VQ 코드의 천이 행렬과 이산 HMM을 이용한 한국어 단어인식

Korean Word Recognition using the Transition Matrix of VQ-Code and DHMM

정 광 우*, 홍 광 석**, 박 병 철*
(Kwang Woo Chung*, Kwang Seok Hong**, Byung Chul Park*)

요 약

본 논문에서는 단어 인식 시스템의 성능 개선을 위하여 다음과 같은 두가지 방법을 제안한다. 첫번째 방법은 VQ 코드간의 천이를 안정화시키기 위하여 음성신호의 특징벡터 시퀀스에 관성을 적용하는 방법이고, 두번째 방법은 이산 HMM 모델에서 인접 프레임 간의 시간 상관성을 고려하기 위하여 VQ 코드의 천이행렬을 출력 심벌의 관측확률에 가중치로 이용하여 새로운 관측확률을 발생하는 방법이다.

특징벡터 시퀀스에 관성을 도입함으로서, SOFM상의 각 단어에 대한 반응경로에서 확률분포가 중첩되는 것을 억제하여 HMM의 상태천이를 안정화 시킬 수 있다. 기존의 이산 HMM에 VQ 코드의 천이행렬을 가중치로 적용함으로써, 특징벡터의 확률분포를 더욱 세분화하고, 특징분포를 적당한 영역으로 제한함으로써 인식시스템의 성능을 개선할 수 있다.

제안한 방법을 평가하기 위하여 50개의 DDD 지역명을 대상으로 인식 실험을 수행하였다. 실험 결과에 의하면, 제안된 방법이 기존의 HMM 모델에 비해 화자종속 실험에서는 4.2%의 인식률 향상과 화자 독립 실험에서는 12.45%의 인식률 향상을 얻을 수 있었다.

Abstract

In this paper, we propose methods for improving the performance of word recognition system. The ray stratey of the first method is to apply the inertia to the feature vector sequences of speech signal to stabilize the transitions between VQ cdoes. The second method is generating the new observation probabilities using the transition matrix of VQ codes as weights at the observation probability of the output symbol, so as to take into account the time relation between neighboring frames in DHMM.

By applying the inertia to the feature vector sequences, we can reduce the overlapping of probability distribution

*성균관 대학교 전자공학과
**제주 대학교 정보공학과
접수일자 : 1994년 4월 14일

of the response paths for each word and stabilize state transitions in the HMM. By using the transition matrix of VQ codes as weights in conventional DHMM, we can divide the probability distribution of feature vectors more and more, and restrict the feature distribution to a suitable region so that the performance of recognition system can improve.

To evaluate the performance of the proposed methods, we carried out experiments for 50 DDD area names. As a result, the proposed methods improved the recognition rate by 4.2% in the speaker-dependent test and 12.45% in the speaker-independent test, respectively, compared with the conventional DHMM.

Ⅰ. 서 론

HMM은 현재 음성인식 시스템에서 가장 널리 이용되는 인식 모델로서, 음성신호의 변동을 통계적으로 처리하고[1], 이 통계량을 확률 형태의 모델에 반영하여 음성을 인식하는 방법이다.[2-4]

화자독립 HMM 모델은 다양한 화자 변동과 음소 환경의 변화에 의해 발생되는 음성학적 특징을 포함하기 위하여 많은 양의 학습 데이터를 필요로 한다. 학습데이터 양의 증가는 모델에 의해 학습되지 않은 영역의 크기를 줄일 수 있어 인식 시스템의 성능을 개선할 수 있는 반면에, 많은 양의 학습 데이터는 각 음소의 스펙트럼 특징 분포를 넓게 퍼지게 함으로써 서로 다른 음소간의 중첩을 발생하게 되고, 이것은 인식 시스템의 성능을 저하시키는 요인으로도 작용하게 된다[5].

음성신호의 특징정보와 개인성에 관련된 정보는 순시적인 스펙트럼뿐만 아니라 스펙트럼의 천이구간 안에 포함되어 있는데 벡터 양자화 기법에서는 이런 정보를 VQ 코드 열로 나타낸다. 그러나 기존의 HMM에서는 비록 VQ 코드의 관측확률이 상태에 의존하여 변화하지만 인접 프레임 간의 VQ 코드의 천이에는 아무런 계약이 없어 두 VQ 코드 사이의 천이가 학습 데이터에서 관측되지 않더라도 두 VQ 코드 사이의 모든 천이가 높은 확률 값을 갖는 것을 허용함으로써 음소식별의 처리능력에 비해 인접 프레임 간의 VQ 코드 천이를 적절히 처리하지 못하는 문제점을 나타내고 있다[6].

본 논문에서는 이러한 문제점들을 보완하기 위해서, 입력벡터의 확률분포를 고려함으로써 양자화 왜곡을 줄일 수 있는 Self Organizing Feature Map (SOFM)을 벡터 양자화기로 사용하였으며[7], SOFM 상의 특징 벡터 시퀀스에 대한 궤적의 안정화를 위해 관성항을 도입하였다. 인접 프레임 간에 관성항을 적용함으로써 입력 특징 벡터 시퀀스에 대한 SOFM상의 확률분포가 중첩되는 것을 억제하고, 일정 수의 뉴런을 갖는 SOFM에 보다 많은 모델을 형성할 수 있도록 하였다.[15] 또한 이산 HMM 모델에서 인접 프레임 간의 시간 상관성을 고려하기 위해서 인접 프레임 간의 VQ 코드의 천이 행렬을 출력 심볼의 관측확률에 가중치로 사용하여 백터 공간상의 특징 분포를 세분화하고, 적당한 영역으로 제한함으로써 인식시스템의 성능을 개선하는 방법을 제안한다.

