한국음향학회지 제21권 제7호 pp. 641~647 (2002)

분절 특징 은닉 마코프 모델에서의 경향 공유에 관한 연구

A Study on Trend Sharing in Segmental-feature HMM

윤 영 선*
(Young-Sun Yun*)

*한남대학교 정보통신 · 멀티미디어공학부

(접수일자: 2002년 6월 14일; 채택일자: 2002년 10월 4일)

본 논문에서는 경향 양자화 기법을 적용하여 분절 특징 은닉 마코프 모델 (HMM: hidden Markov model)의 매개 변수 수를 줄이는 방법을 제안한다. 제안된 방법은 분절 특징 HMM에서 사용하는 분절 특징, 즉 모수적 궤적을 위치 정보와 경향 정보로 분리한 후, 분리된 경향 정보를 경향 코드북을 이용하여 공유한다. 분절 특징에서 위치 정보는 특징의 기준 점을 나타내고, 경향 정보는 분절 특징의 변이를 의미하며 특징의 많은 부분을 차지하고 있다. 따라서 경향 정보가 공유될 수 있다면 분절 특징 HMM의 매개 변수 수를 줄일 수 있을 것이다. 실험 결과 제안된 방식이 기존의 시스템과 비슷한 성능을 보였으며 매개 변수 수를 줄이는 방안으로 고려될 수 있음을 보였다.

핵심용어: 분절 특징 은닉 마코프 모델, 매개 변수 수 축소, 분절 특징, 음성 인식
투고분야: 음성처리 분야 (2,5)

In this paper, we propose the reduction method of the number of parameters in the segmental-feature HMM using trend quantization method. The proposed method shares the trend information of the polynomial trajectories by quantization. The trajectory is obtained by the sequence of feature vectors of speech signals and can be divided by trend and location information. The trend indicates the variation of consequent frame features, while the location points to the positional difference of the trajectories. Since the trend occupies the large portion of SFHMM, if the trend is shared, the number of parameters maybe decreases. To exploit the proposed system, the experiments are performed on TIMIT corpus. The experimental results show that the performance of the proposed system is roughly similar to that of previous system. Therefore, the proposed system can be considered one of parameter reduction method.

Keywords: *Segmental-feature HMM, Reduction method of number of parameters, Segmental feature, Speech recognition*
ASK subject classification: *Speech signal processing (2,5)*

I. 서 론

은닉 마코프 모델 (HMM; hidden Markov model)은 구현하기 쉽고, 유연한 모델링 능력과 높은 성능을 보이고 있기 때문에 다양한 분야에서 널리 사용되고 있다. 그러나 비록 HMM의 성능이 뛰어나다고 할지라도 채택하고 있는 가정으로 인하여 음성 신호의 시간적 종속성을 제대로 표현하지 못한다고 알려지고 있다. 이런 단점을 보완하기 위하여 여러 연구방식이 소개되었으며, 대표적인 방법으로는 분절 모델[1,2]을 이용한 방법과 궤적 방식[3,4]을 적용한 방법, 그리고 선형 회귀 방식에 의하여 동적 특성[5]을 표현한 방법이 있다. 이들 방식은 프레임 특징 (frame feature) 대신에 여러 프레임으로 구성된 분절 특징 (segmental feature)을 이용하거나 여러 프레임의 회귀 함수에 의한 매개 변수 (coefficient) 또는 평균값으로 표현하고 있다. 그러나 기존 연구는 음성 신호의 동적 특성을 반영하기 위하여 분절의 길이에 제한을 두

책임저자: 윤영선 (ysyun@mail.hannam.ac.kr)
306-791 대전광역시 대덕구 오정동 133번지
한남대학교 정보통신 · 멀티미디어공학부 (정보통신공학과)
(전화: 042-629-7569; 팩스: 042-629-7843)

지 않고 궤적의 확률 분포와 추정 오차를 통계적 방법으로 정량화시켰기 때문에 계산량이 많다는 약점과, 분절 길이가 변하는 특성 때문에 경계 문제 (boundary problem)가 발생하여 연속 음성 인식에 적용하기 어렵다는 문제점이 있었다. 이러한 문제점을 완화시키고 성능을 향상시키기 위해 여러 프레임 특징을 모수적 궤적 (parametric trajectory)방식을 이용하여 분절 특징으로 표현하고 인식 모델에 적용한 분절 특징 HMM (SFHMM; segmental-feature HMM)이 제안되었다[6-8]. 그러나 제안된 분절 특징 HMM의 성능이 우수하다 할지라도 기존의 HMM을 이용한 음성인식시스템에 비하여 여전히 계산량과 매개 변수의 수가 많다는 문제점이 지적되었다. 따라서 본 연구에서는 분절 특징 HMM의 매개 변수 수를 줄이는 방법에 대해 살펴보고, 그에 따른 성능의 변화를 살펴보도록 한다.

