

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2022년 1월 13일 (13.01.2022) WIPO | PCT

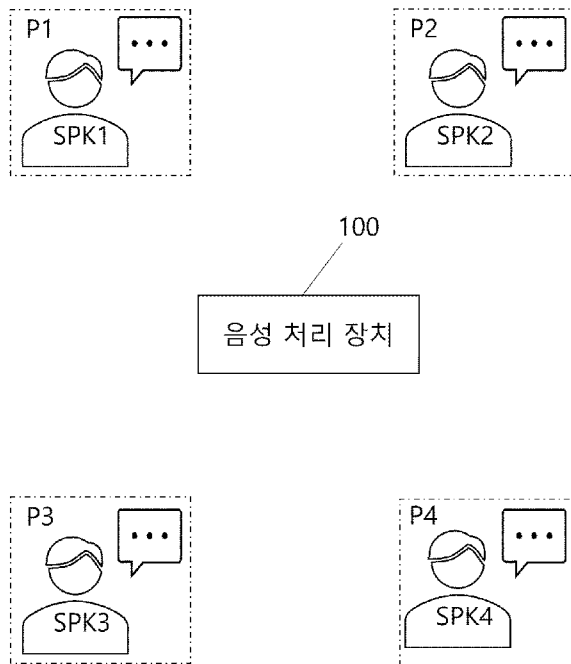


(10) 국제공개번호
WO 2022/010320 A1

- (51) 국제특허분류:
G10L 17/02 (2013.01) H05B 47/12 (2020.01)
H04R 3/00 (2006.01) G06F 40/47 (2020.01)
G01S 3/808 (2006.01) G10L 15/00 (2006.01)
G06K 9/00 (2006.01) G10L 21/0272 (2013.01)
H04N 7/15 (2006.01) H04R 1/40 (2006.01)
- (21) 국제출원번호: PCT/KR2021/008826
- (22) 국제출원일: 2021년 7월 9일 (09.07.2021)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2020-0085369 2020년 7월 10일 (10.07.2020) KR
10-2020-0141311 2020년 10월 28일 (28.10.2020) KR
10-2020-0144524 2020년 11월 2일 (02.11.2020) KR
10-2020-0144534 2020년 11월 2일 (02.11.2020) KR
- (71) 출원인: 주식회사 아모센스 (AMOSENSE CO., LTD.)
[KR/KR]: 31040 충청남도 천안시 서북구 직산읍 4산단 5길 90 천안 제4지방산업단지 19-1블럭, Chungcheongnam-do (KR).
- (72) 발명자: 김정민 (KIM, Jungmin); 31040 충청남도 천안시 서북구 직산읍 4산단5길 90 천안 제4지방산업단지 19-1블럭, Chungcheongnam-do (KR).
- (74) 대리인: 김철진 (KIM, Churchill); 06038 서울시 강남구 강남대로146길 25 전원빌딩 2층, Seoul (KR).
- (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.
- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI

(54) Title: DEVICE FOR PROCESSING VOICE AND OPERATION METHOD THEREOF

(54) 발명의 명칭: 음성을 처리하기 위한 장치 및 이의 작동 방법



100 ... Voice processing device

(57) Abstract: Disclosed is a voice processing device. The voice processing device comprises a memory and a processor configured to perform sound source isolation on voice signals associated with the voices of speakers on the basis of the sound source positions of the respective voices. The processor is configured to: generate sound source position information indicating the sound source positions of the respective voices by using the voice signals associated with the voices; generate isolated voice signals associated with the voices of the respective speakers from the voice signals on the basis of the sound source position information; and match the isolated voice signals and the voice source position information and store the same in the memory.

(57) 요약서: 음성 처리 장치가 개시된다. 음성 처리 장치는 화자들의 음성들과 연관된 음성 신호를 음성들 각각의 음원 위치에 기초하여 음원 분리를 수행하도록 구성되는 프로세서 및 메모리를 포함하고, 프로세서는, 음성들과 연관된 음성 신호를 이용하여 음성들 각각의 음원 위치를 나타내는 음원 위치 정보를 생성하고, 기 음원 위치 정보에 기초하여, 음성 신호로부터 화자들 각각의 음성과 연관된 분리 음성 신호를 생성하고, 분리 음성 신호와 음원 위치 정보를 서로 매칭하여 메모리에 저장하도록 구성된다.

(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML,
MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

명세서

발명의 명칭: 음성을 처리하기 위한 장치 및 이의 작동 방법

기술분야

- [1] 본 발명의 실시 예들은 음성을 처리하기 위한 장치 및 이의 작동 방법에 관한 것이다.

배경기술

- [2] 마이크(microphone)는 음성을 전기적인 신호인 음성 신호로 변환하는 장치이다. 회의실이나 교실과 같이 복수의 화자(speaker)들이 위치하는 공간 내에 마이크가 배치되는 경우, 상기 마이크는 복수의 화자들로부터 나온 음성들을 모두 수신하고, 복수의 화자들의 음성에 연관된 음성 신호들을 생성한다.
- [3] 한편, 복수의 화자들이 동시에 발화하는 경우, 상기 복수의 화자들의 음성들이 모두 혼합될 수 있다. 이 때, 복수의 화자들의 음성들 중에서 특정 화자의 음성을 나타내는 음성 신호를 분리하는 것이 필요하다.

발명의 상세한 설명

기술적 과제

- [4] 본 발명이 해결하고자 하는 과제는 입력된 복수의 음성 신호들을 이용하여 화자의 위치를 파악하고, 복수의 음성 신호들을 화자별로 분리 및 인식할 수 있는 장치 및 이의 작동 방법을 제공하는 것에 있다.
- [5] 본 발명이 해결하고자 하는 과제는 화자들의 음성들에 응답하여, 화자들의 각각의 음성과 연관된 분리 음성 신호를 생성할 수 있는 장치 및 이의 작동 방법을 제공하는 것에 있다.
- [6] 본 발명이 해결하고자 하는 과제는 화자들의 각각의 음성과 연관된 분리 음성 신호를 이용하여, 화자들 각각의 음성에 대한 번역 결과를 생성하고, 생성된 번역 결과를 출력할 수 있는 장치 및 이의 작동 방법을 제공하는 것에 있다.

기술적 해결방법

- [7] 본 발명의 실시 예들에 따른 음성 처리 장치는 화자들의 음성들과 연관된 음성 신호를 음성들 각각의 음원 위치에 기초하여 음원 분리를 수행하도록 구성되는 프로세서 및 메모리를 포함하고, 프로세서는, 음성들과 연관된 음성 신호를 이용하여 음성들 각각의 음원 위치를 나타내는 음원 위치 정보를 생성하고, 기 음원 위치 정보에 기초하여, 음성 신호로부터 화자들 각각의 음성과 연관된 분리 음성 신호를 생성하고, 분리 음성 신호와 음원 위치 정보를 서로 매칭하여 메모리에 저장하도록 구성된다.

발명의 효과

- [8] 본 발명의 실시 예들에 따른 장치는 음성 신호를 이용하여 화자 위치를 파악할 수 있고, 화자 위치를 통해 음성 신호가 어떤 화자의 음성에 대응하는 것인지

구별할 수 있다. 이에 따라, 다수의 화자가 동시에 음성을 발화하더라도, 음성 분리 장치는 음성을 화자 별로 구분하여 분리할 수 있는 효과가 있다.

[9] 본 발명의 실시 예들에 따른 음성 처리 장치는 음성의 음원 위치에 기초하여 특정 음원 위치로부터의 음성과 연관된 분리 음성 신호를 생성할 수 있으므로, 주변 소음의 영향을 최소화된 음성 신호를 생성할 수 있는 효과가 있다.

[10] 본 발명의 실시 예들에 따른 음성 처리 장치는 전송된 음성 신호로부터 화자들 각각의 음성을 추출할 수 있을 뿐만 아니라, 음성들의 음원 위치에 기초하여 음성들의 번역 전 언어인 출발 언어를 판단하고, 판단된 출발 언어에 기초하여 해당 음성을 번역하여 번역 결과를 제공할 수 있는 효과가 있다.

도면의 간단한 설명

[11] 도 1은 본 발명의 실시 예들에 따른 음성 처리 환경을 나타내는 도면이다.

[12] 도 2는 본 발명의 실시 예들에 따른 음성 처리 장치를 나타낸다.

[13] 도 3 내지 도 5는 본 발명의 실시 예들에 따른 음성 처리 장치의 동작을 설명하기 위한 도면이다.

[14] 도 6은 본 발명의 실시 예들에 따른 음성 처리 장치에 의한 음성 분리 방법을 나타내는 플로우 차트이다.

[15] 도 7은 본 발명의 실시 예들에 따른 음성 처리 장치의 번역 기능을 설명하기 위한 도면이다.

[16] 도 8은 본 발명의 실시 예들에 따른 음성 처리 장치의 번역 기능을 설명하기 위한 도면이다.

[17] 도 9는 본 발명의 실시 예들에 따른 음성 처리 장치에 의한 번역 결과의 제공 방법을 나타내는 플로우 차트이다.

[18] 도 10 및 도 11은 본 발명의 실시 예들에 따른 음성 처리 장치의 동작을 설명하기 위한 도면이다.

[19] 도 12는 본 발명의 실시 예들에 따른 음성 처리 장치의 작동을 나타내는 도면이다.

[20] 도 13은 본 발명의 실시 예들에 따른 음성 처리 장치를 나타낸다.

[21] 도 14 및 도 15는 본 발명의 실시 예들에 따른 화자 이동 모드를 설명하기 위한 도면이다.

[22] 도 16은 본 발명의 실시 예들에 따른 음성 처리 장치를 나타낸다.

[23] 도 17 및 도 18은 본 발명의 실시 예들에 따른 음성 처리 장치의 작동을 나타내는 도면이다.

[24] 도 19는 본 발명의 실시 예들에 따른 음성 처리 장치의 작동 방법을 나타내는 플로우 차트이다.

[25] 도 20은 본 발명의 실시 예들에 따른 음성 처리 장치를 나타낸다.

[26] 도 21은 본 발명의 실시 예들에 따른 음성 처리 장치를 나타낸다.

[27] 도 22 및 도 23은 본 발명의 실시 예들에 따른 음성 처리 장치의 동작을

설명하기 위한 도면이다.

- [28] 도 24는 본 발명의 실시 예들에 따른 음성 처리 장치의 작동 방법을 나타내는 플로우 차트이다.
- [29] 도 25는 본 발명의 실시 예들에 따른 음성 처리 장치의 동작을 설명하기 위한 도면이다.
- [30] 도 26은 본 발명의 실시 예들에 따른 음성 처리 장치의 작동을 나타내는 도면이다.
- [31] 도 27은 본 발명의 실시 예들에 따른 음성 처리 장치의 동작을 설명하기 위한 도면이다.

발명의 실시를 위한 형태

- [32] 이하, 첨부된 도면들을 참조하여 본 발명의 실시 예들을 설명한다.
- [33] 도 1은 본 발명의 실시 예들에 따른 음성 처리 환경을 나타내는 도면이다. 도 1을 참조하면, 화자들(SPK1~SPK4)은 공간(예컨대, 회의실, 차량, 강의실 등)에 위치하여 음성을 발화(pronounce)할 수 있다. 실시 예들에 따라, 제1화자(SPK1)는 제1위치(P1)에서 음성을 발화할 수 있고, 제2화자(SPK2)는 제2위치(P2)에 위치에서 음성을 발화할 수 있고, 제3화자(SPK3)는 제3위치(P3)에 위치에서 음성을 발화할 수 있고, 제4화자(SPK4)는 제4위치(P4)에서 음성을 발화 위치할 수 있다.
- [34] 음성 처리 장치(100)는 연산 처리 기능을 갖는 전자 장치일 수 있다. 예컨대, 음성 처리 장치(100)는 스마트폰(smartphone), 노트북(laptop), PDA(personal digital assistance), 웨어러블 디바이스(wearable device), 스마트 워치(smart watch) 또는 태블릿 컴퓨터(tablet computer)일 수 있으나, 본 발명의 실시 예들이 이에 한정되는 것은 아니다.
- [35] 음성 처리 장치(100)는 화자들(SPK1~SPK4)의 음성들과 연관된 음성 신호를 처리함으로써, 화자들(SPK1~SPK4) 각각의 음성에 대한 음성 처리를 수행할 수 있다.
- [36] 음성 처리 장치(100)는 화자들(SPK1~SPK4) 각각의 음성에 응답하여, 화자들(SPK1~SPK4)의 음성과 연관된 음성 신호를 생성할 수 있다. 상기 음성 신호는 특정 시간동안 발화된 음성들과 연관된 신호로서, 복수의 화자들의 음성을 나타내는 신호일 수 있다.
- [37] 실시 예들에 따라, 음성 처리 장치(100)는 화자들(SPK1~SPK4)의 음성과 연관된 음성 신호를 이용하여, 화자들(SPK1~SPK4)의 음성들 각각의 음원 위치를 판단하고, 음원 위치에 기초하여 음원 분리를 수행함으로써, 음성 신호로부터 화자들(SPK1~SPK4) 각각의 음성과 연관된 분리 음성 신호를 추출(또는 생성)할 수 있다.
- [38] 즉, 음성 처리 장치(100)는 음성 신호들에 대응하는 음성들의 음원 위치에 기초하여, 각 위치(P1~P4)에 위치한 화자들(SPK1~SPK4)의 음성과 연관된 분리

음성 신호를 생성할 수 있다. 예컨대, 음성 처리 장치(100)는 음성 신호에 기초하여, 제1위치(P1)에서 발화한 제1화자(SPK1)의 음성과 연관된 제1분리 음성 신호를 생성할 수 있다. 이 때, 제1분리 음성 신호는 화자들(SPK1~SPK4)의 음성들 중 제1화자(SPK1)의 음성과 가장 높은 연관도를 가지는 음성 신호일 수 있다. 다시 말하면, 제1분리 음성 신호에 포함된 음성 성분 중에서 제1화자(SPK1)의 음성 성분의 비중이 가장 높을 수 있다.

[39] 또한, 음성 처리 장치(100)는 화자들(SPK1~SPK4) 각각의 음성에 대한 번역을 제공할 수 있다. 예컨대, 음성 처리 장치(100)는 화자들(SPK1~SPK4) 각각의 음성을 번역하기 위한 출발 언어(source language; 번역 대상 언어)와 도착 언어(target language; 번역 후 언어)를 결정하고, 분리 음성 신호를 이용하여 화자들 각각의 언어에 대한 번역을 제공할 수 있다.

[40] 실시 예들에 따라, 음성 처리 장치(100)는 음성들 각각에 대한 번역 결과를 출력할 수 있다. 상기 번역 결과는 도착 언어로 표현된 화자들(SPK1~SPK4) 각각의 음성과 연관된 텍스트 데이터 또는 음성 신호일 수 있다.

[41] 즉, 본 발명의 실시 예들에 따른 음성 처리 장치(100)는 화자들(SPK1~SPK4)의 음성 각각의 음원 위치에 따라 출발 언어와 도착 언어를 결정하므로, 화자의 음성의 언어가 무엇인지 식별할 필요없이 적은 시간과 적은 리소스로 화자의 음성에 대한 번역을 제공할 수 있는 효과가 있다.

[42] 도 2는 본 발명의 실시 예들에 따른 음성 처리 장치를 나타낸다. 도 2를 참조하면, 음성 처리 장치(100)는 마이크(110), 통신 회로(120), 프로세서(130) 및 메모리(140)를 포함할 수 있다. 실시 예들에 따라, 음성 처리 장치(100)는 스피커(150)를 더 포함할 수 있다.

[43] 마이크(110)는 발생한 음성에 응답하여 음성 신호를 생성할 수 있다. 실시 예들에 따라, 마이크(110)는 음성으로 인한 공기의 진동을 검출하고, 검출 결과에 따라 진동에 대응하는 전기적인 신호인 음성 신호를 생성할 수 있다.

[44] 실시 예들에 따라, 마이크(110)는 복수일 수 있고, 복수의 마이크들(110) 각각은 음성에 응답하여 음성 신호를 생성할 수 있다. 이 때, 복수의 마이크들(110) 각각이 배치된 위치는 서로 다를 수 있으므로, 마이크들(110) 각각으로부터 생성된 음성 신호들은 서로 위상 차(또는 시간 지연)를 가질 수 있다.

[45] 예컨대, 마이크(110)는 각 위치(P1~P4)에 위치한 화자(SPK1~SPK4)의 음성을 수신하고, 화자(SPK1~SPK4)의 음성을 전기적인 신호인 음성 신호들로 변환할 수 있다. 통신 회로(120)는 무선 통신 방식에 따라 외부 장치와 데이터를 주고받을 수 있다. 실시 예들에 따라, 통신 회로(120)는 다양한 주파수의 전파를 이용하여, 외부 장치와 데이터를 주고받을 수 있다. 예컨대, 통신 회로(120)는 근거리 무선 통신, 중거리 무선 통신 및 장거리 무선 통신 중 적어도 하나의 무선 통신 방식에 따라, 외부 장치와 데이터를 주고받을 수 있다.

