

## 과학수사를 위한 한국인 음성 특화 자동화자식별시스템 Forensic Automatic Speaker Identification System for Korean Speakers

김 경 화<sup>1)</sup> · 소 병 민<sup>2)</sup> · 유 하 진<sup>3)</sup>

Kim, Kyung-Wha · So, Byung-Min · Yu, Ha-Jin

### ABSTRACT

In this paper, we introduce the automatic speaker identification system 'SPO(Supreme Prosecutors Office) Verifier'. SPO Verifier is a GMM(Gaussian mixture model)-UBM(universal background model) based automatic speaker recognition system and has been developed using Korean speakers' utterances. This system uses a channel compensation algorithm to compensate recording device characteristics. The system can give the users the ability to manage reference models with utterances from various environments to get more accurate recognition results. To evaluate the performance of SPO Verifier on Korean speakers, we compared this system with one of the most widely used commercial systems in the forensic field. The results showed that SPO Verifier shows lower EER(equal error rate) than that of the commercial system.

**Keywords:** forensic phonetics, automatic speaker recognition system, speaker verification, speaker identification, GMM-UBM

### 1. 서론

범죄 현장 및 범인이 남긴 증거물을 분석하는 과학수사에서는 지문, DNA, 얼굴 등 다양한 생체 정보를 이용하여 범인의 신원을 확인한다. 과학수사의 한 분야인 음성분석은 개개인의 독특한 음성 특징을 이용하여 범인과 피의자 음성이 동일한 음성인지를 분석한다. 이 때 사용되는 분석 기술은 청취 분석, 음향분석 등 전통적인 방법과 최근 음성공학 분야에서 사용되는 화자 인식 기술이다.

범죄 수사와 관련하여 화자 인식은 여러 명의 후보 가운데 범인의 음성과 가장 유사한 음성을 찾아내는 화자 식별(speaker identification)과 범인과 특정 피의자 음성이 동일한가를 비교하는 화자 인증(speaker verification)을 수행한다.

이러한 화자 인식 기술을 이용하여 미국, 독일, 스페인, 러

시아 등에서는 범죄 수사용 자동화자식별시스템을 개발하여 현재 여러 국가의 수사 기관에서 사용하고 있다. 자동화자식별시스템은 텍스트 독립적이어서 범인과 피의자가 말하는 내용이 서로 다르더라도 분석이 가능하고 두 사람이 사용한 언어가 달라도 식별이 가능하다. 또한, 용의자와 피의자가 많아도 빠른 시간 내에 식별 결과를 확인하는 것이 가능하다.

범죄 수사에 활용되는 상용화된 화자 인식 시스템으로는 범인과 피의자 음성을 1대 1로 비교하는 시스템, 지문 자동식별시스템과 같이 저장된 대용량 음성 데이터와 신원 미상인의 음성을 1대 N으로 비교하는 시스템, 그리고 전화 도청 등을 통해 수집된 대량의 음성 파일에서 특정 음성을 탐지하고 선별하는 시스템 등이 있다.

최근 범죄 수사에서는 협박 전화, 보험사기, 대출사기 사건 등과 관련하여 범인과 피의자 음성이 동일한지 여부에 대한 감정 의뢰가 지속적으로 증가하고 있다. 의뢰되는 증거물은 일반 음성 녹음과 일반 전화 녹음, 휴대폰 녹음 등 다양하다. 전화 음성은 일반 음성 녹음과 신호 전송 경로가 다르기 때문에 음성 신호의 손실 및 왜곡이 발생하여 자동화자식별 결과에 영향을 미칠 수 있다. 따라서 다양한 채널로 녹음된 증거 자료에 대한 화자 식별 능력을

1) 대검찰청, savoix@spo.go.kr

2) 대검찰청, sbm1210@spo.go.kr

3) 서울시립대학교 컴퓨터과학부, hjyu@uos.ac.kr, 교신저자

접수일자: 2012년 7월 25일

수정일자: 2012년 9월 14일

게재결정: 2012년 9월 14일

강화할 수 있는 시스템이 필요하다.

