

스코어 기반 관측신뢰도를 이용한 잡음환경하 화자식별

Speaker Identification Using Score-based Confidence in Noisy Environments

민 소 회* · 김 진 영** · 송 민 규* · 나 승 유** · 최 승 호***

So Hee Min · Jin Young Kim · Min Gyu Song · Seung You Na · Seung Ho Choi

ABSTRACT

The performance of speaker identification is severely degraded in noisy environments. Recently probability weighting method based on observation membership was proposed for overcoming the noise problem[1]. In the paper[1] the observation confidence was calculated from SNR with sigmoid function. However, estimating SNR needs additive calculation amount and estimated SNR is corrupted in dynamic noisy environments.

In this paper we propose estimation methods of the observation confidence based on score-based reliabilities (SBR) of entropy and dispersion measures. Generally SBRs are obtained from speaker models' probabilities. The proposed methods are evaluated with ETRI speaker recognition DB. We compared the performances of the proposed methods with those in [1][8]. The experimental results show that the proposed methods can be successfully applied for the case where SNR is not available.

Keywords: Speaker identification, GMM, Score-based reliability, Observation Confidence

1. 서 론

음성인식 및 화자인식과 같은 음성처리 분야에서 현재 활발하게 연구되고 있는 분야는 오인식 단어를 검출하는 기술과 잡음에 강인한 인식기술이다[1]-[7]. 특히 유비쿼터스 기술 개발이 급진전됨에 따라, 모바일 또는 인터넷 환경에서 자동화자인식에 대한 요구가 증대되고 있으나, 유비쿼터스 환경에서 음성은 쉽게 주변 잡음에 노출된다.

이러한 문제점을 극복하기 위한 방법은, 최근 발표된 논문들을 살펴보면, 크게 세 가지로 나뉜다. 첫째는 CMS(cepstrum mean subtraction)[2] 및 RASTA(relative spectra)[3]와 같은 잡음에 강인한 파라미터를 추출하는 방법이고, 둘째는 화자의 모델을 잡음에 맞도록 적응시키는 모델 적응방법이다[4][5]. 그리고 마지막 세 번째 방법은 관측된 신호의 부정확도를 측정하여 이를 확률계산시 가중값으로 사용하는 방법이다. 이 방법은 논문 [1][8]에서 가우시안 혼합모델(Gaussian mixture model,

* 전남대학교 대학원 전자공학과

** 전남대학교 공과대학 전자컴퓨터공학부

*** 동신대학교 공과대학 컴퓨터학과

GMM)에 적용하여 화자식별 영역에서 그 성능이 충분히 검증되었다.

그러나 관측신뢰도(Observation Confidence) 기반 확률 가중방법은 완전히 해결된 문제가 아니며, 개선의 여지가 많다고 할 것이다. 특히 논문 [1][8]에서 관측신뢰도 측정 함수인 멤버십 함수는 신호대잡음비(signal-to-noise ratio, SNR)로부터 결정되는데, SNR을 관측신뢰도의 독립변수로 사용하는 데에 몇 가지 어려운 점이 있다.

첫째, SNR은 정확한 측정이 쉽지 않을 뿐 아니라, SNR을 측정하기 위해서는 많은 부가적인 계산량이 필요하다[9][10].

둘째, 신호에 동적 부가잡음이 더해진 경우에는 SNR의 측정이 더욱 어렵고, 그 오차도 증가한다.

그러므로 SNR에 의존하지 않는 관측신뢰도를 개발하는 것은 매우 의미있는 연구가 될 것이다.

관측신뢰도의 개념이 활발하게 사용되는 분야는 멀티모달(multi-modal) 음성처리 분야이다. 즉 시청각 음성인식 또는 시청각 개인 식별 및 인증에서 계산된 시각정보와 음성정보를 통합(integration)할 때, 적절한 가중값을 구하기 위하여 관측신뢰도가 사용된다[11].

