



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2022년07월05일  
(11) 등록번호 10-2415519  
(24) 등록일자 2022년06월28일

(51) 국제특허분류(Int. Cl.)  
G10L 25/51 (2013.01) G06N 3/08 (2006.01)  
G10L 13/08 (2006.01) G10L 15/02 (2006.01)  
G10L 25/30 (2013.01) G10L 25/69 (2013.01)  
HO4M 3/493 (2006.01)  
(52) CPC특허분류  
G10L 25/51 (2013.01)  
G06N 3/08 (2013.01)  
(21) 출원번호 10-2020-0123985  
(22) 출원일자 2020년09월24일  
심사청구일자 2020년09월24일  
(65) 공개번호 10-2022-0040813  
(43) 공개일자 2022년03월31일  
(56) 선행기술조사문헌  
JP11190996 A\*  
KR1020190131806 A\*  
Hany Farid 외, 'Detecting AI-Synthesized  
Speech Using Bispectral Analysis',  
Proceedings of the IEEE/CVF Conference on  
Computer Vision and Pattern Recognition  
(CVPR) Workshops, 2019, pp. 104-109\*  
JP2021078012 A  
\*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
장원준  
경기도 김포시 태장로 845, 한강센트럴자이 128동  
1803호(장기동, 한강센트럴자이)  
(72) 발명자  
장원준  
경기도 김포시 태장로 845, 한강센트럴자이 128동  
1803호(장기동, 한강센트럴자이)  
(74) 대리인  
신재열, 이영규, 윤병국

전체 청구항 수 : 총 12 항

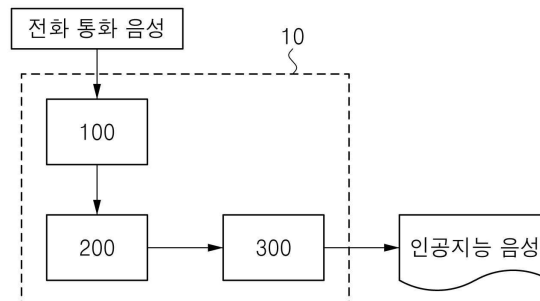
심사관 : 홍경아

(54) 발명의 명칭 인공지능 음성의 컴퓨팅 탐지 장치

(57) 요약

개시되는 발명은 적어도 하나 이상의 프로세서와, 상기 프로세서에 의해 처리되는 명령어들이 저장된 메모리를 포함하는 인공지능 음성의 컴퓨팅 탐지 장치에 관한 것으로서, 전화 통화 음성을 입력받는 음성 입력부;와, 상기 음성 입력부로 입력된 통화 음성의 파형을 분석하여, 실제 음성인지 아니면 인공지능 음성인지를 판별하는 음성 분석부; 및 상기 음성 분석부에서 판별한 결과를 출력하는 출력부;를 포함하는 인공지능 음성의 컴퓨팅 탐지 장치를 제공한다.

대표도 - 도1



(52) CPC특허분류

*G10L 13/08* (2013.01)

*G10L 15/02* (2013.01)

*G10L 25/30* (2013.01)

*G10L 25/69* (2013.01)

*H04M 3/4936* (2013.01)

---

**명세서**

**청구범위**

**청구항 1**

적어도 하나 이상의 프로세서와, 상기 프로세서에 의해 처리되는 명령어들이 저장된 메모리를 포함하는 인공지능 음성의 컴퓨팅 탐지 장치로서,

전화 통화 음성을 입력받는 음성 입력부;

상기 음성 입력부로 입력된 통화 음성의 파형을 분석하여, 실제 음성인지 아니면 인공지능 음성인지를 판별하는 음성 분석부; 및

상기 음성 분석부에서 판별한 결과를 출력하는 출력부;

를 포함하고,

상기 음성 분석부는,

상기 음성 입력부로 입력된 통화 음성의 파형을 각 음절마다 복수의 음편으로 분할하는 음편 생성부와,

상기 음편 생성부에서 만들어진 복수의 음편을 유사한 음편끼리 하나의 음편 그룹으로 묶고, 각 음편 그룹 중의 적어도 어느 한 음편 그룹에 포함된 개체수가 제1 기준치를 초과했을 때 해당 음편 그룹에 포함된 음편 사이의 유사도를 판정하며, 판정된 유사도가 제2 기준치를 초과했을 때 인공지능 음성인 것으로 판정하는 인공지능 모듈을 포함하는, 인공지능 음성의 컴퓨팅 탐지 장치.

**청구항 2**

제1항에 있어서,

상기 음성 분석부는,

텍스트 파일과, 상기 텍스트 파일을 사람이 낭독하는 음성 파일이 함께 있는 학습 파일을 검색하고 취득하는 탐색 모듈;

상기 탐색 모듈이 취득한 텍스트 파일을 인공지능 음성 생성 모듈에 입력하고, 상기 인공지능 음성 생성 모듈에서 출력된 인공지능 음성을 출력하는 인공지능 음성 출력 모듈; 및

상기 음성 파일과 인공지능 음성을 입력받고, 이들 사이의 파형상의 차이에 대해 기계학습하는 인공지능 모듈;

을 포함하는 인공지능 음성의 컴퓨팅 탐지 장치.

