

(71) 출원인
켈컴 인코포레이티드
미국 92121-1714 캘리포니아주 샌 디에고 모어하
우스 드라이브 5775

(72) 발명자
박 수진
미국 92121 캘리포니아주 샌디에고 모어하우스 드
라이브 5775

문 선국
미국 92121 캘리포니아주 샌디에고 모어하우스 드
라이브 5775
(뒷면에 계속)

(74) 대리인
특허법인코리아나

전체 청구항 수 : 총 30 항

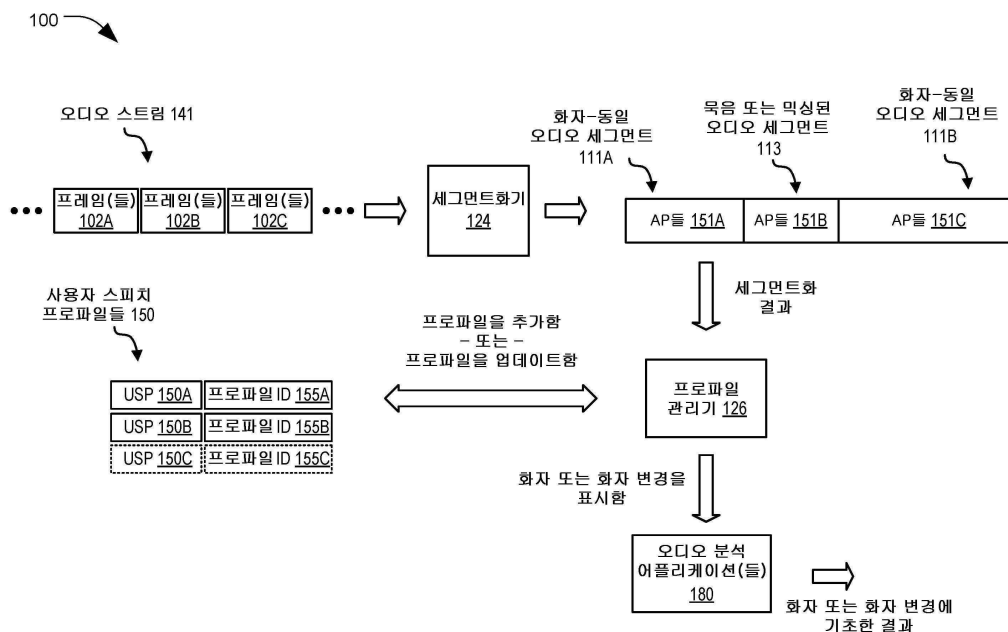
(54) 발명의 명칭 사용자 스피치 프로파일 관리

(57) 요약

디바이스는, 제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 화자들의 스피치에 대응하는지 여부를 결정하도록 구성된 프로세서들을 포함한다. 프로세서들은, 오디오 스트림이 적어도 2명의 화자들의 스피치에 대응함을 결정하는 것에 기초하여, 제 2 전력 모드에서, 세그먼트화 결과를 생성하기 위해 오디오 스트림의 오디오 피

(뒷면에 계속)

대표도



처 데이터를 분석하도록 구성된다. 프로세서들은, 화자-동일 오디오 세그먼트의 복수의 오디오 피쳐 데이터 세트들 중의 오디오 피쳐 데이터 세트에 대한 복수의 사용자 스피치 프로파일들의 비교를, 오디오 피쳐 데이터 세트가 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하도록 구성된다. 프로세서들은, 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 것과 매칭하지 않음을 결정하는 것에 기초하여, 복수의 오디오 피쳐 데이터 세트들에 기초하여 사용자 스피치 프로파일을 생성하도록 구성된다.

(52) CPC특허분류

G10L 15/04 (2013.01)

G10L 15/16 (2013.01)

G10L 17/04 (2013.01)

G10L 17/18 (2013.01)

(72) 발명자

김 래훈

미국 92121 캘리포니아주 샌디에고 모어하우스 드
라이브 5775

비제르 에릭

미국 92121 캘리포니아주 샌디에고 모어하우스 드
라이브 5775

명세서

청구범위

청구항 1

오디오 분석을 위한 디바이스로서,

복수의 사용자들의 복수의 사용자 스피치 프로파일들을 저장하도록 구성된 메모리; 및

하나 이상의 프로세서들을 포함하고,

상기 하나 이상의 프로세서들은,

제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하고;

상기 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응함을 결정하는 것에 기초하여, 제 2 전력 모드에서, 세그먼트화 결과를 생성하기 위해 상기 오디오 스트림의 오디오 피쳐 데이터를 분석하는 것으로서, 상기 세그먼트화 결과는 상기 오디오 스트림의 화자-동일 오디오 세그먼트들을 표시하는, 상기 오디오 스트림의 오디오 피쳐 데이터를 분석하고;

제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들 중의 제 1 오디오 피쳐 데이터 세트에 대한 상기 복수의 사용자 스피치 프로파일들의 비교를, 상기 제 1 오디오 피쳐 데이터 세트가 상기 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하고; 그리고

상기 제 1 오디오 피쳐 데이터 세트가 상기 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여:

상기 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하고;

상기 제 1 사용자 스피치 프로파일을 상기 복수의 사용자 스피치 프로파일들에 추가하도록

구성되는, 오디오 분석을 위한 디바이스.

청구항 2

제 1 항에 있어서,

상기 제 1 오디오 피쳐 데이터 세트는 제 1 오디오 피쳐 벡터를 포함하는, 오디오 분석을 위한 디바이스.

청구항 3

제 1 항에 있어서,

상기 하나 이상의 프로세서들은 스피커 세그먼트화 뉴럴 네트워크를 상기 오디오 피쳐 데이터에 적용함으로써 상기 오디오 피쳐 데이터를 분석하도록 구성되는, 오디오 분석을 위한 디바이스.

청구항 4

제 1 항에 있어서,

상기 하나 이상의 프로세서들은, 상기 세그먼트화 결과가 상기 제 1 오디오 피쳐 데이터 세트가 제 1 화자의 스피치에 대응한다는 것을 표시함을 그리고 상기 제 1 오디오 피쳐 데이터 세트가 상기 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여:

상기 제 1 화자와 연관된 제 1 등록 버퍼에 상기 제 1 오디오 피쳐 데이터 세트를 저장하고; 그리고

정지 조건이 만족될 때까지 상기 제 1 등록 버퍼에 상기 제 1 화자의 스피치에 대응하는 후속 오디오 피쳐 데이터 세트들을 저장하도록

구성되고, 상기 제 1 화자-동일 오디오 세그먼트의 상기 제 1 복수의 오디오 피쳐 데이터 세트들은 상기 제 1

오디오 피쳐 데이터 세트 및 상기 후속 오디오 피쳐 데이터 세트들을 포함하는, 오디오 분석을 위한 디바이스.

청구항 5

제 4 항에 있어서,

상기 하나 이상의 프로세서들은, 임계 묶음보다 긴 것이 상기 오디오 스트림에서 검출됨을 결정하는 것에 응답하여 상기 정지 조건이 만족됨을 결정하도록 구성되는, 오디오 분석을 위한 디바이스.

청구항 6

제 4 항에 있어서,

상기 하나 이상의 프로세서들은, 특정 오디오 피쳐 데이터 세트가 단일 화자의 스피치에 대응함을 결정하는 것에 적어도 부분적으로 기초하여 상기 제 1 등록 버퍼에 상기 특정 오디오 피쳐 데이터 세트를 추가하도록 구성되고, 상기 단일 화자는 상기 제 1 화자를 포함하는, 오디오 분석을 위한 디바이스.

청구항 7

제 1 항에 있어서,

상기 하나 이상의 프로세서들은, 제 1 등록 버퍼에 저장된 상기 제 1 화자-동일 오디오 세그먼트의 상기 제 1 복수의 오디오 피쳐 데이터 세트들의 카운트가 등록 임계치보다 크다고 결정하는 것에 기초하여, 상기 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 상기 제 1 사용자 스피치 프로파일을 생성하도록 구성되는, 오디오 분석을 위한 디바이스.

청구항 8

제 1 항에 있어서,

상기 하나 이상의 프로세서들은, 상기 제 1 오디오 피쳐 데이터 세트가 특정 사용자 스피치 프로파일과 매칭함을 결정하는 것에 기초하여, 상기 제 1 오디오 피쳐 데이터 세트에 기초하여 상기 특정 사용자 스피치 프로파일을 업데이트하도록 구성되는, 오디오 분석을 위한 디바이스.

청구항 9

제 8 항에 있어서,

상기 하나 이상의 프로세서들은, 상기 제 1 오디오 피쳐 데이터 세트가 단일 화자의 스피치에 대응함을 결정하는 것에 적어도 부분적으로 기초하여, 상기 제 1 오디오 피쳐 데이터 세트에 기초하여 상기 특정 사용자 스피치 프로파일을 업데이트하도록 구성되는, 오디오 분석을 위한 디바이스.

청구항 10

제 1 항에 있어서,

상기 하나 이상의 프로세서들은, 제 2 화자-동일 오디오 세그먼트의 제 2 복수의 오디오 피쳐 데이터 세트들 중의 제 2 오디오 피쳐 데이터 세트가 상기 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하도록 구성되는, 오디오 분석을 위한 디바이스.

청구항 11

제 10 항에 있어서,

상기 하나 이상의 프로세서들은, 상기 제 2 오디오 피쳐 데이터 세트가 상기 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여:

상기 제 2 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 2 사용자 스피치 프로파일을 생성하고; 그리고

상기 제 2 사용자 스피치 프로파일을 상기 복수의 사용자 스피치 프로파일들에 추가하도록

구성되는, 오디오 분석을 위한 디바이스.

청구항 12

제 10 항에 있어서,

상기 하나 이상의 프로세서들은, 상기 제 2 오디오 피쳐 데이터 세트가 상기 복수의 사용자 스피치 프로파일들 중의 특정 사용자 스피치 프로파일과 매칭함을 결정하는 것에 기초하여, 상기 제 2 오디오 피쳐 데이터 세트에 기초하여 상기 특정 사용자 스피치 프로파일을 업데이트하도록 구성되는, 오디오 분석을 위한 디바이스.

청구항 13

제 1 항에 있어서,

상기 메모리는 프로파일 업데이트 데이터를 저장하도록 구성되고,

상기 하나 이상의 프로세서들은,

상기 제 1 사용자 스피치 프로파일을 생성하는 것에 응답하여, 상기 제 1 사용자 스피치 프로파일이 업데이트됨을 표시하기 위해 상기 프로파일 업데이트 데이터를 업데이트하고; 그리고

상기 복수의 사용자 스피치 프로파일들의 제 1 카운트가 업데이트되었음을 상기 프로파일 업데이트 데이터가 표시함을 결정하는 것에 기초하여, 상기 제 1 카운트를, 상기 오디오 스트림에서 검출된 화자들의 카운트로서 출력하도록

구성되는, 오디오 분석을 위한 디바이스.

청구항 14

제 1 항에 있어서,

상기 메모리는 사용자 상호작용 데이터를 저장하도록 구성되고,

상기 하나 이상의 프로세서들은,

상기 제 1 사용자 스피치 프로파일을 생성하는 것에 응답하여, 상기 제 1 사용자 스피치 프로파일과 연관된 제 1 사용자가 스피치 지속기간 동안 상호작용하였음을 표시하기 위해 상기 제 1 화자-동일 오디오 세그먼트의 상기 스피치 지속기간에 기초하여 상기 사용자 상호작용 데이터를 업데이트하고; 그리고

적어도 상기 사용자 상호작용 데이터를 출력하도록

구성되는, 오디오 분석을 위한 디바이스.

청구항 15

제 1 항에 있어서,

상기 제 1 전력 모드는 상기 제 2 전력 모드에 비해 더 낮은 전력 모드인, 오디오 분석을 위한 디바이스.

청구항 16

제 1 항에 있어서,

상기 하나 이상의 프로세서들은,

상기 제 1 전력 모드에서, 상기 오디오 스트림의 오디오 정보를 결정하는 것으로서, 상기 오디오 정보는 상기 오디오 스트림에서 검출된 화자들의 카운트, 음성 활성화도 검출 (VAD) 정보, 또는 그 양자 모두를 포함하는, 상기 오디오 스트림의 오디오 정보를 결정하고;

상기 제 2 전력 모드에서 하나 이상의 오디오 분석 어플리케이션들을 활성화하고; 그리고

상기 오디오 정보를 하나 이상의 오디오 분석 어플리케이션들에 제공하도록

구성되는, 오디오 분석을 위한 디바이스.

청구항 17

제 1 항에 있어서,

상기 하나 이상의 프로세서들은, 상기 오디오 스트림의 하나 이상의 제 2 오디오 세그먼트들이 다중 화자들에 대응한다는 것을 상기 세그먼트화 결과가 표시함을 결정하는 것에 응답하여, 상기 하나 이상의 제 2 오디오 세그먼트들에 기초하여 상기 복수의 사용자 스피치 프로파일들을 업데이트하는 것을 억제하도록 구성되는, 오디오 분석을 위한 디바이스.

청구항 18

오디오 분석의 방법으로서,

디바이스에서의 제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하는 단계;

상기 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응함을 결정하는 것에 기초하여, 제 2 전력 모드에서, 세그먼트화 결과를 생성하기 위해 상기 오디오 스트림의 오디오 피쳐 데이터를 분석하는 단계로서, 상기 세그먼트화 결과는 상기 오디오 스트림의 화자-동일 오디오 세그먼트들을 표시하는, 상기 오디오 스트림의 오디오 피쳐 데이터를 분석하는 단계;

상기 디바이스에서, 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들 중의 제 1 오디오 피쳐 데이터 세트에 대한 복수의 사용자 스피치 프로파일들의 비교를, 상기 제 1 오디오 피쳐 데이터 세트가 상기 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하는 단계; 및

상기 제 1 오디오 피쳐 데이터 세트가 상기 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여:

상기 디바이스에서, 상기 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하는 단계; 및

상기 디바이스에서, 상기 제 1 사용자 스피치 프로파일을 상기 복수의 사용자 스피치 프로파일들에 추가하는 단계를 포함하는, 오디오 분석의 방법.

청구항 19

제 18 항에 있어서,

스피커 세그먼트화 뉴럴 네트워크를 상기 오디오 피쳐 데이터에 적용하는 단계를 더 포함하는, 오디오 분석의 방법.

청구항 20

제 18 항에 있어서,

상기 세그먼트화 결과가 상기 제 1 오디오 피쳐 데이터 세트가 제 1 화자의 스피치에 대응한다는 것을 표시함을 그리고 상기 제 1 오디오 피쳐 데이터 세트가 상기 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여:

상기 제 1 화자와 연관된 제 1 등록 버퍼에 상기 제 1 오디오 피쳐 데이터 세트를 저장하는 단계; 및

정지 조건이 만족될 때까지 상기 제 1 등록 버퍼에 상기 제 1 화자의 스피치에 대응하는 후속 오디오 피쳐 데이터 세트들을 저장하는 단계를 더 포함하고,

상기 제 1 화자-동일 오디오 세그먼트의 상기 제 1 복수의 오디오 피쳐 데이터 세트들은 상기 제 1 오디오 피쳐 데이터 세트 및 상기 후속 오디오 피쳐 데이터 세트들을 포함하는, 오디오 분석의 방법.

청구항 21

제 20 항에 있어서,

상기 디바이스에서, 임계 폭음보다 긴 것이 상기 오디오 스트림에서 검출됨을 결정하는 것에 응답하여 상기 정

지 조건이 만족됨을 결정하는 단계를 더 포함하는, 오디오 분석의 방법.

청구항 22

제 20 항에 있어서,

상기 디바이스에서, 특정 오디오 피쳐 데이터 세트가 단일 화자의 스피치에 대응함을 결정하는 것에 적어도 부분적으로 기초하여 상기 제 1 등록 버퍼에 상기 특정 오디오 피쳐 데이터 세트를 추가하는 단계를 더 포함하고, 상기 단일 화자는 상기 제 1 화자를 포함하는, 오디오 분석의 방법.

청구항 23

제 18 항에 있어서,

제 1 등록 버퍼에 저장된 상기 제 1 화자-동일 오디오 세그먼트의 상기 제 1 복수의 오디오 피쳐 데이터 세트들의 카운트가 등록 임계치보다 크다고 결정하는 것에 기초하여, 상기 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 상기 제 1 사용자 스피치 프로파일을 생성하는 단계를 더 포함하는, 오디오 분석의 방법.

청구항 24

제 18 항에 있어서,

상기 제 1 오디오 피쳐 데이터 세트가 특정 사용자 스피치 프로파일과 매칭함을 결정하는 것에 기초하여, 상기 제 1 오디오 피쳐 데이터 세트에 기초하여 상기 특정 사용자 스피치 프로파일을 업데이트하는 단계를 더 포함하는, 오디오 분석의 방법.

청구항 25

제 24 항에 있어서,

상기 제 1 오디오 피쳐 데이터 세트가 단일 화자의 스피치에 대응함을 결정하는 것에 적어도 부분적으로 기초하여, 상기 제 1 오디오 피쳐 데이터 세트에 기초하여 상기 특정 사용자 스피치 프로파일을 업데이트하는 단계를 더 포함하는, 오디오 분석의 방법.

청구항 26

제 18 항에 있어서,

제 2 화자-동일 오디오 세그먼트의 제 2 복수의 오디오 피쳐 데이터 세트들 중의 제 2 오디오 피쳐 데이터 세트가 상기 복수의 사용자 스피치 프로파일들 중의 특정 사용자 스피치 프로파일과 매칭함을 결정하는 것에 기초하여, 상기 제 2 오디오 피쳐 데이터 세트에 기초하여 상기 특정 사용자 스피치 프로파일을 업데이트하는 단계를 더 포함하는, 오디오 분석의 방법.

청구항 27

명령들을 저장하는 비일시적 컴퓨터 판독가능 저장 매체로서,

상기 명령들은, 하나 이상의 프로세서들에 의해 실행될 경우, 상기 프로세서들로 하여금

제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하게 하고;

상기 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응함을 결정하는 것에 기초하여, 제 2 전력 모드에서, 세그먼트화 결과를 생성하기 위해 상기 오디오 스트림의 오디오 피쳐 데이터를 분석하게 하는 것으로서, 상기 세그먼트화 결과는 상기 오디오 스트림의 화자-동일 오디오 세그먼트들을 표시하는, 상기 오디오 스트림의 오디오 피쳐 데이터를 분석하게 하고;

제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들 중의 제 1 오디오 피쳐 데이터 세트에 대한 복수의 사용자 스피치 프로파일들의 비교를, 상기 제 1 오디오 피쳐 데이터 세트가 상기 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하게 하고; 그리고

상기 제 1 오디오 피쳐 데이터 세트가 상기 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여:

상기 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하게 하고;

상기 제 1 사용자 스피치 프로파일을 상기 복수의 사용자 스피치 프로파일들에 추가하게 하는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 28

제 27 항에 있어서,

상기 명령들은, 상기 하나 이상의 프로세서들에 의해 실행될 경우, 상기 프로세서들로 하여금, 제 1 등록 버퍼에 저장된 상기 제 1 화자-동일 오디오 세그먼트의 상기 제 1 복수의 오디오 피쳐 데이터 세트들의 카운트가 등록 임계치보다 크다고 결정하는 것에 기초하여, 상기 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 상기 제 1 사용자 스피치 프로파일을 생성하게 하는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 29

장치로서,

복수의 사용자들의 복수의 사용자 스피치 프로파일들을 저장하는 수단;

제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하는 수단;

제 2 전력 모드에서, 세그먼트화 결과를 생성하기 위해 상기 오디오 스트림의 오디오 피쳐 데이터를 분석하는 수단으로서, 상기 오디오 피쳐 데이터는, 상기 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응함을 결정하는 것에 기초하여 상기 제 2 전력 모드에서 분석되고, 상기 세그먼트화 결과는 상기 오디오 스트림의 화자-동일 오디오 세그먼트들을 표시하는, 상기 오디오 스트림의 오디오 피쳐 데이터를 분석하는 수단;

제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들 중의 제 1 오디오 피쳐 데이터 세트에 대한 상기 복수의 사용자 스피치 프로파일들의 비교를, 상기 제 1 오디오 피쳐 데이터 세트가 상기 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하는 수단;

상기 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하는 수단으로서, 상기 제 1 사용자 스피치 프로파일은, 상기 제 1 오디오 피쳐 데이터 세트가 상기 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여 생성되는, 상기 제 1 사용자 스피치 프로파일을 생성하는 수단; 및

상기 제 1 사용자 스피치 프로파일을 상기 복수의 사용자 스피치 프로파일들에 추가하는 수단을 포함하는, 장치.

청구항 30

제 29 항에 있어서,

상기 저장하는 수단, 상기 결정하는 수단, 상기 분석하는 수단, 상기 수행하는 수단, 상기 생성하는 수단, 및 상기 추가하는 수단은 모바일 통신 디바이스, 스마트 폰, 셀룰러 폰, 스마트 스피커, 스피커 바, 랩탑 컴퓨터, 컴퓨터, 태블릿, 개인용 디지털 보조기, 디스플레이 디바이스, 텔레비전, 게이밍 콘솔, 뮤직 플레이어, 라디오, 디지털 비디오 플레이어, 디지털 비디오 디스크 (DVD) 플레이어, 튜너, 카메라, 내비게이션 디바이스, 비히클, 헤드셋, 증강 현실 헤드셋, 가상 현실 헤드셋, 항공 비히클, 홈 자동화 시스템, 음성 활성화 디바이스, 무선 스피커 및 음성 활성화 디바이스, 휴대용 전자 디바이스, 자동차, 컴퓨팅 디바이스, 통신 디바이스, 사물 인터넷 (IoT) 디바이스, 가상 현실 (VR) 디바이스, 기지국, 모바일 디바이스, 또는 이들의 임의의 조합 중 적어도 하나에 통합되는, 장치.

발명의 설명

기술 분야

관련 출원들에 대한 상호참조

본 출원은 2020년 12월 8일자로 출원된 공동 소유의 미국 정규 특허출원 제17/115,158호로부터의 우선권의 이익을 주장하며, 그 내용들은 전부가 본 명세서에 참조에 의해 명백히 통합된다.

기술분야

본 개시는 일반적으로 사용자 스피치 프로파일들의 관리에 관한 것이다.

배경 기술

기술에서의 진보들은 더 소형이고 더 강력한 컴퓨팅 디바이스들을 발생시켰다. 예를 들어, 소형이고 경량이며 사용자들에 의해 용이하게 휴대되는 모바일 및 스마트 폰들과 같은 무선 전화기들, 태블릿들 및 랩탑 컴퓨터들을 포함한 다양한 휴대용 개인용 컴퓨팅 디바이스들이 현재 존재한다. 이들 디바이스들은 무선 네트워크들 상으로 음성 및 데이터 패킷들을 통신할 수 있다. 추가로, 다수의 그러한 디바이스들은 디지털 스틸 카메라, 디지털 비디오 카메라, 디지털 레코더, 및 오디오 파일 플레이어와 같은 추가 기능성을 통합한다. 또한, 그러한 디바이스들은, 인터넷에 액세스하는데 사용될 수 있는 웹 브라우저 어플리케이션과 같은 소프트웨어 어플리케이션들을 포함한 실행가능 명령들을 프로세싱할 수 있다. 그에 따라, 이들 디바이스들은 현저한 컴퓨팅 능력들을 포함할 수 있다.

그러한 컴퓨팅 디바이스들은 종종, 하나 이상의 마이크로폰들로부터 오디오 신호를 수신하기 위한 기능성을 통합한다. 예를 들어, 오디오 신호는 마이크로폰들에 의해 캡처된 사용자 스피치, 마이크로폰들에 의해 캡처된 외부 사운드들, 또는 이들의 조합을 나타낼 수도 있다. 그러한 디바이스들은, 예컨대, 사용자 인식을 위해 사용자 스피치 프로파일들에 의존하는 어플리케이션들을 포함할 수도 있다. 사용자 스피치 프로파일은 사용자로 하여금 미리결정된 단어들 또는 문장들의 스크립트를 말하게 함으로써 트레이닝될 수 있다. 사용자 스피치 프로파일을 생성하기 위한 그러한 활성 사용자 등록은 시간 소모적이고 불편할 수 있다.

발명의 내용

해결하려는 과제

과제의 해결 수단

본 개시의 일 구현에 따르면, 오디오 분석을 위한 디바이스는 메모리 및 하나 이상의 프로세서들을 포함한다. 메모리는 복수의 사용자들의 복수의 사용자 스피치 프로파일들을 저장하도록 구성된다. 하나 이상의 프로세서들은, 제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하도록 구성된다. 하나 이상의 프로세서들은 또한, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응함을 결정하는 것에 기초하여, 제 2 전력 모드에서, 세그먼트화 결과를 생성하기 위해 오디오 스트림의 오디오 피쳐 (feature) 데이터를 분석하도록 구성된다. 세그먼트화 결과는 오디오 스트림의 화자-동일 (talker-homogenous) 오디오 세그먼트들을 표시한다. 하나 이상의 프로세서들은 추가로, 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들 중의 제 1 오디오 피쳐 데이터 세트에 대한 복수의 사용자 스피치 프로파일들의 비교를, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하도록 구성된다. 하나 이상의 프로세서들은 또한, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여: 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하고 제 1 사용자 스피치 프로파일을 복수의 사용자 스피치 프로파일들에 추가하도록 구성된다.

본 개시의 다른 구현에 따르면, 오디오 분석 방법은, 디바이스에서의 제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하는 단계를 포함한다. 그 방법은 또한, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응함을 결정하는 것에 기초하여, 제 2 전력 모드에서, 세그먼트화 결과를 생성하기 위해 오디오 스트림의 오디오 피쳐 데이터를 분석하는 단계를 포함한다. 세그

먼트화 결과는 오디오 스트림의 화자-동일 오디오 세그먼트들을 표시한다. 그 방법은, 디바이스에서, 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들 중의 제 1 오디오 피쳐 데이터 세트에 대한 복수의 사용자 스피치 프로파일들의 비교를, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하는 단계를 더 포함한다. 그 방법은 또한, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여: 디바이스에서, 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하는 단계, 및 디바이스에서, 제 1 사용자 스피치 프로파일을 복수의 사용자 스피치 프로파일들에 추가하는 단계를 포함한다.

[0009] 본 개시의 다른 구현에 따르면, 비밀시적 컴퓨터 판독가능 매체는 명령들을 포함하고, 그 명령들은, 하나 이상의 프로세서들에 의해 실행될 경우, 하나 이상의 프로세서들로 하여금, 제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하게 한다. 그 명령들은, 하나 이상의 프로세서들에 의해 실행될 경우, 또한, 프로세서들로 하여금, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응함을 결정하는 것에 기초하여, 제 2 전력 모드에서, 세그먼트화 결과를 생성하기 위해 오디오 스트림의 오디오 피쳐 데이터를 분석하게 한다. 세그먼트화 결과는 오디오 스트림의 화자-동일 오디오 세그먼트들을 표시한다. 그 명령들은, 하나 이상의 프로세서들에 의해 실행될 경우, 추가로, 하나 이상의 프로세서들로 하여금 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들 중의 제 1 오디오 피쳐 데이터 세트에 대한 복수의 사용자 스피치 프로파일들의 비교를, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하게 한다. 그 명령들은, 하나 이상의 프로세서들에 의해 실행될 경우, 또한, 하나 이상의 프로세서들로 하여금, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여: 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하게 하고 제 1 사용자 스피치 프로파일을 복수의 사용자 스피치 프로파일들에 추가하게 한다.

[0010] 본 개시의 다른 구현에 따르면, 장치는 복수의 사용자들의 복수의 사용자 스피치 프로파일들을 저장하는 수단을 포함한다. 그 장치는 또한, 제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하는 수단을 포함한다. 그 장치는, 제 2 전력 모드에서, 세그먼트화 결과를 생성하기 위해 오디오 스트림의 오디오 피쳐 데이터를 분석하는 수단을 더 포함한다. 제 2 전력 모드에서 분석되는 오디오 피쳐 데이터는, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응함을 결정하는 것에 기초한다. 세그먼트화 결과는 오디오 스트림의 화자-동일 오디오 세그먼트들을 표시한다. 그 장치는 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들 중의 제 1 오디오 피쳐 데이터 세트에 대한 복수의 사용자 스피치 프로파일들의 비교를, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하는 수단을 더 포함한다. 그 장치는 또한, 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하는 수단을 포함한다. 제 1 사용자 스피치 프로파일은, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여 생성된다. 그 장치는 제 1 사용자 스피치 프로파일을 복수의 사용자 스피치 프로파일들에 추가하는 수단을 더 포함한다.

[0011] 본 개시의 다른 양태들, 이점들, 및 특징들은 다음의 섹션들: 즉, 도면의 간단한 설명, 상세한 설명, 및 청구항들을 포함하여 전체 출원의 검토 후 자명하게 될 것이다.

도면의 간단한 설명

[0012] 도 1 은 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리의 특정 예시적인 예의 블록 다이어그램이다.

도 2a 는 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리를 수행하도록 동작가능한 시스템의 특정 예시적인 양태의 다이어그램이다.

도 2b 는 본 개시의 일부 예들에 따른, 도 2a 의 시스템의 예시적인 컴포넌트들의 다이어그램이다.

도 3 은 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리와 연관된 동작들의 예시적인 양태의 다이어그램이다.

도 4 는 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리와 연관된 동작들의 예시적인 양태의 다이어

그램이다.

도 5 는 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리와 연관된 동작들의 예시적인 양태의 다이어그램이다.

도 6 은 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리와 연관된 동작들의 예시적인 양태의 다이어그램이다.

도 7 은 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리와 연관된 동작들의 예시적인 양태의 다이어그램이다.

도 8 은 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리와 연관된 동작들의 예시적인 양태의 다이어그램이다.

도 9 는 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리와 연관된 동작들의 예시적인 양태의 다이어그램이다.

도 10 은 본 개시의 일부 예들에 따른, 도 2a 의 시스템에 의해 수행될 수도 있는 사용자 스피치 프로파일 관리의 방법의 특정 구현의 다이어그램이다.

도 11 은 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리를 수행하도록 동작가능한 집적 회로의 일 예를 예시한다.

도 12 는 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리를 수행하도록 동작가능한 모바일 디바이스의 다이어그램이다.

도 13 은 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리를 수행하도록 동작가능한 헤드셋의 다이어그램이다.

도 14 는 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리를 수행하도록 동작가능한 웨어러블 전자 디바이스의 다이어그램이다.

도 15 는 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리를 수행하도록 동작가능한 음성 제어식 스피커 시스템의 다이어그램이다.

도 16 은 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리를 수행하도록 동작가능한 가상 현실 또는 증강 현실 헤드셋의 다이어그램이다.

도 17 은 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리를 수행하도록 동작가능한 차량의 제 1 예의 다이어그램이다.

도 18 은 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리를 수행하도록 동작가능한 차량의 제 2 예의 다이어그램이다.

도 19 는 본 개시의 일부 예들에 따른, 사용자 스피치 프로파일 관리를 수행하도록 동작가능한 디바이스의 특정 예시적인 예의 블록 다이어그램이다.

발명을 실시하기 위한 구체적인 내용

[0013]

사용자가 미리결정된 단어들 또는 문장들의 세트를 말하는 활성 사용자 등록을 사용하여 사용자 스피치 프로파일을 트레이닝하는 것은 시간 소모적이고 불편할 수 있다. 예를 들어, 사용자는 사용자 스피치 프로파일을 트레이닝하기 위해 미리 계획하고 시간을 소비해야 한다. 본 명세서에 개시된 사용자 스피치 프로파일 관리의 시스템들 및 방법들은 활성 사용자 등록을 사용하지 않고 다중 화자들 사이를 구별할 수 있게 한다. 예를 들어, 하나 이상의 사용자들의 스피치에 대응하는 오디오 스트림이 세그먼트화기에 의해 수신된다. 세그먼트화기는, 오디오 스트림의 화자-동질 오디오 세그먼트들을 표시하는 세그먼트화 결과를 생성한다. 본 명세서에서 사용되는 바와 같이, "화자-동일 오디오 세그먼트" 는 동일한 화자의 스피치를 나타내는 오디오 부분들 (예컨대, 오디오 프레임들) 을 포함한다. 예를 들어, 세그먼트화 결과는 동일한 화자의 스피치를 나타내는 오디오 프레임들의 세트를 식별한다. 프로파일 관리기는 오디오 프레임들의 세트 중의 오디오 프레임의 오디오 피쳐들을 비교하여, 오디오 피쳐들이 복수의 저장된 사용자 스피치 프로파일들 중 임의의 것과 매칭하는지 여부를 결정한다. 프로파일 관리기는, 오디오 피쳐들이 저장된 사용자 스피치 프로파일들 중 임의의 것

과 매칭하지 않음을 결정하는 것에 응답하여, 오디오 피쳐들에 적어도 부분적으로 기초하여 사용자 스피치 프로파일을 생성한다. 대안적으로, 프로파일 관리기는, 오디오 피쳐들이 저장된 사용자 스피치 프로파일과 매칭함을 결정하는 것에 응답하여, 오디오 피쳐들에 적어도 부분적으로 기초하여 저장된 사용자 스피치 프로파일을 업데이트한다. 따라서, 사용자 스피치 프로파일은, 예컨대, 전화 호출 또는 회의 동안, 수동 등록을 사용하여 생성되거나 업데이트될 수 있다. 프로파일 관리기는 또한, 다중 화자들 사이의 대화 동안 다중의 사용자 스피치 프로파일들을 생성 또는 업데이트할 수 있다. 특정 예에서, 프로파일 관리기는 생성된 또는 업데이트된 스피치 프로파일의 프로파일 식별자를 하나 이상의 추가적인 오디오 어플리케이션들에 제공한다. 예를 들어, 오디오 어플리케이션은, 대응하는 텍스트에 대한 화자들을 표시하는 라벨들을 갖는 트랜스크립트를 생성하기 위해 오디오 스트림에 대한 스피치-대-텍스트 변환을 수행할 수 있다.

