화자 독립 음성 인식을 위한 반연속 HMM과 RBF의 혼합 구조에 관한 연구

A Study on Hybrid Structure of Semi-Continuous HMM and RBF for Speaker Independent Speech Recognition

문 연 주*, 전 선 도*, 강 철 호*
(Yun Joo Moon*, Sun Do June*, Chul Ho Kang*)

※ 이 연구는 1999년도 광운대학교 교내 학술연구비 지원으로 수행된 것임.

요 약

성 인식 알고리즘에서 높은 인식률을 보이는 방법은 hidden Markov model(HMM)과 신경망의 혼합 형태이다. 이것은 통계적인 모델과 신경망 모델의 장점을 혼용하는 방법이다. 본 연구에서 제안하는 인식 알고리듬은 반연속 HMM과 radial basis function(RBF)의 새로운 형태의 혼합 구조로써 반연속 HMM 파라미터 중에서 관측 확률을 결정하는 가중치(혼합확률밀도함수계수)확률을 Baum-Welch 추정 이후 RBF로써 재 추정하는 인식 모델을 제안한다. 제안한 방법은 RBF의 은닉층(hidden layer)의 기본 함수(basis function)와 반연속 HMM의 확률 밀도 함수의 유사함을 고려한 것으로 RBF의 학습 및 추정된 가중치로써 보다 음성 파형을 분별력 있게 구분하고자 하는 것이다. 모의 실험 결과는 반연속 HMM만을 사용 할 때 보다 제안한 반연속 HMM/RBF 혼합 구조가 비 학습 화자에 대한 인식률을 개선함으로써 단순히 반연속 HMM만을 사용하는 것 보다 훨씬 분별력이 높은 방법임을 보여준다.

ABSTRACT

It is the hybrid structure of HMM and neural network(NN) that shows high recognition rate in speech recognition algorithms. And it is a method which has majorities of statistical model and neural network model respectively. In this study, we propose a new style of the hybrid structure of semi-continuous HMM(SCHMM) and radial basis function(RBF), which re-estimates weighting coefficients probability affecting observation probability after Baum-Welch estimation. The proposed method takes account of the similarity of basis function of RBF's hidden layer and SCHMM's probability density functions so as to discriminate speech signals sensibly through the learned and estimated weighting coefficients of RBF. As simulation results show that the recognition rates of the hybrid structure of SCHMM/RBF are higher than those of SCHMM in unlearned speakers' recognition experiment, the proposed method has been proved to be one which has more sensible property in recognition than SCHMM.

I. 서 론

현재 연구되고 있는 음성 인식 방법으로는 신경망에 의한 인식, 시간적인 정합을 이용한 DTW(dynamic time warping)알고리즘, 확률적인 방법으로 알려진 HMM[1] 등이 있다. 그 중에서도 HMM은 음성의 시간적인 변이성을 통계적인 모델에 적용함으로써 높은 인식률을 보이며 신경망과 함께 결합하여 다양하게 연구가 되고 있다.

HMM은 인식방법에 있어서 비학습 화자에 대해서인식 실험 할 경우 인식률이 저하되는 문제를 해결하기 위하여 이산 HMM (discrite HMM), 연속 HMM(continuous HMM)과 신경망을 혼합한 형태를 사용한다. 이렇게 사용하는 신경망 알고리즘으로는 RBF와 TDNN(time delay neural network) 등이 있는데 TDNN을 사용하는 경우에는 TDNN을 연속 HMM에 적용할 때 두개의 독립된 구조로 나눠 학습 과정을 통해 파라미터 값을 구한다. 여기서 각각 얻은 양쪽의 파라미터를 선형적으로 결합하는 방식이다[2]. 또 다른 방법으로는 TDNN을 사용하여 코드북 생성 후 이산 HMM과 혼합한 것이 있다[3]. 그리고 RBF를 사용한 경우는 RBF의 출력값으로 관측 파라미터를 결정

* 광운대학교 전자통신공학과
접수일자 : 1999년 9월 1일

한 후 HMM에 적용한다[4].

