열악한 환경에 강인한 화자인증을 위한 위상 기반 특징 추출 기법

권철홍*

A Phase-related Feature Extraction Method for Robust Speaker Verification

Chul-hong Kwon*

이 논문은 교육과학기술부의 재원으로 2007년도 한국학술진흥재단(No. KRF-2007-D00741(I00101))과 2009년도 한국연구재단(No. 2009-0073337)의 지원을 받아 수행된 연구임

요 약

화자인증 시스템은 훈련 환경과 인식 환경이 다른 경우 인식 성능이 크게 저하된다. 이러한 훈련과 인식 환경의 불일치는 다양한 잡음과 상이한 채널 환경 때문이다. 본 논문은 화자인증 시스템의 강인성 개선을 위하여 음성신호의 위상에 기반한 특징 추출 기법을 제안한다. 이 방법은 음성신호의 위상으로부터 순시 주파수를 계산하여 대역별로 순시 주파수를 모두 모아 구한 히스토그램으로부터 특징 계수를 추출한다. 이 특징 파라미터를 적용한 결과 조용한 환경뿐만 아니라 잡음환경 그리고 채널왜곡 환경에서도 화자인증 시스템의 성능이 개선됨을 알 수 있다.

ABSTRACT

Additive noise and channel distortion strongly degrade the performance of speaker verification systems, as it introduces distortion of the features of speech. This distortion causes a mismatch between the training and recognition conditions such that acoustic models trained with clean speech do not model noisy and channel distorted speech accurately. This paper presents a phase-related feature extraction method in order to improve the robustness of the speaker verification systems. The instantaneous frequency is computed from the phase of speech signals and features from the histogram of the instantaneous frequency are obtained. Experimental results show that the proposed technique offers significant improvements over the standard techniques in both clean and adverse testing environments.

키워드

화자인증, 강인성, 가산잡음, 채널불일치, 위상기반 특징추출

Key word

speaker verification, robustness, additive noise, channel mismatch, phase-related feature extraction

* 대전대학교 정보통신공학과 교수 (교신저자)

접수일자 : 2010. 01. 13
심사완료일자 : 2010. 01. 29

Ⅰ. 서 론

최근 들어 통신 기술과 신호처리 기술의 급속한 발전에 힘입어 다양한 통신 매체에 신호처리 기술의 적용에 대한 관심이 증가하고 있으며, 음성은 인간의 가장 기본적인 의사전달 수단이기 때문에 음성신호처리에 관한 연구는 중요한 정보통신 기술 중의 하나로 인식되고 있다. 음성신호처리 분야 중에서 근래 들어 많은 관심을 불러일으킨 분야 중 하나는 화자인식(speaker recognition) 기술이다. 화자인식이란 발성된 음성 신호에 내재된 개인 고유의 정보를 기반으로 화자를 추정하고 인지하는 기술이다. 화자인식은 응용 분야에 따라 화자식별(speaker identification)과 화자인증(speaker verification)으로 나누어진다[1]. 화자식별이란 여러 후보 중에서 발화한 한 명의 화자를 찾는 방법으로 자동 회의록 작성에 응용될 수 있으며, 화자인증이란 등록된 화자(claimant, 사용자)와 사칭자(impostor)를 구분하는 기법으로 텔레뱅킹 등에서 본인 인증에 사용될 수 있다. 본 논문에서는 화자식별 보다 응용범위가 광범위한 화자인증을 다룬다.

텔레컴 산업에 화자인증 기술의 효과적인 적용을 위해 다양한 잡음 및 채널왜곡에 강인한 화자인증 시스템의 개발이 시급하다. 사용자 음성을 입력 받는 과정에서 추가되는 잡음 외에도 전화선 채널을 통할 경우에 발생하는 채널 차이에 의한 왜곡에 대한 강인성을 우선적으로 해결해야 한다. 본 논문에서는 음성을 이용한 화자인증 기술을 실생활에 적용하기 위해 잡음과 채널왜곡에 강인한 특징 추출 알고리듬의 개발을 다룬다.

