



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2008-0088115
(43) 공개일자 2008년10월02일

(51) Int. Cl.

G10L 19/00 (2006.01) G10L 21/02 (2006.01)

G10L 19/14 (2006.01)

(21) 출원번호 10-2007-0030554

(22) 출원일자 2007년03월28일

심사청구일자 2007년03월28일

(71) 출원인

인하대학교 산학협력단

인천 남구 용현동 253 인하대학교

(72) 발명자

장준혁

서울 서초구 서초동 1660-31 B동 301호

박윤식

서울 은평구 불광1동 268-5 401호

(74) 대리인

김건우

전체 청구항 수 : 총 2 항

(54) 최소 통계법과 소프트 디시전법을 확률적으로 결합하여잡음 전력을 추정하는 방법

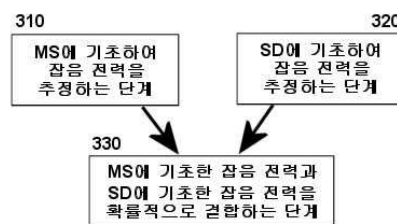
(57) 요약

본 발명은 기존의 최소 통계법(Minimum Statistics; MS)과 소프트 디시전법(Soft Decision; SD)을 확률적으로 결합하여 잡음 전력을 추정하는 새로운 잡음 전력 추정 방법에 관한 것으로서, 보다 구체적으로는 (1) 기존의 최소 통계법(MS)을 이용하여 잡음 전력의 제1 추정치($\hat{\lambda}_n^{MS}$)를 구하는 단계와, (2) 기존의 소프트 디시전법(SD)을 이용하여 잡음 전력의 제2 추정치($\hat{\lambda}_n^{SD}$)를 구하는 단계와, (3) 상기 단계 (1) 및 단계 (2)에서 구해진 잡음 전력의 상기 제1 추정치($\hat{\lambda}_n^{MS}$) 및 상기 제2 추정치($\hat{\lambda}_n^{SD}$)를 다음 수학적식 같이 확률적으로 결합함으로써, 새로운 잡음 전력의 추정치($\hat{\lambda}_n^{CIPE}$)를 구하는 단계를 포함하는 잡음 전력 추정 방법에 관한 것이다.

$$\hat{\lambda}_n^{CIPE} = (1 - p(H_0 | Y(t, k))) \hat{\lambda}_n^{MS} + p(H_0 | Y(t, k)) \hat{\lambda}_n^{SD}$$

본 발명의 잡음 전력 추정 방법에 따르면, 기존의 MS에 기초한 잡음 전력 추정 방법과 기존의 SD에 기초한 잡음 전력 추정 방법을 확률적으로 결합함으로써, 잡음 구간 및 깨끗한 음성 구간에서 모두 우수한 잡음 전력 추정 능력을 보일 수 있다.

대표도 - 도3



특허청구의 범위

청구항 1

잡음신호 n 이 음성신호 x 에 인가된 오염된 음성신호(이하, ‘잡음음성’이라 함) y 에 대하여, n , x , y 의 t 번째 프레임에서의 k 번째 주파수 성분을 $N(t, k)$, $X(t, k)$, $Y(t, k)$ 라 하고,

H_0 과 H_1 을 다음 수학적 식 1과 같이 각각 음성이 존재하지 않을 경우와 음성이 존재할 경우의 가정이라고 하는 경우에 있어서,

<수학적 식 1>

$$H_0 : \text{speech absent} : Y(t, k) = N(t, k)$$

$$H_1 : \text{speech present} : Y(t, k) = X(t, k) + N(t, k)$$

최소 통계법(Minimum Statistics; MS)과 소프트 디시전법(Soft Decision; SD)을 확률적으로 결합하여 잡음 전력을 추정하는 방법으로서,

(1) 다음 수학적 식 2에 의해 MS를 이용하여 잡음 전력의 제1 추정치($\hat{\lambda}_n^{\text{MS}}$)를 구하는 단계;

<수학적 식 2>

$$\hat{\lambda}_n(t, k) = B_{\min}(t, k) \lambda_{\min, n}(t, k)$$

(여기서, $\lambda_{\min, n}(t, k)$ 는 잡음신호 n 의 분산 $\lambda_n(t, k) = E\{|N(t, k)|^2\}$ 의 최소값이고, $B_{\min}(t, k)$ 는 바이어스 보상 함수로서, 다음 수학적 식 3과 같이 근사화될 수 있다.

<수학적 식 3>

$$B_{\min}(t, k) \approx 1 + (D-1) \frac{2}{Q_{\text{eq}}(t, k)}$$

여기서, D 는 한 프레임당 스펙트럼의 수이며, $Q_{\text{eq}}(t, k)$ 는 다음 수학적 식 4와 같이 표현될 수 있다.

<수학적 식 4>

$$\frac{1}{Q_{\text{eq}}(t, k)} \approx \frac{\widehat{\text{var}} P(t, k)}{2 \hat{\lambda}_n^2(t-1, k)}$$

여기서, $\widehat{\text{var}} P(t, k)$ 는 스무딩된 파워 스펙트럼 $P(t, k)$ 의 분산으로서, 다음 수학적 식 5 내지 수학적 식 7과 같이 구해질 수 있다.

<수학적 식 5>

$$\widehat{\text{var}} P(t, k) = \overline{P^2}(t, k) - \overline{P}^2(t, k)$$

<수학적 식 6>

$$\overline{P}(t, k) = \beta(t, k) \overline{P}(t-1, k) + (1 - \beta(t, k)) P(t, k)$$

<수학적 식 7>

$$\overline{P^2}(t, k) = \beta(t, k) \overline{P^2}(t-1, k) + (1 - \beta(t, k)) P^2(t, k)$$

여기서, 스무딩 파라미터인 $\beta(t, k)$ 는 $\beta(t, k) = \alpha^2(t, k)$ 로서 주어지며, $\alpha(t, k)$ 는 다음 수학적 식 8과 같이 주어진다.