제안된 방법의 타당성을 조사하기 위해서 한국어 단어인식을 수행하였다. 특징 파라메터로는 선형예측계수로부터 얻어진 전극형 모델의 스펙트럼 상에서 나타나는 인접한 극점들간의 중첩효과를 제거한 SGDS[8](smoothed group delay spectrum)을 특징벡터로 이용하였고, 9명의 화자가 발성한 53개의 단음절을 이용하여 SOFM을 학습하였으며, 50개의 지역 명을 대상으로 이산 HMM모델과 VQ 코드의 천이 행렬을 구성하여 한국어 단어인식을 수행하였다.

2장에서는 본 논문에서 사용한 SGDS 특징 파라메터 추출방법과 특징 벡터 시퀀스의 안정화를 위하여 도입한 관성에 대하여 설명하고, 3장에서는 인접 프레임간의 시간 상관성을 고려하기 위하여 국부적인 VQ 코드의 천이행렬 작성방법과 이를 이산 HMM 인식모델에서 출력심벌의 관측확률에 가중치로 이용하여 단어인식을 수행하는 가중 출력 DHMM에 대하여 설명하였다. 4장에서는 제안된 방법을 이용한 단어인식 실험의 결과및 고찰에 대하여 설명하고, 5장에서는 전체적인 인식 시스템의 성능 평가에 대하여 기술하였다.

Ⅱ. 음성신호의 특징 파라메터 추출

음성인식에서 사용할 특징 파라메터는 다양한 화자가 같은 발음을 했을 때 화자의 변동에 상관없이

일정하고, 서로 다른 발음은 뚜렷이 구별되도록 표현되는 것이 이상적이다. 그러므로 특징 파라메터 추출은 인식모델과 함께 매우 중요하며 이에 대한 연구가 지난 수십년간 진행되어 왔다.

이 장에서는 여러가지 분석방법중, 본 논문에서 사용한 SGDS 추출방법과 특징에 대하여 설명하고, 특징 벡터 시퀀스의 안정화를 위하여 사용된 관성에 대하여 설명한다.

2-1. SGDS(Smoothed Group Delay Spectrum)

음성신호에 관련된 특징을 추출하는 음성분석은 음성인식 과정에 있어서 매우 중요한 단계로서, 이 단계에서 제거된 중요한 음성정보는 후처리 과정에서 쉽게 복구할 수 없게 된다. 지난 수십년 동안 음성특징벡터 추출을 위한 음성 분석 방법들이 많이 제안되었는데, 이중 널리 이용되는 방법중 하나가 성도 전달함수와 여기(excitation)함수를 분리할 수 있는 LPC cepstrum[9]으로서 많은 인식 시스템에서 좋은 성능을 나타내고 있다. 그러나 이러한 분석방법들은 quefrency 성분의 가중치가 균일하게 주어지기 때문에 전극형 모델의 스펙트럼 상에서 인접한 극점들간의 상호 간섭으로 인하여 극점들을 독립적으로 추출해 내는데 어려움이 있고, 부가적인 잡음이나 가변 주파수 특성 하에서는 심하게 성능이 열화되는 단점이 갖고 있다[8]. 이러한 결점을 보완하기 위하여 SGDS (Smoothed Group Delay Spectrum)가 제시되었다.[10] SGDS는 사람의 청각 시스템에서 스펙 트럼 peaks와 valleys, 스펙트럼 기울기가 서로 다른 중요도를 나타낸다는데 근거하여, cepstral 계수에 Gaussian window 형태를 갖는 비균일 가중치를 곱함으로써 구해진다. SGDS는 음성신호의 특징중에서 포먼트 주파수에 해당하는 스펙 트럼 피크를 더욱 강조하게 되고, 인접한 포먼트들의 영향을 분리할 수 있는 특징을 갖고 있다.

LPC 분석에서 음성의 각 프레임은 전달함수 H(Z)를 갖는 p차 전극형 필터로 표시할 수 있다.

$$H(z)=\frac{1}{1+\sum_{k=1}^{p}\alpha_k z^{-k}} \tag{1}$$

전달함수 H(Z)의 위상 $\theta(w)$는 다음과 같으며,

$$\theta(w)=-\tan^{-1}\frac{Im\{A(w)\}}{\mathrm{Re}\{A(w)\}} \tag{2}$$

여기서 $A(w)=\sum_{k=0}^{p}\alpha_k e^{-jwk}$: 진폭 스펙트럼

GDS(group delay spectrum)은 위상의 미분으로 정의된다.

$$\tau(w_i)=-\frac{\partial\theta(w)}{\partial w}\Big|_{w=w_i} \tag{3}$$

이산적으로 정의된 스펙트럼의 위상성분 $\theta(w_i)$에 대해 GDS는 다음과 같이 정의된다.

$$\hat{\tau}(w_i)=-\frac{\theta(w_i)-\theta(w_{i-1})}{w_i-w_{i-1}}\quad 1\le i\le L \tag{4}$$