본 논문에서는 매개 변수의 수를 줄이기 위하여 관측된 궤적의 경향 (trend) 정보를 공유하는 경향 공유 분절 특징 HMM (trend tied SFHMM)을 제안한다. 일반적으로 궤적은 경향 (trend) 정보와 위치 (location, offset) 정보로 나눌 수 있다. 경향 정보는 궤적으로 표현되는 음성 신호의 변아를 나타내며, 위치 정보는 그러한 변이 형태의 위치를 나타내는데 분절의 중간 프레임에서의 위치로 표현될 수 있다. 만약 SFHMM이 선형 시스템으로 표현된다면 경향 정보는 1차 함수의 기울기로 표현되며, 2차 시스템이라면 포물선 경향 (parabolic tendency)을 보이게 된다. SFHMM은 특징 표현 방법으로 모수적 궤적 방식을 채택하여 쉽게 경향 정보와 위치 정보로 분리할 수 있기 때문에, 궤적 표현 방식에서 많은 부분을 차지하는 경향 정보를 공유한다면 SFHMM의 매개 변수 수를 줄일 수 있을 것이다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 기존 연구에서 제안된 분절 특징 HMM을 간략히 소개하며, 3장에서는 궤적 정보를 경향 정보와 위치 정보로 분리하는 과정과 공유하는 방법에 대하여 설명한다. 4장에서는 제안된 경향 공유 방식을 이용한 시스템의 성능 평가를 의한 모음 분류 실험에 대해 정리하고, 마지막으로 본 연구의 요약 및 결론을 맺도록 하겠다.

II. 분절 특징 은닉 마코프 모델

본 장에서는 기존 연구에서 제안된 분절 특징 HMM에 대해 간략히 요약하도록 한다. 분절 특징 HMM은 분절 분포를 통계적 방식의 확률 분포를 이용하는 것이 아니라, 각 분절을 다항식으로 표현하고 그 다항식의 확률 분포로써 모델링하는 방법이다. 이 방법은 모수적 방법이 평활화 (smoothing) 효과를 내포하고 있기 때문에 잡음 환경에서도 좋은 성능을 보일 것으로 기대되며, 기존의 HMM에서 쉽게 확장될 수 있다. 또한, 기존의 분절 모델 방식이 가변적인 분절 길이를 채택하여 인식 과정에서 많은 계산 시간을 필요하다는 점을 개선하고자 고정된 분절 길이를 이용하여 특징 표현 방법과 인식 모델을 분리하였다[8].

2.1. 분절 특징

Deng은 HMM의 상태에서의 출력 확률을 표현하기 위하여 절대적인 시간에 대한 다항식으로 상태의 평균 변화를 모델링하였으며[9], 모델을 개선하여 상태에서의 지속 시간으로 관측 확률의 변이를 표현하였다[10]. 이 방법은 음성 특징을 모수적인 방법으로 표현한 것이 아니라 특정 상태에서의 관측 확률을 모수적 방법으로 예측하였다. 다음으로 Gish와 Ng가 핵심어 검출 (word spotting)에서 단어의 경계가 결정이 된 경우, 모수적 방식에 의하여 검출된 단어를 검증하는데 사용하였다[3,11]. 전자의 경우는 다항식을 이용하여 HMM의 상태 모델링을 향상시켰으나 여전히 프레임 특징에 기반을 두어 HMM의 약점으로 지적되고 있는 독립 관측 (independence observation) 가정을 완화시키지 못하였다. 후자의 경우에는 주어진 패턴 전체를 하나의 특징 표현으로 모델링하여 패턴 분류에 사용하였다. 이와 같이 여러 프레임 특징으로부터 얻어지는 특징 표현은 분절 특징 (segmental feature)이라 불린다. 분절 특징 HMM은 HMM의 독립 관측 가정을 완화시키기 위하여 음성 신호를 분절 특징으로 모델링한다. Gish의 방식에서는 분절 길이를 알고 있으며, 서로 다른 분절 길이를 갖는 패턴을 비교하기 위하여 패턴의 길이를 0부터 1로 정규화 (normalize)하였다.