본 논문에서는 반연속 HMM과 RBF의 새로운 혼합 구조로써 반연속 HMM의 가중치 확률 (혼합 확률 밀도 계수 b_{jl} :상태 j에서 l번째 가우시안 성분의 상대적인 크기)를 추정 후 이것을 RBF의 목표값(desired value)에 대응시켜 LMS 알고리즘을 이용해 RBF의 가중치를 학습 및 추정하는 방식을 제안한다. 제안한 방식은 반연속 HMM에 대해 보완적인 구조로써 보다 분별력 있게 가중치 확률을 갖게 된다. 모의 실험에서 화자 독립 인식 실험을 하였을 때 실험 결과 비학습 화자에 대해서 4% 정도 인식률 증가를 얻을 수 있었다.

이것은 단순히 반연속 HMM 만을 사용한 것보다 제안한 방식이 더욱 분별력이 있음을 보여주는 것이다.

II. 기존의 반연속 HMM 및 RBF 신경망

2.1 기존의 반연속 HMM

이산 HMM보다 작은 코드북을 사용하고 연속 HMM보다 적은 계산량이 필요하도록 두 경우를 결합한 것을 반연속 HMM이라 한다[5][6][7].

반연속 HMM은 벡터 양자화 코드워드를 가우시안 분포의 평균치들로 생각하며, 각 분포의 공분산 행렬의 대각선값들을 코드북에 포함시키게 된다. 즉, 크기 L인 코드북에서 각 코드워드에 해당하는 D차 평균값 μ와 공분산 행렬의 주대각선 성분 $\sum$ 가 D개 주어지게 된다. 각 코드워드마다 공분산 행렬값이 주어지므로 일반적인 벡터 양자화의 경우와는 달리 유클리디안 (Euclidean) 거리 대신 마할라노비스(Mahalanobis)거리를 사용하게 된다.

확률 밀도 행렬 B는 상태 j에서 l번째 코드워드에 해당하는 가우시안 성분을 발견할 상대적인 크기가 되므로 확률 밀도 행렬 요소 b_{jl}은 연속 밀도 HMM의 C_{jl}(상태 j에서 l번째 가우시안 성분의 상대적인 크기) 와 같은 역할을 한다. 그러면 상태 j에서 관찰값 O_t를 발견할 확률은

$$b_j(o_t) = \sum_{l=1}^{L} b_{jl} p(o_t / \mu_l, \sum_l)^r \quad (1)$$

$b_j(o_t)$: 관측확률, b_{jl} : 가중치 확률

$p(o_t/\mu_l,\Sigma_l)$: 확률 밀도 함수

로 주어진다.

이와 같은 반연속 HMM을 이용하여 음성의 학습 데이터를 잘 표현하기 위해서는, 반연속 HMM의 파라미터 재추정(parameter reestimation) 과정이 필요하다. 이것은 파라미터가 주어졌을 때, 관찰열을 발전할 확률을 반복적으로 최대화시키는 것으로서 EM(Expectation Maximization) 알고리즘이라 한다. 또한 주어진 반연속 HMM 파라미터들로부터 하나의 관찰열에 대응되는 가장 적합한 상태열을 찾는 방법으로 Viterbi 알고리즘이 있다.

반연속 HMM에서 재추정해야할 변수들은 초기 행렬, 상태 천이 A행렬, 출력 확률 행렬 B, 코드북의 μ, $\sum$ 값들이다. 이들은 다음의 재추정식으로부터 구한다.

$$\pi_i = \gamma_1(i) \quad (2)$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \gamma_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (3)$$

$$b_j(k) = \frac{\sum_{t=1}^{T-1} \zeta_t(j,k)}{\sum_{t=1}^{T-1} \gamma_t(j)} \quad (4)$$

$$\mu_j = \frac{\sum_{t=1}^{T-1} \zeta_t(j)\, o_t}{\sum_{t=1}^{T-1} \zeta_t(j)} \quad (5)$$

$$\sum_j = \frac{\sum_{t=1}^{T-1} \zeta_t(i)(o_t - \mu_j)(o_t - \mu_j)'}{\sum_{t=1}^{T-1} \zeta_t(j)} \quad (6)$$

