



2024 년 5 월 23 일 (23.05.2024)

WIPO | PCT

(51) 국제특허분류:

G10L 13/08 (2006.01)

G10L 25/15 (2013.01)

G10L 15/06 (2006.01)

G10L 15/26 (2006.01)

(21) 국제출원번호:

PCT/KR2023/018345

(22) 국제출원일:

2023 년 11 월 15 일 (15.11.2023)

(25) 출원언어:

한국어

(26) 공개언어:

한국어

(30) 우선권정보:

10-2022-0153726 2022 년 11 월 16 일 (16.11.2022) KR

10-2023-0002473 2023 년 1 월 6 일 (06.01.2023) KR

(71) 출원인: 삼성전자 주식회사 (SAMSUNG ELECTRONICS CO., LTD.) [KR/KR]; 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR). 한국과학기술원 (KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY) [KR/KR]; 34141 대전광역시 유성구 대학로 291, Daejeon (KR).

(72) 발명자: 서준식 (SUNG, Junesig); 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR). 정성엽 (JEONG, Seongyeop); 34141 대전광역시 유성구 대학로 291, Daejeon (KR). 최재식 (CHOI, Jaesik); 34141 대전광역시 유성구 대학로 291, Daejeon (KR).

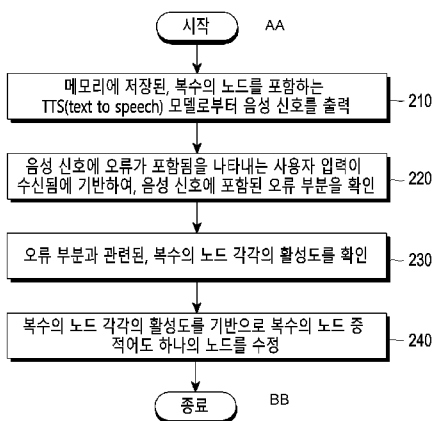
(74) 대리인: 이견주 등 (LEE, Keon-Joo et al.); 03079 서울특별시 종로구 대학로9길 16 미화빌딩, Seoul (KR).

(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.

(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).

(54) Title: ELECTRONIC DEVICE INCLUDING TEXT TO SPEECH (TTS) MODEL AND METHOD FOR CONTROLLING SAME

(54) 발명의 명칭: TTS(TEXT TO SPEECH) 모델을 포함하는 전자 장치 및 이의 제어 방법



(57) Abstract: An electronic device and a method for controlling same are disclosed. The electronic device comprises a memory and at least one processor operatively coupled to the memory, wherein the at least one processor is capable of: acquiring a speech signal on the basis of a text to speech (TTS) model stored in the memory and including a plurality of nodes; on the basis of determining that the speech signal corresponds to an input text and the speech signal includes an error, identifying an error portion of the speech signal that includes the error; determining an activity of each of the plurality of nodes associated with the error portion; and modifying at least one node of the plurality of nodes on the basis of the identified activity of the at least one node.

(57) 요약서: 전자 장치 및 이의 제어 방법이 개시되고, 전자 장치는, 메모리, 메모리와 작동적으로 연결된 적어도 하나의 프로세서를 포함하고, 적어도 하나의 프로세서는, 메모리에 저장된, 복수의 노드를 포함하는 TTS(text to speech) 모델에 기반하여 음성 신호를 획득하고, 음성 신호는 입력 텍스트에 대응되는 것이고, 음성 신호에 오류가 포함됨을 확인함에 기반하여, 오류가 포함된 음성 신호에 포함된 오류 부분을 확인하고, 오류 부분과 관련된, 복수의 노드 각각의 활성도를 확인하고, 적어도 하나의 노드의 식별된 활성도를 기반으로 복수의 노드 중 적어도 하나의 노드를 수정할 수 있다.

210 ... Output speech signal on basis of text to speech (TTS) model stored in memory and including plurality of nodes

220 ... On basis of reception of user input indicating that speech signal includes error, identify error portion included in speech signal

230 ... Determine activity of each of plurality of nodes associated with error portion

240 ... Modify at least one node of plurality of nodes on basis of activity of each of plurality of nodes

AA ... Start

BB ... End



공개:

— 국제조사보고서와 함께 (조약 제21조(3))

명세서

발명의 명칭: TTS(TEXT TO SPEECH) 모 델을 포함하는 전자 장치 및 이의 제어 방법

기술분야

- [1] 본 개시는, TTS(text to speech) 모델을 포함하는 전자 장치 및 이의 제어 방법에 관한 것이다.

배경기술

- [2] 전자 장치, 예를 들어, 스마트폰과 같은 휴대용 전자 장치를 통해 제공되는 다양한 서비스 및 부가 기능들이 점차 증가하고 있다. 이러한 전자 장치의 효용 가치를 높이고, 다양한 사용자들의 욕구를 만족시키기 위해서 통신 서비스 제공자 또는 전자 장치 제조사들은 다양한 기능들을 제공하고 다른 업체와의 차별화를 위해 전자 장치를 경쟁적으로 개발하고 있다. 이에 따라, 전자 장치를 통해서 제공되는 다양한 기능들도 점점 고도화되고 있다.
- [3] 근래에는 사용자 음성 입력에 대한 응답을 제공하는 인공지능 에이전트(예로, 빅스비™, 어시스턴트™, 알렉사™ 등)를 이용한 다양한 서비스들이 제공되고 있다. 특히, TTS(text to speech) 기술을 통해 입력 텍스트와 음향 특징의 관계를 통합적으로 모델링하여 자연스러운 합성음을 출력할 수 있다.

발명의 상세한 설명

과제 해결 수단

- [4] 일 실시 예에 따라, 전자 장치는, 적어도 하나의 메모리, 상기 적어도 하나의 메모리와 작동적으로 연결된 적어도 하나의 프로세서를 포함할 수 있다.
- [5] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 적어도 하나의 메모리에 저장된, 복수의 노드를 포함하는 TTS(text to speech) 모델에 기반하여 음성 신호를 획득하는 명령어들을 저장할 수 있고, 상기 음성 신호는 입력 텍스트에 대응되는 것이다.
- [6] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 음성 신호에 오류가 포함됨을 확인함에 기반하여, 상기 오류가 포함된 상기 음성 신호의 오류 부분을 확인하는 명령어들을 저장할 수 있다.
- [7] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 오류 부분과 관련된, 상기 복수의 노드 각각의 활성도를 확인하는 명령어들을 저장할 수 있다.
- [8] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 적어도 하나의 노드의 상기 식별된 활성도를 기반으로 상기 적어도 하나의 노드를 수정하는 명령어들을 저장할 수 있다.

- [9] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 적어도 하나의 노드와 관련된 가중치를 줄이는 명령어들을 저장할 수 있다.
- [10] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 적어도 하나의 노드를 상기 오류 부분에 대응되는 텍스트와 관련하여 기저장된 적어도 하나의 노드로 교체하는 명령어들을 저장할 수 있다.
- [11] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 기저장된 적어도 하나의 노드는 상기 적어도 하나의 메모리에 저장되고, 상기 오류 부분에 대응되는 텍스트에 대응될 수 있다.
- [12] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 음성 신호가 설정된 길이 이상의 길이를 갖는 적어도 하나의 음소를 포함함을 확인함에 기반하여, 상기 적어도 하나의 음소에 대응되는 상기 음성 신호의 부분을 상기 오류 부분인 것으로 확인하는 명령어들을 저장할 수 있다.
- [13] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 음성 신호가 이상 파형을 가지는 파형 부분을 포함함을 확인함에 기반하여, 상기 파형 부분을 상기 오류 부분인 것으로 확인하는 명령어들을 저장할 수 있다.
- [14] 일 실시 예에 따라, 자동 음성 인식(automatic speech recognition, ASR) 모델은 상기 적어도 하나의 메모리에 저장될 수 있다.
- [15] 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 음성 신호에 상기 ASR 모델을 적용한 결과인 텍스트를 획득하는 명령어들을 저장할 수 있다.
- [16] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 텍스트에 입력 텍스트와 다른 부분을 포함함을 확인함에 기반하여, 상기 입력 텍스트와 다른 부분을 상기 오류 부분인 것으로 확인하는 명령어들을 저장할 수 있다.
- [17] 일 실시 예에 따라, 전자 장치는 디스플레이를 더 포함할 수 있다.
- [18] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 입력 텍스트를 상기 디스플레이에 표시하는 명령어들을 저장할 수 있다.
- [19] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 디스플레이를 통해 수신된 사용자 입력에 기반하여, 상기 오류 부분을 확인하고, 상기 사용자 입력은 상기 입력 텍스트의 부분의 선택을 포함하는 명령어들을 저장할 수 있다.

- [20] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 입력 텍스트의 문장 구조를 확인하는 명령어들을 저장할 수 있다.
- [21] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 문장 구조를 기반으로 적어도 하나의 문자열을 획득하는 명령어들을 저장할 수 있다.
- [22] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 적어도 하나의 문자열을 상기 TTS 모델에 입력함의 결과로 문자열 음성 신호를 획득하는 명령어들을 저장할 수 있다.
- [23] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 문자열 음성 신호에 기반하여 상기 오류 부분의 수정 여부를 확인하는 명령어들을 저장할 수 있다.
- [24] 일 실시 예에 따라, 상기 적어도 하나의 문자열은, 상기 오류 부분에 대응되는 입력 텍스트의 부분 또는 후의 텍스트를 변경하여 획득된 것일 수 있다.
- [25] 일 실시 예에 따라, 전자 장치는, 통신 모듈을 더 포함할 수 있다.
- [26] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 오류 부분 및 상기 적어도 하나의 노드의 수정과 관련된 정보를 서버로 전송하도록 상기 통신 모듈을 제어하는 명령어들을 저장할 수 있다.
- [27] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 통신 모듈을 통해, 상기 서버로부터 수정된 TTS 모델을 수신하는 명령어들을 저장할 수 있다.
- [28] 일 실시 예에 따라, 상기 적어도 하나의 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가, 상기 수정된 TTS 모델을 기반으로 상기 적어도 하나의 메모리에 저장된 TTS 모델을 업데이트하는 명령어들을 저장할 수 있다.
- [29] 일 실시 예에 따라, 전자 장치를 제어하기 위한 방법은, 상기 전자 장치의 적어도 하나의 메모리에 저장된, 복수의 노드를 포함하는 TTS(text to speech) 모델에 기반하여 음성 신호를 획득하는 동작을 포함하고, 상기 음성 신호는 입력 텍스트에 대응되는 것일 수 있다.
- [30] 일 실시 예에 따라, 전자 장치를 제어하기 위한 방법은, 상기 음성 신호에 오류가 포함됨을 확인함에 기반하여, 상기 오류가 포함된 상기 음성 신호의 오류 부분을 확인하는 동작을 포함할 수 있다.
- [31] 일 실시 예에 따라, 전자 장치를 제어하기 위한 방법은, 상기 오류 부분과 관련된, 상기 복수의 노드 각각의 활성도를 확인하는 동작을 포함할 수 있다.
- [32] 일 실시 예에 따라, 전자 장치를 제어하기 위한 방법은, 적어도 하나의 노드의 상기 식별된 활성도를 기반으로 상기 적어도 하나의 노드를 수정하는 동작을 포함할 수 있다.

- [33] 일 실시 예에 따라, 상기 적어도 하나의 노드를 수정하는 동작은, 상기 적어도 하나의 노드와 관련된 가중치를 줄이는 동작을 포함할 수 있다.
- [34] 일 실시 예에 따라, 상기 적어도 하나의 노드를 수정하는 동작은, 상기 적어도 하나의 노드를 상기 오류 부분에 대응되는 텍스트에 대응되는 기저장된 적어도 하나의 노드로 교체하는 동작을 포함할 수 있다.
- [35] 일 실시 예에 따라, 상기 오류 부분을 확인하는 동작은, 상기 음성 신호가 설정된 길이 이상의 길이를 갖는 적어도 하나의 음소를 포함함을 확인함에 기반하여, 상기 적어도 하나의 음소에 대응되는 상기 음성 신호의 부분을 상기 오류 부분인 것으로 확인하는 동작을 포함할 수 있다.
- [36] 일 실시 예에 따라, 상기 오류 부분을 확인하는 동작은, 상기 음성 신호가 이상 파형을 가지는 파형 부분을 포함함을 확인함에 기반하여, 상기 파형 부분을 상기 오류 부분인 것으로 확인하는 동작을 포함할 수 있다.
- [37] 일 실시 예에 따라, 자동 음성 인식(automatic speech recognition, ASR) 모델은 상기 적어도 하나의 메모리에 저장될 수 있다.
- [38] 일 실시 예에 따라, 상기 오류 부분을 확인하는 동작은, 상기 음성 신호에 상기 ASR 모델을 적용한 결과인 텍스트를 획득하는 동작을 포함할 수 있다.
- [39] 일 실시 예에 따라, 상기 오류 부분을 확인하는 동작은, 상기 텍스트에 입력 텍스트와 다른 부분을 포함함에 기반하여, 상기 입력 텍스트와 다른 부분을 상기 오류 부분인 것으로 확인하는 동작을 포함할 수 있다.
- [40] 일 실시 예에 따라, 상기 오류 부분을 확인하는 동작은, 상기 입력 텍스트를 상기 전자 장치의 디스플레이에 표시하는 동작을 포함할 수 있다.
- [41] 일 실시 예에 따라, 상기 오류 부분을 확인하는 동작은, 상기 디스플레이를 통해 수신된 사용자 입력에 기반하여, 상기 오류 부분을 확인하고, 상기 사용자 입력은 상기 입력 텍스트의 부분의 선택을 포함할 수 있다.
- [42] 일 실시 예에 따라, 방법은, 상기 입력 텍스트의 문장 구조를 확인하는 동작을 더 포함할 수 있다.
- [43] 일 실시 예에 따라, 방법은, 상기 문장 구조를 기반으로 적어도 하나의 문자열을 획득하는 동작을 더 포함할 수 있다.
- [44] 일 실시 예에 따라, 방법은, 상기 적어도 하나의 문자열을 상기 TTS 모델에 입력함의 결과로 문자열 음성 신호를 획득하는 동작을 더 포함할 수 있다.
- [45] 일 실시 예에 따라, 방법은, 상기 문자열 음성 신호에 기반하여 상기 오류 부분의 수정 여부를 확인하는 동작을 더 포함할 수 있다.
- [46] 일 실시 예에 따라, 상기 적어도 하나의 문자열은, 상기 오류 부분에 대응되는 입력 텍스트의 부분 또는 후의 텍스트를 변경하여 획득된 것일 수 있다.
- [47] 일 실시 예에 따라, 방법은, 상기 오류 부분 및 상기 적어도 하나의 노드의 수정과 관련된 정보를 서버로 전송하는 동작을 더 포함할 수 있다.
- [48] 일 실시 예에 따라, 방법은, 상기 서버로부터 수정된 TTS 모델을 수신하는 동작을 더 포함할 수 있다.

- [49] 일 실시 예에 따라, 방법은, 상기 수정된 TTS 모델을 기반으로 상기 적어도 하나의 메모리에 저장된 TTS 모델을 업데이트 하는 동작을 더 포함할 수 있다.
- [50] 일 실시 예에 따라, 하나 이상의 프로그램을 저장하는, 비일시적 컴퓨터 판독 가능 기록 매체에 있어서, 상기 하나 이상의 프로그램은, 전자 장치로 하여금, 상기 전자 장치의 적어도 하나의 메모리에 저장된, 복수의 노드를 포함하는 TTS(text to speech) 모델에 기반하여 음성 신호를 획득하는 명령어들을 포함할 수 있고, 상기 음성 신호는 입력 텍스트에 대응되는 것이다.
- [51] 일 실시 예에 따라, 상기 하나 이상의 프로그램은, 전자 장치로 하여금, 상기 음성 신호에 오류가 포함됨을 확인함에 기반하여, 상기 오류가 포함된 상기 음성 신호의 오류 부분을 확인하는 명령어들을 포함할 수 있다.
- [52] 일 실시 예에 따라, 상기 하나 이상의 프로그램은, 전자 장치로 하여금, 상기 오류 부분과 관련된, 상기 복수의 노드 각각의 활성도를 확인하는 명령어들을 포함할 수 있다.
- [53] 일 실시 예에 따라, 상기 하나 이상의 프로그램은, 전자 장치로 하여금, 적어도 하나의 노드의 상기 식별된 활성도를 기반으로 상기 복수의 노드 중 상기 적어도 하나의 노드와 관련된 가중치를 줄이는 명령어들을 포함할 수 있다.