현재 화자인증 시스템은 조용한 환경에서 고성능 마이크로 수집한 음성 데이터로 훈련하고 인식하였을 경우 충분히 좋은 성능을 보여준다. 그러나 훈련 환경과 인식 환경이 달라질 경우 시스템의 인식 성능은 크게 저하된다. 이러한 훈련과 인식 환경의 불일치는 다양한 잡음과 상이한 채널 환경 때문이다[1][2]. TIMIT(clean speech)와 NTIMIT(noisy telephone-bandwidth speech)의 음성 데이터를 사용한 화자인식 성능을 비교해보면, TIMIT의 경우에는 약 95.5%의 인식률을 보인 반면에 NTIMIT의 경우에는 단지 60.7%의 인식률을 보였다[3]. 이를 통하여 기존의 화자인증 기술을 실제 환경에 적용하면 잡음 및 채널왜곡의 영향에 매우 민감한 성능을 보일 것이라는 사실을 알 수 있다.

이러한 문제점을 극복하기 위해, 음질 향상 기법인 spectral subtraction, noise-canceling microphones, pre-processing 등의 다양한 시도가 이루어졌다[4][5]. 그러나 음질 향상 기법은 일관된 성능을 보장해 주지 못하므로 궁극적인 해결 방안이 될 수 없다. 본 논문에서는 다양한 잡음 및 채널왜곡 환경에 강인한 화자인증을 위하여 위상 기반 특징 추출 알고리듬을 제안한다.

논문의 구성은 2장에서는 본 논문에서 제안한 위상 기반 특징 추출 기법에 대해 설명한다. 3장에서는 제안한 화자인증 시스템을 실험하고, 결과를 분석한다. 그리고 4장에서 결론을 맺는다.

Ⅱ. 위상 기반 특징 추출 기법

2.1 위상 기반 특징 파라미터

음성신호는 단구간에서 quasi-stationary하다고 가정하므로 음성신호 $s(t)$ 의 단구간 Fourier 변환(short-time Fourier transform: STFT) $S(f,t)$ 는 다음과 같이 주어진다.

$$S(f,t)=\int_{-\infty}^{\infty} s(\tau)\,w(t-\tau)\,e^{-j2\pi f\tau}d\tau \qquad (1)$$

여기에서 $w(t)$ 는 단구간을 갖는 창함수이다. STFT $S(f,t)$ 는 또한 다음과 같이 표현할 수 있다.

$$S(f,t)=|S(f,t)|e^{j\phi(f,t)} \qquad (2)$$

여기에서 $|S(f,t)|$는 진폭 스펙트럼이고, $\phi(f,t)=\angle S(f,t)$ 는 위상 스펙트럼이다. 이와 같이 음성신호 $s(t)$는 진폭 스펙트럼뿐만 아니라 위상 스펙트럼을 갖고 있다. 즉, 위상 스펙트럼은 음성신호가 갖고 있는 정보의 반을 포함하고 있다.

Ohm은 인간의 청각 시스템이 위상에 둔감하다고 주장했다[6]. 즉, 청각 시스템은 음성을 인지하기 위해 위상 스펙트럼을 무시하고 단지 진폭 스펙트럼만을 사용

한다는 것이다. Helmholtz는 서로 다른 위상 스펙트럼을 가지나 동일한 진폭 스펙트럼을 갖는 신호를 사용한 실험을 통하여 음성인지에 차이가 없다고 주장하였다[6]. 그러나 이들의 주장이 옳지 않다는 사실이 여러 연구자들에 의해 밝혀졌다. 특히, 음성 코딩과 음성 합성 분야에서 위상 스펙트럼은 음성신호의 품질뿐만 아니라 자연스러움에도 영향을 미친다는 것이 알려졌다[7][8]. 음성인식 분야에서도 위상 스펙트럼을 사용함으로써 인식 성능을 향상시킬 수 있다는 연구결과가 일부 발표되어 있다[9]. 그러나 음성인식 분야에서 대부분, 특히 화자인식 분야에서는 위상 스펙트럼이 무시된다. 화자인식 분야에서 위상 스펙트럼이 화자를 인식하는데 어떠한 정보를 제공하는가를 밝히면 위상 스펙트럼은 유용한 정보가 될 수 있다.

