(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2020 년 1 월 9 일 (09.01.2020)

WIPO | PCT

(10) 국제공개번호
WO 2020/009378 A1

(51) 국제특허분류:
H04R 3/00 (2006.01) *G10L 19/04* (2006.01)
G10L 21/0272 (2013.01)

(21) 국제출원번호: PCT/KR2019/007871

(22) 국제출원일: 2019 년 6 월 28 일 (28.06.2019)

(25) 출원언어: 한국어

(26) 공개언어: 한국어

(30) 우선권정보:
10-2018-0077318 2018 년 7 월 3 일 (03.07.2018) KR

(71) 출원인: 삼성전자 주식회사 (**SAMSUNG ELECTRONICS CO., LTD.**) [KR/KR]; 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR).

(72) 발명자: 마르친키에비치피오트르 (**MARCINKIEWICZ, Piotr**); 05-200 보어민 위드나 20/2에이, Wolomin (PL). 그레제시악그제고 파벨 (**GRZESIAK, Grzegorz Pawel**); 01-315 바르샤바 쿠퍼라 4에이/30, Warsaw (PL). 티카주크야쿠브 (**TKACZUK, Jakub**); 84-230 루미아 스탈로바 11, Rumia (PL).

(74) 대리인: 리앤목 특허법인 (**Y.P.LEE,MOCK & PARTNERS**); 06292 서울시 강남구 언주로30길 13 대림아크로텔 12층, Seoul (KR).

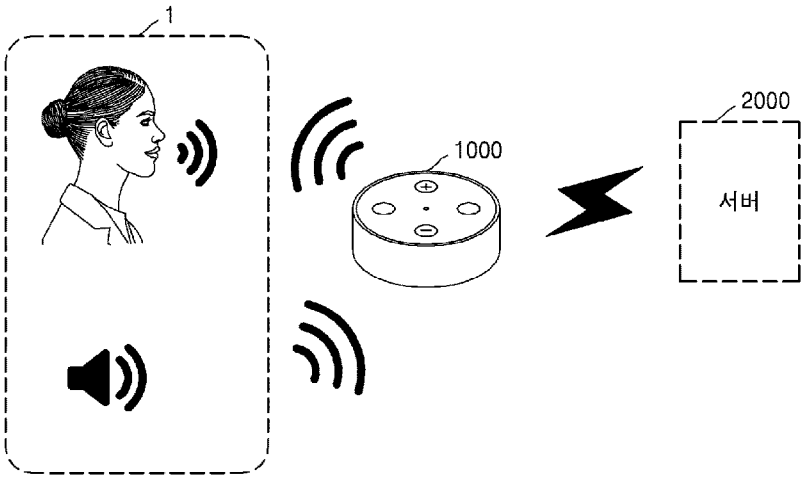
(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 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 (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:
— 국제조사보고서와 함께 (조약 제21조(3))

(54) Title: DEVICE FOR OUTPUTTING SOUND AND METHOD THEREFOR

(54) 발명의 명칭: 소리를 출력하는 디바이스 및 그 방법



2000 ... Server

(57) Abstract: Disclosed are a device for outputting sound and a method therefor. A sound outputting method comprises the steps of: predicting external sound received from an external environment; variably adjusting sound to be outputted from the device on the basis of the predicted external sound; and outputting the adjusted sound.

(57) 요약서: 소리를 출력하는 디바이스 및 그 방법이 개시된다. 소리 출력 방법은 외부 환경으로부터 수신되는 외부 소리를 예측하는 단계, 상기 예측된 외부 소리에 기반하여 상기 디바이스로부터 출력될 소리를 가변적으로 조절하는 단계 및 상기 조절된 소리를 출력하는 단계를 포함한다.

WO 2020/009378 A1

명세서

발명의 명칭: 소리를 출력하는 디바이스 및 그 방법

기술분야

[1] 본 개시는 소리를 출력하는 디바이스 및 그 방법에 관한 것이다. 보다 상세하게는, 출력될 소리를 가변적으로 조절하는 소리 출력 디바이스 및 그 방법에 관한 것이다.

배경기술

[2] 산업이 고도화됨에 따라, 다양한 전자 장치가 아날로그 형태로부터 디지털 형태로 변화되고 있으며, 음향 기기의 경우에도 디지털화가 급속히 보급되고 있다.

[3] 한편, 최근에 스마트폰과 같이 다양한 기능을 복합적으로 수행하는 전자 장치들이 개발됨에 따라, 디지털화에 이어, 음성 인식 기능이 탑재된 다양한 전자 장치들이 출시되고 있다. 음성 인식 기능은, 사용자가 별도의 버튼을 조작하거나 또는 터치 모듈에 대한 접촉을 하지 않고, 음성을 통해 장치를 손쉽게 제어할 수 있도록 한다.

[4] 음성 인식 기능에는 다양한 인공지능 기술이 활용될 수 있다. 언어적 이해를 통해 사용자의 발화 내용을 정확하게 인식하고, 발화 내용에 포함된 사용자의 의도를 파악할 수 있다. 또한, 시각적 이해를 통해 사용자가 사용하는 컨텐츠에 대한 이해가 가능하고, 추론 예측을 통해 사용자의 의도 및 사용자가 실행 중인 컨텐츠에 따른 적절한 응답을 생성할 수도 있다. 나아가, 동작 제어를 통해 생성한 응답을 적절한 형태로 출력할 수도 있다.

발명의 상세한 설명

기술적 과제

[5] 보다 향상된 성능을 제공하기 위하여 출력될 소리를 가변적으로 조절하는 소리 출력 디바이스가 요구된다.

과제 해결 수단

[6] 일부 실시예는, 외부 소리를 예측하고, 예측된 외부 소리에 기초하여 출력될 소리를 조절하는 디바이스 및 방법을 제공할 수 있다.

[7] 또한, 일부 실시예는, 외부 소리를 수신하고, 수신한 외부 소리에 기초하여 미래의 외부 소리를 예측하는 디바이스 및 방법을 제공할 수 있다.

[8] 또한, 일부 실시예는, 미리 저장된 정보에 기초하여 미래의 외부 소리를 예측하는 디바이스 및 방법을 제공할 수 있다.

발명의 효과

[9] 실시예에 의한 소리 출력 디바이스는 출력될 소리를 가변적으로 조절하여 사용자에게 향상된 소리 출력을 경험하게 할 수 있다.

도면의 간단한 설명

[10] 본 발명은, 다음의 자세한 설명과 그에 수반되는 도면들의 결합으로 쉽게 이해될 수 있으며, 참조 번호(reference numerals)들은 구조적 구성요소(structural elements)를 의미한다.

[11] 도 1은 일부 실시예에 따른 디바이스가 소리를 출력하는 예시를 나타내는 도면이다.

[12] 도 2는 일부 실시예에 따른 디바이스가 소리를 출력하는 방법의 흐름도이다.

[13] 도 3은 일부 실시예에 따른 디바이스(1000)가, 예측된 외부 소리에 기반하여, 출력될 소리를 가변적으로 조절하고 출력하는 방법을 나타내는 흐름도이다.

[14] 도 4는 일부 실시예에 따른 디바이스(1000)가, 예측된 외부 소리에 기반하여, 출력될 소리를 가변적으로 조절하고 출력하는 방법을 나타내는 흐름도이다.

[15] 도 5는 일부 실시예에 따른, 사용자의 음성 입력을 포함하는 외부 소리를 설명하기 위한 도면이다.

[16] 도 6은 일부 실시예에 따른 디바이스가 가변적으로 소리를 조절하는 예시를 설명하기 위한 도면이다

[17] 도 7은 일부 실시예에 따른 디바이스가 출력될 소리를 컴포넌트로 분해하고 조절하는 예시를 설명하기 위한 도면이다.

[18] 도 8은 일부 실시예에 따른 디바이스에 의하여 조절된 소리를 설명하기 위한 도면이다.

[19] 도 9는 일부 실시예에 따른 디바이스에 의하여 조절된 소리를 설명하기 위한 도면이다.

[20] 도 10은 일부 실시예에 따른 디바이스가 외부 소리를 예측하는 방법을 설명하기 위한 도면이다.

[21] 도 11은 일부 실시예에 따른 디바이스가 외부 소리를 예측하는 방법을 설명하기 위한 도면이다.

[22] 도 12는 일부 실시예에 따른 디바이스가 외부 소리를 예측하는 방법을 설명하기 위한 도면이다.

[23] 도 13은 일부 실시예에 따른 디바이스가 소리를 출력하는 방법을 설명하기 위한 도면이다.

[24] 도 14는 일부 실시예에 따른 디바이스가 소리를 출력하는 방법을 설명하기 위한 도면이다.

[25] 도 15는 일부 실시예에 따른 디바이스가 소리를 출력하는 방법을 설명하기 위한 도면이다.

[26] 도 16은 일부 실시예에 따른 디바이스가 소리를 출력하는 방법을 설명하기 위한 도면이다.

[27] 도 17은 일부 실시예에 따른 디바이스의 블록도이다.

[28] 도 18은 일부 실시예에 따른 디바이스의 블록도이다.

[29] 도 19는 일부 실시예에 따른 프로세서의 블록도이다.

[30] 도 20은 일부 실시예에 따른 데이터 학습부의 블록도이다.

[31] 도 21은 일부 실시예에 따른 데이터 인식부의 블록도이다.

[32] 도 22는 일부 실시예에 따른 디바이스 및 서버가 서로 연동함으로써 데이터를 학습하고 인식하는 예시를 나타내는 도면이다.

발명의 실시를 위한 최선의 형태

[33] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 개시의 일 실시예에 따른 소리 출력 디바이스는, 소리를 출력하는 스피커, 하나 이상의 인스트럭션을 저장하는 메모리 및 상기 하나 이상의 인스트럭션을 실행하는 프로세서를 포함하며, 상기 프로세서는, 상기 하나 이상의 인스트럭션을 실행함으로써, 외부 환경으로부터 수신될 외부 소리를 예측하고, 상기 예측된 외부 소리에 기반하여 상기 스피커로부터 출력될 소리를 가변적으로 조절하며, 상기 조절된 소리가 출력되도록 상기 스피커를 제어할 수 있다.

[34] 또한 본 개시의 일 실시예에 따른 소리 출력 방법은 외부 환경으로부터 수신되는 외부 소리를 예측하는 단계, 상기 예측된 외부 소리에 기반하여 상기 디바이스로부터 출력될 소리를 가변적으로 조절하는 단계 및 상기 조절된 소리를 출력하는 단계를 포함할 수 있다.

[35] 또한 본 개시의 일 실시예에 따른 컴퓨터로 판독 가능한 기록매체는, 상술된 방법을 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체일 수 있다.

발명의 실시를 위한 형태

[36] 아래에서는 첨부한 도면을 참조하여 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 개시의 실시예를 상세히 설명한다. 그러나 본 개시는 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 또한, 도면에서 본 개시를 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[37] 본 개시의 일부 실시예는 기능적인 블록 구성들 및 다양한 처리 단계들로 나타내어질 수 있다. 이러한 기능 블록들의 일부 또는 전부는, 특정 기능들을 실행하는 다양한 개수의 하드웨어 및/또는 소프트웨어 구성들로 구현될 수 있다. 예를 들어, 본 개시의 기능 블록들은 하나 이상의 마이크로프로세서들에 의해 구현되거나, 소정의 기능을 위한 회로 구성들에 의해 구현될 수 있다. 또한, 예를 들어, 본 개시의 기능 블록들은 다양한 프로그래밍 또는 스크립팅 언어로 구현될 수 있다. 기능 블록들은 하나 이상의 프로세서들에서 실행되는 알고리즘으로 구현될 수 있다. 또한, 본 개시는 전자적인 환경 설정, 신호 처리, 및/또는 데이터 처리 등을 위하여 종래 기술을 채용할 수 있다.

[38] 또한, 도면에 도시된 구성 요소들 간의 연결 선 또는 연결 부재들은 기능적인 연결 및/또는 물리적 또는 회로적 연결들을 예시적으로 나타낸 것일 뿐이다. 실제 장치에서는 대체 가능하거나 추가된 다양한 기능적인 연결, 물리적인 연결,

또는 회로 연결들에 의해 구성 요소들 간의 연결이 나타내어질 수 있다.

[39] 또한, 본 명세서에 기재된 "...부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다. "부", "모듈"은 어드레싱될 수 있는 저장 매체에 저장되며 프로세서에 의해 실행될 수 있는 프로그램에 의해 구현될 수도 있다.

[40] 한편, 본 명세서와 도면에 개시된 본 발명의 실시예들은 본 개시의 기술 내용을 쉽게 설명하고 본 개시의 이해를 돕기 위해 특정 예를 제시한 것일 뿐이며, 본 개시의 범위를 한정하고자 하는 것은 아니다. 즉, 본 개시의 기술적 사상에 바탕을 둔 다른 변형예들이 실시 가능하다는 것은 본 개시의 속하는 기술 분야에서 통상의 지식을 가진 자에게 자명한 것이다. 또한 각각의 실시예는 필요에 따라 서로 조합되어 운용할 수 있다. 예컨대, 본 발명의 일 실시예와 다른 일 실시예의 일부분들이 서로 조합되어 장치의 동작이 운용될 수 있다.

[41] 도 1은 일부 실시예에 따른 디바이스가 소리를 출력하는 예시를 나타내는 도면이다.

[42] 도 1을 참조하면, 디바이스(1000)는 외부 환경(1)으로부터 수신되는 소리인 외부 소리를 수신할 수 있다. 이때 외부 환경(1)은 음성 입력을 제공하는 사용자, 멀티미디어 컨텐츠를 재생하는 외부 장치, 노이즈를 발생시키는 노이즈 소스 등을 포함할 수 있다. 또한 디바이스(1000)는 외부 환경(1)으로 소리를 출력할 수 있다. 디바이스(1000)가 출력하는 소리는 음성 가이드, 멀티미디어 컨텐츠, 경고음 등을 포함할 수 있다. 그러나 이는 예시적인 것으로, 디바이스(1000)가 출력하는 소리는 상술된 예시에 한정되지 않는다.

[43] 디바이스(1000)는, 외부 환경(1)으로부터 수신되는 소리인 외부 소리를 예측할 수 있다. 보다 구체적으로, 디바이스(1000)는, 미래에, 외부 환경(1)으로부터 디바이스(1000)로 수신될 외부 소리를 예측할 수 있다.

[44] 실시 예에 있어서, 디바이스(1000)는 미리 수신된 외부 소리에 기초하여, 미래의 외부 소리를 예측할 수 있다. 혹은 실시 예에 있어서, 디바이스(1000)는 디바이스(1000)가 미리 획득한 정보에 기초하여, 미래의 외부 소리를 예측할 수 있다. 예를 들어, 디바이스(1000)는 현재 디바이스(1000)가 재생 중인 멀티미디어에 관한 정보에 기초하여, 미래에 디바이스(1000)가 출력할 소리를 판별할 수 있다.

[45] 디바이스(1000)는 예측된 외부 소리에 기반하여, 디바이스(1000)로부터 출력될 소리를 가변적으로 조절할 수 있다. 실시 예에 있어서, 디바이스(1000)는 예측된 외부 소리와 출력될 소리가, 주파수 영역 혹은 시간 영역 중 적어도 하나에서 분리되도록, 출력될 소리를 가변적으로 조절할 수 있다.

[46] 이때, 디바이스(1000)는, 출력될 소리를 조절하기 위하여 이용되는 필터의 특성을, 시간 변화에 따라 가변적으로 조절할 수 있다. 예를 들어, 디바이스(1000)는 시간 변화에 따라 지속적으로 외부 소리를 예측하면서,

갱신되는 예측 결과에 기초하여 지속적으로 새로이 필터를 판별할 수 있다.

[47] 한편, 서버(2000)는 디바이스(1000)와 연동함으로써 외부 소리를 예측할 수 있다. 디바이스(1000)는 스마트폰, 태블릿 PC, PC, 스마트 TV, 휴대폰, PDA(personal digital assistant), 랩톱, 미디어 플레이어, 마이크로 서버, GPS(global positioning system) 장치, 전자책 단말기, 디지털방송용 단말기, 네비게이션, 키오스크, MP3 플레이어, 디지털 카메라, 가전기기 및 기타 모바일 또는 비모바일 컴퓨팅 장치일 수 있으나, 이에 한정되지 않는다. 또한, 디바이스(1000)는 통신 기능 및 데이터 프로세싱 기능을 구비한 시계, 안경, 헤어 밴드 및 반지 등의 웨어러블 디바이스일 수 있다. 그러나, 이에 한정되지 않으며, 디바이스(1000)는 외부 소리를 수신하고, 소리를 출력하는 모든 종류의 기기를 포함할 수 있다.

[48] 또한, 디바이스(1000)는 다양한 외부 소리에 관한 정보를 이용하기 위하여, 소정의 네트워크를 통하여 서버 및 다른 장치(미도시)와 통신할 수 있다. 이 경우, 네트워크는 근거리 통신망(Local Area Network; LAN), 광역 통신망(Wide Area Network; WAN), 부가가치 통신망(Value Added Network; VAN), 이동 통신망(mobile radio communication network), 위성 통신망 및 이들의 상호 조합을 포함하며, 각 네트워크 구성 주체가 서로 원활하게 통신을 할 수 있도록 하는 포괄적인 의미의 데이터 통신망이며, 유선 인터넷, 무선 인터넷 및 모바일 무선 통신망을 포함할 수 있다. 무선 통신은 예를 들어, 무선 랜(Wi-Fi), 블루투스, 블루투스 저 에너지(Bluetooth low energy), 지그비, WFD(Wi-Fi Direct), UWB(ultra wideband), 적외선 통신(IrDA, infrared Data Association), NFC(Near Field Communication) 등이 있을 수 있으나, 이에 한정되는 것은 아니다.

[49] 도 2는 일부 실시예에 따른 디바이스가 소리를 출력하는 방법의 흐름도이다.

[50] 단계 S210에서, 디바이스(1000)는, 외부 환경(1)으로부터 수신되는 소리인 외부 소리를 예측할 수 있다. 보다 구체적으로, 디바이스(1000)는, 미래에, 외부 환경(1)으로부터 디바이스(1000)로 수신될 외부 소리를 예측할 수 있다. 실시 예에 있어서, 디바이스(1000)는, 소정의 시간 이후, 외부 환경(1)으로부터 수신될 외부 소리를 예측할 수 있다. 소정의 시간은 디바이스(1000)가 외부 소리를 예측하는 방식에 기초하여 다양한 방식으로 정해질 수 있다.

[51] 실시 예에 있어서, 디바이스(1000)는 미리 수신된 외부 소리에 기초하여, 미래의 외부 소리를 예측할 수 있다. 외부 소리를 예측하기 위하여, 디바이스(1000)는 외부 환경(1)으로부터 수신될 수 있는 다양한 소리에 관한 정보들을 미리 획득할 수 있다. 실시 예에 있어서, 상술된 다양한 소리에 관한 정보들은 사용자의 음성 입력, 발표된 음악, 혹은 안내 방송에 관한 정보를 포함할 수 있다.

[52] 예를 들어, 디바이스(1000)는 "빅스비(Bixby)," 라는 사용자의 음성이 포함된 외부 소리를 수신하면, "볼륨을 올려."라는 사용자의 음성이 포함된 외부 소리가 수신될 것이라고 예측할 수 있다. 혹은 디바이스(1000)는 이미 알고 있는 음악

트랙의 일부분이 포함된 외부 소리를 수신하면, 상술된 음악 트랙의 다음 부분이 포함된 외부 소리가 수신될 것이라고 예측할 수 있다. 혹은 디바이스(1000)는 "열차가 들어오고 있습니다," 라는 안내 방송 소리를 포함하는 외부 소리를 수신하면, "물러서 주시기 바랍니다."라는 다음 안내 방송 소리를 포함하는 외부 소리가 수신될 것이라고 예측할 수 있다.

