(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0030448
(43) 공개일자 2022년03월11일

(51) 국제특허분류(Int. Cl.)
G10L 21/10 (2013.01) *G06N 20/00* (2019.01)
G10L 15/06 (2006.01) *G10L 15/26* (2006.01)
G10L 15/30 (2013.01) *G10L 21/06* (2006.01)
(52) CPC특허분류
G10L 21/10 (2013.01)
G06N 20/00 (2021.08)
(21) 출원번호 10-2020-0110844
(22) 출원일자 2020년09월01일
심사청구일자 2020년09월01일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
임희석
서울특별시 성북구 안암로 145 자연계캠퍼스애기능생활관 311호 NLP&AI연구실
박찬준
서울특별시 송파구 중대로 24, 306동 1303호 (문정동, 올림픽훼밀리타운)
김양희
서울특별시 동대문구 왕산로 52, 102동 1108호
(74) 대리인
김홍석

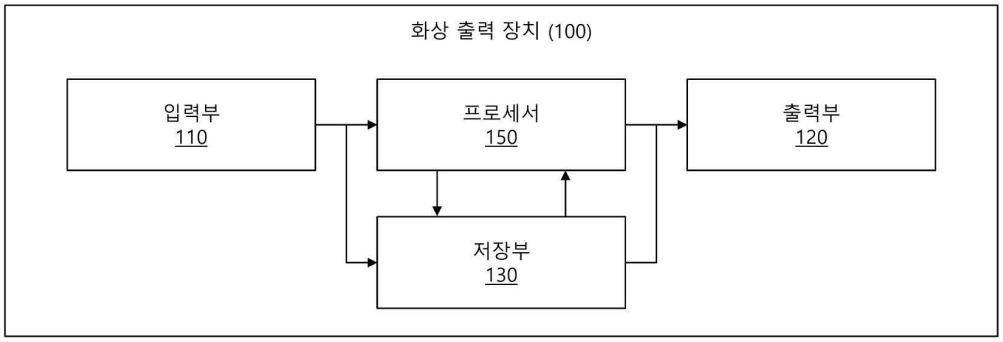
전체 청구항 수 : 총 9 항

(54) 발명의 명칭 **언어 대응 화상 출력 장치, 방법 및 시스템**

(57) 요 약

언어 대응 화상 출력 장치, 방법 및 시스템에 관한 것으로, 언어 대응 화상 출력 장치는 음성을 수신하고 음성 데이터를 획득하는 입력부, 상기 음성 데이터를 텍스트 데이터로 변환하고, 상기 텍스트 데이터를 기반으로 상기 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득하는 프로세서 및 상기 적어도 하나의 화상을 출력하는 출력부를 포함할 수 있다

대 표 도 - 도1



(52) CPC특허분류
G10L 15/063 (2013.01)
G10L 15/26 (2013.01)
G10L 15/30 (2013.01)
G10L 2021/065 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711116144
과제번호	2018-0-01405-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성(R&D)
연구과제명	Human-inspired 복합지능 원천기술 연구 및 인력양성
기 여 율	1/1
과제수행기관명	고려대학교산학협력단
연구기간	2020.01.01 ~ 2020.12.31

명 세 서

청구범위

청구항 1

음성을 수신하고 음성 데이터를 획득하는 입력부;

상기 음성 데이터를 텍스트 데이터로 변환하고, 상기 텍스트 데이터를 기반으로 상기 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득하는 프로세서; 및

상기 적어도 하나의 화상을 출력하는 출력부;를 포함하는 언어 대응 화상 출력 장치.

청구항 2

제1항에 있어서,

상기 프로세서는, 상기 텍스트 데이터에 대한 교정을 수행하여 교정된 텍스트 데이터를 획득하고, 상기 교정된 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득하는 언어 대응 화상 출력 장치.

청구항 3

제2항에 있어서,

상기 프로세서는, 트랜스포머 기반의 시퀀스 투 시퀀스 학습 모델을 이용하여 상기 텍스트 데이터에 대응하는 교정된 텍스트 데이터를 획득하는 언어 대응 화상 출력 장치.

청구항 4

제1항에 있어서,

상기 프로세서는 N-그램을 기반으로 상기 텍스트 데이터에 대응하는 화상을 검출 및 매핑함으로써 상기 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득하는 언어 대응 화상 출력 장치.

청구항 5

음성을 수신하고 음성 데이터를 획득하는 단말 장치; 및

상기 단말 장치와 통신 가능한 서버 장치;를 포함하되,

상기 단말 장치 또는 상기 서버 장치는, 상기 음성 데이터를 텍스트 데이터로 변환하는 과정 및 상기 텍스트 데이터를 기반으로 상기 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득하는 과정 중 적어도 하나를 수행하는 언어 대응 화상 출력 시스템.

청구항 6

음성을 수신하고 음성 데이터를 획득하는 단계;

상기 음성 데이터를 텍스트 데이터로 변환하는 단계;

상기 텍스트 데이터를 기반으로 상기 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득하는 단계; 및

상기 적어도 하나의 화상을 출력하는 단계;를 포함하는 언어 대응 화상 출력 방법.

청구항 7

제6항에 있어서,

상기 텍스트 데이터에 대한 교정을 수행하여 교정된 텍스트 데이터를 획득하는 단계;를 더 포함하는 언어 대응 화상 출력 방법.

청구항 8

제7항에 있어서,

상기 텍스트 데이터에 대한 교정을 수행하여 교정된 텍스트 데이터를 획득하는 단계는,

트랜스포머 기반의 시퀀스 투 시퀀스 학습 모델을 이용하여 상기 텍스트 데이터에 대응하는 교정된 텍스트 데이터를 획득하는 단계;를 포함하는 언어 대응 화상 출력 방법.

청구항 9

제6항에 있어서,

상기 텍스트 데이터를 기반으로 상기 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득하는 단계는,

N-그램을 기반으로 상기 텍스트 데이터에 대응하는 화상을 검출 및 매핑함으로써 상기 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득하는 단계;를 포함하는 언어 대응 화상 출력 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 언어 대응 화상 출력 장치, 방법 및 시스템에 관한 것이다.

배 경 기 술

[0002] 언어 발달 장애를 가진 아동이나 성인(실어증 환자 등), 언어 능력이 저하된 노인 또는 청각 장애인 등은 의사 소통 수단인 언어를 제대로 이해하지 못하거나 제대로 발화하지 못해 타인과의 의사 소통에 어려움을 갖는다. 이에 따라 많은 수의 이들 언어 장애인들은 일상 생활에서 큰 불편함을 겪고 있다. 특히, 이러한 의사 소통의 어려움은 이들의 경제 활동을 제약하고 있으며, 이러한 경제 활동의 제약은 언어 장애인이 경제적 어려움을 겪는 큰 이유가 되고 있다. 실제로 청각 및 언어 장애인의 고용률은 고작 8.5%에 불과하다. 그러나, 청각 및 언어 장애인의 비율은 매년 증가하는 추세이므로, 이들의 의사 소통 능력을 향상시켜 사회적, 경제적 진입 장벽을 낮출 필요성이 존재한다.

[0003] 이들의 의사 소통 능력의 향상을 위해서 보완 대체 의사 소통(AAC: Augmentative and Alternative Communication) 기술이 제시되고 있다. 보완 대체 의사 소통은 발달 장애인이 비언어적 의사 소통을 가능하게 하는 것으로, 구체적으로 상징, 보조도구, 전략 및 기법의 네 가지로 이루어져 있다. 여기서, 상징은 제스처, 사진 또는 수화 체계 등과 같은 개념들을 시각, 청각 또는 촉각적으로 표상화한 것을 의미하고, 보조 도구는 메시지의 전달을 위해 이용되는 도구를 의미한다. 전략은 이들 상징 및 보조기구를 효과적으로 사용하기 위한 계획 및 시스템을 의미하고, 기법은 의도하는 메시지를 전달하는 방법을 의미한다. 보완 대체 의사 소통의 사용은 사회성 향상, 언어 기능 향상을 통해 언어 장애인들이 스스로 의사 소통을 하려는 노력을 증가시키고, 타인과의 의사 소통을 촉진시켜, 이들의 의사 소통 능력 개선에 상당히 도움을 줄 뿐만 아니라, 본인의 의사를 직접 표현할 수 있게 되어 언어 장애인의 삶의 만족도를 개선시키고 있다. 근자에는 이와 같은 보완 대체 의사 소통에 관한 다양한 방법, 소프트웨어 또는 장치 등이 제시되고 있었다. 그러나, 이들 방법이나 장치는 대체적으로 아동 등의 발달 장애인이 사용하기에는 사용 편의성이나 접근성이 낮아, 언어 장애인에게 실질적인 도움을 주기에는 미흡한 부분이 많았다.

