

논문 2007-44SP-6-13

# 복소 라플라시안 확률 밀도 함수에 기반한 음성 향상 기법

## (Noisy Speech Enhancement Based on Complex Laplacian Probability Density Function)

박 윤 식\*, 조 규 행\*, 장 준 혁\*\*

(Yun-Sik Park, Q-Haing Jo, and Joon-Hyuk Chang)

### 요 약

본 논문에서는 복소 라플라시안 확률밀도함수 (PDF, Probability Density Function)에 기반한 새로운 음성 향상 기법을 제시한다. 적용된 복소 라플라시안 PDF가 기존의 가우시안 PDF보다 오염된 음성 분포를 정확하게 표현한다는 것을 Goodness-of-Fit (GOF) 테스트로 확인하였고, 음성 향상 알고리즘의 음성부재확률을 위해 우도비 (LR, Likelihood Ratio)를 적용하였다. 제시된 알고리즘의 성능은 객관적 테스트에 의해 평가하였고 기존의 가우시안 PDF보다 개선된 음성 향상 결과를 나타내었다.

### Abstract

This paper presents a novel approach to speech enhancement based on a complex Laplacian probability density function (pdf). With a use of goodness-of-fit (GOF) test, we show that the complex Laplacian pdf is more suitable to describe the conventional Gaussian pdf. The likelihood ratio (LR) is applied to derive the speech absence probability in the speech enhancement algorithm. The performance of the proposed algorithm is evaluated by the objective test and yields better results compared with the conventional Gaussian pdf-based scheme.

**Keywords :** Complex Laplacian, Speech enhancement, PDF, LR

### I. 서 론

음성 향상 기법은 가변 전송률의 음성 부호화나 자동 음성 인식에서 주요한 요소이다<sup>[1~14]</sup>. 최근에는 soft decision이 적용된 잡음제거이득이 초기의 VAD (Voice Activity Detector)에 의해 각 프레임을 음성/비음성 구분하는 hard decision보다 향상된 성능을 보이면서 음성 향상 기법에 soft decision을 적용하여 음성 향상 성능을 개선하려는 많은 시도가 이루어지고 있다<sup>[2, 11]</sup>. 또

한 이산 푸리에 변환 (DFT, Discrete Fourier Transform)은 오염된 음성 향상을 위해 가장 널리 사용되고 있는 방법으로써 깨끗한 음성과 잡음 스펙트럼이 복소 가우시안 (Complex Gaussian) 확률밀도함수 (PDF, Probability Density Function)를 따른다고 가정하고 있다. 하지만 최근, 깨끗한 음성과 잡음의 DFT 계수가 가우시안 PDF 보다 감마 (Gamma) 혹은 라플라시안 (Laplacian) PDF에 의해 보다 정확히 표현된다고 확인되었으며, 복소 라플라시안 PDF는 다양한 잡음 환경에서 VAD를 위해 오염된 음성의 DFT 계수의 PDF로 사용되었다<sup>[3]</sup>.

본 논문에서는 음성 향상을 위해 DFT 영역에서 복소 라플라시안 PDF를 제시하였다. 다양한 잡음이 부가된 음성신호에 GOF (Goodness-of-Fit) 테스트를 실시하여 기존의 PDF와 라플라시안 PDF를 비교하였으며, 이러한 테스트를 통하여 실험적인 오염 음성신호의 분

\* 학생회원, \*\* 정회원, 인하대학교 전자전기공학부  
(School of Electronic and Electrical Engineering, Inha University)

※ 본 연구는 정보통신부 및 정보통신연구진흥원의 IT 신성장동력핵심기술개발사업의 일환으로 수행하였음. [2005-S096-02, 신체장애인을 위한 착용형 단말 인터페이스 기술]

접수일자: 2007년4월26일, 수정완료일: 2007년10월24일

포가 가우시안 PDF보다 라플라시안 PDF에 더 가까운 것을 확인하였다. 따라서 본 논문은 이러한 통계적 분석을 기반으로, 복소 라플라시안 PDF를 전역음성부재 확률 (GSAP, Global Speech Absence Probability)을 구하기 위해 적용하였으며, 이러한 GSAP가 잡음제거이득의 수정과 잡음 스펙트럼 추정에 이용되는 새로운 방법을 제안하였다. 제안된 알고리즘의 성능은 ITU-T P.862 Perceptual Evaluation of Speech Quality (PESQ)에 의해 평가했고 [1]에서 사용된 방법보다 향상된 결과를 보였다.

