

신뢰성 높은 서브밴드 특징벡터 선택을 이용한 잡음에 강인한 화자검증

김성탁, 지미경, 김희린(ICU)

<차 례>

- | | |
|--|------------|
| 1. 서론 | 5. 실험 및 결과 |
| 2. 다중밴드 MFCC를 이용한 서브밴드 유사도 계산 | 5.1 실험환경 |
| 3. 기존의 화자검증 시스템 | 5.2 실험결과 |
| 4. 신뢰성 높은 서브밴드 특징벡터 선택과 변형된 특징벡터 재결합을 이용한 화자검증 시스템 | 6. 결론 |

<Abstract>

Noise Robust Speaker Verification Using Subband-Based Reliable Feature Selection

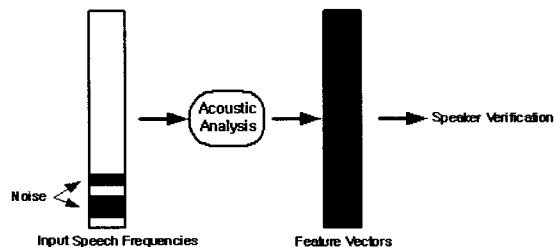
Sungtak Kim, Mikyong Ji, Hoirin Kim

Recently, many techniques have been proposed to improve the noise robustness for speaker verification. In this paper, we consider the feature recombination technique in multi-band approach. In the conventional feature recombination for speaker verification, to compute the likelihoods of speaker models or universal background model, whole feature components are used. This computation method is not effective in a view point of multi-band approach. To deal with non-effectiveness of the conventional feature recombination technique, we introduce a subband likelihood computation, and propose a modified feature recombination using subband likelihoods. In decision step of speaker verification system in noise environments, a few very low likelihood scores of a speaker model or universal background model cause speaker verification system to make wrong decision. To overcome this problem, a reliable feature selection method is proposed. The low likelihood scores of unreliable feature are substituted by likelihood scores of the adaptive noise model. In here, this adaptive noise model is estimated by maximum a posteriori adaptation technique using noise features directly obtained from noisy test speech. The proposed method using subband-based reliable feature selection obtains better performance than conventional feature recombination system. The error reduction rate is more than 31% compared with the feature recombination-based speaker verification system.

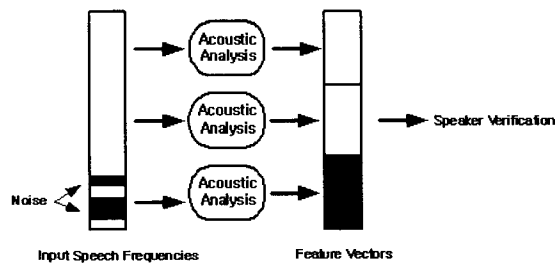
* Keywords: Speaker verification, Feature recombination, Subband likelihood, Modified feature recombination, Reliable feature selection.

1. 서론

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-UBM 기반의 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 극복하게 위해 많은 기술들이 연구되어지고 있다. 이들 기술들은 크게 세 가지로 분류될 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], 잡음의 영향을 최소화하기 위해 화자모델의 출력 스코어를 조절하는 기술. 본 논문에서는 특징벡터를 이용한 기술인 변형된 특징벡터 재결합방법과 스코어를 이용한 기술인 신뢰성 높은 서브밴드 특징벡터 선택방법을 제안한다.



(a)



(b)

<그림 1> 주파수 영역이 제한된 노이즈 환경에서의 화자검증[5]
 (a) 전체 주파수밴드를 이용한 경우 (b) 다중밴드 방법을 이용한 경우

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 다중밴드(multi-band)방법이 제안되었다[5]. <그림 1>은 주파수 영역이 제한된 잡음환경에서의 화자검증을 전체밴드를 이용하는 방법과 다중밴드를 이용하는 방법을 특징벡터 관점에서 보여주고 있다. 다중밴드방법에는 크게 유사도 재결합(likelihood recombination) 방법과 특징벡터 재결합(feature recombination) 방법으로 나뉘어진다. 기존의 다중밴드방법의 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, 서브밴드별로 독립적인 프로세스가 가능한 다중밴드 방법이 광대역 잡음환경에서도 유용한 방법임을 알 수 있다. 기존의 특징벡터 재결합은 유사도를 구하는 과정에서 서브밴드 특징벡터들을 모두 이용하여 유사도를 구한다. 이렇게 구한 유사도는 각 서브밴드들의 왜곡정도가 다른 특성을 반영하지 못하는 단점이 있다. 본 논문에서는 기존 방법의 단점을 극복하기 위해 서브밴드 유사도 계산법을 소개하고, 이 서브밴드 유사도를 이용한 변형된 특징벡터 재결합 방법을 제안한다.

