한국음향학회지 제23권 제3호 pp. 228~234 (2004)

강인한 화자확인 시스템을 위한 채널 불일치 보상 기법에 관한 연구

A Study on Channel Mis-match Compensation Technique for Robust Speaker Verification System

정 희 석*, 강 철 호*
(Hee Suk Jeong*, Chul Ho Kang*)

*광운대학교 대학원 전자통신공학과
(접수일자: 2003년 7월 10일; 수정일자: 2004년 3월 22일; 채택일자: 2004년 4월 9일)

본 논문에서는 공통 코드북의 평균값과 개인 코드북의 평균값 간의 바이어스 제거에 의한 채널 불일치 보상 알고리즘을 제안하였다. 제안한 방식은 학습시 공통 코드북의 센터값과 학습 데이터의 센터값과의 차수별 차를 미리 보상하여 학습하고, 확인시에도 공통 코드북의 센터값과 학습 데이터의 센터값과의 차수별 차를 보상하여 확인함으로써 채널의 불일치에 의한 급격한 본인 인식율 하락을 해결한다. 그러나, 무조건적인 평균값 보상은 사칭자의 인증오류를 가져오게 되므로 채널의 변이에 비례하는 적절한 가중치를 통한 평균값 보상이 필요하다. 따라서, 제안하는 방식은 음성구간을 제외한 묵음구간의 분포를 고려하여 학습시 채널과의 변이차이를 비선형함수에 의한 가중치로 보상해준다. 모의 실험 결과 기존의 캡스트럼 평균 차감법을 사용할 때보다 제안한 알고리즘을 적용했을 때의 본인 거부 오류율이 평균 14.95% 감소함을 알 수 있었다.

핵심용어: 채널 불일치 보상, 바이어스 제거, 화자인식
투고분야: 음성처리 분야 (2.5)

In this paper, we proposed the compensation technique that overcomes the limitations of the conventional approaches through summing up the bias terms between world's codebook and individual codebook vectors of feature parameters. But, mean compensation without condition can bring higher false acceptance. Therefore, the proposed technique compensates the channel mis-match condition by weighted bias sum using nonlinear function regarding to the distortion between speech and silence. The simulation results show that the FRR (flase reject rate) is decreased 14.95% when the proposed algorithm was applied.

Keywords: Mis-match compensation, Bias reduction, speaker verification
ASK subject classification: Speech signal processing (2.5)

I. 서 론

화자확인 시스템의 경우 학습시 사용된 시스템과 동일한 시스템으로 인증시험을 할 경우 그 특성의 차이가 없기 때문에 높은 인식율을 보이게 되나 실생활에서 네트워크를 기반으로 적용하게 될 경우 이러한 예는 극히 찾아보기 힘들게 된다.

즉, 임의의 화자가 네크워크를 기반으로 임의의 시스템으로 학습하여 자신의 모델을 생성한 후 이를 인증하고자 할 때 항상 학습시와 동일한 환경에서의 시스템으로만 확인하게 되지는 않는다.

따라서 이러한 경우 마이크와 같은 음성입력 시스템이나 사운드카드와 같은 채널의 고유한 특성에 영향을 받게 되며 이로 인한 바이어스효과는 심각한 오인식을 가져오는 원인이 된다[1].

이러한 채널 불일치를 보상하기 위한 많은 방식들이 제안되어 왔다[2-4]. 이들 기술들은 대부분 '불일치'의 원인을 발성환경 잡음과 채널의 두 가지 요인으로 나누고 있다. 본 논문에서는 채널에 대한 불일치를 보상하는 기법을 제안한다.

책임저자: 정 희 석 (afe001@korea.com)
139-701 서울시 노원구 월계동 447-1
광운대학교 대학원 전자통신공학과
(전화: 02-940-0370; 팩스: 02-917-9670)

본 논문의 구성은 전체 5장으로 구성되어 있다. 2장에서는 기존의 채널 불일치 보상 기법에 대하여 설명하였다. 3장은 제안한 바이어스 제거에 의한 채널 보상 기법의 이론적 배경과 실제 화자 확인 시스템에 적용하는 기법에 대하여 설명하고, 4장에서는 모의실험 및 성능을 평가했으며 마지막 5장에서 결론을 이끌어 냈다.

