(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2013년12월02일
(11) 등록번호 10-1334991
(24) 등록일자 2013년11월25일

(51) 국제특허분류(Int. Cl.)
G10L 21/02 (2006.01) *G10L 15/28* (2006.01)
(21) 출원번호 10-2012-0067873
(22) 출원일자 2012년06월25일
심사청구일자 2012년06월25일
(56) 선행기술조사문헌
KR100647826 B1*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
박형민
경기도 화성시 반송동 나루마을 반도아파트 638동 902호
정일영
경기도 용인시 수지구 고기동 244번지
(74) 대리인
이지연

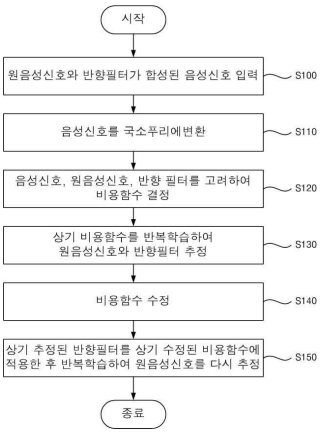
전체 청구항 수 : 총 4 항 심사관 : 권영학

(54) 발명의 명칭 **단일채널 음성신호에 대한 반향신호 제거방법 및 이를 이용한 음성인식장치**

(57) 요 약

본 발명에 따른 반향신호 제거방법은 단일의 마이크 등을 통해 입력되는 단일 채널의 반향된 음성 신호로부터 반향 필터를 추정하고 이를 이용하여 반향 신호를 제거한 원 음성 신호를 추정하는 것을 특징으로 한다. 이를 위하여, 상기 반향된 음성 신호가 우수한 성김 특성을 갖는 경우를 고려하여 원 음성신호(clean speech)와 반향된 음성 신호(reverberated speech) 사이의 성김도(sparseness) 차이를 확대시키기 위하여 신호 세기 스펙트럼(magnitude spectrum)을 사용한다. 또한, 음성 신호의 스펙트럼 성분들이 상호간에 주파수 의존성을 갖게 되는 일반적인 특징을 고려하여 비용 함수를 결정한다. 또한, 본 발명에 따른 방법은 반향 필터의 파워 스펙트럼은 시간 변수에 대하여 신호 세기가 지수함수적으로 감소하는 포락선(envelope)을 갖는 것을 고려하여 원 음성 신호와 반향 필터를 추정하는 것을 특징으로 한다.

대 표 도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호	2010-0025703
부처명	교육과학기술부
연구사업명	일반연구자 지원 사업 기본연구(유형1)
연구과제명	실세계 반향에 강인한 음성 인식 알고리즘 개발
기 여 율	1/1
주관기관	서강대학교 산학협력단
연구기간	2010.09.01 ~ 2013.08.31

특허청구의 범위

청구항 1

반향 환경에서 단일채널로 입력된 음성신호로부터 반향 신호를 제거하는 방법에 있어서,

(a) 원 음성신호(s(t))와 반향 필터(h(t))로 구성되어 단일채널로 입력된 음성신호(x(t))를 주파수 도메인의 신호로 변환하는 단계;

(b) 상기 변환된 음성신호($X_k[t]$)와 측정된 음성신호($Y_k[t]$)의 차이값, 원 음성신호에 대한 확률분포함수(P(S)) 및 반향 필터에 대한 확률분포함수(P(H))를 이용한 비용함수(f(S,H))를 결정하는 단계;

(c) 상기 비용함수를 최소화시키는 방향으로 반복학습하여, 원 음성신호 및 반향 필터를 추정하는 단계;

(d) 상기 변환된 음성신호($X_k[t]$)와 측정된 음성신호($Y_k[t]$)의 차이값만을 이용하여 상기 비용함수를 수정하는 단계;

(e) 상기 추정된 반향 필터를 이용하여, 상기 수정된 비용함수를 최소화시키는 방향으로 반복학습하여 원 음성신호를 재추정하는 단계;

를 구비하고, 상기 (d) 단계에서 수정된 비용함수(f(S,H))는 아래의 수학식에 의해 결정되는 것을 특징으로 하는 단일채널 음성신호에서의 반향신호 제거방법.

$$f(S,H)=\sum_{k,t}(Y_k[t]-X_k[t])^2$$

여기서, $X_k[t]$, $Y_k[t]$는 주파수 빈 k 에서의 시간 프레임 t의 음성신호, 측정된 음성 신호의 각각에 대한 세기 스펙트럼들을 각각 나타냄.

