

# 조건 사후 최대 확률과 음성 스펙트럼 변이 조건을 이용한 통계적 모델 기반의 음성 검출기

## A Statistical Model-Based Voice Activity Detection Employing the Conditional MAP Criterion with Spectral Deviation

김 상 균, 장 준 혁  
(Sang-Kyun Kim, Joon-Hyuk Chang)

한양대학교 융합전자공학부  
(접수일자: 2011년 5월 7일; 채택일자: 2011년 7월 11일)

본 논문에서는 조건 사후 최대 확률 (conditional maximum a posteriori, CMAP)과 음성 스펙트럼 변이 조건을 기반으로 한 새로운 음성 검출기 (voice activity detection, VAD)를 제안한다. 제안된 음성 검출기는 통계적 모델을 기반으로 한 우도비 테스트 (likelihood ratio test, LRT)의 문턱값을 결정 하는데 조건 사후 최대 확률과 스펙트럼 변이의 상태 값을 조건부 확률로 부과한다. 제안된 알고리즘을 다양한 잡음 환경에서 기존의 CMAP 기반의 음성 검출기와 비교한 결과 전체적으로 향상된 성능을 보였으며 특히 SNR이 낮은 조건에서 향상 폭이 컸다.

**핵심용어:** 음성 검출기, 스펙트럼 변이, 조건 사후 최대 확률, 통계적 모델

**투고분야:** 음성처리 분야 (2)

In this paper, we propose a novel approach to improve the performance of a statistical model-based voice activity detection (VAD) which is based on the conditional maximum a posteriori (CMAP) with deviation. In our approach, the VAD decision rule is expressed as the geometric mean of likelihood ratios (LRs) based on adapted threshold according to the speech presence probability conditioned on both the speech activity decisions and spectral deviation in the pervious frame. Experimental results show that the proposed approach yields better results compared to the CMAP-based VAD using the LR test.

**Keywords:** Voice activity detection, Deviation, Conditional maximum a posteriori, Statistical model

**ASK subject classification:** Speech Signal Processing (2)

### I. 서론

음성인식, 음성 부호화 그리고 음성 향상 등 음성신호 처리 기술에서 음성이 존재하는 부분과 아닌 부분을 결정하는 음성 검출기 (voice activity detector, VAD)는 매우 중요한 기술이다. 특히 신호 대 잡음비 (signal-to-noise ratio, SNR)가 낮은 환경에서 음성 구간을 검출하는 정확도는 음성 처리 시스템의 성능에 큰 영향을 준다. 이러한 이유로 음성 검출기의 성능을 향상시키기위한 다양한 알고리즘들이 계속 연구되어지고 있다 [1].

그중 Ephraim과 Malah의 연구에서 시작된 minimum

mean square error (MMSE) 기반의 음성 향상 기법에 사용된 음성의 존재와 부재에 대한 통계적 모델을 우도비 테스트 (likelihood ratio test, LRT)에 적용한 음성

검출기의 성능이 매우 우수한 것으로 알려져 있다 [2-9]. 다시 말해 위의 방법은 음성에 대한 통계모델을 가우시안 분포로 가정하였으며 decision-directed (DD) 기법을 적용 하여 음성의 존재와 부재에 대한 우도비 (likelihood ratio, LR)를 기하평균한 판별식으로부터 음성존재여부를 판단한다 [4].

잡음이 음성신호를 간섭하면 음성의 꼬리와 같은 약한 음성 신호를 검출하기 힘들다. 이러한 취약점을 개선한 방법 중의 하나가 조건 사후 최대 확률 (conditional maximum a posteriori, CMAP) 기반의 음성검출 알고리즘이다. 일반적으로, 음성의 활동은 인접한 프레임들과

강력한 상호 연관성이 있으므로 음성이 활동하는 프레임의 바로 전 프레임이나 바로 다음 프레임은 음성이 활동할 가능성이 높다고 할 수 있고, 그 반대 경우도 성립한다. 이러한 조건을 음성 검출기에 적용한 것이 최근에 제안된 CMAP 기반의 음성검출 알고리즘이며 우수한 성능을 보였다 [10].

