논문 03-28-11C-14 한국통신학회논문지 '03-11 Vol.28 No.11C

음성 향상 전처리와 문턱값 갱신을 적용한 향상된 음성검출 방법

정회원 이 윤 창*, 안 상 식**

An Improved VAD Algorithm Employing Speech Enhancement Preprocessing and Threshold Updating

Yoon-Chang Lee*, Sang-Sik Ahn** *Regular Members*

요 약

본 논문에서는 음성검출의 성능을 향상시킬 목적으로 정합 필터를 이용한 음성향상 전처리 과정을 통하여 SNR을 개선한 후, 이를 LLR(Log Likelihood Ratio) 검사에 의한 최적 결정방법을 적용하여 확률적인 모델을 기준으로 하는 향상된 음성검출 방법을 제안한다. 또한 기존의 음성검출 방법들에서는 제시되지 않았던 문턱값 갱신 알고리즘을 제안하며, 이 방법을 통해서 기존의 방법들에서 성능이 좋지 않았던 낮은 SNR 환경에서도 음성검출을 할 수 있게 되었다. 마지막으로 컴퓨터 시뮬레이션을 통하여 이미 상용화되어 널리 이용중인 G.729B(ITU-T G.729 Annex B)의 음성검출 결과와 비교를 통해서 제안한 음성검출 방법의 성능의 우수성을 검증하며, 실제적인 환경에도 적용이 가능함을 보인다.

Key Words : Voice activity detection, Threshold update, Speech enhancement

ABSTRACT

In this paper, we propose an improved statistical model-based voice activity detection algorithm and threshold update method. We first improve signal-to-noise ratio by using speech enhancement preprocessing algorithm combined power subtraction method and matched filter, then apply it to LLR test optimum decision rule for improving the performance even in low SNR conditions. And we propose an adaptive threshold update method that was not concerned in any papers. We also perform extensive computer simulations to demonstrate the performance improvement of the proposed VAD algorithm employing the proposed speech enhancement preprocessing algorithm and adaptive threshold update method under various background noise environments. Finally we verify our results by comparing ITU-T G.729 Annex B.

I. 서 론

이동통신이나 음성인식 시스템 등의 음성 신호처리 환경에서 잡음이 섞인 신호를 실시간으로 음성 코딩할 목적으로 지난 수년간 음성검출에 대한 많은 연구가 이루어져 왔으며 음성향상 기법이 많이 이용되고 있다. 또한 인터넷 전화와 같은 환경에서는 한정된 자원을 적절히 활용하기 위해서 채널을 통해 전송되는 비트율을 줄이고 채널 대역을 효과적으로 이용하기 위한 가변율 음성 부호화가 응용되고 있으며, 이러한 응용 범위가 점차 증가함에 따라서 음성검출의 필요성이 더욱 커지게 되었다. 음성 부호화의 관점에서 보면 음성신호가 있는 구간

* 고려대학교 전자정보공학과 신호처리 연구실(ychlee@korea.ac.kr) ** 고려대학교 전자및정보공학부(sahn@korea.ac.kr)

논문번호 : 030265-0620, 접수일자 : 2003년 6월 20일

※ 본 연구는 고려대학교 특별연구비에 의하여 수행 되었음.

을 검출하여 이 부분에서만 부호화하면 채널을 통해 전송되는 데이터의 양을 효과적으로 줄일 수 있다. 이와 같은 이유 때문에 음성검출 알고리즘에 대한 연구가 꾸준히 이루어지고 있다.

한편, 음성검출에 필요한 잡음추정 알고리즘은 변화하는 환경에 빠르게 적응할 수 있어야 하며 음성신호의 영향을 덜 받아야 한다. 또한 대부분의 하드웨어가 디지털 방식으로 구현되기 때문에 알고리즘의 계산량도 매우 중요한 문제가 되고있다. 이런 조건들을 만족시키면서 가장 실현 가능한 방법이 잡음 구간에서만 잡음의 파워를 갱신하고 음성이 포함된 구간에서는 갱신을 하지 않는 것이지만 이 방법은 잡음만 존재하는 구간을 찾아내기 위해서 부가적인 음성검출(secondary VAD)이 필요하며 이 과정에 오류가 있을 경우 잡음 추정 오류가 발생되어서 최종적인 음성검출 결과에 많은 영향을 주는 문제점이 있다. 이러한 문제점을 개선하기 위해서 부가적인 음성검출의 도움 없이 잡음을 추정하는 방법이 연구되었으며, 최근에는 확률적인 모델(statistical model)을 이용하여 잡음을 추정하는 연구가 이루어지고 있다 [1][2].

음성검출의 성능은 문턱값에 많은 영향을 받기 때문에 이 값은 변화하는 주변 환경에 대해서 적절한 값을 가지도록 갱신이 이루어져야만 한다. 하지만 지금까지 효과적인 방법이 제안되지 않았다. 따라서 본 논문에서는 낮은 SNR 환경에서도 더 좋은 성능향상을 가져올 수 있는 문턱값 갱신 알고리즘을 제안하고 이를 적용한 개선된 음성검출 방법을 제안한다.

