



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2017년12월14일
 (11) 등록번호 10-1808810
 (24) 등록일자 2017년12월07일

(51) 국제특허분류(Int. Cl.)
 G10L 25/78 (2013.01)
 (21) 출원번호 10-2013-0144979
 (22) 출원일자 2013년11월27일
 심사청구일자 2016년10월06일
 (65) 공개번호 10-2015-0061669
 (43) 공개일자 2015년06월05일
 (56) 선행기술조사문헌
 KR1020080051042 A
 KR1020090002842 A

(73) 특허권자
 한국전자통신연구원
 대전광역시 유성구 가정로 218 (가정동)
 (72) 발명자
 장인선
 대전광역시 유성구 배울2로 78 607동 1101호 (관평동, 운암네오피아)
 임우택
 서울 강남구 광평로19길 10, 1006동 1313호 (수서동, 까치마을아파트)
 (74) 대리인
 성병기, 최윤서

전체 청구항 수 : 총 12 항

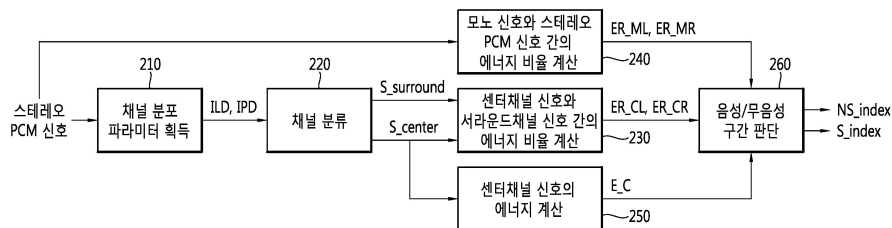
심사관 : 정성윤

(54) 발명의 명칭 **음성/무음성 구간 검출 방법 및 장치**

(57) 요약

음성/무음성 구간 검출 방법 및 장치를 제공한다. 음성/무음성 구간 검출 장치는 스테레오 오디오 신호에서 상기 스테레오 오디오 신호의 채널간 관계 정보를 획득하는 획득부, 상기 채널간 관계 정보를 기초로 상기 스테레오 오디오 신호의 각 요소를 센터채널 성분과 서라운드 성분으로 분류하는 분류부, 상기 센터채널 성분인 요소들로 구성된 센터채널 신호와 상기 서라운드 성분인 요소들로 구성된 서라운드채널 신호 간의 프레임별 에너지 비율 값 및 상기 스테레오 오디오 신호를 기초로 생성된 모노 신호와 상기 스테레오 오디오 신호 간의 프레임별 에너지 비율 값을 계산하는 계산부 및 상기 에너지 비율 값들을 비교하여 상기 스테레오 오디오 신호에서 음성 구간과 무음성 구간을 판단하는 판단부를 포함할 수 있다.

대표도



명세서

청구범위

청구항 1

스테레오 오디오 신호에서 상기 스테레오 오디오 신호의 채널간 관계 정보를 획득하는 획득부;

상기 채널간 관계 정보를 기초로 상기 스테레오 오디오 신호의 각 요소를 센터채널 성분과 서라운드 성분으로 분류하는 분류부;

상기 센터채널 성분인 요소들로 구성된 센터채널 신호와 상기 서라운드 성분인 요소들로 구성된 서라운드채널 신호 간의 프레임별 에너지 비율 값 및 상기 스테레오 오디오 신호를 기초로 생성된 모노 신호와 상기 스테레오 오디오 신호 간의 프레임별 에너지 비율 값을 계산하는 계산부; 및

상기 에너지 비율 값들을 비교하여 상기 스테레오 오디오 신호에서 음성 구간과 무음성 구간을 판단하는 판단부를 포함하는 음성/무음성 구간 검출 장치.

청구항 2

제1항에 있어서,

상기 채널간 관계 정보는,

상기 스테레오 오디오 신호의 채널간 레벨 차이에 대한 정보와 채널간 위상 차이에 대한 정보를 포함하는 것을 특징으로 하는 음성/무음성 구간 검출 장치.

청구항 3

제2항에 있어서,

상기 채널간 관계 정보는,

상기 스테레오 오디오 신호의 채널간 상호연관성 정보를 더 포함하는 것을 특징으로 하는 음성/무음성 구간 검출 장치.