측정된 신호의 신뢰도는 두 가지 방법으로 계산 되는데, 하나는 개별 신호 분석방법으로 입력 음성신호의 SNR을 측정하여 SNR에 따른 관측신뢰도를 계산하는 신호기반 신뢰도이고, 다른 하나는 각 모델들의 확률값 분포로부터 확실성함수(reliability)를 결정하고 이 값들로부터 관측신뢰도를 구하는 스코어(score) 기반 신뢰도이다[11].

관측 신호의 부정확도는 부가잡음에 의해 발생하였으므로 신호로부터 직접 SNR을 구하여 사용하면 신뢰도 추정이 가장 정확할 것이다. 그러나 스코어 기반 신뢰도가 적절한 성능향상을 갖는다면, SNR을 측정하기 어려운 환경에서 유용하게 적용될 수 있다.

본 논문에서는 스코어 기반 확실성 함수로써 엔트로피 척도(entropy measure)와 이산 척도(dispersion measure)를 사용하였고 다양한 관측신뢰도 추정방법을 제시하고 화자식별성능을 검토하였다. 본 논문의 구성은 다음과 같다. 2장에서는 GMM과 관측신뢰도 기반의 화자식별 시스템에 대하여 다루고, 3장에서는 스코어기반 관측신뢰도의 추정방법에 대하여 다룬다. 4장에서는 제시한 방법에 대한 적합성을 화자식별 실험을 통해 검증한다.

2. GMM과 관측신뢰도 기반 화자식별 시스템

2.1 GMM 분류기와 관측신뢰도 함수

화자인식기에서 각 화자별 모델을 만들기 위해 GMM 분류기를 사용한다. GMM은 알려진 바와 같이 특징벡터 공간을 다수의 가우시안 분포의 가중합으로 표현하는 것으로 다음과 같은 식으로 표현된다.

$$f_X(x) = \sum_{i=1}^I w_i N(x; \mu_i, \Sigma_i) \quad (1)$$

여기서 w_i 는 가중값, $N(\cdot)$ 는 정규분포, μ_i 는 i 번째 정규분포의 평균 그리고 Σ_i 는 i 번째 정규분포의 공분산행렬이고 I 는 가우시안 믹스처 개수이다. 위 식 (1)의 GMM 모델 파라미터들은 EM 알고리즘에 의하여 학습되어 진다.

논문 [1][8]에서 제시된 관측신뢰는 가중치로써 확률계산에 반영하게 되는데 식 (2)와 같이 사용된다.

$$P_k(X) = \prod_{n=1}^N p_x^{\mu_n}(x_n|M_k) \quad (2)$$

위 식에서 $P_k(X)$ 는 주어진 관측 열에 대한 k 번째 화자의 확률, $p_x(x_n|M_k)$ 는 k 번째 화자 모델 M_k 에 대한 n 번째 특징벡터 x_n 의 관측확률이고, μ_n 는 n 번째 관측 파라미터의 관측신뢰도이며 N 은 전체 프레임수이다. 관측신뢰도란 주어진 관측파라미터의 신뢰도, 즉 얼마나 정확한 측정인가를 나타내는 값이다. 또 관측신뢰도는 SNR의 함수로 표현하였는데 시그모이드(sigmoid) 함수를 매핑(mapping) 함수로 이용하였다[1][8]. 즉, 관측신뢰도는 식 (3)과 같이 정의된다

$$\mu(t) = \frac{1}{1 + e^{a(SNR(t)-b)}} \quad (3)$$

위 식에서 a 는 스케일 파라미터이고, b 는 이동 파라미터이다. 식(3)은 최소값 0, 최대값 1로서 퍼지이론의 멤버십 함수로 해석될 수 있다.

