



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2009년06월05일
(11) 등록번호 10-0901367
(24) 등록일자 2009년06월01일

(51) Int. Cl.

G10L 21/02 (2006.01)

(21) 출원번호 10-2008-0099002

(22) 출원일자 2008년10월09일

심사청구일자 2008년10월09일

(56) 선행기술조사문헌

KR1020080086298 A

KR1020010019603 A

KR100798056 B1

JP2004302470 A

전체 청구항 수 : 총 2 항

(73) 특허권자

인하대학교 산학협력단

인천 남구 용현동 253 인하대학교

(72) 발명자

장준혁

서울 서초구 서초동 1660-31 B동 301호

김종모

대전광역시 서구 월평동 다모아아파트 107동908호

(74) 대리인

특허법인태동

심사관 : 정성윤

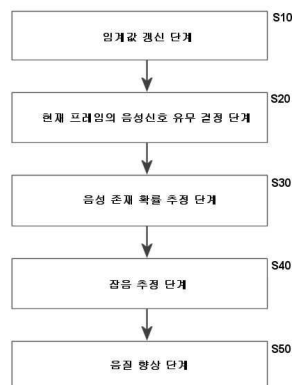
(54) 조건 사후 최대 확률 기반 최소값 제어 재귀평균기법을 이용한 음성 향상 방법

(57) 요약

본 발명은 조건 사후 최대 확률을 기반으로 한 최소값 제어 재귀평균기법을 이용한 음성 향상 방법에 관한 것으로, 보다 구체적으로는 (1) 이전 프레임의 음성신호의 유무에 따라 임계값을 변화시키는 제1 단계; (2) 음성과 잡음이 섞인 신호에서 국부 에너지와 국부 에너지의 최소값의 비를 상기 제1 단계에서 구한 임계값과 비교하여 현재 프레임의 음성신호 유무를 결정하는 제2 단계; (3) 상기 제2 단계에서 결정한 음성신호의 유무에 따라, 음성 존재 확률을 추정하는 제3 단계; (4) 상기 제3 단계에서 추정된 음성 존재 확률을 사용하여 잡음을 추정하는 제4 단계; 및 (5) 상기 제4 단계에서 추정된 잡음을 이용하여 잡음이 섞인 음성의 음질을 향상시키는 제5 단계를 포함하는 것을 그 구성상의 특징으로 한다.

본 발명의 음성 향상 방법에 따르면, 최소값 제어 재귀평균기법에 조건 사후 최대 확률을 추가함으로써, 즉 현재 프레임에 들어온 신호가 이전 프레임에서의 음성의 존재와 부재에 대한 조건을 부여해 주는 방식으로 음성 신호 존재 확률을 수정함으로써, 음성의 품질을 크게 향상시킬 수 있다.

대표도 - 도1



특허청구의 범위

청구항 1

- (1) 이전 프레임의 음성신호의 유무에 따라 임계값을 변화시키는 제1 단계;
- (2) 음성과 잡음이 섞인 신호에서 국부 에너지와 국부 에너지의 최소값의 비를 상기 제1 단계에서 구한 임계값과 비교하여 현재 프레임의 음성신호 유무를 결정하는 제2 단계;
- (3) 상기 제2 단계에서 결정한 음성신호의 유무에 따라, 음성 존재 확률을 추정하는 제3 단계;
- (4) 상기 제3 단계에서 추정한 음성 존재 확률을 사용하여 잡음을 추정하는 제4 단계; 및
- (5) 상기 제4 단계에서 추정한 잡음을 이용하여 잡음이 섞인 음성의 음질을 향상시키는 제5 단계를 포함하는 것을 특징으로 하는 음성 향상 방법.

청구항 2

제1항에 있어서,

상기 제1 단계에서, 이전 프레임의 음성신호의 유무에 따라 다음과 같은 수학적 식 a를 이용하여 임계값에 변화를 주는 것을 특징으로 하는 음성 향상 방법.

<수학적 식 a>

$$\frac{p(S_r|H_n=H_1)}{p(S_r|H_n=H_0)} > \frac{P(H_n=H_0|H_{n-1}=H_i)}{P(H_n=H_1|H_{n-1}=H_i)}, \quad i = 0, 1$$

만약 상기 수학적 식 a에서 이전 프레임에 음성구간이 있지 않을 경우에는, 다음 수학적 식 b와 같이 주어지며,

<수학적 식 b>

$$\frac{p(S_r|H_n=H_1)}{p(S_r|H_n=H_0)} > \frac{P(H_n=H_0|H_{n-1}=H_0)}{P(H_n=H_1|H_{n-1}=H_0)} = \beta'$$

이전 프레임에 음성구간이 있을 경우에는, 다음 수학적 식 c와 같이 주어진다.

