



(19) 대한민국특허청(KR)
 (12) 등록특허공보(B1)

(45) 공고일자 2016년08월12일
 (11) 등록번호 10-1648203
 (24) 등록일자 2016년08월08일

- (51) 국제특허분류(Int. Cl.)
H04R 3/00 (2006.01) *H04M 1/60* (2006.01)
H04M 3/56 (2006.01) *H04N 7/15* (2016.01)
- (21) 출원번호 10-2011-7017092
- (22) 출원일자(국제) 2009년12월17일
 심사청구일자 2014년12월17일
- (85) 번역문제출일자 2011년07월21일
- (65) 공개번호 10-2011-0099750
- (43) 공개일자 2011년09월08일
- (86) 국제출원번호 PCT/IB2009/055803
- (87) 국제공개번호 WO 2010/073193
 국제공개일자 2010년07월01일
- (30) 우선권주장
 08172683.8 2008년12월23일
 유럽특허청(EPO)(EP)

(56) 선행기술조사문현

US20030129956 A1*

JP11242499 A

JP11308591 A

US07162045 B1

*는 심사관에 의하여 인용된 문헌

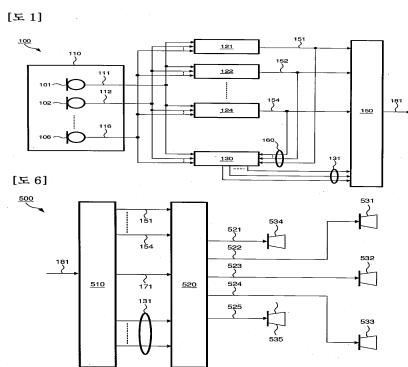
전체 청구항 수 : 총 13 항

심사관 : 송근배

(54) 발명의 명칭 스피치 캡처링 및 스피치 렌더링

(57) 요약

본 발명은 마이크로폰들에 의해 캡처된 사운드 신호들로부터 하나 이상의 주변 신호들(131) 뿐만 아니라, 추출 하나 이상의 스피치 신호들(151 ~ 154)를 제안하며 스피치 신호들 각각은 상이한 화자에 대응한다. 본 발명은 스피치 신호들만을 보내는 것과는 반대로, 렌더링 측에 하나 이상의 스피치 신호들(151 ~ 154) 및 하나 이상의 주변 신호들(131) 양쪽 모두를 송신하는 것을 제안한다. 이것은 렌더링 측에서 공간적으로 다른 방식으로 스피치 및 주변 신호들을 재생할 수 있게 한다. 주변 신호들을 재생함으로써 "함께 있다"는 느낌이 생성된다. 실시예에서, 본 발명은 주변 신호들이 있더라도 스피치 인지성이 증가되게 서로 그리고 주변 신호들과는 공간적으로 다른 2 이상의 스피치 신호들을 재생할 수 있게 한다.

대 표 도

(72) 발명자

벨트, 하름, 예., 베.

네덜란드 엔엘-5656 아에 아인드호펜 하이 테크 캠
페스 빌딩 44 내

사라우크, 바하, 애.

네덜란드 엔엘-5656 아에 아인드호펜 하이 테크 캠
페스 빌딩 44 내

트리키, 마흐디

네덜란드 엔엘-5656 아에 아인드호펜 하이 테크 캠
페스 빌딩 44 내

명세서

청구범위

청구항 1

스피치 캡처링 디바이스(speech capturing device)에 있어서:

- 상이한 공간적 위치들로부터 발생한 복수의 사운드 신호들을 캡처하기 위해 복수의 마이크로폰들을 포함하는 캡처링 회로,
- 각각이 상기 복수의 상기 사운드 신호들로부터 각각의 화자(speaker)에 대응하는 각각의 스피치 신호를 도출하기 위한 하나 이상의 추출 회로들,
- 상기 하나 이상의 추출 회로들에 의해 도출된 하나 이상의 스피치 신호들에 의해 감소된 하나 이상의 주변 신호들을 상기 복수의 사운드 신호들로부터 도출하기 위한 잔차 추출 회로,
- 상기 하나 이상의 스피치 신호들 및 상기 하나 이상의 주변 신호들을 송신하기 위한 송신 회로, 및
- (i) 하나 이상의 상기 화자들의 위치들을 결정하고 (ii) 상기 화자들의 위치들에 대한 공간 정보의 하나 이상의 출력 신호들을 하나 이상의 추출 회로들에 제공하기 위한 오디오비주얼 로케이터(audiovisual locator)를 포함하고,

각각의 추출 회로는 상기 화자들 각각의 위치에 지향되는 공간 정보의 각각의 출력 신호에 응답하여 각각의 스피치 신호를 추가로 추출하는, 스피치 캡처링 디바이스.

청구항 2

삭제

청구항 3

제 1 항에 있어서,

상기 송신 회로는 상기 화자들의 하나 이상의 위치들의 공간 정보의 출력 신호들을 송신하기 위해 추가로 구성되는, 스피치 캡처링 디바이스.

청구항 4

제 1 항에 있어서,

각각의 추출 회로는 대응하는 스피치 신호를 도출하기 위한 일반형 사이드-로브 캔슬러(generalized side-lobe canceller)를 포함하는, 스피치 캡처링 디바이스.

청구항 5

제 1 항에 있어서,

각각의 추출 회로는 대응하는 스피치 신호에서 추가 잡음 감소를 수행하기 위한 포스트-프로세서(post-processor) 회로를 추가로 포함하는, 스피치 캡처링 디바이스.

청구항 6

제 1 항에 있어서,

상기 잔차 추출 회로는 다중-채널 적응형 필터를 추가로 포함하는, 스피치 캡처링 디바이스.

청구항 7

제 6 항에 있어서,

상기 다중-채널 적응형 필터는 상기 마이크로폰들 중 하나에 의해 캡처된 사운드 신호를 기준 신호로서 수신하

도록 결합되는, 스피치 캡처링 디바이스.

청구항 8

스피치 렌더링 디바이스에 있어서:

- 하나 이상의 스피치 신호들 및 하나 이상의 주변 신호들을 수신하기 위한 수신 회로로서, 각각의 스피치 신호는 상이한 공간적 위치에서 각각이 상이한 화자에 대응하고, 상기 상이한 화자들의 위치들에 대한 공간 정보를 추가로 수신하기 위한, 상기 수신 회로, 및
 - (i) 상기 하나 이상의 스피치 신호들 및 (ii) 상기 하나 이상의 주변 신호들을 공간적으로 재생하기 위한 렌더링 회로로서, 공간 정보에 응답하여 청자에 의해 인지되는 하나 이상의 스피치 신호들로부터 각각의 방향들이
 - (a) 화자의 시각화에서 상이한 화자들의 화자의 공간 정보에 의해 표현되는 각각의 상이한 공간 정보에 정렬되고 (b) 공간적으로 재생된 하나 이상의 주변 신호들의 인지되는 방향들과 상이한 방향들을 포함하도록 하나 이상의 스피치 신호들을 공간적으로 재생하는, 상기 렌더링 회로를 포함하는, 스피치 렌더링 디바이스.

청구항 9

제 8 항에 있어서,

상기 렌더링 회로는 2 이상의 스피치 신호들을 공간적으로 재생하도록 구성되고, 상기 청자에 의해 인지되는 공간적으로 재생된 2 이상의 스피치 신호들의 각각의 방향들은 상호 상이한 방향들을 포함하는, 스피치 렌더링 디바이스.

