

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2023-0133864
(43) 공개일자 2023년09월19일

(51) 국제특허분류(Int. Cl.)
H04M 3/56 (2006.01) G10L 13/027 (2013.01)
G10L 13/08 (2006.01) H04N 7/14 (2006.01)
H04N 7/15 (2016.01)

(52) CPC특허분류
H04M 3/568 (2013.01)
G10L 13/027 (2013.01)

(21) 출원번호 10-2023-7025451

(22) 출원일자(국제) 2021년12월09일
심사청구일자 없음

(85) 번역문제출일자 2023년07월25일

(86) 국제출원번호 PCT/US2021/072831

(87) 국제공개번호 WO 2022/169534
국제공개일자 2022년08월11일

(30) 우선권주장
17/166,250 2021년02월03일 미국(US)

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

(72) 발명자
올리비에리 페르디난도
미국 92121 캘리포니아주 샌디에고 모어하우스 드라이브 5775
웨스트버그 레이드
미국 92121 캘리포니아주 샌디에고 모어하우스 드라이브 5775
타가두르 쉬바와 산카르
미국 92121 캘리포니아주 샌디에고 모어하우스 드라이브 5775

(74) 대리인
특허법인코리아나

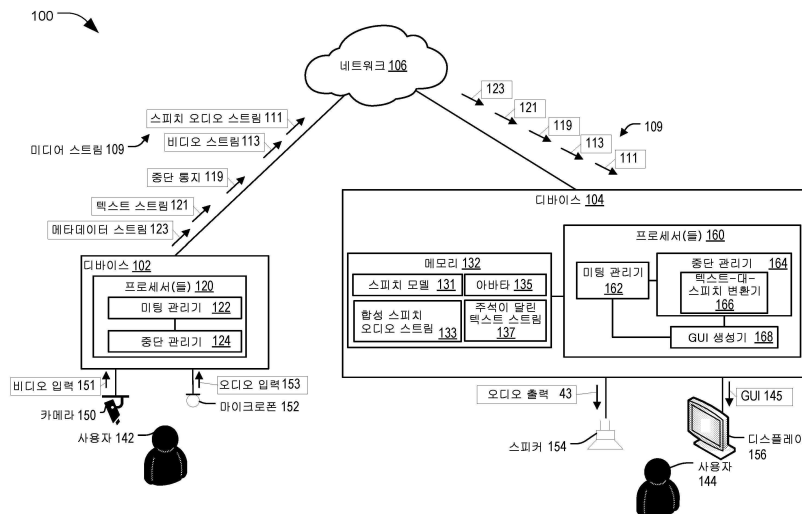
전체 청구항 수 : 총 30 항

(54) 발명의 명칭 스피치 오디오 스트림 중단들을 처리하는 시스템들 및 방법들

(57) 요약

통신을 위한 디바이스는, 온라인 미팅 동안, 제 1 사용자의 스피치를 나타내는 스피치 오디오 스트림을 수신하도록 구성된 하나 이상의 프로세서들을 포함한다. 하나 이상의 프로세서들은 또한 제 1 사용자의 스피치를 나타내는 텍스트 스트림을 수신하도록 구성된다. 하나 이상의 프로세서들은 스피치 오디오 스트림에서의 중단에 응답하여 텍스트 스트림에 기반하여 출력을 선택적으로 생성하도록 추가로 구성된다.

대표도



(52) CPC특허분류

G10L 13/08 (2013.01)

H04N 7/147 (2013.01)

H04N 7/157 (2013.01)

H04M 2201/39 (2013.01)

H04M 2201/40 (2013.01)

H04M 2201/60 (2013.01)

H04M 2203/2088 (2013.01)

명세서

청구범위

청구항 1

통신을 위한 디바이스로서,
하나 이상의 프로세서들을 포함하고,
상기 하나 이상의 프로세서들은,
온라인 미팅 동안, 제 1 사용자의 스피치를 나타내는 스피치 오디오 스트림을 수신하고,
상기 제 1 사용자의 스피치를 나타내는 텍스트 스트림을 수신하고,
상기 스피치 오디오 스트림에서의 중단에 응답하여 상기 텍스트 스트림에 기반하여 출력을 선택적으로 생성하도록
구성되는, 통신을 위한 디바이스.

청구항 2

제 1 항에 있어서,
상기 하나 이상의 프로세서들은 상기 스피치 오디오 스트림의 마지막 수신된 오디오 프레임의 임계 지속기간 내
에 상기 스피치 오디오 스트림의 오디오 프레임들이 수신되지 않는다는 결정에 응답하여 중단을 검출하도록 구
성되는, 통신을 위한 디바이스.

청구항 3

제 1 항에 있어서,
상기 하나 이상의 프로세서들은 상기 텍스트 스트림을 수신하는 것에 응답하여 중단을 검출하도록 구성되는, 통
신을 위한 디바이스.

청구항 4

제 1 항에 있어서,
상기 하나 이상의 프로세서들은 중단 통지를 수신하는 것에 응답하여 상기 중단을 검출하도록 구성되는, 통신을
위한 디바이스.

청구항 5

제 1 항에 있어서,
상기 하나 이상의 프로세서들은 상기 텍스트 스트림을 디스플레이에 대한 출력으로서 제공하도록 구성되는, 통
신을 위한 디바이스.

청구항 6

제 1 항에 있어서,
상기 하나 이상의 프로세서들은, 추가로,
상기 제 1 사용자의 스피치의 역양을 나타내는 메타데이터 스트림을 수신하고,
상기 메타데이터 스트림을 기반으로 상기 텍스트 스트림에 주석을 달도록
구성되는, 통신을 위한 디바이스.

청구항 7

제 1 항에 있어서,
 상기 하나 이상의 프로세서들은, 추가로,
 상기 텍스트 스트림에 대해 텍스트-대-스피치 변환을 수행하여 합성 스피치 오디오 스트림을 생성하고,
 상기 합성 스피치 오디오 스트림을 스피커에 대한 출력으로서 제공하도록
 구성되는, 통신을 위한 디바이스.

청구항 8

제 7 항에 있어서,
 상기 하나 이상의 프로세서들은 추가로 상기 제 1 사용자의 스피치의 억양을 나타내는 메타데이터 스트림을 수신하도록 구성되고, 상기 텍스트-대-스피치 변환은 메타데이터 스트림에 기반하는, 통신을 위한 디바이스.

청구항 9

제 7 항에 있어서,
 상기 하나 이상의 프로세서들은 추가로 상기 합성 스피치 오디오 스트림을 상기 스피커에 제공하는 것과 동시에 아바타를 디스플레이하도록 구성되는, 통신을 위한 디바이스.

청구항 10

제 9 항에 있어서,
 상기 하나 이상의 프로세서들은, 상기 온라인 미팅 동안, 미디어 스트림을 수신하도록 구성되고, 상기 미디어 스트림은 상기 제 1 사용자의 스피치 오디오 스트림 및 비디오 스트림을 포함하는, 통신을 위한 디바이스.

청구항 11

제 10 항에 있어서,
 상기 하나 이상의 프로세서들은, 상기 중단에 응답하여,
 상기 스피치 오디오 스트림의 재생 (playback) 을 중지하고,
 상기 비디오 스트림의 재생을 중지하도록
 구성되는, 통신을 위한 디바이스.

청구항 12

제 10 항에 있어서,
 상기 하나 이상의 프로세서들은, 상기 중단 종료에 응답하여,
 상기 합성 스피치 오디오 스트림을 상기 스피커에 제공하는 것을 억제하고,
 상기 아바타를 디스플레이하는 것을 억제하고,
 상기 비디오 스트림의 재생을 재개하고,
 상기 스피치 오디오 스트림의 재생을 재개하도록
 구성되는, 통신을 위한 디바이스.

청구항 13

제 7 항에 있어서,
 상기 텍스트-대-스피치 변환은 스피치 모델에 기반하여 수행되는, 통신을 위한 디바이스.

청구항 14

제 13 항에 있어서,

상기 스피치 모델은 일반 스피치 모델에 대응하는, 통신을 위한 디바이스.

청구항 15

제 13 항에 있어서,

상기 하나 이상의 프로세서들은, 상기 중단 이전에, 상기 스피치 오디오 스트림에 기반하여 상기 스피치 모델을 업데이트하도록 구성되는, 통신을 위한 디바이스.

청구항 16

제 1 항에 있어서,

상기 하나 이상의 프로세서들은,

상기 온라인 미팅 동안, 제 2 사용자의 스피치를 나타내는 제 2 스피치 오디오 스트림을 수신하고,

상기 출력을 생성하는 것과 동시에 스피커에 상기 제 2 스피치 오디오 스트림을 제공하도록

구성되도록, 통신을 위한 디바이스.

청구항 17

제 1 항에 있어서,

상기 하나 이상의 프로세서들은,

상기 스피치 오디오 스트림에서의 상기 중단에 응답하여 상기 스피치 오디오 스트림의 재생을 중지하고,

상기 중단 종료에 응답하여,

상기 텍스트 스트림에 기반하여 상기 출력을 생성하는 것을 억제하고,

상기 스피치 오디오 스트림의 재생을 재개하도록

구성되는, 통신을 위한 디바이스.

청구항 18

통신 방법으로서,

온라인 미팅 동안 디바이스에서, 제 1 사용자의 스피치를 나타내는 스피치 오디오 스트림을 수신하는 단계,

상기 디바이스에서, 상기 제 1 사용자의 스피치를 나타내는 텍스트 스트림을 수신하는 단계, 및

상기 디바이스에서, 상기 스피치 오디오 스트림에서의 중단에 응답하여 상기 텍스트 스트림에 기반하여 출력을 선택적으로 생성하는 단계

를 포함하는, 통신 방법.

청구항 19

제 18 항에 있어서,

상기 스피치 오디오 스트림의 마지막 수신된 오디오 프레임의 임계 지속기간 내에 상기 스피치 오디오 스트림의 오디오 프레임들이 수신되지 않음을 결정에 응답하여 상기 중단을 검출하는 단계를 더 포함하는, 통신 방법.

청구항 20

제 18 에 있어서,

상기 텍스트 스트림을 수신하는 것에 응답하여 상기 중단을 검출하는 단계를 더 포함하는, 통신 방법.

청구항 21

제 18 에 있어서,

중단 통지를 수신하는 것에 응답하여 상기 중단을 검출하는 단계를 더 포함하는, 통신 방법.

청구항 22

제 18 항에 있어서,

디스플레이에 대한 출력으로서 상기 텍스트 스트림을 제공하는 단계를 더 포함하는, 통신 방법.

청구항 23

제 18 항에 있어서,

상기 제 1 사용자의 스피치의 억양을 나타내는 메타데이터 스트림을 수신하는 단계, 및

상기 메타데이터 스트림을 기반으로 상기 텍스트 스트림에 주석을 다는 단계

를 더 포함하는, 통신 방법.

청구항 24

명령들을 저장하는 비밀시적 컴퓨터 판독가능 저장 매체로서,

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

온라인 미팅 동안, 제 1 사용자의 스피치를 나타내는 스피치 오디오 스트림을 수신하고,

상기 제 1 사용자의 스피치를 나타내는 텍스트 스트림을 수신하고,

상기 스피치 오디오 스트림에서의 중단에 응답하여 상기 텍스트 스트림에 기반하여 출력을 선택적으로 생성하게 하는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 25

제 24 항에 있어서,

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

상기 텍스트 스트림에 대해 텍스트-대-스피치 변환을 수행하여 합성 스피치 오디오 스트림을 생성하고,

상기 합성 스피치 오디오 스트림을 스피커에 대한 출력으로서 제공하게

하는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 26

제 25 항에 있어서,

상기 명령들은, 상기 하나 이상의 프로세서들에 의해 실행될 때, 상기 하나 이상의 프로세서들로 하여금, 상기 제 1 사용자의 스피치의 억양을 나타내는 메타데이터 스트림을 수신하게 하고, 상기 텍스트-대-스피치 변환은 상기 메타데이터 스트림에 기반하는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 27

제 25 항에 있어서,

상기 명령들은, 상기 하나 이상의 프로세서들에 의해 실행될 때, 상기 하나 이상의 프로세서들로 하여금, 상기 합성 스피치 오디오 스트림을 상기 스피커에 제공하는 것과 동시에 아바타를 디스플레이하게 하는, 비밀시적 컴퓨터 판독가능 저장 매체.

청구항 28

제 25 항에 있어서,

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

중단 이전에, 스피치 오디오 스트림에 기반하여 스피치 모델을 업데이트하게 하고, 상기 텍스트-대-스피치 변환은 스피치 모델에 기반하여 수행되는, 비일시적 컴퓨터 판독가능 저장 매체.

청구항 29

장치로서,

온라인 미팅 동안 스피치 오디오 스트림을 수신하기 위한 수단으로서, 상기 스피치 오디오 스트림은 제 1 사용자의 스피치를 나타내는, 상기 수신하기 위한 수단,

상기 제 1 사용자의 상기 스피치를 나타내는 텍스트 스트림을 수신하기 위한 수단, 및

상기 스피치 오디오 스트림에서의 중단에 응답하여 상기 텍스트 스트림에 기반하여 출력을 선택적으로 생성하기 위한 수단

을 포함하는, 장치.

청구항 30

제 29 항에 있어서,

상기 스피치 오디오 스트림을 수신하기 위한 수단, 상기 텍스트 스트림을 수신하기 위한 수단, 및 상기 출력을 선택적으로 생성하기 위한 수단은, 가상 어시스턴트, 가전 제품, 스마트 디바이스, 사물 인터넷 (IoT) 디바이스, 통신 디바이스, 헤드셋, 차량, 컴퓨터, 디스플레이 디바이스, 텔레비전, 게이밍 콘솔, 음악 플레이어, 라디오, 비디오 플레이어, 엔터테인먼트 유닛, 개인용 미디어 플레이어, 디지털 비디오 플레이어, 카메라, 또는 내비게이션 디바이스 중 적어도 하나에 통합되는, 장치.

발명의 설명

기술 분야

[0001] 우선권의 주장

[0002] 본 출원은 2021년 2월 3일자로 출원된 공동 소유의 미국 가특허출원 제 17/166,250 호로부터의 우선권의 이익을 주장하며, 그 내용들은 전부가 본원에 원용에 의해 명시적으로 통합된다.

[0003] 본 개시는 일반적으로 스피치 오디오 스트림 중단들을 처리하는 시스템들 및 방법들에 관한 것이다.

배경 기술

[0004] 기술의 진보는 더 소형이고 더 강력한 컴퓨팅 디바이스들을 발생시켰다. 예를 들어, 소형이고 경량이며 사용자에게 의해 용이하게 휴대되는 모바일 및 스마트 폰들과 같은 무선 전화기들, 태블릿들 및 랩톱 컴퓨터들을 포함하는 다양한 휴대용 개인 컴퓨팅 디바이스들이 현재 존재한다. 이들 디바이스들은 무선 네트워크들을 통해 보이스 및 데이터 패킷들을 통신할 수 있다. 추가로, 다수의 그러한 디바이스들은 디지털 스틸 카메라, 디지털 비디오 카메라, 디지털 레코더, 및 오디오 파일 플레이어와 같은 추가적인 기능성을 통합한다.

또한, 그러한 디바이스들은, 인터넷에 액세스하는데 사용될 수 있는 웹 브라우저 애플리케이션과 같은 소프트웨어 애플리케이션들을 포함한 실행가능 명령들을 프로세싱할 수 있다. 그에 따라, 이들 디바이스들은 상당한 컴퓨팅 능력들을 포함할 수 있다.

[0005] 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. 이러한 디바이스는 온라인 미팅 또는 통화에 사용되는 통신 디바이스들을 포함할 수 있다. 제 1 사용자와 제 2 사용자 사이의 온라인 미팅 동안 네트워크 이슈들은 제 1 사용자의 제 1 디바이스에 의해 전송된 일부 오디오 및 비디오 프레임들이 제 2 사용자의 제 2 디바이스에 의해 수신되지 않도록 프레임 손실을 야기할 수 있다. 네트워크 문제로 인해 프레임 손실되면 온라인 미팅 동안 복구할 수 없는 정보 손실로 이어질 수 있다. 예를 들어, 제 2 사용자는 무엇을 놓쳤는지를 추측하거나 제 1 사용자에게 놓친 것을 반복하도록 요청해야 하며, 이는 사용자 경험에 악영향을 미친다.

발명의 내용

해결하려는 과제

과제의 해결 수단

- [0006] 본 개시의 일 구현에 따르면, 통신을 위한 디바이스는 온라인 미팅 동안, 제 1 사용자의 스피치를 나타내는 스피치 오디오 스트림을 수신하도록 구성된 하나 이상의 프로세서들을 포함한다. 하나 이상의 프로세서들은 또한 제 1 사용자의 스피치를 나타내는 텍스트 스트림을 수신하도록 구성된다. 하나 이상의 프로세서들은 스피치 오디오 스트림에서의 중단에 응답하여 텍스트 스트림에 기반하여 출력을 선택적으로 생성하도록 추가로 구성된다.
- [0007] 본 개시의 다른 구현에 따르면, 통신 방법은 온라인 미팅 동안 디바이스에서, 제 1 사용자의 스피치를 나타내는 스피치 오디오 스트림을 수신하는 단계를 포함한다. 방법은, 또한 디바이스에서, 제 1 사용자의 스피치를 나타내는 텍스트 스트림을 수신하는 단계를 포함한다. 방법은, 디바이스에서, 스피치 오디오 스트림에서의 중단에 응답하여 텍스트 스트림에 기반하여 출력을 선택적으로 생성하는 단계를 더 포함한다.
- [0008] 본 개시의 다른 구현에 따르면, 비일시적 컴퓨터 판독가능 매체는 명령들을 포함하고, 상기 명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 하나 이상의 프로세서들로 하여금, 온라인 미팅 동안, 제 1 사용자의 스피치를 나타내는 스피치 오디오 스트림을 수신하게 한다. 명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 또한 하나 이상의 프로세서들로 하여금, 제 1 사용자의 스피치를 나타내는 텍스트 스트림을 수신하게 한다. 명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 추가로, 하나 이상의 프로세서들로 하여금, 스피치 오디오 스트림에서의 중단에 응답하여 텍스트 스트림에 기반하여 출력을 선택적으로 생성하게 한다.
- [0009] 본 개시의 다른 구현에 따르면, 장치는 온라인 미팅 동안, 제 1 사용자의 스피치를 나타내는 스피치 오디오 스트림을 수신하기 위한 수단을 포함한다. 장치는, 또한 제 1 사용자의 스피치를 나타내는 텍스트 스트림을 수신하기 위한 수단을 포함한다. 장치는, 스피치 오디오 스트림에서의 중단에 응답하여 텍스트 스트림에 기반하여 출력을 선택적으로 생성하기 위한 수단을 더 포함한다.
- [0010] 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

- [0011] 도 1 은 본 개시의 일부 예들에 따라 스피치 오디오 스트림 중단들을 처리하도록 동작가능한 시스템의 특정 예시적인 양태의 블록 다이어그램이다.
- 도 2 는 본 개시의 일부 예들에 따라 스피치 오디오 스트림 중단들을 처리하도록 동작가능한 시스템의 예시적인 양태의 다이어그램이다.
- 도 3a 는 본 개시의 일부 예들에 따라 도 1 의 시스템 또는 도 2 의 시스템에 의해 생성된 예시적인 그래픽 사용자 인터페이스 (GUI) 의 다이어그램이다.
- 도 3b 는 본 개시의 일부 예들에 따라 도 1 의 시스템 또는 도 2 의 시스템에 의해 생성된 예시적인 GUI 의 다이어그램이다.
- 도 3c 는 본 개시의 일부 예들에 따라 도 1 의 시스템 또는 도 2 의 시스템에 의해 생성된 예시적인 GUI 의 다이어그램이다.
- 도 4a 는 본 개시의 일부 예들에 따라 도 1 의 시스템 또는 도 2 의 시스템의 동작들의 예시적인 양태의 다이어그램이다.
- 도 4b 는 본 개시의 일부 예들에 따라 도 1 의 시스템 또는 도 2 의 시스템의 동작들의 예시적인 양태의 다이어그램이다.
- 도 5 는 본 개시의 일부 예들에 따라 스피치 오디오 스트림 중단들을 처리하도록 동작가능한 시스템의 예시적인 양태의 다이어그램이다.
- 도 6a 는 본 개시의 일부 예들에 따라 도 5 의 시스템에 의해 생성된 예시적인 그래픽 사용자 인터페이스 (GUI)

의 다이어그램이다.

도 6b 는 본 개시의 일부 예들에 따라 도 5 의 시스템에 의해 생성된 예시적인 GUI 의 다이어그램이다.

도 6c 는 본 개시의 일부 예들에 따라 도 5 의 시스템에 의해 생성된 예시적인 GUI 의 다이어그램이다.

도 7a 는 본 개시의 일부 예들에 따라 도 5 의 시스템의 동작들의 예시적인 양태의 다이어그램이다.

도 7b 는 본 개시의 일부 예들에 따라 도 5 의 시스템의 동작들의 예시적인 양태의 다이어그램이다.

도 8 은 본 개시의 일부 예들에 따른, 도 1, 도 2 또는 도 5 의 임의의 시스템들에 의해 수행될 수도 있는 스피치 오디오 스트림 중단들을 처리하는 방법의 특정 구현의 다이어그램이다.

도 9 는 본 개시의 일부 예들에 따른, 스피치 오디오 스트림 중단들을 처리하도록 동작가능한 집적 회로의 일 예를 도시한다.

도 10 은 본 개시의 일부 예들에 따른, 스피치 오디오 스트림 중단들을 처리하도록 동작가능한 모바일 디바이스의 다이어그램이다.

도 11 은 본 개시의 일부 예들에 따른, 스피치 오디오 스트림 중단들을 처리하도록 동작가능한 헤드셋의 다이어그램이다.

도 12 는 본 개시의 일부 예들에 따른, 스피치 오디오 스트림 중단들을 처리하도록 동작가능한 웨어러블 전자 디바이스의 다이어그램이다.

도 13 은 본 개시의 일부 예들에 따른, 스피치 오디오 스트림 중단들을 처리하도록 동작가능한 보이스-제어된 스피커 시스템의 다이어그램이다.

도 14 는 본 개시의 일부 예들에 따른, 스피치 오디오 스트림 중단들을 처리하도록 동작가능한 카메라의 다이어그램이다.

도 15 는 본 개시의 일부 예들에 따른, 스피치 오디오 스트림 간섭들을 처리하도록 동작가능한, 가상 현실 또는 증강 현실 헤드셋과 같은 헤드셋의 다이어그램이다.

도 16 은 본 개시의 일부 예들에 따라 스피치 오디오 스트림 중단들을 처리하도록 동작가능한 차량의 제 1 예의 다이어그램이다.

도 17 은 본 개시의 일부 예들에 따라 스피치 오디오 스트림 중단들을 처리하도록 동작가능한 차량의 제 2 예의 다이어그램이다.

도 18 은 본 개시의 일부 예들에 따라 스피치 오디오 스트림 중단들을 처리하도록 동작가능한 디바이스의 특정 예시적인 예의 블록 다이어그램이다.

발명을 실시하기 위한 구체적인 내용

[0012] 온라인 미팅 또는 통화의 일부를 놓치면 사용자 경험에 부정적인 영향을 줄 수 있다. 예를 들어, 제 1 사용자와 제 2 사용자 사이의 온라인 미팅 동안, 제 1 사용자의 제 1 디바이스에 의해 전송된 일부 오디오 프레임들이 제 2 사용자의 제 2 디바이스에 의해 수신되지 않으면, 제 2 사용자는 제 1 사용자의 스피치의 부분을 놓칠 수 있다. 제 2 사용자는 제 1 사용자가 말한 것을 추측하거나 제 1 사용자에게 놓친 것을 반복하도록 요청해야 한다. 이로 인해 의사소통이 잘못되고 대화의 흐름이 방해되며 시간이 낭비될 수 있다.

[0013] 스피치 오디오 스트림 중단들을 처리하는 시스템들 및 방법들이 개시된다. 예를 들어, 각각의 디바이스는 디바이스와 하나 이상의 다른 디바이스들 사이에 온라인 미팅 또는 통화를 설정하도록 구성된 미팅 관리기를 포함한다. (디바이스 또는 서버에서) 중단 관리기는 스피치 오디오 스트림 중단들을 처리하도록 구성된다.

[0014] 제 1 사용자의 제 1 디바이스와 제 2 사용자의 제 2 디바이스 사이의 온라인 미팅 동안, 제 1 디바이스의 미팅 관리기는 미디어 스트림을 제 2 디바이스에 전송한다. 미디어 스트림은 스피치 오디오 스트림, 비디오 스트림, 또는 둘 다를 포함한다. 스피치 오디오 스트림은 미팅 동안 제 1 사용자의 스피치에 대응한다.

[0015] 스트림 관리기 (제 1 디바이스 또는 서버에서) 는 스피치 오디오 스트림에 대해 스피치-대-텍스트 변환을 수행함으로써 텍스트 스트림을 생성하고 이 텍스트 스트림을 제 2 디바이스에 전달한다. 스트림 관리기 (예를 들어, 제 1 디바이스 또는 서버에서 미팅 관리기) 는, 제 1 동작 모드 (예를 들어, 전송 자막 데이터 모드) 에

서, 온라인 미팅 전체에 걸쳐 미디어 스트림과 동시에 텍스트 스트림을 전달한다. 대안적인 예에서, 스트림 관리기 (예를 들어, 제 1 디바이스 또는 서버에서 중단 관리기) 는, 제 2 동작 모드 (예를 들어, 전송 중단 데이터 모드) 에서, 미디어 스트림을 제 2 디바이스에 전송하여 네트워크 이슈들 (예를 들어, 낮은 대역폭, 패킷 손실 등) 을 검출하는 것에 응답하여 텍스트 스트림을 제 2 디바이스에 전달한다.

[0016] 일부 예들에서, 네트워크 이슈들은 텍스트 스트림의 수신에서의 중단 없이 제 2 디바이스에서 미디어 스트림의 수신에서의 중단을 야기한다. 일부 예들에서, 제 2 디바이스는, 제 1 동작 모드 (예를 들어, 디스플레이 자막 데이터 모드) 에서, 네트워크 이슈들을 검출하는 것과 독립적으로 디스플레이에 텍스트 스트림을 제공한다.

다른 예들에서, 제 2 디바이스는, 제 2 동작 모드 (예를 들어, 디스플레이 중단 데이터 모드) 에서, 미디어 스트림에서의 중단을 검출하는 것에 응답하여 텍스트 스트림을 디스플레이한다.

[0017] 특정 예에서, 스트림 관리기 (예를 들어, 미팅 관리기 또는 중단 관리기) 는 텍스트 데이터에 추가하여 메타데이터 스트림을 전달한다. 메타데이터는 제 1 사용자의 스피치의 감정, 억양, 다른 속성들을 나타낸다. 특정 예에서, 제 2 디바이스는 텍스트 스트림에 추가하여 메타데이터 스트림을 디스플레이한다. 예를 들어, 텍스트 스트림은 메타데이터 스트림에 기반하여 주석이 달린다.

[0018] 특정 예에서, 제 2 디바이스는 텍스트 스트림에 대해 텍스트-대-스피치 변환을 수행하여 합성 스피치 오디오 스트림을 생성하고, (예를 들어, 중단된 스피치 오디오 스트림을 대체하기 위해) 합성 스피치 오디오 스트림을 출력한다. 특정 예에서, 텍스트-대-스피치 변환은 메타데이터 스트림에 적어도 부분적으로 기반한다.

