



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2018-0135884
(43) 공개일자 2018년12월21일

- (51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01) G10L 15/18 (2006.01)
G10L 15/22 (2006.01) G10L 17/22 (2013.01)
- (52) CPC특허분류
G06F 17/30026 (2013.01)
G06F 17/30029 (2013.01)
- (21) 출원번호 10-2018-7023580
- (22) 출원일자(국제) 2018년04월17일
심사청구일자 2018년08월16일
- (85) 번역문제출일자 2018년08월16일
- (86) 국제출원번호 PCT/US2018/027879
- (87) 국제공개번호 WO/2018/212885
국제공개일자 2018년11월22일
- (30) 우선권주장
62/506,981 2017년05월16일 미국(US)
(뒷면에 계속)

- (71) 출원인
애플 인크.
미국 캘리포니아 (우편번호 95014) 쿠퍼티노 원
애플 파크 웨이
- (72) 발명자
그레이엄, 데이비드 찬스
미국 95014 캘리포니아주 쿠퍼티노 애플 파크 웨이 1 애플 인크. 내
이라니, 사이러스 다니엘
미국 95014 캘리포니아주 쿠퍼티노 애플 파크 웨이 1 애플 인크. 내
(뒷면에 계속)
- (74) 대리인
장덕순, 백만기

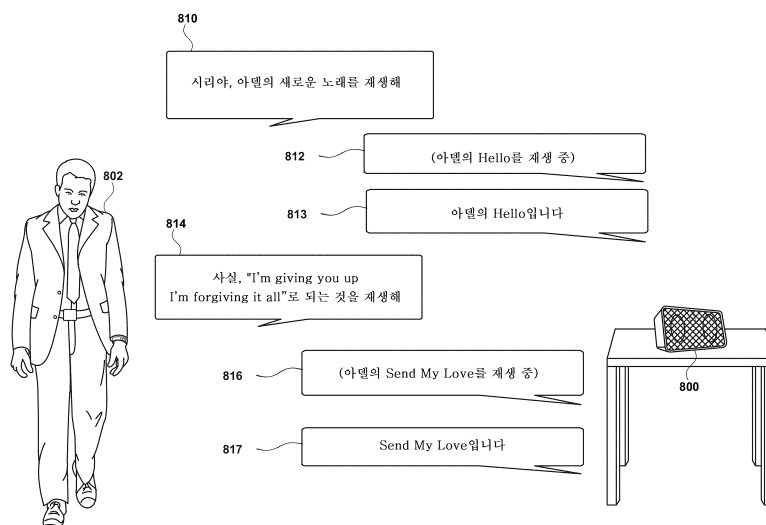
전체 청구항 수 : 총 100 항

(54) 발명의 명칭 미디어 탐색을 위한 지능형 자동화 어시스턴트

(57) 요약

지능형 자동화 어시스턴트를 동작시키기 위한 시스템들 및 프로세스들이 제공된다. 일례에 따르면, 방법은 하나 이상의 프로세서들 및 메모리를 구비한 전자 디바이스에서, 미디어에 대한 요청을 나타내는 제1 자연 언어 스피치 입력을 수신하는 것 - 제1 자연 언어 스피치 입력은 제1 검색 파라미터를 포함함 -; 및 디지털 어시스턴트에 의해, 제1 검색 파라미터에 기초하여 식별된 제1 미디어 아이템을 제공하는 것을 포함한다. 방법은, 제1 미디어 아이템을 제공하는 동안, 제2 자연 언어 스피치 입력을 수신하는 것 및 제2 입력이 미디어에 대한 요청을 정제하려는 사용자 의도에 대응하는지 여부를 결정하는 것을 추가로 포함한다. 방법은, 제2 스피치 입력이 미디어에 대한 요청을 정제하려는 사용자 의도에 대응한다는 결정에 따라, 제1 파라미터 및 제2 스피치 입력에 기초하여, 제2 미디어 아이템을 식별하는 것 및 제2 미디어 아이템을 제공하는 것을 추가로 포함한다.

대표도 - 도8a



(52) CPC특허분류

G06F 17/30976 (2013.01)

G10L 15/1815 (2013.01)

G10L 15/22 (2013.01)

G10L 17/22 (2013.01)

G10L 25/84 (2013.01)

(72) 발명자

피어시, 에이미

미국 95014 캘리포니아주 쿠퍼티노 애플 파크 웨이
1 애플 인크. 내

알시나, 토마스

미국 95014 캘리포니아주 쿠퍼티노 애플 파크 웨이
1 애플 인크. 내

와인버그, 개럿 엘.

미국 95014 캘리포니아주 쿠퍼티노 애플 파크 웨이
1 애플 인크. 내

(30) 우선권주장

PA201770425 2017년06월01일 덴마크(DK)

PA201770426 2017년06월01일 덴마크(DK)

명세서

청구범위

청구항 1

디지털 어시스턴트를 동작시키는 방법으로서,

하나 이상의 프로세서들 및 메모리를 구비한 전자 디바이스에서:

미디어에 대한 요청을 나타내는 제1 자연 언어 스피치 입력을 수신하는 단계 - 상기 제1 자연 언어 스피치 입력은 제1 검색 파라미터를 포함함 -;

상기 디지털 어시스턴트에 의해, 제1 미디어 아이템을 제공하는 단계 - 상기 제1 미디어 아이템은 상기 제1 검색 파라미터에 기초하여 식별됨 -;

상기 제1 미디어 아이템을 제공하는 동안, 제2 자연 언어 스피치 입력을 수신하는 단계;

상기 제2 자연 언어 스피치 입력이 상기 미디어에 대한 요청을 정제하려는 사용자 의도에 대응하는지 여부를 결정하는 단계;

상기 제2 자연 언어 스피치 입력이 상기 미디어에 대한 요청을 정제하려는 사용자 의도에 대응한다는 결정에 따라,

상기 제1 파라미터 및 상기 제2 자연 언어 스피치 입력에 기초하여, 상기 제1 미디어 아이템과 상이한 제2 미디어 아이템을 식별하는 단계; 및

상기 디지털 어시스턴트에 의해, 상기 제2 미디어 아이템을 제공하는 단계를 포함하는, 방법.

청구항 2

제1항에 있어서,

상기 제1 자연 언어 스피치 입력에 기초하여, 텍스트 문자열을 획득하는 단계;

상기 텍스트 문자열에 기초하여, 미디어 아이템들에 대한 추천을 획득하려는 사용자 의도의 표현을 결정하는 단계; 및

상기 사용자 의도의 표현에 기초하여, 태스크 및 상기 태스크를 수행하기 위한 하나 이상의 파라미터들을 결정하는 단계를 추가로 포함하며, 상기 하나 이상의 파라미터들은 상기 제1 검색 파라미터를 포함하는, 방법.

청구항 3

제1항 또는 제2항에 있어서, 상기 제1 미디어 아이템을 제공하는 단계는,

상기 디지털 어시스턴트에 의해, 상기 제1 미디어 아이템과 연관된 구두 응답을 나타내는 스피치 출력을 제공하는 단계; 및

상기 구두 응답을 나타내는 상기 스피치 출력을 제공하는 동안, 상기 디지털 어시스턴트에 의해, 상기 제1 미디어 아이템의 일부분의 재생을 제공하는 단계를 포함하는, 방법.

청구항 4

제1항 또는 제2항에 있어서, 상기 제1 미디어 아이템을 제공하는 단계는,

상기 디지털 어시스턴트에 의해, 상기 제1 미디어 아이템의 재생을 제공하는 단계를 포함하는, 방법.

청구항 5

제1항 또는 제2항에 있어서, 상기 제1 미디어 아이템을 제공하는 단계는,

상기 디지털 어시스턴트에 의해, 복수의 미디어 아이템들을 제공하는 단계를 포함하며, 상기 복수의 미디어 아

이템들은 상기 제1 미디어 아이템을 포함하는, 방법.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서,

상기 제2 자연 언어 스피치 입력을 수신하는 것에 응답하여, 상기 제1 미디어 아이템이 제공되는 방식을 조정하는 단계를 추가로 포함하는, 방법.

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 상기 제2 자연 언어 스피치 입력이 상기 미디어에 대한 요청을 정제하려는 사용자 의도에 대응하는지 여부를 결정하는 단계는, 하나 이상의 미리정의된 구절들 및 상기 하나 이상의 구절들의 자연 언어 등가물들에 기초하여 상기 미디어에 대한 요청을 정제하려는 사용자 의도의 표현을 도출하는 단계를 포함하는, 방법.

청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서,

상기 제2 자연 언어 스피치 입력에 기초하여, 상기 미디어에 대한 요청을 정제하기 위한 하나 이상의 파라미터들을 획득하는 단계를 추가로 포함하는, 방법.

청구항 9

제8항에 있어서, 상기 제2 자연 언어 스피치 입력이 상기 미디어에 대한 요청을 정제하려는 사용자 의도에 대응하는지 여부를 결정하는 단계는, 컨텍스트 정보에 기초하여 상기 미디어에 대한 요청을 정제하려는 사용자 의도의 표현을 도출하는 단계를 포함하는, 방법.

청구항 10

제8항 또는 제9항에 있어서, 상기 하나 이상의 파라미터들 중 일 파라미터는 미디어 아이템의 가사 내용 (lyrical content)에 대응하는, 방법.

청구항 11

제8항 내지 제10항 중 어느 한 항에 있어서, 상기 하나 이상의 파라미터들 중 일 파라미터는 행사 또는 기간에 대응하는, 방법.

청구항 12

제8항 내지 제11항 중 어느 한 항에 있어서, 상기 하나 이상의 파라미터들 중 일 파라미터는 활동에 대응하는, 방법.

청구항 13

제8항 내지 제12항 중 어느 한 항에 있어서, 상기 하나 이상의 파라미터들 중 일 파라미터는 위치에 대응하는, 방법.

청구항 14

제8항 내지 제13항 중 어느 한 항에 있어서, 상기 하나 이상의 파라미터들 중 일 파라미터는 기분에 대응하는, 방법.

청구항 15

제8항 내지 제14항 중 어느 한 항에 있어서, 상기 하나 이상의 파라미터들 중 일 파라미터는 미리결정된 시간 프레임 내의 출시 날짜에 대응하는, 방법.

청구항 16

제8항 내지 제15항 중 어느 한 항에 있어서, 상기 하나 이상의 파라미터들 중 일 파라미터는 의도된 청중에 대

응하는, 방법.

청구항 17

제8항 내지 제16항 중 어느 한 항에 있어서, 상기 하나 이상의 파라미터들 중 일 파라미터는 미디어 아이템들의 컬렉션에 대응하는, 방법.

청구항 18

제8항 내지 제17항 중 어느 한 항에 있어서,

상기 제2 자연 언어 스피치 입력은 제1 사용자와 연관되며;

상기 하나 이상의 파라미터들 중 일 파라미터는 상기 제1 사용자와 상이한 제2 사용자에게 대응하는, 방법.

청구항 19

제8항 내지 제18항 중 어느 한 항에 있어서, 상기 미디어에 대한 요청을 정제하기 위한 하나 이상의 파라미터들을 획득하는 단계는, 컨텍스트 정보에 기초하여 하나 이상의 파라미터들을 결정하는 단계를 포함하는, 방법.

청구항 20

제19항에 있어서, 상기 컨텍스트 정보는 상기 제1 미디어 아이템에 관련된 정보를 포함하는, 방법.

청구항 21

제19항에 있어서,

하나 이상의 사용자들의 물리적 존재를 검출하는 단계를 추가로 포함하며,

상기 컨텍스트 정보는 상기 하나 이상의 사용자들에 관련된 정보를 포함하는, 방법.

청구항 22

제19항에 있어서, 상기 컨텍스트 정보는 상기 전자 디바이스의 하나 이상의 사용자들과 연관된 설정을 포함하는, 방법.

청구항 23

제8항 내지 제22항 중 어느 한 항에 있어서,

상기 제1 자연 언어 스피치 입력에 기초하여, 미디어 아이템들의 제1 세트를 획득하는 단계;

상기 미디어 아이템들의 제1 세트로부터 상기 제1 미디어 아이템을 선택하는 단계;

상기 제2 자연 언어 스피치 입력에 기초하여, 미디어 아이템들의 제2 세트를 획득하는 단계 - 상기 미디어 아이템들의 제2 세트는 상기 미디어 아이템들의 제1 세트의 서브세트임 -; 및

상기 미디어 아이템들의 세트 제2 세트로부터 상기 제2 미디어 아이템을 선택하는 단계를 추가로 포함하는, 방법.

청구항 24

제23항에 있어서, 상기 미디어 아이템들의 제2 세트를 획득하는 단계는,

상기 미디어 아이템들의 제1 세트로부터, 상기 미디어에 대한 요청을 정제하기 위한 상기 하나 이상의 파라미터들에 기초하여 하나 이상의 미디어 아이템들을 선택하는 단계를 포함하는, 방법.

청구항 25

제8항 내지 제24항 중 어느 한 항에 있어서, 상기 제2 미디어 아이템을 식별하는 단계는,

상기 제2 미디어 아이템과 연관된 콘텐츠가 상기 하나 이상의 파라미터들 중 적어도 하나와 매칭되는지 여부를 결정하는 단계를 포함하는, 방법.

청구항 26

제8항 내지 제25항 중 어느 한 항에 있어서, 상기 제2 미디어 아이템을 식별하는 단계는,

상기 제2 미디어 아이템과 연관된 메타데이터가 상기 하나 이상의 파라미터들 중 적어도 하나와 매칭되는지 여부를 결정하는 단계를 포함하는, 방법.

청구항 27

제1항 내지 제26항 중 어느 한 항에 있어서,

미디어 아이템들의 사용자-특정 코퍼스로부터 상기 제2 미디어 아이템을 획득하는 단계를 추가로 포함하며, 상기 미디어 아이템들의 사용자-특정 코퍼스는 사용자와 연관된 데이터에 기초하여 생성되는, 방법.

청구항 28

제27항에 있어서,

상기 제2 자연 언어 스피치 입력과 연관된 음향 정보에 기초하여 상기 미디어 아이템들의 사용자-특정 코퍼스를 식별하는 단계를 추가로 포함하는, 방법.

청구항 29

제27항 또는 제28항에 있어서, 상기 미디어 아이템들의 사용자-특정 코퍼스 내의 미디어 아이템은, 활동; 기분; 행사; 위치; 시간; 큐레이터; 재생목록; 하나 이상의 이전 사용자 입력들; 또는 이들의 임의의 조합을 나타내는 메타데이터를 포함하는, 방법.

청구항 30

제29항에 있어서, 상기 메타데이터의 적어도 일부분은 상기 제1 사용자와 상이한 제2 사용자로부터의 정보에 기초하는, 방법.

청구항 31

제1항 내지 제30항 중 어느 한 항에 있어서,

제3 자연 언어 스피치 입력을 수신하는 단계;

상기 제3 자연 언어 스피치 입력에 기초하여, 상기 제2 미디어 아이템을 미디어 아이템들의 컬렉션과 연관시키려는 사용자 의도의 표현을 결정하는 단계;

상기 제2 미디어 아이템을 상기 미디어 아이템들의 컬렉션과 연관시키는 단계; 및

상기 디지털 어시스턴트에 의해, 상기 연관성을 나타내는 오디오 출력을 제공하는 단계를 추가로 포함하는, 방법.

청구항 32

제1항 내지 제31항 중 어느 한 항에 있어서,

상기 제2 미디어 아이템을 제공하는 동안, 제4 자연 언어 스피치 입력을 수신하는 단계;

상기 제4 자연 언어 스피치 입력에 기초하여, 특정 미디어 아이템에 관련된 정보를 획득하려는 사용자 의도의 표현을 결정하는 단계; 및

상기 디지털 어시스턴트에 의해, 상기 특정 미디어 아이템에 관련된 상기 정보를 제공하는 단계를 추가로 포함하는, 방법.

청구항 33

제32항에 있어서, 상기 특정 미디어 아이템을 컨텍스트 정보에 기초하여 선택하는 단계를 추가로 포함하는, 방법.

청구항 34

제1항 내지 제33항 중 어느 한 항에 있어서,

상기 제2 미디어 아이템을 제공하는 동안, 상기 디지털 어시스턴트에 의해, 제3 미디어 아이템을 나타내는 스피치 출력을 제공하는 단계; 및

상기 제2 미디어 아이템을 제공한 후, 상기 제3 미디어 아이템을 제공하는 단계를 추가로 포함하는, 방법.

청구항 35

제1항 내지 제34항 중 어느 한 항에 있어서, 상기 제2 미디어 아이템을 제공하는 단계는,

상기 디지털 어시스턴트에 의해, 상기 제2 미디어 아이템과 연관된 구두 응답을 나타내는 스피치 출력을 제공하는 단계; 및

상기 구두 응답을 나타내는 상기 스피치 출력을 제공하는 동안, 상기 디지털 어시스턴트에 의해, 상기 제2 미디어 아이템의 일부분의 재생을 제공하는 단계를 포함하는, 방법.

청구항 36

제1항 내지 제34항 중 어느 한 항에 있어서, 상기 제2 미디어 아이템을 제공하는 단계는,

상기 디지털 어시스턴트에 의해, 상기 제2 미디어 아이템의 재생을 제공하는 단계를 포함하는, 방법.

청구항 37

제1항 내지 제34항 중 어느 한 항에 있어서, 상기 제2 미디어 아이템을 제공하는 단계는,

상기 디지털 어시스턴트에 의해, 복수의 미디어 아이템들을 제공하는 단계를 포함하며, 상기 복수의 미디어 아이템들은 상기 제2 미디어 아이템을 포함하는, 방법.

청구항 38

제1항 내지 제37항 중 어느 한 항에 있어서, 상기 제1 미디어 아이템은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 임의의 조합인, 방법.

청구항 39

제1항 내지 제37항 중 어느 한 항에 있어서, 상기 제2 미디어 아이템은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 임의의 조합인, 방법.

청구항 40

제1항 내지 제39항 중 어느 한 항에 있어서, 상기 전자 디바이스는 컴퓨터, 셋톱 박스, 스피커, 스마트 워치, 전화, 또는 이들의 조합인, 방법.

청구항 41

디지털 어시스턴트를 동작시키는 방법으로서,

하나 이상의 프로세서들 및 메모리를 구비한 전자 디바이스에서:

자연 언어 스피치 입력을 수신하는 단계;

상기 디지털 어시스턴트에 의해, 상기 자연 언어 스피치 입력에 기초하여 태스크를 식별하는 단계;

상기 디지털 어시스턴트에 의해, 상기 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력을 제공하는 단계; 및

구두 응답을 나타내는 상기 스피치 출력을 제공하는 동안,

상기 디지털 어시스턴트에 의해, 상기 구두 응답에 대응하는 미디어 아이템의 재생을 제공하는 단계를 포함하는, 방법.

청구항 42

제41항에 있어서, 상기 스피치 입력에 기초하여 태스크를 식별하는 단계는,
 상기 자연 언어 스피치 입력에 기초하여 텍스트 문자열을 획득하는 단계;
 사용자 의도의 표현을 획득하기 위해 상기 텍스트 문자열을 해석하는 단계; 및
 상기 사용자 의도의 표현에 기초하여 상기 태스크를 결정하는 단계를 포함하는, 방법.

청구항 43

제41항 또는 제42항에 있어서,
 상기 자연 언어 스피치 입력은 하나 이상의 미디어 아이템들에 대한 요청을 나타내는 것이며;
 상기 자연 언어 스피치 입력에 기초하여 상기 태스크를 식별하는 단계는 하나 이상의 미디어 아이템들을 제공하는 태스크를 식별하는 단계를 포함하는, 방법.

청구항 44

제43항에 있어서, 상기 미디어 아이템은 제1 미디어 아이템이고, 상기 방법은,
 상기 스피치 입력에 기초하여, 제2 미디어 아이템을 식별하는 단계; 및
 상기 제2 미디어 아이템에 대응하는 정보를 획득하는 단계를 추가로 포함하는, 방법.

청구항 45

제44항에 있어서, 상기 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력을 제공하는 단계는,
 상기 제2 미디어 아이템에 대한 구두 설명을 제공하는 단계를 포함하는, 방법.

청구항 46

제44항 또는 제45항에 있어서, 상기 구두 응답을 나타내는 상기 스피치 출력은,
 상기 디지털 어시스턴트의 음성,
 상기 제2 미디어 아이템과 연관된 음성,
 또는 이들의 조합으로 제공되는, 방법.

청구항 47

제44항 내지 제46항 중 어느 한 항에 있어서, 상기 제1 미디어 아이템은 상기 제2 미디어 아이템의 일부분에 대응하는, 방법.

청구항 48

제44항 내지 제47항 중 어느 한 항에 있어서,
 상기 제1 미디어 아이템의 재생을 제공하는 동안,
 제2 자연 언어 스피치 입력을 수신하는 단계; 및
 상기 제2 자연 언어 스피치 입력을 수신하는 것에 응답하여, 상기 제2 미디어 아이템의 재생을 제공하는 단계를 추가로 포함하는, 방법.

청구항 49

제48항에 있어서, 상기 제2 미디어 아이템의 재생은 상기 제1 미디어 아이템의 상기 재생과 상이한 볼륨에서 제공되는, 방법.

청구항 50

제44항 내지 제49항 중 어느 한 항에 있어서, 상기 식별된 태스크와 연관된 구두 응답을 나타내는 상기 스피치 출력은 제1 스피치 출력이고, 상기 방법은,

상기 제2 미디어 아이템의 상기 재생을 제공하는 동안,

제2 스피치 출력을 제공하는 단계를 추가로 포함하는, 방법.

청구항 51

제44항 내지 제50항 중 어느 한 항에 있어서, 상기 제2 미디어 아이템은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 조합을 포함하는, 방법.

청구항 52

제41항에 있어서, 하나 이상의 결과들을 획득하기 위해 상기 태스크를 수행하는 단계를 추가로 포함하는, 방법.

청구항 53

제52항에 있어서,

상기 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력을 제공하는 단계는, 상기 하나 이상의 결과들 중 일 결과에 대한 구두 설명을 나타내는 스피치 출력을 제공하는 단계, 및

상기 구두 응답에 대응하는 미디어 아이템의 재생을 제공하는 단계는, 상기 결과에 대응하는 사운드 효과의 재생을 제공하는 단계를 포함하는, 방법.

청구항 54

제41항 내지 제53항 중 어느 한 항에 있어서, 상기 스피치 출력은 제1 볼륨에서 제공되고, 미디어 아이템의 재생은 상기 제1 볼륨과 상이한 제2 볼륨에서 제공되는, 방법.

청구항 55

제41항 내지 제54항 중 어느 한 항에 있어서, 상기 전자 디바이스는 컴퓨터, 셋톱 박스, 스피커, 스마트 워치, 전화, 또는 이들의 조합인, 방법.

청구항 56

디지털 어시스턴트를 동작시키는 방법으로서,

하나 이상의 프로세서들 및 메모리를 구비한 전자 디바이스에서:

미디어에 대한 요청을 나타내는 스피치 입력을 수신하는 단계;

상기 스피치 입력을 수신하는 것에 응답하여, 상기 디지털 어시스턴트에 의해, 제1 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하는 단계;

상기 디지털 어시스턴트에 의해, 상기 미디어에 대한 요청에 대응하는 연속적인 비-공정 응답들의 수가 임계치를 만족시키는지 여부를 결정하는 단계;

상기 연속적인 비-공정 응답들의 수가 임계치를 만족시키지 않는다는 결정에 따라,

상기 디지털 어시스턴트에 의해, 상기 제1 미디어 아이템과 상이한 제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하는 단계;

상기 연속적인 비-공정 응답들의 수가 임계치를 만족시킨다는 결정에 따라,

제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하는 것을 중지하는 단계; 및

상기 디지털 어시스턴트에 의해, 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공하는 단계를 포함하는, 방법.

청구항 57

제56항에 있어서,

상기 미디어에 대한 요청을 나타내는 상기 스피치 입력에 기초하여, 텍스트 문자열을 획득하는 단계;

상기 획득된 텍스트 문자열에 기초하여, 사용자 의도의 표현을 결정하는 단계; 및

상기 사용자 의도의 표현에 기초하여, 하나 이상의 미디어 아이템들에 관련된 정보를 획득하는 단계를 추가로 포함하며, 상기 하나 이상의 미디어 아이템들은 상기 제1 미디어 아이템 및 상기 제2 미디어 아이템을 포함하는, 방법.

청구항 58

제56항 또는 제57항에 있어서,

상기 미디어에 대한 요청을 나타내는 상기 스피치 입력은 제1 스피치 입력이고,

상기 미디어에 대한 요청에 대응하는 연속적인 비-긍정 응답들의 수가 임계치를 만족시키는지 여부를 결정하는 단계는,

상기 제1 미디어 아이템의 상기 제안을 나타내는 상기 오디오 출력을 제공한 후에, 제2 스피치 입력을 수신하는 단계;

상기 제2 스피치 입력이 상기 미디어에 대한 요청에 대응하는 비-긍정 응답을 나타내는지 여부를 결정하는 단계;

상기 제2 스피치 입력이 비-긍정 응답을 나타낸다는 결정에 따라, 상기 요청에 대응하는 연속적인 비-긍정 응답들의 수를 업데이트하는 단계; 및

상기 제2 스피치 입력이 비-긍정 응답을 나타내지 않는다는 결정에 따라, 상기 요청에 대응하는 연속적인 비-긍정 응답들의 수를 업데이트하는 것을 중지하는 단계를 포함하는, 방법.

청구항 59

제58항에 있어서, 상기 제2 스피치 입력이 상기 미디어에 대한 요청에 대한 비-긍정 응답을 나타내는지 여부를 결정하는 단계는, 상기 제2 스피치 입력이 거부를 나타내는지 여부를 결정하는 단계를 포함하는, 방법.

청구항 60

제56항 또는 제57항에 있어서, 상기 미디어에 대한 요청에 대응하는 연속적인 비-긍정 응답들의 수가 임계치를 만족시키는지 여부를 결정하는 단계는,

상기 제1 미디어 아이템의 제안을 나타내는 상기 오디오 출력을 제공한 후에, 상기 요청에 대응하는 응답이 미리 정의된 기간 내에 수신되지 않는 것으로 결정하는 단계; 및

상기 요청에 대응하는 상기 연속적인 비-긍정 응답들의 수를 업데이트하는 단계를 포함하는, 방법.

청구항 61

제56항 내지 제60항 중 어느 한 항에 있어서, 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공하는 단계는,

상기 디지털 어시스턴트에 의해, 상기 디지털 어시스턴트에 의해 이전에 제안된 복수의 미디어 아이템들 중에서 사용자 선택에 대한 촉구를 나타내는 스피치 출력을 제공하는 단계를 포함하는, 방법.

청구항 62

제61항에 있어서,

상기 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공한 후에, 사용자 선택을 나타내는 스피치 입력을 수신하는 단계; 및

컨텍스트 정보에 기초하여, 상기 사용자 선택을 나타내는 상기 스피치 입력을 해석하는 단계를 추가로 포함하는, 방법.

청구항 63

제62항에 있어서, 상기 컨텍스트 정보는 상기 디지털 어시스턴트에 의해 이전에 제안된 상기 복수의 미디어 아이템들을 포함하는, 방법.

청구항 64

제56항 내지 제60항 중 어느 한 항에 있어서, 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공하는 단계는,

상기 디지털 어시스턴트에 의해, 상기 미디어에 대한 요청에 대한 하나 이상의 파라미터들에 대한 추구를 나타내는 스피치 출력을 제공하는 단계를 포함하는, 방법.

청구항 65

제64항에 있어서,

상기 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공한 후에, 상기 미디어에 대한 요청에 대한 하나 이상의 파라미터들을 나타내는 스피치 입력을 수신하는 단계를 추가로 포함하는, 방법.

청구항 66

제65항에 있어서, 상기 하나 이상의 파라미터들에 기초하여 제3 미디어 아이템을 획득하는 단계를 추가로 포함하며, 상기 제3 미디어 아이템은 상기 제1 미디어 아이템 및 상기 제2 미디어 아이템과 상이한, 방법.

청구항 67

제56항 내지 제66항 중 어느 한 항에 있어서, 상기 제1 미디어 아이템은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 임의의 조합인, 방법.

청구항 68

제56항 내지 제67항 중 어느 한 항에 있어서, 상기 제2 미디어 아이템은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 임의의 조합인, 방법.

청구항 69

제56항 내지 제68항 중 어느 한 항에 있어서, 상기 전자 디바이스는 컴퓨터, 셋톱 박스, 스피커, 스마트 워치, 전화, 또는 이들의 조합인, 방법.

청구항 70

디지털 어시스턴트를 동작시키는 방법으로서,

하나 이상의 프로세서들 및 메모리를 구비한 전자 디바이스에서:

미디어에 대한 요청을 나타내는 스피치 입력을 수신하는 단계;

상기 디지털 어시스턴트에 의해, 상기 전자 디바이스에 대한 복수의 사용자들의 물리적 존재를 검출하는 단계;

상기 복수의 사용자들의 물리적 존재를 검출하는 것에 응답하여, 상기 복수의 사용자들에 대응하는 복수의 선호도 프로파일들을 획득하는 단계;

상기 디지털 어시스턴트에 의해, 상기 복수의 선호도 프로파일들에 기초하여 병합된 선호도 프로파일을 제공하는 단계;

상기 디지털 어시스턴트에 의해, 상기 병합된 선호도 프로파일에 기초하여 미디어 아이템을 식별하는 단계; 및

상기 디지털 어시스턴트에 의해, 상기 식별된 미디어 아이템을 포함하는 오디오 출력을 제공하는 단계를 포함하는, 방법.

청구항 71

제70항에 있어서,

상기 전자 디바이스는 상기 복수의 사용자들 중 제1 사용자와 연관된 제1 전자 디바이스이고,

상기 전자 디바이스에 대한 상기 복수의 사용자들의 물리적 존재를 검출하는 단계는,

상기 복수의 사용자들 중 제2 사용자와 연관된 제2 전자 디바이스에 대응하는 정보를 수신하는 단계를 포함하는, 방법.

청구항 72

제71항에 있어서, 상기 제2 전자 디바이스에 대응하는 정보를 수신하는 단계는, 상기 제1 전자 디바이스에서, 상기 제2 전자 디바이스로부터 식별 정보를 수신하는 단계를 포함하는, 방법.

청구항 73

제71항에 있어서, 상기 제2 전자 디바이스에 대응하는 정보를 수신하는 단계는 라우팅 디바이스로부터 식별 정보를 수신하는 단계를 포함하며,

상기 라우팅 디바이스는 상기 제1 전자 디바이스 및 상기 제2 전자 디바이스에 접속되는, 방법.

청구항 74

제71항 내지 제73항 중 어느 한 항에 있어서,

원격 디바이스로부터 상기 제2 사용자에게 대응하는 선호도 프로파일을 수신하는 단계를 추가로 포함하는, 방법.

청구항 75

제71항 내지 제73항 중 어느 한 항에 있어서, 상기 제2 사용자에게 대응하는 선호도 프로파일이 상기 전자 디바이스 상에 저장되는, 방법.

청구항 76

제70항 내지 제75항 중 어느 한 항에 있어서, 상기 병합된 선호도 프로파일을 제공하는 단계는,

상기 복수의 선호도 프로파일들 각각에 의해 공유되는 하나 이상의 선호도들을 식별하는 단계를 포함하는, 방법.

청구항 77

제76항에 있어서, 상기 식별된 미디어 아이템은 상기 하나 이상의 선호도들과 매칭되는 메타데이터와 연관되는, 방법.

청구항 78

제71항 내지 제77항 중 어느 한 항에 있어서, 상기 병합된 선호도 프로파일에 기초하여 미디어 아이템을 식별하는 단계는,

상기 복수의 미디어 아이템들로부터 상기 미디어 아이템을 식별하는 단계를 포함하며, 상기 복수의 미디어 아이템들은 상기 제1 사용자와 연관된 미디어 아이템들의 제1 세트 및 상기 제2 사용자와 연관된 미디어 아이템들의 제2 세트를 포함하는, 방법.

청구항 79

제78항에 있어서, 상기 미디어 아이템은 상기 미디어 아이템들의 제1 세트의 일부가 아니며, 상기 미디어 아이템은 상기 미디어 아이템들의 제2 세트의 일부인, 방법.

청구항 80

제78항 또는 제79항에 있어서,

상기 복수의 사용자들의 물리적 존재를 검출한 후에, 상기 디지털 어시스턴트에 의해, 상기 제2 사용자의 부재

를 검출하는 단계; 및

상기 제2 사용자의 부재를 검출한 후에,

상기 복수의 미디어 아이템들을 업데이트하는 단계; 및

상기 병합된 선호 프로파일을 업데이트하는 단계를 추가로 포함하는, 방법.

청구항 81

제78항 내지 제80항 중 어느 한 항에 있어서, 상기 복수의 미디어 아이템들을 업데이트하는 단계는 상기 복수의 미디어 아이템들로부터 상기 미디어 아이템을 제거하는 단계를 포함하는, 방법.

청구항 82

제70항 내지 제81항 중 어느 한 항에 있어서, 상기 오디오 출력은 상기 식별된 미디어 아이템의 구두 설명을 포함하는, 방법.

청구항 83

제70항 내지 제82항 중 어느 한 항에 있어서, 상기 오디오 출력은 상기 병합된 프로파일을 나타내는 스피치 출력을 포함하는, 방법.

청구항 84

제70항 내지 제83항 중 어느 한 항에 있어서, 상기 식별된 미디어 아이템은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 임의의 조합인, 방법.

청구항 85

제70항 내지 제84항 중 어느 한 항에 있어서, 상기 전자 디바이스는 컴퓨터, 셋톱 박스, 스피커, 스마트 워치, 전화, 또는 이들의 조합인, 방법.

청구항 86

전자 디바이스로서,

하나 이상의 프로세서들;

메모리; 및

상기 메모리에 저장되고 상기 하나 이상의 프로세서들에 의해 실행되도록 구성된 하나 이상의 프로그램들을 포함하며, 상기 하나 이상의 프로그램들은,

미디어에 대한 요청을 나타내는 제1 자연 언어 스피치 입력을 수신하고 - 상기 제1 자연 언어 스피치 입력은 제1 검색 파라미터를 포함함 -;

디지털 어시스턴트에 의해, 제1 미디어 아이템을 제공하고 - 상기 제1 미디어 아이템은 상기 제1 검색 파라미터에 기초하여 식별됨 -;

상기 제1 미디어 아이템을 제공하는 동안, 제2 자연 언어 스피치 입력을 수신하고;

상기 제2 자연 언어 스피치 입력이 상기 미디어에 대한 요청을 정제하려는 사용자 의도에 대응하는지 여부를 결정하고;

상기 제2 자연 언어 스피치 입력이 상기 미디어에 대한 요청을 정제하려는 사용자 의도에 대응한다는 결정에 따라,

상기 제1 파라미터 및 상기 제2 자연 언어 스피치 입력에 기초하여, 상기 제1 미디어 아이템과 상이한 제2 미디어 아이템을 식별하고;

상기 디지털 어시스턴트에 의해, 상기 제2 미디어 아이템을 제공하기 위한 명령어들을 포함하는, 전자 디바이스.

청구항 87

전자 디바이스로서,

하나 이상의 프로세서들;

메모리; 및

상기 메모리에 저장되고 상기 하나 이상의 프로세서들에 의해 실행되도록 구성된 하나 이상의 프로그램들을 포함하며, 상기 하나 이상의 프로그램들은,

자연 언어 스피치 입력을 수신하고;

디지털 어시스턴트에 의해, 상기 자연 언어 스피치 입력에 기초하여 태스크를 식별하고;

상기 디지털 어시스턴트에 의해, 상기 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력을 제공하고;

구두 응답을 나타내는 상기 스피치 출력을 제공하는 동안,

상기 디지털 어시스턴트에 의해, 상기 구두 응답에 대응하는 미디어 아이템의 재생을 제공하기 위한 명령어들을 포함하는, 전자 디바이스.

청구항 88

전자 디바이스로서,

하나 이상의 프로세서들;

메모리; 및

상기 메모리에 저장되고 상기 하나 이상의 프로세서들에 의해 실행되도록 구성된 하나 이상의 프로그램들을 포함하며, 상기 하나 이상의 프로그램들은,

미디어에 대한 요청을 나타내는 스피치 입력을 수신하고;

상기 스피치 입력을 수신하는 것에 응답하여, 디지털 어시스턴트에 의해, 제1 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하고;

상기 디지털 어시스턴트에 의해, 상기 미디어에 대한 요청에 대응하는 연속적인 비-긍정 응답들의 수가 임계치를 만족시키는지 여부를 결정하고;

상기 연속적인 비-긍정 응답들의 수가 임계치를 만족시키지 않는다는 결정에 따라,

상기 디지털 어시스턴트에 의해, 상기 제1 미디어 아이템과 상이한 제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하고;

상기 연속적인 비-긍정 응답들의 수가 임계치를 만족시킨다는 결정에 따라,

제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하는 것을 중지하고;

상기 디지털 어시스턴트에 의해, 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공하기 위한 명령어들을 포함하는, 전자 디바이스.

청구항 89

전자 디바이스로서,

하나 이상의 프로세서들;

메모리; 및

상기 메모리에 저장되고 상기 하나 이상의 프로세서들에 의해 실행되도록 구성된 하나 이상의 프로그램들을 포함하며, 상기 하나 이상의 프로그램들은,

미디어에 대한 요청을 나타내는 스피치 입력을 수신하고;

디지털 어시스턴트에 의해, 상기 전자 디바이스에 대한 복수의 사용자들의 물리적 존재를 검출하고;

상기 복수의 사용자들의 물리적 존재를 검출하는 것에 응답하여, 상기 복수의 사용자들에 대응하는 복수의 선호도 프로파일들을 획득하고;

상기 디지털 어시스턴트에 의해, 상기 복수의 선호도 프로파일들에 기초하여 병합된 선호도 프로파일을 제공하고;

상기 디지털 어시스턴트에 의해, 상기 병합된 선호도 프로파일에 기초하여 미디어 아이템을 식별하고;

상기 디지털 어시스턴트에 의해, 상기 식별된 미디어 아이템을 포함하는 오디오 출력을 제공하기 위한 명령어들을 포함하는, 전자 디바이스.

청구항 90

하나 이상의 프로그램들을 저장하는 비일시적 컴퓨터 판독가능 저장 매체로서, 상기 하나 이상의 프로그램들은, 전자 디바이스의 하나 이상의 프로세서들에 의해 실행될 때, 상기 전자 디바이스로 하여금,

미디어에 대한 요청을 나타내는 제1 자연 언어 스피치 입력을 수신하고 - 상기 제1 자연 언어 스피치 입력은 제1 검색 파라미터를 포함함 -;

디지털 어시스턴트에 의해, 제1 미디어 아이템을 제공하고 - 상기 제1 미디어 아이템은 상기 제1 검색 파라미터에 기초하여 식별됨 -;

상기 제1 미디어 아이템을 제공하는 동안, 제2 자연 언어 스피치 입력을 수신하고;

상기 제2 자연 언어 스피치 입력이 상기 미디어에 대한 요청을 정제하려는 사용자 의도에 대응하는지 여부를 결정하고;

상기 제2 자연 언어 스피치 입력이 상기 미디어에 대한 요청을 정제하려는 사용자 의도에 대응한다는 결정에 따라,

상기 제1 파라미터 및 상기 제2 자연 언어 스피치 입력에 기초하여, 상기 제1 미디어 아이템과 상이한 제2 미디어 아이템을 식별하고;

상기 디지털 어시스턴트에 의해, 상기 제2 미디어 아이템을 제공하게 하는 명령어들을 포함하는, 비일시적 컴퓨터 판독가능 저장 매체.

청구항 91

하나 이상의 프로그램들을 저장하는 비일시적 컴퓨터 판독가능 저장 매체로서, 상기 하나 이상의 프로그램들은, 전자 디바이스의 하나 이상의 프로세서들에 의해 실행될 때, 상기 전자 디바이스로 하여금,

자연 언어 스피치 입력을 수신하고;

디지털 어시스턴트에 의해, 상기 자연 언어 스피치 입력에 기초하여 태스크를 식별하고;

상기 디지털 어시스턴트에 의해, 상기 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력을 제공하고;

구두 응답을 나타내는 상기 스피치 출력을 제공하는 동안,

상기 디지털 어시스턴트에 의해, 상기 구두 응답에 대응하는 미디어 아이템의 재생을 제공하게 하는 명령어들을 포함하는, 비일시적 컴퓨터 판독가능 저장 매체.

청구항 92

하나 이상의 프로그램들을 저장하는 비일시적 컴퓨터 판독가능 저장 매체로서, 상기 하나 이상의 프로그램들은, 전자 디바이스의 하나 이상의 프로세서들에 의해 실행될 때, 상기 전자 디바이스로 하여금,

미디어에 대한 요청을 나타내는 스피치 입력을 수신하고;

상기 스피치 입력을 수신하는 것에 응답하여, 디지털 어시스턴트에 의해, 제1 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하고;

상기 디지털 어시스턴트에 의해, 상기 미디어에 대한 요청에 대응하는 연속적인 비-긍정 응답들의 수가 임계치를 만족시키는지 여부를 결정하고;

상기 연속적인 비-긍정 응답들의 수가 임계치를 만족시키지 않는다는 결정에 따라,

상기 디지털 어시스턴트에 의해, 상기 제1 미디어 아이템과 상이한 제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하고;

상기 연속적인 비-긍정 응답들의 수가 임계치를 만족시킨다는 결정에 따라,

제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하는 것을 중지하고;

상기 디지털 어시스턴트에 의해, 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공하게 하는 명령어들을 포함하는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 93

하나 이상의 프로그램들을 저장하는 비밀시적 컴퓨터 판독가능 저장 매체로서, 상기 하나 이상의 프로그램들은, 전자 디바이스의 하나 이상의 프로세서들에 의해 실행될 때, 상기 전자 디바이스로 하여금,

미디어에 대한 요청을 나타내는 스피치 입력을 수신하고;

디지털 어시스턴트에 의해, 상기 전자 디바이스에 대한 복수의 사용자들의 물리적 존재를 검출하고;

상기 복수의 사용자들의 물리적 존재를 검출하는 것에 응답하여, 상기 복수의 사용자들에 대응하는 복수의 선호도 프로파일들을 획득하고;

상기 디지털 어시스턴트에 의해, 상기 복수의 선호도 프로파일들에 기초하여 병합된 선호도 프로파일을 제공하고;

상기 디지털 어시스턴트에 의해, 상기 병합된 선호도 프로파일에 기초하여 미디어 아이템을 식별하고;

상기 디지털 어시스턴트에 의해, 상기 식별된 미디어 아이템을 포함하는 오디오 출력을 제공하게 하는 명령어들을 포함하는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 94

전자 디바이스로서,

미디어에 대한 요청을 나타내는 제1 자연 언어 스피치 입력을 수신하기 위한 수단 - 상기 제1 자연 언어 스피치 입력은 제1 검색 파라미터를 포함함 -;

디지털 어시스턴트에 의해, 제1 미디어 아이템을 제공하기 위한 수단 - 상기 제1 미디어 아이템은 상기 제1 검색 파라미터에 기초하여 식별됨 -;

상기 제1 미디어 아이템을 제공하는 동안, 제2 자연 언어 스피치 입력을 수신하기 위한 수단;

상기 제2 자연 언어 스피치 입력이 상기 미디어에 대한 요청을 정제하려는 사용자 의도에 대응하는지 여부를 결정하기 위한 수단;

상기 제2 자연 언어 스피치 입력이 상기 미디어에 대한 요청을 정제하려는 사용자 의도에 대응한다는 결정에 따라,

상기 제1 파라미터 및 상기 제2 자연 언어 스피치 입력에 기초하여, 상기 제1 미디어 아이템과 상이한 제2 미디어 아이템을 식별하기 위한 수단; 및

상기 디지털 어시스턴트에 의해, 상기 제2 미디어 아이템을 제공하기 위한 수단을 포함하는, 전자 디바이스.

청구항 95

전자 디바이스로서,

자연 언어 스피치 입력을 수신하기 위한 수단;

디지털 어시스턴트에 의해, 상기 자연 언어 스피치 입력에 기초하여 태스크를 식별하기 위한 수단;

상기 디지털 어시스턴트에 의해, 상기 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력을 제공하기 위한 수단; 및

구두 응답을 나타내는 상기 스피치 출력을 제공하는 동안,

상기 디지털 어시스턴트에 의해, 상기 구두 응답에 대응하는 미디어 아이템의 재생을 제공하기 위한 수단을 포함하는, 전자 디바이스.

청구항 96

전자 디바이스로서,

미디어에 대한 요청을 나타내는 스피치 입력을 수신하기 위한 수단;

상기 스피치 입력을 수신하는 것에 응답하여, 디지털 어시스턴트에 의해, 제1 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하기 위한 수단;

상기 디지털 어시스턴트에 의해, 상기 미디어에 대한 요청에 대응하는 연속적인 비-공정 응답들의 수가 임계치를 만족시키는지 여부를 결정하기 위한 수단;

상기 연속적인 비-공정 응답들의 수가 임계치를 만족시키지 않는다는 결정에 따라,

상기 디지털 어시스턴트에 의해, 상기 제1 미디어 아이템과 상이한 제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하기 위한 수단;

상기 연속적인 비-공정 응답들의 수가 임계치를 만족시킨다는 결정에 따라,

제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하는 것을 중지하고;

상기 디지털 어시스턴트에 의해, 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공하기 위한 수단을 포함하는, 전자 디바이스.

청구항 97

전자 디바이스로서,

미디어에 대한 요청을 나타내는 스피치 입력을 수신하기 위한 수단;

디지털 어시스턴트에 의해, 상기 전자 디바이스에 대한 복수의 사용자들의 물리적 존재를 검출하기 위한 수단;

상기 복수의 사용자들의 물리적 존재를 검출하는 것에 응답하여, 상기 복수의 사용자들에 대응하는 복수의 선호도 프로파일들을 획득하기 위한 수단;

상기 디지털 어시스턴트에 의해, 상기 복수의 선호도 프로파일들에 기초하여 병합된 선호도 프로파일을 제공하기 위한 수단;

상기 디지털 어시스턴트에 의해, 상기 병합된 선호도 프로파일에 기초하여 미디어 아이템을 식별하기 위한 수단; 및

상기 디지털 어시스턴트에 의해, 상기 식별된 미디어 아이템을 포함하는 오디오 출력을 제공하기 위한 수단을 포함하는, 전자 디바이스.

청구항 98

전자 디바이스로서,

하나 이상의 프로세서들;

메모리; 및

상기 메모리에 저장되고 상기 하나 이상의 프로세서들에 의해 실행되도록 구성된 하나 이상의 프로그램들을 포함하며, 상기 하나 이상의 프로그램들은 제1항 내지 제85항 중 어느 한 항의 방법을 수행하기 위한 명령어들을 포함하는, 전자 디바이스.

청구항 99

하나 이상의 프로그램들을 저장하는 비밀시적 컴퓨터 판독가능 저장 매체로서, 상기 하나 이상의 프로그램들은, 전자 디바이스의 하나 이상의 프로세서들에 의해 실행될 때, 상기 전자 디바이스로 하여금 제1항 내지 제85항 중 어느 한 항의 방법을 수행하게 하는 명령어들을 포함하는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 100

전자 디바이스로서,

제1항 내지 제85항 중 어느 한 항의 방법을 수행하기 위한 수단을 포함하는, 전자 디바이스.

발명의 설명

기술 분야

[0001] 본 출원은 2017년 5월 16일자로 출원된, 발명의 명칭이 "INTELLIGENT AUTOMATED ASSISTANT FOR MEDIA EXPLORATION"인 미국 가출원 제62/506,981호; 2017년 6월 1일자로 출원된, 발명의 명칭이 "INTELLIGENT AUTOMATED ASSISTANT FOR MEDIA EXPLORATION"인 덴마크 특허 출원 제PA201770425호; 및 2017년 6월 1일자로 출원된, 발명의 명칭이 "INTELLIGENT AUTOMATED ASSISTANT FOR MEDIA EXPLORATION"인 덴마크 특허 출원 제 PA201770426호에 대한 우선권을 주장하며, 그 내용은 그 전문이 본 명세서에 참조로서 편입된다.

[0002] 기술분야

[0003] 이는 일반적으로 지능형 자동화 어시스턴트에 관한 것으로, 보다 구체적으로는 미디어 탐색을 위한 디지털 어시스턴트의 청각 기반 인터페이스를 제공하는 것에 관한 것이다.

배경 기술

[0004] 지능형 자동화 어시스턴트들(또는 디지털 어시스턴트(digital assistant)들)은 인간 사용자들과 전자 디바이스들 사이의 유익한 인터페이스를 제공할 수 있다. 이러한 어시스턴트들은 사용자들이 음성(spoken) 및/또는 텍스트 형태들의 자연 언어를 사용하여 디바이스들 또는 시스템들과 상호작용하도록 할 수 있다. 예를 들어, 사용자는 전자 디바이스 상에서 동작하는 디지털 어시스턴트에게 사용자 요청을 포함하는 스피치 입력을 제공할 수 있다. 디지털 어시스턴트는 스피치 입력으로부터 사용자의 의도를 해석하고 사용자의 의도를 태스크로 운용할 수 있다. 이어서, 태스크들은 전자 디바이스의 하나 이상의 서비스들을 실행함으로써 수행될 수 있고, 사용자 요청에 응답하는 관련 출력은 사용자에게 반환될 수 있다.

[0005] 대부분의 경우, 사용자는 전자 디바이스와 상호작용하기 위해 종래의 그래픽 사용자 인터페이스에 적어도 부분적으로 의존한다. 그러나, 일부 경우에는, 디지털 어시스턴트가 디스플레이 능력이 제한되거나 전혀 없는 전자 디바이스 상에 구현될 수 있다.

발명의 내용

[0006] 예시적인 방법들이 본 명세서에 개시된다. 예시적인 방법은, 하나 이상의 프로세서들 및 메모리를 구비한 전자 디바이스에서, 미디어에 대한 요청을 나타내는 제1 자연 언어 스피치 입력을 수신하는 것 - 제1 자연 언어 스피치 입력은 제1 검색 파라미터를 포함함 -; 디지털 어시스턴트에 의해, 제1 미디어 아이템을 제공하는 것 - 제1 미디어 아이템은 제1 검색 파라미터에 기초하여 식별됨 -; 제1 미디어 아이템을 제공하는 동안, 제2 자연 언어 스피치 입력을 수신하는 것; 및 제2 자연 언어 스피치 입력이 미디어에 대한 요청을 정제하려는 사용자 의도에 대응하는지 여부를 결정하는 것을 포함한다. 방법은, 제2 자연 언어 스피치 입력이 미디어에 대한 요청을 정제하려는 사용자 의도에 대응한다는 결정에 따라, 제1 파라미터 및 제2 자연 언어 스피치 입력에 기초하여, 제1 미디어 아이템과 상이한 제2 미디어 아이템을 식별하는 것; 및 디지털 어시스턴트에 의해, 제2 미디어 아이템을 제공하는 것을 추가로 포함한다.

[0007] 예시적인 방법은, 하나 이상의 프로세서들 및 메모리를 구비한 전자 디바이스에서, 자연 언어 스피치 입력을 수신하는 것; 디지털 어시스턴트에 의해, 자연 언어 스피치 입력에 기초하여 태스크를 식별하는 것; 디지털 어시스턴트에 의해, 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력을 제공하는 것; 및 구두 응답을 나타내는 스피치 출력을 제공하는 동안, 디지털 어시스턴트에 의해, 구두 응답에 대응하는 미디어 아이템의 재생

을 제공하는 것을 포함한다.

