



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년01월20일
(11) 등록번호 10-2205371
(24) 등록일자 2021년01월14일

- (51) 국제특허분류(Int. Cl.)
G10L 15/22 (2006.01) G10L 13/08 (2006.01)
G10L 15/04 (2006.01) G10L 17/04 (2013.01)
- (52) CPC특허분류
G10L 15/22 (2013.01)
G10L 13/086 (2013.01)
- (21) 출원번호 10-2016-0099402(분할)
- (22) 출원일자 2016년08월04일
심사청구일자 2020년01월29일
- (65) 공개번호 10-2018-0010923
- (43) 공개일자 2018년01월31일
- (62) 원출원 특허 10-2016-0092851
원출원일자 2016년07월21일
심사청구일자 2016년07월21일
- (30) 우선권주장
14/805,753 2015년07월22일 미국(US)
- (56) 선행기술조사문헌
US20090319270 A1*
JP2000089780 A*
WO2007111169 A1
*는 심사관에 의하여 인용된 문헌

- (73) 특허권자
구글 엘엘씨
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
- (72) 발명자
게바라 라지엘 알바레스
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
- (74) 대리인
박장원

전체 청구항 수 : 총 20 항

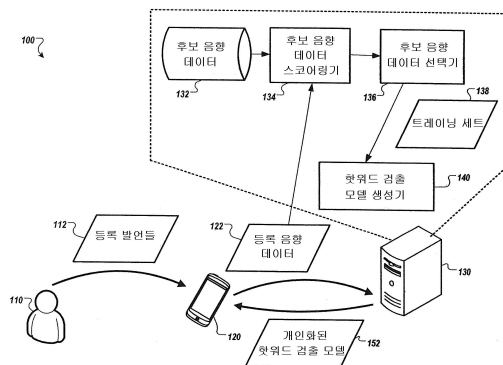
심사관 : 노지명

(54) 발명의 명칭 개별화된 핫워드 검출 모델들

(57) 요약

시스템 내에 알림들을 표시하기 위한, 컴퓨터 저장 매체 상에서 인코딩된 컴퓨터 프로그램들을 포함하는, 방법들, 시스템들 및 장치들이 개시된다. 일 양태에서, 방법은 사용자에게 의해 말해진 등록 발언을 나타내는 등록 음향 데이터를 획득하는 단계; 다른 사용자들에 의해 말해진 발언들을 나타내는 후보 음향 데이터의 세트를 획득하는 단계; 후보 음향 데이터의 세트의 각 후보 음향 데이터에 대하여, 등록 음향 데이터와 후보 음향 데이터의 유사성을 나타내는 유사성 스코어를 결정하는 단계; 유사성 스코어들에 적어도 기초하여 후보 음향 데이터의 세트로부터 후보 음향 데이터의 서브세트를 결정하는 단계; 후보 음향 데이터의 서브세트에 기초하여 검출 모델을 생성하는 단계; 그리고 사용자에게 의해 말해진 발언을 검출하기 위해 상기 검출 모델을 제공하는 단계의 동작들을 포함한다.

대표도



(52) CPC특허분류

G10L 15/04 (2013.01)

G10L 17/04 (2013.01)

명세서

청구범위

청구항 1

컴퓨터로 구현되는 방법에 있어서,

등록 프로세스(enrollment process) 동안, 클라이언트 디바이스에 의해, 사용자가 특정 핫워드(hotword)를 말하도록 프롬프트(prompt)하고, 클라이언트 디바이스에 의해, 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신하는 단계;

등록 프로세스 동안 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신함에 응답하여, 클라이언트 디바이스에 의해, 개인화된(personalized) 핫워드 검출 모델을 획득하는 단계, 상기 개인화된 핫워드 검출 모델은 등록 프로세스 동안 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 (i) 오디오 데이터, 및 (i) 동일한 특정 핫워드에 대한 것이고 (ii) 등록 프로세스 동안 말한 사용자에게 의한 특정 핫워드의 하나의 단일 발언과 음향적으로 유사한 것으로 표시되는 다른(other) 발언에 대응하는 (ii) 다른 저장된 오디오 데이터를 포함하는 생성된 음향 데이터의 트레이닝 세트를 사용하여 상기 사용자에게 의한 특정 핫워드의 있을 수 있는 발언(likely utterance)을 검출하도록 트레이닝되고; 및

생성된 음향 데이터의 트레이닝 세트를 사용하여 사용자가 특정 핫워드를 말할 때를 검출하도록 트레이닝된 개인화된 핫워드 검출 모델을 획득한 후, 클라이언트 디바이스에 의해, 개인화된 핫워드 검출 모델을 사용하여 이후에 수신된 오디오 데이터에서 상기 사용자에게 의한 특정 핫워드의 있을 수 있는 발언을 검출하는 단계를 포함하는 것을 특징으로 하는 컴퓨터로 구현되는 방법.

청구항 2

청구항 1 에 있어서,

상기 등록 프로세스 동안, 사용자가 특정 핫워드를 말하도록 프롬프트하고, 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신하는 단계는:

상기 사용자가 하나 이상의 용어들의 의미론적 해석을 트리거(trigger)하는 하나 이상의 용어들 또는 상기 특정 핫워드에 후속하는 하나 이상의 용어들을 말하도록 프롬프트하는 단계를 포함하는 것을 특징으로 하는 컴퓨터로 구현되는 방법.

청구항 3

청구항 1 에 있어서,

상기 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신함에 응답하여, 개인화된 핫워드 검출 모델을 획득하는 단계는,

상기 사용자에게 의한 상기 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신한 이후에 상기 개인화된 핫워드 검출 모델을 생성하는 단계 및 상기 특정 핫워드의 다른 발언에 대응하는 추가적인 오디오 데이터를 수신하지 않는 단계를 포함하는 것을 특징으로 하는 컴퓨터로 구현되는 방법.

청구항 4

청구항 1 에 있어서,

상기 개인화된 핫워드 검출 모델을 획득한 후 상기 등록 프로세스를 종료(end)하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터로 구현되는 방법.

청구항 5

청구항 1 에 있어서,

상기 개인화된 핫워드 검출 모델은 상기 단일 발언에 적어도 기초하고 상기 특정 핫워드의 다른 발언에 기초하

지 않는 것을 특징으로 하는 컴퓨터로 구현되는 방법.

청구항 6

청구항 1 에 있어서,

상기 개인화된 핫워드 검출 모델을 사용하여 이후에 수신된 오디오 데이터에서 상기 사용자에게 의한 특정 핫워드의 있을 수 있는 발언을 검출하는 단계는,

후속 발언에 대응하는 오디오 데이터를 수신하는 단계; 그리고

상기 개인화된 핫워드 검출 모델에 적어도 기초하여, 상기 후속 발언이 상기 특정 핫워드를 포함할 가능성이 있는지 여부를 결정하는 단계를 포함하는 것을 특징으로 하는 컴퓨터로 구현되는 방법.

청구항 7

청구항 6 에 있어서,

상기 개인화된 핫워드 검출 모델에 적어도 기초하여, 상기 후속 발언이 상기 특정 핫워드를 포함할 가능성이 있는지 여부를 결정하는 것에 응답하여, 상기 후속 발언의 적어도 일부에 대해 의미론적 해석을 수행하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터로 구현되는 방법.

청구항 8

시스템에 있어서,

하나 이상의 컴퓨터 및 동작 가능한 명령어들을 저장하는 하나 이상의 저장 디바이스들을 포함하며, 상기 명령어들은 하나 이상의 컴퓨터에 의해 실행될 때, 상기 하나 이상의 컴퓨터들로 하여금 동작들을 수행하게 하며,

상기 동작들은,

등록 프로세스 동안, 클라이언트 디바이스에 의해, 사용자가 특정 핫워드를 말하도록 프롬프트하고, 클라이언트 디바이스에 의해, 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신하는 동작과;

등록 프로세스 동안 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신함에 응답하여, 클라이언트 디바이스에 의해, 개인화된 핫워드 검출 모델을 획득하는 동작과, 상기 개인화된 핫워드 검출 모델은 등록 프로세스 동안 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 (i) 오디오 데이터, 및 (i) 동일한 특정 핫워드에 대한 것이고 (ii) 등록 프로세스 동안 말한 사용자에게 의한 특정 핫워드의 하나의 단일 발언과 음향적으로 유사한 것으로 표시되는 다른(other) 발언에 대응하는 (ii) 다른 저장된 오디오 데이터를 포함하는 생성된 음향 데이터의 트레이닝 세트를 사용하여 상기 사용자에게 의한 특정 핫워드의 있을 수 있는 발언(likely utterance)을 검출하도록 트레이닝되고; 그리고

생성된 음향 데이터의 트레이닝 세트를 사용하여 사용자가 특정 핫워드를 말할 때를 검출하도록 트레이닝된 개인화된 핫워드 검출 모델을 획득한 후, 클라이언트 디바이스에 의해, 개인화된 핫워드 검출 모델을 사용하여 이후에 수신된 오디오 데이터에서 상기 사용자에게 의한 특정 핫워드의 있을 수 있는 발언을 검출하는 동작을 포함하는 것을 특징으로 하는 시스템.

청구항 9

청구항 8 에 있어서,

상기 등록 프로세스 동안, 사용자가 특정 핫워드를 말하도록 프롬프트하고, 상기 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신하는 동작은,

상기 사용자가 하나 이상의 용어들의 의미론적 해석을 트리거하는 하나 이상의 용어들 또는 상기 특정 핫워드 이후에 하나 이상의 용어들을 말하도록 프롬프트하는 동작을 포함하는 것을 특징으로 하는 시스템.

청구항 10

청구항 8 에 있어서,

상기 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신함에 응답하여, 개인화된 핫워드 검출 모델을 획득하는 동작은,

상기 사용자에게 의한 상기 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신한 이후에 상기 개인화된 핫워드 검출 모델을 생성하는 동작 및 상기 특정 핫워드의 다른 발언에 대응하는 추가적인 오디오 데이터를 수신하지 않는 동작을 포함하는 것을 특징으로 하는 시스템.

청구항 11

청구항 8 에 있어서,

상기 개인화된 핫워드 검출 모델을 획득한 후 상기 등록 프로세스를 종료하는 동작을 더 포함하는 것을 특징으로 하는 시스템.

청구항 12

청구항 8 에 있어서,

상기 개인화된 핫워드 검출 모델은 상기 단일 발언에 적어도 기초하고 상기 특정 핫워드의 다른 발언에 기초하지 않는 것을 특징으로 하는 시스템.

청구항 13

청구항 8 에 있어서,

상기 개인화된 핫워드 검출 모델을 사용하여 이후에 수신된 오디오 데이터에서 상기 사용자에게 의한 특정 핫워드의 있을 수 있는 발언을 검출하는 동작은,

후속 발언에 대응하는 오디오 데이터를 수신하는 동작; 그리고

상기 개인화된 핫워드 검출 모델에 적어도 기초하여, 상기 후속 발언이 상기 특정 핫워드를 포함할 가능성이 있는지 여부를 결정하는 동작을 포함하는 것을 특징으로 하는 시스템.

청구항 14

청구항 13 에 있어서,

상기 개인화된 핫워드 검출 모델에 적어도 기초하여, 상기 후속 발언이 상기 특정 핫워드를 포함할 가능성이 있는지 여부를 결정하는 것에 응답하여, 상기 후속 발언의 적어도 일부에 대해 의미론적 해석을 수행하는 동작을 더 포함하는 것을 특징으로 하는 시스템.

청구항 15

하나 이상의 컴퓨터에 의해 실행가능한 명령어들을 포함하는 비-일시적인 컴퓨터-판독가능한 매체로서, 상기 명령어들은 실행 시, 상기 하나 이상의 컴퓨터로 하여금 동작들을 실행하게 하며,

상기 동작들은,

등록 프로세스 동안, 클라이언트 디바이스에 의해, 사용자가 특정 핫워드를 말하도록 프롬프트하고, 클라이언트 디바이스에 의해, 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신하는 동작과;

등록 프로세스 동안 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신함에 응답하여, 클라이언트 디바이스에 의해, 개인화된 핫워드 검출 모델을 획득하는 동작과, 상기 개인화된 핫워드 검출 모델은 등록 프로세스 동안 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 (i) 오디오 데이터, 및 (i) 동일한 특정 핫워드에 대한 것이고 (ii) 등록 프로세스 동안 말한 사용자에게 의한 특정 핫워드의 하나의 단일 발언과 음향적으로 유사한 것으로 표시되는 다른(other) 발언에 대응하는 (ii) 다른 저장된 오디오 데이터를 포함하는 생성된 음향 데이터의 트레이닝 세트를 사용하여 상기 사용자에게 의한 특정 핫워드의 있을 수 있는 발언(likely utterance)을 검출하도록 트레이닝되고; 그리고

생성된 음향 데이터의 트레이닝 세트를 사용하여 사용자가 특정 핫워드를 말할 때를 검출하도록 트레이닝된 개

개인화된 핫워드 검출 모델을 획득한 후, 클라이언트 디바이스에 의해, 개인화된 핫워드 검출 모델을 사용하여 이후에 수신된 오디오 데이터에서 상기 사용자에게 의한 특정 핫워드의 있을 수 있는 발언을 검출하는 동작을 포함하는 것을 특징으로 하는 컴퓨터-판독가능한 매체.

