



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0026587
(43) 공개일자 2020년03월11일

(51) 국제특허분류(Int. Cl.)
G10L 25/84 (2013.01) G10L 25/18 (2013.01)
G10L 25/78 (2013.01)
(52) CPC특허분류
G10L 25/84 (2013.01)
G10L 25/18 (2013.01)
(21) 출원번호 10-2018-0104732
(22) 출원일자 2018년09월03일
심사청구일자 2018년09월03일

(71) 출원인
국방과학연구소
대전광역시 유성구 북유성대로488번길 160 (수남동)
(72) 발명자
조영우
대전광역시 유성구 북유성대로488번길 160(수남동)
박정식
서울특별시 동대문구 천장산로9길 11, 제2교수연구동 306호 (이문동)
윤정석
경상북도 경산시 대학로 280, IT관 311호 (대동)
(74) 대리인
리엔목특허법인

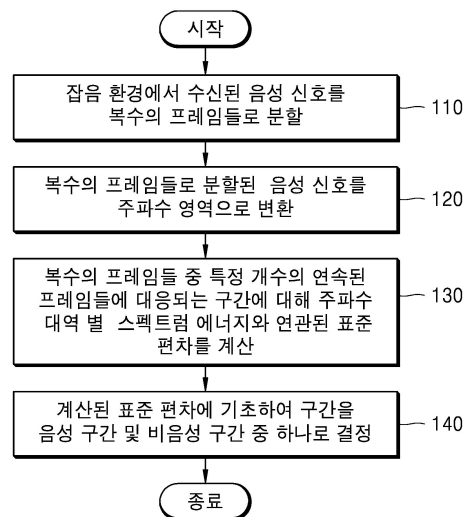
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 음성 구간을 검출하는 방법 및 장치

(57) 요약

일부 실시예에 따르면, 잡음 환경에서 수신된 음성 신호를 복수의 프레임들로 분할하는 단계, 복수의 프레임들로 분할된 음성 신호를 주파수 영역으로 변환하는 단계, 복수의 프레임들 중 특정 개수의 연속된 프레임들에 대응되는 구간에 대해 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차를 계산하는 단계 및 계산된 표준편차에 기초하여 구간을 음성 구간 및 비음성 구간 중 하나로 결정하는 단계를 포함하는, 음성 구간을 검출하는 방법이 개시된다.

대표도 - 도1



(52) CPC특허분류

G10L 25/93 (2013.01)

G10L 2025/783 (2013.01)

명세서

청구범위

청구항 1

음성 구간을 검출하는 방법에 있어서,

잡음 환경에서 수신된 음성 신호를 복수의 프레임들로 분할하는 단계;

상기 복수의 프레임들로 분할된 상기 음성 신호를 주파수 영역으로 변환하는 단계;

상기 복수의 프레임들 중 특정 개수의 연속된 프레임들에 대응되는 구간에 대해 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차를 계산하는 단계; 및

상기 계산된 표준편차에 기초하여 상기 구간을 음성 구간 및 비음성 구간 중 하나로 결정하는 단계를 포함하는, 방법.

청구항 2

제 1항에 있어서,

상기 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차를 계산하는 단계는,

복수의 주파수 대역들 각각에 대해 프레임 간 스펙트럼 에너지의 1차 표준편차를 계산하는 단계; 및

상기 복수의 주파수 대역들에 대해 계산된 1차 표준편차들의 평균을 구하는 단계를 포함하는, 방법.

청구항 3

제 2항에 있어서,

상기 구간을 음성 구간 및 비음성 구간 중 하나로 결정하는 단계는,

상기 1차 표준편차들의 평균이 기 설정된 임계값을 초과하는 경우 상기 구간을 음성 구간으로 결정하는 단계를 포함하는, 방법.

청구항 4

제 3항에 있어서,

상기 기 설정된 임계값은 상기 음성 신호의 SNR(Signal to Noise Ratio)에 따라 상이하게 결정되는, 방법.

청구항 5

제 1항에 있어서,

상기 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차를 계산하는 단계는,

복수의 주파수 대역들 각각에 대해 프레임 별로 스펙트럼 에너지의 1차 편차를 계산하고, 상기 계산된 1차 편차들의 평균을 구하는 단계;

상기 프레임 별로 계산된 1차 편차에서 상기 계산된 1차 편차들의 평균을 감산함으로써 프레임 별 스펙트럼 에너지의 2차 편차를 계산하는 단계;

상기 2차 편차에 기초하여 상기 복수의 주파수 대역들 각각에 대한 2차 표준편차를 계산하는 단계; 및

상기 복수의 주파수 대역들에 대해 계산된 2차 표준편차들의 평균을 구하는 단계를 포함하는, 방법.

청구항 6

제 5항에 있어서,

상기 구간을 음성 구간 및 비음성 구간 중 하나로 결정하는 단계는,

상기 2차 표준편차들의 평균이 기 설정된 임계값을 초과하는 경우 상기 구간을 음성 구간으로 결정하는 단계를 포함하는, 방법.

청구항 7

제 6항에 있어서,

상기 기 설정된 임계값은 상기 음성 신호의 SNR에 상관없이 동일하게 결정되는, 방법.

청구항 8

제 6항에 있어서,

상기 기 설정된 임계값은 음성 구간 검출의 성능 평가 결과에서 False Alarm 오류율 및 False reject 오류율의 합이 최소가 되도록 결정된 것인, 방법.

청구항 9

제 1항에 있어서,

상기 음성 신호는 순수 음성 신호 및 잡음 신호를 포함하는, 방법.

청구항 10

제 1항에 있어서,

상기 복수의 프레임들로 분할된 음성 신호를 주파수 영역으로 변환하는 단계는,

상기 분할된 복수의 프레임들 각각에 해밍 윈도우를 적용하는 단계; 및

상기 해밍 윈도우가 적용된 음성 신호에 대해 고속 푸리에 변환(Fast Fourier Transform: FFT)를 수행하는 단계를 포함하는, 방법.

청구항 11

제 1항의 방법을 실행하는 명령어들을 포함하는 하나 이상의 프로그램이 기록된 컴퓨터로 읽을 수 있는 기록매체.

청구항 12

음성 구간을 검출하는 장치에 있어서,

잡음 환경에서 음성 신호를 수신하는 수신기; 및

상기 수신된 음성 신호를 복수의 프레임들로 분할하고, 상기 복수의 프레임들로 분할된 음성 신호를 주파수 영역으로 변환하며, 상기 복수의 프레임들 중 특정 개수의 연속된 프레임들에 대응되는 구간에 대해 주파수 대역별 스펙트럼 에너지와 연관된 표준편차를 계산하고, 상기 계산된 표준편차에 기초하여 상기 구간을 음성 구간 및 비음성 구간 중 하나로 결정하는 제어부를 포함하는, 장치.

발명의 설명

기술 분야

[0001] 본 개시는 음성 구간을 검출하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] 최근 음성 인터페이스를 이용한 다양한 전자 기기들이 연구되고 있다. 예를 들어, 음성 인식 기술 기반의 스마트 홈 디바이스 또는 개인 비서와 같은 전자기기들이 상용화되고 있다. 스마트 홈 디바이스는 음성인식 스피커를 통해 사용자의 음성을 인식하고, 인식된 사용자의 음성에 기초하여 에어컨, 선풍기, TV 등과 같은 다양한 가전제품의 작동을 제어함으로써 사용자의 편의성을 증대시킬 수 있다. 한편, 이러한 전자 기기들이 올바르게 동

작하기 위해서는 사용자의 음성을 정확하게 인식하는 작업이 선행되어야 한다.