[46] 프로세서(130)는 음성 처리 장치(100)의 전반적인 동작을 제어할 수 있다. 실시 예들에 따라, 프로세서(130)는 연산 처리 기능을 갖는 프로세서를 포함할 수

있다. 예컨대, 프로세서(130)는 CPU(central processing unit), MCU(micro controller unit), GPU(graphics processing unit), DSP(digital signal processor), ADC 컨버터(analog to digital converter) 또는 DAC 컨버터(digital to analog converter)를 포함할 수 있으나, 이에 한정되는 것은 아니다.

- [47] 프로세서(130)는 마이크(110)에 의해 생성된 음성 신호들을 처리할 수 있다. 예컨대, 프로세서(130)는 마이크(110)에 의해 생성된 아날로그 타입의 음성 신호를 디지털 타입의 음성 신호로 변환하고, 변환된 디지털 타입의 음성 신호를 처리할 수 있다. 이 경우, 신호의 타입(아날로그 또는 디지털)이 바뀌는 것이므로, 본 발명의 실시 예들에 대한 설명에 있어서, 디지털 타입의 음성 신호와 아날로그 타입의 음성 신호를 혼용하여 설명하도록 한다.
- [48] 실시 예들에 따라, 프로세서(130)는 마이크(110)에 의해 생성된 음성 신호를 이용하여, 화자들(SPK1~SPK4) 각각의 음성과 연관된 분리 음성 신호를 추출(또는 생성)할 수 있다. 실시 예들에 따라, 프로세서(130)는 각 위치(P1~P4)에 위치한 화자(SPK1~SPK4)의 음성과 연관된 분리 음성 신호를 생성할 수 있다.
- [49] 프로세서(130)는 음성 신호들 사이의 시간 지연(또는 위상 지연)을 이용하여 음성들의 음원 위치(즉, 화자들(SPK1~SPK4)의 위치)를 결정할 수 있다. 예컨대, 프로세서(130)는 음성 처리 장치(100)에 대한 음원(즉, 화자들(SPK1~SPK4))의 상대적인 위치를 결정할 수 있다.
- [50] 프로세서(130)는 결정된 음원 위치에 기초하여, 화자들(SPK1~SPK4) 각각의 음성과 연관된 분리 음성 신호를 생성할 수 있다. 예컨대, 프로세서(130)는 음성들의 음원 위치에 기초하여, 제1화자(SPK1)의 음성과 연관된 제1분리 음성 신호를 생성할 수 있다.
- [51] 실시 예들에 따라, 프로세서(130)는 결정된 음원 위치를 나타내는 음원 위치 정보를 분리 음성 신호와 매칭하여 저장할 수 있다. 예컨대, 프로세서(130)는 제1화자(SPK1)의 음성과 연관된 제1분리 음성 신호 및 제1화자(SPK1)의 음성의 음원 위치를 나타내는 제1음원 위치 정보를 매칭하여 메모리(140)에 저장할 수 있다. 즉, 음원의 위치가 곧 화자들(SPK1~SPK4) 각각의 위치와 대응되므로, 음원 위치 정보는 화자들(SPK1~SPK4) 각각의 위치를 식별하기 위한 화자 위치 정보로 기능할 수도 있다.
- [52] 본 명세서에서 설명되는 프로세서(130) 또는 음성 처리 장치(100)의 동작은 컴퓨팅 장치에 의해 실행 가능한 프로그램의 형태로 구현될 수 있다. 예컨대, 프로세서(130)는 메모리(140)에 저장된 애플리케이션을 실행하고, 애플리케이션의 실행에 따라 특정 작동들을 지시하는 명령어들에 대응하는 작동들을 수행할 수 있다.
- [53] 메모리(140)는 음성 처리 장치(100)의 동작에 필요한 데이터를 저장할 수 있다. 예컨대, 메모리(140)는 비휘발성 메모리 및 휘발성 메모리 중 적어도 하나를 포함할 수 있다.
- [54] 실시 예들에 따라, 메모리(140)는 공간 상의 각 위치(P1~P4)에 대응하는

식별자를 저장할 수 있다. 상기 식별자는 위치(P1~P4)를 구별하기 위한 데이터일 수 있다. 위치들(P1~P4) 각각에는 화자들(SPK1~SPK4) 각각이 위치하므로, 위치(P1~P4)에 대응하는 식별자를 이용하여 화자(SPK1~SPK4)들 각각을 구별할 수 있다. 예컨대, 제1위치(P1)를 나타내는 제1식별자는, 곧, 제1화자(SPK1)를 나타낼 수 있다. 이러한 관점에서, 공간 상의 각 위치(P1~P4)에 대응하는 식별자는 화자들(SPK1~SPK4) 각각을 식별하기 위한 화자 식별자로서 기능할 수도 있다.

- [55] 상기 식별자는 음성 처리 장치(100)의 입력 장치(예컨대, 터치 패드)를 통해 입력될 수 있다.
- [56] 실시 예들에 따라, 메모리(140)는 화자들(SPK1~SPK4) 각각의 위치와 관련된 음원 위치 정보 및 화자들(SPK1~SPK4) 각각의 음성과 연관된 분리 음성 신호를 저장할 수 있다.
- [57] 스피커(150)는 프로세서(130)의 제어에 따라 진동할 수 있고, 상기 진동에 따라 음성이 생성될 수 있다. 실시 예들에 따라, 스피커(150)는 음성 신호에 대응하는 진동을 형성함으로써, 상기 음성 신호와 연관된 음성을 재생할 수 있다.
- [58] 한편, 본 명세서에서는 음성 처리 장치(100)가 마이크(110)를 포함하고, 마이크(110)를 이용하여 화자들(SPK1~SPK4)의 음성들과 연관된 음성 신호를 직접 생성하는 것으로 설명하나, 실시 예들에 따라, 마이크는 음성 처리 장치(100)와 분리되어 외부에 구성될 수 있고, 음성 처리 장치(100)는 분리되어 구성된 마이크로부터 음성 신호를 수신하여, 수신된 음성 신호를 처리 또는 이용할 수 있다. 예컨대, 음성 처리 장치(100)는 분리된 마이크로부터 수신된 음성 신호로부터 분리 음성 신호를 생성할 수 있다.
- [59] 다만, 설명의 편의상, 별도의 언급이 없는 한 음성 처리 장치(100)가 마이크(110)를 포함하는 것을 가정하고 설명한다.
- [60] 도 3 내지 도 5는 본 발명의 실시 예들에 따른 음성 처리 장치의 동작을 설명하기 위한 도면이다. 도 3 내지 도 5를 참조하면, 각 위치(P1~P4)에 위치한 화자들(SPK1~SPK4) 각각이 발화할 수 있다.
- [61] 본 발명의 실시 예들에 따른 음성 처리 장치(100)는 화자들(SPK1~SPK4)의 음성들로부터 각 화자들(SPK1~SPK4)의 음성과 연관된 분리 음성 신호를 생성할 수 있고, 분리 음성 신호와 화자들(SPK1~SPK4) 각각의 위치를 나타내는 위치 정보를 저장할 수 있다.
- [62] 실시 예들에 따라, 음성 처리 장치(100)는 음성 신호들 사이의 시간 지연(또는 위상 지연)을 이용하여 음성들의 음원 위치(즉, 화자들(SPK1~SPK4)의 위치)를 결정할 수 있다. 예컨대, 음성 처리 장치(100)는 음성 처리 장치(100)에 대한 음원(즉, 화자들(SPK1~SPK4))의 상대적인 위치를 결정할 수 있다.
- [63] 음성 처리 장치(100)는 결정된 음원 위치에 기초하여, 화자들(SPK1~SPK4) 각각의 음성과 연관된 분리 음성 신호를 생성할 수 있다.
- [64] 도 3에 도시된 바와 같이, 제1화자(SPK1)가 음성 "AAA"를 발화한다. 음성

"AAA"가 발화되면, 음성 처리 장치(100)는 음성 "AAA"에 응답하여, 음성 "AAA"와 연관된 음성 신호를 생성할 수 있다. 실시 예들에 따라, 음성 "AAA"와 연관된 음성 신호에는 음성 "AAA"이외의 소음들과 관련된 성분도 포함될 수 있다.

- [65] 실시 예들에 따라, 음성 처리 장치(100)는 생성된 음성 신호를 이용하여, 제1화자(SPK1)의 음성 "AAA"와 연관된 분리 음성 신호를 생성할 수 있다. 이 때, 음성 처리 장치(100)는 제1화자(SPK1)의 음성 "AAA"와 연관된 제1분리 음성 신호와 제1화자(SPK1)의 위치인 제1위치(P1)를 나타내는 제1음원 위치 정보를 메모리(140)에 저장할 수 있다. 예컨대, 도 3에 도시된 바와 같이, 제1분리 음성 신호와 제1음원 위치 정보는 서로 매칭되어 저장될 수 있다.
- [66] 도 4에 도시된 바와 같이, 제2화자(SPK2)가 음성 "BBB"를 발화한다. 음성 "BBB"가 발화되면, 음성 처리 장치(100)는 음성 "BBB"에 응답하여, 음성 "BBB"와 연관된 음성 신호를 생성할 수 있다.
- [67] 실시 예들에 따라, 음성 처리 장치(100)는 생성된 음성 신호를 이용하여, 제2화자(SPK2)의 음성 "BBB"와 연관된 제2분리 음성 신호를 생성할 수 있다. 이 때, 음성 처리 장치(100)는 제2화자(SPK2)의 음성 "BBB"와 연관된 제2분리 음성 신호와 제2화자(SPK2)의 위치인 제2위치(P2)를 나타내는 제2음원 위치 정보를 메모리(140)에 저장할 수 있다. 예컨대, 도 4에 도시된 바와 같이, 제2분리 음성 신호와 제2음원 위치 정보는 서로 매칭되어 저장될 수 있다.
- [68] 도 5에 도시된 바와 같이, 제3화자(SPK3)가 음성 "CCC"를 발화하고, 제4화자(SPK4)가 음성 "DDD"를 발화한다. 음성 처리 장치(100)는 음성 "CCC" 및 음성 "DDD"에 응답하여, 음성 "CCC" 및 음성 "DDD"와 연관된 음성 신호를 생성할 수 있다. 즉, 상기 음성 신호는 음성 "CCC" 및 음성 "DDD"와 연관된 성분을 포함하는 음성 신호이다.
- [69] 실시 예들에 따라, 음성 처리 장치(100)는 생성된 음성 신호를 이용하여, 제3화자(SPK3)의 음성 "CCC"와 연관된 제3분리 음성 신호 및 제4화자(SPK4)의 음성 "DDD"와 연관된 제4분리 음성 신호를 생성할 수 있다.
- [70] 이 때, 음성 처리 장치(100)는 제3화자(SPK3)의 음성 "CCC"와 연관된 제3분리 음성 신호와 제3화자(SPK3)의 위치인 제3위치(P3)를 나타내는 제3위치 정보를 메모리(140)에 저장할 수 있다. 또한, 음성 처리 장치(100)는 제4화자(SPK4)의 음성 "DDD"와 연관된 제4분리 음성 신호와 제4화자(SPK4)의 위치인 제4위치(P4)를 나타내는 제4위치 정보를 메모리(140)에 저장할 수 있다.
- [71] 예컨대, 도 4에 도시된 바와 같이, 제3분리 음성 신호와 제3음원 위치 정보는 서로 매칭되어 저장될 수 있고, 제4분리 음성 신호와 제4음원 위치 정보는 서로 매칭되어 저장될 수 있다.
- [72] 즉, 본 발명의 실시 예들에 따른 음성 처리 장치(100)는 화자들(SPK1~SPK4)의 음성들로부터 각 화자들(SPK1~SPK4)의 음성과 연관된 분리 음성 신호를 생성할 수 있고, 분리 음성 신호와 화자들(SPK1~SPK4) 각각의 위치를 나타내는 위치

정보를 저장할 수 있다.

- [73] 도 6은 본 발명의 실시 예들에 따른 음성 처리 장치에 의한 음성 분리 방법을 나타내는 플로우 차트이다. 도 6을 참조하여 설명될 음성 처리 장치의 작동 방법은 비일시적인 저장 매체에 저장되어, 컴퓨팅 장치에 의해 실행 가능한 애플리케이션(예컨대, 음성 분리 애플리케이션)으로서 구현될 수 있다. 예컨대, 프로세서(130)는 메모리(140)에 저장된 애플리케이션을 실행하고, 애플리케이션의 실행에 따라 특정 작동들을 지시하는 명령어들에 대응하는 작동들을 수행할 수 있다.
- [74] 도 6을 참조하면, 음성 처리 장치(100)는 음성에 응답하여, 음성 신호를 생성할 수 있다(S110). 실시 예들에 따라, 음성 처리 장치(100)는 공간에서 감지되는 음성을 전기적인 신호인 음성 신호로 변환할 수 있다.
- [75] 음성 처리 장치(100)는 화자들(SPK1~SPK4)의 음성과 연관된 음성 신호를 이용하여, 음성들 각각에 대한 음원 위치(즉, 화자들(SPK1~SPK4)의 위치)를 판단할 수 있다(S120). 실시 예들에 따라, 음성 처리 장치(100)는 화자들(SPK1~SPK4)의 음성들 각각에 대한 음원 위치(즉, 화자들(SPK1~SPK4)의 위치)를 나타내는 음원 위치 정보를 생성할 수 있다.
- [76] 음성 처리 장치(100)는 음성들 각각에 대한 음원 위치에 기초하여, 화자들(SPK1~SPK4)의 음성들 각각과 연관된 분리 음성 신호를 생성할 수 있다(S130). 실시 예들에 따라, 음성 처리 장치(100)는 생성된 음성 신호를, 음성들 각각에 대한 음원 위치에 기초하여 분리함으로써 화자들(SPK1~SPK4)의 음성들 각각과 연관된 분리 음성 신호를 생성할 수 있다. 예컨대, 음성 처리 장치(100)는 음성 신호에 포함된 성분들을 음원 위치에 기초하여 분리함으로써 화자들(SPK1~SPK4)의 음성들 각각과 연관된 분리 음성 신호를 생성할 수 있다.
- [77] 음성 처리 장치(100)는 음원의 위치를 나타내는 음원 위치 정보와 분리 음성 신호를 저장할 수 있다(S140). 실시 예들에 따라, 음성 처리 장치(100)는 음원의 위치를 나타내는 음원 위치 정보와, 화자들(SPK1~SPK4) 각각의 음성과 연관된 분리 음성 신호를 매칭하여 저장할 수 있다. 예컨대, 음성 처리 장치(100)는 화자들(SPK1~SPK4) 각각의 음성과 연관된 분리 음성 신호에 해당하는 데이터와 음원 위치 정보를 매칭하여 저장할 수 있다.
- [78] 실시 예들에 따라, 본 발명의 실시 예들에 따른 음성 처리 장치(100)(또는 프로세서(130))는 메모리(140)에 저장된 애플리케이션(예컨대, 음성 분리 애플리케이션)을 실행함으로써 화자들(SPK1~SPK4)의 음성들과 연관된 음성 신호로부터 화자들(SPK1~SPK4) 각각의 음성과 연관된 분리 음성 신호를 생성(또는 분리)할 수 있다.
- [79] 일반적으로, 음성 신호에 대한 처리를 수행하기 위해서는 마이크 및 음성 신호를 처리하도록 구성되는 프로세서 등의 하드웨어가 필요하다. 한편, 스마트폰과 같은 모바일 단말기는 스피커 및 프로세서를 기본적으로 포함하므로, 사용자들은 음성 처리 장치(100)를 이용하여 본 발명의 실시 예들에

다른 방법을 수행함으로써, 별도의 하드웨어를 구비하지 않더라도 화자들의 음성을 분리할 수 있는 효과가 있다. 예컨대, 음성 처리 장치(100)의 프로세서(130)는 음성 분리 애플리케이션을 실행하고, 음성 처리 장치(100)에 포함된 하드웨어(예컨대, 스피커)를 이용하여 음성 분리를 수행할 수 있다.

[80] 도 7은 본 발명의 실시 예들에 따른 음성 처리 장치의 번역 기능을 설명하기 위한 도면이다. 도 7을 참조하면, 제1화자(SPK1)는 음성 "AAA"를 한국어(KR)로 발화하고, 제2화자(SPK2)는 음성 "BBB"를 영어(EN)로 발화하고, 제3화자(SPK3)는 음성 "CCC"를 중국어(CN)로 발화하고, 제4화자(SPK4)는 음성 "DDD"를 일본어(JP)로 발화한다.

[81] 본 발명의 실시 예들에 따른 음성 처리 장치(100)는 화자들(SPK1~SPK4)의 음성들로부터 각 화자들(SPK1~SPK4)의 음성과 연관된 분리 음성 신호를 생성할 수 있고, 분리 음성 신호를 이용하여 화자들(SPK1~SPK4) 각각의 음성에 대한 번역을 제공할 수 있다. 이 때, 음성 처리 장치(100)는 화자들(SPK1~SPK4) 각각의 위치와 대응하는 출발 언어 정보를 이용하여, 화자들(SPK1~SPK4)의 음성의 출발 언어를 결정하여, 음성에 대한 번역을 제공할 수 있다.

