

Tolerance Interval Analysis를 이용한 배경화자 없는 간단한 화자인증시스템에 관한 연구*

최홍섭(대진대)

<차 례>

- | | |
|----------------------------------|---------------------|
| 1. 서 론 | 2.4. 제안된 화자인증 알고리즘 |
| 2. 본 론 | 3. 실험 및 결과 |
| 2.1. 화자인증시스템의 구조 | 3.1. 음성DB의 구성 |
| 2.2. GMM 화자모델 | 3.2. 화자인증 실험 및 성능비교 |
| 2.3. Tolerance Interval Analysis | 4. 결론 및 고찰 |

<Abstract>

On the Simple Speaker Verification System Using Tolerance Interval Analysis Without Background Speaker Models

Hong Sub Choi

In this paper, we are focused to develop the simplified speaker verification algorithm without background speaker models, which will be adopted in the portable speaker verification system equipped in portable terminals such as mobile phone and PMP. According to the tolerance interval analysis, the population of someone's speaker model can be represented by a suitable number of selected independent samples of speaker model. So we can make the representative speaker model and threshold under the specified confidence level and coverage. Using proposed algorithm with the number of samples is 40, the experiments show that the false rejection rate is 3.0% and the false acceptance rate 4.3%, worth comparing to conventional method's results, 5.4% and 5.5%, respectively. Next step of research will be on the suitable adaptation methods to overcome speech variation problems due to aging effect and operating environments.

* Keywords : Speaker verification, Tolerance interval analysis, GMM.

1. 서 론

사람이 발성한 음성은 여러 가지 형태의 정보를 지니고 있다. 이것은 음성 신호가 전달하고자 하는 언어적 내용 외에도 화자간의 차이에 의한 음향적 특징을 포함하고 있기 때문이다. 이러한 화자간의 음향적 특징의 변이를 이용하여 발성한 사람의 ID를 확인하는 시스템을 화자인증 시스템이라 하며 신원확인 및 출입통제 시스템에 사용하고 있다[5]. 일반적인 화자인증 시스템에서는 화자 본인의 음성의 시간적 변이와 판정기준으로 사용하는 문턱값의 안정화를 위해서 조건부확률인 우도(likelihood) 값을 정규화 하여 사용한다. 즉, 화자인증시스템이 인증을 받으려는 화자의 모델에 대한 우도값만이 아니라 그 인증화자의 배경화자모델에 대한 우도 값을 구한 후, 그들의 비율인 우도비(likelihood ratio)의 값으로 화자에 대한 진위 판단의 근거로 사용하는 방법이다. 이러한 정규화방법은 배경화자에 대한 의뢰인의 모델 특성이 보다 변별력을 갖도록 하는 동시에 발성음성의 변화에 대해 영향을 적게 받아 시스템이 안정화되고 인식률도 향상되는 장점이 있다[2]. 따라서 화자확인 시스템에서는 어떠한 배경화자모델을 사용하느냐가 시스템의 인식 성능에 많은 영향을 주고 있음을 알 수 있다. 지금까지 이러한 배경화자 모델선정에 많이 이용되는 방법으로는 검증하고자 하는 화자와 가장 유사도가 높은, 또는 가장 낮은 화자모델들을 배경화자 모델로 선정하는 화자기반 코호트(cohort) 방법이 있다. 하지만 이 방법은 데이터베이스에 등록된 화자의 수와 음성데이터의 크기에 따라 시스템의 성능이 좌우되는 경향이 있다[3]. 그러나 지금까지의 화자인증 시스템은 주로 서버급 컴퓨터를 기반으로 구축되었는데, 앞으로는 휴대폰과 같은 단말기의 확대 보급과 함께 휴대용 기기에서의 화자인증 시스템의 구현의 필요성이 급증할 것이 예상된다. 그러나 이러한 휴대용 화자인증 시스템에서는 기본적으로 메모리와 CPU와 같은 하드웨어적인 제약조건으로 인해 서버급의 시스템에서 사용하던 배경화자 모델을 사용할 수가 없기 때문에 가능한 많은 수의 화자로부터 얻은 용량이 큰 음성 데이터를 사용하는 대신 배경화자 없는 간략화 된 화자인증시스템의 구현이 필요하게 되었다. 이러한 시도의 하나로 통계학에서 이용되는 허용구간분석(Tolerance Interval Analysis) 이론을 이용하여 배경화자 없는 화자인증 알고리즘이 소개되었는데[7], 이 방법은 의뢰인을 대표하는 음성 특징벡터를 허용구간분석 방법을 이용하여 구하고, 이를 인증실험 시 우도값을 계산할 때 사용하는 방법이다. 그러나 이 방법은 모델로 사용하는 대표 특징벡터에 따라, 인증 성능의 오거부울의 편차가 많은 문제점을 보이고 있다. 본 논문은 이러한 편차를 줄이기 위해 의뢰인을 대표하는 기준모델로 음성 특징벡터뿐만 아니라 화자모델 또한 허용구간분석 방법을 이용하여 구한 후, 이들로부터 계산된 우도값들을 결합하여 2단계에 걸쳐 인증에 사용하는 방법을 제안하여, 인증 성능의 편차를 줄일 수가 있었다. 이러한 배경화자 없는 인증방법은 인증시스템을 사용하는 화자이외의 다른

