한국음향학회지 제28권 제3호 pp. 279~284 (2009)

# 서브밴드 가중치를 이용한 잡음에 강인한 화자검증

# Noise Robust Speaker Verification Using Sub-Band Weighting

김 성 탁*, 지 미 경*, 김 회 린*
(Sungtak Kim, Mikyong Ji, Hoirin Kim)

*한국정보통신대학교 공학부

(접수일자: 2009년 1월 13일; 수정일자: 2009년 2월 23일; 채택일자: 2009년 3월 2일)

화자검증은 발성화자가 제시화자 (claimed speaker)인지 아닌지를 구별하는 것이다. 기존의 화자검증 시스템인 GMM-UBM 방식의 화자검증 시스템은 무잡음 환경에서는 높은 검증성능을 보이지만, 잡음환경에서는 성능이 급격히 떨어지는 단점이 있다. 이런 단점을 극복하기 위해 멀티밴드를 이용한 방법인 특징벡터 재결합방법이 제안되었지만, 특징벡터 재결합방법은 전체 서브밴드 특징벡터들을 사용하여 유사도를 계산하는 단점이 있다. 이런 단점을 극복하기 위해 기 발표된 이전 논문에서 각 서브밴드 유사도를 독립적으로 계산하는 변형된 특징벡터 재결합방법을 제안하였고, 본 논문에서는 변형된 특징벡터 재결합방법과 각 서브밴드들의 신뢰도를 나타내는 신호 대 잡음비를 이용한 가중치를 이용하여 잡음환경에서 기존의 특징벡터 재결합방법에 비해 에러를 28% 감소시켰다.

**핵심용어:** 화자검증, 변형된 특징벡터 재결합 방법, 서브밴드 신뢰도, 서브밴드 가중치
**투고분야:** 음성처리 분야 (2), 뉴미디어 분야 (13)

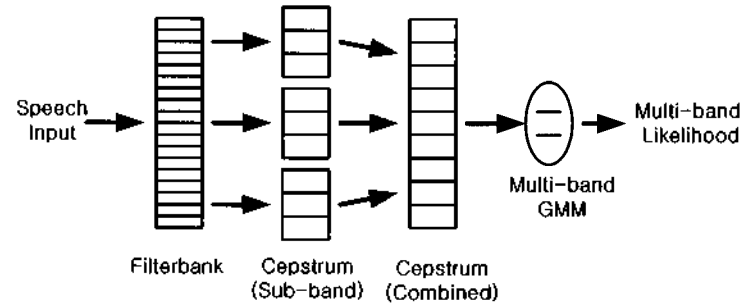
Speaker verification determines whether the claimed speaker is accepted based on the score of the test utterance. In recent years, methods based on Gaussian mixture models and universal background model have been the dominant approaches for text-independent speaker verification. These speaker verification systems based on these methods provide very good performance under laboratory conditions. However, in real situations, the performance of speaker verification system is degraded dramatically. For overcoming this performance degradation, the feature recombination method was proposed, but this method had a drawback that whole sub-band feature vectors are used to compute the likelihood scores. To deal with this drawback, a modified feature recombination method which can use each sub-band likelihood score independently was proposed in our previous research. In this paper, we propose a sub-band weighting method based on sub-band signal-to-noise ratio which is combined with previously proposed modified feature recombination. This proposed method reduces errors by 28% compared with the conventional feature recombination method.

***Keywords:*** Speaker Verification, Modified Feature Recombination, Sub-Band Reliability, Sub-Band Weighting
***ASK subject classification:*** Speech Signal Processing(2), New Media(13)