[0019] 특정 예에서, 제 2 디바이스는 합성 스피치 오디오 스트림의 출력 동안 (예를 들어, 중단된 비디오 스트림을 대체하기 위해) 아바타를 디스플레이한다. 특정 예에서, 텍스트-대-스피치 변환은 일반 스피치 모델에 기반한다. 예를 들어, 제 1 일반 스피치 모델은 하나의 사용자에게 대해 사용될 수 있고, 제 2 일반 스피치 모델은 청취자들이 상이한 사용자들에 대응하는 스피치 사이를 구별할 수 있도록 다른 사용자에게 대해 사용될 수 있다.

다른 특정 예에서, 텍스트-대-스피치 변환은 제 1 사용자의 스피치에 기반하여 생성된 사용자 스피치 모델에 기반한다. 특정 예에서, 사용자 스피치 모델은 온라인 미팅에 앞서 생성된다. 특정 예에서, 사용자 스피치 모델은 온라인 미팅 동안 생성된다 (또는 업데이트된다). 특정 예에서, 사용자 스피치 모델은 일반 스피치 모델로부터 초기화되고 제 1 사용자의 스피치에 기반하여 업데이트된다.

[0020] 특정 예에서, 아바타는 스피치 모델이 트레이닝되고 있음을 나타낸다. 예를 들어, 아바타는 일반 스피치 모델이 사용되고 있음 (또는 사용자 스피치 모델이 준비되지 않음) 을 나타내기 위해 적색으로 초기화되고, 아바타는 스피치 모델이 트레이닝되고 있음을 나타내기 위해 시간에 따라 적색에서 녹색으로 전환된다. 녹색 아바타는 사용자 스피치 모델이 트레이닝되었음을 (또는 사용자 스피치 모델이 준비되었음을) 나타낸다.

[0021] 온라인 미팅은 2 명 초과 사용자들 사이에 있을 수 있다. 제 1 디바이스가 네트워크 문제들을 겪고 있지만 온라인 미팅에서의 제 3 사용자의 제 3 디바이스가 네트워크 문제들을 겪고 있지 않은 상황에서, 제 2 디바이스는 제 3 사용자의 스피치, 비디오, 또는 둘 모두에 대응하는 제 3 디바이스로부터 수신된 제 2 미디어 스트림을 출력하는 것과 동시에 제 1 사용자에게 대해 합성 스피치 오디오 스트림을 출력할 수 있다.

[0022] 본 개시의 특정 양태들은 도면들을 참조하여 이하에 설명된다. 설명에서, 공통 특징들은 공통 도면부호들로 지정된다. 본 명세서에서 사용된 바와 같이, 다양한 용어가 단지 특정 구현들을 설명하는 목적으로 사용되며 구현들을 한정하는 것으로 의도되지 않는다. 예를 들어, 단수 형태들 "a", "an", 및 "the" 는, 문맥이 분명히 달리 표시하지 않는 한, 복수 형태들을 물론 포함하도록 의도된다. 추가로, 본 명세서에서 설명된 일부 피쳐들은 일부 구현들에서 단수이고 다른 구현들에서는 복수이다. 설명하기 위해, 도 1 은 하나 이상의 프로세서들 (도 1 의 "프로세서(들)" (160)) 을 포함하는 디바이스 (104) 를 도시하며, 이는 일부 구현들에서는 디바이스 (104) 가 단일 프로세서 (160) 를 포함하고 다른 구현들에서는 디바이스 (104) 가 다중의 프로세서들 (160) 을 포함함을 나타낸다.

[0023] 본 명세서에서 사용된 바와 같이, 용어들 "포함한다 (comprise)", "포함한다 (comprises)", 및 "포함하는 (comprising)" 은 "포함한다 (include)", "포함한다 (includes)", 또는 "포함하는 (including)" 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, 제한하는 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 (예를 들어, "제 1", "제 2", "제 3" 등) 는 홀로 다른 엘리먼트에 관하여 엘리먼트의 임의의 우선순위 또는 순

서를 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.

[0024] 본 명세서에서 사용된 바와 같이, "결합된" 은 "통신가능하게 결합된", "전기적으로 결합된", 또는 "물리적으로 결합된" 을 포함할 수도 있으며, 또한 (또는 대안적으로) 이들의 임의의 조합들을 포함할 수도 있다. 2 개의 디바이스들 (또는 컴포넌트들) 은 하나 이상의 다른 디바이스, 컴포넌트, 와이어, 버스, 네트워크 (예를 들어, 유선 네트워크, 무선 네트워크, 또는 이들의 조합) 등을 통해 직접적으로 또는 간접적으로 결합될 (예를 들어, 통신적으로 결합될, 전기적으로 결합될, 또는 물리적으로 결합될) 수도 있다. 전기적으로 결합된 2 개의 디바이스들 (또는 컴포넌트들) 은 동일한 디바이스 또는 상이한 디바이스에 포함될 수도 있고, 예시적인, 비-제한적인 예들로서 전자기기들, 하나 이상의 커넥터들 또는 유도 결합을 통해 접속될 수도 있다. 일부 구현들에서, 전기 통신에서와 같이, 통신가능하게 커플링되는 2 개의 디바이스들 (또는 컴포넌트들) 은 하나 이상의 와이어, 버스, 네트워크 등을 통해, 직접 또는 간접적으로 신호들 (예를 들어, 디지털 신호들 또는 아날로그 신호들) 을 전송 및 수신할 수도 있다. 본 명세서에서 사용된 바와 같이, "직접 커플링된" 은 컴포넌트들을 개재하지 않으면서 커플링 (예를 들어, 통신가능하게 커플링, 전기적으로 커플링, 또는 물리적으로 커플링) 되는 2 개의 디바이스들을 포함할 수도 있다.

[0025] 본 개시에서, 용어들 이를 테면 "결정하는 것", "계산하는 것", "추정하는 것", "시프트하는 것", "조정하는 것", 등은 하나 이상의 동작들이 수행되는 방법을 설명하기 위해 사용될 수도 있다. 이러한 용어들이 한정적인 것으로 해석되어서는 안되며 다른 기법들이 유사한 동작들을 수행하는데 이용될 수도 있다는 점에 유의해야 한다. 추가적으로, 본 명세서에서 언급된 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 능동으로 생성하는 것, 추정하는 것, 계산하는 것, 또는 결정하는 것을 지칭할 수도 있거나 또는 다른 컴포넌트 또는 디바이스에 의해서와 같이, 이미 생성된 파라미터 (또는 신호) 를 사용하는 것, 선택하는 것, 또는 액세스하는 것을 지칭할 수도 있다.

[0026] 도 1 을 참조하면, 스피치 오디오 스트림 간섭들을 처리하도록 구성된 시스템의 특정 예시적인 양태가 개시되고 일반적으로 100 으로 지정된다. 시스템 (100) 은 네트워크 (106) 를 통해 디바이스 (104) 에 커플링된 디바이스 (102) 를 포함한다. 네트워크 (106) 는 유선 네트워크, 무선 네트워크, 또는 그 양자 모두를 포함한다. 디바이스 (102) 는 카메라 (150), 마이크로폰 (152), 또는 양자 모두에 커플링된다. 디바이스 (104) 는 스피커 (154), 디스플레이 디바이스 (156), 또는 양자 모두에 커플링된다.

[0027] 디바이스 (104) 는 메모리 (132) 에 커플링된 하나 이상의 프로세서들 (160) 을 포함한다. 하나 이상의 프로세서들 (160) 은 중단 관리기 (164) 에 커플링된 미팅 관리기 (162) 를 포함한다. 미팅 관리기 (162) 및 중단 관리기 (164) 는 그래픽 사용자 인터페이스 (GUI) 생성기 (168) 에 커플링된다. 중단 관리기 (164) 는 텍스트-대-스피치 변환기 (166) 를 포함한다. 디바이스 (102) 는 중단 관리기 (124) 에 커플링된 미팅 관리기 (122) 를 포함하는 하나 이상의 프로세서들 (120) 을 포함한다. 미팅 관리기 (122) 및 미팅 관리기 (162) 는 온라인 미팅 (예를 들어, 오디오 통화, 비디오 통화, 회의 통화 등) 을 확립하도록 구성된다. 특정 예에서, 미팅 관리기 (122) 및 미팅 관리기 (162) 는 통신 애플리케이션 (예를 들어, 온라인 미팅 애플리케이션) 의 클라이언트들에 대응한다. 중단 관리기 (124) 및 중단 관리기 (164) 는 스피치 오디오 중단들을 처리하도록 구성된다.

[0028] 일부 구현들에서, 미팅 관리기 (122) 및 미팅 관리기 (162) 는 중단 관리기 (124) 및 중단 관리기 (164) 에 의해 관리되는 임의의 스피치 오디오 중단을 블라인드 (예를 들어, 인식하지 못함) 한다. 일부 구현들에서, 미팅 관리기 (122) 및 미팅 관리기 (162) 는 각각 디바이스 (102) 및 디바이스 (104) 의 네트워크 프로토콜 스택 (예를 들어, OSI (open systems interconnection) 모델) 의 상위 계층 (예를 들어, 애플리케이션 계층) 에 대응한다. 일부 구현들에서, 중단 관리기 (124) 및 중단 관리기 (164) 는 각각 디바이스 (102) 및 디바이스 (104) 의 네트워크 프로토콜 스택의 하위 레벨 (예를 들어, 전송 계층) 에 대응한다.

[0029] 일부 구현들에서, 디바이스 (102), 디바이스 (104), 또는 둘 다는 다양한 타입들의 디바이스들에 대응하거나 이에 포함된다. 예시적인 예에서, 하나 이상의 프로세서들 (120), 하나 이상의 프로세서들 (160), 또는 이들의 조합은, 도 11 를 참조하여 추가로 설명되는 바와 같이 헤드셋 디바이스에 통합된다. 다른 예들에서, 하나 이상의 프로세서들 (120), 하나 이상의 프로세서들 (160), 또는 이들의 조합은 도 10 을 참조하여 설명된 바

와 같은 모바일 폰 또는 태블릿 컴퓨터 디바이스, 도 12 를 참조하여 설명된 바와 같은 웨어러블 전자 디바이스, 도 13 을 참조하여 설명된 바와 같은 음성 제어된 스피커 시스템, 도 14 을 참조하여 설명된 바와 같은 카메라 디바이스, 또는 도 15 을 참조하여 설명된 바와 같은 가상 현실 헤드셋, 증강 현실 헤드셋, 또는 혼합 현실 헤드셋 중 적어도 하나에 통합된다. 다른 예시적인 예에서, 하나 이상의 프로세서들 (120), 하나 이상의 프로세서들 (160), 또는 이들의 조합은, 도 16 및 도 17 을 참조하여 추가로 설명되는 바와 같이 차량에 통합된다.

[0030] 동작 동안, 미팅 관리기 (122) 및 미팅 관리기 (162) 는 디바이스 (102) 와 디바이스 (104) 사이의 온라인 미팅 (예를 들어, 오디오 통화, 비디오 통화, 회의 통화, 또는 이들의 조합) 을 확립한다. 예를 들어, 온라인 미팅은 디바이스 (102) 의 사용자 (142) 와 디바이스 (104) 의 사용자 (144) 사이의 것이다. 마이크로폰 (152) 은 사용자 (142) 가 말하는 동안 사용자 (142) 의 스피치를 캡처하고 스피치를 나타내는 오디오 입력 (153) 을 디바이스 (102) 에 제공한다. 특정 양태에서, 카메라 (150) (예를 들어, 스틸 카메라, 비디오 카메라, 또는 양자 모두) 는 사용자 (142) 의 하나 이상의 이미지들 (예를 들어, 스틸 이미지들 또는 비디오) 을 캡처하고 하나 이상의 이미지들을 나타내는 비디오 입력 (151) 을 디바이스 (102) 에 제공한다. 특정 양태에서, 카메라 (150) 는 비디오 입력 (151) 을 디바이스 (102) 에 제공하는 동시에 마이크로폰 (152) 은 오디오 입력 (153) 을 디바이스 (102) 에 제공한다.

[0031] 미팅 관리기 (122) 는 오디오 입력 (153), 비디오 입력 (151), 또는 양자 모두에 기반하여 미디어 프레임들의 미디어 스트림 (109) 을 생성한다. 예를 들어, 미디어 스트림 (109) 은 스피치 오디오 스트림 (111), 비디오 스트림 (113), 또는 둘 다를 포함한다. 특정 양태에서, 미팅 관리기 (122) 는 미디어 스트림 (109) 을 네트워크 (106) 를 통해 디바이스 (104) 에 실시간으로 전송한다. 예를 들어, 미팅 관리기 (122) 는 비디오 입력 (151), 오디오 입력 (153), 또는 양자 모두가 수신될 때 미디어 스트림 (109) 의 미디어 프레임들을 생성하고, 미디어 프레임들이 생성될 때 미디어 프레임들의 미디어 스트림 (109) 을 전송 (예를 들어, 전송을 개시) 한다.

[0032] 특정 구현에서, 미팅 관리기 (122) 는, 디바이스 (102) 의 제 1 동작 모드 (예를 들어, 전송 자막 데이터 모드) 동안, 오디오 입력 (153) 에 기반하여 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 양자 모두를 생성한다. 예를 들어, 미팅 관리기 (122) 는 텍스트 스트림 (121) 을 생성하기 위해 오디오 입력 (153) 에 대해 스피치-대-텍스트 변환을 수행한다. 텍스트 스트림 (121) 은 오디오 입력 (153) 에서 검출된 스피치에 대응하는 텍스트를 나타낸다. 특정 양태에서, 미팅 관리기 (122) 는 메타데이터 스트림 (123) 을 생성하기 위해 오디오 입력 (153) 에 대해 스피치 억양 분석을 수행한다. 예를 들어, 메타데이터 스트림 (123) 은 오디오 입력 (153) 에서 검출된 스피치의 억양 (예를 들어, 감정, 음조, 톤 또는 이들의 조합) 을 나타낸다. 디바이스 (102) 의 제 1 동작 모드 (예를 들어, 전송 자막 데이터 모드) 에서, 미팅 관리기 (122) 는 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 (예를 들어, 패쇄 자막 데이터로서) 미디어 스트림 (109) 과 함께 (예를 들어, 네트워크 이슈들 또는 스피치 오디오 중단들과 독립적으로) 디바이스 (104) 에 전송한다. 대안적으로, 미팅 관리기 (122) 는, 디바이스 (102) 의 제 2 동작 모드 (예를 들어, 전송 중단 데이터 모드) 동안, 어떠한 스피치 오디오 중단도 검출되지 않는다고 결정하는 것에 응답하여 텍스트 스트림 (121) 및 메타데이터 스트림 (123) 을 생성하는 것을 억제한다.

[0033] 디바이스 (104) 는 디바이스 (102) 로부터 네트워크 (106) 를 통해 미디어 프레임들의 미디어 스트림 (109) 을 수신한다. 특정 구현에서, 디바이스 (104) 는 미디어 스트림 (109) 의 미디어 프레임들의 세트들 (예를 들어, 버스트들) 을 수신한다. 대안적인 구현에서, 디바이스 (104) 는 미디어 스트림 (109) 의 하나의 미디어 프레임을 한 번에 수신한다. 미팅 관리기 (162) 는 미디어 스트림 (109) 의 미디어 프레임들을 재생한다. 예를 들어, 미팅 관리기 (162) 는 스피치 오디오 스트림 (111) 에 기반하여 오디오 출력 (143) 을 생성하고 스피커 (154) 를 통해 오디오 출력 (143) 을 (예를 들어, 스트리밍 오디오 콘텐츠로서) 재생한다. 특정 양태에서, GUI 생성기 (168) 는 도 3a 를 참조하여 추가로 설명되는 바와 같이, 미디어 스트림 (109) 에 기반하여 GUI (145) 를 생성한다. 예를 들어, GUI 생성기 (168) 는 비디오 스트림 (113) 의 비디오 콘텐츠를 디스플레이하기 위해 GUI (145) 를 생성 (또는 업데이트) 하고 GUI (145) 를 디스플레이 디바이스 (156) 에 제공 (예를 들어, 비디오 콘텐츠를 스트리밍) 한다. 사용자 (144) 는 스피커 (154) 를 통해 사용자 (142) 의 오디오 스피치를 청취하면서 디스플레이 디바이스 (156) 상에서 사용자 (142) 의 이미지들을 볼 수 있다.

[0034] 특정 구현에서, 미팅 관리기 (162) 는 미디어 스트림 (109) 의 미디어 프레임들을 재생 이전에 버퍼에 저장한다. 예를 들어, 미팅 관리기 (162) 는 후속 미디어 프레임이 버퍼에서의 대응하는 재생 시간 (예를 들어, 제 2 재생 시간) 에 이용가능할 가능성을 증가시키기 위해 미디어 프레임을 수신하는 것과 제 1 재생 시

간에서의 미디어 프레임의 재생 사이의 지연을 추가한다. 특정 양태에서, 미팅 관리기 (162) 는 실시간으로 미디어 스트림 (109) 을 재생한다. 예를 들어, 미팅 관리기 (162) 는 미디어 스트림 (109) 의 후속 미디어 프레임들이 디바이스 (104) 에 의해 수신되고 있는 (또는 수신될 것으로 예상되는) 동안 오디오 출력 (143), GUI (145) 의 비디오 콘텐츠, 또는 둘 모두를 재생하기 위해 버퍼로부터 미디어 스트림 (109) 의 미디어 프레임들을 검색한다.

[0035] 미팅 관리기 (162) 는, 디바이스 (104) 의 제 1 동작 모드 (예를 들어, 디스플레이 자막 데이터 모드) 에서, (예를 들어, 스피치 오디오 스트림 (111) 에서의 중단を検출하는 것과 독립적으로) 미디어 스트림 (109) 과 함께 텍스트 스트림 (121) 을 재생한다. 특정 양태에서, 미팅 관리기 (162) 는, 예를 들어 디바이스 (102) 의 제 1 동작 모드 (예를 들어, 전송 자막 데이터 모드) 동안 미디어 스트림 (109) 과 함께 텍스트 스트림 (121), 메타데이터 스트림 (123) 또는 둘 다를 수신한다. 대안적인 양태에서, 미팅 관리기 (162) 는, 예를 들어 디바이스 (102) 의 제 2 동작 모드 (예컨대, 송신 중단 데이터 모드) 동안, 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 모두를 수신하지 않고, 스피치 오디오 스트림 (111), 비디오 스트림 (113), 또는 둘 모두에 기반하여 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 모두를 생성한다. 예를 들어, 미팅 관리기 (162) 는 스피치 오디오 스트림 (111) 에 대해 스피치-대-텍스트 변환을 수행하여 텍스트 스트림 (121) 을 생성하고, 스피치 오디오 스트림 (111) 에 대해 억양 분석을 수행하여 메타데이터 스트림 (123) 을 생성한다.

[0036] 디바이스 (104) 의 제 1 동작 모드 (예를 들어, 디스플레이 자막 데이터 모드) 동안, 미팅 관리기 (162) 는 텍스트 스트림 (121) 을 디스플레이 디바이스 (156) 에 출력으로서 제공한다. 예를 들어, 미팅 관리기 (162) 는 비디오 스트림 (113) 의 비디오 콘텐츠를 디스플레이하는 것, 오디오 출력 (143) 을 스피커 (154) 에 제공하는 것, 또는 둘 다와 동시에 GUI (145) 를 사용하여 텍스트 스트림 (121) 의 텍스트 콘텐츠를 (예를 들어, 패쇄 자막들로서) 디스플레이한다. 설명을 위해, 미팅 관리기 (162) 는 비디오 스트림 (113) 을 GUI 생성기 (168) 에 제공하는 것과 동시에 텍스트 스트림 (121) 을 GUI 생성기 (168) 에 제공한다. GUI 생성기 (168) 는 텍스트 스트림 (121), 비디오 스트림 (113), 또는 둘 다를 디스플레이하도록 GUI (145) 를 업데이트한다. GUI 생성기 (168) 는 미팅 관리기 (162) 가 스피치 오디오 스트림 (111) 을 오디오 출력 (143) 으로서 스피커 (154) 에 제공하는 것과 동시에 GUI (145) 의 업데이트들을 디스플레이 디바이스 (156) 에 제공한다.

[0037] 특정 예에서, 미팅 관리기 (162) 는 텍스트 스트림 (121) 및 메타데이터 스트림 (123) 에 기반하여 주석이 달린 텍스트 스트림 (137) 을 생성한다. 특정 양태에서, 미팅 관리기 (162) 는 메타데이터 스트림 (123) 에 기반하여 텍스트 스트림 (121) 에 주석을 추가함으로써 주석이 달린 텍스트 스트림 (137) 을 생성한다. 미팅 관리기 (162) 는 주석이 달린 텍스트 스트림 (137) 을 디스플레이 디바이스 (156) 에 출력으로서 제공한다. 예를 들어, 미팅 관리기 (162) 는 미디어 스트림 (109) 으로 주석이 달린 텍스트 스트림 (137) 을 재생한다. 설명을 위해, 미팅 관리기 (162) 는 비디오 스트림 (113) 의 비디오 콘텐츠를 디스플레이하는 것, 오디오 출력 (143) 을 스피커 (154) 에 제공하는 것, 또는 둘 다와 동시에 GUI (145) 를 사용하여 주석이 달린 텍스트 스트림 (137) 의 주석이 달린 텍스트 콘텐츠를 (예를 들어, 억양 표시를 가진 패쇄 자막들로서) 디스플레이한다.

[0038] 특정 구현에서, 미팅 관리기 (162) 는 디바이스 (104) 의 제 2 동작 모드 (예를 들어, 디스플레이 중단 데이터 모드 또는 패쇄 자막 디스플레이 모드) 에서 텍스트 스트림 (121) (예를 들어, 주석이 달린 텍스트 스트림 (137)) 을 재생하는 것을 억제한다. 예를 들어, 미팅 관리기 (162) 는 (예를 들어, 디바이스 (102) 의 제 2 동작 모드 동안) 텍스트 스트림 (121) 을 수신하지 않고, 제 2 동작 모드 (예를 들어, 디스플레이 중단 데이터 모드 또는 패쇄 자막 디스플레이 모드) 에서 텍스트 스트림 (121) 을 생성하지 않는다. 다른 예로서, 미팅 관리기 (162) 는 텍스트 스트림 (121) 을 수신하고, 디바이스 (104) 의 제 2 동작 모드 (예를 들어, 디스플레이 중단 데이터 모드 또는 패쇄 자막 디스플레이 모드) 를 검출하는 것에 응답하여, 텍스트 스트림 (121) (예를 들어, 주석이 달린 텍스트 스트림 (137)) 을 재생하는 것을 억제한다. 특정 양태에서, 중단 관리기 (164) 는, 디바이스 (104) 의 제 2 동작 모드 (예를 들어, 디스플레이 중단 데이터 모드) 에서, 미디어 스트림 (109) 에서 중단이 검출되지 않았다는 결정에 응답하여 텍스트 스트림 (121) (예를 들어, 주석이 달린 텍스트 스트림 (137)) 을 재생하는 것을 억제한다 (예를 들어, 텍스트 스트림 (121) 에 대응하는 미디어 스트림 (109) 의 부분들이 수신되었다).

[0039] 특정 양태에서, 중단 관리기 (164) 는 온라인 미팅의 시작 전에 또는 그 근처에서 일반적인 스피치 모델에 기반하여 인공 신경망과 같은 스피치 모델 (131) 을 초기화한다. 특정 양태에서, 중단 관리기 (164) 는 일반 스피치 모델이 사용자의 연령, 위치, 성별, 또는 이들의 조합과 같은 사용자 (142) 의 인구통계학적 데이터와 일치한다고 (예를 들어, 연관된다고) 결정하는 것에 기반하여 복수의 일반 스피치 모델들로부터 일반 스피치 모델

을 선택한다. 특정 양태에서, 중단 관리기 (164) 는 사용자 (142) 의 연락처 정보 (예를 들어, 이름, 위치, 전화 번호, 주소, 또는 이들의 조합) 에 기반하여 온라인 미팅 (예를 들어, 스케줄링된 미팅) 이전에 인구통계 데이터를 예측한다. 특정 양태에서, 중단 관리기 (164) 는 온라인 미팅의 시작 부분 동안 스피치 오디오 스트림 (111), 비디오 스트림 (113), 또는 둘 다에 기반하여 인구통계 데이터를 추정한다. 예를 들어, 중단 관리기 (164) 는 스피치 오디오 스트림 (111), 비디오 스트림 (113) 또는 둘 다를 분석하여, 사용자 (142) 의 나이, 지역 액센트, 성별 또는 이들의 조합을 추정한다. 특정 양태에서, 중단 관리기 (164) 는 사용자 (142) 와 연관된 (예를 들어, 사용자 식별자와 매칭하는) (예를 들어, 이전에 생성된) 스피치 모델 (131) 을 검색한다.

[0040] 특정 양태에서, 중단 관리기 (164) 는 온라인 미팅 동안 (예를 들어, 스피치 오디오 스트림 (111) 에서의 중단 이전에) 스피치 오디오 스트림 (111) 에서 검출된 스피치에 기반하여 스피치 모델 (131) 을 트레이닝 (예를 들어, 생성 또는 업데이트) 한다. 설명을 위해, 텍스트-대-스피치 변환기 (166) 는 텍스트-대-스피치 변환을 수행하기 위해 스피치 모델 (131) 을 사용하도록 구성된다. 특정 양태에서, 중단 관리기 (164) 는 (예를 들어, 디바이스 (102) 의 제 1 동작 모드 동안) 스피치 오디오 스트림 (111) 에 대응하는 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 수신하거나 (예를 들어, 디바이스 (102) 의 제 2 동작 모드 동안) 생성한다. 텍스트-대-스피치 변환기 (166) 는 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다에 대해 텍스트-대-스피치 변환을 수행함으로써 합성 스피치 오디오 스트림 (133) 을 생성하기 위해 스피치 모델 (131) 을 사용한다. 중단 관리기 (164) 는 스피치 오디오 스트림 (111) 및 합성 스피치 오디오 스트림 (133) 의 비교에 기반하여 스피치 모델 (131) 을 업데이트하기 위해 트레이닝 기법들을 사용한다. 스피치 모델 (131) 이 인공 신경망을 포함하는 예시적인 예에서, 중단 관리기 (164) 는 역전파 (backpropagation) 를 사용하여 스피치 모델 (131) 의 가중치 및 바이어스를 업데이트한다. 일부 양태들에 따르면, 스피치 모델 (131) 은, 스피치 모델 (131) 을 사용하는 후속 텍스트-대-스피치 변환이 사용자 (142) 의 스피치 특성들의 더 근접한 매칭인 합성 스피치를 생성할 가능성이 더 많도록 업데이트된다.

[0041] 특정 양태에서, 중단 관리기 (164) 는 사용자 (142) 의 아바타 (135) (예를 들어, 시각적 표현) 를 생성한다. 특정 양태에서, 아바타 (135) 는 도 3a ~ 도 3c 을 참조하여 추가로 설명되는 바와 같이, 스피치 모델 (131) 의 트레이닝의 레벨을 표시하는 트레이닝 표시자를 포함하거나 이에 대응한다. 예를 들어, 중단 관리기 (164) 는, 제 1 트레이닝 기준이 충족되지 않았다고 결정한 것에 응답하여, 아바타 (135) 를 스피치 모델 (131) 이 트레이닝되지 않았음을 나타내는 제 1 시각적 표현으로 초기화한다. 온라인 미팅 동안, 중단 관리기 (164) 는, 제 1 트레이닝 기준이 만족되고 제 2 트레이닝 기준이 만족되지 않는다고 결정하는 것에 응답하여, 스피치 모델 (131) 의 트레이닝이 진행 중임을 나타내기 위해 아바타 (135) 를 제 1 시각적 표현으로부터 제 2 시각적 표현으로 업데이트한다. 중단 관리기 (164) 는, 제 2 트레이닝 기준이 충족된다는 결정에 응답하여, 아바타 (135) 를 (스피치 모델 (131) 의 트레이닝이 완료되었음을 나타내기 위해) 제 3 시각적 표현으로 업데이트한다.