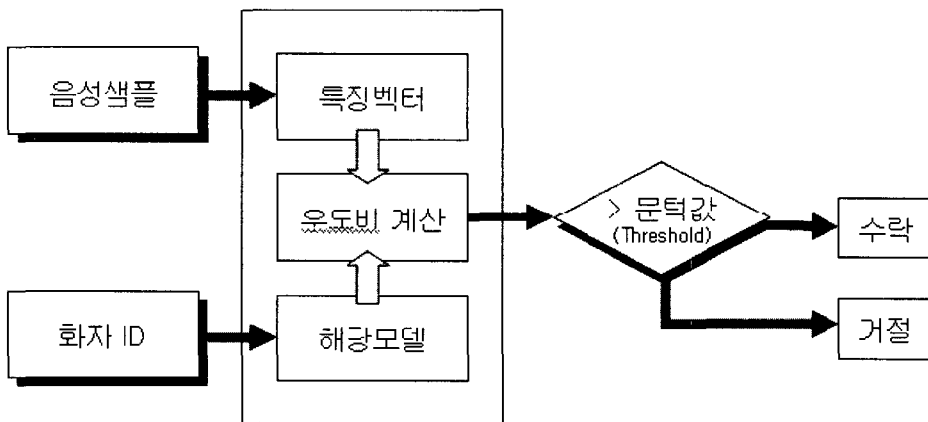
사람의 음성데이터가 필요 없고, 본인의 음성데이터만을 갖고 알고리즘을 수행하기 때문에 메모리 용량 측면에서 보다 효율적이며, 인증실험 시 우도값을 계산할 때도 적은 연산량이 소요되므로 단말기와 같은 휴대용 인증시스템에 효과적으로 적용할 수 있을 것이다.

논문의 구성은 1장 서론에 이어 2장 본론에서는 화자인증시스템의 구조와 GMM 모델방법, 그리고 허용구간분석(Tolerance Interval Analysis)에 대한 설명과 이를 이용한 제안된 화자인증 알고리즘을 설명하였으며, 3장에서는 실험을 위한 음성DB의 구성 및 음성 데이터의 형식, 그리고 실험과정과 결과 및 기존 인증시스템과의 성능 비교를 하고, 그리고 마지막으로 4장 결론 및 고찰로 구성하였다.

2. 본론

2.1. 화자인증 시스템의 개요

GMM을 이용한 화자인증 시스템의 구성은 <그림 1>과 같다.