II. 기존의 채널 불일치 보상 기법

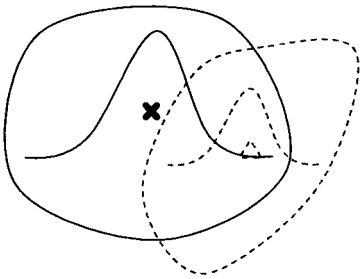
서로 다른 마이크는 상이한 전달 특성을 가진다. 그리고, 심지어 같은 마이크라 할지라도 마이크와의 거리나 실내 잔향에 따라 전달 특성이 크게 변화한다. 이러한 불특정 선형 채널에 의한 채널 왜곡을 제거하는 방법으로 Cepstral Mean Subtraction (CMS) 가 제안 되었다[5-6]. CMS는 현재까지도 음성 인식과 화자 인식 시스템에 가장 보편적이면서도 우수한 성능을 나타내고 있다.

2.1 기존의 채널 불일치 보상 기법 - CMS

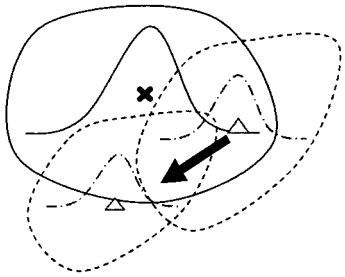
CMS는 채널 왜곡 성분이 천천히 조금씩 변한다고 가정하고 순수한 음성 켑스트럼의 장구간 평균이 0이라면, 켑스트럼의 영역에서 전체구간에 대한 평균을 구하여 차감하면 채널 효과를 제거할 수 있다. 즉, 채널의 영향은 순수한 음성의 켑스트럼에 가산된 형태로 나타나므로 채널 켑스트럼의 추정치는 필터링된 음성의 켑스트럼 평균을 통해 구할 수 있고 결과적으로 추정된 채널 켑스트럼을 제거하여 채널효과를 보상할 수 있다는 것이다.

주어진 신호 x[n]에 대해, 단구간 (short-time) 분석을 통해 얻은 켑스트럴 벡터 $X=x_0, x_1, \ldots, x_{T-1}$의 평균 벡터 $\overline{x}=\frac{1}{T}\sum_{t=0}^{T-1}x_t$ 로 주어진다.

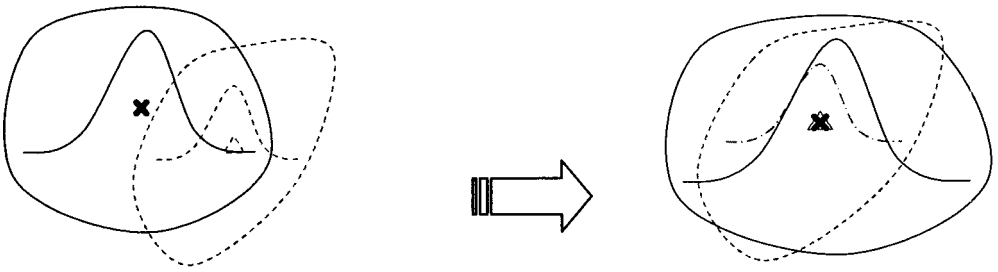
CMS는 식 (1)과 같이 각 벡터 x_t 로부터 평균 벡터 $\overline{x}$



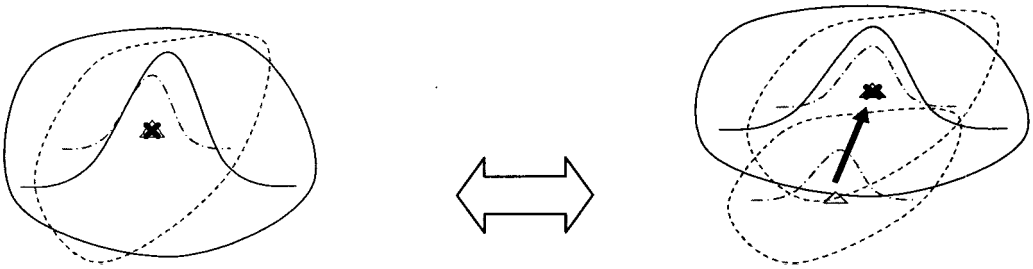
Distribution chart of world's and individual codebook.



Variations of individual codebook resulted from mis-match.



Channel mis-match compensation by proposed method (enrollment).



Channel mis-match compensation by proposed method(verification).