현재 음성인식 및 화자인식 분야에서 가장 널리 사용되는 특징 파라미터는 진폭 스펙트럼에서 추출한 MFCC(Mel-frequency cepstral coefficients)이다. 즉, MFCC를 추출할 때 음성신호가 갖고 있는 정보의 반인 위상 스펙트럼은 이용하지 않고 진폭 스펙트럼만 이용한다. 그런데 진폭 스펙트럼은 그 특성상 배경잡음이나 채널왜곡과 같은 환경의 영향을 많이 받는다[9]. 따라서 MFCC는 이러한 환경 변수에 대하여 강인하지 못하다. 그러므로 본 논문에서는 위상 스펙트럼이 이러한 환경에 강인한가를 밝히고, 음성신호의 위상으로부터 구한 순시 주파수에 기반한 특징 추출 기법을 제안한다.

일반적으로 실수 값을 갖는 시간신호는 주파수 영역으로 변환하면 음의 성분을 갖게 되는데 이 음의 성분은 없어도 된다. 이 음의 성분을 없애기 위해 음의 성분을 0으로 놓고 양의 성분이 두 배가 되는 분석신호를 다음과 같이 정의한다.

$$S_a(f)=\begin{cases}2S(f), & \text{for } f>0,\\ S(f), & \text{for } f=0,\\ 0, & \text{for } f<0\end{cases} \qquad (3)$$
$$= S(f)\cdot 2U(f)$$

여기서 $U(\cdot)$는 계단 함수이다. 이와 같이 정의하면 $S_a(f)$의 역 Fourier 변환 $s_a(t)$는 다음과 같이 나타낼 수 있다[10].

$$\begin{aligned} s_a(t) &= F^{-1}\{S(f)\} * F^{-1}\{2U(f)\} \qquad (4)\\ &= s(t) * [\delta(t)+j\frac{1}{\pi t}]\\ &= s(t)+j[s(t) * \frac{1}{\pi t}]\\ &= s(t)+j\hat{s}(t)\end{aligned}$$

여기에서 F^{-1}와 $*$는 역 Fourier 변환과 컨벌루션 연산이고, $\hat{s}(t)$는 $s(t)$의 Hilbert 변환이다.

이렇게 구한 분석신호 $s_a(t)$를 다음과 같이 표현할 수 있다.

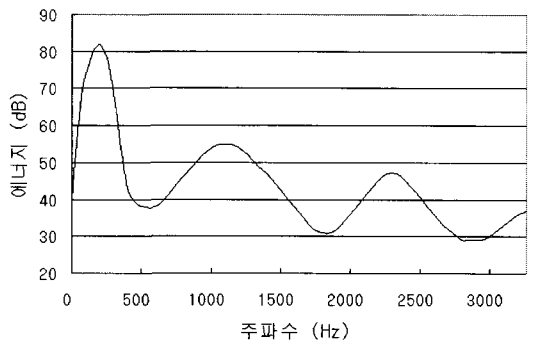
$$s_a(t)=a(t)e^{j\phi(t)} \qquad (5)$$

여기에서 $a(t)=|s_a(t)|$는 순시 진폭이고, $\phi(t)=\angle s_a(t)$는 순시 위상이다. 순시 주파수 $f_i(t)$는 위상으로부터 구할 수 있다.