- [0008] 예시적인 방법은, 하나 이상의 프로세서들 및 메모리를 구비한 전자 디바이스에서, 미디어에 대한 요청을 나타내는 스피치 입력을 수신하는 것; 스피치 입력을 수신하는 것에 응답하여, 디지털 어시스턴트에 의해, 제1 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하는 것; 디지털 어시스턴트에 의해, 미디어에 대한 요청에 대응하는 연속적인 비-긍정 응답들의 수가 임계치를 만족시키는지 여부를 결정하는 것을 포함한다. 방법은, 연속적인 비-긍정 응답들의 수가 임계치를 만족시키지 못한다는 결정에 따라, 디지털 어시스턴트에 의해, 제1 미디어 아이템과 상이한 제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하는 것을 추가로 포함한다. 방법은, 연속적인 비-긍정 응답들의 수가 임계치를 만족시킨다는 결정에 따라, 제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하는 것을 중지하는 것; 및 디지털 어시스턴트에 의해, 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공하는 것을 추가로 포함한다.
- [0009] 예시적인 방법은, 하나 이상의 프로세서들 및 메모리를 구비한 전자 디바이스에서, 미디어에 대한 요청을 나타내는 스피치 입력을 수신하는 것; 디지털 어시스턴트에 의해, 전자 디바이스에 대한 복수의 사용자들의 물리적 존재를 검출하는 것; 복수의 사용자들의 물리적 존재를 검출하는 것에 응답하여, 복수의 사용자들에 대응하는 복수의 선호도 프로파일들을 획득하는 것; 디지털 어시스턴트에 의해, 복수의 선호도 프로파일들에 기초하여 병합된 선호도 프로파일을 제공하는 것; 디지털 어시스턴트에 의해, 병합된 선호도 프로파일에 기초하여 미디어 아이템을 식별하는 것; 및 디지털 어시스턴트에 의해, 식별된 미디어 아이템을 포함하는 오디오 출력을 제공하는 것을 포함한다.
- [0010] 예시적인 비밀시적 컴퓨터 판독가능 매체가 본 명세서에 개시된다. 예시적인 비밀시적 컴퓨터 판독가능 저장 매체는 하나 이상의 프로그램들을 저장한다. 하나 이상의 프로그램들은, 전자 디바이스의 하나 이상의 프로세서들에 의해 실행될 때, 전자 디바이스로 하여금, 미디어에 대한 요청을 나타내는 제1 자연 언어 스피치 입력을 수신하고 - 제1 자연 언어 스피치 입력은 제1 검색 파라미터를 포함함 -; 디지털 어시스턴트에 의해, 제1 미디어 아이템을 제공하고 - 제1 미디어 아이템은 제1 검색 파라미터에 기초하여 식별됨 -; 제1 미디어 아이템을 제공하는 동안, 제2 자연 언어 스피치 입력을 수신하고; 제2 자연 언어 스피치 입력이 미디어에 대한 요청을 정제하려는 사용자 의도에 대응하는지 여부를 결정하게 하는 명령어들을 포함한다. 명령어들은, 추가로 전자 디바이스로 하여금, 제2 자연 언어 스피치 입력이 미디어에 대한 요청을 정제하려는 사용자 의도에 대응한다는 결정에 따라, 제1 파라미터 및 제2 자연 언어 스피치 입력에 기초하여, 제1 미디어 아이템과 상이한 제2 미디어 아이템을 식별하고; 디지털 어시스턴트에 의해, 제2 미디어 아이템을 제공하게 할 수 있다.
- [0011] 예시적인 비밀시적 컴퓨터 판독가능 저장 매체는 하나 이상의 프로그램들을 저장한다. 하나 이상의 프로그램들은, 전자 디바이스의 하나 이상의 프로세서들에 의해 실행될 때, 전자 디바이스로 하여금, 자연 언어 스피치 입력을 수신하고, 디지털 어시스턴트에 의해, 자연 언어 스피치 입력에 기초하여 태스크를 식별하고; 디지털 어시스턴트에 의해, 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력을 제공하고; 구두 응답을 나타내는 스피치 출력을 제공하는 동안, 디지털 어시스턴트에 의해, 구두 응답에 대응하는 미디어 아이템의 재생을 제공하게 하는 명령어들을 포함한다.
- [0012] 예시적인 비밀시적 컴퓨터 판독가능 저장 매체는 하나 이상의 프로그램들을 저장한다. 하나 이상의 프로그램들은, 전자 디바이스의 하나 이상의 프로세서들에 의해 실행될 때, 전자 디바이스로 하여금, 미디어에 대한 요청을 나타내는 스피치 입력을 수신하고, 스피치 입력을 수신하는 것에 응답하여, 디지털 어시스턴트에 의해, 제1 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하고; 디지털 어시스턴트에 의해, 미디어에 대한 요청에 대응하는 연속적인 비-긍정 응답들의 수가 임계치를 만족시키는지 여부를 결정하게 하는 명령어들을 포함한다. 명령어들은, 추가로 전자 디바이스로 하여금, 연속적인 비-긍정 응답들의 수가 임계치를 만족시키지 못한다는 결정에 따라, 디지털 어시스턴트에 의해, 제1 미디어 아이템과 상이한 제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하게 할 수 있다. 명령어들은, 추가로 전자 디바이스로 하여금, 연속적인 비-긍정 응답들의 수가 임계치를 만족시킨다는 결정에 따라, 제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하는 것을 중지하고; 디지털 어시스턴트에 의해, 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공하게 할 수 있다.
- [0013] 예시적인 비밀시적 컴퓨터 판독가능 저장 매체는 하나 이상의 프로그램들을 저장한다. 하나 이상의 프로그램들은, 전자 디바이스의 하나 이상의 프로세서들에 의해 실행될 때, 전자 디바이스로 하여금, 미디어에 대한 요청을 나타내는 스피치 입력을 수신하고, 디지털 어시스턴트에 의해, 전자 디바이스에 대한 복수의 사용자들의 물리적 존재를 검출하고; 복수의 사용자들의 물리적 존재를 검출하는 것에 응답하여, 복수의 사용자들에 대응하는

복수의 선호도 프로파일들을 획득하고; 디지털 어시스턴트에 의해, 복수의 선호도 프로파일들에 기초하여 병합된 선호도 프로파일을 제공하고; 디지털 어시스턴트에 의해, 병합된 선호도 프로파일에 기초하여 미디어 아이템을 식별하고; 디지털 어시스턴트에 의해, 식별된 미디어 아이템을 포함하는 오디오 출력을 제공하게 하는 명령어들을 포함한다.

[0014] 예시적인 전자 디바이스들이 본 명세서에 개시된다. 예시적인 전자 디바이스는 하나 이상의 프로세서들; 메모리; 및

[0015] 메모리에 저장되고 하나 이상의 프로세서들에 의해 실행되도록 구성된 하나 이상의 프로그램들을 포함하며, 하나 이상의 프로그램들은, 미디어에 대한 요청을 나타내는 제1 자연 언어 스피치 입력을 수신하고 - 제1 자연 언어 스피치 입력은 제1 검색 파라미터를 포함함 -; 디지털 어시스턴트에 의해, 제1 미디어 아이템을 제공하고 - 제1 미디어 아이템은 제1 검색 파라미터에 기초하여 식별됨 -; 제1 미디어 아이템을 제공하는 동안, 제2 자연 언어 스피치 입력을 수신하고; 제2 자연 언어 스피치 입력이 미디어에 대한 요청을 정제하려는 사용자 의도에 대응하는지 여부를 결정하기 위한 명령어들을 포함한다. 하나 이상의 프로그램들은, 제2 자연 언어 스피치 입력이 미디어에 대한 요청을 정제하려는 사용자 의도에 대응한다는 결정에 따라, 제1 파라미터 및 제2 자연 언어 스피치 입력에 기초하여, 제1 미디어 아이템과 상이한 제2 미디어 아이템을 식별하고; 디지털 어시스턴트에 의해, 제2 미디어 아이템을 제공하기 위한 명령어들을 추가로 포함한다.

[0016] 예시적인 전자 디바이스는 하나 이상의 프로세서들; 메모리; 및

[0017] 메모리에 저장되고 하나 이상의 프로세서들에 의해 실행되도록 구성된 하나 이상의 프로그램들을 포함하며, 하나 이상의 프로그램들은, 자연 언어 스피치 언어 입력을 수신하고, 디지털 어시스턴트에 의해, 자연 언어 스피치 입력에 기초하여 태스크를 식별하고; 디지털 어시스턴트에 의해, 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력을 제공하고; 구두 응답을 나타내는 스피치 출력을 제공하는 동안, 디지털 어시스턴트에 의해, 구두 응답에 대응하는 미디어 아이템의 재생을 제공하기 위한 명령어들을 포함한다.

[0018] 예시적인 전자 디바이스는 하나 이상의 프로세서들; 메모리; 및

[0019] 메모리에 저장되고 하나 이상의 프로세서들에 의해 실행되도록 구성된 하나 이상의 프로그램들을 포함하며, 하나 이상의 프로그램들은, 미디어에 대한 요청을 나타내는 스피치 입력을 수신하고, 스피치 입력을 수신하는 것에 응답하여, 디지털 어시스턴트에 의해, 제1 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하고; 디지털 어시스턴트에 의해, 미디어에 대한 요청에 대응하는 연속적인 비-공정 응답들의 수가 임계치를 만족시키는지 여부를 결정하기 위한 명령어들을 포함한다. 하나 이상의 프로그램들은 연속적인 비-공정 응답들의 수가 임계치를 만족시키지 못한다는 결정에 따라, 디지털 어시스턴트에 의해, 제1 미디어 아이템과 상이한 제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하기 위한 명령어들을 추가로 포함한다. 하나 이상의 프로그램들은 연속적인 비-공정 응답들의 수가 임계치를 만족시킨다는 결정에 따라, 제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하는 것을 중지하고; 디지털 어시스턴트에 의해, 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공하기 위한 명령어들을 추가로 포함한다.

[0020] 예시적인 전자 디바이스는 하나 이상의 프로세서들; 메모리; 및

[0021] 메모리에 저장되고 하나 이상의 프로세서들에 의해 실행되도록 구성된 하나 이상의 프로그램들을 포함하며, 하나 이상의 프로그램들은, 미디어에 대한 요청을 나타내는 스피치 입력을 수신하고, 디지털 어시스턴트에 의해, 전자 디바이스에 대한 복수의 사용자들의 물리적 존재를 검출하고; 복수의 사용자들의 물리적 존재를 검출하는 것에 응답하여, 복수의 사용자들에 대응하는 복수의 선호도 프로파일들을 획득하고; 디지털 어시스턴트에 의해, 복수의 선호도 프로파일들에 기초하여 병합된 선호도 프로파일을 제공하고; 디지털 어시스턴트에 의해, 병합된 선호도 프로파일에 기초하여 미디어 아이템을 식별하고; 디지털 어시스턴트에 의해, 식별된 미디어 아이템을 포함하는 오디오 출력을 제공하기 위한 명령어들을 포함한다

[0022] 예시적인 전자 디바이스는 미디어에 대한 요청을 나타내는 제1 자연 언어 스피치 입력을 수신하기 위한 수단 - 제1 자연 언어 스피치 입력은 제1 검색 파라미터를 포함함 -; 디지털 어시스턴트에 의해, 제1 미디어 아이템을 제공하기 위한 수단 - 제1 미디어 아이템은 제1 검색 파라미터에 기초하여 식별됨 -; 제1 미디어 아이템을 제공하는 동안, 제2 자연 언어 스피치 입력을 수신하기 위한 수단; 제2 자연 언어 스피치 입력이 미디어에 대한 요청을 정제하려는 사용자 의도에 대응하는지 여부를 결정하기 위한 수단; 제2 자연 언어 스피치 입력이 미디어에 대한 요청을 정제하려는 사용자 의도에 대응한다는 결정에 따라, 제1 파라미터 및 제2 자연 언어 스피치 입력에 기초하여, 제1 미디어 아이템과 상이한 제2 미디어 아이템을 식별하고; 디지털 어시스턴트에 의해, 제2 미디어

아이템을 제공하기 위한 수단을 포함한다.

- [0023] 예시적인 전자 디바이스는 자연 언어 스피치 입력을 수신하기 위한 수단; 디지털 어시스턴트에 의해, 자연 언어 스피치 입력에 기초하여 태스크를 식별하기 위한 수단; 디지털 어시스턴트에 의해, 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력을 제공하기 위한 수단; 및 구두 응답을 나타내는 스피치 출력을 제공하는 동안, 디지털 어시스턴트에 의해, 구두 응답에 대응하는 미디어 아이템의 재생을 제공하기 위한 수단을 포함한다.
- [0024] 예시적인 전자 디바이스는 미디어에 대한 요청을 나타내는 스피치 입력을 수신하기 위한 수단; 스피치 입력을 수신하는 것에 응답하여, 디지털 어시스턴트에 의해, 제1 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하기 위한 수단; 디지털 어시스턴트에 의해, 미디어에 대한 요청에 대응하는 연속적인 비-긍정 응답들의 수가 임계치를 만족시키는지 여부를 결정하기 위한 수단; 연속적인 비-긍정 응답들의 수가 임계치를 만족시키지 못한다는 결정에 따라, 디지털 어시스턴트에 의해, 제1 미디어 아이템과 상이한 제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하기 위한 수단; 연속적인 비-긍정 응답들의 수가 임계치를 만족시킨다는 결정에 따라, 제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하는 것을 중지하고; 디지털 어시스턴트에 의해, 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공하기 위한 수단을 포함한다.
- [0025] 예시적인 전자 디바이스는 미디어에 대한 요청을 나타내는 스피치 입력을 수신하기 위한 수단; 디지털 어시스턴트에 의해, 전자 디바이스에 대한 복수의 사용자들의 물리적 존재를 검출하기 위한 수단; 복수의 사용자들의 물리적 존재를 검출하는 것에 응답하여, 복수의 사용자들에 대응하는 복수의 선호도 프로파일들을 획득하기 위한 수단; 디지털 어시스턴트에 의해, 복수의 선호도 프로파일들에 기초하여 병합된 선호도 프로파일을 제공하기 위한 수단; 디지털 어시스턴트에 의해, 병합된 선호도 프로파일에 기초하여 미디어 아이템을 식별하기 위한 수단; 및 디지털 어시스턴트에 의해, 식별된 미디어 아이템을 포함하는 오디오 출력을 제공하기 위한 수단을 포함한다.
- [0026] 미디어 아이템을 제공하는 동안 자연 언어 스피치 입력을 수신하는 것은 사용자로 하여금 바람직한 콘텐츠를 획득하기 위해 미디어 검색을 쉽게 조정하게 한다. 디지털 어시스턴트는, 사용자로 하여금 현재 재생을 중지하거나 디지털 어시스턴트에 의한 추구를 기다리지 않으면서, 언제든지 미디어 요청을 정제하게 한다. 이와 같이, 디지털 어시스턴트는 사용자에게 미디어 검색 프로세스에 대한 완전하고 유연한 제어를 제공한다. 또한, 미디어 아이템을 제공하는 동안 자연 언어 스피치 입력을 수신하는 것은, 디지털 어시스턴트가 사용자로 하여금 대화에 개입하여 임의의 주어진 시간에서 대화를 조종할 수 있게 함으로써, 디지털 어시스턴트와 사용자 간에 자연스럽게 직관적이며 인간과 유사한 상호작용을 제공한다. 미디어 검색 프로세스의 유연하고 직관적인 제어를 제공하는 것은, 디바이스의 동작성을 향상시키고 디지털 어시스턴트와의 상호작용을 보다 효율적으로 만드는데 (예를 들어, 사용자 의도를 이해하고 사용자에게 전체 제어를 부여함으로써), 이는 추가적으로, 사용자가 디바이스를 보다 신속하고 효율적으로 사용하는 것을 가능하게 함으로써 디바이스의 전력 사용량을 감소시키고 배터리 수명을 개선시킨다.
- [0027] 자연 언어 스피치 입력이 미디어 요청을 정제하고 이에 따라 미디어 아이템을 식별하려는 사용자 의도에 대응하는지 여부를 결정하는 것은 사용자로 하여금 비교적 적은 수의 입력으로 바람직한 콘텐츠를 신속하게 획득하게 한다. 기술은, 예를 들어, 사용자가 미디어 요청을 정제할 때 이전에 특정된 파라미터들을 반복적으로 제공할 필요가 없기 때문에 사용자 입력들의 수를 감소시킨다. 기술은, 또한 예를 들어, 사용자가 점점 더 맞춤형되고 일련의 결정을 통해 바람직한 콘텐츠로 좁혀지는 추천들을 수신할 수 있기 때문에 디지털 어시스턴트와 사용자 간에 자연스럽게 직관적인 상호작용을 제공한다. 사용자 입력의 수를 감소시키고 직관적인 사용자 인터페이스를 제공하는 것은, (예를 들어, 디바이스를 동작시키거나/그와 상호작용할 때 사용자가 적절한 입력들을 제공하는 것을 돕고 사용자 실수들을 감소시킴으로써) 디바이스의 동작성을 향상시키고 사용자-디바이스 인터페이스를 보다 효율적으로 만드는데, 이는 추가적으로, 사용자가 디바이스를 보다 신속하고 효율적으로 사용하는 것을 가능하게 함으로써 디바이스의 전력 사용량을 감소시키고 배터리 수명을 개선시킨다.
- [0028] 사용자 요청에 대한 구두 응답을 나타내는 스피치 출력을 제공하는 동안 또한 관련 미디어 아이템의 재생을 제공하는 것은, 디지털 어시스턴트의 풍부하고 직관적인 청각 인터페이스를 제공한다. 미디어 아이템(예를 들어, 관련 사운드 효과, 콘텐츠의 대표 샘플)의 재생은 사용자가 제시되고 있는 콘텐츠를 신속하게 이해하고, 오디오 출력의 지속기간을 연장하지 않으면서 보다 많은 정보에 입각한 결정을 내리는 데 도움을 준다. 또한, 사용자가 보다 정보에 입각한 결정을 내리는 것을 가능하게 하는 것은 사용자 입력들의 수를 감소시킨다. 풍부하고 직관적인 청각 인터페이스를 제공하는 것은, (예를 들어, 디바이스를 동작시키거나/그와 상호작용할 때 사용자

가 적절한 입력들을 제공하는 것을 돕고 사용자 실수들을 감소시킴으로써) 디바이스의 동작성을 향상시키고 사용자-디바이스 인터페이스를 보다 효율적으로 만드는데, 이는 추가적으로, 사용자가 디바이스를 보다 신속하고 효율적으로 사용하는 것을 가능하게 함으로써 디바이스의 전력 사용량을 감소시키고 배터리 수명을 개선시킨다.

[0029] 추천에 대한 연속적인 비-긍정 응답들의 수가 임계치를 만족시키는지 여부를 결정하고 그렇지 않은 경우 다른 추천을 제공하는 것은, 디지털 어시스턴트로 하여금 사용자에게 옵션을 신속하고 직관적으로 제공하게 한다. 기술은 사용자가 새로운 추천을 반복적으로 요청하지 않아도 되므로 사용자 입력들의 수를 감소시킨다. 추천들을 획득하기 위한 사용자 입력들의 수를 감소시키는 것은, (예를 들어, 디바이스를 동작시키거나/그와 상호작용할 때 사용자가 적절한 입력들을 제공하는 것을 돕고 사용자 실수들을 감소시킴으로써) 디바이스의 동작성을 향상시키고 사용자-디바이스 인터페이스를 보다 효율적으로 만드는데, 이는 추가적으로, 사용자가 디바이스를 보다 신속하고 효율적으로 사용하는 것을 가능하게 함으로써 디바이스의 전력 사용량을 감소시키고 배터리 수명을 개선시킨다.

[0030] 연속적인 비-긍정 응답들의 수가 임계치를 만족시키는지 여부를 결정하고 그렇다면 사용자 입력을 요청하는 것은, 디지털 어시스턴트로 하여금 사용자를 위해 바람직한 콘텐츠를 신속하게 식별하게 한다. 기술은 사용자가 바람직하지 않은 추천을 반복적으로 거부하고, 디지털 어시스턴트가 바람직하지 않은 추천을 제공하지 못하도록 하고/하거나, 새 검색을 시작할 필요가 없기 때문에 사용자 입력들의 수를 감소시킨다. 기술은 또한 디지털 어시스턴트가 사용자 커맨드 없이 적절할 때 정보를 자동으로 촉구하기 때문에 디지털 어시스턴트와 사용자 간에 자연스럽게 직관적인 상호작용을 제공한다. 사용자 입력의 수를 감소시키고 자연 사용자 인터페이스를 제공하는 것은, (예를 들어, 디바이스를 동작시키거나/그와 상호작용할 때 사용자가 적절한 입력들을 제공하는 것을 돕고 사용자 실수들을 감소시킴으로써) 디바이스의 동작성을 향상시키고 사용자-디바이스 인터페이스를 보다 효율적으로 만드는데, 이는 추가적으로, 사용자가 디바이스를 보다 신속하고 효율적으로 사용하는 것을 가능하게 함으로써 디바이스의 전력 사용량을 감소시키고 배터리 수명을 개선시킨다.

[0031] 다수의 사용자들의 물리적 존재를 검출하고 이러한 사용자들의 선호도 프로파일에 기초하여 병합된 선호도 프로파일을 제공하는 것은 다수의 사용자들에 대해 바람직한 콘텐츠의 신속한 식별을 허용한다. 이 기술은 다수의 사용자들에 대한 인지적 부담을 줄여 이들 사이의 공통 선호도를 식별하고, 디지털 어시스턴트에 대한 공통 선호도를 특정하고/하거나 바람직하지 않은 추천을 거부하는 데 필요한 입력들의 수를 감소시킨다. 사용자 입력들의 수 및 인지적 부담을 감소시키는 것은, (예를 들어, 디바이스를 동작시키거나/그와 상호작용할 때 사용자가 적절한 입력들을 제공하는 것을 돕고 사용자 실수들을 감소시킴으로써) 디바이스의 동작성을 향상시키고 사용자-디바이스 인터페이스를 보다 효율적으로 만드는데, 이는 추가적으로, 사용자가 디바이스를 보다 신속하고 효율적으로 사용하는 것을 가능하게 함으로써 디바이스의 전력 사용량을 감소시키고 배터리 수명을 개선시킨다.

도면의 간단한 설명

- [0032] 도 1은 다양한 예들에 따른, 디지털 어시스턴트를 구현하기 위한 시스템 및 환경을 도시하는 블록도이다.
- 도 2a는 다양한 예들에 따른, 디지털 어시스턴트의 클라이언트 측 부분을 구현하는 휴대용 다기능 디바이스를 도시하는 블록도이다.
- 도 2b는 다양한 예들에 따른, 이벤트 처리를 위한 예시적인 컴포넌트들을 도시하는 블록도이다.
- 도 3은 다양한 예들에 따른, 디지털 어시스턴트의 클라이언트 측 부분을 구현하는 휴대용 다기능 디바이스를 도시한다.
- 도 4는 다양한 예들에 따른, 디스플레이 및 터치 감응형 표면을 갖는 예시적인 다기능 디바이스의 블록도이다.
- 도 5a는 다양한 예들에 따른, 휴대용 다기능 디바이스 상의 애플리케이션들의 메뉴에 대한 예시적인 사용자 인터페이스를 도시한다.
- 도 5b는 다양한 예들에 따른, 디스플레이로부터 분리된 터치 감응형 표면을 갖는 다기능 디바이스에 대한 예시적인 사용자 인터페이스를 도시한다.
- 도 6a는 다양한 예들에 따른 개인 전자 디바이스를 도시한다.
- 도 6b는 다양한 예들에 따른 개인 전자 디바이스를 도시하는 블록도이다.
- 도 7a는 다양한 예들에 따른 디지털 어시스턴트 시스템 또는 그의 서버 부분을 도시하는 블록도이다.

도 7b는 다양한 예들에 따른, 도 7a에 도시된 디지털 어시스턴트의 기능들을 도시한다.

도 7c는 다양한 예들에 따른 온톨로지(ontology)의 일부분을 도시한다.

도 8a 및 도 8b는 일부 실시예들에 따른 전자 디바이스의 예시적인 사용자 인터페이스들을 도시한다.

도 9a 및 도 9b는 일부 실시예들에 따른 전자 디바이스의 예시적인 사용자 인터페이스들을 도시한다.

도 10a 및 도 10b는 일부 실시예들에 따른 전자 디바이스의 예시적인 사용자 인터페이스들을 도시한다.

도 11은 일부 실시예들에 따른 전자 디바이스의 예시적인 사용자 인터페이스들을 도시한다.

도 12는 다양한 예들에 따른, 디지털 어시스턴트의 청각 기반 인터페이스를 제공하기 위한 프로세스를 도시한다.

도 13은 다양한 예들에 따른, 디지털 어시스턴트의 청각 기반 인터페이스를 제공하기 위한 프로세스를 도시한다.

도 14는 다양한 예들에 따른, 디지털 어시스턴트의 청각 기반 인터페이스를 제공하기 위한 프로세스를 도시한다.

도 15는 다양한 예들에 따른, 디지털 어시스턴트의 청각 기반 인터페이스를 제공하기 위한 프로세스를 도시한다.

발명을 실시하기 위한 구체적인 내용

[0033] 예들의 다음 설명에서, 첨부된 도면들이 참조되며, 실시될 수 있는 특정 예들이 도면들 내에서 예시로서 도시된다. 다양한 예들의 범주를 벗어나지 않으면서 다른 예들이 이용될 수 있고 구조적 변경이 행해질 수 있다는 것이 이해되어야 한다.

[0034] 상기 논의된 바와 같이, 디지털 어시스턴트는 디스플레이 능력이 제한되거나 전혀 없는 전자 디바이스 상에 구현될 수 있다. 따라서, 주 청각 기반 인터페이스(primarily auditory-based interface)를 제공하는 전자 디바이스(또는 전자 디바이스 상의 디지털 어시스턴트)에 대한 필요성이 있다. 또한, 디지털 어시스턴트와 사용자 간에 자연스럽게 직관적이며 풍부한 상호작용을 제공하기 위한 이러한 주 청각 기반 인터페이스에 대한 필요성이 있다. 이러한 기술들은 디바이스의 동작성을 향상시키고 디지털 어시스턴트와의 상호작용을 보다 효율적으로 만들 수 있는데(예를 들어, 사용자가 제시되고 있는 콘텐츠를 신속하게 획득하고 이해하게 도와줌으로써), 이는 추가적으로, 사용자가 디바이스를 보다 신속하고 효율적으로 사용하는 것을 가능하게 함으로써 디바이스의 전력 사용량을 감소시키고 배터리 수명을 개선시킨다.

[0035] 이하의 설명이 다양한 요소들을 기술하기 위해 "제1", "제2" 등과 같은 용어들을 사용하지만, 이 요소들이 그 용어들에 의해 제한되어서는 안된다. 이들 용어들은 하나의 요소를 다른 요소와 구별하는 데에만 사용된다. 예를 들어, 다양하게 기술된 예들의 범주로부터 벗어남이 없이, 제1 입력이 제2 입력으로 지칭될 수 있을 것이고, 이와 유사하게, 제2 입력이 제1 입력으로 지칭될 수 있을 것이다. 제1 입력 및 제2 입력은 둘 모두 입력들이고, 일부 경우들에서, 별개이고 상이한 입력들이다.

[0036] 본 명세서에서 다양하게 기술된 예들의 설명에 사용되는 용어는 특정 예들을 기술하는 목적만을 위한 것이고, 제한하려는 의도는 아니다. 다양하게 기술된 예들의 설명 및 첨부된 청구범위에 사용되는 바와 같이, 단수의 형태("a", "an", 및 "the")는 문맥상 명백히 달리 나타내지 않는다면 복수의 형태도 마찬가지로 포함하려는 것으로 의도된다. 또한, 본 명세서에서 사용되는 바와 같은 "및/또는"이라는 용어는 열거되는 연관된 항목들 중 하나 이상의 항목들의 임의의 그리고 모든 가능한 조합들을 나타내고 그들을 포괄하는 것임이 이해될 것이다. 용어들 "포함한다(include)", "포함하는(including)", "포함한다(comprise)", 및/또는 "포함하는(comprising)"은, 본 명세서에서 사용될 때, 언급된 특징들, 정수들, 단계들, 동작들, 요소들, 및/또는 컴포넌트들의 존재를 특정하지만, 하나 이상의 다른 특징들, 정수들, 단계들, 동작들, 요소들, 컴포넌트들, 및/또는 이들의 그룹들의 존재 또는 추가를 배제하지 않음이 추가로 이해될 것이다.

[0037] 용어 "~할 경우(if)"는 문맥에 따라 "~할 때(when)" 또는 "~할 시(upon)" 또는 "결정하는 것에 응답하여(in response to determining)" 또는 "검출하는 것에 응답하여(in response to detecting)"를 의미하는 것으로 해석될 수 있다. 유사하게, 구절(phrase) "~라고 결정된 경우" 또는 "[언급된 조건 또는 이벤트가] 검출된 경우"는 문맥에 따라 "~라고 결정할 시" 또는 "~라고 결정하는 것에 응답하여" 또는 "[언급된 조건 또는 이벤트]"

를 검출할 시" 또는 "[언급된 조건 또는 이벤트]를 검출하는 것에 응답하여"를 의미하는 것으로 해석될 수 있다.

[0038] 1. 시스템 및 환경

[0039] 도 1은 다양한 예들에 따른 시스템(100)의 블록도를 도시한다. 일부 예들에서, 시스템(100)은 디지털 어시스턴트를 구현한다. 용어들 "디지털 어시스턴트", "가상 어시스턴트", "지능형 자동화된 어시스턴트", 또는 "자동 디지털 어시스턴트"는 음성 및/또는 텍스트 형태로 입력되는 자연 언어를 해석하여 사용자 의도를 추론하고, 추론된 사용자 의도에 기초하여 행동들을 수행하는 임의의 정보 프로세싱 시스템을 지칭한다. 예를 들어, 추론된 사용자 의도에 따라 동작하기 위해, 시스템은 다음 중 하나 이상을 수행한다: 추론된 사용자 의도를 달성하도록 설계된 단계들 및 파라미터들을 이용하여 태스크 흐름을 식별하는 것; 추론된 사용자 의도로부터의 특정 조건들을 태스크 흐름 내에 입력하는 것; 프로그램들, 메소드들, 서비스들, API들 등을 호출(invoke)함으로써 태스크 흐름을 실행하는 것; 및 청각적(예컨대, 스피치) 및/또는 시각적 형태로 사용자에게 대한 출력 응답들을 생성하는 것.

[0040] 구체적으로, 디지털 어시스턴트는 적어도 부분적으로 자연 언어 커맨드, 요청, 진술, 서술, 및/또는 쿼리의 형태로 사용자 요청을 받아들일 수 있다. 전형적으로, 사용자 요청은 디지털 어시스턴트에 의해 정보제공형 답변 또는 태스크의 수행 중 어느 하나를 구한다. 사용자 요청에 대한 만족스러운 응답은 요청된 정보제공형 답변의 제공, 요청된 태스크의 수행, 또는 이 둘의 조합을 포함한다. 예를 들어, 사용자는 디지털 어시스턴트에게, "내가 지금 어디에 있지?"와 같은 질문을 한다. 사용자의 현재 위치에 기초하여, 디지털 어시스턴트는 "당신은 웨스트 게이트 근처 센트럴 파크에 있습니다"라고 답변한다. 사용자는 또한 태스크의 수행, 예를 들어 "다음주 내 여자친구의 생일 파티에 내 친구들을 초대해 줘"를 요청한다. 이에 응답하여, 디지털 어시스턴트는 "네, 바로 하겠습니다"라고 말함으로써 요청을 승인확인(acknowledge)할 수 있고, 이어서 사용자의 전자 주소록에 열거된 사용자의 친구들 각각에게 사용자를 대신하여 적절한 행사 예정 초대장(calendar invite)을 전송할 수 있다. 요청된 태스크의 수행 동안, 디지털 어시스턴트는 때때로 연장된 기간 내내 다수의 정보 교환들을 수반하는 연속 대화로 사용자와 상호작용한다. 디지털 어시스턴트와 상호작용하여 정보 또는 다양한 태스크의 수행을 요청하는 수많은 다른 방식이 있다. 구두 응답들을 제공하는 것 및 프로그래밍된 행동들을 취하는 것 외에도, 디지털 어시스턴트는 또한 예컨대, 텍스트, 알림들, 음악, 비디오들, 애니메이션들 등과 같은 다른 시각적 또는 청각적 형태들의 응답들을 제공한다.

[0041] 도 1에 도시된 바와 같이, 일부 예들에서, 디지털 어시스턴트는 클라이언트-서버 모델에 따라 구현된다. 디지털 어시스턴트는 사용자 디바이스(104) 상에서 실행되는 클라이언트 측 부분(102)(이하, "DA 클라이언트(102)"), 및 서버 시스템(108) 상에서 실행되는 서버 측 부분(106)(이하, "DA 서버(106)")을 포함한다. DA 클라이언트(102)는 하나 이상의 네트워크들(110)을 통해 DA 서버(106)와 통신한다. DA 클라이언트(102)는 사용자 대면 입력 및 출력 프로세싱, 및 DA 서버(106)와의 통신과 같은 클라이언트 측 기능들을 제공한다. DA 서버(106)는 개개의 사용자 디바이스(104) 상에 각각 존재하는 임의의 수의 DA 클라이언트들(102)에 서버 측 기능들을 제공한다.

[0042] 일부 예들에서, DA 서버(106)는 클라이언트 대면 I/O 인터페이스(112), 하나 이상의 프로세싱 모듈들(114), 데이터 및 모델들(116), 및 외부 서비스들에 대한 I/O 인터페이스(118)를 포함한다. 클라이언트 대면 I/O 인터페이스(112)는 DA 서버(106)에 대한 클라이언트 대면 입력 및 출력 프로세싱을 용이하게 한다. 하나 이상의 프로세싱 모듈들(114)은 데이터 및 모델들(116)을 이용하여 스피치 입력을 프로세싱하고 자연 언어 입력에 기초하여 사용자의 의도를 결정한다. 또한, 하나 이상의 프로세싱 모듈들(114)은 추론된 사용자 의도에 기초하여 태스크 실행을 수행한다. 일부 예들에서, DA 서버(106)는 태스크 완수 또는 정보 획득을 위해 네트워크(들)(110)를 통해 외부 서비스들(120)과 통신한다. 외부 서비스들에 대한 I/O 인터페이스(118)는 그러한 통신을 용이하게 한다.

[0043] 사용자 디바이스(104)는 임의의 적합한 전자 디바이스일 수 있다. 일부 예들에서, 사용자 디바이스는 휴대용 다기능 디바이스(예컨대, 디바이스(200)로서, 도 2a를 참조하여 하기에 기술됨), 다기능 디바이스(예컨대, 디바이스(400), 도 4를 참조하여 하기에 기술됨), 또는 개인 전자 디바이스(예컨대, 디바이스(600), 도 6a 및 도 6b를 참조하여 하기에 기술됨)이다. 휴대용 다기능 디바이스는 PDA 및/또는 음악 재생기 기능들과 같은 다른 기능들을 또한 포함하는, 예를 들어, 모바일 전화기이다. 휴대용 다기능 디바이스들의 특정 예들은 미국 캘리포니아주 쿠퍼티노 소재의 애플 인크.(Apple Inc.)로부터의 아이폰(iPhone)®, 아이팟 터치(iPod Touch)®, 및 아이패드(iPad)® 디바이스들을 포함한다. 휴대용 다기능 디바이스들의 다른 예들은 랩톱 또는 태블릿 컴퓨터들

을 제한없이 포함한다. 또한, 일부 예들에서, 사용자 디바이스(104)는 비휴대용 다기능 디바이스이다. 특히, 사용자 디바이스(104)는 데스크톱 컴퓨터, 게임 콘솔, 텔레비전, 또는 텔레비전 셋톱 박스이다. 일부 예들에서, 사용자 디바이스(104)는 터치 감응형 표면(예컨대, 터치 스크린 디스플레이들 및/또는 터치패드들)을 포함한다. 또한, 사용자 디바이스(104)는 옵션적으로 물리적 키보드, 마우스, 및/또는 조이스틱과 같은 하나 이상의 다른 물리적 사용자 인터페이스 디바이스들을 포함한다. 다기능 디바이스들과 같은 전자 디바이스들의 다양한 예들이 더욱 상세하게 하기에 기술된다.

[0044] 통신 네트워크(들)(110)의 예들은 근거리 통신망(LAN) 및 광역 통신망(WAN), 예컨대 인터넷을 포함한다. 통신 네트워크(들)(110)는, 예를 들어 이더넷(Ethernet), 범용 직렬 버스(Universal Serial Bus; USB), 파이어와이어(FIREWIRE), GSM(Global System for Mobile Communications), EDGE(Enhanced Data GSM Environment), 코드 분할 다중 접속(code division multiple access; CDMA), 시간 분할 다중 접속(time division multiple access; TDMA), 블루투스(Bluetooth), Wi-Fi, VoIP(voice over Internet Protocol), Wi-MAX, 또는 임의의 다른 적합한 통신 프로토콜과 같은 다양한 유선 또는 무선 프로토콜들을 포함하는 임의의 알려진 네트워크 프로토콜을 이용하여 구현된다.

[0045] 서버 시스템(108)은 하나 이상의 독립형 데이터 프로세싱 장치들 또는 분산형 컴퓨터 네트워크 상에 구현된다. 일부 예들에서, 서버 시스템(108)은 또한 서버 시스템(108)의 기본 컴퓨팅 리소스들 및/또는 인프라구조 리소스들을 제공하기 위해 제3자 서비스 제공자들(예컨대, 제3자 클라우드 서비스 제공자들)의 다양한 가상 디바이스들 및/또는 서비스들을 채용한다.

[0046] 일부 예들에서, 사용자 디바이스(104)는 제2 사용자 디바이스(122)를 통해 DA 서버(106)와 통신한다. 제2 사용자 디바이스(122)는 사용자 디바이스(104)와 유사 또는 동일하다. 예를 들어, 제2 사용자 디바이스(122)는 도 2a, 도 4, 및 도 6a와 도 6b를 참조하여 하기에 기술되는 디바이스들(200, 400, 또는 600)과 유사하다. 사용자 디바이스(104)는 블루투스, NFC, BTLE 등과 같은 직접 통신 접속을 통해, 또는 로컬 Wi-Fi 네트워크와 같은 유선 또는 무선 네트워크를 통해 제2 사용자 디바이스(122)에 통신 가능하게 결합되도록 구성된다. 일부 예들에서, 제2 사용자 디바이스(122)는 사용자 디바이스(104)와 DA 서버(106) 사이의 프록시(proxy)로서 동작하도록 구성된다. 예를 들어, 사용자 디바이스(104)의 DA 클라이언트(102)는 제2 사용자 디바이스(122)를 통해 DA 서버(106)로 정보(예컨대, 사용자 디바이스(104)에서 수신된 사용자 요청)를 송신하도록 구성된다. DA 서버(106)는 정보를 프로세싱하고 제2 사용자 디바이스(122)를 통해 사용자 디바이스(104)로 관련 데이터(예컨대, 사용자 요청에 응답하는 데이터 콘텐츠)를 반환한다.

[0047] 일부 예들에서, 사용자 디바이스(104)는 사용자 디바이스(104)로부터 송신되는 정보의 양을 감소시키기 위해 데이터에 대한 단축 요청(abbreviated request)들을 제2 사용자 디바이스(122)로 통신하도록 구성된다. 제2 사용자 디바이스(122)는 DA 서버(106)에 송신할 완전한 요청을 생성하기 위해 단축 요청에 추가하기 위한 보완 정보를 결정하도록 구성된다. 이러한 시스템 아키텍처는, 유리하게도, 제한된 통신 능력들 및/또는 제한된 배터리 전력을 갖는 사용자 디바이스(104)(예컨대, 시계 또는 유사한 소형 전자 디바이스)가, 보다 큰 통신 능력들 및/또는 배터리 전력을 갖는 제2 사용자 디바이스(122)(예컨대, 모바일 전화기, 랩톱 컴퓨터, 태블릿 컴퓨터 등)를 DA 서버(106)에 대한 프록시로서 사용함으로써, DA 서버(106)에 의해 제공되는 서비스들에 액세스하도록 할 수 있다. 2개의 사용자 디바이스들(104, 122)만이 도 1에 도시되어 있지만, 시스템(100)은 일부 예들에서 DA 서버 시스템(106)과 통신하기 위해 이러한 프록시 구성으로 구성되는 임의의 개수 및 유형의 사용자 디바이스들을 포함한다는 것이 이해될 것이다.

[0048] 도 1에 도시된 디지털 어시스턴트가 클라이언트 측 부분(예컨대, DA 클라이언트(102)) 및 서버 측 부분(예컨대, DA 서버(106)) 둘 모두를 포함하지만, 일부 예들에서, 디지털 어시스턴트의 기능들은 사용자 디바이스 상에 설치된 독립형 애플리케이션으로서 구현된다. 게다가, 디지털 어시스턴트의 클라이언트 측 부분과 서버 측 부분 사이의 기능들의 분담은 상이한 구현예들에서 다를 수 있다. 예를 들어, 일부 예들에서, DA 클라이언트는, 로지 사용자 대면 입력 및 출력 프로세싱 기능들만을 제공하고 디지털 어시스턴트의 다른 모든 기능들을 백엔드 서버에 위임하는 썬-클라이언트(thin-client)이다.

[0049] 2. 전자 디바이스

[0050] 이제, 디지털 어시스턴트의 클라이언트 측 부분을 구현하기 위한 전자 디바이스들의 실시예들에 주목한다. 도 2a는 일부 실시예들에 따른, 터치 감응형 디스플레이 시스템(212)을 갖는 휴대용 다기능 디바이스(200)를 도시하는 블록도이다. 터치 감응형 디스플레이(212)는 때때로 편의상 "터치 스크린"이라고 지칭되고, 때때로 "터치 감응형 디스플레이 시스템"으로 알려지거나 지칭된다. 디바이스(200)는 메모리(202)(옵션적으로, 하나 이상의

컴퓨터 판독가능 저장 매체들을 포함함), 메모리 제어기(222), 하나 이상의 프로세싱 유닛들(CPU)(220), 주변기기 인터페이스(218), RF 회로부(208), 오디오 회로부(210), 스피커(211), 마이크로폰(213), 입/출력(I/O) 서브 시스템(206), 다른 입력 제어 디바이스들(216), 및 외부 포트(224)를 포함한다. 디바이스(200)는 옵션적으로 하나 이상의 광 센서(optical sensor)(264)들을 포함한다. 디바이스(200)는, 옵션적으로, 디바이스(200)(예컨대, 디바이스(200)의 터치 감응형 디스플레이 시스템(212)과 같은 터치 감응형 표면) 상의 접촉들의 세기를 검출하기 위한 하나 이상의 접촉 세기 센서들(265)을 포함한다. 디바이스(200)는, 옵션적으로, 디바이스(200) 상의 촉각적 출력들을 생성하기(예컨대, 디바이스(200)의 터치 감응형 디스플레이 시스템(212) 또는 디바이스(400)의 터치패드(455)와 같은 터치 감응형 표면 상의 촉각적 출력들을 생성하기) 위한 하나 이상의 촉각적 출력 생성기들(267)을 포함한다. 이들 컴포넌트는 옵션적으로 하나 이상의 통신 버스들 또는 신호 라인(203)들을 통해 통신한다.

[0051] 명세서 및 청구범위에서 사용되는 바와 같이, 터치 감응형 표면 상의 접촉의 "세기"라는 용어는 터치 감응형 표면 상의 접촉(예컨대, 손가락 접촉)의 힘 또는 압력(단위 면적 당 힘), 또는 터치 감응형 표면 상의 접촉의 힘 또는 압력에 대한 대체물(대용물(proxy))을 지칭한다. 접촉의 세기는, 적어도 4개의 구별되는 값들을 포함하고 더 전형적으로는 수백 개(예컨대, 적어도 256개)의 구별되는 값들을 포함하는 일정 범위의 값들을 갖는다. 접촉의 세기는, 옵션적으로, 다양한 접근법들, 및 다양한 센서들 또는 센서들의 조합들을 이용하여 결정(또는 측정)된다. 예를 들어, 터치 감응형 표면 아래의 또는 그에 인접한 하나 이상의 힘 센서들은 터치 감응형 표면 상의 다양한 지점들에서 힘을 측정하는 데 옵션적으로 사용된다. 일부 구현예들에서는, 다수의 힘 센서들로부터의 힘 측정치들이 접촉의 추정되는 힘을 결정하기 위해 조합(예컨대, 가중 평균)된다. 유사하게, 스타일러스의 압력 감응형 팁(tip)이 터치 감응형 표면 상의 스타일러스의 압력을 결정하는 데 옵션적으로 사용된다. 대안적으로, 터치 감응형 표면 상에서 검출된 접촉 면적의 크기 및/또는 그에 대한 변화들, 접촉 부근의 터치 감응형 표면의 용량 및/또는 그에 대한 변화들, 및/또는 접촉 부근의 터치 감응형 표면의 저항 및/또는 그에 대한 변화들은 터치 감응형 표면 상의 접촉의 힘 또는 압력에 대한 대체물로서 옵션적으로 이용된다. 일부 구현예들에서, 접촉 힘 또는 압력에 대한 대체 측정치들은 세기 임계치가 초과되었는지 여부를 결정하는 데 직접 이용된다(예를 들어, 세기 임계치는 대체 측정치들에 대응하는 단위로 기술된다). 일부 구현예들에서, 접촉 힘 또는 압력에 대한 대체 측정치들은 추정된 힘 또는 압력으로 변환되고, 추정된 힘 또는 압력은 세기 임계치가 초과되었는지 여부를 결정하기 위해 이용된다(예를 들어, 세기 임계치는 압력의 단위로 측정된 압력 임계치이다). 사용자 입력의 속성으로서 접촉의 세기를 사용하는 것은, 그렇지 않았으면 어포던스들을 (예를 들어, 터치 감응형 디스플레이 상에) 디스플레이하고/하거나 (예를 들어, 터치 감응형 디스플레이, 터치 감응형 표면, 또는 노브(knob) 또는 버튼과 같은 물리적/기계적 제어부를 통해) 사용자 입력을 수신하기 위하여 한정된 실면적을 갖는 감소된 크기의 디바이스 상에서 사용자에게 의해 액세스 가능하지 않을 수 있는 부가적인 디바이스 기능에의 사용자 액세스를 가능하게 한다.

[0052] 명세서 및 청구범위에 사용되는 바와 같이, "촉각적 출력"이라는 용어는 디바이스의 이전 위치에 대한 디바이스의 물리적 변위, 디바이스의 다른 컴포넌트(예컨대, 하우징)에 대한 디바이스의 컴포넌트(예컨대, 터치 감응형 표면)의 물리적 변위, 또는 사용자의 촉각을 이용하여 사용자에게 의해 검출된 디바이스의 질량 중심에 대한 컴포넌트의 변위를 지칭한다. 예를 들어, 디바이스 또는 디바이스의 컴포넌트가 터치에 감응하는 사용자의 표면(예컨대, 손가락, 손바닥, 또는 사용자의 손의 기타 부분)과 접촉하는 상황들에서, 물리적 변위에 의해 생성된 촉각적 출력은 사용자에게 의해 디바이스 또는 디바이스의 컴포넌트의 물리적 특성들에서의 인지된 변화에 대응하는 촉감(tactile sensation)으로서 해석될 것이다. 예를 들어, 터치 감응형 표면(예를 들어, 터치 감응형 디스플레이 또는 트랙패드)의 이동은 옵션적으로 물리적 액추에이터 버튼의 "다운 클릭(down click)" 또는 "업 클릭(up click)"으로서 사용자에게 의해 해석된다. 일부 경우에, 사용자는 사용자의 이동에 의해 물리적으로 눌러는 (예를 들어, 변위되는) 터치 감응형 표면과 연관된 물리적 액추에이터 버튼의 이동이 없는 경우에도, "다운 클릭" 또는 "업 클릭"과 같은 촉감을 느낄 것이다. 다른 예로서, 터치 감응형 표면의 이동은, 터치 감응형 표면의 평활도(smoothness)의 변화가 없는 경우에도, 옵션적으로, 사용자에게 의해 터치 감응형 표면의 "거칠기(roughness)"로서 해석 또는 감지된다. 사용자에게 의한 터치의 이러한 해석들이 사용자의 개별화된 감각 인지(sensory perception)에 영향을 받을 것이지만, 대다수의 사용자들에게 보편적인 많은 터치 감각 인지가 있다. 따라서, 촉각적 출력이 사용자의 특정 감각 인지(예컨대, "업 클릭", "다운 클릭", "거칠기")에 대응하는 것으로서 기술될 때, 달리 언급되지 않는다면, 생성된 촉각적 출력은 전형적인(또는 평균적인) 사용자에게 대한 기술된 감각 인지를 생성할 디바이스 또는 그의 컴포넌트의 물리적 변위에 대응한다.

[0053] 디바이스(200)는 휴대용 다기능 디바이스의 일례일 뿐이고, 디바이스(200)는, 옵션적으로, 도시된 것보다 더 많거나 더 적은 컴포넌트들을 갖거나, 옵션적으로, 둘 이상의 컴포넌트들을 조합하거나, 또는 옵션적으로 컴포넌

트들의 상이한 구성 또는 배열을 갖는다는 것이 이해될 것이다. 도 2a에 도시된 다양한 컴포넌트들은 하나 이상의 신호 프로세싱 및/또는 ASIC(application-specific integrated circuit)들을 비롯한, 하드웨어, 소프트웨어, 또는 하드웨어와 소프트웨어 둘 모두의 조합으로 구현된다.

[0054] 메모리(202)는 하나 이상의 컴퓨터 판독가능 저장 매체들을 포함한다. 컴퓨터 판독가능 저장 매체들은 예를 들어 유형적이고 비일시적이다. 메모리(202)는, 고속 랜덤 액세스 메모리를 포함하고, 또한 하나 이상의 자기 디스크 저장 디바이스들, 플래시 메모리 디바이스들, 또는 다른 비휘발성 솔리드 스테이트 메모리 디바이스(non-volatile solid-state memory device)들과 같은 비휘발성 메모리를 포함한다. 메모리 제어기(222)는 디바이스(200)의 다른 컴포넌트들에 의한 메모리(202)에의 액세스를 제어한다.

[0055] 일부 예들에서, 메모리(202)의 비일시적 컴퓨터 판독가능 저장 매체는, 명령어 실행 시스템, 장치, 또는 디바이스, 예컨대 컴퓨터 기반 시스템, 프로세서-포함 시스템(processor-containing system), 또는 명령어 실행 시스템, 장치, 또는 디바이스로부터 명령어들을 페치(fetch)하여 명령어들을 실행할 수 있는 다른 시스템에 의해 또는 그와 관련하여 사용하기 위한(예컨대, 하기에 기술되는 프로세스들의 양태들을 수행하기 위한) 명령어들을 저장하는 데 사용된다. 다른 예들에서, (예컨대, 하기에 기술되는 프로세스들의 양태들을 수행하기 위한) 명령어들은 서버 시스템(108)의 비일시적 컴퓨터 판독가능 저장 매체(도시되지 않음) 상에 저장되거나 또는 메모리(202)의 비일시적 컴퓨터 판독가능 저장 매체와 서버 시스템(108)의 비일시적 컴퓨터 판독가능 저장 매체 사이에서 분담된다.

[0056] 주변기기 인터페이스(218)는 디바이스의 입력 및 출력 주변기기들을 CPU(220) 및 메모리(202)에 결합하는 데 사용된다. 하나 이상의 프로세서들(220)은 디바이스(200)에 대한 다양한 기능들을 수행하기 위해 그리고 데이터를 프로세싱하기 위해 메모리(202)에 저장된 다양한 소프트웨어 프로그램들 및/또는 명령어들의 세트들을 구동 또는 실행시킨다. 일부 실시예들에서, 주변기기 인터페이스(218), CPU(220) 및 메모리 제어기(222)는 칩(204)과 같은 단일 칩 상에 구현된다. 일부 다른 실시예들에서, 이들은 별개의 칩들 상에 구현된다.

[0057] RF(radio frequency) 회로부(208)는 전자기 신호들이라고도 지칭되는 RF 신호들을 수신 및 전송한다. RF 회로부(208)는 전기 신호들을 전자기 신호들로/로부터 변환하고, 전자기 신호들을 통해 통신 네트워크들 및 다른 통신 디바이스들과 통신한다. RF 회로부(208)는, 옵션적으로, 안테나 시스템, RF 송수신기, 하나 이상의 증폭기들, 튜너, 하나 이상의 발진기들, 디지털 신호 프로세서, CODEC 칩셋, 가입자 식별 모듈(SIM) 카드, 메모리 등을 포함하지만 이들로 한정되지 않는, 이러한 기능들을 수행하기 위한 잘 알려진 회로부를 포함한다. RF 회로부(208)는, 옵션적으로, 네트워크들, 예컨대 월드 와이드 웹(WWW)으로도 지칭되는 인터넷, 인트라넷, 및/또는 무선 네트워크, 예컨대 셀룰러 전화 네트워크, 무선 LAN(local area network) 및/또는 MAN(metropolitan area network), 및 다른 디바이스들과 무선 통신에 의해 통신한다. RF 회로부(208)는, 옵션적으로, 예컨대 단거리 통신 무선기기(short-range communication radio)에 의해, 근거리 통신(near field communication, NFC) 필드들을 검출하기 위한 잘 알려진 회로부를 포함한다. 무선 통신은, 옵션적으로, GSM(Global System for Mobile Communications), EDGE(Enhanced Data GSM Environment), HSDPA(high-speed downlink packet access), HSUPA(high-speed uplink packet access), EV-DO(Evolution, Data-Only), HSPA, HSPA+, DC-HSDPA(Dual-Cell HSPA), LTE(long term evolution), NFC(near field communication), W-CDMA(wideband code division multiple access), CDMA(code division multiple access), TDMA(time division multiple access), 블루투스, BTLE(Bluetooth Low Energy), Wi-Fi(Wireless Fidelity)(예컨대, IEEE 802.11a, IEEE 802.11b, IEEE 802.11g, IEEE 802.11n 및/또는 IEEE 802.11ac), VoIP(voice over Internet Protocol), Wi-MAX, 이메일용 프로토콜(예컨대, IMAP(Internet message access protocol) 및/또는 POP(post office protocol)), 인스턴트 메시징(예컨대, XMPP(extensible messaging and presence protocol), SIMPLE(Session Initiation Protocol for Instant Messaging and Presence Leveraging Extensions), IMPS(Instant Messaging and Presence Service)), 및/또는 SMS(Short message Service), 또는 본 문서의 출원일 당시 아직 개발되지 않은 통신 프로토콜들을 비롯한, 임의의 다른 적합한 통신 프로토콜을 포함하지만 이들로 한정되지 않는, 복수의 통신 표준들, 프로토콜들 및 기술들 중 임의의 것을 이용한다.

[0058] 오디오 회로부(210), 스피커(211), 및 마이크روف폰(213)은 사용자와 디바이스(200) 사이에서 오디오 인터페이스를 제공한다. 오디오 회로부(210)는 주변기기 인터페이스(218)로부터 오디오 데이터를 수신하고, 그 오디오 데이터를 전기 신호로 변환하고, 그 전기 신호를 스피커(211)에 송신한다. 스피커(211)는 전기 신호를 사람이 들을 수 있는 음파로 변환한다. 오디오 회로부(210)는 또한 마이크روف폰(213)에 의해 음파로부터 변환된 전기 신호를 수신한다. 오디오 회로부(210)는 전기 신호를 오디오 데이터로 변환하고, 프로세싱을 위해 오디오 데이터를 주변기기 인터페이스(218)에 송신한다. 오디오 데이터는 주변기기 인터페이스(218)에 의해 메모리(202) 및/

또는 RF 회로부(208)로부터 인출(retrieve)되고/되거나 이에 송신된다. 일부 실시예들에서, 오디오 회로부(210)는 또한 헤드셋 잭(예컨대, 도 3의 312)을 포함한다. 헤드셋 잭은 출력-전용 헤드폰들, 또는 출력(예컨대, 한쪽 또는 양쪽 귀용 헤드폰) 및 입력(예컨대, 마이크로폰) 양쪽 모두를 갖는 헤드셋과 같은 분리 가능한 오디오 입/출력 주변기기들과 오디오 회로부(210) 사이의 인터페이스를 제공한다.

[0059] I/O 서브시스템(206)은 터치 스크린(212) 및 다른 입력 제어 디바이스들(216)과 같은, 디바이스(200) 상의 입/출력 주변기기들을 주변기기 인터페이스(218)에 결합한다. I/O 서브시스템(206)은 옵션적으로 디스플레이 제어기(256), 광 센서 제어기(258), 세기 센서 제어기(259), 햅틱 피드백 제어기(261), 및 다른 입력 또는 제어 디바이스들을 위한 하나 이상의 입력 제어기들(260)을 포함한다. 하나 이상의 입력 제어기들(260)은 다른 입력 제어 디바이스들(216)로부터/로 전기 신호들을 수신/전송한다. 다른 입력 제어 디바이스들(216)은 옵션적으로 물리적 버튼들(예컨대, 푸시 버튼(push button), 로커 버튼(rocker button) 등), 다이얼, 슬라이더 스위치, 조이스틱, 클릭 휠 등을 포함한다. 일부 대안적인 실시예들에서, 입력 제어기(들)(260)은 옵션적으로 키보드, 적외선 포트, USB 포트, 및 마우스와 같은 포인터 디바이스 중 임의의 것에 결합된다(또는 어떤 것에도 결합되지 않는다). 하나 이상의 버튼들(예를 들어, 도 3의 308)은 옵션적으로 스피커(211) 및/또는 마이크로폰(213)의 음량 제어를 위한 업/다운 버튼을 포함한다. 하나 이상의 버튼들은 옵션적으로 푸시 버튼(예를 들어, 도 3의 306)을 포함한다.

