(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0129219
(43) 공개일자 2020년11월18일

(51) 국제특허분류(Int. Cl.)
G10L 15/25 (2013.01) *G10L 21/0216* (2013.01)
G10L 25/93 (2013.01)
(52) CPC특허분류
G10L 15/25 (2013.01)
G10L 21/02 (2020.08)
(21) 출원번호 10-2019-0052951
(22) 출원일자 2019년05월07일
심사청구일자 2019년05월07일

(71) 출원인
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
박형민
서울특별시 강남구 삼성로 212 3동 1002호 (대치동, 은마아파트)
제창수
서울특별시 서대문구 통일로25길 30 103동 701호 (홍제동,한양아파트)
조병준
서울시 마포구 마포대로11다길 11, 403호
(74) 대리인
이준영, 성원찬

전체 청구항 수 : 총 15 항

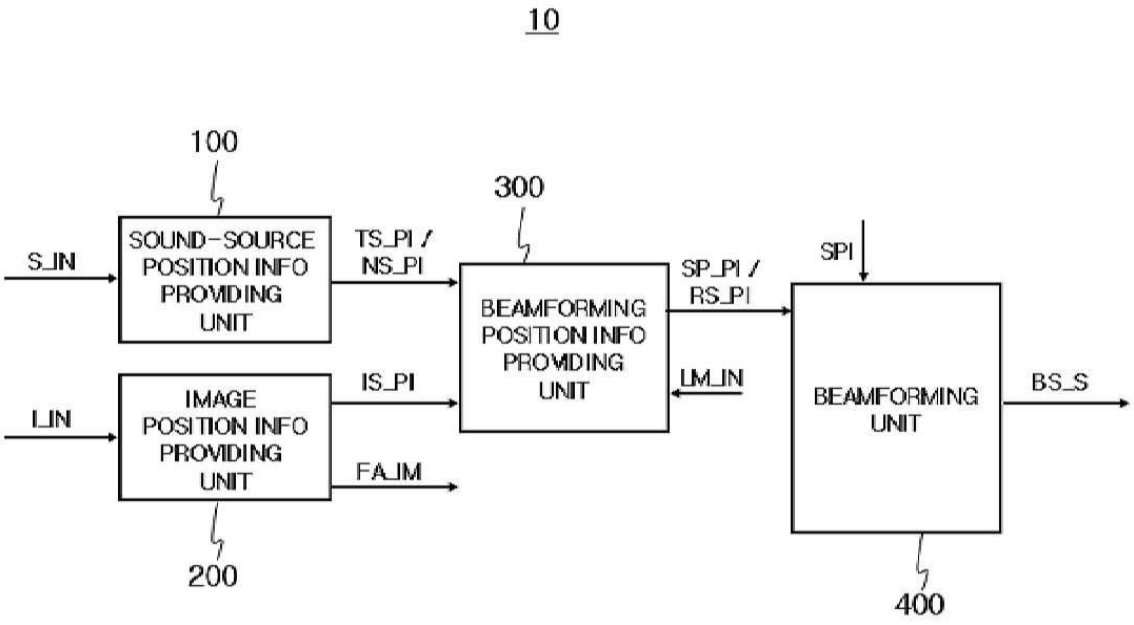
(54) 발명의 명칭 **음성인식 장치 및 음성인식 장치의 동작방법**

(57) 요 약

본 발명의 실시예에 따른 음성인식 장치는 음원 위치정보 제공부, 영상 위치정보 제공부, 빔포밍 위치정보 제공부 및 빔포밍부를 포함할 수 있다. 음원 위치정보 제공부는 소리 입력신호에 기초하여 소리 입력신호에 포함되는 타겟 음원의 위치정보 및 외부 잡음원의 위치정보를 제공할 수 있다. 영상 위치정보 제공부는 영상 입력신호에

(뒷면에 계속)

대 표 도 - 도1



기초하여 영상 입력신호에 포함되는 관심 음원의 위치정보를 제공하고, 관심 음원 중 사람의 얼굴로 판단되는 얼굴 영역 영상을 제공할 수 있다. 빔포밍 위치정보 제공부는 타겟 음원의 위치정보, 외부 잡음원의 위치정보, 관심 음원의 위치정보 및 얼굴 영역 영상으로부터 추출되는 발화자의 입술 움직임 정보에 기초하여 발화자의 위치정보 및 나머지 음원의 위치정보를 제공할 수 있다. 빔포밍부는 발화자의 위치정보 및 나머지 음원의 위치정보에 기초하여 빔포밍된 빔포밍 소리신호를 제공할 수 있다.

본 발명에 따른 음성인식 장치에서는 소리 입력신호 및 영상 입력신호에 기초하여 생성되는 발화자의 위치정보 및 나머지 음원의 위치정보에 따라 빔포밍하여 빔포밍 소리신호를 제공함으로써 음성인식의 성능을 향상시킬 수 있다.

(52) CPC특허분류
G10L 25/93 (2013.01)
G10L 2021/02166 (2013.01)

명 세 서

청구범위

청구항 1

소리 입력신호에 기초하여 상기 소리 입력신호에 포함되는 타겟 음원의 위치정보 및 외부 잡음원의 위치정보를 제공하는 음원 위치정보 제공부;

영상 입력신호에 기초하여 상기 영상 입력신호에 포함되는 관심 음원의 위치정보를 제공하고, 상기 관심 음원 중 사람의 얼굴로 판단되는 얼굴 영역 영상을 제공하는 영상 위치정보 제공부;

상기 타겟 음원의 위치정보, 상기 외부 잡음원의 위치정보, 상기 관심 음원의 위치정보 및 상기 얼굴 영역 영상으로부터 추출되는 발화자의 입술 움직임 정보에 기초하여 상기 발화자의 위치정보 및 나머지 음원의 위치정보를 제공하는 빔포밍 위치정보 제공부; 및

상기 발화자의 위치정보 및 상기 나머지 음원의 위치정보에 기초하여 빔포밍된 빔포밍 소리신호를 제공하는 빔포밍부를 포함하는 음성인식 장치.

청구항 2

제1항에 있어서,

상기 음성인식 장치는,

상기 얼굴 영역 영상으로부터 상기 발화자의 입술 움직임 정보를 추출하여 빔포밍 위치정보 제공부에 제공하고, 상기 발화자의 입술 움직임 정보 및 상기 음원 위치정보 제공부로부터 제공되는 잡음제거 신호에 기초하여 상기 발화자의 발화구간 정보를 제공하는 발화구간 검출부를 더 포함하는 것을 특징으로 하는 음성인식 장치.

청구항 3

제2항에 있어서,

상기 빔포밍부는 상기 발화자의 위치정보, 상기 나머지 음원의 위치정보 및 상기 발화구간 정보에 기초하여 상기 빔포밍 소리신호를 제공하는 것을 특징으로 하는 음성인식 장치.

청구항 4

제3항에 있어서,

상기 빔포밍부는 상기 발화구간 정보에 상응하는 시간간격동안 상기 발화자의 위치정보에 상응하는 위치에서 발생하는 소리를 증폭시켜 상기 빔포밍 소리 신호를 제공하는 것을 특징으로 하는 음성인식 장치.

청구항 5

제4항에 있어서,

상기 빔포밍부는 상기 발화구간 정보에 상응하는 시간간격동안 상기 나머지 음원의 위치정보에 상응하는 위치에서 발생하는 소리를 감소시키는 것을 특징으로 하는 음성인식 장치.

청구항 6

제3항에 있어서,

상기 외부 잡음원의 위치정보는 상기 음성인식 장치가 동작하기 전에 미리 지정한 기지정 잡음원의 위치정보를 포함하는 것을 특징으로 하는 음성인식 장치.

청구항 7

제6항에 있어서,

상기 음성인식 장치는 상기 기지정 잡음원의 위치정보에 상응하는 위치에서 발생하는 소리의 크기를 0으로 미리 설정하는 것을 특징으로 하는 음성인식 장치.