도면의 간단한 설명

- [54] 본 개시의 특정 실시예의 상기 및 다른 측면, 특징 및 장점은 첨부 도면과 함께 취해진 다음의 설명으로부터 더욱 명백해질 것이다.
- [55] 도 1은 일 실시 예에 따른 네트워크 환경 내의 전자 장치의 블록도이다.
- [56] 도 2는 일 실시 예에 따른 전자 장치의 TTS 모델의 수정 동작을 설명하기 위한 흐름도이다.
- [57] 도 3은 일 실시 예에 따른 전자 장치의 TTS 모델의 수정 동작을 설명하기 위한 도면이다.
- [58] 도 4a는 일 실시 예에 따른 전자 장치의 음성 신호의 오류 부분을 확인하는 동작을 설명하기 위한 도면이다.
- [59] 도 4b는 일 실시 예에 따른 전자 장치의 TTS 모델의 적어도 하나의 노드 수정 동작을 설명하기 위한 도면이다.
- [60] 도 4c는 일 실시 예에 따른 전자 장치의 수정된 TTS 모델을 통해 오류 부분이 수정된 음성 신호를 설명하기 위한 도면이다.
- [61] 도 5는 일 실시 예에 따른 전자 장치의 음성 신호의 오류 부분을 사용자 입력에 의해 확인하는 동작을 설명하기 위한 도면이다.
- [62] 도 6은 일 실시 예에 따른 전자 장치의 사용자 입력에 의한 오류 부분을 수정하는 동작을 설명하기 위한 도면이다.
- [63] 도 7은 일 실시 예에 따른 전자 장치의 여러 사용자로부터 수집된 오류 정보를 기반으로 TTS 모델을 업데이트하는 동작을 설명하기 위한 도면이다.

- [64] 도 8은 일 실시 예에 따른 전자 장치에 표시되는 오류 리포트를 설명하기 위한 도면이다. 그리고,
- [65] 도 9는 일 실시 예에 따른 TTS 모델의 관리자가 볼 수 있는 유저 인터페이스를 설명하기 위한 도면이다.

발명의 실시를 위한 형태

- [66] 도 1은, 하나 이상의 실시예들에 따른, 네트워크 환경(100) 내의 전자 장치(101)의 블록도이다. 도 1을 참조하면, 네트워크 환경(100)에서 전자 장치(101)는 제 1 네트워크(198)(예: 근거리 무선 통신 네트워크)를 통하여 전자 장치(102)와 통신하거나, 또는 제 2 네트워크(199)(예: 원거리 무선 통신 네트워크)를 통하여 전자 장치(104) 또는 서버(108) 중 적어도 하나와 통신할 수 있다. 일 실시예에 따르면, 전자 장치(101)는 서버(108)를 통하여 전자 장치(104)와 통신할 수 있다. 일 실시예에 따르면, 전자 장치(101)는 프로세서(120), 메모리(130), 입력 모듈(150), 음향 출력 모듈(155), 디스플레이 모듈(160), 오디오 모듈(170), 센서 모듈(176), 인터페이스(177), 연결 단자(178), 햅틱 모듈(179), 카메라 모듈(180), 전력 관리 모듈(188), 배터리(189), 통신 모듈(190), 가입자 식별 모듈(196), 또는 안테나 모듈(197)을 포함할 수 있다. 하나 이상의 실시예에서는, 전자 장치(101)에는, 이 구성요소들 중 적어도 하나(예: 연결 단자(178))가 생략되거나, 하나 이상의 다른 구성요소가 추가될 수 있다. 하나 이상의 실시예에서는, 이 구성요소들 중 일부들(예: 센서 모듈(176), 카메라 모듈(180), 또는 안테나 모듈(197))은 하나의 구성요소(예: 디스플레이 모듈(160))로 통합될 수 있다.
- [67] 프로세서(120)는, 예를 들면, 소프트웨어(예: 프로그램(140))를 실행하여 프로세서(120)에 연결된 전자 장치(101)의 적어도 하나의 다른 구성요소(예: 하드웨어 또는 소프트웨어 구성요소)를 제어할 수 있고, 다양한 데이터 처리 또는 연산을 수행할 수 있다. 일 실시예에 따르면, 데이터 처리 또는 연산의 적어도 일부로서, 프로세서(120)는 다른 구성요소(예: 센서 모듈(176) 또는 통신 모듈(190))로부터 수신된 명령 또는 데이터를 휘발성 메모리(132)에 저장하고, 휘발성 메모리(132)에 저장된 명령 또는 데이터를 처리하고, 결과 데이터를 비휘발성 메모리(134)에 저장할 수 있다. 일 실시예에 따르면, 프로세서(120)는 메인 프로세서(121)(예: 중앙 처리 장치 또는 어플리케이션 프로세서) 또는 이와는 독립적으로 또는 함께 운영 가능한 보조 프로세서(123)(예: 그래픽 처리 장치, 신경망 처리 장치(NPU: neural processing unit), 이미지 시그널 프로세서, 센서 허브 프로세서, 또는 커뮤니케이션 프로세서)를 포함할 수 있다. 예를 들어, 전자 장치(101)가 메인 프로세서(121) 및 보조 프로세서(123)를 포함하는 경우, 보조 프로세서(123)는 메인 프로세서(121)보다 저전력을 사용하거나, 지정된 기능에 특화되도록 설정될 수 있다. 보조 프로세서(123)는 메인 프로세서(121)와 별개로, 또는 그 일부로서 구현될 수 있다.

- [68] 보조 프로세서(123)는, 예를 들면, 메인 프로세서(121)가 인액티브(예: 슬립) 상태에 있는 동안 메인 프로세서(121)를 대신하여, 또는 메인 프로세서(121)가 액티브(예: 어플리케이션 실행) 상태에 있는 동안 메인 프로세서(121)와 함께, 전자 장치(101)의 구성요소들 중 적어도 하나의 구성요소(예: 디스플레이 모듈(160), 센서 모듈(176), 또는 통신 모듈(190))와 관련된 기능 또는 상태들의 적어도 일부를 제어할 수 있다. 일실시예에 따르면, 보조 프로세서(123)(예: 이미지 시그널 프로세서 또는 커뮤니케이션 프로세서)는 기능적으로 관련 있는 다른 구성요소(예: 카메라 모듈(180) 또는 통신 모듈(190))의 일부로서 구현될 수 있다. 일실시예에 따르면, 보조 프로세서(123)(예: 신경망 처리 장치)는 인공지능 모델의 처리에 특화된 하드웨어 구조를 포함할 수 있다. 인공지능 모델은 기계 학습을 통해 생성될 수 있다. 이러한 학습은, 예를 들어, 인공지능 모델이 수행되는 전자 장치(101) 자체에서 수행될 수 있고, 별도의 서버(예: 서버(108))를 통해 수행될 수도 있다. 학습 알고리즘은, 예를 들어, 지도형 학습(supervised learning), 비지도형 학습(unsupervised learning), 준지도형 학습(semi-supervised learning) 또는 강화 학습(reinforcement learning)을 포함할 수 있으나, 전술한 예에 한정되지 않는다. 인공지능 모델은, 복수의 인공 신경망 레이어들을 포함할 수 있다. 인공 신경망은 심층 신경망(DNN: deep neural network), CNN(convolutional neural network), RNN(recurrent neural network), RBM(restricted boltzmann machine), DBN(deep belief network), BRDNN(bidirectional recurrent deep neural network), 심층 Q-네트워크(deep Q-networks) 또는 상기 중 둘 이상의 조합 중 하나일 수 있으나, 전술한 예에 한정되지 않는다. 인공지능 모델은 하드웨어 구조 이외에, 추가적으로 또는 대체적으로, 소프트웨어 구조를 포함할 수 있다.
- [69] 메모리(130)는, 전자 장치(101)의 적어도 하나의 구성요소(예: 프로세서(120) 또는 센서 모듈(176))에 의해 사용되는 다양한 데이터를 저장할 수 있다. 데이터는, 예를 들어, 소프트웨어(예: 프로그램(140)) 및, 이와 관련된 명령에 대한 입력 데이터 또는 출력 데이터를 포함할 수 있다. 메모리(130)는, 휘발성 메모리(132) 또는 비휘발성 메모리(134)를 포함할 수 있다.
- [70] 프로그램(140)은 메모리(130)에 소프트웨어로서 저장될 수 있으며, 예를 들면, 운영 체제(142), 미들웨어(144) 또는 어플리케이션(146)을 포함할 수 있다.
- [71] 입력 모듈(150)은, 전자 장치(101)의 구성요소(예: 프로세서(120))에 사용될 명령 또는 데이터를 전자 장치(101)의 외부(예: 사용자)로부터 수신할 수 있다. 입력 모듈(150)은, 예를 들면, 마이크, 마우스, 키보드, 키(예: 버튼), 또는 디지털 펜(예: 스타일러스 펜)을 포함할 수 있다.
- [72] 음향 출력 모듈(155)은 음향 신호를 전자 장치(101)의 외부로 출력할 수 있다. 음향 출력 모듈(155)은, 예를 들면, 스피커 또는 리시버를 포함할 수 있다. 스피커는 멀티미디어 재생 또는 녹음 재생과 같이 일반적인 용도로 사용될 수 있다. 리시버는 착신 전화를 수신하기 위해 사용될 수 있다. 일실시예에 따르면, 리시버는 스피커와 별개로, 또는 그 일부로서 구현될 수 있다.

- [73] 디스플레이 모듈(160)은 전자 장치(101)의 외부(예: 사용자)로 정보를 시각적으로 제공할 수 있다. 디스플레이 모듈(160)은, 예를 들면, 디스플레이, 홀로그램 장치, 또는 프로젝터 및 해당 장치를 제어하기 위한 제어 회로를 포함할 수 있다. 일 실시예에 따르면, 디스플레이 모듈(160)은 터치를 감지하도록 설정된 터치 센서, 또는 상기 터치에 의해 발생하는 힘의 세기를 측정하도록 설정된 압력 센서를 포함할 수 있다.
- [74] 오디오 모듈(170)은 소리를 전기 신호로 변환시키거나, 반대로 전기 신호를 소리로 변환시킬 수 있다. 일 실시예에 따르면, 오디오 모듈(170)은, 입력 모듈(150)을 통해 소리를 획득하거나, 음향 출력 모듈(155), 또는 전자 장치(101)와 직접 또는 무선으로 연결된 외부 전자 장치(예: 전자 장치(102))(예: 스피커 또는 헤드폰)를 통해 소리를 출력할 수 있다.
- [75] 센서 모듈(176)은 전자 장치(101)의 작동 상태(예: 전력 또는 온도), 또는 외부의 환경 상태(예: 사용자 상태)를 감지하고, 감지된 상태에 대응하는 전기 신호 또는 데이터 값을 생성할 수 있다. 일 실시예에 따르면, 센서 모듈(176)은, 예를 들면, 제스처 센서, 자이로 센서, 기압 센서, 마그네틱 센서, 가속도 센서, 그립 센서, 근접 센서, 컬러 센서, IR(infrared) 센서, 생체 센서, 온도 센서, 습도 센서, 또는 조도 센서를 포함할 수 있다.
- [76] 인터페이스(177)는 전자 장치(101)가 외부 전자 장치(예: 전자 장치(102))와 직접 또는 무선으로 연결되기 위해 사용될 수 있는 하나 이상의 지정된 프로토콜들을 지원할 수 있다. 일 실시예에 따르면, 인터페이스(177)는, 예를 들면, HDMI(high definition multimedia interface), USB(universal serial bus) 인터페이스, SD카드 인터페이스, 또는 오디오 인터페이스를 포함할 수 있다.
- [77] 연결 단자(178)는, 그를 통해서 전자 장치(101)가 외부 전자 장치(예: 전자 장치(102))와 물리적으로 연결될 수 있는 커넥터를 포함할 수 있다. 일 실시예에 따르면, 연결 단자(178)는, 예를 들면, HDMI 커넥터, USB 커넥터, SD 카드 커넥터, 또는 오디오 커넥터(예: 헤드폰 커넥터)를 포함할 수 있다.
- [78] 햅틱 모듈(179)은 전기적 신호를 사용자가 촉각 또는 운동 감각을 통해서 인지할 수 있는 기계적인 자극(예: 진동 또는 움직임) 또는 전기적인 자극으로 변환할 수 있다. 일 실시예에 따르면, 햅틱 모듈(179)은, 예를 들면, 모터, 압전 소자, 또는 전기 자극 장치를 포함할 수 있다.
- [79] 카메라 모듈(180)은 정지 영상 및 동영상을 촬영할 수 있다. 일 실시예에 따르면, 카메라 모듈(180)은 하나 이상의 렌즈들, 이미지 센서들, 이미지 시그널 프로세서들, 또는 플래시들을 포함할 수 있다.
- [80] 전력 관리 모듈(188)은 전자 장치(101)에 공급되는 전력을 관리할 수 있다. 일 실시예에 따르면, 전력 관리 모듈(188)은, 예를 들면, PMIC(power management integrated circuit)의 적어도 일부로서 구현될 수 있다.

- [81] 배터리(189)는 전자 장치(101)의 적어도 하나의 구성요소에 전력을 공급할 수 있다. 일실시예에 따르면, 배터리(189)는, 예를 들면, 재충전 불가능한 1차 전지, 재충전 가능한 2차 전지 또는 연료 전지를 포함할 수 있다.
- [82] 통신 모듈(190)은 전자 장치(101)와 외부 전자 장치(예: 전자 장치(102), 전자 장치(104), 또는 서버(108)) 간의 직접(예: 유선) 통신 채널 또는 무선 통신 채널의 수립, 및 수립된 통신 채널을 통한 통신 수행을 지원할 수 있다. 통신 모듈(190)은 프로세서(120)(예: 어플리케이션 프로세서)와 독립적으로 운영되고, 직접(예: 유선) 통신 또는 무선 통신을 지원하는 하나 이상의 커뮤니케이션 프로세서를 포함할 수 있다. 일실시예에 따르면, 통신 모듈(190)은 무선 통신 모듈(192)(예: 셀룰러 통신 모듈, 근거리 무선 통신 모듈, 또는 GNSS(global navigation satellite system) 통신 모듈) 또는 유선 통신 모듈(194)(예: LAN(local area network) 통신 모듈, 또는 전력선 통신 모듈)을 포함할 수 있다. 이들 통신 모듈 중 해당하는 통신 모듈은 제 1 네트워크(198)(예: 블루투스, WiFi(wireless fidelity) direct 또는 IrDA(infrared data association)와 같은 근거리 통신 네트워크) 또는 제 2 네트워크(199)(예: 레거시 셀룰러 네트워크, 5G 네트워크, 차세대 통신 네트워크, 인터넷, 또는 컴퓨터 네트워크(예: LAN 또는 WAN)와 같은 원거리 통신 네트워크)를 통하여 외부의 전자 장치(104)와 통신할 수 있다. 이런 여러 종류의 통신 모듈들은 하나의 구성요소(예: 단일 칩)로 통합되거나, 또는 서로 별도의 복수의 구성요소들(예: 복수 칩들)로 구현될 수 있다. 무선 통신 모듈(192)은 가입자 식별 모듈(196)에 저장된 가입자 정보(예: 국제 모바일 가입자 식별자(IMS))를 이용하여 제 1 네트워크(198) 또는 제 2 네트워크(199)와 같은 통신 네트워크 내에서 전자 장치(101)를 확인 또는 인증할 수 있다.
- [83] 무선 통신 모듈(192)은 4G 네트워크 이후의 5G 네트워크 및 차세대 통신 기술, 예를 들어, NR 접속 기술(new radio access technology)을 지원할 수 있다. NR 접속 기술은 고용량 데이터의 고속 전송(eMBB(enhanced mobile broadband)), 단말 전력 최소화 및 다수 단말의 접속(mMTC(massive machine type communications)), 또는 고신뢰도와 저지연(URLLC(ultra-reliable and low-latency communications))을 지원할 수 있다. 무선 통신 모듈(192)은, 예를 들어, 높은 데이터 전송률 달성을 위해, 고주파 대역(예: mmWave 대역)을 지원할 수 있다. 무선 통신 모듈(192)은 고주파 대역에서의 성능 확보를 위한 다양한 기술들, 예를 들어, 빔포밍(beamforming), 거대 배열 다중 입출력(massive MIMO(multiple-input and multiple-output)), 전차원 다중입출력(FD-MIMO: full dimensional MIMO), 어레이 안테나(array antenna), 아날로그 빔형성(analog beam-forming), 또는 대규모 안테나(large scale antenna)와 같은 기술들을 지원할 수 있다. 무선 통신 모듈(192)은 전자 장치(101), 외부 전자 장치(예: 전자 장치(104)) 또는 네트워크 시스템(예: 제 2 네트워크(199))에 규정되는 다양한 요구사항을 지원할 수 있다. 일실시예에 따르면, 무선 통신 모듈(192)은 eMBB 실현을 위한 Peak data rate(예: 20Gbps 이상), mMTC 실현을 위한 손실 Coverage(예: 164dB 이하), 또는 URLLC 실현을 위한 U-plane

latency(예: 다운링크(DL) 및 업링크(UL) 각각 0.5ms 이하, 또는 라운드 트립 1ms 이하)를 지원할 수 있다.

