DSP보드를 이용한 전화음성용 실시간 화자인증 시스템의 구현에 관한 연구*

이현승(대진대), 최홍섭(대진대)

<차 례>

<Abstract>

An Implementation of Real-Time Speaker Verification System on Telephone Voices Using DSP Board

Hyeon Seung Lee, Hong Sub Choi

This paper is aiming at implementation of real-time speaker verification system using DSP board. Dialog/4, which is based on microprocessor and DSP processor, is selected to easily control telephone signals and to process audio/voice signals. Speaker verification system performs signal processing and feature extraction after receiving voice and its ID. Then through computing the likelihood ratio of claimed speaker model to the background model, it makes real-time decision on acceptance or rejection. For the verification experiments, total 15 speaker models and 6 background models are adopted. The experimental results show that verification accuracy rates are 99.5% for using telephone speech-based speaker models.

* Keywords : Speaker verification system, Dialog/4, GMM Speaker model

* 이 논문은 2003학년도 대진대학교 학술연구비지원에 의한 것임.

1. 서 론

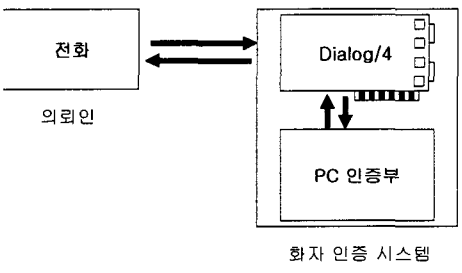
음성신호에는 화자의 음향적 특징들이 포함되어 있으며 이러한 화자간의 음향적 특징의 변이를 이용하여 발성한 사람의 ID를 확인하는 시스템을 화자인증 시스템이라 하며 신원확인 및 출입통제 시스템에 사용하고 있다. 오늘날 전화와 컴퓨터를 결합해 주는 CTI (Computer Telephony Integration) 기술의 발전으로 전화와 컴퓨터 사이에 정보를 교환하는 것이 용이하게 되었다.

본 논문은 이러한 CTI 기술이 적용된 DSP 보드를 이용하여 전화망을 통하여 의뢰인의 ID와 음성을 입력받아 인증 관련 처리를 컴퓨터에서 수행한 후 의뢰인에게 전화로 그 인증 결과를 다시 알려주는 화자인증 시스템의 실시간 구현에 대하여 기술하였다. 논문의 중심은 기존 이론을 응용한 시스템 전체의 실시간 구현에 두었으며, 실험에 사용한 DSP 보드는 Dialogic 사에서 만든 CTI 전용 DSP 보드인 Dialog/4를 이용하였다. 논문의 구성은 1장 서론에 이어, 2장에서 dialog/4 보드의 특징과 화자인증 시스템 구성에 관한 전반적인 내용을 소개하고, 3장에서 화자인증 시스템에 필요한 음성DB의 구성방법 및 내용에 대하여 설명하였으며, 4장에서는 구성된 화자인증 시스템의 성능평가 결과를 제시하고 마지막 5장에서 결과를 고찰하였다.

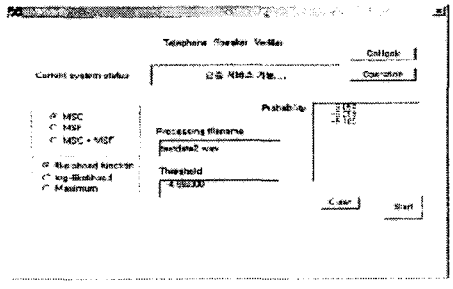
2. 화자인증 시스템의 설계 및 구현

2.1. Dialog/4를 이용한 화자인증 시스템의 개요

논문에서 구현한 화자인증 시스템은 dialog/4 보드를 이용하여 전화기반에서 의뢰인의 ID와 음성을 입력받은 후 인증관련 처리를 PC상에서 수행하여 그 인증결과를 의뢰인의 전화로 다시 알려주는 시스템이다. <그림 1>은 시스템의 개략적인 구성도 이다.