[0003] 음성인식 작업은 전처리 과정, 인식 과정 및 후처리 과정으로 구성된다. 전처리 과정의 주요 목적은 수신된 음성 신호로부터 음성 구간을 검출하고 각 구간으로부터 음성 특징 파라미터를 추출하는 것이다. 인식 과정은 추출된 특징 파라미터를 패턴 인식 기술로 학습함으로써 인식 결과를 얻는 과정이다. 마지막으로, 후처리 과정은 패턴 인식 기술로부터 도출된 결과를 확인하는 과정으로서, 인식 결과를 검증하는 과정이다. 전술한 세가지 과정들은 음성 인식 성능을 결정하는 각 단계별 중요한 요소이고, 독립적으로 음성인식 시스템의 성능에 영향을 미치기 때문에 아직까지 각각의 주제로 연구되고 있다.

[0004] 한편, 전술한 과정들 중 전처리 과정의 핵심인 음성 구간 검출은 인식 과정의 대상이 되는 음성 구간을 검출하는 과정으로서, 음성 인식의 전 과정에 영향을 미칠 뿐만 아니라 인식 성능을 좌우한다. 기존의 대표적인 음성 구간 검출 방법으로는 영교차율을 이용하는 방법 및 신호 에너지를 이용한 방법이 있다. 그러나, 영교차율을 이용하는 방법은 배경 잡음이 포함된 비음성 구간에서도 높은 영교차율이 나타날 수 있기 때문에 음성 구간 검출 성능이 저하되는 문제점이 있고, 신호 에너지를 이용한 방법은 잡음 신호의 에너지와 순수 음성 신호의 에너지 크기가 유사한 환경에서 정확한 음성 구간 검출이 어려운 문제점이 있다. 따라서, 잡음 환경에서도 성능이 저하되지 않고, 음성 구간을 정확하게 검출하기 위한 기술이 요구된다.

발명의 내용

해결하려는 과제

[0005] 다양한 실시예들은 음성 구간을 검출하는 방법 및 장치를 제공하는데 있다. 본 개시가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제들로 한정되지 않으며, 이하의 실시예들로부터 또 다른 기술적 과제들이 유추될 수 있다.

과제의 해결 수단

[0006] 상술한 기술적 과제를 해결하기 위한 수단으로서, 일 측면에 따른 음성 구간을 검출하는 방법은, 잡음 환경에서 수신된 음성 신호를 복수의 프레임들로 분할하는 단계; 상기 복수의 프레임들로 분할된 상기 음성 신호를 주파수 영역으로 변환하는 단계; 상기 복수의 프레임들 중 특정 개수의 연속된 프레임들에 대응되는 구간에 대해 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차를 계산하는 단계; 및 상기 계산된 표준편차에 기초하여 상기 구간을 음성 구간 및 비음성 구간 중 하나로 결정하는 단계를 포함할 수 있다.

[0007] 일 실시예에 따르면, 상기 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차를 계산하는 단계는, 복수의 주파수 대역들 각각에 대해 프레임 간 스펙트럼 에너지의 1차 표준편차를 계산하는 단계; 및 상기 복수의 주파수 대역들에 대해 계산된 1차 표준편차들의 평균을 구하는 단계를 포함할 수 있다.

[0008] 또한, 상기 구간을 음성 구간 및 비음성 구간 중 하나로 결정하는 단계는, 상기 1차 표준편차들의 평균이 기 설정된 임계값을 초과하는 경우 상기 구간을 음성 구간으로 결정하는 단계를 포함할 수 있다.

[0009] 상기 기 설정된 임계값은 상기 음성 신호의 SNR(Signal to Noise Ratio)에 따라 상이하게 결정될 수 있다.

[0010] 다른 실시예에 따르면, 상기 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차를 계산하는 단계는, 복수의 주파수 대역들 각각에 대해 프레임 별로 스펙트럼 에너지의 1차 편차를 계산하고, 상기 계산된 1차 편차들의 평균을 구하는 단계; 상기 프레임 별로 계산된 1차 편차에서 상기 계산된 1차 편차들의 평균을 감산함으로써 프레임 별 스펙트럼 에너지의 2차 편차를 계산하는 단계; 상기 2차 편차에 기초하여 상기 복수의 주파수 대역들 각각에 대한 2차 표준편차를 계산하는 단계; 및 상기 복수의 주파수 대역들에 대해 계산된 2차 표준편차들의 평균을 구하는 단계를 포함할 수 있다.

[0011] 또한, 상기 구간을 음성 구간 및 비음성 구간 중 하나로 결정하는 단계는, 상기 2차 표준편차들의 평균이 기 설정된 임계값을 초과하는 경우 상기 구간을 음성 구간으로 결정하는 단계를 포함할 수 있다.

[0012] 상기 기 설정된 임계값은 상기 음성 신호의 SNR에 상관없이 동일하게 결정될 수 있다.

[0013] 또한, 상기 기 설정된 임계값은 음성 구간 검출의 성능 평가 결과에서 False Alarm 오류율 및 False reject 오류율의 합이 최소가 되도록 결정된 것일 수 있다.

[0014] 상기 음성 신호는 순수 음성 신호 및 잡음 신호를 포함할 수 있다.

[0015] 상기 복수의 프레임들로 분할된 음성 신호를 주파수 영역으로 변환하는 단계는, 상기 분할된 복수의 프레임들 각각에 해밍 윈도우를 적용하는 단계; 및 상기 해밍 윈도우가 적용된 음성 신호에 대해 고속 푸리에 변환(Fast Fourier Transform: FFT)를 수행하는 단계를 포함할 수 있다.

[0016] 또한, 다른 측면에 따른 컴퓨터로 읽을 수 있는 기록매체는 상술한 방법을 실행하는 명령어들을 포함하는 하나 이상의 프로그램이 기록된 기록매체를 포함할 수 있다.

[0017] 또한, 또 다른 측면에 따른 음성 구간을 검출하는 장치는, 잡음 환경에서 음성 신호를 수신하는 수신기; 및 상기 수신된 음성 신호를 복수의 프레임들로 분할하고, 상기 복수의 프레임들로 분할된 음성 신호를 주파수 영역으로 변환하며, 상기 복수의 프레임들 중 특정 개수의 연속된 프레임들에 대응되는 구간에 대해 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차를 계산하고, 상기 계산된 표준편차에 기초하여 상기 구간을 음성 구간 및 비음성 구간 중 하나로 결정하는 제어부를 포함할 수 있다.

발명의 효과

[0018] 본 개시는 음성 구간을 검출하는 방법 및 장치를 제공할 수 있다. 구체적으로, 본 개시에 따른 방법 및 장치는 잡음 환경에서 수신된 음성 신호를 복수의 프레임들로 분할하고, 복수의 프레임들로 분할된 음성 신호를 주파수 영역으로 변환하며, 복수의 프레임들 중 특정 개수의 연속된 프레임들에 대응되는 구간에 대해 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차를 계산하고, 계산된 표준편차에 기초하여 구간을 음성 구간 및 비음성 구간 중 하나로 결정함으로써, 잡음 환경에서도 음성 구간을 정확하게 검출할 수 있다.

[0019] 종래의 영교차율을 이용하는 방법 및 신호의 에너지를 이용하는 방법은 시간 영역의 음성 신호 특성을 이용함에 따라 순수 음성 신호와 유사한 크기를 갖는 잡음 신호가 유입되는 경우 음성 구간 검출의 정확도가 크게 저하되었으나, 본 개시에 따른 방법 및 장치는 시간 영역이 아닌 주파수 영역의 성분을 이용하므로, 음성 구간 검출의 정확도가 증가될 수 있다.

