



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0073150
(43) 공개일자 2010년07월01일

(51) Int. Cl.

G10L 21/02 (2006.01) G10L 15/20 (2006.01)
G10L 19/02 (2006.01)

(21) 출원번호 10-2008-0131743

(22) 출원일자 2008년12월22일

심사청구일자 2008년12월22일

(71) 출원인

인하대학교 산학협력단

인천 남구 용현동 253 인하대학교

(72) 발명자

장준혁

서울 서초구 서초동 1660-31 B동 301호

최재훈

서울특별시 성동구 금호동3가 두산아파트 107동 406호

(74) 대리인

특허법인태동

전체 청구항 수 : 총 2 항

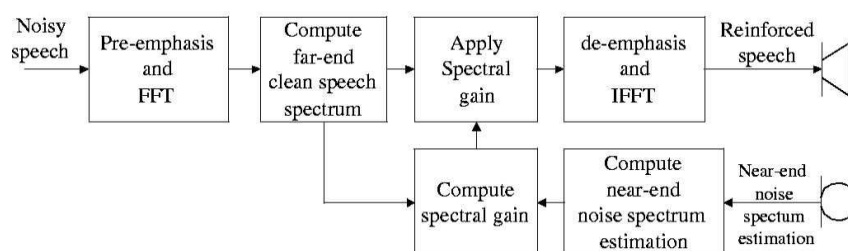
(54) 원단 잡음 환경에서 소프트 디시전을 이용한 음성 강화 기법

(57) 요약

본 발명은 다양한 배경 잡음이 존재하는 근단 환경에서 추정된 근단 화자의 배경 잡음 전력에 기초하여 오염된 원단 화자의 음성 전력을 강화하는 방법에 관한 것으로서, 보다 구체적으로는 (1) 음성 및 잡음 신호의 분포가 복소 가우시안 확률 밀도 함수(PDF)를 따른다고 가정하고, 지역 음성 부재 확률(Local Speech Absence Probability; LSAP)을 구하는 제1 단계; (2) 상기 제1 단계에서 구한 상기 LSAP를 이용하여 잡음이 섞이지 않은 깨끗한 음성 스펙트럼을 추정하는 제2 단계; (3) 근단에 존재하는 배경 잡음을 추정하여 배경 잡음 전력을 추정하는 제3 단계; (4) 상기 제2 단계에서 추정된 상기 깨끗한 음성 스펙트럼 및 상기 제3 단계에서 추정된 상기 배경 잡음 전력을 이용하여 음성 증폭 이득을 수정하는 제4 단계; 및 (5) 상기 제2 단계에서 추정된 상기 깨끗한 음성 스펙트럼, 상기 제3 단계에서 추정된 상기 배경 잡음 전력, 및 상기 제4 단계에서 수정된 상기 음성 증폭 이득을 이용하여 잡음이 섞이지 않은 깨끗한 음성 신호만을 강화시키는 제5 단계를 포함하는 것을 그 구성상의 특징으로 한다.

본 발명의 음성 강화 방법에 따르면, 소프트 디시전에 기초한 음성 존재와 음성 부재의 확률을 적용하여 오염된 원단 음성 신호에서 잡음 신호 전력과 실제 음성 신호 전력을 정확하게 추정하는 것에 의해 깨끗한 음성 신호 전력을 구한 후, 근단의 배경 잡음 전력과의 연산을 통해 얻어진 이득만큼 구해진 깨끗한 음성 신호만을 증폭시킴으로써, 근단 화자가 들게 되는 오염된 음성 신호의 품질을 크게 향상시킬 수 있다.

