행오버를 이용한 SOFT DECISION 음성향상기법

장 준 혁, 김 남 수
서울대학교 전기공학부

Soft Decision Speech Enhancement using Hang-over

Joon-Hyuk Chang and Nam Soo Kim

School of Electrical Engineering, Seoul National University
Kwanak, P.O.Box 34,Seoul 151-742 Korea
Tel: +82-2-880-8439 Fax: +82-2-878-1452
e-mail : {changjh,nkim}@snu.ac.kr

요 약

본 연구에서는 행오버 (hang-over)를 이용한 새로운 soft decision 음성향상기법을 제안한다. 제시된 음성향상기법에서는 global 음성부재확률의 개념을 소개하고 이를 기존의 채널별 음성부재확률과 결합하여 통계적으로 신뢰할 수 있는 음성부재에 대한 확률값을 도출해낸다. 특히 음성의 꼬리 부분에서의 음성부재확률결정의 성능을 향상시키기 위해 행오버의 개념을 도입한다. Hidden Markov model (HMM)에 근거한 행오버를 이용하여 음성부재확률을 수정하는 부분을 소개하고 최종적으로 수정된 음성부재확률을 이용하여 새로운 잡음전력의 갱신 및 이득수정을 통해 향상된 음성을 만들어 낸다. 개발된 음성향상기법은 주관적인 음질평가에서 기존의 방법보다 뛰어난 성능을 나타내었으며, 특히 행오버를 이용한 음성부재확률의 수정에 관련한 성능을 검증하였다.

1 서론

이동환경에서의 음성통신의 중요성이 점차 증가하면서 단일 채널 마이크환경에서의 음성향상 알고리즘에 대한 연구가 주목받고 있다. 특히 저전송율을 위한 음성부호화기의 성능은 배경잡음에 민감하여 전처리기 (preprocesssor)로서 사용되는 음성향상알고리즘의 연구가 수년간 널리 행해지고 있다. 향상된 음성스펙트럼을 얻어내기 위한 여러 시도들은 spectral subtraction [1,2], Wiener filtering [3], soft decision estimation [3,4] 그리고 minimum mean square error (MMSE) estimation [5]방법들로 나뉠 수 있으며 관련 방법들은 구현의 용이성 및 다양한 배경잡음에의 적용가능의 장점을 지니고 있다. 특히 soft decision에 근거한 추정방법을 택했을때 뛰어난 성능을 가진다는 것이 알려져 있다. 이 soft decision 추정방법에서 음성부재확률 (SAP, speech absence probability)은 음성의 존재, 부재에 대한 확률값을 결정하는것으로 기존의 방법에서는 주파수 성분각각에서 확률값을 도출해내는데 [3,4,6,7] 이러한 것은 실제 각 주파수성분의 독립성을 가정한 것으로 견실한 값을 추정하는데는 충분치 않으며 특히 음성의 꼬리 부분은 기존의 통계적가정을 바탕으로 하여 신뢰성있는 값을 구하는데는 한계가 있다.

본 연구에서는 통계적인 신뢰성을 갖는 global 음성부재확률 (GSAP, global SAP)의 개념을 제시하고 이를 기존의 채널별 음성부재확률 (LSAP, local speech absence probability)과 결합하여 새로이 통계적으로 견실한 채널별 음성부재확률을 도출한다. 특히 기존의 통계적 가정을 바탕으로 했을때 음성의 꼬리부분의 부재확률값이 높은 현상을 관찰하고 새로운 제안으로 확률값을 낮추어 주기 위해 행오버의 개념을 소개하고 HMM을 이용하여 행오버 알고리즘을 수행한다.

제안된 행오버를 이용하여 보정된 음성부재확률을 이용하여 soft decision을 통해 이득을 수정할 뿐만아니라 잡음전력의 갱신에 이용하게 된다. 이득수정의 과정에서는 실제 신호대잡음비 (SNR, signal-to-noise ratio)수정을 통한 이득의 수정으로 향상된 성능을 가져오고 특히 잡음전력의 갱신작업이 잡음구간뿐만아니라 음성구간에서도 일어나게 함으로서 비정상 (non-stationary)잡음에 대비하게 된다.

