

궤환구조를 가지는 변별적 가중치 학습에 기반한 음성검출기

Voice Activity Detection Based on Discriminative Weight Training with Feedback

강 상 익*, 장 준 혁*
(Sang-Ick Kang*, Joon-Hyuk Chang*)

*인하대학교 전자공학부

(접수일자: 2008년 8월 25일; 수정일자: 2008년 10월 8일; 채택일자: 2008년 10월 20일)

이동통신에서 배경잡음이 존재하는 실제 환경에서 음성신호처리의 가장 중요한 이슈중의 하나는 강인한 음성검출기를 설계하는 것이다. 상대적으로 간단하면서도 성능이 우수하여 대표적인 음성검출기로 사용되는 통계적모델기반 기법은 각 주파수 채널별 우도비를 이용하여 음성검출 검출식을 만들어내는 방식이다. 최근, 변별적 가중치 학습 (discriminative weight training)을 이용하여 주파수 채널별 가중치가 인가된 우도비를 이용한 음성검출 결정식을 갖는 음성검출기가 제안 되었으며 상대적으로 우수한 성능을 보였다. 본 연구에서는 기존의 변별적 가중치 학습의 입력벡터에 이전프레임의 결정식을 궤환구조형태를 바탕으로 추가하는 새로운 방식을 제안한다. 제안된 기법은 비정상 (non-stationary) 잡음 환경에서 객관적인 방법을 통해 상호비교 분석되었으며 결론적으로 우수한 성능을 보였다.

핵심용어: 음성 검출기, minimum classification error, 통계적 모델, 우도비

투고분야: 음성처리 분야 (2)

One of the key issues in practical speech processing is to achieve robust Voice Activity Detection (VAD) against the background noise. Most of the statistical model-based approaches have tried to employ equally weighted likelihood ratios (LRs), which, however, deviates from the real observation. Furthermore voice activities in the adjacent frames have strong correlation. In other words, the current frame is highly correlated with previous frame. In this paper, we propose the effective VAD approach based on a minimum classification error (MCE) method which is different from the previous works in that different weights are assigned to both the likelihood ratio on the current frame and the decision statistics of the previous frame.

Keywords: Voice activity detection, minimum classification error, Statistical model, Likelihood ratio

ASK subject classification: Speech Signal Processing (2)

I. 서론

음성과 비음성 구간을 검출하는 음성 검출기 (voice activity detector, VAD)는 음성 부호화, 음성인식 그리고 음향학적 반향제거기 등과 같은 다양한 음성 통신 시스템에 적용된다. 특히, 음성검출기는 다중 접속 기술에서 한정된 주파수 대역을 효율적으로 사용하기 위한 가변 전송률 부호화기의 실현을 위해 필수적인 부분을 차지하고 있다. 그러한 이유로 여러 가지 음성 검출기가 제안

되었으며 대표적으로 선형예측 부호화 (linear prediction coding, LPC) 파라미터 [1], 에너지레벨, 포먼트 (formant) 모양 [2], 영교차율 (zero crossing rate, ZCR) [3], 켈스트럴 피쳐 (cepstral feature) [4] 등이 있으며 이와 관련하여 다양한 형태의 알고리즘이 제안되고 있다. 많은 알고리즘 중에 Ephraim과 Malah의 연구에서 시작된 minimum mean square error (MMSE) 기반의 음성 향상 기법에 사용된 음성의 존재와 부재에 대한 통계적 모델을 음성 검출기에 적용한 것이 매우 우수한 성능을 가진 것으로 알려져 있는데 [5-12], 구체적으로 음성에 대한 가우시안 통계모델을 decision-directed (DD) 기법에 도입하여 최적의 음성검출 파라미터 추정에 사용하여 나온 음

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

성의 존재와 부재에 대한 우도비 (likelihood ratio, LR) 를 기하 평균한 결정식으로 음성검출여부를 최종적으로 판단한다 [7].

