한국음향학회지 제25권 제8호 pp. 383-388 (2006)

음성 향상에서 강인한 새로운 선행 SNR 추정 기법에 관한 연구

A Novel Approach to a Robust A *Priori* SNR Estimator in Speech Enhancement

박 윤 식*, 장 준 혁*
(Yun-Sik Park*, Joon-Hyuk Chang*)

*인하대학교 전자전기공학부

(접수일자: 2006년 9월 7일; 수정일자: 2006년 11월 6일; 채택일자: 2006년 11월 17일)

본 논문에서는 잡음 환경에서 단일 마이크로폰의 음성 향상에 대한 새로운 기법을 제시했다. 일반적으로 널리 알려진 스펙트럼 차감법에 근거한 음성 향상 기술은 신호 대 잡음비에 따른 스펙트럼 이득으로 표현 된다. 대표적인 Ephraim 과 Malah의 decision-directed (DD) 추정치는 잡음 구간에서 효율적으로 뮤지컬 잡음을 제거하지만 음성 구간에서는 이전 프레임의 음성 스펙트럼 성분에 더 큰 비중을 두기 때문에 *a priori* SNR의 프레임 지연이 발생한다. 따라서 DD에 의해 추정된 *a priori* SNR이 적용된 잡음 제거 이득은 현재 프레임보다 이전 프레임에 영향을 받으므로 음성 전이 구간에서 잡음 제거 성능을 저하시킨다.

본 논문은 DD의 가중치 파라미터에 Sigmoid Type의 함수를 적용하여 계산적으로는 간단하지만 효과적인 음성 향상 알고리즘을 제안한다. 제안된 접근 방식은 DD의 주요 파라미터인 *a priori* SNR 지연의 문제점을 해결하면서 뮤지컬 잡음 제거에 우수한 DD의 이점은 유지한다. 제안된 알고리즘의 성능은 다양한 잡음 환경에서 ITU-T P.862 Perceptual Evaluation of Speech Quality (PESQ) 와 Mean Opinion Score (MOS), 그리고 음성 스펙트로그램 (Spectrogram)에 의해 평가 했고 기존의 DD의 고정된 가중치 파라미터를 사용 했을 때 보다 향상된 결과를 나타내었다.

핵심용어: 선행 SNR, Decision-Directed 기법, 음성향상, Sigmoid Type

투고분야: 음성 처리 분야 (2.3)

This paper presents a novel approach to single channel microphone speech enhancement in noisy environments. Widely used noise reduction techniques based on the spectral subtraction are generally expressed as a spectral gain depending on the signal-to-noise ratio (SNR). The well-known decision-directed (DD) estimator of Ephraim and Malah efficiently reduces musical noise under the background noise conditions, but generates the delay of the *a prioiri* SNR because the DD weights the speech spectrum component of the previous frame in the speech signal. Therefore, the noise suppression gain which is affected by the delay of the *a priori* SNR, which is estimated by the DD matches the previous frame rather than the current one, so after noise suppression, this degrades the noise reduction performance during speech transient periods. We propose a computationally simple but effective speech enhancement technique based on the sigmoid type function for the weight parameter of the DD. The proposed approach solves the delay problem about the main parameter, the *a priori* SNR of the DD while maintaining the benefits of the DD. Performances of the proposed enhancement algorithm are evaluated by ITU-T P.862 Perceptual Evaluation of Speech Quality (PESQ), the Mean Opinion Score (MOS) and the speech spectrogram under various noise environments and yields better results compared with the fixed weight parameter of the DD.

Key words: a priori SNR, Decision-Directed, Speech Enhancement, Sigmoid Type

ASK subject classification: Speech Signal Processing (2.3)

책임저자: 장 준 혁 (changjh@inha.ac.kr)
인천시 남구 웅현동 253 인하대학교 전자전기공학부
(전화: 032-860-7423; 팩스: 032-868-3654)

I. 서 론

이동환경에서의 음성통신의 중요성이 점차 증가하면서 단일 마이크로폰에서의 음성 향상 기술에 대한 연구가 주목 받고 있다. 특히, 스펙트럼을 바탕으로 한 음성 향상 기술은 일반적으로 신호 대 잡음비에 따른 파라미터 이득으로 표현 된다 [1]-[3]. 하지만 이득에 의한 음성 향상은 잡음 구간의 잡음 제거 과정에서 원하지 않는 뮤지컬 잡음이 생기게 된다. Ephraim과 Malah가 제안한 decision-directed (DD) 추정 방법은 뮤지컬 잡음을 제거하는데 우수한 성능을 보여 준다 [1][4].