[0014] 본 개시의 특정 양태들이 이하에서 도면들을 참조하여 설명된다. 설명에 있어서, 공통 피쳐들은 도면들 전반에 걸쳐 공통 참조 부호들에 의해 지정된다. 일부 도면들에 있어서, 특정 타입의 피쳐의 다중의 인스턴스들이 사용된다. 이들 피쳐들은 물리적으로 및/또는 논리적으로 구별되지만, 각각에 대해 동일한 참조부호가 사용되며, 상이한 인스턴스들은 참조부호에 문자를 추가하는 것에 의해 구별된다. 그룹 또는 타입으로서의 피쳐들이 본 명세서에서 참조될 경우 (예컨대, 피쳐들 중 특정한 피쳐가 참조되고 있지 않을 경우), 참조 부호는 구별 문자 없이 사용된다. 하지만, 동일한 타입의 다중의 피쳐들 중 하나의 특정한 피쳐가 본 명세서에서 참조될 경우, 참조 부호는 구별 문자와 함께 사용된다. 예를 들어, 도 1 을 참조하면, 다중의 프레임들이 예시되고 참조 부호들 (102A, 102B, 및 102C) 과 연관된다. 프레임 (102A) 과 같이 이들 프레임들 중 특정한 프레임을 지칭할 때, 구별 문자 "A" 가 사용된다. 하지만, 이들 프레임 중 어떤 임의의 프레임을 지칭하거나 또는 이들 프레임들을 그룹으로서 지칭할 때는 구별 문자 없이 참조 부호 102 가 사용된다.

[0015] 본 명세서에서 사용된 바와 같이, 다양한 용어는 오직 특정 구현들을 설명할 목적으로 사용되고, 구현들을 한정하는 것으로 의도되지 않는다. 예를 들어, 단수 형태들 ("a", "an", 및 "the") 은, 문맥이 분명하게 달리 표시하지 않으면, 복수의 형태들을 물론 포함하도록 의도된다. 추가로, 본 명세서에서 설명된 일부 피쳐들은 일부 구현들에서 단수이고 다른 구현들에서는 복수이다. 예시하자면, 도 2a 은 하나 이상의 프로세서들 (도 2a 의 "프로세서(들)" (220)) 을 포함하는 디바이스 (202) 를 도시하며, 이는 일부 구현들에서는 디바이스 (202) 가 단일 프로세서 (220) 를 포함하고 다른 구현들에서는 디바이스 (202) 가 다중의 프로세서들 (220) 을 포함함을 나타낸다. 본 명세서에서의 참조의 용어를 위해, 그러한 피쳐들은 일반적으로 "하나 이상의" 피쳐들로서 소개되고, 후속하여, 피쳐들 중 다중의 피쳐들과 관련된 양태들이 설명되지 않는 한, 단수로 지칭된다.

[0016] 본 명세서에서 사용된 바와 같이, 용어들 "구비한다", "구비하다", 및 "구비하는" 은 "포함한다", "포함하다", 또는 "포함하는" 과 상호교환가능하게 사용될 수도 있다. 부가적으로, 용어 "여기서 (wherein)" 는 "여기에서 (where)" 와 상호교환가능하게 사용될 수도 있다. 본 명세서에서 사용된 바와 같이, "예시적인" 은 예, 구현, 및/또는 양태를 나타내며, 한정하는 것으로서 또는 선호도 또는 선호된 구현을 나타내는 것으로서 해석되지 않아야 한다. 본 명세서에서 사용된 바와 같이, 구조, 컴포넌트, 동작 등과 같은 엘리먼트를 수정하는데 사용되는 서수 용어 (예컨대, "제 1", "제 2", "제 3" 등) 는 홀로 다른 엘리먼트에 관하여 엘리먼트의 임의의 우선순위 또는 순서를 표시하는 것이 아니라, 오히려 단지 엘리먼트를 (서수 용어의 사용이 없다면) 동일한 명칭을 갖는 다른 엘리먼트로부터 구별할 뿐이다. 본 명세서에서 사용된 바와 같이, 용어 "세트" 는 특정 엘리먼트의 하나 이상을 지칭하고, 용어 "복수" 는 특정 엘리먼트의 배수 (예컨대, 2 이상) 를 지칭한다.

[0017] 본 명세서에서 사용된 바와 같이, "커플링된" 은 "통신가능하게 커플링된", "전기적으로 커플링된", 또는 "물리적으로 커플링된" 을 포함할 수도 있으며, 또한 (또는 대안적으로) 이들의 임의의 조합들을 포함할 수도 있다. 2개의 디바이스들 (또는 컴포넌트들) 은 하나 이상의 다른 디바이스들, 컴포넌트들, 와이어들, 버스들, 네트워크들 (예컨대, 유선 네트워크, 무선 네트워크, 또는 이들의 조합) 등을 통해 직접적으로 또는 간접적으로 커플링 (예컨대, 통신가능하게 커플링, 전기적으로 커플링, 또는 물리적으로 커플링) 될 수도 있다. 전기적으로 커플링되는 2개의 디바이스들 (또는 컴포넌트들) 은 동일한 디바이스에 또는 상이한 디바이스들에 포함될 수도 있으며, 예시적 비한정적인 예들로서, 전자기기들, 하나 이상의 커넥터들, 또는 유도성 커플링을 통해 연결될 수도 있다. 일부 구현들에 있어서, 전기 통신에서와 같이 통신가능하게 커플링되는 2개의 디바이스들 (또는 컴포넌트들) 은 하나 이상의 와이어들, 버스들, 네트워크들 등을 통해 직접적으로 또는 간접적으로 신호들 (예를 들어, 디지털 신호들 또는 아날로그 신호들) 을 전송 및 수신할 수도 있다. 본 명세서에서 사용된 바와 같이, "직접적으로 커플링된" 은 컴포넌트들을 개재하지 않고 커플링 (예컨대, 통신가능하게 커플링, 전기적으로 커플링, 또는 물리적으로 커플링) 되는 2개의 디바이스들을 포함할 수도 있다.

[0018] 본 개시에 있어서, "결정하는 것", "계산하는 것", "추정하는 것", "시프팅하는 것", "조정하는 것" 등과 같은

용어들은 하나 이상의 동작들이 어떻게 수행되는지를 설명하기 위해 사용될 수도 있다. 그러한 용어들은 한정하는 것으로서 해석되지 않아야 하고 다른 기법들이 유사한 동작들을 수행하는데 활용될 수도 있음을 유의해야 한다. 부가적으로, 본 명세서에서 지칭되는 바와 같이, "생성하는 것", "계산하는 것", "추정하는 것", "사용하는 것", "선택하는 것", "액세스하는 것" 및 "결정하는 것"은 상호교환가능하게 사용될 수도 있다. 예를 들어, 파라미터 (또는 신호)를 "생성하는 것", "계산하는 것", "추정하는 것", 또는 "결정하는 것"은 파라미터 (또는 신호)를 능동적으로 생성하는 것, 추정하는 것, 계산하는 것, 또는 결정하는 것을 지칭할 수도 있거나, 또는 예컨대, 다른 컴포넌트 또는 디바이스에 의해 이미 생성된 파라미터 (또는 신호)를 사용하는 것, 선택하는 것, 또는 액세스하는 것을 지칭할 수도 있다.

[0019] 도 1은 사용자 스펙치 프로파일 관리의 예 (100)를 예시한다. 예 (100)에서, 세그먼트화기 (124) 및 프로파일 관리기 (126)는, 화자들의 활성 사용자 등록을 사용하지 않고 다중 화자들로부터의 스펙치를 구별하기 위해 오디오 스트림 (141)을 프로세싱하도록 협력한다.

[0020] 오디오 스트림 (141)은, 도 1에서 프레임들 (102A, 102B, 102C)로서 나타난 복수의 별개의 부분들을 포함한다. 이 예에서, 각각의 프레임 (102)은 오디오 스트림 (141)의 오디오의 일부를 나타내거나 인코딩한다. 예시하자면, 각각의 프레임 (102)은 오디오 스트림의 오디오의 1/2초를 나타낼 수도 있다. 다른 예들에서, 상이한 사이즈들 또는 지속기간들의 프레임들이 사용될 수도 있다.

[0021] 오디오 스트림 (141)은 세그먼트화기 (124)에 입력으로서 제공된다. 세그먼트화기 (124)는 오디오 스트림 (141)을 세그먼트들로 분할하고, 각각의 세그먼트를 단일 화자로부터의 스펙치, 다중 화자들로부터의 스펙치, 또는 묵음을 포함하는 것으로서 식별하도록 구성된다. 예를 들어, 도 1에서, 세그먼트화기 (124)는, 화자-동일 오디오 세그먼트 (111A)를 함께 형성하는 오디오 부분들의 제 1 세트 (151A)를 식별하였다. 유사하게, 세그먼트화기 (124)는, 제 2 화자-동일 오디오 세그먼트 (111B)를 함께 형성하는 오디오 부분들의 제 2 세트 (151C)를 식별하였다. 세그먼트화기 (124)는 또한, 묵음 또는 믹싱된 화자 오디오 세그먼트 (113)를 함께 형성하는 오디오 부분들의 세트 (151B)를 식별하였다. 묵음 또는 믹싱된 화자 오디오 세그먼트 (113)는 다중 화자로부터의 스펙치를 포함하거나 스펙치를 포함하지 않는 사운드 (예를 들어, 묵음 또는 비-스펙치 잡음)를 나타낸다.

[0022] 특정 예에 있어서, 하기에서 더 상세히 설명되는 바와 같이, 세그먼트화기 (124)는, 화자 세그먼트화를 수행하도록 트레이닝되는 하나 이상의 머신 러닝 세그먼트화 모델들 (예를 들어, 뉴럴 네트워크들)을 사용함으로써 오디오 스트림 (141)을 세그먼트들로 분할한다. 이 예에서, 화자들의 사전 등록은 요구되지 않는다. 오히려, 세그먼트화기 (124)는 오디오 스트림 (141)의 별개의 오디오 프레임들 사이의 화자 특성들을 비교함으로써 2 이상의 이전에 미지의 화자들 사이를 구별하도록 트레이닝된다. 세그먼트화기 (124)가 구별할 수 있는 화자들의 특정 수는 머신 러닝 세그먼트화 모델(들)의 구성 및 트레이닝에 의존한다. 예시하자면, 특정 양태에서, 세그먼트화기 (124)는 3명의 화자들 사이를 구별하도록 구성될 수도 있으며, 이 경우, 머신 러닝 세그먼트화 모델(들)은 화자 1 출력 노드, 화자 2 출력 노드, 화자 3 출력 노드, 묵음 출력 노드, 및 믹싱된 출력 노드에 대응하는 5개의 출력 계층 노드들을 포함할 수도 있다. 이 양태에서, 각각의 출력 노드는, 분석되고 있는 오디오 부분들 (151)의 세트가 개별 출력 노드와 연관된 가능성을 표시하는 세그먼트화 스코어를 출력으로서 생성하도록 트레이닝된다. 예시하자면, 화자 1 출력 노드는 오디오 부분들 (151)의 세트가 제 1 화자의 스펙치를 나타냄을 표시하는 세그먼트화 스코어를 생성하고, 화자 2 출력 노드는 오디오 부분들 (151)의 세트가 제 2 화자의 스펙치를 나타냄을 표시하는 세그먼트화 스코어를 생성하는 등등이다.

[0023] 특정 구현에서, 머신 러닝 세그먼트화 모델(들)은, 세그먼트화기 (124)가 3명의 화자들 사이를 구별하도록 구성될 때 4개의 출력 계층 노드들을 포함할 수도 있다. 예를 들어, 4개의 출력 계층 노드들은 화자 1 출력 노드, 화자 2 출력 노드, 화자 3 출력 노드 및 묵음 출력 노드를 포함하고, 믹싱된 출력 노드를 포함하지는 않는다. 이 구현에서, 믹싱된 스펙치는, 오디오 부분들 (151)의 세트가 대응하는 화자들의 스펙치를 나타냄을 표시하는 다중 화자 출력 노드들의 세그먼트화 스코어들에 의해 표시된다.

[0024] 특정 구현에서, 머신 러닝 세그먼트화 모델(들)은, 세그먼트화기 (124)가 3명의 화자들 사이를 구별하도록 구성될 때 3개의 출력 계층 노드들을 포함할 수도 있다. 예를 들어, 3개의 출력 계층 노드들은 화자 1 출력 노드, 화자 2 출력 노드, 및 화자 3 출력 노드를 포함하고, 묵음 출력 노드를 포함하지는 않는다. 이 구현에서, 묵음은, 오디오 부분들 (151)의 세트가 대응하는 화자의 스펙치를 나타내지 않음을 표시하는 화자 출력 노드들의 각각의 세그먼트화 스코어에 의해 표시된다. 예시하자면, 오디오 부분들 (151)의 세트가 제 1 화자의 스펙치를 나타내지 않음을 표시하는 세그먼트화 스코어를 화자 1 출력 노드가 생성하고, 오디오 부분들

(151)의 세트가 제 2 화자의 스피치를 나타내지 않음을 표시하는 세그먼트화 스코어를 화자 2 출력 노드가 생성하고, 그리고 오디오 부분들 (151)의 세트가 제 3 화자의 스피치를 나타내지 않음을 표시하는 세그먼트화 스코어를 화자 3 출력 노드가 생성할 때, 묵음이 표시된다. 일부 양태들에 있어서, 본 명세서에서 사용된 바와 같이, "묵음"은 "비-스피치 잡음"과 같은 "스피치의 부존재"를 지칭할 수 있다.

[0025] 화자-동일 오디오 세그먼트 (111)의 오디오 부분들 (151)의 각각은 오디오 스트림 (141)의 다중의 프레임들 (102)을 포함한다. 예시하자면, 오디오 부분들 (151A)의 각각은 오(5)초의 사운드를 나타내는 십(10)개의 오디오 프레임들 (102)을 포함할 수도 있다. 다른 예들에서, 상이한 수의 프레임들이 각각의 오디오 부분에 포함되거나, 또는 프레임들은 각각의 오디오 부분 (151A)이 10초 초과 또는 미만의 사운드를 나타내도록 상이한 사이즈를 갖는다. 부가적으로, 각각의 화자-동일 오디오 세그먼트 (111)는 다중의 오디오 부분들 (151)을 포함한다. 화자-동일 오디오 세그먼트 (111)당 오디오 부분들 (151)의 수는 가변적이다. 예를 들어, 화자-동일 오디오 세그먼트 (111)는, 예컨대, 묵음 기간 (예컨대, 임계 지속기간의 묵음)에 의해 또는 다른 화자의 스피치에 의해, 화자의 스피치가 중단될 때까지 계속될 수도 있다.

[0026] 세그먼트화기 (124)는 화자-동일 오디오 세그먼트들 (111)을 식별하는 세그먼트화 결과를 프로파일 관리기 (126)에 제공한다. 프로파일 관리기는 사용자 스피치 프로파일들 (USP들) (150)을 메모리에 유지한다. 각각의 사용자 스피치 프로파일 (150)은 프로파일 식별자 (ID) (155)와 연관된다. 특정 양태에서, 프로파일 ID들 (155) 및 사용자 스피치 프로파일들 (150)은 프로파일 관리기 (126)에 의해 생성된다 (예컨대, 프로파일 ID들 (155) 및 사용자 스피치 프로파일들 (150)은 사용자 사전 등록에 기초하지 않음).

[0027] 세그먼트화 결과에 응답하여, 프로파일 관리기 (126)는 화자-동일 오디오 세그먼트 (111)로부터의 오디오 부분들 (151)을 사용자 스피치 프로파일들 (150)과 비교한다. 오디오 부분들 (151)이 사용자 스피치 프로파일들 (150)중 하나와 매칭하면 (예컨대, 충분히 유사하면), 프로파일 관리기 (126)는 오디오 부분들 (151)에 기초하여 사용자 스피치 프로파일 (150)을 업데이트한다. 예를 들어, 화자-동일 오디오 세그먼트 (111A)의 오디오 부분들 (151A)이 사용자 스피치 프로파일 (150A)과 충분히 유사하면, 프로파일 관리기 (126)는 사용자 스피치 프로파일 (150A)을 업데이트하기 위해 오디오 부분들 (151A)을 사용한다.

[0028] 오디오 부분들 (151)이 사용자 스피치 프로파일들 (150)중 임의의 것과 매칭하지 않으면, 프로파일 관리기 (126)는 오디오 부분들 (151)에 기초하여 사용자 스피치 프로파일 (150)을 추가한다. 예를 들어, 도 1에서, 프로파일 관리기 (126)는 화자-동일 오디오 세그먼트 (111C)의 오디오 부분들 (151C)에 기초하여 사용자 스피치 프로파일 (150C)을 생성하고 프로파일 ID (155C)를 사용자 스피치 프로파일 (150C)에 배정한다.

[0029] 프로파일 관리기 (126)는 또한 오디오 스트림 (141)에서의 화자 또는 화자 변경을 표시하는 출력을 생성한다. 예를 들어, 출력은 화자-동일 오디오 세그먼트 (111)와 매칭하는 사용자 스피치 프로파일 (150)의 프로파일 ID (155)를 포함할 수도 있다. 하나 이상의 오디오 분석 어플리케이션들 (180), 이들은 화자 또는 화자 변경에 기초하여 결과를 생성한다. 예를 들어, 오디오 분석 어플리케이션들 (180)은 검출된 스피치를 트랜스크립트하여 텍스트를 생성할 수도 있고, 화자들에서의 변경이 발생할 때를 텍스트에 표시할 수도 있다.

[0030] 도 2a를 참조하면, 사용자 스피치 프로파일 관리를 수행하도록 구성된 시스템의 특정 예시적인 양태가 개시되고 일반적으로 200으로 지정된다. 시스템 (200)은 마이크로폰 (246)에 커플링되는 디바이스 (202)를 포함한다. 디바이스 (202)는 도 1의 세그먼트화기 (124) 및 프로파일 관리기 (126)를 사용하여 사용자 스피치 프로파일 관리를 수행하도록 구성된다. 특정 양태에서, 디바이스 (202)는, 피쳐 추출기 (222), 세그먼트화기 (124), 프로파일 관리기 (126), 화자 검출기 (278), 하나 이상의 오디오 분석 어플리케이션들 (180), 또는 이들의 조합을 포함하는 하나 이상의 프로세서들 (220)을 포함한다.

[0031] 피쳐 추출기 (222)는 오디오 스트림의 오디오 부분들 (예를 들어, 오디오 프레임들)의 피쳐들을 나타내는 오디오 피쳐 데이터 세트들을 생성하도록 구성된다. 세그먼트화기 (124)는 동일한 화자의 스피치를 나타내는 오디오 부분들 (또는 오디오 피쳐 데이터 세트들)을 표시하도록 구성된다. 프로파일 관리기 (126)는 동일한 화자의 스피치를 나타내는 오디오 부분들 (또는 오디오 피쳐 데이터 세트들)에 기초하여 사용자 스피치 프로파일들을 생성 (또는 업데이트)하도록 구성된다. 화자 검출기 (278)는 오디오 스트림에서 검출된 화자들의 카운트를 결정하도록 구성된다. 특정 구현에서, 화자 검출기 (278)는 오디오 스트림에서 다중 화자들을 검출하는 것에 응답하여 세그먼트화기 (124)를 활성화하도록 구성된다. 이 구현에서, 화자 검출기 (278)가 오디오 스트림에서 단일 화자를 검출하고 프로파일 관리기 (126)가 단일 화자에 대응하는 사용자 스피치 프로파일을 생성 (또는 업데이트)할 경우, 세그먼트화기 (124)는 우회된다. 특정 구현에서, 하나 이상의 오디오 분석 어플리케이션들 (180)은 사용자 스피치 프로파일들에 기초하여 오디오 분석을 수행하도록 구

성된다.

[0032] 특정 양태에서, 디바이스 (202) 는 하나 이상의 프로세서들 (220) 에 커플링된 메모리 (232) 를 포함한다. 특정 양태에서, 메모리 (232) 는 버퍼 (268) 와 같은 하나 이상의 버퍼들을 포함한다. 메모리 (232) 는 세그먼트화 임계치 (257) (도 2a 에서의 "세그먼트화 임계치") 와 같은 하나 이상의 임계치들을 저장하도록 구성된다. 특정 양태에서, 하나 이상의 임계치들은 사용자 입력, 구성 설정들, 디폴트 데이터, 또는 이들의 조합에 기초한다.

[0033] 특정 양태에서, 메모리 (232) 는 피쳐 추출기 (222), 화자 검출기 (278), 세그먼트화기 (124), 프로파일 관리기 (126), 하나 이상의 오디오 분석 어플리케이션들 (180), 또는 이들의 조합에 의해 생성된 데이터를 저장하도록 구성된다. 예를 들어, 메모리 (232) 는 복수의 사용자들 (242) 의 복수의 사용자 스피치 프로파일들 (150), 세그먼트화 결과 (236) (도 2a 에서의 "세그먼트화 결과"), 오디오 피쳐 데이터 세트들 (252), 오디오 부분 (151), 세그먼트화 스코어들 (254) (도 2a 에서의 "세그먼트화 스코어들"), 데이터 세트 세그먼트화 결과 (256) (도 2a 에서의 "데이터 세트 세그먼트화 결과"), 프로파일 ID (155), 또는 이들의 조합을 저장하도록 구성된다. 메모리 (232) 는 프로파일 업데이트 데이터 (272), 사용자 상호작용 데이터 (274) (도 2a 에서의 "사용자 상호작용 데이터"), 또는 이들의 조합을 저장하도록 구성된다.

[0034] 디바이스 (202) 는 모뎀, 네트워크 인터페이스, 입력 인터페이스를 통해, 또는 마이크로폰 (246) 으로부터 오디오 스트림 (141) 을 수신하도록 구성된다. 특정 양태에서, 오디오 스트림 (141) 은 하나 이상의 오디오 부분들 (151) 을 포함한다. 예를 들어, 오디오 스트림 (141) 은 오디오 부분들 (151) 에 대응하는 오디오 프레임들의 세트로 분할될 수도 있으며, 각각의 오디오 프레임은 오디오 스트림 (141) 의 시간-원도우화 부분을 나타낸다. 다른 예에서, 오디오 스트림 (141) 은 오디오 부분들 (151) 을 생성하기 위해 다른 방식으로 분할될 수도 있다. 오디오 스트림 (141) 의 각각의 오디오 부분 (151) 은 묶음, 하나 이상의 사용자들 (242) 로부터의 스피치, 또는 다른 사운드들을 포함하거나 나타낸다. 단일 사용자로부터의 스피치를 나타내는 오디오 부분들 (151) 의 세트는 화자-동일 오디오 세그먼트 (111) 로서 지칭된다. 각각의 화자-동일 오디오 세그먼트 (111) 는 복수의 오디오 부분들 (151) (예를 들어, 복수의 오디오 프레임) 을 포함한다. 특정 양태에서, 화자-동일 오디오 세그먼트 (111) 는 적어도 오디오 프레임들의 임계 카운트 (예를 들어, 5개의 오디오 프레임들) 를 포함한다. 특정 양태에서, 화자-동일 오디오 세그먼트 (111) 는 동일한 사용자의 스피치에 대응하는 오디오 부분들 (151) 의 연속적인 세트를 포함한다. 특정 양태에서, 오디오 부분들 (151) 의 연속적인 세트는 오디오 부분들 (151) 의 하나 이상의 서브세트들을 포함할 수도 있으며, 각각의 서브세트는 스피치에서 자연적인 짧은 일시중지들을 나타내는 임계 묶음 미만에 대응한다.

[0035] 오디오 스트림 (141) 은 화자-동일 오디오 세그먼트들, 묶음에 대응하는 오디오 세그먼트들, 다중 화자들에 대응하는 오디오 세그먼트들, 또는 이들의 조합의 다양한 조합들을 포함할 수 있다. 일 예로서, 도 2a 에서, 오디오 스트림 (141) 은 사용자 (242A) 의 스피치에 대응하는 화자-동일 오디오 세그먼트 (111A) 의 오디오 부분들 (151A), 묶음 (또는 비-스피치 잡음) 에 대응하는 오디오 세그먼트 (113) 의 오디오 부분들 (151B), 및 사용자 (242B) 의 스피치에 대응하는 화자-동일 오디오 세그먼트 (111B) 의 오디오 부분들 (151C) 을 포함한다. 다른 예들에서, 오디오 스트림 (114) 은 오디오 세그먼트들의 상이한 세트 또는 배열을 포함한다. 오디오 부분이 오디오 프레임을 지칭하는 것으로서 설명되지만, 다른 구현들에서, 오디오 부분은 오디오 프레임의 일부, 다중의 오디오 프레임들, 특정 스피치 또는 재생 지속기간에 대응하는 오디오 데이터, 또는 이들의 조합을 지칭한다.

[0036] 피쳐 추출기 (222) 는 오디오 스트림 (141) 의 오디오 피쳐들을 추출 (예를 들어, 결정) 하여 오디오 피쳐 데이터 세트들 (252) 을 생성하도록 구성된다. 예를 들어, 피쳐 추출기 (222) 는 오디오 스트림 (141) 의 오디오 부분 (151) 의 오디오 피쳐들을 추출하여 오디오 피쳐 데이터 세트 (AFDS) (252) 를 생성하도록 구성된다. 특정 양태에서, 오디오 피쳐 데이터 세트 (252) 는 임베딩 벡터와 같은 오디오 피쳐 벡터를 포함한다. 특정 양태에서, 오디오 피쳐 데이터 세트 (252) 는 오디오 부분 (151) 의 멜-주파수 캡스트럴 계수들 (MFCC들) 을 표시한다. 특정 예에서, 피쳐 추출기 (222) 는 오디오 부분들 (151A) 의 오디오 피쳐들을 추출함으로써 하나 이상의 오디오 피쳐 데이터 세트들 (252A) 을 생성한다. 피쳐 추출기 (222) 는 오디오 부분들 (151B) 의 오디오 피쳐들을 추출함으로써 하나 이상의 오디오 피쳐 데이터 세트들 (252B) 을 생성한다. 피쳐 추출기 (222) 는 오디오 부분들 (151C) 의 오디오 피쳐들을 추출함으로써 하나 이상의 오디오 피쳐 데이터 세트들 (252C) 을 생성한다. 오디오 피쳐 데이터 세트들 (252) 은 하나 이상의 오디오 피쳐 데이터 세트들 (252A), 하나 이상의 오디오 피쳐 데이터 세트들 (252B), 하나 이상의 오디오 피쳐 데이터 세트들 (252C), 또는 이들의

조합을 포함한다.

[0037] 예시적인 예에서, 피쳐 추출기 (222) 는 오디오 스트림 (141) 의 각각의 프레임의 오디오 피쳐들을 추출하고, 각각의 프레임의 오디오 피쳐들을 세그먼트화기 (124) 에 제공한다. 특정 양태에서, 세그먼트화기 (124) 는 특정 수의 오디오 프레임들 (예를 들어, 10개의 오디오 프레임들) 의 오디오 피쳐들에 대한 세그먼트화 스코어들의 세트 (예를 들어, 세그먼트화 스코어들 (254)) 를 생성하도록 구성된다. 예를 들어, 오디오 부분 (151) 은 특정 수의 오디오 프레임들 (예컨대, 10개의 오디오 프레임들) 을 포함한다. (예컨대, 세그먼트화 스코어들의 특정 세트를 생성하기 위해 세그먼트화기 (124) 에 의해 사용되는) 특정 수의 오디오 프레임들의 오디오 피쳐들은 오디오 피쳐 데이터 세트 (252) 에 대응한다. 예를 들어, 피쳐 추출기 (222) 는 제 10 오디오 프레임의 제 10 오디오 피쳐들을 포함하여 제 1 오디오 프레임의 제 1 오디오 피쳐들, 제 2 오디오 프레임의 제 2 오디오 피쳐들 등을 추출한다. 세그먼트화기 (124) 는 제 10 오디오 피쳐들을 포함하여 제 1 오디오 피쳐들, 제 2 오디오 피쳐들 등에 기초하여 제 1 세그먼트화 스코어들 (254) 을 생성한다. 예를 들어, 제 1 오디오 피쳐들, 제 2 오디오 피쳐들, 및 제 10 오디오 피쳐들까지는 제 1 오디오 피쳐 데이터 세트 (252) 에 대응한다. 유사하게, 피쳐 추출기 (222) 는 제 20 오디오 프레임의 제 20 오디오 피쳐들을 포함하여 제 11 오디오 프레임의 제 11 오디오 피쳐들, 제 12 오디오 프레임의 제 12 오디오 피쳐들 등을 추출한다. 세그먼트화기 (124) 는 제 20 오디오 피쳐들을 포함하여 제 11 오디오 피쳐들, 제 12 오디오 피쳐들 등에 기초하여 제 2 세그먼트화 스코어들 (254) 을 생성한다. 예를 들어, 제 11 오디오 피쳐들, 제 12 오디오 피쳐들, 및 제 20 오디오 피쳐들까지는 제 2 오디오 피쳐 데이터 세트 (252) 에 대응한다. 10개의 오디오 프레임들에 기초하여 세그먼트화 스코어들의 세트를 생성하는 것이 예시적인 예로서 제공됨이 이해되어야 한다. 다른 예들에서, 세그먼트화기 (124) 는 10개 미만 또는 10개 초과 오디오 프레임들에 기초하여 세그먼트화 스코어들의 세트를 생성한다. 예를 들어, 오디오 부분 (151) 은 10개 미만 또는 10개 초과 오디오 프레임들을 포함한다.

[0038] 세그먼트화기 (124) 는 각각의 오디오 피쳐 데이터 세트에 대한 세그먼트화 스코어들의 세트 (예를 들어, 세그먼트화 스코어들 (254)) 를 생성하도록 구성된다. 예를 들어, 세그먼트화기 (124) 에 대한 오디오 피쳐 데이터 세트 (252) 의 입력에 응답하여, 세그먼트화기 (124) 는 다중의 세그먼트화 스코어들 (254) 을 생성한다. 오디오 피쳐 데이터 세트 (252) 에 응답하여 생성된 세그먼트화 스코어들 (254) 의 수는 세그먼트화기 (124) 가 구별하도록 트레이닝되는 화자들의 수에 의존한다. 일 예로서, 세그먼트화기 (124) 는 K개의 세그먼트화 스코어들 (254) 의 세트를 생성함으로써 K개의 별개의 화자들의 스피치를 구별하도록 구성된다. 이 예에서, 각각의 세그먼트화 스코어 (254) 는, 세그먼트화기 (124) 에 입력된 오디오 피쳐 데이터 세트가 대응하는 화자의 스피치를 나타낼 확률을 표시한다. 예시하자면, 세그먼트화기 (124) 가 화자 (292A), 화자 (292B), 및 화자 (292C) 와 같은 세명(3명)의 상이한 화자들의 스피치를 구별하도록 구성될 때, K 는 삼(3)과 동일하다. 이 예시적인 예에서, 세그먼트화기 (124) 는 세그먼트화기 (124) 에 입력된 각각의 오디오 피쳐 데이터 세트 (252) 에 대해, 세그먼트화 스코어 (254A), 세그먼트화 스코어 (254B), 및 세그먼트화 스코어 (254C) 와 같은 세개(3)의 세그먼트화 스코어들 (254) 을 출력하도록 구성된다. 이 예시적인 예에서, 세그먼트화 스코어 (254A) 는 오디오 피쳐 데이터 세트 (252) 가 화자 (292A) 의 스피치를 나타낼 확률을 표시하고, 세그먼트화 스코어 (254B) 는 오디오 피쳐 데이터 세트 (252) 가 화자 (292B) 의 스피치를 나타낼 확률을 표시하고, 세그먼트화 스코어 (254C) 는 오디오 피쳐 데이터 세트 (252) 가 화자 (292C) 의 스피치를 나타낼 확률을 표시한다. 다른 예들에서, 세그먼트화기 (124) 가 구별하도록 구성되는 화자들의 카운트 (상기의 예들에서 K) 는 3 초과이거나 3 미만이다.