본 논문에서는 범죄 용의자의 음성 식별 결과를 제고하기 위하여 대검찰청에서 개발한 자동화자식별시스템 'SPO (Supreme Prosecutors Office) Verifier'를 소개한다. 본 시스템은 다양한 녹음 채널로 수집된 음성에 대하여 음성 신호의 손실 및 왜곡 정보를 추출하여 이를 보정하는 음성 신호 보정 알고리즘을 사용한다. 또한 한국어인 및 한국어 특징을 반영하여 범인의 음성과 유사한 레퍼런스 모델을 사용자가 지정하여 관리하는 기능을 포함한다.

본 논문의 2장에서는 SPO Verifier 개발에 사용된 자동화자식별 기술을 설명하고, 3장에서는 본 시스템의 설계를 토대로 구현한 내용을 다룬다. 4장에서는 SPO Verifier에 대한 성능 평가 실험 결과를 제시하고 기존의 상용화된 시스템과 비교한다. 그리고 5장에서는 연구 결과를 정리하고 결론을 맺는다.

## 2. SPO Verifier 자동화자식별 기술

자동화자식별시스템은 화자 인식 기술을 이용하여 음성 신호가 누구의 음성인지 자동으로 찾아주는 시스템이다. SPO Verifier에서는 다음과 같은 자동화자식별 기술들을 사용하였다. 먼저 음성 신호로부터 자동화자식별시스템에서 사용가능한 형태인 음성 특징을 추출한다. 또한 음성 신호 간의 서로 다른 채널에 의한 효과를 보정하기 위해 음성 신호 보정 알고리즘을 사용한다. 그리고 잡음 및 환경 변화에 강인한 성능을 갖도록 특징 강화 기법을 적용할 수 있다. 이렇게 처리된 음성 특징으로 화자 모델들을 구성한다. 이후 식별하고자 하는 음성 신호가 있을 때 음성 특징을 추출하고, 추출된 특징과 화자 모델들을 비교한다. 그리고 비교 결과를 사용자에게 나타낸다. 본 연구에서는 음성 특징으로 MFCC(Mel-frequency cepstral coefficients)[1]를 사용하였다.

### 2.1 음성 신호 보정 알고리즘

특정 녹음기기를 이용하여 녹음된 음성 신호는 그 기기가 갖고 있는 채널 잡음이 포함된다. 채널 잡음의 효과를 줄이기 위해 음성 신호 보정 알고리즘을 사용한다. 음성 신호 보정 알고리즘은 cepstrum의 평균을 정규화하는 cepstrum 평균 정규화 기법(cepstrum mean normalization, CMN)[2]과 cepstrum의 평균과 분산을 정규화하는 평균 분산 정규화 기법(mean and variance normalization, MVN)[3]이 있다. cepstrum 평균 정규화 기법은 cepstrum의 평균을 구하여 각 cepstrum에서 빼주는 것으로 평균을 0으로 정규화한다. 그리고 평균 분산 정규화 기법은 cepstrum 평균 정규화 기법에 추가적으로 분산을 1로 만드는 정규화작업을 수행한다. 본 연구에서는 평균 분산 정규화 기법을 사용한다.

### 2.2 특징 강화 기법

본 연구에서 사용한 특징 강화 기법은 주성분 분석(principal component analysis, PCA)[4]이다. 주성분 분석은 고차원의 특징을 저차원으로 줄일 때 정보 손실이 최소가 되는 변환 행렬을 찾는 방법이다. 공분산 행렬을 이용하여 고유벡터(eigenvector)와 고유값(eigenvalue)을 구하고 고유값이 큰 순서대로 행렬을 나열하여 분산을 최대화하는 축으로 차원을 축소한다.

### 2.3 화자 모델 구성 방법

본 연구에서 화자를 모델링하기 위해 사용한 방법은 화자 인식 분야에서 널리 사용되고 있는 가우시안 혼합 모델(Gaussian mixture model, GMM)[5]이다. 가우시안 모델은 관찰된 전체 특징 벡터의 평균과 분산을 이용하여 분포된 모델을 표현하는 방법이다. 단일 가우시안 모델은 수식 (1)과 같다.

$$g(\vec{x}) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp -\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1}(\vec{x} - \vec{\mu}) \quad (1)$$

위의 식에서  $D$ 는 특징 벡터의 차원,  $\vec{\mu}$ 는 특징 벡터의 평균,  $\Sigma$ 는 특징 벡터의 공분산 행렬이다. 단일 가우시안 모델로는 전체 특징 벡터를 잘 표현하지 못하기 때문에 여러 개의 가우시안 분포를 사용하는 가우시안 혼합 모델을 사용한다. 가우시안 모델의 선형 결합으로 이루어진 가우시안 혼합 모델은 수식 (2)와 같다.

$$p(\vec{x}|w, \vec{\mu}, \Sigma) = \sum_{i=1}^M w_i g_i(\vec{x}) \quad (2)$$

이 때  $M$ 은 가우시안 성분의 수,  $w$ 는 각 가우시안 성분의 가중치이다. 가우시안 혼합 모델의 학습에는 EM(expectation maximization) 알고리즘을 사용한다.