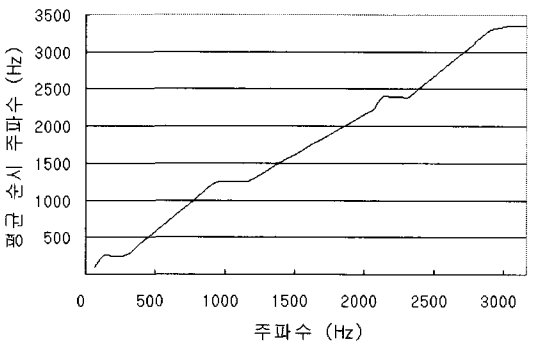
$$f_i(t)=\frac{1}{2\pi}\frac{d\phi(t)}{dt} \qquad (6)$$

순시 주파수의 특성을 보기 위하여, 분석신호 $s_a(t)$를 프레임 별로 멜 스케일 대역통과 필터에 통과시킨 후 각 대역별 신호로부터 순시 주파수 $f_i(t,\lambda)$를 구한다(λ는 필터의 중앙 주파수를 나타낸다). 그리고 각 대역 별로 순시 주파수 $f_i(t,\lambda)$의 평균 $F(\lambda)$를 구한다.

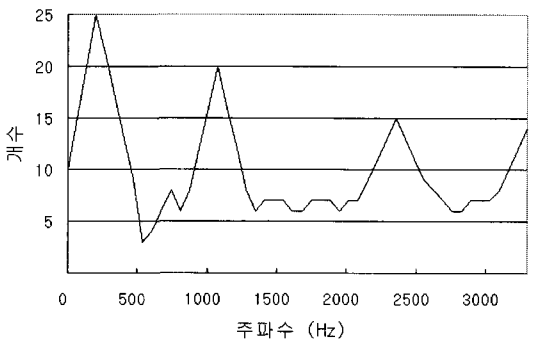
그림 1은 음성신호로부터 MFCC 분석으로 구한 스펙트럼(그림 1 (a))과 평균 순시 주파수 $F(\lambda)$ (그림 1 (b))를 비교한 예이다. MFCC 스펙트럼에서 포먼트가 있는 주파수 영역에서 평균 순시 주파수 $F(\lambda)$가 편평한 모양을 보여 준다는 것을 알 수 있다. 따라서 순시 주파수가 음성의 고유 정보인 포먼트를 포함하고 있다는 사실을 알 수 있다.



(a)



(b)

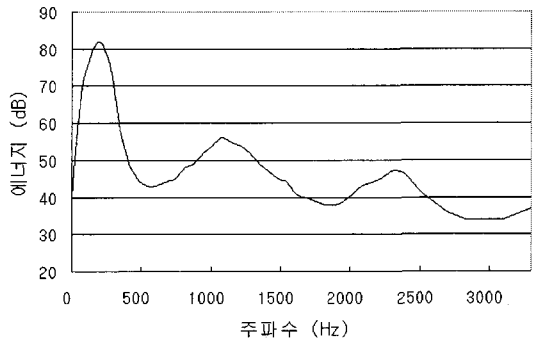


(c)

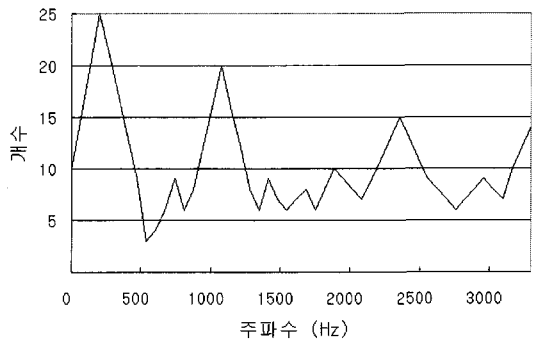
그림 1. 깨끗한 음성신호 (a) MFCC 스펙트럼
(b) 평균 순시 주파수 (c) 순시 주파수의 히스토그램
Fig. 1 Clean speech signal (a) MFCC spectrum
(b) Average instantaneous frequency
(c) Histogram of instantaneous frequency

다음으로, 대역별로 순시 주파수 $f_i(t,\lambda)$ 를 모두 모아 구한 히스토그램 $H(\lambda)$(그림 1 (c))는 음성신호의 포먼트 성분을 보다 분명하게 보여 준다. 그림 1 (a)와 (c)를 비교하여 보면 히스토그램 $H(\lambda)$에서 피크는 포먼트에 해당한다. 따라서 히스토그램 특징 파라미터를 화자인증에 적용하여 성능 개선 여부를 실험할 필요가 있다.

