

# 일반화된 가우시안 분포를 이용한 신호 준공간 기반의 음성검출기법

## Signal Subspace-based Voice Activity Detection Using Generalized Gaussian Distribution

엄용섭, 장준혁\*, 김동국†

(Yong-Sub Um, Joon-Hyuk Chang\*, and Dong Kook Kim†)

전남대학교 전자컴퓨터공학부, \*한양대학교 전자전기공학부

(접수일자: 2012년 9월 14일; 수정일자: 2012년 10월 19일; 채택일자: 2012년 10월 19일)

**초록:** 본 논문에서는 신호준공간(signal subspace) 영역에서 통계적 모델을 이용한 향상된 음성검출기법을 제안한다. 이를 위해 EP(Embedded Prewhitening) 기법에 의해 비상관적인(uncorrelated) 신호준공간을 생성하고, 이 영역에서 잡음음성과 잡음에 대한 통계적 특성을 파악하였다. 이러한 통계적 특성에 근거하여 GGD (Generalized Gaussian Distribution)을 사용하여 보다 효율적인 음성검출 알고리즘을 제안한다. 실험을 통해 제안된 기법이 0-15dB SNR의 시뮬레이션 환경에서 기존 Gaussian을 사용한 신호준공간 기법보다 향상된 음성검출 결과를 보여준다. **핵심용어:** 음성검출, 신호 준공간, 일반적인 가우시안 분포

**ABSTRACT:** In this paper we propose an improved voice activity detection (VAD) algorithm using statistical models in the signal subspace domain. A uncorrelated signal subspace is generated using embedded prewhitening technique and the statistical characteristics of the noisy speech and noise are investigated in this domain. According to the characteristics of the signals in the signal subspace, a new statistical VAD method using GGD (Generalized Gaussian Distribution) is proposed. Experimental results show that the proposed GGD-based approach outperforms the Gaussian-based signal subspace method at 0-15 dB SNR simulation conditions.

**Key words:** Voice activity detection, Signal subspace, Generalized gaussian distribution

**PACS numbers:** 43.60. -c

### 1. 서론

음성검출기법(VAD, Voice Activity Detection)은 잡음이 섞여있는 음성신호에서 음성이 존재하는 부분과 잡음만 존재하는 부분을 판별하는 기술이다. 이 기술은 현재 음성코딩, 음성향상, 음성인식 그리고 이동통신 등 다양한 분야에 널리 사용되고 있다. 최근에 우도비검증(LRT, Likelihood Ratio Test)을 결정 규칙으로 사용하는 통계적 모델에 기반한 음성검출 기법들이 많이 연구되고 있다.<sup>[1-9]</sup> 이런 통계적 음성

검출 알고리즘은 DFT(Discrete Fourier Transform), DCT(Discrete Cosine Transform)와 KLT(Karhunen-Loeve Transform)와 같은 선형영역에서 동작한다. 이러한 영역에서 우도비에 의한 음성검출을 잘 수행하기 위해서는 각 영역에서 음성과 잡음에 대한 통계적 특성을 정확히 모델링이 할 수 있어야 한다.<sup>[10]</sup> 대부분 통계적 음성검출 기법들은 DFT 영역에서 동작하는데, 이 기법은 입력신호의 DFT 계수들이 Gaussian, Laplacian 또는 Gamma 형태의 확률밀도함수(PDF, Probability Density Function)를 갖는다고 가정한다.<sup>[3]</sup> DFT영역에서 음성을 좀 더 정확하게 모델링하기 위해 GGD(Generalized Gaussian Distribution)와 일반화된 감마분포(generalized Gamma distribution)

†Corresponding author: Dong Kook Kim (dkim@jnu.ac.kr)  
Department of Electronics and Computer Engineering,  
Chonnam National University, 300 Yongbong-dong, Buk-gu,  
Gwangju 500-757, Republic of Korea  
(Tel: 82-62-530-1794, Fax: 82-62-530-1759)

분포를 사용한 음성검출 기법이 제안되었다.<sup>[4,5]</sup>

최근에는 KLT와 같은 신호준공간(signal subspace) 영역에서의 음성검출 기법이 제안되었다.<sup>[6,9]</sup> Gazor 등은 DCT와 KLT 영역에서 음성의 분포를 Laplacian, 잡음 분포를 Gaussian으로 하는 기법을 제시하였다.<sup>[6]</sup> 또한 시간영역에서 동작하는 확률주성분분석(principal component analysis)과 EP(Embedded Prewhitening) 기법을 사용하는 신호준공간 기반한 기법이 제안되었다.<sup>[7,8]</sup> 이러한 신호준공간 기법에서는 잡음음성과 잡음에 대해 다변수 가우시안(multivariate Gaussian) 분포를 갖는다고 가정하였다.