[0039] 화자들 (292) 은 가장 최근에, 예를 들어, 세그먼트화 윈도우 동안, 세그먼트화기 (124) 에 의해 검출된 화자들의 세트에 대응한다. 특정 양태에서, 화자들 (292) 은 세그먼트화기 (124) 에 의해 구별되기 위해 사전 등록되어야 하는 것은 아니다. 세그먼트화기 (124) 는, 사전 등록되지 않은 다중의 사용자들의 스피치 사이를 구별함으로써 다중의 사용자들의 수동 등록을 가능케 한다. 세그먼트화 윈도우는 오디오 부분들 (예를 들어, 20개의 오디오 프레임들), 특정 시간 윈도우 (예를 들어, 20 밀리초) 동안 세그먼트화기 (124) 에 의해 프로세싱된 오디오 부분들, 또는 특정 스피치 지속기간 또는 재생 지속기간에 대응하는 오디오 부분들의 특정 카운트까지를 포함한다.

[0040] 도 2a 에 예시된 예에서, 오디오 스트림 (141) 의 오디오 부분들 (151) 의 피쳐들을 나타내는 오디오 피쳐 데이터 세트들 (252) 은 세그먼트화기 (124) 에 대한 입력으로서 제공될 수도 있다. 이 예에서, 오디오 피쳐 데이터 세트들 (252) 은 사용자 (242A) 의 스피치를 나타내는 오디오 피쳐 데이터 세트들 (252A), 목음을 나타내는 오디오 피쳐 데이터 세트들 (252B), 및 사용자 (242B) 의 스피치를 나타내는 오디오 피쳐 데이터 세트

(252C)와 같은, 사용자들 (242) 중 2 이상의 스피치를 나타낸다. 특정 구현에서, 세그먼트화기 (124)는 사용자들 (242)에 관한 사전 정보를 갖지 않는다. 예를 들어, 사용자들 (242)은 디바이스 (202)에 사전 등록하지 않았다. 오디오 피쳐 데이터 세트 (252)의 입력에 응답하여, 세그먼트화기 (124)는 세그먼트화 스코어 (254A), 세그먼트화 스코어 (254B), 및 세그먼트화 스코어 (254C)를 출력한다. 각각의 세그먼트화 스코어 (254)는 오디오 피쳐 데이터 세트 (252)가 개별 화자 (292)의 스피치를 나타낼 확률을 표시하고, 세그먼트화 스코어들 (254)의 각각은 세그먼트화 임계치 (257)와 비교된다. 오디오 피쳐 데이터 세트 (252)에 대한 세그먼트화 스코어들 (254) 중 하나가 세그먼트화 임계치 (257)를 만족하면, 대응하는 화자 (292)의 스피치는 오디오 피쳐 데이터 세트 (252)에서 검출된 것으로서 표시된다. 예시하자면, 오디오 피쳐 데이터 세트 (252)에 대한 세그먼트화 스코어 (254A)가 세그먼트화 임계치 (257)를 만족하면, 화자 (292A)의 스피치는 오디오 피쳐 데이터 세트 (252) (및 오디오 피쳐 데이터 세트 (252)에 의해 나타낸 오디오 부분 (151))에서 검출된 것으로서 표시된다. 유사한 동작이 오디오 피쳐 데이터 세트들 (252A), 오디오 피쳐 데이터 세트들 (252B), 및 오디오 피쳐 데이터 세트들 (252C)의 각각에 대해 수행된다.

[0041] 세그먼트화기 (124)는 세그먼트화 윈도우 동안 (오디오 피쳐 데이터 세트 (252)에 의해 나타낸 스피치와 연관된 바와 같은 세그먼트화기 (124)에 알려지지 않은 사용자 (242)와 같은) 미지의 사용자에게 대한 플레이스홀더로서 화자 (292)를 사용한다. 예를 들어, 오디오 피쳐 데이터 세트들 (252A)은 사용자 (242A)의 스피치에 대응한다. 세그먼트화기 (124)는, 오디오 피쳐 데이터 세트들 (252A)이 화자 (292A) (예를 들어, 사용자 (242A)에 대한 플레이스홀더)의 스피치에 대응함을 표시하기 위해 세그먼트화 임계치 (257)를 만족시키는 오디오 피쳐 데이터 세트들 (252A)의 각각에 대한 세그먼트화 스코어 (254A)를 생성한다. 다른 예로서, 오디오 피쳐 데이터 세트들 (252C)은 사용자 (242B)의 스피치에 대응한다. 세그먼트화기 (124)는, 오디오 피쳐 데이터 세트들 (252C)이 화자 (292B) (예를 들어, 사용자 (242B)에 대한 플레이스홀더)의 스피치에 대응함을 표시하기 위해 세그먼트화 임계치 (257)를 만족시키는 오디오 피쳐 데이터 세트들 (252C)의 각각에 대한 세그먼트화 스코어 (254B)를 생성한다.

[0042] 특정 구현에서, 세그먼트화기 (124)는, 화자 (292A) (예를 들어, 사용자 (242A))의 스피치가 세그먼트화 윈도우의 지속기간 동안 검출되지 않은 때, 예를 들어, 화자 (292A)와 연관된 이전 스피치를 검출한 이후 임계 지속기간이 만료한 때, 화자 (292A) (예를 들어, 세그먼트화 스코어 (254A))를 다른 사용자 (예를 들어, 사용자 (242C))에 대한 플레이스홀더로서 재사용할 수도 있다. 세그먼트화기 (124)는, 화자 플레이스홀더와 연관된 이전 사용자가 세그먼트화 윈도우 동안 말하지 않았을 때 다른 사용자에게 대한 화자 플레이스홀더를 재사용함으로써 오디오 스트림 (141)에서 미리결정된 카운트 초과 화자들 (예를 들어, K 초과 화자들)과 연관된 스피치를 구별할 수 있다. 특정 구현에서, 세그먼트화기 (124)는, 화자 (292A) (예를 들어, 사용자 (242A)), 화자 (292B) (예를 들어, 사용자 (242B)), 및 화자 (292C) (예를 들어, 사용자 (242C))의 각각의 스피치가 세그먼트화 윈도우 내에서 검출됨을 결정하는 것 및 다른 사용자 (예를 들어, 사용자 (242D))와 연관된 스피치가 검출됨을 결정하는 것에 응답하여, 화자 (292A) (예를 들어, 사용자 (242A))의 스피치가 적어도 최근에 검출되었다고 결정하는 것에 기초하여 화자 플레이스홀더 (예를 들어, 화자 (292A))를 재사용한다.

[0043] 특정 양태에서, 세그먼트화기 (124)는 뉴럴 네트워크와 같은 트레이닝된 머신 러닝 시스템을 포함하거나 이에 대응한다. 예를 들어, 오디오 피쳐 데이터 세트들 (252)을 분석하는 것은 스피커 세그먼트화 뉴럴 네트워크 (또는 다른 머신 러닝 기반 시스템)를 오디오 피쳐 데이터 세트들 (252)에 적용하는 것을 포함한다.

[0044] 특정 양태에서, 세그먼트화기 (124)는 세그먼트화 스코어들 (254)에 기초하여 데이터 세트 세그먼트화 결과 (256)를 생성한다. 데이터 세트 세그먼트화 결과 (256)는 오디오 부분 (151)에서 검출되는 화자들 (292) (존재하는 경우)을 표시한다. 예를 들어, 세그먼트화기 (124)에 의해 출력된 데이터 세트 세그먼트화 결과 (256)는, 화자 (292)에 대한 세그먼트화 스코어 (254)가 세그먼트화 임계치 (257)를 만족한다는 (예를 들어, 더 크다는) 결정에 응답하여 화자 (292)의 스피치가 검출됨을 표시한다. 예시하자면, 오디오 피쳐 데이터 세트 (252)의 세그먼트화 스코어 (254A)가 세그먼트화 임계치 (257)를 만족할 때, 세그먼트화기 (124)는 화자 (292A)의 스피치가 오디오 부분 (151)에서 검출됨을 표시하는 오디오 피쳐 데이터 세트 (252)에 대한 데이터 세트 세그먼트화 결과 (256) (예를 들어, "1")를 생성한다. 다른 예에서, 오디오 피쳐 데이터 세트 (252)의 세그먼트화 스코어 (254A) 및 세그먼트화 스코어 (254B)의 각각이 세그먼트화 임계치 (257)를 만족할 때, 세그먼트화기 (124)는 화자 (292A) 및 화자 (292B) (예를 들어, 다중 화자들)의 스피치가 오디오 부분 (151)에서 검출됨을 표시하기 위한 오디오 피쳐 데이터 세트 (252)에 대한 데이터 세트 세그먼트화 결과 (256) (예를 들어, "1, 2")를 생성한다. 특정 예에서, 오디오 피쳐 데이터 세트 (252)에 대한 세그먼트화 스코어 (254A), 세그먼트화 스코어 (254B), 및 세그먼트화 스코어 (254C)의 각각이 세그먼트화

임계치 (257) 를 만족하지 못할 때, 세그먼트화기 (124) 는 묵음 (또는 비-스피치 오디오) 이 오디오 부분 (151) 에서 검출됨을 표시하기 위한 오디오 피쳐 데이터 세트 (252) 에 대한 데이터 세트 세그먼트화 결과 (256) (예를 들어, "0") 를 생성한다. 오디오 부분 (151) (또는 오디오 피쳐 데이터 세트 (252)) 에 대한 세그먼트화 결과 (236) 는 오디오 부분 (151) (또는 오디오 피쳐 데이터 세트 (252)) 에 대한 세그먼트화 스코어들 (254), 데이터 세트 세그먼트화 결과 (256), 또는 그 양자 모두를 포함한다.

[0045] 세그먼트화기 (124) 는 오디오 부분 (151) (예를 들어, 오디오 피쳐 데이터 세트 (252)) 에 대한 세그먼트화 결과 (236) 를 프로파일 관리기 (126) 에 제공하도록 구성된다. 프로파일 관리기 (126) 는, 오디오 피쳐 데이터 세트 (252) 가 복수의 사용자 스피치 프로파일들 (150) 중 임의의 것과 매칭하지 않음을 결정하는 것에 응답하여, 오디오 피쳐 데이터 세트 (252) 에 적어도 부분적으로 기초하여 사용자 스피치 프로파일 (150) 을 생성하도록 구성된다. 특정 양태에서, 프로파일 관리기 (126) 는 화자-동일 오디오 세그먼트 (111) 에 기초하여 사용자 스피치 프로파일 (150) 을 생성하도록 구성된다. 예를 들어, 프로파일 관리기 (126) 는 화자-동일 오디오 세그먼트 (111A) 의 오디오 피쳐 데이터 세그먼트들 (152A) 에 기초하여 화자 (292A) 에 대한 사용자 스피치 프로파일 (150A) (예를 들어, 사용자 (242A) 에 대한 플레이스홀더) 을 생성하도록 구성된다. 사용자 스피치 프로파일 (150A) 은 사용자 (242A) 의 스피치를 나타낸다 (예를 들어, 모델링함). 대안적으로, 프로파일 관리기 (126) 는, 오디오 피쳐 데이터 세트 (252) 가 사용자 스피치 프로파일 (150) 과 매칭함을 결정하는 것에 응답하여, 오디오 피쳐 데이터 세트 (252) 에 기초하여 사용자 스피치 프로파일 (150) 을 업데이트하도록 구성된다. 예를 들어, 프로파일 관리기 (126) 는, 후속 오디오 부분들에 대해 사용자 (242A) 에 대한 플레이스홀더로서 어느 화자 (292) 가 사용되는지와 독립적으로 사용자 스피치 프로파일 (150A) 과 매칭하는 후속 오디오 부분들에 기초하여 사용자 (242A) 의 스피치를 나타내는 사용자 스피치 프로파일 (150A) 을 업데이트하도록 구성된다. 특정 양태에서, 프로파일 관리기 (126) 는 사용자 스피치 프로파일 (150) 을 생성 또는 업데이트하는 것에 응답하여 사용자 스피치 프로파일 (150) 의 프로파일 ID (155) 를 출력한다.

[0046] 특정 구현에서, 화자 검출기 (278) 는 오디오 스트림 (141) 으로부터 추출된 오디오 피쳐들에 기초하여 오디오 스트림 (141) 에서 검출된 화자들의 카운트를 결정하도록 구성된다. 특정 양태에서, 화자 검출기 (278) 는 피쳐 추출기 (222) 에 의해 추출된 오디오 피쳐 데이터 세트들 (252) 에 기초하여 화자들의 카운트를 결정한다. 예를 들어, 화자들의 카운트를 결정하기 위해 화자 검출기 (278) 에 의해 사용되는 오디오 피쳐들은, 세그먼트화 결과 (236) 를 생성하기 위해 세그먼트화기 (124) 에 의해 사용되고 그리고 사용자 스피치 프로파일들 (150) 을 생성 또는 업데이트하기 위해 프로파일 관리기 (126) 에 의해 사용되는 오디오 피쳐들과 동일할 수 있다. 대안적인 양태에서, 화자 검출기 (278) 는 피쳐 추출기 (222) 와는 별개인 제 2 피쳐 추출기에 의해 추출되는 오디오 피쳐들에 기초하여 화자들의 카운트를 결정한다. 이 양태에서, 화자들의 카운트를 결정하기 위해 화자 검출기 (278) 에 의해 사용되는 오디오 피쳐들은, 세그먼트화 결과 (236) 를 생성하기 위해 세그먼트화기 (124) 에 의해 사용되고 그리고 사용자 스피치 프로파일들 (150) 을 생성 또는 업데이트하기 위해 프로파일 관리기 (126) 에 의해 사용되는 오디오 피쳐들과는 별개일 수 있다. 특정 양태에서, 화자 검출기 (278) 는 오디오 스트림 (141) 에서 적어도 2명의 별개의 화자들을 검출하는 것에 응답하여 세그먼트화기 (124) 를 활성화시킨다. 예를 들어, 세그먼트화기 (124) 는, 다중 화자들이 오디오 스트림 (141) 에서 검출될 때 오디오 피쳐 데이터 세트들 (252) 을 프로세싱한다. 대안적으로, 화자 검출기 (278) 가 오디오 스트림 (141) 에서 단일 화자의 스피치를 검출할 때, 세그먼트화기 (124) 는 우회되고 프로파일 관리기 (126) 는 오디오 피쳐 데이터 세트들 (252) 을 프로세싱하여 사용자 스피치 프로파일 (150) 을 생성 또는 업데이트한다.

[0047] 일부 구현들에서, 디바이스 (202) 는 하나 또는 다양한 타입들의 디바이스들에 대응하거나 그에 포함된다. 예시적인 예에서, 하나 이상의 프로세서들 (220) 은, 도 13 을 참조하여 추가로 설명된 바와 같이, 마이크로폰 (246) 을 포함하는 헤드셋 디바이스에 통합된다. 다른 예들에서, 하나 이상의 프로세서들 (220) 은 도 12 를 참조하여 설명된 바와 같은 모바일 폰 또는 태블릿 컴퓨터 디바이스, 도 14 를 참조하여 설명된 바와 같은 웨어러블 전자 디바이스, 도 15 를 참조하여 설명된 바와 같은 음성 제어식 스피커 시스템, 또는 도 16 을 참조하여 설명된 바와 같은 가상 현실 헤드셋 또는 증강 현실 헤드셋 중 적어도 하나에 통합된다. 다른 예시적인 예에서, 하나 이상의 프로세서들 (220) 은, 도 17 및 도 18 을 참조하여 추가로 설명된 바와 같이, 마이크로폰 (246) 을 또한 포함하는 차량에 통합된다.

[0048] 동작 동안, 하나 이상의 프로세서들 (220) 은 하나 이상의 사용자들 (242) (예를 들어, 사용자 (242A), 사용자 (242B), 사용자 (242C), 사용자 (242D) 또는 이들의 조합) 의 스피치에 대응하는 오디오 스트림 (141) 을 수신한다. 특정 예에서, 하나 이상의 프로세서들 (220) 은, 하나 이상의 사용자들의 스피치를 캡처한 마이크로폰 (246) 으로부터 오디오 스트림 (141) 을 수신한다. 다른 예에서, 오디오 스트림 (141) 은 메모리 (232)

에 저장된 오디오 재생 파일에 대응하고, 하나 이상의 프로세서들 (220) 은 메모리 (232) 로부터 오디오 스트림 (141) 을 수신한다. 특정 양태에서, 하나 이상의 프로세서들 (220) 은 다른 디바이스로부터 입력 인터페이스 또는 네트워크 인터페이스 (예를 들어, 모뎀의 네트워크 인터페이스) 를 통해 오디오 스트림 (141) 을 수신한다.

[0049] 피쳐 추출 스테이지 동안, 피쳐 추출기 (222) 는 오디오 스트림 (141) 의 오디오 피쳐 데이터 세트들 (252) 을 생성한다. 예를 들어, 피쳐 추출기 (222) 는 오디오 스트림 (141) 의 오디오 부분들 (151) 의 피쳐들을 결정함으로써 오디오 피쳐 데이터 세트들 (252) 을 생성한다. 특정 예에서, 오디오 스트림 (141) 은 오디오 부분들 (151A), 오디오 부분들 (151B), 오디오 부분들 (151C), 또는 이들의 조합을 포함한다. 피쳐 추출기 (222) 는 오디오 부분들 (151A) 의 피쳐들을 나타내는 오디오 피쳐 데이터 세트들 (252A), 오디오 부분들 (151B) 의 피쳐들을 나타내는 오디오 피쳐 데이터 세트들 (252B), 및 오디오 부분들 (151C) 의 피쳐들을 나타내는 오디오 피쳐 데이터 세트들 (252C), 또는 이들의 조합을 생성한다. 예를 들어, 피쳐 추출기 (222) 는 오디오 부분 (151) 의 오디오 피쳐들을 추출함으로써 오디오 부분 (151) (예컨대, 오디오 프레임) 에 대한 오디오 피쳐 데이터 세트 (252) (예컨대, 피쳐 벡터) 를 생성한다.

[0050] 세그먼트화 스테이지 동안, 세그먼트화기 (124) 는 세그먼트화 결과 (236) 를 생성하기 위해 오디오 피쳐 데이터 세트들 (252) 을 분석한다. 예를 들어, 세그먼트화기 (124) 는 오디오 부분 (151) 에 대한 세그먼트화 스코어들 (254) 을 생성하기 위해 오디오 부분 (151) (예컨대, 오디오 프레임) 의 오디오 피쳐 데이터 세트 (252) (예컨대, 피쳐 벡터) 를 분석한다. 예시하자면, 세그먼트화 스코어들 (254) 은, 오디오 부분 (151) 이 화자 (292A) 의 스피치에 대응할 가능성을 표시하는 세그먼트화 스코어 (254A) (예를 들어, 0.6) 를 포함한다. 세그먼트화 스코어들 (254) 은 또한, 각각, 화자 (292B) 및 화자 (292C) 의 스피치에 대응하는 오디오 부분 (151) 의 가능성을 표시하는 세그먼트화 스코어 (254B) (예를 들어, 0) 및 세그먼트화 스코어 (254C) (예를 들어, 0) 를 포함한다. 특정 양태에서, 세그먼트화기 (124) 는, 세그먼트화 스코어 (254A) 가 세그먼트화 임계치 (257) 를 만족하고 세그먼트화 스코어 (254B) 및 세그먼트화 스코어 (254C) 의 각각이 세그먼트화 임계치 (257) 를 만족하지 못함을 결정하는 것에 응답하여, 오디오 부분 (151) 이 화자 (292A) 의 스피치에 대응하고 화자 (292B) 또는 화자 (292C) 중 어느 하나의 스피치에 대응하지 않음을 표시하는 데이터 세트 세그먼트화 결과 (256) 를 생성한다. 세그먼트화기 (124) 는 오디오 부분 (151) 에 대한 세그먼트화 스코어들 (254), 데이터 세트 세그먼트화 결과 (256), 또는 그 양자 모두를 표시하는 세그먼트화 결과 (236) 를 생성한다.

[0051] 특정 예에서, 세그먼트화 스테이지 동안, 세그먼트화기 (124) 는, 다중의 세그먼트화 스코어들 (예를 들어, 세그먼트화 스코어 (254A) 및 세그먼트화 스코어 (254B)) 의 각각이 세그먼트화 임계치 (257) 를 만족함을 결정하는 것에 응답하여, 오디오 부분 (151) 이 다중 화자들 (예를 들어, 화자 (292A) 및 화자 (292B)) 의 스피치에 대응함을 표시하는 세그먼트화 결과 (236) 를 생성한다.

[0052] 프로파일 관리기 (126) 는, 도 2b 를 참조하여 추가로 설명된 바와 같이, 세그먼트화 결과 (236) 에 기초하여 오디오 부분 (151) (예를 들어, 오디오 피쳐 데이터 세트 (252)) 을 프로세싱한다. 도 2b 에서, 메모리 (232) 는 등록 버퍼들 (234), 프로브 버퍼들 (240), 또는 이들의 조합을 포함한다. 예를 들어, 메모리 (232) 는 화자들 (292) 의 각각에 대해 지정된 등록 버퍼 (234) 및 프로브 버퍼 (240) 를 포함한다. 예시하자면, 메모리 (232) 는 화자 (292A) 에 대해 지정된 등록 버퍼 (234A) 및 프로브 버퍼 (240A), 화자 (292B) 에 대해 지정된 등록 버퍼 (234B) 및 프로브 버퍼 (240B), 및 화자 (292C) 에 대해 지정된 등록 버퍼 (234C) 및 프로브 버퍼 (240C) 를 포함한다. 메모리 (232) 는 등록 임계치 (264), 프로파일 임계치 (258), 묵음 임계치 (294), 또는 이들의 조합을 저장하도록 구성된다. 메모리 (232) 는 정지 조건 (270), 스피치 프로파일 결과 (238), 묵음 카운트 (262) (도 2b 에서의 "Sil 카운트"), 또는 이들의 조합을 표시하는 데이터를 저장하도록 구성된다.

[0053] 프로파일 관리기 (126) 는, 프로파일 체크 스테이지 동안, 오디오 피쳐 데이터 세트 (252) 가 기존의 사용자 스피치 프로파일 (150) 과 매칭하는지 여부를 결정하도록 구성된다. 특정 양태에서, 프로파일 관리기 (126) 는 세그먼트화 결과 (236) 를 생성하기 위해 세그먼트화기 (124) 에 의해 사용된 오디오 피쳐들과 동일한 오디오 피쳐들을, 사용자 스피치 프로파일들 (150) 과의 비교 또는 업데이트를 위해 사용한다. 다른 양태에서, 프로파일 관리기 (126) 는 세그먼트화 결과 (236) 를 생성하기 위해 세그먼트화기 (124) 에 의해 사용된 오디오 피쳐들과는 별개인 제 2 오디오 피쳐들을, 사용자 스피치 프로파일들 (150) 과의 비교 또는 업데이트를 위해 사용한다.

- [0054] 특정 구현에서, 프로파일 관리기 (126) 는 비교의 정확도를 개선하기 위해 사용자 스피치 프로파일들 (150) 과 비교하기 전에 프로브 버퍼 (240) 에서 동일한 화자에 대응하는 오디오 피쳐 데이터 세트들 (252) 을 수집하도록 구성된다. 오디오 피쳐 데이터 세트 (252) 가 기존의 사용자 스피치 프로파일과 매칭하면, 프로파일 관리기 (126) 는, 업데이트 스테이지 동안, 오디오 피쳐 데이터 세트 (252) 에 기초하여 기존의 사용자 스피치 프로파일을 업데이트하도록 구성된다. 오디오 피쳐 데이터 세트 (252) 가 기존의 사용자 스피치 프로파일과 매칭하지 않으면, 프로파일 관리기 (126) 는, 등록 스테이지 동안, 등록 버퍼 (234) 에 오디오 피쳐 데이터 세트 (252) 를 추가하고, 등록 버퍼 (234) 에 저장된 오디오 피쳐 데이터 세트들 (252) 이 등록 임계치 (264) 를 만족함을 결정하는 것에 응답하여, 등록 버퍼 (234) 에 저장된 오디오 피쳐 데이터 세트들 (252) 에 기초하여 사용자 스피치 프로파일 (150) 을 생성하도록 구성된다.
- [0055] 프로파일 체크 스테이지 동안, 프로파일 관리기 (126) 는, 사용자 스피치 프로파일들이 이용가능하지 않다고 그리고 오디오 부분 (151) 이 화자 (예를 들어, 화자 (292A)) 의 스피치에 대응함을 세그먼트화 결과 (236) 가 표시한다고 결정하는 것에 응답하여, 오디오 피쳐 데이터 세트 (252) 를, 화자 (292) 에 대해 지정된 등록 버퍼 (234) (예를 들어, 등록 버퍼 (234A)) 에 추가하고 등록 스테이지로 진행한다.
- [0056] 특정 양태에서, 프로파일 관리기 (126) 는, 적어도 하나의 사용자 스피치 프로파일 (150) 이 이용가능하다고 결정하는 것에 응답하여, 오디오 피쳐 데이터 세트 (252) 가 적어도 하나의 사용자 스피치 프로파일 (150) 중 임의의 것과 매칭하는지 여부를 결정하기 위해 적어도 하나의 사용자 스피치 프로파일 (150) 과의 오디오 피쳐 데이터 세트 (252) 의 비교를 수행한다. 프로파일 관리기 (126) 는, 적어도 하나의 사용자 스피치 프로파일 (150) 이 이용가능하다고 그리고 오디오 부분 (151) 이 화자 (292) (예를 들어, 화자 (292A)) 의 스피치에 대응함을 세그먼트화 결과 (236) 가 표시한다고 결정하는 것에 응답하여, 오디오 피쳐 데이터 세트 (252) 를, 화자 (292) 에 대해 지정된 프로브 버퍼 (240) (예를 들어, 프로브 버퍼 (240A)) 에 추가한다.
- [0057] 프로파일 관리기 (126) 는, 프로브 버퍼 (240) 에 저장된 오디오 피쳐 데이터 세트들 (예를 들어, 오디오 피쳐 데이터 세트 (252) 를 포함) 이 적어도 하나의 사용자 스피치 프로파일 (150) 중 임의의 것과 매칭하는지 여부를 결정한다. 예를 들어, 프로파일 관리기 (126) 는, 프로브 버퍼 (240) (예컨대, 프로브 버퍼 (240A)) 의 오디오 피쳐 데이터 세트들 (예컨대, 오디오 피쳐 데이터 세트 (252) 를 포함) 과 적어도 하나의 사용자 스피치 프로파일 (150) 의 각각의 비교에 기초하여 스피치 프로파일 결과들 (238) 을 생성한다. 예시하자면, 프로파일 관리기 (126) 는, 프로브 버퍼 (240) (예컨대, 프로브 버퍼 (240A)) 의 오디오 피쳐 데이터 세트들 (예컨대, 오디오 피쳐 데이터 세트 (252) 를 포함) 과 사용자 스피치 프로파일 (150A) 의 비교에 기초하여 스피치 프로파일 결과 (238A) 를 생성한다.
- [0058] 특정 양태에서, 프로파일 관리기 (126) 는, 단일 오디오 피쳐 데이터 세트 (예컨대, 오디오 피쳐 데이터 세트 (252)) 가 프로브 버퍼 (240) (예컨대, 프로브 버퍼 (240A)) 에서 이용가능하다고 결정하는 것에 응답하여, 단일 오디오 피쳐 데이터 세트와 사용자 스피치 프로파일 (150A) 의 비교에 기초하여 스피치 프로파일 결과 (238A) 를 생성한다. 대안적으로, 프로파일 관리기 (126) 는, 다중 오디오 피쳐 데이터 세트들 (예컨대, 오디오 피쳐 데이터 세트 (252) 를 포함) 이 프로브 버퍼 (240) (예컨대, 프로브 버퍼 (240A)) 에서 이용가능하다고 결정하는 것에 응답하여, 다중 오디오 피쳐 데이터 세트들과 사용자 스피치 프로파일 (150A) 의 비교들에 기초하여 스피치 프로파일 결과 (238A) 를 생성한다. 예를 들어, 프로파일 관리기 (126) 는 오디오 피쳐 데이터 세트 (252) 와 사용자 스피치 프로파일 (150A) 의 비교에 기초한 제 1 데이터 세트 결과, 프로브 버퍼 (240) 의 제 2 오디오 피쳐 데이터 세트와 사용자 스피치 프로파일 (150A) 의 비교에 기초한 제 2 데이터 세트 결과, 프로브 버퍼 (240) 의 추가적인 오디오 피쳐 데이터 세트들과 사용자 스피치 프로파일 (150A) 의 비교들에 기초한 추가적인 데이터 세트 결과들, 또는 이들의 조합을 생성한다. 프로파일 관리기 (126) 는 제 1 데이터 세트 결과, 제 2 데이터 세트 결과, 추가적인 데이터 세트 결과들, 또는 이들의 조합에 기초하여 (예를 들어, 가중 평균하여) 스피치 프로파일 결과 (238A) 를 생성한다. 특정 양태에서, 프로브 버퍼 (240) 에 더 최근에 추가되는 오디오 피쳐 데이터 세트들의 데이터 세트 결과들에 더 높은 가중치들이 배정된다.
- [0059] 스피치 프로파일 결과 (238A) 는, 오디오 피쳐 데이터 세트들이 사용자 스피치 프로파일 (150A) 과 매칭할 가능성을 표시한다. 유사하게, 프로파일 관리기 (126) 는, 프로브 버퍼 (240) (예컨대, 프로브 버퍼 (240A)) 의 오디오 피쳐 데이터 세트들 (예컨대, 오디오 피쳐 데이터 세트 (252) 를 포함) 과 사용자 스피치 프로파일 (150B) 의 비교에 기초하여 스피치 프로파일 결과 (238B) 를 생성한다.
- [0060] 특정 양태에서, 프로파일 관리기 (126) 는, 오디오 피쳐 데이터 세트 (252) 가 대응하는 사용자 스피치 프로파일 (150) 과 매칭할 가장 높은 가능성을 표시하는 스피치 프로파일 결과 (238) 를 선택한다. 예를 들어, 프

로파일 관리기 (126) 는, 스피치 프로파일 결과 (238A) 가 스피치 프로파일 결과 (238B) 보다 더 높은 (예컨대, 그 이상인) 매칭 가능성을 표시함을 결정하는 것에 응답하여 스피치 프로파일 결과 (238A) 를 선택한다. 프로파일 관리기 (126) 는, 스피치 프로파일 결과 (238A) (예컨대, 가장 높은 매칭 가능성을 표시하는 스피치 프로파일 결과 (238A)) 가 프로파일 임계치 (258) 를 만족함 (예컨대, 그 이상임) 을 결정하는 것에 응답하여, 프로브 버퍼 (240) (예컨대, 프로브 버퍼 (240A)) 에 저장된 오디오 피쳐 데이터 세트들이 사용자 스피치 프로파일 (150A) 과 매칭함을 결정하고 업데이트 스테이지로 진행한다. 대안적으로, 프로파일 관리기 (126) 는, 스피치 프로파일 결과 (238A) (예컨대, 가장 높은 매칭 가능성을 표시하는 스피치 프로파일 결과 (238A)) 가 프로파일 임계치 (258) 를 만족하지 못함 (예컨대, 그 미만임) 을 결정하는 것에 응답하여, 프로브 버퍼 (240) (예컨대, 프로브 버퍼 (240A)) 에 저장된 오디오 피쳐 데이터 세트들이 사용자 스피치 프로파일들 (150) 중 임의의 것과 매칭하지 않음을 결정하고 등록 스테이지로 진행한다.

[0061] 업데이트 스테이지 동안, 프로파일 관리기 (126) 는, 오디오 피쳐 데이터 세트 (252) 가 사용자 스피치 프로파일 (150) (예컨대, 사용자 스피치 프로파일 (150A)) 과 매칭함을 결정하는 것에 응답하여, 사용자 스피치 프로파일 (150) 을 업데이트하고 사용자 스피치 프로파일 (150) 의 프로파일 ID (155) 를 출력한다. 프로파일 관리기 (126) 는, 프로브 버퍼 (240) 에 저장된 오디오 피쳐 데이터 세트들에 기초하여 (프로브 버퍼 (240) 에 저장된 오디오 피쳐 데이터 세트들과 매칭된) 사용자 스피치 프로파일 (150) 을 업데이트한다. 따라서, 사용자 스피치 프로파일 (150A) 은 사용자 스피치에서의 변경들과 매칭하기 위해 시간에 걸쳐 진화한다.

