



## (19) 대한민국특허청(KR)

## (12) 등록특허공보(B1)

(45) 공고일자 2016년04월21일

(11) 등록번호 10-1614647

(24) 등록일자 2016년04월15일

(51) 국제특허분류(Int. Cl.)

*G10L 21/0208* (2013.01)

(21) 출원번호 10-2014-7035393

(22) 출원일자(국제) 2013년04월01일

심사청구일자 2014년12월17일

(85) 번역문제출일자 2014년12월17일

(65) 공개번호 10-2015-0005719

(43) 공개일자 2015년01월14일

(86) 국제출원번호 PCT/CN2013/073584

(87) 국제공개번호 WO 2013/189199

국제공개일자 2013년12월27일

(30) 우선권주장

201210201879.7 2012년06월18일 중국(CN)

(56) 선행기술조사문헌

E.A.P. Habets, 'Single-channel speech dereverberation based on spectral subtraction', Proceedings of the 15th ProRISC, November 2004.\*

Massimiliano Tonelli, 'Blind reverberation cancellation techniques', A thesis submitted for the degree of Master of Philosophy, The University of Edinburgh, October 2011.\*

JP2003533753 A\*

\*는 심사관에 의하여 인용된 문헌

(73) 특허권자

고어텍 인크

중국 산동 프라빈스, 하이테크 인더스트리 디스트릭 웨이팡 시티, 268 동팡로드

(72) 발명자

로우, 샤샤

중국, 산동 프라빈스 261031, 하이테크 인더스트리 디스트릭 웨이팡 시티, 268 동팡로드

우, 시아오지에

중국, 산동 프라빈스 261031, 하이테크 인더스트리 디스트릭 웨이팡 시티, 268 동팡로드

리, 보

중국, 산동 프라빈스 261031, 하이테크 인더스트리 디스트릭 웨이팡 시티, 268 동팡로드

(74) 대리인

김윤배, 이상복, 강철중

전체 청구항 수 : 총 4 항

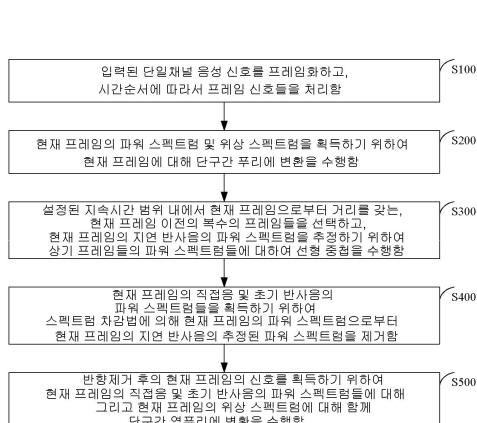
심사관 : 정성윤

(54) 발명의 명칭 단일채널 음성의 반향제거를 위한 방법 및 장치

## (57) 요약

본 발명은 단일채널 음성의 반향제거를 위한 방법 및 장치에 관한 것이다. 본 발명에 따른 방법은, 입력된 단일채널 음성 신호를 프레임화하는 단계, 시간순서에 따라서 프레임 신호들을 처리하는 단계를 포함하되, 상기 처리하는 단계는: 현재 프레임의 파워 스펙트럼 및 위상 스펙트럼을 획득하기 위하여 현재 프레임에 대해 단구간 푸(뒷면에 계속)

대 표 도 - 도1



리에 변환을 수행하는 단계; 설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들을 선택하고, 현재 프레임의 지연 반사음의 파워 스펙트럼을 추정하기 위해서 상기 프레임들의 파워 스펙트럼들에 대하여 선형 중첩을 수행하는 단계; 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들을 획득하기 위해서 스펙트럼 차감법에 의하여 현재 프레임의 파워 스펙트럼으로부터 현재 프레임의 지연 반사음의 추정된 파워 스펙트럼을 제거하는 단계; 및 반향제거 후의 현재 프레임의 신호를 획득하기 위하여 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들에 대하여 그리고 현재 프레임의 위상 스펙트럼에 대하여 함께 단구간 역푸리에 변환을 수행하는 단계;를 포함한다. 반향제거 방법 및 장치는 단일채널 음성의 반향제거에 있어서 반향 환경의 전달 함수의 추정 또는 반향 시간의 추정이 어렵다는 문제점을 해결할 수 있다.

---

## 명세서

### 청구범위

#### 청구항 1

단일채널 음성(single-channel speech)의 반향제거(dereverberation)를 위한 방법으로서,

입력된 단일채널 음성 신호를 프레임화하는(framing) 단계, 및 시간순서에 따라서 프레임 신호들을 처리하는 단계를 포함하되, 상기 처리하는 단계는:

현재 프레임의 파워 스펙트럼(power spectrum) 및 위상 스펙트럼(phase spectrum)을 획득하기 위하여 현재 프레임에 대해 단구간 푸리에 변환(short-time Fourier transform)을 수행하는 단계;

설정된 지속시간 범위(set duration range) 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들을 선택하고, 현재 프레임의 지연 반사음(late reflection sound)의 파워 스펙트럼을 추정하기 위해서 상기 프레임들의 파워 스펙트럼들에 대하여 선형 중첩(linear superposition)을 수행하는 단계;

현재 프레임의 직접음(direct sound) 및 초기 반사음(early reflection sound)의 파워 스펙트럼들을 획득하기 위해서 현재 프레임의 파워 스펙트럼으로부터 현재 프레임의 지연 반사음의 추정된 파워 스펙트럼을 스펙트럼 차감법(spectral subtraction method)에 의하여 제거하는 단계; 및

반향제거 후의 현재 프레임의 신호를 획득하기 위하여 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들에 대하여 그리고 현재 프레임의 위상 스펙트럼에 대하여 함께 단구간 역푸리에 변환(inverse short-time Fourier transform)을 수행하는 단계;를 포함하고,

지속시간 범위의 상한값이 지연 반사음의 감쇠 특성(attenuation characteristics)에 따라서 설정되어, 지속시간 범위의 상한값이 0.3초와 0.5초 사이에서(from 0.3s to 0.5s) 선택되고,

및/또는

지속시간 범위의 하한값이 반향 환경에서의 직접음 및 초기 반사음의 임펄스 응답 분포 영역(shock response distribution areas) 및 음성관련 특성(speech-related characteristics)에 따라서 설정되어, 지속시간 범위의 하한값이 50밀리세컨드와 80밀리세컨드 사이에서(from 50ms to 80ms) 선택되는 것을 특징으로 하는 단일채널 음성의 반향제거를 위한 방법.

#### 청구항 2

삭제

#### 청구항 3

삭제

#### 청구항 4

삭제

#### 청구항 5

청구항 1에 있어서,

현재 프레임의 지연 반사음의 파워 스펙트럼을 추정하기 위해서 상기 프레임들의 파워 스펙트럼들에 대하여 선형 중첩을 수행하는 단계는:

현재 프레임의 지연 반사음의 파워 스펙트럼을 추정하기 위하여 AR(autoregressive) 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 모든 성분들에 대하여 선형 중첩을 수행하는 단계;

또는

현재 프레임의 지연 반사음의 파워 스펙트럼을 추정하기 위하여 MA(Moving Average) 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 직접음 및 초기 반사음 성분들에 대하여 선형 중첩을 수행하는 단계;

또는

AR 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 모든 성분들에 대하여 선형 중첩을 수행한 후에, 현재 프레임의 지연 반사음의 파워 스펙트럼을 추정하기 위하여 MA 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 직접음 및 초기 반사음 성분들에 대하여 선형 중첩을 수행하는 단계;를 포함하는 것을 특징으로 하는 단일채널 음성의 반향제거를 위한 방법.

