

HMM 인식기에서 상태별 다중 특징 파라미터 가중

State-Dependent Weighting of Multiple Feature Parameters in HMM Recognizer

손 종 목*, 배 건 성*

(Jong Mok Son*, Keun Sung Bae*)

요 약

본 논문에서는 특징 파라미터의 분산과 인식성능에 대한 기여도를 고려하여 각 특징 파라미터를 가중시키는 방법을 제안하였다. 각 특징 파라미터의 인식률에 비례하게 전체 기여도를 설정하고, 각 특징 파라미터의 분산에 따라 가중요인을 설정하였다. 전체 기여도와 분산에 따른 가중요인을 사용하여 각 특징 파라미터의 상태별 가중치를 설정하였다. 제안한 방법의 유효성을 살펴보기 위해 유사음소 단위의 HMM 음성인식시스템을 사용하여 인식실험을 하였다. 인식실험에서 제안한 방법으로 가중치를 설정하였을 경우에 인식률이 7.7% 향상됨을 볼 수 있었다.

ABSTRACT

In this paper, we proposed a new approach to weight each feature parameter by considering the dispersion of feature parameters and its degree of contribution to recognition rate. We determined the total distribution factor that is proportional to recognition rate of each feature parameter and the dispersion factor according to the dispersion of each feature parameter. Then, we determined state-dependent weighting using the total distribution factor and dispersion factor. To verify the validity of the proposed approach, recognition experiments were performed using the PLU(Phoneme-Like Unit)-based HMM. Experimental results showed the improvement of 7.7% at the recognition rate using the proposed method.

I. 서 론

인간에게 있어서 가장 자연스러운 정보전달 수단은 음성이다. 컴퓨터의 발달과 더불어 음성을 정보전달 수단으로 이용하려는 시도가 음성인식, 음성합성, 화자인식, 음성부호화 등 여러 분야에서 진행되어 왔다. 이 가운데 음성인식에 관한 연구는 단어단위 인식의 경우 신뢰할 수 있는 단계까지 와 있으며, 이를 이용한 제품들이 상용화되고 있다. 하지만, 보다 다양한 분야에서 음성인식 기술을 사용하기 위해서는 높은 인식성능을 제공하는 음성인식시스템이 요구된다.

음성인식시스템은 음성신호의 특성을 나타내는 방법에 따라 인식성능에 영향을 많이 받는다. 음성의 음향학적 특성을 나타내는 음성 특징 파라미터로는 에너지, 멜 첩스트림 등의 정적 특징 파라미터들과 이들의 시간적 변화특성을 나타내는 차분 에너지, 차분 멜 첩스트림 등의 동적 특징 파라미터들이 많이 사용되고 있으며, 이들 음성 특징 파라미터들이 각각 음성신호의 다른 특성을 표

현하고 있기 때문에 각 음성 특징 파라미터가 가지는 음성 간 분별력은 다르다고 할 수 있다. 일반적으로 HMM을 사용한 음성인식시스템에서는 인식성능 향상을 위해 다수의 특징 파라미터들을 사용하는 경우가 많은데, HMM의 각 상태(State)에 따라 특징지어지는 음성의 특성이 다르기 때문에 각 상태에서 특징 파라미터별 중요도도 다르게 된다. 또한, HMM의 각 상태에서 출력확률은 특징 파라미터 각각의 관측벡터에 대한 출력확률의 곱으로서 나타나므로 출력확률 계산시 HMM의 상태별로 각 특징 파라미터의 중요도에 기반해 가중치를 부여한다면 음성인식시스템의 성능향상을 기대할 수 있을 것이다[1-4]. 특징 파라미터의 가중치를 설정하기 위한 방법으로 discriminative 훈련 알고리즘[1-3]과 ML(Maximum Likelihood) 훈련 알고리즘[4]을 응용한 방법 등이 제안되었다.