<수학적 식 8>

$$\alpha(t, k) = \frac{1}{1 + (P(t-1, k) / \lambda_n(t, k) - 1)^2}$$

(2) 다음 수학적식 9에 의해 SD를 이용하여 잡음 전력의 제2 추정치($\hat{\lambda}_n^{\text{SD}}$)를 구하는 단계;

<수학적식 9>

$$\hat{\lambda}_n(t+1, k) = \zeta_n \hat{\lambda}_n(t, k) + (1 - \zeta_n) E[|N(t, k)|^2 | Y(t, k)]$$

(여기서, ζ_n 은 $0 < \zeta_n < 1$ 의 범위를 갖는 스무딩 파라미터이며, $E[|N(t, k)|^2 | Y(t, k)]$ 은 다음 수학적식 10 내지 수학적식 12를 이용하여 구할 수 있다.

<수학적식 10>

$$E[|N(t, k)|^2 | Y(t, k)] = E[|N(t, k)|^2 | Y(t, k), H_0] p(H_0 | Y(t, k)) + E[|N(t, k)|^2 | Y(t, k), H_1] p(H_1 | Y(t, k))$$

<수학적식 11>

$$E[|N(t, k)|^2 | Y(t, k), H_0] = |Y(t, k)|^2$$

<수학적식 12>

$$E[|N(t, k)|^2 | Y(t, k), H_1] = \left(\frac{\xi(t, k)}{1 + \xi(t, k)} \right) \hat{\lambda}_n(t, k) + \left(\frac{1}{1 + \xi(t, k)} \right)^2 |Y(t, k)|^2$$

여기서, $\xi(t, k)$ 는 선행 SNR(*a priori* SNR)로서, $\xi(t, k) \equiv \frac{\lambda_s(t, k)}{\lambda_n(t, k)}$ 로서 정의되며($\lambda_s(t, k)$ 와 $\lambda_n(t, k)$ 는 각각 음성신호와 잡음신호의 분산을 나타냄), $p(H_1 | Y(t, k)) = 1 - p(H_0 | Y(t, k))$ 이고, $p(H_0 | Y(t, k))$ 는 다음 수학적식 13과 같이 구할 수 있다.

<수학적식 13>

$$p(H_0 | Y(t, k)) = \frac{p(Y(t, k) | H_0) p(H_0)}{p(Y(t, k) | H_0) p(H_0) + p(Y(t, k) | H_1) p(H_1)} = \frac{1}{1 + q A(Y(t, k))}$$

여기서, q는 $q = p(H_1) / p(H_0)$ 로서 주어지며, 우도비(likelihood ratio) $A(Y(t, k))$ 는 다음 수학적식 14와 같다.

<수학적식 14>

$$A(Y(t, k)) = \frac{p(Y(t, k) | H_1)}{p(Y(t, k) | H_0)} = \frac{1}{1 + \xi(t, k)} \exp \left[\frac{\gamma(t, k) \xi(t, k)}{1 + \xi(t, k)} \right]$$

여기서, $\xi(t, k)$ 는 선행 SNR(*a priori* SNR)로서 앞서 정의된 바 있고, $\gamma(t, k)$ 는 사후 SNR(*a posteriori* SNR)로서 다음 수학적식 15와 같이 정의된다.

<수학적식 15>

$$\gamma(t, k) \equiv \frac{|Y(t, k)|^2}{\lambda_n(t, k)}$$

(3) 구해진 잡음 전력의 상기 제1 추정치($\hat{\lambda}_n^{\text{MS}}$) 및 상기 제2 추정치($\hat{\lambda}_n^{\text{SD}}$)를 이용하여, 다음 수학적식 16에 의해 잡음 전력의 추정치($\hat{\lambda}_n^{\text{CnFE}}$)를 구하는 단계

<수학적식 16>

$$\hat{\lambda}_n^{\text{CnFE}} = (1 - p(H_0 | Y(t, k))) \hat{\lambda}_n^{\text{MS}} + p(H_0 | Y(t, k)) \hat{\lambda}_n^{\text{SD}}$$

를 포함하는 방법.

청구항 2

제1항에 있어서,

상기 $\alpha(t,k)$ 의 최대값을 0.96으로 제한하고, 상기 $\beta(t,k)$ 의 최대값은 0.8로 제한하며, 상기 D와 상기 ζ 의 값은 각각 128과 0.99로 설정하는 방법.

