



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2019년01월03일
 (11) 등록번호 10-1935183
 (24) 등록일자 2018년12월27일

(51) 국제특허분류(Int. Cl.)
G10L 21/0316 (2013.01) *G10L 21/0272*
 (2013.01)
 (52) CPC특허분류
G10L 21/0316 (2013.01)
G10L 21/0272 (2013.01)
 (21) 출원번호 10-2017-7007107
 (22) 출원일자(국제) 2014년12월12일
 심사청구일자 2017년03월15일
 (85) 번역문제출일자 2017년03월15일
 (65) 공개번호 10-2017-0042709
 (43) 공개일자 2017년04월19일
 (86) 국제출원번호 PCT/EP2014/077620
 (87) 국제공개번호 WO 2016/091332
 국제공개일자 2016년06월16일
 (56) 선행기술조사문헌
 W02009035615 A1*
 *는 심사관에 의하여 인용된 문헌

(73) 특허권자
후아웨이 테크놀러지 컴퍼니 리미티드
 중국 518129 광둥성 셴젠 룡강 디스트릭트 반티안
 후아웨이 어드미니스트레이션 빌딩
 (72) 발명자
가이저, 위르겐
 독일 80992 뮌헨 리에스트라세 25 화웨이 테크놀
 로지즈 뒤셀도르프 게엠베하 내
그로셰, 페터
 독일 80992 뮌헨 리에스트라세 25 화웨이 테크놀
 로지즈 뒤셀도르프 게엠베하 내
 (74) 대리인
양영준, 김성운, 백만기

전체 청구항 수 : 총 15 항

심사관 : 정성운

(54) 발명의 명칭 **멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 신호 처리 장치**

(57) 요약

본 발명은 멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 신호 처리 장치에 관한 것으로, 멀티-채널 오디오 신호는 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 포함하고, 신호 처리 장치는 필터 및 조합기를 포함하고; 필터는 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호에 기초하여 주파수에 걸쳐 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 결정하고, 중심 채널 오디오 신호의 크기의 척도와 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도 사이의 비에 기초하여 이득 함수를 획득하고, 좌 채널 오디오 신호를 이득 함수로 가중하여 가중된 좌 채널 오디오 신호를 획득하고, 중심 채널 오디오 신호를 이득 함수로 가중하여 가중된 중심 채널 오디오 신호를 획득하고, 우 채널 오디오 신호를 이득 함수로 가중하여 가중된 우 채널 오디오 신호를 획득하도록 구성되고; 조합기는 좌 채널 오디오 신호를 가중된 좌 채널 오디오 신호와 조합하여 조합된 좌 채널 오디오 신호를 획득하고, 중심 채널 오디오 신호를 가중된 중심 채널 오디오 신호와 조합하여 조합된 중심 채널 오디오 신호를 획득하고, 우 채널 오디오 신호를 가중된 우 채널 오디오 신호와 조합하여 조합된 우 채널 오디오 신호를 획득하도록 구성된다.

명세서

청구범위

청구항 1

멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 신호 처리 장치로서, 상기 멀티-채널 오디오 신호는 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 포함하고, 상기 신호 처리 장치는 필터 및 조합기를 포함하고,

상기 필터는

상기 좌 채널 오디오 신호, 상기 중심 채널 오디오 신호, 및 상기 우 채널 오디오 신호에 기초하여 주파수에 걸쳐(over frequency) 상기 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도(measure)를 결정하고,

상기 중심 채널 오디오 신호의 크기의 척도와 상기 멀티-채널 오디오 신호의 상기 전체 크기를 나타내는 척도 사이의 비에 기초하여 제1 이득을 획득하고,

상기 좌 채널 오디오 신호를 상기 제1 이득으로 가중하여 제1 가중된 좌 채널 오디오 신호를 획득하고, 상기 중심 채널 오디오 신호를 상기 제1 이득으로 가중하여 제1 가중된 중심 채널 오디오 신호를 획득하고, 상기 우 채널 오디오 신호를 상기 제1 이득으로 가중하여 제1 가중된 우 채널 오디오 신호를 획득하도록 구성되고,

상기 조합기는

상기 좌 채널 오디오 신호, 상기 중심 채널 오디오 신호, 및 상기 우 채널 오디오 신호를 각각 미리 결정된 제2 이득으로 가중하여 제2 가중된 좌 채널 오디오 신호, 제2 가중된 중심 채널 오디오 신호, 및 제2 가중된 우 채널 오디오 신호를 획득하고,

상기 제1 가중된 좌 채널 오디오 신호를 상기 제2 가중된 좌 채널 오디오 신호와 조합하여 조합된 좌 채널 오디오 신호를 획득하고, 상기 제1 가중된 중심 채널 오디오 신호를 상기 제2 가중된 중심 채널 오디오 신호와 조합하여 조합된 중심 채널 오디오 신호를 획득하고, 상기 제1 가중된 우 채널 오디오 신호를 상기 제2 가중된 우 채널 오디오 신호와 조합하여 조합된 우 채널 오디오 신호를 획득하도록 구성되는, 신호 처리 장치.

청구항 2

제1항에 있어서, 상기 필터는 상기 멀티-채널 오디오 신호의 상기 전체 크기를 나타내는 척도를 상기 중심 채널 오디오 신호의 크기의 척도와 상기 좌 채널 오디오 신호와 상기 우 채널 오디오 신호의 차이의 크기의 척도의 합으로서 결정하도록 구성되는, 신호 처리 장치.

청구항 3

제1항에 있어서, 상기 필터는 다음 식들:

$$G(m, k) = \frac{P_C(m, k)}{P_C(m, k) + P_S(m, k)}$$

$$P_C(m, k) = |C(m, k)|^2$$

$$P_S(m, k) = |L(m, k) - R(m, k)|^2$$

에 따라 상기 제1 이득을 결정하도록 구성되고, G는 상기 제1 이득을 나타내고, L은 상기 좌 채널 오디오 신호를 나타내고, C는 상기 중심 채널 오디오 신호를 나타내고, R은 상기 우 채널 오디오 신호를 나타내고, P_C는 상기 중심 채널 오디오 신호의 크기를 나타내는 척도로서 상기 중심 채널 오디오 신호의 전력을 나타내고, P_S는 상기 좌 채널 오디오 신호와 상기 우 채널 오디오 신호 사이의 차이의 전력을 나타내고, P_C와 P_S의 합은 상기 멀티-채널 오디오 신호의 상기 전체 크기를 나타내는 척도를 나타내고, m은 샘플 시간 인덱스(sample time index)를 나타내고, k는 주파수 빈 인덱스(frequency bin index)를 나타내는 신호 처리 장치.

청구항 4

제1항에 있어서, 상기 멀티-채널 오디오 신호는 좌 서라운드 채널 오디오 신호 및 우 서라운드 채널 오디오 신호를 더 포함하고,

상기 필터는

상기 좌 서라운드 채널 오디오 신호 및 상기 우 서라운드 채널 오디오 신호에 기초하여 부가적으로 주파수에 걸쳐 상기 멀티-채널 오디오 신호의 상기 전체 크기를 나타내는 척도를 결정하고,

상기 멀티-채널 오디오 신호의 상기 전체 크기를 나타내는 척도를 상기 중심 채널 오디오 신호의 크기의 척도와, 상기 좌 채널 오디오 신호와 상기 우 채널 오디오 신호의 차이의 크기의 척도와, 상기 좌 서라운드 채널 오디오 신호와 상기 우 서라운드 채널 오디오 신호의 차이의 크기의 척도의 합으로서 결정하도록 구성되는, 신호 처리 장치.

청구항 5

제1항에 있어서,

상기 좌 채널 오디오 신호, 상기 중심 채널 오디오 신호, 및 상기 우 채널 오디오 신호에 기초하여 음성 활성화도 표시자(voice activity indicator)를 결정하도록 구성되는 음성 활성화도 검출기를 더 포함하고, 상기 음성 활성화도 표시자는 시간에 걸쳐(over time) 상기 멀티-채널 오디오 신호 내의 상기 음성 성분의 크기를 표시하고,

상기 조합기는 상기 제1 가중된 좌 채널 오디오 신호를 상기 음성 활성화도 표시자와 조합하여 상기 조합된 좌 채널 오디오 신호를 획득하고, 상기 제1 가중된 중심 채널 오디오 신호를 상기 음성 활성화도 표시자와 조합하여 상기 조합된 중심 채널 오디오 신호를 획득하고, 상기 제1 가중된 우 채널 오디오 신호를 상기 음성 활성화도 표시자와 조합하여 상기 조합된 우 채널 오디오 신호를 획득하도록 더 구성되는, 신호 처리 장치.

청구항 6

제5항에 있어서, 상기 음성 활성화도 검출기는

상기 좌 채널 오디오 신호, 상기 중심 채널 오디오 신호, 및 상기 우 채널 오디오 신호에 기초하여 상기 멀티-채널 오디오 신호의 전체 스펙트럼 변화(overall spectral variation)를 나타내는 척도를 결정하고,

상기 중심 채널 오디오 신호의 스펙트럼 변화의 척도와 상기 멀티-채널 오디오 신호의 상기 전체 스펙트럼 변화를 나타내는 척도 사이의 비에 기초하여 상기 음성 활성화도 표시자를 획득하도록 구성되는, 신호 처리 장치.

청구항 7

제6항에 있어서, 상기 음성 활성화도 검출기는 다음 식:

$$V = a \times \left(\frac{F_c}{F_c + F_s} - 0.5 \right)$$

에 따라 상기 음성 활성화도 표시자를 결정하도록 구성되고, V는 상기 음성 활성화도 표시자를 나타내고, F_c는 상기 중심 채널 오디오 신호의 스펙트럼 변화의 척도를 나타내고, F_s는 상기 좌 채널 오디오 신호와 상기 우 채널 오디오 신호 사이의 차이의 스펙트럼 변화의 척도를 나타내고, F_c와 F_s의 합은 상기 멀티-채널 오디오 신호의 상기 전체 스펙트럼 변화를 나타내는 척도를 나타내고, a는 미리 결정된 스케일링 팩터(predetermined scaling factor)를 나타내는 신호 처리 장치.

청구항 8

제7항에 있어서, 상기 음성 활성화도 검출기는 다음 식들:

$$F_C(m) = \sum_k \left(|C(m, k)| - |C(m-1, k)| \right)^2$$

$$F_S(m) = \sum_k \left(|S(m, k)| - |S(m-1, k)| \right)^2$$

에 따라 상기 중심 채널 오디오 신호의 스펙트럼 변화의 척도를 스펙트럼 플럭스(spectral flux)로서 그리고 상기 좌 채널 오디오 신호와 상기 우 채널 오디오 신호 사이의 차이의 스펙트럼 변화의 척도(F_S)를 스펙트럼 플럭스로서 결정하도록 구성되고, F_C 는 상기 중심 채널 오디오 신호의 상기 스펙트럼 플럭스를 나타내고, F_S 는 상기 좌 채널 오디오 신호와 상기 우 채널 오디오 신호 사이의 차이의 상기 스펙트럼 플럭스를 나타내고, C 는 상기 중심 채널 오디오 신호를 나타내고, S 는 상기 좌 채널 오디오 신호와 상기 우 채널 오디오 신호 사이의 차이를 나타내고, m 은 샘플 시간 인덱스를 나타내고, k 는 주파수 빈 인덱스를 나타내는 신호 처리 장치.

청구항 9

제5항 내지 제8항 중 어느 한 항에 있어서, 상기 음성 활성화도 검출기는 미리 결정된 저역 통과 필터링 기능에 기초하여 시간에서 상기 음성 활성화도 표시자를 필터링하도록 구성되는, 신호 처리 장치.

청구항 10

제5항에 있어서, 상기 조합기는 음성 활성화도 표시자를 미리 결정된 음성 이득 팩터(speech gain factor)(G_S)로 가중하도록 더 구성되는, 신호 처리 장치.

청구항 11

제5항에 있어서, 상기 조합기는 상기 제1 가중된 좌 채널 오디오 신호와 음성 활성화도 표시자의 조합에 상기 제2 가중된 좌 채널 오디오 신호를 가산하여 상기 조합된 좌 채널 오디오 신호를 획득하고, 상기 제1 가중된 중심 채널 오디오 신호와 상기 음성 활성화도 표시자의 조합에 상기 제2 가중된 중심 채널 오디오 신호를 가산하여 상기 조합된 중심 채널 오디오 신호를 획득하고, 상기 제1 가중된 우 채널 오디오 신호와 상기 음성 활성화도 표시자의 조합에 상기 제2 가중된 우 채널 오디오 신호를 가산하여 상기 조합된 우 채널 오디오 신호를 획득하도록 구성되는, 신호 처리 장치.

청구항 12

제1항에 있어서,

입력 좌 채널 스테레오 오디오 신호(L_{in}) 및 입력 우 채널 스테레오 오디오 신호(R_{in})에 기초하여 상기 좌 채널 오디오 신호, 상기 중심 채널 오디오 신호, 및 상기 우 채널 오디오 신호를 결정하도록 구성되는 업-믹서(up-mixer), 및

상기 조합된 좌 채널 오디오 신호, 상기 조합된 중심 채널 오디오 신호, 및 상기 조합된 우 채널 오디오 신호에 기초하여 출력 좌 채널 스테레오 오디오 신호(L_{out}) 및 출력 우 채널 스테레오 오디오 신호(R_{out})를 결정하도록 구성되는 다운-믹서(down-mixer) 중 적어도 하나를 더 포함하는 신호 처리 장치.

청구항 13

제1항에 있어서, 상기 크기의 척도는 신호의 전력, 대수 전력(logarithmic power), 크기 또는 대수 크기(logarithmic magnitude)를 포함하는 신호 처리 장치.

청구항 14

멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 신호 처리 방법으로서, 상기 멀티-채널 오디오 신호는 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 포함하고, 상기 신호 처리 방법은

상기 좌 채널 오디오 신호, 상기 중심 채널 오디오 신호, 및 상기 우 채널 오디오 신호에 기초하여 주파수에 걸쳐 상기 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 결정하는 단계,

상기 중심 채널 오디오 신호의 크기의 척도와 상기 멀티-채널 오디오 신호의 상기 전체 크기를 나타내는 척도 사이의 비에 기초하여 제1 이득을 획득하는 단계,

상기 좌 채널 오디오 신호를 상기 제1 이득으로 가중하여 제1 가중된 좌 채널 오디오 신호를 획득하는 단계,

상기 중심 채널 오디오 신호를 상기 제1 이득으로 가중하여 제1 가중된 중심 채널 오디오 신호를 획득하는 단계,

상기 우 채널 오디오 신호를 상기 제1 이득으로 가중하여 제1 가중된 우 채널 오디오 신호를 획득하는 단계,

상기 좌 채널 오디오 신호, 상기 중심 채널 오디오 신호, 및 상기 우 채널 오디오 신호를 각각 미리 결정된 제2 이득으로 가중하여 제2 가중된 좌 채널 오디오 신호, 제2 가중된 중심 채널 오디오 신호, 및 제2 가중된 우 채널 오디오 신호를 획득하는 단계;

상기 제1 가중된 좌 채널 오디오 신호를 상기 제2 가중된 좌 채널 오디오 신호와 조합하여 조합된 좌 채널 오디오 신호를 획득하는 단계,

상기 제1 가중된 중심 채널 오디오 신호를 상기 제2 가중된 중심 채널 오디오 신호와 조합하여 조합된 중심 채널 오디오 신호를 획득하는 단계, 및

상기 제1 가중된 우 채널 오디오 신호를 상기 제2 가중된 우 채널 오디오 신호와 조합하여 조합된 우 채널 오디오 신호를 획득하는 단계

를 포함하는 신호 처리 방법.