청구항 16

청구항 15 에 있어서,

상기 등록 프로세스 동안, 사용자가 특정 핫워드를 말하도록 프롬프트하고, 상기 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신하는 동작은,

상기 사용자가 하나 이상의 용어들의 의미론적 해석을 트리거하는 하나 이상의 용어들 또는 상기 특정 핫워드 이후속하는 하나 이상의 용어들을 말하도록 프롬프트하는 동작을 포함하는 것을 특징으로 하는 컴퓨터-판독가능한 매체.

청구항 17

청구항 15 에 있어서,

상기 사용자에게 의한 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신함에 응답하여, 개인화된 핫워드 검출 모델을 획득하는 동작은,

상기 사용자에게 의한 상기 특정 핫워드의 하나의 단일 발언에 대응하는 오디오 데이터를 수신한 이후에 상기 개인화된 핫워드 검출 모델을 생성하는 동작 및 상기 특정 핫워드의 다른 발언에 대응하는 추가적인 오디오 데이터를 수신하지 않는 동작을 포함하는 것을 특징으로 하는 컴퓨터-판독가능한 매체.

청구항 18

청구항 15 에 있어서,

상기 개인화된 핫워드 검출 모델을 획득한 후 상기 등록 프로세스를 종료하는 동작을 더 포함하는 것을 특징으로 하는 컴퓨터-판독가능한 매체.

청구항 19

청구항 15 에 있어서,

상기 개인화된 핫워드 검출 모델은 상기 단일 발언에 적어도 기초하고 상기 특정 핫워드의 다른 발언에 기초하지 않는 것을 특징으로 하는 컴퓨터-판독가능한 매체.

청구항 20

청구항 15 에 있어서,

상기 개인화된 핫워드 검출 모델을 사용하여 이후에 수신된 오디오 데이터에서 상기 사용자에게 의한 특정 핫워드의 있을 수 있는 발언을 검출하는 동작은,

후속 발언에 대응하는 오디오 데이터를 수신하는 동작; 그리고

상기 개인화된 핫워드 검출 모델에 적어도 기초하여, 상기 후속 발언이 상기 특정 핫워드를 포함할 가능성이 있는지 여부를 결정하는 동작을 포함하는 것을 특징으로 하는 컴퓨터-판독가능한 매체.

발명의 설명

기술 분야

[0001] 본 발명은 일반적으로 음성 명령들을 사용하여 컴퓨터들을 제어하는 것에 관한 것이다.

배경 기술

[0002] 컴퓨터는 사용자의 발언(utterance)을 분석하고, 이에 응답하여 동작을 수행한다. 예를 들면, 사용자는 "DRIVE HOME"을 말하고, 이에 응답하여, 컴퓨터는 사용자가 현재 위치로부터 집으로 운전할 방향들을 제시한다.

발명의 내용

- [0003] 일반적으로, 본 명세서에 기술된 본 발명의 양태는 개별화된 핫워드(hotword) 검출 모델을 생성하기 위한 프로세스를 포함한다. 본 명세서에서 사용되는 "핫워드"는 슬립 상태(sleep state) 또는 하이버네이션 상태(hibernation state)로부터 디바이스를 깨우는 용어 및/또는 상기 용어에 대한 의미론적 해석을 트리거(trigger)하는 용어 또는 상기 용어에 후속하는 하나 이상의 용어, 예를 들어, 핫워드에 후속하는 음성 명령들을 지칭한다.
- [0004] 예를 들면, 발언 "OK COMPUTER, DRIVE HOME,"에서, 용어 "OK COMPUTER,"는 후속하는 용어 "DRIVE HOME,"에 대한 의미론적 해석을 트리거 하는 핫워드가 되고, 상기 용어 "DRIVE HOME"은 사용자의 집으로의 방향을 제공하는 음성 명령에 대응한다. 시스템이 발언 "OK COMPUTER, DRIVE HOME,"에 대응하는 소리를 수신하는 경우, 시스템은 상기 발언이 핫워드 "OK COMPUTER"로 시작함을 결정하고, 이에 응답하여, 상기 소리를 표기(transcribe)하고, 음성 명령 "DRIVE HOME"의 표기(transcription)에 대한 의미론적 해석을 수행하며, 사용자가 집으로 운전할 방향들을 출력한다.
- [0005] 핫워드들은 시스템으로 향하지 않은 소리들을 잠재적으로 들을 수 있는(pick up) "항시 온(always on)" 시스템에 유용하다. 예를 들면, 상기 핫워드들의 사용은, 시스템으로 하여금 해당 발언이 그 환경에 존재하는 다른 개인을 향하는 발언이 아닌, 시스템을 향하는 때를 구별하게 하는데 도움을 준다. 이렇게 함으로써, 상기 시스템은 예를 들어, 핫워드를 포함하지 않는 소리들 또는 발언들에 대한 의미론적 해석과 같은, 컴퓨팅적으로 고비용의 프로세싱(computationally expensive processing)을 피할 수 있다.
- [0006] 핫워드들의 사용은 음성-기반 시스템의 신뢰성을 향상시킨다. 예를 들면, 사용자에게 의해 발언된 음성 명령들을 인식하고, 상기 음성 명령들에 따라 액션(action)들이 수행되도록 하게하는 시스템에서, 핫워드들의 사용은 시스템으로 하여금 사용자의 의도된 음성 명령들의 더 높은 퍼센티지를 인식함을 보장한다. 이는 시스템을 좀 더 신뢰할 수 있게 만들고, 또한 사용자에게 대해 시스템의 사용 편의성을 향상시킴을 이해할 수 있을 것이다. 다른 효과는, 의도된 음성 명령이 음성 명령으로서 시스템에 의해 인식되기 위해, 시스템의 사용자가 의도된 음성 명령을 반복해야만 하는 빈도가 감소된다는 것이다. 이는, 시스템이 사용자로부터의 소리들 또는 발언들을 해석 또는 재해석하는 빈도를 잠재적으로 감소시킴으로써, 시스템에서 컴퓨팅적 자원(computational resource)(예를 들면, 메모리 및 프로세싱 자원)을 절약할 수 있다.
- [0007] 시스템은 핫워드 검출 모델에 기초하여, 핫워드를 포함하는 발언을 검출한다. 그러나, 서로 다른 사용자들은 동일한 핫워드를 서로 다른 방법으로 발음한다. 따라서, 어떤 사용자들이 핫워드를 말할 때, 시스템은 검출하지 못할 수 있다. 시스템은 개별화된 핫워드 검출 모델들을 생성하는 것에 기초하여, 핫워드들의 검출을 증가시킨다. 그러나, 핫워드 검출 모델을 생성하는 것은 무수한 발언들을 사용하고, 사용자는 무수한 등록 발언(enrollment utterance)들을 제공하는 것을 원하지 않는다. 따라서, 사용자에게 의해 하나 이상의 등록 발언들을 수신한 후, 시스템은 다른 사용자들에 의한 다른 핫워드의 발언들을 식별하고, 상기 사용자에게 의한 등록 발언들과 유사한 발언들을 선택하고, 상기 선택된 발언들 및 상기 등록 발언들을 사용하는 개별화된 핫워드 검출 모델을 생성한다.
- [0008] 일부 양태들에서, 본 명세서에 기술된 본 발명의 양태는 사용자에게 의해 말해진(spoken) 등록 발언을 나타내는 등록 음향 데이터(enrollment acoustic data)를 획득하는 것, 다른 사용자에게 의해 말해진 발언들을 나타내는 후보 음향 데이터(candidate acoustic data)의 세트(set)를 획득하는 것, 후보 음향 데이터의 세트의 각 후보 음향 데이터에 대해, 등록 음향 데이터와 후보 음향 데이터 사이의 유사성을 나타내는 유사성 스코어(similarity score)를 결정하는 것, 유사성 스코어들에 적어도 기초하여, 후보 음향 데이터의 세트로부터 후보 음향 데이터의 서브세트(subset)를 선택하는 것, 후보 음향 데이터의 서브세트에 기초하여, 검출 모델을 생성하는 것, 그리고 사용자에게 의해 말해진 발언을 검출하는데 사용하기 위해, 검출 모델을 제공하는 것의 동작들을 포함하는 방법들로 구현된다.
- [0009] 다른 버전들은 상기 방법들의 동작들을 수행하기 위해 설정되고, 컴퓨터 저장 디바이스들에 인코딩된(encode), 대응하는 시스템들, 장치들 및 컴퓨터 프로그램들을 포함한다.
- [0010] 이들 및 다른 버전들은 각각 임의로 다음 특징들 중 하나 이상을 포함한다. 예를 들어, 일부 구현들에서, 사용자에게 의해 말해진 등록 발언을 나타내는 등록 음향 데이터를 획득하는 것은, 사용자에게 의해 말해진 미리 결정된 어구(phrase)의 다수의 발언들에 대한 등록 음향 데이터를 획득하는 것을 포함한다.
- [0011] 특정 양태들에서, 다른 사용자에게 의해 말해진 발언들을 나타내는 후보 음향 데이터의 세트를 획득하는 것은, 등

록 발언은 미리 결정된 어구인 것을 결정하는 것, 그리고 다른 사용자들에 의해 말해진 미리 결정된 어구의 발언들을 나타내는 후보 음향 데이터를 식별하는 것을 포함한다.

- [0012] 일부 양태들에서, 후보 음향 데이터의 세트의 각 후보 음향 데이터에 대해, 유사성 스코어를 결정하는 것은, 등록 음향 데이터와 후보 음향 데이터 사이의 거리(distance)를 결정하는 것, 그리고 상기 거리에 기초하여 유사성 스코어를 결정하는 것을 포함한다.
- [0013] 일부 구현들에서, 후보 음향 데이터의 세트의 각 후보 음향 데이터에 대해, 유사성 스코어를 결정하는 것은, 후보 음향 데이터에 의해 나타나는 발언을 말하는 다른 사용자의 인구 통계학적(demographic) 정보 및 등록 발언을 말하는 사용자의 인구 통계학적 정보에 기초하여 유사성 스코어들을 결정하는 것을 포함한다.
- [0014] 특정 양태들에서, 등록 음향 데이터와 후보 음향 데이터 사이의 유사성을 나타내는 후보 음향 데이터의 유사성 스코어들에 적어도 기초하여, 후보 음향 데이터의 세트로부터 후보 음향 데이터의 서브세트를 선택하는 것은 후보 음향 데이터의 미리 결정된 개수(number)를 선택하는 것에 근거한다.
- [0015] 일부 양태들에서, 후보 음향 데이터의 서브세트에 기초하여, 검출 모델을 생성하는 것은 상기 후보 음향 데이터의 서브세트를 사용하여, 뉴럴 네트워크(neural network)를 트레이닝(training)시키는 것을 포함한다. 일부 구현들에서, 추가적인 동작들은 상기 검출 모델을 사용하여, 미리 결정된 어구의 발언을 검출하는 것을 포함한다.
- [0016] 본 명세서에 기술된 본 발명의 하나 이상의 구현들의 세부사항들은 첨부 도면들 및 아래의 서술로 설명된다. 다른 잠재적인 특징들, 양태들 및 본 발명의 이점들은 서술들, 도면들 및 청구 범위들로부터 명확해질 것이다.

