분절 특징의 경향 공유에 관한 연구

윤 영 선
한남대학교 정보통신 · 멀티미디어공학부

A study on trend tying of the segmental-feature

Yun Young-Sun
School of Information Technology and Multimedia Engineering, Hannam University
Email: ysyun@mail.hannam.ac.kr

요 약

본 논문에서는 분절 특징 HMM(SFHMM)의 매개 변수를 줄이는 방법을 제안한다. SFHMM이 HMM보다 우수한 성능을 보이더라도, SFHMM의 매개 변수 수는 HMM보다 많기 때문에 매개 변수 수를 줄이는 방법에 대한 연구가 필요하다. 일반적으로 궤적(trajectory)은 경향(trend) 정보와 위치(location) 정보로 분리될 수 있다. 경향은 분절 특징의 변이를 나타내며, SFHMM 변수의 많은 부분을 담당하기 때문에, 경향 정보를 공유할 수 있다면 SFHMM의 매개 변수 수는 감소될 수 있을 것이다. 제안된 방법은 궤적의 경향 정보를 양자화(quantization)에 의하여 공유한다. 제안된 방법의 성능을 살펴보기 위하여 영어 데이터베이스인 TIMIT 자료를 사용하여 실험하였다. 실험 결과 제안된 방법의 성능은 기존 연구와 거의 유사하나, 궤적의 다양한 정보를 이용한다면 궤적 정보의 공유에 의하여 매개 변수 수를 줄일 수 있을 것으로 보인다.

1. 서론

HMM은 구현하기 쉽고 유연한 모델링 능력을 가지고 있어, 다양한 분야에서 널리 이용되고 있다. 그러나 HMM은 적용된 약한 가정으로 인하여 음성 신호의 동적인 특성을 제대로 반영하지 못한다고 보고 되었다[1]. HMM의 약점을 보완하기 위한 여러 연구가 진행되고 있는데, 대표적인 연구로는 분절 모델(segmental model)[2,3]과 궤적에 의한 접근 방식(trajectory approach)[1,4]을 들 수 있다. 이들 연구는 음성 인식에 많이 사용되는 프레임 특징 대신 분절 특징(segmental feature)을 사용하거나, 프레임 특징들의 회귀 함수(regression function)를 이용한다. 이들 접근 방법에 기초하여 분절 특징 HMM(SFHMM; segmental-feature HMM)[5,6]이 제안되었으며, 분절 특징 HMM은 입력된 음성 신호를 프레임 특징으로 표현하고, 여러 프레임 특징을 모수적 궤적 방식을 이용하여 분절 특징으로 표현하였다. SFHMM은 모든 프레임에 대하여 공통의 분산을 사용하거나, 각 프레임에 대해 개별적인 분산을 적용할 수 있다. 그러나 SFHMM이 HMM보다 성능이 좋다할 지라도, SFHMM을 구성하는 변수의 수는 HMM의 변수 수보다 많다는 단점이 있다. 따라서 SFHMM의 매개 변수 수를 줄이는 연구가 필요하다고 본다.

본 논문에서는 매개 변수를 줄이기 위하여 관측된 궤적의 경향(trend) 정보를 공유하는 경향 공유(trend tied) SFHMM을 제안한다. 일반적으로 궤적은 변이의 형태에 해당되는 경향 정보와 분절의 중앙에 해당되는 위치(offset, location) 정보로 분리할 수 있다. 만약 궤적이 2차 방정식으로 표현된다면 경향은 포물선 형태를 띠게 된다. SFHMM이 모수적 궤적 시스템(parametric trajectory system)을 이용하기 때문에, 경향과 위치 정보는 쉽게 분리될 수 있으며, 이는 매개 변수 수를 줄이는 한 방법으로 여겨질 수 있다.

2. 분절 특징 HMM

음성 신호의 연속된 음향 특징 벡터들(분절 특징)은 특징 공간에서의 궤적 형태로 표현될 수 있다. 이 궤적은 모수적 방법(parametric approach)이나 비모수적 방법(non-parametric approach)에 의하여 표현될 수 있으며, 분절의 길이가 제한 받을 수 있다. 기존 연구에서 제안된 SFHMM은 평활화 효과와 구현이 쉽도록 모수적 방법을 채택하고 고정된 크기의 분절을 이용하여 모

델링하였다. 즉, SFHMM에서는 입력 음성 신호를 일반적인 특징 벡터로 변환하고, 이들 특징으로부터 분절 특징을 추출한 후 인식과정에서 이용한다.