여기서 L은 SGDS의 채널 수
LPC cepstrum의 대수 스펙트럼과 GDS은 다음과 같다.

$$\log|H(w)|=\sum_{k=0}^{\infty}c_k\cos(kw),$$

$$\tau(w)|=\sum_{k=0}^{\infty}k\,c_k\cos(kw) \tag{5}$$

여기서 c_k는 cepstral 계수
위상 $\theta(w)$는 다음과 같다.

$$\theta(w)=-\int_0^{\infty}\tau(x)\,dx=-\sum_{k=0}^{\infty}c_k\sin(kw) \tag{6}$$

각 채널의 대역폭이 2B라 할 때, SGDS는

$$\hat{\tau}(w,B)=\frac{\theta(w+B)-\theta(w-B)}{2B}$$
$$=\sum_{k=0}^{\infty}\frac{\sin(kB)}{B}\,c_k\cos(kw) \tag{7}$$

식(7)은 cepstral 계수 c_k에 sine window함수 $\sin(kB)/B$를 씌운 결과와 동일하다.

SGDS는 낮은 quefrency성분을 제거함으로써 스펙트럼의 피크가 강조되며, 높은 quefrency성분을 제거함으로써 스펙트럼의 작은 변동에 의한 과잉 강조가 제거될 수 있다.

본 논문에서는 16차 LPC cepstrum에 $B=\pi/32$를 적용하였고, 부가적인 smoothing을 위하여 다음과 같이 LPC계수에 exponential window를 부가하였다.

$$\overline{\alpha_k}=A^k\alpha_k(\alpha_k:\text{LPC 계수, A}=0.9) \tag{8}$$

SGDS 추출을 위한 블록다이어그램은 그림 1과 같다.

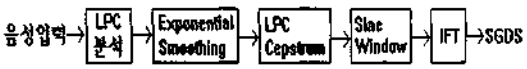


그림. 1. SGDS의 추출

2-2. 관성(Inertia)[15]

음성신호는 특징 벡터들의 시퀀스로 표현되는데 각 프레임간 간격이 작을 경우 벡터 스퀀스는 벡터 공간상에서 그 벡터가 가리키는 질점의 연속적인 운동으로 볼 수 있다. 그러나 일반적으로 사용되는 특징벡터의 시퀀스에서는 안정구간의 인접한 프레임 간에도 변동이 커서, 벡터 공간상에서 질점의 궤적에 대한 연속성을 가정하는 것에는 무리가 있다. 대개의 경우 특징벡터의 시퀀스에 대한 벡터 공간상의 궤적은 국소구간에서 랜덤한 변동을 보이고, 이러한 변동은 각각의 음성 특징벡터에 대한 확률분포를 중첩시키는 하나의 요인이 되기 때문에 특정 음성신호에 대한 특징벡터 시퀀스의 안정성은 이산 HMM모델에 의한 음성인식 시스템의 성능에 많은 영향을 주게 된다. 본 논문에서는 특징벡터의 시퀀스에 대한 궤적을 안정화시키기 위해 특징벡터 시퀀스에 관성을 주어 궤적의 랜덤한 변동을 제거함으로써 SOFM상의 확률분포가 중첩되는 것을 억제하였다. 이는 특징 벡터의 시퀀스에 대한 궤적에 필터를 취하여 smoothing 처리를 하는 것과 동일하다.

본 논문에서는 식(9)로 표현되는 auto-regressive 필터를 사용하여 특징벡터의 시퀀스에 대한 궤적을 smoothing처리하였다.

$$\vec{Y}(t) = a \cdot \vec{X}(t) + b \cdot \vec{Y}(t-1) + c \cdot \vec{Y}(t-2) \qquad (9)$$

식(9)에서 $\vec{X}(t)$는 현재 프레임의 특징벡터로 16채널의 SGDS이고, $\vec{Y}$는 관성이 적용된 SDGS이다.

본 논문에서는 관성을 적용한 SGDS를 음성신호를 특징벡터로 이용하여 자율 학습기능을 갖는 SOFM 신경망의 입력벡터로 사용하였다.

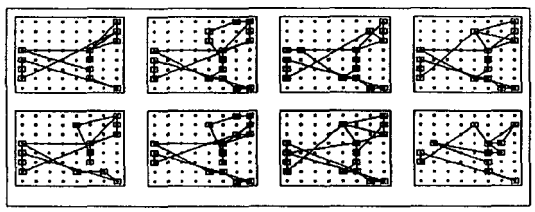
Ⅲ. VQ 코드의 천이 행렬-이산 HMM 인식 모델

3-1. VQ 코드의 천이 행렬

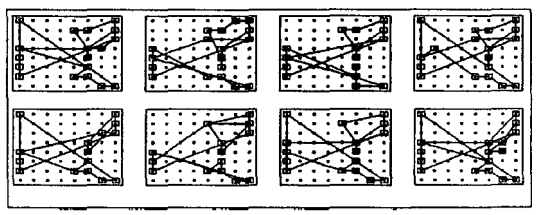
본 논문에서는 감각 신경계에서 흔히 볼 수 있는 뉴런간의 측면작용과 이에 관련된 receptive field의 개념을 단층 신경망 학습 알고리즘에 도입한 SOFM을 벡터양자화기로 이용하였다.[7][14]

