음질향상 기법과 모델보상 방식을 결합한 강인한 음성인식 방식

A Robust Speech Recognition Method Combining the Model Compensation Method with the Speech Enhancement Algorithm

김 희 근* · 정 용 주* · 배 건 성**
Hee-Keun Kim · Yong-Joo Chung · Keun-Seung Bae

ABSTRACT

There have been many research efforts to improve the performance of the speech recognizer in noisy conditions. Among them, the model compensation method and the speech enhancement approach have been used widely. In this paper, we propose to combine the two different approaches to further enhance the recognition rates in the noisy speech recognition. For the speech enhancement, the minimum mean square error-short time spectral amplitude (MMSE-STSA) has been adopted and the parallel model combination (PMC) and Jacobian adaptation (JA) have been used as the model compensation approaches. From the experimental results, we could find that the hybrid approach that applies the model compensation methods to the enhanced speech produce better results than just using only one of the two approaches.

Keywords: Noisy speech recognition, Hidden Markov model, model compensation, speech enhancement

1. 서 론

HMM 기반의 음성인식에서 모델 보상 기법들은 잡음음성인식을 위해서 매우 효과적임이 알려져 있다[1-2]. 그 중에서도 PMC(parallel model combination) 방식은 다른 적응 방식에 비해서도 성능이 뛰어난데, 특히 계산량이 다른 방식에 비해서 비교적 작으며 인식 중에도 잡음신호만 있으면 잡음음성에 대한 적응을 할 수 있다는 장점이 있다. 하지만, PMC 방식에서는 해석의 간편함을 위해서 약간의 통계적 가정을 하는데, 이러한 가정의 오차로 인하여 실제적으로 인식율의 저하를 초래할 수 있음이 알려져 있다[3]. JA 방식은 인식환경과 유사한 환경에서 HMM의 모델 훈련이 가능하다는 가정하에서 우수한 성능을 보이는 것으로 알려져 있다[4]. 위와 같은 잡음음성인식을 위한 모델보상 방식 외에도 잡음음성의 음질개선을 통한 인식성능 향상을 위한 연구들이 많이 이루어지고 있는데, 그 중

* 계명대학교 전자공학과
** 경북대학교 전기전자공학부

에서도 대표적인 방식이 스펙트럼 차감법(spectral subtraction), Wiener 필터링 그리고 MMSE-STSA(Minimum Mean Square Error -Short Time Spectral Amplitude) 기반의 음질개선 기법 등이 다[5-6]. 본 연구에서는 특히 음질개선 방식 중에서 다른 방법들에 비해서 비교적 우수한 성능을 나타내는 MMSE-STSA 기반의 음성개선기법을 사용하였다. 지금까지의 기존의 연구결과들은 주로 모델 보상방식과 잡음음성의 음질개선 방식으로 나뉘어져 독립적으로 많은 연구가 수행되어 왔으나, 이 서로 다른 접근방식을 함께 사용한 연구결과는 그리 많지 않았다. 특히, 잡음음성 음질개선 방식 중에서 가장 우수한 성능을 나타내는 것으로 알려진 MMSE-STSA 방식과 모델보상 방식의 결합은 보다 우수한 음성인식 성능을 보이기 위한 매우 효과적인 방식이라 생각되어 진다. 특히, 본 연구에서는 자동차 환경에서의 잡음음성에 대한 연구를 위해서 자동차 환경에서의 반향과 음악소리 그리고 순수한 자동차 소음 등을 포함한 잡음음성신호를 생성하여 인식실험을 수행함으로써 보다 현실적인 연구결과를 얻고자 노력하였다.

일반적으로 자동차 잡음 환경에서 음성입력을 위해서는 <그림 1(a)>에 나타난 바와 같이 마이크로폰 어레이를 이용하여 beamforming을 통해서 음성을 취득한 후 MMSE 기반의 음성개선방식과 반향제거기를 통해서 음질의 향상을 꾀하게 된다. 또한 <그림 1(b)>에서처럼 인식모델 적응 방식을 통해서도 우수한 성능을 얻을 수 있다. 하지만 음성개선방안과 인식모델보상방식을 동시에 적용한다면 보다 향상된 인식결과를 얻을 수 있으리라 생각된다. 이를 위해서 <그림 1(c)> 와 같은 구조의 인식시스템이 고려될 수 있다. 여기서는 먼저, 마이크로폰 어레이를 통하여 입력된 잡음음성신호에 대해서 기 개발된 MMSE-STSA 기반의 음질개선 알고리듬을 적용한다[7]. 보다 향상된 인식성능을 위해서는 <그림 1(a)>에서처럼 개선된 음성을 인식엔진의 입력으로 바로 사용하는 대신에 <그림 1(c)>에서는 개선된 음성으로부터 잔류(residual)잡음신호를 획득한 후 이를 이용하여 인식모델의 개선을 유도할 수 있으리라 생각된다. 인식모델의 개선을 위해서는 앞에서 언급된 PMC 나 JA방식이 사용된다. 본 연구에서는 인식실험의 편리성을 위해서 실제 환경의 마이크로폰 어레이를 사용하는 대신에 단일 마이크로폰을 이용하여 수집된 음성을 사용하였다.