## II. 통계적 분석

잡음신호  $n$ 이 깨끗한 음성신호  $s$ 에 인가되면 오염된 음성신호  $x$ 가 만들어 지고, 음성과 비음성일 때의 각각의 가정에 대하여 다음과 같이 표현 할 수 있다.

$$H_0 : \text{speech absent} : X(t) = N(t) \quad (1)$$

$$H_1 : \text{speech present} : X(t) = N(t) + S(t) \quad (2)$$

여기서  $X(t) = [X_0(t), X_1(t), \dots, X_{M-1}(t)]^T$ 와  $N(t) = [N_0(t), N_1(t), \dots, N_{M-1}(t)]^T$ ,  $S(t) = [S_0(t), S_1(t), \dots, S_{M-1}(t)]^T$ 는 각각 오염된 음성과 잡음, 깨끗한 음성신호의 DFT 계수를 나타내고,  $M$ 은 주파수 bin의 수를 의미한다. 계수의 통계적 모델은 적절한 DFT 계수 분포를 명시해 줌으로써 이용될 수 있다. 본 논문에서는 적용될 분포를 위해서로 다른 두 개의 PDF를 고려하였다. 첫 번째는 음성 분석에서 DFT 계수 분포를 나타내기 위해 가장 많이 사용되는 기존의 복소 가우시안 PDF이다<sup>[1, 9~10]</sup>.

음성 향상을 위한 적절한 통계적 모델을 적용하기 위해서는 주어진 오염된 음성 스펙트럼에 가까운 모델을 선택해야 한다. 이를 위해 다양한 잡음 조건에서 가정  $H_0$ 과  $H_1$  따라 오염된 음성 스펙트럼의 통계적 적합도 테스트가 수행되었다. 적합도 테스트로는 일련의 깨끗한 음성 분포로부터 최소평균제곱오차 추정치를 구하는 [21]에서 제안된 방법과는 다르게, 일반적으로 GOF로 알려진 Kolmogorov-Smirnov (KS) 테스트를 선택하여 각각의 통계적 가정 하에서 신뢰할 수 있는 분석을 도출하였다.  $X(t) = [X_0(t), X_1(t), \dots, X_{M-1}(t)]^T$ 가 오염된 음성의 DFT 계수를 나타내는 벡터라면, KS 테스트는 실험적 누적분포함수 (CDF, Cumulative Distribution Function)  $F_X$ 를 주어진 분포함수  $F$ 와 비교한다. 실험적

CDF는 다음과 같이 [15]에서 정의된다.

$$F_X(z) = \begin{cases} 0, & z < X_{(n)}, \\ \frac{n}{N}, & X_{(n)} \leq z < X_{(n+1)}, \\ 1, & z \geq X_{(n)}, \end{cases} \quad n = 0, 1, \dots, N-1, \quad (3)$$

여기서,  $X_{(n)}$ ,  $n = 0, \dots, N-1$ 은 데이터  $x$ 의 차수 통계이다. 차수 통계를 구하기 위해  $x$ 의 성분들을 분류 배열하면  $X_{(0)}$ 는 가장 작고  $X_{(N-1)}$ 는 가장 큰 성분이 된다<sup>[15]</sup>. 잡음 환경의 시뮬레이션을 위해 각각 4명의 서로 다른 남성, 여성으로부터 추출한 64초의 깨끗한 음성 데이터에 NOISEX-92 데이터베이스의 white noise와 vehicle noise를 Signal-to-Noise Ratio (SNR) 10 dB로 부가하였다. 이러한 샘플 데이터를 기반으로 샘플의 평균, 분산을 계산하여 라플라시안과 가우시안 분포에 각각 적용하였다. 그림 1은 실험적인 CDF와 제시된 함수를 비교한 것이며 white noise와 vehicle noise 조건에서 라플라시안의 CDF curve가 가우시안의 CDF curve보다 실험적 CDF에 더 근접해 있는 것을 보여주고 있다.

