

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국



(10) 국제공개번호

WO 2020/141643 A1

(43) 국제공개일
2020년 7월 9일 (09.07.2020)

- (51) 국제특허분류: *G10L 13/047* (2013.01) *G10L 15/30* (2013.01)
G10L 13/08 (2006.01) *H04L 9/08* (2006.01)
- (21) 국제출원번호: PCT/KR2019/000118
- (22) 국제출원일: 2019년 1월 3일 (03.01.2019)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (71) 출원인: 엘지전자 주식회사 (LG ELECTRONICS INC.) [KR/KR]; 07336 서울시 영등포구 여의대로 128, Seoul (KR).
- (72) 발명자: 윤정훈 (YOON, Jeonghun); 06772 서울시 서초구 양재대로11길 19 LG전자 특허센터, Seoul (KR). 채종훈 (CHAE, Jonghoon); 06772 서울시 서초구 양재대로 11길 19 LG전자 특허센터, Seoul (KR). 박용철 (PARK, Yongchul); 06772 서울시 서초구 양재대로11길 19 LG 전자 특허센터, Seoul (KR).
- (74) 대리인: 허용록 (HAW, Yong Noke); 06252 서울시 강남구 역삼로 114 현죽빌딩 6층, 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, CZ, DE, DJ, DK, DM, DO, DZ, EC,

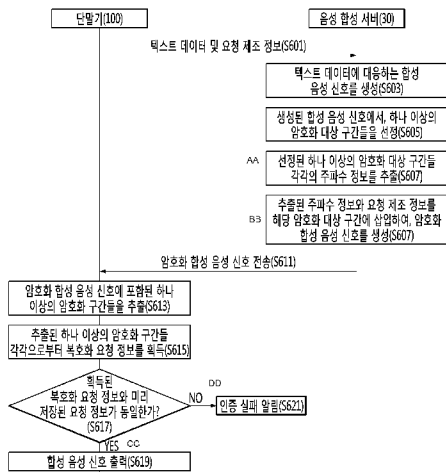
EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, 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))

(54) Title: VOICE SYNTHETIC SERVER AND TERMINAL

(54) 발명의 명칭: 음성 합성 서버 및 단말기



(57) Abstract: A voice synthetic server according to an embodiment of the present invention may comprise: a wireless communication unit for receiving text data and request production information from a terminal; and a processor for generating a synthetic voice signal corresponding to the text data, extracting frequency information from the generated synthetic voice signal, and inserting request information including the extracted frequency information and the request production information into the synthetic voice signal to generate an encoded synthetic voice signal.

(57) 요약서: 본 발명의 일 실시 예에 따른 음성 합성 서버는 단말기로부터 텍스트 데이터 및 요청 제조 정보를 수신하는 무선 통신부 및 상기 텍스트 데이터에 대응하는 합성 음성 신호를 생성하고, 생성된 합성 음성 신호로부터, 주파수 정보를 추출하고, 추출된 주파수 정보와 상기 요청 제조 정보를 포함하는 요청 정보를 상기 합성 음성 신호에 삽입하여, 암호화 합성 음성 신호를 생성하는 프로세서를 포함할 수 있다.

30 ... Voice synthetic server
 100 ... Terminal
 S601 ... Text data and request production information
 S603 ... Generate synthetic voice signal corresponding to text data
 S605 ... Select one or more encoding target sections from generated synthetic voice signal
 S607 ... Transmit encoded synthetic voice signal
 S611 ... Are obtained decoding request information and previously stored request information identical?
 S613 ... Extract one or more encoded sections included in encoded synthetic voice signal
 S615 ... Obtain decoding request information from each of one or more extracted encoded sections
 S617 ... Are obtained decoding request information and previously stored request information identical?
 S619 ... Output synthetic voice signal
 S621 ... Notify authentication failure
 AA ... Extract frequency information from each of one or more selected encoding target sections
 BB ... Insert extracted frequency information and request production information into corresponding encoding target section to generate encoded synthetic voice signal
 CC ... YES
 DD ... NO

WO 2020/141643 A1

명세서

발명의 명칭: 음성 합성 서버 및 단말기

기술분야

- [1] 본 발명은 음성 합성 서버에 관한 것으로, 보다 상세하게는 합성 음성을 암호화할 수 있는 음성 합성 서버에 관한 것이다.

