



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2019년03월20일
 (11) 등록번호 10-1932181
 (24) 등록일자 2018년12월18일

- (51) 국제특허분류(Int. Cl.)
 G10L 15/26 (2006.01) G10L 15/183 (2013.01)
- (21) 출원번호 10-2013-7021258
 (22) 출원일자(국제) 2012년01월13일
 심사청구일자 2017년01월11일
 (85) 번역문제출일자 2013년08월12일
 (65) 공개번호 10-2013-0133832
 (43) 공개일자 2013년12월09일
 (86) 국제출원번호 PCT/US2012/021221
 (87) 국제공개번호 WO 2012/099788
 국제공개일자 2012년07월26일
- (30) 우선권주장
 13/040,553 2011년03월04일 미국(US)
 61/435,022 2011년01월21일 미국(US)
- (56) 선행기술조사문헌
 US20030149561 A1*
 US20050193144 A1*
 US20080221902 A1*
 *는 심사관에 의하여 인용된 문헌
- (73) 특허권자
 구글 엘엘씨
 미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이
 1600 (우:94043)
- (72) 발명자
 로이드 매튜 아이.
 미국 메사추세츠주 02142 캠브리지 씨드 스트리트
 #823 303
 리스벳 판카즈
 인도 방갈로르 560037 돛다나쿤디 아우터 링 로드
 아크메 발렛 씨-704
- (74) 대리인
 박장원, 특허법인태평양

전체 청구항 수 : 총 29 항

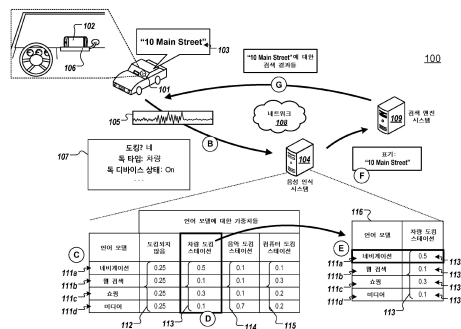
심사관 : 정성윤

(54) 발명의 명칭 **디바이스 도킹 컨텍스트를 이용한 음성인식**

(57) 요약

컴퓨터 저장 매체 상에 인코딩된 컴퓨터 프로그램을 포함하여, 도킹 컨텍스트를 이용하여 음성인식을 수행하기 위한 방법, 시스템, 장치가 개시된다. 일 측면에서, 방법은 인코딩된 음성을 포함하는 오디오 데이터에 액세스하는 단계를 포함한다. 클라이언트 디바이스의 도킹 컨텍스트를 나타내는 정보가 액세스되고, 도킹 컨텍스트는 오디오 데이터와 연관되어 있다. 복수의 언어 모델이 식별된다. 복수개의 언어 모델들 중 적어도 하나가 도킹 컨텍스트에 의해 선택된다. 음성인식은 오디오 데이터의 일부에 대한 표기를 식별하기 위해 선택된 언어 모델을 이용하여 오디오 데이터에 대하여 수행된다.

대표도



명세서

청구범위

청구항 1

디바이스 도킹 컨텍스트를 이용하여 음성 인식(speech recognition)을 수행하는 컴퓨터로 구현되는 방법에 있어서,

서버 시스템에서, 인코딩된 음성을 포함하는 오디오 데이터를 수신하는 단계 와, 상기 인코딩된 음성은 클라이언트 디바이스에 의해 탐지된 것이고;

상기 서버 시스템에서, 상기 클라이언트 디바이스에 의해 상기 오디오 데이터로 인코딩된 음성이 탐지되는 동안, 상기 클라이언트 디바이스의 도킹 컨텍스트(docking context)를 나타내는 정보를 수신하는 단계;

복수의 언어 모델들을 식별하는 단계와, 상기 복수의 언어 모델들 각각은 용어들의 시퀀스에서 상기 시퀀스의 다른 용어들에 기초한 용어의 발생 확률을 나타내고;

상기 복수의 언어 모델들 각각에 대하여, 상기 도킹 컨텍스트와 관련하여 저장된 가중치에 액세스함으로써 상기 도킹 컨텍스트에 기초하여 상기 언어 모델에 할당하기 위한 가중치를 결정하는 단계와, 상기 가중치는 상기 언어 모델이 상기 인코딩된 음성의 정확한 표기(transcription)를 나타낼 확률을 나타내고;

상기 할당된 가중치에 기초하여 상기 복수개의 언어 모델들 중 적어도 하나를 선택하는 단계; 및

상기 오디오 데이터의 일부분에 대한 표기를 식별하기 위해 상기 선택된 언어 모델을 이용하여 상기 오디오 데이터에 대한 음성 인식을 수행하는 단계를 포함하는 방법.

청구항 2

제1항에 있어서,

상기 도킹 컨텍스트는 상기 클라이언트 디바이스에 의해 상기 오디오 데이터로 인코딩된 음성이 탐지되는 동안 상기 클라이언트가 연결된 도킹 스테이션의 타입을 나타내는 것을 특징으로 하는 방법.

청구항 3

제1항에 있어서,

상기 인코딩된 음성은 하나 이상의 쿼리 용어들을 포함하고,

상기 표기는 상기 쿼리 용어들을 포함하고,

상기 방법은,

상기 쿼리 용어들을 포함하는 검색 쿼리를 생성하는 단계;

상기 검색 쿼리를 이용하여 검색을 수행하는 단계; 및

상기 검색의 결과들을 나타내는 정보를 상기 클라이언트 디바이스로 제공하는 단계를 더 포함하는 방법.

청구항 4

제1항에 있어서,

상기 복수의 언어 모델들 각각은 특정 주제 카테고리의 단어들에 대해 훈련된 것을 특징으로 하는 방법.

청구항 5

제1항에 있어서,

상기 도킹 컨텍스트에 기초하여 가중치를 결정하는 단계는,

상기 클라이언트 디바이스가 차량 도킹 스테이션에 연결되어 있는지를 결정하는 단계; 및

상기 클라이언트 디바이스가 차량 도킹 스테이션에 연결되어 있음을 결정함에 응답하여, 주소를 출력하도록 훈련된 네비게이션 언어 모델에 대해서, 상기 네비게이션 언어 모델이 선택되는 확률을 상기 복수의 언어 모델들 내의 다른 언어 모델들에 대하여 증가시키는 가중치를 결정하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 6

디바이스 도킹 컨텍스트를 이용하여 음성 인식(speech recognition)을 수행하는 컴퓨터로 구현되는 방법에 있어서,

인코딩된 음성을 포함하는 오디오 데이터에 액세스하는 단계;

클라이언트 디바이스의 도킹 컨텍스트를 나타내는 정보에 액세스하는 단계, 상기 도킹 컨텍스트는 상기 오디오 데이터와 관련되고;

복수의 언어 모델들을 식별하는 단계;

상기 복수의 언어 모델 각각에 대해, 상기 도킹 컨텍스트에 기초하여 가중치를 결정하는 단계, 상기 가중치는 상기 언어 모델이 상기 인코딩된 음성에 대한 정확한 표기(transcription)를 나타낼 확률을 나타내고;

상기 가중치에 기초하여 상기 복수의 언어 모델들 중에서 적어도 하나를 선택하는 단계; 및

상기 오디오 데이터의 일부분에 대한 표기를 식별하기 위해 상기 선택된 적어도 하나의 언어 모델을 이용하여 상기 오디오 데이터에 대한 음성 인식을 수행하는 단계를 포함하는 방법.

청구항 7

제6항에 있어서,

상기 클라이언트 디바이스의 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스와 상기 클라이언트 디바이스가 물리적으로 연결된 제2 디바이스 사이의 연결을 나타내는 것을 특징으로 하는 방법.

청구항 8

제6항에 있어서,

상기 클라이언트 디바이스의 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스와 상기 클라이언트 디바이스가 무선으로 연결된 제2 디바이스 사이의 연결을 나타내는 것을 특징으로 하는 방법.

청구항 9

삭제

청구항 10

제6항에 있어서,

상기 오디오 데이터로 인코딩된 음성은 상기 클라이언트 디바이스에 의해 탐지되고,

상기 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스에 의해 상기 오디오 데이터로 인코딩된 음성이 탐지되는 동안 상기 클라이언트 디바이스가 도킹 스테이션에 연결되었는지 여부를 나타내는 것을 특징으로 하는 방법.

청구항 11

제6항에 있어서,

상기 오디오 데이터로 인코딩된 음성은 상기 클라이언트 디바이스에 의해 탐지되고,

상기 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스에 의해 상기 오디오 데이터로 인코딩된 음성이 탐지되는 동안 상기 클라이언트 디바이스가 연결된 도킹 스테이션의 타입을 나타내는 것을 특징으로 하는 방법.

법.

청구항 12

제6항에 있어서,

상기 인코딩된 음성은 하나 이상의 발화된(spoken) 쿼리 용어들을 포함하고,

상기 표기는 상기 발화된 쿼리 용어들의 표기를 포함하고,

상기 방법은,

검색 엔진이 상기 하나 이상의 발화된 쿼리 용어들의 상기 표기를 이용하여 검색을 수행하도록 하는 단계; 및

상기 검색 쿼리의 결과들을 나타내는 정보를 상기 클라이언트 디바이스로 제공하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 13

제6항에 있어서,

상기 복수의 언어 모델들 각각에 대하여 가중치를 결정하는 단계는, 상기 도킹 컨텍스트와 관련된 저장된 가중치들에 액세스하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 14

제6항에 있어서,

상기 복수의 언어 모델들 각각에 대하여 가중치를 결정하는 단계는, 저장된 가중치들에 액세스하는 단계 및 상기 도킹 컨텍스트에 기초하여 상기 저장된 가중치를 변경하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 15

제6항에 있어서,

상기 복수의 언어 모델들 각각은 특정 주제 카테고리의 단어들에 대해 훈련된 것을 특징으로 하는 방법.

청구항 16

제6항에 있어서,

상기 도킹 컨텍스트에 기초하여 가중치를 결정하는 단계는,

상기 도킹 컨텍스트가 차량 도킹 스테이션에의 연결을 포함하는지를 결정하는 단계; 및

상기 도킹 컨텍스트가 차량 도킹 스테이션에의 연결을 포함함을 결정함에 응답하여, 주소를 출력하도록 훈련된 네비게이션 언어 모델에 대해서, 상기 네비게이션 언어 모델이 선택되는 확률을 상기 복수의 언어 모델들 내의 다른 언어 모델들에 대하여 증가시키는 가중치를 결정하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 17

디바이스 도킹 컨텍스트를 이용하여 음성 인식(speech recognition)을 수행하는 시스템에 있어서,

하나 이상의 프로세서; 및

상기 하나 이상의 프로세서와 연결되고, 상기 하나 이상의 프로세서에 의해 실행될 때 상기 시스템으로 하여금 작동들(operations)을 수행하도록 하는 명령들(instructions)이 저장된 매체를 포함하고, 상기 작동들은,

인코딩된 음성을 포함하는 오디오 데이터에 액세스하는 단계;

클라이언트 디바이스의 도킹 컨텍스트를 나타내는 정보에 액세스하는 단계, 상기 도킹 컨텍스트는 상기 오디오 데이터와 관련되고;

복수의 언어 모델들을 식별하는 단계;

상기 복수의 언어 모델 각각에 대해, 상기 도킹 컨텍스트에 기초하여 가중치를 결정하는 단계, 상기 가중치는 상기 언어 모델이 상기 인코딩된 음성에 대한 정확한 표기(transcription)를 나타낼 확률을 나타내고; 상기 가중치에 기초하여 상기 복수의 언어 모델들 중에서 적어도 하나를 선택하는 단계; 및

상기 오디오 데이터의 일부분에 대한 표기를 식별하기 위해 상기 선택된 적어도 하나의 언어 모델을 이용하여 상기 오디오 데이터에 대한 음성 인식을 수행하는 단계를 포함하는 것을 특징으로 하는 시스템.

청구항 18

삭제

청구항 19

제17항에 있어서,

상기 오디오 데이터로 인코딩된 음성은 상기 클라이언트 디바이스에 의해 탐지되고,

상기 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스에 의해 상기 오디오 데이터로 인코딩된 음성이 탐지되는 동안 상기 클라이언트 디바이스가 도킹 스테이션에 연결되었는지 여부를 나타내는 것을 특징으로 하는 시스템.

청구항 20

제17항에 있어서,

상기 오디오 데이터로 인코딩된 음성은 상기 클라이언트 디바이스에 의해 탐지되고,

상기 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스에 의해 상기 오디오 데이터로 인코딩된 음성이 탐지되는 동안 상기 클라이언트 디바이스가 연결된 도킹 스테이션의 타입을 나타내는 것을 특징으로 하는 시스템.

