(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년02월24일
(11) 등록번호 10-2219274
(24) 등록일자 2021년02월17일

(51) 국제특허분류(Int. Cl.)
G10L 13/04 (2006.01) *G06F 40/20* (2020.01)
G10L 13/08 (2006.01)
(52) CPC특허분류
G10L 13/00 (2013.01)
G06F 40/253 (2020.01)
(21) 출원번호 10-2018-7021923
(22) 출원일자(국제) 2016년12월29일
심사청구일자 2018년07월27일
(85) 번역문제출일자 2018년07월27일
(65) 공개번호 10-2018-0098654
(43) 공개일자 2018년09월04일
(86) 국제출원번호 PCT/US2016/069182
(87) 국제공개번호 WO 2017/131924
국제공개일자 2017년08월03일
(30) 우선권주장
15/009,432 2016년01월28일 미국(US)
(56) 선행기술조사문헌
JP2010145873 A*
JP2810750 B2*
US20150332665 A1*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
구글 엘엘씨
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
(72) 발명자
샤리피 매튜
스위스 취리히 8002 브랜드스첸케스트라쎄 110
포레스터 제이콥 니콜라스
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
(74) 대리인
박장원

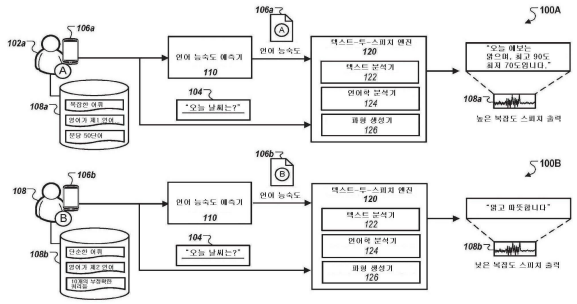
전체 청구항 수 : 총 20 항 심사관 : 정성윤

(54) 발명의 명칭 **적응적 텍스트-투-스피치 출력**

(57) 요 약

일부 구현예들에서, 클라이언트 디바이스의 사용자의 언어 능숙도가 하나 이상의 컴퓨터들에 의해 결정된다. 그 후, 하나 이상의 컴퓨터들은 사용자의 상기 결정된 언어 능숙도에 기초하여 텍스트-투-스피치 모듈에 의한 출력을 위한 텍스트 세그먼트를 결정한다. 출력을 위한 텍스트 세그먼트를 결정한 후에, 하나 이상의 컴퓨터들은 텍스트 세그먼트의 합성된 발언을 포함하는 오디오 데이터를 생성한다. 그 후, 텍스트 세그먼트의 합성된 발언을 포함하는 오디오 데이터는 출력을 위해 클라이언트 디바이스에 제공된다. 개선된 사용자 인터페이스가 보다 나은 텍스트-투-스피치 변환을 통해 제공된다.

대 표 도



(52) CPC특허분류
G06F 40/289 (2020.01)
G10L 13/08 (2013.01)

명 세 서

청구범위

청구항 1

하나 이상의 컴퓨터들에 의해 수행되는 방법으로서,

상기 하나 이상의 컴퓨터들에 의해, (i) 사용자에 의해 제출된 이전의 쿼리들 또는 (ii) 상기 사용자가 음성 인터페이스를 통해 음성 제어 작업을 성공적으로 완료한 비율 중 적어도 하나 이상에 기초하여 클라이언트 디바이스의 사용자의 언어 능숙도를 결정하는 단계;

상기 하나 이상의 컴퓨터들에 의해, 상기 사용자의 상기 결정된 언어 능숙도에 기초하여 텍스트-투-스피치 모듈에 의한 출력을 위한 텍스트 세그먼트를 결정하는 단계, 상기 텍스트 세그먼트를 결정하는 단계는:

상이한 레벨의 언어 복잡도를 표시하는 언어 복잡도 스코어를 각각 가지는 다수의 텍스트 세그먼드들로부터, 상기 클라이언트 디바이스의 사용자의 상기 결정된 언어 능숙도를 기술하는 기준 스코어와 가장 잘 매칭되는 상기 언어 복잡도 스코어를 가지는 텍스트 세그먼트를 선택하는 단계; 또는

(i) 상기 사용자의 상기 결정된 언어 능숙도 및 (ii) 특정한 텍스트 세그먼트의 언어 복잡도 스코어에 적어도 기초하여 상기 사용자에 대한 텍스트-투-스피치 출력을 위한 상기 특정한 텍스트 세그먼트 수정하는 단계를 포함하며;

상기 하나 이상의 컴퓨터들에 의해, 상기 텍스트 세그먼트의 합성된 발언을 포함하는 오디오 데이터를 생성하는 단계;

상기 하나 이상의 컴퓨터들에 의해, 상기 텍스트 세그먼트의 상기 합성된 발언을 포함하는 상기 오디오 데이터를 상기 클라이언트 디바이스에 제공하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 2

청구항 1에 있어서,

상기 클라이언트 디바이스는 텍스트-투-스피치 인터페이스를 사용하는 모바일 어플리케이션을 디스플레이하는 것을 특징으로 하는 방법.

청구항 3

청구항 1에 있어서,

상기 텍스트-투-스피치 모듈에 의한 출력을 위한 상기 텍스트 세그먼트를 결정하는 단계는:

다수의 텍스트 세그먼트들을 상기 사용자에 대한 텍스트-투-스피치 출력을 위한 후보들로서 식별하는 것, 상기 다수의 텍스트 세그먼트들은 상이한 레벨의 언어 복잡도를 가지며; 및

상기 클라이언트 디바이스의 사용자의 상기 결정된 언어 능숙도에 적어도 기초하여 상기 다수의 텍스트 세그먼트들 중에서 선택하는 것을 포함하는 것을 특징으로 하는 방법.

청구항 4

청구항 3에 있어서,

상기 다수의 텍스트 세그먼트들 중에서 선택하는 것은:

상기 다수의 텍스트 세그먼트들 각각에 대해 언어 복잡도 스코어를 결정하는 것; 및

상기 클라이언트 디바이스의 사용자의 상기 언어 능숙도를 기술하는 기준 스코어와 가장 잘 매칭되는 상기 언어 복잡도 스코어를 가지는 상기 텍스트 세그먼트를 선택하는 것을 포함하는 것을 특징으로 하는 방법.

청구항 5

청구항 1에 있어서,

상기 텍스트-투-스피치 모듈에 의한 출력을 위한 상기 텍스트 세그먼트를 결정하는 단계는:

상기 사용자에 대한 텍스트-투-스피치 출력을 위한 텍스트 세그먼트를 식별하는 것;

상기 텍스트-투-스피치 출력에 대한 상기 텍스트 세그먼트의 복잡도 스코어를 계산하는 것; 및

상기 사용자의 상기 결정된 언어 능숙도 및 상기 텍스트-투-스피치 출력을 위한 상기 텍스트 세그먼트의 상기 복잡도 스코어에 적어도 기초하여 상기 사용자에 대한 상기 텍스트-투-스피치 출력을 위한 상기 텍스트 세그먼트를 수정하는 것을 포함하는 것을 특징으로 하는 방법.

청구항 6

청구항 5에 있어서,

상기 사용자에 대한 상기 텍스트-투-스피치 출력을 위한 상기 텍스트 세그먼트를 수정하는 것은:

상기 사용자의 상기 결정된 언어 능숙도에 적어도 기초하여 상기 사용자에 대한 종합적(overall) 복잡도 스코어를 결정하는 것;

상기 사용자에 대한 상기 텍스트-투-스피치 출력을 위한 상기 텍스트 세그먼트 내에서 개별 부분들에 대한 복잡도 스코어를 결정하는 것;

상기 사용자에 대한 상기 종합적 복잡도 스코어보다 큰 복잡도 스코어들을 가지는 상기 텍스트 세그먼트 내의 하나 이상의 개별 부분들을 식별하는 것; 및

복잡도 스코어들을 상기 종합적 복잡도 스코어 미만으로 감소시키기 위해 상기 텍스트 세그먼트 내에서 상기 하나 이상의 개별 부분들을 수정하는 것을 포함하는 것을 특징으로 하는 방법.

청구항 7

청구항 5에 있어서,

상기 사용자에 대한 상기 텍스트-투-스피치 출력을 위한 상기 텍스트 세그먼트를 수정하는 것은:

상기 사용자와 연관된 컨텍스트를 표시하는 데이터를 수신하는 것;

상기 사용자와 연관된 상기 컨텍스트에 대한 종합적 복잡도 스코어를 결정하는 것;

상기 텍스트 세그먼트의 상기 복잡도 스코어가 상기 사용자와 연관된 상기 컨텍스트에 대한 상기 종합적 복잡도 스코어를 초과함을 결정하는 것; 및

상기 복잡도 스코어를 상기 사용자와 연관된 상기 컨텍스트에 대한 상기 종합적 복잡도 스코어 미만으로 감소시키기 위해 상기 텍스트 세그먼트를 수정하는 것을 포함하는 것을 특징으로 하는 방법.

청구항 8

시스템으로서,

하나 이상의 컴퓨터들; 및

명령어들이 저장된 상기 하나 이상의 컴퓨터들에 연결된 비일시적 컴퓨터 판독가능 매체를 포함하며, 상기 명령어들은 상기 하나 이상의 컴퓨터들에 의해 실행될 때, 상기 하나 이상의 컴퓨터들로 하여금 동작들을 수행하게 하며, 상기 동작들은:

상기 하나 이상의 컴퓨터들에 의해, (i) 사용자에 의해 제출된 이전의 쿼리들 또는 (ii) 상기 사용자가 음성 인터페이스를 통해 음성 제어 작업을 성공적으로 완료한 비율 중 적어도 하나 이상에 기초하여 클라이언트 디바이스의 사용자의 언어 능숙도를 결정하는 동작;

상기 하나 이상의 컴퓨터들에 의해, 상기 사용자의 상기 결정된 언어 능숙도에 기초하여 텍스트-투-스피치 모듈에 의한 출력을 위한 텍스트 세그먼트를 결정하는 동작, 상기 텍스트 세그먼트를 결정하는 동작은:

상이한 레벨의 언어 복잡도를 표시하는 언어 복잡도 스코어를 각각 가지는 다수의 텍스트 세그먼드들로

부터, 상기 클라이언트 디바이스의 사용자의 상기 결정된 언어 능숙도를 기술하는 기준 스코어와 가장 잘 매칭되는 상기 언어 복잡도 스코어를 가지는 텍스트 세그먼트를 선택하는 동작; 또는

(i) 상기 사용자의 상기 결정된 언어 능숙도 및 (ii) 특정한 텍스트 세그먼트의 언어 복잡도 스코어에 적어도 기초하여 상기 사용자에 대한 텍스트-투-스피치 출력을 위한 상기 특정한 텍스트 세그먼트 수정하는 동작을 포함하며;

상기 하나 이상의 컴퓨터들에 의해, 상기 텍스트 세그먼트의 합성된 발언을 포함하는 오디오 데이터를 생성하는 동작;

상기 하나 이상의 컴퓨터들에 의해, 상기 텍스트 세그먼트의 상기 합성된 발언을 포함하는 상기 오디오 데이터를 상기 클라이언트 디바이스에 제공하는 동작을 포함하는 것을 특징으로 하는 시스템.

청구항 9

청구항 8에 있어서,

상기 클라이언트 디바이스는 텍스트-투-스피치 인터페이스를 사용하는 모바일 어플리케이션을 디스플레이하는 것을 특징으로 하는 시스템.

청구항 10

청구항 8에 있어서,

상기 텍스트-투-스피치 모듈에 의한 출력을 위한 상기 텍스트 세그먼트를 결정하는 동작은:

다수의 텍스트 세그먼트들을 상기 사용자에게 텍스트-투-스피치 출력을 위한 후보들로서 식별하는 것, 상기 다수의 텍스트 세그먼트들은 상이한 레벨의 언어 복잡도를 가지며; 및

상기 클라이언트 디바이스의 사용자의 상기 결정된 언어 능숙도에 적어도 기초하여 상기 다수의 텍스트 세그먼트들 중에서 선택하는 것을 포함하는 것을 특징으로 하는 시스템.

청구항 11

청구항 10에 있어서,

상기 다수의 텍스트 세그먼트들 중에서 선택하는 것은:

상기 다수의 텍스트 세그먼트들 각각에 대해 언어 복잡도 스코어를 결정하는 것; 및

상기 클라이언트 디바이스의 사용자의 상기 언어 능숙도를 기술하는 기준 스코어와 가장 잘 매칭되는 상기 언어 복잡도 스코어를 가지는 상기 텍스트 세그먼트를 선택하는 것을 포함하는 것을 특징으로 하는 시스템.

청구항 12

청구항 8에 있어서,

상기 텍스트-투-스피치 모듈에 의한 출력을 위한 상기 텍스트 세그먼트를 결정하는 동작은:

상기 사용자에 대한 텍스트-투-스피치 출력을 위한 텍스트 세그먼트를 식별하는 것;

상기 텍스트-투-스피치 출력에 대한 상기 텍스트 세그먼트의 복잡도 스코어를 계산하는 것; 및

상기 사용자의 상기 결정된 언어 능숙도 및 상기 텍스트-투-스피치 출력을 위한 상기 텍스트 세그먼트의 상기 복잡도 스코어에 적어도 기초하여 상기 사용자에 대한 상기 텍스트-투-스피치 출력을 위한 상기 텍스트 세그먼트를 수정하는 것을 포함하는 것을 특징으로 하는 시스템.

청구항 13

청구항 12에 있어서,

상기 사용자에 대한 상기 텍스트-투-스피치 출력을 위한 상기 텍스트 세그먼트를 수정하는 것은:

상기 사용자의 상기 결정된 언어 능숙도에 적어도 기초하여 상기 사용자에 대한 종합적(overall) 복잡도 스코어

를 결정하는 것;

상기 사용자에 대한 상기 텍스트-투-스피치 출력을 위한 상기 텍스트 세그먼트 내에서 개별 부분들에 대한 복잡도 스코어를 결정하는 것;

상기 사용자에 대한 상기 종합적 복잡도 스코어보다 큰 복잡도 스코어들을 가지는 상기 텍스트 세그먼트 내의 하나 이상의 개별 부분들을 식별하는 것; 및

복잡도 스코어들을 상기 종합적 복잡도 스코어 미만으로 감소시키기 위해 상기 텍스트 세그먼트 내에서 상기 하나 이상의 개별 부분들을 수정하는 것을 포함하는 것을 특징으로 하는 시스템.

청구항 14

하나 이상의 프로세싱 디바이스들에 의해 수행되는 방법으로서,

상기 하나 이상의 프로세싱 디바이스들에 의해, 디바이스에 대한 음성 입력의 복잡도 레벨을 결정하는 단계, 상기 디바이스에 대한 음성 입력의 복잡도 레벨을 결정하는 단계는 단어들에 대해 미리 결정된 복잡도와 상기 음성 입력을 연관시키는 것을 포함하며;

제출된 음성 입력에서 미인식 단어들의 수에 기초하여 상기 디바이스에 대한 상기 음성 입력을 제출한 사용자의 언어 능숙도를 결정하는 단계;

상기 하나 이상의 프로세싱 디바이스들에 의해, 상기 음성 입력에 응답하여, 출력을 위한 메시지를 결정하는 단계, 상기 메시지는 상기 음성 입력과 연관된 상기 결정된 복잡도 및 상기 디바이스에 대한 상기 음성 입력을 제출한 사용자의 상기 언어 능숙도에 기초하여 결정되며, 상기 메시지를 결정하는 단계는:

상이한 레벨의 언어 복잡도를 표시하는 언어 복잡도 스코어를 각각 가지는 다수의 텍스트 세그먼드들로부터, 상기 클라이언트 디바이스의 사용자의 상기 결정된 언어 능숙도를 기술하는 기준 스코어와 가장 잘 매칭되는 상기 언어 복잡도 스코어를 가지는 메시지를 선택하는 단계; 또는

(i) 상기 사용자의 상기 결정된 언어 능숙도 및 (ii) 특정한 메시지의 언어 복잡도 스코어에 적어도 기초하여 상기 사용자에 대한 텍스트-투-스피치 출력을 위한 상기 특정한 텍스트 세그먼트 수정하는 단계를 포함하며;

상기 하나 이상의 프로세싱 디바이스들에 의해, 상기 메시지의 합성된 발언을 포함하는 오디오 데이터를 생성하는 단계; 및;

상기 하나 이상의 프로세싱 디바이스들에 의해, 상기 음성 입력에 응답하여, 상기 합성된 발언을 포함하는 상기 오디오 데이터를 출력을 위해 제공하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 15

청구항 14에 있어서,

상기 음성 입력의 상기 복잡도 레벨은 상기 음성 입력의 언어 복잡도를 포함하는 것을 특징으로 하는 방법.