[0042] 트레이닝 기준은 스피치 모델 (131) 을 트레이닝하기 위해 사용되는 오디오 샘플들의 카운트, 스피치 모델 (131) 을 트레이닝하기 위해 사용되는 오디오 샘플들의 재생 지속기간, 스피치 모델 (131) 을 트레이닝하기 위해 사용되는 오디오 샘플들의 커버리지, 스피치 모델 (131) 의 성공 메트릭, 또는 이들의 조합에 기반할 수 있다. 특정 양태에서, 스피치 모델 (131) 을 트레이닝하기 위해 사용되는 오디오 샘플들의 커버리지는 오디오 샘플들에 의해 표현되는 별개의 사운드들 (예를 들어, 모음, 자음 등) 에 대응한다. 특정 양태에서, 성공 메트릭은 스피치 모델 (131) 을 트레이닝하기 위해 사용된 오디오 샘플들과 스피치 모델 (131) 에 기반하여 생성된 합성 스피치의 비교 (예를 들어, 그 사이의 매치) 에 기반한다.

[0043] 일부 구현들에 따르면, 아바타 (135) 의 제 1 컬러, 제 1 음영, 제 1 크기, 제 1 애니메이션, 또는 이들의 조합은 스피치 모델 (131) 이 트레이닝되지 않았음을 나타낸다. 아바타 (135) 의 제 2 컬러, 제 2 음영, 제 2 크기, 제 2 애니메이션, 또는 이들의 조합은 스피치 모델 (131) 이 부분적으로 트레이닝됨을 나타낸다. 아바타 (135) 의 제 3 컬러, 제 3 음영, 제 3 크기, 제 3 애니메이션, 또는 이들의 조합은 스피치 모델 (131) 의 트레이닝이 완료됨을 나타낸다. 특정 양태에서, GUI 생성기 (168) 는 아바타 (135) 의 시각적 표현을 표시하기 위해 GUI (145) 를 생성 (또는 업데이트) 한다.

[0044] 특정 양태에서, 중단 관리기 (124) 는 디바이스 (104) 에 대한 통신 링크에서 네트워크 이슈들 (예를 들어, 감소된 대역폭) 을 검출한다. 중단 관리기 (124) 는, 네트워크 이슈들을 검출하는 것에 응답하여, 스피치 오디오 스트림 (111) 에서의 중단을 나타내는 중단 통지 (119) 를 디바이스 (104) 에 전송하거나, 네트워크 이슈들이 해결됨을 검출할 때까지 미디어 스트림 (109) 의 후속 미디어 프레임들을 디바이스 (104) 에 전송하는 것

을 억제 (예를 들어, 전송을 중단) 하거나, 또는 둘 다를 수행한다. 예를 들어, 중단 관리기 (124) 는, 네트워크 이슈들을 검출하는 것에 응답하여, 중단의 종료까지 스피치 오디오 스트림 (111), 비디오 스트림 (113), 또는 둘 다를 디바이스 (104) 에 전송하는 것 (예를 들어, 전송을 중단함) 을 억제한다.

[0045] 중단 관리기 (124) 는 후속 미디어 프레임들에 대응하는 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 전송한다. 예를 들어, 중단 관리기 (124) 는, 디바이스 (102) 의 제 1 동작 모드 (예를 들어, 전송 자막 데이터 모드) 에서, 후속 미디어 프레임들에 대응하는 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 계속 전송한다. 설명을 위해, 제 1 동작 모드 (예를 들어, 전송 자막 데이터 모드) 에서, 미팅 관리기 (122) 는 미디어 스트림 (109), 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 이들의 조합을 생성한다. 중단 관리기 (124) 는, 제 1 동작 모드 (예를 들어, 전송 자막 데이터 모드) 에서 네트워크 이슈들을 검출하는 것에 응답하여, 미디어 스트림 (109) 의 후속 미디어 프레임들의 전송을 중단하고, 디바이스 (104) 로의 후속 미디어 프레임들에 대응하는 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다의 전송을 계속한다. 대안적으로, 중단 관리기 (124) 는, 디바이스 (102) 의 제 2 동작 모드 (예를 들어, 전송 중단 데이터 모드) 에서 네트워크 이슈들을 검출하는 것에 응답하여, 후속 미디어 프레임들에 대응하는 오디오 입력 (153) 에 기반하여 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 생성한다. 설명을 위해, 제 2 동작 모드 (예를 들어, 전송 중단 데이터 모드) 에서, 미팅 관리기 (122) 는 미디어 스트림 (109) 을 생성하고, 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 생성하지 않는다. 중단 관리기 (124) 는, 디바이스 (102) 의 제 2 동작 모드 (예를 들어, 전송 자막 데이터 모드) 에서 네트워크 이슈들을 검출하는 것에 응답하여, 미디어 스트림 (109) 의 후속 미디어 프레임들의 송신을 중단하고, 디바이스 (104) 로의 후속 미디어 프레임들에 대응하는 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다의 전송을 시작한다. 특정 양태에서, 디바이스 (102) 의 제 2 동작 모드 (예를 들어, 중단 데이터 전송 모드) 에서, 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 디바이스 (104) 에 전송하는 것은 중단 통지 (119) 를 디바이스 (104) 에 전송하는 것에 대응한다.

[0046] 특정 양태에서, 중단 관리기 (164) 는 디바이스 (102) 로부터 중단 통지 (119) 를 수신하는 것에 응답하여 스피치 오디오 스트림 (111) 에서의 중단을 검출한다. 특정 양태에서, 디바이스 (102) 가 제 2 동작 모드 (예를 들어, 전송 중단 데이터 모드) 에서 동작하고 있을 때, 중단 관리기 (164) 는 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 수신하는 것에 응답하여 스피치 오디오 스트림 (111) 에서의 중단을 검출한다.

[0047] 특정 양태에서, 중단 관리기 (164) 는 스피치 오디오 스트림 (111) 의 마지막 수신된 오디오 프레임의 임계 지속기간 내에 스피치 오디오 스트림 (111) 의 어떠한 오디오 프레임들도 수신되지 않는다는 결정에 응답하여 스피치 오디오 스트림 (111) 에서의 중단을 검출한다. 예를 들어, 스피치 오디오 스트림 (111) 의 마지막 수신된 오디오 프레임은 디바이스 (104) 에서 제 1 수신 시간에 수신된다. 중단 관리기 (164) 는 스피치 오디오 스트림 (111) 의 오디오 프레임들이 제 1 수신 시간의 임계 지속기간 내에 수신되지 않는다는 결정에 응답하여 중단을 검출한다. 특정 양태에서, 중단 관리기 (164) 는 중단 통지를 디바이스 (102) 에 전송한다. 특정 양태에서, 중단 관리기 (124) 는 디바이스 (104) 로부터 중단 통지를 수신하는 것에 응답하여 네트워크 이슈들을 검출한다. 중단 관리기 (124) 는, 네트워크 이슈들을 검출하는 것에 응답하여, 전송한 바와 같이, 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 (예를 들어, 미디어 스트림 (109) 의 후속 미디어 프레임들을 전송하는 대신에) 디바이스 (104) 에 전송한다.

[0048] 중단 관리기 (164) 는 중단을 검출하는 것에 응답하여, 텍스트 스트림 (121) 에 기반하여 출력을 선택적으로 생성한다. 예를 들어, 중단 관리기 (164) 는, 중단에 응답하여, 텍스트 스트림 (121), 메타데이터 스트림 (123), 주석이 달린 텍스트 스트림 (137), 또는 이들의 조합을 텍스트-대-스피치 변환기 (166) 에 제공한다. 텍스트-대-스피치 변환기 (166) 는 텍스트 스트림 (121), 메타데이터 스트림 (123), 주석이 달린 텍스트 스트림 (137), 또는 이들의 조합에 기반하여 텍스트-대-스피치 변환을 수행하기 위해 스피치 모델 (131) 을 사용하여 합성 스피치 오디오 스트림 (133) 을 생성한다. 예를 들어, 텍스트 스트림 (121) 에 기반하고 메타데이터 스트림 (123) 과 독립적인 합성 스피치 오디오 스트림 (133) 은 스피치 모델 (131) 에 의해 표현되는 사용자 (142) 의 중심 스피치 특성들을 갖는 텍스트 스트림 (121) 에 의해 표시되는 스피치에 대응한다. 다른 예로서, 주석이 달린 텍스트 스트림 (137) (예컨대, 텍스트 스트림 (121) 및 메타데이터 스트림 (123)) 에 기반한 합성 스피치 오디오 스트림 (133) 은 메타데이터 스트림 (123) 에 의해 표시된 억양을 갖는 스피치 모델 (131) 에 의해 표현되는 바와 같은 사용자 (142) 의 스피치 특성을 갖는 텍스트 스트림 (121) 에 의해 표시되는 스피치에 대응한다. 텍스트-대-스피치 변환을 수행하기 위해 사용자 (142) 의 스피치에 대해 적어도 부분적으로 트레이닝된 스피치 모델 (131)(예를 들어, 스피치 오디오 스트림 (111)) 을 사용하는 것은 합성 스피치 오디오

스트림 (133) 이 사용자 (142) 의 스피치 특성들에 더 근접하게 매칭할 수 있게 한다. 중단에 응답하여, 중단 관리기 (164) 는 합성 스피치 오디오 스트림 (133) 을 오디오 출력 (143) 으로서 스피커 (154) 에 제공하거나, 스피치 오디오 스트림 (111) 의 재생을 중단시키거나, 비디오 스트림 (113) 의 재생을 중단시키거나, 또는 이들의 조합을 수행한다.

[0049] 특정 양태에서, 중단 관리기 (164) 는 합성 스피치 오디오 스트림 (133) 을 오디오 출력 (143) 으로서 스피커 (154) 에 제공하는 것과 동시에 아바타 (135) 를 선택적으로 디스플레이한다. 예를 들어, 중단 관리기 (164) 는 스피치 오디오 스트림 (111) 을 오디오 출력 (143) 으로서 스피커 (154) 에 제공하는 동안 아바타 (135) 를 디스플레이하는 것을 억제한다. 다른 예로서, 중단 관리기 (164) 는 합성 스피치 오디오 스트림 (133) 을 오디오 출력 (143) 으로서 스피커 (154) 에 제공하면서 아바타 (135) 를 디스플레이한다. 설명을 위해, GUI 생성기 (168) 는 합성 스피치 오디오 스트림 (133) 이 스피커 (154) 에 의한 재생을 위해 오디오 출력 (143) 으로서 출력되는 동안 비디오 스트림 (113) 대신에 아바타 (135) 를 디스플레이하도록 GUI (145) 를 업데이트한다. 특정 양태에서, 중단 관리기 (164) 는 스피치 오디오 스트림 (111) 을 오디오 출력 (143) 으로서 스피커 (154) 에 제공하는 것과 동시에 아바타 (135) 의 제 1 표현을 디스플레이하고, 합성 스피치 오디오 스트림 (133) 을 오디오 출력 (143) 으로서 스피커 (154) 에 제공하는 것과 동시에 아바타 (135) 의 제 2 표현을 디스플레이한다. 예를 들어, 도 3c 를 참조하여 추가로 설명되는 바와 같이, 제 1 표현은 아바타 (135) 가 트레이닝되고 있거나 트레이닝되었음을 나타내고 (예를 들어, 스피치 모델 (131) 의 트레이닝 표시자), 제 2 표현은 아바타 (135) 가 말하고 있음을 나타낸다 (예를 들어, 스피치 모델 (131) 은 합성 스피치를 생성하는데 사용되고 있음).

[0050] 특정 구현에서, 중단 관리기 (164) 는 디스플레이 디바이스 (156) 에 대한 출력으로서 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 둘 다를 선택적으로 제공한다. 예를 들어, 중단 관리기 (164) 는, 디바이스 (104) 의 제 2 동작 모드 (예를 들어, 디스플레이 중단 데이터 모드) 동안의 중단에 응답하여, 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 둘 다를 디스플레이하기 위해 GUI (145) 를 업데이트하도록 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 둘 다를 GUI 생성기 (168) 에 제공한다. 대안적인 구현에서, 중단 관리기 (164) 는, 디바이스 (104) 의 제 1 동작 모드 (예를 들어, 디스플레이 자막 데이터 모드) 동안, 디스플레이 디바이스 (156) 에 대한 출력으로서 (예를 들어, 중단과 독립적으로) 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 둘 다를 계속 제공한다. 특정 양태에서, 중단 관리기 (164) 는 합성 스피치 오디오 스트림 (133) 을 오디오 출력 (143) 으로서 스피커 (154) 에 제공하는 것과 동시에, 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 둘 다를 디스플레이 디바이스 (156) 에 제공한다.

[0051] 특정 구현에서, 중단 관리기 (164) 는, 중단 구성 설정에 기반하고 중단에 응답하여, 합성 스피치 오디오 스트림 (133), 텍스트 스트림 (121), 또는 주석이 달린 텍스트 스트림 (137) 중 하나 이상을 출력한다. 예를 들어, 중단 관리기 (164) 는, 중단 그리고 중단 구성 설정이 제 1 값 (예를 들어, 0 또는 "오디오 및 텍스트") 을 갖는다고 결정하는 것에 응답하여, 합성 스피치 오디오 스트림 (133) 을 오디오 출력 (143) 으로서 스피커 (154) 에 제공하는 것과 동시에 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 둘 다를 디스플레이 디바이스 (156) 에 제공한다. 중단 관리기 (164) 는, 중단 그리고 중단 구성 설정이 제 2 값 (예를 들어, 1 또는 "텍스트만") 을 갖는다고 결정하는 것에 응답하여, 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 둘 다를 디스플레이 디바이스 (156) 에 제공하고 오디오 스트림 (143) 을 스피커 (154) 에 제공하는 것을 억제한다. 중단 관리기 (164) 는, 중단 그리고 중단 구성 설정이 제 3 값 (예를 들어, 2 또는 "오디오만") 을 갖는다고 결정하는 것에 응답하여, 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 둘 다를 디스플레이 디바이스 (156) 에 제공하는 것을 억제하고 합성 스피치 오디오 스트림 (133) 을 오디오 출력 (143) 으로서 스피커 (154) 에 제공한다. 특정 양태에서, 중단 구성 설정은 디폴트 데이터, 사용자 입력, 또는 둘 모두에 기초한다.

[0052] 특정 양태에서, 중단 관리기 (124) 는 중단이 종료되었음을 검출하고, 중단 종료 통지를 디바이스 (104) 에 전송한다. 예를 들어, 중단 관리기 (124) 는 디바이스 (104) 와의 통신 링크의 이용가능한 통신 대역폭이 임계치보다 크다는 결정에 응답하여 중단이 종료되었음을 검출한다. 특정 양태에서, 중단 관리기 (164) 는 디바이스 (102) 로부터 중단 종료 통지를 수신하는 것에 응답하여 중단이 종료되었음을 검출한다.

[0053] 다른 특정 양태에서, 중단 관리기 (164) 는 중단이 종료되었음을 검출하고, 중단 종료 통지를 디바이스 (102) 에 전송한다. 예를 들어, 중단 관리기 (164) 는 디바이스 (102) 와의 통신 링크의 이용가능한 통신 대역폭이 임계치보다 크다는 결정에 응답하여 중단이 종료되었음을 검출한다. 특정 양태에서, 중단 관리기 (124)

는 디바이스 (104) 로부터 중단 종료 통지를 수신하는 것에 응답하여 중단이 종료되었음을 검출한다.

[0054] 미팅 관리기 (122) 는, 중단이 종료되었음을 검출하는 것에 응답하여, 스피치 오디오 스트림 (111), 비디오 스트림 (113), 또는 둘 다의 디바이스 (104) 로의 전송을 재개한다. 특정 양태에서, 스피치 오디오 스트림 (111), 비디오 스트림 (113), 또는 둘 다의 전송은 중단 종료 통지의 전송에 대응한다. 중단이 디바이스 (102) 의 제 2 동작 모드 (예를 들어, 전송 중단 데이터 모드) 동안 종료되었음을 검출하는 것에 응답하여, 중단 관리기 (124) 는 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 디바이스 (104) 에 전송하는 것을 억제한다.

[0055] 미팅 관리기 (162) 는, 중단이 종료되었음을 검출하는 것에 응답하여, 텍스트 스트림 (121) 에 기반하여 합성 스피치 오디오 스트림 (133) 을 생성하는 것을 억제하고, 합성 스피치 오디오 스트림 (133) 을 오디오 출력 (143) 으로서 스피커 (154) 에 제공하는 것을 억제하며 (예를 들어, 중단시키고), 스피치 오디오 스트림 (111) 을 오디오 출력 (143) 으로서 스피커 (154) 에 재생하는 것을 재개한다. 미팅 관리기 (162) 는, 중단이 종료되었음을 검출하는 것에 응답하여, 비디오 스트림 (113) 을 디스플레이 디바이스 (156) 에 제공하는 것을 재개한다. 예를 들어, 미팅 관리기 (162) 는 비디오 스트림 (113) 을 디스플레이하기 위해 GUI (145) 를 업데이트하도록 비디오 스트림 (113) 을 GUI 생성기 (168) 에 제공한다.

[0056] 특정 양태에서, 중단이 종료되었음을 검출하는 것에 응답하여, 중단 관리기 (164) 는, 스피치 모델 (131) 이 합성 스피치 오디오를 출력하는데 사용되고 있지 않음 (예를 들어, 아바타 (135) 가 말하고 있지 않음) 을 나타내기 위해 GUI (145) 를 업데이트하라는 제 1 요청을 GUI 생성기 (168) 에 전송한다. GUI 생성기 (168) 는, 제 1 요청을 수신하는 것에 응답하여, 스피치 모델 (131) 이 트레이닝되고 있거나 트레이닝되었고 스피치 모델 (131) 이 합성 스피치 오디오를 출력하는데 사용되고 있지 않음 (예를 들어, 아바타 (135) 가 말하고 있지 않음) 을 나타내는 아바타 (135) 의 제 1 표현을 디스플레이하도록 GUI (145) 를 업데이트한다. 대안적인 양태에서, 중단이 종료되었음을 검출하는 것에 응답하여, 중단 관리기 (164) 는 아바타 (135) 의 디스플레이를 중단하기 위해 제 2 요청을 GUI 생성기 (168) 에 전송한다. 예를 들어, GUI 생성기 (168) 는, 제 2 요청을 수신하는 것에 응답하여, 아바타 (135) 를 디스플레이하는 것을 억제하도록 GUI (145) 를 업데이트한다.

[0057] 특정 양태에서, 중단이 제 2 동작 모드 동안 종료되었음을 검출하는 것 (예를 들어, 중단 데이터를 더 많이 디스플레이하거나 또는 자막이 없는 데이터 모드) 에 응답하여, 중단 관리기 (164) 는 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 둘 다를 디스플레이 디바이스 (156) 에 제공하는 것을 억제한다. 예를 들어, GUI 생성기 (168) 는 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 둘 다를 디스플레이하는 것을 억제하도록 GUI (145) 를 업데이트한다.

[0058] 따라서, 시스템 (100) 은 온라인 미팅 동안 스피치 오디오 스트림 (111) 의 중단 동안 정보 손실을 감소 (예를 들어, 제거) 한다. 예를 들어, 사용자 (144) 는, 네트워크 이슈들이 스피치 오디오 스트림 (111) 이 디바이스 (104) 에 의해 수신되는 것을 방지하더라도 텍스트가 디바이스 (104) 에 의해 수신될 수 있는 경우들에서, 사용자 (142) 의 스피치에 대응하는 오디오 (예를 들어, 합성 스피치 오디오 스트림 (133)), 텍스트 (예를 들어, 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 둘 다), 또는 이들의 조합을 계속해서 수신한다.

[0059] 카메라 (150) 및 마이크로폰 (152) 이 디바이스 (102) 에 커플링되는 것으로서 도시되지만, 다른 구현들에서 카메라 (150), 마이크로폰 (152), 또는 둘 다는 디바이스 (102) 에 통합될 수도 있다. 스피커 (154) 및 디스플레이 디바이스 (156) 가 디바이스 (104) 에 커플링되는 것으로서 도시되지만, 다른 구현들에서 스피커 (154), 디스플레이 디바이스 (156), 또는 둘 다는 디바이스 (104) 에 통합될 수도 있다. 하나의 마이크로폰 및 하나의 스피커가 도시되지만, 다른 구현들에서, 사용자 스피치를 캡처하도록 구성된 하나 이상의 추가적인 마이크 로폰들, 스피치 오디오를 출력하도록 구성된 하나 이상의 추가적인 스피커들, 또는 이들의 조합이 포함될 수도 있다.

[0060] 설명의 용이함을 위해 디바이스 (102) 는 전송 디바이스로서 설명되고 디바이스 (104) 는 수신 디바이스로서 설명된다는 것을 이해해야 한다. 통화 동안, 디바이스 (102) 및 디바이스 (104) 의 역할들은 사용자 (144) 가 말하기 시작할 때 스위칭할 수 있다. 예를 들어, 디바이스 (104) 는 전송 디바이스일 수 있고, 디바이스 (102) 는 수신 디바이스일 수 있다. 설명을 위해, 디바이스 (104) 는 사용자 (144) 의 오디오 및 비디오를 캡처하기 위한 마이크로폰 및 카메라를 포함할 수 있고, 디바이스 (102) 는 사용자 (142) 에게 오디오 및 비디오를 재생하기 위해 스피커 및 디스플레이를 포함하거나 그에 커플링될 수 있다. 특정 양태에서, 예를 들어, 사용자 (142) 및 사용자 (144) 둘 다가 동시에 또는 겹치는 시간들에서 말하고 있을 때, 디바이스 (102)

및 디바이스 (104) 각각은 송신 디바이스 및 수신 디바이스일 수 있다.

[0061] 특정 양태에서, 미팅 관리기 (122) 는 또한 미팅 관리기 (162) 를 참조하여 설명된 하나 이상의 동작들을 수행하도록 구성되며, 그 반대도 마찬가지이다. 특정 양태에서, 중단 관리기 (124) 는 또한 중단 관리기 (164) 를 참조하여 설명된 하나 이상의 동작들을 수행하도록 구성되며, 그 반대도 마찬가지이다. GUI 생성기 (168) 가 미팅 관리기 (162) 및 중단 관리기 (164) 와 별개인 것으로 설명되지만, 다른 구현들에서, GUI 생성기 (168) 는 미팅 관리기 (162), 중단 관리기 (164), 또는 둘 다에 통합된다. 설명을 위해, 일부 예들에서, 미팅 관리기 (162), 중단 관리기 (164), 또는 둘 다는 GUI 생성기 (168) 를 참조하여 설명된 일부 동작들을 수행하도록 구성된다.

[0062] 도 2 를 참조하면, 스피치 오디오 스트림 중단을 처리하도록 동작가능한 시스템이 도시되고 일반적으로 200 으로 지정된다. 특정 양태에서, 도 1 의 시스템 (100) 은 시스템 (200) 의 하나 이상의 컴포넌트들을 포함한다.

[0063] 시스템 (200) 은 네트워크 (106) 를 통해 디바이스 (102) 및 디바이스 (104) 에 결합된 서버 (204) 를 포함한다. 서버 (204) 는 미팅 관리기 (122) 및 중단 관리기 (124) 를 포함한다. 서버 (204) 는 온라인 미팅 데이터를 디바이스 (102) 로부터 디바이스 (104) 로, 그리고 그 반대로 전달하도록 구성된다. 예를 들어, 미팅 관리기 (122) 는 디바이스 (102) 와 디바이스 (104) 간의 온라인 미팅을 설정하도록 구성된다.

[0064] 디바이스 (102) 는 미팅 관리기 (222) 를 포함한다. 온라인 미팅 동안, 미팅 관리기 (222) 는 미디어 스트림 (109) (예를 들어, 스피치 오디오 스트림 (111), 비디오 스트림 (113), 또는 둘 다) 을 서버 (204) 에 전송한다. 서버 (204) 의 미팅 관리기 (122) 는 디바이스 (102) 로부터 미디어 스트림 (109) (예를 들어, 스피치 오디오 스트림 (111), 비디오 스트림 (113), 또는 둘 다) 을 수신한다. 특정 구현에서, 디바이스 (102) 는 미디어 스트림 (109) 을 서버 (204) 에 전송하는 것과 동시에, 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 전송한다.

[0065] 특정 양태에서, 후속 동작들은 도 1 을 참조하여 설명된 바와 같이 수행되고, 서버 (204) 는 디바이스 (102) 를 대신한다. 예를 들어, 미팅 관리기 (122) (도 1 에서와 같이 디바이스 (102) 대신에 서버 (204) 에서 동작함) 는 도 1 을 참조하여 설명된 것과 유사한 방식으로 미디어 스트림 (109), 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 이들의 조합을 디바이스 (104) 에 전송한다. 예를 들어, 미팅 관리기 (122) 는, 서버 (204) 의 제 1 동작 모드 (예를 들어, 전송 자막 데이터 모드) 동안, 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 전송한다. 특정 구현에서, 미팅 관리기 (122) 는 디바이스 (102) 로부터 수신된 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 디바이스 (104) 에 전달한다. 일부 구현들에서, 미팅 관리기 (122) 는 텍스트 스트림 (121), 미디어 스트림 (109), 또는 이들의 조합에 기반하여 메타데이터 스트림 (123) 을 생성한다. 이러한 구현들에서, 미팅 관리기 (122) 는 디바이스 (102) 로부터 수신된 텍스트 스트림 (121) 을 디바이스 (104) 에 전달하거나, 서버 (204) 에서 생성된 메타데이터 스트림 (123) 을 디바이스 (104) 에 전송하거나, 또는 둘 다를 수행한다. 일부 구현들에서, 미팅 관리기 (122) 는 미디어 스트림 (109) 에 기반하여 텍스트 스트림 (121), 메타데이터 스트림 (123) 또는 둘 다를 생성하고, 텍스트 스트림 (121), 메타데이터 스트림 (123) 또는 둘 다를 디바이스 (104) 에 전달한다. 대안적으로, 미팅 관리기 (122) 는, 서버 (204) 의 제 2 동작 모드 (예를 들어, 전송 중단 데이터 모드) 동안, 어떠한 중단도 검출되지 않는다고 결정하는 것에 응답하여 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 전송하는 것을 억제한다. 디바이스 (104) 는 서버 (204) 로부터 네트워크 (106) 를 통해 미디어 스트림 (109), 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 이들의 조합을 수신한다. 미팅 관리기 (162) 는 도 1 을 참조하여 설명된 바와 같이, 미디어 스트림 (109), 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 이들의 조합의 미디어 프레임들을 재생한다. 중단 관리기 (164) 는 도 1 을 참조하여 설명된 바와 같이 스피치 모델 (131) 을 트레이닝하고, 아바타 (135) 를 디스플레이하거나, 또는 둘 다를 수행한다.