$a = -0.51$ 이고 $b = 11.1$ 인 경우의 멤버십 함수를 <그림 1>에 나타내었다. 논문 [1]에서는 문맥독립 화자식별 실험에서 멤버십 함수를 사용한 경우 인식 성능이 개선됨을 확인하였다. 이 때 사용된 시그모이드 함수의 파라미터 a 와 b 는 경험적으로 결정된 값이었고 논문[8]에서는 Particle Swarm Optimization (PSO) 이론을 이용하여 파라미터 a 와 b 를 최적화하였다.

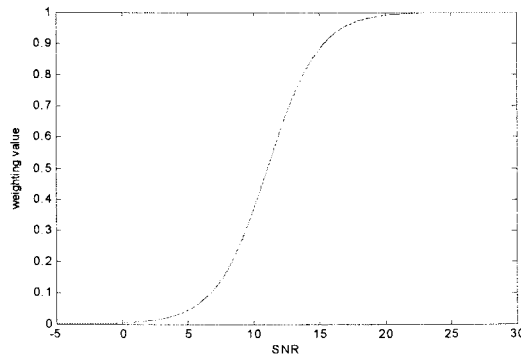


그림 1. 멤버십 함수의 예 ($a = -0.51$ $b = 11.1$)

2.2 관측신뢰도 기반 화자인식 구조

<그림 2>은 논문 [1][8]에서 제시한 관측신뢰도의 개념과 화자식별 과정을 보여주고 있다. 먼저 일반적인 화자인식 과정은 다음과 같다.

- 1) 입력음성의 특징파라미터를 추출한다. 본 논문에서는 지금까지 가장 우수한 성능을 보인다고 알려진 멜켑스트럼(Mel-Cepstrum)을 사용하였다.
- 2) 멜켑스트럼 파라미터에 대하여 CMS를 수행한다. CMS는 채널에 의해 발생하는 채널왜곡을 제거하고 잡음에 의한 파라미터의 오염을 일부 제거하는 성질을 가지고 있다.
- 3) 입력된 CMS 결과 파라미터를 이용하여 GMM 학습을 수행한다. 학습결과는 화자별로 저장한다.
- 4) 화자인식 시에는 입력음성의 열에 대하여 먼저 관측신뢰도를 계산하고 둘째, GMM 모델에 대한 발생확률을 계산하여 가장 높은 확률을 갖는 화자로 판단하게 된다.

<그림 2>에서 점선으로 표시된 부분은 관측신뢰도를 계산하여 가중치로써 확률계산에 반영하기 위한 전처리 과정이다.

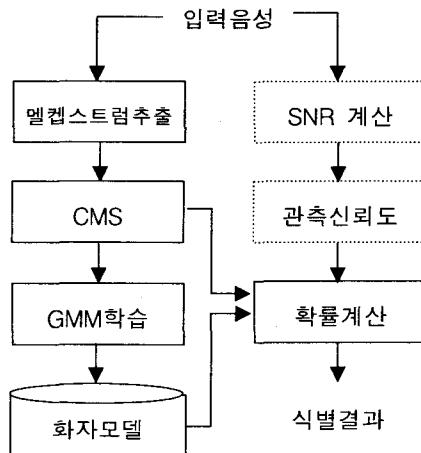


그림 2. 화자식별 알고리즘(논문[1])

3. 스코어 기반 관측신뢰도 추정 및 적용

멀티모달 개인인증에서 다양한 정보를 통합하기 위해, 관측된 각 정보의 신뢰도를 추정하여 적절한 가중값들을 계산한다. 측정된 정보의 신뢰도는 두 가지 접근방법에 의해 계산될 수 있는데, 첫째는 신호 기반 신뢰도이고, 둘째는 스코어 기반 신뢰도이다. 신호 기반 신뢰도는 관측된 신호로부터 적절한 계산을 통해 얻어지는 신뢰도로서, SNR 또는 조화지수 (harmonic index)가 널리 사용된다. 스코어 기반은 인식 모델로부터 계산된 확률분포를 사용하여 신뢰도를 추정하는 방법으로서, 엔

트로피(entropy) 척도, 이산(dispersion) 척도 그리고 차이(difference) 척도 등이 있다[11]. 본 장에서는 스코어 기반 확실성 함수에 대하여 소개하고, 이를 관측신뢰도 기반 화자식별 시스템에 적용하기 위한 방법론에 대하여 설명한다.