최근에 음성검출기의 성능을 높이기 위한 일환으로 각 주파수 채널별 우도비의 단순한 기하 평균을 이용하여 문턱값을 비교하는 방법 대신, 변별적 가중치 학습 (discriminative weight training)을 위한 minimum classification error (MCE) 방법을 이용하여 도출된 최적화된 가중치를 각 주파수 채널별 우도비에 적용하여 기하 평균을 구성하는 방법이 제안되었다 [13]. 즉, 각 주파수채널별 우도비가 음성검출 결정식에 미치는 영향이 다르다는 점에서 착안한 것이다. 본 논문에서는 현재 프레임의 음성의 존재와 부재가 이전 프레임에 많은 영향을 받는다는 것을 고려하여 이전 프레임에서 최적화된 가중치가 적용되어 구해진 음성검출 결정식을 현재 프레임에서의 결정식에 피드백시켜 변별적 가중치 학습에 기반하여 최종적인 음성검출식을 구한다. 제안된 알고리즘은 객관적인 평가방법을 통해 상호비교 되었으며 기존의 알고리즘에 비해 우수한 성능을 보였다.

II. 통계모델 기반의 음성 검출기의 이해

시간축 상에서 원래의 음성신호 $x(t)$ 에 잡음신호 $n(t)$ 가 인가된 입력신호 $y(t)$ 를 discrete Fourier transform (DFT)을 통해 주파수 축으로 변환되어 아래와 같이 표현한다.

$$\mathbf{Y}(t) = \mathbf{X}(t) + \mathbf{N}(t) \quad (1)$$

여기서 $\mathbf{Y}(t) = [Y_1, Y_2, \dots, Y_M]$, $\mathbf{X}(t) = [X_1, X_2, \dots, X_M]$, 그리고 $\mathbf{N}(t) = [N_1, N_2, \dots, N_M]$ 는 각각 잡음에 오염된 음성신호, 원래의 음성신호, 잡음신호의 DFT 계수 벡터를 나타낸다. 주어진 가설 H_0, H_1 이 각각 음성의 부재와 존재를 표현한다고 하면 각 주파수 채널별로 다음과 같이 기술된다.

$$H_0: \text{speech absent} : Y_k(t) = N_k(t) \quad (2)$$

$$H_1: \text{speech present} : Y_k(t) = X_k(t) + N_k(t). \quad (3)$$

음성과 잡음신호의 스펙트럼이 복소 가우시안 분포를 따른다는 가정으로부터 가설 H_0 와 H_1 을 조건으로 한 확

률밀도함수는 아래와 같이 주어진다 [5].

$$p(Y_k|H_0) = \frac{1}{\pi\lambda_{d,k}} \exp\left\{-\frac{|Y_k|^2}{\lambda_{d,k}}\right\} \quad (4)$$

$$p(Y_k|H_1) = \frac{1}{\pi[\lambda_{d,k} + \lambda_{x,k}]} \exp\left\{-\frac{|Y_k|^2}{\lambda_{d,k} + \lambda_{x,k}}\right\} \quad (5)$$

여기서 $\lambda_{x,k}$ 와 $\lambda_{d,k}$ 는 각각 채널별 음성과 잡음의 분산이며, 이 때 k 번째 주파수 밴드에 대한 우도비는 아래와 같이 구한다.

$$A_k \equiv \frac{p(Y_k|H_1)}{p(Y_k|H_0)} = \frac{1}{1 + \xi_k} \exp\left\{\frac{\gamma_k \xi_k}{1 + \xi_k}\right\} \quad (6)$$

여기서 $\xi_k = \lambda_{x,k}/\lambda_{d,k}$ 와 $\gamma_k = Y_k/\lambda_{d,k}$ 는 각각 *a priori* signal-to-noise ratio (SNR)와 *a posteriori* SNR이다 [5]. 음성 부재 구간에서 갱신되는 잡음 신호로부터 구한 잡음 분산 $\lambda_{d,k}$ 를 이용하여 *a posteriori* SNR γ_k 를 추정하며, 또한 *a priori* SNR ξ_k 는 decision-directed (DD) 방식을 이용하여 아래와 같이 추정한다 [5].

$$\hat{\xi}_k(t) = \alpha \frac{|\hat{X}_k(t-1)|^2}{\lambda_{d,k}(t-1)} + (1-\alpha)P[\gamma_k(t) - 1] \quad (7)$$