**청구항 3**

제2항에 있어서,

상기 인공지능 모듈은,

기계학습을 수행한 후, 상기 음성 입력부로 입력된 통화 음성의 파형을 분석하여, 실제 음성인지 아니면 인공지능 음성인지를 판별하는,

인공지능 음성의 컴퓨팅 탐지 장치.

**청구항 4**

제2항에 있어서,

상기 학습 파일은,

오디오 북 파일, 연설문 및 그 낭독 파일을 포함하는,

인공지능 음성의 컴퓨팅 탐지 장치.

**청구항 5**

제2항에 있어서,  
 상기 인공지능 음성 생성 모듈은,  
 통신으로 연결된 외부의 상용화된 어플리케이션 또는 프로그램인 것을 특징으로 하는,  
 인공지능 음성의 컴퓨팅 탐지 장치.

**청구항 6**

제1항 내지 제5항 중 어느 한 항에 있어서,  
 상기 인공지능 음성의 컴퓨팅 탐지 장치는,  
 스마트 단말기의 어플리케이션, 또는 통신사의 기지국이나 중계기의 하드웨어에 탑재되는 프로그램인 것을 특징  
 으로 하는,  
 인공지능 음성의 컴퓨팅 탐지 장치.

**청구항 7**

제6항에 있어서,  
 상기 출력부는,  
 상기 음성 분석부에서 판별한 결과를 앱 알림창, 단문 메시지, 메신저 중의 적어도 어느 하나로서 출력하는,  
 인공지능 음성의 컴퓨팅 탐지 장치.

**청구항 8**

삭제

**청구항 9**

제1항에 있어서,  
 상기 인공지능 모듈은,  
 제3 기준치를 초과하는 복수 개의 음편 그룹에서 제2 기준치를 초과하는 유사도가 나타났을 때 인공지능 음성인  
 것으로 판정하는, 인공지능 음성의 컴퓨팅 탐지 장치.

**청구항 10**

제1항에 있어서,  
 상기 인공지능 모듈은,  
 상기 음성 입력부로 입력된 통화 음성의 파형 중에 포함된 주변 소음, 호흡음, 침 삼키는 소리, 입맛을 다시는  
 소리, 혀 차는 소리를 포함하는 비언어적 음향요소의 빈도수가 제4 기준치 미만인 경우, 이 빈도수를 인공지능  
 음성의 판정에 보조적으로 반영하는, 인공지능 음성의 컴퓨팅 탐지 장치.

**청구항 11**

제1항에 있어서,  
 상기 인공지능 모듈은,  
 상기 음성 입력부로 입력된 통화 음성의 파형을 사전에 정해진 개수의 음절을 포함하는 음절 그룹으로 나누고,  
 각 음절 그룹의 전달 속도 사이의 표준편차가 제5 기준치 미만인 경우, 이 표준편차를 인공지능 음성의 판정에  
 보조적으로 반영하는, 인공지능 음성의 컴퓨팅 탐지 장치.

**청구항 12**

제1항, 제9항 내지 제11항 중 어느 한 항에 있어서,

상기 인공지능 음성의 컴퓨팅 탐지 장치는,

스마트 단말기의 어플리케이션, 또는 통신사의 기지국이나 중계기의 하드웨어에 탑재되는 프로그램인 것을 특징으로 하는, 인공지능 음성의 컴퓨팅 탐지 장치.

**청구항 13**

제12항에 있어서,

상기 출력부는,

상기 음성 분석부에서 판별한 결과를 앱 알림창, 단문 메시지, 메신저 중의 적어도 어느 하나로서 출력하는, 인공지능 음성의 컴퓨팅 탐지 장치.

**발명의 설명**

**기술 분야**

[0001] 본 발명은 인공지능 음성의 컴퓨팅 탐지 장치에 관한 것으로서, 사람의 실제 음성과 기계로 합성된 인공지능 음성을 구별할 수 있고, 인공지능 음성인 경우에 이를 사용자에게 고지함으로써 만일에 발생할 수도 있는 부정행위를 예방할 수 있도록 한다.

**배경 기술**

[0002] 인공지능(AI) 기술의 발전은 하루가 다르게 발전하고 있다. 2016년 알파고를 통해 인공지능이 전 세계의 관심을 끌게 되었고, 인공지능은 오랜 침체기를 거쳐 클라우드 컴퓨팅 환경의 급속한 발전과 빅데이터가 뒷받침되어 딥러닝이 구현되는 돌파구가 열리면서 전환기를 맞았다.

[0003] 인공지능의 연구는 다양한 방향으로 이루어지고 있으며, 그 중 하나로서 스마트폰이나 태블릿, 인공지능 스피커 등에 탑재되는 AI 비서(가상 비서)가 실생활에 깊숙히 접목되고 있다. 예를 들어, 몇 년 전에는 구글 어시스턴트라는 AI 비서가 사용자가 등록한 스케줄 정보에 기반하여 자동으로 미용실을 예약하는 케이스가 소개된 바 있으며, 놀랍게도 미용실 담당자는 통화 상대방이 AI 비서인지를 인지하지 못하였다.

[0004] 이와 같이, 고도화된 AI 비서와 통화하는 상대방이 마치 인간과 통화하는 것이라고 인식한 것은 튜링 테스트를 통과한 것이라 볼 수 있다. 이는 통화자와 AI 비서의 대화 문맥이 아주 자연스럽게 이어졌고, 통화 음성도 기계음성으로 느껴지지 않았기 때문이라고 볼 수 있다.