아래 식에서 중간 변수는 식(7),(8),(9),(10)이다.

$$\gamma_t(i,j) = P(s_t = i,\ s_{t+1} = j \mid O, \lambda) \quad (7)$$

$$\gamma_t(i) = P(s_t = i, \mid O, \lambda) \quad (8)$$

$$\zeta_t(i,k) = P(s_t = i,\ o_t = v_k \mid O, \lambda) \quad (9)$$

$$\zeta_t(k) = P(o_t = v_k \mid O, \lambda) \quad (10)$$

2.2 RBF 신경망 개요

RBF 신경망은 Broomhead와 Lowe에 의해 제안되었으며 그림 1과 같이 은닉층과 출력층의 2층으로 구성된 전방향의 지도 학습 알고리즘을 갖는 신경망이다[8]. 기본적인 구조는 이층 퍼셉트론과 비슷하지만 이층 퍼셉트론과는 달리 은닉층은 단 한층으로만 구성되며 입력층과 은닉층 사이의 연결은 가중치를 갖지 않고 입력값을 그대로 받아들인다. 또 은닉층의 활성화 함수는 시그모이드 함수대신 여러 형태의 radial basis함수로 구성된다.

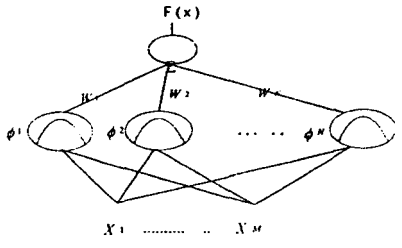

그림 1. RBF 신경망 구조
Fig. 1. RBF network structure.

은닉층의 각 노드는 중심점(center)이라고 불리는 벡터를 포함하고 있는데 이 중심점들은 입력 벡터 세트를 대표하는 벡터들이다. 일반적으로 RBF 신경망의 은닉층 출력은 주로 식 (11)와 같이 가우시안 함수를 활성화 함수로 하여 얻어진다.

$$\Phi_i(x) = \Phi(\ \|x - c_i\|^2 / \rho_i) = \exp(-\|x - c_i\|^2 / \rho_i) \tag{11}$$

for i=1, 2, ... N

N은 중심점의 갯수, ci 는 RBF의 중심점, ρ_i 는 distance scaling 파라미터이다. 그리고 입력벡터와 출력사이의 비선형 함수의 근사는 식 (12)와 같이 비선형 basis 함수 Φ_i 의 선형 조합으로 표현된다.

$$F(x) = \sum_{i=1}^{N} \omega_i \Phi_i(x) \tag{12}$$

은닉층과 출력층 사이의 선형 가중치 ω_i 는 식 (13)과 같은 최소 자승 오차(LMS) 알고리즘을 이용하여 적응된다.

$$\begin{aligned} &\Phi_i(x, t) = \exp(-\|x(t) - c_i(t)\|^2 / \rho_i), \\ &e(t) = d(t) - \sum_{i=1}^{N} \omega_i(t-1)\ \Phi_i(x, t), \\ &\omega_i(t) = \omega_i(t-1) + g_w e(t)\ \Phi_i(x, t) \\ &\text{for} \quad 1 \le i \le N \end{aligned} \tag{13}$$

여기서 d(t)는 현재의 입력 센터에 대한 목표 출력, g_w 는 가중치에 대한 학습 계수이다. 기본적인 RBF 신경망에서 중심점 ci와 distance scaling 파라미터 ρ_i 는 고정되며 단지 가중치 ω_i 만 적응시킨다. 실험적으로 충분한 수의 은닉층 노드를 갖고 중심점이 입력 차원에 적절히 분포되어 있다면 RBF 네트워크는 넓은 범위의 비선형 함수로 근사 시킬 수 있다. .

