



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0021580
(43) 공개일자 2023년02월14일

(51) 국제특허분류(Int. Cl.)
G10L 21/02 (2006.01) G10L 25/03 (2013.01)
G10L 25/78 (2013.01)
(52) CPC특허분류
G10L 21/02 (2021.08)
G10L 25/03 (2013.01)
(21) 출원번호 10-2022-0088509
(22) 출원일자 2022년07월18일
심사청구일자 없음
(30) 우선권주장
202110895493.X 2021년08월05일 중국(CN)

(71) 출원인
하만인터내셔널인더스트리스인코포레이티드
미국 코네티컷 스태포드 애틀랜틱 스트리트400 (우: 06901)
(72) 발명자
시 샤오-푸
미국 캘리포니아 94043 마운틴 뷰 엘리스 스트리트 636
정 지안웬
중국 선전 518067 난산 서커우 공예 써드 로드 넘버 1 차이나 머천츠 포트 프라자 20층
(뒷면에 계속)
(74) 대리인
리앤목특허법인

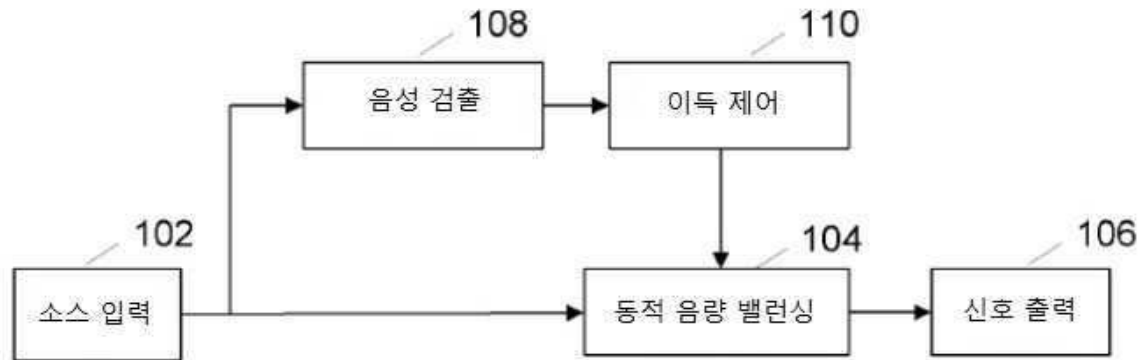
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 **동적 음성 향상을 위한 방법 및 시스템**

(57) 요약

본 개시는 음성 향상을 위한 방법 및 시스템을 제공한다. 본 발명의 방법 및 시스템은 입력 신호에 대해 두 경로의 신호 처리를 동시에 수행할 수 있다. 제1 경로 신호 처리는 오디오 소스 입력을 수신하고 제1 이득 제어 파라미터에 기초하여 오디오 소스 입력에 대해 동적 음량 밸런싱을 수행하는 것을 포함한다. 제2 경로 신호 처리는 오디오 소스 입력에 대해 음성 검출을 수행하고 검출 신뢰도를 계산하는 단계; 및 검출 신뢰도에 기초하여 제2 이득 제어 파라미터를 계산하는 단계를 포함한다. 제1 경로 신호 처리 및 제2 경로 신호 처리는 동기식 또는 비동기식일 수 있다. 본 개시의 방법은 또한 제1 이득 제어 파라미터를 제2 처리 경로에 의해 계산된 제2 이득 제어 파라미터로 업데이트하는 단계, 및 업데이트된 제1 이득 제어 파라미터에 기초하여 제1 경로 신호 처리를 수행하는 단계를 포함한다.

대표도 - 도1



(52) CPC특허분류

G10L 25/78 (2013.01)

(72) 발명자

샤오 이

중국 선전 518067 난산 서커우 공예 씨드 로드 넘
버 1 차이나 머천즈 포트 플라자 20층

자오 예빈

중국 광둥 프라빈스 518052 선전 시티 난산 디스트
릭트 사우스 케위안 로드 넘버 3331 알리 센터 빌
딩 티2 베이스먼트

명세서

청구범위

청구항 1

동적 음성 향상 방법으로서,

오디오 소스 입력을 수신하고 제1 이득 제어 파라미터에 기초하여 오디오 소스 입력에 대해 동적 음량 밸런싱을 수행하는 것을 포함하는 제1 경로 신호 처리를 수행하는 단계;

제2 경로 신호 처리를 수행하는 단계 - 상기 제2 경로 신호 처리는,

오디오 소스 입력에 대해 음성 검출을 수행하고 검출 신뢰도를 계산하는 단계로서, 검출 신뢰도는 오디오 소스 입력의 음성 가능성을 나타내는, 단계; 및

검출 신뢰도에 기초하여 제2 이득 제어 파라미터를 계산하는 단계를 포함함; 그리고

제1 이득 제어 파라미터를 제2 이득 제어 파라미터로 업데이트하고, 업데이트된 제1 이득 제어 파라미터에 기초하여 제1 경로 신호 처리를 수행하는 단계를 포함하는, 동적 음성 향상 방법.

청구항 2

제1항에 있어서,

상기 오디오 소스 입력은 멀티-채널 소스 입력을 포함하고,

상기 오디오 소스 입력에 대해 음성 검출을 수행하고 검출 신뢰도를 계산하는 단계는,

멀티-채널 소스 입력으로부터 중앙 채널 신호를 추출하는 단계;

중앙 채널 신호에 대해 정규화를 수행하는 단계; 그리고

정규화된 중앙 채널 신호에 대해 고속 자기상관을 수행하는 단계를 포함하며,

고속 자기상관의 결과는 검출 신뢰도를 나타내는, 동적 음성 향상 방법.

청구항 3

제1항에 있어서,

상기 검출 신뢰도에 기초하여 제2 이득 제어 파라미터를 계산하는 단계는,

검출 신뢰도의 대수 함수에 기초하여 제2 이득 제어 파라미터를 계산하는 단계;

계산된 제2 이득 제어 파라미터를 평활화하는 단계; 그리고

평활화된 제2 이득 제어 파라미터를 제한하는 단계를 포함하는, 동적 음성 향상 방법.

청구항 4

제1항에 있어서,

상기 오디오 소스 입력은 멀티-채널 소스 입력을 포함하고,

상기 오디오 소스 입력에 대해 동적 음량 밸런싱을 수행하는 단계는,

멀티-채널 소스 입력으로부터 중앙 채널 신호를 추출하는 단계;

제1 이득 제어 파라미터 또는 업데이트된 제1 이득 제어 파라미터에 기초하여 중앙 채널 신호의 음량을 향상시키고 다른 채널 신호의 음량을 감소시키는 단계; 그리고

향상된 중앙 채널 신호와 감소된 다른 채널 신호를 연결 및 혼합하여 출력 신호를 생성하는 단계를 포함하는, 동적 음성 향상 방법.

