



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0085732

(43) 공개일자 2013년07월30일

(51) 국제특허분류(Int. Cl.)

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

(21) 출원번호 10-2012-0006751

(22) 출원일자 2012년01월20일

심사청구일자 2012년01월20일

(71) 출원인

세종대학교산학협력단

서울특별시 광진구 능동로 209 (군자동, 세종대학교)

(72) 발명자

김무영

서울특별시 강남구 도곡2동 도곡렉슬아파트 210동 402호

이명석

서울특별시 구로구 오류동 68-21 미래빌라 1동 209호

(74) 대리인

특허법인태백

전체 청구항 수 : 총 16 항

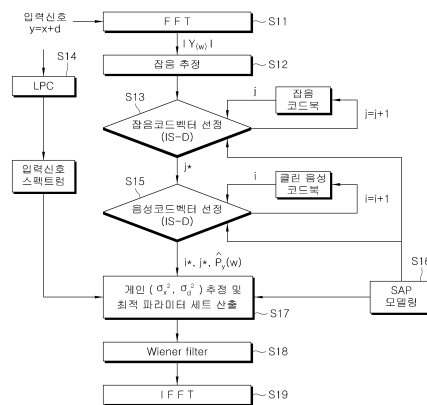
(54) 발명의 명칭 음성 존재 확률을 이용한 코드북 기반 음성 향상 방법 및 그 장치

(57) 요약

본 발명은 기존의 코드북 기반의 음성 향상 방법에 음성 존재 확률을 결합하여, 입력 신호에 음성이 존재하는 구간과 존재하지 않는 구간을 나눠 모델링하고, 이 확률 모델링을 통해, 개인 산정 및 코드북 산정을 진행한다.

본 발명에 따른 음성 존재 확률을 결합한 코드북 기반 음성 향상 방법은 학습된 잡음에 대해서는 기존의 IMCRA나 CDSTP에 비해 더 좋은 성능을 보여주며 학습되지 않는 잡음에 대해서도 기존의 CDSTP에 비해 뛰어난 성능을 보여 준다.

대표도 - 도1



특허청구의 범위

청구항 1

잡음을 포함하는 음성 신호에서 잡음을 제거하는 음성 향상 방법에 있어서,
 입력된 음성 신호를 시간 영역 신호에서 주파수 영역 신호로 변환시키는 단계;
 상기 입력된 음성 신호에 대해서 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하는 음성 부존재 확률(SAP)에 대한 바이너리 추정 모델을 산정하는 단계;
 상기 음성 신호의 잡음 코드북을 선정하는 단계;
 상기 음성 신호의 클린 음성 코드북을 선정하는 단계;
 상기 음성 부존재 확률을 통해 잡음의 클린 음성 부존재 시의 계인과 클린 음성 존재시의 계인을 산출하는 단계;
 상기 음성 부존재 확률을 통해 클린 음성과 잡음의 스펙트럼 형태와 계인을 파라미터 세트를 정하는 단계;
 상기 정해진 파라미터 세트를 이용하여 최대 유사도를 갖는 값을 파라미터 세트를 선정하는 단계;
 상기 선정된 최대 유사도 값을 기초로 바이너(wiener) 필터링을 이용하여 잡음을 제거하는 단계; 및
 상기 잡음이 제거된 음성 신호를 시간 영역 신호로 변환시키는 단계;
 를 포함하고,
 상기 잡음 코드북과 상기 클린 음성 코드북 선정 단계는, 상기 산정된 바이너리 추정 모델을 이용하여, 상기 입력된 음성 신호에 대해서 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하여, 각각 연산을 진행하여 구하며,
 상기 바이너리 모델 산정 단계는, 상기 입력 음성 신호에 클린 음성이 없는 잡음만 존재하는 음성 부존재 추정 모델과, 상기 입력 음성 신호에 클린 음성과 잡음이 존재하는 음성 존재 추정 모델로 구분하는 것을 특징으로 하는 음성 향상 방법

청구항 2

제 1 항에 있어서,
 상기 바이너리 모델은 하기 수학식으로 정의되며,

$$H_0 : |Y(\omega)|^2 = |D(\omega)|^2$$

$$H_1 : |Y(\omega)|^2 = |X(\omega)|^2 + |D(\omega)|^2$$

여기서 H_0 및 H_1 은 각각 음성이 부존재 할 때와 음성이 존재할 때의 추정 모델이며, 추정 모델 H_0 및 H_1 에 대한 확률 밀도 함수(PDF)는 $P(H_0 | y)$ 와 $P(H_1 | y)$ 로 정의되고,

상기 $P(H_0 | y)$ 와 $P(H_1 | y)$ 는 MS, WSA, IMCRA 를 포함하는 잡음 추정 알고리즘 중 어느 한 알고리즘을 통해 계산된 이전 및 상기 이전의 이전 SNR 값을 이용하여 추정하는 것을 특징으로 하는 음성 향상 방법.

청구항 3

제 1 항에 있어서,
 상기 잡음 코드북 선정 및 상기 클린 음성 코드북 선정 단계는, 원 입력된 음성 신호의 스펙트럼과 상기 잡음

코드북의 코드벡터와 상기 클린 음성 코드북의 코드벡터를 이용한 합성 신호 스펙트럼의 IS-D(Itakura-Saito Distortion) 알고리즘을 통해 선정하는 것을 특징으로 하는 음성 향상 방법.

청구항 4

제 1 항에 있어서,

상기 잡음 코드북 선정 및 상기 클린 음성 코드북 선정 단계는, 상기 음성 부존재 확률(SAP) 아래 수학식의 기반 최소 평균 제곱 오류(Minimum Mean Square Error; MMSE) 추정 방법을 통해 선정하며,

$$\begin{aligned}\hat{\theta} &= \sum_{H \in \{H_0, H_1\}} P(H | y) E[\theta | y, H] \\ &= \sum_{H \in \{H_0, H_1\}} p(H | y) \int \theta(H) \frac{p(\theta(H)) p(y | \theta(H))}{p(y | H)} d\theta(H)\end{aligned}$$

여기서 $\theta(H) = [\theta_x(H), \theta_d(H), \sigma_x^2(H), \sigma_d^2(H)]$ 이며, 주어진 모델 H의 파라미터 세트(Θ)이고 $\theta_x(H)$ 와 $\theta_d(H)$ 는 각각 클린 음성 및 잡음의 스펙트럼 형태의 항을 의미하며, $\sigma_x^2(H)$ 와 $\sigma_d^2(H)$ 는 각각 클린 음성과 잡음의 게인의 항이며

상기 파라미터 세트는, $\theta(H) = [\theta_x(H), \theta_d(H), \sigma_x^2(H), \sigma_d^2(H)]$ 로 표현되며, 상기 클린 음성 및 잡음의 스펙트럼 형태는

$$\theta_x(H) = \{a_{x,1}^H, \dots, a_{x,P}^H\}$$

$$\theta_d(H) = \{a_{d,1}^H, \dots, a_{d,Q}^H\}$$

로 표현되며,

상기 클린 음성과 잡음의 게인(gain) 항들($\sigma_x^2(H)$, $\sigma_d^2(H)$)은, 2개의 추정 모델 H_0 및 H_1 에 대해 각각,

$$\begin{aligned}\begin{bmatrix} \sigma_x^{2,ML}(H_0) \\ \sigma_d^{2,ML}(H_0) \end{bmatrix} &= \begin{bmatrix} 0 \\ \left\| |A_d^{H_0}(\omega)|^2 / |Y(\omega)|^2 \right\| \end{bmatrix} \\ \begin{bmatrix} \sigma_x^{2,ML}(H_1) \\ \sigma_d^{2,ML}(H_1) \end{bmatrix} &= C^{-1}D\end{aligned}$$

여기서

$$\begin{aligned}C &= \begin{bmatrix} \left\| 1/|Y(\omega)|^4 |A_x^{H_1}(\omega)|^4 \right\| & \left\| 1/|Y(\omega)|^4 |A_x^{H_1}(\omega)|^2 |A_d^{H_1}(\omega)|^2 \right\| \\ \left\| 1/|Y(\omega)|^4 |A_x^{H_1}(\omega)|^2 |A_d^{H_1}(\omega)|^2 \right\| & \left\| 1/|Y(\omega)|^4 |A_d^{H_1}(\omega)|^4 \right\| \end{bmatrix} \\ D &= \begin{bmatrix} \left\| 1/|Y(\omega)|^2 |A_x^{H_1}(\omega)|^2 \right\| \\ \left\| 1/|Y(\omega)|^2 |A_d^{H_1}(\omega)|^2 \right\| \end{bmatrix}\end{aligned}$$

$$\|\cdot\| = \int |\cdot| d\omega.$$

여기서

의 수학적식을 통해 산정되며 상기 수학적식에서 최대 값을 갖는 $\sigma_x^{2,ML}$ 와 $\sigma_d^{2,ML}$ 가 각각 클린 음성 및 잡음의 계인 ($\sigma_x^2(H)$, $\sigma_d^2(H)$) 이 되고,

상기 정해진 파라미터 세트를 이용하여 최대 유사도를 갖는 값을 파라미터 세트를 선정하는 단계는, 최소 평균 제곱 오류 산정에 대한 아래 수학적식을 통해 선정되며,

$$\hat{\theta} \approx \sum_{H \in \{H_0, H_1\}} p(H|y) \left\{ \frac{1}{N_x N_d} \sum_{i,j=1}^{N_x, N_d} \theta'_{ij}(H) \frac{p(\theta_{x,i}(H)) p(\theta_{d,j}(H)) p(\sigma_{x,ij}^{2,ML}(H)) p(\sigma_{d,ij}^{2,ML}(H)) p(y|\theta'_{ij}(H))}{p(y|H)} \right\}$$

이때,

$$p(y|H) = \frac{1}{N_x N_d} \sum_{i,j=1}^{N_x, N_d} p(\theta_{x,i}(H)) p(\theta_{d,j}(H)) p(\sigma_{x,ij}^{2,ML}(H)) p(\sigma_{d,ij}^{2,ML}(H)) p(y|\theta'_{ij}(H))$$

여기서,

$$p(y|\theta'_{ij}(H)) = C \exp \left(-d_{IS} \left(|Y(\omega)|^2, \frac{\sigma_{x,ij}^{2,ML}(H)}{|A_{x,i}^H(\omega)|^2} + \frac{\sigma_{d,ij}^{2,ML}(H)}{|A_{d,j}^H(\omega)|^2} \right) \right) \text{ 이고,}$$

$$\hat{\theta} = [\hat{\theta}_x, \hat{\theta}_d, \hat{\sigma}_x^2, \hat{\sigma}_d^2]$$

상기 최소 평균 제곱 오류 산정을 통해 얻은 추정 값이 파라미터 세트 경우, 바이너리(wiener) 필터링 단계는, 하기 수학적식을 통해 수행되며,

$$\frac{\hat{\sigma}_x^2}{|\hat{A}_x(\omega)|^2} / \left(\frac{\hat{\sigma}_x^2}{|\hat{A}_x(\omega)|^2} + \frac{\hat{\sigma}_d^2}{|\hat{A}_d(\omega)|^2} \right)$$

여기서 $\hat{A}_x(\omega)$ 와 $\hat{A}_d(\omega)$ 는 각각 추정된 파라미터 세트의 클린 음성과 잡음의 형태($\hat{\theta}_x$ 및 $\hat{\theta}_d$)에 의해 구성된 LPC 스펙트럼인 것을 특징으로 하는 음성 향상 방법.

