(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2025년02월11일
(11) 등록번호 10-2765838
(24) 등록일자 2025년02월05일

(51) 국제특허분류(Int. Cl.)
H04N 21/439 (2011.01) *G06F 3/16* (2018.01)
G10L 15/06 (2006.01) *G10L 15/08* (2006.01)
G10L 15/18 (2006.01) *G10L 15/22* (2006.01)
G10L 15/26 (2006.01) *H04N 21/435* (2011.01)
(52) CPC특허분류
H04N 21/4398 (2013.01)
G06F 3/167 (2013.01)
(21) 출원번호 10-2021-7011130
(22) 출원일자(국제) 2020년06월09일
심사청구일자 2023년06월07일
(85) 번역문제출일자 2021년04월14일
(65) 공개번호 10-2023-0021556
(43) 공개일자 2023년02월14일
(86) 국제출원번호 PCT/US2020/036749
(87) 국제공개번호 WO 2021/251953
국제공개일자 2021년12월16일
(56) 선행기술조사문헌
KR1020200007095 A*
US20090254345 A1*
US20170229040 A1*
WO2019216969 A1*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
구글 엘엘씨
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
(72) 발명자
샤리피 매튜
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
카부네 빅터
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
(74) 대리인
박장원

전체 청구항 수 : 총 40 항 심사관 : 선동국

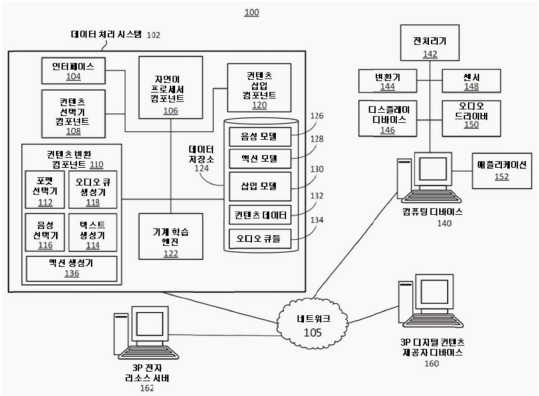
(54) 발명의 명칭 **시각적 컨텐츠로부터 대화형 오디오 트랙 생성**

(57) 요 약

오디오 트랙을 생성하는 것이 제공된다. 시스템은 시각적 출력 포멧을 갖는 디지털 컴포넌트를 선택한다. 시스템은 디지털 컴포넌트 객체를 오디오 출력 포멧으로 변환하기로 결정한다. 시스템은 디지털 컴포넌트 객체에 대한 텍스트를 생성합니다. 시스템은 디지털 컴포넌트 객체의 컨텍스트에 기초하여 텍스트를 렌더링하기 위한 디지털

(뒷면에 계속)

대 표 도



음성을 선택한다. 시스템은 디지털 음성에 의해 렌더링된 텍스트를 사용하여 디지털 컴포넌트 객체의 기준 오디오 트랙을 구성한다. 시스템은 디지털 컴포넌트 객체에 기초하여 비-음성 오디오 큐를 생성한다. 시스템은 디지털 컴포넌트 객체의 오디오 트랙을 생성하기 위해 비-음성 오디오 큐를 디지털 컴포넌트 객체의 기본 오디오 형식과 결합한다. 시스템은 컴퓨팅 디바이스의 스피커를 통해 출력하기 위해 컴퓨팅 디바이스에 디지털 컴포넌트 객체의 오디오 트랙을 제공한다.

(52) CPC특허분류
G10L 15/063 (2013.01)
G10L 15/083 (2013.01)
G10L 15/1822 (2013.01)
G10L 15/22 (2013.01)
G10L 15/26 (2013.01)
H04N 21/435 (2013.01)
H04N 21/4394 (2013.01)
G10L 2015/088 (2013.01)

명 세 서

청구범위

청구항 1

상이한 모달리티간 트랜지션 시스템으로서,

데이터 처리 시스템을 포함하고, 상기 데이터 처리 시스템은:

네트워크를 통해, 데이터 처리 시스템으로부터 멀리 떨어진 컴퓨팅 디바이스의 마이크로폰에 의해 검출된 입력 오디오 신호를 포함하는 데이터 패킷을 수신하고;

요청을 식별하기 위해 입력 오디오 신호를 파싱하고;

요청에 기초하여, 시각적 출력 포맷을 갖는 디지털 컴포넌트 객체를 선택하고 - 상기 디지털 컴포넌트 객체는 메타 데이터와 연관됨 -;

컴퓨팅 디바이스의 유형에 기초하여, 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정하고;

디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 한 결정에 응답하여, 디지털 컴포넌트 객체에 대한 텍스트를 생성하고;

디지털 컴포넌트 객체의 컨텍스트를 처리하는 음성 모델에 의해 생성된 음성 특성 벡터에 기초하여, 복수의 디지털 음성으로부터 텍스트를 렌더링하기 위한 디지털 음성을 선택하고;

디지털 음성에 의해 렌더링된 텍스트로 상기 디지털 컴포넌트 객체의 기본(baseline) 오디오 트랙을 구성하고;

디지털 컴포넌트 객체에 기초하여, 비-음성 오디오 큐를 생성하고;

디지털 컴포넌트 객체의 오디오 트랙을 생성하기 위해 비-음성 오디오 큐를 디지털 컴포넌트 객체의 기본 오디오 형식과 결합하고; 그리고

컴퓨팅 디바이스의 요청에 응답하여, 컴퓨팅 디바이스의 스피커를 통한 출력을 위해 디지털 컴포넌트 객체의 오디오 트랙을 컴퓨팅 디바이스에 제공하는 하나 이상의 프로세서를 포함하는, 상이한 모달리티간 트랜지션 시스템.

청구항 2

제1항에 있어서,

상기 데이터 처리 시스템은,

스마트 스피커를 포함하는 컴퓨팅 디바이스의 유형에 기초하여 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정하는, 상이한 모달리티간 트랜지션 시스템.

청구항 3

제1항에 있어서,

상기 데이터 처리 시스템은,

디지털 어시스턴트를 포함하는 컴퓨팅 디바이스의 유형에 기초하여 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정하는, 상이한 모달리티간 트랜지션 시스템.

청구항 4

제1항에 있어서,

상기 데이터 처리 시스템은,

요청에 응답하여, 실시간 컨텐츠 선택 프로세스에 입력된 컨텐츠 선택 기준에 기초하여 디지털 컴포넌트 객체를 선택하고, 상기 디지털 컴포넌트 객체는 복수의 제3자 컨텐츠 제공자에 의해 제공된 복수의 디지털 컴포넌트 객

체로부터 선택되는, 상이한 모달리티간 트랜지션 시스템.

청구항 5

제1항에 있어서,

상기 데이터 처리 시스템은,

요청 이전에 컴퓨팅 디바이스에 의해 렌더링된 컨텐츠와 관련된 키워드들에 기초하여 디지털 컴포넌트 객체를 선택하고, 상기 디지털 컴포넌트 객체는 복수의 제3자 컨텐츠 제공자에 의해 제공된 복수의 디지털 컴포넌트 객체로부터 선택되는, 상이한 모달리티간 트랜지션 시스템.

청구항 6

제1항에 있어서,

상기 데이터 처리 시스템은,

자연어 생성 모델을 통해, 디지털 컴포넌트 객체의 메타 데이터에 기초하여 디지털 컴포넌트 객체에 대한 텍스트를 생성하는, 상이한 모달리티간 트랜지션 시스템.

청구항 7

제1항에 있어서,

상기 데이터 처리 시스템은,

음성 모델을 통해, 디지털 컴포넌트 객체의 컨텍스트에 기초하여 디지털 음성을 선택하고, 상기 음성 모델은 오디오 및 시각적 미디어 컨텐츠를 포함하는 이력 데이터 세트로 기계 학습 기술에 의해 트레이닝되는, 상이한 모달리티간 트랜지션 시스템.

청구항 8

제1항에 있어서,

상기 데이터 처리 시스템은,

음성 특성 벡터를 생성하기 위해 음성 모델에 디지털 컴포넌트 객체의 컨텍스트를 입력하고, 상기 음성 모델은 오디오 및 시각적 미디어 컨텐츠를 포함하는 이력 데이터 세트로 기계 학습 엔진에 의해 트레이닝되는, 상이한 모달리티간 트랜지션 시스템.

청구항 9

제1항에 있어서,

상기 데이터 처리 시스템은,

메타 데이터에 기초하여, 오디오 트랙에 트리거 단어를 추가하기로 결정하고, 제2 입력 오디오 신호에서 트리거 단어의 검출은 데이터 처리 시스템 또는 컴퓨팅 디바이스로 하여금 트리거 단어에 대응하는 디지털 액션을 수행하게 하는, 상이한 모달리티간 트랜지션 시스템.

청구항 10

제1항에 있어서,

상기 데이터 처리 시스템은,

디지털 컴포넌트 객체의 카테고리를 결정하고;

데이터베이스로부터, 카테고리와 관련된 복수의 디지털 액션에 대응하는 복수의 트리거 단어를 검색하고;

트리거 키워드들의 이력 수행에 기초하여 트레이닝된 디지털 액션 모델을 사용하여, 디지털 컴포넌트 객체의 컨텍스트 및 컴퓨팅 디바이스의 유형에 기초하여 복수의 트리거 단어를 순위 지정하고; 그리고

오디오 트랙에 추가할 가장 높은 순위의 트리거 키워드를 선택하는, 상이한 모달리티간 트랜지션 시스템..

청구항 11

제1항에 있어서,

상기 데이터 처리 시스템은,

디지털 컴포넌트 객체 내의 시각적 객체를 식별하도록 디지털 컴포넌트 객체에 대해 이미지 인식을 수행하고; 그리고

데이터베이스에 저장된 복수의 비-음성 오디오 큐로부터, 시각적 객체에 대응하는 비-음성 오디오 큐를 선택하는, 상이한 모달리티간 트랜지션 시스템.

청구항 12

제1항에 있어서,

상기 데이터 처리 시스템은,

이미지 인식 기술을 통해 디지털 컴포넌트 객체 내의 복수의 시각적 객체를 식별하고;

메타 데이터 및 복수의 시각적 객체에 기초하여, 복수의 비-음성 오디오 큐를 선택하고;

각각의 시각적 객체와 메타 데이터 간의 매칭 레벨을 나타내는 시각적 객체각각에 대해 매칭 스코어를 결정하고;

매칭 스코어에 기초하여 복수의 비-음성 오디오 큐를 순위 지정하고;

텍스트를 렌더링하기 위해 복수의 비-음성 오디오 큐 각각과 상기 컨텍스트에 기초하여 선택된 디지털 음성 간의 오디오 간섭 레벨을 결정하고; 그리고

최고 순위에 기초하여, 임계값 미만의 오디오 간섭 레벨과 관련된 복수의 비-음성 오디오 큐로부터 비-음성 오디오 큐를 선택하는, 상이한 모달리티간 트랜지션 시스템.

청구항 13

제1항에 있어서,

이력 수행 데이터를 사용하여 트레이닝된 삽입 모델에 기초하여, 컴퓨팅 디바이스에 의해 출력된 디지털 미디어 스트림에서 오디오 트랙에 대한 삽입 지점을 식별하고; 그리고

컴퓨팅 디바이스가 디지털 미디어 스트림의 삽입 지점에서 오디오 트랙을 렌더링하게 하는 명령을 컴퓨팅 디바이스로 제공하는, 상이한 모달리티간 트랜지션 시스템.

청구항 14

상이한 모달리티간 트랜지션 방법으로서,

네트워크를 통해 데이터 처리 시스템의 하나 이상의 프로세서에 의해, 데이터 처리 시스템으로부터 멀리 떨어진 컴퓨팅 디바이스의 마이크로폰에 의해 검출된 입력 오디오 신호를 포함하는 데이터 패킷을 수신하는 단계;

데이터 처리 시스템에 의해, 요청을 식별하기 위해 입력 오디오 신호를 파싱하는 단계;

데이터 처리 시스템에 의해 요청에 기초하여, 시각적 출력 포맷을 갖는 디지털 컴포넌트 객체를 선택하는 단계 - 상기 디지털 컴포넌트 객체는 메타 데이터와 연관됨 -;

데이터 처리 시스템에 의해 컴퓨팅 디바이스의 유형에 기초하여, 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정하는 단계;

데이터 처리 시스템에 의해 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 한 결정에 응답하여, 디지털 컴포넌트 객체에 대한 텍스트를 생성하는 단계;

데이터 처리 시스템에 의해, 디지털 컴포넌트 객체의 컨텍스트를 처리하는 음성 모델에 의해 생성된 음성 특성

벡터에 기초하여, 복수의 디지털 음성으로부터 텍스트를 렌더링하기 위한 디지털 음성을 선택하는 단계;

데이터 처리 시스템에 의해, 디지털 음성에 의해 렌더링된 텍스트로 디지털 컴포넌트 객체의 기본 오디오 트랙을 구성하는 단계;

데이터 처리 시스템에 의해 디지털 컴포넌트 객체에 기초하여, 비-음성 오디오 큐를 생성하는 단계;

데이터 처리 시스템에 의해, 디지털 컴포넌트 객체의 오디오 트랙을 생성하기 위해 비-음성 오디오 큐를 디지털 컴포넌트 객체의 기본 오디오 형식과 결합하는 단계; 및

데이터 처리 시스템에 의해 컴퓨팅 디바이스의 요청에 응답하여, 컴퓨팅 디바이스의 스피커를 통한 출력을 위해 디지털 컴포넌트 객체의 오디오 트랙을 컴퓨팅 디바이스에 제공하는 단계를 포함하는, 상이한 모달리티간 트랜지션 방법.

청구항 15

제14항에 있어서,

데이터 처리 시스템에 의해, 스마트 스피커를 포함하는 컴퓨팅 디바이스의 유형에 기초하여 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정하는 단계를 포함하는, 상이한 모달리티간 트랜지션 방법.

청구항 16

제14항에 있어서,

데이터 처리 시스템에 의해 요청에 응답하여, 실시간 컨텐츠 선택 프로세스에 입력된 컨텐츠 선택 기준에 기초하여 디지털 컴포넌트 객체를 선택하는 단계 -상기 디지털 컴포넌트 객체는 복수의 제3자 컨텐츠 제공자에 의해 제공된 복수의 디지털 컴포넌트 객체로부터 선택됨 - 를 포함하는, 상이한 모달리티간 트랜지션 방법.

청구항 17

제14항에 있어서,

데이터 처리 시스템에 의해, 요청 이전에 컴퓨팅 디바이스에 의해 렌더링된 컨텐츠와 관련된 키워드들에 기초하여 디지털 컴포넌트 객체를 선택하는 단계 - 상기 디지털 컴포넌트 객체는 복수의 제3자 컨텐츠 제공자에 의해 제공된 복수의 디지털 컴포넌트 객체로부터 선택됨 - 를 포함하는, 상이한 모달리티간 트랜지션 방법.

청구항 18

상이한 모달리티간 트랜지션 시스템에 있어서,

데이터 처리 시스템을 포함하고, 상기 데이터 처리 시스템은:

컴퓨팅 디바이스에 의해 렌더링된 디지털 스트리밍 컨텐츠와 관련된 키워드들을 식별하고;

키워드들에 기초하여, 시각적 출력 포맷을 갖는 디지털 컴포넌트 객체를 선택하고 - 상기 디지털 컴포넌트 객체는 메타 데이터와 연관됨 -;

컴퓨팅 디바이스의 유형에 기초하여, 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정하고;

디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 한 결정에 응답하여, 디지털 컴포넌트 객체에 대한 텍스트를 생성하고;

디지털 컴포넌트 객체의 컨텍스트를 처리하는 음성 모델에 의해 생성된 음성 특성 벡터에 기초하여, 복수의 디지털 음성으로부터 텍스트를 렌더링하기 위한 디지털 음성을 선택하고 - 상기 음성 모델은 오디오 및 시각적 미디어 컨텐츠를 포함하는 이력 데이터 세트로 기계 학습 엔진에 의해 트레이닝됨 -;

디지털 음성에 의해 렌더링된 텍스트로 디지털 컴포넌트 객체의 기본 오디오 트랙을 구성하고;

디지털 컴포넌트 객체의 메타 데이터에 기초하여, 비-음성 오디오 큐를 생성하고;

디지털 컴포넌트 객체의 오디오 트랙을 생성하도록 비-음성 오디오 큐를 디지털 컴포넌트 객체의 기본 오디오 형식과 결합하고; 그리고

컴퓨팅 디바이스의 스피커를 통해 출력하기 위해 디지털 컴포넌트 객체의 오디오 트랙을 컴퓨팅 디바이스에 제공하는 하나 이상의 프로세서를 포함하는, 상이한 모달리티간 트랜지션 시스템.

청구항 19

제18항에 있어서,

상기 데이터 처리 시스템은,

스마트 스피커를 포함하는 컴퓨팅 디바이스의 유형에 기초하여 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정하는, 상이한 모달리티간 트랜지션 시스템.

청구항 20

제19항에 있어서,

상기 데이터 처리 시스템은,

실시간 컨텐츠 선택 프로세스에 입력된 키워드들에 기초하여 디지털 컴포넌트 객체를 선택하고, 상기 디지털 컴포넌트 객체는 복수의 제3자 컨텐츠 제공자에 의해 제공된 복수의 디지털 컴포넌트 객체로부터 선택되는, 상이한 모달리티간 트랜지션 시스템.

청구항 21

데이터 처리 시스템으로서,

네트워크를 통해 컴퓨팅 디바이스로부터, 요청을 나타내는 데이터 패킷을 수신하고;

요청에 기초하여 시각적 출력 포멧을 갖는 디지털 컴포넌트 객체를 선택하고 - 상기 디지털 컴포넌트 객체는 메타데이터와 연관됨 -;

디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정하고;

디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 한 결정에 응답하여, 디지털 컴포넌트 객체에 대한 텍스트를 획득하고;

디지털 컴포넌트 객체의 컨텍스트를 처리하는 음성 모델에 의해 생성된 음성 특성 벡터에 기초하여, 복수의 디지털 음성으로부터 텍스트를 렌더링하기 위한 디지털 음성을 선택하고;

디지털 음성에 의해 렌더링된 텍스트로 디지털 컴포넌트 객체의 기본(baseline) 오디오 트랙을 구성하고;

디지털 컴포넌트 객체에 기초하여, 비-음성 오디오 큐를 생성하고;

디지털 컴포넌트 객체의 오디오 트랙을 생성하기 위해 비-음성 오디오 큐를 디지털 컴포넌트 객체의 기본 오디오 형식과 결합하고; 그리고

컴퓨팅 디바이스로부터의 요청에 응답하여, 컴퓨팅 디바이스의 스피커를 통한 출력을 위해 디지털 컴포넌트 객체의 오디오 트랙을 컴퓨팅 디바이스에 제공하는 하나 이상의 프로세서를 포함하는, 데이터 처리 시스템.

청구항 22

제21항에 있어서,

상기 하나 이상의 프로세서는,

디지털 컴포넌트 객체에 대해, 액션 또는 상호 작용의 유형을 결정하는, 데이터 처리 시스템..

청구항 23

제22항에 있어서,

상기 하나 이상의 프로세서는,

결정된 유형의 액션 또는 상호 작용에 대해 디지털 컴포넌트 객체를 구성하는, 데이터 처리 시스템.

청구항 24

제23항에 있어서,

상기 하나 이상의 프로세서는,

오디오 트랙과의 상호 작용을 용이하게 하기 위해 실행 가능한 명령을 오디오 트랙에 추가하는, 데이터 처리 시스템.

청구항 25

제24항에 있어서,

상기 하나 이상의 프로세서는,

오디오 트랙에 트리거 워드를 추가하고, 상기 트리거 워드는 사전 결정된 시간 간격 동안 활성 상태로 유지되는, 데이터 처리 시스템.

청구항 26

제25항에 있어서,

상기 사전 결정된 시간 간격은 오디오 트랙의 재생을 포함하는, 데이터 처리 시스템.

청구항 27

제25항 또는 제26항에 있어서,

상기 사전 결정된 시간 간격은,

오디오 트랙 이후의 사전 결정된 시간량을 포함하는, 데이터 처리 시스템.

청구항 28

제21항에 있어서,

상기 하나 이상의 프로세서는,

자연어 생성 모델을 통해, 디지털 컴포넌트 객체의 메타데이터에 기초하여 디지털 컴포넌트 객체에 대한 텍스트를 생성하는, 데이터 처리 시스템.

청구항 29

제21항에 있어서,

상기 하나 이상의 프로세서는,

음성 특성 벡터를 생성하기 위해 음성 모델에 디지털 컴포넌트 객체의 컨텍스트를 입력하고, 상기 음성 모델은 오디오 및 시각적 미디어 컨텐츠를 포함하는 이력 데이터 세트를 사용하여 기계 학습 엔진에 의해 트레이닝되는, 데이터 처리 시스템.

청구항 30

방법으로서,

데이터 처리 시스템에 의해 네트워크를 통해 컴퓨팅 디바이스로부터, 요청을 나타내는 데이터 패킷을 수신하는 단계;

데이터 처리 시스템에 의해 요청에 기초하여, 시각적 출력 포멧을 갖는 디지털 컴포넌트 객체를 선택하는 단계 - 상기 디지털 컴포넌트 객체는 메타데이터와 연관됨 -;

데이터 처리 시스템에 의해, 디지털 컴포넌트 객체를 오디오 출력 포멧으로 변환하기로 결정하는 단계;

데이터 처리 시스템에 의해, 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 한 결정에 응답하여 디지털 컴포넌트 객체에 대한 텍스트를 획득하는 단계;

데이터 처리 시스템에 의해, 디지털 컴포넌트 객체의 컨텍스트를 처리하는 음성 모델에 의해 생성된 음성 특성 벡터에 기초하여, 복수의 디지털 음성으로부터 텍스트를 렌더링하기 위한 디지털 음성을 선택하는 단계;

데이터 처리 시스템에 의해, 디지털 음성에 의해 렌더링된 텍스트로 디지털 컴포넌트 객체의 기본 오디오 트랙을 구성하는 단계;

데이터 처리 시스템에 의해, 디지털 컴포넌트 객체에 기초하여, 비-음성 오디오 큐를 생성하는 단계;

데이터 처리 시스템에 의해, 디지털 컴포넌트 객체의 오디오 트랙을 생성하도록 비-음성 오디오 큐를 디지털 컴포넌트 객체의 기본 오디오 형식과 결합하는 단계; 및

데이터 처리 시스템에 의해 컴퓨팅 디바이스로부터의 요청에 응답하여, 컴퓨팅 디바이스의 스피커를 통한 출력을 위해 디지털 컴포넌트 객체의 오디오 트랙을 컴퓨팅 디바이스에 제공하는 단계를 포함하는, 방법.

청구항 31

제30항에 있어서,

데이터 처리 시스템에 의해, 디지털 컴포넌트 객체에 대해 액션 또는 상호 작용의 유형을 결정하는 단계를 포함하는, 방법.

청구항 32

제31항에 있어서,

데이터 처리 시스템에 의해, 결정된 유형의 액션 또는 상호 작용에 대해 디지털 컴포넌트 객체를 구성하는 단계를 포함하는, 방법.

청구항 33

제32항에 있어서,

데이터 처리 시스템에 의해, 오디오 트랙과의 상호 작용을 용이하게 하기 위해 실행 가능한 명령을 오디오 트랙에 추가하는 단계를 포함하는, 방법.

청구항 34

제33항에 있어서,

데이터 처리 시스템에 의해, 오디오 트랙에 트리거 워드를 추가하는 단계 - 상기 트리거 워드는 사전 결정된 시간 간격 동안 활성 상태로 유지됨 - 를 포함하는, 방법.

청구항 35

제34항에 있어서,

상기 사전 결정된 시간 간격은 오디오 트랙의 재생을 포함하는, 방법.

청구항 36

제34항 또는 제35항에 있어서,

상기 사전 결정된 시간 간격은 오디오 트랙 이후의 사전 결정된 시간량을 포함하는, 방법.

청구항 37

제30항에 있어서,

데이터 처리 시스템에 의해 자연어 생성 모델을 통해, 디지털 컴포넌트 객체의 메타데이터에 기초하여 디지털 컴포넌트 객체에 대한 텍스트를 생성하는 단계를 포함하는, 방법.

청구항 38

제30항에 있어서,

데이터 처리 시스템에 의해, 음성 특성 벡터를 생성하기 위해 음성 모델에 디지털 컴포넌트 객체의 컨텍스트를 입력하는 단계 - 상기 음성 모델은 오디오 및 시각적 미디어 컨텐츠를 포함하는 이력 데이터 세트를 사용하여 기계 학습 엔진에 의해 트레이닝됨 - 를 포함하는, 방법.

청구항 39

명령들을 저장하는 하나 이상의 컴퓨터 판독 가능 저장 매체로서,

상기 명령들은 데이터 처리 시스템으로 하여금:

네트워크를 통해 컴퓨팅 디바이스로부터, 요청을 나타내는 데이터 패킷을 수신하고;

요청에 기초하여, 시각적 출력 포멧을 갖는 디지털 컴포넌트 객체를 선택하고 - 상기 디지털 컴포넌트 객체는 메타데이터와 연관됨 -;

디지털 컴포넌트 객체를 오디오 출력 포멧으로 변환하기로 결정하고;

디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 한 결정에 응답하여, 디지털 컴포넌트 객체에 대한 텍스트를 획득하고;

디지털 컴포넌트 객체의 컨텍스트를 처리하는 음성 모델에 의해 생성된 음성 특성 벡터에 기초하여, 복수의 디지털 음성으로부터 텍스트를 렌더링하기 위한 디지털 음성을 선택하고 - 상기 음성 모델은 오디오 및 시각적 미디어 컨텐츠를 포함하는 이력 데이터 세트로 기계 학습 엔진에 의해 트레이닝됨 -;

디지털 음성에 의해 렌더링된 텍스트로 디지털 컴포넌트 객체의 기본 오디오 트랙을 구성하고;

디지털 컴포넌트 객체에 기초하여, 비-음성 오디오 큐를 생성하고;

디지털 컴포넌트 객체의 오디오 트랙을 생성하기 위해 비-음성 오디오 큐를 디지털 컴포넌트 객체의 기본 오디오 형식과 결합하고; 그리고

컴퓨팅 디바이스로부터의 요청에 응답하여, 컴퓨팅 디바이스의 스피커를 통한 출력을 위해 디지털 컴포넌트 객체의 오디오 트랙을 컴퓨팅 디바이스에 제공하도록 실행 가능한, 컴퓨터 판독 가능 저장 매체.

청구항 40

제39항에 있어서,

상기 명령들은 데이터 처리 시스템으로 하여금:

디지털 컴포넌트 객체에 대해, 액션 또는 상호작용의 유형을 결정하고; 그리고

결정된 유형의 액션 또는 상호 작용에 대해 디지털 컴포넌트 객체를 구성하도록 실행 가능한, 컴퓨터 판독 가능 저장 매체.

발명의 설명

배 경 기 술

[0001] 데이터 처리 시스템은 디지털 컨텐츠를 컴퓨팅 디바이스으로 제공하여 컴퓨팅 디바이스가 디지털 컨텐츠를 제시하게 할 수 있다. 디지털 컨텐츠는 컴퓨팅 디바이스가 디스플레이를 통해 제시할 수 있는 시각적 컨텐츠를 포함할 수 있다. 디지털 컨텐츠는 컴퓨팅 디바이스가 스피커를 통해 출력할 수 있는 오디오 컨텐츠를 포함할 수 있다.

발명의 내용

[0002] 본 기술 솔루션의 적어도 하나의 양태는 오디오 트랙을 생성하는 시스템에 관한 것이다. 시스템에는 데이터 처리 시스템이 포함될 수 있다. 데이터 처리 시스템은 하나 이상의 프로세서를 포함할 수 있다. 데이터 처리 시스템은 네트워크를 통해, 데이터 처리 시스템으로부터 떨어진 컴퓨팅 디바이스의 마이크로폰에 의해 검출된 입력 오디오 신호를 포함하는 데이터 패킷을 수신할 수 있다. 데이터 처리 시스템은 입력 오디오 신호를 파싱(구문 분석)하여 요청을 식별할 수 있다. 데이터 처리 시스템은 요청에 기초하여 시각적 출력 포멧을 갖는 디지털 컴포넌트 객체를 선택할 수 있고, 디지털 컴포넌트 객체는 메타 데이터와 연관된다. 데이터 처리 시스템은 컴퓨팅

디바이스의 유형에 기초하여, 디지털 컴포넌트 객체를 오디오 출력 포멧으로 변환하기로(할 것을) 결정할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체를 오디오 출력 포멧으로 변환하기로 한 결정에 응답하여, 디지털 컴포넌트 객체에 대한 텍스트를 생성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체의 컨텍스트에 기초하여, 텍스트를 렌더링하기 위한 디지털 음성을 선택할 수 있다. 데이터 처리 시스템은 디지털 음성에 의해 렌더링된 텍스트로 디지털 컴포넌트 객체의 기준(baseline) 오디오 트랙을 구성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체의 메타 데이터에 기초하여, 비-음성 오디오 큐(cues)를 생성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체의 기본 오디오 형식과 비-음성 오디오 큐를 결합하여 디지털 컴포넌트 객체의 오디오 트랙을 생성할 수 있다. 데이터 처리 시스템은 컴퓨팅 디바이스의 요청에 응답하여, 컴퓨팅 디바이스의 스피커를 통한 출력을 위해 컴퓨팅 디바이스에 디지털 컴포넌트 객체의 오디오 트랙을 제공할 수 있다.

[0003] 본 기술 솔루션의 적어도 하나의 양태는 오디오 트랙을 생성하는 방법에 관한 것이다. 방법은 데이터 처리 시스템의 하나 이상의 프로세서에 의해 수행될 수 있다. 방법은 데이터 처리 시스템이 데이터 처리 시스템으로부터 떨어진 컴퓨팅 디바이스의 마이크로폰에 의해 검출된 입력 오디오 신호를 포함하는 데이터 패킷을 수신하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 요청을 식별하기 위해 입력 오디오 신호를 파싱하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 요청에 기초하여, 시각적 출력 포맷을 갖는 디지털 컴포넌트 객체를 선택하는 단계를 포함할 수 있으며, 디지털 컴포넌트 객체는 메타 데이터와 연관된다. 방법은 데이터 처리 시스템이 컴퓨팅 디바이스의 유형에 기초하여, 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하로 한 결정에 응답하여, 디지털 컴포넌트 객체에 대한 텍스트를 생성하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 디지털 컴포넌트 객체의 컨텍스트에 기초하여, 텍스트를 렌더링하기 위한 디지털 음성을 선택하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 디지털 음성에 의해 렌더링된 텍스트로 디지털 컴포넌트 객체의 기준 오디오 트랙을 구성하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 디지털 컴포넌트 객체에 기초하여, 비-음성 오디오 큐를 생성하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 디지털 컴포넌트 객체의 오디오 트랙을 생성하기 위해 비-음성 오디오 큐를 디지털 컴포넌트 객체의 기본 오디오 형태와 결합하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 컴퓨팅 디바이스의 요청에 응답하여, 컴퓨팅 디바이스의 스피커를 통한 출력을 위해 컴퓨팅 디바이스에 디지털 컴포넌트 객체의 오디오 트랙을 제공하는 단계를 포함할 수 있다.