본 연구에서는 음성 스펙트럼 변이 (spectral deviation, SD)를 음성검출을 위한 추가정보로서 유용하게 통계모델 기반의 알고리즘에 사용가능하다는 것을 고찰한다. 음성 스펙트럼 변이는 이전 프레임 동안 예측된 평균 long-term 파워 스펙트럼과 현재 파워 스펙트럼의 차이이다 [11]. 본 논문에서는 기존의 CMAP 기반의 음성 검출 판별식에서 문턱값을 결정할 때 음성 스펙트럼 변이 상태를 조건부 확률로 부과한다. 즉, 프레임간 스펙트럼의 변이에 따라 증가하는 상태인 경우, 유지되는 경우, 감소하는 경우의 세가지 경우에 대해 서로 다른 조건을 CMAP에 부가하여 결과적으로 문턱값이 적응적으로 변화하게 되는 새로운 알고리즘을 도입한다. 제안된 음성 검출 기법을 다양한 잡음 환경에서 기존의 CMAP 기반의 음성 검출기와 비교한 결과 전체적으로 향상된 성능을 보였으며 특히 SNR이 낮은 조건에서 향상 폭이 컸다.

본 논문의 II장에서는 CMAP 기반의 음성 검출기에 대해 소개하고, III장에서는 스펙트럼 변이 상태 조건을 부과한 새로운 음성검출 방식에 대해 논하였다. IV장에서는 기존의 방식과 비교실험 결과를 보여주며, 마지막으로 V장에서 결론을 맺어 본 논문을 마친다.

## II. Conditional maximum a posteriori (MAP) 기반의 음성 검출기

음성 검출기는  $n$  번째 프레임에서의  $k$  번째 신호가 잡음만 존재할 경우와 음성과 함께 존재할 경우를 각각 가설  $H_0(k,n)$ ,  $H_1(k,n)$ 로 표현 하면 아래와 같이 나타낼 수 있다.

$$H_0(k,n) : Y(k,n) = D(k,n) \tag{1}$$

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

여기서  $Y(k,n)$ ,  $X(k,n)$  그리고  $D(k,n)$ 는 입력 신호, 음성 신호 그리고 잡음 신호의 이산 푸리에 변환 (discrete Fourier transform, DFT) 계수이다.

음성과 잡음신호의 스펙트럼이 복소 가우시아 분포를

따른다고 가정을 하면, 가설  $H_0(k,n)$ ,  $H_1(k,n)$ 을 조건으로 한 확률밀도함수는 다음과 같이 주어진다 [5].

$$p(Y(k,n)|H_0(k,n)) = \frac{1}{\pi\lambda_d(k,n)} \exp\left\{-\frac{|Y(k,n)|^2}{\lambda_d(k,n)}\right\} \tag{3}$$

$$p(Y(k,n)|H_1(k,n)) = \frac{1}{\pi[\lambda_d(k,n) + \lambda_x(k,n)]} \exp\left\{-\frac{|Y(k,n)|^2}{\lambda_d(k,n) + \lambda_x(k,n)}\right\} \tag{4}$$

여기서  $\lambda_x(k,n)$ 과  $\lambda_d(k,n)$ 는 각각 채널별 음성과 잡음의 분산이며, 이때  $k$  번째 주파수 밴드에 대한 우도비는 아래와 같이 구한다.

$$\Lambda(k,n) \equiv \frac{p(Y(k,n)|H_1(k,n))}{p(Y(k,n)|H_0(k,n))} = \frac{1}{1 + \xi(k,n)} \exp\left\{\frac{\gamma(k,n)\xi(k,n)}{1 + \xi(k,n)}\right\} \tag{5}$$

