



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2012-0005920
(43) 공개일자 2012년01월17일

(51) Int. Cl.

G10L 21/02 (2006.01)

(21) 출원번호 10-2010-0066639

(22) 출원일자 2010년07월09일

심사청구일자 2010년07월09일

(71) 출원인

인하대학교 산학협력단

인천광역시 남구 인하로 100, 인하대학교 (용현동, 인하공업전문대학)

(72) 발명자

장준혁

서울특별시 서초구 서초대로52길 41, B동 301호 (서초동)

박윤식

서울특별시 은평구 통일로68길 4, 401호 (불광동, 원영빌딩)

(74) 대리인

김진우

전체 청구항 수 : 총 4 항

(54) 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법 및 장치

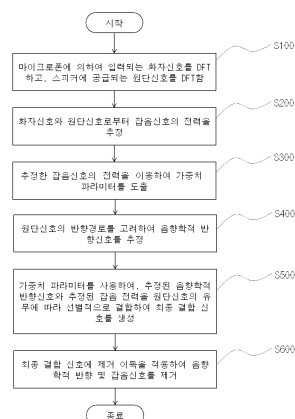
(57) 요약

본 발명은 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법 및 장치에 관한 것으로서, 보다 구체적으로는 (1) 마이크로폰에 의하여 입력되는 화자신호를 이산 푸리에 변환(Discrete Fourier Transform, 이하 'DFT' 라 한다.)하고, 스피커에 공급되는 원단신호를 DFT하는 단계, (2) 상기 화자신호와 상기 원단신호로부터 잡음신호의 전력을 추정하는 단계, (3) 상기 추정된 잡음신호의 전력비를 이용하여 가중치 파라미터를 도출하는 단계, (4) 상기 원단신호의 반향경로를 고려하여 음향학적 반향신호를 추정하는 단계, (5) 상기 가중치 파라미터를 사용하여, 상기 추정된 음향학적 반향신호와 상기 추정된 잡음 전력을 상기 원단신호의 유무에 따라 선별적으로 결합하여 최종 결합 신호를 생성하는 단계 및 (6) 상기 최종 결합 신호에 제거 이득을 적용하여 음향학적 반향 및 잡음신호를 제거하는 단계를 포함하는 것을 그 구성상의 특징으로 한다.

본 발명에서 제안하고 있는 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법 및 장치에 따르면, 선(pre)반향-후(post)잡음 제거 또는 선(pre)잡음-후(post)반향 제거의 통합 구조에서 발생할 수 있는 단점을 보완하기 위해, 가중치 파라미터를 사용하여 결합한 음향학적 반향 및 잡음 전력에 기반하여 반향 및 잡음을 제거함으로써, 종래 기술에서 발생할 수 있는 구조적 문제점을 해결할 수 있다.

또한, 본 발명에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법 및 장치는, 음향학적 반향 및 잡음 전력을 선별적으로 결합한 후, 결합된 신호에 하나의 위너 필터로 제거 이득을 적용하여 음향학적 반향 신호와 배경잡음 신호를 동시에 제거함으로써, 음향학적 반향 제거기와 배경잡음 제거기에 각각 제거필터가 사용되는 종래의 기술에 비하여 반향 및 잡음신호 제거의 성능을 대폭 향상시킬 수 있다.

대표도 - 도3



이 발명을 지원한 국가연구개발사업

과제고유번호 2008-F-045-01

부처명 정보통신부

연구관리전문기관

연구사업명 IT산업경쟁력강화산업

연구과제명 장애인 및 고령자를 위한 Digital Guardian 기술개발

기여율

주관기관 한국과학기술연구원

연구기간 2010년 03월 01일 ~ 2012년 02월 28일

특허청구의 범위

청구항 1

- (1) 마이크로폰에 의하여 입력되는 화자신호를 이산 푸리에 변환(Discrete Fourier Transform, 이하 ‘DFT’ 라 한다.)하고, 스피커에 공급되는 원단신호를 DFT하는 단계;
- (2) 상기 화자신호와 상기 원단신호로부터 잡음신호의 전력을 추정하는 단계;
- (3) 상기 추정된 잡음신호의 전력을 이용하여 가중치 파라미터를 도출하는 단계;
- (4) 상기 원단신호의 반향 경로를 고려하여 음향학적 반향신호를 추정하는 단계;
- (5) 상기 가중치 파라미터를 사용하여, 상기 추정된 음향학적 반향신호와 상기 추정된 잡음 전력을 상기 원단신호의 유무에 따라 선별적으로 결합하여 최종 결합 신호를 생성하는 단계; 및
- (6) 상기 최종 결합 신호에 제거 이득을 적용하여 음향학적 반향 및 잡음신호를 제거하는 단계를 포함하는 것을 특징으로 하는 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법.

청구항 2

제1항에 있어서, 상기 가중치 파라미터는,

다음 수학적식에 의하여 계산되는 것을 특징으로 하는 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법.

$$q(i, k) = \alpha q(i-1, k) + (1 - \alpha) \Gamma(i, k)$$

여기서, γ_{TH} 는 문턱 값, α 는 상수, $\Gamma(i, k)$ 는 아래 수학적식에 의하여 가변 값을 갖게 되는 선행 확률로서 H_0 이면 0, H_1 이면 1을 갖는 인덱스 함수이다.

$$\gamma(i, k) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{TH}$$

청구항 3

제1항에 있어서, 상기 단계 (6)에서,

위너 필터에 기반하여 상기 최종 결합 신호에 제거 이득을 적용하는 것을 특징으로 하는 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법.

청구항 4

마이크로폰에 의하여 입력되는 화자신호를 이산 푸리에 변환(Discrete Fourier Transform, 이하 ‘DFT’ 라 한다.) 또는 역 이산 푸리에 변환(Inverse Discrete Fourier Transform, 이하 ‘IDFT’ 라 한다.)하거나, 스피커에 공급되는 원단신호를 DFT 또는 IDFT하는 푸리에 변환기;

상기 화자신호와 상기 원단신호를 이용하여, 상기 원단신호로부터 반향 경로를 고려한 임펄스응답을 추정하는 에코 경로 응답기(Echo Path Response);

상기 화자신호로부터 잡음신호 전력을 추정하는 잡음 전력 추정기(Noise Power Estimation);

상기 원단신호로부터 음향학적 반향신호를 추정하는 에코 신호 추정기(Estimated Echo Signal);

상기 추정된 음향학적 반향신호와 상기 추정된 잡음신호 전력을 상기 원단신호의 유무에 따라 선별적으로 결합하는 결합 신호 추정기(Unified Signal Estimation); 및

상기 결합 신호 추정기에 의하여 결합된 결합 신호에 위너필터(Wiener Filter)를 적용시킴으로써, 음향학적 반향과 배경잡음을 제거하는 반향 잡음 제거기를 포함하는 것을 특징으로 하는 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 장치.