그러나 분절 특징 HMM에서는 연속 음성 인식의 사용에 용이하도록 음성 패턴을 고정된 길이의 분절들의 열로써 표현하였으며 각 분절은 중첩이 가능하도록 하였다. 또한 각 분절은 인접한 분절들과 중첩될 수 있으므로 기준점을 분절의 중앙에 두었다. 이것을 고려하여 고정길이를 갖는 분절을 표현하면 다음과 같이 나타낼 수 있다.

$$\boldsymbol{C}_t = \boldsymbol{Z}\boldsymbol{B}_t + \boldsymbol{E}_t \qquad (1)$$

위 식에서 $\boldsymbol{C}_t$와 $\boldsymbol{B}_t$는 각각 시간 t에서의 음성 분절과 궤적 계수를 나타낸다. 궤적으로 표현되는 분절 특징은 주

어진 분절 안에서 적용할 프레임의 범위와 표현 형태를 나타내는 디자인 행렬 (design matrix) Z와 궤적 계수 B_t의 곱으로 표현된다. 원래의 음성 분절 C_t와 표현된 궤적 ZB_t의 차이로 인하여 발생되는 잔차 오차 (residual error) E_t는 독립적이며 균일하게 분포 (independent and identically distributed)되어 있다고 가정한다. 이 식에서 각 프레임은 D차원의 특징 벡터로 표현되며, 음성 분절 C_t는 $N\times D$ 행렬, Z와 B_t는 각각 $N\times R$, $R\times D$ 차원의 행렬을 나타낸다.

주어진 음성 분절이 $N=2M+1$의 프레임으로 구성된다고 하면, 입력 벡터 $Y=y_1,\ldots,y_N$은 다음과 같이 분절 단위로 표현될 수 있다.

$$C_t = Y_{t-M}^{t+M} = \begin{bmatrix} y_{t-M} \\ \vdots \\ y_t \\ \vdots \\ y_{t+M} \end{bmatrix} \tag{2}$$
$$y_\tau = [\,y_{\tau,1} \cdots y_{\tau,D}], \quad t-M \le \tau \le t+M.$$

분절 표현에서 알 수 있듯이 현재 시간 t의 분절 C_t의 기준점은 분절 중앙에 있는 프레임 특징 y_t가 되기 때문에, $t-1$ 또는 $t+1$시간의 분절과 중첩될 수 있다. 이와 같은 음성 분절을 표현하기 위하여 음성 신호의 프레임 특징 열의 범위를 조절하는 디자인 행렬 Z는 다음과 같이 정의될 수 있다.

$$Z = \begin{bmatrix} 1 & \left(-\frac{M}{2M}\right) & \cdots & \left(-\frac{M}{2M}\right)^{R-1} \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & \left(\frac{M}{2M}\right) & \cdots & \left(\frac{M}{2M}\right)^{R-1} \end{bmatrix} \tag{3}$$
$$z_\tau = \left[1 \ \left(\frac{\tau-t}{2M}\right) \cdots \left(\frac{\tau-t}{2M}\right)^{R-1}\right],$$

여기에서 z_τ는 디자인 행렬 Z의 τ번째 행 벡터 (row vector)를 나타낸다. 디자인 행렬 Z가 현재 시간 t에 대해서 분절 길이로 정규화된 상대적인 위치 정보를 나타내기 때문에 디자인 행렬을 이용하여 계산된 궤적은 바로 이전에 관측된 특징 벡터와 다음에 오는 관측 벡터를 포함하게 된다. 이와 비슷한 방법으로 궤적 계수 행렬 B_t는 다음과 같이 정의된다.

$$B_t = \begin{bmatrix} b_1^t \\ \vdots \\ b_R^t \end{bmatrix} \tag{4}$$
$$b_i^t = [\,b_{i,1}^t \cdots b_{i,D}^t], \quad 1 \le i \le R.$$

음성 분절 C_t와 디자인 행렬 Z가 주어지면 추정되는 궤적 계수 행렬 $\hat{B}_t$는 선형 회귀 (linear regression) 방정식이나 다음과 같은 행렬 연산에 의하여 계산될 수 있다.