청구항 4

제1항에 있어서,

상기 센터채널 신호는 상기 센터채널 성분인 요소들을 이용하여 역스펙트로그램(inverse spectrogram)을 수행함으로써 생성되고, 상기 서라운드채널 신호는 상기 서라운드 성분인 요소들을 이용하여 역스펙트로그램을 수행함으로써 생성되는 것을 특징으로 하는 음성/무음성 구간 검출 장치.

청구항 5

제1항에 있어서,

상기 판단부는,

상기 센터채널 신호의 프레임별 에너지 값을 기초로 상기 음성 구간으로서 검출한 구간에서의 에너지 값이 임계 값 보다 큰 경우 상기 검출한 구간을 음성 구간으로 판단하는 것을 특징으로 하는 음성/무음성 구간 검출 장치.

청구항 6

음성/무음성 구간 검출 장치에 의한 음성/무음성 구간 검출 방법에 있어서,

스테레오 오디오 신호에서 상기 스테레오 오디오 신호의 채널간 관계 정보를 획득하는 단계;

상기 채널간 관계 정보를 기초로 센터채널 성분인 요소들로 구성된 센터채널 신호와 서라운드 성분인 요소들로 구성된 서라운드채널 신호를 생성하는 단계;

상기 센터채널 신호와 상기 서라운드채널 신호 간의 프레임별 에너지 비율 값 및 상기 스테레오 오디오 신호를 기초로 생성된 모노 신호와 상기 스테레오 오디오 신호 간의 프레임별 에너지 비율 값을 계산하는 단계; 및 상기 에너지 비율 값들을 비교하여 상기 스테레오 오디오 신호에서 음성 구간과 무음성 구간을 검출하는 단계를 포함하는 음성/무음성 구간 검출 방법.

청구항 7

제6항에 있어서,
상기 채널간 관계 정보는,
상기 스테레오 오디오 신호의 채널간 레벨 차이에 대한 정보와 채널간 위상 차이에 대한 정보를 포함하는 것을 특징으로 하는 음성/무음성 구간 검출 방법.

청구항 8

제7항에 있어서,
상기 채널간 관계 정보는,
상기 스테레오 오디오 신호의 채널간 상호연관성 정보를 더 포함하는 것을 특징으로 하는 음성/무음성 구간 검출 방법.

청구항 9

제6항에 있어서,
상기 획득하는 단계 이후에,
상기 채널간 관계 정보를 기초로 상기 스테레오 오디오 신호의 각 요소를 센터채널 성분과 서라운드 성분으로 분류하는 단계를 더 포함하는 것을 특징으로 하는 음성/무음성 구간 검출 방법.

청구항 10

제6항에 있어서,
상기 생성하는 단계는,
상기 센터채널 성분인 요소들을 이용하여 역스펙트로그램(inverse spectrogram)을 수행함으로써 상기 센터채널 신호를 생성하는 단계; 및
상기 서라운드 성분인 요소들을 이용하여 역스펙트로그램을 수행함으로써 상기 서라운드채널 신호를 생성하는 단계를 포함하는 것을 특징으로 하는 음성/무음성 구간 검출 방법.

청구항 11

제6항에 있어서,
상기 계산하는 단계는,
상기 센터채널 신호와 상기 서라운드채널 신호의 프레임별 에너지 값을 각각 계산하고 상기 센터채널 신호와 상기 서라운드 채널 신호의 프레임별 에너지 값을 기초로 상기 센터채널 신호와 상기 서라운드채널 신호 간의 프레임별 에너지 비율 값을 계산하는 단계; 및
상기 스테레오 오디오 신호를 기초로 생성된 모노 신호와 상기 스테레오 오디오 신호의 프레임별 에너지 값을 각각 계산하고 상기 모노 신호와 상기 스테레오 오디오 신호의 프레임별 에너지 값을 기초로 상기 모노 신호와 상기 스테레오 오디오 신호 간의 프레임별 에너지 비율 값을 계산하는 단계를 포함하는 것을 특징으로 하는 음성/무음성 구간 검출 방법.

청구항 12

제6항에 있어서,

상기 검출하는 단계는,

상기 센터채널 신호의 프레임별 에너지 값을 기초로 상기 음성 구간으로서 검출한 구간에서의 에너지 값이 임계 값 보다 큰 경우 상기 검출한 구간을 음성 구간으로 판단하고 상기 검출한 구간에서의 에너지 값이 상기 임계값 이하인 경우 상기 검출한 구간을 비음성 구간으로 판단하는 것을 포함하는 음성/무음성 구간 검출 방법.