[0005] AI 기반 기술의 발전에 힘입어 앞으로는 AI 비서가 스스로 인간에게 전화를 하여 예약 또는 문의를 하는 빈도가 더욱 높아질 것으로 예상된다. 하지만, 인공지능 음성이 인간 목소리에 대한 딥페이크 등에 악용될 우려 또한 제기된다. 예컨대, 채무담당 직원에게 회사 대표의 목소리로 전화가 와서 특정 계좌로 입금을 지시한다거나, 또는 야전 지휘관에게 부대 통수권자로부터 전화가 와서 특정 지역의 공격을 명령한다는 등의 회복하기 어려운 손실과 혼란을 발생시킬 우려도 상상할 수 있다.

[0006] 이러한 인공지능 음성의 부정행위를 방지하고자 매번 당사자에게 확인 전화를 걸어야 하는 상황은 사회적 비용을 크게 증가시킬 것이며, 전화통신 인프라에 대한 신뢰도 하락으로 이어질 가능성이 크므로, 통신 사업자에게는 큰 위험이 된다.

[0007] 따라서, 대화하는 상대방이 인간인지 아닌지 여부를 탐지하고, 인공지능 음성으로 판정되었을 때 이를 고지할 수 있는 기술에 대한 개발이 필요하다.

**선행기술문헌**

**특허문헌**

[0008] (특허문헌 0001) 한국공개특허 제10-2020-0016516호 (2020.02.17 공개)

**발명의 내용**

**해결하려는 과제**

[0009] 본 발명은 통신 단말기를 통해 대화하는 상대방이 인간인지 아닌지 여부를 탐지하고, 인공지능 음성으로 판정되었을 때 이를 고지할 수 있는 효과적인 방안을 제공하는데 그 목적이 있다.

**과제의 해결 수단**

[0010] 본 발명은 적어도 하나 이상의 프로세서와, 상기 프로세서에 의해 처리되는 명령어들이 저장된 메모리를 포함하는 인공지능 음성의 컴퓨팅 탐지 장치로서, 전화 통화 음성을 입력받는 음성 입력부;와, 상기 음성 입력부로 입력된 통화 음성의 파형을 분석하여, 실제 음성인지 아니면 인공지능 음성인지를 판별하는 음성 분석부; 및 상기 음성 분석부에서 판별한 결과를 출력하는 출력부;를 포함하는 인공지능 음성의 컴퓨팅 탐지 장치를 제공한다.

[0011] 본 발명의 일 실시형태에서, 상기 음성 분석부는, 텍스트 파일 및 상기 텍스트 파일을 사람이 낭독하는 음성 파일이 함께 있는 학습 파일을 검색하고 취득하는 탐색 모듈;과, 상기 탐색 모듈이 취득한 텍스트 파일을 인공지능 음성 생성 모듈에 입력하고, 상기 인공지능 음성 생성 모듈에서 출력된 인공지능 음성을 출력하는 인공지능 음성 출력 모듈; 및 상기 음성 파일과 인공지능 음성을 입력받고, 이들 사이의 파형상의 차이에 대해 기계학습하는 인공지능 모듈;을 포함한다.

[0012] 그리고, 상기 인공지능 모듈은, 기계학습을 수행한 후, 상기 음성 입력부로 입력된 통화 음성의 파형을 분석하여, 실제 음성인지 아니면 인공지능 음성인지를 판별한다.

[0013] 여기서, 상기 학습 파일은, 오디오 복 파일이나 연설문 및 그 낭독 파일 등을 포함할 수 있다.

[0014] 그리고, 상기 인공지능 음성 생성 모듈은, 통신으로 연결된 외부의 상용화된 어플리케이션 또는 프로그램일 수 있다.

[0015] 그리고, 이러한 구성을 가진 인공지능 음성의 컴퓨팅 탐지 장치는, 스마트 단말기의 어플리케이션, 또는 통신사의 기지국이나 중계기의 하드웨어에 탑재되는 프로그램인 형태로 구현될 수 있다.

[0016] 그리고, 상기 출력부는, 상기 음성 분석부에서 판별한 결과를 앱 알림창, 단문 메시지, 메신저 중의 적어도 어느 하나로서 출력할 수 있다.

[0017] 한편, 본 발명의 다른 일 실시형태에 따르면, 상기 음성 분석부는, 상기 음성 입력부로 입력된 통화 음성의 파형을 각 음절마다 복수의 음편으로 분할하는 음편 생성부; 및 상기 음편 생성부에서 만들어진 복수의 음편을 유사한 음편끼리 하나의 음편 그룹으로 묶고, 각 음편 그룹 중의 적어도 어느 한 음편 그룹에 포함된 개체수가 제1 기준치를 초과했을 때 해당 음편 그룹에 포함된 음편 사이의 유사도를 판정하며, 판정된 유사도가 제2 기준치를 초과했을 때 인공지능 음성인 것으로 판정하는 인공지능 모듈;을 포함한다.

[0018] 여기서, 상기 인공지능 모듈은, 제3 기준치를 초과하는 복수 개의 음편 그룹에서 제2 기준치를 초과하는 유사도가 나타났을 때 인공지능 음성인 것으로 판정한다.

