

HMM을 기본으로한 집단화 방법의 불특정화자 단어 인식에 응용

The Application of an HMM-based Clustering Method to Speaker Independent Word Recognition

임 현*, 박 순 영*, 방 만 원*
(H. Lim*, S. Y. Park*, M. W. Bang*)

요 약

본 논문에서는 단어를 발음하는 방법이 각각 다른 화자들의 변이성을 잘 흡수하도록 복수개의 통계적인 모델들을 구성하기 위하여 HMM을 기본으로 하는 집단화 방법을 제시한다. 또한 개발된 방법으로 부터 얻어진 HMM집단화된 모델들을 불특정화자 고립단어 인식에 응용된다.

HMM 집단화 방법은 학습용 데이터로부터 어떤 경계치 보다 낮은 유사도를 갖는 관측열들을 분리하여 새로운 집단을 만들고 이 집단내에 있는 관측열들을 이용하여 새로운 모델들을 학습시키는 방법이다. 집단화 과정은 반복되는데 최고의 유사도를 갖는 모델의 집단에 관측열들을 재 분배하고 집단내 관측열들이 변화하면 새로운 모델을 재추정하여 기존의 모델을 대신한다. 그러므로 이 집단화 방법은 집단화 과정과 파라미터 추정이 일체화되어 기존의 패턴에 의한 집단화 방법보다 더욱 효율적이 된다.

실험결과 HMM에 의한 집단화 방법이 기존의 패턴에 의한 집단화 방법보다 고립 숫자음 인식에 있어서 1.43%의 인식률을 향상시킬 수 있었으며 단일 모델의 사용보다는 2.08%의 인식률이 향상되었다.

ABSTRACT

In this paper we present a clustering procedure based on the use of HMM in order to get multiple statistical models which can well absorb the variants of each speaker with different ways of saying words. The HMM-clustered models obtained from the developed technique are applied to the speaker independent isolated word recognition.

The HMM clustering method splits off all observation sequences with poor likelihood scores which fall below threshold from the training set and create a new model out of the observation sequences in the new cluster. Clustering is iterated by classifying each observation sequence as belonging to the cluster whose model has the maximum likelihood score. If any cluster has changed from the previous iteration, the model in that cluster is reestimated by using the Baum-Welch reestimation procedure. Therefore, this method is more efficient than the conventional template-based clustering technique due to the integration capability of the clustering procedure and the parameter estimation.

Experimental data show that the HMM-based clustering procedure leads to 1.43% performance improvements over the conventional template-based clustering method and 2.08% improvements over the single HMM method for the case of recognition of the isolated korean digits.

I. 서 론

음성인식 기술은 초기에는 특정화자에 의한 단독 숫자

음, 지역명, 전화번호등의 고립단어 인식이 주를 이루었으나 최근에는 화자 독립적이며 연속 숫자음, 회화음성 인식을 위한 연구가 진행되고 있다. 화자 독립에 의한 연속음성 인식은 발생자 개개인의 발생기관의 구조와 발생 습관등의 차이에서 비롯된 시간적 및 주파수적 변동을 흡수하는 방법이 요구되어 현재의 기술로는 화자 독립

*목포대학교 전자공학과
접수일자 : 1994년 3월 4일

시스템의 인식률이 화자 종속 시스템의 인식률에 비하여 뒤떨어진다.

음성인식 방법에는 패턴정합(pattern matching)에 의한 방법^[1]과 음성신호의 시간에 따른 특징변화를 통계적으로 모델링하는 Hidden Markov Model(HMM)을 이용한 방법^{[2][3]}, 그리고 신경회로망을 이용하는 방법^[4] 등이 있는데 이 방법들이 불특정화자 단어 인식에 응용될 때에는 여러 계층의 화자로 부터 발생된 기준 음성들을 가지고 각 단어에 대하여 표준 패턴(reference pattern or template)의 작성 또는 모델을 학습시키는 것이 필요하다.

