

화자인증 시스템에서 배경화자 선정 방법에 관한 연구*

A Study on Background Speaker Selection Method in Speaker Verification System

최 홍 섭**
Hong Sub Choi

ABSTRACT

Generally a speaker verification system improves its system recognition ratio by regularizing log likelihood ratio, using a speaker model and its background speaker model that are required to be verified. The speaker-based cohort method is one of the methods that are widely used for selecting background speaker model. Recently, Gaussian-based cohort model has been suggested as a virtually synthesized cohort model, and unlike a speaker-based model, this is the method that chooses only the probability distributions close to basic speaker's probability distribution among the several neighboring speakers' probability distributions and thereby synthesizes a new virtual speaker model. It shows more excellent results than the existing speaker-based method.

This study compared the existing speaker-based background speaker models and virtual speaker models and then constructed new virtual background speaker model groups which combined them in a certain ratio. For this, this study constructed a speaker verification system that uses GMM (Gaussian Mixture Model), and found that the suggested method of selecting virtual background speaker model shows more improved performance.

Keywords: Speaker Verification, Background Speaker, MSC, MSF, Virtually Synthesized Cohort Models

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. 가장 주된 정보는 화자의 배경(출신지역, 교육수준, 모국어 사용여부)이나 화자의 감정 및 건강상태 그리고 성도의 물리적 특성이 음성 신호에 내재되어 있다. 이들 중 많은 부분이 화자에 종속적인 것이고 화자를 구분하는데 사용할 수 있는 정보이다. 아직까지 이들 정보가

* 이 논문은 2001학년도 대진대학교 학술연구비 지원에 의한 것임.

** 대진대학교 이공대학 전자공학과

화자마다 어느 정도까지 독특한 특징을 나타내는지 알려지지 않았지만 화자간의 변이가 화자 내의 변이에 비해 상대적으로 크다는 성질을 이용하여 어느 정도의 신뢰성을 가지고 화자를 구분할 수 있다. 이러한 화자간의 변이를 이용하여 발성한 사람을 알아내는 것을 화자인증시스템이라 한다.

일반적으로 화자인증 시스템에서는 검증하고자 하는 화자모델과 그의 배경화자모델을 이용하여 log likelihood ratio의 값을 정규화 함으로써 시스템의 인식률을 향상시킨다. 이러한 배경화자 모델선정에 많이 이용되는 방법의 하나인 화자기반 cohort 방법에는 검증하고자 하는 화자와 가장 유사도가 높은 화자모델들을 배경화자 모델로 선정하는 MSC(Maximally Spread Close) 방법과 이와 반대로 가장 유사도가 작은 화자모델들을 배경화자모델로 택하는 MSF(Maximally Spread Far) 방법이 있다. 최근에는 가상조합(virtually synthesized) cohort 모델로서 가우시안 확률분포기반의 cohort 모델방법이 제안되었으며, 이는 화자기반과는 달리 주변의 여러 화자들의 확률분포에서 기준화자의 확률분포와 가까운 확률분포들만을 골라서 새로운 가상의 화자모델을 합성하는 방법으로 기존의 화자기반 방법에 비해 우수한 결과를 보여주고 있다.

본 논문에서는 가상 화자모델을 구하고, 이들의 일정한 비율의 조합으로 새로운 가상의 배경화자 모델군을 구성하는 방법을 제안하였으며, 이를 기존의 방법인 화자기반 배경화자 모델과 성능을 비교하는 실험을 하였다. 이를 위해서 GMM(Gaussian Mixture Model)을 사용하는 화자인증 시스템을 구성하였고, 제안된 가상의 배경화자모델 선정방법이 보다 향상된 성능을 보여주고 있음을 확인하였다.