대 표 도 - 도2



이 발명을 지원한 국가연구개발사업

과제고유번호 2008-F-045-01

부처명 산업자원부

연구사업명 IT핵심기술개발사업

연구과제명 장애인 및 고령자를 위한 Digital Guardian 기술개발

주관기관 한국과학기술연구원

연구기간 2008.03~2009.02

특허청구의 범위

청구항 1

원단 잡음 환경에서 소프트 디지전을 이용한 음성 강화 방법으로서,

- (1) 음성 및 잡음 신호의 분포가 복소 가우시안 확률 밀도 함수(PDF)를 따른다고 가정하고, 지역 음성 부재 확률(Local Speech Absence Probability; LSAP)을 구하는 제1 단계;
- (2) 상기 제1 단계에서 구한 상기 LSAP를 이용하여 잡음이 섞이지 않은 깨끗한 음성 스펙트럼을 추정하는 제2 단계;
- (3) 근단에 존재하는 배경 잡음을 추정하여 배경 잡음 전력을 추정하는 제3 단계;
- (4) 상기 제2 단계에서 추정된 상기 깨끗한 음성 스펙트럼 및 상기 제3 단계에서 추정된 상기 배경 잡음 전력을 이용하여 음성 증폭 이득을 수정하는 제4 단계; 및
- (5) 상기 제2 단계에서 추정된 상기 깨끗한 음성 스펙트럼, 상기 제3 단계에서 추정된 상기 배경 잡음 전력, 및 상기 제4 단계에서 수정된 상기 음성 증폭 이득을 이용하여 잡음이 섞이지 않은 깨끗한 음성 신호만을 강화시키는 제5 단계

를 포함하는 음성 강화 방법.

청구항 2

제1항에 있어서,

상기 제1 단계에서, 상기 LSAP($p(H_0|Y(t,k))$)를 다음과 같은 수학적식을 이용하여 구하는 것을 특징으로 하는 음성 강화 방법.

$$\begin{aligned} p(H_0|Y(t,k)) &= \frac{p(Y(t,k)|H_0)p(H_0)}{p(Y(t,k))} \\ &= \frac{p(Y(t,k)|H_0)p(H_0)}{p(Y(t,k)|H_0)p(H_0) + p(Y(t,k)|H_1)p(H_1)} \\ &= \frac{1}{1 + \frac{p(H_1)}{p(H_0)}\Lambda(Y(t,k))} \end{aligned}$$

여기서, $p(H_0)$ 는 음성 부재에 대한 사전 확률(*a priori probability*)이고, $\Lambda(Y(t,k))$ 는 k 번째 주파수 성분의 우도비로서, 아래 수학적식과 같이 주어진다.

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

여기서, $\gamma(t,k) \equiv \frac{\Phi_{YY}(t,k)}{\Phi_{DD}(t,k)}$ 와 $\xi(t,k) \equiv \frac{\Phi_{SS}(t,k)}{\Phi_{DD}(t,k)}$ 로서 정의됨.

명세서

발명의 상세한 설명

기술 분야

[0001] 본 발명은 원단 잡음 환경에서 소프트 디지전을 이용한 음성 강화 방법에 관한 것으로서, 특히 소프트 디지전에 기초한 음성 존재와 음성 부재의 확률을 적용하여 오염된 원단 음성 신호에서 잡음 신호 전력과 실제 음성 신호 전력을 정확하게 추정하는 것에 의해 깨끗한 음성 신호 전력을 구한 후, 근단의 배경 잡음 전력과 연산을 통해 얻어진 이득만큼 구해진 깨끗한 음성 신호만을 증폭시킴으로써, 근단 화자가 듣게 되는 오염된 음성 신호의 품질을 크게 향상시킬 수 있는 음성 강화 방법에 관한 것이다.

배경 기술

[0002] 다양한 배경 잡음이 존재하는 환경에서 휴대폰을 이용해 상대방과 음성 통화를 하게 되면 배경 잡음으로 인해 상대방 음성이 잘 들리지 않게 된다. 예를 들어, 지하철과 같은 소음이 큰 곳에서 휴대폰으로 통화를 하거나, 자동차가 다니는 길거리에서 통화 시에 소음이 없는 조용한 곳에서 통화하는 것보다 상대방의 음성이 매우 작게 들리게 된다. 이와 같은 경우, 사람들은 의식적으로 상대방의 음성을 잘 들을 수 있도록 휴대폰의 볼륨을 최대한으로 키우는 노력을 하게 되는데, 이러한 현상은 심리 음향학 용어로 마스킹 효과(Masking effect)라고 한다. 마스킹 효과(Masking effect)란 특정 큰 신호(Masker)가 존재할 때 함께 존재하는 다른 작은 신호(Maskee)는 전혀 들리지 않게 되는 것을 말하며, 마스킹 효과 중에서도 부분 마스킹 효과(Partial masking effect)에 의하면 비슷한 크기의 두 신호가 존재할 때 두 신호 모두 원래 신호 크기보다 작게 들리게 된다. 이 경우 서로 부분적으로 마스킹 하게 되는데, 이에 따라 잡음이 존재할 경우 함께 존재하는 음성의 크기는 작게 들리게 된다.