3.1 스코어 기반 확실성 함수

본 절에서는 위에서 언급한 스코어 기반 확실성 함수중 본 논문에 적용하고자 하는 엔트로피 척도와 이산 척도에 대하여 설명한다. 왜냐하면, 두 척도가 시청각 개인인증 실험에서 성공적으로 사용되었기 때문이다[11].

• 엔트로피 척도

엔트로피 척도는 주어진 화자 모델로부터 계산된 확률값들의 엔트로피를 구하는 방법으로서 식 (4)과 같이 정의된다.

$$H(t) = -\frac{1}{K} \sum_{k=1}^K P(O_t|\lambda_k) \log P(O_t|\lambda_k) \quad (4)$$

여기서, K 은 전체 화자의 수, O_t 는 시간 t 에서 관측 특징 파라미터, λ_k 는 k 번째 화자의 GMM 모델이다. 신호에 잡음이 추가되면 계산된 특징 파라미터 O_t 를 교란시켜서 $P(O_t|\lambda_k)$ 의 값들이 유사한 값들을 갖게 된다. 결과적으로 엔트로피는 잡음량이 커질수록 증가하게 되므로 엔트로피 $H(t)$ 는 신호의 오염정도를 파악하는 간접적인 파라미터로 사용될 수 있다.

• 이산 척도

로그-유사도(log-likelihood)값들의 이산 척도는 다음 식 (5)와 같이 정의된다.

$$D_1(t) = -\frac{2}{K(K-1)} \sum_{k=1}^K \sum_{l=k+1}^K \log \left(\frac{P(O_t|\lambda_k)}{P(O_t|\lambda_l)} \right) \quad (5)$$

높은 이산 값은 높은 신뢰도를 나타내고, 낮은 이산 값은 낮은 신뢰도를 의미한다. 한편, 시험 신호에 대해 계산되는 최고의 확률값이 나머지 $K-1$ 개의 모델들의 확률값에 대하여 상대적으로 클 경우 신뢰도가 크다고 판단 할 수 있는데, 이러한 의미에서 새로운 이산 척도를 정의할 수 있다. 일단, 로그 확률 중 가장 큰 값과, 두 번째 큰 값을 다음과 같이 정의하자.

$$\begin{aligned} S_{best}(t) &= \max_{1 \leq k \leq K} [\log P(O_t|\lambda_k)] \\ k_{best}(t) &= \arg \max_{1 \leq k \leq K} [\log P(O_t|\lambda_k)] \\ S_{second-best}(t) &= \max_{1 \leq k \leq K, k \neq k_{best}(t)} [\log P(O_t|\lambda_k)] \end{aligned} \quad (6)$$

그러면 이산 척도는 다음과 같이 정의된다.

$$D_2(t) = \alpha(t) |S_{best}(t) - S_{second-best}(t)| \quad (7)$$

위식에서 $\alpha(t)$ 값은 $D_2(t)$ 를 제어하기 위한 파라미터로서, 일반적으로 각 모델들의 로그확률들의 평균값을 사용한다. 또는 $S_{second-best}(t)$ 대신 $S_{worst}(t)$ 를 사용할 때 $D_2(t)$ 가 1이 되도록 다음 식과 같이 정의한다.

$$\alpha(t) = \frac{1}{S_{best}(t) - S_{worst}(t)} \quad (8)$$

본 논문에서는 간단한 계산을 통해 계산되는 식 (6) 및 (7)의 이산척도를 사용하였다.