본 논문에서는 신호준공간 영역에서 향상된 음성검출 기법을 제안한다.<sup>[1]</sup> 이를 위해 EP 기법<sup>[11]</sup>을 사용하여 비상관적인(uncorrelated) 신호준공간을 생성하고, 이 영역에서 잡음음성과 잡음에 대한 Kolmogorov-Smirnov (KS)-test<sup>[12]</sup>를 통해 다양한 통계적 특성을 조사한다. 이러한 신호의 통계적 특성을 바탕으로 GGD를 사용한 음성검출 알고리즘을 제안한다. 실험을 통해 신호준공간 영역에서 제안된 GGD 모델이 기존 방식에 비해 음성검출 성능이 향상됨을 보여준다.

## II. Embedded Prewhitening에 의한 신호 준공간

이 장에서는 음성검출을 위한 EP 기법<sup>[11]</sup>에 의한 신호준공간에 대해 소개한다.<sup>[8]</sup>  $x(t)$ 는 시간  $t$ 에서 영 평균 (zero-mean)을 갖는 깨끗한 음성 신호를 나타내고,  $n(t)$ 는 깨끗한 음성신호와 비상관적인 영 평균 가산잡음이다. 이때 잡음음성은  $y(t) = x(t) + n(t)$ 와 같이 표현된다. 이와 같은 잡음음성신호를 프레임기반으로 처리하는 경우에는 벡터 형식으로 표현할 수 있다. 시간영역에서  $D$ 개의 잡음음성 샘플로 구성된 벡터  $\mathbf{y} \equiv [y(t) \ y(t-1) \ \dots \ y(t-D+1)]^T$ 라 하면 잡음음성은 다음과 같이 표현한다.

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \quad (1)$$

여기서 편의상 시간변수  $t$ 는 생략되었고  $\mathbf{x}$ 와  $\mathbf{n}$ 는  $\mathbf{y}$

와 비슷하게 정의된다. 이때  $\mathbf{x}$ ,  $\mathbf{n}$  그리고  $\mathbf{y}$ 는  $R^D$  공간상에 위치한 벡터들이 된다.  $\mathbf{x}$ 와  $\mathbf{n}$ 는 비상관적이기 때문에 잡음음성  $\mathbf{y}$ 의 상관행렬(correlation matrix)  $\mathbf{R}_y$ 는 깨끗한 음성과 잡음의 상관 행렬의 합으로 표현된다.

$$\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_n \quad (2)$$

본 논문에서 음성검출을 위해 사용되는 신호준공간을 얻기 위해 아래와 같이  $\mathbf{R}_x$ 와  $\mathbf{R}_n$ 를 동시에 대각화할 수 있는 EP 기법을 이용한다.<sup>[8,11]</sup>

$$\mathbf{V}^T \mathbf{R}_x \mathbf{V} = \mathbf{\Lambda}, \quad \mathbf{V}^T \mathbf{R}_n \mathbf{V} = \mathbf{I} \quad (3)$$

여기서  $\mathbf{V}$ 와  $\mathbf{\Lambda}$ 는 행렬  $\mathbf{\Sigma} = \mathbf{R}_n^{-1} \mathbf{R}_x = \mathbf{R}_n^{-1} \mathbf{R}_y - \mathbf{I}$ 의 고유벡터 행렬과 고유값 행렬이다.<sup>[11]</sup> 즉,  $\mathbf{\Sigma} \mathbf{V} = \mathbf{V} \mathbf{\Lambda}$ 와 같은 관계식이 성립한다. 여기서 행렬  $\mathbf{\Sigma}$ 가 rank  $P$  ( $P < D$ ) 이면,  $\mathbf{V}$ 는 행렬  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_P]$  이고,  $\mathbf{\Lambda}$ 는 오직  $P$ 개의 non-zero 고유값  $\lambda_1 \geq \lambda_2 \dots \geq \lambda_P > 0$  을 갖는 대각행렬  $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_P\}$ 이다. 이때 행렬  $\mathbf{V}$ 를 사용하여  $R^D$  공간에 놓여있는  $D$ 차원의 잡음음성 벡터  $\mathbf{y}$ 에 선형변환을 수행하면 다음과 같다.

$$\mathbf{z} = \mathbf{V}^T \mathbf{y} \quad (4)$$

그러면 벡터  $\mathbf{z}$ 는  $R^P$  공간상에 놓인  $P$ 차원의 벡터가 되고, 벡터  $\mathbf{z}$ 가 위치한 이 공간을 EP 기법에 의한 신호준공간이라 부른다.