[0062] 등록 스테이지 동안, 프로파일 관리기 (126) 는, 오디오 피쳐 데이터 세트 (252) 가 화자 (292) (예를 들어, 화자 (292A)) 의 스피치를 나타낸다는 것을 세그먼트화 결과 (236) 가 표시함을 결정하는 것에 응답하여, 오디오 피쳐 데이터 세트 (252) 를, 화자 (292) 에 대응하는 등록 버퍼 (234) (예를 들어, 등록 버퍼 (234A)) 에 추가한다. 프로파일 관리기 (126) 는, 등록 버퍼 (234) 에 저장된 오디오 피쳐 데이터 세트들이 등록 임계치 (264) 를 만족하는지 여부를 결정한다. 특정 양태에서, 프로파일 관리기 (126) 는, 오디오 피쳐 데이터 세트들의 카운트가 등록 임계치 (264) (예컨대, 48개의 오디오 피쳐 데이터 세트들) 이상임을 결정하는 것에 응답하여, 등록 버퍼 (234) 에 저장된 오디오 피쳐 데이터 세트들이 등록 임계치 (264) 를 만족함을 결정한다. 다른 양태에서, 프로파일 관리기 (126) 는, 오디오 피쳐 데이터 세트들의 스피치 지속기간 (예컨대, 재생 지속기간) 이 등록 임계치 (264) (예컨대, 2 초) 이상임을 결정하는 것에 응답하여, 등록 버퍼 (234) 에 저장된 오디오 피쳐 데이터 세트들이 등록 임계치 (264) 를 만족함을 결정한다.

[0063] 프로파일 관리기 (126) 는, 등록 버퍼 (234) 에 저장된 오디오 피쳐 데이터 세트들이 등록 임계치 (264) 를 만족하지 못함을 결정하는 것에 응답하여, 등록 버퍼 (234) 에 저장된 오디오 피쳐 데이터 세트들에 기초하여 사용자 스피치 프로파일 (150) 을 생성하는 것을 억제하고, 오디오 스트림 (141) 의 후속 오디오 부분들을 계속 프로세싱한다. 특정 양태에서, 프로파일 관리기 (126) 는, 정지 조건 (270) 이 만족될 때까지 화자 (292) (예를 들어, 화자 (292A)) 의 스피치를 나타내는 후속 오디오 피쳐 데이터 세트들을 등록 버퍼 (234) (예를 들어, 등록 버퍼 (234A)) 에 계속 추가한다. 예를 들어, 프로파일 관리기 (126) 는, 본 명세서에서 설명된 바와 같이, 등록 버퍼 (234) 에 저장된 오디오 피쳐 데이터 세트들 (예컨대, 오디오 피쳐 데이터 세트 (252) 를 포함) 의 카운트가 등록 임계치 (264) 를 만족함을 결정하는 것, 임계 묶음보다 더 긴 것이 오디오 스트림 (141) 에서 검출됨을 결정하는 것, 또는 그 양자 모두에 응답하여 정지 조건 (270) 이 만족됨을 결정한다. 예시하자면, 정지 조건 (270) 은, 충분한 오디오 피쳐 데이터 세트들이 사용자 스피치 프로파일을 생성하기 위해 등록 버퍼 (234) 에 있을 때 또는 화자 (292) 가 말하기를 정지한 것으로 보일 때, 만족된다.

[0064] 특정 양태에서, 프로파일 관리기 (126) 는, 등록 버퍼 (234) 에 저장된 오디오 피쳐 데이터 세트들 (예컨대, 오디오 피쳐 데이터 세트 (252) 를 포함) 이 등록 임계치 (264) 를 만족함을 결정하는 것에 응답하여, 등록 버퍼 (234) 에 저장된 오디오 피쳐 데이터 세트들에 기초하여 사용자 스피치 프로파일 (150C) 을 생성하고, 등록 버퍼 (234) 를 리셋하고, 사용자 스피치 프로파일 (150C) 을 복수의 사용자 스피치 프로파일들 (150) 에 추가하고, 사용자 스피치 프로파일 (150C) 의 프로파일 ID (155) 를 출력하고, 오디오 스트림 (141) 의 후속 오디오 부분들을 계속 프로세싱한다. 따라서, 프로파일 관리기 (126) 는, 화자 (292) (예컨대, 화자 (292A)) 에 대해 지정된 등록 버퍼 (234) (예컨대, 등록 버퍼 (234A)) 에 저장되는 동일한 화자 (292) (예컨대, 화자 (292A)) 에 대응하는 오디오 부분들의 오디오 피쳐 데이터 세트들에 기초하여 사용자 스피치 프로파일 (150C) 을 생성한다. 사용자 스피치 프로파일 (150C) 을 생성하기 위해 다중의 오디오 피쳐 데이터 세트들을 사용하는 것은 화자 (292A) (예컨대, 사용자 (242A)) 의 스피치를 표현함에 있어서 사용자 스피치 프로파일 (150A) 의 정확도를 개선한다. 따라서, 세그먼트화기 (124) 및 프로파일 관리기 (126) 는, 사용자 스피치 프로파일 생성을 위해 미리결정된 단어들 또는 문장들을 말해야 하는 것은 아니고 그리고 사전 등록되어야 하는 것은 아

닌 사용자들에 대한 사용자 스피치 프로파일들을 생성함으로써 다중의 사용자들의 수동 등록을 가능케 한다.

[0065] 특정 양태에서, 다중 화자들에 대응하는 오디오 부분들은 사용자 스피치 프로파일들 (150) 을 생성 또는 업데이트하기 위해 스킵되거나 무시된다. 예를 들어, 프로파일 관리기 (126) 는, 오디오 부분 (151) 이 다중 화자들의 스피치에 대응한다는 것을 세그먼트화 결과 (236) 가 표시함을 결정하는 것에 응답하여, 오디오 부분 (151) 의 오디오 피쳐 데이터 세트 (252) 를 무시하고 오디오 스트림 (141) 의 후속 오디오 부분들을 계속 프로세싱한다. 예를 들어, 오디오 피쳐 데이터 세트 (252) 를 무시하는 것은 오디오 피쳐 데이터 세트 (252) 를 복수의 사용자 스피치 프로파일들 (150) 과 비교하는 것을 억제하는 것, 오디오 피쳐 데이터 세트 (252) 에 기초하여 사용자 스피치 프로파일 (150) 을 업데이트하는 것을 억제하는 것, 오디오 피쳐 데이터 세트 (252) 에 기초하여 사용자 스피치 프로파일 (150) 을 생성하는 것을 억제하는 것, 또는 이들의 조합을 포함한다.

[0066] 특정 양태에서, (예를 들어, 동일한 사용자의 스피치에서 자연적인 짧은 일시중지들을 나타내는) 임계 묵음보다 짧은 것에 대응하는 오디오 부분들은 사용자 스피치 프로파일 (150) 을 생성 또는 업데이트하는데 사용되지 않지만, 임계 묵음보다 긴 것을 검출하도록 추적된다. 예를 들어, 세그먼트화 스테이지 동안, 세그먼트화기 (124) 는, 오디오 부분 (151) 이 묵음에 대응함을 표시하는 오디오 피쳐 데이터 세트 (252) 에 대한 세그먼트화 결과 (236) 를 생성한다. 프로파일 관리기 (126) 는, 오디오 부분 (151) 이 묵음에 대응함을 결정하는 것에 응답하여 묵음 카운트 (262) 를 (예를 들어, 1 만큼) 증분시킨다. 특정 양태에서, 프로파일 관리기 (126) 는, 묵음 카운트 (262) 가 묵음 임계치 (294) 이상이라고 결정하는 것 (예를 들어, 사용자가 말하기를 종료한 이후 더 긴 일시중지를 나타냄) 에 응답하여, 등록 버퍼들 (234) (예를 들어, 등록 버퍼 (234A), 등록 버퍼 (234B), 및 등록 버퍼 (234C)) 을 리셋 (예를 들어, EMPTY (empty) 로서 마킹) 하고, 프로브 버퍼들 (240) (예를 들어, 프로브 버퍼 (240A), 프로브 버퍼 (240B), 및 프로브 버퍼 (240C)) 을 리셋 (예를 들어, EMPTY 로서 마킹) 하고, 묵음 카운트 (262) 를 (예를 들어, 0 으로) 리셋하고, 또는 이들의 조합을 실시하고, 오디오 스트림 (141) 의 후속 오디오 부분들을 계속 프로세싱한다. 특정 양태에서, 프로파일 관리기 (126) 는, 묵음 카운트 (262) 가 묵음 임계치 (294) 이상이라고 결정하는 것에 응답하여 정지 조건 (270) 이 만족됨을 결정한다. 프로파일 관리기 (126) 는, 정지 조건 (270) 이 만족됨을 결정하는 것에 응답하여, 등록 버퍼들 (234) (예컨대, 등록 버퍼 (234A), 등록 버퍼 (234B), 및 등록 버퍼 (234C)) 을 리셋한다.

[0067] 특정 양태에서, 프로파일 관리기 (126) 는 디바이스 (202) 에 커플링된 디스플레이 디바이스에 통지를 제공한다. 그 통지는 사용자 스피치 분석이 진행 중임을 표시한다. 특정 양태에서, 프로파일 관리기 (126) 는, 사용자 스피치 분석이 수행되어야 하는지 여부를 표시하는 사용자 입력에 기초하여 오디오 스트림 (141) 을 선택적으로 프로세싱한다.

[0068] 도 2a 로 돌아가면, 특정 양태에서, 프로파일 관리기 (126) 는, 얼마나 많은 사용자 스피치 프로파일들 (150) 이 오디오 스트림 (141) 의 프로세싱 동안 생성 또는 업데이트되는지를 추적하기 위해 프로파일 업데이트 데이터 (272) 를 유지한다. 예를 들어, 프로파일 관리기 (126) 는, 사용자 스피치 프로파일 (150) 을 업데이트 (또는 생성) 하는 것에 응답하여 프로파일 업데이트 데이터 (272) 를 업데이트한다. 특정 예에서, 프로파일 관리기 (126) 는, 사용자 스피치 프로파일 (150A) 을 업데이트하는 것에 응답하여, 사용자 스피치 프로파일 (150A) 이 업데이트됨을 표시하기 위해 프로파일 업데이트 데이터 (272) 를 업데이트한다. 다른 예로서, 프로파일 관리기 (126) 는, 사용자 스피치 프로파일 (150C) 을 생성하는 것에 응답하여, 사용자 스피치 프로파일 (150C) 이 업데이트됨을 표시하기 위해 프로파일 업데이트 데이터 (272) 를 업데이트한다. 프로파일 관리기 (126) 는, 복수의 사용자 스피치 프로파일들 (150) 의 제 1 카운트가 오디오 스트림 (141) 의 프로세싱 동안 업데이트되었음을 프로파일 업데이트 데이터 (272) 가 표시함을 결정하는 것에 응답하여, 제 1 카운트를, 오디오 스트림 (141) 에서 검출된 화자들의 카운트로서 출력한다.

[0069] 특정 양태에서, 프로파일 관리기 (126) 는, 복수의 사용자 스피치 프로파일들 (150) 의 각각과 매칭하는 검출된 스피치의 지속기간을 추적하기 위해 사용자 상호작용 데이터 (274) 를 유지한다. 프로파일 관리기 (126) 는, 사용자 스피치 프로파일 (150) 을 업데이트 (또는 생성) 하는 것에 기초하여 사용자 상호작용 데이터 (274) 를 업데이트한다. 예를 들어, 프로파일 관리기 (126) 는, 오디오 부분 (151) 에 기초하여 사용자 스피치 프로파일 (150A) 을 업데이트하는 것에 응답하여, 사용자 스피치 프로파일 (150A) 과 연관된 사용자가 오디오 부분 (151) 의 스피치 지속기간 동안 상호작용하였음을 표시하기 위해 사용자 상호작용 데이터 (274) 를 업데이트한다. 다른 예로서, 프로파일 관리기 (126) 는, 오디오 부분 (151) 에 기초하여 사용자 스피치 프로파일 (150C) 을 생성하는 것에 응답하여, 사용자 스피치 프로파일 (150C) 과 연관된 사용자가 오디오 부분 (151) 의 스피치 지속기간 동안 상호작용하였음을 표시하기 위해 사용자 상호작용 데이터 (274) 를 업데이트한다. 예시하자면, 화자-동일 오디오 세그먼트 (111) 의 오디오 부분들에 기초하여 사용자 스피치 프로파일 (150) 을 생

성 또는 업데이트한 이후, 사용자 상호작용 데이터 (274) 는, 사용자 스피치 프로파일 (150) 과 연관된 사용자가 화자-동일 오디오 세그먼트 (111) 의 스피치 지속기간 동안 상호작용하였음을 표시한다. 특정 양태에서, 프로파일 관리기 (126) 는 사용자 상호작용 데이터 (274) 를 출력한다.

[0070] 특정 양태에서, 프로파일 관리기 (126) 는 프로파일 ID (155), 프로파일 업데이트 데이터 (272), 사용자 상호작용 데이터 (274), 추가적인 정보, 또는 이들의 조합을 하나 이상의 오디오 분석 어플리케이션들 (180) 에 제공한다. 예를 들어, 오디오 분석 어플리케이션 (180) 은, 오디오 스트림 (141) 의 트랜스크립트를 생성하기 위해 오디오 피쳐 데이터 세트들 (252) 에 대해 스피치-대-텍스트 변환을 수행한다. 오디오 분석 어플리케이션 (180) 은, 오디오 피쳐 데이터 세트 (252) 에 대해 프로파일 관리기 (126) 로부터 수신된 프로파일 ID (155) 에 기초하여 트랜스크립트에서의 오디오 피쳐 데이터 세트 (252) 에 대응하는 텍스트를 라벨링한다.

[0071] 특정 양태에서, 하나 이상의 프로세서들 (220) 은 다중의 전력 모드들 중 하나에서 동작하도록 구성된다. 예를 들어, 하나 이상의 프로세서들 (220) 은 전력 모드 (282) (예컨대, 상시 전력 모드) 또는 전력 모드 (284) (예컨대, 주문형 전력 모드) 에서 동작하도록 구성된다. 특정 양태에서, 전력 모드 (282) 는 전력 모드 (284) 에 비해 더 낮은 전력 모드이다. 예를 들어, 하나 이상의 프로세서들 (220) 은 (전력 모드 (284) 에 비해) 전력 모드 (282) 에서 동작함으로써 에너지를 보존하고, 전력 모드 (282) 에서 동작하지 않는 컴포넌트들을 활성화하기 위해 필요에 따라 전력 모드 (284) 로 트랜지션시킨다.

[0072] 특정 예에서, 디바이스 (202) 의 기능들 중 일부는 전력 모드 (284) 에서 활성이지만, 전력 모드 (282) 에서는 활성이 아니다. 예를 들어, 화자 검출기 (278) 는 전력 모드 (282) 에서 그리고 전력 모드 (284) 에서 활성화될 수 있다. 이 예에서, 피쳐 추출기 (222), 세그먼트화기 (124), 프로파일 관리기 (126), 하나 이상의 오디오 분석 어플리케이션들 (180), 또는 이들의 조합은 전력 모드 (282) 에서가 아닌 전력 모드 (284) 에서 활성화될 수 있다. 오디오 스트림 (141) 이 단일 화자의 스피치에 대응할 때, 세그먼트화기 (124) 는 상이한 화자들에 대응하는 오디오 부분들 사이를 구별하기 위해 사용되어야 하는 것은 아니다. 세그먼트화기 (124) 가 사용되어야 하는 것은 아닐 때 전력 모드 (282) 에 머무르는 것 (또는 전력 모드 (282) 로 트랜지션하는 것) 은 전체 리소스 소비를 감소시킨다. 화자 검출기 (278) 는, 전력 모드 (282) 에서, 오디오 스트림 (141) 이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하도록 구성된다. 하나 이상의 프로세서들 (220) 은, 오디오 스트림 (141) 이 적어도 2명의 별개의 화자들의 스피치에 대응한다는 것을 화자 검출기 (278) 의 출력이 표시함을 결정하는 것에 응답하여, 전력 모드 (282) 로부터 전력 모드 (284) 로 트랜지션시키고 세그먼트화기 (124) 를 활성화시키도록 구성된다. 예를 들어, 세그먼트화기 (124) 는, 전력 모드 (284) 에서, 세그먼트화 결과 (236) 를 생성하기 위해 오디오 피쳐 데이터 세트들 (252) 을 분석한다.

[0073] 특정 예에서, 화자 검출기 (278) 및 프로파일 관리기 (126) 는 전력 모드 (282) 에서 그리고 전력 모드 (284) 에서 활성화될 수 있다. 이 예에서, 피쳐 추출기 (222), 세그먼트화기 (124), 하나 이상의 오디오 분석 어플리케이션들 (180), 또는 이들의 조합은 전력 모드 (282) 에서가 아닌 전력 모드 (284) 에서 활성화될 수 있다. 예를 들어, 단일 화자가 검출됨을 표시하는 화자 검출기 (278) 의 출력에 응답하여, 하나 이상의 프로세서들 (220) 은 전력 모드 (282) 에서 유지하거나 전력 모드 (282) 로 트랜지션시킨다. 전력 모드 (282) 에서, 프로파일 관리기 (126) 는 오디오 피쳐 데이터 세트들 (252) 에 기초하여 단일 화자의 사용자 스피치 프로파일 (150) 을 생성 또는 업데이트한다. 대안적으로, 하나 이상의 프로세서들 (220) 은, 오디오 스트림 (141) 이 적어도 2명의 별개의 화자들의 스피치에 대응함을 표시하는 화자 검출기 (278) 의 출력에 응답하여, 전력 모드 (282) 로부터 전력 모드 (284) 로 트랜지션시키고 세그먼트화기 (124) 를 활성화시킨다. 예를 들어, 세그먼트화기 (124) 는, 전력 모드 (284) 에서, 세그먼트화 결과 (236) 를 생성하기 위해 오디오 피쳐 데이터 세트들 (252) 을 분석한다.

[0074] 특정 예에서, 피쳐 추출기 (222), 화자 검출기 (278), 세그먼트화기 (124), 또는 이들의 조합은 전력 모드 (282) 에서 그리고 전력 모드 (284) 에서 활성화될 수 있다. 이 예에서, 프로파일 관리기 (126), 하나 이상의 오디오 분석 어플리케이션들 (180), 또는 이들의 조합은 전력 모드 (282) 에서가 아닌 전력 모드 (284) 에서 활성화될 수 있다. 특정 양태에서, 하나 이상의 프로세서들 (220) 은, 오디오 스트림 (141) 이 적어도 2명의 별개의 화자들의 스피치에 대응한다는 것을 세그먼트화 결과 (236) 가 표시함을 결정하는 것에 응답하여, 전력 모드 (282) 로부터 전력 모드 (284) 로 트랜지션시키고 프로파일 관리기 (126), 하나 이상의 오디오 분석 어플리케이션들 (180), 또는 이들의 조합을 활성화시키도록 구성된다. 예를 들어, 프로파일 관리기 (126) 는, 전력 모드 (284) 에서, 복수의 사용자 스피치 프로파일들 (150) 에 대한 오디오 피쳐 데이터 세트 (252) 의 비교를 수행한다.

- [0075] 특정 양태에서, 하나 이상의 프로세서들 (220) 은, 오디오 스트림 (141) 이 적어도 2명의 별개의 화자들의 스피치에 대응한다는 것을 세그먼트화 결과 (236) 가 표시함을 결정하는 것에 응답하여, 전력 모드 (284) 에서 오디오 스트림 (141) 의 후속 오디오 부분들을 프로세싱한다. 예를 들어, 피쳐 추출기 (222), 세그먼트화기 (124), 또는 그 양자 모두는 후속 오디오 부분들을 프로세싱하기 위해 전력 모드 (284) 에서 동작한다. 특정 양태에서, 피쳐 추출기 (222), 화자 검출기 (278), 세그먼트화기 (124), 또는 이들의 조합은 전력 모드 (282) 에서 오디오 스트림 (141) 의 오디오 정보를 결정하고, 전력 모드 (284) 에서 오디오 정보를 하나 이상의 오디오 분석 어플리케이션들 (180) 에 제공한다. 오디오 정보는 오디오 스트림 (141) 에서 표시된 화자들의 카운트, 음성 활성화도 검출 (VAD) 정보, 또는 그 양자 모두를 포함한다.
- [0076] 특정 구현에서, 오디오 스트림 (141) 의 하나 이상의 부분들, 오디오 피쳐 데이터 세트들 (252), 또는 이들의 조합은 버퍼 (268) 에 저장되고, 하나 이상의 프로세서들 (220) 은 버퍼 (268) 로부터 오디오 스트림 (141) 의 하나 이상의 부분들, 오디오 피쳐 데이터 세트들 (252), 또는 이들의 조합에 액세스한다. 예를 들어, 하나 이상의 프로세서들 (220) 은 오디오 부분 (151) 을 버퍼 (268) 에 저장한다. 피쳐 추출기 (222) 는 버퍼 (268) 로부터 오디오 부분 (151) 을 추출하고, 오디오 피쳐 데이터 세트 (252) 를 버퍼 (268) 에 저장한다. 세그먼트화기 (124) 는 버퍼 (268) 로부터 오디오 피쳐 데이터 세트 (252) 를 추출하고, 오디오 피쳐 데이터 세트 (252) 의 세그먼트화 스코어들 (254), 데이터 세트 세그먼트화 결과 (256), 또는 이들의 조합을 버퍼 (268) 에 저장한다. 프로파일 관리기 (126) 는 버퍼 (268) 로부터 오디오 피쳐 데이터 세트 (252), 세그먼트화 스코어들 (254), 데이터 세트 세그먼트화 결과 (256), 또는 이들의 조합을 추출한다. 특정 양태에서, 프로파일 관리기 (126) 는 프로파일 ID (155), 프로파일 업데이트 데이터 (272), 사용자 상호작용 데이터 (274), 또는 이들의 조합을 버퍼 (268) 에 저장한다. 특정 양태에서, 하나 이상의 오디오 분석 어플리케이션들 (180) 은 버퍼 (268) 로부터 프로파일 ID (155), 프로파일 업데이트 데이터 (272), 사용자 상호작용 데이터 (274), 또는 이들의 조합을 추출한다.
- [0077] 따라서, 시스템 (200) 은 다중 화자들에 대한 수동 사용자 스피치 프로파일 등록 및 업데이트를 가능케 한다. 예를 들어, 복수의 사용자 스피치 프로파일들 (150) 은, 사용자들 (242) 이 스크립트로부터의 미리결정된 단어들 또는 문장들을 말해야 할 필요 없이 디바이스 (202) 의 정규 동작 동안 백그라운드에서 생성 및 업데이트될 수 있다.
- [0078] 마이크로폰 (246) 이 디바이스 (202) 에 커플링되는 것으로서 예시되지만, 다른 구현들에서, 마이크로폰 (246) 은 디바이스 (202) 에 통합될 수도 있다. 단일 마이크로폰 (246) 이 예시되지만, 다른 구현들에서, 사용자 스피치를 캡처하도록 구성된 하나 이상의 추가적인 마이크로폰들 (146) 이 포함될 수도 있다.
- [0079] 시스템 (200) 이 단일 디바이스 (202) 를 포함하는 것으로서 예시되지만, 다른 구현들에서, 디바이스 (202) 에서 수행되는 것으로서 설명된 동작들은 다중의 디바이스들 사이에 분산될 수도 있다. 예를 들어, 피쳐 추출기 (222), 화자 검출기 (278), 세그먼트화기 (124), 프로파일 관리기 (126), 또는 하나 이상의 오디오 분석 어플리케이션들 (180) 중 하나 이상에 의해 수행되는 것으로서 설명된 동작들은 디바이스 (202) 에서 수행될 수도 있고, 피쳐 추출기 (222), 화자 검출기 (278), 세그먼트화기 (124), 프로파일 관리기 (126), 또는 하나 이상의 오디오 분석 어플리케이션들 (180) 중 다른 것들에 의해 수행되는 것으로서 설명된 동작들은 제 2 디바이스에서 수행될 수도 있다.
- [0080] 도 3 을 참조하면, 사용자 스피치 프로파일 관리와 연관된 동작들 (300) 의 예시적인 양태가 도시된다. 특정 양태에서, 동작들 (300) 중 하나 이상은 도 1 의 세그먼트화기 (124), 프로파일 관리기 (126), 도 2a 의 피쳐 추출기 (222), 하나 이상의 프로세서들 (220), 디바이스 (202), 시스템 (200), 또는 이들의 조합에 의해 수행된다.
- [0081] 화자 세그먼트화 (302) 동안, 도 2a 의 피쳐 추출기 (222) 는, 도 2a 를 참조하여 설명된 바와 같이, 오디오 스트림 (141) 에 기초하여 오디오 피쳐 데이터 세트들 (252) 을 생성한다. 세그먼트화기 (124) 는, 도 2a 를 참조하여 설명된 바와 같이, 세그먼트화 결과 (236) 를 생성하기 위해 오디오 피쳐 데이터 세트들 (252) 을 분석한다.
- [0082] 음성 프로파일 관리 (304) 동안, 도 1 의 프로파일 관리기 (126) 는, 306 에서, 오디오 피쳐 데이터 세트 (252) 가 등록된 화자에 대응하는지 여부를 결정한다. 예를 들어, 프로파일 관리기 (126) 는, 도 2b 를 참조하여 설명된 바와 같이, 오디오 피쳐 데이터 세트 (252) 가 임의의 사용자 스피치 프로파일 (150) 과 매칭하는지 여부를 결정한다. 프로파일 관리기 (126) 는, 306 에서, 오디오 피쳐 데이터 세트 (252) 가 프로파일 ID (155) 를 갖는 사용자 스피치 프로파일 (150A) 과 매칭함을 결정하는 것에 응답하여, 308 에서, 오디오 피쳐 테

이터 세트 (252) 에 적어도 부분적으로 기초하여 사용자 스피치 프로파일 (150A) 을 업데이트한다. 대안적으로, 프로파일 관리기 (126) 는, 306 에서, 오디오 피쳐 데이터 세트 (252) 가 복수의 사용자 스피치 프로파일들 (150) 중 임의의 것과 매칭하지 않음을 그리고 오디오 피쳐 데이터 세트 (252) 가 화자 (292A) 의 스피치를 나타낸다는 것을 세그먼트화 결과 (236) 가 표시함을 결정하는 것에 응답하여, 310 에서, 화자 (292A) 에 대해 지정된 등록 버퍼 (234A) 에 오디오 피쳐 데이터 세트 (252) 를 추가한다.

[0083] 프로파일 관리기 (126) 는, 312 에서, 등록 버퍼 (234A) 의 오디오 피쳐 데이터 세트들의 카운트 (또는 등록 버퍼 (234A) 의 오디오 피쳐 데이터 세트들의 스피치 지속기간) 가 등록 임계치 (264) 보다 크다고 결정하는 것에 응답하여, 314 에서, 화자를 등록한다. 예를 들어, 프로파일 관리기 (126) 는, 도 2b 를 참조하여 설명된 바와 같이, 등록 버퍼 (234A) 의 오디오 피쳐 데이터 세트들에 기초하여 사용자 스피치 프로파일 (150C) 을 생성하고 사용자 스피치 프로파일 (150C) 을 복수의 사용자 스피치 프로파일들 (150) 에 추가한다. 프로파일 관리기 (126) 는 오디오 스트림 (141) 의 후속 오디오 부분들을 계속 프로세싱한다.

[0084] 따라서, 화자 세그먼트화 (302) 동안 생성된 세그먼트화 결과 (236) 는 동일한 화자의 스피치에 대응하는 오디오 피쳐 데이터 세트들로 하여금 음성 프로파일 관리 (304) 동안 화자 등록을 위한 동일한 등록 버퍼에서 수집될 수 있게 한다. 다중의 오디오 피쳐 데이터 세트들에 기초하여 사용자 스피치 프로파일 (150C) 을 생성하는 것은 화자의 스피치를 표현함에 있어서 사용자 스피치 프로파일 (150C) 의 정확도를 개선한다.

[0085] 도 4 를 참조하면, 사용자 스피치 프로파일 관리와 연관된 동작들 (400) 의 예시적인 양태가 도시된다. 특정 양태에서, 동작들 (400) 중 하나 이상은 도 1 의 세그먼트화기 (124), 프로파일 관리기 (126), 도 2a 의 피쳐 추출기 (222), 하나 이상의 프로세서들 (220), 디바이스 (202), 시스템 (200), 또는 이들의 조합에 의해 수행된다.

[0086] 오디오 스트림 (141) 은 오디오 부분들 (151A) - 오디오 부분들 (151I) 을 포함한다. 화자 세그먼트화 (302) 동안, 도 1 의 세그먼트화기 (124) 는, 도 2a 를 참조하여 설명된 바와 같이, 오디오 부분들 (151A-I) 의 각각에 대한 세그먼트화 스코어 (254A), 세그먼트화 스코어 (254B), 및 세그먼트화 스코어 (254C) 를 생성한다.

[0087] 세그먼트화 스코어들 (254) 은, 오디오 부분들 (151A) 이 동일한 단일 화자 (예를 들어, 화자 (292A) 로서 지정됨) 의 스피치에 대응함을 표시한다. 예를 들어, 오디오 부분들 (151A) 의 각각의 세그먼트화 스코어 (254A) 는 세그먼트화 임계치 (257) 를 만족한다. 오디오 부분들 (151A) 의 각각의 세그먼트화 스코어 (254B) 및 세그먼트화 스코어 (254C) 는 세그먼트화 임계치 (257) 를 만족하지 않는다.

[0088] 음성 프로파일 관리 (304) 동안, 프로파일 관리기 (126) 는 화자 (292A) 와 연관된 등록 버퍼 (234A) 에 오디오 부분들 (151A) (예를 들어, 대응하는 오디오 피쳐 데이터 세트들) 을 추가한다. 프로파일 관리기 (126) 는 오디오 부분들 (151A) (예를 들어, 대응하는 오디오 피쳐 데이터 세트들) 에 기초하여 사용자 스피치 프로파일 (150A) 을 생성한다.

[0089] 특정 양태에서, 세그먼트화 스코어들 (254) 은, 오디오 부분들 (151B) 이 다중 화자들, 예를 들어, 화자 (292A) 및 다른 화자 (예를 들어, 화자 (292B) 로서 지정됨) 의 스피치에 대응함을 표시한다. 도 4 에서, 프로파일 관리기 (126) 는 오디오 부분들 (151B) (예를 들어, 대응하는 오디오 피쳐 데이터 세트들) 에 기초하여 사용자 스피치 프로파일 (150A) 을 업데이트한다. 특정 양태에서, 프로파일 관리기 (126) 는 또한, 화자 (292B) 와 연관된 등록 버퍼 (234B) 에 오디오 부분들 (151B) 을 추가한다. 대안적인 양태에서, 프로파일 관리기 (126) 는 다중 화자들에 대응하는 오디오 부분들 (151B) 을 무시한다. 예를 들어, 프로파일 관리기 (126) 는, 사용자 스피치 프로파일 (150) 을 업데이트 또는 생성하기 위해 오디오 부분들 (151B) 을 사용하는 것을 억제한다.

[0090] 세그먼트화 스코어들 (254) 은 오디오 부분들 (151C) 이 화자 (292B) (예를 들어, 단일 화자) 의 스피치에 대응함을 표시한다. 프로파일 관리기 (126) 는 오디오 부분들 (151C) 을 등록 버퍼 (234B) 에 추가한다. 프로파일 관리기 (126) 는, 등록 버퍼 (234B) 에 저장된 오디오 부분들 (예를 들어, 대응하는 오디오 피쳐 데이터 세트들) 이 등록 임계치 (264) 를 만족하지 못함을 결정하는 것에 응답하여, 등록 버퍼 (234B) 에 저장된 오디오 부분들 (예를 들어, 대응하는 오디오 피쳐 데이터 세트들) 에 기초하여 사용자 스피치 프로파일 (150) 을 생성하는 것을 억제한다. 특정 양태에서, 등록 버퍼 (234B) 에 저장된 오디오 부분들 (예를 들어, 대응하는 오디오 피쳐 데이터 세트들) 은 오디오 부분들 (151B) (예를 들어, 대응하는 오디오 피쳐 데이터 세트들) 및 오디오 부분들 (151C) (예를 들어, 대응하는 오디오 피쳐 데이터 세트들) 을 포함한다. 대안적인 양태에서, 등록 버퍼 (234B) 에 저장된 오디오 부분들 (예를 들어, 대응하는 오디오 피쳐 데이터 세트들) 은 오디오 부분

들 (151C) (예를 들어, 대응하는 오디오 피쳐 데이터 세트들) 을 포함하고 오디오 부분들 (151B) (예를 들어, 대응하는 오디오 피쳐 데이터 세트들) 을 포함하지 않는다.

