(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년08월30일
(11) 등록번호 10-2572814
(24) 등록일자 2023년08월25일

(51) 국제특허분류(Int. Cl.)
G10L 15/22 (2006.01) *G10L 15/06* (2006.01)
G10L 15/08 (2006.01) *G10L 15/30* (2013.01)
G10L 19/018 (2013.01) *G10L 25/51* (2013.01)
(52) CPC특허분류
G10L 15/22 (2013.01)
G10L 15/063 (2013.01)
(21) 출원번호 10-2023-7002831(분할)
(22) 출원일자(국제) 2019년05월22일
심사청구일자 2023년02월17일
(85) 번역문제출일자 2023년01월25일
(65) 공개번호 10-2023-0018546
(43) 공개일자 2023년02월07일
(62) 원출원 특허 10-2020-7036730
원출원일자(국제) 2019년05월22일
심사청구일자 2020년12월21일
(86) 국제출원번호 PCT/US2019/033571
(87) 국제공개번호 WO 2019/226802
국제공개일자 2019년11월28일
(30) 우선권주장
62/674,973 2018년05월22일 미국(US)
16/418,415 2019년05월21일 미국(US)
(56) 선행기술조사문헌
US09548053 B1
US20150294666 A1
US20180130469 A1

(73) 특허권자
구글 엘엘씨
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
(72) 발명자
그루엔스타인 알렉산더 에이치.
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
조그레카 타랄 프라딥
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
(뒷면에 계속)
(74) 대리인
박장원

전체 청구항 수 : 총 15 항 심사관 : 성민용

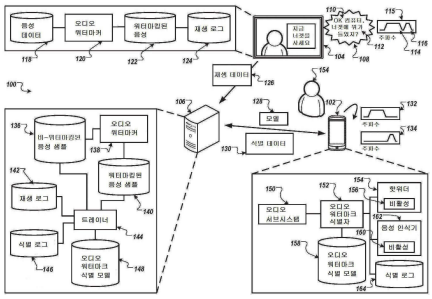
(54) 발명의 명칭 **핫워드 억제**

(57) 요 약

핫워드를 억제하기 위한 컴퓨터 저장 매체에 인코딩된 컴퓨터 프로그램들을 포함하는 방법, 시스템 및 장치가 개시된다. 일 양태에서, 방법은 발언의 재생에 대응하는 오디오 데이터를 수신하는 단계의 액션을 포함한다. 액션은 (i) 주어진 오디오 데이터 샘플이 오디오 워터마크를 포함하는지 여부를 결정하도록 구성되고, 그리고 (ii)

(뒷면에 계속)

대 표 도



오디오 워터마크 샘플을 각각 포함하는 워터마킹된 오디오 데이터 샘플 및 오디오 워터마크 샘플을 각각 포함하지 않는 비-워터마킹된 오디오 데이터 샘플을 사용하여 트레이닝되었던 모델에 대한 입력으로서 상기 오디오 데이터를 제공하는 단계를 더 포함한다. 액션은 모델로부터 오디오 데이터가 오디오 워터마크를 포함하는지 여부를 나타내는 데이터를 수신하는 단계를 더 포함한다. 액션은 상기 오디오 데이터가 상기 오디오 워터마크를 포함하는지 여부를 나타내는 데이터에 기초하여, 상기 오디오 데이터의 프로세싱을 계속하거나 중단하기로 결정하는 단계를 더 포함한다.

(52) CPC특허분류
G10L 15/08 (2013.01)
G10L 15/30 (2013.01)
G10L 19/018 (2013.01)
G10L 25/51 (2013.01)
G10L 2015/088 (2013.01)
G10L 2015/223 (2013.01)

(72) 발명자
페딘티 비자야디트야
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)

바치아니 미첼 에이.유.
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)

명 세 서

청구범위

청구항 1

방법으로서,

데이터 프로세싱 하드웨어에 의해, 복수의 비-워터마킹된 음성 샘플을 획득하는 단계, 각각의 비-워터마킹된 음성은 오디오 워터마크 샘플을 포함하지 않으며;

상기 복수의 비-워터마킹된 음성 샘플의 각각의 비-워터마킹된 음성 샘플로부터, 상기 데이터 프로세싱 하드웨어에 의해, 적어도 하나의 오디오 워터마크를 각각 포함하는 하나 이상의 대응하는 워터마킹된 음성 샘플을 생성하는 단계;

상기 데이터 프로세싱 하드웨어에 의해, 복수의 비-워터마킹된 음성 샘플 및 대응하는 워터마킹된 음성 샘플을 사용하여, 주어진 오디오 데이터 샘플이 오디오 워터마크를 포함하는지 여부를 결정하기 위한 모델을 트레이닝하는 단계; 및

상기 모델을 트레이닝한 후, 상기 데이터 프로세싱 하드웨어에 의해, 상기 트레이닝된 모델을 사용자 컴퓨팅 디바이스에 전송하는 단계를 포함하며, 상기 사용자 컴퓨팅 디바이스는:

발언의 재생에 대응하는 오디오 데이터를 수신하고;

상기 트레이닝된 모델을 사용하여, 상기 오디오 데이터가 상기 오디오 워터마크를 포함하는지 여부를 나타내는 데이터를 획득하고; 그리고

상기 오디오 데이터가 상기 오디오 워터마크를 포함하는지 여부를 나타내는 데이터에 기초하여, 상기 수신된 오디오 데이터의 프로세싱을 계속하거나 중단하기로 결정하도록 구성되는, 방법.

청구항 2

청구항 1에 있어서, 상기 복수의 비-워터마킹된 음성 샘플 각각은 핫워드의 발언을 포함하는, 방법.

청구항 3

청구항 1에 있어서, 상기 복수의 비-워터마킹된 음성 샘플의 일부는 핫워드의 발언을 포함하는, 방법.

청구항 4

청구항 1에 있어서, 상기 복수의 비-워터마킹된 음성 샘플 중 적어도 하나는 상기 복수의 비-워터마킹된 음성 샘플 중 적어도 하나의 다른 음성 샘플을 녹음하는데 사용된 디바이스와 상이한 디바이스에 의해 녹음되는, 방법.

청구항 5

청구항 1에 있어서, 상기 복수의 비-워터마킹된 음성 샘플 중 적어도 하나는 상기 복수의 비-워터마킹된 음성 샘플 중 적어도 하나의 다른 음성 샘플과 상이한 발화된 용어를 포함하는, 방법.

청구항 6

청구항 1에 있어서, 상기 복수의 비-워터마킹된 음성 샘플 중 적어도 일부는 배경 소음을 포함하는, 방법.

청구항 7

청구항 1에 있어서, 적어도 하나의 오디오 워터마크를 각각 포함하는 상기 하나 이상의 대응하는 워터마킹된 음성 샘플을 생성하는 단계는 핫워드의 발언을 포함하는 복수의 비-워터마킹된 음성 샘플 중 적어도 하나로부터, 상기 핫워드와만 오버랩핑되는 적어도 하나의 오디오 워터마크를 포함하는 적어도 하나의 대응하는 워터마킹된 음성 샘플을 생성하는 단계를 포함하는, 방법.

청구항 8

청구항 1에 있어서, 적어도 하나의 오디오 워터마크를 각각 포함하는 상기 하나 이상의 대응하는 워터마킹된 음성 샘플을 생성하는 단계는 핫워드의 발언을 포함하는 복수의 비-워터마킹된 음성 샘플 중 적어도 하나로부터, 상기 핫워드와 오버랩핑되고 상기 핫워드에 선행하는 오디오 워터마크의 시퀀스를 포함하는 적어도 하나의 대응하는 워터마킹된 음성 샘플을 생성하는 단계를 포함하는, 방법.

청구항 9

청구항 1에 있어서, 상기 복수의 비-워터마킹된 음성 샘플 중 적어도 하나로부터, 상기 하나 이상의 대응하는 워터마킹된 음성 샘플을 생성하는 단계는:

동일 간격 오디오 워터마크의 제1 시퀀스를 포함하는 제1 대응하는 워터마킹된 음성 샘플을 생성하는 단계; 및

동일 간격 오디오 워터마크의 제2 시퀀스를 포함하는 제2 대응하는 워터마킹된 음성 샘플을 생성하는 단계를 포함하며, 상기 동일 간격 오디오 워터마크의 제2 시퀀스의 오디오 워터마크 사이의 지속시간은 동일 간격 오디오 워터마크의 제1 시퀀스의 오디오 워터마크 사이의 지속시간과 상이한, 방법.

청구항 10

청구항 1에 있어서, 상기 워터마킹된 음성 샘플 중 하나에서의 상기 적어도 하나의 오디오 워터마크는 상기 워터마킹된 음성 샘플 중 다른 하나에서의 상기 적어도 하나의 오디오 워터마크와 상이한, 방법.

청구항 11

청구항 1에 있어서, 상기 모델을 트레이닝하기 이전에:

상기 데이터 프로세싱 하드웨어에 의해, 상기 모델을 트레이닝하는데 사용되는 복수의 비-워터마킹된 음성 샘플 각각을 오디오 워터마크를 포함하는 것으로 라벨링하는 단계; 및

상기 데이터 프로세싱 하드웨어에 의해, 상기 모델을 트레이닝하는데 사용되는 대응하는 워터마킹된 음성 샘플 각각을 오디오 워터마크를 포함하지 않는 것으로 라벨링하는 단계를 더 포함하는, 방법.

청구항 12

청구항 1에 있어서, 상기 모델을 트레이닝하는 단계는 상기 주어진 오디오 데이터 샘플이 상기 오디오 워터마크를 포함하는지 여부를 결정하기 위해 상기 복수의 비-워터마킹된 음성 샘플 및 상기 대응하는 워터마킹된 음성 샘플에 대해 상기 모델을 트레이닝하기 위해 기계 학습을 사용하는 것을 포함하는, 방법.

청구항 13

청구항 1에 있어서, 상기 오디오 데이터가 특정한 미리 정의된 핫워드의 발언을 포함한다는 결정에 응답하여 상기 오디오 데이터가 상기 오디오 워터마크를 포함하는지 여부를 나타내는 데이터를 획득하기 위해 상기 트레이닝된 모델을 사용하도록 구성되는, 방법.

청구항 14

시스템으로서,

하나 이상의 컴퓨터; 및

동작가능한 명령어를 저장하는 하나 이상의 저장 디바이스를 포함하며, 상기 명령어는 상기 하나 이상의 컴퓨터에 의해 실행될 때, 상기 하나 이상의 컴퓨터로 하여금 청구항 1 내지 13 중 어느 한 항에 따른 방법을 수행하게 하는, 시스템.

청구항 15

하나 이상의 컴퓨터에 의해 실행가능한 명령어를 포함하는 소프트웨어를 저장하는 비일시적 컴퓨터 판독가능 매체로서, 상기 명령어는 실행시 상기 하나 이상의 컴퓨터로 하여금 청구항 1 내지 13 중 어느 한 항의 방법을 수행하게 하는, 비일시적 컴퓨터 판독가능 매체.

발명의 설명

기 술 분 야

[0001] 관련 출원에 대한 상호 참조

[0002] 본 출원은 2018년 5월 22일 출원된 미국 출원 번호 62/674,973에 우선권을 주장하는 2019년 5월 21일 출원된 미국 출원 번호 16/418,415에 우선권을 주장하며, 두 내용 모두 참조로 통합된다.

[0003] 기술분야

[0004] 본 명세서는 자동 음성 프로세싱에 관한 것이다.

배 경 기 술

[0005] 음성 지원 가정 또는 다른 환경의 현실, 즉 사용자가 쿼리 또는 명령을 소리내어 말할 필요가 있고, 컴퓨터 기반 시스템은 쿼리에 응답하고 및/또는 명령이 수행되도록 하는 환경이 우리에게 있다. 음성 지원 환경(예: 가정, 직장, 학교 등)은 상기 환경의 다양한 방 또는 영역에 분산되어 있는 연결된 마이크로폰 디바이스들의 네트워크를 사용하여 구현될 수 있다. 이러한 마이크로폰들의 네트워크를 통해, 사용자는 자신이나 심지어 근처에 컴퓨터 또는 다른 디바이스를 가질 필요없이 본질적으로 환경의 어느 곳에서나 시스템에 구두로 쿼리할 수 있는 권한을 가진다. 예를 들어, 부엌에서 요리하는 동안, 사용자는 시스템에 "세 컵이 몇 밀리리터야?"를 물을 수 있으며, 이에 응답하여, 시스템으로부터 예를 들어, 합성된 음성 출력의 형태로 응답을 받을 수 있다. 대안적으로, 사용자는 시스템에 "가장 가까운 주유소가 언제 문 닫아?" 또는 집을 나갈 준비를 할 때 "오늘 코트를 입어야해?"와 같은 질문을 할 수 있다.

[0006] 또한, 사용자는 시스템에 쿼리를 묻고 및/또는 사용자의 개인 정보와 관련된 명령을 발행할 수 있다. 예를 들어, 사용자는 "존과의 미팅은 언제야?"라고 시스템에 묻거나 "집으로 갈 때 존에게 전화하라고 알려줘"라고 명령할 수 있다.

발명의 내용

[0007] 음성 지원 시스템에서, 시스템과 인터렉션하는 사용자의 방식은 주로, 배타적이지 않은 경우, 음성 입력에 의하도록 구성된다. 결과적으로, 시스템에 지시되지 않은 것들을 포함하여 주변 환경에서 만들어진 모든 발언을 잠재적으로 포착하는 시스템은 주어진 발언이 예를 들어, 환경에서 다른 사람에게 향하는 것과 반대로 시스템에 대해 지시된 것으로 구별하는 방식이 있어야 한다. 이를 수행하는 한 가지 방법은 환경에서 사용자 간에 동의에 의해 시스템의 주의를 끌기 위해 미리 결정된 단어 또는 단어들로 예약된 "핫워드"를 사용하는 것이다. 예시적 환경에서, 시스템의 주의를 끌기 위해 사용되는 핫워드는 "OK computer"이다. 결과적으로, "OK computer"라는 단어가 발화될 때마다, 이는 마이크로폰에 의해 포착되어 시스템에 전달되고, 음성 인식 기법들을 수행하거나 핫워드가 발화되었는지 여부를 결정하기 위해 오디오 피처들 및 신경 네트워크들을 사용하고 그렇다면, 후속 명령 또는 쿼리를 기다린다. 따라서 시스템을 향한 발언은 일반적인 형태인 [핫워드] [쿼리]이며, 여기서 이 예시에서 "핫워드"는 "OK computer"이고 "쿼리"는 임의의 질문, 명령, 선언 또는 단독으로 또는 네트워크를 통해 서버와 함께 음성 인식되고, 파싱되고 시스템에 의해 동작될 수 있는 다른 요청일 수 있다.

[0008] 본 개시는 재녹음된 음성, 예를 들어 방송된 음성 또는 텍스트 음성 변환 오디오를 라이브 음성과 구별하기 위한 오디오 워터마킹 기반 접근법을 논의한다. 이러한 구별은 재녹음된 음성을 포함하는 입력에서 거짓 핫워드 트리거를 검출할 수 있게 하고, 거짓 핫워드 트리거를 억제할 수 있게 한다. 그러나, 사용자의 라이브 음성 입력이 워터마킹되지 않을 것이며, 워터마킹되지 않은 것으로 결정된 음성 입력의 핫워드는 억제되지 않을 수 있다. 워터마크 검출 메커니즘은 소음 및 반향 환경에 견고하며, 작은 풋프린트, 메모리 및 계산, 낮은 대기 시간의 목표를 충족하도록 설계된 컨볼루션 신경 네트워크 기반 검출기를 사용할 수 있다. 이 접근법의 확장성 이점은 대규모 시청 TV 이벤트 중에 수백만 대의 디바이스에서 동시적으로 핫워드 트리거를 방지하는 것에서 강조된다.