[0060] 푸시 버튼의 빠른 누르기(quick press)는 터치 스크린(212)의 잠금을 풀거나, 디바이스를 잠금해제하기 위해 터치 스크린 상의 제스처들을 사용하는 프로세스를 시작하며, 이는 2005년 12월 23일자로 출원된 미국 특허 출원 제11/322,549호(미국 특허 제7,657,849호), "Unlocking a Device by Performing Gestures on an Unlock Image"에 기술된 바와 같으며, 이는 참조 문헌으로서 그 전문이 본 명세서에 편입된다. 푸시 버튼(예컨대, 306)의 더 긴 누르기는 디바이스(200)의 전원을 켜거나 끈다. 사용자는 버튼들 중 하나 이상의 버튼들의 기능을 사용자 지정(customize)할 수 있다. 터치 스크린(212)은 가상 또는 소프트 버튼들 및 하나 이상의 소프트 키보드들을 구현하는 데 사용된다.

[0061] 터치 감응형 디스플레이(212)는 디바이스와 사용자 사이의 입력 인터페이스 및 출력 인터페이스를 제공한다. 디스플레이 제어기(256)는 터치 스크린(212)으로부터/으로 전기 신호들을 수신하고/하거나 전송한다. 터치 스크린(212)은 사용자에게 시각적 출력을 디스플레이한다. 시각적 출력은 그래픽, 텍스트, 아이콘, 비디오 및 이들의 임의의 조합(총칭하여 "그래픽"으로 지칭됨)을 포함한다. 일부 실시예들에서, 시각적 출력의 일부 또는 전부가 사용자-인터페이스 객체들에 대응한다.

[0062] 터치 스크린(212)은 햅틱 및/또는 촉각적 접촉에 기초하는 사용자로부터의 입력을 수용하는 터치 감응형 표면, 센서 또는 센서들의 세트를 갖는다. 터치 스크린(212) 및 디스플레이 제어기(256)는 (메모리(202) 내의 임의의 연관된 모듈들 및/또는 명령어들의 세트들과 함께) 터치 스크린(212) 상의 접촉(및 접촉의 임의의 이동 또는 중단)을 검출하고, 검출된 접촉을 터치 스크린(212) 상에 디스플레이된 사용자 인터페이스 객체들(예컨대, 하나 이상의 소프트 키들, 아이콘들, 웹 페이지들 또는 이미지들)과의 상호작용으로 변환한다. 예시적인 실시예에서, 터치 스크린(212)과 사용자 사이의 접촉 지점은 사용자의 손가락에 대응한다.

[0063] 터치 스크린(212)은, LCD(liquid crystal display) 기술, LPD(light emitting polymer display) 기술, 또는 LED(light emitting diode) 기술을 이용하지만, 다른 실시예들에서는 다른 디스플레이 기술들이 이용될 수 있다. 터치 스크린(212) 및 디스플레이 제어기(256)는, 용량성, 저항성, 적외선, 및 표면 음향파 기술들뿐만 아니라 다른 근접 센서 어레이들, 또는 터치 스크린(212)과의 하나 이상의 접촉 지점들을 결정하기 위한 다른 요소들을 포함하지만 이들로 한정되지 않는, 현재 공지되어 있거나 추후에 개발되는 복수의 터치 감지 기술들 중 임의의 것을 사용하여, 접촉 및 그의 임의의 이동 또는 중단을 검출한다. 예시적인 실시예에서, 미국 캘리포니아주 쿠파티노 소재의 애플 인크.로부터의 아이폰® 및 아이팟 터치®에서 발견되는 것과 같은 투영형 상호 정전용량 감지 기술(projected mutual capacitance sensing technology)이 이용된다.

[0064] 터치 스크린(212)의 일부 실시예들에서의 터치 감응형 디스플레이는 이하의 미국 특허 제6,323,846호(Westerman 등), 제6,570,557호(Westerman 등), 및/또는 제6,677,932호(Westerman), 및/또는 미국 특허 공개 제2002/0015024A1호에 기술된 멀티 터치 감응형 터치패드들과 유사하며, 이 미국 특허들 및 미국 특허 공개 각각은 그 전체적으로 본 명세서에 참조로서 편입된다. 그러나, 터치 스크린(212)은 디바이스(200)로부터의 시각적 출력을 디스플레이하는 반면, 터치 감응형 터치패드들은 시각적 출력을 제공하지 않는다.

[0065] 터치 스크린(212)의 일부 실시예들에서의 터치 감응형 디스플레이는 하기 출원들에 기술되는 바와 같다: (1) 2006년 5월 2일자로 출원되고 발명의 명칭이 "Multipoint Touch Surface Controller"인 미국 특허 출원 제

11/381,313호; (2) 2004년 5월 6일자로 출원된 미국 특허 출원 제10/840,862호, "Multipoint Touchscreen"; (3) 2004년 7월 30일자로 출원된 미국 특허 출원 제10/903,964호, "Gestures For Touch Sensitive Input Devices"; (4) 2005년 1월 31일자로 출원된 미국 특허 출원 제11/048,264호, "Gestures For Touch Sensitive Input Devices"; (5) 2005년 1월 18일자로 출원된 미국 특허 출원 제11/038,590호, "Mode-Based Graphical User Interfaces For Touch Sensitive Input Devices"; (6) 2005년 9월 16일자로 출원된 미국 특허 출원 제 11/228,758호, "Virtual Input Device Placement On A Touch Screen User Interface"; (7) 2005년 9월 16일자로 출원된 미국 특허 출원 제11/228,700호, "Operation Of A Computer With A Touch Screen Interface"; (8) 2005년 9월 16일자로 출원된 미국 특허 출원 제11/228,737호, "Activating Virtual Keys Of A Touch-Screen Virtual Keyboard"; (9) 2006년 3월 3일자로 출원된 미국 특허 출원 제11/367,749호, "Multi-Functional Hand-Held Device". 이 출원들 모두는 그 전문이 본 명세서에 참조로서 편입된다.

- [0066] 터치 스크린(212)은, 예를 들어, 100 dpi를 초과하는 비디오 해상도를 갖는다. 일부 실시예들에서, 터치 스크린은 대략 160 dpi의 비디오 해상도를 갖는다. 사용자는 스타일러스, 손가락 등과 같은 임의의 적합한 물체 또는 부속물을 사용하여 터치 스크린(212)과 접촉한다. 일부 실시예들에서, 사용자 인터페이스는 주로 손가락 기반 접촉들 및 제스처들을 이용하여 동작하도록 설계되는데, 이는 터치 스크린 상에서의 손가락의 더 넓은 접촉 면적으로 인해 스타일러스 기반 입력보다 덜 정밀할 수 있다. 일부 실시예들에서, 디바이스는 대략적인 손가락 기반 입력을 사용자가 원하는 행동들을 수행하기 위한 정밀한 포인터/커서 위치 또는 커맨드로 변환한다.
- [0067] 일부 실시예들에서, 터치 스크린 외에도, 디바이스(200)는, 특정 기능들을 활성화 또는 비활성화시키기 위한 터치패드(도시되지 않음)를 포함한다. 일부 실시예들에서, 터치패드는, 터치 스크린과는 달리, 시각적 출력을 디스플레이하지 않는 디바이스의 터치 감응형 영역이다. 터치패드는 터치 스크린(212)과는 별개인 터치 감응형 표면 또는 터치 스크린에 의해 형성된 터치 감응형 표면의 연장부이다.
- [0068] 디바이스(200)는 또한 다양한 컴포넌트들에 전력을 공급하기 위한 전력 시스템(262)을 포함한다. 전력 시스템(262)은, 전력 관리 시스템, 하나 이상의 전원들(예컨대, 배터리, 교류 전류(alternating current: AC)), 재충전 시스템, 전력 고장 검출 회로, 전력 변환기 또는 인버터, 전력 상태 표시자(예컨대, 발광 다이오드(LED)), 및 휴대용 디바이스들 내에서의 전력의 생성, 관리 및 분배와 연관된 임의의 다른 컴포넌트들을 포함한다.
- [0069] 디바이스(200)는 또한 하나 이상의 광 센서들(264)을 포함한다. 도 2a는 I/O 서브시스템(206) 내의 광 센서 제어기(258)에 결합되는 광 센서를 도시한다. 광 센서(264)는 CCD(charge-coupled device) 또는 CMOS(complementary metal-oxide semiconductor) 포토트랜지스터들을 포함한다. 광 센서(264)는 하나 이상의 렌즈들을 통해 투영되는, 주변환경으로부터의 광을 수광하고, 그 광을 이미지를 표현하는 데이터로 변환한다. 이미징 모듈(243)(카메라 모듈로도 지칭됨)과 함께, 광 센서(264)는 정지 이미지들 또는 비디오를 캡처한다. 일부 실시예들에서, 광 센서는 디바이스 전면 상의 터치 스크린 디스플레이(212)의 반대편인 디바이스(200)의 배면 상에 위치되어, 터치 스크린 디스플레이가 정지 및/또는 비디오 영상 획득을 위한 뷰파인더로서 사용되게 한다. 일부 실시예들에서, 광 센서는 디바이스의 전면 상에 위치됨으로써, 사용자가 터치 스크린 디스플레이 상에서 다른 화상 회의 참가자들을 보는 동안 사용자의 이미지가 화상 회의를 위해 얻어진다. 일부 실시예들에서, 광 센서(264)의 위치는 (예컨대, 디바이스 하우징 내의 렌즈 및 센서를 회전시킴으로써) 사용자에게 의해 변경될 수 있어, 단일 광 센서(264)가 터치 스크린 디스플레이와 함께 화상 회의와 정지 및/또는 비디오 영상 획득 둘 모두에 사용된다.
- [0070] 디바이스(200)는 또한, 옵션적으로, 하나 이상의 접촉 세기 센서들(265)을 포함한다. 도 2a는 I/O 서브시스템(206) 내의 세기 센서 제어기(259)에 결합된 접촉 세기 센서를 도시한다. 접촉 세기 센서(265)는, 옵션적으로, 하나 이상의 압전 저항 변형 게이지들, 용량성 힘 센서들, 전기적 힘 센서들, 압전 힘 센서들, 광학적 힘 센서들, 용량성 터치 감응형 표면들, 또는 다른 세기 센서들(예컨대, 터치 감응형 표면 상의 접촉의 힘(또는 압력)을 측정하는 데 사용되는 센서들)을 포함한다. 접촉 세기 센서(265)는 주변환경으로부터 접촉 세기 정보(예컨대, 압력 정보 또는 압력 정보에 대한 대응물)를 수신한다. 일부 실시예들에서, 적어도 하나의 접촉 세기 센서는 터치 감응형 표면(예컨대, 터치 감응형 디스플레이 시스템(212))과 함께 위치(collocate)되거나 그에 근접한다. 일부 실시예들에서, 적어도 하나의 접촉 세기 센서는 디바이스(200)의 전면 상에 위치된 터치 스크린 디스플레이(212)의 반대편인 디바이스(200)의 배면 상에 위치된다.
- [0071] 디바이스(200)는 또한 하나 이상의 근접 센서들(266)을 포함한다. 도 2a는 주변기기 인터페이스(218)에 결합된 근접 센서(266)를 도시한다. 대안적으로, 근접 센서(266)는 I/O 서브시스템(206) 내의 입력 제어기(260)에 결합된다. 근접 센서(266)는 미국 특허 출원들 제11/241,839호, "Proximity Detector In Handheld Device"; 제

11/240,788호, "Proximity Detector In Handheld Device"; 제11/620,702호, "Using Ambient Light Sensor To Augment Proximity Sensor Output"; 제11/586,862호, "Automated Response To And Sensing Of User Activity In Portable Devices"; 및 제11/638,251호, "Methods And Systems For Automatic Configuration Of Peripherals"에 기술된 바와 같이 수행되며, 이들은 그 전문이 본 명세서에 참조로서 편입된다. 일부 실시예들에서, 근접 센서는 다기능 디바이스가 사용자의 귀 근처에 위치될 때(예컨대, 사용자가 전화 통화를 하고 있을 때) 터치 스크린(212)을 끄고 디스에이블시킨다.

[0072] 디바이스(200)는 또한, 옵션적으로, 하나 이상의 촉각적 출력 생성기들(267)을 포함한다. 도 2a는 I/O 서브시스템(206) 내의 햅틱 피드백 제어기(261)에 결합된 촉각적 출력 생성기를 도시한다. 촉각적 출력 생성기(267)는, 옵션적으로, 스피커들 또는 다른 오디오 컴포넌트들과 같은 하나 이상의 전자음향 디바이스들 및/또는 모터, 솔레노이드, 전기활성 중합체, 압전 액추에이터, 정전 액추에이터, 또는 다른 촉각적 출력 생성 컴포넌트(예컨대, 전기 신호들을 디바이스 상의 촉각적 출력들로 변환하는 컴포넌트)와 같은, 에너지를 선형 모션(linear motion)으로 변환하는 전자기계 디바이스들을 포함한다. 접촉 세기 센서(265)는 햅틱 피드백 모듈(233)로부터 촉각적 피드백 생성 명령어들을 수신하여 디바이스(200)의 사용자에게 의해 감지될 수 있는 디바이스(200) 상의 촉각적 출력들을 생성한다. 일부 실시예들에서, 적어도 하나의 촉각적 출력 생성기는 터치 감응형 표면(예컨대, 터치 감응형 디스플레이 시스템(212))과 함께 위치되거나 그에 근접하며, 옵션적으로, 터치 감응형 표면을 수직으로(예컨대, 디바이스(200)의 표면 내/외로) 또는 측방향으로(예컨대, 디바이스(200)의 표면과 동일한 평면에서 전후로) 이동시킴으로써 촉각적 출력을 생성한다. 일부 실시예들에서, 적어도 하나의 촉각적 출력 생성기 센서는 디바이스(200)의 전면 상에 위치한 터치 스크린 디스플레이(212)의 반대편인 디바이스(200)의 배면 상에 위치된다.

[0073] 디바이스(200)는 또한 하나 이상의 가속도계들(268)을 포함한다. 도 2a는 주변기기 인터페이스(218)에 결합된 가속도계(268)를 도시한다. 대안적으로, 가속도계(268)는 I/O 서브시스템(206) 내의 입력 제어기(260)에 결합된다. 가속도계(268)는 예를 들어, 미국 특허 공개 제20050190059호, "Acceleration-based Theft Detection System for Portable Electronic Devices", 및 미국 특허 공개 제20060017692호, "Methods And Apparatuses For Operating A Portable Device Based On An Accelerometer"에 기술된 바와 같이 수행하며, 이들 둘 모두는 그 전문이 본 명세서에 참조로서 편입된다. 일부 실시예들에서, 하나 이상의 가속도계들로부터 수신된 데이터의 분석에 기초하여 터치 스크린 디스플레이 상에 세로보기(portrait view) 또는 가로보기(landscape view)로 정보가 디스플레이된다. 디바이스(200)는, 옵션적으로, 가속도계(들)(268) 외에도, 자력계(도시되지 않음), 및 디바이스(200)의 위치 및 배향(예컨대, 세로 또는 가로)에 관한 정보를 획득하기 위한 GPS(또는 GLONASS 또는 다른 글로벌 내비게이션 시스템) 수신기(도시되지 않음)를 포함한다.

[0074] 일부 실시예들에서, 메모리(202)에 저장된 소프트웨어 컴포넌트들은 운영 체제(226), 통신 모듈(또는 명령어들의 세트)(228), 접촉/모션 모듈(또는 명령어들의 세트)(230), 그래픽 모듈(또는 명령어들의 세트)(232), 텍스트 입력 모듈(또는 명령어들의 세트)(234), GPS 모듈(또는 명령어들의 세트)(235), 디지털 어시스턴트 클라이언트 모듈(229), 및 애플리케이션들(또는 명령어들의 세트들)(236)을 포함한다. 또한, 메모리(202)는 데이터 및 모델들, 예컨대 사용자 데이터 및 모델들(231)을 저장한다. 게다가, 일부 실시예들에서, 메모리(도 2a의 202 또는 도 4의 470)는 도 2a 및 도 4에 도시된 바와 같이 디바이스/글로벌 내부 상태(257)를 저장한다. 디바이스/글로벌 내부 상태(257)는, 존재하는 경우, 어떤 애플리케이션들이 현재 활성인지를 나타내는 활성 애플리케이션 상태; 어떤 애플리케이션들, 뷰들 또는 다른 정보가 터치 스크린 디스플레이(212)의 다양한 영역들을 점유하는지를 나타내는 디스플레이 상태; 디바이스의 다양한 센서들 및 입력 제어 디바이스(216)들로부터 획득된 정보를 포함하는 센서 상태; 및 디바이스의 위치 및/또는 자세에 관한 위치 정보 중 하나 이상을 포함한다.

[0075] 운영 체제(226)(예컨대, Darwin, RTXC, LINUX, UNIX, OS X, iOS, WINDOWS, 또는 VxWorks와 같은 임베디드 운영 체제)는 일반적인 시스템 태스크들(예컨대, 메모리 관리, 저장 디바이스 제어, 전력 관리 등)을 제어 및 관리하기 위한 다양한 소프트웨어 컴포넌트들 및/또는 드라이버들을 포함하고, 다양한 하드웨어와 소프트웨어 컴포넌트들 사이의 통신을 용이하게 한다.

[0076] 통신 모듈(228)은 하나 이상의 외부 포트들(224)을 통한 다른 디바이스들과의 통신을 용이하게 하고, 또한 RF 회로부(208) 및/또는 외부 포트(224)에 의해 수신되는 데이터를 처리하기 위한 다양한 소프트웨어 컴포넌트들을 포함한다. 외부 포트(224)(예컨대, USB(Universal Serial Bus), 파이어와이어(FIREWIRE) 등)는 다른 디바이스들에 직접적으로 또는 네트워크(예컨대, 인터넷, 무선 LAN 등)를 통해 간접적으로 결합하도록 구성된다. 일부 실시예들에서, 외부 포트는 아이팟®(애플 인크.의 상표) 디바이스들에서 사용되는 30-핀 커넥터와 동일하거나

유사하고/하거나 이와 호환가능한 멀티-핀(예컨대, 30-핀) 커넥터이다.

- [0077] 접촉/모션 모듈(230)은, 옵션적으로, (디스플레이 제어기(256)와 함께) 터치 스크린(212), 및 다른 터치 감응형 디바이스들(예컨대, 터치패드 또는 물리적 클릭 휠)과의 접촉을 검출한다. 접촉/모션 모듈(230)은 접촉이 발생했는지를 결정하는 것(예컨대, 손가락-다운 이벤트(finger-down event)를 검출하는 것), 접촉의 세기(예컨대, 접촉의 힘 또는 압력, 또는 접촉의 힘 또는 압력에 대한 대체물)를 결정하는 것, 접촉의 이동이 있는지를 결정하고 터치 감응형 표면을 가로지르는 이동을 추적하는 것(예컨대, 하나 이상의 손가락-드래그 이벤트(finger-dragging event)들을 검출하는 것), 및 접촉이 중단되었는지를 결정하는 것(예컨대, 손가락-업 이벤트(finger-up event) 또는 접촉 중단을 검출하는 것)과 같은, 접촉의 검출에 관련된 다양한 동작들을 수행하기 위한 다양한 소프트웨어 컴포넌트들을 포함한다. 접촉/모션 모듈(230)은 터치 감응형 표면으로부터 접촉 데이터를 수신한다. 일련의 접촉 데이터에 의해 표현되는 접촉 지점의 이동을 결정하는 것은, 옵션적으로, 접촉 지점의 속력(크기), 속도(크기 및 방향), 및/또는 가속도(크기 및/또는 방향의 변화)를 결정하는 것을 포함한다. 이 동작들은, 옵션적으로, 단일 접촉들(예컨대, 한 손가락 접촉들)에 또는 다수의 동시 접촉들(예컨대, "멀티터치"/다수의 손가락 접촉들)에 적용된다. 일부 실시예들에서, 접촉/모션 모듈(230) 및 디스플레이 제어기(256)는 터치패드 상의 접촉을 검출한다.
- [0078] 일부 실시예들에서, 접촉/모션 모듈(230)은 동작이 사용자에게 의해 수행되었는지 여부를 결정하는 데(예컨대, 사용자가 아이콘에 대해 "클릭"했는지 여부를 결정하는 데) 하나 이상의 세기 임계치들의 세트를 이용한다. 일부 실시예들에서, 적어도 세기 임계치들의 서브세트가 소프트웨어 파라미터들에 따라 결정된다(예컨대, 세기 임계치들은 특정 물리적 액추에이터들의 활성화 임계치들에 의해 결정되지 않으며, 디바이스(200)의 물리적 하드웨어를 변경함이 없이 조정될 수 있다). 예를 들어, 트랙패드 또는 터치 스크린 디스플레이의 마우스 "클릭" 임계치는 트랙패드 또는 터치 스크린 디스플레이 하드웨어를 변경함이 없이 넓은 범위의 미리정의된 임계 값들 중 임의의 것으로 설정될 수 있다. 추가적으로, 일부 구현예들에서, 디바이스의 사용자는 (예컨대, 개별 세기 임계치들을 조정함으로써 그리고/또는 복수의 세기 임계치들을 시스템 레벨 클릭 "세기" 파라미터로 한꺼번에 조정함으로써) 소정 세트의 세기 임계치들 중 하나 이상을 조정하기 위한 소프트웨어 설정들을 제공한다.
- [0079] 접촉/모션 모듈(230)은, 옵션적으로, 사용자에게 의한 제스처 입력을 검출한다. 터치 감응형 표면 상에서의 상이한 제스처들은 상이한 접촉 패턴들(예컨대, 상이한 모션들, 타이밍들, 및/또는 검출된 접촉들의 세기들)을 갖는다. 따라서, 제스처는 옵션적으로 특정 접촉 패턴을 검출함으로써 검출된다. 예를 들어, 손가락 탭 제스처(finger tap gesture)를 검출하는 것은 손가락-다운 이벤트를 검출한 다음에 손가락-다운 이벤트와 동일한 위치(또는 실질적으로 동일한 위치)에서(예컨대, 아이콘의 위치에서) 손가락-업(리프트오프(lift-off)) 이벤트를 검출하는 것을 포함한다. 다른 예로서, 터치 감응형 표면 상에서 손가락 스와이프 제스처(finger swipe gesture)를 검출하는 것은 손가락-다운 이벤트를 검출한 다음에 하나 이상의 손가락-드래깅 이벤트들을 검출하고, 그에 후속하여 손가락-업(리프트오프) 이벤트를 검출하는 것을 포함한다.
- [0080] 그래픽 모듈(232)은, 디스플레이되는 그래픽의 시각적 효과(예컨대, 밝기, 투명도, 채도, 콘트라스트 또는 다른 시각적 속성)를 변경하기 위한 컴포넌트들을 포함하는, 터치 스크린(212) 또는 다른 디스플레이 상에서 그래픽을 렌더링 및 디스플레이하기 위한 다양한 공지된 소프트웨어 컴포넌트들을 포함한다. 본 명세서에 사용되는 바와 같이, 용어 "그래픽"은 텍스트, 웹 페이지들, 아이콘들(예를 들어, 소프트 키들을 포함한 사용자 인터페이스 객체들), 디지털 이미지들, 비디오들, 애니메이션들 등을 제한 없이 포함하는, 사용자에게 디스플레이될 수 있는 임의의 객체를 포함한다.
- [0081] 일부 실시예들에서, 그래픽 모듈(232)은 사용될 그래픽을 표현하는 데이터를 저장한다. 각각의 그래픽에는 옵션적으로 대응하는 코드가 할당된다. 그래픽 모듈(232)은, 필요한 경우 좌표 데이터 및 다른 그래픽 속성 데이터와 함께, 디스플레이될 그래픽을 특정하는 하나 이상의 코드들을 애플리케이션들 등으로부터 수신하며, 이어서 스크린 이미지 데이터를 생성하여 디스플레이 제어기(256)에 출력한다.
- [0082] 햅틱 피드백 모듈(233)은 디바이스(200)와의 사용자 상호작용들에 응답하여 디바이스(200) 상의 하나 이상의 위치들에서 촉각적 출력들을 생성하기 위하여 촉각적 출력 생성기(들)(267)에 의해 이용되는 명령어들을 생성하기 위한 다양한 소프트웨어 컴포넌트들을 포함한다.
- [0083] 일부 예들에서, 그래픽 모듈(232)의 컴포넌트인 텍스트 입력 모듈(234)은 다양한 애플리케이션들(예컨대, 연락처(237), 이메일(240), IM(241), 브라우저(247), 및 텍스트 입력을 필요로 하는 임의의 다른 애플리케이션)에서 텍스트를 입력하기 위한 소프트 키보드들을 제공한다.

- [0084] GPS 모듈(235)은 디바이스의 위치를 결정하고, 이 정보를 다양한 애플리케이션들에서의 사용을 위해 (예컨대, 위치 기반 다이얼링에서 사용하기 위해 전화(238)에; 사진/비디오 메타데이터로서 카메라(243)에; 그리고 날씨 위젯들, 지역 옐로 페이지 위젯들 및 지도/내비게이션 위젯들과 같은 위치 기반 서비스들을 제공하는 애플리케이션들에) 제공한다.
- [0085] 디지털 어시스턴트 클라이언트 모듈(229)은 디지털 어시스턴트의 클라이언트 측 기능들을 제공하기 위한 다양한 클라이언트 측 디지털 어시스턴트 명령어들을 포함한다. 예를 들어, 디지털 어시스턴트 클라이언트 모듈(229)은 휴대용 다기능 디바이스(200)의 다양한 사용자 인터페이스들(예컨대, 마이크로폰(213), 가속도계(들)(268), 터치 감응형 디스플레이 시스템(212), 광 센서(들)(229), 다른 입력 제어 디바이스들(216) 등)을 통하여 음성 입력(예컨대, 스피치 입력), 텍스트 입력, 터치 입력, 및/또는 제스처 입력을 수용할 수 있다. 디지털 어시스턴트 클라이언트 모듈(229)은 또한, 휴대용 다기능 디바이스(200)의 다양한 출력 인터페이스들(예컨대, 스피커(211), 터치 감응형 디스플레이 시스템(212), 촉각적 출력 생성기(들)(267) 등)을 통하여 청각적(예컨대, 스피치 출력), 시각적, 및/또는 촉각적 형태들의 출력을 제공할 수 있다. 예를 들어, 출력은 음성, 사운드, 경고, 텍스트 메시지, 메뉴, 그래픽, 비디오, 애니메이션, 진동, 및/또는 이들 중 둘 이상의 것들의 조합들로서 제공된다. 동작 동안, 디지털 어시스턴트 클라이언트 모듈(229)은 RF 회로부(208)를 사용하여 DA 서버(106)와 통신한다.
- [0086] 사용자 데이터 및 모델들(231)은, 디지털 어시스턴트의 클라이언트 측 기능들을 제공하기 위해 사용자와 연관된 다양한 데이터(예컨대, 사용자-특정 어휘 데이터, 사용자 선호도 데이터, 사용자-특정 이름 발음들, 사용자의 전자 주소록으로부터의 데이터, 할 일 목록들, 쇼핑 목록들 등)를 포함한다. 또한, 사용자 데이터 및 모델들(231)은 사용자 입력을 프로세싱하고 사용자 의도를 결정하기 위해 다양한 모델들(예컨대, 스피치 인식 모델들, 통계 언어 모델들, 자연 언어 프로세싱 모델들, 온톨로지, 태스크 흐름 모델들, 서비스 모델들 등)을 포함한다.
- [0087] 일부 예들에서, 디지털 어시스턴트 클라이언트 모듈(229)은 사용자와 연관된 컨텍스트, 현재 사용자 상호작용, 및/또는 현재 사용자 입력을 확립하기 위해 휴대용 다기능 디바이스(200)의 다양한 센서들, 서브시스템들, 및 주변기기 디바이스들을 이용하여 휴대용 다기능 디바이스(200)의 주위 환경으로부터 추가 정보를 수집한다. 일부 예들에서, 디지털 어시스턴트 클라이언트 모듈(229)은 DA 서버(106)에 사용자 입력과 함께 컨텍스트 정보 또는 그의 서브셋을 제공하여 사용자의 의도를 추론하는 것을 돕는다. 일부 예들에서, 디지털 어시스턴트는 또한 컨텍스트 정보를 이용하여 어떻게 출력들을 준비하여 사용자에게 전달할지를 결정한다. 컨텍스트 정보는 컨텍스트 데이터라고 지칭된다.
- [0088] 일부 예들에서, 사용자 입력을 동반하는 컨텍스트 정보는 센서 정보, 예를 들어, 광(lighting), 주변 소음, 주변 온도, 주위 환경의 이미지 또는 비디오 등을 포함한다. 일부 예들에서, 컨텍스트 정보는 또한 디바이스의 물리적 상태, 예를 들어, 디바이스 배향, 디바이스 위치, 디바이스 온도, 전력 레벨, 속도, 가속도, 모션 패턴, 셀룰러 신호 강도 등을 포함할 수 있다. 일부 예들에서, DA 서버(106)의 소프트웨어 상태에 관련된 정보, 예를 들어, 휴대용 다기능 디바이스(200)의 실행 중인 프로세스, 설치된 프로그램, 과거 및 현재 네트워크 활동성, 백그라운드 서비스, 에러 로그, 리소스 사용 등이 사용자 입력과 연관된 컨텍스트 정보로서 DA 서버(106)에 제공된다.
- [0089] 일부 예들에서, 디지털 어시스턴트 클라이언트 모듈(229)은 DA 서버(106)로부터의 요청들에 응답하여 휴대용 다기능 디바이스(200) 상에 저장된 정보(예컨대, 사용자 데이터(231))를 선택적으로 제공한다. 일부 예들에서, 디지털 어시스턴트 클라이언트 모듈(229)은 또한 DA 서버(106)에 의한 요청에 따라 자연 언어 대화 또는 다른 사용자 인터페이스들을 통해 사용자로부터 추가 입력을 이끌어낸다. 디지털 어시스턴트 클라이언트 모듈(229)은 DA 서버(106)에 추가 입력을 전달하여, 사용자 요청에서 표현되는 사용자의 의도의 의도 추론 및/또는 이행 시에 DA 서버(106)를 돕는다.
- [0090] 디지털 어시스턴트의 보다 상세한 설명은 도 7a 내지 도 7c를 참조하여 하기에 기술된다. 디지털 어시스턴트 클라이언트 모듈(229)이 하기에 기술되는 디지털 어시스턴트 모듈(726)의 임의의 개수의 하위모듈들을 포함할 수 있다는 것이 인식되어야 한다.
- [0091] 애플리케이션들(236)은, 다음의 모듈들(또는 명령어들의 세트들), 또는 이들의 서브셋 또는 수퍼셋(superset)를 포함한다:
- [0092] • 연락처 모듈(237)(때때로 주소록 또는 연락처 목록으로 지칭됨);

- [0093] • 전화 모듈(238);
- [0094] • 화상 회의 모듈(239);
- [0095] • 이메일 클라이언트 모듈(240);
- [0096] • 인스턴트 메시징(IM) 모듈(241);
- [0097] • 운동 지원 모듈(242);
- [0098] • 정지 및/또는 비디오 이미지들을 위한 카메라 모듈(243);
- [0099] • 이미지 관리 모듈(244);
- [0100] • 비디오 재생기 모듈;
- [0101] • 음악 재생기 모듈;
- [0102] • 브라우저 모듈(247);
- [0103] • 캘린더 모듈(248);
- [0104] • 일부 예들에서, 날씨 위젯(249-1), 주식 위젯(249-2), 계산기 위젯(249-3), 알람 시계 위젯(249-4), 사전 위젯(249-5), 및 사용자에게 의해 획득되는 다른 위젯들뿐 아니라 사용자-생성 위젯들(249-6) 중 하나 이상을 포함하는 위젯 모듈들(249);
- [0105] • 사용자-생성 위젯들(249-6)을 만들기 위한 위젯 생성기 모듈(250);
- [0106] • 검색 모듈(251);
- [0107] • 비디오 재생기 모듈 및 음악 재생기 모듈을 통합하는 비디오 및 음악 재생기 모듈(252);
- [0108] • 메모 모듈(253);
- [0109] • 지도 모듈(254); 및/또는
- [0110] • 온라인 비디오 모듈(255).
- [0111] 메모리(202)에 저장된 다른 애플리케이션들(236)의 예들은 다른 워드 프로세싱 애플리케이션들, 다른 이미지 편집 애플리케이션들, 드로잉 애플리케이션들, 프레젠테이션 애플리케이션들, JAVA 인에이블형 애플리케이션들, 암호화, 디지털 권한 관리, 음성 인식 및 음성 복제를 포함한다.
- [0112] 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232) 및 텍스트 입력 모듈(234)과 함께, 연락처 모듈(237)은, 주소록 또는 연락처 목록(예컨대, 메모리(202) 또는 메모리(470) 내의 연락처 모듈(237)의 애플리케이션 내부 상태(292)에 저장됨)을 관리하는 데 사용되며: 메모리는 이름(들)을 주소록에 추가하는 것; 주소록으로부터 이름(들)을 삭제하는 것; 전화번호(들), 이메일 주소(들), 물리적 주소(들) 또는 다른 정보를 이름과 연관시키는 것; 이미지를 이름과 연관시키는 것; 이름들을 분류 및 정렬하는 것; 전화(238), 화상 회의 모듈(239), 이메일(240) 또는 IM(241)에 의한 통신을 개시하고/하거나 용이하게 하기 위해 전화번호들 또는 이메일 주소들을 제공하는 것 등을 포함한다.
- [0113] RF 회로부(208), 오디오 회로부(210), 스피커(211), 마이크로폰(213), 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232), 및 텍스트 입력 모듈(234)과 함께, 전화 모듈(238)은, 전화번호에 대응하는 문자들의 시퀀스를 입력하고, 연락처 모듈(237) 내의 하나 이상의 전화번호들에 액세스하며, 입력된 전화번호를 수정하고, 개개의 전화번호를 다이얼링하며, 대화를 하고, 대화가 완료된 때 접속해제하거나 끊는 데 사용된다. 전송된 바와 같이, 무선 통신은 복수의 통신 표준들, 프로토콜들 및 기술들 중 임의의 것을 사용한다.

- [0114] RF 회로부(208), 오디오 회로부(210), 스피커(211), 마이크로폰(213), 터치 스크린(212), 디스플레이 제어기(256), 광 센서(264), 광 센서 제어기(258), 접촉/모션 모듈(230), 그래픽 모듈(232), 텍스트 입력 모듈(234), 연락처 모듈(237) 및 전화 모듈(238)과 함께, 화상 회의 모듈(239)은 사용자 지시들에 따라 사용자와 한 명 이상의 다른 참여자들 사이의 화상 회의를 개시, 시행 및 종료하도록 하는 실행가능한 명령어들을 포함한다.
- [0115] RF 회로부(208), 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232) 및 텍스트 입력 모듈(234)과 함께, 이메일 클라이언트 모듈(240)은 사용자 지시들에 응답하여 이메일을 작성, 전송, 수신, 및 관리하도록 하는 실행가능한 명령어들을 포함한다. 이미지 관리 모듈(244)과 함께, 이메일 클라이언트 모듈(240)은 카메라 모듈(243)로 촬영된 정지 또는 비디오 이미지들을 갖는 이메일들을 생성 및 전송하는 것을 매우 용이하게 한다.
- [0116] RF 회로부(208), 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232) 및 텍스트 입력 모듈(234)과 함께, 인스턴트 메시징 모듈(241)은, 인스턴트 메시지에 대응하는 문자들의 시퀀스를 입력하고, 이전에 입력된 문자들을 수정하고, (예를 들어, 전화 기반 인스턴트 메시지들을 위한 단문자 메시지 서비스(Short Message Service; SMS) 또는 멀티미디어 메시지 서비스(Multimedia Message Service; MMS) 프로토콜을 이용하거나, 인터넷 기반 인스턴트 메시지들을 위한 XMPP, SIMPLE 또는 IMPS를 이용하여) 개개의 인스턴트 메시지를 송신하고, 인스턴트 메시지들을 수신하고, 수신된 인스턴트 메시지들을 보도록 하는 실행가능한 명령어들을 포함한다. 일부 실시예들에서, 전송 및/또는 수신되는 인스턴트 메시지들은 그래픽, 사진들, 오디오 파일들, 비디오 파일들, 그리고/또는 MMS 및/또는 EMS(Enhanced Messaging Service)에서 지원되는 바와 같은 다른 첨부물들을 포함한다. 본 명세서에서 사용되는 바와 같이, "인스턴트 메시징"은 전화 기반 메시지들(예컨대, SMS 또는 MMS를 이용하여 전송되는 메시지들) 및 인터넷 기반 메시지들(예컨대, XMPP, SIMPLE 또는 IMPS를 이용하여 전송되는 메시지들) 둘 모두를 지칭한다.
- [0117] RF 회로부(208), 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232), 텍스트 입력 모듈(234), GPS 모듈(235), 지도 모듈(254), 및 음악 재생기 모듈과 함께, 운동 지원 모듈(242)은, (예컨대, 시간, 거리, 및/또는 열량 소비 목표와 함께) 운동들을 고안하고; 운동 센서들(스포츠 디바이스들)과 통신하고; 운동 센서 데이터를 수신하고; 운동을 모니터링하는 데 사용되는 센서들을 교정하고; 운동을 위한 음악을 선택 및 재생하고; 운동 데이터를 디스플레이, 저장 및 송신하도록 하는 실행가능한 명령어들을 포함한다.
- [0118] 터치 스크린(212), 디스플레이 제어기(256), 광 센서(들)(264), 광 센서 제어기(258), 접촉/모션 모듈(230), 그래픽 모듈(232) 및 이미지 관리 모듈(244)과 함께, 카메라 모듈(243)은, 정지 이미지들 또는 비디오(비디오 스트림을 포함함)를 캡처하고 이들을 메모리(202) 내에 저장하거나, 정지 이미지 또는 비디오의 특성을 수정하거나, 메모리(202)로부터 정지 이미지 또는 비디오를 삭제하도록 하는 실행가능한 명령어들을 포함한다.
- [0119] 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232), 텍스트 입력 모듈(234) 및 카메라 모듈(243)과 함께, 이미지 관리 모듈(244)은 정지 및/또는 비디오 이미지들을 배열하거나, 수정(예컨대, 편집)하거나, 또는 그렇지 않으면 조작하고, 라벨링하고, 삭제하고, (예컨대, 디지털 슬라이드 쇼 또는 앨범에) 제시하고, 저장하도록 하는 실행가능한 명령어들을 포함한다.
- [0120] RF 회로부(208), 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232) 및 텍스트 입력 모듈(234)과 함께, 브라우저 모듈(247)은, 웹 페이지들 또는 이들의 부분들뿐만 아니라 웹 페이지들에 링크된 첨부물들 및 다른 파일들을 검색하고, 그들에 링크하고, 수신하고, 그리고 디스플레이하는 것을 비롯한, 사용자 지시들에 따라 인터넷을 브라우징하도록 하는 실행가능한 명령어들을 포함한다.
- [0121] RF 회로부(208), 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232), 텍스트 입력 모듈(234), 이메일 클라이언트 모듈(240), 및 브라우저 모듈(247)과 함께, 캘린더 모듈(248)은 사용자 지시들에 따라 캘린더들 및 캘린더들과 연관된 데이터(예컨대, 캘린더 엔트리들, 할 일 목록들 등)를 생성, 디스플레이, 수정, 및 저장하도록 하는 실행가능한 명령어들을 포함한다.
- [0122] RF 회로부(208), 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232), 텍스트 입력 모듈(234) 및 브라우저 모듈(247)과 함께, 위젯 모듈들(249)은 사용자에게 의해 다운로드 및 사용될 수 있거나(예컨대, 날씨 위젯(249-1), 주식 위젯(249-2), 계산기 위젯(249-3), 알람 시계 위젯(249-4) 및 사진 위젯(249-5)), 또는 사용자에게 의해 생성될 수 있는(예컨대, 사용자-생성 위젯(249-6)) 미니-애플리케이션(mini-application)들이다. 일부 실시예들에서, 위젯은 HTML(Hypertext Markup Language) 파일, CSS(Cascading Style Sheets) 파일 및 자바스크립트(JavaScript) 파일을 포함한다. 일부 실시예들에서, 위젯은

XML(Extensible Markup Language) 파일 및 자바스크립트 파일(예컨대, 야후(Yahoo!) 위젯들)을 포함한다.

- [0123] RF 회로부(208), 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232), 텍스트 입력 모듈(234) 및 브라우저 모듈(247)과 함께, 위젯 생성기 모듈(250)은 사용자에게 의해 위젯들을 생성(예를 들어, 웹 페이지의 사용자-특정 부분을 위젯으로 변경)하는 데 사용된다.
- [0124] 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232) 및 텍스트 입력 모듈(234)과 함께, 검색 모듈(251)은 사용자 지시들에 따라 하나 이상의 검색 기준들(예컨대, 하나 이상의 사용자-특정 검색어들)에 매칭되는 메모리(202) 내의 텍스트, 음악, 사운드, 이미지, 비디오, 및/또는 다른 파일들을 검색하도록 하는 실행가능한 명령어들을 포함한다.
- [0125] 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232), 오디오 회로부(210), 스피커(211), RF 회로부(208) 및 브라우저 모듈(247)과 함께, 비디오 및 음악 재생기 모듈(252)은, 사용자가 MP3 또는 AAC 파일들과 같은 하나 이상의 파일 포맷들로 저장된 녹음된 음악 및 다른 사운드 파일들을 다운로드 및 재생할 수 있도록 하는 실행가능한 명령어들, 및 비디오들을 (예컨대, 터치 스크린(212) 상에서 또는 외부 포트(224)를 통해 외부의 접속된 디스플레이 상에서) 디스플레이하도록, 상영하도록, 또는 다른 방식으로 재생하도록 하는 실행가능한 명령어들을 포함한다. 일부 실시예들에서, 디바이스(200)는 옵션적으로 아이팟(애플 인크.의 상표)과 같은 MP3 재생기의 기능을 포함한다.
- [0126] 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232) 및 텍스트 입력 모듈(234)과 함께, 메모 모듈(253)은 사용자 지시들에 따라 메모들, 할 일 목록들 등을 생성 및 관리하도록 하는 실행가능한 명령어들을 포함한다.
- [0127] RF 회로부(208), 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232), 텍스트 입력 모듈(234), GPS 모듈(235), 및 브라우저 모듈(247)과 함께, 지도 모듈(254)은 사용자 지시들에 따라 지도들 및 지도들과 연관된 데이터(예컨대, 운전 길 안내; 특정 위치에서의 또는 그 인근의 상점들 및 다른 관심 지점들에 관한 데이터; 및 다른 위치-기반 데이터)를 수신하고, 디스플레이하고, 수정하고, 저장하는 데 사용된다.
- [0128] 터치 스크린(212), 디스플레이 제어기(256), 접촉/모션 모듈(230), 그래픽 모듈(232), 오디오 회로부(210), 스피커(211), RF 회로부(208), 텍스트 입력 모듈(234), 이메일 클라이언트 모듈(240) 및 브라우저 모듈(247)과 함께, 온라인 비디오 모듈(255)은 사용자가 H.264와 같은 하나 이상의 파일 포맷들의 온라인 비디오들에 액세스하고, 그들을 브라우징하고, (예컨대, 스트리밍 및/또는 다운로드에 의해) 수신하고, (예컨대, 터치 스크린 상에서 또는 외부 포트(224)를 통해 외부의 접속된 디스플레이 상에서) 재생하고, 특정한 온라인 비디오로의 링크와 함께 이메일을 전송하고, 그렇지 않으면 관리하게 하는 명령어들을 포함한다. 일부 실시예들에서, 이메일 클라이언트 모듈(240)보다는 오히려 인스턴트 메시징 모듈(241)이 특정 온라인 비디오로의 링크를 전송하는 데 사용된다. 온라인 비디오 애플리케이션에 대한 추가적 설명은, 2007년 6월 20일자로 출원된 미국 특허 출원 제 60/936,562호, "Portable Multifunction Device, Method, and Graphical User Interface for Playing Online Videos" 및 2007년 12월 31일자로 출원된 미국 특허 출원 제 11/968,067호, "Portable Multifunction Device, Method, and Graphical User Interface for Playing Online Videos"에서 찾아볼 수 있으며, 이들의 내용은 이로써 그 전체가 참조로서 본 명세서에 편입된다.
- [0129] 앞서 식별된 모듈들 및 애플리케이션들 각각은 상술한 하나 이상의 기능들 및 본 출원에 기술되는 방법들(예컨대, 본 명세서에 기술되는 컴퓨터 구현 방법들 및 다른 정보 프로세싱 방법들)을 수행하기 위한 실행가능한 명령어들의 세트에 대응한다. 이들 모듈들(예컨대, 명령어들의 세트들)은 별개의 소프트웨어 프로그램들, 절차들 또는 모듈들로서 구현될 필요는 없으며, 따라서 이들 모듈들의 다양한 서브세트들이 다양한 실시예들에서 조합되거나 달리 재배열될 수 있다. 예를 들어, 비디오 재생기 모듈은 음악 재생기 모듈과 조합되어 단일 모듈(예컨대, 도 2a의 비디오 및 음악 재생기 모듈(252))로 될 수 있다. 일부 실시예들에서, 메모리(202)는 앞서 식별된 모듈들 및 데이터 구조들의 서브세트를 저장한다. 또한, 메모리(202)는 상술되지 않은 추가의 모듈들 및 데이터 구조들을 저장한다.
- [0130] 일부 실시예들에서, 디바이스(200)는 디바이스 상의 미리정의된 세트의 기능들의 동작이 터치 스크린 및/또는 터치패드를 통해 전용으로 수행되는 디바이스이다. 터치 스크린 및/또는 터치패드를 디바이스(200)의 동작을 위한 주 입력 제어 디바이스로서 사용함으로써 디바이스(200) 상의 (푸시 버튼들, 다이얼들 등과 같은) 물리적 입력 제어 디바이스들의 수가 감소된다.

- [0131] 터치 스크린 및/또는 터치패드를 통해 전용으로 수행되는 미리정의된 세트의 기능들은 옵션적으로 사용자 인터페이스들 사이의 내비게이션을 포함한다. 일부 실시예들에서, 터치패드는, 사용자에게 의해 터치될 때, 디바이스(200)를 디바이스(200) 상에 디스플레이되는 임의의 사용자 인터페이스로부터 메인, 홈 또는 루트 메뉴로 내비게이팅한다. 이러한 실시예들에서, "메뉴 버튼"이 터치패드를 이용하여 구현된다. 일부 다른 실시예들에서, 메뉴 버튼은 터치패드 대신에 물리적 푸시 버튼 또는 다른 물리적 입력 제어 디바이스이다.
- [0132] 도 2b는 일부 실시예들에 따른, 이벤트 처리를 위한 예시적인 컴포넌트들을 도시하는 블록도이다. 일부 실시예들에서, 메모리(도 2a의 202 또는 도 4의 470)는 (예컨대, 운영 체제(226)에서의) 이벤트 분류기(270) 및 개개의 애플리케이션(236-1)(예컨대, 전송된 애플리케이션들(237 내지 251, 255, 480 내지 490) 중 임의의 것)을 포함한다.
- [0133] 이벤트 분류기(270)는 이벤트 정보를 수신하고, 이벤트 정보를 전달할 애플리케이션(236-1), 및 애플리케이션(236-1)의 애플리케이션 뷰(291)를 결정한다. 이벤트 분류기(270)는 이벤트 모니터(271) 및 이벤트 디스패처 모듈(event dispatcher module)(274)을 포함한다. 일부 실시예들에서, 애플리케이션(236-1)은 애플리케이션이 활성화되거나 실행 중일 때 터치 감응형 디스플레이(212) 상에 디스플레이되는 현재 애플리케이션 뷰(들)를 나타내는 애플리케이션 내부 상태(292)를 포함한다. 일부 실시예들에서, 디바이스/글로벌 내부 상태(257)는 이벤트 분류기(270)에 의해 어느 애플리케이션(들)이 현재 활성화인지 결정하는 데 이용되며, 애플리케이션 내부 상태(292)는 이벤트 분류기(270)에 의해 이벤트 정보를 전달할 애플리케이션 뷰들(291)을 결정하는 데 이용된다.
- [0134] 일부 실시예들에서, 애플리케이션 내부 상태(292)는 애플리케이션(236-1)이 실행을 재개할 때 이용될 재개 정보, 애플리케이션(236-1)에 의해 디스플레이되고 있거나 디스플레이될 준비가 된 정보를 나타내는 사용자 인터페이스 상태 정보, 사용자가 애플리케이션(236-1)의 이전 상태 또는 뷰로 되돌아가는 것을 가능하게 하기 위한 상태 큐(queue), 및 사용자에게 의해 취해진 이전 행동들의 재실행(redo)/실행취소(undo) 큐 중 하나 이상과 같은 추가 정보를 포함한다.
- [0135] 이벤트 모니터(271)는 주변기기 인터페이스(218)로부터 이벤트 정보를 수신한다. 이벤트 정보는 서브이벤트(예를 들어, 다중 터치 제스처의 일부로서 터치 감응형 디스플레이(212) 상에서의 사용자 터치)에 대한 정보를 포함한다. 주변기기 인터페이스(218)는 I/O 서브시스템(206) 또는 센서, 예컨대 근접 센서(266), 가속도계(들)(268), 및/또는 (오디오 회로부(210)를 통한) 마이크로폰(213)으로부터 수신하는 정보를 송신한다. 주변기기 인터페이스(218)가 I/O 서브시스템(206)으로부터 수신하는 정보는 터치 감응형 디스플레이(212) 또는 터치 감응형 표면으로부터의 정보를 포함한다.
- [0136] 일부 실시예들에서, 이벤트 모니터(271)는 요청들을 미리결정된 간격으로 주변기기 인터페이스(218)에 전송한다. 이에 응답하여, 주변기기 인터페이스(218)는 이벤트 정보를 송신한다. 다른 실시예들에서, 주변기기 인터페이스(218)는 중요한 이벤트(예컨대, 미리결정된 잠음 임계치를 초과하는 입력 및/또는 미리결정된 지속기간 초과 동안의 입력을 수신하는 것)가 있을 때에만 이벤트 정보를 송신한다.
- [0137] 일부 실시예들에서, 이벤트 분류기(270)는 또한 히트 뷰(hit view) 결정 모듈(272) 및/또는 활성 이벤트 인식기 결정 모듈(273)을 포함한다.
- [0138] 히트 뷰 결정 모듈(272)은 터치 감응형 디스플레이(212)가 둘 이상의 뷰를 디스플레이하는 경우에, 하나 이상의 뷰들 내에서 서브이벤트가 발생한 곳을 결정하기 위한 소프트웨어 절차들을 제공한다. 뷰들은 사용자가 디스플레이 상에서 볼 수 있는 제어부들 및 다른 요소들로 구성된다.
- [0139] 애플리케이션과 연관된 사용자 인터페이스의 다른 양태는 본 명세서에서 때때로 애플리케이션 뷰들 또는 사용자 인터페이스 윈도우(user interface window)들로 지칭되는 한 세트의 뷰들이며, 여기서 정보가 디스플레이되고 터치 기반 제스처가 발생한다. 터치가 검출되는 (개개의 애플리케이션의) 애플리케이션 뷰들은 애플리케이션의 프로그램 또는 뷰 계층구조 내의 프로그램 레벨들에 대응한다. 예를 들면, 터치가 검출되는 최하위 레벨의 뷰는 히트 뷰라고 지칭되고, 적절한 입력들로서 인식되는 이벤트들의 세트는 터치 기반 제스처를 시작하는 초기 터치의 히트 뷰에 적어도 부분적으로 기초하여 결정된다.
- [0140] 히트 뷰 결정 모듈(272)은 터치 기반 제스처의 서브이벤트들에 관련된 정보를 수신한다. 애플리케이션이 계층 구조에서 조직화된 다수의 뷰들을 갖는 경우, 히트 뷰 결정 모듈(272)은 히트 뷰를, 서브이벤트를 처리해야 하는 계층구조 내의 최하위 뷰로서 식별한다. 대부분의 상황들에서, 히트 뷰는 개시되는 서브이벤트(예컨대, 이벤트 또는 잠재적 이벤트를 형성하는 서브이벤트들의 시퀀스에서의 제1 서브이벤트)가 발생하는 최하위 레벨 뷰이다. 일단 히트 뷰가 히트 뷰 결정 모듈(272)에 의해 식별되면, 히트 뷰는 전형적으로 그것이 히트 뷰로서 식

별되게 한 것과 동일한 터치 또는 입력 소스에 관련된 모든 서브이벤트들을 수신한다.