청구항 5

잡음을 포함하는 음성 신호에서 잡음을 제거하는 음성 향상 방법에 있어서,

상기 입력된 음성 신호에 대해서 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하는 음성 부존재 확률(SAP)에 대한 바이너리 추정 모델을 산정하는 단계;

상기 음성 부존재 확률에 대한 바이너리 추정 모델을 기반으로 상기 음성 신호의 클린 음성 및 잡음 스펙트럼 형태를 구하는 단계; 및

상기 음성 신호에 대하여 상기 바이너리 추정 모델을 기반으로 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하여, 상기 클린 음성 부존재 시의 계인과 상기 클린 음성 존재시의 계인을 최대 유사도 산정을 통해 별도로 결정하는 단계;

상기 클린 음성 및 잡음의 스펙트럼 형태 및 계인들을 이용하여 CDSTP를 위한 최소 평균 제곱 오류(MMSE) 추정
을 통해 잡음을 제거하는 단계;

상기 CDSTP를 위한 최소 평균 제곱 오류 추정은 상기 입력 음성 신호와 유사한 잡음과 클린 음성의 코드북을 산
정하고 상기 산정된 코드북을 이용하여 상기 클린 음성 및 잡음의 형태 및 계인을 포함하는 파라미터 세트를 산
출하는 단계;를 포함하며,

상기 잡음 코드북과 상기 클린 음성 코드북 선정 단계는, 상기 산정된 바이너리 추정 모델을 이용하여, 상기 입
력된 음성 신호에 대해서 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하여, 각각 연산을
진행하여 구하는 것을 특징으로 하는 음성 향상 방법.

청구항 6

제 5 항에 있어서,

상기 바이너리 모델은 하기 수학식으로 정의되며,

$$H_0 : |Y(\omega)|^2 = |D(\omega)|^2$$

$$H_1 : |Y(\omega)|^2 = |X(\omega)|^2 + |D(\omega)|^2$$

여기서 H_0 및 H_1 은 각각 음성이 부존재 할 때와 음성이 존재할 때의 추정 모델이며, 추정 모델 H_0 및 H_1 에 대한
확률 밀도 함수(PDF)는 $P(H_0 | y)$ 와 $P(H_1 | y)$ 로 정의되고,

상기 잡음 코드북 선정 및 상기 클린 음성 코드북 선정 단계는, 원 입력된 음성 신호의 스펙트럼과 상기 잡음
코드북의 코드벡터와 상기 클린 음성 코드북의 코드벡터를 이용한 합성 신호 스펙트럼의 IS-D(Itakura-Saito
Distortion) 알고리즘을 통해 선정하는 것을 특징으로 하는 음성 향상 방법.

청구항 7

제 6 항에 있어서,

상기 잡음 코드북 선정 및 상기 클린 음성 코드북 선정 단계는, 상기 음성 부존재 확률(SAP) 아래 수학식의 기
반 최소 평균 제곱 오류(Minimum Mean Square Error; MMSE) 추정 방법을 통해 선정하며,

$$\hat{\theta} = \sum_{H \in \{H_0, H_1\}} P(H | y) E[\theta | y, H]$$

$$= \sum_{H \in \{H_0, H_1\}} p(H | y) \int \theta(H) \frac{p(\theta(H)) p(y | \theta(H))}{p(y | H)} d\theta(H)$$

여기서 $\theta(H) = [\theta_x(H), \theta_d(H), \sigma_x^2(H), \sigma_d^2(H)]$ 이며, 주어진 모델 H의 파라미터 세트(θ)이고 $\theta_x(H)$ 와
 $\theta_d(H)$ 는 각각 클린 음성 및 잡음의 스펙트럼 형태의 항을 의미하며, $\sigma_x^2(H)$ 와 $\sigma_d^2(H)$ 는 각각 클린 음성과 잡
음의 계인의 항이고,

상기 파라미터 세트는, $\theta(H) = [\theta_x(H), \theta_d(H), \sigma_x^2(H), \sigma_d^2(H)]$ 로 표현되며, 상기 클린 음성 및 잡음의 스
펙트럼 형태는

$$\theta_x(H) = \{a_{x,1}^H, \dots, a_{x,P}^H\}$$

$$\theta_d(H) = \{a_{d,1}^H, \dots, a_{d,Q}^H\}$$

로 표현되는 것을 특징으로 하는 음성 향상 방법.

청구항 8

제 7 항에 있어서,

상기 클린 음성과 잡음의 게인(gain) 항들($\sigma_x^2(H)$, $\sigma_d^2(H)$)은, 2개의 추정 모델 H_0 및 H_1 에 대해 각각,

$$\begin{bmatrix} \sigma_x^{2,ML}(H_0) \\ \sigma_d^{2,ML}(H_0) \end{bmatrix} = \begin{bmatrix} 0 \\ \left\| |A_d^{H_0}(\omega)|^2 / |Y(\omega)|^2 \right\| \end{bmatrix}$$

$$\begin{bmatrix} \sigma_x^{2,ML}(H_1) \\ \sigma_d^{2,ML}(H_1) \end{bmatrix} = C^{-1}D$$

여기서

$$C = \begin{bmatrix} \left\| 1/|Y(\omega)|^4 |A_x^{H_1}(\omega)|^4 \right\| & \left\| 1/|Y(\omega)|^4 |A_x^{H_1}(\omega)|^2 |A_d^{H_1}(\omega)|^2 \right\| \\ \left\| 1/|Y(\omega)|^4 |A_x^{H_1}(\omega)|^2 |A_d^{H_1}(\omega)|^2 \right\| & \left\| 1/|Y(\omega)|^4 |A_d^{H_1}(\omega)|^4 \right\| \end{bmatrix}$$

$$D = \begin{bmatrix} \left\| 1/|Y(\omega)|^2 |A_x^{H_1}(\omega)|^2 \right\| \\ \left\| 1/|Y(\omega)|^2 |A_d^{H_1}(\omega)|^2 \right\| \end{bmatrix}$$

여기서 $\|\cdot\| = \int |\cdot| d\omega$.

의 수학적식을 통해 산정되며 상기 수학적식에서 최대 값을 갖는 $\sigma_x^{2,ML}$ 와 $\sigma_d^{2,ML}$ 가 각각 클린 음성 및 잡음의 게인 ($\sigma_x^2(H)$, $\sigma_d^2(H)$)이 되고,

상기 정해진 파라미터 세트를 이용하여 최대 유사도를 갖는 값을 파라미터 세트를 선정하는 단계는, 최소 평균 제곱 오류 산정에 대한 아래 수학적식을 통해 선정되며,

$$\hat{\theta} \approx \sum_{H \in \{H_0, H_1\}} p(H|y) \left\{ \frac{1}{N_x N_d} \sum_{i,j=1}^{N_x, N_d} \theta'_{ij}(H) \frac{p(\theta_{x,i}(H))p(\theta_{d,j}(H))p(\sigma_{x,ij}^{2,ML}(H))p(\sigma_{d,ij}^{2,ML}(H))p(y|\theta'_{ij}(H))}{p(y|H)} \right\}$$

이때,

$$p(y|H) = \frac{1}{N_x N_d} \sum_{i,j=1}^{N_x, N_d} p(\theta_{x,i}(H))p(\theta_{d,j}(H))p(\sigma_{x,ij}^{2,ML}(H))p(\sigma_{d,ij}^{2,ML}(H))p(y|\theta'_{ij}(H))$$

여기서,

$$p(y | \theta'_{ij}(H)) = C \exp \left(-d_{IS} \left(|Y(\omega)|^2, \frac{\sigma_{x,ij}^{2,ML}(H)}{|A_{x,i}^H(\omega)|^2} + \frac{\sigma_{x,ij}^{2,ML}(H)}{|A_{d,j}^H(\omega)|^2} \right) \right) \quad \text{이고,}$$

$$\hat{\theta} = [\hat{\theta}_x, \hat{\theta}_d, \hat{\sigma}_x^2, \hat{\sigma}_d^2]$$

상기 최소 평균 제곱 오류 산정을 통해 얻은 추정 값이 파라미터 세트 우, 바이너(wiener) 필터링 단계는, 하기 수학적식을 통해 수행되며,

$$\frac{\hat{\sigma}_x^2}{|\hat{A}_x(\omega)|^2} / \left(\frac{\hat{\sigma}_x^2}{|\hat{A}_x(\omega)|^2} + \frac{\hat{\sigma}_d^2}{|\hat{A}_d(\omega)|^2} \right)$$

여기서 $\hat{A}_x(\omega)$ 와 $\hat{A}_d(\omega)$ 는 각각 추정된 파라미터 세트의 클린 음성과 잡음의 형태($\hat{\theta}_x$ 및 $\hat{\theta}_d$)에 의해 구성된 LPC 스펙트럼인 것을 특징으로 하는 음성 향상 방법.