[0004] 본 기술 솔루션의 적어도 하나의 양태는 오디오 트랙을 생성하는 시스템에 관한 것이다. 시스템은 하나 이상의 프로세서가 있는 데이터 처리 시스템을 포함할 수 있다. 데이터 처리 시스템은 컴퓨팅 디바이스에 의해 렌더링된 디지털 스트리밍 컨텐츠와 관련된 키워드들을 식별할 수 있다. 데이터 처리 시스템은 키워드들에 기초하여, 시각적 출력 포멧을 갖는 디지털 컴포넌트 객체를 선택할 수 있으며, 디지털 컴포넌트 객체는 메타 데이터와 관련된다. 데이터 처리 시스템은 컴퓨팅 디바이스의 유형에 기초하여, 디지털 컴포넌트 객체를 오디오 출력 포멧으로 변환하기로 결정할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체를 오디오 출력 포멧으로 변환하기로 한 결정에 응답하여, 디지털 컴포넌트 객체에 대한 텍스트를 생성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체의 컨텍스트에 기초하여, 텍스트를 렌더링하기 위한 디지털 음성을 선택할 수 있다. 데이터 처리 시스템은 디지털 음성에 의해 렌더링된 텍스트로 디지털 컴포넌트 객체의 기준 오디오 트랙을 구성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체에 기초하여, 비-음성 오디오 큐를 생성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체의 기본 오디오 형식과 비-음성 오디오 큐를 결합하여 디지털 컴포넌트 객체의 오디오 트랙을 생성할 수 있다. 데이터 처리 시스템은 컴퓨팅 디바이스의 스피커를 통한 출력을 위해 컴퓨팅 디바이스에 디지털 컴포넌트 객체의 오디오 트랙을 제공할 수 있다.

[0005] 이들 및 다른 양태 및 구현은 아래에서 상세히 논의된다. 전술한 정보 및 다음의 상세한 설명은 다양한 양태 및 구현의 예시적인 예를 포함하고, 청구된 양태 및 구현의 특성 및 특징을 이해하기 위한 개요 또는 프레임 워크를 제공한다. 도면은 다양한 양태 및 구현에 대한 예시 및 추가 이해를 제공하며, 본 명세서에 통합되고 그 일부를 구성한다.

도면의 간단한 설명

[0006] 첨부된 도면들은 일정한 비율로 그려진 것이 아니다. 다양한 도면의 참조 번호와 명칭은 같은 요소를 나타낸다. 명확성을 위해 모든 컴포넌트가 모든 도면에서 레이블이 지정되는 것은 아니다.

도 1은 구현에 따른 오디오 트랙을 생성하기 위한 시스템의 예시이다.

도 2는 구현에 따른 오디오 트랙을 생성하는 방법의 예시이다.

도 3은 도 1에 도시된 시스템 및 도 2에 도시된 방법을 구현하기 위해 사용될 수 있는 컴퓨터 시스템에 대한 일반적인 아키텍처를 도시하는 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0007] 다음은 오디오 트랙을 생성하는 방법, 장치 및 시스템과 관련된 다양한 개념 및 구현에 대한 보다 상세한 설명이다. 예를 들어, 방법, 장치 및 시스템은 시각적 컨텐츠로부터 오디오 트랙을 생성할 수 있다. 위에서 소개되고 아래에서 더 상세히 논의되는 다양한 개념은 임의의 다양한 방식으로 구현될 수 있다.

[0008] 본 기술 솔루션은 일반적으로 오디오 트랙을 생성하는 것에 관한 것이다. 기술 솔루션의 시스템 및 방법은 시각적 컨텐츠를 처리하여 음성 및 비-음성 큐가 있는 오디오 트랙을 생성할 수 있다. 예를 들어, 특정 유형의 컴퓨팅 디바이스는 오디오 전용 인터페이스를 제공할 수 있다(예를 들어, 사용자로부터 음성 입력 수신, 입력 처리 및 디지털 음성을 통해 오디오 또는 음성 출력을 제공). 특정 컴퓨팅 디바이스는 주로 오디오 사용자 인터페이스를 사용하거나 특정 상황하에서 주로 오디오 인터페이스를 사용할 수 있다. 예를 들어, 모바일 컴퓨팅 디바이스의 사용자는 차량을 운전하거나 달리거나 스트리밍 음악 서비스를 듣는 동안 주로 오디오 전용 인터페이스를 사용할 수 있다. 주(primary) 인터페이스가 오디오 기반인 경우, 데이터 처리 시스템은 오디오 디지털 컴포넌트 객체(예를 들어, 오디오 컨텐츠 아이템)을 제공할 수 있다. 예를 들어, 데이터 처리 시스템은 제3자 오디오 컨텐츠 제공자에 의해 설정되거나 제공된 오디오 컨텐츠 아이템을 선택할 수 있다. 데이터 처리 시스템은 사용자의 컨텐츠 요청에 응답하거나 다른 트리거 이벤트에 기초하여 오디오 컨텐츠 아이템을 제공할 수 있다. 그러나, 제3자 컨텐츠 제공자에 의해 설정된 컨텐츠 아이템은 오디오 컨텐츠 아이템이 아닐 수 있다. 데이터 처리 시스템은 키워드, 관련성 또는 기타 요인과 같은 매칭 기준에 기초하여 그러한 컨텐츠 아이템을 선택하기로 결정할 수 있다. 그러나, 컴퓨팅 디바이스는 오디오 기반 인터페이스만을 가지고 있기 때문에 데이터 처리 시스템은 컴퓨팅 디바이스에 컨텐츠 아이템을 제공하지 못할 수 있다. 또는, 일부 경우에서, 컴퓨팅 디바이스가 주로 오디오 인터페이스를 사용하거나 오디오 인터페이스가 가장 효율적인 인터페이스 인 경우, 데이터 처리 시스템은 시각적 컨텐츠 아이템을 제공하고 컴퓨팅 디바이스가 컴퓨팅 디바이스의 디스플레이를 사용하여 시각적 컨텐츠 아이템을 렌더링하게 함으로써 비효율적이거나 낭비되는 컴퓨팅 이용 또는 부정적인 사용자 경험을 유발할 수 있다. 디스플레이를 사용하면 모바일 컴퓨팅 디바이스(예를 들어, 스마트 폰, 스마트 워치 또는 기타 웨어러블 디바이스)의 배터리 전력이 낭비될 수 있다. 따라서, 오디오 컨텐츠가 선호되는 시각적 컨텐츠를 제공하거나 시각적 포멧으로만 제공되기 때문에 가장 관련성이 높은 컨텐츠 아이템을 제공할 수 없는 데이터 처리 시스템은 모바일 컴퓨팅 디바이스에 의한 컴퓨팅 리소스 사용을 낭비하거나 사용자 경험을 저하시킬 수 있다.

[0009] 또한, 예를 들어, 컨텐츠 아이템을 생성할 포멧 결정, 텍스트를 포함하거나 포함하지 않을 수 있는 시각적 컨텐츠 아이템에 대해 음성 텍스트를 정확하게 생성하는 방법, 생성된 음성 텍스트에 대한 적절한 음성 선택, 비-음성 오디오 큐 추가 등의 다양한 기술적 문제로 인해 상이한 포멧으로 컨텐츠 아이템을 생성하는 것은 기술적으로 어려울 수 있다. 본 기술 솔루션의 시스템 및 방법은 기계 학습 기술 및 이력(과거) 데이터를 사용하여 트레이닝된 자연어 처리 및 모델을 사용하여, 포멧(예를 들어, 오디오 전용, 시청각 포멧 및 컴퓨팅 디바이스 유형 및 컴퓨팅 디바이스의 현재 컨텍스트에 기초한 상호 작용의 모드)을 선택하고, 시각적 컨텐츠 아이템 및 관련 메타 데이터에 기초하여 텍스트를 자동으로 생성하고, 생성된 음성 텍스트에 대한 적절한 디지털 음성 인쇄를 선택하고, 음성 텍스트와 함께 비-음성 오디오 큐를 선택 및 제공할 수 있다.

[0010] 예를 들어, 컴퓨팅 디바이스는 시각적 사용자 인터페이스(예를 들어, 사용자 입력을 위한 터치 스크린을 갖는 디스플레이 화면) 및 오디오 기반 사용자 인터페이스(예를 들어, 마이크로폰 및 스피커) 모두로 구성될 수 있다. 컴퓨팅 디바이스는 현재 컴퓨팅 디바이스와 관련된 스피커를 통해 출력하기 위해 음악을 스트리밍할 수 있다. 데이터 처리 시스템은 요청, 질의 또는 스트리밍 음악과 관련된 정보를 사용하여 제3자 컨텐츠 아이템을 선택할 수 있다. 선택된 타사 컨텐츠 아이템은 시각적 컨텐츠 아이템(예를 들어, 텍스트를 포함할 수 있는 이미지)일 수 있다. 데이터 처리 시스템은 요청, 질의 또는 스트리밍 음악과 관련된 키워드들에 기초하여 이 시각적 컨텐츠 아이템을 선택할 수 있다. 예를 들어, 선택된 시각적 컨텐츠 아이템은 적어도 관련성 스코어에 기초하여 컨텐츠 아이템들을 순위 지정할 수 있는 실시간 컨텐츠 선택 프로세스에 기초하여 결정된 최고 순위 컨텐츠 아이템일 수 있다. 데이터 처리 시스템은 컴퓨팅 디바이스가 시각적 사용자 인터페이스(예를 들어, 디스플레이 스크린 및 터치 스크린 입력) 및 오디오 사용자 인터페이스(예를 들어, 출력용 스피커 및 입력용 마이크로폰)로 구성되어 있다고 결정할 수 있다. 그러나, 데이터 처리 시스템은 컴퓨팅 디바이스의 현재 기능에 기초하여 현재 주로 사용되고있는 인터페이스가 오디오 기반 인터페이스임을 추가로 결정할 수 있다. 따라서 데이터 처리 시스

템은 컴퓨팅 디바이스가 시각 및 오디오 인터페이스로 구성되어 있지만 현재 사용되는 주 인터페이스가 오디오 인터페이스이고, 컨텐츠 아이템에 의한 렌더링을 제공하기 위해 시각적 컨텐츠 아이템에 기초하여 오디오 컨텐츠 아이템을 생성하는 것은 컴퓨팅 디바이스 또는 낭비되는 컴퓨팅 리소스에 의한 배터리 소모를 줄이고(예를 들어, 컴퓨팅 디바이스의 디스플레이를 깨우는 대신 스트리밍 음악과 함께 오디오 컨텐츠 아이템을 제공) 컴퓨팅 디바이스가 제공하는 사용자 경험을 개선할 것이라고 결정할 수 있다(예를 들어, 방해가 되지 않는 방식으로 오디오 컨텐츠 아이템 제공). 따라서, 본 기술 솔루션은 사용자 인터페이스 기능과 사용자 경험을 향상시키면서 배터리 또는 컴퓨팅 리소스 사용률을 줄이기 위해 상이한 모달리티(modalities, 양식)간에 컨텐츠 아이템을 원활하게 트랜지션(전환)할 수 있다.

[0011] 오디오 컨텐츠 아이템을 생성할 때, 데이터 처리 시스템은 오디오 음악 스트림의 삽입 시간을 결정할 수 있다. 데이터 처리 시스템은 임의의 시각 인디케이터와 함께 오디오 컨텐츠 아이템을 동반할지 여부 및 컨텐츠 아이템에 대해 구성할 상호 작용의 유형을 동적으로 결정할 수 있다.

[0012] 도 1은 오디오 트랙을 생성하는 예시적인 시스템(100)을 도시한다. 시스템(100)은 시각적 컨텐츠로부터 오디오 트랙을 생성할 수 있다. 시스템(100)은 컨텐츠 선택 인프라를 포함할 수 있다. 시스템(100)은 데이터 처리 시스템(102)을 포함할 수 있다. 데이터 처리 시스템(102)은 하나 이상의 프로세서(예를 들어, 도 3에 도시된 프로세서(310))를 포함하거나 하나 이상의 프로세서 상에서 실행될 수 있다. 데이터 처리 시스템(102)은 네트워크(105)를 통해 제3자(3P) 디지털 컨텐츠 제공자 디바이스(160) 또는 컴퓨팅 디바이스(140)(예를 들어, 클라이언트 디바이스) 중 하나 이상과 통신할 수 있다. 네트워크(105)는 인터넷, 로컬, 와이드, 메트로 또는 다른 영역 네트워크, 인트라넷, 위성 네트워크, 및 음성 또는 데이터 이동 전화 네트워크와 같은 다른 통신 네트워크를 포함할 수 있다. 네트워크(105)는 랩탑, 데스크탑, 태블릿, 개인용 디지털 어시스턴트, 스마트 폰, 휴대용 컴퓨터 또는 스피커와 같은 적어도 하나의 컴퓨팅 디바이스(140)상에 제시, 출력, 렌더링 또는 디스플레이될 수 있는 웹 페이지, 웹 사이트, 도메인 이름 또는 URL과 같은 정보 리소스에 액세스하는데 사용될 수 있다. 예를 들어, 네트워크(105)를 통해 컴퓨팅 디바이스(140)의 사용자는 3P 디지털 컨텐츠 제공자 디바이스(160)에 의해 제공된 정보 또는 데이터에 액세스할 수 있다. 컴퓨팅 디바이스(140)는 디스플레이 디바이스(146) 및 스피커(예를 들어, 오디오 드라이버(150)에 의해 구동되는 변환기)를 포함할 수 있다. 컴퓨팅 디바이스(140)는 디스플레이를 포함하거나 포함하지 않을 수 있는데, 예를 들어, 컴퓨팅 디바이스는 마이크로폰 및 스피커(예를 들어, 스마트 스피커)와 같은 제한된 유형의 사용자 인터페이스를 포함할 수 있다. 일부 경우, 컴퓨팅 디바이스(140)의 주 사용자 인터페이스는 마이크로폰과 스피커일 수 있다. 컴퓨팅 디바이스(140)는 음성 기반 컴퓨팅 환경과 인터페이스하거나 그에 포함될 수 있다.

[0013] 네트워크(105)는 클라이언트 컴퓨팅 디바이스(140)에 의해 제시, 출력, 렌더링 또는 디스플레이될 수 있는 애플리케이션, 웹 페이지, 웹 사이트, 도메인 이름 또는 URL과 같은 정보 리소스에 액세스하기 위해 데이터 처리 시스템(102)에 의해 사용될 수 있다. 예를 들어, 네트워크(105)를 통해 클라이언트 컴퓨팅 디바이스(140)의 사용자는 3P 디지털 컨텐츠 제공자 디바이스(160)에 의해 제공된 정보 또는 데이터에 액세스할 수 있다. 네트워크(105)는 컨텐츠 배치 또는 검색 엔진 결과 시스템과 연관되거나 디지털 컴포넌트 배치 캠페인의 일부로서 제3자 디지털 컴포넌트를 포함할 수 있는 인터넷상에서 이용 가능한 정보 리소스의 서브 네트워크를 포함하거나 구성할 수 있다.

[0014] 네트워크(105)는 임의의 유형 또는 형태의 네트워크일 수 있으며, 포인트-투-포인트 네트워크, 브로드 캐스트 네트워크, 광역 네트워크, 근거리 통신망, 통신 네트워크, 데이터 통신 네트워크, 컴퓨터 네트워크, ATM(Asynchronous Transfer Mode) 네트워크, SONET(Synchronous Optical Network)) 네트워크, SDH(Synchronous Digital Hierarchy) 네트워크, 무선 네트워크 및 유선 네트워크 중 임의의 것을 포함할 수 있다. 네트워크(105)는 적외선 채널 또는 위성 대역과 같은 무선 링크를 포함할 수 있다. 네트워크(105)의 토폴로지는 버스, 스타 또는 링 네트워크 토폴로지를 포함할 수 있다. 네트워크에는 진화된 이동 전화 프로토콜("AMPS"), 시분할 다중 액세스("TDMA"), 코드 분할 다중 액세스("CDMA"), 이동 통신용 글로벌 시스템("GSM"), 일반 패킷 무선 서비스("GPRS") 또는 범용 이동 통신 시스템("UMTS")을 포함하여 모바일 디바이스간에 통신하는 데 사용되는 임의의 프로토콜 또는 프로토콜들을 사용하는 모바일 전화 네트워크가 포함될 수 있다. 상이한 유형의 데이터가 상이한 프로토콜을 통해 전송되거나 동일한 유형의 데이터가 상이한 프로토콜을 통해 전송될 수 있다.

[0015] 데이터 처리 시스템(102)은 네트워크(105)를 통해 통신하는 프로세서를 갖는 컴퓨팅 디바이스와 같은 적어도 하나의 논리 디바이스를 포함할 수 있다. 데이터 처리 시스템(102)은 적어도 하나의 계산 리소스, 서버, 프로세서 또는 메모리를 포함할 수 있다. 예를 들어, 데이터 처리 시스템(102)은 적어도 하나의 데이터 센터에 위치한 복

수의 계산 리소스 또는 서버를 포함할 수 있다. 데이터 처리 시스템(102)은 논리적으로 그룹화된 다수의 서버를 포함하고 분산 컴퓨팅 기술을 용이하게 할 수 있다. 논리적 서버 그룹은 데이터 센터, 서버 팜 또는 머신 팜으로 지지ㅊ될 수 있다. 서버는 지리적으로 분산될 수도 있다. 데이터 센터 또는 머신 팜을 단일 엔티티로서 관리될 수 있고 머신 팜은 복수의 머신 팜을 포함할 수 있다. 각 머신 팜 내의 서버들은 이기 종일 수 있는데 하나 이상의 서버 또는 머신은 운영 체제 플랫폼의 하나 이상의 유형에 따라 작동할 수 있다.

[0016] 머신 팜에 있는 서버들은 관련 스토리지 시스템과 함께 고밀도 랙 시스템에 저장될 수 있으며 기업 데이터 센터에 위치할 수 있다. 예를 들어, 이러한 방식으로 서버들을 통합하는 것은 로컬화된 고성능 네트워크에 서버들과 고성능 스토리지 시스템을 배치함으로써 시스템 관리 효율성, 데이터 보안, 시스템의 물리적 보안 및 시스템 성능을 향상시킬 수 있다. 서버 및 저장 시스템을 포함하는 데이터 처리 시스템(102) 컴포넌트의 전부 또는 일부를 중앙 집중화하고 이를 진화된 시스템 관리 툴과 결합하는 것은 서버 자원을 보다 효율적으로 사용할 수 있도록 하며, 이는 전력 및 처리 요구 사항을 절약하고 대역폭 사용을 감소시킨다.

[0017] 데이터 처리 시스템(102)은 네트워크(105)를 통해 또는 데이터 처리 시스템(102)의 다양한 컴포넌트 사이에서 데이터 패킷 또는 정보를 수신하고 전송할 수 있는 적어도 인터페이스(104)를 포함할 수 있다. 데이터 처리 시스템(102)은 음성 또는 오디오 입력을 수신하고 그 입력 오디오 신호를 처리하거나 파싱할 수 있는 적어도 하나의 자연어 프로세서 컴포넌트(106)를 포함할 수 있다. 데이터 처리 시스템(102)은 하나 이상의 3P 디지털 컨텐츠 제공자 디바이스(160)에 의해 제공되는 디지털 컴포넌트 아이템(예를 들어, 컨텐츠 아이템)을 선택하도록 설계, 구성 및 동작하는 적어도 하나의 컨텐츠 선택기 컴포넌트(108)를 포함할 수 있다. 데이터 처리 시스템(102)은 제1 모달리티 또는 포맷의 컨텐츠 아이템을 상이한 모달리티 또는 포맷으로 변환할지 여부를 결정하기 위해 적어도 컨텐츠 변환 컴포넌트(108)를 포함할 수 있다. 컨텐츠 아이템을 변환하는 것은 다른 포멧으로 새로운 컨텐츠 아이템을 생성하는 것(예를 들어, 시각적 컨텐츠 아이템으로부터 오디오 트랙을 생성하거나 시각 전용 컨텐츠 아이템으로부터 시청각 컨텐츠 아이템을 생성하는 것)을 지칭하거나 포함할 수 있다. 새로운 컨텐츠 아이템은 오리지널 컨텐츠 아이템의 일부를 포함하거나 포함하지 않을 수 있다. 컨텐츠 변환 컴포넌트(110)는 포맷 선택기(112), 텍스트 생성기(114), 음성 선택기(116), 액션 생성기(136) 또는 오디오 큐 생성기(118)를 포함할 수 있다. 데이터 처리 시스템(102)은 컨텐츠 아이템을 삽입할 시기 또는 위치를 결정할 수 있는 적어도 하나의 컨텐츠 삽입 컴포넌트(120)를 포함할 수 있다. 데이터 처리 시스템(102)은 적어도 하나의 기계 학습 엔진(122)을 포함할 수 있다. 데이터 처리 시스템(102)은 적어도 하나의 데이터 저장소(124)를 포함할 수 있다. 데이터 저장소(124)는 하나 이상의 데이터 구조, 데이터 파일, 데이터베이스 또는 다른 데이터를 포함하거나 저장할 수 있다. 데이터 저장소(124)는 하나 이상의 로컬 또는 분산 데이터베이스를 포함할 수 있고, 데이터베이스 관리 시스템을 포함할 수 있다. 데이터 저장소(124)는 컴퓨터 데이터 저장소 또는 메모리를 포함할 수 있다.

[0018] 데이터 저장소(124)는 음성 모델(126), 액션 모델(128), 삽입 모델(130), 컨텐츠 데이터(132), 또는 오디오 큐(134)를 포함, 저장 또는 유지할 수 있다. 음성 모델(126)은 오디오 또는 시청각 컨텐츠를 포함하는 이력 컨텐츠 아이템 및 이력 컨텐츠 아이템과 관련된 메타 데이터에 기초하여 기계 학습 엔진(122)을 사용하여 트레이닝된 모델을 포함할 수 있다. 음성 모델(126)은 또한 이력 컨텐츠 아이템들과 관련된 성능 정보를 사용하여 트레이닝될 수 있다.

[0019] 액션 모델(128)은 컨텐츠 아이템에 대해 액션 또는 상호 작용의 유형을 결정할 수 있는 기계 학습 엔진(122)을 사용하여 트레이닝된 모델을 포함할 수 있다. 예를 들어, 사용자는 컨텐츠 아이템에 관한 추가 정보를 요청하고, 구매하고, 하이퍼 링크를 선택하고, 컨텐츠 아이템을 일시 중지, 포워딩, 되감기 또는 스킵함으로써 컨텐츠 아이템과 상호 작용할 수 있고, 일부 다른 액션을 수행할 수 있다. 데이터 처리 시스템(102)은 액션 모델(128)을 사용하여 컨텐츠 아이템과의 가능한 상호 작용을 결정하거나 예측할 수 있고, 그런 다음 예측된 상호 작용을 위해 컨텐츠 아이템을 구성할 수 있다. 액션 모델(128)은 또한 사전 결정된 액션들에 매핑되는 컨텐츠 아이템들의 카테고리를 포함할 수 있다.

[0020] 삽입 모델(130)은 기계 학습 엔진(122)을 사용하여 트레이닝되어 예를 들어 디지털 음악 스트림에서의 위치와 같은 생성된 컨텐츠 아이템을 삽입할 위치를 결정할 수 있다. 삽입 모델(130)은 상이한 유형의 컨텐츠 아이템들이 디지털 음악 스트림에 삽입된 위치와 같은 이력 데이터를 사용하여 트레이닝될 수 있다.

[0021] 컨텐츠 데이터(132)는 3P 디지털 컨텐츠 제공자 디바이스들(160)에 의해 제공되는 컨텐츠 아이템 또는 디지털 컴포넌트 객체에 관한 데이터를 포함할 수 있다. 컨텐츠 데이터(132)는 예를 들어 시각적 컨텐츠 아이템 또는 시각적 컨텐츠 아이템의 표시, 컨텐츠 캠페인 파라미터, 키워드, 또는 컨텐츠 선택 또는 컨텐츠 전달을 용이하게 하는 다른 데이터를 포함할 수 있다.

[0022] 오디오 큐(134)는 기준 오디오 트랙에 추가될 수 있는 비-음성 오디오 큐를 지칭할 수 있다. 오디오 큐(134)는 오디오 파일 및 오디오 파일을 기술하는 메타 데이터를 포함할 수 있다. 예시적인 오디오 신호는 대양 파도 소리, 새 지저귐 소리, 스포츠 경기의 응원하는 청중 소리, 바람 부는 소리 또는 자동차 엔진 소리 등이 있다.

[0023] 인터페이스(104), 자연어 프로세서 컴포넌트(106), 컨텐츠 선택기 컴포넌트(108), 컨텐츠 변환 컴포넌트(110), 포맷 선택기 컴포넌트(112), 텍스트 생성기 컴포넌트(114), 음성 선택기 컴포넌트(116), 액션 생성기(136), 오디오 큐 생성기(118), 컨텐츠 삽입 컴포넌트 120, 기계 학습 엔진(122) 또는 데이터 처리 시스템(102)의 다른 컴포넌트는 적어도 하나의 처리 유닛 또는 프로그램 가능한 논리 어레이 엔진과 같은 다른 논리 디바이스, 또는 서로 또는 다른 리소스 또는 데이터베이스와 통신하도록 구성된 모듈을 포함하거나 이용할 수 있다. 인터페이스(104), 자연어 프로세서 컴포넌트(106), 컨텐츠 선택기 컴포넌트(108), 컨텐츠 변환 컴포넌트(110), 포맷 선택기 컴포넌트(112), 텍스트 생성기 컴포넌트(114), 음성 선택기 컴포넌트(116), 오디오 큐 생성기(118), 컨텐츠 삽입 컴포넌트(120), 기계 학습 엔진(122) 또는 데이터 처리 시스템(102)의 다른 컴포넌트는 개별 컴포넌트, 단일 컴포넌트 또는 데이터 처리 시스템(102)의 일부일 수 있다. 데이터 처리 시스템(102)과 같은 시스템(100)과 그의 컴포넌트는 하나 이상의 프로세서, 논리 디바이스 또는 회로와 같은 하드웨어 요소를 포함할 수 있다. 데이터 처리 시스템(102)의 컴포넌트, 시스템 또는 모듈은 데이터 처리 시스템(102)에 의해 적어도 부분적으로 실행될 수 있다.

[0024] 컴퓨팅 디바이스(140)는 적어도 하나의 센서(148), 변환기(144), 오디오 드라이버(150), 전 처리기(142), 또는 디스플레이 디바이스(146)를 포함하거나, 인터페이스하거나 그렇지 않으면 통신할 수 있다. 센서(148)는 예를 들어 주변광 센서, 근접 센서, 온도 센서, 가속도계, 자이로스코프, 모션 검출기, GPS 센서, 위치 센서, 마이크로폰, 또는 터치 센서를 포함할 수 있다. 변환기(144)는 스피커 또는 마이크로폰을 포함할 수 있다. 오디오 드라이버(150)는 하드웨어 변환기(144)에 소프트웨어 인터페이스를 제공할 수 있다. 오디오 드라이버는 대응하는 음향파 또는 음파를 생성하도록 변환기(144)를 제어하기 위해 데이터 처리 시스템(102)에 의해 제공된 오디오 파일 또는 다른 명령을 실행할 수 있다. 디스플레이 디바이스(146)는 도 3에 도시된 디스플레이(335)의 하나 이상의 컴포넌트 또는 기능을 포함할 수 있다. 전 처리기(142)는 트리거 키워드, 사전 결정된 핫 워드, 개시 키워드 또는 활성화 키워드를 검출하도록 구성될 수 있다. 일부 경우, 트리거 키워드는 (액션 모델(128)을 사용하여 액션 생성기(136)에 의해 선택된 액션과 같은) 액션을 수행하기 위한 요청을 포함할 수 있다. 일부 경우, 트리거 키워드는 컴퓨팅 디바이스(140)를 인에이블 또는 활성화하기 위한 사전 결정된 액션 키워드를 포함할 수 있고, 요청 키워드는 트리거 키워드 또는 핫 워드 뒤에 올 수 있다. 전 처리기(142)는 키워드를 검출하고 키워드에 기초하여 액션을 수행하도록 구성될 수 있다. 전 처리기(142)는 웨이크-업 단어 또는 다른 키워드 또는 핫 워드를 검출할 수 있고, 검출에 응답하여 컴퓨팅 디바이스(140)에 의해 실행되는 데이터 처리 시스템(102)의 자연어 프로세서 컴포넌트(106)를 호출할 수 있다. 일부 경우, 전 처리기(142)는 추가 처리를 위해 데이터 처리 시스템(102)으로 용어들을 전송하기 전에 하나 이상의 용어를 필터링하거나 그 용어들을 수정할 수 있다. 전 처리기(142)는 마이크로폰에 의해 검출된 아날로그 오디오 신호를 디지털 오디오 신호로 변환하고, 디지털 오디오 신호를 운반하는 하나 이상의 데이터 패킷을 네트워크(105)를 통해 데이터 처리 시스템(102) 또는 데이터 처리 시스템(102)에 전송하거나 제공할 수 있다. 일부 경우, 전 처리기(142)는 이러한 전송을 수행하라는 명령을 검출하는 것에 응답하여, 입력 오디오 신호의 일부 또는 전부를 운반하는 데이터 패킷을 자연어 프로세서 컴포넌트(106) 또는 데이터 처리 시스템(102)에 제공할 수 있다. 명령은 예를 들어, 입력 오디오 신호를 포함하는 데이터 패킷을 데이터 처리 시스템(102) 또는 데이터 처리 시스템(102)에 전송하기 위한 트리거 키워드 또는 다른 키워드 또는 승인을 포함할 수 있다.

[0025] 클라이언트 컴퓨팅 디바이스(140)는 (센서(148)를 통해) 클라이언트 컴퓨팅 디바이스(140)에 오디오 입력으로서 음성 쿼리들을 입력하고, 변환기(144)(예를 들어, 스피커)로부터 출력된, 데이터 처리 시스템(102)(또는 3P 디지털 컨텐츠 제공자 디바이스(160))으로부터 클라이언트 컴퓨팅 디바이스(140)로 제공될 수 있는 컴퓨터 생성 음성의 형태로 오디오 출력을 수신하는 최종 사용자와 연관될 수 있다. 컴퓨터 생성 음성에는 실제 사람의 녹음 또는 컴퓨터 생성 언어가 포함될 수 있다.