- [0141] 활성 이벤트 인식기 결정 모듈(273)은 뷰 계층구조 내에서 어느 뷰 또는 뷰들이 서브이벤트들의 특정 시퀀스를 수신해야 하는지를 결정한다. 일부 실시예들에서, 활성 이벤트 인식기 결정 모듈(273)은 히트 뷰만이 서브이벤트들의 특정 시퀀스를 수신해야 하는 것으로 결정한다. 다른 실시예들에서, 활성 이벤트 인식기 결정 모듈(273)은 서브이벤트의 물리적 위치를 포함하는 모든 뷰들이 적극 참여 뷰(actively involved view)들이므로 결정하고, 그에 따라 모든 적극 참여 뷰들이 서브이벤트들의 특정 시퀀스를 수신해야 하는 것으로 결정한다. 다른 실시예들에서, 터치 서브이벤트들이 전적으로 하나의 특정 뷰와 연관된 영역으로 한정되었다더라도, 계층구조 내의 상위 뷰들은 여전히 적극 참여 뷰들로서 유지될 것이다.
- [0142] 이벤트 디스패처 모듈(274)은 이벤트 정보를 이벤트 인식기(예컨대, 이벤트 인식기(280))에 디스패치한다. 활성 이벤트 인식기 결정 모듈(273)을 포함하는 실시예들에서, 이벤트 디스패처 모듈(274)은 이벤트 정보를 활성 이벤트 인식기 결정 모듈(273)에 의해 결정된 이벤트 인식기에 전달한다. 일부 실시예들에서, 이벤트 디스패처 모듈(274)은 이벤트 큐 내에 이벤트 정보를 저장하는데, 이벤트 정보는 개개의 이벤트 수신기(282)에 의해 인출된다.
- [0143] 일부 실시예들에서, 운영 체제(226)는 이벤트 분류기(270)를 포함한다. 대안적으로, 애플리케이션(236-1)은 이벤트 분류기(270)를 포함한다. 또 다른 실시예들에서, 이벤트 분류기(270)는 독립형 모듈이거나, 또는 접촉/모션 모듈(230)과 같이 메모리(202)에 저장되는 다른 모듈의 일부이다.
- [0144] 일부 실시예들에서, 애플리케이션(236-1)은 복수의 이벤트 핸들러들(290) 및 하나 이상의 애플리케이션 뷰들(291)을 포함하며, 이들의 각각은 애플리케이션의 사용자 인터페이스의 개개의 뷰 내에 발생하는 터치 이벤트들을 처리하기 위한 명령어들을 포함한다. 애플리케이션(236-1)의 각각의 애플리케이션 뷰(291)는 하나 이상의 이벤트 인식기들(280)을 포함한다. 전형적으로, 개개의 애플리케이션 뷰(291)는 복수의 이벤트 인식기들(280)을 포함한다. 다른 실시예들에서, 이벤트 인식기들(280) 중 하나 이상은 사용자 인터페이스 키트(도시되지 않음) 또는 애플리케이션(236-1)이 방법들 및 다른 속성들을 물려받는 상위 레벨 객체와 같은 별개의 모듈의 일부이다. 일부 실시예들에서, 개개의 이벤트 핸들러(290)는 데이터 업데이터(276), 객체 업데이터(277), GUI 업데이터(278), 및/또는 이벤트 분류기(270)로부터 수신된 이벤트 데이터(279) 중 하나 이상을 포함한다. 이벤트 핸들러(290)는 데이터 업데이터(276), 객체 업데이터(277) 또는 GUI 업데이터(278)를 이용하거나 호출하여 애플리케이션 내부 상태(292)를 업데이트한다. 대안적으로, 애플리케이션 뷰들(291) 중 하나 이상은 하나 이상의 개개의 이벤트 핸들러들(290)을 포함한다. 또한, 일부 실시예들에서, 데이터 업데이터(276), 객체 업데이터(277), 및 GUI 업데이터(278) 중 하나 이상은 개개의 애플리케이션 뷰(291) 내에 포함된다.
- [0145] 개개의 이벤트 인식기(280)는 이벤트 분류기(270)로부터 이벤트 정보(예컨대, 이벤트 데이터(279))를 수신하고 그 이벤트 정보로부터 이벤트를 식별한다. 이벤트 인식기(280)는 이벤트 수신기(282) 및 이벤트 비교기(284)를 포함한다. 일부 실시예들에서, 이벤트 인식기(280)는 또한 적어도 메타데이터(283) 및 이벤트 전달 명령어들(288)(서브이벤트 전달 명령어들을 포함할 수 있음)의 서브셋을 포함할 수 있다.
- [0146] 이벤트 수신기(282)는 이벤트 분류기(270)로부터 이벤트 정보를 수신한다. 이벤트 정보는 서브이벤트, 예를 들어 터치 또는 터치 이동에 관한 정보를 포함한다. 서브이벤트에 따라, 이벤트 정보는 또한 서브이벤트의 위치와 같은 추가 정보를 포함한다. 서브이벤트가 터치의 모션에 관한 것일 때, 이벤트 정보는 또한 서브이벤트의 속력 및 방향을 포함한다. 일부 실시예들에서, 이벤트들은 하나의 배향으로부터 다른 배향으로(예컨대, 세로 배향으로부터 가로 배향으로, 또는 그 반대로)의 디바이스의 회전을 포함하며, 이벤트 정보는 디바이스의 현재 배향(디바이스 자세로도 지칭됨)에 관한 대응하는 정보를 포함한다.
- [0147] 이벤트 비교기(284)는 이벤트 정보를 미리정의된 이벤트 또는 서브이벤트 정의들과 비교하고, 그 비교에 기초하여, 이벤트 또는 서브이벤트를 결정하거나, 이벤트 또는 서브이벤트의 상태를 결정 또는 업데이트한다. 일부 실시예들에서, 이벤트 비교기(284)는 이벤트 정의들(286)을 포함한다. 이벤트 정의들(286)은 이벤트들(예컨대, 서브이벤트들의 미리정의된 시퀀스들), 예를 들어 이벤트 1(287-1), 이벤트 2(287-2) 등의 정의들을 포함한다. 일부 실시예들에서, 이벤트(287) 내의 서브이벤트들은, 예를 들어, 터치 시작, 터치 종료, 터치 이동, 터치 취소, 및 다중 터치를 포함한다. 일례에서, 이벤트 1(287-1)에 대한 정의는 디스플레이된 객체 상의 더블 탭이다. 더블 탭은, 예를 들어, 미리결정된 페이즈(phase) 동안의 디스플레이된 객체 상의 제1 터치(터치 시작), 미리결정된 페이즈 동안의 제1 리프트오프(터치 종료), 미리결정된 페이즈 동안의 디스플레이된 객체 상의 제2 터치(터치 시작), 및 미리결정된 페이즈 동안의 제2 리프트오프(터치 종료)를 포함한다. 다른 예에서, 이벤트 2(287-2)에 대한 정의는 디스플레이된 객체 상의 드래깅이다. 드래깅은, 예를 들어, 미리결정된 페이즈 동안

의 디스플레이된 객체 상의 터치(또는 접촉), 터치 감응형 디스플레이(212)를 가로지르는 터치의 이동, 및 터치의 리프트오프(터치 종료)를 포함한다. 일부 실시예들에서, 이벤트는 또한 하나 이상의 연관된 이벤트 핸들러들(290)에 대한 정보를 포함한다.

- [0148] 일부 실시예들에서, 이벤트 정의(287)는 개개의 사용자 인터페이스 객체에 대한 이벤트의 정의를 포함한다. 일부 실시예들에서, 이벤트 비교기(284)는 어느 사용자 인터페이스 객체가 서브이벤트와 연관되는지 결정하도록 히트 테스트(hit test)를 수행한다. 예를 들어, 3개의 사용자 인터페이스 객체들이 터치 감응형 디스플레이(212) 상에 디스플레이된 애플리케이션 뷰에서, 터치 감응형 디스플레이(212) 상에서 터치가 검출되는 경우, 이벤트 비교기(284)는 3개의 사용자 인터페이스 객체 중 어느 것이 터치(서브이벤트)와 연관되는지를 결정하도록 히트 테스트를 수행한다. 각각의 디스플레이된 객체가 개개의 이벤트 핸들러(290)와 연관되는 경우, 이벤트 비교기는 어느 이벤트 핸들러(290)가 활성화되어야 하는지 결정하는 데 히트 테스트의 결과를 이용한다. 예를 들어, 이벤트 비교기(284)는 히트 테스트를 트리거하는 객체 및 서브이벤트와 연관된 이벤트 핸들러를 선택한다.
- [0149] 일부 실시예들에서, 개개의 이벤트(287)에 대한 정의는 또한 서브이벤트들의 시퀀스가 이벤트 인식기의 이벤트 유형에 대응하는지 대응하지 않는지 여부가 결정된 후까지 이벤트 정보의 전달을 지연하는 지연된 행동들을 포함한다.
- [0150] 개개의 이벤트 인식기(280)가 일련의 서브이벤트들이 이벤트 정의들(286) 내의 이벤트들 중 어떠한 것보다 매칭되지 않는 것으로 결정하면, 개개의 이벤트 인식기(280)는 이벤트 불가능, 이벤트 실패, 또는 이벤트 종료 상태에 진입하고, 그 후 개개의 이벤트 인식기는 터치 기반 제스처의 후속적인 서브이벤트들을 무시한다. 이러한 상황에서, 만일 있다면, 히트 뷰에 대해 활성 상태로 유지되는 다른 이벤트 인식기들이 진행 중인 터치 기반 제스처의 서브이벤트들을 계속해서 추적 및 프로세싱한다.
- [0151] 일부 실시예들에서, 개개의 이벤트 인식기(280)는 이벤트 전달 시스템이 어떻게 적극 참여 이벤트 인식기들에 대한 서브이벤트 전달을 수행해야 하는지를 나타내는 구성가능한 속성들, 플래그(flag)들, 및/또는 목록들을 갖는 메타데이터(283)를 포함한다. 일부 실시예들에서, 메타데이터(283)는 이벤트 인식기들이 어떻게 서로 상호작용하는지, 또는 상호작용할 수 있는지를 나타내는 구성가능한 속성들, 플래그들, 및/또는 목록들을 포함한다. 일부 실시예들에서, 메타데이터(283)는, 서브이벤트들이 뷰 또는 프로그램 계층구조에서의 다양한 레벨들에 전달되는지 여부를 나타내는 구성가능한 속성들, 플래그들, 및/또는 목록들을 포함한다.
- [0152] 일부 실시예들에서, 개개의 이벤트 인식기(280)는 이벤트의 하나 이상의 특정 서브이벤트들이 인식될 때 이벤트와 연관된 이벤트 핸들러(290)를 활성화시킨다. 일부 실시예들에서, 개개의 이벤트 인식기(280)는 이벤트와 연관된 이벤트 정보를 이벤트 핸들러(290)에 전달한다. 이벤트 핸들러(290)를 활성화시키는 것은 개개의 히트 뷰에 서브이벤트들을 전송(및 지연 전송)하는 것과는 별개이다. 일부 실시예들에서, 이벤트 인식기(280)는 인식된 이벤트와 연관된 플래그를 보내고, 그 플래그와 연관된 이벤트 핸들러(290)는 그 플래그를 캐치하고 미리 정의된 프로세스를 수행한다.
- [0153] 일부 실시예들에서, 이벤트 전달 명령어들(288)은 이벤트 핸들러를 활성화시키지 않으면서 서브이벤트에 관한 이벤트 정보를 전달하는 서브이벤트 전달 명령어들을 포함한다. 대신에, 서브이벤트 전달 명령어들은 일련의 서브이벤트들과 연관된 이벤트 핸들러들에 또는 적극 참여 뷰들에 이벤트 정보를 전달한다. 일련의 서브이벤트들 또는 적극 참여 뷰들과 연관된 이벤트 핸들러들은 이벤트 정보를 수신하고 미리결정된 프로세스를 수행한다.
- [0154] 일부 실시예들에서, 데이터 업데이터(276)는 애플리케이션(236-1)에서 이용되는 데이터를 생성 및 업데이트한다. 예를 들어, 데이터 업데이터(276)는 연락처 모듈(237)에서 이용되는 전화 번호를 업데이트하거나, 비디오 재생기 모듈에서 이용되는 비디오 파일을 저장한다. 일부 실시예들에서, 객체 업데이터(277)는 애플리케이션(236-1)에서 이용되는 객체들을 생성 및 업데이트한다. 예를 들어, 객체 업데이터(277)는 새로운 사용자 인터페이스 객체를 생성하거나, 또는 사용자 인터페이스 객체의 위치를 업데이트한다. GUI 업데이터(278)는 GUI를 업데이트한다. 예를 들어, GUI 업데이터(278)는 터치 감응형 디스플레이 상의 디스플레이를 위해 디스플레이 정보를 준비하고 이를 그래픽 모듈(232)에 전송한다.
- [0155] 일부 실시예들에서, 이벤트 핸들러(들)(290)는 데이터 업데이터(276), 객체 업데이터(277), 및 GUI 업데이터(278)를 포함하거나 이들에 액세스한다. 일부 실시예들에서, 데이터 업데이터(276), 객체 업데이터(277), 및 GUI 업데이터(278)는 개개의 애플리케이션(236-1) 또는 애플리케이션 뷰(291)의 단일 모듈 내에 포함된다. 다른 실시예들에서, 이들은 2개 이상의 소프트웨어 모듈 내에 포함된다.
- [0156] 터치 감응형 디스플레이들 상의 사용자 터치들의 이벤트 처리에 관하여 기술한 논의는 또한 입력 디바이스들을

갖는 다기능 디바이스들(200)을 동작시키기 위한 다른 형태들의 사용자 입력들에도 적용되지만, 그 모두가 터치 스크린들 상에서 개시되는 것이 아니라는 것을 이해해야 한다. 예를 들어, 단일 또는 다수의 키보드 누르기 또는 유지(hold)와 옵션적으로 조화된 마우스 이동 및 마우스 버튼 누르기; 터치패드 상에서의, 탭, 드래그, 스크롤 등과 같은 접촉 이동들; 펜 스타일러스 입력들; 디바이스의 이동; 구두 명령어들; 검출된 눈 이동들; 생체 측정 입력들; 및/또는 이들의 임의의 조합은, 인식될 이벤트를 정의하는 서브이벤트들에 대응하는 입력들로서 옵션적으로 이용된다.






[0157] 도 3은 일부 실시예들에 따른, 터치 스크린(212)을 갖는 휴대용 다기능 디바이스(200)를 도시한다. 터치 스크린은, 옵션적으로, 사용자 인터페이스(UI)(300) 내에서 하나 이상의 그래픽들을 디스플레이한다. 이러한 실시예에는 물론 하기에 기술되는 다른 실시예들에서, 사용자는, 예를 들어, 하나 이상의 손가락들(302)(도면에서 축척대로 도시되지 않음) 또는 하나 이상의 스타일러스들(303)(도면에서 축척대로 도시되지 않음)을 이용하여 그래픽 상에 제스처를 행함으로써 그래픽들 중 하나 이상을 선택하는 것이 가능하게 된다. 일부 실시예들에서, 하나 이상의 그래픽들의 선택은 사용자가 하나 이상의 그래픽들과의 접촉을 중단할 때 발생한다. 일부 실시예들에서, 제스처는 옵션적으로 디바이스(200)와 접촉한 손가락의 하나 이상의 탭들, (좌에서 우로, 우에서 좌로, 위로 및/또는 아래로의) 하나 이상의 스와이프들 및/또는 (우에서 좌로, 좌에서 우로, 위로 및/또는 아래로의) 롤링을 포함한다. 일부 구현예들 또는 상황들에서, 그래픽과의 의도하지 않은 접촉은 그래픽을 선택하지 않는다. 예를 들어, 선택에 대응하는 제스처가 탭일 때, 애플리케이션 아이콘 위를 스위프(sweep)하는 스와이프 제스처는 옵션적으로, 대응하는 애플리케이션을 선택하지 않는다.

[0158] 디바이스(200)는 또한 "홈" 또는 메뉴 버튼(304)과 같은 하나 이상의 물리적 버튼들을 포함한다. 전술된 바와 같이, 메뉴 버튼(304)은, 디바이스(200) 상에서 실행되는 애플리케이션들의 세트 내의 임의의 애플리케이션(236)으로 내비게이팅하는 데 사용된다. 대안적으로, 일부 실시예들에서, 메뉴 버튼은 터치 스크린(212) 상에 디스플레이된 GUI에서 소프트 키로서 구현된다.

[0159] 일 실시예에서, 디바이스(200)는 터치 스크린(212), 메뉴 버튼(304), 디바이스의 전원을 온/오프하고 디바이스를 잠그기 위한 푸시 버튼(306), 음량 조절 버튼(들)(308), 가입자 식별 모듈(SIM) 카드 슬롯(310), 헤드셋 잭(312), 및 도킹/충전 외부 포트(224)를 포함한다. 푸시 버튼(306)은 옵션적으로 버튼을 누르고 버튼을 미리 정의된 시간 간격 동안 누른 상태로 유지함으로써 디바이스의 전력을 온/오프시키고; 버튼을 누르고 미리 정의된 시간 간격이 경과하기 전에 버튼을 누름해제함으로써 디바이스를 잠그고; 그리고/또는 디바이스를 잠금해제하거나 잠금해제 프로세스를 개시하는 데 사용된다. 대안적인 실시예에서, 디바이스(200)는 또한 마이크로폰(213)을 통해 일부 기능들의 활성화 또는 비활성화를 위한 구두 입력을 수용한다. 디바이스(200)는 또한, 옵션적으로, 터치 스크린(212) 상의 접촉들의 세기를 검출하기 위한 하나 이상의 접촉 세기 센서들(265) 및/또는 디바이스(200)의 사용자를 위해 촉각적 출력들을 생성하기 위한 하나 이상의 촉각적 출력 생성기들(267)을 포함한다.

[0160] 도 4는 일부 실시예들에 따른, 디스플레이 및 터치 감응형 표면을 갖는 예시적인 다기능 디바이스의 블록도이다. 디바이스(400)가 휴대용일 필요는 없다. 일부 실시예들에서, 디바이스(400)는, 랩톱 컴퓨터, 데스크톱 컴퓨터, 태블릿 컴퓨터, 멀티미디어 재생기 디바이스, 내비게이션 디바이스, (어린이 학습 장난감과 같은) 교육용 디바이스, 게이밍 시스템, 또는 제어 디바이스(예컨대, 가정용 또는 산업용 제어기)이다. 디바이스(400)는 전형적으로 하나 이상의 프로세싱 유닛들(CPU)(410), 하나 이상의 네트워크 또는 다른 통신 인터페이스들(460), 메모리(470), 및 이들 컴포넌트를 상호접속하기 위한 하나 이상의 통신 버스들(420)을 포함한다. 통신 버스들(420)은 옵션적으로 시스템 컴포넌트들을 상호접속하고 이들 사이의 통신을 제어하는 회로부(때때로 칩셋이라고 지칭됨)를 포함한다. 디바이스(400)는 전형적으로 터치 스크린 디스플레이인 디스플레이(440)를 포함하는 입/출력(I/O) 인터페이스(430)를 포함한다. I/O 인터페이스(430)는 또한, 옵션적으로, 키보드 및/또는 마우스(또는 다른 포인팅 디바이스)(450) 및 터치패드(455), 디바이스(400) 상에 촉각적 출력들을 생성하기 위한 촉각적 출력 생성기(457)(예컨대, 도 2a를 참조하여 상술한 촉각적 출력 생성기(들)(267)와 유사함), 및 센서들(459)(예컨대, 도 2a를 참조하여 상술한 접촉 세기 센서(들)(265)와 유사한 광 센서, 가속도 센서, 근접 센서, 터치 감응형 센서, 및/또는 접촉 세기 센서)을 포함한다. 메모리(470)는 DRAM, SRAM, DDR RAM 또는 다른 랜덤 액세스 솔리드 스테이트 메모리 디바이스들과 같은 고속 랜덤 액세스 메모리를 포함하며; 옵션적으로 하나 이상의 자기 디스크 저장 디바이스들, 광 디스크 저장 디바이스들, 플래시 메모리 디바이스들, 또는 다른 비휘발성 솔리드 스테이트 저장 디바이스들과 같은 비휘발성 메모리를 포함한다. 메모리(470)는 옵션적으로 CPU(들)(410)로부터 원격으로 위치된 하나 이상의 저장 디바이스들을 포함한다. 일부 실시예들에서, 메모리(470)는 휴대용 다기능 디바이스(200)(도 2a)의 메모리(202)에 저장된 프로그램들, 모듈들, 및 데이터 구조들과 유사한 프로그램들, 모듈들, 및 데이터 구조들, 또는 이들의 서브세트를 저장한다. 또한, 메모리(470)는, 옵션

적으로, 휴대용 다기능 디바이스(200)의 메모리(202) 내에 존재하지 않는 추가의 프로그램들, 모듈들 및 데이터 구조들을 저장한다. 예를 들어, 디바이스(400)의 메모리(470)는, 옵션적으로, 드로잉 모듈(480), 프레젠테이션 모듈(482), 워드 프로세싱 모듈(484), 웹사이트 제작 모듈(486), 디스크 저장 모듈(488), 및/또는 스프레드시트 모듈(490)을 저장하는 반면, 휴대용 다기능 디바이스(200)(도 2a)의 메모리(202)는, 옵션적으로, 이들 모듈들을 저장하지 않는다.

- [0161] 도 4에서의 앞서 식별된 요소들 각각은, 일부 예들에서, 전송된 메모리 디바이스들 중 하나 이상에 저장된다. 앞서 식별된 모듈들 각각은 상술한 기능을 수행하기 위한 명령어들의 세트에 대응한다. 앞서 식별된 모듈들 또는 프로그램들(예컨대, 명령어들의 세트들)은 별개의 소프트웨어 프로그램들, 절차들 또는 모듈들로서 구현될 필요가 없으며, 따라서 다양한 실시예들에서 이들 모듈의 다양한 서브세트들이 조합되거나 그렇지 않으면 재배열된다. 일부 실시예들에서, 메모리(470)는 앞서 식별된 모듈들 및 데이터 구조들의 서브세트를 저장한다. 또한, 메모리(470)는 상술되지 않은 추가의 모듈들 및 데이터 구조들을 저장한다.
- [0162] 이제, 예를 들어 휴대용 다기능 디바이스(200) 상에서 구현될 수 있는 사용자 인터페이스들의 실시예들을 살펴본다.
- [0163] 도 5a는 일부 실시예들에 따른, 휴대용 다기능 디바이스(200) 상의 애플리케이션들의 메뉴에 대한 예시적인 사용자 인터페이스를 도시한다. 유사한 사용자 인터페이스들이 디바이스(400) 상에 구현된다. 일부 실시예들에서, 사용자 인터페이스(500)는 하기의 요소들, 또는 그들의 서브세트나 슈퍼세트를 포함한다:
- [0164] 셀룰러 및 Wi-Fi 신호들과 같은 무선 통신(들)에 대한 신호 강도 표시자(들)(502);
- [0165]  시간(504);
- [0166]  블루투스 표시자(505);
- [0167]  배터리 상태 표시자(506);
- [0168]  다음과 같은, 빈번하게 사용되는 애플리케이션들에 대한 아이콘들을 갖는 트레이(508):
- [0169] o 옵션적으로, 부재 중 전화들 또는 음성메일 메시지들의 개수의 표시자(514)를 포함하는 "전화"라고 라벨링된 전화 모듈(238)에 대한 아이콘(516);
- [0170] o 옵션적으로, 읽지 않은 이메일들의 개수의 표시자(510)를 포함하는 "메일"이라고 라벨링된 이메일 클라이언트 모듈(240)에 대한 아이콘(518);
- [0171] o "브라우저"라고 라벨링된 브라우저 모듈(247)에 대한 아이콘(520); 및
- [0172] o 아이팟(애플 인크.의 상표) 모듈(252)로도 지칭되는, "아이팟"이라고 라벨링된 비디오 및 음악 재생기 모듈(252)에 대한 아이콘(522); 및
- [0173]  다음과 같은, 다른 애플리케이션들에 대한 아이콘들:
- [0174] o "메시지"라고 라벨링된 IM 모듈(241)에 대한 아이콘(524);
- [0175] o "캘린더"라고 라벨링된 캘린더 모듈(248)에 대한 아이콘(526);
- [0176] o "사진"이라고 라벨링된 이미지 관리 모듈(244)에 대한 아이콘(528);
- [0177] o "카메라"라고 라벨링된 카메라 모듈(243)에 대한 아이콘(530);
- [0178] o "온라인 비디오"라고 라벨링된 온라인 비디오 모듈(255)에 대한 아이콘(532);
- [0179] o "주식"이라고 라벨링된 주식 위젯(249-2)에 대한 아이콘(534);
- [0180] o "지도"라고 라벨링된 지도 모듈(254)에 대한 아이콘(536);
- [0181] o "날씨"라고 라벨링된 날씨 위젯(249-1)에 대한 아이콘(538);
- [0182] o "시계"라고 라벨링된 알람 시계 위젯(249-4)에 대한 아이콘(540);
- [0183] o "운동 지원"이라고 라벨링된 운동 지원 모듈(242)에 대한 아이콘(542);

- [0184] o "메모"라고 라벨링된 메모 모듈(253)에 대한 아이콘(544); 및
- [0185] o 디바이스(200) 및 그의 다양한 애플리케이션들(236)에 대한 설정에의 액세스를 제공하는, "설정"이라고 라벨링된, 설정 애플리케이션 또는 모듈에 대한 아이콘(546).
- [0186] 도 5a에 도시된 아이콘 라벨들은 단지 예시적인 것임에 유의해야 한다. 예를 들어, 비디오 및 음악 재생기 모듈(252)에 대한 아이콘(522)은 옵션적으로 "음악" 또는 "음악 재생기"라고 라벨링된다. 기타 라벨들이 옵션적으로 다양한 애플리케이션 아이콘들에 대해 사용된다. 일부 실시예들에서, 개개의 애플리케이션 아이콘에 대한 라벨은 개개의 애플리케이션 아이콘에 대응하는 애플리케이션의 이름을 포함한다. 일부 실시예들에서, 특정 애플리케이션 아이콘에 대한 라벨은 특정 애플리케이션 아이콘에 대응하는 애플리케이션의 이름과 별개이다.
- [0187] 도 5b는 디스플레이(550)(예컨대, 터치 스크린 디스플레이(212))와 별개인 터치 감응형 표면(551)(예컨대, 태블릿 또는 터치패드(455), 도 4)을 갖는 디바이스(예컨대, 디바이스(400), 도 4) 상의 예시적인 사용자 인터페이스를 도시한다. 디바이스(400)는 또한, 옵션적으로, 터치 감응형 표면(551) 상의 접촉들의 세기를 검출하기 위한 하나 이상의 접촉 세기 센서들(예컨대, 센서들(457) 중 하나 이상) 및/또는 디바이스(400)의 사용자에게 대한 촉각적 출력들을 생성하기 위한 하나 이상의 촉각적 출력 생성기들(459)을 포함한다.
- [0188] 후속하는 일부 예들이(터치 감응형 표면과 디스플레이가 조합된) 터치 스크린 디스플레이(212) 상의 입력들을 참조하여 제공될 것이지만, 일부 실시예들에서, 디바이스는 도 5b에 도시된 바와 같이 디스플레이와 별개인 터치 감응형 표면 상의 입력들을 검출한다. 일부 실시예들에서, 터치 감응형 표면(예컨대, 도 5b의 551)은 디스플레이(예컨대, 550) 상의 주축(예컨대, 도 5b의 553)에 대응하는 주축(예컨대, 도 5b의 552)을 갖는다. 이 실시예들에 따르면, 디바이스는 디스플레이 상의 개개의 위치들에 대응하는 위치들(예컨대, 도 5b에서, 560은 568에 대응하고, 562는 570에 대응함)에서 터치 감응형 표면(551)과의 접촉들(예컨대, 도 5b의 560, 562)을 검출한다. 이러한 방식으로, 터치 감응형 표면(예컨대, 도 5b의 551) 상에서 디바이스에 의해 검출된 사용자 입력들(예컨대, 접촉들(560, 562) 및 그 이동들)은 터치 감응형 표면이 디스플레이와 별개일 때 디바이스에 의해 다기능 디바이스의 디스플레이(예컨대, 도 5b의 550) 상의 사용자 인터페이스를 조작하는 데 사용된다. 유사한 방법들이, 옵션적으로, 본 명세서에 기술된 다른 사용자 인터페이스들에 이용된다는 것이 이해되어야 한다.
- [0189] 추가적으로, 하기의 예들이 손가락 입력들(예컨대, 손가락 접촉들, 손가락 탭 제스처들, 손가락 스와이프 제스처들)을 주로 참조하여 주어지는 반면, 일부 실시예들에서, 손가락 입력들 중 하나 이상은 다른 입력 디바이스로부터의 입력(예컨대, 마우스 기반 입력 또는 스타일러스 입력)으로 대체된다는 것이 이해되어야 한다. 예를 들면, 스와이프 제스처가 옵션적으로 (예컨대, 접촉 대신의) 마우스 클릭 및 뒤이은 (예컨대, 접촉의 이동 대신의) 스와이프의 경로를 따른 커서의 이동으로 대체된다. 다른 예로서, (예컨대, 접촉의 검출 및 뒤이은 접촉을 검출하는 것이 중지되는 것 대신에) 커서가 탭 제스처의 위치 위에 위치되는 동안에 탭 제스처가 옵션적으로 마우스 클릭으로 대체된다. 유사하게, 다수의 사용자 입력들이 동시에 검출되는 경우, 다수의 컴퓨터 마우스들이 옵션적으로 동시에 사용되거나, 또는 마우스와 손가락 접촉들이 옵션적으로 동시에 사용되는 것으로 이해하여야 한다.
- [0190] 도 6a는 예시적인 개인 전자 디바이스(600)를 도시한다. 디바이스(600)는 몸체(602)를 포함한다. 일부 실시예들에서, 디바이스(600)는 디바이스들(200, 400)(예컨대, 도 2a 내지 도 4b)에 대하여 기술된 특징부들의 일부 또는 전부를 포함한다. 일부 실시예들에서, 디바이스(600)는 터치 감응형 디스플레이 스크린(604)(이하, 터치 스크린(604))을 갖는다. 터치 스크린(604)에 대한 대안적으로 또는 부가적으로, 디바이스(600)는 디스플레이 및 터치 감응형 표면을 갖는다. 디바이스들(200, 400)에서와 같이, 일부 실시예들에서, 터치 스크린(604)(또는 터치 감응형 표면)은 가해지는 접촉들(예컨대, 터치들)의 세기를 검출하기 위한 하나 이상의 세기 센서들을 갖는다. 터치 스크린(604)(또는 터치 감응형 표면)의 하나 이상의 세기 센서들은 터치들의 세기를 표현하는 출력 데이터를 제공한다. 디바이스(600)의 사용자 인터페이스는 터치들의 세기에 기초하여 터치들에 응답하고, 이는 상이한 세기들의 터치들이 디바이스(600) 상의 상이한 사용자 인터페이스 동작들을 호출할 수 있다는 것을 의미한다.
- [0191] 터치 세기를 검출 및 프로세싱하기 위한 기술들은, 예를 들어, 관련 출원들: 2013년 5월 8일자로 출원되고 발명의 명칭이 "Device, Method, and Graphical User Interface for Displaying User Interface Objects Corresponding to an Application"인 국제 특허 출원 제PCT/US2013/040061호, 및 2013년 11월 11일자로 출원되고 발명의 명칭이 "Device, Method, and Graphical User Interface for Transitioning Between Touch Input to Display Output Relationships"인 국제 특허 출원 제PCT/US2013/069483호에서 찾으며, 이들 각각은 그 전문

이 본 명세서에 참조로서 편입된다.

- [0192] 일부 실시예들에서, 디바이스(600)는 하나 이상의 입력 메커니즘들(606, 608)을 갖는다. 입력 메커니즘들(606, 608)(포함되어 있는 경우)은 물리적인 것이다. 물리적 입력 메커니즘들의 예들은 푸시 버튼들 및 회전가능 메커니즘들을 포함한다. 일부 실시예들에서, 디바이스(600)는 하나 이상의 부착 메커니즘들을 갖는다. 이러한 부착 메커니즘들(포함되어 있는 경우)은 디바이스(600)가, 예를 들어, 모자, 안경, 귀걸이, 목걸이, 셔츠, 재킷, 팔찌, 시계줄, 쇠줄(chain), 바지, 벨트, 신발, 지갑, 배낭 등에 부착될 수 있게 한다. 이 부착 메커니즘들은 디바이스(600)가 사용자에게 착용되도록 한다.
- [0193] 도 6b는 예시적인 개인 전자 디바이스(600)를 도시한다. 일부 실시예들에서, 디바이스(600)는 도 2a, 도 2b, 및 도 4에 대하여 기술된 컴포넌트들의 일부 또는 전부를 포함한다. 디바이스(600)는 I/O 섹션(614)을 하나 이상의 컴퓨터 프로세서들(616) 및 메모리(618)와 동작가능하게 결합하는 버스(612)를 갖는다. I/O 섹션(614)은 터치 감응형 컴포넌트(622), 및 옵션적으로, 터치 세기 감응형 컴포넌트(624)를 가질 수 있는 디스플레이(604)에 접속된다. 더욱이, I/O 섹션(614)은, Wi-Fi, 블루투스, 근거리 통신(NFC), 셀룰러 및/또는 다른 무선 통신 기술들을 사용하여, 애플리케이션 및 운영 체제 데이터를 수신하기 위해 통신 유닛(630)과 접속된다. 디바이스(600)는 입력 메커니즘들(606 및/또는 608)을 포함한다. 입력 메커니즘(606)은 회전가능 입력 디바이스 또는 예를 들어 누름가능 및 회전가능한 입력 디바이스이다. 일부 예들에서, 입력 메커니즘(608)은 버튼이다.
- [0194] 일부 예들에서, 입력 메커니즘(608)은 마이크로폰이다. 개인 전자 디바이스(600)는, 예를 들어 GPS 센서(632), 가속도계(634), 방향 센서(640)(예컨대, 나침반), 자이로스코프(636), 모션 센서(638), 및/또는 이들의 조합과 같은, 다양한 센서들을 포함하고, 이들 모두는 I/O 섹션(614)에 동작가능하게 접속된다.
- [0195] 개인 전자 디바이스(600)의 메모리(618)는, 예를 들어, 하나 이상의 컴퓨터 프로세서들(616)에 의해 실행될 때, 컴퓨터 프로세서들로 하여금, 하기에 기술되는 기술들 및 프로세스들을 수행하게 하는, 컴퓨터 실행가능한 명령어들을 저장하기 위한 비일시적 컴퓨터 판독가능 저장 매체이다. 예를 들어, 컴퓨터 실행가능한 명령어들은 또한 명령어 실행 시스템, 장치 또는 디바이스로부터 명령어들을 폐지하여 명령어들을 실행할 수 있는 컴퓨터 기반 시스템, 프로세서 포함 시스템, 또는 다른 시스템과 같은, 명령어 실행 시스템, 장치 또는 디바이스에 의해 또는 그와 관련하여 사용하기 위한 임의의 비일시적 컴퓨터 판독가능 저장 매체 내에 저장 및/또는 전달된다. 개인 전자 디바이스(600)는 도 6b의 컴포넌트들 및 구성으로 한정되지 않고, 다수의 구성들에서 다른 또는 추가적인 컴포넌트들을 포함할 수 있다.
- [0196] 여기서 사용되는 바와 같이, "어포던스"라는 용어는, 예를 들어, 디바이스들(200, 400, 600, 800, 900, 1000, 및/또는 1100)(도 2a, 도 4, 도 6a와 도 6b, 도 8a와 도 8b, 도 9a와 도 9b, 도 10a와 도 10b, 및 도 11)의 디스플레이 스크린 상에 디스플레이되는 사용자 상호작용 그래픽 사용자 인터페이스 객체(user-interactive graphical user interface object)를 지칭한다. 예를 들어, 이미지(예컨대, 아이콘), 버튼, 및 텍스트(예컨대, 하이퍼링크)는 각각 어포던스를 구성한다.
- [0197] 본 명세서에 사용되는 바와 같이, "포커스 셀렉터(focus selector)"라는 용어는 사용자와 상호작용하고 있는 사용자 인터페이스의 현재 부분을 나타내는 입력 요소를 지칭한다. 커서 또는 다른 위치 마커(location marker)를 포함하는 일부 구현예들에서, 커서가 특정 사용자 인터페이스 요소(예컨대, 버튼, 창, 슬라이더 또는 다른 사용자 인터페이스 요소) 위에 있는 동안 터치 감응형 표면(예컨대, 도 4의 터치패드(455) 또는 도 5b의 터치 감응형 표면(551)) 상에서 입력(예컨대, 누르기 입력)이 검출될 때, 특정 사용자 인터페이스 요소가 검출된 입력에 따라 조정되도록, 커서가 "포커스 셀렉터"로서 기능한다. 터치 스크린 디스플레이 상의 사용자 인터페이스 요소들과의 직접적인 상호작용을 가능하게 하는 터치 스크린 디스플레이(예컨대, 도 2a의 터치 감응형 디스플레이 시스템(212) 또는 도 5a의 터치 스크린(212))을 포함하는 일부 구현예들에서, 입력(예컨대, 접촉에 의한 누르기 입력)이 특정 사용자 인터페이스 요소(예컨대, 버튼, 창, 슬라이더 또는 다른 사용자 인터페이스 요소)의 위치에 있는 터치 스크린 디스플레이 상에서 검출될 때, 특정 사용자 인터페이스 요소가 검출된 입력에 따라 조정되도록, 터치 스크린 상의 검출된 접촉이 "포커스 셀렉터"로서 기능한다. 일부 구현예들에서, (예를 들어 포커스를 하나의 버튼으로부터 다른 버튼으로 움직이도록 탭 키 또는 화살표 키를 사용함으로써) 터치 스크린 디스플레이 상의 대응하는 커서의 이동 또는 접촉의 이동 없이 포커스가 사용자 인터페이스의 하나의 구역으로부터 사용자 인터페이스의 다른 구역으로 이동되며; 이러한 구현예들에서, 포커스 셀렉터는 사용자 인터페이스의 상이한 영역들 사이에서의 포커스의 이동에 따라 움직인다. 포커스 셀렉터가 갖는 특정 형태와 무관하게, 포커스 셀렉터는 일반적으로 (예컨대, 사용자가 상호작용하고자 하는 사용자 인터페이스의 요소를 디바이스에 나타내는 것에 의해) 사용자 인터페이스와의 사용자의 의도된 상호작용을 전달하기 위해 사용자에게 의해 제어되

는 사용자 인터페이스 요소(또는 터치 스크린 디스플레이 상의 접촉)이다. 예를 들어, 터치 감응형 표면(예컨대, 터치패드 또는 터치 스크린) 상에서 누르기 입력이 검출되는 동안 개개의 버튼 위의 포커스 셀렉터(예컨대, 커서, 접촉 또는 선택 상자)의 위치는 (디바이스의 디스플레이 상에 보여지는 다른 사용자 인터페이스 요소들과 달리) 사용자가 개개의 버튼을 활성화시키려고 하고 있다는 것을 나타낼 것이다.

[0198] 명세서 및 청구범위에서 사용되는 바와 같이, 접촉의 "특성 세기"라는 용어는 접촉의 하나 이상의 세기들에 기초한 접촉의 특성을 지칭한다. 일부 실시예들에서, 특성 세기는 다수의 세기 샘플들에 기초한다. 특성 세기는, 옵션적으로, 미리정의된 수의 세기 샘플들, 또는 미리정의된 이벤트에 대해 (예컨대, 접촉을 검출한 후에, 접촉의 리프트오프를 검출하기 이전에, 접촉의 이동의 시작을 검출하기 이전 또는 이후에, 접촉의 종료를 검출하기 이전에, 접촉의 세기의 증가를 검출하기 이전 또는 이후에, 및/또는 접촉의 세기의 감소를 검출하기 이전 또는 이후에) 미리결정된 기간(예컨대, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10 초) 동안 수집된 세기 샘플들의 세트에 기초한다. 접촉의 특성 세기는, 옵션적으로, 접촉의 세기들의 최대 값, 접촉의 세기들의 중간 값(mean value), 접촉의 세기들의 평균값(average value), 접촉의 세기들의 상위 10 백분위 값(top 10 percentile value), 접촉의 세기들의 최대 값의 절반의 값, 접촉의 세기들의 최대 값의 90 퍼센트의 값 등 중 하나 이상에 기초한다. 일부 실시예들에서, 접촉의 지속시간은 (예컨대, 특성 세기가 시간의 경과에 따른 접촉의 세기의 평균일 때) 특성 세기를 결정하는 데 사용된다. 일부 실시예들에서, 동작이 사용자에게 의해 수행되었는지 여부를 결정하기 위해, 특성 세기가 하나 이상의 세기 임계치들의 세트와 비교된다. 예를 들어, 하나 이상의 세기 임계치들의 세트는 제1 세기 임계치 및 제2 세기 임계치를 포함한다. 이 예에서, 제1 임계치를 초과하지 않는 특성 세기를 갖는 접촉의 결과, 제1 동작이 행해지고, 제1 세기 임계치를 초과하지만 제2 세기 임계치를 초과하지 않는 특성 세기를 갖는 접촉의 결과, 제2 동작이 행해지며, 제2 임계치 초과 특성 세기를 갖는 접촉의 결과, 제3 동작이 행해진다. 일부 실시예들에서, 특성 세기와 하나 이상의 임계치들 간의 비교는, 제1 동작을 수행할 것인지 제2 동작을 수행할 것인지 결정하기 위해 사용되기보다는, 하나 이상의 동작들을 수행할지 여부(예컨대, 개개의 동작을 수행할지 또는 개개의 동작을 수행하는 것을 중지할지 여부)를 결정하기 위해 사용된다.

[0199] 일부 실시예들에서, 특성 세기를 결정하기 위해 제스처의 일부분이 식별된다. 예를 들어, 터치 감응형 표면은 시작 위치로부터 이동하여, 그 지점에서 접촉의 세기가 증가하는 종료 위치에 도달하는 연속적인 스와이프 접촉을 수신한다. 이 예에서, 종료 위치에서의 접촉의 특성 세기는 스와이프 접촉 전체가 아니라 연속적인 스와이프 접촉의 일부분에만(예컨대, 종료 위치에서의 스와이프 접촉의 일부분에만) 기초한다. 일부 실시예들에서, 접촉의 특성 세기를 결정하기 전에 스와이프 접촉의 세기들에 평활화 알고리즘(smoothing algorithm)이 적용된다. 예를 들어, 평활화 알고리즘은, 옵션적으로, 비가중 이동 평균(unweighted sliding-average) 평활화 알고리즘, 삼각(triangular) 평활화 알고리즘, 메디안 필터(median filter) 평활화 알고리즘, 및/또는 지수(exponential) 평활화 알고리즘 중 하나 이상을 포함한다. 일부 상황들에서, 이 평활화 알고리즘들은 특성 세기를 결정하기 위해 스와이프 접촉의 세기들에서의 좁은 급등(spike)들 또는 급감(dip)들을 제거한다.

[0200] 터치 감응형 표면 상에서의 접촉의 세기는 접촉 검출 세기 임계치, 가볍게 누르기 세기 임계치, 깊게 누르기 세기 임계치와 같은 하나 이상의 세기 임계치들, 및/또는 하나 이상의 기타 세기 임계치들에 대해 특징지어진다. 일부 실시예들에서, 가볍게 누르기 세기 임계치는, 디바이스가 물리적 마우스의 버튼 또는 트랙패드를 클릭하는 것과 전형적으로 연관된 동작들을 수행하게 될 세기에 대응한다. 일부 실시예들에서, 깊게 누르기 세기 임계치는, 디바이스가 물리적 마우스의 버튼 또는 트랙패드를 클릭하는 것과 전형적으로 연관된 동작들과 상이한 동작들을 수행하게 될 세기에 대응한다. 일부 실시예들에서, 접촉이 가볍게 누르기 세기 임계치 미만의(예컨대, 그리고 공칭 접촉 검출 세기 임계치(이 미만에서는 접촉이 더 이상 검출되지 않음) 초과) 특성 세기로 검출될 때, 디바이스는 가볍게 누르기 세기 임계치 또는 깊게 누르기 세기 임계치와 연관된 동작을 수행함이 없이 터치 감응형 표면 상의 접촉의 이동에 따라 포커스 셀렉터를 이동시킬 것이다. 일반적으로, 달리 언급되지 않는 한, 이 세기 임계치들은 사용자 인터페이스 도면들의 상이한 세트들 사이에서 일관성이 있다.

[0201] 가볍게 누르기 세기 임계치 미만의 세기로부터 가볍게 누르기 세기 임계치와 깊게 누르기 세기 임계치 사이의 세기로의 접촉의 특성 세기의 증가는 때때로 "가볍게 누르기" 입력으로서 지칭된다. 깊게 누르기 세기 임계치 미만의 세기로부터 깊게 누르기 세기 임계치 초과 세기로의 접촉의 특성 세기의 증가는 때때로 "깊게 누르기" 입력으로서 지칭된다. 접촉 검출 세기 임계치 미만의 세기로부터 접촉 검출 세기 임계치와 가볍게 누르기 세기 임계치 사이의 세기로의 접촉의 특성 세기의 증가는 때때로 터치 표면 상의 접촉을 검출하는 것으로서 지칭된다. 접촉 검출 세기 임계치 초과 세기로부터 접촉 검출 세기 임계치 미만의 세기로의 접촉의 특성 세기의 감소는 때때로 터치 표면으로부터의 접촉의 리프트오프를 검출하는 것으로서 지칭된다. 일부 실시예들에

서, 접촉 검출 세기 임계치는 영(0)이다. 일부 실시예들에서, 접촉 검출 세기 임계치는 0 초과이다.

[0202] 본 명세서에 기술된 일부 실시예들에서, 하나 이상의 동작들은, 개개의 누르기 입력을 포함하는 제스처를 검출하는 것에 응답하여 또는 개개의 접촉(또는 복수의 접촉들)으로 수행되는 개개의 누르기 입력을 검출하는 것에 응답하여 수행되며, 여기서 개개의 누르기 입력은 누르기-입력 세기 임계치 초과와 접촉(또는 복수의 접촉들)의 세기의 증가를 검출하는 것에 적어도 부분적으로 기초하여 검출된다. 일부 실시예들에서, 개개의 동작은, 누르기-입력 세기 임계치 초과와 개개의 접촉의 세기의 증가(예컨대, 개개의 누르기 입력의 "다운 스트로크(down stroke)")를 검출하는 것에 응답하여 수행된다. 일부 실시예들에서, 누르기 입력은 누르기-입력 세기 임계치 초과와 개개의 접촉의 세기의 증가 및 누르기-입력 세기 임계치 미만의 접촉의 세기의 후속하는 감소를 포함하며, 개개의 동작은 누르기-입력 임계치 미만의 개개의 접촉의 세기의 후속하는 감소(예컨대, 개개의 누르기 입력의 "업 스트로크(up stroke)")를 검출하는 것에 응답하여 수행된다.

[0203] 일부 실시예들에서, 디바이스는 때때로 "지터(jitter)"로 지칭되는 돌발적인 입력들을 회피하기 위해 세기 히스테리시스를 채용하며, 여기서 디바이스는 누르기-입력 세기 임계치에 대한 미리정의된 관계를 갖는 히스테리시스 세기 임계치(예컨대, 히스테리시스 세기 임계치는 누르기-입력 세기 임계치보다 X 세기 단위 낮거나, 히스테리시스 세기 임계치는 누르기-입력 세기 임계치의 75%, 90% 또는 어떤 적절한 비율임)를 정의하거나 선택한다. 이와 같이, 일부 실시예들에서, 누르기 입력은 누르기-입력 세기 임계치 초과와 개개의 접촉의 세기의 증가 및 누르기-입력 세기 임계치에 대응하는 히스테리시스 세기 임계치 미만의 접촉의 세기의 후속하는 감소를 포함하며, 개개의 동작은 히스테리시스 세기 임계치 미만의 개개의 접촉의 세기의 후속하는 감소(예컨대, 개개의 누르기 입력의 "업 스트로크")를 검출하는 것에 응답하여 수행된다. 유사하게, 일부 실시예들에서, 누르기 입력은 디바이스가 히스테리시스 세기 임계치 이하에서의 세기로부터 누르기-입력 세기 임계치 이상에서의 세기로의 접촉의 세기의 증가, 및 옵션적으로, 히스테리시스 세기 이하에서의 세기로의 접촉의 세기의 후속하는 감소를 검출하는 경우에만 검출되고, 개개의 동작은 누르기 입력(예컨대, 주변환경에 따른 접촉의 세기의 증가 또는 접촉의 세기의 감소)을 검출하는 것에 응답하여 수행된다.

[0204] 설명의 편의상, 누르기-입력 세기 임계치와 연관된 누르기 입력에 응답하여 또는 누르기 입력을 포함하는 제스처에 응답하여 수행되는 동작들의 설명은, 옵션적으로, 누르기-입력 세기 임계치 초과와 접촉의 세기의 증가, 히스테리시스 세기 임계치 미만의 세기로부터 누르기-입력 세기 임계치 초과와 세기로의 접촉의 세기의 증가, 누르기-입력 세기 임계치 미만의 접촉의 세기의 감소, 및/또는 누르기-입력 세기 임계치에 대응하는 히스테리시스 세기 임계치 미만의 접촉의 세기의 감소 중 어느 하나를 검출하는 것에 응답하여 트리거된다. 추가적으로, 동작이 누르기-입력 세기 임계치 미만의 접촉의 세기의 감소를 검출하는 것에 응답하여 수행되는 것으로서 기술되어 있는 예들에서, 동작은, 옵션적으로, 누르기-입력 세기 임계치에 대응하고 그보다 더 낮은 히스테리시스 세기 임계치 미만의 접촉의 세기의 감소를 검출하는 것에 응답하여 수행된다.

[0205] 3. 디지털 어시스턴트 시스템

[0206] 도 7a는 다양한 예들에 따른 디지털 어시스턴트 시스템(700)의 블록도를 도시한다. 일부 예들에서, 디지털 어시스턴트 시스템(700)은 독립형 컴퓨터 시스템 상에서 구현된다. 일부 예들에서, 디지털 어시스턴트 시스템(700)은 다수의 컴퓨터들에 걸쳐 분산된다. 일부 예들에서, 디지털 어시스턴트의 모듈들 및 기능들 중 일부는 서버 부분과 클라이언트 부분으로 나뉘는데, 여기서 클라이언트 부분은 하나 이상의 사용자 디바이스들(예컨대, 디바이스들(104, 122, 200, 400, 600, 800, 900, 1000, 또는 1100)) 상에 존재하고, 예를 들어, 도 1에 도시된 바와 같은 하나 이상의 네트워크들을 통해 서버 부분(예컨대, 서버 시스템(108))과 통신한다. 일부 예들에서, 디지털 어시스턴트 시스템(700)은 도 1에 도시된 서버 시스템(108)(및/또는 DA 서버(106))의 일 구현예이다. 디지털 어시스턴트 시스템(700)은 단지 디지털 어시스턴트 시스템의 일례일 뿐이라는 것, 및 디지털 어시스턴트 시스템(700)이 도시된 것보다 더 많거나 더 적은 컴포넌트들을 가질 수 있거나, 둘 이상의 컴포넌트를 조합할 수 있거나, 또는 상이한 구성 또는 배열의 컴포넌트들을 가질 수 있다는 것에 유의해야 한다. 도 7a에 도시된 다양한 컴포넌트들은 하나 이상의 신호 프로세싱 및/또는 주문형 집적 회로들을 비롯한, 하드웨어, 하나 이상의 프로세서들에 의한 실행을 위한 소프트웨어 명령어들, 펌웨어, 또는 이들의 조합으로 구현된다.

[0207] 디지털 어시스턴트 시스템(700)은 메모리(702), 하나 이상의 프로세서들(704), 입/출력(I/O) 인터페이스(706), 및 네트워크 통신 인터페이스(708)를 포함한다. 이들 컴포넌트들은 하나 이상의 통신 버스들 또는 신호 라인들(710)을 통해 서로 통신할 수 있다.

[0208] 일부 예들에서, 메모리(702)는 비일시적 컴퓨터 판독가능 매체, 예컨대 고속 랜덤 액세스 메모리 및/또는 비휘발성 컴퓨터 판독가능 저장 매체(예컨대, 하나 이상의 자기 디스크 저장 디바이스들, 플래시 메모리

디바이스들, 또는 기타 비휘발성 솔리드 스테이트 메모리 디바이스들)를 포함한다.

- [0209] 일부 예들에서, I/O 인터페이스(706)는 디스플레이들, 키보드들, 터치 스크린들, 및 마이크로폰들과 같은 디지털 어시스턴트 시스템(700)의 입/출력 디바이스들(716)을 사용자 인터페이스 모듈(722)에 결합한다. I/O 인터페이스(706)는, 사용자 인터페이스 모듈(722)과 함께, 사용자 입력들(예컨대, 음성 입력, 키보드 입력들, 터치 입력들 등)을 수신하고 이에 따라 이들을 프로세싱한다. 일부 예들에서, 예컨대 디지털 어시스턴트가 독립형 사용자 디바이스 상에서 구현되는 경우, 디지털 어시스턴트 시스템(700)은 도 2a, 도 4, 도 6a와 도 6b, 도 8a와 도 8b, 도 9a와 도 9b, 도 10a와 도 10b, 및 도 11 각각의 디바이스들(200, 400, 600, 800, 900, 1000, 또는 1100)에 대하여 기술된 컴포넌트들 및 I/O 통신 인터페이스들 중 임의의 것을 포함한다. 일부 예들에서, 디지털 어시스턴트 시스템(700)은 디지털 어시스턴트 구현예의 서버 부분을 표현하고, 사용자 디바이스(예컨대, 디바이스들(104, 200, 400, 600, 800, 900, 1000, 또는 1100)) 상에 존재하는 클라이언트 측 부분을 통해 사용자와 상호작용할 수 있다.
- [0210] 일부 예들에서, 네트워크 통신 인터페이스(708)는 유선 통신 포트(들)(712) 및/또는 무선 송신 및 수신 회로부(714)를 포함한다. 유선 통신 포트(들)는 하나 이상의 유선 인터페이스들, 예컨대 이더넷, 범용 직렬 버스(USB), 파이어와이어 등을 통해 통신 신호들을 수신하고 전송한다. 무선 회로부(714)는 통신 네트워크들 및 기타 통신 디바이스들로부터/로 RF 신호들 및/또는 광 신호들을 수신하고 전송한다. 무선 통신은 GSM, EDGE, CDMA, TDMA, 블루투스, Wi-Fi, VoIP, Wi-MAX, 또는 임의의 기타 적합한 통신 프로토콜과 같은 복수의 통신 표준들, 프로토콜들, 및 기술들 중 임의의 것을 이용한다. 네트워크 통신 인터페이스(708)는 인터넷, 인트라넷, 및/또는 무선 네트워크, 예컨대 셀룰러 전화 네트워크, 무선 LAN(local area network), 및/또는 MAN(metropolitan area network)과 같은 네트워크들을 이용하는 디지털 어시스턴트 시스템(700)과 기타 디바이스들 간의 통신을 가능하게 한다.
- [0211] 일부 예들에서, 메모리(702) 또는 메모리(702)의 컴퓨터 판독가능 저장 매체들은, 운영 체제(718), 통신 모듈(720), 사용자 인터페이스 모듈(722), 하나 이상의 애플리케이션들(724), 및 디지털 어시스턴트 모듈(726)의 전체 또는 그들의 서브세트를 포함한, 프로그램들, 모듈들, 명령어들, 및 데이터 구조들을 저장한다. 특히, 메모리(702) 또는 메모리(702)의 컴퓨터 판독가능 저장 매체는 하기에 기술되는 프로세스들을 수행하기 위한 명령어들을 저장한다. 하나 이상의 프로세서들(704)은 이러한 프로그램들, 모듈들, 및 명령어들을 실행하고, 데이터 구조들로부터/로 판독/기록한다.
- [0212] 운영 체제(718)(예를 들어, 다윈(Darwin), RTXC, LINUX, UNIX, iOS, OS X, WINDOWS, 또는 VxWorks와 같은 임베디드 운영 체제)는 일반적인 시스템 태스크들(예를 들어, 메모리 관리, 저장 디바이스 제어, 전력 관리 등)을 제어하고 관리하기 위한 다양한 소프트웨어 컴포넌트들 및/또는 드라이버들을 포함하고, 다양한 하드웨어, 펌웨어, 및 소프트웨어 컴포넌트들 간의 통신들을 용이하게 한다.
- [0213] 통신 모듈(720)은 네트워크 통신 인터페이스(708)를 통해 디지털 어시스턴트 시스템(700)과 기타 디바이스들 간의 통신을 용이하게 한다. 예를 들어, 통신 모듈(720)은 도 2a, 도 4, 도 6a와 도 6b 각각에 도시된 디바이스들(200, 400, 600)과 같은 전자 디바이스들의 RF 회로부(208)와 통신한다. 통신 모듈(720)은 또한 무선 회로부(714) 및/또는 유선 통신 포트(712)에 의해 수신된 데이터를 처리하기 위한 다양한 컴포넌트들을 포함한다.
- [0214] 사용자 인터페이스 모듈(722)은 I/O 인터페이스(706)를 통해 사용자로부터(예컨대, 키보드, 터치 스크린, 포인팅 디바이스, 제어기, 및/또는 마이크로폰으로부터) 커맨드들 및/또는 입력들을 수신하고, 디스플레이 상에 사용자 인터페이스 객체들을 생성한다. 사용자 인터페이스 모듈(722)은 또한 출력들(예컨대, 스피치, 사운드, 애니메이션, 텍스트, 아이콘들, 진동들, 햅틱 피드백, 조명 등)을 준비하고, I/O 인터페이스(706)를 통해(예컨대, 디스플레이들, 오디오 채널들, 스피커들, 및 터치패드들 등을 통해) 사용자에게 그들을 전달한다.
- [0215] 애플리케이션들(724)은 하나 이상의 프로세서들(704)에 의해 실행되도록 구성된 프로그램들 및/또는 모듈들을 포함한다. 예를 들어, 디지털 어시스턴트 시스템이 독립형 사용자 디바이스 상에 구현되는 경우, 애플리케이션들(724)은 게임들, 캘린더 애플리케이션, 내비게이션 애플리케이션, 또는 이메일 애플리케이션과 같은 사용자 애플리케이션들을 포함한다. 디지털 어시스턴트 시스템(700)이 서버 상에서 구현되는 경우, 애플리케이션들(724)은, 예를 들어 리소스 관리 애플리케이션들, 진단 애플리케이션들, 또는 스케줄링 애플리케이션들을 포함한다.
- [0216] 메모리(702)는 또한 디지털 어시스턴트 모듈(726)(또는 디지털 어시스턴트의 서버 부분)을 저장한다. 일부 예들에서, 디지털 어시스턴트 모듈(726)은 하기의 하위모듈들, 또는 그들의 서브세트 또는 수퍼세트를 포함한다:

입/출력 프로세싱 모듈(728), 스피치-텍스트(speech-to-text: STT) 프로세싱 모듈(730), 자연 언어 프로세싱 모듈(732), 대화 흐름 프로세싱 모듈(734), 태스크 흐름 프로세싱 모듈(736), 서비스 프로세싱 모듈(738), 및 스피치 합성 모듈(740). 이러한 모듈들 각각은 디지털 어시스턴트 모듈(726)의 하기의 시스템 또는 데이터 및 모델들, 또는 그들의 서브세트 또는 슈퍼세트 중 하나 이상에 대한 액세스를 갖는다: 온톨로지(760), 어휘 인덱스(744), 사용자 데이터(748), 태스크 흐름 모델들(754), 서비스 모델들(756), 및 ASR 시스템.