청구항 21

디바이스 도킹 컨텍스트를 이용하여 음성 인식(speech recognition)을 수행하는 컴퓨터 프로그램이 인코딩된 비-일시적인 컴퓨터 판독가능 저장 매체에 있어서, 상기 프로그램은 하나 이상의 컴퓨터에 의해 실행될 때 상기 하나 이상의 컴퓨터로 하여금 동작들을 수행하도록 하는 명령들을 포함하고, 상기 동작들은,

인코딩된 음성을 포함하는 오디오 데이터에 액세스하는 단계;

클라이언트 디바이스의 도킹 컨텍스트를 나타내는 정보에 액세스하는 단계, 상기 도킹 컨텍스트는 상기 오디오 데이터와 관련되고;

복수의 언어 모델들을 식별하는 단계;

상기 복수의 언어 모델 각각에 대해, 상기 도킹 컨텍스트에 기초하여 가중치를 결정하는 단계, 상기 가중치는 상기 언어 모델이 상기 인코딩된 음성에 대한 정확한 표기(transcription)를 나타낼 확률을 나타내고;

상기 가중치에 기초하여 상기 복수의 언어 모델들 중에서 적어도 하나를 선택하는 단계; 및

상기 오디오 데이터의 일부분에 대한 표기를 식별하기 위해 상기 선택된 적어도 하나의 언어 모델을 이용하여 상기 오디오 데이터에 대한 음성 인식을 수행하는 단계를 포함하는 것을 특징으로 하는 비-일시적인 컴퓨터 판독가능 저장 매체.

청구항 22

삭제

청구항 23

제21항에 있어서,

상기 오디오 데이터로 인코딩된 음성은 상기 클라이언트 디바이스에 의해 탐지되고,

상기 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스에 의해 상기 오디오 데이터로 인코딩된 음성이 탐지되는 동안 상기 클라이언트 디바이스가 도킹 스테이션에 연결되었는지 여부를 나타내는 것을 특징으로 하는 비-일시적인 컴퓨터 판독가능 저장 매체.

청구항 24

제21항에 있어서,

상기 오디오 데이터로 인코딩된 음성은 상기 클라이언트 디바이스에 의해 탐지되고,

상기 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스에 의해 상기 오디오 데이터로 인코딩된 음성이 탐지되는 동안 상기 클라이언트 디바이스가 연결된 도킹 스테이션의 타입을 나타내는 것을 특징으로 하는 비-일시적인 컴퓨터 판독가능 저장 매체.

청구항 25

제21항에 있어서,

상기 인코딩된 음성은 하나 이상의 발화된(spoken) 쿼리 용어들을 포함하고,

상기 표기는 상기 발화된 쿼리 용어들의 표기를 포함하고,

상기 동작들은,

검색 엔진이 상기 하나 이상의 발화된 쿼리 용어들의 상기 표기를 이용하여 검색을 수행하도록 하는 단계; 및

상기 검색 쿼리의 결과들을 나타내는 정보를 상기 클라이언트 디바이스로 제공하는 단계를 더 포함하는 것을 특징으로 하는 비-일시적인 컴퓨터 판독가능 저장 매체.

청구항 26

디바이스 도킹 컨텍스트를 이용하여 음성 인식(speech recognition)을 수행하는 컴퓨터로 구현되는 방법에 있어서,

클라이언트 디바이스에서 음성을 포함하는 오디오를 탐지하는 단계;

상기 탐지된 오디오를 오디오 데이터로 인코딩하는 단계;

상기 오디오 데이터를 서버 시스템으로 전송하는 단계;

상기 클라이언트 디바이스의 도킹 컨텍스트를 식별하는 단계;

상기 도킹 컨텍스트를 나타내는 정보를 상기 서버 시스템으로 전송하는 단계; 및

상기 오디오 데이터의 적어도 일부의 표기를 상기 클라이언트 디바이스에서 수신하는 단계를 포함하고,

상기 서버 시스템은,

복수의 언어 모델 각각에 대해, 상기 도킹 컨텍스트에 기초하여 가중치를 결정하는 단계, 상기 가중치는 상기 언어 모델이 상기 인코딩된 음성에 대한 정확한 표기(transcription)를 나타낼 확률을 나타내고,

상기 가중치에 기초하여 복수의 언어 모델들로부터 언어 모델을 선택하고,

상기 선택된 적어도 하나의 언어 모델을 이용하여 상기 오디오 데이터에 대한 음성 인식을 수행하여 상기 표기를 생성하고,

상기 표기를 상기 클라이언트 디바이스로 전송하는 것을 특징으로 하는 방법.

청구항 27

제26항에 있어서,

상기 식별된 도킹 컨텍스트는 상기 오디오가 탐지된 시점의 상기 클라이언트 디바이스의 도킹 컨텍스트인 것을 특징으로 하는 방법.

청구항 28

제26항에 있어서,

상기 클라이언트 디바이스의 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스와 상기 클라이언트 디바이스가 물리적으로 연결된 제2 디바이스 사이의 연결을 나타내는 것을 특징으로 하는 방법.

청구항 29

제26항에 있어서,

상기 클라이언트 디바이스의 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스와 상기 클라이언트 디바이스가 무선으로 연결된 제2 디바이스 사이의 연결을 나타내는 것을 특징으로 하는 방법.

청구항 30

제6항에 있어서,

상기 도킹 컨텍스트는 제1 도킹 스테이션을 갖는 클라이언트 디바이스의 도킹을 나타내고; 그리고

상기 복수의 언어 모델 각각에 대해, 상기 도킹 컨텍스트에 기초하여 상기 가중치를 결정하는 단계는,

상기 제1 도킹 스테이션이 소정의 도킹 스테이션 타입 세트로부터의 특정 도킹 스테이션 타입을 갖는지 결정하는 단계; 및

상기 특정 도킹 스테이션 타입에 대응하는 가중치를 결정하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 31

제6항에 있어서,

상기 복수의 언어 모델 각각에 대해, 상기 도킹 컨텍스트에 기초하여 상기 가중치를 결정하는 단계는,

상기 언어 모델을 사용하여 상기 오디오 데이터에 대한 표기를 식별하기 전에 각 언어 모델에 대한 상기 가중치를 결정하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 32

제6항에 있어서,

상기 선택된 적어도 하나의 언어 모델을 사용하여 상기 오디오 데이터에 대한 음성 인식을 수행하는 단계는 상기 적어도 하나의 언어 모델을 선택하는 것에 응답하여 발생하는 것을 특징으로 하는 방법.

발명의 설명

기술 분야

[0001] 본 출원은 2011년 1월 21일 출원된 미국 가출원 61/435,022와 2011년 3월 4일에 출원된 미국 출원 13/040,553에 대하여 우선권을 주장한다. 이들 출원의 모든 내용은 여기에 참조로서 통합된다.

배경 기술

[0002] 음성인식의 사용은 점점 더 일반화되고 있다. 기술이 진보함에 따라, 컴퓨팅 디바이스의 사용자들은 음성인식 기능으로의 증가된 액세스를 얻어왔다. 많은 사용자들인 그들의 직업과 일상의 다른 측면들에서 음성인식에 의존한다.

발명의 내용

해결하려는 과제

[0003] 다양한 실시예에서, 클라이언트 디바이스의 도킹 컨텍스트(docking context)는 음성인식(speech recognition)의 정확도를 향상시키기 위해 이용될 수 있다. 음성인식 시스템은, 각각이 다른 주제나 카테고리의 단어들에 대하여 훈련된(trained) 다수의 언어 모델들(language models)을 포함할 수 있다. 인코딩된 음성을 포함하는 오디오 데이터에 액세스할 때, 음성인식도 또한 그 음성과 연관된 도킹 컨텍스트를 나타내는 정보에 액세스할 수 있다. 도킹 컨텍스트는, 예를 들어, 그 음성이 탐지된 시점에, 그 음성을 탐지한 디바이스의 도킹 컨텍스트를 포함할 수 있다. 음성인식 시스템은 그 도킹 컨텍스트에 입력된 음성을 인식하기 위해 사용하는 특정 언어 모델을 선택하기 위해 도킹 컨텍스트를 사용할 수 있다.

[0004] 많은 경우에, 디바이스의 도킹 컨텍스트는, 디바이스가 그 도킹 컨텍스트에 있는 동안 디바이스의 사용자가 이야기 할 것 같은 음성의 타입을 나타낼 수 있다. 예를 들어, 차량 도킹 스테이션(docking station)에 연결된 클라이언트 디바이스에 이야기하고 있는 사용자는 네비게이션이나 주소에 관한 단어들을 사용할 개연성이 있다. 차량 도킹 스테이션의 디바이스 상에 음성이 입력되면, 음성인식 시스템은 네비게이션-관련 단어들에 대하여 훈련된 언어 모델을 선택하고 그 음성을 인식하기 위해 사용할 수 있다. 도킹 컨텍스트에 기초하여 특정 언어 모델을 선택함으로써, 음성인식 시스템은 음성인식 프로세스에게 그 도킹 컨텍스트에서 가장 발화될 개연성이 높은 단어들에 대한 성향을 갖도록(bias) 할 수 있다. 결과적으로, 도킹 컨텍스트에 기초하여 선택된 언어 모델을 사용한 음성인식은 일반화된 언어 모델을 사용한 음성인식보다 더 정확한 표기(transcriptions)를 산출할 수 있다.

과제의 해결 수단

[0005] 일반적인 측면에서, 컴퓨터로 구현되는 방법은 인코딩된 음성을 포함하는 오디오 데이터에 액세스하는 단계; 클라이언트 디바이스의 도킹 컨텍스트를 나타내는 정보에 액세스하는 단계 (상기 도킹 컨텍스트는 상기 오디오 데이터와 관련됨); 복수의 언어 모델들을 식별하는 단계; 상기 도킹 컨텍스트에 기초하여 상기 복수의 언어 모델들 중에서 적어도 하나를 선택하는 단계; 및 상기 오디오 데이터의 일부분에 대한 표기를 식별하기 위해 상기 선택된 언어 모델을 이용하여 상기 오디오 데이터에 대한 음성 인식을 수행하는 단계를 포함한다.

[0006] 실시예들은 하나 이상의 다음 특징들을 포함할 수 있다. 예를 들어, 상기 클라이언트 디바이스의 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스와 상기 클라이언트 디바이스가 물리적으로 연결된 제2 디바이스 사이의 연결을 나타낸다. 상기 클라이언트 디바이스의 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스와 상기 클라이언트 디바이스가 무선으로 연결된 제2 디바이스 사이의 연결을 나타낸다. 상기 방법은, 상기 복수의 언어 모델들 각각에 대하여, 상기 도킹 컨텍스트에 기초하여 상기 언어 모델에 할당하기 위한 가중치를 결정하는 단계(상기 가중치는 상기 언어 모델이 상기 인코딩된 음성의 정확한 표기를 나타낼 확률을 나타냄)를 더 포함하고, 상기 도킹 컨텍스트에 기초하여 상기 복수의 언어 모델들 중에서 적어도 하나를 선택하는 단계는, 상기 할당된 가중치에 기초하여 상기 복수개의 언어 모델들 중 적어도 하나를 선택하는 단계를 포함한다. 상기 오디오 데이터로 인코딩된 음성은 상기 클라이언트 디바이스에 의해 탐지되고, 상기 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스에 의해 상기 오디오 데이터로 인코딩된 음성이 탐지되는 동안 상기 클라이언트 디바이스가 도킹 스테이션에 연결되었는지 여부를 나타낸다. 상기 오디오 데이터로 인코딩된 음성은 상기 클라이언트 디바이스에 의해 탐지되고, 상기 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스에 의해 상기 오디오 데이터로 인코딩된 음성이 탐지되는 동안 상기 클라이언트 디바이스가 연결된 도킹 스테이션의 타입을 나타낸다. 상기 인코딩된 음성은 하나 이상의 발화된(spoken) 쿼리 용어들을 포함하고, 상기 표기는 상기 발화된 쿼리 용어들의 표기를 포함하고, 상기 방법은, 검색 엔진이 상기 하나 이상의 발화된 쿼리 용어들의 상기 표기를 이용하여 검색을 수행하도록 하는 단계; 및 상기 검색의 결과들을 나타내는 정보를 상기 클라이언트 디바이스로 제공하는 단계를 더 포함한다. 상기 복수의 언어 모델들 각각에 대하여 가중치를 결정하는 단계는, 상기 도킹 컨텍스트와 관련된 저장된 가중치들에 액세스하는 단계를 포함한다. 상기 복수의 언어 모델들 각각에 대하여 가중치를 결정하는 단계는, 저장된 가중치들에 액세스하는 단계 및 상기 도킹 컨텍스트에 기초하여 상기 저장된 가중치를 변경하는 단계를 포함한다. 상기 복수의 언어 모델들 각각은 특정 주제 카테고리의 단어들에 대해 훈련된다. 상기 도킹 컨텍스트에 기초하여 가중치를 결정하는 단계는, 상기 클라이언트 디바이스가 차량 도킹 스테이션에 연결되어 있는지를 결정하는 단계; 및 주소를 출력하도록 훈련된 네비게이션 언어 모델에 대해서, 상기 네비게이션 언어 모델이 선택되는 확률을 상기 복수의 언어 모델들 내의 다른 언어 모델들에 대하여

증가시키는 가중치를 결정하는 단계를 포함한다.