현재까지 모든 개인차를 제거할 수 있는 특정 파라미터의 정규화 방법이 개발되지 못했기 때문에 패턴 매칭 방법이나 확률적인 모델링 방법을 사용한 불특정화자 음성 인식에 있어서 학습용 데이터의 양이 충분히 많을 때는 각 인식 대상의 단어에 대하여 한 종류의 기준 패턴이나 모델을 사용하는 것 보다는 복수개의 기준 패턴이나 모델들을 사용하여 인식률을 향상시키려는 노력을 하였다. 여기서 복수개의 패턴이나 모델들을 생성할 때 사용한 기존의 방법들은 먼저 학습 데이터로 사용된 각각의 단어 패턴들 사이에서 발생하는 음성의 신축특성을 동적으로 가장 잘 보상해주는 dynamic time-warping (DTW) 기법^[5]을 사용하여 유사도를 측정하며 이 유사도를 기본으로 학습용 단어들을 비슷한 패턴끼리 여러 집합(set)으로 집단화(clustering)^[6] 한다. 마지막 단계는 각 집합 내에 속하여 있는 단어들을 대표할 수 있는 표준패턴의 생성과 또는 각 집합에 속하여 있는 단어들을 사용하여 모델을 추정한다. 위와 같은 기존의 방법은 패턴 매칭을 기본으로 한 음성 인식 방법에서 우수한 인식 결과를 보여주었으며 이산 HMM 방법에 의한 음성 인식시 필요한 코드북 작성을 위하여도 사용되어지고 있다.

본 연구에서는 각 화자들의 변이성을 잘 흡수할 수 있도록 복수개의 모델을 구성하기 위하여 기존의 패턴사이의 거리 유사도에 의한 집단화 방법을 사용하지 않고 현재의 HMM 파라미터에 대한 관측열들의 확률적인 유사도를 바탕으로 한 집단화 방법^[7]을 사용하여 모델을 구성하였으며, 이의 성능 평가를 위해 기존의 패턴의 집단화에 많이 이용되고 있는 MKM(Modified K-means) 알고리즘^[6]을 이용하여서 구성된 복수개의 모델들과 인식 실험을 비교하였다.

II. Hidden Markov Model(HMM)^{[2][3]}

이산화된 심볼들로 이루어진 관측열 $O_1 O_2 \dots O_T$ 가 Markov chain으로 모델될 수 있다고 가정하면 어떤 stochastic 프로세스가 매 시각에 어느 상태로 천이되며 이 천이된 상태에서 관측심볼이 어떤 확률을 가지고 발생한다고 볼 수 있다. 이러한 과정을 수식으로 표현하면 모델 $\lambda = (A, B, \Pi)$ 가 되며 여기서 A는 상태 천이 확률 분포로서

$$A = \{a_{ij}\}, 1 \leq i \leq N, 1 \leq j \leq N \quad (1)$$

이며 a_{ij} 는 상태 i 에서 상태 j 로 천이될 확률이고 N은 상태수이다. 관측 심볼 확률 분포 B는

$$B = \{b_j(k)\}, 1 \leq j \leq N, 1 \leq k \leq M \quad (2)$$

이며 $b_j(k)$ 는 현재 상태 q_j 에서 관측심볼 v_k 가 나타날 확률이며 M은 관측심볼의 갯수이다. 마지막으로 Π 는 초기 상태 확률 분포로서 $\Pi = \{\pi_1, \pi_2, \dots, \pi_N\}$ 이다.

