MMSE-STSA 기반의 음성개선 기법에서 잡음 및 신호 전력 추정에 사용되는 파라미터 값의 변화에 따른 잡음음성의 인식성능 분석

박철호(경북대), 배건성(경북대)

<차 례>

<Abstract>

Performance Analysis of Noisy Speech Recognition Depending on Parameters for Noise and Signal Power Estimation in MMSE-STSA Based Speech Enhancement

Chul Ho Park, Keun Sung Bae

The MMSE-STSA based speech enhancement algorithm is widely used as a preprocessing for noise robust speech recognition. It weighs the gain of each spectral bin of the noisy speech using the estimate of noise and signal power spectrum. In this paper, we investigate the influence of parameters used to estimate the speech signal and noise power in MMSE-STSA upon the recognition performance of noisy speech. For experiments, we use the Aurora2 DB which contains noisy speech with subway, babble, car, and exhibition noises. The HTK-based continuous HMM system is constructed for recognition experiments. Experimental results are presented and discussed with our findings.

* Keywords: Forgetting factor, Speech enhancement, Speech recognition, MMSE-STSA, HMM.

1. 서 론

잡음음성에 대한 인식 성능을 향상시키기 위한 기술은 크게 음성개선(speech enhancement), 특징보상(feature compensation), 모델적응(model adaptation) 기법으로 구분된다[1]. 그중에서 음성개선 기법[2-4]은 잡음이 섞인 입력음성에서 잡음 성분을 추정하여 빼줌으로써 신호대잡음비(SNR: Signal-to-Noise Ratio)를 높여주는 기술로, 음성인식 시스템의 전처리 과정으로 적용이 용이하므로 잡음음성에 대한 인식 성능을 향상시키기 위한 방법으로 많이 사용된다. 대표적인 음성개선 기법으로는 스펙트럼차감법(spectral subtraction), Wiener 필터링, MMSE-STSA (Minimum Mean Square Error-Short Time Spectral Amplitude) 기법 등을 들 수 있는데, 이중에서 MMSE-STSA 기반의 음성개선 기법[4]이 알고리즘이 간단하면서 우수한 성능을 나타내므로 많이 이용되고 있다.

MMSE-STSA에 기반한 음성개선 기법은 음성과 잡음의 스펙트럼이 통계적으로 서로 독립적인 가우시안(Gaussian) 분포를 갖는다는 가정 하에, 각 스펙트럼 빈(bin)에 대해 추정된 신호 및 잡음 전력을 이용하여 이득함수를 구하여 곱해줌으로써 잡음성분을 줄여준다. 따라서 이득함수의 추정 정확도에 따라 음성개선 성능이 달라지게 되는데, 이득함수는 사전 신호대잡음비(a priori SNR)와 사후 신호대잡음비(a posteriori SNR)의 함수로 주어지기 때문에 정확한 사전 신호대잡음비와 사후 신호대잡음비를 구할 필요가 있다. 사전 신호대잡음비와 사후 신호대잡음비를 구하기 위해서는 각 스펙트럼 빈에 대한 음성신호 및 잡음신호의 전력을 추정해야 한다. 각 프레임마다 문턱치 기반의 VAD(Voice Activity Detection)를 사용하여 음성의 존재 유무를 결정하고 묵음 구간에서만 잡음전력을 추정하는 방법을 hard-decision 방법이라고 한다.

Hard-decision 방법으로 프레임 단위의 각 스펙트럼 빈에 대한 음성신호 및 잡음신호의 전력을 추정할 때, 이전 프레임까지의 추정된 전력 값을 이용하여 재귀적으로 현재 프레임의 전력 값을 추정하게 된다. 이때 이전 프레임까지의 전력 값을 얼마정도 반영할 것인지를 나타내는 망각인자(forgetting factor) 파라미터가 사용된다. 이 파라미터 값은 신호전력 및 잡음전력 추정에 영향을 미치게 되는데, 입력 잡음음성의 신호대잡음비가 낮은 경우에는 VAD의 신뢰성이 떨어질 수 있으므로 현재 프레임의 신호전력 및 잡음전력을 적게 반영하고, 입력 잡음음성의 신호대잡음비가 높은 경우에는 현재 프레임의 신호전력 및 잡음전력을 더 많이 반영하는 게 바람직하다고 할 수 있다.

