논문 2010-47SP-5-26

스펙트럼 변이를 이용한 Soft Decision 기반의 음성향상 기법

(Robust Speech Enhancement Based on Soft Decision Employing Spectral Deviation)

최 재 훈*, 장 준 혁**, 김 남 수***

(Jae-Hun Choi, Joon-Hyuk Chang, and Nam-Soo Kim)

요 약

본 논문에서는 비정상적인 배경 잡음 환경에서 음성향상을 위한 신호의 스펙트럼 변이 (Spectral Deviation)을 적용한 Soft Decision 기반의 잡음전력 수정 기법을 제안한다. 기존의 Soft Decision 기반의 잡음전력 추정에 있어서 잡음신호의 정상성 (Stationarity)을 가정한 스무딩 파라미터를 사용하여 잡음전력을 추정하고 갱신하였지만, 잡음신호의 주파수적인 특성이 상대적으로 빠르게 변하는 비정상적인 환경에서는 강인하지 못한 단점을 가지게 된다. 본 논문에서는 신호의 스펙트럼 변이를 추정하여 정상적인 잡음 환경과 비정상적인 잡음 환경에 따라 적응적으로 잡음전력을 추정하고 갱신하여 잡음신호에 의해 오염된 음성신호를 향상시킨다. 제안된 알고리즘은 다양한 배경 잡음 환경에서 객관적인 음질측정 방법인 ITU-T P.862 perceptual evaluation of speech quality (PESQ)에 의해서 평가되었으며, 기존의 Soft Decision 기반의 음성 향상 기법과 비교하여 보다 향상된 성능을 보여주었다

Abstract

In this paper, we propose a new approach to noise estimation incorporating spectral deviation with soft decision scheme to enhance the intelligibility of the degraded speech signal in non-stationary noisy environments. Since the conventional noise estimation technique based on soft decision scheme estimates and updates the noise power spectrum using a fixed smoothing parameter which was assumed in stationary noisy environments, it is difficult to obtain the robust estimates of noise power spectrum in non-stationary noisy environments that spectral characteristics of noise signal such as restaurant constantly change. In this paper, once we first classify the stationary noise and non-stationary noise environments based on the analysis of spectral deviation of noise signal, we adaptively estimate and update the noise power spectrum according to the classified noise types. The performances of the proposed algorithm are evaluated by ITU-T P. 862 perceptual evaluation of speech quality (PESQ) under various ambient noise environments and show better performances compared with the conventional method.

Keywords: Spectral deviation, Non-stationary noise signal, Soft decision, Speech enhancement

* 학생회원, ** 정회원, 인하대학교 전자공학부
(Dep. of Electronics Engineering, Inha University)
*** 정회원, 서울대학교 전자컴퓨터공학부
(School of Electrical Engineering and Computer Science, Seoul National University)

※ 이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의지원을 받아 수행된 기초연구사업임 (313-2008-2-D00783). 또한 본 연구는 지식경제부 및 한국산업기술평가관리원의 IT핵심기술개발사업의 일환으로 수행하였으며 [2009-S-036- 01, Development of New Virtual Machine Specification and Technology] 그리고 이 논문은 2009년 정부(교육과학기술부)의 재원으로 한국 연구재단의 지원을 받아 수행된 연구(2009-0085162).
접수일자: 2010년3월11일, 수정완료일: 2010년8월2일

Ⅰ. 서 론

언제 어디서나 상대방과 의사소통이 가능하게 하는 이동통신 기술의 발달과 함께 이동통신 단말이 폭넓게 보급되어 현재는 1인 1휴대폰 시대로 발전하게 되었다. 이동하는 환경에서 상대방과 이동통신 단말로 의사소통하게 되는 경우, 주위 잡음에 노출되는 빈도가 커짐에 따라 잡음을 제거하는 음성향상 기술에 대한 연구가 큰 주목을 받고 있다. 음성향상 기술은 크게 잡음신호를 추정하는 부분과 음성신호의 스펙트럼 이득을 추정하는 부분으로 나눌 수 있으며, 잡음신호를 정확하게 추정하는 것은 시스템의 성능에 직접적인 영향을 미치게 되므로 중요한 부분이라 할 수 있다.