[0009] 핫워드 기반 트리거링은 가상 어시스턴트를 활성화하기 위한 메커니즘일 수 있다. 라이브 음성의 핫워드를 녹음된 음성(예: 광고)로부터 구별하는 것은 거짓 핫워드 트리거가 가상 어시스턴트의 의도하지 않은 활성화로 이어지기 때문에 문제가 될 수 있다. 또한 사용자가 다수의 디바이스에 가상 어시스턴트를 설치한 경우, 한 가상 어시스턴트의 음성 출력에 의도치 않게 다른 가상 어시스턴트를 트리거하는 핫워드가 포함될 수 있다. 가상 어시

스턴트의 의도하지 않은 활성화는 일반적으로 바람직하지 않을 수 있다. 예를 들어, 가상 어시스턴트가 홈 오토메이션 디바이스를 제어하는데 사용되는 경우, 가상 어시스턴트를 의도하지 않은 활성화는 예를 들어 조명, 난방 또는 에어컨 장비가 의도치 않게 켜지고, 불필요한 에너지 소비로 이어질 수 있고, 사용자에게 불편할 수 있다. 또한 디바이스가 켜져 있을 때, 다른 디바이스로 메시지를 전송할 수 있고(예: 다른 디바이스로부터 정보 검색, 다른 디바이스에 상태 신호 전달, 검색을 수행하기 위해 검색 엔진과 통신 등), 의도하지 않게 디바이스를 켜면 불필요한 네트워크 트래픽 및/또는 불필요한 프로세싱 능력 사용, 불필요한 전력 소비 등이 발생할 수 있다. 또한, 조명, 난방 또는 에어컨 장비와 같은 장비를 의도하지 않게 활성화하면 장비가 불필요하게 마모되고 안정성이 저하될 수 있다. 또한 가상 어시스턴트 제어 장비 및 디바이스의 범위가 증가함에 따라, 가상 어시스턴트의 의도하지 않은 활성화가 잠재적으로 위험할 수 있는 가능성도 증가한다. 또한 가상 어시스턴트의 의도하지 않은 활성화는 개인 정보 보호에 대한 우려를 유발할 수 있다.

[0010] 본 명세서에 기술된 발명의 일 혁신적인 양태에 따르면, 액션들을 포함하는 핫워드를 억제하기 위한 방법을 포함하며, 상기 액션들은: 컴퓨팅 디바이스에 의해, 발언의 재생에 대응하는 오디오 데이터를 수신하는 단계; 상기 컴퓨팅 디바이스에 의해, (i) 주어진 오디오 데이터 샘플이 오디오 워터마크를 포함하는지 여부를 결정하도록 구성되고, 그리고 (ii) 오디오 워터마크 샘플을 각각 포함하는 워터마킹된 오디오 데이터 샘플 및 오디오 워터마크 샘플을 각각 포함하지 않는 비-워터마킹된 오디오 데이터 샘플을 사용하여 트레이닝되었던 모델에 대한 입력으로서 상기 오디오 데이터를 제공하는 단계; 상기 컴퓨팅 디바이스에 의해, (i) 주어진 오디오 데이터 샘플이 오디오 워터마크를 포함하는지 여부를 결정하도록 구성되고, 그리고 (ii) 오디오 워터마크 샘플을 포함하는 워터마킹된 오디오 데이터 샘플 및 오디오 워터마크 샘플을 포함하지 않는 비-워터마킹된 오디오 데이터 샘플을 사용하여 트레이닝되었던 모델로부터, 상기 오디오 데이터가 상기 오디오 워터마크를 포함하는지 여부를 나타내는 데이터를 수신하는 단계; 및, 상기 컴퓨팅 디바이스에 의해, 상기 오디오 데이터가 상기 오디오 워터마크를 포함하는지 여부를 나타내는 데이터에 기초하여, 상기 오디오 데이터의 프로세싱을 계속하거나 중단하기로 결정하는 단계를 포함하는, 방법.

[0011] 이들 또는 다른 실시예들은 다음 구성들 중 하나 이상을 각각 선택적으로 포함할 수 있다. 상기 오디오 데이터가 상기 오디오 워터마크를 포함하는지 여부를 나타내는 데이터를 수신하는 단계는 상기 오디오 데이터가 상기 오디오 워터마크를 포함함을 나타내는 데이터를 수신하는 것을 포함한다. 상기 오디오 데이터의 프로세싱을 계속하거나 중단하기로 결정하는 단계는 상기 오디오 데이터가 상기 오디오 워터마크를 포함함을 나타내는 데이터를 수신하는 것에 기초하여 상기 오디오 데이터의 프로세싱을 중단하기로 결정하는 것을 포함한다. 상기 액션은, 상기 오디오 데이터의 프로세싱을 중단하기로 결정하는 것에 기초하여, 상기 컴퓨팅 디바이스에 의해, 상기 오디오 데이터의 프로세싱을 중단하는 단계를 더 포함한다. 상기 오디오 데이터가 상기 오디오 워터마크를 포함하는지 여부를 나타내는 데이터를 수신하는 단계는 상기 오디오 데이터가 상기 오디오 워터마크를 포함하지 않음을 나타내는 데이터를 수신하는 것을 포함한다. 상기 오디오 데이터의 프로세싱을 계속하거나 중단하기로 결정하는 단계는 상기 오디오 데이터가 상기 오디오 워터마크를 포함하지 않음을 나타내는 데이터를 수신하는 것에 기초하여 상기 오디오 데이터의 프로세싱을 계속하기로 결정하는 것을 포함한다.

[0012] 상기 액션은, 상기 오디오 데이터의 프로세싱을 계속하기로 결정하는 것에 기초하여, 상기 컴퓨팅 디바이스에 의해, 상기 오디오 데이터의 프로세싱을 계속하는 단계를 더 포함한다. 오디오 데이터의 프로세싱은 오디오 데이터에 대해 음성 인식을 수행함으로써 발언의 전사를 생성하는 것을 포함한다. 오디오 데이터의 프로세싱은 오디오 데이터가 특정한, 미리 정의된 핫워드의 발언을 포함하는지 여부를 결정하는 것을 포함한다. 액션은, (i) 주어진 오디오 데이터 샘플이 오디오 워터마크를 포함하는지 여부를 결정하도록 구성되고, 그리고 (ii) 오디오 워터마크 샘플을 각각 포함하는 워터마킹된 오디오 데이터 샘플 및 오디오 워터마크 샘플을 각각 포함하지 않는 비-워터마킹된 오디오 데이터 샘플을 사용하여 트레이닝되었던 모델에 대한 입력으로서 상기 오디오 데이터를 제공하기 전에, 상기 컴퓨팅 디바이스에 의해, 상기 오디오 데이터가 특정한, 미리 정의된 핫워드의 발언을 포함한다고 결정하는 단계를 포함한다. 액션은 상기 컴퓨팅 디바이스에 의해, 상기 오디오 데이터가 특정한, 미리 정의된 핫워드의 발언을 포함한다고 결정하는 단계를 더 포함한다. (i) 주어진 오디오 데이터 샘플이 오디오 워터마크를 포함하는지 여부를 결정하도록 구성되고, 그리고 (ii) 오디오 워터마크 샘플을 각각 포함하는 워터마킹된 오디오 데이터 샘플 및 오디오 워터마크 샘플을 각각 포함하지 않는 비-워터마킹된 오디오 데이터 샘플을 사용하여 트레이닝되었던 모델에 대한 입력으로서 상기 오디오 데이터를 제공하는 단계는 상기 오디오 데이터가 특정한, 미리 정의된 핫워드의 발언을 포함한다고 결정하는 단계에 대한 응답으로 수행된다.

[0013] 액션은 상기 컴퓨팅 디바이스에 의해, 오디오 워터마크를 각각 포함하는 워터마킹된 상기 오디오 데이터 샘플, 오디오 워터마크를 각각 포함하지 않는 상기 비-워터마킹된 오디오 데이터 샘플, 각각 워터마킹된 및 비-워터마

킹된 오디오 샘플이 오디오 워터마크를 포함하는지 여부를 나타내는 데이터를 수신하는 단계; 및 상기 컴퓨팅 디바이스에 의해, 기계 학습을 사용하여, 오디오 워터마크를 각각 포함하는 워터마킹된 상기 오디오 데이터 샘플, 오디오 워터마크를 각각 포함하지 않는 상기 비-워터마킹된 오디오 데이터 샘플, 각각 워터마킹된 및 비-워터마킹된 오디오 샘플이 오디오 워터마크를 포함하는지 여부를 나타내는 데이터를 사용하여 상기 모델을 트레이닝하는 단계를 포함한다. 상기 워터마킹된 오디오 데이터 샘플의 적어도 일부는 다수의 주기적 위치에서 오디오 워터마크를 각각 포함한다. 상기 워터마킹된 오디오 데이터 샘플 중 하나의 오디오 워터마크는 상기 워터마킹된 오디오 데이터 샘플 중 다른 것의 오디오 워터마크와 상이하다. 액션은 상기 컴퓨팅 디바이스에 의해, 발언의 재생에 대응하는 상기 오디오 데이터의 수신의 제1 시간을 결정하는 단계; 상기 컴퓨팅 디바이스에 의해, 발언의 재생에 대응하는 상기 오디오 데이터 및 상기 오디오 데이터가 워터마크를 포함하는지 여부를 나타내는 데이터를 출력하기 위해 추가 컴퓨팅 디바이스가 제공된 제2 시간을 수신하는 단계; 상기 컴퓨팅 디바이스에 의해, 상기 제1 시간이 상기 제2 시간과 일치한다고 결정하는 단계; 및, 상기 컴퓨팅 디바이스에 의해, 상기 제1 시간이 상기 제2 시간과 일치한다고 결정하는 것에 기초하여, 상기 오디오 데이터가 워터마크를 포함하는지 여부를 나타내는 데이터를 사용하여 상기 모델을 업데이트하는 단계를 더 포함한다.

[0014] 본 양태의 다른 구현예는 대응하는 컴퓨터 시스템, 장치 및 하나 이상의 컴퓨터 저장 디바이스에 기록된 컴퓨터 프로그램을 포함하며, 그 각각은 방법의 동작들을 수행하도록 구성된다. 본 양태의 다른 구현예는 하나 이상의 컴퓨터에 의해 실행가능한 명령어를 포함하는 소프트웨어를 저장하는 비일시적 컴퓨터 판독가능 매체를 포함하며, 상기 명령어는 실행시 상기 하나 이상의 컴퓨터로 하여금 본 명세서에 기술된 임의의 방법을 수행하게 한다.

[0015] 본 명세서에 기술된 본 발명의 특정한 구현예들은 다음의 이점들을 실현하도록 구현될 수 있다. 컴퓨팅 디바이스는 녹음된 미디어에 포함된 핫워드에는 응답하지 않고 라이브 음성에 포함된 핫워드에는 응답할 수 있다. 이것은 디바이스의 의도하지 않은 활성화를 줄이거나 방지할 수 있으므로 배터리 전력과 컴퓨팅 디바이스의 프로세싱 능력을 절약할 수 있다. 네트워크 대역폭은 또한 오디오 워터마크가 있는 핫워드를 수신할 때 검색 쿼리를 수행하는 더 적은 컴퓨팅 디바이스로 보존될 수 있다.

[0016] 본 명세서에 기술된 본 발명의 하나 이상의 구현예들의 세부 사항은 첨부 도면과 아래의 설명에서 기술된다. 본 발명의 다른 구성들, 양태들 및 이점들은 설명, 도면 및 청구항으로부터 명백해질 것이다.

도면의 간단한 설명

[0017] 도 1은 기록 매체에서 핫워드를 검출할 때 핫워드 트리거를 억제하기 위한 예시적 시스템을 도시한다.

도 2는 기록 매체에서 핫워드를 검출할 때 핫워드 트리거를 억제하기 위한 예시적 프로세스의 흐름도이다.

도 3은 워터마킹 영역의 프레임에 대한 예시적 최소 마스킹 임계치, 에너지 및 청력의 절대 임계치를 도시한다.

도 4는 (a) 호스트 신호의 예시적 크기 스펙트로그램, (b) 워터마크 신호의 예시적 크기 스펙트로그램, (c) 워터마크 신호의 예시적 복제 부호 행렬, (d) 복제 부호 행렬 패턴과 부호 행렬을 단일 인스턴스와의 예시적 상관관계(수직선은 복제 간 워터마크 패턴의 예시적 경계를 표현함)를 도시한다.

도 5는 워터마크 검출기에 사용되는 예시적 신경 네트워크 아키텍처를 도시한다.

도 6은 상호-상관 패턴의 복제에 의해 생성된 예시적 매치 필터를 도시한다.

도 7은 예시적 신경 네트워크 출력 및 비-워터마킹된 신호에 대한 예시적 매치-필터링된 신경 네트워크 출력을 도시한다.

도 8는 컴퓨팅 디바이스 및 모바일 컴퓨팅 디바이스의 예를 도시한다.

도면에서, 동일한 번호는 전체에 걸쳐 대응하는 부분들을 표현한다.

발명을 실시하기 위한 구체적인 내용

[0018] 도 1은 기록 매체에서 "핫워드"를 검출할 때 핫워드 트리거를 억제하기 위한 예시적 시스템(100)을 도시한다. 간단히, 아래에서 더 상세히 설명되는 바와 같이, 컴퓨팅 디바이스(104)는 오디오 워터마크(116) 및 미리 정의된 핫워드(110)의 발언을 포함하는 발언(108)을 출력한다. 컴퓨팅 디바이스(102)는 발언(108)을 검출하고, 오디오 워터마크 식별 모델(158)을 사용하여 발언(108)이 오디오 워터마크(134)를 포함한다고 결정한다. 오디오 워터마크(134)를 포함하는 발언(108)에 기초하여, 컴퓨팅 디바이스(102)는 미리 정의된 핫워드(110)에 응답하지

않는다.

[0019] 구체적으로, 컴퓨팅 디바이스(104)는 Nugget World의 광고를 재생하고 있다. 광고를 하는 동안, 광고의 한 배우가 "Ok 컴퓨터, 너겟에 뭐가 들었지?"라는 발언을 한다. 발언(108)은 핫워드(110) "Ok 컴퓨터" 및 "너겟에 뭐가 들었지?"의 다른 용어를 포함하는 쿼리(112)를 포함한다. 컴퓨팅 디바이스(104)는 스피커를 통해 발언(108)을 출력한다. 마이크가 있는 주변의 모든 컴퓨팅 디바이스는 발언(108)을 검출할 수 있다.

[0020] 발언(108)의 오디오는 음성 부분(114) 및 오디오 워터마크(116)를 포함한다. 광고 제작자는 발언(108)을 검출하는 컴퓨팅 디바이스가 핫워드(110)에 응답하지 않도록 하기 위해 오디오 워터마크(116)를 추가할 수 있다. 일부 구현예에서, 오디오 워터마크(116)는 인간의 청력 범위보다 높거나 낮은 오디오 주파수를 포함할 수 있다. 예를 들어, 오디오 워터마크(116)는 20kHz보다 크거나 20Hz보다 작은 주파수를 포함할 수 있다. 일부 구현예에서, 오디오 워터마크(116)는 사람의 청력 범위 내에 있지만 소음과 유사한 소리 때문에 사람에 의해 감지되지 않는 오디오를 포함할 수 있다. 예를 들어, 오디오 워터마크(116)는 8~10kHz 사이의 주파수 패턴을 포함할 수 있다. 서로 다른 주파수 대역의 강도는 사람이 인식할 수 없지만 컴퓨팅 디바이스는 검출할 수 있다. 주파수 도메인 표현(115)에 의해 도시된 바와 같이, 발언(108)은 가청 부분(114)보다 더 높은 주파수 범위에 있는 오디오 워터마크(116)를 포함한다.

[0021] 일부 구현예에서, 컴퓨팅 디바이스(104)는 음성 데이터(118)에 워터마크를 추가하기 위해 오디오 워터마커(120)를 사용할 수 있다. 음성 데이터(118)는 "Ok 컴퓨터, 너겟에 뭐가 들었지?"의 녹음된 발언(108)일 수 있다. 오디오 워터마커(120)는 음성 데이터(118)에 주기적 간격으로 워터마크를 추가할 수 있다. 예를 들어, 오디오 워터마커(120)는 200밀리 초마다 워터마크를 추가할 수 있다. 일부 구현예에서, 컴퓨팅 디바이스(104)는 예를 들어 음성 인식을 수행함으로써 핫워드(110)를 포함하는 음성 데이터(118)의 부분을 식별할 수 있다. 오디오 워터마커(120)는 핫워드(110)의 오디오 위에, 핫워드(110) 이전 및/또는 핫워드(110) 이후에 주기적 워터마크를 추가할 수 있다. 예를 들어, 오디오 워터마커(120)는 "ok 컴퓨터"의 오디오 상에 주기적인 간격으로 3개(또는 임의의 다른 수)의 워터마크를 추가할 수 있다.