청구항 16

청구항 14에 있어서,

출력을 위한 상기 메시지를 결정하는 단계는:

상기 음성 입력에 응답하여 출력을 위한 기본 메시지(baseline message)를 획득하는 것; 및

상기 디바이스에 대한 상기 음성 입력에 대한 상기 결정된 복잡도 레벨에 기초하여 상기 기본 메시지의 복잡도 레벨을 증가시킴으로써 조절된 메시지를 생성하는 것을 포함하는 것을 특징으로 하는 방법.

청구항 17

청구항 14에 있어서,

출력을 위한 상기 메시지를 결정하는 단계는:

상기 음성 입력에 응답하여 출력을 위한 기본 메시지(baseline message)를 획득하는 것; 및

상기 디바이스에 대한 상기 음성 입력에 대한 상기 결정된 복잡도 레벨에 기초하여 상기 기본 메시지의 복잡도 레벨을 감소시킴으로써 조절된 메시지를 생성하는 것을 포함하는 것을 특징으로 하는 방법.

청구항 18

청구항 14에 있어서,

상기 디바이스는 텍스트-투-스피치 인터페이스를 사용하는 모바일 어플리케이션를 실행하는 것을 특징으로 하는 방법.

청구항 19

하나 이상의 프로세싱 디바이스들 및 명령어들을 저장하는 하나 이상의 기계 판독가능 저장 디바이스들을 포함하는 시스템으로서, 상기 명령어들은 상기 하나 이상의 프로세싱 디바이스들에 의해 수행될 때, 상기 시스템으로 하여금 청구항 14 내지 18 중 어느 한 항의 방법을 수행하게 하는 것을 특징으로 하는 시스템.

청구항 20

명령어들을 저장하는 하나 이상의 기계 판독가능 저장 디바이스들로서, 상기 명령어들은 하나 이상의 프로세싱 디바이스들에 의해 수행될 때, 상기 하나 이상의 프로세싱 디바이스들로 하여금 청구항 1 내지 7, 또는 14 내지 18 중 어느 한 항의 방법을 수행하게 하는 것을 특징으로 하는 기계 판독가능 저장 디바이스.

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

발명의 설명

기 술 분 야

[0001] 관련 출원에 대한 상호 참조

[0002] 본 출원은 2016년 1월 28일에 출원된 미국 출원 번호 제15/009,432호 "적응적 텍스트-투-스피치 출력"에 대해 우선권을 주장하며, 그 전체가 본 명세서에 참조로서 통합된다.

[0003] 기술분야

[0004] 본 명세서는 전자 통신을 일반적으로 기술한다.

배 경 기 술

[0005] 스피치 합성은 사람의 스피치의 인공적 생산을 지칭한다. 스피치 합성기들은 텍스트에 대응하는 스피치 출력을 생성하기 위한 소프트웨어 또는 하드웨어 컴포넌트들에서 구현될 수 있다. 예를 들어, 텍스트-투-스피치(TTS) 시스템은 일반적으로 데이터베이스에 저장된 기록된 스피치의 조각들을 연결(concatenating)함으로써 보통의 언어 텍스트를 스피치로 변환한다.

발명의 내용

[0006] 전자 컴퓨팅의 많은 부분이 데스크톱으로부터 모바일 환경으로 이동함에 따라 스피치 합성은 사용자 경험의 중심이 되고 있다. 예를 들면, 디스플레이가 없는 보다 작은 모바일 디바이스들의 사용의 증가는 모바일 디바이스들 상에 디스플레이되는 컨텐츠에 엑세스하고 사용하기 위한 텍스트-투-스피치(TTS) 시스템들의 사용을 증가시켰다.

[0007] 본 명세서는 개선된 사용자 인터페이스들 특히, 개선된 TTS를 통한 보다 나은 컴퓨터 대 사용자 통신을 개시한다.

[0008] 기존 TTS 시스템들이 가지는 하나의 특정한 이슈는 시스템들이 상이한 사용자들의 다양한 언어 능숙도들에 종종 적응할 수 없다는 것이다. 이러한 유연성의 부족은 제한된 언어 능숙도들을 가지는 사용자들이 복잡한 텍스트-투-스피치 출력을 이해하는 것을 종종 방지한다. 예를 들어, TTS 시스템을 사용하는 모국어가 아닌 언어 발화자들(non-native language speakers)은 그들의 제한된 언어 친숙도 때문에 텍스트-투-스피치 추력을 이해하는데 어려움을 겪는다. 기존의 TTS 시스템들이 가지는 다른 이슈는 텍스트-투-스피치 추력을 이해하는 사용자의 즉각적인 능력 또한 특정한 사용자 컨텍스트에 기초하여 다양할 수 있다는 것이다. 예를 들어, 일부 사용자 컨텍스트들은 보다 길거나 더 복잡한 텍스트-투-스피치 출력을 이해하는 것을 더 어렵게 할 수 있는 배경 소음을 포함한다.

[0009] 일부 구현예들에서, 시스템은 사용자가 텍스트-투-스피치 출력을 이해할 수 있는 가능성을 증가시키기 위해 사용자의 언어 능숙도에 기초하여 텍스트-투-스피치 출력에 사용되는 텍스트를 조절한다. 예를 들어, 사용자의 언

어 능숙도는 이전의 사용자 활동으로부터 추론될 수 있고, 사용자의 언어 능숙도에 상응하는 적절한 복잡도로 텍스트-투-스피치 출력을 조절하는데 사용될 수 있다. 일부 예시에서, 시스템은 언어 능숙도의 상이한 레벨들에 대응하는 다수의 후보 텍스트 세그먼트들을 획득한다. 그 후 시스템은 가장 잘 매칭되고 사용자의 언어 능숙도에 가장 가깝게 대응하는 후보 텍스트 세그먼트를 선택하고, 선택된 텍스트 세그먼트의 합성된 발언을 사용자에 대한 출력으로 제공한다. 다른 예시에서, 시스템은 텍스트-투-스피치 출력을 생성하기 전에 사용자의 언어 능숙도에 더 잘 대응되도록 상기 텍스트를 텍스트 세그먼트로 변경한다. 어휘, 문장 구조, 길이 등을 포함하는 텍스트 세그먼트의 다양한 양태들이 조절될 수 있다. 그 후 시스템은 사용자에 출력하기 위한 변경된 텍스트 세그먼트의 합성된 발언을 제공한다.

[0010] 본 명세서에서 논의된 시스템들이 사용자들에 관한 개인 정보를 수집하거나 또는 개인 정보를 사용하는 경우들에 있어서, 사용자들에게 프로그램들 또는 구성들이 개인 정보 예를 들면, 사용자의 소셜 네트워크, 소셜 액션들 또는 활동들, 직업, 사용자의 선호들 또는 사용자의 현재 위치에 관한 정보를 수집할 것인지 여부를 제어할, 사용자와 더 관련된 컨텐츠 서버로부터의 컨텐츠를 수신할지 및/또는 어떻게 수신할지 제어할 기회가 제공될 수 있다. 추가로, 특정 데이터는 그것이 저장되거나 사용되기 전에 하나 이상의 다양한 방식들로 익명화되어, 개인적으로 식별가능한 정보는 제거된다. 예를 들면, 사용자의 신원은 사용자에 관한 개인적으로 식별가능한 정보가 결정될 수 없도록 익명화되거나 또는 사용자의 지리적 위치는 위치 정보가 획득된 곳에서 일반화되어(시, 우편번호 또는 주 수준으로), 사용자의 특정한 위치가 결정될 수 없도록 한다. 따라서, 사용자는 그 또는 그녀에 관한 정보가 어떻게 수집되는지 그리고 컨텐츠 서버에 의해 사용되는지에 관한 제어를 가질 수 있다.

[0011] 일 양태에서, 컴퓨터로 구현되는 방법은 하나 이상의 컴퓨터들에 의해, 클라이언트 디바이스의 사용자의 언어 능숙도를 결정하는 단계; 상기 하나 이상의 컴퓨터들에 의해, 상기 사용자의 상기 결정된 언어 능숙도에 기초하여 텍스트-투-스피치 모듈에 의한 출력을 위한 텍스트 세그먼트를 결정하는 단계; 상기 하나 이상의 컴퓨터들에 의해, 상기 텍스트 세그먼트의 합성된 발언을 포함하는 오디오 데이터를 생성하는 단계; 상기 하나 이상의 컴퓨터들에 의해, 상기 텍스트 세그먼트의 상기 합성된 발언을 포함하는 상기 오디오 데이터를 상기 클라이언트 디바이스에 제공하는 단계를 포함할 수 있다.

[0012] 다른 버젼들은 컴퓨터 저장 디바이스들에 인코딩된 방법들의 액션들을 수행하도록 구성된 대응 시스템들 및 컴퓨터 프로그램들을 포함한다.

[0013] 하나 이상의 구현예들은 다음의 선택적 구성들을 포함할 수 있다. 예를 들면, 일부 구현예들에서, 상기 클라이언트 디바이스는 텍스트-투-스피치 인터페이스를 사용하는 모바일 어플리케이션을 디스플레이한다.

[0014] 일부 구현예들에서, 상기 사용자의 상기 언어 능숙도를 결정하는 단계는 상기 사용자에 의해 제출된 이전의 쿼리들에 적어도 기초하여 상기 사용자의 언어 능숙도를 추론하는 것을 포함한다.

[0015] 일부 구현예들에서, 텍스트-투-스피치 모듈에 의한 출력을 위한 텍스트 세그먼트를 결정하는 단계는 다수의 텍스트 세그먼트들을 상기 사용자에 대한 텍스트-투-스피치 출력을 위한 후보들로서 식별하는 것, 상기 다수의 텍스트 세그먼트들은 상이한 레벨의 언어 복잡도를 가지며; 및 상기 클라이언트 디바이스의 사용자의 상기 결정된 언어 능숙도에 적어도 기초하여 상기 다수의 텍스트 세그먼트들 중에서 선택하는 것을 포함한다.

[0016] 일부 구현예들에서, 상기 다수의 텍스트 세그먼트들 중에서 선택하는 것은 상기 다수의 텍스트 세그먼트들 각각에 대해 언어 복잡도 스코어를 결정하는 것; 및 상기 클라이언트 디바이스의 사용자의 상기 언어 능숙도를 기술하는 기준 스코어와 가장 잘 매칭되는 상기 언어 복잡도 스코어를 가지는 상기 텍스트 세그먼트를 선택하는 것을 포함한다.

[0017] 일부 구현예들에서, 텍스트-투-스피치 모듈에 의한 출력을 위한 텍스트 세그먼트를 결정하는 단계는 상기 사용자에 대한 텍스트-투-스피치 출력을 위한 텍스트 세그먼트를 식별하는 것; 상기 텍스트-투-스피치 출력에 대한 상기 텍스트 세그먼트의 복잡도 스코어를 계산하는 것; 및 상기 사용자의 상기 결정된 언어 능숙도 및 상기 텍스트-투-스피치 출력을 위한 텍스트 세그먼트의 상기 복잡도 스코어에 적어도 기초하여 상기 사용자에 대한 상기 텍스트-투-스피치 출력을 위한 상기 텍스트 세그먼트를 수정하는 것을 포함한다.

[0018] 일부 구현예들에서, 상기 사용자에 대한 상기 텍스트-투-스피치 출력을 위한 상기 텍스트 세그먼트를 수정하는 것은 상기 사용자의 상기 결정된 언어 능숙도에 적어도 기초하여 상기 사용자에 대한 종합적(overall) 복잡도 스코어를 결정하는 것; 상기 사용자에 대한 상기 텍스트-투-스피치 출력을 위한 상기 텍스트 세그먼트 내에서 개별 부분들에 대한 복잡도 스코어를 결정하는 것; 상기 사용자에 대한 상기 종합적 복잡도 스코어보다 큰 복잡도 스코어들을 가지는 상기 텍스트 세그먼트 내의 하나 이상의 개별 부분들을 식별하는 것; 및 복잡도 스코어들

을 상기 종합적 복잡도 스코어 미만으로 감소시키기 위해 상기 텍스트 세그먼트 내에서 상기 하나 이상의 개별 부분들을 수정하는 것을 포함한다.

[0019] 일부 구현예들에서, 상기 사용자에 대한 상기 텍스트-투-스피치 출력을 위한 상기 텍스트 세그먼트를 수정하는 것은 상기 사용자와 연관된 컨텍스트를 표시하는 데이터를 수신하는 것; 상기 사용자와 연관된 상기 컨텍스트에 대한 종합적 복잡도 스코어를 결정하는 것; 상기 텍스트 세그먼트의 상기 복잡도 스코어가 상기 사용자와 연관된 상기 컨텍스트에 대한 상기 종합적 복잡도 스코어를 초과함을 결정하는 것; 및 상기 복잡도 스코어를 상기 사용자와 연관된 상기 컨텍스트에 대한 상기 종합적 복잡도 스코어 미만으로 감소시키기 위해 상기 텍스트 세그먼트를 수정하는 것을 포함한다.

[0020] 다른 양태에서, 컴퓨터 프로그램은 컴퓨팅 장치에 의해 실행될 때 컴퓨팅 장치로 하여금 임의의 상기 방법들을 수행하게 하는 기계 판독가능 명령어들을 포함한다.

[0021] 다른 일반적 양태에서, 컴퓨터로 구현되는 방법은 상기 사용자와 연관된 컨텍스트를 표시하는 데이터를 수신하는 단계; 상기 사용자와 연관된 상기 컨텍스트에 대한 종합적 복잡도 스코어를 결정하는 단계; 상기 사용자에 대한 텍스트-투-스피치 출력을 위한 텍스트 세그먼트를 식별하는 단계; 상기 텍스트 세그먼트의 상기 복잡도 스코어가 상기 사용자와 연관된 상기 컨텍스트에 대한 상기 종합적 복잡도 스코어를 초과함을 결정하는 것; 및 상기 복잡도 스코어를 상기 사용자와 연관된 상기 컨텍스트에 대한 상기 종합적 복잡도 스코어 미만으로 감소시키기 위해 상기 텍스트 세그먼트를 수정하는 단계를 포함한다.

[0022] 일부 구현예들에서, 상기 사용자와 연관된 상기 컨텍스트에 대한 종합적 복잡도 스코어를 결정하는 단계는 상기 사용자가 상기 컨텍스트에 있었던 것으로 결정된 경우 상기 사용자에 의해 이전에 제출된 쿼리들 내에 포함된 용어들을 식별하는 것; 및 상기 식별된 용어들에 적어도 기초하여 상기 사용자와 연관된 상기 컨텍스트에 대한 종합적 복잡도 스코어를 결정하는 것을 포함한다.

[0023] 일부 구현예들에서, 상기 사용자와 연관된 상기 컨텍스트를 표시하는 상기 데이터는 상기 사용자에 의해 이전에 제출되었던 쿼리들을 포함한다.

[0024] 일부 구현예들에서, 상기 사용자와 연관된 상기 컨텍스트를 표시하는 상기 데이터는 상기 사용자와 연관된 현재 위치를 표시하는 GPS 신호를 포함한다.

[0025] 하나 이상의 구현예들의 세부 내용들이 첨부 도면과 아래의 설명에서 기술된다. 다른 잠재적 구성들 및 이점들은 설명, 도면 및 청구항으로부터 명백해질 것이다.