여기  $\xi(k,n) = \lambda_x(k,n)/\lambda_d(k,n)$ 과  $\gamma(k,n) = |Y(k,n)|^2/\lambda_d(k,n)$ 는 각각 *a priori* signal-to-noise ratio (SNR)과 *a posteriori* SNR이다 [4]. 음성 부재 구간에서 갱신되는 잡음 신호로부터 구한 잡음 분산  $\lambda_d(k,n)$ 을 이용하여 *a posteriori* SNR  $\gamma(k,n)$ 을 추정하며, 또한 *a priori* SNR  $\xi(k,n)$ 은 decision-directed (DD) 기법을 이용하여 아래와 같이 추정한다 [2].

$$\hat{\xi}(k,n) = \alpha \frac{|\hat{X}(k,n-1)|^2}{\lambda_d(k,n-1)} + (1-\alpha)P[\gamma(k,n) - 1] \tag{6}$$

여기서  $|\hat{X}(k,n-1)|$ 은 이전 프레임에서 추정된 음성 신호의  $k$  번째 스펙트럼 성분의 크기에 대한 추정치이며, MMSE에 기반하여 구한다 [4]. 또한  $\alpha$ 는 가중치 값이며, 연산자  $P[\cdot]$ 은 아래와 같이 정의된다.

$$P[x] = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

기존의 통계적 모델 기반의 음성검출기에 대한 결정식은 각각의 주파수 채널에서 구해진 우도비를 기하 평균하여 아래와 같이 음성 검출 여부를 판단 한다 [3-9].

$$\log \Lambda(n) = \frac{1}{M} \sum_{k=1}^M \log \Lambda(k,n) \begin{matrix} > \eta \\ < \eta \end{matrix} \begin{matrix} H_1 \\ H_0 \end{matrix} \tag{8}$$

여기서  $M$ 은 전체 주파수 대역의 개수이며,  $\eta$ 는 음성 검출 문턱값이다.

기존의 MAP 기준은 아래와 같은 결정법을 제안한다.

$$\frac{p(H(k,n) = H_1 | Y(k,n))}{p(H(k,n) = H_0 | Y(k,n))} > \frac{H_1}{H_0} \quad (9)$$

위식에 Bayes 법칙을 적용하면 다음과 같이 우도비 테스트 형태로 표현할 수 있다.

$$\frac{p(Y(k,n) | H(k,n) = H_1)}{p(Y(k,n) | H(k,n) = H_0)} > \frac{p(H(k,n) = H_0)}{p(H(k,n) = H_1)} \quad (10)$$

하지만 식 (3)와 (4)에서 알 수 있듯이  $H_0$ 의 분산은  $\lambda_d(k,n)$ 이고  $H_1$ 의 분산은  $\lambda_d(k,n) + \lambda_x(k,n)$ 이다. 때문에  $H_1$ 쪽으로 바이어스된 것을 보상해 주기 위해 위의 식 우변에 상수를 곱해준다 [4,10].

$$\frac{p(Y(k,n) | H(k,n) = H_1)}{p(Y(k,n) | H(k,n) = H_0)} > \alpha \frac{p(H(k,n) = H_0)}{p(H(k,n) = H_1)} \quad (11)$$

여기서  $\alpha \geq 1$  이다.

위의 MAP 결정법은 하나의 문턱값으로 결정을 내린다. 하지만 일반적으로 음성의 활동은 인접 프레임 간에 강력한 상호 연관성이 있다. 이것은 음성이 활동하는 프레임의 바로 전 프레임이나 바로 다음 프레임은 음성이 활동할 확률이 높다는 것이고, 그 반대의 경우에도 같다고 할 수 있다. 이러한 조건은 기존의 행오버를 사용함으로써 통계 모델을 기반으로 한 VAD의 에러를 추가적으로 줄일 수 있다 [4,10]. 음성 활동 전후프레임들의 강력한 상호 연관성의 성질들로 인해 기존의 posterior probability  $P(H(k,n) | Y(k,n))$  대신 현재 프레임의 관찰결과와 이전 프레임의 결정을 조건으로 한 a posteriori probability  $p(H(k,n) | Y(k,n), H(k,n-1))$ 를 계산해 낼 수 있다. 이 조건을 결정식에 부여한 것이 CMAP 기반의 음성 검출기이다. CMAP 기반의 음성 검출기는 직전 프레임의 음성 존재와 부재 확률에 따라 2개의 문턱값을 부여하여 음성의 활동 여부를 판별하며 우수한 성능을 보였다 [10].