본 논문에서는 MMSE-STSA 기반의 음성개선에서 음성신호 및 잡음신호의 전력을 추정할 때 사용되는 파라미터 값의 변화에 따른 잡음음성의 인식성능을 분석한다. 다양한 신호대잡음비를 갖는 잡음음성에 대한 인식실험을 통해 신호 및 잡음 전력 추정에 사용되는 파라미터의 값이 인식성능에 어떠한 형태로 영향을

미치는가를 분석하여 적절한 파라미터 값을 제시하고자 한다. 본 논문의 구성은 다음과 같다. 2장에서 본 연구에서 사용한 hard-decision 방법의 MMSE-STSA 음성개선 기법을 설명하고, 3장에서는 인식실험에 사용된 인식시스템 및 잡음음성 DB와 실험결과를 제시하며, 4장에서 결론을 맺는다.

2. Hard-decision 방법을 이용한 MMSE-STSA 기반의 음성개선

잡음음성 신호 $x(n)$의 k번째 스펙트럼 빈 X_k는 식 (1)과 같이 원 음성신호의 스펙트럼 S_k와 잡음신호의 스펙트럼 V_k의 합으로 표현된다.

$$\begin{aligned} X_k &= S_k + V_k \\ R_k e^{j\vartheta_k} &= A_k e^{j\theta_k} + V_k \end{aligned} \tag{1}$$

여기서 R_k 및 ϑ_k는 잡음음성의 스펙트럼 크기 및 위상을, A_k 및 θ_k는 원 음성신호의 스펙트럼 크기 및 위상을 나타낸다.

음성과 잡음의 스펙트럼이 통계적으로 서로 독립적인 가우시안 랜덤변수라고 가정하면, t번째 프레임의 k번째 스펙트럼 빈 $A_k(t)$에 대한 MMSE 추정치 $\hat{A}_k(t)$는 식 (2)로 주어진다. 이때 사후 신호대잡음비 $\gamma_k(t)$는 식 (3)과 같이 잡음음성과 추정된 잡음의 분산을 이용하여 직접 구하고, 사전 신호대잡음비 $\xi_k(t)$는 식 (4)로 주어지는 decision-directed 방법으로 구한다. 식 (4)에서 $\lambda_{sk}(t)$와 $\lambda_{vk}(t)$는 각각 원 음성과 잡음의 k번째 스펙트럼 빈의 전력 값을 의미하며, α는 이전 프레임까지의 추정된 전력을 반영하는 망각지수이고 $P[\bullet]$는 양의 값을 가지기 위한 연산자이다. 그리고 식 (2)에서 $\Gamma(\bullet)$는 Gamma 함수를 의미하고 $M(\bullet)$은 식 (5)와 같이 정의되는 함수로서 $I_0(\theta)$와 $I_1(\theta)$는 각각 0차와 1차 modified Bessel 함수이다.

$$\begin{aligned} \hat{A}_k(t) &= E\{A_k(t)|X_k(t)\} \\ &= \Gamma(1.5)\frac{\sqrt{\dfrac{\xi_k(t)\gamma_k(t)}{1+\xi_k(t)}}}{\gamma_k(t)} M\left(\frac{\xi_k(t)\gamma_k(t)}{1+\xi_k(t)}\right) R_k(t) \\ &= G_{MMSE}(\xi_k(t),\gamma_k(t))R_k(t) \end{aligned} \tag{2}$$

$$\gamma_k(t) \equiv \frac{R_k^2(t)}{\lambda_{vk}(t)} \tag{3}$$

$$\xi_k(t) \equiv \frac{\lambda_{sk}(t)}{\lambda_{vk}(t)} = \alpha \frac{\hat{A}_k^2(t-1)}{\lambda_{vk}(t-1)} + (1-\alpha)P\left[\gamma_k(t)-1\right] \tag{4}$$

$$M(\theta) = e^{-\frac{\theta}{2}}\left[(1+\theta)I_0\left(\frac{\theta}{2}\right)+\theta I_1\left(\frac{\theta}{2}\right)\right] \tag{5}$$

