



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2016-0149736
(43) 공개일자 2016년12월28일

(51) 국제특허분류(Int. Cl.)
G10L 21/0208 (2013.01) G10L 15/28 (2006.01)
(52) CPC특허분류
G10L 21/0208 (2013.01)
G10L 15/28 (2013.01)
(21) 출원번호 10-2015-0087304
(22) 출원일자 2015년06월19일
심사청구일자 없음

(71) 출원인
엘지전자 주식회사
서울특별시 영등포구 여의대로 128 (여의도동)
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
박형민
서울특별시 강남구 삼성로 212, 3동 1002호 (대치동, 은마아파트)
정규혁
서울특별시 서초구 양재대로11길 19
(뒷면에 계속)
(74) 대리인
김기문

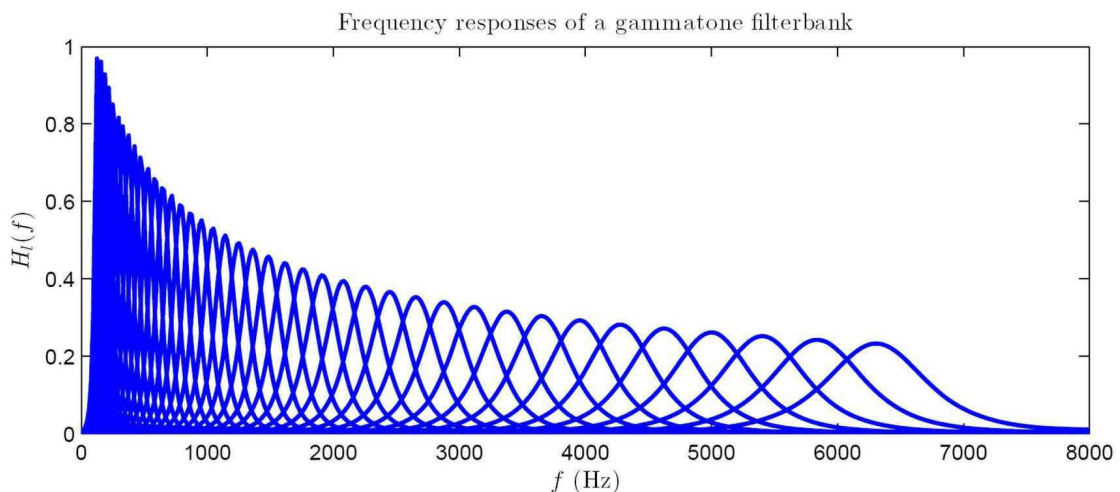
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 음성 인식 장치 및 그 동작 방법

(57) 요약

음성을 인식하는 음성 인식 장치가 개시된다. 음성 인식 장치는 음성 신호를 수신하는 음성 신호 수신부; 및 상기 음성 신호를 주파수 영역의 주파수 스펙트럼으로 변환하고, 하모닉 파워 비율에 기초하여 상기 주파수 스펙트럼의 정상 성분을 억제하는 제어부를 포함하고, 상기 하모닉 파워 비율은 전체 주파수 대역의 파워 대비 하모닉 성분들의 파워 비율이다.

대표도 - 도4



(72) 발명자

최민석

서울특별시 서초구 양재대로11길 19

조병준

서울특별시 서초구 양재대로11길 19

전혜정

서울특별시 서초구 양재대로11길 19

명세서

청구범위

청구항 1

음성 신호를 수신하는 음성 신호 수신부; 및

상기 음성 신호를 주파수 영역의 주파수 스펙트럼으로 변환하고, 하모닉 파워 비율에 기초하여 상기 주파수 스펙트럼의 정상 성분을 억제하는 제어부를 포함하고,

상기 하모닉 파워 비율은 전체 주파수 대역의 파워 대비 하모닉 성분들의 파워 비율인

음성 인식 장치.

청구항 2

제1항에 있어서,

상기 제어부는 상기 주파수 스펙트럼의 미리 정하여진 일정한 기준 주파수 보다 낮은 주파수 대역에 선택적으로 상기 하모닉 파워 비율을 적용하여 상기 주파수 스펙트럼의 정상 성분을 억제하는

음성 인식 장치.

청구항 3

제1항에 있어서,

상기 제어부는 채널 파워 비율에 기초하여 상기 주파수 스펙트럼의 정상 성분을 억제하고,

채널 파워 비율은 채널 파워 비율은 전체 감마톤 채널의 파워 대비 고주파 영역에 해당하는 감마톤 채널들의 파워 비율인

음성 인식 장치.

청구항 4

제3항에 있어서,

상기 제어부는 상기 주파수 스펙트럼의 미리 정하여진 일정한 기준 주파수 보다 높은 주파수 대역에 해당하는 감마톤 채널들에 대해서 선택적으로 채널 파워 비율을 적용하여 정상 성분을 억제하는

음성 인식 장치.

청구항 5

제1항에 있어서,

상기 제어부는 2개의 음성 인식 센서를 이용하여 얻은 2개의 음성 신호를 상호 상관하여 상기 음성 신호를 생성하는

음성 인식 장치.

청구항 6

제5항에 있어서,

상기 제어부는 상기 2개의 음성 신호를 합산 빔포밍 처리하는

음성 인식 장치.

청구항 7

음성 신호를 수신하는 단계;
 상기 음성 신호를 주파수 영역의 주파수 스펙트럼으로 변환하는 단계; 및
 하모닉 파워 비율에 기초하여 상기 주파수 스펙트럼의 정상 성분을 억제하는 단계를 포함하고,
 상기 하모닉 파워 비율은 전체 주파수 대역의 파워 대비 하모닉 성분들의 파워 비율인
 음성 인식 장치의 동작 방법.

청구항 8

제7항에 있어서,
 상기 하모닉 파워 비율에 기초하여 상기 주파수 스펙트럼의 정상 성분을 억제하는 단계는
 상기 주파수 스펙트럼의 미리 정하여진 일정한 기준 주파수 보다 낮은 주파수 대역에 선택적으로 상기 하모닉
 파워 비율을 적용하여 상기 주파수 스펙트럼의 정상 성분을 억제하는 단계를 포함하는
 음성 인식 장치의 동작 방법.

청구항 9

제7항에 있어서,
 채널 파워 비율에 기초하여 상기 주파수 스펙트럼의 정상 성분을 억제하는 단계를 더 포함하고,
 상기 채널 파워 비율은 채널 파워 비율은 전체 감마톤 채널의 파워 대비 고주파 영역에 해당하는 감마톤 채널들
 의 파워 비율인
 음성 인식 장치의 동작 방법.

청구항 10

제9항에 있어서,
 상기 채널 파워 비율에 기초하여 상기 주파수 스펙트럼의 정상 성분을 억제하는 단계는
 상기 주파수 스펙트럼의 미리 정하여진 일정한 기준 주파수 보다 높은 주파수 대역에 해당하는 감마톤 채널들에
 대해서 선택적으로 채널 파워 비율을 적용하여 정상 성분을 억제하는 단계를 포함하는
 음성 인식 장치의 동작 방법.

발명의 설명

기술 분야

[0001] 본 발명은 음성 인식 장치 및 그 동작 방법 에 관한 것으로, 특히 음성 인식 장치가 입력된 음성 신호에서 잔향
 요소들을 억제하여 음성 인식의 정확성 높이는 것에 대한 발명이다.

배경 기술

[0002] 사용자 인터페이스에 대한 관심과 음성 처리 기술의 발달로, 음성 인식 장치를 사용하는 전자제품이 늘어나고
 있다. 예컨대, 사용자의 음성을 인식하여 동작하는 이동 통신 단말기, 디지털 텔레비전이 널리 사용되고 있다.
 다만, 이러한 전자 제품이 음성 인식 과정에서 음성을 정확히 인식하지 못하여 오동작이 이루어지는 경우가 많
 다. 또한, 사용자들은 전자 제품을 동작시키기 위하여 동일한 명령어를 반복적으로 말하여야 하는 경우가 많다.
 이로 인하여 사용자들은 많은 불편을 겪고 있다. 특히, 주위 사물에 의하여 반사되는 이전 음성인 잔향의 영향
 을 많이 받는 실내에서 음성인식이 문제된다. 따라서 잔향이 존재하는 실내 공간에서도 정확하게 음성을 인식할
 수 있는 음성 인식 장치가 필요하다.

발명의 내용

해결하려는 과제

[0003] 본 발명의 일 실시예는 사용자의 음성을 정확하게 인식하는 음성 인식 장치 및 음성 인식 장치의 동작 방법을 제공하기 위한 것이다.

[0004] 특히, 본 발명의 일 실시예는 잔향이 존재하는 실내 공간에서도 사용자의 음성을 정확하게 인식하는 음성 인식 장치 및 음성 인식 장치의 동작 방법을 제공하기 위한 것이다.

과제의 해결 수단

[0005] 본 발명의 일 실시예에 따른 음성 인식 장치는 음성 신호를 수신하는 음성 신호 수신부; 및 상기 음성 신호를 주파수 영역의 주파수 스펙트럼으로 변환하고, 하모닉 파워 비율에 기초하여 상기 주파수 스펙트럼의 정상 성분을 억제하는 제어부를 포함하고, 상기 하모닉 파워 비율은 전체 주파수 대역의 파워 대비 하모닉 성분들의 파워 비율이다.

[0006] 상기 제어부는 상기 주파수 스펙트럼의 미리 정하여진 일정한 기준 주파수 보다 낮은 주파수 대역에 선택적으로 상기 하모닉 파워 비율을 적용하여 상기 주파수 스펙트럼의 정상 성분을 억제할 수 있다.

[0007] 상기 제어부는 채널 파워 비율에 기초하여 상기 주파수 스펙트럼의 정상 성분을 억제하고, 채널 파워 비율은 채널 파워 비율은 전체 감마톤 채널의 파워 대비 고주파 영역에 해당하는 감마톤 채널들의 파워 비율일 수 있다.

[0008] 상기 제어부는 상기 주파수 스펙트럼의 미리 정하여진 일정한 기준 주파수 보다 높은 주파수 대역에 해당하는 감마톤 채널들에 대해서 선택적으로 채널 파워 비율을 적용하여 정상 성분을 억제할 수 있다.