[0022] 워터마크(120)를 추가하기 위한 기술은 도 3 내지 도 7과 관련하여 아래에서 상세히 논의된다. 일반적으로, 각 워터마크(120)는 각 음성 데이터 샘플에 대해 상이하다. 오디오 워터마커(120)는 발언(108)의 오디오에 200밀리 초 또는 300밀리 초마다 오디오 워터마크를 추가하고, 발언의 오디오 "Ok 컴퓨터, 치즈 피자를 주문해"에 200밀리 초 또는 300밀리 초마다 서로 다르거나 동일한 오디오 워터마크를 추가할 수 있다. 오디오 워터마커(120)는 워터마크가 오디오 샘플의 왜곡을 최소화하도록 각 오디오 샘플마다 워터마크를 생성할 수 있다. 이는 오디오 워터마커(120)가 인간이 감지할 수 있는 주파수 범위 내에 있는 워터마크를 추가할 수 있기 때문에 중요할 수 있다. 컴퓨팅 디바이스(104)는 컴퓨팅 디바이스(104)에 의한 추후 출력을 위해 워터마킹된 음성(112)에 워터마킹된 오디오 샘플을 저장할 수 있다.

[0023] 일부 구현예에서, 컴퓨팅 디바이스(104)가 워터마킹된 오디오를 출력할 때마다, 컴퓨팅 디바이스(104)는 출력된 오디오를 나타내는 데이터를 재생 로그(124)에 저장할 수 있다. 재생 로그(124)는 출력된 오디오(108), 오디오(108)를 출력한 날짜 및 시간, 컴퓨팅 디바이스(104), 컴퓨팅 디바이스(104)의 위치, 오디오(108)의 전사 및 워터마크 없는 오디오(108)의 임의의 조합을 식별하는 데이터를 포함할 수 있다.

[0024] 컴퓨팅 디바이스(102)는 마이크로폰을 통해 발언(108)을 검출한다. 컴퓨팅 디바이스(102)는 오디오를 수신할 수 있는 임의의 유형의 디바이스일 수 있다. 예를 들면, 컴퓨팅 디바이스(102)는 데스크톱 컴퓨터, 랩톱 컴퓨터, 태블릿 컴퓨터, 웨어러블 컴퓨터, 셀룰러폰, 스마트폰, 뮤직 플레이어, 전자책 리더, 네비게이션 시스템, 스마트 스피커 및 홈 어시스턴트, 무선(예: 블루투스) 헤드셋, 보청기, 스마트 시계, 스마트 안경, 활동 추적기 또는 임의의 다른 적절한 컴퓨팅 디바이스를 포함할 수 있다. 도 1에 도시된 바와 같이, 컴퓨팅 디바이스(102)는 스마트폰이다. 컴퓨팅 디바이스(104)는 예를 들어 텔레비전, 라디오, 뮤직 플레이어, 데스크톱 컴퓨터, 랩톱 컴퓨터, 태블릿 컴퓨터, 웨어러블 컴퓨터, 셀룰러폰, 또는 스마트폰과 같은 오디오를 출력할 수 있는 임의의 디바이스일 수 있다. 도 1에 도시된 바와 같이, 컴퓨팅 디바이스(104)는 텔레비전이다.

[0025] 컴퓨팅 디바이스(102)의 마이크로폰은 오디오 서브시스템(150)의 일부일 수 있다. 오디오 서브시스템(150)은 마이크로폰을 통해 수신된 오디오를 초기에 프로세싱하도록 각각 설계된 버퍼, 필터, 아날로그-디지털 변환기를 포함할 수 있다. 버퍼는 마이크로폰을 통해 수신되고 오디오 서브시스템(150)에 의해 프로세싱된 현재 오디오를 저장할 수 있다. 예를 들어, 버퍼는 이전 5초의 오디오 데이터를 저장한다.

[0026] 컴퓨팅 디바이스(102)는 오디오 워터마크 식별자(152)를 포함한다. 오디오 워터마크 식별자(152)는 마이크로폰

을 통해 수신 및/또는 버퍼에 저장된 오디오를 프로세싱하고, 오디오에 포함된 오디오 워터마크를 식별하도록 구성된다. 오디오 워터마크 식별자(152)는 프로세싱된 오디오를 오디오 워터마크 식별 모델(158)에 대한 입력으로서 제공하도록 구성될 수 있다. 오디오 워터마크 식별 모델(158)은 오디오 데이터를 수신하고, 오디오 데이터가 워터마크를 포함하는지 여부를 나타내는 데이터를 출력하도록 구성될 수 있다. 예를 들어, 오디오 워터마크 식별자(152)는 오디오 서브시스템(150)을 통해 프로세싱된 오디오를 오디오 워터마크 식별 모델(158)에 지속적으로 제공할 수 있다. 오디오 워터마크 식별자(152)가 더 많은 오디오를 제공함에 따라, 오디오 워터마크 식별 모델(158)의 정확도가 증가할 수 있다. 예를 들어, 300밀리 초 후에, 오디오 워터마크 식별 모델(158)은 하나의 워터마크를 포함하는 오디오를 수신할 수 있다. 500밀리 초 후에, 오디오 워터마크 식별 모델(158)은 2개의 워터마크를 포함하는 오디오를 수신할 수 있다. 임의의 하나의 오디오 샘플의 워터마크가 모두 서로 동일한 실시예에서, 오디오 워터마크 식별 모델(158)은 더 많은 오디오를 프로세싱함으로써 정확성을 향상시킬 수 있다.

[0027] 일부 구현예에서, 오디오 워터마크 식별자(152)는 오디오 서브시스템(150)으로부터 수신된 오디오로부터 임의의 검출된 워터마크를 제거하도록 구성될 수 있다. 워터마크를 제거한 후, 오디오 워터마크 식별자(152)는 워터마크가 없는 오디오를 핫워더(154) 및/또는 음성 인식기(162)에 제공할 수 있다. 일부 구현예에서, 오디오 워터마크 식별자(152)는 오디오 서브시스템(150)으로부터 수신된 오디오를 워터마크를 제거하지 않고 핫워더(154) 및/또는 음성 인식기(162)에 전달하도록 구성될 수 있다.

[0028] 핫워더(154)는 마이크로폰을 통해 수신되고 및/또는 버퍼에 저장된 오디오에서 핫워드를 식별하도록 구성된다. 일부 구현예에서, 핫워더(154)는 컴퓨팅 디바이스(102)의 전원이 켜질 때마다 활성화될 수 있다. 핫워더(154)는 버퍼에 저장된 오디오 데이터를 지속적으로 분석할 수 있다. 핫워더(154)는 버퍼 내의 현재 오디오 데이터가 핫워드를 포함할 가능성을 반영하는 핫워드 신뢰 점수를 계산한다. 핫워드 신뢰 점수를 계산하기 위해, 핫워더(154)는 필터 뱅크 에너지 또는 멜-주파수 셉스트럴 계수(mel-frequency cepstral coefficient)들과 같은 오디오 데이터로부터 오디오 피처들을 추출할 수 있다. 핫워드기(154)는 지원 벡터 머신(support vector machine) 또는 신경 네트워크를 사용하는 것과 같은 이들 오디오 피처들을 프로세싱하기 위해 분류 윈도우들을 사용할 수 있다. 일부 구현예에서, 핫워더(154)는 핫워드 신뢰 점수를 결정하기 위해 음성 인식을 수행하지 않는다(예를 들어, 수신된 오디오로부터 추출된 오디오 피처를 하나 이상의 핫워드에 대한 대응 오디오 피처와 비교하지만, 오디오 데이터에 대한 음성 인식을 수행하기 위해 추출된 오디오 피처를 사용하지 않음). 핫워더(154)는 핫워드 신뢰 점수가 핫워드 신뢰 점수 임계치를 만족하면 오디오가 핫워드를 포함하는 것으로 결정한다. 예를 들면, 핫워더(154)는 핫워드 신뢰 점수가 핫워드 신뢰 점수가 0.8이고 핫워드 신뢰 점수 임계치가 0.7이면, 발언(108)에 대응하는 오디오가 핫워드(110)를 포함하는 것으로 결정한다. 일부 경우에, 핫워드는 깨우기 단어 또는 주의 단어로 지칭될 수 있다.

[0029] 음성 인식기(162)는 들어오는 오디오에 기초하여 전사를 생성하는 임의의 유형의 프로세스를 수행할 수 있다. 예를 들어, 음성 인식기(162)는 버퍼의 오디오 데이터에서 음소를 식별하기 위해 음향 모델을 사용할 수 있다. 음성 인식기(162)는 음소에 대응하는 전사를 결정하기 위해 언어 모델을 사용할 수 있다. 다른 예로서, 음성 인식기(162)는 버퍼의 오디오 데이터를 프로세싱하고 전사를 출력하는 단일 모델을 사용할 수 있다.

[0030] 오디오 워터마크 식별 모델(158)이 오디오가 워터마크를 포함한다고 결정하는 경우, 오디오 워터마크 식별자(152)는 음성 인식기(162) 및/또는 핫워더(154)를 비활성화할 수 있다. 음성 인식기(162) 및/또는 핫워더(154)를 비활성화함으로써, 오디오 워터마크 식별자(152)는 컴퓨팅 디바이스(102)가 핫워드(110) 및/또는 쿼리(112)에 응답하도록 트리거링할 수 있는 오디오의 추가 프로세싱을 방지할 수 있다. 도 1에 도시된 바와 같이, 오디오 워터마크 식별자(152)는 핫워더(154)를 비활성 상태(156)로 설정하고 음성 인식기(162)를 비활성 상태(160)로 설정한다.

[0031] 일부 구현예에서, 핫워더(154)의 디폴트 상태는 활성 상태일 수 있고, 음성 인식기(162)의 디폴트 상태는 활성 상태일 수 있다. 이 경우, 비활성 상태(156) 및 비활성 상태(160)는 미리 결정된 시간 후에 만료될 수 있다. 예를 들어, 5초(또는 다른 미리 결정된 시간) 후에, 핫워더(154) 및 음성 인식기(162) 모두의 상태는 활성 상태로 복귀할 수 있다. 5초 주기는 오디오 워터마크 식별자(152)가 오디오 워터마크를 검출할 때마다 갱신될 수 있다. 예를 들어, 발언(108)의 오디오(115)가 오디오의 지속 기간 동안 워터마크를 포함하는 경우, 핫워더(154) 및 음성 인식기(162)는 비활성 상태(156) 및 비활성 상태(162)로 설정될 수 있고, 발언(108)을 출력하는 컴퓨팅 디바이스(104)의 종료 후 추가 5초 동안 해당 상태를 유지할 수 있다. 다른 예로서, 발언(108)의 오디오(115)가 핫워드(110)의 발언 전반에 걸쳐 워터마크를 포함한다면, 핫워더(154) 및 음성 인식기(162)는 비활성 상태(156) 및 비활성 상태(162)로 설정될 수 있고, 컴퓨팅 디바이스(104)가 쿼리(112)의 출력과 중복될 핫워드(110)를 출

력한 후 추가 5초 동안 해당 상태를 유지할 수 있다.

[0032] 일부 구현예에서, 오디오 워터마크 식별자(152)는 오디오 워터마크 식별자(152)가 워터마크를 식별한 날짜 및 시간을 나타내는 데이터를 식별 로그(164)에 저장할 수 있다. 예를 들어, 오디오 워터마크 식별자(152)는 2019년 6월 10일 오후 3시 15분에 발언(108)의 오디오에서 워터마크를 식별할 수 있다. 식별 로그(164)는 워터마크 수신 시간 및 날짜, 워터마크(134)를 포함하는 발언의 전사, 컴퓨팅 디바이스(102), 워터마크(134), 워터마크 검출시 컴퓨팅 디바이스(102)의 위치, 기본 오디오(132), 결합된 오디오 및 워터마크 및 발언(108) 또는 워터마크(134) 전후에 일정 기간 동안 검출된 임의의 오디오의 임의의 조합을 식별하는 데이터를 저장할 수 있다.

[0033] 일부 구현예에서, 오디오 워터마크 식별자(152)는 오디오 워터마크 식별자(152)가 워터마크를 식별하지 않았고 핫워더(154)가 핫워드를 식별한 날짜 및 시간을 나타내는 데이터를 식별 로그(164)에 저장할 수 있다. 예를 들어, 2019년 6월 20일 오후 7시 15 분에, 오디오 워터마크 식별자(152)는 발언의 오디오에서 워터마크를 식별하지 않을 수 있고, 핫워더(154)는 발언의 오디오에서 핫워드를 식별할 수 있다. 식별 로그(164)는 비-워터마킹된 오디오 및 핫워드의 수신 시간 및 날짜, 발언의 전사, 컴퓨팅 디바이스(102), 컴퓨팅 디바이스의 위치, 오디오가 발언 또는 핫워드 전후의 기간을 검출한 것의 임의의 조합을 식별하는 데이터를 저장할 수 있다.

[0034] 일부 구현예에서, 핫워더(154)는 오디오 워터마크 식별자(152) 이전, 이후 또는 동시에 오디오 서브시스템(150)으로부터 수신된 오디오를 프로세싱할 수 있다. 예를 들어, 오디오 워터마크 식별자(152)는 발언(108)의 오디오가 워터마크를 포함한다고 결정할 수 있고, 동시에 핫워더(154)는 발언(108)의 오디오가 핫워드를 포함한다고 결정할 수 있다. 이 경우, 오디오 워터마크 식별자(152)는 음성 인식기(162)의 상태를 비활성 상태(160)로 설정할 수 있다. 오디오 워터마크 식별자(152)는 핫워더(154)의 상태(156)를 업데이트할 수 없을 수 있다.

[0035] 일부 구현예에서, 오디오 워터마크 식별자(152)가 오디오 워터마크 식별 모델(158)을 사용하기 전에, 컴퓨팅 디바이스(106)는 워터마크 식별 데이터(130)를 생성하고 워터마크 식별 데이터(130)를 컴퓨팅 디바이스(102)에 제공한다. 컴퓨팅 디바이스(106)는 비-워터마킹된 음성 샘플(136), 오디오 워터마커(138) 및 기계 학습을 사용하여 오디오 워터마크 식별 모델(148)을 생성하는 트레이너(144)를 사용한다.

[0036] 비-워터마킹된 음성 샘플(136)은 다양한 조건에서 수집된 다양한 음성 샘플을 포함할 수 있다. 비-워터마킹된 음성 샘플(136)은 서로 다른 사용자가 서로 다른 용어를 말하고, 동일한 용어를 말하고, 서로 다른 유형의 배경 잡음을 갖고 용어를 말하고, 서로 다른 언어로 용어를 말하고, 서로 다른 악센트로 단어를 말하고, 서로 다른 디바이스에 의해 녹음된 용어를 말하는 오디오 샘플을 포함할 수 있다. 일부 구현예에서, 비-워터마킹된 음성 샘플(136)은 핫워드의 발언을 각각 포함한다. 일부 구현예에서, 비-워터마킹된 음성 샘플(136) 중 일부만이 핫워드의 발언을 포함한다.

[0037] 오디오 워터마커(138)는 비-워터마킹된 음성 샘플 각각에 대해 상이한 워터마크를 생성할 수 있다. 오디오 워터마커(138)는 비-워터마킹된 음성 샘플 각각에 대해 하나 이상의 워터마킹된 음성 샘플(140)을 생성할 수 있다. 동일한 비-워터마킹된 음성 샘플을 사용하여, 오디오 워터마커(138)는 매 200밀리 초마다 워터마크를 포함하는 워터마킹된 음성 샘플 및 300밀리 초마다 워터마크를 포함하는 또 다른 워터마킹된 음성 샘플을 생성할 수 있다. 오디오 워터마커(138)는 또한 존재한다면 핫워드와 오버랩핑되는 워터마크만을 포함하는 워터마킹된 음성 샘플을 생성할 수 있다. 오디오 워터마커(138)는 또한 핫워드와 오버랩핑되고 핫워드에 선행하는 워터마크를 포함하는 워터마킹된 음성 샘플을 생성할 수 있다. 이 경우, 오디오 워터마커(138)는 동일한 비-워터마킹된 음성 샘플로 4개의 서로 다른 워터마킹된 음성 샘플을 만들 수 있다. 오디오 워터마커(138)는 또한 4개 보다 많거나 적게 만들 수 있다. 일부 예에서, 오디오 워터마커(138)는 오디오 워터마커(120)와 유사하게 동작할 수 있다.