청구항 15

컴퓨터 상에서 실행될 때 제14항의 방법을 수행하는 프로그램 코드를 포함하고 컴퓨터 판독가능 기록매체에 저장되는 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 오디오 신호 처리의 분야에 관한 것으로, 특히 멀티-채널 오디오 신호들 내의 음성 향상에 관한 것이다.

배경 기술

[0002] 멀티-채널 오디오 신호들, 예를 들어 엔터테인먼트 오디오 신호들 내의 음성 성분을 향상시키기 위해, 상이한 방식들이 현재 이용되고 있다.

[0003] 음성 성분을 향상시키는 간단한 방식은 멀티-채널 오디오 신호로 구성된 중심 채널 오디오 신호를 부스팅(boost)시키거나, 또는 그에 따라 다른 채널들의 모든 오디오 신호를 감쇠(attenuate)시키는 것이다. 이 방식은 음성이 전형적으로 중심 채널 오디오 신호로 패닝된다는 가정을 이용한다. 그러나, 이 방식은 일반적으로 음성 향상의 낮은 성능을 갖고 있다.

[0004] 보다 정교한 방식이 별개의 채널들의 오디오 신호들을 분석하기를 시도한다. 이와 관련하여, 중심 채널 오디오 신호와 다른 채널들의 오디오 신호들 사이의 관계에 관한 정보가 음성 향상을 가능하게 하기 위해 스테레오 다운-믹스와 함께 제공될 수 있다. 그러나, 이 방식은 스테레오 오디오 신호들에 적용될 수 없고 별개의 음성 오디오 채널을 필요로 한다.

[0005] 부드러운 음성 성분들의 레벨을 개선하고 멀티-채널 오디오 신호 내의 큰 비음성 성분들을 감쇠시키는 또 다른 방식은 동적 범위 압축(DRC)이다. 첫째, 이 방식은 큰 음량 성분들을 감쇠시키는 것을 포함한다. 다음에, 전체 음량 레벨이 증가되어, 음성 또는 대화 부스트(voice or dialogue boost)를 초래한다. 그러나, 이 방식은 멀티-채널 오디오 신호의 특성을 감안하지 않으며 변형은 음량 레벨에 대해서만 관련된다.

발명의 내용

- [0006] 본 발명의 목적은 멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 효율적인 개념을 제공하는 것이다.
- [0007] 이 목적은 독립 청구항들의 특징들에 의해 달성된다. 추가 구현 형태가 종속 청구항들, 설명 및 도면으로부터 명백하다.
- [0008] 본 발명은 멀티-채널 오디오 신호가 멀티-채널 오디오 신호의 모든 채널들로부터 결정될 수 있는, 이득 함수에 기초하여 필터링될 수 있다는 발견에 기초한다. 필터링은 위너(Wiener) 필터링 방식에 기초할 수 있고, 멀티-채널 오디오 신호의 중심 채널 오디오 신호는 음성 성분을 포함하는 것으로 간주될 수 있고, 멀티-채널 오디오 신호의 다른 채널들은 비음성 성분들을 포함하는 것으로 간주될 수 있다. 시간에 걸쳐 멀티-채널 오디오 신호 내의 음성 성분의 변화를 고려하기 위해서, 음성 활성화도 검출이 더 수행될 수 있고, 멀티-채널 오디오 신호의 모든 채널들이 음성 활성화도 표시자를 제공하기 위해 처리될 수 있다. 멀티-채널 오디오 신호는 입력 스테레오 오디오 신호의 스테레오 업-믹싱 처리의 결과일 수 있다. 결과적으로, 멀티-채널 오디오 신호 내의 음성 성분의 효율적인 향상이 실현될 수 있다.
- [0009] 제1 양태에 따르면, 본 발명은 멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 신호 처리 장치에 관한 것으로, 멀티-채널 오디오 신호는 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 포함하고, 신호 처리 장치는 필터 및 조합기를 포함하고, 필터는 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호에 기초하여 주파수에 걸쳐 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 결정하고, 중심 채널 오디오 신호의 크기의 척도와 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도 사이의 비에 기초하여 이득 함수를 획득하고, 좌 채널 오디오 신호를 이득 함수로 가중하여 가중된 좌 채널 오디오 신호를 획득하고, 중심 채널 오디오 신호를 이득 함수로 가중하여 가중된 중심 채널 오디오 신호를 획득하고, 우 채널 오디오 신호를 이득 함수로 가중하여 가중된 우 채널 오디오 신호를 획득하도록 구성되고, 조합기는 좌 채널 오디오 신호를 가중된 좌 채널 오디오 신호와 조합하여 조합된 좌 채널 오디오 신호를 획득하고, 중심 채널 오디오 신호를 가중된 중심 채널 오디오 신호와 조합하여 조합된 중심 채널 오디오 신호를 획득하고, 우 채널 오디오 신호를 가중된 우 채널 오디오 신호와 조합하여 조합된 우 채널 오디오 신호를 획득하도록 구성된다. 그러므로, 멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 효율적인 개념이 실현된다.
- [0010] 멀티-채널 오디오 신호는 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 포함한다. 멀티-채널 오디오 신호는 좌 서라운드 채널 오디오 신호 및 우 서라운드 채널 오디오 신호를 더 포함할 수 있다. 멀티-채널 오디오 신호는 LCR/3.0 스테레오 오디오 신호 또는 5.1 서라운드 오디오 신호일 수 있다. 주파수에 걸쳐 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 결정하는 것은 주파수 영역 내의 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 결정하는 것을 포함한다.
- [0011] 이득 함수는 음성 성분의 크기와 멀티-채널 오디오 신호의 전체 크기의 비를 표시할 수 있고, 음성 성분은 중심 채널 오디오 신호로 구성된다고 가정한다. 멀티-채널 오디오 신호의 전체 크기는 주파수에 걸쳐 멀티-채널 오디오 신호 내의 음성 성분 및 비음성 성분의 가산을 사용하여 결정될 수 있다. 이득 함수는 주파수 종속적일 수 있다.
- [0012] 이와 같은 제1 양태에 따른 신호 처리 장치의 제1 구현 형태에서, 필터는 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 중심 채널 오디오 신호의 크기의 척도와 좌 채널 오디오 신호와 우 채널 오디오 신호의 차이의 크기의 척도의 합으로서 결정하도록 구성된다. 그러므로, 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도가 효율적으로 그리고 더욱 적합한 방식으로 결정되어 필터 이득 함수를 획득하기 위해 사용되는데, 왜냐하면 좌 채널 오디오 신호와 우 채널 오디오 신호의 차이는 중심 채널 오디오 신호의 성분들을 포함하지 않는 잔여 신호를 나타내기 때문이다.
- [0013] 이와 같은 제1 양태 또는 제1 양태의 임의의 앞선 구현 형태에 따른 신호 처리 장치의 제2 구현 형태에서, 필터는 다음 식들:

$$G(m, k) = \frac{P_C(m, k)}{P_C(m, k) + P_S(m, k)}$$

$$P_C(m, k) = |C(m, k)|^2$$

$$P_S(m, k) = |L(m, k) - R(m, k)|^2$$

[0014]

[0015]

에 따라 이득 함수를 결정하도록 구성되고, G는 이득 함수를 나타내고, L은 좌 채널 오디오 신호를 나타내고, C는 중심 채널 오디오 신호를 나타내고, R은 우 채널 오디오 신호를 나타내고, P_C는 중심 채널 오디오 신호의 크기를 나타내는 척도로서 중심 채널 오디오 신호의 전력을 나타내고, P_S는 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이의 전력을 나타내고, P_C와 P_S의 합은 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 나타내고, m은 샘플 시간 인덱스를 나타내고, k는 주파수 빈 인덱스를 나타낸다. 그러므로, 이득 함수는 효율적이고 강력한 방식으로 결정된다.

[0016]

이득 함수는 위너 필터링 방식에 따라 결정된다. 중심 채널 오디오 신호는 음성 성분을 포함하는 것으로 간주된다. 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이는 음성 성분들이 중심 채널 오디오 신호로 패닝된다는 가정에 기초하여, 비음성 성분을 포함하는 것으로 간주된다. 위너 필터의 성분들을 이 방식으로 정의함으로써, 신호 대 잡음 비 또는 신호의 잡음 전력 스펙트럼 밀도를 평가하는 고가의 방법들을 이용하는 것이 피해진다.

[0017]

식들 내의 전력을 사용하는 것 대신에, 크기 또는 대수 전력이 이득 함수를 결정하기 위해 이용될 수 있다. 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이는 비중심 채널 오디오 신호들의 조합을 포함하는 잔여 오디오 신호라고 할 수 있고, 중심 채널 오디오 신호를 제외한 모든 오디오 신호들은 또한 비중심 채널 오디오 신호들이라고 할 수 있다. 잔여 오디오 신호는 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이일 수 있다.

[0018]

좌 채널 오디오 신호와 우 채널 오디오 신호의 크기의 합은 중심 채널 추출의 특정 형태인 빔-형성에 대응하고, 본 발명의 실시예들에서 또한 사용될 수 있다. 그러나, 좌 채널 오디오 신호와 우 채널 오디오 신호의 크기의 차이는 중심 채널의 성분의 제거에 대응한다. 그러므로, 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이로서 정의된 잔여 오디오 신호는 필터 이득의 개선된 평가를 가져다 준다.

[0019]

이와 같은 제1 양태 또는 제1 양태의 임의의 앞선 구현 형태에 따른 신호 처리 장치의 제3 구현 형태에서, 멀티-채널 오디오 신호는 좌 서라운드 채널 오디오 신호 및 우 서라운드 채널 오디오 신호를 더 포함하고, 필터는 좌 서라운드 채널 오디오 신호 및 우 서라운드 채널 오디오 신호에 기초하여 부가적으로 주파수에 걸쳐 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 결정하고, 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 중심 채널 오디오 신호의 크기의 척도와 좌 채널 오디오 신호와 우 채널 오디오 신호의 차이의 크기의 척도와 좌 서라운드 채널 오디오 신호와 우 서라운드 채널 오디오 신호의 차이의 크기의 척도의 합으로서 결정하도록 구성된다. 그러므로, 멀티-채널 오디오 신호 내의 서라운드 채널들은 좌 서라운드 채널 오디오 신호와 우 서라운드 채널 오디오 신호의 차이로부터 크기를 획득함으로써, 효율적으로 처리된다. 이 차이 신호는 중심 채널 오디오 신호에 보다 양호한 구별을 제공한다.

[0020]

이와 같은 제1 양태 또는 제1 양태의 임의의 앞선 구현 형태에 따른 신호 처리 장치의 제4 구현 형태에서, 필터는 좌 채널 오디오 신호의 주파수 빈들을 이득 함수의 주파수 빈들로 가중하여 가중된 좌 채널 오디오 신호의 주파수 빈들을 획득하고, 중심 채널 오디오 신호의 주파수 빈들을 이득 함수의 주파수 빈들로 가중하여 가중된 중심 채널 오디오 신호의 주파수 빈들을 획득하고, 우 채널 오디오 신호의 주파수 빈들을 이득 함수의 주파수 빈들로 가중하여 가중된 우 채널 오디오 신호의 주파수 빈들을 획득하도록 구성된다. 그러므로, 멀티-채널 오디오 신호는 주파수 영역에서 효율적으로 처리된다. 동일한 필터로 모든 신호들을 가중하는 것은 스테레오 이미지 내의 오디오 소스 위치들의 시프팅이 발생하지 않는다는 장점을 갖는다. 더구나, 이 방식으로, 음성 성분 이 모든 신호들로부터 추출된다.

[0021]

필터는 주파수 대역들을 획득하기 위해 멜 주파수 스케일(Mel frequency scale)에 따라 주파수 빈들을 그룹화하도록 더 구성될 수 있다. 인덱스 k는 결과적으로 주파수 대역 인덱스에 대응할 수 있다. 필터는 미리 결정된 주파수 범위, 예를 들어, 100Hz 내지 8kHz 내에 배열된 주파수 빈들 또는 주파수 대역들을 단지 처리하도록 더

구성될 수 있다. 이 방식으로, 사람의 음성을 포함하는 주파수들만이 처리된다.

[0022] 이와 같은 제1 양태 또는 제1 양태의 임의의 앞선 구현 형태에 따른 신호 처리 장치의 제5 구현 형태에서, 신호 처리 장치는 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호에 기초하여 음성 활성화도 표시자를 결정하도록 구성되는 음성 활성화도 검출기를 더 포함하고, 음성 활성화도 표시자는 시간에 걸쳐 멀티-채널 오디오 신호 내의 음성 성분의 크기를 표시하고, 조합기는 가중된 좌 채널 오디오 신호를 음성 활성화도 표시자와 조합하여 조합된 좌 채널 오디오 신호를 획득하고, 가중된 중심 채널 오디오 신호를 음성 활성화도 표시자와 조합하여 조합된 중심 채널 오디오 신호를 획득하고, 가중된 우 채널 오디오 신호를 음성 활성화도 표시자와 조합하여 조합된 우 채널 오디오 신호를 획득하도록 더 구성된다. 그러므로, 멀티-채널 오디오 신호 내의 시변 음성 성분의 효율적인 향상이 실현되고, 비음성 신호들이 억제된다.

[0023] 음성 활성화도 표시자는 시간 영역에서의 멀티-채널 오디오 신호 내의 음성 성분의 크기를 표시한다. 음성 활성화도 표시자는 예를 들어, 음성 성분이 신호 내에 존재하지 않을 때 0이고, 음성이 존재할 때 1이다. 0과 1 사이의 값들은 음성이 존재하는 확률로서 해석될 수 있고, 매끄러운 출력 신호를 획득하는 데 도움을 준다.

[0024] 제1 양태의 제5 구현 형태에 따른 신호 처리 장치의 제6 구현 형태에서, 음성 활성화도 검출기는 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호에 기초하여 멀티-채널 오디오 신호의 전체 스펙트럼 변화를 나타내는 척도를 결정하고, 중심 채널 오디오 신호의 스펙트럼 변화의 척도와 멀티-채널 오디오 신호의 전체 스펙트럼 변화를 나타내는 척도 사이의 비에 기초하여 음성 활성화도 표시자를 획득하도록 구성된다. 그러므로, 음성 활성화도 표시자는 스펙트럼 변화의 척도들 사이의 관계를 이용함으로써 효율적으로 결정된다.

[0025] 전체 스펙트럼 변화를 나타내는 척도는 스펙트럼 플럭스 또는 시간 도함수일 수 있다. 스펙트럼 플럭스는 정규화를 위한 상이한 방식들을 사용하여 결정될 수 있다. 스펙트럼 플럭스는 2개 이상의 오디오 신호 프레임들 사이의 전력 스펙트럼들의 차이로서 계산될 수 있다. 전체 스펙트럼 변화를 나타내는 척도는 F_c 와 F_s 의 합일 수 있고, 여기서 F_c 는 중심 채널 오디오 신호의 스펙트럼 변화의 척도를 나타내고, F_s 는 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이의 스펙트럼 변화의 척도를 나타낸다.