[0007] 또 다른 일반적인 측면에서, 컴퓨터로 구현되는 방법은 클라이언트 디바이스에서 음성을 포함하는 오디오를 탐지하는 단계; 상기 탐지된 오디오를 오디오 데이터로 인코딩하는 단계; 상기 오디오 데이터를 서버 시스템으로 전송하는 단계; 상기 클라이언트 디바이스의 도킹 컨텍스트를 식별하는 단계; 상기 도킹 컨텍스트를 나타내는 정보를 상기 서버 시스템으로 전송하는 단계; 및 상기 오디오 데이터의 적어도 일부의 표기를 상기 클라이언트 디바이스에서 수신하는 단계를 포함하고, 상기 서버 시스템은, 상기 도킹 컨텍스트를 나타내는 정보에 기초하여 복수의 언어 모델들로부터 언어 모델을 선택하고, 상기 선택된 언어 모델을 이용하여 상기 오디오 데이터에 대한 음성 인식을 수행하여 상기 표기를 생성하고, 상기 표기를 상기 클라이언트 디바이스로 전송하는 것을 포함한다.

[0008] 실시예들은 하나 이상의 다음 특징들을 포함할 수 있다. 예를 들어, 상기 식별된 도킹 컨텍스트는 상기 오디오가 탐지된 시점의 상기 클라이언트 디바이스의 도킹 컨텍스트이다. 상기 클라이언트 디바이스의 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스와 상기 클라이언트 디바이스가 물리적으로 연결된 제2 디바이스 사이의 연결을 나타낸다. 상기 클라이언트 디바이스의 도킹 컨텍스트를 나타내는 정보는 상기 클라이언트 디바이스와 상기 클라이언트 디바이스가 무선으로 연결된 제2 디바이스 사이의 연결을 나타낸다.

[0009] 이들 측면의 다른 실시예들은 대응하는 시스템, 장치, 및 상기 방법들의 액션을 수행하도록 구성되고 컴퓨터 저장 디바이스에 인코딩된 컴퓨터 프로그램을 포함한다.

발명의 효과

[0010] 하나 이상의 실시예들의 세부 내용은 첨부된 도면과 이하의 상세한 설명에서 개시된다. 다른 특징과 이점은 상세한 설명, 도면, 및 청구항으로부터 명백해질 것이다.

도면의 간단한 설명

[0011] 도 1은 클라이언트 디바이스의 도킹 컨텍스트를 이용하여 음성인식을 수행하기 위한 시스템의 예시를 나타내는 다이어그램이다.

도 2a는 언어 모델의 예시 표현을 나타내는 다이어그램이다.

도 2b는 도 2a에 도시된 언어 모델과 함께 어쿠스틱 모델의 사용 예시를 나타내는 다이어그램이다.

도 3은 클라이언트 디바이스의 도킹 컨텍스트를 이용하여 음성인식을 수행하기 위한 예시 프로세스를 나타내는 흐름도이다.

도 4는 컴퓨팅 디바이스들의 블록 다이어그램이다.

발명을 실시하기 위한 구체적인 내용

[0012] 도 1은 클라이언트 디바이스(102)의 도킹 컨텍스트를 이용하여 음성인식을 수행하기 위한 시스템(100)의 예시를 나타내는 다이어그램이다. 시스템(100)은 클라이언트 통신 디바이스("클라이언트 디바이스")(102), 음성인식 시스템(104)(예를 들어, ASR 엔진(Automated Speech Recognition engine)), 및 검색 엔진 시스템(109)을 포함한다. 클라이언트 디바이스(102), 음성인식 시스템(104), 및 검색엔진 시스템(109)은 하나 이상의 네트워크(108)를 통해 서로 통신한다. 도 1은 또한 상태(A)부터 (G) 동안의 데이터 흐름을 나타낸다.

[0013] 클라이언트 디바이스(102)는 셀룰러폰이나 스마트폰과 같은 모바일 디바이스일 수 있다. 클라이언트 디바이스(102)의 다른 예시들은, GPS 네비게이션 시스템, 태블릿 컴퓨터, 노트북 컴퓨터, 및 데스크톱 컴퓨터를 포함한다.

[0014] 클라이언트 디바이스(102)는 도킹 스테이션(106)에 연결될 수 있다. 도킹 스테이션(106)은 클라이언트 디바이스(102)와 물리적으로 연계될 수 있고 클라이언트 디바이스(102)와 유선 또는 무선 링크를 통해, 예를 들어, 전력 및/또는 데이터를 전송하기 위해 통신할 수 있다. 도킹 스테이션(106)은 클라이언트 디바이스(102)가 도킹 스테이션(106)과 통신하는 동안 클라이언트 디바이스(102)를 (예를 들어 크래들(cradle) 또는 홀스터(holster)로) 물리적으로 잡고 있거나 안정되게 할 수 있다. 클라이언트 디바이스(102)는 도킹 스테이션(106)에 직접 연결되

거나 케이블이나 다른 인터페이스를 통해 연결될 수 있다.

- [0015] 상태 (A)동안, 클라이언트 디바이스(102)의 사용자(101)는 하나 이상의 용어들(terms)을 클라이언트 디바이스(102)의 마이크로폰에 이야기한다. 도시된 예시에서, 사용자(101)는 용어(103) ("10 Main Street")을 검색 쿼리의 일부로서 이야기한다. 발화된 용어(spoken terms)(103)에 대응하는 발음(utterances)이 오디오 데이터(105)로 인코딩된다. 용어(103)는, 예를 들어 클라이언트 디바이스(102) 상에 디스플레이 된 검색 인터페이스나 클라이언트 디바이스(102)의 사용자 인터페이스 상에서 선택된 검색 제어에 기초하여 쿼리 용어로서 식별될 수 있다.
- [0016] 클라이언트 디바이스(102)는 또한 도킹 컨텍스트, 예를 들어 사용자(101)가 용어들(103)을 이야기 할 때 클라이언트 디바이스(102)의 도킹 컨텍스트를 식별할 수 있다. 도시된 예시에서, 클라이언트 디바이스(102)는 사용자(101)가 용어들(103)을 이야기 할 때 차량 도킹 스테이션(106)에 연결되어 있다. 클라이언트 디바이스(102)는, 예를 들어, 클라이언트 디바이스(102)가 도킹 스테이션(106)에 연결되어 있다(예를 들어, 클라이언트 디바이스(102)가 현재 "도킹되어(docked)" 있다)는 것, 도킹 스테이션(106)이 차량 도킹 스테이션이라는 것, 및 도킹 스테이션(106)이 켜져 있다(powered on)는 것을 결정한다.
- [0017] 도킹 컨텍스트는 용어들(103)이 발화되고 있는 컨텍스트일 수 있다. 예를 들어, 도킹 컨텍스트는 클라이언트 디바이스(102)에 의해 발화된 쿼리 용어들(103)에 대응하는 오디오가 탐지된 시점의 클라이언트 디바이스(102)의 상태를 포함할 수 있다. 음성을 탐지하는 것은, 음성을 센싱(sensing), 수신, 또는 녹음하는 것을 포함하지만 이에 제한되는 것은 아니다. 음성을 탐지하는 것은, 비록 일부 실시예에서 발생할 수 있다 하더라도, 수신된 오디오가 음성을 포함하는지를 판단하거나 인코딩된 음성을 포함하는 오디오의 일부분을 식별하는 것을 요구하지 않을 수 있다.
- [0018] 도킹 컨텍스트는 클라이언트 디바이스(102)가 연결된 도킹 스테이션의 아이덴티티(identity)와 특성(characteristics)을 포함할 수 있다. 예를 들어, 도킹 컨텍스트는 (i) 클라이언트 디바이스(102)가 어떤 도킹 스테이션(106)에 연결되었는지 여부, (ii)클라이언트 디바이스(102)가 연결된 도킹 스테이션(106)의 타입(예를 들어, 차량 도킹 스테이션, 컴퓨터, 또는 음악 플레이어), (iii) 도킹 스테이션(106)의 작동 상태(예를 들어, 도킹 스테이션(106)이 온(on), 오프(off), 유희(idle), 또는 절전모드(power saving mode)인지 여부), (iv) 클라이언트 디바이스(102)와 도킹 스테이션(106) 사이의 관계(예를 들어, 클라이언트 디바이스(102)가 충전 중, 정보를 다운로드 중, 정보를 업로드 중, 또는 미디어를 재생 중, 또는 연결이 유희 상태) 중 하나 이상을 포함할 수 있다.
- [0019] 도킹 컨텍스트는 또한, 클라이언트 디바이스(102)와 도킹 스테이션(106)이 연결된 시간의 길이와 같은, 클라이언트 디바이스(102)와 도킹 스테이션(106) 사이의 연결과 관련된 다른 요소들을 포함할 수도 있다. 도킹 컨텍스트는 도킹 스테이션(106)의 하나 이상의 능력(예를 들어, GPS 수신기, 시각적 디스플레이, 오디오 출력, 및 네트워크 액세스)을 포함할 수 있다. 도킹 컨텍스트는 도킹 스테이션(106)의 모델, 제조사, 소프트웨어 버전을 나타내는 하나 이상의 식별자 또한 포함할 수 있다. 도킹 컨텍스트는 클라이언트 디바이스(102)에 연결된 주변 디바이스들(예를 들어, 프린터, 외부 저장 디바이스, 이미징 디바이스)을 포함하는 다수의 디바이스들에 대하여 상술한 요소들을 포함할 수 있다.
- [0020] 일부 실시예에서, 도킹 컨텍스트는 클라이언트 디바이스와, 예를 들어 케이블이나 물리적인 직접 연결을 통해 물리적으로 연결된 도킹 스테이션에 대한 정보를 나타낸다. 일부 실시예에서, 도킹 컨텍스트는 클라이언트 디바이스(102)에 지리적으로 근접해 있는 것으로 판단되고 블루투스과 같은 무선 프로토콜을 통해 연결된 도킹 스테이션(106)을 나타낼 수 있다. 예를 들어, 클라이언트 디바이스(102)가 차량일 때, 클라이언트 디바이스(102)는 차량에 물리적으로 연결된 도킹 스테이션(106)으로 무선으로 연결할 수 있다. 만약 클라이언트 디바이스(102)가 차량 도킹 스테이션(106)에 물리적으로 연결되지 않았다 하더라도, 무선 연결은 도킹 컨텍스트에 포함될 수 있다. 또 다른 예시로서, 도킹 컨텍스트는, 무선으로 연결된 수화기(earpiece)와 같이, 클라이언트 디바이스(102)와 통신하는 하나 이상의 다른 디바이스들을 포함할 수 있다. 도킹 컨텍스트는 클라이언트 디바이스(102)와 통신하는 어떤 디바이스라도 포함할 수 있다.
- [0021] 클라이언트 디바이스(102)는 도킹 컨텍스트의 하나 이상의 특징(aspect)을 나타내는 도킹 컨텍스트 정보(107)를 생성한다. 도킹 컨텍스트 정보(107)는 오디오 데이터(105)와 연관된다. 예를 들어, 도킹 컨텍스트 정보(107)는 클라이언트 디바이스(102)에 의해 오디오 데이터(105)에 인코딩된 음성이 탐지되는 클라이언트 디바이스(102)의 도킹 컨텍스트를 나타낼 수 있다. 클라이언트 디바이스(102) 또는 또 다른 시스템은 오디오 데이터(105)와 연계하여 도킹 컨텍스트 정보(107)를 저장할 수 있다.