III. 제안한 반연속 HMM과 RBF의 혼합구조

3.1 학습 과정

이 논문에서 제안한 전체 구조는 학습 과정과 인식 과정으로 나뉘어지는데 학습 과정의 전체적 모식도는 그림 2와 같다.

위 그림에서 보듯이 입력 데이타를 반연속 HMM과정과 RBF의 두 과정을 거쳐 입력 패턴에 대해 보다 더 분석적인 과정을 갖도록 한다. 먼저 학습 과정 부분에서는 기존의 반연속 HMM과 마찬가지로 L개의 가우시안 확률 밀도들과 각 가우시안 확률 밀도들의 가중치를 결정하는 혼합 밀도 계수에 의해 입력 음성의 특징을 확률적으로 모델링하는 혼합확률을 얻어 Maximum likelihood와 Baum-Welch 알고리즘을 이용해 초기확률, 천이확률, 가중치확률,

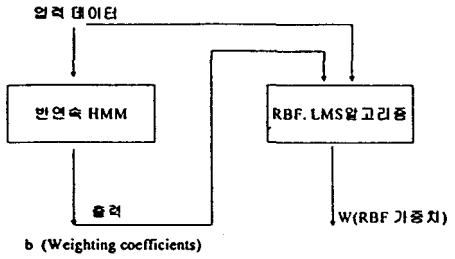

그림 2. 제안한 HMM/RBF 학습 과정 모식도
Fig. 2. Learning block diagram of hybrid HMM/RBF.

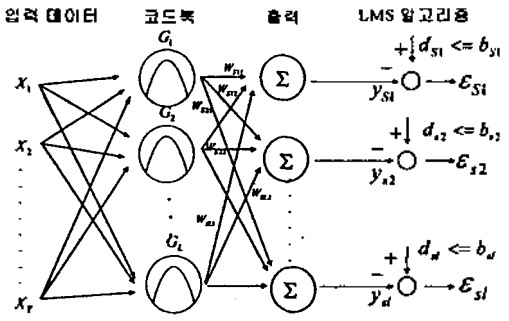

그림 3. RBF 학습 모식도
Fig. 3. Learning block diagram of RBF.

평균벡터 μ, 공분산 행렬 Σ을 학습해 나간다. 그림 2에서 가중치 확률(혼합 확률 밀도 계수 b)을 RBF의 목표값으로 해서 LMS 알고리즘의 적응화 과정을 통해 새로운 파라미터, RBF의 가중치(Weights)를 구한다.

그림 3은 RBF의 새로운 가중치 추정을 위한 학습과정으로서 수식을 전개하면 다음과 같다.

$$\begin{aligned} G_1 &= \sum_{t=1}^{T} \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(X_t - \mu_1)' \Sigma^{-1} (X_t - \mu_1)\right] \\ G_2 &= \sum_{t=1}^{T} \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(X_t - \mu_2)' \Sigma^{-1} (X_t - \mu_2)\right] \\ &\cdot \\ &\cdot \\ G_l &= \sum_{t=1}^{T} \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(X_t - \mu_l)' \Sigma^{-1} (X_t - \mu_l)\right] \\ &\cdot \\ &\cdot \\ G_L &= \sum_{t=1}^{T} \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(X_t - \mu)' \Sigma^{-1} (X_t - \mu)\right] \end{aligned}$$

l : l번째 코드워드 (14)

X_t : d차 행 입력 데이터, μ : d차 행 평균 벡터

Σ : d×d 차 공분산, Σ^{-1} : Σ 의 역 공분산

RBF의 출력값은 다음 식과 같이 표현된다.

$$y_{s1} = \sum_{l=1}^{L} G_l * W_{sl} \quad , y_{s2} = \sum_{l=1}^{L} G_l * W_{s2} \quad \cdots \quad y_{sL} = \sum_{l=1}^{L} G_l * W_{sL}$$

y_{sl} : 출력값, W_{sl} :가중치, l : l 번째 코드워드 (15)