위와 같이 음성검출을 위한 신호준공간을 사용하는 경우의 주요 장점은 신호준공간 영역에서 잡음음성의 성분들은 다른 성분들과 비상관적이기 때문에 이 영역에서 음성신호와 잡음에 대한 최적의 통계적 모델을 이용할 수 있다는 것이다. 이를 살펴보기 위해 신호준공간 영역에서 잡음음성의 성분을 식(1)과 (4)로부터 다음과 같이 표현한다.

$$\mathbf{z} = \mathbf{V}^T (\mathbf{x} + \mathbf{n}) = \mathbf{z}_x + \mathbf{z}_n \quad (5)$$

여기서  $\mathbf{z}_x = \mathbf{V}^T \mathbf{x}$ 와  $\mathbf{z}_n = \mathbf{V}^T \mathbf{n}$ 는 깨끗한 음성과 잡음

1) 본 논문은 기존에 제안된 기법<sup>[9]</sup>에 근거하여 체계적인 설명과 다양한 실험 및 결과를 첨가하였다.

에 대한 신호준공간의 신호성분이다.  $\mathbf{x}$ 와  $\mathbf{n}$ 의 영 평균 가정과 식(5)의 관계로부터  $\mathbf{z}$ 도 영 평균, 즉  $E[\mathbf{z}] = \mathbf{0}$ 임을 알 수 있다. 이때  $\mathbf{z}$ 의 공분산은 식(3)과 (5)로부터 다음과 같이 주어진다.

$$\mathbf{R}_z = E[\mathbf{z}\mathbf{z}^T] = \mathbf{A} + \mathbf{I} \quad (6)$$

여기서  $E[\mathbf{z}_x\mathbf{z}_x^T] = \mathbf{A}$ ,  $E[\mathbf{z}_n\mathbf{z}_n^T] = \mathbf{I}$ 이다. 즉  $\mathbf{z}$ 의 공분산 행렬이 대각 성분만 존재하므로 신호준공간에서  $\mathbf{z}$ 의 성분들이 비상관적임을 알 수 있다. 따라서 원래의 신호공간인  $R^D$  영역 대신 신호 준공간인  $R^P$  영역에서 잡음음성과 잡음을 모델링 하는 것이 훨씬 용이함을 알 수 있다.

### III. 신호준공간에서 신호 모델링

통계적 모델에 기반한 음성 검출기에서는 음성과 잡음에 대한 정확한 확률밀도 함수를 가정하는 것이 중요하다.<sup>[10]</sup> 따라서 신호준공간 영역에서 잡음음성 성분과 잡음 성분에 대한 분포를 조사하는 것이 필요하다. 일반적으로 DFT 영역에서 통계모델로는 Gaussian, Laplacian, Gamma 확률밀도 함수를 많이 사용되고 있다.<sup>[3]</sup> 이 장에서는 신호준공간 영역에서 통계적 모델로서 보다 효율적인 GGD를 사용한 모델링 기법을 소개한다. 스칼라 신호  $z$ 가 영 평균 GGD로 모델링 되는 경우, 이에 대한 확률밀도 함수는 다음과 같다.<sup>[13]</sup>

$$p(z) = \frac{\nu\alpha(\nu)}{2\sigma\Gamma(1/\nu)} \exp\left\{-\left[\alpha(\nu)\frac{|z|}{\sigma}\right]^\nu\right\} \quad (7)$$

여기서  $\alpha(\nu) = \sqrt{\Gamma(3/\nu)/\Gamma(1/\nu)}$ 이며,  $\Gamma(\cdot)$ 는 Gamma 함수이고,  $\nu$ 와  $\sigma$ 는 GGD의 파라미터이다.  $\nu$ 는 shape 파라미터로 지수의 감소율을 나타내며,  $\sigma$ 는 표준편차를 나타낸다. 식(7)에서  $\nu = 1$ 이면 Laplacian,  $\nu = 2$ 이면 Gaussian을 나타낸다.

식(4)와 같이 주어진 신호준공간에서 잡음음성과 잡음에 대한 통계적 특성을 구하기 위해 다양한 잡음환경 하에서 신호들을 조사하였다. Fig. 1과 2는 배블(babble) 잡음 5 dB 조건하에서 신호준공간의 첫 번째

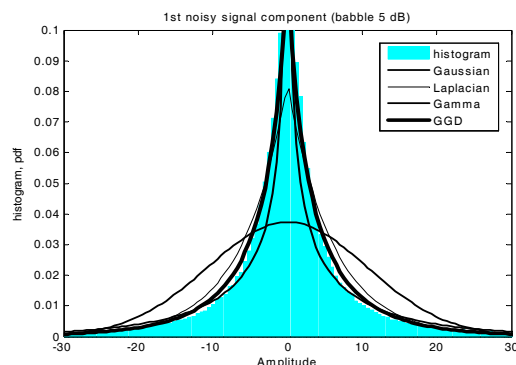


Fig. 1. Histogram and various statistical distributions of the noisy speech for the babble noise at 5 dB SNR.