본 논문에서는 단어단위의 HMM 음성인식시스템을 사용한 인식실험을 통하여 각 특징 파라미터의 전체 기여도를 조사하고, HMM의 상태별로 각 특징 파라미터의 전체 기여도와 분산에 따른 가중요인에 기반해 훈련 후 처리 과정으로 특징 파라미터의 가중치를 설정하는 방법을 제안하였다. 제안한 가중치 설정방법의 유효성을 확인하기 위한 인식실험에서는 인식의 기본단위로 유사음소

* 경북대학교 전자공학부

접수일자: 1999년 1월 12일

를 사용하였으며, 각 유사음소를 위한 HMM의 상태전이 모델은 3상태 Bakis(Left-to-Right)모델을 사용하였다. 각 유사음소 HMM의 훈련은 Segmental K-Means 알고리즘과 Baum-Welch 재추정 알고리즘으로 수행하였으며[5-6], 음성 특징 파라미터로는 에너지, 멜 캡스트럼의 정적 특징 파라미터와 차분 에너지, 차분 멜 캡스트럼의 동적 특징 파라미터를 사용하였다. 단어단위의 인식실험을 통해 구한 전체 기여도와 상태별 특징 파라미터의 분산에 따른 가중치에 따라 제한한 방법으로 가중치를 설정하였을 경우에 가중치를 일정하게 설정하였을 때보다 인식시스템의 성능이 향상되는 것을 음소단위의 인식실험에서 확인할 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서 특징 파라미터별 가중치 방법을 제안하고, 3장에서 인식실험 결과를 비교·분석하며, 마지막으로 4장에서 결론을 맺고 앞으로의 연구방향을 제시한다.

II. 상태별 다중 특징 파라미터의 가중

2.1 다중 특징 파라미터의 사용

음성인식시스템은 음성신호의 특성을 모델링하는 방법에 따라 인식 성능에 영향을 많이 받는다. 음성인식에 사용되는 특징 파라미터들은 각각 고유한 특성을 가지고 있기 때문에, 음성인식에 유효한 신호의 특성을 다른 측면에서 표현하는 음성 특징 파라미터를 다수 사용하는 것이 바람직하다. 많이 사용되는 음성 특징 파라미터로는 에너지, 멜 캡스트럼 등의 정적 특징 파라미터와 이들의 시간적 변화특성을 나타내는 차분 에너지, 차분 멜 캡스트럼 등의 동적 특징 파라미터가 있다.

HMM을 사용한 음성인식시스템에서 다수의 음성 특징 파라미터를 사용하는 것이 단일 특징 파라미터를 사용하는 것에 비해서 높은 인식성능을 얻을 수 있는 것으로 알려져 있다. 음성인식시스템에 다수의 특징 파라미터를 사용하는 방식은 크게 두가지로 나눌 수 있다. 하나는 다수의 특징 파라미터들을 하나의 특징 파라미터로 병합하여 사용하는 방법이고, 나머지는 각각 독립적인 형태로 사용하는 방법이다. 인식성능의 측면에서 다수의 특징 파라미터를 하나의 병합된 특징 파라미터로 사용하는 방법보다 독립적으로 사용하는 방법이 보다 나은 인식성능을 얻을 수 있는 것으로 알려져 있다[7]. 특히, 이산분포 HMM이나 준연속분포 HMM의 경우 다수의 특징 파라미터들을 하나의 특징 파라미터로 병합해 사용할 때, 각 특징 파라미터별로 모델이 설정되는 것이 아니라 하나의 병합된 특징 파라미터에 대해 모델이 설정되기 때문에 오류가 커지게 된다. 이에 비해서, 복수의 특징 파라미터들을 독립적으로 사용하는 경우에는 각각의 특징 파라미터별로 모델이 설정되기 때문에 오류를 줄일 수 있다. 즉, 복수의 특징 파라미터들을 독립적으로 사용하는 경우가 병합된 단일 특징 파라미터를 사용하는 경우보다 상세한 표현이 가능해진다. 그림 1에 각각의 경우를 나타내었다.

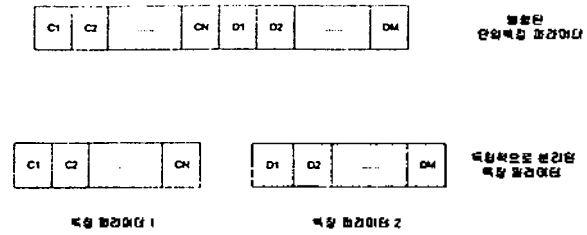


그림 1. 다수의 특징 파라미터 사용 예
Fig. 1. An example of using multiple feature parameters.