여기서 $|\hat{X}_k(t-1)|^2$ 은 이전 프레임에서 추정된 음성 신호의 k 번째 스펙트럼 성분의 크기에 대한 추정치이며, MMSE에 기반하여 구한다 [7]. 또한 α 는 가중치 값이며, 연산자 $P[\cdot]$ 은 아래와 같이 정의된다.

$$P[x] = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

기존의 일반적인 통계적 모델 기반의 음성 검출기에 대한 결정식은 각각의 주파수 채널에서 구해진 우도비를 기하 평균하여 아래와 같이 음성 검출 여부를 판단한다 [6-12].

$$\log \Lambda(t) = \frac{1}{M} \sum_{k=1}^M \log A_k \begin{matrix} > \\ < \\ < \end{matrix} \begin{matrix} H_1 \\ \eta \\ H_0 \end{matrix} \quad (9)$$

여기서 M 은 전체 주파수 대역의 개수이며, η 는 음성 검출 문턱값이다.

III. 제안된 궤환 구조를 가지는 변별적 가중치 학습에 기반한 음성검출기

[13]의 논문에 따르면 기존의 음성의 통계모델에 기반한 음성검출기는 식 (9)에서 보듯이 각 주파수별 성분이 독립이라는 가정에 기반하여 간단한 우도비의 기하 평균식을 이용한 점을 살펴볼 수 있다. 그러나, 각 주파수 채널별 우도비가 음성검출 성능에 균일한 기여를 한다는 것은 음성신호의 주파수특성의 분포 등을 고려하면 실제적이지 않으므로 각 주파수 채널별 우도비인 $\log A_k$ 에 변별적 가중치 학습에 의해 구해진 최적화된 가중치를 적용하여 최종적인 음성검출식을 도출하였다. 한편, 최적화된 가중치를 가지는 [13]의 음성검출식은 음성검출의 프레임 간 강한 상관성을 고려하고 있지는 않는데, 이것은 실제적이라고 보기 어렵다. 본 논문에서는 각 우도비 $\log A_k$ 와 이전 프레임에서 구해진 최적화된 가중치 기반의 음성검출식 전체에 대해 변별적 가중치 학습을 실시하여 최종적으로 가중치를 인가함으로써 보다 효과적인 음성검출기를 제안한다. 그림 1은 제안된 궤환 구조를 가지는 변별적 가중치 학습 알고리즘의 블록도를 보여주는데 이러한 방식은 최적화된 가중치가 현재 프레임에서의 우도비들과 이전 프레임의 궤환 결정식에 적용됨으로서 이전 프레임과의 상관성을 고려하여 성능을 향상시킬 수 있는 장점이 있다. 구체적으로는 아래와 같이 결정식을 제안하는데, [13]과는 이전 프레임의 결정식이 현재 프레임의 결정식에 가중치가 곱해서 들어가 있는 모양이 다르다.

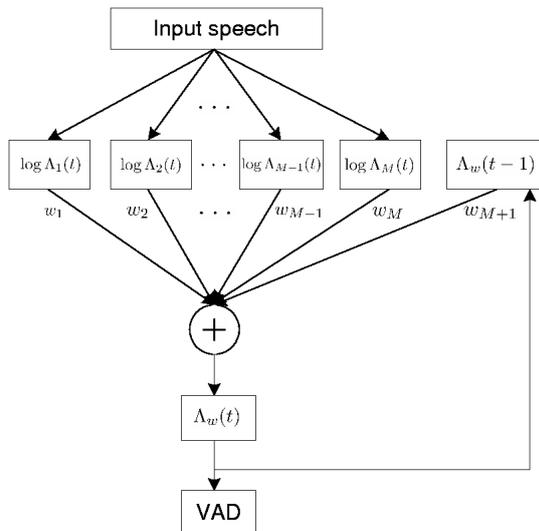


그림 1. 제안된 궤환 구조를 가지는 변별적 가중치 학습 블록도
Fig. 1. Block diagram of the proposed VAD algorithm based on discriminative weight training with feedback.

$$A_w(t) = \frac{1}{M+1} \left(\sum_{k=1}^M w_k \log A_k(t) + w_{M+1} A_w(t-1) \right) \begin{matrix} > \\ < \end{matrix} \eta \begin{matrix} H_1 \\ H_0 \end{matrix} \quad (10)$$