청구항 8

제2항에 있어서,

음원 위치정보 제공부는,

상기 소리 입력신호로부터 잡음을 제거하여 상기 잡음제거 신호를 제공하는 잡음제거 필터;

상기 잡음제거 신호로부터 외부 물체로부터 반사되어 발생하는 반향을 제거하여 반향제거 신호를 제공하는 반향제거기; 및

상기 반향제거 신호로부터 상기 타겟 음원의 위치정보 및 외부 잡음원의 위치정보를 검출하는 음원 위치정보 검출기를 포함하는 것을 특징으로 하는 음성인식 장치.

청구항 9

제8항에 있어서,

상기 음성인식 장치는,

상기 빔포밍 소리신호의 소리특징에 기초하여 잡음 및 반향이 제거된 개선 빔포밍 신호를 제공하는 특징 개선부를 더 포함하는 것을 특징으로 하는 음성인식 장치.

청구항 10

제2항에 있어서,

상기 발화구간 검출부는,

상기 얼굴 영역 영상으로부터 상기 발화자의 입술 움직임 정보를 추출하는 움직임 추출부; 및

상기 발화자의 입술 움직임 정보 및 상기 음원 위치정보 제공부로부터 제공되는 잡음제거 신호에 기초하여 상기 발화자의 발화구간 정보를 제공하는 발화구간 검출기를 포함하는 것을 특징으로 하는 음성인식 장치.

청구항 11

제8항에 있어서,

상기 반향 제거기는 상기 반향제거 신호를 상기 빔포밍부에 제공하는 것을 특징으로 하는 음성인식 장치.

청구항 12

음원 위치정보 제공부가 소리 입력신호에 기초하여 상기 소리 입력신호에 포함되는 타겟 음원의 위치정보 및 외부 잡음원의 위치정보를 제공하는 단계;

영상 위치정보 제공부가 영상 입력신호에 기초하여 상기 영상 입력신호에 포함되는 관심 음원의 위치정보를 제공하고, 상기 관심 음원 중 사람의 얼굴로 판단되는 얼굴 영역 영상을 제공하는 단계;

빔포밍 위치정보 제공부가 상기 타겟 음원의 위치정보, 상기 외부 잡음원의 위치정보, 상기 관심 음원의 위치정보 및 상기 얼굴 영역 영상으로부터 추출되는 발화자의 입술 움직임 정보에 기초하여 상기 발화자의 위치정보 및 나머지 음원의 위치정보를 제공하는 단계; 및

빔포밍부가 상기 발화자의 위치정보 및 상기 나머지 음원의 위치정보에 기초하여 빔포밍된 빔포밍 소리신호를 제공하는 단계를 포함하는 음성인식 장치의 동작방법.

청구항 13

제12항에 있어서,

상기 음성인식 장치의 동작방법은,

발화구간 검출부가 상기 얼굴 영역 영상으로부터 상기 발화자의 입술 움직임 정보를 추출하여 빔포밍 위치정보

제공부에 제공하고, 상기 발화자의 입술 움직임 정보 및 상기 음원 위치정보 제공부로부터 제공되는 잡음제거 신호에 기초하여 상기 발화자의 발화구간 정보를 제공하는 단계를 더 포함하는 것을 특징으로 하는 음성인식 장치의 동작방법.

청구항 14

소리 입력신호에 기초하여 상기 소리 입력신호에 포함되는 타겟 음원의 위치정보 및 외부 잡음원의 위치정보를 제공하는 음원 위치정보 제공부;

영상 입력신호에 기초하여 상기 영상 입력신호에 포함되는 관심 음원의 위치정보를 제공하고, 상기 관심 음원 중 사람의 얼굴로 판단되는 얼굴 영역 영상을 제공하는 영상 위치정보 제공부;

상기 타겟 음원의 위치정보, 상기 외부 잡음원의 위치정보, 상기 관심 음원의 위치정보 및 상기 얼굴 영역 영상으로부터 추출되는 발화자의 입술 움직임 정보에 기초하여 상기 발화자의 위치정보 및 나머지 음원의 위치정보를 제공하는 빔포밍 위치정보 제공부;

상기 발화자의 위치정보 및 상기 나머지 음원의 위치정보에 기초하여 빔포밍된 빔포밍 소리신호를 제공하는 빔포밍부; 및

상기 빔포밍 소리신호에 기초하여 음성을 인식하는 음성인식기를 포함하는 음성인식 시스템.

청구항 15

음원 위치정보 제공부가 소리 입력신호에 기초하여 상기 소리 입력신호에 포함되는 타겟 음원의 위치정보 및 외부 잡음원의 위치정보를 제공하는 단계;

영상 위치정보 제공부가 영상 입력신호에 기초하여 상기 영상 입력신호에 포함되는 관심 음원의 위치정보를 제공하고, 상기 관심 음원 중 사람의 얼굴로 판단되는 얼굴 영역 영상을 제공하는 단계;

빔포밍 위치정보 제공부가 상기 타겟 음원의 위치정보, 상기 외부 잡음원의 위치정보, 상기 관심 음원의 위치정보 및 상기 얼굴 영역 영상으로부터 추출되는 발화자의 입술 움직임 정보에 기초하여 상기 발화자의 위치정보 및 나머지 음원의 위치정보를 제공하는 단계;

빔포밍부가 상기 발화자의 위치정보 및 상기 나머지 음원의 위치정보에 기초하여 빔포밍된 빔포밍 소리신호를 제공하는 단계; 및

음성인식기가 상기 빔포밍 소리신호에 기초하여 음성을 인식하는 단계를 포함하는 음성인식 시스템의 동작방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 음성인식 장치 및 음성인식 장치의 동작방법에 관한 것이다.

배 경 기 술

[0002] 마이크를 통해서 입력되는 소리 입력신호는 음성인식에 필요한 타겟 음성뿐만 아니라 음성인식에 방해가 되는 노이즈들이 포함될 수 있다. 소리 입력신호에서 노이즈를 제거하고, 원하는 타겟 음성만을 추출하여 음성인식의 성능을 높이기 위한 다양하는 연구들이 진행되고 있다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) (한국등록특허) 제10-1133308호 (등록일자, 2012.3.28)

발명의 내용

해결하려는 과제

[0004] 본 발명이 이루고자 하는 기술적 과제는 소리 입력신호 및 영상 입력신호에 기초하여 생성되는 발화자의 위치정보 및 나머지 음원의 위치정보에 따라 빔포밍하여 빔포밍 소리신호를 제공함으로써 음성인식의 성능을 향상시킬 수 있는 음성인식 장치를 제공하는 것이다.

[0005] 본 발명이 이루고자 하는 기술적 과제는 소리 입력신호 및 영상 입력신호에 기초하여 생성되는 발화자의 위치정보 및 나머지 음원의 위치정보에 따라 빔포밍하여 빔포밍 소리신호를 제공함으로써 음성인식의 성능을 향상시킬 수 있는 음성인식 장치의 동작방법을 제공하는 것이다.

과제의 해결 수단

[0006] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 음성인식 장치는 음원 위치정보 제공부, 영상 위치정보 제공부, 빔포밍 위치정보 제공부 및 빔포밍부를 포함할 수 있다. 음원 위치정보 제공부는 소리 입력신호에 기초하여 상기 소리 입력신호에 포함되는 타겟 음원의 위치정보 및 외부 잡음원의 위치정보를 제공할 수 있다. 영상 위치정보 제공부는 영상 입력신호에 기초하여 상기 영상 입력신호에 포함되는 관심 음원의 위치정보를 제공하고, 상기 관심 음원 중 사람의 얼굴로 판단되는 얼굴 영역 영상을 제공할 수 있다. 빔포밍 위치정보 제공부는 상기 타겟 음원의 위치정보, 상기 외부 잡음원의 위치정보, 상기 관심 음원의 위치정보 및 상기 얼굴 영역 영상으로부터 추출되는 발화자의 입술 움직임 정보에 기초하여 상기 발화자의 위치정보 및 나머지 음원의 위치정보를 제공할 수 있다. 빔포밍부는 상기 발화자의 위치정보 및 상기 나머지 음원의 위치정보에 기초하여 빔포밍된 빔포밍 소리신호를 제공할 수 있다.