[53] 실시 예에 있어서, 디바이스(1000)는 디바이스(1000)가 미리 획득한 정보에 기초하여, 미래의 외부 소리를 예측할 수 있다. 예를 들어, 디바이스(1000)는 현재 디바이스(1000)가 재생 중인 멀티미디어에 관한 정보에 기초하여, 미래에 디바이스(1000)가 출력할 소리를 판별할 수 있다. 디바이스(1000) 판별된 미래에 출력될 소리에 기초하여, 미래의 외부 소리를 예측할 수 있다. 혹은 디바이스(1000)는, 네트워크를 통하여, 외부 기기 혹은 서버(2000)로부터, 디바이스(1000) 주변에 위치한 기기들이 재생 중인 멀티미디어에 관한 정보를 획득하고, 획득한 정보에 기초하여 미래의 외부 소리를 예측할 수 있다.

[54] 단계 S220에서, 디바이스(1000)는 예측된 외부 소리에 기반하여, 디바이스(1000)로부터 출력될 소리를 가변적으로 조절할 수 있다.

[55] 실시 예에 있어서, 디바이스(1000)는 예측된 외부 소리와 출력될 소리가, 주파수 영역 혹은 시간 영역 중 적어도 하나에서 분리되도록, 출력될 소리를 가변적으로 조절할 수 있다. 디바이스(1000)는 예측된 외부 소리와 출력될 소리가, 주파수 영역 혹은 시간 영역 중 적어도 하나에서 분리되도록, 출력될 소리에 적용할 적어도 하나의 필터를 판별할 수 있다. 이때, 디바이스(1000)는, 개수, 주파수 범위, 형태, 진폭 등 필터의 특성을, 시간 변화에 따라 가변적으로 조절할 수 있다. 예를 들어, 디바이스(1000)는 시간 변화에 따라 지속적으로 외부 소리를 예측하면서, 갱신되는 예측 결과에 기초하여 지속적으로 새로이 필터를 판별할 수 있다.

[56] 실시 예에 있어서, 디바이스(1000)에서 출력될 소리는 디바이스(1000)가 사용자에게 제공하는 음성 가이드를 포함할 수 있다. 디바이스(1000)는 예측된 외부 소리와 디바이스(1000)에서 출력될 소리가 분리되도록 필터를 적용하여, 사용자에게 보다 명확하게 음성 가이드가 제공되도록 할 수 있다.

[57] 실시 예에 있어서, 디바이스(1000)는 예측된 외부 소리에 사용자의 음성 입력이 포함되는 경우, 예측된 외부 소리와 디바이스(1000)에서 출력될 소리가 분리되도록 필터를 적용하여, 사용자의 음성 입력이 보다 명확하게 인식되도록 할 수 있다.

[58] 실시 예에 있어서, 디바이스(1000)는 예측된 외부 소리에 정보를 포함하는 소리, 예를 들어 경고음, 가 포함되는 경우, 예측된 외부 소리와 디바이스(1000)에서 출력될 소리가 분리되도록 필터를 적용하여, 사용자에게 정보를 포함하는 소리가 보다 명확하게 인식되도록 할 수 있다.

[59] 단계 S230에서, 디바이스(1000)는 조절된 소리를 출력할 수 있다. 디바이스(1000)가 출력하는 소리는 음성 가이드, 멀티미디어 컨텐츠, 경고음

등을 포함하는 다양한 형태일 수 있다.

[60] 도 3은 일부 실시예에 따른 디바이스(1000)가, 예측된 외부 소리에 기반하여, 출력될 소리를 가변적으로 조절하고 출력하는 방법을 나타내는 흐름도이다.

[61] 단계 S310에서, 디바이스(1000)는 출력될 소리가 예측된 외부 소리와 시간 대역에서 분리되도록, 출력될 소리를 조절할 수 있다. 두 소리를 시간 대역에서 분리한다는 것은, 두 소리가 시간 대역에서 중첩되지 않거나 적은 양이 중첩되도록 하는 것일 수 있다. 예를 들어, 디바이스(1000)는 외부 소리의 크기가 작을 것이라 예측되는 시간에 디바이스(1000)로부터 소리가 출력되도록, 출력될 소리를 조절할 수 있다.

[62] 실시 예에 있어서, 디바이스(1000)는 출력될 소리를 복수의 컴포넌트들로 분해할 수 있다. 예를 들어, 디바이스(1000)는 출력될 소리가 오케스트라 음악을 포함하는 경우, 오케스트라 음악을 악기에 기초하는 컴포넌트들로 분해할 수 있다. 디바이스(1000)는 출력될 소리와 예측된 외부 소리를 분리하기 위하여, 출력될 소리의 각 컴포넌트들을 각각 독립적으로 조절할 수 있다. 예를 들어, 디바이스(1000)는 출력될 소리와 예측된 외부 소리를 분리하기 위하여, 출력될 소리의 컴포넌트들 중 드럼에 기초한 컴포넌트만 선택적으로 조절할 수 있다. 그러나 이는 예시적인 것으로, 디바이스(1000)의 컴포넌트 조절 동작은 상술된 예에 한정되지 않는다.

[63] 단계 S320에서, 디바이스(1000)는 조절된 소리를 출력할 수 있다. 디바이스(1000)는 외부 소리와 시간적으로 분리된 소리를 출력하여, 외부 소리 및 출력되는 소리가 사용자에게 명확하게 인식되도록 할 수 있다.

[64] 도 4는 일부 실시예에 따른 디바이스(1000)가, 예측된 외부 소리에 기반하여, 출력될 소리를 가변적으로 조절하고 출력하는 방법을 나타내는 흐름도이다.

[65] 단계 S410에서, 디바이스(1000)는 출력될 소리가 예측된 외부 소리와 주파수 대역에서 분리되도록, 출력될 소리를 조절할 수 있다. 두 소리를 주파수 대역에서 분리한다는 것은, 두 소리가 주파수 대역에서 중첩되지 않거나 적은 양이 중첩되도록 하는 것일 수 있다. 예를 들어, 디바이스(1000)는 외부 소리의 주파수의 전력이 작을 것이라 예측되는 주파수 대역에, 디바이스(1000)로부터 출력될 소리의 주파수가 위치되도록, 출력될 소리를 조절할 수 있다.

[66] 실시 예에 있어서, 디바이스(1000)는 출력될 소리를 복수의 컴포넌트들로 분해할 수 있다. 디바이스(1000)는 출력될 소리와 예측된 외부 소리를 분리하기 위하여, 출력될 소리의 각 컴포넌트들을 각각 독립적으로 조절할 수 있다. 예를 들어, 디바이스(1000)는 출력될 소리와 예측된 외부 소리를 분리하기 위하여, 출력될 소리의 컴포넌트들 중 플루트에 기초한 컴포넌트의 주파수만 선택적으로 조절할 수 있다. 그러나 이는 예시적인 것으로, 디바이스(1000)의 컴포넌트 조절 동작은 상술된 예에 한정되지 않는다.

[67] 단계 S420에서, 디바이스(1000)는 조절된 소리를 출력할 수 있다. 디바이스(1000)는 외부 소리와 주파수적으로 분리된 소리를 출력하여, 외부 소리

및 출력되는 소리가 사용자에게 명확하게 인식되도록 할 수 있다.

[68] 도 5는 일부 실시예에 따른, 사용자의 음성 입력을 포함하는 외부 소리를 설명하기 위한 도면이다.

[69] 도 5를 참조하면, 사용자의 음성 입력은 "six"로 예시되었다. 그러나 이는 예시적인 것으로, 본 개시의 기술적 특징이 사용자의 음성 입력의 내용에 의하여 제한되지 않음은 당업자에게 자명하다.

[70] 도 5를 참조하면, 음성 입력 중 "s"에 해당되는 부분과, "i"에 해당되는 부분은 서로 다른 주파수 및 전력 특성을 나타낸다. 즉, 음성 입력을 포함하는 외부 소리의 주파수 및 전력 특성은, 시간에 따라 변화하며, 디바이스(1000)는 음성 입력의 내용을 예측하여 그 변화를 예측할 수 있다. 디바이스(1000)는 외부 소리를 출력될 소리와 효율적으로 분리하기 위하여, 예측된 음성 입력을 포함하는 외부 소리에 기초하여, 출력될 소리를 시간에 따라 가변적으로 조절할 수 있다.

[71] 도 6은 일부 실시예에 따른 디바이스가 가변적으로 소리를 조절하는 예시를 설명하기 위한 도면이다.

[72] 도 6을 참조하면, 외부 소리는 사용자의 음성 신호(speech signal)를 포함하고, 디바이스(1000)로부터 출력될 소리는 음악 사운드트랙(music soundtrack)을 포함하는 것으로 예시되었다. 그러나 이는 예시적인 것으로, 본 개시의 기술적 특징이 외부 소리 및 디바이스(1000)로부터 출력될 소리의 종류에 의하여 제한되지 않음은 당업자에게 자명하다.

[73] 도 6을 참조하면, 단계 S610에서, 디바이스(1000)는 제 1 음성을 수신한다. 제 1 음성은 음성 신호의 일부분으로, 제 1 수신 시간(t11) 동안의 음성 신호를 포함한다. 제 1 수신 시간(t11)은 디바이스(1000)가 외부 소리를 수신하는 소정의 시간일 수 있다.

[74] 단계 S620 단계에서, 디바이스(1000)는 다음 시간대의 음악 사운드트랙을 예측할 수 있다. 실시 예에 있어서, 디바이스(1000)는 자신이 재생하고 있는 음악 사운드트랙에 관한 정보를 이미 가지고 있으므로, 미리 가지고 있는 정보에 기초하여 앞으로 출력될 음악 사운드트랙을 예측할 수 있다. 예를 들어, 디바이스(1000)는 제 1 예측 시간(t21) 동안 출력될 음악 사운드트랙을 예측할 수 있다. 제 1 예측 시간(t21)은, 디바이스(1000)에 의하여 예측된, 앞으로 출력될 소리가 포함되는 소정의 시간일 수 있다.

[75] 또한, 디바이스(1000)는 다음 시간대의 음성 신호를 예측할 수 있다. 디바이스(1000)는 수신한 제 1 음성에 기초하여, 제 1 예측 시간(t21)동안 수신될 음성 신호를 포함하는 외부 소리를 예측할 수 있다.

[76] 실시 예에 있어서, 디바이스(1000)는 수신한 제 1 음성에 기초하여, 음성 신호를 발성한 사용자의 음성 모델을 예측할 수 있다. 또한, 디바이스(1000)는 수신한 제 1 음성에 기초하여, 제 1 예측 시간(t21)동안 수신될 음성 신호의 내용을 예측할 수 있다. 디바이스(1000)는 예측한 음성 모델 및 음성 신호의 내용 중 적어도

하나를 이용하여, 제 1 예측 시간(t21)동안 수신될 음성 신호를 예측할 수 있다. 예를 들어, 디바이스(1000)는 "빅스비(Bixby)," 라는 제 1 음성에 기초하여, 여성인 사용자의 음성 모델을 판정할 수 있다. 또한 디바이스(1000)는 제 1 음성에 기초하여 "볼륨을 올려."라는 내용을 예측할 수 있다. 디바이스(1000)는 판정된 음성 모델과 내용에 기초하여, 음성 신호를 예측할 수 있다.

[77] 디바이스(1000)는 예측된 외부 소리 및 출력할 소리에 기초하여, 출력될 소리를, 제 1 출력 시간(t31)동안 외부 소리와 분리하기 위하여 적용할 적어도 하나의 필터를 판별할 수 있다. 제 1 출력 시간(t31)은, 판별된 필터가 출력될 소리에 적용되는 시간일 수 있다.

[78] 디바이스(1000)는 상술된 S610, S620 및 S630 단계를 반복하여 수행할 수 있다. 이때, 디바이스(1000)는, 개수, 주파수 범위, 형태, 진폭 등 필터의 특성을, 시간 변화에 따라 가변적으로 조절할 수 있다. 디바이스(1000)는 시간 변화에 따라 지속적으로 외부 소리를 예측하면서, 갱신되는 예측 결과에 기초하여 지속적으로 새로이 필터를 판별하여 효과적으로 외부 소리와 출력될 소리를 분리할 수 있다.

[79] 도 7은 일부 실시예에 따른 디바이스가 출력될 소리를 컴포넌트로 분해하고 조절하는 예시를 설명하기 위한 도면이다.

[80] 단계 S710에서, 디바이스(1000)는 출력될 소리를 복수의 악기 컴포넌트들로 분해(decompose)하고, 복수의 악기 컴포넌트들 중 적어도 하나를 선택할 수 있다. 실시 예에 있어서, 디바이스(1000)는 외부 소리와 출력될 소리를 분리하기 용이하도록, 악기 컴포넌트를 선택할 수 있다. 예를 들어, 디바이스(1000)는 사용자에게 둔감하게 인식되는 악기 컴포넌트를 선택할 수 있다. 혹은, 디바이스(1000)는 외부 소리의 전력이 낮게 예측되는 주파수 대역의 악기 컴포넌트를 선택할 수 있다. 그러나 디바이스(1000)의 악기 컴포넌트 선택 방법은 상술된 예시에 한정되지 않는다.

[81] 단계 S720에서, 디바이스(1000)는 출력될 소리가 예측된 소리와 분리되도록, 선택된 악기 컴포넌트를 조절할 수 있다. 실시 예에 있어서, 디바이스(1000)는, 선택된 악기 컴포넌트를, 시간 대역 및 주파수 대역 중 적어도 하나에 기초하여 조절할 수 있다.

[82] 단계 S730에서, 디바이스(1000)는 조절된 소리를 출력할 수 있다. 보다 구체적으로, 디바이스(1000)는 조절된 각 악기 컴포넌트들을 다시 결합할 수 있다. 디바이스(1000)는 결합된 소리를 출력할 수 있다.

[83] 도 8은 일부 실시예에 따른 디바이스에 의하여 조절된 소리를 설명하기 위한 도면이다.

[84] 도 8의 상단 스펙트로그램(spectrogram)은 조절되지 않은 기존 음악 파일의 전력을, 시간 및 주파수 대역에서 표시한 도면이다. 도 8의 하단 스펙트로그램(spectrogram)은, 디바이스(1000)에 의하여 조절된 음악 파일의 전력을, 시간 및 주파수 대역에서 표시한 도면이다.

[85] 도 8을 참조하면, 음악 파일은, 시간 및 주파수 영역에서 외부 소리와 분리되도록, 조절되어 출력될 수 있다.

[86] 도 9는 일부 실시예에 따른 디바이스에 의하여 조절된 소리를 설명하기 위한 도면이다.

[87] 도 9에 도시된 첫 번째 스펙트로그램(spectrogram)은, 주변 환경 소리(environment)을 포함하는, 예측된 외부 소리의 전력을, 시간 및 주파수 대역에 표시한 도면이다. 두 번째 스펙트로그램(spectrogram)은 디바이스(1000)로부터 출력될, 음성 어시스턴트(voice assistant)를 포함하는, 소리의 전력을, 시간 및 주파수 대역에 표시한 도면이다. 세 번째 스펙트로그램(spectrogram)은 디바이스(1000)로부터 출력될 소리가 조절되지 않은 경우 사용자가 수신하는 소리 의 전력을, 시간 및 주파수 대역에 표시한 도면이다. 네 번째 스펙트로그램(spectrogram)은 디바이스(1000)로부터 출력될 소리가 조절된 경우 사용자가 수신하는 소리의 전력을, 시간 및 주파수 대역에서 표시한 도면이다.

[88] 도 9를 참조하면, 디바이스(1000)는 출력될 소리를 시간 혹은 주파수 대역에서 외부 소리와 분리하여, 사용자로 하여금 음성 어시스턴트를 보다 명료하게 수신하도록 할 수 있다.

[89] 도 10은 일부 실시예에 따른 디바이스가 외부 소리를 예측하는 방법을 설명하기 위한 도면이다.

[90] 도 10을 참조하면, 단계 S1010에서, 디바이스(1000)는 제 2 음성을 수신할 수 있다. 제 2 음성은, 소정의 시간 동안 수신된, 외부 소리에 포함된 사용자의 음성일 수 있다.

[91] 단계 S1020에서, 디바이스(1000)는 제 2 음성을 제 1 학습 모델에 적용하여 제 1 음성을 예측할 수 있다. 제 1 음성은, 소정의 시간 후 수신될, 외부 소리에 포함된 사용자의 음성일 수 있다.

[92] 실시 예에 있어서, 제 1 학습 모델은 사용자의 음성 모델을 결정하기 위하여 학습된 것일 수 있다. 디바이스(1000)는 제 1 학습 모델에, 사용자의 음성에 관한 정보를 적용함으로써, 제 1 학습 모델을 학습시킬 수 있다. 예를 들어, 디바이스(1000)는 음성의 주파수에 관한 정보, 사용자의 성별, 발성 습관에 관한 정보를 제 1 학습 모델에 적용하여 제 1 학습 모델을 학습시킬 수 있다.

[93] 실시 예에 있어서, 제 1 학습 모델은 음성의 내용을 예측하기 위하여 학습된 것일 수 있다. 디바이스(1000)는 제 1 학습 모델에, 음성 입력에 관한 정보를 적용함으로써, 제 1 학습 모델을 학습시킬 수 있다. 예를 들어, 디바이스(1000)는 사용자의 음성 입력 이력에 관한 정보를 제 1 학습 모델에 적용하여 제 1 학습 모델을 학습시킬 수 있다.

[94] 실시 예에 있어서, 제 1 학습 모델은 미리 설정되어 디바이스(1000)에 저장된 것일 수 있다. 이 경우, 제 1 학습 모델을 생성 운영하는 서버(2000)가 디바이스(1000)에게 제 1 학습 모델을 제공할 수 있으며, 디바이스(1000)는

서버(2000)로부터 수신된 제 1 학습 모델을 디바이스(1000) 내에 저장하고 관리할 수 있다.

[95] 또한, 미리 설정된 제 1 학습 모델은 서버(2000)에 저장될 수 있다. 이 경우, 디바이스(1000)는 서버에게 궤적 정보를 제공하고, 궤적 정보에 기초하여 결정된 단어 정보를 서버(2000)로부터 수신할 수 있다. 또한, 이 경우에는, 서버(2000)에 의해 제 1 학습 모델이 사용자 별로 관리될 수 있다.

[96] 도 11은 일부 실시예에 따른 디바이스가 외부 소리를 예측하는 방법을 설명하기 위한 도면이다.

[97] 도 11을 참조하면, 단계 S1110에서, 디바이스(1000)는 제 2 음악 소리를 수신할 수 있다. 제 2 음악 소리는, 소정의 시간 동안 수신된, 외부 소리에 포함된 음악 소리일 수 있다. 제 2 음악 소리는 이미 발행된 음악의 일부분일 수 있다.

[98] 단계 S1120에서, 디바이스(1000)는 제 2 음악 소리에 기초하여 제 1 음악 소리를 예측할 수 있다. 제 1 음악 소리는, 소정의 시간 후 수신될, 외부 소리에 포함된 음악 소리일 수 있다. 즉, 제 1 음악 소리는, 제 2 음악 소리가 포함되는 이미 발행된 음악의, 제 2 음악 소리보다 뒤의 부분일 수 있다.

[99] 실시 예에 있어서, 디바이스(1000)는 제 2 음악 소리를, 음악을 식별하기 위한 제 1 학습 모델에 적용하여 제 1 음악 소리를 예측할 수 있다. 디바이스(1000)는 제 1 학습 모델을 이용하여, 제 2 음악 소리가 포함되는 이미 발행된 음악을 식별할 수 있다. 디바이스(1000)는 식별된 음악에 기초하여 제 1 음악 소리를 예측할 수 있다.

[100] 혹은, 디바이스(1000)는 제 2 음악 소리를, 적어도 하나의 음악에 관한 정보를 포함하는 데이터베이스와 매칭하여 제 1 음악 소리를 예측할 수 있다. 디바이스(1000)는 데이터베이스를 이용하여, 제 2 음악 소리가 포함되는 이미 발행된 음악을 식별할 수 있다. 디바이스(1000)는 식별된 음악에 기초하여 제 1 음악 소리를 예측할 수 있다.