## 청구항 6

단일채널 음성의 반향제거를 위한 장치로서,

입력된 단일채널 음성 신호를 프레임화하고, 시간순서에 따라서 프레임 신호들을 푸리에 변환 유닛에 출력하도록 구성된 프레임화 유닛(framing unit);

현재 프레임의 파워 스펙트럼 및 위상 스펙트럼을 획득하기 위하여 수신된 현재 프레임에 대해 단구간 푸리에 변환을 수행하고, 현재 프레임의 파워 스펙트럼을 스펙트럼 차감 유닛 및 스펙트럼 추정 유닛에 출력하고, 위상 스펙트럼을 역푸리에 변환 유닛에 출력하도록 구성된 푸리에 변환 유닛(Fourier transform unit);

설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들의 파워 스펙트럼들에 대하여 선형 중첩을 수행하고, 현재 프레임의 지연 반사음의 파워 스펙트럼을 추정하고, 현재 프레임의 지연 반사음의 추정된 파워 스펙트럼을 스펙트럼 차감 유닛에 출력하도록 구성된 스펙트럼 추정 유닛(spectral estimation unit);

현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들을 획득하기 위해서, 스펙트럼 추정 유닛으로부터 획득된 현재 프레임의 지연 반사음의 파워 스펙트럼을 푸리에 변환 유닛으로부터 획득된 현재 프레임의 파워 스펙트럼으로부터 스펙트럼 차감법에 의하여 제거하고, 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들을 역푸리에 변환 유닛에 출력하도록 구성된 스펙트럼 차감 유닛(spectral subtraction unit); 및

스펙트럼 차감 유닛에 의해 획득된 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들에 대하여 그리고 푸리에 변환 유닛에 의해 획득된 현재 프레임의 위상 스펙트럼에 대하여 단구간 역푸리에 변환을 수행하고, 반향제거 후의 현재 프레임의 신호를 출력하도록 구성된 역푸리에 변환 유닛(inverse Fourier transform unit);을 포함하고,

지속시간 범위의 상한값이 지연 반사음의 감쇠 특성에 따라서 설정되어, 지속시간 범위의 상한값이 0.3초와 0.5초 사이에서 선택되고,

및/또는

지속시간 범위의 하한값이 반향 환경에서의 직접음 및 초기 반사음의 임펄스 응답 분포 영역 및 음성관련 특성에 따라서 설정되어, 지속시간 범위의 하한값이 50밀리세컨드와 80밀리세컨드 사이에서 선택되는 것을 특징으로 하는 단일채널 음성의 반향제거를 위한 장치.

## 청구항 7

삭제

## 청구항 8

삭제

## 청구항 9

작제

## 청구항 10

청구항 6에 있어서,

스펙트럼 추정 유닛은:

설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들에 대해서, 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하기 위하여 AR 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 모든 성분들에 대하여 선형 중첩을 수행하도록 구성되거나;

또는

설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들에 대해서, 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하기 위하여 MA 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 직접음 및 초기 반사음 성분들에 대하여 선형 중첩을 수행하도록 구성되거나;

또는

설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들에 대해서, AR 모델을 이용하여 상기 프레임들의 파워 스펙트럼들에서의 모든 성분들에 대하여 선형 중첩을 수행한 후에, 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하기 위하여 MA 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 직접음 및 초기 반사음 성분들에 대하여 선형 중첩을 수행하도록 구성되는 것을 특징으로 하는 단일채널 음성의 반향제거를 위한 장치.

## 발명의 설명

### 기술 분야

[0001]

본 발명은 음성 향상 분야, 특히 단일채널 음성의 반향제거를 위한 방법 및 장치에 관한 것이다.

### 배경 기술

[0002]

컨퍼런트 콜(conference call) 또는 스마트(smart) TV VoIP와 같은 음성 통신에서, 말하는 사람이 마이크로폰으로부터 멀리 떨어져 있고 통화 환경이 상대적으로 폐쇄된 공간일 때, 마이크로폰에 의해 수신된 신호는 그 환경 내에서 반향에 의해 쉽게 간섭받을 수 있다. 예를 들어, 실내(room)에서 음성이 벽, 바닥, 및 가구의 표면에 의해 여러 번 반사될 때, 마이크로폰측에 의해 수신된 신호는 직접음(direct sound) 및 반사음(reflection sound)의 하이브리드 신호(hybrid signal)이다. 반사음의 이러한 부분은 반향 신호에 관한 것이다. 심한(heavy) 반향은 불명확한 음성을 초래해서, 통화의 품질에 영향을 줄 것이다. 게다가, 반향으로부터의 간섭(interference)은 음향 수신 시스템(acoustic receiving system)의 성능을 더 저하시키고, 음성 인식 시스템(speech recognition system)의 성능을 더 저하시킨다.

[0003]

종래의 반향제거 방법들은 통상적으로 디콘볼루션(deconvolution)을 채용한다. 이러한 방법들에 있어서, 반향 환경(실내 또는 사무실 등)의 정확한 임펄스 응답(shock response) 또는 전달 함수를 미리 알 필요가 있다. 반향 환경의 임펄스 응답은 특정한 방법 또는 장치에 의해서 미리 측정되거나, 다른 방법들에 의해서 별도로 측정될 수 있다. 이후, 반향 환경의 공지된 임펄스 응답을 가지고, 인버스 필터가 추정되고, 반향 신호들에 대한 디콘볼루션이 이루어지고, 그래서 반향제거가 이루어진다. 이러한 방법들은, 때때로 반향 환경의 임펄스 응답을 미리 획득하기 어렵고, 인버스 필터 자체를 획득하는 과정이 새로운 불안정한 인자들을 유입시킬 수 있다는 문제점을 가진다.

[0004]

다른 반향제거 방법은, 반향 환경의 임펄스 응답의 추정을 요하지 않아서 인버스 필터의 연산 및 인버스 필터링(inverse filtering)의 수행 양쪽 모두를 요하지 않기 때문에, 블라인드(blind) 반향제거 방법이라고도 불린다. 이러한 방법은 통상적으로 음성 모델 가정(speech model assumption)을 기초로 한다. 예를 들어, 반향은 수신된 음성 여기 펄스(received voiced excitation pulse)에 변화를 초래해서 주기성(periodicity)이 두드러지지 않게 된다. 결과적으로, 음성의 명료도(clarity)가 영향을 받는다. 이러한 방법은 통상적으로, 선형 예측 부호화(linear prediction coding: LPC) 모델을 기초로 하는데, 여기서, 음성 발생 모델이 올-폴 모델(all-pole

model)이고, 반향 또는 다른 추가적 잡음(additive noise)이 전체 시스템에 새로운 포인트(zero point)들을 유입시키고, 음성 여기 펄스가 간섭되며(interfered), 올-폴 필터(all-pole filter)는 영향을 받지 않는 것으로 가정된다. 반향제거 방법은 구체적으로는 다음과 같다: 신호의 LPC 잔여(residual)가 추정된 후, 피치-동기 클러스터링 기준(pitch-synchronous clustering criterion) 또는 첨도 최대화 기준(kurtosis maximization criterion)에 따라서 클린 펄스 여기 신호(clean pulse excitation sequence)가 추정되어서, 반향제거를 실현시킨다. 이러한 방법은, 연산이 통상적으로 매우 복잡하고 올-제로 필터(all-zero)만이 반향에 의해 영향을 받는다는 가정이 때때로 실험적인 분석과 일치하지 않는다는 문제점을 가진다.

