

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일

2025 년 5 월 30 일 (30.05.2025)



(10) 국제공개번호

WO 2025/110602 A1

(51) 국제특허분류:

G10L 15/26 (2006.01)

G10L 21/055 (2013.01)

G06F 40/58 (2020.01)

G10L 21/0208 (2013.01)

H04N 21/488 (2011.01)

H04N 21/43 (2011.01)

(72) 발명자: 김홍국 (KIM, Hong Kook); 62254 광주광역시
광산구 첨단중앙로 181번길 42-4, 101동 1101호 (KR).
김미담 (KIM, Midam); 66049 캔자스주 로렌스,
펠리세이즈 드라이브 6120 (US).

(21) 국제출원번호:

PCT/KR2024/017832

(22) 국제출원일:

2024 년 11 월 12 일 (12.11.2024)

(25) 출원언어:

한국어

(26) 공개언어:

한국어

(30) 우선권정보:

10-2023-0165413 2023 년 11 월 24 일 (24.11.2023) KR

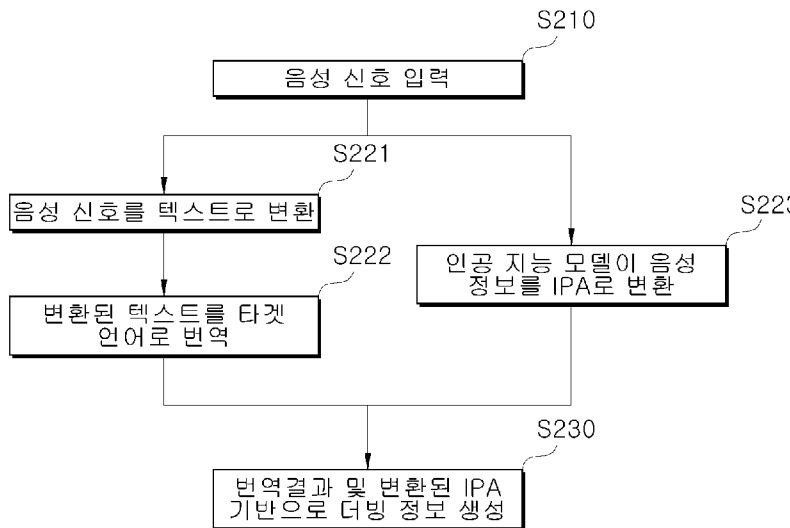
(71) 출원인: 주식회사 오니온에이아이 (AUNIONAI CO.
LTD.) [KR/KR]; 61005 광주광역시 북구 첨단과기로
123, 광주과학기술원 전기전자컴퓨터공학부 씨동
218호 (KR). 광주과학기술원 (GWANGJU INSTITUTE
OF SCIENCE AND TECHNOLOGY) [KR/KR]; 61005
광주광역시 북구 첨단과기로 123 (KR).

(74) 대리인: 특허법인 비엘티 (BLT PATENT & LAW
FIRM); 06595 서울특별시 서초구 법원로3길 25, 1층
(서초동, 태흥빌딩) (KR).

(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의
국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT,
AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH,
CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ,
EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR,
HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH,
KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD,
MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO,
NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,
SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(54) Title: DUBBING GENERATION SYSTEM AND METHOD TO WHICH VOICE TECHNOLOGY BASED ON LINGUISTICS
AND COGNITIVE SCIENCE IS APPLIED USING ARTIFICIAL INTELLIGENCE

(54) 발명의 명칭: 인공지능을 이용하여 언어학 및 인지과학 기반의 음성기술이 적용된 더빙 생성 시스템 및 방법



S210 ... Input voice signal

S221 ... Convert voice signal into text

S222 ... Translate converted text into target language

S223 ... Artificial intelligence model converts voice information into IPA

S230 ... Generate dubbing information on basis of translation result and converted IPA

(57) Abstract: The present disclosure relates to a dubbing generation system and method to which artificial intelligence-based voice technology is applied. The dubbing generation system according to the present disclosure comprises a server and a terminal, wherein the server comprises: a communication unit configured to receive video information from the terminal; and a processor including an artificial intelligence model for converting a voice signal included in the video information into text and generating dubbing information on the basis of the converted text, wherein the text is composed of international voice symbols.

[다음 쪽 계속]

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

(57) 요약서: 본 개시는 인공지능 기반 음성기술이 적용된 더빙 생성 시스템 및 방법에 관한 것이다. 본 개시에 따른 더빙 생성 시스템은 서버 및 단말기를 포함하고, 상기 서버는 상기 단말기로부터 영상 정보를 수신하도록 이루어지는 통신부, 상기 영상 정보에 포함된 음성 신호를 텍스트로 변환하고, 변환된 텍스트를 기반으로 더빙 정보를 생성하는 인공지능 모델을 포함하는 프로세서를 포함하고, 상기 텍스트는 국제음성기호로 구성되는 것을 특징으로 하는 더빙 생성 시스템을 제공할 수 있다.

발명의 설명

발명의 명칭: 인공지능을 이용하여 언어학 및 인지과학 기반의 음성기술이 적용된 더빙 생성 시스템 및 방법 기술분야

- [1] 본 개시는 인공지능 기반 음성기술이 적용된 더빙 생성 시스템 및 방법에 관한 것이다.

배경기술

- [2] 최근 미디어 콘텐츠 시장과 관련된 비디오 스트리밍 시장이 빠른 속도로 성장하고 있다. 또한, 비디오 스트리밍 시장이 국내가 아닌 글로벌 시장으로 확대됨에 따라, 더빙 생성 기술에 대한 니즈가 증가하고 있다.
- [3] 구체적으로, 점차 커지는 미디어 콘텐츠 시장에서 1인 창작자는 경쟁력을 높이기 위해 더빙 생성을 통한 고품질의 동영상을 제공하고 있다. 이에 따라, 더빙 생성 서비스 시장이 확대되고 있다.
- [4] 현재 더빙은 원어를 번역한 결과를 녹음하여 생성되기 때문에, 그 제작 시간이 오래걸릴 뿐만 아니라, 제작자의 언어 및 문화적 이해도에 따라 더빙의 의미 전달력이 상이해질 수 있다.
- [5] 또한, 더빙 생성의 경우, 원어민의 언어 및 감정을 전달하는데 어려움이 있으며, 이로 인해 영상자체가 부자연스러워질 수 있는 문제가 있다.
- [6] 이에, 원작이 제작된 국가의 언어 및 문화적 차이를 반영함과 동시에 영상에 삽입하더라도 부자연스러움이 없는 더빙을 생성할 수 있는 기술에 대한 니즈가 존재한다.

발명의 내용

기술적 과제

- [7] 본 개시는 인공지능을 통해 언어별 특징, 국가간 문화적 차이, 화자의 감정이 반영된 자연스러운 더빙을 생성할 수 있는 시스템 및 방법을 제공하는 것을 그 목적으로 한다.

과제 해결 수단

- [8] 상술한 목적을 달성하기 위해, 본 개시에 따른 더빙 생성 시스템은 서버 및 단말기를 포함하고, 상기 서버는 상기 단말기로부터 영상 정보를 수신하도록 이루어지는 통신부, 상기 영상 정보에 포함된 음성 신호를 텍스트로 변환하고, 변환된 텍스트를 기반으로 더빙 정보를 생성하는 인공지능 모델을 포함하는 프로세서를 포함하고, 상기 텍스트는 국제음성기호로 구성되는 것을 특징으로 하는 더빙 생성 시스템을 제공할 수 있다.
- [9] 일 실시 예에 있어서, 상기 프로세서는 상기 변환된 텍스트 및 상기 음성 신호에 대한 번역 결과를 기반으로, 상기 음성 정보에 대한 더빙 정보를 생성할 수 있다.

- [10] 일 실시 예에 있어서, 상기 프로세서는 상기 변환된 텍스트를 기반으로, 상기 프로세서는 상기 번역 결과를 국제음성기호로 구성된 정보로 변환하고, 변환된 정보를 기반으로 더빙 정보를 생성할 수 있다.
- [11] 일 실시 예에 있어서, 상기 프로세서는 상기 영상 정보에서 상기 음성 신호를 제거하고, 상기 더빙 정보를 결합할 수 있다.
- [12] 일 실시 예에 있어서, 상기 인공지능 모델은 트랜스포머(transformer)일 수 있다.
- [13] 또한, 본 개시에 따른 서버 및 단말기를 포함하는 더빙 생성 시스템의 더빙 생성 방법은, 상기 서버가 상기 단말기로부터 영상 정보를 수신하는 단계; 상기 서버가 상기 영상 정보에 포함된 음성 신호를 텍스트로 변환하는 단계; 및 상기 서버가 상기 변환된 텍스트를 기반으로 더빙 정보를 생성하는 단계를 포함하고, 상기 텍스트는 국제음성기호로 구성될 수 있다.