<그림 1> 화자인증 시스템의 구성

인증을 받기 위하여 입력한 발성음으로부터 특징벡터의 추출이 완료되면 시스템은 의뢰인이 요청한 화자ID에 해당하는 모델과 그의 배경화자모델을 이용하여 입력 데이터에 대한 우도비(유사도)를 계산한 후, 이 값을 미리 지정한 문턱값과 비교하여 의뢰인의 인증요청을 수락 또는 거절을 하게 된다. 발성음과 의뢰인 모델 사이의 우도비는 식(1)과 같이 나타낼 수 있다[2].

$$L(X) = \frac{p(\lambda_c | X)}{p(\lambda_c | X)} \quad (1)$$

식 (1)에 Bayes 정리를 적용한 후 로그를 취하면 식 (2)를 얻게 된다.

$$A(X) = \log p(X | \lambda_c) - \log p(X | \lambda_c) \quad (2)$$

여기서 $p(X | \lambda_c)$ 는 의뢰인 모델에서 발생음이 생성될 확률이고, $p(X | \lambda_c)$ 는 배경화자 모델에서 발생음이 생성될 확률이다. 화자인증 시스템에서 사용되는 문턱값(threshold)은 시스템의 환경에 따라 달라지므로 절대적인 값으로 결정할 수는 없다. 일반적인 실험에서는 판정기준인 문턱값을 오인식(FA)과 오거절(FR)의 비율이 동일한 지점(EER : Equal Error Rate)의 우도비(TH)로 정하게 되는데, 따라서 화자인증 시스템은 식(2)를 계산하여 우도비 $A(X)$ 를 구한 후 그 값을 시스템의 문턱값과 비교하여 $A(X) > TH$ 이면 수락을, $A(X) < TH$ 이면 거절을 한다[3].

2.2. Gaussian Mixture Model (GMM) 화자 모델

GMM은 여러 개의 가우시안 확률밀도(Gaussian probability density) 함수들에 각각의 가중치를 준 다음, 이를 선형 결합함으로써 임의의 모양을 갖는 확률밀도 함수를 표현할 수 있다. 그리고 음성의 특징 벡터의 확률분포는 화자마다 그 모양이 다르며, 이러한 확률분포를 GMM을 이용하여 모델링하여 인식하고자 하는 화자의 모델로 사용함으로써 화자인식에 이용할 수 있다[4][5].

GMM의 혼합 확률분포는 M개의 가우시안 분포의 가중치 합으로 구성되며 다음과 같은 식으로 표시된다.

$$p(\vec{x} | \lambda) = \sum_{i=1}^M p_i b_i(\vec{x}) \quad (3)$$

식 (3)에서 \vec{x} 는 D차원의 랜덤 벡터이며 $b_i(\vec{x})$, $i=1, \dots, M$ 은 성분 가우시안 분포이고 p_i , $i=1, \dots, M$ 은 결합 가중치(mixture weight)라고 불리며 각각의 가우시안 분포에 대한 가중치이다. 각 분포는 D차원의 가우시안 분포이며 식 (4)와 같이 표현된다.

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp \left[-\frac{1}{2} (\vec{x} - \vec{\mu}_i)' \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i) \right] \quad (4)$$

이때 $\vec{\mu}_i$ 는 평균 벡터이며, Σ_i 는 공분산 행렬로 구조는 대각 행렬로 이루어져있다. 그리고 결합 가중치는 식 (5)를 만족한다.

$$\sum_{i=1}^M d_i = 1 \quad (5)$$

가우시안 혼합분포는 위에서 기술한 파라미터들 즉 평균 벡터와 공분산 행렬 그리고 결합가중치에 의해 완전히 표현되며 식 (6)과 같이 표현된다.

$$\lambda = \{d_i, \vec{\mu}_i, \Sigma_i\}, i = 1, \dots, M \quad (6)$$

2.3. 허용구간분석(Tolerance Interval Analysis)

주어진 모집단(population) 내의 어떤 일부 영역 R의 커버리지(coverage) C값이란 이 일부 영역 R이 전체 모집단에 대해서 차지하는 비율로 정의하며, 다음의 식 (7)로 표현할 수 있다.