청구항 10

삭제

청구항 11

삭제

청구항 12

제 8 항에 있어서,

상기 렌더링 회로는 상기 하나 이상의 주변 신호들의 진폭들을 감소시키기 위해 추가로 구성되는, 스피치 렌더링 디바이스.

청구항 13

스피치 신호들을 전달하기 위한 통신 시스템에 있어서:

- 제 1 항에 따른 스피치 캡처링 디바이스, 및
- 스피치 렌더링 디바이스로서,

하나 이상의 스피치 신호들 및 하나 이상의 주변 신호들을 수신하기 위한 수신회로로서, 각각의 스피치 신호는 상이한 화자에 대응하는, 상기 수신회로를 포함하고,

상기 하나 이상의 스피치 신호들 및 상기 하나 이상의 주변 신호들을 공간적으로 재생하기 위한 렌더링 회로로서, 청자에 의해 인지되는 하나 이상의 스피치 신호들로부터 각각의 방향들이 (a) 상이한 화자들의 화자의 각각의 공간적 위치에 정렬되고 (b) 공간적으로 재생된 하나 이상의 주변 신호들의 인지되는 방향들과 상이한 방향들을 포함하도록 하나 이상의 스피치 신호들을 공간적으로 재생하는, 상기 렌더링 회로를 포함하는, 상기 스피치 렌더링 디바이스를 포함하는, 스피치 신호들을 전달하기 위한 통신 시스템.

청구항 14

핸즈프리 회의 단말에 있어서,

제 1 항에 따른 상기 스피치 캡처링 디바이스 및

스피치 렌더링 디바이스로서,

하나 이상의 스피치 신호들 및 하나 이상의 주변 신호들을 수신하기 위한 수신회로로서, 각각의 스피치 신호는 상이한 화자에 대응하는, 상기 수신회로를 포함하고,

상기 하나 이상의 스피치 신호들 및 상기 하나 이상의 주변 신호들을 공간적으로 재생하기 위한 렌더링 회로로서, 청자에 의해 인지되는 하나 이상의 스피치 신호들로부터 각각의 방향들이 (a) 상이한 화자들의 화자의 각각의 공간적 위치에 정렬되고 (b) 공간적으로 재생된 하나 이상의 주변 신호들의 인지되는 방향들과 상이한 방향들을 포함하도록 하나 이상의 스피치 신호들을 공간적으로 재생하는, 상기 렌더링 회로를 포함하는, 상기 스피치 렌더링 디바이스를 포함하는, 핸즈프리 회의 단말.

청구항 15

스피치 캡처링 방법에 있어서:

- 복수의 마이크로폰을 통해 상이한 공간적 위치로부터 발생하는 복수의 사운드 신호들을 캡처하는 단계,
- 하나 이상의 추출 회로들을 통해 상기 복수의 사운드 신호들로부터 하나 이상의 각각의 화자들에 대응하는 하나 이상의 스피치 신호들을 도출하는 단계,
- 잔차 추출 회로를 통해 각각이 상기 하나 이상의 스피치 신호들에 의해 감소된 상기 복수의 사운드 신호들로부터 하나 이상의 주변 신호들을 도출하는 단계, 및
- 송신 회로를 통해 상기 하나 이상의 스피치 신호들 및 상기 하나 이상의 주변 신호들을 송신하는 단계,
- 오디오비주얼 로케이터를 통해 하나 이상의 화자들의 위치들을 결정하는 단계, 및
- 오디오비주얼 로케이터를 통해 화자들의 위치들에 대한 공간 정보의 하나 이상의 출력 신호들을 하나 이상의 추출 회로에 제공하는 단계를 포함하고,

상기 스피치 신호들을 도출하는 단계는 상기 화자들 각각의 위치에 지향되는 공간 정보의 각각의 출력 신호에 응답하여 도출하는 단계를 포함하는, 스피치 캡처링 방법.

청구항 16

스피치 렌더링 방법에 있어서:

- 수신 회로를 통해 상이한 공간적 위치에서 각각이 상이한 화자에 대응하는 하나 이상의 스피치 신호들, 및 하나 이상의 주변 신호들을 수신하는 단계, 및
- 렌더링 회로를 통해 상기 하나 이상의 스피치 신호들 및 상기 하나 이상의 주변 신호들을 공간적으로 재생하는 단계로서, 청자에 의해 인지되는 하나 이상의 스피치 신호들로부터 각각의 방향들이 (a) 상이한 화자들의 화자의 각각의 공간적 위치에 정렬되고 (b) 공간적으로 재생된 하나 이상의 주변 신호들의 인지되는 방향들과 상이한 방향들을 포함하도록 하나 이상의 스피치 신호들을 공간적으로 재생하는, 상기 재생하는 단계를 포함하는, 스피치 렌더링 방법.

발명의 설명

기술 분야

[0001]

본 발명은 예를 들면, 핸즈프리 오디오 또는 오디오비주얼 회의 단말에서 이용하기 위한 스피치 캡처링 방법 및 디바이스에 관한 것이다. 또한, 본 발명은 예를 들면, 핸즈프리 오디오 또는 오디오비주얼 회의 단말에서 이용하기 위한 스피치 렌더링 방법 및 디바이스에 관한 것이다. 또한, 본 발명은 통신 시스템, 핸즈프리 오디오 또는 오디오비주얼 회의 단말에 관한 것이다.

배경 기술

[0002]

최근에 (핸즈프리) 1 대 1 통신 시스템들에서, 스피치(speech)의 송신은 특히 대역폭 제한에 기인하여 모노포닉(monophonic) 송신 채널에 의해 실현된다. 결국, 재생측에서, 모든 사운드들은 동일 방향(또는 복수의 스피커들이 이용된다면 다수 방향들)에서 오며 따라서 양이청취(binaural hearing)에 기초하여 사운드 소스들을 분리해 내는 사람의 능력은 이용될 수 없다. 결국 잡음 및/또는 서로 말하는 화자들로 혼합된 스피치를 경청하는 것

은 어려우며 감소된 스피치 인지성 및 청취자의 피로에 이르게 한다. 이러한 이유로 핸즈프리 전화 시스템들에 있어 송신되는 원하는 스피치 신호는 가능한 한 "명료"한데, 즉, 원하는 직접적인 스피치만을 포함한다. 핸즈프리 통신에서는 변동이 없는 잡음 억제를 반드시 갖추어야만 한다. 스피치를 더욱 향상시키기 위해서 추가의 처리와 더불어 마이크로폰 어레이 빔-포밍(beam-forming)이 이용될 수 있다. 그러나, 알려진 시스템들은 통신 동안에서 대면한 느낌을 제공하지 못하며, 스피치(메시지) 뿐만 아니라, 함께 있다는 느낌도 중요한 일상적 상황들에서 특히 그러하다.

발명의 내용

해결하려는 과제

[0003] 본 발명의 목적은 함께 있다는 느낌이 있는 실생활과 같은 통신을 제공하는 것이다. 본 발명은 독립 청구항들에 의해 규정된다. 종속 청구항들은 이점이 있는 실시예들을 정의한다.