먼저, 입력 신호로부터 구한 각 주파수 채널별 우도비와 이전 프레임 결과에 각각 다른 가중치 w_k 를 적용하여 새로운 특징 벡터를 $\mathbf{A}_w = \{w_1 \log A_1(t), w_2 \log A_2(t), \dots, w_M \log A_M(t), w_{M+1} A_w(t-1)\}$ 라 정의하며 각 가중치는 다음의 조건을 만족한다.

$$\sum_{k=1}^{M+1} w_k = 1, \quad w_k \geq 0 \quad (11)$$

여기서, 가중치를 적용한 각각의 특징 벡터들의 기하 평균값 A_w 를 $\frac{1}{M+1} (\sum_{k=1}^M w_k \log A_k(t) + A_w(t-1))$ 와 같이 정의하고, 훈련할 데이터의 각각의 프레임에서 음성 $g_s(\cdot)$ 와 비음성 $g_n(\cdot)$ 을 구분하는 두 개의 함수를 다음과 같이 정의한다.

$$g_s(\mathbf{A}_w) = A_w - \theta \quad (12)$$

$$g_n(\mathbf{A}_w) = \theta - A_w \quad (13)$$

여기서 θ 는 음성과 비음성을 구분하는 문턱값이다. 이때 문턱값은 음성과 비음성 훈련 데이터의 분포에서 겹치는 경계값을 사용하였다. 제안된 연구에서는 최적화 알고리즘에 기반한 가중치를 구하기 위해 generalized probabilistic descent (GPD) 기반의 MCE 훈련을 적용하며 [14], 실제로 훈련 데이터 프레임 $\mathbf{A}_w(t)$ 의 분류 오류 D 를 다음과 같이 정의한다.

$$D(\mathbf{A}_w(t)) = \begin{cases} -g_s(\mathbf{A}_w(t)) + g_n(\mathbf{A}_w(t)) & \text{if } g_s \text{ is true class} \\ -g_n(\mathbf{A}_w(t)) + g_s(\mathbf{A}_w(t)) & \text{if } g_n \text{ is true class} \end{cases} \quad (14)$$

여기서 식 (14)이 음수인 값을 가질 때 올바른 분류가 되며 이를 기반으로 하는 손실함수 (loss function) L 은 다음과 같이 sigmoid 함수 형태로 정의된다.

$$L = \frac{1}{1 + \exp(-\beta D(\mathbf{A}_w))}, \quad \beta > 0 \quad (15)$$

여기서 β 는 sigmoid 함수의 기울기를 나타낸다. 최적화된 가중치를 구하기 위해선 손실함수가 최소가 되어야 한다. MCE 훈련과정을 통해 가중치를 조정하는 과정에

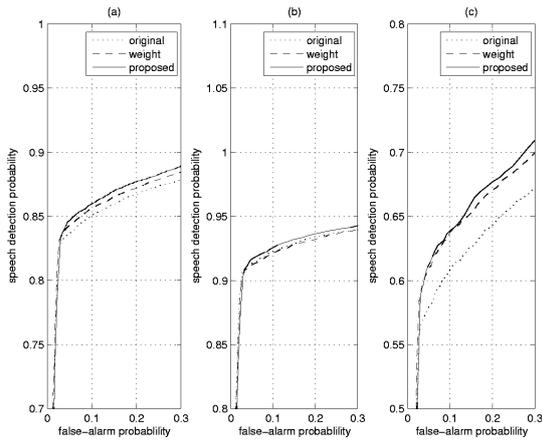


그림 2. Car 잡음 5 dB에서의 ROC 곡선
 Fig. 2. ROC curves for car noise (SNR = 5 dB).
 (a) speech (b) voiced sound (c) unvoiced sound

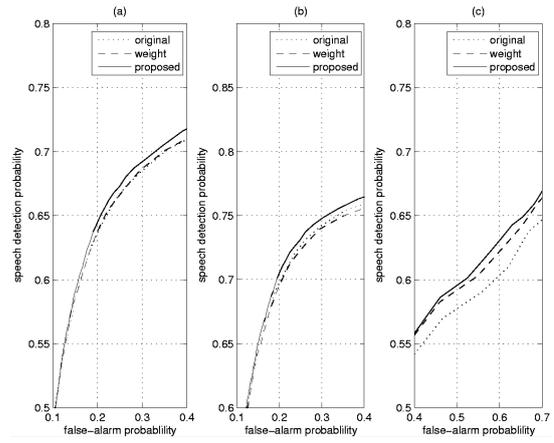


그림 4. Office 잡음 5 dB에서의 ROC 곡선
 Fig. 4. ROC curves for office noise (SNR = 5 dB).
 (a) speech (b) voiced sound (c) unvoiced sound