초기 모델이 주어지면 학습데이터를 사용하여 관측열의 발생확률을 최대화시키도록 모델 파라미터를 개선하기 위한 반복적인 재추정 방법이 이용되는데 만약 학습데이터가 서로 독립적인 K개의 관측열들로 구성되었으며 각각의 관측열들의 길이가 T_k 라고 가정하면 Baum-Welch 알고리즘에 의한 파라미터 재추정식은 다음과 같다.

$$\bar{a}_{ij} = \frac{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \alpha_t^k(i) a_{ij} b_j(o_{t+1}^{(k)}) \beta_{t+1}^k(j)}{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \alpha_t^k(i) \beta_t^k(i)} \quad (3)$$

$$\bar{b}_j(k) = \frac{\sum_{k=1}^K \sum_{t=1, o_t^{(k)}=v_k}^{T_k} \alpha_t^k(j) \beta_t^k(j)}{\sum_{k=1}^K \sum_{t=1}^{T_k} \alpha_t^k(j) \beta_t^k(j)} \quad (4)$$

여기서 $\alpha_t(i)$ 와 $\beta_t(i)$ 는 다음과 같이 forward-backward 알고리즘에 의하여 구할 수 있다.

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N a_{ij} \alpha_t(i) \right] b_j(o_{t+1}), 1 \leq t \leq T-1, 1 \leq j \leq N \quad (5)$$

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), T-1 \geq t \geq 1, 1 \leq i \leq N \quad (6)$$

관측열 O 가 모델 λ 에 의하여 나타날 수 있는 확률은 가장 최적의 경로를 선택하여 관측확률을 구하기 때문에 계산량을 줄일 수 있는 Viterbi 알고리즘^[8]이 이용되는데 간단히 정리하면 먼저 변수 $\delta_t(i)$ 는 다음과 같이 초기 화후 반복 수행하여 얻어진다.

$$\begin{aligned} \delta_1(i) &= \pi_i b_i(o_1), 1 \leq i \leq N \\ \delta_t(j) &= \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(o_t), 2 \leq t \leq T, 1 \leq j \leq N \end{aligned} \quad (7)$$

최적 상태열에 의하여 관측열이 발생할 확률 P^* 는 변수 $\delta_t(i)$ 로부터 다음과 같이 얻을 수 있다.

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (8)$$

III. HMM 집단화 방법

학습데이터가 충분히 많을 때에는 복수개의 통계적 모델을 사용함으로써 인식률을 향상시킬 수 있다. 본 장에서는 복수개의 모델들을 구성하기 위하여 집단내의 모델과 관측열들의 확실적인 유사도를 바탕으로 하는 HMM 집단화 방법의 알고리즘을 기술한다.

HMM 집단화의 기본적인 원리는 그림 1과 같이 표현할 수 있다. N 명의 화자가 어떤 특정 단어를 한번씩 발음한 N 개의 단어들의 특징 벡터들로부터 얻어진 이산화된 유한개의 심볼(symbol)로 구성된 N 개의 관측열(observation sequence)들의 집합 Ω 를 학습데이터로 가정한다.

$$\Omega = \{O^{(1)}, O^{(2)}, \dots, O^{(N)}\} \quad (9)$$

이 학습데이터를 이용하여 HMM의 모델 파라미터를 추정하며 $O^{(k)}$ 는 k 번째의 관측열로서 $O^{(k)} = \{O_1^{(k)}, O_2^{(k)}, \dots, O_{T_k}^{(k)}\}$ 이며 T_k 는 k 번째 관측열의 길이가 된다.

먼저, 전체의 학습 데이터 Ω 로부터 최적 모델 λ 를 추정하는데 $\prod_{k=1}^N P[O^{(k)}|\lambda]$ 를 최대화하기 위하여 Baum-Welch의 재추정 알고리즘을 이용한다.

다음은 현상태의 모델에 대한 모든 학습용 관측열들의 확실적 유사도를 계산하여 어떤 경계치 보다 작은 유사도를 가지고 있는 관측열이 발견되면 이런 관측열로 새로운 집단을 만든다. 즉,

$$O^{(n)} \in (W^+) \text{ if } P[O^{(n)}|\lambda] > \tau$$

$$O^{(n)} \in (W^-) \text{ O.W.} \quad (10)$$

for all $O^{(n)} \in \Omega$.