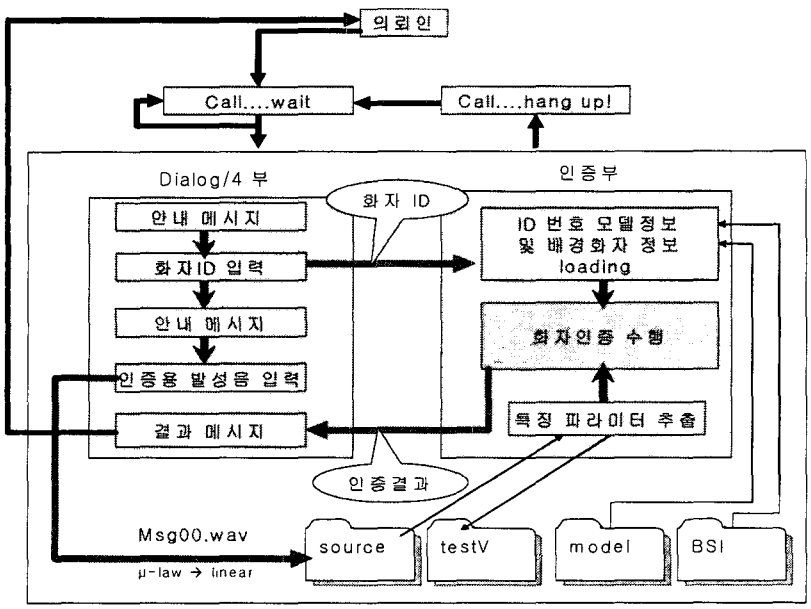
<그림 1> 화자인증 시스템의 구성도



<그림 2> 프로그램 실행 화면

<그림 2>는 프로그램의 실행 화면을 보여준다. 프로그램을 시작하면 인증 시스템의 상태와 정보들을 표시하는 창이 나타나고 대기한다. 실행(operation) 버튼을 누르면 화자인증 서비스가 가능하게 된다. 맨 위쪽의 창은 시스템의 현재 처리 상황이나 상태를 표시하며, 왼쪽의 라디오 버튼들은 배경화자 선정방법과 화자와 등록모델 사이의 유사도 계산 방법을 선택할 수 있는 버튼들이다. 가운데의 창들은 현재 처리중인 파일이름과 현재 시스템에 설정된 문턱 값이 표시되며, 오른쪽의 창은 결과로 나온 유사도 값을 표시하는 부분이다.

<그림 3>은 화자인증 시스템의 처리과정을 보여주는 블록 다이어그램으로 프로그램 내부에서 실제로 수행되는 내용을 보여준다. 프로그램을 시작하면 우선 dialog/4 보드를 초기화하고 콜 관련 메시지들을 처리하도록 설정한다. 프로그램은 내부적으로 크게 두 가지로 구성된다. 하나는 입출력 아날로그 전화망과 직접 연결되어 있는 dialog/4 보드를 제어하는 부분으로서, 보드상의 DSP 프로세서와 일반용 마이크로프로세서의 기능을 효율적으로 이용할 수 있는 C 언어 응용프로그램 인터페이스(API) 함수와 구조체 들을 사용하여 DTMF와 음성신호 등의 모든 전화 관련 신호음들에 대한 신호처리를 수행한다. 두 번째는 PC에 구현된 인증부 로서 전화를 통해 입력된 ID와 음성 데이터를 저장한 후 신호처리를 통해 음성의 특징 벡터를 추출한 후 ID에 해당하는 화자모델과 배경화자 정보를 이용하여 유사도(likelihood ratio)를 계산하여 실시간으로 인증 또는 거절을 알려주는 부분이다.