선행기술문헌

특허문헌

[0004] (특허문헌 0001) 대한민국 등록특허 제1965103호 (2019.04.02. 공고)

(특허문헌 0002) 대한민국 등록특허 제1551714호 (2015.09.10. 공고)

(특허문헌 0003) 대한민국 공개특허 제2006-0012949호 (2006.02.09. 공개)

(특허문헌 0004) 대한민국 공개특허 제2005-0092293호 (2005.09.21. 공개)

발명의 내용

해결하려는 과제

[0005] 본 발명은 의사 소통에 어려움이 존재하거나 언어 학습 능력이 부족한 개개인의 의사 소통 능력을 개선하고 언어 이해력 향상에 도움을 줄 수 있는 언어 대응 화상 출력 장치, 방법 및 시스템을 제공하는 것을 해결하고자 하는 과제로 한다.

과제의 해결 수단

[0006] 상술한 과제를 해결하기 위하여 언어 대응 화상 출력 장치, 언어 대응 화상 출력 시스템 및 언어 대응 화상 출력 방법이 제공된다.

[0007] 언어 대응 화상 출력 장치는, 음성을 수신하고 음성 데이터를 획득하는 입력부, 상기 음성 데이터를 텍스트 데이터로 변환하고, 상기 텍스트 데이터를 기반으로 상기 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득하는 프로세서 및 상기 적어도 하나의 화상을 출력하는 출력부를 포함할 수 있다.

[0008] 상기 프로세서는, 상기 텍스트 데이터에 대한 교정을 수행하여 교정된 텍스트 데이터를 획득하고, 상기 교정된 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득할 수 있다.

[0009] 상기 프로세서는, 트랜스포머 기반의 시퀀스 투 시퀀스 학습 모델을 이용하여 상기 텍스트 데이터에 대응하는 교정된 텍스트 데이터를 획득할 수 있다.

[0010] 상기 프로세서는 N-그램을 기반으로 상기 텍스트 데이터에 대응하는 화상을 검출 및 매핑함으로써 상기 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득할 수도 있다.

[0011] 언어 대응 화상 출력 시스템은, 음성을 수신하고 음성 데이터를 획득하는 단말 장치 및 상기 단말 장치와 통신 가능한 서버 장치를 포함하되, 상기 단말 장치 또는 상기 서버 장치는, 상기 음성 데이터를 텍스트 데이터로 변환하는 과정 및 상기 텍스트 데이터를 기반으로 상기 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득하는 과정 중 적어도 하나를 수행하도록 마련된 것일 수 있다.

[0012] 언어 대응 화상 출력 방법은, 음성을 수신하고 음성 데이터를 획득하는 단계, 상기 음성 데이터를 텍스트 데이터로 변환하는 단계, 상기 텍스트 데이터를 기반으로 상기 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득하는 단계 및 상기 적어도 하나의 화상을 출력하는 단계를 포함할 수 있다.

[0013] 언어 대응 화상 출력 방법은, 상기 텍스트 데이터에 대한 교정을 수행하여 교정된 텍스트 데이터를 획득하는 단계를 더 포함할 수 있다.

[0014] 상기 텍스트 데이터에 대한 교정을 수행하여 교정된 텍스트 데이터를 획득하는 단계는, 트랜스포머 기반의 시퀀스 투 시퀀스 학습 모델을 이용하여 상기 텍스트 데이터에 대응하는 교정된 텍스트 데이터를 획득하는 단계를 포함할 수 있다.

[0015] 상기 텍스트 데이터를 기반으로 상기 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득하는 단계는, N-그램을 기반으로 상기 텍스트 데이터에 대응하는 화상을 검출 및 매핑함으로써 상기 텍스트 데이터에 대응하는 적어도 하나의 화상을 획득하는 단계를 포함할 수 있다.

발명의 효과

[0016] 상술한 언어 대응 화상 출력 장치, 방법 및 시스템에 의하면, 언어의 사용에 어려움이 존재하는 개개인의 의사 소통 능력을 개선하고 이들의 언어 학습 능력이나 언어 이해력 향상을 도모할 수 있게 되고, 이에 따라 인지 능력이 낮아 언어를 제대로 학습할 수 없는 발달 장애 아동들과 같이 언어 발화 및 이해에 어려움을 겪는 사람들이 보다 원활하게 타인과 의사 소통을 하고 언어를 교육받을 수 있게 하는 장점도 얻을 수 있다.

[0017] 상술한 언어 대응 화상 출력 장치, 방법 및 시스템에 의하면, 보조적 의사 소통 수단으로 화상(일례로 픽토그램(pictogram))이 제공되므로, 언어 장애인 등이 자신 또는 타인이 발화 또는 입력한 내용을 보다 직관적이고 신속하면서도 용이하게 인식할 수 있는 장점도 얻을 수 있다.

[0018] 상술한 언어 대응 화상 출력 장치, 방법 및 시스템에 의하면, 의사 소통의 어려움에 기인한 언어 장애인의 욕구 불만 또는 스트레스를 완화시킬 수 있게 되며, 또한 이들 욕구 불만이나 스트레스에서 비롯되는 2차적인 정서

장애나 문제 행동을 감소시킴으로써 언어 장애인의 인성이나 행동을 보다 긍정적으로 개선시킬 수 있는 효과도 얻을 수 있다.

[0019] 상술한 언어 대응 화상 출력 장치, 방법 및 시스템에 의하면, 음성 인식의 정확도의 향상에 따라 사용자는 보다 정확하면서 표준 문법에 부합하는 텍스트를 제공받을 수 있게 되어 사용자는 자신 또는 타인의 발화에 대한 표준 텍스트를 확인할 수 있게 되고, 이에 따라 특정 언어(모국어 또는 비 모국어)의 문법 체계 등을 적절하게 이해하지 못한 아동 또는 성인이 잘못된 문법을 교정 받고 올바른 문장을 익힐 수 있게 되는 효과도 얻을 수 있다.

[0020] 상술한 언어 대응 화상 출력 장치, 방법 및 시스템에 의하면, 화자의 발화에 대응하는 텍스트와 이에 대응하는 화상이 외부로 출력되므로, 소리를 들을 수 없는 청각 장애인도 보다 용이하게 타인과 의사 소통을 할 수 있게 되는 장점도 존재한다.

도면의 간단한 설명

[0021] 본 발명의 상세한 설명에서 인용되는 도면을 보다 충분히 이해하기 위하여 각 도면의 상세한 설명이 제공된다.

도 1은 언어 대응 화상 출력 장치의 일 실시예에 대한 블록도이다.

도 2는 프로세서의 동작의 일 실시예를 설명하기 위한 도면이다.

도 3은 텍스트에 대응하는 화상을 선택 및 배치하는 과정의 일 실시예를 도시한 도면이다.

도 4는 입력된 문장에 대응하는 화상의 일례를 도시한 도표이다.

도 5는 언어 대응 화상 출력 장치에 출력되는 화면의 일 실시예를 도시한 도면이다.

도 6은 언어 대응 화상 출력 시스템의 일 실시예에 대한 도면이다.

도 7은 언어 대응 화상 출력 방법의 일 실시예에 대한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0022] 이하 명세서 전체에서 동일 참조 부호는 특별한 사정이 없는 한 동일 구성요소를 지칭한다. 이하에서 사용되는 '부'가 부가된 용어는, 소프트웨어 또는 하드웨어로 구현될 수 있으며, 실시예에 따라 하나의 '부'가 하나의 물리적 또는 논리적 부품으로 구현되거나, 복수의 '부'가 하나의 물리적 또는 논리적 부품으로 구현되거나, 하나의 '부'가 복수의 물리적 또는 논리적 부품들로 구현되는 것도 가능하다.

[0023] 명세서 전체에서 어떤 부분이 다른 부분과 연결되어 있다고 할 때, 이는 어떤 부분과 다른 부분에 따라서 물리적 연결을 의미할 수도 있고, 또는 전기적으로 연결된 것을 의미할 수도 있다. 또한, 어떤 부분이 다른 부분을 포함한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 부분 이외의 또 다른 부분을 제외하는 것이 아니며, 설계자의 선택에 따라서 또 다른 부분을 더 포함할 수 있음을 의미한다.