HMM을 통해 학습된 결과 값 b와 RBF의 학습을 통해 나온 결과 값 y의 차(ϵ_{sl})를 구한다.

$$\epsilon_{s1}=y_{s1}-d_{s1},\ \epsilon_{s2}=y_{s2}-d_{s2}\ \ldots\ \epsilon_{sL}=y_{sL}-d_{sL}$$
ϵ_{sl} : 에러값 (16)

그 다음 LMS 알고리즘을 이용해 W_{sll} 을 구할 수 있다.

$$W_{sll} = W_{sll} + \text{power}^* \epsilon_{sl}^* G_l \quad (17)$$
$$power=\frac{1}{\sum_{l=1}^{L} G_l * G_l}$$

3.2 인식 과정

인식 과정에서는 그림 4에서 보듯이 인식할 단어의 LPC 계수와 학습 과정을 통해 얻은 RBF의 가중치(W_{sll})가 RBF 입력 데이터로 쓰여 그 결과 나온 출력 값(y_{sl})이 Viterbi 인식 과정의 혼합 확률 밀도 계수(b_{jl})대신 사용된다.

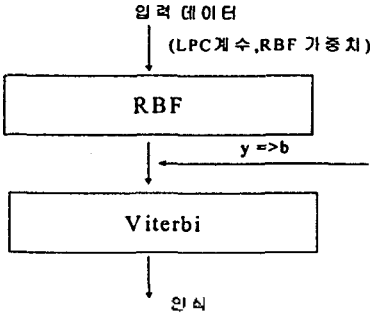

그림 4. 제안한 HMM/RBF 인식과정 모식도
Fig. 4. Recognition block diagram of hybrid HMM/RBF.

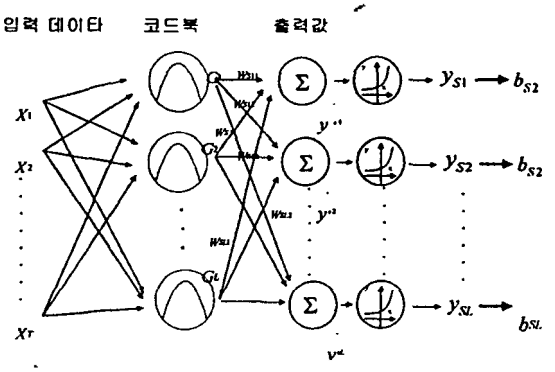

그림 5. RBF 인식 모식도
Fig. 5. Recognition block diagram of RBF.

그림 5에 나타난 수식을 전개하면 다음과 같다.
식 (14), (15)와 동일한 과정을 거쳐 그 결과 나온 값 y'_{sl} 은 양수와 음수 값으로 나타나는데 y'_{sl} 중에서 최소 값이 음수값이므로 그 음수값을 각각 성분에 대해 최소 값을 빼서 y'_{sl} 값을 양수값으로 정규화 시킨다. 그 다음 코드워드의 확률값 분포를 좀더 세밀히 구분하기 위해 식(18) 과 같은 과정을 밟는다. 기본적으로 시그모이드와 같은 함수를 이용하나 실험을 통해서 시그모이드 함수보다 아래에 쓰인 함수가 인식률에 더 적합하므로 다음 식을 이용한다.

$$y_{s1}=\alpha^*(y'_{s1})^k,\ y_{s2}=\alpha^*(y'_{s2})^k\ \ldots\ y_{sL}=\alpha^*(y'_{sL})^k \quad (18)$$

본 논문에서는 실험 결과 α=10, k=5로 하였다. 위 식 결과 y_{sl} 의 모든 값을 더해서 1 이 되어야 하므로 y_{sl} 값을 모두 합해서 각각 성분에 나눔으로써 확률값을 얻을 수 있다.

위 결과 구한 y_{sl}을 Viterbi 인식 과정에서 필요한 파라미터 값(b_{jl})대용으로 취한다.