$$\hat{B}_t = [\,Z'Z]^{-1}Z'\,C_t, \tag{5}$$

여기에서 $'$는 행렬의 전치 (transpose)를 의미한다. 궤적 계수 행렬 $\hat{B}_t$가 추정되면, 최적 적합도 (goodness-of-fit) χ^2은 시간 t의 분절을 구성하는 모든 프레임 특징에 대한 잔차 오차를 더하여 계산된다.

$$\chi_t^2 = \frac{1}{N}\sum_{\tau=t-M}^{t+M}(y_\tau - z_\tau \hat{B}_t)(y_\tau - z_\tau \hat{B}_t)', \tag{6}$$

여기에서 y_τ와 z_τ는 음성 분절과 디자인 행렬의 행 벡터 (row vector)를 의미하며, 분절의 길이는 $N=2M+1$이다. 위 식에서 최적 적합도 χ^2이 작은 값을 나타내면 추정된 궤적 계수가 원래의 음성 분절을 잘 표현하고 있다는 것을 나타낸다. 이와 같은 과정을 통해 입력 음성 분절에 대응되는 궤적 계수 $\hat{B}_t$와 최적 적합도 χ^2이 인식 시스템의 특징으로 사용된다.

2.2. 분절 우도

분절 HMM (segmental HMM)에서는 특정 상태 (state)에서의 분절 관측 확률을 외적 분절 확률 (extra-segmental probability)과 내적 분절 확률 (intra-segmental probability)의 곱으로 표현하였다[4]. 외적 분절 확률은 화자의 특성이나 특정 음에 대한 발음의 변이와 같은 장기적인 변이를 나타내고, 내적 분절 확률은 연속된 조음 현상이나 불안정한 요소에 의해 발생되는 단기적인 변이 현상을 표현한다. 기존의 분절 HMM에서는 분절의 길이가 가변적이기 때문에 외적 분절 변이와 내적 분절 변이를 정확히 추정하기가 어렵다. 먼저 외적 분절 변이를 고정된 표현 방식 (기울기와 중간 값)을 이용하여 음성 신호로부터 추정하고 다시 구해진 외적 분절 변이를 이용하여 내적 분절 변이를 계산하는 과정으로 변수들을 추정한다. 따라서 학습 시간이 오래 걸린다는 단점이 존재한다. 이러한 문제점을 해결하기 위하여 SFHMM에서는 분절 특징 표현과 인식 단계를 분리하여 독립시켰다.

입력 음성으로부터 추출된 분절 특징을 인식 단계에 적용하기 위하여 SFHMM에서는 외적 분절 변이를 상태에서의 평균 궤적으로 표현하고, 내적 분절 변이는 입력 음성을 분절 특징으로 변환한 경우의 궤적 추정 오차로 나타낸다. 시간 t에서 관측 벡터열 C_t가 단일 궤적 ZB_t

로 표현된다면, 모델 λ의 상태 s_i에서 발생하는 C_t의 관측 확률은 외적 분절 확률과 내적 분절 확률로써 다음과 같이 표현된다.

$$P(C_t|s_i,\lambda)=P(Z\hat{B}_t|s_i,\lambda)P(C_t|Z\hat{B}_t,s_i,\lambda). \quad (7)$$

따라서 시간 t에서 상태 j의 분절 관측 확률은 상태 j의 평균 궤적 $Z\beta_j$와 분산 Σ_j를 이용하여 다음과 같이 표현할 수 있다.

$$b_j(C_t)=P(C_t|s_j,\lambda)$$
$$=P(Z\hat{B}_t|Z\beta_j,\Sigma_j)\cdot P(C_t|Z\hat{B}_t), \quad (8)$$

위 식에서 $P(Z\hat{B}_t|s_i,\lambda)\approx P(Z\hat{B}_t|Z\beta_j,\Sigma_j)$는 장기적인 변이의 외적 분절 확률을 나타내고, $P(C_t|Z\hat{B}_t,s_i,\lambda)\approx P(C_t|Z\hat{B}_t)$는 단기적인 변이를 나타내는 내적 분절 확률을 의미한다. 내적 분절 변이는 음성 분절 C_t에서 추정된 궤적 $Z\hat{B}_t$에 관련되고 모델 λ의 상태 j와 무관하기 때문에 모델 관련 변수를 생략할 수 있다.