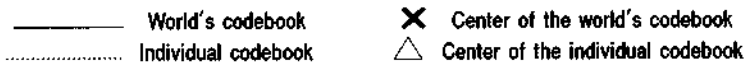


그림 1. 제안한 채널 불일치 보상 기법
Fig. 1. Proposed channel mis-match compensation technique.

를 차감함으로써 차감된 켑스트럼 벡터 $\hat{x}_t$를 얻게 된다.

$$\hat{x}_t = x_t - \bar{x} \quad (1)$$

주어진 신호 x(n)이 선형 채널 h(n)을 통과하고 난 후의 신호를 y(n)이라 하면, 켑스트럴 벡터 $Y = y_0, y_1, \ldots, y_{T-1}$를 얻을 수 있다. 여기서,

$$h = C(\ln|H(\omega_0)|^2 \mathrm{K}\ \ln|H(\omega_M)|^2) \quad (2)$$

, where C 는 DCT matrix라고 하면 $y_t = x_t + h$ 가 된다. 그러므로, 평균벡터 $\bar{y}$는

$$\bar{y} = \frac{1}{T}\sum_{t=0}^{T-1} y_t = \frac{1}{T}\sum_{t=0}^{T-1}(x_t + h) = \bar{x} + h \quad (3)$$

가 되고, 차감된 신호 $\hat{y}_t$는

$$\hat{y}_t = y_t - \bar{y}_t = \hat{x}_t \quad (4)$$

가 되고, 이 차감된 신호는 선형 채널에 강인한 특징 벡터가 되며 학습과정과 인식과정에 동일하게 사용된다.

2.2 문맥 종속 화자 확인 시스템에서 CMS의 문제점

켑스트럼 평균 차감법은 순수한 음성에 대한 켑스트럼 평균이 0 이 되기 위해서 유성음, 무성음, 파열음 등이 음향학적 균형을 이루어야 하므로 이러한 조건이 만족되지 않을 경우 채널성분 이외의 음성성분이 차감되는 단점을 가진다.

단음소로 구성된 짧은 발성음의 경우, 예를 들어, /s/의 경우, /s/는 안정 (stationary) 구간이므로, $\bar{x}$는 x_t와 거의 유사하게 된다. 그러므로, 차감후 벡터 $\hat{x}_t \approx 0$이 된다. 따라서, 일반적으로 2~4초 이상의 음성데이터에서는 CMS는 우수한 성능을 나타내는 것으로 잘 알려져 있다. 그러나, 문맥 종속 화자 확인 시스템의 경우 음성 데이터는 평균적으로 0.5초에서 1초 사이로 음향학적 균형을 이루었다고 볼 수 없으며 CMS의 가정은 성립되지 않게 된다.

III. 제안한 바이어스제거에 의한 채널 불일치 보상 알고리즘

제안한 방식은 학습시와 인식시의 채널 불일치 조건 (Mismatch-condition)을 공통 코드북 센터의 평균값과 개인 코드북 센터의 평균값차의 보상으로 제거한다. 화자가 발성한 음성의 특징 파라미터는 접속하는 시스템의 특성, 즉, 사운드 카드, 마이크, 배경 잡음 등에 의해 크게 변화한다. 이러한 특징 파라미터의 변화는 그림 1에서 보듯이 화자 영역과의 차이를 발생시켜 본인 오거부율을 발생시킨다. 그러나, 이러한 차이는 화자 영역의 분포형태에는 크게 영향을 끼치지 못한다. 즉, 채널에 의해 왜곡된 특징 파라미터의 변화는 화자 영역의 평균 (mean)값에 영향을 미치지만, 분산 (variance)값에는 영향을 미치지 못하게 된다. 따라서, 제안하는 방식은 그림 1과 같이 학습시 공통 코드북의 센터값과 학습 데이터의 센터값과의 차수별 차를 미리 보상하여 학습하고 확인시에도 공통 코드북의 센터값과 학습 데이터의 센터값과의 차수별 차를 보상하여 확인함으로써 채널의 불일치에 의한 급격한 본인 인식율 하락을 해결한다. 그러나, 무조건적인 평균값 보상은 사칭자의 인증오류를 가져오게 되므로 채널의 변이에 비례하는 적절한 가중치를 통한 평균값 보상이 필요하다. 따라서, 제안하는 방식은 음성구간을 제외한 묵음 구간의 분포를 고려하여 학습시 채널과의 변이차이를 비선형함수에 의한 가중치로 보상해준다.

