GMM 기반 실시간 문맥독립화자식별시스템의 성능향상을 위한 프레임선택 및 가중치를 이용한 Hybrid 방법

김민정[†] · 석수영[††] · 김광수[†††] · 정호열[††††] · 정현열[†††††]

요 약

본 논문에서는 GMM(Gaussian Mixture Model)에 기반한 실시간문맥독립화자식별시스템[1][2]의 성능향상을 위하여 프레임선택(Frame Selection)방법과 프레임가중치(Weighting Model Rank)방법을 혼합한 hybrid방법을 제안한다. 본 시스템에서는 GMM의 파라미터를 최적화하기 위하여 MLE(Maximum likelihood estimation)방법과 인식 알고리즘으로 ML(*Maximum Likelihood*)을 기본적으로 사용하였다. 제안한 *hybrid* 방법은 두 단계로 이루어진다. 첫째, 화자모델과 테스트 데이터를 이용하여 프레임단위로 유사도를 계산하고, 가장 큰 유사도 값과 두 번째로 큰 유사도 값의 차를 계산한 후, 차가 문턱치보다 큰 프레임만을 선택한다. 두 번째로, 선택되어진 프레임에서 계산되어진 유사도 값 대신에 가중치 값을 사용하여 전체 스코어를 계산한다. 특징 파라미터로서는 켑스트럼과 회귀계수를 사용하였으며, 학습과 테스트를 위한 데이터베이스는 채집기간이 다른 여러 데이터베이스들로 구성되어 있으며, 실험을 위한 데이터는 임의의 단어를 선택하여 사용하였다. 화자인식실험은 기본 시스템에 프레임선택방법, 프레임가중치방법, 제안한 Hybrid방법을 각각 적용하여 실험하였다. 실험결과, 프레임선택방법에 비해 평균 4%, 프레임가중치방법에 비해 평균 1%의 인식률 향상을 보여, 본 논문에서 적용한 hybrid방법의 유효성을 확인하였다.

Hybrid Method using Frame Selection and Weighting Model Rank to improve Performance of Real-time Text-Independent Speaker Recognition System based on GMM

M.J.Kim[†], S.Y.Suk[††], K.S.Kim[†††], H.Y.Jung[††††] and H.Y.Chung[†††††]

ABSTRACT

In this paper, we propose a hybrid method which is mixed with frame selection and weighting model rank method, based on GMM(gaussian mixture model), for real-time text-independent speaker recognition system. In the system, maximum likelihood estimation was used for GMM parameter optimization, and maximum likelihood was used for recognition basically. Proposed hybrid method has two steps. First, likelihood score was calculated with speaker models and test data at frame level, and the difference is calculated between the biggest likelihood value and second. And then, the frame is selected if the difference is bigger than threshold. The second, instead of calculated likelihood, weighting value is used for calculating total score at each selected frame. Cepstrum coefficient and regressive coefficient were used as feature parameters, and the database for test and training consists of several data which are collected at different time, and data for experience are selected randomly. In experiments, we applied each method to baseline system, and tested. In speaker recognition experiments, proposed hybrid method has an average of 4% higher recognition accuracy than frame selection method and 1% higher than WMR method, implying the effectiveness of it.

Key words: Speaker Recognition, Frame selection, GMM, Weighting, Hybrid

접수일 : 2002년 4월 25일, 완료일 : 2002년 6월 27일

[†] 영남대학교 일반 대학원 정보통신공학과(박사수료)

[††] 영남대학교 일반 대학원 정보통신공학과(박사과정)

[†††] 정회원, 경운대학교 컴퓨터전자정보공학부 전임강사

[††††] 정회원, 영남대학교 전자정보공학부 전임강사

[†††††] 영남대학교 전자정보공학부 교수

1. Introduction

화자식별이란 여러 명의 등록화자 중 발성화자를 식별하는 것을 말한다. 이러한 화자식별 기술은 개인의 음성 특징이 유일하다는 사실을 근거로 하고 있으며 최근의 인터넷 기술의 발전과 더불어 보안을 위한 인증방법으로 각광을 받고 있다. 화자식별시스템은 발성의 종류에 따라 문맥종속 및 문맥독립화자식별로 나눌 수 있는데, 문맥독립화자식별의 경우 보안성이 높아 이에 관해 많은 연구가 진행 중이다. 문맥독립화자식별방법으로서는 장시간(Long-term)통계에 기반한 방법[3], VQ(Vector - quantization)에 기반한 방법[4], HMM(Hidden Markov Model)과 GMM (Gaussian mixture model)에 기반한 방법[5] 등이 연구되고 있으며 이러한 접근방법들 중 화자특성변화의 표현에 있어서나 화자인식률 면에서 좋은 결과를 나타내고 있는 GMM에 의한 접근방법이 가장 유리한 것으로 알려져 있다[6]. 이에 본 연구에서는 GMM을 이용하여 베이스라인시스템을 구현하였다.