청구항 5

제1항에 있어서,

동적 음량 밸런싱을 수행하기 전에 오디오 소스 입력에 대해 크로스오버 필터링을 수행하는 단계를 더 포함하는, 동적 음성 향상 방법.

청구항 6

제5항에 있어서,

오디오 소스 입력의 중간 주파수 범위의 신호에 대해서만 동적 음량 밸런싱을 수행하는 단계; 그리고

상기 동적 음량 밸런싱 후에 상기 오디오 소스 입력의 저주파수 대역 및 고주파수 대역 신호와 상기 오디오 소스 입력의 중간 주파수 대역 신호를 연결 및 혼합하여 출력 신호를 생성하는 단계를 더 포함하는, 동적 음성 향상 방법.

청구항 7

제1항에 있어서,

상기 오디오 소스 입력은 이중-채널 소스 입력을 더 포함하고,

상기 방법은 이중-채널 소스 입력에 기초하여 멀티-채널 소스 입력을 생성하는 단계를 더 포함하는, 동적 음성 향상 방법.

청구항 8

제7항에 있어서,

상기 이중-채널 소스 입력에 기초하여 멀티-채널 소스 입력을 생성하는 단계는,

이중-채널 소스 입력으로부터 좌측 채널 신호와 우측 채널 신호 사이의 교차 상관을 수행하는 단계; 그리고

조합 비율에 따라 멀티-채널 소스 입력을 생성하는 단계를 포함하며,

상기 조합 비율은 교차 상관의 결과에 따라 좌우되는, 동적 음성 향상 방법.

청구항 9

제1항에 있어서,

상기 제1 경로 신호 처리 및 상기 제2 경로 신호 처리가 동기식 또는 비동기식인, 동적 음성 향상 방법.

청구항 10

동적 음성 향상 시스템으로서,

컴퓨터-실행가능 명령어를 저장하도록 구성된 메모리; 그리고

제1항 내지 제9항 중 어느 한 항에 따른 방법을 구현하도록 컴퓨터 실행 가능 명령어를 실행하도록 구성된 프로세서를 포함하는, 동적 음성 향상 시스템.

발명의 설명

기술 분야

[0001] 본 개시는 일반적으로 오디오 신호 처리 분야에 관한 것으로, 보다 상세하게는 오디오 소스의 동적 음성 향상을 위한 방법 및 시스템에 관한 것이다.

배경 기술

[0002] 고화질 케이블 TV 및 온라인 스트리밍과 같은 새로운 미디어 소비 방식과 대형 TV 및 디스플레이의 등장으로 소비자 시장에서 영화 경험이 인기를 얻고 있다. 이러한 미디어 소스에는 종종 멀티-채널 오디오 기술이 수반되거

나 일반적으로 서라운드 기술이라고 한다. Dolby, THX 및 DTS와 같은 서라운드 제공업체에는 소스 콘텐츠에 대해 더 나은 공간 오디오 해상도를 제공하는 자체 멀티-채널 오디오 인코딩 기술이 있다. 영화 형식 콘텐츠의 주요 목적은 몰입형 서라운드 경험을 제공하는 것이므로 서라운드 경험을 위해 음성 명료도를 희생하는 것이 선호되는 경우가 많다. 이는 몰입도와 공간 해상도 측면에서 큰 이점을 제공하지만 종종 음질이 좋지 않고 영화 내용을 이해하는 데 어려움을 겪을 수도 있다. 영화 콘텐츠 소스의 음성 품질을 개선하여 명료도 및 가청도를 향상시키기 위해 영화 콘텐츠에 음성 향상 방법을 적용하는 경우가 많다.

[0003] 기존의 음성 향상을 위한 일반적인 방법은 정적 이퀄라이제이션을 활용하는 것이다. 이 방법은 음성 대역의 음량을 높이기 위해 약 200Hz ~ 4kHz의 오디오 채널에만 정적 이퀄라이제이션을 적용한다. 이 구현에는 시스템 리소스가 거의 필요하지 않지만 이 방법에서 발생하는 왜곡은 분명하다. 이 구현 방식은 클립에 음성이나 대사가 없는 경우에도 항상 작동하므로, 음높이(pitch) 불균형이 발생하고 배경이 증폭된다. 보다 발전된 방법은 먼저 각 시간 프레임 내에서 음성을 감지한 다음, 감지 결과에 따라 오디오 신호를 자동으로 처리하는 것이다. 이 단방향 실행 방식은 음성의 정확한 감지와 시스템 처리의 빠른 응답을 요구한다. 그러나 기존의 일부 방법은 음성을 빠르고 정확하게 감지하지 못하고 신호에 색을 입혀 거칠게 하는 경우가 많다.

[0004] 따라서, 상술한 기존 솔루션의 단점을 극복하기 위한 개선된 기술 솔루션이 필요하다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0005] 본 개시의 일 양태에 따르면, 동적 음성 향상 방법이 제공된다. 방법은 제1 경로 신호 처리를 수행하는 단계를 포함할 수 있고, 제1 경로 신호 처리는 오디오 소스 입력을 수신하는 단계 및 제1 이득 제어 파라미터에 기초하여 오디오 소스 입력에 대해 동적 음량 밸런싱을 수행하는 단계를 포함한다. 방법은 또한 오디오 소스 입력에 대한 음성 검출을 수행하고 검출 신뢰도를 계산하는 것을 포함하는 제2 경로 신호 처리를 수행하는 단계 - 검출 신뢰도는 오디오 소스 입력의 음성 가능성을 표시함; 및 검출 신뢰도에 기초하여 제2 이득 제어 파라미터를 계산하는 단계를 포함한다. 방법은 제1 이득 제어 파라미터를 제2 이득 제어 파라미터로 업데이트하는 단계, 및 업데이트된 제1 이득 제어 파라미터에 기초하여 제1 경로 신호 처리를 수행하는 단계를 더 포함할 수 있다.

[0006] 하나 이상의 실시예에 따르면, 오디오 소스 입력은 멀티-채널 소스 입력을 포함할 수 있고, 오디오 소스 입력에 대해 음성 검출을 수행하고 검출 신뢰도를 계산하는 단계는, 멀티-채널 소스 입력으로부터 중앙 채널 신호를 추출하는 단계; 중앙 채널 신호에 대해 정규화를 수행하는 단계; 및 정규화된 중앙 채널 신호에 대해 고속 자기상관을 수행하는 단계를 포함하며, 고속 자기상관의 결과는 검출 신뢰도를 나타낸다.

[0007] 하나 이상의 실시예에 따르면, 검출 신뢰도에 기초하여 제2 이득 제어 파라미터를 계산하는 단계는, 검출 신뢰도의 대수 함수에 기초하여 제2 이득 제어 파라미터를 계산하는 단계; 계산된 제2 이득 제어 파라미터를 평활화하는 단계; 및 평활화된 제2 이득 제어 파라미터를 제한하는 단계를 포함한다.