[0009] 상기 제어부는 2개의 음성 인식 센서를 이용하여 얻은 2개의 음성 신호를 상호 상관하여 상기 음성 신호를 생성할 수 있다.

[0010] 상기 제어부는 상기 2개의 음성 신호를 합산 빔포밍 처리할 수 있다.

[0011] 본 발명의 일 실시예에 따른 음성 인식 장치의 동작 방법은 음성 신호를 수신하는 단계; 상기 음성 신호를 주파수 영역의 주파수 스펙트럼으로 변환하는 단계; 및 하모닉 파워 비율에 기초하여 상기 주파수 스펙트럼의 정상 성분을 억제하는 단계를 포함하고, 상기 하모닉 파워 비율은 전체 주파수 대역의 파워 대비 하모닉 성분들의 파워 비율이다.

발명의 효과

[0012] 본 발명의 일 실시예는 사용자의 음성을 정확하게 인식하는 음성 인식 장치 및 음성 인식 장치의 동작 방법을 제공한다.

[0013] 특히, 본 발명의 일 실시예는 잔향이 존재하는 실내 공간에서도 사용자의 음성을 정확하게 인식하는 음성 인식 장치 및 음성 인식 장치의 동작 방법을 제공한다.

도면의 간단한 설명

[0014] 도 1은 본 발명의 일 실시예에 따른 음성 인식 장치의 구성을 블록도로 도시한 것이다.

도 2는 본 발명의 일 실시예에 따른 음성 인식 장치가 포함하는 제어부의 구성을 보여준다.

도 3은 본 발명의 일 실시예에 따른 음성 인식 장치의 동작을 보여주는 흐름도이다.

도 4는 본 발명의 일 실시예에 따른 음성 인식 장치가 포함하는 감마톤 필터의 주파수 응답을 보여준다.

도 5는 본 발명의 또 다른 실시예에 따른 음성 인식 장치가 포함하는 제어부의 구성을 보여준다.

도 6은 본 발명의 또 다른 실시예에 따른 음성 인식 장치의 동작을 보여주는 흐름도이다.

도 7은 음성 인식 장치의 제어부가 하모닉 파워 비율을 획득하는 동작을 보여주는 흐름도이다.

도 8은 무향 및 반향 환경에서 평균 파워 차감이 적용되지 않은 시간-주파수 영역에서의 입력 신호 파워와 하모닉 파워 비율을 보여준다.

도 9는 무향 환경, 반향 환경에서 각각 평균 파워 차감을 하였을 때, 각각의 시간-주파수 영역 파워와 하모닉 파워 비율을 보여준다.

도 10은 음성 인식 장치의 제어부가 채널 파워 비율을 획득하는 동작을 보여주는 흐름도이다.

도 11은 무향 및 반향 환경에서 평균 파워 차감을 하지 않았을 때의 감마톤 채널 영역 입력 파워와 채널 파워 비율을 보여준다.

도 12는 무향 및 반향 환경에서 평균 파워 차감을 하였을 때의 감마톤 채널 영역 입력 파워와 채널 파워 비율을 보여준다.

발명을 실시하기 위한 구체적인 내용

- [0015] 아래에서는 첨부한 도면을 참고로 하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0016] 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.
- [0017] 도 1은 본 발명의 일 실시예에 따른 음성 인식 장치의 구성을 블록도로 도시한 것이다.
- [0018] 음성 인식 장치(100)는 음성 신호 수신부(170) 및 제어부(110)를 포함한다.
- [0019] 음성 신호 수신부(170)는 외부로부터 음성 신호를 수신한다. 구체적인 실시예에서 음성 신호 수신부(170)는 하나 또는 복수의 마이크를 포함할 수 있다. 또한, 음성 신호 수신부(170)는 마이크를 통해 음성 신호를 수신할 수 있다. 또 다른 구체적인 실시예에서 음성 신호 수신부(170)는 다른 장치가 인식한 음성 신호를 수신할 수 있다. 구체적으로 음성 신호 수신부(170)는 다른 장치가 인식한 음성 신호를 외부 장치와 통신하는 통신부를 통하여 수신할 수 있다.
- [0020] 제어부(110)는 음성 인식 동작을 제어한다. 구체적으로 제어부(110)는 반향환경에서 음성 인식을 위한 전처리를 수행한다. 구체적으로 제어부(110)는 반향 환경에서 초기 도착 성분을 강조하고 나중에 입력되는 반사 신호를 억제하는 정상 성분 억제 방법을 수행한다. 구체적으로 제어부(110)는 정상 성분의 차감을 통해 음성 신호의 시작점(onset) 성분을 강조하고 하강 에지(falling edge) 성분을 억제한다. 이는 사람의 청각 처리 특성을 모방하는 것이다. 구체적인 제어부(110)의 구성에 대해서는 도 2를 통해 설명한다.
- [0021] 도 2는 본 발명의 일 실시예에 따른 음성 인식 장치가 포함하는 제어부의 구성을 보여준다.
- [0022] 음성 인식 장치(100)의 제어부(110)는 프리-엠포시스부(111), 단구간 푸리에 변환부(113), 크기 제공부(115), 감마톤 필터부(117), 정상 성분 억제부(119), 스펙트럼 재편부(121), 고속 역푸리에 변환부(123), 및 포스트-디엠포시스부(125)를 포함할 수 있다.
- [0023] 프리-엠포시스부(111)는 음성 신호 수신부(170)가 수신한 음성 신호를 프리-엠포시스(Pre-Emphasis)한다.
- [0024] 단구간 푸리에 변환부(113)는 프리-엠포시스된 음성 신호를 주파수 스펙트럼으로 변환한다.
- [0025] 크기 제공부(115)는 변환된 주파수 스펙트럼의 크기를 제공한다.
- [0026] 감마톤 필터부(117)는 크기가 제공된 주파수 스펙트럼을 감마톤 필터링한다.
- [0027] 정상 성분 억제부(119)는 감마톤 필터링한 주파수 스펙트럼의 정상 성분의 파워를 억제한다.
- [0028] 스펙트럼 재편부(121)는 정상 성분 억제된 주파수 스펙트럼을 재편(spectral reshaping)한다.
- [0029] 고속 역푸리에 변환부(123)는 재편된 주파수 스펙트럼을 시간 영역의 음성 신호로 변환한다.
- [0030] 포스트-디엠포시스부(125)는 변환된 음성 신호를 포스트-디엠포시스(Post-Deemphasis)한다.
- [0031] 제어부(110)의 구체적인 동작에 대해서는 도 3을 통해 설명한다.
- [0032] 도 3은 본 발명의 일 실시예에 따른 음성 인식 장치의 동작을 보여주는 흐름도이다.
- [0033] 음성 신호 수신부(170)는 외부로부터 음성 신호를 수신한다(S101). 구체적인 실시예에서 음성 신호 수신부(170)

0)는 하나 또는 복수의 마이크를 포함할 수 있다. 또한, 음성 신호 수신부(170)는 마이크를 통해 음성 신호를 수신할 수 있다. 또 다른 구체적인 실시예에서 음성 신호 수신부(170)는 다른 장치가 인식한 음성 신호를 수신할 수 있다. 구체적으로 음성 신호 수신부(170)는 다른 장치가 인식한 음성 신호를 외부 장치와 통신하는 통신부를 통하여 수신할 수 있다.

[0034] 프리-엠포시스부(111)는 음성 신호 수신부(170)가 수신한 음성 신호를 프리-엠포시스(Pre-Emphasis)한다(S103). 프리-엠포시스부(111)는 음성 신호를 프리-엠포시스하여 고주파 대역의 에너지를 향상한다. 이때, 수신한 음성 신호를 $X[n]$ 이라하면 프리-엠포시스된 신호를 $X_m[n]$ 이라할 수 있다.

[0035] 단구간 푸리에 변환부(113)는 프리-엠포시스된 음성 신호를 단구간 푸리에 변환하여 주파수 스펙트럼으로 변환한다(S105). 이때, 단구간 푸리에 변환부(113)는 프리-엠포시스된 음성 신호에 해밍 윈도우를 곱해주고, 단구간 푸리에 변환하여 주파수 스펙트럼으로 변환할 수 있다. 이때, 생성된 주파수 스펙트럼을 $X[m, k]$ 이라할 수 있다. 이때, m 은 프레임 번호, k 는 주파수 빈(bin). 번호이다. 단구간 푸리에 변환시 시각 프레임은 50ms 길이의 윈도우를 10ms 간격마다 곱하여 생성할 수 있다.

[0036] 크기 제공부(115)는 변환된 주파수 스펙트럼의 크기를 제공한다(S107).

[0037] 감마톤 필터부(117)는 크기가 제공된 주파수 스펙트럼을 감마톤 필터링한다(S109). 이때, 감마톤 필터는 인간의 달팽이관 청각 신경의 주파수 특성을 모사한다. 감마톤 필터의 각 채널별 중심파수를 Equivalent Rectangular Bandwidth(ERB) scale을 따를 수 있다. 감마톤 필터를 통과시킨 감마톤 채널 영역 입력 파워 $P[m, l]$ 은 수학식1과 같이 나타낼 수 있다.

수학식 1

$$P[m, l] = \sum_{k=0}^{N-1} |X[m, k] H_l[k]|^2, \quad 0 \leq l \leq L-1$$

[0038]

[0039] 여기서 N 은 FFT point size이며, 샘플링 주파수(sampling frequency)가 16 KHz이고 50 ms길이의 윈도우를 사용할 경우 N 은 1024일 수 있다. $H_l[k]$ 는 감마톤 필터뱅크의 1번째 채널의 스펙트럼이며 L 은 총 감마톤 채널의 개수로 40 개의 채널을 사용할 수 있다. 도 4는 사용된 감마톤 필터의 주파수 응답을 보여준다.

[0040] 정상 성분 억제부(119)는 감마톤 필터링한 주파수 스펙트럼의 정상 성분 파워를 억제한다(S111). 이때, 정상 성분 억제부(119)는 정상 성분 억제를 위해 저역 통과된 파워 $M[m, l]$ 을 수학식2와 같이 구할 수 있다.

수학식 2

$$M[m, l] = \lambda M[m-1, l] + (1-\lambda) P[m, l]$$

[0041]

[0042] 여기서 λ 는 망각 계수(forgetting factor)로, 저역 통과 필터의 대역폭을 결정한다. 그리고 정상 성분 억제된 파워는 수학식3과 같다.

