화자적응과 군집화를 이용한 화자식별 시스템의 성능 및 속도 향상*

김세현(KAIST), 오영환(KAIST)

<차 례>

<Abstract>

Adaptation and Clustering Method for Speaker Identification with Small Training Data

Se-Hyun Kim, Yung-Hwan Oh

One key factor that hinders the widespread deployment of speaker identification technologies is the requirement of long enrollment utterances to guarantee low error rate during identification. To gain user acceptance of speaker identification technologies, adaptation algorithms that can enroll speakers with short utterances are highly essential. To this end, this paper applies MLLR speaker adaptation for speaker enrollment and compares its performance against other speaker modeling techniques: GMMs and HMM. Also, to speed up the computational procedure of identification, we apply speaker clustering method which uses principal component analysis (PCA) and weighted Euclidean distance as distance measurement. Experimental results show that MLLR adapted modeling method is most effective for short enrollment utterances and that the GMMs performs better when long utterances are available.

* Keywords: Speaker identification, Speaker clustering, MLLR adaptation.

* 본 연구는 과학기술부의 지원을 받아 2006년도 국가지정연구실을 통해 수행되었음

알려져 있다[2]. GMM에서는 각 화자의 음성파라미터의 분포를 가우시안 밀도의 가중 합으로 다음과 같이 표현된다.

$$P(\boldsymbol{x}|\lambda_i) = \sum_{k=1}^{N} w_k p_k(\boldsymbol{x}) \tag{2}$$

위 식에서 w_k는 k번째 밀도 성분의 가중치 값을,

$$p_k(\boldsymbol{x}) = \frac{1}{(2\pi)^{D/2}|\Sigma_k|^{1/2}} exp\left\{\frac{1}{2}(\boldsymbol{x}-\overline{\mu_k})^t \Sigma_k^{-1}(\boldsymbol{x}-\overline{\mu_k})\right\} \tag{3}$$

을 나타낸다. (D는 벡터 $\boldsymbol{x}$의 차수)

위 식에서 각각의 화자모델 $\lambda_i = (w_i, \overline{\mu}_i, \Sigma_i)$는 각 가우시안 분포의 가중치, 평균, 분산으로 표현하게 된다. 표현의 정확성을 위해 화자별로 여러 개의 가우시안을 사용하게 되는데, 이에 따라 정확한 모델링을 위해 학습과정에서 예측해야 되는 파라미터의 수도 같이 증가하게 된다. 파라미터의 값들을 신뢰할 수 있는 수준으로 예측하기 위해서는 각 화자별로 충분한 양의 학습 자료가 필요하게 된다. 그러나 임의의 화자가 발성한 자료를 이용할 수 있는 음성인식과는 달리, 화자 식별에서는 충분한 양의 학습 자료를 확보하는 것은 항상 가능한 것은 아니며, 상황에 따라 학습 자료의 양이 제한될 수도 있다. 따라서 제한된 환경에 적합한 화자모델링 방법이 필요하게 된다. 학습 자료의 양이 적은 경우의 접근방법은 크게 2가지로 나누어 볼 수 있다.

① 학습 자료를 이용해 예측해야 되는 모델의 파라미터의 수를 줄인다.
② 화자독립 모델을 이용하여 각 화자의 특성을 포함하고 있는 유사화자모델을 생성한다.

첫 번째 방법인 파라미터의 수를 줄이는 경우, 정해져 있는 특정 문장이나 단어를 발성하게 하여 화자식별을 행하는 문장 종속형 화자식별 시스템에서는 어느 정도 이상의 시스템 성능을 기대할 수 있으나, 화자의 자유발성을 이용해 화자식별을 행해야하는 문장 독립형 화자식별 시스템의 경우에는 그 성능이 현저히 떨어지게 된다.

따라서 본 논문에서는 화자독립 모델에 화자적응 방법을 적용하여 유사 화자모델을 생성하는 두 번째 방법을 사용하여 학습 자료의 수가 적은 화자식별에 효과적인 화자 모델링을 적용한다. 이 방법은 정확한 화자 모델링 방법 대신 기존의 화자독립 모델을 이용해 유사화자 모델을 생성하는 방법으로 학습 자료의 양이 많아지면 실제 화자모델과 차이가 발생하지만, 충분치 않은 학습 자료로 화자모델을 생성해야 하는 경우에는 효과적으로 적용할 수 있다.