IV. 실험 결과

본 논문에 이용된 음성 데이터의 샘플링주파수는11.025kHz 이고 Preemphasis는 $H(z)=1-0.95z^{-1}$ Frame Blocking은 256 speech sample을 128개씩 천이하고 14차 LPC 켑스트럼 계수를 이용하여 한국어에 존재하는 모든 자음과 모음을 k-means 알고리즘을 이용해 64개의 코드워드로 구성된 코드북 벡터를 생성하였다. 학습에 사용된 단어는 13명의 남성화자가 한국어 15개 고립 단어를 두번씩 발성한 390개이고 테스트에 사용된 단어는 학습 화자에는 위에서 언급한 390개이고 비학습 화자에는 10명 화자가 15개 고립 단어를 두번씩 발성한 300개를 이용하였다. 또한 실험에 사용된 인식 모델은 반연속 HMM이고 반연속 HMM의 상태 갯수를 10개로 하고 left to right 모델이 적용되었다.

4.1 기존의 반연속 HMM에 대한 실험 결과

본 논문에서 학습에 참여한 13명과 학습에 참여하지 않은 10명을 인식 실험에 테스트하였다.

표 1에 보듯이 학습 화자는 93.3%인식률을 보이고 표 2에 비학습 화자는 84.6% 인식률을 보인다. 학습 화자와 비학습 화자의 인식률을 비교 할 경우 새로운 불특정 화자에 대해 인식률이 저조한 것을 알 수 있다.

4.2 제안한 반연속 HMM/RBF에 대한 실험 결과

학습에 참여한 13명과 학습에 참여하지 않은 10명에 대해 제안한 방식으로 테스트 결과 표 3, 표 4 에 알 수 있듯이 학습 화자의 경우 기존 방식보다 제안한 방식이 평균 0.8% 인식률의 상승이 있으나 비학습 화자의 경우 특이성으로 관측되지 않은 심벌에 대해서도 인식이 가능하므로 인식 실험 결과 평균 4%정도 인식률이 상승하였다.

표 1. 학습화자에 대한 기존의 반연속HMM 인식률
Table 1. The recognition rate of a conventional SCHMM for learned speaker.

인식 / 화자	서울	인천	수원	의정부	포천	춘천	원산	해주	개성	평기	강원	충남	제주	한국	미국	계	인식률 (%)
jml	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	28/30	93.3
jhk	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	30/30	100.0
mjl	2/2	2/2	0/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	28/30	93.3
hkd	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	1/2	1/2	2/2	2/2	28/30	93.3
hyh	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	28/30	93.3
khw	2/2	2/2	0/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	27/30	90.0
kdw	2/2	2/2	0/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	28/30	93.3
kmj	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	30/30	100
kmy	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	28/30	93.3
kss	2/2	2/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	29/30	96.7
pkj	2/2	2/2	1/2	2/2	2/2	2/2	1/2	2/2	2/2	1/2	1/2	2/2	2/2	2/2	2/2	26/30	86.7
shj	0/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	1/2	0/2	2/2	2/2	2/2	2/2	2/2	25/30	83.3
jhc	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	1/2	29/30	96.7
계	24/26	26/26	14/26	26/26	26/26	26/26	25/26	25/26	22/26	22/26	25/26	26/26	25/26	25/26	25/26	344/390	93.3
인식률 (%)	92.3	100.0	53.8	100.0	100.0	100.0	96.1	96.1	84.6	84.6	96.1	100.0	96.1	96.1	96.1	93.3	

표 2. 비학습 화자에 대한 기존의 반연속HMM 인식률
Table 2. The recognition rate of a conventional SCHMM for unlearned speaker.