수학식 3

$$\tilde{P}[m, l] = \max(P[m, l] - M[m, l], c_0 M[m, l])$$

[0043]

[0044] 수학식3에서 c_0 는 정상 성분 억제 수행을 하였을 때 파워가 음수가 되는 것을 방지하기 위한 고정 계수로 억제 목적에 맞게 0.01의 값을 가질 수 있다. 이러한 동작을 통해 음성 신호가 변환된 주파수 스펙트럼은 입력 신호의 파워가 급격히 변하는 시작점에서는 고역 통과된 파워를 가지게 된다. 또한, 음성 신호가 변환된 주파수 스펙트럼은 시작점 이후의 하강 에지에서는 저역 통과된 파워의 상수배로 억제되기 때문에 시작점은 강조되고, 하

강 에지는 평활(smoothed)된다. 이를 통해 음성 인식 장치(100)는 음성 신호의 무향, 반향 환경 신호간의 불일치를 줄여줄 수 있다.

- [0045] 스펙트럼 재편부(121)는 정상 성분 억제된 파워를 통해 주파수 스펙트럼을 재편(spectral reshaping)한다(S113). 구체적으로 스펙트럼 재편부(121)는 정상 성분 억제된 파워 $\tilde{P}[m, l]$ 를 통해 주파수 영역에서의 정상 억제된 스펙트럼 $\tilde{X}[m, k]$ 를 얻기 위하여 스펙트럼 재편할 수 있다. 입력된 원본 스펙트럼 $X[m, k]$ 과 정상 성분 억제된 스펙트럼 $\tilde{X}[m, k]$ 의 위상(phase)은 동일한 경우, 스펙트럼 재편부(121)는 크기(magnitude)만을 수정할 수 있다. 스펙트럼 재편부(121)는 가중 계수 $w[m, l]$ 를 수학식4와 같이 정상 성분 억제된 파워 $\tilde{P}[m, l]$ 와 입력 파워 $P[m, l]$ 와의 비율을 통해 각 프레임, 채널에 대해 구할 수 있다.

수학식 4

$$w[m, l] = \frac{\tilde{P}[m, l]}{P[m, l]}, \quad 0 \leq l \leq L-1$$

[0046]

- [0047] 그리고 스펙트럼 재편부(121)는 감마톤 영역에서 구해진 가중 계수 $w[m, l]$ 를 통해 주파수 영역에서의 가중치 $\mu[m, l]$ 를 수학식5와 같이 연산할 수 있다.

수학식 5

$$\mu[m, k] = \frac{\sum_{l=0}^{L-1} w[m, l] |H_l[k]|}{\sum_{l=0}^{L-1} |H_l[k]|}, \quad 0 \leq k \leq N/2, \quad 0 \leq l \leq L-1$$

[0048]

- [0049] 스펙트럼 재편부(121)는 주파수 영역의 하반부에 대해 정상 성분 억제된 스펙트럼 $\tilde{X}[m, k]$ 을 입력 스펙트럼 $X[m, k]$ 에 주파수 영역 가중치 $\mu[m, l]$ 를 곱하여 수학식6과 같이 연산할 수 있다.

수학식 6

$$\tilde{X}[m, k] = \mu[m, k] X[m, k], \quad 0 \leq k \leq N/2$$

[0050]

- [0051] 스펙트럼 재편부(121)는 $N/2+1 \leq k \leq N$ 영역에 대해서는 Hermitian 대칭을 적용할 수 있다.

- [0052] 고속 역푸리에 변환부(123)는 재편된 주파수 스펙트럼을 시간 영역의 음성 신호로 변환한다(S115). 이때, 변환된 시간 영역의 음성 신호를 $\hat{x}_m[n]$ 이라 할 수 있다.

- [0053] 포스트-디엠피시스부(125)는 변환된 음성 신호를 포스트-디엠피시스(Post-Deemphasis)한다(S117). 포스트-디엠피시스부(125)는 변환된 음성 신호를 포스트-디엠피시스하여 프리-엠피시스부(111)가 음성 신호에 프리-엠피시스한 것을 보상한다.

- [0054] 구체적으로 일반적인 두 귀 신호처리의 표준 모델은 주파수별 두 귀 신호간의 상호 상관함수(interaural cross correlation function) 처리에 이은 비선형 정류(nonlinear rectification)와 대역 통과 필터링(bandpass filtering)의 조합으로 이루어진다. 도 3에서 설명한 단일 채널 정상 성분 억제는 이러한 전통적인 두 귀 신호 처리 모델에 근거하여 복수의 채널로 확장될 수 있다. 이때, 제어부(110)는 정상 성분 억제 과정에서 2개의 음성 인식 센서를 이용하여 얻은 두 신호를 상호 상관 하여 각 단계에 입력한다.

수학식 7

$$P_{B_1}[m, l] = \sum_{k=0}^{N-1} \left| X_L[m, k] X_R^*[m, k] \right\| H_l[k] \right|^2, \quad 0 \leq l \leq L-1$$

[0055]

수학식 8

$$P_{B_2}[m, l] = \sqrt{P_L[m, l] P_R[m, l]}, \quad 0 \leq l \leq L-1$$

[0056]

수학식 9

$$\tilde{P}_{B_3}[m, l] = \sqrt{\tilde{P}_L[m, l] \tilde{P}_R[m, l]}, \quad 0 \leq l \leq L-1$$

[0057]

수학식 10

$$\tilde{x}[n] = \tilde{x}_L[n] + \tilde{x}_R[n]$$

[0058]

[0059] 수학식7은 스펙트럼 분석에서의 조합, 수학식8은 감마톤 채널 영역에서 감마톤 채널 파워의 조합, 수학식9는 감마톤 채널영역에서 정상 성분 억제 처리된 파워의 조합이며, 수학식10은 최종적으로 정상 성분 억제 처리 후 시간영역으로 복원된 각 신호를 지연 및 합산 빔포밍(delay-and-sum beamforming; DSBF)을 수행하여 얻은 조합이다. 제어부(110)는 이러한 값을 도 3에 적용하여 복수의 채널 정상 성분을 억제할 수 있다.

[0060] 정상 성분 억제 방법의 처리된 신호는 기본적으로 입력된 감마톤 채널 영역 파워 $P[m, l]$ 에서 저역 통과 성분 $M[m, l]$ 을 차감하는 방식이다. 따라서 설정된 바닥값인 $c_0 M[m, l]$ 으로 정상 성분이 억제됨으로써 이루어진다. 그러나 반향이 심해짐에 따라 느린 반향 성분이 지속적으로 추가됨으로써 $M[m, l]$ 이 점점 큰 값을 가지게 되며, 바닥값인 $c_0 M[m, l]$ 가 상승하기 때문에 하강 에지가 억제되지 않고 길게 늘어지는 현상이 생긴다. 이러한 환경간에 처리된 신호의 불일치는 학습된 음향모델과 실제 반향 환경에서 처리된 신호간의 불일치를 발생시켜 음성인식률에 악영향을 끼친다.

[0061] 또한, 무성음은 고주파 영역에서 대부분의 파워를 가지지만, 저주파 영역에서도 약간의 파워를 가지며 파열음의 경우에는 전 주파수 대역에서 파워를 가진다. 그러나 유성음보다는 상대적으로 낮은 파워를 가지기 때문에 정상 성분 억제 처리되었을 때 낮은 주파수 영역에서 그 특징이 많이 제거된다. 특히, 반향이 심해지는 환경에서는 이전 시간에 발생되었던 유성음의 반향 성분들이 억제되지 않아 무성음의 저주파 영역을 침범하여 왜곡시킨다. 이러한 현상은 무향 환경에서의 인식률도 하락시킬 뿐만 아니라 반향 환경에서의 인식률도 크게 하락시키는 원인이 된다. 따라서 이러한 문제점을 극복할 수 있는 음성 신호의 정상 성분 억제 방법이 필요하다. 이에 대해서는 다음의 도면들을 통해 설명한다.

[0062] 도 5는 본 발명의 또 다른 실시예에 따른 음성 인식 장치가 포함하는 제어부의 구성을 보여준다.

[0063] 본 발명의 또 다른 실시예에 따른 음성 인식 장치(100)가 포함하는 제어부(110)는 프리-엠폰시스부(131), 제1 단구간 푸리에 변환부(133), 제1 크기 제공부(135), 감마톤 필터부(137), 정상 성분 억제부(139), 스펙트럼 재편부(141), 고속 역푸리에 변환부(143), 포스트-디엠폰시스부(145), 제1 평균 파워 차감부(147), 채널 파워 비율 연산부(149), 제2 단구간 푸리에 변환부(153), 제2 크기 제공부(155), 기본 주파수 검출부(157), 제2 평균 파워 차감부(159), 및 하모닉 파워 비율 연산부(161)를 포함한다.