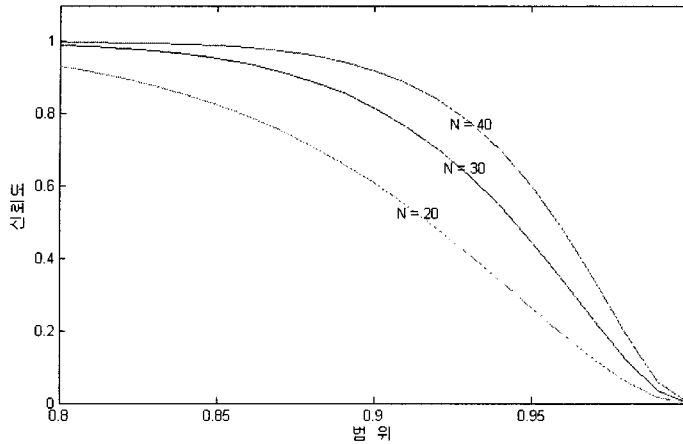
$$C(R) = P(X \in R) \quad (7)$$

여기서 X는 모집단에서 임의로 선택한 표본이며 영역 R의 커버리지는 모집단에서 임의로 하나의 표본을 추출했을 때, 그것이 영역 R에 속할 확률을 의미한다.

허용영역(tolerance region)은 영역의 커버리지 값이 최소한 c 이상이 될 확률이 $1 - \alpha$ 가 되는 임의의 영역을 말한다. 이때 $1 - \alpha$ 를 신뢰도(confidence level)라 하고 c를 허용비율(tolerance proportion)이라고 부른다. 임의의 연속분포 랜덤변수에서 n개의 독립적인 샘플을 선택했을 때, 샘플의 가장 큰 값으로부터 가장 작은 값까지로 주어지는 영역의 커버리지 값 C가 작아도 c 이상이 될 확률, 즉 신뢰도 값 P는 다음과 같이 구해진다[6][7].

$$P(C > c) = \sum_{k=0}^{N-2} {}_N C_k c^k (1-c)^{N-k} \quad (8)$$

허용비율 c값과 신뢰도(confidence level)의 관계는 <그림 2>에서 보면, 훈련용 샘플의 개수(N)가 증가하면 할수록 고정된 신뢰도 수준에서 커버리지 값이 늘어나는 것을 알 수 있다. 예로 신뢰도를 0.9로 고정된 후, 샘플 개수를 N=20, 30 40 으로 증가할 때, 각각의 커버리지값 c를 계산하면 0.82, 0.88, 0.91로 샘플집합이 모집단을 대표할 수 있는 영역의 비율이 증가함을 보이고 있다.