MMSE-STSA 추정치에 음성존재확률(SPP: Speech Present Probability)을 도입하면 식 (6)과 같이 수정된 MMSE-STSA 추정치를 얻을 수 있다. 여기에서 $P\left(H_k^1|X_k(t)\right)$는 주어진 $X_k(t)$에 대한 음성존재확률로서 식 (7)과 같이 likelihood ratio $\Lambda_k(t)$와 사전 음성부재확률 $P(H_k^0)$로 계산되며, H_k^1과 H_k^0는 각각 음성의 존재와 부재에 대한 가설이고 $P\left(H_k^1\right)$는 사전 음성존재확률을 나타낸다. 사전 음성부재확률 $P\left(H_k^0\right)$은 보통 0.2로 설정된다[5].

$$\begin{aligned}\hat{A}_k(t) &= P\left(H_k^1|X_k(t)\right)G_{MMSE}(\xi_k(t),\gamma_k(t))R_k(t) \\ &= \frac{\tilde{\Lambda}_k(t)}{1+\tilde{\Lambda}_k(t)}G_{MMSE}(\xi_k(t),\gamma_k(t))R_k(t)\end{aligned} \tag{6}$$

$$\begin{aligned}\tilde{\Lambda}_k(t) &= \frac{P\left(H_k^1\right)P\left(X_k(t)|H_k^1\right)}{P\left(H_k^0\right)P\left(X_k(t)|H_k^0\right)} = \frac{1-P\left(H_k^0\right)}{P\left(H_k^0\right)}\frac{P\left(X_k(t)|H_k^1\right)}{P\left(X_k(t)|H_k^0\right)} \\ &= \frac{1-P\left(H_k^0\right)}{P\left(H_k^0\right)}\Lambda_k(t)\end{aligned} \tag{7a}$$

$$\Lambda_k(t) \equiv \frac{1}{1+\xi_k(t)}\exp\left[\frac{\gamma_k(t)\xi_k(t)}{1+\xi_k(t)}\right] \tag{7b}$$

잡음구간에서만 잡음전력스펙트럼을 추정하는 hard-decision은 문턱치 기반의 가장 간단한 VAD 방법으로 입력음성 신호에서 프레임 단위로 식 (8)의 VAD 조건에 따라 음성신호가 포함되지 않고 잡음만 존재하는 구간을 찾아내고, 잡음전력스펙트럼 $\lambda_{vk}(t)$는 잡음구간에서만 식 (9)를 이용하여 잡음전력 스펙트럼의 값을

갱신한다. 만약 식 (8)의 조건을 만족하지 않으면, 음성구간으로 판단하여 잡음전력 스펙트럼 추정을 더 이상 수행하지 않고, 식 (10)과 같이 음성구간으로 판정된 현재 프레임을 기준으로 마지막에 추정된 잡음전력 스펙트럼을 그대로 사용한다. 식 (8)에서 M은 스펙트럼 빈의 수를 의미하고 t는 현재 프레임의 인덱스, T는 정규화된 문턱치 계산을 위한 상수 값, β는 잡음전력 스펙트럼을 반복적으로 갱신하기 위한 파라미터 값을 나타낸다.

$$\sum_{k=1}^{M}|R_k(t)|^2 - \sum_{k=1}^{M}\lambda_{vk}(t-1) < T\sum_{k=1}^{M}\lambda_{vk}(t-1) \tag{8}$$

$$\lambda_{vk}(t) = \lambda_{vk}(t-1) + \beta\left(|R_k(t)|^2 - \lambda_{vk}(t-1)\right) \tag{9}$$

$$\lambda_{vk}(t) = \lambda_{vk}(t-1) \tag{10}$$

본 논문에서는 MMSE-STSA 기반의 음성개선 성능에 주요 영향을 미치는 음성신호의 전력 추정에 사용되는 식 (4)의 파라미터 α와 잡음전력 추정에 사용되는 식 (9)의 파라미터 β 값의 변화가 잡음음성의 인식성능에 미치는 영향을 분석 하고자 한다.