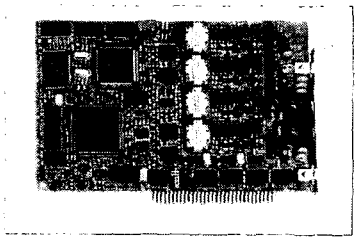
<그림 3> 화자 인증 시스템의 처리과정

dialog/4 보드를 통제하는 부분은 전화벨을 감지하고, 상황에 알맞은 안내문을 읽어주고 인증관련 처리가 끝나면 화자인증 결과를 전송한 후 자동으로 전화를 끊는 일련의 처리 기능을 수행한다. 또한 전화를 통해 의뢰인의 ID와 음성을 시스템의 입력으로 받으면, 의뢰인의 ID는 해당 화자의 배경화자 정보를 참조하기 위해 필요하므로 인증부로 넘겨주고, 의뢰인의 음성은 발성과 동시에 실시간으로 8kHz로 샘플링 되어 μ-law PCM 웨이브 파일 형태로 메모리에 저장한다.

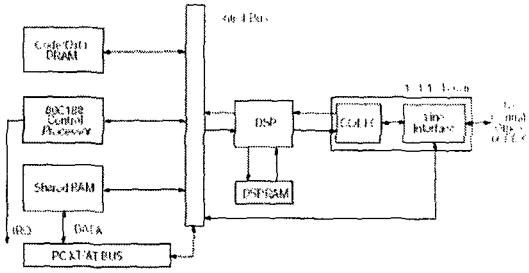
화자의 ID 및 음성파일 저장이 완료되면 인증부는 소스 폴더에 저장된 의뢰인의 음성파일을 읽어와 μ-law PCM으로 저장된 음성을 선형 PCM으로 바꾼 다음, 끝점검출[1]을 이용하여 음성구간을 찾은 후, 필터 H(z) = 1-0.95z-1를 이용하여 프리엠퍼시스 한다. 160샘플(20ms, 8kHz)의 음성신호를 한 프레임으로 하고 80샘플(10ms, 8kHz)마다 중첩하여 각 프레임에 해밍(hamming)창으로 가중치를 주어 매 10ms 마다 특징 파라미터인 12차 MFCC를 추출하며 이를 testV 폴더에 저장한다. 앞의 과정과 동시에 등록화자들의 모델들이 메모리에 적재되며 dialog/4 에서 입력받은 ID에 해당하는 화자모델의 배경화자 정보를 참조하여 처리할 배경화자모델을 결정하는 작업이 수행된다. 두 과정이 완료되면 내부의 인증루틴에서 본인의 모델과 배경화자 모델들이 입력음성을 발생시킬 확률을 계산하여 유사도(likelihood ratio)를 구한 후 산출된 유사도 값을 문턱 값(threshold)과 비교하여 수락(acceptance)할 것인지 거절(rejection)할 것인지 판단하여 해당 안내 메시지를 의뢰인에게 알려준다.

2.2. Dialog/4 보드의 특성

Dialog/4 보드는 Intel® Dialogic®의 제품으로 ISA 슬롯을 통해 PC와 인터페이스 되며, 아날로그 라인을 직접적으로 연결할 수 있는 2개의 RJ-11 포트를 가지고, 포트마다 2개의 채널을 사용할 수 있도록 되어 있다. 하나의 PC에 16개의 보드까지 확장 가능하며, MS-DOS, Windows 95, Windows NT/2000, OS/2 및 UNIX 기반 하에서 운영될 수 있는 C언어 응용 프로그램 인터페이스(API)를 제공한다. 음성데이터 저장 시 각 채널당 24kb/s 에서 64kb/s 까지의 데이터 속도를 선택할 수 있으며, 음성 포맷으로 6kHz와 8kHz의 샘플링 율로 ADPCM 과 μ-law PCM포맷을 지원한다. 또한 펌웨어(firmware) 업그레이드만으로 버그개선 및 성능향상이 가능하다. 실험에 사용된 dialog/4 드라이버는 현재 가장 최신 버전인 System Release 5.1.1이며 <그림 4>는 보드의 전체 모습이다.