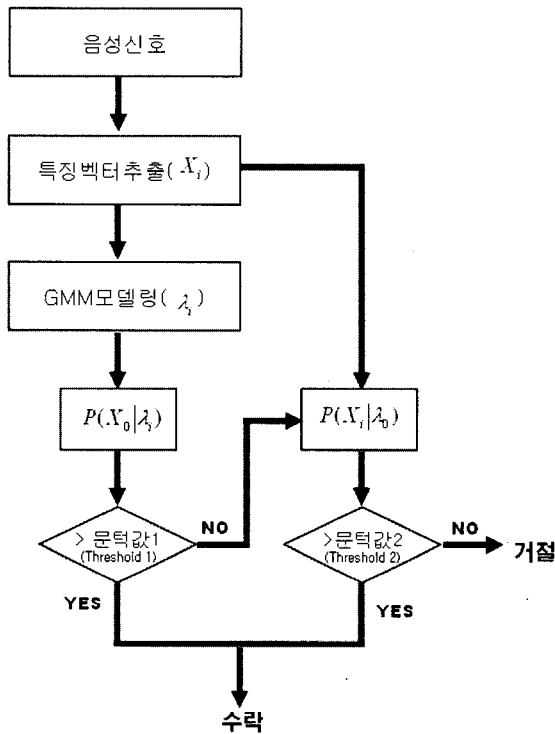


<그림 2> 샘플개수 N에 따른 신뢰도와 커버리지의 관계

2.4. 제안된 화자인증 알고리즘

통계학에서 제시한 허용구간분석(tolerance interval analysis)을 이용하면, 화자인증시스템에서 배경화자 없이 의뢰인의 음성데이터에서 계산한 우도값만을 가지고 인증을 할 수 있다. 즉, 어느 의뢰인의 화자모델의 모집단 전체를 구할 수는 없으므로, 이를 한정된 샘플로 전체 모집단을 대신하도록 하는 것이다. 이를 위해 허용구간분석 이론을 사용하면, 화자인증시스템의 성능을 고려하여, 미리 신뢰도 수준 $1 - \alpha$ 와 허용비율 c 값을 정하면, 이를 만족하는 화자모델 샘플 개수(N)를 식 (8)에서 계산할 수가 있다. 구해진 개수의 샘플에 해당되는 화자의 모델 집합을 구하면 이것이 정해진 신뢰도 수준과 허용비율로 의뢰인 화자의 전체 모집단을 대신한다고 가정할 수 있다. 여기서 화자를 대표하는 화자모델(λ_0)과 특징벡터(X_0)는 다음과 같이 선정한다. 먼저 N개의 의뢰인 음성 데이터를 이용하여 특징벡터(X_j)와 화자모델(λ_j)을 추출한다. 이렇게 구한 각각의 화자모델과 특징벡터들을 이용하여 우도값 $P(X_j|\lambda_j)$, i 와 $j = 1 \dots N$ 을 계산하게 되면, 비록 N^2 개의 한정된 개수의 우도값이지만 특정 의뢰인의 화자모델에 대한 우도값의 분포를 나타나게 된다. 위에서 구한 우도값의 샘플을 이용해서 주어진 화자모델(λ_i)들에 대한 우도값의 분산이 가장 작은 특징벡터를 대표 특징벡터(X_0)로 정한 다음, 이 특징벡터에 대한 화자모델들의 우도값 $P(X_0|\lambda_i)$, $i = 1, \dots, N$ 들의 집합에서 가장 작은 우도값을 1단계의 인증 판정을 위한 문턱값1(Threshold 1)로 정하게 된다. 그리고 대표 특징벡터(X_0)를 찾는 방법과 마찬가지로 주어진 의뢰인의 특징벡터(X_j)들에 대한 우도

값의 분산이 가장 작은 화자모델을 대표 화자모델(λ_0)로 정하고, 이 대표 화자모델에 대한 특징벡터들의 우도값 $P(X_j|\lambda_0)$, $j = 1, \dots, N$ 들 중에서 가장 작은 우도값을 문턱값2(Threshold 2)로 설정한다. 이와 같이 우도값의 샘플집합을 이용해서 정해진 신뢰도 수준과 허용비율로 모집단을 대표하는 화자모델과 특징벡터를 선정하게 되며, 선정된 대표 화자모델과 대표 특징벡터들에 의해 계산되는 우도값들이 결국은 인증시스템의 의뢰인에 대한 인증 수락 판정구간을 나타나게 된다. 따라서 판정구간의 최소 우도값을 의뢰인의 진위여부를 판단하는 기준값인 문턱값으로 사용할 수 있는 것이다.



<그림 3> 제안된 화자인증 알고리즘 흐름도

<그림 3>은 배경화자 없는 간략화된 화자인증 알고리즘의 흐름도이다. 시스템의 입력단에서 의뢰인의 입력음성이 들어오면, 먼저 특징벡터를 추출하고 이로부터 GMM 모델을 학습하여 구한 후, 이를 이용하여 미리 사전에 구한 의뢰인의 대표 특징벡터(X_0)에 대한 우도값을 얻는다. 다음은 구해진 우도값이 사전에 정한 문턱값1과 비교해서 문턱값1 보다 크면 화자인증을 수락하고 아니면 다음 단계로

진행한다. 2단계에서는 대표 특징벡터와 마찬가지로 방법으로 허용구간분석 이론을 적용해서 선정한 의뢰인의 대표 화자모델(λ_n)을 사용해서 우도값을 계산하여 이를 두 번째 문턱값2와 비교하여 최종 인증을 하게 된다. 이차에 걸친 인증단계를 거치는 것은 1단계에서 의뢰인이 인증수락이 안된 오거부인 경우, 실험 데이터의 분포를 보면 우도값이 문턱값1과 거의 차이가 나지 않은 경우가 많아, 이를 다시 한번 검증하기 위해 대표 화자모델에 대한 우도값을 이용하여 최종 판정을 하여 오거부율을 줄일 수 있기 때문이다. 이러한 2단계에 걸친 인증은 알고리즘은 약간 복잡해지지만, 인증 성능이 대표 특징벡터만을 사용하는 경우에 비해 편차도 적으며, 오거부율도 작아져서 보다 안정적인 성능을 보여주고 있다.