주관적인 음질평가에서 제안된 음성향상방법이 북아메리카 CDMA 표준으로 채택되고 있는 IS-127의 음성향상기법보다 전반적으로 우수한 성능을 나타내었으며 특히 HMM에 근거하여 행오버를 soft decision에 적용하였을때의 성능향상을 확인하였다.

2 Soft Decision

잡음신호 n이 음성신호 x에 인가되어 오염된 음성신호 y를 만들어내게 되며 퓨리에 변환을 통해 주파수 축

으로 변환되어

$$Y_k(t) = X_k(t) + N_k(t) \quad (1)$$

가 되는데 여기서 $Y_k(t)$는 t번째 프레임에서의 k번째 주파수성분이 된다. 음성향상기법에서 사용되고 있는 기본가정은 다음과 같은 두 가설이다.

$$H_0 : \text{speech absent} : \mathbf{Y} = \mathbf{N}$$
$$H_1 : \text{speech present} : \mathbf{Y} = \mathbf{N} + \mathbf{X}.$$

위에서 $\mathbf{Y} = [Y_1, Y_2, ..., Y_M], \mathbf{N} = [N_1, N_2, ..., N_M]$, 그리고 $\mathbf{X} = [X_1, X_2, ..., X_M]$를 나타낸다.

음성존재, 부재에 관한 가설을 바탕으로 먼저 주파수채널별 음성부재확률인 LSAP를 구하면

$$\begin{aligned} p(H_0|Y_k(t)) &= \frac{p(Y_k(t)|H_0)p(H_0)}{p(Y_k(t))} \\ &= \frac{p(Y_k(t)|H_0)p(H_0)}{p(Y_k(t)|H_0)p(H_0) + p(Y_k(t)|H_1)p(H_1)} \\ &= \frac{1}{1 + \frac{p(H_1)}{p(H_0)}\Lambda(Y_k(t))} \end{aligned} \quad (2)$$

이 되고 향상된 음성부재확률을 위해 한 프레임에서의 고정된 음성부재확률인 GSAP를 제안하고 현재프레임의 관찰결과를 기반으로 하여 다음과 같이 계산한다.

$$\begin{aligned} p(H_0|Y(t)) &= \frac{p(Y(t)|H_0)p(H_0)}{p(Y(t))} \\ &= \frac{p(Y(t)|H_0)p(H_0)}{p(Y(t)|H_0)p(H_0) + p(Y(t)|H_1)p(H_1)} \end{aligned} \quad (3)$$

위에서 각 주파수 성분들의 통계적인 독립성을 가정하면 (3)은

$$p(H_0|Y(t)) = \frac{1}{1 + \frac{p(H_1)}{p(H_0)}\prod_{k=1}^{M}\Lambda(Y_k(t))} \quad (4)$$

와 같이 변형될 수 있으며 (2)와 (4)에서의 $p(H_0)$는 음성부재에 대한 *a priori* 확률값이 되고 $\Lambda(Y_k(t))$는 k번째 주파수 채널에서의 likelihood ratio가 된다.

음성신호와 잡음의 스펙트럼이 복소가우시안 분포를 따른다는 가정으로부터 가설 H_0와 H_1에 근거한 확률밀도함수는 다음과 같이 주어진다.

$$\begin{aligned} p(Y_k|H_0) &= \frac{1}{\pi\lambda_{n,k}}\exp\{-\frac{|Y_k|^2}{\lambda_{n,k}}\} \\ p(Y_k|H_1) &= \frac{1}{\pi[\lambda_{n,k} + \lambda_{s,k}]} \\ &\quad \cdot \exp\{-\frac{|Y_k|^2}{\lambda_{n,k} + \lambda_{s,k}}\} \end{aligned} \quad (5)$$