3. MLLR 화자적응 기법을 적용한 화자식별

화자들 사이의 음성 변이는 화자들 간의 서로 다른 특성을 나타내는 것으로 동일한 발성에 대해서도 서로 다른 특징을 보인다. 일반적으로 음성인식에서 특정 화자에 특화되어 학습된 화자종속 모델의 경우, 훈련된 화자에 대해서는 더 좋은 인식 성능을 보이게 된다. 그러나 화자종속 모델을 훈련시키기 위해서는 목적 화자의 발성으로 구성된 많은 양의 학습 자료가 필요하게 된다. 이러한 문제점을 극복하기 위해서 많이 사용되는 방법이 화자적응 기법이다. 화자적응은 화자독립 모델로부터 목적화자의 적은 양의 적응 자료를 이용하여 유사화자종속 모델을 만들어, 목적화자에 대한 인식 시스템의 성능을 향상시키는 방법이다. 음성인식에서 사용하는 적응기법에는 MAP(Maximum A Posteriori)[3]에 기반하는 방법, MLLR과 같은 파라미터 변환을 이용하는 방법 그리고 마지막으로 화자군집화에 기반한 방법의 총 3가지 군으로 나누어 볼 수 있다[4].

3.1. MLLR 화자 적응 기법

본 논문에서는 화자모델의 생성을 위한 충분한 자료가 없는 환경을 가정하고 있다. 따라서 화자모델의 생성을 위해 기존의 GMM, HMM 등의 기법들을 적용할 수 없으므로, 음성인식에서 사용하는 화자적응 기법을 이용하여 유사화자 모델을 생성한 후, 이를 화자식별에 이용하려 한다. 위의 3가지 대표적인 화자적응 기법 중 적은 학습자료 환경의 화자 모델링 방법에 적합한 화자적응 기법을 선택하도록 한다. 먼저 MAP에 기반하는 방법의 경우, 화자적응을 위해 필요한 적응 자료의 양이 상대적으로 많아 다른 적응방법에 비해 효과적으로 유사화자 모델을 생성할 수 없다. 3가지 방법 중 상대적으로 가장 적은 양의 적응 자료가 필요한 화자군집화에 기반한 방법의 경우, 화자 적응에는 적은 양의 자료가 필요하지만 학습단계에서 각 화자별로 화자 모델이 생성되어 있어야 한다는 문제가 있다. 즉 적응 자료는 적게 필요하지만, 학습 단계에서 각 화자별로 모델을 생성할 수 있는 충분한 자료가 필요하게 되어 이 역시 학습 자료의 양이 제한된 환경에 적합하지 않다. 따라서 본 논문에서는 학습단계에서 화자 모델을 생성할 필요가 없고, 적은 양의 적응자료만으로 유사화자 모델을 생성할 수 있는 MLLR 화자적응 기법을 이용하여 화자 식별에 필요한 화자모델을 생성하는 식별 방법을 이용한다. 3가지 화자 적응 방법군에 대한 비교 실험결과는 4장에서 보이도록 한다.

MLLR 화자적응은 주어진 모델과 적응자료 사이의 불일치를 줄여줄 수 있는 변환들의 집합을 찾아내는 것이다. 이때, 불일치는 화자간의 불일치나 환경의 불일치 등이 될 수 있다. 좀 더 구체적으로 살펴보면, GMM, HMM의 경우 가우시안 혼합(Gaussian Mixture)들의 평균과 분산 파라미터를 적응 자료의 화자나 환경에

맞도록 선형변환 (linear transformation)하는 기법이다.

이를 식으로 표현하면 다음과 같다.

$$\hat{\mu} = A\mu + b \tag{4}$$

여기서 μ는 적응되기 전 원시 모델의 파라미터이며, 이를 변환행렬 A와 편향 항 b를 이용한 선형변환을 통해 적응 모델 파라미터를 얻게 된다. 변환행렬은 주어진 적응 자료의 양에 따라 전역 적응변환행렬 1개만을 사용하거나, 유사한 특성을 가지는 가우시안들로 나누어 회귀 클래스(regression class)를 생성해 각 클래스 별로 서로 다른 변환행렬을 사용할 수도 있다. 변환행렬은 주어진 적응자료와 EM 알고리듬을 이용해 구할 수 있다[4].

3.2. MLLR 기법을 이용한 화자모델링

본 논문에서 대상으로 하는 환경은 화자식별을 위해 각 화자모델을 생성할 수 있는 충분한 양의 학습 자료가 없으므로, 화자적응 기법을 이용하게 된다. 화자적응을 위해 우선 주어진 모든 화자들의 학습 자료들을 이용해 하나의 화자독립 모델 λ_{SI}를 생성한다. 화자독립 모델을 이용해, 각 화자별로 유사화자 종속모델을 얻는 과정은 다음과 같다. S개의 화자 모델 $\hat{\lambda}_i (i = 1, \cdots, S)$는 모든 화자의 학습 자료를 이용해 생성된 화자독립모델인 λ_{SI}에 i번째 화자의 학습 음성을 화자적응 자료로 이용하고 MLLR 적응기법을 적용하여 얻는다. 이 과정을 통해 화자 i의 음성으로만 학습된 화자종속모델 λ_i대신 이와 유사한 특성을 지니게 되는 유사화자 모델 $\hat{\lambda}_i$을 얻게 된다. 학습 자료의 양이 적은 환경에서는 예측 파라미터의 수가 많은 λ_i의 경우 충분히 모델링되지 못하나, 같은 수의 모델 파라미터를 갖는 $\hat{\lambda}_i$는 직접적인 모델링 대신 λ_{SI}에 적응 기법을 이용해 얻게 되므로, 모델링 파라미터의 예측 대신 변환행렬의 파라미터만을 예측해 화자모델을 얻을 수 있게 되어 상대적으로 적은 학습 자료만으로도 강인한 모델 생성이 가능하다.