배경기술

- [2] 스마트폰에 시작된 음성인식 기술 경쟁은 사물인터넷(IoT)의 본격 확산과 맞물려 이제 집 안에서 본격적으로 불붙을 전망이다.
- [3] 특히, 주목할 만 한 점은 그 기기가 음성을 매개로 명령을 내리고, 대화를 나눌 수도 있는 인공지능(AI) 기기라는 점이다.
- [4] 음성인식 서비스는 막대한 양의 데이터베이스를 활용하여, 사용자의 질문에 최적 답변을 선택하는 구조를 갖고 있다.
- [5] 음성검색 기능 역시 입력된 음성데이터를 클라우드 서버에서 텍스트로 변환하여 분석하고, 그 결과에 따른 실시간 검색결과를 기기로 재전송하는 방식이다.
- [6] 클라우드 서버는 수많은 단어들을 성별, 연령별, 억양별로 구분된 음성 데이터로 구분하여, 저장하고 실시간으로 처리할 수 있는 컴퓨팅 능력을 보유하고 있다.
- [7] 음성 인식은 더 많은 음성데이터가 축적될수록, 인간과 동등한(Human parity) 수준 정도로, 정확해 질 것이다.
- [8] 또한, 최근에는 특정 화자의 목소리로 합성된 합성 음성 인식 서비스를 제공하고 있다.
- [9] 그러나, 합성 음성 기술이 상업화 수준으로 발달함에 따라 합성 음성을 도용하거나, 사칭하는 등 악용하는 사례가 늘고 있다.
- [10] 기존에는 이러한, 합성 음성의 악용을 방지하기 위해, 합성 음성에 배경 음악을 넣거나, 합성 음성의 출력 전, 안내 멘트를 넣었다.
- [11] 그러나, 이러한 방식은 합성 음성의 출력에 방해가 되어, 깨끗한 합성 음성이 제공되지 못하는 문제가 있었다.

발명의 상세한 설명

기술적 과제

- [12] 본 발명은 전술한 문제 및 다른 문제를 해결하는 것을 목적으로 한다.
- [13] 본 발명은 배경 음악 또는 안내 멘트 없는, 깨끗한 합성 음성을 제공하는 것을 그 목적으로 한다.
- [14] 또한, 본 발명은 암호화된 정보를 합성 음성에 내재화하여, 합성 음성의 인증을 통해 합성 음성을 악용하는 것을 방지할 수 있도록 하는 것을 그 목적으로 한다.

과제 해결 수단

- [15] 본 발명의 일 실시 예에 따른 음성 합성 서버는 단말기로부터 텍스트 데이터 및 요청 제조 정보를 수신하는 무선 통신부 및 상기 텍스트 데이터에 대응하는 합성 음성 신호를 생성하고, 생성된 합성 음성 신호로부터, 주파수 정보를 추출하고, 추출된 주파수 정보와 상기 요청 제조 정보를 포함하는 요청 정보를 상기 합성 음성 신호에 삽입하여, 암호화 합성 음성 신호를 생성하는 프로세서를 포함할 수 있다.
- [16] 본 발명의 일 실시 예에 따른 단말기는 메모리와 음향 출력부와 음성 합성 서버로부터 암호화 합성 음성 신호를 수신하는 무선 통신부 및 상기 암호화 합성 음성 신호를 비밀 키 값을 이용하여, 복호화하여, 복호화된 합성 음성 신호 및 복호화된 요청 정보를 획득하고, 복호화된 요청 정보와 상기 메모리에 저장된 요청 정보를 비교하고, 비교 결과에 따라 상기 복호화된 합성 음성 신호의 위조 여부를 결정하는 프로세서를 포함할 수 있다.
- [17] 본 발명의 적용 가능성의 추가적인 범위는 이하의 상세한 설명으로부터 명백해질 것이다. 그러나 본 발명의 사상 및 범위 내에서 다양한 변경 및 수정은 당업자에게 명확하게 이해될 수 있으므로, 상세한 설명 및 본 발명의 바람직한 실시 예와 같은 특정 실시 예는 단지 예시로 주어진 것으로 이해되어야 한다.

발명의 효과

- [18] 본 발명의 실시 예에 따르면, 합성 음성이 위조된 것으로 판단한 경우, 합성 음성을 출력 하지 않아, 의도치 않은 합성 음성이 출력되어, 도용되는 상황이 미연에 방지될 수 있다.
- [19] 본 발명의 실시 예에 따르면, 합성 음성에 암호화된 정보의 일치 여부를 판단하여, 합성 음성을 인증함에 따라, 합성 음성을 이용한 보이스 피싱 범죄를 방지할 수 있다.

도면의 간단한 설명

- [20] 도 1은 본 발명과 관련된 단말기를 설명하기 위한 블록도이다.
- [21] 도 2는 본 발명의 일 실시 예에 따른 음성 시스템을 설명하기 위한 도면이다.
- [22] 도 2는 본 발명에 따른 변형 가능한 이동 단말기의 다른 예를 설명하기 위한 개념도이다.
- [23] 도 3은 본 발명의 일 실시 예에 따라 음성 신호로부터 사용자의 발화 특징을 추출하는 과정을 설명하는 도면이다.
- [24] 도 4는 본 발명의 일 실시 예에 따른 음성 신호를 파워 스펙트럼으로 변환한 예를 설명한 도면이다.
- [25] 도 5는 본 발명의 일 실시 예에 따른 음성 합성 서버의 구성을 설명하기 위한 블록도이다.
- [26] 도 6은 본 발명의 일 실시 예에 따른 합성 음성 암호화 시스템의 합성 음성 암호화 방법을 설명하기 위한 래더 다이어그램이다.
- [27] 도 7은 본 발명의 실시 예에 따라 텍스트 데이터를 합성 음성 신호로 변환하는

과정을 설명하는 도면이다.