또한, 본 시스템에서는 유사도 계산시에 각 프레임에 가중치를 부여하는 WMR(Weighting model rank)방법과 프레임단위로 유사도를 계산하여 유효한 프레임만을 선택하는 프레임선택방법을 각각 적용하여 인식률 향상을 추구하였으며, 이러한 방법들보다 향상된 인식률을 얻기 위하여 두 가지 방법을 혼합한 hybrid방법을 적용하였다. 인식실험결과, 베이스라인에 각각의 방법을 적용한 경우보다, hybrid 방법을 혼합하여 적용한 경우가 가장 높은 인식률을 나타냄으로서 제안한 방법의 유효성을 확인할 수 있었다.

2장에서는 Gaussian mixture model에 대해서 살펴보고, 3장에서 본 시스템에 적용한 화자인식방법을 설명하고, 4장에서 프레임선택방법과 WMR방법에 대해서 언급한 후, 이를 혼합한 hybrid방법에 대해서 설명한다. 5장에서는 인식실험 및 결과에 대해서 검토한 후 6장에서 결론을 맺도록 한다.

2. Gaussian mixture model

GMM(Gaussian mixture model)은 출력확률밀도함수가 가우시안 밀도 혼합(Gaussian density mixture)인 1개의 상태만으로 구성된 CHMM(Continuous HMM)의 한 형태이다.

화자인식에 GMM을 사용하는 이유로 두 가지를 들 수 있다. 첫째로, GMM은 음향학적 클래스(Acoustic class)의 집합을 모델링 할 수 있다는 것이다. 화자의 목소리에 대응되는 음향 공간은 모음이나 비음, 파찰음과 같은 음소를 표현하는 음향학적 클래스의 집합으로 표현될 수 있는데, 이러한 음향학적 클래스는 화자를 구별하는데 이용되는 화자의 성도에 대한 정보를 가지고 있다[7]. i번째 음향학적 클래스의 스펙트럼 형태는 i번째 component 밀도의 평균 μ_i으로 표현되고, 평균 스펙트럼형태의 변화는 공분산행렬 $\sum_i$로 표현된다. 모든 학습 및 테스트 음성은 레이블되지 않기 때문에, 음향학적 클래스는 hidden으로 볼 수 있다. 독립특징벡터를 가정하면, 이러한 hidden 음향학적 클래스로부터 추출된 특징벡터의 관측밀도가 Gaussian mixture이다.

두 번째로, Gaussian basis 함수의 선형조합은 샘플분포(Sample distribution)의 클래스를 표현할 수 있다는 것이다[8]. GMM의 성질 중 하나가 임의의 형태를 가지는 밀도를 부드러운 형태로 근사시키는 것이다. unimodal 가우시안 화자모델은 평균벡터(Mean vector)와 공분산(Covariance)으로 화자의 특징분포를 표현하고, VQ- distortion 모델은 특징벡터의 이산집합으로 화자분포를 표현한다. 이와 같은 점을 고려하여 구성된 GMM은 가우시안 함수의 이산집합을 사용하고, 각각의 평균과 공분산을 가지게 함으로써 이들 두 모델의 특징을 혼합한 형태이다[9].

가우시안 혼합 밀도는 M component 밀도의 가중합계이며, 다음의 식에 의해 얻어진다[5].

$$p(x|\lambda) = \sum_{i=1}^{M} c_i N(x, \mu_i, \sum_i) \tag{1}$$

여기서, x는 a-차원랜덤백터이며, $b_i(x), i=1, \ldots, M$은 component 밀도이고, $c_i, i=1, \ldots, M$은 mixture weight이다.