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

- <11> 본 발명은 잡음 전력을 추정하는 방법에 관한 것으로서, 특히 기존의 최소 통계법(Minimum Statistics; MS)과 기존의 소프트 디시전법(Soft Decision; SD)을 확률적으로 결합하여 잡음 전력을 추정하는 새로운 잡음 전력 추정 방법에 관한 것이다.
- <12> 음성 향상 기술은 다양한 음성신호 처리 시스템의 전처리기(preprocessor)로 사용되어 시스템의 성능을 향상시키며 특히, 음성 통신에 기초한 음성부호화기의 전처리기로서 이동 환경에서의 배경 잡음에 대한 직접적인 음질 향상 및 음성 부호화기의 성능 향상에 중요한 요소로 작용한다. 이러한 음성 향상 기술의 성능에 영향을 미치는 주요 세부 기술로는 (1) 음성신호 및 잡음신호에 대한 통계적 모델 추정 기법, (2) 스펙트럼 이득 수정 기법 및 (3) 잡음신호의 추정 기법 등으로 정리 할 수 있다. 최근 이동 통신 기술의 발달로 이동 중에 발생할 수 있는 비정상적(non-stationary)인 배경 잡음과 음성 구간에서의 배경 잡음 추정의 어려움 때문에 정확한 잡음 추정을 위한 연구가 주목을 받고 있다.
- <13> 현재 잡음 전력을 추정하는 데 있어서 우수한 성능을 보인다고 알려진 대표적인 방법으로는 최소 통계법(Minimum Statistics; MS)에 기초한 방법과 소프트 디시전법(Soft Decision; SD)에 기초한 방법이 있다. MS에 기초한 잡음 전력 추정은 R. Martin에 의해 제안되었으며 G. Doblinger 등에 의해 보완되었다. MS에 기초한 잡음 전력 추정 기법은, 음성신호와 잡음신호가 통계적으로 독립이라는 가정 하에 오염된 음성 구간의 휴지 구간에서 음성 전력이 잡음신호의 전력 레벨까지 감소한다는 점에서 착안되었으며, 음성 구간의 레벨 사이를 연결해 줄 충분히 큰 윈도우를 사용하여 최적화된 전력 스펙트럼 스무딩(smoothing)을 거친 최소 잡음 전력을 추정한 후, 각 스펙트럼 성분에 해당하는 바이어스(bias) 보상을 함으로써 잡음 전력을 추정한다.
- <14> 이에 반하여, SD에 기초한 잡음 전력 추정 기법은, 현재 잡음신호의 추정을 위하여 확률에 기초한 프레임 간 스무딩을 이용하여 이전의 추정신호를 갱신하여 현재의 잡음 전력을 추정하는 방법이다. 다시 말해, SD에 기초한 잡음 전력 추정 기법에서는, 잡음신호로 확실하게 추정된 기존의 잡음신호에 기초하여 통계적 스무딩을 위해 가중치 파라미터로 사용되는 음성 부재 확률(Speech Absence Probability; SAP)에 따라 스펙트럼 성분을 갱신함으로써 잡음 전력을 추정하게 된다.
- <15> 그러나 기존의 MS에 기초한 잡음 전력 추정 방법과 기존의 SD에 기초한 잡음 전력 추정 방법은 잡음 전력을 추정하는 데 있어서 우수한 성능을 보인다고 알려져 있기는 하나, 각각 잡음 구간 및 깨끗한 음성 구간에서는 잡음 전력의 추정 능력이 떨어진다는 문제점이 있다.

발명이 이루고자 하는 기술적 과제

- <16> 본 발명은, 상기와 같은 문제점을 해결하기 위해 제안된 것으로서, 기존의 MS에 기초한 잡음 전력 추정 방법과 기존의 SD에 기초한 잡음 전력 추정 방법을 확률적으로 결합함으로써, 잡음 구간 및 깨끗한 음성 구간에서 모두 우수한 잡음 전력 추정 능력을 나타내는 새로운 잡음 전력 추정 방법을 제안하는 것을 그 목적으로 한다.

발명의 구성 및 작용

- <17> 상기한 목적을 달성하기 위한 본 발명의 특징에 따른, 잡음 전력 추정 방법은,
- <18> 잡음신호 n 이 음성신호 x 에 인가된 오염된 음성신호(이하, ‘잡음음성’이라 함) y 에 대하여, n , x , y 의 t 번째

프레임에서의 k번째 주파수 성분을 $N(t, k)$, $X(t, k)$, $Y(t, k)$ 라 하고,

<19> H_0 과 H_1 을 다음 수학적 식 1과 같이 각각 음성이 존재하지 않을 경우와 음성이 존재할 경우의 가정이라고 하는 경우에 있어서,

수학적 식 1

$$H_0 : \text{speech absent} : Y(t, k) = N(t, k)$$

<20> $H_1 : \text{speech present} : Y(t, k) = X(t, k) + N(t, k)$

<21> 최소 통계법(Minimum Statistics; MS)과 소프트 디시전법(Soft Decision; SD)을 확률적으로 결합하여 잡음 전력을 추정하는 방법으로서,

<22> (1) 다음 수학적 식 2에 의해 MS를 이용하여 잡음 전력의 제1 추정치($\hat{\lambda}_n^{\text{MS}}$)를 구하는 단계;

수학적 식 2

<23> $\hat{\lambda}_n(t, k) = B_{\min}(t, k) \lambda_{\min, n}(t, k)$

<24> (여기서, $\lambda_{\min, n}(t, k)$ 는 잡음신호 n의 분산 $\lambda_n(t, k) = E\{|N(t, k)|^2\}$ 의 최소값이고, $B_{\min}(t, k)$ 는 바이어스 보상 함수로서, 다음 수학적 식 3과 같이 근사화될 수 있다.

수학적 식 3

<25> $B_{\min}(t, k) \approx 1 + (D-1) \frac{2}{Q_{\text{eq}}(t, k)}$

<26> 여기서, D는 한 프레임당 스펙트럼의 수이며, $Q_{\text{eq}}(t, k)$ 는 다음 수학적 식 4와 같이 표현될 수 있다.

수학적 식 4

<27> $\frac{1}{Q_{\text{eq}}(t, k)} \approx \frac{\widehat{\text{var}} P(t, k)}{2 \hat{\lambda}_n^2(t-1, k)}$

<28> 여기서, $\widehat{\text{var}} P(t, k)$ 는 스무딩된 파워 스펙트럼 $P(t, k)$ 의 분산으로서, 다음 수학적 식 5 내지 수학적 식 7과 같이 구해질 수 있다.

수학적 식 5

<29> $\widehat{\text{var}} P(t, k) = \overline{P^2}(t, k) - \overline{P}^2(t, k)$

수학적 식 6

<30> $\overline{P}(t, k) = \beta(t, k) \overline{P}(t-1, k) + (1 - \beta(t, k)) P(t, k)$

수학적 식 7

<31> $\overline{P^2}(t, k) = \beta(t, k) \overline{P^2}(t-1, k) + (1 - \beta(t, k)) P^2(t, k)$

<32> 여기서, 스무딩 파라미터인 $\beta(t, k)$ 는 $\beta(t, k) = \alpha^2(t, k)$ 로서 주어지며, $\alpha(t, k)$ 는 다음 수학적 식 8과 같이 주어진다.