<그림 4> Dialog/4 보드의 외형



<그림 5> dialog/4 보드의 기능적 블록도

<그림 5>는 dialog/4 보드의 기능적 블록도를 보여준다. dialog/4 보드는 신호처리를 수행하는 Motorola 56001(33MHz) DSP 프로세서와 의사결정과 데이터 전송 기능을 수행하는 일반적인 목적의 제어용 마이크로프로세서인 80C188이 결합된 독특한 이중프로세서 구조를 사용한다. 이런 이중프로세서 구조는 호스트 PC가 많은 저수준(low-level)의 의사결정을 하는 부담을 덜어주어 더욱 강력한 응용프로그램의 개발을 가능하게 한다. 드라이버 설치 후 보드를 활성화하면 호스트 PC로부터 보드상의 *code/data RAM*과 *DSP RAM*상에 보드의 모든 운영 및 제어를 위한 기본 환경을 담고 있는 펌웨어를 적재한다.

마이크로프로세서는 로컬버스를 통하여 dialog/4보드의 모든 운영을 제어하며, 호스트 PC로부터의 명령을 번역하고 실행한다. 마이크로프로세서는 실시간 이벤트를 다룰 수 있게 하며, 더욱 빠른 시스템 응답속도를 위하여 호스트 PC의 데이터 흐름을 운영하고 호스트 PC의 처리부담을 감소시키며, DTMF와 전화 제어신호를 처리하며, 걸려오는 전화로부터 DSP를 자유롭게 하는 역할을 한다. 프로세서와 호스트 PC 사이의 통신은 효율적인 파일전송을 위하여 입출력 버퍼로서 할당된 RAM을 사용한다.

DSP 프로세서는 들어오고 나가는 데이터에 대해서 신호분석 및 처리와 오디오 신호의 가변레벨에 대한 자동이득제어(AGC)를, 그리고 디지털 음성을 압축하기 위해서 ADPCM 또는 μ-law PCM 알고리즘을 수행한다. 또한 DTMF와 MF 등의 톤 신호 및 묵음구간을 검출하며, 오디오 데이터 재생을 위하여 압축을 해제하고, 응용프로그램이나 사용자의 요구에 알맞도록 볼륨과 재생속도를 조정하며, DTMF나 MF 등의 톤 신호를 발생시킨다.

아래의 <표 1>은 화자인증 시스템에 사용된 구조체와 C언어 API 함수들의 목록을 정리하였다.

<표 1> 화자인증 시스템에 사용된 구조체 및 함수들의 요약

구조체명	내 용
DV_TPT	Termination Parameter Table
DV_DIGIT	User Digit Buffer Structure
DX_EBLK	Event Block Structure
DX_IOTT	I/O Transfer Table Structure
DX_XPB	I/O Transfer Parameter Block

함 수 명	내 용
dx_open( )	open a board or channel
dx_close( )	close a board or channel
dx_sethook( )	set hook switch state
dx_wtring( )	wait for number of rings
dx_getevt( )	get call status transition event
dx_setevtmsk( )	set Speed/Volume Modification Table
dx_clrtpt( )	clear DV_TPT structure
dx_getdig( )	get digits from channel digit buffer
dx_clrdigbuf( )	clear the firmware digit buffer
dx_reciottdata( )	records voice data to multiple destinations
dx_playf( )	play voice data from a single file
dx_fileopen( )	opens the file specified by file pointer
dx_fileclose( )	closes the file associated with the handle

2.3. GMM 화자 모델

GMM은 통계적 방법으로 화자의 음성을 모델링 하는 방법으로 화자인식에서 많이 사용하는 방법이다[4][5]. 특정 화자가 인증을 위해 발성한 음성신호로부터 얻어낸 특징 벡터들이 지정된 화자의 GMM모델에 의해 발생될 확률을 계산하여 이를 문턱값과 비교하여 화자인증을 하는 방식이다. 화자 s에 대한 GMM 화자모델 λ_s는 다음 식(1)으로 표현된다.