과제의 해결 수단

[0004] 본 발명에 따른 스피치 캡처링 디바이스는:

- 복수의 사운드 신호들을 캡처하기 위해 복수의 마이크로폰들을 포함하는 캡처링 회로,

[0006] - 각각이 복수의 사운드 신호들로부터 각각의 화자에 대응하는 각각의 스피치 신호를 도출하기 위한 하나 이상의 추출 회로들,

[0007] - 각각이 하나 이상의 추출 회로들에 의해 도출된 하나 이상의 스피치 신호들에 의해 감소된 하나 이상의 주변 신호들을 복수의 사운드 신호들로부터 도출하기 위한 잔차 추출 회로, 및

[0008] 하나 이상의 스피치 신호들 및 하나 이상의 주변 신호들을 송신하기 위한 송신 회로를 포함한다.

[0009] 이에 따라 하나 이상의 스피치 신호들은 캡처링 회로에 의해 캡처된 사운드 신호들로부터 추출될 수 있다. 추출 회로는 화자에 대응하는 스피치 신호를 제공한다. 이 스피치 신호는 원하는 화자 전방에 하나 이상의 클로즈 토크 마이크로폰들에 의해 제공되는 신호로서 간주될 수 있다. 잔차 추출 회로는 주변 정보(즉, 원하는 화자들이 없이 사운드 신호들에 포함된 정보)를 포함하는 주변 신호를 도출한다. 이들 스피치 및 주변 신호들을 개별적으로 송신할 때, 렌더링 측은 스피치 및 주변 신호들을 공간적으로 서로 다르게 재생할 수 있다. 주변 신호를 재생함으로써, "함께 있다는" 느낌이 생성된다. 또한, 실시예에서 본 발명은 주변 신호들이 있더라도 스피치 인지성이 증가되고 청취자들의 피로가 감소되도록 서로 그리고 주변 신호들과는 공간적으로 다른 2 이상의 스피치 신호들을 재생할 수 있게 한다.

[0010] 실시예에서, 스피치 캡처링 디바이스는 화자들의 하나 이상의 위치들을 결정하기 위한 오디오비주얼 로케이터를 추가로 포함하고, 각각의 추출 회로는 화자를 각각의 위치에 지향된다. 오디오비주얼 로케이터를 이용하는 이점은 추출 회로들의 견고성을 더욱 개선한다. 추출 회로는 한 동일 화자가 이동하면서 말하고 있지 않을지라도, 이 화자에 계속하여 집중될 수 있다. 이렇게 하여 일단 화자가 (다시) 말하기를 시작하면 추출 회로는 이미 집중되어 있고 또한 첫마디가 고품질로 캡처된다. 그렇지 않다면, 추출 회로는 첫마디 동안 다시 집중해야 하는데, 이것은 첫마디에 대해 떨어지는 품질을 초래한다.

[0011] 다른 실시예에서, 송신 회로는 화자들의 위치들을 포함하는 공간 정보를 더욱 송신한다. 화자(들)의 위치에 관한 이러한 공간 정보는 캡처하는 측에서 청각적 현장을 기술한다. 이러한 공간 정보로 렌더링 측에서 캡처측에서와 동일한 청각적 현장을 재생성하는 것이 가능하다. 이 특성은 오디오가 비디어에 대응해야 하는 오디오비주얼 회의에 있어 특히 중요하다. 예를 들면, 시각적 현장이 좌측, 중간, 및 우측의 위치에 위치한 3 화자들을 포함할 때, 위치 정보는 렌더링 측에서 이들 3 위치들에 화자들을 재생하기 위해(근사적으로) 이용될 수 있다.

[0012] 다른 실시예에서, 각각의 추출 회로는 스피치 신호를 도출하기 위한 일반형 사이드-로드 캔슬러를 포함한다. 일반형 사이드-로드 캔슬러는 적응형 빔-포밍에 특히 적합한 빔-포밍 구현이다. 이러한 캔슬러에서 빔은 원하는 화자에 집중된다.

[0013] 다른 실시예에서, 각각의 추출 회로는 스피치 신호에서 더욱 잡음을 감소시키기 위한 포스트-프로세서 회로를 추가로 포함한다. 포스트-프로세서는 잡음을 더욱 제거하므로 송신 회로에 의해 송신된 스피치 신호엔 잡음이 더 적어진다. 이것은 원하는 화자들이 더 잘 분리될 수 있고 임의의 원하는 위치에 렌더링 측에서 원하는 화자들을 위치시키는 더 많은 자유도가 생성되는 이점이 있다.

- [0014] 다른 실시예에서, 잔차 추출 회로는 각각이 하나 이상의 추출 회로들에 의해 도출된 스피치 신호들에 의해 감소되는 복수의 사운드 신호들로부터 하나 이상의 주변 신호들을 도출하기 위한 다중-채널 적응형 필터를 추가로 포함한다. 다중-채널 적응형 필터는 사운드 신호들에 내포된 스피치 신호들을 추정한다. 이들 스피치 신호들은 사운드 신호들로부터 감하여져, 이에 따라 주변 신호들을 제공한다.
- [0015] 다른 실시예에서, 다중-채널 적응형 필터는 마이크로폰들 중 하나에 의해 캡처된 사운드 신호를 기준 신호로서 수신하도록 결합된다. 이렇게 하여, 한 주변 신호가 비교적 낮은 복잡도로 생성된다.
- [0016] 본 발명의 또 다른 양태에 따라서,
- [0017] 스피치 렌더링 디바이스는:
- 각각이 상이한 화자에 대응하는 하나 이상의 스피치 신호들 및 하나 이상의 주변 신호들을 수신하기 위한 수신 회로, 및
 - 하나 이상의 스피치 신호들이 하나 이상의 주변 신호들과는 다른 방향들로부터 오는 것으로 인지되도록 하나 이상의 스피치 신호들 및 하나 이상의 주변 신호들을 공간적으로 재생하기 위한 렌더링 회로를 포함한다. 렌더링 측에서 주변 신호들을 재생함으로써 "함께 있다는" 느낌이 생성된다. 2 이상의 스피치 신호들이 서로 그리고 주변 신호들과는 공간적으로 다르게 재생되는 실시예에서 스피치 인지성이 유지되며 주변 신호들이 있더라도 증가된다.
- [0019] 실시예에서, 렌더링 회로는 스피치 신호들이 상호 서로 다른 방향들로부터 오는 것으로 인지되도록 2 이상의 스피치 신호들을 공간적으로 재생하도록 구성된다. 공간적으로 다르게 스피치 신호들을 재생함으로써, 렌더링 측에서 청취자들이 서로 다른 화자들 간을 구별하고 이들이 동시에 말할지라도 화자들 중 하나에 집중하기가 더 쉬워진다. 이것은 청취자의 피로가 덜하게 하고 스피치 인지성을 개선시킨다.
- [0020] 다른 실시예에서, 수신 회로는 또한 화자들의 위치들에 관한 공간 정보를 수신하기 위해 구성되고, 렌더링 회로는 스피치 신호들이 공간 정보에 의해 표현된 위치들로부터 오는 것으로 인지되도록 스피치 신호들을 공간적으로 재생하도록 구성된다. 이렇게 하여, 렌더링 측에서 캡처측에서와 동일한 동일한 청각적 현장이 생성될 수 있는데, 이것은 사운드 및 비디오가 일치되어야 하는 오디오비주얼 통신 시스템들에서 특히 중요하다.
- [0021] 다른 실시예에서, 스피치 신호가 인지되는 방향은 화자의 시각화에서 화자의 위치에 정렬된다. 이 실시예에서 같은 화자에 대응하는 오디오 및 비디오 대상들은 한 오디오비주얼 대상으로서 인지되는데, 이것은 청취자가 인지하기가 자연스럽고, 스피치가 비디오 대상의 스피치와는 다른 방향에서 올 경우 발생하게 되는 피로를 감소시킨다.
- [0022] 다른 실시예에서, 하나 이상의 주변 신호들의 진폭들은 감소된다. 이렇게 하여 청취자는 자신의 환경에서 렌더링을 제어하고 이를 자신의 환경에서 발생하는 다른 주변 신호들에 맞게 수정할 수 있다.
- [0023] 본 발명의 또 다른 양태에 따라서, 스피치 신호들, 및 핸즈프리 오디오 또는 오디오비주얼 회의 단말을 전달하기 위한 통신 시스템이 제공된다.
- [0024] 본 발명의 이들 및 다른 면들, 특징들 및 이점들은 이하 기술되는 실시예(들)로부터 명백해질 것이며 이에 관련하여 기술될 것이다.