[0217] 일부 예들에서, 디지털 어시스턴트 모듈(726)에서 구현되는 모델들, 데이터, 및 프로세싱 모듈들을 사용하여, 디지털 어시스턴트는 다음 중 적어도 일부를 수행할 수 있다: 스피치 입력을 텍스트로 변환하는 것; 사용자로부터 수신된 자연 언어 입력에 표현된 사용자의 의도를 식별하는 것; 사용자의 의도를 완전히 추론하는 데 필요한 정보를(예컨대, 단어들, 게임들, 의도들 등의 중의성을 해소함(disambiguating)으로써) 능동적으로 이끌어내고 획득하는 것; 추론된 의도를 이행하기 위한 태스크 흐름을 결정하는 것; 및 태스크 흐름을 실행하여 추론된 의도를 이행하는 것.

[0218] 일부 예들에서, 도 7b에 도시된 바와 같이, I/O 프로세싱 모듈(728)은 도 7a의 I/O 디바이스들(716)을 통해 사용자와 상호작용하거나 또는 도 7a의 네트워크 통신 인터페이스(708)를 통해 사용자 디바이스(예컨대, 디바이스들(104, 200, 400, 또는 600))와 상호작용하여, 사용자 입력(예컨대, 스피치 입력)을 획득하고 사용자 입력에 대한 응답들(예컨대, 스피치 출력들로서)을 제공한다. I/O 프로세싱 모듈(728)은 사용자 입력의 수신과 함께 또는 사용자 입력의 수신 직후에, 사용자 디바이스로부터 사용자 입력과 연관된 컨텍스트 정보를 옵션적으로 획득한다. 컨텍스트 정보는 사용자-특정 데이터, 어휘, 및/또는 사용자 입력에 관련된 선호도들을 포함한다. 일부 예들에서, 컨텍스트 정보는 또한 사용자 요청이 수신된 시간에서의 사용자 디바이스의 소프트웨어 및 하드웨어 상태 및/또는 사용자 요청이 수신되는 시간에서의 사용자의 주위 환경에 관련된 정보를 포함한다. 일부 예들에서, I/O 프로세싱 모듈(728)은 또한 사용자 요청에 관하여 사용자에게 후속 질문들을 전송하고, 그로부터 답변들을 수신한다. 사용자 요청이 I/O 프로세싱 모듈(728)에 의해 수신되고 사용자 요청이 스피치 입력을 포함하는 경우, I/O 프로세싱 모듈(728)은 스피치-텍스트 변환을 위해 스피치 입력을 STT 프로세싱 모듈(730)(또는 스피치 인식기)로 전달한다.

[0219] STT 프로세싱 모듈(730)은 하나 이상의 ASR 시스템들을 포함한다. 하나 이상의 ASR 시스템들은 I/O 프로세싱 모듈(728)을 통해 수신되는 스피치 입력을 프로세싱하여 인식 결과를 생성할 수 있다. 각각의 ASR 시스템은 프론트-엔드 스피치 프리프로세서(front-end speech pre-processor)를 포함한다. 프론트-엔드 스피치 프리프로세서는 스피치 입력으로부터 대표적인 특징을 추출한다. 예를 들어, 프론트-엔드 스피치 프리프로세서는 스피치 입력을 푸리에 변환을 수행하여 대표적인 다차원 벡터들의 시퀀스로서 스피치 입력을 특징짓는 스펙트럼 특징을 추출한다. 또한, 각각의 ASR 시스템은 하나 이상의 스피치 인식 모델들(예컨대, 음향 모델들 및/또는 언어 모델들)을 포함하고, 하나 이상의 스피치 인식 엔진들을 구현한다. 스피치 인식 모델의 예는 은닉 마르코프 모델(hidden Markov models), 가우시안 혼합 모델(Gaussian-Mixture Models), 딥 신경망 모델(Deep Neural Network Models), n-gram 언어 모델, 및 기타 통계 모델을 포함한다. 스피치 인식 엔진의 예는 동적 시간 왜곡 기반 엔진 및 가중치 유한 상태 변환기(WFST) 기반 엔진을 포함한다. 하나 이상의 스피치 인식 모델들 및 하나 이상의 스피치 인식 엔진들은 중간 인식 결과들(예를 들어, 음소들, 음소 문자열들, 및 하위 단어들), 및 궁극적으로 텍스트 인식 결과들(예컨대, 단어들, 단어 문자열들, 또는 토큰들의 시퀀스)을 생성하기 위해 프론트-엔드 스피치 프리프로세서의 추출된 대표 특징들을 프로세싱하는 데 사용된다. 일부 예들에서, 스피치 입력은 제3자 서비스에 의해 또는 사용자의 디바이스(예컨대, 디바이스(104, 200, 400, 또는 600)) 상에서 적어도 부분적으로 프로세싱되어 인식 결과를 생성한다. 일단 STT 프로세싱 모듈(730)이 텍스트 문자열(예를 들어, 단어들, 또는 단어들의 시퀀스, 또는 토큰들의 시퀀스)을 포함하는 인식 결과를 생성하면, 인식 결과는 의도 추론을 위해 자연 언어 프로세싱 모듈(732)로 전달된다. 일부 예들에서, STT 프로세싱 모듈(730)은 스피치 입력의 다수의 후보 텍스트 표현들을 생성한다. 각각의 후보 텍스트 표현은 스피치 입력에 대응하는 단어들 또는 토큰들의 시퀀스이다. 일부 예들에서, 각각의 후보 텍스트 표현은 스피치 인식 신뢰도 점수와 연관된다. 스피치 인식 신뢰도 점수에 기초하여, STT 프로세싱 모듈(730)은 후보 텍스트 표현을 순위를 매기고, 의도 추론을 위해 자연 언어 프로세싱 모듈(732)에 n-베스트(best)(예를 들어, n개의 가장 높은 순위가 매겨짐) 후보 텍스트 표현(들)을 제공하며, 여기서 n은 0보다 큰 미리결정된 정수이다. 예를 들어, 일례에서, 가장 높은 순위가 매겨진(n=1) 후보 텍스트 표현만이 의도 추론을 위해 자연 언어 프로세싱 모듈(732)에 전달된다. 다른 예에서, 5개의 가장 높은 순위가 매겨진(n=5) 후보 텍스트 표현들이 의도 추론을 위해 자연 언어 프로세싱 모듈(732)에 전달된다.

[0220] 스피치-텍스트 프로세싱에 대한 보다 많은 상세사항들은 2011년 9월 20일자로 출원된 "Consolidating Speech Recognition Results"에 대한 미국 특허 출원 제13/236,942호에 기술되어 있으며, 그 전체 개시내용은 본 명세

서에 참조로서 편입된다.

- [0221] 일부 예들에서, STT 프로세싱 모듈(730)은 음성 기호 변환 모듈(731)을 통해 인식가능한 단어들의 어휘를 포함하고/하거나 이에 액세스한다. 각각의 어휘 단어는 스피치 인식 음성 기호로 표현된 단어의 하나 이상의 후보 발음들과 연관된다. 특히, 인식가능한 단어의 어휘는 복수의 후보 발음들과 연관된 단어를 포함한다. 예를 들어, 어휘는 /tə'meɪrʊ/ 및 /tə'matʊ/의 후보 발음과 연관된 "tomato"라는 단어를 포함한다. 또한, 어휘 단어는 사용자로부터의 이전 스피치 입력에 기초한 맞춤형 후보 발음과 연관된다. 이러한 맞춤형 후보 발음은 STT 프로세싱 모듈(730)에 저장되고, 디바이스 상의 사용자의 프로파일을 통해 특정 사용자와 연관된다. 일부 예들에서, 단어에 대한 후보 발음은 단어의 철자 및 하나 이상의 언어 및/또는 음성 규칙들에 기초하여 결정된다. 일부 예들에서, 후보 발음은 예를 들어, 알려진 정준 발음(canonical pronunciations)에 기초하여 수동으로 생성된다.
- [0222] 일부 예들에서, 후보 발음은 후보 발음의 공통성에 기초하여 순위가 매겨진다. 예를 들어, 후보 발음 /tə'meɪrʊ/는 /tə'matʊ/보다 높은 순위가 매겨지는데, 그 이유는 이전 발음이 더 일반적으로 사용되는 발음이기 때문이다(예를 들어, 모든 사용자 중에서, 특정 지리적 지역에 있는 사용자의 경우, 또는 임의의 다른 적절한 사용자들의 서브세트의 경우). 일부 예들에서, 후보 발음은 후보 발음이 사용자와 연관된 맞춤형 후보 발음인지 여부에 기초하여 순위가 매겨진다. 예를 들어, 맞춤형 후보 발음은 정준 후보 발음보다 높은 순위가 매겨진다. 이는 정준 발음에서 벗어나는 고유한 발음을 가진 고유 명사를 인식하는 데 유용할 수 있다. 일부 예들에서, 후보 발음은 지리적인 기원, 국적, 또는 민족성과 같은 하나 이상의 스피치 특성들과 연관된다. 예를 들어, 후보 발음 /tə'meɪrʊ/는 미국과 연관되는 반면, 후보 발음 /tə'matʊ/는 영국과 연관된다. 또한, 후보 발음의 순위는 디바이스 상의 사용자 프로파일에 저장된 사용자의 하나 이상의 특성들(예를 들어, 지리적 기원, 국적, 민족성 등)에 기초한다. 예를 들어, 사용자가 미국과 연관되어 있다는 사용자의 프로파일로부터 결정될 수 있다. 미국과 연관되어 있는 사용자에 기초하여, 후보 발음 /tə'meɪrʊ/(미국과 연관됨)는 후보 발음 /tə'matʊ/(영국과 연관됨)보다 높은 순위가 매겨진다. 일부 예들에서, 순위가 매겨진 후보 발음들 중 하나는 예측된 발음(예를 들어, 가장 가능성이 있는 발음)으로서 선택된다.
- [0223] 스피치 입력이 수신될 때, STT 프로세싱 모듈(730)은 (예를 들어, 음향 모델을 사용하여) 스피치 입력에 대응하는 음소를 결정하는 데 사용되고, 이어서 (예를 들어, 언어 모델을 사용하여) 음소에 매칭되는 단어를 결정하고자 시도한다. 예를 들어, STT 프로세싱 모듈(730)이 스피치 입력의 일부분에 대응하는 음소들의 시퀀스 /tə'meɪrʊ/를 먼저 식별하는 경우, 이어서 어휘 인덱스(744)에 기초하여, 이 시퀀스가 단어 "tomato"에 대응한다고 결정할 수 있다.
- [0224] 일부 예들에서, STT 프로세싱 모듈(730)은 근사 매칭 기술을 사용하여 발화 중인 단어를 결정한다. 따라서, 예를 들어, STT 프로세싱 모듈(730)은 음소들의 특정 시퀀스가 그 단어에 대한 음소들의 후보 시퀀스 중 하나가 아니더라도 음소들의 시퀀스 /tə'meɪrʊ/가 단어 "tomato"에 대응한다고 결정한다.
- [0225] 디지털 어시스턴트의 자연 언어 프로세싱 모듈(732)("자연 언어 프로세서")은 STT 프로세싱 모듈(730)에 의해 생성된 n-베스트 후보 텍스트 표현(들)("단어 시퀀스(들)" 또는 "토큰 시퀀스(들)")을 취하고, 후보 텍스트 표현들 각각을 디지털 어시스턴트에 의해 인식되는 하나 이상의 "행동가능한 의도들"과 연관시키고자 시도한다. "행동가능한 의도" (또는 "사용자 의도")는 디지털 어시스턴트에 의해 수행될 수 있는 태스크를 표현할 수 있고, 태스크 흐름 모델들(754)에서 구현되는 연관 태스크 흐름을 가질 수 있다. 연관 태스크 흐름은 디지털 어시스턴트가 태스크를 수행하기 위해 취하는 일련의 프로그래밍된 행동들 및 단계들이다. 디지털 어시스턴트의 능력들의 범주는 태스크 흐름 모델들(754)에서 구현되고 저장된 태스크 흐름들의 수 및 종류, 또는 다시 말해, 디지털 어시스턴트가 인식하는 "행동가능한 의도들"의 수 및 종류에 의존한다. 그러나, 디지털 어시스턴트의 효율성은 또한 자연 언어로 표현된 사용자 요청으로부터 정확한 "행동가능한 의도(들)"를 추론하는 어시스턴트의 능력에 의존한다.
- [0226] 일부 예들에서, STT 프로세싱 모듈(730)로부터 획득된 단어들 또는 토큰들의 시퀀스에 부가하여, 자연 언어 프로세싱 모듈(732)은 또한 예를 들어, I/O 프로세싱 모듈(728)로부터 사용자 요청과 연관된 컨텍스트 정보를 수신한다. 자연 언어 프로세싱 모듈(732)은 옵션적으로 컨텍스트 정보를 이용하여, STT 프로세싱 모듈(730)로부터 수신된 후보 텍스트 표현들에 포함된 정보를 명확하게 하고, 보완하고/하거나 추가로 정의한다. 컨텍스트

정보는, 예를 들어 사용자 선호도들, 사용자 디바이스의 하드웨어 및/또는 소프트웨어 상태들, 사용자 요청 전, 요청 중, 또는 요청 직후에 수집되는 센서 정보, 디지털 어시스턴트와 사용자 사이의 이전 상호작용들(예컨대, 대화) 등을 포함한다. 본 명세서에 기술하는 바와 같이, 일부 예들에서, 컨텍스트 정보는 동적일 수 있고, 시간의 변화, 위치, 대화의 내용, 및 기타 인자들에 따라 변화한다.

- [0227] 일부 예들에서, 자연 언어 프로세싱은, 예컨대 온톨로지(760)에 기초한다. 온톨로지(760)는 많은 노드들을 포함하는 계층 구조(hierarchical structure)인데, 각각의 노드는 "행동가능한 의도", 또는 "행동가능한 의도들" 중 하나 이상에 관련된 "속성(property)", 또는 기타 "속성들" 중 어느 하나를 표현한다. 전술한 바와 같이, "행동가능한 의도"는 디지털 어시스턴트가 수행할 수 있는, 즉, 그것이 "행동가능"하거나 또는 영향을 미칠 수 있는 태스크를 표현한다. "속성"은 행동가능한 의도 또는 다른 속성의 하위양태와 연관되는 파라미터를 표현한다. 온톨로지(760) 내의 행동가능한 의도 노드와 속성 노드 간의 연결성은 속성 노드에 의해 표현되는 파라미터가 행동가능한 의도 노드에 의해 표현되는 태스크에 어떻게 관련되는지를 정의한다.
- [0228] 일부 예들에서, 온톨로지(760)는 행동가능한 의도 노드들 및 속성 노드들로 구성된다. 온톨로지(760) 내에서, 각각의 행동가능한 의도 노드는 직접적으로 또는 하나 이상의 중간 속성 노드들을 통해 하나 이상의 속성 노드들에 링크된다. 유사하게, 각각의 속성 노드는 직접적으로 또는 하나 이상의 중간 속성 노드들을 통해 하나 이상의 행동가능한 의도 노드들에 링크된다. 예를 들어, 도 7c에 도시된 바와 같이, 온톨로지(760)는 "레스토랑 예약" 노드(즉, 행동가능한 의도 노드)를 포함한다. 속성 노드들 "레스토랑", (예약을 위한) "날짜/시간", 및 "인원수"는 행동가능한 의도 노드(즉, "레스토랑 예약" 노드)에 각각 직접적으로 링크된다.
- [0229] 추가로, 속성 노드들 "요리", "가격대", "전화 번호", 및 "위치"는 속성 노드 "레스토랑"의 하위노드들이고, 중간 속성 노드 "레스토랑"을 통해 "레스토랑 예약" 노드(즉, 행동가능한 의도 노드)에 각각 링크된다. 다른 예를 들어, 도 7c에 도시된 바와 같이, 온톨로지(760)는 또한 "리마인더 설정" 노드(즉, 다른 행동가능한 의도 노드)를 포함한다. 속성 노드들 (리마인더를 설정하기 위한) "날짜/시간" 및 (리마인더를 위한) "주제"는 "리마인더 설정" 노드에 각각 링크된다. 속성 "날짜/시간"이 레스토랑 예약을 하는 태스크 및 리마인더를 설정하는 태스크 둘 모두에 관련되기 때문에, 속성 노드 "날짜/시간"은 온톨로지(760) 내의 "레스토랑 예약" 노드 및 "리마인더 설정" 노드 둘 모두에 링크된다.
- [0230] 행동가능한 의도 노드는, 그의 링크된 속성 노드들과 함께, "도메인"으로 기술된다. 본 논의에서, 각각의 도메인은 개개의 행동가능한 의도와 연관되고, 특정한 행동가능한 의도와 연관된 노드들(및 이들 사이의 관계들)의 그룹을 지칭한다. 예를 들어, 도 7c에 도시된 온톨로지(760)는 온톨로지(760) 내의 레스토랑 예약 도메인(762)의 일례 및 리마인더 도메인(764)의 일례를 포함한다. 레스토랑 예약 도메인은 행동가능한 의도 노드 "레스토랑 예약", 속성 노드들 "레스토랑", "날짜/시간", 및 "인원수", 및 하위속성 노드들 "요리", "가격대", "전화 번호", 및 "위치"를 포함한다. 리마인더 도메인(764)은 행동가능한 의도 노드 "리마인더 설정", 및 속성 노드들 "주제" 및 "날짜/시간"을 포함한다. 일부 예들에서, 온톨로지(760)는 많은 도메인들로 구성된다. 각각의 도메인은 하나 이상의 속성 노드들을 하나 이상의 다른 도메인들과 공유한다. 예를 들어, "날짜/시간" 속성 노드는 레스토랑 예약 도메인(762) 및 리마인더 도메인(764) 외에도, 많은 상이한 도메인들(예컨대, 스케줄링 도메인, 여행 예약 도메인, 영화 티켓 도메인 등)과 연관된다.
- [0231] 도 7c가 온톨로지(760) 내의 2개의 예시적인 도메인들을 도시하지만, 다른 도메인들은, 예를 들어 "영화 찾기", "전화 통화 개시", "길 찾기", "미팅 스케줄링", "메시지 전송", "질문에 대한 답변 제공", "목록 읽기", "내비게이션 명령어들 제공", "태스크에 대한 명령어들 제공" 등을 포함한다. "메시지 전송" 도메인은 "메시지 전송" 행동가능한 의도 노드와 연관되고, "수신자(들)", "메시지 유형", 및 "메시지 본문"과 같은 속성 노드들을 추가로 포함한다. 속성 노드 "수신자"는, 예를 들어 "수신자 이름" 및 "메시지 주소"와 같은 하위속성 노드들에 의해 추가로 정의된다.
- [0232] 일부 예들에서, 온톨로지(760)는 디지털 어시스턴트가 이해할 수 있고 그에 의거하여 동작할 수 있는 모든 도메인들(및 그에 따른 행동가능한 의도들)을 포함한다. 일부 예들에서, 온톨로지(760)는, 예컨대 전체 도메인들 또는 노드들을 추가하거나 제거함으로써 또는 온톨로지(760) 내에서 노드들 사이의 관계를 수정함으로써 수정된다.
- [0233] 일부 예들에서, 다수의 관련 행동가능한 의도들과 연관된 노드들은 온톨로지(760) 내에서 "상위 도메인" 아래에 군집된다. 예를 들어, "여행" 상위 도메인은 여행에 관련된 속성 노드들 및 행동가능한 의도 노드들의 군집을 포함한다. 여행에 관련된 행동가능한 의도 노드들은 "항공 예약", "호텔 예약", "자동차 렌탈", "길찾기", "관심 지점 찾기" 등을 포함한다. 동일한 상위 도메인(예컨대, "여행" 상위 도메인) 아래의 행동가능한 의도 노드

들은 많은 속성 노드들을 공동으로 갖는다. 예를 들어, "항공 예약", "호텔 예약", "자동차 렌탈", "길찾기", 및 "관심 지점 찾기"에 대한 행동가능한 의도 노드들은 속성 노드들 "시작 위치", "목적지", "출발 날짜/시간", "도착 날짜/시간", 및 "인원수" 중 하나 이상을 공유한다.

[0234] 일부 예들에서, 온톨로지(760) 내의 각각의 노드는, 노드에 의해 표현되는 속성 또는 행동가능한 의도와 관련된 단어 및/또는 구절들의 세트와 연관된다. 각각의 노드와 연관된 단어 및/또는 구절들의 개개의 세트는 노드와 연관된 소위 "어휘"이다. 각각의 노드와 연관된 단어 및/또는 구절들의 개개의 세트는 노드에 의해 표현되는 속성 또는 행동가능한 의도와 관련하여 어휘 인덱스(744)에 저장된다. 예를 들어, 도 7b로 돌아와서, "레스토랑"의 속성에 대한 노드와 연관된 어휘는 "음식", "음료", "요리", "배고프다", "먹다", "피자", "패스트푸드", "식사" 등과 같은 단어들을 포함한다. 다른 예를 들어, "전화 통화 개시"의 행동가능한 의도에 대한 노드와 연관된 어휘는 "통화", "전화", "다이얼", "벨소리", "이 번호로 전화", "~에게 전화" 등과 같은 단어 및 구절들을 포함한다. 어휘 인덱스(744)는 상이한 언어들의 단어 및 구절들을 옵션적으로 포함한다.

[0235] 자연 언어 프로세싱 모듈(732)은 STT 프로세싱 모듈(730)로부터 후보 텍스트 표현들(예를 들어, 텍스트 문자열(들) 또는 토큰 시퀀스(들))을 수신하고, 각각의 후보 표현에 대해, 후보 텍스트 표현 내의 단어에 의해 어떤 노드들이 연루되는지를 결정한다. 일부 예들에서, 후보 텍스트 표현 내의 단어 또는 구절이 (어휘 인덱스(744)를 통해) 온톨로지(760) 내의 하나 이상의 노드들과 연관되는 것으로 밝혀지는 경우, 단어 또는 구절은 이들 노드들을 "트리거" 또는 "활성화"시킨다. 활성화된 노드들의 양 및/또는 상대적 중요도에 기초하여, 자연 언어 프로세싱 모듈(732)은 사용자가 디지털 어시스턴트로 하여금 수행하도록 의도했던 태스크로서 행동가능한 의도들 중 하나를 선택한다. 일부 예들에서, 가장 많이 "트리거"된 노드들을 갖는 도메인이 선택된다. 일부 예들에서, 가장 높은 신뢰도 값(예컨대, 그것의 다양한 트리거된 노드들의 상대적 중요도에 기초함)을 갖는 도메인이 선택된다. 일부 예들에서, 도메인은 트리거된 노드들의 수 및 중요도의 조합에 기초하여 선택된다. 일부 예들에서, 디지털 어시스턴트가 사용자로부터의 유사한 요청을 이전에 정확하게 해석했는지 여부와 같은 추가 인자들도 노드를 선택하는 데 있어 마찬가지로 고려된다.

[0236] 사용자 데이터(748)는 사용자 특정 정보, 예컨대 사용자 특정 어휘, 사용자 선호도, 사용자 주소, 사용자의 초기 설정 언어 및 제2 언어, 사용자의 연락처 목록, 및 각각의 사용자에 대한 기타 단기 또는 장기 정보를 포함한다. 일부 예들에서, 자연 언어 프로세싱 모듈(732)은 사용자-특정 정보를 이용하여 사용자 입력에 포함된 정보를 보완하여 사용자 의도를 추가로 정의한다. 예를 들어, "내 생일 파티에 내 친구들을 초대하라"라는 사용자 요청에 대하여, 자연 언어 프로세싱 모듈(732)은, "친구들"이 누구인지 그리고 "생일 파티"가 언제, 어디에서 열릴지를 결정하기 위해 사용자에게 그의/그녀의 요청에서 그러한 정보를 명확하게 제공하도록 요구하는 대신, 사용자 데이터(748)에 액세스할 수 있다.

[0237] 일부 예들에서, 자연 언어 프로세싱 모듈(732)은 하나 이상의 기계 학습 메커니즘들(예를 들어, 신경망)을 사용하여 구현된다는 것을 인식해야 한다. 특히, 하나 이상의 기계 학습 메커니즘들은 후보 텍스트 표현 및 후보 텍스트 표현과 연관된 컨텍스트 정보를 수신하도록 구성된다. 후보 텍스트 표현 및 연관된 컨텍스트 정보에 기초하여, 하나 이상의 기계 학습 메커니즘들은 후보 행동가능한 의도들의 세트에 대한 의도 신뢰도 점수를 결정하도록 구성된다. 자연 언어 프로세싱 모듈(732)은 결정된 의도 신뢰도 점수에 기초하여 후보 행동가능한 의도들의 세트로부터 하나 이상의 후보 행동가능한 의도들을 선택할 수 있다. 일부 예들에서, 온톨로지(예를 들어, 온톨로지(760))는 또한 후보 행동가능한 의도 세트로부터 하나 이상의 후보 행동가능한 의도들을 선택하는 데 사용된다.

[0238] 토큰 문자열에 기초하여 온톨로지를 검색하는 다른 상세 사항들이 2008년 12월 22일자로 출원된 "Method and Apparatus for Searching Using An Active Ontology"에 대한 미국 특허 출원 제12/341,743호에 기술되어 있으며, 이 출원의 전체 개시내용은 본 명세서에 참조로서 편입된다.

[0239] 일부 예들에서, 일단 자연 언어 프로세싱 모듈(732)이 사용자 요청에 기초하여 행동가능한 의도(또는 도메인)를 식별하면, 자연 언어 프로세싱 모듈(732)은 식별된 행동가능한 의도를 표현하기 위해 구조화된 쿼리(structured query)를 생성한다. 일부 예들에서, 구조화된 쿼리는, 행동가능한 의도를 위한 도메인 내의 하나 이상의 노드들에 대한 파라미터들을 포함하고, 파라미터들 중 적어도 일부에는 사용자 요청에 특정되는 특정 정보 및 요건이 채워져 있다. 예를 들어, 사용자는 "7시에 스시 식당에 저녁 예약을 해"라고 말한다. 이러한 경우, 자연 언어 프로세싱 모듈(732)은 사용자 입력에 기초하여 행동가능한 의도가 "레스토랑 예약"이라고 정확하게 식별할 수 있다. 온톨로지에 따라, "레스토랑 예약" 도메인에 대한 구조화된 쿼리는 {요리}, {시간}, {날짜}, {인원수} 등과 같은 파라미터들을 포함한다. 일부 예들에서, 스피치 입력, 및 STT 프로세싱 모듈(730)을 사용

하여 스피치 입력으로부터 도출된 텍스트에 기초하여, 자연 언어 프로세싱 모듈(732)은 레스토랑 예약 도메인에 대한 부분 구조화된 쿼리를 생성하며, 여기서 부분 구조화된 쿼리는 파라미터들 {요리 = "스시"} 및 {시간 = "오후 7시"}을 포함한다. 그러나, 이 예에서, 사용자의 발화는 도메인과 연관된 구조화된 쿼리를 완성하기에 불충분한 정보를 포함한다. 따라서, {인원수} 및 {날짜}와 같은 기타 필수 파라미터들이 현재 이용가능한 정보에 기초하여 구조화된 쿼리에 특정되어 있지 않다. 일부 예들에서, 자연 언어 프로세싱 모듈(732)은 구조화된 쿼리의 일부 파라미터들에 수신된 컨텍스트 정보를 채운다. 예를 들어, 일부 예들에서, 사용자가 "내 근처"의 스시 레스토랑을 요청했다면, 자연 언어 프로세싱 모듈(732)은 구조화된 쿼리 내의 {위치} 파라미터에 사용자 디바이스로부터의 GPS 좌표들을 채운다.

[0240] 일부 예들에서, 자연 언어 프로세싱 모듈(732)은 STT 프로세싱 모듈(730)로부터 수신된 각각의 후보 텍스트 표현에 대해 다수의 후보 행동가능한 의도들을 식별한다. 또한, 일부 예들에서, 각각의 식별된 후보 행동가능한 의도에 대해 개개의 구조화된 쿼리(부분적인 또는 완전한)가 생성된다. 자연 언어 프로세싱 모듈(732)은 각각의 후보 행동가능한 의도에 대한 의도 신뢰도 점수를 결정하고, 의도 신뢰도 점수에 기초하여 후보 행동가능한 의도를 순위를 매긴다. 일부 예들에서, 자연 언어 프로세싱 모듈(732)은 임의의 완성된 파라미터들을 포함하는, 생성된 구조화된 쿼리(쿼리들)를 태스크 흐름 프로세싱 모듈(736)("태스크 흐름 프로세서")로 전달한다. 일부 예들에서, m-베스트(예를 들어, m개의 가장 높은 순위가 매겨짐) 후보 행동가능한 의도들에 대한 구조화된 쿼리(또는 쿼리들)는 태스크 흐름 프로세싱 모듈(736)에 제공되고, 여기서 m은 0보다 큰 미리결정된 정수이다. 일부 예들에서, m-베스트 후보 행동가능한 의도들에 대한 구조화된 쿼리(또는 쿼리들)는 대응하는 후보 텍스트 표현(들)을 갖는 태스크 흐름 프로세싱 모듈(736)에 제공된다.

[0241] 스피치 입력의 다수의 후보 텍스트 표현들로부터 결정된 다수의 후보 행동가능한 의도들에 기초하여 사용자 의도를 추론하는 다른 세부 사항은 2014년 6월 6일에 출원된 "System and Method for Inferring User Intent From Speech Inputs"에 대한 미국 특허 출원 제14/298,725호에 기술되고, 그 전체 개시내용이 본 명세서에 참조로서 편입되어 있다.

[0242] 태스크 흐름 프로세싱 모듈(736)은 자연 언어 프로세싱 모듈(732)로부터 구조화된 쿼리(또는 쿼리들)를 수신하도록, 필요하다면, 구조화된 쿼리를 완성하도록, 그리고 사용자의 궁극적인 요청을 "완성"하는 데 요구되는 행동들을 수행하도록 구성된다. 일부 예들에서, 이들 태스크들을 완수하는 데 필요한 다양한 절차들이 태스크 흐름 모델들(754)에서 제공된다. 일부 예들에서, 태스크 흐름 모델들(754)은 사용자로부터 추가 정보를 획득하기 위한 절차들, 및 행동가능한 의도와 연관된 행동들을 수행하기 위한 태스크 흐름들을 포함한다.

[0243] 상술한 바와 같이, 구조화된 쿼리를 완성하기 위해, 태스크 흐름 프로세싱 모듈(736)은 추가 정보를 획득하고/하거나 잠재적으로 모호한 발화들을 중의성을 해소하기 위해 사용자와의 추가 대화를 개시할 필요가 있다. 그와 같은 상호작용이 필요한 경우, 태스크 흐름 프로세싱 모듈(736)은 대화 흐름 프로세싱 모듈(734)을 호출하여 사용자와의 대화에 참여한다. 일부 예들에서, 대화 흐름 프로세싱 모듈(734)은 어떻게(및/또는 언제) 사용자에게 추가 정보를 물을지를 결정하고, 사용자 응답을 수신하고 프로세싱한다. 질문들은 I/O 프로세싱 모듈(728)을 통해 사용자들에게 제공되고 그들로부터 답변들이 수신된다. 일부 예들에서, 대화 흐름 프로세싱 모듈(734)은 오디오 및/또는 시각적 출력을 통해 사용자에게 대화 출력을 제시하고, 음성 또는 물리적(예컨대, 클릭킹) 응답들을 통해 사용자로부터 입력을 수신한다. 상기 예에 이어서, 태스크 흐름 프로세싱 모듈(736)이 대화 흐름 프로세싱 모듈(734)을 호출하여 도메인 "레스토랑 예약"과 연관된 구조화된 쿼리를 위한 "인원수" 및 "날짜" 정보를 결정하는 경우, 대화 흐름 프로세싱 모듈(734)은 "몇 명입니까?" 및 "어느 요일입니까?"와 같은 질문들을 생성하여 사용자에게 전달한다. 일단 사용자로부터 답변들이 수신되면, 대화 흐름 프로세싱 모듈(734)은 이어서 구조화된 쿼리에 누락 정보를 채우거나, 또는 태스크 흐름 프로세싱 모듈(736)에 정보를 전달하여 구조화된 쿼리로부터 누락 정보를 완성한다.

[0244] 일단 태스크 흐름 프로세싱 모듈(736)이 행동가능한 의도에 대한 구조화된 쿼리를 완성했다면, 태스크 흐름 프로세싱 모듈(736)은 행동가능한 의도와 연관된 궁극적인 태스크를 수행하도록 진행한다. 따라서, 태스크 흐름 프로세싱 모듈(736)은 구조화된 쿼리에 포함된 특정 파라미터들에 따라 태스크 흐름 모델에서 단계들 및 명령어들을 실행한다. 예를 들어, "레스토랑 예약"의 행동가능한 의도에 대한 태스크 흐름 모델은 레스토랑에 연락하기 위한 그리고 특정 시간의 특정 인원수에 대한 예약을 실제로 요청하기 위한 단계들 및 명령어들을 포함한다. 예를 들어, {레스토랑 예약, 레스토랑 = ABC 카페, 날짜 = 3/12/2012, 시간 = 오후 7시, 인원수 = 5}와 같은 구조화된 쿼리를 사용하여, 태스크 흐름 프로세싱 모듈(736)은, (1) OPENTABLE®과 같은 레스토랑 예약 시스템 또는 ABC 카페의 서버에 로그인하는 단계, (2) 웹사이트 상에서 일정 서식으로 날짜, 시간, 및 인원수 정보를 입력하는 단계, (3) 그 서식을 제출하는 단계, 및 (4) 사용자의 캘린더 내에 예약을 위한 캘린더 엔트리를 만드는

단계를 수행한다.

- [0245] 일부 예들에서, 태스크 흐름 프로세싱 모듈(736)은 사용자 입력에서 요청된 태스크를 완수하거나 또는 사용자 입력에서 요청된 정보제공형 답변을 제공하기 위해 서비스 프로세싱 모듈(738)("서비스 프로세싱 모듈")의 어시스턴트를 채용한다. 예를 들어, 서비스 프로세싱 모듈(738)은 태스크 흐름 프로세싱 모듈(736)을 대신해서 전화 통화하도록, 캘린더 엔트리를 설정하도록, 지도 검색을 호출하도록, 사용자 디바이스 상에 설치된 다른 사용자 애플리케이션들을 호출하거나 그들과 상호작용하도록, 그리고 제3자 서비스들(예를 들어, 레스토랑 예약 포탈, 소셜 네트워킹 웹사이트, 은행업무 포탈 등)을 호출하거나 그들과 상호작용하도록 동작한다. 일부 예들에서, 각각의 서비스에 의해 요구되는 프로토콜들 및 애플리케이션 프로그래밍 인터페이스(API)들은 서비스 모델들(756) 중 개개의 서비스 모델에 의해 특정된다. 서비스 프로세싱 모듈(738)은 서비스를 위한 적절한 서비스 모델에 액세스하고, 서비스 모델에 따른 서비스에 의해 요구되는 프로토콜들 및 API들에 따라 서비스에 대한 요청들을 생성한다.
- [0246] 예를 들어, 레스토랑이 온라인 예약 서비스를 가능하게 했다면, 레스토랑은 예약을 하는 데 필요한 파라미터들 및 온라인 예약 서비스에 필요한 파라미터의 값들을 전달하기 위한 API들을 특정하는 서비스 모델을 제출한다. 태스크 흐름 프로세싱 모듈(736)에 의해 요청될 때, 서비스 프로세싱 모듈(738)은 서비스 모델에 저장된 웹 주소를 사용하여 온라인 예약 서비스와의 네트워크 접속을 확립하고, 온라인 예약 서비스의 API에 따른 포맷으로 예약의 필요한 파라미터들(예컨대, 시간, 날짜, 인원수)을 온라인 예약 인터페이스에 전송한다.
- [0247] 일부 예들에서, 자연 언어 프로세싱 모듈(732), 대화 흐름 프로세싱 모듈(734), 및 태스크 흐름 프로세싱 모듈(736)은 사용자의 의도를 추론 및 정의하도록, 사용자 의도를 더 명확히 하고 정제하게 하기 위해 정보를 획득하도록, 그리고 최종적으로 사용자의 의도를 이행하기 위해 응답(즉, 사용자에게로의 출력, 또는 태스크의 완수)을 생성하도록 총체적이고 반복적으로 사용된다. 생성된 응답은 사용자의 의도를 적어도 부분적으로 이행하는 스피치 입력에 대한 대화 응답이다. 또한, 일부 예들에서, 생성된 응답은 스피치 출력으로서 출력된다. 이들 예에서, 생성된 응답은 스피치 합성 모듈(740)(예컨대, 스피치 합성기)에 전송되며, 여기서 스피치 합성 모듈(740)은 스피치 형태로 대화 응답을 합성하도록 프로세싱될 수 있다. 또 다른 예들에서, 생성된 응답은 스피치 입력에서 사용자 요청을 만족시키는 것에 관련된 데이터 콘텐츠이다.
- [0248] 태스크 흐름 프로세싱 모듈(736)이 자연 언어 프로세싱 모듈(732)로부터 다수의 구조화된 쿼리들을 수신하는 예들에서, 태스크 흐름 프로세싱 모듈(736)은 수신된 구조화된 쿼리들 중 제1 구조화된 쿼리를 초기에 프로세싱하여 제1 구조화된 쿼리를 완성하고/하거나 제1 구조화된 쿼리에 의해 표현된 하나 이상의 태스크들 또는 행동들을 실행하도록 시도한다. 일부 예들에서, 제1 구조화된 쿼리는 가장 높은 순위가 매겨진 행동가능한 의도에 대응한다. 다른 예들에서, 제1 구조화된 쿼리는 대응하는 스피치 인식 신뢰도 점수 및 대응하는 의도 신뢰도 점수의 조합에 기초하여 수신된 구조화된 쿼리들로부터 선택된다. 일부 예들에서, 태스크 흐름 프로세싱 모듈(736)은 (예를 들어, 필요한 파라미터를 결정할 수 없기 때문에) 제1 구조화된 쿼리의 프로세싱 동안 에러를 만나면, 태스크 흐름 프로세싱 모듈(736)은 보다 낮은 순위가 매겨진 행동가능한 의도에 대응하는 수신된 구조화된 쿼리들 중 제2 구조화된 쿼리를 선택하고 프로세싱하는 것을 진행할 수 있다. 제2 구조화된 쿼리는, 예를 들어, 대응하는 후보 텍스트 표현의 스피치 인식 신뢰도 점수, 대응하는 후보 행동가능한 의도의 의도 신뢰도 점수, 제1 구조화된 쿼리에서 누락된 필수 파라미터, 또는 이들의 임의의 조합에 기초하여 선택된다.
- [0249] 스피치 합성 모듈(740)은 사용자에게의 제시를 위한 스피치 출력을 합성하도록 구성된다. 스피치 합성 모듈(740)은 디지털 어시스턴트에 의해 제공된 텍스트에 기초하여 스피치 출력을 합성한다. 예를 들어, 생성된 대화 응답은 텍스트 문자열의 형태이다. 스피치 합성 모듈(740)은 텍스트 문자열을 가청 스피치 출력으로 변환한다. 스피치 합성 모듈(740)은, 텍스트로부터의 스피치 출력을 생성하기 위하여 임의의 적절한 스피치 합성 기술을 사용하는데, 이는 연접 합성(concatenative synthesis), 단위 선택 합성, 다이폰 합성, 도메인-특정 합성, 포먼트 합성(formant synthesis), 조음 합성(articulatory synthesis), 은닉 마르코프 모델(hidden Markov model, HMM) 기반 합성, 및 정현파 합성(sinewave synthesis)을 포함하지만 이로 한정되지 않는다. 일부 예들에서, 스피치 합성 모듈(740)은 단어들에 대응하는 음소 문자열에 기초하여 개별 단어들을 합성하도록 구성된다. 예를 들어, 음소 문자열은 생성된 대화 응답의 단어와 연관된다. 음소 문자열은 단어와 연관된 메타데이터에 저장된다. 스피치 합성 모듈(740)은 스피치 형태의 단어를 합성하기 위해 메타데이터 내의 음소 문자열을 직접 프로세싱하도록 구성된다.
- [0250] 일부 예들에서, 스피치 합성 모듈(740)을 사용하는 대신에(또는 그에 부가하여), 스피치 합성은 원격 디바이스(예컨대, 서버 시스템(108)) 상에서 수행되고, 합성된 스피치는 사용자에게의 출력을 위해 사용자 디바이스에

전송된다. 예를 들어, 이는 디지털 어시스턴트에 대한 출력이 서버 시스템에서 생성되는 일부 구현예들에서 발생할 수 있다. 그리고 서버 시스템이 일반적으로 사용자 디바이스보다 많은 프로세싱 능력 또는 리소스를 갖기 때문에, 클라이언트 측 합성에서 실제보다 높은 품질의 스피치 출력을 획득하는 것이 가능하다.

- [0251] 디지털 어시스턴트에 대한 추가적인 상세 사항들은 2011년 1월 10일자로 출원되고 발명의 명칭이 "Intelligent Automated Assistant"인 미국 특허 출원 제12/987,982호, 및 2011년 9월 30일자로 출원되고 발명의 명칭이 "Generating and Processing Task Items That Represent Tasks to Perform"인 미국 특허 출원 제13/251,088호에서 발견될 수 있고, 이 출원들의 전체 개시내용들은 본 명세서에 참조로서 편입된다.
- [0252] 4. 미디어 탐색을 위한 디지털 어시스턴트의 청각 기반 인터페이스를 제공하는 것에 대한 예시적인 기술
- [0253] 도 8a와 도 8b, 도 9a와 도 9b, 도 10a와 도 10b, 및 도 11은 일부 실시예들에 따른 디지털 어시스턴트를 제공하기 위한 예시적인 사용자 인터페이스("UI")를 포함하는 예시적인 기술들을 도시한다. 이 도면들은 또한 도 12 내지 도 15의 프로세스들(1200, 1300, 1400, 1500)을 각각 포함하는, 하기에 기술되는 프로세스들을 설명하는 데 사용된다.
- [0254] 도 8a는 전자 디바이스(800)를 도시한다. 전자 디바이스(800)는 일부 실시예들에서 디바이스들(200, 400, 600)(도 2a, 도 4, 도 6a와 도 6b) 중 임의의 것일 수 있다. 도시된 예에서, 전자 디바이스(800)는 하나 이상의 스피커들을 구비한 전자 디바이스이지만, 전자 디바이스는 전화, 랩톱 컴퓨터, 데스크톱 컴퓨터, 태블릿, 웨어러블 디바이스(예컨대, 스마트 워치), 셋톱 박스, 텔레비전, 스피커, 또는 이들의 임의의 조합 또는 하위 조합과 같은 임의의 유형의 디바이스일 수 있다는 것이 이해될 것이다.
- [0255] 동작 시에, 전자 디바이스(800)는 사용자와 지능형 자동화 어시스턴트(또는 디지털 어시스턴트) 간의 자연 언어 스피치의 교환을 제공한다. 일부 예들에서 교환은 순전히 청각적이다. 일부 예들에서, 교환은 추가적으로 또는 대안적으로 (예를 들어, 그래픽 사용자 인터페이스 및/또는 하나 이상의 광 표시자들에 의해) 시각적 및/또는 촉각적이다.
- [0256] 도 8a를 참조하면, 전자 디바이스(800)는 전자 디바이스(800)의 디지털 어시스턴트에 대한 요청을 나타내는 자연 언어 스피치 입력(810)을 (예를 들어, 마이크를 통해) 수신한다. 자연 언어 스피치 입력(810)은 디지털 어시스턴트로 지향될 수 있는 임의의 요청을 포함할 수 있다. 일부 예들에서, 자연 언어 스피치 입력은 미리결정된 트리거 구절(예를 들어, "시리야")을 포함한다. 일부 예들에서, 자연 언어 스피치 입력은 미디어 아이템들에 대한 요청을 포함한다(예를 들어, "리치 루빈(Rich Rubin)이 프로듀싱한 음악을 재생해", "랩 노래를 재생해", "80년대 뭔가를 재생해", "신나는 뭔가를 재생해"). 도 8a를 참조하면, 사용자(802)는 트리거 구절 및 특정 아티스트의 미디어 아이템들에 대한 요청을 포함하는 자연 언어 스피치 입력(810)을 제공한다: "시리야, 아델(Adele)의 새로운 노래를 재생해."
- [0257] 일부 예들에서, 전자 디바이스(800)는 하나 이상의 태스크들을 수행하기 위해 자연 언어 스피치 입력(810)을 프로세싱한다. 일부 예들에서, 이러한 방식으로 자연 언어 스피치 입력(810)을 프로세싱하는 것은, 예를 들어 STT 프로세싱 모듈(730)을 사용하여 자연 언어 스피치 입력의 하나 이상의 후보 텍스트 표현들(예를 들어, 텍스트 문자열들)을 제공하는 것을 포함한다. 기술된 바와 같이, 각각의 후보 텍스트 표현은 스피치 인식 신뢰도 점수와 연관될 수 있고, 후보 텍스트 표현들은 그에 따라 순위가 매겨질 수 있다. 다른 예들에서, 자연 언어 입력은 텍스트 입력(예를 들어, 전자 디바이스(800)의 터치패드를 통해 입력됨)이고 후보 텍스트 표현으로서 제공되며, 여기서 $n=1$ 이다. 이러한 방식으로 후보 텍스트 표현으로서 제공되는 텍스트 입력은 최대 스피치 인식 신뢰도 점수, 또는 임의의 다른 스피치 인식 신뢰도 점수를 할당받을 수 있다. 도 8a를 참조하면, 디지털 어시스턴트는 후보 텍스트 표현 "시리야, 아델의 새로운 노래를 재생해"를 포함하는 하나 이상의 후보 텍스트 표현들을 제공한다.
- [0258] 일부 예들에서, 전자 디바이스(800)는 예를 들어 자연 언어 프로세싱 모듈(732)을 사용하여 n -베스트(예를 들어, 가장 높은 순위가 매겨짐) 후보 텍스트 표현들에 기초하여 하나 이상의 후보 의도들을 제공한다. 후보 의도들 각각은 의도 신뢰도 점수와 연관될 수 있으며, 후보 의도들은 그에 따라 순위가 매겨질 수 있다. 일부 예들에서, 다수의 후보 의도들이 각각의 후보 텍스트 표현에 대해 식별된다. 또한, 일부 예들에서, 하나 이상의 파라미터들을 갖는 구조화된 쿼리(부분적인 또는 완전한)가 각각의 후보 의도에 대해 생성된다. 도 8a를 참조하면, 전자 디바이스(800)의 디지털 어시스턴트는 후보 텍스트 표현 "시리야, 아델의 새로운 노래를 재생해"에 기초한 "미디어 아이템들에 대한 추천들을 획득하려는" 후보 의도를 포함하는 하나 이상의 후보 의도들을 제공한다. 또한, 전자 디바이스(800)의 디지털 어시스턴트는 다수의 파라미터들을 갖는 구조화된 쿼리를 결정한다.

다: {미디어 아이템들에 대한 추천들을 획득하는 것, 아티스트 = 아델, 미디어 유형 = 노래, 기간 = 새로운}.

- [0259] 그 후, 후보 태스크는 예를 들어 태스크 흐름 프로세싱 모듈(736)을 사용하여 m-베스트(예를 들어, 가장 높은 순위가 매겨짐) 후보 의도들에 기초하여 결정된다. 일부 예들에서, 후보 태스크들은 각각의 m-베스트(예를 들어, 가장 높은 순위가 매겨짐) 후보 의도들에 대한 구조화된 쿼리에 기초하여 식별된다. 예로서, 기술된 바와 같이, 구조화된 쿼리는 하나 이상의 태스크 흐름들(754)과 같은 하나 이상의 태스크 흐름들에 따라 구현될 수 있다.
- [0260] 일부 예들에서, 전자 디바이스(800)는 하나 이상의 결과들을 획득하기 위해 식별된 파라미터에 기초하여 후보 태스크를 수행한다. 예를 들어, 구조화된 쿼리에 기초하여, 전자 디바이스(800)의 태스크 흐름 프로세싱 모듈(예컨대, 태스크 흐름 프로세싱 모듈(736))은 하나 이상의 결과들을 획득하기 위해 프로그램들, 메소드들, 서비스들, API들 등을 호출한다. 결과는, 예를 들어, 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 임의의 조합을 포함하지만 이로 한정되지 않는 하나 이상의 미디어 아이템들에 관한 정보를 포함할 수 있다. 도 8a를 참조하면, 구조화된 쿼리 {미디어 아이템들에 대한 추천들을 획득하는 것, 아티스트 = 아델, 미디어 유형 = 노래, 기간 = 새로운}에 기초하여, 디지털 어시스턴트는 검색 파라미터들 "아델", "노래", 및 "새로운"을 사용하여 미디어 검색을 수행하고 "Hello"라는 제목의 노래(이하, "제1 미디어 아이템")를 식별한다.
- [0261] 그 후, 전자 디바이스(800)는 제1 미디어 아이템을 제공한다. 도 8a를 참조하면, 전자 디바이스(800)의 디지털 어시스턴트는 노래 "Hello"의 재생(812)을 제공한다. 도시된 바와 같이, 디지털 어시스턴트는 또한 제1 미디어 아이템의 설명(예를 들어, 구두 설명)("아델의 Hello입니다")을 포함하는 자연 언어 스피치 출력(813)을 제공하는 동안 제1 미디어 아이템의 재생을 제공한다. 제1 미디어 아이템의 재생 및 제1 미디어 아이템의 설명은 일부 예들에서 동시에 제공될 수 있다.
- [0262] 일부 예들에서, 전자 디바이스(800)는 자연 언어 스피치 입력(810)에 응답하여 제1 미디어 아이템의 일부분의 재생을 제공한다. 제1 미디어 아이템의 일부분은 미디어 아이템의 대표 샘플(예를 들어, 코러스, 제1 절)일 수 있다. 일부 예들에서, 디지털 어시스턴트는 제1 미디어 아이템과 연관된 설명을 나타내는 스피치 출력(예를 들어, "아델의 Hello입니다")을 제공하는 동안 일부분(예를 들어, 코러스)의 재생을 제공한다. 사용자가 스피치 출력에 긍정 응답(예를 들어, 자연 언어 응답)을 제공하면(예를 들어, "그래 이것을 재생해"), 전자 디바이스(800)는 제1 미디어 아이템의 전체(예를 들어, 처음부터)를 재생할 수 있다. 계층화된 오디오 출력을 제공하는 메커니즘에 대한 자세한 내용은 본 명세서에 제공된다.
- [0263] 일부 예들에서, 전자 디바이스(800)는 자연 언어 스피치 입력(810)에 기초하여 식별된 다수의 미디어 아이템들의 요약 및/또는 목록을 제공한다(예를 들어, "당신은 이 아티스트로부터 많은 것을 갖고 있습니다. 우선 10곡 중 3곡입니다: Hello, ..."). 일부 예들에서, 전자 디바이스(800)는 목록을 제공하기 전에 하나 이상의 제안들(예를 들어, "뭔가 듣고 싶은 게 있거나 다음 5곡을 듣고 싶으시면 알려주세요")을 제공한다. 일부 예들에서, 전자 디바이스는 미리결정된 횟수만큼 제안을 제시한 후에 제안을 제공하는 것을 중지한다. 예를 들어, 전자 디바이스는 동일한 미디어 요청에 대하여 그것을 3회 제공한 후 "뭔가 듣고 싶은 게 있으시면 알려주세요"라는 제안을 제공하는 것을 중지할 수 있다. 미디어 추천을 제공하는 추가적 설명은 2016년 9월 23일자로 출원된 미국 특허 출원 제62/399,232호, "INTELLIGENT AUTOMATED ASSISTANT"에 제공되며, 이는 본 명세서에 그 전문이 참조로서 편입된다. 미국 특허 출원 제62/399,232호는 특히, 미디어 콘텐츠를 제시하기 위한 복수의 포맷들(예를 들어, 앨범 명칭, 노래 명칭 등)을 출력하기 위한 예시적인 기술을 기술한다.
- [0264] 전자 디바이스(800)는 미디어 아이템의 재생을 제공하는 동안 자연 언어 스피치 입력을 수신할 수 있다. 도 8a를 참조하면, 노래 "Hello"의 재생을 제공하는 동안, 전자 디바이스(800)는 자연 언어 스피치 입력(814)을 수신한다("사실, 'I'm giving you up I'm forgiving it all'로 되는 것을 재생해"). 일부 예들에서, 자연 언어 스피치 입력(814)을 수신하는 것에 응답하여, 전자 디바이스는 제1 미디어 아이템의 현재 재생이 제공되는 방식을 조정한다(예를 들어, 낮은 볼륨 또는 속도로 재생을 제공하는 것). 오디오 출력을 조정하는 추가의 예시적인 설명은 2016년 9월 23일자로 출원된 미국 특허 출원 제62/399,232호, "INTELLIGENT AUTOMATED ASSISTANT"에 제공되며, 이는 본 명세서에 그 전문이 참조로서 편입된다. 미국 특허 출원 제62/399,232호는 특히, 오디오 출력의 볼륨을 낮추거나/높이는 것을 포함하는 오디오 출력을 조정하고, 오디오 출력을 상이한 속도로 제공하고/하거나 상이한 언어로 오디오를 입력을 제공하기 위한 예시적인 기술을 기술한다. 일부 예들에서, 오디오 출력은 오디오 출력과 연관된 사용자 입력에 응답하여 조정된다.
- [0265] 전자 디바이스(800)는 자연 언어 입력(812)에 대하여 상술한 것과 일치하는 방식으로 자연 언어 입력(814)을 프로세싱한다. 구체적으로, 자연 언어 입력(814)에 기초하여, 전자 디바이스(800)는 하나 이상의 후보 텍스트 표

현들, 하나 이상의 후보 의도들을 제공하고, 가장 높은 순위가 매겨진 후보 의도와 연관된 태스크를 수행한다. 도시된 예에서, 자연 언어 스피치 입력(814)("사실, 'I'm giving you up I'm forgiving it all'로 되는 것을 재생해")에 대응하는 하나의 후보 의도는 "미디어에 대한 요청을 정제하는 것"이다. 일부 예들에서, 후보 의도는 m-베스트 후보 의도들 중 하나이다.