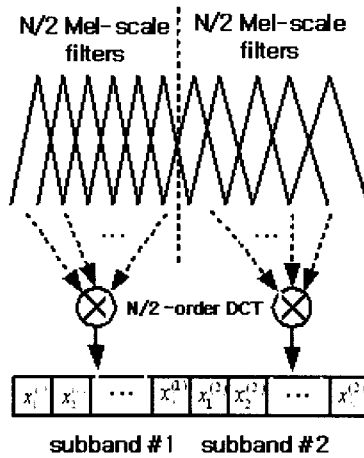
기존에 연구되어진 스코어를 이용한 기술은 world-model이나 cohort model을 이용해서 스코어를 정규화시키는데 중점을 두었다[6][7]. 하지만 본 논문에서는 스코어 정규화보다는 적응형 잡음모델을 이용하여 입력 특징벡터들의 신뢰도를 측정하고 신뢰도가 낮을 경우 스코어 값을 변경하여 불일치 문제를 해결하였다. 화자검증 시스템은 사용자의 입력 특징벡터들을 이용하여 제시된 화자모델에 대한 평균 유사도 값과 UBM에 대한 평균 유사도값의 비를 임계값과 비교하여 임계값보다 크면 수락하고, 작으면 거절한다. 하지만, 잡음환경에서의 화자검증 시, 제시된 화자모델이나 UBM에 대한 유사도들 중에서 매우 낮은 몇 개의 유사도들 때문에 수락해야 할 화자를 거절하거나 거절해야 할 화자를 수락하는 경우가 발생한다. 만약 이런 문제를 발생시키는 신뢰성이 낮은 특징벡터를 선택하고, 낮은 유사도 값을 조정한다면 성능향상에 기여할 수 있을 것이다. 논문에서는 신뢰성이 낮은 특징벡터를 선택하기 위해 적응형 잡음 모델을 이용하였다. GMM-UBM 화자검증 시스템에서 입력특징벡터의 화자모델이나 UBM의 유사도 값이 적응형 잡음모델의 유사도보다 낮으면 신뢰성이 낮은 특징벡터로 간주하고 화자모델이나 UBM의 유사도 대신 잡음모델의 유사도로 대체한다. 더 나아가, 신뢰성 높은 유사도 선택을 변형된 특징벡터 재결합방법에 적용하여서 보다 높은 화자검증 방법을 제안하였다.

2. 다중밴드 MFCC를 이용한 서브밴드 유사도 계산

M 개의 서브밴드와 N 개의 필터를 사용하는 시스템에서 각 서브밴드 당 L 개의 MFCC를 추출한다면, i 번째 서브밴드의 j 번째 MFCC를 구하는 방법은 식 (1)과 같다[8].

$$x_j^{(i)} = \sqrt{\frac{2}{NM}} \sum_{n=1}^{N/M} LFB_n^{(i)} \cos\left((n-0.5)\frac{j\pi}{N/M}\right), 1 \leq j \leq L \leq \frac{N}{M} \quad (1)$$

여기서 $LFB_n^{(i)}$ 는 i 번째 서브밴드의 n 번째 필터뱅크 로그 에너지값이다. <그림 2>는 두 개의 서브밴드와 N 개의 필터를 사용하는 멀티밴드 시스템에서의 멀티밴드 MFCC의 추출방법을 보여준다.



<그림 2> 두 개의 서브밴드를 사용하는 멀티밴드 시스템에서 MFCC 추출방법[6]

<그림 3>은 기존의 특징벡터 재결합 방법과 서브밴드 유사도를 이용한 변형된 특징벡터 재결합 방법을 보여준다. 기존의 특징벡터 재결합 방법의 경우에는 multi-band GMM으로 부터 하나의 multi-band likelihood 값을 계산하지만, 변형된 특징벡터 재결합 방법의 경우에는 모든 서브밴드 특징벡터들을 이용하여 훈련된 multi-band GMM으로 부터 각 서브밴드 유사도 값(subband likelihood)을 구한다. 각 서브밴드 유사도 값을 구하기 위해, 각 서브밴드 특징벡터들이 확률적으로 독립이라는 가정 하에 marginalization을 이용하였다. 자세한 설명은 아래와 같다.