식별 단계에서는 식 (5)와 같이 각 화자의 화자모델 대신 λ_{SI}에 각 화자의 음성과 MLLR 적응기법을 이용해 구한 유사화자모델 $\hat{\lambda}_i$를 대상으로 유사도를 측정한 후, 확률값이 가장 큰 모델을 결과 값으로 출력해 주게 된다.

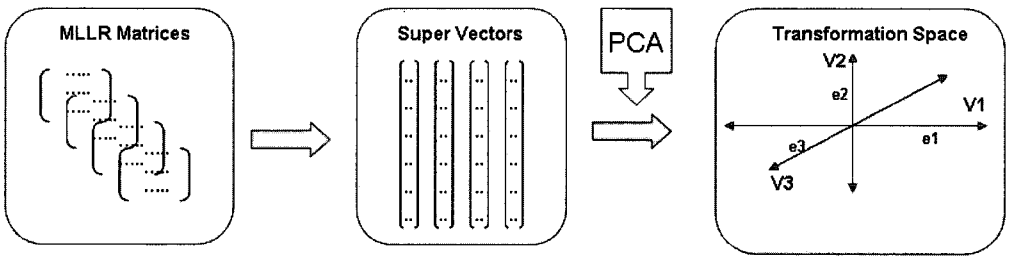
$$\hat{S} = \arg\max_{1 \le k \le S} P(\boldsymbol{X}|\hat{\lambda}_k) \tag{5}$$

3.3. 식별속도 향상을 위한 화자군집화

화자식별을 수행하기 위해서는 S명의 등록된 화자 모델과 입력으로 주어진 관측 벡터열의 유사도를 비교하게 된다. 각 모델과의 유사도 값을 계산한 후, 가장 큰 값을 갖는 모델을 결과로 출력하게 되는데, 이때 비교대상이 되는 화자의 수 S의 값이 커질수록 계산량이 급격이 늘어나게 된다. 실제 화자 식별 시스템의 식별 단계에서 가장 많은 시간이 소요되는 부분이, 각각의 모델과 유사도를 측정하는 과정으로, 본 논문에서는 계산량 감소를 위해 비슷한 특성을 가지는 화자별로 나누어주는 군집화 방법을 이용해 비교 횟수를 줄인다. 모든 등록화자 모델과 비교하는 대신, 선택된 집단내의 화자들과만 비교를 해 결과를 출력하게 되므로, 속도의 개선은 있으나 성능저하를 피할 수 없게 된다. 따라서 기준 시스템에 비해 성능 저하를 최소화하면서 속도를 개선할 수 있는 방법이 필요하다.

3.3.1. 거리척도

화자 군집화를 하기위해서는 먼저 모델들 간의 거리를 측정할 수 있는 거리척도를 정의하여야 한다. 일반적으로 GMM이나 HMM과 같은 모델 기반 방법에서는 우도값에 기반한 거리 척도를 이용한다. 그러나 앞 절에서 제안한 MLLR 적응기법을 적용시켜, 화자모델을 생성하는 방법으로 구한 유사화자 모델을 이용하는 경우에는 생성된 GMM 모델을 이용할 수 있으나 이는 순수한 화자 모델이 아닌 화자 독립 모델로부터 생성된 유사화자 모델이므로, 본 논문에서는 화자의 특성을 나타내고 있는 MLLR 행렬만을 이용하여 군집화하는 방법을 이용한다. 따라서 우도 이외의 행렬간의 유사도를 측정할 수 있는 척도가 필요하다. 본 논문에서 이용한 방법은 다음과 같다.