각 component 밀도는 평균 μ_i과 공분산 $\sum_i$을 가지는 a-variate Gaussian 함수이다.

$$N(x, \mu_i, \sum_i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\sum_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x-\mu_i)' \sum_i^{-1}(x-\mu_i)\right\} \tag{2}$$

여기에서, mixture weight는

$$\sum_{i=1}^{M} c_i = 1 \tag{3}$$

로 제한한다.

Gaussian mixture 밀도는 모든 component밀도의 mixture weight와 공분산행렬, 평균벡터로 구성된다.

$$\lambda = \{c_i, \mu_i, \sum_i\} \quad i = 1, \ldots, M \tag{4}$$

화자모델학습은 주어진 학습음성으로부터 학습 특징벡터의 분포와 가장 잘 맞는 GMM, λ 파라미터를 추정하는 것이다. GMM의 파라미터를 추정하는 방법에는 여러 가지가 있으나, 가장 잘 알려진 방법으로는 MLE(maximum likelihood estimation)가 있다. MLE는 주어진 학습데이터에서 GMM의 유사도를 최대화하는 모델 파라미터를 찾는데 사용된다.

T 학습벡터 $X = x_1, x_2, \ldots, x_T$의 열에서, GMM 유사도는 다음과 같고,

$$P(X|\lambda) = \prod_{t=1}^{T} p(x_t|\lambda) \tag{5}$$

이를 로그영역에서 표현하면 다음과 같다.

$$L(X|\lambda) = \sum_{t=1}^{T} \log p(x_t|\lambda) \tag{6}$$

3. Baseline 시스템에서의 화자식별방법

3.1 일반적인 방법

일반적인 방법에서의 화자식별은 Bayes의 정리에 따라, N명의 화자 중 사후확률 $P(\lambda_i|X)$, $1 \le i \le N$를 최대화하는 모델 λ_{i^*}의 화자 i^*를 찾는 것이다.

$$P(\lambda_i|X) = \frac{p(X|\lambda_i)P(\lambda_i)}{p(X)} \tag{7}$$

여기에서, 사전정보가 없기 때문에, 사전확률 $P(\lambda_i)$는 다음과 같이 표현할 수 있다.

$$P(\lambda_i) = \frac{1}{N}, \quad 1 \le i \le N \tag{8}$$

$\max_i p(X|\lambda_i)$로 사후확률은 최대가 되고, 식별화자는 다음으로 결정된다.

$$i^* = \arg \max_i p(X|\lambda_i) \tag{9}$$

이러한 일반적인 화자식별 방법을 그림으로 나타내면 그림 1과 같다. 입력된 음성은 전처리단을 거치면서 벡터열 X로 변환되고, 각 화자모델들과의 유사도가 계산되어지고, 계산된 유사도중 가장 높은 유사도를 가지는 화자가 인식화자로 결정된다.

3.2 프레임단위 유사도 방법

화자검증시스템에서는 유사도 정규화기법을 적용함으로서 시스템의 성능을 향상시킬 수 있었다[10][11][12]. 화자검증의 일반적인 방법은 요구된 화자 모델 λ_c를 이용하여 입력발성 $X = x_1, x_2, \ldots, x_T$에 유사도 비(Likelihood Ratio) 테스트를 하는 것이다[8].

$$l(X) = \frac{p(\lambda_c|X)}{p(\lambda_{\bar{c}}|X)} \tag{10}$$

여기에 Bayes의 정리를 적용하고, 사전확률이 동일하다고 가정한다면, 로그 영역에서의 유사도 비는 다음으로 표현할 수 있다.

$$\Lambda(X) = \log P(X|\lambda_c) - \log P(X|\lambda_{\bar{c}}) \tag{11}$$

여기에서, $\lambda_{\bar{c}}$는 모든 다른 가능한 화자를 나타낸다.

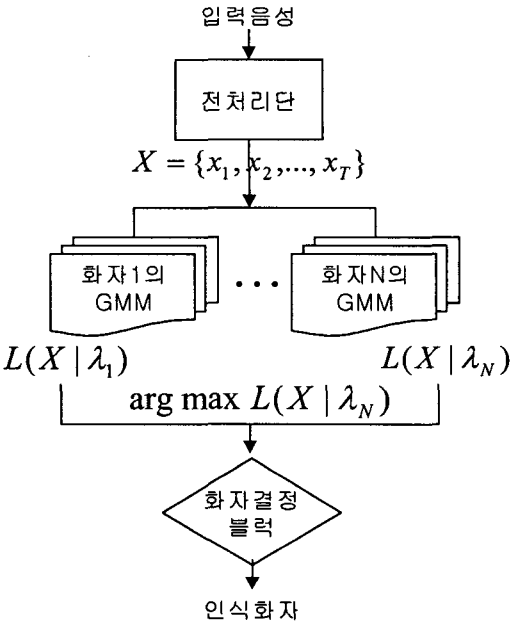

그림 1. 일반적인 화자식별 방법

유사도 $P(X|\lambda_c)$는 식(6)로부터,

$$\log P(X|\lambda_c) = \frac{1}{T}\sum_{t=1}^{T}\log p(x_t|\lambda_c) \quad (12)$$