[101] 도 12는 일부 실시예에 따른 디바이스가 외부 소리를 예측하는 방법을 설명하기 위한 도면이다.

[102] 단계 S1210에서, 디바이스(1000)는 제 2 음악 소리를 수신할 수 있다. 제 2 음악 소리는, 소정의 시간 동안 수신된, 외부 소리에 포함된 음악 소리일 수 있다. 제 2 음악 소리는 이미 발행된 음악의 일부분일 수 있다.

[103] 단계 S1220에서, 디바이스(1000)는 적어도 하나의 음악에 관한 정보를 포함하는 데이터베이스를 획득할 수 있다. 데이터베이스는 디바이스(1000)에 미리 저장될 수 있다. 혹은, 디바이스(1000)는 서버(2000)로부터 데이터베이스 혹은 그에 관한 정보를 수신할 수 있다.

[104] 단계 S1230에서, 디바이스(1000)는 데이터베이스에 포함된 적어도 하나의 음악에 대한 정보와, 수신한 제 2 음악 소리를 매칭할 수 있다. 보다 구체적으로, 디바이스(1000)는 데이터베이스에 포함된 적어도 하나의 음악들에 대한 정보와 제 2 음악 소리를 매칭하여, 제 2 음악 소리가 포함되는 이미 발행된 음악을

식별할 수 있다.

[105] 단계 S1240에서, 디바이스(1000)는 매칭 결과에 기초하여, 제 1 음악 소리를 예측할 수 있다. 보다 구체적으로, 디바이스(1000)는 매칭을 통해 식별된 음악에 기초하여, 제 1 음악 소리를 예측할 수 있다.

[106] 도 13은 일부 실시예에 따른 디바이스가 소리를 출력하는 방법을 설명하기 위한 도면이다.

[107] 도 13을 참조하면, 외부 환경(1)은 외부 노이즈 소스, 사용자 등을 포함하는 다양한 소리 발생원들을 포함한다. 디바이스(1000)는 외부 환경(1)으로부터, 다양한 소리 발생원들로부터 발생된 소리를 포함하는, 외부 소리를 수신한다. 도 13을 참조하면, 외부 소리는 외부 노이즈 및 사용자의 음성 입력을 포함하는 것으로 예시되었다.

[108] 디바이스(1000)는 외부 소리를 수신하는 동안, 반향 소거(echo cancellation)동작을 수행할 수 있다. 실시 예에 있어서, 디바이스(1000)는 네트워크를 통하여 연결된 외부 노이즈 소스로부터 외부 노이즈 정보를 수신하고, 수신한 외부 노이즈 정보를 이용하여 반향 소거 동작을 수행할 수 있다. 또한, 실시 예에 있어서, 디바이스(1000)는 디바이스(1000)가 출력한 소리, 예를 들어 음악, 에 관한 정보에 기초하여 반향 소거 동작을 수행할 수 있다. 디바이스(1000)는 반향 소거 동작을 이용하여, 보다 명료하게 외부 소리를 수신할 수 있다.

[109] 디바이스(1000)는 외부 소리를 예측할 수 있다. 실시 예에 있어서, 디바이스(1000)는 수신한 외부 노이즈 정보 및 수신한 사용자의 음성 입력에 기초하여, 외부 소리를 예측할 수 있다.

[110] 또한 디바이스(1000)는 자연어 이해(NLU: Natural Language Understanding)를 이용하여 음성 입력의 내용을 획득할 수 있다. 디바이스(1000)는 음성 입력의 내용에 대응하여, 음성 어시스턴트를 이용하여, 디바이스(1000)로부터 출력될 음성에 관한 정보를 획득할 수 있다.

[111] 한편, 디바이스(1000)는 디바이스(1000)로부터 출력될 음악에 관한 정보를 획득할 수 있다. 실시 예에 있어서, 디바이스(1000)는 예측된 외부 소리와 출력될 음악에 관한 정보에 기초하여, 출력될 음악을 외부 소리와 분리하기 위한 필터를 판정할 수 있다. 또한 실시 예에 있어서, 디바이스(1000)는 출력될 음성과 출력될 음악에 관한 정보에 기초하여, 출력될 음성과 출력될 음악을 분리하기 위한 필터를 판정할 수 있다.

[112] 디바이스(1000)는, 출력될 음악 및 음성에, 판정된 필터를 적용할 수 있다. 디바이스(1000)는 필터를 이용하여 조절된 음악 및 음성을 출력할 수 있다.

[113] 도 14는 일부 실시예에 따른 디바이스가 소리를 출력하는 방법을 설명하기 위한 도면이다.

[114] S1410 단계에서, 디바이스(1000)는 제 1 음성을 수신할 수 있다. 제 1 음성은, 소정의 시간 동안 수신된, 외부 소리에 포함된 사용자의 음성일 수 있다.

[115] S1411 단계에서, 디바이스(1000)는 수신한 제 1 음성을 이용하여, 사용자를 판정할 수 있다. 사용자를 판정하는 동작은 사용자가 미리 지정된 사용자 중 하나인지 식별하는 동작을 포함할 수 있다.

[116] S1412 단계에서, 디바이스(1000)는 제 1 음성을 이용하여, 사용자의 음성 모델을 획득할 수 있다. 보다 구체적으로, 디바이스(1000)는 사용자의 음성 모델을 결정하기 위하여 학습된 학습 모델에 제 1 음성을 입력하여, 사용자의 음성 모델을 획득할 수 있다

[117] S1420 단계에서, 디바이스(1000)는 외부 노이즈를 수신할 수 있다. 외부 노이즈는, 소정의 시간동안 수신된, 외부 소리에 포함된 노이즈일 수 있다.

[118] S1421 단계에서, 디바이스(1000)는 반향 소거(echo cancellation)동작을 수행할 수 있다. 실시 예에 있어서, 디바이스(1000)는 네트워크를 통하여 연결된 외부 노이즈 소스로부터 외부 노이즈 정보를 수신하고, 수신한 외부 노이즈 정보를 이용하여 반향 소거 동작을 수행할 수 있다. 또한, 실시 예에 있어서, 디바이스(1000)는, 디바이스(1000)가 이미 출력한 소리에 관한 정보에 기초하여 반향 소거 동작을 수행할 수 있다. 디바이스(1000)는 반향 소거 동작을 이용하여, 보다 명료하게 외부 소리를 수신할 수 있다.

[119] S1422 단계에서, 디바이스(1000)는 수신한 제 1 음성에 대한 자연어 이해 동작을 수행할 수 있다. 실시 예에 있어서, 디바이스(1000)는 제 1 음성을, 자연어를 이해하도록 학습된 학습 모델에 적용하여 자연어 이해 동작을 수행할 수 있다. 디바이스(1000)는 자연어 이해 동작을 이용하여, 제 1 음성의 내용을 획득할 수 있다.

[120] S1423 단계에서, 디바이스(1000)는, 획득한 제 1 음성의 내용에 기초하여 , 사용자의 음성 입력의 다음 단어를 예측할 수 있다.

[121] S1424 단계에서, 디바이스(1000)는, 획득한 사용자의 음성 모델과, 예측된 다음 단어에 기초하여, 제 2 음성을 예측할 수 있다. 제 2 음성은 소정의 시간 후 수신될, 외부 소리에 포함된 사용자의 음성일 수 있다.

[122] S1425 단계에서, 디바이스(1000)는 디바이스로부터 출력될 신호를 획득할 수 있다. 또한, S1426 단계에서, 디바이스(1000)는 획득한 출력될 신호에 기초하여, 소정의 시간 후 출력될 신호에 관한 정보를 획득할 수 있다.

[123] S1427 단계에서, 디바이스(1000)는 예측된 제 2 음성과, 획득한 출력될 신호에 관한 정보에 기초하여, 제 2 음성과 출력될 신호가 중첩될 구간을 판정할 수 있다.

[124] S1428 단계에서, 디바이스는(1000)는 중첩된 구간에서, 출력될 신호에 적용될 필터를 판정할 수 있다. 디바이스(1000)는 중첩된 구간에서, 제 2 음성과 출력될 신호가 분리되도록 필터를 판정할 수 있다.

[125] S1429 단계에서, 디바이스(1000)는 출력될 신호에 필터를 적용하여, 소리를 출력할 수 있다.

[126] 도 15는 일부 실시예에 따른 디바이스가 소리를 출력하는 방법을 설명하기

위한 도면이다.

[127] S1510 단계에서, 디바이스(1000)는 외부 소리에 관한 정보를 획득할 수 있다. 실시 예에 있어서, 디바이스(1000)는, 네트워크를 통하여 연결된 외부 기기들로부터 외부 소리에 관한 정보를 획득할 수 있다. 혹은, 디바이스(1000)는 서버로부터 외부 소리에 관한 정보를 획득할 수 있다. 실시 예에 있어서, 외부 소리에 관한 정보는, 외부 기기들로부터 재생되는 멀티미디어 컨텐츠에 관한 정보일 수 있다. 혹은, 실시 예에 있어서, 외부 소리에 관한 정보는, 디바이스(1000)가 위치한 장소에서 발령될 경고음에 관한 정보일 수 있다. 그러나 이는 예시적인 것으로, 본 발명에서 외부 소리에 관한 정보는 상술된 실시예들로 한정되지 않는다.

[128] S1511 단계에서, 디바이스(1000)는 외부 소리의 다음 부분을 획득할 수 있다. 실시 예에 있어서, 디바이스(1000)는 획득한 외부 소리에 관한 정보에 기초하여, 외부 소리의 다음 부분을 예측하고, 획득할 수 있다. 혹은, 디바이스(1000)는 네트워크를 통하여, 외부 소리의 다음 부분에 관한 정보를 직접 수신할 수도 있다.

[129] 한편, S1520 단계에서, 디바이스(1000)는 현재 디바이스(1000)가 재생하고 있는 음악에 관한 정보를 획득할 수 있다.

[130] S1521단계에서, 디바이스(1000)는, 획득한 음악에 관한 정보에 기초하여 , 출력될 음악의 다음 부분을 획득할 수 있다.

[131] S1530 단계에서, 디바이스(1000)는, 획득한 미래의 외부 소리 및 출력될 음악에 기초하여, 디바이스(1000)로부터 출력될 음성의 주파수, 출력될 시간 및 음성 모델을 판정할 수 있다.

[132] 보다 구체적으로, 디바이스(1000)는 획득한 미래의 외부 소리 및 출력될 음악과, 출력될 음성이, 시간 및 주파수 대역 중 적어도 하나에서 분리되도록, 출력될 음성의 출력될 시간 및 음성 모델을 판정할 수 있다.

[133] S1531 단계에서, 디바이스(1000)는, 텍스트 음성 변환 모듈(TSM: Text to Speech Module)을 이용하여, 판정된 주파수, 출력될 시간 및 음성 모델에 기초하여, 출력될 음성을 생성한다.

[134] 한편, S1540 단계에서, 디바이스(1000)는, 디바이스(1000)로부터 출력될 음성과, 출력될 음악이 분리되도록, 출력될 음악에 적용될 필터를 판별할 수 있다.

[135] S1541 단계에서, 디바이스(1000)는, 판별된 필터를 출력될 음악에 적용할 수 있다

[136] S1532 단계에서, 디바이스(1000)는, 필터가 적용된 음악과, 생성된 음성을, 결합하여 출력할 수 있다.

[137] 도 16은 일부 실시예에 따른 디바이스가 소리를 출력하는 방법을 설명하기 위한 도면이다.

[138] S1610 단계에서, 디바이스(1000)는 외부 소리에 관한 정보를 획득할 수 있다.

실시 예에 있어서, 디바이스(1000)는, 네트워크를 통하여 연결된 외부 기기들로부터 외부 소리에 관한 정보를 획득할 수 있다. 혹은, 디바이스(1000)는 서버로부터 외부 소리에 관한 정보를 획득할 수 있다.

[139] S1611 단계에서, 디바이스(1000)는 외부 소리의 다음 부분을 획득할 수 있다. 실시 예에 있어서, 디바이스(1000)는 획득한 외부 소리에 관한 정보에 기초하여, 외부 소리의 다음 부분을 예측하고, 획득할 수 있다. 혹은, 디바이스(1000)는 네트워크를 통하여, 외부 소리의 다음 부분에 관한 정보를 직접 수신할 수도 있다.

[140] 한편, S1620 단계에서, 디바이스(1000)는 현재 디바이스(1000)가 재생하고 있는 음악에 관한 정보를 획득할 수 있다.

[141] S1621단계에서, 디바이스(1000)는, 획득한 음악에 관한 정보에 기초하여 , 출력될 음악의 다음 부분을 획득할 수 있다.

[142] S1630 단계에서, 디바이스(1000)는, 텍스트 음성 변환 모듈(TSM: Text to Speech Module)을 이용하여, 출력될 음성을 생성할 수 있다.

[143] S1631 단계에서, 디바이스(1000)는, 디바이스(1000)로부터 출력될 음성 및 출력될 음악과 외부 소리가 분리되도록, 출력될 음성에 적용될 필터를 판별할 수 있다.

[144] 한편, S1640 단계에서, 디바이스(1000)는, 디바이스(1000)로부터 출력될 음성 및 출력될 음악과 외부 소리가 분리되도록, 출력될 음악에 적용될 필터를 판별할 수 있다.

[145] S1641 단계 및 S1632 단계에서, 디바이스(1000)는, 판별된 필터들을 출력될 음성 및 음악에 적용할 수 있다

[146] S1633 단계에서, 디바이스(1000)는, 필터가 각각 적용된 음악 및 음성을, 결합하여 출력할 수 있다.

[147] 도 17 및 도 18은 일부 실시예에 따른 디바이스(1000)의 블록도이다.

[148] 도 17 및 도 18에 도시된 바와 같이, 일부 실시예에 따른 디바이스(1000)는, 사용자 입력부(1100), 출력부(1200), 제어부(1300), 및 통신부(1500)를 포함할 수 있다. 그러나, 도 17에 도시된 구성 요소 모두가 디바이스(1000)의 필수 구성 요소인 것은 아니다. 도 17에 도시된 구성 요소보다 많은 구성 요소에 의해 디바이스(1000)가 구현될 수도 있고, 도 17에 도시된 구성 요소보다 적은 구성 요소에 의해 디바이스(1000)가 구현될 수도 있다.

[149] 예를 들어, 도 18에 도시된 바와 같이, 일부 실시예에 따른 디바이스(1000)는, 사용자 입력부(1100), 출력부(1200), 제어부(1300), 및 통신부(1500) 이외에 센싱부(1400), A/V 입력부(1600), 및 메모리(1700)를 더 포함할 수도 있다.

[150] 사용자 입력부(1100)는, 사용자가 디바이스(1000)를 제어하기 위한 데이터를 입력하는 수단을 의미한다. 예를 들어, 사용자 입력부(1100)에는 키 패드(key pad), 돔 스위치 (dome switch), 터치 패드(접촉식 정전 용량 방식, 압력식 저항막 방식, 적외선 감지 방식, 표면 초음파 전도 방식, 적분식 장력 측정 방식, 피에조

효과 방식 등), 조그 휠, 조그 스위치, 마이크 등이 있을 수 있으나 이에 한정되는 것은 아니다.

[151] 사용자 입력부(1100)는 외부 환경으로부터 외부 소리를 수신할 수 있다. 실시 예에 있어서, 외부 소리는 사용자의 음성 입력을 포함할 수 있다.

[152] 출력부(1200)는, 오디오 신호 또는 비디오 신호 또는 진동 신호를 출력할 수 있으며, 출력부(1200)는 디스플레이부(1210), 음향 출력부(1220), 및 진동 모터(1230)를 포함할 수 있다.

[153] 디스플레이부(1210)는 디바이스(1000)에서 처리되는 정보를 표시 출력한다. 예를 들어, 디스플레이부(1210)는 사용자에게 음성 입력 인증 결과를 제공하기 위한 사용자 인터페이스를 디스플레이 할 수 있다.

[154] 한편, 디스플레이부(1210)와 터치패드가 레이어 구조를 이루어 터치 스크린으로 구성되는 경우, 디스플레이부(1210)는 출력 장치 이외에 입력 장치로도 사용될 수 있다. 디스플레이부(1210)는 액정 디스플레이(liquid crystal display), 박막 트랜지스터 액정 디스플레이(thin film transistor-liquid crystal display), 유기 발광 다이오드(organic light-emitting diode), 플렉시블 디스플레이(flexible display), 3차원 디스플레이(3D display), 전기영동 디스플레이(electrophoretic display) 중에서 적어도 하나를 포함할 수 있다. 그리고 디바이스(1000)의 구현 형태에 따라 디바이스(1000)는 디스플레이부(1210)를 2개 이상 포함할 수도 있다. 이때, 2개 이상의 디스플레이부(1210)는 힌지(hinge)를 이용하여 마주보게 배치될 수 있다.

[155] 음향 출력부(1220)는 통신부(1500)로부터 수신되거나 메모리(1700)에 저장된 오디오 데이터를 출력한다. 또한, 음향 출력부(1220)는 디바이스(1000)에서 수행되는 기능(예를 들어, 호신호 수신음, 메시지 수신음, 알림음)과 관련된 음향 신호를 출력한다. 이러한 음향 출력부(1220)에는 스피커(speaker), 버저(Buzzer) 등이 포함될 수 있다.

[156] 진동 모터(1230)는 진동 신호를 출력할 수 있다. 예를 들어, 진동 모터(1230)는 오디오 데이터 또는 비디오 데이터(예컨대, 호신호 수신음, 메시지 수신음 등)의 출력에 대응하는 진동 신호를 출력할 수 있다. 또한, 진동 모터(1230)는 터치스크린에 터치가 입력되는 경우 진동 신호를 출력할 수도 있다.

[157] 제어부(1300)는, 통상적으로 디바이스(1000)의 전반적인 동작을 제어한다. 예를 들어, 제어부(1300)는, 메모리(1700)에 저장된 프로그램들을 실행함으로써, 사용자 입력부(1100), 출력부(1200), 센싱부(1400), 통신부(1500), A/V 입력부(1600) 등을 전반적으로 제어할 수 있다.

[158] 구체적으로, 제어부(1300)는, 미래의 외부 소리를 예측하고, 예측된 외부 소리에 기반하여, 출력될 소리를 가변적으로 조절할 수 있다.

[159] 실시 예에 있어서, 제어부(1300)는 미리 수신된 외부 소리에 기초하여, 미래의 외부 소리를 예측할 수 있다. 혹은 실시 예에 있어서, 제어부(1300)는 제어부(1300)가 미리 획득한 정보에 기초하여, 미래의 외부 소리를 예측할 수

있다. 예를 들어, 제어부(1300)는 현재 디바이스(1000)가 재생 중인 멀티미디어에 관한 정보에 기초하여, 미래에 디바이스(1000)가 출력할 소리를 판별할 수 있다.

[160] 제어부(1300)는 예측된 외부 소리에 기반하여, 디바이스(1000)로부터 출력될 소리를 가변적으로 조절할 수 있다. 실시 예에 있어서, 제어부(1300)는 예측된 외부 소리와 출력될 소리가, 주파수 영역 혹은 시간 영역 중 적어도 하나에서 분리되도록, 출력될 소리를 가변적으로 조절할 수 있다.

[161] 이때, 제어부(1300)는, 출력될 소리를 조절하기 위하여 이용되는 필터의 특성을, 시간 변화에 따라 가변적으로 조절할 수 있다. 예를 들어, 제어부(1300)는 시간 변화에 따라 지속적으로 외부 소리를 예측하면서, 갱신되는 예측 결과에 기초하여 지속적으로 새로이 필터를 판별할 수 있다.

[162] 센싱부(1400)는, 디바이스(1000)의 상태 또는 디바이스(1000) 주변의 상태를 감지하고, 감지된 정보를 제어부(1300)로 전달할 수 있다.