위에서 $\lambda_{s,k}$ 와 $\lambda_{n,k}$는 각각이 음성과 잡음의 분산이 된다. 따라서 likelihood ratio $\Lambda(Y_k(t))$는 [8]

$$\begin{aligned} \Lambda(Y_k(t)) &= \frac{p(Y_k(t)|H_1)}{p(Y_k(t)|H_0)} \\ &= \frac{1}{1 + \xi_k(t)}\exp\left[\frac{\gamma_k(t)\xi_k(t)}{1 + \xi_k(t)}\right] \end{aligned} \quad (6)$$

로 나타내어지는데, 여기서

$$\xi_k(t) \equiv \frac{\lambda_{s,k}(t)}{\lambda_{n,k}(t)}, \gamma_k(t) \equiv \frac{|Y_k(t)|^2}{\lambda_{n,k}(t)} \quad (7)$$

이 되고 $\xi_k(t)$와 $\gamma_k(t)$는 각각 predicted SNR과 *a posteriori* SNR로 불린다.

채널별 음성부재확률인 LSAP의 결정에는 음성과 잡음의 분산이 주요한 변수이며 특히 이 값들은 현재값뿐만이 아니라 과거의 관찰들에 의존한다. 이러한 SNR들의 견실한 추정은 LSAP와 GSAP의 결정에 주요한 요소로 작용한다. $\xi_k(t)$와 $\gamma_k(t)$을 추정하는 알고리즘에 대한 상세한 기술은 다음 절로 미루기로 한다.

LSAP와 GSAP는 각각이 주파수 채널과 현재프레임 전체에서의 고정된 음성부재확률값을 나타낸다. LSAP는 주파수채널별로 독립성을 가정한 것으로 그 값이 이전 프레임들의 동일 주파수 성분에만 의존함으로서 통계적인 견실성이 떨어지는 문제점을 가지고 있는 반면 GSAP는 현재 프레임에서의 모든 데이터에 의해 결정되어지는 값으로 신뢰성을 인정할 수 있고 특히 잡음구간의 경우 프레임전체의 고정된 값을 모든 주파수 성분에 균일하게 적용함으로서 이득의 불규칙한 변동을 줄일 수 있다. 이것은 음성향상과정에서의 원하지않는 부가적인 현상인 musical effect를 줄일 수 있는 방안이 된다.

그러나 음성구간에서의 GSAP의 적용은 다음과 같은 마스킹 (masking)현상을 고려하면 개선의 여지가 있다. 주파수축상의 잡음의 전력스펙트럼이 음성전력스펙트럼의 골짜기부분에 위치하게 되면 상대적으로 크게 들리며 이러한 곳은 음질의 향상을 위해서는 상대적인 큰 감쇄를 요구하게 된다. 적절한 방안으로 제안된 방법은 새로이 LSAP를 정의하여 잡음구간에서는 GSAP를 따르고 음성구간에서는 LSAP형태를 그대로 주는것이다. 즉 새로운 음성부재확률의 수정은 LSAP와 GSAP을 결합하여 새로운 향상된 LSAP를 다음과 같이 정의한다.

$$\begin{aligned} \tilde{p}(H_0|Y_k(t)) =& p(H_0|Y(t)) \cdot p(H_0|Y(t)) \\ &+ p(H_1|Y(t)) \cdot p(H_0|Y_k(t)) \end{aligned} \quad (8)$$

위에서 $\tilde{p}(H_0|Y_k(t))$는 수정된 LSAP를 나타내낸다. 통계적으로 그 값을 결정하기 어려운 음성전이구간에서의 음성부재확률값은 $\tilde{p}(H_0|Y_k(t))$의 특성으로부터 통계적신뢰성을 갖는 GSAP의 보상을 받는 LSAP를 사용하게 됨으로서 최대한의 음성왜곡의 방지를 목표로 하게된다.