[82] 도 7에 도시된 바와 같이, 음성 처리 장치(100)는 화자들(SPK1~SPK4) 각각의 음성과 연관된 분리 음성 신호, 화자들(SPK1~SPK4)의 위치를 나타내는 음원 위치 정보 및 화자들(SPK1~SPK4)의 음성의 출발 언어를 나타내는 출발 언어 정보를 저장할 수 있다. 이 때, 출발 언어는 음원 위치 별로 미리 결정되어 저장될 수 있다.

[83] 예컨대, 음성 처리 장치(100)는 제1위치(P1)에 대응하는 출발 언어가 "KR"임을 나타내는 제1출발 언어 정보를 메모리(140)에 저장할 수 있다. 또한, 음성 처리 장치(100)는 제1화자(SPK1)의 음성 "AAA"와 연관된 제1분리 음성 신호, 제1화자(SPK1)의 위치인 제1위치(P1)를 나타내는 제1음원 위치 정보 및 제1화자(SPK1)의 음성 "AAA (KR)"의 출발 언어인 "KR"을 나타내는 제1출발 언어 정보를 메모리(140)에 저장할 수 있다.

[84] 실시 예들에 따라, 화자들(SPK1~SPK4)이 음성을 발화하면, 음성 처리 장치(100)는 화자들(SPK1~SPK4)의 음성에 응답하여, 화자들(SPK1~SPK4) 각각의 음성과 연관된 분리 음성 신호 및 화자들(SPK1~SPK4)의 위치를 나타내는 음원 위치 정보를 생성할 수 있다.

[85] 음성 처리 장치(100)는 출발 언어 정보들을 이용하여, 각 분리 음성 신호에 대응하는 출발 언어를 결정하고, 결정된 출발 언어에 기초하여 화자들(SPK1~SPK4)의 음성에 대한 번역을 제공할 수 있다. 실시 예들에 따라, 음성 처리 장치(100)는 각 분리 음성 신호에 대응하는 음원 위치 정보를 이용하여, 각 음성의 음원 위치와 대응하는 출발 언어를 결정하고, 결정된 출발 언어에 기초하여 분리 음성 신호에 대한 번역 결과를 생성할 수 있다.

[86] 예컨대, 음성 처리 장치(100)는 분리 음성 신호를 텍스트 데이터로 변환하고(예컨대, STT(Speech-To-Text) 변환), 변환된 텍스트 데이터에 대해서

출발 언어로부터 도착 언어로의 번역 결과를 생성하고, 번역 결과를 음성 신호로서 변환(예컨대, TTS(Text-to-Speech) 변환)할 수 있다. 즉, 본 명세서에서 언급되는 번역 결과는 도착 언어로 표현된 화자들(SPK1~SPK4) 각각의 음성과 연관된 텍스트 데이터 또는 음성 신호를 모두 의미할 수 있다.

[87] 실시 예들에 따라, 음성 처리 장치(100)는 생성된 번역 결과를 출력할 수 있다. 예컨대, 음성 처리 장치(100)는 생성된 번역 결과를 스피커(150)를 통해 출력하거나, 또는 다른 외부 장치로 전송할 수 있다.

[88] 도 8은 본 발명의 실시 예들에 따른 음성 처리 장치의 번역 기능을 설명하기 위한 도면이다. 도 8을 참조하면, 음성 처리 장치(100)는 화자들(SPK1~SPK4)의 각각의 음성과 연관된 분리 음성 신호를 생성하고, 분리 음성 신호들을 이용하여 화자들(SPK1~SPK4)의 각각의 음성에 대한 번역 결과를 출력할 수 있다. 이 때, 번역 결과는 화자들(SPK1~SPK4)의 음성들의 언어가 출발 언어로부터 다른 언어(예컨대, 도착 언어)로 변환된 결과를 나타낸다.

[89] 도 8에 도시된 바와 같이, 제1화자(SPK1)는 음성 "AAA"를 한국어(KR)로 발화하고, 제2화자(SPK2)는 음성 "BBB"를 영어(EN)로 발화한다. 이 경우, 1화자(SPK1)의 음성 "AAA"의 출발 언어는 한국어(KR)이고, 제2화자(SPK2)의 음성 "BBB"의 출발 언어는 영어(EN)가 된다.

[90] 음성 처리 장치(100)는 제1화자(SPK1)의 음성 "AAA(KR)"에 응답하여, 제1화자(SPK1)의 음원 위치(예컨대, P1)를 결정하고, 음원 위치에 기초하여 제1화자(SPK1)의 음성 "AAA(KR)"과 연관된 제1분리 음성 신호를 생성할 수 있다. 마찬가지로, 음성 처리 장치(100)는 제2화자(SPK2)의 음성 "BBB(EN)"에 응답하여, 제2화자(SPK2)의 음원 위치(예컨대, P2)를 결정하고, 음원 위치에 기초하여 제2화자(SPK2)의 음성 "BBB(EN)"과 연관된 제2분리 음성 신호를 생성할 수 있다.

[91] 음성 처리 장치(100)는 생성된 분리 음성 신호를 이용하여, 화자들(SPK1~SPK4)의 음성의 언어에 대한 출발 언어로부터 도착 언어로의 번역을 제공할 수 있다. 실시 예들에 따라, 음성 처리 장치(100)는 메모리(140)에 저장된 출발 언어 정보를 이용하여, 화자들(SPK1~SPK4)의 음성의 음원 위치에 따라 결정되는 출발 언어를 결정하고, 결정된 출발 언어에 따라 화자들(SPK1~SPK4) 각각의 음성의 언어에 대한 출발 언어로부터 도착 언어로의 번역 결과를 출력할 수 있다.

[92] 실시 예들에 따라, 음성 처리 장치(100)는 각 위치에 대한 도착 언어를 나타내는 도착 언어 정보를 저장할 수 있고, 저장된 도착 언어 정보를 이용하여 화자들(SPK1~SPK4) 각각의 음성의 음원 위치에 대응하는 도착 언어를 결정할 수 있다. 또한, 실시 예들에 따라, 음성 처리 장치(100)는 사용자로부터의 입력에 기초하여, 화자들(SPK1~SPK4) 각각의 음성에 대한 도착 언어를 결정할 수 있다.

[93] 예컨대, 음성 처리 장치(100)는 제1화자(SPK1)의 음성 "AAA(KR)"의 음원 위치인 제1위치(P1)를 나타내는 제1음원 위치 정보를 이용하여,

- 메모리(140)로부터 제1위치(P1)에 대응하는 제1출발 언어 정보를 리드할 수 있다. 리드 된 제1출발 언어 정보는 제1화자(SPK1)의 음성 "AAA"의 출발 언어가 한국어(KR)임을 지시할 수 있다.
- [94] 상기 번역 결과는 스피커(150)를 통해 출력되거나, 메모리(140)에 저장되거나, 또는, 통신 회로(120)를 통해 외부 장치로 전송될 수 있다.
- [95] 본 명세서에서, 음성 처리 장치(100)에 의해 출력되는 번역 결과는 도착 언어로 표현된 텍스트 데이터이거나 혹은 도착 언어로 발화된 음성과 연관된 음성 신호일 수 있으나, 이에 한정되는 것은 아니다.
- [96] 본 명세서에서, 음성 처리 장치(100)가 번역 결과를 생성한다는 것은, 음성 처리 장치(100)의 프로세서(130) 자체의 연산을 통해 언어를 번역함으로써 번역 결과를 생성하는 것뿐만 아니라, 음성 처리 장치(100)가 번역 기능을 갖는 서버와의 통신을 통해, 상기 서버로부터 번역 결과를 수신함으로써 번역 결과를 생성하는 것을 포함한다.
- [97] 예컨대, 프로세서(130)는 메모리(140)에 저장된 번역 애플리케이션을 실행함으로써, 화자들(SPK1~SPK4) 각각의 음성에 대한 번역 결과를 생성할 수 있다.
- [98] 예컨대, 음성 처리 장치(100)는 분리 음성 신호, 출발 언어 정보 및 도착 언어 정보를 번역기(translator)로 전송하고, 번역기로부터 분리 음성 신호에 대한 번역 결과를 수신할 수 있다. 번역기는 언어에 대한 번역을 제공하는 환경 또는 시스템을 의미할 수 있다. 실시 예들에 따라, 번역기는 분리 음성 신호, 출발 언어 정보 및 도착 언어 정보를 이용하여, 화자들(SPK1~SPK4) 각각의 음성에 대한 번역 결과를 출력할 수 있다.
- [99] 예컨대, 도 8에 도시된 바와 같이, 음성 처리 장치(100)는 제1화자(SPK1)의 음성 "AAA(KR)"에 대한 출발 언어(즉, 한국어(KR)) 및 도착 언어(즉, 영어(EN))를 결정하고, 결정된 출발 언어 및 도착 언어에 따라, 제1화자(SPK1)의 음성 "AAA(KR)"에 대한 번역 결과를 출력할 수 있다. 예컨대, 음성 "AAA(KR)"에 대한 번역 결과는 영어(EN)로 표현된 음성 "AAA(EN)"과 연관된 데이터(예컨대, 음성 데이터 또는 텍스트 데이터 등)일 수 있다. 한편, 비록 도 8에서는 음성 "AAA(KR)"에 대한 도착 언어가 영어(EN)인 것으로 설명되어 있으나, 본 발명의 실시 예들이 이에 한정되는 것은 아니다.
- [100] 상술한 바와 같이, 음성 처리 장치(100)는 화자들(SPK1~SPK4) 음성과 연관된 분리 음성 신호에 기초하여 번역을 제공하므로, 음성 처리 장치(100)는 특정 화자의 음성에 대한 번역 결과를 출력할 수 있는 효과가 있다.
- [101] 마찬가지로, 음성 처리 장치(100)는 제2화자(SPK2)의 음성 "BBB(EN)"에 대한 출발 언어(즉, 영어(EN)) 및 도착 언어(즉, 한국어(KR))를 결정하고, 결정된 출발 언어 및 도착 언어에 따라, 제2화자(SPK2)의 음성 "BBB(EN)"에 대한 번역 결과를 출력할 수 있다. 또한, 음성 처리 장치(100)는 제3화자(SPK3)의 음성 "CCC(CN)" 및 제4화자(SPK4)의 음성 "DDD(CN)"에 대한 번역 결과 또한 출력할

수 있다.

- [102] 도 9는 본 발명의 실시 예들에 따른 음성 처리 장치에 의한 번역 결과의 제공 방법을 나타내는 플로우 차트이다. 도 9를 참조하여 설명될 음성 처리 장치의 작동 방법은 비일시적인 저장 매체에 저장되어, 컴퓨팅 장치에 의해 실행 가능한 애플리케이션(예컨대, 번역 애플리케이션)으로서 구현될 수 있다. 예컨대, 프로세서(130)는 메모리(140)에 저장된 애플리케이션을 실행하고, 애플리케이션의 실행에 따라 특정 작동들을 지시하는 명령어들에 대응하는 작동들을 수행할 수 있다.
- [103] 도 9를 참조하면, 음성 처리 장치(100)는 화자들(SPK1~SPK4) 각각의 음성과 연관된 분리 음성 신호를 생성할 수 있다(S210). 실시 예들에 따라, 음성 처리 장치(100)는 음성에 응답하여 생성된 음성 신호를 음성들 각각의 음원 위치에 기초하여 분리함으로써, 분리 음성 신호를 생성할 수 있다.
- [104] 음성 처리 장치(100)는 화자들(SPK1~SPK4) 각각의 음성을 번역하기 위한 출발 언어를 결정할 수 있다(S220). 실시 예들에 따라, 음성 처리 장치(100)는 화자들(SPK1~SPK4) 각각의 음성의 음원 위치에 기초하여, 화자들(SPK1~SPK4) 각각의 음성을 번역하기 위한 출발 언어를 결정할 수 있다. 또한, 음성 처리 장치(100)는 화자들(SPK1~SPK4) 각각의 음성의 음원 위치에 기초하여, 화자들(SPK1~SPK4) 각각의 음성이 번역될 도착 언어를 결정할 수 있다.
- [105] 음성 처리 장치(100)는 분리 음성 신호를 이용하여, 출발 언어에 따라 화자들(SPK1~SPK4) 각각의 음성에 대한 번역 결과를 출력할 수 있다(S230). 실시 예들에 따라, 음성 처리 장치(100)는 결정된 출발 언어(및 도착 언어)에 기초하여, 화자들(SPK1~SPK4) 각각의 음성에 대한 출발 언어로부터의 도착 언어로의 번역 결과를 출력할 수 있다.
- [106] 본 발명의 실시 예들에 따른 음성 처리 장치(100)는 화자들(SPK1~SPK4)의 음성과 연관된 음성 신호를 생성하고, 음성 신호들을 처리함으로써 화자들(SPK1~SPK4) 각각의 음성과 연관된 분리 음성 신호를 생성할 수 있다.
- [107] 또한, 본 발명의 실시 예들에 따른 음성 처리 장치(100)는 분리 음성 신호를 이용하여, 화자들(SPK1~SPK4)의 음성을 번역하고, 번역 결과를 출력할 수 있다. 이에 따라, 화자들(SPK1~SPK4)의 사용 언어가 다르더라도, 화자들(SPK1~SPK4) 각각은 자신이 사용하는 언어로 발화할 수 있고, 다른 언어를 사용하는 화자들의 음성을 자신이 사용하는 언어로 번역하여 들을 수 있는 효과가 있다.
- [108] 일반적으로, 음성 신호에 대한 처리를 수행하기 위해서는 마이크 및 음성 신호를 처리하도록 구성되는 프로세서 등의 하드웨어가 필요하다. 한편, 스마트폰과 같은 모바일 단말기는 스피커 및 프로세서를 기본적으로 포함하므로, 음성 처리 장치(100)가 스마트폰과 같은 모바일 단말기에 구현된다면, 사용자들은 음성 처리 장치(100)를 이용하여 본 발명의 실시 예들에 따른 방법을 수행함으로써, 별도의 하드웨어를 구비하지 않더라도 화자들의 음성을 분리할 수 있고, 이들을 이용하여 음성에 대한 번역을 제공할 수 있는

효과가 있다.

- [109] 도 10 및 도 11은 본 발명의 실시 예들에 따른 음성 처리 장치의 동작을 설명하기 위한 도면이다. 도 10 및 도 11을 참조하면, 음성 처리 장치(100)는 위치 등록 모드(또는 화자 등록 모드)로 작동할 수 있다. 위치 등록 모드는 화자들(SPK1~SPK4)의 음성들의 음원 위치를 음성 처리 장치(100)에 기준 음원 위치로서 저장하는 모드를 의미한다. 이후, 음성 처리 장치(100)는 저장된 기준 음원 위치를 이용하여, 화자들(SPK1~SPK4)을 식별하여 분리 음성 신호를 생성하거나, 또는 특정 위치에서 발화된 음성과 연관된 분리 음성 신호만을 선택적으로 처리할 수도 있다.
- [110] 프로세서(130)는 외부로부터의 입력에 응답하여, 위치 등록 모드로 작동할 수 있다. 실시 예들에 따라, 프로세서(130)는 특정 문구를 포함하는 음성 신호에 응답하여 위치 등록 모드로 작동하거나, 또는, 음성 처리 장치(100)에 형성된 입력부(예컨대, 버튼 또는 터치 패널)를 통한 입력에 응답하여 위치 등록 모드로 작동할 수 있다.
- [111] 음성 처리 장치(100)는 화자들(SPK1~SPK4)의 음성에 응답하여 화자들(SPK1~SPK4) 각각의 음성에 대한 음원 위치를 결정하고, 음원 위치를 나타내는 음원 위치 정보를 생성할 수 있다.
- [112] 위치 등록 모드에서, 음성 처리 장치(100)는 생성된 음원 위치 정보를 기준 음원 위치 정보로서 메모리(140)에 저장할 수 있다.
- [113] 예컨대, 도 10에 도시된 바와 같이, 위치 등록 모드에서, 제1화자(SPK1)가 "저는 Alice 입니다"라고 발화하면, 음성 처리 장치(100)는 제1화자(SPK1)의 음성에 응답하여 음성 신호를 생성하고, 음성 신호로부터 제1화자(SPK1)의 위치인 제1위치(P1)를 결정할 수 있다. 음성 처리 장치(100)는 제1위치(P1)를 나타내는 제1위치 정보를 생성하고, 제1음원 위치 정보를 기준 음원 위치 정보로서 저장할 수 있다.
- [114] 마찬가지로, 예컨대, 도 11에 도시된 바와 같이, 음성 처리 장치(100)는 나머지 화자들(SPK2~SPK3)의 음성에 응답하여, 나머지 화자들(SPK2~SPK4)의 음성의 음원 위치(P2~P4)를 결정할 수 있다. 한편, 본 발명의 실시 예들에 따른 음성 처리 장치(100)는 화자들(SPK1~SPK4)의 음성이 시간적으로 중첩되어 발화되더라도, 화자들(SPK1~SPK4) 각각의 음성의 음원 위치를 계산할 수 있다.
- [115] 음성 처리 장치(100)는 제2화자(SPK2)의 위치인 제2위치(P2)를 나타내는 제2음원 위치 정보를 생성하고, 제2음원 위치 정보를 기준 음원 위치 정보로서 저장할 수 있고, 제3화자(SPK3)의 위치인 제3위치(P3)를 나타내는 제3음원 위치 정보를 생성하고, 제3음원 위치 정보를 기준 음원 위치 정보로서 저장할 수 있으며, 제4화자(SPK4)의 위치인 제4위치(P4)를 나타내는 제4음원 위치 정보를 생성하고, 제4음원 위치 정보를 기준 음원 위치 정보로서 저장할 수 있다.
- [116] 실시 예들에 따라, 음성 처리 장치(100)는 음원 위치 정보와, 음원 위치 정보에 대응하는 식별자를 저장할 수 있다. 상기 식별자는 음원 위치를 구별하기 위한

데이터로서, 예컨대, 해당하는 음원 위치에 위치한 화자를 나타내는 데이터(예컨대, 이름 등)일 수 있다.