M 개의 서브밴드를 사용하는 멀티밴드 시스템에서 X 는 각 서브밴드 특징벡터로 나누어진다. 즉 $X = \{x^1, x^2, \dots, x^M\}$ 이다. 그리고 각 특징벡터들이 확률적으로 독립

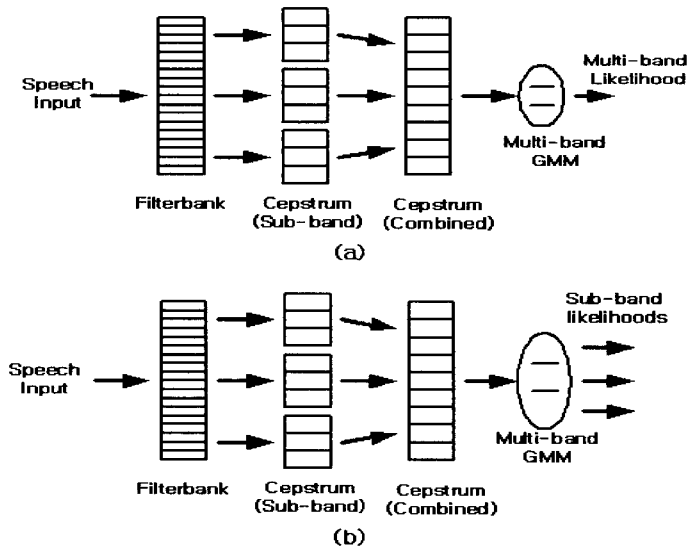
림이라고 가정하면 식 (2)에서와 같이 각 서브밴드들의 유사도 값들의 곱으로 표현된다.

$$p(X|\lambda) = \sum_{w=1}^W C_{w,\lambda} p(x^1, x^2, \dots, x^M|w, \lambda) = \sum_{w=1}^W C_{w,\lambda} \prod_{i=1}^M p(x^i|w, \lambda) \quad (2)$$

여기서 λ 와 $C_{w,\lambda}$ 는 각각 GMM과 mixture weight를 나타내고, W 는 GMM이 가지고 있는 mixture의 수를 나타낸다. 식 (2)에서 구하고자하는 서브밴드를 제외한 나머지 밴드들에 대해 marginalization을 적용하면 특정밴드의 서브밴드 유사도를 구할 수 있다.

$$p(x^i|\lambda) = \sum_{w=1}^W C_{w,\lambda} p(x^i|w, \lambda) \underbrace{\prod_{m=1, m \neq i}^M \int p(x^m|w, \lambda) dx^m}_1 = \sum_{w=1}^W C_{w,\lambda} p(x^i|w, \lambda) \quad (3)$$

식 (3)의 결과를 이용하여 제안된 변형된 특징벡터 재결합 방법으로 화자검증을 수행한다.



<그림 3> (a) 기존의 특징벡터 재결합 방법 (b) 변형된 특징벡터 재결합 방법

3. 기존의 화자검증 시스템

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

$$\frac{P(X|\lambda_C)}{P(X|\lambda_{UBM})} \begin{cases} \geq \theta & \text{accept} \\ < \theta & \text{reject} \end{cases} \quad (4)$$