- [28] 도 8은 합성 음성 신호로부터 하나 이상의 암호화 대상 구간들을 선정하는 과정을 설명하는 도면이다.
- [29] 도 9는 선정된 하나 이상의 암호화 대상 구간들 각각의 주파수 정보를 추출하는 과정을 설명하는 도면이다.
- [30] 도 10 내지 도 12는 본 발명의 일 실시 예에 따라 암호화 합성 음성 신호를 생성하는 과정을 설명하는 도면이다.
- [31] 도 13은 본 발명의 일 실시 예에 따른 복호화 테이블을 설명하는 도면이다.

발명의 실시를 위한 최선의 형태

- [32] 이하, 첨부된 도면을 참조하여 본 명세서에 개시된 실시 예를 상세히 설명하되, 도면 부호에 관계없이 동일하거나 유사한 구성요소는 동일한 참조 번호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다. 이하의 설명에서 사용되는 구성요소에 대한 접미사 "모듈" 및 "부"는 명세서 작성의 용이함만이 고려되어 부여되거나 혼용되는 것으로서, 그 자체로 서로 구별되는 의미 또는 역할을 갖는 것은 아니다. 또한, 본 명세서에 개시된 실시 예를 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 명세서에 개시된 실시 예의 요지를 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다. 또한, 첨부된 도면은 본 명세서에 개시된 실시 예를 쉽게 이해할 수 있도록 하기 위한 것일 뿐, 첨부된 도면에 의해 본 명세서에 개시된 기술적 사상이 제한되지 않으며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [33] 제1, 제2 등과 같이 서수를 포함하는 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되지는 않는다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다.
- [34] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.
- [35] 본 명세서에서 설명되는 단말기에는 휴대폰, 스마트 폰(smart phone), 노트북 컴퓨터(laptop computer), 디지털방송용 단말기, PDA(personal digital assistants), PMP(portable multimedia player), 네비게이션, 슬레이트 PC(slate PC), 태블릿 PC(tablet PC), 울트라북(ultrabook), 웨어러블 디바이스(wearable device, 예를 들어, 워치형 단말기 (smartwatch), 글래스형 단말기 (smart glass), HMD(head mounted display)) 등이 포함될 수 있다.

- [36] 그러나, 본 명세서에 기재된 실시 예에 따른 단말기(100)는 스마트 TV, 데스크탑 컴퓨터, 디지털사이니지 등과 같은 고정 단말기에도 적용될 수도 있다.
- [37] 또한, 본 발명의 실시 예에 따른 단말기(100)는 고정 또는 이동 가능한 로봇에도 적용될 수 있다.
- [38] 또한, 본 발명의 실시 예에 따른 단말기(100)는 음성 에이전트의 기능을 수행할 수 있다. 음성 에이전트는 사용자의 음성을 인식하고, 인식된 사용자의 음성에 적합한 응답을 음성으로 출력하는 프로그램일 수 있다.
- [39] 단말기(100)는 무선 통신부(110), 입력부(120), 러닝 프로세서(130), 센싱부(140), 출력부(150), 인터페이스부(160), 메모리(170), 프로세서(180) 및 전원 공급부(190)를 포함할 수 있다.
- [40] 무선 통신부(110)는, 방송 수신 모듈(111), 이동통신 모듈(112), 무선 인터넷 모듈(113), 근거리 통신 모듈(114), 위치정보 모듈(115) 중 적어도 하나를 포함할 수 있다.
- [41] 방송 수신 모듈(111)은 방송 채널을 통하여 외부의 방송 관리 서버로부터 방송 신호 및/또는 방송 관련된 정보를 수신한다.
- [42] 이동통신 모듈(112)은, 이동통신을 위한 기술표준들 또는 통신방식(예를 들어, GSM(Global System for Mobile communication), CDMA(Code Division Multi Access), CDMA2000(Code Division Multi Access 2000), EV-DO(Enhanced Voice-Data Optimized or Enhanced Voice-Data Only), WCDMA(Wideband CDMA), HSDPA(High Speed Downlink Packet Access), HSUPA(High Speed Uplink Packet Access), LTE(Long Term Evolution), LTE-A(Long Term Evolution-Advanced) 등)에 따라 구축된 이동 통신망 상에서 기지국, 외부의 단말, 서버 중 적어도 하나와 무선 신호를 송수신한다.
- [43] 무선 인터넷 모듈(113)은 무선 인터넷 접속을 위한 모듈을 말하는 것으로, 단말기(100)에 내장되거나 외장될 수 있다. 무선 인터넷 모듈(113)은 무선 인터넷 기술들에 따른 통신망에서 무선 신호를 송수신하도록 이루어진다.
- [44] 무선 인터넷 기술로는, 예를 들어 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) 등이 있다.
- [45] 근거리 통신 모듈(114)은 근거리 통신(Short range communication)을 위한 것으로서, 블루투스(Bluetooth™), RFID(Radio Frequency Identification), 적외선 통신(Infrared Data Association; IrDA), UWB(Ultra Wideband), ZigBee, NFC(Near Field Communication), Wi-Fi(Wireless-Fidelity), Wi-Fi Direct, Wireless USB(Wireless Universal Serial Bus) 기술 중 적어도 하나를 이용하여, 근거리 통신을 지원할 수 있다.