다음 장에서는 MMSE-STSA기반의 음질개선 방식과 PMC/JA기반의 모델보상방식에 대해서 간단히 소개하고 3 장에서는 음질개선방식과 모델보상방식을 사용한 경우에 잡음환경에서의 인식성능을 분석하고 4 장에서 결론을 맺는다.

2. 음질개선과 인식모델 보상 방식의 개요

2.1 MMSE-STSA 기반의 음질개선

$x(t)$와 $d(t)$를 각각 깨끗한 음성신호와 부가 잡음신호라고 하면 부가적으로 오염된 잡음음성신호 $y(t)$는 일반적으로 다음과 같이 표현된다.

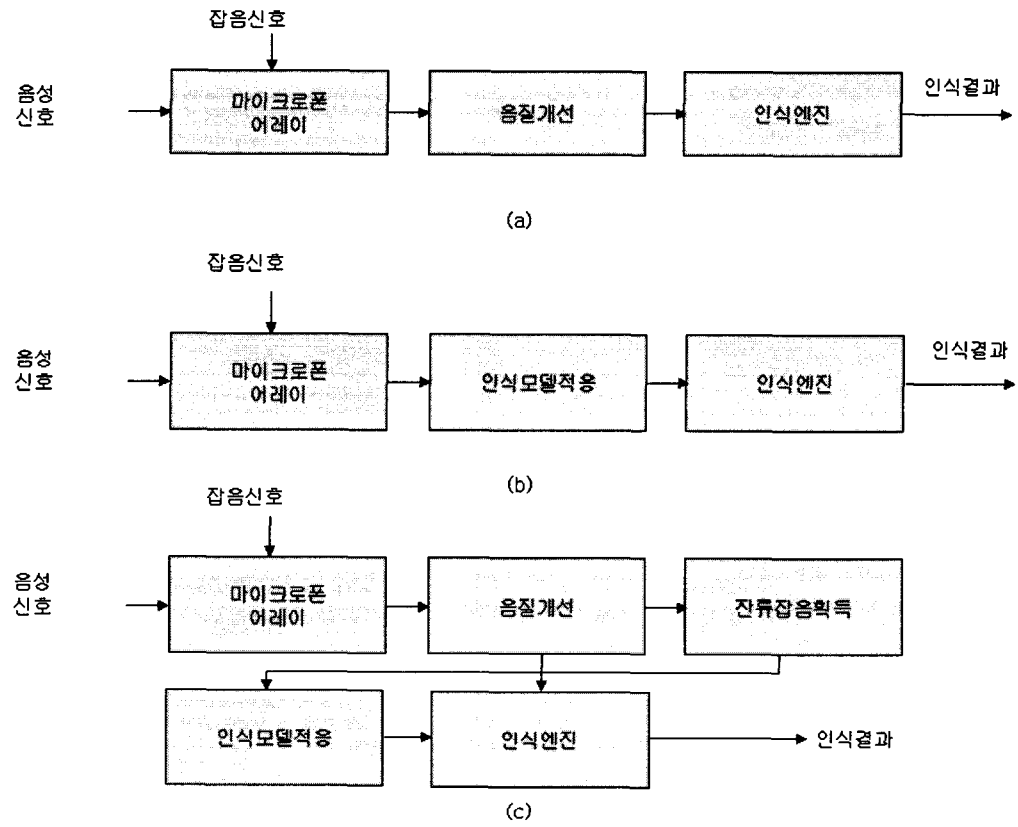


그림 1. 음질개선방식과 인식모델보상 방식의 결합에 의한 향상된 음성인식시스템 구조

$$y(t) = x(t) + n(t) \tag{1}$$

그리고 $x(t)$, $n(t)$, $y(t)$에 대한 k번째 스펙트럼 성분을 각각 $X_k = A_k \exp(j\alpha_k)$, D_k, $Y_k = R_k \exp(j\upsilon_k)$ 등으로 나타낼 수 있다. 잡음음성의 음질을 개선하는 알고리듬인MMSE-STSA에서는 관찰 가능한 $y(t)$ 또는 Y_k 로부터 원래의 음성신호 $x(t)$의 스펙트럼 크기 값 (spectral amplitude) A_k를 MMSE 기준에 의해서 구하게 된다[6].

