



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0119059  
(43) 공개일자 2015년10월23일

- |   |   |
|---|---|
| <p>(51) 국제특허분류(Int. Cl.)<br/> <i>G10L 25/54</i> (2013.01) <i>G10L 21/0208</i> (2013.01)<br/> <i>H04H 60/37</i> (2008.01) <i>H04H 60/58</i> (2008.01)<br/> <i>H04R 29/00</i> (2006.01)</p> <p>(52) CPC특허분류<br/> <i>G10L 25/54</i> (2013.01)<br/> <i>H04H 60/37</i> (2013.01)</p> <p>(21) 출원번호 10-2015-7024566</p> <p>(22) 출원일자(국제) 2014년03월07일<br/>         심사청구일자 2015년09월08일</p> <p>(85) 번역문제출일자 2015년09월08일</p> <p>(86) 국제출원번호 PCT/US2014/022165</p> <p>(87) 국제공개번호 WO 2014/164369<br/>         국제공개일자 2014년10월09일</p> <p>(30) 우선권주장<br/>         13/794,753 2013년03월11일 미국(US)</p> | <p>(71) 출원인<br/> <b>제너럴 인스트루먼트 코포레이션</b><br/>         미국 펜실베이니아 19044 홀삼 토너먼트 드라이브 101</p> <p>(72) 발명자<br/> <b>폰세카, 베네디토, 제이. 주니어</b><br/>         미국 60137 일리노이주 글렌 엘린 로우덴 애비뉴 601<br/> <b>바움, 케빈, 엘.</b><br/>         미국 60008 일리노이주 롤링 메도우즈, 리치니 레인 3450<br/> <i>(뒷면에 계속)</i></p> <p>(74) 대리인<br/> <b>양영준, 백만기</b></p> |
|---|---|

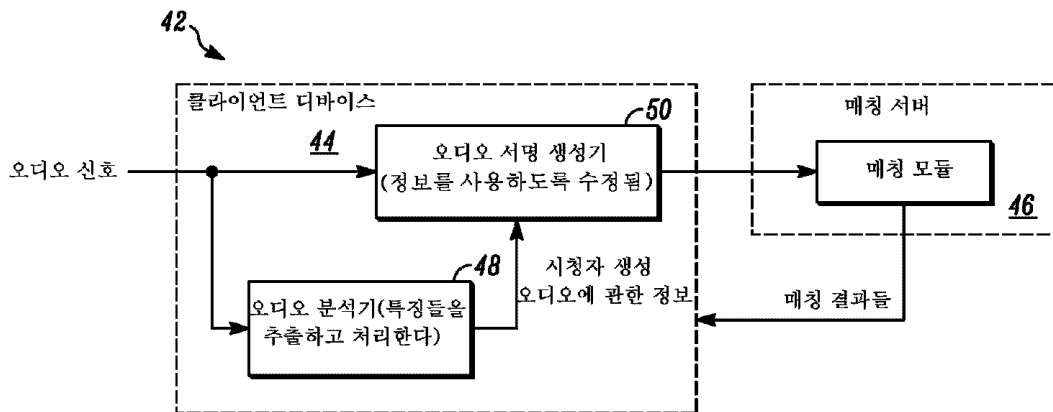
전체 청구항 수 : 총 26 항

(54) 발명의 명칭 **손상된 오디오 신호의 서명 매칭**

(57) 요약

오디오 서명들을 원격 데이터베이스에 저장된 프로그래밍 콘텐츠와 매칭시키는 디바이스들 및 방법들.

대표도



(52) CPC특허분류

*H04H 60/58* (2013.01)

*H04R 29/00* (2013.01)

*G10L 2021/02087* (2013.01)

*H04H 2201/90* (2013.01)

(72) 발명자

**이시티아크, 파이살**

미국 60640 일리노이주 시카고 웨리단 노쓰 5415

**윌리엄스, 제이 제이.**

미국 60025 일리노이주 글렌뷰 크랩트리 레인 317

## 명세서

### 청구범위

#### 청구항 1

장치로서,

(a) 하나 이상의 사용자들에게 미디어 콘텐츠를 출력하는 디바이스로부터의 주 오디오(primary audio)를 포함하는 오디오 신호를 수신할 수 있는 마이크로폰;

(b) (i) 수신된 상기 오디오 신호를 분석하여 상기 오디오 신호 내의 손상의 존재 또는 부재를 식별하고; 그리고 (ii) 식별된 상기 오디오 신호 내의 손상의 존재 또는 부재에 기초하여 시간 구간에 걸쳐 수신된 상기 오디오의 오디오 서명을 생성하는 적어도 하나의 프로세서; 및

(c) 상기 오디오 서명을 서버에 통신하는 송신기 및 상기 서버로부터 응답을 수신할 수 있는 수신기 - 상기 응답은 상기 오디오 서명 및 상기 손상에 기초함 -

를 포함하는 장치.

#### 청구항 2

제1항에 있어서,

상기 마이크로폰은 상기 프라이머리 오디오와 관련없는 오디오를 수신할 수 있고, 상기 프로세서는 상기 관련없는 오디오에 의해 손상된 상기 오디오 서명의 해당 부분들을 무효화함으로써 상기 오디오 서명을 수정하는 장치.

#### 청구항 3

제2항에 있어서,

상기 주 오디오와 관련없는 상기 오디오는 사용자 생성 오디오인 장치.

#### 청구항 4

제2항에 있어서,

상기 프로세서는: (i) 에너지 임계치; (ii) 수집된 오디오의 스펙트럼 특성들에서의 변경; 및 (iii) 상기 오디오 내의 공지된 사용자의 스피치의 존재를 나타내는 스피커 검출기 중 적어도 하나에 기초하여 상기 주 오디오와 관련없는 상기 오디오를 식별하는 장치.

#### 청구항 5

제1항에 있어서,

상기 송신기는 상기 구간의 어느 부분들이 상기 손상을 포함하는지를 상기 서버에 통신하는 장치.

#### 청구항 6

제1항에 있어서,

상기 서버는 데이터베이스 내의 복수의 콘텐츠 중에서 상기 사용자가 보는 콘텐츠를 식별하기 위해 상기 오디오 서명을 사용할 수 있는 장치.

#### 청구항 7

제1항에 있어서,

상기 프로세서는 상기 구간에 걸쳐 복수의 오디오 서명들을 생성하고, 각각의 오디오 서명은 상기 구간의 연속적인 선택된 부분과 연관된 장치.

**청구항 8**

제1항에 있어서,

상기 프로세서는 상기 프로세서에 의해 식별된 손상의 지속기간에 기초하여 상기 마이크로폰에 의해 오디오가 수집되는 기간을 연장하는 장치.

**청구항 9**

제1항에 있어서,

상기 시간 구간의 시작 시간, 종료 시간, 또는 지속기간 중 적어도 하나는 상기 손상에 응답하는 장치.

**청구항 10**

제1항에 있어서,

상기 수신기는 상기 서버가 상기 데이터베이스 내의 콘텐츠에 상기 오디오 서명을 매칭시키는 것에 기초하여 상기 서버로부터 상보적 콘텐츠(complementary content)를 수신하는 장치.

**청구항 11**

장치로서,

(a) 복수의 기준 오디오 서명들을 검색할 수 있는 프로세서 - 각각의 상기 기준 오디오 서명은 제시 디바이스 (presentation device) 상에서 사용자에게 이용가능한 오디오 또는 시청각 프로그램과 연관됨 -;

(b) 상기 사용자에게 근접한 프로세싱 디바이스로부터의 질의 오디오 서명, 및 상기 질의 오디오 서명 내의 손상의 존재를 나타내는 메시지를 수신할 수 있는 수신기

를 포함하고;

(c) 상기 프로세서는 상기 메시지 및 상기 질의 오디오 서명을 사용하여 상기 사용자에게 의해 시청되는 콘텐츠를 식별하는 장치.

**청구항 12**

제11항에 있어서,

상기 질의 오디오 서명은 제1 시간에서 제2 시간까지의 구간을 포함하고, 상기 메시지는 상기 기준 오디오 서명들 중 적어도 하나에 매칭시킬 상기 질의 오디오 서명의 선택적 부분들을 나타내기 위해 상기 프로세서에 의해 사용되는 장치.

**청구항 13**

제12항에 있어서,

상기 메시지는 상기 질의 오디오 서명을 상기 기준 오디오 서명들 중 상기 적어도 하나에 매칭시키는 경우 상기 기준 오디오 서명들 내의 구간들을 무효화(nullify)시키기 위해 사용되는 장치.

**청구항 14**

제11항에 있어서,

상기 메시지는 적어도 하나의 다른 상기 질의 오디오 서명이 수신될 때까지 상기 사용자에게 의해 시청되고 있는 상기 프로그램의 식별을 선택적으로 지연시키기 위해 상기 프로세서에 의해 사용되는 장치.

**청구항 15**

제11항에 있어서,

상기 장치는 적어도 하나의 질의 오디오 서명을 수신하고:

(a) 각각의 상기 질의 오디오 서명을 기준 오디오 서명과 비교하고;

(b) 상기 기준 오디오 서명과의 비교에 기초하여 상기 적어도 하나의 질의 오디오 서명에 대한 각자의 스코어들을 생성하고, 상기 스코어들을 가산하여 전체 스코어를 획득하고;

(c) 적어도 하나의 다른 기준 오디오 서명에 대해 단계들 (a) 및 (b)을 반복하고; 그리고

(d) 가장 높은 전체 스코어를 야기시키는 기준 오디오 서명과 연관된 오디오 또는 시청각 프로그램 세그먼트를, 상기 사용자에게 의해 시청되는 상기 콘텐츠로서 식별

함으로써 상기 사용자에게 의해 시청되는 상기 콘텐츠를 식별하는 장치.

#### 청구항 16

제11항에 있어서,

상기 장치는 적어도 하나의 질의 오디오 서명을 수신하고:

(a) 각각의 상기 적어도 하나의 질의 오디오 서명을 기준 오디오 서명과 비교하고;

(b) 타겟인 상기 기준 오디오 서명과의 비교에 기초하여 상기 적어도 하나의 질의 오디오 서명에 대한 각자의 스코어들을 생성하고, 상기 스코어들을 가산하여 전체 스코어를 획득하고;

(c) 상기 전체 스코어가 임계치를 초과하는 경우, 상기 콘텐츠가 상기 사용자에게 의해 시청되는 경우, 상기 스코어가 상기 임계치를 초과하게 하는 기준 오디오 서명과 연관된 오디오 또는 시청각 프로그램 세그먼트를, 상기 사용자에게 의해 시청되는 상기 콘텐츠로서 식별하고;

(d) 상기 전체 스코어가 상기 임계치를 초과하지 않는 경우, 상기 데이터베이스 내의 또다른 기준 오디오 서명을 타겟 기준 오디오 서명으로서 지정하고, 상기 전체 스코어가 상기 임계치를 초과하거나 상기 데이터베이스 내의 모든 프로그램들이 지정될 때까지 단계들 (a) 및 (b)을 반복

함으로써, 상기 사용자에게 의해 시청되고 있는 상기 콘텐츠를 식별하는 장치.

#### 청구항 17

제11항에 있어서,

상기 프로세서는 상기 사용자에게 의해 시청되고 있는 상기 콘텐츠를 식별하기 위해 스코어들을 사용하고, 상기 스코어들은 상기 질의 오디오 서명을 상기 기준 오디오 서명과 비교함으로써 생성되고, 상기 스코어들은 상기 메시지 내의 정보에 기초하여 정규화되는 장치.

#### 청구항 18

제11항에 있어서,

상기 기준 오디오 서명들 각각은 시간 길이를 가지고, 상기 프로세서는 상기 메시지에 기초하여 상기 길이를 연장할 수 있는 장치.

#### 청구항 19

장치로서,

(a) 오디오 디바이스로부터 제1 마이크로폰에 의해 수집된 제1 오디오 신호에 대응하는 오디오 특징들의 제1 시퀀스를 제1 장치로부터 수신하는 것;

(b) 상기 오디오 디바이스로부터 제2 마이크로폰에 의해 수집된 제2 오디오 신호에 대응하는 오디오 특징들의 제2 시퀀스를 제2 장치로부터 수신하는 것;

(c) 상기 제1 및 제2 오디오 특징들을 사용하여 (i) 상기 제1 오디오 신호 내의 손상의 존재 또는 부재를 식별하고; (ii) 상기 제2 오디오 신호 내의 손상의 존재 또는 부재를 식별하고; (iii) 상기 제1 오디오 신호 및 상기 제2 오디오 신호 각각 내의 식별된 손상의 존재 또는 부재에 기초하여 상기 오디오 디바이스에 의해 생성된 오디오의 오디오 서명을 생성하는 프로세서; 및

(d) 상기 오디오 서명을 서버에 통신하는 송신기

를 포함하는 장치.

**청구항 20**

방법으로서,

- (a) 프로세서를 가지는 디바이스에 근접한 사용자에게 콘텐츠를 제시하는 디바이스로부터 오디오 신호를 수신하는 단계;
- (b) 상기 오디오의 선택 부분들을 손상된 것으로서 식별하는 단계;
- (c) 상기 오디오 및 상기 식별을 사용하여 수신된 상기 오디오의 적어도 하나의 질의 오디오 서명을 생성하는 단계;
- (d) 상기 적어도 하나의 질의 오디오 서명을, 각각이 상기 사용자에게 이용가능한 콘텐츠의 세그먼트를 나타내는 복수의 기준 오디오 서명과 비교하는 단계 - 상기 복수의 기준 오디오 서명들은 상기 디바이스로부터 원격 위치에 있으며, 상기 비교는 상기 적어도 하나의 질의 오디오 서명 내의 손상의 선택적 식별에 기초함 - ;
- (e) 상기 비교에 기초하여, 상기 디바이스로부터 원격인 상기 위치로부터 상기 디바이스에 보조 콘텐츠 (supplementary content)를 송신하는 단계

를 포함하는 방법.

**청구항 21**

제20항에 있어서,

상기 질의 오디오 서명은 상기 질의 오디오 서명의 손상된 부분들을 무효화시킴으로써 생성되는 방법.

**청구항 22**

제20항에 있어서,

상기 질의 오디오 서명의 일부 시간 부분들이 손상되었다는 것을 나타내는 메시지를 상기 디바이스로부터 원격인 상기 위치에 송신하는 단계를 포함하는 방법.

**청구항 23**

제22항에 있어서,

상기 메시지는 상기 질의 오디오 서명에 임베딩되는(embedded) 방법.

**청구항 24**

제22항에 있어서,

상기 메시지는 적어도 하나의 다른 상기 질의 오디오 서명이 수신될 때까지 상기 비교를 선택적으로 지연시키기 위해 사용되는 방법.

**청구항 25**

장치로서,

- (a) 하나 이상의 사용자들에게 미디어 콘텐츠를 출력하는 디바이스로부터의 주 오디오를 포함하는 오디오 신호를 수신할 수 있는 적어도 하나의 마이크론 - 상기 오디오 신호는 사용자 생성 오디오에 의해 손상됨 - ;
- (b) (i) 수신된 상기 오디오 신호의 제1 오디오 서명을 생성하고; (ii) 수신된 상기 오디오 신호를 분석하여 상기 사용자 생성 오디오에 의해 손상되지 않은 상기 오디오 서명 내의 적어도 하나의 구간을 식별하고; 그리고 (iii) 식별된 상기 적어도 하나의 구간을 사용하여 상기 제1 오디오 서명을 데이터베이스에 저장된 제2 오디오 서명과 매칭시키는 적어도 하나의 프로세서

를 포함하는 장치.

**청구항 26**

제25항에 있어서,

상기 적어도 하나의 프로세서는 상기 제2 오디오 서명과의 매칭에 기초하여 상기 제1 오디오 서명을 상기 주 오디오와 동기화시키는 장치.

**발명의 설명**

**기술 분야**

[0001] 관련 출원들의 교차-참조

[0002] 없음

[0003] 이 출원의 발명 대상은 넓게는 사용자에게 의해 시청되는 오디오 또는 시청각 콘텐츠의 원격 식별을 용이하게 하는 시스템들 및 방법들에 관한 것이다.