[0026] 이들 양태들의 다른 구현예들은 상기 방법들의 액션들을 수행하도록 구성된 대응 시스템들, 장치들 및 컴퓨터 저장 디바이스들에 인코딩된 컴퓨터 프로그램들을 포함한다.

도면의 간단한 설명

[0027] 도 1은 언어 능숙도에 기초하여 텍스트-투-스피치 출력을 생성하기 위한 프로세스들의 예시들을 도시하는 도면이다.

도 2은 사용자 컨텍스트에 기초하여 적응적 텍스트-투-스피치 출력을 생성하기 위한 시스템의 예시를 도시하는 도면이다.

도 3은 텍스트-투-스피치 출력 내에서 문장 구조를 수정하기 위한 시스템의 예시를 도시하는 도면이다.

도 4는 클러스터링 기법들을 사용하는 것에 기초하여 적응적 텍스트-투-스피치 출력을 생성하기 위한 시스템의 예시를 도시하는 블록도이다.

도 5는 적응적 텍스트-투-스피치 출력을 생성하기 위한 프로세스의 예시를 도시하는 흐름도이다.

도 6은 본 명세서에 기술된 프로세스들 또는 그 부분들이 구현될 수 있는 컴퓨팅 디바이스들의 블록도이다.

도면에서, 동일한 번호는 전체에 걸쳐 대응하는 부분들을 표현한다.

발명을 실시하기 위한 구체적인 내용

[0028] 도 1은 언어 능숙도에 기초하여 텍스트-투-스피치 출력을 생성하기 위한 프로세스들(100A 및 100B)의 예시들을 도시하는 도면이다. 프로세스들(100A 및 100B)은 텍스트 쿼리(104)에 대해, 각각 높은 언어 능숙도를 가지는 사

용자(102a) 및 낮은 언어 능숙도를 가지는 사용자(102b)에 대한 상이한 텍스트-투-스피치 출력들을 생성하는데 사용된다. 도시된 바와 같이, 사용자 디바이스들(106a 및 106b)에서 쿼리(104)를 수신한 후에, 프로세스(100A)는 사용자(102a)에 대해 높은 복잡도 텍스트-투-스피치 출력(108a)를 생성하는 반면, 프로세스(100B)는 사용자(102b)에 대해 낮은 복잡도 출력(108b)를 생성한다. 추가로, 프로세스들(100A 및 100B)을 실행하는 TTS 시스템들은 언어 능숙도 예측기(110), 텍스트-투-스피치 엔진(120)을 포함할 수 있다. 추가로, 텍스트-투-스피치 엔진(120)은 텍스트 분석기(122), 언어학 분석기(124) 및 파형 생성기(126)를 더 포함할 수 있다.

[0029] 일반적으로, 텍스트-투-스피치 출력을 생성하는데 사용되는 텍스트의 컨텐츠는 사용자의 언어 능숙도에 따라 결정될 수 있다. 추가로 또는 대안으로서, 텍스트-투-스피치 출력를 생성하는데 사용될 텍스트는 사용자의 컨텍스트 예를 들면, 사용자의 위치 또는 활동, 배경 소음 존재, 사용자의 현재 작업 등에 기초하여 결정될 수 있다. 추가로, 청각적 형태로 변환될 텍스트는 사용자가 작업을 완료하기를 실패했거나 액션을 반복하는 중이라는 표시들과 같은 다른 정보를 사용하여 조절되거나 결정될 수 있다.

[0030] 예시에서, 2명의 사용자들, 사용자(102a) 및 사용자(102b)는 각각 사용자 디바이스들(106a 및 106b)에 동일한 쿼리(104)를 어플리케이션, 웹페이지 또는 다른 검색 기능에 대한 입력으로서 제공한다. 예를 들어, 쿼리(104)는 현재 날짜에 대한 일기 예보를 결정하기 위해 사용자 디바이스들(106a 및 106b)에 송신된 음성 쿼리일 수 있다. 그 후, 쿼리(104)는 텍스트-투-스피치 엔진(120)에 전송되어, 상기 쿼리(104)에 대한 응답으로 텍스트-투-스피치 출력을 생성한다.

[0031] 언어 능숙도 예측기(110)는 사용자 데이터(108a)에 기초하여 특정한 사용자(예를 들어, 사용자(102a) 또는 사용자(102b))와 연관된 언어 능숙도 스코어를 결정하는 TTS 시스템 내의 소프트웨어 모듈일 수 있다. 언어 능숙도 스코어는 특정한 언어에서 커뮤니케이션을 이해하는 특히, 특정한 언어에서 스피치를 이해하는 사용자의 능력의 예측치일 수 있다. 언어 능숙도의 일 측정은 음성 제어 작업을 성공적으로 완료하는 사용자의 능력이다. 일정 잡기, 방향 찾기 등 많은 유형의 작업들에는 사용자와 디바이스가 구두 커뮤니케이션을 교환하는 인터렉션들의 시퀀스가 이어진다. 사용자가 음성 인터페이스를 통해 이들 작업의 업무흐름을 성공적으로 완료한 비율은 사용자의 언어 능숙도의 강한 지표일 수 있다. 예를 들면, 사용자가 개시한 10번의 음성 작업들 중 9번을 성공한 사용자는 높은 언어 능숙도를 가졌을 가능성이 크다. 반면에, 사용자가 개시한 음성 작업들 중 대부분을 완료하는데 실패한 사용자는 사용자가 디바이스로부터의 커뮤니케이션을 완전히 이해하지 못했거나 적절한 구두 응답을 제공할 수 없었던 것이므로 낮은 언어 능숙도를 가지는 것으로 추론될 수 있다. 아래에서 추가로 논의될 바와 같이, 사용자가 표준 TTS 출력을 포함하는 업무흐름을 완료하지 않은 경우, 낮은 언어 능숙도 스코어를 결과로 하며, TTS는 다양한 작업들을 이해하고 완료하기 위해 사용자의 능력을 증가시킬 수 있는 적응된, 단순화된 출력을 사용할 수 있다.

[0032] 도시된 바와 같이, 사용자 데이터(108a)는 상기 사용자에 의해 제출된 이전의 텍스트 쿼리들 내에서 사용된 단어들, 영어 또는 TTS 시스템에 의해 활용된 임의의 다른 언어가 사용자의 모국어인지 여부 및 사용자의 언어 이해 능력을 반영하는 활동들 및/또는 행동들의 세트를 포함할 수 있다. 예를 들면, 도 1에 도시된 바와 같이, 사용자의 타이핑 속도는 언어에서 사용자의 언어 유창성을 결정하는데 사용될 수 있다. 추가로, 언어 어휘력 복잡도 스코어 또는 언어 능숙도 스코어가 이전의 텍스트 쿼리들에서 사용자에 의해 사용된 단어들에 미리 결정된 복잡도를 연관시키는 것에 기초하여 사용자에게 할당될 수 있다. 다른 예시에서, 이전 쿼리들에서 미인식 단어들(misrecognized words)의 수 또한 언어 능숙도 스코어를 결정하는데 사용될 수 있다. 예를 들어, 높은 수의 미인식 단어들은 낮은 언어 능숙도를 표시하는데 사용될 수 있다. 일부 구현예들에서, 언어 능숙도 스코어는 쿼리(104)를 제출하기 전에 사용자에 대해 결정되었던 사용자와 연관된 저장된 스코어를 조회함으로써 결정된다.

[0033] 비록 도 1은 언어 능숙도 예측기(110)를 TTS 엔진(120)과 별개의 컴포넌트로서 도시하였지만, 일부 구현예들에서, 도 2에 도시된 바와 같이, 언어 능숙도 예측기(110)는 TTS 엔진(120) 내에서 소프트웨어 모듈에 통합될 수 있다. 상기 예시들에서, 언어 능숙도 예측과 관련된 동작들은 TTS 엔진(120)에 의해 직접적으로 모듈화될 수 있다.

[0034] 일부 구현예들에서, 사용자에게 할당된 언어 능숙도 스코어는 사용자에 대해 예측된 특정한 사용자 컨텍스트에 기초할 수 있다. 예를 들어, 도 2와 관련하여 더 자세히 기술된 바와 같이, 사용자 컨텍스트 결정은 사용자가 일시적으로 제한된 언어 이해 능력을 가지게 할 수 있는 컨텍스트 특정 언어 능숙도들을 결정하는데 사용될 수 있다. 예를 들어, 사용자 컨텍스트가 상당한 배경 소음을 표시하거나 또는 사용자가 운전과 같은 작업에 관계된 경우, 언어 능숙도 스코어는 사용자의 현재 언어 이해 능력이 다른 사용자 컨텍스트들에 비해 일시적으로 감소되었다고 표시하는데 사용될 수 있다.

[0035] 일부 구현예들에서, 이전의 사용자 활동에 기초하여 언어 능숙도를 추론하는 대신에, 언어 능숙도 스코어는 언어 능숙도 예측기(110)의 사용 없이 TTS 엔진(120)에 직접적으로 제공될 수 있다. 예를 들어, 언어 능숙도 스코어는 언어 능숙도의 사용자 레벨을 특정하는 등록 프로세스 동안의 사용자 입력에 기초하여 사용자에게 지정될 수 있다. 예를 들면, 등록 중에, 사용자는 사용자의 능력 레벨을 특정하는 선택을 제공할 수 있고, 이는 그 후 상기 사용자에 대한 적절한 언어 능숙도를 계산하는데 사용될 수 있다. 다른 예시들에서, 사용자는 언어 능숙도의 사용자의 레벨을 특정하는데 사용될 수 있는 인구통계학적 정보, 교육 수준, 거주지 등과 같은 다른 유형의 정보를 제공할 수 있다.

[0036] 상기 기술된 예시들에서, 언어 능숙도 스코어는 최근 생성된 사용자 활동 데이터에 기초하여 주기적으로 조절되는 이산적 값들의 세트이거나 등록 프로세스 동안에 초기적으로 지정된 계속적 스코어일 수 있다. 첫 번째 예시에서, 언어 능숙도 스코어의 값은 사용자의 현재 언어 이해력을 표시하는 하나 이상의 팩터들에 기초하여 편향될 수 있고 능숙도는 약화될 수 있다(예를 들어, 사용자 컨텍스트가 상당한 배경 소음을 표시). 두 번째 예시에서, 언어 능숙도 스코어의 값은 초기 계산 이후에 미리 맞춰질 수 있고, 사용자의 언어 능숙도가 증가되었다는 것을 표시하는(예를 들어, 타이핑 비율의 증가 또는 해당 언어에 대한 보정 비율 감소) 특정 중대 이벤트들 이후에만 조절될 수 있다. 다른 예시에서, 이들 2개의 기법들의 조합이 특정한 텍스트 입력에 기초하여 텍스트-투-스피치 출력을 다양하게 조절하는데 사용될 수 있다. 상기 예시에서, 사용자의 언어 능력의 특정한 양태를 각각 나타내는 다수의 언어 능숙도 스코어들이 사용자에 대한 텍스트-투-스피치 출력을 어떻게 가장 잘 조절할 지 결정하는데 사용될 수 있다. 예를 들면, 하나의 언어 능숙도 스코어는 사용자의 어휘력의 복잡도를 나타낼 수 있는 반면, 다른 언어 능숙도 스코어는 사용자의 문법 능력을 나타내는데 사용될 수 있다.

[0037] TTS 엔진(120)은 언어 능숙도 스코어를 사용하여 사용자의 언어 능숙도 스코어에 의해 표시된 언어 능숙도에 적응된 텍스트-투-스피치 출력를 생성할 수 있다. 일부 예시에서, TTS 엔진(120)은 텍스트 쿼리(104)에 대한 후보 TTS의 세트로부터 특정한 TTS 스트링을 선택하는 것에 기초하여 텍스트-투-스피치 출력을 적응시킬 수 있다. 상기 예시에서, TTS 엔진(120)은 사용자의 언어 능숙도 스코어를 사용하는 것에 기초하여 특정한 TTS 스트링을 선택하여 후보 TTS 스트링들 각각이 사용자에 의해 정확하게 해석될 가능성을 예상한다. 본 기법들과 관련된 더 특정한 설명들이 도 2를 참조하여 제공된다. 대안적으로, 다른 예시에서, TTS 엔진(120)은 기본(baseline) TTS 스트링을 선택할 수 있고, 언어 능숙도 스코어에 의해 표시된 언어 능숙도의 사용자의 레벨에 기초하여 TTS 스트링의 구조를 조절할 수 있다. 상기 예시에서, TTS 엔진(120)은 기본 TTS 스트링의 문법을 조절하여, 대체 단어를 제공하고 및/또는 문장의 복잡도를 감소시켜 사용자에게 이해될 가능성이 더 큰 적응된 TTS 스트링을 생성할 수 있다. 본 기법들과 관련된 더 특정한 설명들이 도 3를 참조하여 제공된다.

[0038] 도 1을 참조하면, TTS 엔진(120)은 사용자들에 대한 언어 능숙도 스코어들이 상이하기 때문에, 사용자들(102a 및 102b)에 대해 상이한 텍스트-투-스피치 출력들을 생성할 수 있다. 예를 들면, 프로세스(100A)에서, 언어 능숙도 스코어(106a)는 높은 영어-언어 능숙도를 표시하며, 이는 사용자(102a)가 복잡한 어휘력를 가졌고, 영어가 모국어이고, 이전의 사용자 쿼리들에서 분당 상대적으로 많은 단어를 가졌다고 표시하는 사용자 데이터(108a)로부터 추론된다. 언어 능숙도 스코어(106a)의 값에 기초하여, TTS 엔진(120)은 복잡한 문법 구조를 포함하는 높은 복잡도의 텍스트-투-스피치 출력(108a)을 생성한다. 도시된 바와 같이, 텍스트-투-스피치 출력(108a)은 오늘의 일기예보가 맑다고 기술하는 독립절(independent clause)을 포함하며, 추가로 일중 최고 온도와 최저 온도에 관한 추가 정보를 포함하는 종속절(subordinate clause)를 포함한다.

[0039] 프로세스(100B)의 예시에서, 언어 능숙도 스코어(106a)는 낮은 영어-언어 능숙도를 표시하며, 이는 사용자(102b)가 단순한 어휘력를 가졌고, 영어가 제2 외국어이고, 이전의 10개의 부정확한 쿼리들을 제공했다고 표시하는 사용자 데이터(108b)로부터 추론된다. 이 예시에서, TTS 엔진(120)은 텍스트-투-스피치 출력(108a)에 비해 보다 단순한 문법 구조를 포함하는 낮은 복잡도의 텍스트-투-스피치 출력(108b)을 생성한다. 예를 들어, 단일의 문장 내에서 다수의 절들을 포함하는 대신에, 텍스트-투-스피치 출력(108b)은 텍스트-투-스피치 출력(108a)와 동일한 주요 정보를 전달하지만, 그 날의 최고 및 최저 온도에 관한 추가 정보는 포함하지 않는 단순한 독립절을 포함한다.

[0040] TTS 출력에 대한 텍스트의 적응은 다양한 상이한 디바이스들 및 소프트웨어 모듈들에 의해 수행될 수 있다. 예를 들면, 서버 시스템의 TTS 엔진은 언어 능숙도 스코어에 기초하여 텍스트를 조절하고 그 후 상기 조절된 텍스트의 합성된 발언을 포함하는 오디오를 출력하는 기능을 포함할 수 있다. 다른 예시로서, 서버 시스템의 프리-프로세싱 모듈은 텍스트를 조절하고 상기 조절된 텍스트를 스피치 합성을 위해 TTS 엔진에 패스할 수 있다. 다른 예시로서, 사용자 디바이스는 적절한 TTS 출력들을 생성할 수 있는 TTS 엔진 또는 TTS 엔진 및 텍스트 프리-

프로세서를 포함할 수 있다.