[0026] 제1 양태의 제6 구현 형태에 따른 신호 처리 장치의 제7 구현 형태에서, 음성 활성화도 검출기는 다음 식:

$$V = a \times \left(\frac{F_c}{F_c + F_s} - 0.5 \right)$$

[0027] 에 따라 음성 활성화도 표시자를 결정하도록 구성되고, V 는 음성 활성화도 표시자를 나타내고, F_c 는 중심 채널 오디오 신호의 스펙트럼 변화의 척도를 나타내고, F_s 는 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이의 스펙트럼 변화의 척도를 나타내고, F_c 와 F_s 의 합은 멀티-채널 오디오 신호의 전체 스펙트럼 변화를 나타내는 척도를 나타내고, a 는 미리 결정된 스케일링 팩터를 나타낸다. 그러므로, 음성 활성화도 표시자가 효율적으로 결정된다. F_c 와 F_s 가 동일한 값들을 갖는 신호들이라면 0의 값을 갖는 음성 활성화도 표시자가 야기된다. F_c 의 값이 높을수록 음성 활성화도 표시자의 값들이 높아진다. 스케일링 팩터는 음성 활성화도 표시자의 크기를 제어할 수 있다.

[0029] 음성 활성화도 표시자의 값들은 척도들의 이전의 정규화에 독립적일 수 있다. 음성 활성화도 표시자의 값들은 간격 [0; 1]로 제한될 수 있다.

[0030] 제1 양태의 제7 구현 형태에 따른 신호 처리 장치의 제8 구현 형태에서, 음성 활성화도 검출기는 다음 식들:

$$F_c(m) = \sum_k \left(|C(m, k)| - |C(m-1, k)| \right)^2$$

$$F_s(m) = \sum_k \left(|S(m, k)| - |S(m-1, k)| \right)^2$$

[0031] 에 따라 스펙트럼 플럭스로서 중심 채널 오디오 신호들의 스펙트럼 변화의 척도 및 스펙트럼 플럭스로서 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이의 스펙트럼 변화의 척도를 결정하도록 구성되고, F_c 는 중심 채널 오디오 신호의 스펙트럼 플럭스를 나타내고, F_s 는 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이

의 스펙트럼 플렉스를 나타내고, C는 중심 채널 오디오 신호를 나타내고, S는 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이를 나타내고, m은 샘플 시간 인덱스를 나타내고, k는 주파수 빈 인덱스를 나타낸다. 그러므로, 스펙트럼 플렉스가 효율적으로 결정된다.

- [0033] 제1 양태의 제5 구현 형태 내지 제8 구현 형태에 따른 신호 처리 장치의 제9 구현 형태에서, 음성 활성화도 검출기는 미리 결정된 저역 통과 필터링 기능에 기초하여 시간에서 음성 활성화도 표시자를 필터링하도록 구성된다. 그러므로, 멀티-채널 오디오 신호 내의 아티팩트들의 효율적인 완화 및/또는 음성 활성화도 표시자의 효율적인 시간적 평활화가 실현된다.
- [0034] 미리 결정된 저역 통과 필터링 기능이 윈-탭 유한 임펄스 응답(FIR) 저역 통과 필터에 의해 실현될 수 있다.
- [0035] 제1 양태의 제5 구현 형태 내지 제9 구현 형태에 따른 신호 처리 장치의 제10 구현 형태에서, 조합기는 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 미리 결정된 입력 이득 팩터로 가중하고, 음성 활성화도 표시자를 미리 결정된 음성 이득 팩터로 가중하도록 더 구성된다. 그러므로, 비음성 성분의 크기에 관련된 음성 성분의 크기의 효율적인 제어가 실현된다.
- [0036] 제1 양태의 제5 구현 형태 내지 제10 구현 형태에 따른 신호 처리 장치의 제11 구현 형태에서, 조합기는 가중된 좌 채널 오디오 신호와 음성 활성화도 표시자의 조합에 좌 채널 오디오 신호를 가산하여 조합된 좌 채널 오디오 신호를 획득하고, 가중된 좌 채널 오디오 신호와 음성 활성화도 표시자의 조합에 중심 채널 오디오 신호를 가산하여 조합된 중심 채널 오디오 신호를 획득하고, 가중된 좌 채널 오디오 신호와 음성 활성화도 표시자의 조합에 우 채널 오디오 신호를 가산하여 조합된 우 채널 오디오 신호를 획득하도록 구성된다. 그러므로, 조합기가 효율적으로 구현된다. 추출된 음성 성분들은 출력 신호들 내의 음성 성분을 향상시키기 위해 원래의 신호들과 조합된다.
- [0037] 제1 양태의 제5 구현 형태 내지 제11 구현 형태에 따른 신호 처리 장치의 제12 구현 형태에서, 멀티-채널 오디오 신호는 좌 서라운드 채널 오디오 신호 및 우 서라운드 채널 오디오 신호를 더 포함하고, 음성 활성화도 검출기는 좌 서라운드 채널 오디오 신호 및 우 서라운드 채널 오디오 신호에 기초하여 부가적으로 음성 활성화도 표시자를 결정하도록 구성된다. 그러므로, 멀티-채널 오디오 신호 내의 서라운드 채널들이 또한 음성 활성화도 표시자를 결정하기 위해 고려되어, 음성 활성화도 표시자의 더 양호한 평가를 가져다 준다.
- [0038] 이와 같은 제1 양태 또는 제1 양태의 임의의 앞선 구현 형태에 따른 신호 처리 장치의 제13 구현 형태에서, 신호 처리 장치는 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 시간 영역으로부터 주파수 영역으로 변환하도록 구성되는 변환기를 더 포함한다. 그러므로, 오디오 신호들의 주파수 영역으로의 효율적인 변환이 실현된다. 이것은 음성 향상 및 음성 활성화도 검출이 주파수 영역에서 수행되는 경우에 요구될 수 있다.
- [0039] 변환기는 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호의 단-시간 이산 푸리에 변환(STFT)을 수행하도록 구성될 수 있다.
- [0040] 이와 같은 제1 양태 또는 제1 양태의 임의의 앞선 구현 형태에 따른 신호 처리 장치의 제14 구현 형태에서, 신호 처리 장치는 조합된 좌 채널 오디오 신호, 조합된 중심 채널 오디오 신호, 및 조합된 우 채널 오디오 신호를 주파수 영역으로부터 시간 영역으로 역 변환하도록 구성되는 역 변환기를 더 포함한다. 그러므로, 오디오 신호들의 시간 영역으로의 효율적인 역 변환이 실현되고, 시간 영역에서의 출력 신호들이 획득된다.
- [0041] 역 변환기는 조합된 좌 채널 오디오 신호, 조합된 중심 채널 오디오 신호, 및 조합된 우 채널 오디오 신호의 역 단-시간 이산 푸리에 변환(ISTFT)을 수행하도록 구성될 수 있다.
- [0042] 이와 같은 제1 양태 또는 제1 양태의 임의의 앞선 구현 형태에 따른 신호 처리 장치의 제15 구현 형태에서, 신호 처리 장치는 입력 좌 채널 스테레오 오디오 신호 및 입력 우 채널 스테레오 오디오 신호에 기초하여 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 결정하도록 구성되는 업-믹서를 더 포함한다. 이 방식으로, 신호 처리 장치는 2-채널, 즉 좌 및 우 채널, 입력 스테레오 오디오 신호를 처리하기 위해 적용될 수 있다.

[0043] 제1 양태의 제15 구현 형태에 따른 신호 처리 장치의 제16 구현 형태에서, 업-믹서는 다음 식들:

$$C = \alpha \times (L_{in} + R_{in})$$

$$L = L_{in} - C$$

$$R = R_{in} - C$$

$$\alpha = \frac{1}{2} \times \left(1 - \sqrt{\frac{(L_r - R_r)^2 + (L_i - R_i)^2}{(L_r + R_r)^2 + (L_i + R_i)^2}} \right)$$

[0044]

[0045]

에 따라 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 결정하도록 구성되고, L_r 은 입력 좌 채널 스테레오 오디오 신호의 실수부를 나타내고, R_r 은 입력 우 채널 스테레오 오디오 신호의 실수부를 나타내고, L_i 는 입력 좌 채널 스테레오 오디오 신호의 허수부를 나타내고, R_i 는 입력 우 채널 스테레오 오디오 신호의 허수부를 나타내고, α 는 직교성 파라미터를 나타내고, L_{in} 은 입력 좌 채널 스테레오 오디오 신호를 나타내고, R_{in} 은 입력 우 채널 스테레오 오디오 신호를 나타내고, L 은 좌 채널 오디오 신호를 나타내고, C 는 중심 채널 오디오 신호를 나타내고, R 은 우 채널 오디오 신호를 나타낸다. 그러므로, 입력 스테레오 오디오 신호의 효율적인 중심 채널 추출이 직교 분해를 사용하여 실현된다. 결과적인 좌 채널 오디오 신호와 우 채널 오디오 신호는 서로 직교한다.

[0046]

이와 같은 제1 양태 또는 제1 양태의 임의의 앞선 구현 형태에 따른 신호 처리 장치의 제17 구현 형태에서, 신호 처리 장치는 조합된 좌 채널 오디오 신호, 조합된 중심 채널 오디오 신호, 및 조합된 우 채널 오디오 신호에 기초하여 출력 좌 채널 스테레오 오디오 신호 및 출력 우 채널 스테레오 오디오 신호를 결정하도록 구성되는 다운-믹서를 더 포함한다. 그러므로, 2-채널, 즉 좌 및 우 채널, 출력 스테레오 오디오 신호가 효율적으로 제공된다.

[0047]

이와 같은 제1 양태 또는 제1 양태의 임의의 앞선 구현 형태에 따른 신호 처리 장치의 제18 구현 형태에서, 크기의 척도는 신호의 전력, 대수 전력, 크기 또는 대수 크기를 포함한다. 그러므로, 크기의 척도는 상이한 스케일들에서 상이한 값들을 표시할 수 있다.

[0048]

멀티-채널 오디오 신호의 크기는 멀티-채널 오디오 신호의 전력, 대수 전력, 크기 또는 대수 크기를 포함한다. 좌 채널 오디오 신호와 우 채널 오디오 신호의 차이의 크기의 척도는 좌 채널 오디오 신호와 우 채널 오디오 신호의 차이의 전력, 대수 전력, 크기 또는 대수 크기를 포함한다. 중심 채널 오디오 신호의 크기는 중심 채널 오디오 신호의 전력, 대수 전력, 크기 또는 대수 크기를 포함한다. 신호는 신호 처리 장치에 의해 처리된 어떤 신호라고 할 수 있다.

[0049]

이와 같은 제1 양태 또는 제1 양태의 앞서 임의의 앞선 구현 형태에 따른 신호 처리 장치의 제19 구현 형태에서, 조합기는 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 미리 결정된 입력 이득 팩터로 가중하고, 가중된 좌 채널 오디오 신호, 가중된 중심 채널 오디오 신호, 및 가중된 우 채널 오디오 신호를 미리 결정된 음성 이득 팩터로 가중하도록 더 구성된다. 그러므로, 비음성 성분의 크기와 관련된 음성 성분의 크기의 효율적인 제어가 실현된다.

[0050]

가중된 오디오 신호들 C_E , L_E , 및 R_E 는 미리 결정된 음성 이득 팩터 G_S 에 의해 가중될 수 있다. 가중은 음성 활성도 검출기를 사용하지 않고 수행될 수 있다.

[0051]

제2 양태에 따르면, 본 발명은 멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 신호 처리 방법에 관한 것으로, 멀티-채널 오디오 신호는 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 포함하고, 신호 처리 방법은 필터에 의해, 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호에 기초하여 주파수에 걸쳐 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 결정하고, 필터에 의해, 중심 채널 오디오 신호의 크기의 척도와 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도 사이의 비에 기초하여 이득 함수를 획득하고, 필터에 의해, 좌 채널 오디오 신호를 이득 함수로 가중하여 가중된 좌 채널 오디오 신호를 획득하고, 필터에 의해, 중심 채널 오디오 신호를 이득 함수로 가중하여 가중된 중심 채널 오디오 신호를 획득

득하고, 필터에 의해, 우 채널 오디오 신호를 이득 함수로 가중하여 가중된 우 채널 오디오 신호를 획득하고, 조합기에 의해, 좌 채널 오디오 신호를 가중된 좌 채널 오디오 신호와 조합하여 조합된 좌 채널 오디오 신호를 획득하고, 조합기에 의해, 중심 채널 오디오 신호를 가중된 중심 채널 오디오 신호와 조합하여 조합된 중심 채널 오디오 신호를 획득하고, 조합기에 의해, 우 채널 오디오 신호를 가중된 우 채널 오디오 신호와 조합하여 조합된 우 채널 오디오 신호를 획득하는 것을 포함한다. 그러므로, 멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 효율적인 개념이 실현된다.

[0052] 신호 처리 방법은 신호 처리 장치에 의해 수행될 수 있다. 신호 처리 방법의 다른 특징들은 신호 처리 장치의 기능성으로부터 직접 생긴다.

[0053] 이와 같은 제2 양태에 따른 신호 처리 방법의 제1 구현 형태에서, 방법은 필터에 의해, 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 중심 채널 오디오 신호의 크기의 척도와 좌 채널 오디오 신호와 우 채널 오디오 신호의 차이의 크기의 척도의 합으로서 결정하는 것을 포함한다. 그러므로, 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도가 효율적으로 그리고 더욱 적합한 방식으로 결정되어 필터 이득 함수를 획득하기 위해 사용되는데, 왜냐하면 좌 채널 오디오 신호와 우 채널 오디오 신호의 차이는 중심 채널 오디오 신호의 성분들을 포함하지 않는 잔여 신호를 나타내기 때문이다.

[0054] 이와 같은 제2 양태 또는 제2 양태의 임의의 앞선 구현 형태에 따른 신호 처리 방법의 제2 구현 형태에서, 방법은 필터에 의해, 다음 식들:

$$G(m, k) = \frac{P_C(m, k)}{P_C(m, k) + P_S(m, k)}$$

$$P_C(m, k) = |C(m, k)|^2$$

$$P_S(m, k) = |L(m, k) - R(m, k)|^2$$

[0055]

에 따라 이득 함수를 결정하는 것을 포함하고, G는 이득 함수를 나타내고, L은 좌 채널 오디오 신호를 나타내고, C는 중심 채널 오디오 신호를 나타내고, R은 우 채널 오디오 신호를 나타내고, P_C는 중심 채널 오디오 신호의 크기를 나타내는 척도로서 중심 채널 오디오 신호의 전력을 나타내고, P_S는 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이의 전력을 나타내고, P_C와 P_S의 합은 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 나타내고, m은 샘플 시간 인덱스를 나타내고, k는 주파수 빈 인덱스를 나타낸다. 그러므로, 이득 함수는 효율적이고 강력한 방식으로 결정된다.

[0057] 이와 같은 제2 양태 또는 제2 양태의 임의의 앞선 구현 형태에 따른 신호 처리 방법의 제3 구현 형태에서, 멀티-채널 오디오 신호는 좌 서라운드 채널 오디오 신호 및 우 서라운드 채널 오디오 신호를 더 포함하고, 방법은 필터에 의해, 좌 서라운드 채널 오디오 신호 및 우 서라운드 채널 오디오 신호에 기초하여 부가적으로 주파수에 걸쳐 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 결정하고, 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 중심 채널 오디오 신호의 크기의 척도와 좌 채널 오디오 신호와 우 채널 오디오 신호의 차이의 크기의 척도와 좌 서라운드 채널 오디오 신호와 우 서라운드 채널 오디오 신호의 차이의 크기의 척도의 합으로서 결정하는 것을 포함한다. 그러므로, 멀티-채널 오디오 신호 내의 서라운드 채널들은 좌 서라운드 채널 오디오 신호와 우 서라운드 채널 오디오 신호의 차이로부터 크기를 획득함으로써, 효율적으로 처리된다. 이 차이 신호는 중심 채널 오디오 신호에 보다 양호한 구별을 제공한다.