[0026] 컴퓨팅 디바이스(140)는 애플리케이션(152)을 실행할 수 있다. 데이터 처리 시스템(102)은 컴퓨팅 디바이스(140)가 애플리케이션(152)을 실행할 수 있는 운영 체제를 포함하거나 실행할 수 있다. 애플리케이션(152)은 클라이언트 컴퓨팅 디바이스(140)가 실행, 운영, 개시 또는 제공하도록 구성된 임의의 유형의 애플리케이션을 포함할 수 있다. 애플리케이션(152)은 멀티미디어 애플리케이션, 음악 플레이어, 비디오 플레이어, 웹 브라우저, 워드 프로세서, 모바일 애플리케이션, 데스크탑 애플리케이션, 태블릿 애플리케이션, 전자 게임, 전자 상거래 애플리케이션, 또는 다른 유형의 애플리케이션을 포함할 수 있다. 애플리케이션(152)은 전자 리소스에 대응하는

데이터를 실행, 렌더링, 로드, 파싱, 처리, 제시 또는 출력할 수 있다. 전자 리소스는 예를 들어 웹 사이트, 웹 페이지, 멀티미디어 웹 컨텐츠, 비디오 컨텐츠, 오디오 컨텐츠, 디지털 스트리밍 컨텐츠, 여행 컨텐츠, 엔터테인먼트 컨텐츠, 상품 또는 서비스의 쇼핑과 관련된 컨텐츠 또는 기타 컨텐츠를 포함할 수 있다.

[0027] 컴퓨팅 디바이스(140)에서 실행되는 애플리케이션(152)은 제3자("3P") 전자 리소스 서버(162)로부터 전자 리소스와 관련된 데이터를 수신할 수 있다. 3P 전자 리소스 서버(162)는 애플리케이션에 의한 실행을 위해 전자 리소스를 제공할 수 있다. 3P 전자 리소스 서버(162)는 파일 서버, 웹 서버, 게임 서버, 멀티미디어 서버, 클라우드 컴퓨팅 환경, 또는 애플리케이션이 컴퓨팅 디바이스(140)를 통해 전자 리소스를 제시하거나 제공하게 하는 데이터를 제공하도록 구성된 기타 백엔드 컴퓨팅 시스템을 포함할 수 있다. 컴퓨팅 디바이스(140)는 네트워크(105)를 통해 3P 전자 리소스 서버(162)에 액세스할 수 있다.

[0028] 3P 전자 리소스 서버(162)의 관리자는 전자 리소스를 개발, 구축, 유지 또는 제공할 수 있다. 3P 전자 리소스 서버(162)는 전자 리소스 요청에 응답하여 전자 리소스을 컴퓨팅 디바이스(140)로 전송할 수 있다. 전자 리소스는 URL, 균일 리소스 식별자, 웹 주소, 파일 이름 또는 파일 경로와 같은 식별자와 연관될 수 있다. 3P 전자 리소스 서버(162)는 애플리케이션(152)으로부터 전자 리소스 요청을 수신할 수 있다. 전자 리소스는 전자 문서, 웹 페이지, 멀티미디어 컨텐츠, 스트리밍 컨텐츠(예를 들어, 음악, 뉴스 또는 팟 캐스트), 오디오, 비디오, 텍스트, 이미지, 비디오 게임 또는 다른 디지털 또는 전자 컨텐츠를 포함할 수 있다.

[0029] 데이터 처리 시스템(102)은 적어도 하나의 3P 디지털 컨텐츠 제공자 디바이스(160)에 액세스하거나 그와 상호작용할 수 있다. 3P 디지털 컨텐츠 제공자 디바이스(160)는 네트워크(105)를 통해, 예를 들어 컴퓨팅 디바이스(140), 데이터 처리 시스템(102) 또는 데이터 처리 시스템(102)과 통신하는 프로세서를 갖는 컴퓨팅 디바이스와 같은 적어도 하나의 논리 디바이스를 포함할 수 있다. 3P 디지털 컨텐츠 제공자 디바이스(160)는 적어도 하나의 계산 리소스, 서버, 프로세서 또는 메모리를 포함할 수 있다. 예를 들어, 3P 디지털 컨텐츠 제공자 디바이스(160)는 적어도 하나의 데이터 센터에 위치한 복수의 계산 리소스 또는 서버를 포함할 수 있다. 3P 디지털 컨텐츠 제공자 디바이스(160)는 광고주 디바이스, 서비스 제공자 디바이스 또는 상품 제공자 디바이스를 포함하거나 지칭할 수 있다.

[0030] 3P 디지털 컨텐츠 제공자 디바이스(160)는 컴퓨팅 디바이스(140)에 의한 프리젠테이션을 위한 디지털 컴포넌트를 제공할 수 있다. 디지털 컴포넌트는 컴퓨팅 디바이스(140)의 디스플레이 디바이스(146)를 통해 프리젠테이션하기 위한 시각적 디지털 컴포넌트일 수 있다. 디지털 컴포넌트에는 검색 질의 또는 요청에 대한 응답이 포함될 수 있다. 디지털 컴포넌트에는 데이터베이스, 검색 엔진 또는 네트워크 리소스의 정보가 포함될 수 있다. 예를 들어, 디지털 컴포넌트는 뉴스 정보, 날씨 정보, 스포츠 정보, 백과 사전 항목(entries), 사전 항목 또는 디지털 텍스트북의 정보를 포함할 수 있다. 디지털 컴포넌트는 광고를 포함할 수 있다. 디지털 컴포넌트는 "스니커즈를 구매 하시겠습니까?"라는 메시지와 같은 상품 또는 서비스에 대한 제안을 포함할 수 있다. 3P 디지털 컨텐츠 제공자 디바이스(160)는 질의에 응답하여 제공될 수 있는 일련의 디지털 컴포넌트를 저장하기 위한 메모리를 포함할 수 있다. 3P 디지털 컨텐츠 제공자 디바이스(160)는 또한 시각적 또는 오디오 기반 디지털 컴포넌트(또는 다른 디지털 컴포넌트)를 컨텐츠 선택기 컴포넌트(108)에 의한 선택을 위해 저장될 수 있는 데이터 처리 시스템(102)에 제공할 수 있다. 데이터 처리 시스템(102)은 디지털 컴포넌트를 선택하고 클라이언트 컴퓨팅 디바이스(140)에 디지털 컴포넌트를 제공(또는 컨텐츠 제공자 컴퓨팅 디바이스(160)에 제공하도록 지시)할 수 있다. 디지털 컴포넌트는 시각적만, 오디오만 또는 오디오 및 시각적 데이터와 텍스트, 이미지 또는 비디오 데이터의 조합일 수 있다. 디지털 컴포넌트 또는 컨텐츠 아이템은 이미지, 텍스트, 비디오, 멀티미디어 또는 하나 이상의 포멧으로 된 다른 유형의 컨텐츠를 포함할 수 있다.

[0031] 데이터 처리 시스템(102)은 적어도 하나의 계산 리소스 또는 서버를 갖는 컨텐츠 배치 시스템을 포함할 수 있다. 데이터 처리 시스템(102)은 적어도 하나의 컨텐츠 선택기 컴포넌트(108)를 포함하거나, 그와 인터페이스하거나 통신할 수 있다. 데이터 처리 시스템(102)은 적어도 하나의 디지털 어시스턴트 서버를 포함하거나, 그와 인터페이스하거나 통신할 수 있다.

[0032] 데이터 처리 시스템(102)은 복수의 컴퓨팅 디바이스(140)와 관련된 익명의 컴퓨터 네트워크 활동 정보를 획득할 수 있다. 컴퓨팅 디바이스(140)의 사용자는 사용자의 컴퓨팅 디바이스(140)에 대응하는 네트워크 활동 정보를 획득하기 위해 데이터 처리 시스템(102)을 긍정적으로 승인할 수 있다. 예를 들어, 데이터 처리 시스템(102)은 하나 이상의 유형의 네트워크 활동 정보를 획득하기 위한 동의 (consent)를 컴퓨팅 디바이스(140)의 사용자에게 프롬프트할 수 있다. 컴퓨팅 디바이스(140)의 사용자의 신원은 익명으로 유지될 수 있으며, 컴퓨팅 디바이스(140)는 고유 식별자(예를 들어, 데이터 처리 시스템 또는 컴퓨팅 디바이스의 사용자에 의해 제공되는 사용자

또는 컴퓨팅 디바이스에 대한 고유 식별자)와 연관될 수 있다. 데이터 처리 시스템(102)은 각각의 관찰(observation)을 대응하는 고유 식별자와 연관시킬 수 있다.

[0033] 3P 디지털 컨텐츠 제공자 디바이스(160)는 전자 컨텐츠 캠페인을 설정할 수 있다. 전자 컨텐츠 캠페인은 컨텐츠 선택기 컴포넌트(108)의 데이터 저장소에 컨텐츠 데이터로서 저장될 수 있다. 전자 컨텐츠 캠페인은 공통 테마에 대응하는 하나 이상의 컨텐츠 그룹을 지칭할 수 있다. 컨텐츠 캠페인은 컨텐츠 그룹, 디지털 컴포넌트 데이터 객체 및 컨텐츠 선택 기준을 포함하는 계층적 데이터 구조를 포함할 수 있다. 컨텐츠 캠페인을 생성하기 위해, 3P 디지털 컨텐츠 제공자 디바이스(160)는 컨텐츠 캠페인의 캠페인 레벨 파라미터에 대한 값들을 지정할 수 있다. 캠페인 레벨 파라미터는 예를 들어 캠페인 이름, 디지털 컴포넌트 객체를 배치하기 위한 선호 컨텐츠 네트워크, 컨텐츠 캠페인에 사용될 리소스들의 값, 컨텐츠 캠페인 시작 및 종료 날짜, 컨텐츠 캠페인의 지속 기간, 디지털 컴포넌트 객체 배치 스케줄, 언어, 지리적 위치, 디지털 컴포넌트 객체를 제공할 컴퓨팅 디바이스 유형을 포함할 수 있다. 일부 경우, 노출은 디지털 컴포넌트 객체가 그의 소스(예를 들어, 데이터 처리 시스템(102) 또는 3P 디지털 컨텐츠 제공자 디바이스(160))로부터 페치되고 카운트될 수 있는 때를 지칭할 수 있다. 일부 경우, 클릭 사기의 가능성으로 인해, 로봇 활동이 노출로서 필터링 및 제외될 수 있다. 따라서 경우에 따라 노출은 로봇 활동 및 오류 코드에서 필터링된 브라우저의 페이지 요청에 대한 웹 서버의 응답 측정을 지칭할 수 있으며, 컴퓨팅 디바이스(140)에 디스플레이하기 위해 디지털 컴포넌트 객체를 렌더링할 기회에 가능한 한 가까운 지점에서 기록될 수 있다. 일부 경우, 노출은 가시적 또는 가청 노출을 지칭할 수 있는데, 예를 들어, 디지털 컴포넌트 객체는 클라이언트 컴퓨팅 디바이스(140)의 디스플레이 디바이스에서 적어도 부분적으로(예를 들어, 20%, 30%, 30%, 40%, 50%, 60%, 70% 이상) 볼 수 있거나, 또는 컴퓨팅 디바이스(140)의 스피커를 통해 들을 수 있다. 클릭 또는 선택은 가청 노출에 대한 음성 응답, 마우스 클릭, 터치 상호 작용, 제스처, 흔들기, 오디오 상호 작용 또는 키보드 클릭과 같은 디지털 컴포넌트 객체와의 사용자 상호 작용을 지칭할 수 있다. 전환(conversion)은 예를 들어, 제품 또는 서비스 구매, 설문 조사 완료, 디지털 컴포넌트에 해당하는 실제 매장 방문 또는 전자 거래 완료와 같이 디지털 컴포넌트 이의(objection)에 대해 원하는 액션을 취하는 사용자를 지칭할 수 있다.

[0034] 3P 디지털 컨텐츠 제공자 디바이스(160)는 컨텐츠 캠페인을 위한 하나 이상의 컨텐츠 그룹을 추가로 설정할 수 있다. 컨텐츠 그룹에는 키워드, 단어, 용어, 문구, 지리적 위치, 컴퓨팅 디바이스 유형, 하루 중 시간, 관심사, 토픽 또는 버티컬(vertical)와 같은 하나 이상의 디지털 컴포넌트 객체 및 대응 컨텐츠 선택 기준이 포함된다. 동일한 컨텐츠 캠페인 하의 컨텐츠 그룹은 동일한 캠페인 레벨 파라미터들를 공유할 수 있지만, 키워드, 제외 키워드(예를 들어, 주요 컨텐츠에 제외 키워드가 있는 경우 디지털 컴포넌트의 배치를 차단함), 키워드 입찰가, 입찰가 또는 컨텐츠 캠페인과 관련된 파라미터와 같은 특정 컨텐츠 그룹 레벨 파라미터에 대한 맞춤형 사양을 가질 수 있다.

[0035] 새로운 컨텐츠 그룹을 생성하기 위해, 3P 디지털 컨텐츠 제공자 디바이스(160)는 컨텐츠 그룹의 컨텐츠 그룹 레벨 파라미터에 대한 값을 제공할 수 있다. 컨텐츠 그룹 레벨 파라미터에는 예를 들어 컨텐츠 그룹 이름 또는 컨텐츠 그룹 테마, 상이한 컨텐츠 배치 기회(예를 들어, 자동 배치 또는 선택 배치) 또는 결과(예를 들어, 클릭, 노출 또는 전환)에 대한 입찰가가 포함된다. 컨텐츠 그룹 이름 또는 컨텐츠 그룹 테마는 3P 디지털 컨텐츠 제공자 디바이스(160)가 디스플레이를 위해 컨텐츠 그룹의 디지털 컴포넌트 객체들이 선택되는 토픽 또는 주제를 캡처하기 위해 사용할 수 있는 하나 이상의 용어일 수 있다. 예를 들어, 자동차 판매점은 운송하는 차량의 각 브랜드에 대해 상이한 컨텐츠 그룹을 생성할 수 있으며, 운송하는 차량의 각 모델에 대해 상이한 컨텐츠 그룹을 추가로 생성할 수 있다. 자동차 판매점에서 사용할 수 있는 컨텐츠 그룹 테마의 예는 예를 들어 "A사의 스포츠카" "B사의 스포츠 카", "C사의 세단", "C사의 트럭", "C사의 하이브리드" 또는 " D사의 하이브리드를 포함할 수 있다. 예시적인 컨텐츠 캠페인 테마는 예를 들어 "C사의 하이브리드" 및 "D사의 하이브리드" 모두의 컨텐츠 그룹을 포함할 수 있다.

[0036] 3P 디지털 컨텐츠 제공자 디바이스(160)는 각각의 컨텐츠 그룹에 하나 이상의 키워드 및 디지털 컴포넌트 객체를 제공할 수 있다. 키워드는 디지털 컴포넌트 객체들과 연관되거나 식별되는 제품 또는 서비스와 관련된 용어들을 포함할 수 있다. 키워드에는 하나 이상의 용어 또는 구문이 포함될 수 있다. 예를 들어 자동차 판매점은 컨텐츠 그룹 또는 컨텐츠 캠페인의 키워드로서 '스포츠카', 'V-6 엔진', '4륜 구동', '연비'를 포함할 수 있다. 일부 경우, 컨텐츠 제공자는 특정 용어 또는 키워드에 대한 컨텐츠 배치를 회피, 방지, 차단 또는 비활성화하기 위해 제외 키워드를 지정할 수 있다. 컨텐츠 제공자는 디지털 컴포넌트 객체를 선택하는데 사용되는 매칭 유형(예를 들어, 정확한(exact) 매치, 구문(phase) 매치 또는 브로드 (broad) 매치)을 지정할 수 있다.

[0037] 3P 디지털 컨텐츠 제공자 디바이스(160)는 3P 디지털 컨텐츠 제공자 디바이스(160)에 의해 제공되는 디지털 컴

포넌트 객체를 선택하기 위해 데이터 처리 시스템(102)에 의해 사용될 하나 이상의 키워드를 제공할 수 있다. 3P 디지털 컨텐츠 제공자 디바이스(160)는 입찰할 하나 이상의 키워드를 식별할 수 있고, 다양한 키워드에 대한 입찰 금액을 추가로 제공할 수 있다. 3P 디지털 컨텐츠 제공자 디바이스(160)는 디지털 컴포넌트 객체를 선택하기 위해 데이터 처리 시스템(102)에 의해 사용될 추가 컨텐츠 선택 기준을 제공할 수 있다. 다수의 3P 디지털 컨텐츠 제공자 디바이스(160)는 동일하거나 상이한 키워드에 입찰할 수 있고, 데이터 처리 시스템(102)은 전자 메시지의 키워드의 표시를 수신하는 것에 응답하여 컨텐츠 선택 프로세스 또는 광고 경매를 실행할 수 있다.

[0038] 3P 디지털 컨텐츠 제공자 디바이스(160)는 데이터 처리 시스템(102)에 의한 선택을 위해 하나 이상의 디지털 컴포넌트 객체를 제공할 수 있다. 데이터 처리 시스템(102)은 (예를 들어, 컨텐츠 선택기 컴포넌트(108)을 통해) 리소스 할당, 컨텐츠 스케줄, 최대 입찰가, 키워드 및 컨텐츠 그룹에 대해 지정된 다른 선택 기준과 매칭하는 컨텐츠 배치 기회가 이용 가능해질 때 디지털 컴포넌트 객체를 선택할 수 있다. 음성 디지털 컴포넌트, 오디오 디지털 컴포넌트, 텍스트 디지털 컴포넌트, 이미지 디지털 컴포넌트, 비디오 디지털 컴포넌트, 멀티미디어 디지털 컴포넌트 또는 디지털 컴포넌트 링크와 같은 상이한 유형의 디지털 컴포넌트 객체가 컨텐츠 그룹에 포함될 수 있다. 디지털 컴포넌트를 선택하면, 데이터 처리 시스템(102)은 컴퓨팅 디바이스(140) 또는 컴퓨팅 디바이스(140)의 디스플레이 디바이스에서 렌더링하면서 컴퓨팅 디바이스(140)를 통해 프리젠테이션하기 위해 디지털 컴포넌트 객체를 전송할 수 있다. 렌더링은 디스플레이 디바이스에 디지털 컴포넌트를 디스플레이하거나 컴퓨팅 디바이스(140)의 스피커를 통해 디지털 컴포넌트를 재생하는 것을 포함할 수 있다. 데이터 처리 시스템(102)은 디지털 컴포넌트 객체를 렌더링하도록 컴퓨팅 디바이스(140)에 명령을 제공할 수 있다. 데이터 처리 시스템(102)은 컴퓨팅 디바이스(140)의 자연어 프로세서 컴포넌트(106) 또는 컴퓨팅 디바이스(140)의 오디오 드라이버(150)에 오디오 신호 또는 음향 파를 생성하도록 지시할 수 있다. 데이터 처리 시스템(102)은 선택된 디지털 컴포넌트 객체를 제시하도록 컴퓨팅 디바이스(140)에 의해 실행되는 애플리케이션에 지시할 수 있다. 예를 들어, 애플리케이션(예를 들어, 디지털 음악 스트리밍 애플리케이션)은 디지털 컴포넌트 객체가 제시될 수 있는 슬롯(예를 들어, 컨텐츠 슬롯)(예를 들어, 오디오 슬롯 또는 시각적 슬롯)을 포함할 수 있다.

[0039] 데이터 처리 시스템(102)은 적어도 하나의 인터페이스(104)를 포함할 수 있다. 데이터 처리 시스템(102)은 예를 들어 데이터 패킷을 사용하여 정보를 수신 및 전송하도록 설계, 구성, 작성 또는 동작하는 인터페이스(104)를 포함할 수 있다. 인터페이스(104)는 네트워크 프로토콜과 같은 하나 이상의 프로토콜을 사용하여 정보를 수신하고 전송할 수 있다. 인터페이스(104)는 하드웨어 인터페이스, 소프트웨어 인터페이스, 유선 인터페이스 또는 무선 인터페이스를 포함할 수 있다. 인터페이스(104)는 하나의 포맷에서 다른 포맷으로 데이터를 변환 또는 포맷하는 것을 용이하게 할 수 있다. 예를 들어, 인터페이스(104)는 소프트웨어 컴포넌트와 같은 다양한 컴포넌트 사이의 통신을 위한 정의를 포함하는 애플리케이션 프로그래밍 인터페이스를 포함할 수 있다. 인터페이스(104)는 자연어 프로세서 컴포넌트(106), 컨텐츠 선택기 컴포넌트(108), 컨텐츠 변환 컴포넌트(110) 및 데이터 저장소(124) 사이와 같이 시스템(100)의 하나 이상의 컴포넌트 사이의 통신을 용이하게 할 수 있다.

[0040] 인터페이스(104)는 네트워크(105)를 통해 데이터 처리 시스템(102)으로부터 떨어진 컴퓨팅 디바이스(140)의 마이크로폰(예를 들어, 센서(148))에 의해 검출된 입력 오디오 신호를 포함하는 데이터 패킷을 수신할 수 있다. 컴퓨팅 디바이스(140)의 사용자는 음성(spech) 또는 음성(voice) 입력을 컴퓨팅 디바이스(140)에 제공하고, 컴퓨팅 디바이스(140)가 입력 오디오 신호 또는 전 처리기(142)에 의한 오디오 신호에 기초하여 생성된 데이터 패킷을 데이터 처리 시스템(102)으로 전송하도록 지시하거나 야기할 수 있다.

[0041] 데이터 처리 시스템(102)은 데이터 패킷 또는 입력 오디오 신호를 파싱하도록 설계, 구성 및 동작하는 자연어 프로세서 컴포넌트(106)를 포함하거나, 그와 인터페이스하거나 통신할 수 있다. 자연어 프로세서 컴포넌트(106)는 데이터 처리 시스템(102)에서 하드웨어, 전자 회로, 애플리케이션, 스크립트 또는 프로그램을 포함할 수 있다. 자연어 프로세서 컴포넌트(106)는 입력 신호, 데이터 패킷 또는 기타 정보를 수신할 수 있다. 자연어 프로세서 컴포넌트(106)는 음성을 포함하는 입력 오디오 신호를 처리하여 음성을 텍스트로 전사한 다음 자연어 처리를 수행하여 전사된 텍스트를 이해하도록 구성된 음성 인식기를 포함하거나 지칭할 수 있다. 자연어 프로세서 컴포넌트(106)는 인터페이스(104)를 통해 데이터 패킷 또는 다른 입력을 수신할 수 있다. 자연어 프로세서 컴포넌트(106)는 데이터 처리 시스템(102)의 인터페이스(104)로부터 입력 오디오 신호를 수신하고, 출력 오디오 신호를 렌더링하기 위해 클라이언트 컴퓨팅 디바이스의 컴포넌트를 구동하기 위한 애플리케이션을 포함할 수 있다. 데이터 처리 시스템(102)은 오디오 입력 신호를 포함하거나 식별하는 데이터 패킷 또는 다른 신호를 수신할 수 있다. 예를 들어, 자연어 프로세서 컴포넌트(106)는 오디오 신호를 수신 또는 획득하고 오디오 신호를 파싱할 수 있는 NLP 기술, 기능 또는 컴포넌트로 구성될 수 있다. 자연어 프로세서 컴포넌트(106)는 인간과 컴퓨터 간의 상호 작용을 제공할 수 있다. 자연어 프로세서 컴포넌트(106)는 자연어를 이해하고 데이터 처리 시스템

(102)이 인간 또는 자연어 입력으로부터 의미를 도출할 수 있도록 하는 기술로 구성될 수 있다. 자연어 프로세서 컴포넌트(106)는 통계적 기계 학습과 같은 기계 학습에 기초한 기술을 포함하거나 그로 구성될 수 있다. 자연어 프로세서 컴포넌트(106)는 입력 오디오 신호를 파싱하기 위해 결정 트리, 통계 모델 또는 확률 모델을 이용할 수 있다. 자연어 프로세서 컴포넌트(106)는 예를 들어 명명된 엔티티 인식(예를 들어, 텍스트 스트림이 주어지면, 텍스트에 있는 어떤 아이템이 사람이나 장소와 같은 적절한 이름에 매핑되는지, 그리고 사람, 위치 또는 조직과 같은 각각의 그러한 이름의 유형이 무엇인지 결정), 자연어 생성(예를 들어, 컴퓨터 데이터베이스 또는 의미론적 의도의 정보를 이해할 수 있는 인간 언어로 변환), 자연어 이해(예를 들어, 텍스트를 컴퓨터 모듈이 조작할 수 있는 1차 논리 구조와 같은 보다 형식적인 표현으로 변환), 기계 번역(예를 들어, 한 인간 언어에서 다른 언어로 텍스트를 자동 번역), 형태학적 세분화(예를 들어, 단어를 개별 형태소로 분리하고 형태소의 클래스를 식별하는데, 이는 고려되는 언어의 단어 형태나 구조의 복잡성에 기초하여 어려울 수 있음), 질문 답변(예를 들어, 구체적이거나 개방형일 수 있는 인간 언어 질문에 대한 답변 결정), 의미론적 처리(예를 들어, 식별된 단어를 유사한 의미를 가진 다른 단어와 연결하기 위해 단어를 식별하고 그 의미를 인코딩한 후 발생할 수 있는 처리)와 같은 기능을 수행할 수 있다.

[0042] 자연어 프로세서 컴포넌트(106)는 (예를 들어, NLP 기술, 기능 또는 컴포넌트를 이용하여) 포함된 트레이닝 데이터에 기초한 기계 학습 모델 트레이닝을 사용하여 오디오 입력 신호를 인식된 텍스트로 변환할 수 있다. 오디오 파형 세트는 데이터 저장소(124) 또는 데이터 처리 시스템(102)에 액세스 가능한 다른 데이터베이스에 저장될 수 있다. 대표 파형은 대규모 사용자 세트에 대해 생성된 다음 사용자로부터의 음성 샘플로 보강될 수 있다. 오디오 신호가 인식된 텍스트로 변환된 후, 자연어 프로세서 컴포넌트(106)는 예를 들어 데이터 처리 시스템(102)이 제공할 수 있는 액션들과 함께, 사용자간에 걸쳐 또는 수동 사양을 통해 트레이닝되었던 데이터 저장소(124)에 저장된 모델을 사용함으로써 텍스트를 연관된 단어들에 매칭시킬 수 있다.

[0043] 오디오 입력 신호는 클라이언트 컴퓨팅 디바이스(140)의 센서(148) 또는 변환기(144)(예를 들어, 마이크로폰)에 의해 검출될 수 있다. 변환기(144), 오디오 드라이버(150) 또는 기타 컴포넌트를 통해 클라이언트 컴퓨팅 디바이스(140)는 오디오 입력 신호를 데이터 처리 시스템(102)에 제공할 수 있는데, 여기서 데이터 처리 시스템(102)에 오디오 입력 신호를 제공할 수 있으며, 이것은 (예를 들어, 인터페이스(104)에 의해) 수신되어 NLP 컴포넌트(106)에 제공되거나 데이터 저장소 (124)에 저장된다.

[0044] 자연어 프로세서 컴포넌트(106)는 입력 오디오 신호를 획득할 수 있다. 입력 오디오 신호로부터, 자연어 프로세서 컴포넌트(106)는 적어도 하나의 요청 또는 적어도 하나의 트리거 키워드, 키워드 또는 요청을 식별할 수 있다. 요청은 입력 오디오 신호의 의도 또는 주제를 나타낼 수 있다. 키워드는 취해질 가능성이 있는 액션 유형을 나타낼 수 있다. 예를 들어, 자연어 프로세서 컴포넌트(106)는 입력 오디오 신호를 파싱하여, 애플리케이션을 호출하기 위한 적어도 하나의 요청, 컨텐츠 아이템과의 상호 작용 또는 컨텐츠 요청을 식별할 수 있다. 자연어 프로세서 컴포넌트(106)는 입력 오디오 신호를 파싱하여, 저녁 식사와 영화에 참석하기 위해 저녁에 집을 떠나라는 요청과 같은 적어도 하나의 요청을 식별할 수 있다. 키워드는 취해야 할 액션을 나타내는 적어도 하나의 단어, 구문, 어근 또는 부분 단어 또는 파생어를 포함할 수 있다. 예를 들어, 입력 오디오 신호로부터의 키워드 "go" 또는 "to go to"는 이동(transport)이 필요함을 나타낼 수 있다. 이 예에서, 입력 오디오 신호(또는 식별된 요청)는 이동 의도를 직접적으로 표현하지 않지만, 그 키워드는 이동이 요청에 의해 표시된 적어도 하나의 다른 액션에 대한 보조 액션임을 나타낸다.

[0045] 자연어 프로세서 컴포넌트(106)는 입력 오디오 신호를 파싱하여 요청 및 키워드를 식별, 결정, 검색 또는 획득할 수 있다. 예를 들어, 자연어 프로세서 컴포넌트(106)는 입력 오디오 신호에 의미론적 처리 기술을 적용하여 키워드 또는 요청을 식별할 수 있다. 자연어 프로세서 컴포넌트(106)는 입력 오디오 신호에 의미론적 처리 기술을 적용하여 하나 이상의 키워드를 식별할 수 있다. 키워드에는 하나 이상의 용어 또는 구문이 포함될 수 있다. 자연어 프로세서 컴포넌트(106)는 의미론적 처리 기술을 적용하여 디지털 액션을 수행하기 위한 의도를 식별할 수 있다.

[0046] 예를 들어, 컴퓨팅 디바이스(140)는 클라이언트 컴퓨팅 디바이스(140)의 센서(148)(예를 들어, 마이크로폰)에 의해 검출된 입력 오디오 신호를 수신할 수 있다. 입력 오디오 신호는 "디지털 어시스턴트, 세탁 및 드라이 클리닝을 할 사람이 필요해"일 수 있다. 클라이언트 컴퓨팅 디바이스(140)의 전 처리기(142)는 "디지털 어시스턴트"와 같이 입력 오디오 신호에 있는 웨이크 업 워드, 핫 워드 또는 트리거 키워드를 검출할 수 있다. 전 처리기(142)는 입력 오디오 신호에 있는 오디오 서명 또는 파형을 트리거 키워드에 대응하는 모델 오디오 서명 또는 파형과 비교함으로써 웨이크 업 단어, 핫 워드 또는 트리거 키워드를 검출할 수 있다. 전 처리기(142)는 입력 오디오 신호가 그 입력 오디오 신호가 자연어 프로세서 컴포넌트(106)에 의해 처리될 것임을 나타내는 웨이크

업 워드, 핫 워드 또는 트리거 키워드를 포함한다고 결정할 수 있다. 핫 워드, 웨이크 업 워드 또는 트리거 키워드를 검출하는 것에 응답하여, 전 처리기(142)는 자연어 프로세서 컴포넌트(106)에 의한 처리를 위해 상기 검출된 입력 오디오 신호를 데이터 처리 시스템(102)에 결정, 승인, 라우팅, 포워딩 또는 제공할 수 있다. .