[0003] 기존의 음성 향상 방법은 원단에서 음성을 전송하기 전에 노이즈 부분의 제거에 초점을 맞추어 왔다. 그러나 배경 잡음이 존재하는 근단(Near-end) 환경에서는 배경 잡음 자체가 근단 화자의 귀에 직접적으로 전달되기 때문에 노이즈가 제거된 원단 음성 신호가 근단에 도달하여도 근단의 배경 잡음으로 인해 원단 음성 신호의 명료성은 저하되게 된다. 이와 같이, 근단의 배경 잡음이 존재하는 상황에서 근단 화자가 듣게 되는 원단 음성 신호의 명료도와 품질을 향상시키기 위해 원단 음성 신호를 강화하는 방법을 제안해 볼 필요성이 있다.

발명의 내용

해결 하고자하는 과제

[0004] 본 발명은, 상기와 같은 필요성의 인식에서 비롯된 것으로서, 소프트 디지전에 기초한 음성 존재와 음성 부재의 확률을 적용하여 오염된 원단 음성 신호에서 잡음 신호 전력과 실제 음성 신호 전력을 정확하게 추정하는 것에 의해 깨끗한 음성 신호 전력을 구한 후, 근단의 배경 잡음 전력과 연산을 통해 얻어진 이득만큼 구해진 깨끗한 음성 신호만을 증폭시킴으로써, 근단 화자가 듣게 되는 오염된 음성 신호의 품질을 크게 향상시킬 수 있는 음성 강화 방법을 제안하는 것을 그 목적으로 한다.

과제 해결수단

[0005] 상기한 목적을 달성하기 위한 본 발명의 특징에 따른 음성 강화 방법은, 원단 잡음 환경에서 소프트 디지전을 이용한 음성 강화 방법으로서,

[0006] (1) 음성 및 잡음 신호의 분포가 복소 가우시안 확률 밀도 함수(PDF)를 따른다고 가정하고, 지역 음성 부재 확률(Local Speech Absence Probability; LSAP)을 구하는 제1 단계;

[0007] (2) 상기 제1 단계에서 구한 상기 LSAP를 이용하여 잡음이 섞이지 않은 깨끗한 음성 스펙트럼을 추정하는 제2 단계;

[0008] (3) 근단에 존재하는 배경 잡음을 추정하여 배경 잡음 전력을 추정하는 제3 단계;

[0009] (4) 상기 제2 단계에서 추정된 상기 깨끗한 음성 스펙트럼 및 상기 제3 단계에서 추정된 상기 배경 잡음 전력을 이용하여 음성 증폭 이득을 수정하는 제4 단계; 및

[0010] (5)상기 제2 단계에서 추정된 상기 깨끗한 음성 스펙트럼, 상기 제3 단계에서 추정된 상기 배경 잡음 전력, 및 상기 제4 단계에서 수정된 상기 음성 증폭 이득을 이용하여 잡음이 섞이지 않은 깨끗한 음성 신호만을 강화시키는 제5 단계를 포함하는 것을 그 구성상의 특징으로 한다.

[0011] 바람직하게는, 상기 제1 단계에서, 상기 LSAP($p(H_0|Y(t,k))$)를 다음과 같은 수학적식을 이용하여 구하는 것을 특징으로 하는 음성 강화 방법.

$$\begin{aligned} p(H_0|Y(t,k)) &= \frac{p(Y(t,k)|H_0)p(H_0)}{p(Y(t,k))} \\ &= \frac{p(Y(t,k)|H_0)p(H_0)}{p(Y(t,k)|H_0)p(H_0) + p(Y(t,k)|H_1)p(H_1)} \\ &= \frac{1}{1 + \frac{p(H_1)}{p(H_0)}\Lambda(Y(t,k))} \end{aligned}$$

[0012]

[0013] 여기서, $p(H_0)$ 는 음성 부재에 대한 사전 확률(*a priori probability*)이고, $\Lambda(Y(t,k))$ 는 k번째 주파수 성분의 우도비로서, 아래 수학적식과 같이 주어진다.

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

[0014]

[0015] 여기서, $\gamma(t,k) \equiv \frac{\Phi_{YY}(t,k)}{\Phi_{DD}(t,k)}$ 와 $\xi(t,k) \equiv \frac{\Phi_{SS}(t,k)}{\Phi_{DD}(t,k)}$ 로서 정의됨.

효 과

[0016] 본 발명의 음성 강화 방법에 따르면, 소프트 디지전에 기초한 음성 존재와 음성 부재의 확률을 적용하여 오염된 원단 음성 신호에서 잡음 신호 전력과 실제 음성 신호 전력을 정확하게 추정하는 것에 의해 깨끗한 음성 신호 전력을 구한 후, 근단의 배경 잡음 전력과의 연산을 통해 얻어진 이득만큼 구해진 깨끗한 음성 신호만을 증폭시킴으로써, 근단 화자가 듣게 되는 오염된 음성 신호의 품질을 크게 향상시킬 수 있다.