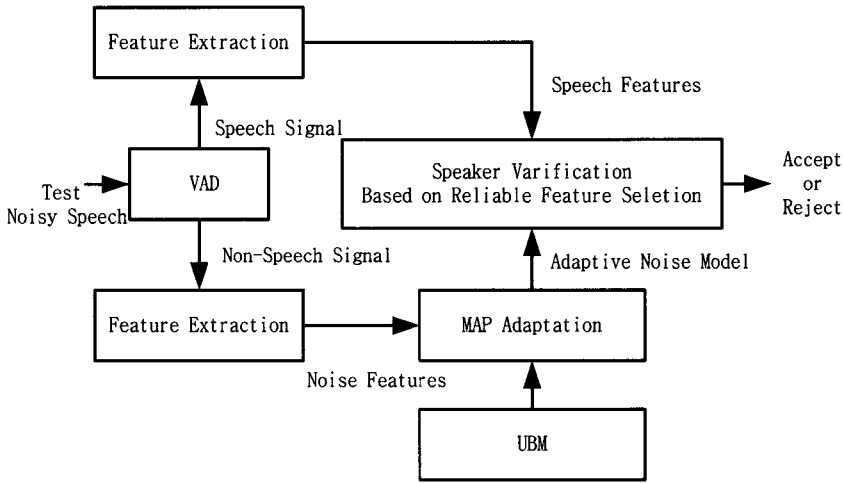
$$P(X|\lambda_C) = \frac{1}{T} \sum_{t=1}^T \log [p(x_t|\lambda_C)] \quad (5)$$

$$P(X|\lambda_{UBM}) = \frac{1}{T} \sum_{t=1}^T \log [p(x_t|\lambda_{UBM})] \quad (6)$$

여기서 λ_C 와 λ_{UBM} 은 각각 제시된 화자모델과 UBM이다.

4. 신뢰성 높은 서브밴드 특징벡터 선택과 변형된 특징벡터 재결합을 이용한 화자검증 시스템

기존의 GMM-UBM 기반의 화자검증 시스템은 무잡음 환경에서는 높은 성능을 보여주지만, 잡음환경에서는 급격히 성능이 저하된다. 이런 성능저하의 주된 원인은 훈련환경과 검증환경의 불일치이다. 불일치가 일어날 경우, 제시된 화자모델이나 UBM의 출력 유사도들 중에서 매우 낮은 유사도들 때문에 틀린 결정을 하는 경우가 발생한다. 만약 이런 매우 낮은 유사도를 선택하고 유사도 값을 보정할 수 있다면 잡음환경에서도 안정된 성능을 보장할 수 있을 것이다. 본 논문에서는 이런 신뢰성이 낮은 유사도 값을 선택하고 유사도 값을 보정하기 위해 적응형 잡음 모델을 이용하였다. 적응형 잡음모델을 얻기 위해, 논문에서는 음성검출기(voice activity detector)를 이용하여 얻은 잡음 특징벡터를 이용하여 적응형 잡음 모델을 생성한다. 논문에서 사용한 음성검출기는 테스트 음성의 처음 10 프레임의 평균 에너지를 이용하였다. 처음 10 프레임의 평균 에너지보다 크면 음성으로 간주하고 작으면 비음성으로 간주하였다. 여기서 비음성으로 간주된 프레임들로부터 추출한 특징벡터는 대략적으로 잡음 특징벡터라고 할 수 있다. 그리고 이 잡음 특징벡터들과 UBM을 MAP(Maximum A Posteriori) 적응기법을 이용하여 적응형 잡음모델을 생성한다.



<그림 4> 신뢰성 높은 특징벡터 선택을 이용한 화자검증 시스템

<그림 4>는 신뢰성 높은 특징 벡터 선택을 이용한 화자검증 시스템을 보여준다. 그림에서 보는 바와 같이 적응형 잡음모델을 현재의 테스트 잡음음성에서 추출한 잡음을 이용하여 생성하므로 잡음에 대한 추측이나 가정이 필요하지 않고 현재의 잡음특성을 잘 반영할 수 있는 장점이 있다. 주어진 특징벡터가 신뢰성 높은 특징벡터인지를 결정하기 위해 제시된 화자모델의 유사도 값이나 UBM의 유사도 값을 적응형 잡음모델의 유사도 값과 비교한다. 만약 주어진 프레임이 잡음에 의한 왜곡이 심할 경우에는 화자모델이나 UBM보다 적응형 잡음모델의 확률적 특성에 더 가깝고, 적응형 잡음 모델이 더 높은 유사도 값을 가질 것이다. 그러므로, 적응형 잡음모델의 유사도 값이 화자모델이나 UBM의 유사도 값보다 클 경우 주어진 특징벡터는 신뢰성이 낮은 것으로 간주한다. 또한 잡음에 의한 영향을 최소화하기 위해 신뢰성이 낮은 특징벡터의 유사도 값 대신 적응형 잡음모델의 유사도 값을 이용하여 화자검증을 수행한다. 전체밴드기반의 신뢰성 높은 특징벡터 선택을 이용한 GMM-UBM 기반의 화자검증 방법은 식 (7)-(11)과 같다.