**배경 기술**

[0004] 많은 경우들에서, 라이브 텔레비전 또는 라디오 상의 방송, DVD 또는 CD 상에서 재생되는 콘텐츠, DVR 상에 레코딩된 시간-시프트된 콘텐츠와 같이, 사람에게 제시되는 오디오 또는 시청각 콘텐츠를 정확하게 식별하는 것이 유용하다. 일 예로서, 텔레비전 또는 다른 방송 등급들을 캡파일링하거나, 특정 시간 슬롯들 동안 어느 광고들이 보여질지를 결정하는 경우, 개별 시청자의 장비 상에 재생되는 콘텐츠를 캡처하고, 특히 지역 방송이 제휴하는 경우 지역에 따라 다른(geographically-varying) 콘텐츠를 디스플레이하거나, 국영 방송 내에 로컬 광고 콘텐츠를 삽입하는 것이 유리하다. 또다른 예로서, 콘텐츠 제공자들은 방송 콘텐츠와 동기화된 보조 자료를 제공하기를 원할 수 있고, 따라서 시청자가 특정 쇼를 시청하는 경우, 보조 자료가 랩톱 컴퓨터, 태블릿 등과 같은 그 사용자의 보조 디스플레이 디바이스에 제공될 수 있다. 이러한 방식으로, 시청자가 야구 생방송을 시청할 것으로 결정되는 경우, 선수가 타격하는 경우 각각의 타자의 통계들이 사용자의 랩톱에 스트리밍될 수 있다.

[0005] 특정 순간에 사용자가 어느 콘텐츠를 시청하는지를 동시에 결정하는 것은 사소한 작업이 아니다. 일부 기법들은 셋톱 박스가 프레임들을 디코딩함에 따라 비디오를 분석하는 셋톱 박스 내의 특수 하드웨어에 의존한다. 그러나, 이러한 시스템들에 대한 필수 프로세싱 능력은 종종 비용-제한적이다. 추가로, 디코딩된 프레임들의 정확한 식별은 사용자가 16:9와 같은 또다른纵横비로 콘텐츠를 시청하고 있을 수 있는 경우 통상적으로 디스플레이에 대한纵横비, 예를 들어, 4:3을 상정하며, 이에 의해, 시청되고 있는 프로그램 콘텐츠의 정확한 식별을 불가능하게 한다. 유사하게, 이러한 시스템들은 또한 시청자의 시스템에 의해 변경될 수 있는 프로그램 프레임율에 대해 너무 민감하며, 또한 시청되는 콘텐츠의 정확한 식별을 저해한다.

[0006] 다른 식별 기법은 추후 식별을 위해 시청각 콘텐츠 내에 보조 코드들을 추가한다. 보조 코드가 통지되지 않도록 신호에 보조 코드를 추가하기 위한 다수의 방법들이 존재한다. 예를 들어, 코드는 그것을 비디오의 수직 블랭킹 구간(vertical blanking interval) 또는 수평 귀선 구간(horizontal retrace interval) 내에 삽입함으로써 텔레비전 비디오의 비-가시적 부분들에 숨겨질 수 있다. 다른 공지된 비디오 인코딩 시스템들은, 그렇지 않은 경우 작은 신호 에너지를 반송하는 신호의 전송 대역폭의 일부분에 보조 코드를 매립한다. 또다른 방법들 및 시스템들은 콘텐츠의 오디오 부분, 예를 들어, 영화 사운드트랙에 보조 코드를 추가한다. 이러한 구성들은 텔레비전 뿐만 아니라, 라디오 및 사전-레코딩된 음악에도 적용가능하다는 장점을 가진다. 또한, 오디오 신호들에 추가된 보조 코드들은 스피커의 출력에서 재생될 수 있고, 따라서, 시청자에 근접한 마이크로폰을 사용하여 코드들을 비강제적으로(non-intrusively) 가로채서 식별할 가능성을 제공한다.

[0007] 시청각 콘텐츠 내의 내장된 코드들의 사용이 사용자에게 제시되는 콘텐츠를 효과적으로 식별할 수 있지만, 이러한 코드들은 실제 사용 시에 단점들을 가진다. 예를 들어, 코드는 소스 인코더에 내장될 필요가 있을 것이며, 코드는 사용자에게 완전히 인지불가능하지 않을 수 있거나, 또는 소비자-등급 카메라들 및 마이크로폰들에서의 센서 왜곡들에 대해 강건하지 않을 수 있다.

**발명의 내용**

**도면의 간단한 설명**

[0008]

발명의 더 양호한 이해를 위해, 그리고 발명이 실제로 어떻게 수행될 수 있는지를 도시하기 위해, 이제 첨부도면들에 대한 참조가 예시에 의해 이루어질 것이다.

도 1은 네트워크 접속을 통해 액세스 가능한 서버의 도움으로, 제1 디바이스 상에서 사용자에게 제시되는 오디오 또는 시청각 콘텐츠를, 제2 디바이스를 통해 사용자에게 제공되는 보조 콘텐츠와 동기화시키는 시스템을 도시한다.

도 2는 도 1의 제2 디바이스에 의해 캡처된 오디오 세그먼트의 스펙트로그램(spectrogram)을, 그 스펙트로그램으로부터 생성된 오디오 서명과 함께 도시한다.

도 3은 기준 스펙트로그램으로부터 생성된 오디오 서명과 함께, 그리고 도 1에 도시된 서버가 액세스가능한 데이터베이스에 저장된, 도 2의 오디오 세그먼트의 기준 스펙트로그램을 도시한다.

도 4는 도 3의 오디오 서명과 도 1의 서버의 데이터베이스 내의 매칭 오디오 서명 간의 비교를 도시한다.

도 5는 외부 잡음에 의해 손상된 오디오 서명과 손상되지 않은 오디오 서명과의 비교를 도시한다.

도 6은 도 5의 손상된 서명이 서버(18)에 의해 수신되는 경우 부정확한 매칭을 초래할 수 있음을 예시한다.

도 7은 텔레비전과 같은 디스플레이 디바이스로부터 클라이언트 디바이스에 의해 캡처된 사용자 재채기 또는 이야기 오디오의 파형들을 도시한다.

도 8은 관련없는 오디오(extraneous audio)의 에너지 레벨에 대해 오디오 서명들을 정확하게 매칭시키는데 있어서의 성능 저하의 다양한 레벨들을 도시한다.

도 9는 손상된 오디오 서명을 정정하는 제1 시스템을 도시한다.

도 10은 손상된 오디오 서명과 도 9의 시스템에 의해 정정된 것과의 비교를 도시한다.

도 11은 도 9의 시스템의 성능을 예시한다.

도 12는 손상된 오디오 서명을 정정하는 제2 시스템을 도시한다.

도 13은 손상된 오디오 서명을 정정하는 제3 시스템을 도시한다.

도 14는 도 13의 시스템의 성능을 도시한다.

도 15 및 16은 손상된 오디오 서명을 정정하는 제4 시스템을 도시한다.

**발명을 실시하기 위한 구체적인 내용**

[0009]

도 1은, 사용자가 제1 디바이스(12) 상에서 시청하는 콘텐츠를 정확하게 식별할 수 있고, 따라서 보조 자료가 사용자에게 근접한 제2 디바이스(14)에 제공될 수 있는 시스템(10)의 아키텍처를 도시한다. 제1 디바이스(12)에 의해 출력된 미디어 콘텐츠로부터의 오디오는 디바이스(12)로부터 수신된 "프라이머리 오디오" 또는 간단히 오디오라고 지칭될 수 있다. 제1 디바이스(12)는 텔레비전일 수 있거나 또는 컴퓨터 디스플레이, 태블릿, PDA, 셀 폰 등과 같이, 사용자에게 시청각 콘텐츠를 제시할 수 있는 임의의 다른 디바이스일 수 있다. 대안적으로, 제1 디바이스(12)는 MP3 플레이어와 같이, 오디오 콘텐츠를 임의의 다른 정보와 함께 사용자에게 제시할 수 있는 디바이스일 수 있거나, 또는 그것은, 라디오 또는 오디오 시스템과 같이, 사용자에게 오디오 콘텐츠만을 제시할 수 있는 디바이스일 수 있다. 제2 디바이스(14)는 태블릿 디바이스로서 도시되었지만, 개인용 컴퓨터, 랩톱, PDA, 셀 폰, 또는 컴퓨터 프로세서 뿐만 아니라 마이크로폰(16)에, 그리고 선택적으로 하나 이상의 추가적인 마이크로폰들(미도시됨)에 동작가능하게 접속된 임의의 다른 유사한 디바이스일 수 있다.

[0010]

제2 디바이스(14)는 바람직하게는 마이크로폰(16) 또는 오디오 신호를 수신할 수 있는 다른 디바이스에 동작가능하게 접속된다. 마이크로폰(16)은 제1 디바이스(12) 상에 제시되는 콘텐츠의 세그먼트와 연관된 프라이머리 오디오 신호를 수신한다. 제2 디바이스(14)는 이후 내부 프로세서 또는 그것이 액세스가능한 임의의 다른 프로세서를 사용하여 수신된 신호의 오디오 서명을 생성한다. 하나 이상의 추가적인 마이크로폰이 사용되는 경우, 제2 디바이스는 바람직하게는 수신된 신호의 오디오 서명을 생성하기 이전에 다수의 마이크로폰들로부터의 수신된 신호를 프로세싱하고 결합시킨다. 제1 디바이스(12) 상에 동시에 디스플레이되는 콘텐츠에 대응하는 오디오 서명이 일단 생성되면, 그 오디오 서명은 인터넷과 같은 네트워크(20), 또는 LAN 또는 WAN과 같은 다른 네트워크를 통해 서버(18)에 송신된다. 서버(18)는 일반적으로는 제1 디바이스(12) 및 제2 디바이스(14)로부터 원격

인 위치에 있을 것이다.

[0011] 때때로 오디오 지문이라 명명될 수 있는 오디오 서명이 임의의 개수의 기법들을 사용하여 표현될 수 있다는 점이 이해되어야 한다. 단지 몇몇 이러한 예들을 인용하자면, 캡처된 오디오 신호의 스펙트로그램 내의 패턴이 오디오 서명을 형성할 수 있고; 스펙트로그램 내의 피크들에 대응하는 시간 및 주파수 쌍들의 시퀀스가 오디오 서명을 형성할 수 있고; 스펙트로그램의 주파수 대역들 내의 피크들 간의 시간차들의 시퀀스들이 오디오 서명을 형성할 수 있고; 각각의 엔트리가 양자화된 시간 구간들 내의 높은 또는 낮은 에너지에 대응하는 이진 행렬이 오디오 서명을 형성할 수 있고 양자화된 주파수 대역들이 오디오 서명을 형성할 수 있다. 종종, 오디오 서명은 서버에 의한 데이터베이스 검색을 용이하게 하기 위해 스트링으로 인코딩된다.

[0012] 서버(18)는 바람직하게는 데이터베이스에 복수의 오디오 서명들을 저장하며, 각각의 오디오 서명은 제1 디바이스(12) 상에 디스플레이될 수 있는 콘텐츠와 연관된다. 저장된 오디오 서명들은 각각 오디오 또는 시청각 콘텐츠의 특정 항목 내의 사전-선택된 구간과 연관될 수 있고, 따라서, 프로그램은 다수의, 시간상으로 순차적인 오디오 서명들에 의해 데이터베이스에 표현된다. 대안적으로, 저장된 오디오 서명들은 각각 프로그램 전체를 연속적으로 스패닝(span)할 수 있고, 따라서, 그 프로그램의 임의의 정의된 구간에 대한 오디오 서명이 생성될 수 있다. 제2 디바이스(14)로부터의 오디오 서명의 수신 시에, 서버(18)는 수신된 서명을 그것의 데이터베이스 내의 서명과 매칭시키려고 시도한다. 성공적인 매칭이 발견되는 경우, 서버(18)는 매칭하는 프로그래밍 세그먼트와 연관된 보조 콘텐츠를 제2 디바이스(14)에 송신할 수 있다. 예를 들어, 어떤 사람이 BMW 또는 다른 자동차의 이미지를 디스플레이하는 순간 제1 디바이스(12) 상에서 제임스 본드 영화를 시청하고 있는 경우, 서버(18)는 수신된 오디오 서명을 사용하여 시청된 세그먼트를 식별하고, 제조, 모델, 가격 정보 등과 같은 그 자동차에 관한 보조 정보를 제2 디바이스(14)에 송신할 수 있다. 이러한 방식으로, 제2 디바이스(14)에 제공되는 보조 자료는 바람직하게는 디바이스(12)에 의해 전체적으로 제시되는 프로그램 또는 다른 콘텐츠에 동기화될 뿐만 아니라, 전송된 보조 콘텐츠가 제1 디바이스(12) 상에서 동시에 디스플레이되고 있는 것에 관련될 수 있도록 콘텐츠의 특정 부분들에 동기화된다.

[0013] 동작 시, 전술한 절차는 바람직하게는, 제2 디바이스(14)에 의해, 수동적 선택 또는 자동 활성화 중 어느 하나에 의해 개시될 수 있다. 후자의 경우, 예를 들어, 많은 기존의 태블릿 디바이스들, PDA들, 랩톱들 등이 텔레비전, 또는 셋톱 박스를 원격으로 동작시키거나, 또는 시청된 프로그래밍 등에 대한 프로그램 가이드에 액세스하기 위해 사용될 수 있다. 따라서, 이러한 디바이스는 이러한 기능들이 디바이스 상에서 수행되는 경우마다 오디오 서명 생성 및 매칭 절차를 시작하도록 구성될 수 있다. 일단 서명 생성 및 매칭 절차가 개시되면, 마이크로폰(16)은 제1 디바이스(12)로부터 오디오를 캡처하기 위해 주기적으로 활성화되고, 스펙트로그램은 마이크로폰이 활성화된 각각의 구간에 걸쳐 캡처된 오디오로부터 근사화된다. 예를 들어,  $S[f, b]$ 를 지속시간  $T$ , 예를 들어,  $T=120$  프레임, 5초 등을 가지는 신호  $s(t)$ 의 프레임 "f" 동안 대역 "b"에서의 에너지를 나타낸다고 하자. 모든 대역들( $b=1, \dots, B$ )이 변경되고, 모든 프레임들( $f=1, \dots, F$ )이 신호  $s(t)$  내에서 변경됨에 따라  $S[f, b]$ 의 집합은  $F \times B$  행렬  $S$ 를 형성하며, 이는 신호의 스펙트로그램과 유사하다. 대역들 "b"가 고속 푸리에 변환(FFT) 빈들(bins)이 아니라 오히려 각각의 FFT 빈 내의 에너지의 선형 조합이기 때문에, 모든  $S[f, b]$ 의 집합이 반드시 스펙트로그램과 등가적이지는 않지만, 이 개시내용의 목적을 위해, 이러한 절차가 스펙트로그램의 등가물을 생성하거나, 또는 오디오 신호로부터 스펙트로그램을 생성하기 위한 일부 대안적인 절차가 사용되며, 이들이 당해 기술분야에 공지되어 있다고 가정될 것이다.

[0014] 오디오의 캡처된 세그먼트로부터 생성된 스펙트로그램을 사용하여, 제2 디바이스(14)는 그 세그먼트의 오디오 서명을 생성한다. 제2 디바이스(14)는 바람직하게는, 스펙트로그램(22) 내의 오디오 에너지에서의 피크들의 위치를 식별하기 위해, 임계치 연산을 스펙트로그램 내에 레코딩된 각자의 에너지들  $S[f, b]$ 에 적용하여 오디오 서명을 생성한다. 임의의 적절한 임계치가 사용될 수 있다. 예를 들어, 전술한 행렬  $S[f, b]$ 이 캡처된 오디오 신호의 스펙트로그램을 나타낸다고 가정하면, 제2 디바이스(14)는 바람직하게는 서명  $S^*$ 을 생성할 수 있는데, 이는  $S[f, b]$ 가  $S$ 의 모든 엔트리들 중에서 가장 높은 에너지를 가지는  $P\%$ (예를 들어,  $P\%=10\%$ ) 피크들 사이에 있는 경우  $S^*[f, b]=1$ 인, 이진  $F \times B$  행렬이다. 오디오 서명을 생성하기 위한 다른 가능한 기법들은 스펙트로그램에 레코딩된 최대 에너지의 백분율로서 선택된 임계치를 포함할 수 있다. 대안적으로, 스펙트로그램에 레코딩된 신호 에너지의 특정된 백분율을 보유하는 임계치가 선택될 수 있다.