[0066] 특정 양태에서, 중단 관리기 (124) 는, 네트워크 이슈들을 검출하는 것에 응답하여, 스피치 오디오 스트림 (111) 에서의 중단을 나타내는 중단 통지 (119) 를 디바이스 (104) 에 전송하거나, 네트워크 이슈들이 해결됨 (예를 들어 중단이 종료되었음) 을 검출할 때까지 미디어 스트림 (109) 의 후속 미디어 프레임들을 디바이스 (104) 에 전송하는 것을 억제 (예를 들어, 전송을 중단) 하거나, 또는 둘 다를 수행한다. 중단 관리기 (124) 는, 도 1 을 참조하여 설명된 바와 같이, 후속 미디어 프레임들에 대응하는 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 디바이스 (104) 에 전송한다. 특정 구현에서, 중단 관리기 (124) 는 디바이스 (102) 로부터 수신된 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 디바이스 (104) 에 전

송한다. 일부 예들에서, 중단 관리기 (124) 는 서버 (204) 에서 생성된, 메타데이터 스트림 (123), 텍스트 스트림 (121), 또는 둘 다를 디바이스 (104) 에 전송한다. 특정 양태에서, 중단 관리기 (124) 는, 서버 (204) 의 제 2 동작 모드 (예를 들어, 전송 중단 데이터 모드) 동안, 스피치 오디오 스트림 (111) 에서의 중단을 검출하는 것에 응답하여, 메타데이터 스트림 (123), 텍스트 스트림 (121), 또는 둘 다를 선택적으로 생성한다.

[0067] 특정 양태에서, 중단 관리기 (164) 는, 도 1 을 참조하여 설명된 것과 유사한 방식으로, (예를 들어, 서버 (204) 에서) 중단 관리기 (124) 로부터 중단 통지 (119) 를 수신하는 것, 서버 (204) 가 제 2 동작 모드 (예를 들어, 전송 중단 데이터 모드) 에서 동작하고 있을 때 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 수신하는 것, 스피치 오디오 스트림 (111) 의 마지막 수신된 오디오 프레임의 임계 지속시간 내에 스피치 오디오 스트림 (111) 의 어떠한 오디오 프레임들도 수신되지 않는다고 결정하는 것, 또는 이들의 조합에 응답하여, 스피치 오디오 스트림 (111) 에서의 중단을 검출한다. 특정 양태에서, 중단 관리기 (164) 는 중단 통지를 서버 (204) 에 전송한다. 특정 양태에서, 중단 관리기 (124) 는 디바이스 (104) 로부터 중단 통지를 수신하는 것에 응답하여 네트워크 이슈들을 검출한다. 중단 관리기 (124) 는, 도 1 을 참조하여 설명된 바와 같이, 후속 미디어 프레임들에 대응하는 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 디바이스 (104) 에 전송한다.

[0068] 중단 관리기 (164) 는, 중단을 검출하는 것에 응답하여, 텍스트 스트림 (121), 메타데이터 스트림 (123), 주석이 달린 텍스트 스트림 (137), 또는 이들의 조합을 텍스트-대-스피치 변환기 (166) 에 제공한다. 텍스트-대-스피치 변환기 (166) 는, 도 1 을 참조하여 설명되는 바와 같이, 텍스트 스트림 (121), 메타데이터 스트림 (123), 주석이 달린 텍스트 스트림 (137), 또는 이들의 조합에 기반하여 텍스트-대-스피치 변환을 수행하기 위해 스피치 모델 (131) 을 사용하여 합성 스피치 오디오 스트림 (133) 을 생성한다. 중단에 응답하여, 중단 관리기 (164) 는, 도 1 을 참조하여 설명된 바와 같이, 합성 스피치 오디오 스트림 (133) 을 오디오 출력 (143) 으로서 스피커 (154) 에 제공하고, 스피치 오디오 스트림 (111) 의 재생을 중단시키고, 비디오 스트림 (113) 의 재생을 중단시키고, 아바타 (135) 를 디스플레이하고, 아바타 (135) 의 특정 표현을 디스플레이하고, 텍스트 스트림 (121) 을 디스플레이하고, 주석이 달린 텍스트 스트림 (137) 을 디스플레이하거나, 또는 이들의 조합을 수행한다.

[0069] 미팅 관리기 (122) 는, 중단이 종료되었음을 검출하는 것에 응답하여, 스피치 오디오 스트림 (111), 비디오 스트림 (113), 또는 둘 다의 디바이스 (104) 로의 전송을 재개한다. 특정 양태에서, 중단이 서버 (204) 의 제 2 동작 모드 (예를 들어, 전송 중단 데이터 모드) 동안 종료되었음을 검출하는 것에 응답하여, 중단 관리기 (124) 는 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 디바이스 (104) 에 전송하는 것을 억제 (예를 들어 송신을 중단) 한다.

[0070] 미팅 관리기 (162) 는, 중단이 종료되었음을 검출하는 것에 응답하여, 텍스트 스트림 (121) 에 기반하여 합성 스피치 오디오 스트림 (133) 을 생성하는 것을 억제하고, 합성 스피치 오디오 스트림 (133) 을 오디오 출력 (143) 으로서 스피커 (154) 에 제공하는 것을 억제 (예를 들어, 중단) 하고, 오디오 출력 (143) 으로서 스피치 오디오 스트림 (111) 의 재생을 스피커 (154) 에 재개하고, 비디오 스트림 (113) 을 디스플레이 디바이스 (156) 에 제공하는 것을 재개하고, 아바타 (135) 의 디스플레이를 중단 또는 조정하고, 텍스트 스트림 (121) 을 디스플레이 디바이스 (156) 에 제공하는 것을 억제하고, 주석이 달린 텍스트 스트림 (137) 을 디스플레이 디바이스 (156) 에 제공하는 것을 억제하고, 또는 이들의 조합을 수행한다.

[0071] 따라서, 시스템 (200) 은 레거시 디바이스 (예를 들어, 중단 관리기를 포함하지 않는 디바이스 (102)) 와의 온라인 미팅 동안 스피치 오디오 스트림 (111) 의 중단 동안 정보 손실을 감소 (예를 들어, 제거) 한다. 예를 들어, 사용자 (144) 는, 네트워크 이슈들이 스피치 오디오 스트림 (111) 이 디바이스 (104) 에 의해 수신되는 것을 방지하더라도 텍스트가 디바이스 (104) 에 의해 수신될 수 있는 경우들에서, 사용자 (142) 의 스피치에 대응하는 오디오 (예를 들어, 합성 스피치 오디오 스트림 (133)), 텍스트 (예를 들어, 텍스트 스트림 (121)), 주석이 달린 텍스트 스트림 (137), 또는 둘 다), 또는 이들의 조합을 계속해서 수신한다.

[0072] 특정 양태에서, 서버 (204) 는 또한 디바이스 (104) 에 더 가까울 수도 있고 (예를 들어, 더 적은 네트워크 홉들), (예를 들어, 디바이스 (102) 로부터 대신에) 서버 (204) 로부터 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 전송하는 것은 전체 네트워크 리소스들을 보존할 수도 있다. 특정 양태에서, 서버 (204) 는 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 디바이스 (104) 에 성공적으로 전송하는데 유용할 수 있는 네트워크 정보에 액세스할 수도 있다. 예로서, 서버 (204) 는 초기에 제 1 네트워크

링크를 통해 미디어 스트림 (109) 을 송신한다. 서버 (204) 는 네트워크 이슈들을 검출하고, 제 1 네트워크 링크가 이용가능하지 않거나 기능하지 않는다는 결정에 적어도 부분적으로 기반하여, 텍스트 송신들을 수용하기 위해 이용가능한 것으로 보이는 제 2 네트워크 링크를 사용하여 텍스트 스트림 (121), 메타데이터 스트림 (123), 또는 둘 다를 전송한다.

[0073] 도 3a 를 참조하면, GUI (145) 의 일 예가 도시된다. 특정 양태에서, GUI (145) 는 도 1 의 시스템 (100), 도 2 의 시스템 (200), 또는 둘 다에 의해 생성된다.

[0074] GUI (145) 는 비디오 디스플레이 (306), 아바타 (135), 및 트레이닝 표시자 (TI) (304) 를 포함한다. 예를 들어, GUI 생성기 (168) 는 온라인 미팅의 시작 동안 GUI (145) 를 생성한다. 비디오 스트림 (113) (예를 들어, 사용자 (142) 의 이미지들 (예를 들어, Jill Pratt)) 은 비디오 디스플레이 (306) 를 통해 디스플레이된다.

[0075] 트레이닝 표시자 (304) 는 스피치 모델 (131) 의 트레이닝 레벨 (예를 들어, 0% 또는 트레이닝되지 않음) 을 나타낸다. 예를 들어, 트레이닝 표시자 (304) 는 스피치 모델 (131) 이 커스텀 트레이닝되지 않았음을 나타낸다. 특정 양태에서, 아바타 (135) 의 표현 (예를 들어, 단색) 은 또한 트레이닝 레벨을 나타낸다. 특정 양태에서, 아바타 (135) 의 표현은 합성 스피치가 출력되고 있지 않음을 나타낸다. 예를 들어, GUI (145) 는 도 3c 를 참조하여 더 설명되는 바와 같이, 합성 스피치 표시자를 포함하지 않는다.

[0076] 특정 구현에서, 스피치 모델 (131) 의 커스텀 트레이닝 이전에 중단이 발생하고 텍스트-대-스피치 변환기 (166) 가 스피치 모델 (131) (예를 들어, 비-커스텀화된 일반 스피치 모델) 을 사용하여 합성 스피치 오디오 스트림 (133) 을 생성하는 경우, 합성 스피치 오디오 스트림 (133) 은 사용자 (142) 의 스피치 특성들과 상이할 수 있는 일반 스피치 특성들을 갖는 오디오 스피치에 대응한다. 특정 양태에서, 스피치 모델 (131) 은 사용자 (142) 의 인구통계학적 데이터와 연관된 일반 스피치 모델을 사용하여 초기화된다. 이 양태에서, 합성 스피치 오디오 스트림 (133) 은 사용자 (142) 의 인구통계학적 데이터 (예를 들어, 연령, 성별, 지역 액센트 등) 와 매칭하는 일반적인 스피치 특성들에 대응한다.

[0077] 도 3b 를 참조하면, GUI (145) 의 일 예가 도시된다. 특정 양태에서, GUI (145) 는 도 1 의 시스템 (100), 도 2 의 시스템 (200), 또는 둘 다에 의해 생성된다.

[0078] 특정 예에서, GUI 생성기 (168) 는 온라인 미팅 동안 GUI (145) 를 업데이트한다. 트레이닝 표시자 (304) 는 스피치 모델 (131) 의 제 2 트레이닝 레벨 (예를 들어, 20% 또는 부분적으로 트레이닝됨) 을 나타낸다. 예를 들어, 트레이닝 표시자 (304) 는 스피치 모델 (131) 이 커스텀 트레이닝되고 있거나 부분적으로 커스텀 트레이닝되었음을 나타낸다. 특정 양태에서, 아바타 (135) 의 표현 (예를 들어, 부분적으로 착색됨) 은 또한 제 2 트레이닝 레벨을 나타낸다. 특정 양태에서, 아바타 (135) 의 표현은 합성 스피치가 출력되고 있지 않음을 나타낸다. 예를 들어, GUI (145) 는 합성 스피치 표시자를 포함하지 않는다.

[0079] 특정 구현에서, 스피치 모델 (131) 의 부분 커스텀 트레이닝에 후속하여 중단이 발생하고 텍스트-대-스피치 변환기 (166) 가 스피치 모델 (131) (예를 들어, 부분 커스텀화된 스피치 모델) 을 사용하여 합성 스피치 오디오 스트림 (133) 을 생성하는 경우, 합성 스피치 오디오 스트림 (133) 은 사용자 (142) 의 스피치 특성들과 일부 유사함을 가진 스피치 특성들을 갖는 오디오 스피치에 대응한다.

[0080] 도 3c 를 참조하면, GUI (145) 의 일 예가 도시된다. 특정 양태에서, GUI (145) 는 도 1 의 시스템 (100), 도 2 의 시스템 (200), 또는 둘 다에 의해 생성된다.

[0081] 특정 예에서, GUI 생성기 (168) 는 중단에 응답하여 GUI (145) 를 업데이트한다. 트레이닝 표시자 (304) 는 스피치 모델 (131) 의 제 3 트레이닝 레벨 (예를 들어, 100% 또는 트레이닝 완료됨) 을 나타낸다. 예를 들어, 트레이닝 표시자 (304) 는 스피치 모델 (131) 이 커스텀 트레이닝되었거나 커스텀 트레이닝이 완료되었음을 나타낸다 (예를 들어, 임계 레벨에 도달함). 특정 양태에서, 아바타 (135) 의 표현 (예를 들어, 완전 착색됨) 은 또한 제 3 트레이닝 레벨을 나타낸다. 특정 양태에서, 아바타 (135) 의 표현은 합성 스피치가 출력되고 있음을 나타낸다. 예를 들어, GUI (145) 는 재생되고 있는 스피치가 합성 스피치임을 나타내기 위해 아바타 (135) 의 일부로서 또는 그와 함께 디스플레이되는 합성 스피치 표시자 (398) 를 포함한다.

[0082] 도 3c 의 예에서, 스피치 모델 (131) 의 커스텀 트레이닝에 후속하여 중단이 발생하고 텍스트-대-스피치 변환기 (166) 가 스피치 모델 (131) (예를 들어, 커스텀화된 스피치 모델) 을 사용하여 합성 스피치 오디오 스트림 (133) 을 생성하기 때문에, 합성 스피치 오디오 스트림 (133) 은 사용자 (142) 의 스피치 특성들과 유사한 스피

치 특성들을 갖는 오디오 스피치에 대응한다.

[0083] 중단 관리기 (164) 는 중단에 응답하여, 비디오 스트림 (113) 의 출력을 중단한다. 예를 들어, 비디오 디스플레이 (306) 는 비디오 스트림 (113) 의 출력이 중단 (예를 들어, 네트워크 문제) 으로 인해 중단되었음을 나타낸다. GUI (145) 는 텍스트 디스플레이 (396) 를 포함한다. 예를 들어, 중단 관리기 (164) 는 중단에 응답하여 텍스트 디스플레이 (396) 를 통해 텍스트 스트림 (121) 을 출력한다.

[0084] 특정 양태에서, 텍스트 스트림 (121) 은 사용자 (144) 가 대화에 계속 참여할 수 있도록 실시간으로 디스플레이 된다. 예를 들어, 사용자 (144) 는 사용자 (142) 가 말한 것을 텍스트 디스플레이 (396) 에서 읽은 후에 사용자 (142) 에게 답장을 말할 수 있다. 특정 양태에서, 네트워크 이슈들이 사용자 (144) 의 스피치에 대응하는 스피치 오디오 스트림이 디바이스 (102) 에 의해 수신되는 것을 방지하는 경우, 중단 관리기 (124) 는 사용자 (144) 의 스피치에 대응하는 텍스트 스트림을 디바이스 (102) 에서 디스플레이할 수 있다. 따라서, 온라인 미팅의 하나 이상의 참가자들은 다른 참가자들의 스피치에 대응하는 텍스트 스트림 또는 스피치 오디오 스트림을 수신할 수 있다.

[0085] 도 4a 를 참조하면, 도 1 의 시스템 (100) 또는 도 2 의 시스템 (200) 의 동작들의 예시적인 양태의 다이어그램이 도시되고 일반적으로 400 으로 지정된다. 도 4a 에 도시된 타이밍 및 동작들은 예시를 위한 것이며 제한적이지 않다. 다른 양태들에서, 추가적인 또는 더 적은 동작들이 수행될 수도 있고 타이밍은 상이할 수도 있다.

[0086] 다이어그램 (400) 은 디바이스 (102) 로부터의 미디어 스트림 (109) 의 미디어 프레임들의 송신의 타이밍을 도시한다. 특정 양태에서, 미디어 스트림 (109) 의 미디어 프레임들은 도 1 을 참조하여 설명된 바와 같이 디바이스 (102) 로부터 디바이스 (104) 로 송신된다. 대안적인 양태에서, 미디어 스트림 (109) 의 미디어 프레임들은 도 2 를 참조하여 설명된 바와 같이, 디바이스 (102) 로부터 서버 (204) 로 그리고 서버 (204) 로부터 디바이스 (102) 로 송신된다.

[0087] 디바이스 (102) 는 제 1 송신 시간에서 미디어 스트림 (109) 의 미디어 프레임 (FR) (410) 을 송신한다. 디바이스 (104) 는 제 1 수신 시간에서 미디어 프레임 (410) 을 수신하고, 제 1 재생 시간에 재생을 위해 미디어 프레임 (410) 을 제공한다. 특정 예에서, 미팅 관리기 (162) 는 제 1 수신 시간과 제 1 재생 시간 사이의 제 1 버퍼링 간격 동안 미디어 프레임 (410) 을 버퍼에 저장한다. 특정 양태에서, 미디어 프레임 (410) 은 비디오 스트림 (113) 의 제 1 부분 및 스피치 오디오 스트림 (111) 의 제 1 부분을 포함한다. 미팅 관리기 (162) 는, 제 1 재생 시간에서, 스피치 오디오 스트림 (111) 의 제 1 부분을 오디오 출력 (143) 의 제 1 부분으로서 스피커 (154) 에 출력하고 비디오 스트림 (113) 의 제 1 부분을 디스플레이 디바이스 (156) 에 출력한다.

[0088] 디바이스 (102) (또는 서버 (204)) 는 제 2 예상 송신 시간에서 미디어 프레임 (411) 을 송신할 것으로 예상된다. 디바이스 (104) 는 제 2 예상 수신 시간에서 미디어 프레임 (411) 을 수신할 것으로 예상된다. 디바이스 (104) 의 중단 관리기 (164) 는, 미디어 스트림 (109) 의 미디어 프레임들이 제 1 수신 시간의 수신 임계 지속기간 내에 수신되지 않았다는 결정에 응답하여, 스피치 오디오 스트림 (111) 에서의 중단을 검출한다. 예를 들어, 중단 관리기 (164) 는 제 1 수신 시간 및 수신 임계 지속기간에 기반하여 제 2 시간을 결정한다 (예를 들어, 제 2 시간 = 제 1 수신 시간 + 수신 임계 지속기간). 중단 관리기 (164) 는, 미디어 스트림 (109) 의 미디어 프레임들이 제 1 수신 시간과 제 2 시간 사이에 수신되지 않았다는 결정에 응답하여, 스피치 오디오 스트림 (111) 에서의 중단을 검출한다. 제 2 시간은 미디어 프레임 (411) 의 제 2 예상 수신 시간에 후속하고 미디어 프레임 (411) 의 예상 재생 시간 이전이다. 예를 들어, 제 2 시간은 미디어 프레임 (411) 의 예상 버퍼링 간격 동안이다.

[0089] 디바이스 (102) (또는 서버 (204)) 는 도 1 ~ 도 2 를 참조하여 설명된 바와 같이, 스피치 오디오 스트림 (111) 에서의 중단을 검출한다. (디바이스 (102) 또는 서버 (204) 의) 중단 관리기 (124) 는, 스피치 오디오 스트림 (111) 에서의 중단에 응답하여, 중단이 종료될 때까지 후속 미디어 프레임들 (예를 들어, 미디어 프레임들 (491) 의 세트) 에 대응하는 텍스트 스트림 (121) 을 디바이스 (104) 에 전송한다. 특정 양태에서, 미디어 프레임 (411) 은 비디오 스트림 (113) 의 제 2 부분 및 스피치 오디오 스트림 (111) 의 제 2 부분을 포함한다. 중단 관리기 (124) (또는 미팅 관리기 (122)) 는 스피치 오디오 스트림 (111) 의 제 2 부분에 대해 스피치-대-텍스트 변환을 수행함으로써 텍스트 스트림 (121) 의 텍스트 (451) 를 생성하고 텍스트 (451) 를 디바이스 (104) 에 전송한다.

[0090] 디바이스 (104) 는 도 1 ~ 도 2 를 참조하여 설명된 바와 같이, 디바이스 (102) 또는 서버 (204) 로부터 텍스트

스트림 (121) 의 텍스트 (451) 를 수신한다. 중단 관리기 (164) 는, 중단에 응답하여, 중단이 종료될 때까지 후속 미디어 프레임들에 대응하는 텍스트 스트림 (121) 의 재생을 개시한다. 예를 들어, 중단 관리기 (164) 는 제 2 재생 시간에서 텍스트 (451) 를 디스플레이 디바이스 (156) 에 제공한다. 특정 양태에서, 제 2 재생 시간은 미디어 프레임 (411) 의 예상 재생 시간에 기반한다 (예를 들어, 동일하다).

[0091] 특정 양태에서, 도 2 의 미팅 관리기 (222) 는 중단을 알지 못하고, 미디어 스트림 (109) 의 미디어 프레임 (413) 을 서버 (204) 에 송신한다. 특정 양태에서, (도 1 의 디바이스 (102) 또는 도 2 의 서버 (204) 의) 중단 관리기 (124) 는, 중단에 응답하여, 디바이스 (104) 로의 미디어 프레임 (413) 의 송신을 중단한다. 특정 양태에서, 미디어 프레임 (413) 은 비디오 스트림 (113) 의 제 3 부분 및 스피치 오디오 스트림 (111) 의 제 3 부분을 포함한다. 중단 관리기 (124) 는 스피치 오디오 스트림 (111) 의 제 3 부분에 기반하여 텍스트 (453) 를 생성한다. 중단 관리기 (124) 는 텍스트 (453) 를 디바이스 (104) 에 송신한다.

[0092] 디바이스 (104) 는 텍스트 (453) 를 수신한다. 중단 관리기 (164) 는, 중단에 응답하여, 제 3 재생 시간에 텍스트 (453) 를 디스플레이 디바이스 (156) 에 제공한다. 특정 양태에서, 제 3 재생 시간은 미디어 프레임 (413) 의 예상 재생 시간에 기반한다 (예를 들어, 동일하다).

[0093] 도 1 ~ 도 2 를 참조하여 설명되는 바와 같이, (디바이스 (102) 또는 서버 (204) 의) 중단 관리기 (124) 는, 중단 종료에 응답하여, 디바이스 (104) 로의 미디어 스트림 (109) 의 후속 미디어 프레임들 (예를 들어, 다음 미디어 프레임들 (493)) 의 송신을 재개한다. 예를 들어, 미팅 관리기 (122) 는 미디어 프레임 (415) 을 디바이스 (104) 에 송신한다. 중단 관리기 (164) 는, 중단 종료에 응답하여, 미디어 스트림 (109) 의 재생을 재개하고 텍스트 스트림 (121) 의 재생을 중지한다. 특정 양태에서, 미디어 프레임 (415) 은 비디오 스트림 (113) 의 제 4 부분 및 스피치 오디오 스트림 (111) 의 제 4 부분을 포함한다. 미팅 관리기 (162) 는, 제 4 재생 시간에서, 스피치 오디오 스트림 (111) 의 제 4 부분을 오디오 출력 (143) 의 부분으로서 스피커 (154) 에 출력하고 비디오 스트림 (113) 의 제 4 부분을 디스플레이 디바이스 (156) 에 출력한다.

[0094] 다른 예로서, 미팅 관리기 (122) 는 미디어 프레임 (417) 을 디바이스 (104) 에 송신한다. 특정 양태에서, 미디어 프레임 (417) 은 비디오 스트림 (113) 의 제 5 부분 및 스피치 오디오 스트림 (111) 의 제 5 부분을 포함한다. 미팅 관리기 (162) 는, 제 5 재생 시간에서, 스피치 오디오 스트림 (111) 의 제 5 부분을 오디오 출력 (143) 의 부분으로서 스피커 (154) 에 출력하고 비디오 스트림 (113) 의 제 5 부분을 디스플레이 디바이스 (156) 에 출력한다.

[0095] 따라서, 디바이스 (104) 는 미디어 스트림 (109) 에서의 중단 동안 텍스트 스트림 (121) 을 다시 재생함으로써 정보 손실을 방지한다. 중단이 종료될 때 미디어 스트림 (109) 의 재생이 재개된다.

[0096] 도 4b 를 참조하면, 도 1 의 시스템 (100) 또는 도 2 의 시스템 (200) 의 동작들의 예시적인 양태의 다이어그램이 도시되고 일반적으로 490 으로 지정된다. 도 4b 에 도시된 타이밍 및 동작들은 예시를 위한 것이며 제한적이지 않다. 다른 양태들에서, 추가적인 또는 더 적은 동작들이 수행될 수도 있고 타이밍은 상이할 수도 있다.

[0097] 다이어그램 (490) 은 디바이스 (102) 로부터의 미디어 스트림 (109) 의 미디어 프레임들의 송신의 타이밍을 도시한다. 도 1 의 GUI 생성기 (168) 는 아바타 (135) 의 트레이닝 레벨을 나타내는 GUI (145) 를 생성한다. 예를 들어, GUI (145) 는 아바타 (135) (예를 들어, 스피치 모델 (131)) 가 트레이닝되지 않았거나 부분적으로 트레이닝되었음을 나타낸다. 디바이스 (104) 는 비디오 스트림 (113) 의 제 1 부분 및 스피치 오디오 스트림 (111) 의 제 1 부분을 포함하는 미디어 프레임 (410) 을 수신한다. 도 4a 를 참조하여 설명되는 바와 같이, 미팅 관리기 (162) 는, 제 1 재생 시간에서, 스피치 오디오 스트림 (111) 의 제 1 부분을 오디오 출력 (143) 의 제 1 부분으로서 스피커 (154) 에 출력하고 비디오 스트림 (113) 의 제 1 부분을 디스플레이 디바이스 (156) 에 출력한다. 중단 관리기 (164) 는, 도 1 을 참조하여 설명된 바와 같이, 미디어 프레임 (410) (예를 들어, 스피치 오디오 스트림 (111) 의 제 1 부분) 에 기반하여 스피치 모델 (131) 을 트레이닝한다. GUI 생성기 (168) 는 아바타 (135) 의 업데이트된 트레이닝 레벨 (예를 들어, 부분적으로 트레이닝되거나 완전히 트레이닝됨) 을 나타내는 GUI (145) 를 업데이트한다.

[0098] 디바이스 (104) 는 도 4a 를 참조하여 설명된 바와 같이, 디바이스 (102) 또는 서버 (204) 로부터 텍스트 스트림 (121) 의 텍스트 (451) 를 수신한다. 중단 관리기 (164) 는, 중단에 응답하여, 미디어 스트림 (109) 의 재생을 중단하고, 스피치 모델 (131) 의 트레이닝을 중단하며, 합성 스피치 오디오 스트림 (133) 의 재생을 개시한다. 예를 들어, 중단 관리기 (164) 는 텍스트 (451) 에 기반하여 합성 스피치 오디오 스트림 (133) 의

합성 스피치 프레임 (471) 을 생성한다. 설명을 위해, 중단 관리기 (164) 는 텍스트 (451) 를 텍스트-대-스피치 변환기 (166) 에 제공한다. 텍스트-대-스피치 변환기 (166) 는 스피치 모델 (131) 을 사용하여 텍스트 (451) 에 대해 텍스트-대-스피치 변환을 수행하여 합성 스피치 프레임 (SFR) (471) 을 생성한다. 중단 관리기 (164) 는, 제 2 재생 시간에서, 합성 스피치 프레임 (471) 을 오디오 출력 (143) 의 제 2 부분으로서 제공한다. GUI 생성기 (168) 는 합성 스피치가 출력되고 있음을 나타내는 합성 스피치 표시자 (398) 를 포함하도록 GUI (145) 를 업데이트한다. 예를 들어, GUI (145) 는 아바타 (135) 가 말하고 있음을 표시한다.

[0099] 디바이스 (104) 는 도 4a 를 참조하여 설명된 바와 같이, 텍스트 (453) 를 수신한다. 중단 관리기 (164) 는, 중단에 응답하여, 텍스트 (453) 에 기반하여 합성 스피치 오디오 스트림 (133) 의 합성 스피치 프레임 (473) 을 생성한다. 중단 관리기 (164) 는, 제 3 재생 시간에서, 합성 스피치 프레임 (473) 을 오디오 출력 (143) 의 제 3 부분으로서 제공한다.

