다. 신경망의 구성은 크게 인공신경세포로 이루어진 신경망 구조와 구조에 적합한 학습 규칙으로 나눌 수 있으며,

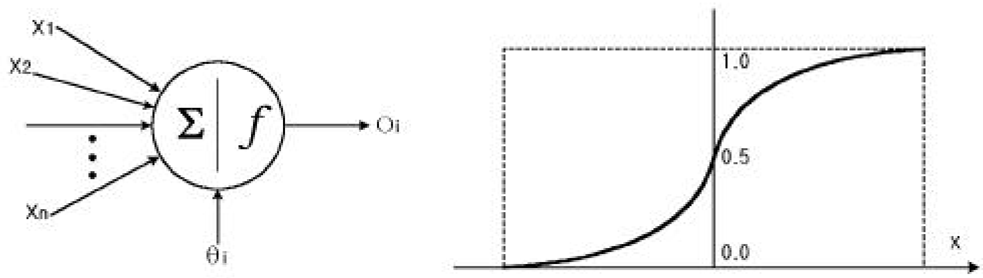
일반적으로 인공신경망(artificial neural network)은 활성함수(activation function)를 가지고 있는 신경 세포와 각 세포를 연결하는 연결가중치로 구성되어 있다. 인공신경세포는 층(layer)을 이루고 있으며 각 층은 입력패턴을 받아들이는 입력층과 입력된 패턴에 대한 신경망의 동작 결과를 출력하는 출력층, 그리고 다층신경망(multi-layer neural network)의 경우, 입력층과 출력층 사이에 하나 이상의 은닉층을 가지고 있다. 입력층 신경세포의 입력값들은 가중치와 곱해져 다음 층의 각 신경세포에서 합산되며 합산값에 활성함수가 취해져 해당 세포의 출력이 발생한다. 출력값의 입력층으로의 재입력 여부에 따라 신경망은 회귀(recurrent)과 비회귀(nonrecurrent) 신경망으로 구분된다.

신경망의 학습과정은 적합한 출력을 내기 위한 연결 가중치의 변화이며 학습 규칙은 가중치의 변화 방법이다. 신경망의 정보는 가중치에 분산되어 표현되며 가중치의 갱신은 각 층에서 동시에 병렬로 수행된다. 따라서, 신경망의 동작 상태에 대한 모든 정보는 가중치의 변화에 포함되었다고 할 수 있다. 신경망의 학습 방법은 크게 목적 패턴을 사용하는 감독 학습방법(supervised learning)과 목적 패턴을 사용하지 않는 무감독 학습방법(nonsupervised learning)이 있다[12],[13],[27],[28].

본 논문에서는 다층인식자(MLP, multi layer perceptron)를 신경망 모델로 이용하였으며, 신경망을 학습하기 위한 학습 규칙으로는 감독학습인 오류역전파법(BP, error backpropagation learning rule)을 적용하였다. MLP는 경사하강법(gradient descent method)의 일종인 역전파법을 이용하여 학습을 진행한다. 그러나, SLP의 경우는 학습수속정리에 의해 학습의 수렴이 보장되지만, 다층으로 확장되면서 학습 진행이 수많은 지역 극소점(local minima)에 고착되는 문제점이 발생하게 된다. MLP의 또다른 문제점은 학습에 상당한 시간이 소요된다는 것이다[29],[30].

3.2.2. 기존 신경망 인식기의 개선

신경망에서 사용되는 뉴런의 구조와 활성함수는 그림 3.4와 같다.



$OUT = F(\sum w_i x_i + \theta_i)$ $\qquad\qquad$ $f(x) = \dfrac{1}{1 + \exp(-\sigma x)}$

그림 3.4 뉴런(neuron)의 구조와 sigmoid 함수

(a) 뉴런 (b) 시그모이드 함수

신경망은 기본적으로 노드(node)와 이들간을 서로 연결하는 가중치로 구성된다. 노드는 뉴런 또는 PE(processing element)라고 하는 개별 처리소자를 가지며 이들은 서로의 연결을 통하여 정보를 주고받는 병렬 분산처리구조를 갖는다. 뉴런 U_i의 총입력량 NT_i은 그림 3.4(a)와 같이 다른뉴런 U_k의 출력 O_k와 뉴런 U_i와 U_k간의 연결세기 W_{ik} 로 부터

$$NT_i = \sum_{k=1=0}^{n} W_{ik} O_k + \theta_i \tag{3.6}$$

로 표시된다. 여기서 θ_i는 뉴런 U_i의 바이어스를 나타낸다. 뉴런망의 전달함수에 위 식에서 구한 총 입력량을 인가하면 뉴런 U_i의 실제 출력 O_i는 다음과 같다.

$$O_i = f_i(NT_i) = \frac{1}{1 + \exp(-NT_i)} \tag{3.7}$$

그리고 일반적으로 전달 함수는 그림 3.4(b)과 같이 단조 증가 함수인 sigmoid함수를 사용한다[12],[26],[32].

다층 신경망의 학습법은 일반화된 델타법으로 불리워지는 BP(back-propagation)법이며 학습을 위해서 외부교사가 존재하여 입력 패턴에 대해 원하는 출력 패턴을 출력하도록 가중치를 조정하여 실제 출력과 원하는 출력과의 오차를 최소로 하는 것이다. 다층 신경망에서 뉴런간의 연결세기를 W라하고 학습시킬 대상패턴을 p=1, 2, ..., n라고 하자. 그리고 입력패턴p에 해당하는 원하는 출력패턴과 실제의 출력패턴을 각각 T_p , O_p라고 하자. 그리고 각 패턴p에 대응하는 에러를

$$E_p = (T_p - O_p) \tag{3.8}$$

로 정의할 때 패턴p에 대한 에러함수는

$$E_p(W) = \frac{1}{2} \mid E_p \mid^2 = \frac{1}{2} \mid T_p - O(W) \mid^2 \tag{3.9}$$

로 표시되며 전체패턴에 대응하는 에러함수는

$$E(W) = \sum_{p=1}^{n} E_p(W) \tag{3.10}$$

로 주어진다.