[0005] 스펙트럼 차감법(spectral subtraction method)에 의한 반향제거가 바람직한 해결책이다. 음성 신호는 직접음, 초기 반사음(early reflection sound), 및 지연 반사음(late reflection sound)을 포함하기 때문에, 전체 음성의 파워 스펙트럼으로부터 지연 반사음의 파워 스펙트럼(power spectrum)을 제거하는 것은 음성의 품질을 향상시킬 수 있다. 하지만, 핵심은 지연 반사음의 스펙트럼의 추정인데, 즉 음성을 왜곡시키지 않으면서 지연 반사음 성분을 효과적으로 제거하기 위하여 지연 반사음의 상대적으로 정확한 파워 스펙트럼을 어떻게 획득하는가이다. 단일채널 음성 반향제거에서, 이용 가능한 마이크로폰 정보의 하나의 경로만이 존재하기 때문에, 반향 환경의 전달 함수의 추정 또는 반향 시간(RT60)의 추정은 상당히 어렵다.

## 발명의 내용

### 해결하려는 과제

[0006] 본 발명은 반향 환경의 전달 함수의 추정 및 반향 시간의 추정이 상당히 어렵다는 문제점을 해결하기 위하여 단일채널 음성의 반향제거를 위한 방법 및 장치를 제공하는 것을 목적으로 한다.

### 과제의 해결 수단

[0007] 본 발명은 반향 환경의 전달 함수의 추정 및 반향 시간의 추정이 상당히 어렵다는 문제점을 해결하기 위하여 단일채널 음성의 반향제거를 위한 방법 및 장치를 제공한다.

[0008] 본 발명은 단일채널 음성의 반향제거를 위한 방법을 공개하고,

[0009] 본 방법은 입력된 단일채널 음성 신호를 프레임화하는(framing) 단계, 및 시간순서에 따라서 프레임 신호들을 처리하는 단계를 포함하되, 상기 처리하는 단계는:

[0010] 현재 프레임의 파워 스펙트럼(power spectrum) 및 위상 스펙트럼(phase spectrum)을 획득하기 위하여 현재 프레임에 대해 단구간 푸리에 변환(short-time Fourier transform)을 수행하는 단계;

[0011] 설정된 지속시간 범위(set duration range) 내에서 현재 프레임으로부터 거리(distance)를 가지는, 현재 프레임 이전의 복수의 프레임들을 선택하고, 현재 프레임의 지연 반사음(late reflection sound)의 파워 스펙트럼을 추정하기 위해서 상기 프레임들의 파워 스펙트럼들에 대하여 선형 중첩(linear superposition)을 수행하는 단계;

[0012] 현재 프레임의 직접음(direct sound) 및 초기 반사음(early reflection sound)의 파워 스펙트럼들을 획득하기 위해서 현재 프레임의 파워 스펙트럼으로부터 현재 프레임의 지연 반사음의 추정된 파워 스펙트럼을 스펙트럼 차감법(spectral subtraction method)에 의하여 제거하는 단계; 및

[0013] 반향제거 후의 현재 프레임의 신호를 획득하기 위하여 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들에 대하여 그리고 현재 프레임의 위상 스펙트럼에 대하여 함께 단구간 역푸리에 변환(inverse short-time Fourier transform)을 수행하는 단계;를 포함한다.

[0014] 바람직하게는, 지속시간 범위의 상한값이 지연 반사음의 감쇠 특성(attenuation characteristics)에 따라서 설정되고,

[0015] 및/또는

[0016] 지속시간 범위의 하한값이 반향 환경에서의 직접음 및 초기 반사음의 임펄스 응답 분포 영역(shock response distribution areas) 및 음성관련 특성(speech-related characteristics)에 따라서 설정된다.

- [0017] 바람직하게는, 지속시간 범위의 상한값이 0.3초와 0.5초에서(from 0.3s to 0.5s) 선택된다.
- [0018] 바람직하게는, 지속시간 범위의 하한값이 50밀리세컨드와 80밀리세컨드 사이에서(from 50ms to 80ms) 선택된다.
- [0019] 바람직하게는, 현재 프레임의 지연 반사음의 파워 스펙트럼을 추정하기 위해서 상기 프레임들의 파워 스펙트럼들에 대하여 선형 중첩을 수행하는 단계는 구체적으로:
- [0020] 현재 프레임의 지연 반사음의 파워 스펙트럼을 추정하기 위하여 AR(autoregressive) 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 모든 성분들에 대하여 선형 중첩을 수행하는 단계;
- [0021] 또는
- [0022] 현재 프레임의 지연 반사음의 파워 스펙트럼을 추정하기 위하여 MA(Moving Average) 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 직접음 및 초기 반사음 성분들에 대하여 선형 중첩을 수행하는 단계;
- [0023] 또는
- [0024] AR(autoregressive) 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 모든 성분들에 대하여 선형 중첩을 수행한 후에, 현재 프레임의 지연 반사음의 파워 스펙트럼을 추정하기 위하여 MA(moving average) 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 직접음 및 초기 반사음 성분들에 대하여 선형 중첩을 수행하는 단계;를 포함한다.
- [0025] 본 발명은 단일채널 음성의 반향제거를 위한 장치를 더 공개하고, 본 장치는:
- [0026] 입력된 단일채널 음성 신호를 프레임화하고, 시간순서에 따라서 프레임 신호들을 푸리에 변환 유닛에 출력하도록 구성된 프레임화 유닛(framing unit);
- [0027] 현재 프레임의 파워 스펙트럼 및 위상 스펙트럼을 획득하기 위하여 수신된 현재 프레임에 대해 단구간 푸리에 변환을 수행하고, 현재 프레임의 파워 스펙트럼을 스펙트럼 차감 유닛 및 스펙트럼 추정 유닛에 출력하고, 위상 스펙트럼을 역푸리에 변환 유닛에 출력하도록 구성된 푸리에 변환 유닛(Fourier transform unit);
- [0028] 설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들의 파워 스펙트럼들에 대하여 선형 중첩을 수행하고, 현재 프레임의 지연 반사음의 파워 스펙트럼을 추정하고, 현재 프레임의 지연 반사음의 추정된 파워 스펙트럼을 스펙트럼 차감 유닛에 출력하도록 구성된 스펙트럼 추정 유닛(spectral estimation unit);
- [0029] 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들을 획득하기 위해서, 스펙트럼 추정 유닛으로부터 획득된 현재 프레임의 지연 반사음의 파워 스펙트럼을 푸리에 변환 유닛으로부터 획득된 현재 프레임의 파워 스펙트럼으로부터 스펙트럼 차감법에 의하여 제거하고, 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들을 역푸리에 변환 유닛에 출력하도록 구성된 스펙트럼 차감 유닛(spectral subtraction unit); 및
- [0030] 스펙트럼 차감 유닛에 의해 획득된 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들에 대하여 그리고 푸리에 변환 유닛에 의해 획득된 현재 프레임의 위상 스펙트럼에 대하여 단구간 역푸리에 변환을 수행하고, 반향제거 후의 현재 프레임의 신호를 출력하도록 구성된 역푸리에 변환 유닛(inverse Fourier transform unit);을 포함한다.
- [0031] 바람직하게는, 스펙트럼 추정 유닛은 구체적으로 지연 반사음의 감쇠 특성에 따라서 지속시간 범위의 상한값을 설정하도록 구성되고; 및/또는 반향 환경에서의 직접음 및 초기 반사음의 임펄스 응답 분포 영역 및 음성관련 특성에 따라서 지속시간 범위의 하한값을 설정하도록 구성된다.
- [0032] 바람직하게는, 스펙트럼 추정 유닛은 구체적으로 0.3초와 0.5초에서 지속시간 범위의 상한값을 선택하도록 구성된다.
- [0033] 바람직하게는, 스펙트럼 추정 유닛은 구체적으로 50밀리세컨드와 80밀리세컨드 사이에서 지속시간 범위의 하한값을 선택하도록 구성된다.
- [0034] 바람직하게는, 스펙트럼 추정 유닛은 구체적으로:
- [0035] 설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들에 대해서, 현재 프레임의 지연 반사음의 파워 스펙트럼을 추정하기 위하여 AR(autoregressive) 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 모든 성분들에 대하여 선형 중첩을 수행하도록 구성되거나;

[0036]

또는

[0037]

설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들에 대해서, 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하기 위하여 MA(moving average) 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 직접음 및 초기 반사음 성분들에 대하여 선형 중첩을 수행하도록 구성되거나;

[0038]

또는

[0039]

설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들에 대해서, AR(autoregressive) 모델을 이용하여 상기 프레임들의 파워 스펙트럼들에서의 모든 성분들에 대하여 선형 중첩을 수행한 후에, 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하기 위하여 MA(moving average) 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 직접음 및 초기 반사음 성분들에 대하여 선형 중첩을 수행하도록 구성된다.