청구항 2

제1항에 있어서, 상기 (b) 단계의 비용함수(f(S,H))는 아래의 수학식에 의해 결정되는 것을 특징으로 하는 단일채널 음성신호에서의 반향신호 제거 방법.

$$f(S,H)=\sum_{k,t}(Y_k[t]-X_k[t])^2+2\lambda\sum_{t}\sqrt{\sum_{k}|S_k[t]|^2}+2\mu\sum_{k,t}\left|\frac{H_k[t]}{\exp(-\varphi t)}\right|^2$$

여기서, k는 주파수 빈(frequency bin)을 나타내며, t는 시간 프레임(time frame)을 나타내며, λ는 원 음성신호의 사전 확률의 비중(weight)을, μ는 반향 필터의 사전 확률의 비중(weight)을, φ는 반향 시간(RT60)에 대한 상수를 나타내며, $S_k[t]$, $H_k[t]$, $X_k[t]$, $Y_k[t]$는 주파수 빈 k 에서의 시간 프레임 t의 원 음성신호, 반향 필터, 음성신호, 측정된 음성 신호의 각각에 대한 세기 스펙트럼들을 각각 나타냄.

청구항 3

삭제

청구항 4

원 음성신호(s(t))와 반향 필터(h(t))로 구성되는 음성신호(x(t))를 입력받는 신호 입력부;

입력된 음성신호(x(t))를 주파수 영역 신호로 변환하는 신호 변환부;

사전 설정된 반향신호 제거 알고리즘을 이용하여 상기 입력된 음성 신호로부터 반향필터와 원 음성신호를 추정하는 원 음성신호 추정부;

를 구비하고, 상기 원음성신호 추정부에 이용되는 상기 반향신호 제거 알고리즘은,

(a) 상기 신호 변환부에 의해 변환된 음성신호($X_k[t]$)와 측정된 음성신호($Y_k[t]$)의 차이값, 원 음성신호에 대한 확률밀도함수(P(S)) 및 반향 필터에 대한 확률밀도함수(P(H))를 이용한 비용함수(f(S,H))를 결정하는 단계;

(b) 상기 비용함수를 최소화시키는 방향으로 반복학습하여, 원 음성신호 및 반향 필터를 추정하는 단계;

(c) 상기 변환된 음성신호($X_k[t]$)와 측정된 음성신호($Y_k[t]$)의 차이값만을 이용하여 상기 비용함수를 수정하는 단계;

(d) 상기 추정된 반향 필터를 이용하여, 상기 수정된 비용함수를 최소화시키는 방향으로 반복학습하여 원 음성신호를 재추정하는 단계;

를 구비하고, 상기 (c) 단계에서 수정된 비용함수(f(S,H))는 아래의 수학식에 의해 결정되는 것을 특징으로 하는 반향신호를 제거한 음성신호 인식장치.

$$f(S,H)=\sum_{k,t}(Y_k[t]-X_k[t])^2$$

여기서, $X_k[t]$, $Y_k[t]$ 는 주파수 빈 k 에서의 시간 프레임 t의 음성신호, 측정된 음성 신호의 각각에 대한 세기 스펙트럼들을 각각 나타냄.

청구항 5

제4항에 있어서, 상기 (a) 단계의 비용함수(f(S,H))는 아래의 수학식에 의해 결정되는 것을 특징으로 하는 반향신호를 제거한 음성신호 인식 장치.

$$f(S,H)=\sum_{k,t}(Y_k[t]-X_k[t])^2+2\lambda\sum_{t}\sqrt{\sum_{k}|S_k[t]|^2}+2\mu\sum_{k,t}\left|\frac{H_k[t]}{\exp(-\varphi t)}\right|^2$$

청구항 6

삭제

명 세 서

기 술 분 야

[0001] 본 발명은 음성인식장치에 관한 것으로서, 더욱 구체적으로는 반향 환경에서 입력되는 단일채널의 음성신호로부터 반향신호를 제거하고 원 음성신호를 추출하는 음성 인식 장치 및 반향신호 제거 방법에 관한 것이다.

배 경 기 술

[0002] 음성 인식 기술은 마이크나 전화를 통하여 얻어진 음향학적 신호를 단어나 단어 집합 또는 문장으로 변환하는 기술로서, 이렇게 인식된 결과들은 명령이나, 제어, 데이터 입력, 문서 준비 등의 응용 분야에서 최종결과로 사용될 수 있다. 이러한 음성 인식 기술을 이용한 음성 인식 시스템에 대한 응용분야가 최근 증가하고 있으며, 이에 대한 다양한 연구와 개발이 진행되고 있다.