[0100] 도 4a 를 참조하여 설명되는 바와 같이, (디바이스 (102) 또는 서버 (204) 의) 중단 관리기 (124) 는, 중단 종료에 응답하여, 디바이스 (104) 로의 미디어 스트림 (109) 의 후속 미디어 프레임들 (예를 들어, 다음 미디어 프레임들 (493)) 의 송신을 재개한다. 예를 들어, 미팅 관리기 (122) 는 미디어 프레임 (415) 을 디바이스 (104) 에 송신한다. 중단 관리기 (164) 는, 중단 종료에 응답하여, 미디어 스트림 (109) 의 재생을 재개하고, 합성 스피치 오디오 스트림 (133) 의 재생을 중단하고, 스피치 모델 (131) 의 트레이닝을 재개한다. GUI 생성기 (168) 는 합성 스피치가 출력되지 않고 있음을 나타내는 합성 스피치 표시자 (398) 를 제거하도록 GUI (145) 를 업데이트한다.

[0101] 특정 예에서, 미팅 관리기 (162) 는 미디어 프레임 (415) 및 미디어 프레임 (417) 을 재생한다. 설명을 위해, 미디어 프레임 (415) 은 비디오 스트림 (113) 의 제 4 부분 및 스피치 오디오 스트림 (111) 의 제 4 부분을 포함한다. 미팅 관리기 (162) 는, 제 4 재생 시간에서, 스피치 오디오 스트림 (111) 의 제 4 부분을 오디오 출력 (143) 의 제 4 부분으로서 스피커 (154) 에 출력하고 비디오 스트림 (113) 의 제 4 부분을 디스플레이 디바이스 (156) 에 출력한다. 특정 양태에서, 미팅 관리기 (162) 는, 제 5 재생 시간에서, 스피치 오디오 스트림 (111) 의 제 5 부분을 오디오 출력 (143) 의 제 5 부분으로서 스피커 (154) 에 출력하고 비디오 스트림 (113) 의 제 5 부분을 디스플레이 디바이스 (156) 에 출력한다.

[0102] 따라서, 디바이스 (104) 는 미디어 스트림 (109) 에서의 중단 동안 합성 스피치 오디오 스트림 (133) 을 다시 재생함으로써 정보 손실을 방지한다. 중단이 종료될 때 미디어 스트림 (109) 의 재생이 재개된다.

[0103] 도 5 를 참조하면, 스피치 오디오 스트림 중단을 처리하도록 동작가능한 시스템이 도시되고 일반적으로 500 으로 지정된다. 특정 양태에서, 도 1 의 시스템 (100) 은 시스템 (500) 의 하나 이상의 컴포넌트들을 포함한다.

[0104] 시스템 (500) 은 네트워크 (106) 를 통해 디바이스 (104) 에 커플링된 디바이스 (502) 를 포함한다. 동작 동안, 미팅 관리기 (162) 는 다수의 디바이스들 (예를 들어, 디바이스 (102) 및 디바이스 (502)) 과의 온라인 미팅을 확립한다. 예를 들어, 미팅 관리기 (162) 는 디바이스 (102) 의 사용자 (142) 및 디바이스 (502) 의 사용자 (542) 와의 사용자 (144) 의 온라인 미팅을 확립한다. 디바이스 (104) 는, 도 1 ~ 도 2 를 참조하여 설명되는 바와 같이, 디바이스 (102) 또는 서버 (204) 로부터, 사용자 (142) 의 스피치, 이미지들, 또는 둘 다를 나타내는 미디어 스트림 (109) (예를 들어, 스피치 오디오 스트림 (111), 비디오 스트림 (113), 또는 둘 다) 을 수신한다. 유사하게, 디바이스 (104) 는 디바이스 (502) 또는 서버 (예를 들어, 서버 (204) 또는 다른 서버) 로부터, 사용자 (542) 의 스피치, 이미지들, 또는 둘 다를 나타내는 미디어 스트림 (509) (예를 들어, 제 2 스피치 오디오 스트림 (511), 제 2 비디오 스트림 (513), 또는 둘 다) 을 수신한다.

[0105] 미팅 관리기 (162) 는, 도 6a 를 참조하여 더 설명되는 바와 같이, 미디어 스트림 (509) 을 재생하는 것과 동시에 미디어 스트림 (109) 을 재생한다. 예를 들어, 미팅 관리기 (162) 는 제 2 비디오 스트림 (513) 을 디스플레이 디바이스 (156) 에 제공하는 것과 동시에 비디오 스트림 (113) 을 디스플레이 디바이스 (156) 에 제공한다. 설명을 위해, 사용자 (144) 는 온라인 미팅 동안 사용자 (542) 의 이미지들을 보는 것과 동시에 사용자 (142) 의 이미지들을 볼 수 있다. 다른 예로서, 미팅 관리기 (162) 는 스피치 오디오 스트림 (111), 제 2 스피치 오디오 스트림 (511), 또는 둘 다를 오디오 출력 (143) 으로서 스피커 (154) 에 제공한다. 설명을 위해, 사용자 (144) 는 사용자 (142) 의 스피치, 사용자 (542) 의 스피치, 또는 둘 다를 들을 수 있다. 특정 양태에서, 중단 관리기 (164) 는 도 1 을 참조하여 설명된 바와 같이, 스피치 오디오 스트림 (111) 에 기반하여 스피치 모델 (131) 을 트레이닝한다. 유사하게, 중단 관리기 (164) 는 제 2 스피치 오디오 스트림 (511) 에 기반하여 사용자 (542) 의 제 2 스피치 모델을 트레이닝한다.

- [0106] 특정 예에서, 디바이스 (104) 는 스피치 오디오 스트림 (111) 의 중단 동안 미디어 스트림 (509) 을 계속 수신한다. 중단 관리기 (164) 는, 도 6c 를 참조하여 더 설명되는 바와 같이, 합성 스피치 오디오 스트림 (133), 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 이들의 조합을 재생하는 것과 동시에 미디어 스트림 (509) 을 재생한다. 예를 들어, 중단 관리기 (164) 는 합성 스피치 오디오 스트림 (133) 을 생성하고 합성 스피치 오디오 스트림 (133) 을 스피커 (154) 에 제공하는 것과 동시에 제 2 스피치 오디오 스트림 (511) 을 제공한다. 다른 예로서, 중단 관리기 (164) 는, 텍스트 스트림 (121) 또는 주석이 달린 텍스트 스트림 (137) 을 포함하는 GUI (145) 에 대한 업데이트들을 생성하고 디스플레이 디바이스 (156) 에 GUI (145) 의 업데이트들을 제공하는 것과 동시에, 디스플레이 디바이스 (156) 에 제 2 비디오 스트림 (513) 을 제공한다. 따라서, 사용자 (144) 는 스피치 오디오 스트림 (111) 의 중단 동안 사용자 (142) 와 사용자 (542) 사이의 대화를 따를 수 있다.
- [0107] 특정 양태에서, 미디어 스트림 (509) 에서의 중단은 스피치 오디오 스트림 (111) 의 중단과 중첩된다. 중단 관리기 (164) 는 제 2 스피치 오디오 스트림 (511) 에 대응하는 제 2 텍스트 스트림, 제 2 메타데이터 스트림 또는 둘 다를 수신한다. 특정 양태에서, 중단 관리기 (164) 는 제 2 텍스트 스트림, 제 2 메타데이터 스트림, 또는 둘 다에 기반하여 제 2 주석이 달린 텍스트 스트림을 생성한다. 중단 관리기 (164) 는 제 2 텍스트 스트림, 제2 메타데이터 스트림, 제 2 주석이 달린 텍스트 스트림, 또는 이들의 조합에 기반하여 텍스트-대-스피치 변환을 수행하기 위해 제 2 스피치 모델을 사용함으로써 제 2 합성 스피치 오디오 스트림을 생성한다. 중단 관리기 (164) 는 합성 스피치 오디오 스트림 (133) 을 재생하는 것과 동시에 스피커 (154) 에 제 2 스피치 오디오 스트림 (511) 을 재생한다. 특정 양태에서, 중단 관리기 (164) 는 디스플레이 디바이스 (156) 에 대한 제 2 텍스트 스트림, 제 2 주석이 달린 텍스트 스트림, 또는 둘 다를 재생하는 것과 동시에, 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 둘 다를 재생한다. 따라서, 사용자 (144) 는 스피치 오디오 스트림 (111) 및 제 2 스피치 오디오 스트림 (511) 의 중단 동안 사용자 (142) 와 사용자 (542) 사이의 대화를 따를 수 있다.
- [0108] 따라서, 시스템 (500) 은 다수의 사용자들과의 온라인 미팅 동안 하나 이상의 스피치 오디오 스트림들 (예를 들어, 스피치 오디오 스트림 (111), 제 2 스피치 오디오 스트림 (511), 또는 둘 다) 의 중단 동안 정보 손실을 감소 (예를 들어, 제거) 시킨다. 예를 들어, 네트워크 이슈들이 하나 이상의 스피치 오디오 스트림들이 디바이스 (104) 에 의해 수신되는 것을 방지하더라도, 사용자 (144) 는 텍스트가 디바이스 (104) 에 의해 수신될 수 있는 경우에 사용자 (142) 의 스피치 및 사용자 (542) 의 스피치에 대응하는 오디오, 텍스트, 또는 이들의 조합을 계속 수신한다.
- [0109] 도 6a 를 참조하면, GUI (145) 의 일 예가 도시된다. 특정 양태에서, GUI (145) 는 도 5 의 시스템 (500) 에 의해 생성된다.
- [0110] GUI (145) 는 온라인 미팅의 다수의 참가자들에 대한 비디오 디스플레이, 아바타, 트레이닝 표시자, 또는 이들의 조합을 포함한다. 예를 들어, GUI (145) 는 도 3a 를 참조하여 설명된 바와 같이, 사용자 (142) 에 대한 비디오 디스플레이 (306), 아바타 (135), 트레이닝 표시자 (304), 또는 이들의 조합을 포함한다. GUI (145) 는 또한 사용자 (542) 에 대한 비디오 디스플레이 (606), 아바타 (635), 트레이닝 표시자 (TI) (604), 또는 이들의 조합을 포함한다. 예를 들어, GUI 생성기 (168) 는 온라인 미팅의 시작 동안 GUI (145) 를 생성한다. 미디어 스트림 (509) 의 제 2 비디오 스트림 (513) (예를 들어, 사용자 (542) (예를 들어, Emily F.) 의 이미지들) 은 비디오 디스플레이 (306) 를 통한 비디오 스트림 (113) (예를 들어, 사용자 (142) (예를 들어, Jill P.) 의 이미지들) 의 디스플레이와 동시에 비디오 디스플레이 (606) 를 통해 디스플레이된다.
- [0111] 트레이닝 표시자 (304) 는 스피치 모델 (131) 의 트레이닝 레벨 (예를 들어, 0% 또는 트레이닝되지 않음) 을 나타내고, 트레이닝 표시자 (604) 는 제 2 스피치 모델의 트레이닝 레벨 (예를 들어, 10% 또는 부분적으로 트레이닝됨) 을 나타낸다. 스피치 모델들의 트레이닝 레벨들은, 하나의 사용자가 다른 사용자보다 더 많이 말하는 경우 또는 하나의 사용자의 스피치가 더 다양한 사운드들을 포함하는 경우 (예를 들어, 모델 커버리지가 더 높음) 상이할 수 있다.
- [0112] 특정 양태에서, 아바타 (135) 의 표현 (예를 들어, 단색) 및 아바타 (635) 의 표현 (예를 들어, 부분적으로 착색됨) 은 또한 각각의 스피치 모델들의 트레이닝 레벨을 나타낸다. 특정 양태에서, 아바타 (135) 의 표현 및 아바타 (635) 의 표현은 합성 스피치가 출력되고 있지 않음을 나타낸다. 예를 들어, GUI (145) 는 임의의 합성 스피치 표시자들을 포함하지 않는다.
- [0113] 특정 구현에서, 중단이 미디어 스트림 (109) 을 수신하는데 발생하면, 텍스트-대-스피치 변환기 (166) 는 스피

치 모델 (131) (예컨대, 비-커스텀화된 일반 스피치 모델) 을 이용하여 합성 스피치 오디오 스트림 (133) 을 생성한다. 중단이 미디어 스트림 (509) 을 수신하는데 발생하면, 텍스트-대-스피치 변환기 (166) 는 제 2 스피치 모델 (예컨대, 부분적으로-커스텀화된 스피치 모델) 을 이용하여 제 2 합성 스피치 오디오 스트림을 생성한다. 특정 양태에서, 중단 관리기 (164) 는, 스피치 모델 (131) 및 제 2 스피치 모델의 트레이닝 (또는 완전한 트레이닝) 이전에 중단이 발생하는 경우, 사용자 (142) 에 대한 합성 스피치가 사용자 (542) 에 대한 합성 스피치와 구별가능하도록 스피치 모델 (131) 을 초기화하는데 사용되는 제 1 일반 스피치 모델과는 별개인 제 2 일반 스피치 모델에 기반하여 제 2 스피치 모델을 초기화한다. 특정 양태에서, 스피치 모델 (131) 은 사용자 (142) 의 데모그래픽 데이터와 연관된 제 1 일반 스피치 모델을 이용하여 초기화되고, 제 2 스피치 모델은 사용자 (542) 의 인구통계학적 데이터와 연관된 제 2 일반 스피치 모델을 이용하여 초기화된다.

[0114] 도 6b 를 참조하면, GUI (145) 의 일 예가 도시된다. 특정 양태에서, GUI (145) 는 도 5 의 시스템 (500) 에 의해 생성된다.

[0115] 특정 예에서, GUI 생성기 (168) 는 온라인 미팅 동안 GUI (145) 를 업데이트한다. 예를 들어, 트레이닝 표시자 (304) 는 스피치 모델 (131) 의 제 2 트레이닝 레벨 (예를 들어, 20% 또는 부분적으로 트레이닝됨) 및 제 2 스피치 모델의 제 2 트레이닝 레벨 (예를 들어, 100% 또는 완전히 트레이닝됨) 을 나타낼 수 있다.

[0116] 도 6c 를 참조하면, GUI (145) 의 일 예가 도시된다. 특정 양태에서, GUI (145) 는 도 5 의 시스템 (500) 에 의해 생성된다.

[0117] 특정 예에서, GUI 생성기 (168) 는 미디어 스트림 (109) 을 수신할 시 중단에 응답하여 GUI (145) 를 업데이트한다. 트레이닝 표시자 (304) 는 스피치 모델 (131) 의 제 3 트레이닝 레벨 (예를 들어, 55% 또는 부분적으로 트레이닝됨) 을 나타내고, 트레이닝 표시자 (604) 는 제 2 스피치 모델의 제 3 트레이닝 레벨 (예를 들어, 100% 또는 완전히 트레이닝됨) 을 나타낸다. 특정 양태에서, 아바타 (135) 의 표현은 합성 스피치가 출력되고 있음을 나타낸다. 예를 들어, GUI (145) 는 합성 스피치 표시자 (398) 를 포함한다. 아바타 (635) 의 표현은 합성 스피치가 사용자 (542) 에 대해 출력되고 있지 않음을 나타낸다. 예를 들어, GUI (145) 는 아바타 (635) 와 연관된 합성 스피치 표시자를 포함하지 않는다.

[0118] 중단 관리기 (164) 는 중단에 응답하여, 비디오 스트림 (113) 의 출력을 중단한다. 예를 들어, 비디오 디스플레이 (306) 는 비디오 스트림 (113) 의 출력이 중단 (예를 들어, 네트워크 문제) 으로 인해 중단되었음을 나타낸다. 중단 관리기 (164) 는 중단에 응답하여 텍스트 디스플레이 (396) 를 통해 텍스트 스트림 (121) 을 출력한다.

[0119] 특정 양태에서, 텍스트 스트림 (121) 은 사용자 (144) 가 대화에 계속 따르고 참여할 수 있도록 실시간으로 디스플레이된다. 예를 들어, 사용자 (144) 는, 사용자 (142) 가 제 1 진술을 했다는 것 (예를 들어, "당신도 축하와 유사한 무언가를 가졌으면 좋겠다") 을, 합성 스피치 오디오 스트림 (133) 으로부터 들을 수 있거나, 텍스트 디스플레이 (396) 상에서 판독할 수 있거나, 또는 둘 다일 수 있다. 사용자 (144) 는 스피커 (154) 에 의해 출력된 미디어 스트림 (509) 의 제 2 스피치 오디오 스트림에서 사용자 (542) 로부터의 답신을 들을 수 있다. 사용자 (144) 는, 사용자 (142) 가 제 2 진술을 했다는 것 (예를 들어, "재밌네요. 즐거웠습니다니 다행이에요.") 을, 합성 스피치 오디오 스트림 (133) 으로부터 들을 수 있거나, 텍스트 디스플레이 (396) 상에서 판독할 수 있거나, 또는 둘 다일 수 있다. 따라서, 사용자 (144) 는 온라인 미팅의 하나 이상의 다른 참가자들에 대한 미디어 스트림을 수신하는 동안 온라인 미팅의 하나 이상의 참가자들에 대해 합성 스피치 오디오 스트림으로부터 오디오를 들을 수 있거나, 텍스트 스트림의 텍스트를 판독하거나, 또는 둘 다일 수 있다.

[0120] 도 7a 를 참조하면, 도 5 의 시스템 (500) 의 동작들의 예시적인 양태의 다이어그램이 도시되고 일반적으로 700 으로 지정된다. 도 7a 에 도시된 타이밍 및 동작들은 설명을 위한 것이며 제한적이지 않다. 다른 양태들에서, 추가적인 또는 더 적은 동작들이 수행될 수도 있고 타이밍은 상이할 수도 있다.

[0121] 다이어그램 (700) 은 디바이스 (102) 로부터의 미디어 스트림 (109) 및 디바이스 (502) 로부터의 미디어 스트림 (509) 의 미디어 프레임들의 송신의 타이밍을 도시한다. 특정 양태에서, 미디어 스트림 (109) 의 미디어 프레임들은 도 1 ~ 도 2 를 참조하여 설명된 바와 같이 디바이스 (102) 또는 서버 (204) 로부터 디바이스 (104) 로 송신된다. 유사하게, 미디어 스트림 (509) 의 미디어 프레임들은 디바이스 (502) 또는 서버 (예를 들어, 서버 (204) 또는 다른 서버) 로부터 디바이스 (104) 로 송신된다.

[0122] 디바이스 (104) 는 미디어 스트림 (109) 의 미디어 프레임 (410) 및 미디어 스트림 (509) 의 미디어 프레임 (710) 을 수신하고, 재생을 위해 미디어 프레임 (410) 및 미디어 프레임 (710) 을 제공한다. 예를 들어, 미

팅 관리기 (162) 는 도 6a 을 참조하여 설명된 바와 같이, 스피치 오디오 스트림 (111) 의 제 1 부분 (예를 들어, 미디어 프레임 (410) 에 의해 표시됨) 및 제 2 스피치 오디오 스트림의 제 1 부분 (예를 들어, 미디어 프레임 (710) 에 의해 표시됨) 을 오디오 출력 (143) 으로서 스피커 (154) 에 출력하고, 비디오 디스플레이 (306) 를 통해 비디오 스트림 (113) 의 제 1 부분 (예를 들어, 미디어 프레임 (410) 에 의해 표시됨) 을 출력하고, 비디오 디스플레이 (606) 를 통해 제 2 비디오 스트림의 제 1 부분 (예를 들어, 미디어 프레임 (710) 에 의해 표시됨) 을 출력한다.

[0123] 디바이스 (104) 는, 도 4a 를 참조하여 설명된 바와 같이, 미디어 스트림 (109) 의 중단 동안, 텍스트 스트림 (121) 의 텍스트 (451) (미디어 프레임 (411) 에 대응함) 를 수신한다. 디바이스 (104) 는 미디어 스트림 (509) 의 미디어 프레임 (711) 을 수신한다. 중단 관리기 (164) 는, 중단에 응답하여, 중단이 미디어 스트림 (509) 의 재생과 동시에 종료될 때까지, 미디어 스트림 (109) 의 후속 미디어 프레임들에 대응하는 텍스트 스트림 (121) 의 재생을 개시한다. 예를 들어, 중단 관리기 (164) 는 재생을 위해 미디어 프레임 (711) 을 제공하는 것과 동시에 디스플레이 디바이스 (156) 에 텍스트 (451) (예를 들어, 미디어 프레임 (411) 에 의해 표시됨) 를 제공한다.

[0124] 디바이스 (104) 는, 도 4a 를 참조하여 설명된 바와 같이, 미디어 스트림 (109) 의 중단 동안, 텍스트 스트림 (121) 의 텍스트 (453) (미디어 프레임 (413) 에 대응함) 를 수신한다. 디바이스 (104) 는 미디어 스트림 (509) 의 미디어 프레임 (713) 을 수신한다. 중단 관리기 (164) 는 재생을 위해 미디어 프레임 (713) 을 제공하는 것과 동시에 텍스트 (453) 를 디스플레이 디바이스 (156) 에 제공한다.

[0125] 중단 관리기 (164) 는, 도 4a 를 참조하여 설명된 바와 같이, 중단 종료에 응답하여, 미디어 스트림 (109) 의 재생을 재개하고 텍스트 스트림 (121) 의 재생을 중지한다. 미팅 관리기 (162) 는 미디어 프레임 (415) 및 미디어 프레임 (715) 을 수신하여 다시 재생한다. 유사하게, 미팅 관리기 (162) 는 미디어 프레임 (417) 및 미디어 프레임 (717) 을 수신하여 다시 재생한다.

[0126] 따라서, 디바이스 (104) 는 미디어 스트림 (509) 의 재생과 동시에 미디어 스트림 (109) 에서의 중단 동안 텍스트 스트림 (121) 을 다시 재생함으로써 정보 손실을 방지한다. 중단이 종료될 때 미디어 스트림 (109) 의 재생이 재개된다.

[0127] 도 7b 를 참조하면, 도 5 의 시스템 (500) 의 동작들의 예시적인 양태의 다이어그램이 도시되고 일반적으로 790 으로 지정된다. 도 7b 에 도시된 타이밍 및 동작들은 예시를 위한 것이며 제한적이지 않다. 다른 양태들에서, 추가적인 또는 더 적은 동작들이 수행될 수도 있고 타이밍은 상이할 수도 있다.

[0128] 다이어그램 (790) 은 디바이스 (102) 로부터의 미디어 스트림 (109) 및 디바이스 (502) 로부터의 미디어 스트림 (509) 의 미디어 프레임들의 송신의 타이밍을 도시한다. 도 1 의 GUI 생성기 (168) 는 아바타 (135) 의 트레이닝 레벨 및 아바타 (635) 의 트레이닝 레벨을 나타내는 GUI (145) 를 생성한다. 예를 들어, GUI (145) 는 아바타 (135) (예를 들어, 스피치 모델 (131)) 가 트레이닝되지 않고 아바타 (635) (예를 들어, 제 2 스피치 모델) 가 부분적으로 트레이닝됨을 나타낸다. 디바이스 (104) 는 미디어 프레임 (410) 및 미디어 프레임 (710) 을 수신하여 다시 재생한다. 중단 관리기 (164) 는 도 4b 를 참조하여 설명된 바와 같이, 미디어 프레임 (410) 에 기반하여 스피치 모델 (131) 을 트레이닝하고, 미디어 프레임 (710) 에 기반하여 제 2 스피치 모델을 트레이닝한다. GUI 생성기 (168) 는 아바타 (135) 의 업데이트된 트레이닝 레벨 (예를 들어, 부분적으로 트레이닝됨) 및 아바타 (635) 의 업데이트된 트레이닝 레벨 (예를 들어, 완전히 트레이닝됨) 을 나타내는 GUI (145) 를 업데이트한다.

[0129] 디바이스 (104) 는 미디어 프레임 (711) 및 텍스트 스트림 (121) 의 텍스트 (451) 를 수신한다. 중단 관리기 (164) 는, 도 4b 를 참조하여 설명된 바와 같이, 텍스트 (451) 에 기반하여 합성 스피치 프레임 (471) 을 생성한다. 중단 관리기 (164) 는 합성 스피치 프레임 (471) 및 미디어 프레임 (711) 을 다시 재생한다. GUI 생성기 (168) 는 사용자 (142) 에 대해서 합성 스피치가 출력되고 있음을 나타내는 합성 스피치 표시자 (398) 를 포함하도록 GUI (145) 를 업데이트한다. 예를 들어, GUI (145) 는 아바타 (135) 가 말하고 있음을 표시한다. GUI (145) 는 사용자 (542) 에 대한 합성 스피치 표시자를 포함하지 않는다 (예를 들어, 아바타 (635) 는 말하는 것으로 표시되지 않는다).

[0130] 디바이스 (104) 는 텍스트 (453) 및 미디어 프레임 (713) 을 수신한다. 중단 관리기 (164) 는, 도 4b 를 참조하여 설명된 바와 같이, 텍스트 (453) 에 기반하여 합성 스피치 프레임 (473) 을 생성한다. 중단 관리기 (164) 는 합성 스피치 프레임 (473) 및 미디어 프레임 (417) 을 다시 재생한다.