- [117] 예컨대, 도 10에 도시된 바와 같이, 음성 처리 장치(100)는 제1화자(SPK1)의 음성에 응답하여, 제1화자(SPK1)를 나타내는 제1식별자(SID1)를 생성하고, 생성된 제1식별자(SID1)를 제1음원 위치 정보와 함께 매칭하여 저장할 수 있다. 즉, 제1식별자(SID1)는 제1화자(SPK1)를 식별하기 위한 수단이 될 수 있다. 예컨대, 음성 처리 장치(100)는 제1화자(SPK1)의 음성의 적어도 일부를 텍스트로 변환하고, 변환된 텍스트에 대응하는 제1식별자(SID1)를 생성할 수 있다. 예컨대, 음성 처리 장치(100)는 제1화자(SPK1)의 음성에 포함된 문구 중 적어도 일부를 제1식별자(SID1)로서 변환할 수 있다.
- [118] 예컨대, 도 11에 도시된 바와 같이, 음성 처리 장치(100)는 나머지 화자들(SPK2~SPK4)의 음성에 응답하여, 나머지 화자들(SPK2~SPK4)을 나타내는 식별자들(SID2~SID4)를 생성하고, 생성된 식별자들(SID2~SID4)을 화자들(SPK2~SPK4)의 음원 위치 정보와 함께 매칭하여 저장할 수 있다.
- [119] 도 12는 본 발명의 실시 예들에 따른 음성 처리 장치의 작동을 나타내는 도면이다. 도 12를 참조하면, 음성 처리 장치(100)는 음성 분리 모드로 작동할 수 있다.
- [120] 실시 예들에 따라, 프로세서(130)는 외부로부터의 입력에 응답하여, 음성 분리 모드로 작동할 수 있다. 실시 예들에 따라, 프로세서(130)는 특정 문구를 포함하는 음성 신호에 응답하여 음성 분리 모드로 작동하거나, 또는, 음성 처리 장치(100)에 형성된 입력부(예컨대, 버튼 또는 터치 패널)를 통한 입력에 응답하여 음성 분리 모드로 작동할 수 있다.
- [121] 음성 분리 모드에서, 음성 처리 장치(100)는 화자들(SPK1~SPK4)의 음성과 연관된 음성 신호를, 음성들의 음원 위치에 기초하여 분리함으로써, 화자들(SPK1~SPK4)의 음성과 관련된 분리 음성 신호를 생성하고, 생성된 분리 음성 신호를 저장할 수 있다.
- [122] 실시 예들에 따라, 음성 처리 장치(100)는 미리 저장된(또는 등록된) 기준 음원 위치와 대응하는 음원 위치에 대응하는 음성과 연관된 분리 음성 신호를 저장할 수 있다. 예컨대, 음성 처리 장치(100)는 음성 신호로부터 분리된 분리 음성 신호들 중에서, 기준 음원 위치로부터 기준 범위 이내에 있는 음원 위치와 대응하는 음성과 연관된 분리 음성 신호를 저장할 수 있다.
- [123] 실시 예들에 따라, 음성 분리 모드에서, 프로세서(130)는 인식되는 음성의 음원 위치가 미리 저장된(또는 등록된) 기준 음원 위치에 대응되지 않는 경우, 위치 등록 모드로 작동할 수 있다. 예컨대, 프로세서(130)는 인식되는 음성의 음원 위치가 미리 저장된 기준 음원 위치와 상이한 경우, 위치 등록 모드로 작동할 수 있고, 이에, 새로운 음원 위치를 등록할 수 있다.
- [124] 음성 처리 장치(100)는 화자들(SPK1~SPK4) 각각의 음성과 연관된 분리 음성 신호와, 대응하는 식별자를 매칭하여 저장할 수 있다. 예컨대, 도 12에 도시된

바와 같이, 음성 처리 장치(100)는 제1화자(SPK1)의 위치인 제1위치(P1)를 나타내는 제1음원 위치 정보에 따라, 제1분리 음성 신호와 제1식별자(SID1)를 매칭하여 저장할 수 있다. 예컨대, 음성 처리 장치(100)는 메모리(140)에 저장된 기준 음원 위치 정보를 참조하여, 제1음원 위치 정보에 대응하는 제1식별자(SID1)를 제1분리 음성 신호와 매칭하여 저장할 수 있다.

[125] 또한, 음성 처리 장치(100)는 분리 음성 신호와, 분리 음성 신호에 대응하는 음성이 수신된 시점을 추가적으로 매칭하여 저장할 수 있다.

[126] 도 13은 본 발명의 실시 예들에 따른 음성 처리 장치를 나타낸다. 도 13을 참조하면, 음성 처리 장치(100A)는 마이크(110), 통신 회로(120), 프로세서(130), 메모리(140) 및 트리거 신호 생성 회로(151)를 포함할 수 있다.

[127] 도 2와 비교할 때, 도 13의 음성 처리 장치(100A)는 트리거 신호 생성 회로(151)를 추가적으로 포함하는 차이가 있다. 이하, 설명의 편의상 중복되는 부분의 설명은 생략하고, 차이점에 대해 설명한다.

[128] 트리거 신호 생성 회로(151)는 외부로부터의 입력에 응답하여 트리거 신호들을 생성할 수 있다. 트리거 신호들은 프로세서(130)로 하여금 특정 동작을 수행하게끔 하는 신호일 수 있다. 트리거 신호들은 화자 등록 트리거 신호 및 화자 이동 트리거 신호를 포함할 수 있다. 이 때, 화자 등록 트리거 신호 및 화자 이동 트리거 신호를 생성하기 위한 입력 조건은 다를 수 있다.

[129] 실시 예들에 따라, 트리거 신호 생성 회로(151)는 터치 패널 또는 버튼과 같은 외부로부터의 물리적인 입력을 감지할 수 있는 입력부를 포함하고, 물리적인 입력에 응답하여 트리거 신호들을 생성할 수 있다. 예컨대, 트리거 신호 생성 회로(151)는 사용자의 터치가 감지되었을 때, 트리거 신호들을 생성할 수 있다.

[130] 실시 예들에 따라, 트리거 신호 생성 회로(151)는 음성 처리 장치(100A)에 의해 수신된 음성 신호들에 포함된 기동어를 인식하여 트리거 신호들을 생성할 수 있다. 예컨대, 트리거 신호 생성 회로(151)는 "화자 등록" 등의 특정 문구를 포함하는 음성 신호가 수신되면, 트리거 신호들을 생성할 수 있다.

[131] 트리거 신호 생성 회로(151)는 생성된 트리거 신호들을 프로세서(130)로 전송할 수 있다.

[132] 실시 예들에 따라, 프로세서(130)는 화자 식별 트리거 신호에 응답하여 화자 등록 모드(또는 위치 등록 모드)로 진입할 수 있다. 실시 예들에 따라, 화자 등록 모드는 화자 등록 트리거 신호가 수신된 시점으로부터 소정의 구간으로 정의되거나 또는 화자 등록 트리거 신호가 수신되는 동안의 구간으로 정의될 수 있으나, 이에 한정되는 것은 아니다.

[133] 도 10 및 도 11을 참조하여 설명한 바와 같이, 음성 처리 장치는 화자 등록 모드에서, 수신된 음성 신호들을 이용하여 기준 음원 위치 정보와 식별자를 생성하고, 또한 분리 음성 신호를 생성하고, 기준음원 위치 정보, 식별자 및 분리 음성 신호를 서로 매칭하여 저장할 수 있다.

[134] 도 14 및 도 15는 본 발명의 실시 예들에 따른 화자 이동 모드를 설명하기 위한

도면이다. 도 14 및 도 15를 참조하여 설명되는 화자 이동 모드는 도 13의 음성 처리 장치(100A)에 의해 수행될 수 있다.

[135] 도 14 및 도 15를 참조하면, 이동 전의 앨리스(Alice)의 위치는 "P1"이고, 화자 등록 모드 이후, 음성 처리 장치(100A)의 메모리(140)에는 앨리스를 식별하기 위한 식별자(SID)가 저장될 수 있고, 앨리스의 위치 "P1"가 기준 음원 위치로서 저장될 수 있다.

[136] 이동 후, 앨리스(Alice)가 위치(P5)에서 "저는 Alice 입니다"라는 음성을 발화한다. 음성 처리 장치(100A)는 화자 이동 모드에서, 앨리스의 음성과 연관된 음성 신호들을 이용하여 이동 후 앨리스의 위치 "P5"를 나타내는 음원 위치 정보를 새롭게 생성할 수 있다.

[137] 음성 처리 장치(100A)는 메모리(140)를 참조하여 화자 식별자(SID)에 매칭되어 저장된 기준 음원 위치 정보를 갱신할 수 있다. 예컨대, 음성 처리 장치(100A)는 화자 식별자(SID) "Alice"에 기존 매칭되어 저장된 기준 음원 위치 정보 "P1"을 이동 후 위치에 대한 기준 음원 위치 정보인 "P5"로 갱신할 수 있다.

[138] 이에 따라, 본 발명의 실시 예들에 따른 음성 처리 장치(100A)는 화자의 이동으로 인해 화자의 위치가 변경되더라도, 화자 식별자에 매칭되어 저장된 화자 위치 정보를 변경된 화자 위치 정보를 갱신할 수 있는 효과가 있다.

[139] 도 16은 본 발명의 실시 예들에 따른 음성 처리 장치를 나타낸다. 도 16을 참조하면, 음성 처리 장치(100B)는 마이크(110), 통신 회로(120), 프로세서(130), 메모리(140) 및 모션 센서(153)를 포함할 수 있다.

[140] 도 2와 비교할 때, 도 16의 음성 처리 장치(100B)는 모션 센서(153)를 추가적으로 포함하는 차이가 있다. 이하, 설명의 편의상 중복되는 부분의 설명은 생략하고, 차이점에 대해 설명한다.

[141] 모션 센서(153)는 음성 처리 장치(100B)와 관련된 물리량을 측정하고, 측정된 물리량에 해당하는 감지 신호를 생성할 수 있다. 예컨대, 모션 센서(153)는 음성 처리 장치(100B)의 위치 또는 움직임을 측정하고, 측정된 위치 또는 움직임에 대응하는 감지 신호를 생성 및 출력할 수 있다.

[142] 실시 예들에 따라, 모션 센서(153)는 음성 처리 장치(100B)의 위치를 측정하고, 음성 처리 장치(100B)의 위치를 나타내는 감지 신호를 출력할 수 있다. 예컨대, 모션 센서(153)는 GPS 센서, LIDAR(Light Detection And Ranging) 센서, 레이더(Radio Detection And Ranging) 센서 또는 UWB(Ultra Wide Band) 센서일 수 있으나, 본 발명의 실시 예들이 이에 한정되는 것은 아니다.

[143] 실시 예들에 따라, 모션 센서(153)는 음성 처리 장치(100B)의 움직임을 측정하고, 음성 처리 장치(100B)의 움직임을 나타내는 감지 신호를 출력할 수 있다. 예컨대, 모션 센서(153)는 자이로 센서, 속도 센서, 또는 가속도 센서일 수 있으나, 본 발명의 실시 예들이 이에 한정되는 것은 아니다.

[144] 한편, 본 명세서에서는 모션 센서(153)가 음성 처리 장치(100B)의 위치 또는 움직임을 측정하는 구성으로 설명되어 있으나, 실시 예들에 따라, 프로세서(130)

및 모션 센서(153)에 의해 음성 처리 장치(100B)의 위치 또는 움직임이 측정될 수도 있다. 예컨대, 모션 센서(153)는 음성 처리 장치(100B)의 위치 또는 움직임과 관련된 신호를 생성 및 출력하고, 프로세서(130)는 모션 센서(153)로부터 출력된 신호에 기초하여 음성 처리 장치(100B)의 위치 또는 움직임과 관련된 값을 생성할 수 있다.

- [145] 도 17 및 도 18은 본 발명의 실시 예들에 따른 음성 처리 장치의 작동을 나타내는 도면이다. 도 17 및 도 18을 참조하여 설명되는 작동은, 도 16를 참조하여 설명된 음성 처리 장치(100B)에 의해 수행될 수 있다.
- [146] 도 17 및 도 18을 참조하면, 음성 처리 장치(100B)는 음성 처리 장치(100B)의 움직임이 감지되는 경우, 변경된 화자들(SPK1~SPK4)의 음성에 대한 음원 위치를 기준 음원 위치 정보로서 저장할 수 있다.
- [147] 도 17에 도시된 바와 같이, 음성 처리 장치(100B)의 움직임에 따라 음성 처리 장치(100B)의 위치가 변화하는 경우, 화자들(SPK1~SPK4)의 음성 처리 장치(100B)에 대한 상대적인 위치가 달라질 수 있다. 나아가, 음성 처리 장치(100B)의 위치가 변화하지 않더라도, 음성 처리 장치(100B)의 움직임(회전, 진동 및 이동 등)이 발생하는 경우, 화자들(SPK1~SPK4)의 음성 처리 장치(100B)에 대한 상대적인 위치가 달라질 수 있다. 즉, 다시 말하면, 화자들(SPK1~SPK4)의 음성들의 음원 위치가 달라질 수 있다.
- [148] 예컨대, 제1화자(SPK1)의 위치는 P1에서 P5로 바뀔 수 있고, 제2화자(SPK2)의 위치는 P2에서 P6로 바뀔 수 있고, 제3화자(SPK3)의 위치는 P3에서 P7로 바뀔 수 있고, 제4화자(SPK4)의 위치는 P4에서 P8로 바뀔 수 있다.
- [149] 본 발명의 실시 예들에 따른 음성 처리 장치(100B)는 음성 처리 장치(100B)의 움직임을 감지할 수 있는 모션 센서(153)를 구비하고, 모션 센서(153)의 검출 결과를 통해 음성 처리 장치(100B)의 위치 변화를 감지할 수 있다. 또한, 음성 처리 장치(100B)는 음성 처리 장치(100B)의 움직임에 의해 변경된 음원 위치를 결정하고, 변경된 음원 위치를 기준 음원 위치 정보로서 저장할 수 있는 효과가 있다.
- [150] 도 18을 참조하면, 음성 처리 장치(100B)는 음성 처리 장치(100B)의 움직임이 감지되면, 위치 등록 모드로 동작할 수 있다. 실시 예들에 따라, 프로세서(130)는 모션 센서(153)의 검출 결과를 이용하여, 음성 처리 장치(100B)의 움직임을 감지할 수 있고, 위치 등록 모드로 동작할지 여부를 결정할 수 있다.
- [151] 즉, 음성 처리 장치(100B)는 위치 등록 모드에 따라 화자들(SPK1~SPK4) 각각의 음원 위치(즉, 화자들(SPK1~SPK4)의 위치))가 기준 음원 위치 정보로서 등록이 완료된 이후라 하더라도, 음성 처리 장치(100B)의 움직임이 감지되면 다시 위치 등록 모드로 작동할 수 있다.
- [152] 도 18에 도시된 바와 같이, 위치 갱신 모드에서, 제1화자(SPK1)가 변경된 위치 "P5"에서 "저는 Alice 입니다"라고 발화하면, 음성 처리 장치(100B)는 제1화자(SPK1)의 음성에 응답하여 음성 신호를 생성하고, 음성 신호로부터

변경된 음원 위치(즉, 제1화자(SPK1)의 변경된 위치)인 "P5"를 결정할 수 있다. 음성 처리 장치(100B)는 변경된 위치 "P5"를 나타내는 음원 위치 정보를 생성하고, 음원 위치 정보를 기준 음원 위치 정보로서 저장할 수 있다.

[153] 실시 예들에 따라, 음성 처리 장치(100B)는 화자들(SPK1~SPK4)의 각각의 변경된 위치를 나타내는 음원 위치 정보를 새롭게 기준 음원 위치 정보로서 저장하거나, 또는, 기존에 저장된 음원 위치 정보를 변경된 위치를 나타내는 음원 위치 정보로서 대체할 수 있다.

[154] 도 19는 본 발명의 실시 예들에 따른 음성 처리 장치의 작동 방법을 나타내는 플로우 차트이다. 도 19를 참조하여 설명될 음성 처리 장치의 작동 방법은 비일시적인 저장 매체에 저장되어, 컴퓨팅 장치에 의해 실행 가능한 프로그램으로서 구현될 수 있다.