본 논문의 정합필터를 이용한 음성향상 전처리는 음성 부호화에 이용하기 위한 향상된 음성 신호를 찾고자 함이 아니라 음성검출 분석에 이용되는 음성성분이 강화된 신호를 생성하기 위함이다.

본 논문의 구성은 다음과 같다. Ⅱ장에서는 확률적인 모델을 적용한 음성검출 알고리즘들에 대해서 살펴보고, Ⅲ장에서는 정합필터를 이용한 음성향상 방법에 대해서 요약하고, Ⅳ장에서는 기존의 방법들보다 음성검출 성능을 개선할 수 있는 향상된 음성검출 알고리즘과 함께 부가적인 음성검출을 사용하지 않는 개선된 잡음 추정방법을 제안하며, 문턱값을 적응적으로 갱신하는 방법을 제안한다. 그리고 Ⅴ장에서는 제안한 음성검출 방법의 성능을 검증하고, 이미 상용화되어 널리 이용중인 G.729B [3] 음성검출 방법과 성능을 비교해 본다. 마지막으로 Ⅵ장에서 결론을 맺는다.

Ⅱ. 음성검출 알고리즘

깨끗한 음성신호 s_n이 상관 관계가 없는 잡음 v_n에 의해 영향을 받는다고 가정했을 때, 잡음이 섞인 신호 x_n을 기준으로 음성검출을 위한 다음의 두 가지의 가설을 가정한다.

H_0 : 음성 신호가 존재하지 않음 : $x_n = v_n$

H_1 : 음성 신호가 존재함 : $x_n = s_n + v_n$

주파수 영역에서의 신호처리 기법의 장점을 이용하기 위하여 M 포인트 DFT 처리를 한 후, 에너지에 기반을 둔 전통적인 음성검출 방법의 단점을 개선하기 위하여 Yang이 제안한 음성검출 방법은 다음과 같다 [4].

$$\sum_{k=0}^{M-1} \frac{|\hat{S}_k|^2}{|\hat{V}_k|^2} \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} \eta. \tag{1}$$

음성신호의 스펙트럼은 $|\hat{S}_k|^2 = |X_k|^2 - |\hat{V}_k|^2$과 같이 maximum likelihood 기법을 이용해 추정되며 S_k, V_k, X_k는 각각 음성신호와 부가적인 잡음 그리고 잡음이 섞인 음성 신호의 DFT를 나타낸다. 또한 ^은 추정된 신호를 의미하며, k는 주파수 공간에서의 DFT 지수를 나타낸다. 이렇게 에너지를 기준으로 하는 음성검출 방법은 낮은 SNR 환경에서 음성 신호가 잡음에 묻히게 되어 음성검출의 성능을 떨어트리는 단점이 있으며, Yang이 제안한 방법은 SNR의 동적 영역이 작아서 음성검출의 성능이 문턱값 η에 많은 영향을 받는 단점이 있다.

잡음의 추정값 $\hat{V}$은 잡음만 존재하는 구간에서 아래의 식과 같이 망각계수 μ를 이용한 전통적인 잡음추정 방법을 이용하여 추정되어진다 [5].

$$|\hat{V}_k(n)| = \mu |\hat{V}_k(n-1)| + (1-\mu)|X_k(n)| \quad , 0 \le \mu \le 1 \tag{2}$$

n은 프레임 지수를 나타낸다. 이러한 부가적인 음성검출의 도움을 받는 잡음추정 방법은 잡음의 특성이 느리게 변화하는 환경에서는 좋은 성능을

보이지만 잡음의 변화가 빠르거나 SNR이 낮은 환경에서는 성능이 떨어지는 문제점이 있다. 이런 문제점을 개선하기 위해서 최근에는 확률적인 모델을 이용하여 모든 프레임에서 잡음의 특성을 갱신할 수 있도록 하는 방법들이 제안되었다.

Sohn *et al.* 은 확률적인 모델을 적용하여 LLR (Log Likelihood Ratio) 검사방법을 이용한 음성검출 방법과 잡음 추정 방법을 아래와 같이 제안하였다 [1].

$$\Lambda = \frac{1}{M}\sum_{k=0}^{M-1}\left\{\frac{|X_k|^2}{\widehat{\lambda_V}(k)} - \log\frac{|X_k|^2}{\widehat{\lambda_V}(k)} - 1\right\} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \quad (3)$$

$$\begin{aligned}\widehat{\lambda_V}(k) &= E[\lambda_V(k)|X_k] \\ &= E[\lambda_V(k)|H_0]P(H_0|X_k) \\ &\quad + E[\lambda_V(k)|H_1]P(H_1|X_k) \qquad (4) \\ &= \frac{1}{1+\varepsilon\Lambda}E[\lambda_V(k)|H_0] \\ &\quad + \frac{\varepsilon\Lambda}{1+\varepsilon\Lambda}E[\lambda_V(k)|H_1],\end{aligned}$$