명세서

기술 분야

- [0001] 본 발명은 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법 및 장치에 관한 것으로서, 특히, 원단 신호에 기반한 가중치 파라미터를 이용하여 음향학적 반향 및 배경잡음 전력을 선별적으로 결합하고 결합된 신호를 하나의 제거 필터에 적용하여 반향 및 잡음을 모두 제거하는 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법 및 장치에 관한 것이다.

배경 기술

- [0002] 이동통신 기술의 발전과 함께 차량용 핸드프리나 화상통화가 가능한 휴대용 단말기의 보급이 확산되면서 관련 음성 통신 방식에서 발생할 수 있는 음향학적 반향(acoustic echo)이나 배경잡음 제거에 대한 연구가 주목받고 있다.

- [0003] 최근에는 이와 같은 음향학적 반향 및 잡음 제거기의 통합 구조가 일반화되면서 각 알고리즘에 대한 상호 특성을 고려해 전체 알고리즘의 성능을 향상시키기 위한 연구가 활발히 진행되고 있다. 일반적인 통합구조는 두 알고리즘이 서로 독립이라고 가정하고 병렬로 설계되어 직렬로 나열하는 방식이지만 실제로는 서로 간의 위치에 따라 성능의 변화가 큰 것으로 알려져 있다.

- [0004] 구체적으로, 반향 제거기가 잡음 제거기의 앞 단에 위치하게 되면, 잡음 제거기가 추정해야할 잡음의 스펙트럼이 반향 제거기에 의해 왜곡이 생겨 잡음추정이 심각히 저하되고, 반대로 잡음 제거기가 반향 제거기의 앞 단에 위치하게 되면, 잡음 제거기의 비선형적인 연산으로 인해 반향 제거기의 성능이 저하되는 구조적 문제점이 존재한다. 특히 이러한 구조적 문제점은 실제로 반향구간에서의 신호 왜곡을 통해 효과적인 반향신호 제거가 어려워 잔여 반향신호를 발생시킬 수 있으며, 또한 부정확한 잡음추정에 의해 음성구간에서의 왜곡 및 신호대 잡음비(SNR, signal-to-noise ratio)를 감소시킬 수 있다는 문제점이 있다.

발명의 내용

해결하려는 과제

- [0005] 본 발명은 기존에 제안된 방법들의 상기와 같은 문제점들을 해결하기 위해 제안된 것으로서, 선(pre)반향-후(post)잡음 제거 또는 선(pre)잡음-후(post)반향 제거의 통합 구조에서 발생할 수 있는 단점을 보완하기 위해, 가중치 파라미터를 사용하여 결합한 음향학적 반향 및 잡음 전력에 기반하여 반향 및 잡음을 제거함으로써, 종래 기술에서 발생할 수 있는 구조적 문제점을 해결할 수 있는 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법 및 장치를 제공하는 것을 그 목적으로 한다.

- [0006] 또한 본 발명은, 음향학적 반향 및 잡음 전력을 선별적으로 결합한 후, 결합된 신호에 하나의 위너 필터로 제거이득을 적용하여 음향학적 반향 신호와 배경잡음 신호를 동시에 제거함으로써, 음향학적 반향 제거기와 배경잡음 제거기에 각각 제거필터가 사용되는 종래의 기술에 비하여 반향 및 잡음신호 제거의 성능을 대폭 향상시킬 수 있는 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법 및 장치를 제공하는 것을 다른 목적으로 한다.

과제의 해결 수단

- [0007] 상기한 목적을 달성하기 위한 본 발명의 특징에 따른, 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법은,
- [0008] (1) 마이크론에 의하여 입력되는 화자신호를 이산 푸리에 변환(Discrete Fourier Transform, 이하 ‘DFT’ 라 한다.)하고, 스피커에 공급되는 원단신호를 DFT하는 단계;
- [0009] (2) 상기 화자신호와 상기 원단신호로부터 잡음신호의 전력을 추정하는 단계;
- [0010] (3) 상기 추정한 잡음신호의 전력을 이용하여 가중치 파라미터를 도출하는 단계;
- [0011] (4) 상기 원단신호의 반향 경로를 고려하여 음향학적 반향신호를 추정하는 단계;
- [0012] (5) 상기 가중치 파라미터를 사용하여, 상기 추정된 음향학적 반향신호와 상기 추정된 잡음 전력을 상기 원단신호의 유무에 따라 선별적으로 결합하여 최종 결합 신호를 생성하는 단계; 및
- [0013] (6) 상기 최종 결합 신호에 제거 이득을 적용하여 음향학적 반향 및 잡음신호를 제거하는 단계를 포함하는 것을 그 구성상의 특징으로 한다.
- [0014] 바람직하게는, 상기 가중치 파라미터는,
- [0015] 다음 수학식에 의하여 계산될 수 있다.
- [0016]
$$q(i, k) = \alpha q(i-1, k) + (1 - \alpha) \Gamma(i, k)$$
- [0017] 여기서, γ_{TH} 는 문턱 값, α 는 상수, $\Gamma(i, k)$ 는 아래 수학식에 의하여 가변 값을 갖게 되는 선행 확률로서 H_0 이면 0, H_1 이면 1을 갖는 인덱스 함수이다.
- [0018]
$$\gamma(i, k) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{TH}$$
- [0019] 바람직하게는, 상기 단계 (6)에서,
- [0020] 위너 필터에 기반하여 상기 최종 결합 신호에 제거 이득을 적용할 수 있다.
- [0021] 상기한 목적을 달성하기 위한 본 발명의 특징에 따른, 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 장치는,
- [0022] 마이크론에 의하여 입력되는 화자신호를 이산 푸리에 변환(Discrete Fourier Transform, 이하 ‘DFT’ 라 한다.) 또는 역 이산 푸리에 변환(Inverse Discrete Fourier Transform, 이하 ‘IDFT’ 라 한다.)하거나, 스피커에 공급되는 원단신호를 DFT 또는 IDFT하는 푸리에 변환기;
- [0023] 상기 화자신호와 상기 원단신호를 이용하여, 상기 원단신호로부터 반향 경로를 고려한 임펄스응답을 추정하는 에코 경로 응답기(Echo Path Response);
- [0024] 상기 화자신호로부터 잡음신호 전력을 추정하는 잡음 전력 추정기(Noise Power Estimation);
- [0025] 상기 원단신호로부터 음향학적 반향신호를 추정하는 에코 신호 추정기(Estimated Echo Signal);
- [0026] 상기 추정된 음향학적 반향신호와 상기 추정된 잡음신호 전력을 상기 원단신호의 유무에 따라 선별적으로 결합하는 결합 신호 추정기(Unified Signal Estimation); 및
- [0027] 상기 결합 신호 추정기에 의하여 결합된 결합 신호에 위너필터(Wiener Filter)를 적용시킴으로써, 음향학적 반향과 배경잡음을 제거하는 반향 잡음 제거기를 포함하는 것을 그 구성상의 특징으로 한다.