[155] 도 19를 참조하여 설명될 작동 방법은 도 16을 참조하여 설명된 음성 처리 장치(100B)에 의해 수행될 수 있다.

[156] 음성 처리 장치(100B)는 화자들(SPK1~SPK4)의 음성에 응답하여, 화자들(SPK1~SPK4) 각각의 음성에 대한 음원 위치를 나타내는 음원 위치 정보를 생성할 수 있다(S310). 실시 예들에 따라, 음성 처리 장치(100B)는 화자들(SPK1~SPK4)의 음성에 응답하여 음성 신호를 생성하고, 음성 신호로부터 화자들(SPK1~SPK4) 각각의 음성에 대한 음원 위치를 나타내는 음원 위치 정보를 생성할 수 있다. 이 때, 음원 위치는 곧 화자들(SPK1~SPK4) 각각의 위치를 나타낸다.

[157] 음성 처리 장치(100B)는 생성된 음원 위치 정보를 기준 음원 위치 정보로서 저장할 수 있다(S320). 실시 예들에 따라, 음성 처리 장치(100B)는 생성된 음원 위치 정보를 메모리(140)에 기준 음원 위치 정보로서 저장할 수 있다.

[158] 음성 처리 장치(100B)는 음성 처리 장치(100B)의 움직임 감지할 수 있다(S330). 실시 예들에 따라, 음성 처리 장치(100B)는 모션 센서(153)를 이용하여, 음성 처리 장치(100B)의 움직임을 감지할 수 있다. 예컨대, 음성 처리 장치(100B)는 모션 센서(153)를 이용하여, 음성 처리 장치(100B)의 위치의 변화, 각도의 변화 또는 속도 및 가속도의 변화를 감지할 수 있다.

[159] 음성 처리 장치(100B)는 감지된 움직임이 기준 움직임을 초과하는지 여부를 판단할 수 있다(S340). 실시 예들에 따라, 음성 처리 장치(100B)는 모션 센서(153)를 이용하여 감지된 물리량이, 미리 지정된 기준 물리량을 초과하는지 판단할 수 있다. 예컨대, 음성 처리 장치(100B)는 음성 처리 장치(100B)의 주기적으로 측정된 위치의 변화가 기준 값을 초과하는지 판단하거나, 또는, 음성 처리 장치(100B)의 가속도가 기준 값을 초과하는지 판단함으로써, 움직임이 기준 움직임을 초과하는지 여부를 판단할 수 있다.

[160] 음성 처리 장치(100B)는 감지된 움직임이 기준 움직임을 초과하는 경우(S340의 Y), 음성 처리 장치(100B)는 화자들(SPK1~SPK4)의 음성에 응답하여 음원 위치 정보를 생성하고, 생성된 음원 위치 정보를 기준 음원 위치 정보로서 저장할 수

있다. 즉, 감지된 움직임이 기준 움직임을 초과하는 경우, 음성 처리 장치(100B)는 화자들(SPK1~SPK4)의 음성의 음원 위치를 다시 결정하고, 변경된 음원 위치를 나타내는 음원 위치 정보를 기준 음원 위치 정보로서 다시 저장할 수 있다. 이를 통해, 음성 처리 장치(100B)의 움직임으로 인해 화자들(SPK1~SPK4)의 상대적인 위치가 변화하더라도, 기준 음원 위치 정보가 갱신될 수 있다. 이에 따라, 음성 처리 장치(100B)의 움직임으로 인한 화자들(SPK1~SPK4)의 상대적인 위치의 변화에 의한 오차가 최소화될 수 있다.

- [161] 도 20은 본 발명의 실시 예들에 따른 음성 처리 장치를 나타낸다. 도 20을 참조하면, 음성 처리 장치(100C)는 마이크(110), 통신 회로(120), 프로세서(130), 메모리(140) 및 발광 장치(155)를 포함할 수 있다.
- [162] 도 2와 비교할 때, 도 20의 음성 처리 장치(100C)는 발광 장치(155)를 추가적으로 포함하는 차이가 있다. 이하, 설명의 편의상 중복되는 부분의 설명은 생략하고, 차이점에 대해 설명한다.
- [163] 발광 장치(150)는 프로세서(130)의 제어에 따라, 빛을 발광할 수 있다. 실시 예들에 따라, 발광 장치(150)는 발광 소자들을 포함하고, 발광 소자들은 전기적인 신호에 따라 특정 파장의 빛을 방출할 수 있다. 예컨대, 발광 장치(150)는 발광 다이오드, LCD(liquid crystal display), OLED(organic lighting emitting diode) 발광 장치, 플렉서블(flexible) 발광 장치, 마이크로 LED 발광 장치 또는 퀀텀닷(quantum dot) 발광 장치일 수 있으나, 본 발명의 실시 예들이 이에 한정되는 것은 아니다.
- [164] 실시 예들에 따라, 발광 장치(150)는 프로세서(130)의 제어에 따라 작동할 수 있다. 예컨대, 발광 장치(150)는 프로세서(130)로부터 전송되는 제어 신호에 기초하여, 특정한 시각적 패턴을 표시할 수 있다.
- [165] 도 21은 본 발명의 실시 예들에 따른 음성 처리 장치를 나타낸다. 도 21을 참조하면, 음성 처리 장치(100C)는 발광 장치(150)를 포함할 수 있다.
- [166] 본 발명의 실시 예들에 따르면, 발광 장치(150)는 복수의 발광 소자들(LED1~LEDn; n은 2이상의 자연수)을 포함할 수 있다. 실시 예들에 따라, 복수의 발광 소자들(LED1~LEDn)은 음성 처리 장치(100C)의 표면에 배치될 수 있으나, 본 발명의 실시 예들이 이에 한정되는 것은 아니며, 복수의 발광 소자들(LED1~LEDn)은 음성 처리 장치(100)의 부분 중 육안으로 보이는 부분에 배치될 수 있다.
- [167] 예컨대, 도 21에 도시된 바와 같이, 음성 처리 장치(100C)는 원형의 단면을 가지는 형태로 구현되고, 복수의 발광 소자들(LED1~LEDn)은 음성 처리 장치(100C)의 표면의 둘레를 따라 연속적으로 배치될 수 있으나, 이에 한정되는 것은 아니다.
- [168] 복수의 발광 소자들(LED1~LEDn) 각각은 서로 다른 위치에 배치될 수 있다.
- [169] 후술하는 바와 같이, 음성 처리 장치(100C)는 화자들(SPK1~SPK4)의 음성에 응답하여 화자들(SPK1~SPK4)의 위치를 판단할 수 있고, 발화하는 화자의

위치에 대응하는 시각적 패턴을 발광 장치(150)를 통해 표시할 수 있다. 예컨대, 음성 처리 장치(100C)는 복수의 발광 소자들(LED1~LEDn) 중에서 발화하는 화자의 위치에 대응하는 발광 소자를 턴-온 할 수 있다. 이에 따라, 사용자는 발광 장치(100C)에 배치된 발광 소자들(LED1~LEDn) 중에서 발광하는 발광 소자들의 위치를 통해, 현재 발화하고 있는 화자(SPK1~SPK4)의 위치를 파악할 수 있는 효과가 있다.

[170] 예컨대, 발광 소자들(LED1~LEDn) 각각은 특정 위치를 나타낼 수 있다.

[171] 도 22 및 도 23은 본 발명의 실시 예들에 따른 음성 처리 장치의 동작을 설명하기 위한 도면이다. 도 22 및 도 23을 참조하면, 본 발명의 실시 예들에 따른 음성 처리 장치(100C)는 화자들(SPK1~SPK4)의 음성에 응답하여, 화자들(SPK1~SPK4) 각각의 음성의 위치를 판단하고, 판단된 위치에 따라 각 위치에 대응하는 시각적 패턴을 출력할 수 있다.

[172] 한편, 도 22 및 도 23을 참조하여 설명될 실시 예에서는 음성 처리 장치(100C)가 복수의 발광 소자들(LED1~LED8)을 이용하여 화자들(SPK1~SPK4)의 위치에 대응하는 시각적 패턴을 출력하는 것을 가정하고 설명한다. 다만, 실시 예들에 따라, 음성 처리 장치(100C)는 다른 시각적 표현 방식에 따라 화자들(SPK1~SPK4)의 위치에 대응하는 시각적 패턴을 출력할 수 있다.

[173] 음성 처리 장치(100C)는 도 3 내지 도 5를 참조하여 설명된 실시 예들에 따라, 화자들(SPK1~SPK4)의 음성으로부터 음성들의 음원 위치(즉, 화자들(SPK1~SPK4)의 위치)를 결정할 수 있다.

[174] 음성 처리 장치(100C)는 발광 소자들(LED1~LED8)를 구별하기 위한 식별자와, 발광 소자들(LED1~LED8) 각각에 대응하는 위치에 관한 정보를 저장할 수 있다. 예컨대, 도 22 및 도 23에 도시된 바와 같이, 제2발광 소자(LED2)에 대응하는 위치는 제2위치(P2)이다. 이 때, 발광 소자들(LED1~LED8) 각각에 대응하는 위치는 발광 소자들(LED1~LED8) 각각의 실제 위치일 수도 있지만, 실제 위치와 무관한 미리 지정된 위치일 수 있다.

[175] 본 발명의 실시 예들에 따르면, 음성 처리 장치(100C)는 화자들(SPK1~SPK4)의 음성에 응답하여, 화자들(SPK1~SPK4) 각각의 음성과 연관된 음원 위치를 결정하고, 발광 소자(LED1~LED8) 중 결정된 음원 위치에 대응하는 위치에 배치된 발광 소자를 작동할 수 있다.

[176] 예컨대, 도 22에 도시된 바와 같이, 제1위치(P1)에 위치한 제1화자(SPK1)가 발화하면, 음성 처리 장치(100C)는 제1화자(SPK1)의 음성으로부터 제1화자(SPK1)의 위치(즉, 음원 위치)를 판단하고, 제1화자(SPK1)의 위치인 제1위치(P1)에 대응하는 발광 소자를 작동시킬 수 있다. 제1위치(P1)에 대응하는 발광 소자는 제8발광 소자(LED8)이므로, 음성 처리 장치(100C)는 제8발광 소자(LED8)를 턴-온할 수 있다. 예컨대, 프로세서(130)는 제8발광 소자(LED8)을 턴-온시키기 위한 제어 신호를 출력할 수 있다.

[177] 마찬가지로, 예컨대, 도 23에 도시된 바와 같이, 제2위치(P2)에 위치한

제2화자(SPK2)가 발화하면, 음성 처리 장치(100C)는 제2발광 소자(LED2)를 턴-온할 수 있다. 예컨대, 프로세서(130)는 제2발광 소자(LED2)를 턴-온시키기 위한 제어 신호를 출력할 수 있다.

[178] 음성 처리 장치(100C)는 화자들(SPK1~SPK4) 각각의 음성이 인식되는 시점에, 화자들(SPK1~SPK4) 각각의 위치에 대응하는 발광 소자를 턴-온할 수 있다. 실시 예들에 따라, 음성 처리 장치(100C)는 화자들(SPK1~SPK4) 각각의 음성이 인식되는 동안 발광 소자를 턴-온하고, 화자들(SPK1~SPK4) 각각의 음성이 인식되지 않을 때 발광 소자를 턴-오프할 수 있다.

[179] 본 발명의 실시 예들에 따른 음성 처리 장치(100C)는 복수의 발광 소자들(LED1~LEDn) 중에서 발화하는 화자의 위치에 대응하는 발광 소자를 턴-온할 수 있다. 이에 따라, 사용자는 발광 장치(100C)에 배치된 발광 소자들(LED1~LEDn) 중에서 발광하는 발광 소자들의 위치를 통해, 현재 발화하고 있는 화자(SPK1~SPK4)의 위치를 파악할 수 있는 효과가 있다.

[180] 도 24는 본 발명의 실시 예들에 따른 음성 처리 장치의 작동 방법을 나타내는 플로우 차트이다. 도 24를 참조하여 설명될 음성 처리 장치의 작동 방법은 비일시적인 저장 매체에 저장되어, 컴퓨팅 장치에 의해 실행 가능한 프로그램으로서 구현될 수 있다.

[181] 도 24를 참조하면, 음성 처리 장치(100C)는 음성에 응답하여, 음성 신호를 생성할 수 있다(S410). 실시 예들에 따라, 음성 처리 장치(100C)는 공간에서 감지되는 음성을 전기적인 신호인 음성 신호로 변환할 수 있다.

[182] 음성 처리 장치(100C)는 화자들(SPK1~SPK4)의 음성과 연관된 음성 신호를 이용하여, 음성들 각각에 대한 음원 위치(즉, 화자들(SPK1~SPK4)의 위치)를 판단할 수 있다(S420). 실시 예들에 따라, 음성 처리 장치(100C)는 화자들(SPK1~SPK4)의 음성들 각각에 대한 음원 위치(즉, 화자들(SPK1~SPK4)의 위치)를 나타내는 음원 위치 정보를 생성할 수 있다.

[183] 음성 처리 장치(100C)는 음성들 각각에 대한 음원 위치에 기초하여, 음원 위치에 대응하는 시각적 패턴을 표시할 수 있다(S430).

[184] 실시 예들에 따라, 음성 처리 장치(100C)는 복수의 발광 소자들(LED1~LEDn)을 포함하는 발광 장치(150)를 포함하고, 복수의 발광 소자들(LED1~LEDn) 중에서 음성의 음원 위치에 대응하는 발광 소자를 턴-온할 수 있다.

[185] 또한, 실시 예들에 따라, 음성 처리 장치(100C)는 디스플레이 장치로 구현되는 발광 장치(150)를 포함할 수 있고, 발광 장치(150)는 화자들(SPK1~SPK4)의 음원 위치를 가리키는 시각적 패턴을 표시할 수 있다. 예컨대, 발광 장치(150)는 화살표, 직선 또는 손가락 등의 도형을 표시함으로써, 화자들(SPK1~SPK4)의 음원 위치를 가리키는 시각적 패턴을 표시할 수 있다.

[186] 도 25는 본 발명의 실시 예들에 따른 음성 처리 장치의 동작을 설명하기 위한 도면이다. 도 25를 참조하면, 음성 처리 장치(100C)는 위치 등록 모드로 작동할 수 있다.

- [187] 도 10 및 도 11을 참조하여 설명된 바와 같이, 위치 등록 모드에서, 음성 처리 장치(100C)는 생성된 음원 위치 정보를 기준 음원 위치 정보로서 메모리(140)에 저장할 수 있다.
- [188] 위치 등록 모드에서, 음성 처리 장치(100C)는 화자들(SPK1~SPK4)의 음성에 응답하여, 화자들(SPK1~SPK4) 각각의 음성의 음원 위치를 판단하고, 판단된 음원 위치에 대응하는 시각적 패턴을 출력할 수 있다. 실시 예들에 따라, 음성 처리 장치(100C)는 음원 위치 정보가 기준 음원 위치 정보로서 메모리(140)에 저장 완료되면, 저장 완료된 음원 위치 정보에 대응하는 시각적 패턴을 출력할 수 있다.
- [189] 예컨대, 도 25에 도시된 바와 같이, 음성 처리 장치(100C)는 제1위치(P1)를 나타내는 제1음원 위치 정보가 기준 음원 위치 정보로서 저장되면, 복수의 발광 소자들(LED1~LED8) 중에서 제1위치(P1)에 대응하는 제8발광 소자(LED8)를 턴-온 할 수 있다. 또한, 음성 처리 장치(100C)는 나머지 화자들(SPK2~SPK4)의 위치들(P2~P4)을 나타내는 음원 위치 정보가 기준 음원 위치 정보로서 저장되면, 복수의 발광 소자들(LED1~LED8) 중에서 제2위치(P2)에 대응하는 제2발광 소자(LED2), 제3위치(P3)에 대응하는 제6발광 소자(LED6) 및 제4위치(P4)에 대응하는 제4발광 소자(LED4)를 턴-온 할 수 있다.
- [190] 이에 따라, 화자들(SPK1~SPK4)은 기준 위치로서 등록된 음원 위치가 어디인지 쉽게 파악할 수 있는 효과가 있다.
- [191] 도 26은 본 발명의 실시 예들에 따른 음성 처리 장치의 작동을 나타내는 도면이다. 도 26을 참조하면, 음성 처리 장치(100C)는 음성 분리 모드로 작동할 수 있다.
- [192] 도 12를 참조하여 설명된 바와 같이, 음성 분리 모드에서, 음성 처리 장치(100C)는 화자들(SPK1~SPK4)의 음성과 연관된 음성 신호를, 음성들의 음원 위치에 기초하여 분리함으로써, 화자들(SPK1~SPK4)의 음성과 관련된 분리 음성 신호를 생성하고, 생성된 분리 음성 신호를 저장할 수 있다.
- [193] 음성 분리 모드에서, 음성 처리 장치(100C)는 화자들(SPK1~SPK4)의 음성에 응답하여, 화자들(SPK1~SPK4) 각각의 음성의 음원 위치를 판단하고, 인식되는 음성의 음원 위치에 대응하는 시각적 패턴을 출력할 수 있다. 예컨대, 도 26에 도시된 바와 같이, 음성 처리 장치(100C)는 복수의 발광 소자들(LED1~LED8) 중에서 발화하는 화자의 위치에 대응하는 발광 소자(제8발광 소자(LED8), 제2발광 소자(LED2) 및 제6발광 소자(LED6))를 턴-온 할 수 있다.
- [194] 이에 따라, 화자들(SPK1~SPK4)은 현재 발화되는 음성의 음원 위치가 어디인지 쉽게 파악할 수 있는 효과가 있다.
- [195] 실시 예들에 따라, 음성 처리 장치(100C)는 음성 분리 모드에서, 위치 등록 모드에서와 다른 표시 방식에 따라 음원 위치에 대응하는 시각적 패턴을 출력할 수 있다. 예컨대, 음성 처리 장치(100C)는 위치 등록 모드에서 제1표시 방식에 따라 음원 위치에 대응하는 시각적 패턴을 출력하고, 음성 분리 모드에서 상기

제1표시 방식과 다른 제2표시 방식에 따라 음원 위치에 대응하는 시각적 패턴을 출력할 수 있다. 상기 표시 방식은 시각적 패턴의 출력 색상, 출력 시간, 출력 주기 등을 의미할 수 있다.