하지만, 최근 Cappé는 DD 추정 방법에 대한 분석을 통해 *a priori* SNR이 잡음 구간에서는 *a posteriori* SNR의 스무딩 (smoothing) 된 형태로 뮤지컬 잡음을 제거하는데 탁월한 성능을 보이지만 음성 구간에서는 모양이 지연되어 따라가는 것을 밝혀냈다 [4]. 즉, MMSE (Minimum Mean Square Error) 잡음 제거 이득은 주로, *a priori* SNR에 좌우되기 때문에 이러한 지연된 파라미터가 적용된 이득은 현재 프레임과 상응되는 값이 아니므로 특히 음성 전이 구간에서 왜곡된 잡음 제거 이득의 적용으로 음성 향상의 성능을 크게 저하 시킨다 [4].

본 논문은 이러한 문제점을 해결하기 위해 고정 가중치 파라미터가 적용된 기존의 DD와는 다르게 *a posteriori* SNR의 변이에 따라 Sigmoid Type 함수의 값이 가중치 파라미터에 적용되도록 하였다. 계산적으로 간단하면서도 실험 결과는 다양한 잡음과 신호 대 잡음비 환경에서 ITU-T P.862 Perceptual Evaluation of Speech Quality (PESQ)와 Mean Opinion Score (MOS), 그리고 음성 스펙트로그램을 테스트하여 기존의 DD보다 향상된 성능을 보였다 [5][6][7].

II. 잡음 제거 이득

잡음에 오염된 음성신호로부터 잡음을 제거하는 음성 향상 과정을 표현하기 위해 $x(t)$와 $d(t)$를 각각 시간 축에서의 음성과 잡음 신호라고 하면, 오염된 음성 신호 $y(t)$는 아래와 같이 표현된다.

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

관련하여, $Y(k)$, $X(k)$, $D(k)$가 각각 $y(t)$, $x(t)$, $d(t)$의 k번째 스펙트럼 성분이라고 한다면 이득 함수에서 파라미터로 작용하는 *a posteriori* SNR 과 *a priori* SNR은 각각 $\gamma(k)$, $\xi(k)$로 아래와 같이 정의 된다 [1].

$$\gamma(k) = \frac{|Y(k)|^2}{E[|D(k)|^2]} \tag{2}$$

$$\xi(k) = \frac{E[|X(k)|^2]}{E[|D(k)|^2]} \tag{3}$$

여기서 $E[\cdot]$는 기대 값 연산자이다. 잡음이 제거된 음성신호의 추정치는 스펙트럼 성분이 통계적으로 독립이라는 가정 하에 아래와 같은 식의 Ephraim과 Malah의 MMSE기법을 도입하면 아래와 같이 구한다.

$$\begin{aligned} |\hat{X}(k)| &= E\{|X(k)| \mid y(t), \quad 0 \le t \le T\} \\ &= E\{|X(k)| \mid Y(0), Y(1), \cdots\} \\ &= E\{|X(k)| \mid Y(k)\}. \end{aligned} \tag{4}$$

위의 MMSE에 기반한 음성신호의 추정치를 실제적으로 구하기 위한 잡음제거 이득은 아래와 같이 오염된 잡음신호와의 곱으로 표현되며,

$$\hat{X}(k) = G(\xi(k), \gamma(k))\, Y(k). \tag{5}$$

여기서, 잡음제거이득 $G(\cdot,\cdot)$는 아래와 같이 (2)와 (3)식의 주요 파라미터가 적용된 형태로 표현되어 진다.

$$\begin{aligned} &G(\xi(k), \gamma(k)) \\ &= \frac{\sqrt{\pi v(k)}}{2\gamma(k)} \exp\left(-\frac{v(k)}{2}\right)\left[(1+v(k))I_0\left(\frac{v(k)}{2}\right) + v(k)I_1\left(\frac{v(k)}{2}\right)\right]. \end{aligned} \tag{6}$$