$$P(Z\hat{B}_t|Z\beta_i,\Sigma_i)$$
$$\approx\prod_{\tau=t-M}^{t+M}\frac{1}{(2\pi)^{D/2}|\Sigma_{\tau-t,i}|^{1/2}}\cdot$$
$$\exp\left\{-\frac{1}{2}(z_\tau(\hat{B}_t-\beta_i))\Sigma_{\tau-t,i}^{-1}(z_\tau(\hat{B}_t-\beta_i))'\right\}, \quad (9)$$

$$P(C_t|Z\hat{B}_t)=\exp\{-\frac{1}{2}\chi_t^2\}, \quad (10)$$

외적 분절 변이에서 사용되는 분산 Σ_j는 채택된 가정에 따라 각 프레임별로 계산되는 분산 수열을 나타내거나 (시변 분산) 분절 내의 모든 프레임에 공통적으로 적용된 단일의 공통 분산 (고정 분산)을 의미한다. 고정 분산을 채택한 경우, 단일 혼합 밀도 (single mixture) 환경에서는 시변 분산 시스템보다 성능이 저하되나 혼합 밀도의 수가 증가할수록 성능이 향상되며 시변 분산 시스템보다 성능이 뛰어난 것으로 보고되고 있다[8].

III. 경향 공유

SFHMM에서 각 분절은 고정된 길이를 갖으며, 다항식에 의한 궤적으로 모델링된다. 이 궤적은 모수적 방법에 의하여 음성 신호의 특징 열로부터 얻어지기 때문에 궤적 계수로부터 쉽게 경향과 위치 정보를 분리할 수 있다. 경향 정보는 음성의 변화 형태를 표현하며 위치 정보는 분절 특징의 기준 위치를 나타낸다.

3.1. 궤적 정보의 분리

궤적 정보는 선형 회귀 방정식으로 표현될 수 있으며 각 특징 차원은 궤적 계수와 디자인 행렬로부터 다음과 같이 복원된다.

$$y_{\tau,i}=b_{1,i}z_{\tau,1}+b_{2,i}z_{\tau,2}+\cdots+b_{R,i}z_{\tau,R},\ 1\le i\le D, \quad (11)$$

여기에서 $y_{\tau,i}$는 분절에서의 τ번째 프레임의 i차 켑스트럼 (cepstrum) 벡터를 의미하며, $b_{\tau,i}$는 τ번째 궤적 계수를 나타낸다. $z_{\tau,r}$은 디자인 행렬의 요소를 나타내며, $\left(\frac{\tau-t}{2M}\right)^{r-1}$로 표현된다.

위 식에서 디자인 행렬의 첫 번째 행 벡터는 1임을 알 수 있다. 즉 $z_{\tau,1}=1$. 따라서 $b_{1,i}$는 켑스트럼 특징 공간에서의 절편 (intercept)을 의미하게 되고 나머지 부분은 분절 특징의 형태를 나타내는 경향 정보로 해석할 수 있다. 따라서 궤적 표현에서 절편을 제외한 나머지 부분을 공유한다면, 다른 궤적 특징과 경향 정보를 공유한다고 할 수 있다.

SFHMM에서는 현재의 프레임 관측 벡터는 분절의 중앙에 존재한다. 따라서 $b_{1,i}$는 궤적 표현에 의해 평활화 (smoothing)된 가운데 점의 위치를 나타낸다. 만약 식 (11)이 행렬 연산으로 변환되면, 궤적 행렬의 첫 번째 행벡터 b_1는 D차원 위치를 의미하고 나머지 부분은 $(R-1)\times D$차원의 경향을 의미하게 된다. 경향을 공유하기 위해서는 궤적 표현으로부터 경향과 위치를 분리하여야 하는데 행렬의 처음 행벡터를 제거하면 경향 벡터가 되며 다음과 같이 표현된다.

$$T_t=\begin{bmatrix}b_{2'}\\ \vdots\\ b_{R'}\end{bmatrix}. \quad (12)$$

이 경향 벡터를 공유하기 위해서 경향 양자화 방법을 이용한다. 경향 양자화 방법을 이용하여 각 분절 특징을 구성하는 경향 벡터는 가장 가까운 코드워드 (codeword)로 교체된다. 경향 벡터가 이미 학습된 코드북 (codebook)의 공유된 경향 $\hat{T}_t$으로 교체된 후, 기존의 행벡터 b_1과 병합되어 최종 특징 벡터로 사용된다. 변수 추정단계에서도 평균 경향은 경향 코드북에서 선택되고 평균 궤적은 조정된 경향과 위치 정보를 병합하여 구해진다.