발명의 설명

기술 분야

[0001] 본 발명의 실시예들은 음성, 음악, 음향 효과 및 잡음 등이 혼재되어 있는 미디어 콘텐츠에서 음성/무음성 구간을 검출하는 방법 및 그 장치에 관한 것이다.

배경 기술

[0002] 종래에는 미디어 콘텐츠에서 음성 구간(speech section)과 음성이 아닌 구간(non-speech section)을 검출하기 위하여 다양한 음성 활성화 측정(Voice Activity Detection) 방법이 사용되어 왔다.

[0003] 일 예로, 한국공개특허공보 특1999-0039422호(공개일 1999년 6월 5일) "G.729 음성 부호화기를 위한 음성 활성화 측정 방법"에는 음성 프레임을 음성정보가 포함된 실음성 구간과 묵음 구간으로 구분하고, 실음성 구간을 다시 유성음과 무성음으로 구분하여 각각 부호화한 후 LPC 파라미터 추출 과정에서 얻어진 음성 프레임의 에너지를 문턱값과 비교하여 음성의 활성도를 측정하는 것이 개시되어 있다.

[0004] 또한, 한국공개특허공보 제10-2013-0085731호(공개일 2013년 7월 30일) "음성 영역 검출 방법 및 그 장치"에는 음성 프레임 간의 자기 상관값을 이용하여 음성 데이터 내의 음성 구간과 묵음 구간을 판별하는 것이 개시되어 있다.

[0005] 그러나, 이러한 종래의 방법들은 단순히 임계치를 이용하여 음성 구간을 검출하므로 에러가 발생할 수 있으며, 잡음이 섞이면서 특징 벡터들이 크게 변함에 따라 정확한 음성 구간의 검출이 어려운 단점이 있다. 또한, 종래의 방법들은 음성과 묵음을 판별하는 것이기 때문에 음악 및 음향 효과 등이 공존하는 미디어 콘텐츠에 적용하는 것은 무리가 있다.

[0006] 한편, 음성과 음악을 구분하는 기술은 대부분 음성 인식 시스템의 성능 향상을 위한 전처리 기술로 개발되고 있다. 기존의 음성/음악 분류 방법을 살펴보면 음악의 주요 특성이라고 생각될 수 있는 시간에 따른 리듬 변화를 이용하여 음성과 음악을 구분하는 방법들이 제안되어 있다. 그러나, 이러한 방법들은 대체로 음악은 음성의 변화에 비해 상대적으로 느리며 비교적 일정한 간격으로 변한다는 원리를 사용하였기 때문에, 음악의 종류에 따라 템포가 빨라지거나 약기가 달라지는 것에 따라 그 성능이 크게 변할 수밖에 없다.

[0007] 이외에 음성 및 음악 데이터베이스(DB: database)를 활용하여 통계적으로 음성/음악 분류 특성을 갖는 특징 벡터를 추출하고 이를 기반으로 학습(training)한 분류기(Classifier)를 이용하여 음성/음악을 분류하는 방법들이 연구되고 있다. 그러나, 이러한 방법들은 높은 성능의 음성/음악 분류를 위해 학습 단계가 필수적이며, 학습을 위해 많은 양의 데이터를 확보하고 이를 기반으로 통계적 특징 벡터를 추출해야 하므로 데이터 확보, 유효 특징 벡터 추출 및 학습 시간에 따른 많은 수고와 시간이 필요하다.

발명의 내용

해결하려는 과제

[0008] 본 발명의 기술적 과제는 사전 학습(training) 없이 오디오 신호에서 음성/무음성 구간을 검출할 수 있는 음성/무음성 구간 검출 방법 및 장치를 제공함에 있다.

[0009] 본 발명의 다른 기술적 과제는 적은 계산량과 메모리 소요(cost)로도 오디오 신호에서 정확하게 음성/무음성 구간을 검출할 수 있는 음성/무음성 구간 검출 방법 및 장치를 제공함에 있다.