[0019] 그리고, 상기 인공지능 모듈은, 상기 음성 입력부로 입력된 통화 음성의 파형 중에 포함된 주변 소음, 호흡음, 침 삼키는 소리, 입맛을 다시는 소리, 혀 차는 소리를 포함하는 비언어적 음향요소의 빈도수가 제4 기준치 미만인 경우, 이 빈도수를 인공지능 음성의 판정에 보조적으로 반영할 수 있다.

[0020] 그리고, 상기 인공지능 모듈은, 상기 음성 입력부로 입력된 통화 음성의 파형을 사전에 정해진 개수의 음절을 포함하는 음절 그룹으로 나누고, 각 음절 그룹의 전달 속도 사이의 표준편차가 제5 기준치 미만인 경우, 이 표준편차를 인공지능 음성의 판정에 보조적으로 반영할 수 있다.

**발명의 효과**

[0021] 위와 같은 구성을 가진 본 발명의 인공지능 음성의 컴퓨팅 탐지 장치는, 통화자에게 현재 자신이 듣고 있는 음성이 인공지능 음성인지 여부에 대해 고지를 함으로써, 통화자가 중요한 판단과 결정을 내릴 때 불의의 피해와

손해를 받지 않을 적절한 안전책을 제공할 수 있다.

[0022] 따라서, 인공지능 음성의 부정행위를 방지하고자 매년 당사자에게 확인 전화를 걸어야 하는 등의 사회적 비용의 증가를 억제하고, 전화통신 인프라에 대한 신뢰도 하락의 위험을 줄일 수 있다.

**도면의 간단한 설명**

[0023] 도 1은 본 발명에 따른 인공지능 음성의 컴퓨팅 탐지 장치의 전체적인 구성을 개략적으로 도시한 도면.

도 2는 본 발명의 제1 실시형태에서의 음성 분석부의 구성을 개략적으로 도시한 도면

도 3은 본 발명의 제2 실시형태에서의 음성 분석부의 구성을 개략적으로 도시한 도면

도 4는 도 3의 음성 분석부에 비언어적 음향요소를 부가적 판단요소로 추가한 실시형태를 개략적으로 도시한 도면.

도 5는 도 3의 음성 분석부에 음질의 전달 속도를 부가적 판단요소로 추가한 실시형태를 개략적으로 도시한 도면.

**발명을 실시하기 위한 구체적인 내용**

[0024] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시 예를 상세히 설명한다. 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시 예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시 예에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시 예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다. 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성 요소를 지칭한다.

[0025] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있을 것이다. 또 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다. 본 명세서에서 사용된 용어는 실시 예들을 설명하기 위한 것이며 본 발명을 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다.

[0026] 본 명세서에서 사용되는 "포함한다 (comprises)" 및/또는 "포함하는 (comprising)"은 언급된 구성 요소, 단계, 동작 및/또는 소자는 하나 이상의 다른 구성 요소, 단계, 동작 및/또는 소자의 존재 또는 추가를 배제하지 않는다.

[0027] 그리고, 본 명세서에서 "컴퓨팅 장치" 또는 "모듈"이라 함은, 예를 들어, 프로세서, 컴퓨터 또는 다중 프로세서나 컴퓨터를 포함하여 데이터를 처리하기 위한 모든 기구, 장치 및 기계를 포함한다. 처리 시스템은, 하드웨어에 부가하여, 예를 들어, 프로세서 펌웨어를 구성하는 코드, 프로토콜 스택, 데이터베이스 관리 시스템, 운영 체제 또는 이들 중 둘 이상의 조합 등, 컴퓨터 프로그램에 대한 실행 환경을 형성하는 모든 요소를 포함할 수 있다. 프로그램, 소프트웨어, 소프트웨어 애플리케이션, 스크립트 또는 코드 등으로 알려진 컴퓨터 프로그램은 컴파일 되거나 해석된 언어 또는 선형적, 절차적 언어를 포함하는 프로그래밍 언어의 어떠한 형태로도 작성될 수 있으며, 독립형 프로그램이나 모듈, 컴포넌트, 서브루틴의 형태는 물론, 컴퓨터 환경에서 사용하기에 적합한 다른 유닛을 더 포함하는 다양한 형태로 구현될 수 있다.

[0029] 본 발명은 적어도 하나 이상의 프로세서와, 상기 프로세서에 의해 처리되는 명령어들이 저장된 메모리를 포함하는 인공지능 음성의 컴퓨팅 탐지 장치(10)로서 구현될 수 있다. 이하에서는, 첨부된 도면을 참조하여 본 발명에 대해 상세히 설명한다.

[0030] 도 1은 본 발명에 따른 인공지능 음성의 컴퓨팅 탐지 장치(10)의 전체적인 구성을 개략적으로 도시한 도면이다. 도면을 참조하면, 본 발명은 음성 입력부(100)와, 음성 분석부(200, 200'), 그리고 출력부(300)를 포함한다.

[0031] 음성 입력부(100)는 전화 통화 음성을 입력받는 구성요소이다. 통화 녹음 기능과 같이 전화 통화 음성을 따로 저장하는 기술은 공지된 기술이며, 이러한 기술을 음성 입력부(100)에 활용할 수 있다.