$\varepsilon = P(H_1)/P(H_0)$, $P(H_0|X_k) = 1/\{1+\varepsilon\Lambda(k)\}$, $P(H_1|X_k) = \varepsilon\Lambda(k)/\{1+\varepsilon\Lambda(k)\}$이며, 주파수 공간에서의 LLR 값은 $\Lambda(k) = P(X_k|H_1)/P(X_k|H_0)$과 같이 정의된다. 프레임 지수 n을 이용하여 정리하면,

$$\begin{aligned}\hat{\lambda}_V^{(n)}(k) &= \frac{1}{1+\varepsilon\Lambda^{(n)}}|X_k^{(n)}(k)|^2 \\ &\quad + \frac{\varepsilon\Lambda^{(n)}}{1+\varepsilon\Lambda^{(n)}}\hat{\lambda}_V^{(n-1)}(k).\end{aligned} \qquad (5)$$

이 잡음 추정 방법은 음성신호가 포함되지 않은 프레임에서만 잡음을 갱신하여 변화가 심한 잡음을 올바르게 추정해 내지 못하는 문제점을 개선하여 부가적인 음성검출의 도움이 필요 없도록 하였다. 그렇지만 잡음 변화율이 매우 크거나 Gaussian 잡음 환경이 아닐 경우 성능이 떨어지는 단점이 있다. 또한 음성 구간이면서 Λ는 음성 신호의 시작 부분과 끝 부분에서는 현재 프레임에 대한 갱신 비율이 커지게 되어서 잡음 추정값에 음성 신호 성분이 많이 포함되는 문제점이 있다.

Cho와 Kondoz는 LLR 검사를 이용할 경우 음성신호의 끝부분이 잡음으로 잘못 결정 내려지는 문제점을 개선하기 위하여 SLR(Smoothed likelihood ratio) 검사를 아래와 같이 새롭게 정의하여 음성검출에 이용하였으며, α는 망각 계수를 나타낸다 [2].

$$\Psi_k^{(n)} = \exp\{\alpha\log\Psi_k^{(n-1)} + (1-\alpha)\log\Lambda_k^{(n)}\}. \quad (6)$$

그리고 주파수 공간에서 스펙트럼의 평균을 이용한 잡음 추정 알고리즘을 제안하였으며, μ는 α와 마찬가지로 망각계수를 나타낸다.

$$\lambda_V^{(n)} = \mu\lambda_V^{(n-1)}(k) + (1-\mu)E[|V_k^{(n)}|^2|X_k^{(n)}]. \quad (7)$$

따라서 잡음의 주파수 스펙트럼은 다음과 같은 방법으로 추정되어질 수 있다.

$$\begin{aligned}E[|V_k^{(n)}|^2|X_k^{(n)}] &= P(H_0|X_k^{(n)})|X_k^{(n)}|^2 \\ &\quad + \{1-P(H_0|X_k^{(n)})\}\lambda_V^{(n-1)}(k).\end{aligned} \qquad (8)$$

이 방법은 음성신호의 시작과 끝부분의 음성 신호의 크기가 작은 부분에서의 검출 에러를 줄인 장점이 있지만 계산량이 많으며, 두 개의 망각 계수를 이용하기 때문에 선택하는 망각계수의 값에 따라서 음성검출 결과에 많은 차이가 있게 되며, 빠르게 변화하는 잡음 환경에서 성능이 저하되는 단점이 있다. 또한 음성신호의 시작부분과 끝 부분의 검출 에러가 줄어드는 대신 기존의 LLR 검사방법에 비해서 잡음 구간을 음성 구간으로 잘못 결정 내리는 false-alarm 확률이 증가하는 단점이 있다.

Lee와 Ahn은 전통적인 잡음 추정방법을 이용하면서도 많은 계산량의 추가 없이 음성검출 성능을 향상시킬 수 있는 음성향상 전처리 과정을 이용한 새로운 음성검출 방법을 제안하였다 [6][7]. 그러나 이 방법은 잡음을 추정하기 위해서 부가적인 음성검출의 도움을 받기 때문에 이의 신뢰도에 따라 최종 음성검출의 성능에 많은 영향을 주어서 변화가 심한 잡음 환경에서는 잡음 추정 오류가 심하여 음성검출의 성능을 떨어트리며, 문턱값이 갱신되지 않기 때문에 실제적인 환경에 적용하기 어려운 문제점이 있다.

앞에서 언급한 문제점들을 개선하기 위해서 본 논문에서는 LLR 검사에 의한 최적 결정 방법을 이용하면서도 추가적인 계산량도 적고, 망각계수 및 문턱값이 변화하는 환경에 적응하여 갱신되는 방법

을 제안한다.