[0041] 일부 구현예들에서, TTS 시스템은 클라이언트 디바이스의 서드 파티 모바일 어플리케이션 또는 웹페이지와 통신을 교환도록 구성된 소프트웨어 모듈들을 포함할 수 있다. 예를 들어, 시스템의 TTS 기능은 어플리케이션 패키지 인터페이스(API)를 통해 서드 파티 모바일 어플리케이션이 사용하도록 만들어질 수 있다. API는 어플리케이션 또는 웹사이트가 TTS 엔진(120)을 실행하는 서버 시스템으로부터 TTS 오디오를 요청하는데 사용할 수 있는 프로토콜들의 정의된 세트를 포함할 수 있다. 일부 구현예들에서, API는 TTS 기능이 사용자의 디바이스에서 로컬적으로 실행하게 할 수 있다. 예를 들면, API는 인터-프로세스 통신(IPC), 원격 프로시저 콜(RPC) 또는 기타 시스템 콜 또는 기능을 통해 어플리케이션 또는 웹페이지에 사용가능해질 수 있다. TTS 엔진 및 연관된 언어 능숙도 분석 또는 텍스트 프리프로세싱은 사용자의 디바이스에서 로컬적으로 실행되어 사용자의 언어 능숙도에 대한 적절한 텍스트를 결정하고 합성된 스피치에 대한 오디오 또한 생성한다.

[0042] 예를 들면, 서드 파티 어플리케이션 또는 웹페이지는 API를 사용하여 서드 파티 어플리케이션 또는 웹페이지의 음성 인터페이스의 작업 흐름에 기초하여 사용자에게 제공되는 음성 명령어들의 세트를 생성한다. API는 어플리케이션 또는 웹페이지가 스피치로 변환될 텍스트를 제공해야 한다는 것을 특정할 수 있다. 일부 예시에서, 사용자 식별자 또는 언어 능숙도 스코어와 같은 다른 정보가 제공될 수 있다.

[0043] TTS 엔진(120)이 API를 사용하여 서드 파티 어플리케이션과 통신을 교환하는 구현예들에서, TTS 엔진(120)은 서드 파티 어플리케이션으로부터의 텍스트 세그먼트가 상기 텍스트에 대한 텍스트-투-스피치 출력를 생성하기 전에 조절되어야만 하는지 결정하는데 사용될 수 있다. 예를 들면, API는 적응적 텍스트-투-스피치 출력의 생성을 개시하는 서드 파티 어플리케이션 내의 조건들을 특정하는 컴퓨터로 구현되는 프로토콜들을 포함할 수 있다.

[0044] 예시로서, 일 API는 어플리케이션이 다수의 상이한 텍스트 세그먼트들을 TTS 출력을 위한 후보들로서 제출하게 하며, 여기서 상기 상이한 텍스트 세그먼트들은 언어 능숙도의 상이한 레벨들에 대응한다. 예를 들면, 상기 후보들은 동일한 의미를 가지지만 상이한 복잡도 레벨들을 가지는 텍스트 세그먼트일 수 있다(예를 들면, 높은 복잡도 응답, 중간 복잡도 응답 및 낮은 복잡도 응답). 그 후, TTS 엔진(120)은 각 후보를 이해는데 필요한 언어 능숙도의 레벨을 결정하고, 상기 사용자에 대한 적절한 언어 능숙도 스코어를 결정하고, 상기 언어 능숙도 스코어에 가장 잘 대응하는 후보 텍스트를 선택할 수 있다. 그 후, TTS 엔진(120)은 상기 선택된 텍스트에 대한 합성된 오디오를 다시 어플리케이션에 예를 들어, API를 사용하여 네트워크를 통해 제공한다. 일부 예시에서, API는 사용자 디바이스들(106a 및 106b)에 로컬적으로 사용가능할 수 있다. 상기 예시에서, API는 다양한 유형의 인터-프로세스 통신(IPC) 을 통해 또는 시스템 콜을 통해 엑세스가능할 수 있다. 예를 들면, API가 사용자 디바이스들(106a 및 106b)에서 로컬적으로 동작하므로, 사용자 디바이스들(106a 및 106b) 상의 API의 출력은 TTS 엔진(120)의 텍스트-투-스피치 출력일 수 있다.

[0045] 다른 예시에서, API는 서드 파티 어플리케이션으로 하여금 단일의 텍스트 세그먼트 및 TTS 엔진(120)이 텍스트 세그먼트를 수정하여 상이한 복잡도를 가지는 텍스트 세그먼트를 생성하는 것이 허용되었는지를 표시하는 값을 제공하게 할 수 있다. 어플리케이션 또는 웹페이지가 변경이 허용되었다고 표시하면, TTS 시스템(120)은 예를 들면, 언어 능숙도 스코어가 원 텍스트가 사용자가 발화된 응답에서 이해할 수 있는 것보다 더 복잡하다고 제안한 경우 텍스트의 복잡도를 감소시키는 것과 같은 다양한 변경을 텍스트에 가할 수 있다. 다른 예시에서, 또한 API는 서드 파티 어플리케이션으로 하여금 사용자 데이터(예를 들어, 서드 파티 어플리케이션에 제출된 이전의 사용자 쿼리들)를 텍스트 세그먼트와 함께 제공하게 하여, TTS 엔진(120)이 사용자와 연관된 사용자 컨텍스트를 결정하고 상기 결정된 사용자 컨텍스트에 기초하여 특정한 텍스트-투-스피치 출력을 생성하도록 조절한다. 유사하게, API는 어플리케이션으로 하여금 사용자 디바이스로부터의 컨텍스트 데이터(예를 들어, GPS, 가속도계 데이터, 주변 소음 레벨 등) 또는 사용자 컨텍스트의 표시를 제공하게 하여, TTS 엔진(120)으로 하여금 서드 파티 어플리케이션을 통해 사용자에게 궁극적으로 제공될 텍스트-투-스피치 출력을 조절하게 한다. 일부 예시에서, 서드 파티 어플리케이션은 API에 사용자의 언어 능숙도를 결정하는데 사용될 수 있는 데이터를 제공할 수 있다.

[0046] 일부 구현예들에서, TTS 엔진(120)은 사용자의 언어 능숙도를 사용하거나 사용자와 연관된 컨텍스트를 결정함 없이 사용자 쿼리에 대한 텍스트-투-스피치 출력을 조절할 수 있다. 상기 구현예에서, TTS 엔진(120)은 사용자가 출력을 이해하지 못한다는 신호들을 수신함에 기초하여(예를 들어, 동일한 쿼리 또는 작업에 대해 다수의 재시도), 초기 텍스트-투-스피치 출력이 사용자에게 너무 복잡하다고 결정할 수 있다. 응답으로, TTS 엔진(120)은 재시도된 쿼리 또는 관련된 쿼리들에 대해 후속 텍스트-투-스피치 응답의 복잡성을 감소시킬 수 있다. 따라서, 사용자가 액션을 성공적으로 완료하는 것을 실패한 경우, TTS 엔진(120)은 세부 내용의 양 또는 TTS 출력을 이해하는데 필요한 언어 능숙도를 그것이 사용자가 이해하는 레벨에 이를때까지 점진적으로 감소시킬 수 있다.

[0047] 도 2는 사용자 컨텍스트에 기초하여 텍스트-투-스피치 출력을 적응적으로 생성하기 위한 시스템(200)의 예시를 도시하는 도면이다. 간략하게, 시스템(200)은 쿼리 분석기(211), 언어 능숙도 예측기(212), 보간기(interpolator)(213), 언어학 분석기(linguistics analyzer)(214), 재-랭커(re-ranker)(215) 및 파형 생성기(216)를 포함하는 TTS 엔진(210)을 포함한다. 시스템(200)은 또한 컨텍스트 프로필들(232)의 세트를 저장하는 컨텍스트 저장소(220) 및 사용자 이력 데이터(234)를 저장하는 사용자 이력 관리자(230)를 포함한다. 일부 예시에서, TTS 엔진(210)은 도 1와 관련하여 기술된 TTS 엔진(120)과 대응된다.

[0048] 예시에서, 사용자(202)는 그 날에 사용자의 첫 번째 미팅과 관련된 정보에 대한 요청을 포함하는 쿼리(204)를 사용자 디바이스(208)에서 초기적으로 제출한다. 그 후 사용자 디바이스(208)는 쿼리(204) 및 사용자(202)와 연관된 컨텍스트 데이터(206)를 쿼리 분석기(211)와 언어 능숙도 예측기(212)에 각각 전송한다. 쿼리들에 대한 응답들이 아닌 다른 유형의 TTS 출력 예를 들어, 일정 리마인더들, 통지들, 작업 업무흐름들 등이 동일한 기법들을 사용하여 적응될 수 있다.

[0049] 컨텍스트 데이터(206)는 반복된 텍스트 쿼리들 간의 시간 간격, 사용자(202)와 연관된 위치, 속도 또는 움직임 패턴을 표시하는 GPS 데이터, 특정한 시간 기간 내에 TTS 엔진(210)에 제출된 이전의 텍스트 쿼리들 또는 다른 유형의 TTS 엔진(210)과 관련된 사용자 활동을 표시할 수 있는 배경 정보와 같은 사용자(202)와 연관된 특정한 컨텍스트와 관련된 정보를 포함할 수 있다. 일부 예시에서, 컨텍스트 데이터(206)는 쿼리(204)가 사용자 액션과 연관된 텍스트 세그먼트인지 또는 텍스트-투-스피치 출력을 생성하기 위해 TTS 엔진(210)에 전송된 명령어인지 여부와 같은 TTS 엔진(210)에 제출된 쿼리(204)의 유형을 표시할 수 있다.

[0050] 쿼리(204)를 수신한 후에, 쿼리 분석기(211)는 쿼리(204)를 파싱하여 쿼리(204)에 대한 응답인 정보를 식별한다. 예를 들면, 쿼리(204)가 음성 쿼리인 일부 예시에서, 쿼리 분석기(211)는 음성 쿼리의 전사(transcription)를 초기적으로 생성하고, 그 후 예를 들면, 상기 쿼리를 검색 엔진에 제공하고 검색 결과들을 수신함으로써 쿼리(204) 내에서 개별 단어들 또는 세그먼트들을 프로세싱하여 쿼리(204)에 대한 응답인 정보를 결정한다. 그 후 상기 쿼리의 전사 및 상기 식별된 정보는 언어학 분석기(214)에 전송될 수 있다.

[0051] 이제 언어 능숙도 예측기(212)를 참조하면, 컨텍스트 데이터(206)를 수신한 후에, 언어 능숙도 예측기(212)는 도 1과 관련하여 기술된 기법들을 사용하여 수신된 컨텍스트 데이터(206)에 기초하여 사용자(202)에 대한 언어 능숙도를 계산한다. 특히, 언어 능숙도 예측기(212)는 저장소(220)에 저장된 다양한 컨텍스트 프로필들(232)을 통해 파싱한다. 컨텍스트 프로필(232)은 특정한 사용자 컨텍스트와 연관되고 텍스트-투-스피치 출력에 포함될 수 있는 정보의 관련된 유형들을 포함하는 보관된 라이브러리일 수 있다. 컨텍스트 프로필(232)은 각 유형의 정보와 연관된, 사용자(202)가 현재 컨텍스트 프로필(232)과 연관된 컨텍스트 내에 있는 경우 각 유형의 정보가 사용자(202)에 의해 이해될 가능성이 높은지를 나타내는 값을 추가적으로 특정한다.

[0052] 도 2에 도시된 예시에서, 컨텍스트 프로필(232)은 사용자(202)가 그/그녀의 직장에 출근하는 중이라고 표시하는 컨텍스트에 사용자(202)가 현재 있다는 것을 특정한다. 추가로, 컨텍스트 프로필(232)은 또한 사용자(202)에 의해 이해될 가능성이 높은 개별 단어들 및 문구들에 대한 값들을 특정한다. 예를 들어, 날짜 또는 시간 정보는 "SINCE"에 대해 "0.9"의 값과 연관되어, 사용자(202)가 미팅과 연관된 상세한 정보(예를 들어, 미팅 참석인 또는 미팅의 위치) 보다는 미팅과 연관된 일반화된 정보(예를 들어, 다음 다가오는 미팅의 시간)를 이해할 가능성이 높다는 것을 표시한다. 이 예시에서, 값들의 차이점들은 특정한 유형의 정보를 이해하는 사용자의 능력의 차이들을 표시하며, 이는 복잡하거나 상세한 정보를 이해하는 사용자의 능력이 감소되었기 때문이다.

[0053] 개별 단어들 및 문구들과 연관된 값은 사용자(202)가 컨텍스트 데이터(206)에 의해 표시된 컨텍스트에 이전에 있었던 경우 이전의 사용자 세션들로부터의 사용자 활동 데이터에 기초하여 결정될 수 있다. 예를 들어, 이력 사용자 데이터가 쿼리 로그(234) 내에 저장된 데이터를 검색하는 사용자 이력 관리자(230)로부터 전송될 수 있다. 예시에서, 날짜 및 시간 정보에 대한 값은 사용자가 일반적으로 미팅들과 연관된 날짜 및 시간 정보에 미팅들의 위치들 보다 더 자주 엑세스한다는 것을 결정함에 기초하여 증가될 수 있다.

[0054] 언어 능숙도 예측기(212)가 상기 수신된 컨텍스트 데이터(206)에 대응하는 특정한 컨텍스트 프로필(232)을 선택한 후에, 언어 능숙도 예측기(212)는 상기 선택된 컨텍스트 프로필(232)을 보간기(213)에 전송한다. 보간기(213)는 상기 선택된 컨텍스트 프로필(232)을 파싱하고, 포함된 개별 단어들 및 문구들 및 그들의 연관된 값들을 추출한다. 일부 예시에서, 보간기(213)는 텍스트-투-스피치 출력 후보들(240a)을 생성하기 위해 상이한 유형의 정보 및 연관된 값들을 언어학 분석기(214)에 직접 전송한다. 상기 예시에서, 보간기(213)는 상기 선택된 컨텍스트 프로필(232)로부터 특정 유형의 정보 및 연관된 값들을 추출하고, 그들을 언어학 분석기(214)에 전송한

다. 다른 예시에서, 보간기(213)는 또한 상기 선택된 컨텍스트 프로필(232)을 재-랭커(215)에 전송할 수 있다.

[0055] 일부 예시에서, TTS 엔진(210)에 구조화된 데이터의 세트(예를 들어, 일정 이벤트의 필드들)가 제공될 수 있다. 상기 예시에서, 보간기(213)는 상기 구조화된 데이터를 상기 컨텍스트 프로필(232)에 의해 표시된 사용자의 능숙도와 매칭되는 레벨에서 텍스트로 변환할 수 있다. 예를 들면, TTS 엔진(210)은 구조화된 데이터에서 정보를 표현하기 위해 상이한 레벨들의 세부사항 또는 복잡도를 표시하는 하나 이상의 문법들을 표시하는 데이터에 엑세스할 수 있고, 사용자의 언어 능숙도 스코어에 기초하여 적절한 문법을 선택할 수 있다. 유사하게, TTS 엔진(210)은 주어진 언어 능숙도 스코어에 적절한 단어들을 선택하기 위해 사전들을 사용할 수 있다.

[0056] 언어학 분석기(214)는 쿼리(204) 내에 포함된 정보에 정규화(normalization)와 같은 동작들을 프로세싱하는 것을 수행한다. 예를 들어, 쿼리 분석기(211)는 쿼리(204) 내에 포함된 각 단어 또는 스니펫에 대한 표음 전사들을 할당할 수 있고, 쿼리(204)를 텍스트-투-페놈 변환을 사용하여 구, 절 및 문장과 같은 운율적 유닛들로 분할할 수 있다. 언어학 분석기(214)는 또는 쿼리(204)에 대한 응답인 것으로 식별된 다수의 텍스트-투-스피치 출력 후보들을 포함하는 리스트(240a)를 생성한다. 예시에서, 리스트(240a)는 상이한 레벨의 복잡도를 가지는 다수의 텍스트-투-스피치 출력 후보들을 포함한다. 예를 들면, 응답 "듀퐁 서클 부근에서 미스터 존과 오후 12시에"는 미팅에 대한 시간, 미팅에 대한 위치, 미팅이 이루어질 상대방을 식별하기 때문에 가장 복잡한 응답이다. 비교하여, 응답 "3시간 후"는 단지 미팅에 대한 시간만 식별하기 때문에 최소한으로 복잡하다.