2. 화자인증 시스템의 기본적 구성

화자인증은 발성된 음성이 원하는 화자, 즉 의뢰인인지 또는 사칭자인지를 구분해 내는 것으로 의뢰인에 대한 초기 등록이 요구된다. 이 방법은 그림 1과 같이 입력 음성신호와 그 음성을 발성한 화자에 대한 인증번호가 같이 주어지며, 시스템은 그 인증번호에 해당되는 화자에 대한 참조 패턴과 입력된 음성신호가 일치하는 지를 검사하여 일치 또는 불일치라는 판정을 내리게 된다. 즉 입력음성에 대한 화자인증은 다음과 같은 기준에 의해 판정을 내리게 된다.

화자 인증 : if $d(T(x), P(x)) \leq$ 임계치

화자 거부 : if $d(T(x), P(x)) >$ 임계치

$T(x)$: 검사할 입력패턴

$P(x)$: 기준화자에 대한 참조패턴

$d(T(x), P(x))$: 입력 패턴과 참조패턴과의 거리

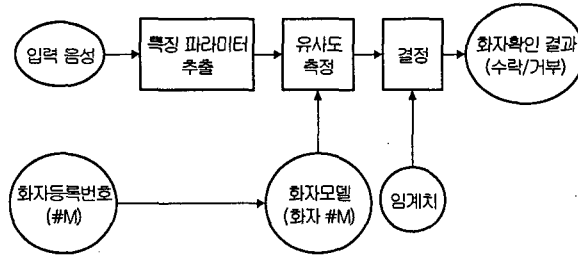


그림 1. 화자인증시스템의 구성

3. 기존의 화자기반 배경화자 선정방법

배경화자를 이용할 때 고려해야 할 문제는 배경화자를 선택하는 방법과 배경화자의 수에 관한 것이다. 직관적으로 볼 때 배경화자(background speaker)는 일반적인 응용에서 명확히 기대되는 사칭자들의 모임으로 구성될 것이다. 이러한 사칭자들은 비슷한 음성의 특징을 가지거나 적어도 동성의 화자들로 구성될 것이다. 반면에 전화 기반의 응용 예에서는 보다 넓게 분포된 사칭자들의 접근을 가정할 수 있는데 예를 들면 남성 사칭자가 여성 사용자를 사칭하는 경우이다. 기존의 시스템에서는 등록된 화자에 최대한 접근한 화자들을 배경화자로 선택하기 때문에 보통의 경우에 시스템의 신뢰도를 확보할 수 있었지만 위와 같이 매우 다른 음성특징을 갖는 사칭자에게는 취약한 약점이 있다. 이런 현상이 발생하는 이유는 매우 다른 음성이 유사도 수식의 분모, 분자들로 잘 모델링이 안되기 때문이다.

그리고 일반적으로 배경화자의 수는 사칭자 집합을 충분히 모델링 할 수 있을 만큼 커야 하지만 실제적으로는 계산량과 컴퓨터의 저장용량을 고려하여 제한된 작은 수의 배경화자를 이용하게 된다. 실시간 계산을 고려하고 일정한 실험 방법을 반복적으로 적용하기 위해 각각의 화자는 사용자 또는 의뢰자가 되는 한편 사칭자로도 이용되어 모든 화자에 대해 돌려가며 실험을 한다. 즉, 배경 화자를 많이 선택할 경우에는 사칭자의 수가 줄어들게 되지만 가능한 많은 실험을 위해 화자들을 최대한 사용하게 된다.

3.1 MSC(Maximal Spread Close) 방법

음성 DB에서 훈련데이터를 이용하여 모든 화자의 GMM모델을 만들고, 화자모델 사이의 거리를 계산한다. 모델 (λ_i, λ_j) 과 훈련음성 (X_i, X_j) 을 가진 화자 i, j 간의 거리는 다음 식(1)과 같이 정의한다.

$$d(\lambda_i, \lambda_j) = \log \frac{p(X_i | \lambda_i)}{p(X_i | \lambda_j)} + \log \frac{p(X_j | \lambda_j)}{p(X_j | \lambda_i)} \quad (1)$$

거리가 가장 가까운 화자 N명($N > B$, B는 최종적인 배경화자군의 크기)을 근접 배경화자군(close cohort)으로 선택하며, 최종 배경화자군 $b(i)$ 는 가장 근접한 화자 중에서 최대한 고루 분포된 화자들로 구성된다.