[0003] 실제 환경에 있어서, 마이크 등의 신호 입력수단으로 입력되는 음성 신호는 스피치와 같은 원 음성신호뿐만 아니라, 벽이나 천장 등의 주변 환경에 따른 여러 경로에 의해 반향된 신호를 포함하게 된다.

[0004] 따라서, 음성 인식 시스템은 입력된 음성신호로부터 반향 신호와 원 음성신호를 추정하고, 이를 이용하여 반향 신호를 제거하기 위한 다양한 방법들을 제안하고 있다. 종래의 음성 인식 시스템은 사전 설정된 비용 함수를 최소화시키도록 반복학습함으로써, 반향 필터와 원 음성신호를 추정하게 된다. 이 때, 종래의 방법들은 비용 함수를 결정함에 있어서, 원 음성신호가 갖는 주파수 의존성을 고려하지 않을 뿐만 아니라, 반향 필터에 대하여도 고려하지 않는 것을 특징으로 한다.

[0005] 하지만, 이러한 알고리즘은 비음수의 파워 스펙트럼을 계산하는 과정이 복잡하다는 문제점이 있으며, 또한 원 음성신호가 반향된 신호보다 일반적으로 더 성김 특성을 갖는 것을 고려함에 따라 간혹 반향 신호가 원 음성신호로 잘못 추정되는 상황이 발생되는 문제점이 있다.

[0006] 이러한 문제점들로 인하여, 반향 환경에 대한 반향 필터를 정확하게 추정하지 못하고, 그 결과 원 음성 신호를 제대로 추정하지 못하거나 계산 과정이 매우 복잡해지게 된다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 1. 한국등록특허 제10-651940호

(특허문헌 0002) 2. 한국공개특허 제10-2012-54087호

(특허문헌 0003) 3. 한국공개특허 제10-2007-121271호

(특허문헌 0004) 4. 한국등록특허 제10-1112154호

(특허문헌 0005) 5. 한국공개특허 제10-2011-70764호

발명의 내용

해결하려는 과제

[0008] 전술한 문제점을 해결하기 위한 본 발명의 목적은 비음수 암묵신호 반향제거알고리즘의 계산을 간단히 하고, 원 음성신호의 주파수 의존성을 고려하고 반향 필터의 지수함수적으로 감소되는 포락선을 고려하여 반향 필터와 원 음성신호를 추정하는 반향 제거 방법을 제공하는 것이다.

[0009] 본 발명의 다른 목적은 전술한 반향 제거 방법을 적용하여 음성인식하는 음성인식장치를 제공하는 것이다.

과제의 해결 수단

[0010] 전술한 기술적 과제를 달성하기 위한 본 발명의 제1 특징은, 반향 환경에서 단일채널로 입력된 음성신호로부터 반향 신호를 제거하는 방법에 관한 것으로서, 상기 반향신호 제거 방법은, (a) 원 음성신호(s(t))와 반향 필터(h(t))로 구성되어 단일채널로 입력된 음성신호(x(t))를 주파수 도메인의 신호로 변환하는 단계; (b) 상기 변환된 음성신호($X_k[t]$)와 측정된 음성신호($Y_k[t]$)의 차이값, 원 음성신호에 대한 확률분포함수(P(S)) 및 반향 필터에 대한 확률분포함수(P(H))를 이용한 비용함수(f(S,H))를 결정하는 단계; (c) 상기 비용함수를 최소화시키는 방향으로 반복학습하여, 원 음성신호 및 반향 필터를 추정하는 단계; (d) 상기 변환된 음성신호

($X_k[t]$)와 측정된 음성신호($Y_k[t]$)의 차이값만을 이용하여 상기 비용함수를 수정하는 단계; (e) 상기 추정된 반향 필터를 이용하여, 상기 수정된 비용함수를 최소화시키는 방향으로 반복학습하여 원 음성신호를 재추정하는 단계;를 구비한다.