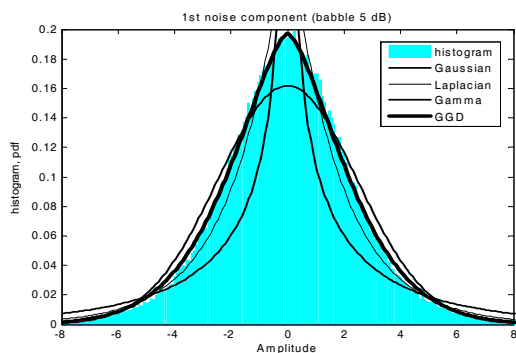


Fig. 2. Histogram and various statistical distributions of the noise for the babble noise at 5 dB SNR.

번째 성분벡터  $\mathbf{v}_1$ 에 의한 잡음음성과 잡음에 대한 크기 성분을 히스토그램(histogram)과 다양한 통계모델로 나타낸 것이다. GGD는 잡음음성에 대해서는  $\nu = 0.7$ , 잡음에 대해서는  $\nu = 1.5$ 를 사용하였다. 잡음음성과 잡음에 대한 표준편차는 식(6)로부터 각각  $\sigma = \sqrt{\lambda + 1}$ 과  $\sigma = 1$ 을 사용하였다. 그림에서 나타난 것 같이 신호준공간에서 최적화된 파라미터 값을 갖는 GGD 모델이 히스토그램에 가장 잘 근사됨을 알 수 있다.

주어진 분포와 실험적인 분포 사이의 객관적인 오차를 측정하기 위해 KS-test<sup>[12]</sup>를 사용하였다. KS-test는 데이터  $\mathbf{X}$ 를 사용하여 실험적인 CDF(Cumulative Distribution Function)  $F_X$ 와 이론적인 CDF  $F$  사이의 최대 거리를 측정하는 테스트로 다음과 같다.<sup>[3,12]</sup>

$$T(\mathbf{X}) = \max |F_X - F| \quad (8)$$

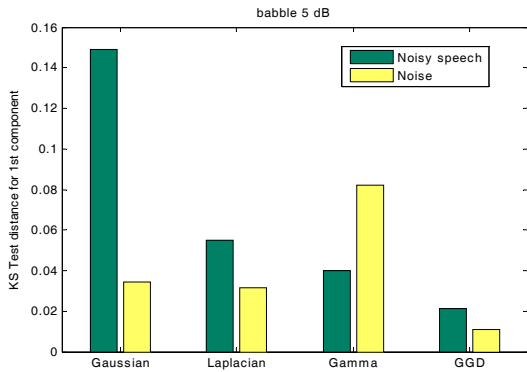


Fig. 3. KS-test results of the noisy speech and noise based on various statistical distributions for the babble noise at 5 dB SNR.

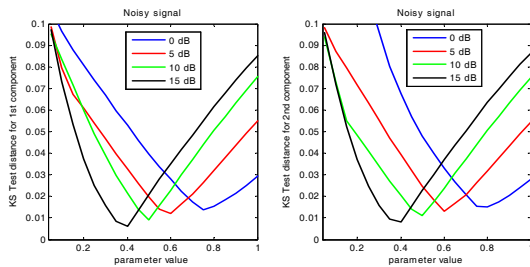


Fig. 4. KS-test results of the noisy speech for the babble noise at 5 dB SNR.

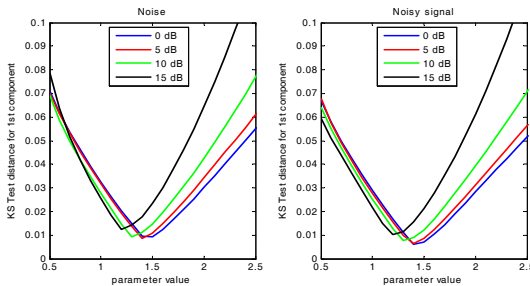


Fig. 5. KS-test results of the noise for the babble noise at 5 dB SNR.

Table 1. Optimal GGD parameters of the noisy speech and noise at various SNRs.

