



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0005617
(43) 공개일자 2020년01월15일

- | | |
|--|---|
| <p>(51) 국제특허분류(Int. Cl.)
 <i>G10L 17/00</i> (2006.01) <i>G10L 15/08</i> (2006.01)
 <i>G10L 15/22</i> (2006.01) <i>G10L 17/24</i> (2013.01)
 <i>G10L 25/93</i> (2013.01)</p> <p>(52) CPC특허분류
 <i>G10L 17/005</i> (2013.01)
 <i>G10L 15/08</i> (2013.01)</p> <p>(21) 출원번호 10-2019-7036225</p> <p>(22) 출원일자(국제) 2018년08월29일
 심사청구일자 2019년12월06일</p> <p>(85) 번역문제출일자 2019년12월06일</p> <p>(86) 국제출원번호 PCT/US2018/048563</p> <p>(87) 국제공개번호 WO 2019/078959
 국제공개일자 2019년04월25일</p> <p>(30) 우선권주장
 15/785,751 2017년10월17일 미국(US)</p> | <p>(71) 출원인
 구글 엘엘씨
 미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이
 1600 (우:94043)</p> <p>(72) 발명자
 크라쿤 알렉산더
 미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이
 1600 (우:94043)
 로즈 리차드 카메론
 미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이
 1600 (우:94043)</p> <p>(74) 대리인
 박장원</p> |
|--|---|

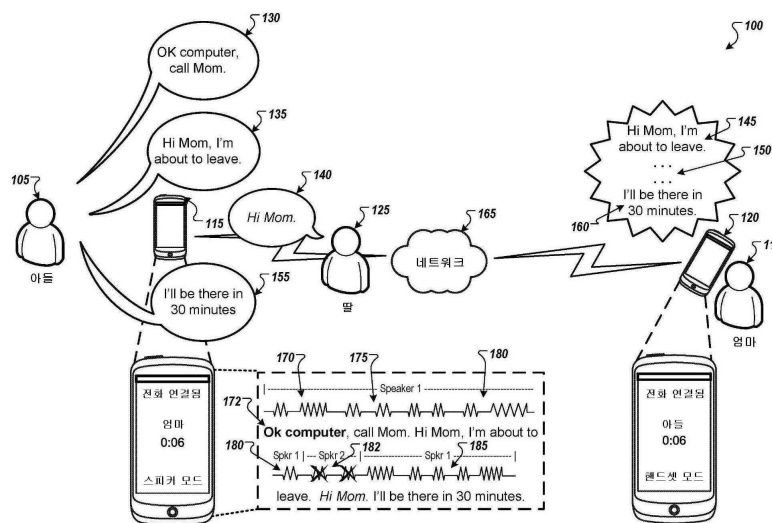
전체 청구항 수 : 총 21 항

(54) 발명의 명칭 화자 구분

(57) 요약

화자 구분(SPEAKER DIARIZATION)을 위한 컴퓨터 저장 매체에 인코딩된 컴퓨터 프로그램들을 포함하는 방법, 시스템 및 장치가 개시된다. 일 양태에서, 방법은 발언에 대응하는 오디오 데이터를 수신하는 단계의 액션을 포함한다. 상기 액션은 상기 오디오 데이터가 제1 화자에 의해 발화된 미리 정의된 핫워드의 발언을 포함한다고 결정하는 단계를 더 포함한다. 상기 액션은 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제1 부분을 식별하는 단계를 더 포함한다. 상기 액션은 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제2 부분을 식별하는 단계를 더 포함한다. 상기 액션은 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 단계를 더 포함한다.

대표도



(52) CPC특허분류

G10L 15/22 (2013.01)

G10L 17/24 (2013.01)

G10L 25/93 (2013.01)

G10L 2015/088 (2013.01)

G10L 2015/223 (2013.01)

G10L 2015/228 (2013.01)

명세서

청구범위

청구항 1

컴퓨터로 구현되는 방법으로서,

컴퓨팅 디바이스에 의해, 발언에 대응하는 오디오 데이터를 수신하는 단계;

상기 오디오 데이터가 제1 화자에 의해 발화된 미리 정의된 핫워드의 발언을 포함한다고 결정하는 단계;

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제1 부분을 식별하는 단계;

상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제2 부분을 식별하는 단계;

상기 오디오 데이터가 상기 제1 화자에 의해 발화된 상기 미리 정의된 핫워드의 발언을 포함한다고 결정하는 것에 기초하여, 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 2

청구항 1에 있어서,

상기 오디오 데이터가 상기 제1 화자에 의해 발화된 미리 정의된 핫워드의 발언을 포함한다고 결정하는 것에 기초하여, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분을 대응하는 묵음 기간으로 교체하는 단계를 더 포함하고,

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 단계는:

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분과 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분 대신에 상기 대응하는 묵음 기간을 전송하는 것을 포함하는 것을 특징으로 하는 방법.

청구항 3

청구항 2에 있어서,

상기 컴퓨팅 디바이스의 포어그라운드에서 실행되는 애플리케이션을 식별하는 단계;

상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분은 상기 컴퓨팅 디바이스의 포어그라운드에서 실행되는 상기 애플리케이션에 기초하여 대응하는 묵음 기간으로 교체되는 것을 특징으로 하는 방법.

청구항 4

청구항 1에 있어서,

상기 오디오 데이터가 상기 제1 화자에 의해 발화된 미리 정의된 핫워드의 발언을 포함한다고 결정하는 것에 기초하여, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분을 제거하는 단계를 더 포함하고,

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 단계는:

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분 대신에 대응하는 묵음 기간 없이 전송하는 것을 포함하는 것을 특징으로 하는 방법.

청구항 5

청구항 4에 있어서,

상기 컴퓨팅 디바이스의 포어그라운드에서 실행되는 애플리케이션을 식별하는 단계;

상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분은 상기 컴퓨팅 디바이스의 포어그라운드에서 실행되는 상기 애플리케이션에 기초하여 대응하는 묵음 기간으로 교체하지 않고 제거되는 것을 특징으로 하는 방법.

청구항 6

임의의 선행하는 청구항에 있어서,

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제3 부분을 식별하는 단계;

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 단계는:

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분과 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제3 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 7

청구항 6에 있어서,

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분은 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분에 선행하며, 그리고

상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분은 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제3 부분에 선행하는 것을 특징으로 하는 방법.

청구항 8

임의의 선행하는 청구항에 있어서, 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분은 상기 미리 정의된 핫워드의 발언을 포함하지 않는 것을 특징으로 하는 방법.

청구항 9

시스템으로서,

하나 이상의 컴퓨터들; 및

동작가능한 명령어들을 저장하는 하나 이상의 저장 디바이스들을 포함하며, 상기 명령어들은 상기 하나 이상의 컴퓨터들에 의해 실행될 때, 상기 하나 이상의 컴퓨터들로 하여금 동작들을 수행하게 하며, 상기 동작들은:

컴퓨팅 디바이스에 의해, 발언에 대응하는 오디오 데이터를 수신하는 동작;

상기 오디오 데이터가 제1 화자에 의해 발화된 미리 정의된 핫워드의 발언을 포함한다고 결정하는 동작;

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제1 부분을 식별하는 동작;

상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제2 부분을 식별하는 동작;

상기 오디오 데이터가 상기 제1 화자에 의해 발화된 상기 미리 정의된 핫워드의 발언을 포함한다고 결정하는 것에 기초하여, 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 동작을 포함하는 것을 특징으로 하는 시스템.

청구항 10

청구항 9에 있어서,

상기 오디오 데이터가 상기 제1 화자에 의해 발화된 미리 정의된 핫워드의 발언을 포함한다고 결정하는 것에 기초하여, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분을 대응하는 묵음 기간으로 교체하는 동작을 더 포함하고,

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 동작은:

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분과 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분 대신에 상기 대응하는 묵음 기간을 전송하는 것을 포함하는 것을 특징으로 하는 시스템.

청구항 11

청구항 10에 있어서,

상기 컴퓨팅 디바이스의 포어그라운드에서 실행되는 애플리케이션을 식별하는 동작;

상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분은 상기 컴퓨팅 디바이스의 포어그라운드에서 실행되는 상기 애플리케이션에 기초하여 대응하는 묵음 기간으로 교체되는 것을 특징으로 하는 시스템.

청구항 12

청구항 9에 있어서,

상기 오디오 데이터가 상기 제1 화자에 의해 발화된 미리 정의된 핫워드의 발언을 포함한다고 결정하는 것에 기초하여, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분을 제거하는 동작을 더 포함하고,

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 동작은:

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분 대신에 대응하는 묵음 기간 없이 전송하는 것을 포함하는 것을 특징으로 하는 시스템.

청구항 13

청구항 12에 있어서,

상기 컴퓨팅 디바이스의 포어그라운드에서 실행되는 애플리케이션을 식별하는 동작;

상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분은 상기 컴퓨팅 디바이스의 포어그라운드에서 실행되는 상기 애플리케이션에 기초하여 대응하는 묵음 기간으로 교체하지 않고 제거되는 것을 특징으로 하는 시스템.

청구항 14

청구항 9 내지 13 중 어느 한 항에 있어서, 상기 동작들은:

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제3 부분을 식별하는 동작;

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 동작은:

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분과 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제3 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 동작을 포함하는 것을 특징으로 하는 시스템.