3 HMM에 근거한 행오버의 적용

각 주파수 채널과 프레임의 고정된 음성부재확률을 나타내는 LSAP와 GSAP는 음성에서 잡음 또는 그 반대의 경우로 전이되는 과정에서는 정확히 추정하는데 한계가 있으며 특히 잡음환경이 비정상 (non-stationary)일때 잡음으로 오판함으로서 전반적인 음

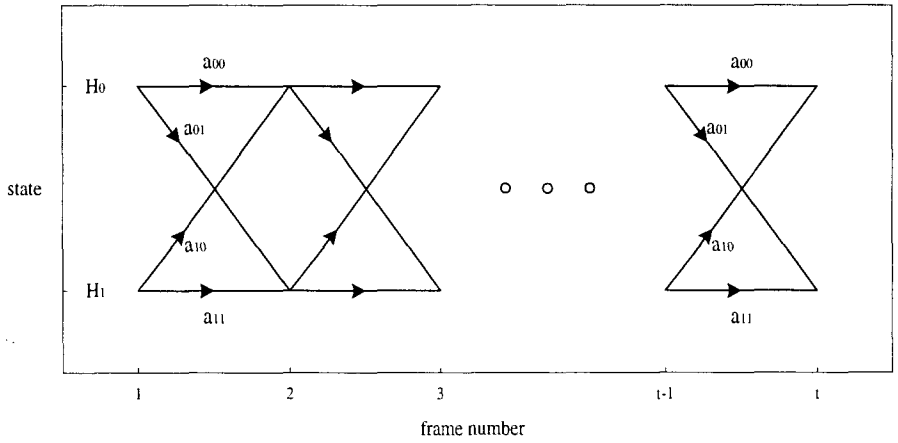

그림 1: 음성의 전이를 표현하는 HMM.

질의 저하를 초래한다. 이러한 문제점을 해결하기 위해 HMM에 근거한 행오버의 개념 [8]을 포함한 soft decision알고리즘을 소개한다. 제안된 방법에서 각 프레임은 H_0 또는 H_1의 두 상태로 가정할 수 있으며 이러한 상태는 일차 Markov process로 표현할 수 있다. 여기서는 각 상태의 전이 확률을 다음과 같은 파라미터로 표현할 수 있다.

$$a_{ij} = p(q_{t,k} = H_j | q_{t-1,k} = H_i), i, j = 0, 1 \quad (9)$$

위에서 $q_{t,k}$는 t프레임에서 k번째 주파수 성분의 상태를 나타낸다. 음성의 꼬리부분에서의 연속성을 강조하기 위해 전이확률에 대한 다음과 같이 제약이 필요하다.

$$p(q_{t,k} = H_1 | q_{t-1,k} = H_1) > p(q_{t,k} = H_1). \quad (10)$$

Markov process는 시불변이라는 가정을 바탕으로 정상상태에 도달하면 다음과 같이 식을 얻을 수 있다.

$$p(q_{t,k} = H_i) = p(H_i), i = 0, 1 \quad (11)$$

여기서 $p(H_0)$와 $p(H_1)$는 $p(H_0) + p(H_1) = 1$의 조건하에서 다음과 같은 방정식을 이용하여 구해낼 수 있다.

$$a_{01} p(H_0) = a_{10} p(H_1). \quad (12)$$

따라서, 총 process는 a_{01}와 a_{10}에 의해 간략히 파라미터화할 수 있으며 그 값으로 각각 0.2와 0.1을 선택한다.

Markov 프레임상태 모델은 현재의 관찰 뿐만 아니라 과거의 관찰결과들에도 의존하므로 음성부재확률계산식의 likelihood ratio식 $\Lambda(Y_k(t)) = p(H_1|Y_k(t))/p(H_0|Y_k(t))$ 을 $\Gamma(Y_k(t)) = p(q_{t,k} = H_1|\mathcal{Y}_k(t))/p(q_{t,k} = H_0|\mathcal{Y}_k(t))$와 같이 수정하여야 하며 여기서 $\mathcal{Y}_k(t) = \{Y_k(t), Y_k(t-1), \ldots, Y_k(1)\}$는 k번째 주파수 성분의 현재 프레임 t까지 관찰결과들의 집합이다. $\Gamma(Y_k(t))$의 계산을 위해서는 전향변수 (forward variable) $\alpha_k(t,i) = p(q_{t,k} = H_i, Y_k(t))$를 정의하여 잘 알려진 forward procedure [9]를 이용하여 다음과 같이 회귀적으로 구해낸다.