### 발명의 효과

[0040]

본 발명의 실시예들은 이하의 유익한 효과들을 가진다: 설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들을 선택하고, 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하기 위하여 상기 프레임들의 파워 스펙트럼들에 대하여 선형 중첩을 수행함으로써, 현재 프레임의 자연 반사음의 파워 스펙트럼은 반향 환경의 전달 함수의 추정 또는 반향 시간의 추정에 대한 필요 없이 추정될 수 있고, 반향제거는 스펙트럼 차감법에 의해서 더 실현된다. 반향제거의 연산 복잡성은 단순화되고, 구현은 더 단순해진다.

[0041]

반향 환경에서 직접음 및 초기 반사음의 임펄스 응답 분포 영역 및 음성관련 특성에 따라서 지속시간 범위의 하한값을 설정함으로써, 반향제거 동안 유용한 직접음 및 초기 반사음이 더 잘 보유될(reserved) 수 있다. 음성의 품질이 향상된다.

[0042]

지연 반사음의 감쇠 특성에 따라서 지속시간 범위의 상한값을 설정함으로써, 지연 반사음의 추정된 파워 스펙트럼의 정확성을 보장하면서 중첩 연산(superposition calculation)의 양이 감소된다.

[0043]

본 발명의 실시예들에서, 상한값은 0.3초와 0.5초 사이에서 선택된다. 이 상한값은 실험에 의해서 획득된 역치값(threshold)이다. 반향 환경이 변할 때, 상한값에 대한 조정이 없이도, 더 양호한 반향제거 효과가 여전히 획득될 수 있다.

[0044]

본 발명의 실시예들에서, 하한값은 50밀리세컨드와 80밀리세컨드 사이에서 선택된다. 반향 환경이 변할 때, 하한값에 대한 조정이 없이도, 직접음 및 초기 반사음으로부터 효과적으로 중첩 연산이 수행될 수 있다. 결과적으로, 중첩의 결과들은 실질적으로 아무런 직접음 및 초기 반사음을 포함하지 않는다. 이렇게 해서, 반향제거 동안 유용한 직접음 및 초기 반사음이 더 잘 보유될 수 있다. 더욱 우수한 음성 품질이 획득된다.

[0045]

반향 환경의 변화는: 반향이 없는 무반향실(anechoic room)에서부터 심한(heavy) 반향을 가진 홀까지 포함한다.

### 도면의 간단한 설명

[0046]

도 1은 본 발명에 따른 단일채널 음성의 반향제거를 위한 방법의 흐름도이다;

도 2는 실제 실내에서의 임펄스 응답을 보이는 개략도이다;

도 3은 본 발명의 구현 효과의 개략도이고, 도 3a는 반향 신호의 시간 영역 다이어그램이고, 도 3b는 반향제거 후의 신호의 시간 영역 다이어그램이고, 도 3c는 반향제거 후의 신호 및 반향 신호의 에너지 포락 곡선이다;

도 4는 본 발명에 따른 단일채널 음성의 반향제거를 위한 장치의 구조도이고;

도 5는 본 발명에 따른 단일채널 음성의 반향제거를 위한 장치의 특정 구현 방식의 구조도이다.

### 발명을 실시하기 위한 구체적인 내용

- [0047] 본 발명의 목적, 기술적 해결책들, 및 이점들을 더욱 명확히 하기 위하여, 본 발명의 실시예들은 도면을 참조하여 이하에서 상세하게 더 설명될 것이다.
- [0048] 도 1을 참조하면, 본 발명에 따른 단일채널 음성의 반향제거를 위한 방법의 흐름도가 도시된다.
- [0049] S100: 입력된 단일채널 음성 신호가 프레임화되고(framed), 프레임 신호들이 시간순서(time sequence)에 따라서 다음과 같이 처리된다.
- [0050] S200: 현재 프레임의 파워 스펙트럼 및 위상 스펙트럼을 획득하기 위하여 현재 프레임에 대해 단구간 푸리에 변환이 수행된다.
- [0051] S300: 설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 갖는, 현재 프레임 이전의 복수의 프레임들이 선택되고, 현재 프레임의 지연 반사음의 파워 스펙트럼을 추정하기 위하여 상기 프레임들의 파워 스펙트럼들에 대하여 선형 중첩(linear superposition)이 수행된다.
- [0052] 복수의 프레임들은 미리 설정된 수의 프레임들을 지칭하고, 이것은 지속시간 범위(duration range) 내의 모든 프레임들이거나 지속시간 범위 내의 프레임들의 일부일 수 있다.
- [0053] S400: 현재 프레임의 지연 반사음의 추정된 파워 스펙트럼은 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들을 획득하기 위하여 스펙트럼 차감법에 의해 현재 프레임의 파워 스펙트럼으로부터 제거된다.
- [0054] S500: 반향제거 후의 현재 프레임의 신호를 획득하기 위하여 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들에 대하여 그리고 현재 프레임의 위상 스펙트럼에 대하여 함께 단구간 역푸리에 변환이 수행된다.
- [0055] 반향 환경(reverberation environment)에서, 마이크로폰에 의해 획득된 신호  $x(t)$ , 즉 단일채널 음성 신호는 직접음 및 반사음의 하이브리드 신호이고, 이것은 이하의 반향 모델로서 표현될 수 있다:
- [0056] 
$$x(t) = h * s(t) + n(t)$$
- [0057] 여기서,  $s(t)$ 는 음원으로부터의 신호이고,  $h$ 는 음원의 위치에서부터 마이크로폰의 위치까지의 두 지점을 간의 실내 임펄스 응답(room shock response)이고,  $*$ 은 콘볼루션 연산(convolution operation)이고,  $n(t)$ 는 반향 환경에서의 다른 추가적 잡음(additive noise)이다.
- [0058] 실제 실내에서의 임펄스 응답은 도 2에서 도시된 바와 같다. 임펄스 응답은 세 부분으로 나눌 수 있는데, 즉, 직접 피크(direct peak)  $hd$ , 초기 반사(early reflection)  $he$ , 및 지연 반사(late reflection)  $hl$ 로 나눌 수 있다.  $hd$  와  $s(t)$ 의 콘볼루션은 일정한 시간 지연 후에 마이크로폰측 상에서 음원으로부터의 신호의 재출현(reappearance)으로서 단순히 고려될 수 있고,  $x(t)$ 에서 직접음 부분에 상응한다. 초기 반사 부분의 임펄스 응답은  $hd$ 에 이어지는 어느 정도의 지속시간의 부분에 상응하고, 지속시간의 종료시점은 50ms와 80ms 사이의 임의의 시점이다. 이 부분과  $s(t)$ 의 콘볼루션에 의해 생성된 초기 반사음이 직접음의 품질을 높이고 향상시킬 수 있다고 일반적으로 고려된다. 지연 반사음 부분의 임펄스 응답은  $hd$  및  $he$ 의 제거 후 실내 임펄스 응답의 나머지 긴 트레일링(trailing) 부분이다. 이 부분과 신호  $s(t)$ 의 콘볼루션에 의해 생성된 반사음은 듣는 효과에 영향을 줄 수 있는 반향 성분이다. 반향제거 알고리즘은 주로 이 부분의 영향을 제거하는 것이다.
- [0059] 그러므로, 반향 모델은 또한 다음과 같이 표현될 수 있다:
- [0060] 
$$x(t) = (hd + he) * s(t) + hl * s(t) + n(t)$$
- [0061]  $hl$  부분은 지수 감쇠(exponential attenuation) 모델과 일치하고, 근사적으로는 다음의 방정식과 일치한다:
- [0062] 
$$hl(t) = b(t)e^{-\frac{3\ln 10}{T_r}t}$$
- [0063] 여기서,  $T_r$ 은 반향 환경의 반향 시간(RT60)이고,  $b(t)$ 는 제로-평균 가우시안 분포 랜덤 변수(zero-mean Gaussian distribution random variable)이다.