실험적인 CDF와 제시된 분포간의 거리측정을 위해 KS 통계  $T$ 를 사용하였으며 [15], [16] 다음과 같이 정의 되어 진다.

$$T = \max |F_X(X_i) - F(X_i)| \quad (4)$$

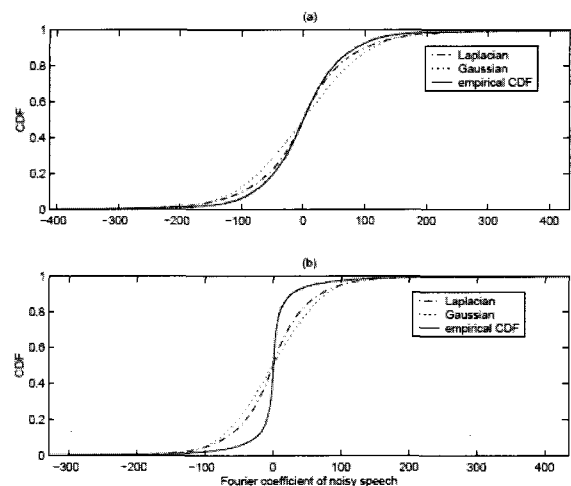


그림 1.  $H_1$ 의 경우에 오염된 음성 스펙트럼 (실수부)에 대한 라플라시안과 가우시안의 CDF 비교 (a) white noise (SNR=10 dB) (b) vehicular noise (SNR=10 dB)

Fig. 1. Comparison of Laplacian and Gaussian CDF of noisy speech spectra (real part) at  $H_1$  (a) white noise (SNR=10 dB) (b) vehicular noise (SNR=10 dB).

표 1. 다양한 잡음 환경에서 오염된 음성의 DFT 계수에 대한 Kolmogorov-Smirnov 테스트 결과 비교  
( $G$ 와  $L$ 은 각각 가우시안과 라플라시안 분포를 의미)

Table 1. Results of Kolmogorov-Smirnov test for the DFT coefficients of noisy speech conditioned on various noise environments.  $G$  and  $L$  denotes the Gaussian and Laplacian distribution, respectively.

noise		white			vehicle			babble		
SNR(dB)		5	10	15	5	10	15	5	10	15
$H_1$	$G; X_{k(R)}$	0.043	0.078	0.129	0.211	0.223	0.231	0.129	0.165	0.198
	$L; X_{k(R)}$	0.031	0.025	0.068	0.164	0.177	0.186	0.071	0.107	0.145
	$G; X_{k(L)}$	0.044	0.081	0.134	0.214	0.225	0.232	0.142	0.173	0.203
	$L; X_{k(L)}$	0.028	0.026	0.073	0.164	0.178	0.187	0.080	0.116	0.149
$H_0$	$G; X_{k(R)}$	0.045	0.052	0.063	0.238	0.270	0.311	0.149	0.127	0.136
	$L; X_{k(R)}$	0.024	0.024	0.023	0.189	0.237	0.277	0.088	0.167	0.078
	$G; X_{k(L)}$	0.051	0.059	0.071	0.243	0.275	0.325	0.153	0.127	0.134
	$L; X_{k(L)}$	0.019	0.016	0.021	0.243	0.237	0.278	0.093	0.067	0.075

여기서 거리는 샘플 포인트  $\{X_i\}$ 에서 평가된  $F_X(\cdot)$ 와  $F(\cdot)$ 사이의 최대 차이를 의미하며, 여러 분포에 대한 데이터를 테스트할 때, 가장 작은  $T$ 를 만드는 PDF가 주어진 데이터에 가장 적합한 분포를 의미한다. 표 1은 테스트에 사용된 모든 잡음 환경에서 라플라시안 분포가 가우시안 분포보다 작은 KS 통계  $T$ 를 갖는 것을 보여 주고 있다. 그러므로 오염된 음성의 DFT 계수에 대한 PDF에는 가우시안보다는 라플라시안 분포가 더 적합하다는 결론을 내릴 수 있다.