[0266] 일부 예들에서, 전자 디바이스(800)는 하나 이상의 미리정의된 구절들 및 하나 이상의 구절들의 자연 언어 등가물들에 기초하여 미디어에 대한 요청을 정제하려는 사용자 의도를 도출한다. 예시적인 미리정의된 구절은, "그래, 그러나", "어떤", "어때", "만", "그 밖의", "다른", "더", "덜", "뭔가 더", "뭔가 덜", "새로운 것", "오래된 것", "되는 것", "같은 소리의 것" "사실", "기다러", "재생해", "아니", "상이한", "건너뛰어", 및 "다음"을 포함하지만, 이로 한정되지 않는다. 이와 같이, 전자 디바이스(800)는 "아니, 그거 말고 뭐가 있어?", "좀 더 신나는 것을 재생해", "다른 것을 재생해", "80년대에 프로듀싱된 것만"과 같은 예시적인 입력을 프로세싱할 수 있다. 도 8a를 참조하면, 전자 디바이스(800)는 자연 언어 스피치 입력(814)의 미리정의된 구절들(예를 들어, "사실", "재생해", "로 되는 것")에 기초하여 미디어 요청을 정제하려는 사용자 의도를 도출할 수 있다. 자연 언어 입력(814)을 프로세싱하기 위한 예시적인 기술은 자연 언어 프로세싱 모듈(732)에 대하여 상술된다. 예를 들어, 전자 디바이스(800)는 자연 언어 입력(814)의 후보 텍스트 표현을 수신하고(텍스트 문자열 "사실 I'm giving you up I'm forgiving it all로 되는 것을 재생해")(예를 들어, STT 프로세싱 모듈(730)로부터), 디지털 어시스턴트의 온톨로지(예를 들어, 온톨로지(760)) 내의 어떤 노드가 후보 텍스트 표현 내의 단어들에 의해 연루되는지를 결정할 수 있다. 활성화된 노드들의 양 및/또는 상대적 중요도에 기초하여, 전자 디바이스(보다 상세하게는, 자연 언어 프로세싱 모듈)는 사용자가 디지털 어시스턴트로 하여금 수행하도록 의도했던 태스크로서 행동가능한 의도들 중 하나를 선택할 수 있다.

[0267] 일부 예들에서, 전자 디바이스(800)는 컨텍스트 정보에 기초하여 미디어에 대한 요청을 정제하려는 사용자 의도를 도출한다. 컨텍스트 정보는 전자 디바이스와의 하나 이상의 이전 사용자 상호작용들(예를 들어, 사용자 세션들)을 포함한다. 예를 들어, 사용자의 이전 요청이 미디어 추천을 획득하려는 사용자 의도(예를 들어, 스피치 입력(812))와 연관되는 경우 그리고/또는 사용자의 현재 입력이 미디어 추천 도메인 내의 하나 이상의 속성들(예를 들어, 속성 노드들)에 대응하는 경우, 전자 디바이스(800)는 (예를 들어, 하나 이상의 특정 속성들을 사용하여) 이전 미디어 요청을 정제하려는 사용자 의도를 도출할 수 있다. 미디어 추천 도메인 내의 속성은 하기에 기술되는 아티스트, 장르, 가사, 출시 날짜, 또는 검색 파라미터들 중 임의의 것에 대응할 수 있다. 도 8a를 참조하면, 전자 디바이스(800)는 사용자의 이전 스피치 입력(912) 및/또는 현재 스피치 입력에서 특정된 속성(가사 "I'm giving you up I'm forgiving it all")에 기초하여 미디어 요청을 정제하려는 사용자 의도를 도출할 수 있다.

[0268] 일부 예들에서, 전자 디바이스(800)는 자연 언어 스피치 입력(814)에 기초하여 하나 이상의 후보 태스크들 및 대응하는 파라미터들을 식별한다. 도 8a를 참조하면, 전자 디바이스(800)는 "이전 미디어 요청을 정제하는 것"의 후보 태스크 및 이전 미디어 요청을 정제하기 위한 "I'm giving you up I'm forgiving it all"의 파라미터를 식별한다.

[0269] 자연 언어 스피치 입력에 기초하여 식별된 파라미터들은 미디어 요청을 정제하는 데 사용될 수 있다. 예시적인 파라미터들이 본 명세서에 제공된다. 일부 예들에서, 파라미터들은 미디어 아이템의 가사 내용(예를 들어, "Hey Jude"), 장르(예를 들어, "힙합"), 노래 또는 앨범 제목(예를 들어, "Hotel California"), 행사 또는 기간(예를 들어, 계절, 휴일, 하루 중 언제, 10년), 활동(예를 들어, 운동, 운전, 수면), 위치(예를 들어, 해변, 직장, 집, 하와이), 기분(예를 들어, 신나는), 아티스트(예컨대, 가수, 프로듀서), 또는 이들의 조합에 대응한다.

[0270] 일부 예들에서, 파라미터들은 미리결정된 시간 프레임 내의 날짜(예를 들어, 출시 날짜)에 대응한다. 예를 들어, 전자 디바이스(800)는 구절(및 이들 구절의 자연 언어 등가물)과 시간 프레임 간의 상관 관계를 저장한다. 예를 들어, 전자 디바이스(800)는 1개월의 시간 프레임을 갖는 "새로운", 3개월의 시간 프레임을 갖는 "최근", 1주일의 시간 프레임을 갖는 "방금 나온" 및 "최신"을 상관시킨다.

[0271] 일부 예들에서, 파라미터들은 하나 이상의 사람들(예를 들어, 의도된 청중)에 대응한다. 예를 들어, 자연 언어 스피치 입력(814)은, "내 친구들은 뭘 듣고 있어?", "제이슨(Jason)은 뭘 재생하고 있어?", "에이미(Amy)로부터 더 많은 음악을 재생해", 내 친구들이 좋아하는 뭔가를 재생해"와 같은 사용자와 연관된 사람들에 대한 언급을 포함할 수 있다. 전자 디바이스(800)는 자연 언어 스피치 입력을 프로세싱하여 사용자를 언급하는 단어 또는 구절(예를 들어, "나", "나를 위해", "내가", "나의"), 사용자 이외의 사람들(예를 들어, "에이미"), 또는 이들

의 조합(예를 들어, "우리", "내 친구와 나")을 식별한다. 이러한 단어들 또는 구절들에 기초하여, 전자 디바이스(800)는 하나 이상의 소스들(예를 들어, 연락처 목록, 소셜 미디어 서비스 및 미디어 서비스와 같은 소프트웨어 서비스들)로부터 식별 정보를 획득한다. 일부 예들에서, 전자 디바이스(800)는 사용자에게 후보 해석들 사이의 중의성을 해소하도록 촉구함으로써(예를 들어, "존 스미스(John Smith) 또는 존 도우(John Doe)를 의미합니까?") 식별 정보를 획득한다. 일부 다른 예들에서, 전자 디바이스(800)는 전자 디바이스 근처의 하나 이상의 사람들의 물리적 존재와 같은 컨텍스트 정보에 기초하여 식별 정보를 획득한다. 하나 이상의 사람들의 물리적 존재를 검출하기 위한 기술은 아래에서 자세히 설명한다.

[0272] 일부 예들에서, 파라미터들은 미디어 아이템의 소스에 대응한다. 예를 들어, 자연 언어 스피치 입력(814)은 미디어 아이템들의 컬렉션에 대한 언급(예를 들어, "내 라이브러리에 뭐가 있어?", "주말 잼 리스트(jam list)로부터 뭔가를 재생해")을 포함할 수 있다. 다른 예로서, 자연 언어 스피치 입력(814)은 미디어 아이템들의 주인에 대한 언급(예를 들어, "제이슨의 컬렉션으로부터 뭔가를 재생해")을 포함할 수 있다. 이에 응답하여, 전자 디바이스(800)는 식별 정보를 획득하고, 하기에 보다 상세히 설명되는 바와 같이, 적절한 허가 설정을 갖는 하나 이상의 미디어 아이템들을 추가로 식별한다.

[0273] 일부 예들에서, 전자 디바이스(800)는 컨텍스트 정보에 적어도 부분적으로 기초하여 미디어 요청을 정제하기 위한 파라미터들을 식별한다. 상기 논의된 바와 같이, 컨텍스트 정보(context information)(또는 컨텍스트 정보(contextual information))는 전자 디바이스(800)의 환경, 예를 들어, 조명, 주변 소음, 주변 온도, 주위 환경의 이미지 또는 비디오 등과 연관된 정보를 포함할 수 있다. 일부 예들에서, 컨텍스트 정보는 전자 디바이스(800)의 물리적 상태, 예를 들어, 디바이스 배향, 디바이스 위치, 디바이스 온도, 전력 레벨, 속도, 가속도, 모션 패턴, 셀룰러 신호 강도 등을 포함한다. 디바이스 위치는 절대적(예를 들어, GPS 좌표에 기초하여) 또는 상대적(예를 들어, 디바이스가 사용자의 거실, 차고, 침실에 있음)일 수 있다. 일부 예들에서, 컨텍스트 정보는 전자 디바이스에서의 현재 시간을 포함한다. 일부 예들에서, 컨텍스트 정보는 실행 중인 프로세스, 설치된 프로그램, 과거 및 현재 네트워크 활동, 백그라운드 서비스, 에러 로그, 리소스 사용 등과 같은 디지털 어시스턴트 서버(예컨대, DA 서버(106))의 상태, 및 전자 디바이스(800)의 상태와 연관된 정보를 포함한다.

[0274] 일부 예들에서, 컨텍스트 정보는 전자 디바이스(800)에 물리적으로 근접한 사람들의 아이덴티티를 포함한다. 일부 예들에서, 전자 디바이스(800)는 하나 이상의 소스들로부터 정보를 획득하고 그 정보를 하나 이상의 사용자들에 관한 알려진 정보와 비교하여 하나 이상의 식별들을 생성함으로써 하나 이상의 사용자들의 물리적 존재 및/또는 아이덴티티를 검출할 수 있다. 예를 들어, 전자 디바이스(800)는 사람의 전자 디바이스의 접속성 정보(예를 들어, 동일한 Wi-Fi 네트워크 상에서, 블루투스 범위 내에서, NFC 범위 내에서)와 같은, 사람과 연관된 전자 디바이스와 관련된 정보에 기초하여 사람의 물리적 존재를 검출할 수 있다. 다른 예로서, 전자 디바이스(800)는 (예를 들어, 카메라 및 마이크로폰을 통해 캡처된) 사람의 얼굴 특성 및/또는 음성 특성에 기초하여 사람의 물리적 존재를 검출할 수 있다. 다른 예로서, 전자 디바이스(800)는 (현재 시간에 대해 설정된) 캘린더 초청 또는 이메일 메시지에 열거된 연락처들과 같이 국부적으로 이용가능한 정보에 기초하여 사람의 물리적 존재를 검출할 수 있다. 또 다른 예로서, 전자 디바이스(800)는 사람에 의해 제공된 크리덴셜(예를 들어, 사용자 이름 및 패스워드)에 기초하여 사람의 물리적 존재를 검출할 수 있다. 일부 예들에서, 전자 디바이스(800)는 사람의 물리적 존재를 검출한 후에 중의성 해소 입력(예를 들어, "내가 듣고 있는 게 그 제이슨 또는 존입니까?") 및/또는 확인("존이 지금 막 파티에 참석했습니까?")을 촉구한다.

[0275] 일부 예들에서, 컨텍스트 정보는 전자 디바이스(800)와 물리적으로 근접한 사람들에 관련된 정보를 포함한다. 예를 들어, 컨텍스트 정보는 선호도, 미디어 컬렉션, 전자 디바이스에 물리적으로 근접한 것으로 검출된 사람들 각각의 이력을 포함할 수 있다. 예를 들어, 전자 디바이스(800)가 사용자의 친구인 에이미가 "내가 또한 좋아하는 뭔가를 재생해"라고 발언했다고 결정하는 경우, 전자 디바이스는 하나 이상의 소스들로부터 에이미의 선호도(좋아하는 장르, 분명한 언어 설정)를 식별하고 미디어 요청을 정제하기 위한 검색 파라미터로서 선호도들을 사용한다. 다수의 사람들에게 대해 병합된 선호도 프로파일을 제공하는 것에 대한 추가 정보는 하기에 제공된다.

[0276] 일부 예들에서, 컨텍스트 정보는 제1 미디어 아이템에 관련된 정보를 포함한다. 예를 들어, 사용자가 제1 미디어 아이템의 추천에 응답하여 "이보다 더 최근의 것을 재생해"라고 발언하는 경우, 전자 디바이스(800)는 제1 미디어 아이템의 출시 날짜에 기초하여 시간 파라미터를 도출한다.

[0277] 미디어 요청을 정제하기 위한 상술한 파라미터는 단지 예시적인 것임이 이해될 것이다. 또한, 전자 디바이스는, 전자 디바이스가 원래의 미디어 요청을 프로세싱하고 있을 때 및/또는 전자 디바이스가 원래의 미디어 요청에 기초하여 하나 이상의 미디어 아이템들을 제공하고(예를 들어, 그에 관련된 정보 또는 그의 재생을

제공하고) 있을 때 언제든지 미디어 요청을 정제하기 위한 사용자 요청을 수신할 수 있음이 추가로 이해될 것이다. 미디어 요청을 정제하기 위해 자연 언어 스피치 입력(예를 들어, 자연 언어 스피치 입력(814))의 사용은 단지 예시적인 것임을 추가로 인식해야 한다. 일부 예들에서, 전자 디바이스(800)는 전자 디바이스의 하나 이상의 센서들(예를 들어, 촉각적 입력, 제스처 입력, 버튼 누름)을 통해 입력을 수신하는 것에 응답하여 미디어 요청을 정제하고/하거나 추가 미디어 아이템을 제공하기 위한 프로세스를 개시할 수 있다. 미디어 검색을 수행하는 추가의 예시적인 설명은 2016년 6월 8일자로 출원된 미국 특허 출원 제62/347,480호, "INTELLIGENT AUTOMATED ASSISTANT FOR MEDIA EXPLORATION"에 제공되며, 이는 본 명세서에 그 전문이 참조로서 편입된다. 미국 특허 출원 제62/347,480호는 특히, 스피치 입력이 미디어 아이템들에 대한 개인화된 추천들을 획득하려는 사용자 의도에 대응하는지 여부를 결정하는 예시적인 기술을 기술한다. 이러한 결정에 응답하여, 일부 예들에서, 미디어 아이템들의 사용자-특정 코퍼스로부터 적어도 하나의 미디어 아이템이 획득된다. 컨텍스트 정보를 획득하는 추가의 예시적인 설명은 2017년 5월 16일자로 출원된 미국 특허 출원 제62/507,056호, "PROVIDING AN AUDITORY-BASED INTERFACE OF A DIGITAL ASSISTANT"에 제공되며, 이는 본 명세서에 그 전문이 참조로서 편입된다. 미국 특허 출원 제62/507,056호는 특히, 자연 언어 스피치 입력을 수신하기 전에, 도중에, 또는 후에 컨텍스트 정보를 획득하기 위한 예시적인 기술을 기술한다. 획득한 컨텍스트 정보는 일부 예들에서 사용자 특정 정보 및 전자 디바이스의 물리적 상태를 포함한다.

[0278] 자연 언어 스피치 입력(814)이 이전의 미디어 요청을 정제하려는 사용자 의도(예를 들어, m-베스트 후보 의도들 중 하나)에 대응한다고 결정한 후에, 전자 디바이스는 제1 미디어 아이템과 상이한 제2 미디어 아이템을 식별한다. 제2 미디어 아이템은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 임의의 조합일 수 있다. 도 8a를 참조하면, 전자 디바이스(800)는 스피치 입력(812) 내의 파라미터들("아델", "새로운", "노래") 및 스피치 입력(814) 내의 파라미터들("I've giving you up I'm forgiving it all")에 기초하여 제2 미디어 아이템(예를 들어, "Send My Love")을 식별한다.

[0279] 일부 예들에서, 자연 언어 스피치 입력(812)("시리야, 아델의 새로운 노래를 재생해")에 기초하여, 전자 디바이스(800)는 미디어 아이템들의 제1 세트(예를 들어, 지난 3개월 간 발매된 아델의 노래들의 세트)를 식별한다. 미디어 아이템들의 제1 세트로부터, 전자 디바이스는 사용자에게 제공하기 위해 노래 "Hello"를 선택한다(예를 들어, 인기 순위에 기초하여). 그 후, 후속하는 자연 언어 스피치 입력(814)에 기초하여, 전자 디바이스(800)는 자연 언어 스피치 입력(814)으로부터 도출된 특정 파라미터에 기초하여 미디어 아이템들의 제1 세트의 서브세트를 식별한다(예를 들어, 노래들의 제1 세트로부터 가사 "I'm giving you up I'm forgiving it all"를 포함하는 노래들만). 일부 예들에서, 미디어 아이템들의 제1 세트의 서브세트를 식별하는 것은 제1 세트의 미디어 아이템이 자연 언어 입력(814) 내의 특정 파라미터와 매칭되는 콘텐츠(예를 들어, 가사, 스크립트) 또는 메타데이터(장르, 출시 날짜)와 연관되는지 여부를 결정하는 것을 포함한다. 그렇다면, 이어서, 전자 디바이스(800)는 서브세트로부터 제2 미디어 아이템(노래 "Send My Love")을 선택하여 사용자에게 제공한다. 그렇지 않은 경우, 전자 디바이스(800)는 사용자에게 제공하기 위해 제2 미디어 아이템을 선택하는 것을 중지한다.

[0280] 일부 예들에서, 전자 디바이스는 미디어 아이템들의 사용자-특정 코퍼스로부터 제1 미디어 아이템 및/또는 제2 미디어 아이템을 식별한다. 일부 예들에서, 전자 디바이스(800)는 사용자 입력(예를 들어, 자연 언어 스피치 입력(814))과 연관된 음향 정보에 기초하여 사용자-특정 코퍼스를 식별한다. 사용자-특정 코퍼스는 사용자와 연관된 데이터(예컨대, 선호도, 설정, 이전 요청, 이전 선택, 이전 거부, 이전 사용자 구매, 사용자-특정 재생 목록)에 기초하여 생성된다. 일부 예들에서, 사용자-특정 코퍼스의 적어도 일부는 소프트웨어 서비스(예를 들어, 미디어 서비스 또는 소셜 미디어 서비스)에 기초하여 생성된다. 예를 들어, 사용자-특정 코퍼스는, 사용자가 이전에 거부하거나 싫어한 미디어 아이템들을 낮은 순위들과 연관시키거나(예를 들어, 소프트웨어 서비스 상에서), 이들 미디어 아이템들을 포함하지 않는다. 다른 예로서, 사용자-특정 코퍼스는 소프트웨어 서비스 상에서 사용자에게 의해 소유/구매된 미디어 아이템에 대응하는 데이터를 포함한다. 또 다른 예로서, 사용자-특정 코퍼스는 소프트웨어 서비스 상에서 사용자에게 의해 생성된 미디어 아이템(예를 들어, 재생목록)에 대응하는 데이터를 포함한다. 상기 논의된 바와 같이, 전자 디바이스는 사용자-특정 코퍼스 내의 미디어 아이템이 특정 검색 파라미터와 매칭되는 메타데이터 또는 콘텐츠와 연관되는지 여부를 결정함으로써 미디어 아이템을 식별할 수 있다. 일부 예들에서, 사용자-특정 코퍼스 내의 적어도 하나의 미디어 아이템은, 활동(예를 들어, 운동, 수면); 기분(예를 들어, 신나는, 차분한, 슬픈); 행사(예를 들어, 생일); 기간(예를 들어, 80년대), 위치; 큐레이터(예를 들어, 롤링 스톤스(Rolling stones) 목록); 컬렉션(예를 들어, 여름 재생목록); 하나 이상의 이전 사용자 입력들(사용자에게 의한 이전 거부, 사용자에게 의한 이전 좋아함); 또는 이들의 임의의 조합을 나타내는 메타데이터를 포함한다. 사용자-특정 코퍼스에 대한 추가의 예시적인 설명은 2016년 6월 8일자로 출원된 미국 특허 출원 제62/347,480호, "INTELLIGENT AUTOMATED ASSISTANT FOR MEDIA EXPLORATION"에 제공되며, 이는 본 명세서에 그

전문이 참조로서 편입된다.

- [0281] 일부 예들에서, 사용자-특정 코퍼스 내의 적어도 하나의 미디어 아이템은 미디어 요청을 제공한 사용자와 상이한 사람으로부터의 정보에 기초한 메타데이터를 포함한다. 예를 들어, 미디어 아이템은 해변에 대응하는 위치들에서 소프트웨어 애플리케이션(예를 들어, 아이튠즈(iTunes)와 같은 미디어 서비스)의 모든 사용자가 재생하는 빈도에 기초하여 "해변" 위치와 연관될 수 있다. 다른 예로서, 미디어 아이템은 사용자의 친구들에 의해(즉, 소셜 미디어 서비스의 사용자와 연관됨) 및/또는 비슷한 인구 통계학적 세그먼트로부터 사람들에 의해 재생된 횟수에 기초하여 활동(예컨대, 파티 참석)과 연관될 수 있다. 일부 예들에서, 메타데이터는 전자 디바이스(800)와 상이한 원격 디바이스에서 생성된다. 일부 예들에서, 사용자-특정 코퍼스 내의 적어도 하나의 미디어 아이템은, 보다 상세히 논의되는 바와 같이, 사용자가 액세스하도록 인가되지 않지만(예를 들어, 구매하지 않은), 전자 디바이스(800)에 물리적으로 근접한 다른 사람에 인가된 미디어 아이템이다.
- [0282] 그 후, 전자 디바이스(800)는 제2 미디어 아이템을 제공한다. 일부 예들에서, 제2 미디어 아이템은 제1 미디어 아이템에 대하여 상술된 것과 일치하는 방식으로 제공된다. 도 8a를 참조하면, 전자 디바이스(800)의 디지털 어시스턴트는 노래 "Send My Love"의 재생(816)을 제공한다. 도시된 바와 같이, 디지털 어시스턴트는 또한 예를 들어 제2 미디어 아이템의 재생을 제공하는 동안 제2 미디어 아이템의 설명을 포함하는 자연 언어 스피치 출력(817)("Send My Love입니다")을 제공한다. 일부 예들(도시되지 않음)에서, 전자 디바이스(800)는 자연 언어 스피치 입력(814)에 응답하여 제2 미디어 아이템의 대표 샘플의 재생을 제공하고 그 전체에서 제2 미디어 아이템을 제공하기 전에 사용자 확인을 요구한다. 일부 예들에서, 전자 디바이스(800)는 자연 언어 스피치 입력(814)에 기초하여 식별된 다수의 미디어 아이템들의 요약 및/또는 목록을 제공한다(예를 들어, "이러한 가사들을 갖는 노래 두 곡을 발견했습니다: Send My Love, Send My Love Acoustic Version...").
- [0283] 일부 예들에서, 도 8b를 참조하면, 전자 디바이스(800)는 제3 자연 언어 스피치 입력(818)을 수신한다("시리야, 내 토요일 오전 재생목록에 이것을 추가해"). 제3 자연 언어 스피치 입력(818)에 기초하여, 전자 디바이스는 미디어 아이템을 미디어 아이템들의 컬렉션과 연관시키려는 사용자 의도를 결정한다. 전자 디바이스는 컨텍스트 정보(예를 들어, 현재/이전에 재생된 미디어 아이템들)에 기초하여 사용자 의도를 결정할 수 있다. 도시된 예에서, 전자 디바이스(800)는 현재 재생된 노래 "Send My Love"를 "토요일 오전"이라는 재생목록에 연관시키고 연관성을 나타내는 스피치 출력(820)("완료했습니다")을 제공한다. 다른 예(도시되지 않음)에서, 전자 디바이스는 자연 언어 스피치 입력 "New Favs라는 새로운 재생목록에 마지막 10곡의 노래를 추가해"를 수신할 수 있다. 이에 응답하여, 전자 디바이스는 "New Favs"라는 미디어 아이템들의 새로운 컬렉션을 생성하고, 이전에 재생된 10곡의 노래를 새로운 컬렉션과 연관시킬 수 있다.
- [0284] 일부 예들에서, 전자 디바이스(800)는 "Send My Love"의 재생을 제공하는 동안 제4 자연 언어 입력(822)("아텔이 투어 중이야?")을 수신한다. 제4 자연 언어 스피치 입력(822)에 기초하여, 전자 디바이스는 특정 미디어 아이템에 관한 정보(예를 들어, 아티스트, 출시 날짜, 관련 인터뷰, 뒷 이야기, 가사의 의미, 투어 정보, 사용자의 친구들이 어느 미디어 아이템을 들었는지)를 획득하려는 사용자 의도를 결정한다. 일부 예들에서, 특정 미디어 아이템은 컨텍스트 정보(재생되고 있는 노래, 이전에 재생된 노래)에 기초하여 식별된다. 도시된 예에서, 전자 디바이스(800)는 아텔, 현재 재생된 노래의 가수과 관련된 투어 정보를 획득하려는 사용자 의도를 결정하고, 정보를 나타내는 스피치 출력(824)("네, 아텔이 다음 달에 당신의 도시에 있을 것입니다. 티켓을 원하십니까?")을 제공한다. 도시된 예에서, 사용자는 부정 응답(826)("지금은 아니야")을 제공한다. 다른 예(도시되지 않음)에서, 사용자는 긍정 응답을 제공할 수 있고, 전자 디바이스(800)는 콘서트 티켓을 구매하기 위한 프로세스를 개시할 수 있다. 일부 예들에서, 전자 디바이스(800)는 제4 자연 언어 스피치 입력(822) 없이 정보를 자동으로 제공한다.
- [0285] 일부 예들에서, 미디어 아이템을 재생하는 동안, 전자 디바이스(800)는 다른 미디어 아이템을 나타내는 스피치 출력을 제공할 수 있다. 예로서, 제2 미디어 아이템("Send My Love")을 제공하는 동안, 전자 디바이스(800)는 재생될 제3 미디어 아이템을 나타내는 스피치 출력(828)("다음은 아텔의 Someone Like You입니다")을 제공한다. 제2 미디어 아이템을 제공한 후에, 전자 디바이스(800)는 제3 미디어 아이템을 제공한다. 일부 예들에서, 미디어 아이템을 재생하는 동안, 전자 디바이스(800)는 위치를 나타내는 자연 언어 스피치 입력(예를 들어, "차고에서 이것을 재생해")을 수신할 수 있다. 이에 응답하여, 전자 디바이스(800)는 특정 위치에 기초하여 다른 전자 디바이스(예를 들어, 사용자의 차고와 연관된 스피커, 사용자의 차고에 물리적으로 위치된 전화기)를 식별하고, 식별된 전자 디바이스로 하여금 미디어 아이템의 재생을 제공하게 한다. 일부 예들에서, 전자 디바이스(800)는 식별된 전자 디바이스에 재생에 관한 정보(예를 들어, 미디어 아이템의 식별 정보, 재생의 진행, 볼륨과 같은 재생 설정)를 전송할 수 있다(예를 들어, 직접 또는 원격 디바이스를 통해). 위치를 나타내는 자연 언어 스피치

치 입력을 프로세싱하고 이에 따라 미디어 요청을 프로세싱하는 것에 대한 추가적 설명은, 예를 들어, 2014년 9월 30일자로 출원되고 발명의 명칭이 "Intelligent Assistant for Home Automation"인 미국 특허 출원 제 14/503,105호(대리인 관리 번호 제106842108200(P23013US1)호), 2016년 6월 9일자로 출원되고 발명의 명칭이 "Intelligent Automated Assistant in a Home Environment,"인 미국 특허 출원 제62/348,015호(대리인 관리 번호 제770003000100(P30331USP1)호), 및 2016년 6월11일자로 출원되고 발명의 명칭이 "Intelligent Device Arbitration and Control"인 미국 특허 출원 제62/348,896호(대리인 관리 번호 제770003001400(P30585USP1)호)에서 발견될 수 있으며, 이들의 전체 개시내용은 본 명세서에 참조로서 편입된다. 미국 특허 출원 제 14/503,105호는 특히, 전자 디바이스들(예를 들어, 도어 록, 서모스탯 등)을 제어하기 위해 가상 어시스턴트를 사용하는 예시적인 기술을 기술한다. 미국 가출원 제62/348,015호는 특히, 사용자 입력이 설정된 위치의 디바이스를 사용하여 태스크를 수행하고 설정된 위치의 디바이스들의 세트를 표현하는 데이터 구조들을 인출하려는 의도에 대응하는지 여부를 결정하기 위한 예시적인 기술을 기술한다. 미국 특허 가출원 제62/348,896호는 특히, 많은 디바이스들 중 어느 디바이스가 음성 입력에 기초하여 브로드캐스팅된 값들을 사용하여 음성 입력에 응답해야 하는지를 결정하는 예시적인 기술을 기술한다.

[0286] 도 9a 및 도 9b는 전자 디바이스(900)를 도시한다. 전자 디바이스(900)는 일부 실시예들에서 디바이스들(200, 400, 600, 800)(도 2a, 도 4, 도 6a와 도 6b, 도 8a와 도 8b) 중 임의의 것일 수 있다. 도시된 예에서, 전자 디바이스(900)는 하나 이상의 스피커들을 구비한 전자 디바이스이지만, 전자 디바이스는 전화, 랩톱 컴퓨터, 데스크톱 컴퓨터, 태블릿, 웨어러블 디바이스(예컨대, 스마트 워치) 셋톱 박스, 텔레비전, 스피커, 또는 이들의 임의의 조합 또는 하위 조합과 같은 임의의 유형의 디바이스일 수 있다는 것이 이해될 것이다.

[0287] 도 9a를 참조하면, 전자 디바이스(900)는 전자 디바이스(900)의 디지털 어시스턴트에 대한 요청을 나타내는 자연 언어 스피치 입력(910)을 (예를 들어, 마이크로폰을 통해) 수신한다. 자연 언어 스피치 입력(910)은 디지털 어시스턴트로 지향될 수 있는 임의의 요청을 포함할 수 있다. 일부 예들에서, 자연 언어 스피치 입력은 미리결정된 트리거 구절(예를 들어, "시리야")을 포함한다. 도 9a에 도시된 예에서, 사용자(902)는 트리거 구절 및 미디어 아이템에 대한 요청을 포함하는 자연 언어 스피치 입력(910)을 제공한다: "시리야, 오늘 나를 위한 음악이 뭐가 있어?"

[0288] 전자 디바이스(900)는 전자 디바이스(800)에 대하여 상기 논의한 것과 일치하는 방식으로 자연 언어 스피치 입력을 프로세싱한다. 예를 들어, 전자 디바이스(900)는 자연 언어 스피치 입력(910)을 프로세싱하여 하나 이상의 후보 텍스트 표현들(예를 들어, 텍스트 표현 "시리야, 오늘 나를 위한 음악이 뭐가 있어") 및 하나 이상의 후보 의도들(예를 들어, "미디어 추천을 획득하려는" 사용자 의도)을 제공한다.

[0289] 전자 디바이스(900)는 자연 언어 스피치 입력(910)에 기초하여 태스크를 식별한다. 일부 예들에서, 전자 디바이스(900)는 상기 논의된 바와 같이, 하나 이상의 후보 의도들(하나 이상의 후보 텍스트 표현들에 기초하여 차례로 식별됨)에 기초하여 하나 이상의 후보 태스크들을 식별한다. 또한, 전자 디바이스는 하나 이상의 결과들을 획득하기 위해 후보 태스크를 수행한다. 하나 이상의 결과들은 노래, 오디오 북, 팟캐스트, 스테이션, 재생 목록, 또는 이들의 조합과 같은 미디어 아이템에 관한 정보를 포함할 수 있다. 도 9a의 도시된 예에서, 전자 디바이스(900)는 자연 언어 스피치 입력(910)에 기초하여 파라미터들 "나를 위한" 및 "음악"을 갖는 "미디어 아이템을 제공하는" 후보 태스크를 식별한다. 또한, 전자 디바이스(900)는 식별된 태스크를 수행하여 재생목록 "Transgressive New Releases"에 관한 정보를 획득한다.

[0290] 도 9a를 참조하면, 전자 디바이스(900)는 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력(914)을 제공한다. 구체적으로, 전자 디바이스(900)는 식별된 재생목록의 구두 설명을 제공한다("재생목록 Transgressive New Releases가 있습니다"). 일부 예들에서, 사용자에게 제공되는 구두 설명은 식별된 미디어 아이템에 대응하는 정보(예를 들어, 메타데이터), 사용자 요청으로부터 식별된 파라미터, 또는 이들의 조합을 포함한다. 예를 들어, "나의 좋아하는 아티스트들로부터의 뭔가를 재생해"라는 사용자 요청에 응답하여, 전자 디바이스(900)는 "지난주에 발매된 당신의 좋아하는 가수들 중 하나인 아델의 노래입니다"라는 스피치 출력을 제공할 수 있다. 전자 디바이스(900)는 사용자-특정 코퍼스의 메타데이터에 기초하여 사용자의 좋아하는 가수들을 결정할 수 있다. 도 8을 참조하여 논의된 바와 같이, 일부 예들에서, 사용자-특정 코퍼스 내의 적어도 하나의 미디어 아이템은 하나 이상의 이전 사용자 입력들(사용자가 이전에 거부한, 사용자가 이전에 좋아한, 사용자가 이전에 검색한)을 나타내는 메타데이터를 포함한다. 대안적으로, 일부 다른 예들에서, 전자 디바이스(900)는 사용자 선호도 데이터(예를 들어, 사용자 데이터 및 모델(231))에 기초하여 사용자의 좋아하는 가수들을 결정할 수 있다.

[0291] 일부 예들에서, 전자 디바이스(900)는 하나 이상의 텍스트-스피치 모드들에 따라 스피치 출력(914)을 제공한다.

예를 들어, 스피치 출력(914)은 디지털 어시스턴트의 음성, 미디어 아이템과 연관된 음성(예를 들어, 아티스트, DJ), 또는 이들의 조합으로 제공될 수 있다. 상이한 텍스트-스피치 모드를 사용하는 추가의 예시적인 설명은 2017년 5월 16일자로 출원된 미국 특허 출원 제62/507,056호, "PROVIDING AN AUDITORY-BASED INTERFACE OF A DIGITAL ASSISTANT(대리인 관리 번호 제770003015700(P34183USP1)호)"에 제공되며, 이는 본 명세서에 그 전문이 참조로서 편입된다. 미국 특허 출원 제62/507,056호는 특허, 자연 언어 스피치 입력에서 특정된 정보와 같은 컨텍스트 정보에 기초하여 상이한 텍스트-스피치 모드를 사용하는 예시적인 기술을 기술한다.

[0292] 구두 응답(예를 들어, 스피치 출력(914))을 나타내는 스피치 출력을 제공하는 동안, 전자 디바이스(900)는 구두 응답에 대응하는 미디어 아이템의 재생인 오디오 출력(912)을 동시에 제공한다. 일부 예들에서, 재생되고 있는 미디어 아이템은 식별된 미디어 아이템의 일부분(예를 들어, 대표 샘플)이다. 예를 들어, 식별된 미디어 아이템이 단일 노래이면, 재생은 노래의 코러스 또는 제1 절을 포함할 수 있다. 다른 예로서, 식별된 미디어 아이템이 재생목록 인 경우, 재생은 재생목록 내의 각각의 노래에 대해 5-초 세그먼트를 포함할 수 있다.

[0293] 일부 예들에서, 전자 디바이스(900)는 스피치 출력과 상이한 볼륨(예를 들어, 그보다 낮은)에서 재생을 제공한다. 일부 예들에서, 전자 디바이스(900)는 스피치 출력과 상이한 충실도(예를 들어, 그보다 낮은)에서 재생을 제공한다. 일부 예들에서, 전자 디바이스(900)는 스피치 출력(914)을 제공하기 전에 오디오 출력(912)을 제공하기 시작한다. 다른 예들에서, 전자 디바이스(900)는 오디오 출력(912) 및 스피치 출력(914)을 동시에 제공하기 시작한다. 또 다른 예들에서, 전자 디바이스(900)는 오디오 출력(912)을 제공하기 전에 스피치 출력(914)을 제공하기 시작한다. 재생 중에 오디오를 조정하는 것의 추가적 설명은 2016년 9월 23일자로 출원된 미국 특허 출원 제62/399,232호, "INTELLIGENT AUTOMATED ASSISTANT"(대리인 관리 번호 제770003001300(P30584USP1)호)에 제공되며, 이는 본 명세서에 그 전문이 참조로서 편입된다.

[0294] 일부 예들에서, 재생(912)을 제공하는 동안, 전자 디바이스는 자연 언어 스피치 입력(916)("그것을 재생해!")을 수신한다. 자연 언어 스피치 입력(916)을 수신하는 것에 응답하여, 전자 디바이스(900)는 그 전체 내에서(예를 들어, 처음부터) 식별된 미디어 아이템의 재생인 오디오 출력(918)을 제공한다. 일부 예들에서, 재생(918)은 재생(912)과 상이한 볼륨 및/또는 충실도로 제공된다.

[0295] 일부 예들에서, 재생(918)을 제공하는 동안, 전자 디바이스(800)는 미디어 아이템에 관련된 정보를 나타내는 스피치 출력(도시되지 않음)을 제공한다. 정보는, 예를 들어, 노래의 사소한 정보("이것은 지난주에 발매되었습니다"), 아티스트의 투어 정보("이 아티스트는 다음 달에 캘리포니아에 올 것입니다. 티켓을 원하십니까?"), 또는 뉴스("이 아티스트는 최근에 약혼했습니다. 그것에 대해 더 알고 싶다면 알려주세요.")에 대응할 수 있다. 미디어 아이템은 재생되거나, 이전에 재생되었거나, 전자 디바이스(900)에 의해 재생될 미디어 아이템일 수 있다.

[0296] 일부 예들에서, 2개의 오디오 층(예를 들어, 구두 설명 및 식별된 미디어 아이템의 대표 샘플)을 동시에 제공하는 대신에, 전자 디바이스(900)는 대표 샘플의 재생을 제공하지 않으면서 식별된 미디어 아이템의 설명을 제공한다. 일부 예들에서, 사용자(902)는 대표 샘플(예를 들어, "어떤 소리가 나?", "어떤 종류의 노래가 재생목록에 있어?")을 듣기 위한 후속 요청을 제공할 수 있다. 이에 응답하여, 전자 디바이스(900)는 대표 샘플(예를 들어, "들어 보자. <30-초 요약>")의 재생을 제공하고, 일부 예들에서 사용자에게 추가 입력을 촉구한다("그것을 재생할까요?").

[0297] 다수의 오디오 층들을 제공하기 위한 상술한 기술은 단지 예시적인 것임이 이해될 것이다. 일반적으로, 전자 디바이스(900)는 디지털 어시스턴트와 사용자 간의 임의의 상호작용의 일부로서 계층화된 및/또는 조정된 오디오 정보를 제공할 수 있다. 예를 들어, 도 9b를 참조하면, 전자 디바이스(900)는 자연 언어 스피치 입력(918)을 수신한다("시리야, 우리 팀은 어떻게 했어?"). 입력에 기초하여, 전자 디바이스(900)는 태스크(예를 들어, "스포츠 이벤트의 점수를 획득하는" 후보 태스크) 및 하나 이상의 파라미터들(예를 들어, "자이언츠(Giants)")을 식별하고, 태스크를 수행하여 하나 이상의 결과들(예를 들어, 점수)을 획득한다. 일부 다른 예들에서, 전자 디바이스(900)는 사용자 선호도 데이터(예를 들어, 사용자 데이터 및 모델(231))에 기초하여 하나 이상의 파라미터들을 결정할 수 있다. 전자 디바이스(900)는 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력(922)을 제공한다. 도시된 예에서, 스피치 출력(922)은 획득된 결과들의 구두 설명을 나타낸다("어제 자이언츠가 이겼습니다, 점수는..").

[0298] 스피치 출력(922)을 제공하는 동안, 전자 디바이스(900)는 또한 구두 응답에 대응하는 미디어 아이템의 재생(920)을 제공한다. 도시된 예에서, 미디어 아이템은 승리 점수에 대응하는 사운드 효과(예를 들어, 군중 응원)이다. 일부 예들에서, 사운드 효과는 미리 녹음된 오디오(예를 들어, 일반적인 사운드 효과, 관련 스포츠 이

벤트에서 녹음된 오디오) 또는 라이브 스트림(예를 들어, 전자 디바이스의 현재 위치에서 비가 내리는 소리)이다. 일부 예들에서, 스피치 출력(922)은 재생(920)과 상이한 볼륨(예를 들어, 그보다 높은) 및/또는 충실도(예를 들어, 그보다 높은)로 제공된다.

[0299] 전자 디바이스(900)의 디지털 어시스턴트는 다양한 텍스트-스피치 모드, 음성, 및 시퀀스로 사용자와 상호작용할(예를 들어, 정보를 제공할) 수 있다는 것이 이해될 것이다. 일반적으로, 디지털 어시스턴트는 직관적이고, 풍부하고, 자연스러운 사용자 인터페이스를 제공하기 위해 오디오 출력의 다양한 층(예를 들어, 백그라운드 오디오, 포어그라운드 오디오) 및 다양한 유형(예를 들어, 사운드 효과, 스피치, 음악)을 조정할 수 있다. 예를 들어, 전자 디바이스는 포어그라운드 오디오의 타이밍, 볼륨, 충실도, 및 콘텐츠에 기초하여 백그라운드 오디오의 타이밍, 볼륨, 충실도, 및 콘텐츠를 조정할 수 있다.

[0300] 도 10a 및 도 10b는 전자 디바이스(1000)를 도시한다. 전자 디바이스(1000)는 일부 실시예들에서 디바이스들(200, 400, 600, 800, 900)(도 2a, 도 4, 도 6a와 도 6b, 도 8a와 도 8b, 도 9a와 도 9b) 중 임의의 것일 수 있다. 도시된 예에서, 전자 디바이스(1000)는 하나 이상의 스피커들을 구비한 전자 디바이스이지만, 전자 디바이스는 전화, 랩톱 컴퓨터, 데스크톱 컴퓨터, 태블릿, 웨어러블 디바이스(예컨대, 스마트 워치), 셋톱 박스, 텔레비전, 스피커, 또는 이들의 임의의 조합 또는 하위 조합과 같은 임의의 유형의 디바이스일 수 있다는 것이 이해될 것이다.

[0301] 도 10a를 참조하면, 전자 디바이스(1000)는 전자 디바이스(1000)의 디지털 어시스턴트에의 미디어에 대한 요청을 나타내는 자연 언어 스피치 입력(1010)("시리야, 내가 지금 뭘 들을까?")을 (예를 들어, 마이크로폰을 통해) 수신한다. 전자 디바이스(1000)는 전자 디바이스들(800, 900)에 대하여 상술된 것과 일치하는 방식으로 자연 언어 스피치 입력을 프로세싱한다. 예를 들어, 전자 디바이스(1000)는 자연 언어 스피치 입력(1010)을 프로세싱하여 하나 이상의 후보 텍스트 표현들(예를 들어, 텍스트 표현 "시리야, 내가 뭘 들을까?") 및 하나 이상의 후보 의도들(예를 들어, "미디어 추천을 획득하려는" 사용자 의도)을 제공한다.

[0302] 전자 디바이스(1000)는 자연 언어 스피치 입력(1010)에 기초하여 태스크를 식별한다. 일부 예들에서, 전자 디바이스(1000)는 상기 논의된 바와 같이 하나 이상의 후보 의도들에 기초하여 하나 이상의 후보 태스크들을 식별하고, 하나 이상의 결과들을 획득하기 위해 가장 높은 순위의 후보 태스크를 수행한다. 일부 예들에서, 하나 이상의 결과들은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 조합에 관한 정보를 포함한다. 도 10a의 도시된 예에서, 전자 디바이스(1000)는 파라미터 "나를 위해"를 갖는 "미디어 아이템을 제공하는" 후보 태스크를 식별하고 제1 미디어 아이템(뱅크스의 "The Altar")을 획득하기 위해 태스크를 수행한다.

[0303] 스피치 입력(1010)을 수신한 것에 응답하여, 전자 디바이스(1000)는 제1 미디어 아이템의 제안을 나타내는 오디오 출력(1012)을 제공한다("얼터너티브를 원한다면,뱅크스의 'The Altar'가 있습니다"). 일부 예들에서, 제1 미디어 아이템의 제안은 미디어 아이템(예를 들어, 장르, 아티스트)의 메타데이터 및 미디어 아이템이 왜 추천되는지에 대한 이유(들)(예를 들어, "당신이 얼터너티브를 원한다면..")와 같은 미디어 추천을 컨텍스트화하도록 추가 정보를 포함한다. 일부 예들에서, 전자 디바이스(1000)는 추천된 미디어 아이템의 일부분의 재생을 동시에 제공한다.

[0304] 일부 예들에서, 오디오 출력(1012)을 제공한 후, 전자 디바이스(1000)는 스피치 입력(1014)("아니")을 수신한다. 전자 디바이스는 스피치 입력(1014)이 미디어에 대한 요청에 대응하는 비-긍정 응답(예를 들어, "아니오", "다음", "그것을 좋아하지 않아", "그것을 싫어해" "몇 개 더" 및 거부를 나타내는 임의의 구절의 자연 언어 등가물)을 나타낸다. 스피치 입력(1014)이 비-긍정 응답을 나타낸다는 결정에 따라, 전자 디바이스는 요청에 대응하는 연속적인 비-긍정 응답들의 수를 업데이트한다. 한편, 스피치 입력(1014)이 비-긍정 응답을 나타내지 않는다는 결정에 따라, 전자 디바이스는 그 수를 업데이트하는 것을 중지한다. 도시된 예에서, 전자 디바이스는 스피치 입력(1014)에 기초하여 0에서 1로 번호를 업데이트한다.

[0305] 도 10a를 참조하면, 일부 예들에서, 전자 디바이스(1000)는 제2 미디어 아이템의 제안을 나타내는 다른 오디오 출력(1016)을 제공한다(재생목록 When Hip-Hop Goes Left는 어떻습니까?). 일부 예들에서, 전자 디바이스(900)는 단일 검색의 일부로서 제1 미디어 아이템 및 제2 미디어 아이템을 식별하고, 제1 미디어 아이템은 제2 미디어 아이템보다 높은 신뢰도 점수를 가지며, 따라서 전자 디바이스에 의해 우선 제안된다. 일부 예들에서, 전자 디바이스는 제1 미디어 아이템 및 제2 미디어 아이템을 각각 식별하기 위해 2개의 별개의 검색을 수행하고, 사용자가 제1 미디어 아이템의 제안(예를 들어, 오디오 입력(1014))에 비-긍정 응답을 제공한 후에 제2 검색이 수행된다.

- [0306] 제2 미디어 아이템의 제안을 제공한 후에, 전자 디바이스(1000)는 스피치 입력(1018)("다음")을 수신한다. 전자 디바이스는 스피치 입력(1018)이 스피치 입력(1014)에 관하여 기술된 것과 일치하는 방식으로 미디어에 대하여 요청에 대응하는 비-긍정 응답을 나타내는지 여부를 결정한다. 도시된 예에서, 전자 디바이스는 스피치 입력(1018)이 비-긍정 응답을 나타내고 1에서 2로 번호를 업데이트한다고 결정한다.
- [0307] 도 10a를 참조하면, 일부 예들에서, 전자 디바이스(1000)는 제3 미디어 아이템의 제안을 나타내는 또 다른 오디오 출력(1020)을 제공한다("재생목록 'If You Like Alabama Shakes'도 있습니다."). 전자 디바이스(900)는 (예를 들어, 상이한 검색 파라미터 및/또는 컨텍스트 정보를 사용하여) 단일 검색 또는 상이한 검색에서 제1 미디어 아이템, 제2 미디어 아이템, 및 제3 미디어 아이템을 식별할 수 있다.
- [0308] 제3 미디어 아이템에 대한 제안을 제공한 후에, 전자 디바이스(1000)는 사용자로부터의 입력들에 대해 (예를 들어, 마이크로폰을 통해) 샘플링한다. 일부 예들에서, 전자 디바이스는 응답이 미리결정된 기간(예를 들어, 5초) 내에 수신되지 않은 것으로 결정한다. 결정에 따라, 전자 디바이스는 요청에 대응하는 연속적인 비-긍정 응답들의 수를 업데이트한다. 도시된 예에서, 전자 디바이스는 미리결정된 기간 내에 응답이 수신되지 않았다고(예를 들어, 침묵) 결정하고 2에서 3으로 번호를 업데이트한다.
- [0309] 전자 디바이스(1000)는 미디어에 대한 요청에 대응하는 연속적인 비-긍정 응답들의 수가 임계치를 만족시키는지 여부를 결정한다. 일부 예들에서, 전자 디바이스는 오디오 출력들(1012, 1016, 1020) 각각을 제공한 후에 결정을 한다. 연속적인 비-긍정 응답들의 수가 임계치를 만족시키지 못한다는 결정에 따라, 전자 디바이스는 다른 미디어 아이템의 제안을 나타내는 오디오 출력을 제공한다. 예를 들어, 오디오 출력(1018)("다음")을 수신한 후, 전자 디바이스는 연속적인 비-긍정 응답의 수(2)가 미리결정된 임계치(예를 들어, 3)와 동일하지 않은 것으로 결정한다. 따라서, 전자 디바이스는 제안된 것과 상이한 다른 미디어 아이템을 제안하기 위해 스피치 출력(1020)을 제공한다.
- [0310] 연속적인 비-긍정 응답들의 수가 임계치를 만족시킨다는 결정에 따라, 전자 디바이스는 다른 미디어 아이템의 제안을 제공하는 것을 중지하고, 대신에 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공한다. 예를 들어, 비-긍정 응답(1022)(예를 들어, 미리결정된 양의 시간 동안 침묵)을 수신한 후, 전자 디바이스는 연속적인 비-긍정 응답의 수(3)가 미리결정된 임계치(예를 들어, 3)와 동일한 것으로 결정한다. 따라서, 도 10b를 참조하면, 전자 디바이스는 스피치 출력(1024)을 제공한다("알겠습니다. 최근에 즐기고 있는 아티스트의 이름을 말씀해 주실 수 있습니까?").
- [0311] 일부 예들에서, 스피치 출력(1024)은 미디어에 대한 요청에 대한 하나 이상의 파라미터들에 대한 촉구를 나타낸다. 도시된 예에서, 전자 디바이스(1000)는 아티스트 파라미터를 사용자에게 촉구하고("당신이 최근에 즐기고 있는 아티스트의 이름을 말씀해 주실 수 있습니까?"), 스피치 입력(1026)("음.. 플룸(Flume)")을 수신한다. 스피치 입력(1026)은 미디어(아티스트 = 플룸)에 대한 요청에 대한 파라미터를 나타낸다. 수신된 파라미터에 기초하여, 전자 디바이스(1000)는 이전에 추천된 미디어 아이템과 상이한 다른 미디어 아이템을 식별한다. 따라서, 전자 디바이스(1000)는 스피치 출력(1028)을 통해 식별된 미디어 아이템을 제공한다("좋습니다. 재생목록 'If You Like Flume'이 있습니다."). 일부 예들에서, 수신된 파라미터에 기초하여, 전자 디바이스는 그에 따라 사용자 선호 데이터(예를 들어, 사용자 데이터 및 모델(231)) 및/또는 사용자-특정 코퍼스를 업데이트한다.
- [0312] 일부 예들에서, 스피치 출력(1024)을 제공하는 것에 대한적으로 또는 추가적으로, 전자 디바이스(1000)는 이전에 제안된 복수의 미디어 아이템들 중에서 사용자 선택을 위한 촉구를 나타내는 스피치 출력을 제공한다(예를 들어, "이것들 중에 괜찮은게 있습니까?"). 일부 예들에서, 전자 디바이스는 사용자 선택(예를 들어, "응, 두 번째 것", "힙합 것", "아델에 의한 것")을 나타내는 스피치 입력을 수신하고, 컨텍스트 정보에 기초하여 스피치 입력을 해석한다. 컨텍스트 정보는 전자 디바이스에 의해 이전에 제안된 복수의 미디어 아이템들을 포함할 수 있다.
- [0313] 도 11은 전자 디바이스(1100)를 도시한다. 전자 디바이스(1100)는 일부 실시예들에서 디바이스들(200, 400, 600, 800, 900, 1000)(도 2a, 도 4, 도 6a와 도 6b, 도 8a와 도 8b, 도 9a와 도 9b, 도 10) 중 임의의 것일 수 있다. 도시된 예에서, 전자 디바이스(1100)는 하나 이상의 스피커들을 구비한 전자 디바이스이지만, 전자 디바이스는 전화, 랩톱 컴퓨터, 데스크톱 컴퓨터, 태블릿, 웨어러블 디바이스(예컨대, 스마트 워치), 셋톱 박스, 텔레비전, 스피커, 또는 이들의 임의의 조합 또는 하위 조합과 같은 임의의 유형의 디바이스일 수 있다는 것이 이해될 것이다.