과제의 해결 수단

- [0010] 본 발명의 일 양태에 따르면, 음성/무음성 구간 검출 장치는 스테레오 오디오 신호에서 상기 스테레오 오디오 신호의 채널간 관계 정보를 획득하는 획득부, 상기 채널간 관계 정보를 기초로 상기 스테레오 오디오 신호의 각 요소를 센터채널 성분과 서라운드 성분으로 분류하는 분류부, 상기 센터채널 성분인 요소들로 구성된 센터채널 신호와 상기 서라운드 성분인 요소들로 구성된 서라운드채널 신호 간의 프레임별 에너지 비율 값 및 상기 스테레오 오디오 신호를 기초로 생성된 모노 신호와 상기 스테레오 오디오 신호 간의 프레임별 에너지 비율 값을 계산하는 계산부 및 상기 에너지 비율 값들을 비교하여 상기 스테레오 오디오 신호에서 음성 구간과 무음성 구간을 판단하는 판단부를 포함할 수 있다.
- [0011] 일 측면에 따르면, 상기 채널간 관계 정보는 상기 스테레오 오디오 신호의 채널간 레벨 차이에 대한 정보와 채널간 위상 차이에 대한 정보를 포함할 수 있다.
- [0012] 다른 측면에 따르면, 상기 채널간 관계 정보는 상기 스테레오 오디오 신호의 채널간 상호연관성 정보를 더 포함할 수 있다.
- [0013] 또 다른 측면에 따르면, 상기 센터채널 신호는 상기 센터채널 성분인 요소들을 이용하여 역스펙트로그램(inverse spectrogram)을 수행함으로써 생성되고, 상기 서라운드채널 신호는 상기 서라운드 성분인 요소들을 이용하여 역스펙트로그램을 수행함으로써 생성될 수 있다.
- [0014] 또 다른 측면에 따르면, 상기 판단부는 상기 센터채널 신호의 프레임별 에너지 값을 기초로 상기 음성 구간으로서 검출한 구간에서의 에너지 값이 상기 임계값 보다 큰 경우 상기 검출한 구간을 음성 구간으로 판단할 수 있다.
- [0015] 본 발명의 다른 양태에 따르면, 음성/무음성 구간 검출 장치에 의한 음성/무음성 구간 검출 방법은 스테레오 오디오 신호에서 상기 스테레오 오디오 신호의 채널간 관계 정보를 획득하는 단계, 상기 채널간 관계 정보를 기초로 센터채널 성분인 요소들로 구성된 센터채널 신호와 서라운드 성분인 요소들로 구성된 서라운드채널 신호를 생성하는 단계, 상기 센터채널 신호와 상기 서라운드채널 신호 간의 프레임별 에너지 비율 값 및 상기 스테레오 오디오 신호를 기초로 생성된 모노 신호와 상기 스테레오 오디오 신호 간의 프레임별 에너지 비율 값을 계산하는 단계 및 상기 에너지 비율 값들을 비교하여 상기 스테레오 오디오 신호에서 음성 구간과 무음성 구간을 검출하는 단계를 포함할 수 있다.

발명의 효과

- [0016] 음성 및 음악에 대한 DB 확보, 통계적 유효 특징 추출 및 사전 학습(training) 등의 시간적, 인적 수고 없이도 오디오 신호에서 음성/무음성 구간을 검출할 수 있다.
- [0017] 오디오 채널 간 특성 및 채널 별 신호의 특성 분석을 위한 간단한 계산량 및 적은 메모리 소요로도 정확한 음성/무음성 구간 검출이 가능하며, 음향 편집 장치, 데이터 검색 방법의 전처리 등에 적용되어 해당 장치의 서비스 품질을 향상시킬 수 있다.

도면의 간단한 설명

- [0018] 도 1은 본 발명의 일실시예에 있어서, 음성/무음성 구간 검출 장치를 나타내는 블록도이다.
- 도 2는 본 발명의 일실시예에 따른 음성/무음성 구간 검출 과정을 설명하기 위한 도면이다.
- 도 3은 본 발명의 일실시예에 따른 음성/무음성 구간에 대한 판별 기준을 나타내는 의사 코드이다.
- 도 4는 본 발명의 일실시예에 있어서, 음성/무음성 구간 검출 방법을 나타내는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0019] 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.