### III. 복소 라플라시안 PDF를 적용한 개선된 음성 향상

가우시안 PDF를 적용하면 아래와 같이 두 가지 경우로 오염된 음성 스펙트럼성분을 표현할 수 있다.

$$p_G(X_k|H_0) = \frac{1}{\pi\lambda_{n,k}} \exp\left[-\frac{|X_k|^2}{\lambda_{n,k}}\right] \quad (5)$$

$$p_G(X_k|H_1) = \frac{1}{\pi[\lambda_{n,k} + \lambda_{s,k}]} \exp\left[-\frac{|X_k|^2}{\lambda_{n,k} + \lambda_{s,k}}\right] \quad (6)$$

여기서,  $\lambda_{n,k}$ 와  $\lambda_{s,k}$ 는 각각  $N_k$ 와  $S_k$ 의 분산을 의미한다<sup>[1]</sup>.  $X_{k(R)}$ 과  $X_{k(I)}$ 을 각각 DFT 계수  $X_k$ 의 실수와 허수부라고 정의하면 복소 라플라시안 PDF에 따라  $X_{k(R)}$ 과  $X_{k(I)}$ 는 다음과 같은 분포를 갖는다고 가정할 수 있다.

$$p(X_{k(R)}) = \frac{1}{\sigma_x} \exp\left[-\frac{2|X_{k(R)}|}{\sigma_x}\right] \quad (7)$$

$$p(X_{k(I)}) = \frac{1}{\sigma_x} \exp\left[-\frac{2|X_{k(I)}|}{\sigma_x}\right] \quad (8)$$

여기서  $\sigma_x^2$ 는  $X_k$ 의 분산이고,  $X_k$ 의 실수와 허수부가 서로 독립이라고 가정하면<sup>[17]</sup>  $p(X_k)$ 는 아래와 같다.

$$p(X_k) = p(X_{k(R)}) \cdot p(X_{k(I)}) \quad (9)$$

$$= \frac{1}{\sigma_x^2} \exp\left[-\frac{2(|X_{k(R)}| + |X_{k(I)}|)}{\sigma_x}\right] \quad (10)$$

실수와 허수부가 독립이라는 가정을 증명하기 위해 DFT 계수의 실수와 허수부에 대해 [21]과 유사한 과정의 산점도 (scatter plots)를 평가하면, 실수와 허수부의 상관관계는 약하고 독립이라는 가정에 의해 발생하는 예러는 무시할 수 있을 정도로 작다는 결론을 내릴 수 있다. 식 (10)에서 오염된 음성의 DFT 계수 분포는 다음과 같이 음성  $H_1$ 과 비음성  $H_0$  경우로 표현된다.

$$p_L(X_k|H_0) = \frac{1}{\lambda_{n,k}} \exp\left[-\frac{2(|X_{k(R)}| + |X_{k(I)}|)}{\sqrt{\lambda_{n,k}}}\right] \quad (11)$$

$$p_L(X_k|H_1) = \frac{1}{\lambda_{n,k} + \lambda_{s,k}} \exp\left[-\frac{2(|X_{k(R)}| + |X_{k(I)}|)}{\sqrt{\lambda_{n,k} + \lambda_{s,k}}}\right] \quad (12)$$

입력 신호  $X(t)$ 의 경우에 GSAP는 다음과 같다.

$$p(H_0|X(t)) = \frac{p(X(t)|H_0)p(H_0)}{p(X(t)|H_0)p(H_0) + p(X(t)|H_1)p(H_1)} \quad (13)$$

여기서  $P(H_0) (= 1 - P(H_1))$ 는 음성 부재에 대한 사전확률 (*a priori probability*)이다. 각 주파수의 스펙트럼 성분은 통계적으로 독립이라고 가정하면 아래와 같이 표현될 수 있다.

$$\begin{aligned} p_L(H_0|X(t)) &= \frac{p(H_0) \prod_{k=1}^M p_L(X_k(t)|H_0)}{p(H_0) \prod_{k=1}^M p_L(X_k(t)|H_0) + p(H_1) \prod_{k=1}^M p_L(X_k(t)|H_1)} \\ &= \frac{1}{1 + q \prod_{k=1}^M \Lambda_k(X_k(t))} \end{aligned} \quad (14)$$