2.1 분절 특징

Deng은 1992년에 HMM 상태에서의 시변 출력 확률 분포를 표현하기 위하여 다항식을 이용하여 상태를 모델링하는 모수적 방법을 제안하였다[7]. 또 다른 연구에서는 Gish와 Ng가 음성 분절의 특징을 표현하기 위하여 다항식의 회귀 함수를 이용하였다[4]. 이들 접근 방식 중에서 Deng의 방법은 분절에서의 특징 표현방법이라기보다 상태에서 생성된 관측을 모델링하는 것이기 때문에, 본 연구에서는 분절 표현 방법으로 Gish의 방법을 선택하였다. 그러나 Gish의 방법은 가변 길이를 갖는 분절을 모델링하기 때문에, 연속 음성 인식에 사용할 경우 경계 문제(boundary problem)가 발생한다. 따라서 SFHMM에서는 각 분절의 길이를 고정시켜 계산 시간 및 복잡도를 줄였다. 고정 길이를 갖는 분절은 다음과 같이 표현된다.

$$C_t = ZB_t + E, \tag{1}$$

여기에서 C_t와 B_t는 시간 t에서의 음성 분절과 분절을 표현하는 궤적의 계수를 나타낸다. 이 식에서 분절 특징은 디자인 행렬 Z를 이용하여 계산된다. 각 프레임은 D차원의 특징 벡터이며, 와 B_t는 각각 $N \times R$과 $R \times D$차원의 2차 행렬을 나타내며, E는 잔차 오차를 나타낸다.

잔차 오차가 독립적이며 균등하게 분포된다고 가정하였기 때문에, 궤적 계수 행렬 $\hat{B}_t$는 선형 회귀 방정식 또는 다음의 행렬 연산에 의하여 계산될 수 있다.

$$\hat{B}_t = [Z'Z]^{-1}Z'C_t, \tag{2}$$

여기에서 $'$는 행렬의 전치(transpose)를 의미한다.

궤적 계수 $\hat{B}_t$가 추정되면, 최대 적합도 (*goodness-of-fit*)는 시간 t에서 분절을 구성하는 각 프레임의 잔차 오차를 합해서 다음과 같이 구한다.

$$\chi_t^2 = \frac{1}{N}\sum_{\tau=t-M}^{t+M}(c_\tau - z_\tau\hat{B}_t)(c_\tau - z_\tau\hat{B}_t)', \tag{3}$$

여기에서 c_τ와 z_τ는 음성 분절과 디자인 행렬의 열벡터(row vector)를 의미하며, 분절의 길이는 $N = 2M + 1$이다. 위 식에서 χ_t^2의 값이 작으면 데이터 정합이 잘 이루어졌다는 것을 의미한다. 분절에 대한 변수가 추정되면, 각 분절은 궤적 계수 행렬 $\hat{B}_t$와 적합도 χ_t^2로 표현된다.

2.2 분절 우도

분절 HMM(segmental HMM)에서 분절의 관측 확률은 외적 분절 확률(extra-segmental probability)과 내적 분절 확률(intra-segmental probability)의 곱으로 표현된다. 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. 반면에 SFHMM에서는 외적 분절 변이를 평균 궤적으로 표현하고, 내적 분절 변이를 궤적의 추정 오차로 표현한다.

SFHMM의 시간 t에서의 관측 벡터 열 C_t가 단일 궤적 ZB_t로 표현된다면, 모델 λ의 상태 s_i에서 발생하는 C_t의 관측 확률은 다음과 같이 표현될 수 있다.

$$P(C_t|s_i,\lambda) = P(Z\hat{B}_t|s_i,\lambda)P(C_t|Z\hat{B}_t,s_i,\lambda). \tag{4}$$

따라서 시간 t에서 상태 j의 분절 관측 확률은 다음과 같이 표현된다.

$$b_j(C_t) = P(C_t|s_j,\lambda) = P(Z\hat{B}_t|Z\beta_j,\Sigma_j)\cdot P(C_t|Z\hat{B}_t), \tag{5}$$

여기에서 β_j와 Σ_j는 상태 j에 해당되는 궤적 모델이다. 위 식에서 외적 분절 확률과 내적 분절 확률은 다음과 같이 정의된다.

$$P(Z\hat{B}_t|Z\beta_i,\Sigma_i) = \prod_{\tau=t-M}^{t+M} b\frac{1}{(2\pi)^{D/2}|\Sigma_{\tau-t,i}|^{1/2}}\cdot \exp\left\{-\frac{1}{2}\{z_\tau(\hat{B}_t-\beta_i)\}\Sigma_{\tau-t,i}^{-1}\{z_\tau(\hat{B}_t-\beta_i)\}'\right\}, \tag{6}$$

$$P(C_t|Z\hat{B}_t) = \exp\{-\frac{1}{2}\chi_t^2\}, \tag{7}$$

Σ_j는 적용된 분산 표현 방법에 따라 각 프레임에 대한 분산 열이거나, 단일의 공통 분산을 의미한다.