- [84] 안테나 모듈(197)은 신호 또는 전력을 외부(예: 외부의 전자 장치)로 송신하거나 외부로부터 수신할 수 있다. 일실시예에 따르면, 안테나 모듈(197)은 서브스트레이트(예: PCB) 위에 형성된 도전체 또는 도전성 패턴으로 이루어진 방사체를 포함하는 안테나를 포함할 수 있다. 일실시예에 따르면, 안테나 모듈(197)은 복수의 안테나들(예: 어레이 안테나)을 포함할 수 있다. 이런 경우, 제 1 네트워크(198) 또는 제 2 네트워크(199)와 같은 통신 네트워크에서 사용되는 통신 방식에 적합한 적어도 하나의 안테나가, 예를 들면, 통신 모듈(190)에 의하여 상기 복수의 안테나들로부터 선택될 수 있다. 신호 또는 전력은 상기 선택된 적어도 하나의 안테나를 통하여 통신 모듈(190)과 외부의 전자 장치 간에 송신되거나 수신될 수 있다. 어떤 실시예에 따르면, 방사체 이외에 다른 부품(예: RFIC(radio frequency integrated circuit))이 추가로 안테나 모듈(197)의 일부로 형성될 수 있다.
- [85] 일 실시예에 따르면, 안테나 모듈(197)은 mmWave 안테나 모듈을 형성할 수 있다. 일실시예에 따르면, mmWave 안테나 모듈은 인쇄 회로 기판, 상기 인쇄 회로 기판의 제 1 면(예: 아래 면)에 또는 그에 인접하여 배치되고 지정된 고주파 대역(예: mmWave 대역)을 지원할 수 있는 RFIC, 및 상기 인쇄 회로 기판의 제 2 면(예: 윗 면 또는 측면)에 또는 그에 인접하여 배치되고 상기 지정된 고주파 대역의 신호를 송신 또는 수신할 수 있는 복수의 안테나들(예: 어레이 안테나)을 포함할 수 있다.
- [86] 상기 구성요소들 중 적어도 일부는 주변 기기들간 통신 방식(예: 버스, GPIO(general purpose input and output), SPI(serial peripheral interface), 또는 MIPI(mobile industry processor interface))를 통해 서로 연결되고 신호(예: 명령 또는 데이터)를 상호간에 교환할 수 있다.
- [87] 일실시예에 따르면, 명령 또는 데이터는 제 2 네트워크(199)에 연결된 서버(108)를 통해서 전자 장치(101)와 외부의 전자 장치(104)간에 송신 또는 수신될 수 있다. 외부의 전자 장치(102, 또는 104) 각각은 전자 장치(101)와 동일한 또는 다른 종류의 장치일 수 있다. 일실시예에 따르면, 전자 장치(101)에서 실행되는 동작들의 전부 또는 일부는 외부의 전자 장치들(102, 104, 또는 108) 중 하나 이상의 외부의 전자 장치들에서 실행될 수 있다. 예를 들면, 전자 장치(101)가 어떤 기능이나 서비스를 자동으로, 또는 사용자 또는 다른 장치로부터의 요청에 반응하여 수행해야 할 경우에, 전자 장치(101)는 기능 또는 서비스를 자체적으로 실행시키는 대신에 또는 추가적으로, 하나 이상의 외부의 전자 장치들에게 그 기능 또는 그 서비스의 적어도 일부를 수행하라고 요청할 수 있다. 상기 요청을 수신한 하나 이상의 외부의 전자 장치들은 요청된 기능 또는 서비스의 적어도 일부, 또는 상기 요청과 관련된 추가 기능 또는 서비스를 실행하고, 그 실행의 결과를 전자 장치(101)로 전달할 수 있다. 전자 장치(101)는 상기 결과를, 그대로 또는 추가적으로 처리하여, 상기 요청에 대한 응답의 적어도 일부로서 제공할 수 있다. 이

를 위하여, 예를 들면, 클라우드 컴퓨팅, 분산 컴퓨팅, 모바일 에지 컴퓨팅(MEC: mobile edge computing), 또는 클라이언트-서버 컴퓨팅 기술이 이용될 수 있다. 전자 장치(101)는, 예를 들어, 분산 컴퓨팅 또는 모바일 에지 컴퓨팅을 이용하여 초저지연 서비스를 제공할 수 있다. 다른 실시예에 있어서, 외부의 전자 장치(104)는 IoT(internet of things) 기기를 포함할 수 있다. 서버(108)는 기계 학습 및/또는 신경망을 이용한 지능형 서버일 수 있다. 일 실시예에 따르면, 외부의 전자 장치(104) 또는 서버(108)는 제 2 네트워크(199) 내에 포함될 수 있다. 전자 장치(101)는 5G 통신 기술 및 IoT 관련 기술을 기반으로 지능형 서비스(예: 스마트 홈, 스마트 시티, 스마트 카, 또는 헬스케어)에 적용될 수 있다.

- [88] 도 2는 일 실시예에 따른 전자 장치의 TTS 모델의 수정 동작을 설명하기 위한 흐름도이다.
- [89] 도 2를 참조하면, 210 동작에서, 전자 장치(예: 도 1의 전자 장치(101) 또는 도 1의 프로세서(120))는 메모리(예: 도 1의 메모리(130))에 저장된, 복수의 노드를 포함하는 TTS(text to speech) 모델에 기반하여 음성 신호를 출력할 수 있다. 일 실시예에 따라, 음성 신호는 스피커(예: 도 1의 음향 출력 모듈(155))를 통해 전자 장치 외부로 출력될 수 있다.
- [90] TTS는 텍스트 형태의 정보를 음성 형태의 정보로 변경하는 기술이다. 일 실시예에 따라, TTS는 {텍스트, 음원}으로 이루어진 데이터 쌍을 학습하여, 임의의 텍스트가 주어졌을 때 이에 가장 적합한 음성을 생성할 수 있다. 텍스트는 각 언어 발화 특성을 모두 포괄할 수 있는 발음열로 변환이 되며, 전자 장치는 주어진 발음열로부터 이를 가장 유사하게 모사할 수 있는 특징을 음향모델로부터 획득하고, 이를 음원(또는 음성 신호)으로 변환하는 과정을 거치게 된다.
- [91] 일 실시예에 따라, 음향모델은 일반적으로 유닛 선택 방식, 통계 기반 방식 및 딥러닝 방식 중 하나로부터 만들어질 수 있다. 예를 들어, 유닛 선택 방식은, 실제 음성 조각을 구조화하여 보유하고 있다가, 요청된 발음열에 가장 맞는 음성 조각을 선택 이후 음성 조각들을 이어붙여 음원으로 생성하는 기술이다.
- [92] 일 실시예에 따라, 통계 기반 방식은 음성 데이터로부터 특징 파라미터를 추출하고, 이를 클러스터링(clustering)하여 각 발음열의 대표 파라미터를 구성하고, 수학적인 소스 필터(source-filter) 모델을 활용하여 음원을 생성하는 기술이다.
- [93] 일 실시예에 따라, 딥러닝 방식은, 통계 기반 방식에서 각 발음열의 대표 파라미터를 구성하는 부분 및 소스 필터(source-filter) 모델을 딥러닝 모델로 대체하는 것일 수 있다.
- [94] 딥러닝 방식은 대량의 모델 파라미터(예: 가중치(weight), 활성화 함수 정보, 모델 구조 등)를 포함하는 신경망(neural net) 모델에 기반하여 합성음을 생성하며, 각 모델 파라미터가 갖는 자유도 및 가중치 간의 관계로부터 대량의 정보를 담고 이를 표현해낼 수 있다. 모델 파라미터들은 레이어(layer)와 같이 계층 형태로 구성될 수 있으며, 하나의 모델은 수십~수백 레이어를 포함할 수 있다. 모델 파라미터의 수는 수십만에서 수십억에 이르기도 하며, 일반적으로 음성 합성/생성 분야

의 모델들은 수백만~수천만 가량의 모델 파라미터를 포함할 수 있다. 텍스트로부터 음원 샘플 사이의 관계를 학습하는 방식도 있으며, 텍스트로부터 음성 파라미터열을 먼저 만들고, 음성 파라미터열에서부터 음원 샘플을 만들어내는 방식도 있다.

- [95] 일 실시 예에 따라, 220 동작에서, 전자 장치는 음성 신호에 오류가 포함됨이 확인됨에 기반하여, 음성 신호에 포함된 오류 부분을 확인할 수 있다.
- [96] 일 실시 예에 따라, 전자 장치는 출력된 음성 신호에 오류가 포함됨을 나타내는 사용자 입력이 수신됨에 기반하여 음성 신호에 포함된 오류 부분을 확인할 수 있다.
- [97] 일 실시 예에 따라, 전자 장치는 음성 신호에 포함된 복수의 음소 중 설정된 길이 이상인 적어도 하나의 음소가 포함됨에 기반하여, 적어도 하나의 음소에 대응되는 부분을 오류 부분인 것으로 확인할 수 있다. 예를 들어, 동일한 음소가 여러 번 중복되는 경우, 오류일 확률이 높으므로, 전자 장치는 설정된 길이 이상인 음소에 대응되는 부분을 오류 부분으로 확인할 수 있다. 예를 들어, /ㅏ/ 발음의 최대 길이를 200ms으로 특정했는데, 이를 넘어서는 길이의 /ㅏ/가 생성문장 내에 존재할 경우, 전자 장치는 발화 오류가 발생했을 것으로 판단할 수 있다. 일 실시 예에 따라, 발화 오류를 판단하기 위한 각 발음의 최대 길이에 해당하는지 여부는 음원 생성 이전의 중간 단계 결과물(예: 발음열)로도 결정 가능하며, 이 경우 음원 생성 중간 단계인 프레임 레벨(frame level)에서 길이 값을 환산하여 오류 부분을 확인할 수 있다.
- [98] 일 실시 예에 따라, 전자 장치는 음성 신호의 파형 중 설정된 범위를 벗어나는 값(예: 이상 파형)을 가지는 파형 부분이 포함됨에 기반하여, 파형 부분을 오류 부분인 것으로 확인할 수 있다. 일 실시 예에 따라, 같은 글자로 구성된 문장이라도 띄어쓰기에 따라 음성의 표현 특성 약간씩 다르게 나타나는 것에서, 특정 음소열로 만들어진 출력이 다른 음소열 대비 파형의 차이가 현저한 경우, 전자 장치는 오류 부분으로 확인할 수 있다.
- [99] 예를 들어, 사용자가 '아버지가 방에 들어가신다' 라는 음원에서 '방에' 부분에 오류가 발생함이 확인되었을때, 전자 장치는 '아버지 가방에 들어가신다' 및/또는 '아버지가 방에 들어 가신다'와 같이 정상적인 문장 내지 '아버 지가방 에들어 가신 다' 및/또는 '아버지 가 방에 들 어가신 다'와 같은 유사한 음소열을 생성할 수 있다. 각 문장에 대해 느낌은 미묘하게 다를 수 있으나, 스펙트럼 파형에서의 차이(예: distance) 관점에서는 그 차이가 크지 않다. 그러나, '방에' 위치에 잡음이 나타났다면 이 부분은 다른 음소열에 대응되는 음성 신호와 대비했을 때 스펙트럼 파형의 차이가 클 것이고, 전자 장치는 이 부분을 오류 부분으로 확인할 수 있다.
- [100] 일 실시 예에 따라, 메모리는 자동 음성 인식(automatic speech recognition, ASR) 모델을 포함할 수 있다. 일 실시 예에 따라, 적어도 하나의 프로세서는, ASR 모델을 이용하여 상기 음성 신호를 인식한 텍스트를 획득하고, 텍스트가 상기 TTS 모

텔에 입력된 상기 음성 신호에 대응되는 입력 텍스트와 다른 부분이 포함됨에 기반하여, 다른 부분을 오류 부분인 것으로 확인할 수 있다.

- [101] 일 실시 예의 자동 음성 인식 모델은 음성 입력을 텍스트 데이터로 변환할 수 있다.
- [102] 일 실시 예에 따라, 음성 인식의 경우 언어 모델에 따라 결과가 달라지는 경향도 있으므로, 전자 장치는 출력된 음성에 대한 음성 인식 결과를 기반으로 오류 부분을 보다 정확히 확인할 수 있다.
- [103] 일 실시 예에 따라, 전자 장치는 사용자 입력을 통해 오류 부분을 확인할 수 있다. 일 실시 예에 따라, 전자 장치는 TTS 모델에 입력된 음성 신호에 대응되는 입력 텍스트를 디스플레이에 표시하고, 디스플레이를 통해 입력 텍스트 중 오류 부분을 선택하는 사용자 입력을 수신할 수 있다.
- [104] 일 실시 예에 따라, 전자 장치는 사용자 입력에 의해 선택된 입력 텍스트의 오류 부분에 대응되는 음성 부분을 음성 신호의 오류 부분으로 확인할 수 있다. 일 실시 예에 따라, 사용자 입력에 의해 오류 부분을 특정하는 동작은 이하 도 6 및 도 8을 참조하여 설명하기로 한다.
- [105] 일 실시 예에 따라, 230 동작에서, 전자 장치는 오류 부분과 관련된, 복수의 노드 각각의 활성도를 확인할 수 있다.
- [106] 일 실시 예에 따라, 전자 장치는 LRP(layer-wise relevance propagation) 기술을 통해 TTS 모델에 포함된 복수의 노드의 활성도(또는 기여도)를 확인할 수 있다.
- [107] 일 실시 예에 따라, 가중치와 노드 활성도를 기반으로 TTS 모델에 포함된 복수의 노드 각각의 출력에 대한 기여도를 노드 단위로 정량화할 수 있다.
- [108] 일 실시 예에 따라, LRP 기술은 모델의 작동 원리를 분석하는 기술, 설명 가능한 AI (eXplainable AI, XAI) 기술 또는 출력에 대한 기여도 정량화 기술로 지칭될 수 있다.
- [109] 일 실시 예에 따라, 240 동작에서, 전자 장치는 복수의 노드 각각의 활성도를 기반으로 복수의 노드 중 적어도 하나의 노드를 수정할 수 있다.
- [110] 일 실시 예에 따라, 전자 장치는 오류 부분에 대한 복수의 노드 각각의 활성도를 획득하고, 오류 부분에 대한 기여도가 높은 적어도 하나의 노드를 수정할 수 있다. 일 실시 예에 따라, 전자 장치는 오류 부분에 대한 기여도가 높은 적어도 하나의 노드와 관련된 가중치를 줄일 수 있다.
- [111] 일 실시 예에 따라, 전자 장치는 오류 부분에 대한 기여도가 높은 적어도 하나의 노드에 대한 가중치를 0으로 수정할 수 있다.
- [112] 일 실시 예에 따라, 전자 장치는 적어도 하나의 노드를 오류 부분에 대응되는 텍스트와 관련하여 기저장된 적어도 하나의 노드로 교체할 수 있다. 예를 들어, '아버지가 방에 들어가신다'라는 입력 텍스트에서 '방'에 오류 부분으로 확인된 경우, '방'을 정상적인 음성으로 변경하는 것으로 검증이 완료된 적어도 하나의 노드(또는 알고리즘)를 오류 부분과 관련된 적어도 하나의 노드로 교체할 수 있다.