청구항 9

잡음을 포함하는 음성 신호에서 잡음을 제거하는 음성 향상 장치에 있어서,

입력된 음성 신호를 시간 영역 신호에서 주파수 영역 신호로 변환시키는 FFT 부;

상기 입력된 음성 신호에 대해서 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하는 음성 부존재 확률(SAP)에 대한 바이너리 추정 모델을 산정하는 SAP 모델링부;

상기 음성 신호의 잡음 코드북을 선정하는 잡음 코드 벡터 선정부;

상기 음성 신호의 클린 음성 코드북을 선정하는 클린 음성 코드 벡터 선정부;

상기 음성 부존재 확률을 통해 잡음의 클린 음성 부존재 시의 계인과 클린 음성 존재시의 계인을 산출하고, 상기 음성 부존재 확률을 통해 클린 음성과 잡음의 스펙트럼 형태와 계인을 파라미터 세트를 정하고 이중 최대 유사도를 갖는 파라미터 세트 값을 선정하는 계인 추정 및 최적 파라미터 선정부;

상기 선정된 최대 유사도 값을 기초로 바이너(wiener) 필터링을 이용하여 잡음을 제거하는 바이너 필터링부; 및

상기 잡음이 제거된 음성 신호를 시간 영역 신호로 변환시키는 IFFT부;

를 포함하고,

상기 잡음 코드 벡터 선정부와 상기 클린 음성 코드 벡터 선정부는, 상기 산정된 바이너리 추정 모델을 이용하여, 상기 입력된 음성 신호에 대해서 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하여, 각각 연산을 진행하여 구하며,

상기 SAP 모델링부는, 상기 입력 음성 신호에 클린 음성이 없는 잡음만 존재하는 음성 부존재 추정 모델과, 상기 입력 음성 신호에 클린 음성과 잡음이 존재하는 음성 존재 추정 모델로 구분하는 것을 특징으로 하는 음성 향상 장치

청구항 10

제 9 항에 있어서,

상기 바이너리 모델은 하기 수학식으로 정의되며,

$$H_0 : |Y(\omega)|^2 = |D(\omega)|^2$$

$$H_1 : |Y(\omega)|^2 = |X(\omega)|^2 + |D(\omega)|^2$$

여기서 H_0 및 H_1 은 각각 음성이 부존재 할 때와 음성이 존재할 때의 추정 모델이며, 추정 모델 H_0 및 H_1 에 대한 확률 밀도 함수(PDF)는 $P(H_0 | y)$ 와 $P(H_1 | y)$ 로 정의되고,

상기 $P(H_0 | y)$ 와 $P(H_1 | y)$ 는 MS, WSA, IMCRA 를 포함하는 잡음 추정 알고리즘 중 어느 한 알고리즘을 통해 계산된 이전 및 상기 이전의 이전 SNR 값을 이용하여 추정하는 것을 특징으로 하는 음성 향상 장치.

청구항 11

제 9 항에 있어서,

상기 잡음 코드 벡터 선정부 및 상기 클린 음성 코드 벡터 선정부는, 원 입력된 음성 신호의 스펙트럼과 상기 잡음 코드부의 코드벡터와 상기 클린 음성 코드부의 코드벡터를 이용한 합성 신호 스펙트럼의 IS-D(Itakura-Saito Distortion) 알고리즘을 통해 선정하는 것을 특징으로 하는 음성 향상 장치.

청구항 12

제 9 항에 있어서,

상기 잡음 코드 벡터 선정부 및 상기 클린 음성 코드 벡터 선정부는, 상기 음성 부존재 확률(SAP) 아래 수학식의 기반 최소 평균 제곱 오류(Minimum Mean Square Error; MMSE) 추정 장치를 통해 선정하며,

$$\begin{aligned} \hat{\theta} &= \sum_{H \in \{H_0, H_1\}} P(H | y) E[\theta | y, H] \\ &= \sum_{H \in \{H_0, H_1\}} p(H | y) \int \theta(H) \frac{p(\theta(H)) p(y | \theta(H))}{p(y | H)} d\theta(H) \end{aligned}$$

여기서 $\theta(H) = [\theta_x(H), \theta_d(H), \sigma_x^2(H), \sigma_d^2(H)]$ 이며, 주어진 모델 H 의 파라미터 세트(θ)이고 $\theta_x(H)$ 와 $\theta_d(H)$ 는 각각 클린 음성 및 잡음의 스펙트럼 형태의 항을 의미하며, $\sigma_x^2(H)$ 와 $\sigma_d^2(H)$ 는 각각 클린 음성과 잡음의 게인의 항이며,

상기 파라미터 세트는, $\theta(H) = [\theta_x(H), \theta_d(H), \sigma_x^2(H), \sigma_d^2(H)]$ 로 표현되며, 상기 클린 음성 및 잡음의 스펙트럼 형태는

$$\theta_x(H) = \{a_{x,1}^H, \dots, a_{x,P}^H\}$$

$$\theta_d(H) = \{a_{d,1}^H, \dots, a_{d,Q}^H\}$$

로 표현되며,

상기 클린 음성과 잡음의 게인(gain) 항들($\sigma_x^2(H)$, $\sigma_d^2(H)$)은, 2개의 추정 모델 H_0 및 H_1 에 대해 각각,

$$\begin{bmatrix} \sigma_x^{2,ML}(H_0) \\ \sigma_d^{2,ML}(H_0) \end{bmatrix} = \begin{bmatrix} 0 \\ \left\| \left| A_d^{H_0}(\omega) \right|^2 / |Y(\omega)|^2 \right\| \right]$$

$$\begin{bmatrix} \sigma_x^{2,ML}(H_1) \\ \sigma_d^{2,ML}(H_1) \end{bmatrix} = C^{-1}D$$

여기서

$$C = \begin{bmatrix} \left\| 1/|Y(\omega)|^4 |A_x^{H_1}(\omega)|^4 \right\| & \left\| 1/|Y(\omega)|^4 |A_x^{H_1}(\omega)|^2 |A_d^{H_1}(\omega)|^2 \right\| \\ \left\| 1/|Y(\omega)|^4 |A_x^{H_1}(\omega)|^2 |A_d^{H_1}(\omega)|^2 \right\| & \left\| 1/|Y(\omega)|^4 |A_d^{H_1}(\omega)|^4 \right\| \end{bmatrix}$$

$$D = \begin{bmatrix} \left\| 1/|Y(\omega)|^2 |A_x^{H_1}(\omega)|^2 \right\| \\ \left\| 1/|Y(\omega)|^2 |A_d^{H_1}(\omega)|^2 \right\| \end{bmatrix}$$

$$\|\cdot\| = \int |\cdot| d\omega.$$

여기서

의 수학적식을 통해 산정되며 상기 수학적식에서 최대 값을 갖는 $\sigma_x^{2,ML}$ 와 $\sigma_d^{2,ML}$ 가 각각 클린 음성 및 잡음의 게인 ($\sigma_x^2(H)$, $\sigma_d^2(H)$)이 되고,

상기 게인 추정 및 최적 파라미터 선정부는, 최소 평균 제곱 오류 산정에 대한 아래 수학적식을 통해 선정되며,

$$\hat{\theta} \approx \sum_{H \in \{H_0, H_1\}} p(H|y) \left\{ \frac{1}{N_x N_d} \sum_{i,j=1}^{N_x, N_d} \theta'_{ij}(H) \frac{p(\theta_{x,i}(H))p(\theta_{d,j}(H))p(\sigma_{x,ij}^{2,ML}(H))p(\sigma_{d,ij}^{2,ML}(H))p(y|\theta'_{ij}(H))}{p(y|H)} \right\}$$

이때,

$$p(y|H) = \frac{1}{N_x N_d} \sum_{i,j=1}^{N_x, N_d} p(\theta_{x,i}(H))p(\theta_{d,j}(H))p(\sigma_{x,ij}^{2,ML}(H))p(\sigma_{d,ij}^{2,ML}(H))p(y|\theta'_{ij}(H))$$

여기서,

$$p(y|\theta'_{ij}(H)) = C \exp \left(-d_{IS} \left(|Y(\omega)|^2, \frac{\sigma_{x,ij}^{2,ML}(H)}{|A_{x,i}^H(\omega)|^2} + \frac{\sigma_{d,ij}^{2,ML}(H)}{|A_{d,j}^H(\omega)|^2} \right) \right) \text{ 이고,}$$

$$\hat{\theta} = [\hat{\theta}_x, \hat{\theta}_d, \hat{\sigma}_x^2, \hat{\sigma}_d^2]$$

상기 최소 평균 제곱 오류 산정을 통해 얻은 추정 값이 파라미터 세트 경우, 상기 바이너리 필터링부는, 하기 수학적식을 통해 수행되며,

$$\frac{\hat{\sigma}_x^2}{|\hat{A}_x(\omega)|^2} / \left(\frac{\hat{\sigma}_x^2}{|\hat{A}_x(\omega)|^2} + \frac{\hat{\sigma}_d^2}{|\hat{A}_d(\omega)|^2} \right)$$

여기서 $\hat{A}_x(\omega)$ 와 $\hat{A}_d(\omega)$ 는 각각 추정된 파라미터 세트의 클린 음성과 잡음의 형태($\hat{\theta}_x$ 및 $\hat{\theta}_d$)에 의해 구성된 LPC 스펙트럼인 것을 특징으로 하는 음성 향상 장치.

청구항 13

잡음을 포함하는 음성 신호에서 잡음을 제거하는 음성 향상 장치에 있어서,

상기 입력된 음성 신호에 대해서 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하는 음성 부존재 확률(SAP)에 대한 바이너리 추정 모델을 산정하는 SAP 모델링부;

상기 음성 부존재 확률에 대한 바이너리 추정 모델을 기반으로 상기 음성 신호의 클린 음성 및 잡음 스펙트럼 형태를 구하는 잡음 코드 벡터 및 클린 음성 코드 벡터 선정부; 및

상기 음성 신호에 대하여 상기 바이너리 추정 모델을 기반으로 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하여, 상기 클린 음성 부존재 시의 계인과 상기 클린 음성 존재시의 계인을 최대 유사도 산정을 통해 별도로 결정하는 계인 추정 및 최적 파라미터 선정부를 포함하며,

상기 잡음 코드 벡터 및 클린 음성 코드 벡터 선정부는,

잡음 코드 벡터 및 클린 음성 코드 벡터 선정부는 상기 클린 음성 및 잡음의 스펙트럼 형태 및 계인들을 이용하여 CDSTP를 위한 최소 평균 제곱 오류(MMSE) 추정을 통해 선정하고, 상기 산정된 바이너리 추정 모델을 이용하여, 상기 입력된 음성 신호에 대해서 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하여, 각각 연산을 진행하여 구하고

상기 계인 추정 및 최적 파라미터 선정부는, 상기 CDSTP를 위한 최소 평균 제곱 오류 추정은 상기 입력 음성 신호와 유사한 잡음과 클린 음성의 코드북을 산정하고 상기 산정된 코드북을 이용하여 상기 클린 음성 및 잡음의 형태 및 계인을 포함하는 파라미터 세트를 산출하는 것을 특징으로 하는 음성 향상 장치.

청구항 14

제 13 항에 있어서,

상기 바이너리 모델은 하기 수학식으로 정의되며,

$$H_0 : |Y(\omega)|^2 = |D(\omega)|^2$$

$$H_1 : |Y(\omega)|^2 = |X(\omega)|^2 + |D(\omega)|^2$$

여기서 H_0 및 H_1 은 각각 음성이 부존재 할 때와 음성이 존재할 때의 추정 모델이며, 추정 모델 H_0 및 H_1 에 대한 확률 밀도 함수(PDF)는 $P(H_0 | y)$ 와 $P(H_1 | y)$ 로 정의되며,

상기 잡음 코드 벡터 및 클린 음성 코드 벡터 선정부는, 원 입력된 음성 신호의 스펙트럼과 상기 잡음 코드북의 코드벡터와 상기 클린 음성 코드북의 코드벡터를 이용한 합성 신호 스펙트럼의 IS-D(Itakura-Saito Distortion) 알고리즘을 통해 선정하는 것을 특징으로 하는 음성 향상 장치.