여기서 경계치 τ 의 선택은 집단화 과정의 수렴속도와 인식률에 영향을 주게 되는데 특히 학습용 데이터가 충분치 않을때는 너무 작거나 큰 경계치 사용에 의하여 어떤 집단은 극소수의 관측열들로만 모델이 학습되게 되고 결과적으로 인식률이 저하될 가능성이 높게 된다. 따라서 본 연구에서는 경계치의 효과적인 결정 방법으로서 매번 모집단내에 있는 학습용 관측열들의 유사도에 대한 산술평균을 사용하여 새로 분리되는 집단(W^+)과 (W^-)에 모집단 관측열들의 절반 정도가 할당되게 한다.

분리된 두 집단(W^+)과 (W^-)을 구성하고 있는 관측열들을 이용하여 새로운 모델 λ^+ 와 λ^- 을 재추정하여서 모델의 갯수를 증가시킨후 새롭게 구성된 모델들이 속한 집단들에 전체 학습용 관측열들을 다음과 같이 재분배한다.

$$O^{(n)} \in (W^+) \text{ iff } P[O^{(n)}|\lambda^+] \geq P[O^{(n)}|\lambda^-],$$

$$O^{(n)} \in (W^-) \text{ O.W.} \quad (11)$$

for all $O^{(n)} \in \Omega$.

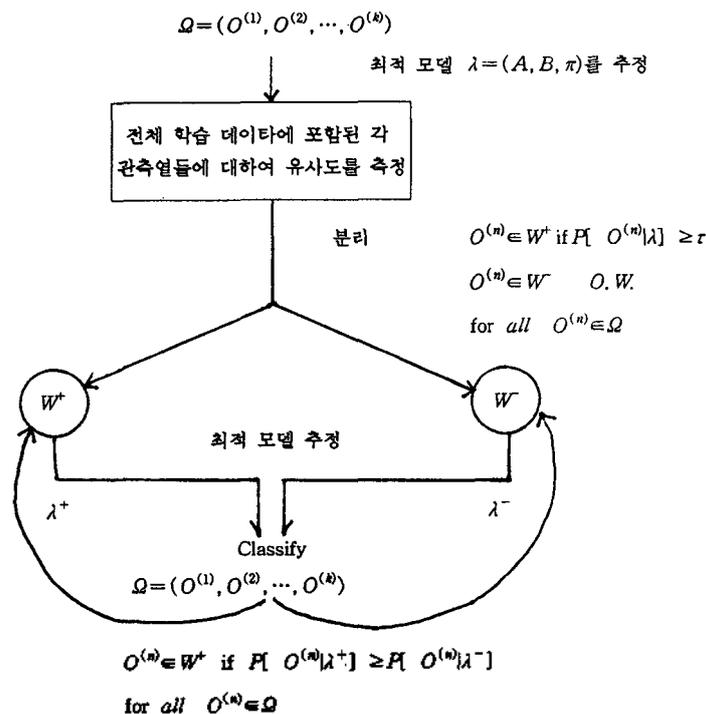


그림 1. HMM을 기본으로 한 집단화 방법의 기본원리
Fig 1. Basic principle of an HMM based clustering method

즉, 각각의 새로운 모델에 대해 모든 학습 관측열들의 유사도를 구하여 유사도가 더 높은 집단에 각 관측열을 재분류한 후 집단내의 새롭게 구성된 관측열들로부터 모델을 재추정하여 기존의 모델을 대신하게 한다. 이 과정은 각 집단의 관측열들의 변화가 없을때 까지 반복된다.