[163] 센싱부(1400)는, 지자기 센서(Magnetic sensor)(1410), 가속도 센서(Acceleration sensor)(1420), 온/습도 센서(1430), 적외선 센서(1440), 자이로스코프 센서(1450), 위치 센서(예컨대, GPS)(1460), 기압 센서(1470), 근접 센서(1480), 및 RGB 센서(illuminance sensor)(1490) 중 적어도 하나를 포함할 수 있으나, 이에 한정되는 것은 아니다. 각 센서들의 기능은 그 명칭으로부터 당업자가 직관적으로 추론할 수 있으므로, 구체적인 설명은 생략하기로 한다.

[164] 통신부(1500)는, 디바이스(1000)와 HMD 장치 또는 디바이스(1000)와 서버 간의 통신을 하게 하는 하나 이상의 구성요소를 포함할 수 있다. 예를 들어, 통신부(1500)는, 근거리 통신부(1510), 이동 통신부(1520), 방송 수신부(1530)를 포함할 수 있다.

[165] 근거리 통신부(short-range wireless communication unit)(1510)는, 블루투스 통신부, BLE(Bluetooth Low Energy) 통신부, 근거리 무선 통신부(Near Field Communication unit), WLAN(와이파이) 통신부, 지그비(Zigbee) 통신부, 적외선(IrDA, infrared Data Association) 통신부, WFD(Wi-Fi Direct) 통신부, UWB(ultra wideband) 통신부, Ant+ 통신부 등을 포함할 수 있으나, 이에 한정되는 것은 아니다.

[166] 이동 통신부(1520)는, 이동 통신망 상에서 기지국, 외부의 단말, 서버 중 적어도 하나와 무선 신호를 송수신한다. 여기에서, 무선 신호는, 음성 호 신호, 화상 통화 호 신호 또는 문자/멀티미디어 메시지 송수신에 따른 다양한 형태의 데이터를 포함할 수 있다.

[167] 방송 수신부(1530)는, 방송 채널을 통하여 외부로부터 방송 신호 및/또는 방송 관련된 정보를 수신한다. 방송 채널은 위성 채널, 지상파 채널을 포함할 수 있다. 구현 예에 따라서 디바이스(1000)가 방송 수신부(1530)를 포함하지 않을 수도 있다.

[168] 또한, 통신부(1500)는, 컨텍스트 정보를 이용하기 위한 정보를, HMD 장치, 서버 및 주변 기기와 송수신할 수 있다.

[169] A/V(Audio/Video) 입력부(1600)는 오디오 신호 또는 비디오 신호 입력을 위한 것으로, 이에는 카메라(1610)와 마이크로폰(1620) 등이 포함될 수 있다. 카메라(1610)은 화상 통화모드 또는 촬영 모드에서 이미지 센서를 통해 정지영상 또는 동영상 등의 화상 프레임을 얻을 수 있다. 이미지 센서를 통해 캡쳐된 이미지는 제어부(1300) 또는 별도의 이미지 처리부(미도시)를 통해 처리될 수 있다.

[170] 카메라(1610)에서 처리된 화상 프레임은 메모리(1700)에 저장되거나 통신부(1500)를 통하여 외부로 전송될 수 있다. 혹은 화상 프레임은 제어부(1300)의 음성 활성 조건 및 음성 비활성 조건 판별 동작에 이용될 수 있다. 카메라(1610)는 단말기의 구성 태양에 따라 2개 이상이 구비될 수도 있다.

[171] 마이크로폰(1620)은, 외부 소리를 수신한다. 마이크로폰(1620)이 수신하는 외부 소리는 음성 입력을 포함할 수 있다. 또한 마이크로폰(1620)은 외부의 음향 신호를 입력 받아 전기적인 음성 데이터로 처리한다. 예를 들어, 마이크로폰(1620)은 외부 디바이스 또는 화자로부터 음향 신호를 수신할 수 있다. 마이크로폰(1620)는 음성 입력을 포함하는 외부의 음향 신호를 입력 받는 과정에서 발생 되는 잡음(noise)를 제거하기 위한 다양한 잡음 제거 알고리즘을 이용할 수 있다.

[172] 메모리(1700)는, 제어부(1300)의 처리 및 제어를 위한 프로그램을 저장할 수 있고, 디바이스(1000)로 입력되거나 디바이스(1000)로부터 출력되는 데이터를 저장할 수도 있다.

[173] 메모리(1700)는 플래시 메모리 타입(flash memory type), 하드디스크 타입(hard disk type), 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD 또는 XD 메모리 등), 램(RAM, Random Access Memory) SRAM(Static Random Access Memory), 롬(ROM, Read-Only Memory), EEPROM(Electrically Erasable Programmable Read-Only Memory), PROM(Programmable Read-Only Memory), 자기 메모리, 자기 디스크, 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다.

[174] 메모리(1700)에 저장된 프로그램들은 그 기능에 따라 복수 개의 모듈들로 분류할 수 있는데, 예를 들어, UI 모듈(1710), 터치 스크린 모듈(1720), 알림 모듈(1730) 등으로 분류될 수 있다.

[175] UI 모듈(1710)은, 애플리케이션 별로 디바이스(1000)와 연동되는 특화된 UI, GUI 등을 제공할 수 있다. 터치 스크린 모듈(1720)은 사용자의 터치 스크린 상의 터치 제스처를 감지하고, 터치 제스처에 관한 정보를 제어부(1300)로 전달할 수 있다. 일부 실시예에 따른 터치 스크린 모듈(1720)은 터치 코드를 인식하고 분석할 수 있다. 터치 스크린 모듈(1720)은 컨트롤러를 포함하는 별도의 하드웨어로 구성될 수도 있다.

[176] 터치스크린의 터치 또는 근접 터치를 감지하기 위해 터치스크린의 내부 또는 근처에 다양한 센서가 구비될 수 있다. 터치스크린의 터치를 감지하기 위한

센서의 일례로 촉각 센서가 있다. 촉각 센서는 사람이 느끼는 정도로 또는 그 이상으로 특정 물체의 접촉을 감지하는 센서를 말한다. 촉각 센서는 접촉면의 거칠기, 접촉 물체의 단단함, 접촉 지점의 온도 등의 다양한 정보를 감지할 수 있다.

[177] 또한, 터치스크린의 터치를 감지하기 위한 센서의 일례로 근접 센서가 있다.

[178] 근접 센서는 소정의 검출면에 접근하는 물체, 혹은 근방에 존재하는 물체의 유무를 전자계의 힘 또는 적외선을 이용하여 기계적 접촉이 없이 검출하는 센서를 말한다. 근접 센서의 예로는 투과형 광전 센서, 직접 반사형 광전 센서, 미러 반사형 광전 센서, 고주파 발진형 근접 센서, 정전용량형 근접 센서, 자기형 근접 센서, 적외선 근접 센서 등이 있다. 사용자의 터치 제스처에는 탭, 터치&홀드, 더블 탭, 드래그, 패닝, 플릭, 드래그 앤드 드롭, 스와이프 등이 있을 수 있다.

[179] 알림 모듈(1730)은 디바이스(1000)의 이벤트 발생을 알리기 위한 신호를 발생할 수 있다. 디바이스(1000)에서 발생되는 이벤트의 예로는 호 신호 수신, 메시지 수신, 키 신호 입력, 일정 알림 등이 있다. 알림 모듈(1730)은 디스플레이부(1210)를 통해 비디오 신호 형태로 알림 신호를 출력할 수도 있고, 음향 출력부(1220)를 통해 오디오 신호 형태로 알림 신호를 출력할 수도 있고, 진동 모터(1230)를 통해 진동 신호 형태로 알림 신호를 출력할 수도 있다.

[180] 도 19는 일부 실시예에 따른 프로세서(1300)의 블록도이다.

[181] 도 19를 참조하면, 일부 실시예에 따른 프로세서(1300)는 데이터 학습부(1310) 및 데이터 인식부(1320)를 포함할 수 있다.

[182] 실시 형태에 따라 데이터 학습부(1310)의 적어도 일부 및 데이터 인식부(1320)의 적어도 일부는 소프트웨어 모듈로 구현되거나 또는 하드웨어 칩 형태로 제작되어 디바이스에 탑재될 수 있다.

[183] 데이터 학습부(1310)는 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나를 판단하기 위한 기준을 학습할 수 있다. 데이터 학습부(1310)는 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나를 판별하기 위하여, 어떤 데이터를 이용할 지에 관한 기준을 학습할 수 있다. 또한, 데이터 학습부(1310)는 제공되는 데이터를 이용하여 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나를 판단하기 위한 기준을 어떻게 결정할 지에 관한 기준을 학습할 수 있다. 데이터 학습부(1310)는 학습에 이용될 데이터를 획득하고, 획득된 데이터를 후술할 데이터 인식 모델에 적용함으로써, 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나에 관한 기준을 학습할 수 있다.

[184] 데이터 학습부(1310)는 도 1 내지 도 10에서의 디바이스가 이용하는 학습 모델들의 기능을 제공할 수 있으며, 1개 이상의 데이터 학습부(1310)에 의해 도 1 내지 도 16에서의 디바이스가 이용하는 학습 모델들의 기능이 구현될 수 있다.

[185] 데이터 인식부(1320)는 데이터에 기초하여 사용자의 음성 모델, 음성 입력의

내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행할 수 있다. 데이터 인식부(1320)는 학습된 데이터 인식 모델을 이용하여, 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행할 수 있다. 데이터 인식부(1320)는 학습에 의한 기 설정된 기준에 따라 소정의 데이터를 획득하고, 획득된 데이터를 입력 값으로 하여 데이터 인식 모델을 이용할 수 있다. 또한, 데이터 인식부(1320)는 이를 이용함으로써, 소정의 데이터에 기초한 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행할 수 있다. 또한, 획득된 데이터를 입력 값으로 하여 데이터 인식 모델에 의해 출력된 결과 값은, 데이터 인식 모델을 갱신하는데 이용될 수 있다.

[186] 데이터 학습부(1310) 및 데이터 인식부(1320) 중 적어도 하나는, 적어도 하나의 하드웨어 칩 형태로 제작되어 디바이스(1000)에 탑재될 수 있다. 예를 들어, 데이터 학습부(1310) 및 데이터 인식부(1320) 중 적어도 하나는 인공 지능(AI; artificial intelligence)을 위한 전용 하드웨어 칩 형태로 제작될 수도 있고, 또는 기존의 범용 프로세서(예: CPU 또는 application processor) 또는 그래픽 전용 프로세서(예: GPU)의 일부로 제작되어 전술한 각종 디바이스(1000)에 탑재될 수도 있다. 이 때, 인공 지능을 위한 전용 하드웨어 칩은 확률 연산에 특화된 전용 프로세서로서, 기존의 범용 프로세서보다 병렬처리 성능이 높아 머신러닝과 같은 인공 지능 분야의 연산 작업을 빠르게 처리할 수 있다.

[187] 데이터 학습부(1310) 및 데이터 인식부(1320)는 하나의 디바이스에 탑재될 수도 있으며, 또는 별개의 장치들에 각각 탑재될 수도 있다. 예를 들어, 데이터 학습부(1310) 및 데이터 인식부(1320) 중 하나는 디바이스에 포함되고, 나머지 하나는 서버에 포함될 수 있다. 또한, 데이터 학습부(1310) 및 데이터 인식부(1320)는 유선 또는 무선으로 통하여, 데이터 학습부(1310)가 구축한 모델 정보를 데이터 인식부(1320)로 제공할 수도 있고, 데이터 인식부(1320)로 입력된 데이터가 추가 학습 데이터로서 데이터 학습부(1310)로 제공될 수도 있다.

[188] 한편, 데이터 학습부(1310) 및 데이터 인식부(1320) 중 적어도 하나는 소프트웨어 모듈로 구현될 수 있다. 데이터 학습부(1310) 및 데이터 인식부(1320) 중 적어도 하나가 소프트웨어 모듈(또는, 인스터력션(instruction) 포함하는 프로그램 모듈)로 구현되는 경우, 소프트웨어 모듈은 컴퓨터로 읽을 수 있는 판독 가능한 비일시적 판독 가능 기록매체(non-transitory computer readable media)에 저장될 수 있다. 또한, 이 경우, 적어도 하나의 소프트웨어 모듈은 OS(Operating System)에 의해 제공되거나, 소정의 애플리케이션에 의해 제공될 수 있다. 또는, 적어도 하나의 소프트웨어 모듈 중 일부는 OS(Operating System)에 의해 제공되고, 나머지 일부는 소정의 애플리케이션에 의해 제공될 수 있다.

[189] 도 20은 일부 실시예에 따른 데이터 학습부(1310)의 블록도이다.

[190] 도 20을 참조하면, 일부 실시예에 따른 데이터 학습부(1310)는 데이터 획득부(1310-1), 전처리부(1310-2), 학습 데이터 선택부(1310-3), 모델

학습부(1310-4) 및 모델 평가부(1310-5)를 포함할 수 있다.

[191] 데이터 획득부(1310-1)는 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위하여 필요한 데이터를 획득할 수 있다. 데이터 획득부(1310-1)는, 예를 들어, 소정의 사용자 음성 혹은 음악에 관한 정보를 획득할 수 있다.

[192] 전처리부(1310-2)는 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위한 학습에 획득된 데이터가 이용될 수 있도록, 획득된 데이터를 전처리할 수 있다. 전처리부(1310-2)는 후술할 모델 학습부(1310-4)가 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위한 학습을 위하여 획득된 데이터를 이용할 수 있도록, 획득된 데이터를 기 설정된 포맷으로 가공할 수 있다.

[193] 학습 데이터 선택부(1310-3)는 전처리된 데이터 중에서 학습에 필요한 데이터를 선택할 수 있다. 선택된 데이터는 모델 학습부(1310-4)에 제공될 수 있다. 학습 데이터 선택부(1310-3)는 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위하여 기 설정된 기준에 따라, 전처리된 데이터 중에서 학습에 필요한 데이터를 선택할 수 있다. 또한, 학습 데이터 선택부(1310-3)는 후술할 모델 학습부(1310-4)에 의한 학습에 의해 기 설정된 기준에 따라 데이터를 선택할 수도 있다.

[194] 모델 학습부(1310-4)는 학습 데이터에 기초하여 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위한 기준을 학습할 수 있다. 또한, 모델 학습부(1310-4)는 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위하여 어떤 학습 데이터를 이용해야 하는 지에 대한 기준을 학습할 수도 있다.

[195] 또한, 모델 학습부(1310-4)는 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위하여 이용되는 데이터 인식 모델을 학습 데이터를 이용하여 학습시킬 수 있다. 이 경우, 데이터 인식 모델은 미리 구축된 모델일 수 있다. 예를 들어, 데이터 인식 모델은 기본 학습 데이터(예를 들어, 샘플 데이터 등)을 입력 받아 미리 구축된 모델일 수 있다.

[196] 데이터 인식 모델은, 인식 모델의 적용 분야, 학습의 목적 또는 장치의 컴퓨터 성능 등을 고려하여 구축될 수 있다. 데이터 인식 모델은 인간의 뇌 구조를 컴퓨터 상에서 모의하도록 설계될 수 있다. 데이터 인식 모델은 인간의 신경망의 뉴런(neuron)을 모의하는, 가중치를 가지는 복수의 네트워크 노드들을 포함할 수 있다. 복수의 네트워크 노드들은 뉴런이 시냅스(synapse)를 통하여 신호를 주고 받는 시냅틱(synaptic) 활동을 모의하도록 각각 연결 관계를 형성할 수 있다.

데이터 인식 모델은, 일 예로, 신경망 모델, 또는 신경망 모델에서 발전한 딥 러닝 모델을 포함할 수 있다. 딥 러닝 모델에서 복수의 네트워크 노드들은 서로 다른 깊이(또는, 레이어)에 위치하면서 컨볼루션(convolution) 연결 관계에 따라 데이터를 주고 받을 수 있다. 데이터 인식 모델에는, 예컨대, DNN(Deep Neural Network), RNN(Recurrent Neural Network), BRDNN(Bidirectional Recurrent Deep Neural Network)과 같은 모델 들이 포함될 수 있으며, 전술한 예에 특별히 한정되지는 않는다.

[197] 다양한 실시예에 따르면, 모델 학습부(1310-4)는 미리 구축된 데이터 인식 모델이 복수 개가 존재하는 경우, 입력된 학습 데이터와 기본 학습 데이터의 관련성이 큰 데이터 인식 모델을 학습할 데이터 인식 모델로 결정할 수 있다. 이 경우, 기본 학습 데이터는 데이터의 타입 별로 기 분류되어 있을 수 있으며, 데이터 인식 모델은 데이터의 타입 별로 미리 구축되어 있을 수 있다. 예를 들어, 기본 학습 데이터는 학습 데이터가 생성된 지역, 학습 데이터가 생성된 시간, 학습 데이터의 크기, 학습 데이터의 장르, 학습 데이터의 생성자, 학습 데이터 내의 오브젝트의 종류 등과 같은 다양한 기준으로 기 분류되어 있을 수 있다.

[198] 또한, 모델 학습부(1310-4)는, 예를 들어, 오류 역전파법(error back-propagation) 또는 경사 하강법(gradient descent)을 포함하는 학습 알고리즘 등을 이용하여 데이터 인식 모델을 학습시킬 수 있다.

[199] 또한, 모델 학습부(1310-4)는, 예를 들어, 학습 데이터를 입력 값으로 하는 지도 학습(supervised learning) 을 통하여, 데이터 인식 모델을 학습시킬 수 있다. 또한, 모델 학습부(1310-4)는, 예를 들어, 별다른 지도없이 상황 판단을 위해 필요한 데이터의 종류를 스스로 학습함으로써, 상황 판단을 위한 기준을 발견하는 비지도 학습(unsupervised learning)을 통하여, 데이터 인식 모델을 학습시킬 수 있다. 또한, 모델 학습부(1310-4)는, 예를 들어, 학습에 따른 상황 판단의 결과가 올바른 지에 대한 피드백을 이용하는 강화 학습(reinforcement learning)을 통하여, 데이터 인식 모델을 학습시킬 수 있다.

[200] 또한, 데이터 인식 모델이 학습되면, 모델 학습부(1310-4)는 학습된 데이터 인식 모델을 저장할 수 있다. 이 경우, 모델 학습부(1310-4)는 학습된 데이터 인식 모델을 데이터 인식부(1320)를 포함하는 전자 장치의 메모리에 저장할 수 있다. 또는, 모델 학습부(1310-4)는 학습된 데이터 인식 모델을 후술할 데이터 인식부(1320)를 포함하는 전자 장치의 메모리에 저장할 수 있다. 또는, 모델 학습부(1310-4)는 학습된 데이터 인식 모델을 전자 장치와 유선 또는 무선 네트워크로 연결되는 서버의 메모리에 저장할 수도 있다.

[201] 이 경우, 학습된 데이터 인식 모델이 저장되는 메모리는, 예를 들면, 전자 장치의 적어도 하나의 다른 구성요소에 관계된 명령 또는 데이터를 함께 저장할 수도 있다. 또한, 메모리는 소프트웨어 및/또는 프로그램을 저장할 수도 있다. 프로그램은, 예를 들면, 커널, 미들웨어, 어플리케이션 프로그래밍 인터페이스(API) 및/또는 어플리케이션 프로그램(또는 "어플리케이션") 등을

포함할 수 있다.

[202] 모델 평가부(1310-5)는 데이터 인식 모델에 평가 데이터를 입력하고, 평가 데이터로부터 출력되는 인식 결과가 소정 기준을 만족하지 못하는 경우, 모델 학습부(1310-4)로 하여금 다시 학습하도록 할 수 있다. 이 경우, 평가 데이터는 데이터 인식 모델을 평가하기 위한 기 설정된 데이터일 수 있다.