MMSE-STSA 에 의해서 추정된 음성신호 $x(t)$의 k번째 스펙트럼 크기 값 $\widehat{A_k}$는 다음과 같다.

$$\begin{aligned} \widehat{A_k} &= E\{A_k \mid Y_k\} \\ &= \frac{\Lambda(Y_k, q_k)}{1 + \Lambda(Y_k, q_k)} E(A_k \mid Y_k, H_k^1) \end{aligned} \tag{2}$$

여기서 $\Lambda(Y_k, q_k)$ 는 다음과 같이 두 개의 우도(likelihood) 값들의 비로 정의된다.

$$\Lambda(Y_k,q_k)=\mu_k\frac{p(Y_k|H_k^1)}{P(Y_k|H_k^0)} \tag{3}$$

위식에서 $\mu_k=\frac{1-q_k}{q_k}$이며 q_k는 k번째 스펙트럼 성분에서 음성신호가 부재(absence)할 확률을 나타낸다. H_k^0와 H_k^1는 각각 k번째 스펙트럼 성분에서 신호가 부재할 가설(hypothesis)과 존재(presence)할 가설을 의미한다.

또한, 식(3)의 $\Lambda(Y_k,q_k)$은 스펙트럼 성분에 대한 Gaussian 모델링을 이용하면 다음과 같이 추정된다(Ephraim 등(1984)).

$$\Lambda(Y_k,q_k)=\frac{1-q_k}{q_k}\frac{\exp(v_k)}{1+\xi_k} \tag{4}$$

$$\text{단, } v_k=\frac{\xi_k}{1+\xi_k}\gamma_k\ ,\ \xi_k=\frac{E\{A_k^2|H_k^1\}}{E(|D_k|^2)},\quad \gamma_k=\frac{R_k^2}{E(|D_k|)^2}\text{이다.} \tag{5}$$

식(2)에서 $E(A_k|Y_k,H_k^1)$ 는 k번째 스펙트럼 성분에서 신호가 존재할 경우의 스펙트럼 크기에 대한 MMSE 값이 되며, 아래식과 같이 추정된다.

$$E(A_k|Y_k,H_1^k)=G_{MMSE}(v_k,\gamma_k)R_k \tag{6}$$

$$G_{MMSE}(\xi_k,\gamma_k)=\Gamma(1.5)\frac{\sqrt{v_k}}{\gamma_k}exp(\frac{-v_k}{2})[(1+v_k)I_0(\frac{v_k}{2})+v_kI_1(\frac{v_k}{2})]R_k \tag{7}$$

여기서$\Gamma(\bullet)$는 gamma 함수를 나타내며 I_0,I_1은 각각 1차와 2차의 Bessel 함수를 나타낸다.

위의 식(5)에서 정의된 ξ_k 와 γ_k 는 각각 사전(prior) 및 사후(posterior) 신호대잡음비로 정의된다. 시간 t에서의 추정된 사전 SNR 값 $\hat{\xi}_{k,t}$을 얻기 위해서는 decision-directed 방식을 사용하며, 아래의 식과 같이 추정된다.

$$\hat{\xi}_{k,t}=\alpha\frac{\widehat{A_k^2}(t-1)}{E\{D_{k,\,t-1}^2\}}+(1-\alpha)P[\gamma_k(t)-1] \tag{8}$$

여기서, operator $P[\bullet]$ 는 아래식과 같이 정의된다.

$$P[x]=\begin{cases}x & \text{if } x\geq 0\\ 0 & otherwise\end{cases} \tag{9}$$

한편, 식(8)에서 요구되는 잡음전력 스펙트럼 $E\{D_{k,t}^2\}$ 값은 VAD(Voice activity detection)을 이용하여 정해진 잡음 구간에 대해서만 재 추정이 이루어지며, 그 구체적 식은 아래식과 같다.

$$E\{D_{k,t}^2\} = E\{D_{k,t-1}^2\} + \beta(|R_k(t)|^2 - E\{D_{k,t-1}^2\}) \tag{10}$$

본 연구에서는 MMSE-STSA 기반의 음성개선을 위해서 사전음성부재 확률 q_k=0.2 로 하고, α는 0.9 그리고 β는 0.4를 사용하였다(박철호 등, 2006).