그림 1은 경향 공유의 전 과정을 보이고 있다. 입력 음

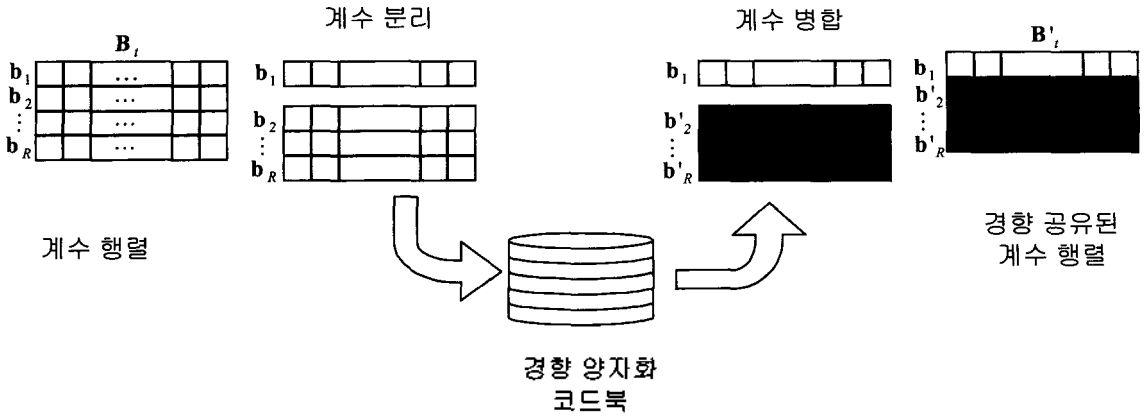

그림 1. 경향 공유 과정: 새로운 궤적 계수 행렬은 원래의 위치 정보와 양자화된 경향 정보를 병합하여 얻어진다
Fig. 1. Flow of trend sharing: The trend-tied trajectory coefficient matrices can be obtained by combining the original location information with the quantized trend information.

성 신호에서 추출된 궤적 벡터는 위치 정보와 경향 정보로 분리된다. 분리된 경향 정보는 경향 양자화 과정을 거쳐 공유된 경향 정보로 대치된다. 그 후 원래의 위치 정보와 다시 결합하여 음성 인식에 사용된다.

3.2. 경향 양자화

경향 양자화 알고리즘은 널리 알려진 벡터 양자화 알고리즘과 유사하다. 그러나 유클리드 거리 (Euclidean distance)로 표현된 거리 척도는 두 경향을 비교하도록 수정되어야 한다. 경향 특성을 반영하기 위하여 유클리드 거리는 다음과 같이 수정된다.

$$D(T_i, T_j) = \frac{1}{N}\sum_{r=1}^{N}\{\tilde{z}_r(T_i - T_j)\}\{\tilde{z}_r(T_i - T_j)\}', \quad (13)$$

여기에서 $\tilde{z}_r$는 경향 벡터에 대응하도록 디자인 행렬에서 첫 번째 열 (column)을 제외한 행 벡터를 나타내고, T_i와 T_j는 경향 벡터를 나타낸다.

IV. 모음 분류 실험

4.1. 실험 조건

제안된 방식의 효과를 검사하기 위하여 16개의 영어 모음에 대해 인식 실험을 하였다. 12차의 MFCC 계수와 정규화된 로그 에너지를 합하여, 13차의 특징 벡터를 만들었으며 1차 미분계수를 더하여 26차의 특징 벡터를 구하였다. 이 26차 벡터는 SFHMM의 분절 특징의 기본 특징과 일반 HMM의 입력 벡터로 사용된다. SFHMM은 분산 표현 방법 중에서 고정 분산을 채택하여 각 분절은 하나의 공통 분산을 이용하도록 하였다. 16개의 영어 모음은 13개의 단모음 /*iy, ih, ey, eh, ae, aa, ah, ao, ow, uw, uh, ux, er*/과 3개의 복모음 /*ay, oy, aw*/으로 구성되었으며, TIMIT 데이터베이스에서 문맥 제약없이 추출하였다. 총 41,429개의 모음이 학습에 사용되었으며, 평가에는 완전 학습 평가용 11,606개의 모음이 사용되었다.