- [113] 일 실시 예에 따라, 전자 장치는 입력 텍스트와 유사한 문자열을 통해 수정된 TTS 모델을 검증할 수 있다. 하기 개시된 하나 이상의 실시예를 통해 획득된 문자열과 수정된 TTS 모델을 이용하여 합성음이 생성되고, 이 합성음에서 오류 여부가 판단될 수 있다. 오류가 없을 경우 수정된 TTS 모델이, 이후 음성 합성에 사용될 수 있다. 오류가 있을 경우, 예를 들어 230 단계와 같은 추가 수정이 수행될 수 있다.
- [114] 일 실시 예에 따라, 전자 장치는 TTS 모델에 입력된 음성 신호에 대응되는 입력 텍스트의 문장 구조를 확인하고, 문장 구조를 기반으로 적어도 하나의 문자열을 획득할 수 있다. 일 실시 예에 따라, 입력 텍스트가 서술형의 문장 구조를 가진 경우, 전자 장치는 서술형인 적어도 하나의 문자열을 획득할 수 있다. 일 실시 예에 따라, 입력 텍스트가 뉴스 헤드라인과 같이 종결어에 명사가 배치된 경우, 전자 장치는 종결어에 명사가 배치된 적어도 하나의 문자열을 획득할 수 있다.
- [115] 일 실시 예에 따라, 적어도 하나의 문자열은, TTS 모델에 입력된 음성 신호에 대응되는 입력 텍스트의 상기 오류 부분 전 및/또는 후의 텍스트를 변경한 것일 수 있다. 예를 들어, 전자 장치는 오류가 발생한 문자열이 포함된 단어를 유지하고, 단어의 전후의 단어를 다른 유사 단어로 수정하여 입력 텍스트와 유사한 적어도 하나의 문자열을 획득할 수 있다.
- [116] 일 실시 예에 따라, 전자 장치는 문자열 내 특정 음소에서 오류가 발생한 경우, 전자 장치는 오류가 발생한 특정 음소를 다른 음소로 변경한 적어도 하나의 문자열을 획득할 수 있다.
- [117] 일 실시 예에 따라, 전자 장치는 기저장된 복수의 문자열 중 오류 부분과 동일한 글자 및/또는 단어를 포함하는 적어도 하나의 문자열을 획득할 수 있다.
- [118] 일 실시 예에 따라, 전자 장치는 기저장된 문장 템플릿에 오류가 발생한 단어를 삽입하여 적어도 하나의 문자열을 획득할 수 있다. 전자 장치는 획득된 적어도 하나의 문자열을 TTS 모델에 입력하고, TTS 모델에 기반하여 출력된 적어도 하나의 문자열에 대한 음성 신호를 기반으로 오류 부분의 수정 여부를 판단할 수 있다.
- [119] 일 실시 예에 따라, 전자 장치는 서버(예: 도 1의 서버(108))를 통해 다른 전자 장치(예: 도 1의 전자 장치(104)) 및/또는 사용자에게 의해 수정된 TTS 모델의 수정 정보를 전자 장치에 저장된 TTS 모델을 업데이트하는데 사용할 수 있다.
- [120] 예를 들어, 전자 장치는 오류 부분 및 상기 적어도 하나의 노드의 수정과 관련된 정보를 서버로 전송하고, 서버로부터 수정된 TTS 모델을 수신하고, 수정된 TTS 모델을 기반으로 메모리에 저장된 TTS 모델을 업데이트할 수 있다.
- [121] 일 실시 예에 따라, 다른 전자 장치 및/또는 사용자에게 의해 수정된 TTS 모델의 수정 정보를 이용하는 전자 장치의 동작은 이하 도 7을 참조하여 설명하기로 한다.
- [122] 도 3은 일 실시 예에 따른 전자 장치의 TTS 모델의 수정 동작을 설명하기 위한 도면이다.

- [123] 일 실시 예에 따라, 전자 장치(예: 도 1의 전자 장치(101) 또는 도 1의 프로세서(120))는, 310 동작에서, 사용자 리포트를 획득할 수 있다. 일 실시 예에 따라, 전자 장치는 TTS 모델을 통해 획득된 음성 신호를 출력하고, 음성 신호에 오류가 있음을 알리는 사용자 입력(예: 사용자 리포트)이 수신할 수 있다.
- [124] 일 실시 예에 따라, 320 동작에서, 전자 장치는 TTS 실패 문자열을 획득할 수 있다.
- [125] 일 실시 예에 따라, 전자 장치는 도 2의 220 동작과 같이, 음성 신호에 설정된 길이 이상인 적어도 하나의 음소 부분, 설정된 범위를 벗어나는 파형 부분, 음성 신호를 자동 음성 인식 했을 때 인식된 텍스트와 입력 텍스트가 다른 부분 및/또는 사용자 입력에 의해 오류인 것으로 선택된 부분을 TTS 실패 문자열로 획득할 수 있다.
- [126] 일 실시 예에 따라, 330 동작에서, 전자 장치는 오류 노드를 특정 및 수정할 수 있다.
- [127] 일 실시 예에 따라, 전자 장치는 도 2의 230 동작과 같이, TTS 모델(340)의 복수의 노드 중 LRP 기술을 통해 오류 부분과 관련된 적어도 하나의 노드를 확인할 수 있다.
- [128] 일 실시 예에 따라, 전자 장치는 도 2의 240 동작과 같이, 확인된 적어도 하나의 노드의 가중치를 줄이거나, 적어도 하나의 노드를 0으로 수정하거나, 신뢰할만한 다른 알고리즘으로 교체하여 적어도 하나의 노드를 수정할 수 있다.
- [129] 일 실시 예에 따라, TTS 모델은 특정한 패턴의 문자열에서 발화 오류가 발생하였을 때, 유사한 패턴에서도 발화 오류를 일으킬 수 있다. 예를 들어, 모든 문장이 서술형으로 끝나는 데이터로 학습한 TTS 모델은, 뉴스 헤드라인 등 명사가 문장 종결 위치에 있을 경우 발화 오류를 만들어낼 가능성이 높다.
- [130] 이에 이하에서는 TTS 모델의 안정성을 향상시키기 위하여 TTS 실패 문자열과 유사한 문자열을 통해 수정된 TTS 모델을 검증하는 동작을 설명한다.
- [131] 일 실시 예에 따라, 350 동작에서, 전자 장치는 유사한 패턴의 문자열을 생성할 수 있다.
- [132] 일 실시 예에 따라, 전자 장치는 수정된 TTS 모델을 검증하기 위해, 오류 문자열과 유사한 문자열들을 생성할 수 있다.
- [133] 예를 들어, '아버지가 방에 들어가신다'라는 입력 텍스트에서 /방/ 부분에 발화 오류가 발생하였고, 오류와 관련된 TTS 모델의 적어도 하나의 노드를 수정하였다면, 전자 장치는 /방/을 유지하고, /방/과 인접한 글자/단어들을 다른 글자/단어들로 변경하여, /방/의 발화가 정상적인지 여부를 확인할 수 있다.
- [134] 예를 들어, 전자 장치는 '아버지가 방을 들어가신다', '아버지의 방에 들어가신다' 및/또는 '아버지는 방에도 들어가신다'와 같은 유사 문자열을 생성할 수 있다. 일 실시 예에 따라, 전자 장치는 상술한 예시 이외에도 많은 유사 문자열을 생성할 수 있다.

- [135] 예를 들어, 전자 장치는 입력 텍스트의 특정 음소에서 오류가 발생함을 확인했다면, 해당 음소를 다른 음소로 변경한 유사 문자열을 획득할 수 있다. 일 실시 예에 따라, 전자 장치는 기 저장된 여러 문장들 중 단어를 포함하는 문장을 유사 문자열로 획득할 수 있다. 일 실시 예에 따라, 전자 장치는 기 저장된 문장 템플릿에 오류 발생 단어를 삽입하여 유사 문자열을 생성할 수 있다.
- [136] 일 실시 예에 따라, 360 동작에서, 전자 장치는 수정한 모델에 기반하여 유사한 패턴의 문자열의 음원을 생성할 수 있다.
- [137] 일 실시 예에 따라, 전자 장치는 생성된 적어도 하나의 유사 문자열을 수정된 TTS 모델에 입력하여 적어도 하나의 유사 문자열에 각각 대응되는 적어도 하나의 음성 신호를 획득할 수 있다.
- [138] 일 실시 예에 따라, 370 동작에서, 전자 장치는 생성된 합성음의 오류 여부를 확인할 수 있다.
- [139] 일 실시 예에 따라, 전자 장치는 적어도 하나의 유사 문자열에 각각 대응되는 적어도 하나의 음성 신호에 설정된 길이 이상인 적어도 하나의 음소 부분, 설정된 범위를 벗어나는 파형 부분, 음성 신호를 자동 음성 인식 했을 때 인식된 텍스트와 입력 텍스트가 다른 부분 및/또는 사용자 입력에 의해 오류인 것으로 선택된 부분과 같은 오류 부분이 포함되는지 여부를 확인할 수 있다.
- [140] 일 실시 예에 따라, 생성된 합성음에 오류가 없으면(370-아니오), 전자 장치는 수정된 노드를 TTS 모델(340)에 반영할 수 있다.
- [141] 일 실시 예에 따라, 생성된 합성음에 오류가 있으면,(370-예), 전자 장치는 320 동작으로 돌아가 오류 부분을 다시 확인하고 수정을 반복할 수 있다.
- [142] 이와 같이 TTS 모델이 특정 오류에 대해 수정을 하였다면, 이와 유사한 패턴의 발화에 대해서도 점검하여, 유사 패턴의 문자열들에도 범용적으로 적용 가능한 수정이 된 것인지 확인하여 모델 안정성을 향상시킬 수 있다.
- [143] 도 4a는 일 실시 예에 따른 전자 장치의 음성 신호의 오류 부분을 확인하는 동작을 설명하기 위한 도면이다.
- [144] 도 4a를 참조하면, 전자 장치(예: 도 1의 전자 장치(101) 또는 도 1의 프로세서(120))는 입력 텍스트(input text)(410)와 TTS 모델(420)을 이용하여 음성 신호를 획득할 수 있다.
- [145] 일 실시 예에 따라, TTS 모델(420)은 발화와 관련된 인코더(encoder)(421) 및 음색과 관련된 디코더(422)를 통해 음성 신호(400)를 획득할 수 있다. 예를 들어, TTS 모델(420)은 sequence-to-sequence 모델에 기반할 수 있다. 예를 들어, TTS 모델(420)은, 인코더(421), 디코더(422) 및/또는 어텐션(423)을 포함할 수 있다. 어텐션(423)은 디코더(422)에서 현재 출력값을 결정하기 위해 이용되는 문맥 정보로서, 인코더(421)에서 생성된 전체 은닉 상태 벡터들(hidden state vectors)을 가중합(weighted sum)하기 위한 가중치(weights)로 이해될 수 있다.
- [146] 일 실시 예에 따라, 전자 장치는 TTS 모델(420)로부터 출력된 음성 신호에 포함된 오류 부분(artifact)(430)을 확인할 수 있다. 예를 들어, 전자 장치는 도 2의 220

동작과 같이, 음성 신호에 설정된 길이 이상인 적어도 하나의 음소 부분, 설정된 범위를 벗어나는 파형 부분, 음성 신호를 자동 음성 인식 했을 때 인식된 텍스트와 입력 텍스트가 다른 부분 및/또는 사용자 입력에 의해 오류인 것으로 선택된 부분을 오류 부분(430)으로 획득할 수 있다.

- [147] 도 4b는 일 실시 예에 따른 전자 장치의 TTS 모델의 적어도 하나의 노드 수정 동작을 설명하기 위한 도면이다. 도 4b에서, 점선으로 표시된 경로는 모델의 활용(또는, inference)을 위한 경로를 의미할 수 있으며, 실선으로 표시된 경로는 전파(예를 들어, LRP(layer-wise relevance propagation)일 수 있지만 제한이 없음)를 위한 경로를 의미할 수 있다.
- [148] 도 4b를 참조하면, 전자 장치는 확인된 오류 부분(예: 도 4a의 오류 부분(430))에 대해 설명 가능한 AI 기법(예: layer-wise relevance propagation, LRP)(450)으로부터 TTS 모델의 인코더(440)(예: 도 4a의 TTS 모델의 인코더(421)) 내의 각 노드들이 정상 부분(460) 및/또는 오류 부분(461)(예: 도 4a의 오류 부분(430))을 생성하는데 기여한 정도를 측정할 수 있다. 예를 들어, ResNet classifier(450)는 인코더(421) 출력에 대해 정상 패턴 및/또는 비정상 패턴을 분류할 수 있으며, 사전에 학습될 수 있다. 예를 들어, 인코더(421), 디코더(422) 및/또는 어텐션(423)에 기반한 TTS 모델에 LRP를 적용하기 위하여 ResNet classifier(450)가 이용될 수 있다. 인코더(421)의 비정상 패턴을 탐지할 수 있도록 ResNet classifier(450)가 이용될 수 있으나, 이는 예시적인 것으로 예를 들어 VGG-16과 같은 분류기로 대체될 수도 있다. ResNet classifier(450)에서 검출된 비정상 패턴에 대해 인코더(421)에 LRP를 적용함으로써 비정상 패턴에 기여도가 상대적으로 큰 인코더 노드를 확인할 수 있으며, 해당 인코더 노드에 대한 수정이 수행될 수 있다.
- [149] 일 실시 예에 따라, 전자 장치는 오류 부분(461)에 기여도가 높은 노드는 오류를 유발하는 요인으로 볼 수 있어, 이를 수정할 수 있다.
- [150] 예를 들어, 전자 장치는 기여도가 높은 노드를 0으로 만들어 제거하거나, 관련 노드 가중치 값들을 줄여서 기여도를 낮출 수 있다. 예를 들어, 전자 장치는 가중치를 절반 이하로 줄일 수 있다. 변경된 관련 노드 가중치를 기반으로 TTS모델의 인코더(440)가 수정될 수 있다.
- [151] 도 4c는 일 실시 예에 따른 전자 장치의 수정된 TTS 모델을 통해 오류 부분이 수정된 음성 신호를 설명하기 위한 도면이다.
- [152] 도 4c를 참조하면, 전자 장치는 입력 텍스트(예: 도 4a의 입력 텍스트(410)) 및/또는 입력 텍스트와 유사한 문자열(470)을 수정된 인코더(440)를 포함하는 TTS 모델에 입력하여 오류 부분이 수정된 음성 신호(480)를 획득할 수 있다.
- [153] 도 5는 일 실시 예에 따른 전자 장치의 음성 신호의 오류 부분을 사용자 입력에 의해 확인하는 동작을 설명하기 위한 도면이다.
- [154] 도 5를 참조하면, 전자 장치(예: 도 1의 전자 장치(101) 또는 도 1의 프로세서(120))는 510 동작에서, TTS 실패 문자열을 표시할 수 있다. 예를 들어, 전자 장치

는 TTS 모델에서 출력된 음성 신호에 오류가 있다는 사용자 입력이 수신되면, TTS 실패 문자열이 포함된 입력 텍스트를 표시할 수 있다.

- [155] 일 실시 예에 따라, 520 동작에서, 전자 장치는 발화 오류가 발생된 문자열을 표시하는 사용자 입력을 수신할 수 있다.
- [156] 일 실시 예에 따라, 530 동작에서, 전자 장치는 사용자 입력에 의해 표시된 부분에 기여한 노드를 수정할 수 있다.
- [157] 예를 들어, 전자 장치는 사용자 입력에 의해 선택된 문자열과 관련된 TTS 모델에 포함된 복수의 노드의 기여도를 획득하고, 기여도가 설정된 값 이상인 적어도 하나의 노드를 수정할 수 있다.
- [158] 이와 같이 사용자에게 의해 오류가 발생된 문자열을 특정하는 경우, 오류 수정의 정확도가 향상될 수 있다.
- [159] 도 6은 일 실시 예에 따른 전자 장치의 사용자 입력에 의한 오류 부분을 수정하는 동작을 설명하기 위한 도면이다.
- [160] 도 6을 참조하면, 전자 장치(예: 도 1의 전자 장치(101) 또는 도 1의 프로세서(120))는 610 동작에서, 음원을 생성할 수 있다. 예를 들어, 생성된 음원은 TTS 모델의 출력 데이터로, 스피커(예: 도 1의 음향 출력 모듈(155))를 통해 출력될 수 있다. 예를 들어, TTS 모델에 기반하여 출력된 음성 신호는 /나는 학교에 갔다/(611)라는 입력 텍스트와 대응되는 것일 수 있다.
- [161] 일 실시 예에 따라, 620 동작에서, 전자 장치는 오류가 감지되었음을 리포트하는 사용자 입력을 수신할 수 있다.
- [162] 일 실시 예에 따라, 630 동작에서 전자 장치는 입력 텍스트 및/또는 출력된 음원의 파형을 유저 인터페이스로 제공할 수 있다. 예를 들어, 전자 장치는 음원의 파형을 발음열 별로 구분하고, 발음열과 대응되는 파형의 부분을 함께 표시할 수 있다. 예를 들어, 전자 장치는 음원의 파형을 입력 텍스트의 소분 단위(예: 어절, 음절, 또는 구)일 수 있지만 제한이 없음)로 구분하고, 소분 단위와 대응되는 파형의 일 부분을 함께(또는 단독으로) 표시할 수 있다.
- [163] 일 실시 예에 따라, 640 동작에서, 전자 장치는 입력 텍스트 및/또는 음원의 파형을 제공하는 유저 인터페이스를 통해 오류 부분(예: /에/)을 선택하는 사용자 입력을 수신할 수 있다.
- [164] 일 실시 예에 따라, 전자 장치는 발화 내 오류를 특정하는 동작(650)을 생략하고, 651 동작에서, 오류 노드 특정 및 수정을 수행할 수 있다. 예를 들어, 전자 장치는 음성 신호에 설정된 길이 이상인 적어도 하나의 음소 부분, 설정된 범위를 벗어나는 파형 부분 및/또는 음성 신호를 자동 음성 인식했을 때 인식된 텍스트와 입력 텍스트가 다른 부분을 확인하는 동작을 생략하고, 사용자 입력에 의해 선택된 부분을 오류 부분으로 확인할 수 있다.
- [165] 일 실시 예에 따라, 전자 장치는 설명가능한 AI 기법을 통해, TTS 모델에 포함된 복수의 노드 각각의 오류 부분에 대한 기여도를 획득하고, 오류 부분에 대한 기여도가 설정된 값 이상인 적어도 하나의 노드를 확인할 수 있다.