- [0131] 중단 관리기 (164) 는, 도 4b 를 참조하여 설명되는 바와 같이, 중단 종료에 응답하여, 미디어 스트림 (109) 의 재생을 재개하고, 합성 스피치 오디오 스트림 (133) 의 재생을 중단하고, 스피치 모델 (131) 의 트레이닝을 재개한다. GUI 생성기 (168) 는 합성 스피치가 출력되지 않고 있음을 나타내는 합성 스피치 표시자 (398) 를 제거하도록 GUI (145) 를 업데이트한다.
- [0132] 특정 예에서, 미팅 관리기 (162) 는 미디어 프레임 (415) 및 미디어 프레임 (715) 을 수신하여 재생한다. 다른 예에서, 미팅 관리기 (162) 는 미디어 프레임 (417) 및 미디어 프레임 (717) 을 수신하여 재생한다.
- [0133] 따라서, 디바이스 (104) 는 미디어 스트림 (509) 의 재생과 동시에 미디어 스트림 (109) 에서의 중단 동안 합성 스피치 오디오 스트림 (133) 을 다시 재생함으로써 정보 손실을 방지한다. 중단이 종료될 때 미디어 스트림 (109) 의 재생이 재개된다.
- [0134] 도 8 을 참조하면, 스피치 오디오 스트림 중단을 처리하는 방법 (800) 의 특정 구현이 도시된다. 특정 양태에서, 방법 (800) 의 하나 이상의 동작들은 도 1 의 미팅 관리기 (162), 중단 관리기 (164), 하나 이상의 프로세서들 (160), 디바이스 (104), 시스템 (100) 또는 이들의 조합에 의해 수행된다.
- [0135] 방법 (800) 은, 802 에서, 온라인 미팅 동안, 제 1 사용자의 스피치를 나타내는 스피치 오디오 스트림을 수신하는 단계를 포함한다. 예를 들어, 도 1 의 디바이스 (104) 는, 도 1 을 참조하여 설명된 바와 같이, 온라인 미팅 동안, 사용자 (142) 의 스피치를 나타내는 스피치 오디오 스트림 (111) 을 수신한다.
- [0136] 방법 (800) 은 또한, 804 에서, 제 1 사용자의 스피치를 나타내는 텍스트 스트림을 수신하는 단계를 포함한다. 예를 들어, 도 1 의 디바이스 (104) 는, 도 1 을 참조하여 설명된 바와 같이, 사용자 (142) 의 스피치를 나타내는 텍스트 스트림 (121) 을 수신한다.
- [0137] 방법 (800) 은, 806 에서, 스피치 오디오 스트림에서의 중단에 응답하여 텍스트 스트림에 기반하여 출력을 선택적으로 생성하는 단계를 더 포함한다. 예를 들어, 도 1 의 중단 관리기 (164) 는, 도 1 을 참조하여 설명된 바와 같이, 스피치 오디오 스트림 (111) 에서의 중단에 응답하여 텍스트 스트림 (121) 에 기반하여 합성된 스피치 오디오 스트림 (133) 을 선택적으로 생성한다. 특정 구현에서, 중단 관리기 (164) 는, 도 1 을 참조하여 설명된 바와 같이, 스피치 오디오 스트림 (111) 에서의 중단에 응답하여, 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 둘 다를 선택적으로 출력한다.
- [0138] 따라서, 방법 (800) 은 온라인 미팅 동안 스피치 오디오 스트림 (111) 의 중단 동안 정보 손실을 감소 (예를 들어, 제거) 한다. 예를 들어, 사용자 (144) 는, 네트워크 이슈들이 스피치 오디오 스트림 (111) 이 디바이스 (104) 에 의해 수신되는 것을 방지하더라도 텍스트가 디바이스 (104) 에 의해 수신될 수 있는 경우들에서, 사용자 (142) 의 스피치에 대응하는 오디오 (예를 들어, 합성 스피치 오디오 스트림 (133)), 텍스트 (예를 들어, 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), 또는 둘 다), 또는 이들의 조합을 계속해서 수신한다.
- [0139] 도 8 의 방법 (800) 은 필드 프로그래밍가능 게이트 어레이 (FPGA) 디바이스, 주문형 집적 회로 (ASIC), 프로세싱 유닛, 이를 테면 중앙 프로세싱 유닛 (CPU), DSP, 제어기, 다른 하드웨어 디바이스, 펌웨어 디바이스, 또는 이들의 임의의 조합에 의해 구현될 수도 있다. 일 예로서, 도 8 의 방법 (800) 은, 도 18 을 참조하여 설명된 바와 같은, 명령들을 실행하는 프로세서에 의해 수행될 수도 있다.
- [0140] 도 9 는 하나 이상의 프로세서들 (160) 을 포함하는 집적 회로 (902) 로서의 디바이스 (104) 의 구현 (900) 을 도시한다. 집적 회로 (902) 는 또한 입력 데이터 (928) (예를 들어, 스피치 오디오 스트림 (111), 비디오 스트림 (113), 미디어 스트림 (109), 중단 통지 (119), 텍스트 스트림 (121), 메타데이터 스트림 (123), 미디어 스트림 (509), 또는 이들의 조합) 가 처리를 위해 수신될 수 있게 하는 입력 (904) (예를 들어, 하나 이상의 버스 인터페이스들) 을 포함한다. 집적 회로 (902) 는 또한 출력 신호 (예를 들어, 스피치 오디오 스트림 (111), 합성 스피치 오디오 스트림 (133), 오디오 출력 (143), 비디오 스트림 (113), 텍스트 스트림 (121), 주석이 달린 텍스트 스트림 (137), GUI (145), 또는 이들의 조합) 의 전송을 가능하게 하는 출력 (906) (예를 들어, 버스 인터페이스) 을 포함한다. 집적 회로 (902) 는 도 10 에 도시된 바와 같은 모바일 폰 또는 태블릿, 도 11 에 도시된 바와 같은 헤드셋, 도 12 에 도시된 바와 같은 웨어러블 전자 디바이스, 도 13 에 도시된 바와 같은 음성 제어 스피커 시스템, 도 14 에 도시된 바와 같은 카메라, 도 15 에 도시된 바와 같은 가상 현실 헤드셋 또는 증강 현실 헤드셋, 또는 도 16 또는 도 17 에 도시된 바와 같은 차량과 같은 시스템에서의 컴포넌트로서 스피치 오디오 스트림 중단을 처리하는 구현을 가능하게 한다.

- [0141] 도 10 은, 디바이스 (104) 가 예시적 비한정적인 예들로서 전화기 또는 태블릿과 같은 모바일 디바이스 (1002) 를 포함하는 구현 (1000) 을 도시한다. 모바일 디바이스 (1002) 는 마이크로폰 (1010), 스피커 (154), 및 디스플레이 스크린 (1004) 을 포함한다. 미팅 관리기 (162), 중단 관리기 (164), GUI 생성기 (168) 또는 이들의 조합을 포함하는 하나 이상의 프로세서들 (160) 의 컴포넌트들은, 모바일 디바이스 (1002) 에 통합되고, 모바일 디바이스 (1002) 의 사용자에게 일반적으로 보이지 않는 내부 컴포넌트들을 표시하기 위해 점선들을 사용하여 도시된다. 특정 예에서, 미팅 관리기 (162) 는 스피치 오디오 스트림 (111) 을 출력하거나, 중단 관리기 (164) 는 합성 스피치 오디오 스트림 (133) 을 출력하고, 이는 그 후 모바일 디바이스 (1002) 에서 하나 이상의 동작을 수행하도록, 예컨대 그래픽 사용자 인터페이스를 시작하거나 또는 그렇지 않으면 디스플레이 스크린 (1004) 에서 사용자의 스피치와 연관된 다른 정보를 디스플레이하도록 (예를 들어, 통합된 "스마트 어시스턴트" 애플리케이션을 통해) 프로세싱된다.
- [0142] 도 11 은, 디바이스 (104) 가 헤드셋 디바이스 (1102) 를 포함하는 구현 (1100) 을 도시한다. 헤드셋 디바이스 (1102) 는 스피커 (154), 마이크로폰 (1110), 또는 둘 다를 포함한다. 미팅 관리기 (162), 중단 관리기 (164) 또는 둘 다를 포함하는 하나 이상의 프로세서들 (160) 의 컴포넌트들은 헤드셋 디바이스 (1102) 에 통합된다. 특정 예에서, 미팅 관리기 (162) 는 스피치 오디오 스트림 (111) 을 출력하거나, 중단 관리자 (164) 는 합성 스피치 오디오 스트림 (133) 을 출력하며, 이는 헤드셋 디바이스 (1102) 로 하여금, 추가 처리를 위해 사용자 스피치에 대응하는 오디오 데이터를 제 2 디바이스 (도시되지 않음) 에 송신하기 위해 헤드셋 디바이스 (1102) 에서 하나 이상의 동작들을 수행하게 할 수 있다.
- [0143] 도 12 는, 디바이스 (104) 가 "스마트 워치" 로서 도시된 웨어러블 전자 디바이스 (1202) 를 포함하는 구현 (1200) 을 도시한다. 미팅 관리기 (162), 중단 관리기 (164), GUI 관리기 (168), 스피커 (154), 마이크로폰 (1210), 또는 이들의 조합은 웨어러블 전자 디바이스 (1202) 에 통합된다. 특정 예에서, 미팅 관리기 (162) 는 스피치 오디오 스트림 (111) 을 출력하거나, 중단 관리기 (164) 는 합성 스피치 오디오 스트림 (133) 을 출력하고, 이는 그 후 웨어러블 전자 디바이스 (1202) 에서 하나 이상의 동작을 수행하도록, 예컨대 GUI (145) 를 시작하거나 또는 그렇지 않으면 웨어러블 전자 디바이스 (1202) 의 디스플레이 스크린 (1204) 에서 사용자의 스피치와 연관된 다른 정보를 디스플레이하도록 프로세싱된다. 설명을 위해, 웨어러블 전자 디바이스 (1202) 는 웨어러블 전자 디바이스 (1202) 에 의해 검출된 사용자 음성에 기반하여 통지를 디스플레이하도록 구성된 디스플레이 스크린을 포함할 수도 있다. 특정 예에서, 웨어러블 전자 디바이스 (1202) 는 사용자 스피치의 검출에 응답하여 햅틱 통지를 제공하는 (예를 들어, 진동하는) 햅틱 디바이스를 포함한다. 예를 들어, 햅틱 통지는 사용자로 하여금 웨어러블 전자 디바이스 (1202) 를 보게 하여, 사용자가 말한 키워드의 검출을 표시하는 디스플레이된 통지를 보게 할 수 있다. 따라서, 웨어러블 전자 디바이스 (1202) 는 청각 장애를 가진 사용자 또는 헤드셋을 착용한 사용자에게, 사용자의 스피치가 검출됨을 알릴 수 있다.
- [0144] 도 13 은, 디바이스 (104) 가 무선 스피커 및 음성 활성화 디바이스 (1302) 를 포함하는 구현 (1300) 이다. 무선 스피커 및 음성 활성화 디바이스 (1302) 는 무선 네트워크 연결성을 가질 수 있고 어시스턴트 동작을 실행하도록 구성된다. 미팅 관리기 (162), 중단 관리기 (164), 또는 둘 다를 포함하는 하나 이상의 프로세서들 (160), 스피커 (154), 마이크로폰 (1310), 또는 이들의 조합은 무선 스피커 및 음성 활성화 디바이스 (1302) 에 포함된다. 동작 동안, 미팅 관리기 (162) 에 의해 출력된 스피치 오디오 스트림 (111) 에서 또는 중단 관리기 (164) 에 의해 출력된 합성 스피치 오디오 스트림 (133) 에서 사용자 스피치로서 식별된 구두 명령을 수신하는 것에 응답하여, 무선 스피커 및 음성 활성화 디바이스 (1302) 는, 예컨대 음성 활성화 시스템 (예를 들어, 통합된 어시스턴트 애플리케이션) 의 실행을 통해 어시스턴트 동작들을 실행할 수 있다. 어시스턴트 동작들은 캘린더 이벤트를 형성하는 것, 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.
- [0145] 도 14 는 디바이스 (104) 가 카메라 디바이스 (1402) 에 대응하는 포터블 전자 디바이스인 구현 (1400) 을 도시한다. 미팅 관리기 (162), 중단 관리기 (164), GUI 관리기 (168), 스피커 (154), 마이크로폰 (1410), 또는 이들의 조합은 카메라 디바이스 (1402) 에 포함된다. 동작 동안, 미팅 관리기 (162) 에 의해 출력된 스피치 오디오 스트림 (111) 에서 또는 중단 관리기 (164) 에 의해 출력된 합성 스피치 오디오 스트림 (133) 에서 사용자 스피치로서 식별된 구두 명령을 수신하는 것에 응답하여, 카메라 디바이스 (1402) 는 예시적인 예로서, 이미지 또는 비디오 캡처 설정들, 이미지 또는 비디오 재생 설정들, 또는 이미지 또는 비디오 캡처 명령들을 조정하는 것과 같이, 음성 사용자 명령들에 응답하여 동작들을 실행할 수 있다.
- [0146] 도 15 는 디바이스 (104) 가 가상 현실, 증강 현실, 또는 혼합 현실 헤드셋 (1502) 에 대응하는 휴대용 전자 디

바이스를 포함하는 구현 (1500) 을 도시한다. 미팅 관리기 (162), 중단 관리기 (164), GUI 관리기 (168), 스피커 (154), 마이크로폰 (1510), 또는 이들의 조합은 헤드셋 (1502) 에 통합된다. 사용자 스피치 검출은 미팅 관리기 (162) 에 의해 출력된 스피치 오디오 스트림 (111) 또는 중단 관리기 (164) 에 의해 출력된 합성 스피치 오디오 스트림 (133) 에 기반하여 수행될 수 있다. 시각적 인터페이스 디바이스는 헤드셋 (1502) 이 착용된 동안 사용자에게 증강 현실 또는 가상 현실 이미지들 또는 장면들의 디스플레이를 가능하게 하기 위해 사용자의 눈들 앞에 포지셔닝된다. 특정 예에서, 시각적 인터페이스 디바이스는, 오디오 스트림에서 검출된 사용자 스피치를 표시하는 통지를 디스플레이하도록 구성된다. 다른 예에서, 시각적 인터페이스 디바이스는 GUI (145) 를 디스플레이하도록 구성된다.

[0147] 도 16 은 디바이스 (104) 가 유인 또는 무인 항공 디바이스 (예를 들어, 수화물 배송 드론) 으로서 예시되는 비히클 (1602) 에 대응하거나 비히클 내에 통합되는 구현 (1600) 을 도시한다. 미팅 관리기 (162), 중단 관리기 (164), GUI 생성기 (168), 스피커 (154), 마이크로폰 (1610), 또는 이들의 조합은 차량 (1602) 에 통합된다. 사용자 스피치 검출은, 예컨대 차량 (1602) 의 인가된 사용자로부터의 전달 명령들을 위해, 미팅 관리기 (162) 에 의해 출력된 스피치 오디오 스트림 (111) 또는 중단 관리기 (164) 에 의해 출력된 합성 스피치 오디오 스트림 (133) 에 기반하여 수행될 수 있다.

[0148] 도 17 은 디바이스 (104) 가 자동차로서 예시된 비히클 (1702) 에 대응하거나 비히클 내에 통합되는 다른 구현 (1700) 을 도시한다. 차량 (1702) 은 미팅 관리기 (162), 중단 관리기 (164), GUI 생성기 (168) 또는 이들의 조합을 포함하는 하나 이상의 프로세서들 (160) 을 포함한다. 차량 (1702) 은 또한 스피커 (154), 마이크로폰 (1710), 또는 둘 다를 포함한다. 사용자 스피치 검출은 미팅 관리기 (162) 에 의해 출력된 스피치 오디오 스트림 (111) 또는 중단 관리기 (164) 에 의해 출력된 합성 스피치 오디오 스트림 (133) 에 기반하여 수행될 수 있다. 예를 들어, 사용자 스피치 검출은 차량 (1702) 의 인가된 사용자로부터 음성 명령을 검출하는데 (예를 들어, 엔진 또는 난방을 시동하기 위해) 사용될 수 있다. 특정 구현에서, 미팅 관리기 (162) 에 의해 출력된 스피치 오디오 스트림 (111) 에서 또는 중단 관리기 (164) 에 의해 출력된 합성 스피치 오디오 스트림 (133) 에서 사용자 스피치로서 식별된 구두 명령을 수신하는 것에 응답하여, 차량 (1702) 의 음성 활성화 시스템은, 예컨대 디스플레이 (1720) 또는 하나 이상의 스피커들 (예를 들어, 스피커 (154)) 을 통해 피드백 또는 정보를 제공함으로써, 스피치 오디오 스트림 (111) 또는 합성 스피치 오디오 스트림 (133) 에서 검출된 하나 이상의 키워드들 (예를 들어, "잠금해제", "엔진 시작", "음악 재생", "날씨 예보 표시", 또는 다른 음성 명령) 에 기반하여 차량 (1702) 의 하나 이상의 동작들을 개시한다. 특정 구현에서, GUI 생성기 (168) 는 온라인 미팅 (예를 들어, 통화) 에 관한 정보를 디스플레이 (1720) 에 제공한다. 예를 들어, GUI 생성기 (168) 는 GUI (145) 를 디스플레이 (1720) 에 제공한다.

[0149] 도 18 을 참조하면, 디바이스의 특정 예시적인 구현의 블록 다이어그램이 도시되고 일반적으로 1800 으로 지칭된다. 다양한 구현들에 있어서, 디바이스 (1800) 는 도 18 에 예시된 것들보다 더 많거나 더 적은 컴포넌트들을 가질 수도 있다. 예시적인 구현에서, 디바이스 (1800) 는 디바이스 (104) 에 대응할 수도 있다. 예시적인 구현에 있어서, 디바이스 (1800) 는 도 1 내지 도 17 을 참조하여 설명된 하나 이상의 동작들을 수행할 수도 있다.

[0150] 특정 구현에 있어서, 디바이스 (1800) 는 프로세서 (1806) (예컨대, 중앙 프로세싱 유닛 (CPU)) 를 포함한다. 디바이스 (1800) 는 하나 이상의 추가적인 프로세서들 (1810) (예컨대, 하나 이상의 DSP들) 을 포함할 수도 있다. 특정 양태에서, 도 1 의 하나 이상의 프로세서들 (160) 은 프로세서 (1806), 프로세서들 (1810), 또는 이들의 조합에 대응한다. 프로세서들 (1810) 은 음성 코더 ("보코더") 인코더 (1836), 보코더 디코더 (1838), 미팅 관리기 (162), 중단 관리기 (164), GUI 생성기 (168), 또는 이들의 조합을 포함하는 스피치 및 음악 코더-디코더 (CODEC) (1808) 를 포함할 수도 있다. 특정 양태에서, 도 1 의 하나 이상의 프로세서들 (160) 은 프로세서 (1806), 프로세서들 (1810), 또는 이들의 조합을 포함한다.

[0151] 디바이스 (1800) 는 메모리 (1886) 및 CODEC (1834) 을 포함할 수도 있다. 메모리 (1886) 는 미팅 관리기 (162), 중단 관리기 (164), GUI 생성기 (168), 또는 이들의 조합을 참조하여 설명된 기능성을 구현하기 위해 하나 이상의 추가적인 프로세서들 (1810) (또는 프로세서 (1806)) 에 의해 실행가능한 명령들 (1856) 을 포함할 수도 있다. 특정 양태에서, 메모리 (1886) 는 미팅 관리기 (162), 중단 관리기 (164), GUI 생성기 (168), 또는 이들의 조합에 의해 사용되거나 생성된 프로그램 데이터 (1858) 를 저장한다. 특정 양태에서, 메모리 (1886) 은 도 1 의 메모리 (132) 에 대응한다. 디바이스 (1800) 는 트랜시버 (1850) 를 통해, 안테나 (1842) 에 결합된 모뎀 (1840) 을 포함할 수도 있다.

- [0152] 디바이스 (1800) 는 디스플레이 제어기 (1826) 에 커플링된 디스플레이 디바이스 (156) 를 포함할 수도 있다. 스피커들 (154) 및 하나 이상의 마이크로폰들 (1832) 이 CODEC (1834) 에 커플링될 수도 있다. CODEC (1834) 은 디지털-대-아날로그 변환기 (DAC) (1802), 아날로그-대-디지털 변환기 (ADC) (1804), 또는 양자 모 두를 포함할 수도 있다. 특정 구현에서, CODEC (1834) 은 하나 이상의 마이크로폰들 (1832) 로부터 아날로그 신호들을 수신하고, 아날로그 신호들을 아날로그-대-디지털 변환기 (1804) 를 사용하여 디지털 신호들로 변환하고, 디지털 신호들을 스피치 및 음악 코덱 (1808) 에 제공할 수도 있다. 스피치 및 음악 코덱 (1808) 은 디지털 신호들을 프로세싱할 수도 있고, 디지털 신호들은 미팅 관리기 (162), 중단 관리기 (164) 또는 둘 다 에 의해 추가로 프로세싱될 수도 있다. 특정 구현에서, 스피치 및 음악 코덱 (1808) 은 CODEC (1834) 에 디지털 신호들을 제공할 수도 있다. 코덱 (1834) 은 디지털 신호들을 디지털-대-아날로그 변환기 (1802) 를 사용하여 아날로그 신호들로 변환할 수도 있고 아날로그 신호들을 스피커 (154) 에 제공할 수도 있다.
- [0153] 특정 구현에서, 디바이스 (1800) 는 시스템-인-패키지 또는 시스템-온-칩 디바이스 (1822) 에 포함될 수도 있다. 특정 구현에서, 메모리 (1886), 프로세서 (1806), 프로세서들 (1810), 디스플레이 제어기 (1826), CODEC (1834), 모뎀 (1840), 및 트랜시버 (1850) 는 시스템-인-패키지 또는 시스템-온-칩 디바이스 (1822) 에 포함된다. 특정의 구현예에서, 입력 디바이스 (1830) 및 전원 (1844) 은 시스템-온-칩 디바이스 (1822) 에 결합된다. 더욱이, 특정 구현에서, 도 18 에 도시된 바와 같이, 디스플레이 디바이스 (156), 입력 디바이스 (1830), 스피커 (154), 하나 이상의 마이크로폰들 (1832), 안테나 (1842), 및 전원 (1844) 은 시스템-온-칩 디바이스 (1822) 외부에 있다. 특정 구현에서, 디스플레이 디바이스 (156), 입력 디바이스 (1830), 스피커 (154), 하나 이상의 마이크로폰들 (1832), 안테나 (1842), 및 전원 (1844) 의 각각은, 인터페이스 또는 제어기와 같은 시스템-온-칩 디바이스 (1822) 의 컴포넌트에 결합될 수도 있다.
- [0154] 디바이스 (1800) 는 가상 어시스턴트, 가전 제품, 스마트 디바이스, 사물 인터넷 (IoT) 디바이스, 통신 디바이스, 헤드셋, 차량, 컴퓨터, 디스플레이 디바이스, 텔레비전, 게이밍 콘솔, 뮤직 플레이어, 라디오, 비디오 플레이어, 엔터테인먼트 유닛, 개인용 미디어 플레이어, 디지털 비디오 플레이어, 카메라, 네비게이션 디바이스, 스마트 스피커, 스피커 파, 모바일 통신 디바이스, 스마트폰, 셀룰러 폰, 랩톱 컴퓨터, 태블릿, 개인용 디지털 어시스턴트, 디지털 비디오 디스크 (DVD) 플레이어, 튜너, 증강 현실 헤드셋, 가상 현실 헤드셋, 항공 비하클, 홈 오토메이션 시스템, 음성-활성화 디바이스, 무선 스피커 및 음성 활성화 디바이스, 휴대용 전자 디바이스, 자동차, 컴퓨팅 디바이스, 가상 현실 (VR) 디바이스, 기지국, 모바일 디바이스, 또는 이들의 임의의 조합을 포함할 수도 있다.
- [0155] 개시된 구현들과 함께, 장치는 온라인 미팅 동안, 제 1 사용자의 스피치를 나타내는 스피치 오디오 스트림을 수신하기 위한 수단을 포함한다. 예를 들어, 스피치 오디오 스트림을 수신하기 위한 수단은 도 1 의 미팅 관리기 (162), 중단 관리기 (164), 하나 이상의 프로세서들 (160), 디바이스 (104), 시스템 (100), 도 2 의 미팅 관리기 (122), 서버 (204), 시스템 (200), 하나 이상의 프로세서들 (1810), 프로세서 (1806), 스피치 및 음악 코덱 (1808), 모뎀 (1840), 트랜시버 (1850), 안테나 (1842), 디바이스 (1800), 온라인 미팅 동안 스피치 오디오 스트림을 수신하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합에 대응할 수 있다.
- [0156] 장치는 또한 제 1 사용자의 스피치를 나타내는 텍스트 스트림을 수신하기 위한 수단을 포함한다. 예를 들어, 텍스트 스트림을 수신하기 위한 수단은 도 1 의 미팅 관리기 (162), 중단 관리기 (164), 텍스트-대-스피치 변환기 (166), 하나 이상의 프로세서들 (160), 디바이스 (104), 시스템 (100), 도 2 의 미팅 관리기 (122), 중단 관리기 (124), 서버 (204), 시스템 (200), 하나 이상의 프로세서들 (1810), 프로세서 (1806), 스피치 및 음악 코덱 (1808), 모뎀 (1840), 트랜시버 (1850), 안테나 (1842), 디바이스 (1800), 텍스트 스트림을 수신하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합에 대응할 수 있다.
- [0157] 장치는, 스피치 오디오 스트림에서의 중단에 응답하여 텍스트 스트림에 기반하여 출력을 선택적으로 생성하기 위한 수단을 더 포함한다. 예를 들어, 출력을 선택적으로 생성하기 위한 수단은 도 1 의 중단 관리기 (164), 텍스트-대-스피치 변환기 (166), GUI 생성기 (168), 하나 이상의 프로세서들 (160), 디바이스 (104), 시스템 (100), 도 2 의 중단 관리기 (124), 서버 (204), 시스템 (200), 하나 이상의 프로세서들 (1810), 프로세서 (1806), 스피치 및 음악 코덱 (1808), 디바이스 (1800), 출력을 선택적으로 생성하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합에 대응할 수 있다.
- [0158] 일부 구현들에서, 비밀식적 컴퓨터 관독가능 매체 (예를 들어, 메모리 (1886) 와 같은 컴퓨터 관독가능 저장 디바이스) 는, 하나 이상의 프로세서들 (예를 들어, 하나 이상의 프로세서들 (1810) 또는 프로세서 (1806)) 에 의

해 실행될 때, 하나 이상의 프로세서들로 하여금, 온라인 미팅 동안, 제 1 사용자 (예를 들어 사용자 (142))의 스피치를 나타내는 스피치 오디오 스트림 (예를 들어, 스피치 오디오 스트림 (111))을 수신하도록 하는 명령들 (예를 들어, 명령어들 (1856))을 포함한다. 명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 또한 하나 이상의 프로세서들로 하여금, 제 1 사용자 (예를 들어, 사용자 (142))의 스피치를 나타내는 텍스트 스트림 (예를 들어, 텍스트 스트림 (121))을 수신하게 한다. 명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 추가로 하나 이상의 프로세서들로 하여금, 스피치 오디오 스트림에서의 중단에 응답하여 텍스트 스트림에 기반하여 출력 (예컨대, 합성 스피치 오디오 스트림 (133), 주석이 달린 텍스트 스트림 (137), 또는 둘 다)을 선택적으로 생성하게 한다.

- [0159] 본 개시의 특정 양태들은 상호관련된 조항들의 제 1 세트로 하기에서 기술된다:
- [0160] 조항 1에 따르면, 통신을 위한 디바이스로서, 온라인 미팅 동안, 제 1 사용자의 스피치를 나타내는 스피치 오디오 스트림을 수신하고; 제 1 사용자의 스피치를 나타내는 텍스트 스트림을 수신하고; 그리고 스피치 오디오 스트림에서의 중단에 응답하여 텍스트 스트림에 기반하여 출력을 선택적으로 생성하도록 구성된 하나 이상의 프로세서들을 포함한다.
- [0161] 조항 2는 조항 1의 디바이스를 포함하고, 하나 이상의 프로세서들은 스피치 오디오 스트림의 마지막 수신된 오디오 프레임의 임계 지속기간 내에 스피치 오디오 스트림의 오디오 프레임들이 수신되지 않는다는 결정에 응답하여 중단을 검출하도록 구성된다.
- [0162] 조항 3은 조항 1의 디바이스를 포함하고, 하나 이상의 프로세서들은 텍스트 스트림을 수신하는 것에 응답하여 중단을 검출하도록 구성된다.
- [0163] 조항 4는 조항 1의 디바이스를 포함하고, 하나 이상의 프로세서들은 중단 통지를 수신하는 것에 응답하여 중단을 검출하도록 구성된다.
- [0164] 조항 5는 조항 1 내지 조항 4 중 임의의 디바이스를 포함하고, 하나 이상의 프로세서들은 텍스트 스트림을 디스플레이에 대한 출력으로서 제공하도록 구성된다.
- [0165] 조항 6은 조항 1 내지 조항 5 중 임의의 디바이스를 포함하고, 하나 이상의 프로세서들은 제 1 사용자의 스피치의 억양을 나타내는 메타데이터 스트림을 수신하고 그리고 메타데이터 스트림에 기반하여 텍스트 스트림에 주석을 달도록 추가로 구성된다.
- [0166] 조항 7은 조항 1 내지 조항 6 중 임의의 디바이스를 포함하고, 하나 이상의 프로세서들은 합성 스피치 오디오 스트림을 생성하기 위해 텍스트 스트림에 대해 텍스트-대-스피치 변환을 수행하고 그리고 합성 스피치 오디오 스트림을 스피커에 대한 출력으로서 제공하도록 추가로 구성된다.
- [0167] 조항 8은 조항 7의 디바이스를 포함하고, 하나 이상의 프로세서들은 제 1 사용자의 스피치의 억양을 나타내는 메타데이터 스트림을 수신하도록 추가로 구성되고, 텍스트-대-스피치 변환은 메타데이터 스트림에 기반한다.
- [0168] 조항 9는 조항 7의 디바이스를 포함하고, 하나 이상의 프로세서들은 합성 스피치 오디오 스트림을 스피커에 제공하는 것과 동시에 아바타를 디스플레이하도록 추가로 구성된다.
- [0169] 조항 10은 조항 9의 디바이스를 포함하고, 하나 이상의 프로세서들은 온라인 미팅 동안 미디어 스트림을 수신하도록 구성되고, 미디어 스트림은 제 1 사용자의 스피치 오디오 스트림 및 비디오 스트림을 포함한다.
- [0170] 조항 11은 조항 10의 디바이스를 포함하고, 하나 이상의 프로세서들은, 중단에 응답하여, 스피치 오디오 스트림의 재생을 중단시키고 비디오 스트림의 재생을 중단하도록 구성된다.
- [0171] 조항 12는 조항 10의 디바이스를 포함하고, 하나 이상의 프로세서들은, 중단 종료에 응답하여, 합성 스피치 오디오 스트림을 스피커에 제공하는 것을 억제하고 아바타를 디스플레이하는 것을 억제하고 비디오 스트림의 재생을 재개하고 스피치 오디오 스트림의 재생을 재개하도록 구성된다.
- [0172] 조항 13은 조항 7의 디바이스를 포함하고, 텍스트-대-스피치 변환은 스피치 모델에 기반하여 수행된다.
- [0173] 조항 14는 조항 13의 디바이스를 포함하고, 스피치 모델은 일반 스피치 모델에 대응한다.
- [0174] 조항 15는 조항 13 또는 조항 14의 디바이스를 포함하고, 하나 이상의 프로세서들은, 중단 이전에, 스피치 오디오 스트림에 기반하여 스피치 모델을 업데이트하도록 구성된다.
- [0175] 조항 16은 조항 1 내지 조항 15 중 임의의 디바이스를 포함하고, 하나 이상의 프로세서들은, 온라인 미팅

동안, 제 2 사용자의 스피치를 나타내는 제 2 스피치 오디오 스트림을 수신하고 출력을 생성하는 것과 동시에 제 2 스피치 오디오 스트림을 스피커에 제공하도록 구성된다.