	SNR ≤ 0	0 < SNR ≤ 3	3 < SNR ≤ 7	7 < SNR ≤ 12	12 < SNR ≤ 17	SNR > 17
$\nu_0$	2.0	2.0	1.5	1.4	1.3	1.2
$\nu_1$	0.9	0.8	0.6	0.5	0.4	0.3

가장 작은 KS-test 거리 값을 나타내는 분포가 실험적인 데이터 분포에 가장 잘 맞는 분포를 나타낸다. Fig. 3은 Fig. 1과 2에 나타난 배블(babble) 잡음에 대한

실험적 분포와 다양한 통계모델을 사용하는 경우의 KS-test 거리 값을 나타낸다. 그림에 나타나듯이 GGD 모델이 잡음음성과 잡음에 대한 모델로서 최적임을 나타낸다.

위의 GGD를 사용하기 위해 다양한 잡음에 대해 최적의  $\nu$  값을 추정하는 것이 필요하다. 이를 위해 babble 잡음 5 dB 환경에서 잡음음성과 잡음에 대해 첫 번째와 두 번째 성분에 대해  $\nu$  값을 변화시키면서 KS-test를 시행하였다. Fig. 4와 5는 잡음음성과 잡음에 대한  $\nu$  값에 따른 KS-test 거리 값을 나타낸다. Fig. 4로부터 잡음음성을 위한 파라미터  $\nu_1$  값은 0.4 ~ 0.8 사이에 분포하고 SNR이 낮아질수록  $\nu_1$  값이 일정하게 커짐을 알 수 있다. 반면에 잡음을 위한 파라미터  $\nu_0$ 는 1.3~1.5 사이에 분포함을 알 수 있다. babble 잡음 뿐 아니라 백색(white)이나 핑크(pink) 등과 같은 다른 잡음환경에서도 비슷한 결과가 나타남을 실험을 통해 확인할 수 있었다. 다양한 잡음환경에 대해 최적의 파라미터를 갖는 GGD를 사용하기 위한 실험적인 방법을 통해 SNR에 따라 Table 1과 같이 추정하였다.

#### IV. GGD 기반 VAD

이장에서는 앞장에서 제시된 통계적 모델에 근거하여 GGD를 사용한 음성검출기법을 제안한다. 먼저 음성검출을 위한 두 가지 가설을 가지고 신호준공간에서 잡음음성을 모델링 할 수 있다.  $H_0$ 와  $H_1$  각각 음성 부재와 음성 존재를 나타낼 때 식(5)에 의해 신호준공간에서 통계적 가설은 다음과 같이 주어진다.

$$H_0 : \text{speech absent} : \mathbf{z} = \mathbf{z}_n \tag{9}$$

$$H_1 : \text{speech present} : \mathbf{z} = \mathbf{z}_x + \mathbf{z}_n \tag{10}$$

식(7)을 사용하여 GGD에 의한 각각의 가설에 대한 잡음음성  $\mathbf{z}$ 의  $k$ 번째 성분,  $z_k$ 에 대한 분포는 다음과 같이 나타낼 수 있다.

$$p(z_k | H_0, \nu_0) = \frac{\nu_0 \alpha(\nu_0)}{2\Gamma(1/\nu_0)} \exp\{- (\alpha(\nu_0) |z_k|)^{\nu_0}\} \tag{11}$$

$$p(z_k|H_1, \nu_1) = \frac{\nu_1 \alpha(\nu_1)}{2\Gamma(1/\nu_1) \sqrt{\lambda_k+1}} \exp\left\{-\left(\frac{\alpha(\nu_1)|z_k|}{\sqrt{\lambda_k+1}}\right)^{\nu_1}\right\} \quad (12)$$

여기서  $\nu_0$ 와  $\nu_1$ 은 GGD가  $H_0$ 와  $H_1$ 일 때 각각의 형태(shape) 파라미터이고 Table 1에 나타내듯이 SNR에 따라 그 값이 다르게 사용되며, 각 분포의 표준편차는 식 (6)로부터 각각 1과  $\sqrt{\lambda+1}$ 이다. 따라서 가정된 분포에 의해 음성검출을 위한 신호준공간 영역에서 GGD 기반의  $k$ 번째 성분에 대한 우도비는 다음과 같이 나타낼 수 있다. [4]

$$\Psi_k^{GGD} = \frac{1}{\sqrt{\lambda_k+1}} \frac{\nu_1 \alpha(\nu_1) \Gamma(\nu_0)}{\nu_0 \alpha(\nu_0) \Gamma(\nu_1)} \cdot \exp\left\{\left(\alpha(\nu_0)|z_k|\right)^{\nu_0} - \left(\alpha(\nu_1) \frac{|z_k|}{\sqrt{\lambda_k+1}}\right)^{\nu_1}\right\} \quad (13)$$