[0024] 제 1이나 제 2 등의 용어는 하나의 부분을 다른 부분으로부터 구별하기 위해 사용되는 것으로, 특별한 기재가 없는 이상 이들이 순차적인 표현을 의미하는 것은 아니다. 또한 단수의 표현은 문맥상 명백하게 예외가 있지 않는 한, 복수의 표현을 포함할 수 있다.

[0025] 이하, 도 1 내지 도 5를 참조하여 언어 대응 화상 출력 장치의 일 실시예에 대해서 설명하도록 한다.

[0026] 도 1은 언어 대응 화상 출력 장치의 일 실시예에 대한 블록도이다.

[0027] 도 1에 도시된 바에 의하면, 언어 대응 화상 출력 장치(100)는 입력부(110), 출력부(120), 저장부(130) 및 프로세서(150)를 포함할 수 있다.

[0028] 입력부(110)은 발화에 따른 음성을 입력 받거나 사용자의 조작에 따른 텍스트 등을 입력 받을 수 있다. 예를 들어, 입력부(110)는 사용자(일례로 언어 장애인, 대화자 또는 설계자 등)가 발화한 음성(음파)을 수신하고, 수신한 음성을 대응하는 전기적 신호로 변환하여 음성 데이터를 획득하는 마이크로 폰 장치를 포함할 수 있다. 여기서, 음성은 영어, 한국어, 일본어, 중국어 또는 이 다양한 언어 중 적어도 하나의 언어를 이용하여 발화된 것일 수 있다. 음성 데이터는 회로나 케이블 등을 통해 프로세서(150) 및 저장부(130) 중 적어도 하나로 전달될 수 있다. 이외에도 입력부(110)는 키보드, 마우스 장치, 태블릿, 터치 스크린, 터치 패드, 트랙볼 및/또는 트랙패드 등을 포함할 수도 있다.

[0029] 출력부(120)는 프로세서(150)의 처리 결과에 따른 텍스트 및 텍스트에 대응하는 화상 중 적어도 하나를 외부로 시각적 또는 청각적으로 출력할 수 있다. 예를 들어, 출력부(120)는 음성 데이터를 변환하여 획득한 텍스트 데이터, 텍스트 데이터를 교정하여 획득한 텍스트 형태의 교정된 텍스트 데이터 및 텍스트 데이터나 교정된 텍스트 데이터를 기반으로 획득된 적어도 하나의 화상 중 적어도 하나를 화면 상에 시각적으로 표시함으로써, 사용자(일례로 언어 장애인)가 발화한 음성에 대응하는 화상 및/또는 텍스트를 사용자 또는 사용자와 의사 소통을 하려는 대화자(청취자를 포함할 수 있다) 등에게 제공할 수 있다. 실시예에 따라서, 출력부(120)는 텍스트를 음성의 형태로 변환하여 출력하는 것도 가능하다. 출력부(120)는 예를 들어 음극선관, 디스플레이 패널이나 스피커 장치 등을 포함할 수 있으나, 이에 한정되는 것은 아니다.

[0030] 저장부(130)는, 화상 출력 장치(100) 또는 프로세서(150) 등의 동작에 필요한 적어도 하나의 애플리케이션(소프트웨어, 프로그램 또는 앱 등으로 지칭 가능함)이나 데이터 등을 일시적 또는 비일시적으로 저장할 수 있다. 예를 들어, 저장부(130)는 음성 데이터를 대응하는 텍스트 데이터로 변환하거나, 텍스트 데이터를 교정하거나, 및/또는 텍스트 데이터에 대응하는 화상을 선택하기 위한 애플리케이션을 저장하거나, 적어도 하나의 단어나 문장에 각각 대응하는 적어도 하나의 화상(일례로 픽토그램(pictogram)이나 아이콘(icon) 등)을 포함하는 데이터베이스 등을 저장하고 있을 수 있다. 여기서, 애플리케이션이나 데이터베이스는 설계자에 의해 직접 작성 및 입력되어 저장부(130)에 저장된 것일 수도 있고, 또는 유선 또는 무선 통신 네트워크를 통해 접속 가능한 유통망(일례로 전자 소프트웨어 유통망(ESD: Electronic Software Distribution)을 통하여 획득 또는 갱신된 것일 수도 있다. 저장부(120)는, 예를 들어, 주기억장치 및 보조기억장치 중 적어도 하나를 포함할 수 있다.

[0031] 프로세서(150)는 화상 출력 장치(100)의 동작 전반에 필요한 연산, 판단, 처리 및/또는 제어 동작 등을 수행하거나 또는 언어에 대응하는 화상의 선택을 위해 필요한 소정의 도작만을 수행하도록 마련된 것일 수 있다. 이 경우, 프로세서(150)는 저장부(130)에 저장된 애플리케이션을 구동시켜, 언어에 대응하는 화상의 선택을 위한 동작을 수행할 수도 있다. 프로세서(150)는, 예를 들어, 중앙 처리 장치(CPU, Central Processing Unit), 마이크로 컨트롤러 유닛(MCU, Micro Controller Unit), 마이컴(Micom, Micro Processor), 애플리케이션 프로세서(AP, Application Processor), 전자 제어 유닛(ECU, Electronic Controlling Unit) 및/또는 각종 연산 처리 및 제어 신호의 생성이 가능한 다른 전자 장치 등을 포함할 수 있다. 이들 장치는 예를 들어 하나 또는 둘 이상의 반도체 칩 및 관련 부품을 이용하여 구현 가능하다. 프로세서(150)의 구체적인 동작은 후술한다.

[0032] 상술한 언어 대응 화상 출력 장치(100)는 데이터의 연산 처리가 가능한 적어도 하나의 정보 처리 장치를 이용하여 구현될 수 있다. 적어도 하나의 정보 처리 장치는, 스마트 폰, 태블릿 피씨, 스마트 시계, 두부 장착형 디스플레이(HMD, Head Mounted Display) 장치, 내비게이션 장치, 개인용 디지털 보조기(PDA), 휴대용 게임기, 데스크톱 컴퓨터, 랩톱 컴퓨터, 디지털 텔레비전, 셋톱 박스, 가전 기기, 인공 지능 음향 재생 장치(인공 지능 스피커) 및/또는 언어에 대응하는 화상을 출력하기 위해 특별히 고안된 장치 등을 포함할 수 있다. 그러나, 이는 예시적인 것으로, 이들 외에도 언어 대응 화상 출력 장치(100)는 음향의 입력, 연산 처리 및 언어에 대응하는 화상의 외부 출력이 가능한 다양한 종류의 정보 처리 장치 중 적어도 하나를 포함할 수 있다.

[0033] 도 2는 프로세서의 동작의 일 실시예를 설명하기 위한 도면이다.

[0034] 도 2에 도시된 바에 의하면, 프로세서(150)는, 일 실시예에 있어서, 인식부(151) 및 보조화상처리부(153)를 포함할 수 있으며, 필요에 따라 교정부(152)를 더 포함할 수 있다. 프로세서(150)의 인식부(151), 교정부(152) 및 보조화상처리부(153)는 논리적으로 구분되는 것일 수도 있고, 또는 물리적으로 구분되는 것일 수도 있다. 물리적으로 구분되는 경우, 인식부(151), 교정부(152) 및 보조화상처리부(153) 중 적어도 하나는 별도로 분리된 다른 반도체 칩 등을 이용하여 구현될 수도 있다.

[0035] 인식부(151)는 음성 데이터를 수신하고 수신한 음성 데이터를 이에 대응하는 텍스트 데이터로 실시간으로 또는 비실시간으로 변환할 수 있다. 인식부(151)는 스피치 투 텍스트(STT: Speech to Text) 기능을 기반으로 음성 데이터의 텍스트 변환을 수행할 수 있다. 이 경우, 인식부(151)는 학습 모델을 기반으로 음성 데이터의 텍스트 변환을 수행할 수 있으며, 여기서 학습 모델은 적어도 하나의 신경망(Neural Network)을 이용하여 구현될 것일 수 있다. 적어도 하나의 신경망은 심층 신경망(DNN: Deep Neural Network), 콘볼루션 신경망(CNN: Convolution Neural Network), 순환 신경망(RNN: Recurrent Neural Network), 심층 신뢰 신경망(DBN: Deep Belief Network), 장단기 메모리(LSTM: Long short term memory), 심층 강화 학습(Deep reinforcement learning) 알고리즘 및/또는 심층 Q-네트워크(Deep Q-Network) 등을 포함할 수 있으나 이에 한정되지는 않는다.