도면의 간단한 설명

- [0025] [0026] 도 1은 본 발명에 따른 스피치 캡처링 디바이스의 실시예를 도시한 도면.
- 도 2는 오디오비주얼 로케이터를 포함하는 스피치 캡처링 디바이스의 실시예를 도시한 도면.
- 도 3은 포스트-프로세서 회로를 포함하는 추출 회로의 실시예를 도시한 도면.
- 도 4는 추출 회로가 포스트-프로세서 회로를 포함하는 스피치 캡처링 디바이스의 실시예를 도시한 도면.
- 도 5는 주변 신호를 도출하기 위한 다중-채널 적응형 필터를 포함하는 잔차 추출 회로의 실시예를 도시한 도면.
- 도 6은 본 발명에 따른 스피치 렌더링 디바이스의 실시예를 도시한 도면.
- 도 7은 본 발명에 따른 스피치 캡처링 디바이스 및 스피치 렌더링 디바이스를 포함하는, 스피치 신호들을 전달하기 위한 통신 시스템의 예를 도시한 도면.

발명을 실시하기 위한 구체적인 내용

[0027]

도 1은 본 발명에 따른 스피치 캡처링 디바이스(100)의 실시예를 도시한 것이다. 스피치 캡처링 디바이스(100)는 핸즈프리 오디오 또는 오디오비주얼 회의 단말에서 이용될 수 있다. 핸즈프리 오디오 또는 오디오비주얼 회의 단말은 2개의 서로 다른 위치에 있는 화자들, 예를 들면 실내에서 서로 통신하는 1대1 통신 어플리케이션을 구현하기 위해 이용된다. 그 위치에서 화자들의 수는 하나 내지는 복수의 화자들로 다양할 수 있다. 실내에서 화자들의 위치도 다양할 수 있다.

[0028]

디바이스(100)는 복수의 사운드 신호들(111, 112 ~ 116)를 캡처하기 위한 캡처링 회로(110)를 포함한다. 캡처링 회로(110)는 복수의 마이크로폰들(101, 102 ~ 106)을 포함한다. 이를 마이크로폰들은 되도록 마이크로폰들의 어레이 형태로 배열된다. 이 예에서 6개의 마이크로폰들만이 도시되었을지라도, 마이크로폰들 의 수는 이러한 개수로 제약되지 않는다.

[0029]

스피치 캡처링 디바이스(100)는 하나 이상의 추출 회로들(121, 122 ~ 124)을 추가로 포함한다. 추출 회로들 각각은 화자들 중 하나에 대응한다. 예로, 추출 회로(121)는 제 1 화자에 대응하며, 추출 회로(122)는 제 2 화자에 대응한다. 등등. 추출 회로들 각각은 다른 화자에 대응한다. 추출 회로들 각각은 예로 신호들(151, 152 또는 154) 중 하나인 화자에 대응하는 스피치 신호를 도출한다. 신호(151)는, 예로, 제 1 화자에 대응하며, 신호(152)는 제 2 화자에 대응한다. 등등. 스피치 신호들(151 ~ 154) 각각은 캡처링 회로(110)에 의해 캡처되는 복수의 사운드 신호들(101, 102 ~ 106)로부터 도출된다. 스피치 캡처링 디바이스(100)는 복수의 사운드 신호들로부터 하나 이상의 주변 신호들(131)을 도출하기 위한 잔차 추출 회로(130)를 추가로 포함한다. 주변 신호들은 사운드 신호들로부터 도출되는, 신호 번들(160)로서 도시된, 스피치 신호들의 보완을 나타낸다. 즉, 주변 신호는 하나 이상의 추출 회로들에 의해 도출되는 스피치 신호들(151, 152 ~ 154)(스피치 신호들은 사운드 신호로부터 감해진다)에 의해 감소된 사운드 신호들로부터 도출된다.

[0030]

스피치 캡처링 디바이스(100)는 하나 이상의 스피치 신호들(151, 152 ~ 154) 및 하나 이상의 주변 신호들(131)을 송신하기 위한 송신 회로(150)를 추가로 포함한다. 송신 회로(150)는 스피치 신호들 및 주변 신호들을 단일의 출력 신호(181)로 합체하며, 이는 스피치 렌더링 디바이스(500)에 송신된다. 합체는 예를 들면, 알려진 종속(concatenating) 또는 멀티플렉싱 방법들에 의해 행해질 수 있음이 당업자에게 명백하다.

[0031]

도 2는 오디오비주얼 로케이터(140)를 포함하는 스피치 캡처링 디바이스(100)의 실시예를 도시한 것이다. 오디오비주얼 로케이터(140)는 하나 이상의 화자들의 위치를 검출한다. 특정 화자의 위치에 추출 회로를 지향시키기 위해서 화자들의 위치에 관한 정보가 추출 회로들에 공급된다. 오디오비주얼 로케이터는 많은 방법들로 실현될 수 있다. 오디오비주얼 로케이터들의 알려진 예들은 예를 들면, US6850265B1 또는 N. Strobel, S. Spors, 및 R. Rabenstein, Joint audio-video object localization and tracking, IEEE Signal Processing Magazine, vol. 18, no. 1, pp. 22-31, Jan. 2001이다. 공간 정보는 예를 들면, US6850265B1에 개시된 바와 같이 각도 또는 N. Strobel, S. Spors, 및 R. Rabenstein, Joint audio-video object localization and tracking, IEEE Signal Processing Magazine, vol. 18, no. 1, pp. 22-31, Jan. 2001에 개시된 바와 같이 위치를 포함할 수도 있을 것이다.

[0032]

도 2에 도시된 스피치 캡처링 디바이스(100)는 오디오비주얼 로케이터(140)에 의해 확장된 도 1의 디바이스이다. 예에서, 제 1 화자(141)의 위치에 관한 정보는 추출 회로(121)에 공급되며, 제 2 화자(142)의 위치에 관한 정보는 추출 회로(122)에 공급된다. 두 화자들만이 있기 때문에 도 2에는 2개의 추출 회로들만이 도시되었지만, 더 많은 추출 회로들이 이용될 수 있고 추출 회로들의 수는 화자들의 수와 일치한다. 송신 회로(150)는 화자들의 위치에 관한 정보를 송신하게 수정된다. 이것은 스피치 신호들(151, 152)을 렌더링할 때 화자들(141, 142)의 위치에 관한 정보를 스피치 렌더링 디바이스(500)가 이용할 수 있게 한다.