도면의 간단한 설명

- [0017] 도 1 및 도 2는 개별화된 핫워드 검출 모델을 생성하기 위한 예시적인 시스템의 블록 다이어그램들의 도면들이다.
 도 3은 개별화된 핫워드 검출 모델을 생성하기 위한 예시적인 프로세스의 흐름도이다.
 도 4는 예시적인 컴퓨팅 디바이스들의 다이어그램이다.
 여러 도면들에서 동일한 도면 부호들은 동일한 요소들을 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0018] 도 1은 개별화된 핫워드 검출 모델을 생성하기 위한 예시적인 시스템(110)의 블록 다이어그램이다. 간략하게, 그리고 아래에서 더 상세히 기술된 바와 같이, 상기 시스템(100)은 클라이언트 디바이스(120) 및 서버(130)를 포함하고, 상기 서버(130)는 후보 음향 데이터 스코어링기(scorer)(134), 후보 음향 데이터 선택기(selector)(136) 및 핫워드 검출 모델 생성기(generator)(140)를 포함한다.
- [0019] 클라이언트 디바이스(120)는 스마트폰, 랩탑 컴퓨터, 태블릿 컴퓨터, 데스크탑 컴퓨터 또는 사용자(110)가 핫워드를 말하는 때를 검출할 수 있도록 설정된 일부 다른 컴퓨팅 디바이스이다. 예를 들면, 상기 클라이언트 디바이스(120)는 사용자(110)가 "OK COMPUTER"라고 말하는 때를 검출할 수 있도록 설정된다.
- [0020] 상기 클라이언트 디바이스(120)는 핫워드 검출 모델을 사용하여, 사용자(110)가 핫워드를 말하는 때를 검출한다. 예를 들면, 클라이언트 디바이스(120)는 핫워드 검출 모델을 사용하여, "OK COMPUTER"를 말하는 사용자를 검출 하고, 상기 핫워드 검출 모델은 핫워드 "OK COMPUTER"가 말해진 때에 대응하는 소리들을 검출하기 위해 트레이닝 되었다.
- [0021] 그러나, 서로 다른 사용자들은 동일한 핫워드를 서로 다른 방법으로 발음한다. 예를 들면, 사용자(110)는 "OK COMPUTER"를 "OK COM-UT-ER"로서 발음하고, 핫워드 검출 모델은 "OK COM-UT-ER"를 "OK COMPUTER"로서 검출하지 못한다. 따라서, 클라이언트 디바이스(120)는 개인화된 핫워드 검출 모델(152)에 기초하여, 사용자(110)에 의해 말해진 핫워드들의 검출을 증가시킬 수 있고, 상기 개인화된 핫워드 검출 모델(152)은 사용자(110)가 핫워드를 말할 때를 검출하기 위해 트레이닝 되었다. 예를 들면, 상기 개인화된 핫워드 검출 모델(152)은 핫워드 "OK COMPUTER"의 사용자(110)의 발음인 "OK COM-UT-ER"를 검출할 수 있도록 트레이닝 된다.
- [0022] 상기 개인화된 핫워드 검출 모델(152)을 획득하기 위해, 클라이언트 디바이스(120)는 사용자가 등록 발언들을 제공하도록 프롬프트(prompt)한다. 예를 들면, 핫워드 "OK COMPUTER"를 검출하기 위한 개인화된 핫워드 검출 모델을 획득하기 위해, 클라이언트 디바이스(120)는 사용자(110)에게 프롬프트 "NOW PERSONALIZING HOTWORD

DETECTION, SAY 'OK COMPUTER' THREE TIMES"를 제공한다. 상기 클라이언트 디바이스(120)는 음향 데이터로서 소리를 캡처(capture)하는 음향 데이터 생성기를 포함한다. 예를 들면, 클라이언트 디바이스(120)는 "OK COMPUTER"를 말하는 사용자(110)를 신호들로서의 "OK COM-UT-ER"로 캡처하고, 상기 신호들을 멜-프리퀀시 캡스트랄 계수들(mel-frequency cepstral coefficients)에 의해 나타나는 등록 음향 데이터(122)로서 인코딩하는 마이크로폰을 포함한다.

[0023] 클라이언트 디바이스(120)는 상기 등록 음향 데이터(122)를 서버(130)에 제공하고, 이에 응답하여 개인화된 핫워드 검출 모델(152)을 수신한다. 예를 들면, 클라이언트 디바이스(120)는 "OK COMPUTER"를 말하는 사용자(110)를 "OK COM-UT-ER"로서 나타내는 등록 음향 데이터(122)를 서버(130)에 제공하고, 이에 응답하여, 상기 등록 음향 데이터에 적어도 기초하여 트레이닝 된, 개인화된 핫워드 검출 모델(152)을 수신한다.

[0024] 그 후 클라이언트 디바이스(120)는 상기 개인화된 핫워드 검출 모델(152)을 사용하여, 사용자가 핫워드를 말하는 때를 검출한다. 예를 들면, 개인화된 핫워드 검출 모델(152)을 사용하는 것은 "OK COMPUTER"를 "OK COM-UT-ER"로서 말하는 사용자(110)를 나타내는 등록 음향 데이터(122)에 기초하여 트레이닝 되며, 클라이언트 디바이스(120)는 사용자가 "OK COM-UT-ER"를 말할 때, 상기 사용자(110)가 핫워드 "OK COMPUTER"를 말하고 있음을 검출한다.

[0025] 서버(130)는 등록 음향 데이터에 기초하여, 개인화된 핫워드 검출 모델을 생성하도록 설정된다. 예를 들면, 서버(130)는 "OK COMPUTER"를 말하는 사용자(110)를 "OK COM-UT-ER"로서 나타내는 등록 음향 데이터(122)를 수신하고, 상기 등록 음향 데이터에 적어도 기초하여 개인화된 핫워드 검출 모델(152)을 트레이닝 시킨다.

[0026] 그러나, 핫워드 검출 모델을 생성하는 것은 무수한 발언들을 사용하고, 사용자는 개인적으로 무수한 등록 발언들을 제공하는 것을 원하지 않는다. 추가적으로, 상기 서버(130)에 의한, 사용자로부터의 무수한 등록 발언들의 수신 및 저장은 비교적 많은 양의 메모리를 요구한다. 따라서, 사용자로부터 하나 이상의 등록 발언들을 수신한 후, 상기 서버(130)는 다른 사용자들에 의한 핫워드의 다른 발언들을 식별하고, 사용자에게 의한 등록 발언들과 유사한 발언들을 선택하고, 상기 선택된 발언들 및 등록 발언들을 사용하여 개인화된 핫워드 검출 모델(152)을 생성한다. 다른 이점들 중에서도, 이러한 방식으로 개인화된 핫워드 검출 모델들을 생성하는 것은 특히, 시스템이 많은 복수의 개별화된 사용자들에 대해 개인화된 검출 모델들을 생성하고 제공할 때, 서버(130)에서 요구되는 메모리의 전체 양을 줄일 수 있는 잠재성을 가진다. 후술하는 바와 같이, 이는 새로운 개별적 음성 검출 모델을 만드는데 있어, 적어도 부분적으로 다른 사용자들로부터의 발언들의 효과적이고 효율적인 사용 때문이다.

[0027] 구체적으로, 서버(130)의 후보 음향 데이터베이스(132)는 다양한 사용자들의 발언들을 나타내는 음향 데이터를 저장한다. 예를 들면, 상기 서버(130)의 후보 음향 데이터베이스(132)는 서로 다른 사용자들의 방대한 수의 발언들을 나타내는 음향 데이터를 저장한다. 후보 음향 데이터베이스(132)는 각 음향 데이터를 발언된 핫워드를 표시하는 데이터와 함께 저장한다. 예를 들면, 후보 음향 데이터베이스(132)는 핫워드 "OK COMPUTER"의 발언으로서 라벨(label)된 음향 데이터의 5만 세트들 및 서로 다른 핫워드 "MY BUTLER"의 발언으로서 라벨된 음향 데이터의 5만 세트들을 저장한다. 일부 구현들에서, 후보 음향 데이터베이스(132)는 음향 데이터와 사용자를 기술하는 인구 통계학적 데이터를 관련시킨다. 예를 들면, 후보 음향 데이터베이스(132)는 음향 데이터와 사용자에게 의해 핫워드가 말해진 때에 사용자가 있었던 위치를 관련시킬 수 있다. 다른 예에서, 후보 음향 데이터베이스(132)는 음향 데이터와 사용자의 성별, 사용자의 나이 범위 또는 사용자를 기술하는 일부 다른 정보를 관련시킬 수 있다.

[0028] 서버(130)의 후보 음향 데이터 스코어링기(134)는 등록 음향 데이터(122) 및 후보 음향 데이터베이스(132)로부터 후보 음향 데이터를 획득하고, 등록 음향 데이터(122)와 후보 음향 데이터 사이의 유사성을 나타내는 유사성 스코어를 생성하도록 설정된다. 예를 들면, 후보 음향 데이터 스코어링기(134)는 "OK COMPUTER"라고 말하는 사용자의 등록 음향 데이터(122) 및 "OK COMPUTER"라고 말하는 다른 사용자를 나타내는 후보 음향 데이터를 수신하고, 90%의 유사성을 결정하고, 0.9의 스코어와 후보 음향 데이터를 관련시킬 수 있다. 상기 예에서, 후보 음향 데이터 스코어링기(134)는 그 후 "OK COMPUTER"라고 말하는 또 다른 사용자를 나타내는 후보 음향 데이터의 두 번째 세트를 획득하고, 상기 등록 음향 데이터(122)에 대해 30%의 유사성을 결정하고, 0.3의 스코어와 후보 음향 데이터의 두 번째 세트를 관련시킬 수 있다.

[0029] 특정 발언을 나타내는 후보 음향 데이터의 유사성 스코어는 특정 발언과 등록 발언 사이의 음향 유사성을 반영한다. 예를 들면, 상기 유사성 스코어는 0에서 1범위 일 수 있고, 상기 범위에서 더 높은 유사성 스코어는 더 큰 음향 유사성을 반영하고, 더 낮은 스코어는 더 낮은 음향 유사성을 반영한다. 다른 예들에서, 예를 들어, 1-5, A-F 또는 0%-100%와 같은, 스코어들 및 범위들의 다른 유형이 사용될 수 있다.

- [0030] 후보 음향 데이터 스코어링기(134)는 등록 음향 데이터와 후보 음향 데이터 사이의 거리에 기초하여 스코어를 생성한다. 예를 들면, 후보 음향 데이터 스코어링기(134)는 등록 음향 데이터의 벨-프린키스 캡스트랄 계수들과 다수의 프레임들에 걸친 후보 음향 데이터 사이의 차이를 종합하고, 유사성 스코어를 결정하고, 상기 유사성 스코어에서 더 큰 종합 거리들은 그 결과 더 적은 유사성을 반영하는 점수들이 되고, 더 낮은 종합 거리들은 그 결과 더 큰 유사성을 반영하는 점수들이 된다.
- [0031] 일부 구현들에서, 후보 음향 데이터 스코어링기(134)는 다른 사용자의 인구 통계학적 정보에 기초하여, 스코어를 결정한다. 예를 들면, 동일 성별 사용자의 발언들을 나타내는 후보 음향 데이터를 선택하는 것 대신, 후보 음향 데이터 스코어링기(134)는 서로 다른 성별들의 사용자들의 발언들을 나타내는 후보 음향 데이터를 획득하고, 상기 후보 음향 데이터에 의해 나타난 발언을 말하는 사용자의 성별이 사용자(110)의 성별과 일치(match)하는지 여부를 결정하고, 일치를 결정하는 것에 응답하여, 사용자(110)와 같은 성별의 사용자들의 발언들을 나타내는 후보 음향 데이터에 더 높은 유사성 스코어를 할당한다.
- [0032] 일부 구현들에서, 후보 음향 데이터 스코어링기(134)는 후보 음향 데이터베이스(132)에 저장된 더 많은 후보 음향 데이터 중에서부터, 후보 음향 데이터를 선택한다. 예를 들면, 후보 음향 데이터 스코어링기(134)는 핫워드 "OK COMPUTER"가 말해진 후보 음향 데이터베이스(132)로부터 음향 데이터를 수신할 것을 선택한다. 후보 음향 데이터 스코어링기(134)는, 등록 음향 데이터와 함께, 말해진 핫워드의 하나 이상의 표시 또는 핫워드를 말하는 사용자의 유형의 표시를 획득하고, 후보 음향 데이터베이스(132)에 쿼리(query)하며, 상기 후보 음향 데이터베이스(132)는 핫워드를 말하는 사용자와 유사한 유형 또는 동일한 핫워드를 말하는 사용자의 음향 데이터에 대한 후보 음향 데이터베이스(132)이다. 예를 들면, 후보 음향 데이터 스코어링기(134)는 여성 사용자에 의해 말해진 핫워드 "OK COMPUTER" 표시를 획득하고, 이에 응답하여, 여성 사용자에 의해 말해진 핫워드 "OK COMPUTER"를 나타내는 음향 데이터에 대한 후보 음향 데이터베이스(132)에 쿼리한다.
- [0033] 후보 음향 데이터 선택기(136)는 후보 음향 데이터 스코어링기(134)으로부터 스코어링 된 후보 음향 데이터 및 등록 음향 데이터(122)를 획득하고, 개인화된 핫워드 검출 모델(152)을 트레이닝 시키기 위해, 음향 데이터의 트레이닝 세트(138)를 생성한다. 예를 들면, 후보 음향 데이터 선택기(136)는 "OK COMPUTER"를 말하는 사용자(110)를 나타내는 등록 음향 데이터를 획득하고, "OK COMPUTER"를 말하는 서로 다른 사용자들을 나타내는 5만개의 후보 음향 데이터를 획득하고, 상기 후보 음향 데이터의 각각은 후보 음향 데이터와 등록 음향 데이터(122) 사이의 유사성을 반영하는 유사성 스코어와 관련되며, 그리고 후보 음향 데이터 선택기(136)는 5만개 중 1만개의 후보 음향 데이터 및 등록 음향 데이터(122)를 포함하는 음향 데이터의 트레이닝 세트를 생성한다.
- [0034] 후보 음향 데이터 선택기(136)는 유사성 스코어들에 적어도 근거한 후보 음향 데이터의 서브세트를 선택하는 것에 기초하여, 트레이닝 세트(138)를 생성한다. 예를 들면, 후보 음향 데이터 선택기(136)는 5만 개의 후보 음향 데이터의 세트를 획득하고, 다른 후보 음향 데이터 보다 후보 음향 데이터와 등록 음향 데이터(122)사이의 더 높은 유사성들을 반영하는 유사성 스코어들을 가진, 상기 세트의 1만 개의 후보 음향 데이터의 서브세트를 선택한다.
- [0035] 후보 음향 데이터 선택기(136)는 예를 들면, 1천, 3천, 1만, 5만과 같은, 후보 음향 데이터의 미리 결정된 개수를 선택하는 것에 기초하여, 후보 음향 데이터의 서브세트를 선택한다. 예를 들면, 후보 음향 데이터 선택기(136)는 "OK COMPUTER"의 단일의 발언을 나타내는 등록 음향 데이터를 획득하고, 후보 음향 데이터와 상기 등록 음향 데이터 사이의 더 큰 유사성을 반영하는 유사성 스코어들을 가진 3천개의 후보 음향 데이터의 서브세트를 선택할 수 있다.
- [0036] 추가적 또는 대안적으로, 후보 음향 데이터 선택기(136)는 임계 유사성 스코어를 만족하는 후보 음향 데이터를 선택하는 것에 기초하여, 후보 음향 데이터의 서브세트를 선택한다. 예를 들면, 후보 음향 데이터 선택기(136)는 0.0 - 1.00의 스코어 범위에서부터 0.8, 0.85, 0.9의 임계 유사성 스코어 보다 위인 유사성 스코어들을 가진 후보 음향 데이터를 선택하고, 선택된 후보 음향 데이터를 트레이닝 세트(138)에 포함시킬 수 있다.
- [0037] 일부 구현들에서, 후보 음향 데이터 선택기(136)는 트레이닝 세트(138) 내의 음향 데이터에 가중치를 준다. 예를 들면, 후보 음향 데이터 선택기(136)는 트레이닝 세트(138) 내에 등록 음향 데이터를 여러 번 포함시키거나, 트레이닝 세트(138) 내의 등록 음향 데이터를 후보 음향 데이터보다 더 큰 가중치와 관련시킨다.
- [0038] 일부 구현들에서, 후보 음향 데이터 선택기(136)는 다수의 등록 음향 데이터에 기초하여, 후보 음향 데이터의 서브세트를 선택한다. 예를 들면, 후보 음향 데이터 선택기(136)는 사용자(110)에 의한 "OK COMPUTER"의 3개 발언들에 대한 등록 음향 데이터를 수신하고, 트레이닝 세트(138)에 포함시키기 위해, 등록 음향 데이터 각각에