- [46] 위치정보 모듈(115)은 이동 단말기의 위치(또는 현재 위치)를 획득하기 위한 모듈로서, 그의 대표적인 예로는 GPS(Global Positioning System) 모듈 또는 WiFi(Wireless Fidelity) 모듈이 있다. 예를 들어, 단말기는 GPS모듈을 활용하면, GPS 위성에서 보내는 신호를 이용하여 이동 단말기의 위치를 획득할 수 있다.
- [47] 입력부(120)는 영상 신호 입력을 위한 카메라(121), 오디오 신호를 수신하기 위한 마이크로폰(122), 사용자로부터 정보를 입력 받기 위한 사용자 입력부(123)를 포함할 수 있다.
- [48] 입력부(120)에서 수집한 음성 데이터나 이미지 데이터는 분석되어 사용자의 제어명령으로 처리될 수 있다.
- [49] 입력부(120)는 영상 정보(또는 신호), 오디오 정보(또는 신호), 데이터, 또는 사용자로부터 입력되는 정보의 입력을 위한 것으로서, 영상 정보의 입력을 위하여, 단말기(100)는 하나 또는 복수의 카메라(121)들을 구비할 수 있다.
- [50] 카메라(121)는 화상 통화모드 또는 촬영 모드에서 이미지 센서에 의해 얻어지는 정지영상 또는 동영상 등의 화상 프레임을 처리한다. 처리된 화상 프레임은 디스플레이부(151)에 표시되거나 메모리(170)에 저장될 수 있다.
- [51] 마이크로폰(122)은 외부의 음향 신호를 전기적인 음성 데이터로 처리한다. 처리된 음성 데이터는 단말기(100)에서 수행 중인 기능(또는 실행 중인 응용 프로그램)에 따라 다양하게 활용될 수 있다. 한편, 마이크로폰(122)에는 외부의 음향 신호를 입력 받는 과정에서 발생하는 잡음(noise)을 제거하기 위한 다양한 잡음 제거 알고리즘이 구현될 수 있다.
- [52] 사용자 입력부(123)는 사용자로부터 정보를 입력 받기 위한 것으로서, 사용자 입력부(123)를 통해 정보가 입력되면,
- [53] 프로세서(180)는 입력된 정보에 대응되도록 단말기(100)의 동작을 제어할 수 있다.
- [54] 사용자 입력부(123)는 기계식 (mechanical) 입력수단(또는, 메커니컬 키, 예를 들어, 단말기(100)의 전/후면 또는 측면에 위치하는 버튼, 돔 스위치 (dome switch), 조그 휠, 조그 스위치 등) 및 터치식 입력수단을 포함할 수 있다. 일 예로서, 터치식 입력수단은, 소프트웨어적인 처리를 통해 터치스크린에 표시되는 가상 키(virtual key), 소프트 키(soft key) 또는 비주얼 키(visual key)로 이루어지거나, 상기 터치스크린 이외의 부분에 배치되는 터치 키(touch key)로 이루어질 수 있다.
- [55] 러닝 프로세서(130)는 데이터 마이닝, 데이터 분석, 지능형 의사 결정, 및 기계 학습 알고리즘 및 기술을 위해 이용될 정보를 수신, 분류, 저장 및 출력하도록 구성 될 수 있다.
- [56] 러닝 프로세서(130)는 단말기에 의해 수신, 검출, 감지, 생성, 사전 정의 또는 다른 방식으로 출력되거나 수신, 검출, 감지, 생성, 사전 정의 또는 다른 방식으로 다른 컴포넌트, 디바이스, 단말기 또는 단말기와 통신하는 장치에 의해 출력되는 데이터를 저장하도록 구성된 하나 이상의 메모리 유닛을 포함 할 수 있다.