지금까지 멀티모달 음성처리에서 여러 모달리티의 통합을 위해 사용되는 스코어 기반 확실성 함수를 소개하였다. 다음 절에서는 위의 확실성 함수를 사용하여 시간 t 의 관측 특징 벡터에 대한 관측신뢰도를 어떻게 추정할 것인지에 대하여 설명하고자 한다.

3.2 관측신뢰도 멤버십 함수 추정 방법

스코어 기반 확실성 함수로부터 논문 [8]에서 정의된 관측신뢰도를 추정하기 위한 방안으로서 다음과 같은 세 가지 방법들을 고려할 수 있다.

- 방법 1: (직접적용) 앞 절에서 정의된 확실성 함수들을 곧바로 관측신뢰도로 사용하는 방법이다. 단, 퍼지이론에서 멤버십 함수는 최대값 1, 최소값 0을 갖는다는 조건을 가지고 있기 때문에, 식 (6)~(8)을 사용한 이산 척도의 경우, 관측신뢰도 추정값으로 바로 사용될 수 있다. 즉,

$$\hat{\mu}(t) = D_2(t) \quad (9)$$

로 정의된다. 또한 엔트로피를 사용하는 경우도 같은 방식으로 정규화하여 사용할 수 있다.

- 방법 2: (SNR 추정) 두 번째 방법은 엔트로피 척도 및 이산 척도로부터 SNR을 추정하여 기존의 SNR 기반 멤버십 함수를 사용하는 방법이다. 즉

$$\widehat{SNR}(t) = f(H(t), D_2(t)) \quad (10)$$

와 같이 스코어 기반 신뢰도들로부터 SNR을 간접적으로 추정하여 식 (11)과 같이 멤버십 함수를 계산한다.

$$\hat{\mu}(t) = \frac{1}{1 + e^{a(\widehat{SNR}(t) - b)}} \quad (11)$$

- 방법 3: (멤버십 함수 추정) 세 번째 방법은 엔트로피 및 이산 척도로부터 관측멤버십 함수를 직접 추정하는 것으로 다음과 같이 정의된다.

$$\hat{\mu}(t) = g(H(t), D_2(t)) \quad (12)$$

단, 함수 $g(\)$ 는 $E[(\mu(SNR(t)) - \hat{\mu}(t))^2]$ 을 최소화 하도록 결정된다.

본 논문에서는 방법 2와 방법 3의 추정 함수로서 다항식을 사용하였는데 예를 들어 함수 g 는 다음과 같이 2차 다항식이다.

$$g(H(t), D_2(t)) = c + a_1 H(t) + a_2 H^2(t) + b_1 D_2(t) + b_2 D_2^2(t) \quad (13)$$

4. 시뮬레이션 및 결과 고찰

4.1 화자인식 DB 및 문맥중속 화자식별 실험개요

본 논문에서는 제안된 방법의 성능을 확인하기 위하여 ETRI에서 만든 한국어 화자인식용 휴대 폰 음성 DB를 사용하여 문맥중속 화자식별 실험을 하였다. 음성데이터의 샘플링 주파수는 8 kHz이며, 8 비트 μ -law PCM 방식으로 코딩되어 제공되었다. 그리고 DB의 전체 화자의 수는 남녀 모두 49 명이고, 화자 당 음성파일은 모두 20 개로 이중 10 개씩을 학습용과 실험용으로 나누어 사용하였다. 문맥중속 인식실험에 사용한 음성데이터는 발성 시간이 약 3 초 정도로 화자모델 학습에 사용된 음성데이터는 파일 10 개를 합친 평균 약 30 초 정도의 분량이다.

실험에서 입력 음성데이터의 한 프레임은 40 ms로 하였고, 20 ms씩 중첩되어 처리되도록 하고, 음성의 특징벡터는 12 차 멜켵스트럼 계수와 로그 에너지를 포함하였으며, 채널 왜곡을 보상하기 위해 CMS 방법을 적용하였다. GMM 화자모델에 포함된 Gaussian 개수는 10 개이고, EM 알고리즘에 의해 GMM 모델 λ_i 의 파라미터를 반복적으로 훈련하여 구하였다. 이 과정에서 공분산 값은 full covariance를 사용하였고, 알고리즘의 초기과정에서는 fuzzy C-means clustering 방법을 사용하였다. 이에 대한 내용은 다음 <표 1>과 같다.