마지막으로 세개 이상의 모델들을 구성하기 위해서는 각 집단 (W^i)에서 추정된 모델 λ^i 와 집단내에 있는 관측열들의 평균 집단내 유사도를 다음과 같이 계산한다.

$$\bar{L}_i = \frac{1}{|W^i|} \sum_{o^m \in W^i} P[o^m | \lambda^i]. \quad (12)$$

여기서, $|W^i|$ 은 집단 (W^i)에 속하는 관측열의 갯수이다. 다음은 가장 작은 \bar{L}_i 를 가지고 있는 집단 (W^i)을 발견하여 이 집단을 두 집단으로 분리시키는데 그 방법은 앞에서 기술한 바와 같이 가장 작은 \bar{L}_i 를 가진 집단내에 있는 각 관측열들이 모델 λ^i 에 대한 유사도를 계산하여 어떤 경계치보다 작은 유사도를 가지고 있는 관측열이 발견되면 이런 관측열로 새로운 집단을 만든후 분리된 각 집단을 Baum-Welch 재추정 알고리즘을 이용하여 모델 파라미터를 재추정한다. 또한 각각의 새로운 모델에 대해 모든 학습용 관측열들의 유사도를 구하여 유사도가 더 높은 집단에 각 관측열을 재분류하고 각 집단에 대한 모델을 재추정하게 되며 이 과정은 각 집단의 관측열들의 변화가 없을때 까지 반복된다. 결국 이러한 과정을 통해 모델들의 갯수를 증가시킬 수 있다.

IV. 실험방법 및 결과

4.1 실험 방법

본 연구에서 사용한 음성 데이터는 남성화자 14명과 여성화자 14명이 조용한 환경의 실험실에서 정상시의 발음으로 표 1과 같은 11개 한국어 숫자음을 10회씩 발성한 것으로서 이 음성신호는 4.5kHz의 차단주파수를 갖는 저역역파기를 통과하여 10kHz로 샘플링 되었으며 각 샘플들은 14비트로 구성되었다.

표 1. 단어인식 실험에 사용된 음성 데이터
Table 1. Speech data used in word recognition

공	일	이	삼	사	오	육	칠	팔	구	영
---	---	---	---	---	---	---	---	---	---	---

음성의 전처리 과정은 음성신호의 입력에서 HMM 학습 이전까지의 과정으로서, 끝점 검출, 특징 추출 그리고 벡터 양자화를 포함한다. 먼저, 본 연구에서는 끝점 검출을 위해 Rabiner와 Sambur에 의해 제안된 단시간 평균 에너지(short-time average energy)와 단시간 영교차율(short-time zero crossing rate)을 이용한 알고리즘을 사용하였다^[9]. 음성데이터의 처음 100ms는 묵음이라 가정하여 파라미터를 구했으며 영교차율에 의한 끝점 보상

은 100ms 이내로 하였다. 다음으로 음성신호의 주파수 스펙트럼이 6dB/oct의 기울기로 감소되는 특성때문에 고주파 성분을 나타내기 위하여 6dB/oct로 입력신호를 보상하여 주는 전처리(preemphasis) 과정이 필요한데 본 실험에서는 $H(z) = 1 - 0.95Z^{-1}$ 과 같이 1차 디지털 필터를 사용하였다.

음성신호의 특징파라미터는 약 20ms 구간내에는 음성신호가 정상(stationary)이라는 가정아래 20ms의 프레임 단위로 구하여지는데 본 연구에서는 leakage 효과를 줄일 수 있도록 20ms 크기의 Hamming window를 사용하며 프레임 양 끝단의 신호정보를 보상하기 위하여 10ms씩 중첩을 시켜서 이동시킨다. 이렇게 Hamming window의 사용으로 프레임 단위로 분할된 데이터에서 14차의 자기상관함수(autocorrelation)를 구한 다음 Durbin의 순환식을 이용하여 14차인 선형예측계수를 구한다. 마지막으로 LPC 켈스트럼계수는 선형예측계수로 부터 순환식을 이용하여 14차항까지 구해지며 이러한 전체적인 특징 추출과정이 그림 2에 나타나 있다.