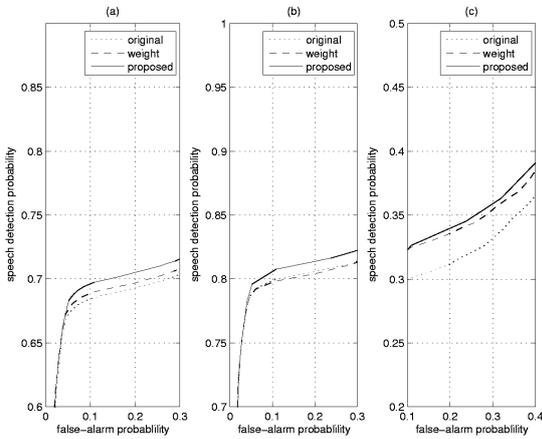


그림 3. Street 잡음 5 dB에서의 ROC 곡선
 Fig. 3. ROC curves for street noise (SNR = 5 dB).
 (a) speech (b) voiced sound (c) unvoiced sound

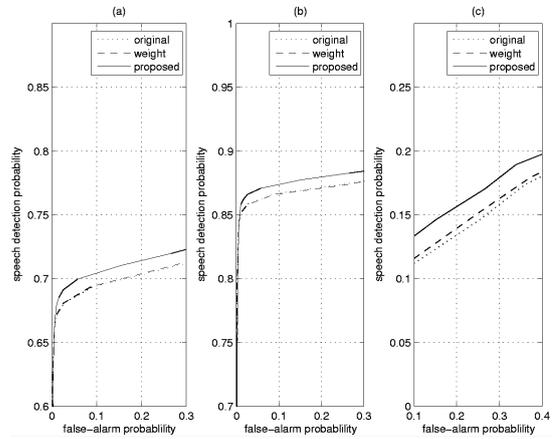


그림 5. White 잡음 5 dB에서의 ROC 곡선
 Fig. 5. ROC curves for white noise (SNR = 5 dB).
 (a) speech (b) voiced sound (c) unvoiced sound

서 식 (10)과 같은 제약조건 때문에 가중치 w 를 \tilde{w} 로 변환한다.

$$\tilde{\mathbf{w}} = \{\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_{M+1}\} \quad (16)$$

$$\tilde{w}_k = \log w_k. \quad (17)$$

가중치 \tilde{w}_k 는 매 프레임마다 연속적으로 존재하는데, 각 주파수 가중치는 다음과 같은 식으로 갱신된다 [13], [15].

$$\tilde{w}_k(n+1) = \tilde{w}_k(n) - \epsilon \frac{\partial L}{\partial w_k} \Big|_{w_k = \tilde{w}_k(n)} \quad (18)$$

여기서 $\epsilon (> 0)$ 는 단조롭게 감소하는 구간의 크기이고, n 은 학습과정에서의 시간단위이다. \tilde{w}_k 를 갱신한 후에 다음의 식과 같이 w_k 로 복원된다.

$$w_k = \frac{\exp(\tilde{w}_k)}{\sum_{i=1}^M \exp(\tilde{w}_i)} \quad (19)$$

식 (19)에서 정규화 된 가중치를 사용했을 때 식 (11)를 만족한다.

기존의 통계적 모델 기반의 음성 검출기의 결정식인 식 (9)와 비교하여, 본 논문에서는 각각의 주파수 채널에 곱해서 구해진 우도비와 이전 프레임의 결과 값을 위에서 제시된 MCE 훈련방법을 이용해 구한 식 (19)의 가중치와 기하 평균하여 (10)과 같이 최종적으로 음성 검출 여부를 판단한다. 결론적으로 음성이 인접한 프레임에서 강한 연관성을 가지고 있다는 것을 고려하여 현재 프레임의 각 우도비 외에 이전 프레임에서의 음성검출 결정식을 추가하여 최적화된 가중치를 적용하는 점이 다르다 [13].