$C(i)$: 화자 i 와 가장 근접한 화자 N 명으로 구성

$b(i)$: 최종적인 배경화자 군

(단계 1) $C(i)$ 에서 가장 근접한 화자를 $b(i)$ 로 이동, $N = N-1$, $B' = 1$

(B' 는 $b(i)$ 에서 현재 화자의 수).

(단계 2) 다음 조건을 만족하는 화자 c 를 $C(i)$ 에서 $b(i)$ 로 이동,

$$c = \arg \max_{c \in C(i)} \left\{ \frac{1}{B'} \sum_{j \in b(i)} \frac{d(\lambda_j, \lambda_c)}{d(\lambda_i, \lambda_c)} \right\} \quad (2)$$

$N = N - 1$, $B' = B' + 1$

(단계 3) 단계(2)를 $B' = B$ 일 때까지 반복한다.

이렇게 함으로써 비슷한 화자를 중복해서 배경화자로 선택하는 것을 방지할 수 있고, 한정된 배경화자수에서 최적의 선택 범위를 얻을 수 있다. 이와 같은 방법을 MSC(Maximally-Spread Close)라고 한다.[4]

3.2 MSF(Maximal Spread Far) 방법

예상되는 사칭자군이 기준화자와 비슷하지 않은 화자들을 포함하고 있을 때, 배경화자 선택은 멀리 떨어져 있는 화자들을 일부 포함하고 있어야 할 것이다. 이 경우 배경화자군의 구성은 거리가 근접한 화자군 일부와 멀리 떨어진 화자군 일부를 혼합하게 되는데, 이때 가장 멀리 떨어져 있는 화자 N 명을 추출한 다음, 이중 고루 분포된 $B/2$ 명을 배경화자의 일부로 선정한다. 선정방법은 다음의 알고리즘을 따른다.[4]

$F(i)$: 화자 i 로부터 가장 멀리 떨어진 화자들의 집합

$b(i)$: 최종적인 배경화자 군

(단계 1) $F(i)$ 중에서 가장 멀리 떨어져 있는 화자를 $b(i)$ 로 이동, $N = N - 1$, $B' = 1$

(단계 2) $F(i)$ 에서 다음 조건을 만족하는 화자 f 를 $b(i)$ 로 이동, 이때 f 는 다음 방법에 의해서 선정,

$$f = \arg \max_{f \in F(i)} \left\{ \frac{1}{B'} \sum_{j \in b(i)} d(\lambda_j, \lambda_f) * d(\lambda_i, \lambda_f) \right\} \quad (3)$$

(단계 3) 단계(2)를 $B' = B/2$ 일 때까지 반복한다.

4. 가상 배경화자 선정방법

그림 2는 가상 배경화자를 구성하는 개념도를 보여주고 있다.[12] 기준화자 I와 근접한 화자 (A, B, C, D)가 각각 믹스처가 3인 GMM 모델을 구성하고 있다. 가상의 배경화자를 선정하기 위해 기준화자 I의 각각의 믹스처와 가장 거리가 가까운 믹스처를 찾는다. 이때 가우시안 믹스