- [57] 러닝 프로세서(130)는 단말기에 통합되거나 구현된 메모리를 포함 할 수 있다. 일부 실시 예에서, 러닝 프로세서(130)는 메모리(170)를 사용하여 구현 될 수 있다.
- [58] 선택적으로 또는 부가 적으로, 러닝 프로세서(130)는 단말기에 직접 결합된 외부 메모리 또는 단말기와 통신하는 서버에서 유지되는 메모리와 같이 단말기와 관련된 메모리를 사용하여 구현 될 수 있다.
- [59] 다른 실시 예에서, 러닝 프로세서(130)는 클라우드 컴퓨팅 환경에서 유지되는 메모리, 또는 네트워크와 같은 통신 방식을 통해 단말기에 의해 액세스 가능한 다른 원격 메모리 위치를 이용하여 구현 될 수 있다.
- [60] 러닝 프로세서(130)는 일반적으로 감독 또는 감독되지 않은 학습, 데이터 마이닝, 예측 분석 또는 다른 머신에서 사용하기 위해 데이터를 식별, 색인화, 카테고리화, 조작, 저장, 검색 및 출력하기 위해 데이터를 하나 이상의 데이터베이스에 저장하도록 구성될 수 있다.
- [61] 러닝 프로세서(130)에 저장된 정보는 다양한 상이한 유형의 데이터 분석 알고리즘 및 기계 학습 알고리즘 중 임의의 것을 사용하여 프로세서(180) 또는 단말기의 하나 이상의 다른 제어기에 의해 이용될 수 있다.
- [62] 이러한, 알고리즘의 예로는, k-최근 인접 시스템, 퍼지 논리 (예: 가능성 이론), 신경 회로망, 볼츠만 기계, 벡터 양자화, 펄스 신경망, 지원 벡터 기계, 최대 마진 분류기, 힐 클라이밍, 유도 논리 시스템 베이지안 네트워크, 페리트넷 (예: 유한 상태 머신, 밀리 머신, 무어 유한 상태 머신), 분류기 트리 (예: 퍼셉트론 트리, 지원 벡터 트리, 마코프 트리, 의사 결정 트리 포리스트, 임의의 포리스트), 판돈 모델 및 시스템, 인공 융합, 센서 융합, 이미지 융합, 보강 학습, 증강 현실, 패턴 인식, 자동화 된 계획 등을 포함한다.
- [63] 프로세서(180)는 데이터 분석 및 기계 학습 알고리즘을 사용하여 결정되거나, 생성된 정보에 기초하여 단말기의 적어도 하나의 실행 가능한 동작을 결정 또는 예측할 수 있다. 이를 위해, 프로세서(180)는 러닝 프로세서(130)의 데이터를 요청, 검색, 수신 또는 활용할 수 있고, 상기 적어도 하나의 실행 가능한 동작 중 예측되는 동작이나, 바람직한 것으로 판단되는 동작을 실행하도록 상기 단말기를 제어할 수 있다.
- [64] 프로세서(180)는 지능적 애플리케이션(즉, 지식 기반 시스템, 추론 시스템 및 지식 획득 시스템)을 구현하는 다양한 기능을 수행 할 수 있다. 이는 적응 시스템, 기계 학습 시스템, 인공 신경망 등을 포함하는, 다양한 유형의 시스템(예컨대, 퍼지 논리 시스템)에 적용될 수 있다.
- [65] 프로세서(180)는, 또한 I/O 처리 모듈, 환경 조건 모듈, 음성 - 텍스트 (STT) 처리 모듈, 자연어 처리 모듈, 작업 흐름 처리 모듈 및 서비스 처리 모듈과 같이, 음성 및 자연 언어 음성 처리를 수반하는 연산을 가능하게 하는 서브 모듈을 포함할 수 있다.
- [66] 이들 서브 모듈들 각각은, 단말기에서의 하나 이상의 시스템 또는 데이터 및