[0032] 음성 분석부(200, 200')는, 음성 입력부(100)로 입력된 통화 음성의 파형을 분석하고, 이를 통해 실제 음성인지 아니면 인공지능 음성인지를 판별하는 구성요소로서, 본 발명에서는 사전에 설계된 인공지능이 담당한다. 음성

분석부(200, 200')의 구체적인 실시형태는 두 가지 방안으로 제시되며, 이에 대해서는 뒤에서 상세히 설명한다.

- [0033] 출력부(300)는 음성 분석부(200, 200')에서 판별한 결과를 출력하는 구성요소다. 출력부(300)가 출력하는 판별 결과는 다양한 형태로 제공될 수 있다. 인공지능 음성인 경우 청각적인 경고음을 발생할 수도 있을 것이며, 그 외에 기록으로 저장할 수 있는 형태로 출력할 수도 있다.
- [0034] 예컨대, 본 발명의 인공지능 음성의 컴퓨팅 탐지 장치(10)는 스마트 단말기(예들 들어, 스마트폰, 통화기능이 있는 태블릿이나 스마트 워치 등)의 어플리케이션, 또는 통신사의 기지국이나 중계기의 하드웨어에 탑재되는 프로그램인 형태로 구현될 수 있다. 또한, 출력부(300)의 출력 형태는 본 발명이 탑재되는 대상이나 형태 등에 따라, 달라질 수 있는데, 예컨대 스마트 단말기의 앱 형태로 탑재된다면 출력부(300)는 음성 분석부(200, 200')에서 판별한 결과를 앱 알림창, 단문 메시지, 메신저 중의 적어도 어느 하나로서 출력할 수 있을 것이다.
- [0035] 도 2는 본 발명의 제1 실시형태에서의 음성 분석부(200)의 구성을 개략적으로 도시한 도면이다.
- [0036] 도 2에 도시된 제1 실시형태에서, 음성 분석부(200)는 탐색 모듈(210), 인공지능 음성 출력 모듈(220), 인공지능 음성 생성 모듈(230) 및 인공지능 모듈(240)을 포함한다.
- [0037] 탐색 모듈(210)은 일종의 "봇"으로서, 텍스트 파일 및 상기 텍스트 파일을 사람이 낭독하는 음성 파일이 함께 있는 학습 파일을 검색하고 취득하는 역할을 한다. 여기서, 학습 파일이란 기계학습에 사용되도록 탐색 모듈(210)이 취득한 파일을 의미하며, 탐색 모듈(210)은 인터넷 상에서 텍스트 파일과 그 음성 파일이 함께 있는 파일을 자동으로 탐색하고 수집한다.
- [0038] 여기서, 중요한 것은, 탐색 모듈(210)이 수집하는 학습 파일이 텍스트 파일과 그 음성 파일을 함께 구할 수 있는 것으로 특정되어 있다는 것이다. 탐색 모듈(210)이 수집한 텍스트 파일은 인공지능 음성 출력 모듈(220)에서 사용된다.
- [0039] 인공지능 음성 출력 모듈(220)은 탐색 모듈(210)이 취득한 텍스트 파일을 인공지능 음성 생성 모듈(230)에 입력하고, 인공지능 음성 생성 모듈(230)에서 출력된 인공지능 음성을 출력하는 역할을 한다. 즉, 인공지능 음성 출력 모듈(220)은, 텍스트 입력 기반의 인공지능 음성 생성 모듈(230)에 대해 탐색 모듈(210)이 취득한 텍스트 파일을 입력하고, 그 결과로서 만들어진 인공지능 음성을 출력하고 저장한다.
- [0040] 그리고, 인공지능 모듈(240)은 사람이 직접 발성한 음성 파일과, 인공지능 음성 출력 모듈(220)에서 생성된 인공지능 음성을 함께 입력받고, 이들 사이의 파형상의 차이에 대해 기계학습을 수행하는 구성요소이다.
- [0041] 제1 실시형태에서의 인공지능 모듈(240)은 기계학습 중에서 지도학습 모델이라고 할 수 있다. 즉, 사람의 음성인지 인공지능 음성인지 그 정답을 아는 상태에서, 다양한 파일을 입력받아 사람의 음성과 인공지능 음성 사이에 나타나는 파형상의 차이에 대해 학습을 수행함으로써, 기계학습을 마친 후에는 임의의 전화 통화 음성을 입력받았을 때 그 파형상의 특징이 어디에 해당하는지를 판단할 수 있는 것이다.
- [0042] 여기서 중요한 점은, 본 발명의 인공지능 모듈(240)은 단순히 사람의 음성과 인공지능 음성을 입력받는 것이 아니라, 동일한 텍스트 지문에 대한 사람의 음성과 인공지능 음성을 입력 받기에 기계학습에 필요한 시간을 절감하고 학습결과의 정확도까지 향상시킬 수 있다는 것에 있다. 다시 말해, 동일한 텍스트에 대한 사람의 음성과 인공지능 음성의 각 파형을 일대일로 비교하기 때문에 그 차이점을 명확히 파악할 수 있으며, 이로 인해 본 발명의 인공지능 음성의 컴퓨팅 탐지 장치(10)는 인공지능 음성을 상당히 정확하게 판정할 수 있게 된다.
- [0043] 이와 같이, 탐색 모듈(210)이 취득하는 학습 파일은 텍스트 파일과 그 음성 파일이 함께 있어야 하는데, 예를 든다면 소설 등의 텍스트를 성우가 읽어주는 오디오 북 파일이라든지, 공공행사 등에서 진행된 연설의 연설문 및 그 낭독 파일 등이 학습 파일로 사용될 수 있다.
- [0044] 그리고, 인공지능 음성 생성 모듈(230)은, 인공지능 음성의 컴퓨팅 탐지 장치(10)에 탑재되는 자체적인 모듈일 수도 있지만, 통신으로 연결된 외부의 상용화된 어플리케이션 또는 프로그램을 인공지능 음성 생성 모듈(230)로 사용하는 것이 바람직할 수 있다. 이는 인공지능 음성 생성 모듈(230)을 탑재하면 구동파일의 크기가 커져 실행에 지연이 발생할 수 있고, 상용화된 외부 어플리케이션 또는 프로그램은 주기적으로 업데이트가 이루어져 최신의 인공지능 음성 기술이 반영될 가능성이 높기에 별도의 개발비용을 들이지 않고도 인공지능 모듈(240)의 성능을 계속 향상시키는데 유리한 점이 있기 때문이다.
- [0045] 도 3은 본 발명의 제2 실시형태에서의 음성 분석부(200')의 구성을 개략적으로 도시한 도면이다.
- [0046] 도 3을 참조하면, 제2 실시형태의 음성 분석부(200')는 음편 생성부(250)와 인공지능 모듈(240')을 포함한다.