[0057] 리스트(240a)는 또한 각 텍스트-투-스피치 출력 후보가 쿼리(204)에 대한 응답일 가망이 있다는 가능성에 기초하여 텍스트-투-스피치 후보들에 대한 기본 랭킹을 포함한다. 예시에서, 리스트(240a)는 가장 복잡한 텍스트-투-스피치 출력 후보가 그것이 쿼리(204)의 컨텐츠와 연관된 가장 많은 양의 정보를 포함하기 때문에 쿼리(204)에 대한 응답일 가능성이 가장 높다고 표시한다.

[0058] 언어학 분석기가 텍스트-투-스피치 출력 후보들의 리스트(240a)를 생성한 후에, 재-랭커(215)는 리스트(240b)를 생성하며, 이는 상기 수신된 컨텍스트 데이터(206)에 기초하여 텍스트-투-스피치 출력 후보들에 대한 랭킹을 조절한다. 예를 들어, 재-랭커(215)는 상기 선택된 컨텍스트 프로필(232) 내에 포함된 특정한 유형의 정보와 연관된 스코어들에 기초하여 랭킹을 조절할 수 있다.

[0059] 예시에서, 재-랭커(215)는 사용자가 출근 중이라고 표시하는 사용자의 현재 컨텍스트를 고려할 때, 텍스트-투-스피치 응답 내에서 날짜와 시간 정보를 이해할 가능성이 크지만 텍스트-투-스피치 응답 내에서 참석자명 또는 위치 정보를 이해할 가능성이 낮다고 표시하는 컨텍스트 프로필(232)에 기초하여, 사용자(202)가 가장 단순한 텍스트-투-스피치 출력을 가장 높게 랭킹할 수 있다. 이 점에서, 수신된 컨텍스트 데이터(206)는 사용자(202)가 TTS 엔진(210)의 텍스트-투-스피치 출력(204c)의 컨텐츠를 이해할 가능성을 증가시키기 위해 특정한 텍스트-투-스피치 출력 후보의 선택을 조절하는데 사용될 수 있다.

[0060] 도 3은 텍스트-투-스피치 출력 내에서 문장 구조를 수정하기 위한 시스템(300)의 예시를 도시하는 도면이다. 간략하게, TTS 엔진(310)은 쿼리(302) 및 사용자(예를 들어, 사용자(202)))에 대한 언어 능숙도 프로필(304)을 수신한다. 그 후, TTS 엔진(310)은 쿼리(302)에 대한 응답인 조절된 텍스트-투-스피치 출력(302c)을 생성하기 위해 동작들(312, 314 및 316)을 수행한다. 일부 예시에서, TTS 엔진(310)은 도 1과 관련하여 기술된 TTS 엔진(120) 또는 도 2와 관련하여 기술된 TTS 엔진(210)과 대응된다.

[0061] 일반적으로, TTS 엔진(310)은 상이한 유형의 조절 기법들을 사용하여 쿼리(302)에 대한 기본 텍스트-투-스피치 출력(306a)의 문장 구조를 수정할 수 있다. 예시로서, TTS 엔진(310)은 개별 단어들 또는 문구들과 연관된 복잡도 스코어가 사용자의 언어 복잡도 프로필(304)에 의해 표시된 임계 스코어보다 크다는 결정에 기초하여 기본 텍스트-투-스피치 출력(306a) 내의 단어들 또는 문구들을 대체할 수 있다. 다른 예시로서, TTS 엔진(310)은 개별적 문장 절들을 재배열하여, 기본 텍스트-투-스피치 출력(306a)의 종합적 복잡도가 언어 복잡도 프로필(304)에 기초한 만족 레벨로 감소되도록 한다. 또한 TTS 엔진(310)은 단어들을 재순서화하고, 문장들을 분할하거나 결합하고, 텍스트의 복잡도를 조절하기 위해 기타 변경들을 가할 수 있다.

[0062] 더 상세히, 동작(312) 동안에, TTS 엔진(310)은 쿼리(302)에 대한 응답인 기본 텍스트-투-스피치 출력(306a)을 초기적으로 생성한다. 그 후, TTS 엔진(310)은 상기 기본 텍스트-투-스피치 출력(306a)을 세그먼트들(312a-312c)로 파싱한다. 또한, TTS 엔진(310)은 개별 세그먼트들 간에 구획점(breakpoint)을 표시하는 구두점들(예를 들어, 쉼표, 마침표, 세미콜론 등)을 검출한다. 또한, TTS 엔진(310)은 세그먼트(312a-312c) 각각에 대한 복잡도 스코어를 계산한다. 일부 예시에서, 복잡도 스코어는 특정한 언어 내에서 특정한 단어의 빈도에 기초하여 계산될 수 있다. 대안적 기법들은 사용자에 의한 사용의 빈도 또는 사용자에 의해 엑세스된 이력적 컨텐츠(예를

들어, 뉴스 기사, 웹페이지 등)에서 출현의 빈도에 기초하여 복잡도 스코어를 계산하는 것을 포함할 수 있다. 이들 예시들 각각에서, 복잡도 스코어는 사용자에 의해 이해될 가능성이 높은 단어들과 사용자에 의해 이해될 가능성이 낮은 다른 단어들을 표시하는데 사용될 수 있다.

[0063] 예시에서, 세그먼트(312a 및 312b)는 각각 "예보(forecast)" 및 "일정(consistent)"과 같은 높은 복잡한 용어들의 포함에 기초하여 상대적으로 복잡한 것으로 결정된다. 그러나, 세그먼트(312c)는 포함된 용어들이 상대적으로 단순하므로 상대적으로 단순한 것으로 결정된다. 이 결정은 세그먼트(312c)에 대한 복잡도 스코어(예를 들어, 0.41)에 비해 높은 복잡도 스코어들(예를 들어, 0.83, 0.75)을 가지는 세그먼트(312a 및 312b)에 의해 표현된다.

[0064] 상기 기술된 바와 같이, 언어 능숙도 프로필(304)은 사용자가 이해가능한 최대 복잡도를 표시하는 임계 복잡도 스코어를 계산하는데 사용될 수 있다. 예시에서, 임계 복잡도 스코어는 "0.7"인 것으로 계산되어, TTS(310)가 세그먼트들(312a 및 312b)은 사용자에 의해 이해될 가능성이 낮다고 결정하도록 한다.

[0065] 언어 능숙도 프로필(304)에 의해 표시된 임계 복잡도 스코어보다 큰 복잡도 스코어들과 연관된 개별 세그먼트를 식별한 후에, 동작(314) 동안에, TTS 엔진(310)은 식별된 단어들을 사용자에 의해 이해될 가능성이 더 높은 것으로 예상되는 대체 단어들로 교체한다. 도 3에서 도시된 바와 같이, "예보"는 "날씨(weather)"로 교체될 수 있고, "일정"은 "변함(change)"으로 교체될 수 있다. 이 예시들에서, 세그먼트들(314a 및 314b)은 언어 능숙도 프로필(304)에 의해 표시된 임계 복잡도 스코어 미만의 보다 낮은 복잡도 스코어를 가지는 보다 단순한 대체재들을 표현한다.

[0066] 일부 구현예들에서, TTS 엔진(310)은 적절히 복잡한 단어들을 매우 복잡한 단어들로 교환하는 것을 결정하기 위해 비감독 기법들을 사용하는 트레이닝된 스킵-그램 모델(skip-gram model)을 사용하여 높은 복잡도 단어들에 대한 단어 교체들을 프로세싱할 수 있다. 또한, 일부 예시에서, TTS 엔진(310)은 동의어 또는 유의어 데이터를 사용하여 높은 복잡도 단어들에 대한 단어 교체를 프로세싱할 수 있다.

[0067] 이제 동작(316)을 참조하면, 쿼리의 문장 절들은 특정한 문장 구조들과 연관된 복잡도들을 계산하는 것과 언어 능숙도 프로필(304)에 의해 표시된 언어 능숙도에 기초하여 사용자가 문장 구조를 이해할 수 있을지 여부를 결정하는 것에 기초하여 조절될 수 있다.

[0068] 예시에서, TTS 엔진(310)은 기본 텍스트-투-스피치 응답(306a)이 3개의 문장 절들(예를 들어, "오늘의 예보는 맑음","그러나 일정하지 않음" 그리고 "따뜻함")을 포함한다고 결정함에 기초하여, 기본 텍스트-투-스피치 응답(306a)이 높은 문장 복잡도를 가진다고 결정한다. 응답으로, TTS 엔진(310)은 종속절 및 독립절을 세그먼트 구두점을 포함하지 않는 단일의 절로 결합하는 조절된 문장 부분들(316a 및 316b)을 생성할 수 있다. 그 결과로서, 상기 조절된 텍스트-투-스피치 응답(306b)은 보다 단순한 어휘(예를 들어, "날씨", "변함")를 포함할 뿐만 아니라 보다 단순한 문장 구조(예를 들어, 분리된 절이 없음)을 포함하며, 사용자가 상기 조절된 텍스트-투-스피치 출력(306b)을 이해할 가능성을 증가시킨다. 그 후, 상기 조절된 텍스트-투-스피치 출력(306b)은 TTS 엔진(310)에 의해 출력(306c)로서 출력하기 위해 생성된다.

[0069] 일부 구현예들에서, TTS 엔진(310)은 사용자에 대해 문제될 것으로 식별된 특정한 문장 구조들을 피하기 위해 가중치 팩터들을 사용한 기본 쿼리(302a) 조절을 포함하는 사용자 특정 재구조화 알고리즘을 사용하여 문장 구조 조절을 수행할 수 있다. 예를 들면, 사용자-특정 재구조화 알고리즘은 종속절들의 포함을 감소시키거나 단순한 주어 동사 목적어 문장들을 가지는 문장 절들을 증가시키기 위한 옵션을 특정할 수 있다.

[0070] 도 4는 클러스터링 기법들을 사용하는 것에 기초하여 텍스트-투-스피치 출력들을 적응적으로 생성하기 위한 시스템(400)의 예시를 도시하는 블록도이다. 시스템(400)은 언어 능숙도 예측기(410), 사용자 유사성 결정기(420), 복잡도 최적화기 및 기계 학습 시스템(400)을 포함한다.

[0071] 간략하게, 언어 능숙도 예측기(410)는 복수의 사용자들(402)로부터 데이터를 수신한다. 그 후, 언어 능숙도 예측기(410)는 복수의 사용자들(402) 각각에 대해 언어 복잡도 프로필들(412)의 세트를 예측하고, 이는 그 후 사용자 유사성 결정기(420)에 보내진다. 사용자 유사성 결정기(420)는 유사한 사용자들의 사용자 클러스터들(424)을 식별한다. 그 후, 복잡도 최적화기(430) 및 기계 학습 시스템(400)은 사용자 클러스터들(424) 내의 각 사용자의 언어 복잡도 프로필(412)과 복수의 사용자들(402)로부터 수신된 컨텍스트 데이터를 분석하여 복잡도 맵핑(442)을 생성한다.

[0072] 일반적으로, 시스템(400)은 사용자들의 인구에 대한 능동 언어 복잡도와 수동 언어 복잡도 간의 관계를 분석하는데 사용될 수 있다. 능동 언어 복잡도는 사용자에 의해 제공된 검출된 언어 입력을 지칭한다(예를 들어, 텍스

트 쿼리들, 음성 입력 등). 수동 언어 복잡도는 사용자에게 제공된 스피치 신호들을 알아듣거나 이해하는 사용자의 능력을 지칭한다. 이 점에서, 시스템(400)은 특정 사용자는 텍스트-투-스피치 출력을 이해할 높은 가능성을 가지는 각 개별 사용자에 대한 적절한 수동 언어 복잡도를 결정하기 위해 다수의 사용자들에 대한 능동 언어 복잡도 및 수동 언어 복잡도 간의 결정된 관계를 사용할 수 있다.

[0073] 복수의 사용자들(402)은 TTS 엔진(예를 들어, TTS 엔진(120))과 연관된 어플리케이션을 사용하는 다수의 사용자들일 수 있다. 예를 들어, 복수의 사용자들(402)은 사용자들에게 텍스트-투-스피치 구성들을 모바일 어플리케이션의 사용자 인터페이스를 통해 제공하기 위해 TTS 엔진을 활용하는 모바일 어플리케이션을 사용하는 사용자들의 세트일 수 있다. 상시 예시에서, 복수의 사용자들(402)로부터의 데이터(예를 들어, 이전의 사용자 쿼리들, 사용자 선택들 등)는 모바일 어플리케이션에 의해 추적될 수 있고, 언어 능숙도 예측기(410)에 의한 분석을 위해 수집될 수 있다.

[0074] 언어 능숙도 예측기(410)는 도 1과 관련하여 이전에 기술된 바와 같이 실질적으로 유사한 기법들을 사용하여 복수의 사용자들(402)에 대한 수동 언어 복잡도들을 초기적으로 측정할 수 있다. 그 후, 언어 능숙도 예측기(410)는 언어 복잡도 프로필(412)을 생성할 수 있고, 이는 복수의 사용자들(402) 각각에 대한 개별적 언어 복잡도 프로필을 포함한다. 각 개별 언어 복잡도 프로필은 복수의 사용자들(402) 각각에 대한 수동 언어 복잡도 및 능동 언어 복잡도를 표시하는 데이터를 포함한다.

[0075] 사용자 유사성 결정기(420)는 언어 능숙도 프로필들(412)의 세트 내에 포함된 언어 복잡도 데이터를 사용하여 복수의 사용자들(402) 내에서 유사한 사용자들을 식별한다. 일부 예시에서, 사용자 유사성 결정기(420)는 유사한 능동 언어 복잡도들(예를 들어, 유사한 언어 입력들, 제공된 스피치 쿼리들 등)을 가지는 사용자들을 그룹화할 수 있다. 다른 예시에서, 사용자 유사성 결정기(420)는 이전의 사용자 제출 쿼리들, 모바일 어플리케이션 상의 특정한 사용자 행동들 또는 사용자 위치들에 포함된 단어들을 비교함으로써 유사한 사용자들을 결정할 수 있다. 그 후, 사용자 유사성 결정기(420)는 사용자 클러스터(424)를 생성하기 위해 유사한 사용자들을 클러스터화한다.

[0076] 일부 구현예들에서, 사용자 유사성 결정기(420)는 특정 클러스터들에서 사용자들에 대해 집합된 데이터를 포함하는 저장된 클러스터 데이터(422)에 기초하여 사용자 클러스터(424)를 생성한다. 예를 들어, 클러스터 데이터(422)는 복수의 사용자들(402)과 연관된 수동 언어 복잡도를 표시하는 특정 파라미터들(예를 들어, 부정확한 쿼리 응답들의 수 등)에 의해 그룹화될 수 있다.

[0077] 사용자 클러스터(424)를 생성한 후에, 복잡도 최적화기(430)는 TTS 시스템에 의한 언어 출력의 복잡도를 다양하게 하고, 사용자 능력을 표시하기 위해 TTS 시스템에 의해 언어 출력을 이해하는 사용자의 성과를 표시하는 파라미터들의 세트를 사용하여 사용자의 수동 언어 복잡도를 측정한다. 예를 들어, 파라미터들은 각 클러스터(424) 내의 사용자들이 얼머나 주어진 텍스트-투-스피치 출력을 이해하는지 나타내는데 사용될 수 있다. 상기 예시에서, 복잡도 최적화기(430)는 사용자에 대한 낮은 복잡도 스피치 신호를 초기적으로 제공하고 복잡도들의 범위 내에서 추가적 스피치 신호들을 반복적으로 제공한다.