청구항 15

청구항 14에 있어서,

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분은 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분에 선행하며, 그리고

상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분은 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제3 부분에 선행하는 것을 특징으로 하는 시스템.

청구항 16

청구항 9 내지 15에 있어서, 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분은 상기 미리 정의된 핫워드의 발언을 포함하지 않는 것을 특징으로 하는 시스템.

청구항 17

하나 이상의 컴퓨터들에 의해 실행가능한 명령어들을 포함하는 소프트웨어를 저장하는 비일시적 컴퓨터 판독가능 매체로서, 상기 명령어들은 실행시 상기 하나 이상의 컴퓨터들로 하여금 동작들을 수행하게 하며, 상기 동작들은:

컴퓨팅 디바이스에 의해, 발언에 대응하는 오디오 데이터를 수신하는 동작;

상기 오디오 데이터가 제1 화자에 의해 발화된 미리 정의된 핫워드의 발언을 포함한다고 결정하는 동작;

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제1 부분을 식별하는 동작;

상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제2 부분을 식별하는 동작;

상기 오디오 데이터가 상기 제1 화자에 의해 발화된 상기 미리 정의된 핫워드의 발언을 포함한다고 결정하는 것에 기초하여, 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 동작을 포함하는 것을 특징으로 하는 비일시적 컴퓨터 판독가능 매체.

청구항 18

청구항 17에 있어서,

상기 오디오 데이터가 상기 제1 화자에 의해 발화된 미리 정의된 핫워드의 발언을 포함한다고 결정하는 것에 기초하여, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분을 대응하는 묵음 기간으로 교체하는 동작을 더 포함하고,

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 동작은:

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분과 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분 대신에 상기 대응하는 묵음 기간을 전송하는 것을 포함하는 것을 특징으로 하는 비일시적 컴퓨터 판독가능 매체.

청구항 19

청구항 17에 있어서,

상기 오디오 데이터가 상기 제1 화자에 의해 발화된 미리 정의된 핫워드의 발언을 포함한다고 결정하는 것에 기초하여, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분을 제거하는 동작을 더 포함하고,

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 동작은:

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분 대신에 대응하는 묵음 기간 없이 전송하는 것을 포함하는 것을 특징으로 하는 비일시적 컴퓨터 판독가능 매체.

청구항 20

청구항 17 내지 19 중 어느 한 항에 있어서, 상기 동작들은:

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제3 부분을 식별하는 동작;

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 동작은:

상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분과 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제3 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 동작을 포함하는 것을 특징으로 하는 비일시적 컴퓨터 판독가능 매체.

청구항 21

컴퓨터 판독가능 명령어들로서, 상기 명령어들은 컴퓨팅 장치에 의해 실행될 때, 상기 컴퓨팅 장치로 하여금 청구항 1 내지 8 중 어느 한 항의 방법을 수행하게 하는 것을 특징으로 하는 컴퓨터 판독가능 명령어들.

발명의 설명

기술 분야

[0001] 본 명세서는 자동 음성 인식에 관한 것이다.

배경 기술

[0002] 음성 지원 가정 또는 다른 환경의 현실, 즉 사용자가 쿼리 또는 명령을 소리내어 말할 필요가 있고, 컴퓨터 기반 시스템은 쿼리에 응답하고 및/또는 명령이 수행되도록 하는 환경이 우리에게 있다. 음성 지원 환경(예: 가정, 직장, 학교 등)은 상기 환경의 다양한 방 또는 영역에 분산되어 있는 연결된 마이크로폰 디바이스들의 네트워크를 사용하여 구현될 수 있다. 이러한 마이크로폰들의 네트워크를 통해, 사용자는 자신이나 심지어 근처에 컴퓨터 또는 다른 디바이스를 가질 필요없이 본질적으로 환경의 어느 곳에서나 시스템에 구두로 쿼리할 수 있는 권한을 가진다. 예를 들어, 부엌에서 요리하는 동안, 사용자는 시스템에 "세 컵이 몇 밀리리터야?"를 물을 수 있으며, 이에 응답하여, 시스템으로부터 예를 들어, 합성된 음성 출력의 형태로 응답을 받을 수 있다. 대안적으로, 사용자는 시스템에 "가장 가까운 주유소가 언제 문 닫아?" 또는 집을 나갈 준비를 할 때 "오늘 코트를 입어야 해?"와 같은 질문을 할 수 있다.

[0003] 또한, 사용자는 시스템에 쿼리를 묻고 및/또는 사용자의 개인 정보와 관련된 명령을 발행할 수 있다. 예를 들어, 사용자는 "준과의 미팅은 언제야?"라고 시스템에 묻거나 "집으로 갈 때 존에게 전화하라고 알려줘"라고 명령할 수 있다.

발명의 내용

[0004] 음성 지원 시스템에서, 시스템과 인터랙션하는 사용자의 방식은 주로, 배타적이지 않은 경우, 음성 입력에 의하도록 구성된다. 결과적으로, 시스템에 지시되지 않은 것들을 포함하여 주변 환경에서 만들어진 모든 발언을 잠재적으로 포착하는 시스템은 주어진 발언이 예를 들어, 환경에서 다른 사람에게 향하는 것과 반대로 시스템에 대해 지시된 것으로 구별하는 방식이 있어야 한다. 이를 수행하는 한 가지 방법은 환경에서 사용자 간에 동의에 의해 시스템의 주의를 끌기 위해 미리 결정된 단어로 예약된 핫워드를 사용하는 것이다. 예시적 환경에서, 시스템의 주의를 끌기 위해 사용되는 핫워드는 "OK computer"이다. 결과적으로, "OK computer"라는 단어가 발화될 때마다, 이는 마이크로폰에 의해 포착되어 시스템에 전달되고, 음성 인식 기법들을 수행하거나 핫워드가 발화되었는지 여부를 결정하기 위해 오디오 피쳐들 및 신경 네트워크들을 사용하고 그렇다면, 후속 명령 또는 쿼리를 기다린다. 따라서 시스템을 향한 발언은 일반적인 형태인 [핫워드] [쿼리]이며, 여기서 이 예시에서 "핫워드"는 "OK computer"이고 "쿼리"는 임의의 질문, 명령, 선언 또는 단독으로 또는 네트워크를 통해 서버와 함께 음성 인식되고, 파싱되고 시스템에 의해 동작될 수 있는 다른 요청일 수 있다.

[0005] 음성 지원 시스템은 핫워드의 발언을 사용자가 시스템과 인터랙션하려는 의도의 표시로서 사용할 수 있다. 음성 지원 시스템이 상이한 사용자들로부터 음성을 검출하는 경우, 시스템은 핫워드를 처음 말한 사용자로부터의 음성을 포함하는 오디오 데이터를 프로세싱하고 전송하며, 핫워드를 말하지 않은 다른 사용자로부터의 음성을 포함하는 오디오 데이터의 프로세싱을 제한하고 전송을 억제할 것이다. 시스템은 핫워드를 사용하여 핫워드를 포

함하는 오디오 데이터의 부분을 식별할 수 있다. 화자 구분(SPEAKER DIARIZATION) 모듈은 핫워드를 포함하는 오디오 데이터의 부분을 분석하여 사용자의 음성 특성을 식별하고, 동일한 사용자로부터의 음성을 포함하는 후속적으로 수신된 오디오 데이터를 식별할 수 있다. 화자 구분 모듈은 이후에 수신된 다른 음성 오디오를 분석하고, 화자가 핫워드 화자와 동일한 화자가 아닌 오디오 부분을 식별할 수 있다. 다른 사용자는 핫워드를 말함으로써 시스템과 인터렉션하려는 의도를 표현하지 않았기 때문에, 시스템은 다른 사용자가 말한 부분을 제거할 수 있다. 다른 사용자가 말한 부분을 제거함으로써, 시스템은 의도하지 않은 음성 지원 시스템과 인터렉션할 수 있는 다른 사용자의 프라이버시를 보호한다.

[0006] 본 명세서에 기술된 발명의 일 혁신적인 양태에 따르면, 액션들을 포함하는 화자 구분을 위한 방법을 포함하며, 상기 액션들은: 컴퓨팅 디바이스에 의해, 발언에 대응하는 오디오 데이터를 수신하는 단계; 상기 오디오 데이터가 제1 화자에 의해 발화된 미리 정의된 핫워드의 발언을 포함한다고 결정하는 단계; 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제1 부분을 식별하는 단계; 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제2 부분을 식별하는 단계; 상기 오디오 데이터가 상기 제1 화자에 의해 발화된 상기 미리 정의된 핫워드의 발언을 포함한다고 결정하는 것에 기초하여, 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 단계를 포함한다.

[0007] 이들 또는 다른 실시예들은 다음 구성들 중 하나 이상을 각각 선택적으로 포함할 수 있다. 상기 액션들은 상기 오디오 데이터가 상기 제1 화자에 의해 발화된 미리 정의된 핫워드의 발언을 포함한다고 결정하는 것에 기초하여, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분을 대응하는 묵음 기간으로 교체하는 단계를 더 포함한다. 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 단계는 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분과 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분 대신에 상기 대응하는 묵음 기간을 전송하는 것을 포함한다. 상기 액션은 상기 컴퓨팅 디바이스의 포어그라운드에서 실행되는 애플리케이션을 식별하는 단계를 더 포함한다.