2.2 모델보상방식 (PMC/JA)

모델보상 방식에서 가장 대표적인 PMC에서는 켑스트럼 영역의 HMM 파라미터들을 선형주파수 영역으로 변환하는 과정이 먼저 이루어진다. 잡음이 섞인 음성 HMM모델을 만들기 위해서 원래의 깨끗한 음성 HMM 파라미터 값과 잡음 신호에 대한 HMM 파라미터 값을 선형주파수 영역에서 서로 결합하여 준다.

이에 대한 자세한 과정은 다음과 같이 요약된다.

1) 로그스펙트럼 영역의 HMM 파라미터 값을 구하기 위해서 역 DCT(discrete cosine transformation) 변환을 취한다(Gales 등, 1993).

$$\mu^{l} = C^{-1}\mu^{c} \quad , \quad \Sigma^{l} = C^{-1}\Sigma^{c}(C^{-1})^{T} \tag{11}$$

여기서 μ^{l} 과 Σ^{l}는 로그 스펙트럼 영역에서의 평균벡터와 공분산 행렬이며 μ^{c} 와 Σ^{c} 는 켑스트럼 영역에서의 값이다.

2) 로그스펙트럼 영역의 HMM 파라미터 값을 선형영역으로 변환한다.

$$\mu_i = \exp(\mu_i^{l} + \frac{\Sigma_{ii}^{l}}{2}), \quad \Sigma_{ij} = \mu_i \mu_j [\exp(\Sigma_{ij}^{l}) - 1] \tag{12}$$

여기서 μ_i 와 Σ_{ij} 는 선형영역의 파라미터 값인 평균벡터 μ 와 공분산행렬 Σ 의 구성 원소이다.

3) 1), 2) 과정에서 얻어진 깨끗한 음성과 잡음신호의 HMM 파라미터 값을 선형영역에서 결합한다.

$$\hat{\mu} = \mu + \bar{\mu} \quad , \quad \hat{\Sigma} = \Sigma + \bar{\Sigma} \tag{13}$$

여기서 $\hat{\mu}$ 와 $\hat{\Sigma}$ 는 선형영역에서의 잡음음성의 평균벡터와 공분산 행렬이 되고, $\bar{\mu}$ 와 $\bar{\Sigma}$ 는 잡음신호에 관한 것이다.

4) 결합된 HMM 파라미터 값에 대해서 다음과 같이 로그변환과 DCT 변환을 취함으로서 최종적으로 켑스트럼 영역에서의 잡음음성의 평균벡터 $\hat{\mu}^c$ 와 공분산행렬 $\hat{\Sigma}^c$ 이 얻어진다.

$$\hat{\mu}_i^l = \log(\hat{\mu}_i) - \frac{1}{2}\log(\frac{\hat{\Sigma}_{ii}}{\hat{\mu}_i^2}+1), \quad \hat{\Sigma}_{ij}^l = \log(\frac{\hat{\Sigma}_{ij}}{\hat{\mu}_i \hat{\mu}_j}+1) \tag{14}$$

$$\hat{\mu}^c = C\ \hat{\mu}^l, \quad \hat{\Sigma}^c = C\ \hat{\Sigma}^l\ C^T \tag{15}$$

PMC 방식과 더불어 널리 이용되는 모델보상방식으로는 JA(Jacobian adaptation) 방식이 있는데, 구체적인 과정은 아래와 같다.

1) 켑스트럼(cepstrum) 영역에서 잡음신호 n 에 의해서 원래의 음성신호 x는 다음과 같이 변환된다고 가정된다.

$$y = C\,[\log\{\exp(C^{-1}x)+\exp(C^{-1}n)\}] \tag{16}$$

여기서 y 는 잡음음성신호이며 C 는 DCT 변환을 나타낸다.

2) 위의 식(16)으로부터 잡음신호 n에 대한 잡음음성신호 y의 변화율을 나타내는 Jacobian 행렬은 다음과 같이 얻어진다(Sagayama 등, 1997).

$$\frac{\partial y}{\partial n} = CR_y C^{-1} \tag{17}$$

여기서 R_y 는 대각행렬이며 k번째 대각원소 $R_{y,k}$는 다음과 같다.

$$R_{y,k} = \frac{(\exp(C^{-1}\mu_n))_k}{(\exp(C^{-1}\mu_x))_k + (\exp(C^{-1}\mu_n))_k} \tag{18}$$

여기서 μ_n 은 잡음신호의 평균벡터이고 μ_x 는 연속밀도 HMM의 각 혼합성분별 평균벡터를 나타낸다. 위의 식 (17) 와 (18) 을 이용하면 주어진 잡음신호에 대해서 기준HMM의 각 혼합성분별로의 Jacobian 행렬을 구할 수 있게 된다.