발명의 효과

[0028] 본 발명에서 제안하고 있는 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법 및 장치에 따르면, 선(pre)반향-후(post)잡음 제거 또는 선(pre)잡음-후(post)반향 제거의 통합 구조에서 발생할 수 있는 단점을 보완하기 위해, 가중치 파라미터를 사용하여 결합한 음향학적 반향 및 잡음 전력에 기반하여 반향 및 잡음을 제거함으로써, 종래 기술에서 발생할 수 있는 구조적 문제점을 해결할 수 있다.

[0029] 또한, 본 발명에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법 및 장치는, 음향학적 반향 및 잡음 전력을 선별적으로 결합한 후, 결합된 신호에 하나의 위너 필터로 제거 이득을 적용하여 음향학적 반향 신호와 배경잡음 신호를 동시에 제거함으로써, 음향학적 반향 제거기와 배경잡음 제거기에 각각 제거필터가 사용되는 종래의 기술에 비하여 반향 및 잡음신호 제거의 성능을 대폭 향상시킬 수 있다.

도면의 간단한 설명

[0030] 도 1은 일반적인 음향학적 반향 및 잡음 제거 시스템의 블록도.
 도 2는 일반적인 음향학적 반향 및 잡음 제거 시스템의 블록도.
 도 3은 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법의 순서도.
 도 4는 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 장치의 블록도.
 도 5는 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법 및 종래 기술에 의한 ERLE와 SA 수치를 비교하여 나타낸 도면.
 도 6은 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법과 종래 기술에 의한 시간에 따른 ERLE 변화를 나타내는 도면.
 도 7은 도 6의 음성신호에 대한 스펙트럼 분석을 위한 음성 스펙트로그램을 나타내는 도면.

발명을 실시하기 위한 구체적인 내용

[0031] 이하에서는 첨부된 도면을 참조하여 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명을 용이하게 실시할 수 있도록 바람직한 실시예를 상세히 설명한다. 다만, 본 발명의 바람직한 실시예를 상세하게 설명함에 있어, 관련된 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략한다. 또한, 유사한 기능 및 작용을 하는 부분에 대해서는 도면 전체에 걸쳐 동일 또는 유사한 부호를 사용한다.

[0032] 덧붙여, 명세서 전체에서, 어떤 부분이 다른 부분과 ‘연결’되어 있다고 할 때, 이는 ‘직접적으로 연결’되어 있는 경우뿐만 아니라, 그 중간에 다른 소자를 사이에 두고 ‘간접적으로 연결’되어 있는 경우도 포함한다. 또한, 어떤 구성요소를 ‘포함’한다는 것은, 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있다는 것을 의미한다.

[0033] 도 1과 도 2는 일반적인 음향학적 반향 및 잡음 제거 시스템의 블록도이다. 도 1은 선반향-후잡음 제거 통합 구조이며, 도 2는 선잡음-후반향 제거 통합 구조이다. 도 1과 도 2에 도시된 바와 같이, 기존의 통합 구조는 음향학적 반향 시스템을 구성하는 주파수 영역 기반 AES(acoustic echo suppression)와 및 잡음 제거 시스템을 구성하는 NS(noise suppression) 알고리즘이 직렬구조로 연결되어 있다. 이러한 통합 구조에서 우선 주파수 영역 기반 AES에 대하여 살펴보면, 반향신호 $d(t)$, 배경잡음 $w(t)$, 근단 (near-end) 화자신호 $s(t)$, 원단 (far-end) 신호와 마이크 입력신호를 각각 $x(t)$, $y(t)$ 라 하고, $Y(i,k)$ 를 $y(t)$ 의 i 번째 프레임의 k 번째 주파수 성분이 라 하면 입력신호 $Y(i,k)$ 로부터 반향신호가 제거된 신호 $\hat{E}(i,k)$ 는 수학식 1과 같이 표현된다.

수학식 1

[0034]

$$\hat{E}(i,k) = G_{AES}(i,k) Y(i,k)$$

[0035]

여기서, $G_{AES}(i,k)$ 는 AES 이득(gain)으로서, 수학식 2와 같은 위너 필터(Wiener Filter)를 사용한다.

수학식 2

[0036]

$$G_{AES}(i,k) = \frac{\max(|Y(i,k)| - |\hat{Y}(i,k)|, 0)}{|Y(i,k)|}$$

[0037]

수학식 2에서 $\hat{Y}(i,k)$ 는 원단신호 $X(i,k)$ 로부터 추정된 음향학적 반향신호이다. 일반적으로 입력 마이크로폰으로 전달되는 음향학적 반향신호는 다양한 반향 경로를 거쳐 입력되기 때문에 원단신호로부터 반향 경로를 고려한 임펄스응답에 대한 추정 과정이 필요하다. 따라서 반향 경로의 임펄스응답을 고려하여 원단신호의 주파수 성분으로부터 추정된 반향신호 $\hat{Y}(i,k)$ 는 least squares(최소 제곱) 추정에 기반하여 수학식 3과 같이 나타낼 수 있다.