$$\alpha_k(t,i) = \begin{cases} p(H_i)p(Y_k(1)|q_{1,k} = H_i), & \text{for } t = 1 \\ \Big(\alpha_k(t-1,0)a_{0j} + \alpha_k(t-1,1)a_{1j}\Big) \cdot & \\ \quad p(Y_k(t)|q_{t,k} = H_i), & \text{for } t \geq 2. \end{cases} \quad (13)$$

위의 결과를 이용하여 $\Gamma(Y_k(t))$을 계산하기 위한 회귀적인 식은 다음과 같다.

$$\Gamma(Y_k(t)) = \frac{\alpha_k(t,1)}{\alpha_k(t,0)} = \frac{a_{01} + a_{11}\Gamma(Y_k(t-1))}{a_{00} + a_{10}\Gamma(Y_k(t-1))} \Lambda(Y_k(t)). \quad (14)$$

그림 1은 HMM이 어떻게 음성의 전이과정을 나타내는데 사용되어 지는지 나타내고 있다.

위에서 구해진 $\Gamma(Y_k(t))$를 이용하여 최종적으로 사용되어질 음성부재확률값으로 사용되어질 LSAP와 GSAP는 다음과 같이 $\Lambda(Y_k(t))$를 대체함으로서 구해진다.

$$p(H_0|Y_k(t)) = \frac{1}{1 + \frac{p(H_1)}{p(H_0)}\Gamma(Y_k(t))} \quad (15)$$

$$p(H_0|Y(t)) = \frac{1}{1 + \frac{p(H_1)}{p(H_0)}\prod_{k=1}^{M}\Gamma(Y_k(t))}. \quad (16)$$

위의 (15)와 (16)을 이용하여 (8)을 바탕으로 행오버를 고려한 수정된 채널별 음성부내확률값이 얻어진다. 행오버를 고려한 GSAP와 원래의 값과 행오버에 의해 수정된 값을 그림 2에서 비교하였으며, 실제 음성전이구간에서 낮아진 GSAP를 관찰할 수 있다.

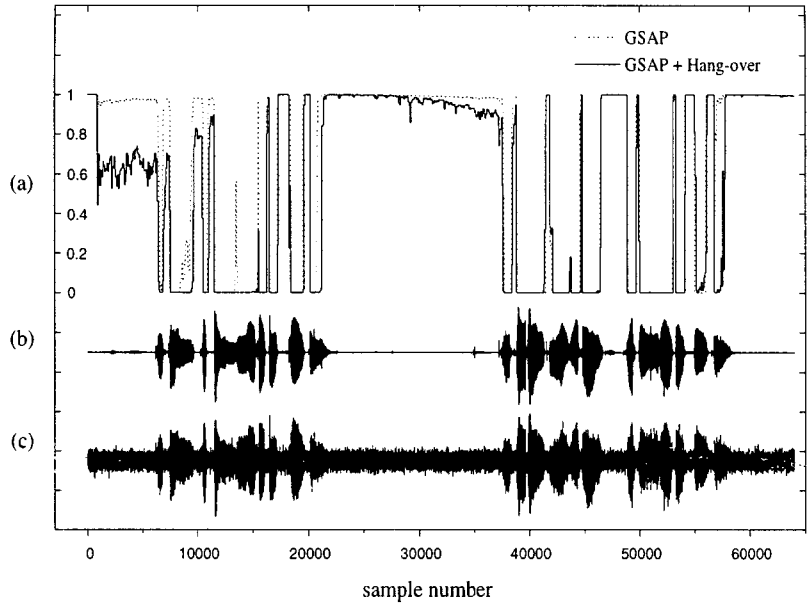

그림 2: 행오버의 영향 (a) GSAP (b) 원래의 음성파형 (c) vehicular 잡음 (SNR = 5dB)에 오염된 음성신호.