- 모델, 또는 이들의 서브셋 또는 슈퍼 셋에 대한 액세스를 가질 수 있다. 또한, 이들 서브 모듈들 각각은, 어휘 색인, 사용자 데이터, 작업 흐름 모델, 서비스 모델 및 자동 음성 인식 (ASR) 시스템을 비롯한 다양한 기능을 제공할 수 있다.
- [67] 다른 실시 예에서, 프로세서(180) 또는 단말기의 다른 양태는 상기 서브 모듈, 시스템, 또는 데이터 및 모델로 구현 될 수 있다.
- [68] 일부 예에서, 러닝 프로세서(130)의 데이터에 기초하여, 프로세서(180)는 사용자 입력 또는 자연 언어 입력으로 표현된 문맥 조건 또는 사용자의 의도에 기초하여 요구 사항을 검출하고 감지하도록 구성 될 수 있다.
- [69] 프로세서(180)는 문맥 조건 또는 사용자의 의도에 기초하여 요구 사항을 완전히 결정하는데 필요한 정보를 능동적으로 이끌어 내고, 획득할 수 있다. 예를 들어, 프로세서(180)는 역사적 입력 및 출력, 패턴 매칭, 모호하지 않은 단어, 입력 의도 등을 포함하는 과거 데이터를 분석함으로써 요구 사항을 결정하는데, 필요한 정보를 능동적으로 이끌어낼 수 있다.
- [70] 프로세서(180)는 문맥 조건 또는 사용자의 의도에 기초하여 요구 사항에 응답하는 기능을 실행하기 위한 태스크 흐름을 결정할 수 있다.
- [71] 프로세서(180)는 러닝 프로세서(130)에서 프로세싱 및 저장을 위한 정보를 수집하기 위해, 단말기에서 하나 이상의 감지 컴포넌트를 통해 데이터 분석 및 기계 학습 작업에 사용되는 신호 또는 데이터를 수집, 감지, 추출, 검출 및/또는 수신하도록 구성 될 수 있다.
- [72] 정보 수집은 센서를 통해 정보를 감지하는 것, 메모리(170)에 저장된 정보를 추출하는 것 또는 통신 수단을 통해 다른 단말기, 엔티티 또는 외부 저장 장치로부터 정보를 수신하는 것을 포함 할 수 있다.
- [73] 프로세서(180)는 단말기에서 사용 히스토리 정보를 수집하여, 저장할 수 있다.
- [74] 프로세서(180)는 저장된 사용 히스토리 정보 및 예측 모델링을 사용하여 특정 기능을 실행하기 위한 최상의 매치를 결정할 수 있다.
- [75] 프로세서(180)는 센싱부(140)를 통해 주변 환경 정보 또는 기타 정보를 수신하거나 감지 할 수 있다.
- [76] 프로세서(180)는 무선 통신부(110)을 통해 방송 신호 및/또는 방송 관련 정보, 무선 신호, 무선 데이터를 수신할 수 있다.
- [77] 프로세서(180)는 입력부(120)로부터 이미지 정보 (또는 해당 신호), 오디오 정보 (또는 해당 신호), 데이터 또는 사용자 입력 정보를 수신 할 수 있다.
- [78] 프로세서(180)는 정보를 실시간으로 수집하고, 정보 (예를 들어, 지식 그래프, 명령 정책, 개인화 데이터베이스, 대화 엔진 등)를 처리 또는 분류하고, 처리 된 정보를 메모리(170) 또는 러닝 프로세서(130)에 저장할 수 있다.
- [79] 단말기의 동작이 데이터 분석 및 기계 학습 알고리즘 및 기술에 기초하여 결정될 때, 프로세서(180)는 결정된 동작을 실행하기 위해 단말기의 구성 요소를 제어 할 수 있다. 그리고 프로세서(180)는 제어 명령에 따라 단말을 제어하여 결정된 동작을 수행 할 수 있다.

- [80] 프로세서(180)는 특정 동작이 수행되는 경우, 데이터 분석 및 기계 학습 알고리즘 및 기법을 통해 특정 동작의 실행을 나타내는 이력 정보를 분석하고, 분석된 정보에 기초하여 이전에 학습 한 정보의 업데이트를 수행 할 수 있다.
- [81] 따라서, 프로세서(180)는 러닝 프로세서(130)과 함께, 업데이트 된 정보에 기초하여 데이터 분석 및 기계 학습 알고리즘 및 기법의 미래 성능의 정확성을 향상시킬 수 있다.
- [82] 센싱부(140)는 이동 단말기 내 정보, 이동 단말기를 둘러싼 주변 환경 정보 및 사용자 정보 중 적어도 하나를 센싱 하기 위한 하나 이상의 센서를 포함할 수 있다.
- [83] 예를 들어, 센싱부(140)는 근접센서(141, proximity sensor), 조도 센서(142, illumination sensor), 터치 센서(touch sensor), 가속도 센서(acceleration sensor), 자기 센서(magnetic sensor), 중력 센서(G-sensor), 자이로스코프 센서(gyroscope sensor), 모션 센서(motion sensor), RGB 센서, 적외선 센서(IR 센서: infrared sensor), 지문인식 센서(finger scan sensor), 초음파 센서(ultrasonic sensor), 광 센서(optical sensor, 예를 들어, 카메라(121 참조)), 마이크로폰(microphone, 122 참조), 배터리 게이지(battery gauge), 환경 센서(예를 들어, 기압계, 습도계, 온도계, 방사능 감지 센서, 열 감지 센서, 가스 감지 센서 등), 화학 센서(예를 들어, 전자 코, 헬스케어 센서, 생체 인식 센서 등) 중 적어도 하나를 포함할 수 있다. 한편, 본 명세서에 개시된 이동 단말기는, 이러한 센서들 중 적어도 둘 이상의 센서에서 센싱되는 정보들을 조합하여 활용할 수 있다.
- [84] 출력부(150)는 시각, 청각 또는 촉각 등과 관련된 출력을 발생시키기 위한 것으로, 디스플레이부(151), 음향 출력부(152), 햅틱 모듈(153), 광 출력부(154) 중 적어도 하나를 포함할 수 있다.
- [85] 디스플레이부(151)는 단말기(100)에서 처리되는 정보를 표시(출력)한다. 예를 들어, 디스플레이부(151)는 단말기(100)에서 구동되는 응용 프로그램의 실행화면 정보, 또는 이러한 실행화면 정보에 따른 UI(User Interface), GUI(Graphic User Interface) 정보를 표시할 수 있다.
- [86] 디스플레이부(151)는 터치 센서와 상호 레이어 구조를 이루거나 일체형으로 형성됨으로써, 터치 스크린을 구현할 수 있다. 이러한 터치 스크린은, 단말기(100)와 사용자 사이의 입력 인터페이스를 제공하는 사용자 입력부(123)로써 기능함과 동시에, 단말기(100)와 사용자 사이의 출력 인터페이스를 제공할 수 있다.
- [87] 음향 출력부(152)는 호신호 수신, 통화모드 또는 녹음 모드, 음성인식 모드, 방송수신 모드 등에서 무선 통신부(110)로부터 수신되거나 메모리(170)에 저장된 오디오 데이터를 출력할 수 있다.
- [88] 음향 출력부(152)는 리시버(receiver), 스피커(speaker), 버저(buzzer) 중 적어도 하나 이상을 포함할 수 있다.
- [89] 햅틱 모듈(haptic module)(153)은 사용자가 느낄 수 있는 다양한 촉각 효과를