발명의 효과

- [14] 본 개시에 따르면, 지역적, 사회적, 문화적, 언어적 차이가 반영된 더빙을 생성할 수 있게 된다. 이에 따라, 본 개시에 따른 더빙 생성 시스템은 원어민이 발음하는 듯한 자연스러운 더빙을 생성할 수 있게 된다.
- [15] 또한, 본 개시에 따르면, 화자가 더빙 생성을 위해 음성을 녹음할 필요없이 빠른 속도로 더빙을 제작하는 것이 가능해진다.

도면의 간단한 설명

- [16] 도 1은 본 개시의 전반적 시스템 도면이다.
- [17] 도 2는 본 개시의 더빙 생성 시스템에 포함된 서버의 블록도이다.
- [18] 도 3은 본 개시의 더빙 생성 시스템에 포함된 단말기의 블록도이다.
- [19] 도 4는 본 개시의 더빙 생성 시스템에 포함된 프로세서의 블록도이다.
- [20] 도 5는 자막 및 더빙 제공 서비스를 나타내는 개념도이다.
- [21] 도 6은 본 개시에 따른 국제음성기호를 활용하여 더빙을 생성하는 방법의 흐름도이다.
- [22] 도 7은 본 개시에 따른 더빙 생성 시스템이 음성 신호를 국제음성기호로 구성된 텍스트로 변환하는 일 실시 예를 나타내는 개념도이다.

발명의 실시를 위한 형태

- [23] 본 개시 전체에 걸쳐 동일 참조 부호는 동일 구성요소를 지칭한다. 본 개시가 실시예들의 모든 요소들을 설명하는 것은 아니며, 본 개시가 속하는 기술분야에서 일반적인 내용 또는 실시예들 간에 중복되는 내용은 생략한다. 명세서에서 사용되는 '부, 모듈, 부재, 블록'이라는 용어는 소프트웨어 또는 하드웨어로 구현될 수 있으며, 실시예들에 따라 복수의 '부, 모듈, 부재, 블록'이 하나의 구성요소로 구현되거나, 하나의 '부, 모듈, 부재, 블록'이 복수의 구성요소들을 포함하는 것도 가능하다.

- [24] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 직접적으로 연결되어 있는 경우뿐 아니라, 간접적으로 연결되어 있는 경우를 포함하고, 간접적인 연결은 무선 통신망을 통해 연결되는 것을 포함한다.
- [25] 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.
- [26] 명세서 전체에서, 어떤 부재가 다른 부재 "상에" 위치하고 있다고 할 때, 이는 어떤 부재가 다른 부재에 접해 있는 경우뿐 아니라 두 부재 사이에 또 다른 부재가 존재하는 경우도 포함한다.
- [27] 제 1, 제 2 등의 용어는 하나의 구성요소를 다른 구성요소로부터 구별하기 위해 사용되는 것으로, 구성요소가 전술된 용어들에 의해 제한되는 것은 아니다.
- [28] 단수의 표현은 문맥상 명백하게 예외가 있지 않는 한, 복수의 표현을 포함한다.
- [29] 각 단계들에 있어 식별부호는 설명의 편의를 위하여 사용되는 것으로 식별부호는 각 단계들의 순서를 설명하는 것이 아니며, 각 단계들은 문맥상 명백하게 특정 순서를 기재하지 않는 이상 명기된 순서와 다르게 실시될 수 있다.
- [30] 이하 첨부된 도면들을 참고하여 본 개시의 작용 원리 및 실시예들에 대해 설명한다.
- [31] 본 명세서에서 '본 개시에 따른 시스템'은 연산처리를 수행하여 사용자에게 결과를 제공할 수 있는 다양한 장치들이 모두 포함된다. 예를 들어, 본 개시에 따른 시스템은, 컴퓨터, 서버 장치 및 휴대용 단말기를 모두 포함하거나, 또는 어느 하나의 형태가 될 수 있다.
- [32] 여기에서, 상기 컴퓨터는 예를 들어, 웹 브라우저(WEB Browser)가 탑재된 노트북, 데스크톱(desktop), 랩톱(laptop), 태블릿 PC, 슬레이트 PC 등을 포함할 수 있다.
- [33] 상기 서버 장치는 외부 장치와 통신을 수행하여 정보를 처리하는 서버로써, 애플리케이션 서버, 컴퓨팅 서버, 데이터베이스 서버, 파일 서버, 게임 서버, 메일 서버, 프록시 서버 및 웹 서버 등을 포함할 수 있다.
- [34] 상기 휴대용 단말기는 예를 들어, 휴대성과 이동성이 보장되는 무선 통신 장치로서, PCS(Personal Communication System), GSM(Global System for Mobile communications), PDC(Personal Digital Cellular), PHS(Personal Handyphone System), PDA(Personal Digital Assistant), IMT(International Mobile Telecommunication)-2000, CDMA(Code Division Multiple Access)-2000, W-CDMA(W-Code Division Multiple Access), WiBro(Wireless Broadband Internet) 단말, 스마트폰(Smart Phone) 등과 같은 모든 종류의 핸드헬드(Handheld) 기반의 무선 통신 장치와 시계, 반지, 팔찌, 발찌, 목걸이, 안경, 콘택트 렌즈, 또는 머리 착용형 장치(head-mounted-device(HMD)) 등과 같은 웨어러블 장치를 포함할 수 있다.
- [35] 이하에서는, 본 개시에 따른 더빙 생성 시스템에 대하여 설명한다.

- [36] 도 1을 참고하면, 본 개시에 따른 더빙 생성 시스템은 서버(10) 및 단말기(20) 중 적어도 하나를 포함할 수 있다. 구체적으로, 본 개시에 따른 더빙 생성 시스템은 서버(10) 또는 단말기(20)에 의해 단독으로 구현되거나, 서버(10) 및 단말기(20) 중 적어도 하나를 포함하는 시스템 형태로 구현될 수 있다. 이하에서 설명하는 더빙 생성 시스템에 관한 설명은 본 개시에 따른 더빙 생성 시스템이 서버(10) 또는 단말기(20)에 의해 단독으로 구현되거나, 서버(10) 및 단말기(20) 중 적어도 하나를 포함하는 시스템 형태로 구현되는 경우에 모두 적용될 수 있다.
- [37] 서버(10)는 적어도 하나의 단말기(20)와 네트워크로 연결되며, 복수의 단말기 각각에 정보를 전송하고, 단말기(20) 중 적어도 하나로부터 수신된 정보에 기반하여 더빙 생성 학습에 필요한 데이터를 생성한다.
- [38] 한편, 상기 단말기(20)는 상술한 휴대용 단말기에 한정되지 않고, 프로세서가 탑재된 노트북, 데스크톱(desktop), 랩톱(laptop), 태블릿 PC, 슬레이트 PC 등을 포함할 수 있는 것은 통상의 기술자에게 자명하다.
- [39] 이하에서는, 본 개시에 따른 더빙 생성 시스템을 구현하기 위한 서버(10) 및 단말기(20) 각각에 대하여 설명한다.
- [40] 도 2는 본 개시의 더빙 생성 시스템에 포함된 서버의 블록도이다.
- [41] 본 개시에 따른 서버(100)는 통신부(110), 저장부(120) 및 프로세서(130) 중 적어도 하나를 포함할 수 있다.
- [42] 통신부(110)는 단말기, 외부 저장소(예를 들어, 데이터베이스(database, 140)), 외부 서버 및 클라우드 서버 중 적어도 하나와 통신을 수행할 수 있다.
- [43] 한편, 외부 서버 또는 클라우드 서버에서는, 프로세서(130)의 적어도 일부의 역할을 수행하도록 구성될 수 있다. 즉, 데이터 처리 또는 데이터 연산 등의 수행은 외부 서버 또는 클라우드 서버에서 이루어지는 것이 가능하며, 본 발명에서는 이러한 방식에 대한 특별한 제한을 두지 않는다.
- [44] 한편, 통신부(110)는 통신하는 대상(예를 들어, 전자기기, 외부 서버, 디바이스 등)의 통신 규격에 따라 다양한 통신 방식을 지원할 수 있다.
- [45] 예를 들어, 통신부(110)는, WLAN(Wireless LAN), Wi-Fi(Wireless-Fidelity), Wi-Fi(Wireless Fidelity) Direct, DLNA(Digital Living Network Alliance), WiBro(Wireless Broadband), WiMAX(World Interoperability for Microwave Access), HSDPA(High Speed Downlink Packet Access), HSUPA(High Speed Uplink Packet Access), LTE(Long Term Evolution), LTE-A(Long Term Evolution-Advanced), 5G(5th Generation Mobile Telecommunication), 블루투스(Bluetooth™), RFID(Radio Frequency Identification), 적외선 통신(Infrared Data Association; IrDA), UWB(Ultra-Wideband), ZigBee, NFC(Near Field Communication), Wi-Fi Direct, Wireless USB(Wireless Universal Serial Bus) 기술 중 적어도 하나를 이용하여, 통신 대상과 통신하도록 이루어질 수 있다.
- [46] 다음으로 저장부(120)는, 본 발명과 관련된 다양한 정보를 저장하도록 이루어질 수 있다. 본 발명에서 저장부(120)는 본 발명에 따른 장치 자체에 구비될 수 있