신경망을 학습시킨다는 것은 $E(W)$를 최소로 하는 W를 찾는 문제로 귀착되는데 Rumelhart등은 급강하법(gradient descent method)을 사용해 오차를 최소로 하는 연결강도를 구하였다[27]. 본 논문에서는 급강하법에 관성항과 바이어스항을 추가하여 다층 신경망(multi-layer perceptron)의

$E(W)$를 최소로 하는 W를 반복적으로 찾는 방법을 다음과 같이 사용하였다[12],[13],[26],[32].

$$\begin{aligned} W_p(n+1) &= W_p(n) + \triangle W_p(n+1) \\ \triangle W_{ij}(n+1) &= \eta\, \delta_i\, O_j + \alpha \triangle W_{ij}(n) \end{aligned} \tag{3.11}$$

$$\begin{aligned} \theta_p(n+1) &= \theta_p(n) + \triangle \theta_p(n+1) \\ \triangle \theta_i\,(n+1) &= \eta\, \delta_i + \alpha \triangle \theta_i(n) \end{aligned} \tag{3.12}$$

$$\begin{aligned} \delta_{p\,i} &= f_i\,'(NT_i)(t_i - o_i) \\ &= f_i\,'(NT_i)\sum W_{k\,i}\,\delta_k \end{aligned} \tag{3.13}$$

여기서 δ_i는 후방향(backward)으로 전달되는 에러를 나타낸다. 바이어스항 θ는 입력 값이 항상 1인 뉴런과 다른 층의 뉴런간을 서로 연결하는 연결강도로 취급하여 θ 도 W와 마찬가지로 학습시킬 수 있다. η는 학습율로서 $0 < \eta < 1$의 범위를 만족하는 값으로 시뮬레이션에서 적절한 값으로 두며, α는 모멘텀항으로 현재의 가중치 변화량을 결정하는데 과거 가중치 변화량을 반영하는 상수이다.

전형적인 다층(multi-layer perceptron)신경망은 입력층과 출력층, 그리고 하나 이상의 히든층을 갖는 구조를 이루고 있으며 입력과 출력층에 위치한 뉴런은 외부와 직접 연결되어 있고 은닉층에 위치한 뉴런은 외부와 직접 연결되어 있지 않다. 그리고 같은 층에 위치하는 뉴런간의 측방향의 연결은 고려하고 있지 않다. 그러나 전형적인 다층 신경망은 인식하여야할 대상의 수가 증가할 때 은닉층의 뉴런수가 증가하게 되고 이로 말미암아 학습시간이 더욱 증가하는 단점이 있다.

이를 개선하기 위해 본 논문에서는 인식대상들 중에서 상대적으로 에러값이 큰 것을 검색하여 따로 보조학습을 시키는 방법과 학습을 효율적으로 완료하기 위하여 제 1 신경망과 제 2 신경망을 개별적으로 구성하여였다. 인식실험을 위해 구현한 다층 신경망 학습기와 인식기를 그림 3.5에 나타내었다.

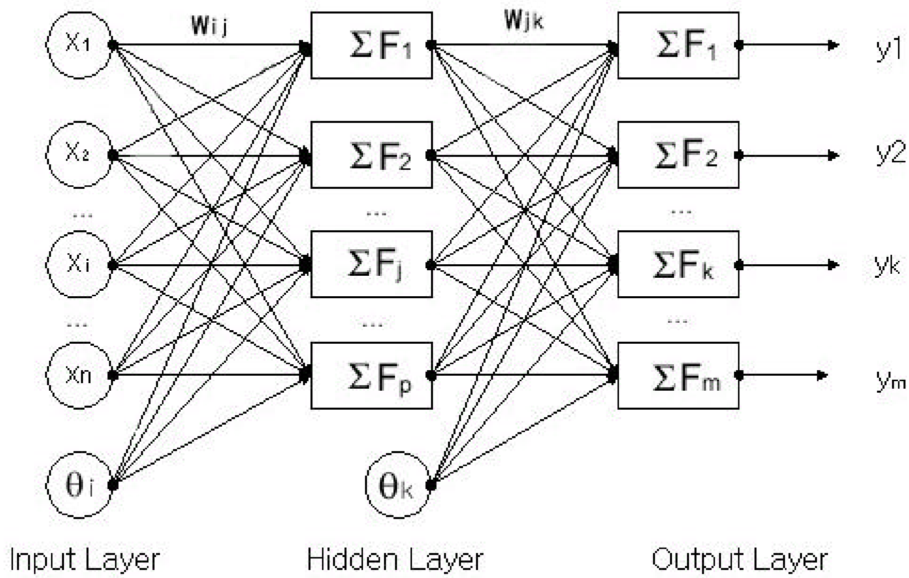


그림 3.5 학습을 위한 다층 신경망 모델

제 2 신경망은 제 1 신경망과 구조와 학습 방법은 같으나, 제 1 신경망의 학습 진행률을 관찰하여 제 1 신경망의 학습이 완료되지 않더라도, 그림 3.6과 같이 지역 국소점에서 전체적으로 신경망이 안정되는 점을 감안하여 각 노드의 연결 가중치를 추출한다. 이때 에러곡선이 상승벡터를 그릴 때 초기치를 추출하는 것으로, LMS 학습방법을 응용한 것이다.

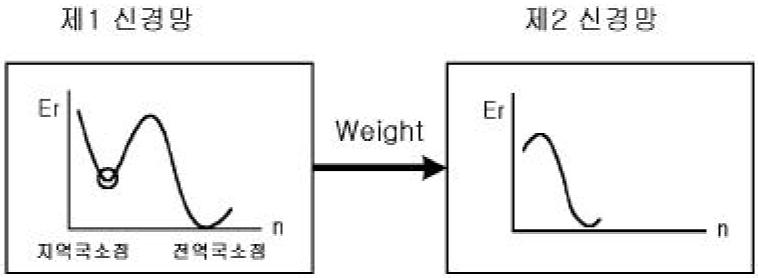


그림 3.6 **효과적인 학습을 위한 제** 1 - **제** 2 **신경망**

그리고 이 값들을 제 2 **신경망의 초기치로 보내는 방법을 시행하여 제** 1 **신경망에서 학습을 완료시키지 못하더라도 제** 2 **신경망에서 더 짧은 시간에 학습을 완료하게 하였다. 이 방법은 망**(network)**의 구조가 같을 경우 제** 1 **신경망으로부터 추출한 초기치를 제** 2 **신경망은 계속해서 사용함으로써, 실험의 양과 횟수가 증가할수록 학습 진행률에서 효과적이다.**

4. 인식 실험 및 검토

4.1 기존 방법에 의한 학습 및 인식 실험

본 연구에서 수행한 실험의 전체적인 흐름도를 그림 4.1에 나타내고 있다.