여기서

$$q = \frac{p(H_1)}{p(H_0)} \quad (15)$$

이고,  $\Lambda_k(X_k(t))$ 는  $k$ 번째 주파수 성분의 우도비로 다음과 같다<sup>[11]</sup>.

$$\begin{aligned} \Lambda_k(X_k(t)) &= \frac{p_L(X_k(t)|H_1)}{p_L(X_k(t)|H_0)} \\ &= \frac{1}{1 + \xi_k(t)} \exp \left[ 2(|X_{k(R)}(t)| + |X_{k(I)}(t)|) \left( \frac{|X_k(t)| - \sqrt{\lambda_{n,k}(t)}}{|X_k(t)| \sqrt{\lambda_{n,k}(t)}} \right) \right] \end{aligned} \quad (16)$$

여기서

$$\xi_k(t) \equiv \frac{\lambda_{s,k}(t)}{\lambda_{n,k}(t)} \quad (18)$$

이다.

#### IV. 음성 향상

##### 1. 잡음과 음성의 전력 추정

식(14)의 global soft decision의 성능 향상 여부는 일반적으로 깨끗한 음성 전력  $\lambda_{s,k}(t)$ 와 배경 잡음 전력  $\lambda_{n,k}(t)$ 의 정확한 추정이 중요한 요소로 작용한다<sup>[23]</sup>. 대체적으로  $N(t)$ 와  $X(t)$ 가 정상 상태라는 가정 하에  $\lambda_{s,k}(t)$ 와  $\lambda_{n,k}(t)$ 를 추정하기 위해 이와 같은 전력 스펙트럼에 long-term smoothing을 적용한다. 따라서  $\hat{\lambda}_{s,k}(t)$ 와  $\hat{\lambda}_{n,k}(t)$ 을 각각  $\lambda_{s,k}(t)$ 과  $\lambda_{n,k}(t)$ 의 추정치라고 한다면 아래와 같다.

$$\begin{aligned} \hat{\lambda}_{n,k}(t+1) &= \zeta_n \hat{\lambda}_{n,k}(t) + (1 - \zeta_n) E[|N_k(t)|^2 | X(t)] \\ \hat{\lambda}_{s,k}(t+1) &= \zeta_s \hat{\lambda}_{s,k}(t) + (1 - \zeta_s) E[|X_k(t)|^2 | X(t)] \end{aligned} \quad (19)$$

여기서  $0 < \zeta_n < 1$ 과  $0 < \zeta_s < 1$ 는 smoothing 파라미터이다. 식(19)는  $\hat{\lambda}_{s,k}(t)$ 와  $\hat{\lambda}_{n,k}(t)$ 가 입력신호  $X(t)$ 에 의존하지 않는 일종의 관련된 파라미터에 의한 예측된 추정치라는 것을 내포하고 있다. 식(19)의  $X(t)$ 와  $N(t)$ 에 적용한 통계적 가정에 기반하여 아래와 같은 식을 도출할 수 있다.

$$\begin{aligned} E[|N_k(t)|^2 | X(t)] &= E[|N_k(t)|^2 | X(t), H_0] p_L(H_0 | X(t)) \\ &\quad + E[|N_k(t)|^2 | X(t), H_1] p_L(H_1 | X(t)) \\ E[|S_k(t)|^2 | X(t)] &= E[|S_k(t)|^2 | X(t), H_0] p_L(H_0 | X(t)) \\ &\quad + E[|S_k(t)|^2 | X(t), H_1] p_L(H_1 | X(t)) \end{aligned} \quad (20)$$

여기서

$$\begin{aligned} E[|N_k(t)|^2 | X(t), H_0] &= |X_k(t)|^2 \\ E[|N_k(t)|^2 | X(t), H_1] &= \left( \frac{\hat{\xi}_k(t)}{1 + \hat{\xi}_k(t)} \right) \hat{\lambda}_{n,k}(t) + \left( \frac{1}{1 + \hat{\xi}_k(t)} \right)^2 |X_k(t)|^2 \\ E[|S_k(t)|^2 | X(t), H_0] &= 0 \\ E[|S_k(t)|^2 | X(t), H_1] &= \left( \frac{1}{1 + \hat{\xi}_k(t)} \right) \hat{\lambda}_{s,k}(t) + \left( \frac{\hat{\xi}_k(t)}{1 + \hat{\xi}_k(t)} \right)^2 |X_k(t)|^2 \end{aligned} \quad (21)$$