위에서 $v(k)$는 아래와 같이 주어진다.

$$v(k) = \frac{\xi(k)}{1+\xi(k)}\gamma(k) \tag{7}$$

특히, (6)식에서 I_0, I_1는 각각 0차, 1차 수정 베셀 (modified Bessel) 함수를 의미한다. 보다 구체적으로, 잡음 제거 이득 함수는 *a posteriori* SNR과 *a priori* SNR을 변수로 갖는 함수로 표현되며, 잡음 신호 $D(k)$는 음성 검출기 (VAD, voice activity detector)를 이용하여 음성 부재 구간에서 갱신되는 잡음 신호를 이용 한다 [8][9]. 여기서, *a posteriori* SNR은 (2)식과 같이 들어

온 잡음 음성 $Y(k)$에서 바로 구할 수 있지만 *a priori* SNR은 잡음 음성에서 잡음이 제거된 음성 $X(k)$의 추정치를 이용하여 계산한다. 실제로, 강인한 *a priori* SNR의 추정은 MMSE에 기반 한 음성향상에서 가장 중요한 부분 중의 하나이며, III절에서 자세히 기술한다.

III. "Decision-Directed" 추정치 접근

Ephraim과 Malah는 *a priori* SNR의 추정치를 구하기 위해 다음과 같이 DD 추정치 방법을 제안했다 [1].

$$\hat{\xi}(k,n)=\alpha\frac{|\hat{X}(k,n-1)|^2}{E[|D(k,n-1)|^2]}+(1-\alpha)P[\gamma(k,n)-1],\quad 0\le\alpha<1. \tag{8}$$

$\hat{X}(k,n-1)$는 $(n-1)$번째 프레임에서 k번째 스펙트럼 성분의 크기를 나타내고 α는 가중치 파라미터, $p[x]$는 $p[x]=x$ if $x\ge 0$이고, $p[x]=0$ if $x<0$을 의미하는 연산자이다.

최근 Cappé는 MMSE 잡음 제거 이득이 *a posteriori* SNR 보다 *a priori* SNR에 의해서 좌우 된다는 것을 보였다 [4]. 또한 이전 프레임에 더 큰 가중치를 주는 DD 추정법에 의해서 주요 파라미터로 작용하는 *a priori* SNR이 노이즈 구간에는 *a posteriori* SNR 보다 매우 작은 분산 값을 가지게 된다. 따라서 *a priori* SNR에 영향을 많이 받는 이득 또한 잡음 구간에서 작은 분산 값을 가지게 되므로 *a priori* SNR을 위해 DD 추정 법을 사용하는 MMSE 잡음 제거 이득이 뮤지컬 잡음 제거에 우수한 성능을 보이게 되는 이유를 밝혀냈다. 하지만 잡음에서 음성으로 바뀌는, 즉 *a posteriori* SNR이 급격히 변하는 음성 전이 구간에서는 DD추정법이 이전 프레임에 더 큰 가중치를 주기 때문에 *a priori* SNR이 프레임 지연을 가지고 *a posteriori* SNR의 모양을 따라가게 된다. 음성 구간에서 *a priori* SNR의 지연은 잡음 제거 이득에 영향을 주므로 전이 구간에서의 음성의 왜곡을 발생시킨다. 결과적으로 이전 프레임에 대한 가중치 파라미터 α가 커질수록 잡음 구간에서 뮤지컬 잡음 제거에는 이점이 있으나 음성 구간에서 음성이 왜곡되는 단점이 생긴다.