3) 위와 같이 얻어진 Jacobian 행렬을 이용하여 다음식과 같이 잡음신호가 n 에서 $\tilde{n}$ 으로 변할 경우의 잡음음성신호 y의 변이를 나타낼 수 있다.

$$\tilde{y} = y + \frac{\partial y}{\partial n}(n - \tilde{n}) \tag{19}$$

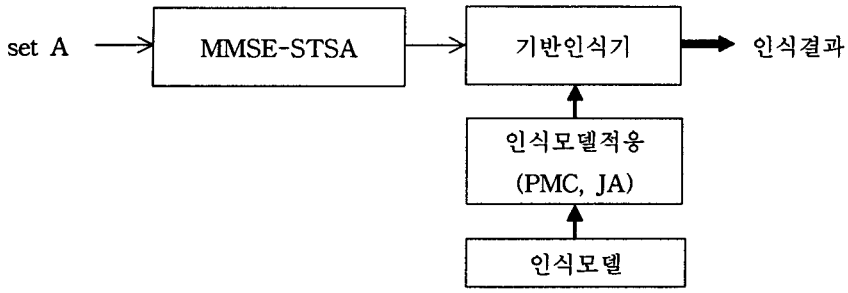
4) 잡음음성신호에 대한 HMM의 각 상태별 혼합성분의 평균값들을 추정하기 위해서는 위식의 양변에 평균자를 취한다.

$$E\{\tilde{y}\} = E\{y\} + \frac{\partial y}{\partial n}(E\{n\} - E\{\tilde{n}\}) \tag{20}$$

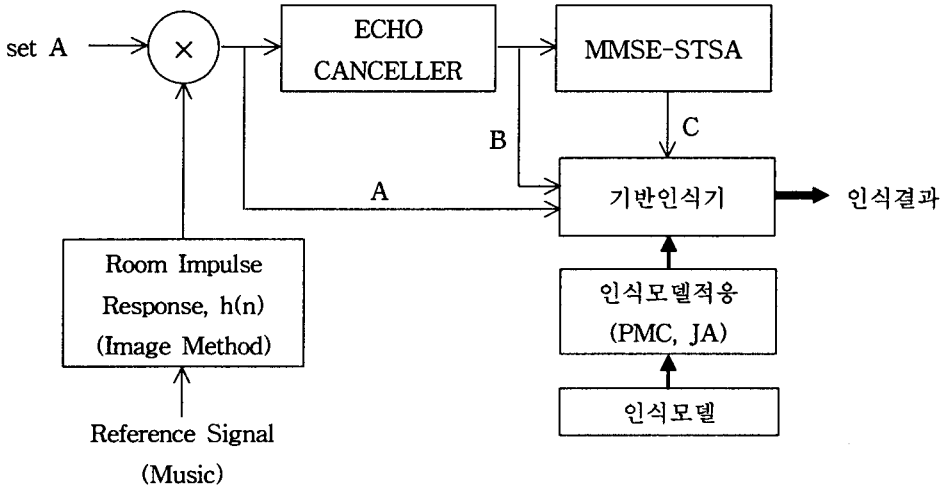
식(20)에서는 식(17)에서 구한 각 혼합성분별의 Jacobian 행렬을 이용하게 된다. 식(20)에서 $E\{y\}$는 미리 그 파라미터 값을 알고 있는 기준 HMM의 평균값이고 $E\{\tilde{y}\}$는 인식시에 사용되는 새로이 추정된 평균값이다. 잡음음성 y의 차분벡터 (delta-cepstrum)와 공분산에 대해서도 위와 비슷한 과정을 거쳐서 모델보상이 이루어지게 된다.