[0015] 도 2는 캡처된 스펙트로그램(22)로부터 생성된 오디오 서명(24)과 함께, 도 1에 도시된 제2 디바이스(14)의 마이크로폰(16)에 의해 캡처된 오디오 신호의 스펙트로그램(22)을 예시한다. 스펙트로그램(22)은 수평축 상에 도시된 시간 구간들에서, 수직축 상에 도시된 정의된 주파수 대역들(kHz) 내에서, 측정된 오디오 신호 내의 에너

지를 레코딩한다. 도 2의 시간축은 프레임들을 표기하지만, 임의의 다른 적절한 메트릭(metric), 예를 들어, 밀리초 등이 사용될 수 있다. 또한, 수직축 상에 도시되고 각자의 필터 बैं크들과 연관된 주파수 범위들이 원하는 경우 다른 구간들로 변경될 수 있거나, 또는 25 kHz를 넘어 확장될 수 있다는 점이 이해되어야 한다. 이러한 예시에서, 오디오 서명(24)은 상대적으로 높은 전력을 가지는 프레임-주파수 대역 쌍들을 나타내는 이진 행렬이다. 일단 생성되면, 오디오 서명(24)은 제1 디바이스(12)에 대해 도시되고 제2 디바이스(14)에 의해 레코딩된 프로그램 세그먼트를 특징으로 하고, 따라서, 오디오 서명(24)은 서버(18)가 액세스가능한 데이터베이스 내의 프로그램의 대응하는 세그먼트에 매칭될 수 있다.

[0016]

구체적으로, 서버(18)는 복수의 오디오 서명들 각각이 추출될 수 있는 데이터베이스에 동작가능하게 접속될 수 있다. 데이터베이스는 복수의 M개의 오디오 신호들  $s(t)$ 을 저장할 수 있고,  $s_m(t)$ 는 m번째 자산(asset)의 오디오 신호를 나타낸다. 각각의 자산 "m"에 대해, 오디오 서명들의 시퀀스  $\{S_m^*[f_n, b]\}$ 가 추출될 수 있고,  $\{S_m^*[f_n, b]\}$ 는 프레임 n과 n+F 사이의 신호  $s_m(t)$ 로부터 추출된 행렬이다. 데이터베이스 내의 대부분의 오디오 신호들이 대략 같은 지속기간을 가지며, 각각의  $s_m(t)$ 가 다수의 프레임들( $N_{max} \gg F$ )를 포함한다고 가정하면, 모든 M개의 자산들을 프로세싱한 이후, 데이터베이스는 대략  $MN_{max}$ 개 서명들을 가질 것이며, 이는 매우 큰 수( $10^7$  개 정도 또는 그 이상)일 것으로 예상된다. 그러나, 현대의 프로세싱 능력으로, 심지어 데이터베이스 내의 추출가능한 이 개수의 오디오 서명들이라도 제2 디바이스(14)로부터 수신된 오디오 서명(24)에 대한 매칭을 찾기 위해 신속하게 검색될 수 있다.

[0017]

데이터베이스에 대한 오디오 서명들이 사전-레코딩된 프로그램들에 대해 먼저, 또는 생방송 텔레비전 프로그램들에 대해 실시간으로 생성될 수 있다는 점이 이해되어야 한다. 또한, 오디오 신호들  $s(t)$ 을 저장하기보다는, 데이터베이스가, 각각이 제1 디바이스(12) 및 제2 디바이스(14)의 사용자에게 이용가능한 프로그래밍의 세그먼트와 연관된, 개별 오디오 서명들을 저장할 수 있다는 점이 이해되어야 한다. 또다른 실시예에서, 서버(18)는 각각이 전체 프로그램에 대응하는 개별 오디오 서명들을 저장할 수 있고, 따라서, 개별 세그먼트들은 서버(18)에 의한 질의 시에 생성될 수 있다. 또다른 실시예는 오디오 서명들이 생성될 오디오 스펙트로그램들을 저장할 것이다. 또한, 일부 실시예들이 오디오 서명들의 데이터베이스를 제2 디바이스(12) 상에 로컬로, 또는 예를 들어, 홈 네트워크 또는 로컬 영역 네트워크(LAN)를 통해 이용가능한 저장소에 저장하여, 원격 서버에 대한 필요성을 제거할 수 있다는 점이 이해되어야 한다. 이러한 실시예에서, 제2 디바이스(12) 또는 일부 다른 프로세싱 디바이스는 이 개시내용에 기술된 서버의 기능들을 수행할 수 있다.

[0018]

도 3은 서버(18)에 의해 기준 오디오 신호  $s(t)$ 로부터 생성된 스펙트로그램(26)을 도시한다. 이 스펙트로그램은, 제2 디바이스(14)에 의해 생성된, 스펙트로그램(22) 및 오디오 서명(24)에 의해 표현된 오디오 세그먼트에 대응한다. 스펙트로그램(26)을 스펙트로그램(22)과 비교함으로써 알 수 있는 바와 같이, 스펙트로그램(22)이 기준 신호와 연관된 오디오를 재생하는 텔레비전으로부터 먼 거리에 위치한 마이크로폰에 의해 레코딩된 오디오 신호로부터 생성되었다는 사실로 인해, 에너지 특성들은 스펙트로그램(22)에 거의 대응하지만, 스펙트로그램(22)에 비해 더 약하다. 도 3은 또한 기준 신호( $s(t)$ )로부터 서버(18)에 의해 생성된 기준 오디오 서명(28)을 도시한다. 서버(18)는 임의의 적절한 절차를 사용하여 오디오 서명(24)을 오디오 서명(28)에 정확하게 매칭시킬 수 있다. 예를 들어, 제2 디바이스(14)에 의해 획득되며 데이터베이스에 질의하기 위해 사용된 오디오 서명을  $S_q^*$  로서 표현하면, 서버 내의 기본 매칭 동작은 후속하는 의사-코드:

```

for m=1,...,M
  for n=1,...,Nmax-F
    score[n,m] = < Sm*[n] , Sq* >
  end
end

```

[0019]

[0020]

를 사용할 수 있고, 여기서, 동일한 차원들(dimensions)의 임의의 2개의 이진 행렬들 A 및 B에 대해, <A,B>는, A의 각각의 엘리먼트가 대응하는 B의 엘리먼트로 곱해지고, 합산된 엘리먼트들의 수로 나누어지는 행렬의 모든 엘리먼트들의 합으로서 정의된다. 이 경우, score[n,m]는  $S_m^*[n]$  및  $S_q^*$  모두에서 1인 엔트리들의 수와 동일하다. 모든 가능한 "m" 및 "n"에 대한 score[n,m]를 수집한 이후, 매칭 알고리즘은 제2 디바이스(14)에 의해

수집된 오디오가 가장 높은 score[n,m]에 대응하는 지연 f에서의 데이터베이스 신호 s<sub>m</sub>(t)에 대응한다고 결정한다.

[0021] 도 4를 참조하면, 예를 들어, 제2 디바이스(14)에 의해 캡처된 오디오로부터 생성된 오디오 서명(24)은 서버(18)에 의해 기준 오디오 서명(28)에 매칭되었다. 구체적으로, 이 도면에 도시된 화살표들은 2개의 오디오 서명들 사이의 오디오 에너지에서의 매칭 피크들을 도시한다. 에너지 내의 이들 매칭 피크들은 score[n,m]=9의 매칭 스코어를 가지고 기준 오디오 서명(28)을 정확하게 식별하기에 충분하였다. 매칭은 다수의 절차들 중 임의의 하나를 사용하여 선언될 수 있다. 위에서 주지된 바와 같이, 오디오 서명(24)은 서버(18)에서의 데이터베이스 내의 모든 오디오 서명, 및 가장 많은 매칭들을 가지는 저장된 서명과 비교될 수 있거나, 또는 그렇지 않은 경우, 임의의 적절한 알고리즘을 사용하여 가장 높은 스코어가 매칭 서명으로 간주될 수 있다. 이러한 기본 매칭 동작에서, 서버(18)는 "m"과 "n"의 모든 가능한 값들을 거침으로써, 가장 높은 score[n,m]를 생성하는 기준 "m" 및 지연 "n"을 검색한다.

[0022] 대안적인 절차에서, 데이터베이스는 사전-정의된 시퀀스에서 검색될 수 있고, 매칭 스코어가 고정된 임계치를 초과하는 경우 매칭이 선언된다. 이러한 기법을 용이하게 하기 위해, 검색 시간을 감소시키기 위해 해싱 연산이 사용될 수 있다. 오디오 서명 방법에 대해 적합한 많은 가능한 해싱 메커니즘들이 존재한다. 예를 들어, 단순한 해싱 메커니즘은 정수들의 세트 1, ..., F(F는 오디오 캡처 내의 프레임들의 수이며, 서명 행렬의 차원들 중 하나를 나타냄)를 G<sub>F</sub>개의 그룹들로 파티셔닝함으로써 시작하는데, 예를 들어, F=100, G<sub>F</sub>=5인 경우, 파티션은 {1, ..., 20}, {21, ..., 40}, ..., {81, ..., 100}일 것이다. 또한, 정수들의 세트 1, ..., B는 또한 G<sub>B</sub>개의 그룹들로 파티셔닝되며, B는 스펙트로그램 내의 대역들의 수이며, 서명 행렬의 또다른 차원을 나타낸다. 해싱 함수 H는 다음과 같이 정의된다: F x B 이진 행렬 S\*에 대해, HS\* = S' 이고, S'는 하나 이상의 엔트리들이 S\*의 대응하는 2차원 파티션 내에서 1과 같은 경우 각각의 엔트리(G<sub>F</sub>, G<sub>B</sub>)가 1과 동일한 G<sub>F</sub> x G<sub>B</sub> 이진 행렬이다.

[0023] 이 절차를 추가로 예시하기 위해 도 4를 참조하여, 그리드 라인들이 특정된 주파수 파티션들을 나타낸다고 가정하면, 디바이스(14)로부터 수신된 질의 서명(28)은 F=130, B=25인 반면 G<sub>F</sub>=13 및 G<sub>B</sub>=10임을 도시한다. 기준 서명(28)의 좌상단 파티션 내에 에너지 피크들이 존재하지 않기 때문에 해싱 연산에서 사용되는 행렬(S')의 엔트리 (1,1)는 0과 같다. 그러나, 파티션 (2.5,5) x (0,10)이 하나의 비제로 엔트리를 가지기 때문에 S'의 엔트리 (2,1)는 1과 같다. G<sub>F</sub>=13 및 G<sub>B</sub>=10가 위의 이러한 예에서 사용되었지만, G<sub>F</sub>=5 및 G<sub>B</sub>=4를 사용하는 것이 더 편리할 수 있다는 점이 이해되어야 한다. 대안적으로, 임의의 다른 값들이 사용될 수 있지만, 이들은 2<sup>{G<sub>F</sub>G<sub>B</sub>}</sup> << MN<sub>max</sub>이 되도록 해야 한다.

[0024] 데이터베이스 내의 모든 MN<sub>max</sub>개 서명들에 해싱 함수를 적용하는 경우, 데이터베이스는, 각각이 0들 및 1들의 행렬 A<sub>j</sub>에 의해 표현될 수 있는 2<sup>{G<sub>F</sub>G<sub>B</sub>}</sup>개 bin들로 파티셔닝되며, 여기서, j=1, ..., 2<sup>{G<sub>F</sub>G<sub>B</sub>}</sup>이다. bin 번호로 인덱싱된 표 T가 생성되고, 2<sup>{G<sub>F</sub>G<sub>B</sub>}</sup>개 bin들 각각에 대해, 표 엔트리 T[j]는 HS<sub>m</sub>\*[n]=A<sub>j</sub>를 만족시키는 서명들의 리스트 S<sub>m</sub>\*[n]를 저장한다. j의 다양한 값들에 대한 표 엔트리들 T[j]은 사전-레코딩된 프로그램들에 대한 시간 이전에, 또는 생방송 텔레비전 프로그램들에 대해 실시간으로 생성된다. 매칭 동작은 HS<sub>q</sub>\*에 의해 주어진 bin 엔트리를 선택함으로써 시작한다. 이후, 스코어는 엔트리 T[HS<sub>q</sub>\*] 내에 열거된 모든 서명들에 대해 S<sub>q</sub>\* 사이에서 계산된다. 충분히 높은 스코어가 발견되는 경우, 프로세스가 종결된다. 대안적으로, 충분히 높은 스코어가 발견되지 않는 경우, 프로세스는 그 행렬 A<sub>j</sub>이 해밍 거리 내의 HS<sub>q</sub>\*에 가장 가까운 bin들의 개별 bin들을 선택하고(해밍 거리는 2개의 이진 객체들 사이의 상이한 비트들의 수를 카운트한다), 스코어들은 엔트리 T[j] 내에 열거된 모든 서명들에 대해 S<sub>q</sub>\* 사이에서 계산된다. 충분히 높은 스코어가 발견되지 않는 경우, 프로세스는 그 행렬 A<sub>j</sub>이 해밍 거리 내의 HS<sub>q</sub>\*에 가장 가까운 다음 bin을 선택한다. 충분히 높은 스코어가 발견될 때까지 또는 검색들의 최대 수에 도달될 때까지 동일한 절차가 반복된다. 프로세스는 가장 높은 스코어를 가지는 기준 서명에 대해 어떠한 매칭도 선언되지 않거나 매칭이 선언된다고 결론짓는다. 위의 절차에서, 데이터베이스 내의 모든 저장된 콘텐츠에 대한 해싱 연산이 먼저 수행되므로(라이브 콘텐츠만이 실시간으로 해싱됨), 그리

고 정확한 서명을 포함할 가능성이 가장 큰 bin들 내에 열거된 서명들에 대해 매칭이 먼저 시도되므로, 매칭 프로세스의 검색 횟수 및 프로세싱 시간이 현저하게 감소된다.

[0025] 직관적으로 말해서, 해싱 연산은 "2-레벨 계층적 매칭"을 수행한다. 행렬  $HS_q^*$ 는 매칭들을 시도할 표 T의 bin들을 우선순위화하기 위해 사용되고, 우선순위는 그것의 연관된 행렬  $A_j$ 이 해밍거리 내의  $HS_q^*$ 에 더 가까운 bin들에 주어진다. 이후, 실제 질의  $S_q^*$ 는 충분히 높은 매칭이 발견될 때까지 우선순위화된 bin들 내에 열거된 서명들 각각에 대해 매칭된다. 매칭을 발견하기 위해 다수의 bin들을 검색하는 것이 필요할 수 있다. 도 4에서, 예를 들어, 실제 서명을 포함하는 bin에 대응하는 행렬  $A_j$ 이 "1"의 25개 엔트리들을 가지는 반면,  $HS_q^*$ 는 "1"의 17개의 엔트리를 가지며,  $HS_q^*$ 가 행렬  $A_j$ 로서 상이한 엔트리들에서 1들을 포함한다는 것을 아는 것이 가능하며, 그 역도 성립한다. 또한, 해싱을 사용하는 매칭 동작들은 초기 콘텐츠 식별 동안 그리고 재동기화동안만 요구된다. 오디오 서명들이 단지 사용자가 여전히 동일한 자산을 보고 있음을 확인하기 위해 캡처되는 경우, 기본 매칭 동작이 사용될 수 있다(이 시점에서  $M=1$ 이므로).

[0026] 캡처된 오디오 신호가 예를 들어, 높은 에너지 잡음에 의해 손상되지 않는 한, 제2 디바이스(14)에 의해 캡처된 오디오 서명을 원격 데이터베이스 내의 대응하는 서명들에 매칭시키는 이전 기법들은 양호하게 작용한다. 일 예로서, 제2 디바이스(14)가 텔레비전 또는 다른 이러한 제1 디바이스(12) 상에서 프로그램을 시청하는 하나 이상의 사람들에 근접할 경우, 사용자로부터의 높은 에너지 잡음(예를 들어, 스피킹, 노래, 또는 박수 잡음들)으로부터의 높은 에너지 잡음이 또한 마이크로폰(16)에 의해 픽업(pick up)될 수 있다. 다른 예들은 문닫기, 기차 지나가는 소리 등과 같은 유사한 우발적 소리들일 수 있다.