<수학적 식 c>

$$\frac{p(S_r|H_n=H_1)}{p(S_r|H_n=H_0)} > \frac{P(H_n=H_0|H_{n-1}=H_1)}{P(H_n=H_1|H_{n-1}=H_1)} = \beta''$$

상기 수학적 식 b와 상기 수학적 식 c를 하나의 수학적 식으로 간략히 표현하면, 아래 수학적 식 d와 같이 주어진다.

<수학적 식 d>

$$\frac{S_r(k,l)}{H_0} > \gamma$$

여기서, $\gamma = \frac{P(H_{n-1}=H_0|Y_{n-1})\beta'}{(1-P(H_{n-1}=H_0|Y_{n-1}))\beta''}$ 로서 정의되고, $S_r=S_r(k, l)$ 은 국부 에너지와 국부 에너지의 최

소값의 비로서, $S(k, 1)/S_{\min}(k, 1)$ 로 정의되며, $S(k, 1)$ 및 $S_{\min}(k, 1)$ 은 각각 국부 에너지 및 국부 에너지의 최소값을 나타낸다. 또한, H_n 은 n 번째 프레임의 정확한 가정을 나타내며, H_0 , H_1 은 각각 음성 부재 및 음성 존재에 대한 가정을 나타낸다. 또한, α 는 바이어스된 것을 보상해 주기 위한 계수로서, $\alpha \geq 1$ 로서 그 범위가 주어진다.

명세서

발명의 상세한 설명

기술 분야

- <1> 본 발명은 음성 향상 방법에 관한 것으로서, 특히 최소값 제어 재귀평균기법에 조건 사후 최대 확률을 추가함으로써, 즉 현재 프레임에 들어온 신호가 이전 프레임에서의 음성의 존재와 부재에 대한 조건을 부여해 주는 방식으로 음성 신호 존재 확률을 수정함으로써, 음성의 품질을 크게 향상시킬 수 있는 음성 향상 방법에 관한 것이다.

배경 기술

- <2> 실제적인 음성 향상 시스템에서 잡음을 정확하게 추정하는 것은 가장 중요한 요소 중 하나이며, 특히 비정상 잡음 신호를 처리할 수 있어야 한다. 잡음 신호의 추정은 음성 향상 시스템의 품질에 미치는 영향이 크기 때문에 추정된 잡음 신호가 너무 작을 경우에는 자연스럽게 못한 잔류 잡음이 감지되며, 너무 클 경우에는 음성 신호가 둔탁하게 들리며 명료도가 떨어지게 된다. 전통적인 잡음 추정은 음성 부재 구간에서 잡음의 평균을 구하는 것이 가장 일반적인데, 이와 같은 음성 검출기(Voice Activity Detection; VAD)에 의존하는 추정은 조정하기가 어렵고 신호대 잡음비가 낮은 응용분야에 사용될 때에는 왜곡된 음성 신호를 출력하게 된다는 문제점이 있다. 최근 진행 중인 연구들은 주로 Soft decision 방식을 적용하여 음성 영역에서도 잡음 신호의 파워 스펙트럼을 추정하고 있으며, 음성 검출기를 사용하는 방법에 대한 대안으로 나온 최소 확률 잡음 신호 추정(Minimum Statistics Noise Estimation; MSNE) 기법은 VAD를 사용하지 않는 독창성을 가진다. 구체적으로 살펴보면, 최적으로 신호의 파워 스펙트럼을 스무딩하고 최소 확률을 적용하는 것을 기본으로 하며, 파워 스펙트럼 스무딩 알고리즘은 시간과 주파수에 종속 관계가 있는 스무딩 매개변수로 1차 회귀 시스템을 사용한다. 이때 스무딩 매개변수는 조건적인 평균 자승 오차를 최소로 하도록 최적화하여 비정상 잡음 신호를 추적한다. 그러나 이러한 잡음 추정은 특이점들에서만 민감하게 반응하고 전통적인 방법들보다 두 배 정도 큰 분산을 가진다. 더욱이 이 방법은 특히 최소 탐색 윈도우를 너무 작게 하면 작은 에너지를 갖는 음소를 더욱 약하게 만드는 단점이 있다. 계산량을 줄이고 효율적인 최소 추적 방법이 제안되기도 하였으나, 잡음 신호의 에너지가 급상승하는 경우엔 잡음 추정이 느려지고 신호를 소멸시키는 문제점이 있다.
- <3> 한편, 최근에 Cohen이 제안한 최소값 제어 재귀평균기법(minima controlled recursive averaging; MCRA)은 서브밴드에서 신호 존재 확률로 조절하는 스무딩 매개변수를 이용하여 파워 스펙트럼에 평균을 취하는 방법으로 이러한 단점을 보완한다. 각 서브밴드에서 신호의 존재는 잡음이 섞인 신호의 국부 에너지와 주어진 윈도우에서의 최소값 사이의 비로 정한다. 이 비율과 특정 임계값과 비교하여, 비율 값이 임계값보다 작으면 음성 신호가 없는 것으로 간주하는 방법을 사용하며, 음성 신호가 있는 부분과 없는 부분 사이에서 발생하는 변동을 줄이기 위해 시간축으로도 평균을 취한다. 이것은 음성 신호의 존재에 있어서 프레임 사이에 높은 상관성이 있기 때문이다.
- <4> 하지만 이러한 MCRA 알고리즘에도 몇 가지 문제점들이 있다. 즉, 효율적인 지역 최소 추적 기술을 사용하여 계산의 복잡성은 줄였지만 갑작스러운 잡음 존재 시 딜레이가 생기고, 또한 각 서브밴드에서 신호의 존재를 잡음이 섞인 신호의 국부 에너지와 주어진 윈도우에서의 최소값 사이의 비로 정한 비율과 특정 임계값만 가지고 비교를 하기 때문에 신뢰성이 떨어진다. 일반적으로, 음성의 활동은 인접한 프레임들과 강력한 상호 연관성이 있으므로, 음성이 활동하는 프레임의 바로 전 프레임이나 바로 다음 프레임은 음성이 활동할 가능성이 크다고 할 수 있고, 그 반대의 가능성도 맞다고 할 수 있다.
- <5> 이와 같은 연구 결과들을 고려해 볼 때, 기존의 통계적 가정을 바탕으로 Hidden Markov Model(HMM)을 이용한 행오버 알고리즘을 조건 사후 최대 확률(Maximum A Posteriori; MAP)에 적용하여 MCRA와 결합시킴으로써 잡음이 섞인 음성의 품질 향상을 시도해 볼 필요가 있다.