[0033]

실시예에서, 추출 회로(121 또는 122)는 빔-포머(beam-former) 회로를 포함한다. 빔-포머 회로의 예들 중 하나는 WO2005/050618 또는 WO2005/106841에 기술된 바와 같은 일반형 사이드-로드 캔슬러(generalized side-lobe canceller)이다. 일반형 사이드-로드 캔슬러는 예를 들면, US7146012에 개시된 바와 같은 필터 합 빔포머(Filtered Sum Beamformer), 및 이에 이어 예를 들면, US7058185에 개시된 바와 같은 다중-채널 잡음 캔슬러(multi-channel noise canceller)를 포함할 수도 있을 것이다.

[0034]

도 3은 예를 들면, US6546099에 기술된 바와 같은 포스트-프로세서(post-processor) 회로(220)를 포함하는 추출 회로(121)의 실시예를 도시한 것이다. 사운드 신호들(111 ~ 116)은 출력에서 스피치 신호(151)를 제공하는 일반형 사이드-로드 캔슬러(210)에 공급된다. 이 스피치 신호(151)는 잔차 추출 회로(130)로의 입력으로서

제공된다. 스피치 신호(151)는 또한 포스트-프로세서(220)에 제공되며, 포스트-프로세서(220)는 스피치 신호(151) 내 잡음을 더욱 감소시킨다. 따라서, 포스트-프로세서(220)는 그 출력에서 스피치 신호(151)보다 더 명료한 다른 스피치 신호(161)를 생성한다. 포스트-프로세서 회로(220)는 송신 회로(150)에 의해 송신하기 위한 다른 스피치 신호(161)를 제공하며, 이것은 도 1 및 도 2에 도시된 바와 같은 스피치 신호(151)를 대신한다.

[0035] 도 4는 추출 회로들(121, 122 ~ 124)가 포스트-프로세서 회로(220)를 포함하는 스피치 캡처링 디바이스(100)의 실시예를 도시한 것이다. 추출 회로들(121 ~ 124) 각각은 2개의 출력들을 가지며, 제 1 출력에서 스피치 신호, 예를 들면, 회로(121)를 위한 스피치 신호(151)가 제공되며, 제 2 출력에서 다른 스피치 신호, 예를 들면, 회로(121)를 위한 다른 스피치 신호(161)가 제공된다. 신호들(161, 162 ~ 164)은 이제 신호들(151, 152 ~ 154) 대신 송신 회로에 제공된다.

[0036] 도 1과 도 2를 비교하면, 잔차 추출 회로(130)는 한 출력 신호(131A)만을 갖는다. 잔차 추출 회로는 2개의 스텝이지들을 갖는다. 인입되는 사운드 신호 각각(신호들(11, 112 ~ 116) 중 하나)을 위한 제 1 스테이지에서, 스피치 신호들(151, 152 ~ 154)이 감해진다. 결국, 추출 회로들에 의해 추출된 스피치 신호들에 관계된 스피치 성분들 중 어느 것도 포함하지 않는 수정된 사운드 신호가 얻어진다. 잔차 추출 회로(130)의 출력에 어떤 신호들이 제공될 것인지에 대한 다양한 선택들이 있다. 예를 들면, 수정된 사운드 신호들은 그대로 회로(130)의 출력에 제공될 수도 있을 것이다. 또 다른 예는 모든 수정된 사운드 신호들에 대한 평균 신호가 회로(130)의 출력에 제공될 수도 있을 것이라는 것이다. 또 다른 대안은 회로(130)의 출력에 제공될 수정된 사운드 신호들 중 하나를 선택하는 것이다. 또 다른 대안이 도 5에 설명된다.

[0037] 도 5는 주변 신호(131A)을 도출하기 위해, 예를 들면, US 7058185에 기술된 바와 같이, 다중-채널 적응형 필터를 포함하는 잔차 추출 회로(130)의 실시예를 도시한 것이다. 이 특정의 실시예에서, 다중-채널 적응형 필터는 간략하게 하기 위해서 2-채널이다. 그러나, 이것은 4-채널일 수도 있을 것이며, 각각의 채널은 추출 회로들(121 ~ 124) 중 하나에 대응한다.

[0038] 도 5의 실시예에서, 마이크로폰(101)에 의해 캡처된 사운드 신호(111)는 기준 신호로 선택된다. 스피치 신호들(151, 152)는 각각 2개의 적응형 필터들(310, 320)에 공급된다. 적응형 필터들(310, 320)은 각각 입력 신호들(151, 152)을 필터링하며, 적응형 필터들(310, 320)은 잔차 신호(131A)에 의해 제어된다. 잔차 신호(131A)는 사운드 신호(111)와 적응형 필터들(310, 320)의 출력 신호들 간에 차이를 산출하는 감산 회로(400)에 의해 제공된다. 적응형 필터들(310, 320)은 잔차 신호(131A)의 에너지가 최소가 되게 이들의 계수들을 조절한다. 필터의 적응을 위해서 S. Haykin, Adaptive Filter Theory, Englewood Cliffs (NJ, USA): Prentice-Hall, 1986, ISBN 0-13-004052-5 025에 기술된 바와 같은 정규화된 최소평균제곱(NLMS) 필터가 이용될 수 있다. 대안은 US 7058185에 기술된 바와 같이 주파수 영역 적응형 필터들을 이용하는 것이다. 잔차 신호(131A)는 또한 회로(130)의 출력 신호이다.

[0039] 도 6은 본 발명에 따른 스피치 렌더링 디바이스(500)의 실시예를 도시한 것이다. 스피치 렌더링 디바이스(500)는 각각의 스피치 신호는 다른 화자에 대응하는 것으로서:

[0040] - 하나 이상의 스피치 신호들(151 ~ 154) 및 하나 이상의 주변 신호들(신호 번들)(131)을 수신하기 위한 수신 회로(510), 및

[0041] - 하나 이상의 스피치 신호들이 하나 이상의 주변 신호들과는 다른 방향들로부터 오는 것으로 인지되며, 하나 이상의 스피치 신호들(151 ~ 154) 및 하나 이상의 주변 신호들(신호 번들)(131)을 공간적으로 재생하기 위한 렌더링 회로(520)를 포함한다.

[0042] 수신 회로(510)는 입력 신호(181)로부터 스피치 신호들(151 ~ 154) 및 주변 신호들(131)을 추출한다. 수신 회로(510)은 송신 회로(150)에서 실행되는 동작을 반대로 하는 알려진 동작을 기능적으로 실행한다. 수신 회로(510)는 입력 신호(181)로부터 스피치 신호들 및 주변 신호들을 추출한다. 추출은 예를 들면, 알려진 비-종속(de-concatenating) 또는 디멀티플렉싱 방법들에 의해 행해질 수 있음이 당업자에게 명백하다.

[0043] 선택적으로 화자들의 위치들에 관한 공간 정보(171)도 입력 신호(181)에 있을 때, 이것은 수신 회로(510)에 의해 추출되어 렌더링 회로(520)에 제공된다.