2.2 상태별 특징 파라미터 가중

HMM을 사용한 음성인식시스템에서는 다수의 음성 특징 파라미터를 사용하는 경우가 많은데, 이들 각 특징 파라미터가 잘 반영하는 음성의 특성이 서로 다르기 때문에 HMM의 각 상태에서 특징 파라미터의 중요도 또한 다르게 된다. 따라서, 출력확률 계산시 HMM의 상태별로 각 특징 파라미터의 중요도에 기반해 가중치를 부여한다면 음성인식시스템의 성능향상을 기대할 수 있을 것이다. 특징 파라미터의 가중치를 설정하기 위한 방법으로 discriminative 훈련 알고리즘[1-3]과 ML 훈련 알고리즘[4]을 이용한 방법 등이 제안되었다.

다수의 특징 파라미터를 사용하는 HMM 음성인식시스템에서 출력확률은 특징 파라미터 각각의 관측벡터에 대한 출력확률 곱의 형태로 나타난다. 일반적으로 HMM에서 특징 파라미터 s 의 관측벡터 O_i^s 가 주어진 i 번째 모델의 j 상태에서 발생할 확률은 식 (1)에 의해 주어진다.

$$b_{ij}^s(O_i^s) = \sum_{k=1}^M c_{ik}^s N(O_i^s; \mu_{ik}^s, \sum_{l=1}^L \sigma_{ik}^s) \quad (1)$$

위의 식에 $N(\cdot)$ 은 평균 μ_{ik}^s 와 분산 $\sum_{l=1}^L \sigma_{ik}^s$ 를 가지는 가우시안 확률밀도 함수를 나타내고, c_{ik}^s 는 가지계수를 나타내며, M 은 가지수를 나타낸다.

HMM 음성인식시스템이 다수의 특징 파라미터를 사용하였을 경우, 관측벡터 O_i 는 각 특징 파라미터의 관측벡터 O_i^s 의 결합으로 이루어진다. 때문에, 관측벡터 O_i 의 관측확률 값은 각 특징 파라미터의 관측벡터 O_i^s 의 관측확률 값을 식 (2)와 같이 그 중요도에 따라 지수적으로 가중시켜 곱해 구할 수 있다[1-4].

$$b_{ij}(O_i) = \prod_{s=1}^S [b_{ij}^s(O_i^s)]^{w_{ij}^s} \quad (2)$$

여기서 S 는 음성인식시스템에서 사용된 특징 파라미터의 수이고, w_{ij}^s 는 i 번째 모델의 j 상태에서 s 번째 특징 파라미터의 가중치를 나타낸다. 가중치를 부여한 관측확률 값의 계산에서 가중치가 작을 경우 그에 따른 관측확

를 낮은 커지게 되고, 최적 상태열에서 가중치의 합이 작은 상태에 많은 관측열이 나타나게 된다. 때문에, 관측열의 발생확률을 최대로 하는 최적 상태열을 찾을 때, 가중치의 합이 미치는 영향을 막기 위해 각 상태에서 특징 파라미터 가중치의 합을 식 (3)에 주어진 것과 같이 일정하게 한다.

$$\sum_j w_{ij}^s = Const. \quad (3)$$

HMM의 각 상태에서 음성 특징 파라미터들의 중요도는 다르다. 본 연구에서는 각 특징 파라미터의 인식능력에 비례하게 설정한 전체 기여도와 각 상태에서의 특징 파라미터 분산에 따른 가중요인에 따라 각 음성 특징 파라미터의 중요도를 결정하고, 이에 따라 가중치를 설정하였다. 각 특징 파라미터가 잘 나타내는 음성의 특징이 서로 다르기 때문에, 각 특징 파라미터가 가지는 음성간 분별력 또한 다르다고 할 수 있다. 가중치 설정에서 각 특징 파라미터의 전체 기여도는 각 특징 파라미터의 인식률에 비례하게 하였다. 또한, 같은 상태에서 특징 파라미터의 분산이 작을수록 분별력이 있다고 할 수 있기 때문에, 특징 파라미터의 분산이 작을수록 가중치를 크게 설정하였다. 일반적으로 에너지나 멜 캡스트럼의 정적 특징 파라미터보다, 차분 에너지나 차분 멜 캡스트럼과 같은 동적 특징 파라미터의 분산이 작기 때문에, 평균적으로 동적 특징 파라미터에 가중치가 많이 부여되는 것을 막기 위해 각 특징 파라미터의 평균 분산 값에 대한 상대적 분산 값을 사용해 가중치를 설정하였다.