3. 실험 및 결과

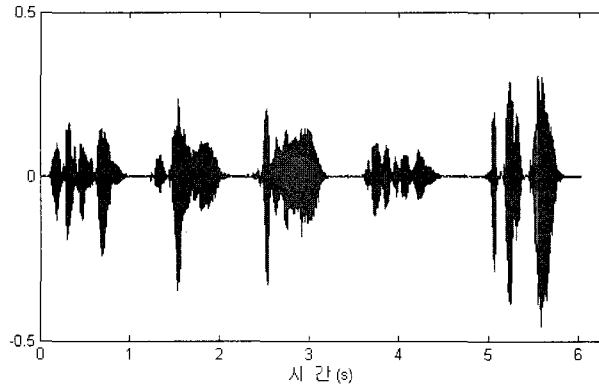
3.1. 음성 DB의 구성

실험에 사용한 음성 DB는 <표 1>과 같으며 ETRI의 음성정보연구센터에서 만든 한국어 화자인식용 DB를 사용하였다. 음성데이터는 SNR이 25dB 이상이 되는 조용한 사무실 환경에서 중저가 마이크로 녹음한 4연 숫자로 구성되어 있으며, 한 화자당 4연 숫자 목록에서 임의로 50개를 선택하여 5회 발생하고, 월차별로 구분하여 각각 4회 반복하여 녹음 수집하였다. 화자는 남녀 각각 10명으로 총 20명의 화자이며, 음성데이터는 16kHz/16비트의 Linear PCM 방식으로 코딩되어 있다. 4연 숫자음 한 개의 평균 지속시간은 1초를 약간 넘는 정도이며, 화자등록 및 실험을 위해서 10개의 데이터를 합친 다음 끝점검출을 하여 평균 10초 정도 분량의 데이터를 사용하였다.

<표 1> 실험에 사용한 음성DB

녹음환경	조용한 사무실 환경
데이터 포맷	16kHz/16bit, linear PCM
발성목록	0~9사이 숫자로 된 4연 숫자음
화자의 수	20명(남녀 10명씩)
발성주기	월차별 각 5회 발생

아래의 <그림 4>는 실험에 사용한 4연 숫자음 5개를 연결한 데이터 샘플의 파형이며, 발성 내용은 “오육공삼 공팔일이 사공일이 칠구칠구 사팔오팔”이다.



<그림 4> 실험에 사용한 4연 숫자 연속음의 파형

<표 2>는 인증에 사용한 화자군의 분류와 음성시료의 개수를 보여주고 있다. 실험에 사용한 음성데이터는 1회 50개의 4연 숫자음을 녹음하고 이를 5회 반복하며, 이를 월차 간격으로 1차, 2차 녹음한 데이터이므로 총 500개의 4연 숫자음으로 구성되어 있다. 이를 화자모델링의 정확도를 고려하여 4연 숫자음 10개를 연결하여 10초 분량의 1개의 음성파일로 만들어 총 50개의 실험용 파일을 만들어 사용하였다. 이중 40개의 음성파일은 훈련용으로 사용하였으며, 나머지 10개는 인증 실험용으로 배정하였다. 특히 월차별로 녹음한 음성들의 특징벡터의 변이가 나타나는 것을 고려하여 1차 녹음데이터와 2차 데이터를 50%씩 균등 배분하여 훈련 및 실험에 적용하여 이에 대한 편차를 줄였다.