IV. 제안된 Sigmoid Type의 가중치 파라미터

기존의 DD에 *a priori* SNR 추정의 문제점을 해결하기 위한 다양한 시도가 최근까지 제시되었다. 이것은 Cappé의 연구 결과에 기반을 둔 것으로 실제로는 DD에 의한 추정 방법은 잡음 구간에서 뮤지컬 잡음 제거와 음성 전이 구간에서 음성 왜곡 사이에 trade off가 발생된다는 사실이 보고되었다. 본 논문에서는 프레임 간의 *a posteriori* SNR의 변이에 따라 아래식의 Sigmoid Type 함수의 값을 DD의 가중치 파라미터 α에 적용하는 알고리즘을 제안했다.

$$\hat{\alpha}(k)=\frac{k\exp[-\beta(s(k)-s_0)]}{1+\exp[-\beta(s(k)-s_0)]}+\sigma \tag{9}$$

여기서 $s(k)$는 아래 식에 의해서 주어진다.

$$s(k)=\log(1/|\Delta\gamma(k)|)=\log\{1/(|\gamma(k)-\gamma(k-1)|)\} \tag{10}$$

(9)식에서 각각의 파라미터는 다양한 잡음 환경과 여러 SNR에 대하여 실험적으로 최적화하여, 기울기 파라미터 $\beta=0.4$, 오프셋 (offset) $s_0=-9$, 상수 $k=-1.5$, $\sigma=0.99$로 설정하였다.

$$\hat{\xi}(k,n)=\hat{\alpha}(k,n)\frac{|\hat{X}(k,n-1)|^2}{E[|D(k,n-1)|^2]}+(1-\hat{\alpha}(k,n))P[\gamma(k,n)-1] \tag{11}$$

(9)식의 $\hat{\alpha}(k)$는 함수에 의해서 구해진 새로운 가중치 파라미터로서 (9)식이 (8)식에 적용되어 위의 (11)식과

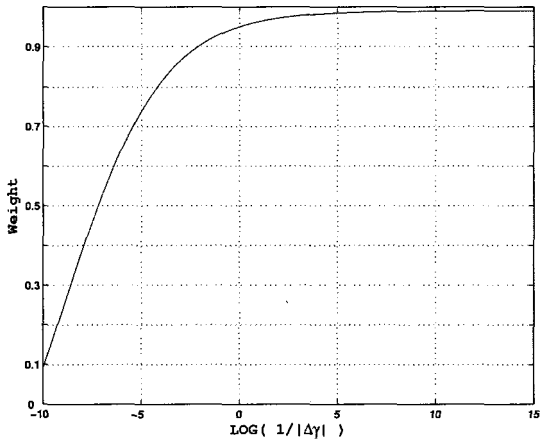

그림 1. 가중치 파라미터 α에 적용되는 Sigmoid Type 함수
Fig 1. The sigmoid type function for the weight parameter α.

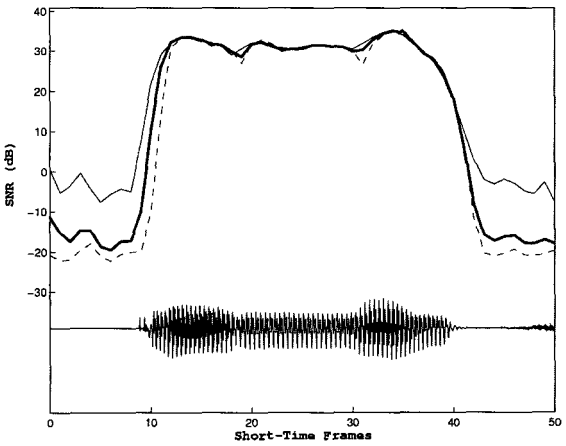

그림 2. 실시간 프레임에서의 SNR: A posteriori SNR (가는선), 기존 DD의 a priori SNR (점선), Sigmoid Type DD의 a priori SNR (굵은선)

Fig 2. SNR in short-time frames. A posteriori SNR (solid line), a priori SNR of the DD (dashed line), a priori SNR of sigmoid type DD (bold line).

같이 $\hat{\alpha}(k,n)$가 적용된 새로운 *a priori* SNR 추정치인 $\hat{\xi}(k,n)$를 유도할 수 있다.