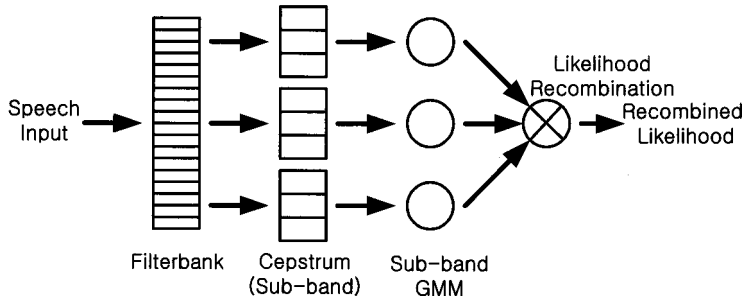
## I. 서 론

화자검증 기술은 주어진 음성신호가 제시된 화자 (claimed speaker)인지 아닌지를 판별하는 것이다. 최근에는 가우시안 혼합모델 (Gaussian mixture model)과 UBM (universal background model)을 이용한 문맥독립 화자검증 기술이 주된 추세이다. GMM-UBM기반의 화자검증 기술 [1]은 무잡음 환경에서는 높은 성능을 보장하지만, 자동차 안이나 공공장소 같은 환경에서는 주변 잡음에 의해 화자검증 시스템의 성능이 많이 저하된다. 이런 성능저하의 주된 원인은 화자모델들의 훈련환경과 검증환경의 불일치 때문에 나타난다. 훈련환경과 검증환경의 불일치 문제를 극복하기 위해 많은 기술들이 연구되고 있다. 이들 기술들은 크게 세 가지로 분류 될 수 있다: 1) 특징벡터를 이용한 기술 [2], 잡음에 왜곡된 음성을 최대한 무잡음 음성과 같게 변형하거나 음성의 질을 높이는 기술; 2) 화자모델을 이용한 기술 [3], 화자모델을 잡음에 왜곡된 음성의 확률적 특성과 일치하도록 변형하거나 적응시키는 기술; 3) 스코어를 이용한 기술 [4], 잡

책임저자: 김 성 탁 (stkim@icu.ac.kr)
대전광역시 유성구 문지로119번지 한국정보통신대학교
(전화 042-866-6221; 팩스: 042-866 6245)




(a) 특징벡터 재결합방법 (Feature recombination)




(b) 유사도 재결합방법 (Likelihood recombination)

그림 1. 다중밴드 방법
Fig. 1. Multi-Band Approach.

음의 영향을 최소화하기 위해 화자모델의 출력 스코어를 조절하는 기술. 여러 가지 기술들 중 특징벡터를 이용한 기술인 변형된 특징벡터 재결합방법 [5-7]과 적응형 잡음모델을 이용한 신뢰성 높은 서브밴드 선택방법 [7]이 제안되었다. 변형된 특징벡터 재결합방법은 기존의 특징벡터 재결합방법의 전체 서브밴드 특징벡터들을 이용하여 유사도를 구하는 단점을 해결한 것이고, 적응형 잡음모델을 이용한 신뢰성 높은 서브밴드 선택방법은 검증환경에서 구한 잡음 특징벡터를 MAP (Maximum A Posteriori) 방법을 이용하여 얻은 적응형 잡음모델의 유사도 값과 화자모델의 유사도 값을 비교하여 화자모델의 유사도 값이 큰 신뢰성 높은 서브밴드를 사용하는 방법이다 [7].