[0008] 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분은 상기 컴퓨팅 디바이스의 포어그라운드에서 실행되는 상기 애플리케이션에 기초하여 대응하는 묵음 기간으로 교체된다. 상기 액션은 상기 오디오 데이터가 상기 제1 화자에 의해 발화된 미리 정의된 핫워드의 발언을 포함한다고 결정하는 것에 기초하여, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분을 제거하는 단계를 더 포함한다. 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 단계는 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분 대신에 대응하는 묵음 기간 없이 전송하는 것을 포함한다. 상기 액션은 상기 컴퓨팅 디바이스의 포어그라운드에서 실행되는 애플리케이션을 식별하는 단계를 더 포함한다. 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분은 상기 컴퓨팅 디바이스의 포어그라운드에서 실행되는 상기 애플리케이션에 기초하여 대응하는 묵음 기간으로 교체하지 않고 제거된다.

[0009] 상기 액션은 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제3 부분을 식별하는 단계를 더 포함한다. 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 단계는 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분과 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제3 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제하는 단계를 포함한다. 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분은 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분에 선행한다. 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분은 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제3 부분에 선행한다. 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분은 상기 미리 정의된 핫워드의 발언을 포함하지 않는다.

[0010] 본 양태의 다른 실시예는 대응하는 컴퓨터 시스템, 장치 및 하나 이상의 컴퓨터 저장 디바이스들에 기록된 컴퓨터 프로그램들을 포함하며, 그 각각은 방법들의 동작들을 수행하도록 구성된다.

[0011] 본 명세서에 기술된 본 발명의 특정한 실시예들은 다음의 이점들을 실현하도록 구현될 수 있다. 핫워드를 말하

여 시스템과의 인터랙션을 초기화하지 않은 사용자로부터의 음성은 인터랙션을 초기화 한 사용자의 음성을 단독 화할 수 있다. 따라서, 인터랙션을 초기화하지 않은 사용자의 음성의 전송을 억제하는 것은 전송된 오디오 데이터가 단독화되거나 이해할 수 없고 결과적으로 시스템에 의해 오해되는 것으로 표현되는 음성의 발생을 감소시키는 역할을 할 수 있다. 전송된 오디오 데이터가 명령 또는 쿼리(예를 들어, 디지털 어시스턴트)를 나타내는 구현예에서, 인터랙션을 초기화하지 않은 사용자의 음성의 전송을 억제함으로써 부적절한 응답이 제공되거나 부적절한 액션이 수행되는 경우를 감소시킬 수 있다. 또한, 사용자의 음성이 시스템의 마이크로폰에 의해 검출되고 사용자가 시스템과의 인터랙션을 개시하기 위해 핫워드를 말하지 않은 경우, 사용자의 프라이버시가 보호될 수 있다. 시스템은 핫워드를 말함으로써 시스템과의 인터랙션을 초기화하지 않은 사용자로부터의 음성을 제거할 수 있다. 다시 말해서, 관계없는 사람의 음성이 마이크에 의해 검출되고 관계없는 사람이 시스템과 명시적으로 인터랙션하기 위해 핫워드를 말하지 않은 경우 관계없는 사람의 프라이버시를 보호할 기회가 있다. 시스템은 핫워드를 말함으로써 시스템과의 인터랙션을 초기화하지 않은 관계없는 사람으로부터의 음성을 제거할 수 있다.

[0012] 본 명세서에서 본 발명의 하나 이상의 실시예들의 세부 사항은 첨부 도면과 아래의 설명에서 기술된다. 본 발명의 다른 구성들, 양태들 및 이점들은 설명, 도면 및 청구항으로부터 명백해질 것이다.

도면의 간단한 설명

[0013] 도 1은 전화 통화 중 화자 구분을 사용하여 오디오를 편집하는 예시적 시스템을 도시한다.

도 2는 화자 구분을 구현하기 위한 예시적 시스템을 도시한다.

도 3는 화자 구분을 위한 예시적 프로세스의 흐름도이다.

도 4는 컴퓨팅 디바이스 및 모바일 컴퓨팅 디바이스의 예를 도시한다.

다양한 도면들에서 동일한 참조 번호 및 기호는 동일한 구성요소를 표시한다.

발명을 실시하기 위한 구체적인 내용

[0014] 도 1은 전화 통화 중 화자 구분을 사용하여 오디오를 편집하는 예시적 시스템(100)을 도시한다. 간략하게, 그리고 아래에서 더 상세히 설명되는 바와 같이, 아들(105)과 엄마(110)는 네트워크(165)를 통해 컴퓨팅 디바이스(115) 및 컴퓨팅 디바이스(120)를 사용하여 전화 통화를 중이다. 아들(105)은 핫워드 "ok computer" 및 "call mom" 명령을 포함하는 발언(130)을 말함으로써 전화 통화를 개시했다. 전화 통화 동안, 컴퓨팅 디바이스(115)의 마이크로폰은 아들(105)과 딸(125) 모두로부터 음성을 검출한다. 아들(105)이 핫워드를 말했다기 때문에, 컴퓨팅 디바이스(115)는 딸(125)과 같은 다른 사용자가 아닌 아들(105)로부터의 음성을 포함하는 오디오 데이터를 전송한다.

[0015] 도 1에 도시된 예에서, 아들(105)은 발언(130) "OK computer, call mom"라고 말함으로써 전화 통화를 개시한다. 컴퓨팅 디바이스(115)는 발언(130)을 검출하고 대응하는 오디오 데이터를 프로세싱한다. 컴퓨팅 디바이스(115)는 전화, 스마트폰, 랩톱 컴퓨터, 데스크톱 컴퓨터, 스마트 스피커, 텔레비전, 스마트 워치 또는 임의의 다른 유사한 디바이스와 같은 전화를 걸도록 구성된 임의의 유형의 디바이스일 수 있다.

[0016] 컴퓨팅 디바이스(115)는 오디오 데이터(170)를 프로세싱하고, 핫워드(172) "ok computer"를 식별한다. 컴퓨팅 디바이스(115)는 오디오 데이터(175)를 계속 프로세싱하고, 명령(172) "call mom"를 식별한다. 컴퓨팅 디바이스(115)는 네트워크(165)를 통해 컴퓨팅 디바이스(120)로 컴퓨팅 디바이스(115)의 연락처에서 "엄마"에 대한 전화 번호에 대응하는 전화 통화를 개시한다.

[0017] 전화 통화는 네트워크(165)를 통해 컴퓨팅 디바이스(115)와 컴퓨팅 디바이스(120) 사이를 연결한다. 전화 통화 중, 사용자(105)는 "Hi Mom, I'm about to leave"라고 말하는 발언(135)을 한다. 컴퓨팅 디바이스(115)의 마이크로폰은 발언(135)을 검출하고, 컴퓨팅 디바이스(115)는 대응하는 오디오 데이터를 프로세싱한다. 컴퓨팅 디바이스(115)는 동일한 사람이 발언(130)에서 핫워드 및 발언(135)을 핫워드를 말한 것으로 결정한다. 컴퓨팅 디바이스(115)는 발언(135)의 오디오 데이터(180)를 네트워크(165)를 통해 컴퓨팅 디바이스(120)에 전송한다. 컴퓨팅 디바이스(120)의 오디오 스피커는 발언(135)에 대응하는 오디오(145)를 출력한다. 엄마(110)는 "Hi Mom, I'm about to leave."를 듣는다.

[0018] 전화 통화가 계속된다. 딸(125)은 아들(105)이 스피커 모드에서 또는 주변 영역의 임의의 소리를 포착하는 스마트 스피커를 사용하여 전화 통화를 하고 있는 방에 들어갈 수 있다. 딸(125)은 아들(105)이 엄마(110)와 대화하고 있음을 인식하고 발언(135)을 말한다. 컴퓨팅 디바이스(115)의 마이크로폰은 "Hi Mom"이라는 발언(140)을 말

하는 딸(125)을 검출한다. 컴퓨팅 디바이스(115)는 발언(140)에 대응하는 오디오 데이터(182)를 프로세싱한다. 컴퓨팅 디바이스(115)는 발언(140)과 발언(130)에서 핫워드(172)를 상이한 사람이 말한 것으로 결정한다. 컴퓨팅 디바이스(115)는 발언(130)에서 핫워드(172)를 말한 사람과 다른 사람이 발언(140)을 말했기 때문에 발언(140)에 대응하는 오디오 데이터(182)의 전송을 억제한다. 이 경우에, 딸(125)은 발언(140)을 발화했다. 딸(125)은 우연의 사용자 또는 의도하지 않은 사용자 또는 둘 모두일 수 있다.

[0019] 컴퓨팅 디바이스(115)가 스마트폰 또는 스마트 스피커와 같은 전화 애플리케이션을 실행하는 컴퓨팅 디바이스인 경우, 컴퓨팅 디바이스(115)는 발언(140)의 오디오 데이터(182)를 전송하는 대신에 대응하는 묵음 기간(150)을 전송할 수 있다. 예를 들어, 딸(125)이 1초 동안 발언(140)을 말하면, 컴퓨팅 디바이스(115)는 대응하는 오디오 데이터(182)가 컴퓨팅 디바이스(120)로 전송되는 것을 억제하고, 대신에 1초의 묵음(150)에 대응하는 오디오 데이터를 전송한다. 발언(135)과 발언(140)의 타이밍을 보존함으로써, 대화는 어머니(110)에게 변경된 소리를 내지 않을 수 있다. 대신에, 어머니(110)는 아들(105)이 대화 중에 일시 중지하였다고 생각할 수 있다.