- [166] 일 실시 예에 따라, 전자 장치는 오류 부분에 대한 기여도가 설정된 값 이상인 적어도 하나의 노드를 0으로 수정하여 제거하거나, 가중치를 줄이거나, 오류 부분의 발음열과 관련하여 신뢰할만한 다른 알고리즘으로 교체할 수 있다.
- [167] 이와 같이 사용자에게 의해 오류가 발생한 문자열을 특정하는 경우, 오류 수정의 정확도가 향상될 수 있다.
- [168] 도 7은 일 실시 예에 따른 전자 장치의 여러 사용자로부터 수집된 오류 정보를 기반으로 TTS 모델을 업데이트하는 동작을 설명하기 위한 도면이다.
- [169] 도 7을 참조하면, 전자 장치(예: 도 1의 전자 장치(101) 또는 도 1의 프로세서(120))는 사용자 A의 오류 수정 노드 정보(710)를 획득할 수 있다. 일 실시 예에 따라, 전자 장치는 사용자 B의 오류 수정 노드 정보(711)를 획득할 수 있다. 일 실시 예에 따라, 전자 장치는 사용자 C의 오류 수정 노드 정보(712)를 획득할 수 있다. 일 실시 예에 따라, 오류 수정 노드 정보(710, 711, 712)는 같은 전자 장치에서 서로 다른 사용자에게 의해 획득되거나, 서로 다른 전자 장치에서 서로 다른 사용자에게 의해 획득될 수 있다. 일 실시 예에 따라, 오류 수정 노드 정보(710, 711, 712)는 오류가 발생한 발음열, 오류가 포함된 문장, 오류와 관련된 노드 정보, 노드의 수정 정보, 및/또는 사용된(예를 들어, 수정 이전의) TTS 모델 버전 정보를 포함할 수 있다.
- [170] 일 실시 예에 따라, 서로 다른 사용자에게 의해 획득된 오류 수정 노드 정보(710, 711, 712)는, 720 동작에서, 서버(예: 도 1의 서버(108))에 의해 취합될 수 있다.
- [171] 일 실시 예에 따라, 서버는 오류 수정 노드 정보(710, 711, 712)를 TTS 모델(730)에 반영하고, 740 동작을 통해 검증 후 업데이트를 수행할 수 있다. 예를 들어, 서버는 오류 수정 노드 정보(710, 711, 712)를 기반으로 수정된 TTS 모델(730)에 오류가 발생한 발음열과 유사한 패턴의 발음열을 입력하거나, 오류가 발생한 음소를 다른 음소로 변경한 발음 열을 입력하거나, 기 저장된 여러 문장들 중 오류가 발생한 단어를 포함하는 문장을 입력하거나, 기 저장된 문장 템플릿에 오류 발생 단어를 삽입한 문장을 입력하여 수정된 TTS 모델(730)을 검증할 수 있다. 일 실시 예에 따라, 수정된 TTS 모델(730)에 의해 오류가 발생되지 않음이 검증되면, 서버는 TTS 모델(730)을 업데이트할 수 있다.
- [172] 일 실시 예에 따라, 서버는 오류 수정 노드 정보(710, 711, 712)를 제공한 사용자에게 대응되는 적어도 하나의 전자 장치 및/또는 오류 리포트하지 않은 사용자X, Y(750, 751)에 대응되는 적어도 하나의 전자 장치에 수정된 TTS 모델(730)에 대한 정보를 전송하여 각 전자 장치에 저장된 TTS 모델을 업데이트할 수 있다.
- [173] 이와 같이, 각 사용자마다 만나는 문제가 서로 다를 경우, 각 사용자들이 발견한 오류들은 아직 그 오류를 만나지 못한 다른 사용자들이 미래에 잠재적으로 만날 수 있는 오류일 수 있다. 따라서 이를 정기적으로 서버에서 모은 후, 유효성을 검증한 후 하나의 모델에 담아 업데이트를 수행하면, 다른 사용자가 수정한 노드 정보까지 획득할 수 있다. 이 경우, 향후 TTS 모델을 사용자가 운용함에 있어, 잠

재적인 발화 오류 이슈를 사전에 제거한 것이 되므로 사용 만족감을 높여줄 수 있다.

- [174] 또한, 업데이트는 오류를 리포트하지 않은 사용자들에게도 미리 적용하여, 현재는 쓰고 있지 않으나 미래에 사용하게 될 사용자들이 발화 오류를 경험할 확률을 크게 낮춰줄 수 있다.
- [175] 도 8은 일 실시 예에 따른 전자 장치에 표시되는 오류 리포트를 설명하기 위한 도면이다.
- [176] 도 8을 참조하면, 전자 장치(예: 도 1의 전자 장치(101) 또는 도 1의 프로세서(120))는 TTS 모델을 통해 합성 음성(810)이 생성되면, 음성 출력을 위한 제1 화면(820)을 표시할 수 있다. 일 실시 예에 따라, 제1 화면(820)은 합성 음성(810)을 플레이 하기 위한 UI(821) 또는 합성 음성(810)에 에러가 포함됨을 리포트하기 위한 UI(822)를 포함할 수 있다.
- [177] 일 실시 예에 따라, 합성 음성(810)에 에러가 포함됨을 리포트하기 위한 UI(822)를 선택하는 사용자 입력이 수신되면, 전자 장치는 에러를 특정하기 위한 제2 화면(830)을 표시할 수 있다.
- [178] 일 실시 예에 따라, 제2 화면(830)은 합성 음성(810)의 파형(waveform)을 표시하는 음성 시각화 UI(831), 합성 음성(810)에 대응되는 입력 텍스트를 표시하는 UI(832), 에러의 종류를 선택하기 위한 UI(833) 및/또는 에러와 관련된 정보를 제출하기 위한 UI(834)를 포함할 수 있다.
- [179] 일 실시 예에 따라, 합성 음성(810)의 파형을 표시하는 음성 시각화 UI(831)는 합성 음성(810)의 파형을 발음열 별로 구분하여 표시하고, 각 발음열은 사용자 입력에 의해 선택될 수 있다.
- [180] 일 실시 예에 따라, 합성 음성(810)에 대응되는 입력 텍스트를 표시하는 UI(832)는 입력 텍스트의 전체를 표시하고, 각 텍스트는 사용자 입력에 의해 선택될 수 있다.
- [181] 일 실시 예에 따라, 에러의 종류를 선택하기 위한 UI(833)는, 중복된 음성, 잡음 및/또는 다른 발음과 같은 에러의 종류를 표시하고, 각 에러의 종류는 사용자 입력에 의해 선택될 수 있다.
- [182] 일 실시 예에 따라, 에러와 관련된 정보를 제출하기 위한 UI(834)를 선택하는 사용자 입력이 수신되면, 전자 장치는 수정 결과를 표시하기 위한 제3 화면(840)을 표시할 수 있다.
- [183] 일 실시 예에 따라, 전자 장치는 제3 화면(840)을 표시하기 전, 합성 음성(810)에 포함된 에러 부분을 특정한 정보 및/또는 제2 화면(830)을 통해 수신된 사용자 입력에 의해 특정된 에러 부분을 기반으로, 에러 부분과 관련된 TTS 모델의 적어도 하나의 노드를 확인하고, 적어도 하나의 노드를 수정할 수 있다.
- [184] 일 실시 예에 따라, 제3 화면(840)은 수정된 음원을 재생할 수 있는 UI(841), 수정된 구간과 관련된 정보를 제공하는 UI(842), 수정된 TTS 모델의 신뢰도를 시각

화한 UI(843) 및/또는 수정된 음원에 에러가 포함됨을 리포트하기 위한 UI(844)를 포함할 수 있다.

[185] 일 실시 예에 따라, 수정된 구간과 관련된 정보를 제공하는 UI(842)는, 수정된 TTS 모델에 의해 수정된 부분을 텍스트 및/또는 과형으로 표시할 수 있다.

[186] 일 실시 예에 따라, 수정된 TTS 모델의 신뢰도를 시각화한 UI(843)는, 오류가 발생한 입력 텍스트와 유사한 적어도 하나의 유사 문자열을 수정된 TTS 모델을 통해 검증한 검증 결과를 표시할 수 있다. 예를 들어, 수정된 TTS 모델의 신뢰도를 시각화한 UI(843)는 100개의 유사 문자열 중 정상적으로 발화된 문장의 비율을 표시할 수 있다.

[187] 도 9는 일 실시 예에 따른 TTS 모델의 관리자가 볼 수 있는 유저 인터페이스를 설명하기 위한 도면이다.

[188] 도 9를 참조하면, 전자 장치(예: 도 1의 전자 장치(101) 또는 도 1의 프로세서(120))는 관리자에게 TTS 모델의 수정 이력과 관련된 UI(910)를 제공할 수 있다.

[189] 일 실시 예에 따라, 관리자에게 제공되는 TTS 모델의 수정 이력과 관련된 UI(910)는 서버(예: 도 1의 서버(108))로부터 검출된 오류 및 수정 결과에 대한 정보를 수신할 수 있다.

[190] 일 실시 예에 따라, 관리자에게 제공되는 TTS 모델의 수정 이력과 관련된 UI(910)는 TTS 모델의 수정 기간, 수정된 샘플 건수, 수정된 샘플 중 정상 처리된 샘플 건수, 수정된 샘플 중 비정상 처리된 샘플 건수 및/또는 정상과 비정상을 분류하기 위한 임계 값(confidence threshold)에 대한 정보를 포함할 수 있다. 예를 들어, 임계 값은 수정될 수도 있으며, 하나의 예에 따라서 UI를 통하여 수정될 수 있으나 그 수정 방식에는 제한이 없다.

[191] 일 실시 예에 따라, 관리자에게 제공되는 TTS 모델의 수정 이력과 관련된 UI(910)는 정상 및 비정상 신뢰도에 대한 시각화된 정보(confidence figure) 및 샘플 리스트를 더 포함할 수 있다. 이로 인해 처리된 샘플 중 임계 값에 따른 정상 및 비정상의 변화를 확인할 수 있다.

[192] 일 실시 예에 따라, 관리자에게 제공되는 TTS 모델의 수정 이력과 관련된 UI(910)는 TTS 모델의 수정 전과 후의 성능 변화에 대한 수치 정보를 포함할 수 있다. 이로 인해, TTS 모델의 수정으로 인해 변화되는 정확도를 제공할 수 있다.

[193] 일 실시 예에 따르면, 전자 장치는, 메모리, 상기 메모리와 작동적으로 연결된 적어도 하나의 프로세서를 포함할 수 있다.

[194] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 메모리에 저장된, 복수의 노드를 포함하는 TTS(text to speech) 모델에 기반하여 음성 신호를 출력할 수 있다.

[195] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 음성 신호에 오류가 포함됨이 확인됨에 기반하여, 상기 음성 신호에 포함된 오류 부분을 확인할 수 있다.

- [196] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 오류 부분과 관련된, 상기 복수의 노드 각각의 활성도를 확인할 수 있다.
- [197] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 복수의 노드 각각의 활성도를 기반으로 상기 복수의 노드 중 적어도 하나의 노드를 수정할 수 있다.
- [198] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 적어도 하나의 노드와 관련된 가중치를 줄일 수 있다.
- [199] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 적어도 하나의 노드를 상기 오류 부분에 대응되는 텍스트와 관련하여 저장된 적어도 하나의 노드로 교체할 수 있다.
- [200] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 음성 신호에 포함된 복수의 음소 중 설정된 길이 이상인 적어도 하나의 음소가 포함됨에 기반하여, 상기 적어도 하나의 음소에 대응되는 부분을 상기 오류 부분인 것으로 확인할 수 있다.
- [201] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 음성 신호의 파형 중 설정된 범위를 벗어나는 값을 가지는 파형 부분이 포함됨에 기반하여, 상기 파형 부분을 상기 오류 부분인 것으로 확인할 수 있다.
- [202] 일 실시 예에 따라, 상기 메모리는 자동 음성 인식(automatic speech recognition, ASR) 모델을 포함할 수 있다.
- [203] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 ASR 모델을 이용하여 상기 음성 신호를 인식한 텍스트를 획득할 수 있다.
- [204] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 텍스트가 상기 TTS 모델에 입력된 상기 음성 신호에 대응되는 입력 텍스트와 다른 부분이 포함됨에 기반하여, 상기 다른 부분을 상기 오류 부분인 것으로 확인할 수 있다.
- [205] 일 실시 예에 따라, 상기 전자 장치는, 디스플레이를 더 포함할 수 있다.
- [206] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 TTS 모델에 입력된 상기 음성 신호에 대응되는 입력 텍스트를 상기 디스플레이에 표시할 수 있다.
- [207] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 디스플레이를 통해 상기 입력 텍스트 중 오류 부분을 선택하는 사용자 입력이 수신됨에 기반하여, 상기 음성 신호의 상기 오류 부분을 확인할 수 있다.
- [208] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 TTS 모델에 입력된 상기 음성 신호에 대응되는 입력 텍스트의 문장 구조를 확인할 수 있다.
- [209] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 문장 구조를 기반으로 적어도 하나의 문자열을 획득할 수 있다.
- [210] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 적어도 하나의 문자열을 상기 TTS 모델에 입력할 수 있다.
- [211] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 TTS 모델에 기반하여 출력된 상기 적어도 하나의 문자열에 대한 음성 신호를 기반으로 상기 오류 부분의 수정 여부를 판단할 수 있다.

- [121] 일 실시 예에 따라, 상기 적어도 하나의 문자열은, 상기 TTS 모델에 입력된 상기 음성 신호에 대응되는 입력 텍스트의 상기 오류 부분 전 및/또는 후의 텍스트를 변경한 것일 수 있다.
- [123] 일 실시 예에 따라, 상기 전자 장치는, 통신 모듈을 더 포함할 수 있다.
- [124] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 오류 부분 및 상기 적어도 하나의 노드의 수정과 관련된 정보를 서버로 전송할 수 있다.
- [125] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 서버로부터 수정된 TTS 모델을 수신할 수 있다.
- [126] 일 실시 예에 따라, 상기 적어도 하나의 프로세서는, 상기 수정된 TTS 모델을 기반으로 상기 메모리에 저장된 TTS 모델을 업데이트할 수 있다.
- [127] 일 실시 예에 따라, 전자 장치의 제어 방법은, 메모리에 저장된, 복수의 노드를 포함하는 TTS(text to speech) 모델에 기반하여 음성 신호를 출력하는 동작을 포함할 수 있다.
- [128] 일 실시 예에 따라, 전자 장치의 제어 방법은, 상기 음성 신호에 오류가 포함됨이 확인됨에 기반하여, 상기 음성 신호에 포함된 오류 부분을 확인하는 동작을 포함할 수 있다.
- [129] 일 실시 예에 따라, 전자 장치의 제어 방법은, 상기 오류 부분과 관련된, 상기 복수의 노드 각각의 활성도를 확인하는 동작을 포함할 수 있다.
- [120] 일 실시 예에 따라, 전자 장치의 제어 방법은, 상기 복수의 노드 각각의 활성도를 기반으로 상기 복수의 노드 중 적어도 하나의 노드를 수정하는 동작을 포함할 수 있다.
- [121] 일 실시 예에 따라, 상기 적어도 하나의 노드를 수정하는 동작은, 상기 적어도 하나의 노드와 관련된 가중치를 줄일 수 있다.
- [122] 일 실시 예에 따라, 상기 적어도 하나의 노드를 수정하는 동작은, 상기 적어도 하나의 노드를 상기 오류 부분에 대응되는 텍스트와 관련하여 기저장된 적어도 하나의 노드로 교체할 수 있다.
- [123] 일 실시 예에 따라, 상기 음성 신호에 포함된 오류 부분을 확인하는 동작은, 상기 음성 신호에 포함된 복수의 음소 중 설정된 길이 이상인 적어도 하나의 음소가 포함됨에 기반하여, 상기 적어도 하나의 음소에 대응되는 부분을 상기 오류 부분인 것으로 확인할 수 있다.
- [124] 일 실시 예에 따라, 상기 음성 신호에 포함된 오류 부분을 확인하는 동작은, 상기 음성 신호의 파형 중 설정된 범위를 벗어나는 값을 가지는 파형 부분이 포함됨에 기반하여, 상기 파형 부분을 상기 오류 부분인 것으로 확인할 수 있다.
- [125] 일 실시 예에 따라, 상기 메모리는 자동 음성 인식(automatic speech recognition, ASR) 모델을 포함할 수 있다.
- [126] 일 실시 예에 따라, 상기 음성 신호에 포함된 오류 부분을 확인하는 동작은, 상기 ASR 모델을 이용하여 상기 음성 신호를 인식한 텍스트를 획득할 수 있다.