다. 이와 다르게, 저장부(120)의 적어도 일부는, 데이터베이스(database: DB, 140) 클라우드 저장소(또는 클라우드 서버) 중 적어도 하나를 의미할 수 있다. 즉, 저장부(120)는 본 발명에 따른 장치 및 방법을 위하여 필요한 정보가 저장되는 공간이면 충분하며, 물리적인 공간에 대한 제약은 없는 것으로 이해될 수 있다. 이에, 이하에서는, 저장부(120), 데이터베이스(140), 외부 저장소, 클라우드 저장소(또는 클라우드 서버)를 별도로 구분하지 않고, 모두 저장부(120)라고 표현하도록 한다.

- [47] 다음으로, 프로세서(130)는 본 발명과 관련된 장치의 전반적인 동작을 제어하도록 이루어질 수 있다. 프로세서(130)는 위에서 살펴본 구성요소들을 통해 입력 또는 출력되는 신호, 데이터, 정보 등을 처리하거나 사용자에게 적절한 정보 또는 기능을 제공 또는 처리할 수 있다.
- [48] 프로세서(130)는 적어도 하나의 CPU(Central Processing Unit, 중앙처리장치)를 포함하여, 본 발명에 따른 기능을 수행할 수 있다.
- [49] 도 2에 도시된 구성 요소들의 성능에 대응하여 적어도 하나의 구성요소가 추가되거나 삭제될 수 있다. 또한, 구성 요소들의 상호 위치는 장치의 성능 또는 구조에 대응하여 변경될 수 있다는 것은 당해 기술 분야에서 통상의 지식을 가진 자에게 용이하게 이해될 것이다.
- [50] 이하, 본 개시의 더빙 생성 시스템에 포함된 단말기에 대하여 구체적으로 설명한다.
- [51] 도 3은 본 개시의 더빙 생성 시스템에 포함된 단말기의 블록도이다.
- [52] 도 3을 참고하면, 본 개시에 따른 단말기(200)는 통신부(210), 입력부(220), 표시부(230) 및 프로세서(240) 등을 포함할 수 있다. 도 3에 도시된 구성요소들은 본 개시에 따른 더빙 생성 시스템을 구현하는데 있어서 필수적인 것은 아니어서, 본 명세서 상에서 설명되는 단말기는 위에서 열거된 구성요소들 보다 많거나, 또는 적은 구성요소들을 가질 수 있다.
- [53] 상기 구성요소들 중 통신부(210)는 외부 장치와 통신을 가능하게 하는 하나 이상의 구성 요소를 포함할 수 있으며, 예를 들어, 방송 수신 모듈, 유선통신 모듈, 무선통신 모듈, 근거리 통신 모듈, 위치정보 모듈 중 적어도 하나를 포함할 수 있다.
- [54] 유선 통신 모듈은, 지역 통신(Local Area Network; LAN) 모듈, 광역 통신(Wide Area Network; WAN) 모듈 또는 부가가치 통신(Value Added Network; VAN) 모듈 등 다양한 유선 통신 모듈뿐만 아니라, USB(Universal Serial Bus), HDMI(High Definition Multimedia Interface), DVI(Digital Visual Interface), RS-1302(recommended standard1302), 전력선 통신, 또는 POTS(plain old telephone service) 등 다양한 케이블 통신 모듈을 포함할 수 있다.
- [55] 무선 통신 모듈은 와이파이(Wifi) 모듈, 와이브로(Wireless broadband) 모듈 외에도, GSM(global System for Mobile Communication), CDMA(Code Division Multiple Access), WCDMA(Wideband Code Division Multiple Access), UMTS(universal

mobile telecommunications system), TDMA(Time Division Multiple Access), LTE(Long Term Evolution), 4G, 5G, 6G 등 다양한 무선 통신 방식을 지원하는 무선 통신 모듈을 포함할 수 있다.

- [56] 입력부(220)는 영상 정보(또는 신호), 오디오 정보(또는 신호), 데이터, 또는 사용자로부터 입력되는 정보의 입력을 위한 것으로서, 적어도 하나의 카메라, 적어도 하나의 마이크론 및 사용자 입력부 중 적어도 하나를 포함할 수 있다. 입력부에서 수집한 음성 데이터나 이미지 데이터는 분석되어 사용자의 제어명령으로 처리될 수 있다.
- [57] 표시부(230)는 시각, 청각 또는 촉각 등과 관련된 출력을 발생시키기 위한 것으로, 디스플레이부, 음향 출력부, 햅틱 모듈 및 광 출력부 중 적어도 하나를 포함할 수 있다. 디스플레이부는 터치 센서와 상호 레이어 구조를 이루거나 일체형으로 형성됨으로써, 터치 스크린을 구현할 수 있다. 이러한 터치 스크린은, 본 장치와 사용자 사이의 입력 인터페이스를 제공하는 사용자 입력부로써 기능함과 동시에, 본 장치와 사용자 간에 출력 인터페이스를 제공할 수 있다.
- [58] 디스플레이부는 본 장치에서 처리되는 정보를 표시(출력)한다. 예를 들어, 디스플레이부는 본 장치에서 구동되는 응용 프로그램(일 예로, 어플리케이션)의 실행화면 정보, 또는 이러한 실행화면 정보에 따른 UI(User Interface), GUI(Graphic User Interface) 정보를 표시할 수 있다.
- [59] 상술한 구성요소 외에, 상술한 단말기는 인터페이스부 및 메모리를 더 포함할 수 있다.
- [60] 인터페이스부는 본 장치에 연결되는 다양한 종류의 외부 기기와의 통로 역할을 수행한다. 이러한 인터페이스부는 유/무선 헤드셋 포트(port), 외부 충전기 포트(port), 유/무선 데이터 포트(port), 메모리 카드(memory card) 포트, 식별 모듈(SIM)이 구비된 장치를 연결하는 포트(port), 오디오 I/O(Input/Output) 포트(port), 비디오 I/O(Input/Output) 포트(port), 이어폰 포트(port) 중 적어도 하나를 포함할 수 있다. 본 장치에서는, 상기 인터페이스부에 연결된 외부 기기와 관련된 적절한 제어를 수행할 수 있다.
- [61] 메모리는 본 장치의 다양한 기능을 지원하는 데이터와, 프로세서의 동작을 위한 프로그램을 저장할 수 있고, 입/출력되는 데이터들(예를 들어, 음악 파일, 정지영상, 동영상 등)을 저장할 있고, 본 장치에서 구동되는 다수의 응용 프로그램(application program 또는 애플리케이션(application)), 본 장치의 동작을 위한 데이터들, 명령어들을 저장할 수 있다. 이러한 응용 프로그램 중 적어도 일부는, 무선 통신을 통해 외부 서버로부터 다운로드 될 수 있다.
- [62] 이러한, 메모리는 플래시 메모리 타입(flash memory type), 하드디스크 타입(hard disk type), SSD 타입(Solid State Disk type), SDD 타입(Silicon Disk Drive type), 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD 또는 XD 메모리 등), 램(random access memory; RAM), SRAM(static random access memory), 롬(read-only memory;

ROM), EEPROM(electrically erasable programmable read-only memory),

PROM(programmable read-only memory), 자기 메모리, 자기 디스크 및 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다. 또한, 메모리는 본 장치와 분리되어 있으나, 유선 또는 무선으로 연결된 데이터베이스가 될 수도 있다.