[0078] 또한, 일부 구현예들에서, 복잡도 최적화기(430)는 각 사용자 클러스터(424)와 연관된 다양한 사용자 컨텍스트들에 대해 최적의 수동 언어 복잡도를 결정할 수 있다. 예를 들어, 파라미터들의 세트를 사용하여 사용자의 언어 능숙도를 측정한 후에, 복잡도 최적화기(430)는 복수의 사용자들(402)로부터 수신된 컨텍스트 데이터에 의해 측정된 데이터를 분류할 수 있고, 따라서 최적의 수동 언어 복잡도가 각 사용자 컨텍스트에 대해 결정될 수 있도록 한다.

[0079] 그 후, 수동 언어 복잡도들의 범위에 대한 성과 데이터를 모은 후에, 기계 학습 시스템(400)은 성과 파라미터들이 사용자의 언어 이해력이 가장 좋다고 표시하는 특정한 수동 언어 복잡도를 결정한다. 예를 들어, 기계 학습 시스템(400)은 특정한 사용자 클러스터(424) 내의 모든 사용자들의 성과 데이터를 수집하여 능동 언어 복잡도, 수동 언어 복잡도 및 사용자 컨텍스트 간의 관계를 결정한다.

[0080] 그 후, 사용자 클러스터(424)에 대한 수집 데이터는 사용자 클러스터(424) 내의 각 사용자에 대한 개별 데이터와 비교되어 사용자 클러스터(424) 내의 각 사용자에 대한 실제 언어 복잡도 스코어를 결정한다. 예를 들어, 도 4에 도시된 바와 같이, 복잡도 맵핑(442)은 최적의 수동 언어 복잡도에 맵핑된 능동 언어 복잡도에 대응하는 실제 언어 복잡도를 추론하기 위해, 능동 언어 복잡도와 수동 언어 복잡도 간의 관계를 표현할 수 있다.

[0081] 복잡도 맵핑(442)은 복수의 사용자(402) 내의 모든 사용자 클러스터들에 대한 능동 언어 복잡도, TTS 복잡도 및 수동 언어 복잡도 간의 관계를 표현하며, 이는 그 후 개별 사용자에 의한 후속 쿼리에 대한 적절한 TTS 복잡도

를 예상하는데 사용될 수 있다. 예를 들어, 상기 기술된 바와 같이, 사용자 입력들(예를 들어, 쿼리들, 텍스트 메시지들, 이메일들 등)은 유사한 사용자들을 사용자 클러스터(424)로 그룹화하는데 사용될 수 있다. 각 클러스터에 대해, 시스템은 이해하기 위해 언어 능숙도의 다양한 레벨들을 요구하는 TTS 출력들을 제공한다. 그 후, 시스템은 사용자들로부터 수신된 응답들, 다양한 TTS 출력들에 대한 작업 완료의 비율을 평가하여, 각 클러스터에서 사용자들에 대해 적절한 언어 복잡도의 레벨을 결정한다. 시스템은 상기 식별된 클러스터들에 대응하는 클러스터 식별자들과 복잡도 스코어들 간의 맵핑(442)을 저장한다. 그 후, 시스템은 복잡도 맵핑(442)을 사용하여 사용자에 대한 TTS 출력에 대한 복잡도의 적절한 레벨을 결정한다. 예를 들면, 시스템은 사용자의 능동 언어 능숙도를 표현하는 클러스터를 식별하고, 맵핑(442)에서 상기 클러스터에 대한 대응 TTS 복잡도 스코어(예를 들어, 수동 언어 이해의 레벨을 표시하는)를 조회하고 및 검색된 TTS 복잡도 스코어에 의해 표시된 복잡도 레벨을 가지는 TTS 출력을 생성한다.

[0082] 그 후 사용자에 대해 결정된 실제 언어 복잡도는 도 1 내지 3과 관련하여 기술된 기법들을 사용하여 TTS 시스템을 조절하는데 사용될 수 있다. 이 점에서, 유사한 사용자들의 그룹(예를 들어, 사용자 클러스터(424))로부터 수집된 언어 복잡도 데이터는 단일의 사용자와 관련된 TTS 시스템의 성능을 지능적으로 조절하는데 사용될 수 있다.

[0083] 도 5는 텍스트-투-스피치 출력을 적응적으로 생성하기 위한 프로세스(500)의 예시를 도시하는 흐름도이다. 간략하게, 프로세스(500)는 클라이언트 디바이스의 사용자의 언어 능숙도를 결정하는 단계(510), 텍스트-투-스피치 모듈에 의한 출력을 위한 텍스트 세그먼트를 결정하는 단계(520), 텍스트 세그먼트의 합성된 발언을 포함하는 오디오 데이터를 생성하는 단계(530) 및 오디오 데이터를 클라이언트 디바이스에 제공하는 단계(540)를 포함할 수 있다.

[0084] 더 상세히, 프로세스(500)는 클라이언트 디바이스의 사용자의 언어 능숙도를 결정하는 단계(510)를 포함한다. 예를 들어, 도 1과 관련하여 기술된 바와 같이, 언어 능숙도 예측기(110)는 다양한 기법들을 사용하여 사용자에 대한 언어 능숙도를 결정할 수 있다. 일부 예시에서, 언어 능숙도는 언어 능숙도의 레벨을 표시하는 할당된 스코어를 표현할 수 있다. 다른 예시에서, 언어 능숙도는 언어 능숙도의 복수의 카테고리들 중 할당된 카테고리를 표현할 수 있다. 다른 예시에서, 언어 능숙도는 사용자의 능숙도 레벨을 표시하는 사용자 입력 및/또는 행동들에 기초하여 결정될 수 있다.

[0085] 일부 구현예들에서, 언어 능숙도는 상이한 사용자 신호들로부터 추론될 수 있다. 예를 들어, 도 1과 관련하여 기술된 바와 같이, 언어 능숙도는 사용자 입력들의 어휘력 복잡도, 사용자의 데이터 입력 비율, 스피치 입력으로부터 미인식 단어들의 수, TTs 복잡도의 상이한 레벨들에 대한 완료된 보이스 액션들의 수 또는 사용자에 의해 열람된 텍스트들의 복잡도의 레벨(예를 들어, 책, 기사 웹페이지 상의 텍스트 등)로부터 추론될 수 있다.

[0086] 프로세스(500)는 텍스트-투-스피치 모듈에 의한 출력을 위한 텍스트 세그먼트를 결정하는 단계(520)를 포함할 수 있다. 예를 들어, TTS 엔진은 사용자의 결정된 언어 능숙도에 기초하여 기본 텍스트 세그먼트를 조절할 수 있다. 일부 예시에서, 도 2와 관련하여 기술된 바와 같이, 출력을 위한 텍스트 세그먼트는 사용자와 연관된 사용자 컨텍스트에 기초하여 조절될 수 있다. 다른 예시에서, 도 3과 관련하여 기술된 바와 같이, 출력을 위한 텍스트 세그먼트는 텍스트 세그먼트의 복잡도를 감소시키기 위해 단어 교체 또는 문장 재구조화에 의해 조절될 수 있다. 예를 들면, 상기 조절은 얼마나 드문 개별 단어들이 텍스트 세그먼트들에 포함되었는지, 사용된 동사들의 유형(예를 들어, 복합 동사, 동사 시제), 텍스트 세그먼트의 언어적 구조(예를 들어, 종속절들의 수, 관련 단어들 간 구분의 수, 구들이 포개진 정도 등에 기초할 수 있다. 또한, 다른 예시에서, 상기 조절은 언어적 특징들(예를 들어, 주어와 동사 간의 평균 구분, 형용사와 명사 간의 구분 등)에 대한 기준 측정과 상기 언어적 측정들에 기초할 수 있다. 상기 예시에서, 기준 측정들은 평균들을 표현할 수 있거나, 상이한 복잡도 레벨들에 대한 범위들 또는 예시들을 포함할 수 있다.

[0087] 일부 구현예들에서, 출력을 위한 텍스트 세그먼트를 결정하는 단계는 사용자의 언어 능숙도 레벨을 기술하는 기준 스코어들과 가장 잘 매칭되는 스코어들을 가지는 텍스트 세그먼트들을 선택하는 것을 포함할 수 있다. 다른 구현예에서, 개별 단어들 또는 문구들은 복잡도에 대해 스코어화될 수 있고, 가장 복잡한 단어들은 교체, 삭제 또는 재구조화되어 종합적 복잡도가 사용자에게 적절한 레벨에 맞도록 될 수 있다.

[0088] 프로세스(500)는 텍스트 세그먼트의 합성된 발언을 포함하는 오디오 데이터를 생성하는 단계(530)를 포함할 수 있다.

[0089] 프로세스(500)는 클라이언트 디바이스에 대한 오디오 데이터를 제공하는 단계(540)를 포함할 수 있다.

[0090] 도 6은 클라이언트로서 또는 서버 또는 복수의 서버들로서 본 문서에서 기술된 시스템들 및 방법들을 구현하는데 사용될 수 있는 컴퓨팅 디바이스들(600, 650)의 블록도이다. 컴퓨팅 디바이스(600)는 랩톱, 데스크톱, 워크스테이션, 개인 휴대 정보 단말기, 서버, 블레이드 서버, 메인 프레임 및 다른 적절한 컴퓨터와 같은 다양한 형태의 디지털 컴퓨터들을 나타내기 위한 것이다. 컴퓨팅 디바이스(650)는 개인 휴대 정보 단말기, 셀룰러 전화기, 스마트폰 및 다른 유사한 컴퓨팅 디바이스들과 같은 다양한 형태의 모바일 디바이스들을 나타내기 위한 것이다. 추가적으로, 컴퓨팅 디바이스(600 또는 650)는 USB 플래시 드라이브들을 포함할 수 있다. USB 플래시 드라이브들은 운영 체제 및 다른 어플리케이션들을 저장할 수 있다. USB 플래시 드라이브들은 무선 전송기 또는 다른 컴퓨팅 디바이스의 USB 포트에 삽입될 수 있는 USB 커넥터와 같은 입력/출력 컴포넌트들을 포함할 수 있다. 여기에 도시된 컴포넌트들, 그들의 연결 및 관계, 및 그들의 기능은 단지 예시적인 것을 의미하며, 본 명세서에 기술된 및/또는 청구된 발명의 구현을 제한하는 것을 의미하지는 않는다.

[0091] 컴퓨팅 디바이스(600)는 프로세서(602), 메모리(604), 저장 디바이스(606), 메모리(604) 및 고속 확장 포트(610)에 연결되는 고속 인터페이스(608) 및 저속 버스(614) 및 저장 디바이스(606)에 연결되는 저속 인터페이스(612)를 포함한다. 컴포넌트들(602, 604, 606, 608, 610 및 612) 각각은 다양한 버스들을 사용하여 상호 연결되고, 공통 마더 보드 상에 또는 적절한 다른 방식으로 장착될 수 있다. 프로세서(602)는 메모리(604) 또는 저장 디바이스(606)에 저장된 명령어들을 포함하는, 컴퓨팅 디바이스(600) 내에서 실행하기 위한 명령어들을 프로세싱하여, 고속 인터페이스(608)에 연결된 디스플레이(616)와 같은 외부 입/출력 디바이스상에 GUI에 대한 그래픽 정보를 디스플레이 할 수 있다. 다른 구현예에서, 다수의 프로세서들 및/또는 다수의 버스들이 다수의 메모리들 및 다수의 유형의 메모리와 함께, 적절하게 사용될 수 있다. 또한, 다수의 컴퓨팅 디바이스들(600)은 필요한 동작의 부분들을 제공하는 각 디바이스와 연결될 수 있다(예를 들어, 서버 뱅크, 블레이드 서버 그룹 또는 멀티 프로세서 시스템).

[0092] 메모리(604)는 컴퓨팅 디바이스(600) 내에 정보를 저장한다. 일 구현예에서, 메모리(604)는 휘발성 메모리 유닛 또는 유닛들이다. 다른 구현예에서, 메모리(604)는 비휘발성 메모리 유닛 또는 유닛들이다. 또한, 메모리(604)는 자기 또는 광학 디스크와 같은 컴퓨터 판독가능 매체의 다른 형태 일 수 있다.

[0093] 저장 디바이스(606)는 컴퓨팅 디바이스(600)에 대한 대형 저장소를 제공할 수 있다. 일 구현예에서, 저장 디바이스(606)는 플로피 디스크 디바이스, 하드 디스크 디바이스, 광 디스크 디바이스 또는 테이프 디바이스, 플래시 메모리 또는 다른 유사한 고체 상태 메모리 디바이스, 또는 저장 영역 네트워크 또는 다른 구성의 디바이스를 포함하는 디바이스의 어레이와 같은 컴퓨터 판독가능 매체이거나 컴퓨터 판독가능 매체를 포함할 수 있다. 컴퓨터 프로그램 제품은 정보 캐리어에 유형적으로 수록될 수 있다. 컴퓨터 프로그램 제품은 또한 실행될 때 상기 기술된 바와 같은 하나 이상의 방법을 수행하는 명령어들을 포함할 수 있다. 정보 캐리어는 메모리(604), 저장 디바이스(606) 또는 프로세서(602)상의 메모리와 같은 컴퓨터 또는 기계 판독가능 매체이다.

[0094] 고속 제어기(608)는 컴퓨팅 디바이스(600)에 대한 대역폭 집중 동작들을 관리하는 반면, 저속 제어기(612)는 낮은 대역폭 집중 동작들을 관리한다. 이러한 기능들의 할당은 단지 예시적인 것이다. 일 구현예에서, 고속 제어기(608)는 메모리(704), 디스플레이(716)(예를 들어, 그래픽 프로세서 또는 가속기를 통해) 및 다양한 확장 카드(도시되지 않음)를 수용할 수 있는 고속 확장 포트(610)에 연결된다. 구현예에서, 저속 제어기(612)는 저장 디바이스(606) 및 저속 확장 포트(614)에 결합된다. 다양한 통신 포트(예를 들어, USB, 블루투스, 이더넷, 무선 이더넷)를 포함할 수 있는 저속 확장 포트는 키보드, 포인팅 디바이스, 마이크로폰/스피커 쌍, 스캐너와 같은 하나 이상의 입력/출력 디바이스 또는 예를 들어 네트워크 어댑터를 통해 스위치 또는 라우터와 같은 네트워킹 디바이스에 결합될 수 있다. 컴퓨팅 디바이스(600)는 도면에 도시된 바와 같이 다수의 상이한 형태로 구현될 수 있다. 예를 들어, 그것은 표준 서버(620)로서 또는 그러한 서버들의 그룹에서 다수로 구현될 수 있다. 또한, 랙 서버 시스템(624)의 일부로서 구현될 수 있다. 또한, 랩톱 컴퓨터(622)와 같은 퍼스널 컴퓨터에서 구현될 수 있다. 대안적으로, 컴퓨팅 디바이스(600)로부터의 컴포넌트들은 디바이스(650)와 같은 모바일 디바이스(도시되지 않음) 내의 다른 컴포넌트들과 결합될 수 있다. 상기 디바이스들 각각은 컴퓨팅 디바이스(600, 650) 중 하나 이상을 포함할 수 있고, 전체 시스템은 서로 통신하는 다수의 컴퓨팅 디바이스들(600, 650)로 구성될 수 있다.

[0095] 컴퓨팅 디바이스(600)는 도면에 도시된 바와 같이 다수의 상이한 형태로 구현될 수 있다. 예를 들어, 그것은 표준 서버(620)로서 또는 그러한 서버들의 그룹에서 다수로 구현될 수 있다. 또한, 랙 서버 시스템(624)의 일부로서 구현될 수 있다. 또한, 랩톱 컴퓨터(622)와 같은 퍼스널 컴퓨터에서 구현될 수 있다. 대안적으로, 컴퓨팅 디바이스(600)로부터의 컴포넌트들은 디바이스(650)와 같은 모바일 디바이스(도시되지 않음) 내의 다른 컴포넌트들과 결합될 수 있다. 상기 디바이스들 각각은 컴퓨팅 디바이스(600, 650) 중 하나 이상을 포함할 수 있고, 전

체 시스템은 서로 통신하는 다수의 컴퓨팅 디바이스들(600, 650)로 구성될 수 있다.