대해, 가장 큰 유사성을 반영하는 유사성 스코어들을 가진 후보 음향 데이터 3천 개를 선택한다. 따라서, 만일 후보 음향 데이터가 다수의 등록 음향 데이터에 대해 선택된다면, 일부 후보 음향 데이터는 트레이닝 세트(138)에 여러 번 나타날 수 있다. 일부 구현들에서, 후보 음향 데이터 선택기(136)는 트레이닝 세트(138)로부터 중복된 후보 음향 데이터를 지우거나, 트레이닝 세트(138)에 중복된 후보 음향 데이터가 포함되는 것을 막을 수 있다.

[0039] 일부 구현들에서, 후보 음향 데이터 선택기(136)는 등록 음향 데이터의 개수에 기초하여 등록 음향 데이터를 선택하기 위해, 후보 음향 데이터의 개수를 결정하고, 상기 등록 음향 데이터의 개수는 후보 음향 데이터 선택기(136)에 의해 수신된다. 예를 들면, 후보 음향 데이터 선택기(136)는 5개의 등록 음향 데이터를 수신하고, 핫워드 검출 모델 생성기가 적어도 1만 음향 데이터의 트레이닝 세트를 수신해야함을 결정하고, 이에 응답하여, 수신된 각 등록 음향 데이터에 대해, 등록 음향 데이터와 함께 트레이닝 세트에 포함시키기 위해 적어도 1999개의 후보 음향 데이터를 선택할 수 있다. 다른 예에서, 후보 음향 데이터 선택기(136)는 10개의 등록 음향 데이터를 수신하고, 핫워드 검출 모델 생성기가 적어도 1만 음향 데이터의 트레이닝 세트를 수신해야함을 결정하고, 이에 응답하여, 수신된 각 등록 음향 데이터에 대해, 등록 음향 데이터와 함께 트레이닝 세트에 포함시키기 위해 적어도 999개의 후보 음향 데이터를 선택할 수 있다.

[0040] 다른 예에서, 후보 음향 데이터 선택기(136)는 다수의 등록 음향 데이터 각각에 대한 서브-유사성 스코어들을 결정하는 것에 기초하여, 후보 음향 데이터에 대한 유사성 스코어를 결정할 수 있다. 예를 들면, 후보 음향 데이터 선택기(136)는 3개의 등록 음향 데이터를 수신하고, 각각의 후보 음향 데이터에 대해, 등록 음향 데이터 중 하나에 각각 대응하는 3개의 서브-유사성 스코어들을 결정하고, 상기 서브-유사성 스코어들을 평균화하는 것에 기초하여 유사성 스코어를 결정할 수 있다. 또 다른 예에서, 후보 음향 데이터 선택기(136)는 유사성 스코어로서 후보 음향 데이터에 대한 서브-유사성 스코어들의 중간값, 최소값 또는 최대값을 가질 수 있다.

[0041] 핫워드 검출 모델 생성기(140)는 후보 음향 데이터 선택기(136)로부터 트레이닝 세트를 수신하고, 개인화된 핫워드 검출 모델(152)을 생성한다. 예를 들면, 핫워드 검출 모델 생성기(140)는 9997개의 선택된 후보 음향 데이터 및 3개의 등록 음향 데이터를 포함하는 트레이닝 세트를 수신하고, 상기 트레이닝 세트에 기초하여 개인화된 핫워드 검출 모델(152)을 생성할 수 있다.

[0042] 핫워드 검출 모델 생성기(140)는 핫워드의 발연들을 나타내는 것으로, 트레이닝 세트(138) 내의 음향 데이터를 검출하기 위해, 뉴럴 네트워크를 트레이닝 시키는 것에 기초하여 개인화된 핫워드 검출 모델(152)을 생성한다. 예를 들면, 핫워드 검출 모델 생성기(140)는 트레이닝 세트(138) 내의 음향 데이터에 기초하여, 핫워드 "OK COMPUTER"를 검출하는 개인화된 핫워드 검출 모델(152)을 생성한다.

[0043] 시스템(100)의 서로 다른 구성들은, 클라이언트 디바이스(120) 및 후보 음향 데이터 스코어링기(134), 후보 음향 데이터 선택기(136) 및 핫워드 검출 모델 생성기(140)를 포함하는 서버(130)의 기능이 결합되고 분리되고, 분산되거나 또는 상호교환 될 수 있는 곳에서 사용될 수 있다. 시스템(100)은 단일 디바이스에서 구현되거나, 다수의 디바이스들에 걸쳐 분산될 수 있다.

[0044] 도 2는 개별화된 핫워드 검출 모델을 생성하기 위한 예시적인 서버(130)의 블록 다이어그램이다. 서버(130)는 도 1에 도시된 서버이다. 상기 기술된 바와 같이, 서버(130)는 후보 음향 데이터베이스(132), 후보 음향 데이터 스코어링기(134), 후보 음향 데이터 선택기(136) 및 핫워드 검출 모델 생성기(140)를 포함한다.

[0045] 후보 음향 데이터베이스(132)는 핫워드 "OK COMPUTER"를 말하는 다양한 사용자들의 다수의 후보 음향 데이터를 포함한다. 예를 들면, 후보 음향 데이터베이스(132)는 "OK COMPUTER"를 "OK COM-PU-TER"로서 말하는 "사용자 A"의 후보 음향 데이터, "OK COMPUTER"를 "OOK COM-PU-TER"로서 말하는 "사용자 B"의 후보 음향 데이터, "OK COMPUTER"를 "OK COP-TER"로서 말하는 "사용자 C"의 후보 음향 데이터, "OK COMPUTER"를 "OK COM-U-TER"로서 말하는 "사용자 D"의 후보 음향 데이터, "OK COMPUTER"를 "OK COM-MUT-ER"로서 말하는 "사용자 E"의 후보 음향 데이터, "OK COMPUTER"를 "OK COM-PUT-EW"로서 말하는 "사용자 F"의 후보 음향 데이터 및 "OK COMPUTER"를 말하는 다른 사용자들의 다른 후보 음향 데이터를 포함할 수 있다.

[0046] 후보 음향 데이터 스코어링기(134)는 사용자의 등록 음향 데이터(202)를 수신하고, 후보 음향 데이터베이스(132)로부터 후보 음향 데이터의 세트를 획득한다. 예를 들면, 후보 음향 데이터 스코어링기(134)는 "OK COMPUTER"를 "OK COM-UT-ER"로서 말하는 사용자의 등록 음향 데이터(202)를 수신하고, 이에 응답하여, 후보 음향 데이터베이스(132)로부터 "OK COMPUTER"를 "OK COM-PU-TER"로서 말하는 "사용자 A"의 후보 음향 데이터, "OK COMPUTER"를 "OOK COM-PU-TER"로서 말하는 "사용자 B"의 후보 음향 데이터, "OK COMPUTER"를 "OK COP-

TER"로서 말하는 "사용자 C"의 후보 음향 데이터, "OK COMPUTER"를 "OK COM-U-TER"로서 말하는 "사용자 D"의 후보 음향 데이터, "OK COMPUTER"를 "OK COM-MUT-ER"로서 말하는 "사용자 E"의 후보 음향 데이터, "OK COMPUTER"를 "OK COM-PUT-EW"로서 말하는 "사용자 F"의 후보 음향 데이터 및 "OK COMPUTER"를 말하는 다른 사용자들의 다른 후보 음향 데이터를 포함하는 후보 음향 데이터의 세트를 획득할 수 있다.

[0047] 후보 음향 데이터 스코어링기(134)는 후보 음향 데이터의 세트의 각각에 대한 유사성 스코어들을 생성한다. 예를 들면, "OK COMPUTER"를 "OK COM-UT-ER"로서 말하는 사용자(110)의 등록 음향 데이터에 대해, 후보 음향 데이터 스코어링기(134)는 "OK COMPUTER"를 "OK COM-PU-TER"로서 말하는 "사용자 A"의 후보 음향 데이터에 대해 중간의 유사성을 반영하는 0.6의 유사성 스코어, "OK COMPUTER"를 "OOK COM-PU-TER"로서 말하는 "사용자 B"의 후보 음향 데이터에 대해 중간의 유사성을 반영하는 0.5의 유사성 스코어, "OK COMPUTER"를 "OK COP-TER"로서 말하는 "사용자 C"의 후보 음향 데이터에 대해 낮은 유사성을 반영하는 0.3의 유사성 스코어, "OK COMPUTER"를 "OK COM-U-TER"로서 말하는 "사용자 D"의 후보 음향 데이터에 대해 높은 유사성을 반영하는 0.9의 유사성 스코어, "OK COMPUTER"를 "OK COM-MUT-ER"로서 말하는 "사용자 E"의 후보 음향 데이터에 대해 높은 유사성을 반영하는 0.8의 유사성 스코어 및 "OK COMPUTER"를 "OK COM-PUT-EW"로서 말하는 "사용자 F"의 후보 음향 데이터에 대해 중간의 유사성을 반영하는 0.5의 유사성 스코어를 생성할 수 있다.

[0048] 후보 음향 데이터 선택기(136)는 후보 음향 데이터 스코어링기(134)로부터 스코어링 된 후보 음향 데이터(204)를 수신하고, 음향 데이터의 트레이닝 세트(138)를 생성한다. 예를 들면, 후보 음향 데이터 선택기(136)는 "OK COMPUTER"를 "OK COM-PU-TER"로서 말하는 "사용자 A"의 후보 음향 데이터에 대해 중간의 유사성을 반영하는 0.6의 유사성 스코어, "OK COMPUTER"를 "OOK COM-PU-TER"로서 말하는 "사용자 B"의 후보 음향 데이터에 대해 중간의 유사성을 반영하는 0.5의 유사성 스코어, "OK COMPUTER"를 "OK COP-TER"로서 말하는 "사용자 C"의 후보 음향 데이터에 대해 낮은 유사성을 반영하는 0.3의 유사성 스코어, "OK COMPUTER"를 "OK COM-U-TER"로서 말하는 "사용자 D"의 후보 음향 데이터에 대해 높은 유사성을 반영하는 0.9의 유사성 스코어, "OK COMPUTER"를 "OK COM-MUT-ER"로서 말하는 "사용자 E"의 후보 음향 데이터에 대해 높은 유사성을 반영하는 0.8의 유사성 스코어 및 "OK COMPUTER"를 "OK COM-PUT-EW"로서 말하는 "사용자 F"의 후보 음향 데이터에 대해 중간의 유사성을 반영하는 0.5의 유사성 스코어, 대응하는 후보 음향 데이터 및 등록 음향 데이터를 수신하고, 이에 응답하여, "OK COMPUTER"를 "OK COM-U-TER"로서 말하는 "사용자 D"의 후보 음향 데이터, "OK COMPUTER"를 "OK COM-MUT-ER"로서 말하는 "사용자 E"의 후보 음향 데이터 및 "OK COMPUTER"를 "OK COM-UT-ER"로서 말하는 사용자의 등록 음향 데이터를 포함하는 음향 데이터의 트레이닝 세트를 생성한다.