표 1. 문장중속 화자식별 실험의 개요

항 목	규 격
음성 DB	ETRI 휴대폰 화자인식용 음성 DB
화샘플링/음성코딩	8kHz/8 bits μ -law PCM
화자 수	49
화자당 학습 음성파일의 개수	10
화자당 테스트 음성파일의 개수	10
프레임길이/중첩	40 ms/20 ms
음성특징벡터	12차 멜 캡스트럼과 로그에너지
채널 보상	Cepstral Mean Subtraction
Gaussian Mixture 개수	10
GMM 모델	EM 알고리즘, full covariance

4.2 실험결과 및 고찰

위 3.2절에서 스코어 기반 확실성 함수를 이용하여 관측신뢰도를 추정하기 위한 세 가지 방법을 설명한 바 있다. 본 절에서는 각 방법을 실험을 통해 비교하고 검토하고자 한다.

ETRI 화자인식용 휴대폰 음성 DB는 잡음이 섞이지 않은 깨끗한 발화들로 구성되어 있다. 따라서 본 논문에서는 잡음음성을 발생시키기 위하여 원 음성에 가우시안 백색 잡음을 더하여 얻었다. 식 (14)은 깨끗한 음성으로부터 잡음 음성을 얻는 과정을 수식으로 표현한 것이다.

$$S_{\eta}(n) = s(n) + \beta\eta(n) \quad (14)$$

위 식에서 $s(n)$ 은 잡음이 섞이지 않은 깨끗한 음성이며, $\eta(n)$ 은 평균전력이 1 인 가우시안 불규칙 잡음이다. 또한 β 는 가산되는 잡음의 양을 결정하는 파라미터이며, $S_{\eta}(n)$ 는 잡음에 오염된 신호를 의미한다.

<표 2>는 잡음량에 따른 화자식별의 성능을 보여주고 있다. 표에 보인 바와 같이 잡음이 없는 경우($\beta = 0$) 최고 식별률 94.9%를 보이고 있다. 그러나 잡음이 증가함에 따라서 화자식별 성능이 크게 저하되고 있다. SNR 기반 관측신뢰도를 사용하는 경우 이미 논문 [8]에서 검증되었듯이 높은 성능향상을 관찰 할 수 있다. 한편 스코어 기반 관측신뢰도를 사용하는 경우를 살펴보면, 방법 1의 경우 성능이 거의 개선되지 않아 멀티모달 인증에 사용되는 확실성 함수를 곧바로 신뢰도값으로 적용하는 것은 적절하지 않음이 판명되었다.

한편 방법 2와 방법 3은 비슷한 성능 개선을 보이지만, 대체적으로 방법 3이 방법 2보다는 좀 더 나은 성능을 보이고 있다. 따라서 스코어 기반 확실성 함수로부터 관측신뢰도의 척도인 멤버십 함수를 중간 매개변수 없이 바로 추정하는 방법이 가장 우수한 성능을 보인다고 결론지을 수 있다.

표 2. 제안된 방법의 비교 (3차 다항식 이용, 식별률 단위: %)

β	평균 SNR (dB)	가중처리 없음 $\mu(t) = 1$	확률가중			
			SNR 기반 [8]	방법1	방법2	방법3
0		94.90	-	-	-	-
0.00625	20.8	85.30	91.83	85.10	86.53	87.34
0.0125	16.3	74.00	84.90	72.04	77.34	79.18
0.025	11.8	52.65	65.70	54.28	62.61	61.22
0.05	7.9	25.71	36.12	25.71	33.06	35.51