$$\frac{\tilde{P}(X|\lambda_C)}{\tilde{P}(X|\lambda_{UBM})} = \begin{cases} \geq \theta & \text{accept} \\ < \theta & \text{reject} \end{cases} \quad (7)$$

$$\tilde{P}(X|\lambda_C) = \frac{1}{T} \sum_{t=1}^T \log[\hat{p}(x_t|\lambda_C)] \quad (8)$$

$$\tilde{P}(X|\lambda_{UBM}) = \frac{1}{T} \sum_{t=1}^T \log[\hat{p}(x_t|\lambda_{UBM})] \quad (9)$$

$$\hat{p}(x_t|\lambda_C) = \begin{cases} p(x_t|\lambda_C) & \text{if } p(x_t|\lambda_C) \geq p(x_t|\lambda_{Noise}) \\ p(x_t|\lambda_{Noise}) & \text{if } p(x_t|\lambda_C) < p(x_t|\lambda_{Noise}) \end{cases} \quad (10)$$

$$\hat{p}(x_t|\lambda_{UBM}) = \begin{cases} p(x_t|\lambda_{UBM}) & \text{if } p(x_t|\lambda_{UBM}) \geq p(x_t|\lambda_{Noise}) \\ p(x_t|\lambda_{Noise}) & \text{if } p(x_t|\lambda_{UBM}) < p(x_t|\lambda_{Noise}) \end{cases} \quad (11)$$

식 (3)의 서브밴드 유사도를 이용한 변형된 특징벡터 재결합 방법과 서브밴드 레벨의 신뢰성 높은 서브밴드 특징벡터를 이용한 화자검증 방법은 아래 식들과 같이 나타낼 수 있다.

$$\frac{\tilde{P}(X|\lambda_C)}{\tilde{P}(X|\lambda_{UBM})} = \begin{cases} \geq \theta & \text{accept} \\ < \theta & \text{reject} \end{cases} \quad (12)$$

$$\tilde{P}(X|\lambda_C) = \frac{1}{T} \sum_{t=1}^T \left[\sum_{i=1}^M \log(\hat{p}(x_t^i|\lambda_C)) \right] \quad (13)$$

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

$$\hat{p}(x_t^i|\lambda_C) = \begin{cases} p(x_t^i|\lambda_C) & \text{if } p(x_t^i|\lambda_C) \geq p(x_t^i|\lambda_{Noise}) \\ p(x_t^i|\lambda_{Noise}) & \text{if } p(x_t^i|\lambda_C) < p(x_t^i|\lambda_{Noise}) \end{cases} \quad (15)$$

$$\hat{p}(x_t^i|\lambda_{UBM}) = \begin{cases} p(x_t^i|\lambda_{UBM}) & \text{if } p(x_t^i|\lambda_{UBM}) \geq p(x_t^i|\lambda_{Noise}) \\ p(x_t^i|\lambda_{Noise}) & \text{if } p(x_t^i|\lambda_{UBM}) < p(x_t^i|\lambda_{Noise}) \end{cases} \quad (16)$$

5. 실험 및 결과

5.1 실험환경

본 논문에선 제안된 알고리즘의 성능을 평가하기 위해 TIMIT 데이터베이스를 사용하였다. 등록화자는 남자 100명과 여자 100명으로 총 200명을 사용하였고, 사칭자를 위해 남자 158명 여자 42명으로 총 200명을 사용하였다. TIMIT 데이터베이스는 화자 당 10문장을 발생하였는데 5문장은 화자모델 훈련에 사용하였고 나머지 5문장은 테스트에 사용하였다. 화자모델 훈련을 위해 논문에서 MAP 적용방법을 이용하였다. 화자검증이나 MAP에 이용한 UBM는 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 UBM는 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.

5.2 실험결과

화자검증 시스템의 동작특성을 나타내기 위해 두 가지 에러를 사용하였다. False acceptance rate와 false rejection rate이다. 화자검증 시스템의 임계값을 변화시킴에 따라 이 두 가지 에러 값이 같은 값을 가질 때 EER(Equal Error Rate)이라 한다. <표 1>은 기존의 특징벡터 재결합 방법과 변형된 특징벡터 재결합 방법의 성능을 보여준다.