SOFM상에서 특징벡터의 랜덤성분을 제거하고, 확률분포의 중첩을 제거하기 위하여 2장에서 설명한 관성을 적용함으로써 벡터 공간상에서 코드 시퀀스의 궤적에 대한 코드천이의 안정화를 통해 인식 시스템의 성능을 개선할 수 있었다.[15] 그러나 이러한 방법을 이용하여 음성신호를 분석하더라도 그림 2과 같이 유사단어 사이에서 코드 시퀀스의 궤적 중 많은 부분이 중첩되어 발생 되는데, 이러한 확률 분포를 더욱 세분화하여 구분하기 위해서 인접 프레임 사이의 상관성을 고려한 VQ 코드의 천이 행렬을 이산 HMM 인식 모델의 심벌 관측확률에 가중치로 이용하는 방법을 제안한다.

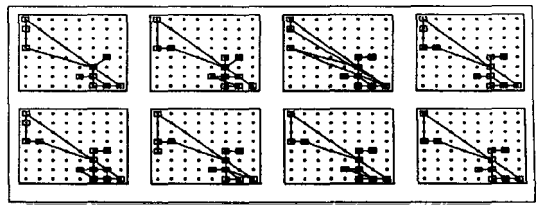
"안양"



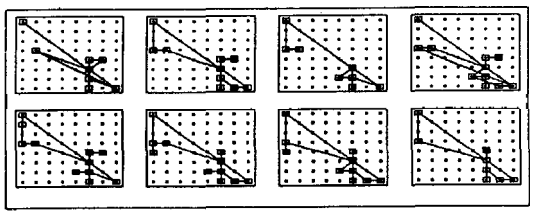
"단양"



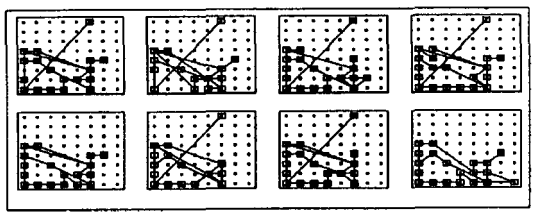
"진해"



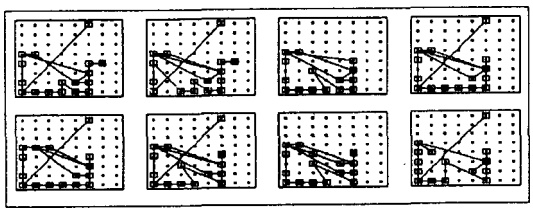
"김해"



"영주"



"경주"



"청주"

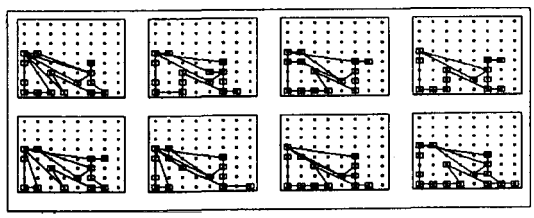


그림 2. 유사음성에 대한 SOFM상의 반응경로

음성 신호내의 인접한 프레임 간에는 매우 높은 상관성이 있으며, 이러한 상관성 정보는 VQ 코드의 시퀀스내에 포함되어 있다. 따라서 이런 천이 정보를 추출하여 음성인식에 이용하는 것이 바람직할 것이다.

VQ 코드의 천이 행렬은, 입력 화자에 대한 스펙트럼 상의 특징분포를 적당한 영역으로 제약하기 위해, SOFM의 출력 코드 열에서 인접 VQ 코드사이의 상관성을 고려하기 위한 것이다. 즉, 인접한 프레임 간의 상관성 정보를 표현하기 위하여 현재 프레임에서 발생된 VQ 코드와 이전 프레임에서 발생된 VQ 코드 간의 상관 정도를 행렬의 형태로 표현한 것이다. 여기서 사용된 VQ 코드 천이 행렬을 계산하는 방법은 SOFM 학습이론[7]에 근거하여 다음과 같은 방법으로 산출한다.

SOFM 학습 알고리즘에서 receptive field내의 뉴런들은 그 중심 뉴런의 활성에 대해 그림 3와 같은 Mexican hap 형태의 lateral interaction을 하는데 이것을 인접 프레임 간의 코드 천이에 적용함으로써 이전 프레임의 VQ 코드에서 현재 프레임의 VQ 코드로의 천이에 대한 행렬을 계산할 수 있다. 본 논문에서는 그림 4에 나타낸 VQ 코드 격자구조에서 Mexican hap 함수 대신에 그림4(b)와 같이 lateral inhibition을 무시한 단순화된 근사함수를 사용하여 인접 프레임 간의 시간 상관성을 계산한다. 인접 프레임 간의 상관성을 고려한 VQ 코드의 천이행렬 요소는 다음과 같이 정의된다.

$$W(c_j|c_i) = \frac{\text{코드 } c_i\text{에서 코드 } c_j\text{로 천이된 횟수}}{\text{코드 } c_i\text{의 발생횟수}} * a \qquad (10)$$