음성향상 시스템에서 잡음신호의 정확한 추정에 따라 음성 품질에 큰 영향을 미치게 되는데, 잡음신호를 작게 추정하게 되는 경우 잔류 잡음으로 인해 부자연스러운 음성을 듣게 되고, 반대로 잡음신호를 크게 추정하는 경우 음성신호의 손실이 발생하여 음성의 명료도가 떨어지게 된다. 기존의 잡음신호의 추정 방법에는 음성 검출기 (Voice Activity Detector: VAD)를 이용하여 음성이 존재하지 않는 구간에서 잡음신호를 평균화하여 구하는 방법이 있다. VAD를 이용한 잡음추정 방법은 신호 대 잡음비 (Signal-to-Noise Ratio: SNR)가 낮은 환경이나 비정상적인 잡음환경에서 정확한 잡음 추정이 어려운 단점들이 존재 한다[1~3].

최근 잡음신호의 추정에서 가장 주목 받고 있는 방법들은 최소 통계 잡음 추정 (Minimum statistics), Soft decision에 기반한 잡음전력 추정법 및 최소값 제어 재귀평균 (Minima Controlled Recursive Averaging: MCRA)으로써, 잡음신호의 추정에 우수한 성능을 보인다고 알려져 있다[4~7]. 최소 통계 잡음 추정 기법은 오염된 음성 구간의 휴지 구간에서 음성신호의 전력 레벨이 잡음신호의 전력 레벨까지 감소한다는 사실에 착안되었다. 충분히 큰 윈도우를 사용하여 최적화된 전력 스펙트럼 스무딩을 거친 최소 잡음 전력을 추정한 후에 바이어스 (bias) 보상을 통하여 잡음신호의 전력을 추정하게 된다[4, 9]. Soft decision 기반의 잡음전력 추정 방법은 음성부재확률 (Speech Absence Probability: SAP)에 기반하여 프레임 사이의 스무딩을 통해 이전 프레임의 추정신호를 갱신함으로써 현재 프레임의 잡음전력을 추정하는 방법이다[5]. 또한 최소 통계 잡음 추정 기법과 Soft Decision 기반의 잡음전력 추정 방법의 장점을 주파수 채널별 SAP를 가중치로 사용하여 음성 구간 보다 효율적인 잡음추정 방법으로 알려진 최소 통계 잡음 추정 기법으로 추정된 잡음전력에 더 큰 가중치를 두고, 반면에 비음성 구간에서 보다 견실한 잡음 추정이 가능한 Soft Decision 기법으로 추정된 잡음전력에 더 큰 가중치를 적용한 방법 또한 우수한 성능을 보임이 알려졌다[6]. 마지막으로 최소값 제어 재귀평균 기법은 각 서브밴드에서 신호의 존재 확률로 조절하는 스무딩 매개변수를 이용하여 스펙트럼에 평균을 취하는 방법으로써, 각 서브밴드에서 신호의 존재는 잡음이 섞인 신호의 국부 에너지와 윈도우에서의 최소값 사이의 비로 결정 되고, 이 비율과 특정 임계값과 비교하여 음성신호의 존재 유무를 결정하는 방법이다. 또한 음성신호가 존재하는 부분과 음성신호가 존재하지 않는 부분 사이에서 발생되는 변동을 줄이기 위해 시간 축으로도 평균을 취한다[7].

위에서 언급된 대표적인 잡음신호의 추정 방법들은 잡음신호의 추정에 있어서 비교적 견실하다는 장점을 가지고 있지만, 보완해야 할 부분이 존재하며 성능 향상을 위한 연구 또한 활발히 진행되고 있다[9~10]. 특히 잡음 구간에서 우수한 성능을 보인다고 알려진 Soft decision 기반의 잡음전력 추정 방법의 경우 스무딩 파라미터에 의한 갱신으로 잡음전력을 추정하게 되는데, 실제 사용되는 스무딩 파라미터의 경우, 잡음의 정상 (Stationarity) 가정을 고려한 $\xi_d = 0.99$을 사용한다[11]. 그러나 잡음신호의 주파수적 특성이 상대적으로 빠르게 변하는 비정상적 잡음 환경에서 성능 저하가 불가피하며, 비정상적 잡음 환경을 고려하여 현재 프레임의 잡음전력에 보다 큰 가중치를 적용하여 빠른 잡음전력의 갱신을 통하여 성능을 향상시킬 수 있는 여지가 존재한다.