3. 실험결과

본 장에서는 음질개선방안과 인식모델 보상방식을 동시에 적용하여 잡음음성에 대한 인식성능을 평가하였다. 성능평가를 위한 기본 인식대상으로는 Aurora2 DB의 테스트 집합인 set A를 선택하였다. set A는 깨끗한 음성(clean)과 6가지의 신호대잡음비(-5 dB~20 dB)를 가지는 음성으로 구성되어있으며, 잡음의 종류로는 지하철, 군중, 자동차, 전시회 배경소음이 있다. 인식실험은 잡음환경에 따라 두 가지의 실험으로 분류되며, <그림 2(a)>와 <그림 2(b)>에 자세히 나타나 있다. <그림 2(a)>에서는 set A의 잡음환경을 그대로 이용하는데 반해, <그림 2(b)>에서는 차량실내의 잡음환경을 이용하였다. 따라서 <그림 2(a)>의 실험에서는 set A의 음성데이터를 그대로 이용하였으며 MMSE- STSA를 이용하여 음질개선을 시도하였다. 이와는 달리 <그림 2(b)>의 실험에서는 인공적으로 생성한 반향음을 set A의 잡음음성에 2차 오염시켜 이용하였다. 반향제거기는 구조가 간단하고 계산량이 작은 NLMS(Normalized Least Mean Square) 알고리즘을 갖는 적응필터를 이용하여 구현하였다[7]. 반향제거기를 거친 후 반향이 제거된 음성신호에서 남아있는 잔여잡음을 줄이기 위해 후처리 과정으로 MMSE-STSA 기반의 음성개선 기법을 적용하였다. 자동차 오디오 출력에 의한 음악 및 음성을 포함하는 반향 신호를 생성하기 위하여 가로×세로×높이를 각각 1.5 × 2.0 × 1 m로 가상의 자동차 실내 환경을 가정하고 image method를 사용하여 룸 임펄스 응답을 구하였다. 음악신호를 위에서 구한 룸 임펄스 응답에 적용하여 반향음을 생성하고, 인식에 사용될 잡음음성에 반향음을 더하여 인식실험에 사용될 테스트 데이터를 생성하였다(박철호 등, 2006).



(a) 부가잡음만이 고려된 경우



(b) 부가잡음 외에도 자동차 반향음이 고려된 경우

그림 2. 전체 인식실험의 모식도

인식특징벡터로는 12 차의 MFCC (mel-frequency cepstrum coefficient)와 로그에너지 그리고 이들에 대한 delta 및 acceleration을 포함하여 총 39 차원의 특징벡터를 사용하였다. 기반인식기의 각 단어모델의 연속밀도 HMM 구조는 left-to-right의 형태를 사용하며, 각 단어모델 별 상태의 개수는 앞뒤의 dummy 상태를 포함하여 묵음 모델이 5 개, sp모델이 3 개, 나머지 숫자 모델들이 18 개이고 상태별 혼합성분의 개수는 3 이다. 여기서 sp모델은 단어 간에 존재 가능한 묵음에 대한 모델이며, sp모델의 2 번째 상태와 묵음 모델의 3번째 상태는 서로 공유된다(tied-state). 한편, 인식모델 보상방식으로는 PMC와 JA를 이용하였으며, 효과적인 인식모델 보상방법을 조사하기 위하여 모델의 파라미터 평균, 델타평균, 분산을 차례대로 추가하면서 보상해주었다.

<표 1>과 <표 2> 그리고 <표 3>은 <그림 2(a)>와 관련된 실험의 인식결과로서 Aurora2 DB

set A 잡음음성을 이용한 인식결과이다. 먼저, <표 1>은 음질개선 방법인 MMSE-STSA를 적용한 경우의 인식결과에 대해서 보다 자세히 보여주고 있는데, 전반적으로 매우 양호한 성능을 나타냄을 알 수 있다. 그리고 <표 2>는 인식모델 보상방식에 대한 결과로 PMC와 JA를 적용하였으며, 0 dB에서 20 dB까지의 인식율의 평균값을 보여주고 있다. <표 3>은 이 두 방식을 결합한 결과이다.

표 1. Aurora2 DB 잡음음성에 대해서 MMSE-STSA 를 적용한 경우의 인식율 (%)

SNR	지하철	군중	자동차	전시회	평균
Clean	98.83	98.88	98.90	99.17	98.94
20 dB	95.92	91.17	97.88	96.30	95.32
15 dB	90.70	81.20	97.11	94.14	90.78
10 dB	80.53	64.21	92.78	87.87	81.35
5 dB	64.63	42.23	79.96	72.76	64.89
0 dB	37.00	14.90	41.96	44.86	34.68
-5 dB	12.40	0.00	11.78	16.08	10.06
평균	73.75	58.74	81.94	79.19	**73.41**