- [0064] 지연 반사음의 파워 스펙트럼을 어떻게 추정하는지가 이하에서 상세하게 설명될 것이다.
- [0065] 파워 스펙트럼의 분석으로부터, 신호의 파워 스펙트럼  $X(t,f)$  이 다음과 같이 표현될 수 있다:
- $$X(t,f) = Y(t,f) + R(t,f)$$
- [0066] 여기서,  $R(t,f)$  는 지연 반사음의 파워 스펙트럼이고,  $Y(t,f)$  는 보유될(reserved) 수 있는 직접음 및 초기 반사음의 파워 스펙트럼들이다. 지연 반사음의 파워 스펙트럼  $R(t,f)$  이 추정된 후에,  $Y(t,f)$  가 스펙트럼 차감법에 의해  $X(t,f)$ 로부터 추정될 수 있어서, 반향제거가 실현될 수 있다.
- [0068] 반향 발생 모델의 분석에 따라서, 지연 반사음의 파워 스펙트럼은 지연 반사음 이전의 신호의 파워 스펙트럼과 선형 관계(linear relationship)를 가지거나, 지연 반사음 이전의 신호의 파워 스펙트럼에서의 일부 성분들을 가질 수 있다. 인간의 음성 특성으로 인하여, 직접음 및 초기 반사음의 파워 스펙트럼들은 직접음 및 초기 반사음 이전의 신호의 파워 스펙트럼과 아무런 선형 관계를 가지지 않거나, 직접음 및 초기 반사음 이전의 신호의 파워 스펙트럼에서의 일부 성분들을 가진다. 그러므로, 설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 갖는, 현재 프레임 이전의 프레임들의 파워 스펙트럼들에서의 성분들에 대하여 선형 중첩을 수행함으로써, 현재 프레임의 지연 반사음의 파워 스펙트럼이 추정될 수 있다. 이후, 스펙트럼 차감법에 의해서 현재 프레임의 파워 스펙트럼으로부터 지연 반사음의 파워 스펙트럼을 제거함으로써, 단일채널 음성의 반향제거가 실현될 수 있다.
- [0069] 바람직하게는, 지속시간 범위의 상한값은 지연 반사음의 감쇠 특성에 따라서 설정된다.
- [0070] 만일 스펙트럼 추정을 위해 더 많은 프레임들이 이용된다면, 추정은 더 정확해질 것이다. 그러나, 너무 많은 프레임들은 연산량의 증가를 초래할 것이다. 도 2 및  $hl$  부분의 지수 감쇠 모델로부터, 현재 프레임으로부터의 거리가 더 클수록 반사음의 에너지가 더 작고, 반사음의 에너지는 어느 순간 이후에 무시될(ignored) 수 있다는 것을 알 수 있다. 그러므로, 반사음의 에너지가 무시될 수 있는 순간이 지연 반사음의 감쇠 특성을 따라서 획득되고, 상한값은 이 순간에서부터 현재 프레임의 순간까지의 지속시간으로 설정된다. 이렇게 해서, 지연 반사음의 추정된 파워 스펙트럼의 정확성을 보장하면서 중첩 연산들의 양이 감소될 수 있다.
- [0071] 바람직하게는, 지속시간 범위의 하한값은 반향 환경에서의 직접음 및 초기 반사음의 임펄스 응답 분포 영역과 음성관련 특성에 따라서 설정된다.
- [0072] 도 2로부터, 직접음 및 초기 반사음 양쪽 모두의 에너지가 현재 프레임에 더 가까운 시간에 집중된다는 것을 알 수 있다. 반향 환경에서의 직접음 및 초기 반사음의 임펄스 응답 분포 영역에 따라서 지속시간 범위의 하한값을 설정함으로써, 직접음 및 초기 반사음의 에너지가 집중되는 시간 기간을 피해서 선형 중첩이 실행될 수 있고, 반향제거 동안 유용한 직접음 및 초기 반사음은 더 잘 보유될 수 있다. 음성의 품질이 향상된다.
- [0073] 바람직하게는, 지속시간 범위의 하한값이 50ms와 80ms 사이에서 선택된다.
- [0074] 다양한 환경에서, 하한값이 50ms와 80ms 사이의 범위에 있는 한, 지연 반사음의 유효한 파워 스펙트럼은 직접음 및 초기 반사음 부분들을 충분히 포함으로써 더 잘 추정될 수 있다는 점이 실험에 의해서 밝혀졌다. 환경이 변할 때, 하한값에 대한 조정이 없이도, 더 좋은 음성 품질이 획득될 수 있다.
- [0075] 바람직하게는, 지속시간 범위의 상한값이 0.3s과 0.5s 사이에서 선택된다.
- [0076] 이론상, 상한값의 세업(setup)은 이 방법을 적용하는 특정한 환경과 관련이 있다. 본 발명과 관련이 있는 지연 반사음의 파워 스펙트럼의 추정에 있어서, 상한값은 이론상 실내 임펄스 응답의 길이에 상응한다. 하지만, 지수 모델(exponential model)에 따라서 감쇠하는 실제 환경에서의 임펄스 응답의  $hl$  부분과 반향 발생 모델의 조합에서, 현재 순간으로부터의 거리가 더 클수록, 반사음의 에너지가 더 작고, 반사음의 에너지는 0.5s를 넘어서는 무시될 수 있다. 그러므로, 실제로, 대략적인(rough) 상한값은 대부분의 반향 환경에 적합할 수 있다. 0.3s과 0.5s 사이의 범위에 있을 때, 상한값은 무반향실 환경(반향 시간: 매우 짧음), 일반적인 사무실 환경(반향 시간: 0.3-0.5s), 또는 심지어 홀(hall)(반향 시간: >1s)과 같은 다양한 반향 환경에서 상당히 적합할 수 있다는 것이 증명되었다. 무반향실 환경에서, 거의 아무런 지연 반사음이 존재하지 않는다. 본 발명에 의해서 제공되는 방법에서, 선형 성분들만이 추정되고 직접음 및 집중된 초기 반사음을 가진 기간이 회피되기 때문에, 상한

값이 무반향설의 반향 시간보다 훨씬 더 길더라도 유효한 음성 성분들이 제거되지 않을 것이다. 홀 환경에 있는 동안, 상한값이 실제 반향 시간보다 더 작을 수 있을지도, 반향제거는 잘 실현될 수 있다. 이것은 임펄스 응답이 빠르게 지수적으로 감쇠될 때 앞선 0.3s에서의 자연 반사음 성분들이 전체 자연 반사음 성분들의 에너지의 대부분을 점유하기 때문이다.

[0077] 특정 구현 방식에서, 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하기 위하여 상기 프레임들의 파워 스펙트럼들에 대하여 선형 중첩을 수행하는 것은 구체적으로: 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하기 위하여 AR(autoregressive) 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 모든 성분에 대하여 선형 중첩을 수행하는 것을 포함한다.