- [0020] 명세서 전체에서, 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. 또한, 명세서에 기재된 "~부" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.
- [0021] 도 1은 본 발명의 일실시예에 있어서, 음성/무음성 구간 검출 장치를 나타내는 블록도이다. 도 1을 참조하면, 본 발명의 일 실시예에 따른 음성/무음성 구간 검출 장치(100)는 획득부(110), 분류부(120), 계산부(130) 및 판단부(140)를 포함한다.
- [0022] 획득부(110)는 오디오 신호에서 상기 오디오 신호의 채널간 관계 정보를 획득한다. 이를 위하여 획득부(110)는 오디오 신호를 입력 받을 수 있다. 상기 오디오 신호는 일 예로, 복수개의 채널을 포함하는 스테레오 신호일 수 있다. 상기 채널간 관계 정보는 상기 오디오 신호의 채널간 레벨 차이(ILD: Inter-channel Level Difference)에 대한 정보와 채널간 위상 차이(IPD: Inter-channel Phase Difference)에 대한 정보를 포함할 수 있다. 또한, 필요에 따라 상기 채널간 관계 정보는 상기 오디오 신호의 채널간 상호연관성(ICC: Inter-Channel Correlation) 정보를 더 포함할 수 있다.
- [0023] 채널간 관계 정보는 STFT(Short Time Fourier Transform)된 좌채널 신호와 우채널 신호를 복소수 스펙트로그램(spectrogram) 행렬로 보았을 때, 특정 프레임 및 주파수 값을 가지는 하나의 요소(element)에 대해 계산된다. 획득부(110)는 오디오 신호의 각 요소 별로 ILD, IPD 등을 추출함으로써 채널간 관계 정보를 획득할 수 있다.
- [0024] 분류부(120)는 획득부(110)에서 획득한 채널간 관계 정보를 기초로 오디오 신호의 각 요소를 센터채널 성분과 서라운드 성분으로 분류한다. 일 예로, 분류부(120)는 각 요소마다의 ILD, IPD가 특정 임계치보다 작으면 센터채널 성분으로 판단하고, 크면 서라운드 성분으로 판단함으로써 상기 각 요소를 분류할 수 있다. 이후, 분류부(120)는 센터채널 성분인 요소들과 서라운드 성분인 요소들을 각각 모아 역 스펙트로그램(inverse spectrogram)을 수행함으로써 센터채널 신호와 서라운드채널 신호를 생성함으로써 오디오 신호를 센터채널 신호와 서라운드채널 신호로 분리한다.
- [0025] 계산부(130)는 분류부(120)로부터 출력된 센터채널 신호와 서라운드채널 신호 간의 프레임별 에너지 비율 값을 계산하고, 상기 오디오 신호를 기초로 생성된 모노 신호와 상기 오디오 신호 간의 프레임별 에너지 비율 값을 계산한다. 이를 위하여 계산부(130)는 센터채널 신호와 서라운드채널 신호의 프레임별 에너지 값을 각각 계산하고, 상기 센터채널 신호와 상기 서라운드 채널 신호의 프레임별 에너지 값을 기초로 상기 센터채널 신호와 상기 서라운드채널 신호 간의 프레임별 에너지 비율 값을 계산하는 한편, 상기 오디오 신호를 기초로 모노 신호를 생성하고 상기 모노 신호와 상기 오디오 신호의 프레임별 에너지 값을 각각 계산한 후 상기 모노 신호와 상기 오디오 신호의 프레임별 에너지 값을 기초로 상기 모노 신호와 상기 오디오 신호 간의 프레임별 에너지 비율 값을 계산할 수 있다.
- [0026] 판단부(140)는 계산부(130)에서 계산된 에너지 비율 값들을 비교하여 오디오 신호에서 음성 구간과 무음성 구간을 판단한다. 일 예로, 판단부(140)는 상기 센터채널 신호와 서라운드채널 신호 간의 에너지 비율 값이 상기 모노 신호와 상기 오디오 신호의 프레임별 에너지 값 보다 큰 경우 해당 구간을 일차적으로 음성 구간으로서 검출할 수 있다. 여기서, 상기 모노 신호와 상기 오디오 신호의 프레임별 에너지 값은 임계치 설정을 위한 이득 값이 부가된 후 상기 센터채널 신호와 서라운드채널 신호 간의 에너지 비율 값과 비교될 수 있다. 그리고, 판단부(140)는 계산부(130)에서 계산된 센터채널 신호의 프레임별 에너지 값을 기초로 상기 음성 구간으로서 검출한 구간에서의 에너지 값이 상기 임계값 보다 큰 경우 상기 검출한 구간을 음성 구간으로 판단할 수 있다.
- [0027] 도 2는 본 발명의 일실시예에 따른 음성/무음성 구간 검출 과정을 설명하기 위한 도면이고, 도 3은 본 발명의 일실시예에 따른 음성/무음성 구간에 대한 판별 기준을 나타내는 의사 코드이다.
- [0028] 먼저 도 2를 참조하면, 스테레오 신호는 획득부(110)로 입력될 수 있다. 그러면, 획득부(110)는 입력된 스테레오 신호에서 복수 채널 간의 관계 정보로 채널간 레벨 차이(ILD)와 채널 간 위상 차이(IPD) 정보를 추출함으로써 채널 분포 파라미터를 획득한다(210). 음성/무음성 구간 판단에는 경우에 따라 채널 간 상호연관성(ICC) 정보 등 채널 간의 정보를 표현하는 데 사용될 수 있는 다양한 파라미터가 활용될 수 있다. 상기 채널 분포 파라미터는 STFT(Short Time Fourier Transform)된 좌채널 신호와 우채널 신호를 복소수 스펙트로그램 행렬로 보았