[0038] 트레이너(144)는 오디오 워터마크 식별 모델(148)을 생성하기 위해 비-워터마킹된 음성 샘플(136) 및 워터마킹된 음성 샘플(140)을 포함하는 기계 학습 및 트레이닝 데이터를 사용한다. 비-워터마킹된 음성 샘플(136) 및 워터마킹된 음성 샘플(140)이 워터마크를 포함하거나 포함하지 않는 것으로 라벨링되기 때문에, 트레이너(148)는 비-워터마킹된 음성 샘플(136)을 포함하는 트레이닝 데이터 및 각 샘플이 워터마크 및 워터마킹된 음성 샘플(140)을 포함하지 않음을 표시하는 라벨 및 각 샘플이 워터마크를 포함한다고 표시하는 라벨을 사용할 수 있다. 트레이너(144)는 기계 학습을 사용하여 오디오 샘플을 수신하고, 오디오 샘플이 워터마크를 포함하는지 여부를 출력할 수 있도록 오디오 워터마크 식별 모델(148)을 생성한다.

[0039] 컴퓨팅 디바이스(106)는 오디오 워터마크 식별 모델(148)에 액세스하고, 수신된 오디오 데이터를 프로세싱하는데 사용하기 위해 컴퓨팅 디바이스(102)에 모델(128)을 제공할 수 있다. 컴퓨팅 디바이스(102)는 오디오 워터마

크 식별 모델(158)에 모델(128)을 저장할 수 있다.

[0040] 컴퓨팅 디바이스(106)는 재생 로그(142) 및 식별 로그(146)에 기초하여 오디오 워터마크 식별 모델(148)을 업데이트할 수 있다. 재생 로그(142)는 컴퓨팅 디바이스(104)로부터 수신되고 재생 로그(124)에 저장된 재생 데이터(126)와 같은 데이터를 포함할 수 있다. 재생 로그(142)는 워터마킹된 오디오를 출력한 다수의 컴퓨팅 디바이스로부터의 재생 데이터를 포함할 수 있다. 식별 로그(146)는 컴퓨팅 디바이스(102)로부터 수신되고 식별 로그(164)에 저장된 식별 데이터(130)와 같은 데이터를 포함할 수 있다. 식별 로그(146)는 오디오 워터마크를 식별하고 워터마킹된 오디오에 포함된 임의의 명령 또는 쿼리의 실행을 방지하도록 구성된 다수의 컴퓨팅 디바이스로부터의 추가 식별 데이터를 포함할 수 있다.

[0041] 트레이너(144)는 재생 로그(142)와 식별 로그(146)를 비교하여 컴퓨팅 디바이스가 워터마킹된 오디오를 출력했고 다른 컴퓨팅 디바이스가 워터마크된 오디오에서 워터마크를 식별했음을 나타내는 매칭 엔트리를 식별할 수 있다. 트레이너(144)는 또한 식별 로그(146) 및 재생 로그(142)에서 워터마크 식별 오류를 식별할 수 있다. 제1 유형의 워터마크 식별 오류는 식별 로그(146)가 컴퓨팅 디바이스가 워터마크를 식별함을 나타내지만 재생 로그(142)는 워터마크된 오디오의 출력을 나타내지 않는 경우 발생할 수 있다. 제2 유형의 워터마크 식별 오류는 재생 로그(142)가 워터마크된 오디오의 출력을 나타내지만, 식별 로그(146)는 워터마크된 오디오 주변의 컴퓨팅 디바이스가 워터마크를 식별하지 않았음을 나타내는 경우에 발생할 수 있다.

[0042] 트레이너(144)는 오류를 업데이트하고, 대응하는 오디오 데이터를 추가 트레이닝 데이터로 사용하여 오디오 워터마크 식별 모델(148)을 업데이트할 수 있다. 트레이너(144)는 컴퓨팅 디바이스가 워터마크를 적절하게 식별한 오디오를 사용하여 오디오 워터마크 식별 모델(148)을 업데이트할 수 있다. 트레이너(144)는 컴퓨팅 디바이스에 의해 출력된 오디오와 컴퓨팅 디바이스에 의해 검출된 오디오를 모두 트레이닝 데이터로 사용할 수 있다. 트레이너(144)는 재생 로그(142) 및 식별 로그(146)에 저장된 오디오 데이터와 기계 학습을 사용하여 오디오 워터마크 식별 모델(148)을 업데이트할 수 있다. 트레이너(144)는 재생 로그(142) 및 식별 로그(146)에서 제공된 워터마킹 라벨 및 위에서 설명한 오류 식별 기술로부터 수정된 라벨을 기계 학습 트레이닝 프로세스의 일부로 사용할 수 있다.

[0043] 일부 구현예에서, 컴퓨팅 디바이스(102) 및 여러 다른 컴퓨팅 디바이스는 서버에서 실행되는 서버 기반 핫워더 및/또는 서버 기반 음성 인식기에 의한 프로세싱을 위해 오디오(115)를 서버에 전송하도록 구성될 수 있다. 오디오 워터마크 식별자(152)는 오디오(115)가 오디오 워터마크를 포함하지 않음을 나타낼 수 있다. 그 결정에 기초하여, 컴퓨팅 디바이스(102)는 서버 기반 핫워더 및/또는 서버 기반 음성 인식기에 의한 추가 프로세싱을 위해 서버로 오디오를 전송할 수 있다. 여러 다른 컴퓨팅 디바이스의 오디오 워터마크 식별자는 또한 오디오(115)가 오디오 워터마크를 포함하지 않음을 나타낼 수 있다. 이들 결정에 기초하여, 각각의 다른 컴퓨팅 디바이스는 서버 기반 핫워더 및/또는 서버 기반 음성 인식기에 의한 추가 프로세싱을 위해 각각의 오디오를 서버에 전송할 수 있다. 서버는 각 컴퓨팅 디바이스로부터의 오디오가 핫워드를 포함하는지 여부를 결정하고 및/또는 오디오의 전사를 생성하고 결과를 각 컴퓨팅 디바이스로 다시 전송할 수 있다.

[0044] 일부 구현예에서, 서버는 워터마크 결정 각각에 대한 워터마크 신뢰 점수를 나타내는 데이터를 수신할 수 있다. 서버는 컴퓨팅 디바이스(102) 및 다른 컴퓨팅 디바이스에 의해 수신된 오디오가 컴퓨팅 디바이스(102) 및 다른 컴퓨팅 디바이스의 위치, 수신된 오디오의 특성, 유사한 시간에 각각의 오디오 부분을 수신 및 임의의 다른 유사한 표시자에 기초하여 동일한 소스로부터 온 것이라고 결정할 수 있다. 일부 예에서, 각 워터마크 신뢰 점수는 범위의 한쪽 끝에 있는 워터마크 신뢰 점수 임계치과 워터마크 신뢰 점수 임계치와의 백분율 차이일 수 있는 다른 신뢰 점수(예: 5% 더 작음)를 포함하는 특정한 범위 내에 있을 수 있다. 예를 들어, 범위는 0.80 ~ 0.76의 워터마크 신뢰 점수 임계치일 수 있다. 다른 예에서, 범위의 다른 쪽 끝은 0.05와 같은 워터마크 신뢰 점수 임계치로부터 고정된 거리일 수 있다. 예를 들어, 범위는 0.80 ~ 0.75의 워터마크 신뢰 점수 임계치일 수 있다.

[0045] 서버가 각 워터마크 신뢰 점수가 워터마크 신뢰 점수 임계치에 가깝지만 임계치를 만족하지 않는 범위 내에 있다고 결정하면, 서버는 워터마크 신뢰 점수 임계치를 조정해야 한다고 결정할 수 있다. 이 경우, 서버는 워터마크 신뢰 점수 임계치를 범위의 하단으로 조정할 수 있다. 일부 구현예에서, 서버는 워터마킹된 음성 샘플(140)에서 각각의 컴퓨팅 디바이스로부터 수신된 오디오를 포함함으로써 워터마킹된 음성 샘플(140)을 업데이트할 수 있다. 트레이너(144)는 기계 학습 및 업데이트된 워터마킹된 음성 샘플(140)을 사용하여 오디오 워터마크 식별 모델(148)을 업데이트할 수 있다.

[0046] 도 1은 위에서 설명된 상이한 기능을 수행하는 3개의 상이한 컴퓨팅 디바이스를 도시하며, 하나 이상의 컴퓨팅 디바이스의 임의의 조합은 기능의 임의의 조합을 수행할 수 있다. 예를 들어, 컴퓨팅 디바이스(102)는 오디오

워터마크 식별 모델(148)을 트레이닝하는 별도의 컴퓨팅 디바이스(106) 대신에 오디오 워터마크 식별 모델(148)을 트레이닝할 수 있다.

[0047] 도 2는 녹음된 매체에서 핫워드를 검출할 때 핫워드 트리거를 억제하기 위한 예시적 프로세스(200)를 도시한다. 일반적으로, 프로세스(200)는 오디오가 오디오 워터마크를 포함하는지 여부를 결정하기 위해 수신된 오디오를 프로세싱한다. 오디오가 오디오 워터마크를 포함하면, 프로세스(200)는 오디오의 추가 프로세싱을 억제할 수 있다. 오디오가 오디오 워터마크를 포함하지 않는 경우, 프로세스(200)는 오디오를 계속 프로세싱하고 오디오에 포함된 임의의 쿼리 또는 명령을 실행한다. 프로세스(200)는 하나 이상의 컴퓨터들 예를 들어, 도 1에 도시된 컴퓨팅 디바이스(102, 104 및/또는 106)를 포함하는 컴퓨터 시스템에 의해 수행되는 것으로 설명될 것이다.

[0048] 시스템은 발언의 재생에 대응하는 오디오 데이터를 수신한다(210). 예를 들어, 텔레비전이 광고를 재생하고 있고, 광고의 배우가 "Ok 컴퓨터, 조명을 켜"라고 말할 수 있다. 시스템은 마이크로폰을 포함하며, 마이크로폰은 배우의 발언을 포함한 광고의 오디오를 검출한다.

[0049] 시스템은 (i) 주어진 오디오 데이터 샘플이 오디오 워터마크를 포함하는지 여부를 결정하도록 구성되고, 그리고 (ii) 오디오 워터마크 샘플을 각각 포함하는 워터마킹된 오디오 데이터 샘플 및 오디오 워터마크 샘플을 각각 포함하지 않는 비-워터마킹된 오디오 데이터 샘플을 사용하여 트레이닝되었던 모델에 대한 입력으로서 상기 오디오 데이터를 제공한다(220). 일부 구현예에서, 시스템은 오디오 데이터가 핫워드를 포함한다고 결정할 수 있다. 핫워드 검출에 기초하여, 시스템은 오디오 데이터를 모델에 대한 입력으로 제공한다. 예를 들어, 시스템은 오디오 데이터가 "ok 컴퓨터"를 포함한다고 결정할 수 있다. "ok 컴퓨터" 검출에 기초하여, 시스템은 모델에 오디오 데이터를 제공한다. 시스템은 핫워드를 포함하는 오디오 데이터의 일부와 핫워드 이후에 수신된 오디오를 제공할 수 있다. 일부 경우에, 시스템은 핫워드 이전의 오디오 일부를 제공할 수 있다.

[0050] 일부 구현예에서, 시스템은 오디오 데이터가 핫워드를 포함하는지 여부를 결정하기 위해 오디오 데이터를 분석할 수 있다. 분석은 오디오 데이터를 모델에 입력으로 제공하기 전이나 후에 발생할 수 있다. 일부 구현예에서, 시스템은 오디오 워터마크를 각각 포함하는 워터마킹된 상기 오디오 데이터 샘플, 오디오 워터마크를 각각 포함하지 않는 상기 비-워터마킹된 오디오 데이터 샘플, 각각 워터마킹된 및 비-워터마킹된 오디오 샘플이 오디오 워터마크를 포함하는지 여부를 나타내는 데이터 및 기계 학습을 사용하여 모델을 트레이닝할 수 있다. 시스템은 모델에 대한 오디오 입력이 워터마크를 포함하는지 또는 워터마크를 포함하지 않는지를 나타내는 데이터를 출력하도록 모델을 트레이닝할 수 있다.

[0051] 일부 구현예에서, 서로 다른 워터마킹된 오디오 신호는 서로 다른 워터마크를 포함할 수 있다(하나의 오디오 샘플의 워터마크는 모두 서로 동일할 수 있지만 한 오디오 신호의 워터마크는 다른 오디오 신호의 워터마크와 다를 수 있다). 시스템은 오디오 신호의 왜곡을 최소화하기 위해 각 오디오 신호에 대해 다른 워터마크를 생성할 수 있다. 일부 구현예에서, 시스템은 오디오 신호에서 주기적 간격으로 워터마크를 배치할 수 있다. 예를 들어, 시스템은 200밀리 초마다 워터마크를 배치할 수 있다. 일부 구현예에서, 시스템은 핫워드 및/또는 핫워드 이전의 기간을 포함하는 오디오에 워터마크를 배치할 수 있다.

[0052] 시스템은 (i) 주어진 오디오 데이터 샘플이 오디오 워터마크를 포함하는지 여부를 결정하도록 구성되고, 그리고 (ii) 오디오 워터마크 샘플을 포함하는 워터마킹된 오디오 데이터 샘플 및 오디오 워터마크 샘플을 포함하지 않는 비-워터마킹된 오디오 데이터 샘플을 사용하여 트레이닝되었던 모델로부터, 상기 오디오 데이터가 상기 오디오 워터마크를 포함하는지 여부를 나타내는 데이터를 수신한다(230). 시스템은 오디오 데이터가 워터마크를 포함한다는 표시를 수신하거나 오디오 데이터가 워터마크를 포함하지 않는다는 표시를 수신할 수 있다.

[0053] 시스템은 오디오 데이터가 오디오 워터마크를 포함하는지 여부를 나타내는 데이터에 기초하여, 오디오 데이터의 프로세싱을 계속하거나 중단한다(240). 일부 구현예에서, 시스템은 오디오 데이터가 오디오 워터마크를 포함하는 경우 오디오 데이터의 프로세싱을 중단할 수 있다. 일부 구현예에서, 시스템은 오디오 데이터가 오디오 워터마크를 포함하지 않는 경우 오디오 데이터의 프로세싱을 계속할 수 있다. 일부 구현예에서, 오디오 데이터의 프로세싱은 오디오 데이터에 대해 음성 인식을 수행하는 것 및/또는 오디오 데이터가 핫워드를 포함하는지 여부를 결정하는 것을 포함할 수 있다. 일부 구현예에서, 프로세싱은 오디오 데이터에 포함된 쿼리 또는 명령을 실행하는 것을 포함할 수 있다.

[0054] 일부 구현예에서, 시스템은 시스템이 오디오 데이터를 수신한 시간 및 날짜를 기록한다. 시스템은 상기 시간 및 날짜를 오디오 데이터를 출력하는 컴퓨팅 디바이스로부터 수신된 시간 및 날짜와 비교할 수 있다. 시스템이 오디오 데이터를 수신한 날짜와 시간이 오디오 데이터를 출력한 날짜와 시간과 일치한다고 결정하면, 시스템은 오

디오 데이터를 추가 트레이닝 데이터로 사용하여 모델을 업데이트할 수 있다. 시스템은 오디오 데이터가 워터마크를 포함하는지 여부를 결정할 때 모델이 올바른지 여부를 식별하고, 오디오 데이터가 트레이닝 데이터에 추가될 때 올바른 워터마크 라벨을 포함하는지 확인할 수 있다.

[0055] 구체적으로, 사용자를 위한 작업을 수행할 수 있는 소프트웨어 에이전트를 일반적으로 "가상 어시스턴트"라고 지칭한다. 예를 들어, 가상 어시스턴트는 사용자의 음성 입력에 의해 작동될 수 있다. 예를 들어, 사용자가 말할 때 가상 어시스턴트가 활성화되게 하고, 발화된 트리거 단어와 연관된 작업을 수행하게 하는 하나 이상의 트리거 단어를 인식하도록 프로그래밍될 수 있다. 이러한 트리거 단어를 종종 "핫워드"라고 한다. 가상 어시스턴트는 예를 들어 사용자의 컴퓨터 모바일 전화 또는 기타 사용자 디바이스에 제공될 수 있다. 대안적으로, 가상 어시스턴트는 소위 "스마트 스피커"(하나 이상의 핫워드의 도움으로 핸즈프리 활성화 및 인터렉션을 제공하는 통합 가상 어시스턴트가 있는 무선 스피커 유형)와 같은 다른 디바이스에 통합될 수 있다.