3. 실험 및 검토

3.1 인식실험 환경

본 논문의 실험은 ETSI(European Telecommunications Standards Institute)에서 배포한 Aurora2 DB를 사용하며, 훈련은 clean condition으로 수행하고 테스트 집합 A에 대해 인식실험을 수행하였다. 테스트 집합 A는 4가지 잡음(subway, babble, car, exhibition)에 대해, 20dB부터 -5dB까지 6가지 잡음레벨에 대한 잡음음성으로 구성되며, 기본(baseline) 음성인식기는 Aurora2-HTK를 사용한다. 이것은 CUED-HTK 음성인식시스템[6]의 단어모델과 훈련절차를 따른 것으로 Aurora2 DB에 적합하도록 구성된 것이다. 단어모델은 one, two, three, four, five, six, seven, eight, nine, zero, oh의 11개로 정의되어 있고, 각 단어모델은 3 mixture, 16개의 state를 갖는 CHMM(Continuous Hidden Markov Model)으로 구성된다. 인식시스템에는 11개의 단어모델 외에 2개의 묵음 모델이 포함되어 있는데, 각각 3 state와 1 state CHMM으로 구성되어 있다. 특징파라미터는 23차 필터뱅크를 이용한 12차 MFCC와 1차 로그에너지, 그리고 각각의 delta 및 acceleration 을 포함한 총 39차로 구성된다. 분석

프레임의 크기는 25ms 이며, 10ms씩 이동시키면서 특징파라미터를 추출하였다.

음성인식에서 주로 사용하는 성능 측정치로 문장인식률(sentence correction), 단어인식률(word correction), 단어정확률(WA: Word Accuracy) 등이 있다. 일반적으로 음성인식의 성능 측정치로 단어인식률 보다는 삽입에러를 인식률에 포함시킨 단어정확률을 더 많이 사용하므로, 본 논문에서는 성능비교를 위한 인식률로 단어정확률을 사용한다. 단어정확률은 식(11)와 같이 정의된다[6].

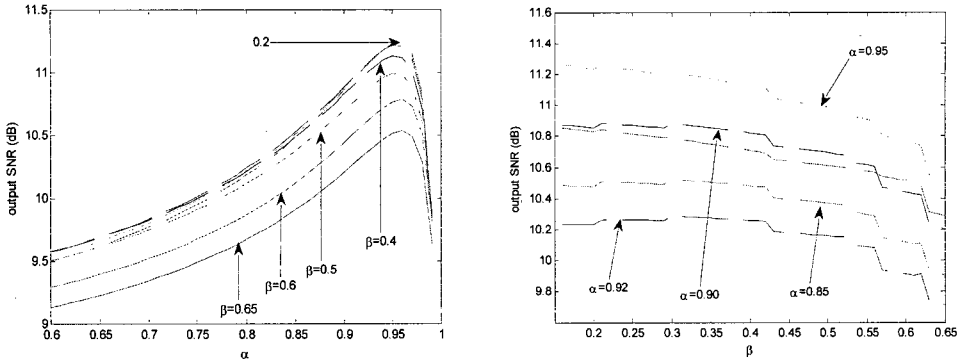
$$WA = \frac{H-I}{N} \times 100 \tag{11a}$$

$$H = N - D - S \tag{11b}$$

여기서 H는 올바르게 인식된 단어 수, I는 삽입된 단어 수, D는 삭제된 단어 수, S는 대치된 단어 수, N은 전체 테스트 단어 수를 의미한다.