가우시안 혼합 모델의 학습에 사용되는 EM 알고리즘은 충분한 데이터가 주어졌을 경우에만 정확한 모델을 추정할 수 있다. 소량의 데이터를 사용하여 모델을 추정하기 위해 MAP(maximum a posteriori)[6] 적용 방법을 사용한다. MAP 적용 방법은 잘 학습된 클래스 독립적인 가우시안 혼합 모델의 파라미터와 소량의 특징 벡터를 사용하여 클래스 종속적인 모델을 얻는 방법이다. 본 연구에서는 한국어인 화자의 음성으로 생성한 클래스 독립적인 모델인 UBM(universal background model)[7]에 시스템의 사용자가 모델로 생성하고자하는 각 화자의 음성을 MAP 적용하여 클래스 종속적인 모델을 생성하였다.

### 2.4 Cohort model 적용

본 연구에서는 유사도비(likelihood ratio, LR)를 계산할 때 cohort model[8]을 이용하였다. 자동화자식별시스템 사용자로 하여금 다수의 레퍼런스 화자 모델을 등록하도록 하고, 유사도비 계산 과정에서 입력 발생과 가장 유사도(likelihood)가 높은 레퍼런스 모델에 대하여 유사도비를 구하였다.

## 3. 시스템의 설계 및 구현

### 3.1 시스템의 설계

SPO Verifier는 크게 off-line단계인 UBM 학습과 용의자 모델 관리 모듈, 레퍼런스 모델 관리 모듈, 화자 식별 모듈로 구성되어있다.

시스템 전체에 대한 구조는 <그림 1>과 같다.



그림 1. SPO Verifier의 구조  
Figure 1. Structure of SPO Verifier

UBM 학습은 용의자 모델 학습과 레퍼런스 모델 학습에 사용되는 UBM을 생성하는 과정이다. 먼저 UBM 학습 음성에서 MFCC 특징을 추출한다. 이후 추출된 특징에 평균 분산 정규화 기법(MVN)을 사용한다. 또한 자동화자식별시스템 사용자가 특징 강화 기법을 사용할 때를 위해 앞서 추출한 특징을 사용하여 캡스트럼의 분산이 최대가 되는 특징 강화 기저를 추정한다. 그리고 추정한 기저로 특징을 강화한다. 특징 강화 기법을 적용한 특징들과 특징 강화 기법을 적용하지 않은 특징들의 두 부류로 나누어 2개의 UBM을 학습하였다. 이렇게 생성된 UBM은 하나는 특징 강화 기법을 적용할 때 사용되고 나머지 하나는 특징 강화 기법을 적용하지 않을 때 사용된다.

SPO Verifier에서 사용하는 UBM의 정보는 <표 1>과 같다.

표 1. SPO Verifier의 UBM  
Table 1. UBM of SPO Verifier

항목	내용
음성 DB	ETRI 휴대폰 화자 인식용 DB ETRI 일반 전화 화자 인식용 DB ETRI 증가 마이크 화자 인식용 DB
음성 특징 벡터	19차 MFCC + 1차 미분계수(delta) 총 38차 특징
채널	휴대폰, 일반 전화, PC 마이크
오디오 채널 수	1 (mono)
샘플링 레이트	8kHz
비트 레이트	16bits/sample
화자 수	휴대폰 : 한국인 남녀 256명 일반 전화 : 한국인 남녀 256명 PC 마이크 : 한국인 남녀 250명
음성 신호 보정 알고리즘	평균 분산 정규화(MVN)
GMM 학습	EM 알고리즘
가우시안 혼합 수	1024
발화 형태	낭독체
총 학습 데이터 길이	500시간 이상

용의자 모델 관리 모듈은 용의자 모델을 생성하는 모듈이다. 이 모듈은 먼저 용의자의 음성에서 특징을 추출한다. 그리고 특징에 평균 분산 정규화 기법을 사용한다. 이 때 사용자가 특징 강화 기법을 사용하고자 한다면 UBM 학습에서 추정된 특징 강화 기저로 특징을 강화한다. 그리고 앞서 생성된 2개의 UBM 중에서 특징 강화 기법 적용 여부에 따라 하나를 택하여 용의자 음성에서 추출한 특징을 MAP 적용하여 모델을 학습한다.