4 음성 및 잡음전력 추정

주파수축상에서의 음성향상을 다루는 문제에서 잡음전력 $\lambda_n(k)$의 추정은 성능에 커다란 영향을 미치는 요소이다. 일반적인 방법으로 음성검출기 (VAD, voice activity detector)를 이용하여 음성부재구간에서만 잡음전력을 갱신한다. 그러나 실제 잡음환경이 비정상 (non-stationary)일경우 잡음전력은 음성구간에서도 변화하게 되며 적절한 추정을 위해서는 음성구간에서의 잡음전력갱신을 고려해야 한다.

우리는 잡음전력 $\lambda_n(k)$과 음성전력 $\lambda_s(k)$의 추정을 위해 배경잡음과 음성 각각에 long-term smoothed 전력스펙트럼을 사용하며 관계식은 다음과 같다.

$$\hat{\lambda}_{n,k}(t+1) = \zeta_n \hat{\lambda}_{n,k}(t) + (1-\zeta_n) E[|N_k(t)|^2 | Y_k(t)]$$
$$\hat{\lambda}_{s,k}(t+1) = \zeta_s \hat{\lambda}_{s,k}(t) + (1-\zeta_s) E[|X_k(t)|^2 | Y_k(t)] \quad (17)$$

위에서 $\hat{\lambda}_{n,k}(t)$와 $\hat{\lambda}_{s,k}(t)$는 $\lambda_{n,k}(t)$와 $\lambda_{s,k}(t)$의 추정치이며 ζ_n와 ζ_s는 음성과 잡음의 정상(stationarity)가정을 고려한 smoothed parameter이다. 위식에 음성의 존재, 부재를 고려하면 현재 프레임에서의 잡음과 음성전력의 추정치는

$$\begin{aligned} E[|N_k(t)|^2|Y_k(t)] &= E[|N_k(t)|^2|Y_k(t), H_0]p(H_0|Y_k(t)) \\ &\quad + E[|N_k(t)|^2|Y_k(t), H_1]p(H_1|Y_k(t)) \\ E[|X_k(t)|^2|Y_k(t)] &= E[|X_k(t)|^2|Y_k(t), H_0]p(H_0|Y_k(t)) \\ &\quad + E[|X_k(t)|^2|Y_k(t), H_1]p(H_1|Y_k(t)) \end{aligned} \quad (18)$$

이며 실제 수식의 주어진 의미에서 잡음전력의 갱신은 음성부재구간 뿐만이 아니라 음성존재구간에서도 일어남을 쉽게 알 수 있다. 음성구간에서는 이미 우리가 정의한 수정된 LSAP의 특성으로부터 채널별 음성부재확률이 정의되므로 각 주파수성분 중 음성부재확률이 낮은 주파수채널에서 잡음전력의 갱신이 일어나게 되며 이것은 음성, 특히 유성음의 유사 (quasi) 주기적인 성질을 고려한 것이다.

(17)에서 추정된 잡음과 음성전력을 이용하여 다음 프레임에서 사용될 predicted SNR값의 추정이 다음과 같은 식에 의해 이루어진다.

$$\hat{\xi}_k(t+1) = \frac{\hat{\lambda}_{s,k}(t+1)}{\hat{\lambda}_{n,k}(t+1).} \quad (19)$$

위 식에서 구한 predicted SNR은 다음 프레임에서 GSAP의 계산과 잡음전력갱신에 사용되어지게 된다.

5 이득수정

$\hat{X}(t) = \left[\hat{X}_1(t), \hat{X}_2(t), \cdots, \hat{X}_M(t)\right]$을 t번째 프레임에서 추정된 음성의 스펙트럼이라고 하자. 얻어진 음성부재확률은 음성향상을 위한 이득의 수정에 이용된다. 관측된 오염된 음성의 스펙트럼을 바탕으로 원래의 음성스펙트럼을 추정하는 방법을 사용하며, 실제로 Ephraim-Malah의 잡음제거방법 (EMSR, Ephraim-Malah noise suppression rule) [5]을 사용한다. EMSR의 스펙트럼향상은 다음과 같이 이루어지며

$$\hat{X}_k(t) = G(\eta_k(t), \gamma_k(t)) Y_k(t) \quad (20)$$

위에서 $\eta_k(t)$는 *a priori* SNR로 불리고 $\gamma_k(t)$는 (7)의 *a posteriori* SNR이다. (20)에서의 이득은 다음과 같이