[0044] 렌더링 회로(520)는 스피치 신호들 및 주변 신호들을 재생 회로(530)에 매핑한다. 재생 회로는 스피커들로 구성될 수도 있을 것이다. 예에서, 재생 회로(530)는 5.1 구성으로 5개의 스피커들(531 ~ 535)을 포함한다. 스피커(532)는 센터 채널 신호를 재생하며, 스피커(533)는 우측 전방 채널 신호를 재생하며, 스피커(531)는 좌측 전방 채널 신호를 재생하며, 스피커(535)는 우측 후방 채널 신호를 재생하며, 스피커(534)는 좌측 후방 채널 신호를

재생한다. 두 화자들만이 있는 경우에, 제 1 화자의 스피치 신호(151)는 스피커(531)에 의해 재생될 수도 있을 것이며, 제 2 화자의 스피치 신호(152)는 스피커(532)에 의해 재생될 수도 있을 것이다. 이때 주변 신호들은 스피커들(535, 534)에 의해 재생될 수 있다. 복수의 스피치 신호들의 경우에, 이들 신호들은 전방 스피커들(531, 532, 533)에 의해 재생될 수 있어 스피치 신호들이 서로 다른 방향들로부터 오는 것으로 인지되는 환상을 줄 수도 있을 것이다. 이러한 인지 효과는 진폭 패닝(panning), 시간 지연 패닝, 및 웨이브-필드 합성 기술들과 같은 많은 방법들로 달성될 수도 있을 것이다.

[0045] 대안적으로, 스피커들 대신에, 헤드폰들이 이용될 수도 있을 것이다. 이러한 경우에, 가상 공간에서 스피치 신호들 및 주변 신호들을 재생하기 위해 HRTFs (Chapter 13 "3D Audio and Virtual Acoustical Environment Synthesis" by Jiashu Chen in the book Acoustical Signal Processing For Telecommunication by Steven L. Gay and Jakob Benesty (Editors) Kluwer Academic Publishers: 2000 ISBN 0-7923-7814-8)이 이용될 수 있다.

[0046] 다른 실시예에서, 수신 회로(510)는 화자들의 위치들에 관한 공간 정보(171)를 더욱 수신하며, 렌더링 회로(520)는 스피치 신호들(151 ~ 154)이 공간 정보(171)로 표현된 위치들로부터 오는 것으로 인지되게 스피치 신호들을 공간적으로 재생한다. 이것은 예를 들면, 진폭 패닝을 이용하여 행해진다. 진폭 패닝으로 화자가 2개의 스피커 위치들 사이에 위치된 것으로 인지되게 스피치 신호를 가상으로 위치시키는 것이 가능하다.

[0047] 다른 실시예에서, 스피치 신호가 인지되는 방향은 화자의 시각화에서 화자의 위치에 정렬된다.

[0048] 다른 실시예에서, 하나 이상의 주변 신호들의 진폭들이 감소된다.

[0049] 도 7은 본 발명에 따른 스피치 캡처링 디바이스(100) 및 스피치 렌더링 디바이스(500)를 포함하는 스피치 신호들을 전달하기 위한 통신 시스템의 예를 도시한 것이다.

[0050] 통신은 2개의 단말들(700, 800)을 포함한다. 이들 단말들 각각은 스피치 캡처링 디바이스(100) 및 스피치 렌더링 디바이스(500)을 포함한다. 이들 단말들은 2개의 서로 다른 위치들, 예를 들면, 다수 실내에 위치된 화자들은 서로 통신하고 있는 1 대 1 통신 어플리케이션을 가진다. 단말(800)에 포함된 스피치 캡처링 디바이스(100)는 단말(700)에 포함된 렌더링 디바이스(500)에 송신되는 스피치 신호들 및 주변 신호들을 포함하는 출력 신호(181A)를 생성한다. 각각, 단말(700)에 포함된 스피치 캡처링 디바이스(100)는 단말(800)에 포함된 렌더링 디바이스(500)에 송신되는 스피치 신호들 및 주변 신호들을 포함하는 출력 신호(181B)를 생성한다.

[0051] 본 발명이 일부 실시예들에 관련하여 기술되었을지라도, 여기에 개시된 특정 형태로 제한되게 한 것은 아니다. 그보다는, 본 본 발명의 범위는 동반된 청구항들에 의해서만 제한된다. 또한, 특징이 특정 실시예들에 관련하여 기술된 것으로 보일 수 있을지라도, 당업자는 기술된 실시예들의 여러 특징들이 본 발명에 따라 조합될 수 있음을 알 것이다. 청구항들에서, 포함하다라는 용어는 다른 요소들 또는 단계들의 존재를 배제하지 않는다.

[0052] 또한, 개별적으로 열거되었을지라도, 복수의 회로, 요소들 또는 방법의 단계들은 예를 들면, 단일의 유닛 또는 프로세서에 의해 구현될 수 있다. 또한, 개개의 특징들이 서로 다른 청구항들에 포함될 수 있을지라도, 이들은 아마도 이점이 있게 조합될 수 있으며, 서로 다른 청구항들에 포함은 특징들의 조합이 가능하지 않고/않거나 이점이 없다는 것을 의미하지 않는다. 또한, 청구항들의 한 카테고리에 특징의 포함은 이 카테고리로의 한정을 의미하지 않으며 그보다는 특징은 동등하게 적합할 때 다른 청구항 카테고리들에 적용될 수 있음을 나타낸다. 또한, 단수 표현들은 복수를 배제하지 않는다. 따라서, "a", "an", "제 1", "제 2" 등의 언급들은 복수를 배제하지 않는다. 청구항들에 참조부호들은 단지 명확하게 하려는 예로서만 제공되며 어떠한 식으로든 청구항들의 범위를 제한하는 것으로서 해석되지 않을 것이다. 본 발명은 몇개의 서로 구별되는 요소들을 포함하는 하드웨어의 회로에 의해서, 그리고 적합하게 프로그래밍된 컴퓨터 또는 다른 프로그램가능한 디바이스의 회로에 의해 구현될 수 있다.