$$\lambda_s = \omega_i^s, \mu_i^s, \Sigma_i^s, i = 1, \ldots, M \quad (1)$$

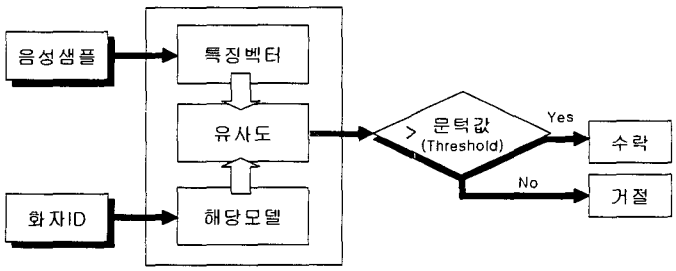
여기서, M은 가우시안 확률분포의 개수, 즉 믹스처의 크기를 의미하고, w_i은 i

번째 가우시안 믹스처의 가중치가 된다. 또, μ_i^s과 Σ_i^s은 각각 가우시안의 평균과 분산이다.

2.4. 배경화자 선정방법

배경화자를 사용하는 목적은 본인과 비슷하지만 본인이 아닌 사람들의 정보를 반영함으로써 화자인증 시스템의 성능을 향상시키기 위함이다. 배경화자를 이용할 때, 두 가지 문제는 배경화자를 생성하는 방법과 배경화자의 수를 결정하는 것이다. 직관적으로 볼 때 배경화자(background speaker)는 일반적인 응용에서 명확히 예상되는 사칭자 들의 집합으로 구성될 것이다. 일반적인 화자인증 시스템에서는 검증하고자 하는 화자모델과 그의 배경화자모델을 이용하여 유사도의 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 배경화자모델에 포함하는 MSF (Maximally Spread Far) 방법이 있다[2].

2.5 화자인증 시스템의 판정기준



<그림 6> 화자 인증 시스템의 구성

GMM을 이용한 화자인증 시스템의 구성은 <그림 6>과 같다. 특징 추출이 완료되면 화자인증 시스템은 의뢰인의 승낙/거절을 결정하기 위해 발성음의 유사도(likelihood ratio)를 계산한다. 발성음과 의뢰인 모델 사이의 유사도는 식(2)와 같이 나타낼 수 있다[2].

$$L(X) = \frac{p(\lambda_c \mid X)}{p(\lambda_{\bar{c}} \mid X)} \quad (2)$$

식(2)에 Bayes정리를 적용한 후 로그를 취하면 식(3)이 나온다.

$$\Lambda(X) = \log p(X \mid \lambda_c) - \log p(X \mid \lambda_{\bar{c}}) \quad (3)$$

$p(X \mid \lambda_c)$는 의뢰인 모델에서 발성음이 생성될 확률이고, $p(X \mid \lambda_{\bar{c}})$는 배경화자 모델에서 발성음이 생성될 확률이다. 화자인증 시스템에서 사용되는 문턱값(threshold)은 시스템의 환경에 따라 달라지므로 절대적인 값으로 결정할 수는 없다. 논문에서는 판정기준으로 T_{EER} 값을 사용한다. 화자인증 시스템은 식(3)를 계산하여 유사도 $\Lambda(X)$를 구한 후 그 값을 시스템의 문턱값 T_{EER}과 비교하여 $\Lambda(X) > T_{EER}$ 이면 수락(acceptance) 메시지를, $\Lambda(X) < T_{EER}$ 이면 거절(rejection) 메시지를 의뢰인에게 알려준다. 이 최종 과정이 끝나면 프로그램은 전화를 끊은 후 다음 인증을 위해 시스템 및 필요한 변수들을 초기화하고 다시 인증가능 상태로 대기하게 된다.

3. 성능평가를 위한 음성DB의 구성

음성DB를 이용하여 모든 화자의 GMM모델을 만들어 모델폴더에 저장하고, 화자모델들 사이의 거리들을 계산하여 알맞은 배경화자 정보를 구하여 BSI 폴더에 저장한다. 음성DB는 15명의 화자에 대하여 크게 두 가지 형태로 수집하였다. 하나는 PC기반에서 고급 마이크를 사용하여 녹음된 깨끗한 음성샘플이며 나머지는 직접 시스템에 전화를 걸어 입력받은 전화음성 샘플이다. 발화내용은 문장과 숫자로 구성되어 있으며 전화음성 샘플은 문장만을 사용하였다. 조용한 환경의 연구실에서 녹음되었으며 마이크는 SHURE사의 VR250BT를 사용하였다. VR250BT는 컴퓨터 음성인식 및 인터넷 폰 사용을 위한 전용 헤드셋 마이크로서 주파수 응답은 100-10,000Hz이며, 단일 지향성의 컨덴서 마이크이다. 음성DB의 수집환경을 <표 2>에 정리하였다.