레퍼런스 모델 관리 모듈은 cohort model 기법에 사용할 레퍼런스 모델을 생성하는 모듈이다. 이 모듈은 먼저 레퍼런스 모델로 등록하고자 하는 레퍼런스 화자의 음성에서 특징을 추출한다. 용의자 모델 관리 모듈과 마찬가지로 음성 신호 보정 알고리즘을 사용하고 특징 강화는 선택적으로 사용한다. 그리고 UBM에 음성 특징을 MAP 적용하여 레퍼런스 모델을 학습한다.

화자 식별 모듈은 테스트 음성의 특징을 용의자 모델 및 레퍼런스 모델과 비교하여 유사도비(LR)를 구하는 모듈이다. 이 모듈은 먼저 테스트 음성에서 특징을 추출한다. 다음으로 추출된 특징에 평균 분산 정규화 기법을 사용하고 특징 강화는 선택적으로 사용한다. 그리고 음성 신호 보정 알고리즘과 특징 강화 기법이 사용된 음성 특징을 용의자 모델 및 레퍼런스 모델과 비교하여 유사도비를 계산한다.

### 3.2 시스템의 구현

SPO Verifier의 개발 환경은 <표 2>와 같다.

표 2. SPO Verifier의 개발 환경  
Table 2. Development environment of SPO Verifier

항목	내용
운영체제	Windows 7
개발언어	Python 버전 2.6.6
	C
개발도구	Eclipse
	Visual Studio 2010
Python 버전	2.6.6

시스템의 사용자 인터페이스는 Python 2.6.6과 Python 2.6용 tkSnack 2.2를 사용하여 개발하였다. 프로그램을 실행하고 화자 인식을 위해 사건을 생성하면 <그림 2>와 같은 화면이 나타난다.

용의자 모델 메뉴에서 사용자는 용의자 모델을 추가할 수 있다. 그리고 용의자 모델 학습 데이터 메뉴에서 모델 생성에 사용할 음성 파일, 즉 학습 데이터를 추가하고 학습할 수 있다. 음성 파일을 추가하면 MFCC 특징을 추출하고 UBM에 MAP 적용을 하여 화자 모델을 생성한다. MFCC 특징을 추출하는 라이브러리와 UBM에 음성 특징을 MAP 적용하는 라이브러리는 C언어로 작성되었다. 그리고 여러 개의 학습 데이터를 일괄적으로 추가하고 파일명으로 용의자 모델을 생성하는 기능을 구현하였다. 또한 사용자는 이미 생성된 용의자 모델을 삭제하거나 추가된 학습 데이터를 삭제할 수 있다. 용의자 모델 학습 데이터 메뉴에서 음성 파일을 선택하면 학습 데이터 정보 메뉴에 파일명, 채널 수, 비트 레이트, 샘플링 레이트, 현재 길이, 순수 발생 길이, 신호 대 잡음비(signal to noise ratio, SNR), 평균 피치 정보가 표시되는 기능을 구현하였다. 그리고 음성 파일을 재생하는 기능도 구현하였다.