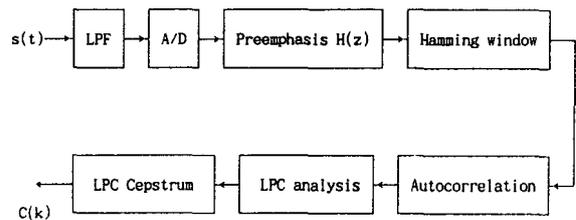


그림 2. 특징 파라미터 추출 과정
Fig 2. The procedure of feature parameter extraction

이산 HMM을 위한 VQ 코드북(codebook) 생성은 패턴의 집단화에 많이 이용되고 있는 MKM 알고리즘을 이용하였으며^[6], 음성의 두 프레임간의 유사도를 측정하기 위하여 프레임내에 있는 특징 벡터들 사이의 거리를 계산하여야 하는데, 본 연구에서는 다음 식과 같이 0차항과 13, 14차항들의 켈스트럼 계수는 포함하지 않는 유클리드 거리 측정법을 이용하였다.

$$d_{cep} = \sum_{i=1}^{12} (C_i^r - C_i^t)^2 \quad (13)$$

여기서 C_i^r 과 C_i^t 는 비교될 프레임의 LPC 켈스트럼 계수이며 0차항과 13, 14차항의 LPC 켈스트럼 계수를 거리 계산에 포함하지 않는 이유는 입력신호의 세기에 의하여 영향을 받기 쉽고 가변성이 크기 때문이다.

위와 같은 방법을 이용해서 남성화자 10명과 여성화자 10명이 11단어를 4회씩 발성한 총 880단어(20명*11단어*4회)를 가지고 128개 코드워드를 갖는 코드북을 만든 후에 14차의 LPC 켈스트럼 계수를 128 codeword 중의 하나인 인덱스로 이산화 시켰다.

본 연구에서 HMM 학습에 사용한 모델의 기본구조는

5개 상태(state)로 구성된 Left-right serial 모델로 그림 3과 같다.

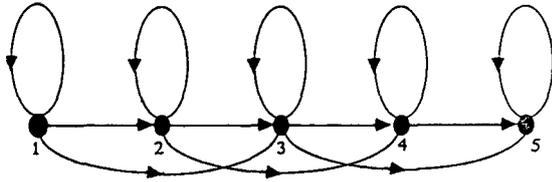


그림 3 단어 HMM의 기본구조
Fig 3. Basic structure of the word HMM

HMM을 기본으로 하는 집단화 방법을 통한 복수개의 모델구성은, 먼저 각 집단내에서 추정된 모델에 대한 관측열들의 유사도를 구하고 경계치를 사용하여 유사도가 낮은 outlier들을 분리하여서 새로운 집단으로 만들었다. 여기서 사용된 경계치는 매 분리때마다 고정되지 않고 각 학습용 관측열들의 유사도에 대한 산술평균을 이용하였다. 이렇게 경계치로서 각 집단내 유사도의 산술평균치를 이용한 결과 새로 분리되는 집단들은 모집단의 관측열들의 절반정도를 각각 할당받게 되었고 모델구성에서의 수렴속도가 빨라졌다. 위와 같이 구성된 새로운 집단들에 대한 모델 추정은 분리되기 전인 모집단의 모델 파라미터를 초기값으로 하고 집단내의 관측열들을 사용하여 복수개의 HMM을 학습시키게 된다.

HMM 집단화 방법을 통한 복수개의 통계적 모델의 성능을 비교, 평가하기 위하여 전체 학습 데이터를 먼저 남녀와 여성화자의 집단으로 분리한 후에 남녀 집단별로 HMM을 학습시켜서 2개의 모델을 구성하였고 또 다른 방법으로서, 기존의 집단화 방법인 패턴사이의 거리 유사도를 기본으로 하는 MKM 집단화 알고리즘을 이용하여 학습용 관측열들을 2개 이상으로 집단화한 후에 각 집

단내의 관측열로부터 HMM을 학습시키는 방법으로 복수개의 모델을 구성하였다. 인식실험은 불특정화자로서 학습에 참가하지 않은 나머지 8명의 화자(즉, 남자 4명과 여자 4명)가 발음한 음성을 인식대상 단어로 선정하였으며, 인식시 사용한 알고리즘은 Viterbi 알고리즘^[8]이다.