[0020] 아들(105)은 계속해서 "I'll be there in 30 minutes"라는 발언(155)을 발화한다. 컴퓨팅 디바이스(115)의 마이크로폰은 발언(155)을 검출한다. 컴퓨팅 디바이스(115)는 발언(155)에 대응하는 오디오 데이터(185)를 프로세싱한다. 컴퓨팅 디바이스(115)는 동일한 사람이 발언(130)에서 핫워드(172) 및 발언(155)을 핫워드를 말한 것으로 결정한다.

[0021] 컴퓨팅 디바이스(115)는 발언(155)의 오디오 데이터(185)를 네트워크(165)를 통해 컴퓨팅 디바이스(120)에 전송한다. 컴퓨팅 디바이스(120)의 오디오 스피커는 발언(155)에 대응하는 오디오(160)를 출력한다. 엄마(110)는 "I'll be there in 30 minutes"을 듣는다. 아들(105) 또는 엄마(110)가 통화를 종료할 때까지 전화 통화는 계속된다.

[0022] 도 2는 화자 구분을 구현하기 위한 예시적 시스템(200)을 도시한다. 시스템(200)은 음성 오디오를 수신, 프로세싱 및 전송하도록 구성된 임의의 유형의 컴퓨팅 디바이스일 수 있다. 예를 들어, 시스템(200)은 도 1의 컴퓨팅 디바이스(115) 또는 컴퓨팅 디바이스(120)와 유사할 수 있다. 시스템(200)의 컴포넌트들은 단일 디바이스 또는 다수의 디바이스들에 걸쳐 분산되어 구현될 수 있다.

[0023] 시스템(200)은 오디오 서브시스템(202)을 포함할 수 있다. 오디오 서브시스템(202)은 마이크로폰(204), 아날로그-디지털 변환기(206), 버퍼(208) 및 다양한 다른 오디오 필터들을 포함할 수 있다. 마이크로폰(204)은 음성과 같은 주변 영역의 사운드를 검출하도록 구성될 수 있다. 아날로그-디지털 변환기(206)는 마이크로폰(204)에 의해 검출된 오디오 데이터를 샘플링하도록 구성될 수 있다. 버퍼(208)는 시스템(200)에 의한 프로세싱을 위해 샘플링된 오디오 데이터를 저장할 수 있다. 일부 구현예에서, 오디오 서브시스템(202)은 지속적으로 활성화될 수 있다. 이 경우, 마이크로폰(204)은 지속적으로 소리를 검출할 수 있다. 아날로그-디지털 변환기(206)는 검출된 오디오 데이터를 지속적으로 샘플링할 수 있다. 버퍼(208)는 마지막 10초의 소리와 같은 최신 샘플링된 오디오 데이터를 저장할 수 있다. 시스템(200)의 다른 컴포넌트들이 버퍼(208)에서 오디오 데이터를 프로세싱하지 않으면, 버퍼(208)는 이전 오디오 데이터를 덮어 쓸 수 있다.

[0024] 도 2에 도시된 예에서, 마이크로폰(204)은 "Hi Mom, I'm about to leave. Hi Mom. I'll be there in 30 minutes"에 대응하는 발언을 검출할 수 있다. 아날로그-디지털 변환기(206)는 수신된 오디오 데이터를 샘플링할 수 있고, 버퍼(208)는 샘플링된 오디오 데이터(212)를 저장할 수 있다.

[0025] 오디오 서브시스템(202)은 오디오 데이터(212)를 핫워드(210)에 제공한다. 핫워드(210)는 마이크로폰(204)을 통해 수신되고 및/또는 버퍼(208)에 저장된 오디오에서 핫워드를 식별하도록 구성된다. 핫워드(210)는 핫워드 검출기, 키워드 스포터 또는 키워드 검출기로 지칭될 수 있다. 일부 구현예에서, 핫워드(210)는 시스템(200)의 전원이 켜질 때마다 활성화될 수 있다. 핫워드(210)는 버퍼(208)에 저장된 오디오 데이터를 지속적으로 분석한다. 핫워드(210)는 버퍼(208) 내의 현재 오디오 데이터가 핫워드를 포함할 가능성을 반영하는 핫워드 신뢰 점수를 계산한다. 핫워드 신뢰 점수를 계산하기 위해, 핫워드(210)는 핫워드 모델(214)을 사용할 수 있다. 핫워드(210)는 필터 뱅크 에너지 또는 멜-주파수 썸스트럴 계수(mel-frequency cepstral coefficient)들과 같은 오디오 데이터(212)로부터 오디오 피쳐들을 추출할 수 있다. 핫워드(210)는 지원 벡터 머신(support vector machine) 또는 신경 네트워크를 사용하는 것과 같은 이들 오디오 피쳐들을 프로세싱하기 위해 분류 윈도우들을 사용할 수 있다. 일부 구현예에서, 핫워드(210)는 핫워드 신뢰 점수를 결정하기 위해 음성 인식을 수행하지 않는다. 핫워드(210)는 핫워드 신뢰 점수가 핫워드 신뢰 점수 임계치를 만족하면 오디오가 핫워드를 포함하는 것으로 결정한다. 예를 들면, 핫워드(210)는 핫워드 신뢰 점수가 핫워드 신뢰 점수가 0.8이고 핫워드 신뢰 점수 임계치가 0.7 이면, 오디오가 핫워드를 포함하는 것으로 결정한다.

- [0026] 도 2에 도시된 예에서, 핫워드(210)는 오디오 데이터(212)의 제1 화자가 핫워드 "ok computer"를 말한 것으로 결정한다. 핫워드(210)는 핫워드를 포함하는 오디오 데이터(212)의 부분을 식별할 수 있다. 예를 들어, 핫워드(210)는 0.0 내지 0.5초 사이의 오디오 데이터가 핫워드(220)를 포함한다는 것을 명령 식별기(216) 및 구분 모듈(218)에 제공할 수 있다. 핫워드(210)는 화자가 핫워드를 발인했음을 나타내는 데이터를 명령 식별기(216)에 제공할 수 있고, 0.0 내지 0.5초 사이의 오디오 데이터가 핫워드를 포함한다는 것을 구분 모듈(218)에 제공할 수 있다.
- [0027] 일부 구현예에서, 핫워드(210)는 화자 식별 능력들을 포함할 수 있다. 이 경우, 핫워드(210)는 핫워드를 말한 특정한 사람을 식별할 수 있다. 예를 들어, 핫워드(210)는 엘리스가 핫워드를 말했다는 가능성이 있다고 결정할 수 있다. 핫워드(210)는 엘리스가 핫워드를 말했다고 표시하는 데이터를 구분 모듈(218)에 제공할 수 있다. 화자 식별을 수행하기 위해, 핫워드(210)는 사용자가 다양한 문구를 반복하도록 요청함으로써 사용자에 대한 음성 데이터를 사전에 수집할 수 있다. 핫워드(210)는 수집된 음성 샘플을 사용하여 화자 식별 모델을 트레이닝할 수 있다.
- [0028] 명령 식별기(216)는 오디오 서브시스템(202) 또는 핫워드(210)로부터 오디오 데이터(212)를 수신한다. 일부 구현예에서, 명령 식별기(216)는 핫워드(210)가 발화된 핫워드를 식별하거나 사용자가 예를 들어 명령 버튼을 선택함으로써 사용자가 명령을 말하고 있다는 것을 표시하지 않으면 명령에 대한 오디오 데이터(212)를 검색하지 않을 수 있다. 명령 식별기(216)는 오디오 데이터(212)에 대해 음성 인식을 수행하는 음성 인식기(222)를 포함할 수 있다. 음성 인식기(222)는 명령 식별기(216)가 명령들에 대해 분석하는 전사를 생성할 수 있다. 명령 식별기(216)는 전사 내의 임의의 용어가 명령들(224)에서의 용어와 일치 하는지 여부를 결정하기 위해 전사를 명령(224)과 비교할 수 있다. 또한, 명령 식별기(216)는 오브젝트들을 요구하는 명령에 대한 명령들의 오브젝트들을 식별할 수 있다. 예를 들어, "call" 명령에는 오브젝트가 필요할 수 있다. 명령 식별기(216)는 "call"에 후속하는 용어가 예를 들어 "mom"과 같은 오브젝트일 수 있다고 결정할 수 있다. 따라서 명령은 "call mom"이다.
- [0029] 일부 구현예에서, 명령 식별기(216)는 타이밍 데이터(226)를 구분 모듈(218)에 제공할 수 있다. 예를 들어, 명령 식별기(216)는 명령 "call mom"이 0.7초 내지 1.3초 사이에서 발생 함을 나타낼 수 있다. 도 2의 예에 따르면, 오디오 데이터(212)는 0.0초와 0.5초 사이의 핫워드 및 0.7초와 1.3초 사이의 명령을 포함한다. 일부 구현예에서, 명령 식별기(216)는 명령 식별기(216)가 명령을 식별할 때까지 음성 인식기(222)를 사용하여 오디오 데이터(212)를 프로세싱한다. 명령 식별기(216)는 명령, 예를 들어 "call mom"을 식별한 후 음성 인식 수행을 중지하도록 음성 인식기(222)에 지시할 수 있다.
- [0030] 구분 모듈(218)은 오디오 데이터(212)를 분석하고, 상이한 사용자들에 의해 발화된 오디오 데이터의 부분을 식별한다. 구분 모듈(218)은 오디오 데이터(212), 오디오 데이터(212)가 0.0 내지 0.5초 사이에 핫워드를 포함한다는 것을 나타내는 데이터 및 오디오 데이터(212)가 0.7 내지 1.3초 사이에 명령을 포함한다는 것을 나타내는 데이터를 수신한다. 구분 모듈(218)은 구분 모델(234)을 적용함으로써 오디오 데이터(212)의 부분을 핫워드 화자에 의해 또는 핫워드 화자에 의해 발화되지 않은 것으로 라벨링한다. 일부 구현예에서, 구분 모듈(218)은 화자에 따라 오디오 데이터(212)의 부분들을 라벨링한다. 구분 모델(234)은 특정한 사람으로부터의 음성을 식별하도록 트레이닝되지 않을 수 있다. 구분 모듈(218)은 구분 모델(234)을 오디오 데이터(212)에 적용하여, 구분 모델(234)이 동일한 화자에 대한 데이터를 포함하지 않더라도 공통된 화자에 의해 발화된 부분들을 식별한다. 구분 모듈(218)은 동일한 사람에 의해 발화된 부분에서 패턴을 식별할 수 있다. 예를 들어, 구분 모듈(218)은 공통 피치를 갖는 부분을 식별할 수 있다.
- [0031] 도 2에 도시된 예에서, 구분 모듈(218)은 오디오 데이터(212)를 분석한다. 구분 모듈(218)은 0.0 내지 0.5초 사이에 핫워드를 포함하는 오디오 데이터(212)의 화자를 화자(1)로서 라벨링한다. 일부 구현예에서, 구분 모듈(218)은 명령을 포함하는 오디오 데이터(212)의 부분들을 프로세싱하지 않을 수 있다. 이 경우, 구분 모듈(218)은 0.7 내지 1.3초 사이의 오디오 데이터 부분에 화자 라벨을 적용하지 않을 수 있다. 구분 모듈(218)은 "Hi Mom, I'm about to leave"에 대응하는 오디오 데이터의 부분(228)이 화자(1)에 의해 발화된 것으로 결정한다. 구분 모듈(218)은 "Hi Mom"에 대응하는 오디오 데이터의 부분(230)이 화자(1)과 상이한 화자(2)에 의해 발화된 것으로 결정한다. 구분 모듈(218)은 "I'll be there in 30 minutes"에 대응하는 오디오 데이터의 부분(232)이 화자(1)에 의해 발화된 것으로 결정한다. 구분 모듈(218)은 오디오 데이터(228, 230 및 232)의 부분들을 서로 지정하기 위해 타이밍 데이터를 생성할 수 있다. 타이밍 데이터는 각 부분의 시작과 종료를 식별할 수 있다.
- [0032] 일부 구현예에서, 구분 모듈(218)은 시스템(200)이 서버 또는 모바일폰과 같은 다른 컴퓨팅 디바이스에 전송할 오디오 데이터를 프로세싱할 수 있다. 예를 들어, 구분 모듈(218)은 프로세싱하여 명령 "call mom" 이후에 시작