- [0314] 동작 중에, 전자 디바이스(1100)는, 사용자(1102)에 의해 발언된, 전자 디바이스(1100)의 디지털 어시스턴트에 의 미디어에 대한 요청을 나타내는 자연 언어 스피치 입력(1110)("시리야, 뭔가를 재생해")을 (예를 들어, 마이크로폰을 통해) 수신한다. 도시된 예에서, 전자 디바이스(1100)는 사용자(1102)와 연관된다. 전자 디바이스(1100)는 전자 디바이스들(800, 900, 1000)에 대하여 상술된 것과 일치하는 방식으로 자연 언어 스피치 입력을 프로세싱한다. 예를 들어, 전자 디바이스(1100)는 자연 언어 스피치 입력(1110)을 프로세싱하여 하나 이상의 후보 텍스트 표현들(예를 들어, 텍스트 표현 "시리야, 뭔가를 재생해") 및 하나 이상의 후보 의도들(예를 들어, "미디어 추천을 획득하려는" 사용자 의도)을 제공한다.
- [0315] 전자 디바이스(1100)는 전자 디바이스에 근접한 복수의 사용자들의 물리적 존재를 검출한다. 일부 예들에서, 전자 디바이스(1100)는 사람의 전자 디바이스의 접속성 상태(예를 들어, 동일한 Wi-Fi 네트워크 상에서, 블루투스 범위 내에서, NFC 범위 내에서)와 같은 사람과 연관된 전자 디바이스와 관련된 정보, 사람의 전자 디바이스 상의 정보 등에 기초하여 사람의 물리적 존재를 검출할 수 있다. 예를 들어, 사용자의 자매가 또한 전자 디바이스(1100)와 물리적으로 근접하고 그녀의 전화기를 그녀가 갖고 있다면, 전자 디바이스(1100)는 자매의 전화기에 대응하는 정보를 수신할 수 있다. 예를 들어, 전자 디바이스(1100)는 (예를 들어, 블루투스 접속을 통해) 자매의 디바이스로부터 식별 정보(예를 들어, 전화 번호, 사용자 이름)를 수신할 수 있다. 다른 예로서, 전자 디바이스(1100)는 라우팅 디바이스(예컨대, 전자 디바이스(1100) 및 자매의 디바이스 둘 모두가 접속된 무선 라우터)로부터 식별 정보를 수신할 수 있다.
- [0316] 일부 예들에서, 전자 디바이스(1100)는 (예를 들어, 카메라 및 마이크로폰을 통해 캡처된) 사람의 얼굴 특성 및 /또는 음성 특성에 기초하여 사람의 물리적 존재를 검출할 수 있다. 다른 예들에서, 전자 디바이스(800)는 캘린더 초청 또는 이메일 메시지에 열거된 연락처들과 같이 국부적으로 이용가능한 정보에 기초하여 사람의 물리적 존재를 검출할 수 있다. 또 다른 일부 예들에서, 전자 디바이스(800)는 사람에 의해 제공된 크리덴셜(예를 들어, 사용자 이름 및 패스워드)에 기초하여 사람의 물리적 존재를 검출할 수 있다. 일부 예들에서, 전자 디바이스(1100)는 사람의 물리적 존재를 검출한 후에 중의성 해소 입력(예를 들어, "내가 듣고 있는 게 그 제이슨 또는 존입니까?") 및/또는 확인 입력(예를 들어, "존이 지금 막 파티에 참석했습니까?")을 촉구한다.
- [0317] 복수의 사용자들(예를 들어, 가족 구성원, 방문자)의 물리적 존재를 검출하는 것에 응답하여, 전자 디바이스(1100)는 복수의 사용자들에 대응하는 복수의 선호도 프로파일들을 획득한다. 일부 예들에서, 전자 디바이스(1100)는 원격 디바이스(예를 들어, 서버 디바이스)로부터 사용자(1102) 이외의 사람(예를 들어, 사용자의 자매)에 대응하는 선호도 프로파일을 수신한다. 일부 예들에서, 전자 디바이스(1100)는 사람의 전자 디바이스(예를 들어, 자매의 전화기)로부터 직접적으로 사용자(1102) 이외의 사람(예를 들어, 사용자(1102)의 자매)에 대응하는 선호도 프로파일을 수신한다. 일부 예들에서, 전자 디바이스(1100)는 사용자(1102) 이외의 사람의 선호도 프로파일을 국부적으로 저장한다. 예를 들어, 사용자(1102)는 디지털 어시스턴트에게 선호도를 국부적으로 저장하도록 이전에 요청했을 수 있다(예를 들어, "시리야, 나의 자매는 비틀즈(The Beatles)를 좋아한다는 것을 기억해 줘").
- [0318] 복수의 선호도 프로파일들에 기초하여, 전자 디바이스(1100)는 병합된 선호도 프로파일을 제공한다. 일부 예들에서, 병합된 선호도 프로파일을 제공하는 것은, 복수의 선호도 프로파일들 각각에 의해 공유되는 하나 이상의 선호도들을 식별하는 것을 포함한다. 도시된 예에서, 전자 디바이스(1100)는 사용자의 선호도 프로파일 및 자매의 선호도 프로파일에 기초하여 병합된 선호도 프로파일을 제공한다. 사용자와 자매 둘 모두가 비틀즈를 선호하기 때문에, 병합된 선호도 프로파일은 비틀즈에 대한 선호도를 포함한다. 한편, 일부 예들에서, 자매가 아니라 사용자만이 밴크스에 대한 선호도를 갖기 때문에, 병합된 선호도 프로파일은 밴크스에 대한 선호도를 포함하지 않을 수 있다.
- [0319] 병합된 선호도 프로파일에 기초하여, 전자 디바이스(1100)는 미디어 아이템을 식별한다. 식별된 미디어 아이템은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 임의의 조합일 수 있다. 예를 들어, 전자 디바이스(1100)는 노래(예를 들어, 아티스트)의 메타데이터가 병합된 프로파일의 하나 이상의 선호도들(예를 들어, 비틀즈)과 매칭되기 때문에 노래 "Hey Jude"를 식별한다. 따라서, 전자 디바이스(1100)는 오디오 출력(1112)을 제공한다("비틀즈로부터 두 분 다 좋아할 만한 것이 있습니다"). 오디오 출력(1112)은 식별된 미디어 아이템의 설명("비틀즈로부터")을 포함하고 병합된 프로파일을 언급한다(예를 들어, "두 분 다 좋아할 만한 것"). 전자 디바이스(1100)는 또한 식별된 미디어를 포함하는 오디오 출력(1113)(노래 "Hey Jude"의 재생)을 제공한다.
- [0320] 일부 예들에서, 병합된 선호도 프로파일에 기초하여 미디어 아이템을 식별하는 것은, 복수의 미디어 아이템들로

부터 미디어 아이템을 식별하는 것을 포함한다. 일부 예들에서, 복수의 미디어 아이템들은 제1 사용자와 연관된 미디어 아이템들의 제1 세트(예를 들어, 제1 사용자가 액세스하도록 인가됨) 및 제2 사용자와 연관된 미디어 아이템들의 제2 세트(예를 들어, 제2 사용자가 액세스하도록 인가됨)를 포함한다. 도시된 예에서, 식별된 미디어 아이템은 미디어 아이템들의 제1 세트의 일부는 아니지만 미디어 아이템들의 제2 세트의 일부이다(즉, 사용자는 노래 "Hey Jude"에 대한 액세스를 갖지 않지만, 사용자의 자매는 액세스를 갖는다).

[0321] 일부 예들에서, 제2 사용자(예를 들어, 자매)를 포함하는 다수의 사용자들의 물리적 존재를 검출한 후에, 전자 디바이스(1100)는 제2 사용자의 부재를 검출한다. 전자 디바이스(1100)는 제2 사용자의 존재를 검출하기 위한 기술과 유사한 기술을 사용하여 제2 사용자의 부재를 검출할 수 있다. 예를 들어, 전자 디바이스(1100)는 제2 사용자의 전자 디바이스와 관련된 정보(예를 들어, 자매의 디바이스가 여전히 무선 네트워크에 접속되어 있는지 여부)를 획득함으로써 부재를 검출할 수 있다. 정보는 제2 사용자의 디바이스로부터 직접적으로 또는 네트워크 라우터로부터 획득될 수 있다. 제2 사용자의 부재를 검출한 후에, 전자 디바이스(1100)는 병합된 선호도 프로파일 및/또는 그로부터 검색할 복수의 미디어 아이템들을 업데이트한다. 예를 들어, 전자 디바이스(1100)가 사용자의 자매의 부재를 검출하면, 전자 디바이스(1100)는 그로부터 검색할 복수의 미디어 아이템으로부터 자매만이 액세스할 수 있는 미디어 아이템들(예를 들어, 자매의 비틀즈 컬렉션)을 제거한다.

[0322] 일부 예들에서, 전자 디바이스(1100)는 사용자 이외의 사람의 선호도 및/또는 활동에 기초하여 미디어에 대한 요청을 나타내는 자연 언어 스피치 입력(1114)을 수신한다. 도 11의 도시된 예에서, 사용자(1102)는 오디오 출력(1114)("내 친구들은 무엇을 듣고 있어?")을 제공한다. 이에 응답하여, 전자 디바이스(1100)는 (예를 들어, 연락처 목록, 소셜 미디어 서비스 및 미디어 서비스와 같은 소프트웨어 서비스, 및 다른 사용자-특정 데이터를 통해) 하나 이상의 사람들을 식별한다. 또한, 전자 디바이스는 하나 이상의 소스들(예를 들어, 소셜 미디어 서비스 및 미디어 서비스와 같은 소프트웨어 서비스)로부터 하나 이상의 사람의 선호도들(예를 들어, 선호 장르, 선호 아티스트) 및/또는 활동들(최근에 재생된 노래들)에 관한 정보를 획득한다. 예를 들어, 전자 디바이스(1100)는 소프트웨어 서비스 상에서 사용자와 연관된 하나 이상의 사람들을 식별할 수 있고 이들 사람들 중 일부 또는 전부가 소프트웨어 서비스를 사용하여 재생한 미디어 아이템을 식별할 수 있다. 대안적으로, 전자 디바이스(1100)는 데이터베이스(예를 들어, 상술한 사용자-특정 코퍼스) 내의 적절한 메타데이터(예를 들어, 친구 태그)를 갖는 미디어 아이템들을 검색함으로써 미디어 아이템을 식별한다. 도시된 예에서, 전자 디바이스(1100)는 식별된 미디어 아이템을 제공하기 위해 오디오 출력(1116)("아델의 Hello입니다")을 제공한다.

[0323] 일부 예들에서, 전자 디바이스는 정보에 대한 요청을 나타내는 자연 언어 스피치 입력(1118)("누가 이것을 듣고 있어?")을 수신한다. 이에 응답하여, 전자 디바이스(1100)는 (예를 들어, 관련 소프트웨어 서비스를 사용하여 사용자-특정 코퍼스를 사용하여) 미디어 아이템과 연관된 하나 이상의 사람들의 식별 정보를 제공한다. 식별 정보는 국부적으로 및/또는 하나 이상의 원격 디바이스들로부터 획득될 수 있다. 도시된 예에서, 전자 디바이스(1100)는 식별 정보를 제공하기 위해 오디오 출력(1120)("당신의 친구들 존과 제인")을 제공한다.

[0324] 4. 미디어 탐색을 위한 디지털 어시스턴트의 청각 기반 인터페이스를 제공하는 것에 대한 프로세스

[0325] 도 12는 다양한 예들에 따른, 디지털 어시스턴트의 청각 기반 인터페이스를 제공하기 위한 프로세스(1200)를 도시한다. 프로세스(1200)는 예를 들어, 디지털 어시스턴트를 구현하는 하나 이상의 전자 디바이스들을 사용하여 수행된다. 일부 예들에서, 프로세스(1200)는 클라이언트-서버 시스템(예컨대, 시스템(100))을 사용하여 수행되고, 프로세스(1200)의 블록들은 서버(예컨대, DA 서버(106))와 클라이언트 디바이스 사이에서 임의의 방식으로 분담된다. 다른 예들에서, 프로세스(1200)의 블록들은 서버와 다수의 클라이언트 디바이스들(예컨대, 모바일 전화기 및 스마트 워치) 사이에서 분담된다. 따라서, 프로세스(1200)의 일부분들이 클라이언트-서버 시스템의 특정 디바이스들에 의해 수행되는 것으로 본 명세서에 기술되지만, 프로세스(1200)는 그렇게 제한되지 않는다는 것이 이해될 것이다. 다른 예들에서, 프로세스(1200)는 단지 하나의 클라이언트 디바이스(예컨대, 사용자 디바이스(104))만을 또는 다수의 클라이언트 디바이스들만을 사용하여 수행된다. 프로세스(1200)에서, 일부 블록들은, 옵션적으로, 조합되고, 일부 블록들의 순서는, 옵션적으로, 변경되고, 일부 블록들은, 옵션적으로, 생략된다. 일부 예들에서, 추가 단계들이 프로세스(1200)와 조합하여 수행될 수 있다.

[0326] 블록(1202)에서, 전자 디바이스는 미디어에 대한 요청을 나타내는 제1 자연 언어 스피치 입력을 수신한다. 제1 자연 언어 스피치 입력은 제1 검색 파라미터를 포함한다. 일부 예들에서, 전자 디바이스는 제1 자연 언어 스피치 입력에 기초하여, 텍스트 문자열을 획득한다. 또한, 전자 디바이스는 텍스트 문자열에 기초하여, 미디어 아이템들에 대한 추천을 획득하려는 사용자 의도의 표현을 결정한다. 또한, 전자 디바이스는, 사용자 의도의 표현에 기초하여, 태스크 및 그 태스크를 수행하기 위한 하나 이상의 파라미터들을 결정하며, 이는 제1 검색 파라

미터를 포함한다.

- [0327] 블록(1204)에서, 전자 디바이스(또는 전자 디바이스의 디지털 어시스턴트)는 제1 미디어 아이템을 제공한다. 제1 미디어 아이템은 제1 검색 파라미터에 기초하여 식별된다. 일부 예들에서, 제1 미디어 아이템은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 임의의 조합이다.
- [0328] 일부 예들에서, 제1 미디어 아이템을 제공하는 것은, 디지털 어시스턴트에 의해, 제1 미디어 아이템과 연관된 구두 응답을 나타내는 스피치 출력을 제공하는 것을 포함한다. 제1 미디어 아이템을 제공하는 것은, 구두 응답을 나타내는 스피치 출력을 제공하는 동안, 디지털 어시스턴트에 의해, 제1 미디어 아이템의 일부분의 재생을 제공하는 것을 추가로 포함한다. 일부 다른 예들에서, 제1 미디어 아이템을 제공하는 것은, 디지털 어시스턴트에 의해, 제1 미디어 아이템의 재생을 제공하는 것을 포함한다. 일부 다른 예들에서, 제1 미디어 아이템을 제공하는 것은, 디지털 어시스턴트에 의해, 복수의 미디어 아이템들을 제공하는 것을 포함하며, 이는 제1 미디어 아이템을 포함한다.
- [0329] 블록(1206)에서, 제1 미디어 아이템을 제공하는 동안, 전자 디바이스는 제2 자연 언어 스피치 입력을 수신한다. 일부 예들에서, 제2 자연 언어 스피치 입력을 수신하는 것에 응답하여, 전자 디바이스는 제1 미디어 아이템이 제공되는 방식을 조정한다.
- [0330] 블록(1208)에서, 전자 디바이스는 제2 자연 언어 스피치 입력이 미디어에 대한 요청을 정제하려는 사용자 의도에 대응하는지 여부를 결정한다. 일부 예들에서, 제2 자연 언어 스피치 입력이 미디어에 대한 요청을 정제하려는 사용자 의도에 대응하는지 여부를 결정하는 것은, 하나 이상의 미리정의된 구절들 및 하나 이상의 구절들의 자연 언어 등가물에 기초하여 미디어에 대한 요청을 정제하려는 사용자 의도의 표현을 도출하는 것을 포함한다. 일부 예들에서, 제2 자연 언어 스피치 입력이 미디어에 대한 요청을 정제하려는 사용자 의도에 대응하는지 여부를 결정하는 것은, 컨텍스트 정보에 기초하여 미디어에 대한 요청을 정제하려는 사용자 의도의 표현을 도출하는 것을 포함한다.
- [0331] 일부 예들에서, 전자 디바이스는 제2 자연 언어 스피치 입력에 기초하여, 미디어에 대한 요청을 정제하기 위한 하나 이상의 파라미터들을 획득한다. 일부 예들에서, 하나 이상의 파라미터들 중 일 파라미터는 미디어 아이템의 가사 내용, 행사 또는 기간, 활동, 위치, 기분, 미리결정된 시간 프레임 내의 출시 날짜, 의도된 청중, 미디어 아이템들의 컬렉션, 또는 이들의 임의의 조합에 대응한다. 일부 예들에서, 제2 자연 언어 스피치 입력은 제1 사용자와 연관되고, 하나 이상의 파라미터들 중 일 파라미터는 제1 사용자와 상이한 제2 사용자에 대응한다.
- [0332] 일부 예들에서, 미디어에 대한 요청을 정제하기 위한 하나 이상의 파라미터들을 획득하는 것은, 컨텍스트 정보에 기초하여 하나 이상의 파라미터들을 결정하는 것을 포함한다. 일부 예들에서, 컨텍스트 정보는 제1 미디어 아이템에 관련된 정보를 포함한다.
- [0333] 일부 예들에서, 전자 디바이스는 하나 이상의 사용자들의 물리적 존재를 검출하고, 컨텍스트 정보는 하나 이상의 사용자들에 관련된 정보를 포함한다. 일부 예들에서, 컨텍스트 정보는 전자 디바이스의 하나 이상의 사용자들과 연관된 설정을 포함한다.
- [0334] 블록(1210)에서, 제2 자연 언어 스피치 입력이 미디어에 대한 요청을 정제하려는 사용자 의도에 대응한다는 결정에 따라, 전자 디바이스(또는 디지털 어시스턴트)는 제1 파라미터 및 제2 자연 언어 스피치 입력에 기초하여, 제1 미디어 아이템과 상이한 제2 미디어 아이템을 식별하고, 제2 미디어 아이템을 제공한다. 제2 미디어 아이템은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 임의의 조합일 수 있다.
- [0335] 일부 예들에서, 전자 디바이스는, 제1 자연 언어 스피치 입력에 기초하여, 미디어 아이템들의 제1 세트를 획득하고 미디어 아이템들의 제1 세트로부터 제1 미디어 아이템을 선택한다. 또한, 전자 디바이스는, 제2 자연 언어 스피치 입력에 기초하여, 미디어 아이템들의 제1 세트의 서브세트인 미디어 아이템들의 제2 세트를 획득하고, 미디어 아이템들의 세트 제2 세트로부터 제2 미디어 아이템을 선택한다. 일부 예들에서, 미디어 아이템들의 제2 세트를 획득하는 것은, 미디어 아이템들의 제1 세트로부터, 미디어에 대한 요청을 정제하기 위한 하나 이상의 파라미터들에 기초하여 하나 이상의 미디어 아이템들을 선택하는 것을 포함한다.
- [0336] 일부 예들에서, 제2 미디어 아이템을 식별하는 것은 제2 미디어 아이템과 연관된 콘텐츠가 하나 이상의 파라미터들 중 적어도 하나와 매칭되는지 여부를 결정하는 것을 포함한다. 일부 다른 예들에서, 제2 미디어 아이템을 식별하는 것은 제2 미디어 아이템과 연관된 메타데이터가 하나 이상의 파라미터들 중 적어도 하나와 매칭되는지 여부를 결정하는 것을 포함한다.

- [0337] 일부 예들에서, 전자 디바이스는 미디어 아이тем들의 사용자-특정 코퍼스로부터 제2 미디어 아이тем을 획득하는데, 미디어 아이тем들의 사용자-특정 코퍼스는 사용자와 연관된 데이터에 기초하여 생성된다. 일부 예들에서, 전자 디바이스는 제2 자연 언어 스피치 입력과 연관된 음향 정보에 기초하여 미디어 아이тем들의 사용자-특정 코퍼스를 식별한다. 일부 예들에서, 미디어 아이тем들의 사용자-특정 코퍼스 내의 미디어 아이тем은, 활동; 기분; 행사; 위치; 시간; 큐레이터; 재생목록; 하나 이상의 이전 사용자 입력들; 또는 이들의 임의의 조합을 나타내는 메타데이터를 포함한다. 일부 예들에서, 메타데이터의 적어도 일부분은 제1 사용자와 상이한 제2 사용자로부터의 정보에 기초한다.
- [0338] 일부 예들에서, 제2 미디어 아이тем을 제공하는 것은, 디지털 어시스턴트에 의해, 제2 미디어 아이тем과 연관된 구두 응답을 나타내는 스피치 출력을 제공하는 것을 포함한다. 또한, 제2 미디어 아이тем을 제공하는 것은, 구두 응답을 나타내는 스피치 출력을 제공하는 동안, 디지털 어시스턴트에 의해, 제2 미디어 아이тем의 일부분의 재생을 제공하는 것을 포함한다.
- [0339] 일부 예들에서, 제2 미디어 아이тем을 제공하는 것은, 디지털 어시스턴트에 의해, 제2 미디어 아이тем의 재생을 제공하는 것을 포함한다. 일부 다른 예들에서, 제2 미디어 아이тем을 제공하는 것은, 디지털 어시스턴트에 의해, 복수의 미디어 아이тем들을 제공하는 것을 포함하며, 이는 제2 미디어 아이тем을 포함한다.
- [0340] 일부 예들에서, 전자 디바이스는, 제3 자연 언어 스피치 입력을 수신하고, 제3 자연 언어 스피치 입력에 기초하여, 제2 미디어 아이тем을 미디어 아이тем들의 컬렉션과 연관시키려는 사용자 의도의 표현을 결정한다. 또한, 전자 디바이스는 제2 미디어 아이тем을 미디어 아이тем들의 컬렉션과 연관시키고, 디지털 어시스턴트에 의해, 그 연관성을 나타내는 오디오 출력을 제공한다.
- [0341] 일부 예들에서, 제2 미디어 아이тем을 제공하는 동안, 전자 디바이스는 제4 자연 언어 스피치 입력을 수신한다. 또한, 전자 디바이스는 제4 자연 언어 스피치 입력에 기초하여, 특정 미디어 아이тем에 관련된 정보를 획득하려는 사용자 의도의 표현을 결정하고, 디지털 어시스턴트에 의해, 특정 미디어 아이тем에 관련된 정보를 제공한다. 일부 예들에서, 전자 디바이스는 컨텍스트 정보에 기초하여 특정 미디어 아이тем을 선택한다.
- [0342] 일부 예들에서, 제2 미디어 아이тем을 제공하는 동안, 전자 디바이스(또는 전자 디바이스의 디지털 어시스턴트)는 제3 미디어 아이тем을 나타내는 스피치 출력을 제공하고, 제2 미디어 아이тем을 제공한 후에 제3 미디어 아이тем을 제공한다.
- [0343] 일부 예들에서, 전자 디바이스는 컴퓨터, 셋톱 박스, 스피커, 스마트 워치, 전화, 또는 이들의 조합이다.
- [0344] 도 12에 관하여 상술한 동작들은, 옵션적으로, 도 1 내지 도 4, 도 6a와 도 6b, 및 도 7a 내지 도 7c에 도시된 컴포넌트들에 의해 구현된다. 예를 들어, 프로세스(1200)의 동작들은, 본 명세서에 기술된, 디바이스들(104, 200, 400, 600, 800, 900, 1000, 1100)을 포함하지만 이로 한정되지는 않는 임의의 디바이스 또는 이들의 컴포넌트에 의해 구현될 수 있다. 다른 프로세스들이 도 1 내지 도 4, 도 6a와 도 6b, 및 도 7a 내지 도 7c에 도시된 컴포넌트들에 기초하여 어떻게 구현되는지는 본 발명이 속한 기술분야에서 통상의 지식을 가진 자에게 자명할 것이다.
- [0345] 도 13은 다양한 예들에 따른, 디지털 어시스턴트의 청각 기반 인터페이스를 제공하기 위한 프로세스(1300)를 도시한다. 프로세스(1300)는 예를 들어, 디지털 어시스턴트를 구현하는 하나 이상의 전자 디바이스들을 사용하여 수행된다. 일부 예들에서, 프로세스(1300)는 클라이언트-서버 시스템(예컨대, 시스템(100))을 사용하여 수행되고, 프로세스(1300)의 블록들은 서버(예컨대, DA 서버(106))와 클라이언트 디바이스 사이에서 임의의 방식으로 분담된다. 다른 예들에서, 프로세스(1300)의 블록들은 서버와 다수의 클라이언트 디바이스들(예컨대, 모바일 전화기 및 스마트 워치) 사이에서 분담된다. 따라서, 프로세스(1300)의 일부분들이 클라이언트-서버 시스템의 특정 디바이스들에 의해 수행되는 것으로 본 명세서에 기술되지만, 프로세스(1300)는 그렇게 제한되지 않는다는 것이 이해될 것이다. 다른 예들에서, 프로세스(1300)는 단지 하나의 클라이언트 디바이스(예컨대, 사용자 디바이스(104))만을 또는 다수의 클라이언트 디바이스들만을 사용하여 수행된다. 프로세스(1300)에서, 일부 블록들은, 옵션적으로, 조합되고, 일부 블록들의 순서는, 옵션적으로, 변경되고, 일부 블록들은, 옵션적으로, 생략된다. 일부 예들에서, 추가 단계들이 프로세스(1300)와 조합하여 수행될 수 있다.
- [0346] 블록(1302)에서, 전자 디바이스는 자연 언어 스피치 입력을 수신한다. 일부 예들에서, 자연 언어 스피치 입력은 하나 이상의 미디어 아이тем들에 대한 요청을 나타낸다.
- [0347] 블록(1304)에서, 전자 디바이스(또는 전자 디바이스의 디지털 어시스턴트)는 자연 언어 스피치 입력에 기초하여

태스크를 식별한다. 일부 예들에서, 태스크를 식별하는 것은, 자연 언어 스피치 입력에 기초하여 텍스트 문자열을 획득하는 것; 사용자 의도의 표현을 획득하기 위해 텍스트 문자열을 해석하는 것; 및 사용자 의도의 표현에 기초하여 태스크를 결정하는 것을 포함한다.

- [0348] 일부 예들에서, 자연 언어 스피치 입력에 기초하여 태스크를 식별하는 것은 하나 이상의 미디어 아이템들을 제공하는 태스크를 식별하는 것을 포함한다. 일부 예들에서, 전자 디바이스는 스피치 입력에 기초하여 미디어 아이템(이후, "제2 미디어 아이템")을 식별하고 미디어 아이템에 대응하는 정보를 획득한다(예를 들어, 식별된 태스크를 수행함으로써). 일부 예들에서, 제2 미디어 아이템은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 조합을 포함한다. 일부 다른 예들에서, 전자 디바이스는 하나 이상의 결과들(예를 들어, 검색 결과들)을 획득하기 위해 태스크를 수행한다.
- [0349] 블록(1306)에서, 전자 디바이스(또는 전자 디바이스의 디지털 어시스턴트)는 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력을 제공한다. 일부 예들에서, 스피치 출력을 제공하는 것은 제2 미디어 아이템의 구두 설명을 제공하는 것을 포함한다. 일부 예들에서, 스피치 출력은 디지털 어시스턴트의 음성, 제2 미디어 아이템과 연관된 음성, 또는 이들의 조합으로 제공된다. 일부 예들에서, 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력을 제공하는 것은 하나 이상의 결과들(예를 들어, 검색 결과들) 중 일 결과에 대한 구두 설명을 나타내는 스피치 출력을 제공하는 것을 포함한다.
- [0350] 블록(1308)에서, 구두 응답을 나타내는 스피치 출력을 제공하는 동안, 전자 디바이스(또는 전자 디바이스의 디지털 어시스턴트)는 구두 응답에 대응하는 미디어 아이템(이하, "제1 미디어 아이템")의 재생을 제공한다. 일부 예들에서, 재생된 미디어 아이템은 제2 미디어 아이템의 일부분에 대응한다. 예를 들어, 재생된 미디어 아이템은 제2 미디어 아이템의 대표 샘플이다.
- [0351] 일부 예들에서, 제1 미디어 아이템의 재생을 제공하는 동안, 전자 디바이스는 제2 자연 언어 스피치 입력을 수신한다. 제2 자연 언어 스피치 입력을 수신하는 것에 응답하여, 전자 디바이스는 제2 미디어 아이템의 재생을 제공한다. 일부 예들에서, 제2 미디어 아이템의 재생은 제1 미디어 아이템의 재생과 상이한 볼륨에서 제공된다.
- [0352] 일부 예들에서, 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력은 제1 스피치 출력이다. 제2 미디어 아이템의 재생을 제공하는 동안, 전자 디바이스는 제2 스피치 출력을 제공한다.
- [0353] 일부 예들에서, 구두 응답에 대응하는 미디어 아이템의 재생을 제공하는 것은, 결과에 대응하는 사운드 효과의 재생을 제공하는 것을 포함한다. 일부 예들에서, 식별된 태스크와 연관된 구두 응답을 나타내는 스피치 출력은 제1 볼륨에서 제공되고, 미디어 아이템(예를 들어, 사운드 효과)의 재생은 제1 볼륨과 상이한 제2 볼륨에서 제공된다.
- [0354] 일부 예들에서, 전자 디바이스는 컴퓨터, 셋톱 박스, 스피커, 스마트 워치, 전화, 또는 이들의 조합이다.
- [0355] 도 13에 관하여 상술한 동작들은, 옵션적으로, 도 1 내지 도 4, 도 6a와 도 6b, 및 도 7a 내지 도 7c에 도시된 컴포넌트들에 의해 구현된다. 예를 들어, 프로세스(1300)의 동작들은, 본 명세서에 기술된, 디바이스들(104, 200, 400, 600, 800, 900, 1000, 1100)을 포함하지만 이로 한정되지는 않는 임의의 디바이스 또는 이들의 컴포넌트에 의해 구현될 수 있다. 다른 프로세스들이 도 1 내지 도 4, 도 6a와 도 6b, 및 도 7a 내지 도 7c에 도시된 컴포넌트들에 기초하여 어떻게 구현되는지는 본 발명이 속한 기술분야에서 통상의 지식을 가진 자에게 자명할 것이다.
- [0356] 도 14는 다양한 예들에 따른, 디지털 어시스턴트의 청각 기반 인터페이스를 제공하기 위한 프로세스(1400)를 도시한다. 프로세스(1400)는 예를 들어, 디지털 어시스턴트를 구현하는 하나 이상의 전자 디바이스들을 사용하여 수행된다. 일부 예들에서, 프로세스(1400)는 클라이언트-서버 시스템(예컨대, 시스템(100))을 사용하여 수행되고, 프로세스(1400)의 블록들은 서버(예컨대, DA 서버(106))와 클라이언트 디바이스 사이에서 임의의 방식으로 분담된다. 다른 예들에서, 프로세스(1400)의 블록들은 서버와 다수의 클라이언트 디바이스들(예컨대, 모바일 전화기 및 스마트 워치) 사이에서 분담된다. 따라서, 프로세스(1400)의 일부분들이 클라이언트-서버 시스템의 특정 디바이스들에 의해 수행되는 것으로 본 명세서에 기술되지만, 프로세스(1400)는 그렇게 제한되지 않는다는 것이 이해될 것이다. 다른 예들에서, 프로세스(1400)는 단지 하나의 클라이언트 디바이스(예컨대, 사용자 디바이스(104))만을 또는 다수의 클라이언트 디바이스들만을 사용하여 수행된다. 프로세스(1400)에서, 일부 블록들은, 옵션적으로, 조합되고, 일부 블록들의 순서는, 옵션적으로, 변경되고, 일부 블록들은, 옵션적으로, 생략된다. 일부 예들에서, 추가 단계들이 프로세스(1400)와 조합하여 수행될 수 있다.

- [0357] 블록(1402)에서, 전자 디바이스는 미디어에 대한 요청을 나타내는 스피치 입력을 수신한다. 일부 예들에서, 전자 디바이스는 미디어에 대한 요청을 나타내는 스피치 입력에 기초하여, 텍스트 문자열을 획득한다. 일부 예들에서, 전자 디바이스는 추가로, 획득된 텍스트 문자열에 기초하여, 사용자 의도의 표현을 결정하고, 사용자 의도의 표현에 기초하여, 하나 이상의 미디어 아이템들에 관한 정보를 획득한다.
- [0358] 블록(1404)에서, 스피치 입력을 수신하는 것에 응답하여, 전자 디바이스(또는 전자 디바이스의 디지털 어시스턴트)는 제1 미디어 아이템의 제안을 나타내는 오디오 출력을 제공한다. 일부 예들에서, 제1 미디어 아이템은 하나 이상의 미디어 아이템들의 일부이다. 일부 예들에서, 제1 미디어 아이템은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 임의의 조합이다.
- [0359] 블록(1406)에서, 전자 디바이스(또는 전자 디바이스의 디지털 어시스턴트)는 미디어에 대한 요청에 대응하는 다수의 연속적인 비-공정 응답들이 임계치를 만족시키는지 여부를 결정한다.
- [0360] 일부 예들에서, 미디어에 대한 요청을 나타내는 스피치 입력은 제1 스피치 입력이다. 일부 예들에서, 미디어에 대한 요청에 대응하는 연속적인 비-공정 응답들의 수가 임계치를 만족시키는지 여부를 결정하는 것은, 제1 미디어 아이템의 제안을 나타내는 오디오 출력을 제공한 후에, 제2 스피치 입력을 수신하는 것; 및 제2 스피치 입력이 미디어에 대한 요청에 대응하는 비-공정 응답을 나타내는지 여부를 결정하는 것을 포함한다. 또한, 제2 스피치 입력이 비-공정 응답을 나타내지 않는다는 결정에 따라, 전자 디바이스는 요청에 대응하는 연속적인 비-공정 응답들의 수를 업데이트한다. 제2 스피치 입력이 비-공정 응답을 나타내지 않는다는 결정에 따라, 전자 디바이스는 요청에 대응하는 연속적인 비-공정 응답들의 수를 업데이트하는 것을 중지한다.
- [0361] 일부 예들에서, 제2 스피치 입력이 미디어에 대한 요청에 대한 비-공정 응답을 나타내는지 여부를 결정하는 것은, 제2 스피치 입력이 거부를 나타내는지 여부를 결정하는 것을 포함한다.
- [0362] 일부 예들에서, 미디어에 대한 요청에 대응하는 연속적인 비-공정 응답들의 수가 임계치를 만족시키는지 여부를 결정하는 것은, 제1 미디어 아이템의 제안을 나타내는 오디오 출력을 제공한 후에, 요청에 대응하는 응답이 미리 정의된 기간 내에 수신되지 않는다고 결정하는 것을 포함한다. 미디어에 대한 요청에 대응하는 연속적인 비-공정 응답들의 수가 임계치를 만족시키는지 여부를 결정하는 것은 요청에 대응하는 연속적인 비-공정 응답들의 수를 업데이트하는 것을 추가로 포함한다.
- [0363] 블록(1408)에서, 연속적인 비-공정 응답들의 수가 임계치를 만족시키지 못한다는 결정에 따라, 전자 디바이스(또는 전자 디바이스의 디지털 어시스턴트)는 제1 미디어 아이템과 상이한 제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공한다. 일부 예들에서, 제2 미디어 아이템은 하나 이상의 미디어 아이템들의 일부이다. 일부 예들에서, 제2 미디어 아이템은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 임의의 조합이다.
- [0364] 블록(1410)에서, 연속적인 비-공정 응답들의 수가 임계치를 만족시킨다는 결정에 따라, 전자 디바이스(또는 전자 디바이스의 디지털 어시스턴트)는 제2 미디어 아이템의 제안을 나타내는 오디오 출력을 제공하는 것을 중지하고; 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공한다.
- [0365] 일부 예들에서, 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공하는 것은, 디지털 어시스턴트에 의해, 디지털 어시스턴트에 의해 이전에 제안된 복수의 미디어 아이템들 중에서 사용자 선택에 대한 추구를 나타내는 스피치 출력을 제공하는 것을 포함한다. 일부 예들에서, 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공한 후에, 전자 디바이스는 사용자 선택을 나타내는 스피치 입력을 수신하고, 컨텍스트 정보에 기초하여 사용자 선택을 나타내는 스피치 입력을 해석한다. 일부 예들에서, 컨텍스트 정보는 디지털 어시스턴트에 의해 이전에 제안된 복수의 미디어 아이템들을 포함한다.
- [0366] 일부 예들에서, 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공하는 것은, 디지털 어시스턴트에 의해, 미디어에 대한 요청에 대한 하나 이상의 파라미터들에 대한 추구를 나타내는 스피치 출력을 제공하는 것을 포함한다. 일부 예들에서, 사용자 입력에 대한 요청을 나타내는 오디오 출력을 제공한 후에, 전자 디바이스는 미디어에 대한 요청에 대한 하나 이상의 파라미터들을 나타내는 스피치 입력을 수신한다. 일부 예들에서, 전자 디바이스는 하나 이상의 파라미터들에 기초하여 제3 미디어 아이템을 획득한다. 제3 미디어 아이템은 제1 미디어 아이템 및 제2 미디어 아이템과 상이하다.
- [0367] 일부 예들에서, 전자 디바이스는 컴퓨터, 셋톱 박스, 스피커, 스마트 워치, 전화, 또는 이들의 조합이다.
- [0368] 도 14에 관하여 상술한 동작들은, 옵션적으로, 도 1 내지 도 4, 도 6a와 도 6b, 및 도 7a 내지 도 7c에 도시된

컴포넌트들에 의해 구현된다. 예를 들어, 프로세스(1400)의 동작들은, 본 명세서에 기술된, 디바이스들(104, 200, 400, 600, 800, 900, 1000, 1100)을 포함하지만 이로 한정되지 않는 임의의 디바이스 또는 이들의 컴포넌트에 의해 구현될 수 있다. 다른 프로세스들이 도 1 내지 도 4, 도 6a와 도 6b, 및 도 7a 내지 도 7c에 도시된 컴포넌트들에 기초하여 어떻게 구현되는지는 본 발명이 속한 기술분야에서 통상의 지식을 가진 자에게 자명할 것이다.

- [0369] 도 15는 다양한 예들에 따른, 디지털 어시스턴트의 청각 기반 인터페이스를 제공하기 위한 프로세스(1500)를 도시한다. 프로세스(1500)는 예를 들어, 디지털 어시스턴트를 구현하는 하나 이상의 전자 디바이스들을 사용하여 수행된다. 일부 예들에서, 프로세스(1500)는 클라이언트-서버 시스템(예컨대, 시스템(100))을 사용하여 수행되고, 프로세스(1500)의 블록들은 서버(예컨대, DA 서버(106))와 클라이언트 디바이스 사이에서 임의의 방식으로 분담된다. 다른 예들에서, 프로세스(1500)의 블록들은 서버와 다수의 클라이언트 디바이스들(예컨대, 모바일 전화기 및 스마트 워치) 사이에서 분담된다. 따라서, 프로세스(1500)의 일부분들이 클라이언트-서버 시스템의 특정 디바이스들에 의해 수행되는 것으로 본 명세서에 기술되지만, 프로세스(1500)는 그렇게 제한되지 않는다는 것이 이해될 것이다. 다른 예들에서, 프로세스(1500)는 단지 하나의 클라이언트 디바이스(예컨대, 사용자 디바이스(104))만을 또는 다수의 클라이언트 디바이스들만을 사용하여 수행된다. 프로세스(1500)에서, 일부 블록들은, 옵션적으로, 조합되고, 일부 블록들의 순서는, 옵션적으로, 변경되고, 일부 블록들은, 옵션적으로, 생략된다. 일부 예들에서, 추가 단계들이 프로세스(1500)와 조합하여 수행될 수 있다.
- [0370] 블록(1502)에서, 전자 디바이스는 미디어에 대한 요청을 나타내는 스피치 입력을 수신한다. 블록(1504)에서, 전자 디바이스(또는 전자 디바이스의 디지털 어시스턴트)는 전자 디바이스에 대한 복수의 사용자들의 물리적 존재를 검출한다. 일부 예들에서, 전자 디바이스는 복수의 사용자들 중 제1 사용자와 연관된 제1 전자 디바이스이다. 또한, 전자 디바이스에 대한 복수의 사용자들의 물리적 존재를 검출하는 것은, 복수의 사용자들 중 제2 사용자와 연관된 제2 전자 디바이스에 대응하는 정보를 수신하는 것을 포함한다.
- [0371] 일부 예들에서, 제2 전자 디바이스에 대응하는 정보를 수신하는 것은, 제1 전자 디바이스에서, 제2 전자 디바이스로부터 식별 정보를 수신하는 것을 포함한다. 일부 예들에서, 제2 전자 디바이스에 대응하는 정보를 수신하는 것은, 제1 전자 디바이스 및 제2 전자 디바이스에 접속된 라우팅 디바이스로부터 식별 정보를 수신하는 것을 포함한다.
- [0372] 블록(1506)에서, 복수의 사용자들의 물리적 존재를 검출하는 것에 응답하여, 전자 디바이스는 복수의 사용자들에 대응하는 복수의 선호도 프로파일들을 획득한다. 일부 예들에서, 전자 디바이스는 원격 디바이스로부터 제2 사용자에게 대응하는 선호도 프로파일을 수신한다. 다른 예들에서, 제2 사용자에게 대응하는 선호도 프로파일은 전자 디바이스 상에 저장된다.
- [0373] 블록(1508)에서, 전자 디바이스(또는 전자 디바이스의 디지털 어시스턴트)는 복수의 선호도 프로파일들에 기초하여 병합된 선호도 프로파일을 제공한다. 일부 예들에서, 병합된 선호도 프로파일을 제공하는 것은, 복수의 선호도 프로파일들 각각에 의해 공유되는 하나 이상의 선호도들을 식별하는 것을 포함한다.
- [0374] 블록(1510)에서, 전자 디바이스(또는 전자 디바이스의 디지털 어시스턴트)는 병합된 선호도 프로파일에 기초하여 미디어 아이템을 식별한다. 일부 예들에서, 식별된 미디어 아이템은 하나 이상의 선호도들과 매칭되는 메타데이터와 연관된다. 식별된 미디어 아이템은 노래, 오디오 북, 팟캐스트, 스테이션, 재생목록, 또는 이들의 임의의 조합일 수 있다.
- [0375] 일부 예들에서, 병합된 선호도 프로파일에 기초하여 미디어 아이템을 식별하는 것은, 복수의 미디어 아이템들로부터 미디어 아이템을 식별하는 것을 포함한다. 복수의 미디어 아이템들은 제1 사용자와 연관된 미디어 아이템들의 제1 세트 및 제2 사용자와 연관된 미디어 아이템들의 제2 세트를 포함한다. 일부 예들에서, 식별된 미디어 아이템은 미디어 아이템들의 제1 세트의 일부가 아니지만, 미디어 아이템들의 제2 세트의 일부이다.
- [0376] 블록(1512)에서, 전자 디바이스(또는 전자 디바이스의 디지털 어시스턴트)는 식별된 미디어 아이템을 포함하는 오디오 출력을 제공한다. 일부 예들에서, 오디오 출력은 식별된 미디어 아이템의 구두 설명을 포함한다. 일부 예들에서, 오디오 출력은 병합된 프로파일을 나타내는 스피치 출력을 포함한다.
- [0377] 일부 예들에서, 복수의 사용자들의 물리적 존재를 검출한 후에, 전자 디바이스(또는 전자 디바이스의 디지털 어시스턴트)는 제2 사용자의 부재를 검출한다. 제2 사용자의 부재를 검출한 후, 전자 디바이스는 복수의 미디어 아이템들을 업데이트하고 병합된 선호도 프로파일을 업데이트한다. 일부 예들에서, 복수의 미디어 아이템들을 업데이트하는 것은 복수의 미디어 아이템들로부터 미디어 아이템을 제거하는 것을 포함한다.

- [0378] 일부 예들에서, 전자 디바이스는 컴퓨터, 셋톱 박스, 스피커, 스마트 워치, 전화, 또는 이들의 조합이다.
- [0379] 도 15에 관하여 상술한 동작들은, 옵션적으로, 도 1 내지 도 4, 도 6a와 도 6b, 및 도 7a 내지 도 7c에 도시된 컴포넌트들에 의해 구현된다. 예를 들어, 프로세스(1500)의 동작들은, 본 명세서에 기술된, 디바이스들(104, 200, 400, 600, 800, 900, 1000, 1100)을 포함하지만 이로 한정되지는 않는 임의의 디바이스 또는 이들의 컴포넌트에 의해 구현될 수 있다. 다른 프로세스들이 도 1 내지 도 4, 도 6a와 도 6b, 및 도 7a 내지 도 7c에 도시된 컴포넌트들에 기초하여 어떻게 구현되는지는 본 발명이 속한 기술분야에서 통상의 지식을 가진 자에게 자명할 것이다.
- [0380] 일부 구현예들에 따르면, 전자 디바이스의 하나 이상의 프로세서들에 의한 실행을 위한 하나 이상의 프로그램들을 저장하는 컴퓨터 판독가능 저장 매체(예를 들어, 비일시적 컴퓨터 판독가능 저장 매체)가 제공되며, 여기서 하나 이상의 프로그램들은 본 명세서에 기술된 방법들 또는 프로세스들 중 임의의 것을 수행하기 위한 명령어들을 포함한다.
- [0381] 일부 구현예들에 따르면, 본 명세서에 기술된 방법들 또는 프로세스들 중 임의의 것을 수행하기 위한 수단을 포함하는 전자 디바이스(예를 들어, 휴대용 전자 디바이스)가 제공된다.
- [0382] 일부 구현예들에 따라, 본 명세서에 기술된 방법들 또는 프로세스들 중 임의의 것을 수행하도록 구성된 프로세스 유닛을 포함하는 전자 디바이스(예를 들어, 휴대용 전자 디바이스)가 제공된다.
- [0383] 일부 구현예들에 따르면, 하나 이상의 프로세서들 및 하나 이상의 프로세서들에 의한 실행을 위한 하나 이상의 프로그램들을 저장하는 메모리를 포함하는 전자 디바이스(예를 들어, 휴대용 전자 디바이스)가 제공되며, 여기서 하나 이상의 프로그램들은 본 명세서에 기술된 방법들 또는 프로세스들 중 임의의 것을 실행하기 위한 명령어들을 포함한다.
- [0384] 전술한 설명은, 설명의 목적을 위해, 특정 실시예들을 참조하여 기술되었다. 그러나, 상기의 예시적인 논의들은 본 발명을 개시된 정확한 형태들로 규명하거나 제한하려는 의도는 아니다. 많은 수정들 및 변형들이 상기 교시 내용들의 관점에서 가능하다. 실시예들은 기술들의 원리 및 그것들의 실제적인 응용을 가장 잘 설명하기 위하여 선택되고 기술되었다. 따라서, 본 발명이 속한 기술분야에서 통상의 지식을 가진 자들은 고려된 특정 사용에 적합한 바와 같이 다양한 수정을 이용하여 기술들 및 다양한 실시예들을 최상으로 활용하는 것이 가능하게 된다.
- [0385] 본 개시내용 및 예들이 첨부된 도면들을 참조하여 충분히 기술되었지만, 본 발명이 속한 기술분야에서 통상의 지식을 가진 자에게 다양한 변경들 및 수정들이 명백할 것이라는 것에 주목하여야 한다. 그러한 변경들 및 수정들은 청구항들에 의해 정의되는 바와 같은 개시내용 및 예들의 범주 내에 포함되는 것으로 이해되어야 한다.
- [0386] 상술한 바와 같이, 본 기술의 일 양태는, 초청 콘텐츠 또는 사용자가 관심을 가질만한 임의의 다른 콘텐츠의 사용자에게로의 전달을 향상시키기 위한, 다양한 소스들로부터 입수가능한 데이터의 수집 및 사용이다. 본 개시내용은, 일부 경우들에 있어서, 이러한 수집된 데이터가 특정 개인을 고유하게 식별하거나 또는 그와 연락하거나 그의 위치를 확인하는 데 이용될 수 있는 개인 정보 데이터를 포함할 수 있음을 고려한다. 그러한 개인 정보 데이터는 인구통계 데이터, 위치 기반 데이터, 전화 번호들, 이메일 주소들, 집 주소들, 또는 임의의 다른 식별 정보를 포함할 수 있다.
- [0387] 본 개시내용은 본 기술에서의 그러한 개인 정보 데이터의 이용이 사용자들에게 이득을 주기 위해 사용될 수 있음을 인식한다. 예를 들어, 개인 정보 데이터는 더 큰 관심이 있는 타겟 콘텐츠를 사용자에게 전달하는 데 이용될 수 있다. 따라서, 그러한 개인 정보 데이터의 이용은 전달된 콘텐츠의 계산된 제어를 가능하게 한다. 게다가, 사용자에게 이득을 주는 개인 정보 데이터에 대한 다른 이용들이 또한 본 개시내용에 의해 고려된다.
- [0388] 본 개시내용은 그러한 개인 정보 데이터의 수집, 분석, 공개, 전달, 저장, 또는 다른 이용을 책임지고 있는 엔티티들이 잘 확립된 프라이버시 정책들 및/또는 프라이버시 관례들을 준수할 것이라는 것을 추가로 고려한다. 특히, 그러한 엔티티들은, 대체로 개인 정보 데이터를 사적이고 안전하게 유지시키기 위한 산업적 또는 행정적 요건들을 만족시키거나 넘어서는 것으로 인식되는 프라이버시 정책들 및 관례들을 구현하고 지속적으로 이용해야 한다. 예를 들어, 사용자들로부터의 개인 정보는 엔티티의 적법하며 적절한 사용들을 위해 수집되어야 하고, 이들 적법한 사용들을 벗어나서 공유되거나 판매되지 않아야 한다. 또한, 그러한 수집은 단지 사용자들의 고지에 입각한 동의를 수신한 후에만 발생해야 한다. 추가적으로, 그러한 엔티티들은 그러한 개인 정보 데이터에 대한 액세스를 보호하고 안전하게 하며 개인 정보 데이터에 대한 액세스를 갖는 다른 사람들이 그들의

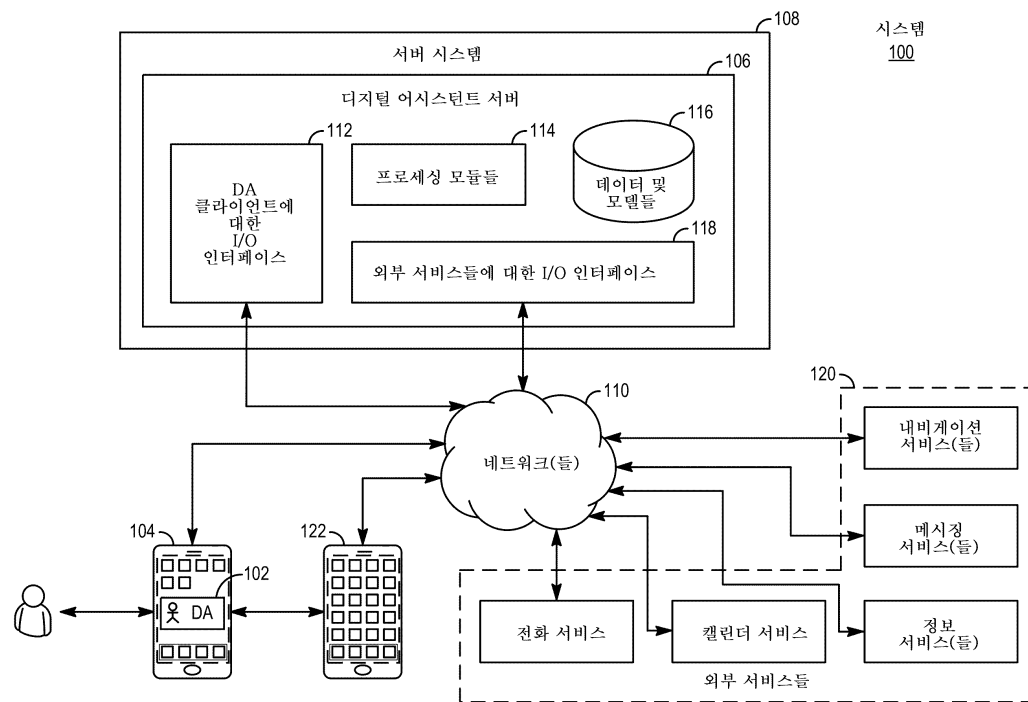
프라이버시 정책들 및 절차들을 고수한다는 것을 보장하기 위한 임의의 필요한 단계들을 취할 것이다. 게다가, 그러한 엔티티들은 널리 인정된 프라이버시 정책들 및 관례들에 대한 그들의 고수를 증명하기 위해 제3자들에 의해 그들 자신들이 평가를 받을 수 있다.

[0389] 전술한 것에도 불구하고, 본 개시내용은 또한 사용자가 개인 정보 데이터의 사용, 또는 그에 대한 액세스를 선택적으로 차단하는 실시예들을 고려한다. 즉, 본 개시내용은 그러한 개인 정보 데이터에 대한 액세스를 방지하거나 차단하기 위해 하드웨어 및/또는 소프트웨어 요소들이 제공될 수 있다는 것을 고려한다. 예를 들어, 광고 전달 서비스들의 경우에, 본 기술은 사용자들이 서비스를 위한 등록 중에 개인 정보 데이터의 수집 시의 참여의 "동의함" 또는 "동의하지 않음"을 선택하는 것을 허용하도록 구성될 수 있다. 다른 예에서, 사용자들은 타깃 콘텐츠 전달 서비스들을 위한 위치 정보를 제공하지 않도록 선택할 수 있다. 또 다른 예에서, 사용자들은 정확한 위치 정보를 제공하지 않지만 위치 구역 정보의 전달을 허용하도록 선택할 수 있다.