[0056] 스마트 스피커의 광범위한 채택으로 추가 문제가 발생한다. 1억 명 이상의 시청자를 끌어들이는 스포츠 이벤트와 같이 많은 청중이 있는 이벤트에서 핫워드가 포함된 광고는 가상 어시스턴트의 동시 트리거로 이어질 수 있다. 시청률이 높기 때문에 음성 인식 서버에 대한 동시 쿼리가 크게 증가하여 서비스 거부(DOS)로 이어질 수 있다.

[0057] 거짓 핫워드 필터링을 위한 두 가지 가능한 메커니즘은 (1) 쿼리 오디오로부터의 핑거프린트가 거짓 트리거를 필터링하기 위해 광고와 같은 알려진 오디오로부터의 핑거프린트의 데이터베이스에 대해 검사되는 오디오 핑거프린팅 (2) 오디오가 퍼블리셔에 의해 워터마킹되고, 가상 어시스턴트에 의해 기록된 쿼리가 필터링을 위해 워터마크가 검사되는 오디오 워터마킹에 기초하는 것이다.

[0058] 본 개시는 컨볼루션 신경 네트워크을 사용하는 저-지연, 작은 풋프린트 워터마크 검출기의 설계를 설명한다. 이 워터마크 검출기는 관심 시나리오에서 빈번하게 발생할 수 있는 소음 및 반향 환경에 견고하도록 트레이닝된다.

[0059] 오디오 워터마킹은 저작권 보호 및 제2 화면 애플리케이션에서 사용될 수 있다. 저작권 보호에서 워터마크 검출은 일반적으로 전체 오디오 신호를 검출할 수 있으므로 지연 시간에 민감할 필요가 없다. 제2 화면 애플리케이션의 경우, 높은 지연 시간 워터마크 검출로 인한 지연이 허용될 수 있다. 이 두 가지 시나리오와 달리 가상 어시스턴트의 워터마크 검출은 지연 시간에 매우 민감하다.

[0060] 워터마크 검출을 포함하는 알려진 애플리케이션에서, 워터마크를 구성하는 내장 메시지는 일반적으로 미리 알려지지 않았으며, 워터마크 검출기는 메시지 시퀀스가 워터마크를 포함하는지 여부를 결정하기 전에 메시지 시퀀스를 디코딩해야 하고, 포함한다면 워터마크를 결정해야 한다. 그러나, 본 명세서에 설명된 일부 애플리케이션에서, 워터마크 검출기는 디코더/워터마크 검출기에 의해 정확히 알려진 워터마크 패턴을 검출할 수 있다. 즉, 재녹음된 음성 콘텐츠의 퍼블리셔 또는 제공자는 이를 워터마크로 워터마킹할 수 있으며, 예를 들어 가상 어시스턴트의 제공자 및/또는 가상 어시스턴트를 포함하는 디바이스의 제공자에게 워터마크의 세부 사항을 제공할 수 있다. 유사하게, 가상 어시스턴트의 제공자는 가상 어시스턴트로부터 출력된 음성에 워터마크가 제공되도록 구성하고, 워터마크의 세부 사항을 이용 가능하게 할 수 있다. 그 결과, 수신된 메시지에서 워터마크가 검출되면, 수신된 메시지가 사용자로부터 실시간 음성 입력이 아니라는 것을 알 수 있으며, 수신된 메시지에서 핫워드로 인한 가상 어시스턴트의 활성화를 억제할 수 있고, 전체 메시지가 수신 및 프로세싱될 때까지 기다릴 필요가 없다. 이렇게하면 지연 시간이 줄어든다.

[0061] 핫워드 억제를 위한 일부 구현예는 오디오 핑거프린팅 접근법을 사용한다. 이 접근법은 알려진 오디오의 핑거프린트 데이터베이스가 필요하다. 디바이스에서 이 데이터베이스를 유지 관리하는 것이 사소한 일이 아니므로 이러한 솔루션을 온-디바이스 배포는 불가능하다. 그러나 오디오 핑거프린팅 접근법의 중요한 장점은 오디오 퍼블리싱 프로세스를 수정할 필요가 없다는 것이다. 따라서 오디오 퍼블리셔가 공동 작업자가 아닌 적대적인 시나리오도 다룰 수 있다.

[0062] 본 개시는 워터마크 기반 핫워드 억제 메커니즘을 설명한다. 핫워드 억제 메커니즘은 메모리 및 계산 풋프린트의 설계 제약을 가져오는 온-디바이스 배포를 사용할 수 있다. 또한 사용자 경험에 영향을 주지 않도록 지연 시간에 대한 제약이 있다.

[0063] 워터마크 기반 접근법은 워터마크를 추가하기 위해 오디오 퍼블리싱 프로세스를 수정해야 할 수 있다. 따라서 때로는 공동 작업자가 퍼블리싱한 오디오를 검출하는데만 사용할 수 있다. 그러나 핑거프린트 데이터베이스의 유지 관리가 필요하지 않을 수 있다. 이 구성은 몇 가지 이점을 제공한다.

[0064] 제1 이점은 온-디바이스 배포의 가능성이다. 이는 여러 가상 어시스턴트가 동시에 트리거될 수 있는 높은 시청

률 이벤트 중에 이점이 될 수 있다. 이러한 잘못된 트리거를 검출하기 위한 서버 기반 솔루션은 동시 트리거의 규모로 인해 서비스 거부로 이어질 수 있다. 제2 이점은 퍼블리셔가 공동 작업할 수 있지만 오디오가 미리 알려지지 않은 TTS(text-to-speech) 합성기 출력과 같이 공동 작업자의 의해 퍼블리시된 알려지지 않은 오디오를 검출하는 것이다. 제3 이점은 확장성일 수 있다. 온라인 플랫폼의 오디오/비디오 퍼블리셔와 같은 엔터티는 가상 어시스턴트를 트리거하지 않도록 오디오에 워터마크를 표시할 수 있다. 일부 구현예에서, 이러한 플랫폼은 오디오 핑거프린팅 기반 접근법을 사용하여 실제로 처리할 수 없는 수백만 시간의 콘텐츠를 호스팅한다.

[0065] 일부 구현예에서, 본 명세서에 설명된 워터마크 기반 접근법은 적대적 에이전트를 다루는 능력을 가질 수 있는 오디오 핑거프린팅 기반 접근법과 결합될 수 있다.

[0066] 아래 설명은 워터마크 임베더 및 워터마크 검출기에 대해 설명한다.

[0067] 워터마크 임베더는 FFT 도메인에서 확산 스펙트럼 기반 워터마킹에 기초할 수 있다. 워터마크 임베더는 워터마크 신호의 진폭을 형성하는데 사용되는 최소 마스킹 임계치(MMT)를 추정하기 위해 심리 음향 모델을 사용할 수 있다.

[0068] 이 기법을 요약하면, 워터마크 추가가 가능한 호스트 신호의 영역은 최소 에너지 기준에 기초하여 선택된다. 이산 푸리에 변환(DFT) 계수는 이러한 영역의 모든 호스트 신호 프레임(25ms창-12.5ms홉)에 대해 추정된다. 이러한 DFT 계수는 심리 음향 모델을 사용하여 최소 마스킹 임계치(MMT)를 추정하는데 사용된다. MMT는 워터마크 신호의 프레임에 대한 크기 스펙트럼을 형성하는데 사용된다. 도 3은 추정된 MMT를 호스트 신호 에너지 및 청력의 절대 임계치와 함께 제공한다. 호스트 신호의 위상은 워터마크 신호에 사용될 수 있으며, DFT 계수의 부호는 메시지 페이로드로부터 결정된다. 메시지 비트 페이로드는 다수의 스크램블링을 사용하여 프레임 청크에 분산될 수 있다. 일부 구현예에서, 시스템은 쿼리가 워터마킹되었는지 여부를 검출할 수 있고, 어떠한 페이로드도 전송할 필요가 없을 수 있다. 따라서, 시스템은 프레임 청크(예를 들어, 16 프레임 또는 200ms)에 걸쳐 부호 행렬을 무작위로 선택하고, 워터마킹 영역에 걸쳐 이 부호 행렬을 반복할 수 있다. 이러한 부호 행렬의 반복은 워터마크 검출기 출력을 후-프로세싱하고 검출 성능을 향상시키기 위해 이용될 수 있다. 개별 워터마크 프레임의 오버랩 추가는 워터마크 신호를 생성할 수 있다. 도 2의 서브 플롯 (a) 및 (b)는 호스트 신호와 워터마크 신호의 크기 스펙트럼을 나타내고, 서브 플롯(c)은 부호 행렬을 나타낸다. 수직선은 행렬의 두 복제 사이의 경계를 나타낸다.

[0069] 워터마크 신호는 워터마크의 비가청성을 더욱 보장하기 위해, 팩터(예: $\alpha \in [0, 1]$)로 스케일링한 후 시간 도메인에서 호스트 신호에 추가될 수 있다. 일부 구현예에서, α는 PEAQ(Perceptual Evaluation of Audio Quality)와 같은 객관적인 평가 메트릭을 사용하여 반복적으로 결정된다. 일부 구현예에서, 시스템은 보수적 스케일링 팩터(예를 들어, $\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$)를 사용하고, 이러한 스케일링 팩터 각각에서 검출 성능을 평가할 수 있다.

[0070] 일부 구현예에서, 워터마크 검출기에 대한 설계 요구 사항은 모델의 메모리 풋프린트와 계산 복잡성 모두에 상당한 제약을 두는 온-디바이스 배포일 수 있다. 아래 설명은 온-디바이스 키워드 검출을 위한 컨벌루션 신경 네트워크 기반 모델 아키텍처를 설명한다. 일부 구현예에서, 시스템은 시간적 컨볼루션 신경 네트워크를 사용할 수 있다.

[0071] 일부 구현예에서, 신경 네트워크는 200ms 패턴의 한 인스턴스와 동일한 200ms 패턴의 복제일 수 있는 임베디드 워터마크 부호 행렬(도 4, 서브 플롯(c))의 상호-상관을 추정하도록 트레이닝된다. 도 4의 서브 플롯(d)는 크로스-상관을 보여준다. 상호 상관은 각 부호 행렬 블록의 시작에 대한 정보를 인코딩할 수 있으며, 호스트 신호 내에서 워터마크 신호의 전체 기간 동안 0이 아닐 수 있다.

[0072] 시스템은 다중 작업 손실 함수를 사용하여 신경 네트워크을 트레이닝할 수 있다. 주요 작업은 GT 상호 상관의 추정일 수 있으며, 보조 작업은 에너지 섭동 패턴 및/또는 워터마크 크기 스펙트럼의 추정일 수 있다. 평균 제곱 오차는 GT(들)과 네트워크 출력(들) 사이에서 계산될 수 있다. 정규화 상수를 사용하여 보조 손실을 스케일링한 후 손실의 일부 또는 전부를 보간할 수 있다. 일부 구현예에서, 대응하는 GT의 동적 범위를 커버하기 위해 각 네트워크 출력을 경계화하면 성능이 향상될 수 있다.

[0073] 일부 구현예에서, 시스템은 네트워크 출력을 후 프로세싱할 수 있다. 일부 구현예에서, 워터마크는 페이로드 메시지를 갖지 않을 수 있으며, 단일 부호 매트릭스는 워터마킹 영역 전체에 복제된다. 이것은 주기적인 상호 상관 패턴을 초래할 수 있다(도 4, 서브 플롯(d)). 이 양태는 네트워크 출력에서 거짓 피크를 제거하는데 이용될 수 있다. 일부 구현예에서, 성능을 개선하기 위해, 시스템은 관심 주파수를 분리하는 대역 통과 필터를 통해 상

호 상관 패턴을 복제함으로써 생성된 매치 필터를 사용할 수 있다(도 5 참조). 도 7은 매치-필터링 전후에 비-워터마킹된 신호에 대해 생성된 네트워크 출력을 비교한다. 일부 구현예에서, 주기성을 갖지 않는 거짓된 피크는 상당히 억제될 수 있다. GT(Ground Truth)(705)는 대략 0.0(예를 들어, -0.01과 0.01 사이)일 수 있고, 네트워크 출력(710) 및 매치 필터링된 네트워크 출력(720)보다 x-축을 더 가깝게 추적할 수 있다. 네트워크 출력(710)은 GT(705) 및 매치 필터링된 네트워크 출력(720)보다 x-축에 대해 더 다양할 수 있다. 매치 필터링된 네트워크 출력(720)은 네트워크 출력(710)보다 x-축을 더 가깝게 추적할 수 있고, GT(705)만큼 가깝게 x-축을 추적하지 않을 수 있다. 매치 필터링된 네트워크 출력(720)은 네트워크 출력(710)보다 더 부드러울 수 있다. 매치 필터링된 네트워크 출력(720)은 네트워크 출력(710)보다 더 작은 범위 내에 남아 있을 수 있다. 예를 들어, 매치 필터링된 네트워크 출력(720)은 -0.15와 0.15 사이에 머물 수 있다. 네트워크 출력(710)은 -0.30과 0.60 사이에 머물 수 있다.

[0074] 신경 네트워크가 트레이닝되면, 신경 네트워크를 구현하는 모델을 오디오 데이터 샘플에 적용함으로써, 주어진 오디오 데이터 샘플에 오디오 워터마크가 포함되어 있는지 여부를 결정하는 방법에 사용될 수 있다. 방법은 오디오 데이터가 오디오 워터마크를 포함할 가능성을 반영하는 신뢰 점수를 결정하는 단계; 오디오 데이터가 오디오 워터마크를 포함할 가능성을 반영하는 신뢰 점수를 신뢰 점수 임계치와 비교하는 단계; 및 오디오 데이터가 오디오 워터마크를 포함할 가능성을 반영하는 신뢰 점수를 신뢰 점수 임계치와 비교함에 기초하여, 오디오 데이터에 대해 추가 프로세싱을 수행할지 여부를 결정하는 단계를 포함한다.

[0075] 일부 실시예에서 상기 방법은: 오디오 데이터가 오디오 워터마크를 포함할 가능성을 반영하는 신뢰 점수를 신뢰 점수 임계치와 비교함에 기초하여, 신뢰 점수가 신뢰 점수 임계치를 만족한다고 결정하는 단계를 포함하며, 상기 오디오 데이터에 대해 추가 프로세싱을 수행할지 여부를 결정하는 단계는 상기 오디오 데이터에 대해 추가 프로세싱의 수행을 억제하기로 결정하는 것을 포함한다. 일부 실시예에서 상기 방법은: 발언이 오디오 워터마크를 포함할 가능성을 반영하는 신뢰 점수를 신뢰 점수 임계치와 비교함에 기초하여, 신뢰 점수가 신뢰 점수 임계치를 만족하지 않는다고 결정하는 단계를 포함하며, 상기 오디오 데이터에 대해 추가 프로세싱을 수행할지 여부를 결정하는 단계는 상기 오디오 데이터에 대해 추가 프로세싱의 수행하기로 결정하는 것을 포함한다. 일부 실시예에서 상기 방법은: 사용자로부터, 상기 오디오 데이터에 대한 추가 프로세싱의 수행을 확인하는 데이터를 수신하는 단계; 및 상기 오디오 데이터에 대한 추가 프로세싱의 수행을 확인하는 데이터를 수신함에 기초하여, 상기 모델을 업데이트하는 단계를 포함한다. 일 실시예에서, 오디오 데이터에 대한 추가 프로세싱은 오디오 데이터의 전사에 기초한 액션을 수행하는 단계; 또는 상기 오디오 데이터가 특정한, 미리 정의된 핫워드를 포함하는지 여부를 결정하는 단계를 포함한다. 일부 실시예에서 상기 방법은: (i) 주어진 오디오 데이터 샘플이 오디오 워터마크를 포함하는지 여부를 결정하도록 구성되고, 그리고 (ii) 오디오 워터마크 샘플을 포함하는 워터마킹된 오디오 데이터 샘플 및 오디오 워터마크 샘플을 포함하지 않는 비-워터마킹된 오디오 데이터 샘플을 사용하여 트레이닝되었던 모델에, 상기 오디오 데이터를 적용하기 전에, 상기 오디오 데이터가 특정한, 미리 정의된 핫워드를 포함한다고 결정하는 단계를 포함한다. 일부 실시예에서 상기 방법은: 상기 오디오 데이터가 특정한, 미리 정의된 핫워드를 포함한다고 결정하는 단계를 포함하며, (i) 주어진 오디오 데이터 샘플이 오디오 워터마크를 포함하는지 여부를 결정하도록 구성되고, 그리고 (ii) 오디오 워터마크 샘플을 포함하는 워터마킹된 오디오 데이터 샘플 및 오디오 워터마크 샘플을 포함하지 않는 비-워터마킹된 오디오 데이터 샘플을 사용하여 트레이닝되었던 모델에, 상기 오디오 데이터를 적용하는 것은, 상기 오디오 데이터가 특정한, 미리 정의된 핫워드를 포함한다고 결정하는 것에 대한 응답으로 수행된다. 일부 실시예에서 상기 방법은: 오디오 워터마크를 포함하는 워터마킹된 오디오 데이터 샘플 및 오디오 워터마크를 포함하지 않는 비-워터마킹된 오디오 데이터 샘플을 수신하는 단계; 및 기계 학습을 사용하여, 오디오 워터마크를 포함하는 워터마킹된 오디오 데이터 샘플 및 오디오 워터마크를 포함하지 않는 비-워터마킹된 오디오 데이터 샘플을 사용하여 모델을 트레이닝하는 단계를 포함한다. 일 실시예에서, 방법은: 워터마킹된 오디오 데이터 샘플의 적어도 일부는 다수의 주기적 위치에서 오디오 워터마크를 포함한다.