[0007] 일 실시예에 있어서, 상기 음성인식 장치는 발화구간 검출부를 더 포함할 수 있다. 발화구간 검출부는 상기 얼굴 영역 영상으로부터 상기 발화자의 입술 움직임 정보를 추출하여 빔포밍 위치정보 제공부에 제공할 수 있다. 발화구간 검출부는 상기 발화자의 입술 움직임 정보 및 상기 음원 위치정보 제공부로부터 제공되는 잡음제거 신호에 기초하여 상기 발화자의 발화구간 정보를 제공할 수 있다.

[0008] 일 실시예에 있어서, 상기 빔포밍부는 상기 발화자의 위치정보, 상기 나머지 음원의 위치정보 및 상기 발화구간 정보에 기초하여 상기 빔포밍 소리 신호를 제공할 수 있다.

[0009] 일 실시예에 있어서, 상기 빔포밍부는 상기 발화구간 정보에 상응하는 시간간격동안 상기 발화자의 위치정보에 상응하는 위치에서 발생하는 소리를 증폭시켜 상기 빔포밍 소리 신호를 제공할 수 있다.

[0010] 일 실시예에 있어서, 상기 빔포밍부는 상기 발화구간 정보에 상응하는 시간간격동안 상기 나머지 음원의 위치정보에 상응하는 위치에서 발생하는 소리를 감소시킬 수 있다.

[0011] 일 실시예에 있어서, 상기 외부 잡음원의 위치정보는 상기 음성인식 장치가 동작하기 전에 미리 지정한 기지정 잡음원의 위치정보를 포함할 수 있다.

[0012] 일 실시예에 있어서, 상기 음성인식 장치는 상기 기지정 잡음원의 위치정보에 상응하는 위치에서 발생하는 소리의 크기를 0으로 미리 설정할 수 있다.

[0013] 일 실시예에 있어서, 음원 위치정보 제공부는 잡음제거 필터, 반향 제거기 및 음원 위치정보 검출기를 포함할 수 있다. 잡음제거 필터는 상기 소리 입력신호로부터 장치 자체에서 발생하는 잡음을 제거하여 상기 잡음제거 신호를 제공할 수 있다. 반향 제거기는 상기 잡음제거 신호로부터 외부 물체로부터 반사되어 발생하는 반향을 제거하여 반향제거 신호를 제공할 수 있다. 음원 위치정보 검출기는 상기 반향제거 신호로부터 상기 타겟 음원의 위치정보 및 외부 잡음원의 위치정보를 검출할 수 있다.

[0014] 일 실시예에 있어서, 상기 음성인식 장치는 특징 개선부를 더 포함할 수 있다. 특징 개선부는 상기 빔포밍 소리신호의 소리특징에 기초하여 잡음 및 반향이 제거된 개선 빔포밍 신호를 제공할 수 있다.

[0015] 일 실시예에 있어서, 상기 발화구간 검출부는 움직임 추출부 및 발화구간 검출기를 포함할 수 있다. 움직임 추출부는 상기 얼굴 영역 영상으로부터 상기 발화자의 입술 움직임 정보를 추출할 수 있다. 발화구간 검출기는 상기 발화자의 입술 움직임 정보 및 상기 음원 위치정보 제공부로부터 제공되는 잡음제거 신호에 기초하여 상기 발화자의 발화구간 정보를 제공할 수 있다.

[0016] 일 실시예에 있어서, 상기 반향 제거기는 상기 반향제거 신호를 상기 빔포밍부에 제공할 수 있다.

[0017] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 음성인식 장치의 동작방법에서는, 음원 위치정보 제공

부가 소리 입력신호에 기초하여 상기 소리 입력신호에 포함되는 타겟 음원의 위치정보 및 외부 잡음원의 위치정보를 제공할 수 있다. 영상 위치정보 제공부가 영상 입력신호에 기초하여 상기 영상 입력신호에 포함되는 관심 음원의 위치정보를 제공하고, 상기 관심 음원 중 사람의 얼굴로 판단되는 얼굴 영역 영상을 제공할 수 있다. 빔포밍 위치정보 제공부가 상기 타겟 음원의 위치정보, 상기 외부 잡음원의 위치정보, 상기 관심 음원의 위치정보 및 상기 얼굴 영역 영상으로부터 추출되는 발화자의 입술 움직임 정보에 기초하여 상기 발화자의 위치정보 및 나머지 음원의 위치정보를 제공할 수 있다. 빔포밍부가 상기 발화자의 위치정보 및 상기 나머지 음원의 위치정보에 기초하여 빔포밍된 빔포밍 소리신호를 제공할 수 있다.

[0018] 일 실시예에 있어서, 발화구간 검출부가 상기 얼굴 영역 영상으로부터 상기 발화자의 입술 움직임 정보를 추출하여 빔포밍 위치정보 제공부에 제공하고, 상기 발화자의 입술 움직임 정보 및 상기 음원 위치정보 제공부로부터 제공되는 잡음제거 신호에 기초하여 상기 발화자의 발화구간 정보를 제공할 수 있다.

[0019] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 음성인식 시스템은 음원 위치정보 제공부, 영상 위치정보 제공부, 빔포밍 위치정보 제공부, 빔포밍부 및 음성인식기를 포함할 수 있다. 음원 위치정보 제공부는 소리 입력신호에 기초하여 상기 소리 입력신호에 포함되는 타겟 음원의 위치정보 및 외부 잡음원의 위치정보를 제공할 수 있다. 영상 위치정보 제공부는 영상 입력신호에 기초하여 상기 영상 입력신호에 포함되는 관심 음원의 위치정보를 제공하고, 상기 관심 음원 중 사람의 얼굴로 판단되는 얼굴 영역 영상을 제공할 수 있다. 빔포밍 위치정보 제공부는 상기 타겟 음원의 위치정보, 상기 외부 잡음원의 위치정보, 상기 관심 음원의 위치정보 및 상기 얼굴 영역 영상으로부터 추출되는 발화자의 입술 움직임 정보에 기초하여 상기 발화자의 위치정보 및 나머지 음원의 위치정보를 제공할 수 있다. 빔포밍부는 상기 발화자의 위치정보 및 상기 나머지 음원의 위치정보에 기초하여 빔포밍된 빔포밍 소리신호를 제공할 수 있다. 음성인식기는 상기 빔포밍 소리신호에 기초하여 음성을 인식할 수 있다.

[0020] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 음성인식 시스템의 동작방법에서는, 음원 위치정보 제공부가 소리 입력신호에 기초하여 상기 소리 입력신호에 포함되는 타겟 음원의 위치정보 및 외부 잡음원의 위치정보를 제공할 수 있다. 영상 위치정보 제공부가 영상 입력신호에 기초하여 상기 영상 입력신호에 포함되는 관심 음원의 위치정보를 제공하고, 상기 관심 음원 중 사람의 얼굴로 판단되는 얼굴 영역 영상을 제공할 수 있다. 빔포밍 위치정보 제공부가 상기 타겟 음원의 위치정보, 상기 외부 잡음원의 위치정보, 상기 관심 음원의 위치정보 및 상기 얼굴 영역 영상으로부터 추출되는 발화자의 입술 움직임 정보에 기초하여 상기 발화자의 위치정보 및 나머지 음원의 위치정보를 제공할 수 있다. 빔포밍부가 상기 발화자의 위치정보 및 상기 나머지 음원의 위치정보에 기초하여 빔포밍된 빔포밍 소리신호를 제공할 수 있다. 음성인식기가 상기 빔포밍 소리신호에 기초하여 음성을 인식할 수 있다.

[0021] 위에서 언급된 본 발명의 기술적 과제 외에도, 본 발명의 다른 특징 및 이점들이 이하에서 기술되거나, 그러한 기술 및 설명으로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

발명의 효과

[0022] 이상과 같은 본 발명에 따르면 다음과 같은 효과가 있다.