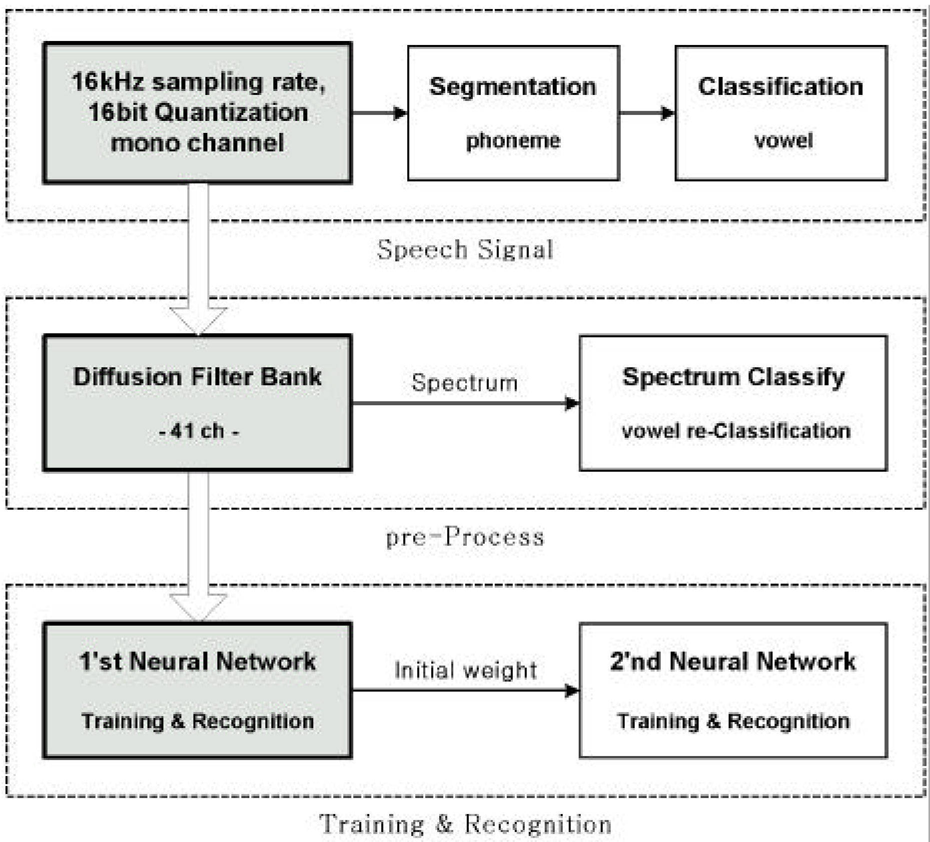


그림 4.1 전체 실험 과정

실험 단계는 크게 원시 음향 데이터 처리단계, 전처리과정, 그리고 학습 및 인식과정으로 나눌 수 있는데, 원시 음향 데이터 처리 단계에서는

각 구성 음소별로 세그멘테이션(segmentation)을 하고, 음성의 특징 파라미터를 추출하는 전처리과정에서는 확산망필터뱅크를 이용하고, 이에 따른 스펙트럼 특징을 분류하게된다. 그리고 제 2 신경망을 이용해서 학습과 인식 실험이 이루어진다.

실험에 사용되는 시료는 한국전자 통신연구소의 ETRI 445 PBW DB로서 16kHz sampling rate, 16bit quantization된 22명의 남자화자가 445개 단어를 발음한 것이다. 이 중 4명의 남자화자의 시료를 실험에 사용한다. 각 음소를 구별하기 위하여 시간의 추이에 따라 음성의 각 샘플의 크기를 비교하여 임계값 이상이 되는 시점을 음소의 시작점으로 하였다. 이 때 임계값은 묵음구간의 최대 신호 값보다 큰 값으로 정하였다. 음성신호 중 모음은 자음에 비해 주파수가 낮게 나타나므로 시각적으로 구분이 가능하여 레벨링 표에 의해 1차적인 음소의 분절을 하고, 그림 4.2와 그림 4.3과 같이 분리된 음소데이터를 편집할 수 있는 프로그램(Cool Edit 2000)을 이용하여 시각과 청각에 의해 재분절하고, 분류한다. 이때 샘플 개수는 1000 ±200 개의 샘플로 하였다.

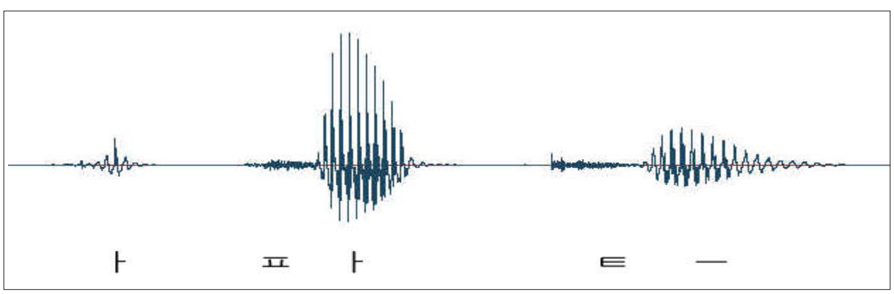
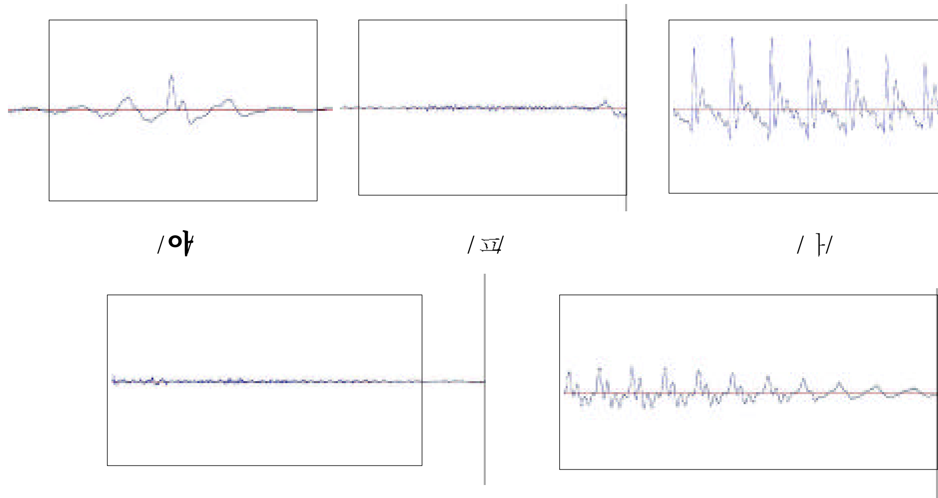