주파수 공간에서 DFT 계수들은 서로 독립적인 Gaussian 확률 분포를 따르기 때문에 각각의 가설에 따라 다음과 같은 확률 분포를 갖는다.

$$P(X_k|H_{0,k}) = \frac{1}{\pi|\widehat{V}_k|^2}\exp\left\{-\frac{|X_k|^2}{|\widehat{V}_k|^2}\right\}, \quad (9)$$

$$P(X_k|H_{1,k},|\widehat{S}_k|^2) = \frac{1}{\pi(|\widehat{S}_k|^2+|\widehat{V}_k|^2)}\exp\left\{\frac{-|X_k|^2}{|\widehat{S}_k|^2+|\widehat{V}_k|^2}\right\}, \quad (10)$$

위 가설을 기준으로 한 LLR 검사 값은 다음과 같이 표현되며, 음성검출의 기준으로 사용된다.

$$\Lambda = \log\sum_{k=0}^{M-1}\frac{P(X_k|H_{1,k},|\widehat{S}_k|^2)}{P(X_k|H_{0,k})} = \sum_{k=0}^{M-1}\left\{\frac{|X_k|^2}{|\widehat{V}_k|^2} - \log\frac{|X_k|^2}{|\widehat{V}_k|^2} - 1\right\} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \quad (11)$$

위 식을 보면 SNR을 개선하면 음성검출의 성능을 향상시킬 수 있음을 알 수 있다. 이를 위해서 음성향상 전처리 과정을 다음 장에서 제안한다.

Ⅲ. 정합필터를 이용한 음성 향상

잡음이 섞인 음성 신호로부터 잡음의 영향을 줄이는 음성 향상 방법으로는 Power Subtraction [4], Wiener Filtering [5] 방법 등이 있다. 이중 Power Subtraction 방법이 계산량이 적고 구현이 간단하기 때문에 많이 이용되고 있으며 다음과 같다.

$$\hat{s}_n = \frac{1}{M}\sum_{k=0}^{M-1}\widehat{S}_k\exp\left\{j\frac{2\pi}{M}kn\right\}, \quad (12)$$

$$\widehat{S}_k = \sqrt{|X_k|^2-|\widehat{V}_k|^2}\frac{X_k}{|X_k|}. \quad (13)$$

본 논문에서는 Power Subtraction 방법을 이용하여 음성 향상된 신호에 정합필터를 적용한 전처리 과정을 추가하였다. 이를 통해서 SNR을 최대화하고 이를 음성검출 알고리즘에 사용함으로써 성능을 향상한다.

각 주파수에서의 DFT 값 $\widehat{S}_k$에 α_k만큼의 가중치를 주어 선형 결합한 출력 신호는 다음과 같다.

$$\zeta_M = \sum_{k=0}^{M-1}\alpha_k\widehat{S}_k. \quad (14)$$

한편, Power Subtraction 방법으로 음성 향상된 신호의 크기 $|\widehat{S}_k|$는 다음과 같이 근사 시킬 수 있으며,

$$|\widehat{S}_k| = \sqrt{|S_k|^2+|V_k|^2-|\widehat{V}_k|^2} \approx |S_k|+\omega_k, \quad (15)$$

잡음의 파워가 잘 추정되어서 잔여 잡음 ω_k의 파워 σ_ω^2가 모든 주파수 대역에서 동일하다고 가정하면, 정합필터 출력에서의 잡음의 파워 ${\sigma_{\omega_T}}^2$는 $\sigma_\omega^2\sum_{k=0}^{M-1}|\alpha_k|^2$ 이고, SNR γ_M은 $\alpha_k = \widehat{S}_k^*$ 일때 식 (16)과 같이 최대가 된다. 여기서 *는 복소쌍을 나타낸다.

$$\gamma_M = \sum_{k=0}^{M-1}\frac{|\widehat{S}_k|^2}{\sigma_\omega^2}. \quad (16)$$

이와 같이 정합필터를 이용한 전처리 과정을 통해서 음성 향상된 신호는 다음과 같다.

$$\widehat{X}_k = (|X_k|^2-|\widehat{V}_k|^2)\frac{X_k}{|X_k|}. \quad (17)$$

한편, 변화하는 환경에 따른 잡음의 파워를 추정하기 위해서 현재 프레임의 음성검출 결과를 기준으로 잡음 구간이라고 판단될 경우에만 다음과 같은 방법으로 잡음의 파워가 갱신된다.

$$|\widehat{V}_k(n+1)|^2 = \mu|\widehat{V}_k(n)|^2+(1-\mu)|X_k(n)|^2, \quad 0\le\mu\le1 \quad (18)$$

μ는 망각 계수를 나타내며, 이 값은 SNR에 반비례하여 초기화 구간에서 주어진 잡음 환경에 맞게 적응적으로 값이 갱신되어진다. 부가적인 음성검출 과정을 필요로 하는 전통적인 잡음 추정방법은 잡음추정 결과가 부가적인 음성검출 결과에 많은 영향을 받지만 제안한 방법은 잡음의 특성이 프레임마다 심하게 변화하지 않으며, 음성신호보다 변화가 느린 일반적인 특성을 이용하여 현재 프레임의 음성검출 결과를 기준으로 다음 프레임의 잡음을 추

정해 내도록 하여 부가적인 음성검출의 도움이 없이도 잡음을 추정할 수 있게 하였다.