[0047] 자연어 프로세서 컴포넌트(106)는 입력 오디오 신호를 수신하고, 의미론적 처리 기술 또는 기타 자연어 처리 기술을 문장을 포함하는 입력 오디오 신호에 적용함으로써 트리거 문구인 "세탁 해줘" 및 "드라이 클리닝 해줘"을 식별할 수 있다. 일부 경우, 자연어 프로세서 컴포넌트(106)는 입력 오디오 신호에 대응하는 데이터 패킷을 데이터 처리 시스템(102)에 제공하여 자연어 프로세서 컴포넌트(106)가 입력 오디오 신호를 처리하게 할 수 있다. 자연어 프로세서 컴포넌트(106)는 디지털 어시스턴트 서버와 함께 또는 그를 통해 입력 오디오 신호를 처리할 수 있다. 자연어 프로세서 컴포넌트(106)는 세탁 및 드라이 클리닝과 같은 다수의 키워드를 추가로 식별할 수 있다.

[0048] 자연어 프로세서 컴포넌트(106)는 정보 검색 또는 다른 정보 요청을 수행하는 것에 대응하는 검색 질의, 키워드, 의도 또는 구문을 식별할 수 있다. 자연어 프로세서 컴포넌트(106)는 입력 오디오 신호가 토픽, 이벤트, 현재 이벤트, 뉴스 이벤트, 사전적 정의, 이력 이벤트, 사람, 장소 또는 사물에 관한 정보 요청에 대응한다고 결정할 수 있다. 예를 들어, 자연어 프로세서 컴포넌트(106)는 입력 오디오 신호가 여행 준비, 차량 예약, 정보 획득, 웹 검색 수행, 주가 확인, 애플리케이션 시작, 뉴스 확인, 음식 주문 또는 기타 제품, 상품 또는 서비스 쇼핑을 위한 질의, 요청, 의도 또는 액션에 대응한다고 결정할 수 있다.

[0049] 자연어 프로세서 컴포넌트(106)는 하나 이상의 기술을 사용하여 입력 오디오 신호를 파싱하거나 처리할 수 있다. 기술에는 규칙 기반 기술 또는 통계 기술이 포함될 수 있다. 기술은 기계 학습 또는 딥 러닝을 이용할 수 있다. 예시적인 기술은 명명된 엔티티 인식, 감정 분석, 텍스트 요약, 양태 추출(mining) 또는 토픽 추출을 포함할 수 있다. 기술은 텍스트 임베딩(예를 들어, 문자열의 실제값 벡터 표현), 기계 번역(예를 들어, 언어 분석 및 언어 생성) 또는 다이얼로그 및 대화(예를 들어, 인공 지능에서 사용되는 모델)를 포함하거나 이를 기반으로 할 수 있다. 기술은 표제어 추출(lemmatization), 형태학적 분할, 단어 분할, 품사 태깅, 파싱, 문장 분리 또는 형태소 분석과 같은 구문 기술(예를 들어, 문법에 기초하여 문장에서 단어 배열)을 결정하거나 이용하는 것을 포함할 수 있다. 기술은 명명된 엔티티 인식(예를 들어, 애플리케이션(152)의 이름, 사람 또는 장소와 같이 현재 그룹으로 식별되고 분류될 수 있는 텍스트 부분 결정), 단어 의미 명확화 또는 자연어 생성과 같은 의미론적 기술을 결정하거나 이용하는 것을 포함할 수 있다.

[0050] 일부 경우, 자연어 프로세서 컴포넌트(106)는 애플리케이션(152)을 시작 (launch)하기 위한 요청을 식별할 수 있고, 애플리케이션(152)을 시작하기 위한 명령을 컴퓨팅 디바이스(140)에 제공할 수 있다. 일부 경우, 애플리케이션(152)은 자연어 프로세서 컴포넌트(106)가 입력 오디오 신호를 수신하기 전에 이미 시작되었을 수 있다. 예를 들어, 입력 오디오 신호의 처리 또는 파싱에 기초하여, 자연어 프로세서 컴포넌트(106)는 호출, 시작, 개방 또는 활성화할 애플리케이션(152)을 식별할 수 있다. 자연어 프로세서 컴포넌트(106)는 용어, 키워드, 트리거 키워드 또는 문구를 식별하기 위해 입력 오디오 신호를 파싱하는 것에 기초하여 애플리케이션(152)을 식별할 수 있다. 자연어 프로세서 컴포넌트(106)는 애플리케이션(152)을 식별하기 위해 식별된 용어, 키워드, 트리거 키워드 또는 문구를 사용하여 데이터 저장소(124)에서 검색(lookup)을 수행할 수 있다. 일부 경우, 키워드는 "애플리케이션_이름_A" 또는 "애플리케이션_이름_B"와 같은 애플리케이션(152)의 식별자를 포함할 수 있다. 일부 경우, 키워드는 승차 공유 애플리케이션, 레스토랑 예약 애플리케이션, 영화 티켓 애플리케이션, 뉴스 애플리케이션, 날씨 애플리케이션, 내비게이션 애플리케이션, 스트리밍 음악 애플리케이션, 스트리밍 비디오 애플리케이션, 레스토랑 리뷰 애플리케이션, 또는 다른 유형 또는 카테고리의 애플리케이션 (152)과 같은 애플리케이션(152)의 유형 또는 카테고리를 나타낼 수 있다. 입력 오디오 신호의 수신 전에 애플리케이션(152)이 이미 시작되어 실행될 수 있는 경우, 자연어 프로세서 컴포넌트(106)는 입력 오디오 신호를 처리하여 애플리케이션(152)에서 수행할 액션을 결정하거나 애플리케이션(152)에 의해 렌더링된 전자 리소스를 통해 제시되는 콜투 액션(call-to-action)에 응답할 수 있다.

[0051] 데이터 처리 시스템(102)은 컴퓨터 네트워크를 통해 컴퓨팅 디바이스(140)에 프리젠테이션하기 위한 컨텐츠 요청을 수신할 수 있다. 데이터 처리 시스템(102)은 클라이언트 컴퓨팅 디바이스(140)의 마이크로폰에 의해 검출된 입력 오디오 신호를 처리함으로써 요청을 식별할 수 있다. 요청은 디바이스 유형, 위치 및 그 요청과 관련된 키워드와 같은 요청의 선택 기준을 포함할 수 있다. 선택 기준은 컴퓨팅 디바이스(140)의 컨텍스트에 관한 정보를 포함할 수 있다. 컴퓨팅 디바이스(140)의 컨텍스트는 컴퓨팅 디바이스(140)에서 실행중인 애플리케이션에 관한 정보, 컴퓨팅 디바이스(140)의 위치에 관한 정보, 컴퓨팅 디바이스(140)를 통해(예를 들어, 애플리케이션(152)을 통해) 렌더링, 제시, 제공 또는 액세스되는 컨텐츠에 관한 정보를 포함할 수 있다. 예를 들어, 컨텐츠

선택 기준은 아티스트, 노래 제목, 또는 디지털 스트리밍 음악 애플리케이션(152)을 통해 재생되는 음악과 관련된 장르와 같은 정보 또는 키워드를 포함할 수 있다. 일부 경우, 컨텐츠 선택 기준은 애플리케이션(152)의 브라우징 히스토리와 관련된 키워드를 포함할 수 있다.

[0052] 데이터 처리 시스템(102)은 3P 디지털 컨텐츠 제공자 디바이스(160)에 의해 제공되는 디지털 컴포넌트를 선택하도록 결정할 수 있다. 데이터 처리 시스템(102)은 컴퓨팅 디바이스(140)로부터의 요청에 응답하여 디지털 컴포넌트를 선택하도록 결정할 수 있다. 데이터 처리 시스템(102)은 애플리케이션(152)에 있는 컨텐츠 슬롯을 식별하는 것에 응답하여 디지털 컴포넌트를 선택하도록 결정할 수 있다. 데이터 처리 시스템(102)은 이벤트, 조건, 트리거에 응답하거나 시간 간격에 기초하여 디지털 컴포넌트를 선택하도록 결정할 수 있다.

[0053] 데이터 처리 시스템(102)은 데이터 저장소(124) 또는 하나 이상의 3P 디지털 컨텐츠 제공자 디바이스(160)에 의해 제공되는 컨텐츠를 포함할 수 있는 데이터베이스로부터 디지털 컴포넌트 객체를 선택하고, 네트워크(105)를 경유하여 컴퓨팅 디바이스(140)를 통해 프리젠테이션하기 위한 디지털 컴포넌트를 제공할 수 있다. 컴퓨팅 디바이스(140)는 디지털 컴포넌트 객체와 상호 작용할 수 있다. 컴퓨팅 디바이스(140)는 디지털 컴포넌트에 대한 오디오 응답을 수신할 수 있다. 컴퓨팅 디바이스(140)는 컴퓨팅 디바이스(140)가 상품 또는 서비스 제공자를 식별하고, 상품 또는 서비스 제공자의 상품 또는 서비스를 요청하고, 서비스 제공자에게 서비스 수행을 지시하고, 서비스 제공자에게 정보를 전송하거나 상품 또는 서비스 제공자 디바이스에 질의하게 하거나 할 수 있는 하이퍼링크 또는 디지털 컴포넌트 객체와 관련된 다른 버튼을 선택하라는 표시를 수신할 수 있다.

[0054] 데이터 처리 시스템(102)은 컨텐츠 선택기 컴포넌트(108)를 포함, 실행 또는 그와 통신하여 요청, 질의, 키워드 또는 컨텐츠 선택 기준을 수신하고 수신된 정보에 기초하여 디지털 컴포넌트를 선택할 수 있다. 데이터 처리 시스템(102)은 실시간 컨텐츠 선택 프로세스에 입력된 컨텐츠 선택 기준에 기초하여 디지털 컴포넌트 객체를 선택할 수 있다. 데이터 처리 시스템(102)은 다수의 제3자 컨텐츠 제공자(160)에 의해 제공되는 다수의 디지털 컴포넌트 객체를 저장하는 데이터 저장소(124)로부터 디지털 컴포넌트 객체를 선택할 수 있다.

[0055] 데이터 처리 시스템(102)은 컨텐츠 데이터(132) 데이터 구조 또는 데이터 저장소(124)의 데이터베이스에서 디지털 컴포넌트 객체를 선택하는데 사용되는 정보를 저장할 수 있다. 컨텐츠 데이터(132)는 컨텐츠 선택 기준, 디지털 컴포넌트 객체, 이력 수행 정보, 선호도, 또는 디지털 컴포넌트 객체를 선택하고 전달하는데 사용되는 기타 정보를 포함할 수 있다.

[0056] 컨텐츠 선택기 컴포넌트(108)는 실시간 컨텐츠 선택 프로세스를 통해 디지털 컴포넌트를 선택할 수 있다. 컨텐츠 선택 프로세스는 예를 들어 검색 엔진을 통해 검색을 수행하거나 원격 서버 또는 3P 디지털 컨텐츠 제공자 디바이스(160)와 같은 디바이스에 저장된 데이터베이스에 액세스하는 것을 포함할 수 있다. 컨텐츠 선택 프로세스는 제3자 컨텐츠 제공자(160)에 의해 제공되는 스폰서 디지털 컴포넌트 객체를 선택하는 것을 지칭하거나 포함할 수 있다. 실시간 컨텐츠 선택 프로세스는 컴퓨팅 디바이스(140)에 제공할 하나 이상의 디지털 컴포넌트를 선택하기 위해 다수의 컨텐츠 제공자에 의해 제공되는 디지털 컴포넌트가 파싱, 처리, 가중 또는 매칭되는 서비스를 포함할 수 있다. 컨텐츠 선택기 컴포넌트(108)는 실시간으로 컨텐츠 선택 프로세스를 수행할 수 있다. 컨텐츠 선택 프로세스를 실시간으로 수행하는 것은 클라이언트 컴퓨팅 디바이스(140)를 통해 수신된 컨텐츠 요청에 응답하여 컨텐츠 선택 프로세스를 수행하는 것을 지칭할 수 있다. 실시간 컨텐츠 선택 프로세스는 요청을 수신한 시간 간격(예를 들어, 1초, 2초, 5초, 10초, 20초, 30초, 1 분, 2분, 3분, 5분, 10분 또는 20분)내에서 수행(예를 들어, 시작 또는 완료)될 수 있다. 실시간 컨텐츠 선택 프로세스는 클라이언트 컴퓨팅 디바이스(140)와의 통신 세션 동안 또는 통신 세션이 종료된 후 시간 간격 내에 수행될 수 있다. 실시간 컨텐츠 선택 프로세스는 온라인 컨텐츠 아이템 경매를 지칭하거나 포함할 수 있다.

[0057] 음성 기반 환경에서 프리젠테이션하기 위한 디지털 컴포넌트들을 선택하기 위해, 데이터 처리 시스템(102)은 (예를 들어, 자연어 프로세서 컴포넌트(106)의 NLP 컴포넌트를 통해) 입력 오디오 신호를 파싱하여 쿼리, 키워드를 식별하고, 키워드 및 다른 컨텐츠 선택 기준을 사용하여 매칭되는 디지털 컴포넌트를 선택할 수 있다. 컨텐츠 선택기 컴포넌트(108)는 그 요청 전에 컴퓨팅 디바이스(140) 디바이스에서 실행하는 애플리케이션(152)에 의해 렌더링된 컨텐츠와 관련된 키워드들에 기초하여 디지털 컴포넌트 객체를 선택할 수 있다. 데이터 처리 시스템(102)은 브로드 매치, 정확한 매치 또는 구문 매치에 기초하여 매치되는 디지털 컴포넌트를 선택할 수 있다. 예를 들어, 컨텐츠 선택기 컴포넌트(108)는 후보 디지털 컴포넌트의 주제를 분석, 파싱 또는 처리하여, 후보 디지털 컴포넌트의 주제가 클라이언트 컴퓨팅 디바이스(140)의 마이크로폰에 의해 검출된 입력 오디오 신호의 키워드 또는 문구의 주제에 대응하는지 여부를 결정할 수 있다. 컨텐츠 선택기 컴포넌트(108)는 이미지 처리 기술, 문자 인식 기술, 자연어 처리 기술, 또는 데이터베이스 조회를 사용하여 후보 디지털 컴포넌트의

음성, 오디오, 용어, 문자, 텍스트, 심볼 또는 이미지를 식별, 분석 또는 인식할 수 있다. 후보 디지털 컴포넌트는 후보 디지털 컴포넌트의 주제를 나타내는 메타 데이터를 포함할 수 있으며, 이 경우 컨텐츠 선택기 컴포넌트(108)는 메타 데이터를 처리하여 후보 디지털 컴포넌트의 주제가 입력 오디오 신호에 대응하는지 여부를 결정할 수 있다.

[0058] 3P 디지털 컨텐츠 제공자(160)는 디지털 컴포넌트를 포함하는 컨텐츠 캠페인을 설정할 때 추가 인디케이터(indicators)를 제공할 수 있다. 컨텐츠 제공자는 컨텐츠 선택기 컴포넌트(108)가 후보 디지털 컴포넌트에 관한 정보를 사용하여 조회을 수행함으로써 식별할 수 있는 컨텐츠 캠페인 또는 컨텐츠 그룹 수준에서 정보를 제공할 수 있다. 예를 들어, 후보 디지털 컴포넌트는 컨텐츠 그룹, 컨텐츠 캠페인 또는 컨텐츠 제공자에 매핑될 수 있는 고유 식별자를 포함할 수 있다. 컨텐츠 선택기 컴포넌트(108)는 데이터 저장소(124)의 컨텐츠 캠페인 데이터 구조에 저장된 정보에 기초하여, 3P 디지털 컨텐츠 제공자 디바이스(160)에 관한 정보를 결정할 수 있다.

[0059] 컨텐츠 선택기 컴포넌트(108)에 의해 선택된 디지털 컴포넌트 객체의 포맷 또는 모달리티(양식)는 시각적, 시청각적 또는 오디오 전용일 수 있다. 시각적 전용 포멧을 가진 디지털 컴포넌트 객체는 이미지 또는 텍스트가 있는 이미지일 수 있다. 오디오 전용 포멧을 가진 디지털 컴포넌트 객체는 오디오 트랙일 수 있다. 시청각적 포멧을 갖는 디지털 컴포넌트 객체는 비디오 클립일 수 있다. 디지털 컴포넌트 객체들은 디지털 컴포넌트 객체의 포멧에 기초하여 상이한 유형의 상호 작용에 대해 구성될 수 있다. 예를 들어, 시각적 전용 디지털 컴포넌트 객체는 키보드, 마우스 또는 터치 스크린 입력을 통한 상호 작용을 위해 구성될 수 있다(예를 들어, 디지털 컴포넌트 객체에 내장된 하이퍼 링크 선택하도록 구성됨). 오디오 전용 디지털 컴포넌트 객체는 음성 입력을 통한 상호 작용을 위해 구성될 수 있다(예를 들어, 액션을 수행하기 위해 사전 결정된 키워드를 감지하도록 구성됨).

[0060] 그러나, 컨텐츠 선택기 컴포넌트(108)에 의해 선택된 디지털 컴포넌트 객체의 포맷은 컴퓨팅 디바이스(140)와 호환되지 않을 수 있거나 컴퓨팅 디바이스(140)에 의한 프리젠테이션을 위해 최적화되지 않을 수 있다. 일부 경우, 데이터 처리 시스템(102)은 컴퓨팅 디바이스(140)에 대해 호환되거나 최적화된 디지털 컴포넌트 객체들을 필터링할 수 있다. 포멧 또는 모달리티에 기초한 필터링은 컴퓨팅 디바이스(140)에 대해 호환되거나 최적화된 포멧을 갖는 디지털 컴포넌트 객체를 선택하게 할 수 있다. 포맷 또는 모달리티에 기초한 필터링은 디지털 컴포넌트 객체들의 포맷이 컴퓨팅 디바이스(140)에 대해 호환되거나 최적화되지 않을 수 있기 때문에 관련성이 있거나 컨텐츠 선택 기준과 더 잘 매칭할 수 있는 디지털 컴포넌트 객체의 선택을 방지할 수 있다. 예를 들어, 컴퓨팅 디바이스(140)가 디스플레이 디바이스(146)가 없는 스마트 스피커인 경우, 시각적 컨텐츠 아이템들은 필터링되거나 선택이 금지되어 선택 아이템이 오디오 전용 컨텐츠 아이템으로 제한될 수 있다. 시각적 컨텐츠 아이템이 선택된 오디오 전용 컨텐츠 아이템보다 컨텐츠 선택 기준과 더 잘 매칭하는 키워드들이 포함된 경우, 데이터 처리 시스템(102)은 상위(top) 매칭 컨텐츠 아이템을 제공하지 못할 수 있다. 상위 매칭 컨텐츠 아이템을 제공하지 않음으로써, 데이터 처리 시스템(102)은 프레젠테이션을 위해 관련없는 컨텐츠 아이템을 제공함으로써 컴퓨팅 디바이스(140)의 컴퓨팅 리소스 소비, 네트워크 대역폭 또는 배터리 전력을 낭비할 수 있다.

[0061] 따라서, 본 기술 솔루션은 포맷에 관계없이 컨텐츠 선택 기준에 기초하여 가장 매칭되는 디지털 컴포넌트 객체를 선택한 다음, 디지털 컴포넌트 객체를 컴퓨팅 디바이스(140)와 최적화되거나 호환되는 포맷으로 변환할 수 있다. 포멧이나 모달리티로 인해 선택되는 컨텐츠 아이템을 제거하거나 방지하지 않음으로써 본 기술 솔루션은 관련성과 다른 컨텐츠 선택 기준에 기초하여 가장 높은 순위의 컨텐츠 아이템을 선택한 다음 실시간으로 컨텐츠 아이템을 원하는 포멧으로 변환할 수 있다.

[0062] 이를 위해 데이터 처리 시스템(102)은 3P 디지털 컨텐츠 제공자 디바이스(160)에 의해 제공되는 디지털 컴포넌트 객체의 오리지널 포맷과 상이한 포맷으로 디지털 컴포넌트 객체를 생성하도록 설계, 구성 및 동작하는 컨텐츠 변환 컴포넌트(110)를 포함할 수 있다. 3P 디지털 컨텐츠 제공자 디바이스(160)는 제1 포맷으로 오리지널 컨텐츠 아이템을 제공할 수 있고, 데이터 처리 시스템(102)은 오리지널 컨텐츠 아이템에 기초한 제2 포맷으로 제2 컨텐츠 아이템을 생성할 수 있다. 예를 들어, 컨텐츠 선택기 컴포넌트(108)는 요청 또는 컨텐츠 선택 기준에 기초하여, 시각적 출력 포맷을 갖는 디지털 컴포넌트 객체를 선택할 수 있다. 컨텐츠 변환 컴포넌트(110)는 컴퓨팅 디바이스(140)에 디스플레이 디바이스가 없지만 오디오 인터페이스가 있음을 결정할 수 있다. 컨텐츠 변환 컴포넌트(110)는 오디오 전용 포맷을 갖는 새로운 디지털 컴포넌트 객체를 생성할 수 있다. 컨텐츠 변환 컴포넌트(110)는 디지털 컴포넌트 객체와 연관된 메타 데이터를 사용하여 오디오 전용 포맷으로 새로운 디지털 컴포넌트 객체를 생성할 수 있다. 컨텐츠 변환 컴포넌트(110)는 새로운 컨텐츠 아이템에 대한 포멧을 선택하고, 오리지널 컨텐츠 아이템에 기초하여 텍스트를 생성하고, 새로운 컨텐츠 아이템에 대한 음성을 선택하고, 새로운 컨텐츠 아이템에 대한 비-음성 오디오 큐를 생성하고, 새로운 컨텐츠 아이템과 상호 작용하는데 사용되는 액션을

생성하고, 그런 다음 컴퓨팅 디바이스(140)에 제공할 새로운 컨텐츠 아이템을 생성할 수 있다.

[0063] 컨텐츠 변환 컴포넌트(110)는 컨텐츠 선택기 컴포넌트(108)에 의해 선택된 디지털 컴포넌트 객체에 기초하여 디지털 컴포넌트 객체를 생성하기 위한 포맷을 선택하도록 설계, 구성 및 동작하는 포맷 선택기(112)를 포함할 수 있다. 포맷 선택기(112)는 다양한 기술 또는 요인(factors)를 사용하여 디지털 컴포넌트 객체를 변환할 포맷을 결정할 수 있다. 요인는 예를 들어 컴퓨팅 디바이스(140)의 유형, 컴퓨팅 디바이스(140)의 가용 인터페이스, 컴퓨팅 디바이스(140)의 잔여 배터리 전력, 컴퓨팅 디바이스(140)의 위치, 컴퓨팅 디바이스(140)와 관련된 이동 수단(예를 들어, 운전, 기차, 비행기, 도보, 달리기, 자전거 타기 또는 정지), 컴퓨팅 디바이스(140)의 포어그라운드에서 실행되는 애플리케이션(152)의 유형, 컴퓨팅 디바이스(140)의 상태 또는 다른 요인을 포함할 수 있다. 일부 경우, 요인는 요리, 업무, 휴식과 같은 사용자 활동을 포함할 수 있다. 데이터 처리 시스템(102)은 하루 중 시간 또는 최근 검색 활동(예를 들어, 레시피를 찾는 것)에 기초하여 사용자가 요리중임을 결정할 수 있다. 데이터 처리 시스템(102)은 하루 중 시간, 요일 및 위치(예를 들어, 사업장)에 기초하여 사용자가 일하고 있는지 여부를 결정할 수 있다. 데이터 처리 시스템(102)은 컴퓨팅 디바이스(140)상의 하루 중 시간, 위치 및 활동(예를 들어, 영화 스트리밍)에 기초하여 사용자가 휴식을 취하고 있는지 여부를 결정할 수 있다.

[0064] 포맷 선택기(112)는 컴퓨팅 디바이스(140)의 유형에 기초하여 상기 선택된 디지털 컴포넌트 객체를 변환할 포맷을 선택할 수 있다. 포맷 선택기(112)는 컨텐츠 요청과 함께 컴퓨팅 디바이스(140)의 유형에 관한 정보를 수신할 수 있다. 예를 들어, 컴퓨팅 디바이스(140)에 의해 제공되는 컨텐츠 요청은 컴퓨팅 디바이스(140)의 유형을 나타낼 수 있다. 컨텐츠 요청이 수신되지 않은 경우, 포맷 선택기(112)는 컴퓨팅 디바이스(140)와 관련된 계정 정보 또는 프로파일 정보, 또는 컴퓨팅 디바이스(140)에서 실행되는 애플리케이션(152)으로부터 수신된 정보에 기초하여 컴퓨팅 디바이스(140)의 유형을 결정할 수 있다. 일부 경우, 포맷 선택기(112)는 컴퓨팅 디바이스(140)의 유형에 관한 정보를 컴퓨팅 디바이스(140)에 질의할 수 있다. 컴퓨팅 디바이스(140)의 예시적인 유형은 랩탑, 태블릿, 스마트 시계, 웨어러블 디바이스, 스마트 폰, 스마트 스피커, 스마트 텔레비전, 또는 사물 인터넷 디바이스(예를 들어, 스마트 기기 또는 스마트 조명)를 포함할 수 있다. 디바이스의 유형은 컴퓨팅 디바이스(140)에서 이용 가능한 인터페이스의 유형(예를 들어, 시각적 출력 인터페이스, 오디오 출력 인터페이스, 오디오 입력 인터페이스, 터치 입력 인터페이스 또는 키보드 및 마우스 인터페이스)을 나타낼 수 있다. 예를 들어, 컴퓨팅 디바이스(140)의 유형이 스마트 스피커인 경우, 데이터 처리 시스템(102)은 컴퓨팅 디바이스(140)에 대한 주 인터페이스가 오디오 인터페이스이고 컴퓨팅 디바이스(140)에 디스플레이 디바이스가 없다고 결정할 수 있다. 포맷 선택기(112)는 디바이스 유형에 대한 주 인터페이스가 오디오 전용 인터페이스라는 것에 응답하여 오리지널 시각적 디지털 컴포넌트 객체를 오디오 전용 포맷 디지털 컴포넌트 객체로 변환하도록 결정할 수 있다. 다른 예에서, 컴퓨팅 디바이스의 유형이 스마트 텔레비전인 경우, 데이터 처리 시스템(102)은 주 인터페이스가 시청각 인터페이스라고 결정할 수 있다. 포맷 선택기(112)는 주 인터페이스가 시청각 인터페이스라는 결정에 응답하여, 오리지널 시각 전용 디지털 컴포넌트 객체를 시청각 디지털 컴포넌트 객체로 변환하도록 결정할 수 있다. 디지털 컴포넌트 객체를 컴퓨팅 디바이스(140)의 유형에 대한 주 포맷으로 변환함으로써, 데이터 처리 시스템(102)은 컴퓨팅 디바이스(140)상의 디지털 컴포넌트 객체의 렌더링 또는 프리젠테이션을 최적화할 수 있다. 렌더링 또는 프리젠테이션을 최적화하는 것은 컴퓨팅 디바이스(140)의 주 사용자 인터페이스 또는 사용자 인터페이스들의 주 조합을 사용하여 디지털 컴포넌트 객체를 출력하는 것을 지칭할 수 있다.

[0065] 포맷 선택기(112)는 컴퓨팅 디바이스(140)의 이용 가능한 인터페이스들에 기초하여 변환을 위한 포맷을 선택할 수 있다. 컴퓨팅 디바이스(140)의 유형은 컴퓨팅 디바이스(140)가 포함하는 인터페이스들의 유형을 나타낼 수 있다. 그러나, 하나 이상의 인터페이스가 이용 가능하지 않을 수 있으며, 이 경우 포맷 선택기(112)는 이용 가능한 인터페이스를 식별한 다음 디지털 컴포넌트 객체를 이용 가능한 인터페이스에 대응하는 포맷으로 변환할 수 있다. 예를 들어, 컴퓨팅 디바이스(140)는 디스플레이 디바이스(146)를 비활성화하거나 끄는 동안 디지털 음악 스트리밍과 같은 오디오를 출력할 수 있으며, 이는 전력 소비를 줄일 수 있다. 이 예에서, 포맷 선택기(112)는 디스플레이 디바이스(146)가 꺼졌기 때문에 시각적 인터페이스가 현재 이용 가능하지 않다고 결정할 수 있지만, 오디오 출력 인터페이스가 현재 오디오를 출력하고 있기 때문에 이용 가능하다고 결정할 수 있다. 다른 예에서, 포맷 선택기(112)는 오디오가 음소거된 경우 오디오 인터페이스가 사용 가능하지 않다고 결정할 수 있고, 디스플레이 디바이스(146)가 시각적 출력을 능동적으로 제공하고 있는 경우 시각적 출력 인터페이스가 사용 가능하다고 결정할 수 있다. 따라서, 오디오 인터페이스를 사용할 수 없는 경우, 포맷 선택기(112)는 디지털 컴포넌트 객체에 대한 시각적 출력 포맷을 선택할 수 있고; 시각적 인터페이스를 사용할 수 없는 경우, 포맷 선택기(112)는 디지털 컴포넌트 객체에 대한 오디오 출력 포맷을 선택할 수 있다. 시각적 또는 오디오 출력 인터페이스를 사용할 수 없는 경우, 포맷 선택기(112)는 컨텐츠 변환을 종료하고 디지털 컴포넌트 객체의 전달을 차단

하여 낭비되는 컴퓨팅 리소스 사용 및 네트워크 대역폭 사용을 방지할 수 있다.

[0066] 포맷 선택기(112)는 컴퓨팅 디바이스(140)의 잔여 배터리 전력에 기초하여 출력 인터페이스를 결정할 수 있다. 예를 들어, 잔여 배터리 전력이 임계값(예를 들어, 10%, 15%, 20%, 25% 또는 기타 임계값) 미만인 경우, 포맷 선택기(112)는 디스플레이 디바이스에 비해 더 적은 에너지를 소비할 수 있는 오디오 출력과 같이 렌더링할 에너지의 최소량을 이용하는 포맷을 선택하도록 결정할 수 있다.