수학식 3

[0038]

$$\hat{Y}(i,k) = \left| \frac{E\{X_d^*(i,k) Y(i,k)\}}{E\{X_d^*(i,k) X_d(i,k)\}} \right| |X(i,k)|$$

[0039]

여기서, *는 complex conjugate를 의미하고, $X_d(i,k)$ 는 샘플 지연된(delayed) 원단신호의 주파수 성분을 나타낸다.

[0040]

잡음 제거(Noise Suppression; NS)와 관련하여 살펴보면, 일반적으로 주파수 영역에 기반한 잡음 제거는 수학식 4와 같이 잡음 제거 이득 $G_{NS}(i,k)$ 과 잡음이 섞여 있는 입력신호 $\hat{E}(i,k)$ 의 곱으로 표현된다.

수학식 4

[0041]

$$\hat{S}(i,k) = G_{NS}(i,k) \hat{E}(i,k)$$

[0042]

여기서, $G_{NS}(i,k)$ 는 추정기법에 따라 위너 필터 또는 MMSE(Minimum Mean Square Error, 최소평균제곱오차) 추정에 기반한 다양한 제거 이득이 적용가능하며, 본 발명에서는 상용 코덱인 IS-127 EVRC(Enhanced Variable Rate Codec)에 적용되는 잡음 제거기법을 통합구조의 잡음 제거기로서 적용하였다.

[0043]

이와 같이 기존의 잡음 및 반향 제거 구조는 AES 및 NS 알고리즘이 독립적으로 설계되고 직렬로 연결되어 선반향-후잡음 또는 선잡음-후반향의 통합구조로 각각의 신호를 순차적으로 제거하게 된다. 그러나 선반향-후잡음 통합구조에서는 반향 제거기에 의해 선처리된 신호는 잡음 스펙트럼이 반향 제거기에 의해 왜곡이 생겨 잡음 제거 알고리즘에서 정확한 잡음전력 추정이 어려우며, 반대로 선잡음-후반향 통합구조에서는 잡음 제거기의 비선형적인 연산으로 인해 반향 제거기의 성능이 저하되는 구조적 문제점이 있다.

[0044] 도 3은 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법의 순서도이다. 도 3에 도시된 바와 같이, 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법은, 마이크로폰에 의하여 입력되는 화자신호를 이산 푸리에 변환(Discrete Fourier Transform, 이하 ‘DFT’라 한다.)하고, 스피커에 공급되는 원단신호를 DFT하는 단계(S100), 화자신호와 원단신호로부터 잡음신호의 전력을 추정하는 단계(S200), 추정된 잡음신호의 전력을 이용하여 가중치 파라미터를 도출하는 단계(S300), 원단신호의 반향경로를 고려하여 음향학적 반향신호를 추정하는 단계(S400), 가중치 파라미터를 사용하여, 추정된 음향학적 반향신호와 추정된 잡음 전력을 원단신호의 유무에 따라 선별적으로 결합하여 최종 결합 신호를 생성하는 단계(S500), 최종 결합 신호에 제거 이득을 적용하여 음향학적 반향 및 잡음신호를 제거하는 단계(S600)를 포함하여 구성될 수 있다.

[0045] 단계 S100에서는, 마이크로폰에 의하여 입력되는 화자신호를 DFT하고, 스피커에 공급되는 원단신호를 DFT한다. 화자신호와 원단신호를 푸리에 변환하는 것은, 각 주파수 별로 잡음과 반향을 제거하기 위함이다.

[0046] 단계 S200에서는, 화자신호와 원단신호로부터 잡음신호의 전력을 추정한다. 종래 기술과 달리, 원단(Far-end)신호와 마이크로폰 입력신호인 화자신호로부터, 잡음신호의 전력을 추정하고 이를 이용하여 가중치 파라미터를 도출한다. 잡음신호의 전력비를 사용함으로써 종래 기술에서 대두되는 구조적 문제점을 해결할 수 있다.

[0047]

[0048] 단계 S300에서는, 추정된 잡음신호의 전력을 이용하여 가중치 파라미터를 도출한다. 이때 가중치 파라미터는, 수학적 식 5에 의하여 계산될 수 있다.

수학적 식 5

$$q(i, k) = \alpha q(i-1, k) + (1 - \alpha) \Gamma(i, k)$$

[0049]

[0050] 여기서, γ 는 문턱 값, α 는 상수, $\Gamma(i, k)$ 는 수학적 식 6에 의하여 가변 값을 갖게 되는 선행 확률로서 H_0 이면 0, H_1 이면 1을 갖는 인덱스 함수이다.

수학적 식 6

$$\gamma(i, k) \underset{H_0}{\overset{H_1}{\geq}} \gamma_{TH}$$

[0051]

[0052] 단계 S400에서는, 원단신호의 반향경로를 고려하여 음향학적 반향신호를 추정한다. 입력 마이크로폰으로 전달되는 음향학적 반향신호는 다양한 반향 경로를 거쳐 입력되기 때문에, 원단신호로부터 반향 경로를 고려한 임펄스 응답에 대한 추정 과정이 필요하다. 따라서 반향 경로의 임펄스 응답을 고려하여 원단신호의 주파수 성분 $X(i, k)$ 로부터 추정된 반향신호 $\hat{Y}(i, k)$ 는 least squares 추정에 기반하여 수학적 식 3으로 나타낼 수 있다.

[0053] 단계 S500에서는, 가중치 파라미터를 사용하여, 추정된 음향학적 반향신호와 추정된 잡음 전력을 원단신호의 유무에 따라 선별적으로 결합하여 최종 결합 신호를 생성한다. 최종 결합 신호를 생성함으로써, 하나의 제거 이득을 적용하는 것만으로도 반향 및 잡음의 제거가 가능하므로, 본 발명은 반향과 잡음에 각각 제거 이득을 적용하는 종래의 기술에 비하여 효과적으로 음성을 처리할 수 있다.

[0054] 단계 S600에서는, 최종 결합 신호에 제거 이득을 적용하여 음향학적 반향 및 잡음신호를 제거한다. 이때 위너 필터에 기반하여 제거 이득을 적용할 수 있으며, 음향학적 반향 및 잡음신호 제거를 위한 제거 이득을 $G_{pro}(i, k)$ 라 하면 제거된 최종 신호 $E(i, k)$ 는 수학식 7과 같이 입력신호 $Y(i, k)$ 와 제거 이득의 곱으로 나타낼 수 있다.