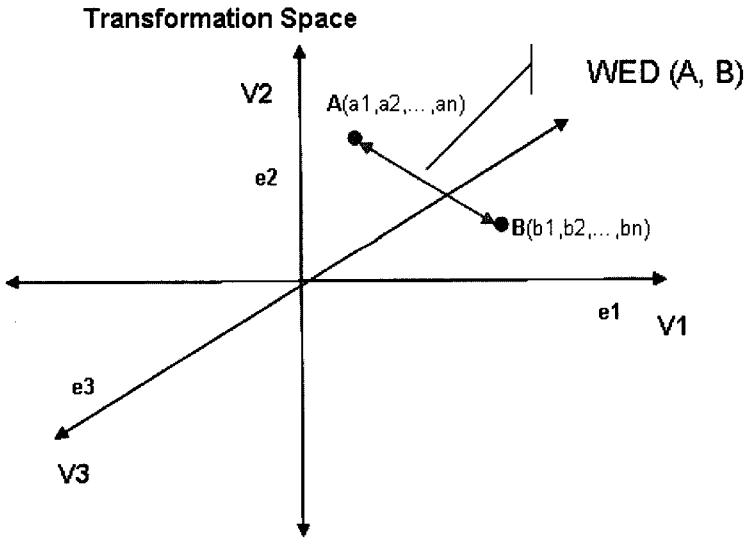
<그림 2> MLLR행렬을 이용한 변환 공간 구성도

먼저 MLLR 행렬로 표현된 모든 화자모델들을 슈퍼벡터 형태로 바꾼 후 PCA (principal component analysis) 방법을 이용해 저차원의 변환공간 (Transformation space)을 생성한다[5]. 변환공간은 n개의 고유벡터($v_1, v_2, \cdots, v_n$)와 고유치

$(e_1, e_2, \cdots, e_n)$으로 표현된다.

위와 같은 과정을 통해 구성된 변환 공간상에서 2개의 점 사이의 거리를 구하는 과정은 다음과 같다. 두 화자 모델간의 거리측정을 위해서는 두 화자의 특성을 나타내는 MLLR 행렬을 슈퍼벡터 형태로 바꾼 후, 위에서 구한 저차원의 변환공간으로 <그림 3>과 같이 사영시킨다. 변환 공간에서 두 점 A, B 사이의 가중 유클리드 거리(WED, Weighted Euclidean Distance, 각 차원에 대응하는 고유치를 가중치로 이용)는 다음 식과 같이 계산 할 수 있다.

$$WED(A,B) = \sqrt{e_1(a_1 - b_1)^2 + e_2(a_2 - b_2)^2 + \cdots + e_n(a_n - b_n)^2} \quad (6)$$



<그림 3> 가중 유클리드 거리를 이용한 A, B의 거리 측정

3.3.2. 화자 군집화

위에서 설명한 거리 척도를 이용해 식별 대상 화자와 비교할 집단을 선택하는 방법에는 2가지 접근 방법이 있다. 먼저 학습 단계에서 K-means 알고리듬을 사용해 각각의 화자별로 군집화를 하고, 각 집단별로 통합모델을 생성하는 방법이다. 앞 절에서 설명한 방법으로 생성된 변환 공간상에서 K-means 알고리듬과 가중 유클리드 거리, WED를 이용해 화자 군집화를 수행하는 알고리듬을 정리하면 다음과 같다.

<단계 1>
S명의 화자를 표현하는 MLLR 행렬을 이용해 각 화자별로 한 개씩 S개의 슈퍼벡터를 생성한다.
<단계 2>
S개의 슈퍼벡터를 변환 공간으로 사영시켜 S개의 점 $P_1, P_2, \cdots, P_s$를 얻는다.
<단계 3>
변환 공간상의 S개의 점을 다음 과정을 통해 K개의 클러스터로 나누어 준다.
1. K개의 클러스터를 초기화 : $C_1, C_2, \cdots, C_K$
2. $P_1, P_2, \cdots, P_s$ 각각에 대해서
각 클러스터의 중심값과 WED를 이용한 거리 비교를 통해 가장 가까운 클러스터에 할당해준다.
For(j=1,...,S)
$\arg\min_i WED(C_i, P_j)$
3. 각 클러스터의 중심값을 클러스터에 할당된 점들의 평균값을 이용해 갱신해준다.
4. 전체 왜곡값이 임계값보다 크면 2.로 돌아가 반복. 임계값보다 작으면 군집화를 종료한다.

식별 단계에서는, 요구 화자의 음성이 입력되면 이를 이용해 MLLR 행렬을 구한다. MLLR 행렬을 슈퍼벡터 형태로 바꾼 후 앞 절에서 구한 변환 공간으로 사영시켜 T=$(t_1, t_2, \cdots, t_n)$ 얻는다. 이 값을 각 클러스터의 중심값 C_i와 비교해 비교할 집단을 선택하게 된다. 선택된 집단 내에서 점수를 계산하는 것은 기존 식별방법과 동일하다.

두 번째 방법은 학습 단계에서 군집화를 행하지 않고 식별 요청이 들어오면 이를 이용해 I명의 비교대상 화자를 선택하는 방법이다.