화자모델링을 위한 특징벡터로는 MFCC (Mel-frequency Cepstral Coefficients)를 많이 이용한다. 기존의 MFCC를 구하는 방법은 필터뱅크 출력 전체를 이용한다. 전체 필터뱅크 출력들을 이용하여 특징벡터를 구하는 경우, 음성신호가 비록 주파수 영역이 제한된 잡음 (band-limited noise)으로 왜곡이 되더라도 전체 특징벡터 성분에 영향을 주게 된다. 이런 문제점을 극복하기 위해 멀티밴드 (multi-band)방법이 제안되었다. 그림 1에서와 같이 멀티밴드방법에는 크게 특징벡터 재결합 (feature recombination)방법과 유사도 재결합 (likelihood recombination)방법으로 나누어진다. 그림에서 보듯이 특징벡터 재결합방법은 각 서브밴드 별로 구한 특징벡터를 재결합한 후, 화자모델을 구하고 검증한다. 유사도 재결합방법은 각 서브밴드 별로 특징벡터를 구하고, 각 서브밴드 별로 구한 특징벡터를 이용하여 서브밴드마다 화자모델을 구축한다. 화자검증은 각 서브밴드 별로 구한 유사도들을 재결합하여 이용한다. 기존의 멀티밴드방법의 단점은 음성신호가 주파수 영역이 제한된 잡음환경에서는 높은 성능향상을 보여주지만, 광대역 잡음 (broad-band noise)에 의해 왜곡된 경우에는 성능향상에 기여를 못한다. 하지만, 비록 음성신호가 광대역 잡음에 의해 왜곡된 경우라도 각 서브밴드마다 왜곡의 정도는 다르다. 이런 점을 감안하면, 서브밴드 별로 독립적인 프로세스가 가능한 멀티밴드 방법이 광대역 잡음환경에서도 여전히 유용한 방법임을 알 수 있다. 하지만, 기존의 특징벡터 재결합은 유사도를 구하는 과정에서 서브밴드 특징벡터들을 모두 이용하여 유사도를 구한다. 이렇게 구한 유사도는 각 서브밴드들의 왜곡 정도를 반영하지 못한다. 본 논문에서는 기존 방법의 단점을 극복하기 위해 기 발표된 변형된 특징벡터 재결합방법 [7]을 소개하고, 서브밴드 신호 대 잡음비를 이용한 서브밴드 가중치를 이용한 잡음에 강인한 화자검증 방법을 제안한다.

## II. 기존의 화자검증 시스템

GMM-UBM 기반의 문맥독립 화자검증 방법은 간단하면서도 성능이 우수한 것으로 알려져 있다. 화자가 발성한 음성의 특징벡터 $X=\{x_1,x_2,\cdots,x_T\}$가 주어졌을 때, 화자검증은 특징벡터 $X$가 제시된 화자로부터 발성되었는지를 결정하는 것이다. 제시된 화자를 수락하거나 거절하기 위해, GMM-UBM 기반의 화자검증에서는 아래와 같은 유사도 비를 이용한다.

$$\frac{P(X\mid\lambda_C)}{P(X\mid\lambda_{UBM})}\begin{cases}\geq accept\\ < reject\end{cases} \tag{1}$$

$$P(X\mid\lambda_C)=\frac{1}{T}\sum_{t=1}^{T}\log\left[p(x_t\mid\lambda_C)\right] \tag{2}$$

$$P(X\mid\lambda_{UBM})=\frac{1}{T}\sum_{t=1}^{T}\log\left[p(x_t\mid\lambda_{UBM})\right] \tag{3}$$

여기서 $\lambda_C$와 $\lambda_{UBM}$은 각각 제시된 화자모델과 UBM이다.

## III. 서브밴드 가중치와 변형된 특징벡터 재결합을 이용한 화자검증 시스템

기존의 GMM-UBM기반의 화자검증 시스템은 무잡음 환경에서는 높은 성능을 보여주지만, 잡음환경에서는 급격히 성능이 저하된다. 이런 성능저하의 주된 원인은 잡음으로 인한 훈련환경과 검증환경의 불일치 때문이다. 본 논문에서는 잡음에 의한 불일치문제를 해결하기 위해 서브밴드 신호 대 잡음비 (SNR: signal-to-noise ratio)를 이용한 서브밴드 가중치방법을 사용하였다. 잡음 에너지는 비 음성 프레임들의 평균 에너지를 사용하였다. 각 프레임을 음성과 비 음성 프레임으로 판별하는 방법은 처음 10개의 프레임들의 평균 에너지보다 크면 음성 프레임으로, 작으면 비 음성 프레임으로 간주하였다. $t$ 번째 프레임의 신호 대 잡음비는 식 (4)와 같다. 식 (5)는 주파수 차감법 (Spectral Subtraction)에서 사용하는 방법을 이용하여 잡음이 섞인 음성신호의 에너지 $|X_t(k)|$와 잡음의 평균에너지 $|\overline{N}(k)|$를 이용하여 무잡음 음성신호의 에너지 $|S_t(k)|$를 구하는 방법을 보여준다. 잡음신호의 평균에너지, $|\overline{N}(k)|$는 처음 10개 프레임들의 평균 에너지를 사용하였다. 여기서 $k$, $|X_t(k)|$, $|S_t(k)|$, 그리고 $|N(k)|$는 각각 주파수 인덱스, 잡음신호의 에너지 절대값, 추정된 무잡음 신호의 에너지 절대값, 그리고 잡음의 에너지 절대값을 나타낸다.