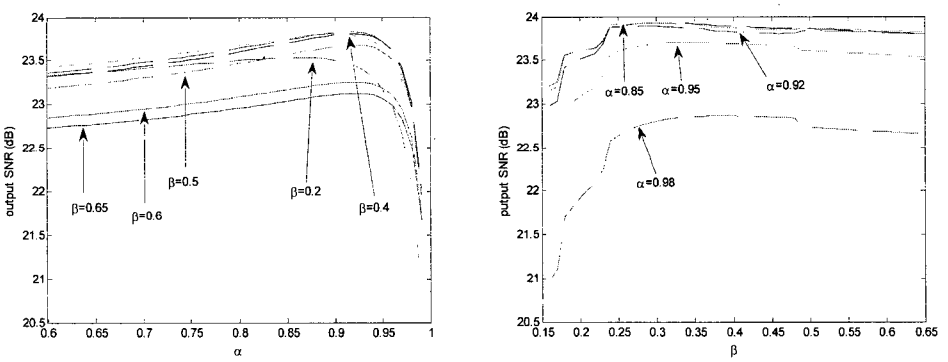
3.2 신호 및 잡음 전력 추정 파라미터에 따른 음성개선 결과

먼저 MMSE-STSA 기법에서의 잡음 및 신호전력 추정에 사용되는 파라미터 값에 따른 잡음제거 성능을 분석하기 위하여 영어문장 음성신호에 다양한 잡음(subway, babble, car, exhibition)을 인가하여 20dB부터 0dB까지 5dB간격으로 잡음음성 신호를 만든 후 파라미터 값의 변화에 따라 개선된 음성의 SNR을 측정하였다. VAD를 위한 문턱치인 T는 0.6으로 설정하였다. <그림 1>은 자동차 잡음의 입력 잡음음성의 SNR이 5dB 일 때의 파라미터 α, β 값의 변화에 따른 잡음이 제거된 출력신호의 SNR을 보인 것이다. 입력신호의 SNR이 상대적으로 낮으면, α는 0.95보다 크거나, β는 0.25보다 작을수록 더 개선된 결과를 나타냄을 볼 수 있다. <그림 2>는 입력신호의 SNR이 20dB인 자동차 잡음음성에서 파라미터 α, β의 변화에 따른 출력신호의 SNR을 나타낸 것인데, 이때는 α 값이 0.95 또는 그보다 크거나 β가 0.2보다 작아지면 음성개선 성능이 더 저하하게 되고, 대략 α는 0.90 이하에서 β는 0.3~0.5근처에서 더 개선된 결과를 보임을 알 수 있다. Subway, exhibition, babble과 같은 잡음의 경우에도 앞에서와 비슷한 양상을 보였었다.



(a) (b)

<그림 1> 입력음성의 SNR이 5dB인 자동차 잡음음성의 음성개선 결과



(a) (b)

<그림 2> 입력음성의 SNR이 20dB인 자동차 잡음음성의 음성개선 결과

3.3 인식실험 결과

<표 1>은 음성개선 기법을 적용하지 않은 잡음음성에 대한 기본인식기의 인식결과를 보인 것이다. 입력음성의 SNR이 20dB 및 15dB인 경우 babble 잡음을 제외하고는 90% 이상의 인식률을 보이고 있는데, 사람의 목소리와 비슷한 특성을 갖는 babble 잡음의 경우가 인식성능이 가장 나쁘게 나타남을 볼 수 있다. <표 2>부터 6은 기본 인식기의 전처리 과정으로 MMSE-STSA 기반의 음성개선 기법을 적용하여 음성개선을 한 후에 인식실험을 한 결과를 나타낸 것이다. 사전 음성부재 확률 $P(H_k^0)$는 0.2, VAD를 위한 문턱치 T는 0.6으로 설정하였다[5]. 백색잡음이 첨가된 음성의 음성개선을 위한 예비실험에서도 <그림 1>, <그림 2>에서와 비슷한 양상으로 입력음성의 SNR이 높은 경우에는 α 값을 작게, β 값을 크게 할수록,

SNR이 낮은 경우에는 상대적으로 α 값을 크게, β 값은 작게 해주는 것이 더 좋은 잡음제거 개선 효과가 있었다. 따라서 입력음성의 SNR에 따른 인식성능의 차이를 분석하기 위하여 신호 및 잡음 전력의 추정에 사용되는 파라미터 (α , β) 값을 (0.85 , 0.5), (0.90 , 0.4), (0.92 , 0.3), (0.95 , 0.25), (0.98, 0.2)의 5가지 경우에 대해서 실험하였다.

<표 1> 잡음음성에 대한 기본 인식시스템의 인식률 (%)

SNR	Subway	Babble	Car	Exhibition	평균
Clean	98.93	99.00	98.96	99.20	99.02
20 dB	97.05	90.15	97.41	96.39	95.25
15 dB	93.49	73.76	90.04	92.04	87.33
10 dB	78.72	49.43	67.01	75.66	67.71
5 dB	52.20	26.81	34.09	44.83	39.48
0 dB	26.01	9.28	14.46	18.05	16.95
-5 dB	11.18	1.57	9.39	9.60	7.94
평균	69.49	49.89	60.60	65.39	61.34

<표 2> MMSE-STSA 기반의 음성개선 후의 인식률 ($\alpha = 0.85$, $\beta = 0.5$) (%)