[0091] 세그먼트화 스코어들 (254) 은, 오디오 부분들 (151D) 이 다른 단일 화자 (예를 들어, 화자 (292C) 로서 지정됨) 의 스피치에 대응함을 표시한다. 프로파일 관리기 (126) 는 오디오 부분들 (151D) 의 제 1 서브세트 (예를 들어, 대응하는 오디오 피쳐 데이터 세트들) 를 등록 버퍼 (234C) 에 추가한다. 프로파일 관리기 (126) 는, 등록 버퍼 (234C) 에 저장된 오디오 부분들 (151D) 의 제 1 서브세트 (예를 들어, 대응하는 오디오 피쳐 데이터 세트들) 가 등록 임계치 (264) 를 만족함을 결정하는 것에 응답하여, 등록 버퍼 (234C) 에 저장된 오디오 부분들 (151D) 의 제 1 서브세트 (예를 들어, 대응하는 오디오 피쳐 데이터 세트들) 에 기초하여 사용자 스피치 프로파일 (150B) 을 생성한다. 프로파일 관리기 (126) 는 오디오 부분들 (151D) 의 제 2 서브세트에 기초하여 사용자 스피치 프로파일 (150B) 을 업데이트한다.

[0092] 세그먼트화 스코어들 (254) 은 오디오 부분들 (151E) 이 임계 목음보다 큰 것에 대응함을 표시한다. 예를 들어, 오디오 부분들 (151E) 의 카운트는 목음 임계치 (294) 이상이다. 프로파일 관리기 (126) 는, 오디오 부분들 (151E) 이 임계 목음보다 큰 것에 대응함을 결정하는 것에 응답하여 등록 버퍼들 (234) 을 리셋한다.

[0093] 세그먼트화 스코어들 (254) 은, 오디오 부분들 (151F) 이 단일 화자 (예를 들어, 화자 (292A) 로서 지정됨) 의 스피치에 대응함을 표시한다. 프로파일 관리기 (126) 는, 오디오 부분들 (151F) 의 각각이 사용자 스피치 프로파일 (150B) 과 매칭함을 결정하는 것에 응답하여, 오디오 부분들 (151F) 에 기초하여 사용자 스피치 프로파일 (150B) 을 업데이트한다. 화자 지정 (예를 들어, 화자 (292A)) 이 재사용되고 있기 때문에, 오디오 부분 (151D) 및 오디오 부분 (151F) 은 상이한 지정된 화자들, 예를 들어, 화자 (292C) 및 화자 (292A) 와 각각 연관되지만, 오디오 부분 (151D) 및 오디오 부분 (151F) 은 동일한 화자 (예를 들어, 도 2a 의 사용자 (242C)) 의 스피치에 대응하고 동일한 사용자 스피치 프로파일 (예를 들어, 사용자 스피치 프로파일 (150B)) 과 매칭한다.

[0094] 세그먼트화 스코어들 (254) 은, 오디오 부분들 (151G) 이 단일 화자 (예를 들어, 화자 (292B) 로서 지정됨) 의 스피치에 대응함을 표시한다. 프로파일 관리기 (126) 는, 오디오 부분들 (151G) 의 제 1 서브세트가 사용자 스피치 프로파일들 (150) 중 임의의 것과 매칭하지 않음을 결정하는 것에 응답하여, 오디오 부분들 (151G) 의 제 1 서브세트를, 화자 (292B) 와 연관된 등록 버퍼 (234B) 에 추가한다. 프로파일 관리기 (126) 는 오디오 부분들 (151G) 의 제 1 서브세트에 기초하여 사용자 스피치 프로파일 (150C) 을 생성하고, 오디오 부분들 (151G) 의 제 2 서브세트에 기초하여 사용자 스피치 프로파일 (150C) 을 업데이트한다. 화자 지정 (예를 들어, 화자 (292B)) 이 재사용되고 있기 때문에, 오디오 부분 (151C) 및 오디오 부분 (151G) 은 동일한 지정된 화자, 예를 들어, 화자 (292B) 와 연관되며, 오디오 부분 (151C) 및 오디오 부분 (151G) 은 동일한 사용자 또는 상이한 사용자들의 스피치에 대응할 수 있다.

[0095] 세그먼트화 스코어들 (254) 은 오디오 부분들 (151H) 이 임계 목음보다 큰 것에 대응함을 표시한다. 프로파일 관리기 (126) 는, 오디오 부분들 (151H) 이 임계 목음보다 큰 것에 대응함을 결정하는 것에 응답하여 등록 버퍼들 (234) 을 리셋한다.

[0096] 세그먼트화 스코어들 (254) 은, 오디오 부분들 (151I) 이 단일 화자 (예를 들어, 화자 (292C) 로서 지정됨) 의 스피치에 대응함을 표시한다. 프로파일 관리기 (126) 는, 오디오 부분들 (151I) 의 각각이 사용자 스피치 프로파일 (150A) 과 매칭함을 결정하는 것에 응답하여, 오디오 부분들 (151I) 에 기초하여 사용자 스피치 프로파일 (150A) 을 업데이트한다. 화자 지정 (예를 들어, 화자 (292C)) 이 재사용되고 있기 때문에, 오디오 부분 (151A) 및 오디오 부분 (151I) 은 상이한 지정된 화자들, 예를 들어, 화자 (292A) 및 화자 (292C) 와 각각 연관되지만, 오디오 부분 (151A) 및 오디오 부분 (151I) 은 동일한 사용자 (예를 들어, 도 2a 의 사용자 (242A)) 의 스피치에 대응하고 동일한 사용자 스피치 프로파일 (예를 들어, 사용자 스피치 프로파일 (150A)) 과 매칭한다. 대안적인 양태에서, 프로파일 관리기 (126) 는, 오디오 부분들 (151I) 이 복수의 사용자 스피치 프로파일들 (150) 중 임의의 것과 매칭하지 않음을 결정하는 것에 응답하여, 화자 (292C) 와 연관된 등록 버퍼 (234C) 에 오디오 부분들 (151I) 의 제 1 서브세트를 추가하고, 오디오 부분들 (151I) 의 제 1 서브세트에 기초하여 사용자 스피치 프로파일 (150D) 을 생성한다. 화자 지정 (예를 들어, 화자 (292C)) 을 재사용함으로써, 프로파일 관리기 (126) 는, 세그먼트화기 (124) 에 의해 구별될 수 있는 화자들 (292) 의 미리 결정된 카운트 (예를 들어, K) 보다 더 큰 카운트의 사용자 프로파일들을 생성 (또는 업데이트) 할 수 있다.

[0097] 도 5 를 참조하면, 사용자 스피치 프로파일 관리와 연관된 동작들 (500) 의 예시적인 양태가 도시된다. 특정 양태에서, 동작들 (500) 중 하나 이상은 도 1 의 세그먼트화기 (124), 프로파일 관리기 (126), 도 2a 의 피

처 추출기 (222), 하나 이상의 프로세서들 (220), 디바이스 (202), 시스템 (200), 또는 이들의 조합에 의해 수행된다.

[0098] 오디오 스트림 (141)은 오디오 부분들 (151A), 오디오 부분들 (151B), 및 오디오 부분들 (151C)을 포함한다. 예를 들어, 오디오 부분들 (151A)은 오디오 부분 (151D) (예컨대, 오디오 프레임), 하나 이상의 추가적인 오디오 부분들, 및 오디오 부분 (151E)을 포함한다. 오디오 부분들 (151B)은 오디오 부분 (151F), 하나 이상의 추가적인 오디오 부분들, 오디오 부분 (151G)을 포함한다. 오디오 부분들 (151C)은 오디오 부분 (151H), 하나 이상의 추가적인 오디오 부분들, 및 오디오 부분 (151I)을 포함한다.

[0099] 특정 양태에서, 오디오 부분들 (151A)의 각각의 데이터 세트 세그먼트화 결과 (256A)는 오디오 부분 (151A)이 화자 (292A)의 스피치에 대응함을 표시한다. 예를 들어, 오디오 부분 (151D)의 데이터 세트 세그먼트화 결과 (256D) (예컨대, "1")는 오디오 부분 (151D)이 화자 (292A)의 스피치를 나타냄을 표시한다. 다른 예로서, 오디오 부분 (151E)의 데이터 세트 세그먼트화 결과 (256E) (예컨대, "1")는 오디오 부분 (151E)이 화자 (292A)의 스피치를 나타냄을 표시한다.

[0100] 오디오 부분들 (151B)의 각각의 데이터 세트 세그먼트화 결과 (256B)는 오디오 부분 (151B)이 묵음 (또는 비-스피치 잡음)에 대응함을 표시한다. 예를 들어, 오디오 부분 (151F)의 데이터 세트 세그먼트화 결과 (256F) (예컨대, "0")는 오디오 부분 (151F)이 묵음 (또는 비-스피치 잡음)을 나타냄을 표시한다. 다른 예로서, 오디오 부분 (151G)의 데이터 세트 세그먼트화 결과 (256G) (예컨대, "0")는 오디오 부분 (151G)이 묵음 (또는 비-스피치 잡음)을 나타냄을 표시한다.

[0101] 오디오 부분들 (151C)의 각각의 데이터 세트 세그먼트화 결과 (256C)는 오디오 부분 (151C)이 화자 (292B)의 스피치에 대응함을 표시한다. 예를 들어, 오디오 부분 (151H)의 데이터 세트 세그먼트화 결과 (256H) (예컨대, "2")는 오디오 부분 (151H)이 화자 (292B)의 스피치를 나타냄을 표시한다. 다른 예로서, 오디오 부분 (151I)의 데이터 세트 세그먼트화 결과 (256I) (예컨대, "2")는 오디오 부분 (151I)이 화자 (292B)의 스피치를 나타냄을 표시한다.

[0102] 그래프 (590)는 세그먼트화 결과 (236)의 일 예의 시각적 묘사이다. 예를 들어, 오디오 부분들 (151A)은 화자 (292A) (예를 들어, 단일 화자)의 스피치를 나타내어, 오디오 부분들 (151A)은 오디오 스트림 (141)의 화자-동일 오디오 세그먼트 (111A)에 대응한다. 오디오 부분들 (151B)은 묵음을 나타내어, 오디오 부분들 (151B)은 오디오 스트림 (141)의 오디오 세그먼트 (113A) (예를 들어, 화자-동일 오디오 세그먼트가 아님)에 대응한다. 오디오 부분들 (151C)은 화자 (292B) (예를 들어, 단일 화자)의 스피치를 나타내어, 오디오 부분들 (151C)은 오디오 스트림 (141)의 화자-동일 오디오 세그먼트 (111B)에 대응한다.

[0103] 그래프 (592)는 스피치 프로파일 결과 (238)의 일 예의 시각적 묘사이다. 프로파일 관리기 (126)는 오디오 부분들 (151A)의 제 1 서브세트에 기초하여 사용자 스피치 프로파일 (150A)을 생성한다. 프로파일 관리기 (126)는, 사용자 스피치 프로파일 (150A)의 생성 이후, 후속 오디오 부분들 (예를 들어, 후속 오디오 피쳐 데이터 세트들)을 사용자 스피치 프로파일 (150A)과 비교함으로써 스피치 프로파일 결과 (238A)를 결정한다. 오디오 부분 (151)의 스피치 프로파일 결과 (238A)는, 오디오 부분 (151)이 사용자 스피치 프로파일 (150A)과 매칭할 가능성을 표시한다. 프로파일 관리기 (126)는 오디오 부분들 (151C)의 제 1 서브세트를 사용자 스피치 프로파일 (150A)과 비교함으로써 오디오 부분들 (151C)의 제 1 서브세트의 스피치 프로파일 결과 (238A)를 결정한다. 프로파일 관리기 (126)는, 오디오 부분들 (151C)의 제 1 서브세트의 스피치 프로파일 결과 (238A)가 프로파일 임계치 (258) 미만을 결정하는 것에 응답하여, 오디오 부분들 (151C)의 제 1 서브세트가 사용자 스피치 프로파일 (150A)과 매칭하지 않음을 결정한다.

[0104] 프로파일 관리기 (126)는, 오디오 부분들 (151C)의 제 1 서브세트가 사용자 스피치 프로파일 (150A)과 매칭하지 않음을 결정하는 것에 응답하여, 오디오 부분들 (151C)의 제 1 서브세트에 기초하여 사용자 스피치 프로파일 (150B)을 생성한다. 프로파일 관리기 (126)는, 사용자 스피치 프로파일 (150B)의 생성 이후, 후속 오디오 부분들을 사용자 스피치 프로파일 (150B)과 비교함으로써 스피치 프로파일 결과 (238B)를 결정한다. 스피치 프로파일 결과 (238B)는, 오디오 부분들이 사용자 스피치 프로파일 (150B)과 매칭할 가능성을 표시한다. 예를 들어, 오디오 부분들 (151C)의 제 2 서브세트의 스피치 프로파일 결과 (238B)는, 오디오 부분들 (151C)의 제 2 서브세트가 사용자 스피치 프로파일 (150B)과 매칭함을 표시한다. 특정 양태에서, 프로파일 관리기 (126)는 그래프 (590), 그래프 (592), 또는 그 양자 모두를 포함하는 그래픽 사용자 인터페이스 (GUI)를 생성하고, GUI를 디스플레이 디바이스에 제공한다.

- [0105] 도 6 을 참조하면, 사용자 스피치 프로파일 관리와 연관된 동작들 (600) 의 예시적인 양태가 도시된다. 특정 양태에서, 동작들 (600) 중 하나 이상은 도 1 의 세그먼트화기 (124), 프로파일 관리기 (126), 도 2a 의 피처 추출기 (222), 하나 이상의 프로세서들 (220), 디바이스 (202), 시스템 (200), 또는 이들의 조합에 의해 수행된다.
- [0106] 오디오 스트림 (141) 은 다중 화자들의 스피치에 대응하는 오디오 부분들 (151J) 을 포함한다. 예를 들어, 오디오 부분들 (151J) 은 오디오 부분 (151K) (예컨대, 오디오 프레임), 하나 이상의 추가적인 오디오 부분들, 및 오디오 부분 (151L) 을 포함한다. 특정 양태에서, 오디오 부분들 (151J) 의 각각의 데이터 세트 세그먼트화 결과 (256D) 는 오디오 부분 (151J) 이 화자 (292A) 및 화자 (292B) 의 스피치에 대응함을 표시한다. 예를 들어, 오디오 부분 (151K) 의 데이터 세트 세그먼트화 결과 (256K) (예컨대, "1, 2") 는 오디오 부분 (151K) 이 화자 (292A) 및 화자 (292B) 의 스피치를 나타냄을 표시한다. 다른 예로서, 오디오 부분 (151L) 의 데이터 세트 세그먼트화 결과 (256L) (예컨대, "1, 2") 는 오디오 부분 (151L) 이 화자 (292A) 및 화자 (292B) 의 스피치를 나타냄을 표시한다. 오디오 부분들 (151J) 은 다중 화자들의 스피치를 나타내어, 오디오 부분들 (151J) 은 오디오 세그먼트 (113B) (예를 들어, 화자-동일 오디오 세그먼트가 아님) 에 대응한다.
- [0107] 프로파일 관리기 (126) 는, 사용자 스피치 프로파일 (150A) 의 생성 이후, 후속 오디오 부분들 (예를 들어, 후속 오디오 피처 데이터 세트들) 을 사용자 스피치 프로파일 (150A) 과 비교함으로써 스피치 프로파일 결과 (238A) 를 결정한다. 프로파일 관리기 (126) 는 오디오 부분들 (151J) 을 사용자 스피치 프로파일 (150A) 과 비교함으로써 오디오 부분들 (151J) 의 스피치 프로파일 결과 (238A) 를 결정한다. 특정 양태에서, 오디오 부분들 (151J) 이 화자 (292A) 의 스피치에 추가하여 화자 (292B) 의 스피치를 포함하기 때문에, 오디오 부분들 (151J) 에 대한 스피치 프로파일 결과 (238A) 는 오디오 부분들 (151A) 에 대한 스피치 프로파일 결과 (238A) 보다 낮다.
- [0108] 도 7 을 참조하면, 사용자 스피치 프로파일 관리와 연관된 동작들 (700) 의 예시적인 양태가 도시된다. 특정 양태에서, 동작들 (700) 중 하나 이상은 도 2a 의 피처 추출기 (222), 세그먼트화기 (124), 프로파일 관리기 (126), 하나 이상의 프로세서들 (220), 디바이스 (202), 시스템 (200), 또는 이들의 조합에 의해 수행된다.
- [0109] 오디오 스트림 (141) 은 오디오 부분들 (151J) 및 오디오 부분들 (151K) 을 포함한다. 예를 들어, 오디오 부분들 (151J) 은 오디오 부분 (151L) (예컨대, 오디오 프레임), 하나 이상의 추가적인 오디오 부분들, 및 오디오 부분 (151M) 을 포함한다. 오디오 부분들 (151K) 은 오디오 부분 (151N) (예컨대, 오디오 프레임), 하나 이상의 추가적인 오디오 부분들, 및 오디오 부분 (151O) 을 포함한다.
- [0110] 특정 양태에서, 오디오 부분들 (151J) 의 각각의 데이터 세트 세그먼트화 결과 (256J) 는, 오디오 부분 (151J) 이 화자 (292C) (예를 들어, 단일 화자) 의 스피치를 나타내어 오디오 부분들 (151J) 이 화자-동일 오디오 세그먼트 (111C) 에 대응함을 표시한다. 오디오 부분들 (151K) 의 각각의 데이터 세트 세그먼트화 결과 (256K) 는, 오디오 부분 (151K) 이 묵음 (또는 비-스피치 잡음) 을 나타내어 오디오 부분들 (151K) 이 오디오 세그먼트 (113C) 에 대응함을 표시한다.
- [0111] 프로파일 관리기 (126) 는, 사용자 스피치 프로파일 (150A) 의 생성 이후, 오디오 부분들 (151J) 을 사용자 스피치 프로파일 (150A) 과 비교함으로써 오디오 부분들 (151J) 의 스피치 프로파일 결과 (238A) 를 결정한다. 프로파일 관리기 (126) 는, 스피치 프로파일 결과 (238A) 가 프로파일 임계치 (258) 미만임을 결정하는 것에 응답하여, 오디오 부분들 (151J) 이 사용자 스피치 프로파일 (150A) 과 매칭하지 않음을 결정한다.
- [0112] 프로파일 관리기 (126) 는, 오디오 부분들 (151J) 이 사용자 스피치 프로파일 (150A) 과 매칭하지 않음을 결정하는 것에 응답하여, 화자 (292C) 와 연관된 등록 버퍼 (234C) 에 오디오 부분들 (151J) 을 저장한다. 프로파일 관리기 (126) 는, 등록 버퍼 (234C) 에 저장된 오디오 부분들 (151J) 이 등록 임계치 (264) 를 만족하지 못함을 결정하는 것에 응답하여, 등록 버퍼 (234C) 에 저장된 오디오 부분들 (151J) 에 기초하여 사용자 스피치 프로파일 (150) 을 생성하는 것을 억제한다. 프로파일 관리기 (126) 는, 오디오 부분들 (151K) 이 임계 묵음보다 큰 것을 표시함을 결정하는 것에 응답하여 등록 버퍼들 (234) 을 리셋한다 (예를 들어, 앰프티로서 마킹함). 따라서, 오디오 부분들 (151J) 은, 화자 (292C) 가 말하기를 중지한 것으로 보일 때 등록 버퍼 (234C) 로부터 제거된다.
- [0113] 도 8 을 참조하면, 사용자 스피치 프로파일 관리와 연관된 동작들 (800) 의 예시적인 양태가 도시된다. 특정 양태에서, 동작들 (800) 중 하나 이상은 도 1 의 세그먼트화기 (124), 프로파일 관리기 (126), 도 2a 의 피처 추출기 (222), 하나 이상의 프로세서들 (220), 디바이스 (202), 시스템 (200), 또는 이들의 조합에 의해 수

행된다.

- [0114] 도 1 의 세그먼트화기 (124) 는, 804 에서, 화자 세그먼트화 (302) 를 수행한다. 예를 들어, 세그먼트화기 (124) 는, 도 2a 를 참조하여 설명된 바와 같이, 시간 T 에서 피쳐 추출기 (222) 로부터 오디오 피쳐 데이터 세트 (252) 를 수신하고, 오디오 부분 (151) 의 오디오 피쳐 데이터 세트 (252) 에 대한 세그먼트화 스코어들 (254) 을 생성한다.
- [0115] 도 1 의 프로파일 관리기 (126) 는, 806 에서, 세그먼트화 스코어들 (254) 중 임의의 것이 세그먼트화 임계치 (257) 를 만족하는지 여부를 결정한다. 예를 들어, 프로파일 관리기 (126) 는, 세그먼트화 스코어들 (254) 중 어느 것도 세그먼트화 임계치 (257) 를 만족하지 않음을 결정하는 것에 응답하여, 오디오 피쳐 데이터 세트 (252) 가 묵음 (또는 비-스피치 잡음) 을 나타냄을 결정하고 묵음 카운트 (262) 를 (예를 들어, 1 만큼) 증분시킨다. 프로파일 관리기 (126) 는, 묵음 카운트 (262) 를 증분시키는 것에 후속하여, 808 에서, 묵음 카운트 (262) 가 묵음 임계치 (294) 보다 큰지 여부를 결정한다.
- [0116] 프로파일 관리기 (126) 는, 808 에서 묵음 카운트 (262) 가 묵음 임계치 (294) 보다 크다고 결정하는 것에 응답하여, 810 에서 리셋을 수행한다. 예를 들어, 프로파일 관리기 (126) 는 등록 버퍼들 (234) 을 리셋 (예컨대, 엠프티로서 마킹) 하는 것, 프로브 버퍼들 (240) 을 리셋 (예컨대, 엠프티로서 마킹) 하는 것, 묵음 카운트 (262) 를 리셋 (예컨대, 0 으로 리셋) 하는 것, 또는 이들의 조합에 의해 리셋을 수행하고, 오디오 스트림 (141) 의 후속 오디오 피쳐 데이터 세트들을 프로세싱하기 위해 804 로 리턴한다. 대안적으로, 프로파일 관리기 (126) 는, 808 에서 묵음 카운트 (262) 가 묵음 임계치 (294) 이하라고 결정하는 것에 응답하여, 오디오 스트림 (141) 의 후속 오디오 피쳐 데이터 세트들을 프로세싱하기 위해 804 로 리턴한다.
- [0117] 프로파일 관리기 (126) 는, 806 에서 세그먼트화 스코어들 (254) 중 적어도 하나가 세그먼트화 임계치 (257) 를 만족함을 결정하는 것에 응답하여, 812 에서, 오디오 피쳐 데이터 세트 (252) 를 프로브 버퍼들 (240) 중 적어도 하나에 추가한다. 예를 들어, 프로파일 관리기 (126) 는, 화자 (292A) 와 연관된 세그먼트화 스코어 (254A) 가 세그먼트화 임계치 (257) 를 만족함을 결정하는 것에 응답하여, 오디오 피쳐 데이터 세트 (252) 가 화자 (292A) 의 스피치를 나타냄을 결정하고, 오디오 피쳐 데이터 세트 (252) 를, 화자 (292A) 와 연관된 프로브 버퍼 (240A) 에 추가한다. 특정 구현에서, 다중 화자들 (292) 의 스피치를 나타내는 오디오 피쳐 데이터 세트 (252) 는 다중 화자들 (292) 에 대응하는 다중의 프로브 버퍼들 (240) 에 추가된다. 예를 들어, 프로파일 관리기 (126) 는, 세그먼트화 스코어 (254A) 및 세그먼트화 스코어 (254B) 의 각각이 세그먼트화 임계치 (257) 를 만족함을 결정하는 것에 응답하여, 오디오 피쳐 데이터 세트 (252) 를 프로브 버퍼 (140A) 및 프로브 버퍼 (140B) 에 추가한다. 대안적인 구현에서, 다중 화자들 (292) 의 스피치를 나타내는 오디오 피쳐 데이터 세트 (252) 는 무시되고 프로브 버퍼들 (240) 에 추가되지 않는다.
- [0118] 프로파일 관리기 (126) 는, 816 에서, 대응하는 화자 (예를 들어, 화자 (292A)) 가 등록되는지 여부를 결정한다. 예를 들어, 프로파일 관리기 (126) 는, 대응하는 프로브 버퍼 (240) (예컨대, 프로브 버퍼 (240A)) 의 오디오 피쳐 데이터 세트들 (예컨대, 오디오 피쳐 데이터 세트 (252) 를 포함) 을 복수의 사용자 스피치 프로파일들 (150) 과 비교함으로써 화자 (292) (예컨대, 화자 (292A)) 가 등록되는지 여부를 결정한다.
- [0119] 프로파일 관리기 (126) 는, 816 에서 화자 (292) (예컨대, 화자 (292A)) 가 등록되지 않음을 결정하는 것에 응답하여, 818 에서, 오디오 피쳐 데이터 세트 (252) 가 품질 체크를 통과하는지 여부를 결정한다. 예를 들어, 프로파일 관리기 (126) 는, 오디오 피쳐 데이터 세트 (252) 가 다중 화자들 (292) 에 대응함을 결정하는 것에 응답하여, 오디오 피쳐 데이터 세트 (252) 가 품질 체크를 실패한 것을 결정한다. 대안적으로, 프로파일 관리기 (126) 는, 오디오 피쳐 데이터 세트 (252) 가 단일 화자에 대응함을 결정하는 것에 응답하여, 오디오 피쳐 데이터 세트 (252) 가 품질 체크를 통과한 것을 결정한다.
- [0120] 프로파일 관리기 (126) 는, 818 에서 오디오 피쳐 데이터 세트 (252) 가 품질 체크를 통과하지 못함을 결정하는 것에 응답하여, 오디오 스트림 (141) 의 후속 오디오 피쳐 데이터 세트들을 프로세싱하기 위해 804 로 리턴한다. 대안적으로, 프로파일 관리기 (126) 는, 818 에서 오디오 피쳐 데이터 세트 (252) 가 품질 체크를 통과함을 결정하는 것에 응답하여, 820 에서, 화자 (292) (예컨대, 화자 (292A)) 의 스피치를 나타내는 오디오 피쳐 데이터 세트 (252) 를 화자 (292) 와 연관된 등록 버퍼 (234) (예컨대, 등록 버퍼 (234A)) 에 추가한다.
- [0121] 프로파일 관리기 (126) 는, 822 에서, 등록 버퍼 (234) (예컨대, 등록 버퍼 (234A)) 에 저장된 오디오 피쳐 데이터 세트들의 카운트가 등록 임계치 (264) 보다 큰지 여부를 결정한다. 프로파일 관리기 (126) 는, 822 에

서 등록 버퍼들 (234) (예컨대, 등록 버퍼 (234))의 각각에 저장된 오디오 피쳐 데이터 세트들의 카운트가 등록 임계치 (264) 이하라고 결정하는 것에 응답하여, 오디오 스트림 (141)의 후속 오디오 피쳐 데이터 세트들을 프로세싱하기 위해 804로 리턴한다. 대안적으로, 프로파일 관리기 (126)는, 등록 버퍼 (234) (예컨대, 등록 버퍼 (234A))의 오디오 피쳐 데이터 세트들의 카운트가 등록 임계치 (264)보다 크다고 결정하는 것에 응답하여, 824에서, 사용자 스피치 프로파일 (150A)을 생성하고, 사용자 스피치 프로파일 (150A)을 복수의 사용자 스피치 프로파일들 (150)에 추가하고, 오디오 스트림 (141)의 후속 오디오 피쳐 데이터 세트들을 프로세싱하기 위해 804로 리턴한다.

[0122] 프로파일 관리기 (126)는, 816에서 화자 (292A)가 등록됨을 결정하는 것에 응답하여, 826에서, 오디오 피쳐 데이터 세트 (252) (또는 스피치가 오디오 피쳐 데이터 세트 (252)에 의해 표현되는 화자 (292)와 연관된 프로브 버퍼 (240)의 오디오 피쳐 데이터 세트들)가 품질 체크를 통과하는지 여부를 결정한다. 프로파일 관리기 (126)는, 826에서 오디오 피쳐 데이터 세트 (252) (또는 프로브 버퍼 (240)의 오디오 피쳐 데이터 세트들)가 품질 체크를 통과하지 못함을 결정하는 것에 응답하여, 오디오 스트림 (141)의 후속 오디오 피쳐 데이터 세트들을 프로세싱하기 위해 804로 리턴한다. 프로파일 관리기 (126)는, 826에서 오디오 피쳐 데이터 세트 (252) (또는 프로브 버퍼 (240)의 오디오 피쳐 데이터 세트들)가 품질 체크를 통과함을 결정하는 것에 응답하여, 오디오 피쳐 데이터 세트 (252) (또는 프로브 버퍼 (240)의 오디오 피쳐 데이터 세트들)에 기초하여 (오디오 피쳐 데이터 세트 (252)와 매칭하는) 사용자 스피치 프로파일 (150A)을 업데이트하고, 오디오 스트림 (141)의 후속 오디오 피쳐 데이터 세트들을 프로세싱하기 위해 804로 리턴한다. 대안적인 양태에서, 826에서, 품질 체크는 오디오 피쳐 데이터 세트 (252)를 프로브 버퍼 (240)에 추가하기 전에 수행된다. 예를 들어, 프로파일 관리기 (126)는, 오디오 피쳐 데이터 세트 (252)가 품질 체크를 통과하지 못함을 결정하는 것에 응답하여, 오디오 피쳐 데이터 세트 (252)를 프로브 버퍼 (240)에 추가하는 것을 억제하고, 오디오 스트림 (141)의 후속 오디오 피쳐 데이터 세트들을 프로세싱하기 위해 804로 리턴한다.

[0123] 도 9를 참조하면, 사용자 스피치 프로파일 관리와 연관된 동작들 (900)의 예시적인 양태가 도시된다. 특정 양태에서, 동작들 (900)중 하나 이상은 도 1의 세그먼트화기 (124), 프로파일 관리기 (126), 도 2a의 피쳐 추출기 (222), 화자 검출기 (278), 하나 이상의 프로세서들 (220), 디바이스 (202), 시스템 (200), 또는 이들의 조합에 의해 수행된다.

[0124] 하나 이상의 프로세서들 (220)은, 전력 모드 (282)에서, 시간 T에서 오디오 피쳐들 (예를 들어, 오디오 피쳐 데이터 세트 (252))을 버퍼 (268)에 추가한다. 도 2a의 화자 검출기 (278)는, 904에서, 다중 화자들이 이 오디오 스트림 (141)에서 검출되는지 여부를 결정한다. 예를 들어, 화자 검출기 (278)는, 오디오 피쳐들 (예컨대, 오디오 피쳐 데이터 세트 (252))이 다중 화자들의 스피치를 나타냄을 결정하는 것에 응답하여 다중 화자들이 검출됨을 결정한다. 다른 예에서, 화자 검출기 (278)는, 오디오 피쳐들 (예컨대, 오디오 피쳐 데이터 세트 (252))이 이전 오디오 피쳐들 (예컨대, 이전 오디오 피쳐 데이터 세트)에서 검출된 제 1 화자의 스피치에 후속하는 제 2 화자의 스피치를 나타냄을 결정하는 것에 응답하여 다중 화자들이 검출됨을 결정한다.

[0125] 화자 검출기 (278)는, 904에서 다중 화자들이 오디오 스트림 (141)에서 검출되지 않았음을 결정하는 것에 응답하여, 오디오 스트림 (141)의 후속 오디오 피쳐들을 계속 프로세싱한다. 대안적으로, 화자 검출기 (278)는, 904에서 다중 화자들이 오디오 스트림 (141)에서 검출됨을 결정하는 것에 응답하여, 906에서, 하나 이상의 프로세서들 (220)을 전력 모드 (282)로부터 전력 모드 (284)로 트랜지션시키고 하나 이상의 어플리케이션들 (920)을 활성화시킨다. 특정 양태에서, 하나 이상의 어플리케이션들 (920)은 피쳐 추출기 (222), 세그먼트화기 (124), 프로파일 관리기 (126), 하나 이상의 오디오 분석 어플리케이션들 (180), 또는 이들의 조합을 포함한다. 특정 양태에서, 화자 검출기 (278)는 하나 이상의 어플리케이션들 (920)을 활성화하기 위해 하나 이상의 프로세서들 (220)을 전력 모드 (282)로부터 전력 모드 (284)로 트랜지션시키기 위한 웨이크업 신호 또는 인터럽트 중 적어도 하나를 생성한다.

[0126] 화자 검출기 (278)는, 910에서, 전력 모드 (284)에서, 다중 화자들이 검출되는지 여부를 결정한다. 예를 들어, 화자 검출기 (278)는, 다중 화자들이 검출되는지 여부의 이전 결정 이후 임계 시간이 만료한 이후에 다중 화자들이 검출되는지 여부를 결정한다. 화자 검출기 (278)는, 다중 화자들이 검출됨을 결정하는 것에 응답하여, 전력 모드 (282)로 트랜지션시키는 것을 억제한다. 대안적으로, 화자 검출기 (278)는, 다중 화자들이 오디오 피쳐 데이터 세트들의 임계 카운트 내에서 검출되지 않음을 결정하는 것에 응답하여, 하나 이상의 프로세서들 (220)을 전력 모드 (284)로부터 전력 모드 (282)로 트랜지션시킨다.