[0020] 또한, 본 개시에 따른 방법 및 장치는 계산된 표준편차와 기 설정된 임계값과의 비교만을 통해 음성 구간 검출을 수행하므로, 심층 신경망 계산 등과 같은 부가적인 연산이 필요하지 않고, 고성능의 하드웨어를 갖추지 않더라도 실시간으로 음성 구간 검출이 가능하다.

[0021] 또한, 본 개시에 따른 방법 및 장치는 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차로서, 2차 표준편차를 이용하여 기 설정된 임계값과의 비교를 수행함으로써, 음성 신호의 잡음 수준(예를 들어, SNR)에 상관없이 단일한 임계값을 적용할 수 있다. 2차 표준편차는 1차 편차로부터 다시 한번 계산된 2차 편차를 이용하여 정의되는 표준편차를 의미할 수 있다.

도면의 간단한 설명

[0022] 도 1은 일부 실시예에 따른 음성 구간을 검출하는 방법을 나타내는 흐름도이다.

도 2는 시간 영역 및 주파수 영역 각각에서 음성 구간 및 비음성 구간을 분류한 결과의 예시를 나타내는 도면이다.

도 3은 일부 실시예에 따른 음성 구간 검출의 성능 평가 결과에 기초하여 임계값을 결정하는 과정을 설명하기 위한 도면이다.

도 4는 일부 실시예에 따른 1차 표준편차를 이용하여 음성 구간을 검출한 결과 및 주파수 에너지를 이용하여 음성 구간을 검출한 결과를 비교하는 도면이다.

도 5는 일부 실시예에 따른 2차 표준편차를 이용하여 음성 구간을 검출한 결과, 1차 표준편차를 이용하여 음성 구간을 검출한 결과 및 주파수 에너지를 이용하여 음성 구간을 검출한 결과를 비교하는 도면이다.

도 6은 일부 실시예에 따른 주파수 대역 별 스펙트럼 에너지의 편차를 이용하는 방법 및 종래 기술에 따른 시간 영역 에너지를 이용하는 방법 각각에 대해 임계값을 변경해가면서 음성 구간 검출의 성능을 평가한 결과를 나타내는 표이다.

도 7은 일부 실시예에 따른 주파수 대역 별 스펙트럼 에너지의 편차를 이용하는 방법 및 종래 기술에 따른 시간 영역 에너지를 이용하는 방법 각각에 대해 음성 신호의 SNR을 변경해가면서 음성 구간 검출의 성능을 평가한 결과를 나타내는 표이다.

도 8은 일부 실시예에 따른 음성 구간을 검출하는 장치의 구성을 나타내는 블록도이다.

발명을 실시하기 위한 구체적인 내용

- [0023] 이하 첨부된 도면을 참조하면서 오로지 예시를 위한 실시예들을 상세히 설명하기로 한다. 하기 설명은 실시예들을 구체화하기 위한 것일 뿐 발명의 권리 범위를 제한하거나 한정하는 것이 아님은 물론이다. 상세한 설명 및 실시예로부터 당해 기술분야의 전문가가 용이하게 유추할 수 있는 것은 권리범위에 속하는 것으로 해석된다.
- [0024] 본 명세서에서 사용되는 '구성된다' 또는 '포함한다' 등의 용어는 명세서 상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다.
- [0025] 또한, 본 명세서에서 사용되는 '제 1' 또는 '제 2' 등과 같이 서수를 포함하는 용어는 다양한 구성 요소들을 설명하는데 사용할 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로만 사용된다.
- [0026] 본 명세서에서 사용되는 용어는 본 발명에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도 또는 관례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 발명에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 발명의 전반에 걸친 내용을 토대로 정의되어야 한다.
- [0027] 본 실시예들은 음성 구간을 검출하는 방법 및 장치에 관한 것으로서 이하의 실시예들이 속하는 기술 분야에서 통상의 지식을 가진 자에게 널리 알려져 있는 사항들에 관해서는 자세한 설명을 생략한다.
- [0028] 도 1은 일부 실시예에 따른 음성 구간을 검출하는 방법을 나타내는 흐름도이다. 도 1의 방법은 음성 구간 검출 장치에 의해 수행될 수 있다. 음성 구간 검출 장치는 음성 인터페이스를 이용한 다양한 전자 기기들에 포함되어 음성 인식 작업의 전처리 과정을 수행하는 장치의 일부 구성요소를 의미할 수 있다.
- [0029] 단계 110에서, 음성 구간 검출 장치는 잡음 환경에서 수신된 음성 신호를 복수의 프레임들로 분할할 수 있다. 잡음 환경에서 수신된 음성 신호는 순수 음성 신호 및 잡음 신호(또는 비음성 신호)를 포함할 수 있다. 순수 음성 신호는 사용자의 발화에 의해 발생하는 신호로서 음성 인식의 대상이 되는 신호를 의미하고, 잡음 신호는 순수 음성 신호 외의 신호로서 음성 인식의 대상이 되지 않는 신호를 의미할 수 있다. 음성 구간 검출 장치는 음성 신호를 일정한 시간 단위, 즉 프레임 단위로 분할할 수 있다. 음성 구간 또는 비음성 구간은 특정 개수의 연속된 프레임들로 정의될 수 있다.
- [0030] 단계 120에서, 음성 구간 검출 장치는 복수의 프레임들로 분할된 음성 신호를 주파수 영역으로 변환할 수 있다. 음성 구간 검출 장치는 신호의 왜곡을 방지하기 위해 분할된 복수의 프레임들 각각에 해밍 윈도우를 적용할 수 있다. 또한, 음성 구간 검출 장치는 해밍 윈도우가 적용된 음성 신호에 대해 고속 푸리에 변환(Fast Fourier Transform: FFT)를 수행할 수 있다. FFT가 수행됨에 따라 시간 영역의 음성 신호가 주파수 영역으로 변환될 수 있다. 이와 같이, 본 개시에 따른 음성 구간 검출 장치는 시간 영역이 아닌 주파수 영역의 성분을 이용함으로써 음성 구간 검출의 정확도를 증가시킬 수 있다. 이하 도 2를 참조하여 주파수 영역의 성분을 이용함에 따라 음성 구간 검출의 정확도가 증가되는 이유를 상세히 설명한다.
- [0031] 도 2는 시간 영역 및 주파수 영역 각각에서 음성 구간 및 비음성 구간을 분류한 결과의 예시를 나타내는 도면이다.
- [0032] 도 2를 참조하면, 시간 영역에서 시간에 따른 음성 신호의 신호 값을 나타내는 그래프(210) 및 주파수 영역에서 시간에 따른 주파수 대역 별 에너지의 변화를 나타내는 그래프(220)가 도시되어 있다.
- [0033] 그래프(210)를 참조하면, 시간 영역에서 음성 구간의 신호 값이 일반적으로 비음성 구간의 신호 값보다 크다는 점을 이용하여 음성 구간 및 비음성 구간이 구별될 수 있음을 알 수 있다. 다만, 잡음 신호의 신호 값 및 음성 신호의 신호 값의 크기가 유사한 경우라면, 그래프(210)에 도시된 것처럼 신호 값의 차이가 크게 나는 구간들이 존재하지 않게 되고, 음성 구간 및 비음성 구간이 명확하게 구분되지 않을 수 있다.
- [0034] 한편, 시간 영역에서 잡음 신호의 신호 값 및 음성 신호의 신호 값의 크기가 유사한 경우라도, 주파수 영역에서 시간에 따른 주파수 대역 별 에너지를 이용하면 음성 구간 및 비음성 구간의 명확한 구분이 가능함을 그래프

(220)를 통해 확인할 수 있다. 그래프(220)를 참조하면, 음성 구간과 비교하여 비음성 구간에서 주파수 대역 별 스펙트럼 에너지의 편차가 작음을 알 수 있다. 잡음 신호가 시간 영역에서 큰 신호 값을 갖는 경우라도 잡음 신호가 존재하는 주파수 대역에서의 스펙트럼 에너지 편차는 크지 않을 수 있기 때문이다. 이와 같이, 시간 영역이 아닌 주파수 영역의 성분을 이용하게 되면 시간 영역에서 잡음 신호의 신호 값이 얼마나 큰지에 상관 없이 정확한 음성 구간 검출이 수행될 수 있다.