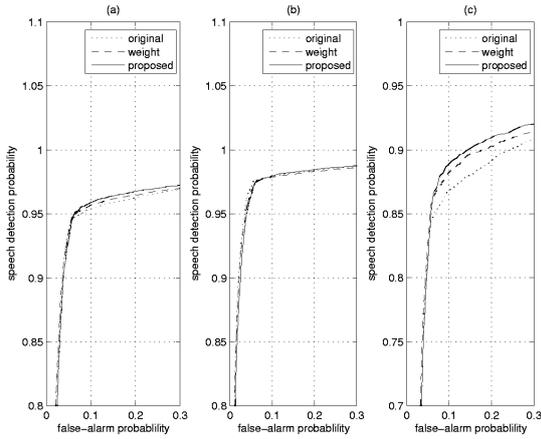


그림 6. Car 잡음 15 dB에서의 ROC 곡선
 Fig. 6. ROC curves for car noise (SNR = 15 dB)
 (a) speech (b) voiced sound (c) unvoiced sound

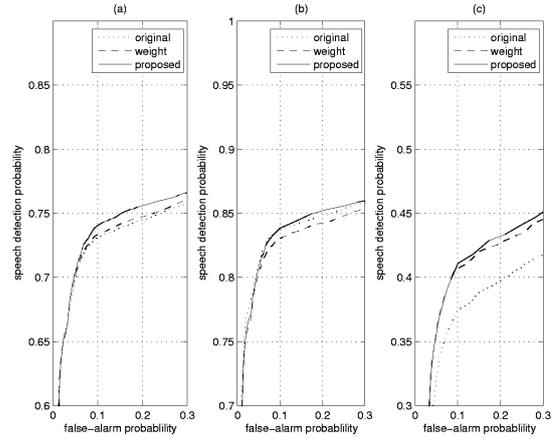


그림 8. Office 잡음 15 dB에서의 ROC 곡선
 Fig. 8. ROC curves for office noise (SNR = 15 dB).
 (a) speech (b) voiced sound (c) unvoiced sound

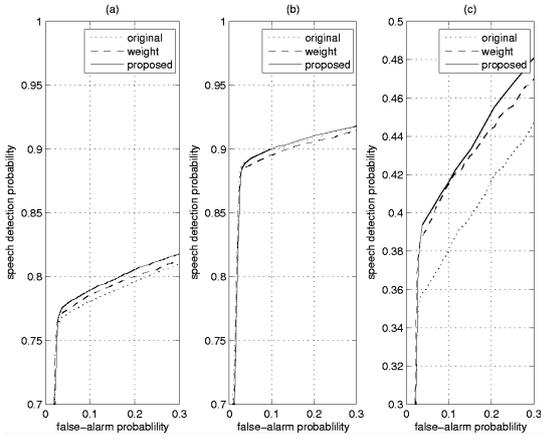


그림 7. Street 잡음 15 dB에서의 ROC 곡선
 Fig. 7. ROC curves for street noise (SNR = 15 dB).
 (a) speech (b) voiced sound (c) unvoiced sound

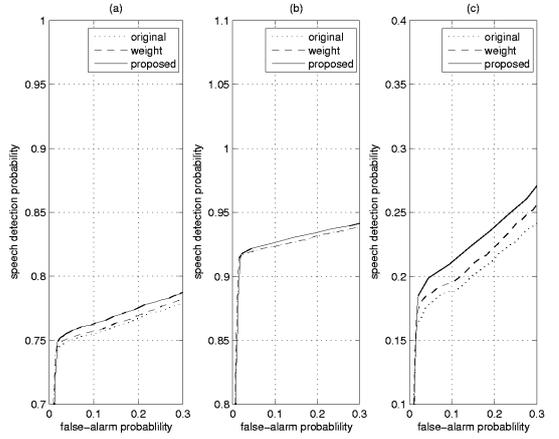


그림 9. White 잡음 15 dB에서의 ROC 곡선
 Fig. 9. ROC curves for white noise (SNR = 15 dB).
 (a) speech (b) voiced sound (c) unvoiced sound