<표 2> 음성 DB 수집 환경

발화내용	발성횟수	채 널	녹음환경	녹음방법
문장	10번	조용한 환경	연구실	고성능 마이크
숫자	10번	조용한 환경	연구실	고성능 마이크
문장	10번	전화	연구실	휴대전화(핸드폰)

각각의 발화내용은 <표 3>에 있다. 문장과 숫자 모두 약 15초 정도의 길이이며 숫자의 경우 20-99 사이에서 중복을 허락하여 무작위로 선정하였다. 문장의 경우 같은 내용을 10번 발음함으로써 10개의 샘플을 취하였고, 숫자는 서로 다른 10개의 내용을 한번씩 발성함으로써 10개의 샘플을 취하였다.

<표 3> 음성 DB 발화내용

문 장
안녕하십니까 여기는 대진대학교 전자공학과 멀티미디어 통신연구실 입니다. 화자 인증기의 성능평가를 위한 검증용 음성을 녹음하는 중입니다. 이 문장은 약 15초간 녹음됩니다.

숫 자
1. 스물일곱- 여든 둘- 서른아홉- 예순 하나- 일흔 다섯- 쉰 셋 예순 둘- 아흔 하나- 마흔 아홉- 일흔 다섯- 마흔 셋- 서른 넷
2. 스물여섯- 여든 하나- 쉰 셋- 서른다섯- 마흔 셋- 아흔 둘 마흔 여섯- 일흔 하나- 아흔 다섯- 일흔 넷- 쉰일곱- 예순 아홉
⋮
10. 서른 넷- 예순 셋- 일흔 여섯- 마흔 하나- 스물일곱- 예순 둘 여든 일곱- 마흔 여섯- 쉰 둘- 아흔 하나- 스물다섯- 일흔 아홉

또한 기본 인증 성능을 평가할 때 인증시스템의 객관성을 확보하기 위해 YOHO 표준 DB에 대한 실험을 수행하였다. YOHO DB의 특징은 <표 4>와 같다.

<표 4> YOHO DB의 구성

총화자 수	발성화자	채 널	녹음환경	녹음방법
138명	등록 세션 4개 테스트 세션 10개	조용한 환경	사무실	고성능 마이크

4. 성능 평가 및 결과

화자인증 시스템의 성능은 15명의 화자를 대상으로 수집한 음성과 YOHO 음성 DB를 이용하여 실험하였으며, 기본 인증 성능과 전화 기반에서의 성능으로 구분하여 평가하였다. 음성의 특징벡터로 12차 MFCC를 사용하였으며, 화자모델을 위해 사용된 GMM의 차수는 32차이다[3]. 배경화자의 수는 6명으로 하였으며, 배경화자 구성 방법으로는 MSC만 사용한 경우와 MSC와 MSF를 모두 사용한 경우에 대하여 평가하였으나 두 방법의 결과가 뚜렷한 차이를 나타내지 않아 MSC 방법의 결과만을 제시하였다. 실험결과는 EER을 백분율로 표현하고, 화자인증 시스템의 인증률은 (100-EER)%로 계산하였으며, 시스템에 사용된 실제 문턱값 T_{EER} 을 함께 표시하였다.