그림 4.2 원시 음향 데이터 - /아파트/ - 16kHz, 16bit, mono

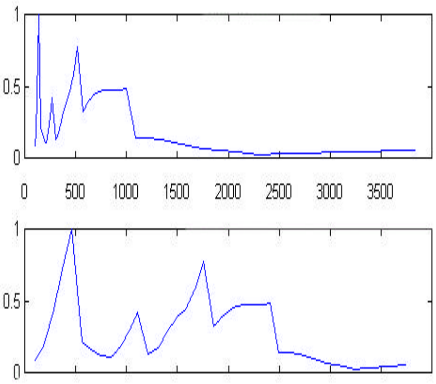


/아/ /ㅍ/ /ㅏ/

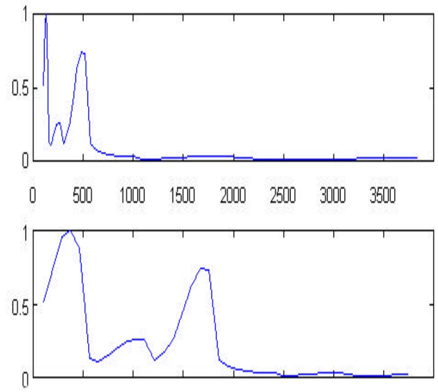
/ㅌ/ /ㅡ/

그림 4.3 **원시 음향 데이터를 각 음소 단위로 분리**

이렇게 얻어진 음소 데이터는 전처리기인 확산필터뱅크를 통하여 음소의 확산스펙트럼을 얻게 된다. 그림 4.4**에 나타낸 스펙트럼은 그림** 4.3 **에서** /ㅏ/**와** /ㅡ/**에 대한 스펙트럼이다. 실험에서는 모음** /ㅏ/, /ㅓ/, /ㅗ/, /ㅜ/, /ㅡ/, /ㅣ/, /ㅔ/, /ㅐ/**의 스펙트럼을 추출하였다. 사용된 필터 뱅크는 확산회수와 차분회수를 가변시켜 중심주파수 간격을 세분화 할 수 있는 특성에 따라 표** 4.1**와 같이 저역에 집중되도록** 41**채널의 중심주파수를 갖는 구조로 설계되었다. 각 음소의 확산 스펙트럼은 그 특징을 확인하고, 특징별로 분리한다. 얻어진 확산스펙트럼의 특성을 그래프로 출력하여 특성이 유사한 그룹으로 음소를 재분류한다. 그리고 학습 목록을 만든 후 분류되지 않은 학습데이터와, 재분류된 학습데이터를 제** 2 **신경망의 입력 노드에 입력하여 학습을 에러값이 설정값 이하로 나타날 때까지 실행한다.**



(a)



(b)

그림 4.4 **그림** 4.3)**에서** /ㅏ/**와** /ㅡ/**의 확산망필터뱅크를 통한 스펙트럼**
(a) /ㅏ/**의 스펙트럼** (b) /ㅡ/**의 스펙트럼**

표 4.1 **확산 필터 뱅크의 각 중심 주파수**

Filter Channel No.	Center Freq.[㎐]	Filter Channel No.	Center Freq.[㎐]	Filter Channel No.	Center Freq.[㎐]
1	100	15	367	29	1335
2	108	16	402	30	1468
3	120	17	435	31	1609
4	130	18	480	32	1742
5	143	19	521	33	1921
6	158	20	574	34	2085
7	172	21	632	35	2296
8	191	22	691	36	2531
9	211	23	765	37	2765
10	232	24	847	38	3062
11	250	25	929	39	3390
12	272	26	1000	40	3593
13	302	27	1093	41	3843
14	333	28	1210		

학습기와 인식기는 연속음 인식으로의 확장이 용이한 음소단위 인식을 위해 메모리를 적게 사용하고 구조가 단순한 다층 신경망을 구성하여 사용하였다. 학습과 인식을 위해 이러한 확산망필터뱅크를 출력한 스펙트럼 결과를 이 다층 신경망을 이용한다. 신경망은 제 1 신경망과 제 2 신경망으로 구성되어 제 1 신경망의 학습 진행과정을 연속적으로 관찰하여 에러 곡선이 지역 국소점에 머물러 학습이 완료되지 않은 상황에서 안정이 되어 더 이상의 학습이 진행이 되지 않은 점을 감안하여, 각 노드의 연결 가중치를 제 2 신경망의 초기 조건으로 전달하여 초기치로 사용한다. 그리고 제 2 신경망에서 학습을 재 진행하여 완료하게 하였다. 입력층, 은닉층, 그리고 출력층의 노드가 변하기 않은 신경망에는 이 초기치를 실험시마다 적용이 가능하여, 전체적으로 실험의 횟수가 증가할수록 학습의 시간이 감소하며, 효율적이었다. 그리고 다층 신경망은 인식하여야할 대상의 수가 증가 할 때 은닉층의 뉴런수가 증가하므로 학습시간이 증가하는 단점이 있으므로 이를 개선하기 위하여 인식 대상들 중에서 상대적으로 에러값이 큰 것을 검색하여 보조학습을 시키는 방법을 행하여 전체 학습회수를 줄여 빨리 수렴하도록 하였다. 결과적으로 보조학습기와 제 2 신경망은 학습의 효율을 높이기 위해서 사용된 방법이다. 수정된 학습과정을 그림 4.5에 나타내었다.

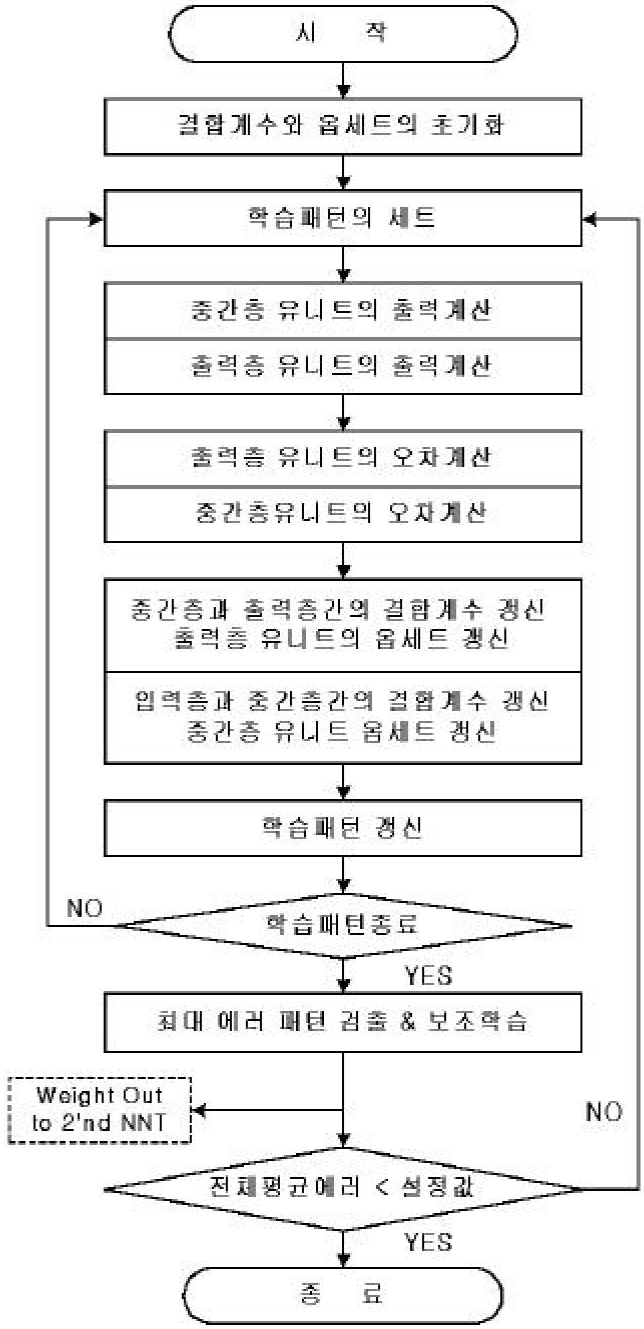
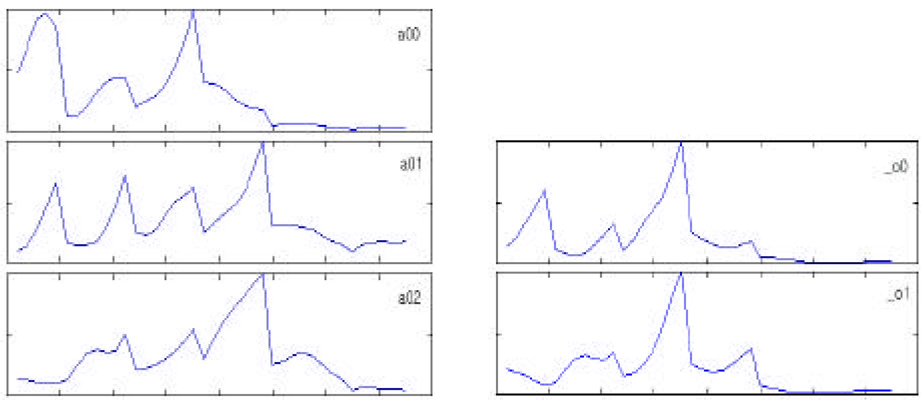