[0058] 이와 같은 제2 양태 또는 제2 양태의 임의의 앞선 구현 형태에 따른 신호 처리 방법의 제4 구현 형태에서, 방법은 필터에 의해, 좌 채널 오디오 신호의 주파수 빈들을 이득 함수의 주파수 빈들로 가중하여 가중된 좌 채널 오디오 신호의 주파수 빈들을 획득하고, 필터에 의해, 중심 채널 오디오 신호의 주파수 빈들을 이득 함수의 주파수 빈들로 가중하여 가중된 중심 채널 오디오 신호의 주파수 빈들을 획득하고, 필터에 의해, 우 채널 오디오 신호의 주파수 빈들을 이득 함수의 주파수 빈들로 가중하여 가중된 우 채널 오디오 신호의 주파수 빈들을 획득하는 것을 포함한다. 그러므로, 멀티-채널 오디오 신호는 주파수 영역에서 효율적으로 처리된다. 동일한 필터로 모든 신호들을 가중하는 것은 스테레오 이미지 내의 오디오 소스 위치들의 시프팅이 발생하지 않는다는 장점을 갖는다. 더구나, 이 방식으로, 음성 성분이 모든 신호들로부터 추출된다.

[0059] 이와 같은 제2 양태 또는 제2 양태의 임의의 앞선 구현 형태에 따른 신호 처리 방법의 제5 구현 형태에서, 방법은 음성 활성화도 검출기에 의해, 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호에 기초하여 음성 활성화도 표시자를 결정하고 - 음성 활성화도 표시자는 시간에 걸쳐 멀티-채널 오디오 신호 내의 음성 성분의 크기를 표시함 -, 조합기에 의해, 가중된 좌 채널 오디오 신호를 음성 활성화도 표시자와 조합하여 조합된 좌 채널 오디오 신호를 획득하고, 조합기에 의해, 가중된 중심 채널 오디오 신호를 음성 활성화도 표시자와 조합하여 조합된 중심 채널 오디오 신호를 획득하고, 조합기에 의해, 가중된 우 채널 오디오 신호를 음성 활성화도 표시자와 조합하여 조합된 우 채널 오디오 신호를 획득하는 것을 포함한다. 그러므로, 멀티-채널 오디오 신호들 내의 시변 음성 성분의 효율적인 향상이 실현되고, 비음성 신호들이 억제된다.

[0060] 제2 양태의 제5 구현 형태에 따른 신호 처리 방법의 제6 구현 형태에서, 방법은 음성 활성화도 검출기에 의해, 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호에 기초하여 멀티-채널 오디오 신호의 전체 스펙트럼 변화를 나타내는 척도를 결정하고, 음성 활성화도 검출기에 의해, 중심 채널 오디오 신호의 스펙트럼 변화의 척도와 멀티-채널 오디오 신호의 전체 스펙트럼 변화를 나타내는 척도 사이의 비에 기초하여 음성 활성화도 표시자를 획득하는 것을 포함한다. 그러므로, 음성 활성화도 표시자는 스펙트럼 변화의 척도들 사이의 관계를 이용함으로써 효율적으로 결정된다.

[0061] 제2 양태의 제6 구현 형태에 따른 신호 처리 방법의 제7 구현 형태에서, 방법은 음성 활성화도 검출기에 의해, 다음 식:

$$V = a \times \left(\frac{F_c}{F_c + F_s} - 0.5 \right)$$

[0062] 에 따라 음성 활성화도 표시자를 결정하는 것을 포함하고, V는 음성 활성화도 표시자를 나타내고, F_c는 중심 채널 오디오 신호의 스펙트럼 변화의 척도를 나타내고, F_s는 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이의 스펙트럼 변화의 척도를 나타내고, F_c와 F_s의 합은 멀티-채널 오디오 신호의 전체 스펙트럼 변화를 나타내는 척도를 나타내고, a는 미리 결정된 스케일링 팩터를 나타낸다. 그러므로, 음성 활성화도 표시자가 효율적으로 결정된다. F_c와 F_s가 동일한 값들을 갖는 신호들이라면 0의 값을 갖는 음성 활성화도 표시자가 야기된다. F_c의 값이 높을수록 음성 활성화도 표시자의 값들이 높아진다. 스케일링 팩터는 음성 활성화도 표시자의 크기를 제어할 수 있다.

[0064] 제2 양태의 제7 구현 형태에 따른 신호 처리 방법의 제8 구현 형태에서, 방법은 음성 활성화도 검출기에 의해, 다음 식들:

$$F_c(m) = \sum_k \left(|C(m, k)| - |C(m-1, k)| \right)^2$$

$$F_s(m) = \sum_k \left(|S(m, k)| - |S(m-1, k)| \right)^2$$

[0065] 에 따라 스펙트럼 플럭스로서 중심 채널 오디오 신호들의 스펙트럼 변화의 척도 및 스펙트럼 플럭스로서 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이의 스펙트럼 변화의 척도를 결정하는 것을 포함하고, F_c는 중심 채널 오디오 신호의 스펙트럼 플럭스를 나타내고, F_s는 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이의 스펙트럼 플럭스를 나타내고, C는 중심 채널 오디오 신호를 나타내고, S는 좌 채널 오디오 신호와 우 채널 오디오 신호 사이의 차이를 나타내고, m은 샘플 시간 인덱스를 나타내고, k는 주파수 빈 인덱스를 나타낸다. 그러므로, 스펙트럼 플럭스가 효율적으로 결정된다.

[0067] 제2 양태의 제5 구현 형태 내지 제8 구현 형태에 따른 신호 처리 방법의 제9 구현 형태에서, 방법은 음성 활성화도 검출기에 의해, 미리 결정된 저역 통과 필터링 기능에 기초하여 시간에서 음성 활성화도 표시자를 필터링하는 것을 포함한다. 그러므로, 멀티-채널 오디오 신호 내의 아티팩트들의 효율적인 완화 및/또는 음성 활성화도 표시자의 효율적인 시간적 평활화가 실현된다.

[0068] 제2 양태의 제5 구현 형태 내지 제9 구현 형태에 따른 신호 처리 방법의 제10 구현 형태에서, 방법은 조합기에 의해, 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 미리 결정된 입력 이득 팩터로

가중하고, 조합기에 의해, 음성 활성화도 표시자를 미리 결정된 음성 이득 팩터로 가중하는 것을 포함한다. 그러므로, 비음성 성분의 크기에 관련한 음성 성분의 크기의 효율적인 제어가 실현된다.

[0069] 제2 양태의 제5 구현 형태 내지 제10 구현 형태에 따른 신호 처리 방법의 제11 구현 형태에서, 방법은 조합기에 의해, 가중된 좌 채널 오디오 신호와 음성 활성화도 표시자의 조합에 좌 채널 오디오 신호를 가산하여 조합된 좌 채널 오디오 신호를 획득하고, 조합기에 의해, 가중된 좌 채널 오디오 신호와 음성 활성화도 표시자의 조합에 중심 채널 오디오 신호를 가산하여 조합된 중심 채널 오디오 신호를 획득하고, 조합기에 의해, 가중된 좌 채널 오디오 신호와 음성 활성화도 표시자의 조합에 우 채널 오디오 신호를 가산하여 조합된 우 채널 오디오 신호를 획득하는 것을 포함한다. 그러므로, 조합이 효율적으로 수행된다. 추출된 음성 성분들은 출력 신호들 내의 음성 성분을 향상시키기 위해 원래의 신호들과 조합된다.

[0070] 제2 양태의 제5 구현 형태 내지 제11 구현 형태에 따른 신호 처리 방법의 제12 구현 형태에서, 멀티-채널 오디오 신호는 좌 서라운드 채널 오디오 신호 및 우 서라운드 채널 오디오 신호를 더 포함하고, 방법은 음성 활성화도 검출기에 의해, 좌 서라운드 채널 오디오 신호 및 우 서라운드 채널 오디오 신호에 기초하여 부가적으로 음성 활성화도 표시자를 결정하는 것을 포함한다. 그러므로, 멀티-채널 오디오 신호 내의 서라운드 채널들이 또한 음성 활성화도 표시자를 결정하기 위해 고려되어, 음성 활성화도 표시자의 더 양호한 평가를 가져다 준다.

[0071] 이와 같은 제2 양태 또는 제2 양태의 임의의 앞선 구현 형태에 따른 신호 처리 방법의 제13 구현 형태에서, 방법은 변환기에 의해, 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 시간 영역으로부터 주파수 영역으로 변환하는 것을 포함한다. 그러므로, 오디오 신호들의 주파수 영역으로의 효율적인 변환이 실현된다. 이것은 예를 들어, 음성 향상 및 음성 활성화도 검출이 주파수 영역에서 수행되는 경우에 요구될 수 있다.

[0072] 이와 같은 제2 양태 또는 제2 양태의 임의의 앞선 구현 형태에 따른 신호 처리 방법의 제14 구현 형태에서, 방법은 역 변환기에 의해, 조합된 좌 채널 오디오 신호, 조합된 중심 채널 오디오 신호, 및 조합된 우 채널 오디오 신호를 주파수 영역으로부터 시간 영역으로 역 변환하는 것을 포함한다. 그러므로, 오디오 신호들의 시간 영역으로의 효율적인 역 변환이 실현되고, 시간 영역에서의 출력 신호들이 획득된다.

[0073] 이와 같은 제2 양태 또는 제2 양태의 임의의 앞선 구현 형태에 따른 신호 처리 방법의 제15 구현 형태에서, 방법은 업-믹서에 의해, 입력 좌 채널 스테레오 오디오 신호 및 입력 우 채널 스테레오 오디오 신호에 기초하여 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 결정하는 것을 포함한다. 이 방식으로, 신호 처리 방법은 입력 스테레오 오디오 신호를 처리하기 위해 적용될 수 있다.

[0074] 제2 양태의 제15 구현 형태에 따른 신호 처리 방법의 제16 구현 형태에서, 방법은 업-믹서에 의해, 다음 식들:

$$C = \alpha \times (L_{in} + R_{in})$$

$$L = L_{in} - C$$

$$R = R_{in} - C$$

$$\alpha = \frac{1}{2} \times \left(1 - \sqrt{\frac{(L_r - R_r)^2 + (L_i - R_i)^2}{(L_r + R_r)^2 + (L_i + R_i)^2}} \right)$$

[0075] .

[0076] 에 따라 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 결정하는 것을 포함하고, L_r 은 입력 좌 채널 스테레오 오디오 신호의 실수부를 나타내고, R_r 은 입력 우 채널 스테레오 오디오 신호의 실수부를 나타내고, L_i 는 입력 좌 채널 스테레오 오디오 신호의 허수부를 나타내고, R_i 는 입력 우 채널 스테레오 오디오 신호의 허수부를 나타내고, α 는 직교성 파라미터를 나타내고, L_{in} 은 입력 좌 채널 스테레오 오디오 신호를 나타내고, R_{in} 은 입력 우 채널 스테레오 오디오 신호를 나타내고, L 은 좌 채널 오디오 신호를 나타내고, C 는 중심 채널 오디오 신호를 나타내고, R 은 우 채널 오디오 신호를 나타낸다. 그러므로, 입력 스테레오 오디오 신호의 효율적인 중심 채널 추출이 직교 분해를 사용하여 실현된다. 결과적인 좌 채널 오디오 신호와 우 채널 오디오 신호는 서로 직교한다.

- [0077] 이와 같은 제2 양태 또는 제2 양태의 임의의 앞선 구현 형태에 따른 신호 처리 방법의 제17 구현 형태에서, 방법은 다운-믹서에 의해, 조합된 좌 채널 오디오 신호, 조합된 중심 채널 오디오 신호, 및 조합된 우 채널 오디오 신호에 기초하여 출력 좌 채널 스테레오 오디오 신호 및 출력 우 채널 스테레오 오디오 신호를 결정하는 것을 포함한다. 그러므로, 2-채널, 즉 좌 및 우 채널, 출력 스테레오 오디오 신호가 효율적으로 제공된다.
- [0078] 이와 같은 제2 양태 또는 제2 양태의 임의의 앞선 구현 형태에 따른 신호 처리 방법의 제18 구현 형태에서, 크기의 척도는 신호의 전력, 대수 전력, 크기 또는 대수 크기를 포함한다. 그러므로, 크기의 척도는 상이한 스케일들에서 상이한 값들을 표시할 수 있다.
- [0079] 이와 같은 제2 양태 또는 제2 양태의 임의의 앞선 구현 형태에 따른 신호 처리 방법의 제19 구현 형태에서, 방법은 조합기에 의해, 좌 채널 오디오 신호, 중심 채널 오디오 신호, 및 우 채널 오디오 신호를 미리 결정된 입력 이득 팩터로 가중하고, 조합기에 의해, 가중된 좌 채널 오디오 신호, 가중된 중심 채널 오디오 신호, 및 가중된 우 채널 오디오 신호를 미리 결정된 음성 이득 팩터로 가중하는 것을 포함한다. 그러므로, 비음성 성분의 크기와 관련된 음성 성분의 크기의 효율적인 제어가 실현된다.
- [0080] 제3 양태에 따르면, 본 발명은 컴퓨터 상에서 실행될 때 이와 같은 제2 양태 또는 제2 양태의 구현 형태들 중 어느 것에 따른 방법을 수행하는 프로그램 코드를 포함하는 컴퓨터 프로그램에 관한 것이다. 그러므로, 이 방법은 자동으로 수행될 수 있다.
- [0081] 신호 처리 장치는 컴퓨터 프로그램 및/또는 프로그램 코드를 실행하도록 프로그램가능하게 구성될 수 있다.
- [0082] 본 발명은 하드웨어 및/또는 소프트웨어에서 구현될 수 있다.

도면의 간단한 설명

- [0083] 본 발명의 실시예들이 다음의 도면과 관련하여 설명될 것이다:
 - 도 1은 실시예에 따른 멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 신호 처리 장치의 다이어그램을 도시하고;
 - 도 2는 실시예에 따른 멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 신호 처리 방법의 다이어그램을 도시하고;
 - 도 3은 실시예에 따른 멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 신호 처리 장치의 다이어그램을 도시하고;
 - 도 4는 실시예에 따른 신호 처리 장치의 업-믹서의 다이어그램을 도시하고;
 - 도 5는 실시예에 따른 신호 처리 장치의 필터의 다이어그램을 도시하고;
 - 도 6은 실시예에 따른 신호 처리 장치의 음성 활성화 검출기의 다이어그램을 도시하고;
 - 도 7은 실시예에 따른 멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 신호 처리 장치의 다이어그램을 도시한다.

동일한 참조 부호는 동일한 또는 동등한 특징들에 대해 사용된다.