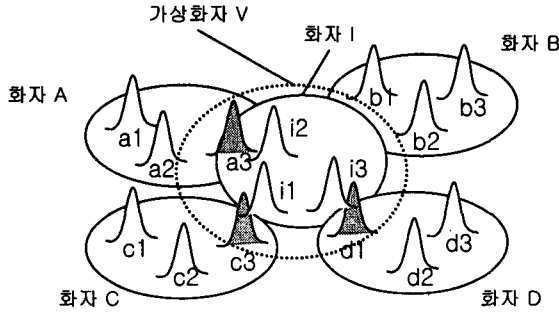


그림 2. 가상화자의 개념도

처 간의 거리는 Bhattacharyya 거리를 사용한다.[11] 그림의 예에서 기준화자의 믹스처 i1과 근접한 믹스처는 화자C의 c3이고, i2와는 화자A의 a3, i3과는 화자D의 d1이 선택된다. 그리고 각각의 선택된 믹스처 a3, c3, d1으로 가상의 배경화자 모델 λ_v 를 구성하는 것이다.

가상의 배경화자 모델 $\lambda_v = \{p_m, \mu_m, \Sigma_m\}$ 는 다음과 같다.

p_m : 기준화자 i 의 가중치 p_m

μ_m : 선택된 화자의 믹스처 평균 μ_v

Σ_m : 선택된 화자의 믹스처 표준편차 Σ_v

가상으로 구한 화자모델이 기준화자의 모델에 얼마나 근접한 지를 입증하기 위하여 모델들의 확률밀도함수의 모양을 비교하여 보았다. 그림 3의 실선부분이 기준화자이고 점선부분이 가상의 화자모델이며, 이는 12 차 특징벡터의 일부에 대한 확률모델 GMM의 확률밀도함수로 서로 매우 유사한 분포를 갖고 있음을 알 수 있으며 이로서 위의 방법으로 구성한 가상의 화자에 대한 확률모델이 유용함을 확인할 수 있겠다.

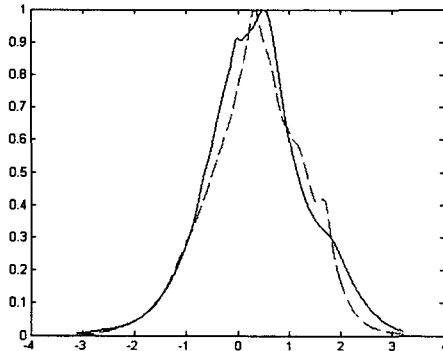


그림 3. 기준화자의 모델(실선)과 가상의 모델(점선)의 비교 그림

가상의 배경화자모델을 화자인증 시스템에 적용한 결과는 다음 표 1과 같다.

표 1. 단일 가상화자를 선택한 결과

	maximum log-likelihood
MSC	4.10
MSC+MSF	6.15

이는 화자기반에서 MSC, MSC+MSF로 선택된 화자모델의 믹스처로부터 가상의 배경화자 한 명을 구성하여 실험한 결과로 EER의 값이 4.10과 6.15로 여러 명으로 구성되는 화자기반보다는 좋지 못한 결과이다.

따라서 본 논문에서는 이에 대한 개선방법으로 화자기반의 방법에서와 같이 여러 명의 가상 화자모델들로 가상의 배경화자군을 형성하는 방법을 제안하였다.

5. 실험 및 결과

5.1 음성 DB

음성 DB로는 YOHO 표준 DB를 사용하였다. YOHO DB는 문장중속 화자인식 실험을 위해 만들어졌으며, 총 138 명의 화자 데이터로 구성되었다. 각각의 화자마다 4 번의 등록세션과 10번의 테스트세션이 있으며, (35-72-41)과 같이 숫자를 발음한 파일이 등록세션에는 24 개씩, 그리고 화자인증실험을 위한 테스트세션에는 4 개씩의 음성파일이 있다. 데이터들은 21-97까지의 56 개의 숫자들의 조합으로 구성되었으며, 고급 전화 마이크를 사용하여 조용한 환경의 사무실에서 녹음하였다.