발명의 실시를 위한 구체적인 내용

[0017] 이하에서는 첨부된 도면들을 참조하여, 본 발명에 따른 실시예에 대하여 상세하게 설명하기로 한다.

[0018] 본 발명의 상세한 설명에서는, 먼저 원단의 오염된 음성 신호 전력으로부터 잡음이 섞이지 않은 깨끗한 음성 신호 전력만을 추정할 수 있다는 것을 확인한 후, 근단에서 추정된 배경 잡음 전력과의 연산을 통해 이득을 구하여 오염된 원단 음성 신호 중에서 깨끗한 음성 신호 전력만을 증폭하는 본 발명에 따른 음성 강화 방법을 상세히 설명하기로 한다.

[0019] 1. 본 발명에서 제안하고 있는 음성 강화 방법의 개요

[0020] 도 1은 오염된 음성 전력 스펙트럼에서 실제 잡음이 섞이지 않은 깨끗한 음성 신호 전력만을 추정하여 근단에 존재하는 배경 잡음 전력과 일정한 SNR을 유지하도록 이득을 수정하여, 원단의 실제 음성 신호 전력을 증폭시키는 방식의 음성 강화 방법의 개념을 나타내는 도면이다. 도 2는, 도 1의 개념을 구현한 구체적인 음성 강화 방

법의 블록도로서, 각 단계별로 원단 음성 신호의 주파수 변환 단계, 잡음이 섞이지 않은 깨끗한 음성 추정 단계, 근단의 배경 잡음 전력과의 SNR을 이용한 이득 수정 단계, 원단의 음성 신호로부터 추정된 잡음이 섞이지 않은 깨끗한 음성 신호를 증폭시키는 단계를 포함하고 있다. 이하에서 도 2의 구성을 보다 상세하게 설명하기로 한다.

[0021] 도 2의 구성은, 잡음이 섞인 오염된 음성 신호에서 잡음 신호와 실제 음성 신호를 구분하기 위해 소프트 디시전에 기초한 음성 존재와 음성 부재의 확률을 적용하여 오염된 원단 음성 신호에서 잡음 신호 전력과 실제 음성 신호 전력을 정확하게 추정된 후, 근단의 배경 잡음 전력과의 연산을 통해 얻어진 이득만큼 추정된 깨끗한 음성 신호만을 증폭하여 근단 화자가 듣게 되는 오염된 음성 신호의 명료성을 강화시킨다는 것이다. 따라서 원단에서 입력되는 오염된 음성 신호로부터 깨끗한 음성 신호 전력을 정확하게 추정하는 것이 무엇보다 중요하다.

[0022] 원단에서의 깨끗한 음성 신호, 잡음 신호, 오염된 음성 신호를 각각 $S(t)$ 와 $d(t)$, $y(t)$ 라고 하면, 각각의 성분을 이산 푸리에 변환(Discrete Fourier Transform; DFT)을 통해 다음 수학적 식 1과 같이 주파수 축으로 변환하여 나타낼 수 있다.

수학적 식 1

[0023] $Y(t, k) = S(t, k) + D(t, k)$

[0024] 여기서, $Y(t, k)$, $S(t, k)$, $D(t, k)$ 는 각각 $y(t)$, $s(t)$, $d(t)$ 의 t 번째 프레임에 대한 k 번째 주파수 성분을 표시한다.

[0025] 추정된 근단의 배경 잡음 전력과 오염된 원단 음성 신호에서 깨끗한 음성 신호 전력의 추정치로 구해진 이득 $G(t, k)$ 를 곱해서 강화된 음성 신호 $\hat{y}(t)$ 을 다음과 수학적 식 2와 같이 나타낼 수 있다.

수학적 식 2

[0026] $\hat{Y}(t, k) = G(t, k) \cdot Y(t, k)$

[0027] 추정된 근단 화자의 배경 잡음 전력과 오염된 원단 음성 신호에서 잡음 신호를 제외한, 깨끗한 음성 신호 전력의 추정을 통하여 이득 $G(t, k)$ 를 다음 수학적 식 3과 같이 나타낼 수 있다.