[0078] 예를 들어, 현재 프레임의 자연 반사음의 파워 스펙트럼은 다음 방정식에 따라서 AR 모델을 이용하여 추정된다:

$$R(t, f) = \sum_{j=J_0}^{J_{AR}} \alpha_{j,f} \cdot X(t - j \cdot \Delta t, f)$$

[0079] 여기서,  $R(t, f)$ 는 자연 반사음의 추정된 파워 스펙트럼이고,  $J_0$ 는 설정된 지속시간 범위의 하한값으로부터 획득된 시작 순번이고,  $J_{AR}$ 는 설정된 지속시간 범위의 상한값으로부터 획득된 AR 모델의 차수이고,  $\alpha_{j,f}$ 는 AR 모델의 추정 파라미터이고,  $X(t - j \cdot \Delta t, f)$ 는 현재 프레임 이전의  $j$  프레임의 파워 스펙트럼이고,  $\Delta t$ 는 프레임들 간의 간격이다.

[0080] 특정 구현 방식에서, 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하기 위하여 상기 프레임들의 파워 스펙트럼들에 대하여 선형 중첩을 수행하는 것은 구체적으로: 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하기 위하여 MA(Moving Average) 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 직접음 및 초기 반사음 성분들에 대하여 선형 중첩을 수행하는 것을 포함한다.

[0081] 예를 들어, 현재 프레임의 자연 반사음의 파워 스펙트럼은 다음 방정식에 따라서 MA 모델을 이용하여 추정된다:

$$R(t, f) = \sum_{j=J_0}^{J_{MA}} \beta_{j,f} \cdot Y(t - j \cdot \Delta t, f)$$

[0082] 여기서,  $R(t, f)$ 는 자연 반사음의 추정된 파워 스펙트럼이고,  $J_0$ 는 설정된 지속시간 범위의 하한값으로부터 획득된 시작 순번이고,  $J_{MA}$ 는 설정된 지속시간 범위의 상한값으로부터 획득된 MA 모델의 차수이고,  $\beta_{j,f}$ 는 MA 모델의 추정 파라미터이고,  $Y(t - j \cdot \Delta t, f)$ 는 현재 프레임 이전의  $j$  프레임의 직접음 및 초기 반사음의 파워 스펙트럼들이고,  $\Delta t$ 는 프레임들 간의 간격이다.

[0083] 특정 구현 방식에서, 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하기 위하여 상기 프레임들의 파워 스펙트럼들에 대하여 선형 중첩을 수행하는 것은 구체적으로: AR 모델을 이용하여 상기 프레임들의 파워 스펙트럼들에서의 모든 성분들에 대하여 선형 중첩을 수행한 후에, 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하기 위하여 MA 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 직접음 및 초기 반사음 성분들에 대하여 선형 중첩을 수행하는 것을 포함한다.

[0084] 예를 들어, 현재 프레임의 자연 반사음의 파워 스펙트럼은 다음 방정식에 따라서 ARMA 모델을 이용하여 추정된다:

$$R(t, f) = \sum_{j=J_0}^{J_{AR}} \alpha_{j,f} \cdot X(t - j \cdot \Delta t, f) + \sum_{j=J_0}^{J_{MA}} \beta_{j,f} \cdot Y(t - j \cdot \Delta t, f)$$

[0085] 여기서,  $R(t, f)$ 는 자연 반사음의 추정된 파워 스펙트럼이고,  $J_0$ 는 설정된 지속시간 범위의 하한값으로부터 획득된 시작 순번이고,  $J_{AR}$ 는 설정된 지속시간 범위의 상한값으로부터 획득된 AR 모델의 차수이고,  $\alpha_{j,f}$ 는 AR 모델

의 추정 파라미터이고,  $J_{MA}$  는 설정된 지속시간 범위의 상한값으로부터 획득된 MA 모델의 차수이고,  $\beta_{j,f}$ 는 MA 모델의 추정 파라미터이고,  $Y(t-j \cdot \Delta t, f)$ 는 현재 프레임 이전의 j 프레임의 직접음 및 초기 반사음의 파워 스펙트럼이고,  $X(t-j \cdot \Delta t, f)$ 는 현재 프레임 이전의 j 프레임의 파워 스펙트럼이고,  $\Delta t$ 는 프레임들 간의 간격이다.

[0089] AR 모델, MA 모델, 및 ARMA 모델의 구체적인 해들을 위한 잘 알려져 있는 알고리즘들, 예를 들어, 윌-워커 방정식(Yule-Walker equations) 또는 버그 알고리즘(Burg algorithm)이 존재한다.

[0090] 스펙트럼 차감법에 의한 반향제거의 핵심은 자연 반사음의 파워 스펙트럼의 추정이다. 선행기술에서 언급된 자연 반사음의 파워 스펙트럼의 추정은 통상적으로 상술한 AR 또는 MA 또는 ARMA 모델의 일정한 구체적인 예이다. 게다가, 자연 반사음의 파워 스펙트럼의 추정의 다른 방법들은 통상적으로 음성 단속 단계(speech intermittent stage)의 반향 환경에서 반향 시간(RT60)의 추정을 요하며, 이것은 자연 반사음의 파워 스펙트럼의 추정에서 중요한 파라미터로서 취급된다. 본 발명에서는 다양한 환경들에서의 반향 시간의 추정 또는 임펄스 응답의 추정을 요하지 않고, 본 방법은 반향 환경에서 말하고 있는 사람의 움직임으로 인하여 반향 임펄스 응답 또는 반향 시간이 변하는 다양한 상이한 반향 환경들 및 경우들에 대한 적합할 수 있다.

[0091] 특정 구현 방식에서, 스펙트럼 차감법에 의해서 프레임의 파워 스펙트럼으로부터 반향 성분들을 제거하는 것은 구체적으로:

[0092] 자연 반사음의 파워 스펙트럼에 따라서 스펙트럼 차감법에 의하여 이득 함수(gain function)를 획득하는 것; 및

[0093] 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들을 획득하기 위하여 현재 프레임의 파워 스펙트럼을 이득 함수에 곱하는 것;을 포함한다.

[0094] 자연 반사음의 파워 스펙트럼  $R(t, f)$ 의 추정을 끝낸 후에, 반향제거 후의 음성 신호  $Y(t, f)$ 가 스펙트럼 차감법에 의해 획득될 수 있다:

$$Y(t, f) = G(t, f) \cdot X(t, f)$$

$$G(t, f) = \frac{X(t, f) - R(t, f)}{X(t, f)}$$

[0096] 여기서,  $G(t, f)$ 는 스펙트럼 차감법에 의해 획득되는 이득 함수이다.

[0097] 본 발명의 구현 효과가 도 3에서 도시된다. 반향 신호(단일채널 음성 신호)가 회의실(conference room)로부터 획득되고, 음원에서부터 마이크로폰까지의 거리가 2m이고, 반향 시간(RT60)은 약 0.45s이다. 자연 반사음의 파워 스펙트럼은 본 발명에서 제시된 AR 모델에 따라서 추정되고, 하한값은 80ms로서 설정되고, 상한값은 0.5s로서 설정된다. 도시된 바와 같이, 본 발명에 의해 제공되는 방법을 이용한 반향제거 후에, 반향 트레일링(reverberation trailing)은 눈에 띄게 감소하고, 음성의 품질은 현저하게 향상된다.