4.2. 실험 및 결과

실험 평가를 하기 위하여 SFHMM의 분절 길이와 회귀 차수, 그리고 혼합 밀도의 수를 변경하며 실험하였으며, 경향 양자화를 위해서 다양한 크기의 코드북을 사용하였다. 성능 변화를 평가하기 위한 기본 시스템의 성능은 표 1에 정리되어 있다. 실험에 사용된 SFHMM은 고정 분산 (fixed variance)을 사용하였다. SFHMM을 표현하는 방식에는 시변 분산 (time-varying variance)과 고정 분산 방식이 있는데, 고정 분산 방식이 혼합 밀도의 수가 증가하면서 시변 분산 방식과 비슷한 성능을 보이며 표현하는데 필요한 매개 변수의 수가 작기 때문에 고정 분산 방식

표 1. 모음 분류 실험을 위한 기본 시스템의 성능 평가
Table 1. The performance of baseline systems for vowel classification.

시스템	조건	*M=1*	*M=2*
HMM	-	52.09	54.45
SFHMM (고정 분산)	*N=3, R=2*	53.33	55.51
	N=3, R=3	53.32	55.53
	N=5, R=2	54.22	56.31
	N=5, R=3	54.03	56.44

표 2. 경향 공유 방식을 적용한 SFHMM의 성능 비교. N: 분절 길이, R: 회귀 차수, D: 경향 코드 북 크기
Table 2. The performance comparison of SFHMMs using trend tying approach. N: segment length, R: regression order, D: codebook size

조건		M=1	M=2	조건		M=1	M=2
D=16	N=3, R=2	52.67	54.75	D=32	N=3, R=2	52.46	54.78
	N=3, R=3	52.00	54.28		N=3, R=3	52.20	54.40
	N=5, R=2	52.88	53.54		N=5, R=2	53.08	54.17
	N=5, R=3	51.69	53.54		N=5, R=3	51.87	52.91
D=64	N=3, R=2	52.55	54.78	D=128	N=3, R=2	52.73	55.06
	N=3, R=3	[illegible]	54.28		N=3, R=3	52.21	54.38
	N=5, R=2	[illegible]	54.20		N=5, R=2	53.22	54.88
	N=5, R=3	51.85	53.60		N=5, R=3	52.60	54.53
D=256	N=3, R=2	52.69	54.99	최대 성능	조건	N=5 R=2	N=3 R=2
	N=3, R=3	52.16	53.69		크기	64	128
	N=5, R=2	53.22	[illegible]		성능	53.82	55.06
	N=5, R=3	52.26	53.64				

을 택했다.

경향을 공유한 경우의 SFHMM의 성능을 비교하기 위하여 경향 양자화 과정에 필요한 코드북의 크기를 16에서부터 256까지 변화시키면서 실험을 하였다. 실험 결과는 표 2에 정리되어 있다. 표 2에서 N과 R은 각각 SFHMM을 표현하는 분절 길이와 회귀 차수를 나타내며, D는 코드북의 크기를 표현한다. 또한 표에서 음영지역 ([illegible])으로 표현된 부분은 각 조건에서의 최고 성능을 나타낸다.

실험 결과, 제안된 시스템은 일반 HMM보다 우수한 성능을 보이나 기존의 SFHMM에 비해 성능저하가 크지 않음을 알 수 있다. 동일 혼합 밀도를 사용하는 경우, 분절 길이 $N=3$인 경우에서 분절 길이가 확장될수록 ($N=5$) 더욱 더 많은 코드북이 필요하고, 혼합 밀도의 수가 증가할수록 마찬가지로 더 많은 코드북이 필요하다는 것을 성능의 변화를 통하여 알 수 있다. 이것은 혼합 밀도의 수가 증가하면서 매개 변수의 수가 증가하였기 때문으로 보인다. 즉, 일반 HMM에서 혼합 밀도의 수가 증가하면 그만큼 매개 변수의 수도 증가하게 된다. 그러나 SFHMM의 경우 혼합 밀도의 수는 증가하더라도 경향 정보에 해당되는 변수의 수는 고정된다. 따라서 혼합 밀도에 대응되는 코드북을 이용하면 성능이 향상될 수 있을 것이다. 또는 궤적 표현 중에서 경향과 위치 정보의 비율 때문에 뚜렷한 성능 향상이 없을 수 있다. 경향 정보는 $N-1$ 프레임에 대해서 계산되나, 위치 정보는 분절의 중앙 프레임에 대해서만 계산되기 때문이다. 따라서 경향과 위치 정보에 대한 비율을 조정한다면 성능 차이는 커질 수 있을 것이다.