[0076] 도 8은 본 명세서에 기술된 기법들을 구현하는데 사용될 수 있는 컴퓨터 디바이스(800)와 모바일 컴퓨터 디바이스(850)의 예시를 도시한다. 컴퓨팅 디바이스(800)는 랩톱, 데스크톱, 워크 스테이션, 개인 휴대 정보 단말기, 서버, 블레이드 서버, 메인 프레임 및 다른 적절한 컴퓨터와 같은 다양한 형태의 디지털 컴퓨터들을 나타내기 위한 것이다. 모바일 컴퓨팅 디바이스(850)는 개인 휴대 정보 단말기, 셀룰러 전화기, 스마트폰 및 다른 유사한 컴퓨팅 디바이스들과 같은 다양한 형태의 모바일 디바이스들을 나타내기 위한 것이다. 여기에 도시된 컴포넌트들, 그들의 연결 및 관계, 및 그들의 기능은 단지 예시적인 것을 의미하며, 제한하는 것을 의미하지는 않는다.

[0077] 컴퓨팅 디바이스(800)는 프로세서(802), 메모리(804), 저장 디바이스(806), 메모리(804) 및 다수의 고속 확장

포트(810)에 연결되는 고속 인터페이스(808) 및 저속 확장 포트(814) 및 저장 디바이스(806)에 연결되는 저속 인터페이스(812)를 포함한다. 프로세서(802), 메모리(804), 저장 디바이스(806), 고속 인터페이스(808), 고속 확장 포트(810) 및 저속 인터페이스(812) 각각은 다양한 버스들을 사용하여 상호연결되며, 공통 마더보드에 또는 적절한 다른 방식으로 장착된다. 프로세서(802)는 메모리(804) 또는 저장 디바이스(806)에 저장된 명령어들을 포함하는, 컴퓨팅 디바이스(800) 내에서 실행하기 위한 명령어들을 프로세싱하여, 고속 인터페이스(808)에 연결된 디스플레이(816)와 같은 외부 입/출력 디바이스상에 GUI에 대한 그래픽 정보를 디스플레이할 수 있다. 다른 구현예에서, 다수의 프로세서들 및/또는 다수의 버스들이 다수의 메모리들 및 다수의 유형의 메모리와 함께, 적절하게 사용될 수 있다. 또한, 다수의 컴퓨팅 디바이스들은 필요한 동작의 부분들을 제공하는 각 디바이스와 연결될 수 있다(예를 들어, 서버 뱅크, 블레이드 서버 그룹 또는 멀티 프로세서 시스템).

[0078] 메모리(804)는 컴퓨팅 디바이스(800) 내에 정보를 저장한다. 일부 구현예에서, 메모리(804)는 휘발성 메모리 유닛 또는 유닛들이다. 일부 구현예에서, 메모리(804)는 비휘발성 메모리 유닛 또는 유닛들이다. 또한, 메모리(804)는 자기 또는 광학 디스크와 같은 컴퓨터 판독가능 매체의 다른 형태일 수 있다.

[0079] 저장 디바이스(806)는 컴퓨팅 디바이스(800)에 대한 대형 스토리지를 제공할 수 있다. 일부 구현예에서, 저장 디바이스(806)는 플로피 디스크 디바이스, 하드 디스크 디바이스, 광 디스크 디바이스 또는 테이프 디바이스, 플래시 메모리 또는 다른 유사한 고체 상태 메모리 디바이스, 또는 저장 영역 네트워크 또는 다른 구성의 디바이스를 포함하는 디바이스의 어레이와 같은 컴퓨터 판독가능 매체이거나 컴퓨터 판독가능 매체를 포함할 수 있다. 명령어는 정보 캐리어에 저장될 수 있다. 명령어는 하나 이상의 프로세싱 디바이스(예를 들어, 프로세서(802))에 의해 실행될 때, 상술한 것과 같은 하나 이상의 방법을 수행한다. 명령어는 또한 컴퓨터 또는 기계 판독가능 매체(예를 들어, 메모리(804), 저장 장치(806) 또는 프로세서(802)상의 메모리)와 같은 하나 이상의 저장 디바이스에 의해 저장될 수 있다.

[0080] 고속 인터페이스(808)는 컴퓨팅 디바이스(800)에 대한 대역폭 집중 동작들을 관리하는 반면, 저속 인터페이스(812)는 낮은 대역폭 집중 동작들을 관리한다. 이러한 기능들의 할당은 단지 예시적인 것이다. 일부 구현예에서, 고속 인터페이스(808)는 메모리(804), 디스플레이(816)(예를 들어, 그래픽 프로세서 또는 가속기를 통해) 및 다양한 확장 카드를 수용할 수 있는 고속 확장 포트(810)에 연결된다. 구현예에서, 저속 인터페이스(812)는 저장 디바이스(806) 및 저속 확장 포트(814)에 연결된다. 다양한 통신 포트(예를 들어, USB, 블루투스, 이더넷, 무선 이더넷)를 포함할 수 있는 저속 확장 포트(814)는 키보드, 포인팅 디바이스, 스캐너와 같은 하나 이상의 입력/출력 디바이스 또는 예를 들어 네트워크 어댑터를 통해 스위치 또는 라우터와 같은 네트워킹 디바이스에 결합될 수 있다.

[0081] 컴퓨팅 디바이스(800)는 도면에 도시된 바와 같이 다수의 상이한 형태로 구현될 수 있다. 예를 들어, 그것은 표준 서버(820)로서 또는 그러한 서버들의 그룹에서 다수로 구현될 수 있다. 또한, 랩톱 컴퓨터(822)와 같은 퍼스널 컴퓨터에서 구현될 수 있다. 또한, 랙 서버 시스템(824)의 일부로서 구현될 수 있다. 대안적으로, 컴퓨팅 디바이스(800)로부터의 컴포넌트들은 모바일 컴퓨팅 디바이스(850)와 같은 모바일 디바이스 내의 다른 컴포넌트들과 결합될 수 있다. 상기 디바이스들 각각은 컴퓨팅 디바이스(800), 모바일 컴퓨팅 디바이스(850) 중 하나 이상을 포함할 수 있고, 전체 시스템은 서로 통신하는 다수의 컴퓨팅 디바이스들로 구성될 수 있다.

[0082] 모바일 컴퓨팅 디바이스(850)는 다른 여러 컴포넌트들 중에서도 특히, 프로세서(852), 메모리(864), 디스플레이(854)와 같은 입/출력 디바이스, 통신 인터페이스(866) 및 송수신기(868)를 포함한다. 모바일 디바이스(850)에는 또한 추가적 저장을 제공하기 위해 마이크로 드라이브 또는 다른 디바이스와 같은 저장 디바이스가 제공될 수 있다. 프로세서(852), 메모리(864), 디스플레이(854), 통신 인터페이스(866) 및 트랜스시버(868) 각각은 다양한 버스들을 사용하여 상호 연결되고, 몇몇 컴포넌트들은 공통 마더 보드 상에 또는 적절한 다른 방식으로 장착될 수 있다.

[0083] 프로세서(852)는 메모리(864)에 저장된 명령어들을 포함하는 모바일 컴퓨팅 디바이스(850) 내의 명령어들을 실행할 수 있다. 프로세서(852)는 별개의 그리고 다수의 아날로그 및 디지털 프로세서들을 포함하는 칩들의 칩셋으로서 구현될 수 있다. 프로세서(852)는 예를 들어 사용자 인터페이스들, 모바일 컴퓨팅 디바이스(850)에 의해 실행되는 애플리케이션 및 모바일 컴퓨팅 디바이스(850)에 의한 무선 통신과 같은 모바일 컴퓨팅 디바이스(850)의 다른 컴포넌트들의 조정을 제공할 수 있다.

[0084] 프로세서(852)는 제어 인터페이스(858) 및 디스플레이(854)에 연결된 디스플레이 인터페이스(856)를 통해 사용자와 통신할 수 있다. 디스플레이(854)는 예를 들어, TFT LCD(박막 트랜지스터 액정 디스플레이) 또는 OLED(유기 발광 다이오드) 디스플레이 또는 다른 적절한 디스플레이 기술을 포함할 수 있다. 디스플레이 인터페이스

(856)는 사용자에게 그래픽 및 다른 정보를 제공하기 위해 디스플레이(854)를 구동하기 위한 적절한 회로를 포함할 수 있다. 제어 인터페이스(858)는 사용자로부터 명령을 수신하고, 프로세서(852)에 제출하기 위해 그들을 변환할 수 있다. 추가로, 외부 인터페이스(862)는 프로세서(852)와의 통신에 제공되어 다른 디바이스들과 모바일 컴퓨팅 디바이스(850)의 근거리 통신을 가능하게할 수 있다. 외부 인터페이스(862)는 예를 들면, 일부 구현예들에서는 유선 통신을 위해 또는 다른 구현예들에서는 무선 통신을 위해 제공될 수 있고, 다수의 인터페이스들도 사용될 수 있다.

[0085] 메모리(864)는 모바일 컴퓨팅 디바이스(850) 내에 정보를 저장한다. 메모리(864)는 컴퓨터 판독가능 매체 또는 매체들, 휘발성 메모리 유닛 또는 유닛들, 비휘발성 메모리 유닛 또는 유닛들 중 하나 이상으로 구현될 수 있다. 또한 확장 메모리(874)는 예를 들어 SIMM(Single In Line Memory Module) 카드 인터페이스를 포함할 수 있는 확장 인터페이스(872)를 통해 모바일 컴퓨팅 디바이스(850)에 제공되고 접속될 수 있다. 상기 확장 메모리(874)는 모바일 컴퓨팅 디바이스(850)에 대해 추가 저장 공간을 제공할 수 있거나, 또는 애플리케이션들 또는 모바일 컴퓨팅 디바이스(850)에 대한 다른 정보를 저장할 수 있다. 특히, 확장 메모리(874)는 전술한 프로세스들을 수행하거나 보충하는 명령어들을 포함할 수 있으며, 또한 보안 정보를 포함할 수 있다. 따라서, 예를 들면, 확장 메모리(874)는 모바일 컴퓨팅 디바이스(850)에 대한 보안 모듈로서 제공될 수 있고, 모바일 컴퓨팅 디바이스(850)의 보안 사용을 허용하는 명령어들로 프로그래밍될 수 있다. 또한, 보안 어플리케이션들은 SIMM 카드 상의 식별 정보를 해킹할 수 없는 방식으로 배치하는 것과 같이, SIMM 카드들을 통해 추가 정보와 함께 제공될 수 있다.

[0086] 메모리는 예를 들어, 후술되는 바와 같이, 플래시 메모리 및/또는 NVRAM 메모리((non-volatile random access memory))를 포함할 수 있다. 일부 구현예들에서, 명령어들은 정보 캐리어에 저장되어, 하나 이상의 프로세싱 디바이스들(예를 들어, 프로세서(852))에 의해 실행될 때, 상술한 것과 같은 하나 이상의 방법들을 수행한다. 명령어들은 또한 하나 이상의 컴퓨터 또는 기계 판독가능 매체(예를 들어, 메모리(864), 확장 메모리(874) 또는 프로세서(852)상의 메모리)와 같은 하나 이상의 저장 디바이스들에 의해 저장될 수 있다. 일부 구현예에서, 명령어들은 예를 들어 트랜스시버(868) 또는 외부 인터페이스(862)를 통해 전파된 신호로 수신될 수 있다.

[0087] 모바일 컴퓨팅 디바이스(850)는 필요에 따라 디지털 신호 프로세싱 회로를 포함할 수 있는 통신 인터페이스(866)를 통해 무선으로 통신할 수 있다. 통신 인터페이스(866)는 GSM 음성 호출(이동 통신용 글로벌 시스템), SMS(단문 메시지 서비스), EMS(확장 메시징 서비스) 또는 MMS 메시징(멀티미디어 메시징 서비스), CDMA(코드 분할 다중 접속), TDMA(시분할 다중 접속), PDC(개인 디지털 셀룰러), WCDMA(광대역 코드 분할 다중 접속), CDMA2000 또는 GPRS(일반 패킷 무선 서비스)와 같은 다양한 모드 또는 프로토콜 하에서 통신을 제공할 수 있다. 이러한 통신은 예를 들어, 무선 주파수를 사용하는 트랜스시버(868)를 통해 발생될 수 있다. 추가로, 블루투스, Wi-Fi 또는 다른 트랜시버를 사용하는 것과 같은 단거리 통신이 발생될 수 있다. 추가로, GPS(Global Positioning System) 수신기 모듈(870)은 모바일 컴퓨팅 디바이스(850)상에서 실행되는 애플리케이션들에 의해 적절히 사용될 수 있는 추가적인 네비게이션 및 위치 관련 무선 데이터를 모바일 컴퓨팅 디바이스(850)에 제공할 수 있다.

[0088] 모바일 컴퓨팅 디바이스(850)는 또한 사용자로부터 발화된 정보를 수신하고 그것을 이용가능한 디지털 정보로 변환할 수 있는 오디오 코덱(860)을 사용하여 청각적으로 통신할 수 있다. 오디오 코덱(860)은 마찬가지로, 예를 들어 모바일 컴퓨팅 디바이스(850)의 핸드셋 내의 스피커를 통하는 것과 같이, 사용자를 위한 가청 사운드를 생성할 수 있다. 이러한 사운드는 보이스 전화 콜들로부터의 사운드 포함할 수 있고, 기록된 사운드(예를 들어, 음성 메시지, 음악 파일 등)를 포함할 수 있고, 또한 모바일 컴퓨팅 디바이스(850)상에서 동작하는 애플리케이션들에 의해 생성된 사운드를 포함할 수 있다.

[0089] 모바일 컴퓨팅 디바이스(850)는 도면에 도시된 바와 같이 다수의 상이한 형태로 구현될 수 있다. 예를 들면, 그것은 셀룰러 전화(880)로서 구현될 수 있다. 또한 스마트폰(882), 개인 휴대 정보 단말기(PDA), 또는 다른 유사한 모바일 디바이스의 일부로서 구현될 수 있다.

[0090] 본 명세서에 기술된 시스템들 및 기법들의 다양한 구현예들은 디지털 전자 회로, 집적 회로, 특수하게 설계된 ASIC들(application specific integrated circuits), 컴퓨터 하드웨어, 펌웨어, 소프트웨어 및/또는 이들의 조합으로 구현될 수 있다. 이들 다양한 구현예들은 적어도 하나의 프로그래머블 프로세서를 포함하는 프로그래머블 시스템 상에서 실행가능하고 및/또는 인터프리트가능한 하나 이상의 컴퓨터 프로그램들에서의 구현예를 포함할 수 있고, 이는 전용 또는 범용일 수 있고, 저장 시스템, 적어도 하나의 입력 디바이스 및 적어도 하나의 출력 디바이스로부터 데이터 및 명령어들을 수신하고 그에 데이터 및 명령어들을 전송하기 위해 연결될 수 있다.