### III. 제안된 음성 검출기

예측된 음성의 스펙트럼 변이는 아래처럼 계산된다.

$$\Delta(n) = \sum_{i=0}^M |E(n,i) - \bar{E}(n,i)| \quad (12)$$

여기서  $E(n,i)$ 는 현재 파워 스펙트럼을 나타내고  $\bar{E}(n,i)$ 는 이전 프레임 동안 예측된 평균 long-term 파워 스펙트럼이다. 초기  $\bar{E}(1,i)$  값은  $E(1,i)$ 로 정의한다.

다음 프레임에 사용될  $\bar{E}(n+1,i)$ 는 아래의 식에 의해 갱신된다 [12].

$$\bar{E}(n+1,i) = \beta \bar{E}(n,i) + (1-\beta) E(n,i) \quad (13)$$

여기서  $\beta$ 는 가중치 값이며 본 논문에서는  $0.5 \leq \beta \leq 0.8$  범위에서 적용하였다.

본래의 스펙트럼 변이는 식 (12)에서처럼 절대 값을 취하여 평균 long-term 파워 스펙트럼과의 차이만을 정보로 활용한다. 현재 파워 스펙트럼의 증가와 감소에 대한 추가 정보를 얻기 위해 본 논문에서는 아래와 같이 변형한 후 문턱값  $\gamma$ 와 비교한다.

$$\Delta'(n) = \sum_{i=0}^M E(n,i) - \bar{E}(n,i) \quad (14)$$

여기서  $\Delta'(n)$ 는  $\gamma$ 에 의해 다음과 같이 3 가지 경우의 조건으로 분류된다.

$$\Delta'(n) \leq -\gamma \quad (15)$$

$$-\gamma < \Delta'(n) < \gamma \quad (16)$$

$$\gamma \leq \Delta'(n) \quad (17)$$

위의 3 가지 경우의 조건에 대한 표현을 간단히 하기 위해 앞으로  $D_i$ 로 표현하며 식 (15), 식 (16) 그리고 식 (17)에 대한 조건은  $i = -1, 0, 1$ 으로 각각 대응시킨다.

위에서 구한 3가지 조건과 CMAP 조건을 식 (9)에 적용하면 아래와 같다.

$$\frac{p(H(k,n) = H_1 | Y(k,n), D(k,n) = D_i, H(k,n-1) = H_j)}{p(H(k,n) = H_0 | Y(k,n), D(k,n) = D_i, H(k,n-1) = H_j)} > \psi \quad (18)$$

여기서  $\psi$ 는 문턱값이고  $j$ 는 음성의 존재와 부재를 1, 0으로 각각 나타낸다. 위 식에 Bayes 법칙을 적용하여 식 (10)을 도출한 것처럼 우도비 테스트를 유도하면 다음과 같다.

$$\frac{p(Y(k,n)|H(k,n) = H_1, D(k,n) = D_i, H(k,n-1) = H_j)}{p(Y(k,n)|H(k,n) = H_0, D(k,n) = D_i, H(k,n-1) = H_j)} > \psi \frac{p(H(k,n) = H_0|D(k,n) = D_i, H(k,n-1) = H_j)}{p(H(k,n) = H_1|D(k,n) = D_i, H(k,n-1) = H_j)} \quad (19)$$

기존의 CMAP 기준과 같은 이유로 위의 식을 아래와 같이 근사화 시킬 수 있다.