수학적 식 8

$$\alpha(t, k) = \frac{1}{1 + (P(t-1, k)/\lambda_n(t, k) - 1)^2}$$

(2) 다음 수학적 식 9에 의해 SD를 이용하여 잡음 전력의 제2 추정치($\hat{\lambda}_n^{SD}$)를 구하는 단계;

수학적 식 9

$$\hat{\lambda}_n(t+1, k) = \zeta_n \hat{\lambda}_n(t, k) + (1 - \zeta_n) E[|N(t, k)|^2 | Y(t, k)]$$

(여기서, ζ_n 은 $0 \leq \zeta_n \leq 1$ 의 범위를 갖는 스무딩 파라미터이며, $E[|N(t, k)|^2 | Y(t, k)]$ 은 다음 수학적 식 10 내지 수학적 식 12를 이용하여 구할 수 있다.

수학적 식 10

$$E[|N(t, k)|^2 | Y(t, k)] = E[|N(t, k)|^2 | Y(t, k), H_0] p(H_0 | Y(t, k)) + E[|N(t, k)|^2 | Y(t, k), H_1] p(H_1 | Y(t, k))$$

수학적 식 11

$$E[|N(t, k)|^2 | Y(t, k), H_0] = |Y(t, k)|^2$$

수학적 식 12

$$E[|N(t, k)|^2 | Y(t, k), H_1] = \left(\frac{\xi(t, k)}{1 + \xi(t, k)} \right) \hat{\lambda}_n(t, k) + \left(\frac{1}{1 + \xi(t, k)} \right)^2 |Y(t, k)|^2$$

여기서, $\xi(t, k)$ 는 선행 SNR(a priori SNR)로서, $\xi(t, k) \equiv \frac{\lambda_s(t, k)}{\lambda_n(t, k)}$ 로서 정의되며($\lambda_s(t, k)$ 와 $\lambda_n(t, k)$ 는 각각 음성신호와 잡음신호의 분산을 나타냄), $p(H_1 | Y(t, k)) = 1 - p(H_0 | Y(t, k))$ 이고, $p(H_0 | Y(t, k))$ 는 다음 수학적 식 13과 같이 구할 수 있다.

수학적 식 13

$$p(H_0 | Y(t, k)) = \frac{p(Y(t, k) | H_0) p(H_0)}{p(Y(t, k) | H_0) p(H_0) + p(Y(t, k) | H_1) p(H_1)} = \frac{1}{1 + q \Lambda(Y(t, k))}$$

여기서, q 는 $q = p(H_1)/p(H_0)$ 로서 주어지며, 우도비(likelihood ratio) $\Lambda(Y(t, k))$ 는 다음 수학적 식 14와 같다.

수학적 식 14

$$\Lambda(Y(t, k)) = \frac{p(Y(t, k) | H_1)}{p(Y(t, k) | H_0)} = \frac{1}{1 + \xi(t, k)} \exp \left[\frac{\gamma(t, k) \xi(t, k)}{1 + \xi(t, k)} \right]$$

여기서, $\xi(t, k)$ 는 선행 SNR(a priori SNR)로서 앞서 정의된 바 있고, $\gamma(t, k)$ 는 사후 SNR(a posteriori SNR)로

서 다음 수학적식 15와 같이 정의된다.

수학적식 15

$$\gamma(t, k) \equiv \frac{|Y(t, k)|^2}{\lambda_n(t, k)}$$

<45>

(3) 구해진 잡음 전력의 상기 제1 추정치($\hat{\lambda}_n^{MS}$) 및 상기 제2 추정치($\hat{\lambda}_n^{SD}$)를 이용하여, 다음 수학적식 16에 의해 잡음 전력의 추정치($\hat{\lambda}_n^{CNPE}$)를 구하는 단계

수학적식 16

$$\hat{\lambda}_n^{CNPE} = (1 - p(H_0 | Y(t, k))) \hat{\lambda}_n^{MS} + p(H_0 | Y(t, k)) \hat{\lambda}_n^{SD}$$

<47>

를 포함하는 것을 그 특징으로 한다.

<48>

본 발명의 다른 특징에 따른 잡음 전력 추정 방법은, 상기 $\alpha(t, k)$ 의 최대값을 0.96으로 제한하고, 상기 $\beta(t, k)$ 의 최대값을 0.8로 제한하며, 상기 D와 상기 ζ_n 의 값은 각각 128과 0.99로 설정하는 것을 그 특징으로 한다.

<49>

이하에서는 본 발명의 실시예를 도면을 참조하여 상세히 설명한다.

<50>

본 발명의 상세한 설명에서는, 먼저 기존의 MS에 기초한 잡음 전력 추정 방법과 기존의 SD에 기초한 잡음 전력 추정 방법을 각각 살펴본 후, 본 발명에 따른 새로운 잡음 전력 추정 방법을 설명한다. 마지막으로, 본 발명의 잡음 전력 추정 방법을 기존의 MS에 기초한 잡음 전력 추정 방법과 기존의 SD에 기초한 잡음 전력 추정 방법과 비교하도록 한다.

<51>

1. MS(Minimum Statistics)에 기초한 잡음 전력 추정 방법

<52>

잡음신호 n이 음성신호 x에 인가된 오염된 음성신호 y에 대하여, n, x, y 사이의 관계를 푸리에 변환을 통해 주파수 축에서 다음 수학적식 17과 같이 표현할 수 있다.