수학적 식 3

[0028] $G(t, k) = \min \left\{ \max \left\{ \sqrt{\xi \cdot \frac{\Phi_{NN}(t, k)}{\hat{\Phi}_{SS}(t, k)}}, 1 \right\}, G_{\max} \right\}$

[0029] 수학적 식 3에서 나타낸 $\hat{\Phi}_{SS}(t, k)$ 는 오염된 원단 음성 신호에서 잡음이 섞이지 않은 깨끗한 음성 신호 전력 $\Phi_{SS}(t, k)$ 의 추정 값을 나타낸다. 본 발명에서 제안하고 있는 음성 강화 방법에서는 오염된 음성 신호로부터 잡음 신호와 실제 음성 신호를 정확하게 추정하고, 잡음 신호를 제외한 추정된 깨끗한 음성 신호 전력만을 증폭시키게 된다. 따라서 상기 수학적 식 3의 $\hat{\Phi}_{SS}(t, k)$ 는 오염된 원단 음성 신호 중에서 실제 잡음이 섞이지 않은 깨끗한 음성 신호 전력만을 추정한 값을 나타낸다. 오염된 원단 음성 신호의 깨끗한 음성 신호 전력의 추정 값 $\hat{\Phi}_{SS}(t, k)$ 는 소프트 디시전에 기초하여 아래 설명하는 바와 같이 구할 수 있다.

[0030] 2. 소프트 디시전을 이용한 깨끗한 음성 신호의 추정

[0031] 원단 신호의 음성 부재와 음성 존재에 대한 가설을 각각 H_0 , H_1 이라고 하면, 주파수 채널에 따라 다음 수학적 4와 같이 가정할 수 있다.

수학적 4

[0032] H_0 (음성 부재): $Y(t) = D(t)$

[0033] H_1 (음성 존재): $Y(t) = S(t) + D(t)$

[0034] 음성 신호와 잡음 신호의 스펙트럼이 zero-mean 복소 가우시안 분포를 따른다고 가정하면, 수학적 4에서 제시한 가설 H_0 , H_1 에 따라 다음 수학적 5와 같은 확률 밀도 함수로 표현할 수 있다.

수학적 5

$$p(Y(t,k)|H_0) = \frac{1}{\pi\lambda_d(t,k)} \exp\left\{-\frac{\Phi_{YY}(t,k)}{\lambda_d(t,k)}\right\}$$

$$p(Y(t,k)|H_1) = \frac{1}{\pi[\lambda_d(t,k) + \lambda_s(t,k)]} \cdot \exp\left\{-\frac{\Phi_{YY}(t,k)}{\lambda_d(t,k) + \lambda_s(t,k)}\right\}$$

[0035]

[0036] 여기서, $\lambda_s(t,k)$ 와 $\lambda_d(t,k)$ 는 각각 t번째 프레임에 대한 k번째 주파수 성분에서의 음성과 잡음의 분산을 나타낸다.

[0037] 따라서 음성의 존재와 부재에 관한 가설로부터 주파수 채널별 음성부재확률인 LSAP는 다음 수학적 6과 계산할 수 있다.

수학적 6

$$p(H_0|Y(t,k)) = \frac{p(Y(t,k)|H_0)p(H_0)}{p(Y(t,k))}$$

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

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

[0038]

[0039] 여기서, $p(H_0)$ 는 음성 부재에 대한 사전 확률(a priori probability)이고, $\Lambda(Y(t,k))$ 는 k번째 주파수 대역의 우도비(likelihood ratio)로써 다음 수학적 7과 같이 표현될 수 있다.

수학식 7

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

여기서, $\gamma(t,k)$, $\xi(t,k)$ 는 각각 *a posteriori* SNR과 predicted SNR로서 다음 수학식 8과 같이 정의될 수 있다.

수학식 8

$$\gamma(t,k) \equiv \frac{\Phi_{YY}(t,k)}{\Phi_{DD}(t,k)}$$

$$\xi(t,k) \equiv \frac{\Phi_{SS}(t,k)}{\Phi_{DD}(t,k)}$$

본 발명에서 제안하고 있는 음성 강화 방법의 경우 오염된 원단 음성 신호 중에서 잡음이 섞이지 않은 깨끗한 음성 신호만이 증폭의 대상이기 때문에, 잡음 전력 $\Phi_{DD}(t,k)$ 과 음성 전력 $\Phi_{SS}(t,k)$ 의 추정이 알고리즘 성능에 중요한 역할을 하게 된다. 일반적으로, 음성 검출기(Voice Activity Detector; VAD)를 사용하여 음성 부재 구간에서 잡음 전력을 갱신한다. 그러나 실제 잡음 환경이 비정상(non-stationary)인 경우, 잡음 전력은 음성의 존재 구간에서도 갱신되어야 한다. 따라서 신뢰성 있는 음성 전력 $\Phi_{SS}(t,k)$ 와 잡음 전력 $\Phi_{DD}(t,k)$ 을 추정하기 위하여, 음성 전력과 잡음 전력 각각에 long-term smoothed 전력 스펙트럼을 사용하여 다음 수학식 9와 같이 나타낼 수 있다.