하는 오디오 데이터의 화자들을 식별할 수 있는데, 이는 그것이 전화 통화에서 다른 사용자의 컴퓨팅 디바이스에 전송되는 오디오 데이터이기 때문이다.

[0033] 구분 모듈(218)은 오디오 데이터(236)를 오디오 편집기(238)에 제공한다. 오디오 편집기(238)는 애플리케이션 설정(240) 및 사용자 설정(242)에 따라 핫워드 화자가 아닌 화자에 의해 발화된 오디오 데이터의 부분을 제거하기 위해 오디오 데이터(236)를 업데이트한다. 도 2의 예에서, 오디오 편집기(238)는 핫워드를 발화하지 않은 화자, 화자(2)에 의해 발화되었던 오디오 부분(248)을 제거하고, 핫워드 화자, 화자(1)에 의해 발화되었던 오디오 부분(246 및 250)을 유지함으로써 오디오 데이터(244)를 생성한다. 일부 구현예에서, 오디오 편집기(238)는 오디오 부분(248)을 대응하는 묵음 기간으로 교체한다. 일부 구현예에서, 오디오 편집기(238)는 오디오 부분(248)에 대한 대응 묵음 기간 없이 오디오 부분(246)과 오디오(250)를 이어붙인다.

[0034] 오디오 편집기(238)는 애플리케이션 설정(240) 및 사용자 설정(242)에 액세스하여 비핫워드 화자의 오디오를 대응하는 묵음 기간으로 교체할지 또는 핫워드 화자의 오디오를 함께 이어 붙일 것인지 여부를 결정한다. 예를 들어, 애플리케이션은 전화 애플리케이션일 수 있다. 전화 애플리케이션에 대한 애플리케이션 설정(240)은 오디오 데이터(236)의 타이밍을 보존하기 위해 지시할 수 있다. 이 경우, 오디오 편집기(238)는 오디오(248)를 대응하는 묵음 부분으로 교체함으로써 오디오 데이터(244)를 생성한다. 다른 예시로서, 애플리케이션은 검색 애플리케이션일 수 있다. 음성 메모 애플리케이션에 대한 애플리케이션 설정(240)은 전화 애플리케이션과 같은 타이밍에 의존하지 않을 수 있다. 이 경우, 오디오 편집기(238)는 오디오 데이터(246)와 오디오 데이터(250)를 함께 이어 붙임으로써 오디오 데이터(244)를 생성한다.

[0035] 일부 구현예에서, 오디오 편집기(238)는 다른 컴퓨팅 디바이스로 전사를 전송하기 전에 오디오 데이터(236)의 전사를 편집할 수 있다. 오디오 편집기(238)는 오디오 부분(248)에 대한 전사를 포함하지 않는 오디오 데이터(244)의 전사를 전송할 수 있다. 예를 들어, 오디오 편집기(238)는 "Hi Mom, I'm about to leave. I'll be there in 30 minutes"을 전송할 수 있다. 시스템(200)은 메시징 애플리케이션과 같은 애플리케이션에 대한 전사를 전송할 수 있다.

[0036] 오디오 편집기(238)는 오디오 데이터(236)를 편집하는 방법을 결정하기 위해 사용자 설정(242)에 액세스할 수 있다. 사용자 설정(244)은 핫워드 화자 및 비핫워드 화자 모두의 음성을 포함하는 오디오 데이터를 편집하는 방법과 관련된 설정을 포함할 수 있다. 일부 구현예에서, 사용자 설정(244)은 핫워드 화자 및 비핫워드 화자 모두의 음성을 포함하는 오디오 데이터를 유지하도록 표시할 수 있다. 일부 구현예에서, 사용자 설정(244)은 핫워드 화자 및 비핫워드 화자 모두의 음성을 포함하는 오디오 데이터를 제거하도록 표시할 수 있다. 일부 구현예에서, 사용자 설정(240)은 핫워드 화자 및 비핫워드 화자 모두의 음성을 포함하는 오디오 데이터에 대한 편집 규칙을 포함할 수 있다.

[0037] 시스템은 사용자 인터페이스 생성기(252)를 포함한다. 사용자 인터페이스 생성기(252)는 오디오 데이터(236)를 편집하기 위한 옵션을 사용자에게 제시하는 사용자 인터페이스를 생성할 수 있다. 예를 들어, 사용자 인터페이스는 오디오 데이터(244)를 생성하기 위한 옵션을 포함할 수 있다. 사용자는 오디오 데이터를 전송하기 전에 오디오 부분(248)을 제거하는 옵션을 선택할 수 있다. 사용자는 오디오 부분(248)의 제거를 무시하고, 오디오 데이터를 다른 컴퓨팅 디바이스로 전송할 때 오디오 부분(248)이 오디오 부분(246 및 250)에 포함되도록 선택할 수 있다.

[0038] 일부 구현예에서, 사용자 인터페이스 생성기(252)는 오디오 편집기(238)에 의해 오디오 데이터(236)의 편집을 표시할 수 있다. 예를 들어, 사용자 인터페이스 생성기(252)는 오디오 편집기(238)가 전송하고 있었던 오디오 데이터의 전사 및 오디오 편집기(238)가 제거하고 있는 오디오 데이터의 전사를 나타내는 인터페이스를 생성할 수 있다. 사용자 인터페이스는 또한 각각의 오디오 부분에 대해 식별된 화자들과 관련된 데이터를 포함할 수 있다.

[0039] 일부 구현예에서, 구분 모듈(218)은 핫워드(210) 이후 및 명령 식별기(216) 전에 오디오 데이터(212)를 프로세싱할 수 있다. 이 경우에, 핫워드(210)는 핫워드를 포함하는 오디오 데이터(212)의 부분을 식별한다. 구분 모듈(218)은 핫워드 화자로부터의 음성을 포함하는 오디오 데이터를 전송하고, 핫워드 화자와 다른 화자로부터의 음성을 포함하는 오디오 데이터의 전송을 억제한다. 구분 모듈(218)은 오디오 데이터를 명령 식별기(216)로 전송할 수 있고, 명령 식별기(216)는 데이터를 오디오 편집기(238)에 전송한다. 구분 모듈(218)은 오디오 서브시스템(202)이 검출된 오디오를 프로세싱 함에 따라 오디오 데이터(212)의 화자 변경을 실시간 또는 거의 실시간으로 식별할 수 있다.