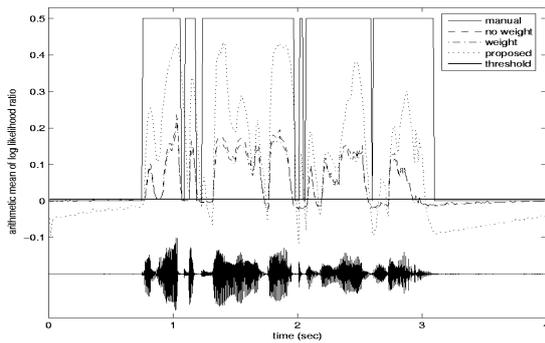


그림 10. white 잡음 5 dB SNR에서 각 음성 검출기의 출력 결과 비교
 Fig. 10. Comparison of VAD output for white noise (SNR = 5 dB).

V. 실험 결과 비교 및 분석

본 논문에서 제안된 MCE 기법을 이용한 최적화된 가중치 기반의 음성 검출기의 성능을 평가하기 위해 기존의

표 1. 다양한 잡음환경과 SNR에서 제안된 새로운 통계모델기반 음성검출기와 기존의 알고리즘들의 P_d 와 P_{fa} 성능 비교
 Table 1. Compare the proposed method with the conventional methods for various noise environments at various SNR.

		Proposed		Weight		No weight	
Noise	SNR	P_d (%)	P_{fa} (%)	P_d (%)	P_{fa} (%)	P_d (%)	P_{fa} (%)
Car	5 dB	88.33	25.61	87.96	25.61	87.47	25.69
	15 dB	97.01	25.06	96.79	25.20	96.64	25.11
Street	5 dB	71.00	25.41	70.27	25.77	69.81	25.42
	15 dB	81.25	25.51	80.68	25.57	80.44	25.75
Office	5 dB	68.05	26.37	67.31	26.38	67.23	26.35
	15 dB	76.13	25.46	75.33	25.00	75.34	25.89
White	5 dB	71.98	26.63	70.92	25.85	70.77	25.28
	15 dB	78.22	26.0	77.81	26.53	77.43	26.13

우도비 테스트를 이용한 통계적 모델 기반의 음성 검출기의 성능과 Receiver Operating Characteristic (ROC) 곡

선을 이용하여 비교하였다 [7]. 최적화된 가중치를 도출하기 위해 손실함수 L 에서 정의된 기울기 파라미터 $\beta=1$ 로 결정하였다. 실험에 사용된 데이터는 성능 평가를 위해 총 230초의 깨끗한 음성 데이터에 음성과 비음성 부분을 10 ms마다 수동으로 표시하였다. 분류된 음성 데이터의 음성 구간은 총 57.1%로 유성음 44.0%, 무성음 13.1%로 구성되었으며 잡음 환경은 음성 데이터에 street, car, office, white 잡음이 각각 5, 15 dB SNR로 부과되었다. 그림 2~9의 (a)는 각각 street, car, office, white 잡음 환경에 따라 음성 검출기의 문턱값을 변경하면서 실제 음성을 음성이라고 판단한 음성 검출 확률 (Pd)과 비음성에 대해 음성이라고 판단한 오경보 확률 (Pf)을 5, 15 dB SNR에서 측정한 ROC 곡선이며, (b)와 (c)는 (a)와 동일한 SNR과 동일한 잡음 환경에서 각각 유성음과 무성음에서의 ROC 곡선이다. 그림 10은 제한한 음성의 시간연속성을 고려한 음성검출 방법의 성능을 효과적으로 검증하기 위해서 white 잡음 5 dB SNR에서 각 음성 검출기의 출력 결과 비교하였다. 비교의 편의를 위해 원래의 음성신호의 파형에 비교하여 도시하였다. 구체적인 성능 평가를 위해 그림 10에 도시된 실제 문턱값을 대입하여 제안된 음성검출기의 성능을 구하고, 기존의 VAD 알고리즘 결과들과 비교 분석하였다. 결과는 표 1에 나타나 있다. 실험 분석 결과 동일한 SNR의 주어진 잡음 조건에서 이전 프레임의 결과 값을 귀환시켜 가중치를 적용한 음성 검출기의 경우 전체적으로 기존의 음성 검출기와 각 채널별 우도비에 가중치를 적용시킨 음성 검출기보다 향상된 성능을 보여준다. 결론적으로 기존의 두 방법 보다 이전 프레임의 결과값을 귀환하여 이용한 음성 검출 방법이 우수한 것을 확인할 수 있다.