[0008] 하나 이상의 실시예에 따르면, 오디오 소스 입력은 멀티-채널 소스 입력을 포함할 수 있고, 오디오 소스 입력에 대해 동적 음량 밸런싱을 수행하는 단계는 멀티-채널 소스 입력으로부터 중앙 채널 신호를 추출하는 단계; 제1 이득 제어 파라미터 또는 업데이트된 제1 이득 제어 파라미터에 기초하여 중앙 채널 신호의 음량을 향상시키고 다른 채널 신호의 음량을 감소시키는 단계; 및 향상된 중앙 채널 신호와 감소된 다른 채널 신호를 연결 및 혼합하여 출력 신호를 생성하는 단계를 포함한다.

[0009] 하나 이상의 실시예에 따르면, 방법은 동적 음량 밸런싱을 수행하기 전에 오디오 소스 입력에 대해 크로스오버 필터링을 수행하는 단계를 또한 포함할 수 있다.

[0010] 하나 이상의 실시예에 따르면, 방법은 또한 오디오 소스 입력의 중간 주파수 범위의 신호에 대해서만 동적 음량 밸런싱을 수행하는 단계; 및 상기 동적 음량 밸런싱 이후에, 상기 오디오 소스 입력의 저주파수 대역 및 고주파수 대역의 신호와 상기 오디오 소스 입력 중 중주파수 대역의 신호를 연결 및 믹싱하여 출력 신호를 생성하는 단계를 포함할 수 있다.

[0011] 하나 이상의 실시예에 따르면, 오디오 소스 입력은 또한 이중-채널 소스 입력을 포함하고, 방법은 또한 이중-채

널 소스 입력에 기초하여 멀티-채널 소스 입력을 생성하는 단계를 포함한다.

[0012] 하나 이상의 실시예에 따르면, 이중-채널 소스 입력에 기초하여 멀티-채널 소스 입력을 생성하는 단계는: 이중-채널 소스 입력으로부터 좌측 채널 신호와 우측 채널 신호 사이의 교차 상관을 수행하는 단계; 및 조합 비율에 따라 멀티-채널 소스 입력을 생성하는 단계를 포함한다. 조합 비율은 교차 상관의 결과에 따라 다르다.

[0013] 하나 이상의 실시예에 따르면, 제1 경로 신호 처리 및 제2 경로 신호 처리는 동기식 또는 비동기식이다.

[0014] 본 개시의 다른 양태에 따르면, 메모리 및 프로세서를 포함하는 음성 향상을 위한 시스템이 제공된다. 메모리는 컴퓨터-실행가능 명령어를 저장하도록 구성된다. 프로세서는 위에서 설명된 방법을 구현하기 위해 명령어를 실행하도록 구성된다.

도면의 간단한 설명

[0015] 본 개시내용은 첨부 도면을 참조하여 다음의 비제한적 구현의 설명을 읽음으로써 더 잘 이해될 수 있으며, 여기서:

도 1은 본 개시 내용의 구현의 하나 이상의 실시예들에 따른 음성 향상의 개략적인 블록도를 개략적으로 도시한다.

도 2는 본 개시내용의 하나 이상의 실시예에 따른 음성 검출의 개략적인 블록도를 예시적으로 도시한다.

도 3은 본 개시내용의 하나 이상의 실시예에 따른 음성 검출에 기초한 이득 추정의 개략적인 블록도를 예시적으로 도시한다.

도 4는 본 개시내용의 하나 이상의 실시예에 따른 동적 음량 밸런싱 프로세스의 개략도를 예시적으로 도시한다.

도 5는 본 개시내용의 다른 구현의 하나 이상의 실시예에 따른 음성 향상의 개략도를 도시한다.

도 6은 도 5의 구현의 하나 이상의 실시예에 따른 동적 음량 밸런싱 프로세스의 개략도를 도시한다.

도 7은 본 발명의 일 실시예에 따른 소스 입력이 이중-채널 소스 입력인 경우 이중-채널 소스 입력을 기반으로 멀티-채널 소스 입력을 생성하는 과정을 개략적으로 도시한 도면이다. 그리고

도 8은 본 개시내용의 하나 이상의 실시예에 따른 동적 음성 향상을 위한 방법을 개략적으로 도시한다.

발명을 실시하기 위한 구체적인 내용

[0016] 실시예에 대한 다음 설명은 제한이 아니라 예시의 목적으로만 제공된다는 것을 이해해야 한다. 도면에 도시된 기능 블록, 모듈 또는 유닛의 예의 구분은 이러한 기능 블록, 모듈 또는 유닛이 물리적으로 분리된 유닛으로 구현되어야 함을 의미하는 것으로 해석되어서는 안 된다. 도시되거나 설명된 기능 블록, 모듈 또는 유닛은 별도의 유닛, 회로, 칩, 기능 블록, 모듈 또는 회로 요소로서 구현될 수 있다. 하나 이상의 기능 블록 또는 유닛은 또한 공통 회로, 칩, 회로 요소 또는 유닛에서 구현될 수 있다.

[0017] 단수 용어(예: "a")의 사용은 항목 수를 제한하기 위한 것이 아니다. 관계 용어, 예를 들어 "위쪽", "아래쪽", "왼쪽", "오른쪽", "상부", "하부", "하향", "상향", "옆쪽", "제1", "제2"("제3" 등), "입구", "출구" 등은 도면에 대한 특정 참조의 명확성을 위해 서면 설명에서 사용된 것으로, 본 개시의 범위를 제한하려는 의도가 아니다. 달리 명시되지 않는 한, 첨부된 청구범위. 용어 "결합", "결합하는", "결합되는", "결합된", "커플러" 및 유사한 용어는 본 명세서에서 광범위하게 사용되며, 고정, 결합, 접촉, 조임, 부착, 연계(associating), 삽입, 그 위에 또는 내부에 형성, 통신 또는 그렇지 않으면 직접 또는 간접적으로 기계, 자기, 전기, 화학적 및 작동적으로 중간 요소 및 하나 이상의 부재와 연관되어, 또는, 일 부재가 다른 부재와 통합되는 방식으로 일체형으로 형성되는 경우를 또한 포함할 수 있지만 이에 국한되지 않는다. 커플링은 회전을 포함하여 모든 방향에서 발생할 수 있다. "포함하는" 및 "~와 같은"이라는 용어는 제한적인 것이 아니라 예시적인 것이며, "~할 수 있는"이라는 단어는 달리 명시되지 않는 한 "~할 수 있지만 반드시 그런 것은 아님"을 의미한다. 본 개시내용에서 임의의 다른 언어가 사용되지만, 도면에 도시된 실시예는 예시 및 설명을 위해 제공된 예이며 본 명세서의 주제의 유일한 실시예는 아니다.

[0018] 기존 기술 솔루션의 단점을 극복하고 음성 출력의 품질을 개선하여 사용자에게 더 나은 경험을 제공하기 위해, 본 발명은 사람의 음성을 능동적으로 감지하고, 오디오 소스 입력에서 음성의 가능성을 나타내는 감지 신뢰도를 기반으로, 오디오 소스(예를 들어 극장 오디오 소스)에서 음성 크기를 동적으로 향상시키는 솔루션을 제안한다.