다음은 3.1절과 3.2절은 제안한 채널불일치 보상알고리즘의 수행단계이다.

3.1 등록 과정

step 1〉 초기화 : 음성구간의 모든 입력벡터에 대한 하나의 중심값을 설정한다.

$$\mu_{individual}^{(p)} = \frac{\sum_{all} x^{(p)}}{N_T} \quad (5)$$

where , N_T : 총프레임수

step 2〉 공통 코드북의 평균 센터값과 입력벡터에 대한 중심값의 차를 구한다.

$$Bias^{(p)} = \mu_{world}^{(p)} - \mu_{individual}^{(p)} \quad (6)$$
$$p = 0,1,2,\cdots k$$

step 3〉 공통 코드북의 평균 센터값과 입력벡터에 대한 중심값의 차를 보상한다.

$$\bar{x}^{(p)} = Bias^{(p)} + x^{(p)} \quad (7)$$
$$p = 0,1,2,\cdots k$$

step 4〉 묵음 구간의 입력벡터에 대한 평균값을 구하

고, 화자 모델에 기록한다.

$$N_{sp}^{(p)} = \frac{\sum_{silence} x^{(p)}}{N_{silence}} \tag{8}$$

step 5) 등록과정을 수행한다.

3.2 확인 과정

step 1) 초기화 : 묵음구간의 입력벡터에 대한 평균값을 구하고, 시그모이드 (Sigmoid)함수에 의해 가중치를 결정한다.

$$distortion = \sum_{p=0}^{p=k} N_{sp}^{(p)} - N_{bg}^{(p)} \tag{9}$$

$$w = \frac{1}{1+\exp(10*(-0.5+distortion))} \tag{10}$$

$where, N_{sp}$ 는 화자모델의 묵음구간 LPCC

$where, N_{bg}$ 는 입력데이터의 묵음구간 LPCC

step 2)음성구간의 모든 입력벡터에 대한 하나의 중심값을 설정한다.

$$\mu_{individual}^{(p)} = \frac{\sum_{all} x^{(p)}}{N_T} \tag{11}$$

step 3) 공통 코드북의 평균 센터값과 입력벡터에 대한 중심값의 차를 구한다.

$$Bias^{(p)} = \mu_{world}^{(p)} - \mu_{individual}^{(p)} \tag{12}$$

step 4) 공통 코드북의 평균 센터값과 입력벡터에 대한 중심값의 차를 적절한 가중치를 주어 보상한다.

$$\bar{x}^{(p)} = w * Bias^{(p)} + x^{(p)} \tag{13}$$

step 5) 화자 확인과정을 수행한다.

그림 2는 제안한 채널 불일치 보상 기법을 적용한 화자 확인 시스템의 전체 구성도이다. 그림과 같이 제안한 보상 기법은 등록 과정과 확인 과정에서 모두 사용된다.

IV. 실험 환경 및 결과

본 실험에 사용된 배경 모델을 만드는데 사용된 음성 데이터는 20~30대의 100명의 남/여 화자로부터 5가지 종류의 헤드셋 마이크로 수집한 10번씩 발음한 데이터로 사전에 구성하였다. 인증 실험에 사용된 단어는 ("안녕하세요")로 각 30명의 남/여 화자에 의해 한달에 걸쳐 5가지 특성이 다른 헤드셋 마이크로 수집되었다.

실험에 사용된 음성 데이터는 11.025kHz 16bit 로 샘플링 되었고, 음성 분석 구간은 한 프레임을 20msec로 하고 1/3 중첩시켜 해밍 윈도우 (hamming window)을 취한 후 1차의 에너지 캡스트럼과 19차의 LPC 캡스트럼을 특징 파라미터로 구성하였다.