각 상태에서의 특징 파라미터의 분산은 식 (4)와 같이 정의된다[8].

$$\sigma_{ij}^s = \sum_{l=1}^K p_{il}^s \text{trac}(\Sigma_{ijl}^s) + \sum_{l=1}^K \sum_{m=1}^K p_{il}^s p_{jm}^s \| \mu_{ijl}^s - \mu_{ijm}^s \|^2 \quad (4)$$

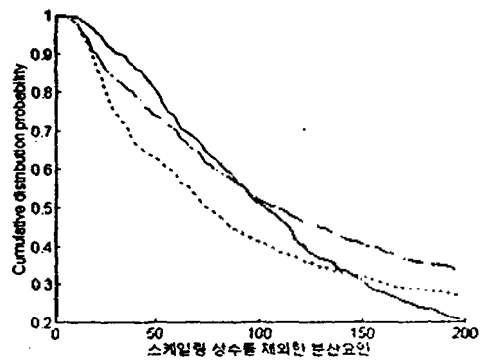
여기서 $\sum_{l=1}^K$ 는 i 번째 모델의 j 상태에서 k 번째 가지의 분산 행렬이고, p_{il}^s 는 k 번째 가지의 발생확률이다. 또한, 분산에 따른 가중요인 d_{ij}^s 는 식 (6)와 같이 정의하였다.

$$n_{ij}^s = \| m_{ij}^s - m_{ij}^s \| \quad (5)$$

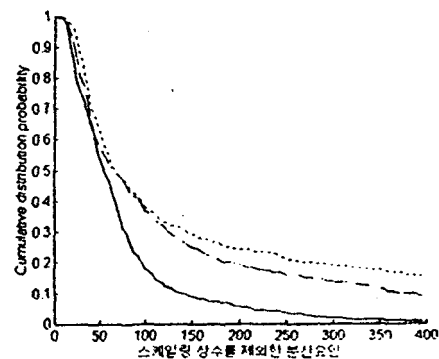
$$d_{ij}^s = Sc^s \cdot \frac{\sigma_{ij}^s}{n_{ij}^s} \quad (6)$$

여기서, m_{ij}^s 는 특징 파라미터 s 의 전체 평균 값이고, m_{ij}^s 는 특징 파라미터 s 의 특정상태에서의 평균값이다. n_{ij}^s 는 i 번째 모델의 j 상태에서 특징 파라미터 s 의 분산에 대한 가중치이다. 이는, 전체 모델에 대해 그 상태

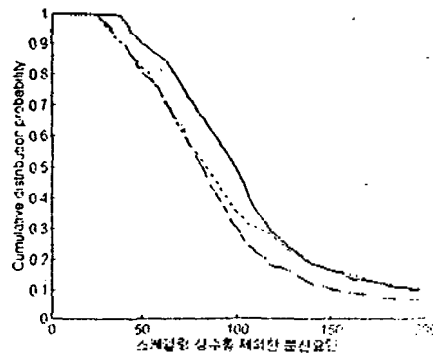
에서의 특징 파라미터가 지니는 분별력의 정도를 나타내는 인자인데, 실험에서는 특징 파라미터의 특정상태에서의 평균 값과 전체 평균 값과의 거리에 비례하게 식 (5)와 같이 설정하여 사용하였다. Sc^s 는 각각의 특징 파라미터 분산에 따른 가중요인의 전체 평균 값을 모두 같게 해주기 위한 스케일링 상수이다. 그림 2에 스케일링 상수를 제외한 분산에 따른 가중요인의 누적분포를 나타내었다. 평균적으로 상태 2에서 정적 파라미터들의 분산에 따른 가중 요인이 상태 1, 3에 비해 작은 값을 나타내고, 동적 파라미터들의 분산에 따른 가중요인이 큰 값을 보임을 알 수 있다.