[0035] 다시 도 1로 돌아오면, 단계 130에서, 음성 구간 검출 장치는 복수의 프레임들 중 특정 개수의 연속된 프레임들에 대응되는 구간에 대해 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차를 계산할 수 있다. 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차는 1차 표준편차 및 2차 표준편차를 포함할 수 있으나, 반드시 이에 제한되는 것은 아니다. 이하에서는 음성 구간 검출 장치가 복수의 주파수 대역들 각각에 대해 프레임 간 스펙트럼 에너지의 1차 표준편차를 계산하는 과정에 대해 상세히 설명한다.

[0036] 음성 구간 검출 장치는 표준편차 계산에 필요한 주파수 대역 별 스펙트럼 에너지를 다음과 같은 수학적 식 1에 따라 계산할 수 있다.

수학적 식 1

[0037]
$$P_t[m] = \sqrt{Re_t[m]^2 + Im_t[m]^2}$$

[0038] 수학적 식 1에서 $P_t[m]$ 는 t 번째 프레임의 m번째 주파수 대역의 스펙트럼 에너지를 의미하고, $Re_t[m]$ 및 $Im_t[m]$ 각각은 주파수 대역에 대한 실수 값 및 허수 값을 나타낼 수 있다.

[0039] 음성 구간 검출 장치는 각 주파수 대역 별 스펙트럼 에너지로부터 편차를 계산하기 위해 주파수 대역 별 스펙트럼 에너지의 평균을 다음과 같은 수학적 식 2에 따라 계산할 수 있다.

수학적 식 2

[0040]
$$\mu^{(1)}[m] = \frac{1}{L} \sum_{k=0}^{L-1} (P_{t-k}[m])$$

[0041] 수학적 식 2에서 $\mu^{(1)}[m]$ 는 특정 개수의 프레임들에 대한 m 번째 주파수 대역의 스펙트럼 에너지의 평균을 의미하고, 숫자 (1)은 주파수 대역 별 스펙트럼 에너지로부터 1차적으로 계산한 평균임을 나타내는 식별자일 수 있다. 또한, L은 특정 구간에 대응되는 프레임들의 개수이며, k는 t, t-1, ..., t-(L-1) 프레임들을 나타내기 위한 상수일 수 있다.

[0042] 음성 구간 검출 장치는 앞의 수학적 식 2에 따라 평균을 계산한 후 각 주파수 대역 별 스펙트럼 에너지로부터 편차를 계산할 수 있다. 편차는 다음과 같은 수학적 식 3에 따라 계산될 수 있다.

수학적 식 3

[0043]
$$D_t^{(1)}[m] = P_t[m] - \mu^{(1)}[m]$$

[0044] 수학적 식 3에서 $D_t^{(1)}[m]$ 은 t 번째 프레임 m 번째 주파수 대역의 스펙트럼 에너지로부터 평균을 차감한 편차를 의미할 수 있다. t, t-1, ..., t-(L-1) 프레임들 간 주파수 대역 별 스펙트럼 에너지가 유사할 경우, 앞의 수학적 식 3에 따라 계산되는 편차는 작아질 수 있다. 반대로, t, t-1, ..., t-(L-1) 프레임들 간 주파수 대역 별 스펙트럼 에너지가 상이할 경우 앞의 수학적 식 3에 따라 계산되는 편차는 커질 수 있고, 이는 주파수 대역 별 스펙트럼

럼 에너지의 차이가 크다는 것을 의미할 수 있다.

[0045] 한편, 편차는 양수 값 및 음수 값을 모두 가지므로, 편차를 모두 더할 경우 의도하지 않은 값인 0이 나올 수 있다. 이와 같은 문제를 해결하기 위해 음수의 값을 양수화해야 하고, 다양한 방법들 중 하나로서 제곱이 적용될 수 있다. 음성 구간 검출 장치는 다음과 같은 수학적 식 4에 따라 편차 제곱의 합인 분산을 계산할 수 있다.

수학적 식 4

$$V^{(1)}[m] = \frac{1}{L} \sum_{k=0}^{L-1} (D_{t-k}^{(1)}[m])^2$$

[0046]

[0047] 수학적 식 4에서 $V^{(1)}[m]$ 은 m 번째 주파수 대역의 편차에 대한 분산을 의미할 수 있다. 분산은 t, t-1, ..., t-(L-1) 프레임들 각각의 편차를 제공하여 더한 뒤 프레임들의 총수인 L로 나눔으로써 계산될 수 있다. 분산을 통해 주파수 대역 별 스펙트럼 에너지의 평균에 대한 오차를 알 수 있으나, 분산은 편차를 제공하여 계산함에 따라 실제 값과의 차이가 큰 문제가 있다. 따라서, 실제 값으로 근접시키기 위해 제곱근이 적용될 수 있고, 이는 표준편차에 대응된다. 음성 구간 검출 장치는 다음과 같은 수학적 식 5에 따라 프레임 간 스펙트럼 에너지의 1차 표준편차를 계산할 수 있다.

수학적 식 5

$$Std^{(1)}[m] = \sqrt{V^{(1)}[m]}$$

$$Std_E^{(1)} = \frac{1}{N/d} \sum_{m=0}^{N/d} (Std^{(1)}[m])$$

[0048]

[0049] 수학적 식 5에서 $Std^{(1)}[m]$ 은 m 번째 주파수 대역의 분산을 제곱근하여 계산한 표준편차를 의미할 수 있다. 표준편차는 주파수 대역 별 스펙트럼 에너지의 평균으로부터 실제 값과 비교 가능한 오차에 대응될 수 있다. 한편, 음성 구간 검출 장치는 복수의 주파수 대역들에 대해 계산된 1차 표준편차들의 평균을 구할 수 있다. $Std_E^{(1)}$ 은 각각의 주파수 대역 별 1차 표준편차를 모두 합한 후 특정 주파수 대역의 FFT 계수인 N/d로 나눈 1차 표준편차들의 평균을 의미할 수 있다.

[0050] 단계 140에서, 음성 구간 검출 장치는 계산된 표준편차에 기초하여 구간을 음성 구간 및 비음성 구간 중 하나로 결정할 수 있다. 예를 들어, 음성 구간 검출 장치는 1차 표준편차들의 평균이 기 설정된 임계값을 초과하는 경우 구간을 음성 구간으로 결정할 수 있다. 또한, 음성 구간 검출 장치는 1차 표준편차들의 평균이 기 설정된 임계값 이하인 경우 구간을 비음성 구간으로 결정할 수 있다. 기 설정된 임계값은 음성 구간 검출의 성능 평가 결과에 기초하여 실험적으로 추정될 수 있다. 이하 도 3을 참조하여 기 설정된 임계값이 결정되는 과정을 상세히 설명한다.

[0051] 도 3은 일부 실시예에 따른 음성 구간 검출의 성능 평가 결과에 기초하여 임계값을 결정하는 과정을 설명하기 위한 도면이다.

[0052] 도 3을 참조하면, 음성 구간 검출의 성능 평가 결과에 따른 False alarm(FA) 오류율 및 False reject(FR) 오류율 간의 관계를 나타내는 그래프가 도시되어 있다. False alarm은 비음성 구간을 음성 구간으로 잘못 판단하는 오류를 의미하고, False reject는 음성 구간을 비음성 구간이라고 잘못 판단하는 오류를 의미한다.