순시 주파수의 히스토그램 $H(\lambda)$가 가산잡음 왜곡에 강인하다는 것을 다음 예에서 알 수 있다. SNR 10 dB로 조정된 백색 가우시안 잡음을 음성신호에 더한 잡음음성 신호에 대해, MFCC 스펙트럼과 히스토그램 $H(\lambda)$이 그림 2에 보인다.



(a)



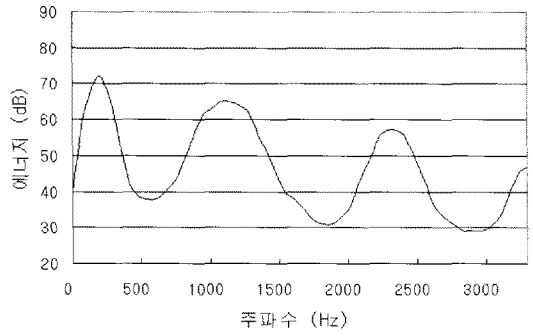
(b)

그림 2. 잡음 음성신호 (a) MFCC 스펙트럼
(b) 순시 주파수의 히스토그램
Fig. 2 Noisy speech signal (a) MFCC spectrum
(b) Histogram of instantaneous frequency

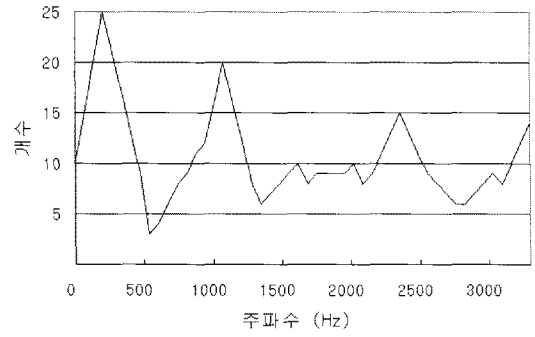
그림 1 (c)와 그림 2 (b)를 비교하여 보면 차이가 별로 없다는 사실로부터 이 히스토그램은 가산잡음 왜곡에 의해 크게 변하지 않는다는 것을 알 수 있다.

그런데 그림 1 (a)와 그림 2 (a)를 비교하면 MFCC 스펙트럼은 차이가 존재함을 볼 수 있으므로, 히스토그램 $H(\lambda)$가 가산잡음에 더 강인하다는 것을 알 수 있다.

그림 3은 채널왜곡된 음성신호에 대해 MFCC 스펙트럼과 순시 주파수의 히스토그램 $H(\lambda)$을 보여 준다. 가산잡음 왜곡인 경우와 같이 이 경우에도 히스토그램 $H(\lambda)$(그림 1 (c)와 그림 3 (b) 비교)는 MFCC 스펙트럼(그림 1 (a)와 그림 3 (a) 비교) 보다 채널왜곡에 더 강인함을 알 수 있다.



(a)



(b)

그림 3. 채널왜곡 음성신호 (a) MFCC 스펙트럼 (b) 순시 주파수의 히스토그램

Fig. 3 Channel-distorted speech signal (a) MFCC spectrum (b) Histogram of instantaneous frequency

앞에서 살펴 본 바와 같이 순시 주파수의 히스토그램 $H(\lambda)$가 잡음 및 채널왜곡에 강인하다는 사실을 이용하여, 본 논문에서는 이 특징 파라미터를 화자인증 시스템에 적용하여 열악한 환경에서 성능 개선을 이루고자 한다.