[0067] 포맷 선택기(112)는 컴퓨팅 디바이스(140)의 이동 수단에 기초하여 디지털 컴포넌트 객체에 대한 포맷을 선택할 수 있다. 이동 수단의 예로는 운전, 기차, 비행기, 도보, 달리기, 자전거 타기 또는 정지(예를 들어, 움직이지 않거나 이동하지 않음)가 포함될 수 있다. 포맷 선택기(112)는 이동 수단이 운전, 달리기 또는 자전거 타기인 경우 오디오 전용 출력 포멧을 선택하여, 사용자가 이러한 이동 수단에서 시각적 출력을 인식하지 못할 수 있으므로 사용자의 주의를 분산시키고 하고 낭비되는 에너지 소비를 방지할 수 있다. 이동 수단이 도보, 정지, 대중교통 수단 또는 비행기인 경우, 포멧 선택기(112)는 시각적 출력이 사용자의 주의를 분산시키지 않고 사용자가 시각적 출력을 인식할 수 있기 때문에 시각적 출력 또는 시청각 출력 포멧을 선택할 수 있다.

[0068] 포맷 선택기(112)는 컴퓨팅 디바이스(140)의 포 그라운드에서 실행되는 애플리케이션(152)의 유형에 기초하여 디지털 컴포넌트 객체의 포맷을 선택할 수 있다. 애플리케이션(152)의 주 출력 인터페이스가 디지털 음악 스트리밍 애플리케이션(152)과 같이 오디오 전용인 경우, 포맷 선택기(112)는 예를 들어 오디오 출력 포맷을 선택할 수 있다. 애플리케이션(152)의 주 출력 인터페이스가 시각적 전용 포맷인 경우, 포맷 선택기(112)는 시각적 전용 출력을 선택할 수 있다. 애플리케이션(152)의 주 출력 인터페이스가 디지털 비디오 스트리밍 애플리케이션(152)에서와 같이 시청각 출력의 조합인 경우, 포맷 선택기(112)는 디지털 컴포넌트 객체에 대한 시청각 출력 포맷을 선택할 수 있다.

[0069] 포맷 선택기(112)는 디지털 어시스턴트 디바이스인 컴퓨팅 디바이스(140)의 유형에 기초하여 또는 디지털 어시스턴트 애플리케이션을 포함하는 애플리케이션(152)을 실행하는 컴퓨팅 디바이스(140)에 기초하여 디지털 컴포넌트 객체에 대한 포맷을 선택할 수 있다. 디지털 어시스턴트 애플리케이션은 가상 어시스턴트를 지칭하거나 포함할 수 있다. 디지털 어시스턴트 애플리케이션은 커맨드 또는 질문에 기초하여 태스크 또는 서비스를 수행할 수 있는 소프트웨어 에이전트를 포함할 수 있다. 디지털 어시스턴트 애플리케이션(152)은 (예를 들어, 사용자에 의해 발화된) 자연어 입력을 수신하고 처리한 다음 태스크, 액션을 수행하거나 입력에 대한 응답을 제공하도록 구성될 수 있다. 포맷 선택기(112)는 애플리케이션 또는 디지털 어시스턴트인또는 컴퓨팅 디바이스(140)의 유형에 응답하여, 디지털 어시스턴트 애플리케이션(152)의 주 인터페이스가 음성 기반(또는 오디오 기반) 인터페이스일 수 있기 때문에 디지털 컴포넌트 객체에 대해 오디오 전용 포맷을 선택하도록 결정할 수 있다.

[0070] 컨텐츠 변환 컴포넌트(110)는 디지털 컴포넌트 객체에 기초하여 텍스트를 생성하도록 설계, 구성 및 동작하는 텍스트 생성기(114)를 포함할 수 있다. 예를 들어, 포맷 선택기(112)가 시각적 디지털 컴포넌트 객체를 오디오 전용 디지털 컴포넌트 객체로 변환하기로 결정하는 것에 응답하여, 텍스트 생성기(114)는 디지털 컴포넌트 객체를 처리하여 오디오를 통해 출력될 수 있는 텍스트를 생성할 수 있다. 시각적 컴포넌트 객체에 기초하여 텍스트를 생성하기 위해, 텍스트 생성기(114)는 시각적 컴포넌트의 텍스트를 파싱하고, 시각적 디지털 컴포넌트 객체를 처리하기 위해 이미지 처리 기술을 적용하거나 광학 문자 인식 기술을 적용할 수 있다. 텍스트 생성기(114)는 시각적 컴포넌트 객체와 관련된 메타 데이터를 획득하고 그 메타 데이터를 파싱하거나 처리하여 텍스트를 생성할 수 있다. 메타 데이터에는 예를 들어 제품 사양 또는 제품 설명이 포함될 수 있다. 따라서, 텍스트 생성기(114)는 예를 들어, 시각적 디지털 컴포넌트, 디지털 컴포넌트 객체에 내장된 하이퍼 링크 또는 URL, 제품에 대한 링크 또는 제품 설명에 텍스트의 튜플을 사용할 수 있다.

[0071] 텍스트 생성기(114)는 시각적 디지털 컴포넌트, 메타 데이터 또는 대응하는 링크로부터 획득된 텍스트의 튜플을 자연어 생성 모델에 입력하여 텍스트를 생성할 수 있다. 텍스트 생성기(114)는 자연어 생성 엔진 또는 컴포넌트를 포함하거나, 이들로 구성되거나, 이들에 액세스할 수 있다. 자연어 생성은 구조화된 데이터를 자연어로 변환하는 프로세스를 지칭할 수 있다. 자연어 생성을 사용하는 텍스트 생성기(114)는 텍스트-음성 변환 시스템에 의해 판독될 수 있는 텍스트를 생성할 수 있다.

[0072] 자연어 생성 기술로 구성된 텍스트 생성기(114)는 내용 결정(예를 들어, 텍스트에서 언급할 정보 결정); 문서 구조화(예를 들어, 전달할 정보의 전체 구성); 집계(예를 들어, 가독성과 자연스러움을 개선하기 위해 유사한 문장 병합); 어휘 선택(예를 들어, 개념에 단어 삽입); 참조 표현 생성(예를 들어, 객체 및 영역을 식별하는 참조 표현 생성) 및 실현(예를 들어, 구문(syntax), 형태(morphology) 및 맞춤법의 규칙에 따라 정확할 수 있는 실제 텍스트 생성)과 같은 다수의 단계 (stage)에서 텍스트를 생성할 수 있다.

[0073] 텍스트 생성기(114)는 인간이 작성한 텍스트의 큰 코퍼스(말뭉치)에서와 같이 기계 학습을 사용하여 통계 모델을 트레이닝함으로써 자연어 생성을 수행할 수 있다. 기계 학습은 예를 들어 모델을 트레이닝시키기 위해 3P 디지털 컨텐츠 제공자 디바이스(160)에 의해 제공되는 디지털 컴포넌트 객체에 대응하는 사람 작성의 텍스트를 처리할 수 있다.

[0074] 텍스트 생성기(114)는 시퀀스 대 시퀀스 모델을 사용하여 텍스트를 생성할 수 있다. 시퀀스 대 시퀀스 모델에는 인코더와 디코더의 두 부분이 포함될 수 있다. 인코더와 디코더는 하나의 네트워크로 결합된 2개의 상이한 신경망 모델일 수 있다. 신경망은 장단기 메모리("LSTM") 블록들과 같은 순환(반복) 신경망("RNN")일 수 있다. 네트워크의 인코더 부분은 입력 시퀀스(예를 들어, 시각적 디지털 컴포넌트의 텍스트에 해당하는 튜플, 디지털 컴포넌트 객체에 내장된 하이퍼 링크 또는 URL, 제품에 대한 링크 또는 제품 설명)를 이해한 다음 입력의 더 작은 차원 표현을 생성하도록 구성될 수 있다. 인코더는 이 표현을 출력을 나타내는 시퀀스를 생성하도록 구성될 수 있는 디코더 네트워크로 포워딩할 수 있다. 디코더는 디코더의 반복의 각 시간 단계에서 하나씩 단어를 생성할 수 있다.

[0075] 텍스트 생성기(114)는 생성적 적대 네트워크("GAN")를 사용하여 텍스트를 생성할 수 있다. GAN은 생성된 텍스트가 "진짜"인지 "가짜"인지 검출하도록 구성된 공격자(예를 들어, 판별자 네트워크)를 도입함으로써 실제 샘플을 생성하도록 트레이닝된 생성기 네트워크를 지칭할 수 있다. 예를 들어, 판별자는 생성기를 조정하는데 사용되는 동적으로 업데이트되는 평가 메트릭일 수 있다. GAN에 있는 생성기와 판별기는 평형점(equilibrium point)에 도달할 때까지 지속적으로 개선될 수 있다.

[0076] 따라서, 텍스트 생성기(114)는 자연어 생성 기술을 사용하여 시각적 디지털 컴포넌트 객체에 기초하여 텍스트를 생성할 수 있다. 컨텐츠 변환 컴포넌트(110)는 텍스트를 음성으로 출력하기 위해 사용할 디지털 음성 인쇄를 선택할 수 있다. 컨텐츠 변환 컴포넌트(110)는 텍스트를 렌더링하기 위해 디지털 음성을 선택하도록 설계, 구성 및 동작하는 음성 선택기(116)를 포함할 수 있다. 컨텐츠 변환 컴포넌트(110)는 디지털 컴포넌트 객체의 컨텍스트 또는 생성된 텍스트에 기초하여 디지털 음성을 선택할 수 있다. 음성 선택기(116)는 디지털 컴포넌트 객체의 유형 또는 텍스트의 컨텍스트와 매칭하는 디지털 음성을 선택할 수 있다. 예를 들어, 음성 선택기(116)는 액션 영화에 대한 광고와 비교하여 베개 광고에 대해 다른 디지털 음성을 선택할 수 있다.

[0077] 디지털 음성 인쇄를 선택하여 텍스트 생성기(114)에 의해 생성된 텍스트에 대한 오디오 트랙을 생성하기 위해, 음성 선택기(116)는 이력 데이터를 사용하여 기계 학습 엔진(122)에 의해 트레이닝된 음성 모델(126)을 사용할 수 있다. 음성 모델(126)을 트레이닝시키는데 사용된 이력 데이터는 예를 들어 컴퓨팅 디바이스(140) 또는 다른 매체를 통한 프리젠테이션을 위해 3P 디지털 컨텐츠 제공자에 의해 생성된 오디오 디지털 컴포넌트 객체를 포함할 수 있다. 이력 데이터에는 3P 디지털 컨텐츠 제공자들이 생성한 각 오디오 디지털 컴포넌트 객체와 관련된 메타 데이터 또는 컨텍스트 정보가 포함될 수 있다. 메타 데이터 또는 컨텍스트 정보는 예를 들어 주제, 개념, 키워드, 지리적 위치, 브랜드 이름, 버티컬 카테고리, 제품 카테고리, 서비스 카테고리, 또는 오디오 디지털 컴포넌트 객체의 양태을 기술하는 기타 정보를 포함할 수 있다. 이력 데이터에는 오디오 디지털 컴포넌트 객체와 관련된 성능 정보가 포함될 수 있다. 성능 정보는 오디오 디지털 컴포넌트 객체에 대한 선택 또는 변환과 같이 최종 사용자가 오디오 디지털 컴포넌트와 상호 작용했는지 여부를 나타낼 수 있다.

[0078] 예를 들어, 이력 디지털 컴포넌트는 텔레비전, 라디오 또는 컴퓨팅 디바이스에서 방송하기 위해 3P 컨텐츠 제공자에 의해 생성된 라디오 광고(예를 들어, 방송 라디오 또는 디지털 스트리밍 라디오 방송국), 텔레비전 광고(예를 들어, 방송 또는 케이블 텔레비전 또는 디지털 스트리밍 텔레비전 채널)를 포함할 수 있다. 이러한 이력 디지털 컴포넌트는 TV에 제시되는 경우 오디오 및 시각적 컴포넌트를 포함할 수 있다. 텔레비전 광고와 관련된 메타 데이터 또는 컨텍스트 정보는 제품 유형(예를 들어, 자동차, 여행, 가전 제품 또는 식품), 서비스 유형(예를 들어, 세금 서비스, 전화 서비스, 인터넷 서비스, 식당, 배달 서비스 또는 가정 서비스), 제품 또는 서비스에 관한 설명 정보, 제품 또는 서비스를 제공하는 회사 또는 엔티티에 관한 정보, 광고가 제공될 지리적 위치(예를 들어, 주, 지리적 지역, 도시 또는 우편 번호) 또는 기타 키워드를 포함할 수 있다. 따라서, 이력 데이터는 오디오(또는 오디오-비디오) 3P 디지털 컴포넌트 객체에 대응하는 오디오 트랙 및 그 오디오 트랙과 관련된 메타 데이터를 포함할 수 있다. 이력 3P 디지털 컴포넌트 객체들을 저장하는 예시적인 데이터 구조가 [표 1]에 도시되어 있다.

표 1

[0079] 예시적인 이력 데이터

고유 ID	오디오 파일	제품/서비스	버티컬	위치	브랜드	설명
1	오디오_1. mp3	제품	자동차	미국	회사_A	고급 스포츠 카
2	오디오_2,mp3	서비스	은행	뉴잉글랜드	회사_B	저금리 신용 카드 제공

[0080] [표 1]은 텍스트 생성기(114)에 의해 생성된 텍스트를 렌더링하는데 사용할 디지털 음성을 선택하기 위해 음성 선택기(116)에 의해 사용되는 음성 모델(126)을 트레이닝시키기 위해 기계 학습 엔진(122)에 의해 사용되는 이력 데이터의 예시적인 예를 제공한다. 표 1에 도시된 바와같이 각 이력 3P 디지털 컴포넌트 객체는 오디오 트랙(예를 들어, 오디오_1.mp3 및 오디오_2.mp3), 광고가 제품 또는 서비스 용인지 여부의 표시, 버티컬 마켓의 표시(예를 들어, 자동차 또는 은행), 광고가 제공되는 위치의 표시(예를 들어, 미국 또는 뉴 잉글랜드와 같은 지리적 지역), 광고의 브랜드 또는 제공자(예를 들어, 회사_A 또는 회사_B), 디지털 컴포넌트 객체와 관련된 추가 설명 또는 키워드(예를 들어, 고급 스포츠 카 또는 저금리 신용 카드 제공)를 포함할 수 있다. 오디오 파일은 .wav, .mp3, .aac 또는 임의의 다른 오디오 포멧을 포함하는 임의의 포멧일 수 있다. 일부 경우, 이력 디지털 컴포넌트 객체에는 오디오 및 비디오가 모두 포함될 수 있으며, 이 경우 오디오 파일 포멧은 .mp4, .mov, .wmv, .flv 또는 다른 파일 포멧과 같은 오디오 및 시각적 파일 포멧을 지칭할 수 있다.

[0081] 데이터 처리 시스템(102)은 이력 디지털 컴포넌트 데이터를 전처리하여, 음성 모델(126)을 트레이닝하도록 데이터를 처리하기 위해 기계 학습 엔진(122)에 적합한 포맷으로 데이터를 배치할 수 있다. 예를 들어, 음성 선택기(116) 또는 기계 학습 엔진(122)은 오디오 처리 기술 또는 파싱 기술로 구성되어 데이터의 특징을 식별하기 위해 이력 디지털 컴포넌트 데이터를 처리할 수 있다. 특징은 예를 들어 오디오 파일의 오디오 특성, 제품/서비스, 버티컬 카테고리, 설명의 키워드 또는 기타 정보가 포함될 수 있다. 오디오 특성의 예로는 음성의 성별, 음성의 연령대, 피치(음높이), 주파수, 진폭 또는 볼륨, 억양, 방언, 언어, 악센트, 단어가 발화되는 속도 또는 기타 특성이 포함될 수 있다.

[0082] 기계 학습 엔진(122)은 이력 데이터를 분석하여 음성 모델(126)을 트레이닝시키기 위해 임의의 기계 학습 또는 통계 기술을 사용할 수 있다. 기계 학습 엔진(122)은 예측 또는 결정을 내리기 위해 샘플 데이터 또는 트레이닝 데이터(예를 들어, 이력 디지털 컴포넌트 객체들)에 기초하여 모델을 구축할 수 있는 학습 기술 또는 기능으로 구성될 수 있다. 기계 학습 엔진(122)은 지도 또는 비지도 학습 기술, 반-지도 학습, 강화 학습, 자기 학습, 특징 학습, 희소 사전 학습 또는 관련 규칙으로 구성될 수 있다. 기계 학습을 수행하기 위해, 기계 학습 엔진(122)은 트레이닝 데이터에 대해 트레이닝된 음성 모델(126)을 생성할 수 있다. 이 모델은 예를 들어 인공 신경망, 의사 결정 트리, 지원 벡터 머신, 회귀 분석, 베이지안 네트워크 또는 유전 알고리즘을 기반으로 할 수 있다.

[0083] 음성 선택기(116)는 시각적 디지털 컴포넌트에 기초하여 텍스트 생성기(114)에 의해 생성된 텍스트를 수신할 때, 시각적 디지털 컴포넌트와 연관된 메타 데이터와 함께 텍스트를 사용하여 기계 학습 엔진(122)을 통해 트레이닝된 음성 모델(126)을 사용하여 디지털 음성 인쇄를 선택할 수 있다. 예를 들어, 데이터 처리 시스템(102)은 음성 모델(126)을 통해, 디지털 컴포넌트 객체의 컨텍스트에 기초하여 디지털 음성을 선택할 수 있다. 컨텍스트는 디지털 컴포넌트 객체와 관련된 텍스트, 메타 데이터 또는 기타 정보를 포함하거나 지칭하거나 포함할 수 있다. 컨텍스트는 컴퓨팅 디바이스(140)와 관련된 정보를 지칭하거나 포함할 수 있다. 일부 경우, 음성 선택기(116)는 이동 수단, 위치, 선호도, 성능 정보 또는 컴퓨팅 디바이스(140)와 관련된 다른 정보와 같은 컴퓨팅 디바이스(140)의 컨텍스트에 기초하여 디지털 음성을 선택할 수 있다.

[0084] 음성 선택기(116)는 디지털 컴포넌트 객체의 컨텍스트를 음성 모델(126)에 입력하여 음성 특성 벡터를 생성한 다음, 텍스트를 렌더링하기 위해 디지털 음성을 선택할 수 있다. 텍스트 및 메타 데이터는 제품, 서비스, 버티컬 카테고리, 키워드 또는 음성 모델(126)에 입력될 수 있는 다른 정보에 관한 정보를 나타낼 수 있다. 음성 모델(126)에 대한 입력은 시각적 디지털 컴포넌트에 기초하여 생성된 텍스트 또는 시각적 디지털 컴포넌트의 텍스트 및 메타 데이터의 조합일 수 있다. 음성 모델(126)의 출력은 텍스트를 렌더링하는데 사용할 디지털 음성 인쇄의 특성을 예측하는 음성 특성 벡터일 수 있다. 출력은 디지털 음성 인쇄의 성별(예를 들어, 남성 또는 여성), 억양(예를 들어, 억양은 텍스트의 각 음절에 대해 변경될 수 있음), 악센트, 발성, 피치, 음량, 말하기 속도, 톤(음조), 색조(texture), 음량 또는 기타 정보를 나타낼 수 있다. 음성 모델(126)의 출력은 베이스, 바리톤, 테너, 알토, 메조-소프라노 및 소프라노와 같은 다른 음성 유형을 포함할 수 있다.

[0085] 음성 선택기(116)는 음성 모델(126)에 의해 출력된 음성 특성 벡터를 데이터 저장소(124)에 저장된 가용 디지털 음성 인쇄와 비교하여 매칭되는 디지털 음성 인쇄 또는 가장 근접하게 매치되는 디지털 음성 인쇄를 식별할 수 있다. 디지털 음성 인쇄는 성별, 악센트, 발성, 피치, 음량, 말하기 속도 또는 기타 정보에 기초하여 분류될 수 있다. 음성 선택기(116)는 음성 모델(126)의 출력을 저장된 또는 가용 디지털 음성 인쇄와 비교하여 텍스트를 렌더링하는데 사용할 가장 근접하게 매칭되는 디지털 음성 인쇄를 선택할 수 있다. 음성 선택기(116)는 매칭되는 음성 인쇄를 선택하기 위해 특성에 가중치를 부여할 수 있다. 예를 들어, 성별과 같은 특성은 악센트와 같은 특성보다 가중치가 더 클 수 있다. 말하기 속도와 같은 특성은 악센트와 같은 특성보다 더 많은 가중치를 부여할 수 있다. 일부 경우, 음성 선택기(116)는 가장 특성에 매칭하는 디지털 음성을 선택할 수 있다. 음성 선택기(116)는 음성 모델(126)의 출력에 기초하여 디지털 음성 인쇄를 선택하기 위해 임의의 매칭 기술을 사용할 수 있다.

[0086] 선택된 디지털 음성 인쇄는 그 디지털 음성 인쇄를 식별하는 고유 식별자를 포함할 수 있다. 디지털 음성 인쇄는 컨텐츠 변환 컴포넌트(110)가 텍스트-음성 변환을 수행하는데 사용할 수 있는 정보를 포함할 수 있다. 디지털 음성 인쇄에는 텍스트-음성 변환 엔진에 대한 명령이 포함될 수 있다. 컨텐츠 변환 컴포넌트(110)는 디지털 음성 인쇄에 의해 표시된 음성 특성을 사용하여 텍스트를 렌더링하기 위해 임의의 유형의 텍스트-음성 변환 기술을 사용할 수 있다. 예를 들어, 컨텐츠 변환 컴포넌트(110)는 디지털 음성 인쇄에 의해 정의된 인간과 유사한 음성을 사용하여 텍스트를 렌더링하기 위해 신경망 기술을 사용할 수 있다.

[0087] 컨텐츠 변환 컴포넌트(110)는 텍스트 음성 변환 기술을 사용하여 그리고 선택된 디지털 음성 인쇄에 기초하여 렌더링된 텍스트의 기본(baseline) 오디오 트랙을 생성할 수 있다. 예를 들어, 컨텐츠 변환 컴포넌트(110)는 디지털 음성에 의해 렌더링된 텍스트로 디지털 컴포넌트 객체의 기준 오디오 트랙을 구성하도록 설계, 구성 및 동작하는 오디오 큐 생성기(118)를 포함할 수 있다. 오디오 큐 생성기(118)는 텍스트-음성 변환 엔진을 사용하여 디지털 음성에 따라 텍스트를 렌더링하거나 합성할 수 있다.

[0088] 컨텐츠 변환 컴포넌트(110)는 기준 오디오 트랙에 비-음성 오디오 큐를 추가하기로 결정할 수 있다. 오디오 큐 생성기(118)는 기준 오디오 트랙에 추가하기 위해 비-음성 오디오 큐를 생성하도록 설계, 구성 및 동작할 수 있다. 비-음성 큐는 음향 효과를 지칭하거나 포함할 수 있다. 비-음성에는 예를 들어 해양 파도 소리, 바람 소리, 나뭇잎이 바스락 거리는 소리, 자동차 엔진, 운전, 비행기 이륙, 군중 응원, 스포츠, 액션 영화 효과(예를 들어, 고속 자동차 추격, 헬리콥터 등). 달리기, 자전거 타기 또는 기타 음향 효과가 포함될 수 있다. 따라서 비-음성 오디오 큐는 단어나 숫자(예를 들어, 발화된 단어)가 포함된 음성 없는 소리 또는 음향 효과를 지칭할 수 있다.

[0089] 오디오 큐 생성기(118)는 디지털 컴포넌트 객체의 텍스트 또는 메타 데이터에 기초하여 하나 이상의 비-음성 오디오 큐를 생성할 수 있다. 오디오 큐 생성기(118)는 텍스트에 대해 선택된 디지털 음성에 기초하여 하나 이상의 비-음성 오디오 큐를 생성할 수 있다. 오디오 큐 생성기(118)는 컴퓨팅 디바이스(140)의 컨텍스트(예를 들어, 이동 수단, 컴퓨팅 디바이스의 유형, 컴퓨팅 디바이스(140)의 포어그라운드에서 실행되는 애플리케이션(152) 유형, 애플리케이션(152)에 제시되는 컨텐츠, 또는 컴퓨팅 디바이스로부터 수신된 요청)140)에 기초하여 비-음성 오디오 큐를 선택할 수 있다.

[0090] 오디오 큐 생성기(118)는 오디오 큐 모델(134) 또는 오디오 큐 데이터 저장소를 사용하여 기준 오디오 트랙에 추가할 하나 이상의 비-음성 오디오 큐를 선택할 수 있다. 오디오 큐(134) 데이터 저장소는 음향 효과가 있다는 인디케이터와 같은 메타 데이터로 태깅된 음향 효과를 포함할 수 있다. 예를 들어 해양 파도의 음향 효과는 "해양 파도"와 같은 음향 효과 설명으로 태킹될 수 있다. 오디오 큐(134) 데이터 저장소는 다수 유형의 해양 파도 음향 효과를 포함할 수 있고 각 해양 파도를 구별하는 대응하는 태그 또는 설명을 포함할 수 있다.

[0091] 일부 경우, 오디오 큐는 디지털 음성의 특성에 맞게 구성되거나 최적화될 수 있다. 특정 음성 큐는 특정 음성 특성 벡터를 갖는 디지털 음성에 대한 배경 음향 효과로 프리젠테이션하기 위해 최적화될 수 있다. 최적화된 음향 효과는 텍스트를 가리거나 주의를 분산시키지 않고 디지털 음성으로 텍스트를 렌더링과 함께 렌더링될 수 있는 음향 효과를 지칭하며, 텍스트를 사용자가 이해할 수 있도록 하여 향상된 사용자 인터페이스를 제공하는 것을 목표로 한다. 예를 들어, 디지털 음성과 동일한 주파수 및 진폭을 가진 음향 효과는 음향 효과에 대한 디지털 음성을 인식, 식별 또는 구별하는 것을 어렵게 만들 수 있으며, 이로 인해 사용자 경험이 저하되고 비효율적인 출력을 제공하여 컴퓨팅 리소스 사용이 낭비될 수 있다.

[0092] 오디오 큐 생성기(118)는 시각적 디지털 컴포넌트에 대한 이미지 인식을 수행하여 디지털 컴포넌트 객체에서 시각적 객체를 식별할 수 있다. 오디오 큐 생성기(118)는 시각적 디지털 컴포넌트와 관련된 임의의 텍스트를 무시

할 수 있고, 이미지 인식을 수행하여 객체를 검출할 수 있다. 오디오 큐 생성기(118)는 임의의 이미지 처리 기술 또는 물체 검출 기술을 사용할 수 있다. 오디오 큐 생성기(118)는 기계 학습 엔진(122)에 의해 트레이닝된 모델을 사용할 수 있고, 객체의 설명이 태깅된 객체의 이미지를 포함하는 트레이닝 데이터 세트에 기초할 수 있다. 기계 학습 엔진(122)은 오디오 큐 생성기(118)가 모델에 입력되는 새로운 이미지에서 객체를 검출하기 위해 모델을 사용할 수 있도록 트레이닝 데이터로 모델을 트레이닝시킬 수 있다. 오디오 큐 생성기(118)는 트레이닝된 모델을 사용하여 시각적 디지털 컴포넌트 객체에서 이미지를 검출할 수 있다. 따라서, 오디오 큐 생성기(118)는 시각적 디지털 컴포넌트 객체에 대한 이미지 인식을 수행하여 디지털 컴포넌트 객체에서 시각적 객체를 식별할 수 있다. 오디오 큐 생성기(118)는 데이터 저장소(124)에 저장된 비-음성 오디오 큐(134)로부터, 시각적 객체에 대응하는 비-음성 오디오 큐를 선택할 수 있다.

[0093] 오디오 큐 생성기(118)는 또한 디지털 컴포넌트 객체에 대응하는 랜딩 웹 페이지에 대한 링크와 같은 시각적 디지털 컴포넌트에 내장된 링크에 액세스함으로써 객체들을 식별할 수 있다. 오디오 큐 생성기(118)는 웹 페이지를 파싱하여 추가 컨텍스트 정보, 키워드 또는 메타 데이터뿐만 아니라 시각적 객체를 식별할 수 있다. 오디오 큐 생성기(118)는 객체, 컨텍스트 정보, 키워드 또는 메타 데이터에 기초하여 오디오 큐를 선택할 수 있다. 예를 들어, 랜딩 웹 페이지의 텍스트에는 '해변, 휴가, 크루즈'라는 키워드가 포함될 수 있다. 오디오 큐 생성기(118)는 하나 이상의 키워드에 대응하는 오디오 큐를 선택할 수 있다.

[0094] 오디오 큐 생성기(118)가 시각적 객체의 이미지 또는 디지털 컴포넌트에 링크된 웹 페이지 또는 시각적 디지털 컴포넌트의 메타 데이터와 연관된 다른 키워드 또는 컨텍스트 정보에 기초하여 다수의 후보 오디오 큐를 식별하는 경우, 오디오 큐 생성기(118)는 하나 이상의 비-음성 오디오 큐를 선택할 수 있다. 오디오 큐 생성기(118)는 얼마나 많은 비-음성 오디오 큐를 선택할지를 결정하기 위해 정책을 사용할 수 있다. 정책은 식별된 모든 오디오 큐를 선택하거나, 사전 결정된 수의 오디오 큐를 무작위로 선택하거나, 오디오 트랙 전체에서 상이한 오디오 큐를 번갈아 가며, 하나 이상의 오디오 큐를 오버레이 또는 혼합하거나, 사전 결정된 수의 최고 순위 오디오 큐를 선택하는 것이다. .

[0095] 예를 들어, 오디오 큐 생성기(118)는 시각적 디지털 컴포넌트에서 가장 두드러진 객체를 식별하고 가장 두드러진 객체에 대응하는 오디오 큐를 선택할 수 있다. 객체의 현저성(Prominence, 두드러짐)은 시각적 디지털 컴포넌트에서 객체의 크기를 지칭할 수 있다(예를 들어, 시각적 디지털 컴포넌트에서 가장 큰 객체가 가장 두드러진 객체일 수 있음). 두드러짐은 배경이 아닌 이미지의 전경에 있는 객체에 기초할 수 있다. 오디오 큐 생성기(118)는 텍스트 생성기(114)에 의해 생성된 텍스트와 가장 관련이 있는 시각적 디지털 컴포넌트 내의 객체를 식별할 수 있다. 관련성은 객체 및 텍스트의 설명에 기초하여 결정될 수 있다. 예를 들어, 생성된 텍스트가 객체의 이름이나 객체 설명에 키워드를 포함하는 경우, 객체는 텍스트와 관련이 있는 것으로 결정될 수 있다. 오디오 큐 생성기(118)는 텍스트와 가장 관련이 있는 키워드 또는 개념을 결정할 수 있고, 오디오 큐에 대해 이러한 키워드를 선택할 수 있다.