- [0047] 음편 생성부(250)는 음성 입력부(100)로 입력된 통화 음성의 파형을 각 음절마다 복수의 음편으로 분할하는 구성요소다. 여기서, 음편이란 하나의 음절 안에 포함된 최소의 음성요소라 말할 수 있다. 음편은 언어마다 다른데, 우리말은 한 음절을 형성하는 초성, 중성, 종성 각각으로 음편을 나눌 수 있다.
- [0048] 제2 실시형태의 음성 분석부(200')도 인공지능 모듈(240')을 포함하는데, 그 구조는 제1 실시형태와는 다르다. 제2 실시형태에서의 인공지능 모듈(240')은 먼저 음편 생성부(250)에서 만들어진 복수의 음편을 유사한 음편끼리 하나의 음편 그룹으로 묶는다. 이는 기계학습의 여러 유형 중 비지도학습 모델에 해당한다. 비지도학습 모델은 정답은 없이 수 많은 데이터들을 입력했을 때, 어떤 기준에 따라 데이터를 유형별로 묶는데 사용하기에 적합하다. 예를 들어, 비지도학습 모델을 이용하면 수 많은 종류와 개수의 사진을 입력했을 때, 형태나 색상 등의 기준에 따라 동물 사진, 과일 사진, 풍경 사진, 자동차 사진 등으로 유형화할 수 있다. 본 발명에서는 음편의 파형상 유사성을 기준으로 하여 복수의 음편 그룹으로 유형화한다.
- [0049] 음편 생성부(250)에서 만들어진 복수의 음편이 복수의 음편 그룹으로 묶인 다음에는, 각 음편 그룹 중의 적어도 어느 한 음편 그룹에 포함된 개체수가 제1 기준치를 초과했을 때 해당 음편 그룹에 포함된 음편 사이의 유사도를 판정한다. 이는 하나 이상의 음편 그룹 안에 충분한 개수의 표본이 모였을 때 그 음편들 사이의 유사도를 판정한다는 것이며, 그 유사도 판정 개시에 필요한 표본 개수의 문턱값이 제1 기준치에 해당함을 말한다.
- [0050] 그리고, 해당 음편 그룹(제1 기준치를 초과한 음편 그룹)에서의 음편 사이의 유사도, 즉 파형상의 유사도가 제2 기준치를 초과한다면, 이를 기준으로 음성 입력부(100)로 입력된 통화 음성을 인공지능 음성으로 판정하게 된다. 이는 해당 음편 그룹으로 묶인 음편들이 얼마나 유사한지를 한층 더 깊게 판단함을 의미한다. 사람의 음성은 동일한 음편을 발음한다고 하여도 통화의 길이가 길어질수록 기계가 아닌 이상 파형상으로는 차이가 발생하기 마련이다. 따라서, 하나의 음편 그룹에 속한 다수의 음편들이 파형상으로 별다른 차이가 없이 매우 유사하다면, 이는 해당 통화가 기계적으로 합성된 인공지능 음성임을 강하게 시사한다고 볼 수 있다.
- [0051] 여기서, 제1 기준치와 제2 기준치는 지도학습 모델의 기계학습을 통해 결정될 수 있다. 즉, 인공지능 모듈(240')의 기계학습 과정 중에는 사람의 음성인지 인공지능 음성인지를 알려준 상태에서 음편의 분석을 수행하게 되고, 이러한 학습 과정 중에 제1 기준치와 제2 기준치에 대한 최적의 값이 결정될 수 있다.
- [0052] 또한, 위와 같은 본 발명의 제2 실시형태는, 음편의 유사도 분석에 있어서 제3 기준치를 초과하는 복수 개의 음편 그룹에서 제2 기준치를 초과하는 유사도가 나타났을 때 인공지능 음성인 것으로 판정할 수 있다. 이는 제3 기준치를 초과하는 개수의 여러 음편 그룹에서 제2 기준치를 넘어서는 유사성이 발견되었을 때 좀더 확실히 인공지능 음성일 것이라 판정한다는 것이며, 제3 기준치도 지도학습 모델의 기계학습 과정에서 결정될 수 있다.
- [0053] 나아가, 제2 실시형태의 인공지능 모듈(240')은 음편이라는 언어적 요소 외에 다른 요소를 부가적인 판단요소로 사용할 수 있으며, 이러한 부가적 판단요소는 인공지능 음성인지의 판단을 강화하는 요소로 기능한다.
- [0054] 도 4는 이러한 부가적 판단요소로서, 비언어적 음향요소를 추가적으로 고려한 음성 분석부(200')의 구성을 개략적으로 보여준다. 도 4의 음성 분석부(200')는, 음성 입력부(100)로 입력된 통화 음성의 파형 중에 포함된 주변 소음, 호흡음, 침 삼키는 소리, 입맛을 다시는 소리, 혀 차는 소리와 같은 비언어적 음향요소의 빈도수가 제4 기준치 미만인 경우, 이 빈도수를 인공지능 음성의 판정에 보조적으로 반영하고 있다.
- [0055] 통상적인 통화 환경에는 주변 소음이 섞이는 경우가 비일비재하며, 통화 시간이 길어질수록 인간의 발성 메커니즘상 호흡음, 침 삼키는 소리, 입맛을 다시는 소리, 혀 차는 소리와 같은 비언어적 음향요소가 발생하기 쉽다. 따라서, 이와 같은 비언어적 음향요소가 매우 적게, 다시 말해 비언어적 음향요소의 빈도수가 제4 기준치 미만인 경우라면, 이는 해당 통화가 기계적으로 합성된 인공지능 음성이라는 판단을 강화하는 요소로 작용할 수 있다.
- [0056] 그리고, 도 5는 사람이 복수의 문장을 발성할 때의 말하는 속도(전달 속도)가 일정하기 어렵다는 사실을 또 다른 부가적 판단요소로 반영한 음성 분석부(200')의 실시형태를 도시하고 있다. 즉, 사람은 길게 말하는 동안에 폐에 채워놓은 공기를 많이 써버리면 문장을 끝맺기 위해 말하는 속도가 빨라지는 경향이 있고, 또한 통화 중 감정의 기복에 따라 말하는 속도에 변화가 생기기 마련이다.
- [0057] 도 5에 도시된 인공지능 모듈(240')은 이러한 특성을 반영한 것으로서, 음성 입력부(100)로 입력된 통화 음성의 파형을 사전에 정해진 개수(예를 들면, 10개나 20개)의 음절을 포함하는 음절 그룹으로 나누고, 각 음절 그룹의 전달 속도 사이의 표준편차가 제5 기준치 미만인 경우라면, 이 표준편차를 인공지능 음성의 판정에 보조적으로 반영할 수 있다.