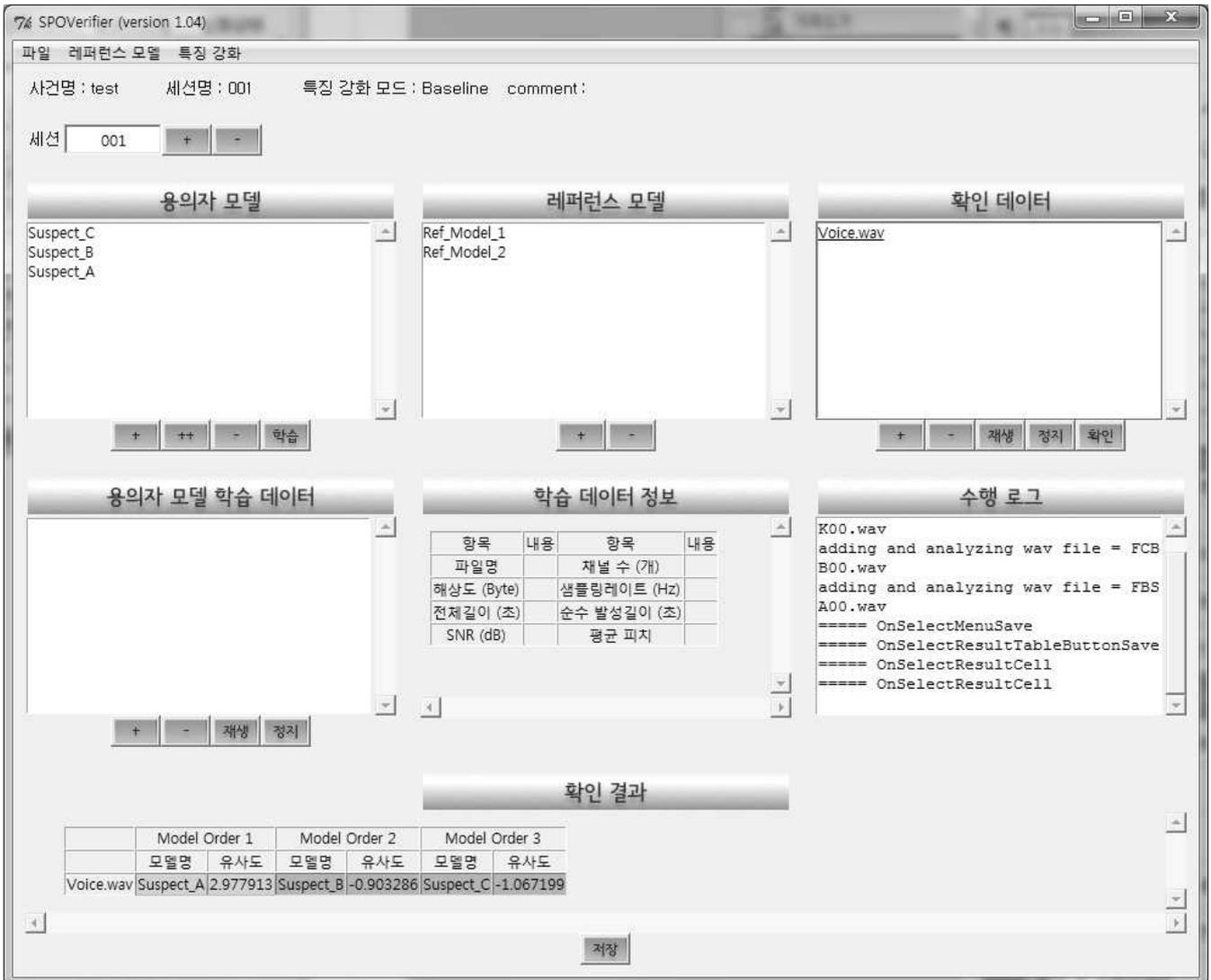


그림 2. SPO Verifier의 화자 인식 화면  
Figure 2. Speaker identification screen of SPO Verifier