전체 시스템은 그림 2와 같이 끝점 추출, 특징 파라미터 추출 및 개인 가중치 함수에 의해 화자의 개인성을 강조하고 채널 불일치를 보상하기 위해 제안된 보상 기법을 포함하는 전처리 과정과 화자간 변별력을 향상시키기 위한 공통 코드북을 이용한 개인 코드북 생성 및 화자 모델을 생

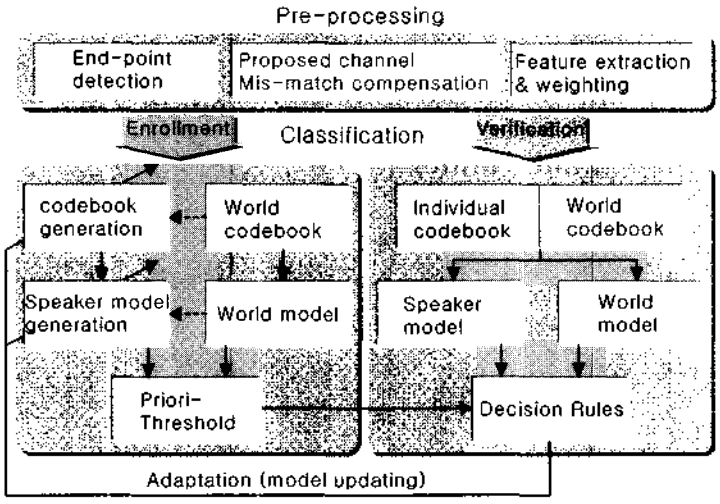


그림 2. 채널 불일치 보상 기법을 적용한 화자 확인 시스템의 구성도
Fig. 2. Block diagram of the proposed speaker verification system.

표 1. 기존의 방법과 제안한 방법의 인식율 비교 (마이크 1로 학습/마이크 1로 인식)

Tabel 1. Performance comparison between CMS and proposed method (Mic. 1-Enrollment, Mic. 1-verification).

Mic. 1(Enrollment) → Mic. 1(verification)				
CMS(Cepstral Mean Subtraction)				
Each 30 speakers	Male		Female	
	FR	FA	FR	FA
total	37/600	113 /17400	32/600	164 /17400
%	6.16	0.65	5.33	0.94
total error rate	FR: 5.75%, FA: 0.8%			
Proposed channel mis-match compensation				
Each 30 speakers	Male		Female	
	FR	FA	FR	FA
total	32/600	133 /17400	33/600	166 /17400
%	5.33	0.76	5.5	0.95
total error rate	FR: 5.42%, FA: 0.86%			

표 2. 기존의 방법과 제안한 방법의 인식율 비교(마이크 1로 학습/마이크 2로 인식)

Tabel 2. Performance comparison between CMS and proposed method(Mic. 1-Enrollment, Mic. 2-verification).

Mic. 1(Enrollment) → Mic. 2(verification)				
CMS(Cepstral Mean Subtraction)				
Each 30 speakers	Male		Female	
	FR	FA	FR	FA
total	52/600	109 /17400	48/600	165 /17400
%	8.66	0.63	8.0	0.95
total error rate	FR: 8.33%, FA: 0.79%			
Proposed channel mis-match compensation				
Each 30 speakers	Male		Female	
	FR	FA	FR	FA
total	41/600	106 /17400	35/600	156 /17400
%	6.83	0.61	5.83	0.89
total error rate	FR: 6.33%, FA: 0.75%			

성하고 우도비를 계산하여 인증/거부하는 학습 및 인식과정, 마지막으로, 인식시 화자내 변화에 적응할 수 있도록 코드북 및 화자 모델을 변화시키는 화자 적응 과정으로 구성된다. 공통 코드북은 LBG 알고리즘을 이용하여 128개의 클러스터로 구성되었고 개인 코드북은 modified LBG 알고리즘을 이용한 가변길이 개인 코드북을 생성하였다.

학습 과정에서 실험 대상자들은 마이크 1번으로 3번 발성함으로써 화자의 모델을 만들고 매일 5종류의 마이크로 1번씩 인증 실험을 수행하였다.

표 1-5는 남/여 각 30명의 실험 화자에 대한 헤드셋 마이크 종류별 실험 결과 비교표 이다. 켑스트럼 평균 차감법 (CMS)의 경우 특성이 다른 헤드셋 마이크 3, 4, 5로 발성한 경우 본인을 거부하는 거부 오류율이 평균

표 3. 기존의 방법과 제안한 방법의 인식율 비교 (마이크 1로 학습/마이크 3로 인식)

Tabel 3. Performance comparison between CMS and proposed method (Mic. 1-Enrollment, Mic. 3-verification).