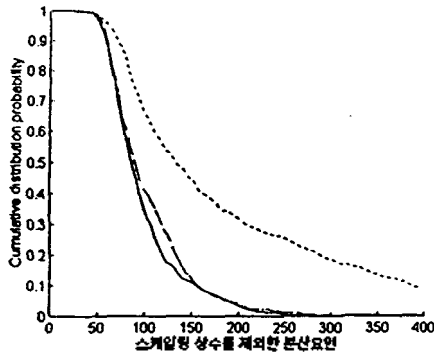
(a) 에너지



(b) 차분 에너지



(c) 멜 캡스트럼



(d) 차분 멜캡스트럼

상태 1 : —, 상태 2 : — —, 상태 3 : ····

그림 2. 스케일링 상수를 제외한 분산 요인의 누적분포
Fig. 2. Cumulative distribution of the dispersion factor except the scaling constant.

특징 파라미터의 전체 기여도는 단어단위의 인식실험을 통하여 구하였는데, 각 특징 파라미터만을 사용하였을 때의 인식률에 비례하게 설정하였다.

HMM의 각 상태별 특징 파라미터의 가중은 각 특징 파라미터의 전체 기여도가 클수록, 분산에 따른 가중요인이 작을수록 가중치를 크게 설정하였다. 또한, 식 (3)의 조건을 만족시키도록 식 (7)과 같이 가중치를 설정하였다.

$$w'_s = R^s(1-\alpha) + \left(\sum_{i=1}^s R^i \right) \frac{(1/d'_s)}{\sum_{i=1}^s (1/d'_i)} \alpha \quad (7)$$

여기서, R^s 는 s 번째 특징 파라미터의 인식성능에 비례하게 설정한 전체 기여도에 따른 가중요인이고, d'_s 는 특징 파라미터의 분산에 따른 가중요인이다. α 는 전체 기여도에 의한 가중요인과 분산에 따른 가중요인의 영향을 조절하는 상수이다.

III. 실험 및 고찰

3.1 음성인식시스템의 구성

HMM의 훈련과 인식실험을 위한 데이터로는 ETRI(Electronics and Telecommunications Research Institute)의 611DB와 445DB를 사용하였다. 실험에 사용한 데이터는 16KHz, 16bits로 샘플링하였다. 이 실험 데이터를 Pre-emphasis 계수 0.95로 전처리 한 후, 16ms 길이의 해밍 윈도우를 8ms간격으로 이동하며 구간단위로 분석하였다. 각 구간에서 1차의 에너지와 13차의 멜 캡스트럼을 구하고, 이로부터 현재 구간을 포함한 전후 각 3구간(전체 7구간)의 정보를 이용하여 1차의 차분 에너지와 13차의 차분 멜 캡스트럼을 구하였다. 모델의 구성시 각 특징 파

라미터들은 독립적으로 다루어졌다. 표 1은 음성 데이터의 분석조건을 나타내고 있다.

표 1. 음성 데이터 분석조건
Table 1. Analysis condition of speech data.

Sampling Frequnc	16 KHz
Quantization	16 bits
Hamming Window	16 ms(256 points)
Frame Rate	8 ms(128 points)
Feature Parameters	Energy(1 order) △ -Energy(1 order) MFCC(13 order) △ -MFCC(13 order)

인식실험은 크게 두가지로 나누어서 행하였다. 하나는 각 특징 파라미터의 전체 기여도를 살펴보기 위해 행한 단어단위의 인식실험이고, 다른 하나는 가중치 설정의 유효성을 확인하기 위해 행한 유사음소단위의 인식실험이다. 유사음소의 경우 인식 대상을 구성하는 방법에 따라 인식시스템의 성능이 달라지기 때문에, 특징 파라미터의 분별력 이외의 요인이 전체 기여도 측정에 영향을 끼치는 것을 줄이기 위하여 전체 기여도를 살펴보기 위한 실험은 단어단위로 하였다.