[63] 한편, 상술한 단말기는 프로세서(240)를 포함한다. 프로세서는 본 장치 내의 구성요소들의 동작을 제어하기 위한 알고리즘 또는 알고리즘을 재현한 프로그램에 대한 데이터를 저장하는 메모리, 및 메모리에 저장된 데이터를 이용하여 전술한 동작을 수행하는 적어도 하나의 프로세서(미도시)로 구현될 수 있다. 이때, 메모리와 프로세서는 각각 별개의 칩으로 구현될 수 있다. 또는, 메모리와 프로세서는 단일 칩으로 구현될 수도 있다.

[64] 한편, 프로세서는 이하의 도면에서 설명되는 본 개시에 따른 다양한 실시 예들을 본 장치 상에서 구현하기 위하여, 위에서 살펴본 구성요소들을 중 어느 하나 또는 복수를 조합하여 제어할 수 있다.

[65] 한편, 도 1 내지 3에 도시된 구성 요소들의 성능에 대응하여 적어도 하나의 구성요소가 추가되거나 삭제될 수 있다. 또한, 구성 요소들의 상호 위치는 장치의 성능 또는 구조에 대응하여 변경될 수 있다는 것은 당해 기술 분야에서 통상의 지식을 가진 자에게 용이하게 이해될 것이다.

[66] 한편, 도 4와 같이, 서버 및 단말기 중 적어도 하나에 포함된 프로세서는 후술할 더빙 생성 시스템을 구현하기 위한 복수의 모듈을 포함할 수 있다. 구체적으로, 프로세서(300)는 음성 인식 모듈(310) 및 인공지능 모듈(320)을 포함할 수 있다. 후술하는 더빙 생성 방법은 상기 모듈들의 동작에 의해 구현되는 것으로 서술하나, 후술하는 각 단계의 동작의 수행이 반드시 상기 모듈들에 의해 수행될 필요는 없다.

[67] 이하에서는, 본 발명에서 서술되는 인공지능에 대하여 구체적으로 설명한다.

[68] 본 개시에 따른 인공지능과 관련된 기능은 상술한 서버 및 단말기에 탑재된 프로세서와 메모리를 통해 동작된다. 프로세서는 하나 또는 복수의 프로세서로 구성될 수 있다. 이때, 하나 또는 복수의 프로세서는 CPU, AP, DSP(Digital Signal Processor) 등과 같은 범용 프로세서, GPU, VPU(Vision Processing Unit)와 같은 그래픽 전용 프로세서 또는 NPU와 같은 인공지능 전용 프로세서일 수 있다. 하나 또는 복수의 프로세서는, 메모리에 저장된 기 정의된 동작 규칙 또는 인공지능 모델에 따라, 입력 데이터를 처리하도록 제어한다. 또는, 하나 또는 복수의 프로세서가 인공지능 전용 프로세서인 경우, 인공지능 전용 프로세서는, 특정 인공지능 모델의 처리에 특화된 하드웨어 구조로 설계될 수 있다.

[69] 기 정의된 동작 규칙 또는 인공지능 모델은 학습을 통해 만들어진 것을 특징으로 한다. 여기서, 학습을 통해 만들어진다는 것은, 기본 인공지능 모델이 학습 알고리즘에 의하여 다수의 학습 데이터들을 이용하여 학습됨으로써, 원하는 특성(또는, 목적)을 수행하도록 설정된 기 정의된 동작 규칙 또는 인공지능 모델이 만들어짐을 의미한다. 이러한 학습은 본 개시에 따른 인공지능이 수행되는 기기 자

체에서 이루어질 수도 있고, 별도의 서버 및/또는 시스템을 통해 이루어질 수도 있다. 학습 알고리즘의 예로는, 지도형 학습(supervised learning), 비지도형 학습(unsupervised learning), 준지도형 학습(semi-supervised learning) 또는 강화 학습(reinforcement learning)이 있으나, 전술한 예에 한정되지 않는다.

[70] 인공지능 모델은, 복수의 신경망 레이어들로 구성될 수 있다. 복수의 신경망 레이어들 각각은 복수의 가중치들(weight values)을 갖고 있으며, 이전(previous) 레이어의 연산 결과와 복수의 가중치들 간의 연산을 통해 신경망 연산을 수행한다. 복수의 신경망 레이어들이 갖고 있는 복수의 가중치들은 인공지능 모델의 학습 결과에 의해 최적화될 수 있다. 예를 들어, 학습 과정 동안 인공지능 모델에서 획득한 로스(loss) 값 또는 코스트(cost) 값이 감소 또는 최소화되도록 복수의 가중치들이 갱신될 수 있다. 인공 신경망은 심층 신경망(DNN: Deep Neural Network)를 포함할 수 있으며, 예를 들어, CNN (Convolutional Neural Network), DNN (Deep Neural Network), RNN (Recurrent Neural Network), RBM (Restricted Boltzmann Machine), DBN (Deep Belief Network), BRDNN(Bidirectional Recurrent Deep Neural Network) 또는 심층 Q-네트워크 (Deep Q-Networks) 등이 있으나, 전술한 예에 한정되지 않는다.

[71] 본 개시의 예시적인 실시예에 따르면, 프로세서는 인공지능을 구현할 수 있다. 인공지능이란 사람의 신경세포(biological neuron)를 모사하여 기계가 학습하도록 하는 인공신경망(Artificial Neural Network) 기반의 기계 학습법을 의미한다. 인공지능의 방법론에는 학습 방식에 따라 훈련데이터로서 입력데이터와 출력데이터가 같이 제공됨으로써 문제(입력데이터)의 해답(출력데이터)이 정해져 있는 지도학습(supervised learning), 및 출력데이터 없이 입력데이터만 제공되어 문제(입력데이터)의 해답(출력데이터)이 정해지지 않는 비지도학습(unsupervised learning), 및 현재의 상태(State)에서 어떤 행동(Action)을 취할 때마다 외부 환경에서 보상(Reward)이 주어지는데, 이러한 보상을 최대화하는 방향으로 학습을 진행하는 강화학습(reinforcement learning)으로 구분될 수 있다. 또한, 인공지능의 방법론은 학습 모델의 구조인 아키텍처에 따라 구분될 수도 있는데, 널리 이용되는 딥러닝 기술의 아키텍처는, 합성곱신경망(CNN; Convolutional Neural Network), 순환신경망(RNN; Recurrent Neural Network), 트랜스포머(Transformer), 생성적 대립 신경망(GAN; generative adversarial networks) 등으로 구분될 수 있다.

[72] 본 장치와 시스템은 인공지능 모델을 포함할 수 있다. 인공지능 모델은 하나의 인공지능 모델일 수 있고, 복수의 인공지능 모델로 구현될 수도 있다. 인공지능 모델은 뉴럴 네트워크(또는 인공 신경망)로 구성될 수 있으며, 기계학습과 인지과학에서 생물학의 신경을 모방한 통계학적 학습 알고리즘을 포함할 수 있다. 뉴럴 네트워크는 시냅스의 결합으로 네트워크를 형성한 인공 뉴런(노드)이 학습을 통해 시냅스의 결합 세기를 변화시켜, 문제 해결 능력을 가지는 모델 전반을 의미할 수 있다. 뉴럴 네트워크의 뉴런은 가중치 또는 바이어스의 조합을 포함할 수 있다. 뉴럴 네트워크는 하나 이상의 뉴런 또는 노드로 구성된 하나 이상

의 레이어(layer)를 포함할 수 있다. 예시적으로, 장치는 input layer, hidden layer, output layer를 포함할 수 있다. 장치를 구성하는 뉴럴 네트워크는 뉴런의 가중치를 학습을 통해 변화시킴으로써 임의의 입력(input)으로부터 예측하고자 하는 결과(output)를 추론할 수 있다.

[73] 프로세서는 뉴럴 네트워크를 생성하거나, 뉴럴 네트워크를 훈련(train, 또는 학습(learn)하거나, 수신되는 입력 데이터를 기초로 연산을 수행하고, 수행 결과를 기초로 정보 신호(information signal)를 생성하거나, 뉴럴 네트워크를 재훈련(retrain)할 수 있다. 뉴럴 네트워크의 모델들은 GoogleNet, AlexNet, VGG Network 등과 같은 CNN(Convolution Neural Network), R-CNN(Region with Convolution Neural Network), RPN(Region Proposal Network), RNN(Recurrent Neural Network), S-DNN(Stacking-based deep Neural Network), S-SDNN(State-Space Dynamic Neural Network), Deconvolution Network, DBN(Deep Belief Network), RBM(Restricted Boltzman Machine), Fully Convolutional Network, LSTM(Long Short-Term Memory) Network, Classification Network 등 다양한 종류의 모델들을 포함할 수 있으나 이에 제한되지는 않는다. 프로세서는 뉴럴 네트워크의 모델들에 따른 연산을 수행하기 위한 하나 이상의 프로세서를 포함할 수 있다. 예를 들어 뉴럴 네트워크는 심층 뉴럴 네트워크 (Deep Neural Network)를 포함할 수 있다.