[0011] 전술한 제1 특징에 따른 반향 신호 제거 방법에 있어서, 상기 (b) 단계의 비용함수(f(S,H))는 아래의 수학식에 의해 결정되며,

$$f(S,H)=\sum_{k,t}(Y_k[t]-X_k[t])^2+2\lambda\sum_t\sqrt{\sum_k|S_k[t]|^2}+2\mu\sum_{k,t}\left|\frac{H_k[t]}{\exp(-\varphi t)}\right|^2$$

[0012]

[0013] 여기서, k는 주파수 빈(frequency bin)을 나타내며, t는 시간 프레임(time frame)을 나타내며, λ는 원 음성신호의 사전 확률의 비중(weight)을, μ는 반향 필터의 사전 확률의 비중(weight)을, φ는 반향 시간(RT60)에 대한 상수를 나타내며, $S_k[t]$, $H_k[t]$, $X_k[t]$, $Y_k[t]$는 주파수 빈 k 에서의 시간 프레임 t의 원 음성신호, 반향 필터, 음성신호, 측정된 음성 신호의 각각에 대한 세기 스펙트럼들을 각각 나타낸다.

[0014] 전술한 제1 특징에 따른 반향 신호 제거 방법에 있어서, 상기 (d) 단계에서 수정된 비용함수(f(S,H))는 아래의 수학식에 의해 결정되는 것이 바람직하다.

$$f(S,H)=\sum_{k,t}(Y_k[t]-X_k[t])^2$$

[0015]

[0016] 본 발명의 제2 특징에 따른 음성 신호 인식 장치는, 원 음성신호(s(t))와 반향 필터(h(t))로 구성되는 음성신호(x(t))를 입력받는 신호 입력부; 입력된 음성신호(x(t))를 주파수 영역 신호로 변환하는 신호 변환부; 사전 설정된 반향신호 제거 알고리즘을 이용하여 상기 입력된 음성 신호로부터 반향필터와 원 음성신호를 추정하는 원 음성신호 추정부;를 구비하고,

[0017] 상기 원음성신호 추정부에 이용되는 상기 반향신호 제거 알고리즘은, (a) 상기 신호 변환부에 의해 변환된 음성신호($X_k[t]$)와 측정된 음성신호($Y_k[t]$)의 차이값, 원 음성신호에 대한 확률밀도함수(P(S)) 및 반향 필터에 대한 확률밀도함수(P(H))를 이용한 비용함수(f(S,H))를 결정하는 단계;(b) 상기 비용함수를 최소화시키는 방향으로 반복학습하여, 원 음성신호 및 반향 필터를 추정하는 단계;(c) 상기 변환된 음성신호($X_k[t]$)와 측정된 음성신호($Y_k[t]$)의 차이값만을 이용하여 상기 비용함수를 수정하는 단계; (d) 상기 추정된 반향 필터를 이용하여, 상기 수정된 비용함수를 최소화시키는 방향으로 반복학습하여 원 음성신호를 재추정하는 단계;를 구비한다.

발명의 효과

[0018] 본 발명에 의하여, 반향 환경에서의 음성 인식을 정확하게 수행하게 된다.

[0019] 일반적으로 음성 인식 성능을 비교하기 위하여, 원 음성신호와 추정된 음성신호의 멜-주파수 캡스트럼 계수(Mel-frequency cepstral coefficient; 'MFCC') 차이값(CD)을 각 방법에 따라 추출하여 비교한다. 전술한 CD는 수학식 12에 의해 구할 수 있다.

수학식 12

$$CD = \frac{1}{T}\sum_{t}\sum_{c}(M_c[t] - \hat{M}_c[t])^2$$

[0020]

[0021] 여기서, $M_c[t]$와 $\hat{M}_c[t]$는 프레임 t에 있어서의 원 음성신호와 추정된 음성 신호의 c-번째 MFCC 값을 각각 의미한다.

[0022] 도 2는 종래의 방법과 본 발명에 따른 방법을 이용하여 추정된 음성신호에 대한 CD들을 비교한 그래프이다. 도 2에 있어서, 실선은 반향 필터를 제거하지 않은 상태에서 추정된 음성신호와 원 음성신호의 CD를 도시한 것이며, 점선은 종래의 방법에 의해 반향 필터를 제거한 상태에서 추정된 음성 신호와 원 음성신호의 CD를 도시한 것이며, 일점 쇄선은 본 발명에 의해 반향 필터를 제거한 상태에서 추정된 음성 신호와 원 음성 신호와의 CD를 도시한 것이다.

[0023] 도 2를 통해, 본 발명에 따른 방법에 의해 추정된 음성 신호와 원 음성 신호의 CD가 가장 낮은 것을 알 수 있다. 따라서, 본 발명에 따른 반향 제거 방법을 사용하는 음성 인식 장치의 성능이 우수함을 알 수 있다.

도면의 간단한 설명

[0024] 도 1은 본 발명의 바람직한 실시예에 따른 반향 제거 방법을 순차적으로 도시한 흐름도이다.