로 계산할 수 있다.

유사도 $P(X|\lambda_{\bar{c}})$는 백그라운드 화자들의 모델을 사용하여 계산되어지며, B개의 백그라운드 화자모델을 $\{\lambda_1,\ldots,\lambda_B\}$라고 하면, 백그라운드 화자들의 로그 유사도는 다음과 같이 계산할 수 있다.

$$\log P(X|\lambda_{\bar{c}}) = \log\left\{\frac{1}{B}\sum_{b=1}^{B}P(X|\lambda_b)\right\} \quad (13)$$

백그라운드 모델에 의한 유사도 정규화는 발성문장의 변화에 따른 변화를 최소화 할 수 있기 때문에 시스템의 성능을 향상시킬 수 있다[10].

기존의 일반적인 화자식별시스템에서는 식별화자를 결정하는데 단일 발성으로부터 유사도를 계산하기 때문에 정규화 과정이 필요 없지만[10], 본 시스템에서는 프레임단위 유사도를 사용하므로 정규화 과정이 필요하며, 이러한 유사도 정규화는 다음과 같이 적용시킬 수 있다.

$$p_{norm}(x_t|\lambda_i) = \frac{p(x_t|\lambda_i)}{\frac{1}{B}\sum_{b=1}^{B}p(x_t|\lambda_b)} \quad (14)$$

모든 벡터 x_t, $t=1,2,\ldots,T$에서, 계산되어진 유사도의 전체 합계를 구하면 각 화자 모델 i에 대한 새로운 스코어가 계산되고,

$$Sc_i(X|\lambda_i) = \frac{1}{T}\sum_{t=1}^{T}\log p_{norm}(x_t|\lambda_i) \quad (15)$$

인식화자는 가장 높은 스코어 $Sc_i(X|\lambda_i)$를 가지는 화자로 결정된다.

그림2는 일반적인 화자식별방법에 프레임단위 유사도를 적용한 방법을 그림으로 나타낸 것이다. 입력된 음성은 전처리단을 거치면서 벡터열 X로 변환되고 모든 화자모델들과의 유사도 $p(x_t|\lambda_i)$, $i=1,2,\ldots,N$가 각각 계산되어진 후, 계산된 유사도는 다음 단인 유사도 및 스코어계산에서 $t=1,2,\ldots,T$에 대한 각각의 합계가 이루어지며 결과로서 $Sc(X|\lambda_i)$를 출력한다. 인식화자는 가장 높은 스코어 $Sc(X|\lambda_i)$를 가지는 화자로 결정된다.