또한 신호준공간 영역에서 다양한 통계적 모델들을 적용하여 여러 가지 우도비를 구할 수 있다. 만약  $\nu_0 = \nu_1 = 2$ 이면 Gaussian에 의한 우도비  $\Psi_k^G$ ,  $\nu_0 = \nu_1 = 1$ 이면 Laplacian에 의한 우도비  $\Psi_k^L$ 을 각각 다음과 같이 구할 수 있다.

$$\Psi_k^G = \frac{1}{\sqrt{\lambda_k+1}} \exp\left\{\frac{\lambda_k}{2(\lambda_k+1)} z_k^2\right\} \quad (14)$$

$$\Psi_k^L = \frac{1}{\sqrt{\lambda_k+1}} \exp\left\{\sqrt{2} \frac{\sqrt{\lambda_k+1}-1}{\sqrt{\lambda_k+1}} |z_k|\right\} \quad (15)$$

신호준공간에서 음성검출을 위한 최종적인 결정 규칙은 다음과 같이 우도비들의 기하평균으로 수행된다.

$$\log \Psi = \frac{1}{P} \sum_{k=1}^P \log \Psi_k \begin{matrix} > \\ < \end{matrix} \eta \begin{matrix} H_1 \\ H_0 \end{matrix} \quad (16)$$

여기서  $\eta$ 는 음성검출을 위한 임계값을 나타낸다.

### V. 실험 및 결과고찰

제안된 신호준공간 기반 음성검출 알고리즘을 구

현하기 위해 잡음음성과 잡음에 대한 공분산 행렬  $\mathbf{R}_y$ 와  $\mathbf{R}_n$ 에 대한 정확한 추정이 필요하다. 잡음음성은 직사각형 윈도우(rectangular window)을 사용하여 프레임(frame) 별로 처리하였다. 8 kHz로 샘플링된 음성에 대해 한 프레임으로 20 msec(160샘플)을 사용하였고 50% overlap 하였다.  $D=20$  준프레임(sub-frame)을 사용하여 자기상관 계수 추정을 통해 각각의  $20 \times 20$  잡음음성과 잡음의 공분산 행렬을 추정하였다. 이 공분산 행렬을 바탕으로 행렬  $\mathcal{J} = \mathbf{R}_n^{-1} \mathbf{R}_y - \mathbf{I}$ 에 고유값 분해를 수행하여 고유값과 고유벡터를 얻었다. 신호 준공간의 차원  $P$  값을 결정하기 위해서 행렬  $\mathcal{J}$ 의 고유값이 영보다 큰 고유값의 수로 설정하였다. 그리고 GGD의 SNR에 따른 최적의 파라미터 값을 사용하기 위해 각 프레임마다 SNR 값을 추정하는 것이 필요하다.

이를 위해 SNR의 추정은 신호준공간에서 식  $SNR_{dB} = 10 \log(\sum_{k=1}^P \lambda_k / D)$ 을 사용하였고, 이전 프레임의 SNR 값과 현재 SNR값을 스무딩(smoothing)하여 사용하였다. [14] 그리고 한프레임 내에서  $(2 \times 160) / 20 - 1 = 15$  잡음음성을 생성하여 음성검출에 이용하였다.

시간  $l$ 에서 잡음 공분산은 음성신호가 부재한 구간의 신호를 사용하여 다음과 같은 회귀적 방식에 의해 갱신되었다.

$$\mathbf{R}_n(l) = \alpha_n \mathbf{R}_n(l-1) + (1-\alpha_n) \widetilde{\mathbf{R}}_n(l) \quad (18)$$

여기서  $\alpha_n$ 는 현재의 잡음 공분산 추정값  $\widetilde{\mathbf{R}}_n(l)$ 와 전 데이터 샘플의 영향을 조절하는 망각 인자(forgetting factor)이다.

제안된 알고리즘의 실험결과를 평가하기 위해 여러 가지 음성검출 알고리즘에 대한 탐색(detection)과 False-Alarm(FA) 확률  $P_D$ 와  $P_{FA}$ 를 조사하였다.  $P_D$ 는 실제로 정확하게 음성이라고 판단 할 확률을 뜻하고,  $P_{FA}$ 는 비음성을 음성이라 잘못 판단할 확률을 뜻한다.  $P_D$ 와  $P_{FA}$ 의 값을 계산하기 위하여 456초의 음성을 10ms단위로 수동으로 된 레이블을 기준으로 삼았다. 전체 신호 중에서 음성의 비율은 58.2%이고, 이중 44.85%는 유성음(voiced sounds)이며 13.4%는 무성음(unvoiced sound)이었다. 잡음에 오염된 음성신호를