도 2는 종래의 방법과 본 발명에 따른 방법을 이용하여 추정된 음성신호에 대한 CD들을 비교한 그래프이다.

발명을 실시하기 위한 구체적인 내용

[0025] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예에 따른 단일채널 로 입력된 음성신호에서의 반향 제거 방법에 대하여 구체적으로 설명한다. 본 발명에 따른 반향제거방법은 단일의 마이크 등을 통해 입력되는 단일 채널의 반향된 음성 신호로부터 반향 필터를 추정하고 이를 이용하여 반향 신호를 제거한 원 음성 신호를 추정하는 것을 특징으로 한다. 이를 위하여, 본 발명에 따른 반향 제거 방법은, 상기 반향된 음성 신호가 우수한 성김 특성을 갖는 경우를 고려하여 원 음성신호(clean speech)와 반향된 음성 신호(reverberated speech) 사이의 성김도(sparseness) 차이를 확대시키기 위하여 신호 세기 스펙트럼(magnitude spectrum)을 사용하는 것을 특징으로 한다. 또한, 본 발명에 따른 방법은, 음성 신호의 스펙트럼 성분들이 상호간에 주파수 의존성을 갖게 되는 일반적인 특징을 고려하여 비용 함수를 결정하는 것을 특징으로 한다. 또한, 본 발명에 따른 방법은 반향 필터의 파워 스펙트럼은 시간 변수에 대하여 신호 세기가 지수함수적으로 감소하는 포락선(envelope)을 갖는 것을 고려하여 원 음성 신호와 반향 필터를 추정하는 것을 특징으로 한다.

[0026] 본 발명에 따른 음성 인식 장치는 신호 입력부, 신호 변환부 및 음성신호 추정부를 구비하여, 음성 신호로부터 원 음성신호 및 반향 필터를 추정함으로써, 반향 신호가 제거된 원 음성신호를 정확하게 추정하는 것을 특징으로 한다.

[0027] 상기 음성 인식 장치는 원 음성신호, 및 상기 원 음성신호가 반향된 반향 신호를 나타내는 반향 필터가 합성되어 이루어지는 음성신호가 마이크로폰과 같은 신호 입력부를 통해 입력되면, 상기 신호 변환부에 의해 입력된 음성 신호를 국소 퓨리에 변환한 후, 상기 음성신호 추정부가 사전 설정된 반향 제거 알고리즘을 적용하여 원 음성신호 및 반향 필터를 추정하게 된다. 상기 음성신호 추정부에서 사용하는 반향 제거 알고리즘은 본 발명에 따른 반향 제거 방법과 동일하다.

[0028] 도 1은 본 발명의 바람직한 실시예에 따른 반향제거방법을 순차적으로 도시한 흐름도이다.

[0029] 단일의 마이크 등과 같은 신호 입력부를 통해, 원음성신호와 반향필터가 합성된 음성 신호(the reverberated

speech)(x(t))가 입력되면, 국소 푸리에 변환(Short-Time Furier Transform; 'STFT')) 한다. STFT 변환된 주파수 도메인에 있어서, [t] 번째 프레임에서의 k-번째 주파수 빈(frequency bin)의 음성 신호($x_k[t]$)를 수학식 1과 같이 모델링한다.

수학식 1

$$x_k[t] = s_k[t] * h_k[t] = \sum_{\tau} s_k[\tau] h_k[t-\tau]$$

[0030]

[0031] 여기서, $s_k[t]$ 및 $h_k[t]$ 는 각각 원 음성신호 및 반향 필터의 k 번째 서브밴드 신호를 나타내며, * 는 컨볼류션 연산자를 나타낸다.

[0032] 본 발명에 따른 방법에 있어서, 연산을 간단하게 하기 위하여 $s_k[t]$ 와 $h_k[t]$ 는 서로 상관되어 있지 않다(uncorrelated)고 가정함에 따라, 음성 신호($x_k[t]$)의 세기 스펙트럼(magnitude spectrum)은 음성 신호와 원 음성신호의 성김도 차이(sparsity difference)를 크게 하기 위하여 수학식 2와 같이 $s_k[t]$ 와 $h_k[t]$ 의 세기들의 컨볼류션에 의해 근사될 수 있다.

수학식 2

$$\begin{aligned} X_k[t] &= \left| \sum_{\tau} s_k[\tau] h_k[t-\tau] \right| \\ &\approx \sum_{\tau} S_k[\tau] H_k[t-\tau] \end{aligned}$$

[0033]

[0034] 여기서, $X_k[t]$, $S_k[t]$, $H_k[t]$ 는 각각 $|x_k[t]|$, $|s_k[t]|$, $|h_k[t]|$ 를 나타낸다.