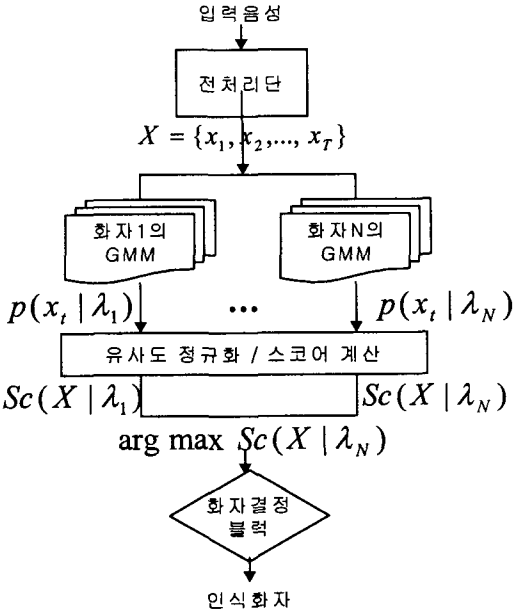

그림 2. 프레임단위 화자식별방법

4. 인식성능향상을 위해 적용한 방법들

4.1 유사도 차를 이용한 프레임 선택방법

유사도 차를 이용한 프레임선택방법은 현재 입력되고 있는 테스트음성이 인식되고자 하는 화자의 정보를 많이 포함하고 있다고 가정한다면, 테스트음성과 인식되고자하는 화자모델과의 유사도는 다른 화자모델들과의 유사도보다는 훨씬 클 것이며, 반대로 인식되고자하는 화자의 특성을 많이 포함하고 있지 않다면, 다른 화자들과의 유사도와는 크게 차이가 나지 않을 것이다. 그러므로, 다른 화자들과의 유사도 차가 큰 프레임만을 선택한다면, 인식되고자 하는 화자의 특성이 많이 포함되어 있는 프레임들을 선택할 수 있을 것이다. 유사도 차를 이용한 프레임선택방법은 각 프레임에서 가장 큰 유사도를 가지는 화자와 두 번째로 큰 유사도를 가지는 화자의 유사도 값이 일정 값 이상의 차이가 나는 프레임은 해당프레임에서의 각 화자들의 유사도를 전체 스코어의 계산에 사용하고, 일정 값 이하의 값을 나타내는 프레임은 계산에서 제외시킴으로서, 결과적으로, 전체 스코어를 계산할 때 화자의 정보를 많이 포함하고 있는, 즉, 변별력이 큰 프레임만을 사용하는 것이다. 이때 문턱치는 실험을 통해 얻은 값으로서, 최소 문턱치 값에

서부터 일정 값만큼 증가시켜가며 가장 높은 인식률이 나오는 값으로 설정하였다.

유사도 차를 이용한 프레임선택방법은 다음의 과정으로 이루어진다. 식(14)으로부터 얻어진 프레임 유사도중 가장 큰 값과 두 번째로 큰 값을 가지는 벡터를 다음과 같이 찾고,

$$w^t_{m1} = \arg\max\left(p_{norm}(x_t|\lambda_1), p_{norm}(x_t|\lambda_2), \ldots, p_{norm}(x_t|\lambda_N)\right) \quad (16)$$

$$w^t_{m2} = \arg\max\left(p_{norm}(x_t|\lambda_1), p_{norm}(x_t|\lambda_2), \ldots, p_{norm}(x_t|\lambda_{m1-1}), p_{norm}(x_t|\lambda_{m1+1}), \ldots, p_{norm}(x_t|\lambda_N)\right) \quad (17)$$

다음의 식으로 가장 큰 값을 가지는 프레임과 두 번째로 큰 값을 가지는 프레임과의 유사도 차를 구할 수 있다.

$$W^t = w^t_{m1} - w^t_{m2} \quad (18)$$

위의 식으로부터 얻어진 값은 화자모델들의 유사도 변이를 나타내기 때문에, 문턱치를 설정하여 문턱치보다 큰 값을 나타내는 프레임은 식(15)의 스코어 계산시에 사용되고, 문턱치보다 작은 값을 가지는 프레임은 계산에서 제외된다.

4.2 Weighting Model Rank(WMR) Method

WMR방법은 인식화자를 결정하는 스코어를 계산할 때 테스트음성과 화자모델들과의 유사도를 사용하지 않고, 각 프레임에서 계산된 유사도들의 상대적 위치에 따라 가중치를 부여하고, 이러한 가중치를 인식화자를 결정하는 스코어를 계산하는데 사용하는 것이다[13]. 이렇게 함으로서, 프레임단위에서 높은 유사도 값을 가지는 화자모델은 더 높은 값을, 낮은 유사도 값을 가지는 화자모델은 더 낮은 값을 부여하여 화자들 간의 변별력을 더 높일 수가 있다.

WMR방법은 첫 번째로, 식 (14)를 이용하여 프레임단위 유사도를 계산하고, 이를 내림순으로 정렬한다. 즉, 가장 큰 프레임 유사도를 가지는 화자모델은 최상위에 위치하고, 가장 낮은 프레임 유사도를 가지는 화자모델은 최하위에 위치하게 된다. 표1은 각 프레임에서의 화자모델의 순위와 가중치의 관계를 나타낸 것이다.