[0096] 오디오 큐 생성기(118)는 현저성, 관련성 또는 현저성과 관련성 둘 모두에 기초하여 객체의 순위 지정할 수 있다. 오디오 큐 생성기(118)는 순위에 기초하여 하나 이상의 오디오 큐를 선택하기로 결정할 수 있다. 예를 들어, 오디오 큐 생성기(118)는 가장 높은 순위의 오디오 큐, 상위 2개의 가장 높은 순위의 큐, 상위 3 개의 가장 높은 순위의 큐, 또는 일부 다른 수의 오디오 큐를 선택할 수 있다.

[0097] 일부 경우, 오디오 큐 생성기(118)는 디지털 음성에 의해 렌더링된 텍스트로부터 주의를 분산시키는 오디오 큐에 기초하여 오디오 큐가 기준 오디오 트랙에 추가되는 것을 필터링, 제거 또는 방지할 수 있다. 예를 들어, 오디오 큐 생성기(118)는 이미지 인식 기술을 통해 디지털 컴포넌트 객체내의 다수의 시각적 객체를 식별할 수 있다. 오디오 큐 생성기(118)는 메타 데이터(예를 들어, 3P 디지털 컨텐츠 제공자 디바이스(160)에 의해 제공된 메타 데이터 또는 디지털 컴포넌트 객체의 링크에 대응하는 랜딩 페이지와 관련된 키워드) 및 텍스트에 기초하여 다수의 비-음성 오디오 큐를 식별할 수 있다. 오디오 큐 생성기(118)는 시각적 객체 각각과 메타 데이터 간의 매칭 레벨을 수준을 나타내는 시각적 객체 각각에 대한 매칭 스코어를 결정할 수 있다. 오디오 큐 생성기(118)는 임의의 매칭 기술을 사용하여 관련성, 브로드 매치, 구문 매치 또는 정확한 매치와 같은 매칭 스코어를 결정할 수 있다. 오디오 큐 생성기(118)는 컨텐츠 선택기 컴포넌트(108)와 유사한 기술을 사용하여 매칭 스코어를 결정할 수 있다. 오디오 큐 생성기(118)는 매칭 스코어에 기초하여 비-음성 오디오 큐를 순위 지정할 수 있다. 일부 경우, 오디오 큐 생성기(118)는 하나 이상의 가장 높은 순위의 오디오 큐를 선택할 수 있다.

[0098] 일부 경우, 오디오 큐 생성기(118)는 디지털 음성을 사용하여 합성된 텍스트를 간섭, 융합, 방해 또는 부정적 영향을 미치지 않는 가장 높은 순위의 하나 이상의 오디오 큐를 선택할 수 있다. 예를 들어, 오디오 큐 생성기

(118)는 텍스트를 렌더링하기 위해 선택된 디지털 음성과 비-음성 오디오 큐 각각 사이의 오디오 간섭 레벨을 결정할 수 있다. 간섭 레벨은 예를 들어 진폭, 주파수, 피치 또는 타이밍과 같은 하나 이상의 요인을 사용하여 결정될 수 있다. 예시적인 예에서, 합성된 텍스트와 동일한 주파수 및 진폭을 갖는 음향 효과는 최종 사용자가 렌더링된 텍스트를 정확하게 인식하는 것을 방해할 수 있는 높은 레벨의 간섭을 유발할 수 있다. 다른 예에서, 큰 충돌 소리가 텍스트로부터 주의를 분산시킬 수 있다. 그러나, 음성 오디오 트랙 전체에 걸쳐 낮은 진폭의 부드러운 바람 소리는 텍스트로부터 주의를 분산시키지 않을 수 있다.

[0099] 간섭 레벨을 결정하기 위해, 오디오 큐 생성기(118)는 합성된 텍스트에 대한 비-음성 오디오 큐에 의해 야기되는 간섭의 양을 결정할 수 있다. 이 양은 텍스트의 백분율, 연속 지속 시간 또는 텍스트에 대한 데시벨 레벨이 될 수 있다. 일부 경우, 간섭은 신호 대 잡음비 또는 텍스트 신호 대 비-음성 오디오 큐 신호의 비에 기초할 수 있다. 간섭 레벨은 등급(예를 들어, 낮음, 중간 또는 높음) 또는 숫자 값(예를 들어, 1에서 10까지의 척도(scale) 또는 척도의 한쪽 끝은 간섭을 나타내지 않고 척도의 반대쪽 끝은 전체 간섭을 나타내는 임의의 다른 척도)를 사용하여 표시될 수 있다. 전체 간섭은 합성된 텍스트를 완전히 취소할 수 있는 파괴적인 간섭을 의미할 수 있다.

[0100] 일부 경우, 오디오 큐 생성기(118)는 합성된 텍스트 및 비-음성 오디오 큐에 대응하는 오디오 파형을 결합한 다음 결합된 신호를 처리하여 최종 사용자가 합성된 텍스트를 인식할 수 있는지 여부를 결정함으로써 간섭 레벨을 결정할 수 있다. 오디오 큐 생성기(118)는 인터페이스(104) 또는 자연어 프로세서 컴포넌트(106)와 유사한 오디오 처리 기술을 사용하여, 합성된 텍스트가 데이터 처리 시스템(102) 자체에 의해 정확하게 인식될 수 있는지 여부를 확인, 검증 또는 결정할 수 있는데, 이는 최종 사용자가 오디오 트랙을 능동적으로 인식할 수 있는지 여부를 나타낼 수 있다.

[0101] 사전 결정된 임계값(예를 들어, 낮은 간섭, 5, 6, 7 미만의 간섭 스코어 또는 기타 메트릭) 미만의 간섭 레벨을 갖는 오디오 큐를 식별할 때, 오디오 큐 생성기(118)는 임계값 미만의 오디오 간섭 레벨을 갖는 가장 높은 순위의 비-음성 오디오 큐를 선택할 수 있다.

[0102] 오디오 큐 생성기(118)는 선택된 비-음성 오디오 큐를 기준 오디오 트랙과 결합하여 디지털 컴포넌트 객체의 오디오 트랙을 생성할 수 있다. 오디오 트랙은 시각적 디지털 컴포넌트 객체에 기초한 오디오 전용 디지털 컴포넌트 객체에 해당할 수 있다. 오디오 큐 생성기(118)는 임의의 오디오 믹싱(혼합) 기술을 사용하여 비-음성 큐를 기준 오디오 트랙과 결합할 수 있다. 예를 들어, 오디오 큐 생성기(118)는 기준 오디오 트랙 위에 비-음성 오디오 큐를 오버레이하고, 비-음성 오디오 큐를 배경 오디오로 추가하고, 합성된 텍스트 전후 또는 합성 텍스트 사이에 비-음성 오디오 큐를 삽입할 수 있다. 데이터 처리 시스템(102)은 오디오 트랙을 생성하기 위해 2개 이상의 입력 오디오 신호의 특성을 결합, 변경, 등화 또는 기타 변경하도록 구성된 디지털 믹싱 컴포넌트를 포함할 수 있다. 데이터 처리 시스템(102)은 비-음성 오디오 큐 및 기준 오디오 트랙을 수신하고 2개의 신호를 합산하여 결합된 오디오 트랙을 생성할 수 있다. 데이터 처리 시스템(102)은 디지털 믹싱 프로세스를 사용하여 입력 오디오 신호를 결합하므로 원하지 않는 노이즈 또는 왜곡의 도입을 피할 수 있다.

[0103] 따라서, 일단 기본 오디오 포맷이 합성되면, 데이터 처리 시스템(102)은 메타 데이터로부터 결정될 수 있는 비-음성 오디오 큐 또는 백킹(backing) 트랙을 삽입하는 제2 생성 단계를 수행할 수 있다. 예를 들어, 시각적 디지털 컴포넌트가 야자수가 있는 해변 휴양지처럼 보이는 경우, 데이터 처리 시스템(102)은 파도와 바람에 움직이는 나무 잎의 오디오를 합성할 수 있고, 합성된 오디오를 텍스트-음성 변환 기준 오디오 트랙에 추가하여 오디오 트랙을 생성할 수 있다.

[0104] 일부 경우, 액션 생성기(136)는 사전 결정되거나 고정된 오디오를 기준 오디오 트랙 또는 비-음성 큐로 생성된 오디오 트랙에 추가할 수 있다. 예를 들어, 데이터 처리 시스템(102)은 휴리스틱 또는 규칙 기반 기술을 사용하여, "웹 사이트에서 이것에 대해 자세히 알아보기"와 같은 문구를 추가하기로 결정할 수 있다. 이것은 사용자가 디지털 컴포넌트 객체와 독립적으로 액션을 수행하도록 프롬프트할 수 있다. 데이터 처리 시스템(102)은 이력 수행 정보, 오디오 트랙의 시간 길이에 기초하거나 구성 또는 설정(예를 들어, 데이터 처리 시스템(102)의 관리자에 의해 설정된 디폴트 설정 또는 3P 디지털 컨텐츠 제공자 디바이스(160)에 의해 제공될 수 있는 설정)에 기초하여 고정(된) 오디오를 자동으로 추가하기로 결정할 수 있다. 일부 경우, 고정 오디오에는 "웹 사이트에서 이것에 대해 자세히 알아 보시겠습니까?"와 같은 음성 입력 프롬프트가 포함될 수 있다. 데이터 처리 시스템(102)은 프롬프트에 대응하는 디지털 액션을 자동으로 수행하기 위해, 이 경우 "예"와 같은 프롬프트에 대한 응답에서 트리거 단어를 검출하도록 오디오 디지털 컴포넌트 객체를 구성할 수 있다.

[0105] 데이터 처리 시스템(102)은 생성된 오디오 트랙을 컴퓨팅 디바이스(140)에 제공하여 컴퓨팅 디바이스(140)가 스

피커(예를 들어, 변환기(144))를 통해 오디오 트랙을 출력하거나 제시하게 할 수 있다. 일부 경우, 데이터 처리 시스템(102)은 오디오 트랙에 실행 가능한 커맨드를 추가할 수 있다. 컨텐츠 변환 컴포넌트(110)는 트리거 단어를 오디오 트랙에 추가하도록 설계, 구성 및 동작하는 액션 생성기(136)를 포함할 수 있다. 트리거 단어는 오디오 트랙과의 상호 작용을 용이하게할 수 있다. 전 처리기(142)는 트리거 단어를 청취한 후 트리거 단어에 응답하는 액션을 수행할 수 있다. 트리거 단어는 오디오 트랙을 재생하는 동안 및 오디오 트랙 이후의 사전 결정된 시간량(예를 들어, 1초, 2초, 5초, 10초, 15초 또는 기타 적절한 시간 간격))과 같이 사전 결정된 시간 간격 동안 활성 상태로 유지되는 새로운 웨이크 업 또는 핫 워드가 될 수 있다. 컴퓨팅 디바이스(140)의 마이크로폰(예를 들어, 센서(148))에 의해 검출된 입력 오디오 신호에 있는 트리거 단어의 검출에 응답하여, 데이터 처리 시스템(102) 또는 컴퓨팅 디바이스(140)는 트리거 단어에 대응하는 디지털 액션을 수행할 수 있다.

[0106] 예시적인 예에서, 오디오 트랙은 크루즈 티켓을 구매하고 해변이 있는 섬으로 휴가를 가기 위한 광고를 포함할 수 있다. 오디오 트랙은 "크루즈 티켓의 가격을 알고 싶으십니까?"와 같은 프롬프트를 포함할 수 있다. 트리거 단어는 "예" 또는 "가격은 얼마입니까", "비용은 얼마입니까" 또는 사용자가 티켓 가격을 요청하려는 의도를 전달하는 일부 다른 변형일 수 있다. 데이터 처리 시스템(102)은 전 처리기(142)가 사용자에 의해 제공되는 후속 음성 입력에서 트리거 키워드를 검출할 수 있도록 컴퓨팅 디바이스(140) 또는 컴퓨팅 디바이스(140)의 전 처리기(142)에 트리거 단어를 제공할 수 있다. 음성 입력에서 트리거 단어를 검출하는 것에 응답하여, 전 처리기(142)는 음성 입력을 데이터 처리 시스템(102)으로 포워딩할 수 있다. NLP 컴포넌트(106)는 음성 입력을 파싱하고, 사용자에게 제시된 디지털 컴포넌트와 관련된 랜딩 페이지로 사용자를 안내하거나 요청된 정보에 액세스하여 제공하는 것과 같이 음성 입력에 대응하는 액션을 수행할 수 있다.

[0107] 트리거 키워드는 다양한 디지털 액션에 링크될 수 있다. 예시적인 디지털 액션은 정보 제공, 애플리케이션 실행, 내비게이션 애플리케이션 실행, 음악 또는 비디오 재생, 제품 또는 서비스 주문, 기기 제어, 조명 디바이스 제어, 사물 인터넷 지원 디바이스 제어, 식당에서 음식 주문, 예약, 승차 공유 주문, 영화 티켓 예약, 항공권 예약, 스마트 TV 제어 또는 다른 디지털 액션을 포함할 수 있다.

[0108] 액션 생성기(136)는 하나 이상의 기술을 사용하여 디지털 컴포넌트 객체에 대한 액션을 선택할 수 있다. 액션 생성기(136)는 휴리스틱 기법을 사용하여 사전 결정된 액션 세트로부터 액션을 선택할 수 있다. 액션 생성기(136)는 생성된 텍스트를 입력으로 수신하여 예측된 액션을 출력하도록 구성된 액션 모델(128)을 사용할 수 있다.

[0109] 예를 들어, 액션 생성기(136)는 디지털 컴포넌트 객체의 카테고리를 결정할 수 있다. 카테고리는 자동차, 은행, 스포츠, 의류 등과 같은 버티컬 카테고리를 지칭할 수 있다. 액션 생성기(136)는 카테고리에 대해 설정된 하나 이상의 트리거 단어 및 디지털 액션을 검색하기 위해 카테고리를 이용하여 데이터베이스를 조회하거나 질의할 수 있다. 데이터베이스에는 키워드 및 디지털 액션을 트리거하기 위한 카테고리 매핑이 포함될 수 있다. [표 2]는 단어와 액션을 트리거하기 위한 예시적인 카테고리 매핑을 도시한다.

표 2

[0110] 단어와 디지털 액션을 트리거하기 위한 카테고리 매핑의 예시적인 예

카테고리	트리거 단어	디지털 액션
승차 공유	예; 승차(ride); 승차 주문; 가기(go to)	컴퓨팅 디바이스(140)에서 승차 공유 애플리케이션 시작. 사용자를 픽업하도록 승차 공유 주문.
여행	항공편 예약; 항공편 가격 확인; [도시]행은 얼마입니까; [도시]행 다음 항공편은 언제입니까	항공편 예약 옵션 제공; 항공편 예약 애플리케이션 실행; 항공편을 검색 및 결과 제공.
차량 쇼핑	자동차 비용은 얼마입니까; 자동차에 사용할 수 있는 옵션은 무엇입니까; 가장 가까운 대리점은 어디입니까?	컨텐츠 아이템 제공자의 랜딩 페이지에 액세스; 요청된 정보 제공; 가장 가까운 자동차 판매점으로 가는 길을 안내하는 내비게이션 애플리케이션 시작

[0111] [표 2]는 단어와 디지털 액션을 트리거하기 위한 예시적인 카테고리 매핑을 도시한다. [표 2]에 도시된 바와 같이, 카테고리에는 승차 공유, 여행 및 차량 쇼핑이 포함될 수 있다. 액션 생성기(136)는 디지털 컴포넌트 객체에 대해 생성된 텍스트, 디지털 컴포넌트 객체와 관련된 메타 데이터, 또는 디지털 컴포넌트 객체에 내장된 링크들과 관련된 데이터 파싱에 기초하여 상기 선택된 디지털 컴포넌트 객체의 카테고리를 결정할 수 있다. 일부

경우, 3P 디지털 컨텐츠 제공자 디바이스(160)는 메타 데이터와 함께 카테고리 정보를 제공할 수 있다. 일부 경우, 액션 생성기(136)는 의미론적 처리 기술을 사용하여 디지털 컴포넌트 객체와 관련된 정보에 기초하여 카테고리를 결정할 수 있다.

[0112] 액션 생성기(136)는 디지털 컴포넌트 객체에 대한 카테고리를 식별하거나 결정할 때, 조회를 수행하거나 매핑을 질의할 수 있다(예를 들어, [표 2]). 액션 생성기(136)는 매핑으로부터, 카테고리와 관련된 하나 이상의 디지털 액션에 대응하는 하나 이상의 트리거 단어를 검색한다. 예를 들어, 액션 생성기(136)가 카테고리를 "승차 공유"로 식별하는 경우, 액션 생성기(136)는 질의 또는 조회에 응답하여 트리거 키워드 "예", "승차", "승차 주문" 또는 "가기"을 검색할 수 있다. 액션 생성기(136)는 '컴퓨팅 디바이스(140)에서 승차 공유 애플리케이션 시작' 또는 '사용자를 픽업하도록 승차 공유 주문'과 같은 디지털 동작을 추가로 식별할 수 있다. 액션 생성기(136)는 모든 트리거 키워드를 검출하고 트리거 키워드의 검출에 응답하여 해당 디지털 액션들을 수행하기 위한 명령으로 오디오 디지털 컴포넌트 객체를 구성할 수 있다.

[0113] 일부 경우, 액션 생성기(136)는 전부는 아니지만 검색된 트리거 키워드들 및 디지털 액션들 중 하나 이상을 디지털 컴포넌트 객체에 추가하기로 결정할 수 있다. 예를 들어, 액션 생성기(136)는 트리거 키워드들의 이력 수행에 기초하여 트레이닝된 디지털 액션 모델(128)을 사용하여, 디지털 컴포넌트 객체의 컨텍스트 및 클라이언트 디바이스의 유형에 기초하여 트리거 단어들을 순위 지정할 수 있다. 액션 생성기(136)는 오디오 트랙에 추가할 최고 순위의 트리거 키워드를 선택할 수 있다.

[0114] 액션 모델(128)은 트레이닝 데이터를 사용하여 기계 학습 엔진(122)에 의해 트레이닝될 수 있다. 트레이닝 데이터에는 트리거 키워드와 관련된 이력 수행 정보가 포함될 수 있다. 이력 수행 정보는 각 트리거 키워드에 대해, 트리거 키워드가 상호 작용을 유발했는지 여부(예를 들어, 컴퓨팅 디바이스(140)가 오디오 트랙의 프리젠테이션 이후에 수신된 음성 입력에서 트리거 키워드를 검출했는지 여부), 디지털 컴포넌트 객체와 관련된 컨텍스트 정보(예를 들어, 상호 작용 흐름의 범주, 키워드, 개념 또는 상태) 및 상호 작용의 컨텍스트 정보를 포함할 수 있다. 컴퓨팅 디바이스(140)의 컨텍스트는 컴퓨팅 디바이스(140)의 유형(예를 들어, 모바일 디바이스, 랩탑 디바이스, 스마트 폰 또는 스마트 스피커), 컴퓨팅 디바이스(140)의 사용 가능한 인터페이스, 이동 수단(예를 들어, 도보, 운전, 정지, 자전거 타기 등), 또는 컴퓨팅 디바이스(140)의 위치를 포함할 수 있다. 예를 들어, 이동 수단이 달리기, 자전거 타기 또는 운전인 경우, 데이터 처리 시스템(102)은 시각적 또는 터치 입력이 필요하지 않은 상호 작용 유형을 선택할 수 있으며 사용자 경험을 개선하기 위해 오디오 출력을 생성할 수 있다.

[0115] 기계 학습 엔진(122)은 액션 모델(128)이 디지털 컴포넌트 객체 및 컴퓨팅 디바이스(140)의 컨텍스트에 기초하여 상호 작용을 유발할 가능성이 가장 높은 트리거 키워드를 예측할 수 있도록 이 트레이닝 데이터에 기초하여 액션 모델(128)을 트레이닝할 수 있다. 따라서, 액션 생성기(136)는 상호 작용을 유발할 가능성이 가장 높은 트리거 키워드를 제공하기 위해 실시간 프로세스에서 디지털 컴포넌트 객체에 추가할 액션을 맞춤화하거나 조정(tailor)할 수 있다. 트리거 키워드의 수를 액션 모델(128)에 의해 결정된 상호 작용을 유발할 가능성이 가장 높은 키워드로 제한함으로써, 액션 생성기(136)는 컴퓨팅 디바이스(140)의 사용자가 의도치 않게 원하지 않는 액션 수행할 가능성을 감소시키면서, 전 처리기(142) 또는 NLP 컴포넌트(106)가 트리거를 정확하고 신뢰성있게 검출할 가능성을 개선할 수 있다. 또한, 트리거 단어의 수를 제한함으로써, 액션 생성기(136)는 네트워크(105)를 통해 전송되는 커맨드 또는 데이터 패킷의 수를 감소시킬뿐만 아니라 전 처리기(142)가 처리하는 트리거 단어의 수를 감소시킴으로써 네트워크 대역폭 통신 및 컴퓨팅 리소스 이용을 감소시킬 수 있다.

[0116] 데이터 처리 시스템(102)은 컴퓨팅 디바이스(140)의 스피커를 통해 출력하기 위해 컴퓨팅 디바이스(140)에 디지털 컴포넌트 객체의 오디오 트랙을 제공할 수 있다. 일부 경우, 데이터 처리 시스템(102)은 디지털 컴포넌트 객체의 오디오 트랙에 대한 삽입 지점을 결정할 수 있다. 삽입 지점은 컴퓨팅 디바이스(140)의 오디오 출력과 관련된 시간 지점을 지칭할 수 있다. 오디오 출력은 디지털 스트리밍 음악 또는 컴퓨팅 디바이스(140)에서 실행되는 애플리케이션(152)을 통해 제공되는 다른 오디오(또는 시청각) 출력에 대응할 수 있다. 데이터 처리 시스템(102)은 컴퓨팅 디바이스(140)에 의해 출력되는 메인 오디오 컨텐츠의 난독화 또는 왜곡을 방지하는 동시에 사용자 경험 및 상기 생성된 오디오 트랙이 최종 사용자에 의해 인식되고 궁극적으로 상호 작용을 수신할 가능성을 개선하기 위해 삽입 시간을 결정할 수 있다.

[0117] 데이터 처리 시스템(102)은 오디오 트랙에 대한 삽입 지점을 식별하도록 설계, 구성 및 동작하는 컨텐츠 삽입 컴포넌트(120)를 포함할 수 있다. 컨텐츠 삽입 컴포넌트(120)는 컴퓨팅 디바이스(140)에 의해 출력된 디지털 미디어 스트림에서 오디오 트랙에 대한 삽입 지점을 식별할 수 있다. 컨텐츠 삽입 컴포넌트(120)는 삽입 모델(130)을 사용하여 삽입 지점을 식별할 수 있다. 기계 학습 엔진(122)은 이력 수행 데이터를 사용하여 삽입 모델

(130)을 트레이닝할 수 있다. 이력 수행 데이터는 트레이닝 데이터를 포함하거나 트레이닝 데이터로 지지ㅊ될 수 있다. 삽입 모델(130)을 트레이닝하는데 사용되는 이력 수행 데이터는 디지털 미디어 스트림에 삽입된 오디오 트랙에 대한 이력 삽입 지점들에 관한 데이터를 포함할 수 있다. 데이터는 오디오 트랙이 삽입된 시기, 오디오 트랙에 관한 컨텍스트 정보, 디지털 미디어 스트림에 관한 컨텍스트 정보, 사용자가 오디오 트랙과 상호 작용했는지 여부, 사용자가 오디오 트랙과 상호 작용한 방법(예를 들어, 사용자가 취한 액션 또는 상호 작용이 긍정적이었는지 부정적이었는지 여부), 또는 컴퓨팅 디바이스(140)에 관한 컨텍스트 정보(예를 들어, 컴퓨팅 디바이스의 유형, 컴퓨팅 디바이스의 가용 인터페이스 또는 컴퓨팅 디바이스의 위치)를 포함할 수 있다.

[0118] 기계 학습 엔진(122)은 이 트레이닝 데이터를 사용하여 삽입 모델(130)을 트레이닝할 수 있다. 기계 학습 엔진(122)은 삽입 모델(130)이 디지털 컨텐츠 스트림(예를 들어, 음악, 뉴스, 팟 캐스트, 비디오 또는 기타 미디어 스트리밍)에서 시각적 디지털 컴포넌트 객체에 기초하여 생성된 오디오 트랙을 삽입할 시기를 예측하는데 사용될 수 있도록 임의의 기술을 사용하여 삽입 모델(130)을 트레이닝할 수 있다.

[0119] 컨텐츠 삽입 컴포넌트(120)는 이력 수행 데이터를 사용하여 트레이닝된 삽입 모델(130)에 기초하여 삽입 지점을 식별할 수 있다. 삽입 지점은 예를 들어 현재 스트리밍 미디어 세그먼트 이후 및 다음 세그먼트 시작 전일 수 있다. 각 세그먼트는 다른 노래에 해당할 수 있다. 팟 캐스트와 같은 다른 예에서, 컨텐츠 삽입 컴포넌트(120)는 삽입 모델(130)을 사용하여 세그먼트 동안 오디오 트랙을 삽입할 것을 결정할 수 있다. 컨텐츠 삽입 컴포넌트(120)는 세그먼트가 시작된 후 및 세그먼트가 끝나기 전에 오디오 트랙을 삽입할 수 있다. 예를 들어, 세그먼트는 30분의 지속 시간을 가질 수 있고, 컨텐츠 삽입 컴포넌트(120)는 삽입 모델(130)을 사용하여 세그먼트 재생 15분 후에 오디오 트랙을 삽입할 것을 결정할 수 있다.

[0120] 컨텐츠 삽입 컴포넌트(120)는 현재 컨텍스트(예를 들어, 생성된 오디오 트랙, 스트리밍 미디어 및 컴퓨팅 디바이스(140)의 컨텍스트)에 기초하여 커스텀 삽입 지점을 결정할 수 있다. 컨텐츠 삽입 컴포넌트(120)는 실시간으로 커스텀 삽입 지점을 결정할 수 있다. 컨텐츠 삽입 컴포넌트(120)는 제1 및 제2 컴퓨팅 디바이스가 상이한 유형의 컴퓨팅 디바이스(예를 들어, 랩탑 대 스마트 스피커)인 경우 제2 컴퓨팅 디바이스(140)와 비교하여 제1 컴퓨팅 디바이스(140)에 대해 다른 삽입 지점을 결정할 수 있다. 컨텐츠 삽입 컴포넌트(120)는 상이한 컴퓨팅 디바이스와 연관된 이동 수단(예를 들어, 도보 대 운전 대 정지)에 기초하여 제1 컴퓨팅 디바이스(140)와 제2 컴퓨팅 디바이스(140)에 대해 상이한 삽입 지점을 결정할 수 있다.

[0121] 컨텐츠 삽입 컴포넌트(120)는 디지털 미디어 스트림에 있는 키워드, 용어 또는 개념에 근접한 오디오 트랙을 삽입하도록 결정할 수 있다. 컨텐츠 삽입 컴포넌트(120)는 디지털 미디어 스트림을 모니터링하여 오디오 트랙과 관련된 디지털 미디어 스트림에서 트리거 단어를 검출한 다음 디지털 미디어 스트림에서 검출된 트리거 단어에 후속하거나 그에 응답하여 오디오 전용 디지털 컴포넌트 객체를 삽입하도록 결정할 수 있다.

[0122] 컨텐츠 삽입 컴포넌트(120)는 3P 전자 리소스 서버(162)로부터 디지털 미디어 스트림의 세그먼트의 사본을 획득할 수 있다. 컨텐츠 삽입 컴포넌트(120)는 디지털 미디어 스트림의 세그먼트를 분석하여 세그먼트 내의 모든 토큰(예를 들어, 키워드, 토픽 또는 개념) 및 문장을 식별할 수 있다. 컨텐츠 삽입 컴포넌트(120)는 토큰 또는 문장이 디지털 컴포넌트 객체에 얼마나 관련이 있는지를 결정하기 위해 각 토큰 및 문장에 대한 관련성 스코어를 결정할 수 있다. 컨텐츠 삽입 컴포넌트(120)는 가장 높은 관련성 스코어를 갖는 토큰을 선택한 다음, 선택된 토큰에 인접한 삽입을 위해 (예를 들어, 토큰이 제시되기 전 또는 후에) 오디오 전용 디지털 컴포넌트 객체를 제공할 수 있다.

[0123] 일부 경우, 컨텐츠 삽입 컴포넌트(120)는 디지털 미디어의 세그먼트에서 모든 토큰을 식별하고, 오디오 트랙이 각 토큰에 인접하게 삽입되는 몬테카를로 시뮬레이션을 실행할 수 있다. 컨텐츠 삽입 컴포넌트(120)는 어떤 삽입 지점이 가장 잘 들리는지 결정하기 위해 변형(variations)을 신경망 엔진에 입력할 수 있다. 신경망은 기계 학습 기술을 사용하여 디지털 미디어 스트림에 삽입된 인간 평가 오디오 트랙을 포함하는 트레이닝 데이터에 기초하여 트레이닝될 수 있다. 예를 들어, 컨텐츠 삽입 컴포넌트(120)는 삽입 모델(130)을 사용하여 삽입 지점을 결정할 수 있다. 트레이닝 데이터는 삽입 지점에서 오디오 트랙을 사용하여 디지털 미디어 스트림의 등급을 매기는 인간 평가자(human raters)를 포함할 수 있다. 등급은 좋음 또는 나쁨과 같은 이진법이거나 척도의 스코어일 수 있다(예를 들어, 0에서 10은 최고의 사운드 트랙을 나타내고 0은 최악의 사운드 트랙을 나타낸다).

[0124] 일부 경우, 컨텐츠 삽입 컴포넌트(120)는 휴리스틱 기술을 사용하여 상기 생성된 오디오 트랙에 대한 삽입 지점을 결정할 수 있다. 휴리스틱 기법은 디지털 미디어 스트림의 유형에 기초하여 상이할 수 있다. 디지털 미디어 스트림의 컨텐츠가 노래인 경우, 휴리스틱 규칙은 노래 재생이 완료된 후 상기 생성된 오디오 트랙을 삽입하는 것일 수 있다. 디지털 미디어 스트림의 컨텐츠가 팟 캐스트인 경우, 휴리스틱 규칙은 관련 토큰이 포함된 문장

뒤에 오디오 트랙을 삽입하는 것일 수 있다.

[0125] 삽입 지점을 선택하면, 데이터 처리 시스템(102)은 컴퓨팅 디바이스(140)가 디지털 미디어 스트림의 삽입 지점에서 오디오 트랙을 렌더링하도록 하는 명령을 컴퓨팅 디바이스(140)에 제공할 수 있다.