본 발명의 방법 및 시스템은 입력 신호에 대해 두 경로의 신호 처리를 동시에 수행할 수 있다. 제1 경로 신호 처리는 오디오 소스 입력을 수신하고 제1 이득 제어 파라미터에 기초하여 오디오 소스 입력에 대해 동적 음량 밸런싱을 수행하는 것을 포함한다. 제2 경로 신호 처리는 오디오 소스 입력에 대해 음성 검출을 수행하고 검출 신뢰도를 계산하는 단계; 및 검출 신뢰도에 기초하여 제2 이득 제어 파라미터를 계산하는 단계를 포함한다. 제1 경로 신호 처리 및 제2 경로 신호 처리는 동기식 또는 비동기식일 수 있다. 본 개시의 방법은 또한 제1 이득 제어 파라미터를 제2 처리 경로에 의해 계산된 제2 이득 제어 파라미터로 업데이트하는 단계, 및 업데이트된 제1 이득 제어 파라미터에 기초하여 제1 경로 신호 처리를 수행하는 단계를 포함한다. 본 개시내용의 방법 및 시스템은 음성의 명료성을 더 향상시키고 오디오 제품을 사용하는 사용자의 경험을 향상시킬 수 있다.

[0019] 이하, 첨부된 도면을 참조하여 본 발명의 다양한 실시예에 따른 동적 음성 향상 방법 및 시스템을 상세히 설명한다. 도 1은 본 개시 내용의 구현의 하나 이상의 실시예들에 따른 음성 방법 및 시스템의 개략적인 블록도를 도시한다. 이해의 편의를 위해, 본 발명은 방법 및 시스템의 주요 처리 절차에 따라 여러 모듈을 참조하여 설명될 것이다. 모듈 설명에 대한 참조는 솔루션을 보다 명확하게 설명하기 위한 것이지 제한하기 위한 것이 아님을 당업자는 이해할 것이다.

[0020] 도 1은 본 개시 내용의 구현의 하나 이상의 실시예들에 따른 개략도를 도시한다. 도 1에 도시된 하나 이상의 실시예에서, 도 1에 도시된 바와 같이, 본 개시내용에서 오디오 소스 입력 신호를 처리하는 방법 및 시스템은 소스 입력 모듈(102), 동적 음량 밸런싱 모듈(104), 신호 출력 모듈(106), 음성 검출 모듈(108), 및 이득 제어 모듈(110)의 부분들을 포함한다. 도 1에서 알 수 있듯이, 본 발명의 방법 및 시스템은 입력 신호에 대해 두 경로의 신호 처리를 동시에 수행할 수 있다. 제1 경로 신호 처리는 주로 수신된 소스 입력 신호에 대한 동적 음량 균형을 수행하는 데 사용된다. 제2 경로 신호 처리는 수신된 소스 입력 신호에 대한 음성 감지를 수행하고 이득을 추정하는 데 사용된다. 제1 경로 신호 처리 및 제2 경로 신호 처리는 동기식 또는 비동기식으로 수행될 수 있다. 이는 실제 시스템의 처리 능력 및 대기 시간 요건에 따라 다르다. 소스 입력 신호에 대한 이 이중-경로 처리 설계는 지연을 최소화하고 오디오 왜곡을 방지한다. 예를 들어, 제1 경로 신호 처리와 제2 경로 신호 처리가 비동기식으로 수행되는 경우, 한편으로 신호는 전체 시스템을 빠르게 그리고 적은 지연으로 통과할 수 있다. 반면에 상대적으로 낮은 속도로 이득을 추정할 수 있으므로, 추정된 이득이 더 높은 정확도와 평활도를 갖게 되어 오디오 왜곡을 방지하는 데 큰 도움이 된다.

[0021] 도 1을 참조하면, 예를 들어, 제1 경로 신호 처리는 소스 입력 모듈(102)을 통해 오디오 소스 입력 신호를 수신하고, 동적 음량 밸런싱 모듈(104)을 통해 현재 이득 제어 파라미터에 기초하여 수신된 오디오 소스 입력 신호에 대해 동적 밸런싱을 수행하는 것을 포함할 수 있다. 제2 경로 처리는 음성 검출 모듈(108)에서 입력 모듈(102)로부터 수신된 오디오 소스 입력 신호를 검출하고 검출 신뢰도를 계산하는 것을 포함할 수 있다. 예를 들어, 제2 경로 처리는 또한 다음을 포함한다: 이득 제어 모듈(110)이 계산된 검출 신뢰도에 기초하여 새로운 이득 제어 파라미터를 추정할 수 있다.

[0022] 이득 제어 모듈(110)에 의해 추정된 새로운 이득 제어 파라미터는 동적 음량 밸런싱 모듈(104)에 의해 현재 사용되는 이득 제어 파라미터를 업데이트하는데 사용될 수 있다. 따라서, 동적 음량 밸런싱 모듈(104)은 업데이트된 이득 제어 파라미터에 기초하여 제1 경로 신호 처리를 수행할 수 있다. 즉, 동적 음량 밸런싱 모듈(104)은 업데이트된 이득 제어 파라미터에 기초하여 수신된 오디오 소스 입력 신호에 대해 동적 음량 밸런싱을 수행할 수 있다. 동적 음량 밸런싱 후의 오디오 신호는 신호 출력 모듈(106)을 통해 출력될 수 있다.

[0023] 오디오 소스 입력은 멀티-채널 소스 입력, 이중-채널 소스 입력 및 단일-채널 소스 입력을 포함할 수 있다. 서로 다른 소스 입력의 처리 절차는 첨부된 도면을 참조하여 각각 후술된다. 도 2는 오디오 입력 소스가 멀티-채널 소스 입력을 포함하는 본 개시내용의 하나 이상의 실시예들에 따른 음성 검출의 개략적인 블록도를 예시적으로 도시한다. 도 2에 도시된 음성 검출 과정은 예를 들어, 도 1의 음성 검출 모듈(108)에 의해, 수행될 수 있다. 도 2에 도시된 바와 같이, 먼저 중앙 채널 추출을 수행한다. 즉, 멀티-채널 소스 입력으로부터 중앙 채널 신호를 추출한다. 일반적으로 대부분의 음성 신호는 중앙 채널에 존재한다. 그런 다음 추출된 중앙 채널 신호에 대해 정규화를 수행하여 입력 신호가 유사한 레벨로 스케일링된다. 정규화된 신호는 예를 들어 다음 방정식(1)으로 표시된다.