$$SNR_t^{Full}=10\log_{10}\left[\frac{\sum_{k=1}^{K}|S_t(k)|^2}{\sum_{k=1}^{K}|\overline{N}(k)|^2}\right] \tag{4}$$

$$|S_t(k)|=\max\left\{|X_t(k)|-1.1|\overline{N}(k)|,0.001|\overline{N}(k)|\right\} \tag{5}$$

서브밴드 시스템에서 $t$ 번째 프레임의 $i$ 번째 서브밴드 신호 대 잡음비는 식 (6)를 이용하여 구할 수 있다.

$$SNR_t^{(i)}=10\log_{10}\left[\frac{\sum_{k\in Sub\text{-}band\ i}|S_t(k)|^2}{\sum_{k\in Sub\text{-}band\ i}|\overline{N}(k)|^2}\right] \tag{6}$$

식 (4)와 (6)에서 구한 신호 대 잡음비를 이용하여 식 (7)과 (8)과 같이 가중치를 구하는 간단한 방법인 sigmoid 함수를 이용하여 신호 대 잡음비에 따른 전체밴드 가중치와 서브밴드 가중치를 구하였다. 신호 대 잡음비가 30 dB이상이면 거의 1.0에 가까운 가중치를 가지고, 0 dB이하에서 거의 0에 가까운 가중치를 가진다.

$$\rho_t^{(Full)}=\frac{1.0}{1.0+\exp[-0.5(SNR_t^{Full}-15.0)]} \tag{7}$$

$$\rho_t^{(i)}=\frac{1.0}{1.0+\exp[-0.5(SNR_t^{(i)}-15.0)]} \tag{8}$$

식 (1)에서 (3)과 식 (7)을 이용한 전체밴드 가중치를 이용한 화자검증은 식 (9)에서 식 (11)과 같다.

$$\frac{\widetilde{P}(X\mid\lambda_C)}{\widetilde{P}(X\mid\lambda_{UBM})}\begin{cases}\geq accept\\ < reject\end{cases} \tag{9}$$

$$\widetilde{P}(X\mid\lambda_C)=\frac{1}{T}\sum_{t=1}^{T}\log\left[p(x_t\mid\lambda_C)\right]\cdot\rho_t^{(Full)} \tag{10}$$

$$\widetilde{P}(X\mid\lambda_{UBM})=\frac{1}{T}\sum_{t=1}^{T}\log\left[p(x_t\mid\lambda_{UBM})\right]\cdot\rho_t^{(Full)} \tag{11}$$

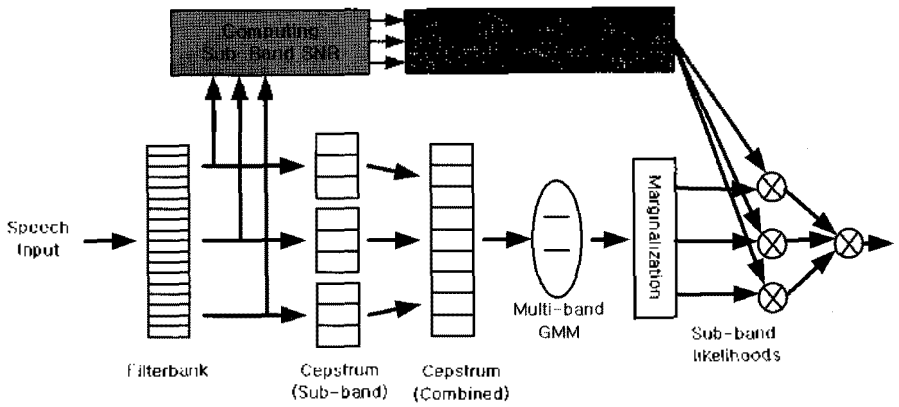



그림 2. 서브밴드 가중치를 이용한 변형된 특징벡터 재결합방법
Fig. 2. Modified Feature Recombination using Sub-Band Weighting.