[0027] 도 5-6은 이러한 관련없는 잡음이 어떻게 캡처된 오디오의 오디오 서명을 손상시키고, 데이터베이스 내의 대응하는 서명에 대한 매칭에 악영향을 줄 수 있는지를 예시한다. 구체적으로, 도 5는 디바이스(14)의 마이크로폰(16)에 의해 캡처된, 동일한 프로그램 세그먼트의 오디오 서명(30)과 함께, 텔레비전 프로그램의 세그먼트에 대한 기준 오디오 서명(28)을 도시하며, 여기서 마이크로폰(16)은 또한 세그먼트 동안 사용자로부터의 잡음을 캡처한다. 참작될 수 있는 바와 같이, 사용자 생성 오디오는 마이크로폰(16)에 의해 레코딩된 세그먼트의 오디오 서명을 마스킹하고, 도 6에서 알 수 있는 바와 같이, 사용자 생성 오디오는 데이터베이스 내의 부정확한 서명이 매칭되는(또는 대안적으로, 매칭 서명이 발견되지 않음) 결과를 초래할 수 있다.

[0028] 도 7은 예시적인 파형들(34 및 40)을 도시하고, 오디오 세그먼트 각각은 제2 디바이스(14)의 마이크로폰(16)에 의해 캡처되고, 사용자는 구간들(36) 동안 각자 재채기하거나 이야기한다. 이들 구간들(36) 동안 사용자 생성 오디오는 서명이 요구되는 세그먼트의 오디오보다 통상적으로 약 40 dB 더 높은 피크들(38)을 가진다. 사용자 생성 오디오와 텔레비전으로부터의 오디오 신호 사이의 오디오 에너지에서의 이러한 통상적인 차이의 영향은 서명들이 피크들 사이의 시간차들의 다양한 시퀀스들에 의해 형성된 오디오 서명 추출 방법으로 평가되었으며, 각각의 시퀀스는 스펙트로그램의 특정 주파수 대역으로부터 기인한다. 도 8을 참조하면, 사용자 생성 오디오와 텔레비전 또는 다른 오디오 디바이스로부터의 오디오 신호 사이의 약 40dB의 이러한 통상적인 차이는 원격 데이터베이스에서 매칭 서명을 찾으려고 시도하는 경우 대략 65%의 성능 저하를 초래한다. 이 도면으로부터 알 수 있는 바와 같이, 심지어 단지 10dB의 차이가 성능을 50% 이상만큼 저하시킨다.

[0029] 사용자의 위치에서 생성된 오디오 서명과 원격 데이터베이스 내의 대응하는 기준 오디오 서명 사이의 정확한 매칭을 제공하는 것은, 오디오 캡처 서명을 손상시키는 관련없는 잡음의 존재시, 문제가 된다. 스펙트로그램으로부터 유도된 오디오 서명이 단지 신호 에너지 내의 피크들을 보존하며, 레코딩된 오디오 내의 잡음의 소스가 종종 레코딩된 것으로 발견된 신호보다 더 큰 에너지를 가지기 때문에, 스펙트로그램에 표현된 그리고 잡음에 의해 명백하게 손상된 오디오 신호의 일부분이, 만약 존재하는 경우, 쉽게 복원될 수 없다. 가능하게는, 마이크로폰(16)에 의해 캡처된 오디오 신호는 스펙트로그램을 생성하기 이전에 신호로부터 임의의 관련없는 잡음을 필터링할 것을 시도하도록 프로세싱될 수 있지만, 이러한 솔루션을 자동화시키는 것은 잡음의 존재를 예측불가능한 경우 어려울 것이다. 또한, 실제 프로그램 세그먼트들이 잡음(고함지르기 또는 폭발 등을 포함하는 세그먼트들)에 대해 잘못 취해질 가능성이 있는 경우, 임의의 효과적인 잡음 필터는 잡음을 정확하게 모델링할 능력에 의존할 가능성이 있다. 이는, 예를 들어, 하나의 마이크로폰이 (예를 들어, 사용자에서 지시됨으로써) 주로 잡음을 캡처하도록 구성되도록 제2 디바이스(14) 내에 다수의 마이크로폰들을 포함시킴으로써 달성될 수 있다. 따라서, 각자의 마이크로폰들에 의해 캡처된 오디오는 잡음을 모델링하고 그것을 필터링해내도록 사용될 수 있다. 그러나, 이러한 해결책은 증가한 비용 및 복잡도를 수반할 수 있고, 사용자 생성 오디오와 같은 잡음,

제2 디바이스(14)와 사용자 사이에 근접성을 고려하면, 레코딩되도록 의도된 오디오 신호를 여전히 손상시킨다.

[0030]

이러한 어려움들의 견지에서, 도 9는 (서버(18)에서와 같은) 먼 위치에 있는 데이터베이스 내의 기준 서명들과 (예를 들어, 디바이스(12)와 같은 제시 디바이스(presentation device)로부터 오디오 출력을 수신함으로써) 로컬로 생성된 오디오 서명들 사이에, 그리고 심지어 오디오 서명들이 손상된 스펙트로그램들, 예를 들어, 사용자 생성 오디오를 포함하는 오디오의 스펙트로그램들로부터 생성되는 경우, 정확한 매칭들을 가능하게 하는 신규한 시스템의 예를 예시한다. 용어 "손상"이 단지, 예를 들어, 마이크로폰(16)에 의해 수신된 임의의 오디오, 또는 디스플레이 디바이스(12)로부터의 주 오디오가 아닌 무언가로부터 기인한 스펙트로그램 또는 오디오 서명, 신호 또는 잡음에서 반영된 임의의 다른 정보를 지칭한다는 점이 이해되어야 한다. 후속하는 설명들이 일반적으로 사용자 생성 오디오를 지칭하지만, 이 발명의 실시예들이, 소비될 프로그램과 관련없는 임의의 다른 오디오에 적용되며, 이것이 사용자 생성 오디오에 의해 야기되는 손상을 다루기 위한 방법들 중 임의의 것이 또한 기기(appliances), 경적, 팡 닫히는 문, 장난감 등과 같은 잡음들에 의해 야기되는 손상을 다루도록 적용될 수 있음을 의미한다는 것이 이해되어야 한다. 일반적으로, 관련없는 오디오는 주 오디오가 아닌 임의의 오디오를 지칭한다. 구체적으로, 도 9는 클라이언트 디바이스(44)에 의해 송신된 오디오 서명들을 서버(46)에 동작가능하게 접속된 데이터베이스 내의 서명들에 매칭시키는, 클라이언트 디바이스(44)와 서버(46)를 포함하는 시스템(42)을 도시한다. 클라이언트 디바이스(44)는 태블릿, 랩톱, PDA 또는 다른 이러한 제2 디바이스(14)일 수 있고, 바람직하게는 오디오 서명 생성기(50)를 포함한다. 오디오 서명 생성기(50)는 클라이언트 디바이스(44)에 근접한 하나 이상의 마이크로폰들(16)에 의해 수신된 오디오로부터 스펙트로그램을 생성한다. 하나 이상의 마이크로폰들(16)은 바람직하게는 클라이언트 디바이스(44) 내로 통합되지만, 선택적으로, 클라이언트 디바이스(44)는 하나 이상의 외부 마이크로폰들에 접속할 수 있는 마이크로폰 잭 또는 무선 트랜시버와 같은 입력을 포함할 수 있다.

[0031]

이전에 주지된 바와 같이, 오디오 서명 생성기(50)에 의해 생성된 스펙트로그램은 예를 들어, 사용자로부터의 잡음에 의해 손상될 수 있다. 이 잡음을 정정하기 위해, 시스템(42)은 바람직하게는 하나 이상의 마이크로폰들(16)에 의해 수신된 오디오 신호를 입력으로서 가지는 오디오 분석기(48)를 또한 포함한다. 오디오 분석기(48)가 단순히 마이크로폰(16)으로부터 오디오 신호를 수신하는 것으로서 도시되지만, 마이크로폰(16)이 오디오 분석기(48)의 제어하에 있을 수 있는데, 이는 마이크로폰(16)을 활성화시키고 비활성화시키기 위한 명령들을 발행하여, 오디오 분석기(48) 및 오디오 서명 생성기(50)에 의해 후속적으로 다루어지는 오디오 신호를 초래한다. 오디오 분석기(48)는 임의의 잡음, 예를 들어, 사용자 생성 오디오의 존재 및 시간적 위치 모두를 식별하도록 오디오 신호를 프로세싱한다. 도 7과 관련하여 이전에 주지된 바와 같이, 신호 내의 잡음은 종종 신호 그 자체보다 훨씬 더 높은 에너지를 가질 수 있고, 따라서, 예를 들어, 오디오 분석기(48)는 신호 에너지에 대해 임계치 연산을 적용하여, 평균 신호 에너지의 일부 백분율보다 더 큰 오디오 서명의 부분들을 식별하고, 그 부분들을 잡음에 의해 손상된 것으로서 식별할 수 있다. 대안적으로, 오디오 분석기는 일부 고정 임계치보다 높은 수신된 오디오의 임의의 부분들을 잡음에 의해 손상된 것으로서 식별할 수 있거나, 또는 대안적으로, 예를 들어, 전용 제2 마이크로폰(16) 등으로부터의 잡음 모델 또는 오디오를 사용함으로써, 잡음의 오디오 신호 내의 존재 및 시간적 위치를 식별하기 위해 또다른 메커니즘을 사용할 수 있다. 오디오 분석기(48)가 사용자 생성 오디오의 존재 및 시간적 위치를 결정하기 위해 사용할 수 있는 대안적인 메커니즘은 수집된 오디오의 스펙트럼 특성들 내의 예상되지 않은 변경들을 관측하는 것일 수 있다. 예를 들어, 이전의 이력이 텔레비전에 의해 캡처된 오디오가 특정 스펙트럼 특성들을 가진다는 것을 나타내는 경우, 이러한 특성들에서의 변경들은 사용자 생성 오디오의 존재를 나타낼 수 있다. 오디오 분석기(48)가 사용자 생성 오디오의 존재 및 시간적 위치를 결정하기 위해 사용할 수 있는 또다른 대안적인 메커니즘은 스피커 검출 기법들을 사용하는 것일 수 있다. 예를 들어, 오디오 분석기(48)는 가정의 하나 이상의 사용자들에 대한 스피커 모델들을 구축할 수 있고, 캡처된 모델을 분석하는 경우, 수집된 오디오가 모델링된 스피커들로부터의 스피치(speech)를 포함하는 이들 스피커 모델들을 통해, 이들이 오디오 수집 프로세스 동안 발화하고 있으며, 따라서 텔레비전으로부터 수신된 오디오 내의 사용자-생성된 손상을 생성하고 있음을 나타낸다고 결정할 수 있다.

[0032]

일단 오디오 분석기(48)가 하나 이상의 마이크로폰들(16)에 의해 수신된 오디오 신호 내의 임의의 검출된 잡음의 시간적 위치를 식별하면, 오디오 분석기(48)는 그 정보를 오디오 서명 생성기(50)에 제공하며, 오디오 서명 생성기(50)는 그 정보를 사용하여 그것이 생성한 스펙트로그램의, 잡음에 의해 손상된 해당 부분들을 무효화할 수 있다. 이 프로세스는 도 10에 관련하여 일반적으로 기재될 수 있고, 통지되기에는 너무 약하게 만드는, 신호의 사용자 생성 오디오 대즐링(audio dazzling) 부분들을 포함하는 제1 스펙트로그램(52)을 도시한다. 이전에 지시된 바와 같이, 오디오 서명이 단순히 스펙트로그램(52)로부터 생성되는 경우, 오디오 서명은 도 10에 도시된 서버(46)에 의해 정확하게 매칭되지 않을 수 있다. 그러나, 오디오 서명을 생성하는 경우, 오디오 서명

생성기(50)는 오디오 분석기(48)로부터의 정보를 사용하여 세그먼트들(56)을 무효화시키거나 배제시킬 수 있다. 이것을 수행하기 위한 하나의 절차는 다음과 같다.  $S[f,b]$ 가 지속기간  $T$ , 예를 들어,  $T=120$  프레임, 5초 등을 가지는 신호  $s(t)$ 의 프레임 "f" 동안 대역 "b" 내의 에너지를 나타낸다고 하자. 모든 대역들( $b=1, \dots, B$ )이 변경되고 모든 프레임들( $f=1, \dots, F$ ) 이 신호  $s(t)$ 내에서 변경됨에 따라,  $S[f,b]$ 의 집합은  $F \times B$  행렬  $S$ 를 형성하며, 이는 신호의 스펙트로그램과 유사하다.  $F^{\wedge}$ 가 사용자 생성 오디오 또는 신호를 손상시키는 다른 이러한 잡음을 포함함에 따라 오디오 분석기(48)에 의해 식별된 영역들 내에 위치한 프레임들에 대응하는  $\{1, \dots, F\}$ 의 부분집합을 표기하도록 하고,  $S^{\wedge}$ 를 다음과 같이 정의된 행렬이라고 하자:  $f$ 가  $F^{\wedge}$  내에 있지 않은 경우, 모든  $b$ 에 대해  $S^{\wedge}[f,b]=S[f,b]$ 이고; 그렇지 않은 경우, 모든  $b$ 에 대해,  $S^{\wedge}[f,b]=0$ 이다.  $S^{\wedge}$ 로부터, 오디오 서명 생성기(50)는 이진  $F \times B$  행렬인 서명  $S_q^*$ 을 생성하며, 여기서,  $S^{\wedge}[f,b]$ 가  $S^{\wedge}$ 의 모든 엔트리들 중에서 가장 높은 에너지를 가진  $P\%$ (예를 들어,  $P=10\%$ ) 피크들 사이에 있는 경우,  $S_q^*[f,b]=1$ 이다. 단일 서명  $S_q^*$ 은 이후 오디오 서명 생성기(50)에 의해 매칭 서버(46)에 송신된다. 대안적으로, 오디오 서명 생성기가 세그먼트들(56)을 배제시키는 절차는 오디오 세그먼트에 대한 다수의 서명들(58)을 생성하는 것이며, 각각은 잡음에 의해 손상되지 않은 인접한 오디오 세그먼트들을 포함한다. 클라이언트 디바이스(44)는 이후, 데이터베이스에 저장된 기준 오디오 서명들에 별도로 매칭될 수 있는 이들 서명들(58) 각각을 서버(46)에 전송할 수 있으며, 매칭 결과들은 클라이언트 디바이스(44)에 반환된다. 클라이언트 디바이스(44)는 이후 매칭이 발견되었는지에 대한 결정을 수행하기 위해 매칭 결과들을 사용할 수 있다. 예를 들어, 서버(46)는 프로그램 내의 어디에서 매칭이 발견되었는지를 나타내는 그 프로그램 내의 시간 오프셋과 함께, 만약 존재하는 경우, 서명이 매칭된 프로그램의 식별 모두를 나타내는 하나 이상의 매칭 결과들을 반환할 수 있다. 클라이언트 디바이스는 이후, 이 경우에서, 서명들의 일부 정의된 백분율이 동일한 프로그램에 그리고 서로 충분히 가까운 시간적 구간들 내에 모두 매칭되는 경우 매칭을 선언할 수 있다. 매칭 세그먼트들이 이격되어야 하는 시간 구간들의 충분성을 결정할 시에, 클라이언트 디바이스(44)는 무효화된 세그먼트들의 시간 길이에 관한 정보, 즉, 동일한 프로그램에 대한 상이한 매칭들이 서버(46)에 송신된 오디오 서명들로부터 무효화된 세그먼트들의 지속기간과 거의 동일한 시간만큼 시간적으로 분리되는지를 선택적으로 사용할 수 있다. 대안적인 실시예가 서버(46)가 이 분석을 수행하게 하며, 만약 하나가 발견되는 경우, 클라이언트 디바이스(44)에 의해 송신된 서명들의 세트에 단일 매칭 프로그램을 단순히 반환할 수 있다는 점이 이해되어야 한다.