수학식 9

$$\hat{\Phi}_{SS}(t+1,k) = \zeta_s \hat{\Phi}_{SS}(t,k) + (1-\zeta_s) \Phi_{SS}(t,k) \\ \hat{\Phi}_{DD}(t+1,k) = \zeta_d \hat{\Phi}_{DD}(t,k) + (1-\zeta_d) \Phi_{DD}(t,k)$$

상기 수학식 9에서, $\hat{\Phi}_{DD}(t,k)$ 와 $\hat{\Phi}_{SS}(t,k)$ 는 각각 $\Phi_{DD}(t,k)$ 과 $\Phi_{SS}(t,k)$ 의 추정 값이고, ζ_d 과 ζ_s 는 정상(stationary) 상태를 가정한 스무딩 파라미터로서, $0 < \zeta_d, \zeta_s < 1$ 의 범위를 갖는다.

음성 신호 $S(t)$ 와 잡음 신호 $D(t)$ 의 통계적 가정과 상기 수학식 9를 이용하여 현재 프레임에 대한 음성 신호와 잡음 신호의 전력 추정 값을 계산하면 다음 수학식 10 및 11과 같이 나타낼 수 있다.

수학식 10

$$\Phi_{SS}(t,k) = E[|S(t,k)|^2 | Y(t,k), H_0] p(H_0 | Y(t,k)) \\ + E[|S(t,k)|^2 | Y(t,k), H_1] p(H_1 | Y(t,k))$$

수학식 11

$$\begin{aligned} \Phi_{DD}(t,k) &= E[|D(t,k)|^2 | Y(t,k), H_0] p(H_0 | Y(t,k)) \\ &\quad + E[|D(t,k)|^2 | Y(t,k), H_1] p(H_1 | Y(t,k)) \end{aligned}$$

[0049]

[0050] 상기 수학식 10 및 11은, 다시 다음 수학식 12 및 13으로 나타낼 수 있다.

수학식 12

$$\begin{aligned} E[|D(t,k)|^2 | Y(t,k), H_0] &= \Phi_{YY}(t,k) \\ E[|D(t,k)|^2 | Y(t,k), H_1] \\ &= \left(\frac{\hat{\xi}(t,k)}{1 + \hat{\xi}(t,k)} \right) \hat{\Phi}_{DD}(t,k) + \left(\frac{1}{1 + \hat{\xi}(t,k)} \right)^2 \Phi_{YY}(t,k) \end{aligned}$$

[0051]

수학식 13

$$\begin{aligned} E[|S(t,k)|^2 | Y(t,k), H_0] &= 0 \\ E[|S(t,k)|^2 | Y(t,k), H_1] \\ &= \left(\frac{1}{1 + \hat{\xi}(t,k)} \right) \hat{\Phi}_{SS}(t,k) + \left(\frac{\hat{\xi}(t,k)}{1 + \hat{\xi}(t,k)} \right)^2 \Phi_{YY}(t,k) \end{aligned}$$

[0052]

[0053] 상기 수학식 12와 수학식 13에서 사용된 $\hat{\xi}(t,k)$ 는 predicted SNR로 정의되며, 추정된 잡음 전력과 음성 전력의 비로 다음 수학식 14와 같이 나타내어진다.

수학식 14

$$\hat{\xi}(t,k) \equiv \frac{\Phi_{SS}(t,k)}{\Phi_{DD}(t,k)}$$

[0054]

[0055] 수학식 14와 같이 구해진 predicted SNR은 잡음 전력의 갱신을 위해 사용될 수 있다.