[0176] 조항 17 은 조항 1 내지 조항 16 중 임의의 디바이스를 포함하고, 하나 이상의 프로세서들은, 스피치 오디오 스트림에서의 중단에 응답하여 스피치 오디오 스트림의 재생을 중지하고 중단 종료에 응답하여, 텍스트 스트림에 기반하여 출력을 생성하는 것을 억제하고, 스피치 오디오 스트림의 재생을 재개하도록 구성된다.

[0177] 본 개시의 특정 양태들은 상호관련된 조항들의 제 2 세트로 하기에서 기술된다:

[0178] 조항 18 에 따르면, 통신 방법으로서, 온라인 미팅 동안 디바이스에서, 제 1 사용자의 스피치를 나타내는 스피치 오디오 스트림을 수신하는 단계, 디바이스에서, 제 1 사용자의 스피치를 나타내는 텍스트 스트림을 수신하는 단계, 및 스피치 오디오 스트림에서의 중단에 응답하여 텍스트 스트림에 기반하여 출력을 선택적으로 생성하는 단계를 포함한다.

[0179] 조항 19 는 조항 18 의 방법을 포함하고, 스피치 오디오 스트림의 마지막 수신된 오디오 프레임의 임계 지속기간 내에 스피치 오디오 스트림의 오디오 프레임들이 수신되지 않는다는 결정에 응답하여 중단을 검출하는 단계를 더 포함한다.

[0180] 조항 20 은 조항 18 의 방법을 포함하고, 텍스트 스트림을 수신하는 것에 응답하여 중단을 검출하는 단계를 더 포함한다.

[0181] 조항 21 은 조항 18 의 방법을 포함하고, 중단 통지를 수신하는 것에 응답하여 중단을 검출하는 단계를 더 포함한다.

[0182] 조항 22 는 조항 18 내지 조항 21 의 임의의 방법을 포함하고, 디스플레이에 대한 출력으로서 텍스트 스트림을 제공하는 단계를 더 포함한다.

[0183] 조항 23 은 조항 18 내지 조항 22 중 임의의 방법을 포함하고, 제 1 사용자의 스피치의 역양을 나타내는 메타데이터 스트림을 수신하는 단계, 및 메타데이터 스트림에 기반하여 텍스트 스트림에 주석을 다는 단계를 더 포함한다.

[0184] 본 개시의 특정 양태들은 상호관련된 조항들의 제 3 세트로 하기에서 기술된다:

[0185] 조항 24 에 따르면, 비일시적 컴퓨터 판독가능 저장 매체는 명령들을 저장하고, 그 명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 하나 이상의 프로세서들로 하여금, 온라인 미팅 동안, 제 1 사용자의 스피치를 나타내는 스피치 오디오 스트림을 수신하게 하고; 제 1 사용자의 스피치를 나타내는 텍스트 스트림을 수신하게 하고; 그리고 스피치 오디오 스트림에서의 중단에 응답하여 텍스트 스트림에 기반하여 출력을 선택적으로 생성하게 한다.

[0186] 조항 25 는 조항 24 의 비일시적 컴퓨터 판독가능 저장 매체를 포함하고, 명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 하나 이상의 프로세서들로 하여금, 합성 스피치 오디오 스트림을 생성하기 위해 텍스트 스트림에 대해 텍스트-대-스피치 변환을 수행하게 하고, 그리고 스피커에 대한 출력으로서 합성 스피치 오디오 스트림을 제공하게 한다.

[0187] 조항 26 은 조항 25 의 비일시적 컴퓨터 판독가능 저장 매체를 포함하고, 명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 하나 이상의 프로세서들로 하여금, 제 1 사용자의 음성의 역양을 나타내는 메타데이터 스트림을 수신하게 하고, 텍스트-대-스피치 변환은 메타데이터 스트림에 기반한다.

[0188] 조항 27 은 조항 25 또는 조항 26 의 비일시적 컴퓨터 판독가능 저장 매체를 포함하고, 명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 하나 이상의 프로세서들로 하여금, 합성 스피치 오디오 스트림을 스피커에 제공하는 것과 동시에 아바타를 디스플레이하게 한다.

[0189] 조항 28 은 조항 25 내지 조항 27 중 임의의 비일시적 컴퓨터 판독가능 저장 매체를 포함하고, 명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 하나 이상의 프로세서들로 하여금, 중단 이전에, 스피치 오디오 스트림에 기반하여 스피치 모델을 업데이트하게 하고, 텍스트-대-스피치 변환은 스피치 모델에 기반하여 수행된다.

[0190] 본 개시의 특정 양태들은 상호관련된 조항들의 제 4 세트로 하기에서 기술된다:

[0191] 조항 29 에 따르면, 장치로서, 온라인 미팅 동안, 제 1 사용자의 스피치를 나타내는 스피치 오디오 스트림을 수신하기 위한 수단, 제 1 사용자의 스피치를 나타내는 텍스트 스트림을 수신하기 위한 수단, 및 스피치 오디오

스트림에서의 중단에 응답하여 텍스트 스트림에 기반하여 출력을 의미 선택적으로 생성하기 위한 수단을 포함한다.

[0192] 조항 30 은 조항 29 의 장치를 포함하고, 스피치 오디오 스트림을 수신하기 위한 수단, 텍스트 스트림을 수신하기 위한 수단, 및 출력을 선택적으로 생성하기 위한 수단은, 가상 어시스턴트, 가전 제품, 스마트 디바이스, 사물 인터넷 (IoT) 디바이스, 통신 디바이스, 헤드셋, 차량, 컴퓨터, 디스플레이 디바이스, 텔레비전, 게이밍 콘솔, 음악 플레이어, 라디오, 비디오 플레이어, 엔터테인먼트 유닛, 개인용 미디어 플레이어, 디지털 비디오 플레이어, 카메라, 또는 내비게이션 디바이스 중 적어도 하나에 통합된다.

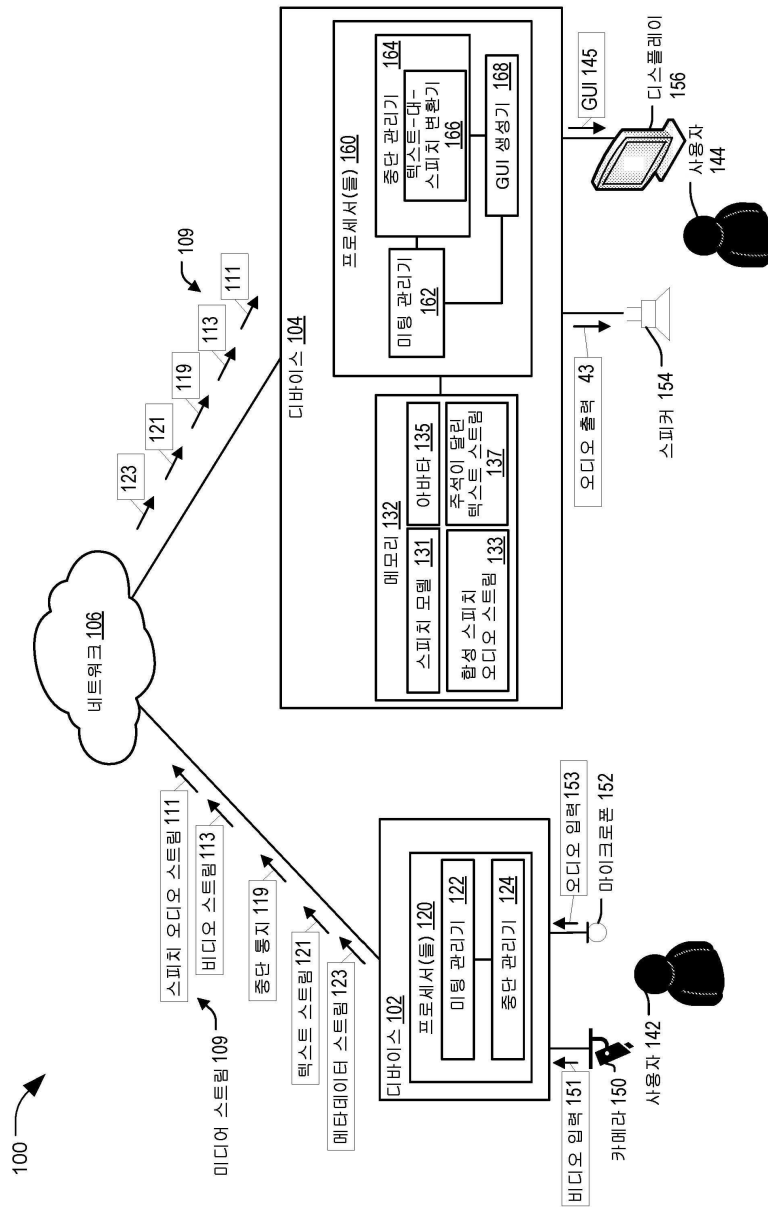
[0193] 당업자는 본 명세서에 개시된 구현들과 관련하여 설명된 다양한 예시적인 논리 블록들, 구성들, 모듈들, 회로들, 및 알고리즘 단계들이 전자 하드웨어, 프로세서에 의해 실행되는 컴퓨터 소프트웨어, 또는 이들 양자의 조합들로서 구현될 수도 있음을 추가로 인식할 것이다. 다양한 예시적인 컴포넌트들, 블록들, 구성들, 모듈들, 회로들 및 단계들이 일반적으로 그들의 기능성의 관점에서 상기 기술되었다. 그러한 기능성이 하드웨어로서 구현될지 또는 프로세서 실행가능 명령들로서 구현될지는 전체 시스템에 부과된 설계 제약들 및 특정 어플리케이션에 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.

[0194] 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. 다르게는, 저장 매체는 프로세서에 통합될 수도 있다. 프로세서 및 저장 매체는 ASIC (application-specific integrated circuit) 에 상주할 수도 있다. ASIC 은 컴퓨팅 디바이스 또는 사용자 단말기에 상주할 수도 있다. 대안으로, 프로세서 및 저장 매체는 컴퓨팅 디바이스 또는 사용자 단말기에서 이산 컴포넌트들로서 상주할 수도 있다.

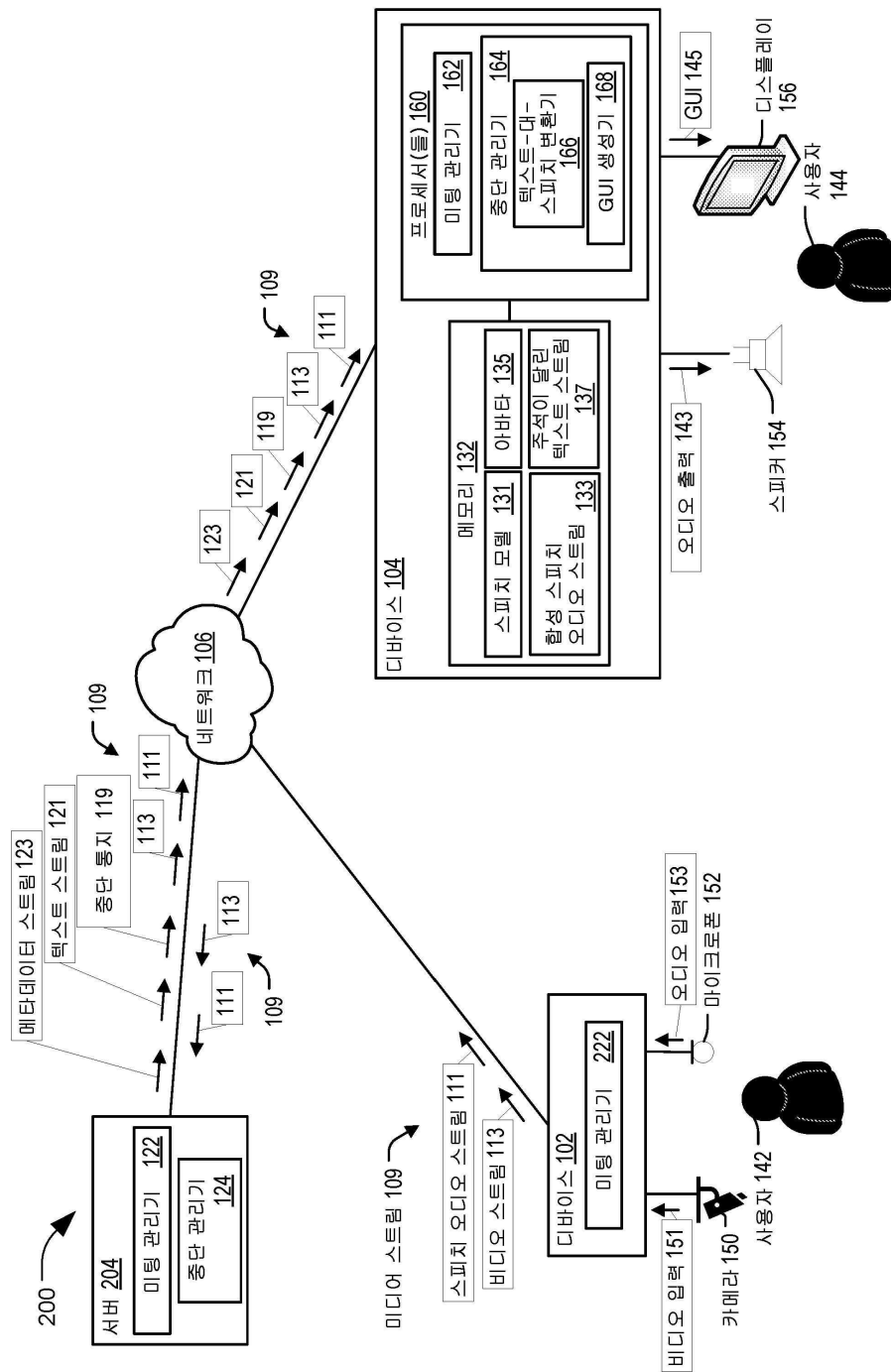
[0195] 개시된 실시형태들의 상기 설명은 당업자로 하여금 개시된 실시형태들을 제조 또는 이용할 수 있도록 제공된다. 이들 양태들에 대한 여러 변경들은 당업자들에게 명백할 것이며, 본원에서 정의된 원리들은 본 개시물의 범위로부터 일탈함이 없이 다른 양태들에 적용될 수도 있다. 따라서, 본 개시는 본원에서 나타낸 양태들에 한정하려는 것이 아니라, 다음 청구항들에 의해 정의되는 바와 같은 원리들 및 신규한 특징들과 가능한 부합하는 최광의의 범위를 부여하려는 것이다.