그림 1은 Sigmoid Type 함수가 가중치 파라미터에 적용되는 값을 보여 주고 있다. 잡음 구간에서 음성 구간으로 바뀌는 음성 전이 구간에서 $|\Delta\gamma|$가 상대적으로 급격히 커지는 특성을 이용해 $\log(1/|\Delta\gamma|)$값을 Sigmoid Type 함수의 변수로 적용하여 $|\Delta\gamma|$가 매우 커지는 구간에서는 이전 프레임에 대한 가중치를 작게 갖도록 하였다. 즉 음성 전이 구간에서는 DD의 가중치 파라미터 비중을 현재 프레임의 *a posteriori* SNR에 더 크게 주어 *a priori* SNR의 추정치 값에 *a posteriori* SNR의 값이 더 반영 되도록 하였다. 따라서 음성 전이 구간에서 *a priori* SNR 파라미터의 지연에 의한 잡음 제거 이득의 왜곡을 감소시켰다. 또한, $|\Delta\gamma|$이 적은 잡음 구간에서는 $\log(1/|\Delta\gamma|)$이 커지므로 이전 프레임의 *a priori* SNR에 큰 가중치가 적용되어 뮤지컬 잡음 제거에 우수한 기존 DD의 이점도 유지할 수 있다.

그림 2는 새로운 알고리즘에 따른 *a priori* SNR의 변화를 보여주고 있다. 음성 전이 구간에서 Sigmoid Type DD에 의한 *a priori* SNR은 지연 없이 *a posteriori* SNR과 같은 프레임에서 음성 구간으로 바뀌고 음성 전이 구간에서 기존의 고정 가중치 파라미터가 적용된 DD의 *a priori* SNR 보다 *a posteriori* SNR에 더 가까운 것을 볼 수 있다.

표 1. 다양한 노이즈 환경에서 기존의 DD와 Sigmoid Type DD (Proposed)의 PESQ 수치 비교

Table 1. PESQ scores of the DD and sigmoid type DD (Proposed) under the various noise type.

Noise type	Method	SNR (dB)			
		5	10	15	20
WGN	DD	1.915	2.303	2.694	3.080
	Proposed	1.977	2.364	2.752	3.133
Babble noise	DD	2.136	2.530	2.921	3.279
	Proposed	2.167	2.561	2.952	3.318
Vehicle noise	DD	3.437	3.647	3.766	3.819
	Proposed	3.532	3.746	3.871	3.924

표 2. 다양한 노이즈 환경에서 기존의 DD와 Sigmoid Type DD (Proposed)의 MOS 비교

Table 2. The MOS of the conventional DD and sigmoid type DD (Proposed) under the various noise type.

Noise type	Method	SNR (dB)		
		5	10	15
WGN	DD	1.80	2.49	2.80
	Proposed	1.95	2.51	2.94
Babble noise	DD	1.95	2.76	3.57
	Proposed	2.06	2.88	3.61
Vehicle noise	DD	4.17	4.47	4.62
	Proposed	4.19	4.53	4.66

V. 실험 및 결과고찰

본 논문에서는 기존 DD 알고리즘에서 발생되는 음성 전이 구간에서 *a priori* SNR 파라미터의 지연에 의한 잡음 제거 이득의 왜곡을 감소시킴으로써 개선된 음성 향상을 유도 하였다 [10][11]. 제안된 음성 향상 알고리즘의 음질 평가를 위해 널리 적용되고 있는 ITU-T P.862 PESQ 와 MOS, 그리고 음성 스펙트로그램을 수행하였으며 표 1과 표 2, 그림 3은 각각 추출된 PESQ 수치와 MOS, 그리고 음성 스펙트로그램을 보여주고 있다.

표 1의 PESQ 테스트를 위해 샘플은 남성, 여성화자 각각이 100개의 문장을 발음하도록 한 음성을 한 프레임의 크기가 10ms에서 8kHz로 샘플링 한 데이타에 세가지 형태의 잡음이 부가 되었다. 잡음은 NOISE-X92 데이터베이스의 white gaussian noise (WGN), babble noise, vehicle noise를 사용 하였으며 SNR을 5, 10, 15, 20 dB로 달리하여 조사하였다. PESQ값은 이들 샘플에 대한 평균 수치로 나타냈고, 기존 DD에 의한 PESQ를 위해 가중치 파라미터 $\alpha=0.99$로 설정하여 PESQ 수치를 추출하였다. 표 1은 기존의 DD 방법보다 논문에서 제안한 Sigmoid Type DD방법이 PESQ 수치로 white gaussian, babble, vehicle noise에서 각각 평균 0.06, 0.03, 0.1 정도 향상된 수치를 나타내고 있다.