[0023] 본 발명에 따른 음성인식 장치에서는 소리 입력신호 및 영상 입력신호에 기초하여 생성되는 발화자의 위치정보 및 나머지 음원의 위치정보에 따라 빔포밍하여 빔포밍 소리신호를 제공함으로써 음성인식의 성능을 향상시킬 수 있다.

[0024] 본 발명에 따른 음성인식 장치의 동작방법에서는 소리 입력신호 및 영상 입력신호에 기초하여 생성되는 발화자의 위치정보 및 나머지 음원의 위치정보에 따라 빔포밍하여 빔포밍 소리신호를 제공함으로써 음성인식의 성능을 향상시킬 수 있다.

[0025] 이 밖에도, 본 발명의 실시 예들을 통해 본 발명의 또 다른 특징 및 이점들이 새롭게 파악될 수도 있을 것이다.

도면의 간단한 설명

[0026] 도 1은 본 발명의 실시예들에 따른 음성인식 장치를 나타내는 도면이다.

도 2는 도 1의 음성인식 장치에 포함되는 음원 위치정보 제공부의 동작을 설명하기 위한 도면이다.

도 3은 도 1의 음성인식 장치에 포함되는 영상 위치정보 제공부의 동작을 설명하기 위한 도면이다.

도 4는 도 1의 음성인식 장치의 일 실시예를 설명하기 위한 도면이다.

도 5는 도 4의 발화구간 검출부의 일 예를 설명하기 위한 도면이다.

도 6 및 7은 발화구간 동안의 빔포밍 동작을 설명하기 위한 도면이다.

도 8은 도 1의 음성인식 장치에 포함되는 음원 위치정보 제공부의 일 예를 나타내는 도면이다.

도 9는 도 1의 음성인식 장치의 일 실시예를 설명하기 위한 도면이다.

도 10은 본 발명의 실시예들에 따른 음성인식 장치의 동작방법을 나타내는 순서도이다.

도 11은 도 10의 음성인식 장치의 동작방법의 일 실시예를 나타내는 순서도이다.

도 12는 본 발명의 실시예들에 따른 음성인식 시스템을 나타내는 도면이다.

도 13은 본 발명의 실시예들에 따른 음성인식 시스템의 동작방법을 나타내는 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0027] 본 명세서에서 각 도면의 구성 요소들에 참조번호를 부가함에 있어서 동일한 구성 요소들에 한해서는 비록 다른 도면상에 표시되더라도 가능한한 동일한 번호를 가지도록 하고 있음에 유의하여야 한다.

[0028] 한편, 본 명세서에서 서술되는 용어의 의미는 다음과 같이 이해되어야 할 것이다.

[0029] 단수의 표현은 문맥상 명백하게 다르게 정의하지 않는 한, 복수의 표현을 포함하는 것으로 이해되어야 하는 것으로, 이들 용어들에 의해 권리범위가 한정되어서는 아니 된다.

[0030] "포함하다" 또는 "가지다" 등의 용어는 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0031] 이하, 첨부되는 도면을 참고하여 상기 문제점을 해결하기 위해 고안된 본 발명의 바람직한 실시예들에 대해 상세히 설명한다.

[0032] 도 1은 본 발명의 실시예들에 따른 음성인식 장치를 나타내는 도면이고, 도 2는 도 1의 음성인식 장치에 포함되는 음원 위치정보 제공부의 동작을 설명하기 위한 도면이고, 도 3은 도 1의 음성인식 장치에 포함되는 영상 위치정보 제공부의 동작을 설명하기 위한 도면이다.

[0033] 도 1 내지 3을 참조하면, 본 발명의 실시예에 따른 음성인식 장치(10)는 음원 위치정보 제공부(100), 영상 위치정보 제공부(200), 빔포밍 위치정보 제공부(300) 및 빔포밍부(400)를 포함할 수 있다. 음원 위치정보 제공부(100)는 소리 입력신호(S_IN)에 기초하여 소리 입력신호(S_IN)에 포함되는 타겟 음원(TS)의 위치정보(TS_PI) 및 외부 잡음원(NS)의 위치정보(NS_PI)를 제공할 수 있다. 예를 들어, 소리 입력신호(S_IN)는 타겟 음원(TS)으로부터 발생하는 소리 및 외부 잡음원으로부터 발생하는 소리 및 외부 물체로부터 반사되어 들어오는 반향소리를 포함할 수 있다. 외부 잡음원은 제1 외부 잡음원(NS1) 및 제2 외부 잡음원(NS2)을 포함할 수 있다. 음원 위치정보 제공부(100)는 소리 입력신호(S_IN)를 분석하여 타겟 음원(TS)의 위치정보(TS_PI), 제1 외부 잡음원(NS1)의 위치정보(NS1_PI) 및 제2 외부 잡음원(NS2)의 위치정보(NS2_PI)를 제공할 수 있다.

[0034] 영상 위치정보 제공부(200)는 영상 입력신호(I_IN)에 기초하여 영상 입력신호(I_IN)에 포함되는 관심 음원(ITS)의 위치정보(IS_PI)를 제공할 수 있다. 관심 음원(ITS)은 영상에서 소리가 발생할 수 있는 사람얼굴, 스피커(SPK), 스마트폰, 청소기, 동물 및 전자기기(DE)들일 수 있다. 예를 들어, 관심 음원(ITS)은 관심 타겟 음원(ITS) 및 관심 노이즈 음원을 포함할 수 있다. 관심 타겟 음원(ITS)은 제1 사람(M1)의 얼굴일 수 있고, 관심 노이즈 음원은 제2 사람(M2)의 얼굴, 스피커(SPK) 및 전자기기(DE)일 수 있다. 영상 위치정보 제공부(200)는 영상 입력신호(I_IN)를 분석하여 관심 타겟 음원(ITS)에 해당하는 제1 사람(M1)의 위치정보 및 제1 관심 노이즈 음원(INS1)에 해당하는 제2 사람(M2)의 위치정보, 제2 관심 노이즈 음원(INS2)에 해당하는 스피커(SPK)의 위치정보 및 제3 관심 노이즈 음원에 해당하는 전자기기(DE)의 위치정보를 제공할 수 있다.

[0035] 영상 위치정보 제공부(200)는 관심 음원(ITS) 중 사람의 얼굴로 판단되는 얼굴 영역 영상(FA_IM)을 제공할 수 있다. 예를 들어, 관심 음원(ITS) 중 사람의 얼굴로 판단되는 관심 음원(ITS)은 제1 사람(M1)의 얼굴 및 제2 사람(M2)의 얼굴일 수 있다. 제1 사람(M1)의 얼굴 및 제2 사람(M2)의 얼굴 중 카메라를 응시하고 있는 얼굴인 제1

사람(M1)의 얼굴 영상은 영상 위치정보 제공부(200)에 제공될 수 있다. 이 밖에도 다른 얼굴의 특징을 이용해서 영상 위치정보 제공부(200)는 얼굴영역 영상을 제공할 수도 있다.

[0036] 빔포밍 위치정보 제공부(300)는 타겟 음원(TS)의 위치정보(TS_PI), 외부 잡음원(NS)의 위치정보(NS_PI), 관심 음원(ITS)의 위치정보(IS_PI) 및 얼굴 영역 영상(FA_IM)으로부터 추출되는 발화자의 입술 움직임 정보(LM_IN)에 기초하여 발화자의 위치정보(SP_PI) 및 나머지 음원의 위치정보(RS_PI)를 제공할 수 있다. 예를 들어, 음원 위치정보 제공부(100)는 소리 입력신호(S_IN)를 분석하여 타겟 음원(TS)의 위치정보(TS_PI), 제1 외부 잡음원(NS1)의 위치정보(NS1_PI) 및 제2 외부 잡음원(NS2)의 위치정보(NS2_PI)를 제공할 수 있다. 또한, 영상 위치정보 제공부(200)는 영상 입력신호(I_IN)를 분석하여 관심 타겟 음원(ITS)에 해당하는 제1 사람(M1)의 위치정보 및 제1 관심 노이즈 음원(INS1)에 해당하는 제2 사람(M2)의 위치정보, 제2 관심 노이즈 음원(INS2)에 해당하는 스피커(SPK)의 위치정보 및 제3 관심 노이즈 음원에 해당하는 전자기기(DE)의 위치정보를 제공할 수 있다.