그러나 방법 3을 사용하더라도 SNR 기반의 관측신뢰도를 이용하는 경우에 비하여 성능 개선이 우수하지 못하다. <그림 3>은 SNR 기반 방식의 성능 개선량을 100%라고 할 때, 방법 3의 성능 개선정도를 β 값에 따라 보여주고 있다. 그림에서 볼 수 있듯이, SNR 이 20.8dB 일 때 약 31%의 성능을 보이며, 7.9dB에서는 94%의 성능을 보이고 있다. 따라서 제안한 스코어 기반 관측신뢰도 방법은 SNR의 크기가 낮아질수록, 논문 [8]의 성능에 근접해 감을 알 수 있다.

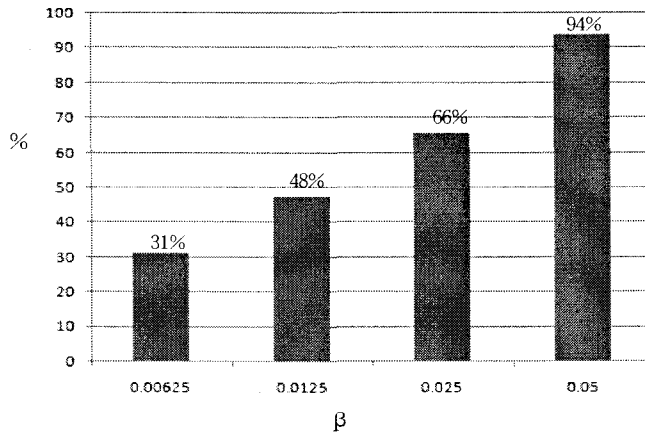


그림 3. SNR 방식 대비 성능 개선 정도

다음 <표 3>은 스코어기반 확실성함수를 이용하여 관측신뢰도를 추정하는 방법 3에 대하여 다항식의 차수가 3 차, 2 차 그리고 1차였을 때의 성능을 보여주고 있다. <표 3>에 의하면, 다항식의 차수가 1 차와 2 차일 때에는 비슷한 성능을 보이나 3 차일 때는 상대적으로 약간 성능이 향상됨을 알 수 있다.

표 3. SNR 추정 다항식의 차수에 따른 성능 비교

β	방법3		
	3차 다항식	2차 다항식	1차 다항식
0.00625	87.34	87.14	87.14
0.0125	79.18	77.96	77.95
0.025	61.22	60.20	60.20
0.05	35.51	32.04	32.04

지금까지 본 논문에서 제안한 방법들의 성능을 검증하였다. 성능검증 결과, SNR기반의 관측신뢰도를 이용하는 경우[8]에 비해 우수하지 못함을 보았다. 그럼에도 불구하고 본 논문에서 제안하는 방식은 SNR의 계산 없이 사용할 수 있다는 장점을 가지고 있기 때문에, SNR을 계산하기 위해 부가적 계산량이 많은 시스템, 또는 잡음의 특성이 매우 동적이어서 SNR 추정이 매우 어려운 경우 충분히 활용이 가능하다고 사료된다.

5. 결 론

화자인식에서 잡음문제를 해결하기 위해 제시된 멤버십 함수 기반 확률가중 방법이 최근에 제시된 바 있다[1][8]. 관측신뢰도 측정 함수인 멤버십 함수는 SNR로부터 시그모이드 함수 꼴로 결정되었는데, 실제 SNR을 정확하게 측정하는 것은 쉽지 않은 문제일 뿐 아니라, 부가적인 계산량을 요구한다.

본 논문에서는 신호로부터 직접 구해지는 SNR 대신, 화자 모델들의 계산된 확률분포로부터 결정되는 스코어 기반 관측신뢰도를 이용하는 방법을 제안하고 실험을 통하여 검증하였다. 실험 결과, SNR기반의 관측신뢰도를 이용하는 경우[8]에 비해 성능 저하가 있었지만, 여전히 잡음환경 아래에서 식별 성능을 향상시킬 수 있음을 보였다. 따라서 SNR을 추정하기 어려운 환경에서, 제안한 방법이 유용하게 사용되리라고 판단된다.