- [196] 예컨대, 위치 등록 모드에서, 음성 처리 장치(100C)는 음원 위치 정보가 기준 음원 위치 정보로서 메모리(140)에 저장 완료되면, 저장 완료된 음원 위치 정보에 대응하는 시각적 패턴을 출력할 수 있고, 음성 분리 모드에서, 음성 처리 장치(100C)는 화자들(SPK1~SPK4)의 음성이 인식되는 동안, 인식된 음성의 음원 위치 정보에 대응하는 시각적 패턴을 출력할 수 있다.
- [197] 도 27은 본 발명의 실시 예들에 따른 음성 처리 장치의 동작을 설명하기 위한 도면이다. 도 27의 동작 방법은 음성 처리 장치(100, 100A, 100B, 100C)에 의해 수행될 수 있다.
- [198] 음성 처리 장치(100)는 화자들(SPK1~SPK4) 각각의 음성과 연관된 분리 음성 신호들을 이용하여 회의록(MIN)을 생성할 수 있다. 생성된 회의록(MIN)은 문서 파일, 이미지 파일 또는 음성 파일의 형태로 저장될 수 있으나, 이에 한정되는 것은 아니다.
- [199] 음성 처리 장치(100)는 서로 매칭되어 저장된 분리 음성 신호에 기초하여, 화자들(SPK1~SPK4) 각각의 음성을 나타내는 데이터를 생성할 수 있고, 생성된 화자들(SPK1~SPK4) 각각의 음성을 나타내는 데이터를 이용하여 회의록(MIN)을 생성할 수 있다. 실시 예들에 따라, 음성 처리 장치(100)는 화자들(SPK1~SPK4) 각각의 음성이 인식된 시점에 따라, 각 화자의 음성과 연관된 데이터를 시간 순으로 정렬하여 회의록(MIN)을 생성할 수 있다.
- [200] 실시 예들에 따라, 음성 처리 장치(100)는 화자들(SPK1~SPK4)을 식별하기 위한 식별자들을 이용하여, 회의록(MIN)에 특정 음성을 발화한 화자(SPK1~SPK4)를 나타내는 식별자를 함께 표시할 수 있다. 이를 통해, 회의록(MIN)에서의 발언들이 화자별로 구별될 수 있다.
- [201] 도 27에 도시된 바와 같이, 화자들(SPK1~SPK4)가 순차적으로 "AAA1", "BBB2", "AAA3", "CCC4", "DDD5", "CCC6" 및 "BBB7"을 발화한다. 상술한 바와 같이, 음성 처리 장치(100)는 "AAA1" 및 "AAA3"에 대응하는 제1분리 음성 신호와 제1화자(SPK1)를 나타내는 제1식별자(SID1)를 매칭하여 저장하고, "BBB2" 및 "BBB7"에 대응하는 제2분리 음성 신호와 제2식별자(SID2)를 매칭하여 저장하고, "CCC4" 및 "CCC6"에 대응하는 제3분리 음성 신호와 제3식별자(SID3)를 매칭하여 저장하고, "DDD5"에 대응하는 제4분리 음성 신호와 제4식별자(SID4)를 매칭하여 저장할 수 있다.
- [202] 이상과 같이 실시 예들이 비록 한정된 실시 예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여

대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

[203] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

[204]

[205]

[206]

산업상 이용가능성

[207] 본 발명의 실시 예들은 음성을 처리하기 위한 장치 및 이의 작동 방법에 관한 것이다.

청구범위

- [청구항 1] 화자들의 음성들과 연관된 음성 신호를 상기 음성들 각각의 음원 위치에 기초하여 음원 분리를 수행하도록 구성되는 프로세서; 및
메모리를 포함하고,
상기 프로세서는,
상기 음성들과 연관된 음성 신호를 이용하여 상기 음성들 각각의 음원 위치를 나타내는 음원 위치 정보를 생성하고,
상기 음원 위치 정보에 기초하여, 상기 음성 신호로부터 상기 화자들 각각의 음성과 연관된 분리 음성 신호를 생성하고,
상기 분리 음성 신호와 상기 음원 위치 정보를 서로 매칭하여 상기 메모리에 저장하도록 구성되는,
음성 처리 장치.
- [청구항 2] 제1항에 있어서, 상기 음성 처리 장치는,
상기 화자들의 음성들에 응답하여 상기 음성 신호를 생성하도록 구성되는 마이크를 더 포함하는,
음성 처리 장치.
- [청구항 3] 제2항에 있어서,
상기 마이크는 어레이를 이루도록 배치된 복수의 마이크들을 포함하고,
상기 복수의 마이크들은 상기 음성들에 응답하여 상기 음성 신호를 생성하도록 구성되는,
음성 처리 장치.
- [청구항 4] 제3항에 있어서, 상기 프로세서는,
상기 복수의 마이크들로부터 생성된 복수의 음성 신호들 사이의 시간 지연에 기초하여, 상기 음성들 각각의 음원 위치를 판단하고,
판단된 음원 위치에 기초하여, 상기 분리 음성 신호를 생성하는,
음성 처리 장치.
- [청구항 5] 제1항에 있어서,
상기 메모리는 상기 화자들의 음성들의 발화된 언어인 출발 언어를 나타내는 출발 언어 정보를 저장하고,
상기 프로세서는 상기 출발 언어 정보와 상기 분리 음성 신호에 기초하여,
상기 화자들의 음성의 언어가 상기 출발 언어로부터 번역될 언어인 도착 언어로 번역된 번역 결과를 출력하는,
음성 처리 장치.
- [청구항 6] 제5항에 있어서,
상기 프로세서는 상기 출발 언어 정보에 기초하여, 상기 음성들 각각의 음원 위치에 따라 상기 음성들의 위치에 대응하는 출발 언어를 결정하고,
결정된 출발 언어에 따라 상기 음성들 각각에 대한 번역 결과를 출력하는,

- 음성 처리 장치.
- [청구항 7] 제1항에 있어서, 상기 프로세서는,
위치 등록 모드에서,
상기 음성 신호를 이용하여 상기 음성들 각각의 음원 위치를 나타내는
음원 위치 정보를 생성하고, 상기 음원 위치 정보를 기준 음원 위치
정보로서 상기 메모리에 저장하고,
음성 분리 모드에서,
상기 기준 음원 위치로부터 기준 범위 이내에 있는 음원 위치와 대응하는
음성과 연관된 분리 음성 신호를 상기 메모리에 저장하는,
음성 처리 장치.
- [청구항 8] 제7항에 있어서, 상기 음성 처리 장치는,
화자 등록 트리거 신호를 생성하도록 구성되는 트리거 신호 생성 회로를
더 포함하고,
상기 프로세서는 상기 화자 등록 트리거 신호에 응답하여, 상기 위치 등록
모드에서 작동하는,
음성 처리 장치.
- [청구항 9] 제7항에 있어서,
상기 메모리는 상기 화자들 각각을 식별하기 위한 식별자를 저장하고,
상기 식별자는 상기 기준 음원 위치 정보와 서로 매칭되어 저장된,
음성 처리 장치.
- [청구항 10] 제9항에 있어서, 상기 프로세서는,
상기 분리 음성 신호와 대응하는 음원 위치 정보와 상기 기준 음원 위치
정보를 비교하고, 상기 분리 음성 신호와 대응하는 음원 위치 정보와
대응하는 식별자를 결정하고, 상기 분리 음성 신호와 상기 식별자를 서로
매칭하여 상기 메모리에 저장하는,
음성 처리 장치.
- [청구항 11] 제8항에 있어서,
상기 트리거 신호 생성 회로는 화자 이동 트리거 신호를 생성하고,
상기 프로세서는,
상기 화자 이동 트리거 신호에 응답하여, 상기 음성 신호들을 이용하여
화자 위치 정보와 화자 식별자를 생성하고, 상기 화자 식별자와 매칭된
기준 화자 식별자를 결정하고, 상기 기준 화자 식별자와 매칭되어 저장된
기준 화자 위치 정보를 생성된 화자 위치 정보로 갱신하는,
음성 처리 장치.
- [청구항 12] 제8항에 있어서,
상기 음성 처리 장치는 상기 음성 처리 장치의 움직임 감지하도록
구성되는 모션 센서를 더 포함하고,
상기 프로세서는,

상기 센서에 의해 감지된 상기 음성 처리 장치의 움직임이 기준 움직임을 초과하는지 판단하고,

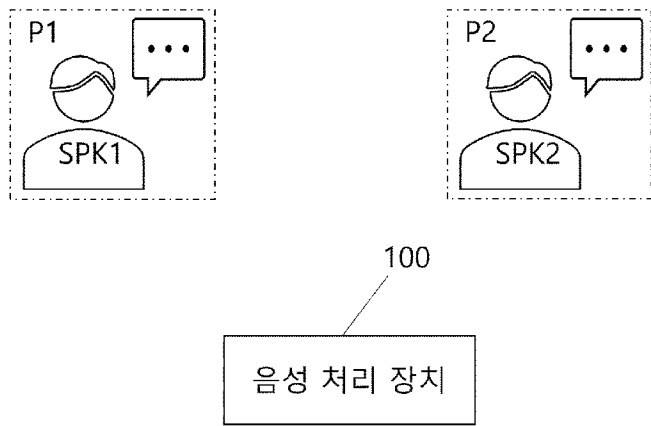
상기 음성 처리 장치의 움직임이 기준 움직임을 초과할 때, 상기 화자들의 음성들에 기초하여, 상기 화자들의 음성들의 변경된 음원 위치를 나타내는 음원 위치 정보를 생성하고, 변경된 음원 위치를 나타내는 음원 위치 정보를 상기 기준 음원 위치 정보로서 상기 메모리에 저장하는, 음성 처리 장치.

[청구항 13] 제1항에 있어서,
상기 음성 처리 장치는 상기 프로세서의 제어에 따라 발광하도록 구성되는 발광 장치를 더 포함하고,
상기 프로세서는,
상기 발광 장치를 통해 상기 음원 위치에 대응하는 시각적 패턴이 표시되도록, 상기 발광 장치를 제어하기 위한 발광 제어 신호를 출력하는, 음성 처리 장치.

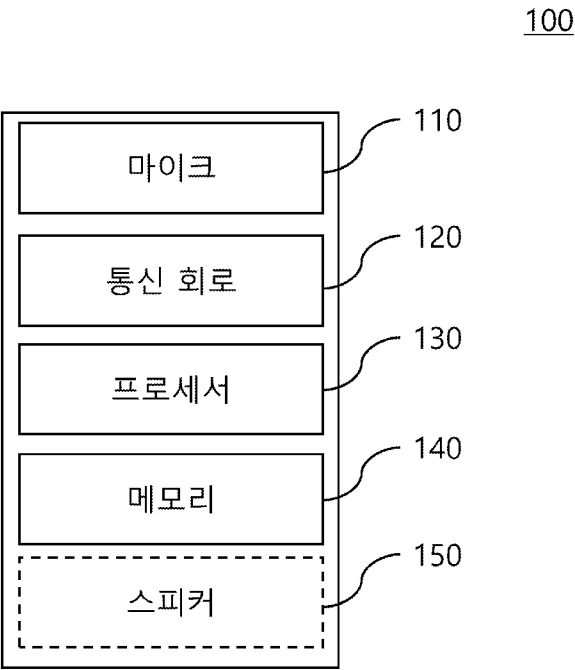
[청구항 14] 제13항에 있어서,
상기 발광 장치는 각각이 빛을 발광하도록 구성되는 복수의 발광 소자들을 포함하고,
상기 프로세서는,
상기 복수의 발광 소자들 중에서 결정된 음원 위치에 대응하는 발광 소자를 선택적으로 턴-온시키기 위한 상기 발광 제어 신호를 출력하는, 음성 처리 장치.

[청구항 15] 제14항에 있어서,
상기 메모리는 상기 발광 소자들 각각의 식별자 및 위치를 나타내는 정보를 저장하고,
상기 프로세서는,
상기 메모리를 참조하여, 상기 복수의 발광 소자들 중에서 결정된 음원 위치에 대응하는 발광 소자의 식별자를 리드하고, 리드된 식별자를 이용하여 결정된 음원 위치에 대응하는 발광 소자를 선택적으로 턴-온시키기 위한 상기 발광 제어 신호를 출력하는, 음성 처리 장치.

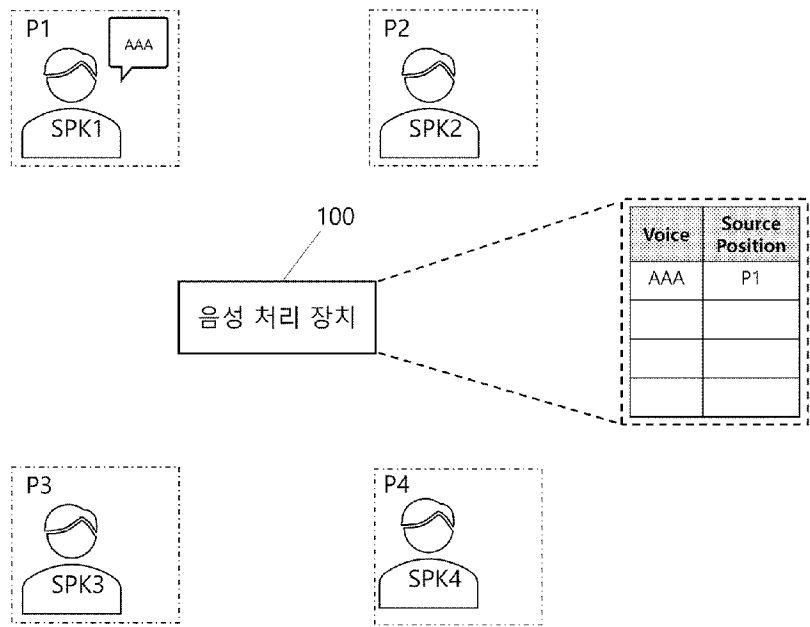
[도1]



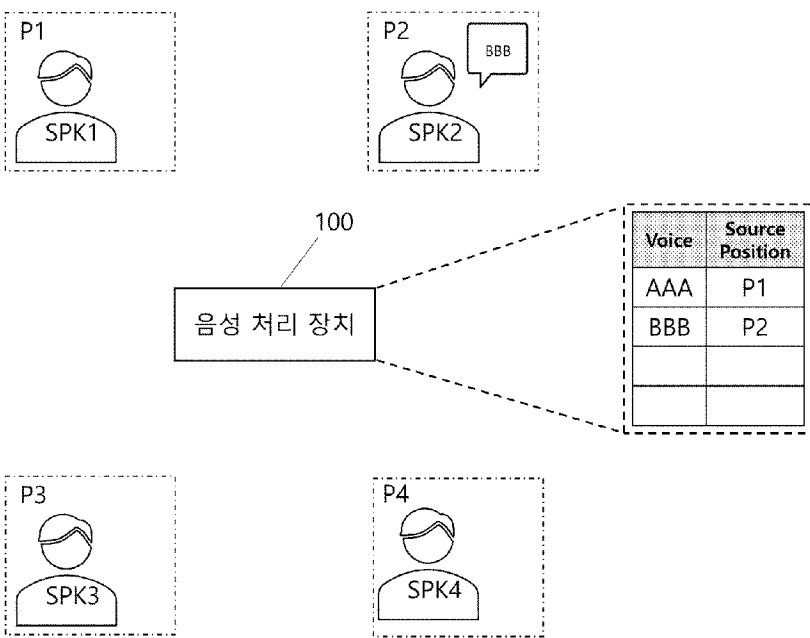
[도2]