- [0022] 상태 (B) 동안, 음성인식 시스템(104)는 도킹 컨텍스트 정보(107)에 액세스한다. 음성인식 시스템(104)은 또한 오디오 데이터(105)에도 액세스한다. 예를 들어, 클라이언트 디바이스(102)는 도킹 컨텍스트 정보(107)와 오디오 데이터(105)를 음성인식 시스템(104)으로 전송할 수 있다. 부가적으로, 또는 대안적으로, 도킹 컨텍스트 정보(107), 오디오 데이터(102), 또는 두 가지 모두 음성인식 시스템(104)에 연결된 저장 디바이스나 또 다른 시스템으로부터 액세스될 수 있다.
- [0023] 일부 실시예에서, 도킹 컨텍스트 정보(107)는 오디오 데이터(105), 또는 심지어 오디오 데이터(105)에 인코딩된 용어들(103)이 발화되기 전에 액세스될 수 있다. 예를 들어, 클라이언트 디바이스(102)는 클라이언트 디바이스(102)의 도킹 컨텍스트가 변경될 때 업데이트된 도킹 컨텍스트 정보(107)를 음성인식 시스템(104)로 제공하도록 구성될 수 있다. 결과적으로, 가장 최근에 수신된 도킹 컨텍스트 정보(107)가 현재의 도킹 컨텍스트를 나타내는 것으로 추정된다. 음성인식 시스템(104)은 음성 시퀀스에서 첫 번째 단어를 인식하기 위해 사용하는 언어 모델을 선택하기 위해 도킹 컨텍스트 정보(107)를 사용할 수 있다. 일부 실시예에서, 음성인식 시스템(104)은 심지어 사용자(101)가 말하기 전에도 도킹 컨텍스트 정보(107)에 기초하여 언어 모델을 선택할 수 있다.
- [0024] 상태 (C) 동안, 음성인식 시스템(104)은 다수의 언어 모델들(111a-111d)을 식별한다. 언어 모델들(111a-111d)은, 예를 들어, 시퀀스 내의 다른 용어들에 기초하여 용어들의 시퀀스 내의 한 용어의 발생 확률을 나타낼 수 있다. 언어 모델들과 그것들이 어떻게 사용될 수 있는지는 도 2a와 2b를 참조하여 자세하게 설명된다.
- [0025] 언어 모델들(111a-111d)은 각기 별도로 특정 주제(예를 들어, 네비게이션 또는 쇼핑)나 용어들의 타입(예를 들어, 이름 또는 주소들)에 대해 집중될 수 있다. 일부 예시에서, 언어 모델들(111a-111d)은 특정 액션(예를 들어, 음성 다이얼 또는 미디어 재생) 또는 특정 도킹 컨텍스트(예를 들어, 도킹되지 않음(undocked), 차량 도킹 스테이션에 연결됨, 또는 미디어 도킹 스테이션에 연결됨)에 특화될 수 있다. 결과적으로, 언어 모델들(111a-111d)은 일반-목적의 언어 모델(general-purpose language model)에 포함된 단어의 서브세트(subset)를 포함할 수 있다. 예를 들어, 네비게이션 용 언어 모델(111a)은 네비게이션에 사용되는, 숫자와 주소들과 같은 용어들을 포함할 수 있다.
- [0026] 음성인식 시스템은 도시된 것보다 더욱 세세한(fine-grained) 언어 모델들을 식별할 수 있다. 예를 들어, 미디어용 단일 언어 모델(111d) 대신에, 음성인식 시스템(104)은 비디오, 오디오, 또는 이미지와 관련된 별개의 언어 모델들(또는 언어 모델(111d)의 부분들)을 식별할 수 있다.
- [0027] 일부 실시예에서, 식별된 언어 모델들(111a-111d)은 더 크고 일반적인 언어 모델에 포함된 하위모델(submodels)일 수 있다. 일반적인 언어 모델은 특정 타입 단어들의 정확한 예측을 위해 특별하게 훈련된 몇 개의 언어 모델들을 포함할 수 있다. 예를 들어, 하나의 언어 모델은 이름을 예측하기 위해, 다른 것은 숫자를 예측하기 위해, 또 다른 것은 주소를 예측하는 등을 위해 훈련될 수 있다.
- [0028] 음성인식 시스템(104)은 도킹 컨텍스트 정보(107)에 나타난 도킹 컨텍스트와 연관된 언어 모델들(111a-111d)을 식별할 수 있다. 예를 들어, 음성인식 시스템(104)은 사용자(101)에 의해 발화된 용어들(103)을 매칭하는 적어도 임계 확률을 갖는 언어 모델들(111a-111d)을 식별할 수 있다. 또 다른 예시로서, 언어 모델들(111a-111d)의 특정 세트는 특정 도킹 컨텍스트에 대응하도록 미리 결정될 수 있다.
- [0029] 부가적으로, 또는 대안적으로, 음성인식 시스템(104)은 이전에 인식된 음성에 기초하여 언어 모델들(111a-111d)을 식별할 수 있다. 예를 들어, 음성 인식 시스템(104)은, 이전에 인식된 단어 "play"에 기초하여 게임용 언어 모델 및 미디어 용 언어 모델이 시퀀스에서 다음에 오는 용어들을 매칭시킬 개연성이 가장 높다고 판단할 수 있다. 결과적으로, 음성인식 시스템(104)은 게임용 언어 모델과 미디어용 언어 모델을 오디오 데이터(105)에 인코딩된 음성을 인식하기 위해 사용될 수 있는 언어 모델들로 식별할 수 있다.
- [0030] 상태 (D) 동안, 음성인식 시스템(104)은 도킹 컨텍스트 정보(107)에 나타난 도킹 컨텍스트에 기초하여 각각의 식별된 언어 모델들(111a-111d)에 대한 가중치(weighting values)를 결정할 수 있다. 일부 실시예에서, 각각의 식별된 언어 모델들(111a-111d)에 대한 가중치는 또한 음성 시퀀스에서 이미 인식된 용어들에 기초한 언어 모델로부터의 출력과 같은, 다른 정보에 기초할 수 있다. 결정되는 가중치들은 각각의 언어 모델들(111a-111d)로 할당된다.
- [0031] 가중치들은 사용자(101)에 의해 발화된 용어들(103)이 각각의 언어 모델들(111a-111d)에 포함된 용어들의 타입들과 매칭되고, 그에 따라 언어 모델들(111a-111d)이 용어들(103)의 정확한 표기를 나타낼 확률을 나타낸다. 예를 들어, 네비게이션 언어 모델(111a)에 할당된 가중치는 오디오 데이터(105)에 인코딩된 음성이 네비게이션 용

어들을 포함하는 확률을 나타낼 수 있다. 웹 검색 언어 모델(111b)에 할당된 가중치는 오디오 데이터에 인코딩된 음성이 웹 검색에서 일반적으로 사용되는 흔한 용어들을 포함하는 확률을 나타낼 수 있다.

[0032] 일부 예시들에서, 음성인식 시스템(104)은 다수의 저장된 가중치 세트들(112, 113, 114, 115)에서 선택할 수 있다. 각각의 가중치 세트(112, 113, 114, 115)는 특정 도킹 컨텍스트에 대응할 수 있다. 도시된 예시에서, 가중치 세트(113)는 차량 도킹 스테이션(106)에 대응한다. 도킹 컨텍스트 정보(107)가 클라이언트 디바이스(102)가 차량 도킹 스테이션(106)에 연결되어 있다는 것을 나타내기 때문에, 음성인식 시스템은 차량 도킹 스테이션(106)에 대응하는 가중치 세트(113)를 선택한다. 세트(113) 내의 가중치들은 각 언어 모델들(111a-111d)에 할당된다.

[0033] 다양한 세트(112, 113, 114, 115)에서 가중치들은, 예를 들어, 다양한 도킹 컨텍스트에서 다양한 사용자들에 의해 발화되는 많은 수의 용어들에 대한 통계적 분석을 수행함으로써 결정될 수 있다. 특정 도킹 컨텍스트가 주어진 특정 언어 모델에 대한 가중치는 그 언어 모델이 그 도킹 컨텍스트에서 산출하는 정확한 결과들의 관측된 빈도에 기초할 수 있다. 예를 들어, 만약 네비게이션 언어 모델(111a)이 클라이언트 디바이스(102)가 차량 도킹 스테이션에 있을 때 발생하는 음성의 50%에 대해 정확하게 발음을 예측하면, 세트(113)에서 네비게이션 언어 모델(111a)에 대한 가중치는 0.5가 될 수 있다. 어떻게 언어 모델이 음성을 예측하는지에 대한 예시가 도 2a와 도 2b를 참조하여 이하에서 설명된다.

[0034] 일부 실시예에서, 음성인식 시스템(104)은 가중치의 초기 세트(initial set)를 조절하여 언어 모델들(111a-111d)에 대한 가중치를 결정할 수 있다. 예를 들어, 가중치 세트(112)는 도킹 컨텍스트 정보(107)가 클라이언트 디바이스(102)가 도킹되지 않았다는 것을 나타낼 때, 또는 클라이언트 디바이스(102)의 도킹 컨텍스트가 알려지지 않았을 때 사용될 수 있다. 도킹 컨텍스트 정보(107)가 클라이언트 디바이스(102)가 도킹되었다는 것을 나타낼 때, 세트(112)의 개별 가중치들은 도킹 컨텍스트의 다양한 특성에 기초하여 변경될 수 있다. 가중치들은 공식, 룩업 테이블(look-up tables), 및 다른 방법들을 이용하여 결정될 수 있다. 일부 실시예에서, 음성인식 시스템(104)은 저장된 가중치 세트들 사이에서 각 핵심 구문(key phrase)에 대응하는 것을 선택하기 위해 도킹 컨텍스트를 사용할 수 있다. 가중치 세트들(112, 113, 114, 115)은 단일 도킹 컨텍스트와 직접적으로 연관될 것이 요구되지 않는다. 예를 들어, 세트(112)는 핵심 구문 "navigate to"와 연관될 수 있다. 사용자(101)가 용어 "navigate to"를 이야기할 때, 세트(112)가 도킹 컨텍스트가 알려져 있거나 그렇지 않은 경우에도 선택된다. 또한, 클라이언트 디바이스(102)가 차량 도킹 스테이션(106)에 있는 것으로 알려져 있을 때, 세트(112)는, 사용자가 핵심 구문 "navigate to"를 이야기 했던 것처럼, 심지어 사용자(101)가 그 핵심 구문을 이야기하지 않았더라도, 선택될 수 있다.

[0035] 도킹 컨텍스트는, 시작 상태부터 하나 이상의 핵심 구문들과 연관된 상태까지, 또는 핵심 구문과 연관된 가중치로부터 특정 언어 모델의 선택까지와 같이, 다양한 결정 및 언어 모델을 선택하기 위해 궁극적으로 사용되는 가중치의 타입에 영향을 미칠 수 있다. 도킹 컨텍스트는 핵심 구문에 대응하는 하나 이상의 상태들을 선택하기 위해 사용되는 가중치들을 결정하기 위해 사용될 수 있고, 상기 핵심 구문들에 대응하는 상태들은 언어 모델들(111a-111d)에 대한 가중치와 적절히 연관될 수 있다. 예를 들어, 차량 도킹 컨텍스트는 구문 "navigate to"에 대응하는 상태에 대해서는 "0.6"의 가중치를, 구문 "call"에 대응하는 상태에 대해서는 "0.4"의 가중치를 결정하기 위해 사용될 수 있다. 각 핵심 구문 상태(key phrase state)는 그 상태에서 다양한 언어 모델들의 가능성을 나타내는 가중치 세트와 연관될 수 있다.

[0036] 핵심 구문에 대응하는 상태가 선택된 이후에도, 다양한 언어 모델들(111a-111d)의 확률을 나타내는 가중치 세트가 선택되고, 도킹 컨텍스트는 가중치를 수정하기 위해 사용될 수 있다. 예를 들어, 구문 "navigate to"와 연관된 상태는 네비게이션 언어 모델이 비즈니스 언어 모델보다 2배의 개연성을 나타내는 가중치를 포함할 수 있다. 도킹 컨텍스트는 현재 구술(dictation)의 인식을 위해, 네비게이션 언어 모델이 비즈니스 언어 모델보다 3배의 개연성이 있도록 가중치를 수정하기 위해 사용될 수 있다.

[0037] 상태 (E) 동안, 음성인식 시스템(104)은 할당된 가중치에 기초하여 언어 모델을 선택한다. 표(116)에 도시된 바와 같이, 세트(113)에서 가중치(113a-113d)가 언어 모델들(111a-111d)로 할당된다. 이 가중치들(113a-113d)은, 도킹 컨텍스트 정보(107)에 의해 나타나는 도킹 컨텍스트에 기초하여, 대응하는 언어 모델들(111a-111d)이 사용자(101)에 의해 발화된 용어들(103)과 매칭되는 확률을 나타낸다. 네비게이션 용어 언어 모델(111a)은 가장 높은 가중치(113a)를 포함하고, 이는 도킹 컨텍스트에 기초하여, 언어 모델(111a)이 오디오 데이터(105)에 인코딩된 용어들(103)의 내용을 가장 정확하게 예측할 것을 나타낸다. 가중치에 기초하여, 음성인식 시스템(104)은 오디오 데이터(105)의 음성 인식에 사용하기 위해 언어 모델(111a)을 선택한다.