[0127] 따라서, 하나 이상의 프로세서들 (220)은 (전력 모드 (284)에 비해) 전력 모드 (282)에서 동작함으로써 에너지

지를 보존하고, 전력 모드 (282) 에서 동작하지 않는 컴포넌트들을 활성화하기 위해 필요에 따라 전력 모드 (284) 로 트랜지션시킨다. 전력 모드 (284) 로 선택적 트랜지션시키는 것은 디바이스 (202) 의 전체 전력 소비를 감소시킨다.

[0128] 도 10 을 참조하면, 사용자 스피치 프로파일 관리의 방법 (1000) 의 특정 구현이 도시된다. 특정 양태에서, 방법 (1000) 의 하나 이상의 동작들은 도 1 의 세그먼트화기 (124), 프로파일 관리기 (126), 도 2a 의 화자 검출기 (278), 하나 이상의 프로세서들 (220), 디바이스 (202), 시스템 (200), 또는 이들의 조합 중 적어도 하나에 의해 수행된다.

[0129] 방법 (1000) 은, 1002 에서, 제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하는 단계를 포함한다. 예를 들어, 도 2a 를 참조하여 설명된 바와 같이, 도 2a 의 화자 검출기 (278) 는, 전력 모드 (282) 에서, 오디오 스트림 (141) 이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정한다.

[0130] 방법 (1000) 은, 1004 에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응함을 결정하는 것에 기초하여, 제 2 전력 모드에서, 세그먼트화 결과를 생성하기 위해 오디오 스트림의 오디오 피쳐 데이터를 분석하는 단계를 포함한다. 예를 들어, 도 2a 를 참조하여 설명된 바와 같이, 도 2a 의 하나 이상의 프로세서들 (220) 은, 오디오 스트림 (141) 이 적어도 2명의 별개의 화자들의 스피치에 대응함을 결정하는 것에 기초하여, 전력 모드 (284) 로 트랜지션시키고 세그먼트화기 (124) 를 활성화시킨다. 도 2a 를 참조하여 설명된 바와 같이, 세그먼트화기 (124) 는, 전력 모드 (284) 에서, 세그먼트화 결과 (236) 를 생성하기 위해 오디오 스트림 (141) 의 오디오 피쳐 데이터 세트들 (252) 을 분석한다. 도 2a 를 참조하여 설명된 바와 같이, 세그먼트화 결과 (236) 는 오디오 스트림 (141) 의 화자-동일 오디오 세그먼트들 (예를 들어, 화자-동일 오디오 세그먼트 (111A) 및 화자-동일 오디오 세그먼트 (111B)) 을 표시한다.

[0131] 방법 (1000) 은 또한, 1006 에서, 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들 중의 제 1 오디오 피쳐 데이터 세트에 대한 복수의 사용자 스피치 프로파일들의 비교를, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하는 단계를 포함한다. 예를 들어, 도 2b 를 참조하여 설명된 바와 같이, 도 1 의 프로파일 관리기 (126) 는, 화자-동일 오디오 세그먼트 (111A) 의 하나 이상의 오디오 피쳐 데이터 세트들 (252A) 중의 오디오 피쳐 데이터 세트 (252) 에 대한 복수의 사용자 스피치 프로파일들 (150) 의 비교를, 오디오 피쳐 데이터 세트 (252) 가 복수의 사용자 스피치 프로파일들 (150) 중 임의의 것과 매칭하는지 여부를 결정하기 위해 수행한다.

[0132] 방법 (1000) 은, 1008 에서, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여: 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하는 단계, 및 제 1 사용자 스피치 프로파일을 복수의 사용자 스피치 프로파일들에 추가하는 단계를 더 포함한다. 예를 들어, 도 2b 를 참조하여 설명된 바와 같이, 도 1 의 프로파일 관리기 (126) 는, 오디오 피쳐 데이터 세트 (252) 가 복수의 사용자 스피치 프로파일들 (150) 중 임의의 것과 매칭하지 않음을 결정하는 것에 기초하여, 하나 이상의 오디오 피쳐 데이터 세트들 (252A) 의 적어도 서브세트에 기초하여 사용자 스피치 프로파일 (150C) 을 생성하고, 사용자 스피치 프로파일 (150C) 을 복수의 사용자 스피치 프로파일들 (150) 에 추가한다.

[0133] 방법 (1000) 은 화자-동일 오디오 세그먼트의 오디오 피쳐 데이터 세트들에 기초하여 사용자 스피치 프로파일의 생성을 가능케 한다. 동일한 화자의 스피치에 대응하는 다중의 오디오 피쳐 데이터 세트들을 사용하는 것은, 단일 오디오 피쳐 데이터에 기초하여 사용자 스피치 프로파일을 생성하는 것에 비해, 화자의 스피치를 표현함에 있어서 사용자 스피치 프로파일의 정확도를 개선한다. 수동 등록은, 사용자가 미리결정된 단어 또는 문장들을 말해야 하는 것 또는 사전 등록되어야 하는 것 없이 사용자 스피치 프로파일을 생성하는데 사용될 수 있다.

[0134] 도 10 의 방법 (1000) 은 필드 프로그래밍가능 게이트 어레이 (FPGA) 디바이스, 주문형 집적 회로 (ASIC), 프로세싱 유닛, 예컨대, 중앙 프로세싱 유닛 (CPU), DSP, 제어기, 다른 하드웨어 디바이스, 펌웨어 디바이스, 또는 이들의 임의의 조합에 의해 구현될 수도 있다. 일 예로서, 도 10 의 방법 (1000) 은, 도 19 를 참조하여 설명된 바와 같이, 명령들을 실행하는 프로세서에 의해 수행될 수도 있다.

[0135] 도 11 은 하나 이상의 프로세서들 (220) 을 포함하는 집적 회로 (1102) 로서의 디바이스 (202) 의 구현 (1100)

을 도시한다. 하나 이상의 프로세서들 (220) 은 복수의 어플리케이션들 (1122) 을 포함한다. 어플리케이션들 (1122) 은 피쳐 추출기 (222), 화자 검출기 (278), 세그먼트화기 (124), 프로파일 관리기 (126), 프로파일 관리기 (126), 하나 이상의 오디오 분석 어플리케이션들 (180), 또는 이들의 조합을 포함한다. 집적 회로 (1102) 는 또한, 오디오 스트림 (141) 이 프로세싱을 위해 수신될 수 있게 하기 위해 하나 이상의 버스 인터페이스들과 같은 오디오 입력부 (1104) 를 포함한다. 집적 회로 (1102) 는 또한, 프로파일 ID (155) 와 같은 출력 신호 (1143) 의 전송을 가능케 하기 위해 버스 인터페이스와 같은 신호 출력부 (1106) 를 포함한다.

집적 회로 (1102) 는 도 12 에 도시된 바와 같은 모바일 폰 또는 태블릿, 도 13 에 도시된 바와 같은 헤드셋, 도 14 에 도시된 바와 같은 웨어러블 전자 디바이스, 도 15 에 도시된 바와 같은 음성 제어식 스피커 시스템, 도 16 에 도시된 바와 같은 가상 현실 헤드셋 또는 증강 현실 헤드셋, 또는 도 17 또는 도 18 에 도시된 바와 같은 차량과 같이 마이크로폰들을 포함하는 시스템에서의 컴포넌트로서 사용자 스피치 프로파일 관리의 구현을 가능케 한다.

[0136] 도 12 는, 디바이스 (202) 가 예시적 비한정적인 예들로서 전화기 또는 태블릿과 같은 모바일 디바이스 (1202) 를 포함하는 구현 (1200) 을 도시한다. 모바일 디바이스 (1202) 는 마이크로폰 (246) 및 디스플레이 스크린 (1204) 을 포함한다. 어플리케이션들 (1122) 을 포함하는 하나 이상의 프로세서들 (220) 의 컴포넌트들은 모바일 디바이스 (1202) 에 통합되고, 모바일 디바이스 (1202) 의 사용자에게 일반적으로 보이지 않는 내부 컴포넌트들을 표시하기 위해 점선들을 사용하여 예시된다. 특정 예에서, 어플리케이션들 (1122) 의 피쳐 추출기 (222), 세그먼트화기 (124), 및 프로파일 관리기 (126) 는 사용자 스피치 프로파일들을 관리하도록 동작하며, 이들은, 그 다음, 그래픽 사용자 인터페이스를 런칭하기 위한 것 또는 그렇지 않으면 (예를 들어, 통합 "스마트 어시스턴트" 어플리케이션을 통해) 디스플레이 스크린 (1204) 에서 사용자의 스피치와 연관된 다른 정보 (예를 들어, 대화 트랜스크립트) 를 디스플레이하기 위한 것과 같이 모바일 디바이스 (1202) 에서 하나 이상의 동작들을 수행하기 위해 사용된다.

[0137] 도 13 은, 디바이스 (202) 가 헤드셋 디바이스 (1302) 를 포함하는 구현 (1300) 을 도시한다. 헤드셋 디바이스 (1302) 는 마이크로폰 (246) 을 포함한다. 어플리케이션들 (1122) 을 포함하는 하나 이상의 프로세서들 (220) 의 컴포넌트들은 헤드셋 디바이스 (1302) 에 통합된다. 특정 예에서, 어플리케이션들 (1122) 의 피쳐 추출기 (222), 세그먼트화기 (124), 및 프로파일 관리기 (126) 는 사용자 스피치 프로파일들을 관리하도록 동작하며, 이들은 헤드셋 디바이스 (1302) 로 하여금 추가의 프로세싱을 위해, 사용자 스피치에 대응하는 정보 (예를 들어, 도 2b 의 프로파일 업데이트 데이터 (272), 사용자 상호작용 데이터 (274), 또는 그 양자 모두) 를 제 2 디바이스 (도시 안됨) 로 송신하는 것과 같이, 헤드셋 디바이스 (1302) 에서 하나 이상의 동작들을 수행하게 할 수도 있다.

[0138] 도 14 는, 디바이스 (202) 가 "스마트 워치" 로서 예시된 웨어러블 전자 디바이스 (1402) 를 포함하는 구현 (1400) 을 도시한다. 어플리케이션들 (1122) 및 마이크로폰 (246) 은 웨어러블 전자 디바이스 (1402) 에 통합된다. 특정 예에서, 어플리케이션들 (1122) 의 피쳐 추출기 (222), 세그먼트화기 (124), 및 프로파일 관리기 (126) 는 사용자 스피치 프로파일들을 관리하도록 동작하며, 이들은, 그 다음, 그래픽 사용자 인터페이스를 런칭하기 위한 것 또는 그렇지 않으면 웨어러블 전자 디바이스 (1402) 의 디스플레이 스크린 (1404) 에서 사용자의 스피치와 연관된 다른 정보를 디스플레이하기 위한 것과 같이 웨어러블 전자 디바이스 (1402) 에서 하나 이상의 동작들을 수행하기 위해 사용된다. 예시하자면, 웨어러블 전자 디바이스 (1402) 는, 웨어러블 전자 디바이스 (1402) 에 의해 검출된 사용자 스피치에 기초하여 통지 (예컨대, 캘린더 이벤트를 추가하기 위한 옵션) 를 디스플레이하도록 구성되는 디스플레이 스크린 (1404) 을 포함할 수도 있다. 특정 예에서, 웨어러블 전자 디바이스 (1402) 는 사용자 스피치의 검출에 응답하여 햅틱 통지를 제공하는 (예를 들어, 진동하는) 햅틱 디바이스를 포함한다. 예를 들어, 햅틱 통지는 사용자로 하여금 웨어러블 전자 디바이스 (1402) 를 보게 하여, 사용자가 말한 키워드의 검출을 표시하는 디스플레이된 통지를 보게 할 수 있다. 따라서, 웨어러블 전자 디바이스 (1402) 는 청각 장애를 가진 사용자 또는 헤드셋을 착용한 사용자에게, 사용자의 스피치가 검출됨을 알릴 수 있다. 특정 예에서, 웨어러블 전자 디바이스 (1402) 는 스피치의 검출에 응답하여 대화의 트랜스크립트를 디스플레이할 수 있다.

[0139] 도 15 는, 디바이스 (202) 가 무선 스피커 및 음성 활성화 디바이스 (1502) 를 포함하는 구현 (1500) 이다. 무선 스피커 및 음성 활성화 디바이스 (1502) 는 무선 네트워크 접속성을 가질 수 있고 어시스턴트 동작을 실행하도록 구성된다. 어플리케이션들 (1122) 을 포함하는 하나 이상의 프로세서들 (220), 마이크로폰 (246), 또는 이들의 조합은 무선 스피커 및 음성 활성화 디바이스 (1502) 에 포함된다. 무선 스피커 및 음성 활성화 디바이스 (1502) 는 또한, 스피커 (1504) 를 포함한다. 동작 동안, 어플리케이션들 (1122) 의 피쳐 추출

기 (222), 세그먼트화기 (124), 및 프로파일 관리기 (126)의 동작을 통해 사용자 스피치 프로파일 (150A)과 연관된 사용자의 사용자 스피치로서 식별된 구두 커맨드를 수신하는 것에 응답하여, 무선 스피커 및 음성 활성화 디바이스 (1502)는, 예컨대, 음성 활성화 시스템 (예를 들어, 통합 어시스턴트 어플리케이션)의 실행을 통해 어시스턴트 동작들을 실행할 수 있다. 어시스턴트 동작들은 온도를 조정하는 것, 음악을 재생하는 것, 조명을 턴온하는 것 등을 포함할 수 있다. 예를 들어, 어시스턴트 동작들은 키워드 또는 핵심 구문 (예를 들어, "헬로 어시스턴트 (hello assistant)") 이후 커맨드를 수신하는 것에 응답하여 수행된다. 특정 양태에서, 어시스턴트 동작들은, 사용자 스피치 프로파일 (150A)과 연관된 사용자에 대한 사용자 특정 커맨드 (예를 들어, "내일 오후 2시에 내 캘린더에서 약속을 설정할 것" 또는 "내 방에서 난방을 증가시킬 것")를 수행하는 것을 포함한다.

[0140] 도 16은 디바이스 (202)가 가상 현실, 증강 현실, 또는 혼합 현실 헤드셋 (1602)에 대응하는 휴대용 전자 디바이스를 포함하는 구현 (1600)을 도시한다. 어플리케이션들 (1122), 마이크로폰 (246), 또는 이들의 조합은 헤드셋 (1602)에 통합된다. 시각적 인터페이스 디바이스 (1620)는, 헤드셋 (1602)이 착용되는 동안 사용자로서의 증강 현실 또는 가상 현실 이미지들 또는 장면들의 디스플레이를 가능케 하기 위해 사용자의 눈 앞에 포지셔닝된다. 특정 예에서, 시각적 인터페이스 디바이스는, 마이크로폰 (246)으로부터 수신된 오디오 신호에서 검출된 사용자 스피치를 표시하는 통지를 디스플레이하도록 구성된다. 특정 양태에서, 시각적 인터페이스 디바이스는, 마이크로폰 (246)에 의해 캡처된 대화의 대화 트랜스크립트를 디스플레이하도록 구성된다.

[0141] 도 17은 디바이스 (202)가 유인 또는 무인 항공 디바이스 (예를 들어, 수화물 배송 드론)으로서 예시되는 비히클 (1702)에 대응하거나 비히클 내에 통합되는 구현 (1700)을 도시한다. 어플리케이션들 (1122), 마이크로폰 (246), 또는 이들의 조합은 비히클 (1702)에 통합된다. 스피치 분석은, 예컨대, 마이크로폰 (246)에 의해 캡처된 대화의 트랜스크립트를 생성하기 위해, 비히클 (1702)의 마이크로폰 (246)으로부터 수신된 오디오 신호들에 기초하여 수행될 수 있다.

[0142] 도 18은 디바이스 (202)가 자동차로서 예시된 비히클 (1802)에 대응하거나 비히클 내에 통합되는 다른 구현 (1800)을 도시한다. 비히클 (1802)은, 어플리케이션들 (1122)을 포함하는 하나 이상의 프로세서들 (220)을 포함한다. 비히클 (1802)은 또한, 마이크로폰 (246)을 포함한다. 마이크로폰 (246)은 비히클 (1802)의 하나 이상의 승객들의 발언들을 캡처하도록 포지셔닝된다. 사용자 스피치 분석은, 비히클 (1802)의 마이크로폰 (246)으로부터 수신된 오디오 신호들에 기초하여 수행될 수 있다. 일부 구현들에서, 사용자 스피치 분석은, 비히클 (1802)의 승객들 사이의 대화와 같은 내부 마이크로폰들 (예를 들어, 마이크로폰 (246))로부터 수신된 오디오 신호에 기초하여 수행될 수 있다. 예를 들어, 사용자 스피치 분석은, 비히클 (1802)에서 검출된 대화에 기초하여 특정 사용자 스피치 프로파일들과 연관된 사용자들에 대한 캘린더 이벤트를 설정하는데 사용될 수 있다 (예컨대, "토요일 오후에 피크닉 가자" 및 "물론, 그것 좋겠다"). 일부 구현들에서, 사용자 스피치 분석은, 사용자들이 비히클 (1802) 외부에서 말하는 것과 같은 외부 마이크로폰들 (예를 들어, 마이크로폰 (246))로부터 수신된 오디오 신호에 기초하여 수행될 수 있다. 특정 구현에서, 특정 스피치 프로파일들과 연관된 사용자들 사이의 특정 대화를 검출하는 것에 응답하여, 어플리케이션들 (1122)은, 디스플레이 (1820) 또는 하나 이상의 스피커들 (예를 들어, 스피커 (1830))을 통해, 예컨대, 피드백 또는 정보 (예를 들어, "사용자 1은 토요일 오후 3시까지 사전 약속 있음, 오후 4시에 피크닉 스케줄?")를 제공함으로써, 검출된 대화, 검출된 사용자들, 또는 그 양자 모두에 기초하여 비히클 (1802)의 하나 이상의 동작들을 개시한다.

[0143] 도 19를 참조하면, 디바이스의 특정 예시적인 구현의 블록 다이어그램이 도시되고 일반적으로 1900으로 지정된다. 다양한 구현들에 있어서, 디바이스 (1900)는 도 19에 예시된 것들보다 더 많거나 더 적은 컴포넌트들을 가질 수도 있다. 예시적인 구현에 있어서, 디바이스 (1900)는 디바이스 (202)에 대응할 수도 있다. 예시적인 구현에 있어서, 디바이스 (1900)는 도 1 내지 도 18을 참조하여 설명된 하나 이상의 동작들을 수행할 수도 있다.

[0144] 특정 구현에 있어서, 디바이스 (1900)는 프로세서 (1906) (예컨대, 중앙 프로세싱 유닛 (CPU))를 포함한다. 디바이스 (1900)는 하나 이상의 추가적인 프로세서들 (1910) (예컨대, 하나 이상의 DSP들)을 포함할 수도 있다. 특정 양태에서, 도 2a의 하나 이상의 프로세서들 (220)은 프로세서 (1906), 프로세서들 (1910), 또는 이들의 조합에 대응한다. 프로세서들 (1910)은 피쳐 추출기 (222), 화자 검출기 (278), 세그먼트화기 (124), 프로파일 관리기 (126), 하나 이상의 오디오 분석 어플리케이션들 (180), 또는 이들의 조합을 포함할 수

도 있다.

- [0145] 디바이스 (1900) 는 메모리 (1986) 및 코텍 (1934) 을 포함할 수도 있다. 특정 양태에서, 메모리 (1986) 은 도 2a 의 메모리 (232) 에 대응한다. 메모리 (1986) 는, 피쳐 추출기 (222), 화자 검출기 (278), 세그먼트화기 (124), 프로파일 관리기 (126), 하나 이상의 오디오 분석 어플리케이션들 (180), 또는 이들의 조합을 참조하여 설명된 기능성을 구현하기 위해 하나 이상의 추가적인 프로세서들 (1910) (또는 프로세서 (1906)) 에 의해 실행가능한 명령들 (1956) 을 포함할 수도 있다. 디바이스 (1900) 는, 트랜시버 (1950) 를 통해, 안테나 (1952) 에 커플링된 무선 제어기 (1940) 를 포함할 수도 있다. 특정 양태에서, 디바이스 (1900) 는 트랜시버 (1950) 에 커플링된 모델을 포함한다.
- [0146] 디바이스 (1900) 는 디스플레이 제어기 (1926) 에 커플링된 디스플레이 (1928) 를 포함할 수도 있다. 하나 이상의 스피커들 (1992), 마이크로폰 (246), 또는 이들의 조합은 코텍 (1934) 에 커플링될 수도 있다. 코텍 (1934) 은 디지털-투-아날로그 컨버터 (DAC) (1902), 아날로그-투-디지털 컨버터 (ADC) (1904), 또는 그 양자 모두를 포함할 수도 있다. 특정 구현에서, 코텍 (1934) 은 마이크로폰 (246) 으로부터 아날로그 신호들을 수신하고, 아날로그-투-디지털 컨버터 (1904) 를 사용하여 아날로그 신호들을 디지털 신호들로 변환하고, 디지털 신호들을 하나 이상의 프로세서들 (1910) 에 제공할 수도 있다. 하나 이상의 프로세서들 (1910) 은 디지털 신호들을 프로세싱할 수도 있다. 특정 구현에서, 하나 이상의 프로세서들 (1910) 은 디지털 신호들을 코텍 (1934) 에 제공할 수도 있다. 코텍 (1934) 은 디지털 신호들을 디지털-투-아날로그 컨버터 (1902) 를 사용하여 아날로그 신호들로 변환할 수도 있고 아날로그 신호들을 스피커 (1992) 에 제공할 수도 있다.
- [0147] 특정 구현에서, 디바이스 (1900) 는 시스템-인-패키지 또는 시스템-온-칩 디바이스 (1922) 에 포함될 수도 있다. 특정 구현에서, 메모리 (1986), 프로세서 (1906), 프로세서들 (1910), 디스플레이 제어기 (1926), 코텍 (1934), 무선 제어기 (1940), 및 트랜시버 (1950) 가 시스템-인-패키지 또는 시스템-온-칩 디바이스 (1922) 에 포함된다. 특정 구현에서, 입력 디바이스 (1930) 및 전력 공급부 (1944) 가 시스템-온-칩 디바이스 (1922) 에 커플링된다. 더욱이, 특정 구현에서, 도 19 에 예시된 바와 같이, 디스플레이 (1928), 입력 디바이스 (1930), 스피커 (1992), 마이크로폰 (246), 안테나 (1952), 및 전력 공급부 (1944) 는 시스템-온-칩 디바이스 (1922) 외부에 있다. 특정 구현에서, 디스플레이 (1928), 입력 디바이스 (1930), 스피커 (1992), 마이크로폰 (246), 안테나 (1952), 및 전력 공급부 (1944) 의 각각은, 인터페이스 또는 제어기와 같은 시스템-온-칩 디바이스 (1922) 의 컴포넌트에 커플링될 수도 있다.
- [0148] 디바이스 (1900) 는 스마트 스피커, 스피커 바, 모바일 통신 디바이스, 스마트 폰, 셀룰러 폰, 랩탑 컴퓨터, 컴퓨터, 태블릿, 개인용 디지털 보조기, 디스플레이 디바이스, 텔레비전, 게이밍 콘솔, 뮤직 플레이어, 라디오, 디지털 비디오 플레이어, 디지털 비디오 디스크 (DVD) 플레이어, 튜너, 카메라, 내비게이션 디바이스, 비히클, 헤드셋, 증강 현실 헤드셋, 가상 현실 헤드셋, 항공 비히클, 홈 자동화 시스템, 음성 활성화 디바이스, 무선 스피커 및 음성 활성화 디바이스, 휴대용 전자 디바이스, 자동차, 컴퓨팅 디바이스, 통신 디바이스, 사물 인터넷 (IoT) 디바이스, 가상 현실 (VR) 디바이스, 기지국, 모바일 디바이스, 또는 이들의 임의의 조합을 포함할 수도 있다.
- [0149] 설명된 구현들과 함께, 장치는 복수의 사용자들의 복수의 사용자 스피치 프로파일들을 저장하는 수단을 포함한다. 예를 들어, 저장하는 수단은 도 2a 의 메모리 (232), 디바이스 (202), 시스템 (200), 메모리 (1986), 디바이스 (1900), 복수의 사용자 스피치 프로파일들을 저장하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합을 포함한다.
- [0150] 그 장치는, 제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하는 수단을 더 포함한다. 예를 들어, 결정하는 수단은 도 2a 의 화자 검출기 (278), 하나 이상의 프로세서들 (220), 디바이스 (202), 시스템 (200), 프로세서 (1906), 하나 이상의 프로세서들 (1910), 디바이스 (1900), 제 1 전력 모드에서 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합을 포함한다.
- [0151] 그 장치는 또한, 세그먼트화 결과를 생성하기 위해 오디오 스트림의 오디오 피쳐 데이터를 분석하는 수단을 포함한다. 예를 들어, 분석하는 수단은 도 2a 의 세그먼트화기 (124), 하나 이상의 프로세서들 (220), 디바이스 (202), 시스템 (200), 프로세서 (1906), 하나 이상의 프로세서들 (1910), 디바이스 (1900), 오디오 피쳐 데이터를 분석하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합을 포함한다. 세그먼트화 결과 (236) 는 오디오 스트림 (141) 의 화자-동일 오디오 세그먼트들을 표시한다.

- [0152] 그 장치는 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들 중의 제 1 오디오 피쳐 데이터 세트에 대한 복수의 사용자 스피치 프로파일들의 비교를, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하는 수단을 더 포함한다. 예를 들어, 비교를 수행하는 수단은 도 2a의 프로파일 관리기 (126), 하나 이상의 프로세서들 (220), 디바이스 (202), 시스템 (200), 프로세서 (1906), 하나 이상의 프로세서들 (1910), 디바이스 (1900), 비교를 수행하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합을 포함한다.
- [0153] 그 장치는 또한, 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하는 수단을 포함한다. 예를 들어, 제 1 사용자 스피치 프로파일을 생성하는 수단은 도 2a의 프로파일 관리기 (126), 하나 이상의 프로세서들 (220), 디바이스 (202), 시스템 (200), 프로세서 (1906), 하나 이상의 프로세서들 (1910), 디바이스 (1900), 제 1 사용자 스피치 프로파일을 생성하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합을 포함한다. 사용자 스피치 프로파일 (150A)은, 오디오 피쳐 데이터 세트 (252)가 복수의 사용자 스피치 프로파일들 (150) 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여 생성된다.
- [0154] 그 장치는 제 1 사용자 스피치 프로파일을 복수의 사용자 스피치 프로파일들에 추가하는 수단을 더 포함한다. 예를 들어, 제 1 사용자 스피치 프로파일을 추가하는 수단은 도 2a의 프로파일 관리기 (126), 하나 이상의 프로세서들 (220), 디바이스 (202), 시스템 (200), 프로세서 (1906), 하나 이상의 프로세서들 (1910), 디바이스 (1900), 제 1 사용자 스피치 프로파일을 추가하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합을 포함한다.
- [0155] 일부 구현들에서, 비밀시적 컴퓨터 판독가능 매체 (예를 들어, 메모리 (1986)와 같은 컴퓨터 판독가능 저장 디바이스)는, 하나 이상의 프로세서들 (예를 들어, 하나 이상의 프로세서들 (1910) 또는 프로세서 (1906))에 의해 실행될 경우, 하나 이상의 프로세서들로 하여금, 제 1 전력 모드 (예를 들어, 전력 모드 (282))에서, 오디오 스트림 (예를 들어, 오디오 스트림 (141))이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하게 하는 명령들 (예를 들어, 명령들 (1956))을 포함한다. 명령들은, 하나 이상의 프로세서들에 의해 실행될 경우, 또한, 프로세서들로 하여금, 세그먼트화 결과 (예를 들어, 세그먼트화 결과 (236))를 생성하기 위해 오디오 스트림의 오디오 피쳐 데이터 (예를 들어, 오디오 피쳐 데이터 세트들 (252))를 분석하게 한다. 세그먼트화 결과는 오디오 스트림의 화자-동일 오디오 세그먼트들 (예를 들어, 화자-동일 오디오 세그먼트 (111A) 및 화자-동일 오디오 세그먼트 (111B))을 표시한다. 명령들은, 하나 이상의 프로세서들에 의해 실행될 경우, 또한, 프로세서들로 하여금, 제 1 화자-동일 오디오 세그먼트 (예를 들어, 화자-동일 오디오 세그먼트 (111A))의 제 1 복수의 오디오 피쳐 데이터 세트들 (예를 들어, 오디오 피쳐 데이터 세트들 (252A)) 중의 제 1 오디오 피쳐 데이터 세트 (예를 들어, 오디오 피쳐 데이터 세트 (252))에 대한 복수의 사용자 스피치 프로파일들 (예를 들어, 복수의 사용자 스피치 프로파일들 (150))의 비교를, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하게 한다. 명령들은, 하나 이상의 프로세서들에 의해 실행될 경우, 추가로, 프로세서들로 하여금, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여: 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일 (예를 들어, 사용자 스피치 프로파일 (150A))을 생성하게 하고 제 1 사용자 스피치 프로파일을 복수의 사용자 스피치 프로파일들에 추가하게 한다.
- [0156] 본 개시의 특정 양태들은 상호관련된 조항들의 제 1 세트로 하기에서 기술된다:
- [0157] 조항 1에 따르면. 오디오 분석을 위한 디바이스는, 복수의 사용자들의 복수의 사용자 스피치 프로파일들을 저장하도록 구성된 메모리; 및 하나 이상의 프로세서들을 포함하고, 그 하나 이상의 프로세서들은, 제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하고; 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응함을 결정하는 것에 기초하여, 제 2 전력 모드에서, 세그먼트화 결과를 생성하기 위해 오디오 스트림의 오디오 피쳐 데이터를 분석하는 것으로서, 세그먼트화 결과는 오디오 스트림의 화자-동일 오디오 세그먼트들을 표시하는, 상기 오디오 스트림의 오디오 피쳐 데이터를 분석하고; 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들 중의 제 1 오디오 피쳐 데이터 세트에 대한 복수의 사용자 스피치 프로파일들의 비교를, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하고; 그리고 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결

정하는 것에 기초하여: 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하고; 제 1 사용자 스피치 프로파일을 복수의 사용자 스피치 프로파일들에 추가하도록 구성된다.