을 때, 특정 프레임 및 주파수 값을 가지는 하나의 요소에 대해 계산된다. 이후, 획득부(110)는 각 요소에 따른 ILD, IPD 등을 출력하고, 출력된 요소 별 ILD, IPD는 분류부(120)에 입력된다.

[0029] 분류부(120)는 각 요소에 대해 ILD, IPD가 특정 임계치보다 작으면 센터 채널 성분으로 분류하고, 특정 임계치보다 크면 서라운드 성분으로 분류한다(220). 이후, 센터채널인 요소들과 서라운드 성분인 요소들을 각각 모아 역 스펙트로그램을 수행함으로써 센터채널 신호(S_center)와 서라운드채널 신호(S_surround)를 구성한 후 출력한다. 그러면, 계산부(130)는 센터채널 신호(S_center)와 서라운드채널 신호(S_surround)의 프레임별 에너지 값을 계산하고, 계산된 프레임별 에너지의 비율 값을 다음의 수학적 식 1을 이용하여 계산한다(230).

수학적 식 1

$$ER_CL[i] = E(S_center[i]) / E(LS_surround[i]),$$

$$ER_CR[i] = E(S_center[i]) / E(RS_surround[i])$$

[0030]

[0031] 여기서, ER_CL[i], ER_CR[i]은 각각 i-번째 프레임에서의 센터채널 신호와 좌서라운드(Left Surround) 신호 간의 에너지 비율 값 및 센터채널 신호와 우서라운드(Right Surround) 신호 간의 에너지 비율 값을 나타낸다. E(.)는 에너지 값을 계산하는 함수이며, LS_surround와 RS_surround는 각각 서라운드채널 신호의 좌채널 신호 및 우채널 신호를 나타낸다.

[0032] 한편, 계산부(130)는 스테레오 신호를 입력 받아 모노 신호를 생성한다. 그리고, 생성한 모노 신호와 스테레오 신호의 프레임별 에너지 값을 계산하고, 계산된 프레임별 에너지의 비율 값을 다음의 수학적 식 2를 이용하여 계산한다(240).

수학적 식 2

$$ER_ML[i] = E(M[i]) / E(L[i]),$$

$$ER_MR[i] = E(M[i]) / E(R[i])$$

[0033]

[0034] 여기서, ER_ML[i], ER_MR[i]은 i-번째 프레임에서의 각각 모노 신호(M)와 스테레오 신호 내 좌채널 신호(L) 간의 에너지 비율 값 및 모노 신호(M)와 스테레오 신호 내 우채널 신호의(R) 간의 에너지 비율 값을 나타낸다. E(.)는 에너지 값을 계산하는 함수이며, 다음의 수학적 식 3과 같이 계산된다.

수학적 식 3

$$E(L[i]) = \frac{1}{N} \sum_{k=1}^N L(k)$$

[0035]

[0036] 여기서, k는 샘플 인덱스이며, N은 프레임 길이이다.

[0037] 또한, 계산부(130)는 다음의 수학적 식 4를 이용하여 센터채널 신호(S_center)의 프레임별로 에너지 값을 계산한다

(250).

수학적 식 4

$$E_C[i] = E(S_center[i])$$

[0038]

여기서, $E_C[i]$ 는 i -번째 프레임에서의 센터채널 신호의 에너지 값을 나타낸다.