- [0040] 시스템(200)의 컴포넌트는 클라이언트 디바이스, 예를 들어 컴퓨팅 디바이스(115) 또는 서버(클라우드라고도 함)에서 구현될 수 있다. 예를 들어, 명령 식별기(216)는 서버에서 구현될 수 있지만, 시스템(200)의 다른 컴포넌트는 클라이언트 디바이스에서 구현될 수 있다. 클라이언트 디바이스는 오디오를 서버에 전송하기 전에 핫워드를 말하지 않은 사용자로부터의 음성 데이터를 포함하는 오디오를 제거하도록 구성될 수 있다. 비핫워드 화자의 음성 데이터를 제거함으로써, 클라이언트 디바이스는 비핫워드 화자의 프라이버시를 보호한다. 비핫워드 화자는 방관자이거나 의도하지 않은 사용자일 수 있으며, 시스템과 인터랙션을 의도하지 않았을 가능성이 크다.
- [0041] 도 3는 화자 구분을 위한 예시적 프로세스(300)을 도시한다. 일반적으로, 프로세스(300)는 핫워드 및 워터마크를 포함하는 오디오에 기초하여 미디어 콘텐츠에 대응하는 오디오에 대해 음성 인식을 수행한다. 프로세스(300)는 하나 이상의 컴퓨터들 예를 들어, 도 1의 컴퓨팅 디바이스(115) 또는 도 2의 시스템(200)을 포함하는 컴퓨터 시스템에 의해 수행되는 것으로 설명될 것이다.
- [0042] 시스템은 발언에 대응하는 오디오 데이터를 수신한다(310). 발언은 하나 이상의 화자로부터의 음성을 포함할 수 있다. 예를 들어, 발언은 "Ok computer, call Mom. Hi Mom, I'm about to leave. Hi Mom. I'll be there in 30 minutes"일 수 있다. 시스템은 상기 오디오 데이터가 제1 화자에 의해 발화된 미리 정의된 핫워드의 발언을 포함한다고 결정한다(320). 시스템은 핫워드를 사용하여 오디오 데이터에서 핫워드를 식별할 수 있다. 예를 들어, 시스템은 오디오 데이터가 핫워드 "ok computer"를 포함할 가능성이 있다고 결정할 수 있다.
- [0043] 시스템은 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제1 부분을 식별한다(330). 일부 구현예에서, 시스템은 핫워드와 상기 핫워드에 뒤따르는 임의의 명령도 포함하지 않는 오디오 데이터를 분석한다. 예를 들어, 시스템은 "ok computer, call mom"에 대응하는 오디오 데이터 부분을 분석하지 않을 수 있다. 시스템은 명령이 뒤따르는 오디오 데이터를 분석할 수 있다. 시스템은 핫워드를 발화한 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 부분을 식별한다. 예를 들어, 시스템은 "Hi Mom, I'm about to leave"에 대응하는 오디오 데이터가 핫워드를 말한 제1 화자에 의해 발화된 것으로 결정할 수 있다.
- [0044] 시스템은 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 제2 부분을 식별한다(340). 시스템은 오디오 데이터를 분석하여 핫워드 화자가 아닌 화자로부터의 음성을 포함하는 오디오 데이터의 부분을 식별한다. 예를 들어, 시스템은 "Hi Mom"이 핫워드 화자가 아닌 화자에 의해 발화되었다고 결정할 수 있다.
- [0045] 일부 구현예에서, 시스템은 오디오 데이터의 제3 부분의 화자를 식별할 수 있다. 예를 들어, 시스템은 "I'll be there in 30 minutes"에 대응하는 오디오 데이터가 핫워드 화자에 의해 발화된 것으로 결정할 수 있다.
- [0046] 시스템은 상기 오디오 데이터가 상기 제1 화자에 의해 발화된 상기 미리 정의된 핫워드의 발언을 포함한다고 결정하는 것에 기초하여, 상기 제1 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제1 부분을 전송하고, 상기 상이한 제2 화자로부터의 음성을 포함하는 상기 오디오 데이터의 상기 제2 부분의 전송을 억제한다. 일부 구현예에서, 시스템은 예를 들어, 전화 통화에서 음성의 타이밍이 보존되어야 하는 경우에 오디오 데이터의 제2 부분을 대응하는 묵음 기간으로 교체할 수 있다. 일부 구현예에서, 시스템은 핫워드 화자로부터의 음성을 포함하는 오디오 데이터의 부분들을 함께 이어붙여 교체하고, 핫워드 화자가 아닌 화자로부터 음성을 포함하는 오디오 데이터를 제거할 수 있다. 시스템의 타이밍 유지 여부는 애플리케이션 설정 또는 사용자 설정에 기초할 수 있다.
- [0047] 핫워드 화자가 아닌 화자로부터의 음성을 포함하는 오디오 데이터를 전송하지 않음으로써, 시스템은 시스템의 마이크로폰이 검출할 수 있는 다른 화자의 프라이버시를 보호한다. 시스템은 핫워드 발언을 사용하여 핫워드 화자가 시스템과 인터랙션하려고 한다는 것을 표시한다. 시스템은 핫워드 화자가 아닌 화자에 의해 발화된 오디오 데이터 또는 오디오 데이터의 전사를 저장하지 않을 수 있다. 시스템이 오디오 데이터의 전사를 생성하는 경우에, 시스템은 핫워드 화자가 아닌 화자에 의해 발화된 전사의 부분을 삭제할 수 있다.
- [0048] 도 4은 본 명세서에 기술된 기법들을 구현하는데 사용될 수 있는 컴퓨터 디바이스(400)와 모바일 컴퓨터 디바이스(450)의 예시를 도시한다. 컴퓨팅 디바이스(400)는 랩톱, 데스크톱, 워크 스테이션, 개인 휴대 정보 단말기, 서버, 블레이드 서버, 메인 프레임 및 다른 적절한 컴퓨터와 같은 다양한 형태의 디지털 컴퓨터들을 나타내기 위한 것이다. 모바일 컴퓨팅 디바이스(450)는 개인 휴대 정보 단말기, 셀룰러 전화기, 스마트폰 및 다른 유사한 컴퓨팅 디바이스들과 같은 다양한 형태의 모바일 디바이스들을 나타내기 위한 것이다. 여기에 도시된 컴포넌트들, 그들의 연결 및 관계, 및 그들의 기능은 단지 예시적인 것을 의미하며, 제한하는 것을 의미하지는 않는다.
- [0049] 컴퓨팅 디바이스(400)는 프로세서(402), 메모리(404), 저장 디바이스(406), 메모리(404) 및 다수의 고속 확장 포트(410)에 연결되는 고속 인터페이스(408) 및 저속 확장 포트(414) 및 저장 디바이스(406)에 연결되는 저속

인터페이스(412)를 포함한다. 프로세서(402), 메모리(404), 저장 디바이스(406), 고속 인터페이스(408) 및 저속 인터페이스(412) 각각은 다양한 버스들을 사용하여 상호연결되며, 공통 마더보드에 또는 적절한 다른 방식으로 장착된다. 프로세서(402)는 메모리(404) 또는 저장 디바이스(406)에 저장된 명령어들을 포함하는, 컴퓨팅 디바이스(400) 내에서 실행하기 위한 명령어들을 프로세싱하여, 고속 인터페이스(408)에 연결된 디스플레이(416)와 같은 외부 입/출력 디바이스상에 GUI에 대한 그래픽 정보를 디스플레이할 수 있다. 다른 구현예에서, 다수의 프로세서들 및/또는 다수의 버스들이 다수의 메모리들 및 다수의 유형의 메모리와 함께, 적절하게 사용될 수 있다. 또한, 다수의 컴퓨팅 디바이스들은 필요한 동작의 부분들을 제공하는 각 디바이스와 연결될 수 있다(예를 들어, 서버 뱅크, 블레이드 서버 그룹 또는 멀티 프로세서 시스템).

[0050] 메모리(404)는 컴퓨팅 디바이스(400) 내에 정보를 저장한다. 일부 구현예에서, 메모리(404)는 휘발성 메모리 유닛 또는 유닛들이다. 일부 구현예에서, 메모리(404)는 비휘발성 메모리 유닛 또는 유닛들이다. 또한, 메모리(404)는 자기 또는 광학 디스크와 같은 컴퓨터 판독가능 매체의 다른 형태 일 수 있다.

[0051] 저장 디바이스(406)는 컴퓨팅 디바이스(400)에 대한 대형 스토리지를 제공할 수 있다. 일부 구현예에서, 저장 디바이스(406)는 플로피 디스크 디바이스, 하드 디스크 디바이스, 광 디스크 디바이스 또는 테이프 디바이스, 플래시 메모리 또는 다른 유사한 고체 상태 메모리 디바이스, 또는 저장 영역 네트워크 또는 다른 구성의 디바이스를 포함하는 디바이스의 어레이와 같은 컴퓨터 판독가능 매체이거나 컴퓨터 판독가능 매체를 포함할 수 있다. 명령어들은은 정보 캐리어에 저장될 수 있다. 명령어들은 하나 이상의 프로세싱 디바이스들(예를 들어, 프로세서(402))에 의해 실행될 때, 상술한 것과 같은 하나 이상의 방법들을 수행한다. 명령어들은 또한 컴퓨터 또는 기계 판독가능 매체(예를 들어, 메모리(404), 저장 장치(406) 또는 프로세서(402)상의 메모리)와 같은 하나 이상의 저장 디바이스들에 의해 저장될 수 있다.