$$\frac{p(Y(k,n)|H(k,n) = H_1)}{p(Y(k,n)|H(k,n) = H_0)} > \psi \frac{p(H(k,n) = H_0|D(k,n) = D_i, H(k,n-1) = H_j)}{p(H(k,n) = H_1|D(k,n) = D_i, H(k,n-1) = H_j)} \quad (20)$$

여기서 표현을 간단히 하기위해 식 (20)의 우변을  $\psi_{i,j}$ 로 표현한다.

$$\frac{p(Y(k,n)|H(k,n) = H_1)}{p(Y(k,n)|H(k,n) = H_0)} > \psi_{i,j} \quad (21)$$

여기서 각각의 주파수 채널에서 구해진 우도비를 기하 평균하여 아래와 같이 음성 검출 여부를 판단한다.

$$\log \Lambda'(n) = \frac{1}{M} \sum_{k=1}^M \log \Lambda'(k,n) > \log \psi_{i,j} \quad (22)$$

여기서  $M$ 은 전체 주파수 대역의 개수이다. 최종적으로 위의 식이 음성 스펙트럼 변이 조건과 직전 프레임에서 음성의 존재와 부재에 대한 조건에 따른 문턱값을 적용한 결정식이다.

### IV. 실험 결과

본 논문에서 제안된 음성 검출기의 성능을 평가하기 위해 CMAP 기반의 음성 검출기와 음성 검출 성능을 비교하였다 [10]. 실험에 사용된 데이터는 기존의 음성 검출

알고리즘에서 성능 평가 비교를 위해 사용된 음성 데이터의 길이를 고려하여 각각 4 명의 남성, 여성화자가 말한 총 456초의 음성을 8 kHz로 샘플링하였다. 또한 평가를 위해 깨끗한 음성 데이터에 음성과 비음성 부분을 10 ms 마다 수동으로 표시하였다. 분류된 음성 데이터의 음성 구간은 총 58.2 %로 유성음 44.5 %, 무성음 13.4 %로 구성되었다. 또한 기존의 음성검출 평가에 사용된 car, street, office 잡음과 새로 추가된 babble, white 잡음을 5, 10 dB 그리고 15 dB SNR로 부과했다. 계산된  $\log \psi_{i,j}$ 의 범위는  $\log \psi_{i,j} \in [2, 32]$ 이다. 입력 음성의 특성상 처음 0.04초를 노이즈로 가정하기 때문에 초기 값의 조건은 이전 프레임노이즈이고 음성 스펙트럼 변이는  $D_0$ 라고 가정하여  $\log \psi_{0,0}$ 를 초기 문턱값으로 정하였다.

표 1은 기존의 CMAP 기반의 음성 검출기와 음성 검출 성능을 비교하기 위해  $P_e$  (probability of error),  $P_m$  (probability of miss),  $P_{fa}$  (false alarm probability)를 나타낸 것이다. 표 1을 보면 모든 상황에서 제안된 알고리즘이 기존의 음성 검출기보다 성능이 우수한 것을 볼 수 있다. 특히 비정상 잡음인 street와 office 잡음에서의  $P_e$  성능 향상 폭은 SNR 5 dB [0.41, 0.43]로 나머지 잡음에서의 향상 폭인 [0.24, 0.27]보다 높았으며 SNR 10 dB 그리고 15 dB에서도 street와 office 잡음에서 성능 향상

표 1. 기존의 CMAP 기반의 음성 검출기와 제안된 음성 검출기의 성능 비교

Table 1. Comparison of voice activity detection probability of error ( $P_e$ ), probability of miss ( $P_m$ ) and false alarm probability ( $P_{fa}$ ) between the method of the CMAP-based and the proposed technique.