여기서 c_i는 이전 프레임에서 발생된 VQ 코드이고, c_j는 현재 프레임에서 발생된 VQ 코드와 receptive field 내의 VQ 코드 값 모두를 포함한다. a는 가중치로서 receptive field내의 중심 뉴런인 경우, 즉 현재 프레임에서 발생된 VQ 코드인 경우는 1, receptive field내의 뉴런으로서 중심뉴런이 아닌 경우는 0.4의 값을 갖는다. 이와 같이 현재 프레임에서 발생된 코드 뿐만 아니라 receptive field 내의 코드까지 포함한 것은, 제한된 학습 데이터를 이용하여 신뢰성 있

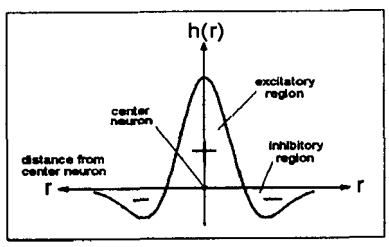


그림 3. Mexican hap 함수

는 천이 행렬을 얻기 위해서이다. 본 논문에서는 receptive field내에 존재하는 코드값에 일정한 값을 가중치로 주어 천이 행렬을 작성하였다. 이것은 Fuzzy VQ와 같은 효과를 얻을 수 있다.

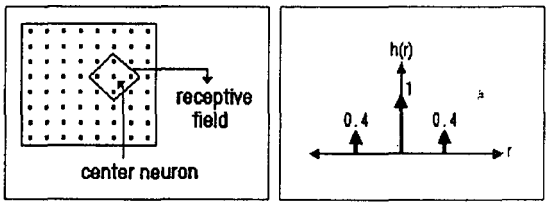


그림 4. (a) SOFM상의 VQ 코드 배열
(b) VQ 코드 천이에 도입된 가중치

3-2. 시간 상관성을 갖는 HMM 음성인시기

HMM은 천이들에 의해 서로 연결된 상태들의 모임으로서 각 천이에는 2가지 종류의 확률이 관련되어 있다. 하나는 현재 상태에서 다른 상태로 천이가 이루어질 상태 천이확률이고, 다른 하나는 천이가 이루어졌을 때 유한개의 관측대상으로로 부터 각 출력 심벌이 방출되는 조건부 확률을 규정하는 출력 확률밀도 함수로 정의된다. 그러나 기존의 HMM 모델은 현재 상태에서의 출력 심벌 발생확률이 이전에 발생된 출력 심벌에는 무관하게 주어짐으로써 음성신호와 같이 인접 프레임간의 의존성이 높은 경우에는 적합하지 못한 점이 있다.[6][12] 본 논문에서는 인접프레임 간의 상관성을 고려하기 위하여 VQ 코드의 천이 행렬을 이산 HMM 모델의 심벌 관측확률에 가중치로 사용하여 새로운 형태의 심벌 관측확률을 계산하는 방법을 사용한다. 즉, 현재 프레임에서 발생된 VQ 코드의 발생 확률을 이전 프레임에서 발생된 VQ 코드에 의해 조절함으로써 인접 프레임간의 시간상관성을 고려할 수 있는 방법이다. HMM 모델에 시간 상관성을 고려함으로써 모델에서의 출력심벌의 관측확률 분포는 같은 상태에서조차 이전 프레임에서 발생된 VQ코드에 의존하게되어 서로 다른 확률 값을 갖게 되며, 국부적인 VQ 코드의 천이에 제약을 줌으로써 스펙트럼상에서 특징분포의 중첩을 줄일 수 있다.

시간 상관성을 고려한 가중 출력 HMM모델에서 출력 심벌의 관측 확률을 계산하는 과정은 다음과 같다.

1단계) 입력 화자로부터 발성된 음성 데이터를 이용하여 SOFM을 학습 한다.
SOFM의 출력 코드 열을 이용하여 각 단어별 DHMM을 작성 한다.

2단계) 학습 데이터의 VQ 코드 열을 이용하여 인접 프레임 간의 VQ 코드의 천이에 대한 행렬을 계산 한다.

3단계) 이산 HMM의 출력확률을 가중확률로 수정한다.
출력 심벌의 관측확률 $P(c_j)$을 이전 프레임의 VQ-code c_i에 따라 변화시킨다.

$$WP(c_j|c_i) = \frac{W(c_j|c_i)P(c_j)}{\sum_{m=1}^{N} W(c_m|c_i)P(c_m)} \tag{11}$$

c_j : 현재 프레임의 VQ-code
c_i : 이전 프레임의 VQ-code
N : codebook size
$P(c_j)$: 기존의 이산 HMM에서 VQ-code c_j가 발생될 출력 심볼 확률
$W(c_j|c_i)$: VQ 코드의 천이행렬에서 VQ-code c_i에서 VQ-code c_j로 천이가 일어날 확률
$WP(c_j|c_i)$: 출력 심벌의 가중 관측확률
이산 HMM모델에서 VQ-code c_i에서 VQ-code c_j로 천이가 일어날 때의 가중 출력 확률

식(11)에서 이전 프레임의 VQ-code로부터 현재 프레임의 모든 VQ-code로의 천이확률은 새로운 확률분포를 생성하기 위하여 이산 HMM의 출력 확률 분포에 대한 가중치로 이용되었다. 예를 들면, 만약 이전 프레임에서 현재 프레임으로의 VQ-code 천이가 입력화자 혹은 인식하고자하는 단어에서 발생했다면 현재 프레임의 출력 확률은 증가될 것이고, 그렇지 않다면 작아질 것이다. 이렇게 새로운 확률분포를 구성함으로써 특징벡터의 중첩을 줄일 수 있다.