청구항 15

제 14 항에 있어서,

상기 잡음 코드 벡터 및 클린 음성 코드 벡터 선정부는, 상기 음성 부존재 확률(SAP) 아래 수학식의 기반 최소 평균 제곱 오류(Minimum Mean Square Error; MMSE) 추정 장치를 통해 선정하며,

$$\begin{aligned}\hat{\theta} &= \sum_{H \in \{H_0, H_1\}} P(H | y) E[\theta | y, H] \\ &= \sum_{H \in \{H_0, H_1\}} p(H | y) \int \theta(H) \frac{p(\theta(H)) p(y | \theta(H))}{p(y | H)} d\theta(H)\end{aligned}$$

여기서 $\theta(H) = [\theta_x(H), \theta_d(H), \sigma_x^2(H), \sigma_d^2(H)]$ 이며, 주어진 모델 H의 파라미터 세트(θ)이고 $\theta_x(H)$ 와 $\theta_d(H)$ 는 각각 클린 음성 및 잡음의 스펙트럼 형태의 항을 의미하며, $\sigma_x^2(H)$ 와 $\sigma_d^2(H)$ 는 각각 클린 음성과 잡음의 게인의 항인 것을 특징으로 하는 음성 향상 장치.

청구항 16

제 15 항에 있어서,

상기 파라미터 세트는, $\theta(H) = [\theta_x(H), \theta_d(H), \sigma_x^2(H), \sigma_d^2(H)]$ 로 표현되며, 상기 클린 음성 및 잡음의 스펙트럼 형태는

$$\theta_x(H) = \{a_{x,1}^H, \dots, a_{x,P}^H\}$$

$$\theta_d(H) = \{a_{d,1}^H, \dots, a_{d,Q}^H\}$$

로 표현되고,

상기 클린 음성과 잡음의 게인(gain) 항들($\sigma_x^2(H)$, $\sigma_d^2(H)$)은, 2개의 추정 모델 H_0 및 H_1 에 대해 각각,

$$\begin{aligned}\begin{bmatrix} \sigma_x^{2,ML}(H_0) \\ \sigma_d^{2,ML}(H_0) \end{bmatrix} &= \begin{bmatrix} 0 \\ \left\| A_d^{H_0}(\omega) \right\|^2 / |Y(\omega)|^2 \end{bmatrix} \\ \begin{bmatrix} \sigma_x^{2,ML}(H_1) \\ \sigma_d^{2,ML}(H_1) \end{bmatrix} &= C^{-1} D\end{aligned}$$

여기서

$$C = \begin{bmatrix} \left\| \frac{1}{|Y(\omega)|^4} |A_x^{H_1}(\omega)|^4 \right\| & \left\| \frac{1}{|Y(\omega)|^4} |A_x^{H_1}(\omega)|^2 |A_d^{H_1}(\omega)|^2 \right\| \\ \left\| \frac{1}{|Y(\omega)|^4} |A_x^{H_1}(\omega)|^2 |A_d^{H_1}(\omega)|^2 \right\| & \left\| \frac{1}{|Y(\omega)|^4} |A_d^{H_1}(\omega)|^4 \right\| \end{bmatrix}$$

$$D = \begin{bmatrix} \left\| \frac{1}{|Y(\omega)|^2} |A_x^{H_1}(\omega)|^2 \right\| \\ \left\| \frac{1}{|Y(\omega)|^2} |A_d^{H_1}(\omega)|^2 \right\| \end{bmatrix}$$

여기서 $\|\cdot\| = \int |\cdot| d\omega$.

의 수학적식을 통해 산정되며 상기 수학적식에서 최대 값을 갖는 $\sigma_x^{2,ML}$ 와 $\sigma_d^{2,ML}$ 가 각각 클린 음성 및 잡음의 계인 ($\sigma_x^2(H)$, $\sigma_d^2(H)$)이 되며,

상기 계인 추정 및 최적 파라미터 선정부는, 최소 평균 제곱 오류 산정에 대한 아래 수학적식을 통해 상기 정해진 파라미터 세트를 이용하여 최대 유사도를 갖는 값을 파라미터 세트를 선정하며,

$$\hat{\theta} \approx \sum_{H \in \{H_0, H_1\}} p(H|y) \left\{ \frac{1}{N_x N_d} \sum_{i,j=1}^{N_x, N_d} \theta'_{ij}(H) \frac{p(\theta_{x,i}(H))p(\theta_{d,j}(H))p(\sigma_{x,ij}^{2,ML}(H))p(\sigma_{d,ij}^{2,ML}(H))p(y|\theta'_{ij}(H))}{p(y|H)} \right\}$$

이때,

$$p(y|H) = \frac{1}{N_x N_d} \sum_{i,j=1}^{N_x, N_d} p(\theta_{x,i}(H))p(\theta_{d,j}(H))p(\sigma_{x,ij}^{2,ML}(H))p(\sigma_{d,ij}^{2,ML}(H))p(y|\theta'_{ij}(H))$$

여기서,

$$p(y|\theta'_{ij}(H)) = C \exp \left(-d_{IS} \left(|Y(\omega)|^2, \frac{\sigma_{x,ij}^{2,ML}(H)}{|A_{x,i}^H(\omega)|^2} + \frac{\sigma_{d,ij}^{2,ML}(H)}{|A_{d,j}^H(\omega)|^2} \right) \right) \text{ 이고,}$$

상기 최소 평균 제곱 오류 산정을 통해 얻은 추정 값이 파라미터 세트 $\hat{\theta} = [\hat{\theta}_x, \hat{\theta}_d, \hat{\sigma}_x^2, \hat{\sigma}_d^2]$ 로 계산되는 경우, 바이너 필터링부는, 하기 수학적식을 통해 수행되며,

$$\frac{\hat{\sigma}_x^2}{|\hat{A}_x(\omega)|^2} / \left(\frac{\hat{\sigma}_x^2}{|\hat{A}_x(\omega)|^2} + \frac{\hat{\sigma}_d^2}{|\hat{A}_d(\omega)|^2} \right)$$

여기서 $\hat{A}_x(\omega)$ 와 $\hat{A}_d(\omega)$ 는 각각 추정된 파라미터 세트의 클린 음성과 잡음의 형태($\hat{\theta}_x$ 및 $\hat{\theta}_d$)에 의해 구성된 LPC 스펙트럼인 것을 특징으로 하는 음성 향상 장치.

명세서

기술분야

- [0001] 본 발명은 음성 신호 처리 방법 및 그 장치에 관한 것이다. 구체적으로는 코드북 기반으로 음성에 혼합되어있는 잡음을 제거하여 향상된 음성을 얻기 위한 방법 및 그 장치에 관한 것이다.

배경 기술

- [0002] 음성향상 기법은 휴대기기를 이용한 통신 및 음악정보 처리 분야는 물론이고 로봇제어를 위한 음성인식 등 다양한 분야에서 필요로 하는 기술이다. 음성향상 기법의 성능을 향상시키기 위해서는 잡음을 제거하는 기술이 중요하다.
- [0003] 잡음을 제거하는 기술에는 잡음으로 오염된 음성이 깨끗한 음성과 배경 잡음의 합이라는 가정에서 시작된 스펙트럼 제거(spectral subtraction; SS) 방법이 있다. 또한 잡음의 변화가 주파수 밴드 별로 독립적이라는 가정을 이용하여 성능을 개선한 MBSS(multi-bank spectral subtraction) 방법 등이 있다. 또한 잡음 추정 알고리즘은 음성 인식(voice activity detection; VAD)을 이용한 방법이나 최소 통계(Minimum statistics; MS)를 이용한 방법 등 다양한 알고리즘이 연구되어 있다. 또한 단기 예측계수 코드북 기법(Codebook Driven Short-Term Predictor parameter estimation; CDSTP)과 같은 방법은 음성과 잡음의 선형 예측 계수를 데이터베이스로 이용하고, 입력된 음성신호와와의 최대 유사도(Maximum Likelihood; ML) 또는 최소 평균 제곱 오류(Minimum Mean Square Error; MMSE) 추정을 통하여 잡음을 제거하고 향상된 음성 신호를 얻는다. 이 기법은 코드북에 포함된 다양한 선형 예측 계수를 이용하기 때문에 정상(stationary) 배경 잡음 환경뿐만 아니라 비정상(non-stationary) 배경 잡음 환경에서도 뛰어난 성능을 보인다.

- [0004] MS(Minimum statistics), WSA(Weighted spectral averaging), IMCRA(improve minima control recursive average) 등은 이전의 프레임의 정보를 이용하여 잡음 추정하여 급변하는 잡음환경에서 취약점을 보인다. 또한 코드북 기반의 잡음 추정방법의 경우에는 급변하는 잡음 환경에서는 강인하지만, 학습되지 않은 잡음에 대하여 취약점을 보인다.

- [0005] 이에 따라 급변하는 잡음환경에서도 강인하며 학습되지 않은 잡음에도 강인한 잡음 추정 알고리즘을 구현하는 개선 방법이 요구되었다.

발명의 내용

해결하려는 과제

- [0006] 본 발명의 목적은 급변하는 잡음 환경과 학습되지 않은 잡음에도 강인한 알고리즘을 구현하는 음성 향상 방법을 제공하는 것이다.

과제의 해결 수단

- [0007] 상기한 바와 같은 목적을 달성하기 위한 본 발명의 사상에 따르면, 음성 향상 방법은, 입력된 음성 신호를 시간 영역 신호에서 주파수 영역 신호로 변환시키는 단계, 상기 입력된 음성 신호에 대해서 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하는 음성 부존재 확률(SAP)에 대한 바이너리 추정 모델을 산정하는 단계, 상기 음성 신호의 잡음 코드북을 선정하는 단계, 상기 음성 신호의 클린 음성 코드북을 선정하는 단계, 상기 음성 부존재 확률을 통해 잡음의 클린 음성 부존재 시의 계인과 클린 음성 존재시의 계인을 산출하는 단계, 상기 음성 부존재 확률을 통해 클린 음성과 잡음의 스펙트럼 형태와 계인을 파라미터 세트를 정하는 단계, 상기 정해진 파라미터 세트를 이용하여 최대 유사도를 갖는 값을 파라미터 세트를 선정하는 단계, 상기 선정된 최대 유사도 값을 기초로 바이너리(wiener) 필터링을 이용하여 잡음을 제거하는 단계, 및 상기 잡음이 제거된 음성 신호를 시간 영역 신호로 변환시키는 단계를 포함한다.
- [0008] 바람직하게는, 상기 잡음 코드북과 상기 클린 음성 코드북 선정 단계는, 상기 산정된 바이너리 추정 모델을 이용하여, 상기 입력된 음성 신호에 대해서 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하여, 각각 연산을 진행하여 구한다.
- [0009] 더욱 바람직하게는, 상기 바이너리 모델 산정 단계는, 상기 입력 음성 신호에 클린 음성이 없는 잡음만 존재하는 음성 부존재 추정 모델과, 상기 입력 음성 신호에 클린 음성과 잡음이 존재하는 음성 존재 추정 모델로 구분

한다.