본 논문에서는 비정상적 잡음환경에서 잡음신호의 스펙트럼 변이 (Spectral Deviation)를 이용한 Soft decision 기반의 잡음전력 추정 수정 기법을 제안한다. 구체적으로, 먼저 다양한 잡음신호마다 고유한 스펙트럼 변이를 분석하고, 분석된 잡음신호별 스펙트럼 변이 값에 따라 잡음전력 추정에 적용되는 스무딩 파라미터를 정상적 잡음 환경에서는 이전 프레임의 잡음전력에 더 큰 가중치가 적용되도록 하고, 비정상적 잡음 환경에서는 현재 프레임의 추정된 잡음전력에 보다 큰 가중치를 적용함으로써 잡음신호의 종류에 따라 잡음전력 추정에 사용되는 스무딩 파라미터가 적응적으로 변하는

알고리즘을 제안한다. 제안된 알고리즘의 성능은 다양한 잡음환경에서 객관적인 음질평가 방법인 ITU-T P.862 Perceptual Evaluation of Speech Quality (PESQ)를 이용하여 평가되었으며, 기존의 제안된 Soft Decision 방법보다 향상된 결과를 나타내었다[12].

본 논문의 구성은 다음과 같다. 먼저 Ⅱ장에서는 기존의 Soft decision 기반 잡음전력 추정 기법에 대하여 서술하고, Ⅲ장에서는 제안된 잡음신호의 스펙트럼 변이를 이용한 Soft Decision 기반의 잡음전력 수정 기법에 대해서 기술하였다. Ⅳ장에서는 실험 결과 비교 및 분석을 기술하였으며, 마지막 Ⅴ장에서는 결론을 맺는다.

Ⅱ. 본 론

1. Soft Decision 기반 잡음전력 추정

주파수축 기반의 음성향상에 있어서 잡음전력의 추정은 성능에 직접적인 영향을 미치며, 음성향상 및 음성 부호화기등의 성능 향상에 중요한 요소로 작용한다. 가장 널리 사용되는 방법은 음성검출기 (Voice Activity Detector: VAX)를 사용하여 음성이 없는 구간에서만 잡음전력을 갱신한다. 그러나 실제 잡음전력은 음성부재 구간뿐만 아니라 음성이 존재하는 구간에서도 변화하게 되며, 음성구간에서도 잡음전력이 갱신되어야 한다[11]. 본 장에서는 대표적인 잡음전력 추정 방법으로 Soft decision에 기반한 잡음전력 추정 법에 대하여 설명한다.

시간 축에서 음성신호 $x(n)$에 잡음신호 $d(n)$이 인가되어 오염된 음성신호 $y(n)$을 만들게 된다면, 각각의 성분을 DFT(discrete Fourier transform)을 통해서 주파수 축으로 다음과 같이 나타낼 수 있다.

$$Y(t,k) = X(t,k) + D(t,k) \tag{1}$$

여기서 $Y(t,k)$, $X(t,k)$ 그리고 $D(t,k)$는 각각 t번째 프레임에 대한 k번째 주파수 성분을 의미한다. 또한, 잡음신호 $D(t,k)$는 음성신호인 $X(t,k)$와 통계적으로 독립이라고 가정한다. 음성의 통계 모델에 기반한 Soft Decision 추정을 위해 음성 부재와 존재에 대한 가설을 각각 H_0와 H_1이라 한다면, 주파수 채널에 따라 다음과 같이 가정할 수 있다[11].

$$H_0 : speech\ absence :\ Y(t,k) = D(t,k)$$

$$H_1 : speech\ present :\ Y(t,k) = X(t,k) + D(t,k) \tag{2}$$

음성신호 $X(t,k)$와 잡음신호 $D(t,k)$가 통계적으로 독립이라는 가정과 음성신호와 잡음신호의 스펙트럼이 zero-mean 복소 가우시안 분포를 보인다고 가정하면, 제시된 가설 H_0와 H_1에 따라 다음과 같은 확률밀도함수로 표현할 수 있다[11].