3. 경향 공유

SFHMM에서 각 분절은 고정된 길이를 갖으며, 다항식에 의한 궤적으로 모델링된다. 이 궤적은 음성 신호의 특징 열로부터 얻어지며, 경향과 위치 정보로 분리될 수 있다. 경향 정보는 연속된 프레임 특징 벡터의 변이를 표현하며, 위치 정보는 궤적의 기준 위치를 나타낸다.

3.1 궤적 정보의 분리

궤적은 선형 회귀 방정식으로 표현될 수 있는데, 각 특징 차원에 대하여 다음의 다항식이 고려될 수 있다.

$$y_{\tau,i} = b_{1,i}z_{\tau,1} + b_{2,i}z_{\tau,2} + \cdots + b_{R,i}z_{\tau,R}, 1 \le i \le D, \tag{8}$$

여기에서 $y_{\tau,i}$는 분절에서의 τ번째 프레임의 i차 캡스트럼 벡터를 의미하며, $b_{\tau,i}$는 τ번째 궤적 계수를 나타낸다. 마지막으로 $z_{\tau,r}$은 디자인 행렬의 요소를 나타내며,

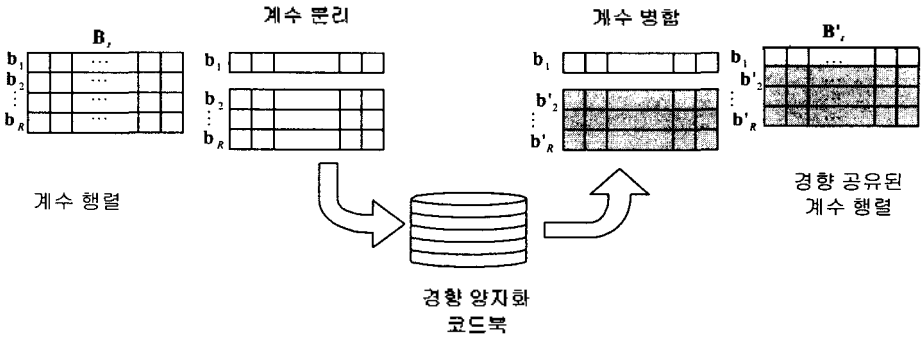

그림 1 경향 공유 과정 : 새로운 궤적 계수 행렬은 원래의 위치 정보와 양자화된 경향 정보를 병합하여 얻어진다.

$\left(\frac{\tau-t}{2M}\right)^{i}$로 표현된다.

위 식에서 디자인 행렬의 첫 번째 행벡터는 1임을 알 수 있다, 즉 $z_{\tau,1}=1$. 따라서 $b_{1,i}$는 캡스트럼 특징 공간에서의 절편(intercept)을 의미하게 되고 나머지 부분은 분절 변이에 해당되는 경향과 관련된다. 따라서 궤적 표현에서 절편을 제외한 나머지 부분을 공유한다면, 다른 궤적과 경향 정보를 공유한다고 할 수 있다.

SFHMM에서는 현재의 프레임 관측 벡터는 분절의 중앙에 존재한다. 따라서 $b_{1,i}$는 궤적 표현에 의해 평활화된 가운데 점의 위치를 나타낸다. 만약 식 (8)이 행렬 연산으로 변환되면, 궤적 행렬의 첫 번째 열벡터 $\boldsymbol{b}_1$는 D차원 위치를 의미하고, 나머지 부분은 $(R-1)\times D$ 차원의 경향을 의미하게 된다. 경향을 공유하기 위해서는 궤적 표현으로부터 경향과 위치를 분리하여야 하는데, 행렬의 처음 열벡터를 제거하면 경향 벡터가 되며, 다음과 같이 표현된다.

$$\boldsymbol{T}_i=\begin{bmatrix}\boldsymbol{b}_{2'}\\ \vdots \\ \boldsymbol{b}_{R'}\end{bmatrix}. \qquad (9)$$

이 경향 계수는 경향 양자화 과정을 거쳐서 가장 가까운 코드워드로 교체된다. 경향 계수가 이미 학습된 코드북의 새로운 경향 $\hat{\boldsymbol{T}}_i$으로 교체된 후, 기존의 열벡터 $\boldsymbol{b}_1$과 병합되어 최종 특징 벡터로 사용된다. 변수 추정단계에서도 평균 경향은 경향 코드북에서 선택되고, 평균 궤적은 조정된 경향과 위치 정보를 병합하게 된다.