[0096] 컴퓨팅 디바이스(650)는 다른 여러 컴포넌트들 중에서도 특히, 프로세서(652), 메모리(664), 디스플레이(654)와 같은 입/출력 디바이스, 통신 인터페이스(666) 및 송수신기(668)를 포함한다. 디바이스(650)에는 또한 추가적 저장을 제공하기 위해 마이크로 드라이브 또는 다른 디바이스와 같은 저장 디바이스가 제공될 수 있다. 컴포넌트들(650, 652, 664, 654, 666 및 668) 각각은 다양한 버스들을 사용하여 상호 연결되고, 몇몇 컴포넌트들은 공통 마더 보드 상에 또는 적절한 다른 방식으로 장착될 수 있다.

[0097] 프로세서(652)는 메모리(664)에 저장된 명령어들을 포함하는 컴퓨팅 디바이스 (650)내의 명령어들을 실행할 수 있다. 프로세서는 별개의 그리고 다수의 아날로그 및 디지털 프로세서들을 포함하는 칩들의 칩셋으로서 구현될 수 있다. 추가적으로, 프로세서는 임의의 수의 아키텍처들을 사용하여 구현될 수 있다. 예를 들면, 프로세서(610)는 CISC(Complex Instruction Set Computers) 프로세서, RISC(Reduced Instruction Set Computer) 프로세서 또는 MISC(Minimal Instruction Set Computer)프로세서일 수 있다. 프로세서는 예를 들어 사용자 인터페이스들, 디바이스(650)에 의해 실행되는 어플리케이션 및 디바이스(650)에 의한 무선 통신과 같은 디바이스(650)의 다른 컴포넌트들의 조정을 제공할 수 있다.

[0098] 프로세서(652)는 제어 인터페이스(658) 및 디스플레이(656)에 연결된 디스플레이 인터페이스(656)를 통해 사용자와 통신할 수 있다. 디스플레이(654)는 예를 들어, TFT LCD(박막 트랜지스터 액정 디스플레이) 또는 OLED(유기 발광 다이오드) 디스플레이 또는 다른 적절한 디스플레이 기술을 포함할 수 있다. 디스플레이 인터페이스(786)는 사용자에게 그래픽 및 다른 정보를 제공하기 위해 디스플레이(654)를 구동하기 위한 적절한 회로를 포함할 수 있다. 제어 인터페이스(658)는 사용자로부터 명령을 수신하고, 프로세서(652)에 제출하기 위해 그들을 변환할 수 있다. 추가로, 외부 인터페이스(662)는 프로세서(652)와의 통신에 제공되어 다른 디바이스들과 디바이스(650)의 근거리 통신을 가능하게 할 수 있다. 외부 인터페이스(662)는 예를 들면, 일부 구현예들에서는 유선 통신을 위해 또는 다른 구현예들에서는 무선 통신을 위해 제공될 수 있고, 다수의 인터페이스들도 사용될 수 있다.

[0099] 메모리(664)는 컴퓨팅 디바이스(650) 내에 정보를 저장한다. 메모리(664)는 컴퓨터 판독가능 매체 또는 매체들, 휘발성 메모리 유닛 또는 유닛들, 비휘발성 메모리 유닛 또는 유닛들 중 하나 이상으로 구현될 수 있다. 또한 확장 메모리(774)는 예를 들어 SIMM(Single In Line Memory Module) 카드 인터페이스를 포함할 수 있는 확장 인터페이스(772)를 통해 디바이스(650)에 제공되고 접속될 수 있다. 상기 확장 메모리(674)는 디바이스(650)에 대해 추가 저장 공간을 제공할 수 있거나, 또는 어플리케이션들 또는 디바이스(650)에 대한 다른 정보를 저장할 수 있다. 특히, 확장 메모리(674)는 전술한 프로세스들을 수행하거나 보충하는 명령어들을 포함할 수 있으며, 또한 보안 정보를 포함할 수 있다. 따라서, 예를 들면, 확장 메모리(674)는 디바이스(650)에 대한 보안 모듈로서 제공될 수 있고, 디바이스(650)의 보안 사용을 허용하는 명령어들로 프로그래밍될 수 있다. 또한, 보안 어플리케이션들은 SIMM 카드 상의 식별 정보를 해킹할 수 없는 방식으로 배치하는 것과 같이, SIMM 카드들을 통해 추가 정보와 함께 제공될 수 있다.

[0100] 메모리는 예를 들어, 후술되는 바와 같이, 플래시 메모리 및/또는 NVRAM 메모리를 포함할 수 있다. 일 구현예에서, 컴퓨터 프로그램 제품은 정보 캐리어에 유형적으로 수록된다. 컴퓨터 프로그램 제품은 또한 실행될 때 상기 기술된 바와 같은 하나 이상의 방법을 수행하는 명령어들을 포함한다. 정보 캐리어는 예를 들어 송수신기(668) 또는 외부 인터페이스(662)를 통해 수신될 수 있는 메모리(664), 확장 메모리(674) 또는 프로세서(652)상의 메모리와 같은 컴퓨터 또는 기계 판독가능 매체이다.

[0101] 디바이스(650)는 필요에 따라 디지털 신호 처리 회로를 포함할 수 있는 통신 인터페이스(666)를 통해 무선으로 통신할 수 있다. 통신 인터페이스(666)는 다른 것들 중에서도 GSM 보이스 콜들, SMS, EMS 또는 MMS 메시징, CDMA, TDMA, PDC, WCDMA, CDMA2000 또는 GPRS와 같은 다양한 모드들 또는 프로토콜들 하에서의 통신을 위해 제공될 수 있다. 이러한 통신은 예를 들어, 무선 주파수 송수신기(668)를 통해 발생될 수 있다. 추가로, 블루투스, Wi-Fi 또는 다른 트랜시버(도시되지 않음)를 사용하는 것과 같은 단거리 통신이 발생될 수 있다. 추가로, GPS(Global Positioning System) 수신기 모듈(670)은 디바이스(780)상에서 실행되는 어플리케이션들에 의해 적절히 사용될 수 있는 추가적인 네비게이션 및 위치 관련 무선 데이터를 디바이스(650)에 제공할 수 있다.

[0102] 디바이스(650)는 또한 사용자로부터 발화된 정보를 수신하고 그것을 이용가능한 디지털 정보로 변환할 수 있는 오디오 코덱(660)을 사용하여 청각적으로 통신할 수 있다. 오디오 코덱(660)은 마찬가지로, 예를 들어 디바이스(650)의 핸드셋 내의 스피커를 통하는 것과 같이, 사용자를 위한 가청 사운드를 생성할 수 있다. 이러한 사운드

는 보이스 전화 콜들로부터의 사운드 포함할 수 있고, 기록된 사운드(예를 들어, 음성 메시지, 음악 파일 등)를 포함할 수 있고, 또한 디바이스(650)상에서 동작하는 어플리케이션들에 의해 생성된 사운드를 포함할 수 있다.

[0103] 컴퓨팅 디바이스(650)는 도면에 도시된 바와 같이 다수의 상이한 형태로 구현될 수 있다. 예를 들면, 그것은 셀룰러 전화(480)로서 구현될 수 있다. 또한 스마트폰(682), 개인 휴대 정보 단말기(PDA), 또는 다른 유사한 이동 디바이스의 일부로서 구현될 수 있다.

[0104] 본 명세서에 기술된 시스템들 및 방법들의 다양한 구현예들은 디지털 전자 회로, 집적 회로, 특수하게 설계된 ASIC들(application specific integrated circuits), 컴퓨터 하드웨어, 펌웨어, 소프트웨어 및/또는 이들의 조합으로 구현될 수 있다. 이들 다양한 구현예들은 적어도 하나의 프로그래머블 프로세서를 포함하는 프로그래머블 시스템 상에서 실행가능하고 및/또는 인터프리트가능한 하나 이상의 컴퓨터 프로그램들에서의 구현예를 포함할 수 있고, 이는 전용 또는 범용일 수 있고, 저장 시스템, 적어도 하나의 입력 디바이스 및 적어도 하나의 출력 디바이스으로부터 데이터 및 명령어들을 수신하고 그에 데이터 및 명령어들을 전송하기 위해 연결될 수 있다.

[0105] 이들 컴퓨터 프로그램들(프로그램, 소프트웨어, 소프트웨어 어플리케이션 또는 코드로도 알려짐)은 프로그래머블 프로세서에 대한 기계 명령어들을 포함하며, 하이레벨 절차어 및/또는 객체 지향 프로그래밍 언어 및/또는 어셈블리/기계어에서 구현될 수 있다. 본 명세서에서 사용된 바와 같이, 용어 "기계 판독가능 매체", "컴퓨터 판독가능 매체"는 기계 판독가능 신호로서 기계 명령어들을 수신하는 기계 판독가능 매체를 포함하여, 기계 명령어들 및/또는 데이터를 프로그래머블 프로세서에 제공하는데 사용되는 임의의 컴퓨터 프로그램 물, 장치 및/또는 디바이스 예를 들어, 자기 디스크, 광학 디스크, 메모리, 프로그래머블 로직 디바이스(PLD)를 지칭한다. 용어 "기계 판독가능 신호"는 기계 명령어들 및/또는 데이터를 프로그래머블 프로세서에 제공하는데 사용되는 임의의 신호를 지칭한다.

[0106] 사용자와의 인터렉션을 제공하기 위해, 본 명세서에서 기술된 시스템들 및 기법들은 사용자에게 정보를 디스플레이하기 위해 예를 들어, CRT(cathode ray tube) 또는 LCD(liquid crystal display) 모니터와 같은 디스플레이 디바이스 및 사용자가 컴퓨터에 입력을 제공할 수 있는 키보드 및 포인팅 디바이스 예를 들어, 마우스 또는 트랙볼을 갖는 컴퓨터에서 구현될 수 있다. 다른 종류의 디바이스들도 사용자와의 인터렉션을 제공하는데 사용될 수 있다. 예를 들어, 사용자에게 제공되는 피드백은 시각 피드백, 청각 피드백 또는 촉각 피드백과 같은 임의의 형태의 감각적 피드백일 수 있고, 사용자로부터의 입력은 음향, 음성 또는 촉각 입력을 포함하는 임의의 형태로 수신될 수 있다.

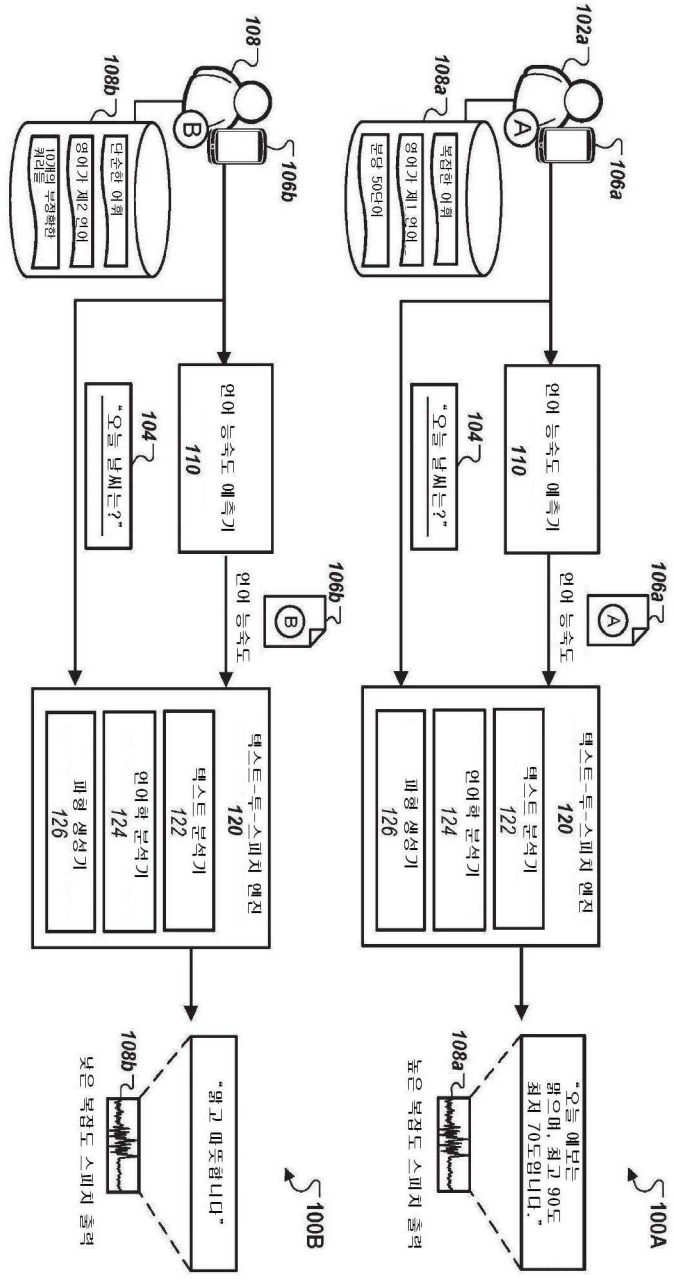
[0107] 본 명세서에서 기술된 시스템들 및 기법들은 예를 들어 데이터 서버와 같은 백엔드 컴포넌트, 애플리케이션 서버와 같은 미들웨어 컴포넌트 또는 그래픽 사용자 인터페이스를 가지는 사용자 컴퓨터 또는 사용자가 본 명세서에 기술된 시스템들 및 기법들의 구현예와 인터렉션할 수 있는 웹 브라우저와 같은 프론트엔드 컴포넌트 또는 하나 이상의 상기 백엔드, 미들웨어 또는 프론트엔드 컴포넌트들의 임의의 조합을 포함하는 컴퓨팅 시스템에서 구현될 수 있다. 시스템의 컴포넌트들은 디지털 데이터 통신의 임의의 형태 또는 매체, 예를 들어 통신 네트워크에 의해 상호연결될 수 있다. 통신 네트워크들의 예시들은 LAN(local area network), WAN(wide area network) 및 인터넷을 포함한다.

[0108] 컴퓨팅 시스템은 사용자들 및 서버들을 포함할 수 있다. 사용자와 서버는 일반적으로 서로 멀리 떨어져 있으며, 일반적으로 통신 네트워크를 통해 인터렉션한다. 사용자와 서버의 관계는 각각의 컴퓨터에서 실행되고 서로 사용자-서버 관계를 갖는 컴퓨터 프로그램에 의해 발생한다.