- [0010] 더욱 바람직하게는, 상기 잡음 코드북 선정 및 상기 클린 음성 코드북 선정 단계는, 원 입력된 음성 신호의 스펙트럼과 상기 잡음 코드북의 코드벡터와 상기 클린 음성 코드북의 코드벡터를 이용한 합성 신호 스펙트럼의 IS-D(Itakura-Saito Distortion) 알고리즘을 통해 선정한다.
- [0011] 본 발명의 다른 특징에 의하면, 잡음을 포함하는 음성 신호에서 잡음을 제거하는 음성 향상 방법은, 상기 입력된 음성 신호에 대해서 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하는 음성 부존재 확률(SAP)에 대한 바이너리 추정 모델을 산정하는 단계, 상기 음성 부존재 확률에 대한 바이너리 추정 모델을 기반으로 상기 음성 신호의 클린 음성 및 잡음 스펙트럼 형태를 구하는 단계, 및 상기 음성 신호에 대하여 상기 바이너리 추정 모델을 기반으로 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하여, 상기 클린 음성 부존재 시의 계인과 상기 클린 음성 존재시의 계인을 최대 유사도 산정을 통해 별도로 결정하는 단계를 포함한다.
- [0012] 일 실시예에서 상기 음성 향상 방법은, 상기 클린 음성 및 잡음의 스펙트럼 형태 및 계인들을 이용하여 CDSTP를 위한 최소 평균 제곱 오류(MMSE) 추정을 통해 잡음을 제거하는 단계를 더 포함한다.
- [0013] 다른 실시예에서, 상기 음성 향상 방법은, 상기 클린 음성 및 잡음의 스펙트럼 형태 및 계인들을 이용하여 상기 입력된 음성 신호와의 최대 유사도 추정을 통해 잡음을 제거하는 단계를 더 포함한다.
- [0014] 바람직하게는, 상기 CDSTP를 위한 최소 평균 제곱 오류 추정은 상기 입력 음성 신호와 유사한 잡음과 클린 음성의 코드북을 산정하고 상기 산정된 코드북을 이용하여 상기 클린 음성 및 잡음의 형태 및 계인을 포함하는 파라미터 세트를 산출하는 단계를 더 포함한다.
- [0015] 본 발명의 또 다른 특징에 의하면, 잡음을 포함하는 음성 신호에서 잡음을 제거하는 음성 향상 장치는, 상기 입력된 음성 신호에 대해서 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하는 음성 부존재 확률(SAP)에 대한 바이너리 추정 모델을 산정하는 SAP 모델링부, 상기 음성 부존재 확률에 대한 바이너리 추정 모델을 기반으로 상기 음성 신호의 클린 음성 및 잡음 스펙트럼 형태를 구하는 잡음 코드 벡터 및 클린 음성 코드 벡터 선정부, 및 상기 음성 신호에 대하여 상기 바이너리 추정 모델을 기반으로 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하여, 상기 클린 음성 부존재 시의 계인과 상기 클린 음성 존재시의 계인을 최대 유사도 산정을 통해 별도로 결정하는 계인 추정 및 최적 파라미터 선정부를 포함하며, 상기 잡음 코드 벡터 및 클린 음성 코드 벡터 선정부는, 잡음 코드 벡터 및 클린 음성 코드 벡터 선정부는 상기 클린 음성 및 잡음의 스펙트럼 형태 및 계인들을 이용하여 CDSTP를 위한 최소 평균 제곱 오류(MMSE) 추정을 통해 선정하고, 상기 계인 추정 및 최적 파라미터 선정부는, 상기 CDSTP를 위한 최소 평균 제곱 오류 추정은 상기 입력 음성 신호와 유사한 잡음과 클린 음성의 코드북을 산정하고 상기 산정된 코드북을 이용하여 상기 클린 음성 및 잡음의 형태 및 계인을 포함하는 파라미터 세트를 산출하고, 상기 잡음 코드 벡터 및 클린 음성 코드 벡터 선정부는, 상기 산정된 바이너리 추정 모델을 이용하여, 상기 입력된 음성 신호에 대해서 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하여, 각각 연산을 진행하여 구한다.

$$H_0 : |Y(\omega)|^2 = |D(\omega)|^2$$

$$H_1 : |Y(\omega)|^2 = |X(\omega)|^2 + |D(\omega)|^2$$

- [0016] 일 실시예에서, 상기 바이너리 모델은 하기 수학식으로 정의되며,
- [0017] 여기서 H_0 및 H_1 은 각각 음성이 부존재 할 때와 음성이 존재할 때의 추정 모델이며, 추정 모델 H_0 및 H_1 에 대한 확률 밀도 함수(PDF)는 $P(H_0 | y)$ 와 $P(H_1 | y)$ 로 정의되며, 상기 잡음 코드 벡터 및 클린 음성 코드 벡터 선정부는, 원 입력된 음성 신호의 스펙트럼과 상기 잡음 코드북의 코드벡터와 상기 클린 음성 코드북의 코드벡터를 이용한 합성 신호 스펙트럼의 IS-D(Itakura-Saito Distortion) 알고리즘을 통해 선정된다.

발명의 효과

- [0018] 본 발명에 따른 음성 향상 방법은, 음성 존재 확률을 코드북 기반의 음성 향상 알고리즘에 적용하여 급변하는 잡음 환경에서도 강인할 뿐만 아니라 학습되지 않는 잡음에 대해서도 강인한 효과를 얻을 수 있다.

도면의 간단한 설명

- [0019] 도 1은 본 발명의 실시예에 따른 음성 인식 향상 방법의 프로세스를 나타낸다.
 도 2는 본 발명의 실시예에 따른 음성 인식 향상 장치의 구성을 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0020] 기타 실시예들의 구체적인 사항들은 상세한 설명 및 도면들에 포함되어 있다.
- [0021] 본 발명의 이점 및, 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 후술 되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 것이며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하며, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다. 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성 요소를 지칭한다.
- [0022] 도 1은 본 발명의 실시예에 따른 음성 인식 향상 방법의 프로세스를 나타낸다.
- [0023] 도 1을 참조하면, 입력되는 신호는 신호 처리를 위해 신호 영역의 음성 신호를 주파수 영역의 신호로 바꾸기 위해 고속 푸리에 변환(Fast Fourier Transform; FFT)을 수행한다(S11). 입력되는 음성 신호(y)는 클린 음성(x)과 잡음(d)이 섞여서 들어온다. 그런 다음 잡음 추정 단계(S12)를 거쳐 입력된 음성 신호(x+d)에서 잡음(d)을 추정하게 된다.
- [0024] 한편, 잡음으로 오염된 $y(n)$ 이 클린 음성 $x(n)$ 와 잡음 $d(n)$ 의 합으로 이루어졌다고 추정하면 다음과 같이 표현할 수 있다.

수학식 1

$$y(n) = x(n) + d(n)$$

- [0025] 여기서 $x(n)$ 은 클린 음성 신호이며 $d(n)$ 은 잡음 신호이다.
- [0026] 이를, 주파수 파워 스펙트럼으로 보면

수학식 2

$$|Y(\omega)|^2 = |X(\omega)|^2 + |D(\omega)|^2$$

- [0027] 으로 표시될 수 있다.
- [0028] 여기서, $|X(\omega)|^2$ 및 $|D(\omega)|^2$ 는 각각 클린 음성 신호와 잡음 신호의 파워 스펙트럼이며, ω 는 주파수 값이다.
- [0029] 이러한 추정에 기초하면, $D(\omega)$ 가 완전하게 특정되기만 한다면 $X(\omega)$ 를 완벽하게 재구성될 수 있을 것이다. 따라서 잡음 추정은 음성 향상에 있어서 결정적인 부분이라고 할 수 있다.
- [0030] 잡음 추정 방법으로는 MS(Minimum statistics), WSA(Weighted spectral averaging), IMCRA(improve minima control recursive average), CDSTP(Codebook driven short-term parameter estimation) 기법 등이 있다. 이 중에서, MS, WSA, IMCRA 등은 이전의 프레임의 정보를 이용하여 잡음 추정하여 급변하는 잡음환경에서 취약점을 보인다. 이러한 문제점을 극복하기 위해 CDSTP가 제안되었는데, CDSTP는 이전의 프레임으로부터 이전 정보를 이용하지 않고 음성 및 잡음 코드북들에 기초하여 잡음 스펙트럼을 추정한다.
- [0031] CDSTP 알고리즘은 음성과 잡음의 스펙트럼을 LPC(Linear predictive coding) 계수로 표현하고, 대표적인 스펙트럼 형태(shape) 들을 LPC 형태로 코드북에 저장한다. 이후 최대 유사도 추정(maximum-likelihood estimates)

방식을 사용하여 음성과 잡음의 코드북 파라미터와 각각의 게인 값을 추정하여 잡음을 추정하는 알고리즘이다.

[0034] 그러나 CDSTP의 잡음 추정 성능은 코드북 형태에 상당한 영향을 받는다. 만일 입력된 신호가 사전에 학습된 코드북에 정의된 잡음 신호에 의해 오염된 신호라면 CDSTP는 MS, WSA 및 IMCRA보다 훨씬 뛰어난 성능을 보여준다. 그러나 현실 세계에서 잡음의 형태에 대한 경우의 수는 무한대라고 볼 수 있기에 모든 종류에 대한 잡음 코드북을 갖는 것은 불가능하다고 볼 수 있을 것이다. 또한 코드북에 대한 수가 증가하게 되면 메모리 요구량도 증가하게 되는 문제점도 발생된다.

[0035] 따라서 본 발명에서는 음성 부재 확률(SAP; speech absence probability)를 이용한 CDSTP 알고리즘을 제한한다.

[0036] CDSTP에서는 음성과 잡음 스펙트럼 형태(shape) 코드북에서 음성과 잡음 형태 코드벡터 후보를 선정하여 likelihood(유사도) 값을 계산하게 되고, likelihood 값을 최소화하는 코드북 엔트리와 게인(gain)을 추정하여 최종적으로 바이너리 필터(Wiener filter) 계수로 사용하게 된다. CDSTP에서는 Itakura-Saito distortion(IS-D)을 이용하여 음성과 잡음에 대한 적절한 코드벡터를 찾는다. 그리고 프레임마다 이와 같은 과정의 반복을 통해서 파라미터들을 측정하게 된다.

[0037] CDSTP에서 사용하는 파라미터는 LPC 계수와 게인(gain)이다. 잡음과 음성 형태(shape) 코드벡터에 따른 게인 추정은 non-stationary 배경잡음 추정에 중요한 역할을 한다.