$$\begin{aligned} p(Y(t,k)|H_0) &= \frac{1}{\pi\lambda_d(t,k)} exp\left\{ -\frac{|Y(t,k)|^2}{\lambda_d(t,k)} \right\} \\ p(Y(t,k)|H_1) &= \frac{1}{\pi(\lambda_x(t,k)+\lambda_d(t,k))} \\ &\quad \cdot \exp\left\{ -\frac{|Y(t,k)|^2}{\lambda_x(t,k)+\lambda_d(t,k)} \right\} \end{aligned} \tag{3}$$

식 (3)에서 $\lambda_x(t,k)$와 $\lambda_d(t,k)$는 각각 t번째 프레임에 대한 k번째 주파수 성분에서의 음성과 잡음의 분산을 의미한다. 음성의 존재와 부재에 관한 위의 가설로부터 주파수 채널별 음성부재확률 (Speech Absence Probability: SAP)을 구하면, Bayes' rule에 의하여 아래와 같이 나타낼 수 있다.

$$\begin{aligned} p(H_0|Y(t,k)) &= \frac{p(Y(t,k)|H_0)p(H_0)}{p(Y(t,k))} \\ &= \frac{p(Y(t,k)|H_0)p(H_0)}{p(Y(t,k)|H_0)p(H_0)+p(Y(t,k)|H_1)p(H_1)} \\ &= \frac{1}{1+\frac{p(H_1)}{p(H_0)}\Lambda(Y(t,k))} \end{aligned} \tag{4}$$

여기서 $p(H_0)$는 음성 부재에 대한 *a priori* 확률이고, $\Lambda(Y(t,k))$는 k번째 주파수 대역의 우도비 (likelihood ratio)로써 다음과 표현된다.

$$\begin{aligned} \Lambda(Y(t,k)) &= \frac{p(Y(t,k)|H_1)}{p(Y(t,k)|H_0)} \\ &= \frac{1}{1+\xi(t,k)} exp\left[\frac{\gamma(t,k)\xi(t,k)}{1+\xi(t,k)} \right] \end{aligned} \tag{5}$$

식 (5)에서 $\gamma(t,k)$와 $\xi(t,k)$는 각각 *a posteriori SNR*과 *a priori SNR*로 아래와 같이 정의된다.

$$\gamma(t,k) \equiv \frac{|Y(t,k)|^2}{\lambda_d(t,k)} \tag{6}$$

$$\xi(t,k) \equiv \frac{\lambda_x(t,k)}{\lambda_d(t,k)} \tag{7}$$

Soft decision 기반의 잡음전력 추정은 스무딩 파라미터에 의한 갱신으로 잡음전력을 추정하게 되며, Long-term 스무딩된 전력스펙트럼 추정치 $\hat{\lambda}_d(t,k)$는 다음과 같이 정의된다[11].

$$\hat{\lambda}_d(t+1,k)=\zeta_d\hat{\lambda}_d(t,k)+(1-\zeta_d)E\left[|D(t,k)|^2|Y(t,k)\right] \quad (8)$$

여기서 ξ_d는 정상(stationary) 가정을 고려한 스무딩 파라미터로 $0<\xi_d<1$의 범위를 갖으며, 일반적으로 $\xi_d=0.99$로 설정된다. 음성의 존재와 부재를 고려하여 현재 프레임에서의 잡음전력의 추정치 $E\left[|D(t,k)|^2|Y(t,k)\right]$에 주파수 채널별 SAP를 적용하여 나타내면 다음과 같이 표현된다.

$$\begin{aligned}E\left[|D(t,k)|^2|Y(t,k)\right]&=E\left[|D(t,k)|^2|Y(t,k),H_0\right]p(H_0|Y(t,k))\\&+E\left[|D(t,k)|^2|Y(t,k),H_1\right]p(H_1|Y(t,k))\end{aligned} \quad (9)$$