noise	SNR (dB)	CMAP-based			proposed		
		$P_e$	$P_m$	$P_{fa}$	$P_e$	$P_m$	$P_{fa}$
car	5	5.86	1.85	11.46	5.62	1.78	10.94
	10	5.61	1.75	10.99	5.51	1.69	10.82
	15	5.40	1.65	10.61	5.34	1.60	10.53
street	5	11.08	5.13	19.34	10.67	5.12	18.38
	10	7.69	4.12	12.68	7.43	3.77	12.54
	15	6.64	2.81	11.98	6.46	2.71	11.67
office	5	16.21	12.24	21.72	15.78	11.93	21.13
	10	11.27	7.48	16.65	10.84	6.79	16.48
	15	8.40	2.92	16.03	8.16	2.81	15.57
babble	5	12.41	9.99	15.76	12.28	9.84	15.69
	10	7.76	4.53	12.64	7.52	4.22	12.13
	15	5.92	2.13	11.21	5.78	2.04	10.97
white	5	6.61	3.43	11.02	6.34	3.06	10.92
	10	6.12	2.25	11.53	5.86	1.94	11.32
	15	5.84	1.80	11.45	5.71	1.75	11.21

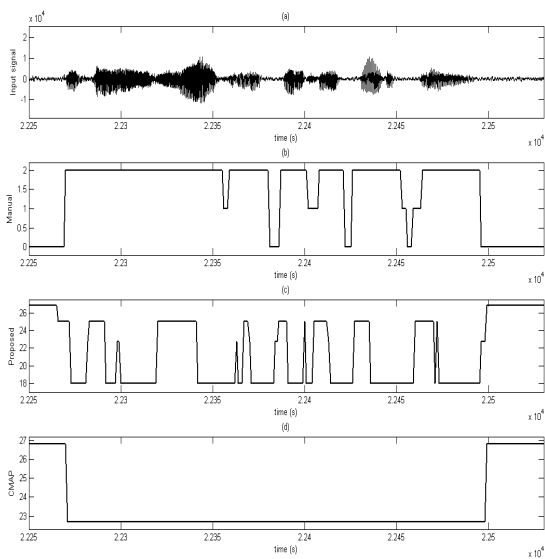


그림 1. (a) 테스트 파일의 파형, (b) 테스트 파일의 매뉴얼 결정, (c) 제안된 알고리즘의 문턱값, (d) 기존의 CMAP의 문턱값  
 Fig. 1. (a) Waveform of the test file, (b) Manual decision of the test file, (c) Threshold of the proposed method, (d) Threshold of the CMAP.

이 높았다. 또한 모든 잡음에서 SNR이 낮을수록 성능 향상 폭이 높게 나오는 경향을 볼 수 있다.

그림 1은 CMAP과 제안된 알고리즘이 실시간으로 입력된 음성 파형에 따른 문턱값의 변화를 보여준다. 그림 1에 b는 테스트 파일의 매뉴얼이며 0은 노이즈, 1은 무성음 그리고 2는 유성음을 뜻한다. 보시는 바와 같이 기존의 CMAP보다 제안한 알고리즘의 문턱값이 음성의 시작과 끝부분에서 약간 낮게 떨어짐으로써 음성의 시작과 끝부분을 보다 효율적으로 검출이 가능해 지는 것을 볼 수 있다. 실험결과에서 보여주는 것처럼 기존의 방법보다 제안된 음성 검출 방법이 우수하다.

## V. 결론

본 논문에서는 음성의 CMAP 기반의 음성 검출기 성능 향상을 위해 음성 스펙트럼 변이 상태를 조건부확률로 우도비 테스트에 적용하는 방법을 제안하였다. 즉, 음성의 활동은 인접 프레임 간에 강력한 상호 연관성이 있다는 성질을 기존의 MAP 기준과 음성 스펙트럼 변이는 이전 프레임 동안 예측된 평균 long-term 파워 스펙트럼과 현재 파워 스펙트럼의 차이를 우도비 테스트에 음성검출 결정법을 제안하였다. 제안된 알고리즘의 성능 평가를 위해  $P_e$ ,  $P_m$  그리고  $P_{fa}$  값을 표 1에서 비교하였고 그림 1에서는 입력 음성 파형에 따른 문턱값 변화를 기존의 방법과 비교하였다. 표 1에서는 제안된 음성 검출 방법이 우수

하다는 것을 수치로 알 수 있으며 특히 비정상 잡음에서 성능이 향상이 크다는 것을 확인할 수 있었고 SNR이 낮을수록 성능 향상이 높아진다는 것을 보았다. 그림 1에서는 제안한 음성 검출 방법의 문턱값이 음성의 시작과 끝부분에서 보다 효율적으로 변한다는 것을 보았다. 모든 결과에서 알 수 있듯이 제안된 음성 검출 방법이 우수하다는 것을 알 수 있다.