[0056] 3. 실험 결과

[0057] 본 발명에서 제안하고 있는 음성 강화 방법의 성능을 평가하기 위해서 ITU-T P.800의 표준 음질 측정 방법인 CCR(Comparison Category Rating) 테스트를 실시하였다. 제안된 음성 강화 방법은 추정된 근단 화자의 배경 잡음 전력에 기초하여 오염된 원단 음성 신호 중 잡음을 제외한 실제 음성 구간을 증폭시켜서 배경 잡음이 존재하는 근단 화자가 듣게 되는 원단 음성 신호의 명료성 향상을 목적으로 한다. 따라서 원단 음성 신호가 오염된 신호임을 가정하기 위해, 남성과 여성 화자가 각각 10개의 문장을 발음한 음성을 8kHz로 샘플링한 후에 NOISEX-92 데이터베이스의 백색 가우시안 잡음(white gaussian noise; WGN)을 SNR 10dB로 부가하였다. 실제 배경 잡음이 존재하는 환경에서 실험을 진행하는 대신에 오염된 원단 음성 신호와 강화된 근단 음성 신호에 근단의 배경 잡음으로서 NOISEX-92 데이터베이스의 백색 가우시안 잡음(WGN), 배블 잡음(babble noise), 및 차량 잡음(vehicle noise)을 잡음 레벨이 SNR 5dB, 10dB, 15dB, 20dB로 섞이도록 하였고, 헤드폰을 통해 들을 수 있게 실험을 진행하였다.

[0058] 총 14명의 청자들에 의해 주관적 테스트 방법인 CCR 테스트가 진행되었으며, 14명의 청자들은 3가지 종류의 음성 파일을 듣게 된다. 첫 번째 파일은 reference 파일로서, 잡음이 없는 깨끗한 음성 파일을 참고용으로 듣게

된다. 이후, 14명의 청자는 2개의 파일을 각각 듣게 되는데, 하나의 파일은 제안된 알고리즘이 적용된 파일이고, 다른 하나의 파일은 제안된 알고리즘과 비교 대상이 되는 파일이다. 이때 참고용 reference 파일을 제외한 비교 대상 2개의 파일은 순서 없이 재생된다. reference 파일을 기준으로 2개의 파일 중 어떤 음성 신호가 얼마나 음질이 좋은지, 배경 잡음에 대해 음성 신호가 얼마나 명료하게 들리는지, 얼마나 크게 잘 들리는지, 왜곡은 없는지를 평가 기준으로 -3점부터 +3점까지 주관적인 점수를 매기게 된다. 표에 기록된 총 점수는 평균값으로 나타내었는데, 결과에 나타난 점수가 0보다 큰 양수 값을 가질수록 제안된 알고리즘이 성능이 더 좋다는 것을 의미하고, 반대로 0보다 작은 음수의 평균값을 가지게 된다면 강화된 음성 신호의 성능이 나쁘다는 것을 의미하게 된다.

[0059]

첫 번째 실험의 목적은 근단 배경 잡음이 존재하는 환경에서 근단 화자가 오염된 원단 음성을 들었을 경우, 어떠한 음성 처리 기법도 적용되지 않고 전송된 음성 신호 대비 제안된 음성 강화 기법에 의해 강화 원단 음성 신호가 얼마나 더 명료하게 들리는가를 평가하기 위한 것이다. 표 1에 나타난 결과에 따르면, 모든 SNR에서 본 발명에서 제안하고 있는 음성 강화 알고리즘을 적용하여 강화된 음성 신호가 어떠한 음성 처리 기법도 거치지 않고 근단에 전송된 오염된 원단 음성 신호 대비 평균값이 0보다 큰 양수 값을 가짐을 볼 수 있는데, 이를 통해 제안된 음성 강화 알고리즘의 성능이 우수함을 확인할 수 있었다.

표 1

[0060]

잡음	SNR (dB)	점수
백색 잡음	5	1.49 ± 0.14
	10	1.46 ± 0.16
	15	1.36 ± 0.19
	20	1.33 ± 0.20
배틀 잡음	5	1.49 ± 0.14
	10	1.50 ± 0.15
	15	1.48 ± 0.19
	20	1.55 ± 0.18
차량 잡음	5	1.61 ± 0.14
	10	1.87 ± 0.15
	15	1.70 ± 0.19
	20	1.68 ± 0.21

[0061]

두 번째 실험은 기존의 Sauert와 Vary가 제안한 SNR 복구 방법이 적용된 오염된 원단 음성 신호 대비 본 발명에서 제안하고 있는 음성 강화 방법이 적용된 오염된 원단 음성 신호와의 비교를 통해 근단 화자가 듣게 되는 음성의 명료성을 평가하기 위한 것이다. 표 2에 나타난 결과를 보면, 배틀 잡음, 차량 잡음, 백색 잡음에서 모두 0보다 큰 양수의 평균값을 나타냄을 확인할 수 있다. 이것은 기존의 SNR 복구 방법보다 본 발명에서 제안하고 있는 음성 강화 방법이 더 우수한 성능을 나타냄을 보여준다. 그러나 표 2에 나타난 점수가 일관된 결과를 보여주지 못하고 있는데, 이것은 잡음의 주파수 성격에 따라 음성 강화의 효과가 다르기 때문이라고 추정된다.