[0049] 후보 음향 데이터 선택기(136)는 유사성 스코어들에 기초하여 후보 음향 데이터의 세트의 서브세트를 선택함으로써, 트레이닝 세트를 생성한다. 예를 들면, 후보 음향 데이터 선택기(136)는 핫워드 검출 모델 생성기가 3개의 음향 데이터의 트레이닝 세트를 수신해야 함을 결정하고, 하나의 등록 음향 데이터가 있음을 결정하고, 총 3개의 음향 데이터를 얻기 위해 2개의 후보 음향 데이터를 선택하는 것을 결정하고, 그리고 모든 후보 음향 데이터 중에서 등록 음향 데이터와 가장 큰 유사성을 반영하는 0.9 및 0.8의 유사성 스코어들을 가진 후보 음향 데이터를 선택할 수 있다.

[0050] 핫워드 검출 모델 생성기(140)는 트레이닝 세트(138)를 수신하고, 개인화된 핫워드 검출 모델(152)을 생성한다. 예를 들면, 핫워드 검출 모델 생성기(140)는 "OK COMPUTER"를 "OK COM-U-TER"로서 말하는 "사용자 D"의 후보 음향 데이터 및 "OK COMPUTER"를 "OK COM-MUT-ER"로서 말하는 "사용자 E"의 후보 음향 데이터 및 "OK COMPUTER"를 "OK COM-UT-ER"로서 말하는 사용자의 등록 음향 데이터를 포함하는 트레이닝 세트를 수신하고, 상기 사용자(110)에 의해 말해진 핫워드 "OK COMPUTER"를 나타내는 그 음향 데이터를 검출하기 위해, 뉴럴 네트워크를 트레이닝 시킬 수 있다.

[0051] 도 3은 개별화된 핫워드 검출 모델을 생성하기 위한 예시적인 프로세스의 흐름도이다. 다음은, 도 1을 참조하여 기술된 시스템(100)의 컴포넌트(component)들에 의해 수행되는, 프로세스(300)를 기술한다. 그러나, 상기 프로세스(300)는 다른 시스템들 또는 시스템 구성들에 의해 수행될 수도 있다.

[0052] 프로세스(300)는 사용자에 의해 말해진 등록 발언을 나타내는 등록 음향 데이터를 획득하는 것(310)을 포함한다. 예를 들면, 후보 음향 데이터 스코어(134)는 사용자가 핫워드 "MY BUTLER"를 말한 때를 검출하고자, 클라이언트 디바이스(120)를 트레이닝 시키기 위한 샘플 등록 발언을 제공하기 위해 클라이언트 디바이스(120)에 의해 유도된 후, 클라이언트 디바이스(120)로부터 핫워드 "MY BUTLER"를 말하는 사용자를 나타내는 등록 음향 데이터를 획득할 수 있다.

[0053] 프로세스(300)는 후보 음향 데이터의 세트를 획득하는 것(320)을 포함하고, 상기 후보 음향 데이터의 세트는 다

른 사용자들에 의해 말해진 발언들을 나타낸다. 예를 들면, 후보 음향 데이터 스코어링기(134)는 등록 음향 데이터가 20세와 30세 사이의 남자에 의해 말해진 핫워드 "MY BUTLER"에 대한 것이라는 것을 결정하고, 이에 응답하여, 후보 음향 데이터베이스(132)로부터, 핫워드 "MY BUTLER"를 말하는 20세와 30세 사이의 다른 남자 사용자들을 나타내는 후보 음향 데이터를 획득할 수 있다.

[0054] 상기 프로세스는, 후보 음향 데이터의 세트의 각 후보 음향 데이터에 대한, 유사성 스코어를 결정하는 것(330)을 포함하고, 상기 유사성 스코어는 등록 음향 데이터와 후보 음향 데이터 사이의 유사성을 나타낸다. 예를 들면, 후보 음향 데이터베이스(132)로부터 획득된 각 후보 음향 데이터에 대해, 후보 음향 데이터 스코어링기(134)는 핫워드 "MY BUTLER"를 말하는 사용자들 나타내는 등록 음향 데이터 및 핫워드 "MY BUTLER"를 말하는 다른 사용자 나타내는 후보 음향 데이터를 결정한다.

[0055] 상기 프로세스는 유사성 스코어들에 적어도 기초하여 후보 음향 데이터의 세트로부터, 후보 음향 데이터의 서브세트를 선택하는 것(340)을 포함한다. 예를 들면, 후보 음향 데이터 선택기(136)는 등록 음향 데이터와 가장 큰 유사성을 반영하는 유사성 스코어들을 가진 후보 음향 데이터의, 예를 들어, 1천, 5천, 2만 또는 어떤 다른 개수와 같은, 미리 결정된 개수를 선택할 수 있다. 다른 예에서, 후보 음향 데이터 선택기(136)는, 예를 들어, 0.7, 0.8, 0.9 또는 어떤 다른 양과 같은, 임계 유사성 스코어를 만족하는 유사성 스코어들을 가진 후보 음향 데이터를 선택할 수 있다.

[0056] 상기 프로세스는 후보 음향 데이터의 서브세트에 기초하여 검출 모델을 생성하는 것(350)을 포함한다. 예를 들면, 핫워드 검출 모델 생성기(140)는 "MY BUTLER"를 말하는 다른 사용자들의 선택된 후보 음향 데이터를 사용하여, 사용자가 핫워드 "MY BUTLER"를 소리 낼 때를 검출하기 위해, 뉴럴 네트워크를 트레이닝 시키는 것에 기초하여, 개인화된 핫워드 검출 모델을 생성한다.

[0057] 상기 프로세스는 사용자에게 의해 말해진 발언을 검출하는 데 사용하기 위해, 검출 모델을 제공하는 것(360)을 포함한다. 예를 들면, 서버(130)는 클라이언트 디바이스(120)에 개인화된 핫워드 검출 모델(152)을 제공하고, 상기 개인화된 핫워드 검출 모델(152)은 핫워드 검출 모델 생성기(140)에 의해 생성된다. 클라이언트 디바이스(120)는 그 후 사용자(110)가 핫워드 "MY BUTLER"를 말하는 때를 검출하기 위해, 상기 개인화된 핫워드 검출 모델(152)을 사용한다.

[0058] 도 4는 컴퓨팅 디바이스(400)의 예시 및 여기에 기술된 기술들을 구현하기 위해 사용될 수 있는 모바일 컴퓨팅 디바이스(450)를 도시한다. 컴퓨팅 디바이스(400)는 다양한 형태들의 디지털 컴퓨터들, 예를 들어, 랩탑들, 데스크탑들, 워크스테이션들, PDA들, 서버들, 블레이드 서버들, 메인 프레임들 및 다른 적합한 컴퓨터들을 나타내도록 의도된다. 컴퓨팅 디바이스(450)는, 여러 형태들의 모바일 컴퓨팅 디바이스들, 예를 들어, PDA들, 셀룰러 폰들, 스마트 폰들 및 다른 적합한 컴퓨팅 디바이스들을 나타내도록 의도된다. 여기에서 도시된 컴포넌트들, 그것들의 접속들 및 관계들, 그리고 그것들의 기능들은, 단지 예시적인 것을 의미하고, 제한하는 것을 의미하지 않는다.

[0059] 컴퓨팅 디바이스(400)는 프로세서(402), 메모리(404), 저장 디바이스(406), 상기 메모리(404) 및 다수의 고속 확장 포트들(410)과 연결되는 고속 인터페이스(408) 및 저속 확장 포트(414) 및 상기 저장 디바이스(406)와 연결되는 저속 인터페이스(412)를 포함한다. 상기 프로세서(402), 메모리(404), 저장 디바이스(406), 고속 인터페이스(408), 고속 확장 포트들(410) 및 저속 인터페이스(412) 각각은 다양한 버스들을 사용하여 상호접속되고, 공통의 마더보드 상에 또는 적합한 다른 방식들로 장착될 수 있다. 상기 프로세서(402)는 외부 입력/출력 디바이스, 예를 들어, 고속 인터페이스(408)에 연결된 디스플레이(416) 상의 그래픽 사용자 인터페이스(GUI)에 대한 그래픽 정보를 디스플레이 하기 위해 메모리(404) 내에 저장되거나 또는 저장 디바이스(406) 상에 저장된 명령어들을 포함하는, 컴퓨팅 디바이스(400) 내에서 실행을 위한 명령어들을 처리할 수 있다. 다른 구현들에서, 다수의 프로세서들 및/또는 다수의 버스들은, 다수의 메모리들 및 메모리의 타입들에 따라, 적합하게 사용될 수 있다. 또한, 다수의 컴퓨팅 디바이스들은 필요한 동작들의 부분들을 제공하는 각 디바이스(예를 들어, 서버 뱅크, 블레이드 서버들의 그룹 또는 멀티-프로세서 시스템으로서)와 연결될 수 있다.

[0060] 메모리(404)는 컴퓨팅 디바이스(400) 내에 정보를 저장한다. 일부 구현들에서, 메모리(404)는 휘발성 메모리 또는 유닛들이다. 일부 구현들에서, 메모리(404)는 비-휘발성 메모리 유닛 또는 유닛들이다. 상기 메모리(404)는 또한 컴퓨터-판독가능한 매체, 예를 들어, 자기 또는 광 디스크의 다른 형태일 수 있다.

[0061] 저장 디바이스(406)는 컴퓨팅 디바이스(400)에 대량 스토리지(storage)를 제공할 수 있다. 일부 구현들에서, 저장 디바이스(406)는 컴퓨터-판독가능한 매체, 예를 들어, 플로피 디스크 디바이스, 하드 디스크 디바이스, 광

디스크 디바이스 또는 테이프 디바이스, 플래시 메모리 또는 다른 유사한 고체 상태 메모리 디바이스 또는 디바이스들의 어레이(저장 영역 네트워크 또는 다른 구성들 내의 디바이스를 포함)일 수 있거나 상기 컴퓨터-판독가능한 매체를 포함할 수 있다. 명령어들은 정보 캐리어 내에 저장될 수 있다. 상기 명령어들은, 하나 이상의 프로세싱 디바이스들(예를 들면, 프로세서(402))에 의해 실행될 때, 예를 들어, 상기 기술된 것들과 같은, 하나 이상의 방법들을 수행한다. 상기 명령어들은 또한, 예를 들어, 컴퓨터-판독가능한 매체 또는 기계-판독가능한 매체와 같은, 하나 이상의 저장 디바이스들(예를 들어, 메모리(404), 저장 디바이스(406) 또는 상기 프로세서(402)상의 메모리)에 의해 저장될 수 있다.

[0062] 고속 인터페이스(408)는 컴퓨팅 디바이스(400)에 대한 대역폭-집약적인 동작들을 관리하는 반면에, 저속 인터페이스(412)는 더 낮은 대역폭-집약적인 동작들을 관리한다. 이와 같은 기능들의 할당은 단지 예시적이다. 일부 구현들에서, 고속 인터페이스(408)는 메모리(404), 디스플레이(416) 및 고속 확장 포트들(410)에 연결 (예를 들면, 그래픽 프로세서 또는 액셀레이터를 통해) 되고, 상기 고속 확장 포트들(410)은 여러 확장 카드들(미도시)을 수용할 수 있다. 상기 구현에서, 저속 인터페이스(412)는 저장 디바이스(406) 및 저속 확장 포트(414)에 연결된다. 여러 통신 포트들(예를 들어, USB, 블루투스, 이더넷, 무선 이더넷)을 포함할 수 있는, 저속 확장 포트(414)는, 하나 이상의 입력/출력 디바이스들, 예를 들어, 키보드, 포인팅 디바이스, 스위치 또는 네트워크 어댑터를 통한 스위치 또는 라우터와 같은, 네트워킹 디바이스에 연결될 수 있다.

[0063] 컴퓨팅 디바이스(400)는, 도면에서 도시된 것처럼, 복수의 각각 다른 형태들로 구현될 수 있다. 예를 들면, 표준 서버(420)로서, 또는 그와 같은 서버들의 그룹으로 여러 번 구현될 수 있다. 또한, 컴퓨팅 디바이스(400)는 랩탑 컴퓨터(422)와 같은 개인용 컴퓨터로 구현될 수 있다. 컴퓨팅 디바이스(400)는 또한 랙 서버 시스템(424)의 부분으로서 구현될 수 있다. 대안적으로, 컴퓨팅 디바이스(400)로부터의 컴포넌트들은, 모바일 컴퓨팅 디바이스(450)와 같은 모바일 컴퓨팅 디바이스(미도시) 내의 다른 컴포넌트들과 결합될 수 있다. 그와 같은 디바이스들 각각은, 하나 이상의 컴퓨팅 디바이스(400) 및 모바일 컴퓨팅 디바이스(450)를 포함할 수 있고, 전체 시스템은 서로 통신하는 다수의 컴퓨팅 디바이스들로 구성될 수 있다.