[0052] 고속 인터페이스(408)는 컴퓨팅 디바이스(400)에 대한 대역폭 집중 동작들을 관리하는 반면, 저속 인터페이스(412)는 낮은 대역폭 집중 동작들을 관리한다. 이러한 기능들의 할당은 단지 예시적인 것이다. 일부 구현예에서, 고속 인터페이스(408)는 메모리(404), 디스플레이(416)(예를 들어, 그래픽 프로세서 또는 가속기를 통해) 및 다양한 확장 카드(도시되지 않음)를 수용할 수 있는 고속 확장 포트(410)에 연결된다. 구현예에서, 저속 인터페이스(412)는 저장 디바이스(406) 및 저속 확장 포트(414)에 연결된다. 다양한 통신 포트(예를 들어, USB, 블루투스, 이더넷, 무선 이더넷)를 포함할 수 있는 저속 확장 포트(414)는 키보드, 포인팅 디바이스, 스캐너와 같은 하나 이상의 입력/출력 디바이스 또는 예를 들어 네트워크 어댑터를 통해 스위치 또는 라우터와 같은 네트워킹 디바이스에 결합될 수 있다.

[0053] 컴퓨팅 디바이스(400)는 도면에 도시된 바와 같이 다수의 상이한 형태로 구현될 수 있다. 예를 들어, 그것은 표준 서버(420)로서 또는 그러한 서버들의 그룹에서 다수로 구현될 수 있다. 또한, 랩톱 컴퓨터(422)와 같은 퍼스널 컴퓨터에서 구현될 수 있다. 또한, 랙 서버 시스템(424)의 일부로서 구현될 수 있다. 대안적으로, 컴퓨팅 디바이스(400)로부터의 컴포넌트들은 모바일 컴퓨팅 디바이스(450)와 같은 모바일 디바이스(도시되지 않음) 내의 다른 컴포넌트들과 결합될 수 있다. 상기 디바이스들 각각은 컴퓨팅 디바이스(400), 모바일 컴퓨팅 디바이스(450) 중 하나 이상을 포함할 수 있고, 전체 시스템은 서로 통신하는 다수의 컴퓨팅 디바이스들로 구성될 수 있다.

[0054] 모바일 컴퓨팅 디바이스(450)는 다른 여러 컴포넌트들 중에서도 특히, 프로세서(452), 메모리(464), 디스플레이(454)와 같은 입/출력 디바이스, 통신 인터페이스(466) 및 송수신기(468)를 포함한다. 모바일 디바이스(450)에는 또한 추가적 저장을 제공하기 위해 마이크로 드라이브 또는 다른 디바이스와 같은 저장 디바이스가 제공될 수 있다. 프로세서(452), 메모리(464), 디스플레이(454), 통신 인터페이스(466) 및 트랜스미터(468) 각각은 다양한 버스들을 사용하여 상호 연결되고, 몇몇 컴포넌트들은 공통 마더 보드 상에 또는 적절한 다른 방식으로 장착될 수 있다.

[0055] 프로세서(452)는 메모리(464)에 저장된 명령어들을 포함하는 모바일 컴퓨팅 디바이스(450) 내의 명령어들을 실행할 수 있다. 프로세서(452)는 별개의 그리고 다수의 아날로그 및 디지털 프로세서들을 포함하는 칩들의 칩셋으로서 구현될 수 있다. 프로세서(452)는 예를 들어 사용자 인터페이스들, 모바일 컴퓨팅 디바이스(450)에 의해 실행되는 어플리케이션 및 모바일 컴퓨팅 디바이스(450)에 의한 무선 통신과 같은 모바일 컴퓨팅 디바이스(450)의 다른 컴포넌트들의 조정을 제공할 수 있다.

[0056] 프로세서(452)는 제어 인터페이스(458) 및 디스플레이(456)에 연결된 디스플레이 인터페이스(456)를 통해 사용자와 통신할 수 있다. 디스플레이(454)는 예를 들어, TFT LCD(박막 트랜지스터 액정 디스플레이) 또는 OLED(유기 발광 다이오드) 디스플레이 또는 다른 적절한 디스플레이 기술을 포함할 수 있다. 디스플레이 인터페이스

(456)는 사용자에게 그래픽 및 다른 정보를 제공하기 위해 디스플레이(454)를 구동하기 위한 적절한 회로를 포함할 수 있다. 제어 인터페이스(458)는 사용자로부터 명령을 수신하고, 프로세서(452)에 제출하기 위해 그들을 변환할 수 있다. 추가로, 외부 인터페이스(462)는 프로세서(452)와의 통신에 제공되어 다른 디바이스들과 모바일 컴퓨팅 디바이스(450)의 근거리 통신을 가능하게할 수 있다. 외부 인터페이스(462)는 예를 들면, 일부 구현예들에서는 유선 통신을 위해 또는 다른 구현예들에서는 무선 통신을 위해 제공될 수 있고, 다수의 인터페이스들도 사용될 수 있다.

[0057] 메모리(464)는 모바일 컴퓨팅 디바이스(450) 내에 정보를 저장한다. 메모리(464)는 컴퓨터 판독가능 매체 또는 매체들, 휘발성 메모리 유닛 또는 유닛들, 비휘발성 메모리 유닛 또는 유닛들 중 하나 이상으로 구현될 수 있다. 또한 확장 메모리(474)는 예를 들어 SIMM(Single In Line Memory Module) 카드 인터페이스를 포함할 수 있는 확장 인터페이스(472)를 통해 모바일 컴퓨팅 디바이스(450)에 제공되고 접속될 수 있다. 상기 확장 메모리(474)는 모바일 컴퓨팅 디바이스(450)에 대해 추가 저장 공간을 제공할 수 있거나, 또는 어플리케이션들 또는 모바일 컴퓨팅 디바이스(450)에 대한 다른 정보를 저장할 수 있다. 특히, 확장 메모리(474)는 전술한 프로세스들을 수행하거나 보충하는 명령어들을 포함할 수 있으며, 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 보안 사용을 허용하는 명령어들로 프로그래밍될 수 있다. 또한, 보안 어플리케이션들은 SIMM 카드 상의 식별 정보를 해킹할 수 없는 방식으로 배치하는 것과 같이, SIMM 카드들을 통해 추가 정보와 함께 제공될 수 있다.

[0058] 메모리는 예를 들어, 후술되는 바와 같이, 플래시 메모리 및/또는 NVRAM 메모리(non-volatile random access memory)를 포함할 수 있다. 일부 구현예들에서, 명령어들은 정보 캐리어에 저장되어, 하나 이상의 프로세싱 디바이스들(예를 들어, 프로세서(452))에 의해 실행될 때, 상술한 것과 같은 하나 이상의 방법들을 수행한다. 명령어들은 또한 하나 이상의 컴퓨터 또는 기계 판독가능 매체(예를 들어, 메모리(464), 확장 메모리(474) 또는 프로세서(452)상의 메모리)와 같은 하나 이상의 저장 디바이스들에 의해 저장될 수 있다. 일부 구현예에서, 명령어들은 예를 들어 트랜스시버(468) 또는 외부 인터페이스(462)를 통해 전파된 신호로 수신될 수 있다.

[0059] 모바일 컴퓨팅 디바이스(450)는 필요에 따라 디지털 신호 처리 회로를 포함할 수 있는 통신 인터페이스(466)를 통해 무선으로 통신할 수 있다. 통신 인터페이스(466)는 GSM 음성 호출(이동 통신용 글로벌 시스템), SMS(단문 메시지 서비스), EMS(확장 메시징 서비스) 또는 MMS 메시징(멀티미디어 메시징 서비스), CDMA(코드 분할 다중 접속), TDMA(시분할 다중 접속), PDC(개인 디지털 셀룰러), WCDMA(광대역 코드 분할 다중 접속), CDMA2000 또는 GPRS(일반 패킷 무선 서비스)와 같은 다양한 모드 또는 프로토콜 하에서 통신을 제공할 수 있다. 이러한 통신은 예를 들어, 무선 주파수를 사용하는 트랜스시버(468)를 통해 발생될 수 있다. 추가로, 블루투스, Wi-Fi 또는 다른 트랜시버(도시되지 않음)를 사용하는 것과 같은 단거리 통신이 발생될 수 있다. 추가로, GPS(Global Positioning System) 수신기 모듈(470)은 모바일 컴퓨팅 디바이스(450)상에서 실행되는 어플리케이션들에 의해 적절히 사용될 수 있는 추가적인 네비게이션 및 위치 관련 무선 데이터를 모바일 컴퓨팅 디바이스(450)에 제공할 수 있다.

[0060] 모바일 컴퓨팅 디바이스(450)는 또한 사용자로부터 발화된 정보를 수신하고 그것을 이용가능한 디지털 정보로 변환할 수 있는 오디오 코덱(460)을 사용하여 청각적으로 통신할 수 있다. 오디오 코덱(460)은 마찬가지로, 예를 들어 모바일 컴퓨팅 디바이스(450)의 핸드셋 내의 스피커를 통하는 것과 같이, 사용자를 위한 가청 사운드를 생성할 수 있다. 이러한 사운드는 보이스 전화 콜들로부터의 사운드 포함할 수 있고, 기록된 사운드(예를 들어, 음성 메시지, 음악 파일 등)를 포함할 수 있고, 또한 모바일 컴퓨팅 디바이스(450)상에서 동작하는 어플리케이션들에 의해 생성된 사운드를 포함할 수 있다.