4.2 인식 결과

표 2는 전체 20명의 화자가 11단어를 4회씩 발성한 학습용 음성 데이터를 가지고 각 단어별로 여러종류의 집단화 방법에 의하여 복수개의 모델들을 학습시킨 후 학습에 참가하지 않은 8명의 불특정화자에 대한 인식실험을 수행한 결과이다. 여기서 기존의 단일 HMM 모델과 비교된 집단화 방법들은 먼저 HMM을 기본으로한 집단화 및 학습을 통하여 3개의 모델까지 구성하였으며 다음 방법은 남녀로 각각 집단화한 후 각 집단으로부터 2개의 모델들을 학습시켰고 마지막 집단화 방법은 패턴사이의 거리 유사도를 기본으로한 MKM알고리즘을 이용하여 학습데이터를 3개까지 집단화하였고 이렇게 집단화된 관측열들을 이용하여 복수개의 HMM들을 구성하였다.

인식 결과에서 알 수 있듯이 복수개의 모델 사용은 전체적으로 인식률을 향상시켰으며 그 중에서 남녀로 집단화 할때는 0.57%, MKM알고리즘을 이용하여 구성된 2개와 3개의 모델일때는 0.65%와 0.52%, 그리고 HMM 집단화 방법을 이용하여 구성된 모델중에서 2개의 모델일때 2.08%, 3개의 모델일때 0.72%의 인식률이 향상되었다. 그러나 집단을 4개 이상으로 분리하였을때는 인식률이 향상되지 않았는데 그 원인은 학습데이터의 양이 충분하지 않은 상황에서 모델갯수의 증가로 인하여 어떤 집단들은 극소수의 관측열들로만 이루어지게 되어 정확히 모델을 학습시킬 수 없게 되고 만약 이런 극소수의 관측열들이 여러 화자들과는 발음특성이 아주 특이한 화자로부터 발생된 것이라면 이런 관측열들로부터 학습된 모

표 2 3종류의 집단화 방법에 대한 인식실험 결과
Table 2. Recognition results for 3 different clusering methods

집 단 화 방 법	모 델 수	시 험 화 자								인 식 륜 (%)
		M1	M2	M3	M4	F1	F2	F3	F4	
HMM 단일모델	1	73.74	87.50	96.24	85.86	85.86	94.32	84.85	86.67	86.88
HMM을 기본으로한 집단화 방법	2	78.79	89.77	98.49	86.67	89.90	94.32	81.82	91.92	88.96
	3	78.79	85.23	95.96	84.85	89.90	94.32	88.89	82.83	87.60
남 녀 집단화 방법	2	83.84	89.77	92.93	90.91	77.78	97.73	88.89	77.78	87.45
MKM 알고리즘을 이용한 집단화 방법	2	78.79	94.32	97.65	86.67	87.88	94.32	78.79	81.82	87.53
	3	78.79	87.50	94.95	85.86	83.84	96.59	81.82	89.90	87.40

델은 다른 단어의 모델과 근접할 수도 있게 되어 전체적인 인식률에 나쁜 영향을 미치게 된다. 결국, 최적 모델의 갯수는 학습데이터의 양, 학습에 참가한 화자들의 변이성, 인식대상의 단어 종류등에 의하여 의존하게 된다. 본 실험에서는 세계의 집단화 방법들중에서 HMM을 기본으로 하여 구성된 2개의 모델이 가장 높게 인식률을 향상시킬 수 있었다.

V. 결 론

본 연구에서는 HMM을 기본으로 한 집단화 방법을 이용하여 복수개의 통계적 모델을 만드는 방법을 연구하였으며 이 연구된 방법은 불특정 고립단어(한국어 숫자음) 인식에 응용되고 기존의 방법인 패턴들의 거리 유사도에 의존한 집단화 방법과 비교되었다.