Ⅳ. 실험 및 고찰

4.1 실험 조건

제안된 인식방법을 평가하기 위하여 50개의 지역명을 대상으로 실험하였다.

본 논문에서 사용된 데이터는 50개의 지역명을 9명

의 화자가 10번씩 발성한 총 4500개의 데이터를 가지고 실험하였다.

음성 데이터의 특징 파라메터 추출에 대한 분석 조건은 표 1과 같다.

표 1. 분석 조건

샘플링 주파수	10 kHz
LPF	4.5 kHz
분석 프레임 길이	20 msec
프레임 간격	10 msec
SGDS 채널	16 채널

벡터 양자화기로 사용된 SOFM은 입력노드가 16개이고 출력노드가 8x8(화자 종속) 혹은 16x16(화자독립)인 것을 사용하였다.

SOFM상의 반응경로에 관성을 주기 위해 사용된 필터의 계수 값은 음소의 평균길이를 고려하여 적절하게 선택해야 하는데, 본 연구에서는 분석조건에 대해 많은 실험을 통해 식(9)의 필터계수를 a=0.25, b=0.72, c=0.1로 각각 선택하여 특징벡터를 추출하였다. SOFM 학습을 위해서는 50개의 단어에 포함된 53개의 단음절을 각 화자가 2회씩 발성한 데이터를 이용하였다.

이산 HMM에 의한 인식은 Forward-Backward 알고리즘을 사용하였으며, zero확률을 방지하기 위해 HMM의 모든 파라메터 값에 대해 임의의 최저치 이하의 값을 최저치(여기서는 10^{-6}을 이용)로 대치하여 계산하였으며, 상태수는 단어 당 음소의 수를 고려하여 7개로 하였다. 본 연구에서 사용된 이산 HMM모델의 구조를 그림 5에 나타내었다.

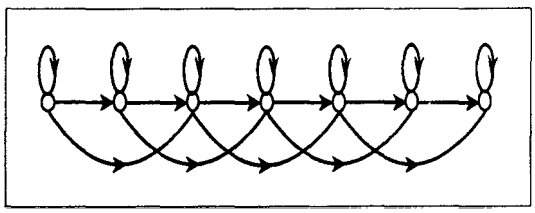


그림 5. HMM의 구조

그림 6에는 본 논문에서 사용된 전체인식 시스템의 블록도를 나타내었다.

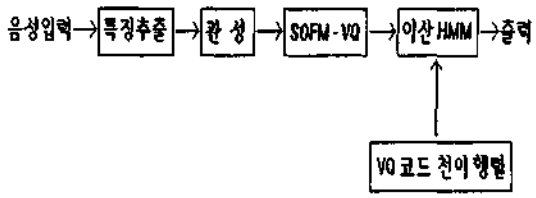


그림 6. 인식 시스템의 블록도

4.2 화자종속 HMM

음성특징벡터 시퀀스에 관성을 도입하여 특징 벡터 시퀀스의 안정화를 기하고, VQ 코드의 천이 행렬을 이용하여 인접 프레임간의 시간상관성을 고려한 제안된 가중 출력 DHMM과 기존의 DHMM 모델에 의한 인식성능을 비교 검토하기 위하여, 먼저 화자종속 실험을 수행하였다.

화자 종속 실험에서는 9명의 화자에 대해 8x8의 SOFM으로 벡터양자화기를 구성하였으며, 53개의 단음절을 이용하여 SOFM-VQ을 학습하였다.

각 파라메터의 재추정은 Baum-Welch 재추정 알고리즘으로 모델을 학습하였다.

VQ 코드 천이 행렬은 각 화자의 단어별 특징파라메터의 집합으로부터 독립적으로 구성하였다. 화자종속 실험에 사용된 화자 종속 코드 천이 행렬은 각 화자가 5회 발성한 음성 데이타로부터 인접 프레임간의 VQ 코드 천이의 빈도 수를 계산하여 작성하였으며, 인접 프레임간의 VQ 코드 천이의 빈도 수는 이전 프레임에서 현재 프레임의 VQ 코드로의 천이 뿐만 아니라 SOFM 학습이론에 근거하여 receptive field 내의 VQ 코드에도 일정한 빈도수를 주어 계산하였다. 또한 각 단어의 시간 길이차에 의한 발생 빈도수의 영향을 줄이기 위하여 전체 프레임수로 정규화 하였다.

실험내용으로는 음성 특징 파라메터에 관성을 주지 않은 경우, 관성을 준 경우, 관성 뿐만 아니라 인접 프레임간의 시간 상관성을 고려한 경우로 각각 나누어 50개의 지역명 인식실험을 수행하였으며, 그 결과를 표 2에 나타내었다.