<53>

수학적식 17

$$Y(t, k) = X(t, k) + N(t, k)$$

<54>

여기서, $Y(t, k)$, $X(t, k)$, $N(t, k)$ 는 각각 y, x, n의 t번째 프레임에서의 k번째 주파수 성분이다.

<55>

MS 추정을 위한 최소 잡음신호를 얻기 위해 스무딩이 적용된 전력 스펙트럼 성분을 다음 수학적식 18과 같이 구한다.

<56>

수학적식 18

$$P(t, k) = \alpha P(t-1, k) + (1 - \alpha) |Y(t, k)|^2$$

<57>

여기서, $P(t, k)$ 는 반복적 스무딩으로 구해진 전력 스펙트럼 성분이다(기존의 MD 추정의 일예에서는 수식 내의 고정 스무딩 파라미터 α 를 0.85로 설정했다).

<58>

그러나 1에 가까운 스무딩 파라미터에 의해 추정된 $P(t, k)$ 는 넓은 동적 범위의 음성 구간을 가지게 되며 이는 부정확한 잡음 전력 추정을 초래할 수 있다. 이러한 문제를 해결하기 위해 최적화된 스무딩을 위한 가변 스무딩 파라미터 $\alpha_{opt}(t, k)$ 가 다음 수학적식 19와 같이 제시된다.

<59>

수학적식 19

<60>
$$\alpha_{opt}(t,k) = \frac{1}{1 + (P(t-1,k)/\lambda_n(t,k) - 1)^2} \quad (=수학식 8)$$

<61> 여기서, $\lambda_n(t,k) = E\{|N(t,k)|^2\}$ 는 잡음신호의 분산이며, $\alpha_{opt}(t,k)$ 는 $P(t-1,k)/\lambda_n(t,k)$ 가 1일 때 발생하는 교착상태를 고려하여 그 최대값을 0.96으로 제한한다.

<62> 가변 스무딩 파라미터를 적용하여 구해진 전력 스펙트럼의 최소값은 최종적인 잡음신호의 추정을 위해 다음과 수학식 20과 같이 바이어스 보상을 한다.

수학식 20

<63>
$$\hat{\lambda}_n(t,k) = B_{min}(t,k)\lambda_{min,n}(t,k) \quad (=수학식 2)$$

<64> 여기서, $\hat{\lambda}_n(t,k)$ 는 바이어스 보상이 적용된 추정된 잡음신호이고, $B_{min}(t,k)$ 는 바이어스 보상 함수이며, $\lambda_{min,n}(t,k)$ 는 최소 잡음 전력이다. 바이어스 보상 함수 $B_{min}(t,k)$ 는 다음 수학식 21과 같이 근사식으로 표현할 수 있다.

수학식 21

<65>
$$B_{min}(t,k) \approx 1 + (D-1) \frac{2}{Q_{eq}(t,k)} \quad (=수학식 3)$$

<66> 여기서, D는 한 프레임당 스펙트럼 개수로 128로 설정하며, $Q_{eq}(t,k)$ 는 다음 수학식 22와 같이 표현된다.

수학식 22

<67>
$$\frac{1}{Q_{eq}(t,k)} \approx \frac{\widehat{\text{var}} P(t,k)}{2\hat{\lambda}_n^2(t-1,k)} \quad (=수학식 4)$$

<68> 여기서, $\widehat{\text{var}} P(t,k)$ 는 스무딩된 전력 스펙트럼 $P(t,k)$ 의 분산으로 다음 수학식 23 내지 수학식 25와 같이 구해진다.

수학식 23

<69>
$$\widehat{\text{var}} P(t,k) = \overline{P^2}(t,k) - \overline{P}^2(t,k) \quad (=수학식 5)$$

수학식 24

<70>
$$\overline{P}(t,k) = \beta(t,k)\overline{P}(t-1,k) + (1-\beta(t,k))P(t,k) \quad (=수학식 6)$$

수학식 25

<71>
$$\overline{P^2}(t,k) = \beta(t,k)\overline{P^2}(t-1,k) + (1-\beta(t,k))P^2(t,k) \quad (=수학식 7)$$

<72> 여기서, 스무딩 파라미터인 $\beta(t,k)$ 는 $\beta(t,k) = \alpha^2(t,k)$ 로서 주어지며, 최대값으로 0.8을 갖는다.

<73> 2. SD(Soft Decision)에 기초한 잡음 전력 추정 방법

<74> 다음으로, SD에 기초한 잡음 전력 추정 방법에 대하여 살펴본다. 음성이 존재하는 경우와 음성이 존재하지 않는 경우 각각의 가정 H_0 과 H_1 에 대하여 오염된 음성은 다음 수학식 26과 같이 표현할 수 있다.

수학식 26

H_0 : speech absent: $Y(t,k) = N(t,k)$

<75> H_1 : speech present: $Y(t,k) = X(t,k) + N(t,k) \quad (=수학식 1)$

<76> 여기서, 잡음신호 $N(t,k)$ 와 음성신호 $X(t,k)$ 는 통계적으로 독립이라고 가정한다. 음성신호와 잡음신호가 복소가우시안 분포를 따른다고 가정하면, H_0 과 H_1 의 확률밀도함수는 다음 수학식 27 및 수학식 28과 같이 주어질 수 있다.

수학식 27

<77>
$$p(Y(t,k)|H_0) = \frac{1}{\pi\lambda_n(t,k)} \exp\left[-\frac{|Y(t,k)|^2}{\lambda_n(t,k)}\right]$$

수학식 28

<78>
$$p(Y(t,k)|H_1) = \frac{1}{\pi(\lambda_s(t,k) + \lambda_n(t,k))} \cdot \exp\left[-\frac{|Y(t,k)|^2}{\lambda_s(t,k) + \lambda_n(t,k)}\right]$$

<79> 여기서, $\lambda_s(t,k)$ 과 $\lambda_n(t,k)$ 는 각각 음성신호와 잡음신호의 분산을 나타내며, Bayes' rule에 의하여 각 주파수 채널별 음성부재확률인 SAP를 구하면 다음 수학식 29와 같다.