- [0064] 프리-엠포시스부(131)는 음성 신호 수신부(170)가 수신한 음성 신호를 프리-엠포시스(Pre-Emphasis)한다.
- [0065] 제1 단구간 푸리에 변환부(113)와 제2 단구간 푸리에 변환부(153)는 프리-엠포시스된 음성 신호를 주파수 스펙트럼으로 변환한다. 구체적인 실시예에 따라서 제1 단구간 푸리에 변환부(113)와 제2 단구간 푸리에 변환부(153)는 하나의 구성으로 통합될 수 있다.
- [0066] 제1 크기 제공부(135)와 제2 크기 제공부(155)는 변환된 주파수 스펙트럼의 크기를 제공한다. 구체적인 실시예에 따라서 제1 크기 제공부(135)와 제2 크기 제공부(155)는 하나의 구성으로 통합될 수 있다.
- [0067] 감마톤 필터부(137)는 크기가 제공된 주파수 스펙트럼을 감마톤 필터링한다.
- [0068] 정상 성분 억제부(139)는 감마톤 필터링한 주파수 스펙트럼의 정상 성분의 파워를 억제한다. 구체적으로 정상 성분 억제부(139)는 하모닉 파워 비율 및 채널 파워 비율 중 적어도 어느 하나에 기초하여 감마톤 필터링된 스펙트럼을 정상 성분 억제할 수 있다.
- [0069] 스펙트럼 재편부(141)는 정상 성분 억제된 주파수 스펙트럼을 재편(spectral reshaping)한다.
- [0070] 고속 역푸리에 변환부(143)는 재편된 주파수 스펙트럼을 시간 영역의 음성 신호로 변환한다.
- [0071] 포스트-디엠포시스부(145)는 변환된 음성 신호를 포스트-디엠포시스(Post-Deemphasis)한다.
- [0072] 제1 평균 파워 차감부(147)는 감마톤 채널 영역에서 평균 파워 차감을 수행한다.
- [0073] 채널 파워 비율 연산부(149)는 감마톤 채널 영역에서의 평균 파워 차감 후의 파워에 기초하여 채널 파워 비율을 연산한다. 구체적으로 채널 파워 비율은 전체 감마톤 채널의 파워 대비 고주파 영역에 해당하는 감마톤 채널들의 파워 비율로 정의할 수 있다.
- [0074] 기본 주파수 검출부(157)는 하모닉 성분들의 파워를 측정하기 위하여 입력된 음성신호에서 기본 주파수를 검출한다. 구체적으로 기본 주파수 검출부(157)는 자기상관함수를 기반으로 한 방법(autocorrelation method) 및 하모닉 지배도(harmonic dominance)를 기반으로 한 방법 중 어느 하나에 기초하여 기본 주파수를 검출할 수 있다. 기본 주파수 검출부(157)는 자기상관함수를 기반으로 한 방법에 기초하여 비터비(Viterbi) 탐색이 제거된 방법으로 기본 주파수를 검출할 수 있다.
- [0075] 제2 평균 파워 차감부(159)는 주파수 영역에서의 평균 파워를 차감한다. 구체적으로 제2 평균 파워 차감부(159)는 기본주파수를 구한 후 하모닉 파워 비율을 구하기 전, 주파수 영역에서의 평균 파워를 차감할 수 있다.
- [0076] 하모닉 파워 비율 연산부(161)는 주파수 스펙트럼의 하모닉 파워 비율을 연산한다. 구체적으로 하모닉 파워 비율은 전체 주파수 대역의 파워 대비 하모닉 성분들의 파워 비율로 정의할 수 있다.
- [0077] 제어부(110)의 구체적인 동작은 도 6을 통해 설명한다.
- [0078] 도 6은 본 발명의 또 다른 실시예에 따른 음성 인식 장치의 동작을 보여주는 흐름도이다.
- [0079] 음성 신호 수신부(170)는 외부로부터 음성 신호를 수신한다(S301). 구체적인 동작은 도 3의 음성 신호 수신 단계(S101)와 동일할 수 있다.
- [0080] 프리-엠포시스부(131)는 음성 신호 수신부(170)가 수신한 음성 신호를 프리-엠포시스(Pre-Emphasis)한다(S303). 구체적인 동작은 도 3의 프리-엠포시스 단계(S103)와 동일할 수 있다.
- [0081] 제1 단구간 푸리에 변환부(133)와 제2 단구간 푸리에 변환부(135)는 프리-엠포시스된 음성 신호를 단구간 푸리에 변환하여 주파수 스펙트럼으로 변환한다(S305). 이때, 제2 단구간 푸리에 변환부(135)는 하모닉 파워 비율 연산을 위해 음성 신호를 단구간 푸리에 변환하여 주파수 스펙트럼으로 변환할 수 있다. 제1 단구간 푸리에 변환부(133)와 제2 단구간 푸리에 변환부(135)의 구체적인 동작은 도 3의 단구간 푸리에 변환 단계(S105)와 동일할 수 있다. 또한, 앞서 설명한 바와 같이 제1 단구간 푸리에 변환부(133)와 제2 단구간 푸리에 변환부(135)는 하나의 구성으로 통합될 수 있다.
- [0082] 제1 크기 제공부(135)와 제2 크기 제공부(155)는 변환된 주파수 스펙트럼의 크기를 제공한다(S307). 구체적인 동작은 도 3의 주파수 스펙트럼의 크기를 제공 단계(S107)와 동일할 수 있다. 또한, 앞서 설명한 바와 같이 제1 크기 제공부(135)와 제2 크기 제공부(155)는 하나의 구성으로 통합될 수 있다.

- [0083] 감마톤 필터부(337)는 크기가 제공된 주파수 스펙트럼을 감마톤 필터링한다(S309). 구체적인 동작은 도 3의 주파수 스펙트럼의 감마톤 필터링 단계(S109)와 동일할 수 있다.
- [0084] 하모닉 파워 비율 연산부(161)는 하모닉 파워 비율을 획득한다(S311).
- [0085] 채널 파워 비율 연산부(149)는 채널 파워 비율을 획득한다(S313).
- [0086] 정상 성분 억제부(139)는 하모닉 파워 비율 및 채널 파워 비율 중 적어도 어느 하나에 기초하여 감마톤 필터링된 스펙트럼을 정상 성분 억제한다(S315).
- [0087] 하모닉 파워 비율 연산부(161)가 하모닉 파워 비율을 연산하여 정상 성분 억제부(139) 하모닉 파워 비율에 기초하여 감마톤 필터링된 스펙트럼을 정상 성분을 억제하는 동작에 대해서는 도 7를 통해 설명한다.
- [0088] 도 7은 음성 인식 장치의 제어부가 하모닉 파워 비율을 획득하는 동작을 보여주는 흐름도이다.
- [0089] 기본 주파수 검출부(157)는 주파수 스펙트럼의 기본 주파수를 검출한다(S501). 구체적으로 기본 주파수 검출부(157)는 자기상관함수를 기반으로 한 방법 및 하모닉 지배도를 기반으로 한 방법 중 어느 하나에 기초하여 기본 주파수를 검출할 수 있다. 기본 주파수 검출부(157)는 자기상관함수를 기반으로 한 방법에 기초하여 비터비(Viterbi) 탐색이 제거된 방법으로 기본 주파수를 검출할 수 있다. 이때, 기본 주파수 검출부(157)는 먼저 다음과 같이 일반적인 인간이 낼 수 있는 최소, 최대 기본주파수 $F_{0,min}$, $F_{0,max}$ 를 수학식11과 같이 설정하고 그에 따른 최소, 최대 시간지연 $\tau_{0,min}$, $\tau_{0,max}$ 을 수학식12와 같이 구할 수 있다.

수학식 11

[0090]
$$F_{0,min} = 70Hz, \quad F_{0,max} = 400Hz$$

수학식 12

[0091]
$$\tau_{0,min} = \text{round}(F_s / F_{0,min}), \quad \tau_{0,max} = \text{round}(F_s / F_{0,max})$$

- [0092] 그리고 기본 주파수 검출부(157)는 자기상관함수 $r_{xx}[m, \tau]$ 을 수학식13과 같이 구하며, $r_{xx}[m, \tau]$ 가 최대가 되는 시간지연 $\tau_0[m]$ 을 최소, 최대 시간지연 범위 T내에서 수학식14와 같이 구할 수 있다.

수학식 13

[0093]
$$r_{xx}[m, \tau] = \sum_{k=0}^{N-1} X[m, k] X^*[m, k] e^{j2\pi k \tau / N}$$

수학식 14

[0094]
$$\tau_0[m] = \arg \max_{\tau \in T} r_{xx}[m, \tau], \quad T = \{\tau_{0,min}, \tau_{0,min} + 1, \dots, \tau_{0,max} - 1, \tau_{0,max}\}$$

- [0095] 그리고 기본 주파수 검출부(157)는 구해진 최대 시간지연 $\tau_0[m]$ 을 통해 기본주파수 추정값 $\hat{F}_0[m]$ 을 수학식15와 같이 구할 수 있다.

수학식 15

$$\hat{F}_0[m] = \frac{F_s}{\tau_0[m]}$$

[0096]

[0097] 기본주파수 추정값 $\hat{F}_0[m]$ 은 프레임에 따라 불안정한 값으로 급격히 변할 수도 있기 때문에 기본 주파수 검출부 (157)는 수학식16과 같이 앞, 뒤 프레임들과의 평균을 통해 $F_0[m]$ 을 구할 수 있다. δ_h 는 평균을 낼 때 사용하고자 하는 앞, 뒤 프레임 수를 의미하며 실험을 통해 1로 정할 수 있다.

수학식 16

$$F_0[m] = \frac{1}{2\delta_h + 1} \sum_{\delta=-\delta_h}^{\delta_h} \hat{F}_0[m + \delta]$$

[0098]

[0099] 제2 평균 파워 차감부(159)는 주파수 영역에서의 주파수 스펙트럼에 평균 파워를 차감한다(S503). 무향 환경에서의 입력신호는 현재 발음되고 있는 음성만이 입력되지만 반향 환경에서의 현재 시간에 입력되고 있는 신호는 현재 발음되는 음성과 그 이전 시간에 발음되었던 음성의 반향 성분이 함께 입력되게 된다. 특히 이전 시간의 발음에 존재하는 반향 성분 중 큰 파워를 가지고 있는 유성음의 하모닉 성분들이 다시 현재 프레임에 반복되어 나타나며, 파워가 모두 감쇄되기 전까지 지속적으로 나타나기 때문에 긴 꼬리 형태로 남게 된다. 이에 대해서는 도 8을 참조하여 구체적으로 설명한다.

[0100] 도 8은 무향 및 반향 환경에서 평균 파워 차감이 적용되지 않은 시간-주파수 영역에서의 입력 신호 파워와 하모닉 파워 비율을 보여준다.