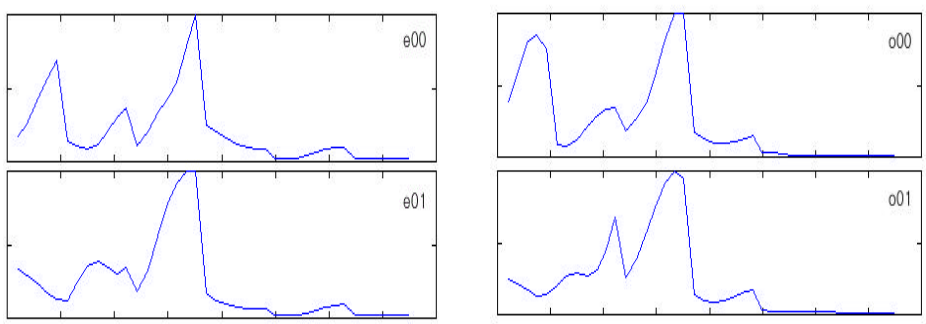
그림 4.5 **신경망 학습 진행 과정 - 제**1, **제**2**신경망**

4.2 개선된 방법에 의한 인식 실험

본 논문에서 확산스펙트럼 특징에 따른 음소의 재분류를 제안하였는데, 그 특징에 따라 음소사전을 재구성하여 인식 실험을 한 음소는 /ㅏ/, /ㅓ/, /ㅗ/, /ㅜ/, /ㅡ/, /ㅣ/, /ㅔ/이다. 여기서 /ㅔ/와 /ㅐ/는 발음상 구분이 어렵고, 확산망 필터 뱅크를 통과한 스펙트럼이 유사하게 나타났기 때문에 /ㅔ/를 대표음으로 한다.



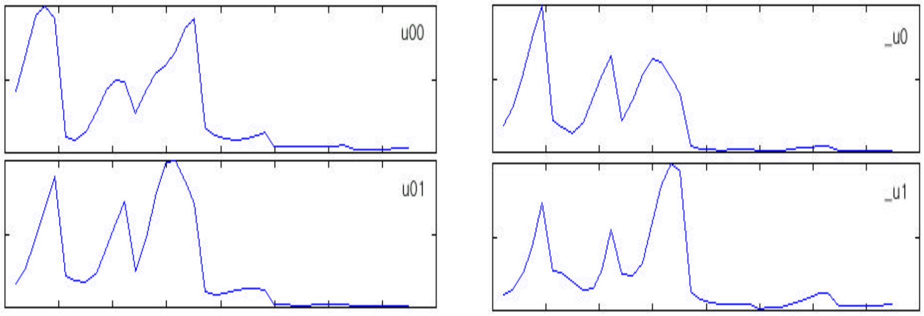
(a)



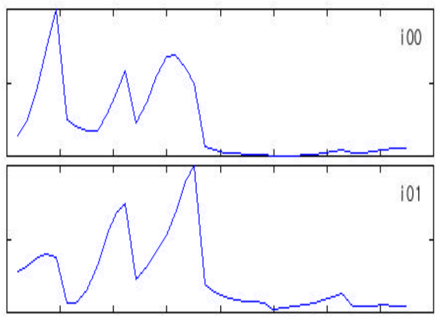
(b)

그림 4.6 재분류된 각음소의 대표적인 스펙트럼 (계속)

(a) /ㅏ/,/ㅓ/ (b) /ㅔ/,/ㅗ/



(c)



(d)

그림 4.6 (**계속**) **재분류된 각음소의 대표적인 스펙트럼**

(c) /ㅜ/, /ㅡ/ (d) /ㅣ/

각 모음은 그림 4.6**과 같이** 2**개 이상의 대표 특징으로 분류할 수 있었다**. **이런 현상은 단일 화자가 같은 음소를 발음했을 때에도 나타나는 현상이다**. **결과적으로 한 음소에 대해 단일 음소 사전을 가졌던 기존 구성에 비해서 음소 사전이 표** 4.2**와 같이** 2**배 이상으로 증가하였으며**, **학습과 인식과정에서는 음소 사전의 증가로 인한 계산량의 증가**, **학습 및 인식 시간이 증가하게 되는데**, **인식 대역이 증가하여 인식률을 높이게 된다**.

표 4.2 실험을 위한 스펙트럼 특성에 의한 음소 사전의 재구성

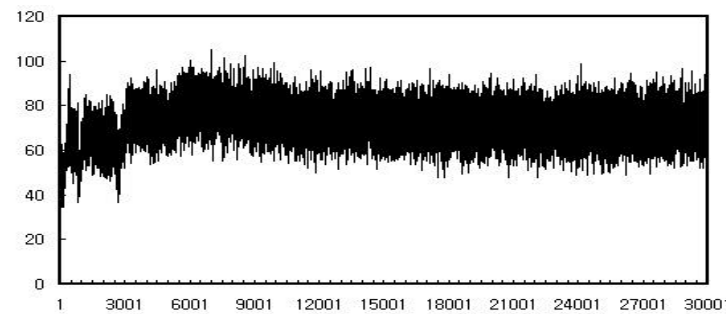
기존 분류명		제안된 분류명	갯수
아(aa)	⇒	a00, a01, a02	3
어(axr)		_o0, _o1	2
오(ao)		o00, o01	2
우(uh)		u00, u01	2
으(U)		_u0, _u1	2
이(ih)		i00, i01	2
에(eh)/애(ae)		e00, e01	2
8			15

전처리기로 사용한 확산필터뱅크를 통해 추출된 스펙트럼을 각각 5개 × 8음소, 6개 × 8음소, 10개 × 8음소를 다층 신경망 학습기 입력으로 하여 학습시킨다. 이 후 학습에 참가한 시료와 참가하지 않은 시료를 인식기에 입력하여 인식 결과를 얻고, 제안한 확산스펙트럼에 따라 재분류된 음소를 학습시켜서 인식 결과를 얻는다. 그리고 두 실험 결과를 다음 실험1과 2에서 얻어서 비교할 것이다.