그 과정은 다음과 같다. 식별 요청 화자의 음성이 들어오면 K-means 알고리듬 방법과 동일하게 변환 공간으로 사영시켜 T을 구하게 된다. 점 T와 S명의 화자가 사영된 점들 $P_1, P_2, \cdots, P_s$사이의 가중 유클리드 거리를 이용해 식별 요청 화자와 거리가 가까운 I명의 화자를 선택한다. 두 번째 방법은 학습단계에서 비교대상 집단의 생성이 이루어지는 첫 번째 방법과 달리 인식단계에서 I개의 모델을 선택해 비교대상 집단을 생성하게 되므로 추가적인 시간이 필요하다. 그러나 입력 화자에 맞추어 동적으로 비교대상 집단을 생성할 수 있는 장점이 있으며, S명의 화자 모델을 슈퍼벡터로 바꾸고 변환공간에 사영시키는 작업은 오프라인으로 가능하고, 온라인상에서 필요한 작업은 식별 요청 화자의 MLLR 행렬을 구해 이를 변환공간으로 사영시킨 후 거리를 구하는 과정만 추가된다. I명의 화자를 포함하는 집단이 선택되어지면, 실제 인식 단계에서는 기존 화자식별 단계의 점수 계산과 동일하게 I개의 화자 모델과 입력 화자의 유사도 측정을 해 결과를 출력한다.

4. 실험 및 결과

4.1. 실험 환경

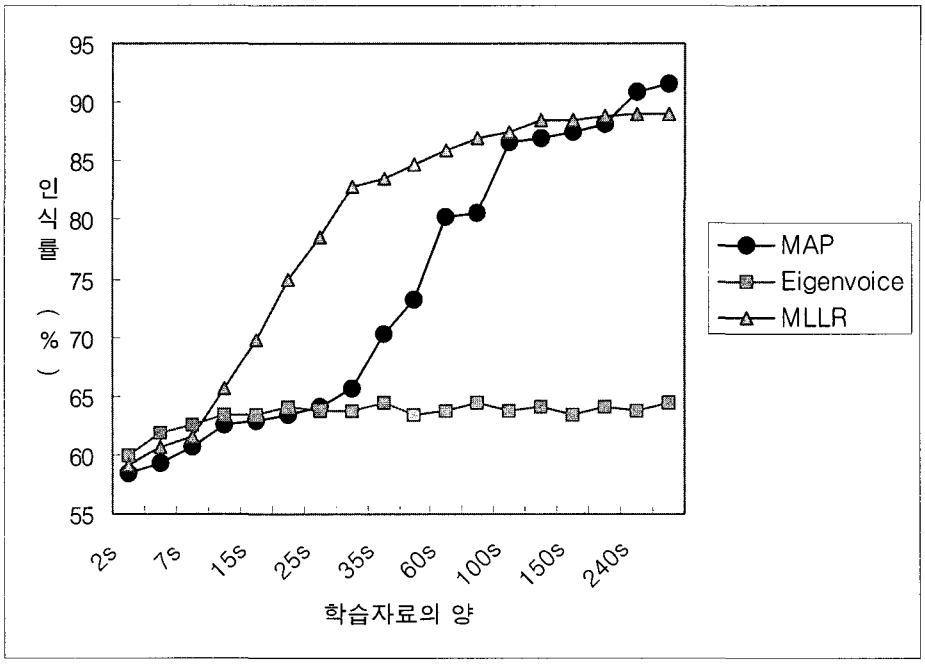
제안한 방법의 유효성을 검증하기 위해 문장 독립형 화자식별 시스템에 대해 성능을 비교하였다. 남성 20명, 여성 20명으로 구성된 총 40명의 화자를 대상으로 실험하였으며, 학습에는 화자별로 30초~300초 정도의 자유발화 음성을 이용하였다. 특징벡터는 에너지와 12차 멜-켑스트럼(MFCC) 특징벡터들과 그 미분값과 2차 미분값을 사용하였고, 인식기는 HTK(HMM Tool Kit)를 이용해 구현하였다[6].

GMM의 경우, 학습 자료의 양에 따라 16~128개 가우시안들을 사용하였고, HMM은 6개의 상태를 가지는 에르고딕 HMM으로 구현하고 각 상태별로 4~16개의 가우시안을 사용하여 학습 모델을 얻었다. MLLR 행렬의 수도 학습 자료의 양에 따라 1~8개까지 나누어 학습하였다.

식별실험에는 등록화자 40명이 발성한 2~3초 정도의 음성을 화자별로 8번씩 실험하였다. 제안한 방법의 유효성을 검증하기 위해 문장 독립형 화자식별 시스템에 대해 성능을 비교하였다.