[203] 예를 들어, 모델 평가부(1310-5)는 평가 데이터에 대한 학습된 데이터 인식 모델의 인식 결과 중에서, 인식 결과가 정확하지 않은 평가 데이터의 개수 또는 비율이 미리 설정된 임계치를 초과하는 경우 소정 기준을 만족하지 못한 것으로 평가할 수 있다. 예컨대, 소정 기준이 비율 2%로 정의되는 경우, 학습된 데이터 인식 모델이 총 1000개의 평가 데이터 중의 20개를 초과하는 평가 데이터에 대하여 잘못된 인식 결과를 출력하는 경우, 모델 평가부(1310-5)는 학습된 데이터 인식 모델이 적합하지 않은 것으로 평가할 수 있다.

[204] 한편, 학습된 데이터 인식 모델이 복수 개가 존재하는 경우, 모델 평가부(1310-5)는 각각의 학습된 동영상 인식 모델에 대하여 소정 기준을 만족하는지를 평가하고, 소정 기준을 만족하는 모델을 최종 데이터 인식 모델로서 결정할 수 있다. 이 경우, 소정 기준을 만족하는 모델이 복수 개인 경우, 모델 평가부(1310-5)는 평가 점수가 높은 순으로 미리 설정된 어느 하나 또는 소정 개수의 모델을 최종 데이터 인식 모델로서 결정할 수 있다.

[205] 한편, 데이터 학습부(1310) 내의 데이터 획득부(1310-1), 전처리부(1310-2), 학습 데이터 선택부(1310-3), 모델 학습부(1310-4) 및 모델 평가부(1310-5) 중 적어도 하나는, 적어도 하나의 하드웨어 칩 형태로 제작되어 전자 장치에 탑재될 수 있다. 예를 들어, 데이터 획득부(1310-1), 전처리부(1310-2), 학습 데이터 선택부(1310-3), 모델 학습부(1310-4) 및 모델 평가부(1310-5) 중 적어도 하나는 인공 지능(AI; artificial intelligence)을 위한 전용 하드웨어 칩 형태로 제작될 수도 있고, 또는 기존의 범용 프로세서(예: CPU 또는 application processor) 또는 그래픽 전용 프로세서(예: GPU)의 일부로 제작되어 전술한 각종 전자 장치에 탑재될 수도 있다.

[206] 또한, 데이터 획득부(1310-1), 전처리부(1310-2), 학습 데이터 선택부(1310-3), 모델 학습부(1310-4) 및 모델 평가부(1310-5)는 하나의 전자 장치에 탑재될 수도 있으며, 또는 별개의 전자 장치들에 각각 탑재될 수도 있다. 예를 들어, 데이터 획득부(1310-1), 전처리부(1310-2), 학습 데이터 선택부(1310-3), 모델 학습부(1310-4) 및 모델 평가부(1310-5) 중 일부는 전자 장치에 포함되고, 나머지 일부는 서버에 포함될 수 있다.

[207] 또한, 데이터 획득부(1310-1), 전처리부(1310-2), 학습 데이터 선택부(1310-3), 모델 학습부(1310-4) 및 모델 평가부(1310-5) 중 적어도 하나는 소프트웨어 모듈로 구현될 수 있다. 데이터 획득부(1310-1), 전처리부(1310-2), 학습 데이터 선택부(1310-3), 모델 학습부(1310-4) 및 모델 평가부(1310-5) 중 적어도 하나가 소프트웨어 모듈(또는, 인스터력션(instruction) 포함하는 프로그램 모듈)로

구현되는 경우, 소프트웨어 모듈은 컴퓨터로 읽을 수 있는 판독 가능한 비일시적 판독 가능 기록매체(non-transitory computer readable media)에 저장될 수 있다. 또한, 이 경우, 적어도 하나의 소프트웨어 모듈은 OS(Operating System)에 의해 제공되거나, 소정의 애플리케이션에 의해 제공될 수 있다. 또는, 적어도 하나의 소프트웨어 모듈 중 일부는 OS(Operating System)에 의해 제공되고, 나머지 일부는 소정의 애플리케이션에 의해 제공될 수 있다.

[208] 도 21은 일부 실시예에 따른 데이터 인식부(1320)의 블록도이다.

[209] 도 21을 참조하면, 일부 실시예에 따른 데이터 인식부(1320)는 데이터 획득부(1320-1), 전처리부(1320-2), 인식 데이터 선택부(1320-3), 인식 결과 제공부(1320-4) 및 모델 갱신부(1320-5)를 포함할 수 있다.

[210] 데이터 획득부(1320-1)는 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위하여 필요한 데이터를 획득할 수 있으며, 전처리부(1320-2)는 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위해 획득된 데이터가 이용될 수 있도록, 획득된 데이터를 전처리할 수 있다. 전처리부(1320-2)는 후술할 인식 결과 제공부(1320-4)가 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위하여 획득된 데이터를 이용할 수 있도록, 획득된 데이터를 기 설정된 포맷으로 가공할 수 있다.

[211] 인식 데이터 선택부(1320-3)는 전처리된 데이터 중에서 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위해 필요한 데이터를 선택할 수 있다. 선택된 데이터는 인식 결과 제공부(1320-4)에게 제공될 수 있다. 인식 데이터 선택부(1320-3)는 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위하여 기 설정된 기준에 따라, 전처리된 데이터 중에서 일부 또는 전부를 선택할 수 있다. 또한, 인식 데이터 선택부(1320-3)는 후술할 모델 학습부(1310-4)에 의한 학습에 의해 기 설정된 기준에 따라 데이터를 선택할 수도 있다.

[212] 인식 결과 제공부(1320-4)는 선택된 데이터를 데이터 인식 모델에 적용하여 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행할 수 있다. 인식 결과 제공부(1320-4)는 데이터의 인식 목적에 따른 인식 결과를 제공할 수 있다. 인식 결과 제공부(1320-4)는 인식 데이터 선택부(1320-3)에 의해 선택된 데이터를 입력 값으로 이용함으로써, 선택된 데이터를 데이터 인식 모델에 적용할 수 있다. 또한, 인식 결과는 데이터 인식 모델에 의해 결정될 수 있다.

[213] 모델 갱신부(1320-5)는 인식 결과 제공부(1320-4)에 의해 제공되는 인식 결과에 대한 평가에 기초하여, 데이터 인식 모델이 갱신되도록할 수 있다. 예를 들어, 모델 갱신부(1320-5)는 인식 결과 제공부(1320-4)에 의해 제공되는 인식 결과를

모델 학습부(1310-4)에게 제공함으로써, 모델 학습부(1310-4)가 데이터 인식 모델을 갱신하도록 할 수 있다.

[214] 한편, 데이터 인식부(1320) 내의 데이터 획득부(1320-1), 전처리부(1320-2), 인식 데이터 선택부(1320-3), 인식 결과 제공부(1320-4) 및 모델 갱신부(1320-5) 중 적어도 하나는, 적어도 하나의 하드웨어 칩 형태로 제작되어 전자 장치에 탑재될 수 있다. 예를 들어, 데이터 획득부(1320-1), 전처리부(1320-2), 인식 데이터 선택부(1320-3), 인식 결과 제공부(1320-4) 및 모델 갱신부(1320-5) 중 적어도 하나는 인공 지능(AI; artificial intelligence)을 위한 전용 하드웨어 칩 형태로 제작될 수도 있고, 또는 기존의 범용 프로세서(예: CPU 또는 application processor) 또는 그래픽 전용 프로세서(예: GPU)의 일부로 제작되어 전술한 각종 전자 장치에 탑재될 수도 있다.

[215] 또한, 데이터 획득부(1320-1), 전처리부(1320-2), 인식 데이터 선택부(1320-3), 인식 결과 제공부(1320-4) 및 모델 갱신부(1320-5)는 하나의 전자 장치에 탑재될 수도 있으며, 또는 별개의 전자 장치들에 각각 탑재될 수도 있다. 예를 들어, 데이터 획득부(1320-1), 전처리부(1320-2), 인식 데이터 선택부(1320-3), 인식 결과 제공부(1320-4) 및 모델 갱신부(1320-5) 중 일부는 전자 장치에 포함되고, 나머지 일부는 서버에 포함될 수 있다.

[216] 또한, 데이터 획득부(1320-1), 전처리부(1320-2), 인식 데이터 선택부(1320-3), 인식 결과 제공부(1320-4) 및 모델 갱신부(1320-5) 중 적어도 하나는 소프트웨어 모듈로 구현될 수 있다. 데이터 획득부(1320-1), 전처리부(1320-2), 인식 데이터 선택부(1320-3), 인식 결과 제공부(1320-4) 및 모델 갱신부(1320-5) 중 적어도 하나가 소프트웨어 모듈(또는, 인스터럭션(instruction) 포함하는 프로그램 모듈)로 구현되는 경우, 소프트웨어 모듈은 컴퓨터로 읽을 수 있는 판독 가능한 비일시적 판독 가능 기록매체(non-transitory computer readable media)에 저장될 수 있다. 또한, 이 경우, 적어도 하나의 소프트웨어 모듈은 OS(Operating System)에 의해 제공되거나, 소정의 애플리케이션에 의해 제공될 수 있다. 또는, 적어도 하나의 소프트웨어 모듈 중 일부는 OS(Operating System)에 의해 제공되고, 나머지 일부는 소정의 애플리케이션에 의해 제공될 수 있다.

[217] 도 22는 일부 실시예에 따른 디바이스(1000) 및 서버(2000)가 서로 연동함으로써 데이터를 학습하고 인식하는 예시를 나타내는 도면이다.

[218] 도 22를 참조하면, 서버(2000)는 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위한 기준을 학습할 수 있으며, 디바이스(1000)는 서버(2000)에 의한 학습 결과에 기초하여 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행할 수 있다.

[219] 이 경우, 서버(2000)의 모델 학습부(2340)는 도 29에 도시된 데이터 학습부(1310)의 기능을 수행할 수 있다. 서버(2000)의 모델 학습부(2340)는 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의

식별 중 적어도 하나의 동작을 수행하기 위하여 어떤 데이터를 이용할 지에 관한 기준을 학습할 수 있다. 또한, 서버의 모델 학습부(2340)는 데이터를 이용하여 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위한 기준을 학습할 수 있다. 모델 학습부(2340)는 학습에 이용될 데이터를 획득하고, 획득된 데이터를 후술할 데이터 인식 모델에 적용함으로써, 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위한 기준을 학습할 수 있다.

[220] 또한, 디바이스(1000)의 인식 결과 제공부(1320-4)는 인식 데이터 선택부(1320-3)에 의해 선택된 데이터를 서버(2000)에 의해 생성된 데이터 인식 모델에 적용하여 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행할 수 있다. 예를 들어, 인식 결과 제공부(1320-4)는 인식 데이터 선택부(1320-3)에 의해 선택된 데이터를 서버(2000)에게 전송하고, 서버(2000)가 인식 데이터 선택부(1320-3)에 의해 선택된 데이터를 인식 모델에 적용하여 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행할 것을 요청할 수 있다. 또한, 인식 결과 제공부(1320-4)는 서버(2000)에 의해 판단된 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행하기 위한 정보를 서버(2000)로부터 수신할 수 있다.

[221] 또는, 디바이스(1000)의 인식 결과 제공부(1320-4)는 서버(2000)에 의해 생성된 인식 모델을 서버(2000)로부터 수신하고, 수신된 인식 모델을 이용하여 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행할 수 있다. 이 경우, 디바이스(1000)의 인식 결과 제공부(1320-4)는 인식 데이터 선택부(1320-3)에의해 선택된 데이터를 서버(2000)로부터 수신된 데이터 인식 모델에 적용하여 소정의 데이터로부터 사용자의 음성 모델, 음성 입력의 내용 및 입력된 음악의 식별 중 적어도 하나의 동작을 수행할 수 있다.

[222] 또한, 디바이스(1000) 및 서버(2000)는 데이터 인식 모델의 학습 및 데이터 인식을 위한 작업을 효과적으로 분배하여 수행할 수 있으며, 이를 통하여, 사용자의 의도에 부합하는 서비스를 제공하기 위하여 데이터 처리를 효율적으로 수행하고, 사용자의 프라이버시를 효과적으로 보호할 수 있다.

[223] 일부 실시예는 컴퓨터로 읽을 수 있는 저장 매체(computer-readable storage media)에 저장된 명령어들을 포함하는 S/W 프로그램으로 구현될 수 있다.

[224] 예로, 컴퓨터는, 저장 매체로부터 저장된 명령어를 호출하고, 호출된 명령어에 따라 개시된 실시예에 따른 동작이 가능한 장치로서, 개시된 실시예들에 따른 디바이스 또는 디바이스와 통신 연결된 외부의 서버를 포함할 수 있다.

[225] 컴퓨터로 읽을 수 있는 저장매체는, 비일시적(non-transitory) 저장매체의

형태로 제공될 수 있다. 여기서, '비일시적'은 저장매체가 신호(signal), 전류(current)를 포함하지 않으며, 실재(tangible)한다는 것을 의미할 뿐 데이터가 저장매체에 반영구적 또는 임시적으로 저장됨을 구분하지 않는다. 예로, 비일시적 저장 매체는 CD, DVD, 하드 디스크, 블루레이 디스크, USB, 내장 메모리, 메모리 카드, ROM 또는 RAM 등과 같은 비일시적 판독가능 저장매체뿐만 아니라 레지스터, 캐쉬, 버퍼 등과 같이 임시적으로 저장되는 매체를 포함할 수 있다.

[226] 또한, 개시된 실시예들에 따른 방법은 컴퓨터 프로그램 제품(computer program product)으로 제공될 수 있다.

[227] 컴퓨터 프로그램 제품은 S/W 프로그램, S/W 프로그램이 저장된 컴퓨터로 읽을 수 있는 저장 매체 또는 판매자 및 구매자 간에 거래되는 상품을 포함할 수 있다.

[228] 예를 들어, 컴퓨터 프로그램 제품은 디바이스 제조사 또는 전자 마켓(예, 구글 플레이 스토어, 앱 스토어)을 통해 전자적으로 배포되는 S/W 프로그램 형태의 상품(예, 다운로더블 앱)을 포함할 수 있다. 전자적 배포를 위하여, S/W 프로그램의 적어도 일부는 저장 매체에 저장되거나, 임시적으로 생성될 수 있다. 이 경우, 저장 매체는 제조사 또는 전자 마켓의 서버, 또는 중계 서버의 저장매체가 될 수 있다.

청구범위

[청구항 1] 소리를 출력하는 디바이스에 있어서:
상기 소리를 출력하는 스피커;
하나 이상의 인스트럭션을 저장하는 메모리; 및
상기 하나 이상의 인스트럭션을 실행하는 프로세서;
를 포함하며,
상기 프로세서는, 상기 하나 이상의 인스트럭션을 실행함으로써, 외부 환경으로부터 수신될 외부 소리를 예측하고, 상기 예측된 외부 소리에 기반하여 상기 스피커로부터 출력될 소리를 가변적으로 조절하며, 상기 조절된 소리가 출력되도록 상기 스피커를 제어하는, 디바이스.

[청구항 2] 제 1항에 있어서,
상기 프로세서는 상기 예측된 외부 소리와, 상기 출력될 소리의 적어도 한 부분을, 주파수 및 시간 중 적어도 하나의 대역에서 분리하여, 상기 출력될 소리를 가변적으로 조절하는 디바이스.

[청구항 3] 제 2항에 있어서,
상기 프로세서는 상기 출력될 소리를 필터링하기 위한 적어도 하나의 주파수 대역을 설정하고, 상기 적어도 하나의 주파수 대역에 기초하여 필터링된 소리가 출력되도록 상기 스피커를 제어하되, 상기 적어도 하나의 주파수 대역은 시간 대역에서 동적으로 조절되는, 디바이스.

[청구항 4] 제 3항에 있어서,
상기 프로세서는 상기 적어도 하나의 주파수 대역의 수를 설정하고, 상기 적어도 하나의 주파수 대역의 폭 및 중심 주파수 중 적어도 하나를 설정하여, 상기 출력될 소리를 필터링하기 위한 적어도 하나의 주파수 대역을 설정하는, 디바이스.

[청구항 5] 제 2항에 있어서,
상기 출력될 소리는 적어도 하나의 악기 컴포넌트를 포함하는 음악 소리를 포함하며, 상기 프로세서는 상기 적어도 하나의 악기 컴포넌트 중 적어도 하나의 악기 컴포넌트를 주파수 및 시간 중 적어도 하나의 대역에서 조절하는, 디바이스.

[청구항 6] 제 1항에 있어서,
상기 예측된 외부 소리는 사용자가 발성할 제 1 음성을 포함하며,
상기 프로세서는 상기 사용자가 발성한 제 2 음성을 획득하고, 상기 획득한 제 2 음성에 기초하여 상기 제 1 음성을 예측하는, 디바이스.

[청구항 7] 제 1항에 있어서,
상기 예측된 외부 소리는 제 1 음악 소리를 포함하며,
상기 프로세서는 상기 외부 환경으로부터 제 2 음악 소리를 획득하고, 상기 제 2 음악 소리에 기초하여, 상기 제 2 음악 소리가 포함되는 이미

발행된 음악을 식별하며, 상기 식별된 음악에 기초하여 상기 제 1 음악 소리를 예측하는, 디바이스.

[청구항 8] 제 7항에 있어서,
상기 프로세서는 적어도 하나의 음악에 관한 정보를 포함하는 데이터베이스를 획득하고, 상기 데이터베이스에 포함된 적어도 하나의 음악에 관한 정보와 상기 획득한 제 2 음악 소리를 매칭하며, 상기 매칭 결과에 기초하여 상기 발행된 음악을 식별하는, 디바이스.

[청구항 9] 제 1항에 있어서,
상기 스피커로부터 출력될 소리는 음성 소리 및 음악 소리를 포함하고,
상기 프로세서는 상기 예측된 외부 소리, 상기 음성 소리 및 상기 음악 소리의 적어도 한 부분을, 주파수 및 시간 중 적어도 하나의 대역에서 분리하는, 디바이스.

[청구항 10] 제 9항에 있어서,
상기 프로세서는, 상기 음성 소리 및 상기 음악 소리에 서로 다른 필터를 적용하여, 상기 예측된 외부 소리, 상기 음성 소리 및 상기 음악 소리의 적어도 한 부분을, 주파수 및 시간 중 적어도 하나의 대역에서 분리하는, 디바이스.

[청구항 11] 디바이스를 이용하여 소리를 출력하는 방법에 있어서:
외부 환경으로부터 수신되는 외부 소리를 예측하는 단계;
상기 예측된 외부 소리에 기반하여 상기 디바이스로부터 출력될 소리를 가변적으로 조절하는 단계; 및
상기 조절된 소리를 출력하는 단계를 포함하는, 소리 출력 방법.

[청구항 12] 제 11항에 있어서,
상기 예측된 외부 소리에 기반하여 상기 출력될 소리를 가변적으로 조절하는 단계는
상기 예측된 외부 소리와, 상기 출력될 소리의 적어도 한 부분을, 주파수 및 시간 중 적어도 하나의 대역에서 분리하는 단계인 소리 출력 방법.