여기서

$$\begin{aligned}E\left[|D(t,k)|^2|Y(t,k),H_0\right]&=|Y(t,k)|^2\\E\left[|D(t,k)|^2|Y(t,k),H_1\right]&=\left(\frac{\xi(t,k)}{1+\xi(t,k)}\right)\hat{\lambda}_d(t,k)\\&+\left(\frac{1}{1+\xi(t,k)}\right)^2|Y(t,k)|^2\end{aligned} \quad (10)$$

2. 스펙트럼 변이 (Spectral Deviation)를 이용한 Soft Decision 잡음 전력 추정 수정 기법

기존의 Soft decision 방법은 음성부재확률에 의하여 잡음구간에서 Long-term 스무딩을 적용함으로써 보다 정확한 잡음전력을 추정한다고 알려져 있다[11]. 특히 Long-term 스무딩을 적용한 식에서 보듯이, 스무딩 파라미터 ζ_d의 값이 1에 가까운 가중치 값 (실제, $\zeta_d=0.99$)을 통하여 잡음전력을 추정하게 된다. 가중치 스무딩 파라미터 $\zeta_d=0.99$는 잡음신호의 정상성을 가정한 값으로써, 정상적 잡음환경에서는 견실한 추정이 가능하지만, 잡음신호의 주파수적 특성이 정상상태의 잡음신호와 비교해서 상대적으로 빠르게 변하는 비정상적 잡음 환경에서는 성능 저하가 불가피하다. 구체적으로 정상상태 잡음신호의 경우 주파수적 특성이 상대적으로 천천히 변하기 때문에 이전 프레임에 보다 큰 가중치 스무딩 파라미터를 적용함으로써 잡음신호의 전력을 추정하게 되고, 상대적으로 천천히 잡음신호의 전력을 갱신하는 것이 필요하다. 반대로, 주파수적 특성이

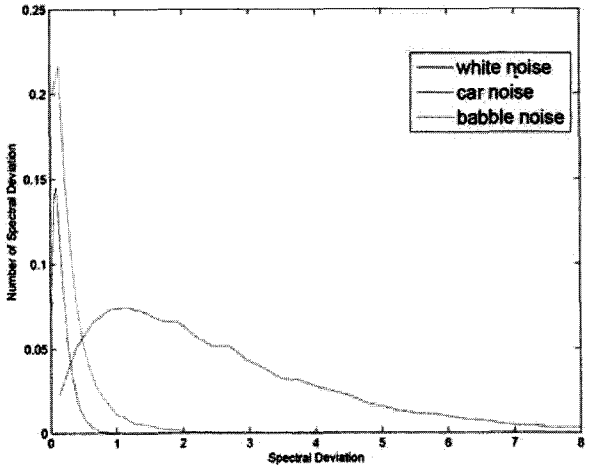

그림 1. 잡음신호별 스펙트럼 변이의 분포도
Fig. 1. Normalized distributions of spectral deviation of the various noise signal.

상대적으로 빠르게 변하는 비정상적인 잡음 환경에서는 잡음신호의 추정뿐만 아니라, 잡음신호 전력의 갱신이 보다 빠르게 이루어지도록 하는 것이 필요하다.

따라서 본 논문에서는 정상적인 잡음 환경과 비정상적인 잡음 환경에 따라 가중치 스무딩 파라미터를 각각의 잡음환경에서 최적화되어 적응적으로 적용함으로써 잡음신호의 전력을 추정하고 갱신하는 Soft decision 기반의 잡음전력 추정 수정 알고리즘을 제안한다. 정상적인 잡음신호와 비정상적인 잡음신호를 분류하기 위해 본 논문에서는 잡음신호의 주파수적인 특성중의 하나인 스펙트럼 변이 (Spectral Deviation)를 도입한다[13]. 먼저 t번째 프레임에서의 k번째 주파수 성분에 대한 분산은 다음과 같이 나타낼 수 있다.

$$\Delta'(t,k)=\frac{\Delta(t,k)}{|P(t,k)|} \quad (11)$$

여기서 $P(t,k)$와 $P(t-1,k)$는 각각 현재 t번째 프레임의 k번째 주파수 성분의 전력과 이전 $t-1$번째 프레임의 k번째의 주파수 성분의 전력을 의미하며, $\Delta(t,k)$는 현재 프레임의 전력과 이전 프레임의 전력의 차로써 다음 식에 의해 나타낼 수 있다.