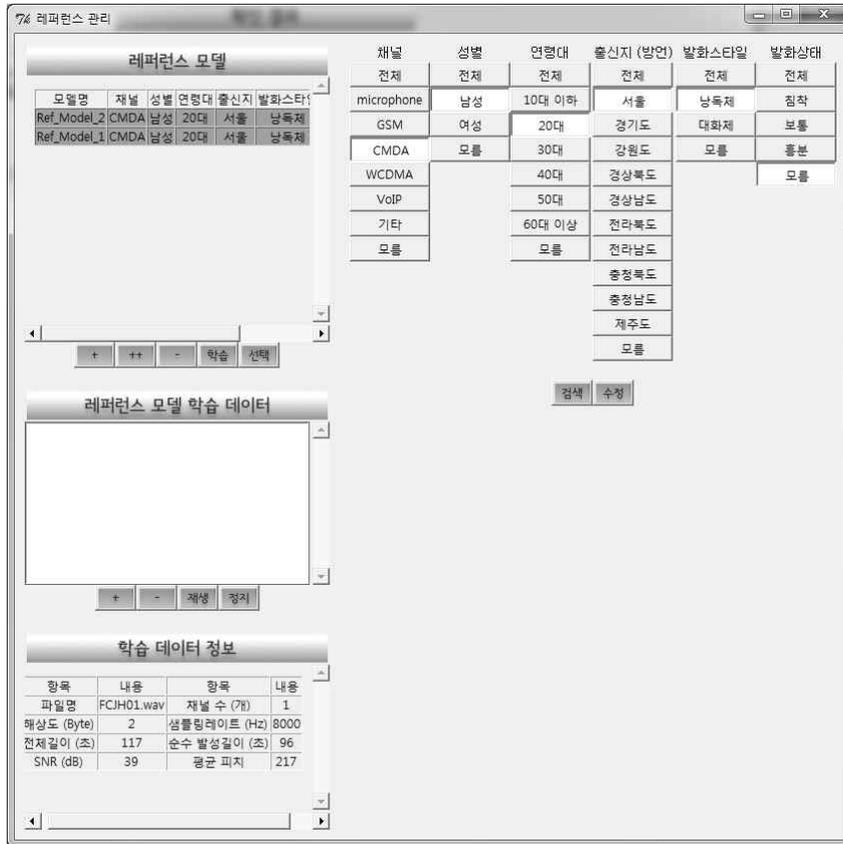


그림 3. SPO Verifier의 레퍼런스 모델 관리 화면  
Figure 3. Reference model management screen of SPO Verifier

레퍼런스 모델 관리 작업은 <그림 3>과 같은 레퍼런스 모델 관리 메뉴에서 수행된다.

레퍼런스 모델 메뉴에서 사용자는 성별, 연령대, 출신 지역, 발화 형태, 감정 상태를 지정하여 레퍼런스 모델을 관리할 수 있다. 용의자 모델 메뉴와 마찬가지로 모델을 추가하고 학습하는 기능을 갖고 있다. 그리고 여러 개의 학습 데이터를 선택하고 각각의 파일명으로 모델들을 일괄 추가하는 기능과 추가된 모델을 삭제하는 기능도 구현되어 있다. 사용자는 생성된 레퍼런스 모델들 중에서 식별에 사용하고자 하는 모델들을 선택하여 식별에 사용할 수 있다.

<그림 2>의 확인 데이터 메뉴에서 사용자는 테스트 음성을 추가하거나 추가된 테스트 음성을 목록에서 제거할 수 있다. 그리고 확인 데이터 메뉴는 용의자 모델 및 레퍼런스 모델과 테스트 음성을 비교하여 유사도비를 계산하는 기능을 갖고 있다. 이 때 가우시안 혼합 모델 디코딩(GMM decoding)을 수행하는 라이브러리는 C언어로 작성되었다. 확인 결과는 <그림 2>의 화자 인식 화면 아래 쪽에 각 용의자 모델과 테스트 음성의 유사도비가 높은 순서대로 사용자에게 제시된다.

본 시스템은 사용자가 특징 강화 사용 여부를 선택하는 기능을 구현하였다. 사용자가 특징 강화 모드를 설정하면 비교 작업을 수행할 때 특징 강화 기법이 적용된 음성 특징과 UBM을 사용한다.

#### 4. 성능 평가 실험

본 시스템의 성능을 평가하기 위해 현재 상용화된 자동화자식별시스템 중 하나인 Agnitio사의 BATVOX(ver 3.1)[9]와 성능을 비교하였다. BATVOX는 현재 미국, 독일, 영국, 스페인, 멕시코 등의 수사 기관에서 범죄자 음성 식별 업무에 사용 중인 시스템으로서 미국국립표준기술연구소(NIST)로부터 성능을 검증받았다.

본 평가 실험에서 성능 평가의 기준으로는 동일 오류율(equal error rate, EER)을 사용하였다. 동일 오류율은 화자 인증에서 사용하는 기준으로 등록된 화자(true speaker)를 사칭자(imposter)로 판정하는 오류 발생률인 오거부율(false rejection rate, FRR)과 사칭자를 등록된 화자로 판정하는 오류 발생률인 오수락율(false alarm rate, FAR)이 같아질 때의 최소 오류율이다. 그리고 본 실험은 남녀 화자를 구분하여 진행하였다.

##### 4.1 실험 설계

###### 4.1.1 실험 DB

본 연구에서 개발한 시스템의 성능을 평가하기 위해 음성 데이터베이스를 수집하였다. 수집한 음성 DB는 <표 3>과 같다.