수학식 7

$$[0055] \quad E(i, k) = G_{pro}(i, k) Y(i, k)$$

[0056] 여기서, $G_{pro}(i, k)$ 는 위너 필터로서 수학식 8과 같이 표현된다.

수학식 8

$$[0057] \quad G_{pro}(i, k) = \frac{\xi(i, k)}{1 + \xi(i, k)}$$

[0058] 또한, $\xi(i, k)$ 는 반향 및 잡음신호가 제거된 결과신호의 전력 $\lambda_{output}(i, k)$ 과 결합 전력 $\lambda_{combined}(i, k)$ 의 비로서, 수학식 9와 같다.

수학식 9

$$[0059] \quad \xi(i, k) = \frac{\lambda_{output}(i, k)}{\lambda_{combined}(i, k)}$$

[0060] 여기서, $\lambda_{output}(i, k)$ 은 수학식 3으로부터 유도된 반향 경로를 고려하여 추정된 음향학적 반향신호의 전력 $\lambda_{echo}(i, k)$ 과 EVRC의 잡음 제거 알고리즘으로부터 추정된 잡음전력 $\lambda_{noise}(i, k)$ 의 가중치 합으로 수학식 10과 같이 유도된다.

수학식 10

$$[0061] \quad \lambda_{combined}(i, k) = q(i, k)\lambda_{echo}(i, k) + (1 - q(i, k))\lambda_{noise}(i, k)$$

[0062] 여기서, $q(i, k)$ 는 음향학적 반향과 잡음 전력을 선별적으로 결합하기 위한 가중치 파라미터로서 단계 S300에서 이미 설명한 바와 같다.

[0063] 반향 및 잡음 제거의 성능을 더욱 개선시키기 위하여, 수학식 9는 Decision-Directed 추정 기법에 기반하여 수학식 11과 같이 도출될 수 있다.

수학식 11

$$[0064] \quad \hat{\xi}(i, k) = \alpha_d \frac{|E(i-1, k)|^2}{\lambda_{combined}(i-1, k)} + (1 - \alpha_d) \left[\frac{|Y(i, k)|^2}{\lambda_{combined}(i, k)} - 1 \right]$$

- [0065] 여기서, α_d 는 스무딩 파라미터(Smoothing Parameter)이다.
- [0066] 도 4는 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 장치의 블록도이다. 도 4에 도시된 바와 같이, 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 장치는, 푸리에 변환기, 에코 경로 응답기, 잡음 전력 추정기, 에코 신호 추정기, 결합 신호 추정기, 반향 잡음 제거기를 포함하여 구성될 수 있다.
- [0067] 푸리에 변환기는, 마이크로폰에 의하여 입력되는 화자신호를 DFT 또는 역 이산 푸리에 변환(Inverse Discrete Fourier Transform, 이하 'IDFT' 라 한다.)하거나, 스피커에 공급되는 원단신호를 DFT 또는 IDFT한다. 푸리에 변환기는, 단계 S100에서의 과정을 동일하게 수행하므로 자세한 설명은 생략하도록 한다.
- [0068] 에코 경로 응답기(Echo Path Response)는, 화자신호와 원단신호를 이용하여, 원단신호로부터 반향 경로를 고려한 임펄스응답을 추정한다. 마이크로폰으로 입력되는 음향학적 반향신호의 반향 경로가 매우 다양하기 때문에, 임펄스 응답에 대한 추정을 하여야 하는데 이때 원단신호로부터의 반향 경로를 고려할 수 있다. 임펄스 응답 추정은, 단계 S400에서 이미 설명한 바와 같으므로 자세한 설명은 생략하도록 한다.
- [0069] 잡음 전력 추정기(Noise Power Estimation)는, 화자신호를 DFT하여 얻은 특정 주파수 성분으로부터 잡음신호 전력을 추정한다. 잡음 전력 추정기는, 단계 S200에서의 과정을 동일하게 수행하므로 자세한 설명은 생략하도록 한다.
- [0070] 에코 신호 추정기(Estimated Echo Signal)는, 원단신호를 DFT하여 얻은 특정 주파수 성분으로부터 음향학적 반향신호를 추정한다. 에코 신호 추정기는, 단계 S400에서의 과정을 동일하게 수행하므로 자세한 설명은 생략하도록 한다.
- [0071] 결합 신호 추정기는, 추정된 음향학적 반향신호와 추정된 잡음신호 전력을 원단신호의 유무에 따라 선별적으로 결합한다. 이때 가중치 파라미터를 사용할 수 있으며, 가중치 파라미터는 단계 S300에서 이미 설명한 바와 같으므로 자세한 설명은 생략하도록 한다. 결합 신호 추정기가 음향학적 반향신호와 잡음신호 전력을 결합하기 때문에, 하나의 제거 이득을 적용하더라도 음향학적 반향과 배경잡음을 모두 제거할 수 있다. 결합 신호 추정기는, 단계 S500에서의 과정을 동일하게 수행하므로 자세한 설명은 생략하도록 한다.
- [0072] 반향 잡음 제거기는, 결합 신호 추정기에 의하여 결합된 결합 신호에 위너필터(Wiener Filter)를 적용시킴으로써, 음향학적 반향과 배경잡음을 제거한다. 위너 필터는 수학적 식 8에 의하여 계산될 수 있으며, 반향 잡음 제거기는, 단계 S600에서의 과정을 동일하게 수행하므로 자세한 설명은 생략하도록 한다.
- [0073] 본 발명의 성능 평가를 위해 다양한 잡음 및 반향 환경에서 객관적인 실험을 수행하였다. 성능 평가는 근단화자신호가 없을 때 원단신호 구간에서 반향 제거 후 여전히 남아 있는 잔여 반향의 제거 정도를 측정하는 ERLE(Echo Return Loss Enhancement), 동시 통화 구간에서의 음성의 보존도를 평가하는 SA(Speech Attenuation), 그리고 스펙트럼 성분 분석을 위한 음성 스펙트로그램(Spectrogram) 테스트를 실시하였다. 우선 음향학적 반향신호의 감쇠 정도와 음성 보존도를 측정하기 위해 수학적 식 12 및 수학적 식 13과 같은 ERLE(t)와 SA의 수치를 비교하였다.