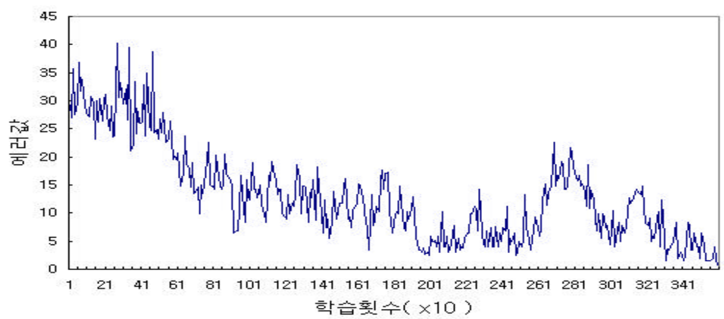
4.3. 실험 결과

4.3.1. 실험 1 - 음소패턴의 재분류 없는 기존의 방법

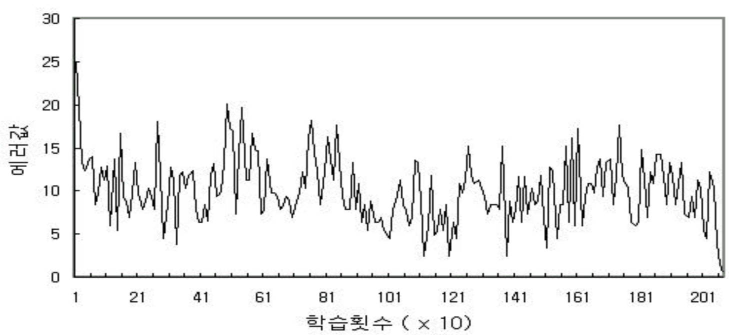
그림 4.7은 인식 실험을 위한 음소의 재분류 없이, 즉 한 음소에 대해 단일음소사전을 갖는 방법으로 학습을 완료한 신경망의 에러변화 그래프이다. 그림 4.7(a)는 제 1 신경망에서 학습했을 때의 에러변화 그래프이다. 이것은 지역국소점에서 망(network)이 안정이 되어 학습을 완료하지 못하는 것을 나타낸다. 그림 4.7(b) (d)는 제 2 신경망에 의해 학습을 보다 짧은 시간에 완료한 것을 보이고 있다.



(a)



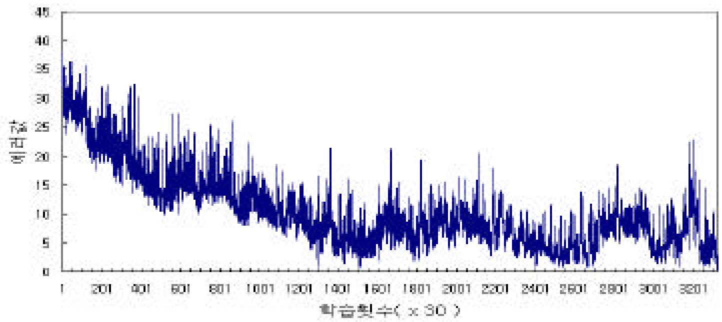
(b)



(c)

그림 4.7 **신경망의 에러 변화 그래프**(**계속**)

(a) **학습 미완료** (b) 5**개**×8**음소를 학습한 신경망** (c) 6**개**×8**음소를 학습한 신경망**



(d)

그림 4.7 (계속) 신경망의 에러 변화 그래프

(d) 10개×8음소를 학습한 신경망

표 4.3 학습에 참여한 시료의 인식 결과

	학습된 데이터 인식률(%)		
	5개×8음소	6개×8음소	10개×8음소
전체에러(≤0.5) / 데이터	0.02	0.00041	0.00567
아	100	100	100
에	100	100	100
오	100	100	100
어	100	100	100
애	100	100	100
이	100	100	100
으	100	100	100
우	100	100	100

표 4.3에서처럼 학습에 참여한 시료를 인식기의 입력으로 해서 인식 실험을 했을 때에는 인식결과가 모두 100%를 나타내었다. 실제 실험을

위해서는 학습에 참여하지 않은 시료만을 선택해서 시행되었고, 인식 실험 결과를 다음 표 4.4에 나타내었다. 인식기에 사용되는 신경망은 학습을 위해 무작위로 일정한 양으로 각 음소 데이터를 선택하여 학습을 완료하였고, 인식을 확인하기 위해 입력되는 각 음소의 데이터는 학습에 참여하지 않은 미학습 데이터이고 각각 100개씩 입력된다. 표 4.4 표 4.6는 확산망 필터 뱅크 출력 스펙트럼 결과의 분류 없이 학습과 인식 실험이 이루어진 결과이다. 표 4.4 표 4.6에서 인식불은 무반응과 과반응을 나타내는데 무반응은 데이터가 입력되고 출력 노드의 반응이 한 군데에서도 나타나지 않은 상태이고, 과반응은 출력 노드의 반응이 2개 이상 나타난 것을 의미한다. 실험에서 인식률은 전체적으로 30% 이하로 나타났으며, 이런 인식불가판단의 경우가 많이 나타나므로 우회 처리가 필요하다.