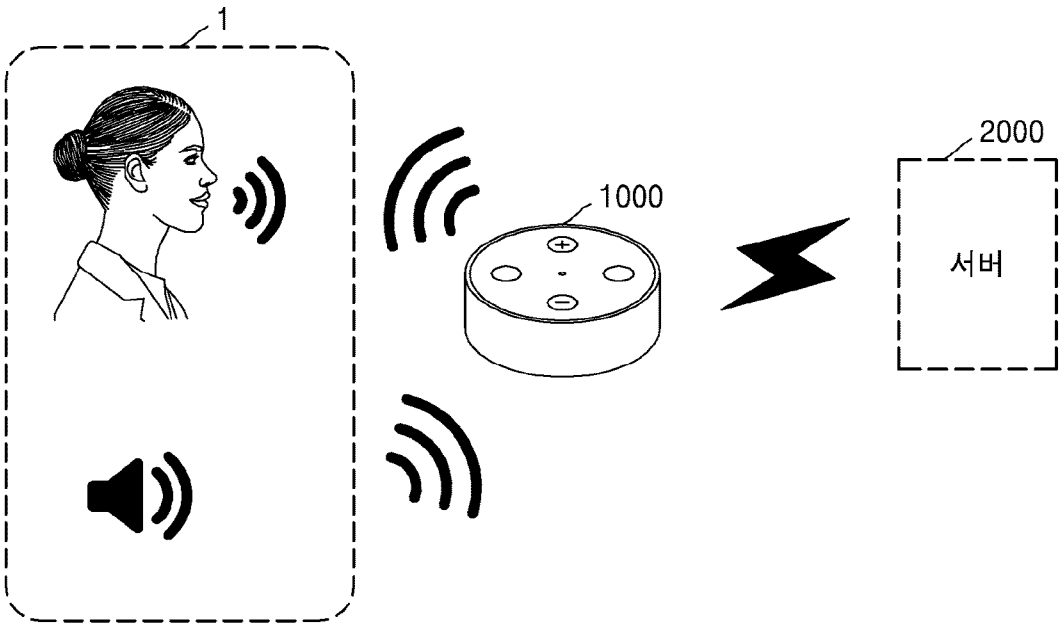
[청구항 13] 제 12항에 있어서,
상기 예측된 외부 소리와, 상기 출력될 소리의 적어도 한 부분을, 주파수 및 시간 중 적어도 하나의 대역에서 분리하는 단계는
상기 출력될 소리를 필터링하기 위한 적어도 하나의 주파수 대역을 설정하는 단계; 및
상기 적어도 하나의 주파수 대역에 기초하여 필터링된 소리를 출력하는 단계를 포함하며,
상기 적어도 하나의 주파수 대역은 시간 대역에서 동적으로 조절되는 소리 출력 방법.

[청구항 14] 제 13항에 있어서,
상기 출력될 소리를 필터링하기 위한 적어도 하나의 주파수 대역을

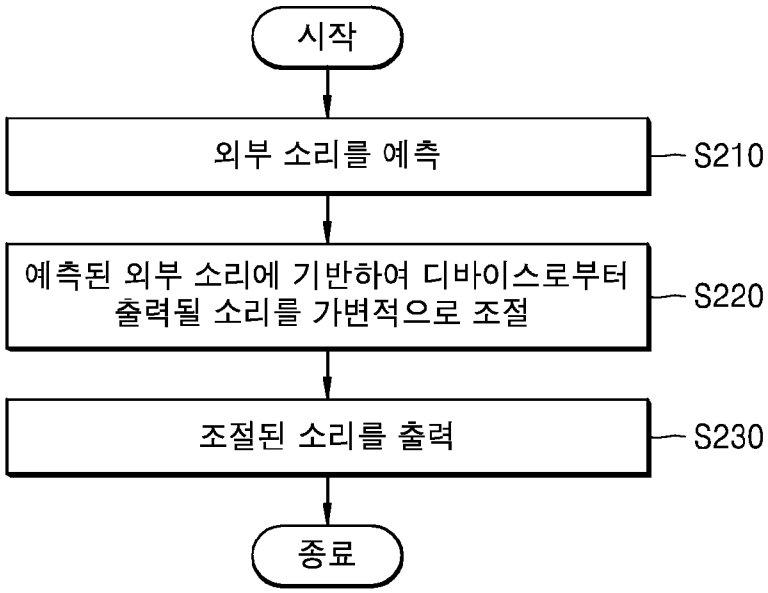
설정하는 단계는,
상기 적어도 하나의 주파수 대역의 수를 설정하는 단계; 및
상기 적어도 하나의 주파수 대역의 폭 및 중심 주파수 중 적어도 하나를 설정하는 단계를 포함하는 소리 출력 방법.

[청구항 15] 제 12항에 있어서,
상기 출력될 소리는 적어도 하나의 악기 컴포넌트를 포함하는 음악 소리를 포함하고, 상기 예측된 외부 소리와, 상기 출력될 소리의 적어도 한 부분을, 주파수 및 시간 중 적어도 하나의 대역에서 분리하는 단계는
상기 적어도 하나의 악기 컴포넌트 중 적어도 하나의 악기 컴포넌트를 주파수 및 시간 중 적어도 하나의 대역에서 조절하는 단계를 포함하는 소리 출력 방법.

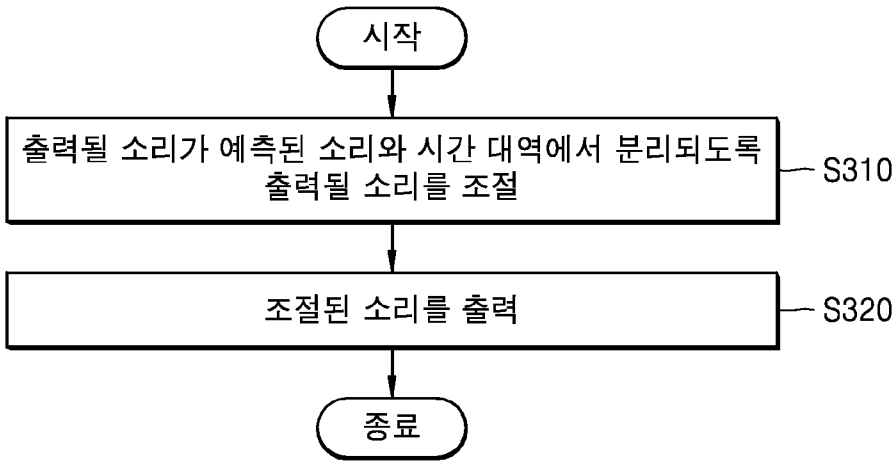
[도1]


1
1000
2000
서버

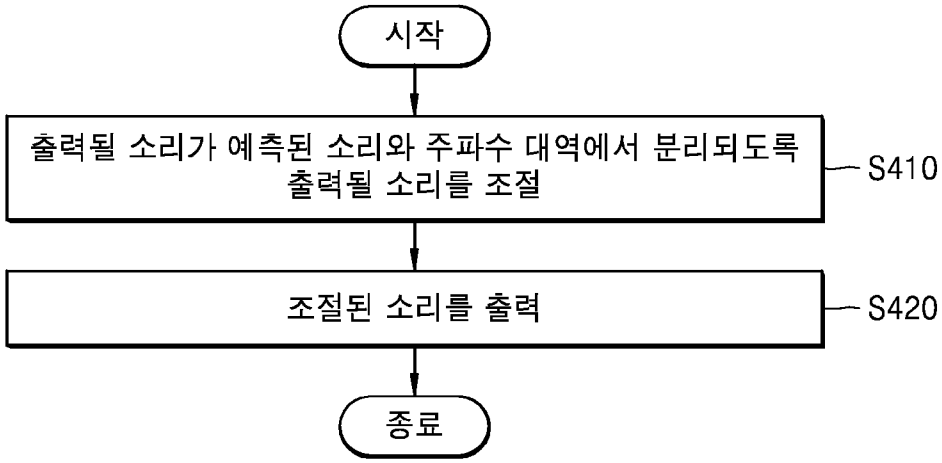
[도2]


시작
외부 소리를 예측
S210
예측된 외부 소리에 기반하여 디바이스로부터 출력될 소리를 가변적으로 조절
S220
조절된 소리를 출력
S230
종료

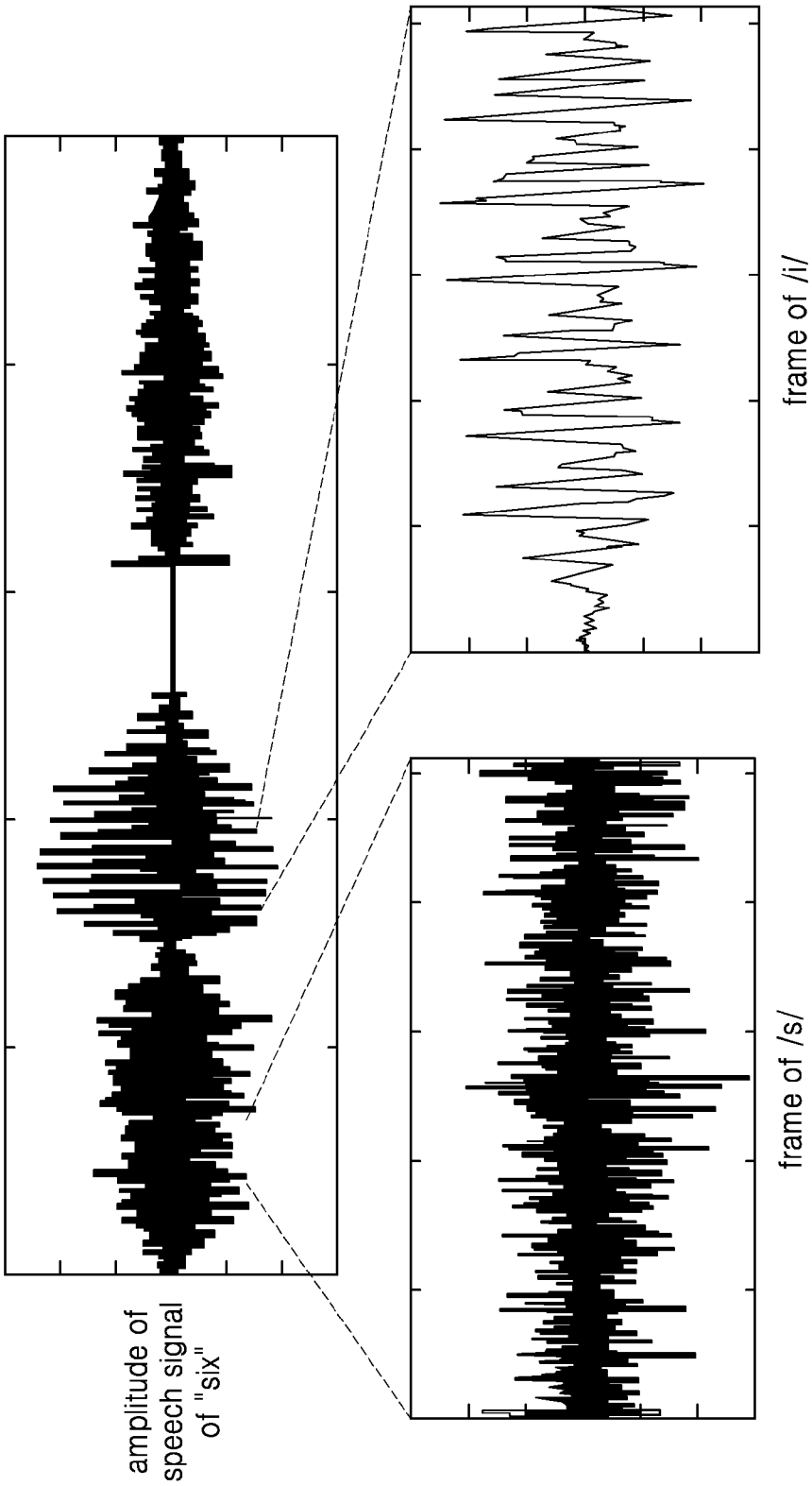
[도3]


시작
출력될 소리가 예측된 소리와 시간 대역에서 분리되도록 출력될 소리를 조절
S310
조절된 소리를 출력
S320
종료

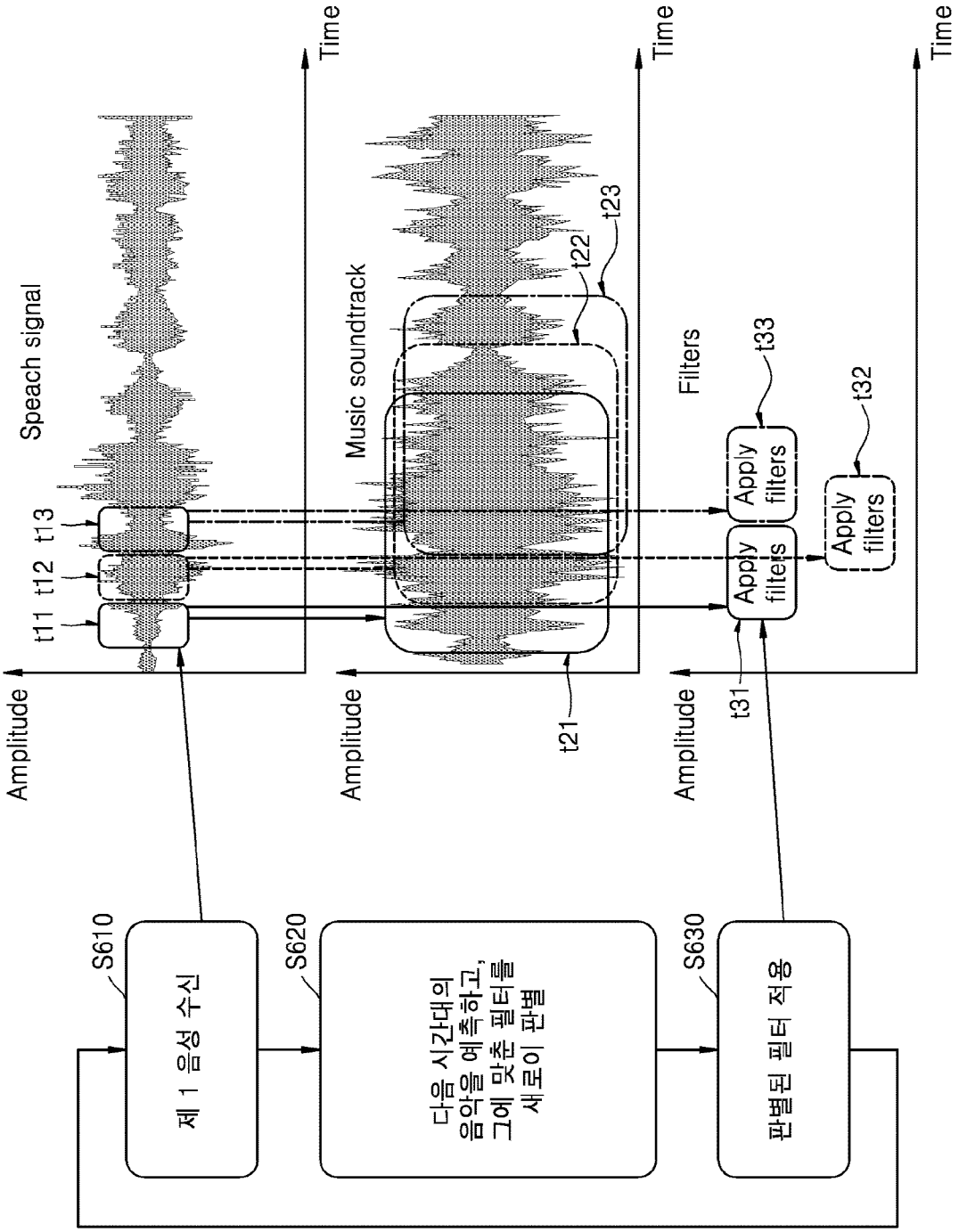
[도4]


시작
출력될 소리가 예측된 소리와 주파수 대역에서 분리되도록 출력될 소리를 조절
S410
조절된 소리를 출력
S420
종료

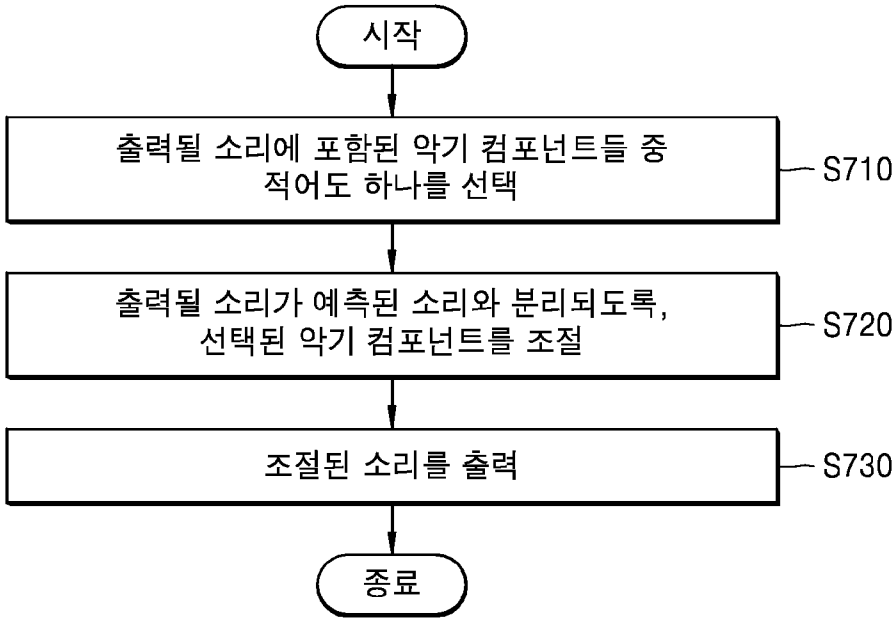
[도5]



[도6]

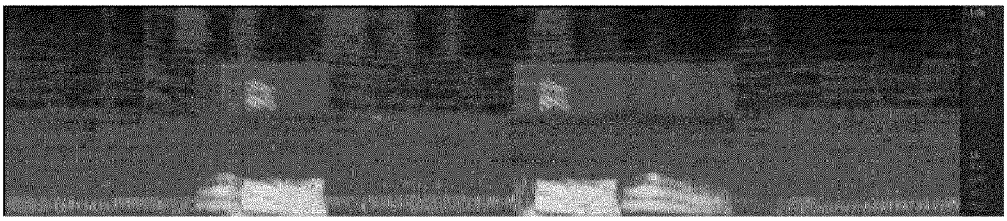

Speach signal
Music soundtrack
Filters
Time
Amplitude
t11
t12
t13
t21
t22
t23
t31
t32
t33
Apply filters
S610
S620
S630
제 1 음성 수신
다음 시간대의 음악을 예측하고, 그에 맞춘 필터를 새로이 판별
판별된 필터 적용

[도7]


시작
출력될 소리에 포함된 악기 컴포넌트들 중 적어도 하나를 선택
S710
출력될 소리가 예측된 소리와 분리되도록, 선택된 악기 컴포넌트를 조절
S720
조절된 소리를 출력
S730
종료

[도8]

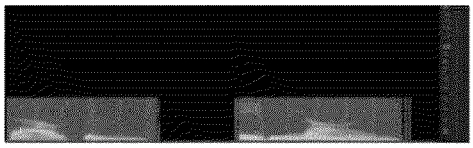




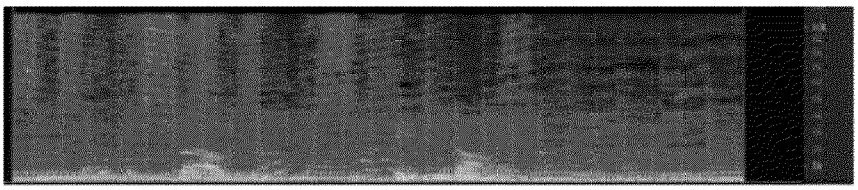
[도9]



Environment



Voice Assistant

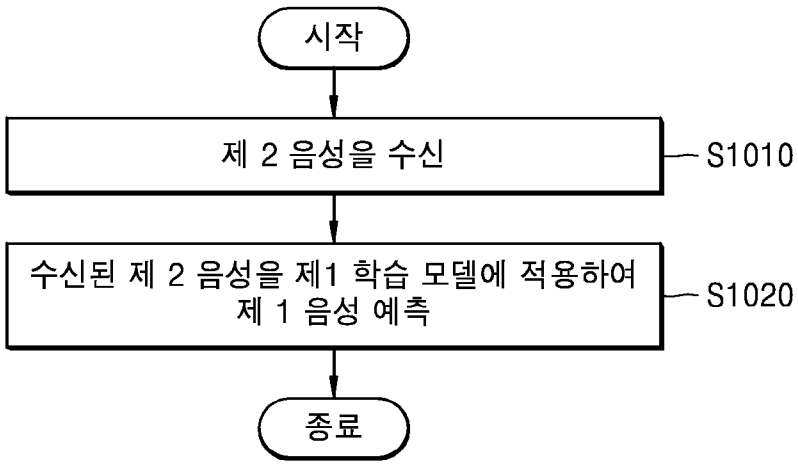


Signal „captured" by user (wihtout any modification)