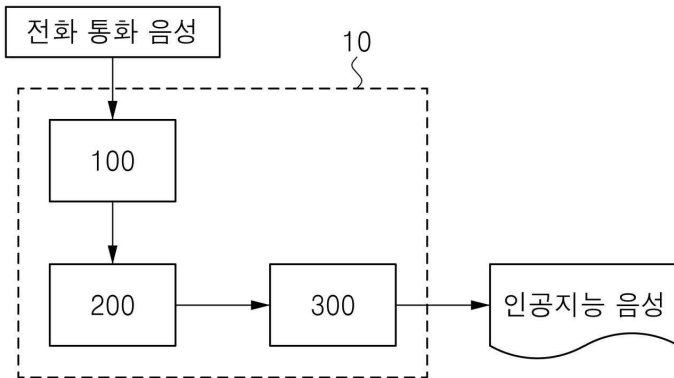
- [0058] 각 음절 그룹의 전달 속도 사이의 표준편차가 제5 기준치 미만이라는 것은 말하는 속도가 상당히 일정하다는 것을 의미하는바, 이러한 경우는 해당 통화가 기계적으로 합성된 인공지능 음성이라는 판단을 강화하는데 이용할 수 있다.
- [0059] 여기서, 도 4와 도 5의 실시형태를 서로 병렬적으로 양립할 수 있으며, 제4 기준치와 제5 기준치 역시 지도학습 모델의 기계학습 과정에서 결정될 수 있다.
- [0061] 이상과 같이 본 명세서에서 기술한 기술적 특징과 이를 실행하는 구현물은 디지털 전자 회로로 구현되거나, 본 명세서에서 기술하는 구조 및 그 구조적인 등가물 등을 포함하는 컴퓨터 소프트웨어, 펌웨어 또는 하드웨어로 구현되거나, 이들 중 하나 이상의 조합으로 구현될 수 있다. 또한 본 명세서에서 기술한 기술적 특징을 실행하는 구현물은 컴퓨터 프로그램 제품, 다시 말해 처리 시스템의 동작을 제어하기 위하여 또는 이것에 의한 실행을 위하여 유형의 프로그램 저장매체 상에 인코딩된 컴퓨터 프로그램 명령어에 관한 모듈로서 구현될 수도 있다.
- [0062] 본 발명의 장치는 하나 이상의 프로세서를 포함한다. 상기 프로세서는 프로그램의 명령어들을 실행하기 위한 것으로, 디지털 시그널 프로세서(digital signal processor(DSP), 마이크로 프로세서(microprocessor) 등으로 구현될 수 있는 중앙처리 장치(central processing unit(CPU)), MCU(Micro Controller Unit), MPU(micro processing unit), 컨트롤러 (controller), 어플리케이션 프로세서(application processor(AP)), 또는 커뮤니케이션 프로세서 (communication processor(CP)), ARM 프로세서 중 하나 또는 그 이상을 포함하거나, 해당 용어로 정의될 수 있다. 또한, 프로세서는 프로세싱 알고리즘이 내장된 SoC(System on Chip), LSI(large scale integration), FPGA(Field Programmable gate array) 형태로 구현될 수 있다.
- [0063] 본 발명은 메모리와 스토리지 같은 저장부를 포함할 수 있다. 상기 저장부에는 시스템의 처리 및 제어를 위한 프로그램 명령어와, 처리 과정 및 처리 후의 데이터가 저장될 수 있으며, ROM(Read Only Memory), RAM(Random Access Memory), EPROM(Erasable Programmable Read Only Memory), EEPROM(Electrically Erasable Programmable Read-Only Memory), 플래쉬(flash) 메모리, SRAM(Static RAM), HDD(Hard Disk Drive), SSD(Solid State Drive) 등으로 구현될 수 있다.
- [0064] 이상 첨부된 도면을 참조하여 본 발명의 실시 예들을 설명하였지만, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 그 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시 예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다.