Mic. 1(Enrollment) → Mic. 3(verification)				
CMS(Cepstral Mean Subtraction)				
Each 30 speakers	Male		Female	
	FR	FA	FR	FA
total	156/600	42/17400	139/600	83/17400
%	26.0	0.24	23.16	0.48
total error rate	FR: 24.58%, FA: 0.36%			
Proposed channel mis-match compensation				
Each 30 speakers	Male		Female	
	FR	FA	FR	FA
total	64/600	68/17400	42/600	89/17400
%	10.66	0.39	7.0	0.51
total error rate	FR: 8.83%, FA: 0.45%			

표 4. 기존의 방법과 제안한 방법의 인식율 비교 (마이크 1로 학습/마이크 4로 인식)

Tabel 4. Performance comparison between CMS and proposed method (Mic. 1-Enrollment, Mic. 4-verification).

Mic. 1(Enrollment) → Mic. 4(verification)				
CMS(Cepstral Mean Subtraction)				
Each 30 speakers	Male		Female	
	FR	FA	FR	FA
total	234/600	22/17400	228/600	26/17400
%	39.0	0.13	38.0	0.15
total error rate	FR: 38.5%, FA: 0.14%			
Proposed channel mis-match compensation				
Each 30 speakers	Male		Female	
	FR	FA	FR	FA
total	72/600	34/17400	68/600	41/17400
%	12.0	0.19	11.3	0.23
total error rate	FR: 11.66%, FA: 0.21%			

표 5. 기존의 방법과 제안한 방법의 인식율 비교 (마이크 1로 학습/마이크 5로 인식)

Tabel 5. Performance comparison between CMS and proposed method (Mic. 1-Enrollment, Mic. 5-verification).

Mic. 1(Enrollment) → Mic. 5(verification)				
CMS(Cepstral Mean Subtraction)				
Each 30 speakers	Male		Female	
	FR	FA	FR	FA
total	302/600	18/17400	247/600	23/17400
%	50.33	0.10	41.16	0.13
total error rate	FR: 45.75%, FA: 0.12%			
Proposed channel mis-match compensation				
Each 30 speakers	Male		Female	
	FR	FA	FR	FA
total	103/600	18/17400	88/600	22/17400
%	12.0	0.19	11.3	0.23
total error rate	FR: 15.91%, FA: 0.11%			

36.27% 정도로 급격히 증가하는 것을 볼 수 있다. 반면, 제안한 채널 불일치 보상 기법을 사용한 경우 특성이 다른 헤드셋 마이크를 사용하여도 평균 12.13%의 거부 오류율을 을 나타냄을 확인할 수 있다. 또한, 사칭자 인식 오류율은 0.45%에서 0.49%로 다소 증가하나 EER (Equal Error Rate)의 관점에서 보아 성능이 크게 향상 되었음을 알 수 있다.

그림 3과 4는 켑스트럼 평균 차감법과 제안한 채널 불일치 보상 기법을 사용한 마이크별 본인 거부 오류율 및 사칭자 인식 오류율을 나타낸 것이다. 마이크 3, 4, 5번에 대해 본인 거부 오류율이 제안한 기법을 사용했을 때 평균 24.14% 감소한 것을 확인할 수 있다.

V. 결 론

실제 화자확인 시스템에서 불일치조건 (Mis-match condition)은 인식율의 급격한 하락을 가져오게 된다. 특히, 마이크와 같은 불특정 선형 채널의 불일치는 화자확인 시스템에서 주변 환경 잡음이나 화자내 변이와 함께 가장 큰 오인식 요인이 된다.

따라서, 화자확인 시스템에서 이동 환경을 고려한 채널 보상에 대한 연구는 매우 중요한 요소이다. 본 논문에서는 학습과정에서 공통 코드북의 평균값과 개인 코드북의 평균값과의 바이어스를 보상하고, 인식과정에서 시그모이드 함수를 이용한 비선형적인 바이어스 보상기법을 제안하였다.

모의실험 결과, 5가지 종류의 헤드셋 마이크에서 수집한 데이터베이스에 대해, 학습한 마이크와 인증시험에 사용되는 마이크가 동일한 경우에는 기존의 채널 보상 방법인 켑스트럼 평균 차감법을 사용한 경우나 채널보상 알고리즘을 적용한 경우나 거의 비슷한 인식율을 보인다. 그러나, 인증 시험 시 다른 마이크를 사용하여 테스트한 결과 둘 사이에는 상당한 성능의 차이를 보인다. 특히, 마이크1로 학습한 모델에 대해 마이크 3, 4, 5로 인증 실험을 할 경우 본인 거부 오류율이 제안한 기법을 사용했을 때 평균 24.14% 감소한 것을 확인할 수 있다. 결론적으로, 1초 미만의 소량의 음성 데이터 만으로 확인 과정을 수행해야 하는 문맥 종속 화자 확인 시스템에서는 대표적 채널 보상 기법인 CMS보다 본 논문에서 제안한 채널 보상 기법이 우수한 성능을 나타냄을 확인할 수 있다.