표 2

[0062]

잡음	SNR (dB)	점수
백색 잡음	5	0.81 ± 0.17
	10	0.87 ± 0.18
	15	1.12 ± 0.20
	20	1.09 ± 0.23
배틀 잡음	5	0.46 ± 0.17
	10	0.41 ± 0.18
	15	0.39 ± 0.20
	20	0.31 ± 0.20

차량 잡음	5	0.00 ± 0.15
	10	0.19 ± 0.15
	15	0.10 ± 0.16
	20	0.17 ± 0.19

[0063] 본 발명에서 제안하고 있는 음성 강화 방법이 기존의 SNR 복구 방법보다 더 우수한 성능을 보인다는 것은 도 3을 통해서도 확인할 수 있다. 도 3에서, (a)는 오염된 원단 음성 신호를 나타내고, (b)는 SNR 복구 방법에 의해 음성과 잡음 구간이 모두 증폭된 음성 신호 파형을 나타내며, (c)는 본 발명에서 제안하고 있는 음성 강화 방법에 의해 잡음이 섞이지 않은 깨끗한 음성 신호 전력만 증폭시킨 것을 나타내고 있다. 도 3에 도시된 결과로부터, 기존의 방법과 비교해 본 발명에서 제안하고 있는 음성 강화 방법이 음질을 향상시키고 있음을 확인할 수 있다.

[0064] 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.

도면의 간단한 설명

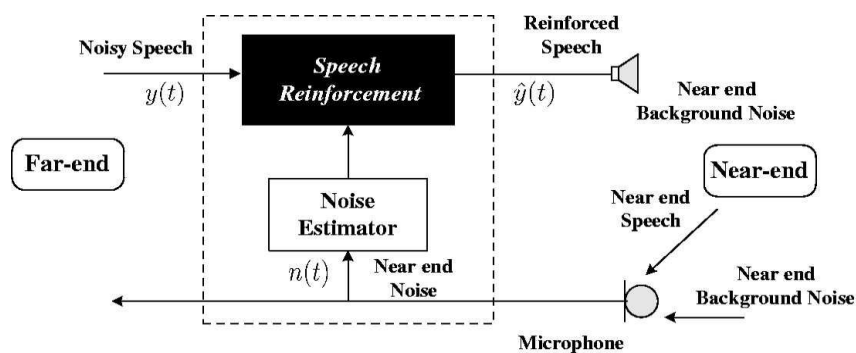
[0065] 도 1은 오염된 음성 전력 스펙트럼에서 실제 잡음이 섞이지 않은 깨끗한 음성 신호 전력만을 추정하여 근단에 존재하는 배경 잡음 전력과 일정한 SNR을 유지하도록 이득을 수정하여 원단의 실제 음성 신호 전력을 증폭시키는 방식의 음성 강화 방법의 개념을 나타내는 도면.

[0066] 도 2는 본 발명의 일 실시예에 따른 음성 강화 방법의 순서도를 나타내는 도면.

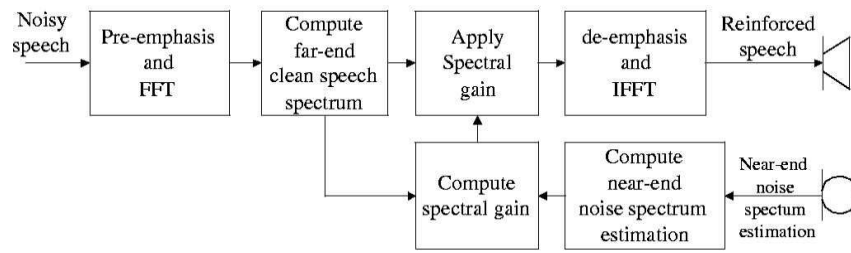
[0067] 도 3은 본 발명의 일 실시예에 따른 음성 강화 방법을 종래 기술과 비교하여 도시한 도면으로서, (a)는 원단의 잡음이 섞인 오염된 음성 신호의 경우를, (b)는 기존의 다른 기술에 의해 잡음과 음성이 동시에 증폭한 경우를, (c)는 본 발명에서 제안하고 있는 음성 강화 방법에 의해 잡음이 섞이지 않은 음성 신호만 강화한 경우를 나타내는 도면이다.

도면

도면1



도면2



도면3