[0035] 반향 필터와 원 음성 신호로 구성된 음성신호가 마이크로폰으로 입력되어 측정되면, 측정된 음성 신호의 서브밴드 파워($y_k[t]$)는 수학식 3과 같이 모델링된다.

수학식 3

$$Y_k[t] = X_k[t] + \varepsilon_k[t]$$

[0036]

[0037] 여기서, $\varepsilon_k[t]$ 는 부가적인 측정시의 잡음 또는 모델 오류를 나타낸다.

[0038] 본 발명에 따른 방법에서, 수학식 4에 의한 maximun - a- posteriori(MAP) 추정에 의해, $Y \equiv \{Y_k[1], \cdots, Y_k[T]\}_{k=1}^{K}$ 과 함께 $S \equiv \{S_k[1], \cdots, S_k[T]\}_{k=1}^{K}$, $H \equiv \{H_k[1], \cdots, H_k[T]\}_{k=1}^{K}$ 을 구하게 된다. 여기서, k는 주파수 빈(frequency bin)의 인덱스를, K는 주파수 빈의 총 개수를, T는 Time Frame의 총 개수를 의미한다.

수학식 4

$$\{S^*, H^*\} = \arg\max_{S,H} P(S,H|Y)$$
$$= \arg\max_{S,H} P(Y|S,H)P(S)P(H)$$

[0039]

[0040] 여기서 S^*, H^* 은 원 음성신호 및 반향 필터의 최적값을 나타낸다.

[0041] 여기서, P(S)는 원 음성신호에 대한 확률 분포(probability distribution)함수를, P(H)는 반향 필터의 확률 분포 함수를, P(S,H|Y)는 측정된 음성신호(Y)에 대한 원 음성 신호(S)와 반향 필터(H)의 사후 확률 함수를, P(Y|S,H)는 측정된 음성신호(Y)에 대한 원 음성 신호(S)와 반향 필터(H)의 우도(likelihood) 함수를 의미한다.

[0042] 한편, 본 발명에 따른 방법에 있어서, 원 음성신호(Clean speech)는 주파수에 따라 상호간의 의존성을 갖는 것을 제1 전제 조건으로 함으로써, 원 음성신호에 대한 확률분포함수 (P(S))는 수학식 5와 같이 표현한다.

수학식 5

$$P(S) \propto \prod_t \exp\left(-\sqrt{\sum_k \left|\frac{S_k[t]}{\sigma_k}\right|^2}\right)$$

[0043]

[0044] 여기서, σ_k 는 k 번째 주파수에서의 편차를 나타낸다.

[0045] 또한, 주파수 빈에서 추정된 원 음성신호가 임의로 쉬프트됨에 따라 발생되는 모호성을 회피하기 위하여, 본 발명에 따른 방법에서는 반향 필터의 세기 스펙트럼이 시간 변수에 대하여 지수 함수적으로 감소하는 포락선을 갖는다는 제2 전제조건을 적용함으로써, 반향 필터(H)에 대한 확률분포함수(P(H))를 수학식 6과 같이 표현한다.

수학식 6

$$P(H) \propto \prod_{k,t} \exp\left(-\left|\frac{H_k[t]}{\exp(-\varphi t)}\right|^2\right)$$

[0046]

[0047] 여기서, Φ는 반향 시간(RT60)에 대한 상수이다.

[0048] 전술한 제1 및 제2 전제 조건들을 적용하여 수학식 7로 표현되는 비용함수(f(S,H))를 최소화시키는 방향으로 반복학습하여 원 음성신호(S)와 반향 필터(H)를 추정한다.

수학식 7

$$f(S,H) = \sum_{k,t}(Y_k[t] - X_k[t])^2 + 2\lambda \sum_t \sqrt{\sum_k |S_k[t]|^2} + 2\mu \sum_{k,t} \left|\frac{H_k[t]}{\exp(-\varphi t)}\right|^2$$

[0049]

[0050] 여기서, k는 주파수 빈(frequency bin)을 나타내며, t는 시간 프레임(time frame)을 나타내며, λ는 원 음성신호의 사전 확률의 비중(weight)을, μ는 반향 필터의 사전 확률의 비중(weight)을, ϕ는 반향 시간(RT60)에 대한 상수를 나타내며, $S_k[t]$, $H_k[t]$, $X_k[t]$, $Y_k[t]$는 주파수 빈 k 에서의 시간 프레임 t의 원 음성신호, 반향 필터, 음성신호, 측정된 음성 신호의 각각에 대한 세기 스펙트럼들을 각각 나타낸다.