인식 / 화자	서울	인천	수원	의정부	포천	춘천	원산	해주	개성	평기	강원	충남	제주	한국	미국	계	인식률 (%)
jmj	2/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	1/2	1/2	1/2	2/2	2/2	2/2	2/2	26/30	86.6
kjh	2/2	2/2	1/2	2/2	2/2	2/2	2/2	0/2	0/2	1/2	2/2	2/2	1/2	2/2	2/2	23/30	76.6
kbw	2/2	2/2	2/2	1/2	1/2	2/2	2/2	1/2	2/2	1/2	2/2	2/2	1/2	1/2	2/2	25/30	83.3
kch	2/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	0/2	0/2	0/2	2/2	2/2	2/2	2/2	24/30	80.0
kkj	2/2	2/2	2/2	2/2	2/2	1/2	2/2	2/2	2/2	0/2	2/2	2/2	1/2	2/2	1/2	26/30	86.6
kmy	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	2/2	2/2	2/2	1/2	2/2	27/30	90.0
kjh	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	0/2	0/2	2/2	2/2	2/2	2/2	2/2	26/30	86.6
lyc	1/2	2/2	0/2	2/2	2/2	2/2	2/2	2/2	0/2	2/2	2/2	2/2	2/2	2/2	2/2	25/30	83.3
nhw	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	0/2	1/2	2/2	1/2	2/2	25/30	83.3
ssj	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	2/2	1/2	1/2	2/2	27/30	90.0
계	19/20	19/20	16/20	18/20	18/20	20/20	20/20	17/20	9/20	10/20	15/20	19/20	17/20	15/20	19/20	254/300	84.6
인식률 (%)	95.0	95.0	80.0	90.0	90.0	100.0	100.0	85.0	45.0	50.0	75.0	95.0	85.0	75.0	95.0	84.6	

표 3. 학습화자에 대한 제안한 HMM/RBF의 인식률
Table 3. The recognition rate of a proposed HMM/RBF for learned speaker

인식 / 화자	서울	인천	수원	의정부	포천	춘천	원산	해주	개성	평기	강원	충남	제주	한국	미국	계	인식률 (%)
jml	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	0/2	1/2	2/2	2/2	2/2	2/2	2/2	26/30	86.3
jhk	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	30/30	100.0
mjl	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	29/30	96.7
hkd	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	29/30	96.7
hyh	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	29/30	96.7
khw	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	2/2	1/2	1/2	2/2	2/2	28/30	93.3
kdw	2/2	2/2	0/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	28/30	93.3
kmj	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	30/30	100
kmy	2/2	2/2	0/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	28/30	93.3
kss	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	31/30	98.6
pkj	2/2	2/2	1/2	2/2	2/2	2/2	1/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	30/30	96.7
shj	1/2	2/2	2/2	2/2	2/2	2/2	2/2	1/2	1/2	2/2	2/2	2/2	2/2	2/2	2/2	27/30	90.0
jhc	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	30/30	100.0
계	25/26	26/26	14/26	26/26	26/26	26/26	25/26	25/26	20/26	24/26	24/26	24/26	24/26	24/26	24/26	347/390	95.5
인식률 (%)	96.1	100.0	61.5	100.0	100.0	100.0	96.1	96.1	76.9	92.3	92.3	100.0	100.0	92.3	100.0	94.1	

표 4. 비학습 화자에 대한 제안한 HMM/RBF의 인식률
Table 4. The recognition rate of a proposed HMM/RBF for unlearned speaker.

인식 / 화자	서울	인천	수원	의정부	포천	춘천	원산	해주	개성	평기	강원	충남	제주	한국	미국	계	인식률 (%)
jmj	2/2	1/2	2/2	2/2	2/2	2/2	1/2	2/2	1/2	1/2	2/2	2/2	2/2	2/2	2/2	25/30	83.3
kjh	2/2	2/2	1/2	1/2	2/2	2/2	2/2	1/2	0/2	1/2	2/2	1/2	1/2	2/2	2/2	25/30	83.3
kbw	2/2	2/2	2/2	1/2	1/2	1/2	2/2	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	26/30	86.6
kch	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	0/2	0/2	2/2	2/2	2/2	2/2	2/2	26/30	86.6
kkj	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	1/2	1/2	2/2	2/2	1/2	2/2	2/2	27/30	90.0
kmy	2/2	2/2	1/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	1/2	2/2	1/2	2/2	2/2	25/30	83.3
kjh	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	0/2	0/2	2/2	2/2	2/2	2/2	2/2	26/30	86.6
lyc	2/2	2/2	0/2	2/2	2/2	2/2	2/2	2/2	1/2	1/2	2/2	2/2	2/0	2/2	1/2	26/30	86.6
nhw	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	2/2	2/2	2/2	1/2	2/2	25/30	83.3
ssj	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	1/2	2/2	2/2	2/2	2/2	1/2	2/2	25/30	83.3
계	20/20	19/20	14/20	19/20	19/20	19/20	19/20	19/20	8/20	10/20	19/20	20/20	18/20	18/20	20/20	255/300	85.0
인식률 (%)	100.0	95.0	70.0	95.0	95.0	95.0	95.0	95.0	40.0	55.0	95.0	100.0	90.0	90.0	100.0	88.3	