[0037] 예를 들어, 음원 위치정보 제공부(100)로부터 제공되는 타겟 음원(TS)의 위치정보(TS_PI)는 영상 위치정보 제공부(200)로부터 제공되는 관심 타겟 음원(ITS)에 해당하는 제1 사람(M1)의 위치정보와 매칭될 수 있고, 음원 위치정보 제공부(100)로부터 제공되는 제1 외부 잡음원(NS1)의 위치정보(NS1_PI)는 영상 위치정보 제공부(200)로부터 제공되는 제1 관심 노이즈 음원(INS1)에 해당하는 제2 사람(M2)의 위치정보와 매칭될 수 있다. 또한, 음원 위치정보 제공부(100)로부터 제공되는 제2 외부 잡음원(NS2)의 위치정보(NS2_PI)는 영상 위치정보 제공부(200)로부터 제공되는 제2 관심 노이즈 음원(INS2)에 해당하는 스피커(SPK)의 위치정보와 매칭될 수 있다.

[0038] 예를 들어, 발화자의 위치정보(SP_PI)는 타겟 음원(TS)의 위치정보(TS_PI) 및 제1 사람(M1)의 위치정보에 기초하여 생성될 수 있다. 나머지 음원의 위치정보(RS_PI)는 제1 나머지 음원의 위치정보(RS_PI) 및 제2 나머지 음원의 위치정보(RS_PI)를 포함할 수 있다. 제1 나머지 음원의 위치정보(RS_PI)는 제1 외부 잡음원(NS1)의 위치정보(NS1_PI) 및 제2 사람(M2)의 위치정보에 기초하여 생성될 수 있다. 제2 나머지 음원의 위치정보(RS_PI)는 제2 외부 잡음원(NS2)의 위치정보(NS2_PI) 및 스피커(SPK)의 위치정보에 기초하여 생성될 수 있다.

[0039] 또한, 전자기기(DE)가 동작하지 않는 경우, 소리 입력신호(S_IN)는 전자기기(DE)에서 발생하는 소리는 포함될 수 없다. 소리 입력신호(S_IN)가 전자기기(DE)에서 발생하는 소리를 포함하지 않을 경우, 음원 위치정보 제공부(100)는 전자기기(DE)의 위치정보를 외부 잡음원으로 제공하지 않을 수 있다. 반면에, 스피커(SPK)가 동작하지 않는 경우에도 영상 입력신호(I_IN)를 통해서 전자기기(DE)의 위치정보가 판단할 수 있다. 영상 입력신호(I_IN)를 통해서 판단되는 전자기기(DE)의 위치정보를 미리 인지하고 있는 경우, 전자기기(DE)가 작동할 경우, 발생하는 전자기기(DE)의 소리를 제거하기 위하여 음성인식 장치(10)는 빔포밍을 통하여 기지정 잡음원의 위치정보에 상응하는 위치에서 발생하는 소리의 크기를 0으로 미리 설정할 수 있다.

[0040] 빔포밍부(400)는 발화자의 위치정보(SP_PI) 및 나머지 음원의 위치정보(RS_PI)에 기초하여 빔포밍된 빔포밍 소리신호(BS_S)를 제공할 수 있다. 예를 들어, 빔포밍부(400)는 발화자의 위치정보(SP_PI), 제1 나머지 음원의 위치정보(RS_PI) 및 제2 나머지 음원의 위치정보(RS_PI)에 기초하여 빔포밍된 빔포밍 소리신호(BS_S)를 제공할 수 있다. 빔포밍은 발화자의 위치정보(SP_PI)에 상응하는 위치에서 발생하는 소리를 증폭시키고, 나머지 음원의 위치정보(RS_PI)에 상응하는 위치에서 발생하는 소리를 감소시키는 동작일 수 있다. 따라서, 빔포밍부(400)가 발화자의 위치정보(SP_PI), 제1 나머지 음원의 위치정보(RS_PI) 및 제2 나머지 음원의 위치정보(RS_PI)에 기초하여 빔포밍된 빔포밍 소리신호(BS_S)를 제공하는 경우, 음성인식기는 빔포밍 소리신호(BS_S)를 수신할 수 있다. 빔포밍 소리신호(BS_S)를 이용하여 음성인식을 수행하는 경우, 음성인식의 성능은 향상될 수 있다.

[0041] 본 발명에 따른 음성인식 장치(10)에서는 소리 입력신호(S_IN) 및 영상 입력신호(I_IN)에 기초하여 생성되는 발화자의 위치정보(SP_PI) 및 나머지 음원의 위치정보(RS_PI)에 따라 빔포밍하여 빔포밍 소리신호(BS_S)를 제공함으로써 음성인식의 성능을 향상시킬 수 있다.

[0042] 도 4는 도 1의 음성인식 장치의 일 실시예를 설명하기 위한 도면이고, 도 5는 도 4의 발화구간 검출부의 일 예를 설명하기 위한 도면이고, 도 6 및 7은 발화구간 동안의 빔포밍 동작을 설명하기 위한 도면이다.

[0043] 도 1, 및 4 내지 7을 참조하면, 본 발명의 실시예에 따른 음성인식 장치(10)는 음원 위치정보 제공부(100), 영상 위치정보 제공부(200), 빔포밍 위치정보 제공부(300) 및 빔포밍부(400)를 포함할 수 있다. 음원 위치정보 제공부(100)는 소리 입력신호(S_IN)에 기초하여 소리 입력신호(S_IN)에 포함되는 타겟 음원(TS)의 위치정보(TS_PI) 및 외부 잡음원(NS)의 위치정보(NS_PI)를 제공할 수 있다. 영상 위치정보 제공부(200)는 영상 입력신호(I_IN)에 기초하여 영상 입력신호(I_IN)에 포함되는 관심 음원(ITS)의 위치정보(IS_PI)를 제공하고, 관심 음원(ITS) 중 사람의 얼굴로 판단되는 얼굴 영역 영상(FA_IM)을 제공할 수 있다. 빔포밍 위치정보 제공부(300)는 타

겟 음원(TS)의 위치정보(TS_PI), 외부 잡음원(NS)의 위치정보(NS_PI), 관심 음원(ITS)의 위치정보(IS_PI) 및 얼굴 영역 영상(FA_IM)으로부터 추출되는 발화자의 입술 움직임 정보(LM_IN)에 기초하여 발화자의 위치정보(SP_PI) 및 나머지 음원의 위치정보(RS_PI)를 제공할 수 있다. 빔포밍부(400)는 발화자의 위치정보(SP_PI) 및 나머지 음원의 위치정보(RS_PI)에 기초하여 빔포밍된 빔포밍 소리신호(BS_S)를 제공할 수 있다.

[0044] 일 실시예에 있어서, 음성인식 장치(10)는 발화구간 검출부(500)를 더 포함할 수 있다. 발화구간 검출부(500)는 얼굴 영역 영상(FA_IM)으로부터 발화자의 입술 움직임 정보(LM_IN)를 추출하여 빔포밍 위치정보 제공부(300)에 제공할 수 있다. 발화구간 검출부(500)는 움직임 추출부(510) 및 발화구간 검출기(530)를 포함할 수 있다. 움직임 추출부(510)는 얼굴 영역 영상(FA_IM)으로부터 발화자의 입술 움직임 정보(LM_IN)를 추출할 수 있다. 예를 들어 빔포밍 위치정보 제공부(300)로 제공되는 입술 움직임 정보(LM_IN)는 발화자의 위치정보(SP_PI)를 파악하기 위해서 사용될 수 있다.