[0051] 전술할 비용함수를 최소화시키는 원 음성신호의 세기 스펙트럼($S_k[t]$) 및 반향 필터의 세기 스펙트럼($H_k[t]$)을 수학식 8 및 9를 이용하여 반복학습하여 추정한다.

수학식 8

$$\begin{aligned} S_k[\tau] &= S_k'[\tau] \frac{\sum_t H_k'[t-\tau] Y_k[t]}{\sum_t H_k'[t-\tau] X_k'[t] + \lambda \frac{S_k'[\tau]}{\sqrt{\sum_k |S_k'[\tau]|^2}}} \\ &= S_k'[\tau] \frac{F^{-1}\{F\{H_k'[t]\}_u^* F\{Y_k[t]\}_u\}_\tau}{F^{-1}\{F\{S_k'[t]\}_u | F\{H_k'[t]\}_u|^2\}_\tau + \lambda \frac{S_k'[\tau]}{\sqrt{\sum_k |S_k'[\tau]|^2}}} \end{aligned}$$

[0052]

수학식 9

$$H_k[\tau] = H_k'[\tau]\frac{\sum_t S_k'[t-\tau]Y_k[t]}{\sum_t S_k'[t-\tau]X_k'[t] + 2\mu\frac{H_k[t]}{\exp(-\varphi t)}} = H_k'[\tau]\frac{F^{-1}\{F\{S_k'[t]\}_u^* F\{Y_k[t]\}_u\}_\tau}{F^{-1}\{|F\{S_k'[t]\}_u|^2 F\{H_k'[t]\}_u\}_\tau + 2\mu\frac{H_k[t]}{\exp(-\varphi t)}}$$

[0053]

[0054] 다음, 본 발명에 따른 방법에 의하여 추정된 원 음성신호들이 원 음성 신호에 대한 부정확한 전제 조건들로 인하여 발생될 수 있는 오류들을 가지는 것을 방지하기 위하여, 원 음성신호의 재추정한다.

[0055] 원 음성신호를 재추정하기 위하여, 제1 및 제2 전제조건들을 적용하지 않은비용함수를 다시 설정한다. 수학식 10은 재설정된 비용함수를 나타낸 것이다.

수학식 10

$$f(S,H) = \sum_{k,t}(Y_k[t] - X_k[t])^2$$

[0056]

[0057] 재설정된 비용 함수에 상기 수학식 9에 의해 추정된 반향 필터의 세기 스펙트럼($H_k[t]$)을 대입하여, 상기 재설정된 비용 함수를 최소화시키는 원 음성신호의 세기 스펙트럼을 수학식 11을 이용하여 반복학습하여 재추정한다.

수학식 11

$$S_k[\tau] = S_k'[\tau]\frac{\sum_t H_k[t-\tau]Y_k[t]}{\sum_t H_k[t-\tau]X_k'[t]} = S_k'[\tau]\frac{F^{-1}\{F\{H_k[t]\}_u^* F\{Y_k[t]\}_u\}_\tau}{F^{-1}\{F\{S_k'[t]\}_u|F\{H_k[t]\}_u|^2\}_\tau}$$

[0058]

[0059] 전술한 전제 조건들에 의하여 원 음성신호의 스펙트럼 성분들이 과도하게 왜곡되더라도, 전술한 원 음성신호에 대한 재추정과정을 통하여 보상될 수 있게 된다.

[0060] 이상에서 본 발명에 대하여 그 바람직한 실시예를 중심으로 설명하였으나, 이는 단지 예시일 뿐 본 발명을 한정하는 것이 아니며, 본 발명이 속하는 분야의 통상의 지식을 가진 자라면 본 발명의 본질적인 특성을 벗어나지