[74] 뉴럴 네트워크는 CNN(Convolutional Neural Network), RNN(Recurrent Neural Network), 퍼셉트론(perceptron), 다층 퍼셉트론(multilayer perceptron), FF(Feed Forward), RBF(Radial Basis Network), DFF(Deep Feed Forward), LSTM(Long Short Term Memory), GRU(Gated Recurrent Unit), AE(Auto Encoder), VAE(Variational Auto Encoder), DAE(Denoising Auto Encoder), SAE(Sparse Auto Encoder), MC(Markov Chain), HN(Hopfield Network), BM(Boltzmann Machine), RBM(Restricted Boltzmann Machine), DBN(Deep Belief Network), DCN(Deep Convolutional Network), DN(Deconvolutional Network), DCIGN(Deep Convolutional Inverse Graphics Network), GAN(Generative Adversarial Network), LSM(Liquid State Machine), ELM(Extreme Learning Machine), ESN(Echo State Network), DRN(Deep Residual Network), DNC(Differentiable Neural Computer), NTM(Neural Turing Machine), CN(Capsule Network), KN(Kohonen Network) 및 AN(Attention Network)를 포함할 수 있으나 이에 한정되는 것이 아닌 임의의 뉴럴 네트워크를 포함할 수 있음은 통상의 기술자가 이해할 것이다.

[75] 본 개시의 예시적인 실시예에 따르면, 프로세서는 GoogleNet, AlexNet, VGG Network 등과 같은 CNN(Convolution Neural Network), R-CNN(Region with Convolution Neural Network), RPN(Region Proposal Network), RNN(Recurrent Neural Network), S-DNN(Stacking-based deep Neural Network), S-SDNN(State-Space Dynamic Neural Network), Deconvolution Network, DBN(Deep Belief Network), RBM(Restricted Boltzman Machine), Fully Convolutional Network, LSTM(Long Short-Term Memory) Network, Classification Network, Generative

Modeling, eXplainable AI, Continual AI, Representation Learning, AI for Material Design, 자연어 처리를 위한 BERT, SP-BERT, MRC/QA, Text Analysis, Dialog System, GPT-3, GPT-4, 비전 처리를 위한 Visual Analytics, Visual Understanding, Video Synthesis, ResNet 데이터 지능을 위한 Anomaly Detection, Prediction, Time-Series Forecasting, Optimization, Recommendation, Data Creation 등 다양한 인공지능 구조 및 알고리즘을 이용할 수 있으며, 이에 제한되지 않는다. 이하, 첨부된 도면을 참조하여 본 개시의 실시예를 상세하게 설명한다.

- [76] 이하에서는, 본 개시에 따른 더빙 생성 시스템 및 방법이 제공되는 자막 및 더빙 제공 서비스에 대하여 설명한다.
- [77] 도 5는 자막 및 더빙 제공 서비스를 나타내는 개념도이다.
- [78] 도 5를 참조하면, 음성 정보를 입력 받는 단계가 진행된다. 여기서 음성 정보를 입력 받는 단계는, 영상 정보를 입력 받는 단계일 수 있다. 본 개시에 따른 자막 더빙 생성 시스템은 영상 정보를 입력받고, 영상 정보에 포함된 음성 정보를 더빙 생성에 활용한다. 이와 달리, 상기 음성 정보를 입력 받는 단계는 영상 정보가 아니라 음성 정보 자체를 입력 받는 단계일 수 있다.
- [79] 본 명세서에서 음성 정보는 영상에 포함된 소리와 관련된 모든 데이터를 포함할 수 있다. 일 실시예로, 음성 정보는 영상 정보에 포함된 음성 신호 및 배경음을 포함할 수 있다.
- [80] 프로세서는 영상 정보를 입력 받아 영상 정보에 포함된 음성 정보를 가공한다.
- [81] 다음으로, 음성 정보에서 음성 신호 및 배경음 정보를 분리하는 단계가 진행된다.
- [82] 프로세서는 음성 정보에서 음성 신호와 배경음 정보를 분리하여 별도의 파일을 생성할 수 있다.
- [83] 음성 신호는 영상 정보에 포함된 화자의 음성이며, 더빙 생성의 대상이 되는 정보이다. 추후, 음성 정보에서 분리된 음성 신호는 더빙 중 적어도 하나로 생성된다.
- [84] 상기 음성 신호와 배경음 정보의 분리는 음성 정보로부터 음성 신호 및 배경음 정보를 분리하도록 훈련된 별도의 인공지능 모델을 통해 수행될 수 있다.
- [85] 다음으로, 분리된 음성 신호를 텍스트로 변환하는 단계가 진행된다.
- [86] 상기 음성 신호는 상기 음성 인식 모듈(310)을 통해 텍스트로 변환될 수 있다. 여기서, 상기 텍스트는 상기 음성 신호를 분리한 영상 제작의 기반이 된 언어로 이루어질 수 있다.
- [87] 본 명세서에서 영상 제작의 기반이 된 언어를 '원어'라하며, 더빙을 생성하고자 하는 언어를 "타겟 언어"라 한다.
- [88] 다음으로, 변환된, 원어로 구성된 텍스트를 타겟 언어로 번역하는 단계가 진행된다.
- [89] 상기 번역 결과는 타겟 언어로 구성된 텍스트 형태이며, 원본 영상의 자막으로 제공될 수 있다.

- [90] 상기 타겟 언어의 종류는 영상 제작자 또는 영상을 시청하는 사용자로부터 입력 받을 수 있다.
- [91] 일 실시 예에 있어서, 영상 제작자는 상기 타겟 언어를 미리 지정하여 영상 제작 단계에서부터 적어도 하나의 타겟 언어에 대한 자막을 생성할 수 있다.
- [92] 일 실시 예에 있어서, 영상 제작자는 자막을 생성하지 않고, 영상을 시청하는 시청자가 원하는 타겟 언어를 지정하여 자막을 생성할 수 있다. 특정, 영상 시청자에 의해 제작된 자막은 해당 타겟 언어의 자막을 요청하는 영상 시청자에게 다시 제공될 수 있다.
- [93] 상기 번역 결과는 이후, 더빙 생성에 활용될 수 있다. 이에 대하여는 후술한다.
- [94] 한편, 상기 번역은 원어로 이루어진 텍스트를 타겟 언어로 번역하도록 훈련된 별도의 인공지능 모델을 통해 수행될 수 있다. 여기서, 번역을 위한 인공지능 모델의 종류는 별도로 한정하지 않는다.
- [95] 한편, 프로세서는 상기 자막을 생성한 후 원본 영상에 합성한다. 이 때, 프로세서는 영상 정보 내에서 음성 신호가 발생하는 시간 구간을 고려하여, 자막을 삽입할 시간 구간을 설정할 수 있다. 구체적으로, 프로세서는 음절, 어절 또는 문장 단위로 영상 정보내에서 음성 신호가 발생된 시간 구간을 매칭시킬 수 있다. 이러한 방식으로 매칭된 시간 정보는 음성 신호를 텍스트로 변환한 후, 변환된 텍스트에도 매칭된다. 이후, 상기 변환된 텍스트를 번역한 결과에도 상기 시간 정보가 매칭될 수 있다. 프로세서는 상기 시간 정보를 기반으로 자막을 상기 영상 정보에 자막을 삽입할 수 있다.
- [96] 다음으로, 분리된 음성 신호, 변환된 텍스트 및 번역 결과 중 적어도 하나를 기반으로 타겟 언어로 이루어진 음성 신호(이하, 더빙 정보)를 생성하는 단계가 진행된다.
- [97] 본 개시에 따른 더빙 생성 방법은 상술한 음성 신호를 생성하는 단계에 구현될 수 있다.
- [98] 이하, 도 6 및 7을 참조하여, 타겟 언어로 이루어진 더빙 정보를 생성하는 방법에 대하여 구체적으로 설명한다.
- [99] 도 6을 참조하면, 인공지능 모듈(320)이 음성 신호를 입력 받는 단계가 진행된다(S210).
- [100] 여기서, 인공지능 모듈(320)이 입력받는 정보는 영상 정보에 포함된 음성 정보이거나, 상기 음성 정보에서 배경음 정보가 분리된 음성 신호일 수 있다.
- [101] 다음으로, 분리된 음성 신호를 텍스트로 변환하는 단계가 진행된다(S221).
- [102] 상기 음성 신호는 상기 음성 인식 모듈(310)을 통해 텍스트로 변환될 수 있다. 여기서, 상기 텍스트는 상기 원어로 이루어질 수 있다.
- [103] 다음으로, 변환된 텍스트를 타겟 언어로 번역하는 단계가 진행된다(S222).
- [104] 원어로 구성된 상기 변환된 텍스트는 번역을 통해, 타겟 언어로 구성된 텍스트로 변환된다.