[0033] 위의 절차는 서명들이 이진 행렬들에 의해 형성되는 오디오 서명 추출 방법들에서 뿐만 아니라, 피크들 간의 시간차들의 다양한 시퀀스들에 의해 서명들이 형성될 있는 방법들에서 사용될 수 있고, 각각의 시퀀스는 스펙트로그램의 특정 주파수 대역으로부터 기인한다. 도 11은 후자의 경우 시스템(42)을 사용함으로써 획득된 성능 상의 개선을 일반적으로 도시한다. 알 수 있는 바와 같이, 시스템(42)이 사용되지 않는 경우, 성능은 신호 대 잡음 비에 따라 약 49% 내지 약 33% 사이의 임의의 곳으로 떨어진다. 그러나, 시스템(42)이 사용되는 경우, 사용자 생성 오디오와 같은 잡음의 존재시의 성능은 대략 79%로 증가한다.

[0034] 도 12는 클라이언트 디바이스(62) 및 매칭 서버(64)를 가지는 대안적인 시스템(60)을 도시한다. 클라이언트 디바이스(62)는 다시 태블릿, 랩톱, PDA, 또는 오디오 신호를 수신하고 그것을 프로세싱할 수 있는 임의의 다른 디바이스일 수 있다. 클라이언트 디바이스(62)는 바람직하게는 오디오 서명 생성기(66) 및 오디오 분석기(68)를 포함한다. 오디오 서명 생성기(66)는 클라이언트 디바이스(62)에 통합되거나 근접한 하나 이상의 마이크로폰들(16)에 의해 수신된 오디오로부터 스펙트로그램을 생성하고, 오디오 서명을 매칭 서버(64)에 제공한다. 이전에 언급된 바와 같이, 마이크로폰(16)은 오디오 분석기(68)의 제어 하에 있을 수 있고, 이는 마이크로폰(16)을 활성화 및 비활성화시키기 위한 명령들을 발행하여, 오디오 분석기(68) 및 오디오 서명 생성기(66)에 의해 후속적으로 처리되는 오디오 신호를 초래한다. 오디오 분석기(68)는 임의의 잡음, 예를 들어, 사용자 생성 오디오의 존재 및 시간적 위치 모두를 식별하도록 오디오 신호를 프로세싱한다. 오디오 분석기(68)는 그것의 분석에 의해 발견된 임의의 잡음의 존재 및 시간적 위치를 나타내는 정보를 서버(64)에 제공한다.

[0035] 서버(64)는 오디오 분석기(68)에 의해 제공된 결과들을 사용하여 오디오 서명 생성기(66)에 의해 제공된 오디오 서명을 매칭시키는 매칭 모듈(70)을 포함한다. 일 예로서, 이전에 설명된 바와 같이,  $S[f,b]$ 를 신호( $s(t)$ )의 프레임 "f" 동안 대역 "b" 내의 에너지를 나타낸다고 하고,  $F^{\wedge}$ 를 사용자 생성 오디오 또는 신호를 손상시키는 다른 이러한 잡음을 포함하는 것으로서 오디오 분석기(68)에 의해 식별된 영역들 내에 위치한 프레임들에 대응하는  $\{1, \dots, F\}$ 의 부분집합을 나타낸다고 하자; 매칭 모듈(70)은 잡음을 포함하는 것으로 결정된 수신된 오디오 서명의 부분들을 폐기할 수 있는데, 즉, 잡음에 의해 손상되지 않은 시간 구간 동안만 수신된 서명과 데이터베이스 내의 서명 간의 매칭 분석을 수행할 수 있다. 더 정확하게는, 매칭 스코어에서 사용된 질의 오디오 서명

$S_q^*$  은 다음과 같이 정의된  $S_q^{**}$  로 대체된다:  $f$ 가  $F^{\wedge}$ 내에 있지 않은 경우, 모든  $b$ 에 대해  $S_q^{**}[f,b]=S_q^*[f,b]$  이고;  $f$ 가  $F^{\wedge}$ 내에 있는 경우, 모든  $b$ 에 대해  $S_q^{**}[f,b]=0$  이고; 이전에 정의된 바와 같이, 최종 매칭 스코어는 연산  $\langle \dots \rangle$ 를 이용하여  $\langle S_m^*[n], S_q^{**} \rangle$  에 의해 주어진다. 이러한 예에서, 서버는 가장 높은 매칭 스코어(즉, 가장 많은 매칭들)를 가지는 데이터베이스로부터의 오디오 서명을 매칭 서명으로서 선택할 수 있다. 대안적으로, 매칭 모듈(70)은 일시적으로 상이한 매칭 스코어 함수를 채택할 수 있으며; 즉, 연산  $\langle S_m^*[n], S_q^* \rangle$  을 사용하는 것 대신, 매칭 모듈(70)은 대안적인 매칭 연산  $\langle S_m^*[n], S_q^* \rangle_{1^{\wedge}}$  을 사용하며, 여기서 2개의 이진 행렬들  $A$  및  $B$  사이의 연산  $\langle A, B \rangle_{1^{\wedge}}$  은 행렬의  $F^{\wedge}$ 에 포함되지 않은 열들(columns) 내의 모든 엘리먼트들의 합으로서 정의되며,  $A$ 의 각각의 엘리먼트는  $B$ 의 대응하는 엘리먼트로 곱해지고, 합산된 엘리먼트들의 수로 나누어진다. 이러한 후자의 대안에서, 매칭 모듈(70)은 실제로, 시간적으로 정규화된 스코어를 사용하여 임의의 배제된 구간들을 보상한다. 다시 말해, 정규화된 스코어는 서명의 전체 시간 구간으로 간주되고 있는(배제되지 않은) 서명의 시간 구간들의 비로 나누어진 매칭들의 수로서 계산되며, 정규화된 스코어는 임계치와 비교된다. 대안적으로, 정규화 절차는 단순히 단위 시간 당 매칭들의 임계치를 표현할 수 있다. 위의 예들 모두에서, 매칭 모듈(70)은 그 위로 매칭이 선언되는, 상이한 임계치 스코어를 채택할 수 있다. 일단 매칭 모듈(70)이 매칭을 식별하거나 매칭이 발견되지 않았다고 결정한 경우, 결과들은 클라이언트 디바이스(62)에 반환될 수 있다.

[0036]

도 9의 시스템은, 오디오 서명 생성 절차의 제어를 가지며 리거시 매칭 서버와 함께 작업해야 하는 경우 유용한 반면, 도 12의 시스템은, 매칭 절차의 제어를 가지며 리거시 오디오 서명 생성 절차들과 함께 작업해야 하는 경우 유용하다. 도 9 및 도 12의 시스템들이 일부 상황에서 양호한 결과들을 제공할 수 있지만, 사용자 생성 오디오의 존재에 관한 정보가 오디오 서명 생성기 및 매칭 모듈 모두에 제공되는 경우 추가적인 개선이 획득될 수 있다. 이러한 이점을 이해하고자, 이진 행렬이 스펙트로그램 내의  $P\%$  가장 강력한 피크들로부터 생성되는 위에 주어진 오디오 서명 알고리즘을 고려하며,  $F^{\wedge}$ 를 사용자 생성 오디오를 포함하는 것으로서 오디오 분석기에 의해 식별된 영역들 내에 위치한 프레임들에 대응하는  $\{1, \dots, F\}$ 의 부분 집합을 나타낸다고 하자.  $F^{\wedge}$ 가 오디오 서명 생성기에만 제공되는 경우, 도 9의 시스템에서와 같이,  $F^{\wedge}$  내의 프레임들은 이후 매칭 서버에 송신되는 서명을 생성하도록 무효화된다. 서명의 무효화된 부분들은 잘못된 프로그램을 이용한 높은 매칭 스코어의 생성을 회피한다. 결과적인 매칭 스코어는, 유실된 매칭을 초래하는 최소 매칭 스코어 임계치 아래에서 종료할 수 있다. 매칭 서버가 무효화된 부분들을 오디오 서명 내의 묵음(silence)인 것으로서 부정확하게 해석할 수 있기 때문에 잘못된 매칭이 또한 발생할 수 있다. 다시 말해, 오디오 서명의 부분들이 무효화되었음을 알지 않고는, 매칭 서버는 무효화된 부분들을 무효화된 구간들 동안 묵음 또는 다른 저-에너지 오디오를 가지는 서명들과 잘못 매칭시키려고 할 수 있다. 반면,  $F^{\wedge}$ 가 매칭 서버에만 제공되는 경우, 도 12에 관련하여 설명된 바와 같이, 서버는, 존재하는 경우, 어느 세그먼트들이 무효화되어야 할지를 결정하고, 따라서, 무효화된 시간 세그먼트들을 데이터베이스 내의 서명들에 매칭시키려고 시도하지 않음을 알 수 있지만;  $F^{\wedge}$  내의 프레임들 내의 피크들이 서명의 생성 동안 배제되지 않기 때문에, 이후  $P\%$  가장 강력한 피크들의, 전부가 아닌 경우 대부분이 사용자 생성 오디오를 포함하는 프레임들(즉,  $F^{\wedge}$  내의 프레임들) 내에 포함될 것이며, 생성된 오디오 서명 내의 "1"들의, 전부가 아닌 경우 대부분이  $F^{\wedge}$  내의 프레임들에 집중될 것이다. 후속적으로, 매칭 모듈이 서명 및  $F^{\wedge}$ 에 관한 정보를 수신함에 따라, 그것은  $F^{\wedge}$  내의 프레임들에 포함된 서명의 부분들을 무시한다. 이러한 프레임들이 무시됨에 따라, 서명 내의 나머지 프레임들 중 일부가 매칭 절차에서 사용될 "1"들을 포함하는 것이 발생할 수 있으며, 다시, 매칭 스코어는 감소된다. 이상적으로는,  $F^{\wedge}$ 는 오디오 서명 생성기와 매칭 모듈 모두에 제공되어야 한다. 이러한 경우, 오디오 서명 생성기는  $F^{\wedge}$  밖의 프레임들 내에서  $P\%$  가장 강력한 프레임들의 분포를 집중시킬 수 있고, 매칭 모듈은  $F^{\wedge}$  내의 프레임들을 무시하고, 높은 매칭 스코어들을 허용하기 위해 서명 내에 충분한 "1"들을 가질 수 있다. 또한, 매칭 모듈은  $F^{\wedge}$  내의 프레임들의 수에 관한 정보를 사용하여, 서명에서 제외된 프레임들을 설명할 정규화 상수를 생성할 수 있다.

[0037]

도 13은 오디오 서명 생성기와 매칭 모듈 모두에 사용자 생성 오디오에 관한 정보를 제공할 수 있는 또다른 대안적인 시스템(72)을 도시한다. 시스템(72)은 클라이언트 디바이스(74) 및 매칭 서버(76)를 가진다. 클라이언트 디바이스(72)는 다시, 태블릿, 랩톱, PDA, 또는 오디오 신호를 수신하고 그것을 프로세싱할 수 있는 임의의 다른 디바이스일 수 있다. 클라이언트 디바이스(72)는 바람직하게는 오디오 서명 생성기(78) 및 오디오 분석기(80)를 포함한다. 오디오 분석기(80)는, 이미 논의된 기법들을 사용하여, 클라이언트 디바이스(72)에 통합되거나 근접한 하나 이상의 마이크로폰들(16)에 의해 수신된 오디오를 프로세싱하여 임의의 잡음, 예를 들어, 사용

자 생성 오디오의 존재 및 시간적 위치 모두를 식별한다. 오디오 분석기(80)는 이후 오디오 서명 생성기(78) 및 매칭 모듈(82)에 정보를 제공한다. 이전에 언급된 바와 같이, 마이크로폰(16)은 오디오 분석기(80)의 제어 하에 있을 수 있고, 이는 마이크로폰(16)을 활성화하거나 비활성화하기 위한 명령들을 발행하여, 오디오 분석기(80) 및 오디오 서명 생성기(78)에 의해 후속적으로 처리되는 오디오 신호를 초래한다.

[0038] 오디오 서명 생성기(78)는 오디오 분석기(80)로부터 오디오 및 정보 모두를 수신한다. 오디오 서명 생성기(78)는 도 9의 시스템(42)의 기재에서 설명된 바와 같이, 단일 오디오 서명을 생성하는 경우 오디오 분석기(80)로부터의 정보를 사용하여 사용자 생성 오디오를 가진 세그먼트들을 무효화시키고, 단일 서명  $S_q^*$ 이 이후 오디오 서명 생성기(78)에 의해 매칭 서버(76)에 송신된다.

[0039] 매칭 모듈(82)은 오디오 서명 생성기(78)로부터 오디오 서명  $S_q^*$ 을 수신하고, 오디오 분석기(80)로부터 사용자 생성 오디오에 관한 정보를 수신한다. 이 정보는 사용자 생성 오디오를 포함하는 것으로서 오디오 분석기(80)에 의해 식별되었던 영역들 내에 위치한 프레임들의 집합  $F^*$ 에 의해 표현될 수 있다. 다른 기법들이 오디오 서명 내의 손상의 존재 및 위치를 나타내는 정보를 서버(76)에 송신하기 위해 사용될 수 있다는 점이 이해되어야 한다. 예를 들어, 오디오 서명 생성기(78)는  $F^*$ 에 포함된 프레임들에 걸쳐 오디오 서명  $S_q^*$  내의 모든 엔트리들을 "1"과 같도록 함으로써 매칭 모듈(82)에 집합  $F^*$ 를 통지할 수 있고; 따라서, 매칭 서버(76)가 열(column)이 "1"로서 마크된 모든 엔트리들을 가지는 이진 행렬을 수신하는 경우, 그것은 이러한 열에 대응하는 프레임들 매칭 절차로부터 배제될 프레임들의 집합  $F^*$ 의 일부분으로서 식별할 것이다.

[0040] 매칭 서버(76)는 클라이언트 디바이스(74)에 의해 수신된 오디오 서명을 매칭시킬 복수의 기준 오디오 서명들을 저장하는 데이터베이스에 동작가능하게 접속된다. 데이터베이스는 바람직하게는 도 2와 관련하여 기재된 것과 동일한 방식으로 구성될 수 있다. 매칭 서버(76)는 바람직하게는 매칭 모듈(82)을 포함한다. 매칭 모듈(82)은 오디오 서명  $S_q^*$  및 도 12의 시스템(60)에서 기술된 바와 같은 사용자 생성 오디오를 포함하는 프레임들의 집합  $F^*$ 에 관한 정보를 처리하며; 즉, 매칭 모듈(82)은 일시적으로 상이한 매칭 스코어 함수를 채택한다. 따라서, 전술된 바와 같은 기본 매칭 절차의  $score[n,m]$ 를 계산하기 위해 연산  $\langle S_m^*[n], S_q^* \rangle$ 을 사용하는 것 대신, 매칭 모듈(82)은 매칭 스코어 계산에 대해  $F^*$  내의 프레임들을 무시하는 대안적인 매칭 동작  $\langle S_m^*[n], S_q^* \rangle_{F^*}$ 를 사용할 수 있다.

[0041] 대안적으로, 해싱 절차가 매칭 동작 동안 요구되는 경우, 도 4에 대해 전술된 절차는 사용자 생성 오디오 정보를 다음과 같이 고려하도록 수정될 수 있다. 절차는 그것의 대응하는 행렬  $A_j$ 이  $HS_q^*$ 에 대해 가장 작은 해밍 거리를 가지는 빈 엔트리를 선택함으로써 시작하고, 해밍 거리는 이제  $F^*$  외부의 프레임들만을 고려하여 계산된다. 매칭 스코어는 이후  $S_q^*$ 와 선택된 빈에 대응하는 엔트리에 열거된 모든 서명들 사이에서 계산된다. 충분히 높은 스코어가 발견되지 않는 경우, 프로세스는 해밍 거리의 내림차순으로 다음 빈을 선택하고, 프로세스는 충분히 높은 스코어가 발견되거나 계산들의 최대 수에서의 제한에 도달될 때까지 반복된다.

[0042] 프로세스는 "매칭 없음" 선언으로 종결될 수 있거나, 또는 가장 높은 스코어를 가지는 기준 서명이 매칭으로 선언될 수 있다. 이 절차의 결과들은 클라이언트 디바이스(74)에 반환될 수 있다.