Ⅳ. 향상된 음성검출 방법

Ⅲ 장에서 제안한 정합필터를 적용한 음성향상 전처리 과정을 통해서 얻어진 향상된 음성신호의 주파수 스펙트럼은 $|\widehat{X_k}|=|\widehat{S_k}|^2=|X_k|^2-|\widehat{V_k}|^2$ 과 같이 나타낼 수 있으며, 이를 식 (11)의 LLR 검사에 적용하면 다음과 같다.

$$\Lambda=\sum_{k=0}^{M-1}\left\{\frac{|\widehat{X_k}|^2}{|\widehat{V_k}|^2}-\log\frac{|\widehat{X_k}|^2}{|\widehat{V_k}|^2}-1\right\}\begin{matrix}H_1\\ >\\ <\\ H_0\end{matrix}\eta, \qquad (19)$$

여기서 $|\widehat{V_k}|^2$은 정합 필터를 이용한 음성향상 전처리 이후의 잡음 파워에 대한 추정치이며, 다음의 식과 같이 갱신이 이루어진다.

$$|\widehat{V_k}(n+1)|^2=\mu|\widehat{V_k}(n)|^2+(1-\mu)|\widehat{X_k}(n)|^2 \qquad (20)$$
$$,0\le\mu\le1$$

식 (19)에서 볼 수 있듯이 SNR을 개선함으로서 LLR 값 Λ의 동적 영역을 충분히 증가시킬 수 있어서 음성검출의 성능을 향상시킬 수 있다.

이동통신 환경에서 통화가 시작된 후 짧은 구간 동안에는 대화가 이루어지지 않고 배경잡음만 존재한다고 가정하고 이 구간을 초기화 구간으로 정의한다. 이 구간에서 현재의 잡음 환경에 알맞도록 파라미터의 초기값이 계산되고 이 값을 기준으로 갱신이 이루어진다. 이를 통해서 주어진 잡음 환경에 대한 사전 정보가 없더라도 성능 파라미터를 초기화 할 수 있다. 또한 음성검출의 기준이 되는 문턱값 η를 변화하는 잡음의 환경에 따라서 적응적으로 갱신하는 알고리즘을 그림 1과 같이 제안한다. 초기화 구간에서는 LLR 값 Λ를 일정 크기의 메모리 (*MEM*)에 저장을 하며, 초기화 구간의 마지막 프레임(*initial_frame*)에서 LLR 값들의 평균 ($\overline{MEM}$)과 표준편차(σ_{MEM})를 기준으로 하여 초기 문턱값이 계산된다. 그리고 초기화 구간을 벗어나면, 현재 프레임이 잡음이라고 결정이 내려진 경우에만 메모리에 저장된 최근의 값들을 기준으로 다음과 같이 문턱값 갱신이 이루어진다.

$$\eta(n+1)=\mu\eta(n)+(1-\mu)(\overline{MEM}+\gamma\sigma_{MEM}), \qquad (21)$$

문턱값 갱신 알고리즘에서는 μ가 1 에서 먼 값을 가질 때 갱신 속도가 빨라지게 되어서 변화하는 환경에 빨리 적응할 수 있다.

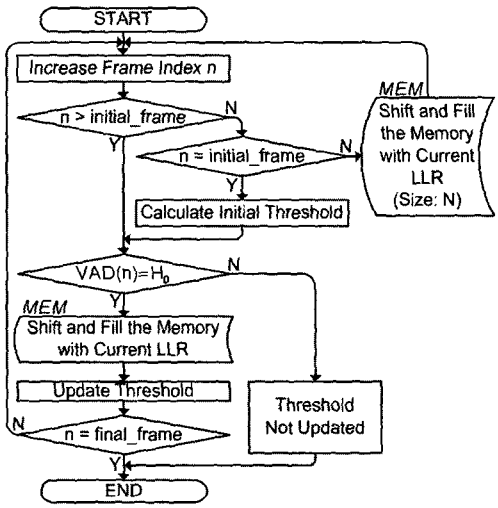

그림 1. 문턱값 갱신 알고리즘

또한 가중치 γ 의 값은 초기화 구간 안에서 그림 2와 같은 방법으로 잡음 환경에 알맞는 적절한 값이 자동적으로 결정되어 문턱값 갱신 알고리즘에 이용된다.