발명의 내용

해결 하고자하는 과제

- <6> 본 발명은, 상기와 같은 필요성의 인식에서 비롯된 것으로서, 특정 임계값만을 사용하여 각 서브밴드에서 음성 신호의 유무를 추정하는 기존의 MCRA 방법 대신에, 이전 서브밴드의 음성신호 유무에 대한 조건을 추가하여 더욱 신뢰성 있는 비교를 통해 음성 존재 확률의 성능을 향상시켜 우수한 잡음 추정 방법을 도출하고, 이를 이용하여 잡음 스펙트럼을 추정하고 또한 잡음이 섞인 음성의 품질을 향상시킬 수 있는 음성 향상 방법을 제안하는 것을 그 목적으로 한다.

과제 해결수단

- <7> 상기한 목적을 달성하기 위한 본 발명의 특징에 따른, 조건 사후 최대 확률 기반 최소값 제어 재귀평균기법을 이용한 음성 향상 방법은,
- <8> (1) 이전 프레임의 음성신호의 유무에 따라 임계값을 변화시키는 제1 단계;
- <9> (2) 음성과 잡음이 섞인 신호에서 국부 에너지와 국부 에너지의 최소값의 비를 상기 제1 단계에서 구한 임계값과 비교하여 현재 프레임의 음성신호 유무를 결정하는 제2 단계;
- <10> (3) 상기 제2 단계에서 결정한 음성신호의 유무에 따라, 음성 존재 확률을 추정하는 제3 단계;
- <11> (4) 상기 제3 단계에서 추정한 음성 존재 확률을 사용하여 잡음을 추정하는 제4 단계; 및
- <12> (5) 상기 제4 단계에서 추정한 잡음을 이용하여 잡음이 섞인 음성의 음질을 향상시키는 제5 단계를 포함하는 것을 그 구성상의 특징으로 한다.
- <13> 바람직하게는, 상기 제1 단계에서, 이전 프레임의 음성신호의 유무에 따라 다음과 같은 수학적식을 이용하여 임계값에 변화를 주는 것을 특징으로 한다.

$$\frac{p(S_r|H_n = H_1)}{p(S_r|H_n = H_0)} > \frac{P(H_n = H_0|H_{n-1} = H_i)}{P(H_n = H_1|H_{n-1} = H_i)}, \quad i = 0, 1$$

- <14> 만약 상기 수학적식에서 이전 프레임에 음성구간이 있지 않을 경우에는,

$$\frac{p(S_r|H_n = H_1)}{p(S_r|H_n = H_0)} > \frac{P(H_n = H_0|H_{n-1} = H_0)}{P(H_n = H_1|H_{n-1} = H_0)} = \beta'$$

- <15> 이전 프레임에 음성구간이 있을 경우에는,

$$\frac{p(S_r|H_n = H_1)}{p(S_r|H_n = H_0)} > \frac{P(H_n = H_0|H_{n-1} = H_1)}{P(H_n = H_1|H_{n-1} = H_1)} = \beta''$$