본 실험에서는 반복적인 실험을 위하여 계산량과 데이터량을 고려하여 대상 화자의 수를 20 명으로 제한하였다. 기준화자의 등록을 위하여 등록세션에 있는 24 개 파일을 모두 합친 약 1 분 40 초 길이의 데이터를 사용하였고, 테스트에는 YOHO DB의 각 테스트세션의 4 개의 파일을 합친 20 초 데이터를 만들어 세션별로 10 회 실험하였다.

5.2 화자기반 배경화자 선정방법 실험

배경화자를 선정하기 위해서 먼저 그림 4의 과정을 통해 각각의 화자를 모델링하는데, 이때 사용한 GMM 믹스처는 32 차로 하였으며[8] 12차 MFCC(Mel-Frequency Cepstral Coefficient)를 특징 파라미터로 사용하였다.[4]

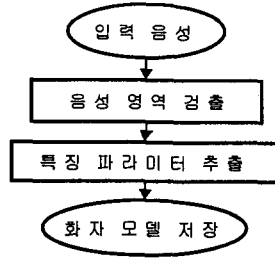


그림 4. 화자모델

다음으로 화자모델들로부터 기준 화자 i에 대한 배경화자를 선택하기 위해 그림 5와 같이 화자 i의 배경화자 선택을 위해 각각의 화자모델들과 거리를 계산한다.

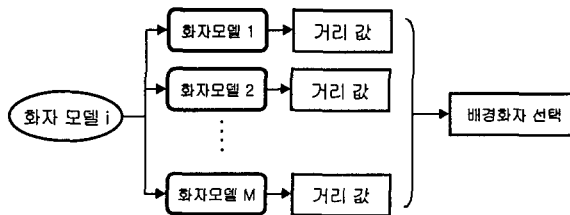


그림 5. 화자 기반 배경화자 선정

이때 배경화자 선택은 MSC 방법과 MSC+MSF 방법 2 가지로 한다. 배경화자 선택시 예비 배경화자 수는 N=12로 하고 최종 배경화자의 수는 B=6으로 정하였다. 다음으로 화자 i의 테스트용 데이터 10 개와 화자 i와 배경화자로 선택된 6명의 화자를 제외한 나머지 13 명의 테스트 데이터 130 개를 사칭자로 사용하여 화자인증 실험을 하였으며, 결과로 나온 likelihood 값들을 기준값과 비교하여 화자인증 수락여부를 결정하였다.

배경화자에 대한 계산에는 log-likelihood 방법과 normalized log-likelihood 방법 그리고 maximum 방법 등을 사용하였으며, log-likelihood 방법은 식(5)와 같으며, normalized log-likelihood 방법은 식 (6), 그리고 maximum log-likelihood 방법은 식(7)과 같이 계산하였다.

$$\log p(X | \lambda_{\bar{c}})_{ll} = \log \left\{ \frac{1}{B} \sum_{b=1}^B p(X | \lambda_b) \right\} \tag{5}$$

$$\log p(X | \lambda_{\bar{c}})_{nll} = \frac{1}{B} \sum_{b=1}^B \log p(X | \lambda_b) \tag{6}$$

$$\log p(X | \lambda_{\bar{c}})_{mll} = \arg \max_{\lambda_{\bar{c}} = b(i)} \log p(X | \lambda_{\bar{c}}) \tag{7}$$

화자기반 배경화자 선정방법에 의한 화자인증 실험의 결과는 다음 표 1에 있으며, MSC 경우에는 normalized log-likelihood 방법이 EER값이 3.03으로 가장 좋았으며, MSC+MSF 경우에는 maximum log-likelihood 방법이 3.29로 성능이 우수했다.