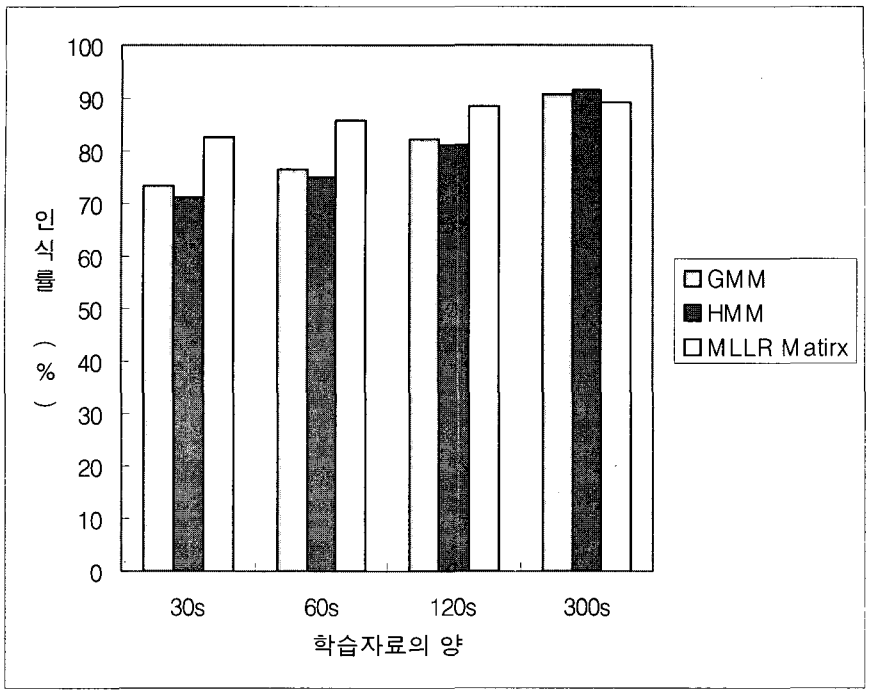
4.2. 실험 결과

화자적응 방법들 중 제한된 학습 자료만으로 학습을 할 때 적합한 화자적응 방법을 찾기 위해 3가지 화자적응 군에 대한 실험을 수행하였다. 다음의 <그림 4>는 3가지 화자 적응군을 대표하는 MAP, MLLR, Eigenvoice[7] 방법으로 화자모델을 구성한 후, 화자식별 실험을 수행한 결과이다. 학습 자료의 양이 7초 이하인 경우 효과적이었으나 학습 자료의 양이 증가하면서 성능이 급격히 수렴하는 결과를 보였다. 주어진 3분 미만의 학습자료 환경에서는 MLLR 적응기법을 이용한 화자표현 방법이 가장 효과적이었으며, 학습자료의 양이 4분 이상 증가하면 MAP 방법이 더 효과적이었다. 그러나 학습자료의 양이 4분 이상인 경우에는 화자적응 기법을 이용한 유사화자모델 생성보다는 GMM을 이용한 모델링 방법이 더 좋은 성능을 보임을 다음 실험을 통해 알 수 있다.



<그림 4> 화자적응 방법에 대한 화자식별 비교실험 결과

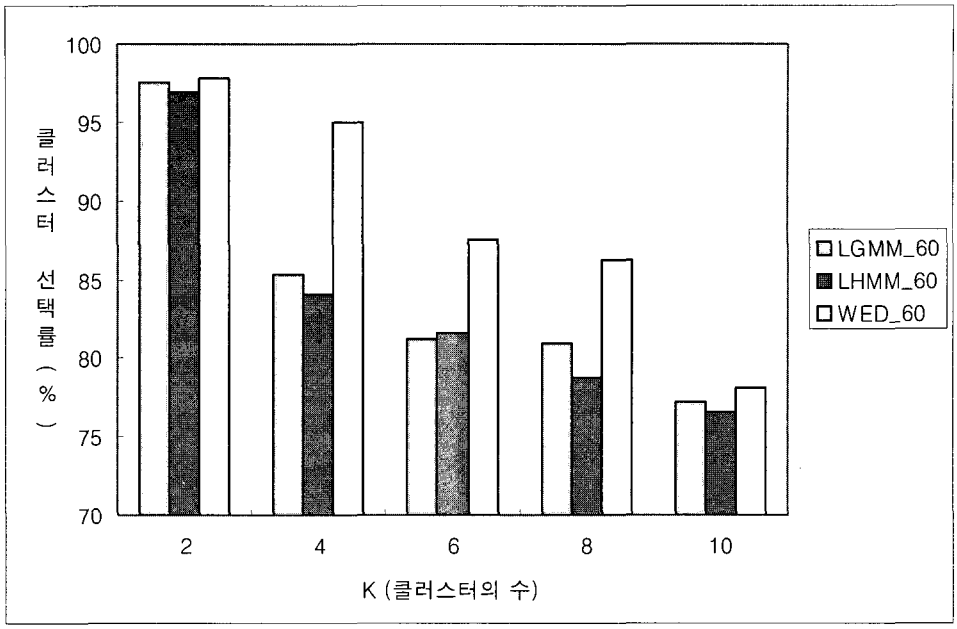
<그림 5>는 학습 자료의 양을 변화시킴에 따라 GMM, HMM, MLLR 행렬 모델링 방법의 식별 성능을 보인 것이다. <그림 5>에서 보는 바와 같이 학습 자료의 양이 적은 경우에는 제안한 MLLR 적응기법에 기반한 방법의 성능이 더 우수함을 알 수 있다. 이는 학습자료 양의 부족으로 인해 GMM, HMM의 경우 통계적으로 충분히 모델링 되지 않아서 나타나는 결과로, 학습 자료의 양이 늘어남에 따라 인식성능의 차이의 폭이 줄어들고 충분한(5분 이상) 학습 자료가 제공되는 경우 GMM, HMM의 성능이 MLLR 적응 방법보다 더 좋은 결과를 보이게 됨을 볼 수 있다.