[0091] 이들 컴퓨터 프로그램들(프로그램, 소프트웨어, 소프트웨어 애플리케이션 또는 코드로도 알려짐)은 프로그래머블 프로세서에 대한 기계 명령어들을 포함하며, 하이레벨 절차어 및/또는 객체 지향 프로그래밍 언어 및/또는 어셈블리/기계어에서 구현될 수 있다. 본 명세서에서 사용된 바와 같이, 용어 “기계 판독가능 매체”, “컴퓨터 판독가능 매체”는 기계 판독가능 신호로서 기계 명령어들을 수신하는 기계 판독가능 매체를 포함하여, 기계 명령어들 및/또는 데이터를 프로그래머블 프로세서에 제공하는데 사용되는 임의의 컴퓨터 프로그램 물, 장치 및/또는 디바이스 예를 들어, 자기 디스크, 광학 디스크, 메모리, 프로그래머블 로직 디바이스(PLD)를 지칭한다. 용어 "기계 판독가능 신호"는 기계 명령어 및/또는 데이터를 프로그래머블 프로세서에 제공하는데 사용되는 임의의 신호를 지칭한다.

[0092] 사용자와의 인터렉션을 제공하기 위해, 본 명세서에서 기술된 시스템들 및 기법들은 사용자에게 정보를 디스플레이하기 위해 예를 들어, CRT(cathode ray tube) 또는 LCD(liquid crystal display) 모니터와 같은 디스플레이 디바이스 및 사용자가 컴퓨터에 입력을 제공할 수 있는 키보드 및 포인팅 디바이스 예를 들어, 마우스 또는 트랙볼을 갖는 컴퓨터에서 구현될 수 있다. 다른 종류의 디바이스들도 사용자와의 인터렉션을 제공하는데 사용될 수 있다. 예를 들어, 사용자에게 제공되는 피드백은 시각 피드백, 청각 피드백 또는 촉각 피드백과 같은 임의의 형태의 감각적 피드백일 수 있고, 사용자로부터의 입력은 음향, 음성 또는 촉각 입력을 포함하는 임의의 형태로 수신될 수 있다.

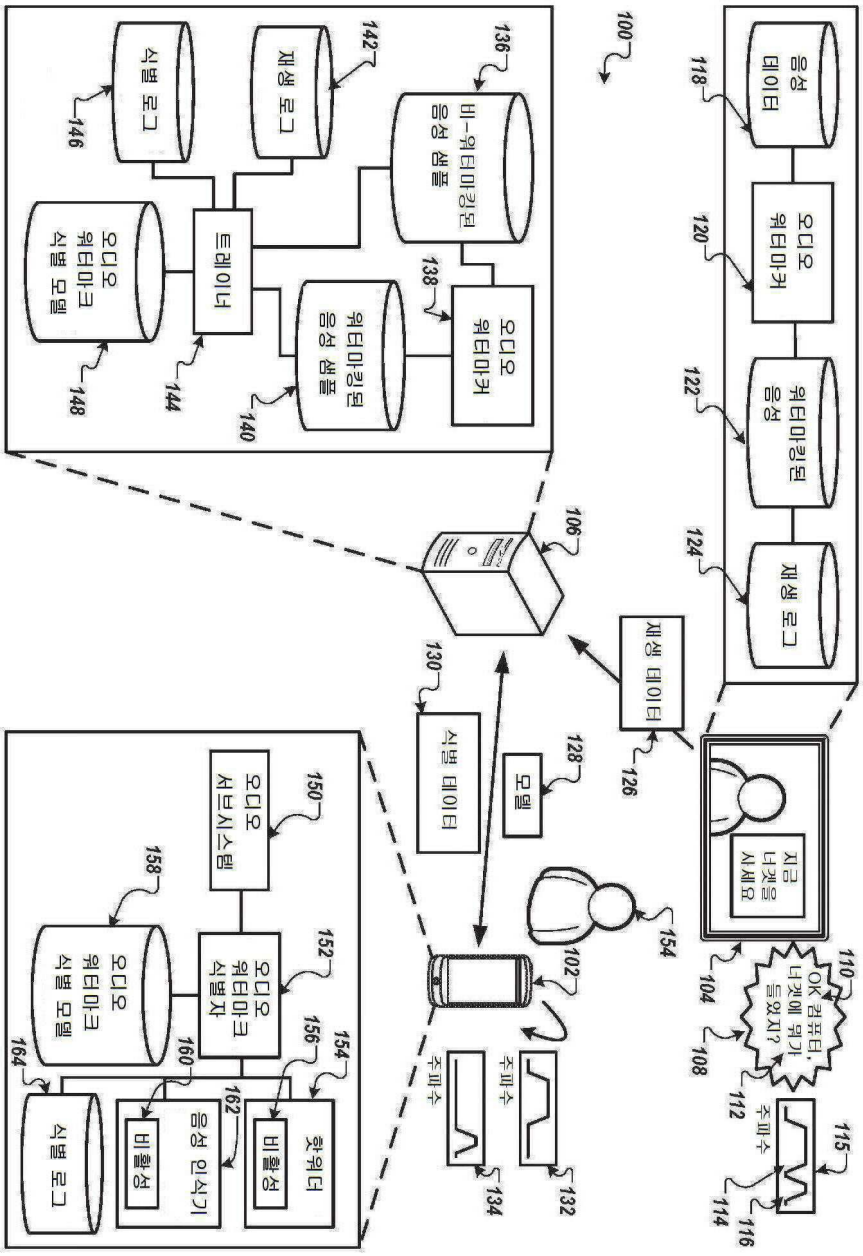
[0093] 본 명세서에서 기술된 시스템들 및 기법들은 예를 들어 데이터 서버와 같은 백엔드 컴포넌트, 애플리케이션 서버와 같은 미들웨어 컴포넌트 또는 그래픽 사용자 인터페이스를 가지는 사용자 컴퓨터 또는 사용자가 본 명세서에 기술된 시스템들 및 기법들의 구현예와 인터렉션할 수 있는 웹 브라우저와 같은 프론트엔드 컴포넌트 또는 하나 이상의 상기 백엔드, 미들웨어 또는 프론트엔드 컴포넌트들의 임의의 조합을 포함하는 컴퓨팅 시스템에서 구현될 수 있다. 시스템의 컴포넌트들은 디지털 데이터 통신의 임의의 형태 또는 매체, 예를 들어 통신 네트워크에 의해 상호연결될 수 있다. 통신 네트워크들의 예시들은 LAN(local area network), WAN(wide area network) 및 인터넷을 포함한다.

[0094] 컴퓨팅 시스템은 사용자들 및 서버들을 포함할 수 있다. 사용자와 서버는 일반적으로 서로 멀리 떨어져 있으며, 일반적으로 통신 네트워크를 통해 인터렉션한다. 사용자와 서버의 관계는 각각의 컴퓨터에서 실행되고 서로 사용자-서버 관계를 갖는 컴퓨터 프로그램에 의해 발생한다.

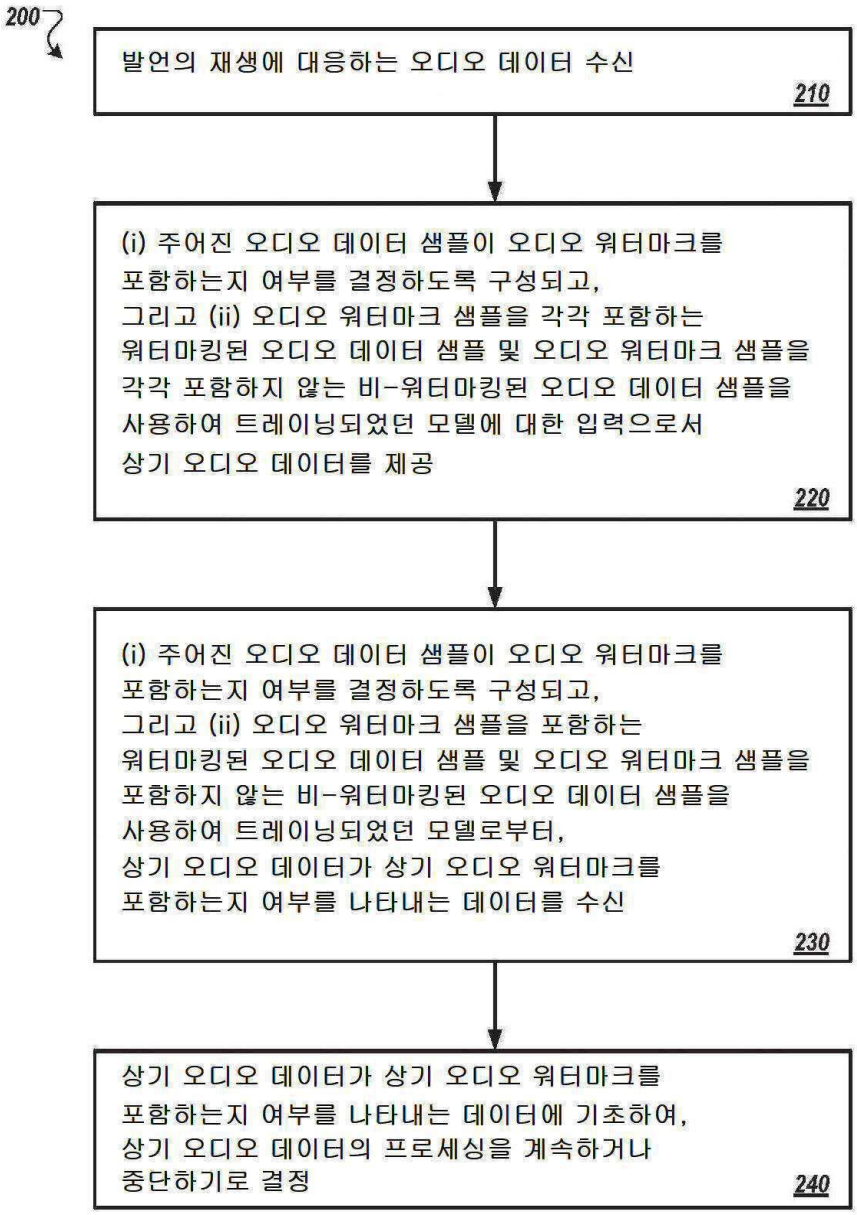
[0095] 비록 일부 구현예들이 상술되었지만, 다른 수정들도 가능하다. 예를 들면, 출원에 도시된 논리 흐름은 원하는 결과를 달성하기 위해 특정한 도시된 순서, 또는 시계열적 순서를 반드시 필요로 하지 않는다. 추가로, 다른 단계들이 제공될 수 있거나, 단계들이 기술된 흐름으로부터 생략될 수 있고, 다른 컴포넌트들이 기술된 시스템에 추가되거나 그로부터 제거될 수 있다. 따라서, 다른 구현예들도 다음의 청구항들의 범위 내에 있다. 또한, 일 양태 또는 구현예에서 설명된 구성은 임의의 다른 양태 또는 구현예에서 적용될 수 있다.

도면

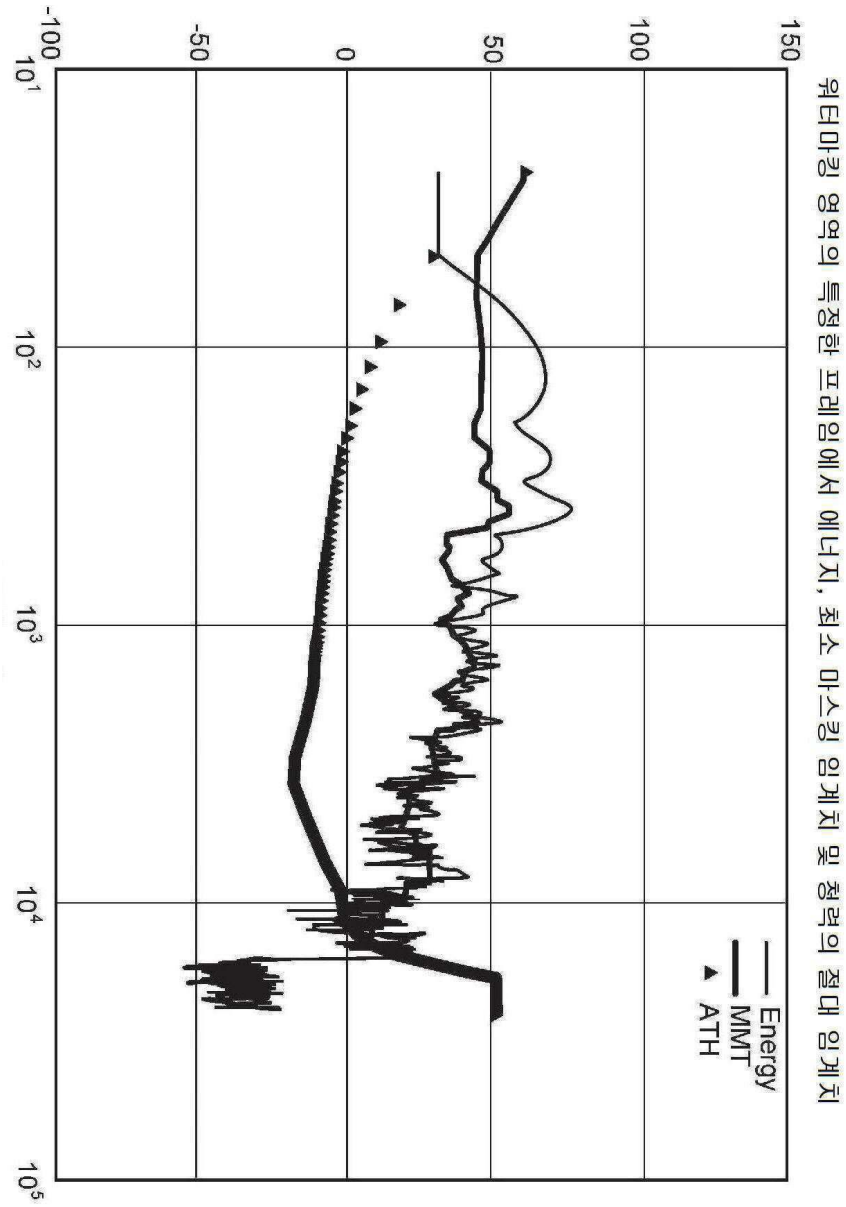
도면1


100
118
음성 데이터
120
오디오 워터마커
122
워터마킹된 음성
124
재생 로그
106
재생 데이터
126
130
식별 데이터
128
모델
136
비-워터마킹된 음성 샘플
142
재생 로그
146
식별 로그
138
오디오 워터마커
140
워터마킹된 음성 샘플
144
트레이너
148
오디오 워터마크 식별 모델
102
154
104
지금 너겟을 사세요
110
OK 컴퓨터, 너겟에 뭐가 들었지?
108
112
주파수
115
114
116
주파수
134
132
150
오디오 서브시스템
158
오디오 워터마크 식별 모델
152
오디오 워터마크 식별자
154
156
160
162
164
식별 로그
음성 인식기
비활성
핫워더
비활성