[0024]
$$x_{i_norm}(n) = (x_i(n) - \mu_i) / \sigma_i \quad (1)$$

[0025] 여기서,

[0026] $x_i(n)$ 은 i 번째 시간 프레임의 n 번째 샘플링 시점에서의 입력 신호를 나타내고, $x_{i_norm}(n)$ 은 i 번째 시간 프레임의 n 번째 샘플링 시점에서의 출력 신호, 즉 정규화된 신호를 나타낸다. μ_i 및 σ_i 는 i 번째 시간 프레임에 해당하는 입력 신호의 평균 및 분산이다.

[0027] 다음으로, 정규화된 신호에 대해 고속 자기상관 처리를 수행하여 자기상관 결과를 출력한다. 예를 들어, 고속 자기상관 처리는 먼저 정규화된 입력 신호에 대해 STFT(short-time Fourier Transform) 방법을 이용하여 푸리에 변환을 수행하고, 푸리에 변환된 신호에 대해 고속 자기상관을 수행할 수 있다. 예를 들어, 고속 자기상관 처리 절차는 다음 수학적 (2)-(4)와 같다.

$$X_i(z) = \text{STFT}(x_{i_norm}(n)) \quad (2)$$

$$c_i(n) = \text{iSTFT}(X_i(z) * \overline{X_i(z)}) \quad (3)$$

$$C_i = \text{norm}(c_i(n)) \quad (4)$$

[0028]

[0029] 여기서 $X_i(z)$ 는 푸리에 변환된 신호이고, $\overline{X_i(z)}$ 는 $X_i(z)$ 의 켤레를 나타내고, iSTFT는 역-단시간 푸리에 변환이고, $c_i(n)$ 은 i 번째 시간 프레임의 신호의 자기상관이다. 다음으로, $c_i(n)$ 의 노름(norm)을 계산하여 C_i 를 구한다. 예를 들어, 최종 자기상관 결과의 출력 C_i 는 유클리드 노름(Euclidean norm)에 기초하여 얻어진다. 자기 상관 결과의 출력 C_i 는 검출 신뢰도를 나타내며, 이는 중앙 채널 신호에서 음성의 가능성을 나타낼 수 있다.

[0030] 도 3은 본 개시내용의 하나 이상의 실시예에 따른 음성 검출에 기초하여 동적 이득을 추정하는 방법 및 시스템의 개략적인 블록도를 예시적으로 도시한다. 도 3에 도시된 음성 검출을 기반으로 동적 이득을 추정하는 과정은 예를 들어, 도 1의 이득 제어 모듈(110)에 의해 수행될 수 있다. 예를 들어, 도 2에 도시된 프로세스를 참조하여 음성 검출 모듈(108)을 통해 생성된 검출 신뢰도 C_i 는 이득 제어 모듈(110)에 대한 입력으로 작용한다. 이 입력에 기초하여, 음성에 대한 이득(이후 이득 제어 파라미터로도 지칭될 수 있음)은 동적 음량 밸런싱 모듈(104)에 대한 입력으로서 이득 제어 모듈(110)에서 처리 후 출력된다. 일부 예에서 이득의 동적 범위는 다음 등식(5)에 의해 계산된다.

$$G_i = D_0 * \ln(C_i + D_1) \quad (5)$$

[0031] 여기서 G_i 는 동적 제어 모듈의 출력을 나타낸다. D_0 및 D_1 은 0보다 큰 실수일 수 있는 동적 이득 변동 범위의 제어 파라미터이다. $\ln(\cdot)$ 은 자연 로그 함수이다. 일부 예들에서, G_i 는 이득 제어 모듈(110)로부터의 출력으로서 동적 음량 밸런싱 모듈(104)에 제공될 수 있다.

[0032] 일부 다른 예들에서, G_i 는 추가로 처리된 다음, 이득 제어 모듈(104)로부터의 출력으로서 기능할 수 있다. 예를 들어, G_i 는 오디오 왜곡을 줄이기 위해 평활화된다. 또한, 이득 G_{i_lim} 이 적절한 크기 범위 내에 있도록 하기 위해 소프트 리미터(soft limiter)가 사용될 수도 있다. 예를 들어, 다음 수학적 6의 탄젠트 함수가 소프트 리미터로 사용될 수 있다.

$$G_{i_lim} = \tanh(\alpha G_i + \beta) + \gamma \quad (6)$$

[0033] 여기서 α , β 및 γ 는 시스템 구성에 따라 달라지는 리미터 파라미터이고 α 는 0보다 큰 실수일 수 있고 β 및 γ 는 0이 아닌 실수일 수 있다. 이 때, G_{i_lim} 은 이득 제어 모듈(110)의 출력으로 작용할 수 있다.

[0034] 도 4는 본 개시내용의 하나 이상의 실시예에 따른 각 채널의 동적 음량 밸런싱 프로세스의 개략도를 예시적으로 도시한다. 도 4의 동적 음량 균형 처리는 동적 음량 밸런싱 모듈(104)에 의해 수행될 수 있다. 예를 들어, 멀티-채널 소스 입력을 수신한 후, 동적 음량 밸런싱 모듈(104)은 먼저 채널 추출을 수행하여 중앙 채널 신호를 추출한다. 그런 다음, 중앙 채널 신호의 음량이 향상되고, 다른 채널 신호의 음량은 이득 제어 파라미터를 기반으로 감소된다. 그런 다음 향상된 중앙 채널 신호와 축소된 다른 채널 신호를 연결 및 혼합하여 출력 신호를 생성한다. 이득 제어 파라미터는 현재 이득 제어 파라미터 또는 업데이트된 이득 파라미터일 수 있다. 예를 들어, 제1 경로 신호 처리 및 제2 경로 신호 처리가 동기적인 경우, 현재 시간 프레임(예: i 번째 시간 프레임)의 신

호의 동적 음량 밸런싱에 사용되는 이득 제어 파라미터가 계산되어 이득 제어 파라미터가 실시간으로 업데이트된다. 예를 들어, G_i 또는 $G_{i,lim}$ 가 실시간으로 업데이트된다. 제1 경로 신호 처리 및 제2 경로 신호 처리가 비동기식인 경우, 음성 검출 및 이득 추정을 포함하는 제2 경로 신호 처리의 속도가 상대적으로 낮기 때문에, 현재 시간 프레임(예를 들어, i 번째 시간 프레임)의 신호의 동적 음량 밸런싱에 사용되는 이득 제어 파라미터는 G_{i-n} 또는 $G_{i-n,lim}$ 과 같은, 이전 시간 프레임의 신호의 동적 음량 밸런싱에 사용되는 이득 제어 파라미터일 수 있다. n 은 0보다 큰 정수이며, 그 값은 시스템의 실제 처리 능력이나 엔지니어의 실제 경험에 따라 달라질 수 있다. 또한, 현재/업데이트된 이득 제어 파라미터를 기반으로 중앙 채널의 신호와 다른 채널의 신호는 각각 상이한 비율로 증가 및 감소될 수 있다. 즉, 중앙 채널 신호의 음량을 향상시키기 위한 향상 제어 파라미터와 중앙 채널 신호의 음량을 줄이기 위한 감쇠 제어 파라미터는 각각 현재/업데이트된 이득 제어 파라미터에 기초하여 더 결정될 수 있다. 예를 들어, 향상 제어 파라미터 및 감쇠 제어 파라미터는 시스템 요구 사항 또는 경험에 따라 엔지니어가 설정한 비례 계산, 함수 계산 또는 기타 계산 방법에 의해 결정될 수 있다. 결과적으로 시스템의 전체 음량은 그대로 유지되지만 각 채널의 음량은 동적으로 균형을 이룬다.