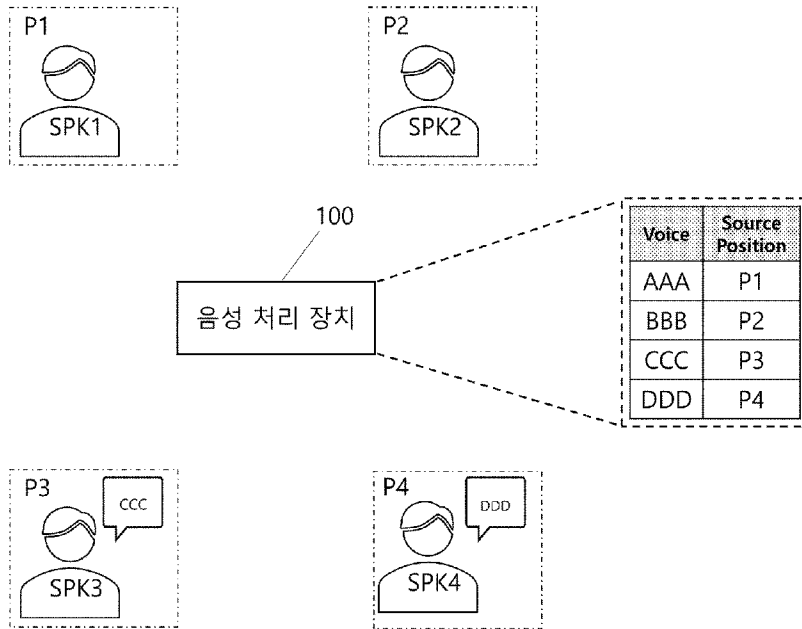
[도3]



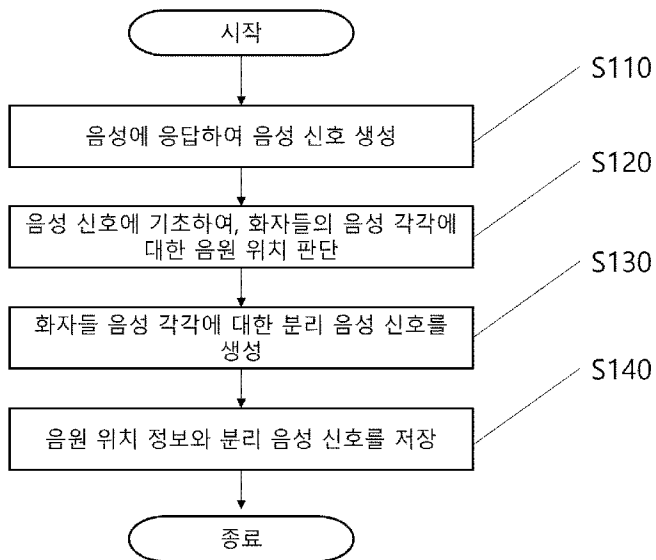
[도4]



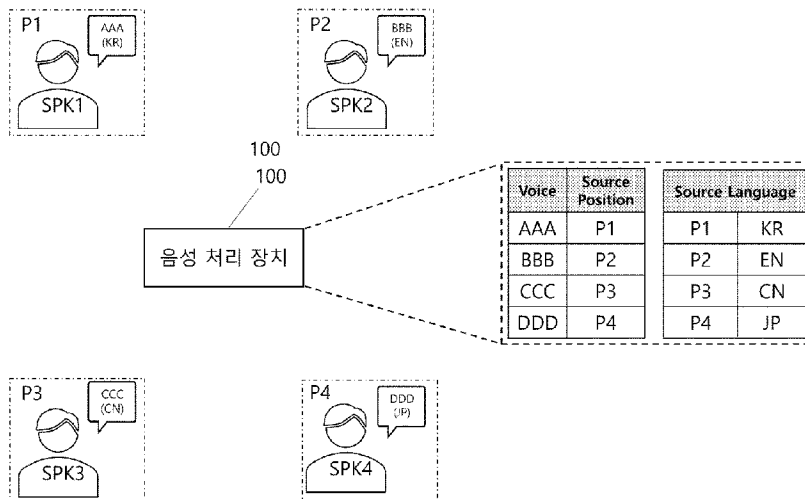
[도5]



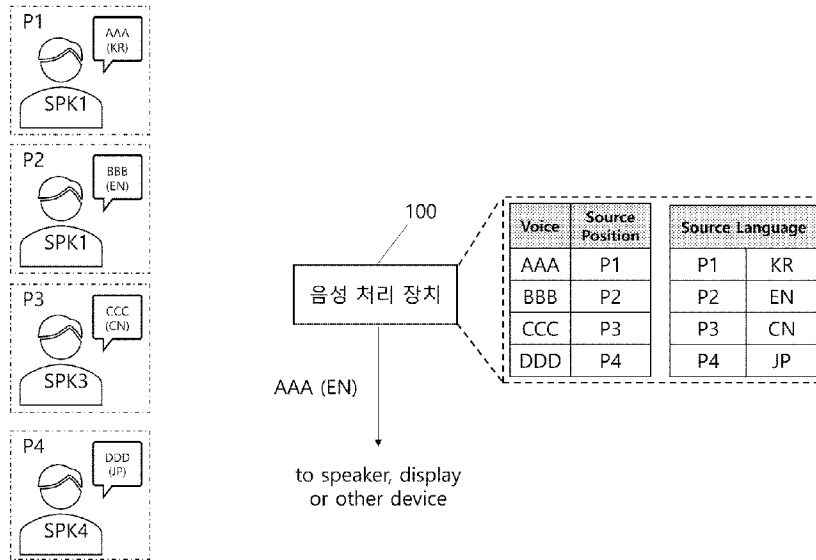
[도6]



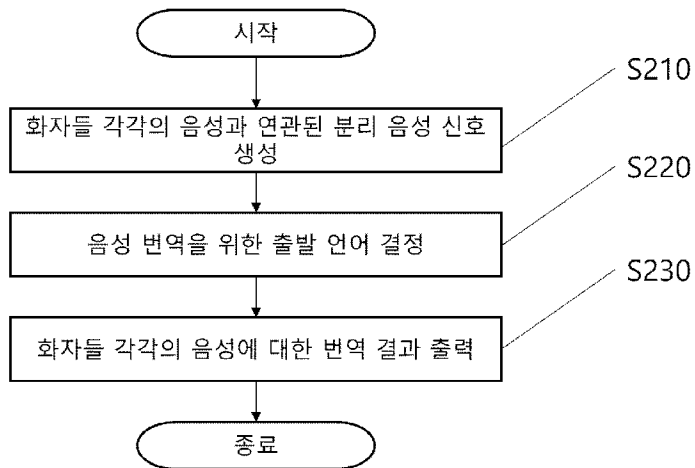
[도7]



[도8]

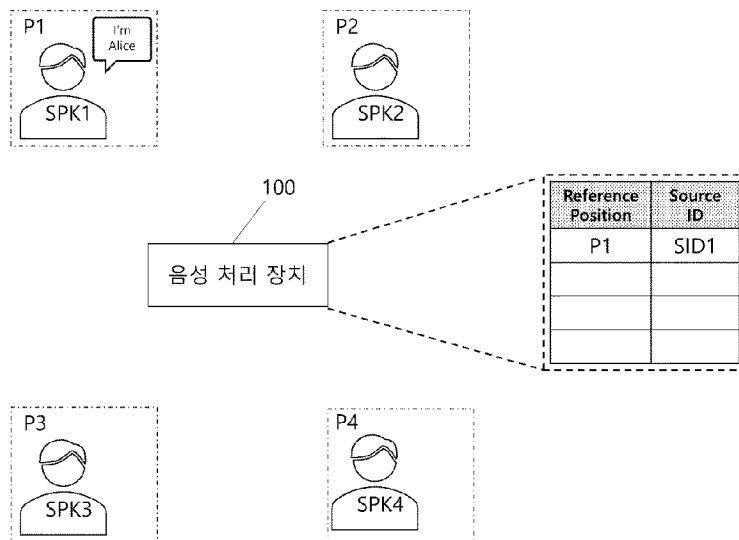


[도9]



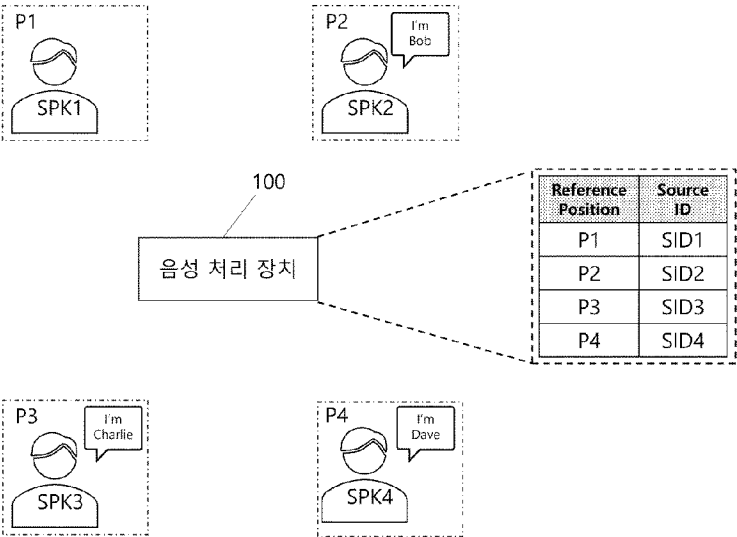
[도10]

REGISTER MODE



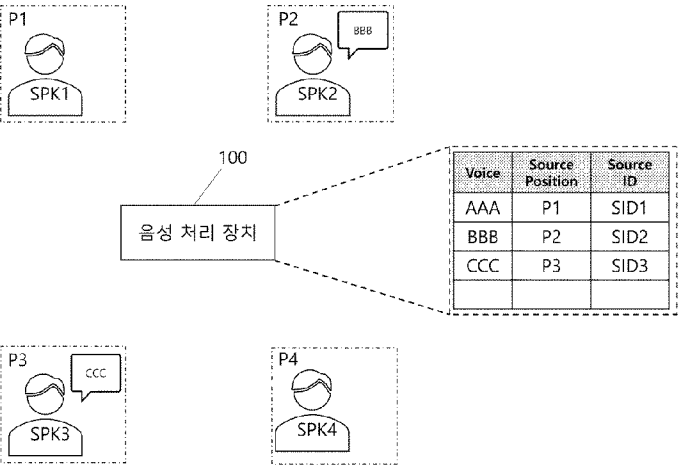
[도11]

REGISTER MODE



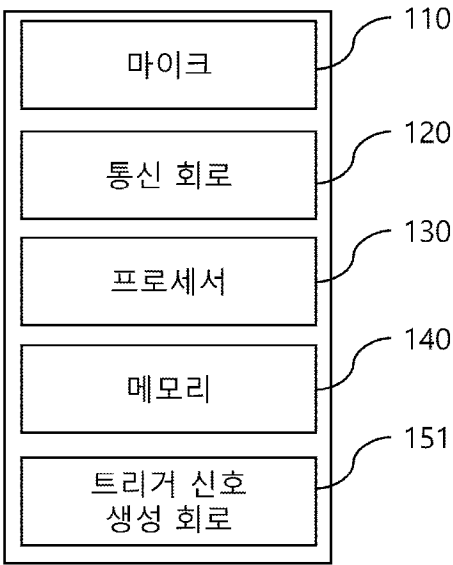
[도12]

SEPARATION MODE

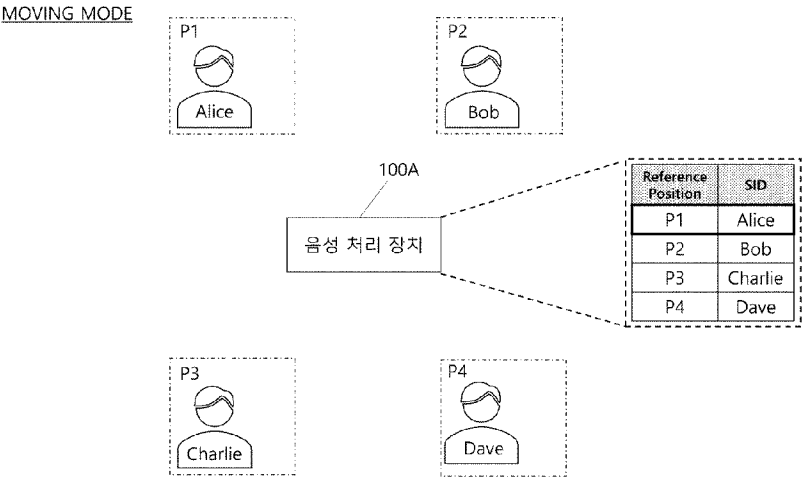


[도13]

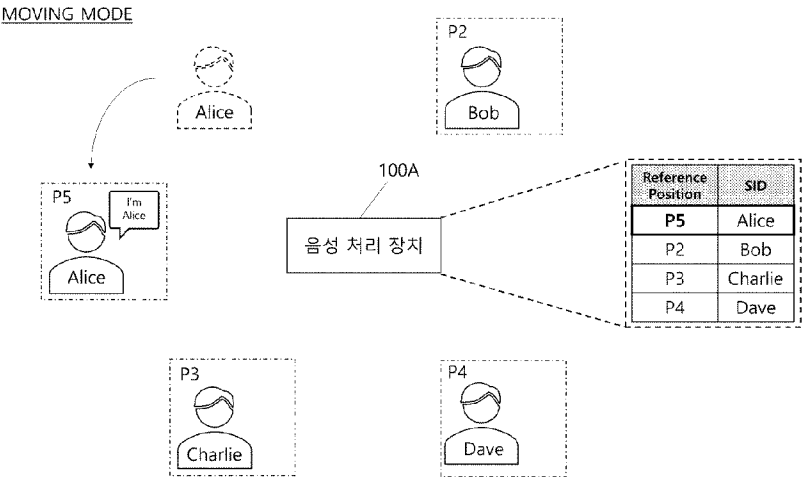
100A



[도14]

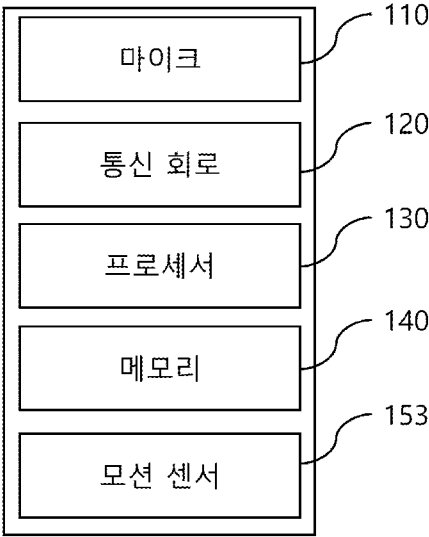


[도15]

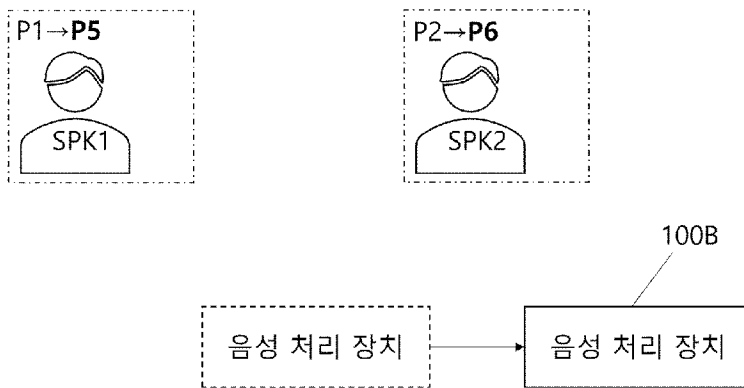


[도16]

100B

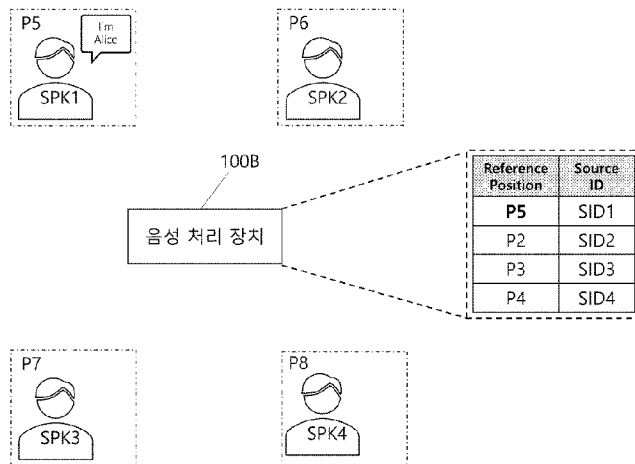


[도17]

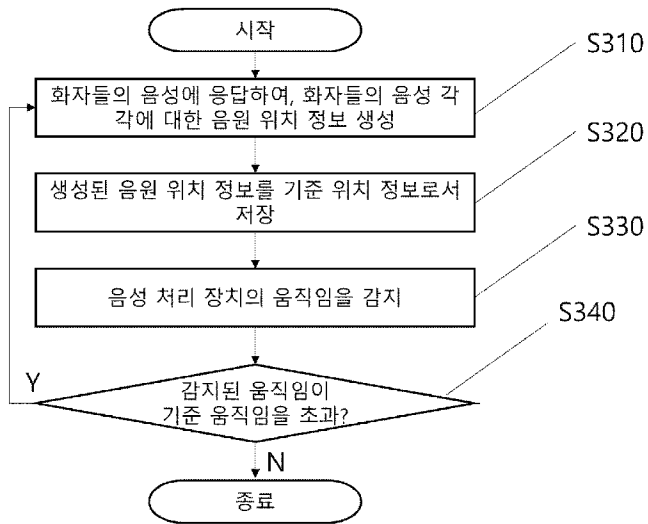


[도18]