부호의 설명

[0053] 100: 스피치 캡처링 디바이스 101, 102 ~ 106: 마이크로폰

110: 캡처링 회로 121, 122 ~ 124: 추출 회로

130: 잔차 추출 회로 140: 오디오비주얼 로케이터

150: 송신 회로 220: 포스트-프로세서 회로

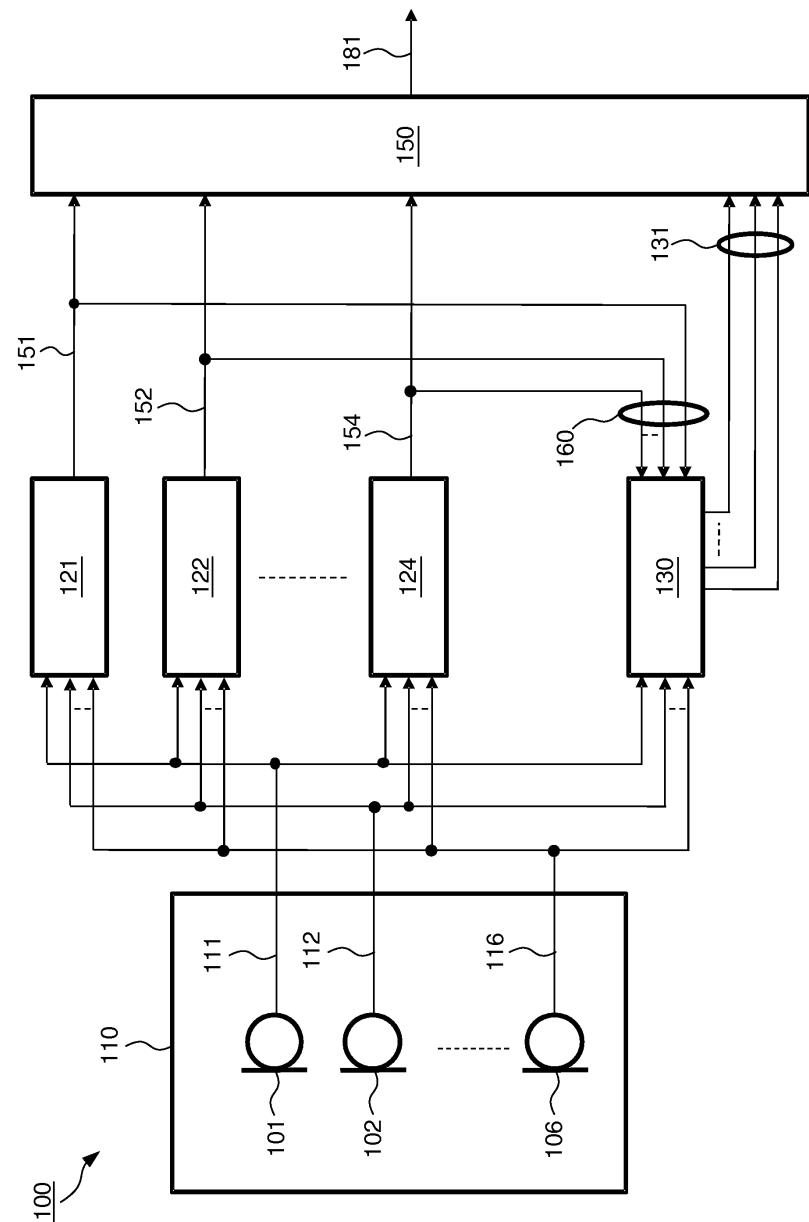
310, 320: 적응형 필터 500: 스피치 렌더링 디바이스

510: 수신 회로

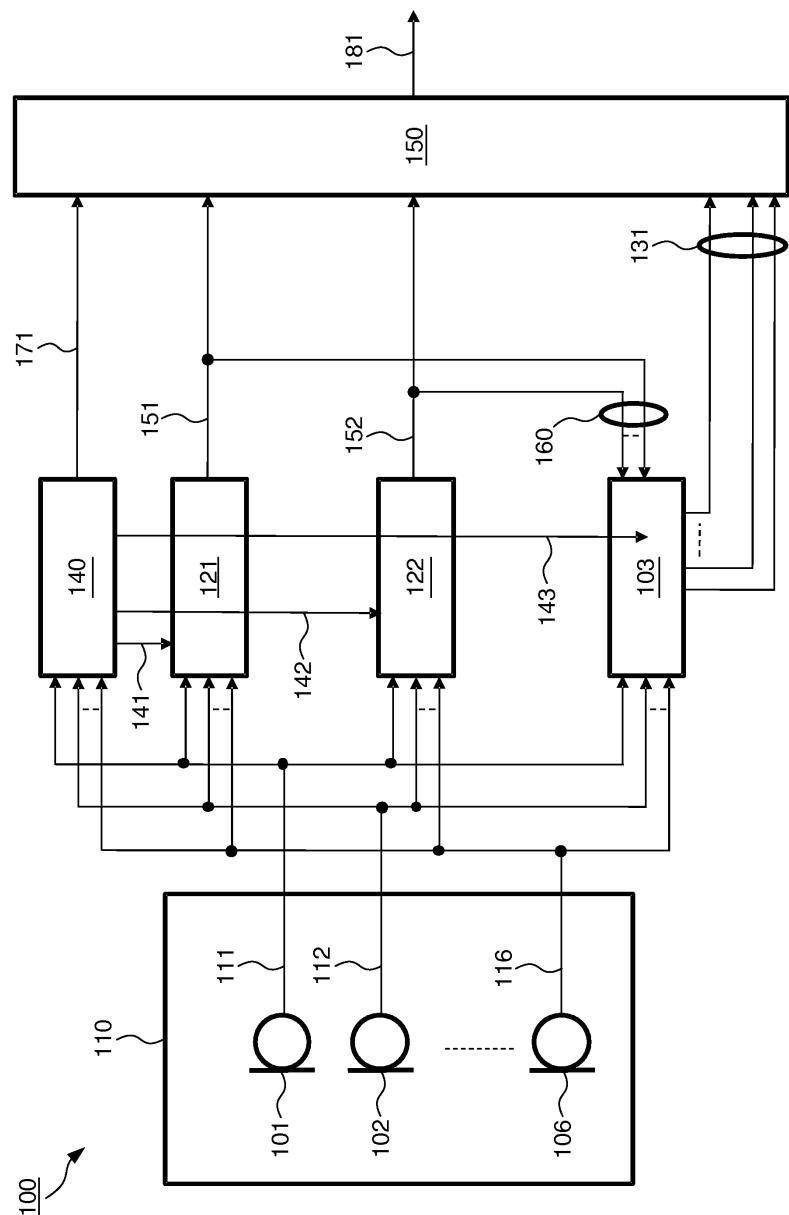