- [227] 일 실시 예에 따라, 상기 텍스트가 상기 TTS 모델에 입력된 상기 음성 신호에 대응되는 입력 텍스트와 다른 부분이 포함됨에 기반하여, 상기 다른 부분을 상기 오류 부분인 것으로 확인할 수 있다.
- [228] 일 실시 예에 따라, 상기 음성 신호에 포함된 오류 부분을 확인하는 동작은, 상기 TTS 모델에 입력된 상기 음성 신호에 대응되는 입력 텍스트를 디스플레이에 표시할 수 있다.
- [229] 일 실시 예에 따라, 상기 음성 신호에 포함된 오류 부분을 확인하는 동작은, 상기 디스플레이를 통해 상기 입력 텍스트 중 오류 부분을 선택하는 사용자 입력이 수신됨에 기반하여, 상기 음성 신호의 상기 오류 부분을 확인할 수 있다.
- [230] 일 실시 예에 따라, 전자 장치의 제어 방법은, 상기 TTS 모델에 입력된 상기 음성 신호에 대응되는 입력 텍스트의 문장 구조를 확인하는 동작을 더 포함할 수 있다.
- [231] 일 실시 예에 따라, 전자 장치의 제어 방법은, 상기 문장 구조를 기반으로 적어도 하나의 문자열을 획득하는 동작을 더 포함할 수 있다.
- [232] 일 실시 예에 따라, 전자 장치의 제어 방법은, 상기 적어도 하나의 문자열을 상기 TTS 모델에 입력하는 동작을 더 포함할 수 있다.
- [233] 일 실시 예에 따라, 전자 장치의 제어 방법은, 상기 TTS 모델에 기반하여 출력된 상기 적어도 하나의 문자열에 대한 음성 신호를 기반으로 상기 오류 부분의 수정 여부를 판단하는 동작을 더 포함할 수 있다.
- [234] 일 실시 예에 따라, 상기 적어도 하나의 문자열은, 상기 TTS 모델에 입력된 상기 음성 신호에 대응되는 입력 텍스트의 상기 오류 부분 전 및/또는 후의 텍스트를 변경한 것일 수 있다.
- [235] 일 실시 예에 따라, 전자 장치의 제어 방법은, 상기 오류 부분 및 상기 적어도 하나의 노드의 수정과 관련된 정보를 서버로 전송하는 동작을 더 포함할 수 있다.
- [236] 일 실시 예에 따라, 전자 장치의 제어 방법은, 상기 서버로부터 수정된 TTS 모델을 수신하는 동작을 더 포함할 수 있다.
- [237] 일 실시 예에 따라, 전자 장치의 제어 방법은, 상기 수정된 TTS 모델을 기반으로 상기 메모리에 저장된 TTS 모델을 업데이트 하는 동작을 더 포함할 수 있다.
- [238] 일 실시 예에 따라, 하나 이상의 프로그램을 저장하는, 비일시적 컴퓨터 판독 가능 기록 매체에 있어서, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 메모리에 저장된, 복수의 노드를 포함하는 TTS(text to speech) 모델에 기반하여 음성 신호를 출력하는 명령어들을 포함할 수 있다.
- [239] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 음성 신호에 오류가 포함됨이 확인됨에 기반하여, 상기 음성 신호에 포함된 오류 부분을 확인하는 명령어들을 포함할 수 있다.
- [240] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 오류 부분과 관련된, 상기 복수의 노드 각각의 활성도를 확인하는 명령어들을 포함할 수 있다.

- [241] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 복수의 노드 각각의 활성도를 기반으로 상기 복수의 노드 중 적어도 하나의 노드를 수정하는 명령어들을 포함할 수 있다.
- [242] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 적어도 하나의 노드와 관련된 가중치를 줄이는 명령어들을 포함할 수 있다.
- [243] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 적어도 하나의 노드를 상기 오류 부분에 대응되는 텍스트와 관련하여 저장된 적어도 하나의 노드로 교체하는 명령어들을 포함할 수 있다.
- [244] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 음성 신호에 포함된 복수의 음소 중 설정된 길이 이상인 적어도 하나의 음소가 포함됨에 기반하여, 상기 적어도 하나의 음소에 대응되는 부분을 상기 오류 부분인 것으로 확인하는 명령어들을 포함할 수 있다.
- [245] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 음성 신호의 파형 중 설정된 범위를 벗어나는 값을 가지는 파형 부분이 포함됨에 기반하여, 상기 파형 부분을 상기 오류 부분인 것으로 확인하는 명령어들을 포함할 수 있다.
- [246] 일 실시 예에 따라, 상기 메모리는 자동 음성 인식(automatic speech recognition, ASR) 모델을 포함할 수 있다.
- [247] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 ASR 모델을 이용하여 상기 음성 신호를 인식한 텍스트를 획득하는 명령어들을 포함할 수 있다.
- [248] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 텍스트가 상기 TTS 모델에 입력된 상기 음성 신호에 대응되는 입력 텍스트와 다른 부분이 포함됨에 기반하여, 상기 다른 부분을 상기 오류 부분인 것으로 확인하는 명령어들을 포함할 수 있다.
- [249] 일 실시 예에 따라, 상기 전자 장치는, 디스플레이를 더 포함할 수 있다.
- [250] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 TTS 모델에 입력된 상기 음성 신호에 대응되는 입력 텍스트를 상기 디스플레이에 표시하는 명령어들을 포함할 수 있다.
- [251] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 디스플레이를 통해 상기 입력 텍스트 중 오류 부분을 선택하는 사용자 입력이 수신됨에 기반하여, 상기 음성 신호의 상기 오류 부분을 확인하는 명령어들을 포함할 수 있다.
- [252] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 TTS 모델에 입력된 상기 음성 신호에 대응되는 입력 텍스트의 문장 구조를 확인하는 명령어들을 포함할 수 있다.
- [253] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 문장 구조를 기반으로 적어도 하나의 문자열을 획득하는 명령어들을 포함할 수 있다.

- [254] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 적어도 하나의 문자열을 상기 TTS 모델에 입력하는 명령어들을 포함할 수 있다.
- [255] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 TTS 모델에 기반하여 출력된 상기 적어도 하나의 문자열에 대한 음성 신호를 기반으로 상기 오류 부분의 수정 여부를 판단하는 명령어들을 포함할 수 있다.
- [256] 일 실시 예에 따라, 상기 적어도 하나의 문자열은, 상기 TTS 모델에 입력된 상기 음성 신호에 대응되는 입력 텍스트의 상기 오류 부분 전 및/또는 후의 텍스트를 변경한 것일 수 있다.
- [257] 일 실시 예에 따라, 상기 전자 장치는, 통신 모듈을 더 포함할 수 있다.
- [258] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 오류 부분 및 상기 적어도 하나의 노드의 수정과 관련된 정보를 서버로 전송하는 명령어들을 포함할 수 있다.
- [259] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 서버로부터 수정된 TTS 모델을 수신하는 명령어들을 포함할 수 있다.
- [260] 일 실시 예에 따르면, 상기 하나 이상의 프로그램은, 전자 장치가, 상기 수정된 TTS 모델을 기반으로 상기 메모리에 저장된 TTS 모델을 업데이트하는 명령어들을 포함할 수 있다.
- [261] 본 문서에 개시된 하나 이상의 실시예들에 따른 전자 장치는 다양한 형태의 장치가 될 수 있다. 전자 장치는, 예를 들면, 휴대용 통신 장치(예: 스마트폰), 컴퓨터 장치, 휴대용 멀티미디어 장치, 휴대용 의료 기기, 카메라, 웨어러블 장치, 또는 가전 장치를 포함할 수 있다. 본 문서의 실시예에 따른 전자 장치는 전술한 기기들에 한정되지 않는다.
- [262] 본 문서의 일 실시예들 및 이에 사용된 용어들은 본 문서에 기재된 기술적 특징들을 특정한 실시예들로 한정하려는 것이 아니며, 해당 실시예의 다양한 변경, 균등물, 또는 대체물을 포함하는 것으로 이해되어야 한다. 도면의 설명과 관련하여, 유사한 또는 관련된 구성요소에 대해서는 유사한 참조 부호가 사용될 수 있다. 아이템에 대응하는 명사의 단수 형은 관련된 문맥상 명백하게 다르게 지시하지 않는 한, 상기 아이템 한 개 또는 복수 개를 포함할 수 있다. 본 문서에서, "A 또는 B", "A 및 B 중 적어도 하나", "A 또는 B 중 적어도 하나", "A, B 또는 C", "A, B 및 C 중 적어도 하나", 및 "A, B, 또는 C 중 적어도 하나"와 같은 문구들 각각은 그 문구들 중 해당하는 문구에 함께 나열된 항목들 중 어느 하나, 또는 그들의 모든 가능한 조합을 포함할 수 있다. "제 1", "제 2", 또는 "첫째" 또는 "둘째"와 같은 용어들은 단순히 해당 구성요소를 다른 해당 구성요소와 구분하기 위해 사용될 수 있으며, 해당 구성요소들을 다른 측면(예: 중요성 또는 순서)에서 한정하지 않는다. 어떤(예: 제 1) 구성요소가 다른(예: 제 2) 구성요소에, "기능적으로" 또는 "통신적으로"라는 용어와 함께 또는 이런 용어 없이, "커플드" 또는 "커넥티드"라고 언급된 경우, 그것은 상기 어떤 구성요소가 상기 다른 구성요소에 직접적으로

(예: 유선으로), 무선으로, 또는 제 3 구성요소를 통하여 연결될 수 있다는 것을 의미한다.

- [263] 본 문서의 하나 이상의 실시예들에서 사용된 용어 "모듈"은 하드웨어, 소프트웨어 또는 펌웨어로 구현된 유닛을 포함할 수 있으며, 예를 들면, 로직, 논리 블록, 부품, 또는 회로와 같은 용어와 상호 호환적으로 사용될 수 있다. 모듈은, 일체로 구성된 부품 또는 하나 또는 그 이상의 기능을 수행하는, 상기 부품의 최소 단위 또는 그 일부가 될 수 있다. 예를 들면, 일실시예에 따르면, 모듈은 ASIC(application-specific integrated circuit)의 형태로 구현될 수 있다.
- [264] 본 문서의 하나 이상의 실시예들은 기기(machine)(예: 전자 장치(101))의 헤 읽을 수 있는 저장 매체(storage medium)(예: 내장 메모리(136) 또는 외장 메모리(138))에 저장된 하나 이상의 명령어들을 포함하는 소프트웨어(예: 프로그램(140))로서 구현될 수 있다. 예를 들면, 기기(예: 전자 장치(101))의 프로세서(예: 프로세서(120))는, 저장 매체로부터 저장된 하나 이상의 명령어들 중 적어도 하나의 명령을 호출하고, 그것을 실행할 수 있다. 이것은 기기가 상기 호출된 적어도 하나의 명령어에 따라 적어도 하나의 기능을 수행하도록 운영되는 것을 가능하게 한다. 상기 하나 이상의 명령어들은 컴파일러에 의해 생성된 코드 또는 인터프리터에 의해 실행될 수 있는 코드를 포함할 수 있다. 기기로 읽을 수 있는 저장 매체는, 비일시적(non-transitory) 저장 매체의 형태로 제공될 수 있다. 여기서, '비일시적'은 저장 매체가 실제(tangible)하는 장치이고, 신호(signal)(예: 전자기파)를 포함하지 않는다는 것을 의미할 뿐이며, 이 용어는 데이터가 저장 매체에 반영구적으로 저장되는 경우와 일시적으로 저장되는 경우를 구분하지 않는다.
- [265] 일실시예에 따르면, 본 문서에 개시된 하나 이상의 실시예들에 따른 방법은 컴퓨터 프로그램 제품(computer program product)에 포함되어 제공될 수 있다. 컴퓨터 프로그램 제품은 상품으로서 판매자 및 구매자 간에 거래될 수 있다. 컴퓨터 프로그램 제품은 기기로 읽을 수 있는 저장 매체(예: compact disc read only memory(CD-ROM))의 형태로 배포되거나, 또는 어플리케이션 스토어(예: 플레이 스토어™)를 통해 또는 두 개의 사용자 장치들(예: 스마트폰들) 간에 직접, 온라인으로 배포(예: 다운로드 또는 업로드)될 수 있다. 온라인 배포의 경우에, 컴퓨터 프로그램 제품의 적어도 일부는 제조사의 서버, 어플리케이션 스토어의 서버, 또는 중계 서버의 메모리와 같은 기기로 읽을 수 있는 저장 매체에 적어도 일시 저장되거나, 일시적으로 생성될 수 있다.
- [266] 하나 이상의 실시예들에 따르면, 상기 기술한 구성요소들의 각각의 구성요소(예: 모듈 또는 프로그램)는 단수 또는 복수의 개체를 포함할 수 있으며, 복수의 개체 중 일부는 다른 구성요소에 분리 배치될 수도 있다. 하나 이상의 실시예들에 따르면, 전술한 해당 구성요소들 중 하나 이상의 구성요소들 또는 동작들이 생략되거나, 또는 하나 이상의 다른 구성요소들 또는 동작들이 추가될 수 있다. 대체적으로 또는 추가적으로, 복수의 구성요소들(예: 모듈 또는 프로그램)은 하나의 구성요소로 통합될 수 있다. 이런 경우, 하나 이상의 실시예들에 따르면, 통

합된 구성요소는 상기 복수의 구성요소들 각각의 구성요소의 하나 이상의 기능들을 상기 통합 이전에 상기 복수의 구성요소들 중 해당 구성요소에 의해 수행되는 것과 동일 또는 유사하게 수행할 수 있다. 하나 이상의 실시예들에 따르면, 모듈, 프로그램 또는 다른 구성요소에 의해 수행되는 동작들은 순차적으로, 병렬적으로, 반복적으로, 또는 휴리스틱하게 실행되거나, 상기 동작들 중 하나 이상이 다른 순서로 실행되거나, 생략되거나, 또는 하나 이상의 다른 동작들이 추가될 수 있다.

청구범위

- [청구항 1] 전자 장치에 있어서,
메모리;
상기 메모리와 작동적으로 연결된 적어도 하나의 프로세서;를 포함하고,
상기 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가,
상기 메모리에 저장된, 복수의 노드를 포함하는 TTS(text to speech) 모델에 기반하여 음성 신호를 획득하고, 상기 음성 신호는 입력 텍스트에 대응되는 것임,
상기 음성 신호에 오류가 포함됨을 확인함에 기반하여, 상기 오류가 포함된 상기 음성 신호의 오류 부분을 확인하고,
상기 오류 부분과 관련된, 상기 복수의 노드 각각의 활성도를 확인하고,
적어도 하나의 노드의 상기 식별된 활성도를 기반으로 상기 적어도 하나의 노드를 수정하는 명령어들을 저장하는 전자 장치.
- [청구항 2] 제1항에 있어서,
상기 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가,
상기 적어도 하나의 노드와 관련된 가중치를 줄이는 명령어들을 저장하는 전자 장치.
- [청구항 3] 제1항 또는 제2항에 있어서,
상기 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가,
상기 적어도 하나의 노드를 상기 오류 부분에 대응되는 텍스트와 관련하여 기저장된 적어도 하나의 노드로 교체하는 명령어들을 저장하고,
상기 기저장된 적어도 하나의 노드는 상기 메모리에 저장되고, 상기 오류 부분에 대응되는 텍스트에 대응되는 전자 장치.
- [청구항 4] 제1항 내지 제3항 중 어느 한 항에 있어서,
상기 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가,
상기 음성 신호가 설정된 길이 이상의 길이를 갖는 적어도 하나의 음소를 포함함을 확인함에 기반하여, 상기 적어도 하나의 음소에 대응되는 상기 음성 신호의 부분을 상기 오류 부분인 것으로 확인하는 명령어들을 저장하는 전자 장치.
- [청구항 5] 제1항 내지 제4항 중 어느 한 항에 있어서,
상기 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가,

상기 음성 신호가 이상 파형을 가지는 파형 부분을 포함함을 확인함에 기반하여, 상기 파형 부분을 상기 오류 부분인 것으로 확인하는 명령어들을 저장하는 전자 장치.

[청구항 6] 제1항 내지 제5항 중 어느 한 항에 있어서,
자동 음성 인식(automatic speech recognition, ASR) 모델은 상기 메모리에 저장되고,
상기 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가,
상기 음성 신호에 상기 ASR 모델을 적용한 결과인 텍스트를 획득하고,
상기 텍스트에 입력 텍스트와 다른 부분을 포함함을 확인함에 기반하여,
상기 입력 텍스트와 다른 부분을 상기 오류 부분인 것으로 확인하는 명령어들을 저장하는 전자 장치.