[0038] 한편, 음성신호 $y(n)$ 는 효율적인 음성 코딩을 위해 사람의 vocal tract 필터를 선형 예측 코딩(Linear Predictive Coding; LPC)계수에 의해 모델링 된다(S14). 보다 나은 LPC계수의 코딩을 위해 선형 스펙트럼 주파수(Line Spectral Frequency; LSF)가 제안되었는데, 음성의 각 프레임은 올-폴(all-pole) 필터,

$H(z)=1/A(z)$ 로써 일반화 될 수 있다. 여기서 $A(z)$ 는 LPC로 구성된 인버스 필터이고, 그 값은 Z-변환(Z-transform)을 이용하여 나타낸다.

수학식 3

$$A(z)=1+a_1z^{-1}+a_2z^{-2}+...+a_pz^{-p}$$

[0039]

[0040] 여기서 P 는 LPC 차수이고, a 는 LPC 계수이다. LSF를 정의하기 위해 inverse 필터는 2개의 다항식으로 나타낼 수 있다.

수학식 4

$$P(z)=A(z)+z^{-(P+1)}A(z^{-1})$$

$$Q(z)=A(z)-z^{-(P+1)}A(z^{-1})$$

[0041]

[0042] $P(z)$ 는 인버스 필터의 짝수 항이고, $Q(z)$ 는 홀수 항을 나타낸다. 코딩을 위해 이용되는 LSF계수는 이 $P(z)$ 와

$Q(z)$ 의 근을 말한다. (2)번식으로 구한 LSF는 오름차순으로 정렬된다. 그 성질은 효율적인 LSF계산을 도와주기 때문에, 압축 효율을 높일 수 있게 된다. 그 결과 입력된 음성 신호를 하나의 프레임이 M개의 선형 스펙트럼 주파수로 이루어진 선형 스펙트럼 주파수를 얻을 수 있다.

[0043] 도 1을 참조하면, 음성과 잡음의 형태 코드 벡터를 통해 게인 값을 구하고, IS-D를 이용하여 최적의 파라미터를 찾는 과정을 나타낸다. 잡음으로 오염된 입력신호의 스펙트럼과 코드북에 저장되어 있는 정보들을 이용해 클린

음성과 잡음의 형태와 계인을 측정한다.

[0044] 본 발명의 일 실시예에서는 잡음 코드북 데이터베이스에 있는 잡음 코드북들을 이용하여 잡음 코드북을 선정하는 데(S13), 본 발명의 일 실시예에서는 IS-D 알고리즘을 이용하여 잡음 코드북을 선정한다.

[0045] IS-D를 이용한 방법에서 잡음신호의 코드북 타입의 선정은 다음의 식에 의해 이루어진다.

수학식 5

$$n^* = \arg \max_n \left\{ \max_{1 \leq j \leq S_n} p(\tilde{a}_w | a_w^j(n)) \right\}, 1 \leq n \leq N$$

[0046]

[0047] $\tilde{a}_w$ 는 MS(minimum statistics) 등의 잡음 추정 기법으로 예측한 잡음신호 샘플벡터이며, $a_w^j(n)$ 은 n 번째 잡음 코드북의 j 번째 샘플 벡터, S_n 은 n 번째 코드북의 크기, N은 잡음 코드북의 수를 나타낸다.

[0048] 수학식 5에 의해 선정된 잡음신호 코드북과 음성신호 코드북의 조합을 이용하여 IS-D(Itakura-Saito Distortion)를 최소화하는 파라미터를 찾는다. 그리고 잡음 제거를 위해 바이너 필터(Wiener filter)를 만들 때에 이 파라미터들이 사용되며 이 바이너 필터링(Wiener filtering) 과정을 거쳐 음성 향상을 이룰 수 있다.

[0049] 이와 같은 방법은 잡음 코드북의 i 번째 코드 벡터를 변경해 가면서 최대 유사도 값을 갖는 잡음 코드 벡터를 찾게 된다. 마찬가지로 IS-D 알고리즘을 통해 클린 음성의 코드북에서 클린 음성 코드 벡터를 찾는다(S15).

[0050] 클린 음성과 잡음 코드북에서 각각 i 번째 음성(클린 음성) 코드 벡터 $\{a_x^i\}$ 와 j 번째 잡음 벡터 $\{a_w^j\}$ 를 선택하는 과정을 나타내면 다음 수학식과 같다.

수학식 6

$$i^* j^* = \arg \max_{i,j} \max_{\sigma_x^2, \sigma_w^2} p_y(y | a_x^i, a_w^j; \sigma_x^2, \sigma_w^2)$$

[0051]

[0052] 여기서 σ_x^2 과 σ_w^2 는 음성과 잡음의 excitation variance를 나타낸다. 위 수학식의 확률밀도함수(PDF)

$p_y(y | a_x^i, a_w^j; \sigma_x^2, \sigma_w^2)$ 가 가우시안이라 가정하고 유사도(likelihood) 값을 로그 도메인에서 표현하면 다음 수학식으로 표현될 수 있다.

수학식 7

$$L = \int_0^{2\pi} -\frac{P_y(w)}{\frac{\sigma_x^2}{|A_x^i(w)|^2} + \frac{\sigma_w^2}{|A_w^j(w)|^2}} + \ln \left(\frac{1}{\frac{\sigma_x^2}{|A_x^i(w)|^2} + \frac{\sigma_w^2}{|A_w^j(w)|^2}} \right) dw$$

[0053]

[0054] 여기서 $P_y(w)$ 는 입력신호의 스펙트럼, $A_x^i(w)$ 는 i 번째 음성 코드벡터의 스펙트럼, $A_w^j(w)$ 는 j 번째 잡음 코드벡터의 스펙트럼이다.

[0055] 위 두 수학식을 결합하면 다음과 같은 수학식으로 나타낼 수 있다.

수학식 8

[0056]

$$i^* j^* = \arg \min_{i, j} \left\{ \min_{\sigma_x^2, \sigma_w^2} d_{\mathbb{S}}(P_y(w), \hat{P}_y(w)) \right\}$$

[0057]

여기서 $\hat{P}_y(w)$ 는 음성과 잡음 코드벡터를 통해 표현한 스펙트럼으로 $\frac{\sigma_x^2}{|A_x^i(w)|^2} + \frac{\sigma_w^2}{|A_w^j(w)|^2}$ 와 같이 표현할 수 있다.

[0058]

$d_{\mathbb{S}}(P_y(w), \hat{P}_y(w))$ 는 원신호와 합성신호 스펙트럼의 IS-D(Itakura-Saito Distortion)을 의미하며, 아래 수학식 9와 같이 표현할 수 있다.

수학식 9

[0059]

$$d_{\mathbb{S}}(P_y(w), \hat{P}_y(w)) = \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{P_y(w)}{\hat{P}_y(w)} - \ln \left(\frac{P_y(w)}{\hat{P}_y(w)} \right) - 1 \right) dw$$

[0060]

일반적인 CDSTP에서 계인은 위 수학식 8을 최소화하는 과정에서 구할 수 있을 것이다. 그리고 $\{i^*, j^*, \sigma_x^2, \sigma_w^2\}$ 와 같이, 최적의 LPC와 계인이 선택되었다면, 아래 수학식 10과 같은 Wiener filter(바이너 필터)를 구현하여 잡음 제거에 적용할 수 있을 것이다.

수학식 10

[0061]

$$H(w) = \frac{\frac{\sigma_x^2}{|A_x^{i^*}(w)|^2}}{\frac{\sigma_x^2}{|A_x^{i^*}(w)|^2} + \frac{\sigma_w^2}{|A_w^{j^*}(w)|^2}}$$

[0062]

그러나 본 발명의 음성 향상 방법은 음성 부존재 확률을 이용하여 잡음의 계인을 산정하는데, 음성 부존재 확률(SAP)을 이용하기 위해, 본 발명에서는 음성이 존재하지 않는 묵음 구간과 음성이 존재하는 구간을 나누어 음성 부존재 확률에 대한 아래와 같은 수학식의 바이너리 추정 모델을 구한다(S16).

수학식 11

[0063]

$$H_0: |Y(\omega)|^2 = |D(\omega)|^2$$

$$H_1: |Y(\omega)|^2 = |X(\omega)|^2 + |D(\omega)|^2$$

[0064]

여기서 H_0 및 H_1 은 각각 음성이 부존재 할 때와 음성이 존재할 때의 추정 모델이다. 즉, H_0 는 잡음(D)만이 존

제하는 경우이며, H_1 은 클린 음성(X)과 잡음(D)이 혼합된 경우이다. 입력된 음성의 목음 구간에서의 H_0 의 확률 즉 $P(H_0)$ 는 1이 되고 H_1 의 확률 즉 $P(H_1)$ 는 0이 되며, 음성이 존재하는 구간에서의 H_1 의 확률 즉 $P(H_1)$ 는 1이 되고 H_0 의 확률 즉 $P(H_0)$ 는 0이 된다. 그리고 $P(H_0) + P(H_1)$ 는 언제나 1의 값을 갖게 된다.

[0065] 주어진 잡음 관측 입력 신호 y 에 대해 추정 모델 H 에 대한 확률 밀도 함수(PDF)는 $P(H | y)$ 로 정의된다. 바람직한 실시예에서, $P(H | y)$ 는 MS, WSA, IMCRA 등과 같은 종래의 잡음 추정 알고리즘을 통해 계산된 이전 및 그 이전 SNR 값을 이용하여 추정한다. 즉, 본 SAP를 이용한 발명에서는, 2개의 추정 모델 H_0 및 H_1 에 대한 확률 밀도 함수(PDF)를 $P(H_0 | y)$ 와 $P(H_1 | y)$ 로 정의하며, $P(H_0 | y)$ 와 $P(H_1 | y)$ 는 종래의 잡음 추정 알고리즘을 통해 계산된 이전 및 상기 이전의 이전 SNR 값을 이용하여 추정하게 된다.

[0066] 그리고 CDSTP를 위한 SAP 기반 최소 평균 제곱 오류(Minimum Mean Square Error; MMSE) 추정은 다음 수학식으로 정의된다.

수학식 12

$$\begin{aligned}\hat{\theta} &= \sum_{H \in \{H_0, H_1\}} P(H | y) E[\theta | y, H] \\ &= \sum_{H \in \{H_0, H_1\}} p(H | y) \int \theta(H) \frac{p(\theta(H)) p(y | \theta(H))}{p(y | H)} d\theta(H)\end{aligned}$$

[0067]

[0068] 여기서 $\theta(H) = [\theta_x(H), \theta_d(H), \sigma_x^2(H), \sigma_d^2(H)]$ 이며, 주어진 모델 H 의 파라미터 세트(θ)이다. 그리고 $\theta_x(H)$ 와 $\theta_d(H)$ 는 각각 클린 음성 및 잡음의 스펙트럼 형태의 항을 의미하며, $\sigma_x^2(H)$ 와 $\sigma_d^2(H)$ 는 각각 클린 음성과 잡음의 계인의 항을 나타낸다. 위 수학식에서 알 수 있는 바와 같이, 본 발명에 따른 SAP 기반 CDSTP에서는 클린 음성이 존재하는 구간(H_1)과 클린 음성이 부존재하는 목음 구간(H_0)을 구분하여 각각 연산을 별도로 진행한다.