2.2 화자인증 시스템에의 적용

순시 주파수의 히스토그램 $H(\lambda)$을 화자인증 시스템의 특징 파라미터로 사용하기 위해, 히스토그램에 DCT(Discrete Cosine Transform)를 적용하여 켑스트럼 계수로 변환한다. 특징 계수를 구하는 절차는 다음과 같다. 한 프레임의 음성신호 $s(t)$에 대하여 Hilbert 변환을 통해 $\hat{s}(t)$를 구하여 분석신호 $s_a(t)$를 구성한다. 분석신호 $s_a(t)$를 프레임 별로 멜 스케일 대역통과 필터에 통과시킨 후 각 대역별 신호로부터 순시 주파수 $f_i(t,\lambda)$를 구한다. 그리고 대역별로 순시 주파수를 모두 모아 히스토그램을 구한다. 여기에 DCT를 적용하여 켑스트럼 계수로 변환한 뒤 이 계수를 화자인증 시스템의 특징 파라미터로 삼는다.

현재 대부분의 문장 독립형 화자인증 시스템에서는 GMM-UBM 시스템[11]을 사용한다. 본 논문에서도 GMM-UBM 시스템을 사용하였다. 이 시스템에서는 사용자 음성으로 훈련하는 화자별 GMM(Gaussian Mixture Model) 이외에, UBM(Universal Background Model)이라는 모델로 사칭자를 모델링 한다. UBM은 음성 특징의 화자 독립 분포를 표현하기 위해 훈련되는 하나의 큰 GMM이다. 그리고 화자인증 시스템에서는 화자 모델을 만들 때 대부분 훈련 음성 자료를 충분히 얻기 어려운 경우가 많다. 적은 훈련 음성 자료를 효과적으로 이용하는 방법으로 화자적응이 있다. 이는 UBM으로부터 화자적응을 통하여 각각의 화자 모델을 생성하는 것이다. 본 논문에서는 화자적응 방법 중 MAP(Maximum A Posteriori) 화자적응 기법[11]을 사용하였다.

Ⅲ. 실험 방법 및 결과

3.1 실험 방법

본 논문의 실험에서 사용한 음성 DB는, ETRI 음성정보 연구센터에서 구축한 한국어 화자인식용 영리용 음성 DB로, SNR 25 dB 이상 확보 가능한 조용한 사무실 PC 환경에서 중가의 마이크(모델명: Sennheiser MD425)를 사용하여 수집하였고, 16 kHz, 16 bit, Linear PCM으로 저장되었다.

잡음 보상에 대한 실험을 위해 잡음 환경의 음성을 제작하였다. 화자인증 시 사용된 테스트 음성에 잡음을 추

가하였다. 추가된 잡음은 Noisex-92[12]를 근거로 백색 잡음(white), 차량 잡음(volvo), 군중 잡음(babble)에 5 dB, 10 dB, 15 dB, 20 dB의 SNR을 적용하였다. 잡음 환경의 실험에서 훈련 모델은 SNR이 25 dB 이상의 조용한 사무실 환경에서 수집한 음성 DB로 작성하였다.

또한, 채널왜곡 실험을 위해 다양한 종류의 전화기에서 수집한 음성 DB도 사용하였다. 이 음성 DB도 ETRI에서 수집한 것으로, 테스트 음성은 훈련 음성과 다른 종류의 전화기로 수집한 것이다.

음성 DB는 250명(기간별: 주차 100명, 월차 100명, 3개월차 50명)의 화자가 발성한 2연 숫자, 4연 숫자, 문장으로 구성되어 있다. 문장 음성의 발성목록은 개인정보와 관련된 10개의 질문과 3어절 이내로 구성된 단문 10개로 구성되며, 한 화자당 동일한 목록을 5회 발성하고, 녹음 간격에 따라 주차/월차/3개월차로 구분하여 4회 반복한 것이다.

GMM은 EM(Expectation-Maximization) 알고리즘을 적용하여 2, 4, 8, 16, 32, 64, 128, 256, 512 순으로 mixture를 증가시킨 ML(Maximum Likelihood) 모델로 작성하였다. 실험 대상 화자 모델을 작성하기 위하여 MAP 화자적응을 할 때 각 화자의 0주차(주차 화자의 음성 중 기간별 첫 녹음시점) 6개 단문의 5회 발성음성으로 60초 분량을 사용하였다.