표 2. 화자기반에서의 화자인증 결과.

	log-likelihood	normalized log-likelihood	maximum log-likelihood
MSC	3.73	3.03	3.65
MSC +MSF	4.09	3.33	3.29

표 3. 화자기반에서 배경화자수에 따른 결과 (MSC)

	log-likelihood	normalized log-likelihood	maximum log-likelihood
B=6	3.73	3.03	3.65
B=8	3.69	3.00	3.64

그리고 배경화자 수의 증감에 따른 시스템의 성능을 비교하기 위해 배경화자 수를 B=6, 8에 대하여 각각 실험하였으며, 표 2를 통해서 일반적으로 배경화자의 수를 증가하면 보다 나은 인식 성능을 보임을 알 수 있었다. 그러나 본 실험의 경우에는 전체 대상이 되는 화자가 20 명으로 적은 인원이어서 실험의 신뢰도를 고려하려면 보다 많은 대상 화자의 수에 대한 실험이 필요하겠다.

5.3 가상화자 기반 배경화자 선정방법 실험

가상화자기반 배경화자 선정방법 역시 화자기반과 마찬가지로 그림 4와 같이 각각의 화자 모델을 구하며 화자의 GMM 모델은 다음과 같이 구성한다.

$$\begin{aligned}
 \text{GMM의 평균} & \begin{bmatrix} m_1^1 & m_2^1 & \dots & m_{12}^1 \\ m_1^2 & m_2^2 & \dots & m_{12}^2 \\ \vdots & \vdots & \ddots & \vdots \\ m_1^{32} & m_2^{32} & \dots & m_{12}^{32} \end{bmatrix} \\
 \text{GMM의 분산} & \begin{bmatrix} \sigma_1^1 & \sigma_2^1 & \dots & \sigma_{12}^1 \\ \sigma_1^2 & \sigma_2^2 & \dots & \sigma_{12}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1^{32} & \sigma_2^{32} & \dots & \sigma_{12}^{32} \end{bmatrix} \\
 \text{GMM의 가중치} & [w_1 \quad w_2 \quad \dots \quad w_{32}]
 \end{aligned}$$

가상의 화자모델을 구성하기 위하여 그림 6과 같이 기준화자 i의 GMM의 각각의 가우시안 믹스처와 나머지 화자들의 GMM의 가우시안 믹스처간의 거리를 Bhattacharyya 방법[11]으로 계산한 후, 가장 거리가 가까운 가우시안 믹스처 32 개를 이용하여 새로운 가상의 화자 모델을 구성한다. 기준화자 GMM의 각각의 가우시안 믹스처에 대해서 가장 근접한 화자들의 믹스처들을 순서대로 나열한 다음, 이들 중에서 근접한 순서대로 32 개의 믹스처를 선택하여 가상화자모델로 정하며, 이러한 가상화자 B 개로 가상의 화자군을 만들어 배경화자로 선정한다. 이때 가장 근접한 순서대로만 가상화자군을 구성하면 기준화자의 믹스처들과 거리가 아