$$\Delta(t,k)=|P(t,k)-P(t-1,k)| \quad (12)$$

따라서 t번째 프레임의 k번째 주파수 성분에 대한 스펙트럼 변이는 다음과 같이 나타낼 수 있다.

$$\hat{\Delta}(t,k)=\frac{\Delta'(t,k)}{E(P(t))} \quad (13)$$

표 1. 잡음신호별 스펙트럼 변이 값에 따른 잡음추정 및 갱신에 적용되는 스무딩 파라미터 값

Table 1. The smoothing parameter applied to the estimation and update of the noise power spectrum according to spectral deviation under various noise signal.

	Noise Type	$\hat{\Delta}(t,k)$ 값	스무딩 파라미터 값
$\zeta_d^{SD}(t,k)$	white noise	0.017~0.7	0.99
	car noise	0.7~1.2	0.98
	babble noise	1.2~10.0	0.97

여기서 $E(P(t))$는 t번째 프레임 전체의 잡음신호에 대한 전력의 기대 값을 의미한다. 그림 1에는 식에 의해서 구해진 white, car, babble 잡음 신호에 대한 스펙트럼 변이의 분포도를 나타내었다. 그림 1에서 보듯이 white 잡음 신호의 경우 0.017에서 1.699 사이에 스펙트럼 변이 값이 주로 분포하며, car 잡음의 경우 0.0443에서 4.350, babble 잡음의 경우는 0.130에서 12.715 사이로 스펙트럼 변이 값이 가장 넓게 분포됨을 확인할 수 있다. 따라서 본 논문에서는 표 1에서와 같이 세 가지 잡음신호별 스펙트럼 변이 값에 따라 잡음신호별 가중치 스무딩 파라미터의 값을 제안한다.

기존의 제안된 Long-term 스무딩된 전력스펙트럼 추정치 $\hat{\lambda}_d(t,k)$에 새롭게 제안된 잡음신호별 가중치 스무딩 파라미터 $\zeta_d^{SD}(t,k)$와 결합하면 다음과 같이 정의할 수 있다.

$$\hat{\lambda}_d^{SD}(t+1,k) = \zeta_{SD}(t,k)\hat{\lambda}_d{}^{SD}(t,k) + (1-\zeta_{SD}(t,k))E[|D(t,k)|^2|Y(t,k)] \quad (14)$$

식 (14)에서 새롭게 정의된 잡음전력 $\hat{\lambda}_d^{SD}$의 적용으로써 본 논문에서는 musical 잡음의 제거에 탁월한 성능을 가진 것으로 알려져 있는 Ephraim-Malah noise suppression (EMSR)을 이득함수로 선택한다. 일반적으로 오염된 음성신호에 잡음제거이득을 곱함으로써 잡음이 제거된 음성신호를 추정하게 되는데, 다음과 같이 나타낼 수 있다[1].

$$\hat{X}(t,k) = G(\xi(t,k),\gamma(t,k))\,Y(t,k) \quad (15)$$

MMSE (Minimum Mean Square Error)에 기반한 잡음제거이득 $G(\xi(t,k),\gamma(t,k))$는 다음과 같이 주어진다.

$$G(\xi(t,k),\gamma(t,k)) = \frac{\sqrt{\pi v(t,k)}}{2\gamma(t,k)} exp\left(-\frac{v(t,k)}{2}\right) \cdot \left[(1+v(t,k)I_0\left(\frac{v(t,k)}{2}\right)+v(t,k)I_1\left(\frac{v(t,k)}{2}\right)\right] \quad (16)$$

여기서 $v(t,k)$는 다음 식에 의해 주어진다.

$$v(t,k) = \frac{\xi(t,k)}{1+\xi(t,k)}\gamma(t,k) \quad (17)$$

식 (17)에서 보듯이 잡음제거이득의 주요 파라미터인 *a priori SNR*은 깨끗한 음성신호의 전력으로부터 구해지기 때문에, 다음과 식과 같이 깨끗한 음성신호의 전력을 추정하게 된다.

$$\hat{\lambda}_x(t+1,k) = \zeta_x\hat{\lambda}_x(t,k) + (1-\zeta_x)E[|X(t,k)|^2|Y(t,k)|] \quad (18)$$