[0036] 인식부(151)가 획득한, 음성 데이터에 대응하는 텍스트 데이터는 교정부(152)로 전달되거나 또는 보조화상처리부(153)로 전달될 수 있다. 실시예에 따라 텍스트 데이터는, 텍스트 데이터의 전달 이전에, 전달 이후에 또는

전달과 동시에 저장부(130)로도 전달된 후, 저장부(130)에 일시적 또는 비일시적으로 저장될 수도 있다.

[0037] 교정부(152)는 음성 데이터에 대응하는 텍스트 데이터(즉, 소스 문장)를 교정하여 교정된 텍스트 데이터(즉, 목표 문장)를 획득할 수 있다. 구체적으로 교정부(152)는 변환된 텍스트 데이터(일례로 노이즈 문장)의 문법적 오류를 교정하여 교정된 텍스트 데이터(일례로 올바른 문장)를 획득할 수 있다. 이 경우, 텍스트 데이터는 표준 문법 체계를 기준으로 교정될 수도 있으나, 실시예에 따라서 의사 소통의 상황에 따라서 설계자가 원하는 체계를 기준으로 교정되도록 하는 것도 가능하다. 교정부(152)는 적어도 하나의 학습 모델을 기반으로 텍스트 데이터를 교정할 수 있다. 교정부(152)가 이용하는 적어도 하나의 학습 모델은, 인식부(151)가 이용하는 학습 모델과 동일할 수도 있고 상이할 수도 있다. 일 실시예에 의하면, 교정부(151)는 트랜스포머(Transformer) 기반의 시퀀스 투 시퀀스(seq2seq: Sequence to Sequence) 학습 모델을 이용 또는 응용하여 텍스트 데이터에 대한 교정을 수행할 수도 있다. 트랜스포머란 콘볼루션(Convolution)이나 순환(Recurrent) 없이 오직 어텐션(attention)만을 이용하여 학습을 수행하는 모델로, 적어도 하나의 인코더(encoder)-디코더(decoder)의 구조로 형성되어 있다. 트랜스포머에서는 인코더 및 디코더 각각에 의해 입력 및 출력 각각에 대한 셀프 어텐션(self-attention)이 학습된 후, 입력과 출력 간의 어텐션이 학습된다. 인코더 및 디코더 각각은 멀티 헤드 어텐션(Multi Head Attention) 층을 포함하고 있으며, 멀티 헤드 어텐션 층에 쿼리(Q, query), 키(K, key) 및 값(V, value) 이라는 3개의 벡터가 학습을 위해 입력될 수 있다. 이와 같은 트랜스포머에 의하면 보다 우수하게 오문장이 올바른 문장으로 번역 및 교정될 수 있게 된다. 이외에도 교정부(151)는 심층 신경망, 합성곱 신경망, 순환 신경망, 심층 신뢰 신경망, 장단기 메모리, 심층 강화 학습 알고리즘 및/또는 심층 Q-네트워크 등과 같이 텍스트 교정에 이용 가능한 다양한 종류의 학습 알고리즘 중 적어도 하나를 이용하여 텍스트 데이터의 교정을 수행할 수도 있다.

[0038] 보조화상처리부(153)는 인식부(151)가 획득한 텍스트 데이터 및 교정부(152)가 획득한 교정된 텍스트 데이터 중 적어도 하나를 기반으로, 텍스트 데이터 또는 교정된 텍스트 데이터에 대응하는 보조 화상을 획득할 수 있다. 구체적으로 보조화상처리부(153)는 텍스트 데이터 또는 교정된 텍스트 데이터 내의 적어도 하나의 단어(명사, 동사, 형용사 또는 부사 등) 각각에 대응하는 적어도 하나의 화상, 적어도 하나의 구(명사구, 동사구, 형용사구, 관형사구 또는 부사구 등) 각각에 대응하는 적어도 하나의 화상 및/또는 적어도 하나의 문장에 대응하는 적어도 하나의 화상을 획득하고, 획득한 적어도 하나의 화상을 적절하게 배치할 수 있다. 이 경우, 보조화상처리부(135)는 획득한 적어도 하나의 화상을, 원하는 언어의 어순(일례로 한국어 어순이나 영어 어순 등)에 따라 배치함으로써 사용자가 보다 직관적으로 발화한 단어나 문장의 의미를 알 수 있게 할 수도 있다. 실시예에 따라서, 설계자가 미리 선택에 따라서 이들 적어도 하나의 화상은 배치될 수도 있다. 여기서, 보조 화상은 정지 화상 및 동화상 중 적어도 하나를 포함할 수 있다. 보다 구체적으로 보조 화상은 픽토그램이나 아이콘 등과 같은 도안이나, 촬영된 사진이나 동영상 등을 포함할 수 있다. 이들 픽토그램, 사진 또는 동영상 등은 소정의 데이터베이스의 형태로 저장부(130)에 저장되어 있을 수 있으며, 각각의 픽토그램, 사진 또는 동영상에는 해당 픽토그램, 사진 또는 동영상 등을 설명하기 위한 메타 텍스트(단어, 구 혹은 문장 형태)가 부가되어 있을 수 있다.

[0039] 일 실시예에 의하면, 보조화상처리부(153)는 텍스트 데이터에 적합한 적어도 하나의 보조 화상을 획득하기 위해 멀티모덜 처리(Multimodal processing)를 이용할 수도 있다. 멀티모덜 처리는 텍스트 및 화상과 같이 서로 상이한 형태의 정보를 포함하는 멀티모덜 데이터(Multimodal date)를 효과적으로 다룰 수 있는 처리 방법을 의미한다.

[0040] 도 3은 텍스트에 대응하는 화상을 선택 및 배치하는 과정의 일 실시예를 도시한 도면이다.

[0041] 보다 구체적으로 보조화상처리부(153)는, 도 3에 도시된 바와 같이, 상호 참조 해결(co-reference resolution)를 이용하여 텍스트 데이터 또는 교정된 텍스트 데이터 내에서의 대용어 문제를 해결할 수 있다(155). 상호 참조(대용어)란 특정한 대상을 표현하는 특정한 단어 대신에 특정한 대상을 지칭하는 다른 단어(대명사나 약칭 등) 등을 의미하며, 상호 참조 해결이란 이와 같이 특정 개체에 대한 다수의 표현 간의 참조 관계를 밝히는 것이다. 상호 참조 해결에 의해 텍스트 데이터 또는 교정된 텍스트 데이터 내에 대용어가 존재한다고 하더라도 보조화상처리부(153)는 텍스트 데이터 또는 교정된 텍스트 데이터 내의 단어나 어구에 대한 적절한 화상을 획득할 수 있게 된다. 상호 참조 과정(155)은 실시예에 따라 생략 가능하다.

[0042] 또한, 보조화상처리부(153)는 형태적 유사성을 기반으로 텍스트 데이터 또는 교정된 텍스트 데이터 내의 단어 또는 어구에 대응하는 보조 화상을, 화상에 대한 데이터베이스(일례로 픽토그램 데이터베이스)로부터 검출하여 매핑할 수 있다. 이 경우, 보조화상처리부(153)는 N-그램(N-gram)을 기반으로 텍스트 데이터 또는 교정된 텍스

트 데이터 내의 단어나 어구에 대응하는 화상을 매핑할 수 있다. N-그램은 주어진 텍스트 또는 음성 샘플 내에서 추출된 연속적으로 나열된 N개의 아이템(문자, 단어나 구 등)을 의미한다. N-그램은 N개의 단어의 집합을 하나로 단위로 이용하여 텍스트 데이터(일례로 문장)를 처리하게 된다. 예를 들어, 텍스트 데이터는 1개의 단어의 집합(유니그램), 2개의 단어의 집합(바이그램) 또는 3개의 단어의 집합(트라이그램) 각각을 기반으로 처리될 수 있다. 보조화상처리부(153)는, 일 실시예에 있어서, N-그램부터 1-그램(유니그램)을 이용하여 전체 문장을 분석함으로써 단어나 문장 등에 가장 적절한 화상을 선택할 수 있다. 형태적 유사성 매핑에 따라 적어도 하나의 화상이 획득되면, 보조화상처리부(153)는 획득한 적어도 하나의 화상을 텍스트 데이터 내의 단어나 구에 대응하는 화상을 판단하고, 이를 출력할 화상으로 결정할 수 있다(158).