520: 렌더링 회로

530: 재생 회로

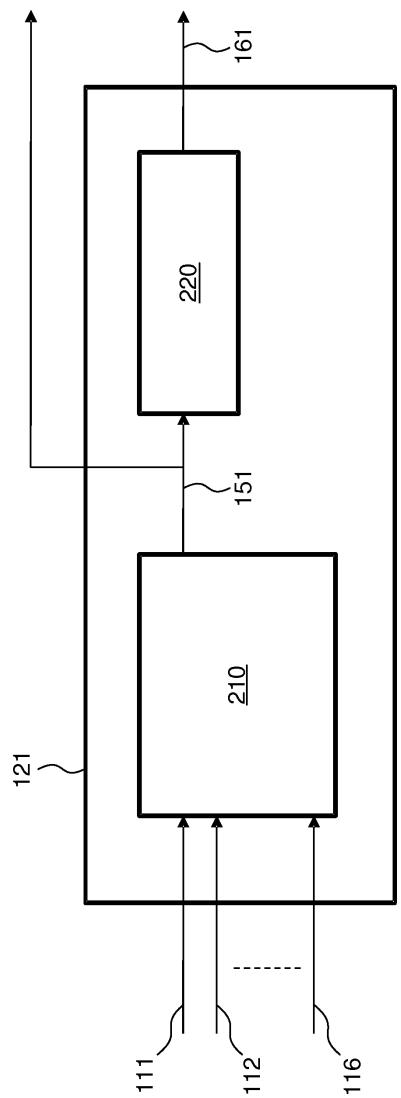
700, 800: 단말

도면**도면1**

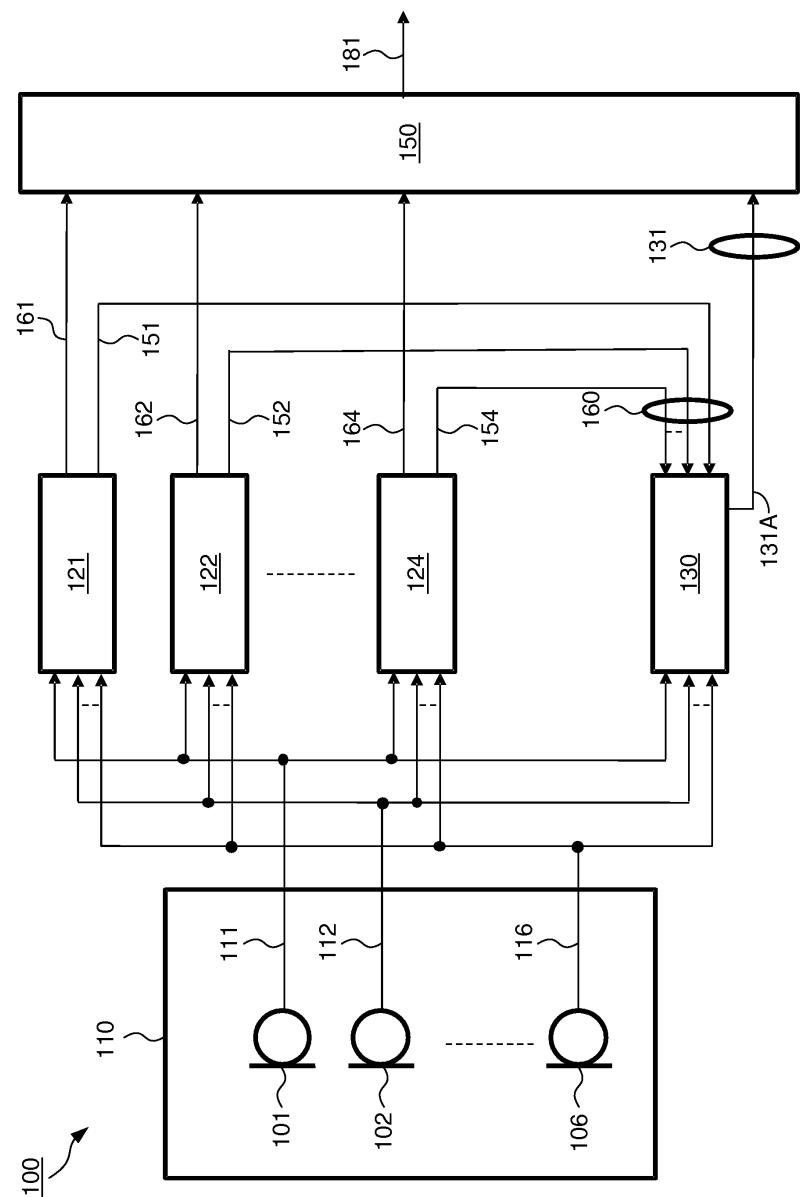
도면2



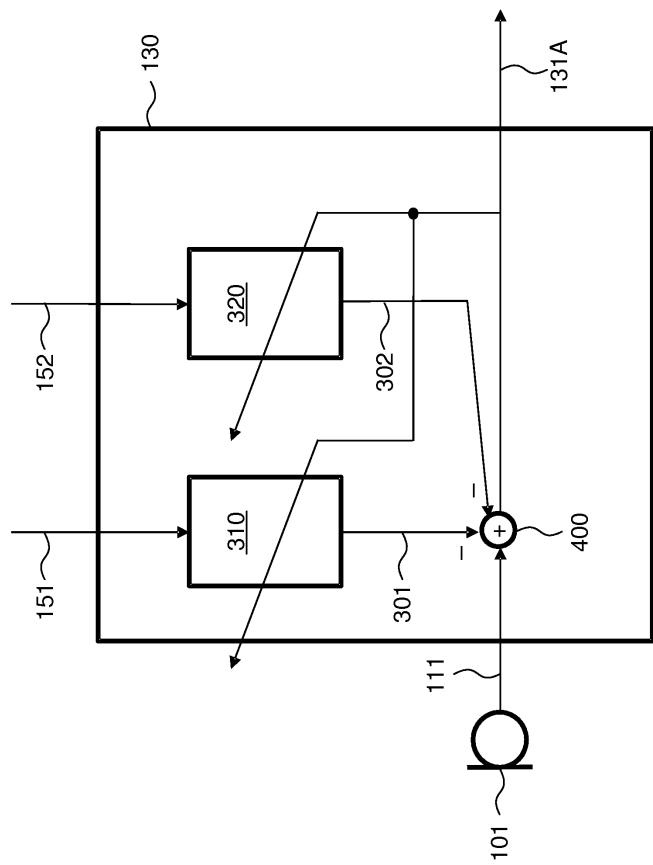
도면3



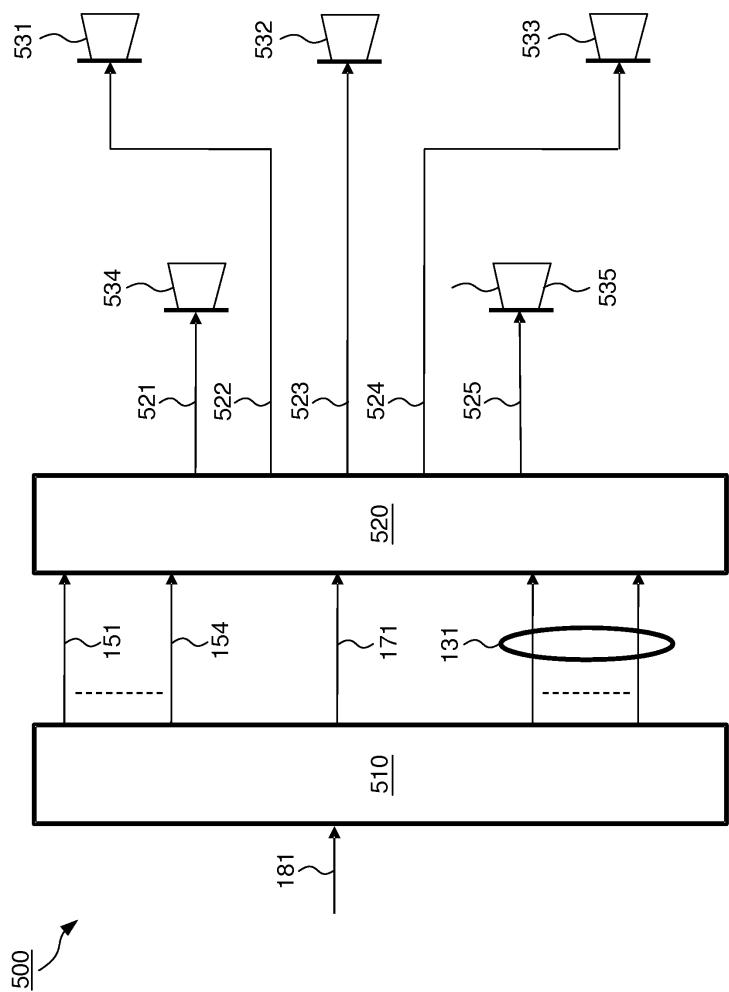
도면4



도면5



도면6



도면7