발명을 실시하기 위한 구체적인 내용

- [0084] 도 1은 실시예에 따른 멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 신호 처리 장치(100)의 다이어그램을 도시한다. 멀티-채널 오디오 신호는 좌 채널 오디오 신호 L, 중심 채널 오디오 신호 C, 및 우 채널 오디오 신호 R을 포함한다. 신호 처리 장치(100)는 필터(101) 및 조합기(103)를 포함한다.
- [0085] 필터(101)는 좌 채널 오디오 신호 L, 중심 채널 오디오 신호 C, 및 우 채널 오디오 신호 R에 기초하여 주파수에 걸쳐 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 결정하고, 중심 채널 오디오 신호 C의 크기의 척도와 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도 사이의 비에 기초하여 이득 함수 G를 획득하고, 좌 채널 오디오 신호 L을 이득 함수 G로 가중하여 가중된 좌 채널 오디오 신호 L_E 를 획득하고, 중심 채널 오디오 신호 C를 이득 함수 G로 가중하여 가중된 중심 채널 오디오 신호 C_E 를 획득하고, 우 채널 오디오 신호 R을 이득 함수 G로 가중하여 가중된 우 채널 오디오 신호 R_E 를 획득하도록 구성된다.

- [0086] 조합기(103)는 좌 채널 오디오 신호 L을 가중된 좌 채널 오디오 신호 L_E 와 조합하여 조합된 좌 채널 오디오 신호 L_{EV} 를 획득하고, 중심 채널 오디오 신호 C를 가중된 중심 채널 오디오 신호 C_E 와 조합하여 조합된 중심 채널 오디오 신호 C_{EV} 를 획득하고, 우 채널 오디오 신호 R을 가중된 우 채널 오디오 신호 R_E 와 조합하여 조합된 우 채널 오디오 신호 R_{EV} 를 획득하도록 구성된다.
- [0087] 멀티-채널 오디오 신호들은 예를 들어, 좌 채널 오디오 신호 L, 우 채널 오디오 신호 및 중심 채널 오디오 신호 C만을 포함하고, 또한 LCR 스테레오 또는 3.0 스테레오 오디오 신호들이라고 할 수 있는, 3-채널 스테레오 오디오 신호들, 좌 채널 오디오 신호 L, 우 채널 오디오 신호 R, 중심 채널 오디오 신호 C, 좌 서라운드 채널 오디오 신호 L_S , 우 서라운드 채널 오디오 신호 R_S , 및 베이스 채널 신호 B, 또는 중심 채널 오디오 신호 및 적어도 2개의 다른 채널 오디오 신호들을 갖는 다른 멀티-채널 신호들을 포함하는 5.1 멀티-채널 오디오 신호들을 포함할 수 있다. 중심 채널 오디오 신호 C 이외의 오디오 신호들, 예를 들어, 좌 채널 오디오 신호 L, 우 채널 오디오 신호 R, 좌 서라운드 채널 오디오 신호 L_S , 우 서라운드 채널 오디오 신호 R_S 및 베이스 채널 신호 B는 또한 비중심 채널 오디오 신호들이라고 할 수 있다. 5.1 멀티-채널 오디오 신호의 경우에, 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도는 중심 채널 오디오 신호의 크기의 양, 좌 채널 오디오 신호와 우 채널 오디오 신호의 차이의 크기의 척도, 좌 서라운드 채널 오디오 신호와 우 서라운드 채널 오디오 신호의 차이의 크기의 척도, 및 저-주파수 효과들 채널 오디오 신호의 크기의 척도의 합으로서 획득될 수 있다. 5.1 멀티-채널 오디오 신호의 경우에, 획득된 필터는 포함된 오디오 신호들 모두를 가중하기 위해 사용될 수 있다.
- [0088] 도 2는 실시예에 따른 멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 신호 처리 방법(200)의 다이어그램을 도시한다. 멀티-채널 오디오 신호는 좌 채널 오디오 신호 L, 중심 채널 오디오 신호 C, 및 우 채널 오디오 신호 R을 포함한다.
- [0089] 신호 처리 방법(200)은 좌 채널 오디오 신호 L, 중심 채널 오디오 신호 C, 및 우 채널 오디오 신호 R에 기초하여 주파수에 걸쳐 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도를 결정하고(201), 중심 채널 오디오 신호 C의 크기의 척도와 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도 사이의 비에 기초하여 이득 함수 G를 획득하고(203), 좌 채널 오디오 신호 L을 이득 함수 G로 가중하여 가중된 좌 채널 오디오 신호 L_E 를 획득하고(205), 중심 채널 오디오 신호 C를 이득 함수 G로 가중하여 가중된 중심 채널 오디오 신호 C_E 를 획득하고(207), 우 채널 오디오 신호 R을 이득 함수 G로 가중하여 가중된 우 채널 오디오 신호 R_E 를 획득하고(209), 좌 채널 오디오 신호 L을 가중된 좌 채널 오디오 신호 L_E 와 조합하여 조합된 좌 채널 오디오 신호 L_{EV} 를 획득하고(211), 중심 채널 오디오 신호 C를 가중된 중심 채널 오디오 신호 C_E 와 조합하여 조합된 중심 채널 오디오 신호 C_{EV} 를 획득하고(213), 우 채널 오디오 신호 R을 가중된 우 채널 오디오 신호 R_E 와 조합하여 조합된 우 채널 오디오 신호 R_{EV} 를 획득하는(215) 것을 포함한다.
- [0090] 신호 처리 방법(200)은 신호 처리 장치(100)에 의해, 예를 들어, 필터(101) 및 조합기(103)에 의해 수행될 수 있다.
- [0091] 다음에서, 신호 처리 장치(100) 및 신호 처리 방법(200)의 다른 구현 형태들 및 실시예들이 설명될 것이다.
- [0092] 본 발명은 오디오 신호 처리의 분야에 관한 것이다. 신호 처리 장치(100) 및 신호 처리 방법(200)은 음성 향상, 예를 들어, 오디오 신호들, 예를 들어, 스테레오 오디오 신호들 내의 대화 향상을 위해 적용될 수 있다. 특히, 신호 처리 장치(100) 및 신호 처리 방법(200)은 업-믹서(301)와 조합하여 또는 업-믹서(301) 및 다운-믹서(303)와 조합하여, 대화 선명성을 개선하도록 스테레오 오디오 신호들을 처리하기 위해 적용될 수 있다.
- [0093] TV들, 랩탑들, 태블릿 컴퓨터들, 이동 전화기들, 및 스마트폰들과 같은, 2개의 스피커를 갖는 상이한 디바이스들이 있다. 스테레오 오디오 신호들이 이러한 디바이스들을 사용하여 재생될 때, 예를 들어, 영화들로부터의 사운드트랙들의 음성 성분들은 정상인 및 청각 장애 청취자들에게는 알아듣기가 어려울 수 있다. 이것은 특히 잡음이 있는 환경들에서 또는 음성 성분이 비음성 성분들 또는 음악 또는 사운드 효과들과 같은 사운드들에 의해 중첩될 때 그렇다.
- [0094] 본 발명의 실시예들은 특히, 대화 선명성을 개선하기 위해 스테레오 오디오 신호들의 음성 성분을 향상시키는 것을 목적으로 한다. 하나의 중요한 가정은 음성, 또는 상응하여 말이 스테레오 오디오 신호들의 대부분에 대해 일반적으로 그러한, 멀티-채널 오디오 신호 내에서 중심-패닝된다는 것이다. 목적은 비음성 성분은 변화하

지 않은 채로 남기면서, 음성 품질에 영향을 주지 않고 음성 성분들의 음량을 향상시키는 것이다. 이것은 특히 동시의 음성 및 비음성 성분들을 갖는 시간 간격들 동안 가능하여야 한다. 본 발명의 실시예들은 예를 들어, 단지 스테레오 오디오 신호만을 사용하게 하고 별개의 음성 오디오 채널 또는 원래의 5.1 멀티-채널 오디오 신호로부터의 추가 지식을 필요로 하지 않거나 이용하지 않는다. 이 목적들은 설명된 신호 처리 장치(100) 또는 신호 처리 방법(200)을 사용하여 가상 중심 채널 오디오 신호를 추출하고 다른 오디오 신호들뿐만 아니라 이 중심 채널 오디오 신호를 향상시킴으로써 달성된다. 게다가, 음성 활성화도 검출을 위한 방식이 비음성 성분들이 처리에 의해 영향받지 않을 수 있도록 하기 위해 이용될 수 있다. 본 발명의 다른 실시예들은 5.1 멀티-채널 오디오 신호와 같은, 다른 멀티-채널 오디오 신호들을 처리하기 위해 사용될 수 있다.

[0095] 본 발명의 실시예들은 다음의 방식에 기초하는데, 스테레오 오디오 신호 기록으로부터, 중심 채널 오디오 신호가 업-믹싱 방식을 사용하여 추출된다. 이 중심 채널 오디오 신호는 원래의 음성 성분의 평가를 획득하기 위해, 음성 향상 및 음성 활성화도 검출을 사용하여 더 처리될 수 있다. 이 방식의 특징은 음성 성분이 중심 채널 오디오 신호로부터 뿐만 아니라, 나머지 채널 오디오 신호들로부터 추출될 수 있다. 업-믹싱 처리는 완벽히 될 수 없기 때문에, 이들 나머지 채널 오디오 신호는 여전히 음성 성분을 포함할 수 있다. 음성 성분들이 또한 추출되고 부스팅될 때, 결과적인 출력 오디오 신호들은 개선된 음성 품질 및 폭을 갖는다.

[0096] 다음에서, 2-대-3 업-믹싱에 의해 2-채널 스테레오 오디오 신호로부터 획득된, (중심 채널 오디오 신호, 좌 채널 오디오 신호, 및 우 채널 오디오 신호를 포함하는) 멀티-채널 오디오 신호 LCR의 음성 성분을 향상시키는 본 발명의 특정한 실시예들이 도 3 내지 7에 기초하여 설명된다.

[0097] 그러나, 본 발명의 실시예들은 이러한 멀티-채널 오디오 신호들로 제한되지 않고 또한 예를 들어, 다른 디바이스들로부터 수신된 LCR 3개 채널 오디오 신호들의 처리, 또는 예를 들어, 5.1 또는 7.1 멀티채널 신호들의 중심 채널 오디오 신호를 포함하는 다른 멀티-채널 오디오 신호들의 처리를 포함할 수 있다. 다른 실시예들이 심지어 음성 활성화도 검출이 있거나 없이 음성 또는 대화 향상을 적용하기 전에 가상 중심 채널 오디오 신호를 획득하기 위해 멀티-채널 신호를 업-믹싱함으로써, 중심 채널 오디오 신호를 포함하지 않는, 멀티-채널 신호들, 예를 들어, 좌 및 우 오디오 채널 신호 및 좌 및 우 서라운드 채널 신호를 포함하는 4.0 멀티채널 신호를 처리하도록 구성될 수 있다.

[0098] 도 3은 실시예에 따른 멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 신호 처리 장치(100)의 다이어그램을 도시한다. 신호 처리 장치(100)는 필터(101), 조합기(103), 업-믹서(301), 및 다운-믹서(303)를 포함한다. 필터(101) 및 조합기(103)는 좌 채널 처리기(305), 중심 채널 처리기(307), 및 우 채널 처리기(309)를 포함한다.

[0099] 업-믹서(301)는 입력 좌 채널 스테레오 오디오 신호 L_{in} 및 입력 우 채널 스테레오 오디오 신호 R_{in} 에 기초하여 좌 채널 오디오 신호 L, 중심 채널 오디오 신호 C, 및 우 채널 오디오 신호 R을 결정하도록 구성된다. 바꾸어 말하면, 업-믹서(301)는 도 4에 기초하여 보다 상세히 예시적으로 설명되는 바와 같이, 2-대-3 업-믹스를 제공한다.

[0100] 좌 채널 처리기(305)는 조합된 좌 채널 오디오 신호 L_{EV} 를 제공하기 위해 좌 채널 오디오 신호 L을 처리하도록 구성된다. 중심 채널 처리기(307)는 조합된 중심 채널 오디오 신호 C_{EV} 를 제공하기 위해 중심 채널 오디오 신호 C를 처리하도록 구성된다. 우 채널 처리기(309)는 조합된 우 채널 오디오 신호 R_{EV} 를 제공하기 위해 우 채널 오디오 신호 R을 처리하도록 구성된다. 좌 채널 처리기(305), 중심 채널 처리기(307), 및 우 채널 처리기(309)는 도 5에 기초하여 보다 상세히 예시적으로 설명되는 바와 같이, 음성 향상, ENH를 수행하도록 구성된다. 좌 채널 처리기(305), 중심 채널 처리기(307), 및 우 채널 처리기(309)는 도 6에 기초하여 보다 상세히 예시적으로 설명되는 바와 같이, 음성 활성화도 검출, VAD에 의해 제공된 음성 활성화도 표시자를 처리하도록 추가적으로 구성될 수 있다.

[0101] 다운-믹서(303)는 조합된 좌 채널 오디오 신호 L_{EV} , 조합된 중심 채널 오디오 신호 C_{EV} , 및 조합된 우 채널 오디오 신호 R_{EV} 에 기초하여 출력 좌 채널 스테레오 오디오 신호 L_{out} 및 출력 우 채널 스테레오 오디오 신호 R_{out} 를 결정하도록 구성된다. 바꾸어 말하면, 다운-믹서(303)는 3-대-2 다운-믹스를 제공한다.

[0102] 그러므로, 음성-향상된 오디오 신호들이 다운-믹스된 2-채널 스테레오 신호 L_{out} 및 R_{out} 가 통상적인 2-채널 스테레오 재생 디바이스, 예를 들어, 통상적인 스테레오 TV 세트에 직접 출력될 수 있도록 하는 방식으로 처리된다.

[0103] 본 발명의 한 실시예에서, 공통 방식이 입력 좌 채널 스테레오 오디오 신호 L_{in} 및 입력 우 채널 스테레오 오디오 신호 R_{in} 을 포함하는 입력 스테레오 오디오 신호로부터의 중심 채널 추출을 위해 업-믹서(301)에 의해 사용된다. 이것은 L, C, 및 R로서 표시된, 좌, 중심, 및 우 채널 오디오 신호를 초래한다. 본 발명의 다른 실시예들은 업-믹싱을 위해 다른 방식들을 사용할 수 있다. 본 발명의 다른 실시예들이 상상가능하고, 예를 들어, 5.1 멀티-채널 오디오 신호들이 가용하고 포함된 좌, 중심 및 우 채널들이 직접 사용된다.

[0104] 좌, 중심, 및 우 채널 오디오 신호들 L, C, 및 R은 다음에 멀티-채널 오디오 신호의 모든 채널들 상에 적용될 수 있는 시간 및/또는 주파수 종속 음성 향상 필터(101)를 평가하기 위해 개선된 방식으로 처리된다. 이 필터(101)는 음성 성분과 동시에 존재할 수 있는 비음성 성분들을 감쇠시키도록 구성된다. 다른 방식들과의 차이는 중심 채널 오디오 신호뿐만 아니라, 다른 오디오 신호들, 예를 들어, 도 3에 도시된 것과 같은 LCR 경우에서의 좌 채널 오디오 신호 및 우 채널 오디오 신호가 동일한 필터(101)로 처리된다는 것이다. 본 발명의 실시예들은 음성 향상 필터(101)를 정의하기 위해 개선된 방식을 사용한다.