[0101] 도 8(a)에서 보여지는 무향 환경에서의 시간-주파수 영역 파워를 살펴봤을 때, 발화 음성만 존재하기 때문에 하모닉 구조가 깨끗하게 표현되고 유성음과 무성음의 구별이 뚜렷하게 나타난다. 그러나 도 8(b)에서 보여지는 반향 환경에서는 이전 발화 음성에서의 하모닉 성분들이 감쇄되지 않은 상태로 남아 하모닉 구조 사이에 파워가 존재하여 깨끗하게 보이지 않는다. 따라서 반향 성분이 감쇄되지 않아 긴 꼬리 형태로 무성음의 저주파 영역을 침범하게 된다. 이러한 반향 성분들로 인해 생기는 왜곡은 하모닉 파워 비율이 무향 환경과 반향 환경에서 차이를 보이게 되는 원인이 된다. 무향 환경에서 첫 번째 유성음에서 무성음으로 넘어가는 구간의 시간-주파수 영역을 살펴보면 발화가 없는 짧은 정지 구간이지만 하강 에지 부근에서 전체적인 주파수 영역의 파워가 하모닉 파워보다 작아지게 된다. 때문에 하모닉 파워 비율이 오버샷 하는 것을 확인할 수 있다. 두 번째 유성음에서 무성음으로 넘어가는 구간에서도 하강 에지 부근에서 비슷한 오버샷이 발생하는 것을 확인할 수 있다. 반향 환경에서의 첫 번째 유성음에서 무성음으로 넘어가는 구간의 시간-주파수 영역을 살펴보면 반향 성분으로 인해 하모닉 성분들이 긴 꼬리를 가지게 되며, 하모닉 파워 비율도 무성음에서 큰 값을 가져 정상 성분 억제에 위한 가중치로 사용하였을 때 반향 성분들을 억제하기 힘들어진다. 두 번째 영역에서는 반향 성분들로 인해 하모닉 성분 이외의 주파수 영역에서 생긴 파워로 인해 무향 환경에 비해 하모닉 파워 비율이 전체적으로 줄어들며 첫 번째 영역과 마찬가지로 유성음의 반향 성분들이 무성음을 침범하여 무성음 영역에서도 큰 하모닉 파워 비율을 가지는 문제점을 보인다. 따라서 이러한 문제점을 해결하기 위하여 제어부(110)는 일반적인 스펙트럼 차감(spectral subtraction) 방법에 근거하여 현재 시간에 입력되고 있는 이전 발화의 반향 성분의 파워를 제거할 수 있다. 이때, 스펙트럼 차감 방법은 잡음 환경에서 입력된 음성 신호에서 잡음을 제거하기 위해 비음성 구간에서 잡음의 파워 레벨을 추정하고 잡음으로 인한 파워 상승을 차감하는 방법이다. 이와 마찬가지로 제2 파워 차감부(159)는 감쇄되지 않은 이전 발화의 반향 성분의 파워로 인해 파워 상승이 생겨나는 것을 제거하기 위하여 이전 발화에 대한 반향 성분의 파워를 추정하여 현재 시간에 입력된 음성의 파워에서 차감할 수 있다. 제2 파워 차감부(15

9)는 이전 발화에 대한 반향 성분 파워 추정값은 수학식17과 같이 이전 δ_m 개 프레임에 대한 파워의 평균으로 구할 수 있다.

수학식 17

$$B_f[m, k] = \frac{1}{\delta_m} \sum_{\delta=1}^{\delta_m} |X[m - \delta, k]|^2$$

[0102]

[0103] δ_m 은 반향 환경의 정도에 따른 실험을 통해 10으로 결정할 수 있다. 그리고 제2 파워 차감부(159)는 구해진 반향 성분의 파워 추정 값을 수학식18과 같이 입력신호의 파워에서 차감할 수 있다.

수학식 18

$$P_{fn}[m, k] = \max(|X[m, k]|^2 - B_f[m, k], \varepsilon_f)$$

[0104]

[0105] 수학식17에서 ε_f 는 파워가 음수가 되지 않도록 하는 작은 상수값이며, 주파수 영역에서 입력신호의 동적 범위(dynamic range)를 고려하여 1.0×10^{-3} 의 값으로 설정할 수 있다.

[0106]

도 9는 무향 환경, 반향 환경에서 각각 평균 파워 차감을 하였을 때, 각각의 시간-주파수 영역 파워와 하모닉 파워 비율을 보여준다. 도 9(a)에서 보여주는 무향 환경에서의 시간-주파수 영역을 살펴보면 유성음에서 무성음으로 넘어가는 첫 번째 및 두 번째 구간에서의 짧은 정지 구간이 평균 파워 차감을 통해 전체적인 파워가 제거되어 확연히 구분 가능해졌으며, 반향 환경에서는 무성음으로 침범하는 유성음의 반향 성분의 파워가 크게 제거되었음을 확인할 수 있다. 무향 환경에서의 하모닉 파워 비율을 살펴보면 평균 파워 차감 전에 발생하였던 오버슛 현상이 사라지고 하강 에지에서 서서히 줄어드는 것을 확인할 수 있다. 또한, 도 9(b)에서 보여주는 반향 환경에서도 마찬가지로 오버슛 현상이 완화되고 유성음의 반향 성분이 무성음으로 침범함으로써 무성음에서 하모닉 파워가 커지는 현상도 크게 줄어들었음을 확인할 수 있다. 따라서 음성 인식 장치(100)는 평균 파워 차감을 통해 하모닉 파워 비율이 환경에 따른 불일치를 겪는 문제를 크게 해결할 수 있다. 또한, 음성 인식 장치(100)는 유성음의 초기 시작부분에서 하모닉 파워 비율이 높고 하강 에지에서 서서히 줄어들 수 있다. 또한, 하모닉 파워 비율의 특성은 정상 성분 억제 방법이 음성의 초기 시작 부분을 강조하고 정상 성분을 억제하는 동작 특성과 일치하기 때문에 정상 성분 억제를 위한 가중치로 사용할 수 있다.

[0107]

다시 도 7로 돌아가 하모닉 파워 비율을 획득하는 동작을 설명한다.

[0108]

하모닉 파워 비율 연산부(161)는 평균 파워 차감된 주파수 스펙트럼에서 기본 주파수에 기초하여 하모닉 파워 비율을 연산한다(S505). 이때, 하모닉 파워 비율은 전체 주파수 대역의 파워 대비 하모닉 성분들의 파워 비율로 정의할 수 있다.

[0109]

하모닉 파워 비율 연산부(161)는 수학식18과 같이 하모닉 파워 비율의 추정값 $\hat{\zeta}_h[m]$ 을 계산할 수 있다. 하모닉 파워 비율을 계산 시 환경간의 불일치를 해결하기 위해 하모닉 파워 비율 연산부(161)는 평균 파워 차감 후의 파워 $P_{fn}[m, k]$ 을 이용하여 전체 주파수 대역 파워와 하모닉 성분들의 파워를 구할 수 있다. 그리고 수학식19와 같이 프레임에 따라 불안정한 값으로의 급격한 변화를 방지하기 위해 하모닉 파워 비율 연산부(161)는 구해진 하모닉 파워 비율의 추정값을 앞, 뒤 프레임에서의 평균을 수행하여 하모닉 파워 비율 $\hat{\zeta}_h[m]$ 을 구할 수 있다.

수학식 19

$$\hat{\zeta}_h[m] = \frac{\sum_{\kappa=1}^{\kappa_{\max}} \sum_{\gamma=-\gamma_k}^{\gamma_k} P_{fn}[m, \rho(\kappa) + \delta_k + \gamma]}{\sum_{k=0}^{N/2} P_{fn}[m, k]}$$

[0110]

수학식 20

$$\zeta_h[m] = \frac{1}{2\delta_h + 1} \sum_{\delta=-\delta_h}^{\delta_h} \hat{\zeta}_h[m + \delta]$$

[0111]

[0112] 수학식18에서 하모닉 성분들의 파워는 기본주파수 $F_0[m]$ 에서의 파워부터 $F_0[m]$ 의 정수배에 위치한 하모닉 성분들의 파워로 구성되는데, 하모닉 파워 비율 연산부(161)는 하모닉 성분들이 지배적인 주파수 대역인 $F_s/4 = 4000\text{Hz}$ 대역까지의 하모닉 성분들에 대한 파워를 구할 수 있다. 수학식20의 $K_{\max}$ 는 m번째 프레임에서 $F_s/4$ 범위 안에 존재하는 하모닉 성분의 개수이다.

수학식 21

$$\kappa_{\max} = \text{floor}((F_s/4)/F_0[m])$$

[0113]

[0114] 수학식18에서 하모닉 파워 비율 연산부(161)는 $\rho(\kappa)$ 는 기본주파수 $F_0[m]$ 의 k번째 하모닉 성분의 주파수 빈 번호를 계산할 수 있고, 수학식22와 같다.

수학식 22

$$\rho(\kappa) = \text{round}(\kappa \cdot F_0[m] \cdot N / F_s)$$

[0115]

[0116] 그리고 한 프레임 내에서의 발음 변화로 인해 하모닉 성분들이 정확히 기본주파수의 정수배에 위치하지 않는 경우가 존재하기 때문에 하모닉 파워 비율 연산부(161)는 Δ 범위 내에 존재하는 주파수 빈 중 파워가 최대가 되는 주파수 빈으로 보정을 하는 보정 상수 δ_k 를 수학식22와 같이 계산할 수 있다.

수학식 23

$$\delta_k = \arg \max_{\delta \in \Delta} P_{jm}[m, \rho(\kappa) + \delta],$$

$$\Delta = \{-\delta_{k,\max}, -\delta_{k,\max} + 1, \dots, -1, 0, 1, \dots, \delta_{k,\max} - 1, \delta_{k,\max}\}$$

[0117]

[0118] 보정 범위 Δ 에서 주변 주파수를 보는 범위는 $\delta_{k,\max}$ 를 통해 정해지며, $\delta_{k,\max}$ 는 하모닉 파워 비율 연산부(161)는 최소 기본주파수 $F_{0,\min}$ 일 때의 하모닉 성분들간의 거리의 절반 범위로 정의하여 수학식23과 같이 구할 수 있으며 기본주파수의 정수배 앞, 뒤 주파수 빈에서 하모닉 성분의 파워를 찾을 수 있다.

수학식 24

$$\delta_{k,\max} = \text{round}(N \cdot F_{0,\min} / F_s)$$

[0119]

[0120] 일반적으로 정상 성분 억제 방법은 50ms의 윈도우에서 최적의 성능을 나타내며, 이 경우 전체 주파수 대역의 파워에서 하모닉 성분들이 차지하는 파워의 비율은 최대 60%정도의 값을 보인다. 그러나 더 큰 윈도우 길이를 사용하게 되면 주파수 영역의 해상도가 높아지게 되기 때문에 하모닉 성분을 각 정수배마다 1개의 파워만을 사용하게 되면 하모닉 파워 비율의 동적 범위가 달라지게 된다. 따라서 하모닉 파워 비율 연산부(161)는 윈도우의

길이 길어질 때에는 각 정수배 위치에서 γ_k 개의 앞, 뒤 주파수 bin의 파워까지 더해줄 수 있다. 하모닉 파워 비율 연산부(161)는 100ms 미만의 윈도우를 사용하게 될 경우에 $\gamma_k = 0$, 그 이상의 윈도우의 경우 $\gamma_k \geq 1$ 을 사용할 수 있다.