- [0038] 일부 실시예에서, 단일 언어 모델(111a)가 가중치(113a-113d)에 기초하여 선택된다. 일부 실시예에서, 다수의 언어 모델들(111a-111d)이 가중치(113a-113d)에 기초하여 선택될 수 있다. 예를 들어, 상위 N개의 언어 모델들(111a-111d)를 포함하는 서브세트가 선택될 수 있고 오디오 데이터(105)에 대한 표기 후보(candidate transcriptions)를 식별하기 위해 나중에 사용될 수 있다.
- [0039] 음성인식 시스템(104)은 또한 가중치를 다른 요소들과 결합하여 이용하여 언어 모델을 선택할 수도 있다. 예를 들어, 음성인식 시스템(104)은 가중치(113a-113d)와, 음성 시퀀스에서 인식된 이전 단어들에 기초하거나 이전 표기들에 기초한 가중치와 같은 다른 가중치들의 가중치 결합(weighted combination)을 결정할 수 있다.
- [0040] 일 예시로서, 음성인식 시스템(104)은 시퀀스에서 첫 번째 용어를 "play"로 표기할 수 있다. 도킹 컨텍스트에만 기초한 가중치는, 후속 음성을 인식하기 위해서는 네비게이션 언어 모델 또는 미디어 언어 모델이 사용되어야 한다고 나타낼 수 있다. (첫 번째 용어 "play"를 인식하기 위해 이전에 사용되었던 언어 모델의 출력과 같은) 다른 정보에 기초한 제2 가중치 세트는 게임 언어 모델이나 미디어 언어 모델이 사용되어야 한다고 나타낼 수 있다. 양쪽 가중치 세트를 고려하면, 음성인식 시스템(104)은 시퀀스에서 다음 용어의 정확한 표기를 산출할 개연성이 가장 높은 것으로 미디어 언어 모델을 선택할 수 있다. 이 예시에서 설명한 바와 같이, 일부 경우에, 도킹 컨텍스트가 각 용어에 대해 동일할 수 있다 하더라도, 시퀀스 내의 다른 용어들을 인식하기 위해 다른 언어 모델들이 사용될 수 있다.
- [0041] 상태 (F) 동안, 음성인식 시스템(104)은 선택된 언어 모델(111a)을 이용하여 오디오 데이터(105)에 대한 음성인식을 수행한다. 음성인식 시스템(104)은 오디오 데이터(105)의 적어도 일부분에 대한 표기를 식별한다. 음성인식 시스템(104)은 일반적인 언어 모델보다 선택된 언어 모델(111a)를 이용하여 용어들(103)을 더 정확하게 인식할 수 있다. 이는 도킹 컨텍스트가 오디오 데이터(105)에 인코딩 될 가장 높은 개연성의 용어들의 타입을 나타내고, 선택된 언어 모델(111a)은 이러한 용어들을 가장 잘 예측하도록 선택되었기 때문이다.
- [0042] 선택된 언어 모델(111a)을 이용함으로써, 음성인식 시스템(104)은 용어(103)에 대해 가능한 표기의 범위를 선택된 언어 모델(111a)에 의해 지시되는 것으로 좁힐 수 있다. 이는, 특히 구문에서 첫 번째 단어에 대하여, 음성인식을 실질적으로 향상시킬 수 있다. 일반적으로, 음성 시퀀스의 시작 시점에 발행할 수 있는 매우 큰 용어들의 세트가 있다. 시퀀스의 첫 번째 용어에 대해서, 음성인식 시스템은 다음에 나올 것 같은 용어들을 나타내는 시퀀스 내의 이전 단어들의 이점을 갖지 못한다. 그럼에도 불구하고, 주제(예를 들어, "driving directions to" 또는 "show map at")를 나타내는 이전 단어의 부재에도, 도킹 컨텍스트에 기초하여 선택된, 선택된 언어 모델(111a)이 이미 용어들(103)의 그럴듯한 내용으로 조정되었기 때문에, 음성인식 시스템(104)은 여전히 용어들의 정확한 세트를 미리 인식할 수 있다. 도킹 컨텍스트에 기초하여 선택된 언어 모델을 이용하는 것은, 음성인식을 정확하게, 또는 심지어 사용자가 이전 구문(prefix phrase)에서 음성의 주제를 특정했던 것보다 더욱 정확하게 할 수 있다.
- [0043] 같은 이유로, 음성인식은, 음성인식을 안내할 단어들 사이의 연관성이 거의 없는, 단일 용어들 및 짧은 용어 시퀀스에 대해서도 향상될 수 있다. 검색 쿼리들은 종종 짧은 용어 시퀀스를 포함하기 때문에, 도킹 컨텍스트에 기초한 언어 모델을 사용하는 것은 이 적용에 있어서 정확성을 상당히 향상시킬 수 있다.
- [0044] 이 예시에서, 발화된 용어들(103)은 주소 "10 Main Street"를 포함하고, 용어들(103)이 주소를 포함한다는 것을 나타내는 발화된 이전 구문(예를 들어, "navigation to")은 없다. 여전히, 용어들(103)이 발화된 도킹 컨텍스트에 기초하여, 음성인식 시스템(104)은 주소에 특화된(예를 들어, 최적화되거나 특화된) 특화된 언어 모델(111a)를 선택할 수 있다. 이 언어 모델(111a)은 오디오 데이터(105)에 인코딩된 첫 번째 용어가 숫자일 것이고, 첫 번째 용어에는 거리 이름이 뒤따를 것이라는 높은 확률을 나타낼 수 있다. 선택된 언어 모델(111a)에 포함된 특화된 단어 및 패턴들은 오디오 데이터(105)의 음성인식 정확도를 향상시킬 수 있다. 예를 들어, 선택된 언어 모델(111a)의 포커스 바깥에 있는 용어들(예를 들어, 네비게이션과 관련 없는 용어들)은 언어모델로부터 배제될 수 있고, 따라서 그것들은 용어들(103)에 대한 가능한 표기로서 배제된다. 이와 대조적으로, 이러한 용어들은 일반적인 언어모델에서 유효한 표기 가능성으로 포함될 수 있고, 이는 유효한 가능성이 있는 것처럼 보이는, 하지만 사실상 현재 용어들(103)을 인식하는 데에는 관련 없는 많은 용어들을 포함한다.
- [0045] 선택된 언어 모델을 사용하여, 음성인식 시스템(104)은 오디오 데이터(105)에 대해 표기 "10 Main Street"를 선택한다. 상기 표기는 검색엔진 시스템(109)로 전달될 수 있다. 상기 표기는 또한, 사용자(101)가 표기의 정확성을 확인하고 필요한 경우 정정을 할 수 있도록 하기 위해 클라이언트 디바이스(102)로도 전송될 수 있다.
- [0046] 상태 (G) 동안, 검색엔진 시스템(109)은 발화된 쿼리 용어들(103)의 표기를 이용하여 검색을 수행한다. 이 검색

은 웹 검색, 네비게이션 안내 검색, 또는 다른 타입의 검색이 될 수 있다. 검색 쿼리의 결과를 나타내는 정보가 클라이언트 디바이스(102)로 전송된다. 상기 표기는 도킹 컨텍스트에 기초하여 선택된 특화된 언어 모델(111a)를 이용하여 결정된다. 따라서, 상기 표기가 사용자(101)에 의해 발화된 쿼리 용어들(103)과 매치될 가능성은 일반적인 언어 모델을 사용한 가능성보다 더 크다. 결과적으로, 상기 표기를 포함하는 검색 쿼리는 사용자(101)가 의도했던 검색일 개연성이 더 높다.

- [0047] 용어들(103)의 표기가 검색에 사용되는 것으로 기술되었지만, 표기의 다양한 다른 사용도 가능하다. 다른 실시예에서, 표기는, 예를 들어, 지도나 길안내(directions)를 검색하거나, 음악 또는 다른 미디어를 검색하고 재생하거나, 연락처를 식별하고 통신을 시작하거나, 애플리케이션을 선택하고 실행하거나, 문서를 위치시키고 오픈하거나, 모바일 디바이스(102)의 기능(예를 들어 카메라)을 활성화 시키는 등을 위해 사용될 수 있다. 각각의 이러한 사용에 대해서, 표기를 이용하여 검색된 정보는 하나 이상의 서버 시스템, 클라이언트 디바이스(102), 또는 도킹 스테이션(106)에 의해 식별될 수 있다.
- [0048] 일부 실시예에서, 오디오 데이터(105)의 서로 다른 부분의 음성을 인식하기 위해 서로 다른 언어 모델(111a-111d)이 선택되고 사용될 수 있다. 심지어 오디오 데이터(105)가 단일 도킹 컨텍스트와 연관될 때도, 다른 정보(예를 들어 시퀀스에서 인식된 다른 단어들)가 언어 모델(111a-111d)의 선택에 영향을 줄 수 있다. 결과적으로, 시퀀스에서 다른 용어들은 다른 언어 모델들(111a-111d)을 이용하여 인식될 수 있다.
- [0049] 도 2a는 언어 모델(200)의 예시 표현을 나타내는 다이어그램이다. 일반적으로, 음성인식 시스템은 음성을 포함하는 오디오 데이터를 수신하고 그 오디오 데이터에 가장 잘 매칭되는 하나 이상의 표기를 출력한다. 음성인식 시스템은 오디오 데이터로부터 하나 이상의 용어들을 인식하기 위한 다수의 기능을 동시에 또는 순차적으로 수행할 수 있다. 예를 들어, 음성인식 시스템은 어쿠스틱 모델(acoustic model) 및 언어 모델(200)을 포함할 수 있다. 언어 모델(200)과 어쿠스틱 모델은 오디오 데이터 내 음성의 하나 이상의 표기를 선택하기 위해 함께 사용될 수 있다.
- [0050] 어쿠스틱 모델은 오디오 데이터의 일부와 매칭되는 용어들을 식별하기 위해 사용될 수 있다. 오디오 데이터의 특정 부분에 대해서, 어쿠스틱 모델은 오디오 데이터의 다양한 특성과 매칭되는 용어들 및 각 용어가 오디오 데이터와 매칭되는 정도를 나타내는 가중치 또는 신뢰값(confidence score)을 출력할 수 있다.
- [0051] 언어 모델(200)은 음성 패턴 내의 용어들 사이의 관계에 관한 정보를 포함할 수 있다. 예를 들어, 언어 모델(200)은 흔히 사용되는 용어들의 시퀀스 및 문법 규칙과 다른 언어 관습을 따르는 시퀀스들에 대한 정보를 포함할 수 있다. 언어 모델(200)은 시퀀스 내의 하나 이상의 다른 용어들에 기초하여 음성 시퀀스에서 한 용어의 발생 가능성을 나타내기 위해 사용될 수 있다. 예를 들어, 언어 모델(200)은 시퀀스 내의 이전 단어들에 기초하여, 단어 시퀀스의 특정 부분에서 어느 단어가 가장 높은 발생 확률을 갖는지를 식별할 수 있다.
- [0052] 언어 모델(200)은 노드들(nodes)(201a-201i)과 노드들(201a-201i) 사이의 트랜지션(transitions)(202a-202h)의 세트를 포함한다. 각 노드(201a-201i)는 음성 시퀀스에서 단일 용어(예를 들어 한 단어)가 선택되는 결정 포인트(decision point)를 나타낸다. 노드(201a-201i)로부터의 각 트랜지션(202a-202h)은 시퀀스의 구성요소로 선택될 수 있는 용어와 연관된다. 각 트랜지션(202a-202h)은 또한, 예를 들어 트랜지션(202a-202h)과 연관된 용어가 시퀀스에서 그 시점에 발생하는 확률을 나타내는 가중치와 연관된다. 가중치는 시퀀스 내의 다수의 이전 용어들에 기초하여 설정된다. 예를 들어, 각 노드에서의 트랜지션 및 트랜지션에 대한 가중치는 음성 시퀀스 내의 그 노드 이전에 발생한 N개의 용어들에 의해 결정될 수 있다.
- [0053] 일 예시로서, 첫 번째 노드(201a)는 음성 시퀀스에서 첫 번째 용어가 선택되는 결정 포인트를 나타낸다. 노드(201a)에서의 유일한 트랜지션은 용어 "the"와 연관된 트랜지션(202a)이다. 트랜지션(202a)를 따르는 것은 음성 시퀀스에서 첫 번째 용어로서 용어 "the"를 선택하는 것을 의미하고, 이는 노드(201b)에서의 다음 결정으로 유도한다.
- [0054] 노드(201b)에서는 2가지 트랜지션이 가능하다: (1) 용어 "hat"과 연관되고 0.6의 가중치를 갖는 트랜지션(202b); 및 (2) 용어 "hats"와 연관되고 0.4의 가중치를 갖는 트랜지션(202c)이다. 트랜지션(202b)은 트랜지션(202c)보다 더 높은 가중치를 갖고, 이는 이 시점의 음성 시퀀스에서는 용어 "hat"이 용어 "hats"보다 발생할 개연성이 더 높다는 것을 나타낸다. 각 노드(201a-201i)에서 가장 높은 가중치를 갖는 트랜지션(202a-202h)을 선택함으로써, 가장 개연성이 높은 용어들의 시퀀스를 나타내는 경로(204)가 생성되고, 이 예시에서는 "the hat is black"이다.
- [0055] 언어 모델에서 트랜지션의 가중치들은 용어들의 유효한 시퀀스를 보여주는 예시 텍스트 코퍼스(a corpus of

example text)에서의 언어 패턴들에 기초하여 결정될 수 있다. 식별 훈련(discriminative training)과 같은 기계 학습 기술(machine learning techniques)은 히든 마르코프 모델(Hidden Markov Models ("HMMs"))을 이용하여 트랜지션의 확률을 설정하기 위해 이용될 수 있다. 문법 모델을 수동으로 특정하고 구축하기 위해 가중 유한 상태 변환기(weighted finite-state transducers)가 사용될 수 있다. 예시 구문의 코퍼스에서 n-그램의 발생을 카운트하고 그 카운트로부터의 트랜지션 확률을 유도하기 위해 n-그램 스무딩(N-gram smoothing)이 이용될 수 있다. 예시 텍스트의 코퍼스를 이용하여 HMMs에서의 확률을 설정하기 위해 Baum-Welch 알고리즘과 같은 기대값-최대화 기술(expectation-maximization techniques)이 이용될 수 있다.