<표 1> 특징벡터 재결합 방법과 변형된 특징벡터 재결합 방법의 화자검증 성능

System SNR (dB)	Full-band system EER (%)	Error Reduction Rate (%) over full-band system	
		Feature Recombination	Modified Feature Recombination
20	4.93	-10.91	5.83
15	8.13	2.92	18.00
10	13.66	10.61	22.60
5	20.83	9.66	21.07
0	28.71	5.00	14.37
Average ERR		3.49	16.37

<표 1>의 실험결과를 보면, 기존의 특징벡터 재결합 방법의 경우는 전체밴드를 이용한 방법에 비해 평균 에러 감소율이 3.49%인 반면, 변형된 특징벡터 재결합 방법은 16.37%로 잡음환경에서 화자검증 성능향상에 기여함을 알 수 있다.

제안한 신뢰성 높은 특징벡터 선택방법의 성능을 기존의 방법과 비교하기 위해 프레임 SNR을 이용한 신뢰성 높은 프레임 선택방법을 이용한 경우의 화자검증 성능을 <표 2>에 나타내었다. 결과를 보면 프레임 SNR 기반의 신뢰성 높은 특징벡터 선택방법이 잡음 환경에서 화자검증 성능향상에 기여함을 알 수 있다. 본 논문에서는 프레임 SNR이 30 dB 이상의 특징벡터를 신뢰성 높은 특징벡터로 간주

하고 화자검증에 사용하였다. 프레임 SNR은 아래 식들을 이용하였다.

$$SNR_t = 10 \log_{10} \left[\frac{\sum_{k=1}^K |S_t(k)|^2}{\sum_{k=1}^K |\bar{N}(k)|^2} \right] \quad (17)$$

$$|S_t(k)| = \max\{|X_t(k)| - 1.1|\bar{N}(k)|, 0.001|\bar{N}(k)|\} \quad (18)$$

여기서 k , $|X_t(k)|$, $|S_t(k)|$, 그리고 $|\bar{N}(k)|$ 는 각각 주파수 인덱스, 잡음에 왜곡된 음성의 에너지, 추정된 무잡음 음성의 에너지, 평균 잡음 에너지를 나타낸다. 프레임 SNR을 구하기 위해선 무잡음 음성 추정이 필요하다. 본 논문에서는 기존의 스펙트럼 차감방법(spectral subtraction)에서 이용한 방법인 식 (18)을 사용하였다[10]. 평균 잡음 에너지는 VAD를 이용하여 음성이 아닌 프레임들로 부터 구하였다. 본 실험에선 프레임 SNR이 30 dB 이상의 특징벡터를 신뢰성 높은 특징벡터로 간주하였다.

<표 2> SNR 기반의 신뢰성 높은 프레임 선택을 이용한 특징벡터 재결합 방법과 변형된 특징벡터 재결합 방법의 화자검증 성능

System SNR (dB)	Feature Recombination EER (%)	Error Reduction Rate (%) over full-band system	
		Feature Recombination + SNR-Based Reliable Frame Selection	Modified Feature Recombination + SNR-Based Reliable Frame Selection
20	5.46	-6.18	18.99
15	7.89	0.63	25.83
10	12.21	4.30	29.58
5	18.81	2.79	23.06
0	27.28	2.34	14.12
Average ERR		0.77	22.31

<표 3>의 실험결과를 보면 알 수 있듯이, 적응형 잡음 모델을 이용한 신뢰성 높은 특징벡터 선택이 SNR 기반의 신뢰성 높은 프레임 선택보다 성능이 우수함을

알 수 있다. <표 4>는 적응형 잡음 모델을 이용한 신뢰성 높은 특징벡터 선택에서 신뢰성이 낮은 특징벡터의 유사도를 잡음모델의 유사도로 대체하지 않고 입력에서 제외시킬 경우의 성능을 보여준다. 실험결과, 제안한 잡음모델의 유사도로 대체하는 경우가 제외시키는 경우보다 성능이 우수하였다.