표 4.4 8음소 × 5개 데이터를 학습한 신경망의 인식 결과

출력 / 데이터	아	에*	오	어	이	으	우	인식불		인식률(%)
								무반응	과반응	
아	48	3	1	9	1	-	-	30	8	48
애	2	36	1	2	5	4	1	30	4	36
오	1	6	24	11	1	5	5	33	9	24
어	9	5	8	39	2	-	-	32	1	39
에	1	24	2	5	3	7	-	38	9	24
이	-	2	2	-	24	5	6	26	21	24
으	-	8	7	3	5	1	12	36	12	1
우	-	6	11	4	6	10	16	33	10	16
										26.50

에*는 /에/와 /애/의 대표음

표 4.5 8음소 × 6개 데이터를 학습한 신경망의 인식 결과

출력 / 데이터	아	에*	오	어	이	으	우	인식불 무반응	인식불 과반응	인식률(%)
아	41	1	1	8	-	-	4	31	13	41
에	-	16	4	-	2	5	-	47	10	16
오	-	-	38	2	-	2	15	28	11	38
어	7	3	14	36	1	-	4	25	10	36
애	-	18	4	-	-	5	3	46	12	18
이	-	5	-	-	20	8	3	20	37	20
으	-	3	4	1	5	19	17	33	10	19
우	-	7	6	2	5	5	19	31	12	19
										25.87

표 4.6 8음소 × 10개 데이터를 학습한 신경망의 인식 결과

출력 / 데이터	아	에*	오	어	이	으	우	인식불 무반응	인식불 과반응	인식률(%)
아	59	4	-	5	-	-	-	1	31	59
에	-	40	-	2	1	3	-	13	17	40
오	8	7	13	8	3	1	7	19	22	13
어	16	3	-	38	2	2	2	17	17	38
애	1	28	2	3	2	4	1	21	21	28
이	-	1	-	-	22	2	-	7	52	22
으	1	9	1	6	11	7	10	20	19	7
우	-	5	4	2	11	4	22	13	22	22
										28.63

4.3.2. 실험 2 - 음소패턴의 재분류에 의한 방법

다음은 확산망 필터 뱅크의 출력 스펙트럼 결과의 특성에 따라 음소를 세부 그룹으로 재분류한 실험이다. 그림4.5(a) (d)는 각각의 음소를 대표할만한 스펙트럼이며 분류되는 특징의 기준이 된다. 각각의 음소를 그림 4.5와 같은 특징에 따라 표 4.1과 같이 재분류하였고, 학습용으로 각 음소는 5개씩 선택하여 신경망 학습기를 통해 학습을 그림 4.8과 같이 완료하였다. 음소의 재분류 없는 인식실험에서 사용한 각각의 음소가 100개로 이루어진 미학습 시료를 입력으로 인식실험을 한 결과를 표 4.7에 나타낸다. 표에서 인식률에 괄호안의 수치는 과반응 중에서 출력이 인식을 하면서 과반응 때문에 에러로 간주 된 개수이다. 다시 말해 교사패턴과 같이 출력노드에서 반응이 보이며 거기에 다른 하나 이상의 출력노드에서 같이 반응을 보인 것이다.

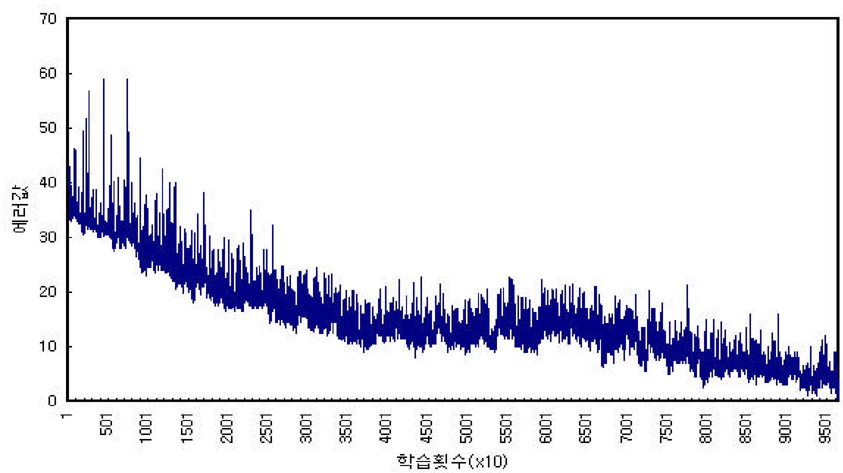


그림 4.8 재분류된 음소의 학습과정에 따른 에러 변화 그래프

표 4.7 재분류된 음소를 학습한 신경망의 인식 결과

출력 / 데이터	아	에	오	어	이	으	우	인식불		인식률
								무반응	과반응	
아	**74**	9	-	-	-	-	-	16	7	**74**(6)
에(애)	-	**49**	-	-	3	14	2	14	18	**49**(10)
오	2	-	**34**	35	-	-	-	12	11	**34**(11)
어	8	-	12	**55**	-	-	-	12	10	**55**(8)
이	-	4	-	-	**58**	8	10	8	12	**58**(10)
으	-	17	-	-	17	**25**	10	9	21	**25**(16)
우	-	18	-	-	15	20	**18**	14	24	**18**(6)
										44.7

표 4.8 실험1과 2의 결과 비교

	인 식 률						비고
	기존 방법			제안한 방법			
아	**59**	41	48	**74**	77	62	+15 %
에*	**34**	17.0	36	**49**	56	61	+15 %
오	**13**	38	24	**34**	10	10	+21 %
어	**38**	36	39	**55**	31	43	+17 %
이	**22**	20	24	**58**	24	24	+36 %
으	**7**	19	1	**25**	12	16	+18 %
우	**22**	19	16	**18**	32	27	-4 %
평균인식률	**27.86**	27.14	26.86	**44.71**	34.57	34.71	+16.85

위 표 4.8에서 '기존 방법'은 확산스펙트럼에 따른 음소의 재분류를 하지 않은채 각 음소에 단일 교사패턴을 가진 채 학습한 것이고, '제안한

방법'이란 재분류에 따른 음소사전을 2개 이상으로 증가하여 학습해서 인식 실험을 한 것을 말한다. 그리고 기존 방법의 평균 인식률이 실험 1의 결과와 다른 이유는, /ㅔ/와 /ㅐ/의 평균치를 위 표의 /ㅔ/에 대입했기 때문이다. 결과 비교를 쉽게 하기 위해 표에 음영처리 된 것을 보자. 각 방법 중 한가지 실험을 선택하여 비교한 것이다. 여기서 인식률의 증가를 볼 수 있는데 /ㅏ/의 경우는 15% 향상되었고, /ㅔ/는 15% 증가하였다. 이때 /ㅔ/는 /ㅔ/와 /ㅐ/의 대표음이다. /ㅗ/는 21%, /ㅓ/는 17%, /ㅣ/는 36%, /ㅡ/는 18% 증가, 그리고 /ㅜ/는 4% 감소된 결과를 보였다. 전체 인식률은 16.85% 향상되었다. 따라서 제안한 방법의 인식률이 향상되었음을 확인하였다.

5. 결　론

본 연구에서는 학습의 시간을 줄이고, 효율적으로 완료하기 위해서 제 2 신경망을 제안하였고, 인식률을 높이기 위해 확산필터의 스펙트럼 특성에 따른 음소의 재분류를 제안하였다.

제 2 신경망은 제 1 신경망의 학습진행 경과를 관찰하여 초기조건을 추출하여 사용한다. 기존의 일반적인 BP 신경망의 단점을 보완한 제 2 신경망은 학습 시간을 줄였고, 모든 학습을 원하는 에러 값(1.0) 이하로 완료하였다. 확산필터의 스펙트럼에 따른 음소의 재분류는 학습과 인식 과정에서 작업량을 증가하는 반면에 인식률을 높이게 된다. 같은 발음에 대해 스펙트럼 특징을 분류하여 음소 사전의 양이 증가하여 학습의 시간과 계산량이 증가하였지만, 한 음소에 대한 인식 가능한 대역이 넓어져서 인식률이 높아졌다.