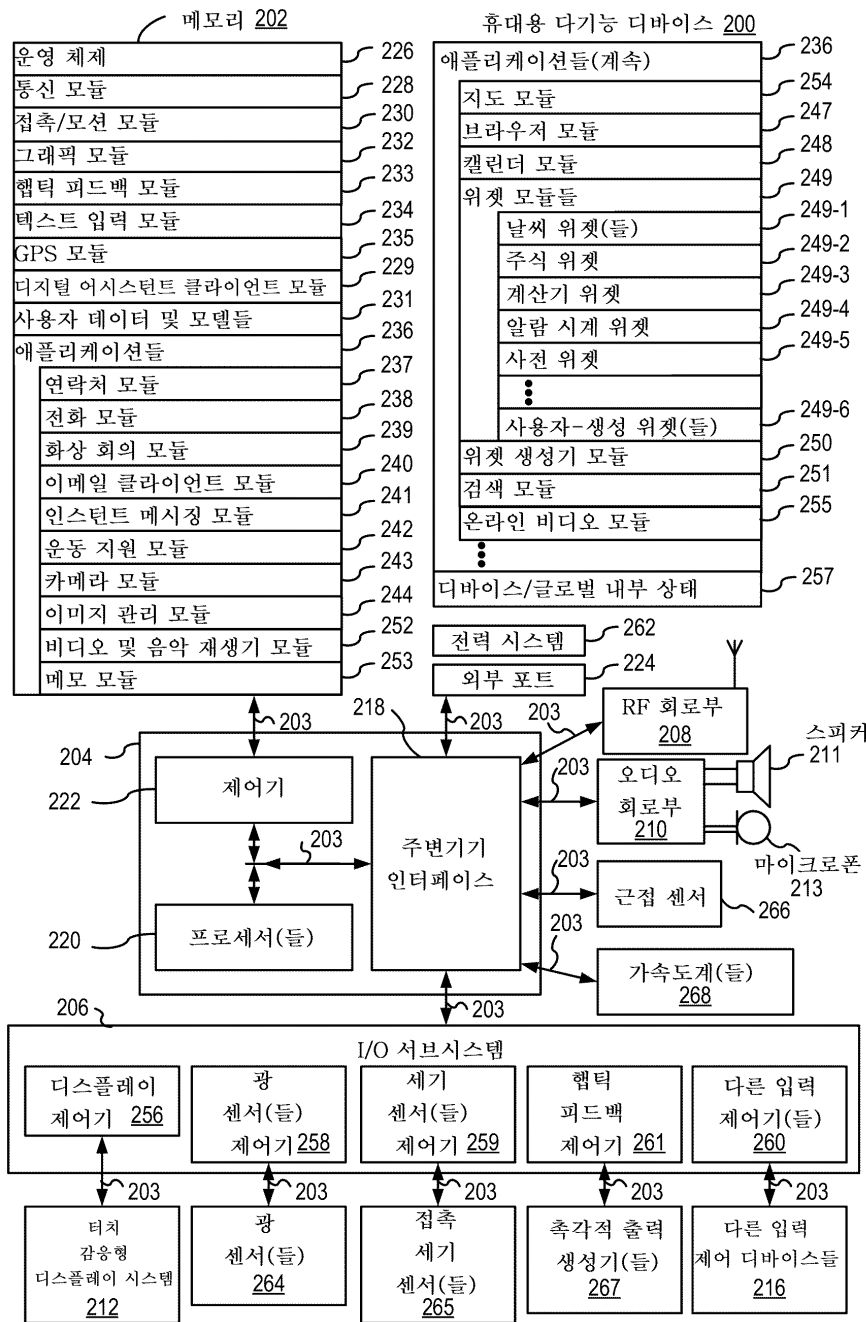
[0390] 따라서, 본 개시내용이 하나 이상의 다양한 개시된 실시예들을 구현하기 위해 개인 정보 데이터의 사용을 광범위하게 커버하지만, 본 개시내용은 다양한 실시예들이 또한 그러한 개인 정보 데이터에 액세스할 필요 없이 구현될 수 있다는 것을 또한 고려한다. 즉, 본 기술의 다양한 실시예들은 그러한 개인 정보 데이터의 모두 또는 일부분의 결여로 인해 동작 불가능하게 되지 않는다. 예를 들어, 콘텐츠는, 사용자와 연관된 디바이스에 의해 요청되는 콘텐츠, 콘텐츠 전달 서비스들에 대해 이용가능한 다른 비-개인 정보, 또는 공개적으로 입수가능한 정보와 같은 비-개인 정보 데이터 또는 최소량의 개인 정보에 기초하여 선호도를 추론함으로써 선택되고 사용자들에게 전달될 수 있다.

도면

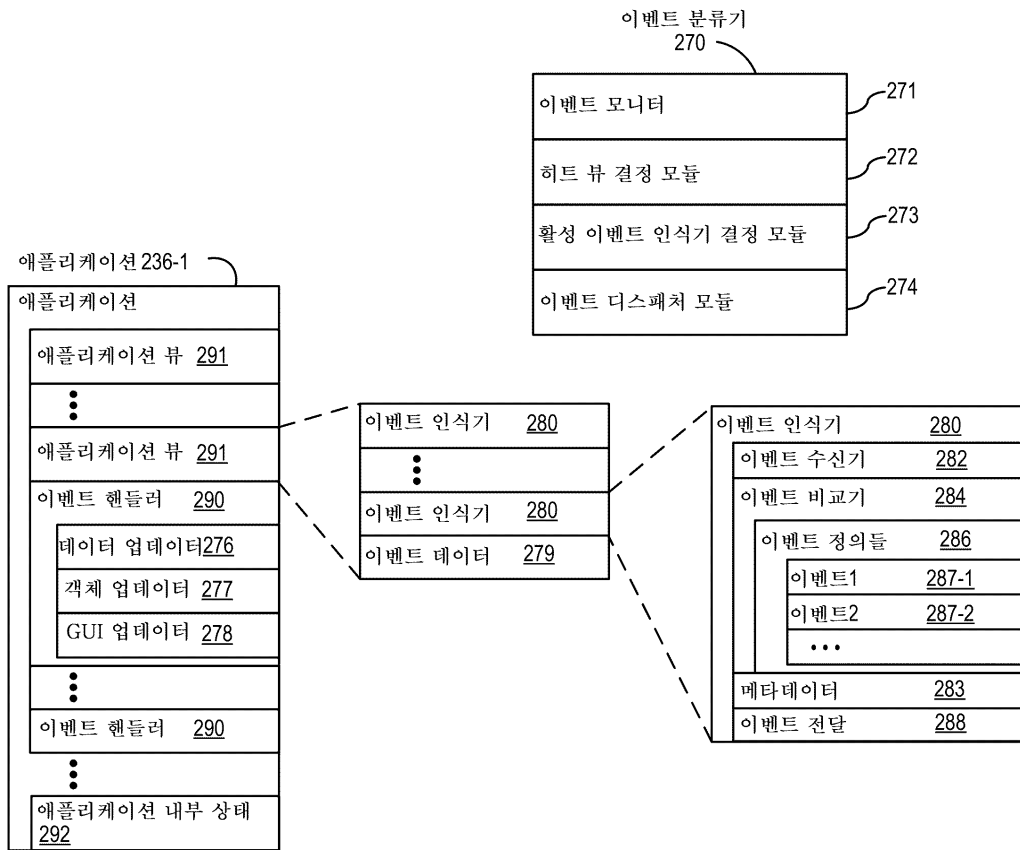
도면1



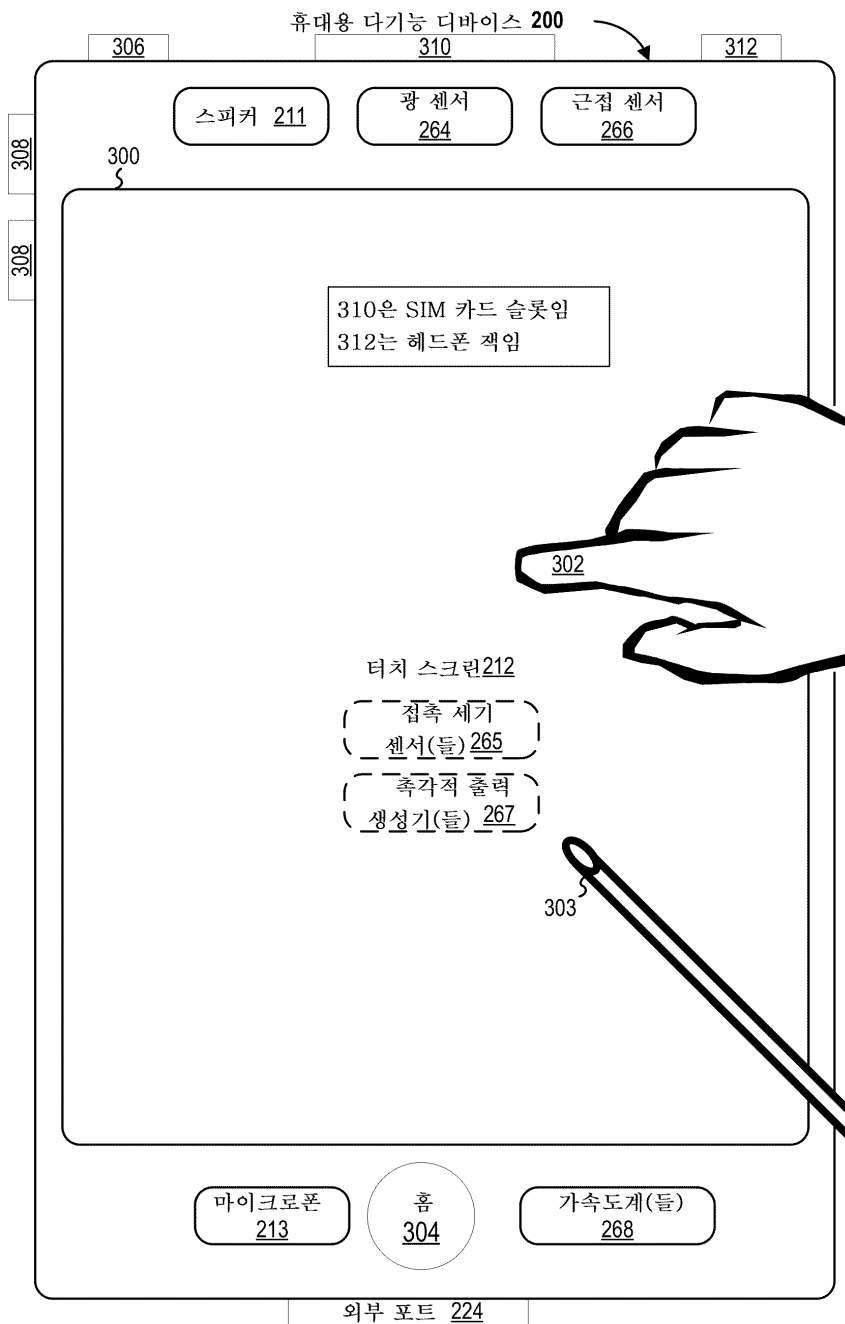
도면2a



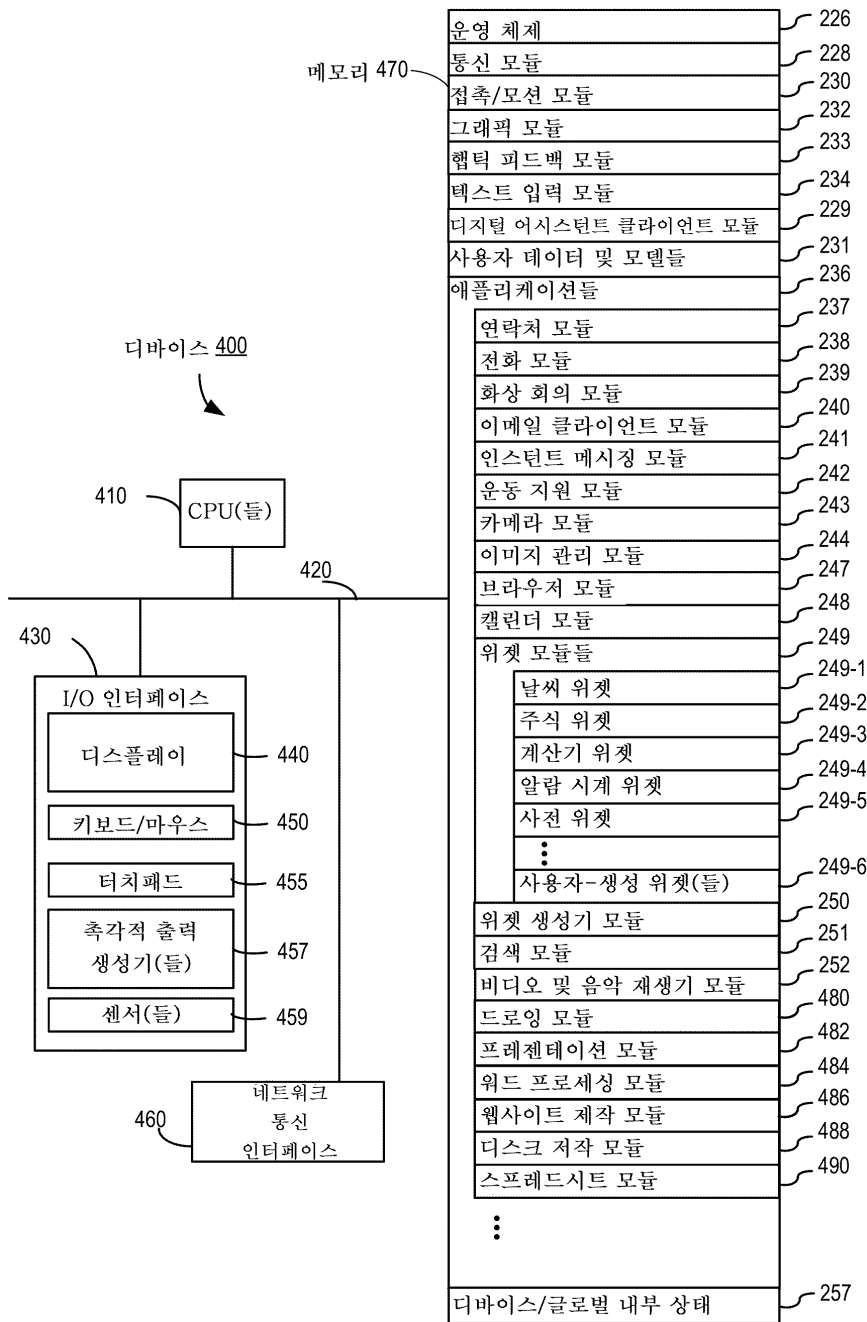
도면2b



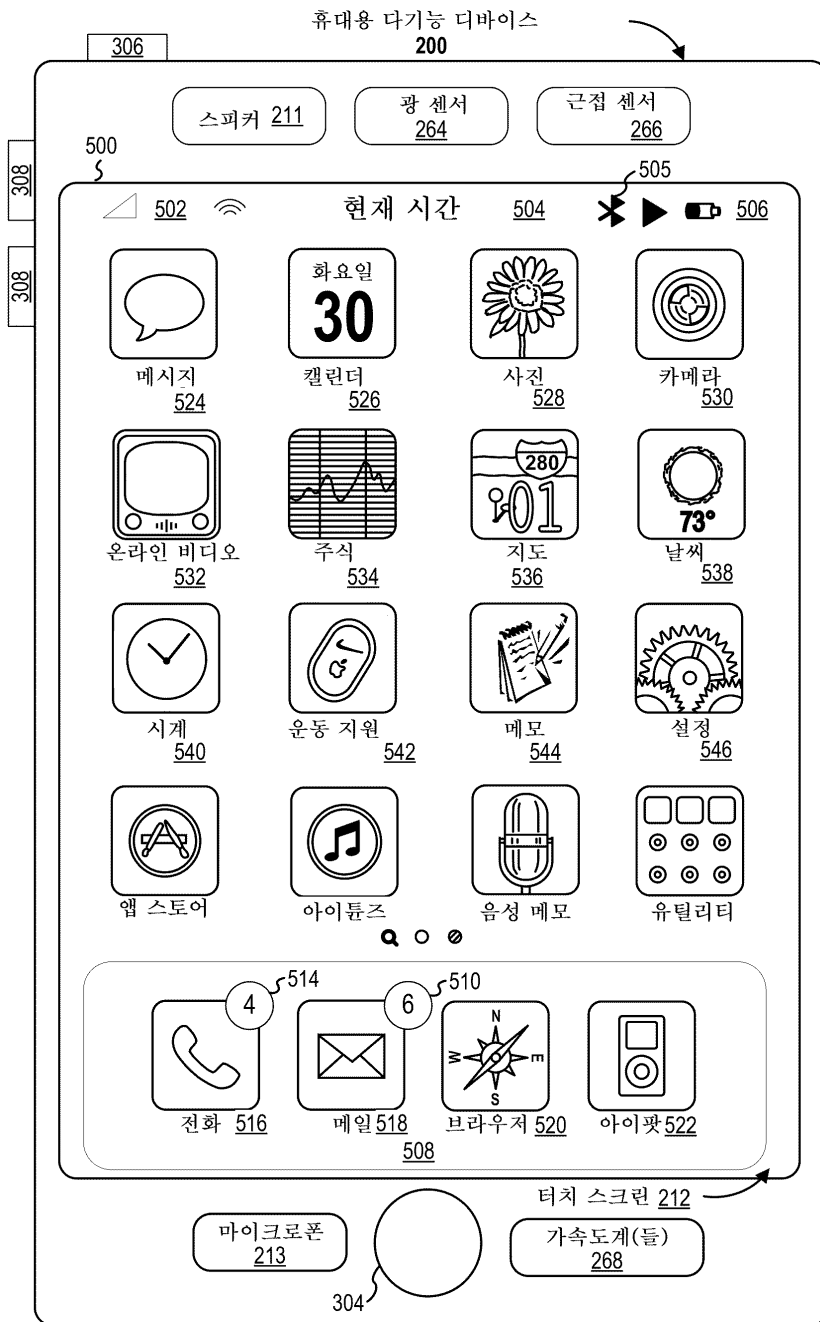
도면3



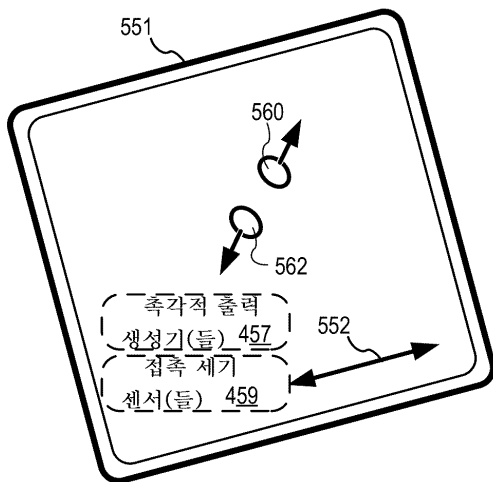
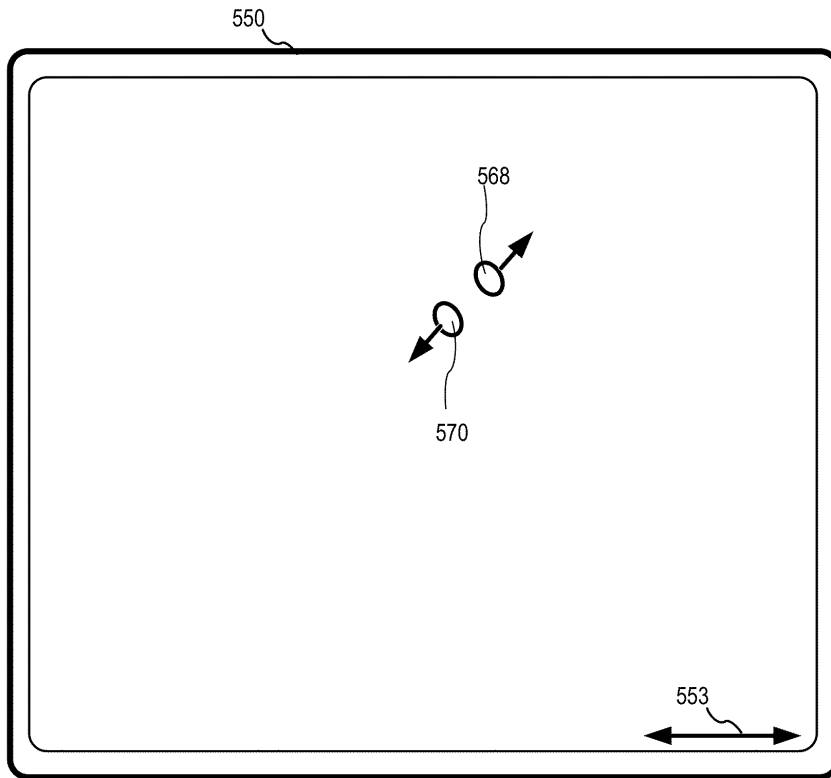
도면4



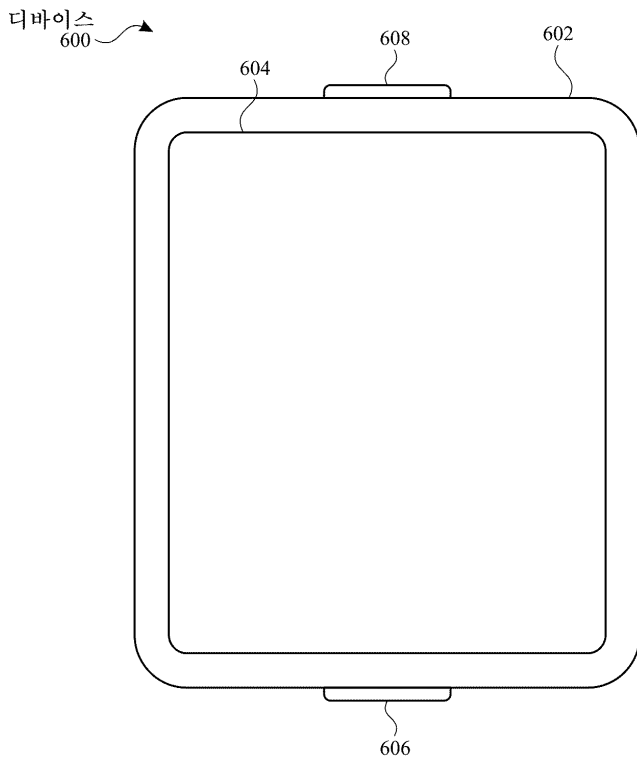
도면5a



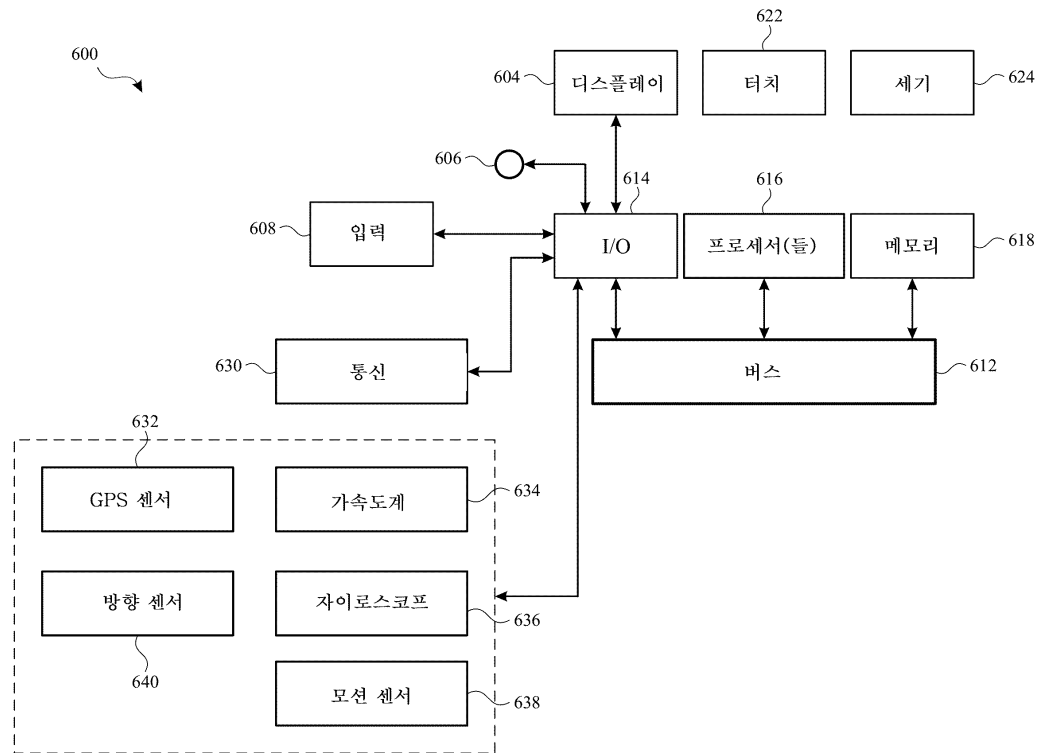
도면5b



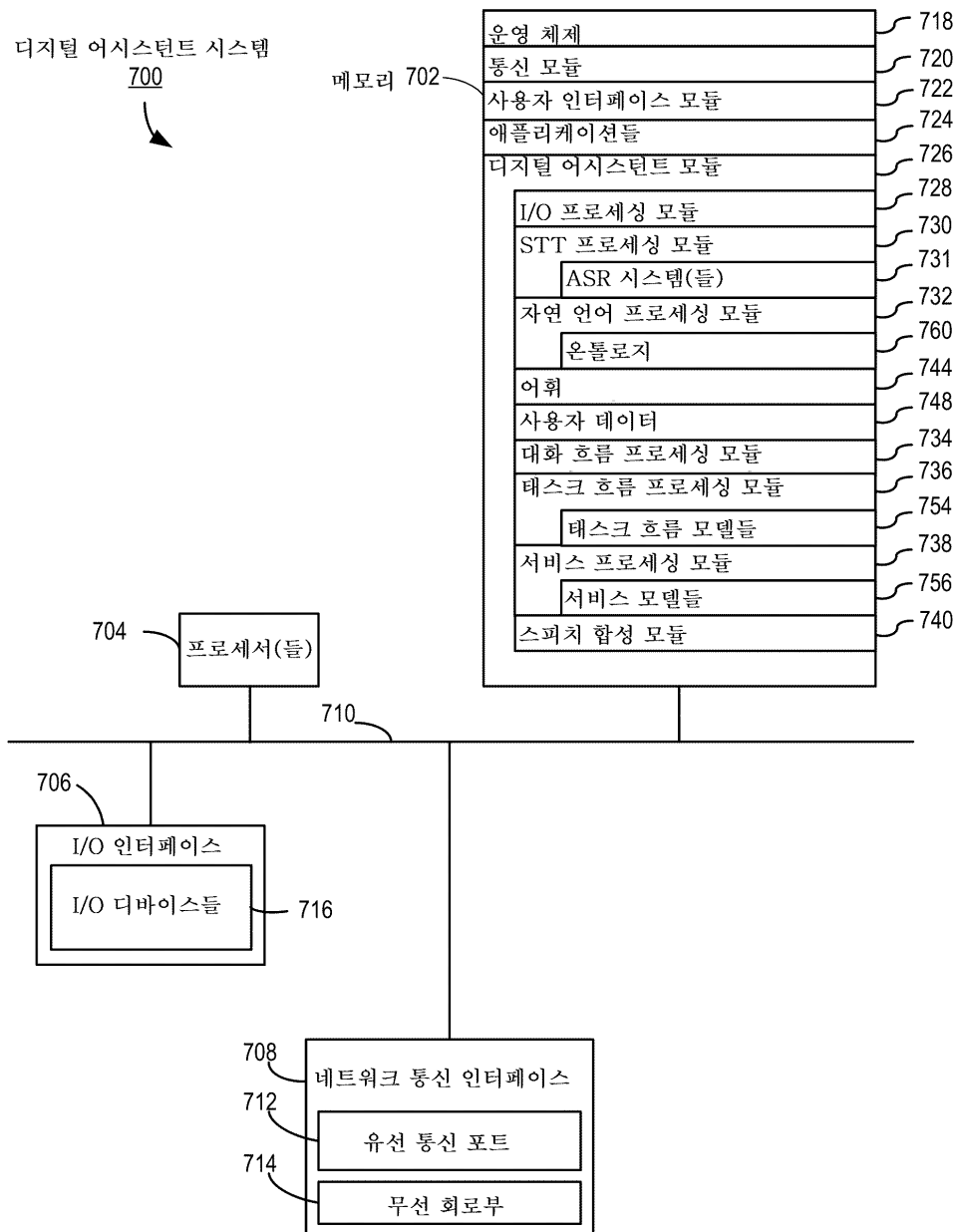
도면6a



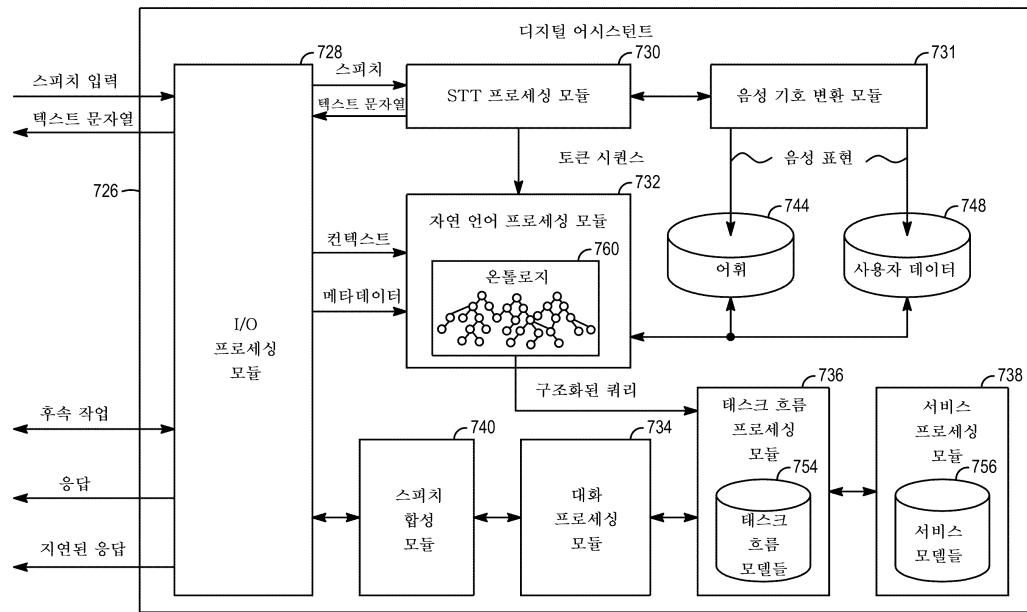
도면6b



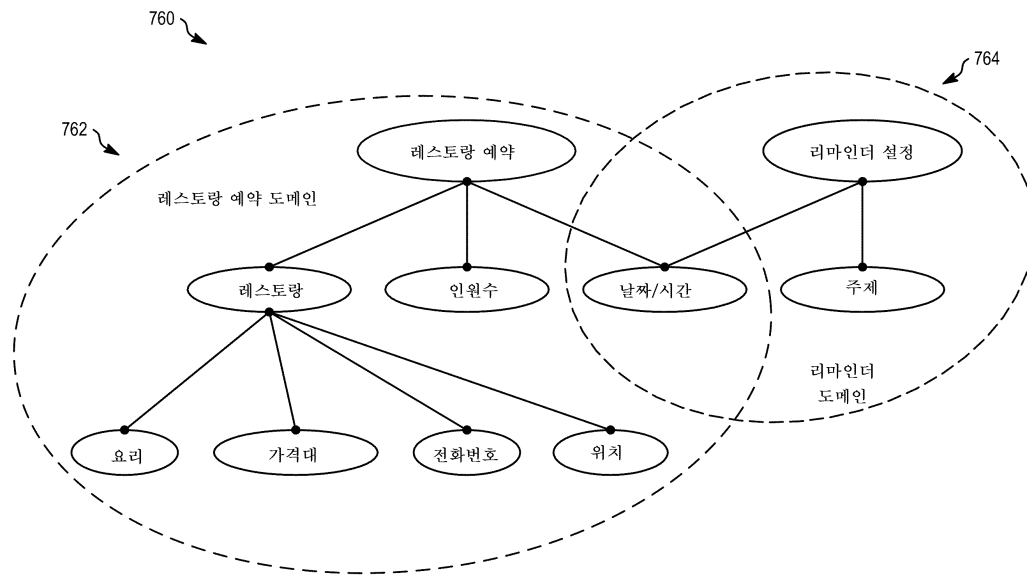
도면7a



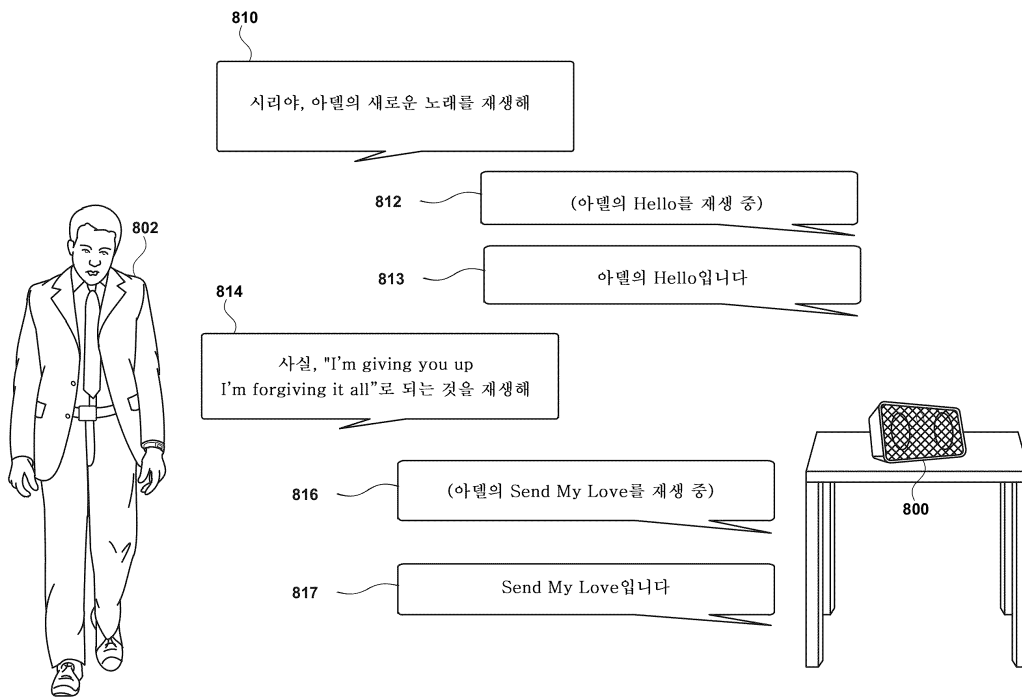
도면7b



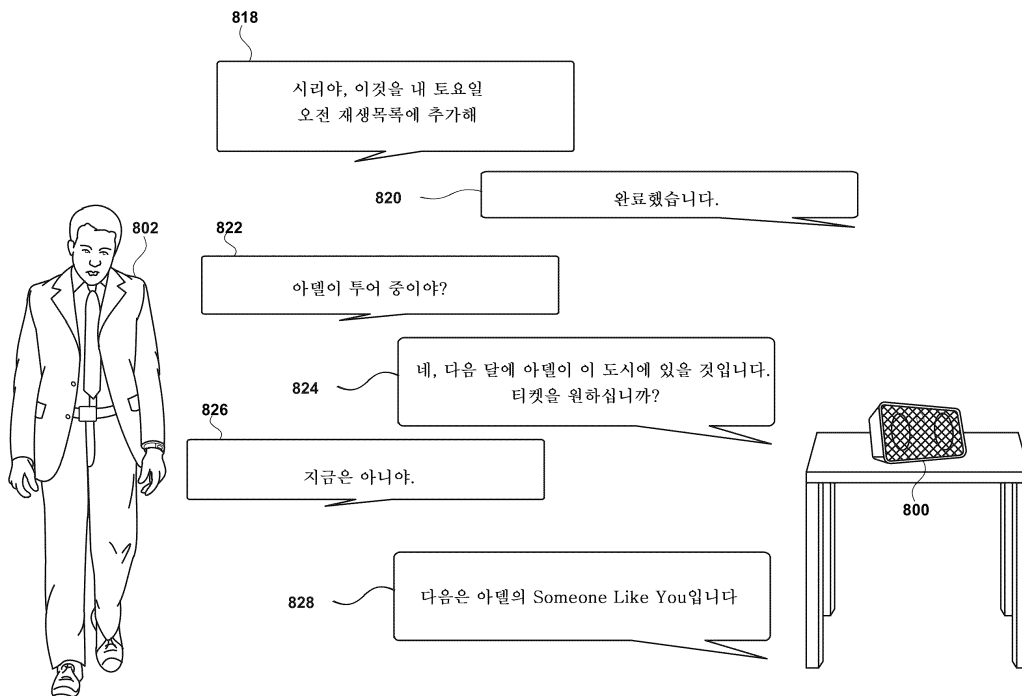
도면7c



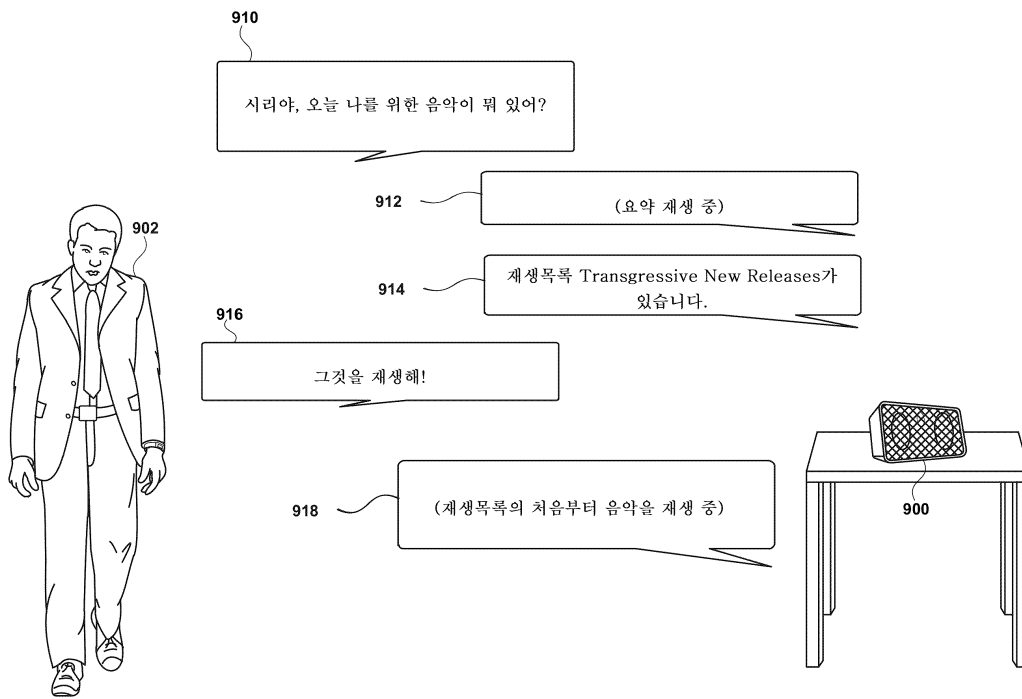
도면8a



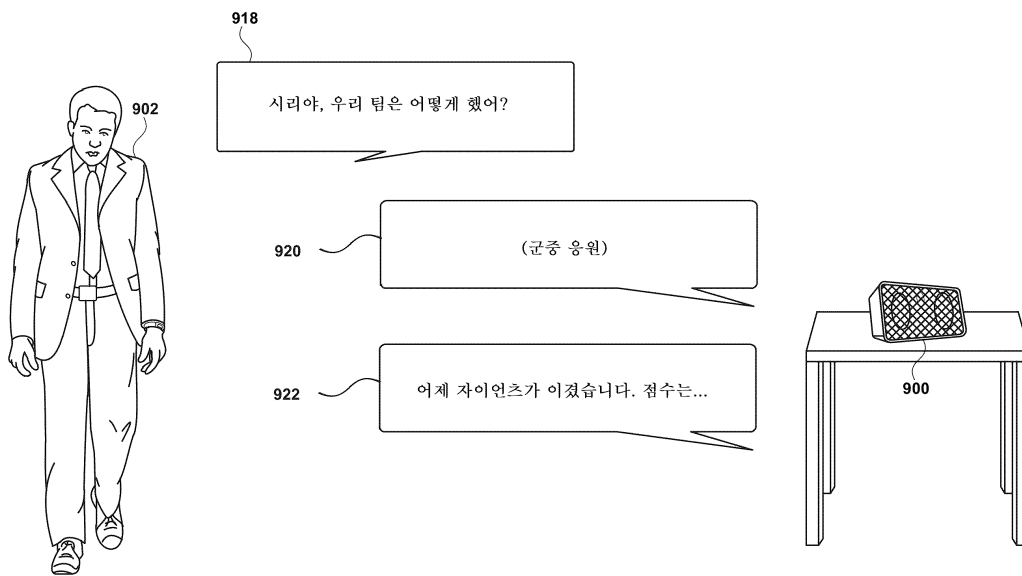
도면8b



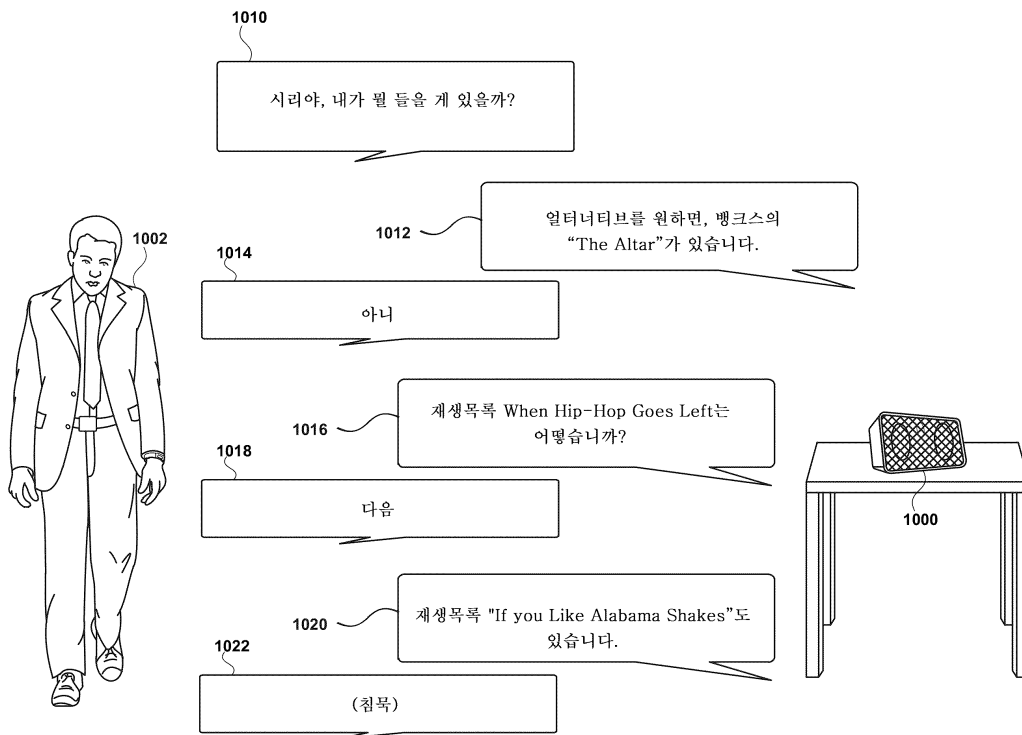
도면9a



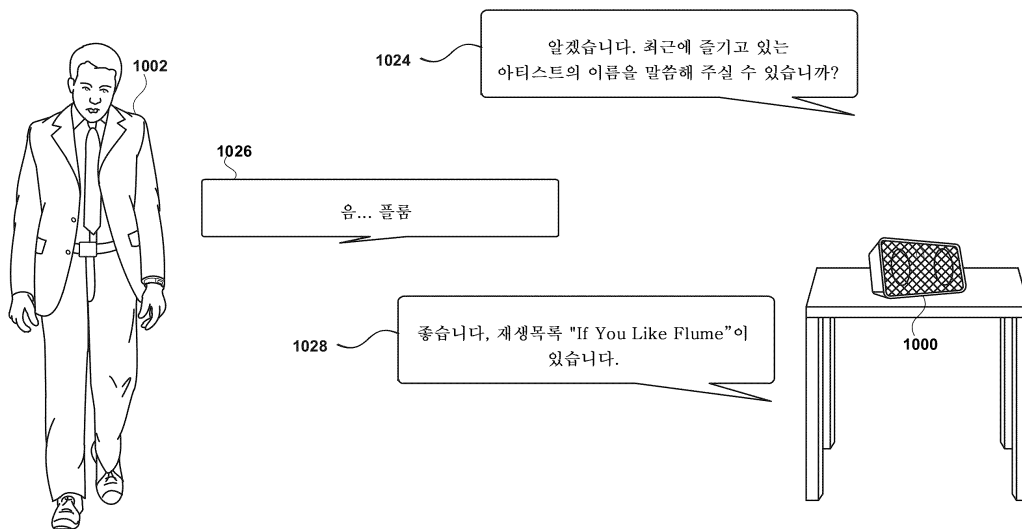
도면9b



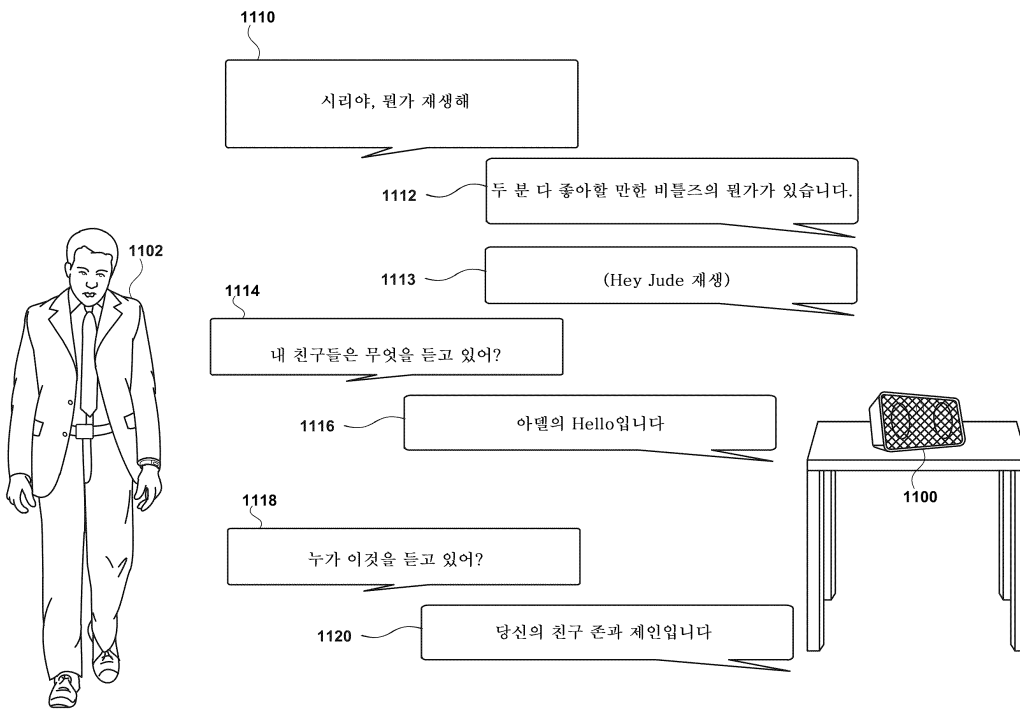
도면10a



도면10b

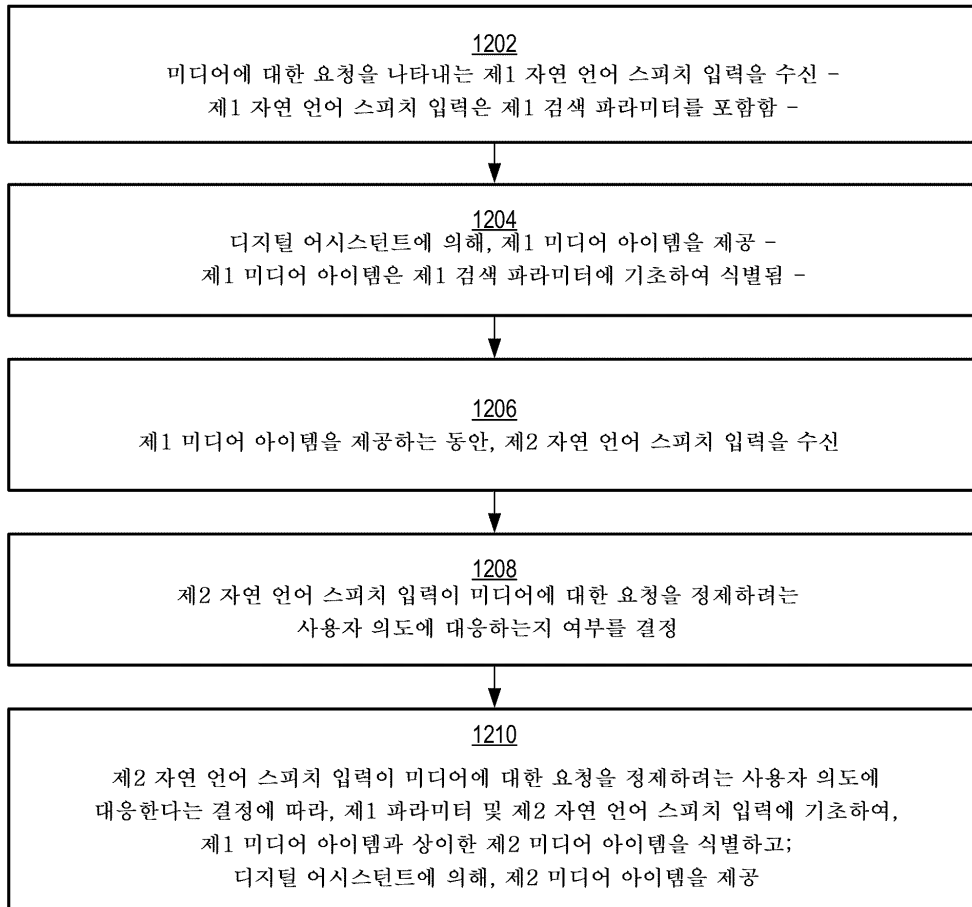


도면11



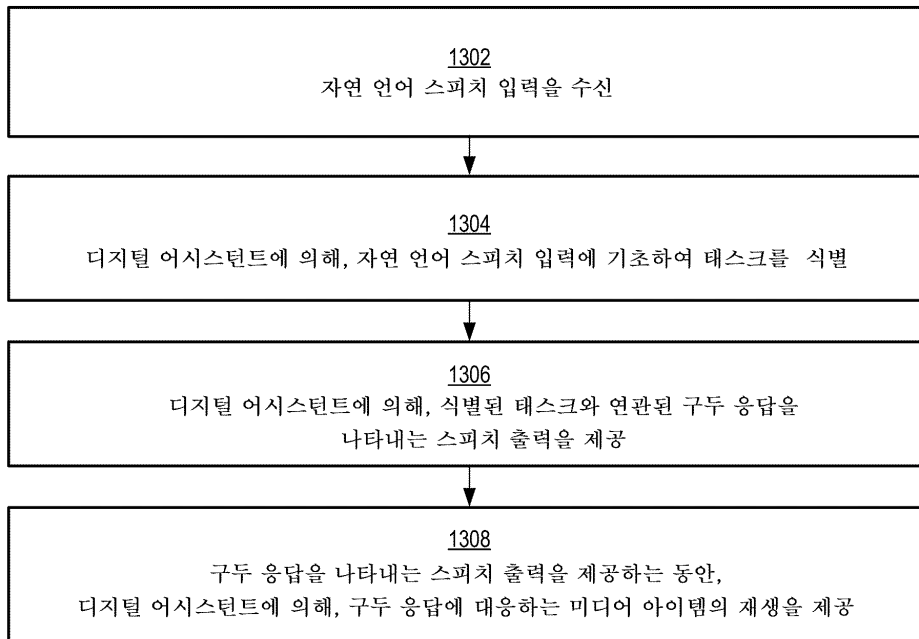
도면12

프로세스
1200 →



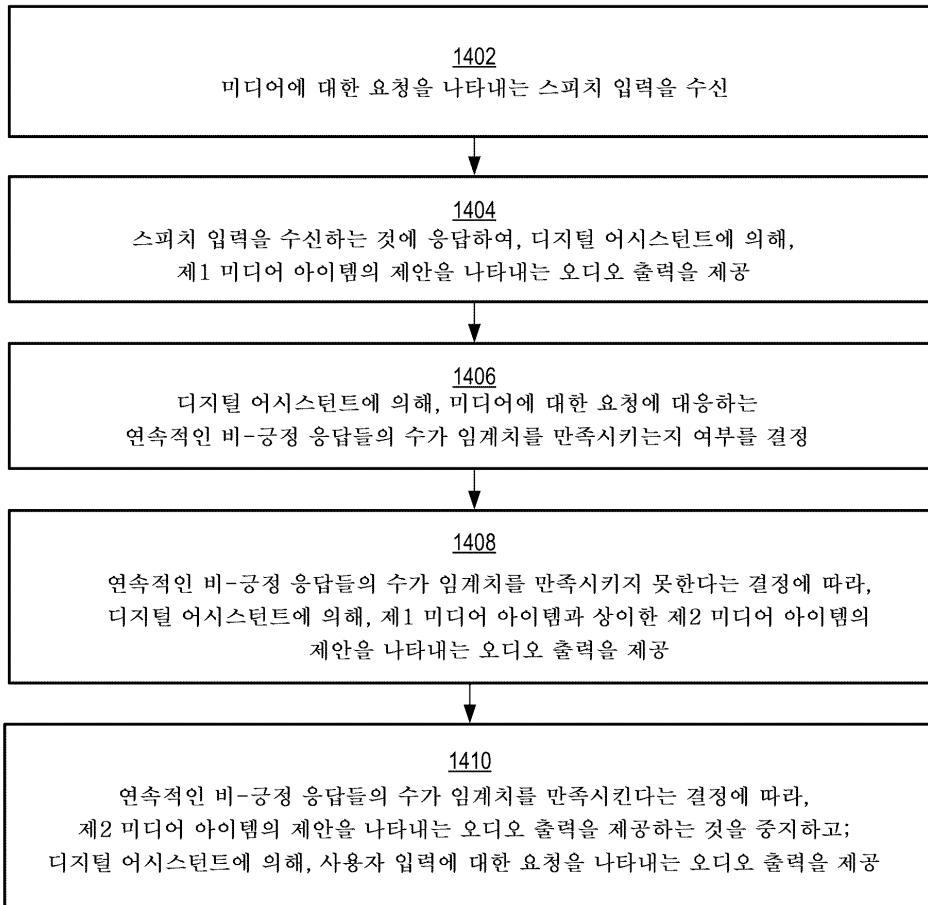
도면13

프로세스
1300 →



도면14

프로세스
1400



도면15

프로세스
1500