4.1. 인증부 자체 성능

이 실험은 화자인증 시스템의 기본 성능을 확인하기 위한 실험이다. PC에서 고성능 마이크를 사용하여 입력받은 깨끗한 음성으로 화자모델을 구한 후 같은 환경에서 녹음된 테스트 음성으로 화자인증을 하는 경우의 성능평가이다. <표 5>는 등록용 음성으로는 문장을 발성한 각각의 샘플을 모두 더한 약 2분 40초 길이의 음성을 이용하고, 테스트 음성으로는 숫자를 발성한 15초 길이의 샘플 10개를 사용하여 인증 실험한 결과이다.

<표 5> 문장 모델에 대한 숫자 테스트 발성음 적용 결과

인증률(%)	EER(%)	T_{EER}
92.62	7.38	-4.34

<표 6>은 등록용 음성으로는 숫자를 발성한 각각의 샘플을 모두 더한 약 2분

40초 길이의 음성을 이용하고, 테스트 음성으로는 문장을 발성한 15초 길이의 샘플 10개를 사용하여 인증실험한 결과이다.

<표 6> 숫자 모델에 대한 문장 테스트 발성음 적용 결과

인증률(%)	EER(%)	TEER
95.92	4.08	-5.14

YOHO 표준 DB를 이용한 실험은 직접 구성한 DB와 비슷한 환경 하에서의 비교를 위해 화자를 15명으로 제한하여 실험하였다. 등록용 음성으로 화자마다 4개의 등록 세션 중 한 개 세션의 24개 파일을 합친 약 1분 40초 길이의 음성을 사용하며, 테스트 음성으로는 10개의 테스트 세션 각각의 4개 파일을 합친 약 16초 길이의 샘플 10개를 사용하였다.

<표 7> YOHO DB 결과

인증률(%)	EER(%)	TEER
96.42	3.58	-5.10

인증부 자체 성능평가에서 인증률은 각각 92.62%와 95.92%를 보여주었으며, 객관적 성능 검증을 위해 수행한 YOHO 표준 DB의 실험 결과 또한 <표 7>에서 보듯이 96.42%로 비슷한 결과를 보여주었다. 주목할 점은 화자 등록(훈련)에 사용한 음성발성 내용과 테스트에 사용한 음성발성 내용이 다른 경우, 즉 문장내용으로 등록하고 숫자 음으로 테스트하는 경우보다 YOHO 실험의 경우처럼 모두 숫자 음으로 등록과 테스트를 하면 인증률이 향상되었음을 보인다는 것이다. 이는 화자인증 시스템이 문장종속인 경우가 문장독립인 경우에 비해 인식성능이 좋아짐을 의미한다.

4.2. 전화 기반에서의 성능

이 실험은 전화 기반에서의 성능을 알아보기 위한 실험으로 실제 전화음성으로 화자인증을 하는 경우의 성능평가이다. <표 8>은 등록용 음성으로 마이크를 이용하여 문장을 발성한 각각의 샘플을 모두 더한 약 2분 40초 길이의 음성을 이용하고, 테스트 음성으로는 휴대전화를 통해 입력받은 15초 길이의 문장발성 샘플 10개를 사용하여 인증 실험한 결과이다.

<표 8> 문장 모델에 대한 전화 음성 적용 결과

인증률(%)	EER(%)	TEER
73.00%	27.00	-4.88

<표 9>는 등록용 음성으로 휴대전화를 통해 입력받은 15초 길이의 문장발성 샘플 5개를 합친 약 1분 15초 음성을 이용하고, 테스트 음성으로 역시 휴대전화를 통해 입력받은 15초 길이의 문장발성 샘플 5개를 사용하여 인증 실험한 결과이다.

<표 9> 전화 음성 모델에 대한 전화 음성 적용 결과

인증률(%)	EER(%)	TEER
99.5	0.50	-3.00

실제의 전화음성을 이용한 화자인증 시스템의 성능평가에서는, PC에서 고성능 마이크로 입력받은 음성을 사용하여 화자를 모델링한 경우 73%, 전화음성으로 화자를 모델링 하였을 경우는 99.5%의 인증률을 보여주었다.