수학식 12

$$ERLE(t) = 10\log_{10}\left[\frac{E\{y^2(t)\}}{E\{e^2(t)\}}\right]$$

[0074]

수학식 13

$$SA = \frac{1}{N} \sum_{t=1}^N 10\log_{10}\left[\frac{E\{s^2(t)\}}{E\{\tilde{s}^2(t)\}}\right]$$

[0075]

[0076] 여기서, N은 이중 발생 구간의 샘플수이고 $\tilde{s}$ 는 출력신호의 E(t)에서의 근단화자신호 성분을 의미한다.

[0077] 테스트 샘플을 위해 7명의 화자로 부터 얻은 8kHz로 샘플링 된 20개의 문장을 수집하고 각 문장을 원단화자와 근단화자신호로 분류하여 합성하였다. 원단화자신호로 분류된 음성은 섞기 전에 반사 경로를 고려한 실제 환경을 모델링하기 위해 임펄스응답 필터를 통과시키고, 입력 마이크로폰으로 들어가는 반향신호는 근단화자신호보다 3.5dB 작게 하였다. 모델링 환경의 장소는 5×4×3m 크기로 설정하였으며, 잡음 환경을 위해서 vehicular와 babble 잡음을 다양한 SNR(signal-to-noise ratio)로 부가하였다. 그리고 본 발명에 적용된 가중치 파라미터와 문턱 값은 성능 평가에 사용된 테스트 샘플에 기반하여 $\gamma_{TH}=1.0$, $\alpha=0.4$, $\alpha_d=0.97$ 로 설정하였다.

[0078] 도 5는 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법 및 종래 기술에 의한 ERLE와 SA 수치를 비교하여 나타낸 도면이다. structure 1은 기존의 선반향-후잡음 제거의 통합 구조이며, structure 2는 선잡음-후반향 제거의 통합 구조이다. 도 5에 도시된 바와 같이, 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법이, 다양한 SNR 환경에서 기존의 기법보다 향상된 ERLE 수치를 보임을 알 수 있다(a). 또한, 음성 보존도에서도 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법이, 기존의 선반향-후잡음 제거 통합구조보다 우수한 성능을 나타내며, 선잡음-후반향 제거 통합구조와는 대체적으로 비슷한 음성보존도를 유지하는 것을 확인할 수 있다(b).

[0079] 도 6은 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법과 종래 기술에 의한 시간에 따른 ERLE 변화를 나타내는 도면이다. 배경 잡음으로서는 white 잡음이 SNR=15dB로 부가되었으며, 음향학적 반향신호가 섞여 있는 입력신호에 대하여 실험하였다. 도 6에 도시된 바와 같이, 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법이 기존의 방법보다 동시통화 구간에서 근단화자신호를 잘 보존하면서 효과적으로 잡음 및 반향신호를 제거함으로써 향상된 ERLE 변화를 나타내고 있음을 알 수 있다.

[0080] 도 7은 도 6의 음성신호에 대한 스펙트럼 분석을 위한 음성 스펙트로그램을 나타내는 도면이다. 이때 (a)는 깨끗한 근단화자신호의 스펙트로그램, (b)는 structure 1에 의한 출력신호 스펙트로그램, (c)는 structure 2에 의한 출력신호 스펙트로그램이며, (d)는 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법에 의한 출력신호 스펙트로그램을 나타낸다. 도 7에 도시된 바와 같이, 기존의 방법에 의한 스펙트로그램 (b), (c)와, 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성 처리 방법에 의한 스펙트로그램 (d)를 반향신호와 잡음이 섞이기 전의 깨끗한 근단화자신호의 스펙트로그램 (a)와 비교하여 볼 때, 본 발명의 일실시예에 따른 결합된 음향학적 반향 및 배경 잡음 전력에 기반한 음성

처리 방법이 기존의 방법보다 효과적으로 잡음 및 잔여 반향신호를 제거하는 것을 확인할 수 있다.

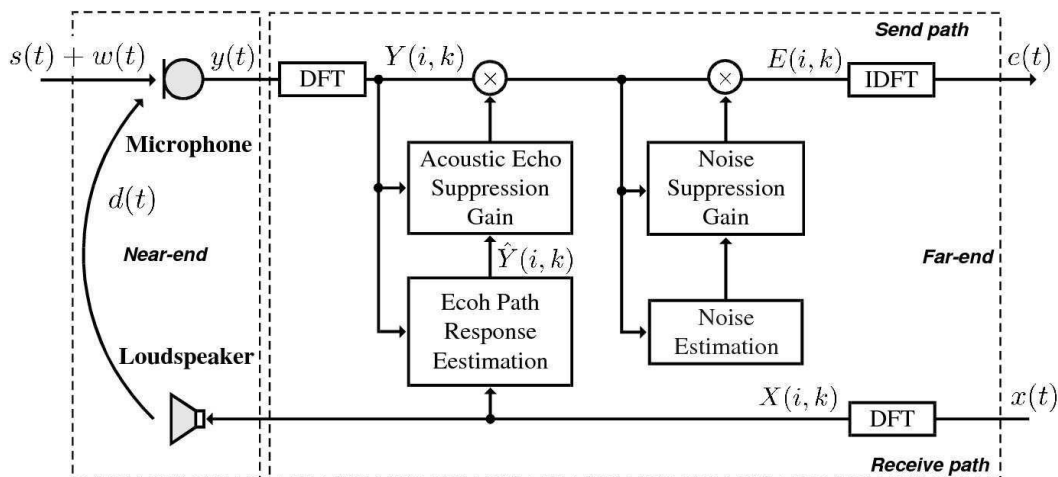
[0081] 이상 설명한 본 발명은 본 발명이 속한 기술분야에서 통상의 지식을 가진 자에 의하여 다양한 변형이나 응용이 가능하며, 본 발명에 따른 기술적 사상의 범위는 아래의 특허청구범위에 의하여 정해져야 할 것이다.