수학식 29

<80>
$$p(H_0|Y(t,k)) = \frac{p(Y(t,k)|H_0)p(H_0)}{p(Y(t,k)|H_0)p(H_0) + p(Y(t,k)|H_1)p(H_1)}$$

$$= \frac{1}{1 + qA(Y(t,k))} \quad (= \text{수학식 13})$$

<81> 여기서, $q = p(H_1)/p(H_0)$ 이고, 상기 수학식 27과 수학식 28을 상기 수학식 29에 대입하면 우도비(likelihood ratio) $A(Y(t,k))$ 는 다음 수학식 30과 같다.

수학식 30

<82>
$$A(Y(t,k)) = \frac{p(Y(t,k)|H_1)}{p(Y(t,k)|H_0)}$$

$$= \frac{1}{1 + \xi(t,k)} \exp\left[\frac{\gamma(t,k)\xi(t,k)}{1 + \xi(t,k)}\right] \quad (= \text{수학식 14})$$

<83> 여기서, $\gamma(t,k)$ 와 $\xi(t,k)$ 는 각각 사후 SNR(a posteriori SNR)과 선행 SNR(a priori SNR)로서 다음 수학식 31과 정의된다.

수학식 31

<84>
$$\gamma(t,k) \equiv \frac{|Y(t,k)|^2}{\lambda_n(t,k)}, \quad \xi(t,k) \equiv \frac{\lambda_s(t,k)}{\lambda_n(t,k)}$$

<85> SD에 기초한 Long-Term 스무딩된 전력 스펙트럼 $\hat{\lambda}_n(t,k)$ 는 다음 수학식 32와 같이 주어진다.

수학식 32

<86>
$$\hat{\lambda}_n(t+1,k) = c_n \hat{\lambda}_n(t,k) + (1 - c_n) E[|N(t,k)|^2 | Y(t,k)] \quad (= \text{수학식 9})$$

<87> 여기서, c_n 은 $0 < c_n < 1$ 의 범위를 갖는 스무딩 파라미터로서, 그 값을 0.99로 설정하며, $E[|N(t,k)|^2 | Y(t,k)]$ 은 주파수 채널별 SAP를 적용하여 다음 수학식 33 내지 수학식 35를 이용하여 구할 수 있다.

수학식 33

<88>
$$E[|N(t,k)|^2 |Y(t,k)] = E[|N(t,k)|^2 |Y(t,k), H_0] p(H_0 | Y(t,k)) + E[|N(t,k)|^2 |Y(t,k), H_1] p(H_1 | Y(t,k)) \quad (=수학식 10)$$

수학식 34

<89>
$$E[|N(t,k)|^2 |Y(t,k), H_0] = |Y(t,k)|^2 \quad (=수학식 11)$$

수학식 35

<90>
$$E[|N(t,k)|^2 |Y(t,k), H_1] = \left(\frac{\xi(t,k)}{1+\xi(t,k)} \right) \hat{\lambda}_n(t,k) + \left(\frac{1}{1+\xi(t,k)} \right)^2 |Y(t,k)|^2 \quad (=수학식 12)$$

<91> 3. 본 발명에 따른 잡음 전력 추정 방법 CNPE(Combined Noise Power Estimation)

<92> 도 1 및 도 2는 각각, 음성 구간 및 잡음 구간에서 MS에 기초한 잡음 전력 추정 방법과 SD에 기초한 잡음 전력 추정 방법에 의해 추정된 신호를 실제 음성 및 잡음과 비교한 결과를 나타내는 도면이다.

<93> 도 1은 베블 잡음(babble noise)이 SNR 5 dB로 인가된 잡음 음성에 대하여, 각각의 추정 방법에 의해 잡음이 제거된 음성신호를 실제의 깨끗한 음성신호와 음성 구간에서 비교한 도면이다. 도 1에서 확인할 수 있는 바와 같이, MS에 기초한 잡음 전력 추정 방법을 이용한 음성 향상 알고리즘에 의해 잡음이 제거된 음성신호가, 특히 음성 구간의 휴지 구간에서 SD에 기초한 잡음 전력 추정 방법을 이용한 음성 향상 알고리즘에 의해 잡음이 제거된 음성신호보다 본래의 깨끗한 음성신호에 더 가까운 것을 볼 수 있다. 이와 같은 결과는, MS에 기초한 잡음 전력 추정 방법이 음성신호의 짧은 휴지 기간의 최소값을 이용하여 잡음 전력을 추정하기 때문이다. 보다 구체적으로 설명하면, SD에 기초한 잡음 전력 추정 방법은 음성 부재 확률(SAP)에 근거한 스무딩을 이용하므로 음성이 지속되는 경우, 전력이 급격히 감소하는 휴지 구간에서는 이전 전력 스펙트럼 성분에 영향을 받는 스무딩에 의해서 정확한 잡음 전력을 추정하기가 어려운데 반하여, MS에 기초한 잡음 전력 추정 방법은 이러한 휴지 구간의 최소 잡음 전력을 추정한 상태에서 바이어스 보상이 되기 때문에 휴지 구간에서도 효과적으로 잡음 추정을 할 수 있기 때문이다.