<그림 5> 화자표현 방법에 대한 화자식별 비교실험결과

다음 <그림 6>과 <표 1>은 화자모델링 방법과 각 거리 척도를 이용한 군집화에 대한 비교 실험결과이다. 아래 그래프와 표에서 "LGMM"은 GMM 화자 모델링 방법과 우도(Likelihood)를 거리척도로 이용한 방법이며, "LHMM"은 HMM 모델링 방법과 거리척도는 우도를 이용한 경우, 마지막으로 "WED"는 MLLR 행렬을 화자 모델링방법으로 사용하고 가중 유클리드 거리를 거리척도로 이용한 실험을 나타낸다. 각각의 방법으로 비교대상 집단을 구성한 후, 입력화자의 음성에 의해 선택된 집단에 정답 화자가 속해있는지 여부에 대한 정확도를 나타낸다.

<그림 6>은 학습 자료의 양이 60초인 경우에 대해서, 비교대상 집단의 수 즉 클러스터의 수를 2개~10개까지 조정하면서 군집화를 수행한 결과를 보인다. 클러스터의 수가 2에서 10까지 변함에 따라서 모든 경우에 인식률의 저하가 있지만 가중 유클리드 거리를 사용한 방법이 인식률의 저하가 상대적으로 낮음을 볼 수 있다. 군집화를 통해 나누어진 집단들 중 1개의 집단만을 선택해 선택된 집단 내에서만 식별 결과를 찾게 되므로 실제 인식률은 클러스터 선택률보다 높아질 수 없다. 따라서 클러스터 선택률을 어느 정도 이상의 성능을 보장해 줄 수 있는 K값을 선택해야 한다.



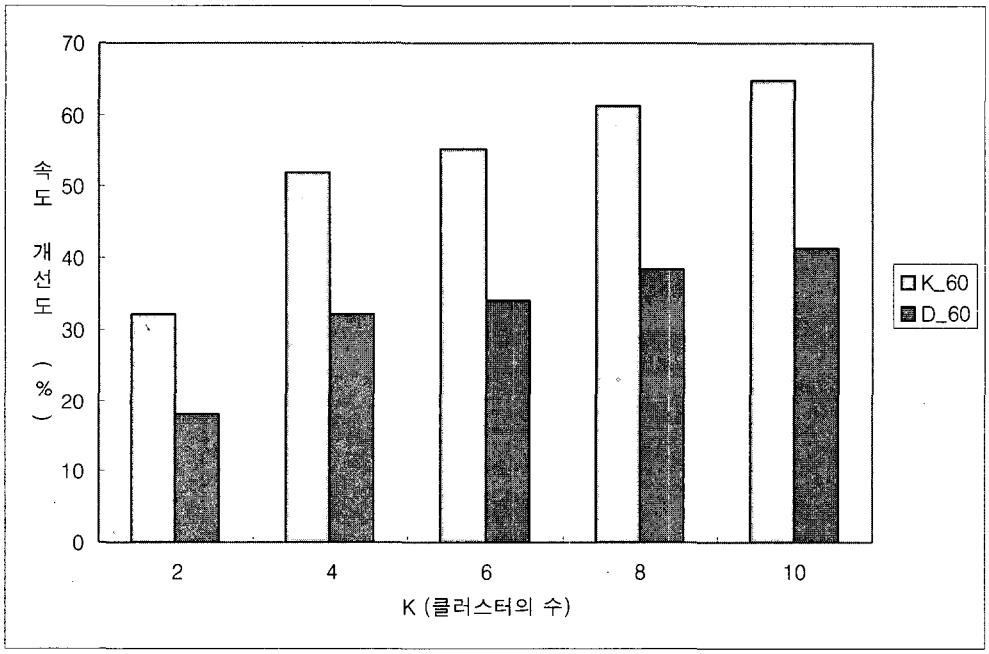
<그림 6> 클러스터 수에 따른 집단 인식률 변화 (학습자료 60초인 경우)