- 발생시킨다. 햅틱 모듈(153)이 발생시키는 촉각 효과의 대표적인 예로는 진동이 될 수 있다.
- [90] 광출력부(154)는 단말기(100)의 광원의 빛을 이용하여 이벤트 발생을 알리기 위한 신호를 출력한다. 단말기(100)에서 발생 되는 이벤트의 예로는 메시지 수신, 호 신호 수신, 부재중 전화, 알람, 일정 알림, 이메일 수신, 애플리케이션을 통한 정보 수신 등이 될 수 있다.
- [91] 인터페이스부(160)는 단말기(100)에 연결되는 다양한 종류의 외부 기기와의 통로 역할을 수행한다. 이러한 인터페이스부(160)는, 유/무선 헤드셋 포트(port), 외부 충전기 포트(port), 유/무선 데이터 포트(port), 메모리 카드(memory card) 포트, 식별 모듈이 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. 단말기(100)에서는, 상기 인터페이스부(160)에 외부 기기가 연결되는 것에 대응하여, 연결된 외부 기기와 관련된 적절할 제어를 수행할 수 있다.
- [92] 한편, 식별 모듈은 단말기(100)의 사용 권한을 인증하기 위한 각종 정보를 저장한 칩으로서, 사용자 인증 모듈(user identify module; UIM), 가입자 인증 모듈(subscriber identity module; SIM), 범용 사용자 인증 모듈(universal subscriber identity module; USIM) 등을 포함할 수 있다. 식별 모듈이 구비된 장치(이하 '식별 장치')는, 스마트 카드(smart card) 형식으로 제작될 수 있다. 따라서 식별 장치는 상기 인터페이스부(160)를 통하여 단말기(100)와 연결될 수 있다.
- [93] 메모리(170)는 단말기(100)의 다양한 기능을 지원하는 데이터를 저장한다.
- [94] 메모리(170)는 단말기(100)에서 구동되는 다수의 응용 프로그램(application program 또는 애플리케이션(application)), 단말기(100)의 동작을 위한 데이터들, 명령어들을, 러닝 프로세서(130)의 동작을 위한 데이터들(예를 들어, 머신러닝을 위한 적어도 하나의 알고리즘 정보 등)을 저장할 수 있다.
- [95] 프로세서(180)는 상기 응용 프로그램과 관련된 동작 외에도, 통상적으로 단말기(100)의 전반적인 동작을 제어한다. 프로세서(180)는 위에서 살펴본 구성요소들을 통해 입력 또는 출력되는 신호, 데이터, 정보 등을 처리하거나 메모리(170)에 저장된 응용 프로그램을 구동함으로써, 사용자에게 적절한 정보 또는 기능을 제공 또는 처리할 수 있다.
- [96] 또한, 프로세서(180)는 메모리(170)에 저장된 응용 프로그램을 구동하기 위하여, 도 1와 함께 살펴본 구성요소들 중 적어도 일부를 제어할 수 있다. 나아가, 프로세서(180)는 상기 응용프로그램의 구동을 위하여, 단말기(100)에 포함된 구성요소들 중 적어도 둘 이상을 서로 조합하여 동작시킬 수 있다.
- [97] 전원공급부(190)는 프로세서(180)의 제어 하에서, 외부의 전원, 내부의 전원을 인가 받아 단말기(100)에 포함된 각 구성요소들에 전원을 공급한다. 이러한 전원공급부(190)는 배터리를 포함하며, 상기 배터리는 내장형 배터리 또는 교체 가능한 형태의 배터리가 될 수 있다.