도면

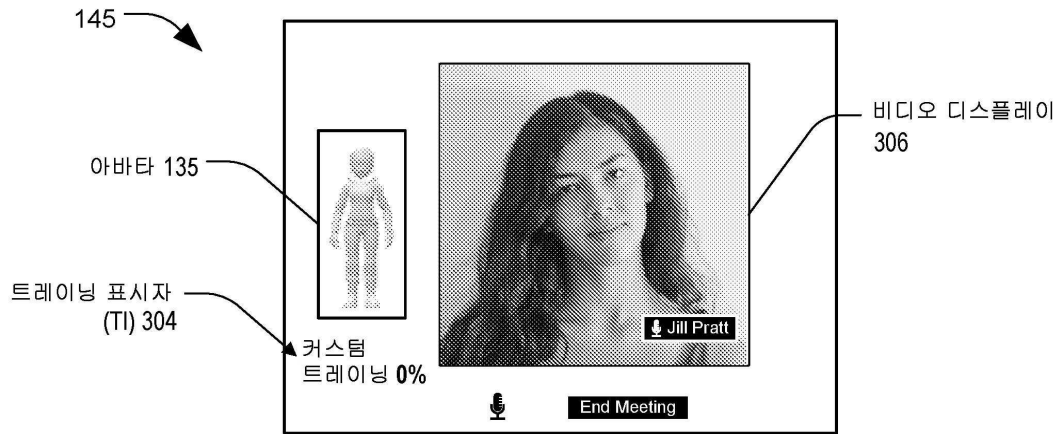
도면1



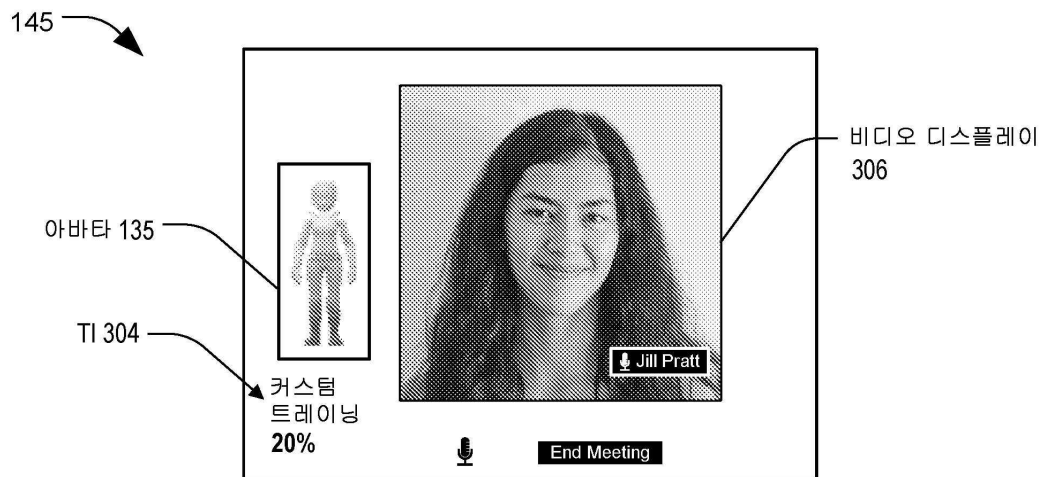
도면2



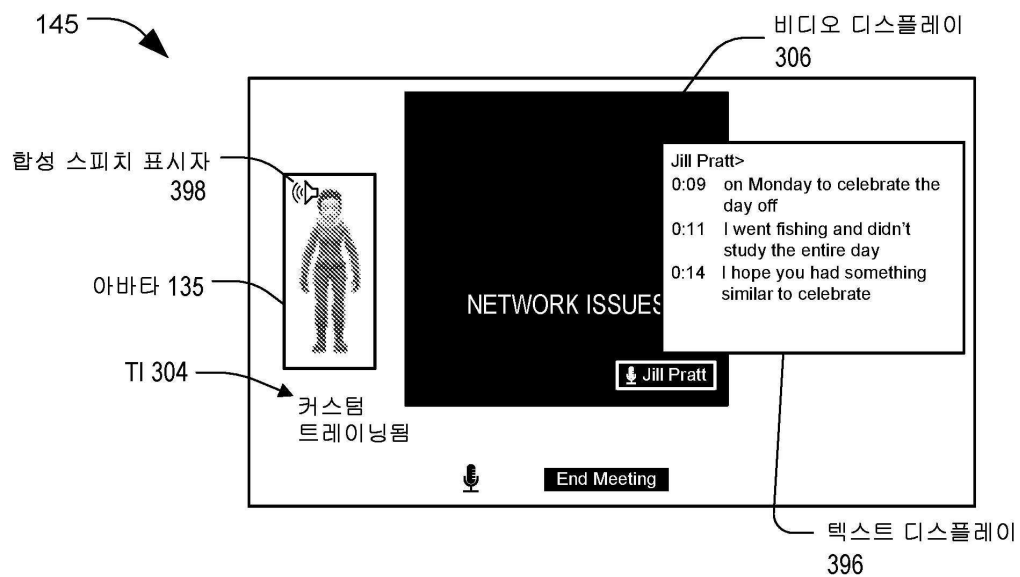
도면3a



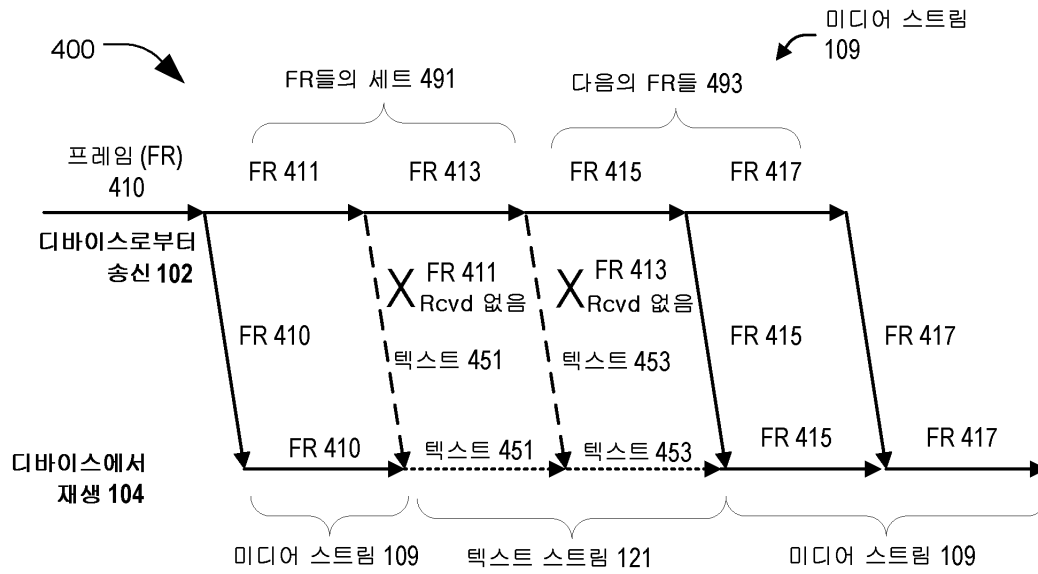
도면3b



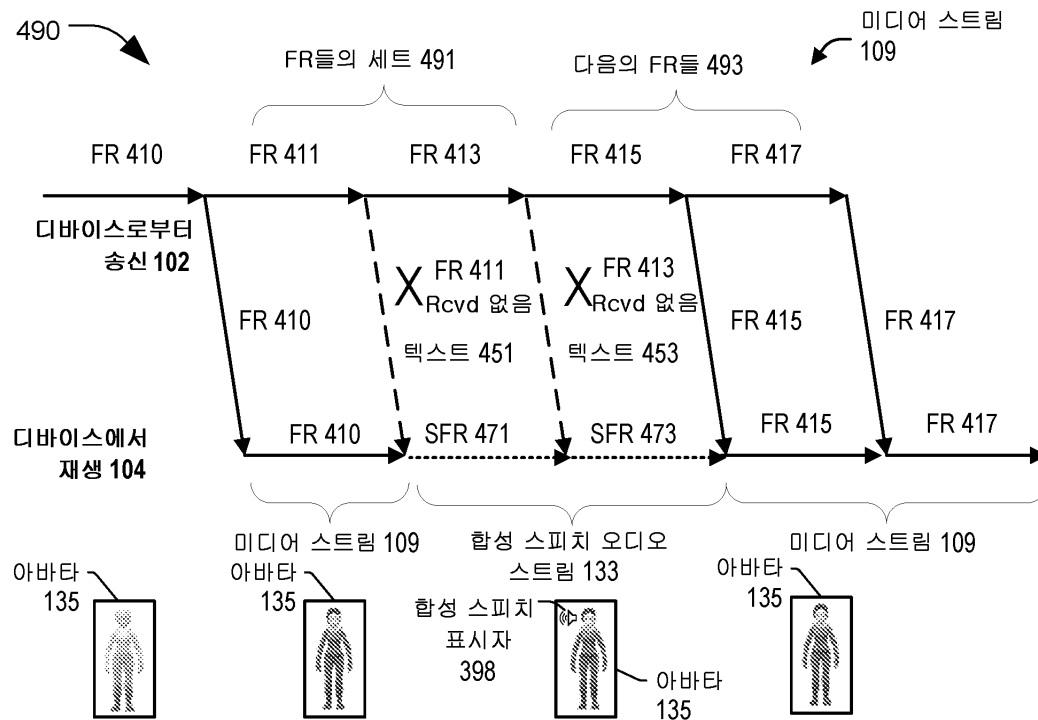
도면3c



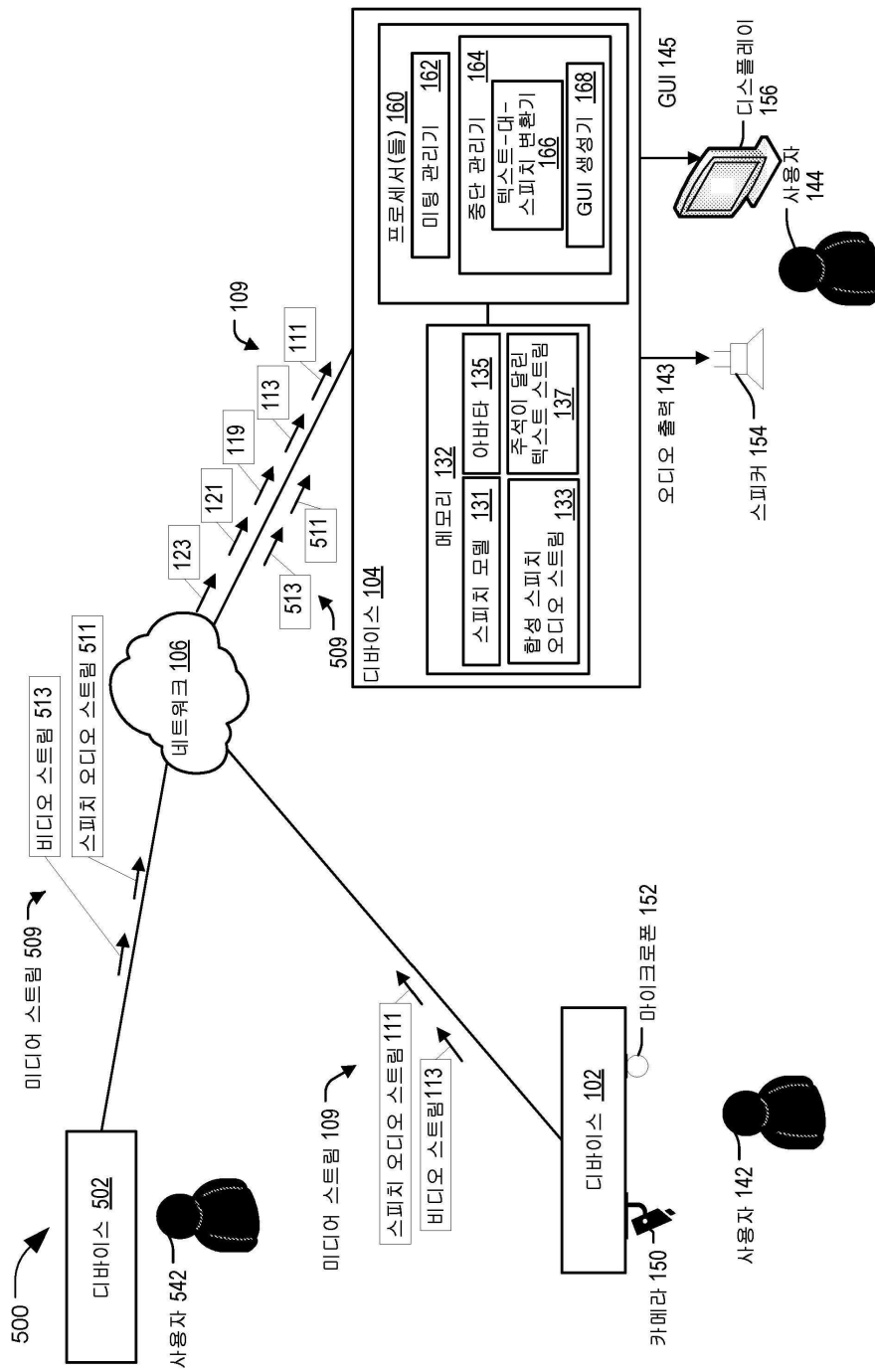
도면4a



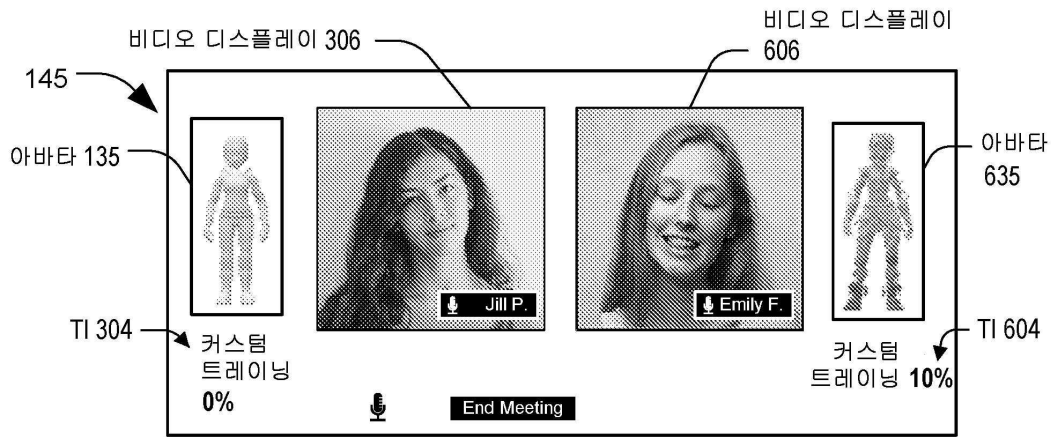
도면4b



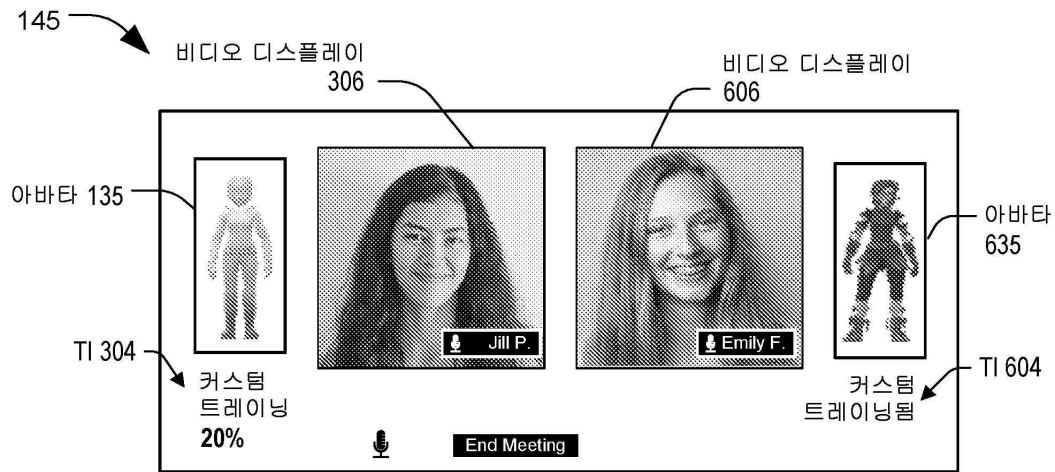
도면5



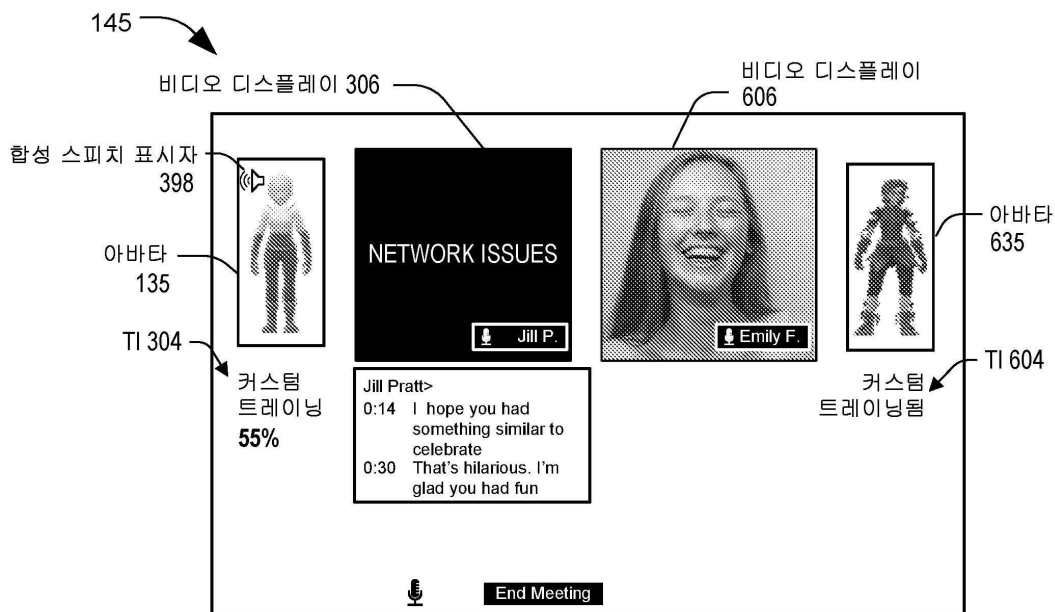
도면6a



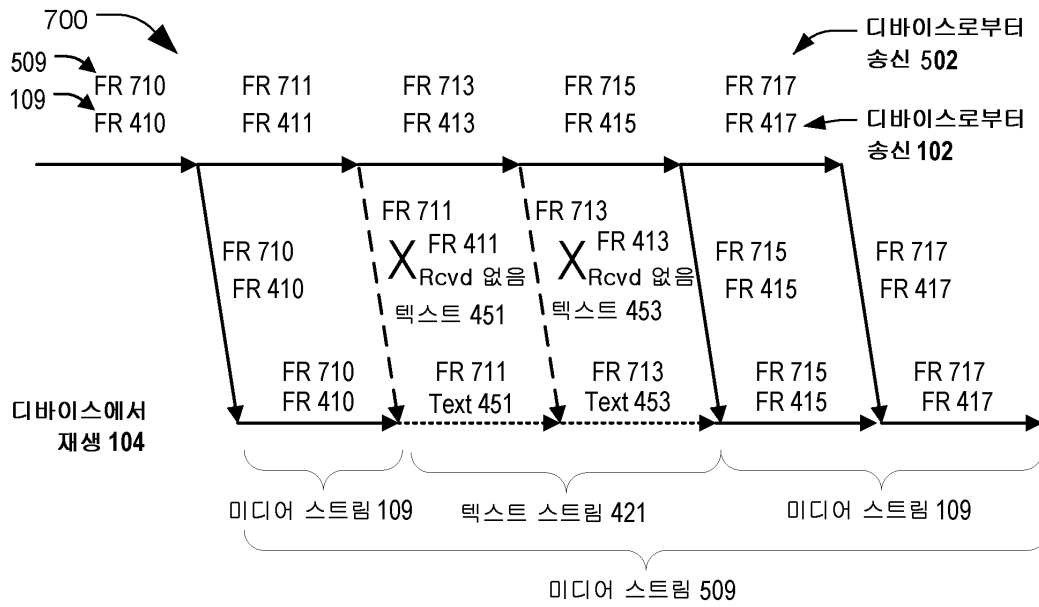
도면6b



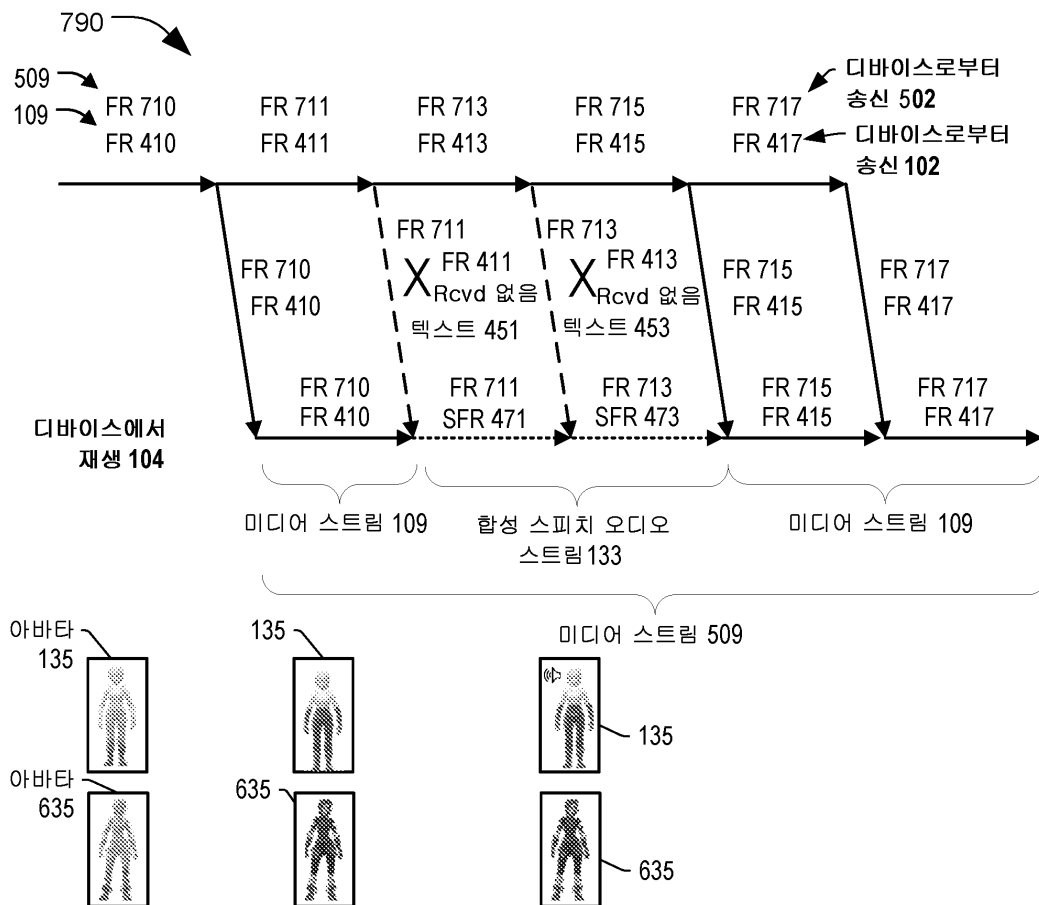
도면6c



도면7a

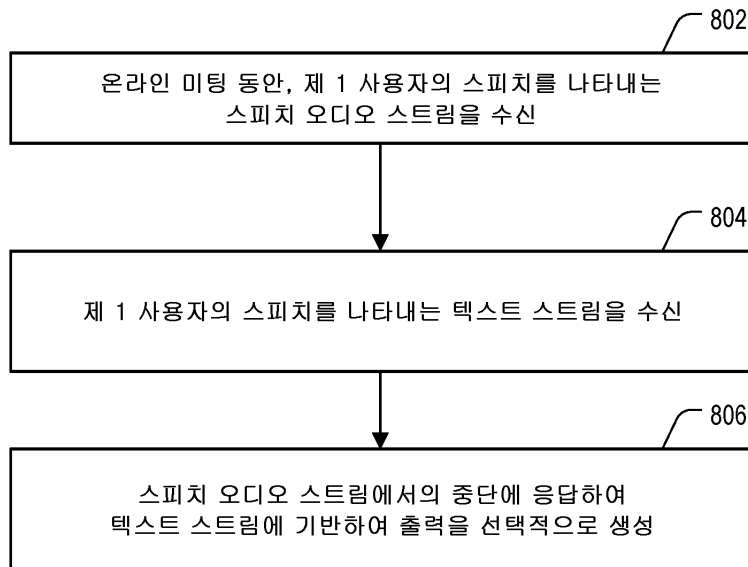


도면7b



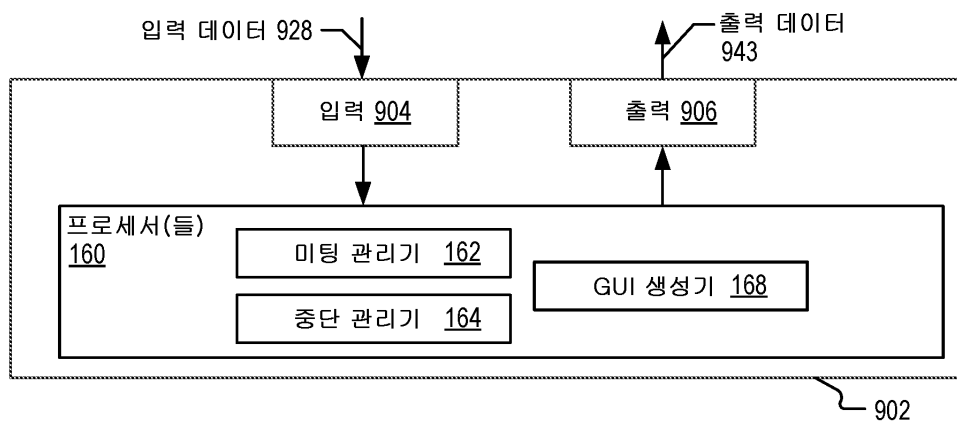
도면8

800

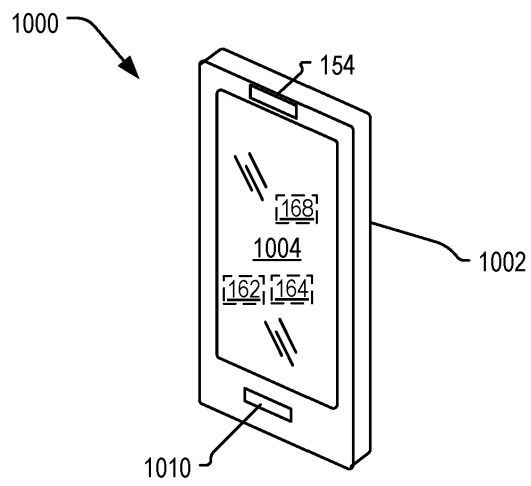


도면9

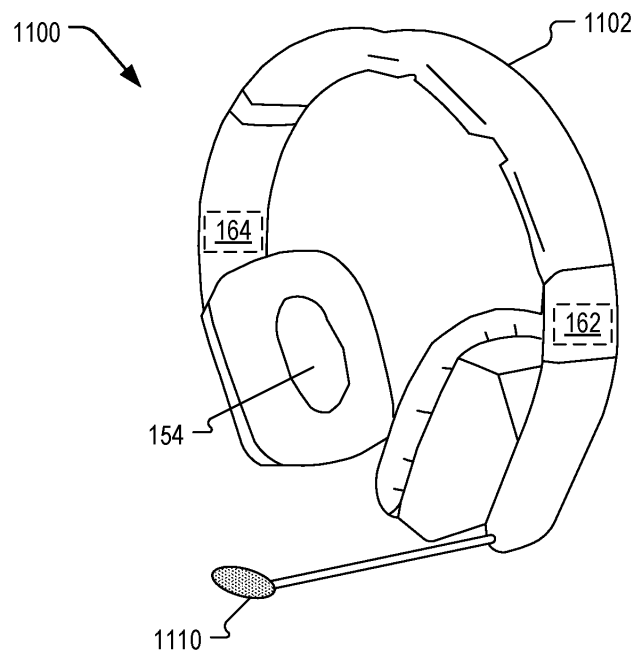
900



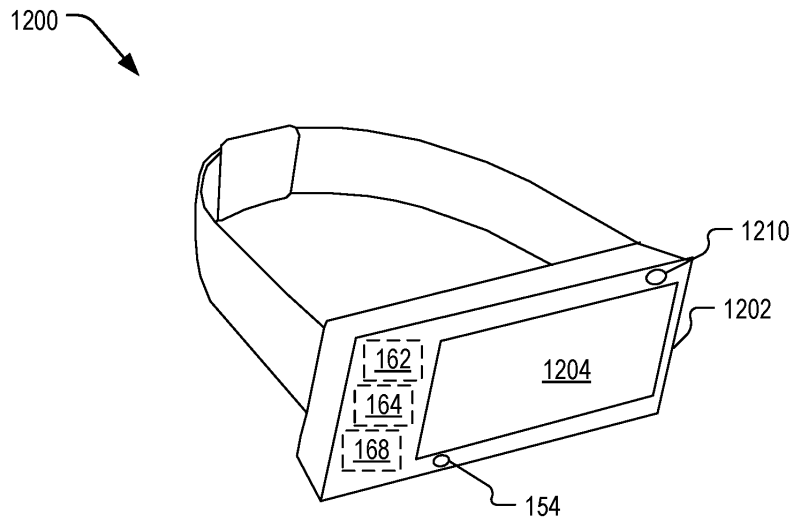
도면10



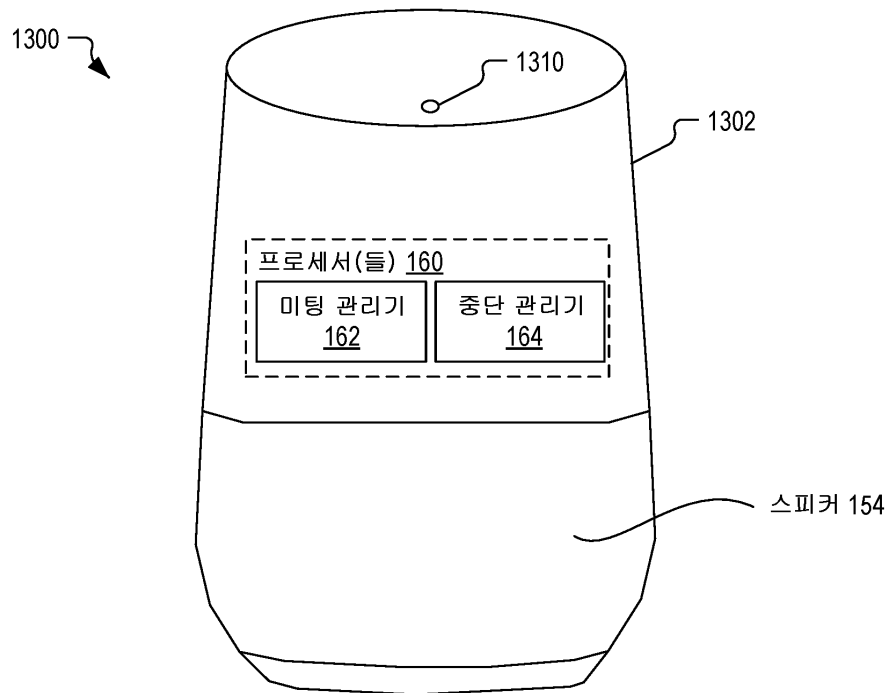
도면11



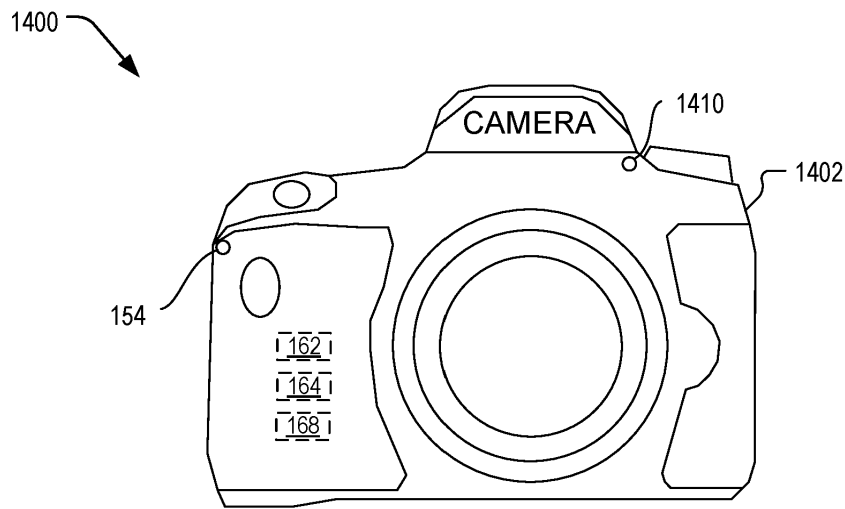
도면12



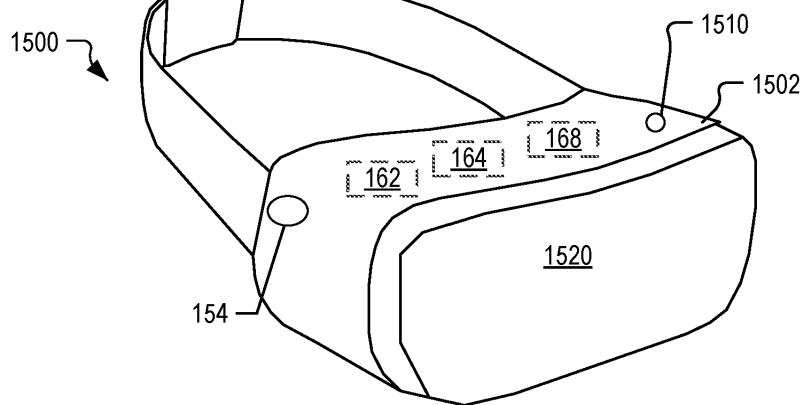
도면13



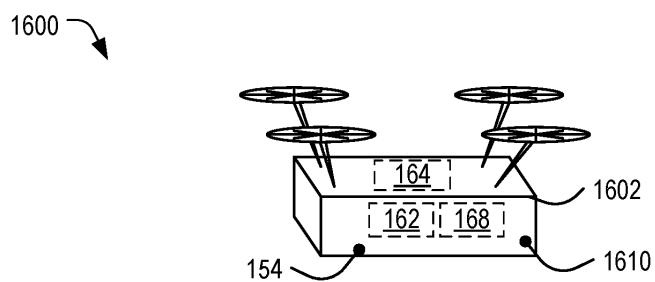
도면14



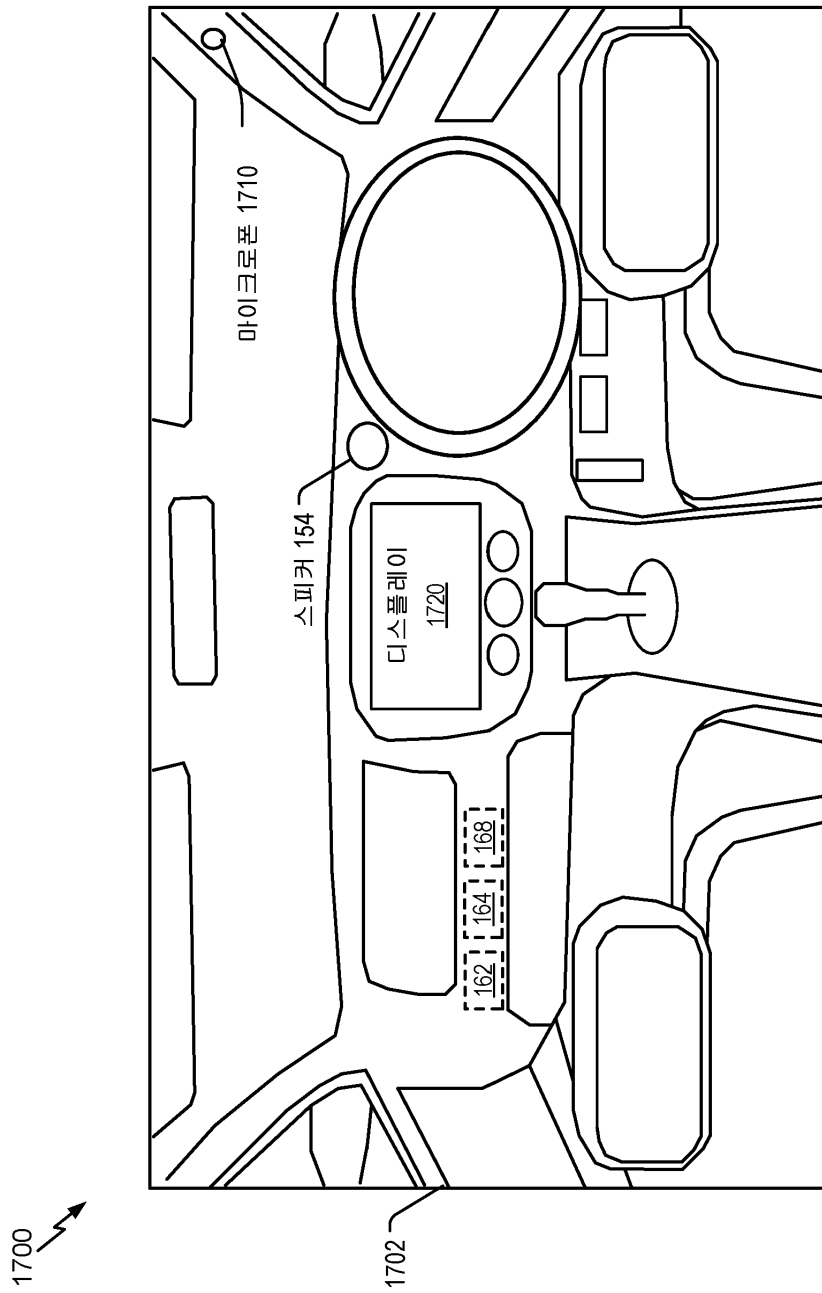
도면15



도면16



도면17



도면18