<표 2> 인증에 사용한 화자군의 분류와 음성시료의 개수

실험용 음성 DB의 구성	화자 수	화자당 의뢰인 수	화자당 사칭자 수	총 의뢰인 파일 수	총 사칭자 파일 수
	A	B	C	$D = A \times B$	$E = A \times C$
	20	10	190	200	3800

3.2. 화자인증 실험 및 성능 비교

<표 3> 화자인증 실험환경

실험환경	매개변수 값
GMM 차수	16차
특징벡터	MFCC 12차
프레임 길이	20ms(320샘플)
프레임 주기	10ms(160샘플)
Window	Hamming

화자인증시스템에서 화자들에 대한 실험 환경은 <표 3>과 같이 가우시안 성분의 개수가 16개인 GMM모형을 사용하였고, 음성의 특징벡터는 12차 MFCC로, 음성데이터는 20ms 구간을 한 프레임으로 하고 10ms 중첩하여 프레임율(frame rate)을 초당 100 프레임으로 하였다[1]. 실험에서 기존의 배경화자모형을 이용하는 화자인증시스템의 구성은 배경화자 수를 8개로 고정하고 식(2)를 이용하여 우도비를 계산하였다. 이 기준시스템의 오거부율(FR)과 오인식율(FA)은 각각 5.5%와 5.4%로 구해졌다. 기존의 인증시스템에서는 인증판정에 적용하는 문턱값에 따라서 오거부율과 오인식율의 비율이 달라질 수 있으므로 일반적인 성능기준으로 사용되는 EER(Equal Error Rate)를 구한 후, 이때의 우도값으로 문턱값을 정하였다. 위의 인식률의 차는 문턱값을 정확히 50대 50으로 정하기 어렵기 때문에 가장 근접한 경우의 각각의 인식률로 차이가 발생하게 된다. 그러나 허용구간분석 이론을 적용하여 제안한 화자인증시스템의 성능은 사전처리에 의해서 구한 고정된 문턱값들을 사용하므로, 오인식율과 오거부율이 각각 다르게 나올 수밖에 없다.

<표 4> 기존 화자인증시스템과 제안된 화자인증시스템의 성능 비교

샘플수	FR(오거부율)			FA(오인식율)		
	N=20	N=30	N=40	N=20	N=30	N=40
제안된 방법(2단계)	9.8%	8.0%	3.0%	5.5%	4.5%	4.3%
기존 1단계 인증방법	11.2%	9.5%	4.2%	5.5%	4.5%	4.3%
Baseline	5.5%			5.4%		

<표 4>의 실험결과를 보면 모집단을 대표하기 위해서 선택하는 샘플의 개수가

N=20, 30, 40으로 증가하면서 제안된 방법에서 전반적인 인식률의 향상을 가져오고 있음을 확인할 수 있다. 이 결과는 앞의 허용구간분석 이론에서 고정된 신뢰도에서 샘플 개수의 증가에 따른 샘플집합의 커버리지가 증가하기 때문이라고 볼 수 있다. 그리고 오거부율의 경우는 N=20일 때 9.8%에서, N=40일 때, 3.0%로, 오인식율의 경우는 각각 5.5%에서 4.3%로 샘플의 개수의 증가와 함께 인식오류가 줄어들고 있음을 알 수 있다. 특히 2단계로 화자모델을 이용하여 인증판단을 하는 방법이 그렇지 않은 경우에 비해 오거부율에서 1-2% 정도 개선됨을 알 수 있었다. 이를 기존의 배경화자 있는 인증시스템에서 구한 것과 비교했을 때, N=20일 경우에는 기존의 방법에 비해 매우 낮은 인식성능을 보이나, N=40 정도로 샘플 개수를 확장하는 경우에는 성능이 기존의 방법보다 2% 정도의 개선효과를 볼 수 있었다.