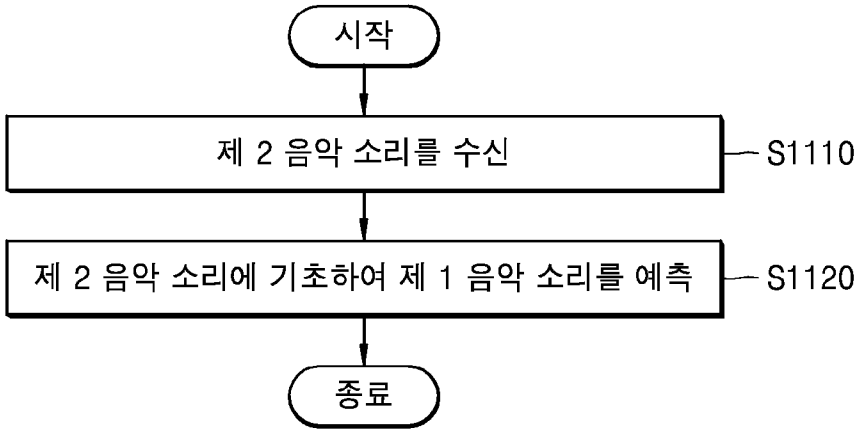


Signal „captured" by user (with modification of syntesised utterance)

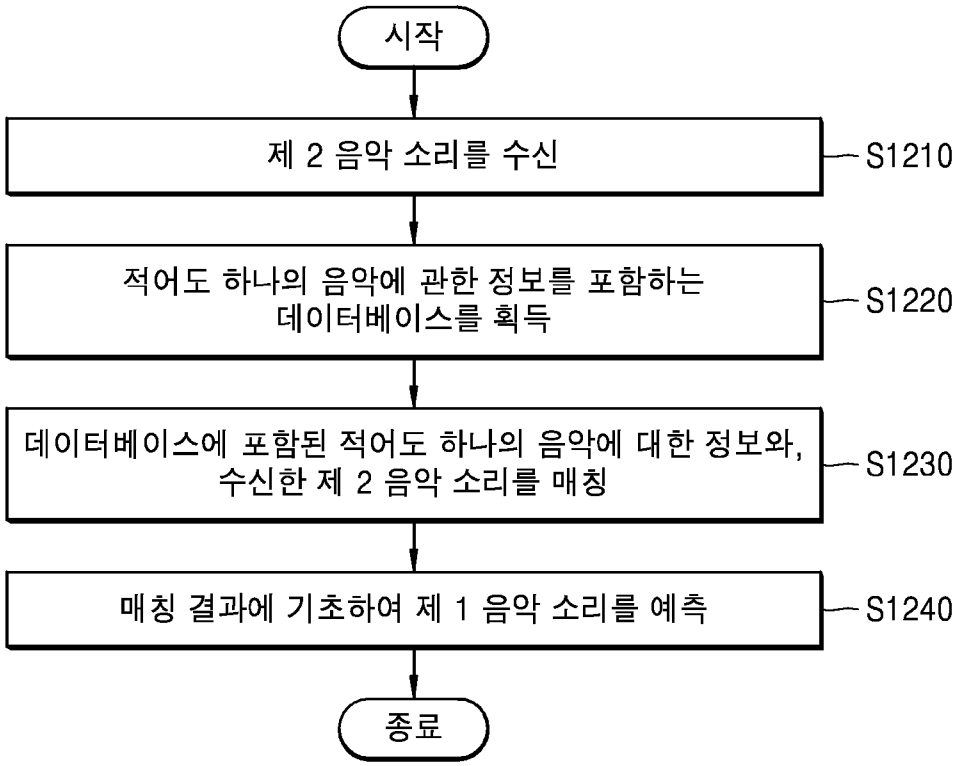
[도10]



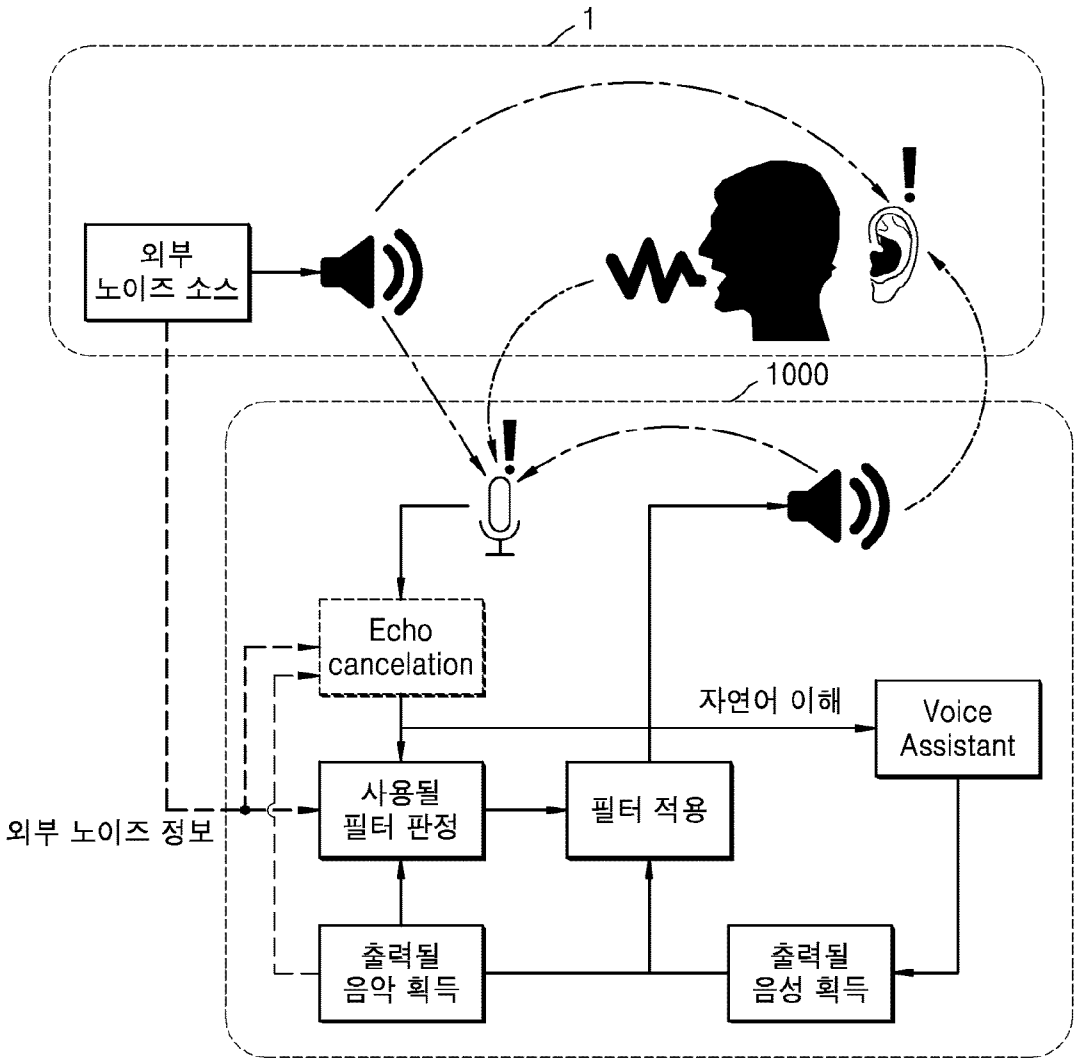
[도11]


시작
제 2 음악 소리를 수신
S1110
제 2 음악 소리에 기초하여 제 1 음악 소리를 예측
S1120
종료

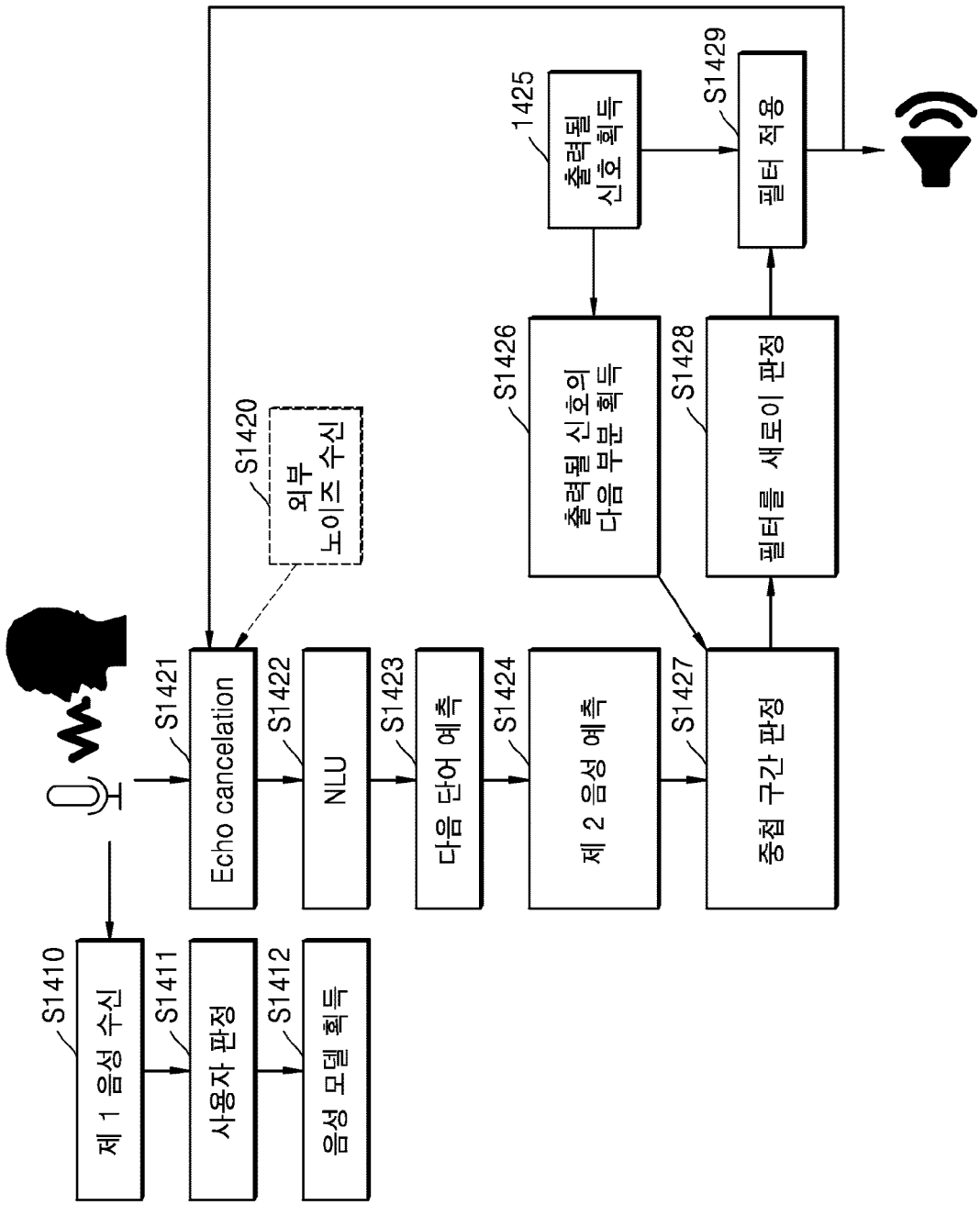
[도12]


시작
제 2 음악 소리를 수신
S1210
적어도 하나의 음악에 관한 정보를 포함하는
데이터베이스를 획득
S1220
데이터베이스에 포함된 적어도 하나의 음악에 대한 정보와,
수신한 제 2 음악 소리를 매칭
S1230
매칭 결과에 기초하여 제 1 음악 소리를 예측
S1240
종료

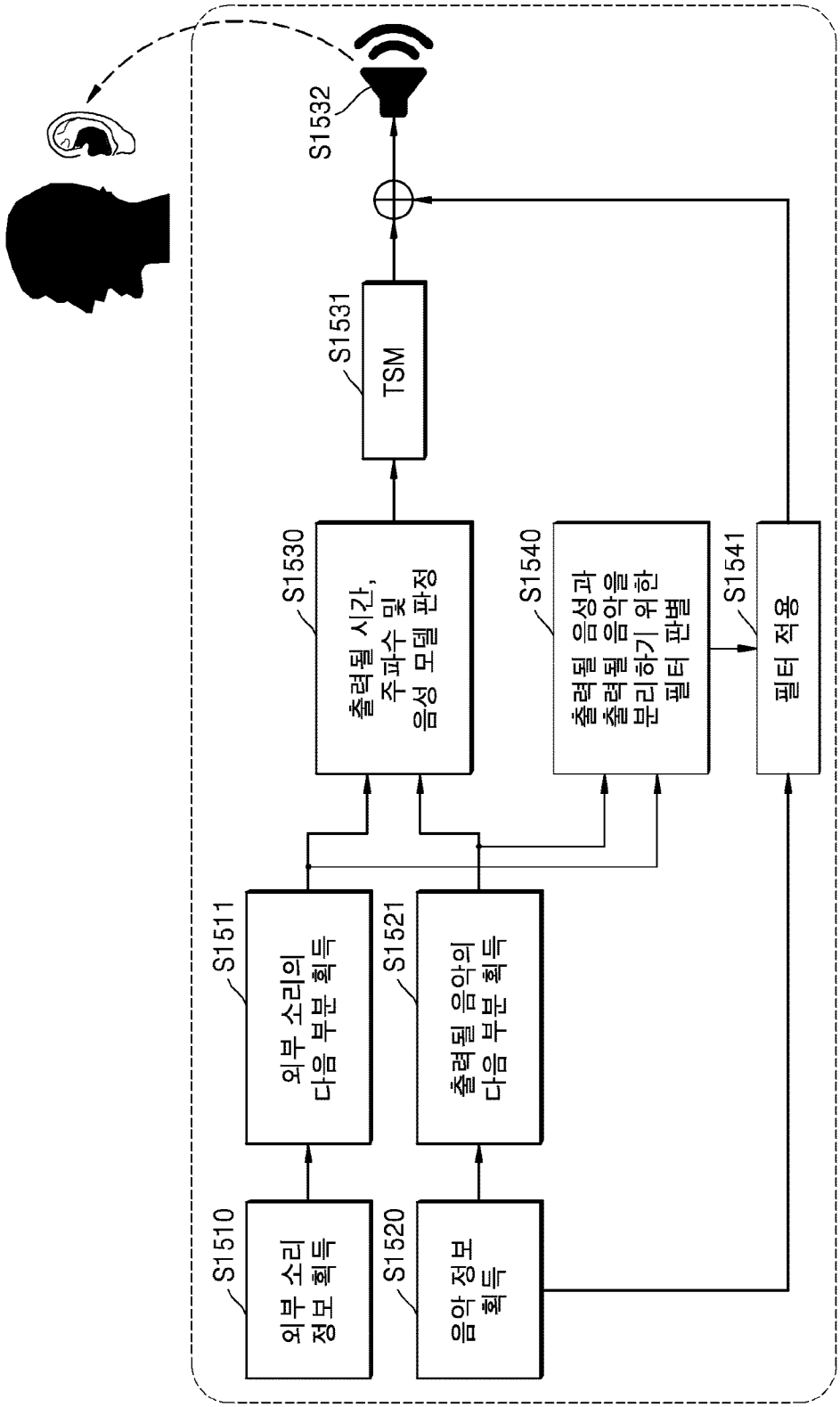
[도13]


1
외부
노이즈 소스
1000
Echo
cancelation
자연어 이해
Voice
Assistant
사용될
필터 판정
필터 적용
외부 노이즈 정보
출력될
음악 획득
출력될
음성 획득

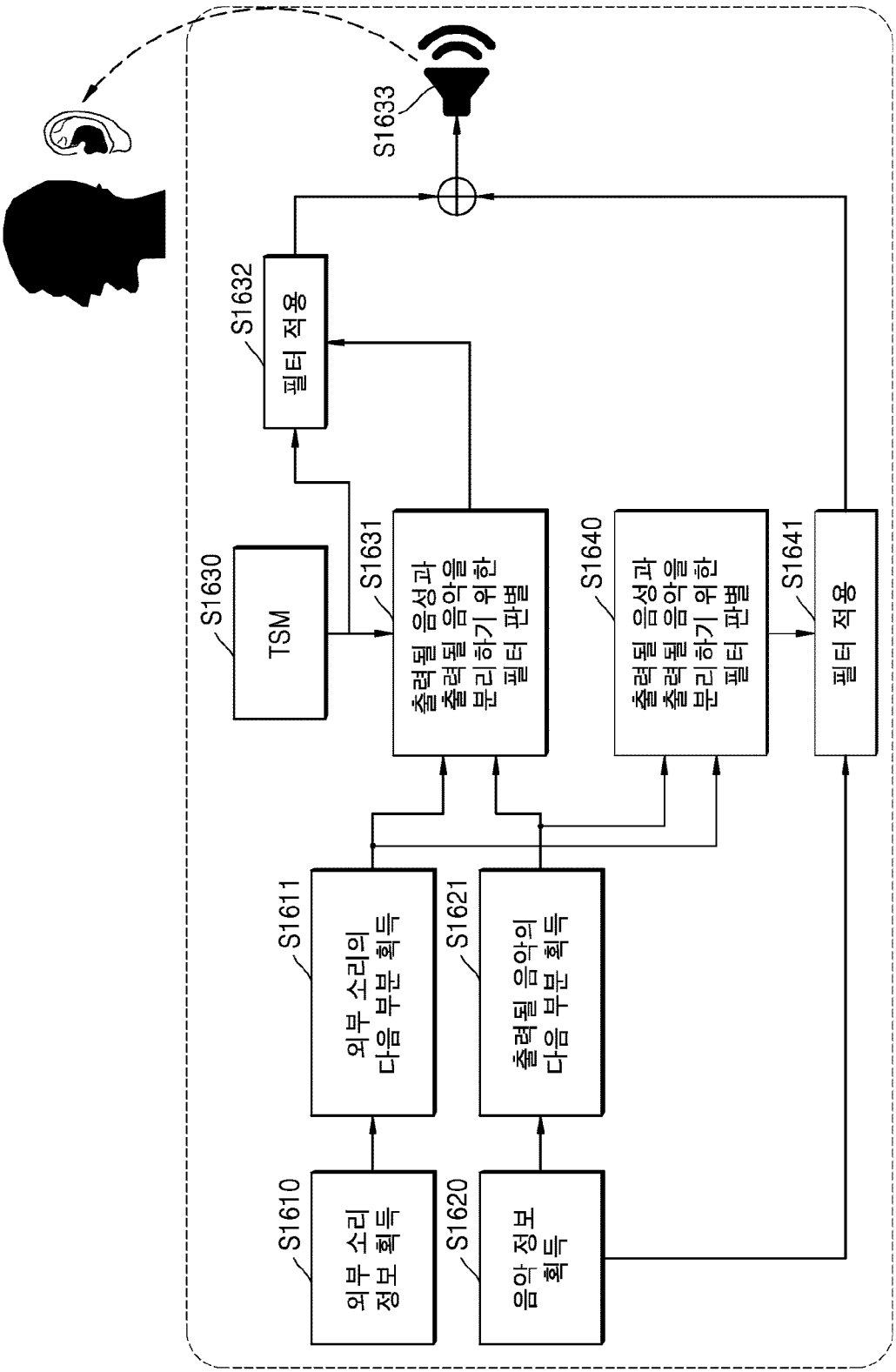
[도14]


S1410
제 1 음성 수신
S1411
사용자 판정
S1412
음성 모델 획득
S1421
Echo cancelation
S1422
NLU
S1423
다음 단어 예측
S1424
제 2 음성 예측
S1427
중첩 구간 판정
S1428
필터를 새로이 판정
S1426
출력될 신호의 다음 부분 획득
S1420
외부 노이즈 수신
1425
출력될 신호 획득
S1429
필터 적용