이러한 화자모델에 따른 인증률의 차이는 마이크 특성의 영향을 받은 결과로 생각된다. 즉, PC에서 고성능 마이크를 사용하여 입력받은 깨끗한 음성으로 화자모델한 후 전화음성으로 인증을 시도하는 경우, PC에서 입력받을 때 사용한 마이크와 테스트할 때 사용된 전화기 마이크의 특성이 다른 것에 대한 보상이 되지 않아 인증률이 낮게 되며, 전화음성으로 모델한 후 전화음성으로 인증을 시도하는 경우, 화자모델을 위한 등록음성을 입력할 때 사용한 전화기로 또한 인증 테스트를 하였으므로 마이크 특성이 같아서 결과가 향상되었음을 확인하였다.

5. 결론 및 고찰

본 논문은 DSP 보드를 이용한 전화음성 기반의 실시간 화자인증 시스템의 구현과 그 성능 평가에 대하여 서술하였다.

논문에서 사용된 DB는 15명의 화자를 대상으로 음성을 수집하여 구성하였으며 배경화자의 수는 6명으로 하였다. 인증부 자체 성능평가에서 인증률은 약 95%를 보여주었다. 객관적 평가를 위해 실시한 표준 YOHO DB의 실험 결과 또한 비슷한 결과를 보여주었다. 배경화자 선정방법에 있어 MSC만 사용할 경우와 MSC와 MSF 모두 사용한 방법의 결과 차이는 미미하여 MSC 방법의 결과만을 제시하였

다. 이는 실험에 사용된 화자 수와 테스트용 음성 데이터 양이 두 방법의 차이를 뚜렷이 드러내기에는 적은 양이기 때문이다.

실제의 전화음성을 이용한 화자인증 시스템의 성능평가에서는 PC에서 고성능 마이크로 입력받은 음성을 사용하여 화자를 모델링한 경우 약 73%의 인증률을 나타내었지만, 전화음성으로 화자를 모델 하였을 경우는 99.5%의 인증률을 보여주었다. 이러한 화자모델에 따른 인증률의 차이는 마이크 특성의 영향을 받은 결과로 생각된다. 즉, 전자의 경우 PC에서 입력받을 때 사용한 마이크와 테스트할 때 사용된 전화기 마이크의 특성이 다른 것에 대한 보상이 되지 않아 인증률이 낮게 되며, 후자의 경우 화자모델을 위한 등록음성을 입력할 때 사용한 전화기를 인증실험에서도 사용을 하였으므로 결과가 향상된 것으로 판단된다. 또한 전자의 경우에 채널보상을 위한 켑스트럼 평균 차감법(CMS: Cepstral Mean Subtraction)을 적용하여 보았으나 뚜렷한 성능의 향상을 얻을 수 없었는데 이는 전화음성 샘플을 수집할 때 모두 동일한 핸드셋을 사용했기 때문인 것으로 추정된다. 차후 서로 다른 핸드셋에서 수집된 전화음성을 이용한 실험이 필요하며 핸드셋과 채널보상을 위한 방법들이 추가적으로 구현된다면 더 나은 성능을 보여 줄 수 있을 것이다.

참 고 문 헌

[1] L. R. Rabiner & R. W. Schafer (1978), Digital processing of speech signals, Prentice Hall.

[2] D. A. Reynolds (1995), "Speaker identification and verification using gaussian mixture speaker models", *Speech Communication*, Vol. 17 pp.91~108.

[3] D. A. Reynolds (1992), "A gaussian mixture modeling approach to text-independent speaker identification", *박사학위논문*, Georgia Institute of Technology.

[4] D. A. Reynolds, 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, pp.72-83.

[5] G. H, S. M (1994), "Text-independent speaker identification", *IEEE Signal Processing Magazine*, pp.18-32.

접수일자 : 2004년 1월 27일
게재결정 : 2004년 3월 9일

▶이현승(Hyeon Seung Lee)
주소: 487-711 경기도 포천시 선단동 대진대학교 공과대학 전자공학과
소속: 대진대학교 공과대학 전자공학과
전화: +82-31-535-9522, Fax: +82-31-539-1900
E-mail: empasland@empal.com

▶ 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