표 2에 나타낸 인식 결과에 따르면, 기존의 이산 HMM만을 이용한 방법 1의 평균인식률이 93.2%를 보인 반면, 관성을 적용한 방법 2의 경우 평균 인식률은 96.6%로 3.4%의 향상된 인식률을 보이고 있다. 방법 1의 경우, 같은 음성신호에 대해 SOFM상의 뉴런의 반응경로가 불안정 궤적을 나타내어 HMM모델

표 2. 화자종속 인식 결과

방법 / 화자	방법 1	방법 2	방법 3
A 화자	95.6%	96.0%	96.0%
B 화자	86.8%	96.0%	97.6%
C 화자	95.2%	96.8%	98.6%
D 화자	95.2%	98.0%	98.8%
E 화자	94.8%	96.8%	97.2%
F 화자	91.6%	95.2%	96.4%
G 화자	88.4%	93.6%	94.8%
H 화자	96.4%	98.0%	98.4%
I 화자	94.8%	98.8%	100%
평균인식률	93.2%	96.6%	97.5%

방법 1 : 관성을 주지 않은 경우
방법 2 : 관성을 준 경우
방법 3 : 관성과 시간 상관성을 고려한 경우

간의 확률분포가 중첩되어 많은 오인식을 발생한 반면, 방법 2에서는 관성을 적용하여 SOFM상의 뉴런의 반응경로에서 발생되는 랜덤성분을 제거함으로써 궤적을 안정화시킬 수 있었으며, 결과적으로 HMM 모델의 확률분포가 중첩되는 것을 억제하여 인식률의 향상을 가져올 수 있었다. 방법 3은 이산 HMM 모델에서 시간 상관성을 고려하기 위하여 인접 프레임간의 VQ 코드의 천이 행렬을 출력 심벌의 관측확률에 가중치로 사용하여 새로운 출력 심벌의 관측확률을 구성한 경우로 평균인식률이 97.5%로 가장 좋은 인식률을 나타내고 있다. 이는 음소신호내의 천이정보를 인식시에 이용함으로써 유사단어 간의 오인식이 줄어들어 방법 2에 비해 인식률이 1.1% 향상되었다.

4.2 화자독립 HMM

화자독립 인식실험에서는 5명의 화자가 발성한 음성 데이터를 16x16 SOFM 벡터 양자화기의 학습과 이산 HMM 모델을 학습하는데 사용하였으며, 4명의 화자가 발성한 음성 데이터를 테스트 데이터로 이용하였다. 여기서 16x16 SOFM을 이용한 이유는 많은 화자가 발성한 음성 데이터 내에 특징벡터의 변동이 커져서 8x8 SOFM으로 구성하기에는 VQ 코드 북의 크기가 너무 작아 화자독립 HMM에서는 16x16 SOFM으로 확장하였다.

화자 독립 실험에 사용된 천이 행렬은 5명의 화자가 발성한 음성 데이터의 단어별 특징 파라메터 집합으로부터 인접 프레임간의 VQ 코드 천이의 빈도수를 계산하여 화자 독립 VQ 코드 천이 행렬을 작성하였다.

인식에 사용된 이산 HMM모델은 각 단어당 2개씩 작성하여 인식실험에 사용하였다. 화자독립 인식실험 결과를 그림 7에 나타내었다.

그림 7에서, 방법 3은 본 논문에서 제안한 관성과 VQ 코드 천이 행렬을 이용한 경우의 인식결과로 93.85%의 인식률을 나타내고 있으며, 기존의 이산 HMM 모델의 인식결과와 비교하여 12.45%의 인식률 향상을 얻을 수 있었다. 결과적으로 관성에 의해 인접 프레임간의 VQ코드의 천이를 안정화 시키고, 이 정보를 인식 단계에서 이산 HMM모델의 출력심벌의 관측확률에 가중치로 적용함으로써 유사음성에 대한 변별능력을 개선할 수 있음을 알 수 있다.

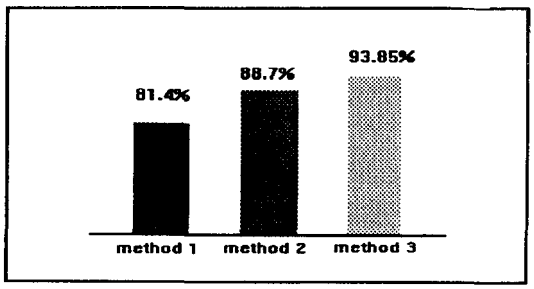


그림 7. 화자독립 인식결과

V. 결 론

본 연구에서는 두가지 방법을 이용하여 음성인식 시스템의 성능을 개선하고자 하였다. 첫번째는, 음성신호의 특징벡터 시퀀스에 관성을 도입하는 방법이고, 두번째는, VQ 코드의 천이 행렬을 이산 HMM 모델의 각 상태에서 출력 심벌의 관측확률에 대한 가중치로서 사용하는 방법이다.

제안된 인식 알고리즘은 특징벡터 시퀀스에 관성을 줌으로써, 특징벡터 시퀀스의 궤적에 포함된 랜덤한 변동성분을 줄여 궤적을 안정화시킬 수 있었으며, 결과적으로 SOFM상의 특징분포가 중첩되는 것을 억제할 수 있는 특징을 갖고 있다. 또한 인접 프레임간의 상관성을 나타내는 천이 행렬을 DHMM 모델의 가중치로 이용함으로써, 출력 심벌의 관측확률 분

포는 같은 상태에서 조차 이전 프레임에서 발생된 VQ 코드에 의존하게되어 서로 다른 확률 값을 갖게 되며, 이는 특징벡터의 확률분포를 더욱 세분화시키는 역할을 한다. 따라서 제안된 인식 알고리즘은 스펙트럼상에서 서로 다른 입력 음성간의 특징분포의 중첩을 감소시켜 인식시스템의 성능을 개선할 수 있는 특징을 갖고 있다.