향후, 좀 더 다양한 스코어 기반 신뢰도를 대상으로 실험을 확장하여 검증할 예정이며, 관측신뢰도 측정함수로서, 다항식 추정기외에 신경회로망과 같은 학습기도 적용하여 성능을 검증하고자 한다.

참 고 문 헌

- [1] Kim, Jinyoung et. al. 2007. "Modified GMM training for inexact observation and its application to speaker identification." *Speech Science* 14(1), 163-175.
- [2] Rosenberg, A. et. al. 1994. "Cepstral channel normalization techniques for HMM-based speaker verification." *Proc. ICSLP-94*, 1835-1838.

- [3] Zhen, Bin, Wu, Xihong, Liu, Zhimin & Chi, Huisheng. 2000. "An enhanced RASTA processing for speaker identification." *Proc of 2000 ICSLP*, 251-254.
- [4] Mengusoglu, E. 2003. "Confidence measure based model adaptation for speaker verification." *Proc. of the 2nd IASTED International Conference on Communications, Internet and Information Technology*.
- [5] Sit, Chin-Hung, Mak, Man-Wai & Kung, Sun-Yuan. 2004. "Maximum likelihood and maximum a posteriori adaptation for distributed speaker recognition systems." *Proc of 1st Int. Conf. on biometric authentication*.
- [6] Mammone, R. J., Zhang, X. & Ramachandran, R. P. 1996. "Robust speaker recognition: A feature-based approach." *IEEE Signal Processing Magazine* 13(5), 58-71.
- [7] Eberhart, R. & Kennedy, J. 1995. "A new optimizer using particle swarm theory." *Proc. of Sixth International Symposium on Micro Machine and Human Science*, 39-43.
- [8] 민소희, 김진영, 송민규, 나승유. 2007. "Particle Swarm 기반 최적화 멤버쉽 함수에 의한 잡음 환경에서의 화자인식 성능향상." *음성과학* 14(2), 105-114.
- [9] Vondrasek, M. & Pollak, P. 2005. "Methods for speech SNR estimation: Evaluation tool and analysis of VAD dependency." *Radioengineering* 14(1), 6-11.
- [10] Nember, E., Goubraïn, R. & Mahmoud, S. 1999. "SNR estimation of speech signals using subband and fourth-order statistics." *IEEE Signal Processing Letters*. 7(7), 171-174.
- [11] Fox, Nial Andrew. 2005. Audio and video based person identification, *Ph.D dissertation. University College Dublin*.

접수일자: 2007. 10. 8

게재결정: 2007. 11. 26

▲ 민소희

광주광역시 북구 용봉동 300번지 전남대학교 (우: 500-757)
 전남대학교 대학원 전자공학과
 Tel: +82-62-530-0370
 E-mail: minsh@chonnam.ac.kr

▲ 김진영(교신저자)

광주광역시 북구 용봉동 300번지 전남대학교 (우: 500-757)
 전남대학교 공과대학 전자컴퓨터공학부
 Tel: +82-62-530-1757
 E-mail: beyondi@chonnam.ac.kr

▲ 송민규

광주광역시 북구 용봉동 300번지 전남대학교 (우: 500-757)
 전남대학교 대학원 전자공학과
 Tel: +82-62-530-0472
 E-mail: smg686@lycos.co.kr

▲ 나승유

광주광역시 북구 용봉동 300번지 전남대학교 (우: 500-757)

전남대학교 공과대학 전자컴퓨터공학부

Tel: +82-62-530-1753

E-mail: syna@chonnam.ac.kr

▲ 최승호

전남 나주시 건재로 253번지 (우: 520-714)

동신대학교 공과대학 컴퓨터학과

Tel: +82-62-330-3194

E-mail: shchoi@dongshinu.ac.kr