- <16> 위의 식을 간략히 표현하면, 아래 수학적식과 같이 주어짐.

$$S_r(k,l) > \gamma$$

$$\gamma = \frac{P(H_{n-1} = H_0|Y_{n-1})\beta' + (1 - P(H_{n-1} = H_0|Y_{n-1}))\beta''}{P(H_{n-1} = H_0|Y_{n-1})\beta' + (1 - P(H_{n-1} = H_0|Y_{n-1}))\beta''}$$

- <17> 여기서,

효 과

<22> 본 발명의 음성 향상 방법에 따르면, 최소값 제어 재귀평균기법에 조건 사후 최대 확률을 추가함으로써, 즉 현재 프레임에 들어온 신호가 이전 프레임에서의 음성의 존재와 부재에 대한 조건을 부여해 주는 방식으로 음성 신호 존재 확률을 수정함으로써, 음성의 품질을 크게 향상시킬 수 있다.

발명의 실시를 위한 구체적인 내용

- <23> 이하에서는 첨부된 도면들을 참조하여, 본 발명에 따른 실시예에 대하여 상세하게 설명하기로 한다.
- <24> 도 1은 본 발명의 일 실시예에 따른 음성 향상 방법의 구성을 나타내는 도면이다. 도 1에 도시된 바와 같이, 본 발명의 일 실시예에 따른 음성 향상 방법은, 임계값 갱신 단계(S10), 현재 프레임의 음성신호 유무 결정 단계(S20), 음성 존재 확률 추정 단계(S30), 잡음 추정 단계(S40), 및 음질 향상 단계(S50)를 포함한다.
- <25> 먼저, 임계값 갱신 단계(S10)에서는, 이전 프레임의 음성신호의 유무에 따라 임계값을 변화시킨다. 구체적으로 어떤 원리에 의하여, 어떤 수학적식을 적용하여 임계값을 변화시키는지의 추후 상세히 설명하기로 한다.
- <26> 다음으로, 현재 프레임의 음성신호 유무 결정 단계(S20)에서는, 음성과 잡음이 섞인 신호에서 국부 에너지와 국부 에너지의 최소값의 비를 단계 S10에서 구한 임계값과 비교하여 현재 프레임의 음성신호 유무를 결정하고, 음성 존재 확률 추정 단계(S30)에서는, 단계 S20에서 결정한 음성신호의 유무에 따라 음성 존재 확률을 추정하며, 잡음 추정 단계(S40)에서는, 단계 S30에서 추정한 음성 존재 확률을 사용하여 잡음을 추정한다.
- <27> 마지막으로, 음질 향상 단계(S50)에서는, 단계 S40에서 추정한 잡음을 이용하여 잡음이 섞인 음성의 음질을 향상시킨다.
- <28> 이하에서는, 먼저 잡음을 추정하기 위해 음성 존재 확률로 조정되는 시변 스무딩 매개변수를 구하기 위한 과정을 설명한 후, 음성 존재 확률을 구하기 위한 국부 에너지와 국부 에너지의 최소값의 비를 이전프레임의 음성의 유무에 따른 임계값과 비교를 하는 과정을 상세히 설명하기로 한다.
- <29> 1. 최소값 제어 재귀평균기법에 의한 잡음 추정
- <30> $x(n)$ 과 $d(n)$ 을 각각 음성신호와, 상관성이 없는 가산 잡음 신호라고 하자. 여기서, n 은 이산시간을 나타낸다. 이때 관측되는 신호 $y(n)$ 은 $y(n)=x(n)+d(n)$ 으로 주어지고, 부분적으로 중복으로 나누어 윈도우를 취한 다음 실 시간 푸리에 변환(short-time Fourier transform; STFT)을 이용하면 다음 수학적식 1과 같이 나타낼 수 있다.

수학적식 1

- <31> $Y(k, l) = X(k, l) + D(k, l)$
- <32> 여기서, $Y(k, l)$ 은 l 번째 프레임에서의 k 번째 주파수 성분이 된다. 음성 향상 기법에서 사용되고 있는 기본 가설을 음성의 부재와 존재 각각에 대해 $H_0(k, l)$ 과 $H_1(k, l)$ 이라고 하면, 다음 수학적식 2와 같이 나타낼 수 있다.

수학적식 2

- <33> $H_0(k, l): Y(k, l) = D(k, l)$
- <34> $H_1(k, l): Y(k, l) = X(k, l) + D(k, l)$
- <35> 여기서, $X(k, l)$ 와 $D(k, l)$ 는 각각 원래 음성 신호와 잡음 신호의 푸리에 변환 계수를 나타낸다. 여기서, 을 번째 서브밴드에서 잡음 신호의 분산이라고 하면, 추정하기 위해 음성 신호 부재 구간에서 관측된 신호에 시간의 반복 스무딩을 적용하면 다음 수학적식 3과 같이 나타낼 수 있다.