[0053] 임계값을 작은 값에서 큰 값으로 변경하면서 각각의 임계값에 대해 음성 구간 검출의 성능 평가를 수행할 경우 False alarm 및 False reject의 발생 비율이 계산될 수 있다. 예를 들어, False alarm 오류율 및 False reject 오류율은 도 3에 도시된 바와 같이 반비례 관계를 갖는다. 이 때, 그래프의 원점에서 가장 가까운 지점(310)은 False Alarm 오류율 및 False reject 오류율의 합이 최소가 되는 지점으로서, 지점(310)에 대응되는 임계값이

최적의 임계값으로 추정될 수 있다.

- [0054] 이하 도 4를 참조하여 1차 표준편차들의 평균을 이용하여 음성 구간을 검출하는 방법의 유효성에 대해 상세히 설명한다.
- [0055] 도 4는 일부 실시예에 따른 1차 표준편차를 이용하여 음성 구간을 검출한 결과 및 주파수 에너지를 이용하여 음성 구간을 검출한 결과를 비교하는 도면이다.
- [0056] 도 4를 참조하면, 잡음 수준에 따른 편차를 확인하기 위해 0dB, 5dB, 10dB 및 15dB의 다양한 SNR(Signal to Noise Ratio)을 갖는 음성 신호들에 대해 시간에 따른 주파수 스펙트럼 에너지를 나타내는 그래프(410) 및 시간에 따른 주파수 대역 별 스펙트럼 에너지의 1차 표준편차를 나타내는 그래프(420)가 도시되어 있다.
- [0057] 그래프(410) 및 그래프(420)를 비교하면, 주파수 에너지를 계산한 결과보다 1차 표준편차를 계산한 결과에서 시간에 따른 값들의 분포가 더욱 일정함을 알 수 있다. 또한, 이러한 특징은 비음성 구간에서 보다 명확히 나타남을 알 수 있다. 1차 표준편차를 이용하여 음성 구간을 검출하는 경우 주파수 에너지를 이용하여 음성 구간을 검출하는 경우보다 음성 구간 및 비음성 구간 간의 차이가 명확하고, 이에 따라 음성 구간이 더 정확하게 검출될 수 있다.
- [0058] 한편, 음성 구간 및 비음성 구간을 구분하기 위한 임계값은 앞에서 도 3을 참조하여 설명한 방식에 의해 결정될 수 있다. 도 4를 참조하면, 음성 신호의 SNR에 따라 상이한 임계값들이 결정됨을 알 수 있다. 주파수 에너지를 이용하는 경우 15Db 결과와 0dB 결과 간의 임계값 차이는 약 5,600,000이다. 1차 표준편차를 이용하는 경우 15Db 결과와 0dB 결과 간의 임계값 차이는 약 22,000으로서, 주파수 에너지를 이용하는 경우보다 감소될 수 있다. 다만, 1차 표준편차를 이용하는 경우에도 여전히 잡음 수준 별로 임계값의 차이가 존재한다.
- [0059] 한편, 음성 구간 검출 장치가 어떠한 잡음 환경에서도 음성 구간을 정확하게 검출하는 범용성을 갖기 위해서는 잡음 수준에 따라 상이한 임계값들을 적용하는 것보다는 잡음 수준에 상관없이 단일한 임계값을 적용하는 것이 바람직할 수 있다. 본 개시에 따른 방법 및 장치는 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차로서, 2차 표준편차를 이용하여 기 설정된 임계값과의 비교를 수행함으로써, 음성 신호의 잡음 수준(예를 들어, SNR)에 상관없이 단일한 임계값을 적용할 수 있다. 이하에서는 음성 구간 검출 장치가 복수의 주파수 대역들 각각에 대해 프레임 간 스펙트럼 에너지의 2차 표준편차를 계산하는 과정에 대해 상세히 설명한다.
- [0060] 음성 구간 검출 장치는 복수의 주파수 대역들 각각에 대해 프레임 별로 스펙트럼 에너지의 1차 편차를 계산할 수 있다. 1차 편차는 앞의 수학적 식 3에 따라 계산되는 편차를 의미할 수 있다. 음성 구간 검출 장치는 1차 편차의 값으로부터 다시 한번 계산된 편차를 의미하는 2차 편차를 계산하기 위해 계산된 1차 편차들의 평균을 다음과 같은 수학적 식 6에 따라 구할 수 있다.

수학적 식 6

$$\mu^{(2)}[m] = \frac{1}{L} \sum_{k=0}^{L-1} (D_{t-k}^{(1)}[m])$$

- [0061]
- [0062] 수학적 식 6에서 $\mu^{(2)}[m]$ 는 특정 개수의 프레임들에 대해 m 번째 주파수 대역의 1차 편차(즉, $D_t^{(1)}[m]$)의 평균을 의미할 수 있다. 숫자 (2)는 주파수 대역 별 스펙트럼 에너지의 1차 편차로부터 2차 편차를 도출하기 위해 계산된 평균임을 나타내는 식별자일 수 있다.

- [0063] 음성 구간 검출 장치는 수학적 식 6에 따라 계산된 평균을 이용하여 각 주파수 대역 별 2차 편차를 계산할 수 있다. 2차 편차는 다음과 같은 수학적 식 7에 따라 계산될 수 있다.

수학적 식 7

$$D_t^{(2)}[m] = D_t^{(1)}[m] - \mu^{(2)}[m]$$

[0064]

[0065] 수학식 7에서 $D_t^{(2)}[m]$ 는 t 번째 프레임의 m 번째 대역 별 $D_t^{(1)}[m]$ 에서 수학식 6에 따른 평균을 감산한 2차 편차를 의미할 수 있다. 이와 같이, 음성 구간 검출 장치는 프레임 별로 계산된 1차 편차에서 계산된 1차 편차들의 평균을 감산함으로써 프레임 별 스펙트럼 에너지의 2차 편차를 계산할 수 있다.

[0066] 음성 구간 검출 장치는 2차 편차에 기초하여 복수의 주파수 대역들 각각에 대한 2차 분산 및 2차 표준편차를 다음과 같은 수학식 8에 따라 계산할 수 있다.

수학식 8

$$V^{(2)}[m] = \frac{1}{L} \sum_{k=0}^{L-1} (D_{t-k}^{(2)}[m])^2$$

$$Std^{(2)}[m] = \sqrt{V^{(2)}[m]}$$

[0067]

[0068] 또한, 음성 구간 검출 장치는 복수의 주파수 대역들에 대해 계산된 2차 표준편차들의 평균을 다음과 같은 수학식 9에 따라 구할 수 있다.

수학식 9

$$Std_E^{(2)} = \frac{1}{N/d} \sum_{m=0}^{N/d} (Std^{(2)}[m])$$

[0069]

[0070] 수학식 9에서 $Std_E^{(2)}$ 은 각각의 주파수 대역 별 2차 표준편차를 모두 합한 후 특정 주파수 대역의 FFT 계수인 N/d로 나눈 2차 표준편차들의 평균을 의미할 수 있다.

[0071] 음성 구간 검출 장치는 2차 표준편차들의 평균이 기 설정된 임계값을 초과하는 경우 구간을 음성 구간으로 결정할 수 있다. 또한, 음성 구간 검출 장치는 2차 표준편차들의 평균이 기 설정된 임계값을 이하인 경우 구간을 비음성 구간으로 결정할 수 있다. 이하 도 5를 참조하여 2차 표준편차들의 평균을 이용하여 음성 구간을 검출하는 방법의 유효성에 대해 상세히 설명한다.