각 특징 파라미터의 전체 기여도를 살펴보기 위한 단어단위 인식실험에서는 상태의 생략이 없는 Bakis 모델을 사용하였다. 각 HMM의 상태수는 인식대상어가 포함하고 있는 음소의 수를 고려하여 단어모델마다 다르게 설정하였다. 인식대상어로는 ETRI와 445DB중 40개의 단어를 선정하였고, 445DB의 훈련용 남성 화자들의 데이터를 이용하여 훈련하였다. 각 모델은 상태의 수에 따라 등분할하여 상태전이 확률을 초기화한 후, Segmental K-Means 알고리즘으로 관측빈도 함수를 초기화하였다. 초기화된 모델을 Baum-Welch 재추정 과정을 사용하여 훈련하였다.

특징 파라미터의 전체 기여도와 분산에 따른 가중치 부여의 유효성을 확인하기 위해 유사음소단위의 음성인식시스템을 구현하여 인식실험을 하였다. 인식대상어 구성을 위한 기본 모델로는 문맥종속(Context Dependent) 유사음소단위의 SCHMM(Semi-Continuous HMM)을 사용하였다. 앞에서 제시한 바와 같이 하나의 유사음소를 모델링하기 위해 3상태 Bakis 모델을 사용하였는데, 기본 음소 중 그 길이가 짧은 음소를 모델링하기 위해 상태의 생략을 허용하였다. 그림 3은 유사음소 HMM을 위해 사용한 상태전이 모델의 구성을 나타내고 있다. 음성 특징 파라미터의 동적 범위와 정밀도를 높이기 위해 각 특징 파라미터들로부터 독립적인 기본 관측단위를 생성하여 인식에 사용하였다. 각 특징 파라미터의 기본 관측단위 생성을 위해 Modified K-Means 알고리즘을 사용하였고, 기본 관측단위를 사용한 각 유사음소 모델의 훈련은 Baum-Welch 재추정 알고리즘을 이용하여 하였다. 훈련

데이터로는 ETRI의 음소 레이블링된 DB인 611DB를 사용하였다. 43개의 기본 유사음소를 사용하여 Reduction Rule을 적용한 결과 1,392개의 유사음소 모델을 얻었다 [9]. 인식대상어를 구성하기 위해 유사음소 모델을 연결한 인식망을 구성하여 인식실험에 사용하였고, 인식 과정으로는 Viterbi 알고리즘을 사용하였으며, 인식시 소요되는 시간을 줄이기 위해 Beam 탐색법과 Pruning 기법을 사용하였다.

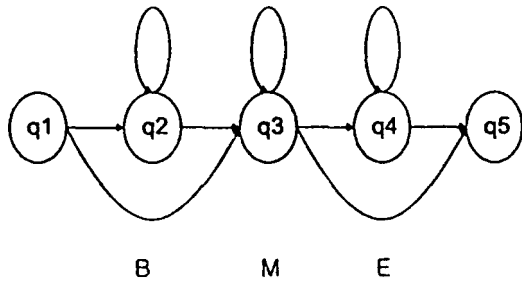


그림 3. 유사음소 HMM을 위한 상태전이 모델의 구성
Fig. 3. The topology of a state transition model for a PLU HMM.

3.2 실험 결과 및 분석

음성인식시스템에서 각 특징 파라미터의 인식성능에 대한 전체 기여도를 살펴보기 위해, 먼저 HMM을 사용한 단어기반 인식실험을 하였다. 단어기반 인식실험을 위한 음성 자료로는 ETRI의 445DB에서 40개의 단어를 선정하여 사용하였다. 445DB의 훈련용 남성 화자에 대한 1,280개 음성 데이터를 사용하여 모델을 훈련하고, 나머지 테스트용 남성 화자의 480개 음성 데이터를 평가용 자료로 사용하였다.

그림 4는 각 특징 파라미터의 전체 기여도를 살펴보기 위한 인식실험의 결과인데, 멜 캡스트럼이 99.2%로 인식률이 가장 높고, 차분 멜 캡스트럼 98.95%, 차분 에너지 35.42%, 에너지 29.5%의 순으로 나타났다. 이로부터 각 특징 파라미터가 음성인식시스템의 성능에 미치는 영향이 다양을 알 수 있다. 또한, 전체 기여도는 멜 캡스트럼, 차분 멜 캡스트럼, 차분 에너지, 에너지의 순서로 높음을 알 수 있으며, 식 (3)의 Const. 값을 4로 하여 각 특징 파라미터별 인식률에 비례하게 설정한 에너지와 멜 캡스트럼, 차분 에너지, 차분 멜 캡스트럼의 전체 기여도에 따른 가중 요인은 각각 0.4486, 1.5080, 0.5386, 1.504이다. 표 2에 각 특징 파라미터별 인식률과 그에 비례하게 설정한 전체 기여도에 따른 가중요인을 특징 파라미터별로 나타내었다.