[0043] 오디오 서명 생성기(78) 및 매칭 모듈(82) 모두에 정보를 제공하는 것의 이점이 도 14에서 평가된다. 이 평가는 매칭 모듈(82)에서 사용자 생성 오디오를 포함하는 프레임들의 집합  $F^*$ 에 관한 지식을 가지는 이점에 초점을 두었다. 위에서 설명된 바와 같이, 이 정보가 이용가능하지 않고, 무효화된 엔트리들을 가지는 서명이 도달하는 경우, 서명의 일부분들의 무효화가 주어지면, 매칭 스코어가 감소한다. 도 14는  $F^*$ 에 관한 정보가 매칭 모듈(82)에 제공되지 않는 경우, 평균 매칭 스코어가 스코어 스케일에서 약 52임을 도시한다.  $F^*$ 에 관한 정보가 매칭 모듈(82)에 제공되어, 그것이  $F^*$  내의 프레임들의 수에 기초하여 매칭 스코어를 정규화하도록 하는 경우, 평균 매칭 스코어는 약 79로 증가한다. 따라서, 다른 방식으로는 오디오 캡처가 식별된 콘텐츠에 대응한다는 낮은 증거를 나타내는 낮은 매칭 스코어를 생성할 질의들이 이제 더 높은 매칭 스코어를 생성하고, 오디오 서명에 대한 무효화된 부분을 조정할 것이다.

[0044] 시스템(72)이 각자 도 9 및 도 12에서의 시스템들(42 및 60)에 대해 기술된 특징들 중 다수를 포함할 수 있다는 점이 이해되어야 한다. 비-제한적인 예들에서, 매칭 모듈(82)은 일련의 "1"들에 의해 손상된 부분들을 식별하

는 오디오 서명을 수신할 수 있고, 그 부분들을 사용하여 수신된 오디오 서명을 다수의 인접한 서명들로 세그먼트화하고, 해당 서명들을 데이터베이스 내의 기준 서명들에 별도로 매칭시킨다. 또한, 마이크로폰(16)이 도 9 및 도 12에 각자 표현된 시스템들의 오디오 분석기들(48 및 68)의 제어 하에 있음을 고려하면, 시스템(72)은, 데이터베이스에 질의하기 위해 사용된 오디오 서명의 시간 길이를 무효화된 부분들의 시간 길이와 동일한 구간, 또는 일부 다른 구간만큼 자동으로 그리고 선택적으로 연장함으로써(그리고 질의 서명이 대응하는 양에 의해 비교되는 기준 오디오 서명들의 길이를 연장함으로써) 오디오 서명의 무효화된 세그먼트들을 보상할 수 있다. 오디오 서명의 시간 길이의 연장은 오디오 서명 생성기 및 매칭 모듈 모두에 전달될 것이며, 이는 그에 따라 이들의 각자의 동작들을 연장할 수 있다.

[0045] 도 15 및 16은 일반적으로 사용자 생성 오디오의 형태로 잡음의 존재 시에 개선된 오디오 서명 생성이 가능한 시스템을 예시하며, 여기서, 2명의 사용자들은 텔레비전 세트와 같은 오디오 또는 시청각 디바이스(84)에 근접하며, 각각의 사용자는 전송된 방식들 중 임의의 방식으로, 손상(잡음)을 보상하는 시스템들이 구비된, 각각이 태블릿, 랩톱 등일 수 있는 상이한 디바이스(86 및 88)를 각자 가진다. 일반적으로 한번에 단 한 사람만 말하는 동안에 대화에 둘 이상의 사람들이 참여하는 경우 많은 사용자 생성 오디오가 발생하는 것이 관측된다. 이러한 환경에서, 말하는 사람에 의해 사용되는 디바이스(86 또는 88)는, 경우에 따라, 일반적으로 말하지 않는 사람에 의해 사용되는 디바이스보다 훨씬 더 많은 잡음을 픽업할 것이고, 따라서, 손상된 오디오에 관한 정보는 말하지 않는 사람의 디바이스(86 또는 88)로부터 복원될 수 있다.

[0046] 구체적으로, 도 16은 제1 클라이언트 디바이스(92a) 및 제2 클라이언트 디바이스(92b)를 포함하는 시스템(90)을 도시한다. 클라이언트 디바이스(92a)는 오디오 서명 생성기(94a) 및 오디오 분석기(96a)를 가질 수 있는 반면, 클라이언트 디바이스(92b)는 오디오 서명 생성기(94b) 및 오디오 분석기(96b)를 가질 수 있다. 따라서, 클라이언트 디바이스들 각각은 도 1, 9, 12 및 13에 대해 전송된 시스템들 중 임의의 것에 따라 매칭 서버(100) 및 기능과 독립적으로 통신할 수 있다. 다시 말해, 단독으로 동작하는 디바이스들 중 어느 것이라도, 이전에 개시된 기법들 중 임의의 것을 사용하여, 디바이스(84)로부터 오디오를 수신하고, 그것의 내부 오디오 분석기(96a 또는 96b)의 도움으로 또는 도움없이 서명을 생성하고, 그 서명을 매칭 서버에 전달하고, 응답을 수신할 수 있다.

[0047] 추가로, 그러나, 시스템(90)은, 오디오 분석기(92a) 및 오디오 분석기(92b) 모두의 결과들을 사용하여, 각자의 디바이스들(92a 및 92b)에 의해 생성된 오디오 서명들을 합성할 수 있는 적어도 하나의 그룹 오디오 서명 생성기(98)를 포함한다. 구체적으로, 시스템(90)은 각자의 디바이스들에 의해 생성된 오디오 서명들이 동일한 시간 구간들을 포함하도록 2개의 디바이스들(92a 및 92b)을 동기화할 수 있다. 이러한 동기화를 통해, 그룹 오디오 서명 생성기(98)는 하나의 디바이스(92a 또는 92b)에 의해 생성된 오디오 서명의 임의의 부분들이 잡음으로서 분석된 시간 세그먼트들을 가지는지를 결정할 수 있지만, 다른 디바이스(92a 또는 92b)의 오디오 서명 내의 동일한 구간은 잡음이 아닌 것(즉, 신호)으로서 분석되며, 그 역도 성립한다. 이러한 방식으로, 그룹 오디오 서명 생성기(98)는 각자의 디바이스들(92a 및 92b) 각각에 의해 인입 오디오 신호의 개별 분석들을 생성하여, 디바이스들(92a 및 92b) 중 어느 하나가 단독으로 생성할 수 있는 것보다 더 깨끗한 오디오 서명을 구간에 걸쳐 생성할 수 있다. 그룹 오디오 서명 생성기(98)는 이후, 개선된 서명을 매칭 서버(100)에 포워딩하여, 데이터베이스 내의 기준 서명들을 비교할 수 있다. 이러한 작업을 수행하기 위해, 오디오 분석기들(96a 및 96b)은 오디오 서명들의 조합을 수행하고 전송한 더 깨끗한 오디오 서명을 생성하게 하기 위해 그룹 오디오 서명 생성기(98)에 미가공(raw) 오디오 특징들을 포워딩할 수 있다. 이러한 미가공 오디오 특징들은 디바이스들(92a 및 92b)에 의해 캡처된 실제 스펙트로그램들, 또는 이러한 스펙트로그램들의 함수를 포함할 수 있고; 또한, 이러한 미가공 오디오 특징들은 또한 실제 오디오 샘플들을 포함할 수 있다. 이러한 마지막 대안에서, 그룹 오디오 서명 생성기는 오디오 서명을 생성하기 이전에 오디오 소거 기법들을 사용할 수 있다. 더 정확하게는 그룹 오디오 서명 생성기(98)는 더 적은 사용자 생성 오디오를 포함하는 단일 오디오 세그먼트를 생성하고, 매칭 모듈에 송신될 단일 오디오 서명을 생성하기 위해, 디바이스들(92a 및 92b) 모두에 의해 캡처된 오디오 세그먼트의 샘플들을 사용할 수 있다.

[0048] 그룹 오디오 서명 생성기(98)는 디바이스들(92a 및 92b) 중 어느 하나, 또는 둘 모두에 존재할 수 있다. 하나의 예에서, 디바이스들(92a 및 92b) 각각은 그룹 오디오 서명 생성기(98)를 호스팅할 수 있고, 디바이스들(92a 및 92b)의 사용자들은 사용자 인터페이스를 통해, 어느 디바이스가 그룹 오디오 서명 생성기(98)를 호스팅할지를 선택하도록 촉구되고, 선택 시, 매칭 서버와의 모든 통신은, 이 협업적 모드가 어느 한 사용자에게 의해 선택 해제될 때까지, 또는 디바이스들(92a 및 92b)이 서로와의 통신을 중단할 때까지(예를 들어, 하나의 디바이스가 턴오프되거나, 또는 다른 방에 가져가지는 등), 선택된 호스트 디바이스(92a 또는 92b)를 통해 진행할 수 있다. 대안적으로, 자동화된 절차는 어느 디바이스(92a 또는 92b)가 그룹 오디오 서명 생성기를 호스팅할지를 무작위

로 선택할 수 있다. 또한, 그룹 오디오 서명 생성기는 디바이스들 모두(92a 및 92b)와 통신하는 독립형 디바이스일 수 있다. 통상의 기술자는 또한 이 시스템이 둘보다 많은 클라이언트 디바이스들을 포함하도록 용이하게 확장될 수 있음을 이해할 것이다.

[0049] 또한, 도 9, 도 12, 도 13, 또는 도 16의 시스템들 중 임의의 것에서, 대안적인 실시예가 상이한 디바이스들 내에 오디오 분석기 및 오디오 서명 생성기를 위치시킬 수 있다는 점이 이해되어야 한다. 이러한 실시예에서, 오디오 분석기 및 오디오 서명 생성기 각각은 자신만의 마이크로폰을 가질 것이며, 이들이 매칭 서버와 통신하는 동일한 방식으로 서로 많이 통신할 수 있을 것이다. 추가적인 대안적인 실시예에서, 오디오 분석기 및 오디오 서명 생성기는 동일한 디바이스에 위치되지만, 서로 통신하는 별도의 소프트웨어 프로그램들 또는 프로세스들이다.

[0050] 또한, 기준 서명들에 대해 오디오 서명들을 매칭시키는 전술한 시스템들 중 몇몇이 손상된 세그먼트들을 무효화 시킴으로써 오디오 서명들 내의 손상들을 보상하지만, 본 개시내용과 부합하는 다른 시스템들은 손상을 다루기 위한 대안적인 기법들을 사용할 수 있다는 점이 이해되어야 한다. 일 예로서, 도 1의 디바이스(14), 도 9의 디바이스(44), 또는 도 12의 디바이스(62)와 같은 클라이언트 디바이스는, 후속하는 질의된 오디오 서명들을 이전에 매칭된 프로그램으로부터의 오디오 서명들과 초기에 비교함으로써, 일단 매칭 프로그램이 초기에 발견되면 프로세싱 능력을 절감하도록 구성될 수 있다. 다시 말해, 사용자가 채널들을 스위칭하지 않거나 트릭 플레이 모드, 예를 들어, 빨리 감기 등에 진입하지 않았다는 가정이 주어지는 경우, 매칭 프로그램이 초기에 발견된 이후, 후속적으로 수신된 오디오 서명들은 클라이언트 디바이스에 전송되고, 그 서명을 해당 시점에 예상된 기준 서명과 비교함으로써 동일한 프로그램이 사용자에게 여전히 제시됨을 확인하기 위해 사용된다. 수신된 서명이 기대된 세그먼트에 대한 매칭이 아닌 경우에만, 사용자가 트릭 플레이 모드에 진입했는지를 먼저 결정하고, 그렇지 않은 경우, 수신된 서명을 다른 프로그램들의 기준 서명들과 비교함으로써 다른 어느 프로그램이 사용자에게 의해 시청될 수 있는지를 결정하도록 시도하는 것이 필요해진다. 이 기법은 본 출원의 양수인에 의해 2012년 6월 26일에 출원된 공동-계류중인 출원 일련 번호 제131/533,309호에 개시되어 있으며, 그 개시내용은 전체가 참조로 본원에 포함된다.

[0051] 이러한 기법들이 주어지면, 클라이언트 디바이스는, 사용자에게 의해 시청되거나 청취되는 프로그램을 초기에 식별한 이후, 프로그램으로부터 다음에 올(still-to-come) 오디오 세그먼트들에 대응하는 오디오 서명들의 시퀀스를 수신할 수 있다. 이러한 다음에 올 오디오 서명들은 프로그램이 사전-레코딩되었을 때 원격 서버로부터 용이하게 이용가능하다. 그러나, 심지어 프로그램이 라이브일 때도, 방송 네트워크를 통한 프로그램의 전송에서 제로가 아닌(non-zero) 지연이 존재하고; 따라서, 다음에 올 오디오 서명들을 생성하여 이들을 클라이언트 디바이스의 매칭 동작이 시도되기 이전에 클라이언트 디바이스에 전송하는 것이 여전히 가능하다. 이러한 다음에 올 오디오 서명들은 사용자가 선형 방식으로 동일한 프로그램을 계속 시청하는 경우 클라이언트 디바이스에서 생성되도록 예상되는 오디오 서명들이다. 이들 다음에 올 오디오 서명들을 수신한 경우, 클라이언트 디바이스는 오디오 샘플들을 수집하고, 오디오 특징들을 추출하고, 오디오 서명들을 생성하고, 이들을 저장된 예상 오디오 서명들과 비교하여 사용자가 동일한 프로그램을 여전히 시청하거나 청취하고 있음을 확인할 수 있다. 다시 말해, 오디오 서명 생성 및 매칭 절차들 모두는 이 절차 동안 클라이언트 디바이스 내에서 수행된다. 이 절차 동안 생성된 오디오 서명들이 또한 사용자 생성 오디오에 의해 손상될 수 있으므로, 도 9, 도 12 또는 도 13의 시스템들의 방법들은, 심지어 오디오 서명 생성기, 오디오 분석기, 및 매칭 모듈이 클라이언트 디바이스에 위치되더라도, 여전히 적용될 수 있다.

[0052] 대안적으로, 이러한 기법들에서, 오디오 신호의 손상은 사용자 생성 오디오와 같은 손상의 존재 또는 부재를 먼저 식별함으로써 보상될 수 있다. 이러한 잡음 또는 다른 손상이 식별되는 경우, 오디오 서명이 수신될 때까지 매칭에서의 어떠한 초기 시도도 이루어지지 않을 수 있으며, 여기서 오디오의 분석은 잡음이 존재하지 않음을 나타낸다. 유사하게, 일단 초기 매칭이 이루어지면, 잡음을 포함하는 임의의 후속적인 오디오 서명들은 무시될 수 있거나, 또는 대안적으로, 매칭을 검증하기 위해 그 시점에서 예상된 세그먼트의 오디오 서명과 비교될 수 있다. 그러나, 어느 경우든, 예를 들어, 잡음에 의해 손상된 오디오 서명 사이에 "매칭 없음"이 선언되는 경우, 사용자가 트릭 플레이모드에 진입했는지 또는 채널들을 스위칭했는지에 대한 결정은 잡음을 포함하지 않는 서명이 수신될 때까지 연기된다.