[0039]

판단부(140)는 우선 입력받은 에너지 비율 값(ER_CL , ER_ML , ER_CR , ER_MR)을 비교하여 음성/무음성 구간을 검출한다. 일반적으로, 음성과 같이 사용자에게 중요 정보를 주는 음원은 센터채널에 위치한다. 따라서, 판단부(140)는 ER_CL 이 ER_ML 보다 크거나 ER_CR 이 ER_MR 보다 큰 경우 해당 구간을 음성 구간으로 판단할 수 있다(260).

[0040]

일 예로, 실제 방송 콘텐츠 제작 시 오디오는 모노 혹은 스테레오 마이크를 사용하여 현장 녹음되며, 녹화 후 프로듀서가 녹화본을 보며 음악 추가 및 효과음 증폭 등 스튜디오에서의 믹싱 작업을 수행함으로써 프로그램이 제작된다. 현장 녹화에서는 붐 마이크 등 초지향성 혹은 지향성 마이크를 사용하여 배우의 목소리를 중심으로 녹음하므로, 음성 신호는 방송 콘텐츠 내 센터 채널에 분포하게 된다.

[0041]

스튜디오에서는 현장 녹화 오디오에 스테레오 음악 및 음향 효과를 추가한다. 그러므로, 음성에 해당하는 프레임에서는 센터채널 신호와 서라운드채널 신호 간의 에너지 비율이 모노 신호와 스테레오 신호 간의 에너지 비율에 비해 크다. 반면, 음악 등 스튜디오에서의 믹싱 작업을 통해 추가된 음성이 아닌 신호의 경우에는 센터채널 신호와 서라운드채널 신호 간의 에너지 비율이 모노 신호와 스테레오 신호 간의 에너지 비율에 비해 작게 된다. 이는 생방송으로 제작되는 뉴스 프로그램도 마찬가지이다. 판단부(140)는 이를 기준으로 일차적으로 음성 구간 여부를 판별하며, 음성 구간으로 판별된 경우 센터채널 음성에 위치하는 음성의 활성 정도를 좀 더 정확하게 판단하고자 프레임 별 에너지 값을 계산하여 특정 프레임에서의 에너지 값이 임계값 보다 큰 경우 음성 구간으로 판단하고, 임계값 보다 작은 경우 무음성 구간으로 판단할 수 있다.

[0042]

음성/무음성 구간 판별 기준의 의사 코드(pseudo code)는 도 3과 같다. 도 3에서 알파(alpha)는 에너지 비율 임계치 설정을 위한 이득값을 나타내고, 베타(beta)는 프레임별 에너지의 임계값을 나타낸다. 판단부(140)는 도 3의 판별 기준에 따라 음성 구간 여부를 판단하고 그 결과를 출력할 수 있다.

[0043]

도 4는 본 발명의 실시시에 있어서, 음성/무음성 구간 검출 방법을 나타내는 흐름도이다.

[0044]

음성/무음성 구간 검출 장치는 오디오 신호에서 음성 구간과 무음성 구간을 검출하기 위하여 먼저 오디오 신호에서 ILD, IPD 등을 추출함으로써 상기 오디오 신호의 채널간 관계 정보를 획득한다(410). 여기서, 상기 오디오 신호는 복수개의 채널을 포함하는 스테레오 신호일 수 있다. 음성/무음성 구간 검출 장치는 필요에 따라 상기 채널간 관계 정보로서 채널간 상호연관성 정보를 추출할 수도 있다.

[0045]

이후, 음성/무음성 구간 검출 장치는 추출된 채널간 관계 정보를 기초로 오디오 신호의 각 요소를 센터채널 성분과 서라운드 성분으로 분류하여 센터채널 성분인 요소들로 구성된 센터채널 신호(S_center)와 서라운드 성분인 요소들로 구성된 서라운드채널 신호($S_surround$)를 생성한다(420). 이 때, 상기 센터채널 신호(S_center)와 상기 서라운드채널 신호($S_surround$)는 각각 센터채널 성분인 요소들을 이용하여 역스펙트로그램(inverse spectrogram)을 수행하고 상기 서라운드 성분인 요소들을 이용하여 역스펙트로그램을 수행함으로써 생성될 수 있다.