이고

$$\hat{\xi}_k(t) \equiv \frac{\hat{\lambda}_{s,k}(t)}{\hat{\lambda}_{n,k}(t)} \quad (22)$$

이다. 식(21)은 잡음 전력 스펙트럼의 추정치가 비음성 구간뿐만 아니라 음성 구간에서도 갱신되는 특성을 보여주고 있다. 이것은 일반적으로 비음성 구간에서만 잡음 전력을 갱신해주는 기존의 방법과는 구별되는 점이다. 그러나 부정확한 잡음 전력이 음성의 왜곡을 크게 발생시킬 수 있기 때문에 SAP가 threshold  $p_{th}$ 보다 클 경우에만  $\hat{\lambda}_{n,k}(t)$ 을 갱신되도록 하였으며, 본 논문의 실험에서는  $p_{th} = 0.7$ 로 설정하였다.

##### 2. GSAP가 결합된 잡음제거이득

$\hat{S}(t) = [\hat{S}_1(t), \hat{S}_2(t), \dots, \hat{S}_M(t)]$ 는  $t$ 번째 프레임에서 추정된 깨끗한 음성의 스펙트럼을 의미한다. 기존의 일반적인 스펙트럼 향상 기법은 오염된 음성신호의 각각의 스펙트럼 성분에 특정 잡음제거이득을 적용함으로써 추정된 깨끗한 음성신호  $\hat{X}(t)$ 를 얻었다. 본 논문에서는 잡음제거이득을 구하는 여러 방법 중에서 음성 향상 후 음질을 저해하는 뮤지컬 잡음을 제거하는데 우수한 성능을 보이는 Ephraim과 Malah<sup>[9]</sup>에 의해 제안된 EMSR

(Ephraim and Malah Suppression Rule)을 선택하였으며 아래와 같다.

$$\hat{S}_k(t) = G(\eta_k(t), \gamma_k(t)) X_k(t) \quad (23)$$

여기서  $\eta_k(t)$ 와  $\gamma_k(t)$ 는 각각 *a priori* SNR과 *a posteriori* SNR을 나타내며<sup>[1]</sup>, 식(23)의 잡음제거이득  $G(\cdot, \cdot)$ 은 아래식과 같다.

$$G(\eta, \gamma) = \frac{\sqrt{\pi}}{2} \sqrt{\frac{\eta}{\gamma(1+\eta)}} \times M\left[\frac{\gamma\eta}{1+\eta}\right] \quad (24)$$

여기서

$$M[\theta] = \exp\left(-\frac{\theta}{2}\right) \left[ (1+\theta) I_0\left(\frac{\theta}{2}\right) + \theta I_1\left(\frac{\theta}{2}\right) \right] \quad (25)$$

이며, 식(25)에서  $I_0, I_1$ 는 각각 0차, 1차 수정 베셀(modified Bessel) 함수를 의미한다. EMSR에서 지배적 파라미터로 사용되는 *a priori* SNR을 추정하는 decision-directed 방법은 musical noise를 제거하는데 우수한 성능을 가지고 있다<sup>[20]</sup>.  $\hat{\gamma}_k(t)$ 와  $\hat{\eta}_k(t)$ 가 각각  $\gamma_k(t)$ 와  $\eta_k(t)$ 의 추정치라고 하면 decision-directed 방법은 다음과 같다.

$$\hat{\gamma}_k(t) = \frac{|X_k(t)|^2}{\hat{\lambda}_{n,k}(t)}$$

$$\hat{\eta}_k(t) = \alpha \frac{|\hat{S}(t-1)|^2}{\hat{\lambda}_{n,k}(t-1)} + (1-\alpha) P[\hat{\gamma}_k(t) - 1] \quad (26)$$

여기서  $0 < \alpha < 1$ 는 가중치 파라미터로  $\alpha=0.99$ ,  $P[x]$ 는  $P[x]=x$  if  $x \geq 0$ 이고,  $P[x]=0$  if  $x < 0$ 을 의미하는 연산자이다.