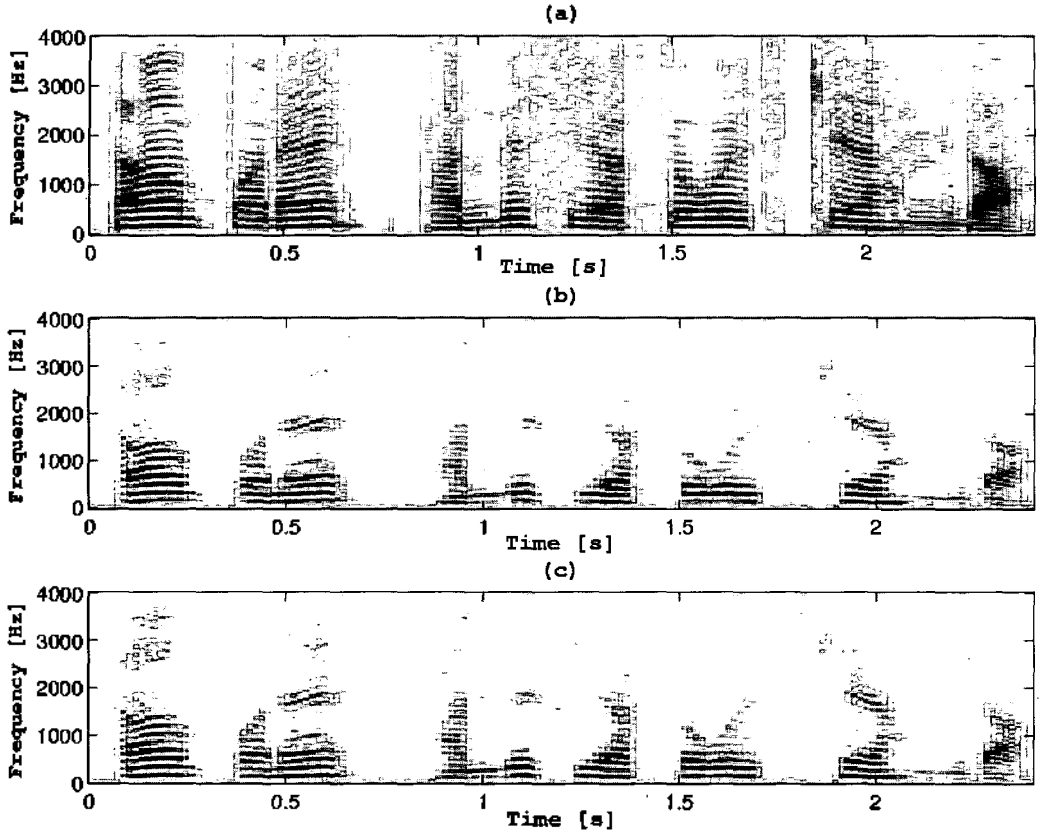

그림 3. 음성 스펙트로그램. (a) 깨끗한 음성 신호. (b) 기존 DD 방법에 의해 잡음이 제거된 음성 신호. (c) Sigmoid Type DD 방법에 의해 잡음이 제거된 음성 신호

Fig 3. Speech spectrograms. (a) Original clean speech. (b)Enhanced speech with the DD algorithm. (c) Enhanced speech with the sigmoid type DD algorithm.

따라서 실험에 사용된 모든 잡음 환경에서 기존의 DD보다 제안된 방법의 PESQ 수치가 향상된 것을 보여주고 있다.

표 2의 MOS는 남성, 여성 화자 각각이 10개의 문장을 발음하도록 한 음성에 white gaussian, babble, vehicle noise가 SNR이 5, 10, 15 dB로 포함된 오염된 음성을 대상으로 10명의 청취자에 의하여 결정하였다. 표 2는 세 가지 잡음 환경과 신호 대 잡음비에 대하여 향상된 MOS를 보여주고 있다.