식 (18)에서 ζ_x는 $0<\zeta_x<1$의 범위를 갖는 스무딩 파라미터이고, 음성부재를 고려한 $E[|X(t,k)|^2|Y(t,k)|]$는 주파수 채널별 음성부재확률을 적용하면 다음과 같이 표현된다.

$$E[|X(t,k)|^2|Y(t,k)|] = E[|X(t,k)|^2 Y(t,k), H_0]p(H_0|Y(t,k)) + E[|X(t,k)|^2 Y(t,k), H_1]p(H_1|Y(t,k)) \quad (19)$$

잡음제거이득의 주요 파라미터는 *a priori SNR* $\xi(t,k)$과 *a posteriori SNR* $\gamma(t,k)$이며, 새롭게 구해진 잡음전

표 2. 다양한 배경 잡음 환경에서 제안된 알고리즘 대비 기존의 Soft Decision 기반의 잡음전력 추정 기법의 PESQ 비교

Table 2. The PESQ results for the proposed algorithm with respect to the conventional method based on Soft Decision under various background noise environments.

Noise type	SNR (dB)	Method	
		Conventional algorithm	Proposed algorithm
babble noise	5 dB	2.344	2.387
	10 dB	2.668	2.700
	15 dB	2.958	2.973
car noise	5 dB	3.315	3.320
	10 dB	3.600	3.603
	15 dB	3.837	3.832
white noise	5 dB	2.088	2.093
	10 dB	2.432	2.435
	15 dB	2.756	2.758

력 $\hat{\lambda}_d^{SD}$을 적용하면 각각 다음과 같이 나타낼 수 있다.

$$\hat{\xi}(t,k) = \frac{\hat{\lambda}_x(t,k)}{\hat{\lambda}_d^{SD}} \tag{20}$$

$$\hat{\gamma}(t,k) = \frac{|Y(t,k)|^2}{\hat{\lambda}_d^{SD}} \tag{21}$$

Ⅲ. 실 험 분석 및 비교

본 논문에서는 제안된 스펙트럼 변이를 이용한 Soft decision 기반의 잡음전력 수정 기법의 성능을 평가하기 위해서, 다양한 배경 잡음 환경에서 ITU-T P.862의 객관적인 음질 측정 방법인 perceptual evaluation of speech quality (PESQ)로 실험을 진행하였다[12]. 먼저, NTT 한국어 음성 데이터베이스에서 20세에서 35세 사이의 남녀 각각 4명씩의 음성 샘플을 추출하였으며, 음성 샘플의 총 길이는 8초이고, 8 kHz로 샘플링 하였다. 원단의 다양한 배경 잡음을 위해서, NOISEX-92 데이터베이스로부터 추출된 white, babble, 그리고 vehicle 잡음이 SNR 5, 10, 15 dB로 섞이도록 하였다.

실험에서 사용된 음성 데이터에는 먼저 제안된 기법에 의한 잡음전력의 추정치가 MMSE 기반의 잡음제거 알고리즘에 적용되어, 잡음이 제거된 음성 데이터와 잡음이 섞이지 않은 깨끗한 음성 데이터와의 PESQ를 수행하였다. 또한 기존의 정상적인 잡음환경을 가정한 Soft decision 기반의 기법과의 비교를 위해, Soft decision 기반의 기법을 MMSE 기반의 잡음제거 알고리즘에 적용하였으며, 잡음전력의 추정 및 갱신을 위한 스무딩 파라미터의 경우 정상상태를 가정한 $\zeta_d = 0.99$를 사용하여 실험을 진행하였다.

표 2에는 다양한 잡음 환경과 SNR에 대해서 제안된 스펙트럼 변이를 이용한 Soft decision 기반의 잡음전력 추정 수정 기법과 기존의 제안된 기법과의 PESQ 결과를 보여준다. 표 2의 결과에 의하면 babble 잡음, car 잡음, white 잡음에서 평균적으로 향상된 수치를 확인할 수 있었다. 실험 결과를 통하여 제안된 스펙트럼 변이를 이용한 Soft decision 기반의 잡음전력 추정 수정 기법은 기존의 정상상태를 가정한 Soft decision 기법과 비교하여 특히 가장 대표적인 비정상적인 잡음신호인 babble noise에서 보다 향상된 결과를 보여줌을 확인할 수 있었다.