[0064] 모바일 컴퓨팅 디바이스(450)는, 다른 컴포넌트들 사이에서, 프로세서(452), 메모리(464), 디스플레이(454)와 같은 입력/출력 디바이스, 통신 인터페이스(466) 및 트랜시버(468)를 포함한다. 모바일 컴퓨팅 디바이스(450)는 또한 추가적인 저장을 제공하기 위해, 마이크로-드라이브와 같은 저장 디바이스 또는 다른 디바이스가 제공될 수 있다. 상기 프로세서(452), 메모리(464), 디스플레이(454), 통신 인터페이스(466) 및 트랜시버(468) 각각은 여러 버스들을 사용하여 상호접속되고, 상기 컴포넌트들 중 다수는 공통의 마더보드 상에 또는 적합한 다른 방식들로 장착될 수 있다.

[0065] 프로세서(452)는 메모리(464) 내에 저장되는 명령어들을 포함하는, 모바일 컴퓨팅 디바이스(450) 내의 명령어들을 실행할 수 있다. 상기 프로세서는 독립된 다수의 아날로그 및 디지털 프로세서들을 포함하는 칩들의 칩셋으로 구현될 수 있다. 상기 프로세서는, 예를 들면, 사용자 인터페이스들의 제어와 같은, 모바일 컴퓨팅 디바이스(450)의 다른 컴포넌트들의 조정을 위해, 모바일 컴퓨팅 디바이스(450)에 의해 구동되는 어플리케이션들 및 모바일 컴퓨팅 디바이스(450)에 의한 무선 통신을 제공할 수 있다.

[0066] 프로세서(452)는 디스플레이(454)와 연결된 제어 인터페이스(458) 및 디스플레이 인터페이스(456)를 통해, 사용자와 통신할 수 있다. 디스플레이(454)는, 예를 들면, TFT(Thin-Film-Transistor Liquid Crystal Display) 디스플레이 또는 OLED(Organic Light Emitting Diode) 디스플레이 또는 다른 적합한 디스플레이 기술일 수 있다. 디스플레이 인터페이스(456)는 그래픽 및 다른 정보를 사용자에게 제공하기 위해, 디스플레이(454)를 구동하는 적합한 회로를 포함할 수 있다. 상기 제어 인터페이스(458)는 사용자로부터 명령어들을 수신할 수 있고, 프로세서(452)로의 제공을 위해 명령어들을 변환할 수 있다. 또한, 외부 인터페이스(462)는 다른 디바이스들과 모바일 컴퓨팅 디바이스(450)의 근거리 통신을 가능하게 하도록, 프로세서(452)와의 통신에 제공될 수 있다. 외부 인터페이스(462)는, 예를 들면, 일부 구현들에서 유선 통신을 제공할 수 있거나, 또는 구현들에서 무선 통신을 제공할 수 있으며, 다수의 인터페이스들이 또한 사용될 수 있다.

[0067] 메모리(464)는 모바일 컴퓨팅 디바이스(450) 내에 정보를 저장한다. 상기 메모리(464)는 하나 이상의 컴퓨터-판독가능한 매체 또는 매체들, 휘발성 메모리 유닛 또는 유닛들, 또는 비-휘발성 메모리 유닛 또는 유닛들로서 구현될 수 있다. 확장 인터페이스(474)는 또한 확장 인터페이스(472)를 통해 디바이스(450)에 제공될 수 있거나 접속될 수 있고, 상기 확장 인터페이스(472)는, 예를 들면, SIMM(Single In Line Memory Module) 카드 인터페이스를 포함할 수 있다. 상기 확장 메모리(474)는 모바일 컴퓨팅 디바이스(450)를 위한 여분의 저장 공간을 제공할 수 있거나, 또는 모바일 컴퓨팅 디바이스(450)를 위한 어플리케이션 또는 다른 정보를 저장할 수 있다. 특

히, 상기 확장 메모리(474)는 상기 기술된 프로세스들을 실행하거나 보충하기 위한 명령어들을 포함할 수 있고, 그리고 또한 보안 정보를 포함할 수 있다. 따라서, 예를 들면, 확장 메모리(474)는 모바일 컴퓨팅 디바이스(450)에 대한 보안 모듈로서 제공될 수 있고, 모바일 컴퓨팅 디바이스(450)의 안전한 사용을 허용하는 명령어들로 프로그래밍 될 수 있다. 부가적으로, 보안 어플리케이션들은, 해킹할 수 없는 방식으로 SIMM 카드 상의 정보를 식별하는 것을 배치하는 것처럼, 추가 정보에 따라, SIMM 카드들을 통해 제공될 수 있다.

[0068] 상기 메모리는, 예를 들면, 하기에서 논의된 것처럼, 플래시 메모리 및/또는 NVRAM 메모리(non-volatile random access memory)를 포함할 수 있다. 일부 구현들에서, 상기 기술된 것들과 같이, 명령어들은 정보 캐리어 내에 저장되며, 상기 정보 캐리어는 명령어들이, 하나 이상의 프로세싱 디바이스들(예를 들면, 프로세서(452))에 의해 실행될 때, 하나 이상의 방법들을 수행한다. 명령어들은 또한, 예를 들어, 하나 이상의 컴퓨터-판독가능한 매체 또는 기계-판독가능한 매체들(예를 들면, 메모리(464), 확장 메모리(474) 또는 프로세서(452)상의 메모리)과 같은, 하나 이상의 저장 디바이스들에 의해 저장될 수 있다. 일부 구현들에서, 상기 명령어들은, 예를 들면, 트랜시버(468) 또는 외부 인터페이스(462)를 통해, 전파된 신호 내에서 수신될 수 있다.

[0069] 모바일 컴퓨팅 디바이스(450)는 통신 인터페이스(466)을 통해 무선으로 통신할 수 있고, 상기 통신 인터페이스(466)는 필요한 디지털 신호 처리 회로를 포함할 수 있다. 통신 인터페이스(466)는 다른 것들 중에서, GSM 음성 호출들(Global System for Mobile communications), SMS(Short Message Service), EMS(Enhanced Messaging Service) 또는 MMS 메시징(Multimedia Messaging Service), CDMA(code division multiple access), TDMA(time division multiple access), PDC(Personal Digital Cellular), WCDMA(Wideband Code Division Multiple Access) CDMA2000 또는 GPRS(General Packet Radio Service)와 같은, 여러 모드들 또는 프로토콜들 하에서 통신들을 제공할 수 있다. 이러한 통신은, 예를 들면, 무선-주파수를 사용하는 트랜시버(468)를 통해 발생할 수 있다. 부가적으로, 단거리 통신은, 예를 들어 블루투스, WiFi 또는 다른 이러한 트랜시버(미도시)를 사용하여, 발생할 수 있다. 부가적으로, GPS(Global Positioning System) 수신기 모듈(470)은 모바일 컴퓨팅 디바이스(450)에 추가적인 내비게이션-관련 무선 데이터 및 위치-관련 무선 데이터를 제공할 수 있고, 상기 모바일 컴퓨팅 디바이스(450)는 모바일 컴퓨팅 디바이스(450) 상에서 구동하는 어플리케이션들에 의해 적합하게 사용될 수 있다.

[0070] 모바일 컴퓨팅 디바이스(450)는 또한 오디오 코덱(460)을 사용하여 오디오적으로 통신할 수 있고, 상기 오디오 코덱(460)은 사용자로부터 음성 정보를 수신할 수 있고, 상기 음성 정보를 사용가능한 디지털 정보로 변환한다. 오디오 코덱(460)은 또한, 예를 들어, 모바일 컴퓨팅 디바이스(450)의 핸드세트(hand set) 내에서, 스피커를 통해, 사용자를 위한 들을 수 있는 소리를 발생시킬 수 있다. 이러한 소리는 음성 전화 호출들로부터 소리를 포함할 수 있고, 기록된 소리(예를 들면, 음성 메시지들, 음악 파일들, 등)를 포함할 수 있으며, 또한 모바일 컴퓨팅 디바이스(450) 상에서 동작하는 어플리케이션들에 의해 발생된 소리를 포함할 수 있다.

[0071] 모바일 컴퓨팅 디바이스(450)는, 도면에서 도시된 것처럼, 다수의 각각 다른 형태들로 구현될 수 있다. 예를 들면, 상기 모바일 컴퓨팅 디바이스(450)는 셀룰러 텔레폰(480)으로서 구현될 수 있다. 모바일 컴퓨팅 디바이스(450)는 또한 스마트-폰(482), PDA 또는 다른 유사한 모바일 컴퓨팅 디바이스의 일부로서 구현될 수 있다.

[0072] 본 명세서에서 기술된 본 발명의 실시예들, 기능적인 동작들 및 프로세스들은 디지털 전자 회로에서, 유형적으로 수록된 컴퓨터 소프트웨어 또는 펌웨어에서, 컴퓨터 하드웨어 또는 그들의 하나 이상의 조합들에서 구현되며, 이들은 본 명세서에서 개시된 구조들 및 그들의 구조적 균등물들을 포함한다. 본 명세서에서 기술된 본 발명의 실시예들은 하나 이상의 컴퓨터 프로그램들, 즉, 데이터 프로세싱 장치에 의한 실행을 위해, 또는 데이터 프로세싱 장치의 동작을 제어하기 위해, 유형적 비휘발성 프로그램 캐리어 상에 인코딩된 컴퓨터 프로그램 명령들의 하나 이상의 모듈들로서 구현될 수 있다. 대안적으로 또는 부가적으로, 프로그램 명령어들은 예를 들어, 데이터 프로세싱 장치에 의한 실행을 위해 적합한 수신 장치들로 전송하기 위한 정보를 인코딩 하기 위해 생성되는 기계-생성 전기, 광학 또는 전자기 신호와 같은, 인위적으로 생성된 전파 신호상에서 인코딩될 수 있다. 컴퓨터 저장 매체는 기계-판독가능한 저장 디바이스, 기계-판독가능한 저장 기관, 랜덤 또는 시리얼 액세스 메모리 디바이스 또는 그들의 하나 이상의 조합일 수 있다.

[0073] 용어 "데이터 프로세싱 장치"는, 예시적인 프로그램 가능한 프로세서, 컴퓨터 또는 복수의 프로세서들 또는 컴퓨터들을 거치는 것을 포함하여, 데이터를 처리하기 위한 장치들, 디바이스들 및 기계들의 모든 종류를 포함한다. 상기 장치는 예를 들면, FPGA(field programmable gate way array) 또는 ASIC(application specific integrated circuit)과 같은, 특수 목적 논리 회로를 포함한다. 상기 장치는 또한, 하드웨어에 부가하여, 예를 들면, 프로세서 펌웨어, 프로토콜 스택, 데이터베이스 관리 시스템, 작동 시스템, 또는 그들의 하나 이상의 조

합을 구성하는 코드와 같은, 문제되는 컴퓨터 프로그램에 대한 실행 환경을 만드는 코드를 포함한다.

[0074] 컴퓨터 프로그램(또한 프로그램, 소프트웨어, 소프트웨어 어플리케이션, 모듈, 소프트웨어 모듈, 스크립트 또는 코드로서 언급 또는 기술될 수 있는)은 컴파일(compile)되거나 해석된 언어들 또는 서술문의 또는 절차상의 언어들 포함하여, 프로그래밍 언어의 임의의 형태로 쓰여질 수 있고, 상기 컴퓨터 프로그램은, 독립형 프로그램 또는 모듈, 컴포넌트, 서브루틴(subroutine) 또는 컴퓨팅 환경에서 사용하기에 적합한 다른 유닛을 포함하여, 임의의 형태로 전개될 수 있다. 컴퓨터 프로그램은, 필요하지는 않지만, 파일 시스템 내의 파일에 대응할 수 있다. 프로그램은 다른 프로그램들 또는 데이터를 유지하는 파일의 부분(예를 들어, 마크업 언어 문서 내에 저장된 하나 이상의 스크립트들)내, 문제되는 프로그램에 전용되는 단일 파일 내 또는 다수의 조직화된 파일들(예를 들어, 하나 이상의 모듈들, 서브 프로그램들 또는 코드의 부분들을 저장하는 파일들) 내에 저장될 수 있다. 컴퓨터 프로그램은 하나의 컴퓨터 또는 다수의 컴퓨터들 상에서 실행될 수 있도록 전개될 수 있고, 상기 하나의 컴퓨터 또는 다수의 컴퓨터들은 하나의 사이트에 위치해 있거나 또는 다수의 사이트에 걸쳐 분산되어 있고 통신 네트워크에 의해 상호 접속되어 있다.