[0053] 오디오 서명에서의 손상의 보상에 대한 전술한 논의가 신호 내에 잡음을 유입한 사용자 생성 오디오의 예를 사용하여 예시되었지만, 다른 형태들의 손상이 가능하며, 전술된 기법들을 사용하여 용이하게 보상될 수 있다는 점이 또한 이해되어야 한다. 프로그래밍 콘텐츠를 전달하는 위성 안테나(satellite dish) 시스템들은 강풍, 비 등으로 인해 짧은 신호 끊김을 빈번하게 겪으며, 오디오 신호들은 간단히 산발적일 수 있다. 또다른 예로서,

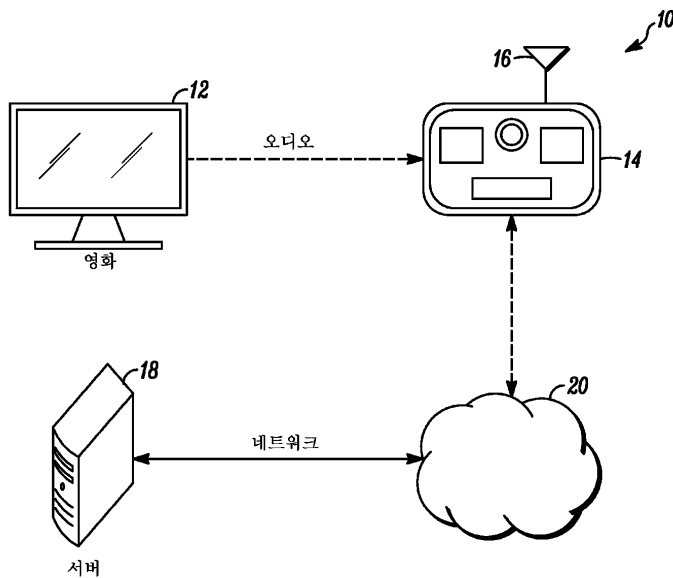
DRV 상에 저장되거나 DVD 상에서 재생되는 프로그래밍 콘텐츠가 데이터베이스 내의 프로그래밍 콘텐츠에 매칭되는 경우, 오디오 신호는 디지털 저장 매체의 결함으로 인해 손상될 수 있다. 그러나, 어느 경우든, 이러한 손상은 모델링될 수 있고, 따라서, 전술한 바와 같이 식별되고 보상될 수 있다.

[0054]

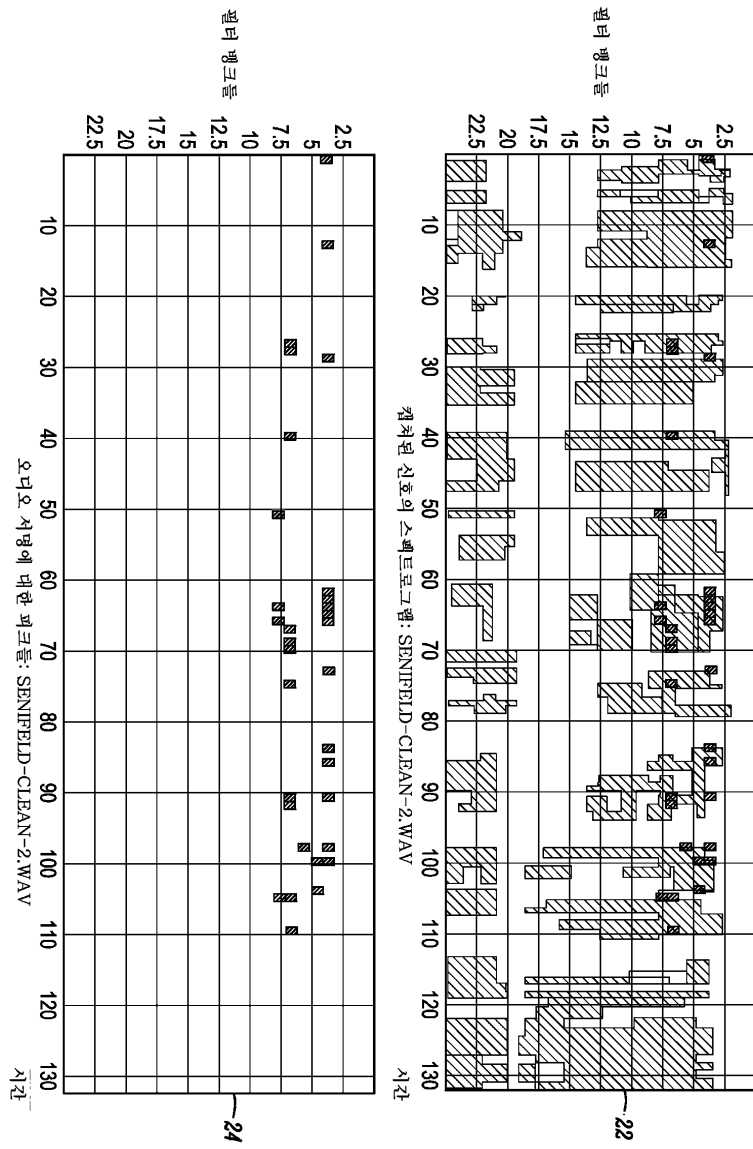
개시내용이 기재된 특정 실시예로 제한되지 않으며, 등가물들 또는 문자상의 범위를 초과하여 청구항의 실시가능한 범위를 확장시키는 임의의 다른 원리들의 원칙을 포함하여, 통용법의 원리들에 따라 해석되는 바와 같이, 개시내용 및 첨부된 청구항들의 범위로부터 벗어나지 않고 변경들이 이루어질 수 있다는 점이 이해될 것이다. 문맥이 달리 지시하지 않는 한, 구성요소의 인스턴스의 수에 대한 청구항에서의 인용은 하나의 인스턴스 또는 하나보다 많은 인스턴스에 대한 인용이며, 그것이 구성요소의 적어도 언급된 개수의 인스턴스들을 요구하지만, 언급된 것이 아닌 구성요소의 더 많은 인스턴스들을 가지는 구조 또는 방법이 청구항의 범위로부터 배제되도록 의도되지 않는다. 용어 "포함하다" 또는 그 파생어는, 청구항에서 사용되는 경우, 청구된 구조 또는 방법 내의 다른 구성요소들 또는 단계들의 존재를 배제하도록 의도되지 않는 비배타적 의미로 사용된다.