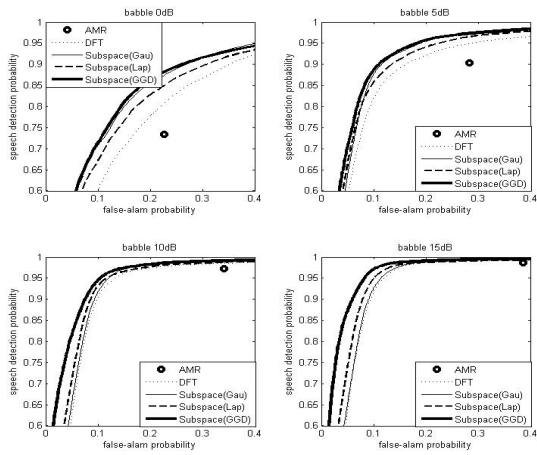


Fig. 6. ROC curves for the babble noise at various SNR conditions.

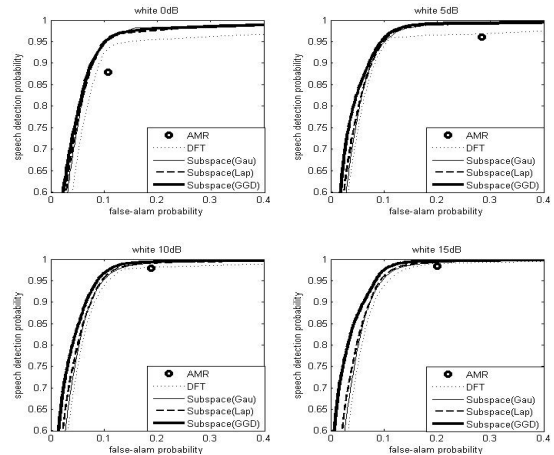


Fig. 8. ROC curves for the white noise at various SNR conditions.

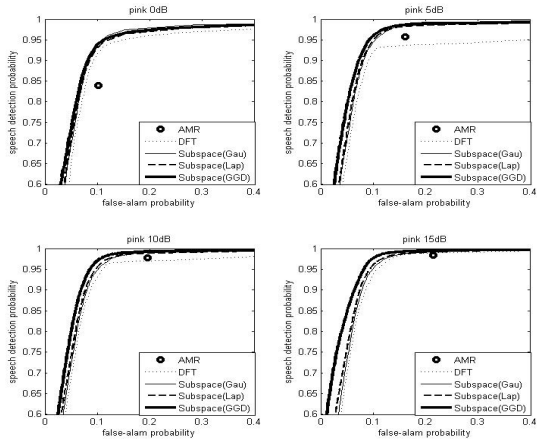


Fig. 7. ROC curves for the pink noise at various SNR conditions.

만들기 위해 white, babble, pink잡음을 NOISEX-92 잡음으로부터 SNR을 변화하면서 원래의 음성신호에 첨가하였다.

제안된 신호준공간 기반 음성검출기의 성능을 평가하기 위해 기존의 Gaussian 분포를 사용한 DFT 기반과 신호준공간 기법을 비교하였다. 새로운 기법으로는 신호 준 공간에서 Laplacian과 GGD를 사용하였다. 모든 알고리즘에는 같은 hang-over 기법을 적용하였다.<sup>[2]</sup> 또한 ETSI AMR VAD option 2와 같은 표준 음성검출기를 비교를 위해 포함하였다. Fig. 6-8은 babble, pink 및 white 잡음에서 AMR, Gaussian 기반 DFT와 신호준공간 기법 그리고 제안된 Laplacian와 GGD 기법에 대한 ROC(Receiver Operating Characteristic) 곡선을 나타낸다. 실제 실험 결과 신호준공간 기반

GGD 음성검출기는 Gaussian이나 Laplacian 기반 기법보다 뛰어난 성능을 보여준다. 그리고 특히 SNR이 더 높을수록 GGD를 사용한 신호준공간 기법이 다른 기법에 비해 성능 향상이 더 크게 나타남을 확인할 수 있다. GGD 기법이 다른 기법에 비해 성능이 향상된 이유로는 신호준공간에서 음성신호와 잡음의 통계치와 잘 맞는 GGD의 파라미터 값을 추정하여 이용했기 때문이다.