[0098] 도 4에서 도시된 바와 같이, 단일채널 음성의 반향제거를 위한 장치는 이하의 유닛들을 포함한다:

[0099] 입력된 단일채널 음성 신호를 프레임화하고(frame), 시간순서에 따라서 프레임 신호들을 푸리에 변환 유닛(200)에 출력하도록 구성된 프레임화 유닛(100);

[0100] 현재 프레임의 파워 스펙트럼 및 위상 스펙트럼을 획득하기 위하여 수신된 현재 프레임에 대해 단구간 푸리에 변환을 수행하고, 현재 프레임의 파워 스펙트럼을 스펙트럼 차감 유닛(400) 및 스펙트럼 추정 유닛(300)에 출력하고, 위상 스펙트럼을 역푸리에 변환 유닛(500)에 출력하도록 구성된 푸리에 변환 유닛(200);

[0101] 설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들의 파워 스펙트럼들에 대하여 선형 중첩을 수행하고, 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하고, 현재 프레임의 자연 반사음의 추정된 파워 스펙트럼을 스펙트럼 차감 유닛(400)에 출력하도록 구성된 스펙트럼 추정 유닛(300);

[0102] 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들을 획득하기 위하여, 스펙트럼 추정 유닛(300)으로부터 획득된 현재 프레임의 자연 반사음의 파워 스펙트럼을 푸리에 변환 유닛(200)으로부터 획득된 현재 프레임의 파워 스펙트럼으로부터 스펙트럼 차감법에 의해 제거하고, 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼

들을 역푸리에 변환 유닛(500)에 출력하도록 구성된 스펙트럼 차감 유닛(400); 및

[0103] 스펙트럼 차감 유닛(400)에 의해 획득된 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들에 대하여 그리고 푸리에 변환 유닛(200)에 의해 획득된 현재 프레임의 위상 스펙트럼에 대하여 단구간 역푸리에 변환을 수행하고, 반향제거 후의 현재 프레임의 신호를 출력하도록 구성된 역푸리에 변환 유닛(500).

[0104] 바람직하게는, 스펙트럼 추정 유닛(300)은 구체적으로 자연 반사음의 감쇠 특성에 따라서 지속시간 범위의 상한값을 설정하도록 구성된다.

[0105] 바람직하게는, 스펙트럼 추정 유닛(300)은 구체적으로 반향 환경에서의 직접음 및 초기 반사음의 임펄스 응답 분포 영역 및 음성관련 특성에 따라서 지속시간 범위의 하한값을 설정하도록 구성된다.

[0106] 바람직하게는, 스펙트럼 추정 유닛(300)은 구체적으로 0.3s과 0.5s 사이의 지속시간 범위의 상한값을 선택하도록 구성된다.

[0107] 바람직하게는, 스펙트럼 추정 유닛(300)은 구체적으로 50ms과 80ms 사이의 지속시간 범위의 하한값을 선택하도록 구성된다.

[0108] 특정 구현 방식에서의 장치가 도 5에서 도시된다. 스펙트럼 추정 유닛(300)은 구체적으로: 설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들에 대해서, 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하기 위하여 AR 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 모든 성분들에 대하여 선형 중첩을 수행하도록 구성된다.

[0109] 예를 들어, 현재 프레임의 자연 반사음의 파워 스펙트럼은 이하의 방정식에 따라서 AR 모델을 이용해서 추정된다:

$$R(t, f) = \sum_{j=J_0}^{J_{AR}} \alpha_{j,f} \cdot X(t - j \cdot \Delta t, f)$$

[0110] 여기서,  $R(t, f)$  는 자연 반사음의 추정된 파워 스펙트럼이고,  $J_0$  는 설정된 지속시간 범위의 하한값으로부터 획득된 시작 순번이고,  $J_{AR}$  는 지속시간 범위의 상한값으로부터 획득된 AR 모델의 차수이고,  $\alpha_{j,f}$  는 AR 모델의 추정 파라미터이고,  $X(t - j \cdot \Delta t, f)$  는 현재 프레임 이전의  $j$  프레임의 파워 스펙트럼이고,  $\Delta t$  는 프레임들 간의 간격이다.

[0112] 다른 특정 구현 방식에서, 스펙트럼 추정 유닛(300)은 구체적으로: 설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들에 대해서, 현재 프레임의 자연 반사음의 파워 스펙트럼을 추정하기 위하여 MA 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 직접음 및 초기 반사음 성분들에 대하여 선형 중첩을 수행하도록 구성된다.

[0113] 예를 들어, 현재 프레임의 자연 반사음의 파워 스펙트럼은 이하의 방정식에 따라서 MA 모델을 이용하여 추정된다:

$$R(t, f) = \sum_{j=J_0}^{J_{MA}} \beta_{j,f} \cdot Y(t - j \cdot \Delta t, f)$$

[0114] 여기서,  $R(t, f)$  는 자연 반사음의 추정된 파워 스펙트럼이고,  $J_0$  는 설정된 지속시간 범위의 하한값으로부터 획득된 시작 순번이고,  $J_{MA}$  는 설정된 지속시간 범위의 상한값으로부터 획득된 MA 모델의 차수이고,  $\beta_{j,f}$  는 MA 모델의 추정 파라미터이고,  $Y(t - j \cdot \Delta t, f)$  는 현재 프레임 이전의  $j$  프레임의 직접음 및 초기 반사음의 파워 스펙트럼들이고,  $\Delta t$  는 프레임들 간의 간격이다.

[0116] 다른 특정 구현 방식에서, 스펙트럼 추정 유닛(300)은 구체적으로: 설정된 지속시간 범위 내에서 현재 프레임으로부터 거리를 가지는, 현재 프레임 이전의 복수의 프레임들에 대해서, AR 모델을 이용하여 상기 프레임들의 파워 스펙트럼들에서의 모든 성분들에 대하여 선형 중첩을 수행한 후에, 현재 프레임의 자연 반사음의 파워 스펙

트럼을 추정하기 위하여 MA 모델을 이용해서 상기 프레임들의 파워 스펙트럼들에서의 직접음 및 초기 반사음 성분들에 대하여 선형 중첩을 수행하도록 구성된다.

[0117] 예를 들어, 현재 프레임의 지연 반사음의 파워 스펙트럼은 이하의 방정식에 따라서 ARMA 모델을 이용하여 추정된다:

$$R(t, f) = \sum_{j=J_0}^{J_{AR}} \alpha_{j,f} \cdot X(t - j \cdot \Delta t, f) + \sum_{j=J_0}^{J_{MA}} \beta_{j,f} \cdot Y(t - j \cdot \Delta t, f)$$

[0118]

[0119] 여기서,  $R(t, f)$ 는 지연 반사음의 추정된 파워 스펙트럼이고,  $J_0$ 는 설정된 지속시간 범위의 하한값으로부터 획득된 시작 순번이고,  $J_{AR}$ 는 설정된 지속시간 범위의 상한값으로부터 획득된 AR 모델의 차수이고,  $\alpha_{j,f}$ 는 AR 모델의 추정 파라미터이고,  $J_{MA}$ 는 설정된 지속시간 범위의 상한값으로부터 획득된 MA 모델의 차수이고,  $\beta_{j,f}$ 는 MA 모델의 추정 파라미터이고,  $X(t - j \cdot \Delta t, f)$ 는 현재 프레임 이전의  $j$  프레임의 직접음 및 초기 반사음의 파워 스펙트럼들이고,  $Y(t - j \cdot \Delta t, f)$ 는 현재 프레임 이전의  $j$  프레임의 파워 스펙트럼이고,  $\Delta t$ 는 프레임들 간의 간격이다.

[0120] AR 모델, MA 모델, 및 ARMA 모델의 구체적인 해들을 위한 잘 알려져 있는 알고리즘들, 예를 들어, 윌-워커 방정식 또는 버그 알고리즘이 존재한다.

[0121] 스펙트럼 차감 유닛(400)은 구체적으로: 지연 반사음의 파워 스펙트럼에 따라서 스펙트럼 차감법에 의해 이득함수를 획득하고; 현재 프레임의 직접음 및 초기 반사음의 파워 스펙트럼들을 획득하기 위하여 현재 프레임의 파워 스펙트럼을 이득 함수에 곱하도록 구성된다.