<표 1>의 결과는 비교대상 집단(cluster) 수 K를 4개로 두었을 때의 올바른 클러스터를 선택한 정도를 보이고 있다. 실험결과를 보면 역시 학습 자료의 양이 적은 경우, GMM이나 HMM과 우도를 척도로 이용한 방법들이 화자를 충분히 모델링하지 못하고, 변별력도 떨어짐을 볼 수 있다. 반면, MLLR 행렬과 가중 유클리드 거리척도를 이용한 방법은 적은 학습자료 환경에서 좋은 성능을 보임을 알 수 있다. 클러스터 선택률이 인식 성능에 직접적인 영향을 미치게 되므로, 90% 이하의 클러스터 선택률을 보이는 GMM이나 HMM 모델링과 우도를 사용한 방법은 시스템의 성능에 근접한 성능을 보이며 속도를 개선하려는 방법으로는 적당하지 않다. 가중 유클리드 거리를 이용한 방법인 WED의 경우에는 학습 자료의 양에 상관없이 94%이상의 클러스터 선택률을 보이고 있다.

<표 1> 비교대상 집단 선택을 위한 거리척도 비교 실험 결과 (K=4)

학습 자료의 양	LGMM	LHMM	WED
30초	84.36%	83.44%	94.06%
60초	85.31%	84.06%	95.00%
120초	91.56%	90.63%	95.94%
300초	99.4%	99.07%	98.44%

다음은 학습단계에서 비교대상 집단을 생성하는 방법(K-means)과 인식단계에서 동적으로 비교대상 집단을 생성하는 방법(Dynamic)에 대한 비교실험 결과이다. 계산 속도와 비교대상 집단의 수가 밀접한 관계가 있으므로, 비교대상 집단의 수인 클러스터의 수를 2개에서 10개까지 변화시켜 실험하였다. <그림 7>은 60초의 학습 자료를 등록 자료로 이용한 경우 클러스터의 수 K의 변화에 대한 계산속도 개선 정도를 보이고 있다. 인식단계에서의 속도개선을 측정하게 되므로 학습단계에서 군집화를 하는 K-means 알고리듬을 사용하는 방법(K_60)이 인식단계에서 비교대상 집단을 선택하는 방법(D_60)에 비해 큰 속도 개선을 얻을 수 있다.



<그림 7> 군집수에 따른 식별 속도의 향상정도 (학습자료 60초인 경우)

<그림 8>은 위 실험과 동일한 학습 자료가 60초인 경우의 클러스터 수 K에 따른 인식률을 보이고 있다. 동적 군집화 방법이 속도 개선은 낮지만 인식률의 저하 역시 상대적으로 낮음을 알 수 있다. 속도 개선뿐 아니라 인식률의 저하도 고려해야 하므로 인식률과 속도 개선 정도를 고려하여 적당한 클러스터의 수와 군집화 방법을 선택하여야 한다.

[3] J. L. Gauvain and C. H. Lee, "Maximum a-posteriori estimation for multivariate Gaussian mixture observations of Markov chains", *IEEE Trans. on Speech and Audio Processing*, Vol. 2, No.2 pp. 291-298, 1994.

[4] P. C. Woodland, "Speaker adaptation: techniques and challenges", *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)* 1999, Vol. 1, pp.85--90, 1999.

[5] C. Huang, T. Chen and E. Chang, "Speaker selection training for large vocabulary continuous speech recognition", *Proc. ICASSP 2002*, Vol. 1, pp. 609-612, 2002.

[6] S. Young, "*The HTK Book (for HTK version 3.1)*", Cambridge University Engineering Department 2001.

[7] R. Kuhn, J. C. Junqua, P. Nguyen et al., "Rapid speaker adaptation in eigenvoice space", *IEEE Trans. on Speech and Audio Processing*, Vol. 8, No. 6, pp. 695-707, 2000.

접수일자: 2006년 5월 15일
게재결정: 2006년 6월 19일

▶ 김세현(Se-hyun Kim) : 교신저자
주소: 305-701 대전시 유성구 구성동 373-1 한국과학기술원 전자전산학과 전산학전공
소속: 한국과학기술원(KAIST) 전자전산학과 전산학전공 음성인터페이스 연구실
전화: 042) 869-3556
E-mail: shkim@speech.kaist.ac.kr

▶ 오영환(Yung-hwan Oh)
주소: 305-701 대전시 유성구 구성동 373-1 한국과학기술원 전자전산학과 전산학전공
소속: 한국과학기술원(KAIST) 전자전산학과 전산학전공 음성인터페이스 연구실
전화: 042) 869-3516
E-mail: yhoh@cs.kaist.ac.kr