[0121] 수학식19를 통해 구해진 하모닉 파워 비율을 정상 성분 억제 방법에 도입하기 위해 정상 성분 억제부(139)는 수학식25와 같이 정상 성분 억제된 파워 수학식3의 바닥값의 계수로 적용할 수 있다.

수학식 25

$$\tilde{P}[m, l] = \begin{cases} \max(P[m, l] - M[m, l], c_0 M[m, l]), & l > l_{th} \\ \max(P[m, l] - M[m, l], \max(c_h \zeta_h[m], c_0) M[m, l]), & l \leq l_{th} \end{cases}$$

[0122]

[0123] 여기서 c_h 는 하모닉 파워 비율의 동적 범위를 맞추기 위한 고정 상수이며 c_0 는 기존 정상 성분 억제 방법에서의 고정 상수이다.

[0124] 하모닉 파워 비율을 바닥값에 적용하게 되면 하강 에지의 억제를 완화시키기 때문에 전체 감마톤 채널 영역의 하강 에지가 길게 늘어난다. 그러나 실제 반향 환경에서 길게 늘어지는 하강 에지는 큰 파워를 가지는 유성음의 저주파 대역이기 때문에 모든 채널에 하모닉 파워 비율을 적용하게 되면 음성인식률의 향상이 적어진다. 따라서 제어부(110)는 정상 성분 억제부(139)를 통하여 유성음이 지배적인 저주파 대역에 해당하는 감마톤 채널들에 대해서만 선택적으로 하모닉 파워를 적용하며, 고주파 대역에 해당하는 채널들은 기존 정상 성분 억제 방법과 같이 고정 상수 c_0 로 바닥값을 설정하여 억제할 수 있다. 이때, 저주파 대역은 미리 정하여진 일정한 기준 주파수보다 낮은 주파수를 나타낼 수 있다. 하모닉 파워 비율 연산부(161)는 수학식24에서 l_{th} 를 기준으로 하모닉 파워 비율을 적용하는 채널 구간을 설정할 수 있다.

[0125] 채널 파워 비율 연산부(149)가 채널 파워 비율을 연산하여 정상 성분 억제부(139)가 채널 파워 비율에 기초하여 감마톤 필터링된 스펙트럼을 정상 성분을 억제하는 동작에 대해서는 도 10을 통해 설명한다.

[0126] 도 10은 음성 인식 장치의 제어부가 채널 파워 비율을 획득하는 동작을 보여주는 흐름도이다.

[0127] 제1 평균 파워 차감부(147)는 평균 파워를 차감한다(S701). 감마톤 채널 영역에서 평균 파워 차감을 수행한다. 비록 무성음의 고주파 반향 성분은 파워가 적기 때문에 유성음에 비해 크게 늘어지지 않고 음성인식률에도 영향을 미치지 않는 성분이라고 알려져 있지만, 분명히 반향 환경에서 파워가 존재하는 성분이기 때문에 채널 파워 비율이 환경간의 불일치를 겪는다. 따라서 제1 평균 파워 차감부(147)는 평균 파워를 차감하여 채널 파워 비율의 환경에 대한 불일치를 완화할 수 있다. 수학식25는 이전 발화에 대한 반향 성분의 감마톤 채널 파워 추정값이며 제1 평균 파워 차감부(147)는 이전 δ_m 개 프레임에 대한 감마톤 채널 파워의 평균으로 구할 수 있다.

수학식 26

$$B_g[m, l] = \frac{1}{\delta_m} \sum_{\delta=1}^{\delta_m} P[m - \delta, l]$$

[0128]

수학식 27

$$P_{gn}[m, l] = \max(P[m, l] - B_g[m, l], \varepsilon_g)$$

[0129]

[0130] 그리고 수학식25를 통해 제1 평균 파워 차감부(147)는 구해진 반향 성분의 감마톤 채널 파워 추정값 $B_g[m, l]$ 을 수학식26과 같이 감마톤 채널 입력신호의 파워에서 차감해줄 수 있다. 제1 평균 파워 차감부(147)는 ε_g 는 동적

범위를 고려하여 1.0×10^{-6} 으로 설정할 수 있다.

[0131] 도 11은 무향 및 반향 환경에서 평균 파워 차감을 하지 않았을 때의 감마톤 채널 영역 입력 파워와 채널 파워 비율을 보여준다. 도 11(a)에서 보여지는 무향 환경의 경우 무성음 구간에 대해 채널 파워 비율이 커짐으로써 목표하는 역할을 수행하고는 있지만 유성음-무성음-유성음으로의 상태천이가 이루어 질 때 무성음 구간이 정밀하지 않은 모습을 보인다. 도 11(b)에서 보여지는 반향 환경에서는 반향 성분으로 인해 중간 유성음 구간에서도 채널 파워 비율이 높아지며, 특히 중간 무성음-유성음-무성음의 상태천이 구간은 반향 성분들로 인해 유성음인 구간에서도 채널 파워 비율이 매우 높아져 무성음 구간이 정밀하지 않은 환경간의 불일치가 발생하는 것을 알 수 있다. 도 12는 무향 및 반향 환경에서 평균 파워 차감을 하였을 때의 감마톤 채널 영역 입력 파워와 채널 파워 비율을 보여준다. 도 12(a)에서 보여지는 무향 환경에서는 좀 더 정밀한 무성음 구간에서 채널 파워 비율이 상승하며, 도 12(b)의 반향 환경에서도 무성음-유성음-무성음 구간의 사이에 생겨난 반향 성분들이 제거되어 무성음 구간이 확연히 구분됨으로써 환경간의 불일치를 해결할 수 있음을 보여준다.

[0132] 채널 파워 비율 연산부(149)는 차감된 평균 파워를 가지는 주파수 스펙트럼에서 채널 파워 비율을 획득한다 (S703). 이때, 채널 파워 비율은 전체 감마톤 채널의 파워 대비 고주파 영역에 해당하는 감마톤 채널들의 파워 비율로 정의할 수 있다. 이때, 채널 파워 비율 연산부(149)는 채널 파워 비율을 수학식28과 같이 계산할 수 있다.

수학식 28

$$\zeta_c[m] = \frac{\sum_{l=l_u}^{L-1} P_{gn}[m, l]}{\sum_{l=0}^{L-1} P_{gn}[m, l]}$$

[0133]

[0134] 채널 파워 비율을 계산 시 환경간의 불일치를 해결하기 위해 채널 파워 비율 연산부(149)는 감마톤 채널 영역에서의 평균 파워 차감 후의 파워 $P_{gm}[m, k]$ 을 사용할 수 있다. 수학식27에서 l_u 는 고주파 영역에 해당하는 감마톤 채널의 범위를 설정할 수 있다. 구체적으로 채널 파워 비율 연산부(149)는 일정 기준 값 이상의 주파수 영역을 감마톤 채널 범위로 설정할 수 있다. 대부분의 무성음의 고주파 영역 특징은 34번째 이상의 채널에서 지배적으로 나타나기 때문에 채널 파워 비율 연산부(149)는 $l_u = 34$ 로 정할 수 있다.

[0135] 기존 정상 성분 억제 처리에서 무성음의 특징을 살리기 위해서는 반향 환경에서 상승하는 $M[m, l]$ 의 영향력을 줄여야 하므로 정상 성분 억제부(139)는 수학식28과 같이 고역 통과된 파워 $P[m, l] - M[m, l]$ 연산에서 $M[m, l]$ 의 가중치로 채널 파워 비율을 적용할 수 있다.

수학식 29

$$\tilde{P}[m, l] = \begin{cases} \max(P[m, l] - (1 - c_c \zeta_c[m])M[m, l], c_0 M[m, l]), & l > l_{th} \\ \max(P[m, l] - (1 - c_c \zeta_c[m])M[m, l], \max(c_h \zeta_h[m], c_0)M[m, l]), & l \leq l_{th} \end{cases}$$

[0136]

[0137] 채널 파워 비율은 무성음일 때 큰 값, 유성음일 때 작은 값을 나타내므로 채널 파워 비율 연산부(149)는 $1 - c_c \zeta_c[m]$ 형태로 가중할 수 있다. 이때, C_c 는 채널 파워 비율의 동적 범위를 조절하는 고정 상수일 수 있다. 하모닉 파워 비율의 경우 유성음에 해당하는 감마톤 채널에 대해 적용하였다. 그러나 채널 파워 비율의 경우에는 무성음의 전체 채널에 대한 특징을 향상시키는 것이므로 채널 파워 비율 연산부(149)는 모든 채널에 대한 고역 통과된 파워 $P[m, l] - M[m, l]$ 의 연산에서 $M[m, l]$ 의 가중치로 적용할 수 있다.

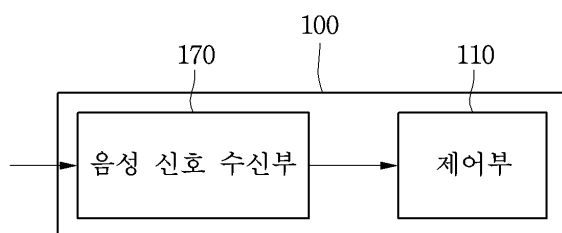
[0138] 도 4 내지 도 10을 통해 설명한 방법은 도 2 내지 도 3을 통해 설명한 정상 성분 억제 방법에 비해 유성음의 하강 에지를 확장시키고 무성음의 특징을 향상시킴으로써 환경간의 차이를 완화시킬 수 있다. 다시 도 6으로 돌아가 음성 인식 장치(100)의 동작을 설명한다.

[0139] 스펙트럼 재편부(141)는 정상 성분 억제된 파워를 통해 주파수 스펙트럼을 재편(spectral reshaping)한다 (S317). 스펙트럼 재편부(141)의 구체적인 동작은 도 3의 스펙트럼 재편 단계(S113)와 같을 수 있다.