표 2. Aurora2 DB 잡음음성에 대해서 인식모델 보상을 적용한 경우의 인식율(%)

전체 평균 (20 dB~0 dB)	보상방법		
	평균	평균+델타평균	평균+델타평균+분산
PMC	**65.29**	**65.21**	**63.44**
JA	**65.77**	**65.83**	**66.09**

표 3. Aurora2 DB 잡음음성에 대해서 MMSE-STSA와 인식모델 보상을 동시에 적용한 경우의 인식율 (%)

전체 평균 (20 dB~0 dB)	보상방법		
	평균	평균+델타평균	평균+델타평균+분산
PMC	**76.71**	**77.31**	**76.34**
JA	**77.17**	**77.95**	**77.82**

본 연구에서 사용한 기반인식기의 Aurora2 DB set A에서의 베이스라인 인식율은 평균 62.54(%)였다. 이와 비교하여 <표 1>의 결과를 보면 MMSE-STSA기법으로 음질을 개선할 경우, 73.41(%)로 인식률이 증가하는 것을 볼 수 있다. <표 2>를 보면 인식모델 보상방식을 기반인식기에 적용할 경우에 인식률은 최대 66.09(%)로 인식률이 증가하는 것을 볼 수 있다. 이는 JA 방식에서 평균과 델타 평균, 분산을 보상해주는 경우였다. 위의 결과에서 우리는 MMSE-STSA 기반의 음질개선 방식이 인식모델 보상방식에 비해서 우수한 성능을 나타냄을 알 수 있다. 이것은 기본적으로 MMSE-STSA 방식이 잡음음성인식에 있어서 그 성능이 매우 효과적인 알고리듬임을 의미하며. 상대적으로 PMC방식에서 잡음성분의 추정시에 잡음음성신호의 처음 몇 개의 프레임만을 이용함으로서 잡음신호에 통계정보가 충분히 정확하게 추출되지 못하는 점도 있다고 생각된다. 한편 이 두 방식의 결합에 의한 성능은 <표 3>에 나타나 있으며, 최대 인식율은 77.95(%)로 MMSE-STSA 기반

의 음질개선을 적용한 후에, 잔류잡음신호를 기반으로 JA 방식을 통하여 평균과 델타평균을 보상해 줄 때 가장 좋은 성능이 나타났다. 이와 같은 실험결과를 종합해서 볼 때, 음질개선 방식과 인식모델 보상방식을 단독으로 사용하는 것 보다 두 방식의 결합을 통한 접근 방식을 사용함으로서 인식성능이 크게 향상되는 것을 알 수 있다.

다음에 나타낼 <표 4>와 <표 5>는 <그림 2(b)>와 관련된 실험이다. 여기서는 차량실내의 반향음을 고려하고 있으며, 음질개선 방법으로 MMSE-STSA기법과 반향제거기를 사용하고 있다. 그리고 음질개선에 따라 변화하는 성능을 알아보기 위하여, 기반인식기의 입력신호를 A, B, C로 나누어 실험하였다. 여기서 신호 A는 아무런 음질개선처리도 하지 않은 오염된 음성신호이며, B는 A신호를 입력으로 하여 반향제거를 한 신호이다. 그리고 C는 다시 신호 B를 입력으로 하여 최종적으로 MMSE-STSA기법을 적용한, 최종적으로 음질이 개선된 신호이다.

표 4. Aurora2 DB에서 차량실내 반향음을 고려한 경우의 인식율 (%)

전체 평균 (20 dB~0 dB)	기반인식기
A	24.26
B	52.77
C	71.15

표 5. Aurora2 DB에서 차량실내 반향음을 고려한 경우의 인식율 (%) (인식모델 보상이 적용된 경우)

전체 평균 (20dB~0dB)		보상방법		
		평균	평균+델타평균	평균+델타평균+분산
A	PMC	26.20	26.39	24.21
	JA	25.01	25.02	26.89
B	PMC	54.81	54.30	49.22
	JA	53.54	52.70	51.42
C	PMC	75.33	75.84	72.80
	JA	74.17	74.90	74.02