[0046]

센터채널 신호(S_center)와 서라운드채널 신호($S_surround$)가 생성되면, 음성/무음성 구간 검출 장치는 센터채널 신호와 서라운드채널 신호 간의 프레임별 에너지 비율 값(ER_CL , ER_CR) 및 상기 오디오 신호를 기초로 생성된 모노 신호와 상기 오디오 신호 간의 프레임별 에너지 비율 값(ER_ML , ER_MR)을 계산한다.

[0047]

구체적으로, 음성/무음성 구간 검출 장치는 센터채널 신호(S_center)와 서라운드채널 신호($S_surround$)의 프레임별 에너지 값을 각각 계산하고, 계산한 프레임별 에너지 값들을 기초로 상기 센터채널 신호와 상기 서라운드채널 신호 간의 프레임별 에너지 비율 값(ER_CL , ER_CR)을 계산할 수 있다(430). 또한, 상기 오디오 신호를 기

[0048]

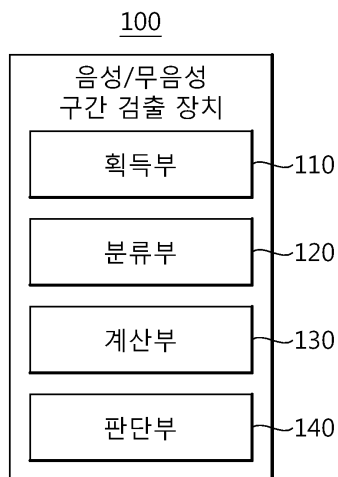
초로 생성된 모노 신호와 상기 오디오 신호의 프레임별 에너지 값을 각각 계산하고, 계산한 프레임별 에너지 값을 기초로 상기 모노 신호와 상기 오디오 신호 간의 프레임별 에너지 비율 값(ER_ML, ER_MR)을 계산할 수 있다(440).

[0049] 상술한 과정을 통해 각 신호들 간의 에너지 비율 값들(ER_CL, ER_CR, ER_ML, ER_MR)이 계산되면, 음성/무음성 구간 검출 장치는 상기 에너지 비율 값들(ER_CL, ER_CR, ER_ML, ER_MR)을 비교하여 일차적으로 오디오 신호에서 음성 구간과 무음성 구간을 검출한다(450). 그리고, 센터채널 신호의 프레임별 에너지 값(E_C)을 기초로 음성 구간으로서 검출한 구간에서의 에너지 값이 상기 임계값 보다 큰 경우 상기 검출한 구간을 음성 구간으로 판단하고, 상기 검출한 구간에서의 에너지 값이 상기 임계값 이하인 경우 상기 검출한 구간을 비음성 구간으로 판단한다(460).

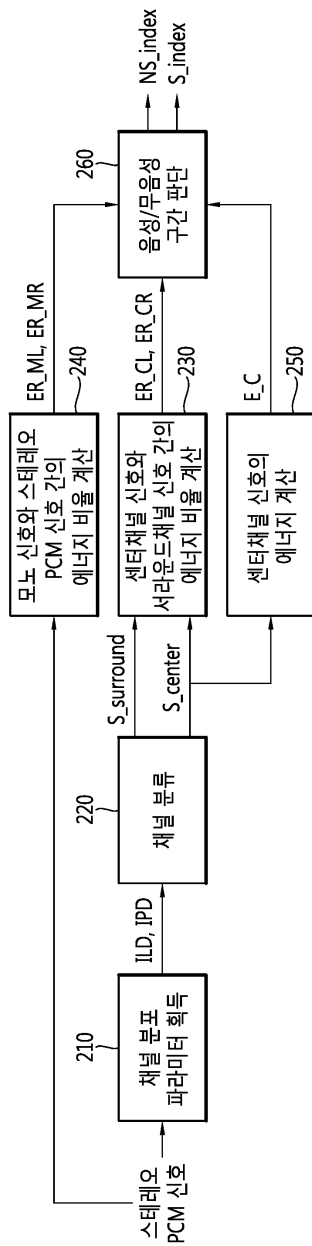
[0050] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 발명에 개시된 실시 예들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시 예에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

도면

도면1



도면2



도면3

```

if (ER_CL[i] > alpha * ER_ML[i]) OR (ER_CR[i] > alpha * ER_MR[i])
    if (E_C[i] > beta)
        Speech
    else
        non-Speech
    end
else
    non-Speech
end
    
```

도면4