표 1. 화자모델의 N-best 유사도 list

Rank r	유사도	Weight $w(r)$	Model
1	p^t_i	$w(1)$	Model λ_i (max. likelihood)
2	p^t_j	$w(2)$	Model λ_j
...	...	...	...
m	p^t_k	$w(m)$	model λ_k
...	...	...	...
N	p^t_p	$w(N)$	Model λ_p (min. likelihood)

다음으로, 각 순위에 해당하는 모델의 프레임 유사도를 가중치 값으로 대치한다. 이때, 가중치는 지수함수를 사용함으로서 선형함수보다 화자들 사이에 변별력을 더 주었다[13].

$$w(r_\lambda) = \exp(A - Br_\lambda), r_\lambda = 1, \ldots, N \quad (19)$$

여기에서, A와 B는 $w(1) \approx N$이 되도록 설정하였다. 그림3은 가중치와 순위의 관계를 나타낸 것이다. 두 번째로, 각 모델 λ_i의 순위에 따라 프레임유사도 $p(x_t|\lambda_i)$ 대신에 부여된 가중치 $w_t(r_{\lambda i})$를 사용하여 전체 스코어 $Sc(X|\lambda_i)$를 계산한다. 전체 스코어 $Sc(X|\lambda_i)$는 $t=1, \ldots, T$에서 모든 가중치를 더함으로서 얻을 수 있다.

$$\log Sc(X|\lambda_i) = \sum_{t=1}^{T} w_t(r_{\lambda_i}) \quad (20)$$

여기에서, $w_t(r_{\lambda i})$는 시간 t에서 순위 $r_{\lambda i}$의 모델 i의 가중치이다.

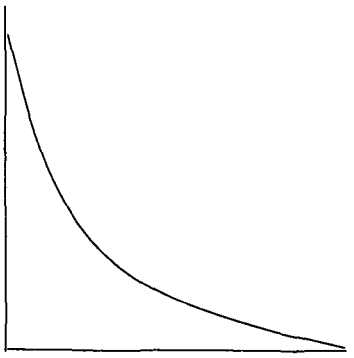

그림 3. 가중치와 유사도 순위

4.3 Hybrid Method

Hybrid 방법은 앞에서 살펴본 두 방법을 혼합하는 것이다. 이러한 Hybrid 방법은 프레임들 중에서 화자정보를 많이 포함하고 있는 프레임을 선택할 수 있다는 것과, 선택된 프레임에서 화자들의 유사도 값 대신에 가중치 값을 사용함으로서 화자들 간의 변별력을 더 높일 수 있다는 장점이 있다.

Hybrid방법은 먼저, 식(14)으로부터 얻어지는 프레임단위 유사도에서 가장 높은 유사도 값을 가지는 화자의 유사도 값과 두 번째로 높은 유사도 값을 가지는 화자의 유사도 값의 차이를 식(18)을 이용하여 결정한 후, 이 값이 문턱치를 넘어서는 프레임만을 선택하게 되는데, 이때 문턱치는 최소 문턱치 값에서부터 일정 값만큼 증가시켜가며 가장 높은 인식률이 나오는 값으로 설정하였다. 다음으로, 선택된 프레임에서 화자들의 유사도 값을 표1과 같이 정렬한 후 가중치를 유사도 값 대신에 이용하여 인식화자를 결정하는 전체스코어계산에 사용하는 것이다. 이때, 가중치를 결정하는 식(19)에서 A, B 값은 등록화자 35명을 기준으로 하였을 경우, $A=3.7, B=0.15$ 일 때 가장 높은 인식률을 나타내었으나, 이 값들은 등록화자의 수가 달라질 경우에는 매번 새로이 계산되어야 한다. 따라서, 본 논문에서는 등록화자가 달라졌을 경우 A와 B값을 화자 수에 비례하게 계산하여 가중치를 결정하였다.

5. 인식실험 및 결과

인식실험에서 GMM의 Mixture의 수는 계산량을 고려하여 16으로 고정하였으며, 특징파라미터는 켑스트럼 계수와 회귀계수 값만을 사용하였으며, 전처리단에서 사용된 분석조건은 표2와 같다.