[0043] 만약 화상에 대한 데이터베이스에 매핑될 수 있는 화상이 부재한 경우라면, 보조화상처리부(153)는 의미론적 유사성을 기반으로 매핑을 수행할 수도 있다(157). 다시 말해서, 텍스트 데이터 내의 어휘에 대응하는 이미지가 데이터베이스에 존재하지 않되, 이와 같은 어휘와 의미가 동일 또는 유사한 화상이 데이터베이스에 존재하는 경우에는, 보조화상처리부(153)는 의미가 동일 또는 유사한 화상을 검출하여 매핑시킬 수 있다. 의미적으로 얼마나 유사한지 여부는 단어 임배딩(Word Embedding) 기술 등을 이용하여 판단될 수 있다. 단어 임배딩 알고리즘은, Word2Vec이나 WordNet 등을 포함할 수 있다. 의미론적 유사성을 기반으로 획득된 화상은, 형태적 유사성을 기반으로 획득된 화상과 마찬가지로 텍스트 데이터 내의 단어나 구에 대응하는 화상으로 판단되고, 보조화상처리부(153)에 의해 외부로 출력될 화상으로 결정될 수 있다(158).

[0044] 한편, 상술한 과정에서 상호 참조 해결(155), 형태적 유사도 분석(156) 또는 의미론적 유사도 분석(157)을 통해서도 처리되지 못하는 다른 단어나 어구(일례로 알려지지 않은 단어나, 조사, 접속사 및/또는 관사 등과 같은 문법 요소 등)은 처리되지 않거나 또는 필요에 따라 제거될 수 있으며, 이에 따라 이들 단어나 어구에 대응하는 화상은 외부로 출력되지 않게 된다. 이 경우, 처리되지 않은 단어는 그대로 출력부(120)를 통해 출력될 수도 있고 또는 출력부(120)를 통해 출력되지 않도록 처리될 수도 있다.

[0045] 보조화상처리부(153)는, 도 3에 도시된 바와 같이 상호 참조 해결(155) 및 형태적 유사성 매핑(156)을 순차적으로 이용하여 적절한 화상을 획득하되, 필요에 따라 의미론적 유사성 매핑(157)을 더 이용할 수도 있으나, 실시예에 따라서 형태적 유사성 매핑(156)을 먼저 수행하고 상호 참조 해결(155)을 수행할 수도 있고, 또는 형태적 유사성 매핑(156) 대신에 의미론적 유사성 매핑(157)만을 수행할 수도 있다. 이외에도 설계자가 고려 가능한 방법에 따라서 보조화상처리부(153)는 주어진 텍스트 데이터 또는 교정된 텍스트 데이터에 대응하는 적어도 하나의 화상(일례로 픽토그램)을 획득할 수 있다.

[0046] 도 4는 입력된 문장에 대응하는 화상의 일례를 도시한 도표이다. 도 4의 제1 행은 인식부(151)의 출력 결과를, 제2 행은 교정부(152)의 출력 결과를, 제3 행은 보조화상처리부(153)의 출력 결과를 의미한다. 동일한 열은 순차적으로 서로 대응되는 텍스트 데이터, 교정된 텍스트 데이터 및 적어도 하나의 보조 화상을 의미한다.

[0047] 인식부(151), 교정부(152) 및 보조화상처리부(153)의 동작에 의해, 도 4에 도시된 바와 같이, 입력된 문장(일례로 언어 장애인이 발화한 문장)에 대응하는 적어도 하나의 화상(p11 내지 p13)이 결정될 수 있다. 구체적으로 사용자가 발화하고, 이에 응하여 인식부(151)가 사용자의 발화에 대응하는 적어도 하나의 텍스트 데이터(일례로 I love danceing, I goes home 또는 You is happy)를 획득하면, 교정부(152)는 주어진 적어도 하나의 텍스트 데이터를 교정하여, 적어도 하나의 텍스트 데이터에 대응하는 적어도 하나의 교정된 텍스트 데이터(일례로 I love dancing, I go home 또는 You are happy)를 획득할 수 있다. 보조화상처리부(153)는 인식부(151)의 텍스트 데이터 또는 교정부(152)의 교정 텍스트 데이터를 기반으로 대응하는 적어도 하나의 보조 화상, 일례로 픽토그램을 결정할 수 있다. 예를 들어, 보조화상처리부(153)는 [I love dancing]이라는 문장을 수신하면, 이에 대응하여 [나(I)]에 대응하는 픽토그램(p11), [사랑(love)]에 대응하는 픽토그램(p12) 및 [춤(dancing)]에 대응하는 픽토그램(p13)을 각각 획득할 수 있다. 만약 be 동사처럼 대응하는 픽토그램이 부재한 경우에는 단어 그 자체(p31)를 픽토그램에 대체하여 선택할 수도 있다.

[0048] 도 5는 언어 대응 화상 출력 장치에 출력되는 화면의 일 실시예를 도시한 도면이다.

[0049] 도 5에 도시된 바를 참조하면, 일 실시예에 있어서, 인식부(151)가 획득한 텍스트 데이터나 교정부(152)가 교정한 교정 텍스트 데이터는, 보조화상처리부(153)가 획득한 적어도 하나의 보조 화상과 함께 출력부(121), 일례로 디스플레이(121a)에 표시될 수 있다. 구체적으로 예를 들어, 출력부(121)는 보조 화상(d1) 및 텍스트(d2, d3)를 함께 한 화면에 표시할 수 있다. 물론 실시예에 따라서 보조 화상(d1) 및 텍스트(d2, d3)가 순차적으로 디스플레이(121a)에 표시되는 것도 가능하다. 실시예에 따라서, 하나의 화면 상에는 서로 대응하되 상이한 언어로 표현된 복수의 텍스트(d2, d3)가 표시될 수도 있다. 일례로 동일한 의미를 갖되 하나는 영어로, 다른 하나는 한국

어로 표현된 복수의 텍스트(d2, d3)가 함께 디스플레이(121a)에 표시될 수도 있다. 물론, 실시예에 따라 디스플레이(121a)에는 오직 하나의 언어의 텍스트만 표시되는 것도 가능하다. 이 경우, 오직 하나의 언어는 대화자 양자 모두가 사용하는 언어를 포함할 수 있으며, 예를 들어, 발화를 한 대화자(일례로 언어 장애인, 보호자 또는 교사 등)가 사용한 언어를 포함할 수도 있다. 실시예에 따라서, 복수의 텍스트(d2, d3)는 비시각적 형태(일례로 청각적 형태)로 출력되는 것도 가능하다. 일 실시예에 의하면, 디스플레이(121a)에는 대화에 대응하는 화상 획득 및 출력 기능이 수행되고 있음을 알리거나 또는 화상을 획득하고 싶은 문장을 발화하라는 메시지(d4) 등이 더 표시될 수도 있다.

[0050] 이와 같이 발화자가 발화를 진행하면, 발화된 문장(d2, d3)과 발화된 문장(d2, d3)에 대응하는 화상, 일례로 픽토그램(d1)이 발화에 응하여 자동적으로 생성되어 출력부(121), 일례로 디스플레이(121a)에 표시되므로, 언어 장애인과 타인이 상호 대화하는 경우에도 보다 용이하고 신속하게 의사 소통이 가능해진다. 또한, 발화된 문장(d2, d3)과 발화된 문장(d2, d3)에 대응하는 화상(d1)은 동시에 또는 순차적으로 디스플레이(121a)에 표시되므로, 발화자는 문장(d2, d3)과 화상(d1)을 상호 비교하여 발화한 문장의 구체적인 의미를 보다 용이하게 파악할 수 있게 되고 이에 따라 언어 학습의 편의성도 개선될 수 있게 된다. 이에 따라 상술한 언어 대응 화상 출력 장치(100)는 발달 장애 아동을 비롯하여 언어를 구사하거나 학습하지 못하는 아동이나 청소년, 실어증 환자 및/또는 치매 환자 등을 치료하거나 재활하는 데에 이용될 수 있다. 또한, 병원, 복지관 또는 공공기관 등에서 언어 장애인의 접근성 향상을 위한 보조적 장치로 이용될 수도 있다.

[0051] 이하 도 6을 참조하여 언어에 대응하는 화상을 출력하기 위한 시스템의 일 실시예를 설명하도록 한다.

[0052] 도 6은 언어 대응 화상 출력 시스템의 일 실시예에 대한 도면이다.