- [105] 상기 번역은 기 공지된 인공 지능 모델을 통해 수행되거나 별도로 훈련된 인공 지능 모델을 통해 수행될 수 있다.
- [106] 상기 S221 및 S222 단계는 도 5에서 설명한 자막 및 더빙 제공 서비스에서 설명한 번역 결과를 활용함으로써 생략될 수 있다.
- [107] 상기 텍스트 번역과는 별개로, 다음으로, 인공지능 모델이 음성 신호를 국제음성기호(IPA)로 구성된 텍스트로 변환하는 단계가 진행된다(S223).
- [108] 상기 인공지능 모듈(320)은 음성 정보를 국제음성기호(International Phonetic Association, IPA)로 변환하도록 훈련된 인공지능 모델을 포함할 수 있다.
- [109] 일 실시 예에 있어서, 상기 인공지능 모델은 트랜스포머(transformer)일 수 있으나, 이에 한정되지는 않는다.
- [110] 일 실시 예에 있어서, 도 7을 참조하면, 상기 인공지능 모델은 "만 원만 빌려주 시면 안될까요"라는 음성 정보를 입력받아 아래와 같은 국제음성기호로 변환할 수 있다.
- [111] /manwʌnman biljəʊ dʒʊfɪmjən
andwɛlgajəʊ/
- [112] 구체적으로, 상기 인공지능 모델은 'src'가 한국어로 구성된 음성 신호이고, 'tgt'가 영문 IPA embedding이고, 'output'으로 영문 IPA embedding이 되도록 훈련될 수 있다.
- [113] 이후, 상기 인공지능 모델은 상기 번역 결과 및 국제음성기호로 구성된 텍스트를 기반으로 더빙 정보를 생성한다(S230).
- [114] 구체적으로, 인공지능 모델은 상기 변환된 텍스트 및 상기 음성 신호에 대한 번역 결과를 기반으로, 상기 음성 정보에 대한 더빙 정보를 생성할 수 있다.
- [115] 인공지능 모델은 상기 변환된 텍스트를 기반으로, 상기 번역 결과를 국제음성기호로 구성된 정보로 변환한다. 이를 위해, 인공지능 모델은 원어로 이루어진 음성 신호를 기반으로 생성된, 국제음성기호로 구성된 텍스트와 상기 원어로 이루어진 음성 신호에 대한 번역 결과를 입력 받아 국제음성기호로 구성되며, 타겟 언어로 번역된 텍스트를 생성하도록 훈련될 수 있다. 훈련 과정에서, 국제음성기호를 활용하는 경우, 원어 및 타겟 언어의 지역적, 사회적, 문화적, 언어적 차이가 반영된 더빙 정보를 생성할 수 있게 된다.
- [116] 또한, 상술한 바와 같이, 더빙 정보 생성을 위한 인공지능 모델 훈련 시 국제음성기호를 활용하는 경우, 자연스럽게 타겟 언어를 발음하는 더빙 정보를 생성할 수 있게 된다.
- [117] 다시, 도 5를 참조하면, 생성된 음성 신호(더빙 정보)와 분리된 배경음 정보를 합성하는 단계가 진행된다.
- [118] 프로세서는 상기 번역 결과를 활용하여, 상기 영상 정보에 자막을 생성하고, 상기 자막 생성 결과를 활용하여, 상기 생성된 음성 신호와 상기 분리된 배경음 정보를 합성할 수 있다.

- [119] 구체적으로, 프로세서는 상기 영상 정보에서 상기 자막이 출력되는 시간 정보를 탐색하고, 탐색된 시간 정보를 활용하여, 상기 생성된 음성 신호와 상기 분리된 배경음 정보를 합성할 수 있다.
- [120] 프로세서는 음절, 어절 또는 문장 단위로 영상 정보내에서 음성 신호가 발생된 시간 구간을 매칭시킬 수 있다. 이러한 방식으로 매칭된 시간 정보는 음성 신호를 텍스트로 변환한 후, 변환된 텍스트에도 매칭된다. 이후, 상기 변환된 텍스트를 번역한 결과에도 상기 시간 정보가 매칭될 수 있다. 프로세서는 상기 시간 정보를 기반으로 자막을 상기 영상 정보에 자막을 삽입할 수 있다. 또한, 프로세서는 상술한 시간 정보를 더빙 생성에도 활용할 수 있다.
- [121] 한편, 프로세서는 영상 정보에 포함된 음성 정보에서 상기 음성 신호 및 상기 배경음 정보를 분리한 결과를 활용하여, 상기 생성된 음성 신호와 상기 분리된 배경음 정보를 합성할 수 있다. 구체적으로, 프로세서는 상기 음성 정보에서 상기 음성 신호가 분리된 시간 정보를 활용하여, 상기 생성된 음성 신호와 상기 분리된 배경음 정보를 합성할 수 있다.
- [122] 프로세서는 음성 신호를 배경음 정보에서 분리할 때, 음성 신호가 분리된 시간 구간을 정의하는 정보를 배경음 정보에 매칭시킬 수 있다. 이러한 시간 구간 정보는 음절, 어절 또는 문장 단위로 정의되어, 배경음 정보에 매칭될 수 있다. 상기 시간 정보는 배경음 정보에 더빙 정보를 합성할 때 활용될 수 있다.
- [123] 상술한 바와 같이, 본 개시에 따른 더빙 생성 시스템에 따르면, 더빙을 생성한 후 음성 정보에서 분리된 배경음을 합성하기 때문에, 음원 손실 없이 더빙을 생성할 수 있게 된다.
- [124] 또한, 본 개시에 따르면, 수작업으로 더빙을 생성하는 종래의 방법 대비 빠른 속도로 더빙을 생성할 수 있다.
- [125] 또한, 본 개시에 따르면, 지역적, 사회적, 문화적, 언어적 차이가 반영된 더빙을 생성할 수 있게 된다.
- [126] 한편, 개시된 실시예들은 컴퓨터에 의해 실행 가능한 명령어를 저장하는 기록 매체의 형태로 구현될 수 있다. 명령어는 프로그램 코드의 형태로 저장될 수 있으며, 프로세서에 의해 실행되었을 때, 프로그램 모듈을 생성하여 개시된 실시예들의 동작을 수행할 수 있다. 기록매체는 컴퓨터로 읽을 수 있는 기록매체로 구현될 수 있다.
- [127] 컴퓨터가 읽을 수 있는 기록매체로는 컴퓨터에 의하여 해독될 수 있는 명령어가 저장된 모든 종류의 기록 매체를 포함한다. 예를 들어, ROM(Read Only Memory), RAM(Random Access Memory), 자기 테이프, 자기 디스크, 플래쉬 메모리, 광 데이터 저장장치 등이 있을 수 있다.
- [128] 이상에서와 같이 첨부된 도면을 참조하여 개시된 실시예들을 설명하였다. 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자는 본 개시의 기술적 사상이나 필수적인 특징을 변경하지 않고도, 개시된 실시예들과 다른 형태로 본 개시가

실시될 수 있음을 이해할 것이다. 개시된 실시예들은 예시적인 것이며, 한정적으로 해석되어서는 안 된다.