[0122] 지연 반사음의 파워 스펙트럼  $R(t, f)$ 의 추정을 끝낸 후에, 반향제거 후의 음성 신호  $Y(t, f)$ 가 스펙트럼 차감법에 의해 획득될 수 있다:

$$Y(t, f) = G(t, f) \cdot X(t, f)$$

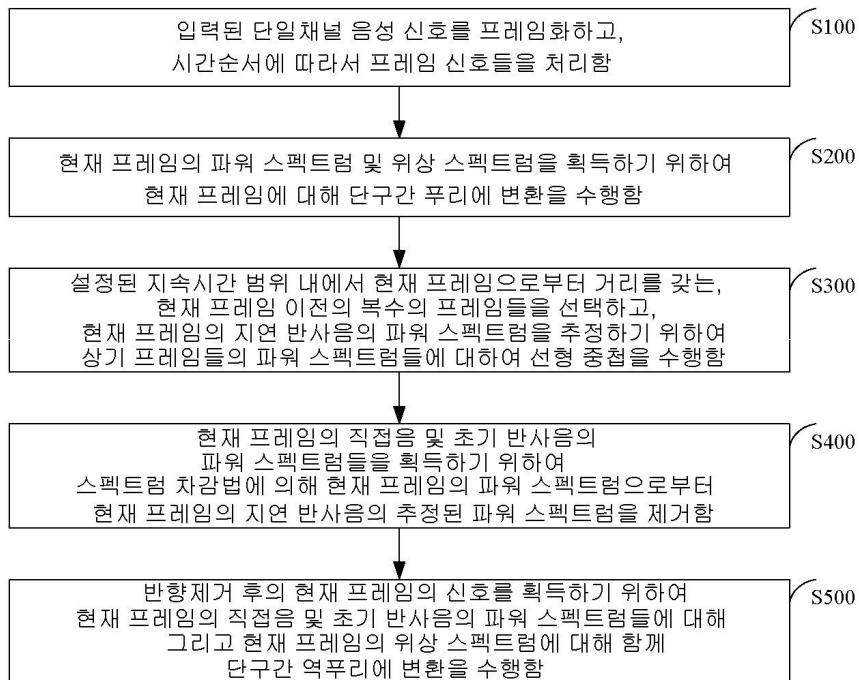
$$G(t, f) = \frac{X(t, f) - R(t, f)}{X(t, f)}$$

[0123] 여기서,  $G(t, f)$ 는 스펙트럼 차감법에 의해 획득된 이득 함수이다.

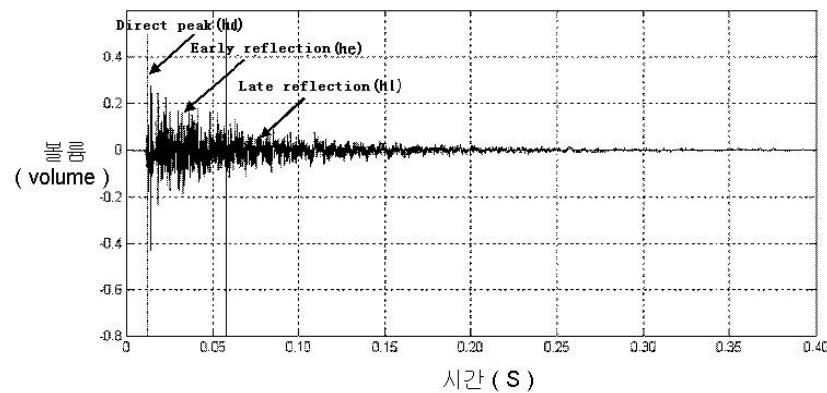
[0125] 상술한 설명은 단지 본 발명의 바람직한 실시예들을 설명하는 것이고, 본 발명의 보호 범위를 제한하려고 의도된 것이 아니다. 본 발명의 사상 및 원리 내에서 만들어진 임의의 변형(modification), 등가적 교체(equivalent replacement), 및 개선(improvement)은 본 발명의 보호 범위 내에 속할 것이다.

## 도면

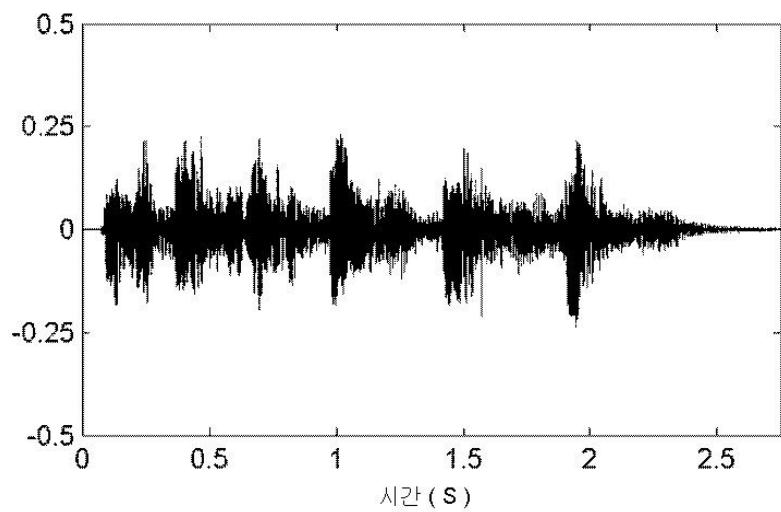
### 도면1



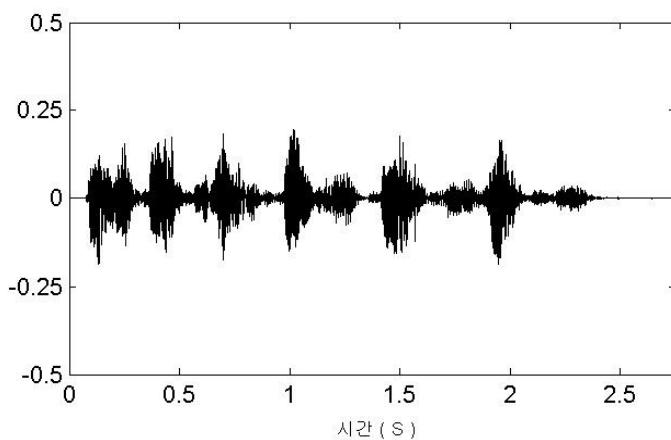
### 도면2



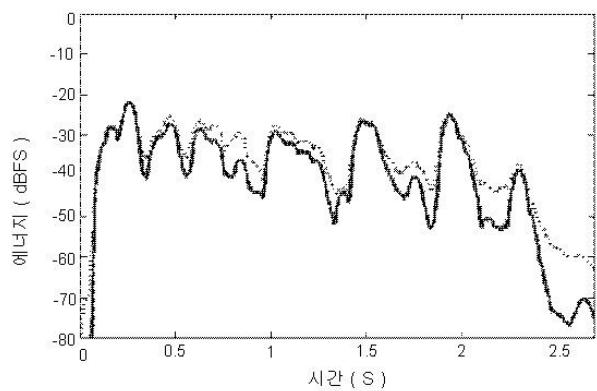
도면3a



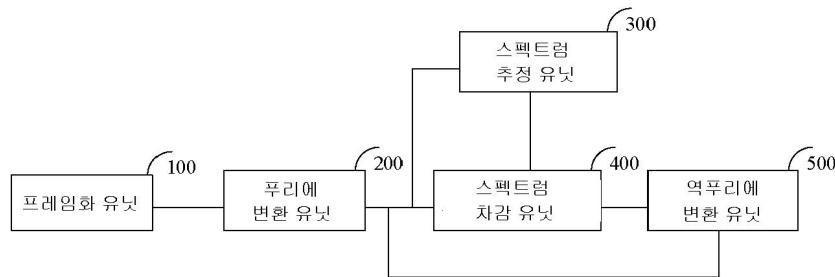
도면3b



도면3c



도면4



도면5