표 3. 본 연구에서 수집한 음성 DB  
Table 3. Collected speech database in this research

항목	내용
채널	휴대전화
오디오 채널 수	1 (mono)
샘플링 레이트	8kHz
비트 레이트	16bits/sample
화자 수	한국인 남녀 101명 (남자 50명, 여자 51명)
발화 형태	낭독체
화자별 음성 길이	약 3분 (약 1분 길이의 음성 3개로 구성)

문장 독립적인 실험을 하기 위하여 발성 내용은 2,458개의 문장 중 1분 이상의 발성이 가능하도록 화자에 따라 임의로 여러 문장을 선택하여 만든 실험자료를 사용하였다. 사용한 실험자료의 예는 <표 4>와 같다.

표 4. 수집한 음성 DB의 실험자료 예  
Table 4. Examples of collected speech corpus

가장 귀중한 사랑의 가치는 희생과 헌신이다.
경험을 바탕으로 지혜의 답을 쌓아 가도록 하자.
군자는 스스로 재능이 없음을 근심하며 남이 알아주지 않음을 근심하지 않는다.
기쁘게 일하고 해 놓은 일을 기뻐하는 사람은 행복하다.
나태는 자기의 실패뿐 아니라 타인의 성공에 의해서도 벌을 받는다.
대가의 기대를 행위의 결과로 생각하는 사람이 되지 말라.

수집한 전체 음성 데이터베이스 중 남녀 81명(남자 36명, 여자 45명)의 데이터를 성능 평가에 사용하였다. 그리고 이 중에서 남녀 41명(남자 18명, 여자 23명)의 데이터를 학습과 테스트에 사용하였고 나머지 40명(남자 18명, 여자 22명)의 데이터는 레퍼런스 모델 학습에 사용하였다.

4.1.2 학습 데이터 및 테스트 데이터

본 연구에서 수집한 음성 데이터베이스는 화자별로 약 1분 길이의 음성 3개로 구성되어있다. SPO Verifier는 3개의 발성 중 첫 번째 발성을 UBM에 MAP 적용하여 모델을 학습하였고 나머지 2개의 발성을 테스트에 사용하였다. 그리고 BATVOX도 동일하게 첫 번째 발성을 모델 학습에 사용하고 나머지 2개의 발성을 테스트에 사용하였다.

4.2 실험 결과

4.2.1 BATVOX(ver 3.1) 실험 결과

BATVOX는 매뉴얼 상에서 자체적인 음성 신호 보정 알고

리즘을 사용한다고 밝히고 있고 특징 강화에 대해서는 언급을 하지 않고 있다[10]. 앞서 언급한 한국인 남녀 41명에 대한 BATVOX(ver 3.1) 실험 결과는 <표 5>와 같다. 표에서 FAR는 오거부율, FRR은 오수락율, EER은 동일 오류율이다.

표 5. BATVOX(ver 3.1) 실험 결과  
Table 5. Experimental results using BATVOX(ver 3.1)

평균 EER	남자			여자		
	FAR	FRR	EER	FAR	FRR	EER
4.41%	0.00%	11.11%	5.56%	0.00%	6.52%	3.26%

실험 결과 남자 음성에 대한 동일 오류율은 5.56%, 여자 음성에 대한 동일 오류율은 3.26%로 평균 4.41%의 동일 오류율을 보였다.

4.2.2 SPO Verifier 실험 결과

SPO Verifier를 사용한 실험은 특징 강화 기법을 사용하지 않은 실험과 특징 강화 기법으로 주성분 분석(PCA)을 사용한 실험, 두 가지로 진행하였다. 그리고 MFCC 특징의 평균과 분산을 정규화하는 평균 분산 정규화 기법을 사용하였다.

특징 강화를 사용하지 않은 실험 결과는 <표 6>과 같다.

표 6. SPO Verifier 실험 결과  
Table 6. Experimental results using SPO Verifier

평균 EER	남자			여자		
	FAR	FRR	EER	FAR	FRR	EER
0.82%	0.40%	2.78%	1.59%	0.10%	0.0%	0.05%

실험 결과 남자 음성에 대한 동일 오류율은 1.59%, 여자 음성에 대한 동일 오류율은 0.05%로 평균 0.82%의 동일 오류율을 보였다. 이는 BATVOX보다 3.59% 낮은 수치이다.