**부호의 설명**

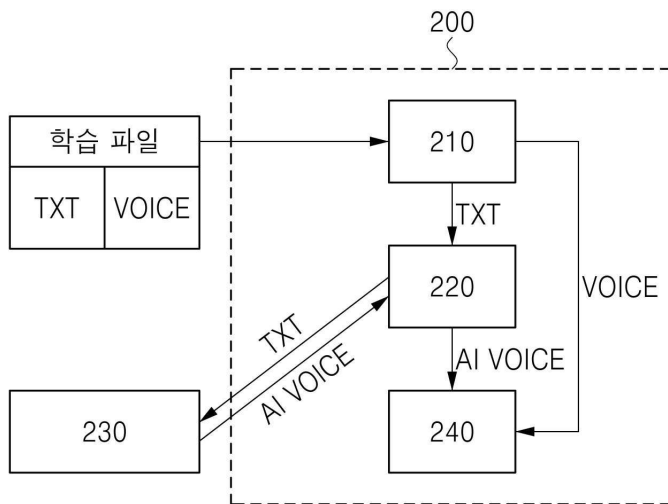
- [0065] 10: 인공지능 음성의 컴퓨팅 탐지 장치
- 100: 음성 입력부
- 200: 음성 분석부
- 210: 탐색 모듈
- 220: 인공지능 음성 출력 모듈
- 230: 인공지능 음성 생성 모듈
- 240: 인공지능 모듈
- 200': 음성 분석부
- 240': 인공지능 모듈
- 250: 음편 생성부
- 300: 출력부

도면

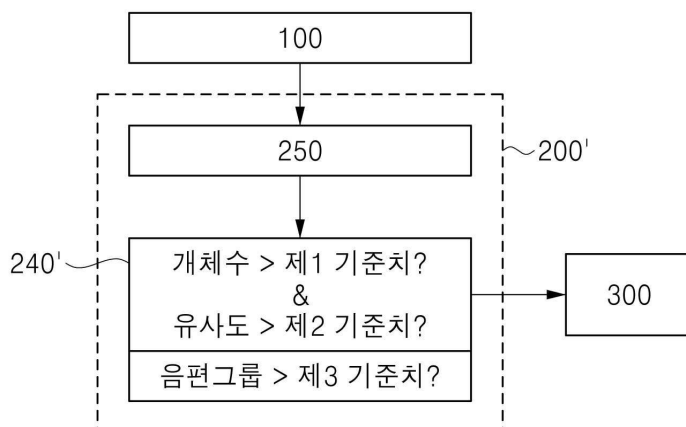
도면1



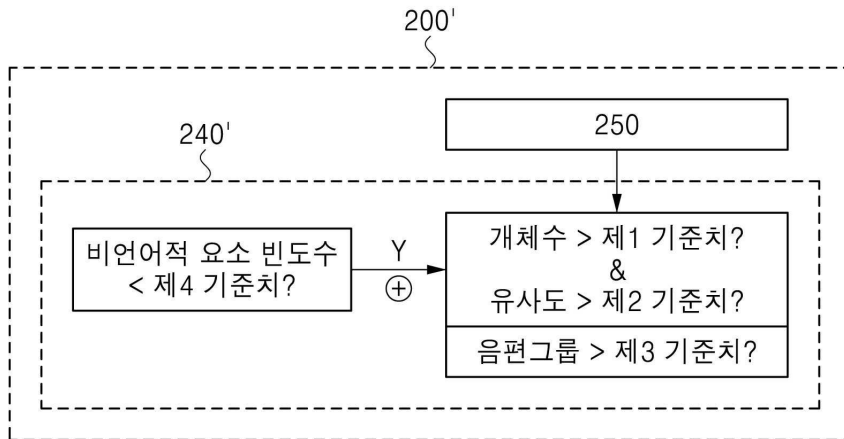
도면2



도면3



도면4



도면5