SNR	Subway	Babble	Car	Exhibition	평균
Clean	98.74	98.94	98.81	99.07	98.89
20 dB	96.01	90.99	97.97	95.62	95.15
15 dB	90.30	80.38	96.48	93.49	90.16
10 dB	79.67	63.57	89.92	86.89	80.01
5 dB	61.96	41.48	70.83	68.78	60.76
0 dB	33.68	15.78	31.76	37.80	29.76
-5 dB	11.15	0	8.71	12.00	7.97
평균	72.32	58.44	77.39	76.52	71.17

<표 3> MMSE-STSA 기반의 음성개선 후의 인식률 ($\alpha = 0.90$, $\beta = 0.4$) (%)

SNR	Subway	Babble	Car	Exhibition	평균
Clean	98.74	98.91	98.84	99.07	98.89
20 dB	95.89	90.78	97.79	95.50	94.99
15 dB	90.64	80.38	96.78	92.93	90.18
10 dB	80.07	64.21	91.17	86.21	80.42
5 dB	63.03	42.44	73.93	69.92	62.33
0 dB	35.00	16.08	34.83	39.37	31.32
-5 dB	10.87	0	9.04	12.43	8.09
평균	72.93	58.78	78.90	76.79	71.85

<표 4> MMSE-STSA 기반의 음성개선 후의 인식률 ($\alpha = 0.92, \beta = 0.3$) (%)

SNR	Subway	Babble	Car	Exhibition	평균
Clean	98.71	98.91	98.84	99.07	98.88
20 dB	95.67	89.84	97.61	95.37	94.62
15 dB	90.54	79.87	96.63	92.72	89.94
10 dB	79.74	63.42	91.62	85.50	80.07
5 dB	63.37	41.93	74.50	69.64	62.36
0 dB	35.22	15.87	36.18	39.52	31.70
-5 dB	10.84	0	9.42	12.50	7.34
평균	72.91	58.19	79.31	76.55	71.74

<표 5> MMSE-STSA 기반의 음성개선 후의 인식률 ($\alpha = 0.95, \beta = 0.25$) (%)

SNR	Subway	Babble	Car	Exhibition	평균
Clean	98.74	98.85	98.78	99.07	98.86
20 dB	95.67	90.18	97.38	94.94	94.54
15 dB	90.73	80.93	96.69	91.73	90.02
10 dB	80.93	64.63	92.69	84.88	80.78
5 dB	64.78	43.11	77.72	70.93	64.14
0 dB	38.19	16.69	40.29	42.30	34.37
-5 dB	11.48	0	10.29	13.67	8.86
평균	74.06	59.11	80.95	76.96	72.77

<표 6> MMSE-STSA 기반의 음성개선 후의 인식률 ($\alpha = 0.98, \beta = 0.2$) (%)

SNR	Subway	Babble	Car	Exhibition	평균
Clean	98.68	98.76	98.81	99.04	98.82
20 dB	95.27	90.72	96.63	93.52	94.04
15 dB	90.21	81.77	95.29	89.79	89.27
10 dB	80.84	67.90	92.13	83.15	81.01
5 dB	67.21	46.43	81.21	69.58	66.11
0 dB	41.36	19.71	49.18	45.08	38.83
-5 dB	14.25	0	13.39	17.12	11.19
평균	74.98	61.31	82.89	76.22	73.85

<표 7> MMSE-STSA의 파라미터 값 변화에 따른 평균 인식률 (%)

SNR \ (α, β)	(0.85, 0.5)	(0.90, 0.4)	(0.92, 0.3)	(0.95, 0.25)	(0.98, 0.2)
20 dB	95.15	94.99	94.62	94.54	94.04
15 dB	90.16	90.18	89.94	90.02	89.27
10 dB	80.01	80.42	80.07	80.78	81.01
5 dB	60.76	62.33	62.36	64.14	66.11
0 dB	29.76	31.32	31.70	34.37	38.83