표 2. 전체 기여도에 따른 가중요인

Table 2. Weighting factor for the degree of contribution to recognition rate.

특징 파라미터	에너지	멜 캡스트럼	차분 에너지	차분 멜 캡스트럼
인식률	29.5%	99.2%	35.42%	98.95%
가중요인	0.4486	1.5080	0.5386	1.5048

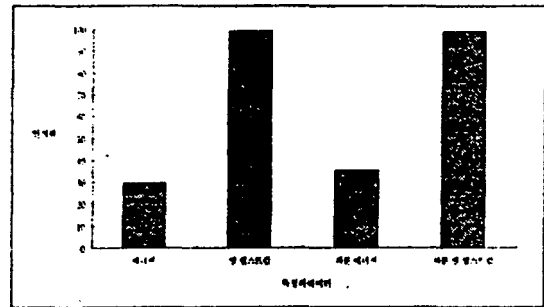


그림 4. 단어단위 인식실험에서 특징 파라미터별 인식률
Fig. 4. Recognition rate for each feature parameter in word-based recognition experiment.

특징 파라미터의 전체 기여도와 상태별 분산에 따른 가중치 부여의 유효성 확인을 위해 유사음소 기반 SCHMM 음성인식시스템을 이용한 단어단위의 인식실험을 하였다. 대상단어로는 ETRI의 445DB중 앞의 단어기반 인식실험에서 사용하지 않은 40개의 단어를 선정하였다. 선정된 대상단어에 대한 유사음소 인식망을 구성하여 인식실험에 사용하였는데, 445DB의 훈련용 남성 화자와 테스트용 남성 화자의 총 1,760개 음성을 평가용 자료로 사용하였다.

분산에 기반한 가중요인은 에너지나 멜 캡스트럼 등 정적 파라미터의 경우 상태 2의 값이 상태 1, 3의 값보다 낮게 나타나고, 차분 에너지나 차분 멜 캡스트럼 등 동적 파라미터의 경우 상태 2보다 상태 1, 3의 값이 낮게 나타나는 경향을 보였다. 이에 따라, 상태 1, 3에서의 특징 파라미터 가중은 동적 특징 파라미터를 강조하고, 상태 2에서의 특징 파라미터 가중은 정적 특징 파라미터를 강조하는 경향을 보였다. 이는 상태 1, 3에서 상태 2보다 음성의 변화 특성이 중요시 됨을 나타낸다고 할 수 있다.

표 3은 특징 파라미터 가중치 설정에 따른 인식률을 나타내었다. 모든 특징 파라미터의 가중치를 균일하게 두었을 경우 85.63%의 인식률을 나타내었으나, $\alpha=0.1$ 의 값으로 식 (6), (7)을 사용하여 가중치를 부가하였을 경우 93.35%의 인식률을 보여, 7.7%의 인식성능 향상을 확인할 수 있었다.

표 3. 가중치 설정에 따른 인식률
Table 3. Recognition rate for the setting of weighting value.