청구범위

- [청구항 1] 서버 및 단말기를 포함하는 더빙 생성 시스템에 있어서,
상기 서버는,
상기 단말기로부터 영상 정보를 수신하도록 이루어지는 통신부;
상기 영상 정보에 포함된 음성 신호를 텍스트로 변환하고, 변환된 텍스트를 기반으로 더빙 정보를 생성하는 인공지능 모델을 포함하는 프로세서를 포함하고,
상기 텍스트는 국제음성기호로 구성되는 것을 특징으로 하는 더빙 생성 시스템.
- [청구항 2] 제1항에 있어서,
상기 프로세서는,
상기 변환된 텍스트 및 상기 음성 신호에 대한 번역 결과를 기반으로, 상기 음성 정보에 대한 더빙 정보를 생성하는 것을 특징으로 하는 더빙 생성 시스템.
- [청구항 3] 제2항에 있어서,
상기 프로세서는,
상기 변환된 텍스트를 기반으로, 상기 프로세서는 상기 번역 결과를 국제음성기호로 구성된 정보로 변환하고, 변환된 정보를 기반으로 더빙 정보를 생성하는 것을 특징으로 하는 더빙 생성 시스템.
- [청구항 4] 제3항에 있어서,
상기 프로세서는,
상기 영상 정보에서 상기 음성 신호를 제거하고, 상기 더빙 정보를 결합하는 것을 특징으로 하는 더빙 생성 시스템.
- [청구항 5] 제4항에 있어서,
상기 인공지능 모델은 트랜스포머(transformer)인 것을 특징으로 하는 더빙 생성 시스템.
- [청구항 6] 서버 및 단말기를 포함하는 더빙 생성 시스템의 더빙 생성 방법에 있어서,
상기 서버가 상기 단말기로부터 영상 정보를 수신하는 단계;
상기 서버가 상기 영상 정보에 포함된 음성 신호를 텍스트로 변환하는 단계; 및
상기 서버가 상기 변환된 텍스트를 기반으로 더빙 정보를 생성하는 단계를 포함하고,
상기 텍스트는 국제음성기호로 구성되는 것을 특징으로 하는 더빙 생성 방법.
- [청구항 7] 제6항에 있어서,
상기 더빙 정보를 생성하는 단계는,

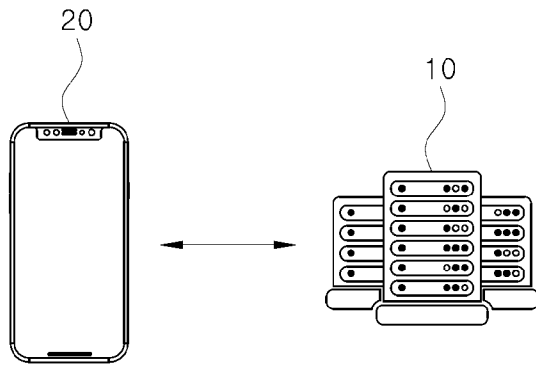
상기 변환된 텍스트 및 상기 음성 신호에 대한 번역 결과를 기반으로, 상기 음성 정보에 대한 더빙 정보를 생성하는 것을 특징으로 하는 더빙 생성 방법.

[청구항 8] 제7항에 있어서,
상기 더빙 정보를 생성하는 단계는,
상기 변환된 텍스트를 기반으로, 상기 프로세서는 상기 번역 결과를 국제 음성기호로 구성된 정보로 변환하고, 변환된 정보를 기반으로 더빙 정보를 생성하는 것을 특징으로 하는 더빙 생성 방법.

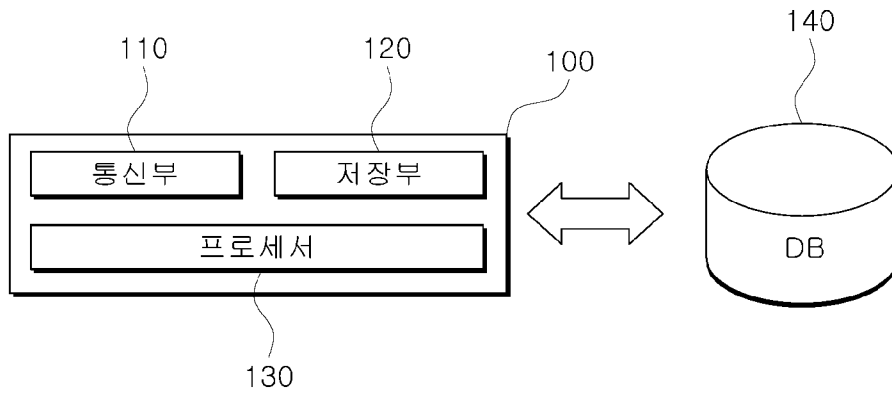
[청구항 9] 제8항에 있어서,
상기 영상 정보에서 상기 음성 신호를 제거하고, 상기 더빙 정보를 결합하는 단계를 더 포함하는 것을 특징으로 하는 더빙 생성 방법.

[청구항 10] 제9항에 있어서,
상기 인공지능 모델은 트랜스포머(transformer)인 것을 특징으로 하는 더빙 생성 방법.

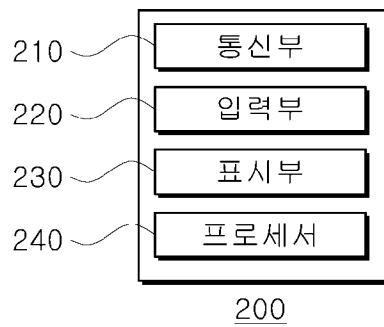
[도1]



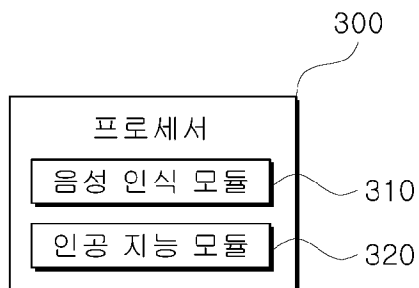
[도2]



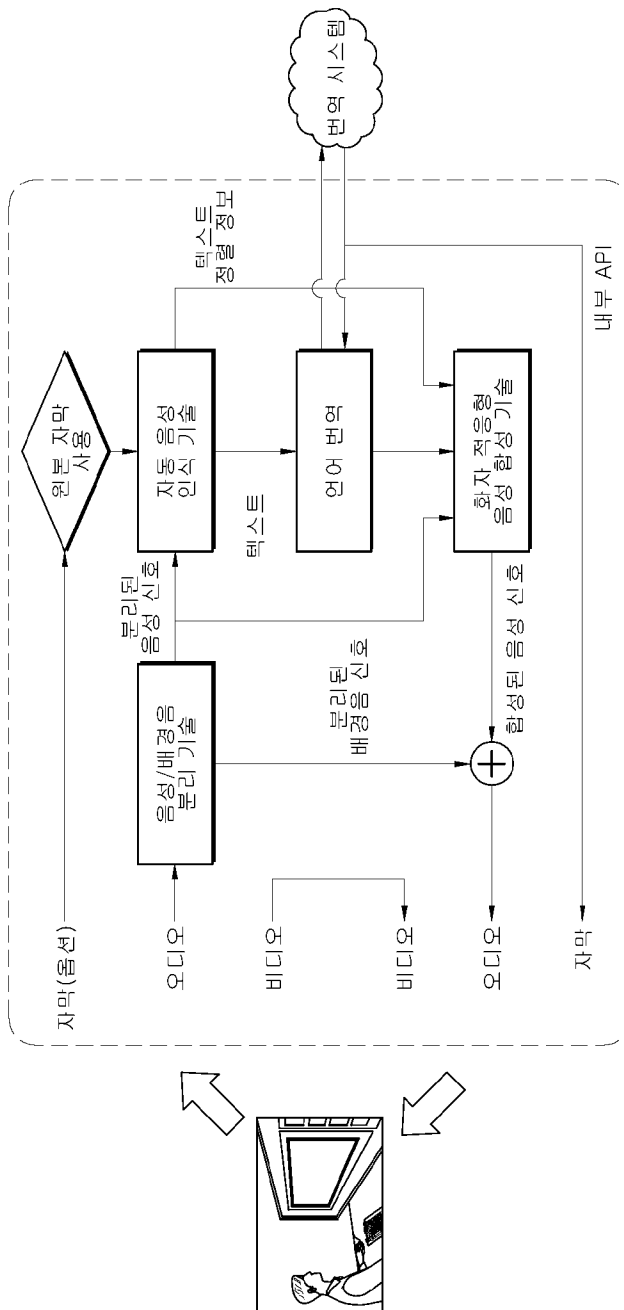
[도3]