[0053] 도 6에 도시된 바에 의하면, 언어 대응 화상 출력 시스템(200)은, 일 실시예에 있어서, 적어도 하나의 단말 장치(210)와, 적어도 하나의 단말 장치(210)와 통신 네트워크(201)를 통해 상호 데이터를 송수신할 수 있도록 마련된 서버 장치(220)를 포함할 수 있다.

[0054] 적어도 하나의 단말 장치(210)는 사용자의 발화나 조작에 따라 단어나 문장 등을 수신하여 획득하도록 마련된다. 적어도 하나의 단말 장치(210)는, 예를 들어, 스마트폰, 셀룰러폰, 휴대용 게임기, 태블릿 피씨, 랩톱 컴퓨터, 데스크톱 컴퓨터, 두부 장착형 디스플레이 장치, 스마트 시계, 인공 지능 스피커 또는 이외 발화나 조작에 따라 언어를 입력받을 수 있는 적어도 하나의 전자 장치를 포함할 수 있다.

[0055] 서버 장치(220)는 적어도 하나의 단말 장치(210)로부터 수신한 데이터에 대한 처리를 수행하고, 처리 결과를 단말 장치(210)로 전송할 수 있도록 마련된다. 여기서, 적어도 하나의 단말 장치(210)로부터 수신한 데이터는 음성 데이터를 포함할 수도 있고, 음성 데이터를 텍스트로 변환한 데이터를 포함할 수도 있으며, 또는 텍스트로 변환한 데이터를 교정한 교정된 텍스트 데이터를 포함할 수도 있다. 다시 말해서, 단말 장치(210)는 음성 데이터를 수신하고 이를 그대로 또는 압축 변환하여 서버 장치(220)로 전송할 수도 있고, 음성 데이터를 텍스트 데이터로 변환한 후 서버 장치(220)로 전송할 수도 있으며, 교정된 텍스트 데이터를 획득하고 이를 서버 장치(220)로 전송할 수도 있다.

[0056] 서버 장치(220)는 수신한 데이터에 따라서 필요한 처리를 수행할 수 있다. 즉, 서버 장치(220)는 음성 데이터를 수신한 경우에는 음성 데이터를 텍스트 데이터로 변환하고, 필요에 따라 이에 대응하는 교정된 텍스트 데이터를 획득하고, 텍스트 데이터 또는 교정된 텍스트 데이터에 대응하는 보조 화상(일례로 적어도 하나의 픽토그램)을 획득 및 배열할 수도 있다. 만약 텍스트 데이터를 수신한 경우에는 서버 장치(220)는 텍스트 데이터에 대응하는 보조 화상을 획득 및 비열하거나 또는 필요에 따라 텍스트 데이터를 교정된 텍스트 데이터로 변환하고 교정된 텍스트 데이터에 대응하는 보조 화상을 획득 및 배열할 수도 있다. 또한, 만약 교정된 텍스트 데이터를 수신한 경우에는 서버 장치(220)는 교정된 텍스트 데이터에 대응하는 보조 화상을 획득할 수도 있다. 획득된 보조 화상은 단말 장치(210)로 전송될 수 있다. 실시예에 따라, 텍스트 데이터 및/또는 텍스트 데이터에 대응하는 교정된 텍스트 데이터도 단말 장치(210)로 전송될 수 있다.

[0057] 일 실시예에 의하면, 서버 장치(220)는, 서버 장치(220)로 음성 데이터, 텍스트 데이터 및/또는 교정된 텍스트 데이터를 전송한 단말 장치(211)와 동일한 단말 장치(211)로 보조 화상을 전송하거나 보조 화상에 더불어 텍스트 데이터 및/또는 교정된 텍스트 데이터를 더 전송할 수도 있다. 또한, 다른 실시예에 의하면, 서버 장치(220)는 텍스트 데이터 및/또는 교정된 텍스트 데이터를 전송한 단말 장치(211)와 상이한 단말 장치(212)로 보조 화상을 전송하거나 텍스트 데이터 및/또는 교정된 텍스트 데이터를 더 전송하는 것도 가능하다. 또 다른 실시예에 의하면, 서버 장치(220)는 양 단말 장치(211, 212) 모두로 보조 화상을 전송하거나, 텍스트 데이터 및/또는

교정된 텍스트 데이터를 더 전송할 수도 있다. 이에 따라 대화자 중 적어도 한 명이 언어 장애인인 상황에서, 적어도 하나의 언어 장애인, 적어도 하나의 다른 대화자 또는 이들 모두가 보조 화상, 텍스트 데이터 및/또는 교정된 텍스트 데이터를 수신하여 확인할 수 있게 된다. 서버 장치(220)는 서버용 컴퓨팅 장치, 랩톱 컴퓨터, 데스크톱 컴퓨터, 스마트폰, 태블릿피씨, 인공 지능 스피커, 셋톱 박스 및/또는 휴대용 게임기 등 외부와 통신을 수행할 수 있으면서, 데이터의 연산 처리가 가능한 적어도 하나의 전자 장치를 이용하여 구현 가능하다.

[0058] 통신 네트워크(201)는 유선 통신 네트워크, 무선 통신 네트워크 또는 이들의 조합을 이용하여 구축된 것일 수 있다. 여기서, 무선 통신 네트워크는 근거리 통신 네트워크(와이 파이(Wi-Fi), 캔(CAN) 통신, 지그비(zigbee) 또는 블루투스(Bluetooth) 등) 및 원거리 통신 네트워크(3GPP, 3GPP2 또는 와이맥스 계열 등의 이동 통신 표준을 기반으로 구현된 이동 통신 네트워크 등) 중 적어도 하나를 포함할 수 있다.

[0059] 단말 장치(210: 211, 212) 및 서버 장치(220) 적어도 하나가 수행하는 음성 데이터의 텍스트 데이터로의 변환, 텍스트 데이터의 교정 및/또는 보조 화상의 선택 및 배치 등의 처리 과정은 화상 출력 장치(100)를 통해 설명한 바와 동일하거나 일부 변형된 과정을 통하여 수행될 수 있다. 따라서, 이에 대한 자세한 설명은 이하 생략한다.

[0060] 이하 도 7을 참조하여 언어 대응 화상 출력 방법의 일 실시예를 설명하도록 한다.

[0061] 도 7은 언어 대응 화상 출력 방법의 일 실시예에 대한 흐름도이다.

[0062] 도 7에 도시된 언어 대응 화상 출력 방법의 일 실시예에 의하면, 먼저 사용자(일례로 언어 장애인이나 법적 보호자) 등이 발화를 하면, 언어에 대응하는 화상을 출력하기 위한 장치(일례로 스마트폰, 텔레비전 또는 컴퓨터 장치 등)는 발화에 따른 파동을 수신하고 수신한 발화에 대응하는 음성 데이터를 생성할 수 있다(300).

[0063] 음성 데이터가 생성되면, 음성 데이터에 대응하는 텍스트 데이터가 획득될 수 있다(302). 음성 데이터의 텍스트 데이터로의 변환은 스피치 투 텍스트 기능을 기반으로 수행 가능하다. 실시예에 따라서, 음성 데이터의 텍스트 데이터로의 변환을 수행하기 위해 딥 러닝과 같은 머신 러닝이 이용될 수 있다.

[0064] 순차적으로 필요에 따라 음성 데이터에 대응하는 텍스트 데이터에 대한 교정이 더 수행될 수 있다(304). 이에 따라 텍스트 데이터의 문법적 오류를 교정한 교정된 텍스트 데이터가 획득될 수 있다. 일 실시예에 의하면, 트랜스포머 기반의 시퀀스 투 시퀀스 학습 모델이 텍스트 데이터의 교정에 이용될 수 있으며, 이외에도 심층 신경망이나 합성곱 신경망 등 다양한 학습 알고리즘이 텍스트 데이터의 교정을 위해 이용될 수 있다. 텍스트 데이터의 교정 과정은 생략 가능하다.