본 논문에서 제안된 서브밴드 가중치와 변형된 특징벡터 재결합방법을 이용한 화자검증은 아래 식 (12)에서 식 (14)와 같다.

$$\frac{\tilde{P}(X \mid \lambda_c)}{\tilde{P}(X \mid \lambda_{UBM})} \begin{cases} \geq accept \\ < reject \end{cases} \tag{12}$$

$$\tilde{P}(X \mid \lambda_C) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{M}\log\left[p(x_t^{(i)} \mid \lambda_C)\right]\cdot \rho_t^{(i)} \tag{13}$$

$$\tilde{P}(X \mid \lambda_{UBM}) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{M}\log\left[p(x_t^{(i)} \mid \lambda_{UBM})\right]\cdot \rho_t^{(i)} \tag{14}$$

위 식 (12)와 (14)에서 서브밴드 유사도는 기존의 특징벡터 재결합방법에서 각 서브밴드가 독립적이라는 가정과 전체 서브밴드 특징벡터들로 훈련된 화자모델이나 UBM모델로부터 특정 서브밴드의 유사도값은 특정서브밴드를 제외한 나머지 서브밴드들을 marginalization을 이용하여 구할 수 있다 [5]-[7]. 그림 2는 서브밴드 가중치를 이용한 변형된 특징벡터 재결합방법을 보여준다.

그림 2에서 보듯이 변형된 특징벡터 재결합방법은 그림 1의 기존의 특징벡터 재결합방법과 달리 각 서브밴드가 독립적이라는 가정과 특정 서브밴드를 제외한 나머지 서브밴드들을 marginalization한 후, 특정 서브밴드 유사도를 구할 수 있다. 그리고 본 논문에서는 서브밴드 신호 대 잡음비를 이용한 서브밴드 가중치를 변형된 특징벡터 재결합방법에 적용하여 잡음환경에 강인한 화자검증 시스템을 제안하였다.

# IV. 실험 및 결과

## 4.1. 실험환경

본 논문에선 제안된 알고리즘의 성능을 평가하기 위해 TIMIT 데이터베이스 [8]를 사용하였다. 등록화자는 남자 100명과 여자 100명으로 총 200명을 사용하였고, 사칭자를 위해 남자 158명 여자 42명으로 총 200명을 사용하였다. TIMIT 데이터베이스는 화자 당 10문장을 발성하였는데 5문장은 화자모델 훈련에 사용하였고 나머지 5문장은 테스트에 사용하였다. 화자모델 훈련을 위해 논문에서 MAP 적응방법을 이용하였다. 화자검증이나 MAP에 이용한 UBM은 남자 50명과 여자 50명으로 총 100명으로 훈련하였다. 잡음환경에서의 화자검증을 위해 TIMIT 데이터베이스의 음성들을 8 kHz로 downsampling한 후, Aurora 2 데이터베이스 [9]의 8가지 잡음 (airport, babble, car, exhibition, restaurant, street, subway, train)을 여러 가지 신호 대 잡음비 (SNR)로 음성을 왜곡시켰다. 실험에 사용한 화자모델과 UBM은 160개의 mixture를 가지는 가우시안 혼합모델을 사용하였다. 기존의 전체 주파수 밴드를 이용한 특징벡터 추출방법에선 33개의 필터를 사용하는 필터뱅크를 이용하여 18차의 MFCC를 추출하였다. 기존의 특징벡터 재결합 방법과 제안된 변형된 특징벡터 재결합 방법에서는 33개의 필터를 사용하는 필터뱅크를 3개의 서브밴드로 나누고 각 서브밴드 당 6차의 MFCC를 추출하였다.