[0037] 도 5는 본 개시내용의 다른 구현의 하나 이상의 실시예에 따른 방법 및 시스템의 개략도를 도시한다. 도 5에 도시된 하나 이상의 실시예에서, 본 발명의 오디오 소스 입력 신호를 처리하는 방법 및 시스템은 소스 입력 모듈(502), 동적 음량 밸런싱 모듈(504), 신호 출력 모듈(506), 음성 검출 모듈(508) 및 이득 제어 모듈(510)의 부분들을 포함한다. 이 모듈들은 도 1의 대응하는 모듈(102-110)과 실질적으로 동일한 원리로 동작한다. 또한, 도 5에 도시된 방법 및 시스템은 크로스오버 필터링 모듈(512)을 더 포함할 수 있다. 도 5에 도시된 처리 절차와 도 1-4를 참조하여 앞서 설명한 처리 절차 간의 차이점은, 제1 신호 경로에 크로스오버 필터링이 추가된 것이다. 따라서, 입력 모듈(502)로부터 수신된 소스 입력 신호는 먼저 크로스오버 필터링 모듈(512)에 의해 처리되고, 그 다음 동적 음량 밸런싱을 위해 동적 음량 밸런싱 모듈(504)에 의해 처리된다. 사람의 음성의 주파수 범위는 기본적으로 중간 주파수 범위이므로, 크로스오버 필터를 선택하여 입력 신호를 처리하여 상이한 주파수 범위의 신호들을 구별할 수 있다. 따라서 입력 신호에서 중간 주파수 범위의 신호에만 이득 제어가 적용되고, 입력 신호의 다른 주파수 범위에 있는 신호는 변경되지 않은 상태로 유지된다. 추가된 크로스오버 필터링을 통해 소스 입력 신호에서 중간 주파수 범위의 신호에만 동적 음량 밸런싱을 수행하여, 비음성 주파수 범위의 왜곡을 최대한 방지할 수 있다. 공간을 절약하기 위해, 도 5 및 도 1에 도시된 실시예의 상이한 부분들만이 아래에서 설명된다. 다른 동일한 부분에 대해서는 도 1-4 및 관련 설명을 참조할 수 있다.

[0038] 도 6은 도 5의 구현의 하나 이상의 실시예에 따른 동적 음량 밸런싱 프로세스의 개략도를 도시한다. 도 6에 도시된 바와 같이, 크로스오버 필터링 후의 소스 입력 신호는 중주파수, 고주파수 및 저주파수 범위의 신호를 포함할 수 있다. 다음으로, 동적 음량 밸런싱은 중간 주파수 범위의 신호에 대해서만 수행된다. 동적 음량 밸런싱에는 중앙 채널 신호를 추출하기 위한 채널 추출이 포함된다. 그런 다음, 현재/업데이트된 이득 제어 파라미터를 기반으로 중앙 채널 신호의 음량이 향상되고 다른 채널 신호의 음량이 감소한다. 멀티-채널 소스 입력 신호의 저주파수 범위 및 고주파수 범위의 신호들은 동적 음량 밸런싱을 거치지 않지만 동적 음량 밸런싱 후 중간 주파수 범위 신호와 직접 연결 및 혼합되어 출력 신호를 생성한다. 따라서, 비음성 신호로 인한 왜곡을 더 잘 피할 수 있다.

[0039] 소스 입력이 중앙 채널을 갖는 멀티-채널 소스 입력인 경우에 수행되는 다수의 처리 절차가 도 1-6을 참조하여 위에서 설명되었다. 당업자는 소스 입력이 단일 채널 입력인 경우 도 1-6에 도시된 처리 절차를 또한 수행할 수 있으며, 중앙 채널 추출 프로세스가 생략될 것임을 이해할 수 있을 것이다. 즉, 위에서 설명한 두 경로의 신호 처리는 단일-채널 소스 입력에서 직접 수행된다.

[0040] 소스 입력이 이중-채널 소스 입력인 경우, 앞서 설명한 방법 및 시스템을 구현하기 전에 미리 중심 추출 프로세스를 추가하여 이중-채널 소스 입력을 기반으로 멀티-채널 소스 입력이 생성되도록 해야 한다. 도 7은 본 발명의 하나 이상의 실시예에 따라, 소스 입력이 이중-채널 소스 입력인 경우 이중-채널 소스 입력을 기반으로 멀티-채널 소스 입력을 생성하는 과정을 개략적으로 도시한 도면이다.

[0041] 도 7에 도시된 업믹싱 과정은 중심 추출 알고리즘을 채택하여, 이중-채널 소스 입력을 기반으로 멀티-채널 소스 입력을 출력할 수 있다. 중심 추출 알고리즘은 예를 들어 왼쪽 및 오른쪽 채널 입력 신호 간의 교차 상관을 계산하고, 왼쪽 및 오른쪽 채널 입력 신호를 중앙 채널 신호로 결합하는 것을 포함할 수 있으며, 그 조합 비율은 교차상관에 좌우되며, 다음 방정식 (7)을 참조한다:

[0042]
$$\text{center}(n) = \theta * \text{corr}(\text{left}(n), \text{right}(n)) * (\text{left}(n) + \text{right}(n)) \quad (7)$$