부호의 설명

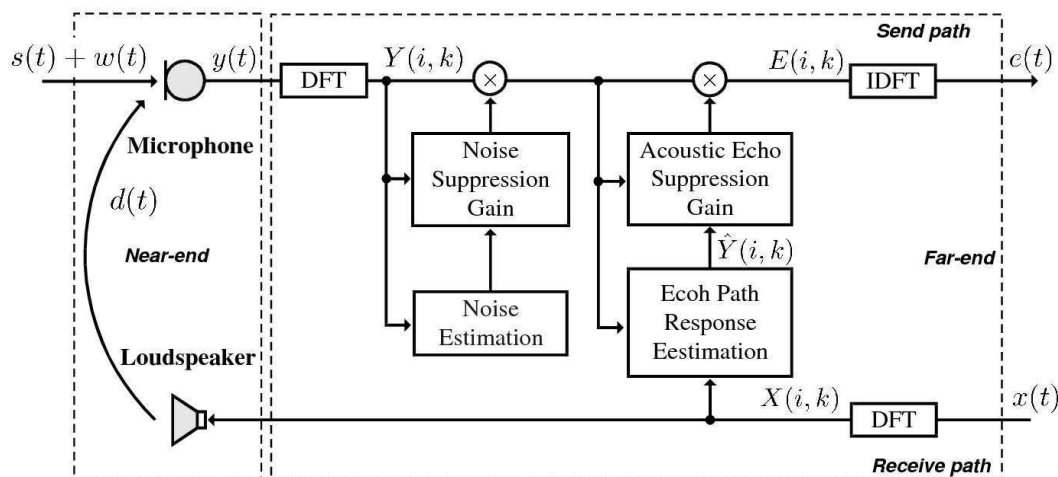
[0082] S100: 마이크로폰에 의하여 입력되는 화자신호를 DFT하고, 스피커에 공급되는 원단신호를 DFT하는 단계
S200: 화자신호와 원단신호로부터 잡음신호의 전력을 추정하는 단계
S300: 추정된 잡음신호의 전력비를 이용하여 가중치 파라미터를 도출하는 단계
S400: 원단신호의 반향경로를 고려하여 음향학적 반향신호를 추정하는 단계
S500: 가중치 파라미터를 사용하여, 추정된 음향학적 반향신호와 추정된 잡음 전력을 원단신호의 유무에 따라 선별적으로 결합하여 최종 결합 신호를 생성하는 단계
S600: 최종 결합 신호에 제거 이득을 적용하여 음향학적 반향 및 잡음신호를 제거하는 단계