[청구항 7] 제1항 내지 제6항 중 어느 한 항에 있어서,
디스플레이;를 더 포함하고,
상기 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가,
상기 입력 텍스트를 상기 디스플레이에 표시하고,
상기 디스플레이를 통해 수신된 사용자 입력에 기반하여, 상기 오류 부분을 확인하고, 상기 사용자 입력은 상기 입력 텍스트의 부분의 선택을 포함하는 명령어들을 저장하는 전자 장치.

[청구항 8] 제1항 내지 제7항 중 어느 한 항에 있어서,
상기 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가,
상기 입력 텍스트의 문장 구조를 확인하고,
상기 문장 구조를 기반으로 적어도 하나의 문자열을 획득하고,
상기 적어도 하나의 문자열을 상기 TTS 모델에 입력함의 결과로 문자열 음성 신호를 획득하고,
상기 문자열 음성 신호에 기반하여 상기 오류 부분의 수정 여부를 확인하는 명령어들을 저장하는 전자 장치.

[청구항 9] 제8항에 있어서,
상기 적어도 하나의 문자열은,
상기 오류 부분에 대응되는 입력 텍스트의 부분 또는 후의 텍스트를 변경하여 획득된 것인 전자 장치.

[청구항 10] 제1항 내지 제9항 중 어느 한 항에 있어서,
통신 모듈;을 더 포함하고,
상기 메모리는 상기 적어도 하나의 프로세서에 의해 수행될 때, 상기 전자 장치가,

상기 오류 부분 및 상기 적어도 하나의 노드의 수정과 관련된 정보를 서버로 전송하도록 상기 통신 모듈을 제어하고,
 상기 통신 모듈을 통해, 상기 서버로부터 수정된 TTS 모델을 수신하고,
 상기 수정된 TTS 모델을 기반으로 상기 메모리에 저장된 TTS 모델을 업데이트 하는 명령어들을 저장하는 전자 장치.

- [청구항 11] 전자 장치를 제어하기 위한 방법에 있어서,
 상기 전자 장치의 메모리에 저장된, 복수의 노드를 포함하는 TTS(text to speech) 모델에 기반하여 음성 신호를 획득하는 동작, 상기 음성 신호는 입력 텍스트에 대응되는 것임;
 상기 음성 신호에 오류가 포함됨을 확인함에 기반하여, 상기 오류가 포함된 상기 음성 신호의 오류 부분을 확인하는 동작;
 상기 오류 부분과 관련된, 상기 복수의 노드 각각의 활성도를 확인하는 동작; 및
 적어도 하나의 노드의 상기 식별된 활성도를 기반으로 상기 적어도 하나의 노드를 수정하는 동작;을 포함하는, 방법.

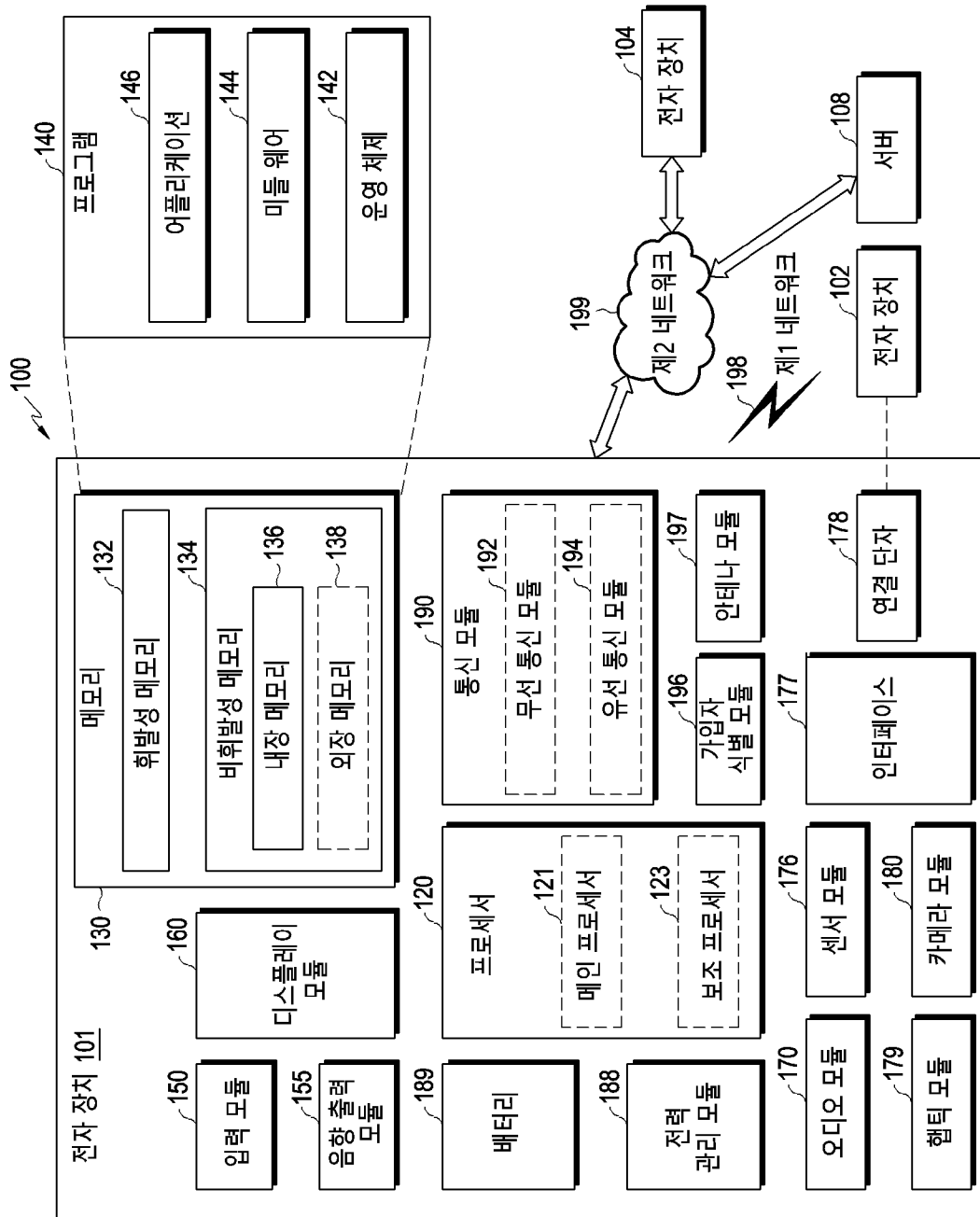
- [청구항 12] 제11항에 있어서,
 상기 적어도 하나의 노드를 수정하는 동작은,
 상기 적어도 하나의 노드와 관련된 가중치를 줄이는, 방법.

- [청구항 13] 제11항 또는 제12항에 있어서,
 상기 적어도 하나의 노드를 수정하는 동작은,
 상기 적어도 하나의 노드를 상기 오류 부분에 대응되는 텍스트에 대응되는 기저장된 적어도 하나의 노드로 교체하는, 방법.

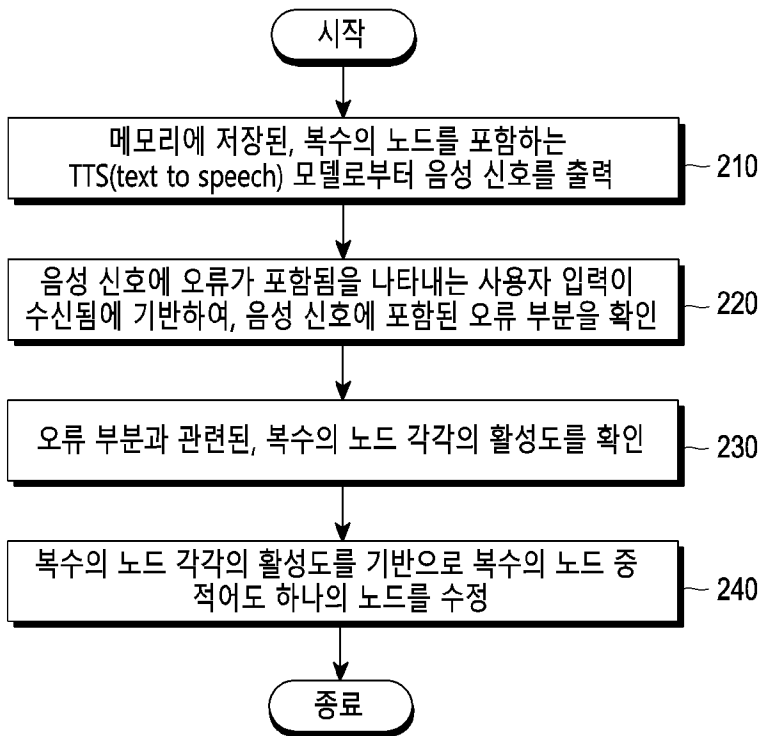
- [청구항 14] 제11항 내지 제13항 중 어느 한 항에 있어서,
 상기 오류 부분을 확인하는 동작은,
 상기 음성 신호가 설정된 길이 이상의 길이를 갖는 적어도 하나의 음소를 포함함을 확인함에 기반하여, 상기 적어도 하나의 음소에 대응되는 상기 음성 신호의 부분을 상기 오류 부분인 것으로 확인하는, 방법.

- [청구항 15] 하나 이상의 프로그램을 저장하는, 비일시적 컴퓨터 판독 가능 기록 매체에 있어서, 상기 하나 이상의 프로그램은, 전자 장치로 하여금:
 상기 전자 장치의 메모리에 저장된, 복수의 노드를 포함하는 TTS(text to speech) 모델에 기반하여 음성 신호를 획득하고, 상기 음성 신호는 입력 텍스트에 대응되는 것임,
 상기 음성 신호에 오류가 포함됨을 확인함에 기반하여, 상기 오류가 포함된 상기 음성 신호의 오류 부분을 확인하고,
 상기 오류 부분과 관련된, 상기 복수의 노드 각각의 활성도를 확인하고,
 적어도 하나의 노드의 상기 식별된 활성도를 기반으로 상기 복수의 노드 중 상기 적어도 하나의 노드와 관련된 가중치를 줄이는 명령어들을 포함하는, 기록 매체.

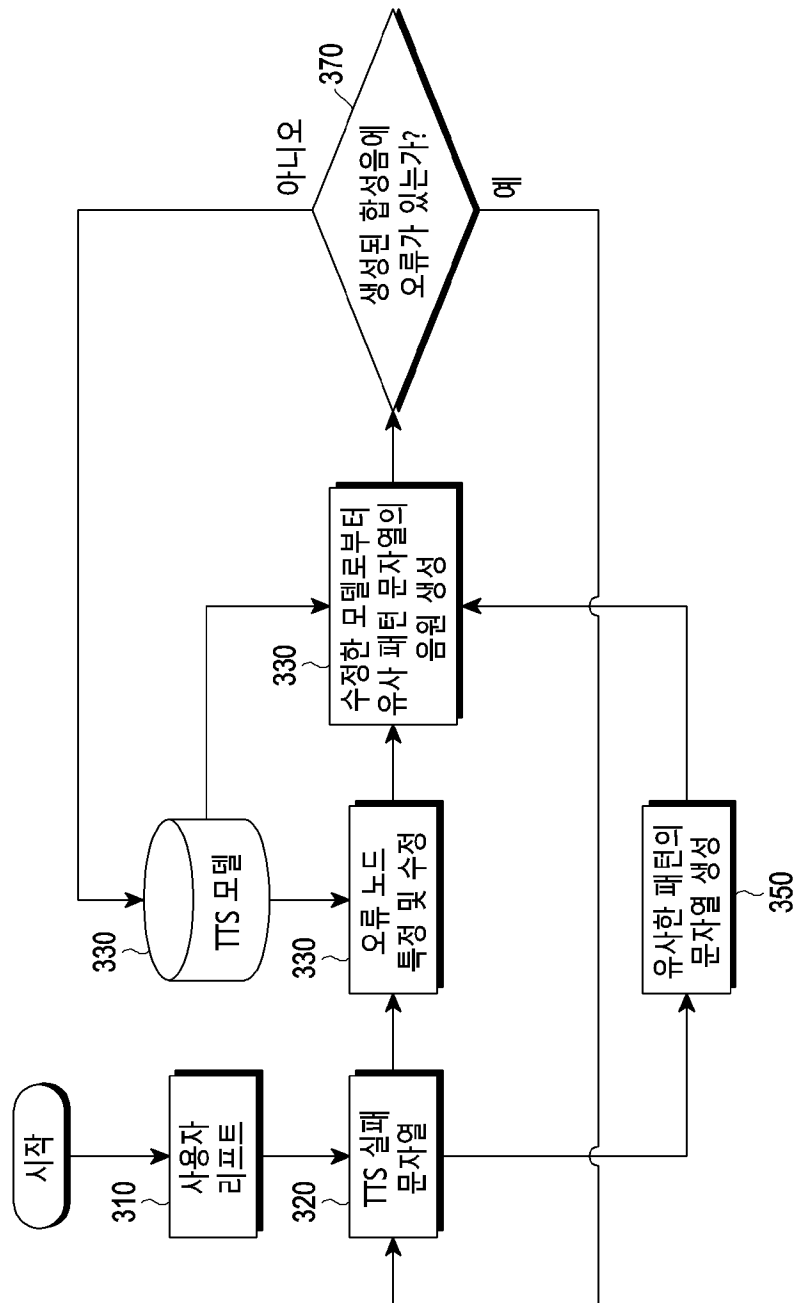
[도 1]



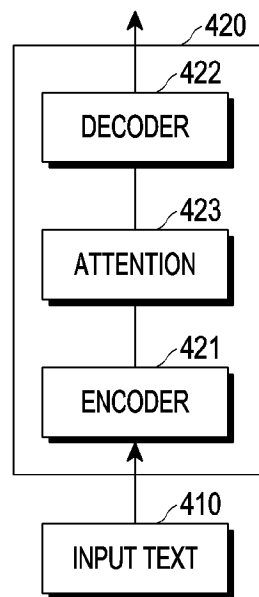
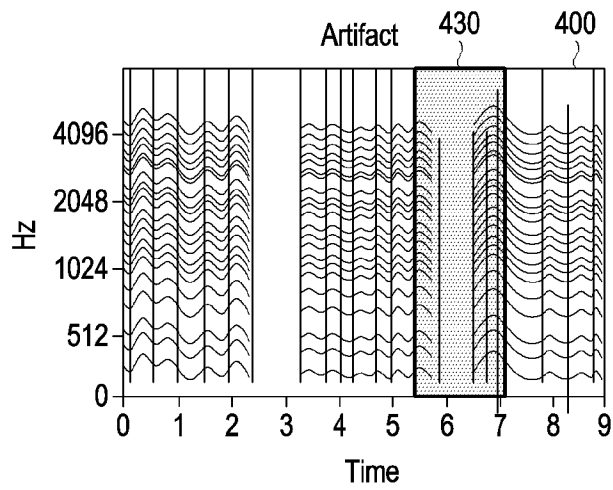
[도2]



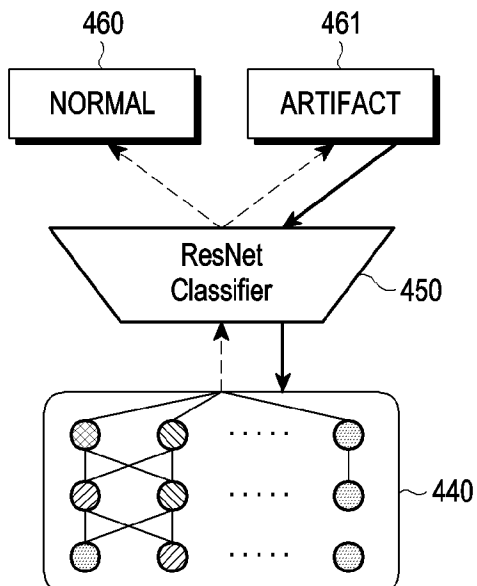
[도3]