[0126] 도 2는 구현에 따른 오디오 트랙을 생성하는 방법의 예시이다. 방법(200)은 예를 들어 데이터 처리 시스템, 인터페이스, 컨텐츠 선택기 컴포넌트, 자연어 프로세서 컴포넌트, 컨텐츠 변환 컴포넌트 또는 컴퓨팅 디바이스를 포함하여 도 1 또는 도 3에 도시된 하나 이상의 시스템, 컴포넌트 또는 모듈에 의해 수행될 수 있다. 결정 블록(202)에서, 데이터 처리 시스템은 입력 신호가 수신되었는지 여부를 결정할 수 있다. 입력 신호는 데이터 처리 시스템에서 떨어진 컴퓨팅 디바이스에 의해 검출된 음성 입력에 해당할 수 있다. 입력 신호는 컴퓨팅 디바이스의 마이크로폰에 의해 검출된 음성 입력과 같은 오디오 입력 신호를 전달하는 데이터 패킷을 포함할 수 있다. 데이터 처리 시스템은 데이터 처리 시스템의 인터페이스를 통해 입력 신호를 수신할 수 있다. 데이터 처리 시스템은 네트워크를 통해 컴퓨팅 디바이스로부터 입력 신호를 수신할 수 있다.

[0127] 결정 블록(202)에서, 데이터 처리 시스템이 입력 신호가 수신되었다고 결정하는 경우, 데이터 처리 시스템은 ACT(204)로 진행하여 입력 신호를 파싱하고 요청을 검출할 수 있다. 데이터 처리 시스템은 자연어 처리 기술을 사용하여 입력 신호를 파싱하고, 입력 신호에서 하나 이상의 키워드, 용어, 개념, 구문 또는 기타 정보를 검출할 수 있다.

[0128] 데이터 처리 시스템은 컨텐츠 선택 여부를 결정하기 위해 결정 블록(206)으로 진행할 수 있다. 컨텐츠를 선택하는 것은 제3자 디지털 컴포넌트 공급자에 의해 제공된 디지털 컴포넌트 객체를 선택하기 위해 실시간 컨텐츠 선택 프로세스를 수행하는 것을 지칭할 수 있다. 컨텐츠를 선택하는 것은 제3자 디지털 컴포넌트 제공자에 의해 제공되는 컨텐츠 선택 기준을 사용하여 실시간 온라인 경매를 수행하는 것을 지칭할 수 있다.

[0129] 데이터 처리 시스템은 결정 블록에서(206)에서, 데이터 처리 시스템이 ACT(204)에서 입력 신호에 있는 컨텐츠 요청을 검출하는 경우 컨텐츠를 선택하기로 결정할 수 있다. 데이터 처리 시스템이 결정 블록(202)에서, 입력 신호가 수신되지 않았다고 결정하는 경우, 데이터 처리 시스템은 또한 결정 블록(206)에서 컨텐츠를 선택하기로 결정할 수 있다. 예를 들어, 데이터 처리 시스템은 온라인 컨텐츠 선택 프로세스를 수행하고, 컴퓨팅 디바이스로부터 디지털 컴포넌트 객체에 대한 명시적인 요청을 수신하지 않고 디지털 컴포넌트 객체를 컴퓨팅 디바이스로 푸시하기로 사전에 결정할 수 있다. 데이터 처리 시스템은 컴퓨팅 디바이스에 의해 출력된 디지털 음악 스트림에서 프레젠테이션 기회(예를 들어, 미디어 세그먼트 또는 노래 사이)를 식별하고 이 기회에 디지털 컴포넌트 객체를 제공하기로 자동으로 결정할 수 있다. 따라서, 일부 경우 데이터 처리 시스템은 컨텐츠 요청을 수신한 다음 컨텐츠 선택 프로세스를 수행할 수 있는 반면, 다른 경우 데이터 처리 시스템은 컨텐츠를 수신하지 않고 컨텐츠 선택 프로세스를 수행하기로 사전에 결정할 수 있다. 데이터 처리 시스템이 컨텐츠 요청을 수신한 경우에도, 그 요청은 메인 컨텐츠(예를 들어, 입력 오디오 신호의 질의에 응답하는 검색 결과)에 대한 것일 수 있으며, 데이터 처리 시스템은 온라인 경매를 수행하여 상기 요청에 응답할 수 있지만 입력 질의에 직접 응답하는 유기적 검색 결과와는 다른 보조 컨텐츠를 선택할 수 있다.

[0130] 결정 블록(206)에서, 데이터 처리 시스템이 3P 디지털 컴포넌트 제공자의 디지털 컴포넌트 객체를 선택하기 위해 컨텐츠 선택 프로세스를 수행하지 않기로 결정한 경우, 데이터 처리 시스템은 결정 블록(202)으로 돌아가서 입력 신호가 수신되었는지 여부를 결정할 수 있다. 그러나, 데이터 처리 시스템이 결정 블록(206)에서 컨텐츠를 선택하기로 결정한 경우, 데이터 처리 시스템은 ACT(208)로 진행하여 컨텐츠를 선택하기 위한 컨텐츠 선택 프로세스를 실행할 수 있다. 데이터 처리 시스템은 (예를 들어, 컨텐츠 선택기 컴포넌트를 통해) 입력 요청, 컴퓨팅 디바이스 또는 디지털 스트리밍 컨텐츠와 관련된 컨텐츠 선택 기준 또는 다른 컨텍스트 정보를 사용하여 디지털 컴포넌트 객체를 선택할 수 있다.

[0131] 데이터 처리 시스템은 컴퓨팅 디바이스의 디스플레이 디바이스를 통해 디스플레이하도록 구성된 시각 전용 디지털 컴포넌트 객체, 컴퓨팅 디바이스의 스피커를 통해 재생하도록 구성된 오디오 전용 디지털 컴포넌트 객체, 또는 컴퓨팅 디바이스의 디스플레이와 스피커를 통해 출력되도록 구성된 시청각 디지털 컴포넌트와 같은 포맷을 갖는 디지털 컴포넌트 객체들을 선택할 수 있다.

[0132] 결정 블록(210)에서, 데이터 처리 시스템은 선택된 디지털 컴포넌트를 다른 포맷으로 변환할지 여부를 결정할 수 있다. 예를 들어, 선택된 디지털 컴포넌트 객체가 시각 전용 포맷인 경우, 데이터 처리 시스템은 컴퓨팅 디바이스의 디스플레이 디바이스를 통해 프리젠테이션하기 위해 컴퓨팅 디바이스에 시각적 포맷으로 디지털 컴포넌트 객체를 제공할지 또는 스피커와 같은 컴퓨팅 디바이스의 다른 출력 인터페이스를 통해 프리젠테이션하기

위해 디지털 컴포넌트 객체를 다른 포멧으로 변환할지 여부를 결정할 수 있다.

[0133] 데이터 처리 시스템(예를 들어, 포맷 선택기)은 디지털 컴포넌트 객체를 변환할지 여부를 결정할 수 있다. 데이터 처리 시스템은 컴퓨팅 디바이스의 사용 가능한 인터페이스, 컴퓨팅 디바이스의 주 인터페이스, 컴퓨팅 디바이스의 컨텍스트(예를 들어, 이동 수단), 컴퓨팅 디바이스의 유형 또는 기타 요인에 기초하여 결정을 내릴 수 있다. 결정 블록(210)에서, 데이터 처리 시스템이 상기 선택된 디지털 컴포넌트 객체를 다른 포맷으로 변환하지 않기로 결정한 경우, 데이터 처리 시스템은 ACT(212)로 진행하여 상기 선택된 디지털 컴포넌트 객체를 오리지널 포멧으로 컴퓨팅 디바이스에 전송할 수 있다.

[0134] 그러나, 데이터 처리 시스템이 결정 블록(210)에서, 디지털 컴포넌트 객체를 다른 포맷으로 변환하기로 결정한 경우, 데이터 처리 시스템은 ACT(214)로 진행하여 텍스트를 생성할 수 있다. 예를 들어, 오리지널 포멧이 시각 전용 포멧이고 데이터 처리 시스템이 디지털 컴포넌트를 오디오 전용 포멧으로 변환하기로 결정한 경우. 데이터 처리 시스템은 ACT(214)로 진행하여 시각적 디지털 컴포넌트 객체에 대한 텍스트를 생성할 수 있다. 데이터 처리 시스템은 (예를 들어, 텍스트 생성기를 통해) 자연어 생성 기술을 사용하여 시각적 디지털 컴포넌트 객체에 기초하여 텍스트를 생성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체의 시각적 컨텐츠만에 기초하여 텍스트를 생성할 수 있다. 시각적 컨텐츠는 이미지를 지칭할 수 있다. 일부 경우, 데이터 처리 시스템은 시각적 디지털 컴포넌트 객체와 관련된 메타 데이터에 기초하여 텍스트를 생성할 수 있다.

[0135] ACT(216)에서, 데이터 처리 시스템은 텍스트를 합성하거나 렌더링하는데 사용할 디지털 음성을 선택할 수 있다. 데이터 처리 시스템은 선택된 디지털 음성을 사용하여 상기 생성된 텍스트의 텍스트-음성 변환을 수행할 수 있다. 데이터 처리 시스템은 생성된 텍스트, 디지털 컴포넌트 객체와 관련된 컨텍스트 정보(예를 들어, 키워드, 토픽, 개념, 버티컬 카테고리), 메타 데이터, 컴퓨팅 디바이스와 관련된 컨텍스트 정보에 기초하여 디지털 음성을 선택할 수 있다. 데이터 처리 시스템은 기계 학습 기술 및 이력 데이터에 기초하여 트레이닝된 모델을 사용하여 상기 생성된 텍스트를 합성하는데 사용할 디지털 음성을 선택할 수 있다. 예를 들어, 데이터 처리 시스템은 디지털 컴포넌트 객체(예를 들어, 메타 데이터)와 관련된 컨텍스트 정보를 모델에 입력할 수 있으며 모델은 음성 특성 벡터를 출력할 수 있다. 음성 특성 벡터는 성별, 말하기 속도, 억양, 음량 또는 기타 특성을 나타낼 수 있다.

[0136] 데이터 처리 시스템은 음성 특성 벡터와 매칭하는 디지털 음성을 선택할 수 있다. 데이터 처리 시스템은 선택된 디지털 음성을 사용하여 기준 오디오 트랙을 구성할 수 있다. 데이터 처리 시스템은 음성 특성 벡터로 표시되는 기준 오디오 트랙을 구성할 수 있다. 예를 들어, 디지털 음성은 성별과 같은 고정 특성은 물론 말하기 속도, 억양 또는 음량과 같은 동적 특성을 포함할 수 있다. 동적 특성은 음절 단위로 다를 수 있다. 데이터 처리 시스템은 음절 단위로 음성 특성 벡터에 해당하는 고정 특성 및 동적 특성을 사용하여 텍스트를 합성하도록 구성된 텍스트-음성 변환 엔진을 사용할 수 있다.

[0137] ACT(218)에서, 데이터 처리 시스템은 비-음성 오디오 큐를 생성할 수 있다. 데이터 처리 시스템은 ACT(216)에서 생성된 기준 오디오 트랙과 비-음성 오디오 큐를 결합할 수 있다. 비-음성 오디오 큐를 생성하기 위해, (예를 들어, 오디오 큐 생성기를 통해) 데이터 처리 시스템은 시각적 디지털 컴포넌트에서 객체들을 식별할 수 있다. 데이터 처리 시스템은 시각적 컴포넌트만 식별할 수 있다. 데이터 처리 시스템은 시각적 컴포넌트와 텍스트 컴포넌트(예를 들어, 디지털 컴포넌트 객체와 관련된 메타 데이터)를 모두 식별할 수 있다. 객체를 식별할 때 데이터 처리 시스템은 객체를 식별하거나 나타내는 오디오 큐를 식별할 수 있다. 예를 들어, 데이터 처리 시스템이 대양 파도와 야자수를 식별하는 경우, 데이터 처리 시스템은 파도 소리와 나뭇잎을 통과하는 바람 소리를 선택할 수 있다.

[0138] 데이터 처리 시스템은 임의의 오디오 믹싱 기술을 사용하여 상기 선택된 오디오 큐를 기준 오디오 트랙과 결합할 수 있다. 데이터 처리 시스템은 기준 오디오 트랙의 일부 또는 오디오 트랙 전체에 비-음성 오디오 큐를 추가할 수 있다. 데이터 처리 시스템은 기준 오디오 트랙의 음성 텍스트를 왜곡하거나 난독화하지 않는 방식으로 비-음성 오디오 큐를 기본 트랙에 추가하여 사용자 경험을 향상시킬 수 있다. 일부 경우, 데이터 처리 시스템은 결합된 오디오 트랙을 시뮬레이션하고 품질을 테스트할 수 있다. 예를 들어, 데이터 처리 시스템은 결합된 오디오 트랙을 수신하고 결합된 오디오 트랙에서 자연어 처리를 수행하는 것을 시뮬레이션할 수 있다. 데이터 처리 시스템은 데이터 처리 시스템의 텍스트 생성기에 의해 생성된 텍스트와 파싱된 텍스트를 비교함으로써 데이터 처리 시스템의 NLP 컴포넌트가 음성 텍스트를 정확하게 검출할 수 있는지 여부를 확인할 수 있다. 데이터 처리 시스템이 결합된 오디오 트랙의 텍스트를 정확하게 해독할 수 없는 경우, 데이터 처리 시스템은 비-음성 오디오 큐가 음성 텍스트에 부정적인 영향을 미치며 최종 사용자가 음성 텍스트를 정확하게 식별하지 못하게 할 수 있

다고 결정할 수 있다. 따라서, 사용자 경험을 개선하기 위해, 데이터 처리 시스템은 하나 이상의 비-음성 오디오 큐를 제거한 다음 결합된 오디오 트랙을 재생성하고 재 테스트할 수 있다. 데이터 처리 시스템은 데이터 처리 시스템이 음성 텍스트를 정확하게 해석할 수 있을 때까지 비-음성 오디오 큐의 이러한 제거를 수행하고 재생성 및 재 테스트를 수행할 수 있다. 결합된 오디오 트랙의 음성 텍스트가 인식될 수 있다는 결정에 응답하여, 데이터 처리 시스템은 결합된 오디오 트랙을 프레젠테이션용으로 승인할 수 있다.

[0139] 일부 경우, 데이터 처리 시스템은 결합된 오디오 트랙을 프리젠테이션을 위해 컴퓨팅 디바이스로 전송할 수 있다. 일부 경우, 데이터 처리 시스템은 ACT(220)로 진행하여 오디오 트랙에 대한 삽입 지점을 결정할 수 있다. 데이터 처리 시스템은 기계 학습 기술과 이력 데이터를 통해 트레이닝된 삽입 모델을 사용하여, 컴퓨팅 디바이스에 의해 출력되는 디지털 미디어 스트림에 오디오 트랙을 삽입할 위치를 결정할 수 있다. 데이터 처리 시스템은 컴퓨팅 리소스 사용률, 네트워크 대역폭 소비를 줄이고 디지털 미디어 스트림의 레이턴시 또는 지연을 방지하거나 사용자 경험을 향상시키는 삽입 지점을 결정할 수 있다. 예를 들어, 데이터 처리 시스템은 디지털 미디어 스트림의 세그먼트 시작, 도중 또는 이후에 오디오 트랙을 삽입하도록 결정할 수 있다.

[0140] 삽입 지점을 결정하면, 데이터 처리 시스템은 ACT(222)로 진행하여 변환된 컨텐츠(또는 오디오 전용 디지털 컴포넌트 객체)를 컴퓨팅 디바이스로 제공하여 컴퓨팅 디바이스가 변환된 디지털 컴포넌트를 렌더링, 재생 또는 제시하도록 할 수 있다. 일부 경우, 데이터 처리 시스템은 변환된 디지털 컴포넌트를 구성하여 디지털 액션을 호출, 시작 또는 수행할 수 있다. 예를 들어, 데이터 처리 시스템은 사용자의 후속 음성 입력에서 트리거 단어들을 검출하고 그 트리거 단어들에 응답하여 디지털 액션을 수행하도록 컴퓨팅 디바이스 또는 데이터 처리 시스템을 구성하는 명령을 제공할 수 있다.

[0141] 일부 경우, 데이터 처리 시스템은 컨텐츠 요청을 수신하지 못할 수 있다. 예를 들어, 데이터 처리 시스템은 컴퓨팅 디바이스에 의해 렌더링된 디지털 스트리밍 컨텐츠와 관련된 키워드를 사전에 식별할 수 있다. 데이터 처리 시스템은 결정 블록(206)에서, 키워드들에 응답하여 컨텐츠를 선택하도록 결정할 수 있다. 그런 다음 데이터 처리 시스템은 키워드들에 기초하여 시각적 출력 포맷을 가진 디지털 컴포넌트 객체를 선택할 수 있다. 데이터 처리 시스템은 컴퓨팅 디바이스의 유형에 기초하여 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 한 결정에 응답하여 디지털 컴포넌트 객체에 대한 텍스트를 생성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체의 컨텍스트에 기초하여 텍스트를 렌더링하기 위한 디지털 음성을 선택할 수 있다. 데이터 처리 시스템은 디지털 음성에 의해 렌더링된 텍스트로 디지털 컴포넌트 객체의 기준 오디오 트랙을 구성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체에 기초하여 비-음성 오디오 큐를 생성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체의 기본 오디오 형식과 비-음성 오디오 큐를 결합하여 디지털 컴포넌트 객체의 오디오 트랙을 생성할 수 있다. 데이터 처리 시스템은 컴퓨팅 디바이스의 스피커를 통한 출력을 위해 컴퓨팅 디바이스에 디지털 컴포넌트 객체의 오디오 트랙을 제공할 수 있다.

[0142] 도 3은 예시적인 컴퓨터 시스템(300)의 블록도이다. 컴퓨터 시스템 또는 컴퓨팅 디바이스(300)는 시스템(100) 또는 데이터 처리 시스템(102)과 같은 그의 컴포넌트를 포함하거나 구현하는데 사용될 수 있다. 컴퓨팅 시스템(300)은 정보를 전달하기 위한 버스(305) 또는 다른 통신 컴포넌트 및 정보를 처리하기 위해 버스(305)에 결합된 프로세서(310) 또는 처리 회로를 포함한다. 컴퓨팅 시스템(300)은 또한 정보를 처리하기 위해 버스에 결합된 하나 이상의 프로세서(310) 또는 처리 회로를 포함할 수 있다. 컴퓨팅 시스템(300)은 또한 정보를 저장하기 위해 버스(305)에 결합된 랜덤 액세스 메모리(RAM) 또는 다른 동적 저장 디바이스와 같은 메인 메모리(315) 및 프로세서(310)에 의해 실행될 명령을 포함한다. 메인 메모리(315)는 데이터 저장소이거나 이를 포함할 수 있다. 메인 메모리(315)는 또한 프로세서(310)에 의한 명령의 실행 동안 위치 정보, 임시 변수, 또는 다른 중간 정보를 저장하기 위해 사용될 수 있다. 컴퓨팅 시스템(300)은 프로세서(310)에 대한 정적 정보 및 명령을 저장하기 위해 버스(305)에 결합된 판독 전용 메모리(ROM) (320) 또는 다른 정적 저장 디바이스를 더 포함할 수 있다. 솔리드 스테이트 디바이스, 자기 디스크 또는 광 디스크와 같은 저장 디바이스(325)는 정보 및 명령을 지속적으로 저장하기 위해 버스(305)에 연결될 수 있다. 저장 디바이스(325)는 데이터 저장소를 포함하거나 그의 일부일 수 있다.

[0143] 컴퓨팅 시스템(300)은 사용자에게 정보를 디스플레이하기 위해 버스(305)를 통해 액정 디스플레이 또는 액티브 매트릭스 디스플레이와 같은 디스플레이(335)에 연결될 수 있다. 영숫자 및 기타 키들을 포함하는 키보드와 같은 입력 디바이스(330)는 정보 및 커맨드 선택을 프로세서(310)에 전달하기 위해 버스(305)에 연결될 수 있다. 입력 디바이스(330)는 터치 스크린 디스플레이(335)를 포함할 수 있다. 입력 디바이스(330)는 또한 프로세서(310)에 방향 정보 및 커맨드 선택을 전달하고 디스플레이(335)상의 커서 이동을 제어하기 위한 마우스, 트랙볼

또는 커서 방향 키와 같은 커서 컨트롤을 포함할 수 있다. 디스플레이(335)는 예를 들어 데이터 처리 시스템(102), 클라이언트 컴퓨팅 디바이스(140) 또는 도 1의 다른 컴포넌트의 일부일 수 있다.

[0144] 본 명세서에 설명된 프로세스, 시스템 및 방법은 메인 메모리(315)에 포함된 명령(명령어)의 배열을 실행하는 프로세서(310)에 응답하여 컴퓨팅 시스템(300)에 의해 구현될 수 있다. 이러한 명령은 저장 디바이스(325)와 같은 다른 컴퓨터 판독 가능 매체로부터 메인 메모리(315)로 판독될 수 있다. 메인 메모리(315)에 포함된 명령 배열의 실행은 컴퓨팅 시스템(300)으로 하여금 본 명세서에 설명된 예시적인 프로세스를 수행하게 한다. 다중 처리 배열에서 하나 이상의 프로세서는 또한 메인 메모리(315)에 포함된 명령을 실행하기 위해 사용될 수 있다. 하드-와이어 회로는 본 명세서에 설명된 시스템 및 방법과 함께 소프트웨어 명령 대신 또는 이와 조합하여 사용될 수 있다. 본 명세서에 설명된 시스템 및 방법은 하드웨어 회로와 소프트웨어의 특정 조합으로 제한되지 않는다.

[0145] 예시적인 컴퓨팅 시스템이 도 3에서 설명되었지만, 본 명세서에 기술된 동작들을 포함하는 주제는 다른 유형의 디지털 전자 회로, 또는 본 명세서에 개시된 구조 및 그의 구조적 등가물을 포함하는 컴퓨터 소프트웨어, 펌웨어 또는 하드웨어 또는 이들 중 하나 이상의 조합으로 구현될 수 있다.

[0146] 본 명세서에 설명된 시스템이 사용자 또는 사용자 디바이스에 설치된 애플리케이션에 관한 개인 정보를 수집하거나 개인 정보를 사용하는 상황에서, 사용자에게는 프로그램이나 기능이 사용자 정보(예를 들어, 사용자의 소셜 네트워크, 소셜 액션 또는 활동, 직업, 사용자의 선호도 또는 사용자의 현재 위치에 관한 정보)를 수집하는지 여부를 제어할 수 있는 기회가 제공된다. 추가적으로 또는 대안적으로, 특정 데이터는 개인 정보가 제거되도록 저장 또는 사용되기 전에 하나 이상의 방법으로 처리될 수 있다.

[0147] 본 명세서에 기술된 주제 및 동작들은 디지털 전자 회로, 또는 본 명세서에 개시된 구조 및 그의 구조적 등가물을 포함하는 컴퓨터 소프트웨어, 펌웨어 또는 하드웨어, 또는 이들 중 하나 이상의 조합으로 구현될 수 있다. 본 명세서에 설명된 주제는 하나 이상의 컴퓨터 프로그램, 예를 들어, 데이터 처리 장치에 의해 실행되거나 데이터 처리 장치의 동작을 제어하기 위해 하나 이상의 컴퓨터 저장 매체에 인코딩된 컴퓨터 프로그램 명령의 하나 이상의 회로로 구현될 수 있다. 대안적으로 또는 추가적으로, 프로그램 명령은 인위적으로 생성된 전파 신호, 예를 들어 데이터 처리 장치에 의한 실행을 위해 적절한 수신기 장치로의 전송을 위해 정보를 인코딩하도록 생성된 기계 생성 전기, 광학 또는 전자기 신호에 인코딩될 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독 가능 저장 디바이스, 컴퓨터 판독 가능 저장 기판, 랜덤 또는 직렬 액세스 메모리 어레이 또는 디바이스, 또는 이들 중 하나 이상의 조합일 수 있거나 그에 포함될 수 있다. 컴퓨터 저장 매체는 전파 신호가 아니지만, 컴퓨터 저장 매체는 인위적으로 생성된 전파 신호로 인코딩된 컴퓨터 프로그램 명령의 소스 또는 대상일 수 있다. 컴퓨터 저장 매체는 또한 하나 이상의 개별 컴포넌트 또는 매체(예를 들어, 다중 CD, 디스크 또는 기타 저장 디바이스)일 수 있거나 이에 포함될 수 있다. 본 명세서에 설명된 동작들은 하나 이상의 컴퓨터 판독 가능 저장 디바이스에 저장되거나 다른 소스로부터 수신된 데이터에 대해 데이터 처리 장치에 의해 수행되는 동작으로 구현될 수 있다.

[0148] "데이터 처리 시스템", "컴퓨팅 디바이스", "컴포넌트" 또는 "데이터 처리 장치"라는 용어는 예를 들어 프로그램 가능 프로세서, 컴퓨터, 시스템 온 칩, 또는 전술한 것 중 다수 또는 이들의 조합을 포함하여, 데이터를 처리하기 위한 다양한 장치, 디바이스 및 기계를 포함한다. 이 장치는 특수 목적 논리 회로, 예를 들어 FPGA(필드 프로그래밍 가능 게이트 어레이) 또는 ASIC(애플리케이션 특정 집적 회로)를 포함할 수 있다. 장치는 또한 하드웨어에 추가하여 문제의 컴퓨터 프로그램에 대한 실행 환경을 생성하는 코드, 예를 들어, 프로세서 펌웨어, 프로토콜 스택, 데이터베이스 관리 시스템, 운영 체제, 크로스-플랫폼 런타임 환경, 가상 머신 또는 이들 중 하나 이상의 조합을 구성하는 코드를 포함할 수 있다. 장치 및 실행 환경은 웹 서비스, 분산 컴퓨팅 및 그리드 컴퓨팅 인프라와 같은 다양한 다른 컴퓨팅 모델 인프라를 구현할 수 있다. 자연어 프로세서 컴포넌트(106) 및 다른 데이터 처리 시스템(102) 또는 데이터 처리 시스템(102) 컴포넌트는 하나 이상의 데이터 처리 장치, 시스템, 컴퓨팅 디바이스 또는 프로세서를 포함하거나 공유할 수 있다. 컨텐츠 변환 컴포넌트(110)와 컨텐츠 선택기 컴포넌트(108)는 예를 들어 하나 이상의 데이터 처리 장치, 시스템, 컴퓨팅 디바이스 또는 프로세서를 포함하거나 공유할 수 있다.

[0149] 컴퓨터 프로그램(프로그램, 소프트웨어, 소프트웨어 애플리케이션, 앱, 스크립트 또는 코드라고도 함)은 컴파일 또는 해석 언어, 선언적 또는 절차적 언어를 포함하여 모든 형태의 프로그래밍 언어로 작성될 수 있으며, 독립 실행형 프로그램이나 모듈, 컴포넌트, 서브 루틴, 객체, 또는 컴퓨팅 환경에서 사용하기에 적합한 기타 유닛을 포함하여 모든 형태로 배포될 수 있다. 컴퓨터 프로그램은 파일 시스템에 있는 파일에 해당할 수 있다. 컴퓨터

프로그램은 다른 프로그램이나 데이터(예를 들어, 마크 업 언어 문서에 저장된 하나 이상의 스크립트)를 보유하는 파일의 일부, 해당 프로그램 전용 단일 파일 또는 다수의 조정된 파일(예를 들어, 하나 이상의 모듈, 하위 프로그램 또는 코드일부를 저장하는 파일)에 저장될 수 있다. 컴퓨터 프로그램은 하나의 컴퓨터 또는 하나의 사이트에 위치하거나 다수의 사이트에 분산되고 통신 네트워크로 상호 연결된 다수의 컴퓨터에서 실행되도록 배포될 수 있다.

[0150] 본 명세서에 설명된 프로세스 및 로직 흐름은 입력 데이터에 대해 동작하여 출력을 생성함으로써 액션들을 수행하기 위해 하나 이상의 컴퓨터 프로그램(예를 들어, 데이터 처리 시스템(102)의 컴포넌트들)을 실행하는 하나 이상의 프로그래밍 가능 프로세서에 의해 수행될 수 있다. 프로세스 및 로직 흐름은 또한 FPGA 또는 ASIC과 같은 특수 목적 로직 회로에 의해 수행될 수 있으며, 장치도 특수 목적 로직 회로로 구현될 수 있다. 컴퓨터 프로그램 명령 및 데이터 저장에 적합한 디바이스는 예를 들어 반도체 메모리 디바이스(예를 들어, EPROM, EEPROM 및 플래시 메모리 디바이스); 자기 디스크(예를 들어 내부 하드 디스크 또는 이동식 디스크); 광 자기 디스크; 및 CD ROM과 DVD-ROM 디스크를 포함하여 모든 형태의 비-휘발성 메모리, 매체 및 메모리 디바이스를 포함한다. 프로세서와 메모리는 특수 목적 논리 회로에 의해 보완되거나 그에 통합될 수 있다.

[0151] 본 명세서에 설명된 주제는 예를 들어 백엔드 컴포넌트(예를 들어, 데이터 서버), 미들웨어 컴포넌트(예를 들어, 애플리케이션 서버), 프론트 엔드 컴포넌트(예를 들어, 사용자가 본 명세서에 설명된 주제의 구현과 상호작용할 수 있는 그래픽 사용자 인터페이스 또는 웹 브라우저가 있는 클라이언트 컴퓨터), 또는 이러한 백 엔드, 미들웨어 또는 프런트 엔드 컴포넌트 중 하나 이상의 조합을 포함하는 컴퓨팅 시스템에서 구현될 수 있다. 시스템의 컴포넌트들은 디지털 데이터 통신의 모든 형태 또는 매체, 예를 들어 통신 네트워크에 의해 상호 연결될 수 있다. 통신 네트워크의 예는 근거리 통신망("LAN") 및 광역 통신망("WAN"), 네트워크 간(예를 들어, 인터넷) 및 피어-투-피어 네트워크(예를 들어, 애드 혹 피어-투-피어 네트워크)를 포함한다.

[0152] 시스템(100) 또는 시스템(300)과 같은 컴퓨팅 시스템은 클라이언트와 서버를 포함할 수 있다. 클라이언트와 서버는 일반적으로 서로 떨어져 있고 일반적으로 통신 네트워크(예를 들어, 네트워크(105))를 통해 상호작용한다. 클라이언트와 서버의 관계는 개별 컴퓨터에서 실행되고 서로 클라이언트-서버 관계를 갖는 컴퓨터 프로그램으로 인해 발생한다. 일부 구현에서, 서버는 (예를 들어, 클라이언트 디바이스와 상호 작용하는 사용자에게 데이터를 디스플레이하고 사용자로부터 사용자 입력을 수신하기 위해) 클라이언트 디바이스로 데이터(예를 들어, 디지털 컴포넌트를 나타내는 데이터 패킷)를 전송한다. 클라이언트 디바이스에서 생성된 데이터(예를 들어, 사용자 상호 작용의 결과)는 서버에서 클라이언트 디바이스로부터 수신될 수 있다(예를 들어, 데이터 처리 시스템(102)의 인터페이스(104)에 의해 수신됨).