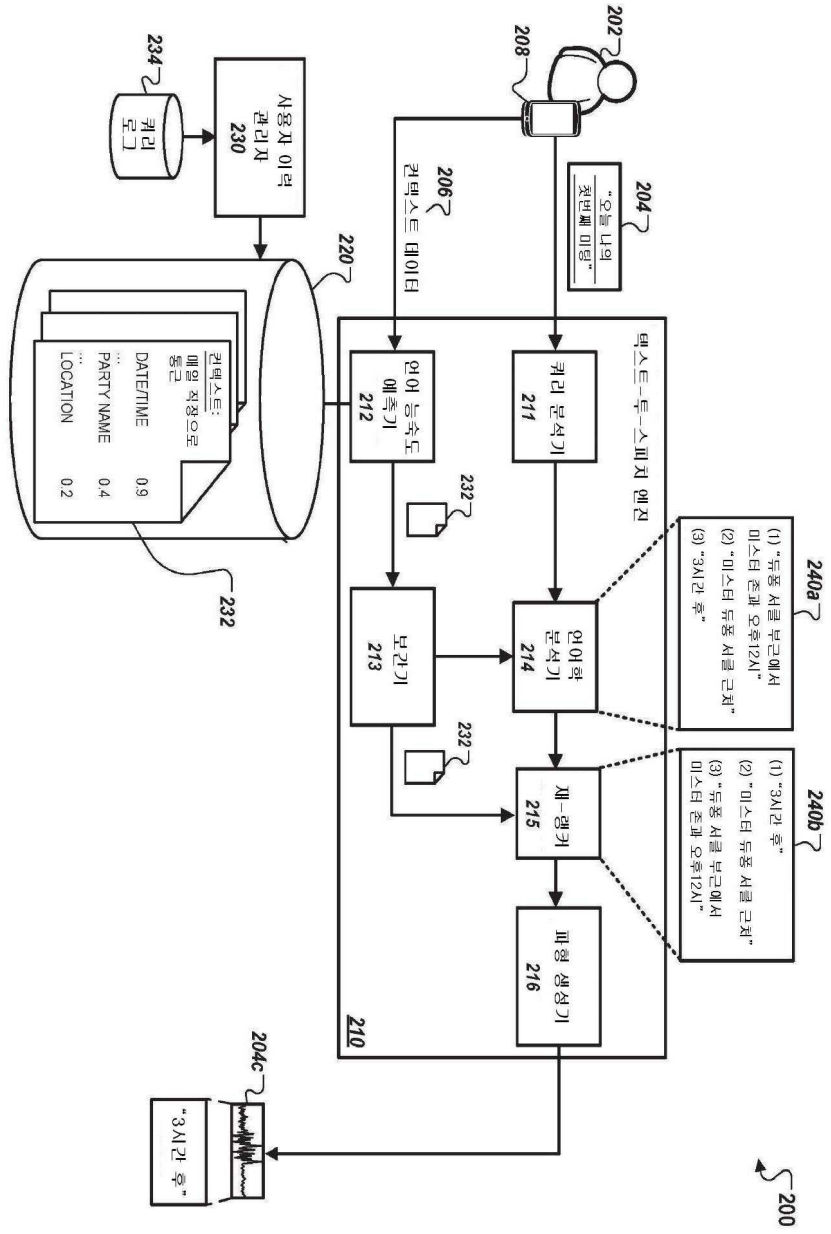
[0109] 다수의 실시예들이 기술되었다. 그럼에도 불구하고, 다양한 수정들이 본 발명의 정신과 범위로부터 벗어나지 않고 이루어질 수 있다는 것이 이해될 것이다. 추가로, 도면들에 도시된 논리 흐름들은 원하는 결과들을 달성하기 위해 특정한 도시된 순서, 또는 시계열적 순서를 반드시 필요로 하지 않는다. 추가로, 다른 단계들이 제공될 수 있거나, 단계들이 기술된 흐름으로부터 생략될 수 있고, 다른 컴포넌트들이 기술된 시스템에 추가되거나 그로부터 제거될 수 있다. 따라서, 다른 실시예들도 다음의 청구항들의 범위 내에 있다.

도면

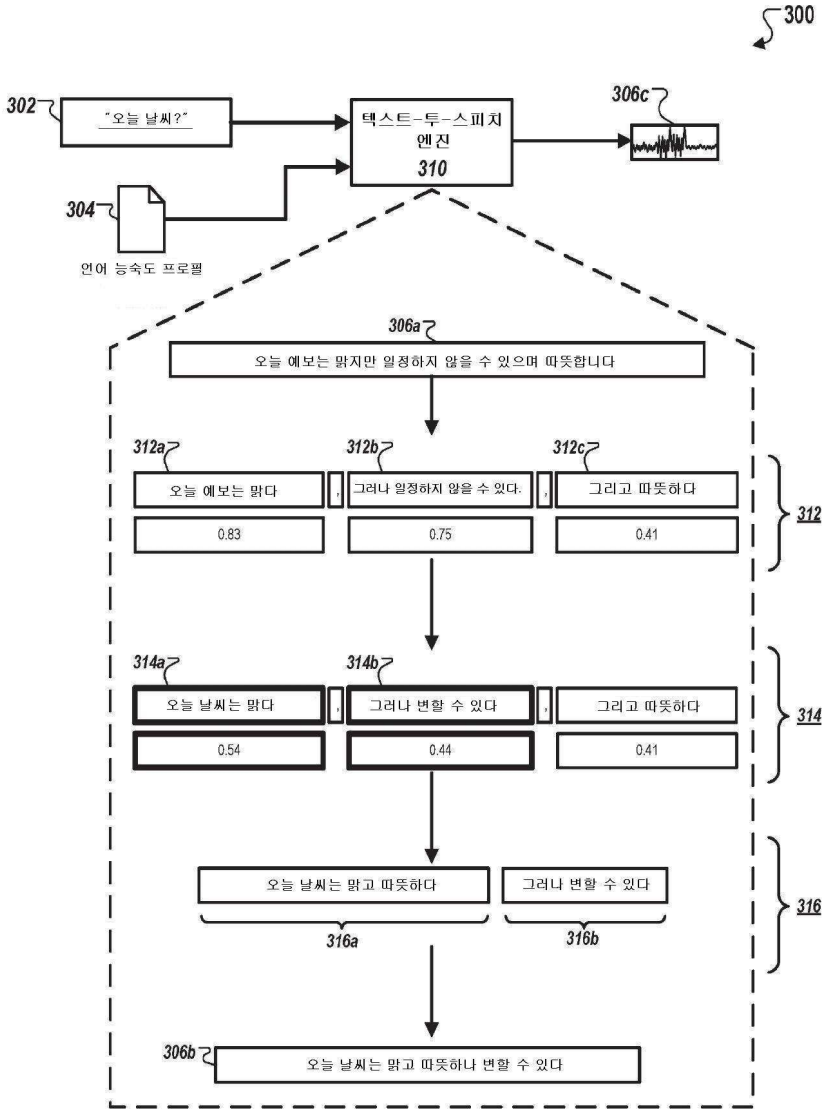
도면1


102a
108a
106a
A
복잡한 어휘
영어가 제1 언어
분당 50단어
언어 능숙도 예측기
110
104
"오늘 날씨는?"
언어 능숙도
106a
텍스트-투-스피치 엔진
120
텍스트 분석기
122
언어학 분석기
124
파형 생성기
126
"오늘 에보는 맑으며, 최고 90도 최저 70도입니다."
108a
높은 복잡도 스피치 출력
100A
108
108b
106b
B
단순한 어휘
영어가 제2 언어
10개의 부정확한 커리큘
언어 능숙도 예측기
110
104
"오늘 날씨는?"
언어 능숙도
106b
텍스트-투-스피치 엔진
120
텍스트 분석기
122
언어학 분석기
124
파형 생성기
126
"맑고 따뜻합니다"
108b
낮은 복잡도 스피치 출력
100B

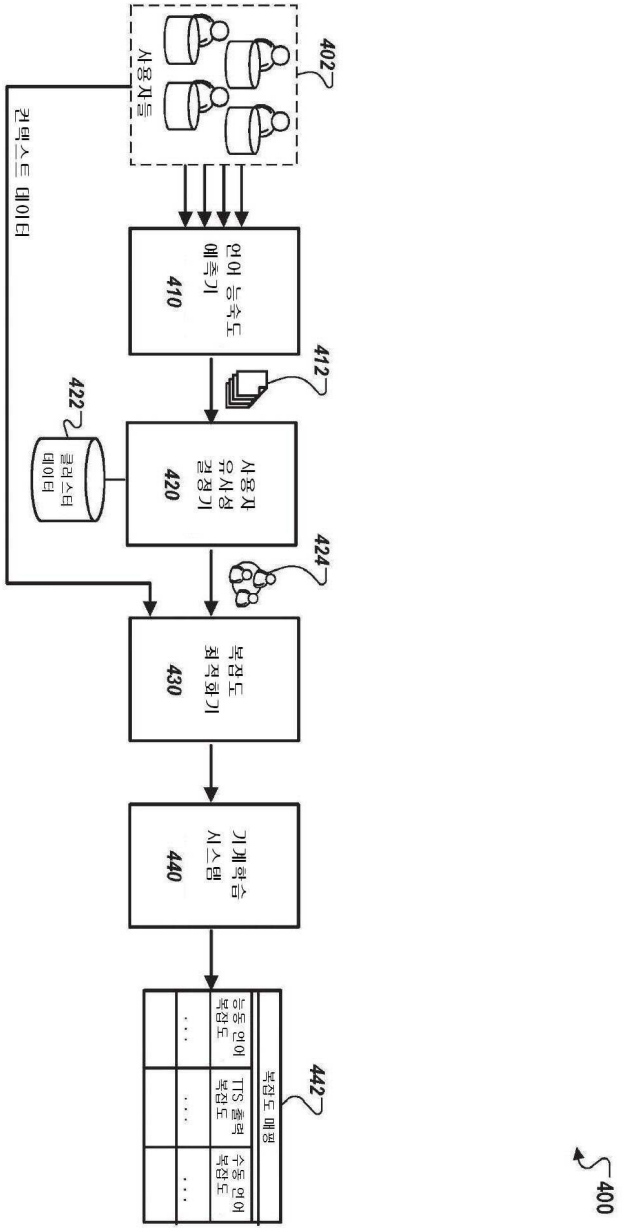
도면2


202
208
204
"오늘 나의 첫번째 미팅"
206
컨텍스트 데이터
234
쿼리 로그
사용자 이력 관리자 230
220
컨텍스트: 메일 작성으로 돌근
DATE/TIME 0.9
PARTY NAME 0.4
LOCATION 0.2
232
텍스트-투-스피치 엔진
언어 능숙도 예측기 212
쿼리 분석기 211
232
보간기 213
언어학 분석기 214
232
재-랭커 215
파형 생성기 216
210
240a
(1) "듀퐁 서클 부근에서 미스터 존과 오후12시"
(2) "미스터 듀퐁 서클 근처"
(3) "3시간 후"
240b
(1) "3시간 후"
(2) "미스터 듀퐁 서클 근처"
(3) "듀퐁 서클 부근에서 미스터 존과 오후12시"
204c
"3시간 후"
200

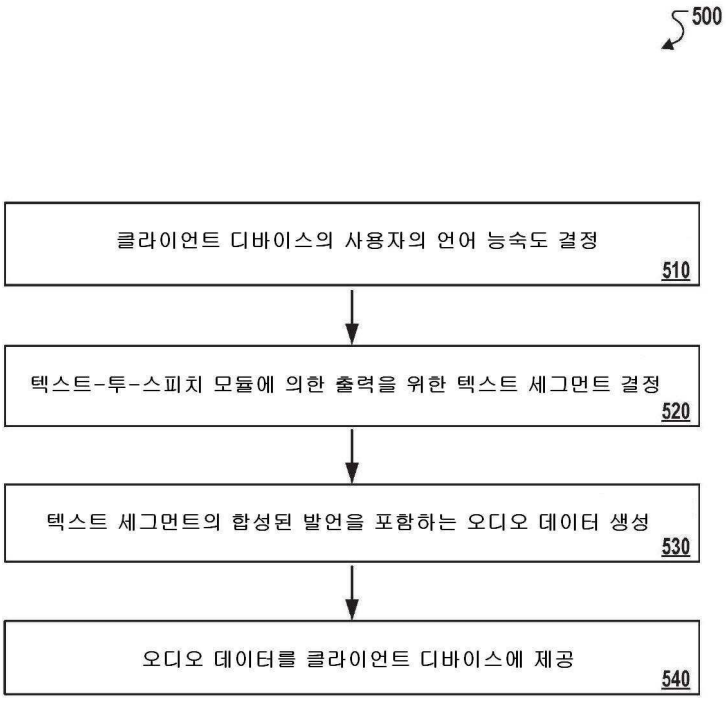
도면3


300
302
"오늘 날씨?"
텍스트-투-스피치
엔진
310
306c
304
언어 능숙도 프로필
306a
오늘 예보는 맑지만 일정하지 않을 수 있으며 따뜻합니다
312a
312b
312c
오늘 예보는 맑다
그러나 일정하지 않을 수 있다.
그리고 따뜻하다
0.83
0.75
0.41
312
314a
314b
오늘 날씨는 맑다
그러나 변할 수 있다
그리고 따뜻하다
0.54
0.44
0.41
314
오늘 날씨는 맑고 따뜻하다
그러나 변할 수 있다
316a
316b
316
306b
오늘 날씨는 맑고 따뜻하나 변할 수 있다

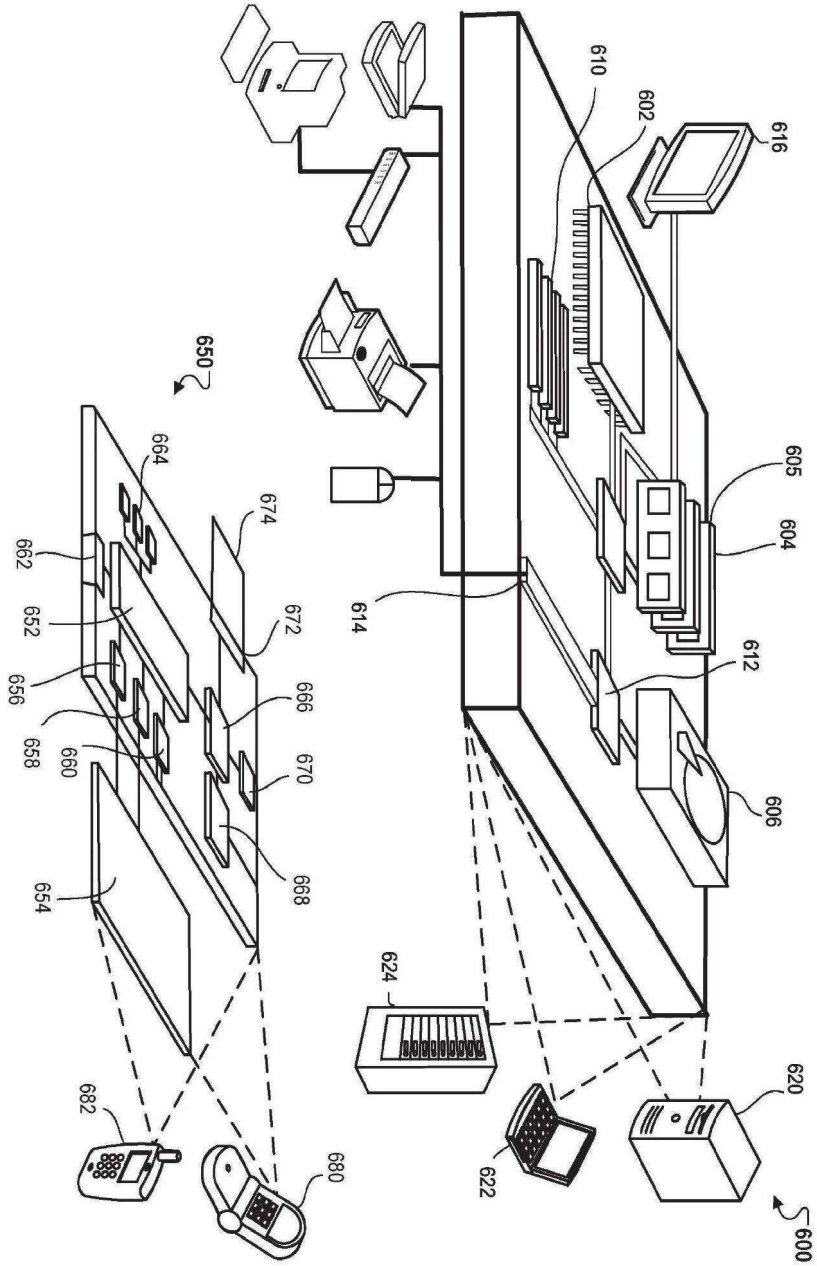
도면4


사용자들
402
컨텍스트 데이터
언어 능숙도 예측기
410
412
422
클러스터 데이터
사용자 유사성 결정기
420
424
복잡도 최적화기
430
기계학습 시스템
440
442
복잡도 매핑
능동 언어 복잡도
TTS 출력 복잡도
수동 언어 복잡도
400

도면5


500
클라이언트 디바이스의 사용자의 언어 능숙도 결정
510
텍스트-투-스피치 모듈에 의한 출력을 위한 텍스트 세그먼트 결정
520
텍스트 세그먼트의 합성된 발언을 포함하는 오디오 데이터 생성
530
오디오 데이터를 클라이언트 디바이스에 제공
540

도면6


610
602
616
650
664
674
662
652
672
656
666
658
660
670
654
668
605
604
614
612
606
624
620
622
682
680
600