- [0158] 조항 2 는 조항 1 의 디바이스를 포함하고, 여기서, 제 1 오디오 피쳐 데이터 세트는 제 1 오디오 피쳐 벡터를 포함한다.
- [0159] 조항 3 은 조항 1 또는 조항 2 의 디바이스를 포함하고, 여기서, 하나 이상의 프로세서들은 스피커 세그먼트화 뉴럴 네트워크를 오디오 피쳐 데이터에 적용함으로써 오디오 피쳐 데이터를 분석하도록 구성된다.
- [0160] 조항 4 는 조항 1 내지 조항 3 중 어느 하나의 디바이스를 포함하고, 여기서, 하나 이상의 프로세서들은, 세그먼트화 결과가 제 1 오디오 피쳐 데이터 세트가 제 1 화자의 스피치에 대응한다는 것을 표시함을 그리고 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여: 제 1 화자와 연관된 제 1 등록 버퍼에 제 1 오디오 피쳐 데이터 세트들을 저장하고; 그리고 정지 조건이 만족될 때까지 제 1 등록 버퍼에 제 1 화자의 스피치에 대응하는 후속 오디오 피쳐 데이터 세트들을 저장하도록 구성되고, 여기서, 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들은 제 1 오디오 피쳐 데이터 세트 및 후속 오디오 피쳐 데이터 세트들을 포함한다.
- [0161] 조항 5 는 조항 4 의 디바이스를 포함하고, 여기서, 하나 이상의 프로세서들은 임계 폭음보다 긴 것이 오디오 스트림에서 검출됨을 결정하는 것에 응답하여 정지 조건이 만족됨을 결정하도록 구성된다.
- [0162] 조항 6 은 조항 4 내지 조항 5 중 어느 하나의 디바이스를 포함하고, 여기서, 하나 이상의 프로세서들은 특정 오디오 피쳐 데이터 세트가 단일 화자의 스피치에 대응함을 결정하는 것에 적어도 부분적으로 기초하여 제 1 등록 버퍼에 특정 오디오 피쳐 데이터 세트를 추가하도록 구성되고, 여기서, 단일 화자는 제 1 화자를 포함한다.
- [0163] 조항 7 은 조항 1 내지 조항 6 중 어느 하나의 디바이스를 포함하고, 여기서, 하나 이상의 프로세서들은, 제 1 등록 버퍼에 저장된 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들의 카운트가 등록 임계치보다 크다고 결정하는 것에 기초하여, 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하도록 구성된다.
- [0164] 조항 8 은 조항 1 내지 조항 7 중 어느 하나의 디바이스를 포함하고, 여기서, 하나 이상의 프로세서들은, 제 1 오디오 피쳐 데이터 세트가 특정 사용자 스피치 프로파일과 매칭함을 결정하는 것에 기초하여, 제 1 오디오 피쳐 데이터 세트에 기초하여 특정 사용자 스피치 프로파일을 업데이트하도록 구성된다.
- [0165] 조항 9 는 조항 8 의 디바이스를 포함하고, 여기서, 하나 이상의 프로세서들은, 제 1 오디오 피쳐 데이터 세트가 단일 화자의 스피치에 대응함을 결정하는 것에 적어도 부분적으로 기초하여, 제 1 오디오 피쳐 데이터 세트에 기초하여 특정 사용자 스피치 프로파일을 업데이트하도록 구성된다.
- [0166] 조항 10 은 조항 1 내지 조항 9 중 어느 하나의 디바이스를 포함하고, 여기서, 하나 이상의 프로세서들은 제 2 화자-동일 오디오 세그먼트의 제 2 복수의 오디오 피쳐 데이터 세트들 중의 제 2 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하도록 구성된다.
- [0167] 조항 11 은 조항 10 의 디바이스를 포함하고, 여기서, 하나 이상의 프로세서들은, 제 2 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여: 제 2 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 2 사용자 스피치 프로파일을 생성하고; 제 2 사용자 스피치 프로파일을 복수의 사용자 스피치 프로파일들에 추가하도록 구성된다.
- [0168] 조항 12 는 조항 10 의 디바이스를 포함하고, 여기서, 하나 이상의 프로세서들은, 제 2 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중의 특정 사용자 스피치 프로파일과 매칭함을 결정하는 것에 기초하여, 제 2 오디오 피쳐 데이터 세트에 기초하여 특정 사용자 스피치 프로파일을 업데이트하도록 구성된다.
- [0169] 조항 13 은 조항 1 내지 조항 12 중 어느 하나의 디바이스를 포함하고, 여기서, 메모리는 프로파일 업데이트 데이터를 저장하도록 구성되고, 하나 이상의 프로세서들은, 제 1 사용자 스피치 프로파일을 생성하는 것에 응답하여, 제 1 사용자 스피치 프로파일이 업데이트됨을 표시하기 위해 프로파일 업데이트 데이터를 업데이트하고; 그리고 복수의 사용자 스피치 프로파일들의 제 1 카운트가 업데이트되었음을 프로파일 업데이트 데이터가 표시함을 결정하는 것에 기초하여, 제 1 카운트를, 오디오 스트림에서 검출된 화자들의 카운트로서 출력하도록 구성된다.
- [0170] 조항 14 는 조항 1 내지 조항 13 중 어느 하나의 디바이스를 포함하고, 여기서, 메모리는 사용자 상호작용 데이

터를 저장하도록 구성되고, 하나 이상의 프로세서들은, 제 1 사용자 스피치 프로파일을 생성하는 것에 응답하여, 제 1 사용자 스피치 프로파일과 연관된 제 1 사용자가 스피치 지속기간 동안 상호작용하였음을 표시하기 위해 제 1 화자-동일 오디오 세그먼트의 스피치 지속기간에 기초하여 사용자 상호작용 데이터를 업데이트하고; 그리고 적어도 사용자 상호작용 데이터를 출력하도록 구성된다.

[0171] 조항 15 는 조항 1 내지 조항 14 중 어느 하나의 디바이스를 포함하고, 여기서, 제 1 전력 모드는 제 2 전력 모드에 비해 더 낮은 전력 모드이다.

[0172] 조항 16 은 조항 1 의 디바이스를 포함하고, 여기서, 하나 이상의 프로세서들은, 제 1 전력 모드에서, 오디오 스트림의 오디오 정보를 결정하는 것으로서, 오디오 정보는 오디오 스트림에서 검출된 화자들의 카운트, 음성 활성화도 검출 (VAD) 정보, 또는 그 양자 모두를 포함하는, 상기 오디오 스트림의 오디오 정보를 결정하고; 제 2 전력 모드에서 하나 이상의 오디오 분석 어플리케이션들을 활성화하고; 그리고 오디오 정보를 하나 이상의 오디오 분석 어플리케이션들에 제공하도록 구성된다.

[0173] 조항 17 은 조항 1 내지 조항 16 중 어느 하나의 디바이스를 포함하고, 여기서, 하나 이상의 프로세서들은, 오디오 스트림의 하나 이상의 제 2 오디오 세그먼트들이 다중 화자들에 대응한다는 것을 세그먼트화 결과가 표시함을 결정하는 것에 응답하여, 하나 이상의 제 2 오디오 세그먼트들에 기초하여 복수의 사용자 스피치 프로파일들을 업데이트하는 것을 억제하도록 구성된다.

[0174] 본 개시의 특정 양태들은 상호관련된 조항들의 제 2 세트로 하기에서 기술된다:

[0175] 조항 18 에 따르면, 오디오 분석의 방법은, 디바이스에서의 제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하는 단계; 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응함을 결정하는 것에 기초하여, 제 2 전력 모드에서, 세그먼트화 결과를 생성하기 위해 오디오 스트림의 오디오 피쳐 데이터를 분석하는 단계로서, 세그먼트화 결과는 오디오 스트림의 화자-동일 오디오 세그먼트들을 표시하는, 상기 오디오 스트림의 오디오 피쳐 데이터를 분석하는 단계; 디바이스에서, 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들 중의 제 1 오디오 피쳐 데이터 세트에 대한 복수의 사용자 스피치 프로파일들의 비교를, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하는 단계; 및 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여: 디바이스에서, 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하는 단계; 및 디바이스에서, 제 1 사용자 스피치 프로파일을 복수의 사용자 스피치 프로파일들에 추가하는 단계를 포함한다.

[0176] 조항 19 는 조항 18 의 방법을 포함하고, 스피커 세그먼트화 뉴럴 네트워크를 오디오 피쳐 데이터에 적용하는 단계를 더 포함한다.

[0177] 조항 20 은 조항 18 또는 조항 19 의 방법을 포함하고, 세그먼트화 결과가 제 1 오디오 피쳐 데이터 세트가 제 1 화자의 스피치에 대응한다는 것을 표시함을 그리고 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여: 제 1 화자와 연관된 제 1 등록 버퍼에 제 1 오디오 피쳐 데이터 세트를 저장하는 단계; 및 정지 조건이 만족될 때까지 제 1 등록 버퍼에 제 1 화자의 스피치에 대응하는 후속 오디오 피쳐 데이터 세트들을 저장하는 단계를 더 포함하고, 여기서, 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들은 제 1 오디오 피쳐 데이터 세트 및 후속 오디오 피쳐 데이터 세트들을 포함한다.

[0178] 조항 21 은 조항 20 의 방법을 포함하고, 디바이스에서, 임계 묵음보다 긴 것이 오디오 스트림에서 검출됨을 결정하는 것에 응답하여 정지 조건이 만족됨을 결정하는 단계를 더 포함한다.

[0179] 조항 22 는 조항 20 또는 조항 21 의 방법을 포함하고, 디바이스에서, 특정 오디오 피쳐 데이터 세트가 단일 화자의 스피치에 대응함을 결정하는 것에 적어도 부분적으로 기초하여 제 1 등록 버퍼에 특정 오디오 피쳐 데이터 세트를 추가하는 단계를 더 포함하고, 여기서, 단일 화자는 제 1 화자를 포함한다.

[0180] 조항 23 은 조항 18 내지 조항 22 중 어느 하나의 방법을 포함하고, 제 1 등록 버퍼에 저장된 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들의 카운트가 등록 임계치보다 크다고 결정하는 것에 기초하여, 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하는 단계를 더 포함한다.

- [0181] 조항 24 는 조항 18 내지 조항 23 중 어느 하나의 방법을 포함하고, 제 1 오디오 피쳐 데이터 세트가 특정 사용자 스피치 프로파일과 매칭함을 결정하는 것에 기초하여, 제 1 오디오 피쳐 데이터 세트에 기초하여 특정 사용자 스피치 프로파일을 업데이트하는 단계를 더 포함한다.
- [0182] 조항 25 는 조항 24 의 방법을 포함하고, 제 1 오디오 피쳐 데이터 세트가 단일 화자의 스피치에 대응함을 결정하는 것에 적어도 부분적으로 기초하여, 제 1 오디오 피쳐 데이터 세트에 기초하여 특정 사용자 스피치 프로파일을 업데이트하는 단계를 더 포함한다.
- [0183] 조항 26 은 조항 18 내지 조항 25 중 어느 하나의 방법을 포함하고, 제 2 화자-동일 오디오 세그먼트의 제 2 복수의 오디오 피쳐 데이터 세트들 중의 제 2 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중의 특정 사용자 스피치 프로파일과 매칭함을 결정하는 것에 기초하여, 제 2 오디오 피쳐 데이터 세트에 기초하여 특정 사용자 스피치 프로파일을 업데이트하는 단계를 더 포함한다.
- [0184] 본 개시의 특정 양태들은 상호관련된 조항들의 제 3 세트로 하기에서 기술된다:
- [0185] 조항 27 에 따르면, 비밀시적 컴퓨터 판독가능 저장 매체는 명령들을 저장하고, 그 명령들은, 하나 이상의 프로세서들에 의해 실행될 경우, 프로세서들로 하여금, 제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하게 하고; 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응함을 결정하는 것에 기초하여, 제 2 전력 모드에서, 세그먼트화 결과를 생성하기 위해 오디오 스트림의 오디오 피쳐 데이터를 분석하게 하는 것으로서, 세그먼트화 결과는 오디오 스트림의 화자-동일 오디오 세그먼트들을 표시하는, 상기 오디오 스트림의 오디오 피쳐 데이터를 분석하게 하고; 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들 중의 제 1 오디오 피쳐 데이터 세트에 대한 복수의 사용자 스피치 프로파일들의 비교를, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하게 하고; 그리고 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여: 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하게 하고; 제 1 사용자 스피치 프로파일을 복수의 사용자 스피치 프로파일들에 추가하게 한다.
- [0186] 조항 28 은 조항 27 의 비밀시적 컴퓨터 판독가능 저장 매체를 포함하고, 여기서, 그 명령들은, 하나 이상의 프로세서들에 의해 실행될 경우, 프로세서들로 하여금, 제 1 등록 버퍼에 저장된 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들의 카운트가 등록 임계치보다 크다고 결정하는 것에 기초하여, 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하게 한다.
- [0187] 본 개시의 특정 양태들은 상호관련된 조항들의 제 4 세트로 하기에서 기술된다:
- [0188] 조항 29 에 따르면, 장치는 복수의 사용자들의 복수의 사용자 스피치 프로파일들을 저장하는 수단; 제 1 전력 모드에서, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응하는지 여부를 결정하는 수단; 제 2 전력 모드에서, 세그먼트화 결과를 생성하기 위해 오디오 스트림의 오디오 피쳐 데이터를 분석하는 수단으로서, 오디오 피쳐 데이터는, 오디오 스트림이 적어도 2명의 별개의 화자들의 스피치에 대응함을 결정하는 것에 기초하여 제 2 전력 모드에서 분석되고, 세그먼트화 결과는 오디오 스트림의 화자-동일 오디오 세그먼트들을 표시하는, 상기 오디오 스트림의 오디오 피쳐 데이터를 분석하는 수단; 제 1 화자-동일 오디오 세그먼트의 제 1 복수의 오디오 피쳐 데이터 세트들 중의 제 1 오디오 피쳐 데이터 세트에 대한 복수의 사용자 스피치 프로파일들의 비교를, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하는지 여부를 결정하기 위해 수행하는 수단; 제 1 복수의 오디오 피쳐 데이터 세트들에 기초하여 제 1 사용자 스피치 프로파일을 생성하는 수단으로서, 제 1 사용자 스피치 프로파일은, 제 1 오디오 피쳐 데이터 세트가 복수의 사용자 스피치 프로파일들 중 임의의 사용자 스피치 프로파일과 매칭하지 않음을 결정하는 것에 기초하여 생성되는, 상기 제 1 사용자 스피치 프로파일을 생성하는 수단; 및 제 1 사용자 스피치 프로파일을 복수의 사용자 스피치 프로파일들에 추가하는 수단을 포함한다.
- [0189] 조항 30 은 조항 29 의 장치를 포함하고, 여기서, 저장하는 수단, 결정하는 수단, 분석하는 수단, 수행하는 수단, 생성하는 수단, 및 추가하는 수단은 모바일 통신 디바이스, 스마트 폰, 셀룰러 폰, 스마트 스피커, 스피커 바, 랩탑 컴퓨터, 컴퓨터, 태블릿, 개인용 디지털 보조기, 디스플레이 디바이스, 텔레비전, 게이밍 콘솔, 뮤직 플레이어, 라디오, 디지털 비디오 플레이어, 디지털 비디오 디스크 (DVD) 플레이어, 튜너, 카메라, 내비게이션 디바이스, 비히클, 헤드셋, 증강 현실 헤드셋, 가상 현실 헤드셋, 항공 비히클, 홈 자동화 시스템, 음성 활성화 디바이스, 무선 스피커 및 음성 활성화 디바이스, 휴대용 전자 디바이스, 자동차, 컴퓨팅 디바이스, 통신 디바

이스, 사물 인터넷 (IoT) 디바이스, 가상 현실 (VR) 디바이스, 기지국, 모바일 디바이스, 또는 이들의 임의의 조합 중 적어도 하나에 통합된다.

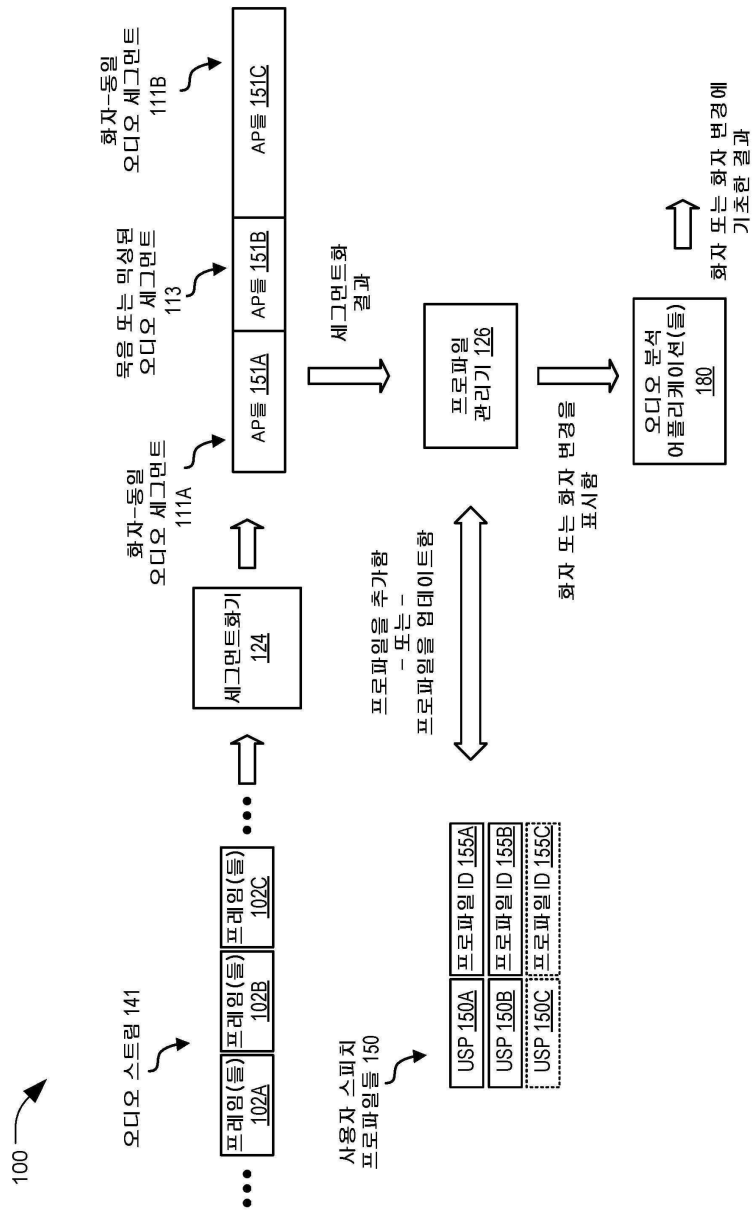
[0190] 당업자는 본 명세서에 개시된 구현들과 관련하여 설명된 다양한 예시적인 논리 블록들, 구성들, 모듈들, 회로들, 및 알고리즘 단계들이 전자 하드웨어, 프로세서에 의해 실행되는 컴퓨터 소프트웨어, 또는 이들 양자의 조합들로서 구현될 수도 있음을 추가로 인식할 것이다. 다양한 예시적인 컴포넌트들, 블록들, 구성들, 모듈들, 회로들 및 단계들이 일반적으로 그들의 기능성의 관점에서 상기 기술되었다. 그러한 기능성이 하드웨어로서 구현될지 또는 프로세서 실행가능 명령들로서 구현될지는 전체 시스템에 부과된 설계 제약들 및 특정 어플리케이션에 의존한다. 당업자는 설명된 기능성을 각각의 특정 어플리케이션에 대하여 다양한 방식으로 구현할 수도 있으며, 그러한 구현의 결정들은 본 개시의 범위로부터의 일탈을 야기하는 것으로서 해석되지는 않아야 한다.

[0191] 본 명세서에 개시된 구현들과 관련하여 설명된 방법 또는 알고리즘의 단계들은 하드웨어에서, 프로세서에 의해 실행되는 소프트웨어 모듈에서, 또는 이들 양자의 조합에서 직접 구현될 수도 있다. 소프트웨어 모듈은 랜덤 액세스 메모리 (RAM), 플래시 메모리, 판독 전용 메모리 (ROM), 프로그래밍가능 판독 전용 메모리 (PROM), 소거가능한 프로그래밍가능 판독 전용 메모리 (EPROM), 전기적으로 소거가능한 프로그래밍가능 판독 전용 메모리 (EEPROM), 레지스터들, 하드 디스크, 착탈가능 디스크, 콤팩트 디스크 판독 전용 메모리 (CD-ROM), 또는 당업계에 공지된 임의의 다른 형태의 비일시적 저장 매체에 상주할 수도 있다. 예시적인 저장 매체는, 프로세서가 저장 매체로부터 정보를 판독할 수도 있고 저장 매체에 정보를 기입할 수도 있도록 프로세서에 커플링된다. 대안적으로, 저장 매체는 프로세서에 통합될 수도 있다. 프로세서 및 저장 매체는 주문형 집적 회로 (ASIC) 에 상주할 수도 있다. ASIC 는 컴퓨팅 디바이스 또는 사용자 단말기에 상주할 수도 있다. 대안적으로, 프로세서 및 저장 매체는 컴퓨팅 디바이스 또는 사용자 단말기에 별개의 컴포넌트들로서 상주할 수도 있다.

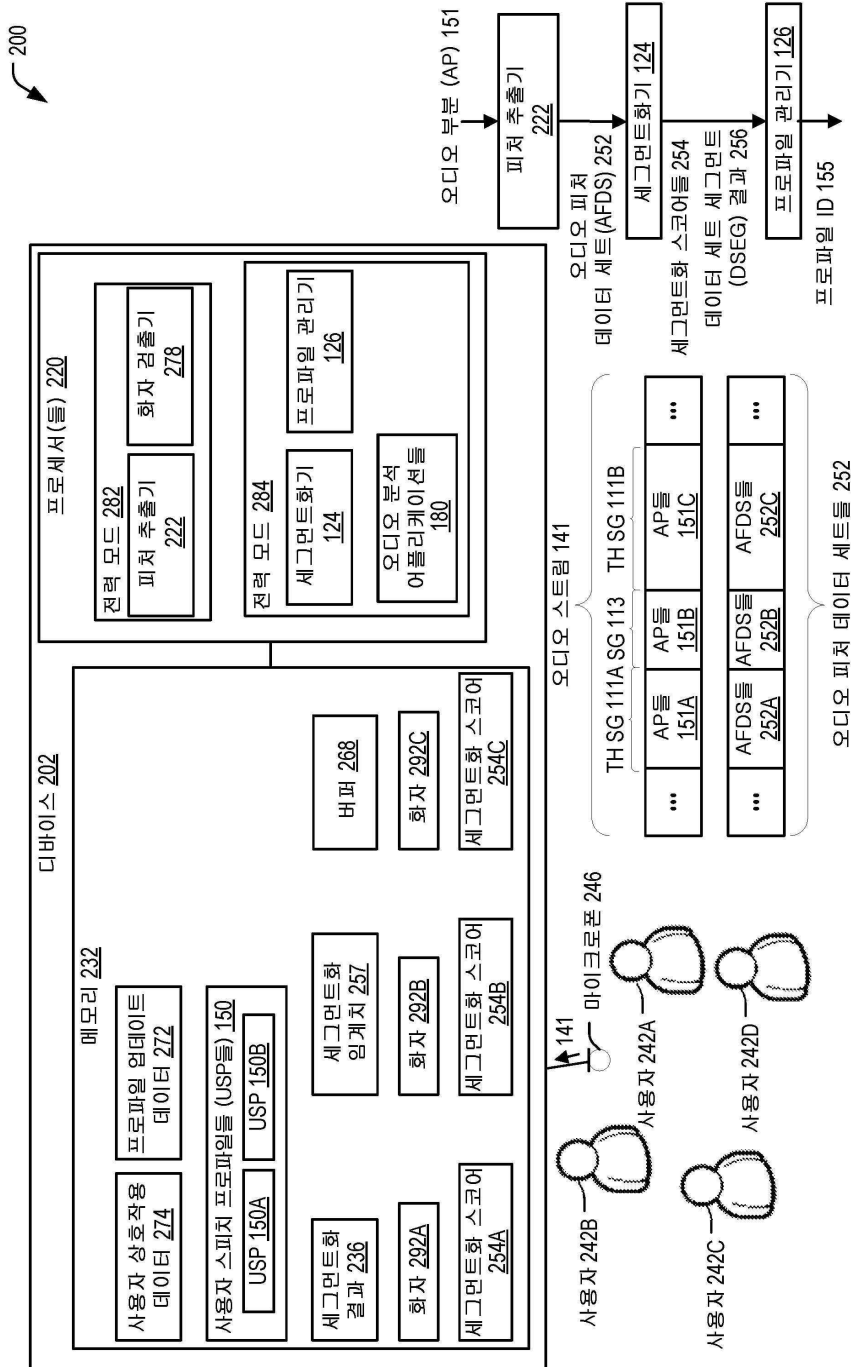
[0192] 개시된 양태들의 상기 설명은 당업자로 하여금 개시된 양태들을 제조 또는 이용할 수 있도록 제공된다. 이들 양태들에 대한 다양한 수정들은 당업자에게 용이하게 자명할 것이며, 본 명세서에서 정의된 원리들은 본 개시의 범위로부터 일탈함없이 다른 양태들에 적용될 수도 있다. 따라서, 본 개시는 본 명세서에서 설명된 양태들로 한정되도록 의도되지 않으며, 다음의 청구항들에 의해 정의된 바와 같은 원리들 및 신규한 특징들과 부합하는 가능한 최광의 범위를 부여받아야 한다.