[0045] 발화구간 검출부(500)에 포함되는 발화구간 검출기(530)는 발화자의 입술 움직임 정보(LM_IN) 및 음원 위치정보 제공부(100)로부터 제공되는 잡음제거 신호(NR_S)에 기초하여 발화자의 발화구간 정보(SPI)를 제공할 수 있다. 예를 들어, 발화구간 검출기(530)는 발화자의 입술 움직임 정보(LM_IN) 및 잡음제거 신호(NR_S)에 기초하여 발화자가 발화하는 시간간격을 제공할 수 있다. 발화시작시간이 제1 시간(T1)이고, 발화종료시간이 제2 시간(T2)인 경우, 발화구간(SP)은 제1 시간(T1) 및 제2 시간(T2) 사이의 시간간격일 수 있다.

[0046] 일 실시예에 있어서, 빔포밍부(400)는 발화자의 위치정보(SP_PI), 나머지 음원의 위치정보(RS_PI) 및 발화구간 정보(SPI)에 기초하여 빔포밍 소리 신호(BS_S)를 제공할 수 있다.

[0047] 일 실시예에 있어서, 빔포밍부(400)는 발화구간 정보(SPI)에 상응하는 시간간격동안 발화자의 위치정보(SP_PI)에 상응하는 위치에서 발생하는 소리를 증폭시켜 빔포밍 소리신호(BS_S)를 제공할 수 있다.

[0048] 일 실시예에 있어서, 빔포밍부(400)는 발화구간 정보(SPI)에 상응하는 시간간격동안 나머지 음원의 위치정보(RS_PI)에 상응하는 위치에서 발생하는 소리를 감소시킬 수 있다. 예를 들어, 제1 나머지 음원의 위치정보(RS_PI)에 상응하는 위치에서 발생하는 소리 및 제2 나머지 음원의 위치정보(RS_PI)에 상응하는 위치에서 발생하는 소리는 감소시킬 수 있다.

[0049] 일 실시예에 있어서, 외부 잡음원(NS)의 위치정보(NS_PI)는 음성인식 장치(10)가 동작하기 전에 미리 지정한 기지정 잡음원의 위치정보를 포함할 수 있다. 예를 들어, 음성인식 장치(10)가 동작하기 이전에, 사용자는 전자기기(DE)의 위치를 기지정 잡음원으로 미리 지정할 수 있다. 이 경우, 기지정 잡음원의 위치정보에 상응하는 위치에서 발생하는 소리의 크기를 0으로 미리 설정함으로써 간헐적으로 발생하는 잡음을 제거할 수 있다.

[0050] 도 8은 도 1의 음성인식 장치에 포함되는 음원 위치정보 제공부의 일 예를 나타내는 도면이고, 도 9는 도 1의 음성인식 장치의 일 실시예를 설명하기 위한 도면이다.

[0051] 도 8 및 9를 참조하면, 음원 위치정보 제공부(100)는 잡음제거 필터(110), 반향 제거기(130) 및 음원 위치정보 검출기(150)를 포함할 수 있다. 잡음제거 필터(110)는 소리 입력신호(S_IN)로부터 잡음을 제거하여 잡음제거 신호(NR_S)를 제공할 수 있다. 반향 제거기(130)는 잡음제거 신호(NR_S)로부터 외부 물체로부터 반사되어 발생하는 반향을 제거하여 반향제거 신호(DE_S)를 제공할 수 있다. 음원 위치정보 검출기(150)는 반향제거 신호(DE_S)로부터 타겟 음원(TS)의 위치정보(TS_PI) 및 외부 잡음원(NS)의 위치정보(NS_PI)를 검출할 수 있다. 예를 들어, 반향 제거기(130)는 반향제거 신호(DE_S)를 빔포밍부(400)에 제공할 수 있다.

[0052] 일 실시예에 있어서, 음성인식 장치(10)는 특징 개선부(600)를 더 포함할 수 있다. 특징 개선부(600)는 빔포밍 소리신호(BS_S)의 소리특징에 기초하여 잡음 및 반향이 제거된 개선 빔포밍 신호(EB_S)를 제공할 수 있다.

[0053] 본 발명에 따른 음성인식 장치(10)에서는 소리 입력신호(S_IN) 및 영상 입력신호(I_IN)에 기초하여 생성되는 발화자의 위치정보(SP_PI) 및 나머지 음원의 위치정보(RS_PI)에 따라 빔포밍하여 빔포밍 소리신호(BS_S)를 제공함으로써 음성인식의 성능을 향상시킬 수 있다.

[0054] 도 10은 본 발명의 실시예들에 따른 음성인식 장치의 동작방법을 나타내는 순서도이고, 도 11은 도 10의 음성인식 장치의 동작방법의 일 실시예를 나타내는 순서도이다.

[0055] 도 10 및 11을 참조하면, 본 발명의 실시예에 따른 음성인식 장치(10)의 동작방법에서는, 음원 위치정보 제공부(100)가 소리 입력신호(S_IN)에 기초하여 소리 입력신호(S_IN)에 포함되는 타겟 음원(TS)의 위치정보(TS_PI) 및 외부 잡음원(NS)의 위치정보(NS_PI)를 제공할 수 있다(S100). 영상 위치정보 제공부(200)가 영상 입력신호(I_IN)에 기초하여 영상 입력신호(I_IN)에 포함되는 관심 음원(ITS)의 위치정보(IS_PI)를 제공하고, 관심 음원

(ITS) 중 사람의 얼굴로 판단되는 얼굴 영역 영상(FA_IM)을 제공할 수 있다(S200). 빔포밍 위치정보 제공부(300)가 타겟 음원(TS)의 위치정보(TS_PI), 외부 잡음원(NS)의 위치정보(NS_PI), 관심 음원(ITS)의 위치정보(IS_PI) 및 얼굴 영역 영상(FA_IM)으로부터 추출되는 발화자의 입술 움직임 정보(LM_IN)에 기초하여 발화자의 위치정보(SP_PI) 및 나머지 음원의 위치정보(RS_PI)를 제공할 수 있다(S300). 빔포밍부(400)가 발화자의 위치정보(SP_PI) 및 나머지 음원의 위치정보(RS_PI)에 기초하여 빔포밍된 빔포밍 소리신호(BS_S)를 제공할 수 있다(S400).

[0056] 일 실시예에 있어서, 발화구간 검출부(500)가 얼굴 영역 영상(FA_IM)으로부터 발화자의 입술 움직임 정보(LM_IN)를 추출하여 빔포밍 위치정보 제공부(300)에 제공하고, 발화자의 입술 움직임 정보(LM_IN) 및 음원 위치정보 제공부(100)로부터 제공되는 잡음제거 신호(NR_S)에 기초하여 발화자의 발화구간 정보(SPI)를 제공할 수 있다(S430).

[0057] 도 12는 본 발명의 실시예들에 따른 음성인식 시스템을 나타내는 도면이다.

[0058] 도 12를 참조하면, 본 발명의 실시예에 따른 음성인식 시스템은 음원 위치정보 제공부(100), 영상 위치정보 제공부(200), 빔포밍 위치정보 제공부(300), 빔포밍부(400) 및 음성인식기(700)를 포함할 수 있다. 음원 위치정보 제공부(100)는 소리 입력신호(S_IN)에 기초하여 소리 입력신호(S_IN)에 포함되는 타겟 음원(TS)의 위치정보(TS_PI) 및 외부 잡음원(NS)의 위치정보(NS_PI)를 제공할 수 있다. 영상 위치정보 제공부(200)는 영상 입력신호(I_IN)에 기초하여 영상 입력신호(I_IN)에 포함되는 관심 음원(ITS)의 위치정보(IS_PI)를 제공하고, 관심 음원(ITS) 중 사람의 얼굴로 판단되는 얼굴 영역 영상(FA_IM)을 제공할 수 있다. 빔포밍 위치정보 제공부(300)는 타겟 음원(TS)의 위치정보(TS_PI), 외부 잡음원(NS)의 위치정보(NS_PI), 관심 음원(ITS)의 위치정보(IS_PI) 및 얼굴 영역 영상(FA_IM)으로부터 추출되는 발화자의 입술 움직임 정보(LM_IN)에 기초하여 발화자의 위치정보(SP_PI) 및 나머지 음원의 위치정보(RS_PI)를 제공할 수 있다. 빔포밍부(400)는 발화자의 위치정보(SP_PI) 및 나머지 음원의 위치정보(RS_PI)에 기초하여 빔포밍된 빔포밍 소리신호(BS_S)를 제공할 수 있다. 음성인식기(700)는 빔포밍 소리신호(BS_S)에 기초하여 음성을 인식할 수 있다.