[0105] 또한, 음성 활성화도 검출이 멀티-채널 오디오 신호의 모든 채널들로부터의 정보를 이용하는 개선된 방식을 사용하여 수행될 수 있다. 음성 활성화도 검출기의 출력, 예를 들어, 음성 활성화도 표시자는 음성 활성화도를 표시할 수 있는 연 판정(soft decision)일 수 있다. 음성 향상과 음성 활성화도 검출의 조합은 음성 성분만 또는 음성 성분만을 적어도 거의 포함하는 멀티-채널 오디오 신호를 제공한다. 이 음성 성분 멀티-채널 오디오 신호는 조합된 채널 오디오 신호들 L_{EV} , C_{EV} , 및 R_{EV} 를 획득하기 위해 조합기(103)에 의해 부스팅되어 원래의 멀티-채널 오디오 신호에 가산될 수 있다. 스테레오로의 다운-믹스는 최종 출력 채널 스테레오 오디오 신호들 L_{out} 및 R_{out} 를 제공하기 위해 다운-믹서(303)에 의해 수행될 수 있다.

[0106] 도 4는 실시예에 따른 신호 처리 장치(100)의 업-믹서(301)의 다이어그램을 도시한다. 업-믹서(301)는 입력 좌 채널 스테레오 오디오 신호 L_{in} 및 입력 우 채널 스테레오 오디오 신호 R_{in} 에 기초하여 좌 채널 오디오 신호 L, 중심 채널 오디오 신호 C, 및 우 채널 오디오 신호 R을 결정하도록 구성된다. 업-믹서(301)는 2-대-3 업-믹싱을 제공한다. 업-믹서(301)는 업-믹싱 방식을 사용하여 입력 2-채널 스테레오 오디오 신호로부터의 중심 채널 오디오 신호 C의 추출을 수행하도록 구성된다.

[0107] 예를 들어, 2-채널 입력 스테레오 오디오 신호로부터 가상 중심 채널 오디오 신호 C를 획득하는 처리를 또한 중심 추출이라고 한다. 이것은 단지 기록의 통상적인 스테레오 오디오 신호가 가용할 때만 요구될 수 있다. 중심 추출을 달성하는 상이한 방식들이 있다. 일군의 업-믹싱 방식들은 매트릭스 디코딩에 기초한다. 이들 방식은 업-믹싱하는 선형 신호-독립 방식들이다. 그들은 매트릭스 디코더와 결합될 수 있고 시간 영역에서 동작한다. 기하학적 방식들은 반면에, 신호-종속적이다. 이들 방식은 좌 채널 오디오 신호 L과 우 채널 오디오 신호 R이 서로에 대하여 비상관된다는 가정에 의존할 수 있다. 이들 방식은 주파수 영역에서 동작한다.

[0108] 다음에서, 특정한 방식이 본 발명의 임의의 실시예에서 사용될 수 있는, 중심 추출에 대해 예로서 설명된다. 이 방식은 주파수 영역에서 수행된다. 이것은 입력 스테레오 오디오 신호가 예를 들어, 단-시간 윈도우들 상에서의 이산 푸리에 변환(DFT) 알고리즘을 적용함으로써 주파수 영역으로 변환된다는 것을 의미한다. 이산 푸리에 변환(DFT)의 블록 크기에 대한 적절한 선택은 48000Hz의 샘플링 주파수가 사용될 때 1024일 수 있다.

[0109] 이 방식은 좌 및 우 채널 오디오 신호들 L 및 R이 서로에 관하여 직교하다는 가정에서 성립된다. 이 아이디어는

[0110]
$$C = \alpha \times (L_{in} + R_{in}) \quad (1)$$

[0111] 로서 중심 채널 오디오 신호 C를 획득하는 것이고 여기서 α 는 결정된 파라미터이다. 좌 및 우 채널 오디오 신호들 L 및 R은 다음에, 결과적인 중심 채널 오디오 신호 C로부터

[0112]
$$L = L_{in} - C \quad (2)$$

[0113]
$$R = R_{in} - C \quad (3)$$

[0114] 로서 도출될 수 있다. 파라미터 α 는 오디오 신호들의 직교성을 설명하는 구속

[0115]
$$L \times R^* = 0 \quad (4)$$

[0115] 를 이행하도록 하는 방식으로 최적화될 수 있다. 이 문제에 대한 수학적 해가 도출될 수 있고, 다음과 같은 결과가 산출되고

$$\alpha = \frac{1}{2} \times \left(1 - \sqrt{\frac{(L_r - R_r)^2 + (L_i - R_i)^2}{(L_r + R_r)^2 + (L_i + R_i)^2}} \right) \quad (5)$$

[0117] 여기서 L_r , L_i , R_r 및 R_i 는 각각, 입력 좌 및 우 스테레오 오디오 신호들 L_{in} 및 R_{in} 의 스펙트럼 성분들의 실수부 및 허수부를 나타낸다. 파라미터 α 는 시간 종속 및 주파수 종속이고 그러므로 오디오 신호 샘플들의 주어진 프레임의 모든 주파수 빈들에 대해 계산될 수 있다.

[0118] 중심 추출을 위한 다른 특정한 기하학적 방식들이 적용될 수 있다. 다른 특정한 방식들은 예를 들어, 중심 추출을 위한 주요 성분 분석을 사용한다.

[0119] 도 5는 실시예에 따른 신호 처리 장치(100)의 필터(101)의 다이어그램을 도시한다. 필터(101)는 감산기(501), 결정기(503), 결정기(505), 결정기(507), 가중기(509), 가중기(511), 및 가중기(513)를 포함한다. 다이어그램은 음성 향상 방식을 도시한다.

[0120] 감산기(501)는 잔여 오디오 신호 S를 획득하기 위해 좌 채널 오디오 신호 L로부터 우 채널 오디오 신호 R을 감산하도록 구성된다.

[0121] 결정기(503)는 중심 채널 오디오 신호 C의 크기의 척도 P_c 를 획득하기 위해 중심 채널 오디오 신호 C의 제공된 크기 또는 전력을 결정하도록 구성된다. 결정기(505)는 잔여 오디오 신호 S의 크기의 척도 P_s 를 획득하기 위해 잔여 오디오 신호 S의 제공된 크기 또는 전력을 결정하도록 구성된다.

[0122] 결정기(507)는 이득 함수 G를 획득하기 위해 중심 채널 오디오 신호 C의 크기의 척도 P_c 와 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도 사이의 비를 결정하도록 구성된다. 멀티-채널 오디오 신호의 전체 크기를 나타내는 척도는 중심 채널 오디오 신호 C의 크기의 척도 P_c 와 잔여 오디오 신호 S의 크기의 척도 P_s 의 합에 의해 형성된다. 이득 함수 G는 시간-종속 및/또는 주파수-종속일 수 있다. 샘플 시간 인덱스는 m으로서 표시된다. 주파수 빈 인덱스는 k로서 표시된다.

[0123] 가중기(509)는 가중된 좌 채널 오디오 신호 L_e 를 획득하기 위해 좌 채널 오디오 신호 L을 이득 함수 G로 가중하도록 구성된다. 가중기(511)는 가중된 중심 채널 오디오 신호 C_e 를 획득하기 위해 중심 채널 오디오 신호 C를 이득 함수 G로 가중하도록 구성된다. 가중기(513)는 가중된 우 채널 오디오 신호 R_e 를 획득하기 위해 우 채널 오디오 신호 R을 이득 함수 G로 가중하도록 구성된다.

[0124] 본 발명의 실시예들은 음성 향상을 위한 위너 필터링 방식에 따라 이득 함수 G를 평가하기 위해 좌, 중심, 및 우 채널 오디오 신호들 L, C, 및 R로부터의 정보를 사용한다. 위너 필터링 방식은 비음성 성분들을 제거하기 위해 멀티-채널 오디오 신호의 모든 채널들 상에 적용될 수 있다. 중심 채널 오디오 신호 C가 음성 성분을 포함하는 경우에, 위너 필터링 방식은 멀티-채널 오디오 신호의 모든 채널들의 음성 성분들만을 (거의) 유지한다.

[0125] 일반적으로, 이용된 음성 향상 방식은 부가 잡음을 다룰 수 있다. 그러므로, 임의의 채널의 입력 신호 Y는 $Y = X + N$ 으로서 간주될 수 있고, 여기서 X는 깨끗한 음성 성분을 포함하고 N은 부가 잡음으로서 간주될 수 있다. X와 N은 서로에 대하여 비상관된다고 가정한다. 관찰된 오디오 신호 Y로부터 N을 제거하기 위해, 부가 잡음 N의 잡음 전력 스펙트럼 밀도 또는 선형적인 신호 대 잡음 비 X/N이 평가될 수 있다. 주파수-종속 이득 함수 G 또는 $G(m,k)$ 는 다음에

$$G = \frac{\frac{X}{N}}{1 + \frac{X}{N}} = \frac{X}{X + N} \quad (6)$$

[0127] 로서 획득될 수 있고 깨끗한 음성 성분을 포함하는 오디오 신호의 평가는 오디오 신호의 모든 주파수 빈들 상에서 동작하는, $\hat{X} = G \times Y$ 로서 결정될 수 있다.

[0128] 음성 향상 방식은 중심 채널 오디오 신호 C가 대부분 음성을 포함한다는 가정을 이용한다. 일반적으로 중심 추출 방식은 완벽한 중심 추출을 제공하지 않기 때문에, 중심 채널 오디오 신호 C는 비음성 성분들을 포함할 수 있고 멀티-채널 오디오 신호의 다른 채널들은 음성 성분들을 포함할 수 있다. 그러므로, 목적은 중심 채널 오디오 신호 C에서 비음성 성분들을 제거하고 멀티-채널 오디오 신호의 다른 채널들 내의 음성 성분들을 분리하는 것이다. 이 목적을 달성하기 위해, 위너 필터링 방식이 이득 함수 G를 평가하기 위해 적용될 수 있다. 추가 잡음 N의 잡음 전력 스펙트럼 밀도를 평가하기 위해 복잡한 방식들을 사용하는 것 대신에, 위너 필터링 방식에 대한 X 및 N을 정의하기 위해 간단하면서 효율적인 방식이 식들 (7), (8), 및 (9)에 의해 정의된 것과 같이, 사용된다. 중심 채널 오디오 신호 C는 X에 대응하는, 음성 성분을 포함하는 것으로 간주되고, 멀티-채널 오디오 신호의 다른 채널들의 내용은 N에 대응하는, 잡음을 포함하는 것으로 간주된다.

[0129] 실시예에서, 잔여 오디오 신호 S는 감산기(501)에 의해, 예를 들어, $S = L - R$ 에 따라 좌 및 우 채널 오디오 신호들로부터 획득된다. 이 방식으로, 중심 성분들이 잔여 신호로부터 제거된다. 전력들은

$$P_C(m, k) = |C(m, k)|^2 \quad (7)$$

[0130]
$$P_S(m, k) = |L(m, k) - R(m, k)|^2 \quad (8)$$

[0131] 에 따라 결정기(503)에 의한 중심 채널 오디오 신호 C의 스펙트럼 및 결정기(505)에 의한 잔여 오디오 신호 S의 스펙트럼으로부터 결정될 수 있고 여기서 m은 샘플 시간 인덱스이고 k는 주파수 빈 인덱스이다. 또 하나의 가능한 방식은 전력, 또는 대수 크기 또는 전력 대신에 크기를 사용하는 것이다. 다른 실시예들에서, 전력들은 처리 아티팩트들을 감소시키기 위해 시간에 걸쳐 평활화될 수 있다.

[0132] 다음에 이득 함수 G는 다음 식에 따른 위너 필터링 방식에 따라 결정기(507)에 의해 결정된다.

$$G(m, k) = \frac{P_C(m, k)}{P_C(m, k) + P_S(m, k)} \quad (9)$$

[0133] 이득 함수 G는 후속하여 가중기들(509-513)에 의해, 각각, 좌, 중심, 및 우 채널 오디오 신호들 L, C, 및 R에 적용된다. 이것은 가중된 좌 채널 오디오 신호 L_E , 가중된 중심 채널 오디오 신호 C_E , 및 가중된 우 채널 오디오 신호 R_E 를 초래한다.

[0135] 원래의 중심 채널 오디오 신호 C가 단지 음성 성분만을 포함하는 경우에, 향상된 가중된 오디오 신호들은 또한 단지 음성 성분들만을 포함한다.

[0136] 본 발명의 실시예에서, 상이한 멀티-채널 오디오 신호 포맷이 사용된다. 예시적인 5.1 멀티-채널 오디오 신호에 대해, 잔여 오디오 신호 S를 결정하는 옵션은

[0137]
$$S = L - R + L_S - R_S, \quad (10)$$

[0138] 이고, 여기서 L은 좌 채널 오디오 신호를 나타내고, R은 우 채널 오디오 신호를 나타내고, L_S 는 좌 서라운드 채널 오디오 신호를 나타내고, R_S 는 우 서라운드 채널 오디오 신호를 나타낸다. 또 하나의 실시예에서, 전력 P_S 는 $L-R$ 의 전력과 L_S-R_S 의 전력의 합으로서 결정될 수 있다.

[0139] 잔여 오디오 신호 S 및 잔여 오디오 신호의 전력 P_S 는 7.1 멀티-채널 오디오 신호 포맷과 같은, 다른 멀티-채널 오디오 신호 포맷들을 사용하여 그에 따라 결정될 수 있다.

[0140] 계산 복잡성을 더욱 감소시키기 위해, 오디오 신호들의 주파수 빈들은 예를 들어 벨 주파수 스케일에 따라, 주파수 대역들로 함께 그룹화될 수 있다. 이 경우에, 이득 함수 G는 각각의 주파수 빈에 대해 결정될 수 있다.

[0141] 또한, 예를 들어, 100Hz 내지 8000Hz의 주파수 범위 내의 사람의 음성을 가능하게 포함할 수 있는 주파수들만을 처리하는 것은 비음성 성분들을 필터링하여 제거하는 데 도움을 준다.

[0142] 음성 향상의 실시예들은 업-믹싱 과정 동안에 중심 채널 오디오 신호 C 내로 누설된 원하지 않은 비음성 성분들을 제거한다. 또한, 그것은 멀티-채널 오디오 신호의 다른 채널들 내로 누설된 직접적인 성분들을 부스팅시킨다.

[0143] 도 6은 실시예에 따른 신호 처리 장치(100)의 음성 활성화도 검출기(601)의 다이어그램을 도시한다. 음성 활성화도 검출기(601)는 좌 채널 오디오 신호 L, 중심 채널 오디오 신호 C, 및 우 채널 오디오 신호 R에 기초하여 음성 활성화도 표시자 V를 결정하도록 구성되고, 여기서, 음성 활성화도 표시자 V는 시간에 걸쳐 멀티-채널 오디오 신호 내의 음성 성분의 크기를 표시한다. 음성 활성화도 검출기(601)는 감산기(603), 결정기(605), 결정기(607), 지연기(609), 지연기(611), 감산기(613), 감산기(615), 결정기(617), 결정기(619), 및 결정기(621)를 포함한다.

[0144] 감산기(603)는 잔여 오디오 신호 S를 획득하기 위해 좌 채널 오디오 신호 L로부터 우 채널 오디오 신호 R을 감산하도록 구성된다. 결정기(605)는 $|C(m,k)|$ 를 획득하기 위해 중심 채널 오디오 신호 C의 크기를 결정하도록 구성되고, 여기서 m은 샘플 시간 인덱스를 나타내고 k는 주파수 빈 인덱스를 나타낸다. 결정기(607)는 $|S(m,k)|$ 를 획득하기 위해 잔여 오디오 신호 S의 크기를 결정하도록 구성되고, 여기서 m은 샘플 시간 인덱스를 나타내고 k는 주파수 빈 인덱스를 나타낸다. 지연기(609)는 $|C(m-1,k)|$ 를 획득하기 위해 $|C(m,k)|$ 를 샘플 기간(sample time period)만큼 지연하도록 구성된다. 지연기(611)는 $|S(m-1,k)|$ 를 획득하기 위해 $|S(m,k)|$ 를 샘플 기간만큼 지연하도록 구성된다. 감산기(613)는 $|C(m,k)| - |C(m-1,k)|$ 를 획득하기 위해 $|C(m,k)|$ 로부터 $|C(m-1,k)|$ 를 감산하도록 구성된다. 감산기(615)는 $|S(m,k)| - |S(m-1,k)|$ 를 획득하기 위해 $|S(m,k)|$ 로부터 $|S(m-1,k)|$ 를 감산하도록 구성된다.