수학적식 3

- $H'_0(k, l): \hat{\lambda}_d(k, l+1) = \alpha_d \hat{\lambda}_d(k, l) + (1-\alpha_d) |Y(k, l)|^2$
- <36> $H'_1(k, l): \hat{\lambda}_d(k, l+1) = \hat{\lambda}_d(k, l)$
- <37> 여기서, $\alpha_d(0 < \alpha_d < 1)$ 는 스무딩 매개변수이며, H'_0 , H'_1 은 각각 가설에 근거하여 잡음 전력 갱신을 목적으로 한

음성 신호의 부재와 존재를 나타낸다. 음성 신호를 추정하는데 사용되는 수학적 2에서의 가설과 잡음 신호의 스펙트럼 갱신을 조절하는데 사용되는 수학적 3은 구별해야 한다. 즉, 음성신호가 존재(H_1)할 때 음성신호의 부재(H_0)라고 결정하는 것이 잡음 신호를 추정할 때보다 음성 신호를 추정할 때 더 위험스럽다. 그러므로 서로 다른 결정 법칙이 사용되고, 일반적으로 H_1 보다는 H_0 에 더 높은 신뢰를 두고 있으며, $P(H_1|Y) \geq P(H_0|Y)$ 인 것이다.

<38> $p(k,l)=P(H_1(k,l)|Y(k,l))$ 가 음성 존재의 조건 확률을 나타낸다고 하면, 수학적 3은 다음 수학적 4와 같다.

수학적 4

$$\begin{aligned}\hat{\lambda}_d(k,l+1) &= \hat{\lambda}_d(k,l)p(k,l) \\ &\quad + (\alpha_d \hat{\lambda}_d(k,l) + (1-\alpha_d)|Y(k,l)|^2)(1-p(k,l)) \\ &= \tilde{\alpha}_d(k,l)\hat{\lambda}_d(k,l) + (1-\tilde{\alpha}_d(k,l))|Y(k,l)|^2\end{aligned}$$

<39>

<40> 여기서, $\tilde{\alpha}_d$ 는 음성 존재 확률로 조정하는 시변 스무딩 매개변수로서 다음 수학적 5와 같이 주어진다.

수학적 5

$$\tilde{\alpha}_d(k,l) = \alpha_d + (1-\alpha_d)p(k,l)$$

<41>

<42> $p(k,l)$ 에 대한 추정은 다음 수학적 6을 사용한다.

수학적 6

$$\hat{p}(k,l) = \begin{cases} \alpha_p \hat{p}(k,l-1) + (1-\alpha_p) & \text{if } \frac{S(k,l)}{S_{\min}(k,l)} > \delta \\ \alpha_p \hat{p}(k,l-1) & \text{otherwise} \end{cases}$$

<43>

<44> 여기서 $\alpha_p(0 < \alpha_p < 1)$ 는 스무딩 매개변수이고, δ 는 음성신호 존재의 임계값이다.

<45> 특히, 국부 에너지는 다음 수학적 7과 같이 표시된다.

수학적 7

$$S(k,l) = \alpha_s S(k,l-1) + (1-\alpha_s)|Y(k,l)|^2$$

<46>

<47> 여기서, $\alpha_s(0 < \alpha_s < 1)$ 는 매개변수이다.

<48> 국부 에너지의 최소값 $S_{\min}(k,l)$ 과 임시변수 $S_{\text{tmp}}(k,l)$ 은 $S_{\min}(k,0)=S(k,0)$ 과 $S_{\text{tmp}}(k,0)=S(k,0)$ 으로 초기화한다. 그리고 다음 수학적 8과 같이, 국부 에너지와 이전 프레임의 최소값을 비교하여 현재 프레임에 대한 최소값을 구한다.

수학적 8

$$S_{\min}(k,l) = \min\{S_{\min}(k,l-1), S(k,l)\}$$

<49>

$$S_{\text{tmp}}(k,l) = \min\{S_{\text{tmp}}(k,l-1), S(k,l)\}$$

<50>

<51> L개의 프레임을 모두 처리하고 나면, 다음 수학적 9로 초기화시킨다.

수학적 9

$$S_{\min}(k,l) = \min\{S_{\text{tmp}}(k,l-1), S(k,l)\}$$

<52>

<53> $S_{\text{tmp}}(k,1) = S(k,1)$

<54> 상기 수학식 8 및 9의 과정을 반복함으로써, 다음 수학식 10과 같이 최소값을 탐색한다.

수학식 10

$$\begin{array}{c} H_1 \\ S_r(k,l) > \delta \\ H_0 \end{array}$$

<55>

<56> 여기서, $S_r(k,l) = \frac{S(k,l)}{S_{\min}(k,l)}$ 이다.