[0056] 도 2b는 도 2a에 도시된 언어 모델과 함께 어쿠스틱 모델의 사용 예시를 나타내는 다이어그램이다. 언어 모델의 출력은 오디오 데이터에 대한 표기를 선택하기 위해 어쿠스틱 모델의 출력과 결합될 수 있다. 예를 들어, 도 2b는 단일 용어에 대응하는 오디오 데이터의 부분에 대한 어쿠스틱 모델 및 언어 모델로부터의 출력의 결합을 나타낸다. 특히, 도 2b는 도 2a에서 노드(201d)로부터의 트랜지션(202f-202h)에 의해 선택된 용어에 대응하는 오디오 데이터에 대한 출력을 나타낸다. 언어 모델은 용어들(212a-212c) 및 노드(201d)에서 가장 높게 가중된(highest-weighted) 트랜지션과 연관되는 대응 가중치(213a-213c)를 출력한다. 어쿠스틱 모델은 오디오 데이터와 가장 잘 매칭되는 용어들(216a-216c)을, 용어들(216a-216c)이 오디오 데이터와 매칭되는 정도를 나타내는 대응 가중치(217a-217c)와 함께 출력한다.

[0057] 가중치(213a-213c 및 217a-217c)는 결합 가중치(223a-223e)를 생성하기 위해 결합되고, 이는 용어들(222a-222e)의 결합된 세트를 서열화하기 위해 사용된다. 도시된 바와 같이, 어쿠스틱 모델과 언어 모델의 출력에 기초하여, 용어(222a) "black"이 가장 높은 결합 가중치(223a)를 찾고 따라서 오디오 데이터의 대응하는 부분에 대한 가장 개연성이 높은 표기이다. 비록 어쿠스틱 모델과 언어 모델에 의해 출력된 가중치(213a-213c, 217a-217c)가 결합 가중치(223a-223e)를 결정하는데 동일한 영향을 갖는 것으로 보여지지만, 가중치(213a-213c, 217a-217c)는 또한 동등하지 않게 결합될 수도 있고 다른 타입의 데이터와 결합될 수도 있다.

[0058] 도 3은 클라이언트 디바이스의 도킹 컨텍스트를 이용하여 음성인식을 수행하기 위한 예시 프로세스(300)을 나타내는 흐름도이다. 간략하게, 프로세스(300)은 인코딩된 음성을 포함하는 오디오 데이터에 접근하는 것을 포함한다. 클라이언트 디바이스의 코딩 컨텍스트를 나타내는 정보가 액세스된다. 다수의 언어 모델이 식별된다. 언어 모델들 중 적어도 하나가 도킹 컨텍스트에 기초하여 선택된다. 선택된 언어 모델을 이용하여 오디오 데이터에 대한 음성인식이 수행된다.

[0059] 보다 자세하게, 인코딩된 음성을 포함하는 오디오 데이터가 액세스된다(302). 오디오 데이터는 클라이언트 디바이스로부터 수신될 수 있다. 인코딩된 음성은 클라이언트 디바이스에 의해 녹음된 음성과 같은, 클라이언트 디바이스에 의해 탐지된 음성일 수 있다. 인코딩된 음성은 하나 이상의 발화된 쿼리 용어들을 포함할 수 있다.

[0060] 클라이언트 디바이스의 도킹 컨텍스트를 나타내는 정보가 액세스된다(304). 도킹 컨텍스트는 오디오 데이터와 연관될 수 있다. 도킹 컨텍스트를 나타내는 정보는 클라이언트 디바이스로부터 수신될 수 있다. 예를 들어, 도킹 컨텍스트를 나타내는 정보는, 오디오 데이터에 인코딩된 음성이 클라이언트 디바이스에 의해 탐지되는 동안, 클라이언트 디바이스가 도킹 스테이션에 연결되었는지 여부를 나타낼 수 있다. 도킹 컨텍스트를 나타내는 정보는 또한, 오디오 데이터에 인코딩된 음성이 클라이언트 디바이스에 의해 탐지되는 동안, 클라이언트 디바이스가 연결된 도킹 스테이션의 타입을 나타낼 수도 있다.

[0061] 도킹 컨텍스트를 나타내는 정보는 클라이언트 디바이스와 클라이언트 디바이스가 무선으로 연결된 제2 디바이스 사이의 연결을 나타낼 수 있다. 도킹 컨텍스트를 나타내는 정보는 클라이언트 디바이스와 그 클라이언트 디바이스가 물리적으로 연결된 제2 디바이스 사이의 연결을 나타낼 수 있다.

[0062] 다수의 언어 모델이 식별된다(306). 각각의 다수 언어 모델은 시퀀스 내의 다른 용어들에 기초하여 용어들의 시퀀스에서 하나의 용어의 발생 확률을 나타낼 수 있다. 각각의 다수 언어 모델은 특정 주제의 단어 카테고리들에 대해 훈련될 수 있다. 특정 단어 카테고리들은 각 언어 모델에 대해 서로 다를 수 있다. 하나 이상의 다수 언어 모델들은 언어 모델의 일부 또는 서브세트를 포함할 수 있다. 예를 들어, 하나 이상의 다수 언어 모델들은 또 다른 언어 모델의 하위모델일 수 있다.

[0063] 적어도 하나의 식별된 언어 모델이 도킹 컨텍스트에 기초하여 선택된다(308). 예를 들어, 각각의 식별된 언어 모델에 대한 가중치는 도킹 컨텍스트에 기초하여 결정될 수 있다. 가중치는 각 언어 모델로 할당될 수 있다. 각 가중치는 그것이 할당된 언어 모델이 인코딩된 음성의 정확한 표기를 나타낼 것인지의 확률을 나타낼 수 있다. 각각의 언어 모델에 대한 가중치를 결정하는 것은 도킹 컨텍스트와 연관되어 저장된 가중치에 액세스하는 것을

포함한다. 각각의 언어 모델에 대한 가중치를 결정하는 것은 저장된 가중치에 역세상하는 것과 도킹 컨텍스트에 기초하여 그 저장된 가중치를 변경하는 것을 포함한다.

- [0064] 도킹 컨텍스트에 기초하여 가중치를 결정하는 것은, 예를 들어, 클라이언트 디바이스가 차량 도킹 스테이션에 연결되어 있는 것을 결정하는 것과, 주소를 출력하도록 훈련된 네비게이션 언어 모델에 대해서, 네비게이션 언어 모델이 다른 식별된 언어 모델과 관련하여 선택되는 확률을 증가시키는 가중치를 결정하는 것을 포함할 수 있다.
- [0065] 음성인식이 선택된 언어 모델을 이용하여 오디오 데이터에 대해 수행된다(310). 오디오 데이터의 적어도 일부분에 대하여 표기가 식별된다. 예를 들어, 오디오 데이터로 인코딩된 하나 이상의 발화된 용어들에 대한 표기가 생성될 수 있다.
- [0066] 오디오 데이터에 인코딩된 음성은 발화된 쿼리 용어들을 포함할 수 있고, 오디오 데이터의 일부의 표기는 그 발화된 쿼리 용어들의 표기를 포함할 수 있다. 프로세스(300)는 검색엔진이 하나 이상의 발화된 쿼리 용어들의 표기를 이용하여 검색을 수행하도록 하는 것과 검색 쿼리의 결과를 식별하는 정보를 클라이언트 디바이스로 제공하는 것을 포함할 수 있다.
- [0067] 도 4는, 클라이언트나 서버 또는 복수개의 서버로서, 본 명세서에서 설명된 시스템 및 방법을 구현하기 위해 사용될 수 있는 컴퓨팅 디바이스(400, 450)의 블록 다이어그램이다. 컴퓨팅 디바이스(400)는 랩탑, 데스크톱, 워크스테이션, PDA, 서버, 블레이드(blade) 서버, 메인 프레임, 및 그 밖의 적절한 컴퓨터들과 같은 다양한 형태의 디지털 컴퓨터를 나타내기 위해 사용된다. 모바일 컴퓨팅 디바이스(450)는 PDA, 셀룰러 폰, 스마트폰, 및 그 밖의 유사한 컴퓨팅 디바이스와 같은 다양한 형태의 모바일 디바이스를 나타내기 위해 사용된다. 여기에 보여지는 컴포넌트들, 그 연결 및 관계, 및 그 기능들은 단지 예시를 의미하고, 본 명세서에서 설명하거나 또는 청구된 발명의 실시예를 제한하는 것을 의미하지 않는다.
- [0068] 컴퓨팅 디바이스(400)는 프로세서(402), 메모리(404), 저장 디바이스(406), 메모리(404)에 접속하는 고속 인터페이스(408)와 고속 확장 포트(410), 및 저속 버스(414)와 저장 디바이스(406)에 접속하는 저속 인터페이스(412)를 포함한다. 각 구성요소(402, 404, 406, 408, 410, 및 412)는 다양한 버스들을 사용하여 서로 접속되고, 일반적인 마더보드 또는 적절한 경우 다른 방식으로 탑재될 수 있다. 프로세서(402)는 컴퓨팅 디바이스(400) 내에서 실행하기 위한 명령어를 처리할 수 있으며, 이러한 명령어에는, 고속 인터페이스(408)에 연결된 디스플레이(416)와 같은 외장 입/출력 디바이스상에서 GUI용 그래픽 정보를 디스플레이 하기 위해, 메모리(404) 또는 저장 디바이스(406)에 저장되는 명령어가 포함된다. 다른 실시예에서, 다중 프로세서 및/또는 다중 버스는 적절한 경우, 다중 메모리 및 메모리 타입과 함께 사용될 수 있다. 또한, 다중 컴퓨팅 디바이스(400)는 각 디바이스가 필요 동작의 부분을 제공하는 형태(예를 들어, 서버 뱅크, 블레이드 서버의 그룹, 또는 다중 프로세서 시스템)로 접속될 수 있다.
- [0069] 메모리(404)는 컴퓨팅 디바이스(400) 내에 정보를 저장한다. 일 실시예에서, 메모리(404)는 비 휘발성 유닛 또는 유닛들이다. 또 다른 실시예에서, 메모리(404)는 휘발성 유닛 또는 유닛들이다. 메모리(404)는 또한, 자기 또는 광 디스크와 같은 또 다른 형태의 컴퓨터-판독가능 매체일 수 있다.
- [0070] 저장 디바이스(406)는 컴퓨팅 디바이스(400)를 위한 대용량 저장소(mass storage)를 제공할 수 있다. 일 실시예에서, 저장 디바이스(406)는 컴퓨터-판독 가능한 매체이다. 다양한 다른 실시예들에서, 저장 디바이스(406)는 플로피 디스크 디바이스, 하드 디스크 디바이스, 광 디스크 디바이스, 또는 테이프 디바이스, 플래시 메모리 또는 다른 유사한 고체 상태 메모리 디바이스, 또는 저장 영역 네트워크 또는 다른 구성에 존재하는 디바이스를 포함하는 디바이스 어레이일 수 있다. 일 실시예에서, 컴퓨터 프로그램 제품은 정보 캐리어(information carrier) 내에 유형적으로 구체화된다. 또한, 컴퓨터 프로그램 제품은 실행될 때, 상술한 것과 같은 하나 이상의 방법을 수행하는 명령어를 포함한다. 정보 캐리어는 메모리(404), 저장 디바이스(406), 또는 프로세서(402)상의 메모리와 같은 컴퓨터 또는 기계 판독가능 매체이다.
- [0071] 부가적으로, 컴퓨팅 디바이스(400 또는 450)는 USB 플래시 드라이브를 포함할 수 있다. USB 플래시 드라이브는 운영체제와 다른 애플리케이션을 저장할 수 있다. USB 플래시 드라이브는 무선 트랜스미터 또는 다른 컴퓨팅 디바이스의 USB 포트에 삽입될 수 있는 USB 커넥터와 같은 입/출력 컴포넌트를 포함할 수 있다.
- [0072] 고속 제어부(408)는 컴퓨팅 디바이스(900)에 대한 대역-집약적 동작을 관리하고, 반면에 저속 제어부(412)는 저대역-집약적 동작(lower bandwidth-intensive operations)을 관리한다. 이러한 기능(duties)들의 배치는 단지 예시적인 것이다. 일 실시예에서, 고속 제어부(408)는 메모리(404), 디스플레이(416)(예를 들어, 그래픽 프로세

서 또는 가속기를 통합)에 연결되고, 다양한 확장 카드(도시되지 않음)을 수용할 수 있는 고속 확장 포트(410)에 연결된다. 일부 실시예에서는, 저속 제어부(412)는 저속 디바이스(406) 및 저속 확장 포트(414)에 연결된다. 다양한 통신 포트(예를 들어, USB, 블루투스, 이더넷, 무선 이더넷)를 포함할 수 있는 저속 확장 포트는 키보드, 포인팅 디바이스, 스캐너와 같은 하나 이상의 입/출력 디바이스들에 연결되거나, 또는 예컨대 네트워크 어댑터를 통하여, 스위치나 라우터와 같은 네트워크 디바이스에 연결될 수 있다.