[0072] 도 5는 일부 실시예에 따른 2차 표준편차를 이용하여 음성 구간을 검출한 결과, 1차 표준편차를 이용하여 음성 구간을 검출한 결과 및 주파수 에너지를 이용하여 음성 구간을 검출한 결과를 비교하는 도면이다.

[0073] 도 5를 참조하면, 도 4와 비교하여 시간에 따른 주파수 대역 별 스펙트럼 에너지의 2차 표준편차를 나타내는 그래프(510)가 추가적으로 도시되어 있다. 그래프(510)를 참조하면, 2차 표준편차를 이용하는 경우 1차 표준편차를 이용하는 경우와 비교하여 15Db 결과와 0dB 결과 간의 임계값 차이가 약 22,000에서 약 200으로 감소됨을 알 수 있고, 0dB, 5dB 및 10dB 간에는 동일한 임계값이 적용됨을 알 수 있다. 이는 2차 표준편차를 이용하는 경우 주파수 에너지를 이용하는 경우보다 비음성 구간에서 값이 분산되지 않는 특성이 유지되면서도 대부분의 잡음 수준에 적용될 수 있는 고정된 임계값이 추정될 수 있음을 나타낸다. 2차 표준편차를 이용하는 경우 음성 구간과 비음성 구간을 구분하기 위한 임계값이 음성 신호의 SNR에 상관없이 동일하게 결정될 수 있다.

[0074] 이하 도 6 및 도 7을 참조하여 주파수 대역 별 스펙트럼 에너지의 편차를 이용하는 방법이 종래 기술에 따른 시간 영역 에너지를 이용하는 방법보다 개선된 성능을 가진다는 점을 보여주는 실험 결과를 설명한다. 실험에는 남자 5명이 실시간으로 입력한 음성 신호로서, "마이크 입력", "오케이 음성"등의 단어의 발화에 의한 음성 신호가 사용되었다. 또한, 음성 구간 검출에 사용된 하나의 프레임의 샘플 수는 320개이며, 16kHz의 샘플링 주파수에 의해 표본화되었다. 또한, 주파수 영역에서 FFT 크기는 512로 설정되었으며, 각 음성 신호에 대해 해밍 윈도우가 적용되었다. 해당 실험에서는 주파수 영역의 특성을 고려한 음성 구간 검출의 성능이 평가되었으며, 실시간으로 입력된 음성 신호를 대상으로 시간 영역의 에너지에 기초한 종래의 방법과의 성능 비교가 수행되었다.

[0075] 도 6은 일부 실시예에 따른 주파수 대역 별 스펙트럼 에너지의 편차를 이용하는 방법 및 종래 기술에 따른 시간

영역 에너지를 이용하는 방법 각각에 대해 임계값을 변경해가면서 음성 구간 검출의 성능을 평가한 결과를 나타내는 표이다.

- [0076] 도 6을 참조하면, 주파수 대역 별 스펙트럼 에너지의 편차를 이용하는 방법(도 6의 주파수 에너지 기반 VAD(Voice Activity Detection)) 및 종래 기술에 따른 시간 영역 에너지를 이용하는 방법(도 6의 시간 영역 에너지 기반 VAD)을 각각 이용하였을 때, False alarm 오류 및 False reject 오류의 빈도수에 대한 결과가 도시되어 있다. 다양한 잡음 수준을 갖는 잡음 환경에서 발화한 화자 5명의 임계값 별 결과를 참조하면, 일부 실시예에 따른 주파수 대역 별 스펙트럼 에너지의 편차를 이용하는 방법을 이용하는 경우 종래 기술에 따른 시간 영역 에너지를 이용하는 방법을 이용하는 경우보다 더 낮은 False alarm 및 False reject 오류 빈도가 관찰된다. 이러한 결과는 주파수 대역 별 스펙트럼 에너지의 편차를 이용하는 방법이 높은 음성 구간 검출 성능을 가진다는 것을 의미한다.
- [0077] 도 7은 일부 실시예에 따른 주파수 대역 별 스펙트럼 에너지의 편차를 이용하는 방법 및 종래 기술에 따른 시간 영역 에너지를 이용하는 방법 각각에 대해 음성 신호의 SNR을 변경해가면서 음성 구간 검출의 성능을 평가한 결과를 나타내는 표이다. 실험을 위해 자동차 잡음 신호가 사용되었으며, 모든 SNR에 대해 공통된 임계값이 적용되었다. 또한, 공정한 평가를 위해 각 잡음 수준에 따라 달라지는 3개의 프레임에 대한 에너지 값으로부터 평균이 계산되었다. 한편, 음성 구간 검출 성능을 판단하기 위한 기준으로서 False reject 오류가 사용되었다.
- [0078] 도 7을 참조하면, 음성 신호가 어떠한 SNR을 갖는 경우라도 주파수 대역 별 스펙트럼 에너지의 편차를 이용하는 방법이 종래 기술에 따른 시간 영역 에너지를 이용하는 방법보다 낮은 빈도수의 False reject를 가짐을 알 수 있다. 이러한 결과는 주파수 대역 별 스펙트럼 에너지의 편차를 이용하는 방법이 높은 음성 구간 검출 성능을 가진다는 것을 의미한다.
- [0079] 한편, 음성 구간을 검출하는 방법은 그 방법을 실행하는 명령어들을 포함하는 하나 이상의 프로그램이 기록된 컴퓨터로 읽을 수 있는 기록 매체에 기록될 수 있다. 컴퓨터로 읽을 수 있는 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령어의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.
- [0080] 도 8은 일부 실시예에 따른 음성 구간을 검출하는 장치의 구성을 나타내는 블록도이다.
- [0081] 도 8를 참조하면, 음성 구간을 검출하는 장치(80)는 수신기(810) 및 제어부(820)를 포함할 수 있다. 도 8에 도시된 장치(80)는 도 1에 도시된 방법을 시계열적으로 처리할 수 있다. 따라서, 이하에서 생략된 내용이라고 하더라도 도 1의 방법에 관하여 이상에서 기술된 내용은 도 8의 장치(80)에 의해 수행될 수 있음을 알 수 있다.
- [0082] 한편, 도 8에 도시된 장치(80)에는 본 실시예와 관련된 구성들만이 도시되어 있다. 따라서, 도 8에 도시된 구성요소들 외에 다른 범용적인 구성요소들이 장치(80)에 더 포함될 수 있음을 본 실시예와 관련된 기술분야에서 통상의 지식을 가진 자라면 이해할 수 있다. 예를 들어, 장치(80)는 메모리(미도시)를 더 포함할 수 있다.
- [0083] 메모리는 장치(80) 내에서 처리되는 각종 데이터들을 저장하는 하드웨어로서, 예를 들어, 메모리는 장치(80)에서 처리된 데이터들 및 처리될 데이터들을 저장할 수 있다. 또한, 메모리는 장치(80)에 의해 구동될 애플리케이션들, 드라이버들 등을 저장할 수 있다.
- [0084] 메모리는 DRAM(dynamic random access memory), SRAM(static random access memory) 등과 같은 RAM(random access memory), ROM(read-only memory), EEPROM(electrically erasable programmable read-only memory), CD-ROM, 블루레이 또는 다른 광학 디스크 스토리지, HDD(hard disk drive), SSD(solid state drive), 또는 플래시 메모리를 포함하며, 나아가서, 장치(80)에 액세스될 수 있는 외부의 다른 스토리지 디바이스를 포함할 수 있다.
- [0085] 수신기(810)는 음성 신호를 수신하는 임의의 적절한 장치를 의미할 수 있다. 예를 들어, 수신기(810)는 잡음 환경에서 음성 신호를 수신할 수 있다. 잡음 환경에서 수신된 음성 신호는 순수 음성 신호 및 잡음 신호(또는 비음성 신호)를 포함할 수 있다.
- [0086] 제어부(820)는 하나 또는 복수 개의 프로세서에 의하여 구현될 수 있다. 예를 들어, 제어부(820)는 다수의 논리 게이트들의 어레이로 구현될 수 있고, 범용적인 마이크로 프로세서와 마이크로 프로세서에서 실행될 수 있는 프

로그래밍이 저장된 메모리의 조합으로 구현될 수도 있다.