도면

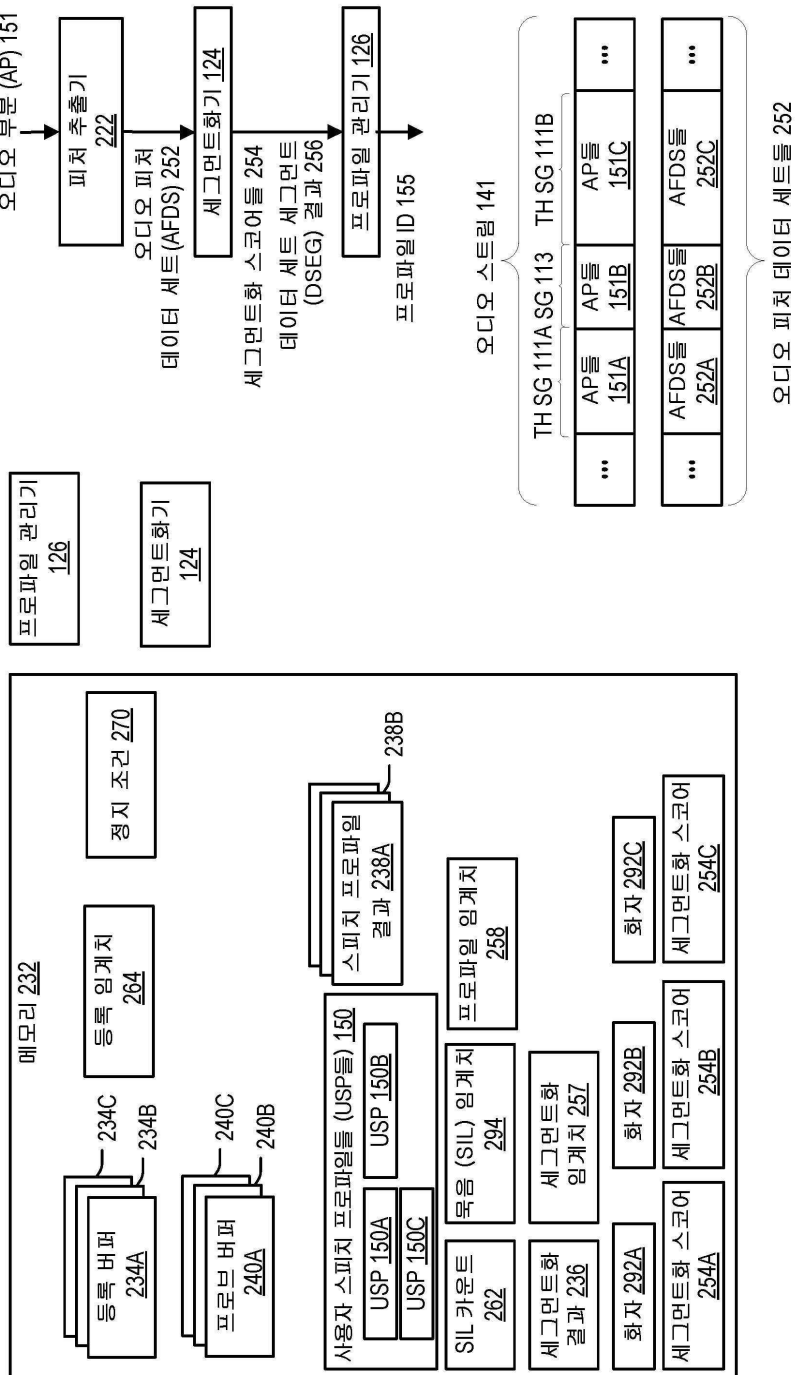
도면1



도면2a

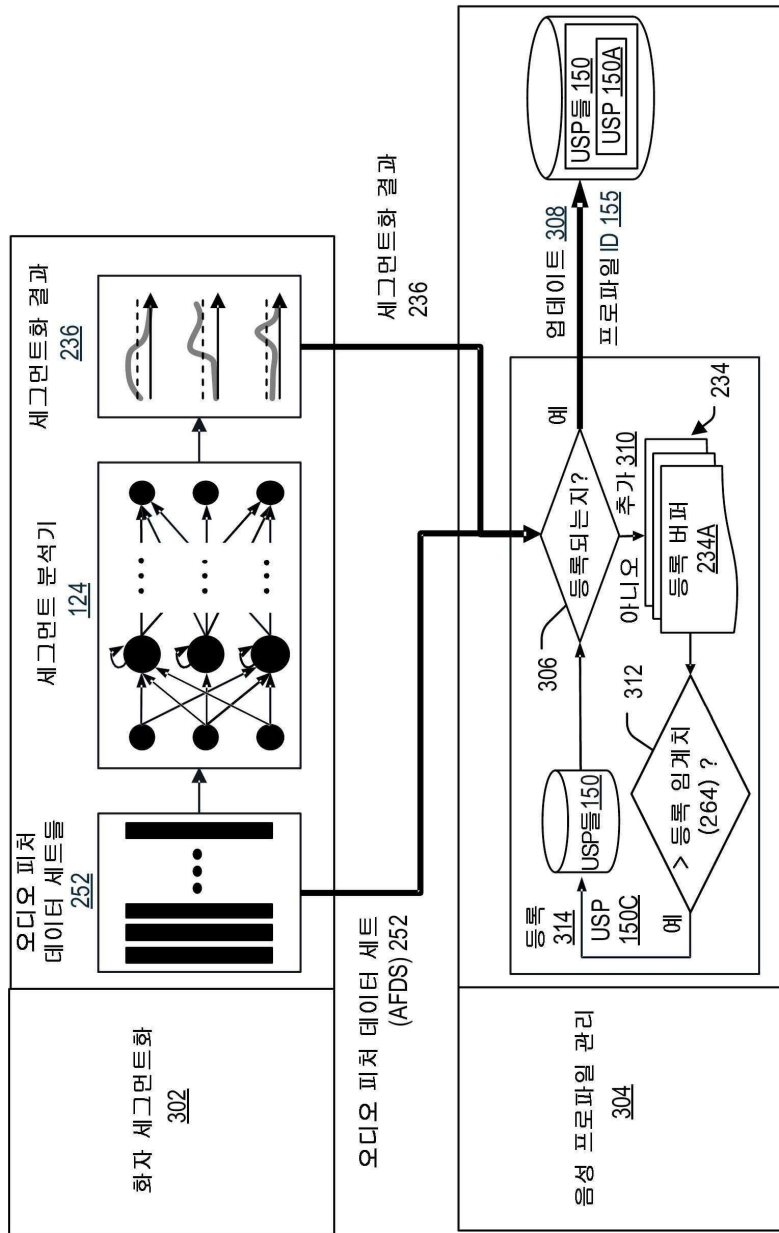


도면2b

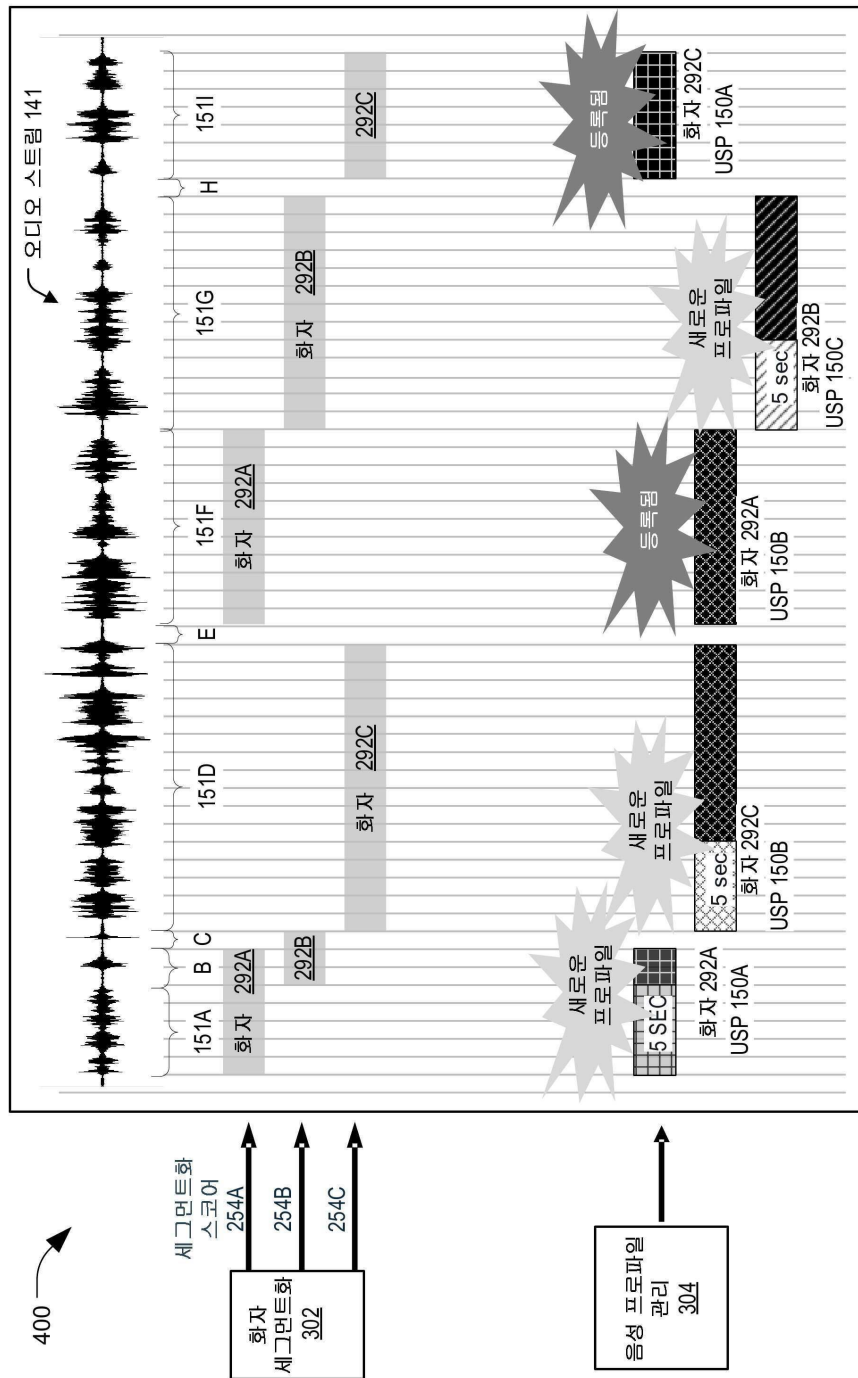


도면3

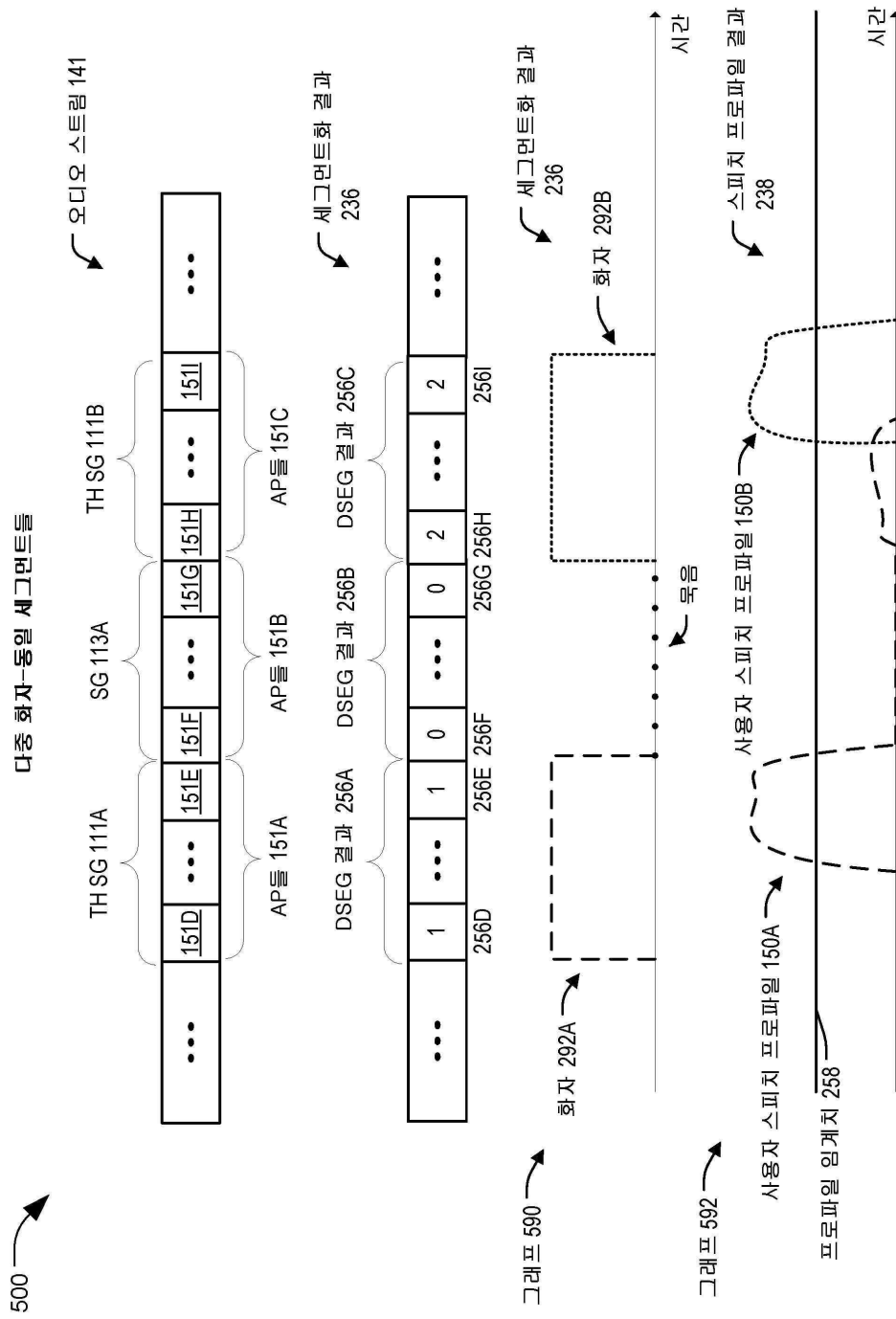
300 →



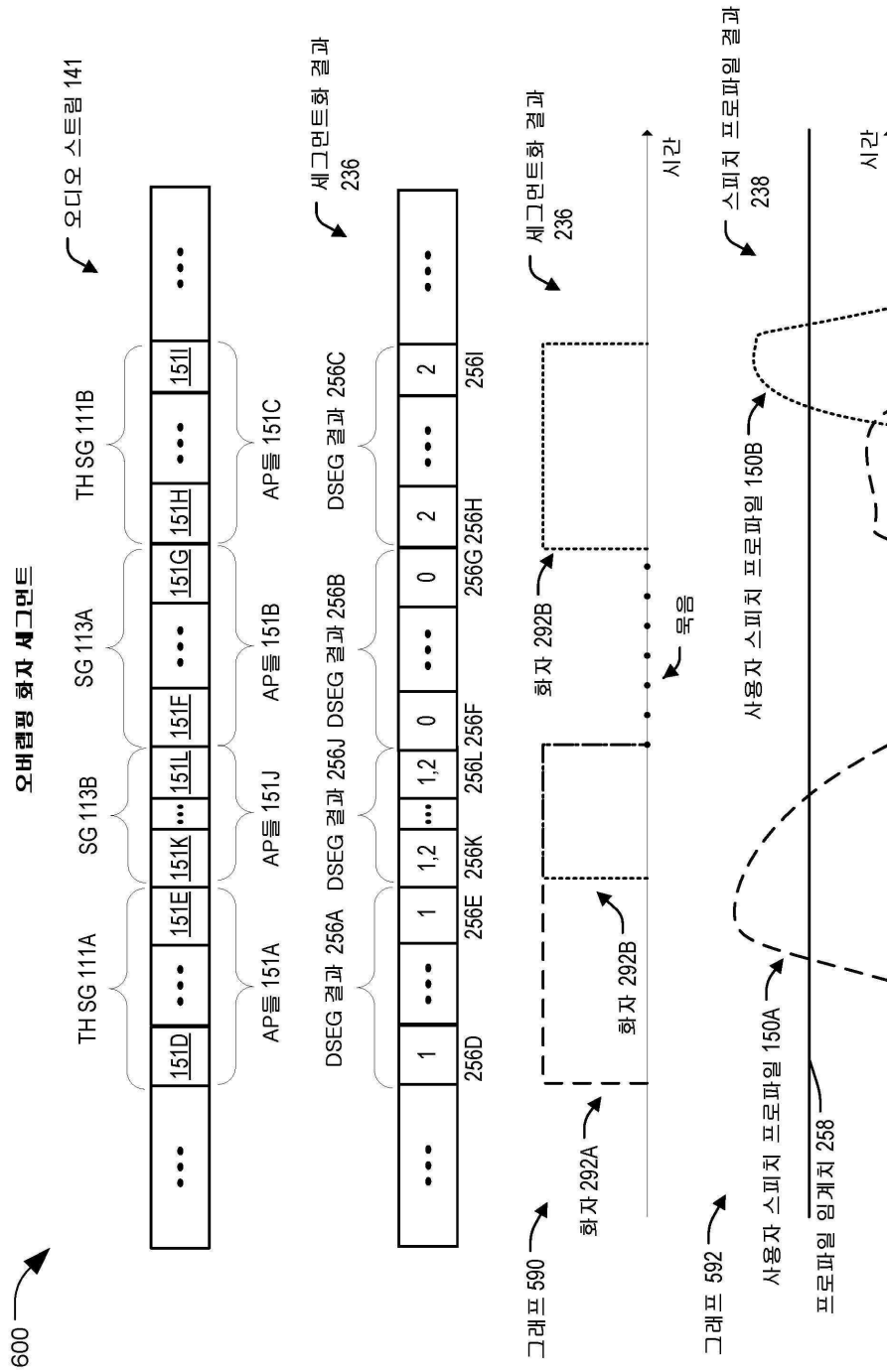
도면4



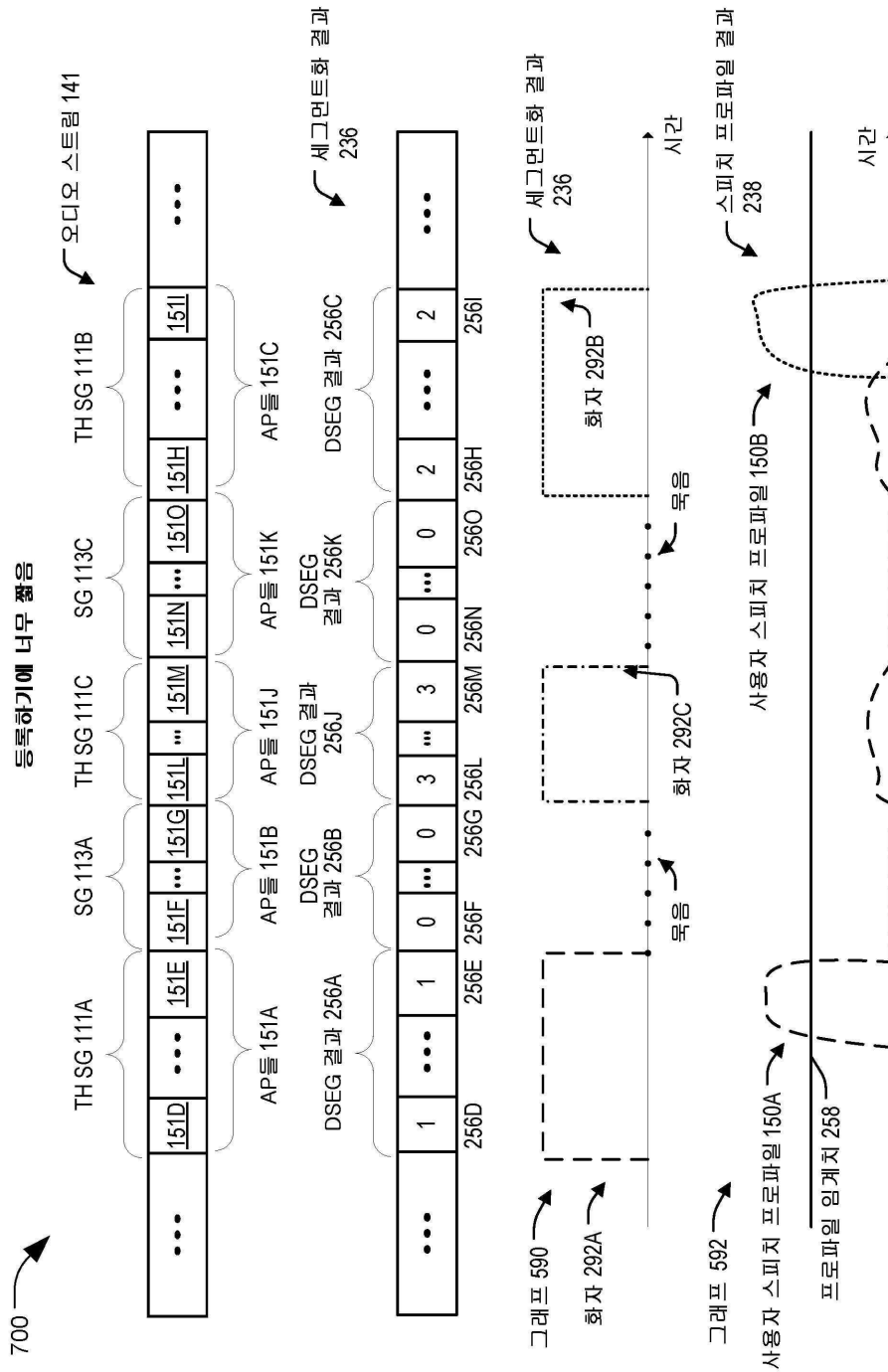
도면5



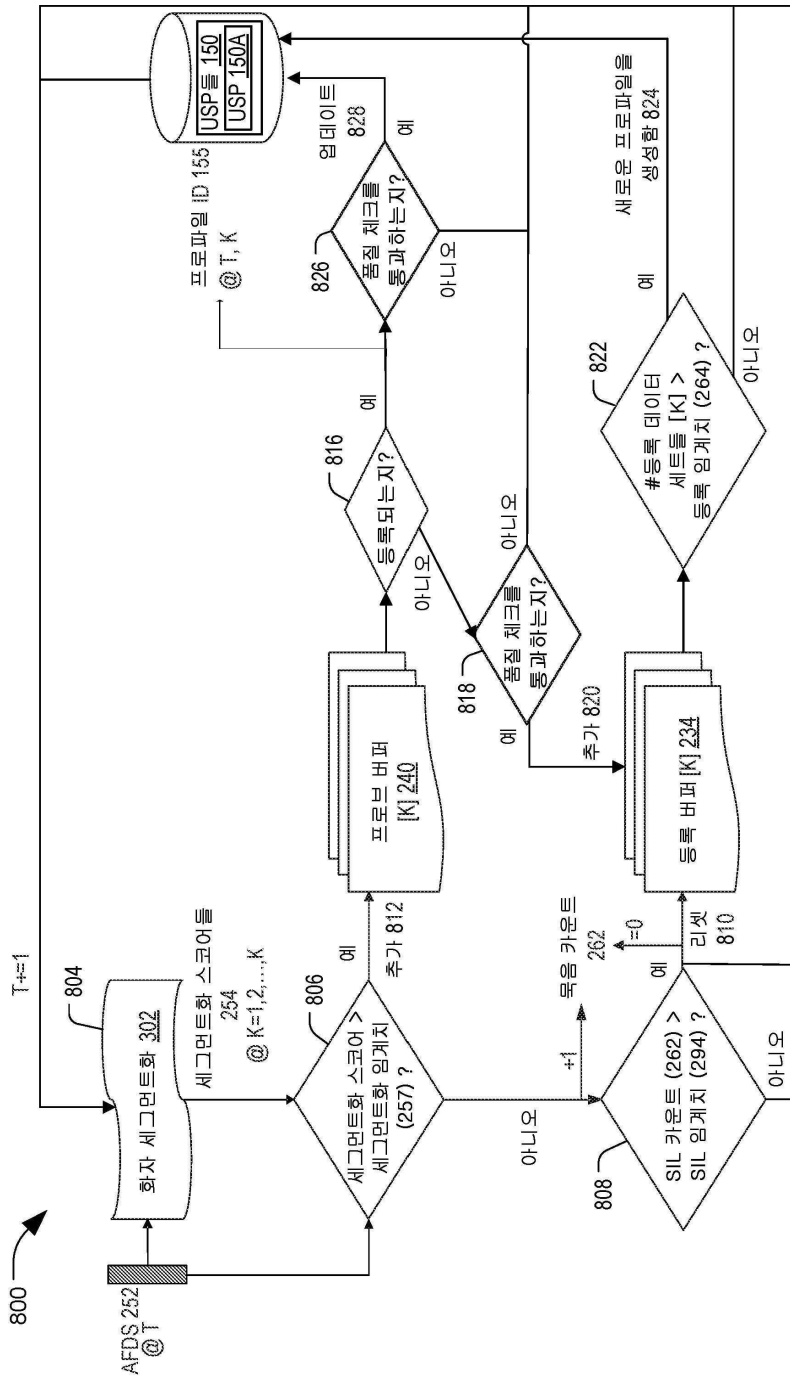
도면6



도면7

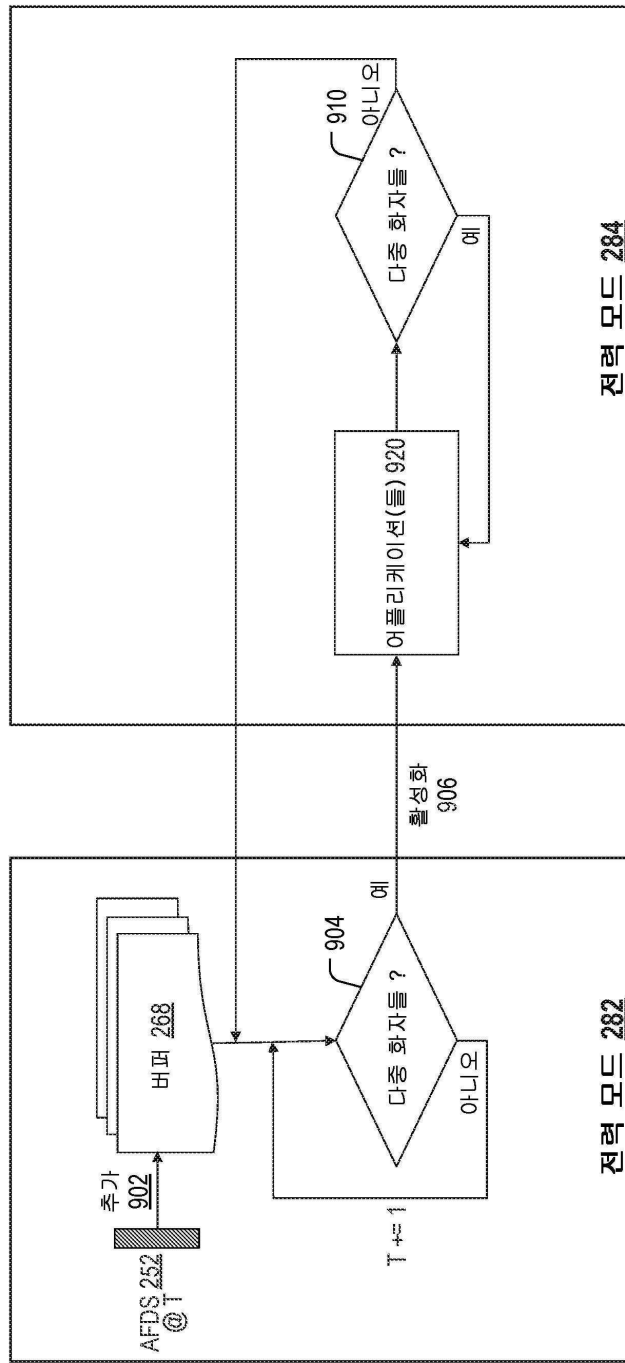


도면8



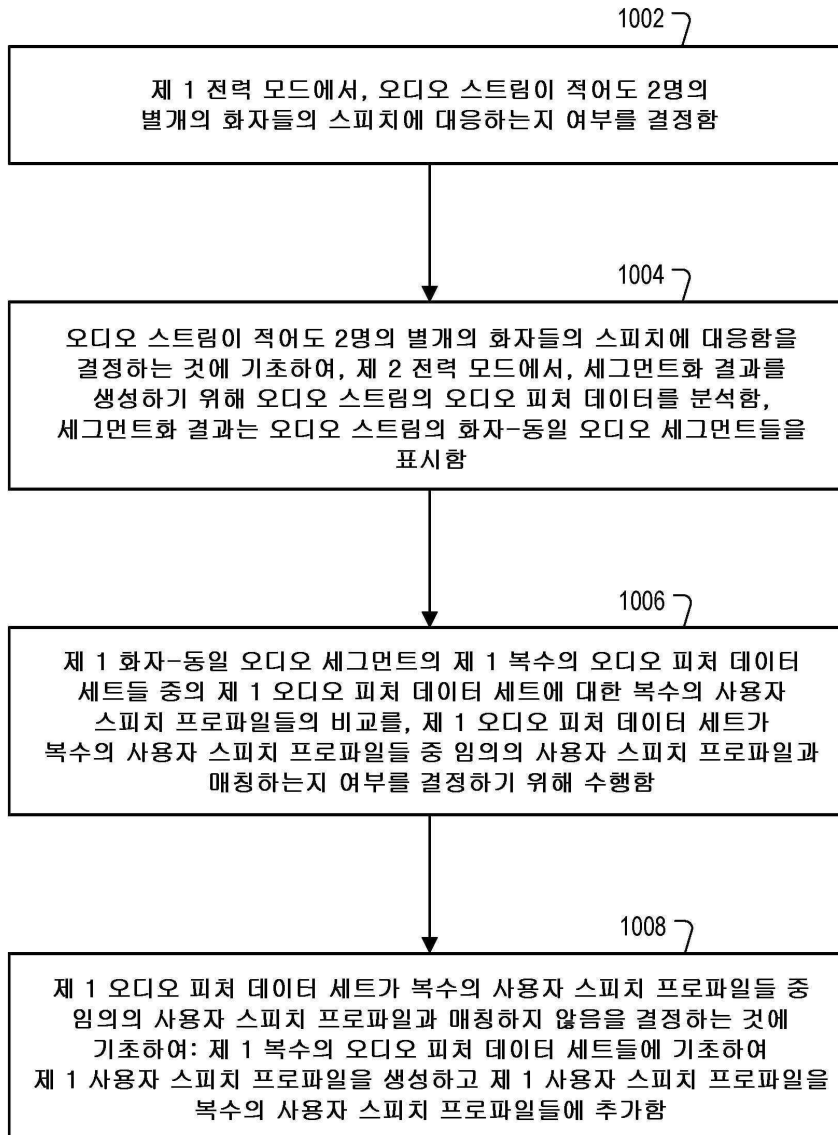
도면9

900

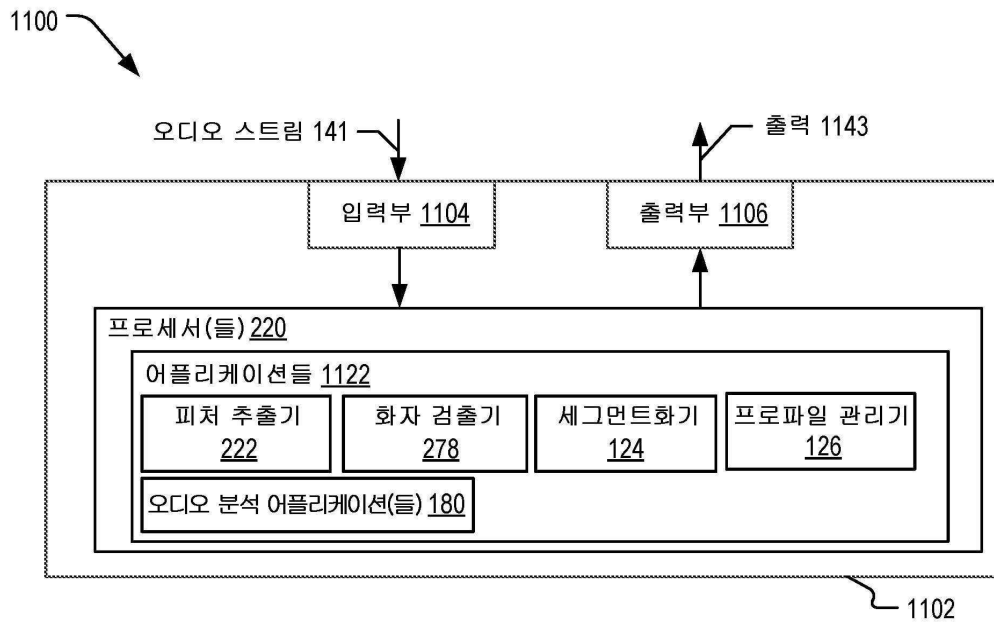


도면10

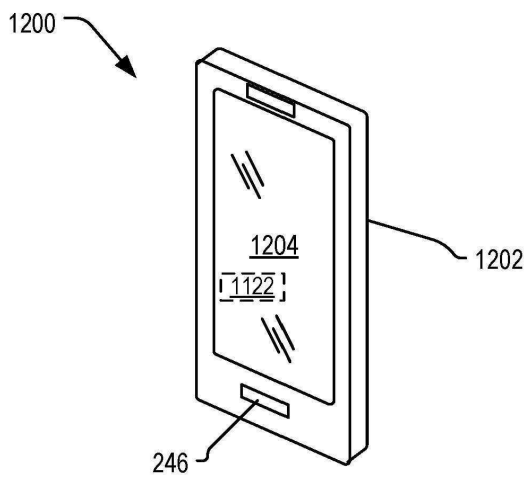
1000



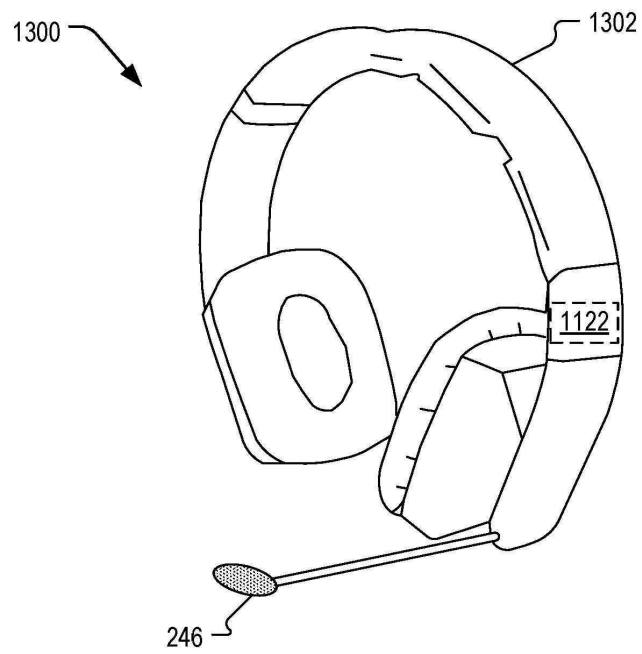
도면11



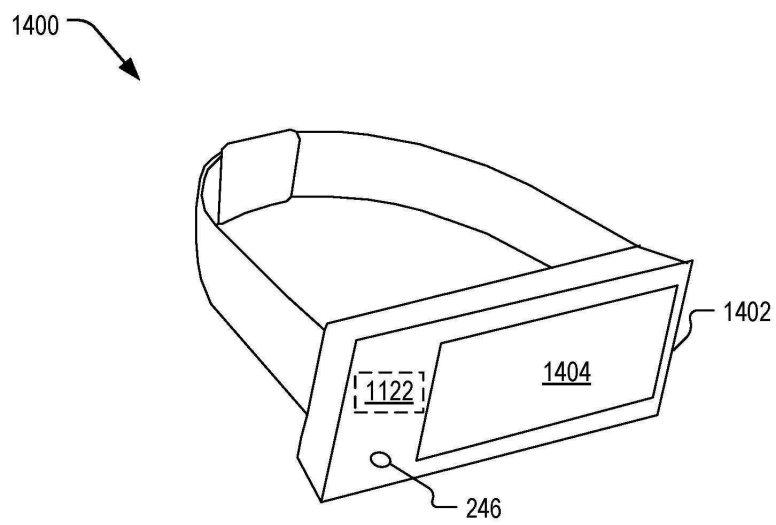
도면12



도면13

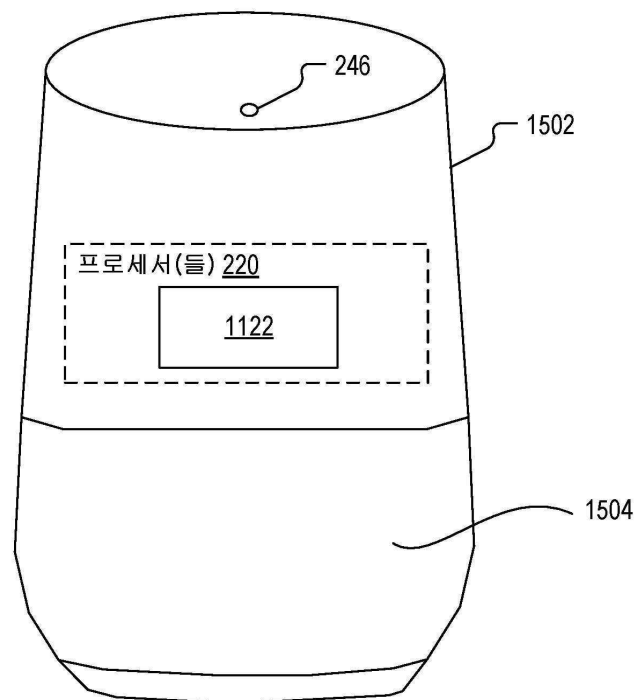


도면14



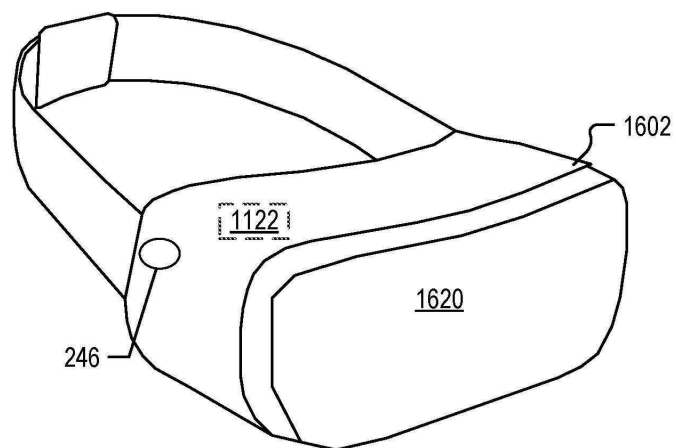
도면15

1500

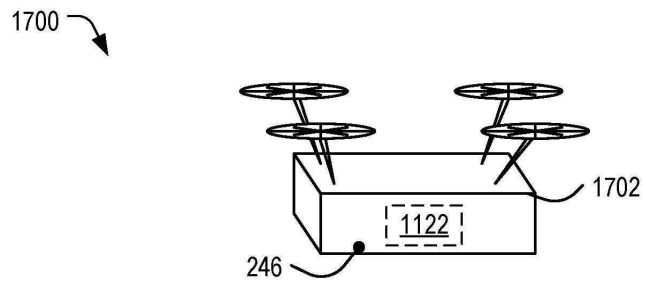


도면16

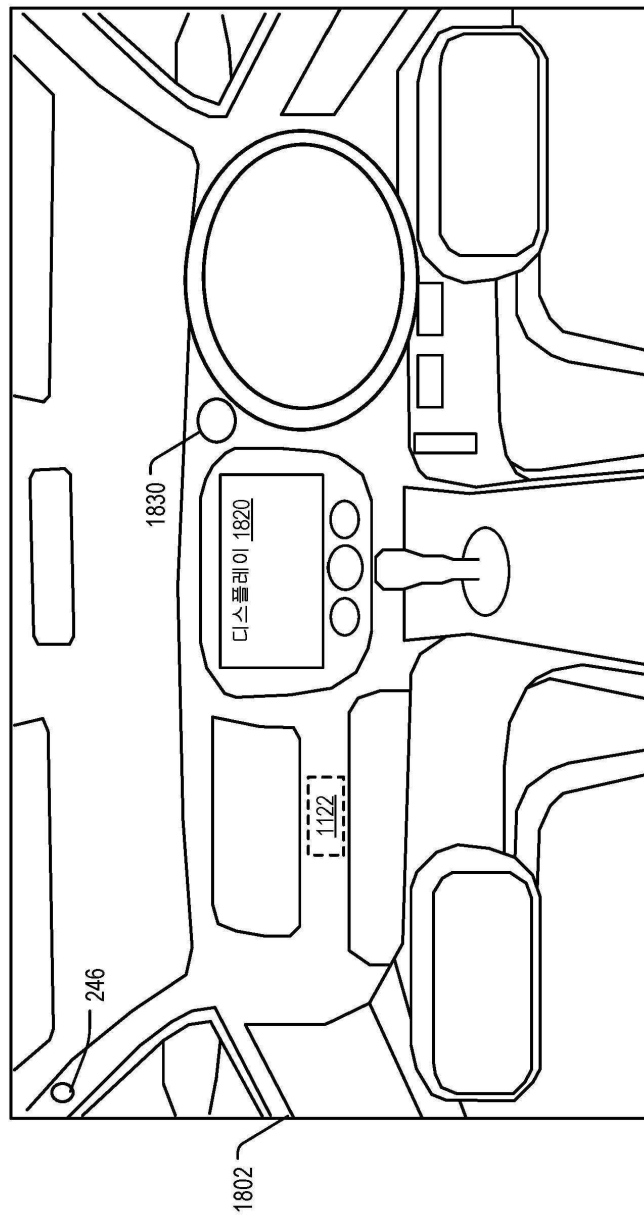
1600



도면17



도면18



도면 19