## VI. 결 론

본 논문에서는 잡음과 음성을 판별하기 위해 신호준공간 영역에서 GGD 기반 음성검출 알고리즘을 제안하였다. 비상관적인 신호준공간을 생성하기 위해 EP 기법을 사용하였고, 이 영역에서 잡음음성과 잡음에 대한 통계적 특성을 조사하였다. 이러한 통계적 특성에 따라 신호준공간에서 GGD를 사용한 음성검출 알고리즘을 제안하였고, GGD의 최적의 파라미터를 추정하기 위해 KS-test를 사용하였다. 다양한 잡음환경에서의 실험 결과 제안한 신호준공간 기반 GGD 음성검출 알고리즘이 기존의 알고리즘보다 우수한 성능을 나타내었다.

## 감사의 글

이 논문은 2011년도 전남대학교 학술연구비의 지

원에 의하여 연구되었음. 또한 이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2012R1A1A 2043554).

## 참고 문헌

1. J. S. Sohn, N. S. Kim, and W. Y. Sung, "A statistical model-based voice activity detection," IEEE Sig. Proc. Lett. **6**, 1-3 (1999).
2. Y. D. Cho and A. Kondoz, "Analysis and improvement of a statistical model-based voice activity detector," IEEE Sig. Proc. Lett. **8**, 276-278 (2001).
3. J. H. Chang, N. S. Kim, and S. K. Mitra, "Voice activity detection based on multiple statistical models," IEEE Trans. Signal Proc. **54**, 1965-1976 (2006).
4. J. H. Chang, J. W. Shin, and N. S. Kim, "Voice activity detector employing generalized Gaussian distribution," IEE Electronics Lett. **40**, 1561-1563 (2004).
5. J. W. Shin, J. H. Chang, and N. S. Kim, "Statistical modeling of speech signals based on generalized Gamma distribution," IEEE Sig. Proc. Lett. **12**, 258-261 (2005).
6. S. Gazor and W. Zhang, "A Soft Voice Activity Detector Based on a Laplacian-Gaussian Model," IEEE Trans. Speech and Audio Proc. **11**, 498-505 (2003).
7. K. C. Ryu and D. K. Kim, "Statistical voice activity detector based on signal subspace model," (In Korean), J. Acoust. Soc. Kr. **27**, 372-378 (2008).
8. D. K. Kim and J.H. Chang, "A subspace approach based on embedded prewhitening for voice activity detection," J. Acoust. Soc. Am. **130**, 304-310 (2011).
9. Y. S. Um, J. H. Chang, and D. K. Kim "An improved VAD approach based on generalized Gaussian distribution in signal subspace domain," (In Korean), J. Acoust. Soc. Kr. 2(s) **30**, 136-139 (2011).
10. S. Gazor and W. Zhang "Speech probability distribution," IEEE Sig. Proc. Lett. **10**, 204-207 (2003).
11. Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," IEEE Trans. Speech Audio Proc. **11**, 334-341 (2003).
12. A. G. Glen, L. M. Leemis, and D. R. Barr, "Order statistics in goodness-of-fit testing," IEEE Trans. Reliab. **50**, 209-213 (2001).
13. S. Nadarajah, "A generalized Gaussian distribution," J. Appl. Stat. **32**, 685- 694 (2005).

## 저자 약력

### ▶ 엄 용 섭(Yong-Sub Um)



2009년 2월: 광주대학교 광통신전자공학부  
학사  
2012년 8월: 전남대학교 전자컴퓨터공학부  
석사  
2012년 7월 ~ 현재: 한국알프스 펌웨어 엔  
지니어

### ▶ 김 동 국(Dong Kook Kim)



1989년 2월: 전남대학교 전자공학과 학사  
1991년 2월: 포항공과대학 전자전기공학과  
석사  
2003년 2월: 서울대학교 전기컴퓨터공학부  
박사  
1991년 2월 ~ 1999년 2월: 삼성종합기술원  
전문연구원  
2003년 4월~2004년 2월: 한국전자통신연  
구원 선임연구원  
2004년 2월~현재: 전남대학교 전자컴퓨  
터공학부 부교수

### ▶ 장 준 혁(Joon-Hyuk Chang)



1998년 2월: 경북대학교 전자공학과 학사  
2000년 2월: 서울대학교 전기공학부 석사  
2004년 2월: 서울대학교 전기컴퓨터공  
학부 박사  
2000년 3월 ~ 2005년 4월: 넷더스 연구  
소장  
2004년 5월 ~ 2005년 4월: 캘리포니아  
주립대학, 산타바바라(UCSB) 박사후  
연구원  
2005년 5월 ~ 2005년 8월: 한국과학기술  
연구원(KIST) 연구원  
2005년 9월 ~ 2011년: 인하대학교 전자  
공학부 조교수  
2011년 ~ 현재: 한양대학교 융합전자공  
학부 부교수