제안된 성능 알고리즘을 평가하기 위하여 50개의 지역명을 대상으로 화자종속과 화자 독립 음성인식 실험을 수행하여 다음과 같은 결론을 얻을 수 있었다.

첫째, SOFM상의 특징벡터 시퀀스에 대한 궤적의 안정화를 위해 인접 프레임간에 관성 항을 적용함으로써 입력 특징벡터 시퀀스에 대한 SOFM상의 확률분포가 중첩되는 것을 억제하여 7.3%의 인식률 향상을 얻을 수 있었다. 둘째, 인접 프레임 간의 시간 상관성을 고려하기 위해, 이산 HMM 모델의 각 상태에서의 심벌 관측확률에 각 단어별 VQ 코드의 천이 행렬을 가중치로 이용함으로써 벡터공간상의 특징 분포를 더욱 세분화하여 인식 시스템의 성능이 5.15% 향상되었다.

따라서, 관성과 VQ 코드 천이 행렬을 이용한 음성인식 방법이 기존의 이산 HMM 보다 우수한 방법임을 확인하였다.

제안된 알고리즘의 계산량는 기존의 HMM보다 다소 증가하지만 거의 무시할 수 있을 정도로 증가폭이 작다. 반면에 메모리는 각 단어별로 VQ 코드의 천이행렬을 작성하여야 하기 때문에 인식 단어의 증가와 더불어 많은 양의 메모리를 필요로 하지만 소용량의 단어 인식의 경우에는 메모리에 대한 고려없이 제안된 알고리즘을 적용 가능하다.

차후 연구과제로는 대용량 단어 인식을 수행하기 위해 음소별 천이 행렬을 작성하여 음소단위의 HMM 인식 모델로 확장하는 것이 필요하다.

참 고 문 헌

1. L.R.Rabiner and B.H.Juang, "An introduction to Hidden Markov Models," IEEE ASSP MAGAZIE, Jan., pp.4-16, 1986.
2. L.R.Rabiner, "A Tutorial on HIdden Markov Models and selected applications in speech recognition," Proc. IEEE, pp.257-268, Feb., 1989.
3. L.R.Rabiner, et al., "Recognition of isolated digits using Hidden Markov Models with continuous mixture densities," BSTJ. vol.64, No.6, pp.1211-1234, July-Aug., 1985.
4. L.R.Rabiner, B.H.Juang, S.E.Levinson and M.M. Sondhi, "Some properties of continuous hidden Markov Model repesentation," AT&T Tech.J., vol 64, pp.1251-1270, July-Aug., 1985.
5. S.Takahashi, T.Matsuoka, Y.Minami and K. Shikano, "Phoneme HMM constrained by frame correlations," Proc. ICASSP93, pp. 219-222, 1993.
6. S.Takahashi, "T.Natsuoka and K.Shikano, "Phonemic HMM constrainey by statistical VQ-code transition," Proc. ICASSP92, pp. 553-556, 1992.
7. T.Kohonen, Self-Organization and Associative Memory, 1987.
8. H.Singer, T.Umezaki and F.Itakura, "Low bit quantization of the smoothed group delay spectrum for speech recognition," Proc. ICASSP90, pp. 761-764. 1990.
9. A.H.Gray and J.D.Markel, "Distance measures for speech processing," IEEE ASSP, Vol.ASSP-24, 5, Ocr., 1976.
10. F.Itakura and T.Umezaki, "Distance Measure for speech recognition based on the smoothed gorup delay spectrum," Proc. ICASSP87, pp.1257-1260, 1987.
11. Y.Linde, A.Buzo and R.M.GRAY, "An Algorithm for Vector Quantizer design," IEEE COM., pp. 84-95, 1980.
12. C.J.Wellekens, "Explicit correlation in hidden Markov model for speech recognition," Proc. ICASSP87, pp.384-386, 1987.
13. S.Furui and M.M.Sondhi, Advances in speech signal processing, Marcel Dekker, Inc.1992.
14. T.Kohonen, "The Self-Organization Map," Proc. IEEE, vol 78, No.9, pp. 1464-1480, 1990.
15. 정광우, 윤석현, 홍광석, 박병철, "관성과 SOFM-HMM을 이용한 고립단어 인식," 전자공학회 논문지 제31권 B편 제6호, pp17-24, 1994.

▲鄭 光 宇(Kwang Woo Chung) 1966년 12월 1일생

1989년 2월 : 성균관대학교 전자공학과 졸업(공학사)

1991년 2월 : 성균관대학교 대학원 전자공학과 졸업(공학석사)

1992년 3월~현재 : 성균관대학교 대학원 전자공학과 박사과정

※주관심분야 : 음성인식 및 신호처리, 신경회로망 등임

▲洪 光 錫(Kwang Seok Hong)

9권 5호 참조

현재 : 제주대학교 정보공학과 교수

▲朴 炳 哲(Byung Chul Park)

9권 5호 참조

현재 : 성균관대학교 전자공학과 교수