[0069] 이때, 클린 음성 및 잡음의 스펙트럼 형태(shape)는 각각 다음의 수학식과 같이 P차 및 Q차 LPC 파라미터들에 의해 나타내진다.

수학식 13

$$\theta_x(H) = \{a_{x,1}^H, \dots, a_{x,P}^H\}$$

[0070]

$$\theta_d(H) = \{a_{d,1}^H, \dots, a_{d,Q}^H\}$$

[0071]

[0072] 그리고 대응되는 스펙트럼은 각각 다음의 수학식으로 나타내질 수 있다.

수학식 14

$$A_x^H(\omega) = \sum_{k=0}^P a_{x,k}^H e^{-j\omega k}$$

[0073]

$$A_d^H(\omega) = \sum_{k=0}^Q a_{d,k}^H e^{-j\omega k}$$

[0074]

[0075] 그리고 $\theta_x(H)$, $\theta_d(H)$ 및 잡음 섞은 음성 y 가 주어진 경우, 클린 음성과 잡음의 게인(gain) 항들 ($\sigma_x^2(H)$, $\sigma_d^2(H)$)은 아래의 수학의 최대 유사도(ML; Maximum Likelihood) 산정을 통해 결정될 수 있다 (S17).

수학식 15

$$\begin{bmatrix} \sigma_x^{2,ML}(H_0) \\ \sigma_d^{2,ML}(H_0) \end{bmatrix} = \begin{bmatrix} 0 \\ \left\| \left| A_d^{H_0}(\omega) \right|^2 / |Y(\omega)|^2 \right\| \end{bmatrix}$$

$$\begin{bmatrix} \sigma_x^{2,ML}(H_1) \\ \sigma_d^{2,ML}(H_1) \end{bmatrix} = C^{-1}D$$

[0076] 여기서

$$C = \begin{bmatrix} \left\| 1/|Y(\omega)|^4 |A_x^{H_1}(\omega)|^4 \right\| & \left\| 1/|Y(\omega)|^4 |A_x^{H_1}(\omega)|^2 |A_d^{H_1}(\omega)|^2 \right\| \\ \left\| 1/|Y(\omega)|^4 |A_x^{H_1}(\omega)|^2 |A_d^{H_1}(\omega)|^2 \right\| & \left\| 1/|Y(\omega)|^4 |A_d^{H_1}(\omega)|^4 \right\| \end{bmatrix}$$

$$D = \begin{bmatrix} \left\| 1/|Y(\omega)|^2 |A_x^{H_1}(\omega)|^2 \right\| \\ \left\| 1/|Y(\omega)|^2 |A_d^{H_1}(\omega)|^2 \right\| \end{bmatrix}$$

[0077] 여기서 $\|\cdot\| = \int |\cdot| d\omega$ 이다.

[0078] 즉 위 수학식에서 최대 값을 갖는 $\sigma_x^{2,ML}$ 와 $\sigma_d^{2,ML}$ 가 각각 클린 음성 및 잡음의 게인 ($\sigma_x^2(H)$, $\sigma_d^2(H)$)이 된다.

[0079] 위 수학식에서 알 수 있는 바와 같이, 게인은 클린 음성이 부존재할 때의 게인($\sigma_x^{2,ML}(H_0)$, $\sigma_d^{2,ML}(H_0)$)과 클린 음성이 존재할 때의 게인($\sigma_x^{2,ML}(H_1)$, $\sigma_d^{2,ML}(H_1)$)을 구분하여 산출한다.

[0080] 한편, 수학식 1, 2에서 클린 음성과 잡음이 서로 독립적인 것이라고 가정하는데, 이 경우 클린 음성 및 잡음의 스펙트럼 형태와 게인의 항들도 서로 독립적인 것이라고 한다면,

수학식 16

$$p(\theta(H)) = p(\theta_x(H))p(\theta_d(H))p(\sigma_x^{2,ML}(H))p(\sigma_d^{2,ML}(H))$$

[0081] 와 같은 확률식이 도출될 수 있다.

[0082] 그러면, 수학식 12의 SAP 기반 최소 평균 제곱 오류(Minimum Mean Square Error; MMSE) 추정식은 다음 식과 같

이 변형될 수 있다.

수학식 17

$$\hat{\theta} \approx \sum_{H \in \{H_0, H_1\}} p(H|y) \int_{\Theta(H)} \theta(H) \frac{p(\theta_x(H))p(\theta_d(H))p(\sigma_x^2(H))p(\sigma_d^2(H))p(y|\theta(H))\delta(\sigma_x^2 - \sigma_x^{2,ML}(H))\delta(\sigma_d^2 - \sigma_d^{2,ML}(H))}{p(y|H)} d\theta(H)$$

[0086]

[0087] 여기서 $\delta(\cdot)$ 는 델타 함수이다.

[0088] 또한 계인 항들은 주어진 형태 항들에 의해 결정되기 때문에, 수학식 17은 다음과 같이 바뀔 수 있다.

수학식 18

$$\hat{\theta} \approx \sum_{H \in \{H_0, H_1\}} p(H|y) \int_{\Theta_d(H)} \int_{\Theta_x(H)} \theta'(H) \frac{p(\theta_x(H))p(\theta_d(H))p(\sigma_x^{2,ML}(H))p(\sigma_d^{2,ML}(H))p(y|\theta'(H))}{p(y|H)} d\theta_x(H) d\theta_d(H)$$

[0089]

[0090] 이 때,

$$p(y|H) = \int_{\Theta_x(H)} \int_{\Theta_d(H)} p(\theta_x(H))p(\theta_d(H))p(\sigma_x^{2,ML}(H))p(\sigma_d^{2,ML}(H))p(y|\theta'(H))d\theta_d(H)d\theta_x(H).$$

[0091]

[0092] 여기서 $\theta'(H) = [\theta_x(H), \theta_d(H), \sigma_x^{2,ML}(H), \sigma_d^{2,ML}(H)]$ 이다.

[0093] 수학식 18에서의 적분은 아래의 수학식에 나타낸 수치 적분을 통해 계산될 수 있다.

수학식 19

$$\hat{\theta} \approx \sum_{H \in \{H_0, H_1\}} p(H|y) \left\{ \frac{1}{N_x N_d} \sum_{i,j=1}^{N_x, N_d} \theta'_{ij}(H) \frac{p(\theta_{x,i}(H))p(\theta_{d,j}(H))p(\sigma_{x,ij}^{2,ML}(H))p(\sigma_{d,ij}^{2,ML}(H))p(y|\theta'_{ij}(H))}{p(y|H)} \right\}$$

[0094]

[0095] 이 때,

$$p(y|H) = \frac{1}{N_x N_d} \sum_{i,j=1}^{N_x, N_d} p(\theta_{x,i}(H))p(\theta_{d,j}(H))p(\sigma_{x,ij}^{2,ML}(H))p(\sigma_{d,ij}^{2,ML}(H))p(y|\theta'_{ij}(H))$$

[0096]

[0097] 여기서,

$$p(y|\theta'_{ij}(H)) = C \exp \left(-d_{IS} \left(\left| Y(\omega) \right|^2, \frac{\sigma_{x,ij}^{2,ML}(H)}{\left| A_{x,i}^H(\omega) \right|^2} + \frac{\sigma_{d,ij}^{2,ML}(H)}{\left| A_{d,j}^H(\omega) \right|^2} \right) \right)$$

[0098] 이다.

[0099] 여기서 C 와 $d_{IS}(\cdot)$ 는 각각 임의 상수와 Itakura-Saito distortion 측정 값이다. 클린 음성과 잡음의 스펙트럼 형태를 표현하기 위해 2개의 독립적인 LPC 코드북들이 학습된다(S13, S15). N_x 및 N_d 는 클린 음성과 잡음의 코드벡터이다.

[0100] i 번째 클린 음성 및 j 번째 잡음의 코드벡터들, $\theta_{x,i}(H)$ 와 $\theta_{d,j}(H)$ 의 주어진 조합들에 대해서, 대응하는 계인 항들인, $\sigma_{x,ij}^{2,ML}(H)$ 및 $\sigma_{d,ij}^{2,ML}(H)$ 는 수학식 15를 통해 얻어진다. 그 결과 i 번째 클린 음성 및 j 번째 잡음에 대한 파라미터 세트인 $\theta'_{ij}(H) = [\theta_{x,i}(H), \theta_{d,j}(H), \sigma_{x,ij}^{2,ML}(H), \sigma_{d,ij}^{2,ML}(H)]$ 를 얻을 수 있다.

[0101] 이를 간략히 하기 위해, i 번째 클린 음성 및 j 번째 잡음에 대한 파라미터 세트 $\theta'_{ij}(H)$ 의 각 성분들의 확률 밀도 함수(PDF)는 고르게 분포되어 있다고 가정한다. 이 경우, 수학식 19에서 분자와 분모 사이의 항들이 상쇄되기 때문에 각 확률 밀도 함수의 간격은 특정될 필요가 없다.

[0102] 수학식 19에서 최적의 최소 평균 제곱 오류(MMSE) 추정 값인 $\hat{\theta} = [\hat{\theta}_x, \hat{\theta}_d, \hat{\sigma}_x^2, \hat{\sigma}_d^2]$ 로 계산되면, 음성 향상을 위해 다음과 바이너리 필터(wiener filter) 과정을 거친다(S18).

수학식 20

$$\frac{\hat{\sigma}_x^2}{|\hat{A}_x(\omega)|^2} / \left(\frac{\hat{\sigma}_x^2}{|\hat{A}_x(\omega)|^2} + \frac{\hat{\sigma}_d^2}{|\hat{A}_d(\omega)|^2} \right)$$

[0103]

[0104] 여기서 $\hat{A}_x(\omega)$ 와 $\hat{A}_d(\omega)$ 는 각각 추정된 파라미터 세트 클린 음성과 잡음의 형태인 $\hat{\theta}_x$ 및 $\hat{\theta}_d$ 에 의해 구현된 LPC 스펙트럼이다.

[0105] 위와 같은 바이너리 필터링을 통해 잡음은 제거되고 클린 음성만 남게 되어 음성이 향상되게 된다. 그런 다음 필터링된 신호를 주파수 영역에서 시간 영역으로 변환시키는 단계를 거치게 된다(S19).