<표 4>는 인식모델 보상 방식을 이용하지 않은 경우의 인식결과이며, A에서 C로 갈수록 24.26(%)에서 71.15(%)로 성능이 크게 향상되는 것을 볼 수 있다. <표 5>에서는 인식모델 보상 방식이 기반인식기에 적용된 경우의 인식성능을 나타내고 있다. 입력이 A일 때, 최대 인식률은 26.89(%)로 이는 JA를 이용하고 평균과 델타평균, 분산을 보상해 주었을 때 얻어진 값이다. 신호 A에는 잡음음성과 반향음이 아무런 처리과정을 거치지 않고 포함되어 있으므로 인식율이 상당히 저조하게 나타나는 것을 알 수 있다. 한편 입력이 신호 B인 경우에는 PMC를 이용하고 평균만 보상해줄 때 54.81(%)로 가장 높은 인식률을 얻을 수 있었다. 그리고 입력신호가 C일 때는 PMC를 이용하고 평균, 델타평균을 보상해줌으로써 75.84(%)의 최고 성능을 얻을 수 있었다. <표 5>에서 우리는 최고 인식율이 75.84(%)가 됨을 알 수 있으며 이는 <표 4>에서 얻어진 최고 인식율 71.15(%) 비해서 상당한 향상이 있음을 의미한다. 이는 앞의 <표 1>, <표 2>, <표 3>의 결과에서 보았듯이, 음질개선

방식 및 반향제거기만을 사용하는 것보다는 인식모델 보상방식을 결합할 경우에 성능이 보다 향상됨을 의미한다. 한편, 모델 보상방법에서 분산을 보상해주는 경우에 PMC에서는 오히려 인식성능이 떨어지고, JA에서는 성능이 비슷하게 유지되는 것을 알 수 있다. 그러나 평균, 델타평균을 보상해주는 경우에는 PMC와 JA 모두 인식률이 좋아진다. 따라서 모델 보상에서는 평균과 델타평균 파라미터만을 적응하는 것이 비교적 좋은 성능을 기대할 수 있다고 생각된다.

4. 결 론

본 연구에서는 기존의 음질개선 방식 중 우수하다고 알려진 MMSE-STSA 방식과 모델보상 방식에서 가장 대표적으로 사용되는 JA/PMC 방식을 결합하여 잡음 환경 하에서 인식실험을 수행하였다. MMSE-STSA 방식은 부가잡음이 단독으로 존재하는 경우와 반향음이 존재하는 경우 모두에 있어서 매우 우수한 성능을 보임을 알 수 있었다. PMC나 JA 등의 모델보상방식은 MMSE-STSA 방식에 비해서는 다소 성능이 떨어지는 현상을 볼 수 있었다. 하지만, MMSE-STSA를 적용한 후에 모델보상방식을 부가적으로 적용할 경우에는 상당한 인식성능의 향상을 볼 수 있었고, 따라서 이러한 두 가지 방식의 결합이 매우 효과적임을 알 수 있었다.

참 고 문 헌

[1] Gales, M. & Young, S. 1993. "Parallel model combination for speech recognition in noise." *Tech. Rep.* 135, Cambridge University.

[2] Moreno, P. 1996. *Speech recognition in noisy environments.* Ph.D. Thesis. Carnegie Mellon University.

[3] Hung, J-W., Shen, J-L. & Lee, L-S. 2001. "New approaches for domain transformation and parameter combination for improved accuracy in parallel model combination (PMC) techniques." *IEEE Trans. Speech and Audio Processing* 9(8), 842-855.

[4] Sagayama, S., Yamaguchi, Y. & Takahashi, S. 1997. "Jacobian adaptation of noisy speech models." *IEEE Workshop on Automatic Speech Recognition and Understanding*, 396-403.

[5] Boll, S. 1979. "Suppression of acoustic noise in speech using spectral subtraction." *IEEE Trans. Acoust., Speech, Signal Processing* 27(2), 113-120.

[6] Ephraim, Y. & Malah, D. 1984. "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator." *IEEE Trans. on ASSP* 32(6), 1109-1121.

[7] 박철호, 배건성. 2006. "자동차 환경에서의 음성인터페이스를 위한 핵심 기반기술 개발." *경북대학교 연구보고서.*

접수일자: 2007. 4. 26
게재결정: 2007. 5. 30

▲ 김희근
서울특별시 구로구 구로동 188-5번지 (우: 152-050)
(주) 글로벌테크
Tel: +82-2-6300-4111
E-mail: hkkim@globalteq.com

▲ 정용주
대구광역시 달서구 신당동 1000번지 (우:704-701)
계명대학교 전자공학과
Tel: +82-53-580-5925
E-mail: yjjung@kmu.ac.kr

▲ 배건성
대구광역시 북구 산격동 1370번지 (우: 702-701)
경북대학교 전자전기컴퓨터학부
Tel: +82-53-950-5527
E-mail: ksbae@ee.knu.ac.kr