[0059] 도 13은 본 발명의 실시예들에 따른 음성인식 시스템의 동작방법을 나타내는 순서도이다.

[0060] 도 13을 참조하면, 본 발명의 실시예에 따른 음성인식 시스템의 동작방법에서는, 음원 위치정보 제공부(100)가 소리 입력신호(S_IN)에 기초하여 소리 입력신호(S_IN)에 포함되는 타겟 음원(TS)의 위치정보(TS_PI) 및 외부 잡음원(NS)의 위치정보(NS_PI)를 제공할 수 있다(S100). 영상 위치정보 제공부(200)가 영상 입력신호(I_IN)에 기초하여 영상 입력신호(I_IN)에 포함되는 관심 음원(ITS)의 위치정보(IS_PI)를 제공하고, 관심 음원(ITS) 중 사람의 얼굴로 판단되는 얼굴 영역 영상(FA_IM)을 제공할 수 있다(S200). 빔포밍 위치정보 제공부(300)가 타겟 음원(TS)의 위치정보(TS_PI), 외부 잡음원(NS)의 위치정보(NS_PI), 관심 음원(ITS)의 위치정보(IS_PI) 및 얼굴 영역 영상(FA_IM)으로부터 추출되는 발화자의 입술 움직임 정보(LM_IN)에 기초하여 발화자의 위치정보(SP_PI) 및 나머지 음원의 위치정보(RS_PI)를 제공할 수 있다(S300). 빔포밍부(400)가 발화자의 위치정보(SP_PI) 및 나머지 음원의 위치정보(RS_PI)에 기초하여 빔포밍된 빔포밍 소리신호(BS_S)를 제공할 수 있다(S400). 음성인식기(700)가 빔포밍 소리신호(BS_S)에 기초하여 음성을 인식할 수 있다(S500).

[0061] 본 발명의 소리 입력신호(S_IN) 및 영상 입력신호(I_IN)에 기초하여 생성되는 빔포밍 소리신호(BS_S)는 화자 인식, 음원 구간 검출, 음원 물체 검출, 음원 분류와 같은 다양한 장치에서 사용 가능하며, 빔포밍 소리신호(BS_S)의 단독 사용뿐만 아니라 영상 입력신호(I_IN) 및 이에 기초하여 생성된 음원의 위치정보등의 결합을 통해 다양한 기능을 수행할 수 있다.

[0062] 본 발명에 따른 음성인식 장치(10)에서는 소리 입력신호(S_IN) 및 영상 입력신호(I_IN)에 기초하여 생성되는 발화자의 위치정보(SP_PI) 및 나머지 음원의 위치정보(RS_PI)에 따라 빔포밍하여 빔포밍 소리신호(BS_S)를 제공함으로써 음성인식의 성능을 향상시킬 수 있다.

[0063] 위에서 언급된 본 발명의 기술적 과제 외에도, 본 발명의 다른 특징 및 이점들이 이하에서 기술되거나, 그러한 기술 및 설명으로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