식(23)에서  $\hat{S}_k(t)$ 는 입력된 오염 음성의 스펙트럼에서 잡음이 제거된 스펙트럼 추정치이다. 음성/비음성의 경우를 고려한 soft decision이 적용 된다면 아래와 같이 나타낼 수 있다.

$$\hat{S}_k(t) = E[|S_k(t)||X(t)]$$

$$= \frac{E[|S_k(t)||X(t), H_0] p_L(H_0|X(t)) + E[|S_k(t)||X(t), H_1] p_L(H_1|X(t))}{E[|S_k(t)||X(t), H_1] p_L(H_1|X(t))} \quad (27)$$

$$= E[|S_k(t)||X(t), H_1] p_L(H_1|X(t)).$$

식(27)에 의해서 기존의 식(23)의 잡음제거이득을 수정하면 soft decision이 적용된 잡음제거이득  $\tilde{G}(\cdot, \cdot)$ 는 다음과 같다.

$$\tilde{G}(\eta_k(t), \gamma_k(t)) = p_L(H_1|X(t)) G(\eta_k(t), \gamma_k(t)). \quad (28)$$

## V. 실험

본 논문에서는 제안된 복소 라플라시안 PDF 기반의 soft decision이 적용된 음성 향상 알고리즘의 성능 평가를 위해 다양한 잡음 조건에서 객관적인 테스트를 수행하였다. 각각 5명의 남성과 여성 화자가 발음하도록 한 10개의 문장이 테스트에 사용되었으며 NOISEX-92 데이터베이스의 white noise, babble noise, vehicle noise가 다양한 SNR로 깨끗한 음성에 추가되었다.

표 2는 ITU-T P.862 PESQ에 의한 객관적인 수치들을 보여주고 있다. PESQ 결과로부터 대부분의 잡음 환경에서 제안된 복소 라플라시안 알고리즘이 기존의 방법들보다 향상된 결과를 보였다<sup>[1]</sup>.

표 2. 제안된 음성 향상 알고리즘 (Proposed)과 기존의 SEGSD 기법의 PESQ 수치 비교

Table 2. PESQ result for the proposed enhancement algorithm (Proposed) and conventional SEGSD technique.

noise	method	SNR(dB)		
		5	10	15
white	none	1.50	1.80	2.61
	SEGSD	1.71	1.85	2.81
	Proposed	1.81	1.90	2.87
vehicle	none	2.37	2.50	2.64
	SEGSD	2.60	2.65	2.81
	Proposed	2.68	2.71	2.84
babble	none	1.81	2.03	2.21
	SEGSD	1.97	2.22	2.36
	Proposed	2.09	2.30	2.44

## VI. 결론

본 논문에서는 복소 라플라시안 PDF를 이용한 soft decision기반의 음성 향상 알고리즘을 제안하였다. 통계적 모델에 대한 다양한 연구와 실험이 복소 라플라시안 PDF가 오염된 음성 스펙트럼 분포를 표현하기에 우수한 모델로 사용될 수 있다는 것을 보여주었다. 제안된 알고리즘의 성능은 PESQ 평가를 통하여 기존의 방법들<sup>[1]</sup>보다 향상된 결과를 보였다.