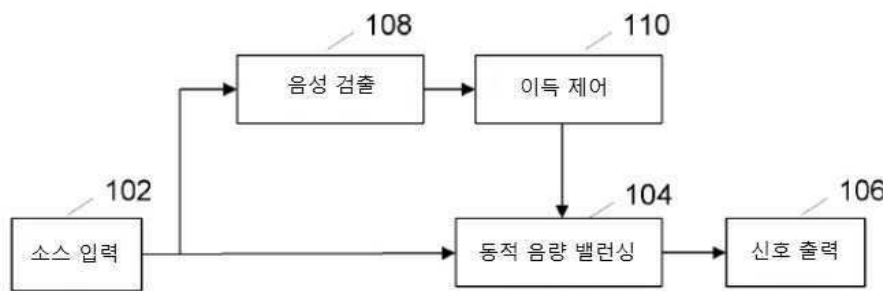
- [0043] 여기서 left(n)은 왼쪽 채널 입력 신호, right(n)은 오른쪽 채널 입력 신호, center(n)은 중앙 채널 신호, corr()은 교차 상관 함수, θ 는 실제의 튜닝 파라미터이고, θ 는 0보다 크고 1보다 작거나 같다.
- [0044] 도 8은 본 개시내용의 하나 이상의 실시예에 따른 동적 음성 향상을 위한 방법을 개략적으로 도시한다. 도 8에 도시된 바와 같이, 방법은 제1 경로 신호 처리를 수행하는 단계를 포함한다. 제1 경로 신호 처리는 오디오 소스 입력을 수신하고 제1 이득 제어 파라미터에 기초하여 오디오 소스 입력에 대해 동적 음량 밸런싱을 수행하는 것을 포함한다(S802). 방법은 또한 제2 경로 신호 처리를 수행하는 단계를 포함한다. 제2 경로 신호 처리는 오디오 소스 입력에 대해 음성 검출을 수행하고 검출 신뢰도를 계산하는 단계(S804); 및 상기 검출 신뢰도에 기초하여 제2 이득 제어 파라미터를 계산하는 단계(S806)를 포함한다. 방법은 또한 제1 이득 제어 파라미터를 제2 이득 제어 파라미터로 업데이트하는 단계(S808), 및 업데이트된 제1 이득 제어 파라미터에 기초하여 제1 경로 신호 처리를 수행하는 단계(S802)를 포함할 수 있다. 도 8에 도시된 방법은 프로세서에 의해 수행될 수 있다.
- [0045] 본 개시에서 제공하는 방법 및 시스템은 사운드바, 스테레오 스피커와 같은 소비자 제품뿐만 아니라 극장, 콘서트 홀과 같은 영화관 응용 제품에도 적용될 수 있다. 본 개시에 의해 제공되는 방법 및 시스템은 음성의 명료도를 더 향상시키고 오디오 제품 및 애플리케이션을 사용하는 사용자의 경험을 개선할 수 있다. 첨부된 도면을 참조하여 본 개시에서 설명된 상세한 방법 및 시스템 모두가 프로세서에 의해 구현될 수 있다.
- [0046] 항목 1. 동적 음성 향상 방법으로서, 오디오 소스 입력을 수신하고 제1 이득 제어 파라미터에 기초하여 오디오 소스 입력에 대해 동적 음량 밸런싱을 수행하는 것을 포함하는 제1 경로 신호 처리를 수행하는 단계; 제2 경로 신호 처리를 수행하는 단계 - 상기 제2 경로 신호 처리는: 오디오 소스 입력에 대해 음성 검출을 수행하고 검출 신뢰도를 계산하는 단계 - 검출 신뢰도는 오디오 소스 입력의 음성 가능성을 나타냄; 그리고 검출 신뢰도에 기초하여 제2 이득 제어 파라미터를 계산하는 단계 - 를 포함함; 그리고 제1 이득 제어 파라미터를 제2 이득 제어 파라미터로 업데이트하고, 업데이트된 제1 이득 제어 파라미터에 기초하여 제1 경로 신호 처리를 수행하는 단계를 포함하는, 동적 음성 향상 방법.
- [0047] 항목 2. 제1항에 있어서, 상기 오디오 소스 입력은 멀티-채널 소스 입력을 포함하고, 상기 오디오 소스 입력에 대해 음성 검출을 수행하고 검출 신뢰도를 계산하는 단계는: 멀티-채널 소스 입력으로부터 중앙 채널 신호를 추출하는 단계; 중앙 채널 신호에 대해 정규화를 수행하는 단계; 그리고 정규화된 중앙 채널 신호에 대해 고속 자기상관을 수행하는 단계를 포함하며, 고속 자기상관의 결과는 검출 신뢰도를 나타내는, 동적 음성 향상 방법.
- [0048] 항목 3. 전술한 항 중 어느 한 항에 있어서, 상기 검출 신뢰도에 기초하여 제2 이득 제어 파라미터를 계산하는 단계는: 검출 신뢰도의 대수 함수에 기초하여 제2 이득 제어 파라미터를 계산하는 단계; 계산된 제2 이득 제어 파라미터를 평활화하는 단계; 그리고 평활화된 제2 이득 제어 파라미터를 제한하는 단계를 포함하는, 동적 음성 향상 방법.
- [0049] 항목 4. 전술한 항 중 어느 한 항에 있어서, 상기 오디오 소스 입력은 멀티-채널 소스 입력을 포함하고, 상기 오디오 소스 입력에 대해 동적 음량 밸런싱을 수행하는 단계는: 멀티-채널 소스 입력으로부터 중앙 채널 신호를 추출하는 단계; 제1 이득 제어 파라미터 또는 업데이트된 제1 이득 제어 파라미터에 기초하여 중앙 채널 신호의 음량을 향상시키고 다른 채널 신호의 음량을 감소시키는 단계; 그리고 향상된 중앙 채널 신호와 감소된 다른 채널 신호를 연결 및 혼합하여 출력 신호를 생성하는 단계를 포함하는, 동적 음성 향상 방법.
- [0050] 항목 5. 전술한 항 중 어느 한 항에 있어서, 동적 음량 밸런싱을 수행하기 전에 오디오 소스 입력에 대해 크로스오버 필터링을 수행하는 단계를 더 포함하는, 동적 음성 향상 방법.
- [0051] 항목 6. 전술한 항 중 어느 한 항에 있어서, 오디오 소스 입력의 중간 주파수 범위의 신호에 대해서만 동적 음량 밸런싱을 수행하는 단계; 그리고 상기 동적 음량 밸런싱 후에 상기 오디오 소스 입력의 저주파수 대역 및 고주파수 대역 신호와 상기 오디오 소스 입력의 중간 주파수 대역 신호를 연결 및 혼합하여 출력 신호를 생성하는 단계를 더 포함하는, 동적 음성 향상 방법.
- [0052] 항목 7. 전술한 항 중 어느 한 항에 있어서, 상기 오디오 소스 입력은 이중-채널 소스 입력을 더 포함하고, 상기 방법은 이중-채널 소스 입력에 기초하여 멀티-채널 소스 입력을 생성하는 단계를 더 포함하는, 동적 음성 향상 방법.
- [0053] 항목 8. 전술한 항 중 어느 한 항에 있어서, 상기 이중-채널 소스 입력에 기초하여 멀티-채널 소스 입력을 생성하는 단계는, 이중-채널 소스 입력으로부터 좌측 채널 신호와 우측 채널 신호 사이의 교차 상관을 수행하는 단계; 그리고 조합 비율에 따라 멀티-채널 소스 입력을 생성하는 단계를 포함하며, 상기 조합 비율은 교차 상관의

결과에 따라 좌우되는, 동적 음성 향상 방법.