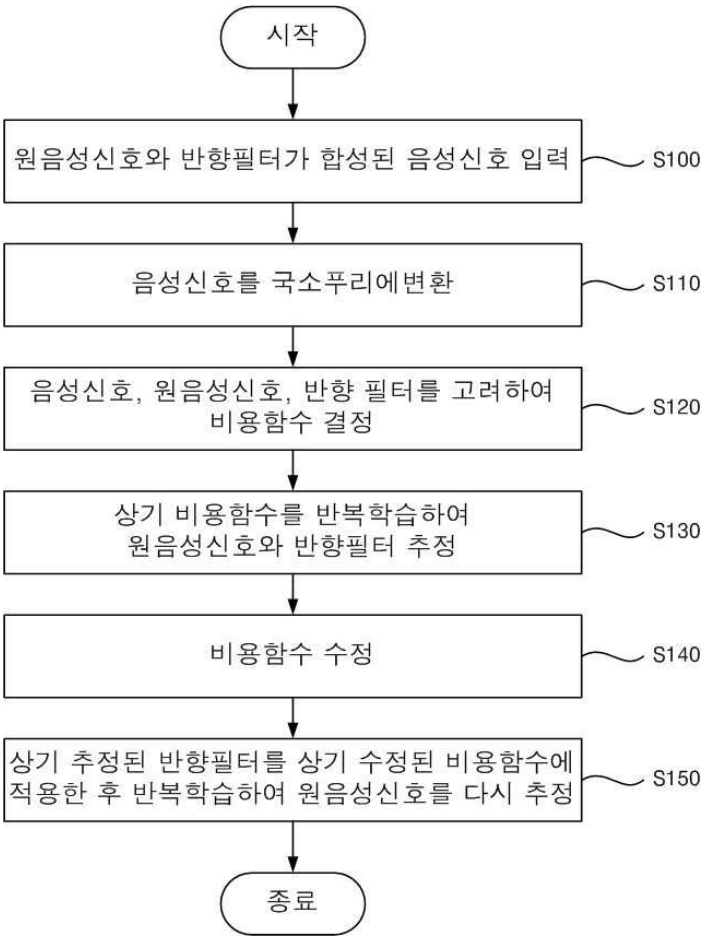
않는 범위에서 이상에 예시되지 않은 여러 가지의 변형과 응용이 가능함을 알 수 있을 것이다. 그리고, 이러한 변형과 응용에 관계된 차이점들은 첨부된 청구 범위에서 규정하는 본 발명의 범위에 포함되는 것으로 해석되어야 할 것이다.

산업상 이용가능성

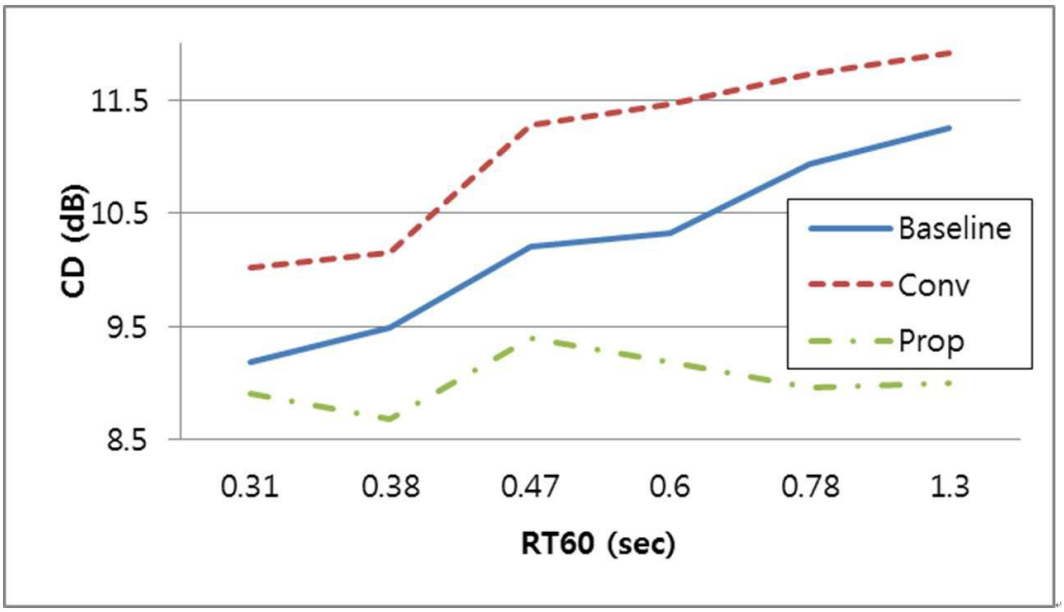
[0061] 본 발명에 따른 반향 제거 방법은 반향 환경에서의 단일 채널 음성 인식 시스템에 사용될 수 있다.

도면

도면1



도면2


11.5
10.5
9.5
8.5
CD (dB)
0.31
0.38
0.47
0.6
0.78
1.3
RT60 (sec)
Baseline
Conv
Prop