<표 3> 신뢰성 높은 특징벡터 선택을 이용한 특징벡터 재결합 방법과 변형된 특징벡터 재결합 방법의 화자검증 성능

System SNR (dB)	Feature Recombination EER (%)	Error Reduction Rate (%) over full-band system	
		Feature Recombination + Fullband-based Reliable Feature Selection	Modified Feature Recombination + Subband-based Reliable Feature Selection
20	5.46	6.18	21.97
15	7.89	7.77	29.00
10	12.21	8.50	33.47
5	18.81	8.64	35.81
0	27.28	10.68	35.61
Average ERR		8.35	31.17

<표 4> 신뢰성이 높은 서브밴드 특징벡터 선택에서 유사도를 제외시키는 경우와 대체시키는 경우의 성능비교

System SNR (dB)	신뢰성 높은 서브밴드 특징벡터 선택 (EER (%))	
	제외시킬 경우	대체시킬 경우
20	4.38	4.26
15	5.96	5.60
10	8.84	8.13
5	13.36	12.08
0	18.16	17.56

6. 결 론

본 논문에서는 기존의 특징벡터 재결합 방법에서 모든 서브밴드 특징벡터를 이용하여 하나의 유사도를 구하는 과정의 단점을 극복하고자, marginalization을 이용하여 각 서브밴드 유사도를 계산하는 변형된 특징벡터 재결합 방법을 제안하였다. 또한, 적응형 잡음모형을 이용한 서브밴드 특징벡터 선택방법을 이용하여 변형된 특징벡터 재결합 기반의 화자검증 시스템을 잡음에 더욱 강인하게 하였다. 실험결과 프레임 SNR 기반의 신뢰성 높은 특징벡터 선택방법보다 적응형 잡음모형을 이용한 특징벡터 선택방법이 더 높은 성능을 보였다.

참 고 문 헌

- [1] D. Reynold, T. Quatieri, R. Dunn, "Speaker verification using adapted Gaussian mixture models", *Digital Signal Processing*, Vol. 10, pp. 19-41, 2000.
- [2] A. Drygajlo, M. El-Maliki, "Speaker verification in noisy environments with combined spectral subtraction and missing feature theory", *Proc. ICASSP*, Vol. 2, pp. 121-124, 1998.
- [3] K. Yiu, M. Mak, S. Kung, "Environment adaptation for robust speaker verification", *Proc. EUROSPEECH*, pp. 2973-2976, 2003.
- [4] C. Barras, J. Gauvain, "Feature and score normalization for speaker verification of cellular data", *Proc. ICASSP*, Vol. 2, pp. 49-52, 2003.
- [5] S. Okawa, E. Bocchieri, A. Potamianos, "Multi-band speech recognition in noise environment", *Proc. ICASSP*, pp. 641-644, 1998.
- [6] A. E. Rosenberg, J. DeLong, C. H. Lee, B. H. Juang, F. K. Soong, "The use of cohort normalized scores for speaker recognition", *Proc. ICSLP*, pp. 599-602, 1992.
- [7] R. Auchenthaler, J. S. Mason, "Score normalization in multi-band speaker verification", *Proc. RL42C*, pp. 102-106, 1998.
- [8] B. Mak, "A mathematical relationship between full-band and multiband mel-frequency cepstral coefficients", *IEEE Signal Processing Letters*, Vol. 9, No. 8, pp. 241-244, 2002.
- [9] D. Pearce, H. Hirsch, "The Aurora experimental framework for the performance evaluation of speech recognition systems under noise conditions", *Proc. ICSLP*, Vol. 4, pp. 29-32, 2000.
- [10] M. Berouti, R. Schwartz, J. Makhoul, "Enhancement of speech corrupted by acoustic noise", *Proc. ICASSP*, pp. 208-211, 1979.

▶ 김성탁(Sungtak Kim) : 교신저자

주소: 305-732 대전광역시 유성구 문지로 119번지
소속: 한국정보통신대학교 공학부 음성인식기술연구실
전화: 042) 866-6221
FAX: 042) 866-6245
E-mail: stkim@icu.ac.kr

▶ 지미경(Mikyong Ji)

주소: 305-732 대전광역시 유성구 문지로 119번지
소속: 한국정보통신대학교 공학부 음성인식기술연구실
전화: 042) 866-6221
FAX: 042) 866-6245
E-mail: lindaji@icu.ac.kr

▶ 김희린(Hoirin Kim)

주소: 305-732 대전광역시 유성구 문지로 119번지
소속: 한국정보통신대학교 공학부 음성인식기술연구실
전화: 042) 866-6139
FAX: 042) 866-6245
E-mail: hrkim@icu.ac.kr