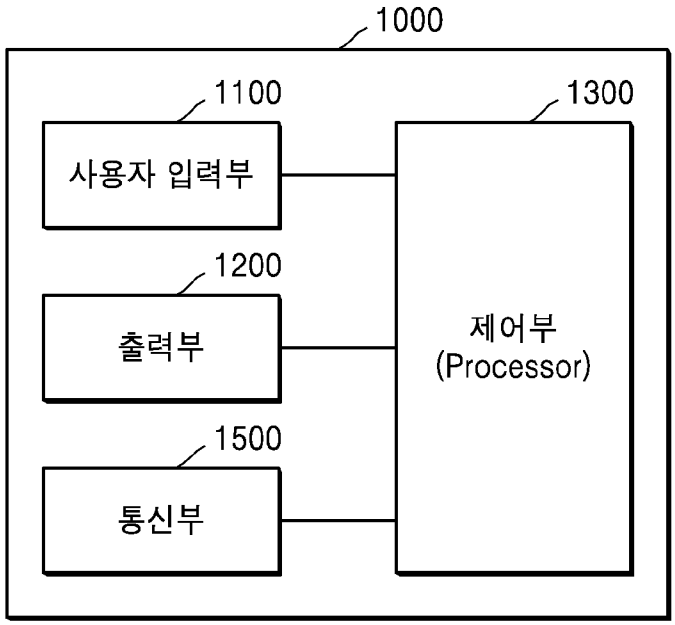
[도15]


S1532
S1531
TSM
S1530
출력될 시간,
주파수 및
음성 모델 판정
S1540
출력될 음성과
출력될 음악을
분리하기 위한
필터 판별
S1541
필터 적용
S1511
외부 소리의
다음 부분 획득
S1521
출력될 음악의
다음 부분 획득
S1510
외부 소리
정보 획득
S1520
음악 정보
획득

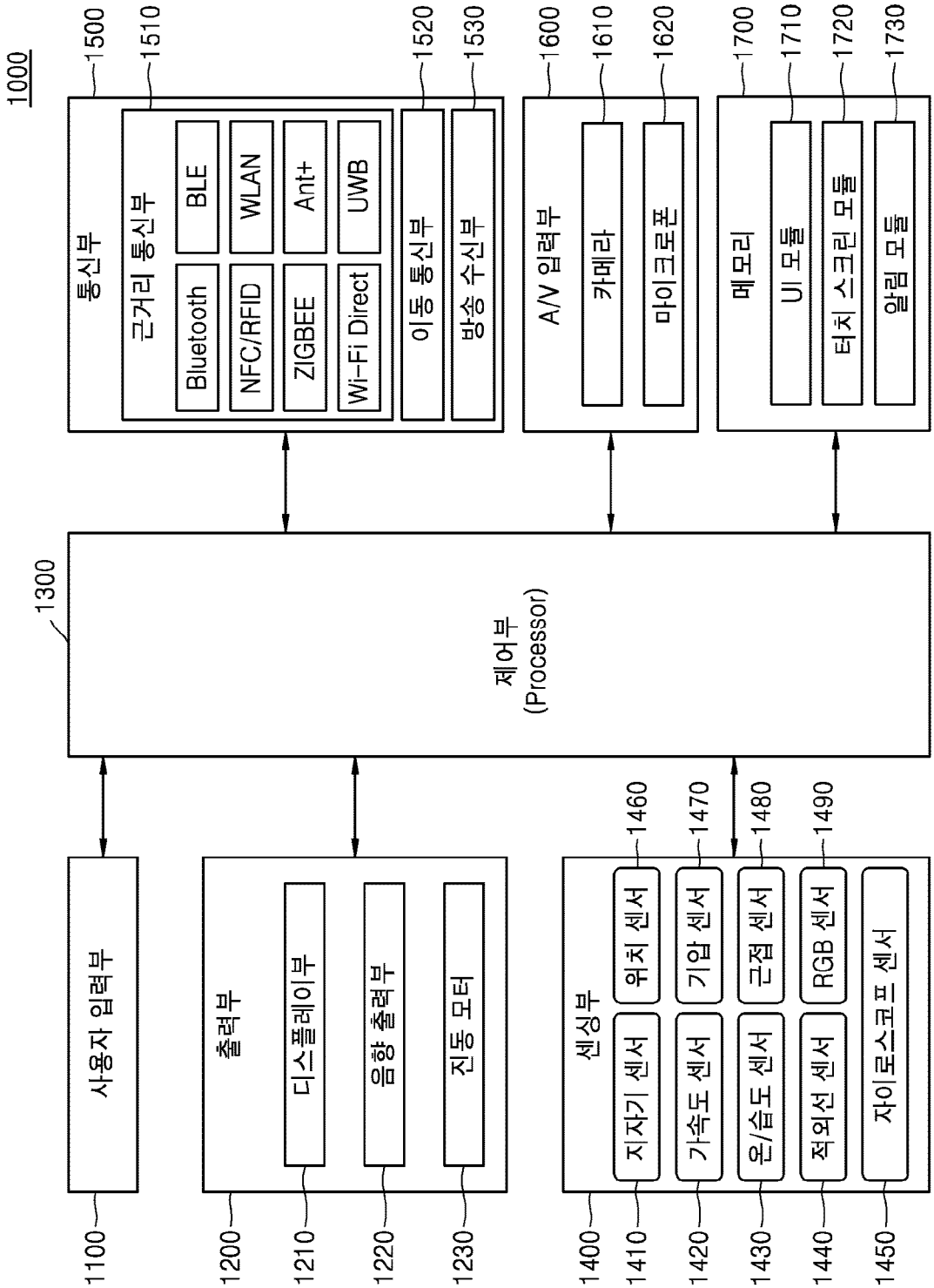
[도16]


S1633
S1632
필터 적용
S1630
TSM
S1631
출력될 음성과 출력될 음악을 분리하기 위한 필터 판별
S1640
출력될 음성과 출력될 음악을 분리하기 위한 필터 판별
S1641
필터 적용
S1611
외부 소리의 다음 부분 획득
S1621
출력될 음악의 다음 부분 획득
S1610
외부 소리 정보 획득
S1620
음악 정보 획득

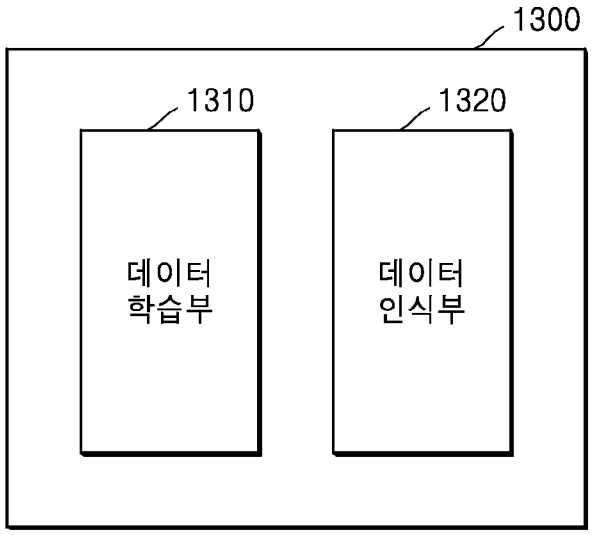
[도17]



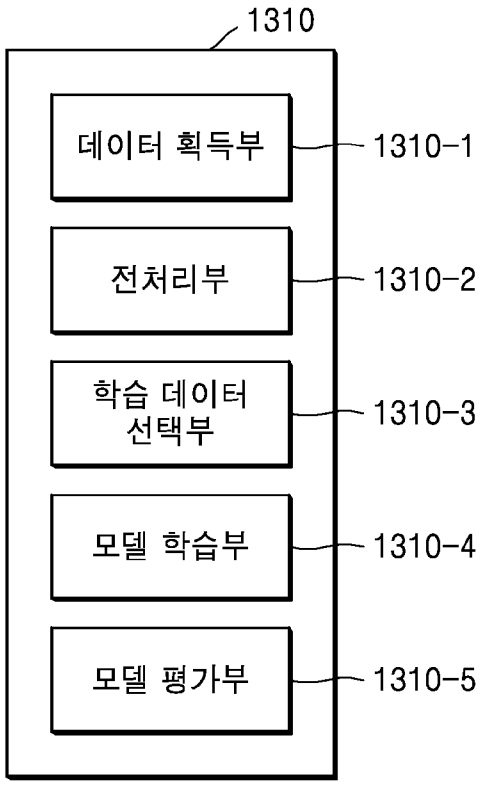
[도18]


1000
통신부
1500
근거리 통신부
1510
Bluetooth
BLE
NFC/RFID
WLAN
ZIGBEE
Ant+
Wi-Fi Direct
UWB
이동 통신부
1520
방송 수신부
1530
A/V 입력부
1600
카메라
1610
마이크로폰
1620
메모리
1700
UI 모듈
1710
터치 스크린 모듈
1720
알림 모듈
1730
1300
제어부
(Processor)
사용자 입력부
1100
출력부
1200
디스플레이부
1210
음향 출력부
1220
진동 모터
1230
센싱부
1400
지자기 센서
1410
위치 센서
1460
가속도 센서
1420
기압 센서
1470
온/습도 센서
1430
근접 센서
1480
적외선 센서
1440
RGB 센서
1490
자이로스코프 센서
1450

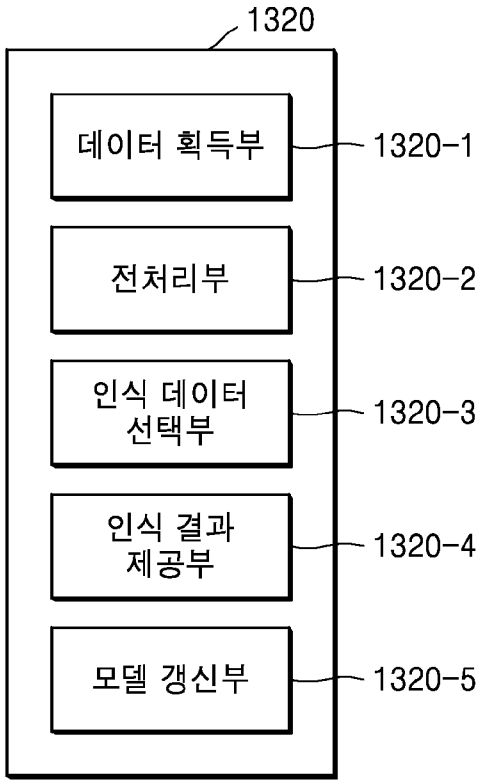
[도19]


1300
1310
1320
데이터 학습부
데이터 인식부

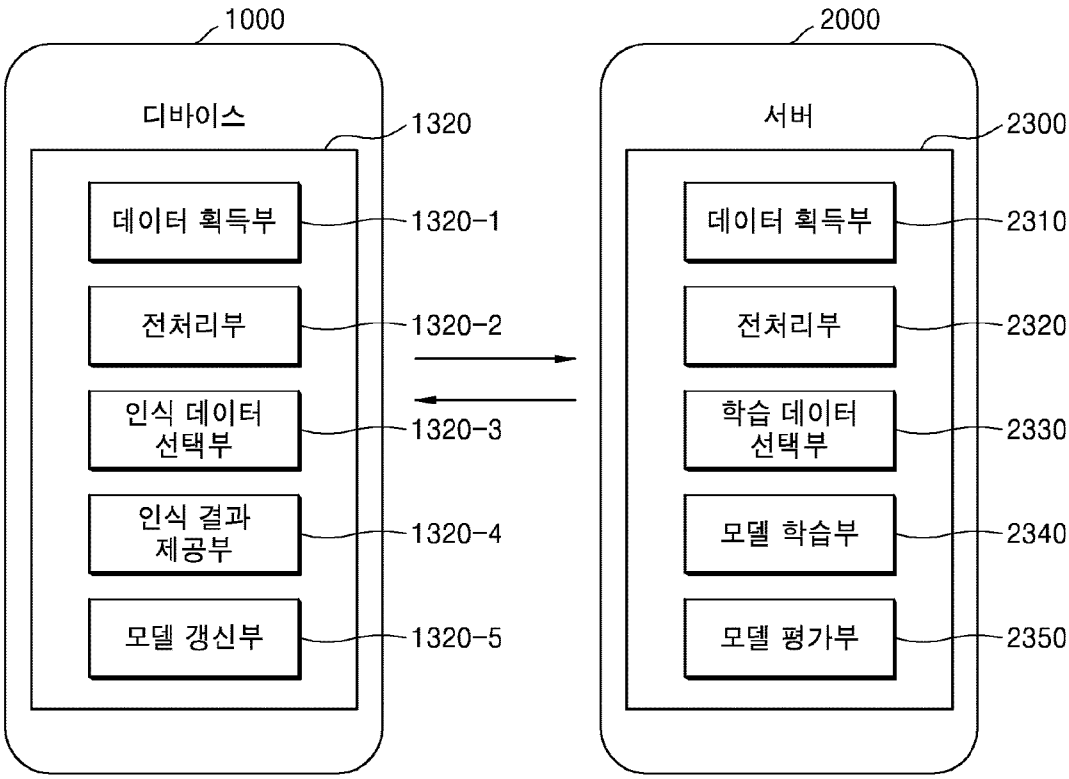
[도20]


1310
데이터 획득부
1310-1
전처리부
1310-2
학습 데이터 선택부
1310-3
모델 학습부
1310-4
모델 평가부
1310-5

[도21]


1320
데이터 획득부
1320-1
전처리부
1320-2
인식 데이터 선택부
1320-3
인식 결과 제공부
1320-4
모델 갱신부
1320-5

[도22]


1000
디바이스
1320
데이터 획득부
1320-1
전처리부
1320-2
인식 데이터 선택부
1320-3
인식 결과 제공부
1320-4
모델 갱신부
1320-5
2000
서버
2300
데이터 획득부
2310
전처리부
2320
학습 데이터 선택부
2330
모델 학습부
2340
모델 평가부
2350

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2019/007871

A. CLASSIFICATION OF SUBJECT MATTER

H04R 3/00(2006.01)i, G10L 21/0272(2013.01)i, G10L 19/04(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)

H04R 3/00; G06F 17/00; G10L 15/00; G10L 15/10; G10L 15/183; G10L 15/28; G10L 17/02; G10L 21/02; H04M 1/00; H04M 1/60; G10L 21/0272; G10L 19/04

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: electronic device, voice, speaker, predict

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KR 10-2018-0070970 A (SAMSUNG ELECTRONICS CO., LTD.) 27 June 2018 See paragraphs [0041]-[0058], [0147]-[0192] and figures 2-3, 9.	1,6,11
Y		2,5,9-10,12,15
A		3-4,7-8,13-14
Y	JP 3982823 B2 (TOSHIBA CORP.) 26 September 2007 See paragraphs [0033]-[0037] and figure 1.	2,5,9-10,12,15
A	KR 10-2009-0094576 A (OK, Jong Suk et al.) 08 September 2009 See paragraphs [0020]-[0052] and figures 1-3.	1-15
A	JP 2006-171077 A (NISSAN MOTOR CO., LTD.) 29 June 2006 See paragraphs [0011]-[0032] and figures 1-8.	1-15
A	US 2005-0102142 A1 (SOUFFLET, Frederic et al.) 12 May 2005 See paragraphs [0093]-[0149] and figures 1-3.	1-15

☐ 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
"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	Date of mailing of the international search report
04 OCTOBER 2019 (04.10.2019)	**04 OCTOBER 2019 (04.10.2019)**

Name and mailing address of the ISA/KR	Authorized officer
Korean Intellectual Property Office Government Complex Daejeon Building 4, 189, Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea Facsimile No. +82-42-481-8578	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/KR2019/007871

Patent document cited in search report	Publication date	Patent family member	Publication date
KR 10-2018-0070970 A	27/06/2018	EP 3501023 A1	26/06/2019
		US 2018-0174580 A1	21/06/2018
		WO 2018-117532 A1	28/06/2018
JP 3982823 B2	26/09/2007	JP 2005-198093 A	21/07/2005
KR 10-2009-0094576 A	08/09/2009	KR 10-1004913 B1	28/12/2010
JP 2006-171077 A	29/06/2006	JP 4608670 B2	12/01/2011
US 2005-0102142 A1	12/05/2005	CN 1228762 C	23/11/2005
		CN 1491412 A	21/04/2004
		DE 60222093 T2	05/06/2008
		EP 1362343 A1	19/11/2003
		EP 1362343 B1	29/08/2007
		ES 2291440 T3	01/03/2008
		FR 2820872 A1	16/08/2002
		FR 2820872 B1	16/05/2003
		JP 2004-530149 A	30/09/2004
		JP 4751569 B2	17/08/2011
		KR 10-0908358 B1	20/07/2009
		KR 10-2003-0076661 A	26/09/2003
		MX PA03007178 A	04/12/2003
		US 7983911 B2	19/07/2011
		WO 02-065454 A1	22/08/2002

국 제 조 사 보 고 서

국제출원번호
PCT/KR2019/007871

A. 발명이 속하는 기술분류(국제특허분류(IPC))

H04R 3/00(2006.01)i, G10L 21/0272(2013.01)i, G10L 19/04(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)
H04R 3/00; G06F 17/00; G10L 15/00; G10L 15/10; G10L 15/183; G10L 15/28; G10L 17/02; G10L 21/02; H04M 1/00; H04M 1/60; G10L 21/0272; G10L 19/04

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: 전자장치(electronic device), 음성(voice), 스피커(speaker), 예측(predict)

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X	KR 10-2018-0070970 A (삼성전자주식회사) 2018.06.27 단락 [0041]-[0058], [0147]-[0192] 및 도면 2-3, 9 참조.	1,6,11
Y		2,5,9-10,12,15
A		3-4,7-8,13-14
Y	JP 3982823 B2 (TOSHIBA CORP.) 2007.09.26 단락 [0033]-[0037] 및 도면 1 참조.	2,5,9-10,12,15
A	KR 10-2009-0094576 A (옥종석 등) 2009.09.08 단락 [0020]-[0052] 및 도면 1-3 참조.	1-15
A	JP 2006-171077 A (NISSAN MOTOR CO., LTD.) 2006.06.29 단락 [0011]-[0032] 및 도면 1-8 참조.	1-15
A	US 2005-0102142 A1 (FREDERIC SOUFFLET 등) 2005.05.12 단락 [0093]-[0149] 및 도면 1-3 참조.	1-15

☐ 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	국제조사보고서 발송일
2019년 10월 04일 (04.10.2019)	2019년 10월 04일 (04.10.2019)

ISA/KR의 명칭 및 우편주소
대한민국 특허청
(35208) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)
팩스 번호 +82-42-481-8578

심사관
강민정
전화번호 +82-42-481-8131

국제조사보고서
대응특허에 관한 정보

국제출원번호
PCT/KR2019/007871

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2018-0070970 A	2018/06/27	EP 3501023 A1	2019/06/26
		US 2018-0174580 A1	2018/06/21
		WO 2018-117532 A1	2018/06/28
JP 3982823 B2	2007/09/26	JP 2005-198093 A	2005/07/21
KR 10-2009-0094576 A	2009/09/08	KR 10-1004913 B1	2010/12/28
JP 2006-171077 A	2006/06/29	JP 4608670 B2	2011/01/12
US 2005-0102142 A1	2005/05/12	CN 1228762 C	2005/11/23
		CN 1491412 A	2004/04/21
		DE 60222093 T2	2008/06/05
		EP 1362343 A1	2003/11/19
		EP 1362343 B1	2007/08/29
		ES 2291440 T3	2008/03/01
		FR 2820872 A1	2002/08/16
		FR 2820872 B1	2003/05/16
		JP 2004-530149 A	2004/09/30
		JP 4751569 B2	2011/08/17
		KR 10-0908358 B1	2009/07/20
		KR 10-2003-0076661 A	2003/09/26
		MX PA03007178 A	2003/12/04
		US 7983911 B2	2011/07/19
		WO 02-065454 A1	2002/08/22