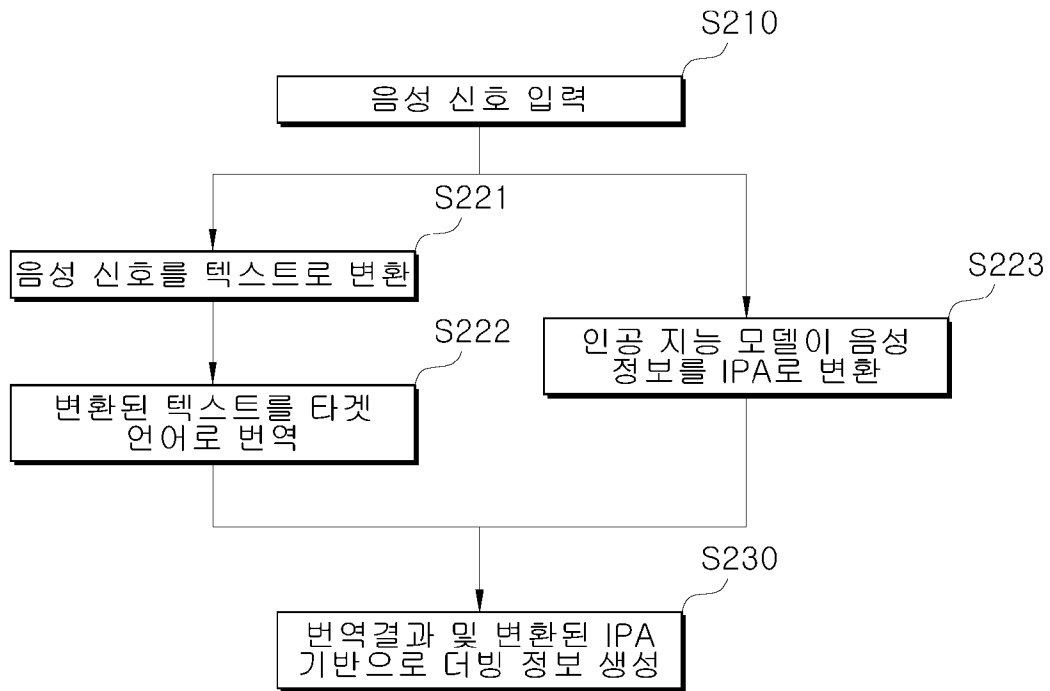
[도4]



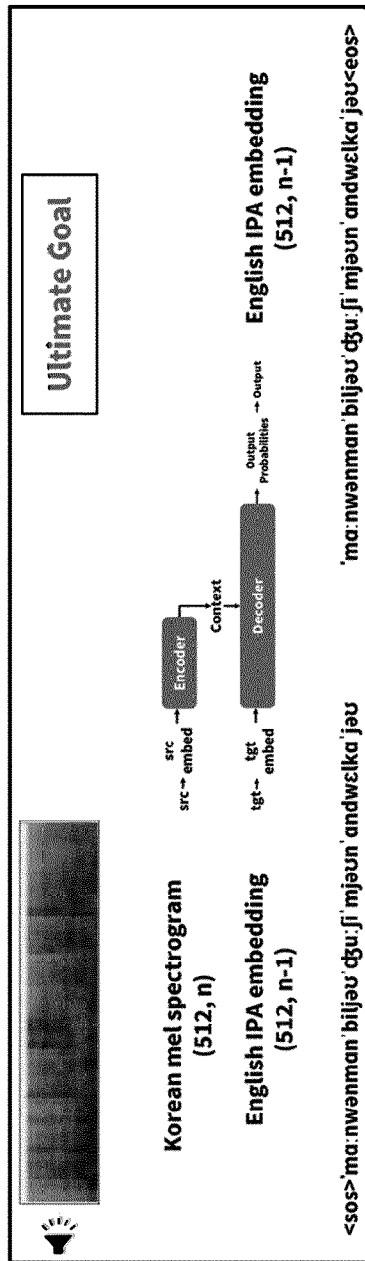
[도5]



[도6]



[도7]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2024/017832

A. CLASSIFICATION OF SUBJECT MATTER**G10L 15/26**(2006.01)i; **G10L 21/055**(2013.01)i; **G06F 40/58**(2020.01)i; **G10L 21/0208**(2013.01)i; **H04N 21/488**(2011.01)i; **H04N 21/43**(2011.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L 15/26(2006.01); G06F 40/40(2020.01); G10L 13/10(2013.01); G10L 15/187(2013.01); G10L 17/02(2013.01); G10L 25/30(2013.01)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above

Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & keywords: 영상(video), 음성(voice), 텍스트(text), 더빙(dubbing), 번역(translation), 국제음성기호(International Phonetic Alphabet, IPA)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KR 10-2546559 B1 (ELGEN CO., LTD.) 26 June 2023 (2023-06-26) See paragraphs [0029] and [0051]; and claim 1.	1-10
Y	KR 10-2022-0065483 A (SRUNIVERSE CO., LTD.) 20 May 2022 (2022-05-20) See paragraphs [0043] and [0069]; and claims 1 and 4.	1-10
A	KR 10-1735195 B1 (NAVER CORPORATION) 12 May 2017 (2017-05-12) See paragraphs [0030]-[0045]; and figures 2-3.	1-10
A	KR 10-2168529 B1 (SUPERTONE INC.) 22 October 2020 (2020-10-22) See paragraphs [0047]-[0081]; and figures 2-5.	1-10
A	JP 2826215 B2 (INTERNATIONAL BUSINESS MACHINES CORP.) 18 November 1998 (1998-11-18) See claims 1-2.	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

21 February 2025

Date of mailing of the international search report

21 February 2025

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon Building 4, 189 Cheongsaro, Seo-gu, Daejeon 35208

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2024/017832

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
KR	10-2546559	B1	26 June 2023	None			
KR	10-2022-0065483	A	20 May 2022	None			
KR	10-1735195	B1	12 May 2017	KR	10-2017-0017545	A	15 February 2017
KR	10-2168529	B1	22 October 2020	KR	10-2168529	B9	19 August 2021
JP	2826215	B2	18 November 1998	EP	0481107	A1	22 April 1992
				EP	0481107	B1	06 September 1995
				JP	04-313034	A	05 November 1992
				US	5230037	A	20 July 1993

A. 발명이 속하는 기술분류(국제특허분류(IPC)) G10L 15/26(2006.01)i; G10L 21/055(2013.01)i; G06F 40/58(2020.01)i; G10L 21/0208(2013.01)i; H04N 21/488(2011.01)i; H04N 21/43(2011.01)i		
B. 조사된 분야		
조사된 최소문헌(국제특허분류를 기재) G10L 15/26(2006.01); G06F 40/40(2020.01); G10L 13/10(2013.01); G10L 15/187(2013.01); G10L 17/02(2013.01); G10L 25/30(2013.01)		
조사된 기술분야에 속하는 최소문헌 이외의 문헌 한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC		
국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우)) eKOMPASS(특허청 내부 검색시스템) & 키워드: 영상(video), 음성(voice), 텍스트(text), 더빙(dubbing), 번역(translation), 국제음성기호(International Phonetic Alphabet, IPA)		
C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	KR 10-2546559 B1 (주식회사 엔젠) 2023.06.26 단락 [0029], [0051]; 및 청구항 1	1-10
Y	KR 10-2022-0065483 A (주식회사 에스알유니버스) 2022.05.20 단락 [0043], [0069]; 및 청구항 1, 4	1-10
A	KR 10-1735195 B1 (네이버 주식회사) 2017.05.12 단락 [0030]-[0045]; 및 도면 2-3	1-10
A	KR 10-2168529 B1 (주식회사 수퍼톤) 2020.10.22 단락 [0047]-[0081]; 및 도면 2-5	1-10
A	JP 2826215 B2 (INTERNATIONAL BUSINESS MACHINES CORP.) 1998.11.18 청구항 1-2	1-10
☐ 추가 문헌이 C(계속)에 기재되어 있습니다. ☒ 대응특허에 관한 별지를 참조하십시오.		
* 인용된 문헌의 특별 카테고리: “A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌 “D” 본 국제출원에서 출원인이 인용한 문헌 “E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌 “L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌 “O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌 “P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌 “T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌 “X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다. “Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다. “&” 동일한 대응특허문헌에 속하는 문헌		
국제조사의 실제 완료일 2025년02월21일 (21.02.2025)		국제조사보고서 발송일 2025년02월21일 (21.02.2025)
ISA/KR의 명칭 및 우편주소 대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578		심사관 고재용 전화번호 +82-42-481-8131

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2546559 B1	2023/06/26	없음	
KR 10-2022-0065483 A	2022/05/20	없음	
KR 10-1735195 B1	2017/05/12	KR 10-2017-0017545 A	2017/02/15
KR 10-2168529 B1	2020/10/22	KR 10-2168529 B9	2021/08/19
JP 2826215 B2	1998/11/18	EP 0481107 A1	1992/04/22
		EP 0481107 B1	1995/09/06
		JP 04-313034 A	1992/11/05
		US 5230037 A	1993/07/20