<57> 2. 조건 사후 확률기반 최소값 제어 재귀평균기법

<58> 전통적인 MAP 기준은, 다음 수학식 11과 같은 결정법을 따른다.

수학식 11

$$\begin{array}{c} H_1 \\ \frac{P(H_n = H_1 | Y)}{P(H_n = H_0 | Y)} > 1 \\ H_0 \end{array}$$

<59>

<60> 여기서, H_n 은 n 번째 프레임의 정확한 가정을 나타낸다. Bayes 법칙에 따르면 수학식 11의 기준을 다음 수학식 12와 같이 바꿀 수 있다.

수학식 12

$$\begin{array}{c} H_1 \\ \frac{p(Y|H_n = H_1)}{p(Y|H_n = H_0)} > \frac{P(H_n = H_0)}{P(H_n = H_1)} \\ H_0 \end{array}$$

<61>

<62> 바이어스된 것을 보상해 주기 위해, 상기 수학식 12를 변형시키면 다음 수학식 13과 같이 나타낼 수 있다.

수학식 13

$$\begin{array}{c} H_1 \\ \frac{p(Y|H_n = H_1)}{p(Y|H_n = H_0)} > \alpha \frac{P(H_n = H_0)}{P(H_n = H_1)} \\ H_0 \end{array}$$

<63>

<64> 여기서 $\alpha \geq 1$ 이다.

<65> 본 발명에서는, MCRA의 하나의 임계값만을 사용하여 각 서브밴드에서 신호의 존재를 추정할 때 이전 서브밴드의 신호가 음성을 포함하고 있는지 아닌지에 대한 조건을 추가하여 이 조건에 의해 두 가지의 임계값을 두어 더욱 신뢰성 있는 비교를 통해 신호 존재 확률의 성능을 더욱 좋게 한다. 다음은 이전 서브밴드의 신호가 음성을 포함하고 있는지 아닌지에 대한 조건을 추가한 새로운 음성 신호 존재 확률의 정의를 유도하는 과정이다.

<66> 상기 수학식 13을 Bayes의 룰에 의해 바꾸어 표현하면, 다음 수학식 14와 같다.

수학식 14

$$\frac{p(S_r|H_1)}{p(S_r|H_0)} > \frac{P(H_0)}{P(H_1)}$$

<67>

<68>

일반적으로, 음성의 활동은 인접한 프레임들과 강력한 상호 연관성이 있다. 즉, 음성이 활동하는 프레임의 바로 전 프레임이나 바로 다음 프레임은 음성이 활동할 가능성이 크다고 할 수 있고, 그 반대의 가능성도 맞다고 할 수 있다. 이러한 상호 연관성에 기초하여 HMM을 이용한 행 오버를 사용함으로써 통계 모델을 기반으로 한 VAD의 에러를 효과적으로 줄일 수 있다. 음성 활동 전후 프레임들의 강력한 상호 연관성의 성질들로 인해 기존의 사후 확률(a posterior probability) $P(H_n|X)$ 대신에 현재 프레임의 관찰 결과와 이전 프레임의 결정을 조건으로 한 사후 확률 $P(H_n|X, H_{n-1})$ 를 계산할 수 있다. 상기 수학식 14에 이전 프레임의 조건을 추가한 후 Bayes 법칙에 의해 다음 수학식 15와 같이 바꿀 수 있다.

수학식 15

$$\frac{p(H_n = H_1|S_r, H_{n-1} = H_i)}{p(H_n = H_0|S_r, H_{n-1} = H_i)} > \frac{P(H_0)}{P(H_1)}$$

<69>

<70>

상기 수학식 15는 다시 다음 수학식 16과 같이 바꿀 수 있다.

수학식 16

$$\frac{p(S_r|H_n = H_1, H_{n-1} = H_i)}{p(S_r|H_n = H_0, H_{n-1} = H_i)} > \frac{P(H_n = H_0|H_{n-1} = H_i)}{P(H_n = H_1|H_{n-1} = H_i)}, i = 0, 1$$

<71>

<72>

비록 현재 프레임의 음성 활동이 이전 프레임에 의존할지라도, 현재 프레임의 음성 활동은 현재 프레임에서 관찰된 잡음 섞인 음성 신호의 DFT 계수의 분포에 지배적인 영향을 받는다. 더욱이, $p(S_r|H_n=H_1, H_{n-1}=H_0)$ 와 $p(S_r|H_n=H_0, H_{n-1}=H_1)$ 에 분포된 매개변수들은 샘플 데이터의 부족으로 신뢰성 있는 추정이 불가하다. 이러한 이유로, 우리는 다음 수학식 17과 같이 간단한 가정을 할 수 있다.