[0145] 결정기(617)는 예를 들어, $|C(m,k)| - |C(m-1,k)|$ 에 대해 모든 주파수 빈들에 걸쳐 제공된 합 Σ^2 에 기초하여, 중심 채널 오디오 신호 C의 스펙트럼 변화의 척도 F_c , 예를 들어, 스펙트럼 플럭스를 결정하도록 구성된다. 결정기(619)는 예를 들어, $|S(m,k)| - |S(m-1,k)|$ 에 대해 모든 주파수 빈들에 걸쳐 제공된 합 Σ^2 에 기초하여, 좌 채널 오디오 신호 L과 우 채널 오디오 신호 R 사이의 차이의 스펙트럼 변화의 척도 F_s , 예를 들어, 스펙트럼 플럭스를 결정하도록 구성된다. 결정기(621)는 스펙트럼 변화의 척도 F_c 및 스펙트럼 변화의 척도 F_s 에 기초하여, 예를 들어 계수 $F_c / (F_c + F_s)$ 에 기초하여 음성 활성화도 표시자 V를 결정하도록 구성된다.

[0146] 음성 활성화도 검출은 음성의 일시적 검출 및 세그멘테이션의 과정을 포함한다. 음성 활성화도 검출의 목적은 조용한 가운데서 또는 다른 사운드들 중에서 음성을 검출하는 것이다. 이러한 방식은 음성 기술의 거의 어느 종류에 대해서도 바람직할 수 있다.

[0147] 음성 활성화도 검출에 대한 다양한 다른 방식들이 본 발명의 실시예들에서 적용될 수 있다. 간단한 방식은 예를 들어 에너지-기반이다. 에너지 임계가 음성을 검출하기 위해 사용될 수 있다. 전형적으로, 이러한 방식은 단지 조용한 가운데에서의 음성에 대해서만 효과적이다. 다른 방식들은 신호 대 잡음비(SNR) 평가에 기초하고 통계적 음성 향상 방식들과 유사한 통계적 모델-기반 방식들을 포함한다. 파라메트릭 모델-기반 방식들은 보통 저-레벨 오디오 특징들을 가우션 혼합 모델과 같은 분류기와 결합시킨다. 가능한 오디오 특징들은 4Hz 변조 에너지, 제로 크로싱 레이트, 스펙트럼 중심, 또는 스펙트럼 플럭스이다.

[0148] 본 발명의 실시예에서, 음성 활성화도 검출은 단지 음성 또는 대화 성분들만이 부스팅되고 비음성 성분들은 변화되지 않은 채로 남게 하기 위해 이용된다. 음성 향상 방식의 개관이 도 6에 주어진다.

[0149] 음성 활성화도 표시자 V는 그것이 음성 향상 방식 내에서 행해질 수 있음에 따라, 중심 채널 오디오 신호 C 및 잔여 오디오 신호 $S = L - R$ 로부터 도출된다. 이들 오디오 신호로부터, 스펙트럼 플럭스가 추출된다. 스펙트럼 플럭스는 스펙트럼의 시간적 변화의 척도이다. DFT 또는 주파수 영역 신호 X의 스펙트럼 플럭스는 다음 식으로서 정의될 수 있다.

[0150]
$$F_x(m) = \sum_k (|X(m,k)| - |X(m-1,k)|)^2 \quad (11)$$

[0151] 스펙트럼 플럭스의 다른 유사한 정의들이 또한 본 발명의 다른 실시예들에서 이용될 수 있다. 스펙트럼 플럭스는 스펙트럼 에너지 분포의 변화를 표시하고 시간에 따른 시간적 도함수를 나타낸다. 차이가 2개의 연속하는 오디오 신호 프레임들에 걸쳐 결정되는, 식(11)에서의 정의 대신에, 스펙트럼 플럭스는 또한 다수의 오디오 신호 프레임을 포함하는 2개의 연속하는 블록들에 대한 차이로서 결정될 수 있다. 음성 성분들을 갖는 오디오 신호들에 대해, 음악 및 다른 사운드들에 비해 스펙트럼 플럭스의 보다 높은 값들이 예상된다.

[0152] 본 발명의 실시예에서, 예를 들어, 멀티-채널 오디오 신호의 한 채널이 주로 음성을 포함하는, 특정한 채널 셋업이 주파수-중속 연속적 음성 활성화도 표시자 V를 도출하기 위해 이용된다. 중심 채널 오디오 신호 C의 스펙트럼 플럭스 F_c 및 잔여 오디오 신호 S의 스펙트럼 플럭스 F_s 는 다음에 식(11)에 따라 결정될 수 있다.

[0153] 임의의 정규화 과정에 독립한 음성 활성화도 표시자 V 를 획득하기 위해, 음성 활성화도 표시자 V 는 예를 들어, 다음 식으로서 계산될 수 있다.

$$V = a \times \left(\frac{F_c}{F_c + F_s} - 0.5 \right) \quad (12)$$

[0154]

[0155] 음성 활성화도 표시자 V 의 이 정의는 $F_c = F_s$ 인 경우에 $V=0$ 인 것을 보장한다. 마지막으로, V 는 $V \in [0;1]$ 로 제한된다. 파라미터 a 는 V 의 동적 범위를 제어하는 미리 결정된 스케일링 팩터를 나타내고, 여기서 $a=4$ 는 다음 식을 산출하는 허용가능한 값일 수 있다.

$$V = 4 \times \left(\frac{F_c}{F_c + F_s} - 0.5 \right) \quad (13)$$

[0156]

[0157] 또한, 음성 활성화도 표시자 V 는 F_c 가 소정의 임계값 t 을 초과하지 않는 경우에 $V=0$ 으로 설정될 수 있다. 시간에 걸쳐 매끄러운 음성 활성화도 표시자 곡선을 획득하기 위해, 시간적 평활화가 V 에 적용될 수 있다.

[0158] 음성 향상 방식과 유사하게, 음성 활성화도 검출 방식은 주파수 빈들이 예를 들어 멜 주파수 스케일에 따라, 주파수 대역들로 그룹화될 때 또한 수행될 수 있다. 또한, 고려된 주파수들을 사람의 음성의 주파수 범위, 예를 들어, 100 내지 8000Hz로 제한하면, 성능이 더욱 개선된다.

[0159] 음성 활성화도 검출 방식의 결과는 간단하고 효율적인 알고리즘을 사용하여 획득된 주파수-독립 연속적 판정이다. 그것은 단지 몇개의 조정가능한 파라미터들을 이용할 수 있고 예를 들어 모델을 학습하기 위해, 더 이상의 어떤 데이터를 사용하지 않을 수 있다. 이 방식은 음성과 음악과 같은 다른 사운드들 사이를 강건하게 구별할 수 있다.

[0160] 도 7은 실시예에 따른 멀티-채널 오디오 신호 내의 음성 성분을 향상시키는 신호 처리 장치(100)의 다이어그램을 도시한다. 다이어그램은 믹싱 과정을 도시한다. 신호 처리 장치(100)는 도 1과 관련하여 설명된 것과 같은 신호 처리 장치의 가능한 구현을 형성한다. 신호 처리 장치(100)는 필터(101), 조합기(103), 및 음성 활성화도 검출기(601)를 포함한다.

[0161] 필터(101)는 도 5의 필터(101)와 관련하여 설명된 기능성을 제공한다. 음성 활성화도 검출기(601)는 도 6 내의 음성 활성화도 검출기(601)와 관련하여 설명된 기능성을 제공한다.

[0162] 실시예에서, 조합기(103)는 좌 채널 오디오 신호 L 를 가중된 좌 채널 오디오 신호 L_E 와 조합하여 조합된 좌 채널 오디오 신호 L_{EV} 를 획득하고, 중심 채널 오디오 신호 C 를 가중된 중심 채널 오디오 신호 C_E 와 조합하여 조합된 중심 채널 오디오 신호 C_{EV} 를 획득하고, 우 채널 오디오 신호 R 를 가중된 우 채널 오디오 신호 R_E 와 조합하여 조합된 우 채널 오디오 신호 R_{EV} 를 획득하도록 구성된다. 조합기는 가산기(701), 가산기(703), 가산기(705), 가중기(707), 가중기(709), 가중기(711), 및 가중기(713)를 포함한다.

[0163] 실시예에서, 가중기(713)는 음성 활성화도 표시자 $V(m)$ 을 미리 결정된 음성 이득 팩터 G_S 로 가중하여 가중된 음성 활성화도 표시자 $V_G = G_S V(m)$ 을 획득하도록 구성되고, 여기서 m 은 샘플 시간 인덱스를 나타낸다. 조합기는 좌 채널 오디오 신호 L , 중심 채널 오디오 신호 C , 및 우 채널 오디오 신호 R 을 미리 결정된 입력 이득 팩터 G_{in} 으로 가중하도록 구성되는 도면에 도시하지 않은 추가 가중기를 포함할 수 있다.

[0164] 가중기(707)는 가중된 좌 채널 오디오 신호 L_E 를 가중된 음성 활성화도 표시자 $V_G = G_S V(m)$ 로 가중하도록 구성되고, 가산기(701)는 조합된 좌 채널 오디오 신호 L_{EV} 를 획득하기 위해 좌 채널 오디오 신호 L 에 그 결과를 가산하도록 구성된다. 가중기(709)는 가중된 중심 채널 오디오 신호 C_E 를 가중된 음성 활성화도 표시자 $V_G = G_S V(m)$ 로 가중하도록 구성되고, 가산기(703)는 조합된 중심 채널 오디오 신호 C_{EV} 를 획득하기 위해 중심 채널 오디오 신호 C 에 그 결과를 가산하도록 구성된다. 가중기(711)는 가중된 우 채널 오디오 신호 R_E 를 가중된 음성 활성화도 표시자 $V_G = G_S V(m)$ 로 가중하도록 구성되고, 가산기(705)는 조합된 우 채널 오디오 신호 R_{EV} 를 획득하기 위해 우 채널 오디오 신호 R 에 그 결과를 가산하도록 구성된다.

[0165] 실시예에서, 가중기(713)는 가중된 좌 채널 오디오 신호 L_E , 가중된 중심 채널 오디오 신호 C_E , 및 가중된 우 채널 오디오 신호 R_E 를 미리 결정된 음성 이득 팩터 G_S 로 가중하도록 구성된다. 조합기(103)는 좌 채널 오디오 신호 L , 중심 채널 오디오 신호 C , 및 우 채널 오디오 신호 R 를 미리 결정된 입력 이득 팩터 G_{in} 으로 가중하도록 구성되는 도면에 도시하지 않은 추가 가중기를 포함할 수 있다.

[0166] 미리 결정된 음성 이득 팩터 G_S 는 또한 음성 활성화도 검출기(601)가 사용되지 않는 경우에 적용될 수 있다. 간단히 하기 위해, 가중기(713)는 도면에 단일의 가중기(713)로서 도시된다. 가능한 구현에서, 가중기(713)는 3번, 특정하게, 가중기(709)와 가산기(703) 사이에서, 가중기(707)와 가산기(701) 사이에서, 및 가중기(711)와 가산기(705) 사이에서 사용된다. 음성 활성화도 검출기(601)가 사용되지 않는 경우에, $V = 1$ 이 가정될 수 있고, G_S 가 V 를 수정하기 위해 사용될 수 있다.

[0167] 음성 향상 및 음성 활성화도 검출의 결과들은 그러므로 깨끗한 음성 오디오 신호의 평가를 획득하기 위해 조합될 수 있다. 음성 향상과 음성 활성화도 검출은 설명된 것과 같이 동시에 수행될 수 있다. 음성 활성화도 표시자 V 는 음성 이득 팩터 G_S 로 가중기(713)에 의해 가중 또는 승산될 수 있고, 여기서 $V_G = V G_S$ 는 음성 부스트를 제어하기 위해 사용될 수 있다. V_G 는 가중된 오디오 신호들 L_E , C_E , 및 R_E 로 증배식으로 가중기들(707, 709, 711)에 의해 조합될 수 있고 결과적인 오디오 신호들은 다음 식들:

$$C_{EV}(m,k) = G_{in} \times C + G_S \times V(m) \times G(m,k) \times C(m,k) \quad (14)$$

$$L_{EV}(m,k) = G_{in} \times L + G_S \times V(m) \times G(m,k) \times L(m,k) \quad (15)$$

$$R_{EV}(m,k) = G_{in} \times R + G_S \times V(m) \times G(m,k) \times R(m,k) \quad (16)$$

[0168] 에 따라 신호 처리 장치(100)의 최종 조합된 오디오 신호들 L_{EV} , C_{EV} , 및 R_{EV} 를 획득하기 위해 원래의 오디오 신호들 L , C , 및 R 에 가산기들(701, 703, 705)에 의해 가산될 수 있고, G_{in} 은 원래의 오디오 신호들 상에 적용된 입력 이득 팩터이다. 이 팩터는 멀티-채널 오디오 신호로 구성된 비음성 성분들의 이득을 제어한다. G_{in} 및 G_S , 예를 들어, $G_{in} = 1$ 및 $G_S = -1$ 의 특정한 조합들이 멀티-채널 오디오 신호로부터 음성 성분을 제거하기 위해 사용될 수 있다. 음성 성분을 부스팅시키기 위한 적절한 설정들은 $G_{in} = 1$ 일 수 있고 G_S 는 1과 4 사이의 범위에 있을 수 있다. 최종 조합된 오디오 신호들 L_{EV} , C_{EV} , 및 R_{EV} 는 다음에 시간 영역으로 다시 변환될 수 있고 스테레오 다운-믹스를 생성하기 위해 사용될 수 있다.

[0170] 결과적으로, 음성 또는 대화 향상의 문제에 대한 계산적으로 저렴하면서 효율적인 해결책이 제공된다. 모든 성분들은 DFT 주파수 영역에서 동작할 수 있다. 예를 들어, 5.1 서라운드 오디오 신호 내의 중심 채널 오디오 신호 C 가 부스팅되고 중심 채널 오디오 신호 C 내의 모든 사운드들이 향상되는 간단한 방식에 비교하여, 본 발명의 실시예들에서 중심 채널 오디오 신호 C 내의 음성 성분만이 예를 들어 음성 활성화도 검출로 인해 부스팅된다. 게다가, 본 발명의 실시예들은 또한 동시의 음성 및 비음성 성분들을 처리하고, 여기서 음성 성분들만이 예를 들어, 음성 향상 방식으로 인해 부스팅된다.

[0171] 중심 채널 오디오 신호 C 뿐만 아니라, 다른 오디오 신호들(예를 들어, L 및 R)이 음성 향상 및 음성 활성화도 검출을 사용하여 처리된다는 사실은 최종 오디오 신호들이 고 품질의 공간적으로 넓은 음성 성분을 포함하는 것을 보장한다. 이것은 중심 채널 오디오 신호 C 만이 처리될 때만 그런 것은 아니다. 본 발명의 실시예들은 특정한 코덱, 믹스, 또는 5.1 서라운드 오디오 신호와 같은, 멀티-채널 오디오 신호 포맷에 독립적이고, 상이한 채널 구성들로 확장될 수 있다.