인식결과를 확인해보면 /ㅏ/와 /ㅔ/의 경우는 15%, /ㅗ/는 21%, /ㅓ/는 17%, /ㅣ/는 36%, /ㅡ/는 18% 증가, 그리고 /ㅜ/는 4% 감소된 결과를 보여서 평균 인식률은 16.85% 향상되었다. 인식률이 기존의 단일 음소사전을 갖는 각 음소를 통해 학습된 인식률보다 전체적으로 상승되어 제안된 방법이 인식률을 높일 수 있다는 것을 실험을 통해 확인하였다.

참 고 문 헌

[1] 임현, 박순영, 방만원 등, “HMM 집단화 방법을 이용한 불특정화자 단어인식”, 제6회 신호처리합동 학술대회 논문집, Vol. 6, No. 1, pp. 757 760, 1993.

[2] 김성낙, 남시병, 이상범, “음정단위를 이용한 한국어 연속단어 인식”, 대한전자공학회논문지, Vol.28-B, No. 8, pp. 9 13, 1991.

[3] 이형근, 조용수, 양원영, "유색 잡음 환경에서의 한국어 단모음 인식", 제6회 신호처리 합동학술대회 논문집, Vol. 6, No.1 pp. 748 751, 1993.

[4] 이영호, 정 홍, “음절을 기반으로 한 한국어 음성인식”, 대한 전자공학회논문지, Vol.31-B, No. 1, pp. 11 22, 1994.

[5] 은종관 외, “음성 인식 기술의 최근 동향과 국내 연구 개발 현황”, 음성통신 및 신호처리 workshop, pp. 21 62, 1990.

[6] D. R. Reddy, "Computer Recognition of Connected Speech," J. Acoust. Soc. Amer., Vol. 42, pp. 329 347, Aug. 1967.

[7] 권율, 남기곤, 윤태훈, 김재창, “확산뉴런망을 이용한 밝기 변화 추출”, 전자공학회논문지, 제29권 B편, 제11호, pp. 85-92, 1992.

[8] 권율, 남기곤, 윤태훈, 김재창, “ 확산 뉴런망을 이용한 영상처리”, 전자공학회논문지, 제30권 B편, 제1호, pp.90-98, 1993.

[9] 허만탁, “확산망과 차분연산에 의한 대역통과필터 구현과 특성해석”, 부산대학교, 공학박사학위논문, 1994.

[10] 강정희, “확산망을 이용한 필터뱅크의 구현”, 부산공업대학교 공학석사학위논문, 1995.

[11] 석칠호, “확산필터뱅크를 이용한 한국어 단모음 인식”, 부산공업대학교 공학석사 학위논문, 1996.

[12] 이종혁, “ $\nabla^2 G$ 연산자에 의한 신호 분석 및 음성인식을 위한 신경회로망의 구현”, 부산대학교, 공학박사학위논문, 1991.

[13] 이종혁, 정용근, 남기곤, 윤태훈, 김재창, 박의열, 이양성, “ $\nabla^2 G$ 연산자의 신호분석 특성을 이용한 음성인식 신경회로망에 관한연구”, 전자공학회논문지, 제29권 B편, 제10호, pp. 90-98, 1992.

[14] 이웅용, 조창호, 등, “DTW를 이용한 200 단어 실시간 인식 시스템의 구현”, 제6회 신호처리합동 학술대회 논문집, Vol. 6, No. 1, pp. 739 742, 1993.

[15] 조영광, 강철호, “Self-Organizing Feature Map과 ART모델을 이용한 단어인식”, 제6회 신호처리합동 학술대회 논문집, Vol. 6, No. 1, pp. 743 747, 1993.

[16] 정영호, 김종교, “TDNN을 이용한 한국어 숫자음 인식,” 제6회 신호처리합동 학술대회논문집, Vol. 6, No. 1, pp. 769 772, 1993.

[17] 이종석, 이상욱, “신경망과 구문분석을 이용한 한국어연결 숫자음인식”, 대한전자공학회논문지, Vol. 30-B, No. 12, pp. 21 30, 1993.

[18] 김순협, “음성인식 기술현황 및 실용화 전망”, 한국음향학회지, Vol. 13, No 2, pp. 86 95, 1994.

[19] 패턴 인식론, 오영환저, 정익사, pp. 169 264.

[20] 현대 국어음운론, Linear & Non-linear phonology, 오정란저, 형설출판사, pp. 3 14.

[21] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Leaning Internal Representations by Error Propagation," in D. E. Rumelhart and J. L. McClelland (Eds.) Parallel Distributed Processing : Explorations in the Microstructure of Cognition, Vol. 1: Foundations, MIT Press, pp. 318 362, 1986.

[22] Hyun-Yeol Chung, Shozo Makino, and Ken'iti Kido, "Analysis

and Recognition of Isolated Korean Vowels using Formant Frequency", J. Acoust. Soc. Jpn. (E) Vol. 9, No. 5, pp. 225 232, 1988.

[23] S. B. Dames and P. Mermelstein, "Comparision of Parametric Representations for Monosyllable Word Recognition in Continuously Spoken Sentences, " IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-28, No. 4, Aug. 1980.

[24] L. R. Rabiner and R. W. S chafer, "Digital Processing of Speech Signals", Englewood Cliffs, N. J. : Prentice Hall, 1978.

[25] L. R. Rabiner and B. H. Juang, "Fundamentals of Speech Recognition", Prentice Hall, 1993.

[26] **전용구**, **양진우**, **김순협**, “SOFM**신경회로망을 이용한 한국어 음소인식**”, **한국음향학회지**, Vol. 14, No. 2, pp. 101 112, 1995.

[27] **合原一辛 편저**, **정 호선**, **여 진경 공역**, **뇌와 카오스**, Ohm**사**. 1994.

[28] Brain Lab. **뇌전위와 카오스**, **한국과학기술원 물리학과 센서 연구실**, Sep. 1. 1995.

[29] Scott E. Fahlman, 'An Empirical Studyof Learning Speed in Back-Propagation Networks, *CM U- CS-88-162 '*, 1988.

[30] Martin Riedmiller/ Heinrich Braun, 'RPROP-A Fast Adaptive Learning Algorithm', Proceedings of ISCIS VII.

[31] **장종칠**, “**확산필터뱅크와 신경망을 이용한 한국어의 음소단위 복모음 인식에 관한 연구**”, **부경대학교 공학석사 학위논문**, 1999.

[32] Freeman, James A. "Simulating Neural Networks", Addison-Wesley, pp. 67 114, 1994.