## 참고 문헌

- [1] 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.
- [2] J.-H. Chang and N. S. Kim, "Speech enhancement : new approaches to soft decision," *em IEICE Trans. Inf. and Syst.*, vol. 27, E84-D, pp. 1231-1240, Sep. 2001.
- [3] J.-H. Chang and N. S. Kim, "Voice activity detection based on complex Laplacian model," *Electronics Letters*, vol. 39, no. 7, pp. 632-634, Apr. 2003.
- [4] J.-H. Chang, N. S. Kim and S. K. Mitra, "Voice activity detection based on multiple statistical models," *IEEE Trans. Signal Processing*, June 2006.
- [5] J.-H. Chang and N. S. Kim, "A new structural approach in system identification with generalized analysis-by-synthesis for Robust Speech Coding," *IEEE Trans. Speech and Audio Processing*, vol. 14, no. 3, pp. 747-751, May 2006
- [6] J.-H. Chang, "Perceptual weighting filter for robust speech modification," *Signal Processing*, vol. 86, Issue 5, pp. 1089-1093, May 2006.
- [7] J.-H. Chang, "Warped discrete cosine transform-based noisy speech enhancement," *IEEE Trans. Circuit and Systems II*, vol. 52, issue 9, pp. 535-539, Sept. 2005.
- [8] J.-H. Chang and N. S. Kim, "Distorted speech rejection for automatic speech recognition in wireless communication," *IEICE Trans. Info. and Syst.*, vol. E87-D, no. 7, pp. 1978-1981, July 2004.
- [9] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, no. 6, pp. 1109-1121, Dec. 1984.
- [10] J. Sohn and 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.
- [11] J. Sohn, N. S. Kim and W. Sung, "A statistical model-based voice activity detection," *IEEE Signal Processing Letters*, vol. 6, no. 1, pp. 1-3, Jan. 1999.
- [12] I. Cohen and B. Berdugo, "Speech enhancement for nonstationary noise environments," *Signal Processing*, vol 81, pp. 2403-2418, Nov. 2001.
- [13] I. Cohen and 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.
- [14] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *IEEE Signal Processing Letters*, vol. 9, no. 4, pp. 113-116, Apr. 2002.
- [15] A. G. Glen, L. M. Leemis, and D. R. Barr, "Order statistics in goodness-of-fit testin," *IEEE Trans. Reliability.*, vol. 50, no. 2, pp. 209-213, June 2001.
- [16] R. C. Reininger and J. D. Gibson, "Distributions of the two dimensional DCT coefficients for images," *IEEE Trans. Commuications.*, vol. Com-31, no. 6, pp. 835-839, June 1983.
- [17] D. R. Brillinger, *Time Series: Data Analysis and Theory*, New York: Holden-Day, 1981.
- [18] TIA/EIA/IS-127, "Enhanced variable rate codec, speech service option 3 for wideband spectrum digital systems," 1996.
- [19] R. J. McAulary and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.28, pp. 137-145, Apr. 1980.
- [20] O. Cappé, "Elimination of musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. Speech and Audio Processing*, vol. 2, no. 2, pp. 345-349, Apr. 1994.
- [21] R. Martin, "Speech enhancement using MMSE short time spectral estimation with gamma distributed speech priors," *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, pp. 1253-1256, Orlando, FL. , May 2002.
- [22] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Processing Letters*, vol. 10, no. 7, pp. 204-207, July 2003.
- [23] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech and Audio Processing*, vol. 9, no. 5, pp. 504-512, July 2001.
- [24] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol 12, no. 3, pp. 247-251, July 1993.
- [25] N. Ma, M. Bouchard and R. Goubran, "Perceptual Kalman filtering for speech enhancement in colored noise," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 1, pp. 717-720, Montreal, May 2004.

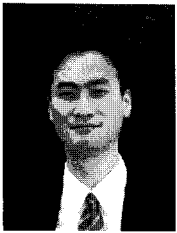
저 자 소 개



**박 윤 식**(학생회원)  
 2006년 인하대학교 전자공학과  
 학사 졸업  
 2006년~현재 인하대학교  
 전자공학과 대학원  
 석사과정  
 <주관심분야 : 잡음제거, 음향학  
 적 반향제거>



**조 규 행**(학생회원)  
 2004년 인하대학교 전자공학과  
 학사 졸업  
 2004년~2006년 LG.Philips LCD  
 연구원  
 2006년~현재 인하대학교  
 전자공학과 석사과정  
 <주관심분야 : 음성검출, 잡음제거>



**장 준 혁**(정회원)  
 1998년 경북대학교 전자공학과  
 학사 졸업  
 2000년 서울대학교 전기공학부  
 석사 졸업  
 2004년 서울대학교 전기컴퓨터  
 공학부 박사 졸업  
 2000년~2005년 (주)넷더스 연구소장  
 2004년~2005년 캘리포니아 주립대학,  
 산타바바라 (UCSB) 박사후연구원  
 2005년~2005년 한국과학기술연구원 (KIST)  
 연구원  
 2005년~현재 인하대학교 전자전기공학부 조교수  
 <주관심분야 : 음성/오디오 신호처리, 통신 신호  
 처리, 휴먼/컴퓨터 인터페이스 등>