수학식 17

<73>

$p(S_r|H_n=H_j, H_{n-1}=H_i) = p(S_r|H_n=H_j), i=0,1 j=0,1$

<74>

이를 이용하여 상기 수학식 16을 다음 수학식 18과 같이 간략화할 수 있다.

수학식 18

$$\frac{p(S_r|H_n = H_1)}{p(S_r|H_n = H_0)} > \frac{P(H_n = H_0)}{P(H_n = H_1)}, i = 0, 1$$

<75>

<76>

상기 수학식 18에서 만약 이전 프레임에 음성구간이 있지 않을 경우에는, 다음 수학식 19와 같이 되며,

수학식 19

$$\frac{p(S_r|H_n=H_1)}{p(S_r|H_n=H_0)} > \frac{P(H_n=H_0|H_{n-1}=H_0)}{P(H_n=H_1|H_{n-1}=H_0)} = \beta'$$

- <77>
- <78> 이전 프레임에 음성구간이 있을 경우에는, 다음 수학식 20과 같이 된다.

수학식 20

$$\frac{p(S_r|H_n=H_1)}{p(S_r|H_n=H_0)} > \frac{P(H_n=H_0|H_{n-1}=H_1)}{P(H_n=H_1|H_{n-1}=H_1)} = \beta''$$

- <79>
- <80> 이러한 임계값 개수의 증가는 VAD의 성능이 향상될 수 있도록 추가적인 자유도를 제공한다. 인접한 프레임들 간 음성 활동의 강력한 상호 연관성으로 인해 일반적으로 다음 수학식 21이 성립된다.

수학식 21

$$\frac{P(H_n=H_0|H_{n-1}=H_0)}{P(H_n=H_1|H_{n-1}=H_0)} > \frac{P(H_n=H_0|H_{n-1}=H_1)}{P(H_n=H_1|H_{n-1}=H_1)}$$

- <81>
- <82> 상기 수학식 21은 이전 프레임에서 음성 활동이 관측되었다면 현재 프레임에서 음성 활동이 관측될 확률이 높고 관측되지 않을 확률이 낮으며, 그 반대의 경우도 가능하기 때문에 결론적으로 β' 이 β'' 보다 크다는 것을 알 수 있다.
- <83> 상기 결과들을 정리해 보면, 다음 수학식 22와 같이 나타낼 수 있다.

수학식 22

$$S_r(k,l) > \gamma$$

- <84>
- <85> 여기서, γ 는 다음 수학식 23과 같이 주어진다.

수학식 23

$$\gamma = \frac{P(H_{n-1}=H_0|Y_{n-1})\beta' + (1 - P(H_{n-1}=H_0|Y_{n-1}))\beta''}{P(H_{n-1}=H_0|Y_{n-1})\beta' + (1 - P(H_{n-1}=H_0|Y_{n-1}))\beta''}$$

- <86>
- <87> 위와 같이, 본 발명에서는, 이전 프레임의 음성신호의 유무에 따라 임계값에 변화를 주어 이를 가지고 음성 존재 확률을 유동적으로 추정함으로써, 더 정확한 잡음 추정을 할 수가 있다.
- <88> 도 2는 MAP를 기반으로 한 임계값 테스트를 통한 음성 존재 확률을, F16 잡음(SNR=10dB)의 경우를 예를 들어 비교하여 보여주고 있는 도면이다. 도 2a는 깨끗한 음성 파형을 나타내는 도면이며, 도 2b는 실시간 프레임에서의 음성 존재 확률을 나타내는 도면이다. 도 2로부터, 본 발명에서 제안하고 있는 음성 향상 방법이, 기존의 MCRA 방법보다 음성이 시작하는 부분은 더 빨리 음성임을 알아내고, 음성이 끝나는 부분에서는 급격히 떨어지지 않아 음성임에도 불구하고 음성이 아니라고 판단해서 음성 정보를 잃어버리는 것을 줄여줌을 분명하게 확인할 수 있다.

- <89> 3. 실험 결과
- <90> 본 발명에서는 기존의 MCRA 알고리즘에서 각 서브밴드에서 신호의 존재를 잡음이 섞인 신호의 국부 에너지와 주

어진 윈도우에서의 최소값 사이의 비로 정해 이 비율과 특정 임계값만을 가지고 비교하기 때문에 신뢰성이 떨어지는 문제를 조건 MAP를 적용함으로써 임계값을 지속적으로 갱신하여 음성 존재 확률값을 더 신뢰할 수 있도록 하여 음성 향상을 유도하였다. 제안된 음성 향상 알고리즘의 음질 평가를 위해 널리 적용되고 있는 ITU-T P.862 PESQ, 주관적 음질 평가를 수행하였으며, 표 1과 표 2는 각각 추출된 PESQ수치와, 주관적 음질 평가 결과를 보여주고 있다.