Ⅳ. 결 론

본 논문에서는 비정상적인 잡음 환경에서, 잡음신호의 고유한 주파수적인 특성 중의 하나인 스펙트럼 변이를 이용한 Soft decision 기반의 잡음전력 추정 수정 기법을 제시하였다. 먼저 잡음신호별 스펙트럼 변이를 구한 후에, 잡음신호에 따른 스펙트럼 변이에 따라 각각의 잡음신호에 최적화되어 적응적으로 Long-term 스무딩 파라미터가 적용되도록 기존의 Soft decision 잡음전력 추정 기법에 결합하였다.

객관적인 음질 측정 방법인 PESQ를 통하여 기존의 정상적인 잡음 환경을 가정한 잡음전력 추정 기법과 비교하여 제안된 알고리즘은 babble 잡음에 대해서는 0.030, car 잡음에 대해서는 0.001, white 잡음에 대해서는 0.003 향상된 결과를 보여주었다.

감 사 의 글

이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (313-2008-2-D00783). 또한 본 연구는 지식경제부 및 한국산업기술평가관리원의 IT핵심기술개발사업의 일환으로 수행하였으며[2009-S-036- 01, Development of New Virtual Machine Specification and Technology] 그리고 이 논문은 2009년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구 (2009-0085162).

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. Acous., Speech, Signal Process.*, vol. ASSP-32, no. 6, pp. 1109-1121, Dec. 1984.

[2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, 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.*, ASSP-28, 137-145, Apr. 1980.

[4] R. Martin, "Spectral subtraction based on minimum statistics," in *Proc. 7th EUSIPCO'94, Edinburgh, U.K.*, pp. 1182-1185, Sept. 1994.
[5] J. Sohn, W. Sung, "A voice activity detector employing soft decision based noise spectrum adaptation," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing.*, pp. 365-368, 1998.
[6] Y. -S. Park, J. -H. Chang, "A probabilistic combination method of minimum statistics and soft decision for robust noise power estimation in speech enhancement," *IEEE Signal Processing Letters*, vol. 15, pp. 95-98, Jan. 2008.
[7] I. Cohen, B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Processing Letters*, vol. 9, no. 1, pp. 12-15, Jan. 2002.
[9] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Trans. On Speech and Audio Processing.*, 9 (5) pp. 504-512, July 2001.
[10] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 466 - 475, Sep. 2003.
[11] N. S. Kim and J. H. Chang, 'Spectral enhancement based on global soft decision," *IEEE Signal Processing Letters*, vol. 7, no. 5, pp. 108-110, May 2000.
[12] Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs 2001, ITU-T P.862.
[13] TIA/EIA/IS-127, "Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems," 1996.

저 자 소 개

최 재 훈(학생회원)
2007년 인하대학교 전자전기공학부 학사.
2008년 삼성전자 정보통신 총괄 연구원.
2009년 인하대학교 전자공학과 석사.
2010년 인하대학교 전자공학과 박사과정.
<주관심분야 : 디지털 음성신호처리>

장 준 혁(정회원)
1998년 경북대하교 전자공학과 학사.
2000년 서울대학교 전기공학부 석사.
2004년 서울대학교 전기컴퓨터 공학부 박사.
2000년~2005년 (주)넷더스 연구소장
2004년~2005년 캘리포니아 주립대학, 산타바바라(UCSB) 박사후연구원
2005년 한국과학기술연구원(KIST) 연구원
2005년~현재 인하대학교 전자공학부 조교수
<주관심분야 : 음성신호처리, 오디오 신호처리, 통신 신호처리, 휴먼/컴퓨터 인터페이스>

김 남 수(정회원)
1988년 서울대학교 전자공학과 학사
1990년 한국과학기술원 전기 및 전자공학과 석사
1994년 한국과학기술원 전기 및 전자공학과 박사
1994년~1998년 삼성종합기술원 전문연구원
1998년~현재 서울대학교 전기공학부 교수