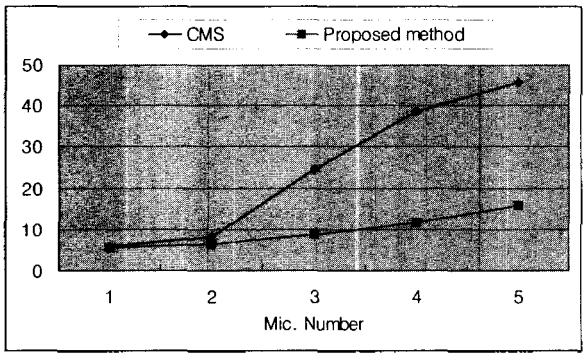


그림 3. CMS와 제안한 방법의 마이크별 본인 거부 오류율
Fig. 3. FRR in CMS and proposed method per Mic.

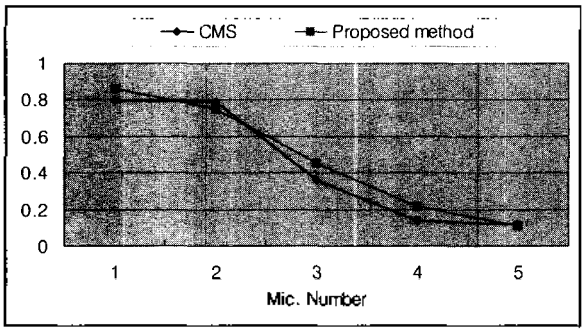


그림 4. CMS와 제안한 방법의 마이크별 사칭자 오인식율
Fig. 4. FAR in CMS and proposed method per Mic.

감사의 글

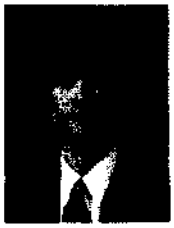
본 논문은 2002년도 광운대학교 교내학술연구비 지원으로 수행되었음.

참 고 문 헌

1. M. Omologo, M. Matassoni, P. Svaizer, D. Giuliani, "Microphone array based speech recognition with different talker-array positions", Proc. ICASSP97, 227-230, 1997.
2. A.E.Rosenberg and F.K.Soong, "Recent research in automatic speaker recognition," Advances in Speech Signal Processing, 1992, 701-738.
3. Jayant M. Naik, "Speaker Verification: A Turorial", IEEE Communication Magazine, January 1990.
4. M. W. Mak and S. Y. Kung, "Robust speaker verification over the telephone by feature recuperation", Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of 2001 International Symposium on, 433-436, 2001.
5. Atal, B.S., "Effectiveness of Linear Prediction Characteristics of the Speech Wave for Automatic Speaker Identification and Verification", Journal of the Acoustical Society of America, 1974, **55**(6), 1304-1312.
6. D. Naik, "Pole-filtered cepstral mean subtraction", Proceedings ICASSP-1995, 1:157-160, 1995.

저자 약력

• 정 희 석 (Hee-Suk Jeong)



1996년 8월: 광운대학교 전자통신공학과(공학사)
1998년 8월: 광운대학교 일반대학원 전자통신공학과(공학석사)
1999년 3월~현재: 광운대학교 전자통신공학과 박사과정
2002년 1월~현재: (주)한국파워보이스 대표이사
※주요관심분야: 음성인식, 화자인식, 적응신호처리

• 강 철 호 (Chul-Ho Kang)

1975년 2월: 한양대학교 전자공학과 졸업 (공학사)
1979년 2월: 서울대학교 대학원 전자공학과 졸업 (공학석사)
1988년 2월: 서울대학교 대학원 전자공학과 졸업 (공학박사)
1977년 3월~1982년 2월: 국방과학연구소 연구원
1991년 1월~1992년 1월: 미국 일리노이대학교 객원교수
2000년 3월~2001년 2월: 중국 연변 과학기술대학교 교환교수
1983년 3월~현재: 광운대학교 전자통신공학과 정교수
※주요관심분야: 음성신호처리, 적응신호처리, 통신신호처리