표 2. 전처리 분석조건

Sampling Rate	16 kHz
Pre-emphasis coefficient	0.98
Hamming Windows	yes
Frame length	256 points
Frame Shift	120 points
Cepstrum vector dimension	10

실험을 위한 데이터베이스로는 10대에서 50대까지의 남녀화자가 혼합된 데이터베이스들로서 채집시간과, 발성목록이 다른 데이터베이스들을 사용하였다. 화자모델생성을 위한 데이터베이스는 무작위로 추출된 단어로서, 추출된 단어는 4000프레임에서는 약 80단어, 10000프레임에서는 약 200단어정도가 사용되었으며, 테스트를 위한 단어 역시 무작위로 추출된 단어로서 약 3개의 단어가 사용되었다. 테스트 화자는 102명과 316명의 화자를 대상으로 실험하였으며, 인식실험은 베이스라인시스템, 베이스라인시스템에 프레임선택방법을 적용한 경우, 베이스라인시스템에 WMR방법을 적용한 경우, 그리고 베이스라인시스템에 두 가지 방법 모두 사용하는 Hybrid방법을 적용한 경우에 대해서 인식실험을 하였다.

인식실험결과, 베이스라인에 프레임선택방법만을 적용한 경우와 WMR방법만을 적용한 경우 모두, 인식률 향상이 있었지만, 베이스라인에 Hybrid방법을 적용한 경우가 가장 높은 인식률을 나타냄으로서 제안한 방법의 유효성을 확인할 수 있었다.

표3과 그림4, 그림 5는 화자수와 프레임 수에 따라 각각의 방법을 적용하여 인식 실험한 결과를 비교한 것이다. 표3에서 프레임선택방법과 hybrid방법에서 괄호 안의 수는 문턱치를 의미한다.

표 3. 테스트화자수와 및 훈련 프레임 수에 따른 각 방법의 인식률

화자수	모델 프레임수	Baseline	Frame Selection (Threshold)	WMR	Hybrid (Threshold)
102	4000	85.29	85.29(0.5)	94.12	95.1(0.35)
	10000	97.06	98.03(1.1)	99.02	100(1.1)
316	4000	87.66	87.66(0.5)	89.97	90.82(0.17)
	10000	93.35	93.35(0.5)	97.78	97.78(0.5)

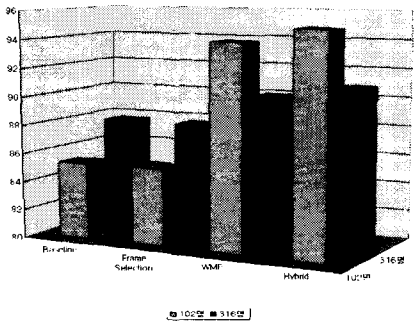

그림 4. 4000 훈련프레임에서의 인식률

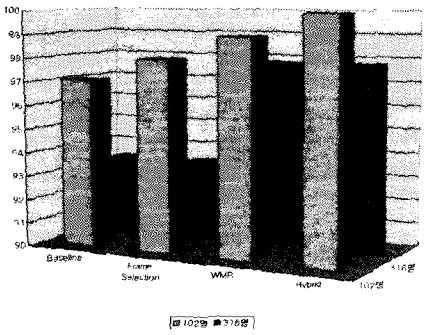

그림 5. 10000 훈련프레임에서의 인식률

6. 결 론

본 논문에서는 일반적인 화자식별방법과 프레임 선택방법, WMR방법과 제안한 hybrid방법으로 인식 실험을 수행하였으며, 인식실험결과, WMR방법과 프레임선택방법 모두 일반적인 화자인식방법보다 인식률 향상이 있었으나, 본 논문에서 제안한 Hybrid 방법을 사용한 경우, 각각의 방법보다 대상화자수 및 훈련 프레임 수에 따라 최고 10%에서 최저 1%의 인식률 향상이 있었으며, 제안한 방법이 가장 높은 인식률을 나타내어 제안한 방법의 유효성을 확인할 수 있었다. 제안한 hybrid방법은 프레임선택방법을 이용하여 화자의 정보를 많이 포함하고 있는 프레임만을 선택한 후, 선택된 프레임에 지수함수를 이용한 가중치를 줌으로서 화자들 사이의 변별력이 약한 프레임에서도 높은 변별력을 줄 수 있어, 향상된 인식률을 얻을 수 있었다. 화자식별이란, 거의 완벽한 인식률을 얻어야만 되는 점과 실시간으로 이루어 져야 한다는 조건이 있기 때문에, 향후 좀더 강건한 화자인식 알고리즘과 고속화를 연구할 계획이다.