## 4.2. 실험결과

화자검증 시스템의 동작특성을 나타내기 위해 2 가지 에러를 사용하였다. False acceptance rate와 false rejection rate이다. 화자검증 시스템의 임계값을 변화시킴에 따라 이

표 1. 특징벡터 재결합방법과 변형된 특징벡터 재결합방법의 화자검증 성능
Table 1. The Performances of Conventional Feature Recombination and a Proposed Modified Feature Recombination.

| System / SNR | Full-Band System EER (%) | Multi-Band Systems EER (%) | |
|---|---|---|---|
| | | Feature Recombination | Modified Feature Recombination |
| 20 dB | 4.93 | 5.46 | 4.64 |
| 15 dB | 8.13 | 7.89 | 6.66 |
| 10 dB | 13.66 | 12.21 | 10.58 |
| 5 dB | 20.83 | 18.81 | 16.44 |
| 0 dB | 28.71 | 27.30 | 24.59 |
| Average Error Reduction Rate (%) | | 3.44 | 16.37 |

표 2. 전체밴드 가중치를 이용한 특징벡터 재결합방법과 서브밴드 가중치를 이용한 변형된 특징벡터 재결합방법의 화자검증 성능
Table 2. The Performances of Conventional Feature Recombination using Full-Band Weighting and a Proposed Modified Feature Recombination using Sub-Band Weighting.

| Sys. / SNRs | Feature Recombination EER (%) | Feature Recombination EER (%) | | Modified Feature Recombination | |
|---|---|---|---|---|---|
| | | + Full-Band Dropping | + Full-Band Weighting | + Sub-Band Dropping | + Sub-Band Weighting |
| 20 dB | 5.46 | 5.80 | 5.16 | 4.06 | 4.05 |
| 15 dB | 7.89 | 7.84 | 7.33 | 5.63 | 5.46 |
| 10 dB | 12.21 | 11.69 | 11.30 | 8.88 | 8.59 |
| 5 dB | 18.81 | 18.29 | 17.93 | 14.03 | 12.89 |
| 0 dB | 27.30 | 26.64 | 26.45 | 23.88 | 19.89 |
| Average Error Reduction Rate (%) | | 0.77 | 5.57 | 23.91 | 28.99 |

두 가지 에러 값이 같은 값을 가질 때 EER (Equal Error Rate)이라 한다. 표 1은 기존의 특징벡터 재결합 방법과 기 발표된 이전 논문[7]에서 제안된 변형된 특징벡터 재결합 방법의 성능을 보여준다.

표 1의 실험결과를 보면, 기존의 특징벡터 재결합 방법의 경우는 전체밴드를 이용한 방법에 비해 평균 에러 감소율이 3.44%인 반면, 변형된 특징벡터 재결합 방법은 16.37%로 잡음환경에서 화자검증 성능향상에 기여함을 알 수 있다. 표 2는 제안한 서브밴드 가중치를 이용한 변형된 특징벡터 재결합방법의 화자검증 성능을 보여준다. 서브밴드 가중치를 이용한 화자검증 기술과 비교실험을 하기 위해 신호 대 잡음비가 30 dB이상인 프레임이나 서브밴드만 화자검증에 사용하는 전체밴드 제외 (full-band dropping)와 서브밴드 제외 (sub-band dropping)를 실험하였다. 표 2에서 잡음환경이 20 dB이하인 경우만 있지만, 이 수치는 평균 신호 대 잡음비를 나타내는 것이기 때문에 30 dB이상인 프레임이나 서브밴드가 테스트 발성에 존재한다. 실험결과를 보면 변형된 특징벡터 재결합방법에 서브밴드 제외나 서브밴드 가중치를 사용할 경우가 기존의 특징벡터 재결합방법에 프레임 제외나 전체밴드 가중치를 사용할 때보다 높은 성능을 보여주었다. 실험을 통하여 제안한 변형된 특징벡터 재결합방법과 서브밴드 가중치가 잡음환경에서 화자검증 시스템의 성능을 향상시켜줌을 알 수 있었다.