5. 결론 및 고찰

앞으로 휴대폰과 같은 휴대용 단말기의 보급 확대에 따른 휴대용 인증시스템의 수요가 급증할 것으로 예상되나, 휴대용 단말기의 경우, 메모리용량, 중앙처리장치의 연산능력이 서버급에 못 미치는 등의 하드웨어적인 제약조건으로 가능한 작은 메모리와 연산능력을 요하는 인증알고리즘이 필요하게 될 것이다. 기존의 배경화자를 이용한 우도비를 기준으로 의뢰인의 인증여부를 결정하는 화자인증시스템에서는 배경화자를 선정하기 위한 많은 수의 화자와 화자별로 큰 용량의 음성데이터를 필요로 하게 되는 문제가 있다. 그러나 본 논문에서 제안하는 알고리즘의 경우는 배경화자가 필요 없기 때문에 인증시스템을 사용하는 화자이외의 다른 사람의 음성데이터가 필요 없고, 본인의 음성데이터만을 갖고 알고리즘을 수행하기 때문에 메모리 용량 측면에서 보다 효율적이며, 인증실험 시 우도값을 계산할 때도 적은 연산량이 소요되므로 단말기와 같은 휴대용 인증시스템에 효과적으로 적용할 수 있을 것이다. 제안된 알고리즘의 성능을 비교하기 위하여 GMM기반 화자인증시스템을 구현한 후, 중저가 마이크로 녹음한 4연 숫자 연속음으로 이루어진 10초 분량의 음성데이터들을 사용하여 실험하였다. 실험결과를 보면 기존의 시스템에서 얻은 성능은 오거부율(FR)과 오인식율(FA)은 각각 5.5%와 5.4%로 나타났으며, 제안된 인증시스템의 경우는 모집단을 대표하는 샘플의 개수가 N=20, 30, 40으로 증가하면서 전반적인 인식률의 향상을 가져오고 있음을 확인할 수 있었다. 즉, 오거부율의 경우는 N=40일 때, 3.0%로, 오인식율의 경우는 4.3%로 샘플의 개수의 증가와 함께 인식오류가 줄어들고 있으며, N=20일 경우에는 기존의 방법에 비해 매우 낮은 인식성능을 보이나, N=40 정도로 샘플 개수를 확장하는 경우에는 성능이 기존의 방법보다 2% 정도의 개선효과를 보았다. 이와 같이 허용구간분석

이론을 이용하여 화자모델 및 문턱값을 원하는 신뢰도 및 커버리지 범위 내에서 산출하여 이용하면 배경화자 없이 비교적 간단한 구조의 인증시스템으로 기존의 성능을 얻을 수 있음을 확인할 수 있었다. 앞으로 연구방향은 시간에 따른 음성특징의 변화와 계속되는 사용자의 입력음성을 적응적으로 표준 화자모델 및 문턱값에 적절하게 반영할 수 있는 방법을 모색해야 할 것이다.

참 고 문 헌

- [1] L. R. Rabiner, R. W. Schafer, Digital processing of speech signals, Prentice Hall, 1978.
- [2] D. A. Reynolds, "Speaker identification and verification using gaussian mixture speaker models", *Speech Communication*, Vol. 17 pp.91~108, 1995.
- [3] D. A. Reynolds, *A Gaussian Mixture Modeling Approach to Text-independent Speaker Identification*, Ph. D. Thesis, Georgia Institute of Technology, 1992.
- [4] D. A. Reynolds, R. C. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models", *IEEE. Trans. On Speech and Audio Processing*, Vol. 3, No. 1, pp.72-83, 1995.
- [5] H. Gish, M. Schmidt, "Text-independent speaker identification", *IEEE Signal Processing Magazine*, pp.18-32, 1994.
- [6] E. C. Real, A. H. Baumann, "Open set classification using tolerance intervals", *the 34th Asilomar Conference on Signals, Systems and Computers*, Monterey, CA, USA, 2000.
- [7] C.-N. Hsu, H.-C. Yu, B.-H. Yang, "Speaker verification without background speaker models", *Proc. ICASSP*, pp.II-233-236, 2003.

접수일자 : 2005년 11월 15일

게재결정 : 2005년 12월 20일

▶ 최홍섭(Hong Sub Choi)

주소: 487-711 경기도 포천시 선단동 대진대학교 공과대학 전자공학과

소속: 대진대학교 공과대학 전자공학과

전화: +82-31-539-1903, Fax: +82-31-539-1900

E-mail: hschoi@daejin.ac.kr