도면

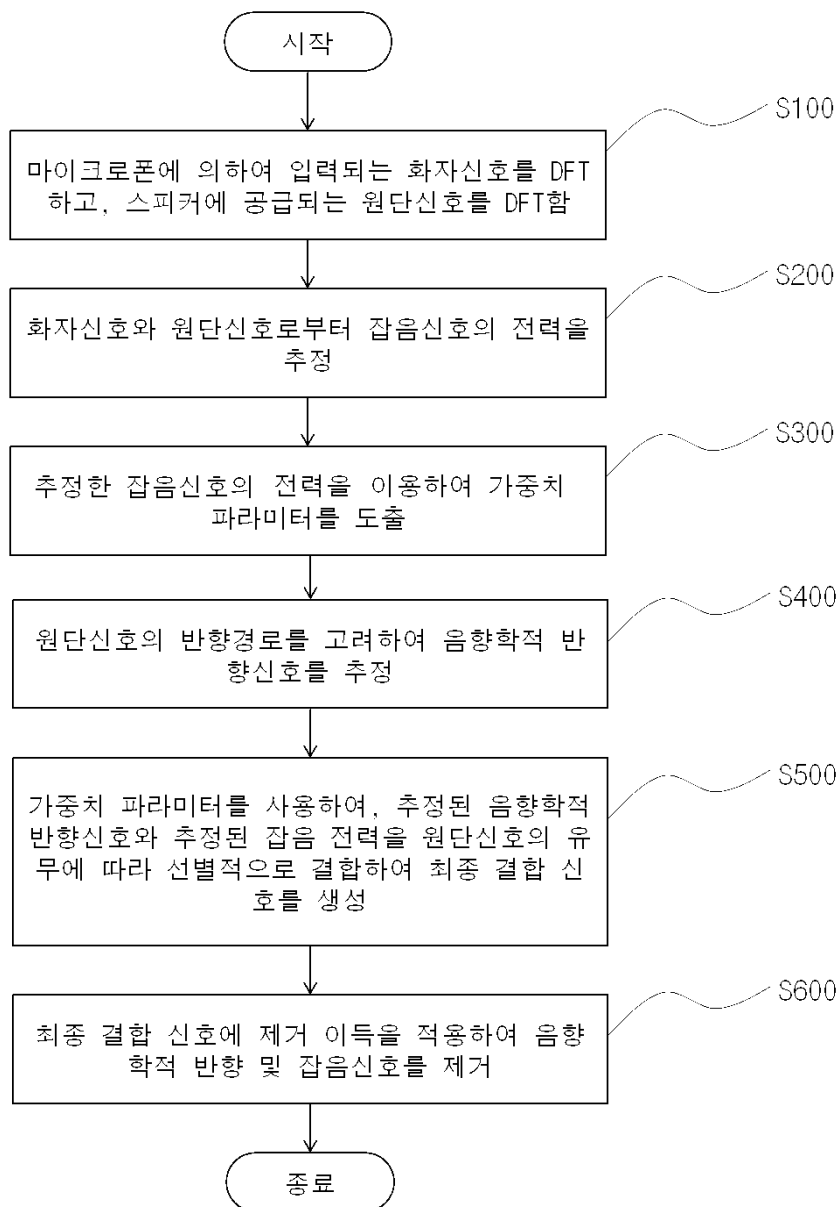
도면1



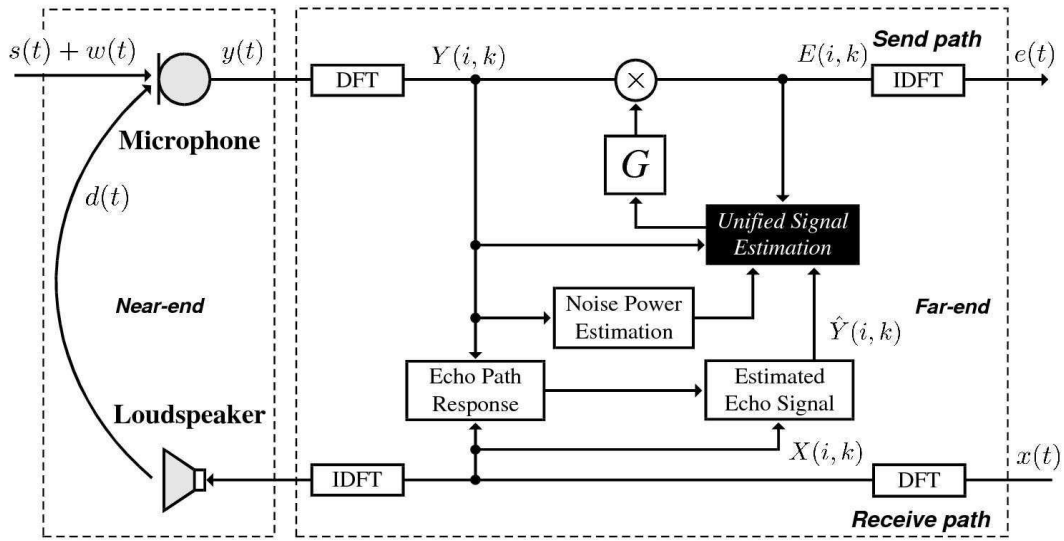
도면2



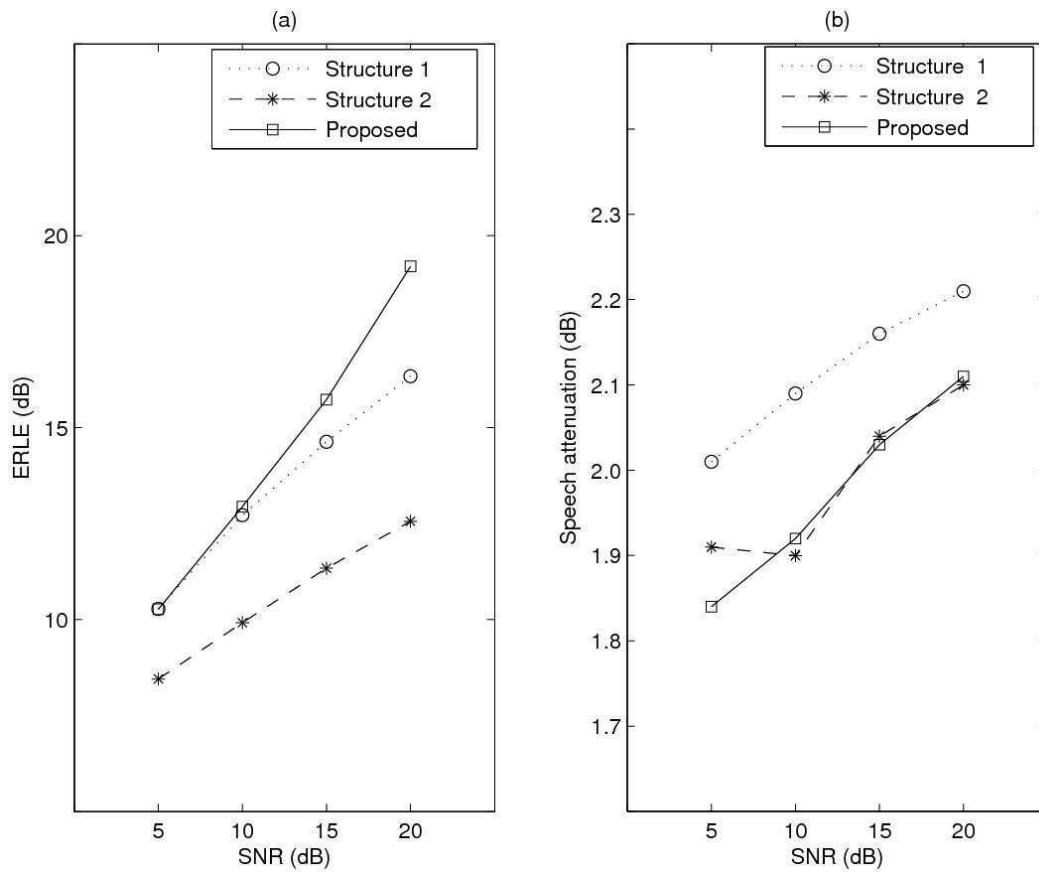
도면3



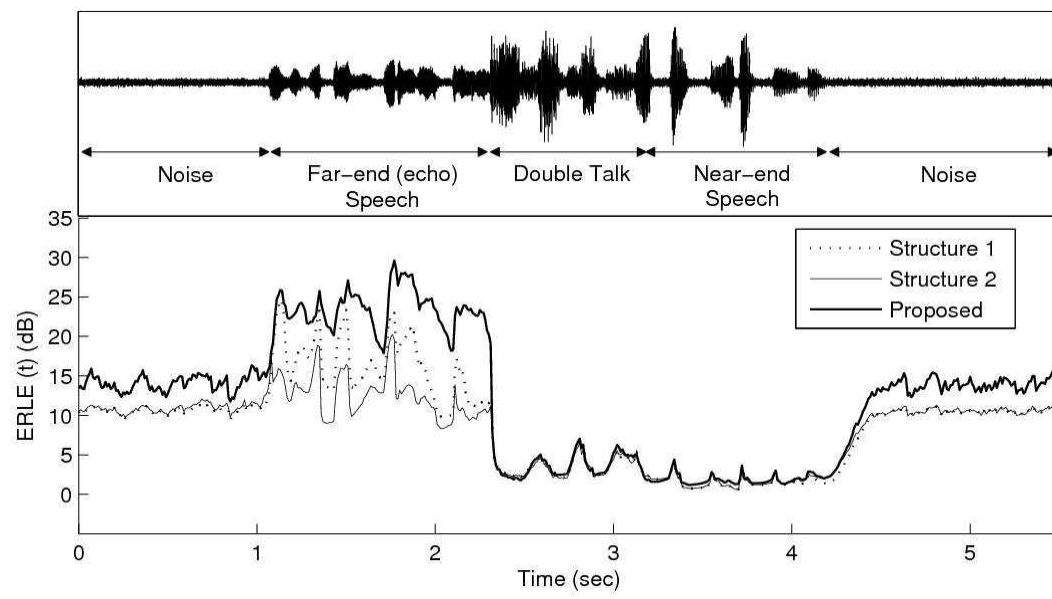
도면4



도면5



도면6



도면7