주어진다.

$$G(\eta,\gamma) = \frac{\sqrt{\pi}}{2}\sqrt{\frac{\eta}{\gamma(1+\eta)}} \times M\left[\frac{\gamma\eta}{1+\eta}\right] \quad (21)$$

여기서

$$M[\theta] = \exp\left(-\frac{\theta}{2}\right)\left[(1+\theta)I_0\left(\frac{\theta}{2}\right)+\theta I_1\left(\frac{\theta}{2}\right)\right] \quad (22)$$

이고, I_0와 I_1은 각각 제1종, 제2종 변형 베셀 (modified Bessel)함수이다. *a priori* SNR은 decision directed 방법에 의해 계산되어 지며, 실제로 다음과 같다.

$$\hat{\eta}_k(t) = \alpha\frac{|\hat{X}_k(t-1)|^2}{\hat{\lambda}_{n,k}(t-1)} + (1-\alpha)P[\hat{\gamma}_k(t)-1] \quad (23)$$

여기서 $\alpha \in [0,1]$ 이고 만약 $x \geq 1$이면 $P[x]=x$ 그렇지 않으면 $P[x]=0$이다.

잡음제거이득은 (18)에서와 같이 음성의 존재, 부재 여부를 고려하여 갱신되어져야 하며 그것은 soft decision의 개념을 바탕으로 하여 다음과 같이 구해낼 수 있다.

$$\begin{aligned}\hat{X}_k(t) &= E[|X_k(t)||Y_k(t)] \\ &= E[|X_k(t)||Y_k(t),H_0]p(H_0|Y_k(t)) \\ &\quad + E[|X_k(t)||Y_k(t),H_1]p(H_1|Y_k(t)) \\ &= E[|X_k(t)||Y_k(t),H_1]p(H_1|Y_k(t)). \end{aligned} \quad (24)$$

위의 식으로부터 우리는 사용되어 지는 잡음제거이득 $G(\cdot,\cdot)$에 다음과 같이 음성존재확률 $P(H_1|Y_k(t))$을 인가함으로서 새로운 이득함수를 구해낸다.

$$\tilde{G}(\eta_k(t),\gamma_k(t)) = p(H_1|Y_k(t))G(\eta_k(t),\gamma_k(t)). \quad (25)$$

여기서 우리는 위의 이득의 직접수정을 바탕으로 잡음제거이득의 직접적인 수정이 아니라 이득함수의 주요 매개변수인 $\eta_k(t)$와 $\gamma_k(t)$를 수정함으로써 새로운 SNR의 추정치를 이득함수에 인가하는 다음의 방법을 택한다.

$$\tilde{G}(\eta_k(t),\gamma_k(t)) = G(\tilde{\eta}_k(t),\tilde{\gamma}_k(t)) \quad (26)$$

위에서

$$\tilde{\eta}_k(t) = p(H_1|Y_k(t))\eta_k(t) \quad (27)$$

$$\tilde{\gamma}_k(t) = p(H_1|Y_k(t))\gamma_k(t) \quad (28)$$

이 되며, 이러한 SNR의 수정을 통한 이득수정은 향상된 음질로서 검증되었다.

6 실험결과 및 결론

제안된 음성향상기법 (SESD, speech enhancement based on soft decision)의 성능을 평가하기 위해 mean opinion score (MOS) 평가방법을 택하였다. 남성, 여성화자 각각이 10개의 문장을 발음하도록한 음성을 8kHz로 샘플링한 데이터에 세가지형태의 잡음이 부가되었다. 잡음은 NOISEX-92 데이터베이스의 babble, pink, white잡음을 사용하였으며 SNR을 5, 10, 15dB로 달리하여 조사하였다. 평가에 참가한 인원은 총 10명이며 그 결과가 표1에 나타나 있다. 제안된 SESD방법은 IS-127 standard에서 쓰이고 있는 음성향상기법보다 전반적으로 우수한 성능을 나타내었으며, 행오버를 이용한 음성부재확률의 수정역시 전체적으로 성능향상을 가져왔다.