UBM 작성은 월차 화자 음성 중 기간별 첫 녹음 시점인 0개월차 월차 화자 100명으로 남자 50명과 여자 50명으로 구성하였다. 훈련 DB의 환경적 데이터가 균형이 맞으므로 남녀 화자 모두의 음성을 사용하여 하나의 UBM을 작성하였다. UBM 훈련 시 각 화자당 6단문의 5회 발성 음성으로, 화자당 약 72초 정도로 총 2시간 분량을 사용하였다.

본 논문에서 화자인증 실험의 테스트 음성 DB는 주차화자 100명을 대상으로 하였는데, 남자 50명과 여자 50명으로 구성하였다. 사용된 테스트 음성은 훈련 시점과 1주일 차이의 음성인 주차화자의 1주차 음성이다. 이 테스트 DB는 훈련과 독립적인(훈련에 사용되지 않은 단문) 3개 단문의 5회 음성을 사용하여 각 화자당 15 단문으로 하였다.

화자인증 시스템은 GMM-UBM으로 화자와 사칭자의 비율을 1:10으로 하였다. 표 1은 화자인증에 사용한 화자군의 분류와 음성시료의 개수를 보여주고 있다.

표 1. 실험에 사용한 화자 및 음성시료의 구성
Table 1. Speaker and speech database

	남자	여자	전체
화자 수	50 명	50 명	100 명
사용자 테스트 음성시료	750 단문	750 단문	1,500 단문
사칭자 테스트 음성시료	7,500 단문	7,500 단문	15,000 단문

사용된 모든 음성은 음성구간을 검출하여 앞, 뒤의 묵음을 제거한 단문의 음성을, 프레임 길이는 25ms, 프레임 주기는 10ms이며 Hamming 창함수를 사용하였다.

MFCC 기반 화자인증 시스템을 기본으로 하여, 순시 주파수의 히스토그램으로 구한 켑스트럼 계수(HIFCC), 그리고 이 둘의 결합으로 구성된 특징 파라미터를 사용하여 화자인증 시스템의 성능을 비교한다.

3.2 실험 결과

화자인증 분야에서 가장 널리 사용되는 성능 지표인 EER(Equal Error Rate)로 성능을 비교하였다. 표 2에 조용한 훈련 및 테스트 환경에서 화자인증 시스템의 성능을 보여주는 EER을 정리하였다. 실험 결과의 EER을 비교하면, HIFCC를 이용한 화자인증 시스템은 MFCC를 사용한 화자인증 시스템보다 EER이 증가하여 성능이 나빠졌다. 그러나 MFCC와 HIFCC를 결합한 화자인증 시스템은 MFCC 또는 HIFCC만을 이용한 화자인증 시스템보다 EER이 감소하여 성능이 개선됨을 알 수 있다.

표 2. 조용한 환경에서 화자인증 EER 비교
Table 2. Comparison of EERs for speaker versification systems in clean testing environments

	남자	여자	평균
MFCC	3.61%	4.48%	4.05%
HIFCC	4.16%	4.93%	4.55%
MFCC+HIFCC	2.88%	3.60%	3.24%

잡음 환경에 대한 화자인증 실험결과는 표 3과 같다. 이 경우에도 HIFCC를 이용한 화자인증 시스템은 MFCC를 사용한 화자인증 시스템보다 EER이 증가하였으나, MFCC와 HIFCC를 결합한 화자인증 시스템은 MFCC 또는 HIFCC만을 이용한 화자인증 시스템보다 EER이 감소하였다.

표 3. 잡음 환경에서 화자인증 EER 비교
Table 3. Comparison of EERs for speaker versification systems in noisy testing environments

	백색 잡음	차량 잡음	군중 잡음	평균
MFCC	28.39%	13.92%	15.96%	19.42%
HIFCC	30.16%	15.93%	18.96%	21.68%
MFCC+HIFCC	24.58%	11.60%	12.96%	16.38%

채널왜곡 환경에 대한 화자인증 실험결과는 표 4와 같다. 이 경우에도 앞의 실험과 비슷한 성능 경향을 보였다.