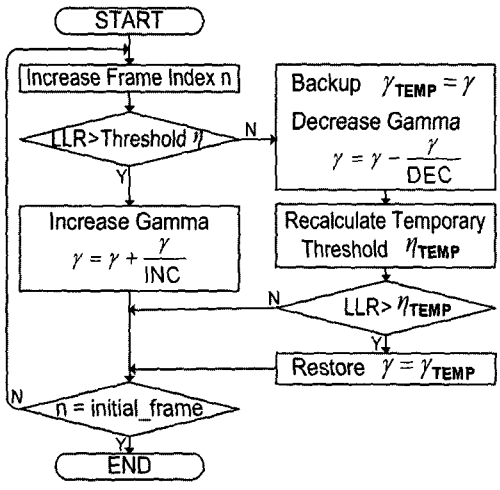

그림 2. 가중치(γ) 갱신 알고리즘

초기화 구간에서 Λ가 문턱값보다 큰 값을 가진

다면 초기화 구간의 가정에 위배되므로 문턱값을 증가시키기 위해 γ를 증가시킨다. 그렇지 않을 경우에는 그 값을 먼저 감소시킨 후에 검증과정을 거쳐서 가정을 만족하면 감소시킨 값을 사용하고 만족하지 못하면 이전의 값을 되돌린다. 이와 같은 방법으로 문턱값 갱신에 이용되는 γ값을 주어진 잡음 환경에 적응적으로 변화할 수 있도록 하였다. 이때 는 동적 영역이 크기 때문에 음성검출 확률보다 검출 오류 확률이 문턱값에 의한 영향을 많이 받으며, 음성검출 에러의 경우에는 Hangover 의 도움을 받아 성능을 개선할 수 있기 때문에 감소 계수 *DEC*가 증가 계수 *INC*보다 4배 크도록 각각 32와 8의 값을 이용하였다 [8].

V. 컴퓨터 시뮬레이션

컴퓨터 시뮬레이션을 위해서 SiPro Lab. DB [9]와 NOISEX-92 DB [10]에서 제공하고 있는 음성신호와 잡음의 표본을 이용하였으며, 실제적인 환경에 가까운 잡음 표본을 이용하기 위해서 babble 잡음과 vehicular 잡음에 -3dB 의 white 잡음이 더해졌다. 음성 신호의 표본은 8KHz 16Bit 표본화 되었으며, 각 프레임마다 음성 신호에 더해지는 잡음은 Gaussian 분포를 가진다. 음성 신호와 잡음은 대부분 stationary 하지 않아서 짧은 시간 구간일 경우에만 stationary 하다는 가정을 한다. 이를 위해서 30ms 길이의 Hamming 윈도우를 고려하여 프레임을 구분하였으며 128 포인트 DFT 처리를 하였다.

현재의 음성검출 알고리즘에서 Hangover의 사용은 일반적이며, 이를 통해서 낮은 에너지를 가지는 음성 신호들이 잡음으로 잘못 인식되는 확률을 줄일 수 있다. 이는 현재의 검출 결과를 바로 이웃하는 주변 프레임의 검출 결과와 비슷해지도록 현재 프레임의 검출 결정을 일정 구간 동안 보류하는 방법으로 이루어진다. 본 논문에서는 G.729B 와 동일한 4 프레임의 Hangover 구간을 정의하였다. 또한 현재 잡음 환경에 맞는 여러 가지 파라미터의 초기값을 결정하기 위해서 초기화 구간을 가정하였다. 초기화 구간은 32 프레임 이상 자유롭게 정의할 수 있으며, 초기화 구간이 길수록 파라미터들이 최적화되어서 음성검출의 성능향상을 기대할 수 있지만 너무 길 경우 대화가 없는 초기 상태가 길어지게 되므로 실제상황에 알맞지 않다. 본 논문에서는 128 프레임의 초기화 구간을 정의하였다.

그림 3은 vehicular 잡음 환경에서 여러 SNR 환경에 따라서 γ값이 초기값을 기준으로 자동으로 갱신되는 모습을 보여주고 있다. 이를 통해서 문턱값 결정을 위한 파라미터가 주변의 환경에 맞도록 적응적으로 결정된다.

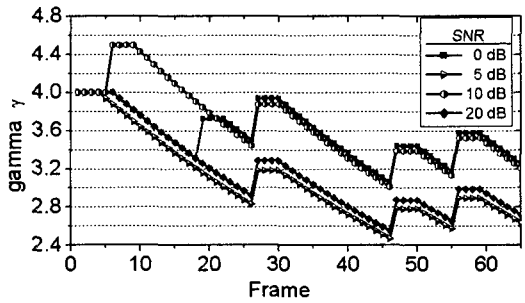

그림 3. 가중치(γ) 갱신 결과

그림 4는 제안한 문턱값 갱신 알고리즘으로 문턱값이 갱신되고 있는 모습을 보여준 그림이다. 초기화 구간과 음성이 포함되지 않은 구간으로 결정된 구간에서만 LLR 검사 값을 기준으로 문턱값이 변화되는 모습을 확인할 수 있다.

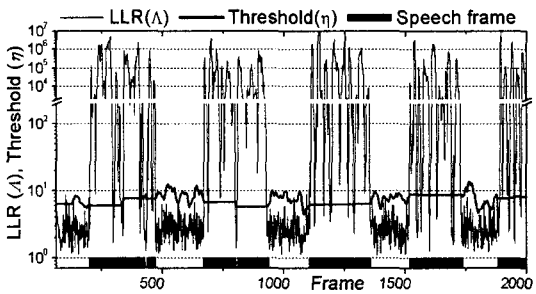

그림 4. 문턱값의 갱신 결과

그림 5, 6, 7은 각각 white, vehicular, babble 잡음 환경에서의 제안한 음성검출 방법의 문턱값 갱신 알고리즘을 적용한 경우와 그렇지 않은 경우를 기존에 제안되었던 음성검출 방법들과 비교한 결과이다. Sohn *et al.*의 결과는 참고논문의 값을 인용하였으며 점선은 최적의 성능치로 예상한 결과이다. 이 시뮬레이션 결과로부터, 제안한 문턱값 갱신 알고리즘을 적용한 음성검출 방법이 기존의 방법들보다 특히 낮은 SNR에서 더 좋은 성능을 나타냄을 확인할 수 있으며, 문턱값 갱신 알고리즘을 사용했을 때 성능이 좀더 개선되는 모습을 볼 수 있다.