[0075] 본 명세서에 기술된 프로세스들 및 논리 흐름들은 입력 데이터를 작동하고 출력을 생성함으로써, 하나 이상의 프로그램 가능한 컴퓨터들에 의해 수행될 수 있고, 상기 프로그램 가능한 컴퓨터들은 기능들을 수행하기 위해 하나 이상의 컴퓨터 프로그램들을 실행한다. 상기 프로세스들 및 논리 흐름들 및 구현될 수 있는 장치는 또한 예를 들면, FPGA(field programmable gate way array) 또는 ASIC(application specific integrated circuit) 과 같은, 특수 목적 논리 회로에 의해 수행될 수 있다.

[0076] 컴퓨터 프로그램의 실행을 위해 적합한 컴퓨터들은, 예시로서, 일반적 또는 특수 목적 마이크로프로세서들 또는 둘다 또는 중앙 프로세싱 유닛의 임의의 다른 유형을 포함하거나 이에 기초한다. 일반적으로, 중앙 프로세싱 유닛은 읽기-전용 메모리 또는 랜덤 액세스 메모리 또는 모두로부터 명령어들 및 데이터를 받을 것이다. 컴퓨터의 필수 요소들은 명령어들을 수행 또는 실행하는 중앙 프로세싱 유닛 및 명령어들 및 데이터를 저장하기 위한 하나 이상의 메모리 디바이스들이다. 또한, 일반적으로, 컴퓨터는 하나 이상의 대량 저장 디바이스들, 예를 들면, 자기, 자기 광학 디스크들 또는 광학 디스크들을 포함하거나, 그로부터 데이터를 수신하거나 그에 데이터를 전송하거나 또는 둘다를 위해 하나 이상의 대량 저장 디바이스들에 동작적으로 연결될 수 있다. 그러나, 컴퓨터는 이러한 디바이스들을 포함할 필요가 없다. 게다가, 컴퓨터는, 몇 가지만 예를 들자면, 예를 들어, 모바일 텔레폰, PDA, 모바일 오디오 또는 비디오 플레이어, 게임 콘솔, Global Positioning System(GPS) 수신기 또는 휴대용 저장 장치(예를 들어, universal serial bus(USB) 플래시 드라이브)와 같은, 다른 디바이스 내에 포함될 수 있다.

[0077] 컴퓨터 프로그램 명령어들 및 데이터를 저장하기 위해 적합한 컴퓨터 판독가능 미디어는, 예를 들면, EPROM, EEPROM 및 플래시 메모리 디바이스들, 자기 디스크들, 예를 들면, 내장 하드 디스크 또는 탈착 가능한 디스크들, 자기 광학 디스크들 및 CD-ROM 및 DVD-ROM 디스크들과 같은, 반도체 메모리 디바이스들의 예시적인 방법에 의한 것을 포함하여 비휘발성 메모리, 미디어 및 메모리 디바이스들의 모든 형태들을 포함한다. 상기 프로세서 및 메모리는 특수 목적 논리 회로에 의해 보충되거나, 특수 목적 논리 회로에 통합될 수 있다.

[0078] 사용자와의 상호작용을 제공하기 위해, 본 명세서에 기술된 본 발명의 실시 예들은 예를 들어, CRT(cathode ray tube) 또는 LCD(liquid crystal display) 모니터와 같은, 사용자에게 정보를 디스플레이 하기 위한 디스플레이 디바이스 및 사용자가 컴퓨터에 입력을 제공할 수 있는 키보드 및 예를 들면, 마우스 또는 트랙볼과 같은, 포인팅 디바이스를 가지는 컴퓨터 상에서 구현될 수 있다. 디바이스들의 다른 종류들은 또한 사용자와의 상호작용을 제공하기 위해 사용될 수 있다. 예를 들면, 사용자에게 제공된 피드백은 예를 들면, 시각 피드백, 청각 피드백 또는 촉각 피드백과 같은, 감각 피드백의 임의의 종류일 수 있고, 사용자로부터의 입력은 음향, 음성 또는 촉각 입력을 포함하여, 임의의 형태로 수신될 수 있다. 부가적으로, 예를 들면, 웹 브라우저로부터 수신된 요청들에 대응하여 사용자의 클라이언트 디바이스 상의 웹 브라우저에 웹 페이지들을 보내는 것과 같이, 컴퓨터는 서류들을 보내고 사용자에게 의해 사용되는 디바이스로부터 서류들을 수신함으로써, 사용자와 상호작용을 할 수 있다.

[0079] 본 명세서에 기술된 본 발명의 실시 예들은 예를 들면, 데이터 서버로서, 컴퓨팅 시스템 내에서 구현될 수 있고, 상기 컴퓨팅 시스템은 백 엔드(back end) 컴포넌트를 포함하거나 또는 예를 들면, 어플리케이션 서버와 같은, 미들웨어(middleware) 컴포넌트를 포함하거나 또는 예를 들면, 그래픽 사용자 인터페이스 또는 사용자가 본 명세서에 기술된 주체의 구현들과 상호작용하는 것을 통한 웹 브라우저를 가지는 클라이언트 컴퓨터와 같은, 프론트 엔드(front end) 컴포넌트를 포함하거나 또는 이러한 하나 이상의 백 엔드, 미들웨어, 또는 프론트 엔드 컴포넌트들의 임의의 조합을 포함한다. 상기 시스템의 컴포넌트들은 예를 들면, 통신 네트워크와 같은, 임의의

형태 또는 디지털 데이터 통신의 매체에 의해 상호연결 될 수 있다. 통신 네트워크들의 예시들은 예를 들면, 인터넷과 같은, local area network("LAN") 및 wide area network("WAN")을 포함한다.

[0080] 컴퓨팅 시스템은 클라이언트들 및 서버들을 포함할 수 있다. 클라이언트 및 서버는 일반적으로 서로로부터 원격이며 전형적으로 통신 네트워크를 통해 상호작용한다. 클라이언트와 서버의 관계는 각각의 컴퓨터들 상에서 실행되고 서로에 대하여 클라이언트-서버 관계를 가지는 컴퓨터 프로그램들 덕분에 발생한다.

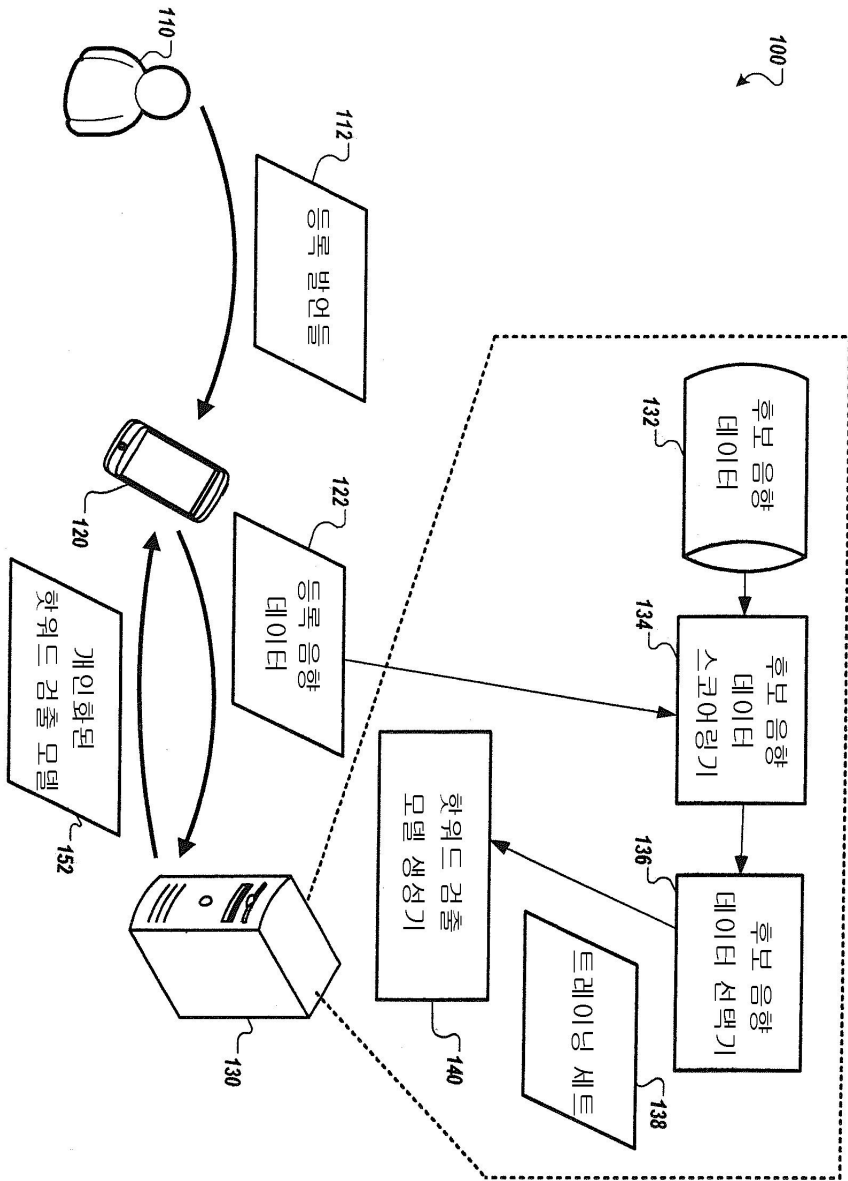
[0081] 본 명세서는 많은 특정 구현 세부사항들을 포함하나, 이들은 청구 범위에 대한 제한으로서 해석되어서는 안되며, 오히려 특정 실시예들의 구체적인 특징들의 기술들로서 해석되어야만 한다. 개별 실시예들의 문맥으로 본 명세서에 기술된 어떤 특징들은 또한 단일 실시 예에서 조합으로 구현될 수 있다. 역으로, 단일 실시예의 문맥들에서 기술된 다양한 특징들은 또한 다수의 구현들에서 개별적으로 또는 임의의 적합한 서브컴비네이션(subcombination)으로 구현될 수 있다. 게다가, 비록 구성들이 임의의 조합들로 동작하는 것으로서 상기 기술되었고, 이와 같이 초기에 청구되었지만, 청구된 조합으로부터의 하나 이상의 구성들은 일부 경우들에서 상기 조합으로부터 제외될 수 있고, 상기 주장된 조합은 서브컴비네이션 또는 서브컴비네이션의 변형에 관한 것일 수 있다.

[0082] 마찬가지로, 비록 도면들에서 동작들이 특정 순서로 표시되었지만, 이는 원하는 결과들을 달성하기 위해 이러한 동작들이 보여진 특정 순서 또는 순차적인 순서로 수행되어야 하는 것을 요구하거나 또는 모든 도시된 동작들이 수행되어야 하는 것을 요구하는 것으로 이해되어서는 안된다. 특정 상황들에서, 멀티태스킹 및 병렬 프로세싱이 유리할 수 있다. 또한, 상기 기술된 실시 예들에서 다양한 시스템 컴포넌트들의 분리는 모든 실시 예에서 이러한 분리를 요구하는 것으로 이해되어서는 안되고, 기술된 프로그램 컴포넌트들 및 시스템들은 일반적으로 단일 소프트웨어 제품으로 통합되거나 또는 다수의 소프트웨어 제품들로 패키징(package)될 수 있다는 것이 이해되어야 한다.