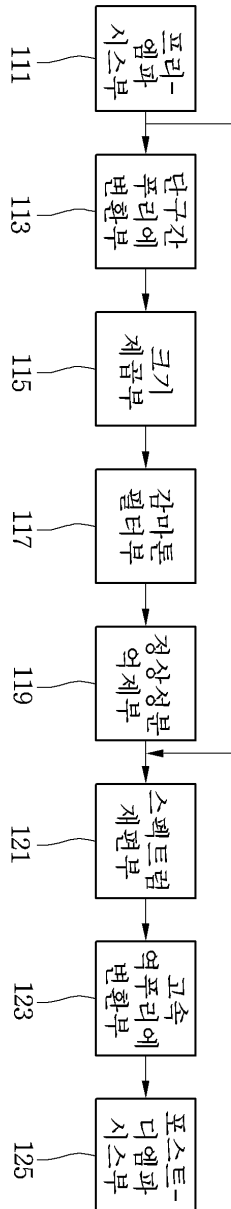
- [0140] 고속 역푸리에 변환부(143)는 재편된 주파수 스펙트럼을 시간 영역의 음성 신호로 변환한다(S319). 고속 역푸리에 변환부(143)의 구체적인 동작은 도 3의 재편된 주파수 스펙트럼을 시간 영역의 음성 신호로 변환하는 단계(S115)와 같을 수 있다.
- [0141] 포스트-디엠퍼시스부(145)는 변환된 음성 신호를 포스트-디엠퍼시스(Post-Deemphasis)한다(S321). 포스트-디엠퍼시스부(145)는 변환된 음성 신호를 포스트-디엠퍼시스하여 프리-엠퍼시스부(131)가 음성 신호에 프리-엠퍼시스한 것을 보상한다.
- [0142] 앞서 설명한 바와 같이 일반적인 두 귀 신호처리의 표준 모델은 주파수별 두 귀 신호간의 상호 상관함수(interaural cross correlation function) 처리에 이은 비선형 정류(nonlinear rectification)와 대역 통과 필터링(bandpass filtering)의 조합으로 이루어진다. 도 6 내지 도 10에서 설명한 단일 채널 정상 성분 억제에 이러한 전통적인 두 귀 신호처리 모델에 근거하여 복수의 채널로 확장될 수 있다. 이때, 제어부(110)는 정상 성분 억제 과정에서 2개의 마이크를 이용하여 얻은 두 신호를 상호 상관 하여 각 단계에 입력할 수 있다. 구체적으로 제어부(110)는 앞서 설명한 수학적 7 내지 수학적 10을 도 6 내지 도 10에서 설명한 단일 채널 정상 성분 억제에 적용하여 복수의 채널로 확장할 수 있다.
- [0143] 또한, 제어부(110)는 전처리 방법을 통해 음성 신호로부터 추출한 음성 신호의 특징을 나타내는 특징 벡터와의 가중 합을 통해 음성 인식의 정확도를 향상할 수 있다. 특히, 가산 잡음 환경과 무향 환경에서의 음성 인식 정확도를 향상할 수 있다. 구체적으로 제어부(110)는 1채널 음성 인식 센서 환경에서 바이패스 필터를 사용할 수 있다. 또한, 제어부(110)는 2채널 이상 음성 인식 센서 환경에서 합산 빔포밍 처리를 사용할 수 있다. 이때, 제어부(110)는 음원의 위치를 아는 경우일 수 있다.
- [0144] 이상에서 실시예들에 설명된 특징, 구조, 효과 등은 본 발명의 적어도 하나의 실시예에 포함되며, 반드시 하나의 실시예에만 한정되는 것은 아니다. 나아가, 각 실시예에서 예시된 특징, 구조, 효과 등은 실시예들이 속하는 분야의 통상의 지식을 가지는 자에 의해 다른 실시예들에 대해서도 조합 또는 변형되어 실시 가능하다. 따라서 이러한 조합과 변형에 관계된 내용들은 본 발명의 범위에 포함되는 것으로 해석되어야 할 것이다.
- [0145] 이상에서 실시예를 중심으로 설명하였으나 이는 단지 예시일 뿐 본 발명을 한정하는 것이 아니며, 본 발명이 속하는 분야의 통상의 지식을 가진 자라면 본 실시예의 본질적인 특성을 벗어나지 않는 범위에서 이상에 예시되지 않은 여러 가지의 변형과 응용이 가능함을 알 수 있을 것이다. 예를 들어, 실시예에 구체적으로 나타난 각 구성 요소는 변형하여 실시할 수 있는 것이다. 그리고 이러한 변형과 응용에 관계된 차이점들은 첨부된 청구 범위에서 규정하는 본 발명의 범위에 포함되는 것으로 해석되어야 할 것이다.