<94> 도 2는 백색 가우시안 잡음(white gaussian noise; WGN)이 SNR 10 dB로 인가된 잡음 음성에 대하여, 각각의 추정 방법에 의해 추정된 잡음 전력을 실제의 잡음 전력과 비교한 도면이다. 도 2에서 확인할 수 있는 바와 같이, 잡음 구간에서는 음성 구간에서와 달리, SD에 기초한 잡음 전력 추정 방법에 의해 추정된 잡음 전력이 MS에 기초한 잡음 전력 추정 방법에 의해 추정된 잡음 전력보다 실제의 잡음 전력에 더 가까운 결과를 나타낸다. 이와 같은 결과는, 잡음 구간에서는, MS에 기초한 잡음 전력 추정 방법에서 추정된 최소 잡음 전력에 분산에 의한 바이어스 보상이 유사한 값을 보여서 추정 에러가 상대적으로 크지만, SD에 기초한 잡음 추정 방법은 SAP에 의해 확실히 잡음 구간이라고 판별이 되면 수학식 32에 기초하여 도 2에서 보듯이 Long-Term 스무딩된 형태로 실제 잡음 전력의 궤적을 그대로 따라가는 구조이기 때문이다.

<95> 본 출원에서는 도 1 및 도 2에서 관찰되는 상기 비교 결과에 기초하여, MS에 기초한 잡음 추정 방법과, SD에 기초한 잡음 추정 방법을 확률적으로 결합하는 방법을 새롭게 제안한다. 도 3은 본 발명의 일 실시예에 따른 잡음 추정 방법을 나타내는 도면이다. 도 3에 도시한 바와 같이, 본 발명의 일 실시예에 따른 잡음 추정 방법은, (1) 기존의 최소 통계법(MS)을 이용하여 잡음 전력의 제1 추정치를 구하는 단계(310)와, (2) 기존의 소프트 디시전법(SD)을 이용하여 잡음 전력의 제2 추정치를 구하는 단계(320)와, (3) 상기 단계 310과 상기 단계 320에서 구해진 잡음 전력의 상기 제1 추정치와 제2 추정치를 확률적으로 결합함으로써, 새로운 잡음 전력의 추정치를 구하는 단계(330)를 포함한다.

<96> 단계 310과 단계 320의 구체적인 구현 방법의 일례에 대해서는, 1. MS(Minimum Statistics)에 기초한 잡음 전력 추정 방법 단락 및 2. SD(Soft Decision)에 기초한 잡음 전력 추정 방법 단락에서 이미 상세히 설명한 바 있다. 이하에서는, 단계 330의 구체적인 구현 방법의 일례에 대하여 수학식들을 통하여 보다 상세히 설명하기로 한다.

<97> 본 출원에서는, 효과적인 잡음 추정을 위해 SD 추정에서 구해진 주파수 채널별 SAP를 가중치 파라미터로 적용하여 음성 구간에서는 MS에 의한 잡음 전력 추정치($\hat{\lambda}_n^{MS}$)에 비중을 두고 비음성 구간에서는 SD에 의한 잡음 전력 추정치($\hat{\lambda}_n^{SD}$)에 비중을 두도록 기존의 알고리즘을 확률적으로 결합하여 새로운 잡음 전력 추정치($\hat{\lambda}_n^{CNPE}$)를 구하는 결합 잡음 전력 추정(Combined Noise Power Estimation, CNPE) 알고리즘을 다음 수학식 36과 같이

제안한다.

수학식 36

<98>
$$\hat{\lambda}_n^{\text{CNPE}} = (1 - p(H_0 | Y(t, k))) \hat{\lambda}_n^{\text{MS}} + p(H_0 | Y(t, k)) \hat{\lambda}_n^{\text{SD}} \quad (= \text{수학식 16})$$

<99> 위와 같이 CNPE에 의하여 구해진 잡음 전력 추정치는 음성 향상의 주요 파라미터인 선행 SNR과 사후 SNR을 구하는데 사용되어 질 수 있다. 먼저, 선행 SNR을 구하기 위해 깨끗한 음성 신호 전력은 다음 수학식 37과 같이 추정된다.

수학식 37

<100>
$$\hat{\lambda}_s(t+1, k) = \zeta_s \hat{\lambda}_s(t, k) + (1 - \zeta_s) E[|X(t, k)|^2 | Y(t, k)]$$

<101> 여기서, ζ_s 은 $0 < \zeta_s < 1$ 의 범위를 갖는 스무딩 파라미터로서, 그 값이 0.98로 설정되고, 주파수 채널별 SAP 적용하면 $E[|X(t, k)|^2 | Y(t, k)]$ 은 다음 수학식 38과 같이 표현될 수 있다.

수학식 38

<102>
$$E[|X(t, k)|^2 | Y(t, k)] = E[|X(t, k)|^2 | Y(t, k), H_0] p(H_0 | Y(t, k)) + E[|X(t, k)|^2 | Y(t, k), H_1] p(H_1 | Y(t, k))$$

<103> 여기서,

수학식 39

<104>
$$E[|X(t, k)|^2 | Y(t, k), H_0] = 0$$

수학식 40

<105>
$$E[|X(t, k)|^2 | Y(t, k), H_1] = \left(\frac{1}{1 + \hat{\xi}(t, k)} \right) \hat{\lambda}_s(t, k) + \left(\frac{\hat{\xi}(t, k)}{1 + \hat{\xi}(t, k)} \right)^2 |Y(t, k)|^2$$

<106> 따라서 새로운 잡음 전력 추정치($\hat{\lambda}_n^{\text{CNPE}}$)가 적용된 새로운 선행 SNR $\hat{\xi}(t, k)$ 는 수학식 36 및 수학식 37을 이용하여 다음 수학식 41과 같이 다시 표현될 수 있다.

수학식 41

<107>
$$\hat{\xi}(t, k) = \frac{\hat{\lambda}_s(t, k)}{\hat{\lambda}_n^{\text{CNPE}}(t, k)}$$

<108> 다음으로, 사후 SNR도 다음 수학식 42와 같이 다시 주어질 수 있다.