VI. 결론

본 논문에서는 음성의 존재와 부재에 대한 통계적 모델에 기반한 각 주파수 채널별 우도비를 단순히 기하 평균을 취하여 문턱값과 비교하는 기존의 방법 대신, MCE를 이용하여 도출한 최적화된 가중치를 각 주파수 채널별 우도비와 이전 프레임의 음성검출 결정식까지 고려하여 최종적인 음성검출식을 만들어내는 새로운 방법으로 기존의 방식보다 향상된 통계모델기반의 음성 검출기를 제시하였으며, 객관적인 실험 결과로부터 제안된 음성 검출기의 성능이 우수함을 알 수 있었다.

감사의 글

본 연구는 지식경제부와 한국산업기술재단의 전략기술인력양성사업과 지식경제부 출연금으로 ETRI, SoC산업진흥센터에서 수행한 IT SoC 핵심설계인력양성사업의 연구결과로 수행되었음

참고 문헌

1. L. R. Rabiner and M. R. Sambur, "Voiced-unvoiced-silence detection using Itkura LPC distance measure," Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, 323-326, May 1977.
2. J. D. Hoyt and H. Wechsler, "Detection of human speech in structured noise," Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, 237-240, May 1994.
3. J. C. Junqua, B. Reaves, and B. Mark, "A study of endpoint detection algorithms in adverse conditions: Incidence on a DTW and HMM recognize," Proc. Eurospeech, 1371-1374, 1991.
4. J. A. Haigh and J. S. Mason, "Robust voice activity detection using cepstral feature," Proc. IEEE TELCON, 321-324, China, 1993.
5. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoustics, Speech, Sig. Process., **ASSP-32**(6), 1190-1121, Dec. 1984.
6. J. Sohn and W. Sung, "A voice activity detector employing soft decision based noise spectrum adaptation," Proc. Int. Conf. Acoustics, Speech, and Sig. Process., **1**, 365-368, May 1998.
7. J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," IEEE Sig. Process. Lett., **6**(1), 1-3, Jan. 1999.
8. Y. D. Cho and A. Kondoz, "Analysis and improvement of a statistical model-based voice activity detector," IEEE Sig. Process. Lett., **8**(10), 276-278, Oct. 2001.
9. J. -H. Chang, J. W. Shin, and N. S. Kim, "Voice activity detector employing generalised gaussian distribution," Electron. Lett., **40**(24), 1561-1563, Nov. 2004.
10. J. -H. Chang, N. S. Kim, and S. K. Mitra, "Voice activity detection based on multiple statistical models," IEEE Trans. Sig. Process., **54**(6), 1965-1976, June 2006.
11. Y. C. Lee and S. S. Ahn, "Statistical model-based VAD algorithm with wavelet Transform," IEICE Trans. Fundamentals., **E89-A**(6), 1594-1600, June 2006.
12. J. Ramirez, J. M. Gorriz, 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., **13**(8), 497-500, Aug. 2006.
13. 강상익, 조규행, 장준혁, 박승섭, "통계적 모델 기반의 음성 검출기를 위한 변별적 가중치 학습," 한국음향학회지, **26**(5), 173-181, 2007.
14. B. -H. Juang, W. Chou, and C. -H. Lee, "Minimum classification error rate methods for speech recognition," IEEE Trans. Speech Audio Processing, **5**(3), 257-265, May 1997.
15. Y. Kida, T. Kawahara, "Voice activity detection based on

optimally weighted combination of multiple feature," Proc. Interspeech, 2621-2624, Sep. 2005.

저자 약력

•강 상 익 (Sang-Ick Kang)



2007년 2월: 인하대학교 전자공학과 학사
2007년 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: 주넷더스 연구소장
2004년 5월~2005년 4월: 캘리포니아 주립대학 산타바바라 (UCSB) 박사후연구원
2005년 5월~2005년 8월: 한국과학기술연구원 (KIST) 연구원
2005년 9월~현재: 인하대학교 전자공학부 조교수