V. 결 론

본 논문에서는 다항식의 회귀 함수를 이용하여 분절 특징을 표현하는 SFHMM의 매개 변수 수를 줄이는 방안에 대하여 연구를 하였다. 여러 프레임에 해당되는 분절 특징의 표현으로 모수적 궤적 방식을 이용하였기 때문에 궤적 정보는 간단하게 경향 정보와 위치 정보로 분리될 수 있다. 경향은 분절 특징의 변이를 나타내고, 위치 정보는 궤적의 물리적인 이동을 의미한다. 제안된 방식은 벡터 양자화 알고리즘과 비슷한 방식으로 경향 양자화 과정을 거쳐 경향 정보를 공유한다. SFHMM에서의 경향 정보에 따른 효과를 살펴보기 위하여 영어 데이터베이스인 TIMIT 자료를 이용하여 영어 모음 분류 실험을 하였다. 실험 결과 제안된 방식은 기존의 방식과 현저한 성능 차이를 보이지 않았다. 따라서 제안된 방식은 큰 성능의 저하없이 매개 변수 수를 줄이는 연구로서 고려될 수 있으므로, 계속해서 경향 정보와 위치 정보의 비율 조정이나 여러 경향 코드북의 사용을 통한 성능 향상에 대한 연구가 필요하겠다.

감사의 글

본 연구는 한남대학교 2001년도 교비 학술연구비와 한국전자통신연구원 2002년도 위탁과제의 지원으로 이루어졌습니다.

참 고 문 헌

1. M. J. F. Gales and S. J. Young, "Segmental hidden markov models," *In Proceedings of European Conference on Speech Communication and Technology*, 1579-1582, 1993.
2. M. Ostendorf, V. Digalakis, and O. A. Kimball, "From HMMs to segment models: A unified view of stochastic modeling for speech recognition," *IEEE Tr. on Speech and Audio Processing*, **4** (5), 360-378, 1996.
3. H. Gish and K. Ng, "Parametric trajectory models for speech recognition," *In Proceedings of International Conference on Spoken Language Processing*, i-466-469, 1996.
4. W. J. Holmes and M. J. Russell, "Probabilistic trajectory segmental HMMs," *Computer Speech and Language*, **13**, 3-37, 1999.
5. S. Furui, "Speaker-independent isolated word recognition using dynamic features of speech spectrum," *IEEE Trans. on Acoustics, Speech and Signal Processing*, **34** (1), 52-59, 1986.
6. 윤영선, 오영환, "모수적 궤적 기반의 분절 HMM을 이용한 연속 음성 인식," 한국음향학회지, **19** (3), 35-44, 2000.
7. 윤영선, 오영환, "분절 특징 HMM의 매개 변수 수의 감소에 관한 연구," 한국음향학회지, **19** (7), 48-52, 2000.
8. 윤영선, "분절 특징 HMM을 이용한 영어 음소 인식," 한국정보과학회지, **29** (3), 167-179, 2002.
9. L. Deng, "A generalized hidden markov model with state conditioned trend functions of time for the speech signal," *Signal Processing*, **27**, pp. 65-78, 1992.
10. L. Deng and M. Aksmanovic and D. Sun and J. Wu, "Speech recognition using hidden markov models with polynomial regression functions as non-stationary states," *IEEE Trans. on Speech and Audio Processing*, **2** (4), 507-520, 1994.
11. H. Gish and K. Ng, "A segmental speech model with application to word spotting," *In Proc. of Int. Conf. on Acoustics, Speech and Signal Proc.*, II-447-450, 1993.

저자 약력

● 윤 영 선 (Young-Sun Yun)

1990년 2월 한국과학기술원 전산학과 (학사)
1992년 2월 한국과학기술원 전산학과 (석사)
1992년 3월 ~ 1995년 7월: ㈜한디소프트 주임연구원
1995년 9월 ~ 2001년 2월: 한국과학기술원 전산학과 (박사)
2001년 3월 ~ 현재: 한남대학교 정보통신 · 멀티미디어 공학부 조교수
※ 주관심분야: 음성 인식, 패턴 인식, 음성 정보 검색 등