Reterences

[1] 김민정, 석수영, 정현열, "Gaussian Mixture Model을 이용한 실시간문맥독립화자인식에 대한 고찰," 한국음향학회 하계학술발표대회 논문집 제20권, pp123-126, 2001

[2] 김민정, 석수영, 정현열, 정호열, "프레임단위유사도정규화를 이용한 문맥독립화자식별시스템의 성능 향상," 제14회 신호처리합동학술대회 논문집 vol.14, pp487-490, 2001

[3] S. Furui, F. Itakura, and S. Saito, "Talker recognition by longtime averaged speech spectrum," *Trans. IECE*, Vol. 55-A, No. 1, pp. 549-556, 1972..

[4] A. E. Rosenberg and F. K. Soong, "Evaluation of a vector quantization talker recognition system in text independent and text dependent models," *Computer Speech and Language*, Vol. 2, pp. 143-157, 1987.

[5] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. on SAP*, Vol. 3, No. 1, pp. 72-83, 1995.

[6] S. Furui, "*An overview of speaker recognition technology*," in *Acoustic speech and speaker recognition*(C.-H. Lee, F. K. Soong, and K. K. Paliwal, eds.), Ch. 2, pp. 31-56, Kluwer Acad. Pub., 1996.

[7] H. Matsumoto and H. Wakita, "Vowel normalization by frequency warped spectral matching," *Speech Communication*, Vol. 5, No. 2, pp. 239-251, 1986.

[8] K. Fukunaga, *Introduction to statistical pattern recognition*, Academic Press Inc., 1990.

[9] H. Gish and M. Schmidt, "Text-independent speaker identification," *IEEE Signal Processing Magazine*, pp. 18-32, Oct. 1994.

[10] D.A. Reynolds, "Speaker identification and verification using Gaussian mixture speaker models," *Speech Communication*, Vol. 17, No. 1-2,pp.91-108, 1995.

[11] A. Rosenberg, J. DeLong, C.Lee, B.Juang and F. Soong, "The use of cohort normalized scores for speaker verification," proc. *ICSLP*, pp.599-602, 1992.

[12] T. Matsui and S. Furui, "Likelihood normalization for speaker verification using a phoneme- and speaker-independent model," *Speech Communication*, Vol. 17, pp. 109-116, Aug. 1995.

[13] K.Markov and S.Nakagawa, "Text-Independent speaker identification on TIMIT database."

Proceedings, *Acous.Soc.Jap.*pp.83-84, March 1995.

김 민 정

1999년 영남대학교 일반대학원 멀티미디어 통신공학과 (공학석사)
1999년 현재 영남대학교 일반 대학원 정보통신공학과 (박사수료)

관심분야 : 디지털신호처리, 음성처리, 음성인식, 화자인식

석 수 영

1998년 계명대학교 물리학과 (이학사)
2000년 영남대학교 일반대학원 멀티미디어 통신공학과 (공학석사)
2000년 3월-현재 영남대학교 일반대학원 정보통신공학과 (박사과정)
관심분야 : 디지털신호처리, 문자인식, 음성인식

김 광 수

1994년 경남대학교 전자공학과 (공학사)
1998년 영남대학교 일반대학원 전자공학과(공학석사)
1998년 3월-현재 영남대학교 일반대학원 전자공학과(박사수료)
2001년 3월-현재 경운대학교 컴퓨터전자정보공학부 전임강사
관심분야 : 음성분석 및 인식, 음성 및 오디오 신호처리, 음질평가

정 호 열

1988년 2월 아주대학교 전자공학과(공학사)
1990년 2월 아주대학교 전자공학과(공학석사)
1993년 2월 아주대학교 전자공학과(박사수료)
1998년 (프)리옹국립응용과학원 (INSA de Lyon) 전자공학전공(공학박사)
1998년 4월~1998년 12월 (프)CREATIS 박사후 과정
1999년 3월~현재 : 영남대학교 전자정보공학부 전임강사
관심분야 : 음성, 영상 신호처리, 인공지능, 디지털 워터마킹

정 현 일

1975년 영남대학교 전자공학과 (공학사)
1989년 일본 동북대학교 정보공학과(공학박사)
1989년 3월-현재 영남대학교 전자정보공학부 교수
1992년 7월-1993년 7월 미국 CMU Robotics 연구소 객원연구원
1994년 12월-1995년 2월 일본 토요하시기술과학대학 외국인 연구자
2000년 6월-2000년 8월 미국 Qualcomm Inc. 수석 엔지니어
관심분야 : 음성인식, 화자인식, 음성합성 및 DSP 응용분야