[0106] 아울러, 본 발명의 상세한 설명에서는 입력된 음성에서 잡음을 제거하여 음성을 향상시키는 방법을 설명하였지만, 본 발명은 잡음 제거 방법뿐만 아니라 본 방법을 적용한 잡음 제거 장치도 포함한다.

[0107] 도 2는 본 발명의 실시예에 따른 음성 인식 향상 장치의 구성을 나타낸다.

[0108] 도 2를 참조하면 음성 인식 향상 장치(20)는 FFT부(21), 잡음 추정부(22), 잡음 코드벡터 선정부(23), 잡음 코드 데이터베이스(24), 클린 음성 코드벡터 선정부(25), 클린 음성 코드 데이터베이스(26), 계인 추정 및 최적 파라미터 선정부(27), LPC부(28), SAP 모델링부(29), wiener filtering 부(30) 및 IFFT부(31)를 포함한다. FFT부(21)는 입력되는 신호의 신호 처리를 위해 신호 영역의 음성 신호를 주파수 영역의 신호로 바꾸는 고속 푸리에 변환(Fast Fourier Transform; FFT)을 수행을 하며, 잡음 추정부(22)는 입력된 음성 신호(x+d)에서 잡음(d)을 추정하는 기능을 하고, 잡음 코드 벡터 선정부(23)와 클린 음성 코드 벡터 선정부(25)는 음성 부존재 확률에 대한 바이너리 추정 모델을 기반으로 상기 음성 신호의 클린 음성 및 잡음 스펙트럼 형태를 구하는 기능을 하며, 잡음 코드 데이터베이스(24)와 클린 음성 코드 데이터베이스(26)는 각각 잡음 코드와 클린 음성 코드를 저장한다. 계인 추정 및 최적 파라미터 선정부(27)는 음성 신호에 대하여 상기 바이너리 추정 모델을 기반으로 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하여, 상기 클린 음성 부존재 시의 계인과 상기 클린 음성 존재시의 계인을 최대 유사도 산정을 통해 별도로 결정하며, 입력 음성 신호와 유사한 잡음과 클

린 음성의 코드북을 산정하여 클린 음성 및 잡음의 형태 및 계인을 포함하는 파라미터 세트를 산출하는 기능을 한다. LPC부(28)는 음성과 잡음의 스펙트럼을 LPC(Linear predictive coding) 계수로 표현하고, 대표적인 스펙트럼 형태(shape) 들을 LPC 형태로 만드는 기능을 수행하며, SAP 모델링부(29)는 입력된 음성 신호에 대해서 클린 음성이 존재하는 구간과 클린 음성이 부존재하는 구간을 구분하는 음성 부존재 확률(SAP)에 대한 바이너리 추정 모델을 산정하는 기능을 한다. 또한 Wiener filtering 부(30)는 클린 음성 및 잡음의 스펙트럼 형태 및 계인들을 이용하여 입력된 음성 신호와의 최대 유사도 추정을 통해 잡음을 Wiener filtering 알고리즘을 통해 제거하는 기능을 하며, IFFT부(31)는 주파수 영역의 신호를 시간 영역의 신호로 바꾸기 위해 역 고속 푸리에 변환을 수행한다.

[0109] 각 구성요소에 대한 구체적인 기능은 도 1의 음성 향상 알고리즘에서 설명한 바와 같아 반복되는 설명은 생략한다.

[0110] 본 발명에 따른 음성 향상 시스템의 성능을 알아보기 위해 기존의 IMCRA와 CDSTP 와 본 발명에 따른 SAP-CDSTP 를 이용한 잡음 추정 알고리즘을 다양한 잡음 환경에서의 LSD(log-spectral distortion) 과 SSNR(Signal Separation and Noise Reduction) 값을 비교한 결과는 다음과 같다.

표 1

		dB	LSD score			SSNR score		
			IMCRA	CDSTP	SAP-CDSTP	IMCRA	CDSTP	SAP-CDSTP
Inside noise	Pink	-5	8.37	7.82	6.36	-5.09	-3.52	-1.80
		0	6.20	5.72	4.90	-1.09	0.28	1.34
		5	4.43	4.09	3.73	2.90	4.03	4.61
	White	-5	10.83	9.11	7.29	-4.71	-3.78	-1.85
		0	8.35	6.86	5.70	-0.67	-0.02	1.23
		5	6.22	5.10	4.50	3.25	3.64	4.40
	Babble	-5	7.82	7.23	6.72	-7.29	-5.86	-4.97
		0	5.80	5.36	5.09	-2.89	-1.92	-1.39
		5	4.12	3.90	3.77	1.42	1.88	2.17
	F16	-5	7.79	7.53	6.30	-5.54	-4.93	-3.24
		0	5.64	5.18	4.58	-1.11	0.00	0.87
		5	4.02	3.69	3.39	2.96	3.78	4.27
	Factory2	-5	5.56	5.32	4.82	-3.93	-1.76	-0.77
		0	4.17	3.95	3.71	0.04	1.92	2.46
		5	2.94	2.77	2.67	4.11	5.62	5.91
	avg.		6.15	5.58	4.90	-1.18	-0.04	0.88
Outside noise	Factory1	-5	7.96	8.79	7.98	-5.04	-6.60	-5.53
		0	6.23	6.61	6.05	-1.31	-2.56	-1.71
		5	4.50	4.72	4.43	2.71	1.81	2.34
	Buccaneer1	-5	9.60	9.95	8.44	-5.92	-6.88	-4.94
		0	7.25	7.61	6.56	-1.75	-2.59	-1.20
		5	5.29	5.59	4.96	2.30	1.62	2.49
	Car	-5	6.84	8.58	7.58	-4.22	-6.16	-4.86
		0	4.56	5.69	5.21	0.05	-1.35	-0.72
		5	3.22	4.09	3.87	3.93	2.43	2.74
	Street	-5	7.93	8.39	7.68	-3.09	-4.50	-3.58
		0	5.48	5.37	5.06	3.21	2.74	3.22
		5	4.07	3.84	3.70	7.34	6.88	7.14
	Restaurant	-5	8.72	9.73	9.33	-8.50	-10.07	-9.63
		0	6.17	6.83	6.57	-3.11	-4.38	-4.07
		5	4.44	4.85	4.71	1.37	0.34	0.54
	avg.		6.15	6.71	6.14	-0.80	-1.95	-1.18

[0112] 위의 표 1을 살펴보면, 내부 잡음 환경에는 NOISEX-92 데이터베이스 중 핑크 잡음, 화이트 잡음, babble(와글와글) 잡음, F16 잡음, 공장2 잡음에서 테스트하였고, 외부 잡음 환경에는 NOISEX-92 데이터베이스와 AURORA2 데

이터베이스 중 공장 1, buccaneer 1 잡음, 자동차 잡음, 거리 잡음, 식당 잡음에서 테스트 하였다.

[0113] 테스트 결과, 본 발명에 따른 SAP를 이용한 CDSTP 잡음 추정 알고리즘을 이용한 음성 향상 방법은 기존의 IMCRA 나 CDSTP 잡음 추정 방법에 비해 훨씬 낮은 LSD값을 가졌고, SSNR 성능도 본 발명에 따른 방법이 기존의 IMCRA 나 CDSTP 방법에 비해 더 뛰어난 성능을 가지고 있음이 확인되었다.

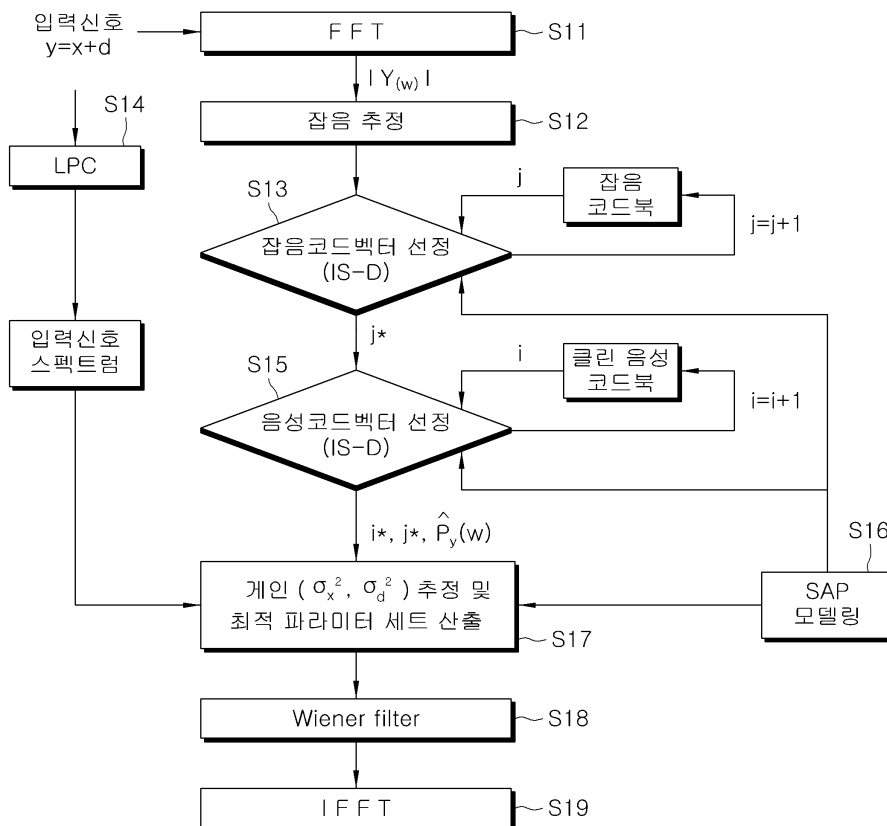
[0114] 한편, 외부 잡음 환경에서 기존의 CDSTP 의 경우는 제한된 잡음 코드벡터들의 한계로 인해, IMCRA 보다 못한 성능을 갖지만, 본 발명에 따른 SAP를 결합한 CDSTP 잡음 추정 방법을 통한 음성 향상 방법은 IMCRA보다 동등하거나 더 좋은 성능을 가짐을 확인할 수 있다. 따라서 본 발명에 따른 음성 향상 방법은 급변하는 잡음 환경뿐만 아니라 학습되지 않는 잡음에 대해서도 강인한 효과를 갖는다.

[0115] 본 발명은 첨부된 도면에 도시된 실시예를 참고하여 설명되었으나, 이는 예시적인 것에 불과하며, 본 발명과 관련된 기술 분야에서의 통상적인 지식을 가진 자라면, 이로부터 다양한 변형 또는 균등한 타 실시예가 존재 가능하다는 점을 이해할 것이다.

[0116] 따라서 본 발명의 진정한 기술적 보호 범위는 첨부된 특허청구범위의 기술적 사상에 의해 정해져야 할 것이다.

도면

도면1



도면2