**도면**

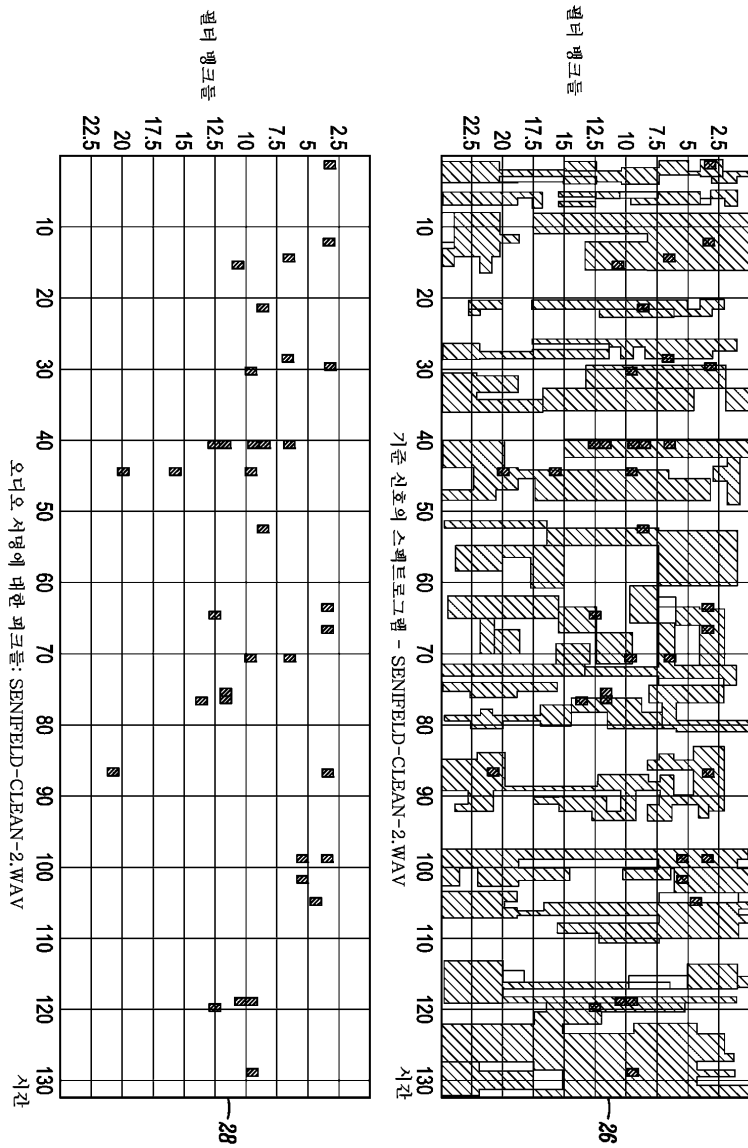
**도면1**



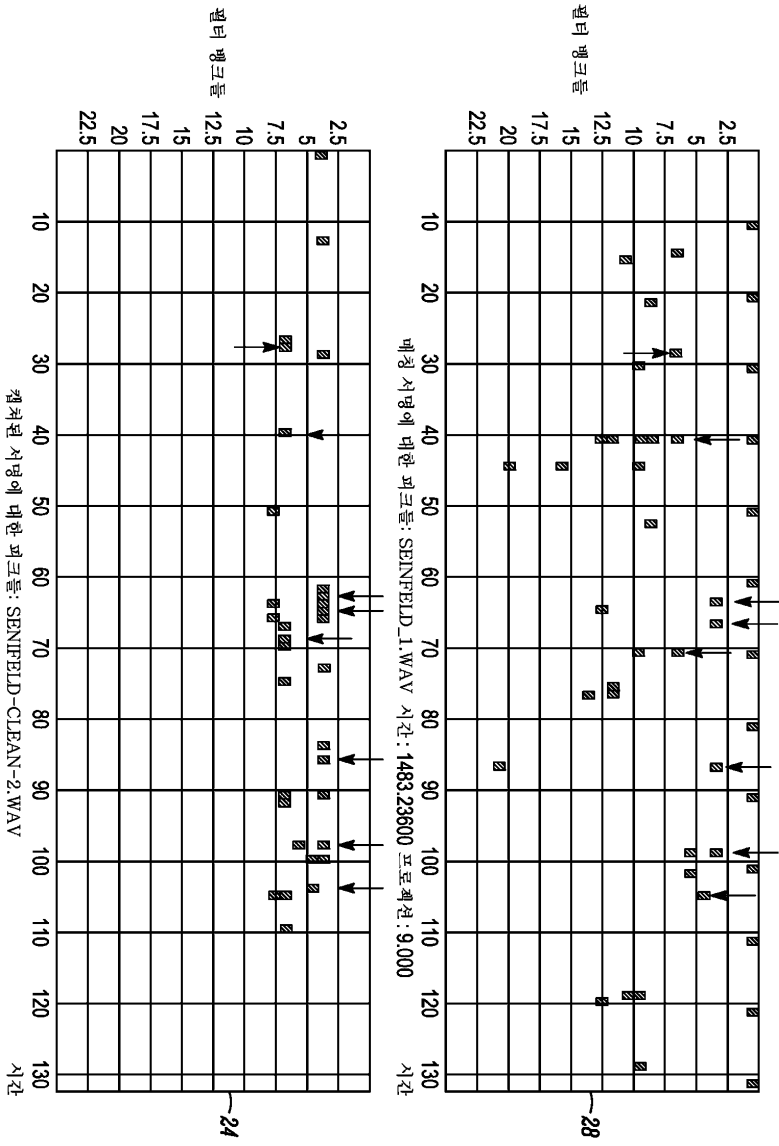
도면2



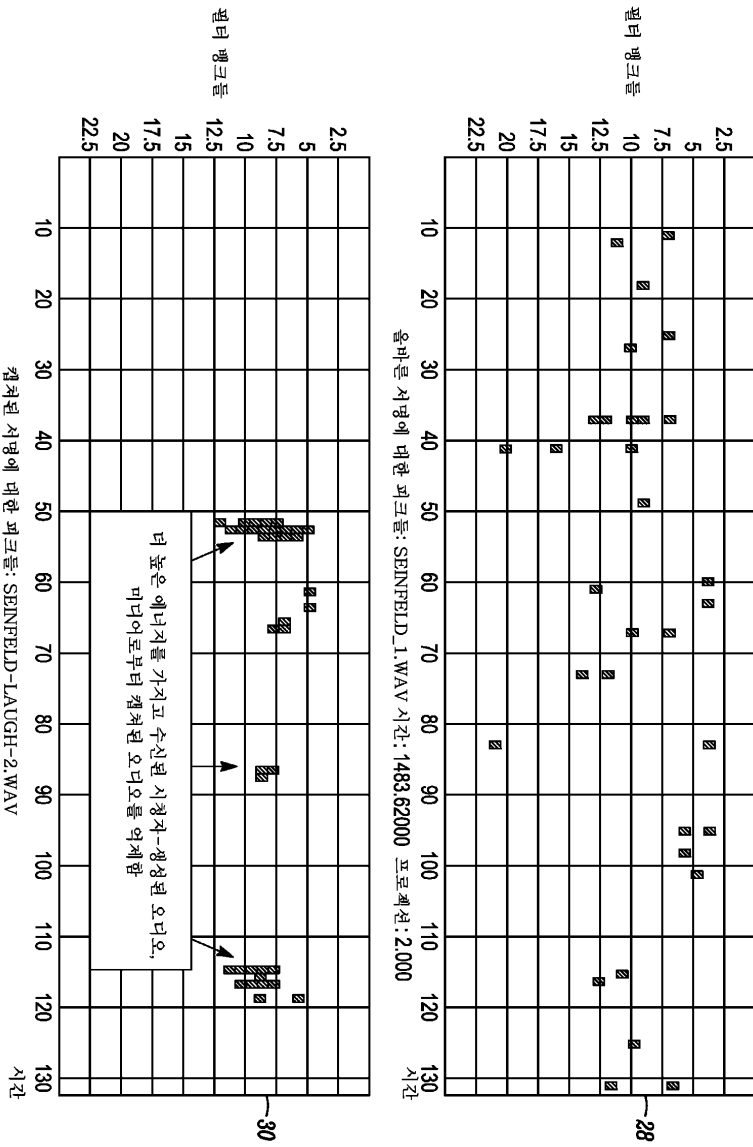
도면3



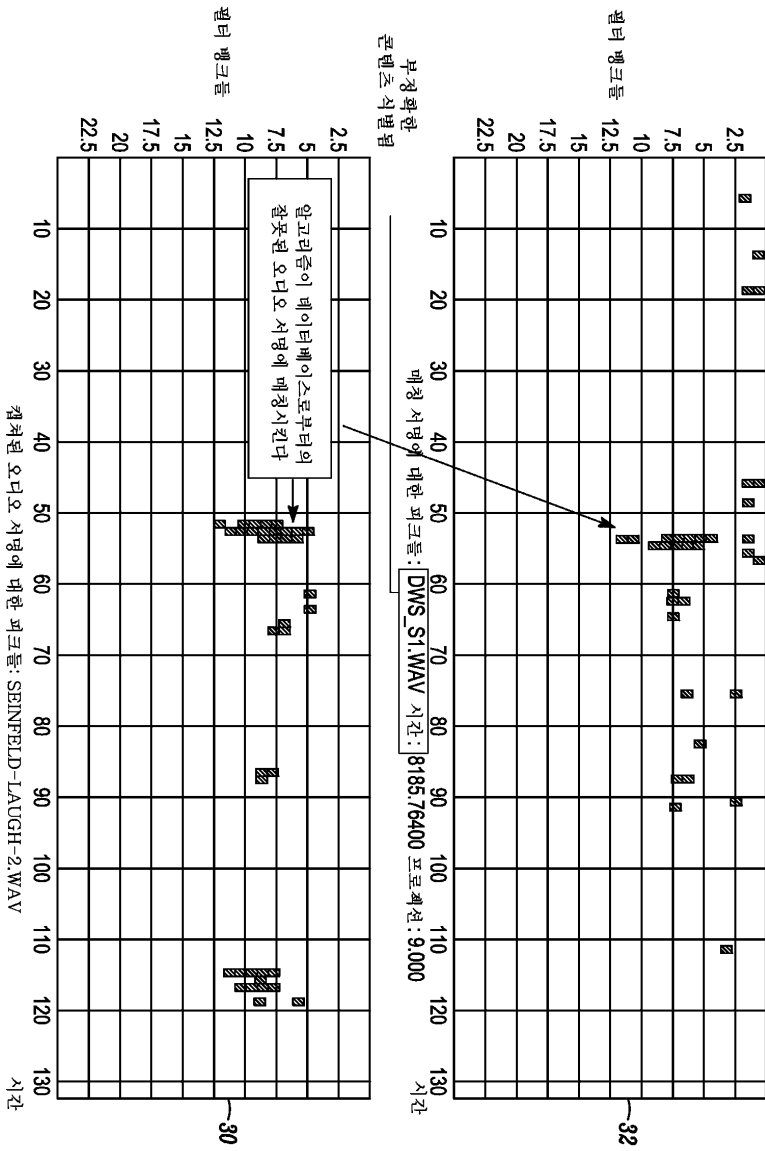
도면4



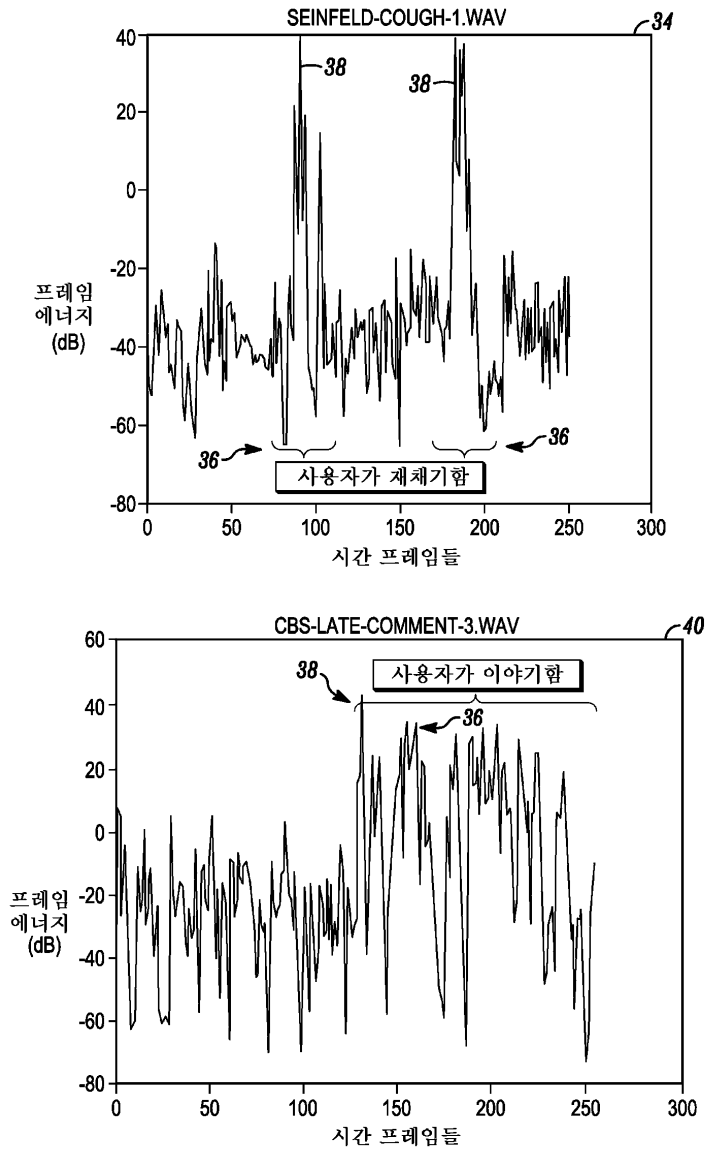
도면5



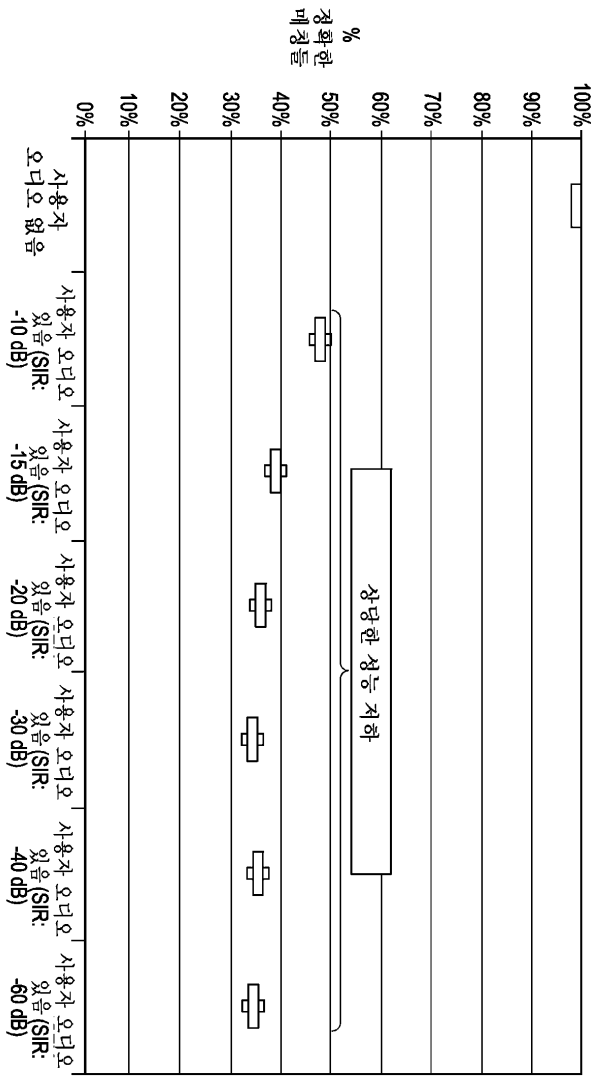
도면6



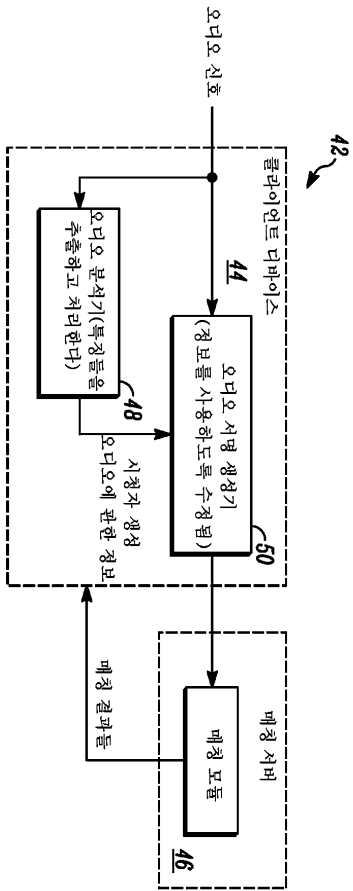
도면7



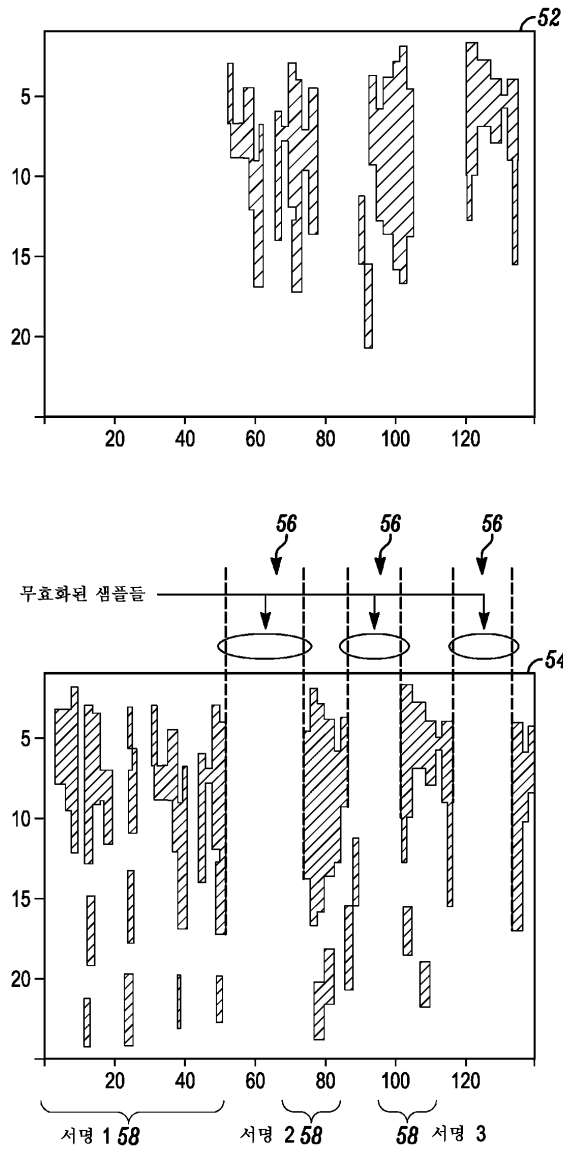
도면8



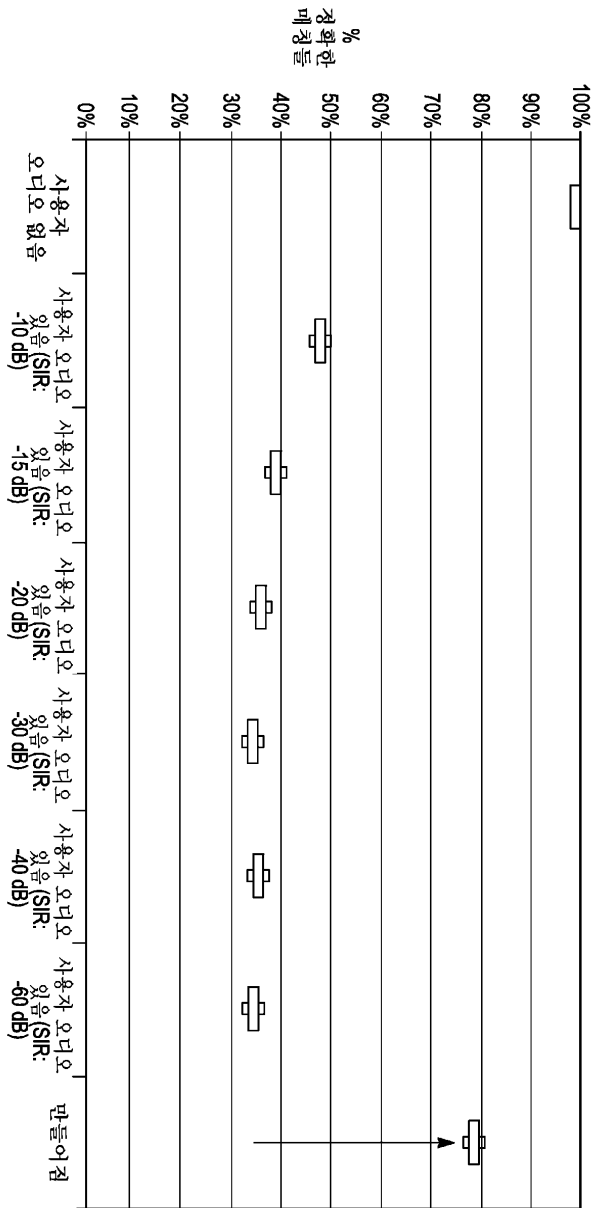
도면9



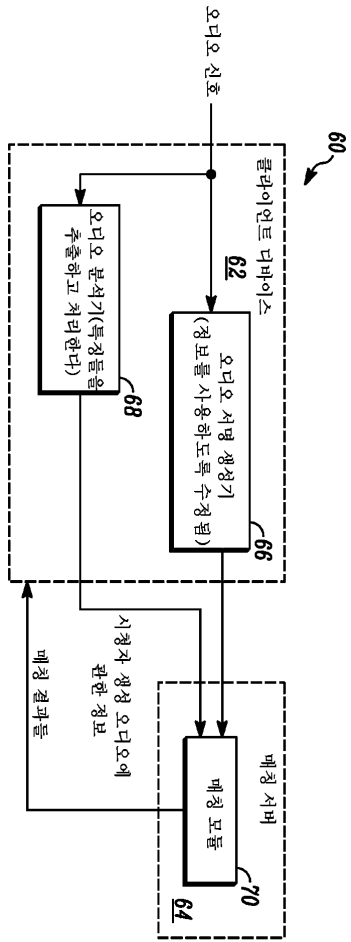
도면10



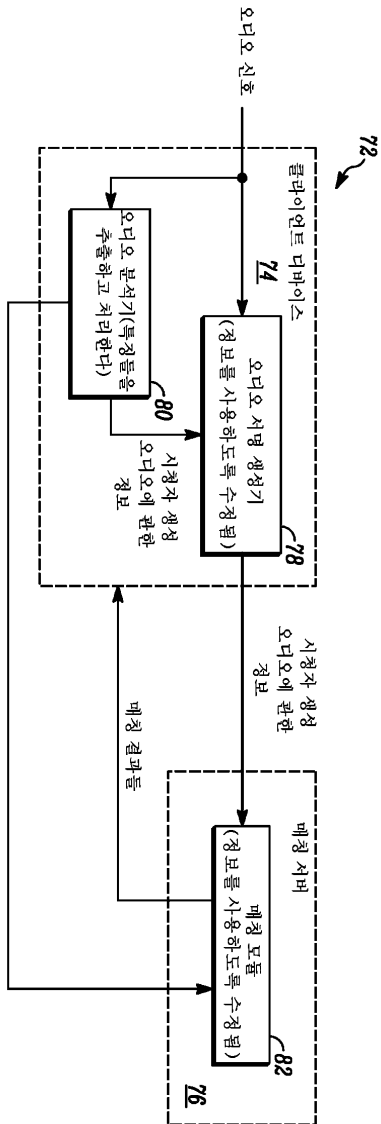
도면11



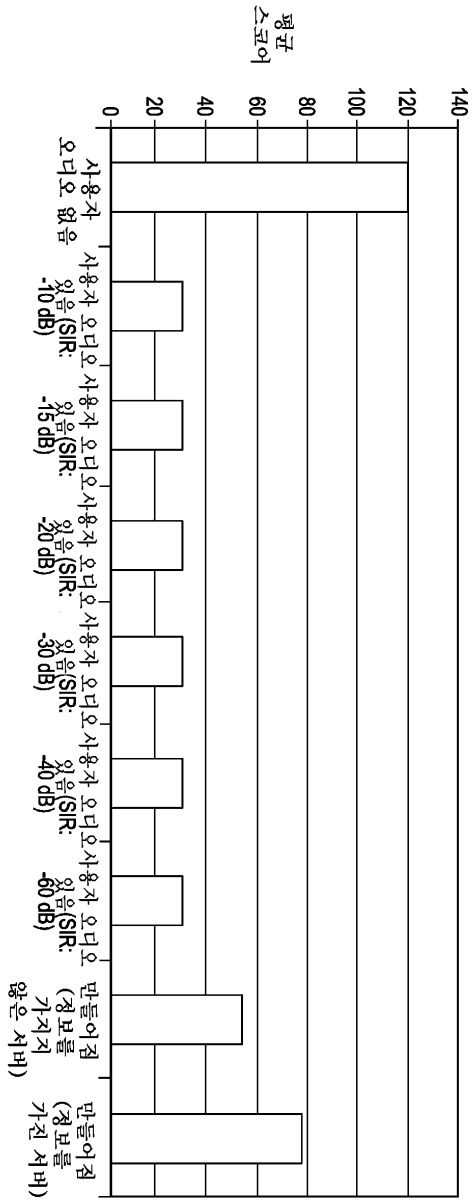
도면12



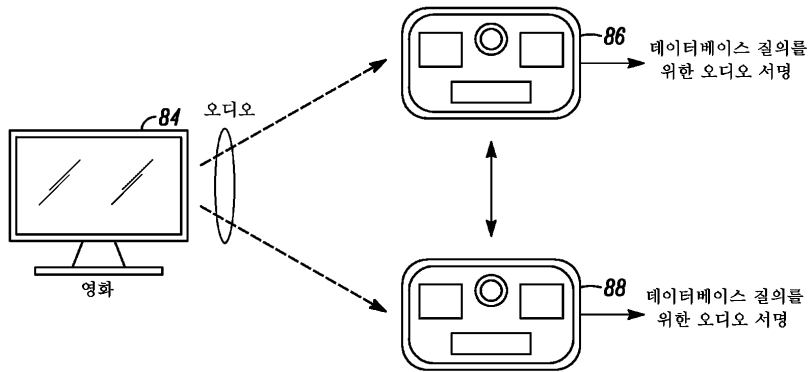
도면13



도면14



도면15



도면16