도면

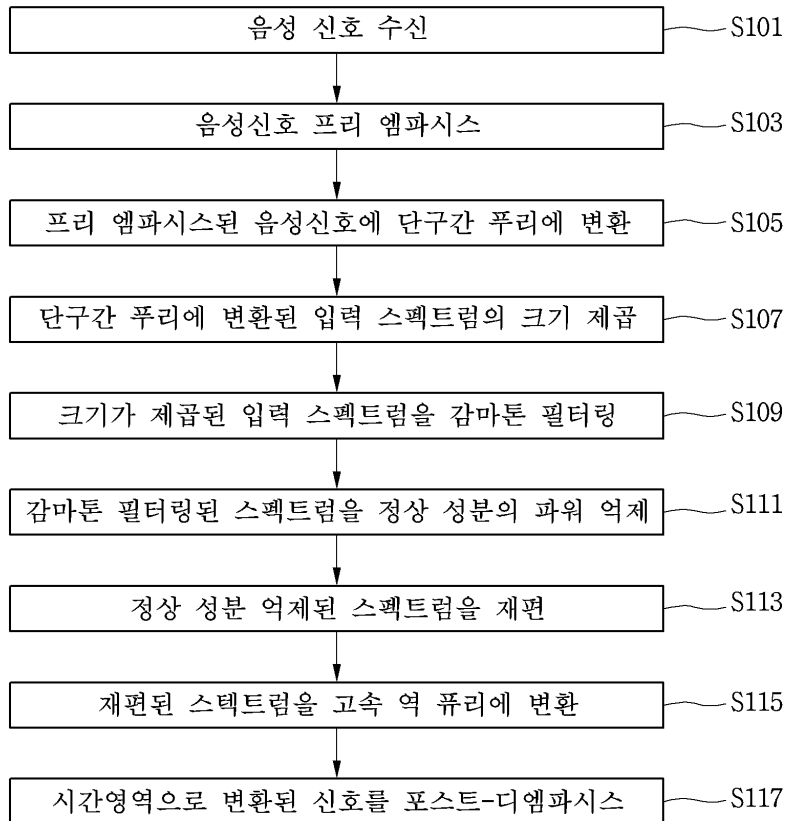
도면1



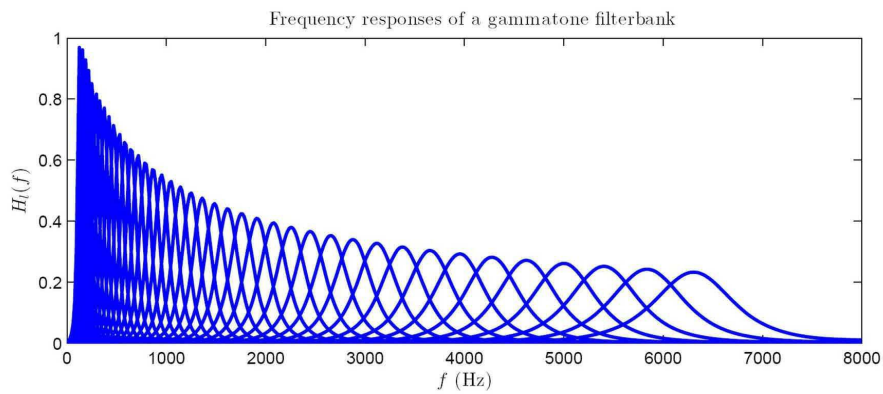
도면2



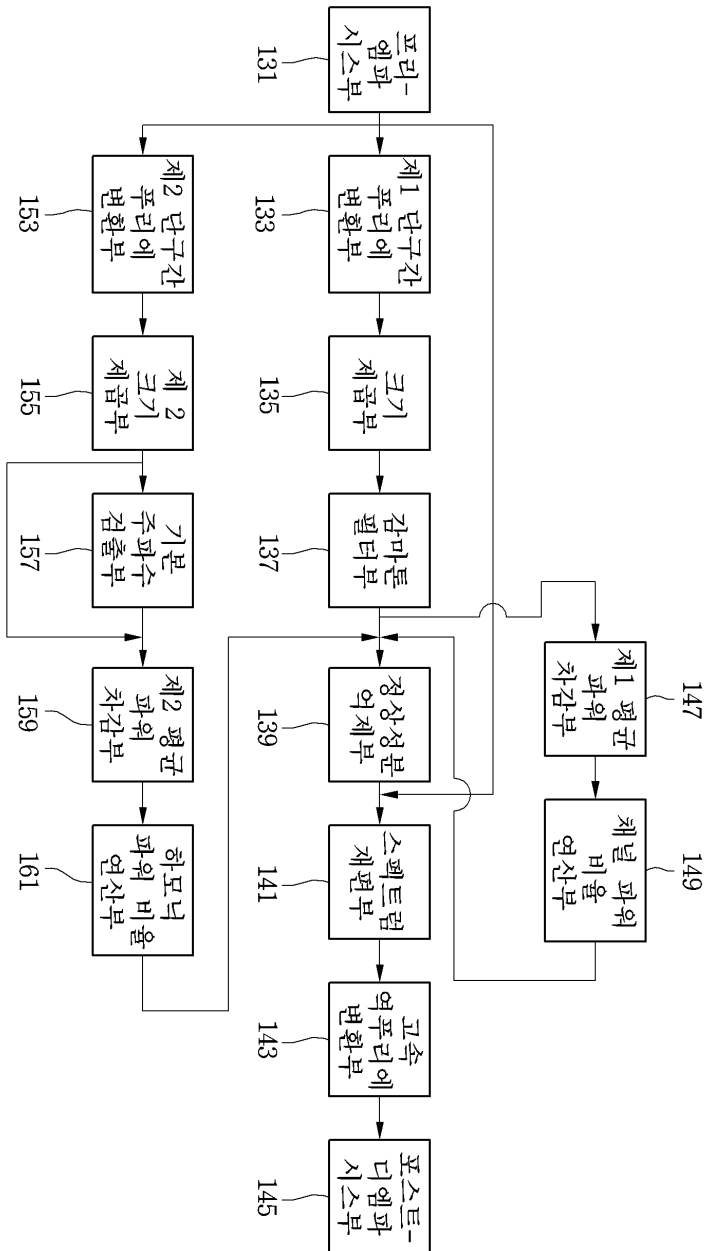
도면3



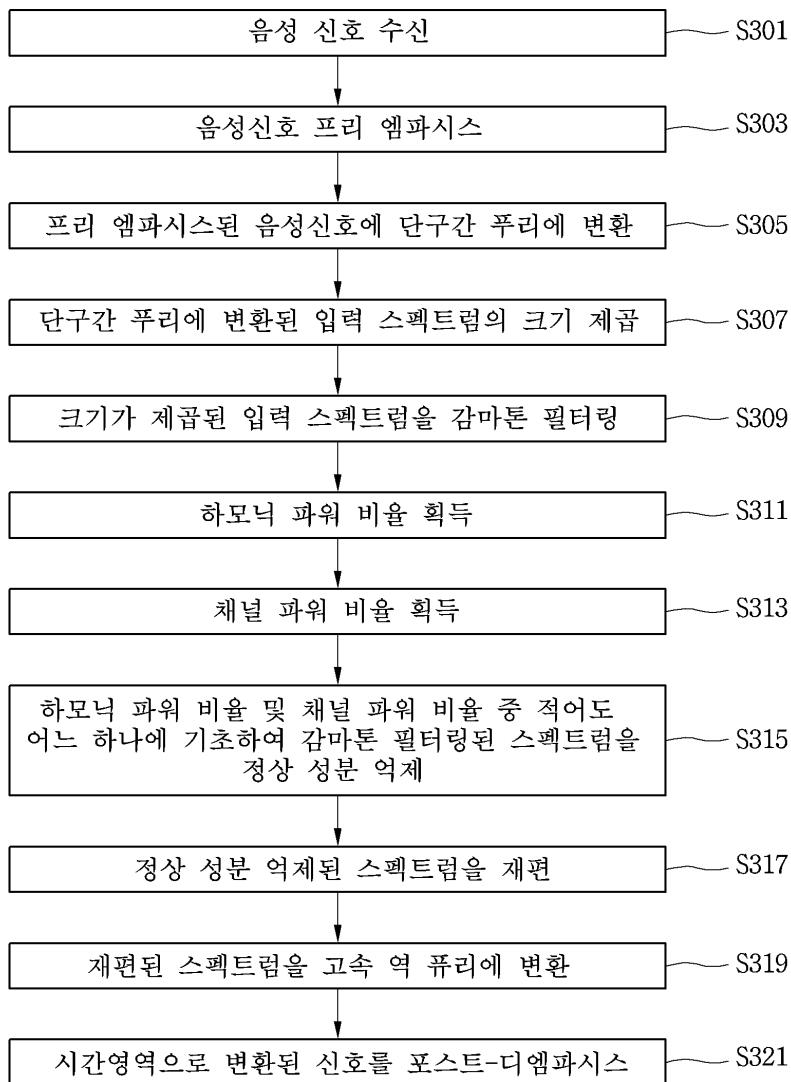
도면4



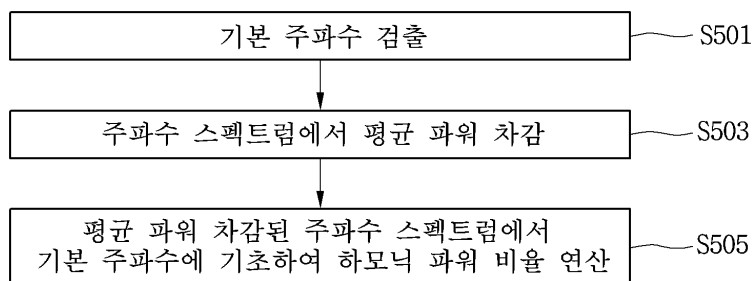
도면5



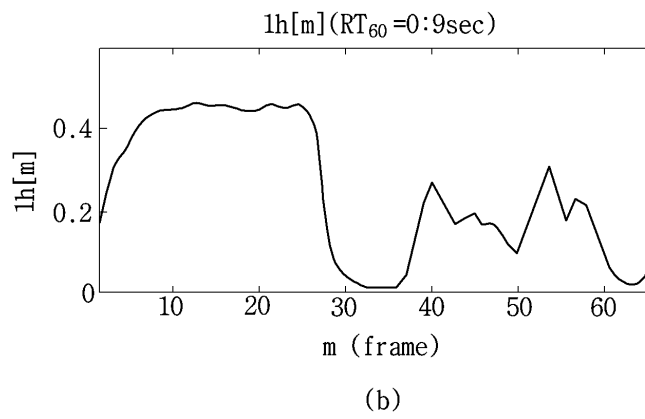
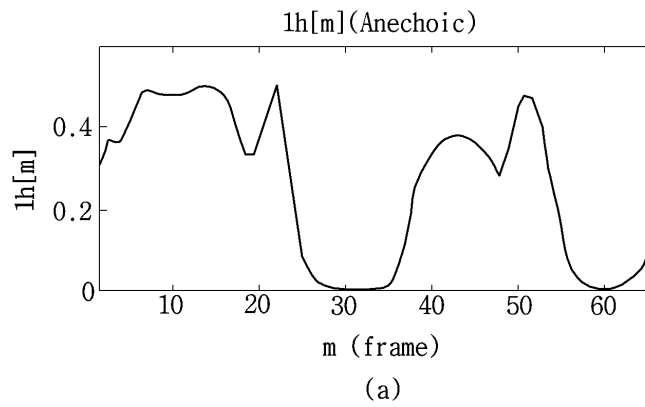
도면6



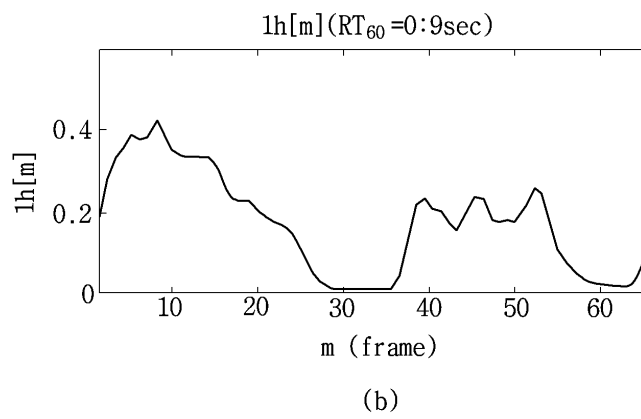
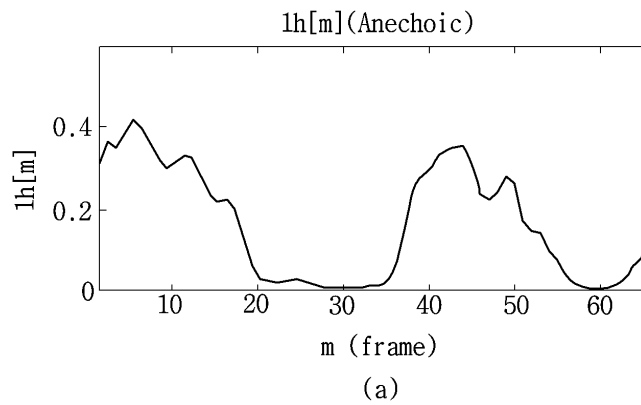
도면7



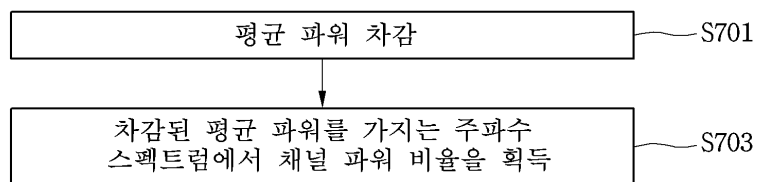
도면8



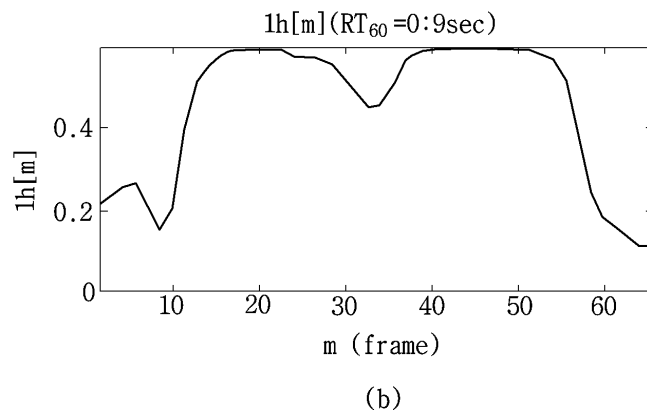
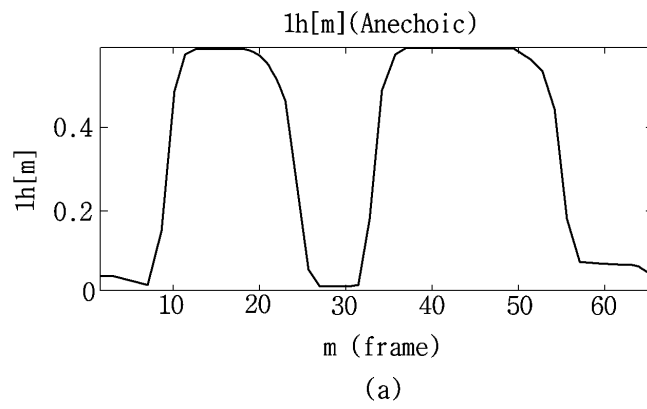
도면9



도면10



도면11



도면12