그림 3의 음성 스펙트로그램의 추출을 위해 128 overlap 샘플을 갖는 256 샘플의 Hanning window와 WGN이 SNR=10 dB로 포함된 잡음 음성을 사용 하였다. 기존 DD에 의해 잡음이 제거된 그림 (b)보다 제안된 DD에 의해 잡음이 제거된 그림 (c)가 음성구간에서의 음성 스펙트럼 성분이 더 선명한 것을 볼 수 있다. 따라서 표 1과 표 2, 그림 3은 기존의 고정 가중치 파라미터를 사용하는 DD 방법보다 본 논문에서 제안한 Sigmoid Type의 가중치 파라미터를 갖는 DD 방법이 음성 향상에 우수한 성능을 가지고 있음을 보여주고 있다.

VI. 결 론

본 논문에서는 Sigmoid Type의 함수를 기존 DD의 고정 가중치 파라미터에 적용하는 새로운 알고리즘을 제안하였다. 기존의 방법은 이전 프레임에 큰 가중치를 주어 잡음 구간에서의 뮤지컬 잡음 제거에는 탁월한 성능을 보였지만 음성 전이 구간에서는 주요 파라미터인 *a priori* SNR의 지연이 생기게 된다. 따라서 지연된 파라미터에 의한 왜곡된 잡음 제거 이득으로 음성 전이 구간에서 음성이 왜곡되는 단점이 있었다.

본 논문에서 제시하는 방법은 Sigmoid Type 함수 값이 a posteriori SNR의 변이에 따라 가변적으로 DD의 가중치 파라미터에 적용되므로 음성 전이 구간에서의 음성 왜곡을 줄이고 잡음 구간에서 뮤지컬 잡음 제거의 이점은 유지 하였다. 따라서 실험에 사용된 모든 잡음과 신호대 잡음 비 환경에서 기존의 DD보다 음성 향상에서 우수한 성능을 보였다.

감사의 글

본 논문은 정통부 및 정보통신연구진흥원의 정보통신선도기반기술개발사업의 연구결과로 수행되었습니다.

참 고 문 헌

1. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-32, 6 1109-1121, Dec. 1984.
2. S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-27, 2 113-120, Apr. 1979.
3. R. J. McAualy and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-28, 137-145, Apr. 1980.
4. O. Cappé, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," IEEE Trans. Speech Audio Process., 2(2) 345-349, Apr.1994.
5. N. Ma, M. Bouchard and R. Goubran, "Perceptual Kalman filtering for speech enhancement in colored noise," in Proc. IEEE Int. Conf. on Acoustic, Speech and Signal Processing, 1 717-720, Montreal, May 2004.
6. C. You, S. N. Koh, and S. Rahardja "Signal subspace speech enhancement for audible noise reduction," in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, I 145-148, Mar. 2005.
7. N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," IEEE Trans. Speech and Audio Processing, 7(2) 126-137, Mar. 1999.
8. N. S. Kim, J.-H. Chang, "Spectral enhancement based on global soft decision," IEEE Signal Processing Letters, 7(5) May 2000, 108-110.
9. J. Sohn, N. S. Kim, W. Sung, "A statistical model-based voice activity detection," IEEE Signal Processing Letters, 6(1) 1-3, Jan. 1999.
10. C. Plapous, C. Marro, P. Scalart, and L. Mauuary, "A two-step noise reduction technique," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Montreal, QC, Canada, May 2004, 1 289-292.
11. I. Cohen, "Speech enhancement using a noncausal a priori SNR estimator," IEEE Signal Processing Letters, 11(9) Sept. 2004, 725-728.

저자 약력

• **박 윤 식 (Yun-Sik Park)**

2006년 2월 : 인하대학교 전자공학과 학사
2006년 3월~현재 : 인하대학교 전자공학과 석사과정

• **장 준 혁 (Joon-Hyuk Chang)**

1998년 2월 : 경북대학교 전자공학과 학사
2000년 2월 : 서울대학교 전기공학부 석사
2004년 2월 : 서울대학교 전기컴퓨터공학부 박사
2000년 3월~2005년 4월 : ㈜넷더스 연구소장
2004년 5월~2005년 4월 : 캘리포니아 주립대학, 산타바바라 (UCSB) 박사후연구원
2005년 5월~2005년 8월 : 한국과학기술연구원 (KIST) 연구원
2005년 9월~현재 : 인하대학교 전자전기공학부 조교수