## 감사의 글

본 연구는 “지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (NIPA-2011-C1090-1121-0007).” 또한 본 연구는 지식경제부 및 한국산업기술평가관리원의 IT핵심기술개발사업의 일환으로 수행하였음 [KI001824, 장애인 및 고령자를 위한 Digital Guardian 기술개발].

## 참고 문헌

1. K. Srinivasant and Allen Gersho, "Voice activity detection for cellular networks," *Proc. IEEE Speech Coding Workshop*, pp. 85-86, Oct. 1993.
2. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoustics, Speech, Sig. Process.*, vol. ASSP-32, no. 6, pp. 1190-1121, Dec. 1984.
3. J. Sohn and W. Sung, "A voice activity detector employing soft decision based noise spectrum adaptation," *Proc. Int. Conf. Acoustics, Speech, and Sig. Process.*, vol. 1, pp. 365-368, May 1998.
4. J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *IEEE Sig. Process. Lett.*, vol. 6, no. 1, pp. 1-3, Jan. 1999.
5. Y. D. Cho and A. Kondoz, "Analysis and improvement of a statistical model-based voice activity detector," *IEEE Sig. Process. Lett.*, vol. 8, no.10, pp. 276-278, Oct. 2001.
6. J.-H. Chang, J. W. Shin, and N. S. Kim, "Voice activity detector employing generalised gaussian distribution," *Electron. Lett.*, vol. 40, no. 24, pp. 1561-1563, Nov. 2004.
7. J.-H. Chang, N. S. Kim, and S. K. Mitra, "Voice activity detection based on multiple statistical models," *IEEE Trans. Sig. Process.*, vol. 54, no. 6, pp. 1965-1976, June 2006.
8. Y. C. Lee and S. S. Ahn, "Statistical model-based VAD algorithm with Wavelet Transform," *IEICE Trans. Fundamentals*, vol. E89-A, no. 6, pp. 1594-1600, June 2006.
9. J. Ramirez, J. M. Gorrioz, J. C. Segura, C. G. Puntonet, and A. J. Rubio, "Speech/non-speech discrimination based on contextual information integrated bispectrum LRT," *IEEE Sig. Process. Lett.*, vol. 13, no. 8, pp. 497-500, Aug. 2006.
10. J. W. Shin, H. J. Kwon, S. H. Jin and N. S. Kim, "Voice activity detection based on conditional MAP criterion," *IEEE Signal Processing Letters*, vol. 15, pp. 257-260, Feb. 2008.
11. 3GPP2 Spec., "Enhanced Variable Rate Codec (EVRC)," 3Gpp2-C.S0014-0, vol. 1.0 Apr. 2004.

---

## 저자 약력

---

• 김 상 균 (Sang-Kyun Kim)



2008년 2월: 인하대학교 전자공학과 학사  
2011년 2월: 인하대학교 전자공학부 석사  
2011년 9월 ~ 현재: 한양대학교 전자통신컴퓨터  
공학부 박사과정

• 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: (주)넷더스 연구소장  
2004년 5월 ~ 2005년 4월: 캘리포니아 주립대학,  
산타바바라 (UCSB) 박사후연구원  
2005년 5월 ~ 2005년 8월: 한국과학기술연구원  
(KIST) 연구원  
2005년 9월 ~ 2011년 2월: 인하대학교 전자전기공  
학부 조교수  
2011년 3월 ~ 현재: 한양대학교 융합전자공학부 부  
교수