[0073] 컴퓨팅 디바이스(400)는 도면에 도시된 바와 같이, 복수의 다른 형태로 구현될 수 있다. 예를 들어, 컴퓨팅 디바이스(400)는 표준 서버(420)로 구현되거나 이러한 서버들의 그룹에서 여러 번 구현될 수 있다. 또한, 컴퓨팅 디바이스(400)는 랙 서버 시스템(424)의 부분으로서 구현될 수 있다. 추가적으로, 컴퓨팅 디바이스(400)는 랩탑 컴퓨터(422)와 같은 개인용 컴퓨터 내에 구현될 수 있다. 대안적으로, 컴퓨팅 디바이스(400)로부터의 구성요소는 디바이스(450)와 같은 모바일 디바이스(도시되지 않음) 내 다른 구성요소와 조합될 수 있다. 이러한 디바이스 각각은 하나 이상의 컴퓨팅 디바이스(400, 450)를 포함하고, 전체 시스템은 서로 통신하는 다중 컴퓨팅 디바이스(400, 450)로 구성될 수 있다.

[0074] 컴퓨팅 디바이스(450)는 여러 구성요소 중에서 프로세서(452), 메모리(464), 디스플레이(454)와 같은 입/출력 디바이스, 통신 인터페이스(466), 및 트랜스시버(468) 등을 포함한다. 또한, 디바이스(450)에는 추가적인 저장소를 제공하기 위하여, 마이크로 드라이브 또는 다른 디바이스와 같은 저장 디바이스가 제공될 수 있다. 구성요소(450, 452, 464, 454, 466, 및 468) 각각은 다양한 버스를 이용하여 서로 접속되고, 구성요소의 몇몇은 공통의 마더보드에 탑재되거나 적절한 다른 방법으로 탑재될 수 있다.

[0075] 프로세서(452)는 컴퓨팅 디바이스(450) 내에서의 실행을 위해 명령어를 실행하며, 이 명령어에는 메모리(464)에 저장된 명령어가 포함된다. 프로세서는 별개이며 다수의 아날로그 및 디지털 프로세서들을 포함하는 칩들의 칩셋으로서 구현될 수 있다. 부가적으로, 프로세서는 어떤 다수의 아키텍처를 이용하여 구현될 수 있다. 예를 들어, 프로세서(402)는 CISC(Complex Instruction Set Computers) 프로세서, RISC(Reduced Instruction Set Computer) 프로세서, 또는 MISC(Minimal Instruction Set Computer) 프로세서 일 수 있다. 프로세서는, 예를 들어, 사용자 인터페이스의 컨트롤, 디바이스(450)에 의해 실행되는 애플리케이션, 및 컴퓨팅 디바이스(450)에 의한 무선 통신과 같은 디바이스(450)의 다른 구성요소들 사이에 조정을 제공할 수 있다.

[0076] 프로세서(452)는 제어 인터페이스(458) 및 디스플레이(454)에 연결된 디스플레이 인터페이스(456)를 통해 사용자와 통신할 수 있다. 디스플레이(454)는, 예를 들어, TFT LCD 디스플레이 또는 OLED 디스플레이, 또는 다른 적절한 디스플레이 기술일 수 있다. 디스플레이 인터페이스(456)는 그래픽 및 다른 정보를 사용자에게 나타내기 위해 디스플레이(454)를 구동하는 적절한 회로를 포함할 수 있다. 제어 인터페이스(458)는 사용자로부터 명령들을 수신하고, 프로세서(452)에 제출하기 위해 그 명령들을 변환한다. 더욱이, 확장 인터페이스(462)는 디바이스(450)와 다른 디바이스들 간에 근거리 통신이 가능하도록 하기 위해, 프로세서(452)와의 통신에 제공될 수 있다. 확장 인터페이스(462)는, 예를 들어, 일부 실시예에서는 유선 통신을 제공하고 다른 실시예에서 무선 통신을 제공하며, 또한 다중 인터페이스가 사용될 수 있다.

[0077] 메모리(464)는 컴퓨팅 디바이스(450) 내에 정보를 저장한다. 메모리(464)는 컴퓨터 판독가능 매체 또는 미디어, 휘발성 메모리 유닛 또는 유닛들, 또는 비 휘발성 메모리 유닛 또는 유닛들 중 하나 이상으로서 구현될 수 있다. 또한, 확장 메모리(474)가 제공되어, 예를 들어 SIMM 카드 인터페이스를 포함하는 확장 인터페이스(474)를 통해 디바이스(450)에 접속될 수 있다. 이러한 확장 메모리(474)는 디바이스(450)를 위한 여분의 저장 공간을 제공할 수 있고, 또한 애플리케이션 또는 디바이스(450)를 위한 다른 정보를 저장할 수 있다. 특히, 확장 메모리(474)는 상술된 프로세스를 실행하거나 보조하기 위한 명령어를 포함하고, 또한 보안 정보를 포함할 수 있다. 따라서, 예를 들어, 확장 메모리(474)는 디바이스(450)용 보안 모듈로서 제공될 수 있고, 디바이스(450)의 안전한 사용을 가능하게 하는 명령어로 프로그램 될 수 있다. 더욱이, 보안 애플리케이션은, 해킹할 수 없는 방식(non-hackable manner)으로 SIMM 카드 상에 식별 정보를 위치시킨 것과 같은 추가적 정보와 함께 SIMM 카드를 통해 제공될 수 있다.

[0078] 메모리는 아래에서 논의되는 것과 같이 예를 들어, 플래시 메모리 및/또는 NVRAM 메모리를 포함할 수 있다. 일 실시예에서, 컴퓨터 프로그램 제품은 정보 캐리어에 유형적으로 구체화된다. 컴퓨터 프로그램 제품은 실행될 때, 상술된 것과 같은 하나 이상의 방법을 수행하는 명령어를 포함한다. 정보 캐리어는 메모리(464), 확장 메모리(474), 또는, 예를 들어 트랜스시버(468) 또는 확장 인터페이스(462)를 통해 수신될 수 있는 프로세서(452)상의 메모리와 같은, 컴퓨터- 또는 기계-판독가능 매체이다.

[0079] 디바이스(450)는 디지털 신호 처리 회로를 필요에 따라 포함하는 통신 인터페이스(466)를 통해 무선으로 통신할

수 있다. 통신 인터페이스(466)는 GSM 음성 호, SMS, EMS, 또는 MMS 메시징, CDMA, TDMA, PDC, WCDMA, CDMA2000, 또는 GPRS 등과 같은 다양한 모드 또는 프로토콜 하에서의 통신을 제공할 수 있다. 이러한 통신은 예를 들어, 무선-주파수 트랜스미터(468)를 통해 수행될 수 있다. 또한, 단거리 통신은 예를 들어, 블루투스, Wi-Fi, 또는 다른 이러한 트랜스미터(도시되지 않음)를 사용하여 수행될 수 있다. 이에 더하여, GPS 수신기 모듈(470)은 추가적인 네비게이션- 및 위치- 관련 무선 데이터를 디바이스(450)에 제공할 수 있으며, 이 무선 데이터는 디바이스(450)에서 실행중인 애플리케이션에 의해 적절하게 사용될 수 있다.

[0080] 디바이스(450)는 또한 사용자로부터의 발화 정보(spoken information)를 수신하고, 그 발화 정보를 사용 가능한 디지털 정보로 변환하는 오디오 코덱(460)을 이용하여, 청취 가능하게(audibly) 통신할 수 있다. 또한, 오디오 코덱(460)은 예를 들어, 디바이스(450)의 핸드셋 내의 스피커를 통하는 것과 같이 해서, 사용자가 들을 수 있는 음성을 생성한다. 이러한 음성은 음성 전화 호로부터의 음성을 포함할 수 있고, 녹음된 음성(예를 들어, 음성 메시지, 뮤직 파일 등)은 포함할 수 있고, 또한 디바이스(450) 상에서 동작하는 애플리케이션에 의해 생성된 음성을 포함할 수 있다.

[0081] 컴퓨팅 디바이스(450)는 도면에 도시된 것처럼, 다수의 다양한 형태로 구현될 수 있다. 예를 들어, 컴퓨팅 디바이스(450)는 셀룰러 전화(480)로서 구현될 수 있다. 또한, 컴퓨팅 디바이스(450)는 스마트폰(482), PDA, 또는 다른 유사한 모바일 디바이스의 일부로서 구현될 수 있다.

[0082] 본 명세서에 기재된 시스템의 다양한 실시예와 기술은 디지털 전자 회로, 집적 회로, 특별하게 설계된 ASICs(application specific integrated circuit), 컴퓨터 하드웨어, 펌웨어, 소프트웨어, 및/또는 그것의 조합물로 실현될 수 있다. 이러한 다양한 실시예는 하나 이상의 컴퓨터 프로그램으로 된 실시예를 포함하며, 이 컴퓨터 프로그램은 적어도 하나의 프로그램 가능한 프로세서를 포함하는 프로그램 가능한 시스템에서 실행가능하고 및/또는 해석 가능하다. 또한, 전용 또는 범용 프로세서일 수 있는 이 프로그램 가능한 프로세서는 데이터와 명령어를 송수신하기 위해, 저장 시스템, 적어도 하나의 입력 디바이스 및 적어도 하나의 수신 디바이스에 연결된다.

[0083] 컴퓨터 프로그램(또한 프로그램, 소프트웨어, 소프트웨어 애플리케이션, 또는 코드로 알려짐)은 프로그램 가능한 프로세서를 위한 기계 명령어를 포함하고, 고레벨 절차 및/또는 객체 지향 프로그램 언어 및/또는 어셈블리/기계 언어로 구현될 수 있다. 본 명세서에서 사용되는 바와 같이, 용어 "기계 판독가능 매체(machine-readable medium)"와 "컴퓨터 판독가능 매체(computer-readable medium)"는 기계 명령어 및/또는 데이터를 프로그램 가능한 프로세서에 제공하기 위해 이용되는 임의의 컴퓨터 프로그램 제품, 장치, 및/또는 디바이스(예를 들어, 마그네틱 디스크, 광학 디스크, 메모리, PLDs(Programmable Logic Devices))를 가리키며, 기계 판독가능 신호와 같은 기계 명령어를 수신하는 기계 판독가능 매체를 포함한다. 용어 "기계 판독가능 신호(machine-readable signal)"는 기계 명령어 및/또는 데이터를 프로그램 가능한 프로세서에 제공하기 위해 사용되는 임의의 신호를 가리킨다.

[0084] 사용자와의 상호 작용을 제공하기 위하여, 본 명세서에 기술된 시스템과 기술은, 정보를 사용자에게 디스플레이 하기 위한 디스플레이 디바이스(예를 들어, CRT 또는 LCD 모니터)와 사용자가 컴퓨터에 입력을 제공할 수 있는 키보드 및 포인팅 디바이스(예를 들어, 마우스 또는 트랙볼)를 구비한 컴퓨터 상에서 구현될 수 있다. 사용자와의 상호 작용을 제공하기 위하여 다른 종류의 디바이스가 또한 사용될 수 있다; 예를 들어, 사용자에게 제공되는 피드백은 임의의 형태의 감각 피드백(예를 들어, 시각 피드백, 청각 피드백 또는 촉각 피드백)일 수 있고, 사용자로부터의 입력은 음향(acoustic), 음성(speech) 또는 촉각(tactile) 입력을 포함하는 임의의 형태로 수신될 수 있다.

[0085] 본 명세서에서 설명한 시스템과 기술은, 백 엔드(back end) 구성요소(예를 들어, 데이터 서버와 같은), 또는 미들웨어 구성요소(예를 들어, 애플리케이션 서버), 또는 프론트 엔드(front end) 구성요소(예를 들어, 본 명세서에서 설명된 시스템 및 기술의 실시예와 사용자가 상호 작용할 수 있는 그래픽 사용자 인터페이스 또는 웹브라우저를 구비한 클라이언트 컴퓨터), 또는 이러한 백 엔드, 미들웨어, 또는 프론트 엔드 구성요소들의 임의의 조합을 포함하는 컴퓨팅 시스템으로 구현될 수 있다. 시스템의 구성요소는 디지털 데이터 통신의 임의의 형태 또는 매체(예를 들어, 통신 네트워크)에 의해 상호 접속될 수 있다. 통신 네트워크의 예시는 LAN, WAN, 및 (애드혹(ad-hoc) 또는 정적 구성원을 포함하는) 피어-투-피어(peer-to-peer) 네트워크, 그리드 컴퓨팅 인프라스트럭처(grid computing infrastructures), 및 인터넷을 포함한다.