## V. 결 론

본 논문에서는 기존의 특징벡터 재결합 방법에서 모든 서브밴드 특징벡터를 이용하여 하나의 유사도를 구하는 과정의 단점을 극복하고자, 기 발표된 각 서브밴드가 독립적이라는 가정과 marginalization을 이용하여 각 서브밴드 유사도를 계산하는 변형된 특징벡터 재결합 방법과 서브밴드의 신뢰도를 나타내는 서브밴드 신호 대 잡음비를 이용한 가중치를 사용하여 잡음에 강인한 화자검증 시스템을 제안하였다. 그 결과 기존의 특징벡터 재결합 방법에 가중치를 이용하는 것보다 각 서브밴드 별로 가중치를 이용할 수 있는 변형된 특징벡터 재결합방법이 잡음에 더 강인함을 알 수 있었다.

## 참 고 문 헌


1. D. Reynold, T. Quatieri, and R. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, Nos. 1-3, vol. 10, pp. 19-41, 2000.
2. A. Drygajlo and M. El-Maliki, "Speaker verification in noisy environments with combined spectral subtraction and missing feature theory," *In Proc. ICASSP*, vol. 2, pp. 121-124,1998.
3. K. Yiu, M. Mak, and S. Kung, "Environment adaptation for robust speaker verification," In *Proc. EUROSPEECH*, pp. 2973-2976, 2003.
4. C. Barras and J. Gauvain, "Feature and score normalization for speaker verification of cellular data," In *Proc. ICASSP*, vol. 2, pp. 49-52, 2003.
5. S. Kim, M. Ji, Y. Suh, and H. Kim, "Noise Robust Speaker Identification using Sub-Band Weighting in Multi Band Approach," *IEICE Trans. Inf. & Syst.*, E90-D vol. 12, pp. 2110-2114, 2007.
6. S. Kim, M. Ji, and H. Kim, "Noise Robust Speaker Recognition using Sub-Band Likelihoods and Reliable Feature Selection," *ETRI Journal*, vol. 30, no. 1, pp. 89-100, 2008.
7. 김성탁, 지미경, 김회린,"신뢰성 높은 서브밴드 특징벡터 선택을 이용한 잡음에 강인한 화자검증," *말소리*, 제63호, 125-137쪽, 2007.
8. TIMIT database, *TIMIT acoustic-phonetic speech corpus*, National Institute of Standards and Technology (NIST), NIST speech disk, 1990.
9. D. Pearce and H. Hirsch, "The aurora experimental framework for the performance evaluation of speech recognition systems under noise conditions," in *Proc. ICSLP*, vol. 4, pp. 29-32, 2000.


## 저자 약력


**•김 성 탁 (Sungtak Kim)**

2000년 2월: 울산대학교 전자공학과 (학사)
2003년 8월: 한국정보통신대학교 공학부 (석사)
2008년 8월: 한국정보통신대학교 공학부 (박사)
※주관심분야: 음성처리, 음성인식, 화자인식

**•지 미 경 (Mikyong Ji)**

2000년 2월: 한성대학교 전산과 (학사)
2002년 3월: 한국정보통신대학교 공학부 (석사)
2008년 4월: 한국정보통신대학교 공학부 (박사)
※주관심분야: 음성처리, 음성인식, 화자인식

**•김 회 린 (Hoirin Kim)**

1984년 2월: 한양대학교 전자공학과 (학사)
1987년 2월: 한국과학기술연구원 전자공학과 (석사)
1992년 2월: 한국과학기술연구원 전자공학과 (박사)
1987년 10월~1999년 12월: ETRI 선임연구원
1994년 6월~1995년 5월: 일본 ATR-ITL 방문연구원
2006년 8월~2007년 7월: 미국 UCSD 방문교수
2001년 1월~현재: 한국정보통신대학교 공학부 부교수
※주관심분야: 음성인식, 화자인식, 음향코딩, 음향정보 인덱싱