[0087] 제어부(820)는 수신된 음성 신호를 복수의 프레임들로 분할하고, 복수의 프레임들로 분할된 음성 신호를 주파수 영역으로 변환하며, 복수의 프레임들 중 특정 개수의 연속된 프레임들에 대응되는 구간에 대해 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차를 계산하고, 계산된 표준편차에 기초하여 구간을 음성 구간 및 비음성 구간 중 하나로 결정할 수 있다.

[0088] 제어부(820)는 앞의 수학식 5에 따라 계산되는 1차 표준편차들의 평균 또는 앞의 수학식 9에 따라 계산되는 2차 표준편차들의 평균을 계산하고, 1차 표준편차들의 평균 또는 2차 표준편차들의 평균이 기 설정된 임계값을 초과하는 경우 판정 대상이 되는 구간을 음성 구간으로 결정할 수 있다. 기 설정된 임계값은 음성 구간 검출의 성능 평가 결과에서 False Alarm 오류율 및 False reject 오류율의 합이 최소가 되도록 결정될 수 있다. 또한, 기 설정된 임계값은 음성 신호의 SNR에 따라 상이하게 결정될 수 있고, 음성 신호의 SNR에 상관없이 동일하게 결정될 수도 있다.

[0089] 본 개시에 따른 음성 구간 검출 장치(80)는 시간 영역이 아닌 주파수 영역의 성분을 이용하므로, 음성 구간 검출의 정확도가 증가될 수 있다.

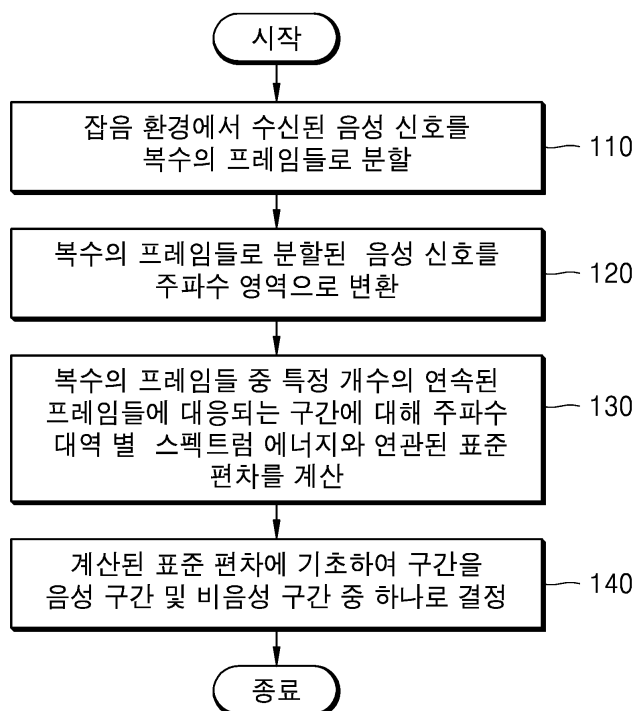
[0090] 또한, 본 개시에 따른 음성 구간 검출 장치(80)는 계산된 표준편차와 기 설정된 임계값과의 비교만을 통해 음성 구간 검출을 수행하므로, 심층 신경망 계산 등과 같은 부가적인 연산이 필요하지 않고, 고성능의 하드웨어를 갖추지 않더라도 실시간으로 음성 구간 검출이 가능하다.

[0091] 또한, 본 개시에 따른 음성 구간 검출 장치(80)는 주파수 대역 별 스펙트럼 에너지와 연관된 표준편차로서, 2차 표준편차를 이용하여 기 설정된 임계값과의 비교를 수행함으로써, 음성 신호의 잡음 수준(예를 들어, SNR)에 상관없이 단일한 임계값을 적용할 수 있다. 이에 따라, 음성 구간 검출 장치(80)는 다양한 잡음 수준을 갖는 잡음 환경에서 음성 구간 검출을 위해 이용되는 범용성을 가질 수 있다.

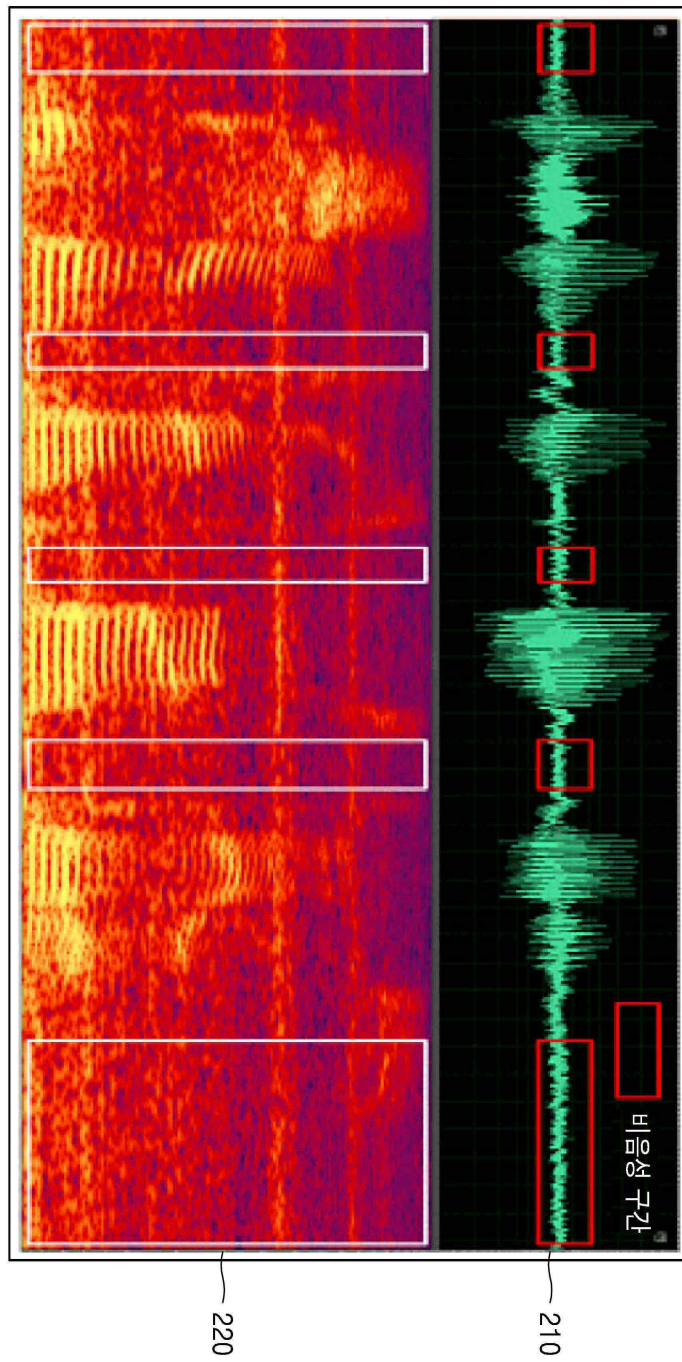
[0092] 이상에서 실시예들에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속한다.