표 1. 여러 가지 환경에서의 음성검출 성능 비교

Environments		Detection Probability P_D					Total Error Probability P_T				
Noise	SNR	A	B	C	D	E	A	B	C	D	E
White	0 dB	90.32	75.36	N/A	69.97	57.50	15.08	28.04	N/A	44.02	42.44
	5 dB	95.56	86.28	84.58	80.89	72.66	9.96	17.49	16.76	31.10	27.48
	10 dB	98.43	92.55	90.76	88.47	84.48	7.68	10.56	11.55	22.32	15.86
	20 dB	99.77	98.45	98.40	97.48	94.24	6.31	3.58	5.82	12.90	6.32
Vehicular	0 dB	92.31	90.10	N/A	87.96	81.94	10.52	16.24	N/A	27.03	49.58
	5 dB	97.85	94.60	97.30	93.30	88.80	6.19	10.14	7.54	21.10	41.40
	10 dB	99.32	97.69	98.46	96.57	95.24	5.02	6.04	7.56	17.69	34.26
	20 dB	99.90	99.56	99.75	99.44	97.44	5.52	3.36	7.74	15.27	24.20
Babble	0 dB	94.20	90.63	N/A	86.10	65.42	12.43	26.13	N/A	35.76	43.74
	5 dB	97.46	94.82	93.04	91.17	76.40	9.62	21.07	30.14	29.67	32.42
	10 dB	98.96	97.35	95.74	94.96	86.78	8.56	17.69	27.75	25.20	22.96
	20 dB	99.94	99.37	99.09	98.21	95.48	8.14	14.11	25.19	21.58	16.06

A : 제안하는 음성검출 방법의 음성검출 결과 (문턱값 갱신 알고리즘을 적용한 경우)
B : 제안하는 음성검출 방법의 음성검출 결과 (문턱값 갱신 알고리즘을 적용하지 않았을 경우)
C : Sohn *et al.* 음성검출 결과 [1]
D : Yang 음성검출 결과 [4]
E : ITU-T G.729 Annex B 음성검출 결과 [3]
N/A : Not Available

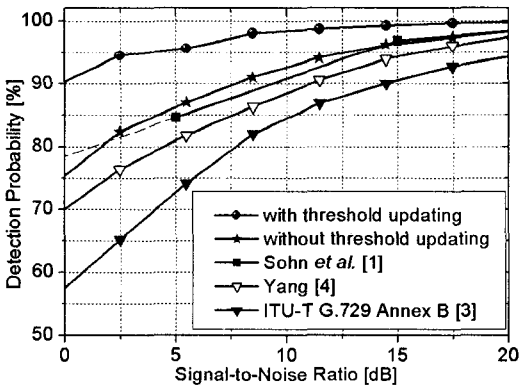

그림 5. white 잡음 환경에서의 음성검출 결과

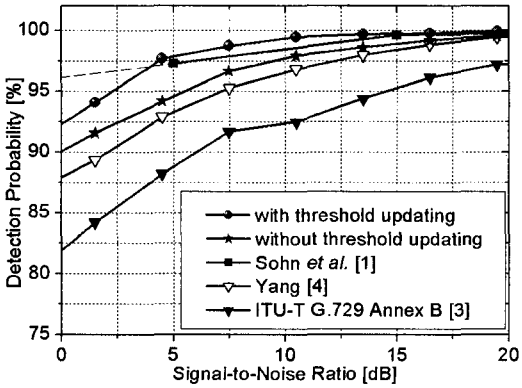

그림 6. vehicular 잡음 환경에서의 음성검출 결과

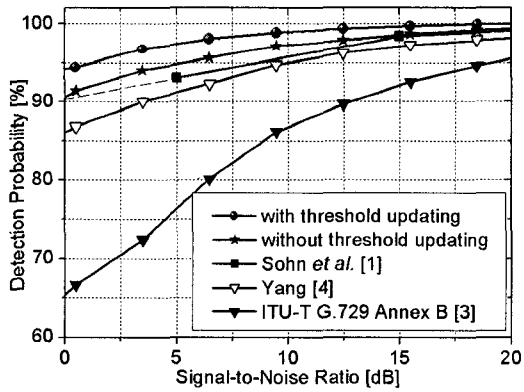

그림 7. babble 잡음 환경에서의 음성검출 결과

표 1은 각각의 알고리즘들에 대해서 음성검출에 대한 시뮬레이션 결과를 정리한 것이다. P_D는 올바른 음성검출이 내려질 확률을 의미하며, P_T는 음성검출의 음성부호화 응용에서 음질을 결정하는데 중요한 역할을 하는 검출 오류가 발생할 확률을 나타낸다. 음성구간을 잡음구간으로 결정을 내려서 음성이 부호화되어 전달되지 못한다면 음성 복호화 후 음질에 왜곡이 발생하게 된다. 이러한 관점에서 보면 제안하는 알고리즘이 다른 알고리즘들에 비해서 낮은 SNR 환경에서 상대적으로 적은 음성검출 오류를 가지므로 더 좋은 성능을 나타냄을 확인할 수 있다.