[0065] 텍스트 데이터 또는 교정된 텍스트 데이터를 이용하여, 텍스트 데이터 또는 교정된 텍스트 데이터에 대응하는 적어도 하나의 화상이 검출 및 획득될 수 있다(306). 이 경우, 텍스트 데이터 또는 교정된 텍스트 데이터 내의 각각의 단어 또는 어구에 대해 적어도 하나의 화상이 결정될 수 있되, 일부의 단어 또는 어구(예를 들어, 관사나 사전에 없는 단어 등)에 대해서는 대응하는 화상이 결정되지 않을 수도 있다. 이들 화상은 언어의 어순에 따라 배치될 수도 있고 또는 설계자가 선택한 순서에 따라서 배치될 수도 있다. 보조 화상은 정지 화상 및/또는 동화상을 포함 가능하며, 정지 화상은 예를 들어 픽토그램을 포함할 수 있다. 적어도 하나의 화상의 획득은, 멀티모덜 처리를 기반으로 수행될 수 있다. 보다 구체적으로 텍스트 데이터 또는 교정된 텍스트 데이터 내의 단어나 어구에 대응하는 화상은, 적어도 하나의 화상은 N-그램을 이용하여 검출 및 매핑될 수 있다. 보다 구체적으로 예를 들어, 대응하는 화상의 검출은, 형태적 유사도 분석을 기반으로 수행될 수 있으며, 상호 참조 및/또는 의미론적 유사도 분석을 더 이용하여 수행될 수도 있다. 만약 형태적 유사도 분석, 상호 참조 및/또는 의미론적 유사도 분석에 의해서도 대응하는 화상이 검색되지 않으면, 텍스트 데이터 또는 교정된 텍스트 데이터 내의 해당 단어나 구 등에 대한 화상은 외부로 출력되지 않으며, 이 경우, 실시예에 따라서 해당 단어나 구 등이 그대로 외부로 출력될 수도 있고 또는 출력되지 않을 수도 있다. 실시예에 따라서, 적어도 하나의 화상의 획득은 복수의 화상으로 구축된 데이터베이스를 검색하여 텍스트 데이터 또는 교정된 텍스트 데이터 내의 단어나 어구에 대응하는 화상을 검출함으로써 수행될 수도 있다.

[0066] 획득된 화상은 외부로 출력될 수 있다(308). 이 경우, 획득된 화상과 화상에 대응하는 단어나 문장도 함께 외부로 시각적 또는 청각적으로 출력될 수 있다. 이에 따라 발화자 또는 발화자에 대한 대화자는, 발화자의 발화 내용을 시각적으로 용이하게 파악할 수 있게 되어, 양자 간의 의사 소통이 보다 원활해지게 된다. 또한, 언어 장애자가 발화자인 경우, 언어 장애자는 자신이 발화한 문장이 발화 의도나 내용에 부합하는 여부를 쉽게 확인할 수 있게 되므로, 언어 장애자의 언어 능력의 향상을 도모할 수 있게 된다.

[0067] 상술한 실시예에 따른 언어 대응 화상 출력 방법은, 컴퓨터 장치에 의해 구동될 수 있는 프로그램의 형태로 구현될 수 있다. 여기서 프로그램은, 프로그램 명령, 데이터 파일 및 데이터 구조 등을 단독으로 또는 조합하여

포함할 수 있다. 프로그램은 기계어 코드나 고급 언어 코드를 이용하여 설계 및 제작된 것일 수 있다. 프로그램은 상술한 방법을 구현하기 위하여 특별히 설계된 것일 수도 있고, 컴퓨터 소프트웨어 분야에서 통상의 기술자에게 기 공지되어 사용 가능한 각종 함수나 정의를 이용하여 구현된 것일 수도 있다. 또한, 여기서, 컴퓨터 장치는, 프로그램의 기능을 실현 가능하게 하는 프로세서나 메모리 등을 포함하여 구현된 것일 수 있으며, 필요에 따라 통신 장치를 더 포함할 수도 있다.

[0068] 상술한 언어 대응 화상 출력 방법을 구현하기 위한 프로그램은, 컴퓨터에 의해 판독 가능한 기록 매체에 기록될 수 있다. 컴퓨터에 의해 판독 가능한 기록 매체는, 예를 들어, 솔리드 스테이트 드라이브(SSD), 롬, 램 또는 플래시 메모리 등과 같은 반도체 저장 장치, 하드 디스크나 플로피 디스크 등과 같은 자기 디스크 저장 매체, 콤팩트 디스크나 디브이디 등과 같은 광 기록 매체, 플롭티컬 디스크 등과 같은 자기-광 기록 매체 및 자기 테이프 등 컴퓨터 등의 호출에 따라 실행되는 특정 프로그램을 저장 가능한 적어도 한 종류의 물리적 장치를 포함할 수 있다.

[0069] 이상 언어 대응 화상 출력 장치, 언어 대응 화상 출력 방법 및 언어 대응 화상 출력 시스템을, 여러 실시예를 들어 설명하였으나, 장치 및 방법은 오직 상술한 실시예에 한정되는 것은 아니다. 해당 기술 분야에서 통상의 지식을 가진 자가 상술한 실시예(들)를 기초로 수정 및 변형하여 구현 가능한 다양한 장치, 방법 또는 시스템 역시 상술한 언어 대응 화상 출력 장치, 언어 대응 화상 출력 방법 및 언어 대응 화상 출력 시스템의 일례가 될 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성 요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나 다른 구성 요소 또는 균등물에 의하여 대치되거나 또는 치환되더라도 상술한 언어 대응 화상 출력 장치, 언어 대응 화상 출력 방법 및 언어 대응 화상 출력 시스템의 일 실시예가 될 수 있다.