다음으로 주성분 분석을 사용한 실험 결과는 <표 7>과 같다.

표 7. 주성분 분석을 사용한 SPO Verifier 실험 결과  
Table 7. Experimental results using SPO Verifier and the feature enhancement(PCA)

평균 EER	남자			여자		
	FAR	FRR	EER	FAR	FRR	EER
1.40%	2.54%	0.00%	1.27%	0.89%	2.17%	1.53%

실험 결과 남자 음성에 대한 동일 오류율은 1.27%, 여자 음성에 대한 동일 오류율은 1.53%로 평균 1.40%의 동일 오류율을 보였다. 이는 BATVOX보다 3.01% 낮은 수치이다. 특징 강화를 사용한 결과가 특징 강화를 사용하지 않은 결과보다 낮은 성능을 보이는데 이는 본 실험에서 사용한 DB가 잡음이 없는 음성이기 때문으로 추정된다. DB에 잡음을 추가한 실험

을 진행하려 했으나 모든 음성 파일에 일정한 잡음을 추가했을 경우 일부 음성 파일이 BATVOX에서 요구하는 최소 신호대 잡음비를 달성하지 못해 실험이 불가능하였다.

### 5. 결론

본 연구에서 과학수사를 위한 한국인 음성 특화 자동화자식별시스템을 개발하였다. 개발한 시스템의 성능을 평가하기 위해 한국인 남녀 41명의 음성을 사용하여 기존의 상용화된 시스템과 성능을 비교하였다. 성능 평가 실험 결과 기존의 상용화된 시스템보다 3.59% 낮은 동일 오류율을 보여 한국인 음성에 대해 더 높은 성능을 보이는 것을 확인하였다.

추후 SPO Verifier의 화자 인식 성능 검증을 위해 미국국립 표준기술연구소나 한국전자통신연구원(ETRI)과 같은 공신력 있는 기관에 의뢰하고 성능 검증 이후 대검찰청의 음성분석 업무에 활용할 예정이다.

### 참고문헌

[1] Huang, X., Acero, A. & Hon, H. W. (2001). *Spoken language processing: a guide to theory algorithm, and system development*. NJ: Prentice Hall PTR.

[2] Atal, B. S. (1974). Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification, *Journal of the Acoustical Society of America*. Vol. 55, No. 6, 1304-1312.

[3] Viikki, O. & Laurila, K. (1998). Cepstral domain segmental feature vector normalization for noise robust speech recognition, *Speech Communication*. Vol. 25, 133-147.

[4] Smith, L. I. (2002). *A tutorial on principal components analysis*. [http://www.cs.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf).

[5] Reynolds, D. A. & Rose, R. C. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models, *IEEE Trans. Speech Audio Processing*. Vol. 3, 72-83.

[6] Gauvain, J. L. & Lee, C. H. (1994). Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains, *IEEE Trans. Speech Audio Processing*. Vol. 2, 291-298.

[7] Reynolds, D. A., Quatieri, T. F. & Dunn, R. B. (2000). Speaker verification using adapted Gaussian mixture models, *Digital Signal Processing*. Vol. 10, 19-41.

[8] Rosenberg, A. E., DeLong, J., Lee, C. H., Juang, B. H. & Soong, F. K. (1992). The use of cohort normalized scores for

speaker verification, *International Conference on Speech and Language Processing*, 599-602.

[9] BATVOX webpage. [http://www.agnitio-corp.com/producto.php?id\\_producto=2](http://www.agnitio-corp.com/producto.php?id_producto=2).

[10] BATVOX user manual (2005). Agnitio.

• **김경화 (Kim, Kyung Wha)**

대검찰청 과학수사담당관실  
서울시 서초구 반포대로 157  
Tel: 02-3480-2150  
Email: savoix@spo.go.kr  
관심분야: 법음성학, 화자 식별

• **소병민 (So, Byung-Min)**

대검찰청 과학수사담당관실  
서울시 서초구 반포대로 157  
Tel: 02-3480-2150  
Email: sbm1210@spo.go.kr  
관심분야: 법음성학, 화자 식별

• **유하진 (Yu, Ha-Jin)**

서울시립대학교 컴퓨터과학부  
서울시 동대문구 서울시립대로 163  
Tel: 02-2210-5322  
Email: hjyu@uos.ac.kr  
관심분야: 화자 인식, 음성 인식