그림 1은 경향 공유의 전 과정을 보이고 있다. 제안된 시스템에서, 학습 단계에 사용되는 모든 경향 정보는 가장 가까운 코드워드로 조정된다.

3.2 경향 양자화

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

$$D(\boldsymbol{T}_i,\boldsymbol{T}_j)=\frac{1}{N}\sum_{\tau=1}^{N}\left[\tilde{z}_\tau(\boldsymbol{T}_i-\boldsymbol{T}_j)\right]\left[\tilde{z}_\tau(\boldsymbol{T}_i-\boldsymbol{T}_j)\right]', \qquad (10)$$

여기에서 $\tilde{z}_\tau$는 디자인 행렬에서 첫 번째 행을 제외한 열벡터를 의미하고, $\boldsymbol{T}_i$와 $\boldsymbol{T}_j$는 경향 계수 행렬을 나타낸다.

4. 실험 결과

제안된 방식의 효과를 검사하기 위하여 16개의 영어 모음에 대해 인식 실험하였다. 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 사용한다. SFHMM은 분산 표현 방법 중에서 고정 분산을 채택하여 각 분절은 공통된 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 제약 없이 추출하였다. 총 41,429개의 모음이 학습에 사용되었으며, 평가에는 완전 학습 평가용 11,606개의 모음이 사용되었다.

실험 평가를 하기 위하여 SFHMM의 분절 길이와 회귀 차수, 그리고 혼합 밀도의 수를 변경하며 실험하였으며, 경향 양자화를 위해서는 256 단계의 코드북을 사용하였다. 실험 결과는 표 1에 정리되어 있다.

실험 결과, 제안된 시스템은 일반 HMM보다 뚜렷한 성능 향상을 보이지 않았다. 단일 혼합 밀도를 이용하

표 1 다양한 분절 조건에서의 모음 인식률 비교 (M은 혼합 밀도의 수를 나타내며, 경향 양자화 단계는 256이다.)

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
SFHMM (경향 공유)	*N=3, R=2*	53.25	54.95
	N=3, R=3	52.90	54.26
	N=5, R=2	53.32	54.44
	N=5, R=3	53.06	55.01

는 경우에는 HMM보다 성능 향상이 있었으나, 두 개의 혼합 밀도를 이용하는 경우에는 거의 유사하였다. 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.

5. 결론

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 뚜렷한 성능 차이를 보이지 않았다. 그러나 성능 차이가 뚜렷하지 않더라도 제안된 방식은 매개 변수 수를 줄이는 연구로서 고려될 수 있으므로, 계속해서 성능을 향상시키기 위한 방안으로써 경향 정보와 위치 정보의 비율 조정이나 여러 경향 코드북의 사용에 대한 연구가 필요하겠다.

참고문헌

1. W.J.Holmes, M.J.Russell, "Probabilistic trajectory segmental HMMs," *Computer Speech and Language*, vol. 13, pp. 3-37, 1999

2. M.J.F.Gales, S.J.Young, "Segmental Hidden Markov Models," In *Proc. of European Conf. on Speech Comm. and Tech.*, pp. 1579-1582, 1993

3. M.Ostendorf, V.Digalakis, O.A.Kimball, "From HMMs to Segment models: A Unified View of Stochastic Modeling for Speech Recognition," *IEEE Tr. on Speech and Audio Processing*, vol.4, no. 5, pp. 360-378, 1996

4. H.Gish, K.Ng, "Parametric trajectory models for speech recognition," In *Proc. of Int. Conf. on Spoken Lang. Proc.*, pp. I-466-469, 1996

5. Y.-S.Yun, Y.-H.Oh, "A Segmental-Feature HMM for Speech Pattern Modeling," *IEEE Signal Processing Letters*, vol. 7, no. 6, pp. 135-137, 2000

6. Y.-S.Yun, Y.-H.Oh, "A Segmental-Feature HMM for Continuous Speech Recognition Based On a Parametric Trajectory Model," *Speech Communication*, (to appear)

7. L.Deng,"A generalized hidden Markov model with state-conditioned trend functions of time for the speech signal," *Signal Processing*, vol. 27, pp.65-78, 1992