[0172] 본 발명, 및 특히 신호 처리 장치의 실시예들은 예를 들어, 도 1 내지 7에 기초하여 여기에 설명된 필터(101), 조합기(103) 및/또는 다른 유닛들 또는 단계들의, 여기에 설명된 장치 및 방법들의 다양한 기능성들을 구현하도록 구성되는 단일 또는 다수의 프로세서를 포함할 수 있다.

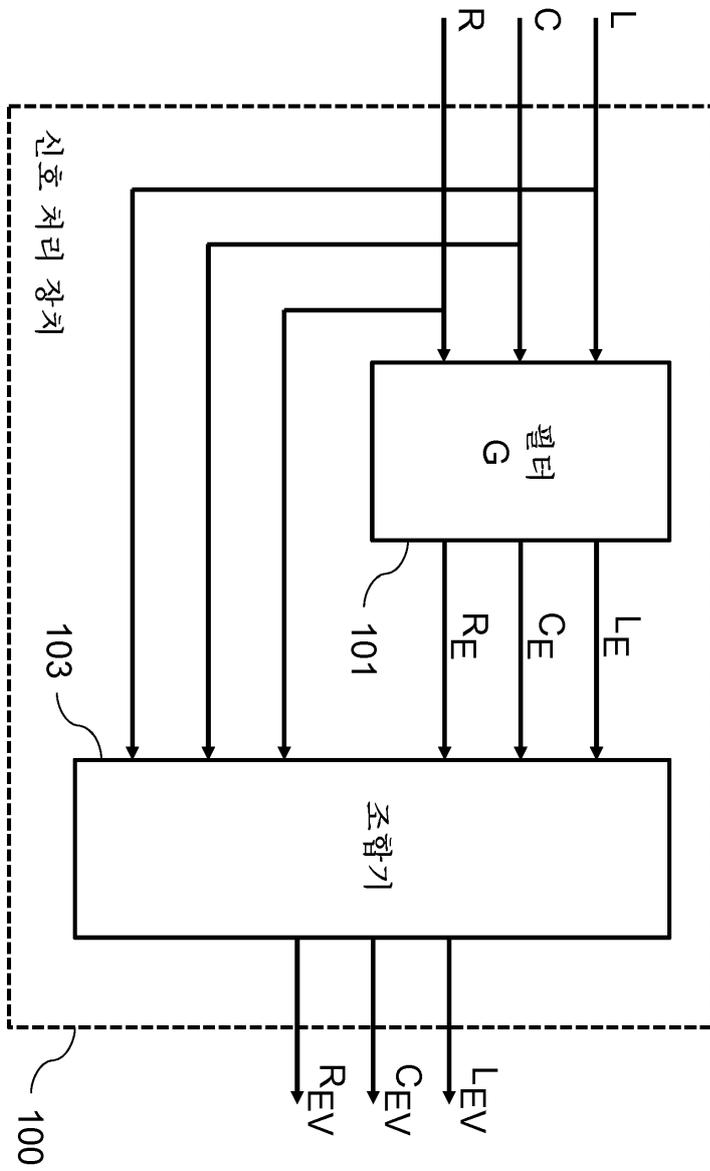
[0173] 본 발명의 방법들의 소정의 구현 조건들에 따라, 본 발명의 방법들은 하드웨어 또는 소프트웨어 또는 이들의 임의의 조합에서 구현될 수 있다.

[0174] 상기 구현들은 본 발명의 방법들 중 적어도 하나의 실시예가 수행되도록 프로그램가능한 컴퓨터 시스템과 협력하거나 협력할 수 있는 전자적으로 판독가능한 제어 신호들이 그에 저장되어 있는 디지털 저장 매체, 특히 플로

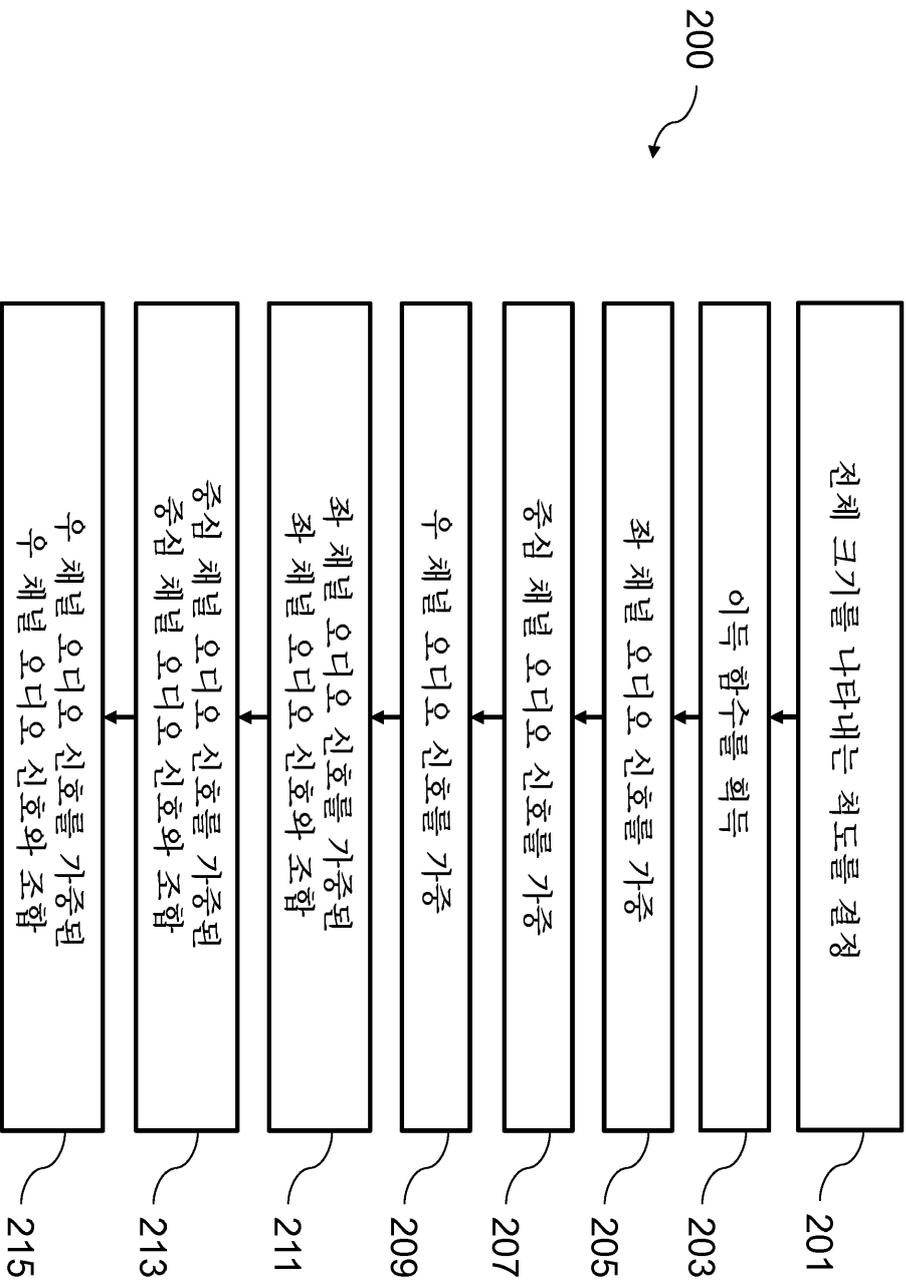
피 디스크, CD, DVD, 또는 블루-레이 디스크, ROM, PROM, EPROM, EEPROM 또는 플래시 메모리를 사용하여 수행될 수 있다.

- [0175] 본 발명의 추가 실시예는 머신 판독가능 캐리어 상에 저장된 프로그램을 갖는 컴퓨터 프로그램 제품이거나, 따라서 그를 포함하고, 프로그램 코드는 컴퓨터 프로그램 제품이 컴퓨터 상에서 실행할 때 본 발명의 방법들 중 적어도 하나를 수행하기 위해 이용가능하다.
- [0176] 바꾸어 말하면, 본 발명의 방법들의 실시예들은 컴퓨터 프로그램이 컴퓨터 상에서, 또는 프로세서 등 상에서 실행할 때 본 발명의 방법들 중 적어도 하나를 수행하는 프로그램 코드를 갖는 컴퓨터 프로그램이거나, 따라서 그를 포함한다.
- [0177] 본 발명의 추가 실시예는 컴퓨터 프로그램 제품이 컴퓨터 상에서, 또는 프로세서 등 상에서 실행할 때 본 발명의 방법들 중 적어도 하나를 수행하기 위해 이용가능한 컴퓨터 프로그램이 그에 저장되어 있는 머신 판독가능 디지털 저장 매체이거나, 따라서 그를 포함한다.
- [0178] 본 발명의 추가 실시예는 컴퓨터 프로그램 제품이 컴퓨터 상에서, 또는 프로세서 등 상에서 실행할 때 본 발명의 방법들 중 적어도 하나를 수행하기 위해 이용가능한 컴퓨터 프로그램을 나타내는 데이터 스트림 또는 신호들의 시퀀스이거나, 따라서 그들을 포함한다.
- [0179] 본 발명의 추가 실시예는 본 발명의 방법들 중 적어도 하나를 수행하도록 적응된 컴퓨터 프로세서 또는 기타 프로그램가능한 논리 디바이스이거나, 따라서 그들을 포함한다.
- [0180] 본 발명의 추가 실시예는 컴퓨터 프로그램 제품이 컴퓨터, 프로세서 또는 기타 프로그램가능한 논리 디바이스, 예를 들어, FPGA(필드 프로그램가능한 게이트 어레이) 또는 ASIC(주문형 집적 회로) 상에서 실행할 때 본 발명의 방법들 중 적어도 하나를 수행하기 위해 이용가능한 컴퓨터 프로그램이 그에 저장되어 있는 컴퓨터 프로세서 또는 기타 프로그램가능한 논리 디바이스이거나, 따라서 그들을 포함한다.
- [0181] 상기 내용이 그 특정한 실시예들을 참조하여 특정하게 도시되고 설명되었지만, 형태 및 상세들에서의 다양한 다른 변화들이 본 발명의 취지 및 범위에서 벗어나지 않고서 이루어질 수 있다는 것이 본 기술 분야의 통상의 기술자들에 의해 이해될 것이다. 그러므로 다양한 변화들이 여기에 개시되고 다음의 청구범위에 의해 이해되는 폭넓은 개념에서 벗어나지 않고서 상이한 실시예들에 적응하여 이루어질 수 있다는 것을 이해할 것이다.

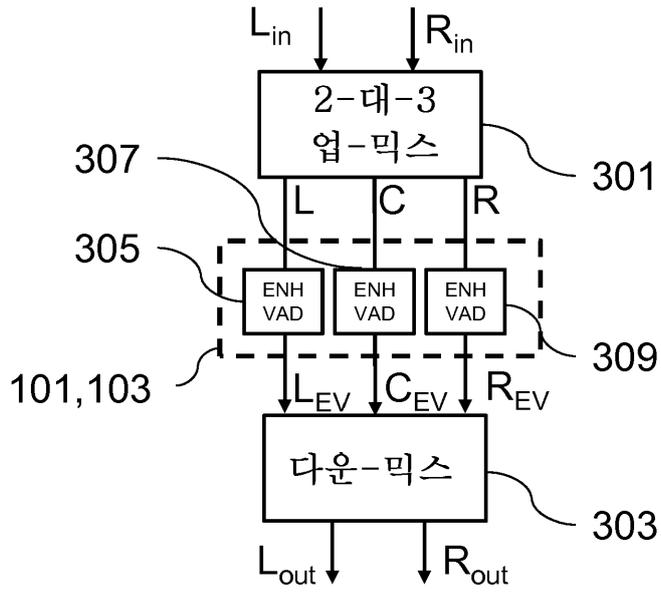
도면
도면1



도면2

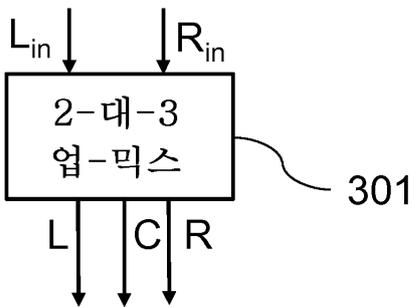


도면3

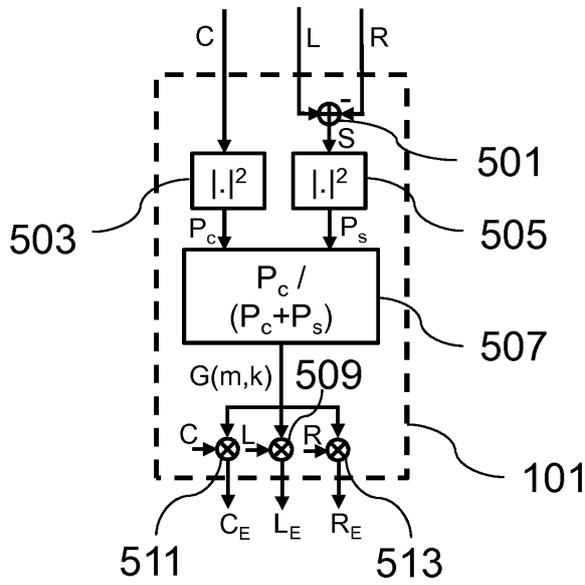


100

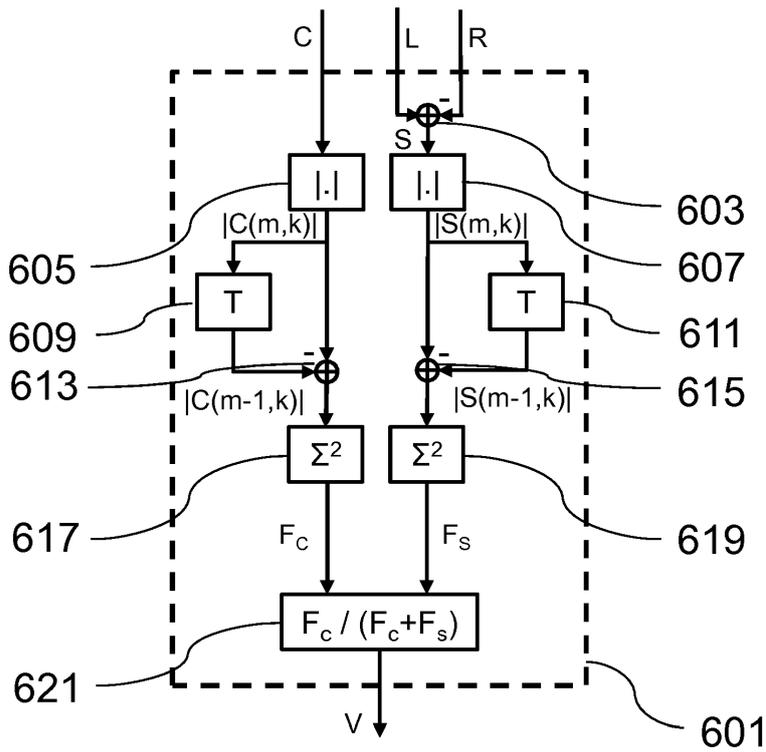
도면4



도면5



도면6



도면7