도면2

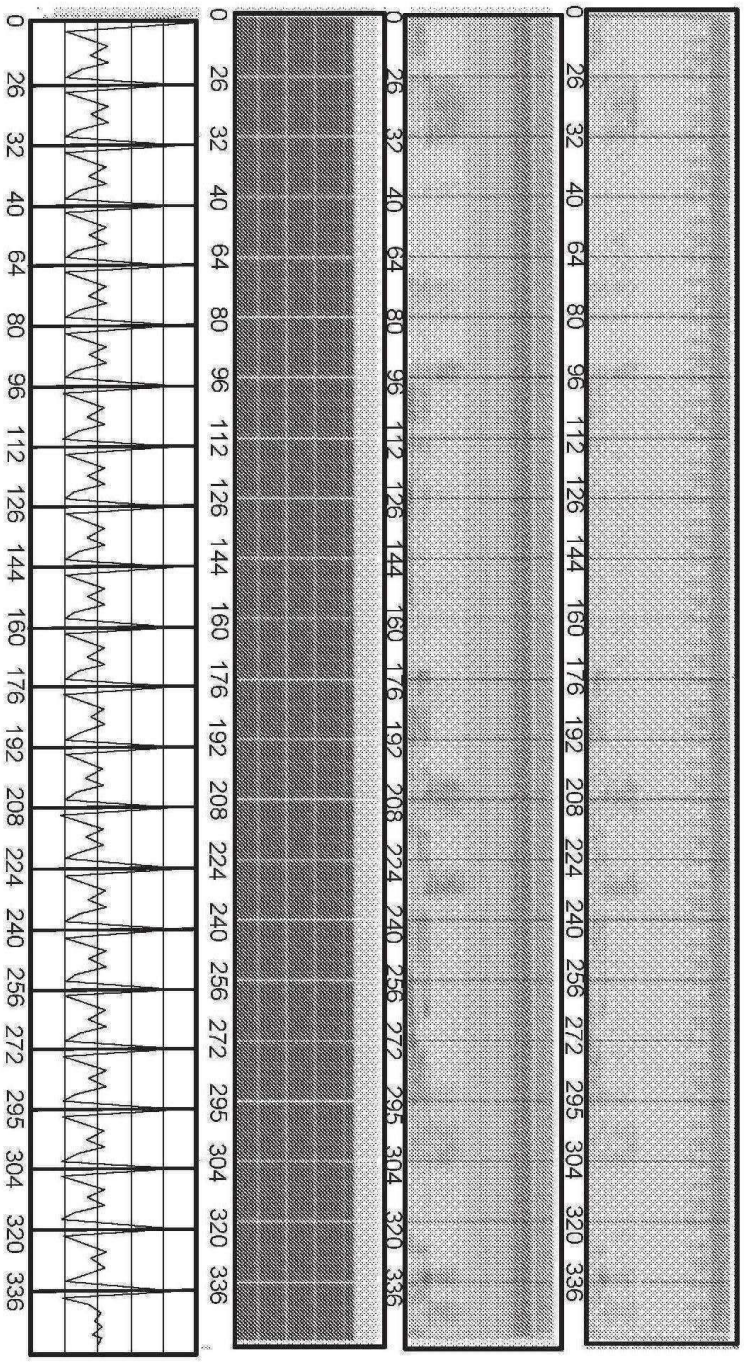

200
발언의 재생에 대응하는 오디오 데이터 수신
210
(i) 주어진 오디오 데이터 샘플이 오디오 워터마크를 포함하는지 여부를 결정하도록 구성되고, 그리고 (ii) 오디오 워터마크 샘플을 각각 포함하는 워터마킹된 오디오 데이터 샘플 및 오디오 워터마크 샘플을 각각 포함하지 않는 비-워터마킹된 오디오 데이터 샘플을 사용하여 트레이닝되었던 모델에 대한 입력으로서 상기 오디오 데이터를 제공
220
(i) 주어진 오디오 데이터 샘플이 오디오 워터마크를 포함하는지 여부를 결정하도록 구성되고, 그리고 (ii) 오디오 워터마크 샘플을 포함하는 워터마킹된 오디오 데이터 샘플 및 오디오 워터마크 샘플을 포함하지 않는 비-워터마킹된 오디오 데이터 샘플을 사용하여 트레이닝되었던 모델로부터, 상기 오디오 데이터가 상기 오디오 워터마크를 포함하는지 여부를 나타내는 데이터를 수신
230
상기 오디오 데이터가 상기 오디오 워터마크를 포함하는지 여부를 나타내는 데이터에 기초하여, 상기 오디오 데이터의 프로세싱을 계속하거나 중단하기로 결정
240

도면3

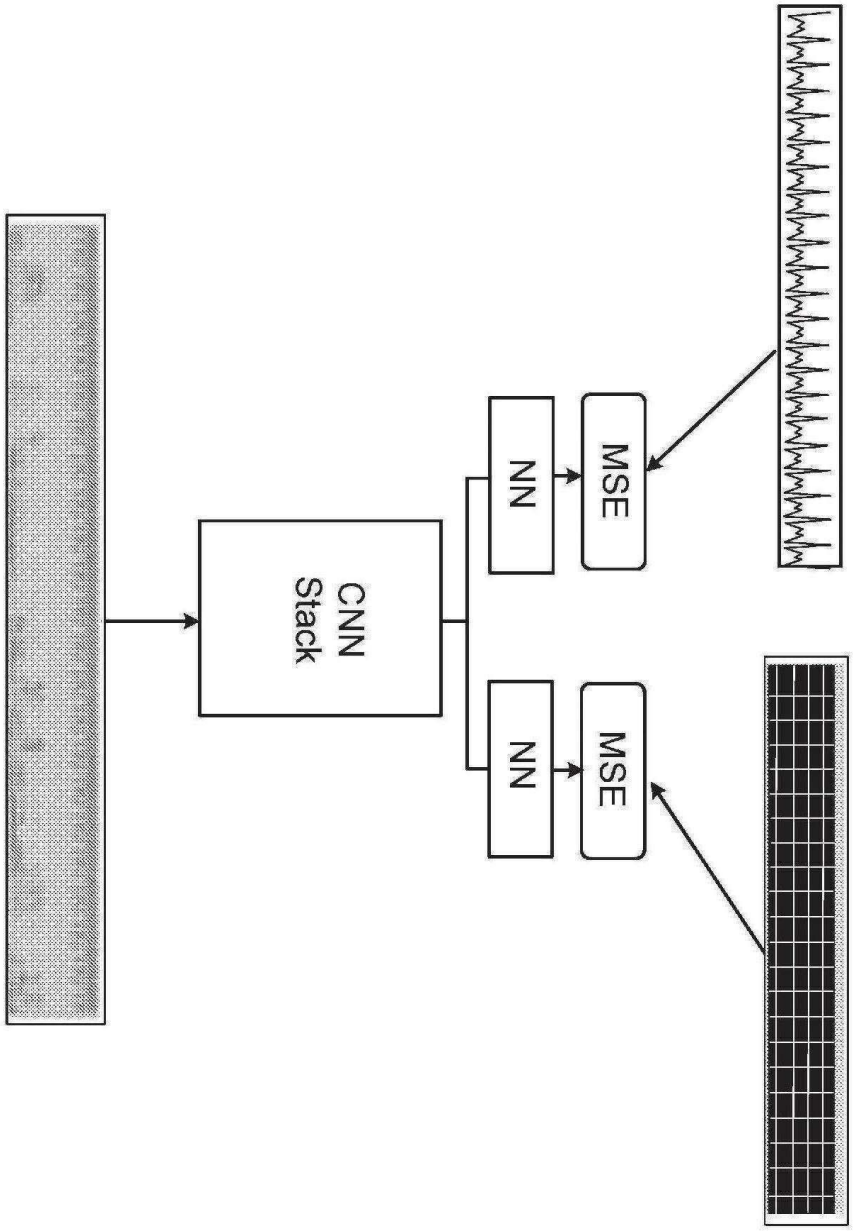


워터마킹 영역의 특정한 프레임에서 에너지, 최소 마스킹 임계치 및 청력의 절대 임계치

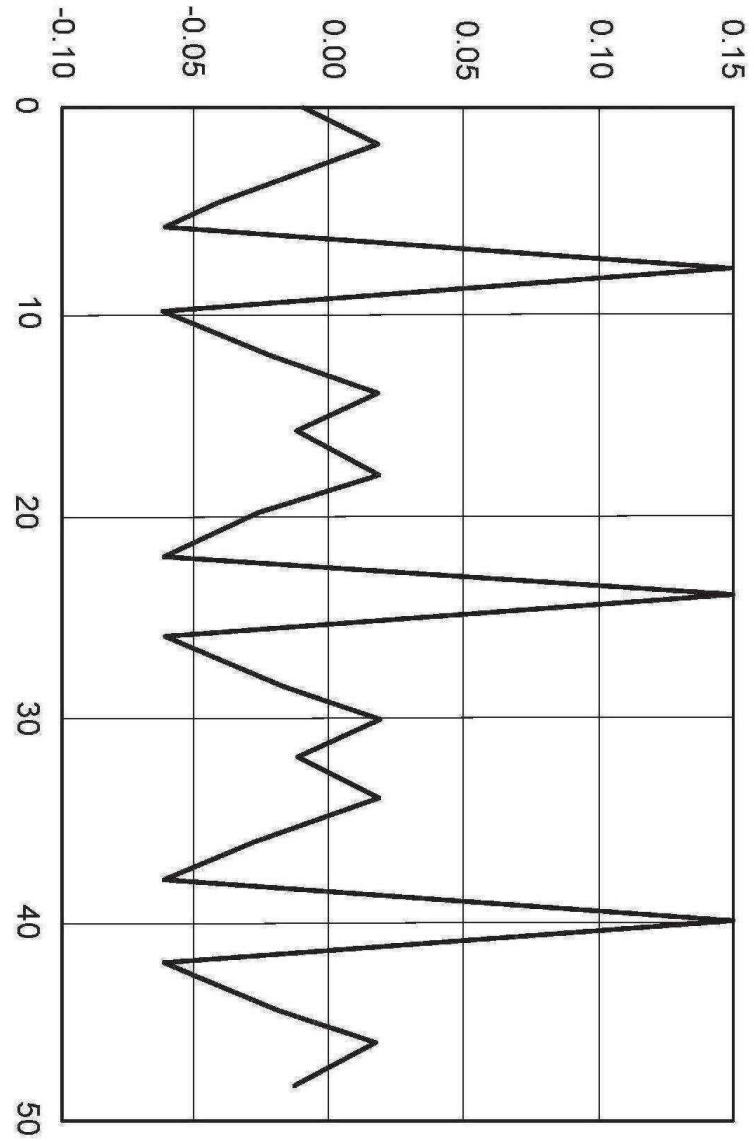
도면4


0 26 32 40 64 80 96 112 126 144 160 176 192 208 224 240 256 272 295 304 320 336
0 26 32 40 64 80 96 112 126 144 160 176 192 208 224 240 256 272 295 304 320 336
0 26 32 40 64 80 96 112 126 144 160 176 192 208 224 240 256 272 295 304 320 336
0 26 32 40 64 80 96 112 126 144 160 176 192 208 224 240 256 272 295 304 320 336

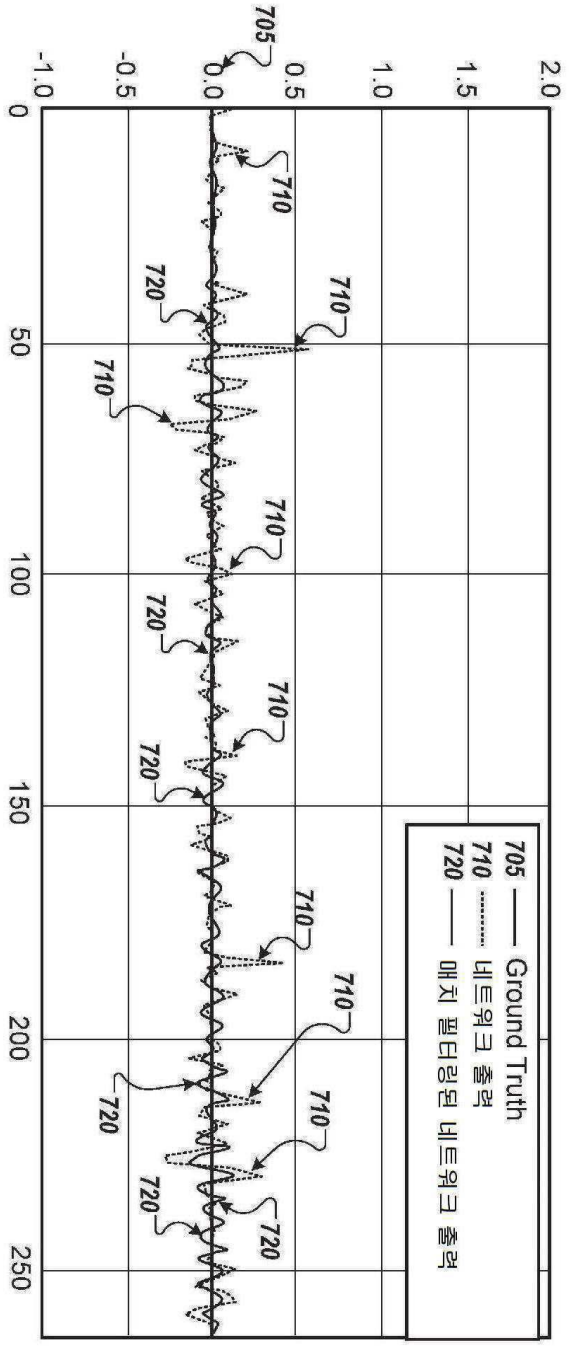
도면5



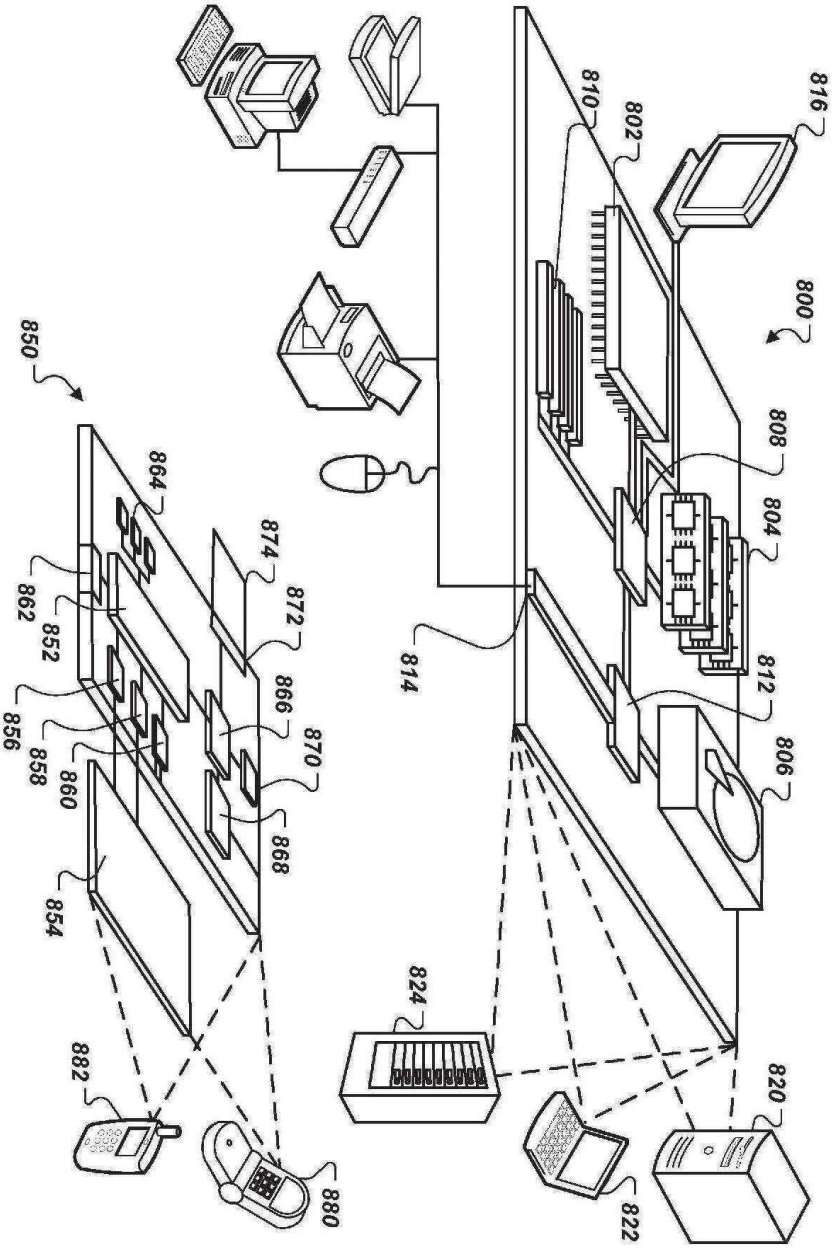
도면6



도면7


705
710
720
705 —— Ground Truth
710 ..... 네트워크 출력
720 —— 매치 필터링된 네트워크 출력
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
0
50
100
150
200
250

도면8


800
802
804
806
808
810
812
814
816
820
822
824
850
852
854
856
858
860
862
864
866
868
870
872
874
880
882