부호의 설명

[0064] 10: 음성인식 장치 100: 음원 위치정보 제공부

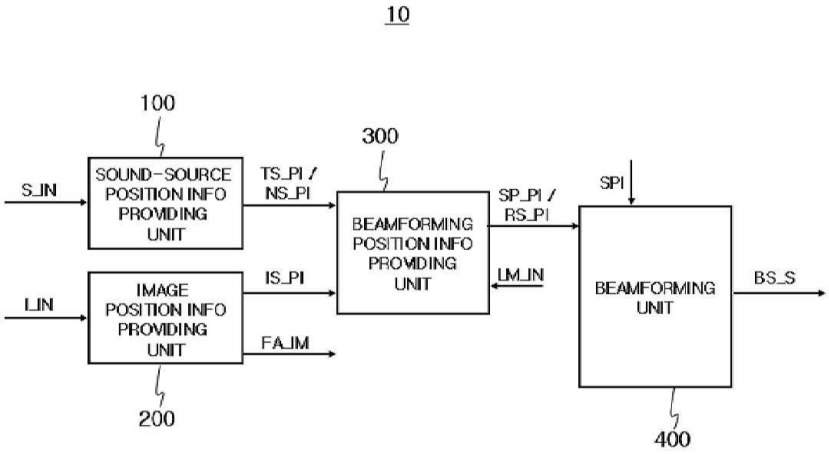
200: 영상 위치정보 제공부 300: 빔포밍 위치정보 제공부

400: 빔포밍부 500: 발화구간 검출부

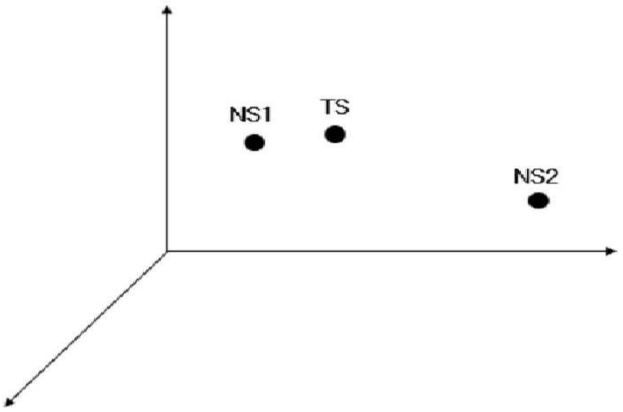
600: 특징 개선부

도면

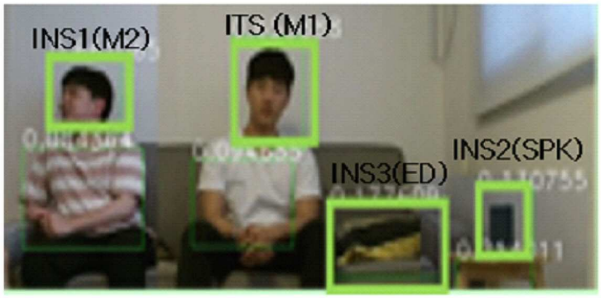
도면1



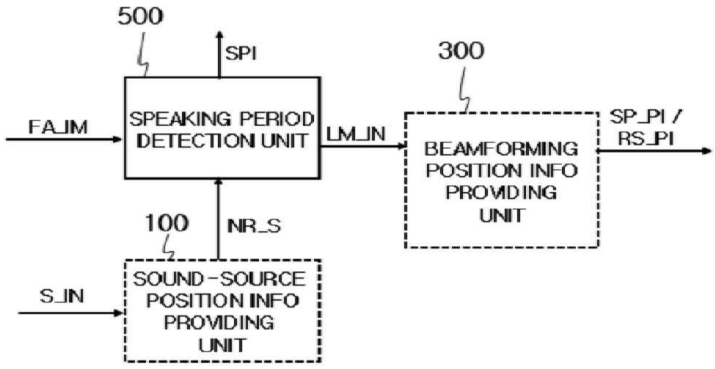
도면2



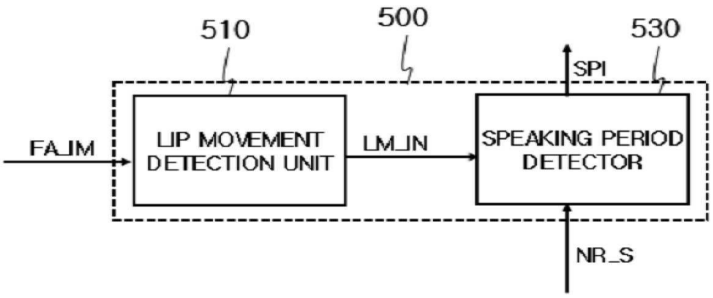
도면3



도면4



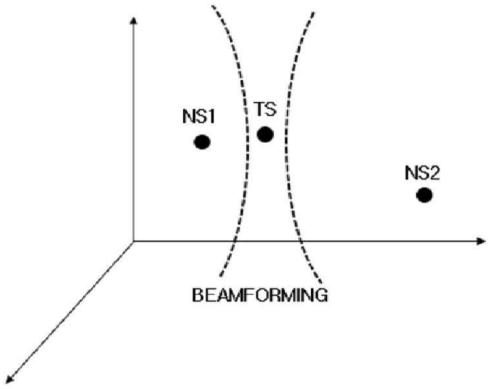
도면5



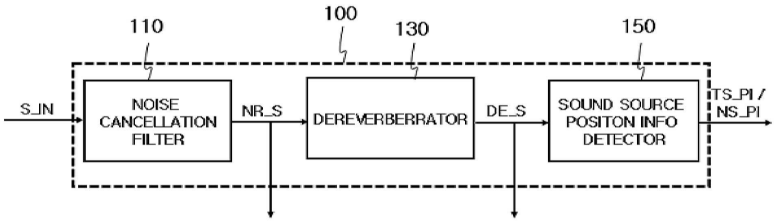
도면6



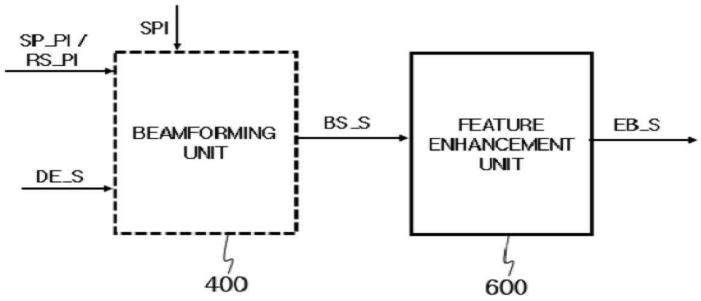
도면7



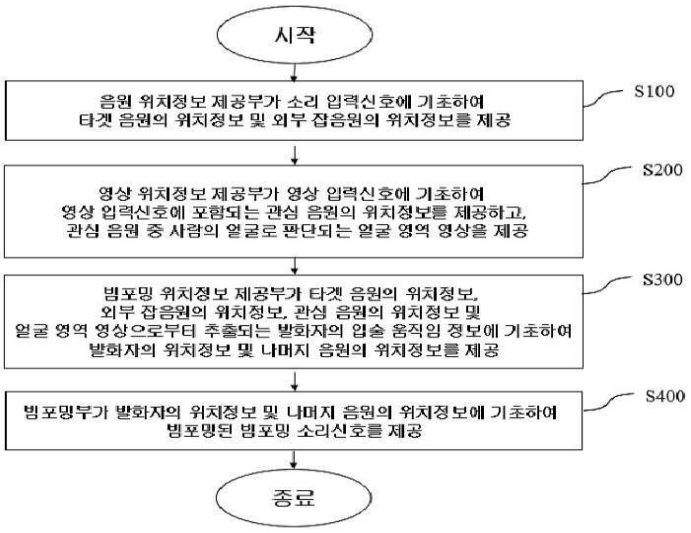
도면8



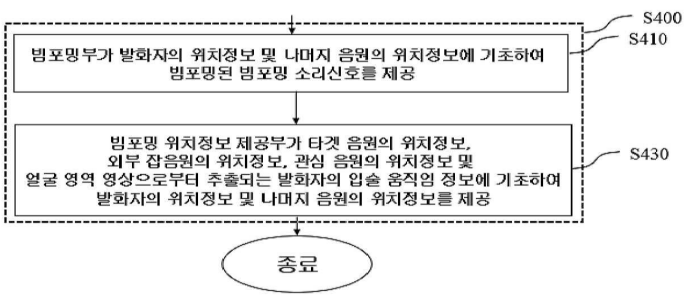
도면9



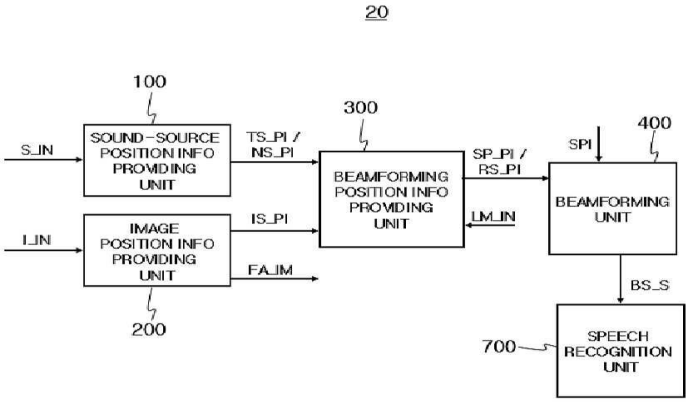
도면10



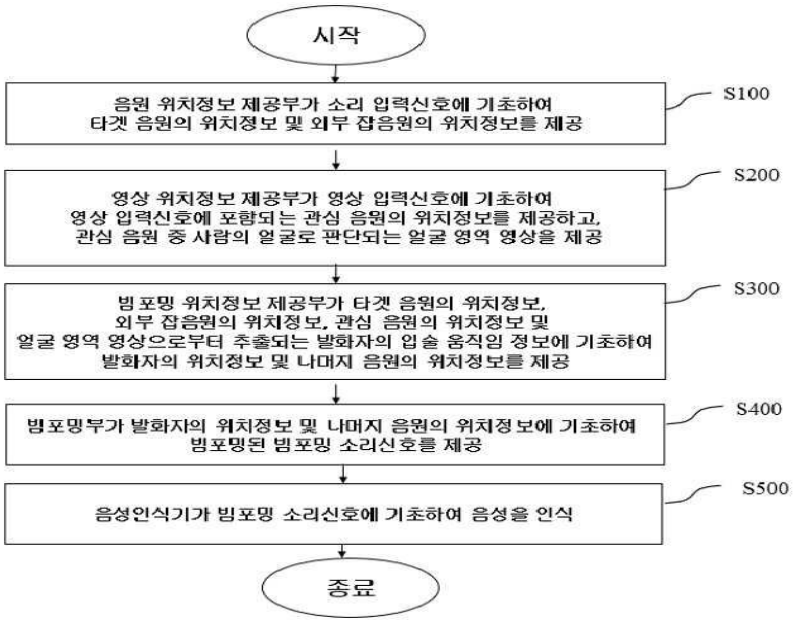
도면11



도면12


20
100
S_IN
SOUND-SOURCE POSITION INFO PROVIDING UNIT
TS_PI / NS_PI
300
BEAMFORMING POSITION INFO PROVIDING UNIT
SP_PI / RS_PI
LM_IN
SPI
400
BEAMFORMING UNIT
I_IN
IMAGE POSITION INFO PROVIDING UNIT
IS_PI
FA_IM
200
BS_S
700
SPEECH RECOGNITION UNIT

도면13


시작
음원 위치정보 제공부가 소리 입력신호에 기초하여 타겟 음원의 위치정보 및 외부 잡음원의 위치정보를 제공
S100
영상 위치정보 제공부가 영상 입력신호에 기초하여 영상 입력신호에 포함되는 관심 음원의 위치정보를 제공하고, 관심 음원 중 사람의 얼굴로 판단되는 얼굴 영역 영상을 제공
S200
빔포밍 위치정보 제공부가 타겟 음원의 위치정보, 외부 잡음원의 위치정보, 관심 음원의 위치정보 및 얼굴 영역 영상으로부터 추출되는 발화자의 입술 움직임 정보에 기초하여 발화자의 위치정보 및 나머지 음원의 위치정보를 제공
S300
빔포밍부가 발화자의 위치정보 및 나머지 음원의 위치정보에 기초하여 빔포밍된 빔포밍 소리신호를 제공
S400
음성인식기가 빔포밍 소리신호에 기초하여 음성을 인식
S500
종료