[0153] 동작들은 특정 순서로 도면에 도시되어 있지만, 이러한 동작들은 도시된 특정 순서 또는 순차적인 순서로 수행될 필요가 없으며, 도시된 모든 동작들이 수행될 필요는 없다. 본 명세서에 설명된 동작들은 다른 순서로 수행될 수 있다.

[0154] 다양한 시스템 컴포넌트의 분리는 모든 구현에서 분리를 요구하지 않으며, 설명된 프로그램 컴포넌트들은 단일 하드웨어 또는 소프트웨어 제품에 포함될 수 있다. 예를 들어, 컨텐츠 변환 컴포넌트(110)와 컨텐츠 삽입 컴포넌트(120)는 단일 컴포넌트, 앱 또는 프로그램, 또는 하나 이상의 처리 회로를 갖는 논리 디바이스 일 수 있거나, 데이터 처리 시스템(102)의 하나 이상의 프로세서에 의해 실행될 수 있다.

[0155] 본 기술 솔루션의 적어도 하나의 양태는 오디오 트랙을 생성하는 시스템에 관한 것이다. 시스템은 데이터 처리 시스템을 포함할 수 있다. 데이터 처리 시스템은 하나 이상의 프로세서를 포함할 수 있다. 데이터 처리 시스템은 네트워크를 통해, 데이터 처리 시스템으로부터 떨어진 컴퓨팅 디바이스의 마이크로폰에 의해 검출된 입력 오디오 신호를 포함하는 데이터 패킷을 수신할 수 있다. 데이터 처리 시스템은 입력 오디오 신호를 파싱하여 요청을 식별할 수 있다. 데이터 처리 시스템은 요청에 기초하여 시각적 출력 포멧을 갖는 디지털 컴포넌트 객체를 선택할 수 있으며, 디지털 컴포넌트 객체는 메타 데이터와 관련된다. 데이터 처리 시스템은 컴퓨팅 디바이스의 유형에 기초하여, 디지털 컴포넌트 객체를 오디오 출력 포멧으로 변환하기로 결정할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체를 오디오 출력 포멧으로 변환하기로 한 결정에 응답하여, 디지털 컴포넌트 객체에 대한 텍스트를 생성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체의 컨텍스트에 기초하여, 텍스트를 렌더링할 디지털 음성을 선택할 수 있다. 데이터 처리 시스템은 디지털 음성으로 렌더링된 텍스트로 디지털 컴포넌트 객체의 기준 오디오 트랙을 구성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체의 메타 데이터에 기초하여 비-음성 오디오 큐를 생성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체의 기본 오디오 형식과 비-음성 오디오 큐를 결합하여 디지털 컴포넌트 객체의 오디오 트랙을 생성할 수 있다. 데이터 처리 시

스템은 컴퓨팅 디바이스의 요청에 응답하여, 컴퓨팅 디바이스의 스피커를 통한 출력을 위해 컴퓨팅 디바이스로 디지털 컴포넌트 객체의 오디오 트랙을 제공할 수 있다.

[0156] 데이터 처리 시스템은 스마트 스피커를 포함하는 컴퓨팅 디바이스의 유형에 기초하여 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정할 수 있다. 데이터 처리 시스템은 디지털 어시스턴트를 포함하는 컴퓨팅 디바이스의 유형에 기초하여 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정할 수 있다.

[0157] 데이터 처리 시스템은 요청에 응답하여, 실시간 컨텐츠 선택 프로세스에 입력된 컨텐츠 선택 기준에 기초하여 디지털 컴포넌트 객체를 선택할 수 있으며, 디지털 컴포넌트 객체는 복수의 제3자 컨텐츠 제공자에 의해 제공된 복수의 디지털 컴포넌트 객체로부터 선택될 수 있다. 데이터 처리 시스템은 요청 이전에 컴퓨팅 디바이스에 의해 렌더링된 컨텐츠와 관련된 키워드들에 기초하여 디지털 컴포넌트 객체를 선택할 수 있다. 디지털 컴포넌트 객체는 복수의 제3자 컨텐츠 제공자에 의해 제공된 복수의 디지털 컴포넌트 객체로부터 선택될 수 있다.

[0158] 데이터 처리 시스템은 자연어 생성 모델을 통해, 디지털 컴포넌트 객체의 메타 데이터에 기초하여 디지털 컴포넌트 객체에 대한 텍스트를 생성할 수 있다. 데이터 처리 시스템은 음성 모델을 통해, 디지털 컴포넌트 객체의 컨텍스트에 기초하여 디지털 음성을 선택할 수 있다. 음성 모델은 오디오 및 시각적 미디어 컨텐츠를 포함하는 이력 데이터 세트를 사용하여 기계 학습 기술로 트레이닝될 수 있다.

[0159] 데이터 처리 시스템은 음성 특성 벡터를 생성하기 위해 음성 모델에 디지털 컴포넌트 객체의 컨텍스트를 입력할 수 있다. 음성 모델은 오디오 및 시각적 미디어 컨텐츠를 포함하는 이력 데이터 세트를 사용하여 기계 학습 엔진에 의해 트레이닝될 수 있다. 데이터 처리 시스템은 음성 특성 벡터에 기초하여 복수의 디지털 음성으로부터 디지털 음성을 선택할 수 있다.

[0160] 데이터 처리 시스템은 메타 데이터에 기초하여 오디오 트랙에 트리거 단어를 추가하기로 결정할 수 있다. 제2 입력 오디오 신호에서 트리거 단어의 검출에 응답하여, 데이터 처리 시스템 또는 컴퓨팅 디바이스가 트리거 단어에 대응하는 디지털 액션을 수행하게 한다.

[0161] 데이터 처리 시스템은 디지털 컴포넌트 객체의 카테고리를 결정할 수 있다. 데이터 처리 시스템은 데이터베이스로부터 카테고리와 관련된 복수의 디지털 액션에 대응하는 복수의 트리거 단어를 검색할 수 있다. 데이터 처리 시스템은 트리거 키워드들의 이력(과거) 수행에 기초하여 트레이닝된 디지털 액션 모델을 사용하여, 디지털 컴포넌트 객체의 컨텍스트 및 컴퓨팅 디바이스의 유형에 기초하여 복수의 트리거 단어를 순위 지정할 수 있다. 데이터 처리 시스템은 가장 높은 순위의 트리거 키워드를 선택하여 오디오 트랙에 추가할 수 있다.

[0162] 데이터 처리 시스템은 디지털 컴포넌트 객체에서 시각적 객체를 식별하기 위해 디지털 컴포넌트 객체에 대해 이미지 인식을 수행할 수 있다. 데이터 처리 시스템은 데이터베이스에 저장된 복수의 비-음성 오디오 큐로부터, 시각적 객체에 대응하는 비-음성 오디오 큐를 선택할 수 있다.

[0163] 데이터 처리 시스템은 이미지 인식 기술을 통해 디지털 컴포넌트 객체에서 복수의 시각적 객체를 식별할 수 있다. 데이터 처리 시스템은 복수의 시각적 객체에 기초하여, 복수의 비-음성 오디오 큐를 선택할 수 있다. 데이터 처리 시스템은 시각적 객체 각각과 메타 데이터 간의 매칭 레벨을 나타내는 각 시각적 객체에 대한 매칭 스코어를 결정할 수 있다. 데이터 처리 시스템은 매칭 스코어에 기초하여 복수의 비-음성 오디오 큐를 순위 지정할 수 있다. 데이터 처리 시스템은 복수의 비-음성 오디오 큐 각각과 텍스트를 렌더링하기 위해 컨텍스트에 기초하여 선택된 디지털 음성 간의 오디오 간섭 레벨을 결정할 수 있다. 데이터 처리 시스템은 최고 순위에 기초하여, 임계값 미만의 오디오 간섭 레벨과 관련된 복수의 비-음성 오디오 큐로부터 비-음성 오디오 큐를 선택할 수 있다.

[0164] 데이터 처리 시스템은 이력 수행 데이터를 사용하여 트레이닝된 삽입 모델에 기초하여, 컴퓨팅 디바이스에 의해 출력된 디지털 미디어 스트림의 오디오 트랙에 대한 삽입 지점을 식별할 수 있다. 데이터 처리 시스템은 컴퓨팅 디바이스로 하여금 디지털 미디어 스트림의 삽입 지점에서 오디오 트랙을 렌더링하도록 하는 명령을 컴퓨팅 디바이스로 제공할 수 있다.

[0165] 본 기술 솔루션의 적어도 하나의 양태는 오디오 트랙을 생성하는 방법에 관한 것이다. 이 방법은 데이터 처리 시스템의 하나 이상의 프로세서에 의해 수행될 수 있다. 이 방법은 데이터 처리 시스템이 데이터 처리 시스템으로부터 떨어진 컴퓨팅 디바이스의 마이크로폰에 의해 검출된 입력 오디오 신호를 포함하는 데이터 패킷을 수신하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 요청을 식별하기 위해 입력 오디오 신호를 파싱하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 요청에 기초하여 시각적 출력 포맷을 갖는 디지털 컴포넌트 객체를 선택하는 단계를 포함할 수 있고, 디지털 컴포넌트 객체는 메타 데이터와 관련된다. 방법은 데이터

처리 시스템이 컴퓨팅 디바이스의 유형에 기초하여, 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 한 결정에 응답하여, 디지털 컴포넌트 객체에 대한 텍스트를 생성하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 디지털 컴포넌트 객체의 컨텍스트에 기초하여, 텍스트를 렌더링하기 위한 디지털 음성을 선택하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 디지털 음성에 의해 렌더링된 텍스트로 디지털 컴포넌트 객체의 기본 오디오 트랙을 구성하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 디지털 컴포넌트 객체에 기초하여 비-음성 오디오 큐를 생성하는 단계를 포함할 수 있다. 이 방법은 데이터 처리 시스템이 디지털 컴포넌트 객체의 오디오 트랙을 생성하기 위해 비-음성 오디오 큐를 디지털 컴포넌트 객체의 기본 오디오 형식과 결합하는 단계를 포함할 수 있다. 이 방법은 데이터 처리 시스템이 컴퓨팅 디바이스로부터의 요청에 응답하여, 컴퓨팅 디바이스의 스피커를 통한 출력을 위해 컴퓨팅 디바이스에 디지털 컴포넌트 객체의 오디오 트랙을 제공하는 단계를 포함할 수 있다.

[0166] 방법은 데이터 처리 시스템이 스마트 스피커를 포함하는 컴퓨팅 디바이스의 유형에 기초하여 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정하는 단계를 포함할 수 있다. 방법은 데이터 처리 시스템이 요청에 응답하여, 실시간 컨텐츠 선택 프로세스에 입력된 컨텐츠 선택 기준에 기초하여 디지털 컴포넌트 객체를 선택하는 단계를 포함할 수 있고, 디지털 컴포넌트 객체는 복수의 제3자 컨텐츠 제공자에 의해 제공된 복수의 디지털 컴포넌트 객체로부터 선택될 수 있다.

[0167] 방법은 데이터 처리 시스템이 요청 이전에 컴퓨팅 디바이스에 의해 렌더링된 컨텐츠와 관련된 키워드들에 기초하여 디지털 컴포넌트 객체를 선택하는 단계를 포함할 수 있다. 디지털 컴포넌트 객체는 복수의 제3자 컨텐츠 제공자에 의해 제공된 복수의 디지털 컴포넌트 객체로부터 선택될 수 있다.

[0168] 본 기술 솔루션의 적어도 하나의 양태는 오디오 트랙을 생성하는 시스템에 관한 것이다. 시스템은 하나 이상의 프로세서를 갖는 데이터 처리 시스템을 포함할 수 있다. 데이터 처리 시스템은 컴퓨팅 디바이스에 의해 렌더링된 디지털 스트리밍 컨텐츠와 관련된 키워드들을 식별할 수 있다. 데이터 처리 시스템은 키워드들에 기초하여, 시각적 출력 포맷을 갖는 디지털 컴포넌트 객체를 선택할 수 있고, 디지털 컴포넌트 객체는 메타 데이터와 관련된다. 데이터 처리 시스템은 컴퓨팅 디바이스의 유형에 기초하여, 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 한 결정에 응답하여, 디지털 컴포넌트 객체에 대한 텍스트를 생성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체의 컨텍스트에 기초하여 텍스트를 렌더링하기 위한 디지털 음성을 선택할 수 있다. 데이터 처리 시스템은 디지털 음성에 의해 렌더링된 텍스트로 디지털 컴포넌트 객체의 기준 오디오 트랙을 구성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체에 기초하여 비-음성 오디오 큐를 생성할 수 있다. 데이터 처리 시스템은 디지털 컴포넌트 객체의 기본 오디오 형식과 비-음성 오디오 큐를 결합하여 디지털 컴포넌트 객체의 오디오 트랙을 생성할 수 있다. 데이터 처리 시스템은 컴퓨팅 디바이스의 스피커를 통한 출력을 위해 컴퓨팅 디바이스에 디지털 컴포넌트 객체의 오디오 트랙을 제공할 수 있다.

[0169] 데이터 처리 시스템은 스마트 스피커를 포함하는 컴퓨팅 디바이스의 유형에 기초하여 디지털 컴포넌트 객체를 오디오 출력 포맷으로 변환하기로 결정할 수 있다. 데이터 처리 시스템은 실시간 컨텐츠 선택 프로세스에 입력된 키워드에 기초하여 디지털 컴포넌트 객체를 선택할 수 있으며, 디지털 컴포넌트 객체는 복수의 제3자 컨텐츠 제공자에 의해 제공된 복수의 디지털 컴포넌트 객체로부터 선택된다.

[0170] 이제 일부 예시적인 구현을 설명 하였지만, 전술한 내용은 예시적인 것이며 제한적이지 않으며 예로서 제시됨 것임이 명백하다. 특히, 본 명세서에 제시된 많은 예가 방법 동작 또는 시스템 요소의 특정 조합을 포함하지만, 이러한 동작 및 이러한 요소는 동일한 목적을 달성하기 위해 다른 방식으로 결합될 수 있다. 하나의 구현과 관련하여 논의된 동작(Act), 요소 및 기능은 다른 구현이나 구현들에서 유사한 역할에서 제외되지 않는다.

[0171] 본 명세서에서 사용된 어법 및 용어는 설명을 위한 것이며 제한하는 것으로 간주되어서는 안된다. "포함하는(including)", 포함하는(comprising)", "갖는", "포함하는(containing)", "포함하는(involving)" "특징되는(characterized by)", "것을 특징으로 하는(characterized in that)" 및 이의 변형은 이후에 나열된 항목, 그와 동등한 항목 및 추가 항목뿐만 아니라 이후에 독점적으로 나열된 항목으로 구성된 대체 구현을 포함하는 것을 의미한다. 일 구현에서, 본 명세서에 설명된 시스템 및 방법은 설명된 요소, 동작 또는 컴포넌트 중 하나 이상, 또는 모두의 각각의 조합으로 구성된다.

[0172] 본 명세서에서 단수로 언급된 시스템 및 방법의 구현 또는 요소 또는 동작에 대한 모든 참조는 또한 복수의 이러한 요소를 포함하는 구현을 포함할 수 있으며, 본 명세서에서 임의의 구현 또는 요소 또는 동작에 대한 복수

의 참조는 또한 단일 요소만을 포함하는 구현을 포함할 수 있다. 단수 또는 복수 형태의 참조는 현재 개시된 시스템 또는 방법, 이들의 컴포넌트, 동작 또는 요소를 단일 또는 복수의 구성으로 제한하려는 것이 아니다. 임의의 정보, 동작 또는 요소에 기초하는 임의의 동작 또는 요소에 대한 참조는 동작 또는 요소가 정보, 동작 또는 요소에 적어도 부분적으로 기초하는 구현을 포함할 수 있다.

[0173] 본 명세서에 개시된 임의의 구현은 임의의 다른 구현 또는 실시예와 결합될 수 있으며, "구현", "일부 구현", "하나의 구현"등에 대한 언급은 반드시 상호 배타적인 것은 아니며, 그 구현과 관련하여 설명된 특정 특징, 구조 또는 특성이 적어도 하나의 구현 또는 실시예에 포함될 수 있음을 나타내도록 의도된다. 본 명세서에서 사용되는 이러한 용어는 반드시 모두 동일한 구현을 지칭하는 것은 아니다. 임의의 구현은 본 명세서에 개시된 양태 및 구현과 일치하는 임의의 방식으로 포괄적으로 또는 배타적으로 임의의 다른 구현과 결합될 수 있다.

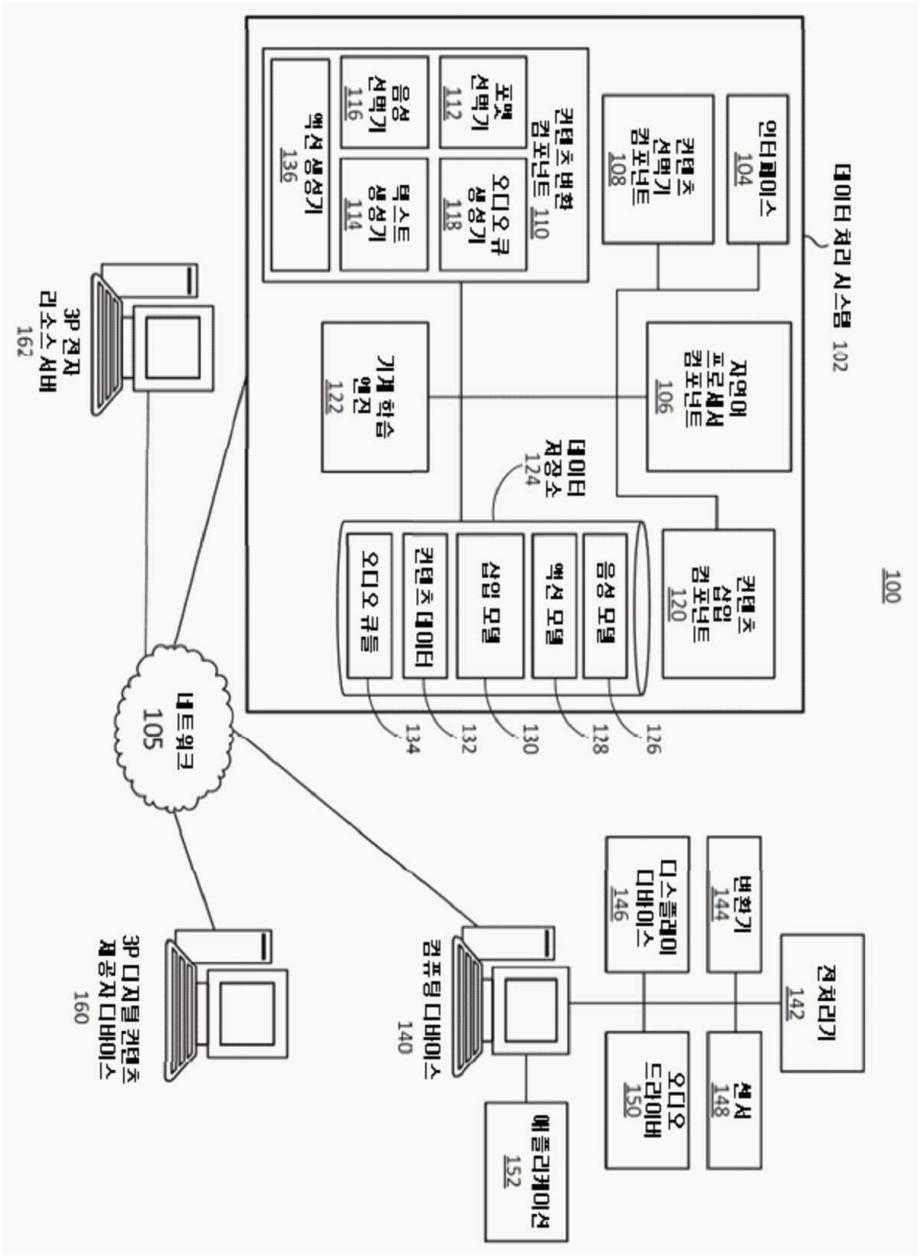
[0174] "또는"에 대한 언급은 "또는"을 사용하여 설명된 임의의 용어가 설명된 용어들 중 임의의 하나, 둘 이상 및 모두를 나타낼 수 있도록 포괄적인 것으로 해석될 수 있다. 용어의 결합 리스트 중 적어도 하나에 대한 언급은 설명된 용어의 단일, 둘 이상 및 모두를 나타내는 포괄적 OR로 해석될 수 있다. 예를 들어, " 'A'와 'B'중 적어도 하나"에 대한 언급은 'A'만, 'B'만, 'A'와 'B' 모두를 포함할 수 있다. "포함하는" 또는 기타 공개 용어와 함께 사용되는 이러한 언급은 추가 아이템을 포함할 수 있다.

[0175] 도면의 기술적 특징, 상세한 설명 또는 청구항 다음에 참조 부호가 오는 경우, 도면, 상세한 설명 및 청구 범위의 이해도를 높이기 위해 참조 부호가 포함되었다. 따라서, 참조 부호나 그의 부재는 청구 요소의 범위에 제한적인 영향을 미치지 않는다.

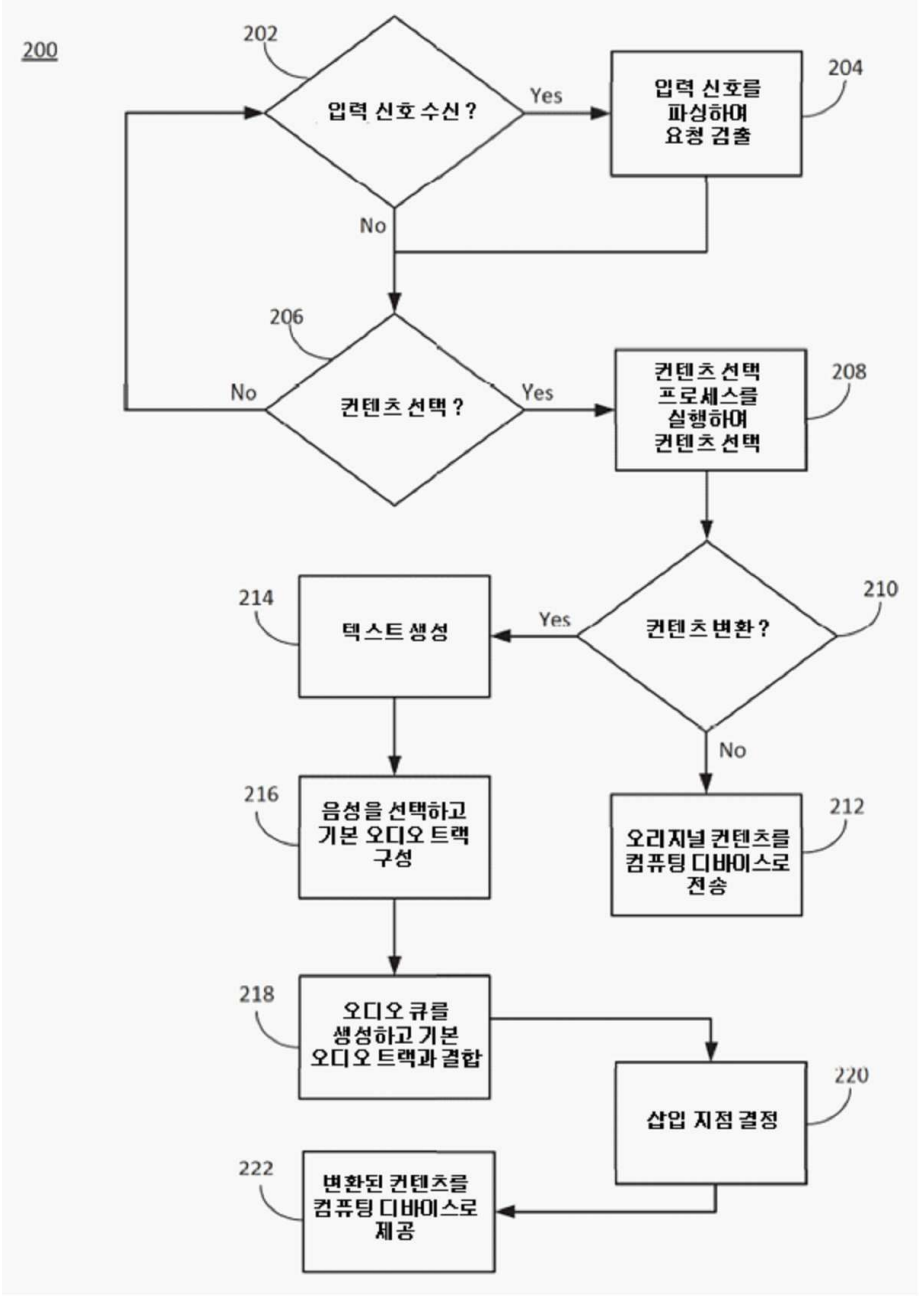
[0176] 본 명세서에 설명된 시스템 및 방법은 특성에서 벗어나지 않고 다른 특정 형태로 구현될 수 있다. 예를 들어, 3P 또는 3P 디지털 컨텐츠 제공자 디바이스(160)과 같은 제3자로 설명된 디바이스, 제품 또는 서비스는 부분적으로 또는 전체적으로 제1자 디바이스, 제품 또는 서비스이거나 이를 포함할 수 있고, 데이터 처리 시스템(102) 또는 다른 컴포넌트와 관련된 엔티티에 의해 공통적으로 소유될 수 있다. 전술한 구현들은 설명된 시스템 및 방법을 제한하기 보다는 예시적인 것이다. 따라서 본 명세서에 설명된 시스템 및 방법의 범위는 전술한 설명이 아니라 첨부된 청구 범위에 의해 표시되고, 청구 범위의 동가성의 의미 및 범위 내에 있는 변경이 본 명세서에 포함된다.

도면

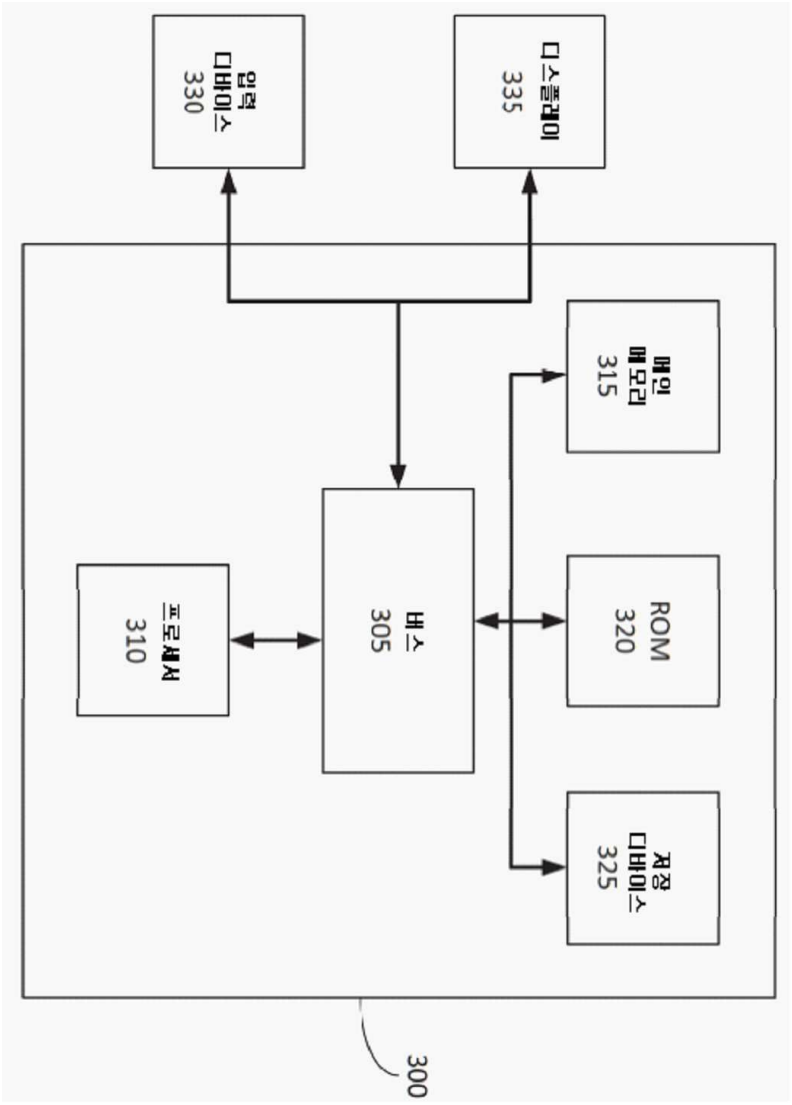
도면1


데이터 처리 시스템 102
100
인터페이스 104
컨텐츠 선택기 컴포넌트 108
컨텐츠 변환 컴포넌트 110
포맷 선택기 112
오디오 큐 생성기 118
음성 선택기 116
텍스트 생성기 114
액션 생성기 136
자연어 프로세서 컴포넌트 106
기계 학습 엔진 122
데이터 저장소 124
음성 모델 126
액션 모델 128
삽입 모델 130
컨텐츠 데이터 132
오디오 큐 클 134
컨텐츠 삽입 컴포넌트 120
네트워크 105
3P 컨텐츠 리소스 서버 162
3P 디지털 컨텐츠 제공자 디바이스 160
컴퓨팅 디바이스 140
프리젠터 142
변환기 144
디스플레이 디바이스 146
센서 148
오디오 드라이버 150
애플리케이션 152

도면2


200
202
입력 신호 수신?
Yes
204
입력 신호를 파싱하여 요청 검출
No
206
컨텐츠 선택?
No
Yes
208
컨텐츠 선택 프로세스를 실행하여 컨텐츠 선택
210
컨텐츠 변환?
Yes
214
텍스트 생성
No
212
오리지널 컨텐츠를 컴퓨팅 디바이스로 전송
216
음성을 선택하고 기본 오디오 트랙 구성
218
오디오 큐를 생성하고 기본 오디오 트랙과 결합
220
삽입 지점 결정
222
변환된 컨텐츠를 컴퓨팅 디바이스로 제공

도면3


입력 디바이스
330
디스플레이
335
메인 메모리
315
버스
305
ROM
320
프로세서
310
저장 디바이스
325
300