<표 2> ~ <표 6>에서, 음성개선 기법을 적용함으로써 기본 인식시스템에 비해 평균 9.83%에서 12.51% 까지의 인식률 향상이 얻어짐을 볼 수 있다. 특히 인식률이 급격히 떨어지는 10dB 입력음성에 대해서도 음성개선 기법을 적용함으로써 67.71%에서 81.01%로 평균 인식률이 증가함을 볼 수 있다. 하지만 입력음성의 SNR이 20dB 이상인 양호한 음질의 음성에서는 음성개선 기법을 적용함으로써 다소 발생하게 되는 왜곡으로 인해 인식률이 기본 인식시스템에 비해 약간 저하됨을 볼 수 있다. <표 7>은 신호 및 잡음 전력의 추정에 사용되는 파라미터 (α , β) 값에 따른 인식성능을 비교하기 위하여 20dB에서 0dB 까지의 입력음성에 대한 평균 인식률을 보인 것이다. 입력음성의 SNR이 높은 20dB 및 15dB인 경우에는 대체로 α 값이 작고 β 값이 큰 경우에 인식률이 다소 높음을 볼 수 있으며, 입력음성의 SNR이 낮은 5dB 및 0dB인 경우에는 α 값이 클수록, β 값이 작을수록 인식률이 향상됨을 볼 수 있다. 이것은 입력음성의 SNR이 낮은 경우에는 잡음구간의 검출이 어려우므로 이전 프레임까지에서 추정된 값을 더 많이 반영하고 현재 프레임에서의 신호 및 잡음 전력 값을 적게 반영하는 것이 바람직하다는 것을 의미하며, <그림 1> 및 <그림 2>의 음성개선 결과와도 부합된다고 판단된다.

4. 결 론

본 논문에서는 MMSE-STSA 기반의 음성개선 기법에서 음성신호 및 잡음신호의 전력을 추정할 때 사용되는 파라미터 α 및 β 값의 변화에 따른 잡음음성의 인식성능을 분석하였다. 이를 위해 ETSI에서 제공하는 Aurora2 DB 및 인식시스템을 사용하여 다양한 SNR을 갖는 잡음음성에 대한 인식실험을 수행하였다. 입력음성의 SNR이 높은 20dB 및 15dB인 경우에는 대체로 α 값이 작고 β 값이 큰 경우에 인식률이 다소 높았으며, 입력음성의 SNR이 낮은 5dB 및 0dB인 경우에는 α 값이 클수록, β 값이 작을수록 인식률이 향상됨을 볼 수 있었다. 실험결과를 종합해 볼 때, 잡음이 상당히 존재하는 10dB 이상의 실제 환경에서의 사용 가능한 인식률을 고려할 경우 MMSE-STSA 기반의 음성개선 기법에서 신호전력 추정에 사용되는 α는 0.9 내외, 잡음전력 추정에 사용되는 파라미터 β는 0.3 내외의 값을 설정하는 것이 적절하다고 생각된다.

참 고 문 헌

[1] 김남수, "잡음 환경에서의 음성인식", *Telecommunication Review,* 제13권, 5호, pp. 650-661, 2003년 10월.

[2] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction", *in Proc. ICASSP*, Vol. ASSP-27, no. 2, pp. 113-120, Apr. 1979.

[3] Y. Ephraim, D. Malah, and B. H. Juang, "Speech enhancement based on hidden Markov modeling," *in Proc. ICASSP*, pp. 353-356, 1989.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," *in Proc. ICASSP*, Vol. ASSP-32, no. 6, pp. 1109-1121, Dec. 1984.

[5] 정성윤, *잡음환경에 강인한 음성인식을 위한 MMSE-STSA 추정치 기반의 특징파라미터 추출*, 경북대학교 박사학위 논문, 2004.

[6] S. Young, D. Kershaw, J. Odell et al., *The HTK Book version 3.0*, 2000.

접수일자: 2006년 2월 15일
게재결정: 2006년 3월 13일

▶ 박철호(Chul-Ho Park)
주소: 702-701 대구광역시 북구 산격동 1370번지 경북대학교
소속: 경북대학교 공과대학 전자공학과
전화: 053) 940-8627
Fax : 053) 940-8827
E-mail: chilo@mir.knu.ac.kr

▶ 배건성(Keun-Sung Bae) : 교신저자
주소: 702-701 대구광역시 북구 산격동 1370번지 경북대학교
소속: 경북대학교 전자전기공학부
전화: 053) 950-5527
Fax : 053) 940-8827
E-mail: ksbae@ee.knu.ac.kr