부호의 설명

[0070] 100: 화상 출력 장치 110: 입력부

120: 출력부 130: 저장부

150: 프로세서 151: 인식부

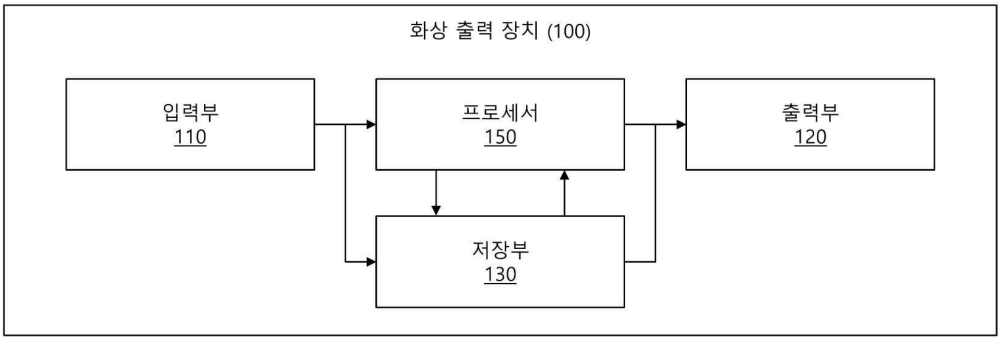
152: 교정부 153: 보조화상처리부

200: 화상 출력 시스템 210: 단말기

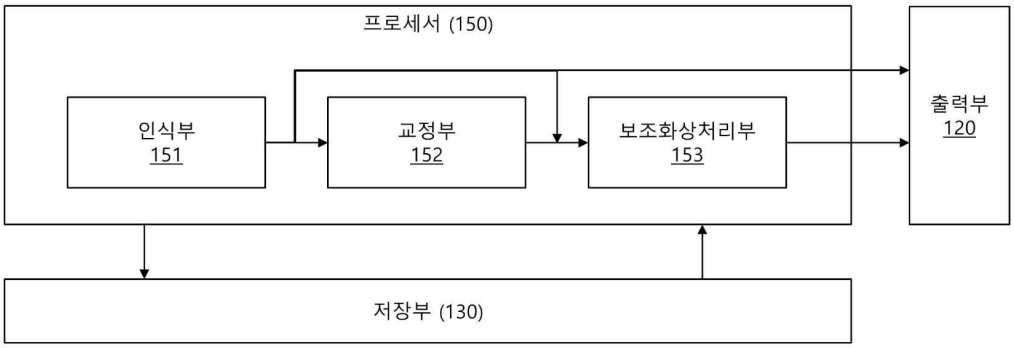
220: 서버 장치

도면

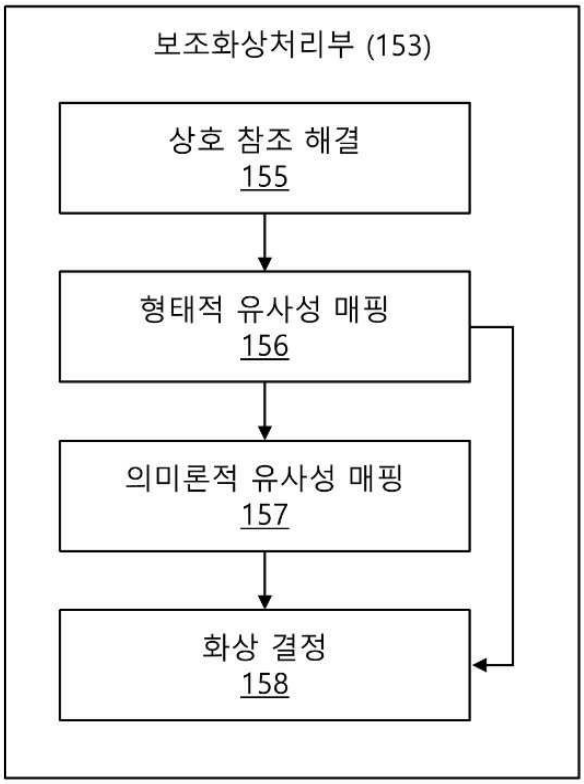
도면1



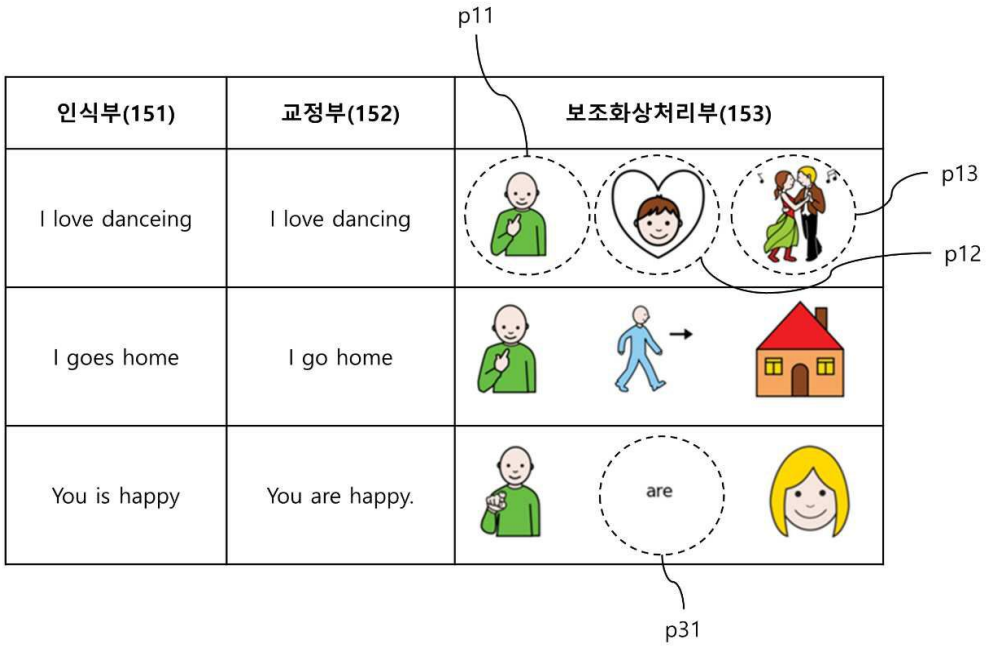
도면2


프로세서 (150)
인식부
151
교정부
152
보조화상처리부
153
출력부
120
저장부 (130)

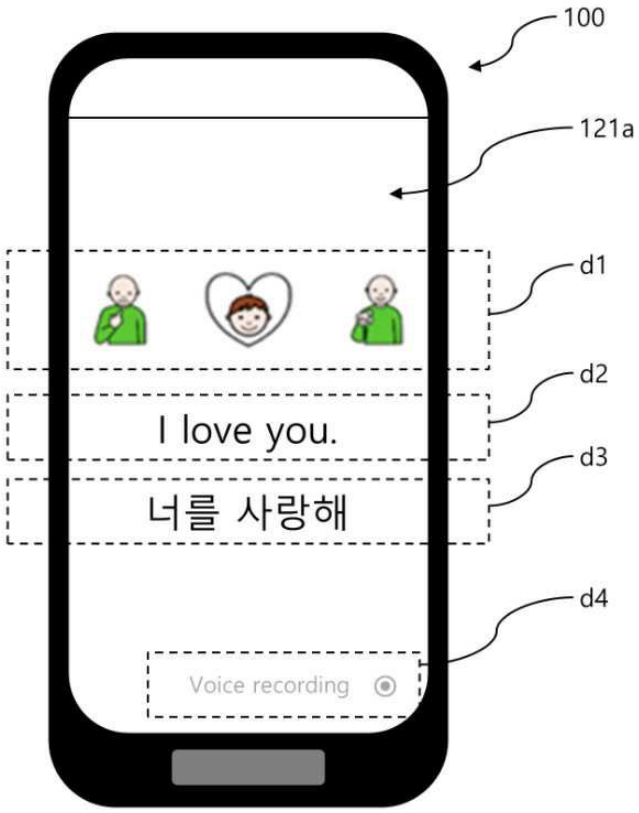
도면3


보조화상처리부 (153)
상호 참조 해결
155
형태적 유사성 매핑
156
의미론적 유사성 매핑
157
화상 결정
158

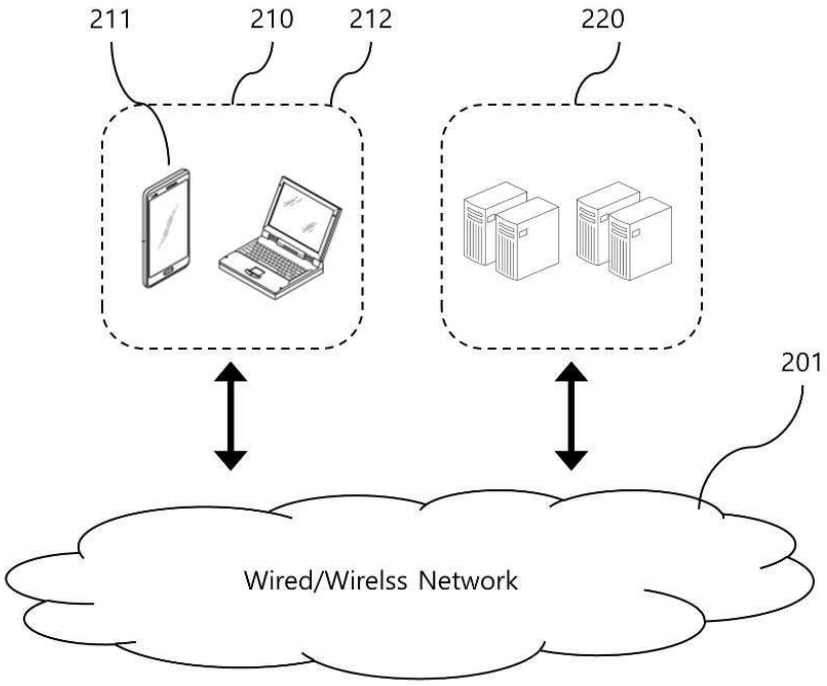
도면4


p11
인식부(151)
교정부(152)
보조화상처리부(153)
I love danceing
I love dancing
I goes home
I go home
You is happy
You are happy.
are
p13
p12
p31

도면5

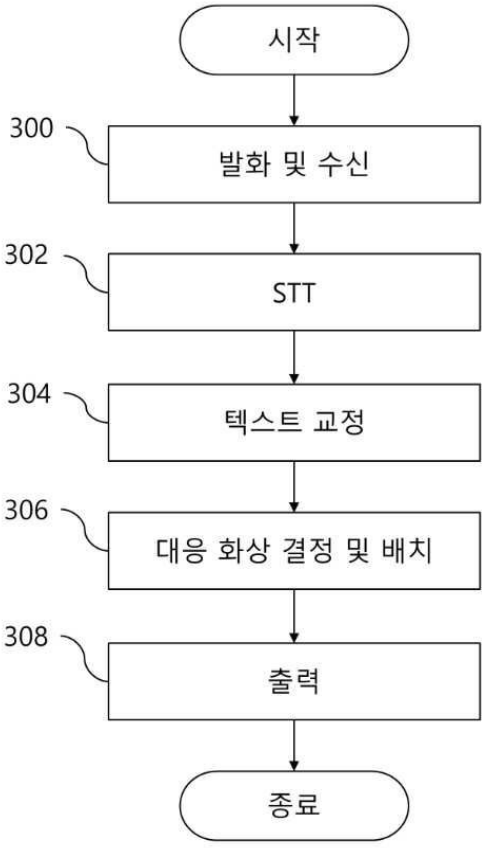

100
121a
d1
d2
I love you.
d3
너를 사랑해
d4
Voice recording

도면6


211
210
212
220
201
Wired/Wirelss Network

200

도면7


시작
300
발화 및 수신
302
STT
304
텍스트 교정
306
대응 화상 결정 및 배치
308
출력
종료