제안된 HMM 집단화 방법은 현 상태의 모델에 대한 집단내의 관측열들의 확률적 유사도를 측정하며 현 모델에 대하여 적용도가 낮은 관측열들의 집단인 outlier들로 새로운 모델을 학습시키는 방법으로서 학습용 관측열들의 분리 및 복수개의 모델구성을 효율적으로 가능케 하였다.

실험 결과, 복수개의 모델사용은 전체적으로 인식률을 향상시켰으며 HMM 집단화를 통하여 구성된 복수개의 모델들이 기존의 방법인 패턴사이의 유사도를 기본으로 구성된 복수개의 모델 및 남녀별로 구성된 복수개의 모델보다 인식률의 향상 정도가 높았다.

<본 연구는 한국전자통신연구소 연구비와 음성 데이터 베이스 지원에 의한 것입니다.>

참 고 문 헌

1. L. R. Rabiner, S. E. Levinson, A. E. Rosenberg and J. G. Wilpon, "Speaker-Independent Recognition of Isolated Words Using Clustering Techniques," *IEEE Trans. Acoust., Speech, and Signal Processing*, Vol. ASSP-27, No.4, pp. 336-349, Aug. 1979.
2. L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech recognition," *Proceedings of IEEE*, Vol. 77, No. 2, pp. 257-286, Feb. 1989.
3. 은종관, "대어휘 연속음성 인식을 위한 음소인식 기술 개발," 과학기술처 최종보고서, May, 1989.
4. 김석동, 이행세, "신경망을 이용한 우리말 음성의 인식에 대한 연구," *한국음향학회지*, Vol. 11, No. 3, pp.14-24, June 1992.
5. Hiroaki Sakoe and Seibi Chiba, "Dynamic Programming Algorithm Optimization for Spoken Words Recognition," *IEEE Trans. Acoust., Speech, and Signal Processing*, Vol. ASSP-26, No.1, pp. 43-49, Feb. 1978.
6. J. G. Wilpon, L. R. Rabiner, "A modified K-means clustering algorithm for use in isolated word recognition," *IEEE Trans. Acoust., Speech, and Signal Processing*, Vol. ASSP-33, No. 3, June 1985.
7. L. R. Rabiner, C. H. Lee, B. H. Juang and G. Wilpon, "HMM Clustering for Connected Word Recognition," *IEEE Int. Conf.*, pp. 405-408, 1989.
8. G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268-278, Mar. 1973.
9. L. R. Rabiber and M. R. Sambur, "An Algorithm for Determining the Endpoints of Isolated Utterances," *Bell Syst. Tech. J.*, Vol. 54, No. 2, pp. 297-315, Feb. 1975.

▲임 현



1970년 9월 17일생

1992년 : 목포대학교 전자공학과(공학사)

1994년 : 목포대학교 대학원 전자공학과(공학석사)

1995년~현재 : 군북무중

※주관심분야 : 음성인식, 영상처리

▲박 순 영

1957년 1월 25일생

1982년 : 연세대학교 전자공학과(공학사)

1984년 : 연세대학교 대학원 전자공학과(공학석사)

1986년 : State Univ. of New York at Buffalo, 전기 및 컴퓨터공학과(공학석사)

1989년 : State Univ. of New York at Buffalo, 전기 및 컴퓨터공학과(Ph. D.)

1989년~1990년 : State Univ. of New York at Buffalo, 전기 및 컴퓨터공학과 연구원

1990년~현재 : 목포대학교 전자공학과 조교수

※주관심분야 : 음성인식, 음성 및 진동신호처리, 영상 및 비디오신호처리

▲방 만 원

1951년 8월 25일생

1974년 : 명지대학교 전자공학과(공학사)

1977년 : 명지대학교 대학원 전자공학과(공학석사)

1980년 : 명지대학교 대학원 전자공학과(공학박사)

1982년~1983년 : 일본동경대학교 전기공학과 객원연구원

1979년~1988년 : 관동대학 전자공학과 부교수

1988년~현재 : 목포대학교 전자공학과 교수

※주관심분야 : 음성인식, 디지털신호처리, 통화품질