수학식 42

<109>
$$\gamma(t, k) = \frac{|Y(t, k)|^2}{\hat{\lambda}_n^{\text{CNPE}}(t, k)}$$

4. 실험 결과

<111> 본 출원에서 제안된 잡음 전력 추정 방법에 기초한 음성 향상 알고리즘의 음질 평가를 위해 MOS(Mean Opinion Score) 테스트를 수행하였다. MOS 테스트는 각 음성 재료 샘플에 대해 1(나쁨)에서 5(우수함)까지의 등급을 매기는 청취자 그룹을 대상으로 시행되며, 그런 후에 점수의 평균을 내 MOS 점수를 구하게 된다. MOS 테스트를 위해 남성, 여성화자 각각이 5개의 문장을 발음하도록 한 총 10개의 음성 데이터를 한 프레임의 크기를 10ms에서 8kHz로 샘플링 하여 3가지 형태의 잡음이 부가된 오염된 음성을 사용하였다. 잡음은 NOISE-X92 데이터베이스

의 백색 가우시안 잡음(WGN), 배블 잡음, 차량 잡음(vehicle noise)을 사용하였으며, SNR을 5, 10, 15 dB로 달리하여 10명의 청취자에 대해 실험하였다. 다음 표 1은 다양한 잡음 환경과 SNR에 대한 각 알고리즘의 MOS 결과를 보여주고 있다.

표 1

<112>

잡음유형	방법	SNR(dB)		
		5	10	15
백색 가우시안 잡음	MS	1.57	2.20	2.69
	SD	1.78	2.34	2.91
	CNPE	1.86	2.51	2.91
배블 잡음	MS	2.40	2.98	3.39
	SD	2.27	2.90	3.42
	CNPE	2.45	3.02	3.70
차량 잡음	MS	3.68	4.19	4.42
	SD	3.87	4.25	4.59
	CNPE	4.03	4.46	4.73

표 1에서 알 수 있는 바와 같이, 대부분의 잡음 환경에 대해서 제안된 CNPE 방법이 MS에 기초한 추정 방법 및 SD에 기초한 추정 방법보다 음성 향상에 우수한 성능을 나타내었다.

이상 설명한 본 발명은 본 발명이 속한 기술분야에서 통상의 지식을 가진 자에 의하여 다양한 변형이나 응용이 가능하며, 본 발명에 따른 기술적 사상의 범위는 아래의 특허청구범위에 의하여 정해져야 할 것이다.

발명의 효과

본 발명의 잡음 전력 추정 방법에 따르면, 기존의 MS에 기초한 잡음 전력 추정 방법과 SD에 기초한 잡음 전력 추정 방법을 확률적으로 결합함으로써, 잡음 구간 및 깨끗한 음성 구간에서 모두 우수한 잡음 전력 추정 능력을 보일 수 있다.

도면의 간단한 설명

도 1 및 도 2는 각각, 음성 구간 및 잡음 구간에서 MS에 기초한 잡음 전력 추정 방법과 SD에 기초한 잡음 전력 추정 방법에 의해 추정된 신호를 실제 음성 및 잡음과 비교한 결과를 나타내는 도면으로서, 도 1은 배블 잡음(babble noise)이 SNR 5 dB로 인가된 잡음 음성에 대하여, 각각의 추정 방법에 의해 잡음이 제거된 음성신호를 실제의 깨끗한 음성신호와 음성 구간에서 비교한 도면이고, 도 2는 백색 가우시안 잡음(white gaussian noise; WGN)이 SNR 10 dB로 인가된 잡음 음성에 대하여, 각각의 추정 방법에 의해 추정된 잡음 전력을 실제의 잡음 전력과 비교한 도면.

도 3은 본 발명의 일 실시예에 따른 잡음 추정 방법을 나타내는 도면.

<도면 중 주요 부분에 대한 부호의 설명>

clean : 실제의 깨끗한 음성

true noise : 실제의 잡음

MS : MS에 기초한 잡음 전력 추정 방법에 의한 추정 결과

SD : SD에 기초한 잡음 전력 추정 방법에 의한 추정 결과

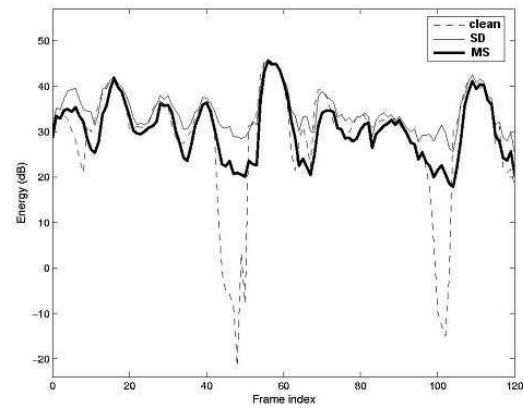
310 : MS에 기초하여 잡음 전력을 추정하는 단계

320 : SD에 기초하여 잡음 전력을 추정하는 단계

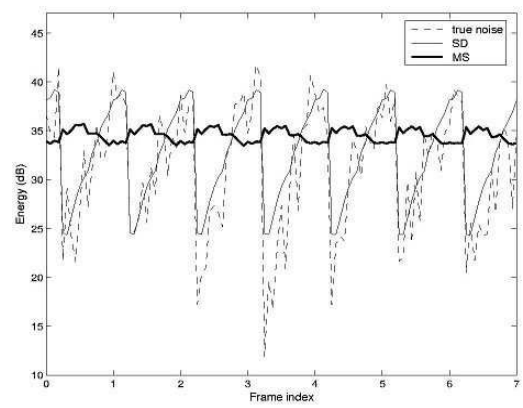
330 : MS에 기초한 잡음 전력과 SD에 기초한 잡음 전력을 확률적으로 결합하는 단계

도면

도면1



도면2



도면3