- [0054] 항목 9. 전술한 항 중 어느 한 항에 있어서, 상기 제1 경로 신호 처리 및 상기 제2 경로 신호 처리가 동기식 또는 비동기식인, 동적 음성 향상 방법.
- [0055] 항목 10. 동적 음성 향상 시스템으로서, 컴퓨터-실행가능 명령어를 저장하도록 구성된 메모리; 그리고 제1항 내지 제9항 중 어느 한 항에 따른 방법을 구현하도록 컴퓨터 실행 가능 명령어를 실행하도록 구성된 프로세서를 포함하는, 동적 음성 향상 시스템.
- [0056] 구현예들에 대한 설명은 예시 및 설명을 위해 제공되었다. 구현예들은 상기 설명에 따라 적절하게 수정 및 변경될 수 있고 또는 이러한 수정 및 변경이 방법을 실행함으로써 얻어질 수 있다. 예를 들어, 달리 표시되지 않는 한, 설명된 방법 중 하나 이상은 적절한 장치 및/또는 장치의 조합에 의해 수행될 수 있다. 방법은 저장된 명령어를 수행하기 위해 하나 이상의 추가 하드웨어 요소(예: 저장 장치, 메모리, 하드웨어 네트워크 인터페이스/안테나, 스위치, 액추에이터, 클록 회로 등)와 함께 하나 이상의 논리 장치(예: 프로세서)를 사용하여 수행될 수 있다. 설명된 방법 및 관련 동작은 또한 본 출원에 설명된 순서 이외의 다양한 순서로 병렬 및/또는 동시에 실행될 수 있다. 설명된 시스템은 본질적으로 예시적이며 추가 요소를 포함하거나 요소를 생략할 수 있다. 본 개시내용의 주제는 개시된 다양한 시스템 및 구성뿐만 아니라 다른 특징, 기능 및/또는 특성의 모든 신규 및 비명백한 조합을 포함한다.
- [0057] 시스템은 추가적이거나 상이한 로직을 포함할 수 있고, 많은 상이한 방식으로 구현될 수 있다. 프로세서는 마이크로프로세서, 마이크로컨트롤러, ASIC(Application Specific Integrated Circuit), 디지털 신호 프로세서 DSP, 이산 로직, 또는 이들 및/또는 다른 유형의 회로 또는 로직의 조합으로 구현될 수 있다. 유사하게, 메모리는 DRAM, SRAM, 플래시 메모리, 또는 다른 유형의 메모리일 수 있다. 파라미터(예: 조건 및 임계값) 및 기타 데이터 구조는 별도로 저장 및 관리될 수 있으며, 단일 메모리 또는 데이터베이스로 결합될 수 있으며, 또는 논리적 및 물리적으로 다양한 방식으로 구성될 수 있다. 프로그램 및 명령어 세트는 단일 프로그램의 일부이거나 별도의 프로그램이거나 복수의 메모리 및 프로세서에 분산될 수 있다.
- [0058] 본 출원에서 사용된 바와 같이, 단수 형태로 나열되고 단어 "a/one"이 앞에 오는 요소 또는 단계는 그러한 배제가 표시되지 않는 한 상기 복수의 요소 또는 단계를 배제하지 않는 것으로 이해되어야 한다. 또한, 본 개시내용의 "일 구현" 또는 "일 예"에 대한 언급은 인용된 특징을 또한 포함하는 추가 구현의 존재를 배제하는 것으로 해석되도록 의도되지 않는다. 본 발명은 특정 구현을 참조하여 위에서 설명되었다. 그러나, 당업자는 첨부된 특허청구범위에 기재된 본 발명의 보다 넓은 정신 및 범위를 벗어나지 않고 다양한 수정 및 변경이 이루어질 수 있음을 이해할 것이다.

도면

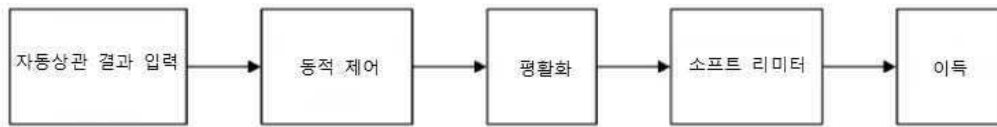
도면1



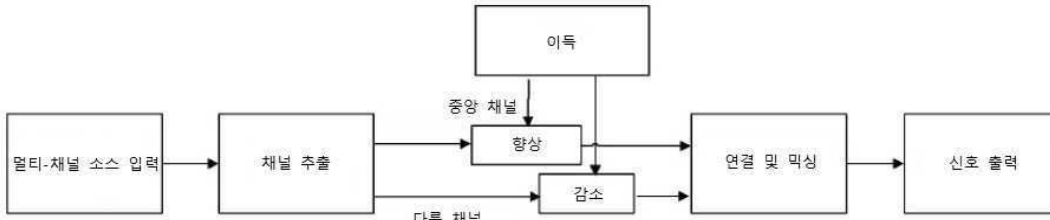
도면2



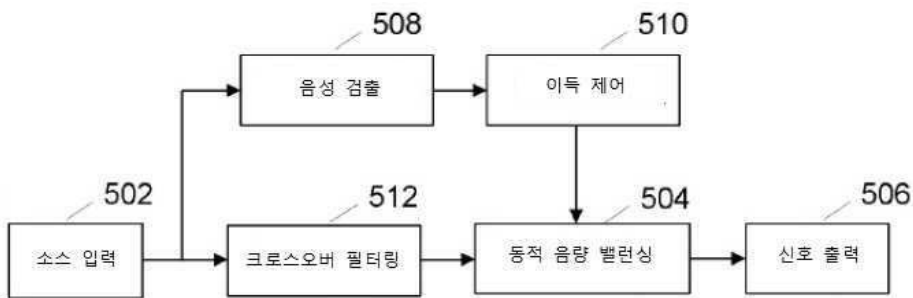
도면3



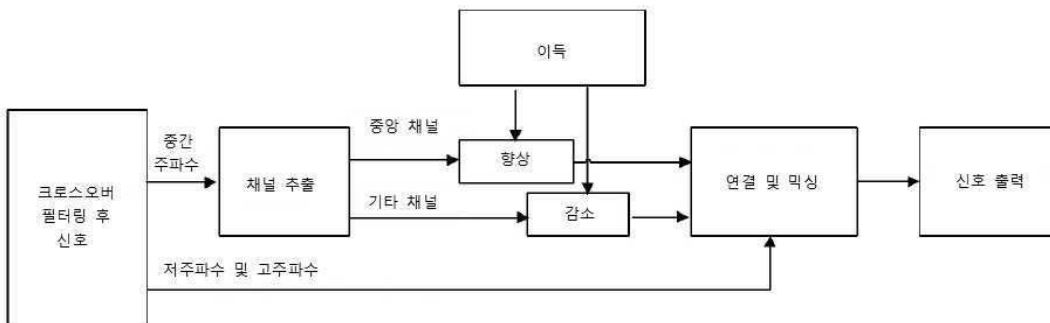
도면4



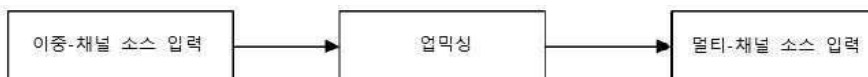
도면5



도면6



도면7



도면8