도면

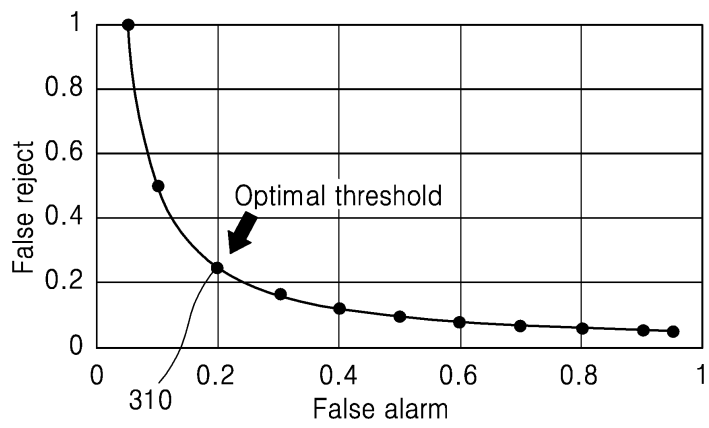
도면1



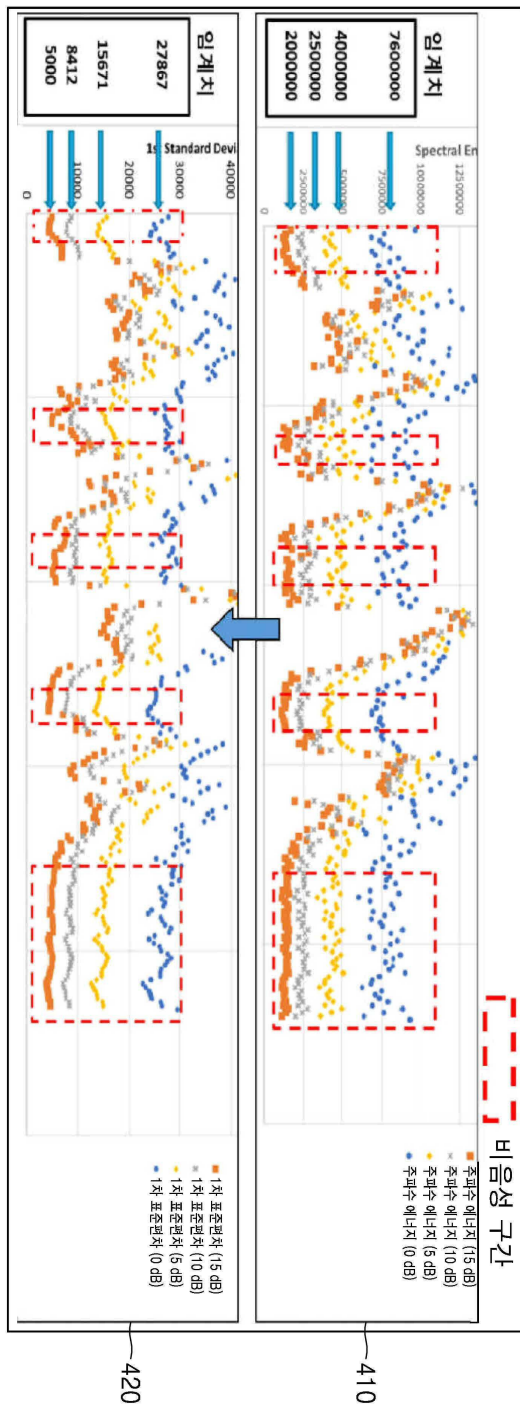
도면2



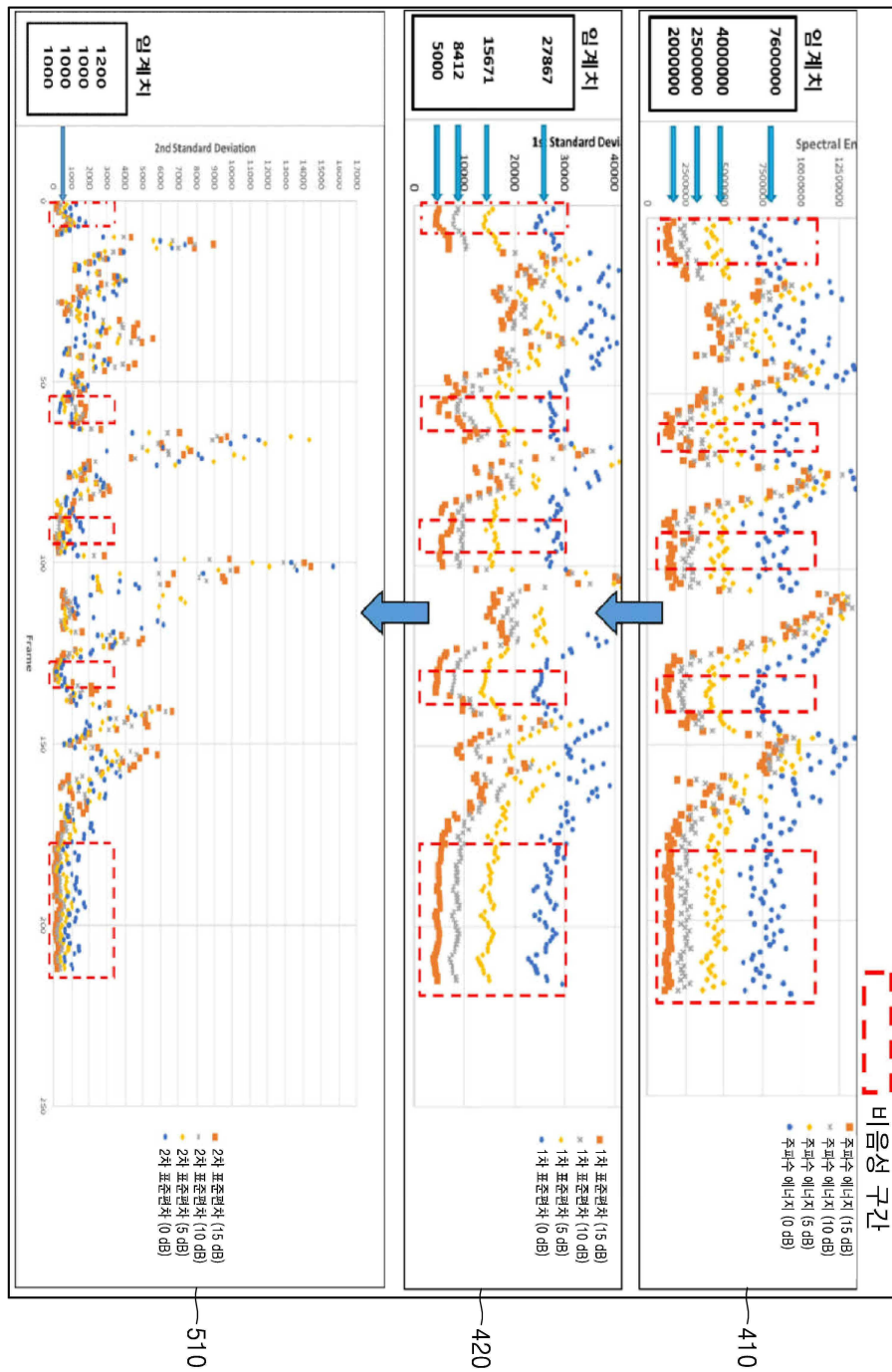
도면3



도면4



도면5



도면6

THR	주파수 에너지 기반 VAD		THR	시간 영역 에너지 기반 VAD	
	FR	FA		FR	FA
θ_1	11	92	θ_1	19	111
θ_2	23	73	θ_2	27	83
θ_3	32	42	θ_3	36	73
θ_4	37	39	θ_4	44	60

도면7

	주파수 에너지 기반 VAD	시간 영역 에너지 기반 VAD
SNR(dB)	FR	FR
0	19	32
5	17	30
10	9	24
15	5	19

도면8