표 1

Noise type	Method	SNR (dB)		
		5	10	15
White noise	MCRA	1.936	2.296	2.641
	Proposed	2.050	2.391	2.715
Babble noise	MCRA	2.379	2.697	2.972
	Proposed	2.397	2.724	3.004
F16 noise	MCRA	2.056	2.467	2.776
	Proposed	2.152	2.548	2.846

<91>

표 1의 PESQ 테스트를 위해 샘플은 남성, 여성 화자 각각이 100개의 문장을 발음하도록 한 음성을 한 프레임의 크기가 10ms에서 8kHz로 샘플링한 데이터에 네 가지 형태의 잡음이 추가되었다. 잡음은 NOISEX-92 데이터베이스의 white noise, babble noise, F16 noise를 사용하였으며, SNR을 5, 10, 15dB로 달리하여 테스트하였다. PESQ 값은 이들 샘플에 대한 평균 수치로 나타내었고, 기존 MCRA에 의한 PESQ를 위하여 가중치 파라미터는 $\alpha_d=0.95$, $\alpha_p=0.2$, $\alpha_s=0.45$ 로 각각 설정하였으며, 임계값 β' , β'' 는 다양한 잡음 환경에서 최적화된 실험치로 구하여 $\beta'=9$, $\beta''=3.5$ 로 설정하였다.

<92>

표 1로부터, 제안하고 있는 방법이 기존 방법에 비하여, 특히 white noise와 F16 noise의 5dB와 10 dB에서 두드러진 성능 향상을 보임을 확인할 수 있다. 이는 제안한 알고리즘이 도 1에서와 같이 낮은 SNR에서도 음성 존재 확률을 더욱 잘 추정함에 따라 개선된 음성 향상을 보이고 있음을 알 수 있다. babble noise의 경우, 잡음의 특성상 제안한 알고리즘의 음성 존재 확률이 기존과 비슷하지만 약간의 향상된 결과를 보여줄 수 있음을 확인할 수 있다.

<93>

표 2

Noise type	Method	SNR (dB)		
		5	10	15
White noise	MCRA	2.17	2.56	3.03
	Proposed	2.23	2.65	3.07
Babble noise	MCRA	2.93	3.32	3.69
	Proposed	2.95	3.42	3.76
F16 noise	MCRA	2.23	2.77	3.20
	Proposed	2.33	2.90	3.46

<94>

표 2의 주관적 음질 평가는 남성, 여성 화자 각각이 10개의 문장을 발음하도록 한 음성에 white, babble, F16 noise가 SNR이 5, 10, 15dB로 부가된 잡음 신호를 20명의 청취자를 대상으로 평가하였다. 표 2로부터, 제안한 방법이 기존 방법에 비하여, 세 가지 잡음 환경과 각 SNR에 대하여 모두 향상된 결과를 보여주고 있음을 확인할 수 있다. 결론적으로, 제안된 조건부 MAP를 기반으로 한 MCRA 알고리즘이 다양한 잡음 환경에서 우수함을 확인할 수 있다.

<95>

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.

<96>

도면의 간단한 설명

<97>

도 1은 본 발명의 일 실시예에 따른 음성 향상 방법의 구성을 나타내는 도면.

<98> 도 2는 F16 잡음(SNR=10dB)에서의 확률을 비교하여 나타내는 도면으로서, 도 2a는 깨끗한 음성 파형을 나타내는 도면이며, 도 2b는 실시간 프레임에서의 음성 존재 확률을 나타내는 도면.

<99> <도면 중 주요 부분에 대한 부호의 설명>

<100> S10: 임계값 갱신 단계

<101> S20: 현재 프레임의 음성신호 유무 결정 단계

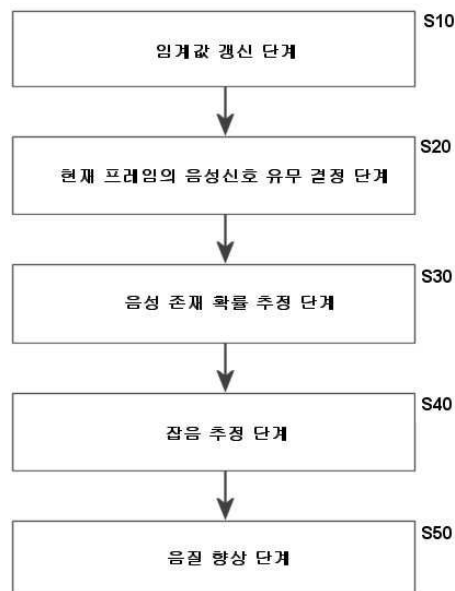
<102> S30: 음성 존재 확률 추정 단계

<103> S40: 잡음 추정 단계

<104> S50: 음질 향상 단계

도면

도면1



도면2