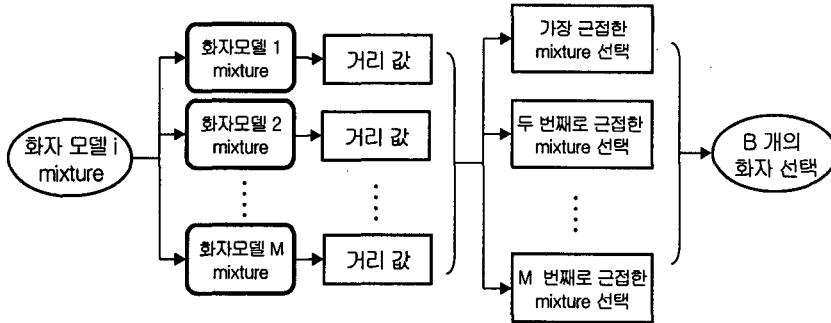


그림 6. 가상의 화자군 구성도

주 근접한 믹스처들로만 가상화자를 구성하게 되므로 이는 유사한 확률분포를 갖는 비슷한 배경화자를 중복해서 설정하는 경우와 동일하므로 가능한 이들 간의 적당한 차이를 주기 위해 L씩 (L=1, 2, 3, 4, 5) 차이를 두면서 가상화자군을 형성해 보았다. 표 4는 이에 대한 결과를 요약한 것이다. 표 2의 화자기반 선정방법의 실험에서는 6 명의 배경화자의 경우에 normalized log-likelihood에서 3.03의 EER로 가장 좋은 결과를 보이지만, 가상배경화자군의 실험에서는 EER값이 2.10-2.94 정도로 모든 방법에서 보다 좋은 결과를 보여주고 있다. 특히 maximum log-likelihood를 사용할 경우에는 1.68-2.39의 결과로 log-likelihood를 제외하고는 모든 경우에서 화자기반의 배경화자 선정방법보다 가상화자들로 구성된 배경화자군을 이용하면 성능이 개선됨을 알 수 있다. L값의 변화에 따라서 결과에 미치는 영향은 일정한 관계를 갖지는 않지만 일반적으로 L=1의 경우처럼순차적으로 거리가 가까운 가우시안 믹스처로 배경화자를 선정하는 방법보다는 L=2 또는 3 정도로 거리간격을 준 가우시안 믹스처들로 배경화자군을 구성하는 것이 좋다는 것을 알 수 있다. 특히 maximum log-likelihood에서 배경화자수를 B=6으로 할 때, L=1에서는 EER값이 2.07이지만 L=2, L=3으로 하면 EER값이 1.68로 감소되었다.

표 4(a). 가상화자군 결과 (L=1)

	log-likelihood	normalized log-likelihood	maximum log-likelihood
B = 1	7.50	4.34	4.34
B = 2	5.47	4.52	3.65
B = 3	5.10	3.78	3.10
B = 4	4.92	2.42	2.78
B = 5	2.94	2.63	2.31
B = 6	2.97	2.86	2.07

표 4(b). 가상화자군 결과 (L=2)

	log-likelihood	normalized log-likelihood	maximum log-likelihood
B = 1	7.50	4.34	4.34
B = 2	4.84	2.44	2.71
B = 3	4.84	3.15	2.23
B = 4	3.89	2.31	1.78
B = 5	4.36	2.42	1.81
B = 6	3.81	2.44	1.68

표 4(c). 가상화자군 결과 (L=3)

	log-likelihood	normalized log-likelihood	maximum log-likelihood
B = 1	7.50	4.34	4.34
B = 2	5.42	3.42	3.15
B = 3	4.23	2.81	2.36
B = 4	4.13	2.07	1.92
B = 5	4.15	2.62	1.81
B = 6	4.10	2.26	1.68

표 4(d). 가상화자군 결과 (L=4)

	log-likelihood	normalized log-likelihood	maximum log-likelihood
B = 1	7.50	4.34	4.34
B = 2	4.89	3.81	3.02
B = 3	4.21	3.36	2.39
B = 4	4.36	3.36	2.13
B = 5	3.60	2.89	2.02
B = 6	3.63	2.94	1.92

표 4(e). 가상화자군 결과 (L=5)

	log-likelihood	normalized log-likelihood	maximum log-likelihood
B = 1	7.50	4.34	4.34
B = 2	4.78	3.15	2.78
B = 3	3.65	2.65	2.73
B = 4	3.73	2.94	2.50
B = 5	3.44	2.76	2.39
B = 6	4.26	2.10	2.39

6. 결론 및 향후 과제

본 논문에서는 문장독립형 화자인증시스템에서 배경화자 선정에 관한 방법에 대해서 화자 기반과 가상 화자기반에서의 성능비교를 하였다. 결과는 화자기반보다 가상 화자기반에서 좋은 성능을 볼 수 있었고, 특히 가상의 화자 한 명을 사용할 때보다 다수의 가상화자를 선정하여 배경화자로 선정하였을 때 더 좋은 성능 향상을 얻을 수 있었다. 본 논문에서 사용한 화자의 수는 20 명으로 이를 이용하여 배경화자를 선정한다면 19 명 중에서 선정하게 된다. 가상화자에서는 화자모델의 믹스처 즉 32×19 에서 32 개의 믹스처를 선정하는 것이 기준화자와 더욱 유사한 화자 모델을 선정할 수 있고, 더욱 비슷한 사칭자를 처리할 수 있게 된다. 그리고 사칭자의 유사함이 단일 방향으로 유사하다고 할 수 없기 때문에 여러 각도에서의 화자 모델들로 구성된 배경화자군이 더욱 이상적인 배경화자임을 알 수 있었다.