본 연구에서는 새로운 음성부재확률을 정의하여 음성존재, 부재구간 각각에 통계적으로 견실한 값을 도출해 내었다. 더욱이 제안된 HMM에 근거한 행오버를 음성부재확률을 구하는데 이용함으로서 음성의 전이구간에서도 향상된 성능을 보였다. 이와 같이 수정된 음성부재확률을 음성부재구간 뿐만아니라 음성존재구간에서도 잡음전력의 갱신이 일어나도록 하였으며 이득수정방법에 적용함으로서 성능향상을 꾀하였다. 특히 새로운 선택방법으로서 주요 파라미터인 SNR을 수정함으로서 이득을 수정하도록 한 방법이 제시되었다. 그림1에 제안된 SESD알고리즘에 대한 전체블록도가 도시되었다.

7 참고문헌

[1] S. F. Boll, "Suppression of Acoustic Noise in Speech using Spectral Subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.29, April. 1979.

[2] J. S. Lim, A. V. Oppenheim, "Enhancement and Bandwidth compression of Noisy Speech," in *Proc.IEEE.*, vol. 67, December. 1979.

[3] 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.

[4] P. Scalart and J. Vieira Filho, "Speech Enhancement Based on A Priori Signal to Noise Estimation," *Proc. of Int. Conf. Acoust., Speech, Signal Processing*, 1996.

[5] 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.

[6] D. Malah, R. Cox and A. J. Accardi, "Tracking speech-presence uncertainty to improve speech enhancement in non-stationary noise environments," *Proc. of Int. Conf. Acoust., Speech, Signal Processing*, Phoenix, AZ, Mar. 1999.

[7] J. Yang, "Frequency Domain Noise suppression Approaches in Mobile Telephone Systems," *Proc. of Int. Conf. Acoust., Speech, Signal Processing*, pp. II

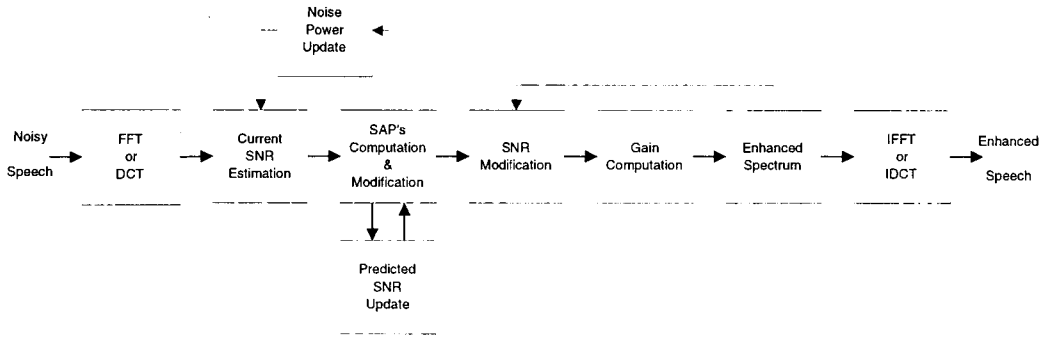

그림 3: 제안된 SESD알고리즘에 대한 전체블록도.

noise	white			babble			pink		
SNR(dB)	5	10	15	5	10	15	5	10	15
none	2.00	2.14	2.29	1.13	1.68	1.92	1.19	1.49	2.11
IS-127	2.60	3.09	3.54	2.34	2.89	3.51	2.31	3.04	3.38
SESD	2.95	3.45	3.70	2.51	3.16	3.76	2.45	3.17	3.60
SESD+hang-over	2.99	3.45	3.79	2.65	3.20	3.78	2.47	3.17	3.61

표 1: 제안된 음성향상기법(SESD)과 IS-127 음성향상기법에 대한 MOS평가비교.

363-366, 1993.
[8] 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.
[9] L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition* Prentice Hall, Englewood Cliffs, NJ, 1993.