[0086] 컴퓨팅 시스템은 클라이언트와 서버를 포함할 수 있다. 클라이언트와 서버는 보통 서로 떨어져 있으며, 일반적으로는 통신 네트워크를 통하여 상호 작용한다. 클라이언트와 서버의 관계는 각각의 컴퓨터 상에서 실행되고 상

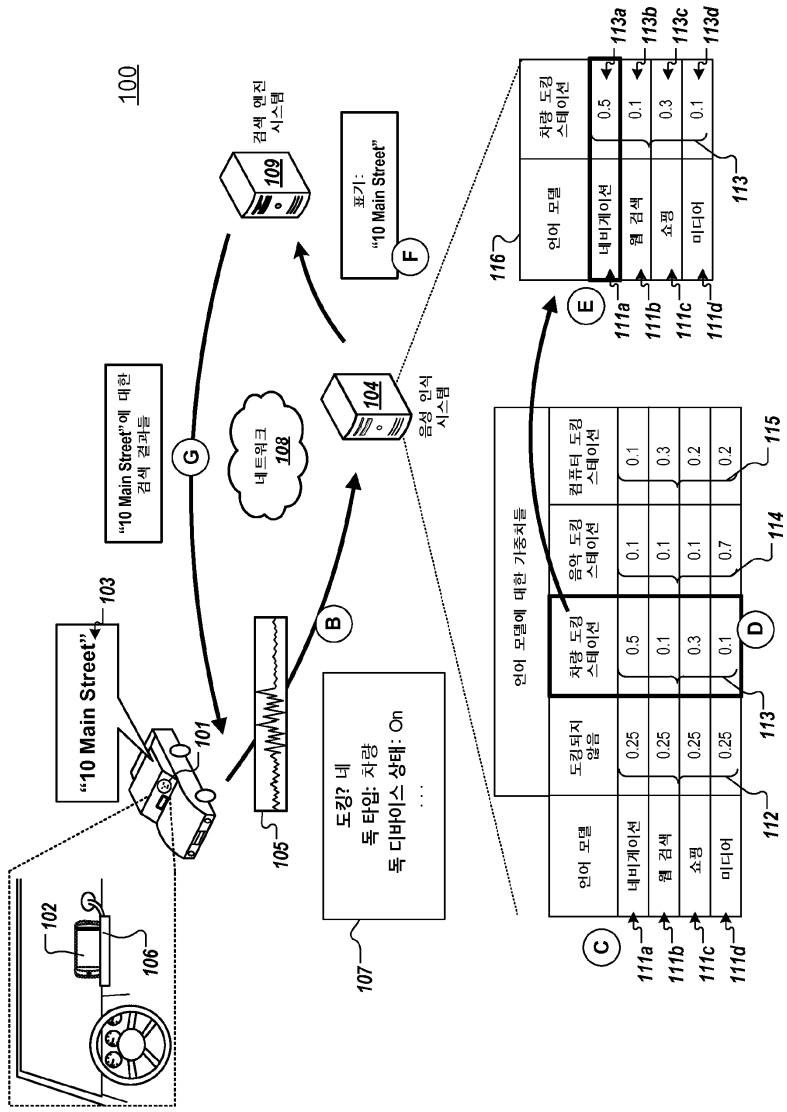
호 클라이언트-서버 관계를 갖는 컴퓨터 프로그램에 의하여 발생한다.

[0087]

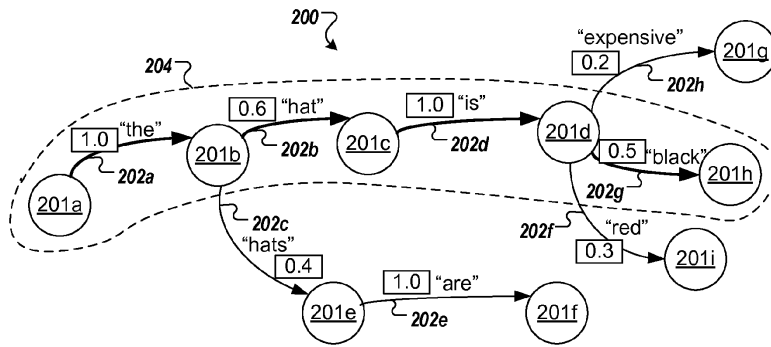
다수의 실시예들이 기술되었다. 그럼에도 불구하고, 본 발명의 범위 및 사상으로부터 벗어나지 않고 다양한 변형들이 가해질 수 있음이 이해될 것이다. 예를 들어, 친구들 또는 그룹들과 공유하기 위한 대상들은 다양한 방식으로 액세스 하기 위해 이용될 수 있고, 그러한 정보를 디스플레이 하는 특정 방법들은 여기에 도시된 특정 예시들로부터 다양해질 수 있다. 따라서, 다른 실시예들도 후술하는 청구항들의 범위 내에 속한다.

도면

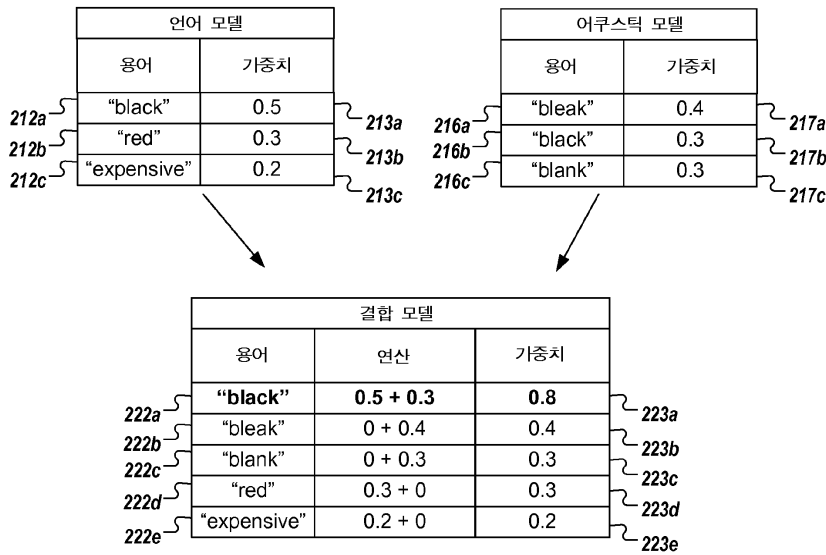
도면1



도면2a

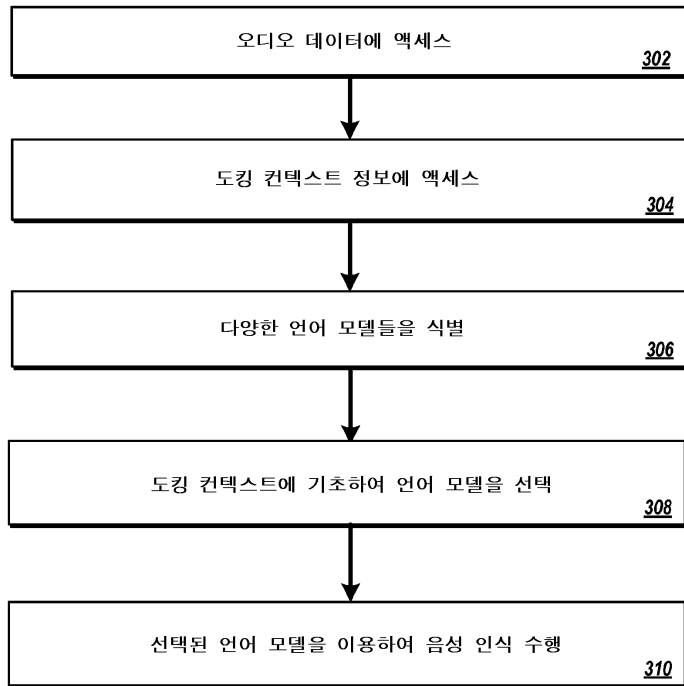


도면2b

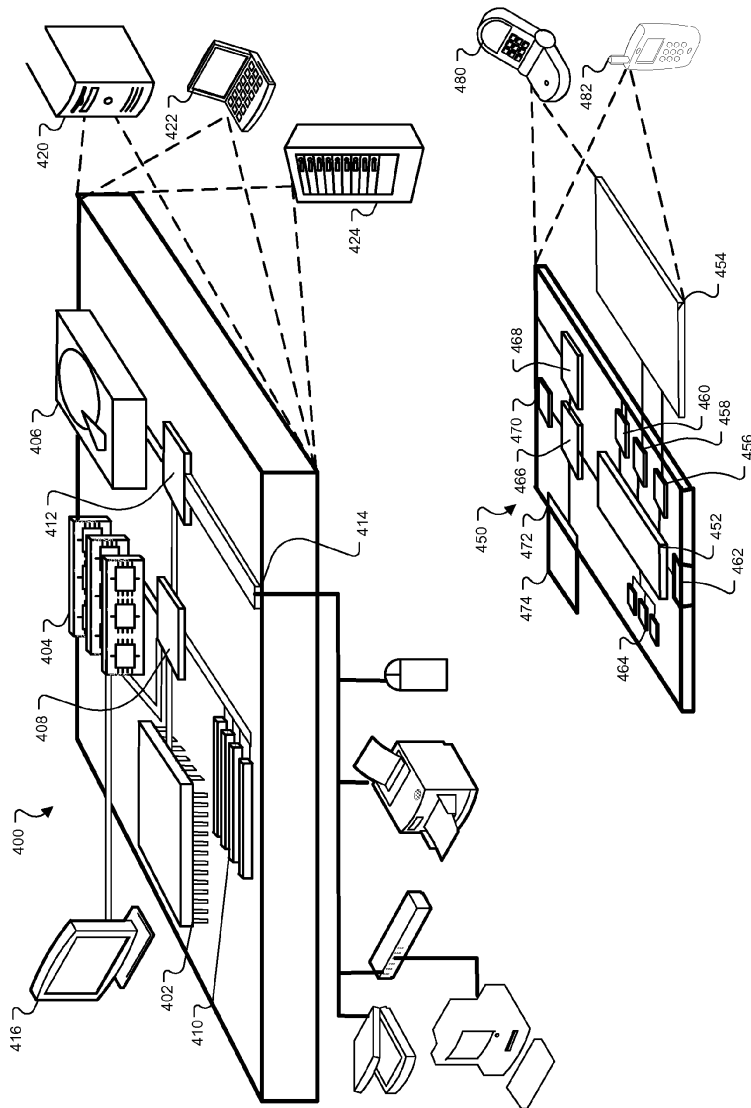


도면3

300 ↘



도면4



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 6의 7번째 줄

【변경전】

상기 도킹 콘텍스트에 기초하여

【변경후】

상기 도킹 컨텍스트에 기초하여

【직권보정 2】

【보정항목】 청구범위

【보정세부항목】 청구항 17의 11번째 줄

【변경전】

상기 도킹 콘텍스트에 기초하여

【변경후】

상기 도킹 컨텍스트에 기초하여

【직권보정 3】

【보정항목】 청구범위

【보정세부항목】 청구항 21의 9번째 줄

【변경전】

상기 도킹 컨텍스트에 기초하여

【변경후】

상기 도킹 컨텍스트에 기초하여

【직권보정 4】

【보정항목】 청구범위

【보정세부항목】 청구항 31의 2번째 줄

【변경전】

상기 도킹 컨텍스트에 기초하여

【변경후】

상기 도킹 컨텍스트에 기초하여

【직권보정 5】

【보정항목】 청구범위

【보정세부항목】 청구항 26의 12번째 줄

【변경전】

상기 복수의 언어 모델 각각에 대해

【변경후】

복수의 언어 모델 각각에 대해

【직권보정 6】

【보정항목】 청구범위

【보정세부항목】 청구항 30의 2번째 줄

【변경전】

상기 도킹 컨텍스트는

【변경후】

상기 도킹 컨텍스트는

【직권보정 7】

【보정항목】 청구범위

【보정세부항목】 청구항 30의 4번째 줄

【변경전】

상기 도킹 컨텍스트에 기초하여

【변경후】

상기 도킹 컨텍스트에 기초하여

【직권보정 8】

【보정항목】 청구범위

【보정세부항목】 청구항 26의 12번째 줄

【변경전】

상기 도킹 콘택스트에 기초하여

【변경후】

상기 도킹 콘택스트에 기초하여