표 4. 채널왜곡 환경에서 화자인증 EER 비교
Table 4. Comparison of EERs for speaker versification systems in mismatched telephone environments

	남자	여자	평균
MFCC	28.11%	29.07%	28.59%
HIFCC	29.66%	30.63%	30.15%
MFCC+HIFCC	24.18%	25.30%	24.74%

Ⅳ. 결 론

본 논문에서 잡음 및 채널왜곡 환경에서 화자인증 시스템의 강인성을 개선하기 위하여 위상 기반 특징 추출 기법을 제안했다. 이 방법은 음성신호의 위상으로부터 순시 주파수를 계산하여 대역별로 순시 주파수를 모두 모아 구한 히스토그램으로부터 특징 계수를 추출한다.

본 논문에서는 음성신호의 순시 주파수가 음성의 고유 정보인 포먼트를 포함하고 있다는 사실을 밝혔고, 순시 주파수의 히스토그램은 잡음 및 채널왜곡에 강인하다는 사실을 확인하고, 이 특징 파라미터를 화자인증 시스템에 적용하여 열악한 환경에서 성능 개선을 이루고자 하였다. 이 방법을 적용한 결과 조용한 환경뿐만 아니라 잡음환경 그리고 채널왜곡 환경에서도 화자인증 시스템의 성능이 개선됨을 알 수 있다.

감사의 글

이 논문은 교육과학기술부의 재원으로 2007년도 한국학술진흥재단(No. KRF-2007-D00741 (I00101))과 2009년도 한국연구재단(No. 2009- 0073337)의 지원을 받아 수행된 연구임.

참고문헌

[1] J. Campbell, "Speaker Recognition: a Tutorial," *Proc. IEEE*, vol. 85, pp. 1437-1462, 1997.

[2] J.M. Naik, "Speaker Verification," *IEEE Communication Magazine*, pp. 42-49, 1990.

[3] D.A. Reynolds and R.C. Rose, "Robust Text-independent Speaker Identification Using Gaussian Mixture Speaker Models," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 1, pp. 72-83, 1995.

[4] R.J. Mammone, X. Zhang and R.P. Ramachandran, "Robust Speaker Recognition : a Feature-based Approach," *IEEE Signal Processing Magazine*, pp. 58-70, 1996.

[5] J. Ortega-Garcia and J. Gonzalez-Rodriguez, "Overview of Speech Enhancement Techniques for Automatic Speaker Recognition," *IEEE Trans. Speech and Audio Processing*, pp. 929-932, 1996.

[6] L.R. Rabiner and R.W. Schafer, *Discrete-time Speech Signal Processing, Principles and Practice*, Prentice Hall, NJ, 1978.

[7] H. Pobloth and W.B. Kleijn, "On Phase Prception in Speech," *Proc. ICASSP*, pp. 29-32, 1999.

[8] D.S. Kim, "Perceptual Phase Redundancy in Speech," *Proc. ICASSP*, pp. 1383-1386, 2000.

[9] H.A. Murthy and V. Gadde, "The Modified Group Delay Function and its Application to Phoneme Recognition," *Proc. ICASSP*, pp. 68-71, 2003.

[10] P. Maragos, J.F. Kaiser and T.F. Quatieri, "Energy Separation in Signal Modulations with Application to Speech Analysis," *IEEE Trans. on Signal Processing*, vol. 41, pp. 3024-3051, 1993.

[11] D.A. Reynolds, T.F. Quatieri and R.B. Dunn, "Speaker Verification Using Adapted Gaussian Mixture Models," *Digital Signal Processing*, vol. 10, pp. 19-41, 2000.

[12] Noisex-92, http://www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html.

저자소개

권철홍(Chul-Hong Kwon)

1987년 서울대학교 전자공학과 (공학사)
1989년 한국과학기술원 전자공학과(공학석사)
1994년 한국과학기술원 전자공학과(공학박사)
1997년~현재 대전대학교 정보통신공학과 교수
1999년~2000년 미국 벨연구소 초빙연구원
2007년~2008년 호주 NSW대학 객원교수
※ 관심분야 : 화자인식, 음성인식, 모바일 멀티미디어