표 5. 기존 방식과 제안한 방식의 인식률 비교
Table 5. The comparison of recognition rate for a conventional method and a proposed method.

	인식률(%)	
학습화자	기존의 방식	93.3
	제안한 방식	94.1
비학습화자	기존의 방식	84.6
	제안한 방식	88.3

V. 결 론

본 연구에서 제안한 방법의 구조는 기존의 반연속 HMM과 신경망의 새로운 혼합 형태를 갖는다. 이러한 방법은 기존의 반연속 HMM 만을 사용했을 때보다 화자 독립 인식 실험에 적용했을 경우 비 화자 학습 대해서 인식률의 향상을 보여주었다. 이러한 결과는 제안한 반연속 HMM/RBF 혼합 구조에서 가중치 확률(b_{jk})이 RBF의 신경망 구조로 표현됨으로써 반연속 HMM만을 사용하는 경우보다 분별력이 높은 형태가 됨을 알 수 있다.

결국 이러한 혼합 구조가 기존의 반연속 HMM의 통계적 모델 중 관측 확률을 신경망의 학습능력으로 재표현함으로써 반연속 HMM 만을 사용한 것 보다 불특정 화자의 음성 인식에 상당한 향상을 가져오게 된 것이다.

참 고 문 헌

1. Lawrence Rabiner, Biing-Hwang Juang, *Fundamentals of Speech Recognition*, PrenticeHall International Inc, pp.321-389 , 1993
2. Christian Dugast, Laurence Delvillers, and Xavier Aubert, "Combining TDNN and HMM in a Hybrid System for Improved Continuous-Speech Recognition," IEEE Transactions on Speech and Audio Processing, Vol .2 No.1 Part II, January 1994
3. 방영조, " TDNN 과 HMM을 결합한 새로운 단어 방식에 관한 연구 ", 신호처리합동학술대회, 제 4권 제 1호 1991
4. Eliot Singer, Richard P. Lippmann, "A Speech Recognition Using Radial Basis Function Neural Networks In an HMM Framework," Proc. ICASSP, pp.629-632, 1992.
5. X.D Huang, Y. Ariki, M.A. Jack. *Hidden Markov Models for Speech Recognition* Edinburgh University Press, 1990
6. X.D. Huang, M.A. Jack, "Semi-Continuous Hidden Markov Models for Speech Recognition," Computer Speech and Language, vol.3, pp. 239-251, 1989.
7. X.D. Huang, "Semi-Continuous Hidden Markov models for speech recognition," Ph.D. thesis, Department of Electrical Engineering, University of Edinburgh, 1989.
8. Simon Haykin, *Neural Networks*, Prentice-Hall, pp.256-317, 1998

▲문 연 주(Yun Joo Moon) 1970년 12월 16일생

1998년 : 광운대학교 전자통신과(공학사)
1998년 3월~현재 : 광운대학교 전자통신과 석사과정 재학중
※주관심 분야 : 음성압축, 음성인식, 디지털 필터설계

▲전 선 도(Sun Do June)
제 17권 8호 참조

▲강 철 호(Chul Ho Kang)
제 17권 8호 참조