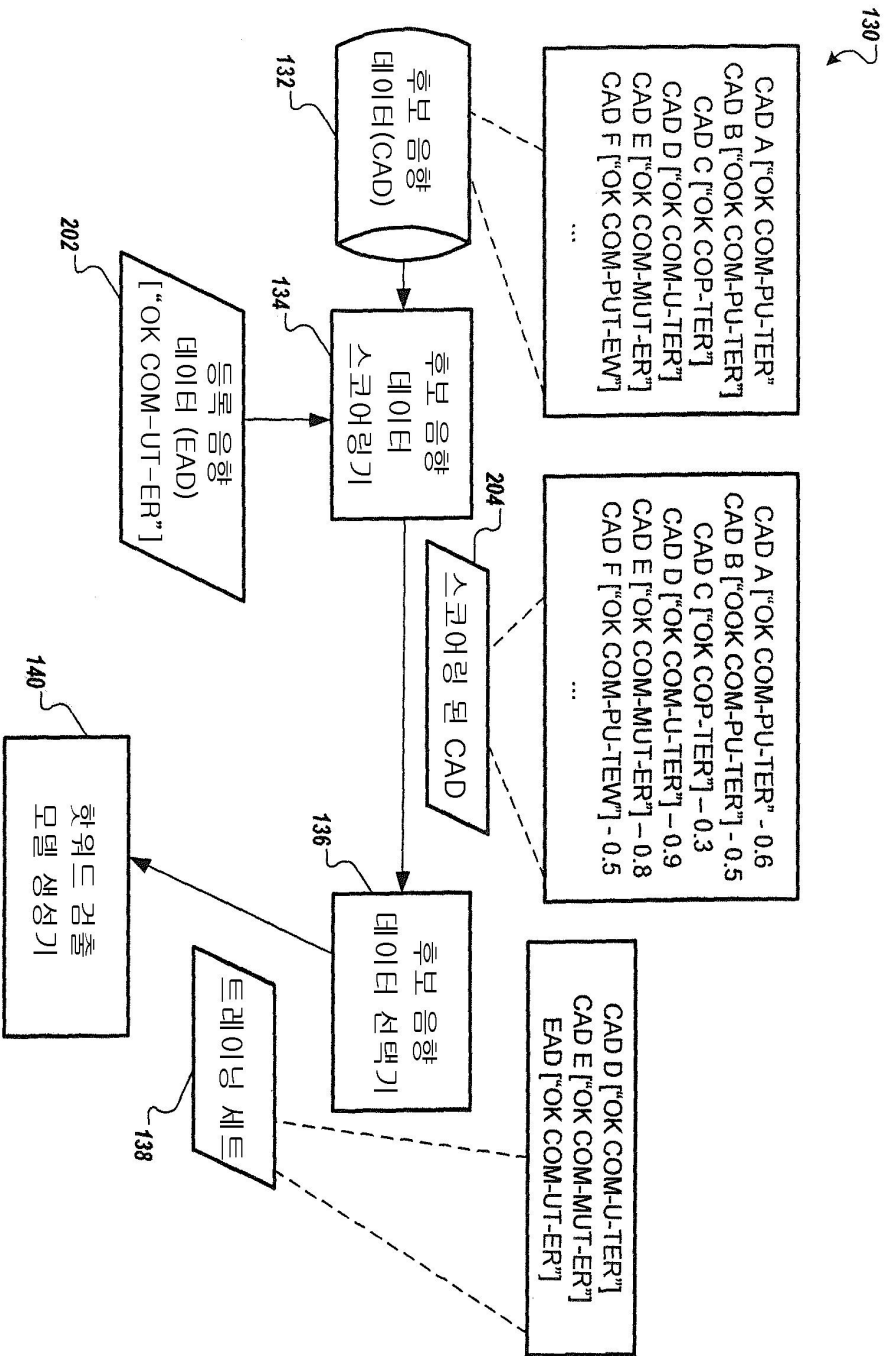
[0083] 본 발명의 특정 실시 예들이 기술되었다. 다른 실시 예들도 다음의 청구 범위 내에 있다. 예를 들면, 청구 범위에 기재된 동작들은 각각 다른 순서로 수행될 수 있고, 여전히 바람직한 결과를 달성할 수 있다. 일 예시로서, 첨부된 도면에 도시된 프로세스들은 바람직한 결과들을 달성하기 위해, 반드시 보여진 특정 순서 또는 순차적인 순서를 필요로 하지 않는다. 특정 실시 예들에서, 멀티태스킹 및 병렬 프로세싱이 유리할 수 있다. 다른 단계들이 제공될 수 있으며, 또는 기술된 프로세스들로부터 단계들이 제거될 수 있다. 따라서, 다른 실시 예들은 다음의 청구 범위 내에 있다.

도면

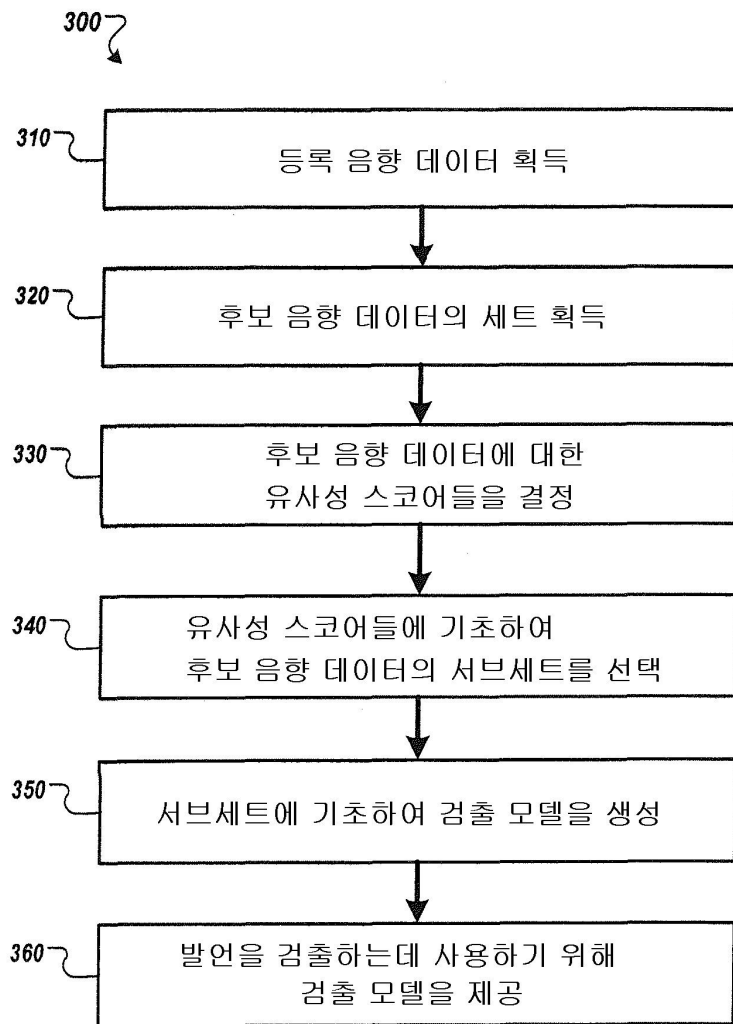
도면1



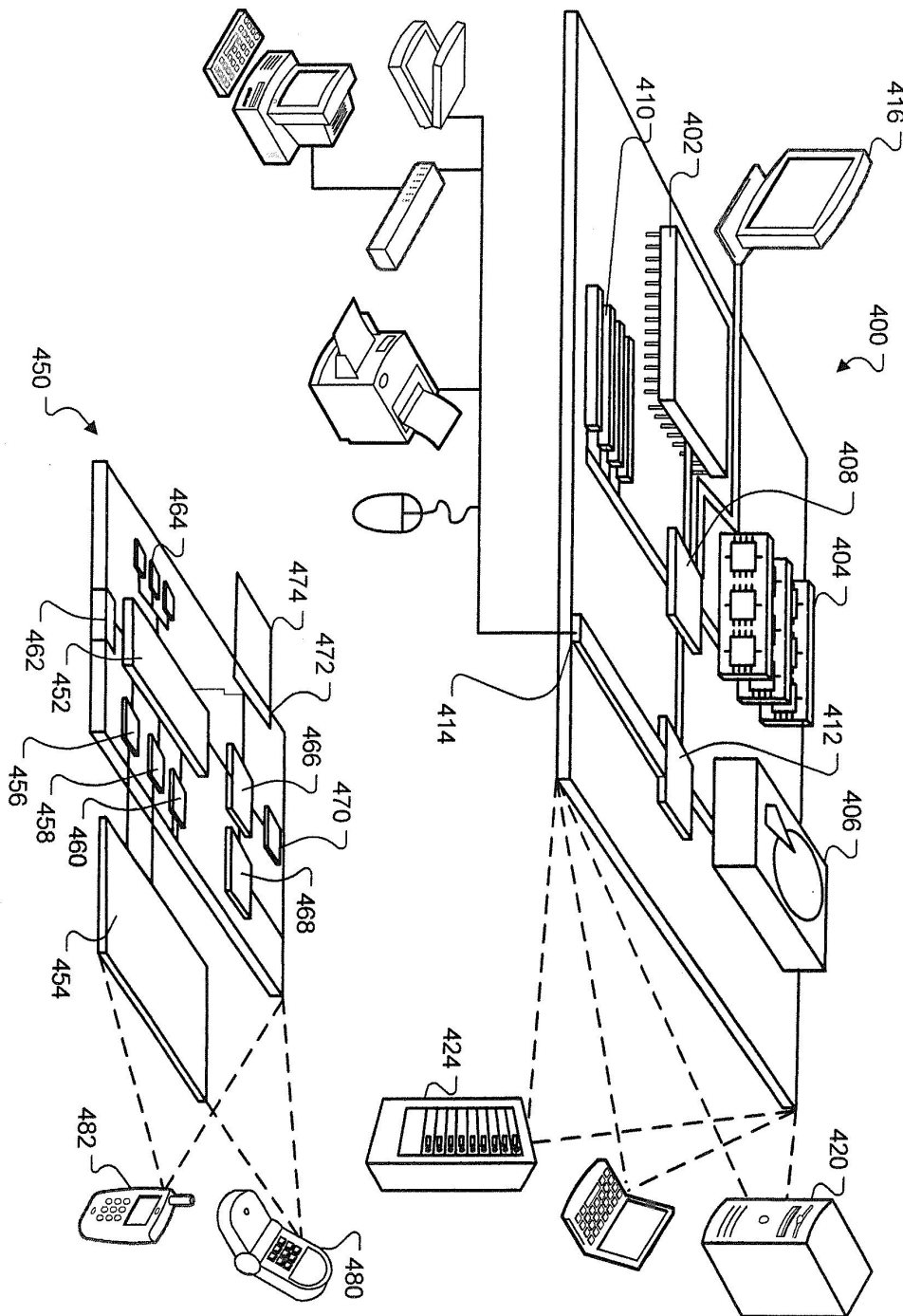
도면2



도면3



도면4



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 6

【변경전】

청구항 1 에 있어서,

상기 후속적으로 수신된 오디오 데이터에서 상기 특정 핫워드의 있을 수 있는 발언을 검출하기 위해 상기 개인화된 핫워드 검출 모델을 사용하는 단계는,

후속 발언에 대응하는 오디오 데이터를 수신하는 단계; 그리고

상기 개인화된 핫워드 검출 모델에 적어도 기초하여, 상기 후속 발언이 상기 특정 핫워드를 포함할 가능성이 있는지 여부를 결정하는 단계를 포함하는 것을 특징으로 하는 컴퓨터로 구현되는 방법.

【변경후】

청구항 1 에 있어서,

상기 개인화된 핫워드 검출 모델을 사용하여 이후에 수신된 오디오 데이터에서 상기 사용자에게 의한 특정 핫워드의 있을 수 있는 발언을 검출하는 단계는,

후속 발언에 대응하는 오디오 데이터를 수신하는 단계; 그리고

상기 개인화된 핫워드 검출 모델에 적어도 기초하여, 상기 후속 발언이 상기 특정 핫워드를 포함할 가능성이 있는지 여부를 결정하는 단계를 포함하는 것을 특징으로 하는 컴퓨터로 구현되는 방법.

【직권보정 2】

【보정항목】 청구범위

【보정세부항목】 청구항 13

【변경전】

청구항 8 에 있어서,

상기 후속적으로 수신된 오디오 데이터에서 상기 특정 핫워드의 있을 수 있는 발언을 검출하기 위해 상기 개인화된 핫워드 검출 모델을 사용하는 동작은,

후속 발언에 대응하는 오디오 데이터를 수신하는 동작; 그리고

상기 개인화된 핫워드 검출 모델에 적어도 기초하여, 상기 후속 발언이 상기 특정 핫워드를 포함할 가능성이 있는지 여부를 결정하는 동작을 포함하는 것을 특징으로 하는 시스템.

【변경후】

청구항 8 에 있어서,

상기 개인화된 핫워드 검출 모델을 사용하여 이후에 수신된 오디오 데이터에서 상기 사용자에게 의한 특정 핫워드의 있을 수 있는 발언을 검출하는 동작은,

후속 발언에 대응하는 오디오 데이터를 수신하는 동작; 그리고

상기 개인화된 핫워드 검출 모델에 적어도 기초하여, 상기 후속 발언이 상기 특정 핫워드를 포함할 가능성이 있는지 여부를 결정하는 동작을 포함하는 것을 특징으로 하는 시스템.

【직권보정 3】

【보정항목】 청구범위

【보정세부항목】 청구항 20

【변경전】

청구항 15 에 있어서,

상기 후속적으로 수신된 오디오 데이터에서 상기 특정 핫워드의 있을 수 있는 발언을 검출하기 위해 상기 개인화된 핫워드 검출 모델을 사용하는 동작은,

후속 발언에 대응하는 오디오 데이터를 수신하는 동작; 그리고

상기 개인화된 핫워드 검출 모델에 적어도 기초하여, 상기 후속 발언이 상기 특정 핫워드를 포함할 가능성이 있는지 여부를 결정하는 동작을 포함하는 것을 특징으로 하는 컴퓨터-관독가능한 매체.

【변경후】

청구항 15 에 있어서,

상기 개인화된 핫워드 검출 모델을 사용하여 이후에 수신된 오디오 데이터에서 상기 사용자에게 의한 특정 핫워드의 있을 수 있는 발언을 검출하는 동작은,

후속 발언에 대응하는 오디오 데이터를 수신하는 동작; 그리고

상기 개인화된 핫워드 검출 모델에 적어도 기초하여, 상기 후속 발언이 상기 특정 핫워드를 포함할 가능성

이 있는지 여부를 결정하는 동작을 포함하는 것을 특징으로 하는 컴퓨터-관독가능한 매체.