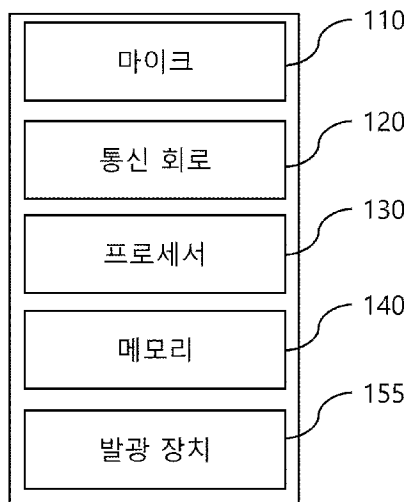
REGISTER MODE



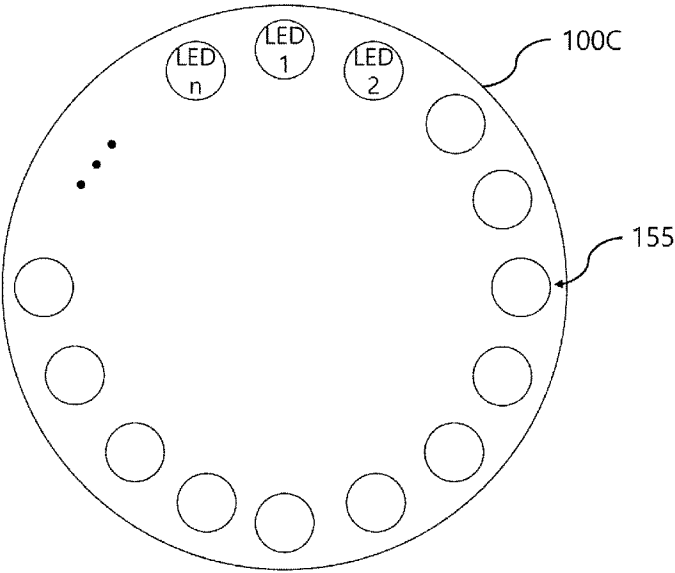
[도19]



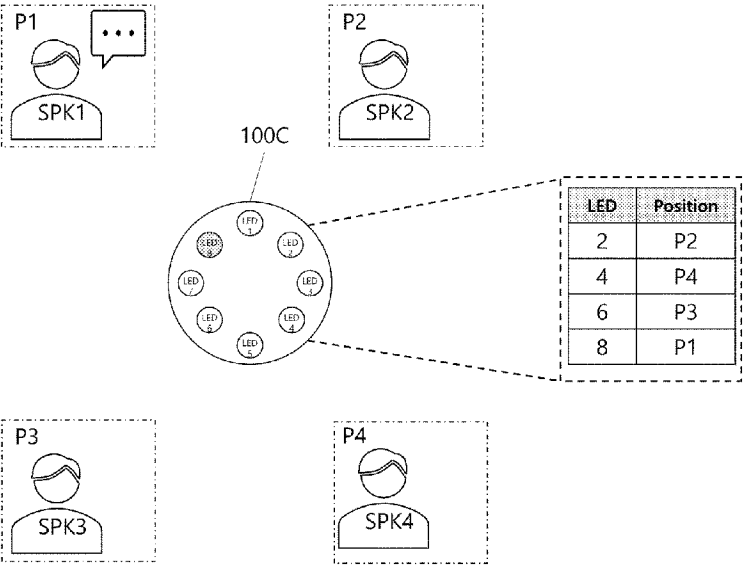
[도20]

100C

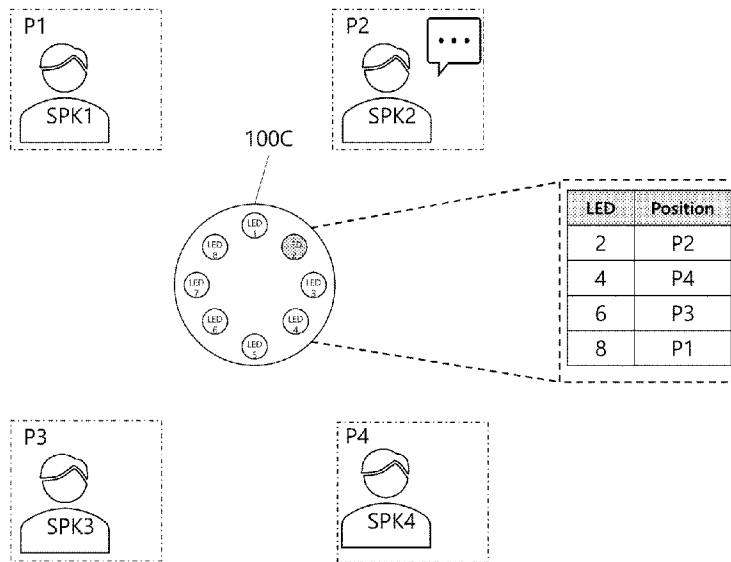
[도21]



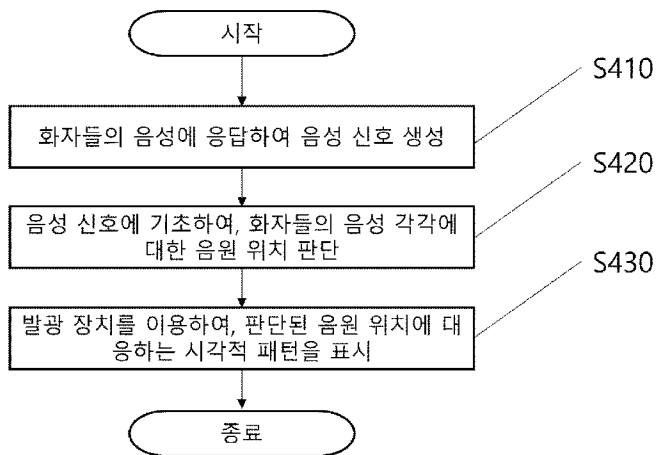
[도22]



[도23]

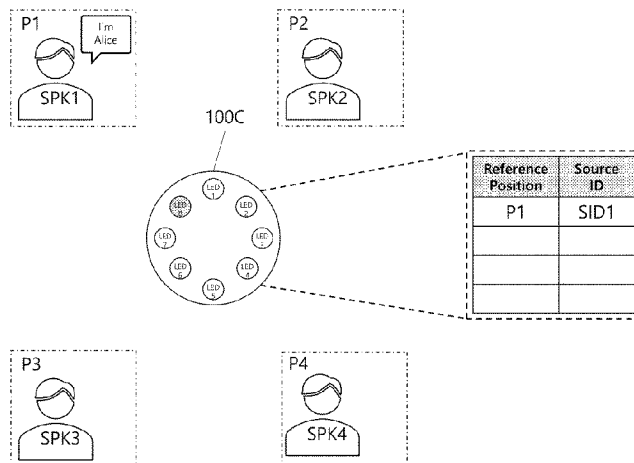


[도24]



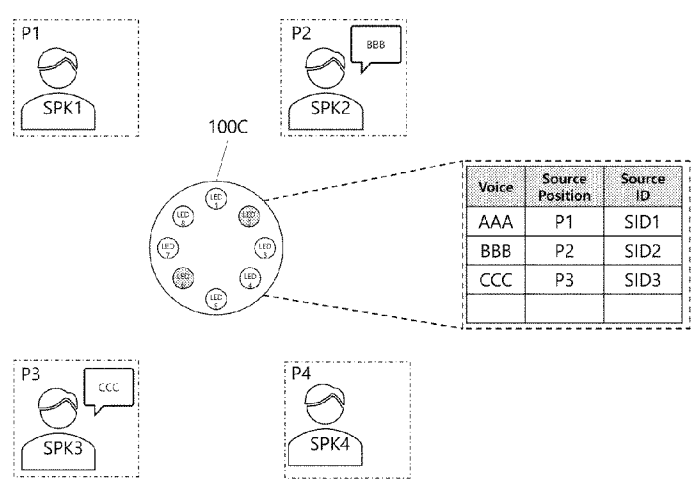
[도25]

REGISTER MODE

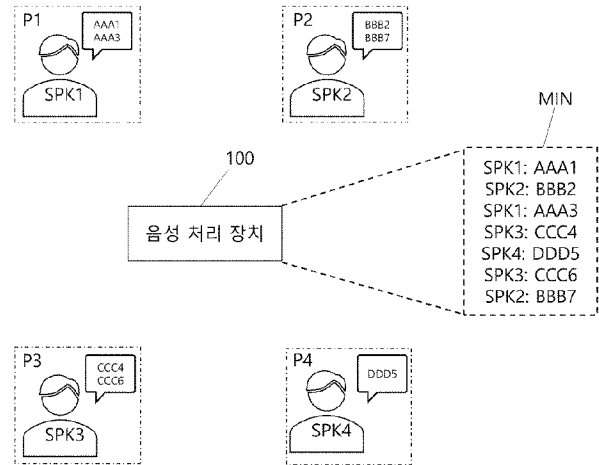


[도26]

SEPARATION MODE



[도27]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2021/008826

A. CLASSIFICATION OF SUBJECT MATTER

G10L 17/02(2013.01)i; **H04R 3/00**(2006.01)i; **G01S 3/808**(2006.01)i; **G06K 9/00**(2006.01)i; **H04N 7/15**(2006.01)i;
H05B 47/12(2020.01)i; **G06F 40/47**(2020.01)i; **G10L 15/00**(2006.01)i; **G10L 21/0272**(2013.01)i; **H04R 1/40**(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L 17/02(2013.01); G01S 3/808(2006.01); G01S 5/02(2010.01); G01S 5/22(2006.01); G01S 5/28(2006.01);
 G10L 15/04(2006.01); G10L 17/00(2006.01); G10L 21/0208(2013.01); G10L 21/0216(2013.01); G10L 21/0308(2013.01);
 G10L 25/90(2013.01); H04B 1/40(2006.01)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above
 Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & keywords: 음성(voice), 음원(sound source), 위치(location), 분리(separation), 매칭(matching), 번역(translation), 식별자(identifier)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 5339501 B2 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 13 November 2013 (2013-11-13) See paragraphs [0009] and [0061] and claims 1, 6 and 7.	1-4
Y		5-15
Y	KR 10-1983310 B1 (YOON, Seong Yong) 29 May 2019 (2019-05-29) See claim 1.	5,6
Y	JP 5724125 B2 (ADVANCED TELECOMMUNICATION RESEARCH INSTITUTE INTERNATIONAL) 27 May 2015 (2015-05-27) See paragraph [0023] and figure 2.	7-12
Y	JP 4964204 B2 (NIPPON TELEGR. & TELEPH. CORP. <NTT>) 27 June 2012 (2012-06-27) See paragraph [0017].	11,12



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance
 “D” document cited by the applicant in the international application
 “E” earlier application or patent but published on or after the international filing date
 “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 “O” document referring to an oral disclosure, use, exhibition or other means
 “P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

19 October 2021

Date of mailing of the international search report

20 October 2021

Name and mailing address of the ISA/KR

**Korean Intellectual Property Office
 Government Complex-Daejeon Building 4, 189 Cheongsaro, Seo-gu, Daejeon 35208**

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2021/008826**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KR 10-2127640 B1 (SAMSUNG ELECTRONICS CO., LTD.) 30 June 2020 (2020-06-30) See claim 12.	13-15
A	KR 10-1889809 B1 (MEDI AZEN, INC.) 20 August 2018 (2018-08-20) See paragraphs [0014]-[0028] and claim 1.	1-15
A	US 2020-0217919 A1 (NOKIA TECHNOLOGIES OY) 09 July 2020 (2020-07-09) See paragraphs [0008]-[0053].	1-15

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2021/008826

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
JP	5339501	B2	13 November 2013	JP	2010-026361	A	04 February 2010
KR	10-1983310	B1	29 May 2019	KR	10-2019-0005617	A	16 January 2019
JP	5724125	B2	27 May 2015	JP	2012-211768	A	01 November 2012
JP	4964204	B2	27 June 2012	JP	2010-054733	A	11 March 2010
KR	10-2127640	B1	30 June 2020	CN	105103457	A	25 November 2015
				CN	105103457	B	07 July 2017
				CN	107257525	A	17 October 2017
				CN	107257525	B	07 July 2020
				EP	2979364	A1	03 February 2016
				EP	3591849	A1	08 January 2020
				KR	10-2014-0118060	A	08 October 2014
				US	10091599	B2	02 October 2018
				US	10869146	B2	15 December 2020
				US	2014-0294183	A1	02 October 2014
				US	2017-0019745	A1	19 January 2017
				US	2019-0037329	A1	31 January 2019
				US	9467786	B2	11 October 2016
				WO	2014-157846	A1	02 October 2014
KR	10-1889809	B1	20 August 2018	None			
US	2020-0217919	A1	09 July 2020	EP	3642643	A1	29 April 2020
				WO	2018-234626	A1	27 December 2018

A. 발명이 속하는 기술분류(국제특허분류(IPC)) G10L 17/02(2013.01)i; H04R 3/00(2006.01)i; G01S 3/808(2006.01)i; G06K 9/00(2006.01)i; H04N 7/15(2006.01)i; H05B 47/12(2020.01)i; G06F 40/47(2020.01)i; G10L 15/00(2006.01)i; G10L 21/0272(2013.01)i; H04R 1/40(2006.01)i		
B. 조사된 분야 조사된 최소문헌(국제특허분류를 기재) G10L 17/02(2013.01); G01S 3/808(2006.01); G01S 5/02(2010.01); G01S 5/22(2006.01); G01S 5/28(2006.01); G10L 15/04(2006.01); G10L 17/00(2006.01); G10L 21/0208(2013.01); G10L 21/0216(2013.01); G10L 21/0308(2013.01); G10L 25/90(2013.01); H04B 1/40(2006.01) 조사된 기술분야에 속하는 최소문헌 이외의 문헌 한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우)) eKOMPASS(특허청 내부 검색시스템) & 키워드: 음성(voice), 음원(sound source), 위치(location), 분리(separation), 매칭(matching), 번역(translation), 식별자(identifier)		
C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X	JP 5339501 B2 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2013.11.13 단락 [0009], [0061] 및 청구항 1, 6, 7	1-4
Y		5-15
Y	KR 10-1983310 B1 (윤성용) 2019.05.29 청구항 1	5,6
Y	JP 5724125 B2 (ADVANCED TELECOMMUNICATION RESEARCH INSTITUTE INTERNATIONAL) 2015.05.27 단락 [0023] 및 도면 2	7-12
Y	JP 4964204 B2 (NIPPON TELEGR. & TELEPH. CORP. <NTT>) 2012.06.27 단락 [0017]	11,12
☒ 추가 문헌이 C(계속)에 기재되어 있습니다. ☒ 대응특허에 관한 별지를 참조하십시오.		
* 인용된 문헌의 특별 카테고리: “A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌 “D” 본 국제출원에서 출원인이 인용한 문헌 “E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌 “L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌 “O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌 “P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌 “T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌 “X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다. “Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다. “&” 동일한 대응특허문헌에 속하는 문헌		
국제조사의 실제 완료일 2021년10월19일(19.10.2021)		국제조사보고서 발송일 2021년10월20일(20.10.2021)
ISA/KR의 명칭 및 우편주소 대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578		심사관 양정록 전화번호 +82-42-481-5709

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	KR 10-2127640 B1 (삼성전자주식회사) 2020.06.30 청구항 12	13-15
A	KR 10-1889809 B1 (미디어젠(주)) 2018.08.20 단락 [0014]-[0028] 및 청구항 1	1-15
A	US 2020-0217919 A1 (NOKIA TECHNOLOGIES OY) 2020.07.09 단락 [0008]-[0053]	1-15

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
JP 5339501 B2	2013/11/13	JP 2010-026361 A	2010/02/04
KR 10-1983310 B1	2019/05/29	KR 10-2019-0005617 A	2019/01/16
JP 5724125 B2	2015/05/27	JP 2012-211768 A	2012/11/01
JP 4964204 B2	2012/06/27	JP 2010-054733 A	2010/03/11
KR 10-2127640 B1	2020/06/30	CN 105103457 A	2015/11/25
		CN 105103457 B	2017/07/07
		CN 107257525 A	2017/10/17
		CN 107257525 B	2020/07/07
		EP 2979364 A1	2016/02/03
		EP 3591849 A1	2020/01/08
		KR 10-2014-0118060 A	2014/10/08
		US 10091599 B2	2018/10/02
		US 10869146 B2	2020/12/15
		US 2014-0294183 A1	2014/10/02
		US 2017-0019745 A1	2017/01/19
		US 2019-0037329 A1	2019/01/31
		US 9467786 B2	2016/10/11
		WO 2014-157846 A1	2014/10/02
KR 10-1889809 B1	2018/08/20	없음	
US 2020-0217919 A1	2020/07/09	EP 3642643 A1	2020/04/29
		WO 2018-234626 A1	2018/12/27