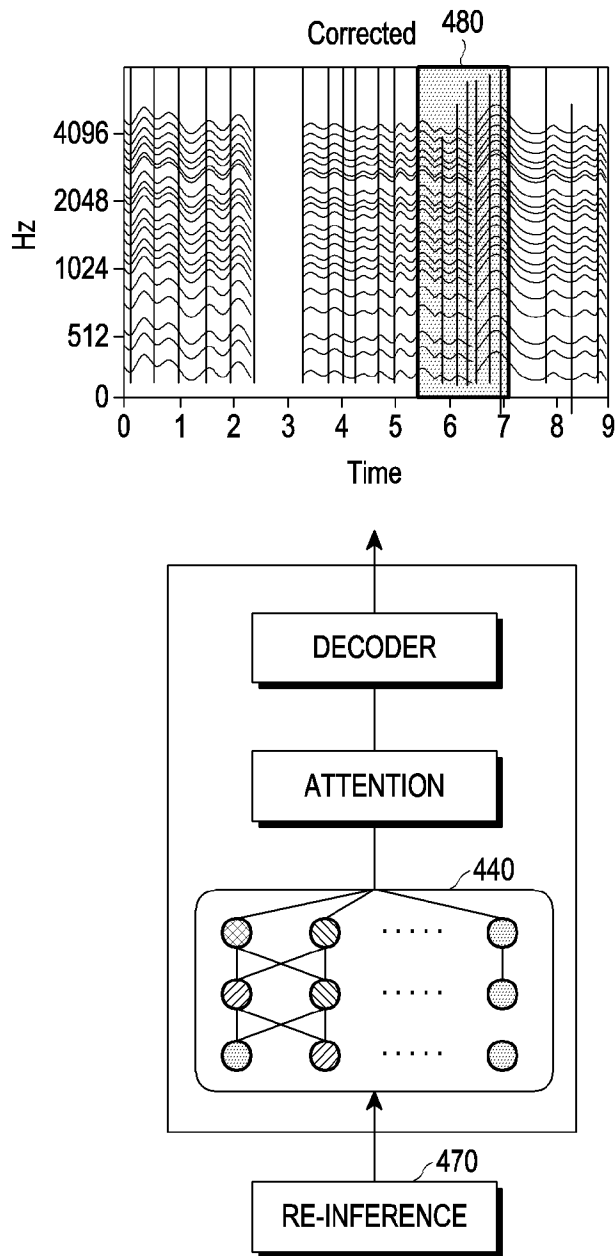
[도4a]



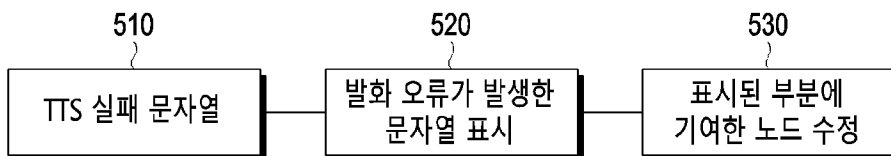
[도4b]



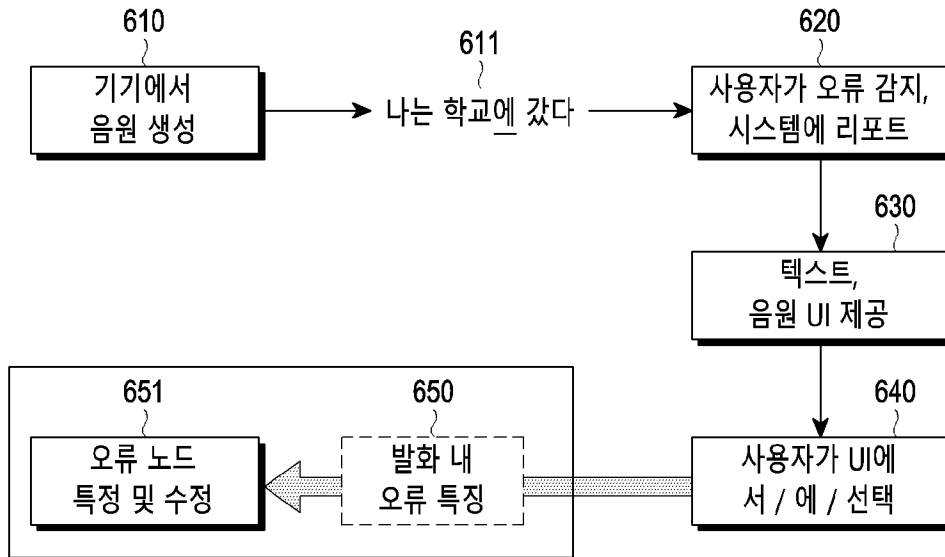
[도4c]



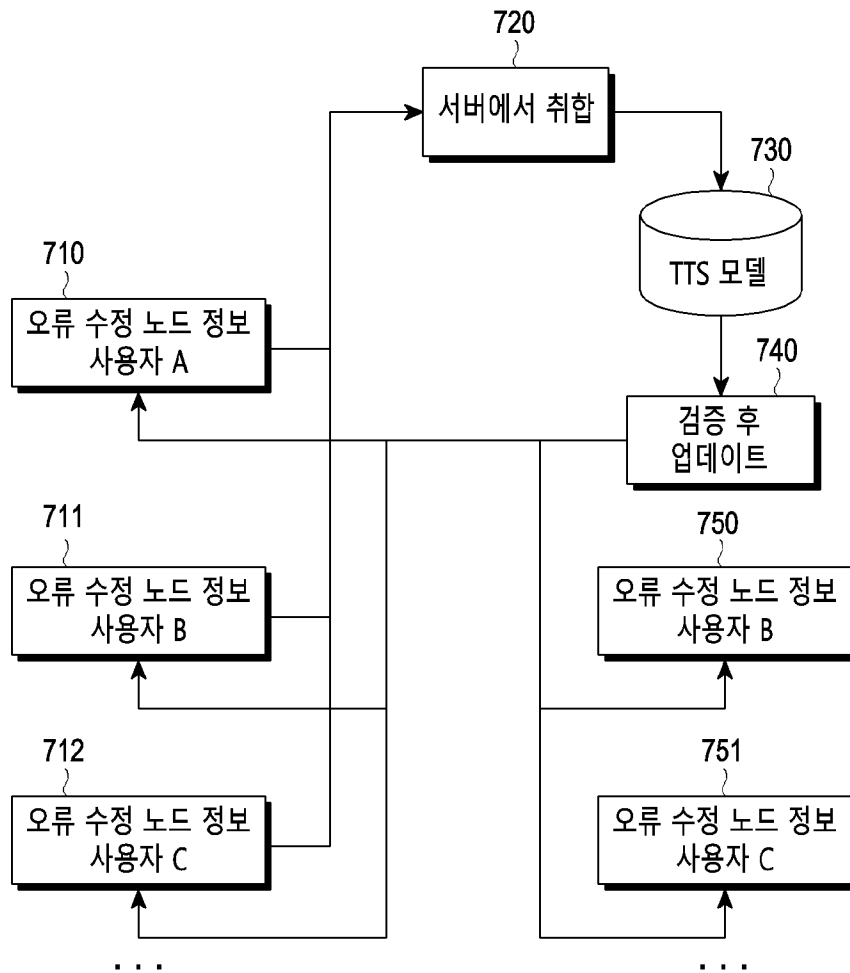
[도5]



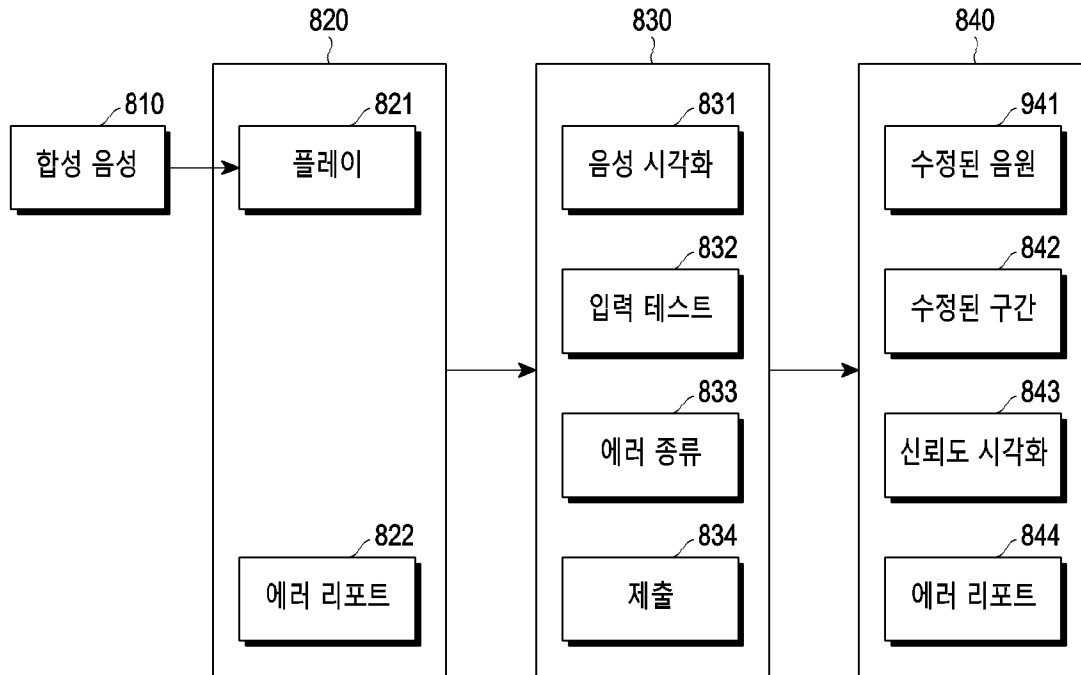
[도6]



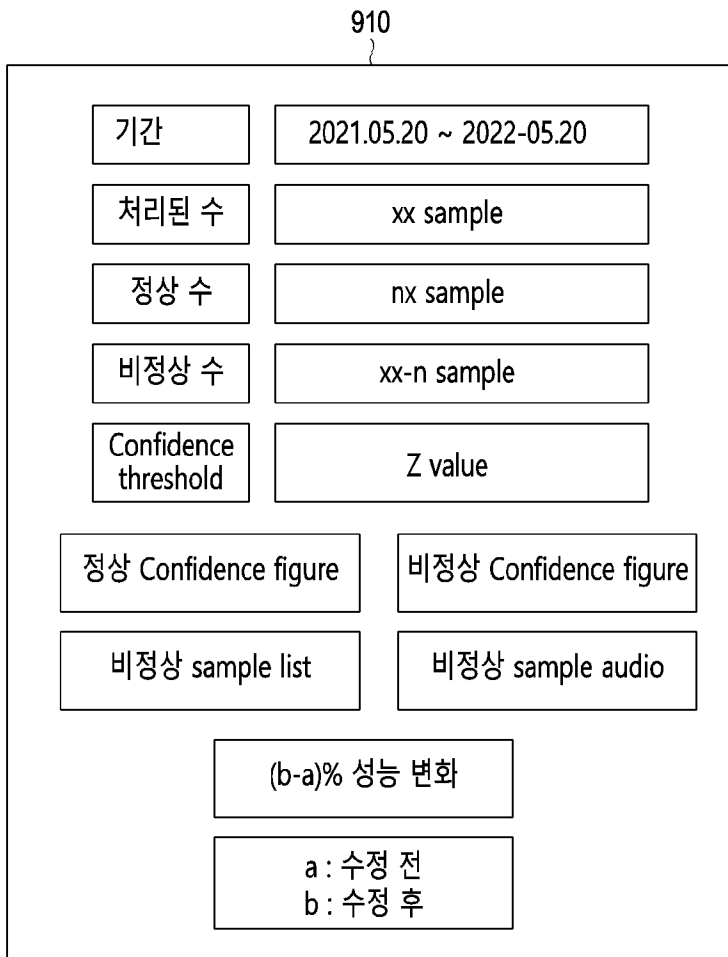
[도7]



[도8]



[도9]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2023/018345

A. CLASSIFICATION OF SUBJECT MATTER

G10L 13/08(2006.01)i; G10L 15/06(2006.01)i; G10L 25/15(2013.01)i; G10L 15/26(2006.01)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 13/08(2006.01); G06F 17/28(2006.01); G06K 9/64(2006.01); G06N 3/02(2006.01); G10L 15/06(2006.01);
G10L 15/16(2006.01); G10L 15/22(2006.01); G10L 15/26(2006.01); G10L 25/69(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: TTS(text to speech), 신경망(neural network), 노드(node), 가중치(weight), 오류(error), 활성도(relevance)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KR 10-2019-0101333 A (LG ELECTRONICS INC.) 30 August 2019 (2019-08-30) See paragraphs [0039]-[0040], [0085], [0127], [0130] and [0141]-[0142]; and claim 1.	1-15
Y	CN 103559537 B (NANJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 23 March 2016 (2016-03-23) See paragraph [0021]; and claim 1.	1-15
Y	KR 10-2386635 B1 (KAKAO ENTERPRISE CORP.) 14 April 2022 (2022-04-14) See paragraphs [0058] and [0077]; and claim 1.	4-5,7,14
Y	KR 10-2018-0039371 A (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 18 April 2018 (2018-04-18) See paragraphs [0051]-[0054]; and claim 1.	8-9
A	US 2021-0304769 A1 (MICROSOFT TECHNOLOGY LICENSING, LLC) 30 September 2021 (2021-09-30) See claim 1.	1-15



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

16 February 2024

Date of mailing of the international search report

16 February 2024

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/KR2023/018345

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
KR	10-2019-0101333	A	30 August 2019	US	2019-0385592	A1	19 December 2019
CN	103559537	B	23 March 2016	CN	103559537	A	05 February 2014
KR	10-2386635	B1	14 April 2022	KR	10-2021-0128255	A	26 October 2021
KR	10-2018-0039371	A	18 April 2018	None			
US	2021-0304769	A1	30 September 2021	CN	111410629	A	14 July 2020
				CN	111430567	A	17 July 2020
				CN	111467886	A	31 July 2020
				CN	111467886	B	19 November 2021
				CN	113470662	A	01 October 2021
				EP	4128214	A1	08 February 2023
				US	11587569	B2	21 February 2023
				US	11655231	B2	23 May 2023
				US	2021-0300892	A1	30 September 2021
				US	2023-0095257	A1	30 March 2023
				US	2023-0148089	A1	11 May 2023
				US	2023-0186919	A1	15 June 2023
				WO	2021-196374	A1	07 October 2021
				WO	2021-196460	A1	07 October 2021
				WO	2021-196654	A1	07 October 2021
				WO	2021-196843	A1	07 October 2021
				WO	2021-197377	A1	07 October 2021
				WO	2021-201970	A1	07 October 2021

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G10L 13/08(2006.01)i; G10L 15/06(2006.01)i; G10L 25/15(2013.01)i; G10L 15/26(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G10L 13/08(2006.01); G06F 17/28(2006.01); G06K 9/64(2006.01); G06N 3/02(2006.01); G10L 15/06(2006.01);
G10L 15/16(2006.01); G10L 15/22(2006.01); G10L 15/26(2006.01); G10L 25/69(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: TTS(text to speech), 신경망(neural network), 노드(node), 가중치
(weight), 오류(error), 활성화도(relevance)

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	KR 10-2019-0101333 A (엔지전자 주식회사) 2019.08.30 단락 [0039]-[0040], [0085], [0127], [0130], [0141]-[0142]; 및 청구항 1	1-15
Y	CN 103559537 B (NANJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 2016.03.23 단락 [0021]; 및 청구항 1	1-15
Y	KR 10-2386635 B1 (주식회사 카카오엔터프라이즈) 2022.04.14 단락 [0058], [0077]; 및 청구항 1	4-5,7,14
Y	KR 10-2018-0039371 A (한국전자통신연구원) 2018.04.18 단락 [0051]-[0054]; 및 청구항 1	8-9
A	US 2021-0304769 A1 (MICROSOFT TECHNOLOGY LICENSING, LLC) 2021.09.30 청구항 1	1-15

☐ 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

2024년02월16일(16.02.2024)

국제조사보고서 발송일

2024년02월16일(16.02.2024)

ISA/KR의 명칭 및 우편주소

대한민국 특허청
(35208) 대전광역시 서구 청사로 189, 4동 (둔산동,
정부대전청사)

팩스 번호 +82-42-481-8578

심사관

양정록

전화번호 +82-42-481-5709

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2019-0101333 A	2019/08/30	US 2019-0385592 A1	2019/12/19
CN 103559537 B	2016/03/23	CN 103559537 A	2014/02/05
KR 10-2386635 B1	2022/04/14	KR 10-2021-0128255 A	2021/10/26
KR 10-2018-0039371 A	2018/04/18	없음	
US 2021-0304769 A1	2021/09/30	CN 111410629 A	2020/07/14
		CN 111430567 A	2020/07/17
		CN 111467886 A	2020/07/31
		CN 111467886 B	2021/11/19
		CN 113470662 A	2021/10/01
		EP 4128214 A1	2023/02/08
		US 11587569 B2	2023/02/21
		US 11655231 B2	2023/05/23
		US 2021-0300892 A1	2021/09/30
		US 2023-0095257 A1	2023/03/30
		US 2023-0148089 A1	2023/05/11
		US 2023-0186919 A1	2023/06/15
		WO 2021-196374 A1	2021/10/07
		WO 2021-196460 A1	2021/10/07
		WO 2021-196654 A1	2021/10/07
		WO 2021-196843 A1	2021/10/07
		WO 2021-197377 A1	2021/10/07
		WO 2021-201970 A1	2021/10/07