	동일한 가중치	$\alpha=0$	$\alpha=0.1$	$\alpha=0.2$
인식률	85.63%	91.66%	93.35%	93.24%

IV. 결 론

본 논문에서는 다수의 특징 파라미터를 사용한 HMM 음성인식시스템에서, 각 특징 파라미터의 인식능력에 따른 전체 기여도와 상태별 분산에 따른 가중요인을 이용하여 가중치를 설정하는 방법을 제안하였다. 제안한 가중치 설정 방법의 유효성을 확인하기 위하여, 유사음소 기반의 SCHMM 음성인식시스템을 구현하고 인식실험을 하였다. 각 음성 특징 파라미터의 전체 기여도를 구하기 위하여, 먼저 단어단위의 HMM 음성인식시스템을 구현하였다. 각각의 특징 파라미터를 사용하여 단어단위 인식실험을 한 결과 델 캡스트립, 차분 델 캡스트립, 차분 에너지, 에너지의 순서로 인식성능이 높음을 알 수 있었다. 이에 따라, 전체 기여도에 의한 가중요인을 설정하고, 이를 가중치 부여에 사용하였다. 또한, HMM의 각 상태에서 특징 파라미터별 분산에 따른 가중요인을 구하여 특징 파라미터별 전체 기여도와 함께 가중치 설정에 이용함으로써 인식성능을 개선시키고자 하였다. 가중치 설정의 유효성을 확인하기 위해 유사음소 단위의 음성인식시스템을 구현하고 인식실험을 하였다. 유사음소 단위의 음성인식시스템을 사용한 인식실험은 인식 대상어에 따른 인식방울 구성하여 단어단위로 이루어졌다.

인식실험 결과 모든 특징 파라미터들에 동일한 가중치를 사용하였을 경우보다 본 논문에서 제안한 각 특징 파라미터의 인식능력에 따른 전체 기여도와 상태별 분산에 따라 혼련 후처리 과정으로 가중치를 설정하였을 경우 향상된 인식성능을 나타내었다. 음성인식시스템의 인식성능 향상을 위해서는 특징 파라미터 선정과 각 특징 파라미터가 가지는 특성을 인식시스템에 반영하는 방법에 관한 연구가 계속 이루어져야 할 것이다.

참 고 문 헌

1. C.M. del Alamo, F.J. Caminero-Gil, C. dela Torre-Munilla, and L. Hernandez-Gomez, "Codebook Weights Adaptation for Discriminative Training of SCHMM-Based Speech Recognition Systems," *EUROSPEECH*, vol. 1, pp. 93-96, 1995.
2. I. Rogina, and A. Waibel, "Learning State-Dependent Stream Weights for Multi-Codebook HMM Speech Recognition System," *ICASSP*, vol. 1, pp. 217-220, 1994.
3. J. Hernando, J. Ayarte, and E. Mouic, "Optimization of Speech Parameter Weighting for CDHMM Word Recognition," *EUROSPEECH*, vol. 1, pp. 105-108, 1995.
4. J. Hernando, "Maximum Likelihood Weighting of Dynamic Speech Features for CDHMM Speech Recognition,"

ICASSP, vol. 2, pp. 1267-1270, 1997.

5. L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. of IEEE*, vol. 77, no. 2, pp. 257-286, February 1989.
6. B.-H. Juang, and L.R. Rabiner, "The Segmental K-Means Algorithm for Estimation Parameters of Hidden Markov Models," *IEEE Trans. on ASSP*, vol. 38, no. 9, pp. 1639-1641, September 1990.
7. K.F. Lee, H.W. Hon, R. Reddy, "An Overview of the SPHINX Speech Recognition System," *IEEE Trans. ASSP*, vol. 38, no. 1, pp. 35-44, 1990.
8. Y. Zhao, "A Speaker-Independent Continuous Speech Recognition System Using Continuous Mixture Gaussian Density HMM of Phoneme-Sized Units," *IEEE Trans. on SAP*, vol. 1, no. 3, pp. 345-361, July 1993.
9. C.-H. Lee, L.R. Rabiner, R. Pieraccini, and Jay G. Wilpon, "Acoustic Modeling of Subword Units for Speech Recognition," *ICASSP*, vol. 2, pp. 721-724, 1990.

▲손 증 목(Jong Mok Son)



1997년 2월: 경북대학교 전자공학과 (공학사)
1999년 2월: 경북대학교 전자공학과 (공학석사)
1999년 3월~현재: 경북대학교 전자공학과 박사과정 재학 중
※ 주관심분야: 디지털 신호처리, 음성신호처리, 음성 인식

▲배 건 성(Keun Sung Bac)



1977년 2월: 서울대학교 전자공학과 (공학사)
1979년 2월: 한국과학기술원 전기 및 전자공학과(공학석사)
1989년 5월: University of Florida (공학박사)
1979년 3월~현재: 경북대학교 전자공학과 교수
※ 주관심분야: 음성분석 및 인식, 디지털 신호처리, 지능통신, 음성 부호화, 웨이브렛 분석 등