또한 실험을 통해 알 수 있었던 것은 EER에서 근접한 사칭자에서 에러가 많이 나타났고 기준화자와 유사도가 떨어진 화자에서도 에러가 발생하였다. 이 에러는 기준화자와 유사도가 많이 떨어져 있고 배경화자와도 유사도가 많이 떨어져 있기 때문에 발생하는 에러이다. 이와 같은 문제는 시스템에서 기준화자와 유사도가 많이 떨어져 있는 사칭자를 사전에 처리하게 되면 해결될 것이다.

향후 연구과제로는 가상의 화자군을 기준화자와 근접한 믹스처로 구성되어 있다는 것에서 착안하여 배경화자군을 기준화자의 믹스처로부터 선정하는 방법을 모색하는 것이다. 이렇게 함으로 많은 배경화자를 선정하기 위해 사칭자군을 형성하지 않아도 되는 장점이 생길 수 있을 것이다.

참 고 문 헌

- [1] 김상경. 1999. *GMM에 기반한 화자 인식에 모음을 이용한 인증 발생 감축에 관한 연구*. 석사학위논문, 한양대학교.
- [2] 강관욱. 2000. *컴퓨터 접근 통제를 위한 화자 인식 기법적용에 관한 연구*. 연세대 석사학위논문, 연세대학교.
- [3] Rabiner L. R. & R. W. Schafer. 1978. *Digital Processing of Speech Signals*.
- [4] Reynolds, Douglas A. 1995. "Speaker Identification and Verification Using Gaussian Mixture Speaker Models." *Speech Communication*, Vol. 17, 91-108.
- [5] Horikawa-cho, Saiwai-ku, Kawasski-shi. 1999. "A New Cohort Normalization Using Local Acoustic Information for Speaker Verification." *ICASSP*.
- [6] 정종순. 1996. *대표 평균패턴과 가중 캡스트럼을 이용한 화자인식의 성능 향상에 관한 연구*. 석사학위논문, 한국 과학기술원.
- [7] 임창현, 1995. "화자 인식 기술." *제 12회 음성 통신 및 신호 처리 워크샵 논문집*, 42-46.
- [8] Reynolds, Douglas A. 1992. *A Gaussian Mixture Modeling Approach to Text-independent Speaker Identification*. 박사학위논문, Georgia Institute of Technology.
- [9] Reynolds, D. A. & R. C. Rose. 1995. "Robust Text-independent Speaker Identification Using Gaussian Mixture Speaker Models." *IEEE. Trans. On Speech and Audio Processing*, Vol. 3, No. 1, 72-83.

- [10] Gish, H. & M. Schmidt. 1994. "Text-independent Speaker Identification." *IEEE Signal Processing Magazine*, 18-32.
- [11] Therrien, Charles W. 1989. *Decision, Estimation and Classification*, Wiley.
- [12] Isobe, Toshihiro & Jun-ichi Takahashi. 1999. "Text-independent Speaker Verification Using Virtual Speaker Based Cohort Normalization." *Eurospeech*.

접수일자: 2002. 3. 5.

게재결정: 2002. 4. 6.

▲ 최홍섭

경기도 포천군 포천읍 선단리 (우: 487-711)

대진대학교 이공대학 전자공학과

Tel: +82-31-539-1903 Fax: +82-31-539-1900

E-mail: hschoi@road.daejin.ac.kr