[0061] 모바일 컴퓨팅 디바이스(450)는 도면에 도시된 바와 같이 다수의 상이한 형태로 구현될 수 있다. 예를 들면, 그것은 셀룰러 전화(480)로서 구현될 수 있다. 또한 스마트폰(482), 개인 휴대 정보 단말기(PDA), 또는 다른 유사한 모바일 디바이스의 일부로서 구현될 수 있다.

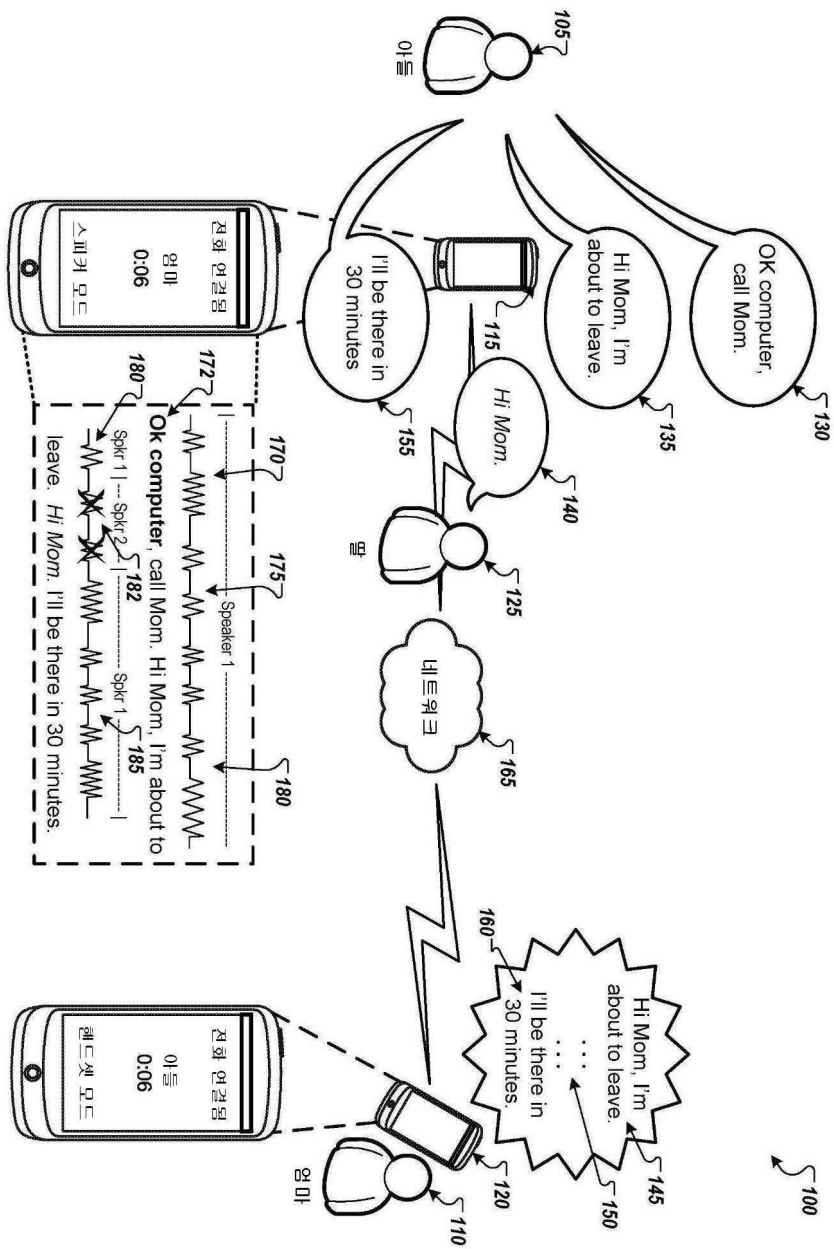
[0062] 본 명세서에 기술된 시스템들 및 기법들의 다양한 구현예들은 디지털 전자 회로, 집적 회로, 특수하게 설계된 ASIC들(application specific integrated circuits), 컴퓨터 하드웨어, 펌웨어, 소프트웨어 및/또는 이들의 조합으로 구현될 수 있다. 이들 다양한 구현예들은 적어도 하나의 프로그래머블 프로세서를 포함하는 프로그래머블 시스템 상에서 실행가능하고 및/또는 인터프리트가능한 하나 이상의 컴퓨터 프로그램들에서의 구현예를 포함할 수 있고, 이는 전용 또는 범용일 수 있고, 저장 시스템, 적어도 하나의 입력 디바이스 및 적어도 하나의 출력 디바이스으로부터 데이터 및 명령어들을 수신하고 그에 데이터 및 명령어들을 전송하기 위해 연결될 수

있다.

- [0063] 이들 컴퓨터 프로그램들(프로그램, 소프트웨어, 소프트웨어 어플리케이션 또는 코드로도 알려짐)은 프로그래머블 프로세서에 대한 기계 명령어들을 포함하며, 하이레벨 절차어 및/또는 객체 지향 프로그래밍 언어 및/또는 어셈블리/기계어에서 구현될 수 있다. 본 명세서에서 사용된 바와 같이, 용어 "기계 판독가능 매체", "컴퓨터 판독가능 매체"는 기계 판독가능 신호로서 기계 명령어들을 수신하는 기계 판독가능 매체를 포함하여, 기계 명령어 및/또는 데이터를 프로그래머블 프로세서에 제공하는데 사용되는 임의의 컴퓨터 프로그램 물, 장치 및/또는 디바이스 예를 들어, 자기 디스크, 광학 디스크, 메모리, 프로그래머블 로직 디바이스(PLD)를 지칭한다. 용어 "기계 판독가능 신호"는 기계 명령어 및/또는 데이터를 프로그래머블 프로세서에 제공하는데 사용되는 임의의 신호를 지칭한다.
- [0064] 사용자와의 인터랙션을 제공하기 위해, 본 명세서에서 기술된 시스템들 및 기법들은 사용자에게 정보를 디스플레이하기 위해 예를 들어, CRT(cathode ray tube) 또는 LCD(liquid crystal display) 모니터와 같은 디스플레이 디바이스 및 사용자가 컴퓨터에 입력을 제공할 수 있는 키보드 및 포인팅 디바이스 예를 들어, 마우스 또는 트랙볼을 갖는 컴퓨터에서 구현될 수 있다. 다른 종류의 디바이스들도 사용자와의 인터랙션을 제공하는데 사용될 수 있다. 예를 들어, 사용자에게 제공되는 피드백은 시각 피드백, 청각 피드백 또는 촉각 피드백과 같은 임의의 형태의 감각적 피드백일 수 있고, 사용자로부터의 입력은 음향, 음성 또는 촉각 입력을 포함하는 임의의 형태로 수신될 수 있다.
- [0065] 본 명세서에서 기술된 시스템들 및 기법들은 예를 들어 데이터 서버와 같은 백엔드 컴포넌트, 어플리케이션 서버와 같은 미들웨어 컴포넌트 또는 그래픽 사용자 인터페이스를 가지는 사용자 컴퓨터 또는 사용자가 본 명세서에 기술된 시스템들 및 기법들의 구현예와 인터랙션할 수 있는 웹 브라우저와 같은 프론트엔드 컴포넌트 또는 하나 이상의 상기 백엔드, 미들웨어 또는 프론트엔드 컴포넌트들의 임의의 조합을 포함하는 컴퓨팅 시스템에서 구현될 수 있다. 시스템의 컴포넌트들은 디지털 데이터 통신의 임의의 형태 또는 매체, 예를 들어 통신 네트워크에 의해 상호연결될 수 있다. 통신 네트워크들의 예시들은 LAN(local area network), WAN(wide area network) 및 인터넷을 포함한다.
- [0066] 컴퓨팅 시스템은 사용자들 및 서버들을 포함할 수 있다. 사용자와 서버는 일반적으로 서로 멀리 떨어져 있으며, 일반적으로 통신 네트워크를 통해 인터랙션한다. 사용자와 서버의 관계는 각각의 컴퓨터에서 실행되고 서로 사용자-서버 관계를 갖는 컴퓨터 프로그램에 의해 발생한다.
- [0067] 비록 일부 구현예들이 상술되었지만, 다른 수정들도 가능하다. 예를 들어, 클라이언트 어플리케이션은 텔리게이트(들)에 액세스하는 것으로 기술되지만, 다른 구현예에서는 텔리게이트(들)는 하나 이상의 서버에서 실행되는 어플리케이션과 같은 하나 이상의 프로세서들에 의해 구현되는 다른 어플리케이션에 의해 이용될 수 있다. 추가로, 도면들에 도시된 논리 흐름들은 원하는 결과들을 달성하기 위해 특정한 도시된 순서, 또는 시계열적 순서를 반드시 필요로 하지 않는다. 추가로, 다른 단계들이 제공될 수 있거나, 단계들이 기술된 흐름으로부터 생략될 수 있고, 다른 컴포넌트들이 기술된 시스템에 추가되거나 그로부터 제거될 수 있다. 따라서, 다른 구현예들도 다음의 청구항들의 범위 내에 있다.

도면

도면1



도면3

300 ↗