Ⅵ. 결론

잡음이 섞인 음성신호의 SNR을 개선하기 위하여 정합필터를 이용한 음성향상 전처리 과정을 적용하였고 음성검출의 기준이 되는 문턱값을 변화하는 주변환경에 적응적으로 갱신되도록 하는 문턱값 갱신 알고리즘을 제안하였다.

V장의 컴퓨터 시뮬레이션 결과에서 제안한 음성검출 방법이 G.729B 보다 뛰어난 성능을 나타냄을 볼 수 있으며, 통화품질에 중요한 영향을 미치는 음성검출 오류확률이 G.729B나 다른 알고리즘들에 비해 여러 SNR 환경에서 더 낮은 값을 가지고 있는 것을 확인할 수 있다. 이와 같은 결과로써 SNR이 낮은 환경에서도 잡음이 섞인 음성신호에서 음성신호가 존재하는 구간을 찾아낼 수 있었고, 이 결과를 음성 부호화에 이용할 경우 음성 부분에 대해서만 부호화 과정을 거치고 잡음으로 결정된 부분에 대해서는 잡음의 기본적인 특징만을 부호화 함으로써 실제적으로 채널을 통해서 전송되는 데이터의 양을 효과적으로 줄일 수 있게 되었다. 그리고 음성검출의 성능을 결정짓는 파라미터인 문턱값 가중치나 잡음 갱신 가중치 값을 실험적인 값이 아닌 초기화 구간에서의 신호의 특성을 기준으로 적응적으로 갱신된 값을 이용하므로 실제적인 환경에의 적용도 가능해졌다. 또한 문턱값 갱신을 통해서 변화하는 잡음에 적응하여 문턱값을 변화시켜서 음성검출 성능을 향상시켰을 뿐만 아니라 낮은 SNR 환경에서도 음성검출 오류를 줄일 수 있게 되었다

참 고 문 헌

[1] Jongseo Sohn, Wonyong Sung, "A statistical model-based voice activity detection," *IEEE Signal processing Letters*, Vol.6, No.1, 1999.

[2] Yong Duk Cho, Ahmet M. Kondoz, "Analysis and improvement of a statistical model-based voice activity detector," *IEEE Signal Processing Letters* , Vol.8, Issue 10, pp.276-278, Oct. 2001.

[3] ITU-T Recommendation, "G.729 Annex B: A silence compression scheme for G.729 optimized for terminals conforming to Recommendation V.70," Nov. 1996.

[4] Jin Yang, "Frequency domain noise suppression approaches in mobile telephone system," *ICASSP-93*, Vol.2, pp. 363-366. 1993.

[5] H. G. Hirsch, C. Ehrlicher, "Noise estimation techniques for robust speech recognition," *ICASSP-95*, pp. 153-156. May. 1995.

[6] Yoon-Chang Lee, Sang-Sik Ahn, "An improved voice activity detection algorithm employing speech enhancement preprocessing," *IEICE Transaction on Fundamentals*, Vol. E84-A, No.6, Jun. 2001.

[7] 이윤창, 안상식, "정합필터를 이용한 음성검출 방법," *한국통신학회 하계종합학술발표회 논문초록집*, Vol. 25, pp. 6, Aug. 2002.

[8] Ahmet M. Kondoz, "Digital speech coding for low bit rate communications systems," John Wiley & Sons, pp. 337-341.

[9] http://www.sipro.com , Sipro Lab., Telecom Inc.

[10] http://spib.rice.edu , Rice Univ., DSP group.

이 윤 창 (Yoon-Chang Lee) 정회원

1998년 2월 : 고려대학교 응용전자공학과 공학사

2000년 2월 : 고려대학교 전자정보공학과 공학석사

2000년 3월 ~ 현재 : 고려대학교 전자정보공학과 박사과정 수료

<주관심분야> 디지털 신호처리, 이동통신 알고리즘, 디지털 하드웨어 구현

안 상 식 (Sang-Sik Ahn) 정회원

1983년 2월 : 고려대학교 전자공학과 공학사

1985년 2월 : 고려대학교 전기공학과 공학석사

1984년 12월 ~ 1987년 8월 : LG 중앙 연구소 주임 연구원

1994년 1월 : Polytechnic Univ. 전기공학과 Ph. D.

1994년 2월 ~ 1995년 2월 : LG 중앙 연구소 책임 연구원

1995년 3월 ~ 현재 : 고려대학교 전자및정보공학부 부교수

<주관심분야> 신호처리, 이동통신 알고리즘