- [98] 한편, 앞서 살펴본 것과 같이, 프로세서(180)는 응용 프로그램과 관련된 동작과, 통상적으로 단말기(100)의 전반적인 동작을 제어한다. 예를 들어, 프로세서(180)는 상기 이동 단말기의 상태가 설정된 조건을 만족하면, 애플리케이션들에 대한 사용자의 제어 명령의 입력을 제한하는 잠금 상태를 실행하거나, 해제할 수 있다.
- [99] 도 2는 본 발명의 일 실시 예에 따른 음성 시스템을 설명하기 위한 도면이다.
- [100] 도 2를 참조하면, 음성 시스템(1)은 단말기(100), 음성 텍스트 변환(Speech To Text, STT) 서버(10), 자연어 처리(Natural Language Processing, NLP) 서버(20) 및 음성 합성 서버(30)를 포함할 수 있다.
- [101] 단말기(100)는 음성 데이터를 STT 서버(10)에 전송할 수 있다.
- [102] STT 서버(10)는 단말기(100)로부터 수신된 음성 데이터를 텍스트 데이터로 변환할 수 있다.
- [103] STT 서버(10)는 언어 모델을 이용하여 음성-텍스트 변환의 정확도를 높일 수 있다.
- [104] 언어 모델은 문장의 확률을 계산하거나, 이전의 단어들에 주어졌을 때 다음 단어가 나올 확률을 계산할 수 있는 모델을 의미할 수 있다.
- [105] 예컨대, 언어 모델은 유니그램(Unigram) 모델, 바이그램(Bigram) 모델, N-그램(N-gram) 모델 등과 같은 확률론적 언어 모델들을 포함할 수 있다.
- [106] 유니그램 모델은 모든 단어의 활용이 완전히 서로 독립적이라고 가정하는 모델로, 단어 열의 확률을 각 단어의 확률의 곱으로 계산하는 모델이다.
- [107] 바이그램 모델은 단어의 활용이 이전 1개의 단어에만 의존한다고 가정하는 모델이다.
- [108] N-그램 모델은 단어의 활용이 이전 (n-1)개의 단어에 의존한다고 가정하는 모델이다.
- [109] 즉, STT 서버(10)는 언어 모델을 이용하여 음성 데이터로부터 변환된 텍스트 데이터가 적합하게 변환된 것인지 판단할 수 있고, 이를 통해 텍스트 데이터로의 변환의 정확도를 높일 수 있다.
- [110] NLP 서버(20)는 STT 서버(10)로부터 텍스트 데이터를 수신할 수 있다. NLP 서버(20)는 수신된 텍스트 데이터에 기초하여, 텍스트 데이터에 대한 의도 분석을 수행할 수 있다.
- [111] NLP 서버(20)는 의도 분석의 수행 결과를 나타내는 의도 분석 정보를 단말기(100)에 전송할 수 있다.
- [112] NLP 서버(20)는 텍스트 데이터에 대해, 형태소 분석 단계, 구문 분석 단계, 화행 분석 단계, 대화 처리 단계를 순차적으로, 수행하여, 의도 분석 정보를 생성할 수 있다.
- [113] 형태소 분석 단계는 사용자가 발화한 음성에 대응하는 텍스트 데이터를 의미를 지닌 가장 작은 단위인 형태소 단위로 분류하고, 분류된 각 형태소가 어떤 품사를 가지는지를 결정하는 단계이다.

- [114] 구문 분석 단계는 형태소 분석 단계의 결과를 이용하여, 텍스트 데이터를 명사구, 동사구, 형용사 구 등으로 구분하고, 구분된 각 구들 사이에, 어떠한 관계가 존재하는지를 결정하는 단계이다.
- [115] 구문 분석 단계를 통해, 사용자가 발화한 음성의 주어, 목적어, 수식어들이 결정될 수 있다.
- [116] 화행 분석 단계는 구문 분석 단계의 결과를 이용하여, 사용자가 발화한 음성에 대한 의도를 분석하는 단계이다. 구체적으로, 화행 분석 단계는 사용자가 질문을 하는 것인지, 요청을 하는 것인지, 단순한 감정 표현을 하는 것인지와 같은 문장의 의도를 결정하는 단계이다.
- [117] 대화 처리 단계는 화행 분석 단계의 결과를 이용하여, 사용자의 발화에 대해 대답을 할지, 호응을 할지, 추가 정보를 문의하는 질문을 할지를 판단하는 단계이다.
- [118] NLP 서버(20)는 대화 처리 단계 후, 사용자가 발화한 의도에 대한 답변, 호응, 추가 정보 문의 중 하나 이상을 포함하는 의도 분석 정보를 생성할 수 있다.
- [119] 한편, NLP 서버(20)는 단말기(100)로부터 텍스트 데이터를 수신할 수도 있다. 예를 들어, 단말기(100)가 음성 텍스트 변환 기능을 지원하는 경우, 단말기(100)는 음성 데이터를 텍스트 데이터로 변환하고, 변환된 텍스트 데이터를 NLP 서버(20)에 전송할 수 있다.
- [120] 음성 합성 서버(30)는 기 저장된 음성 데이터들을 조합하여, 합성 음성을 생성할 수 있다.
- [121] 음성 합성 서버(30)는 모델로 선정된 한 사람의 음성을 녹음하고, 녹음된 음성을 음절 또는 단어 단위로 분할할 수 있다. 음성 합성 서버(30)는 음절 또는 단어 단위로, 분할된 음성을 내부 또는 외부의 데이터 베이스에 저장할 수 있다.
- [122] 음성 합성 서버(30)는 주어진 텍스트 데이터에 대응하는 음절 또는 단어를 데이터 베이스로부터 검색하고, 검색된 음절 또는 단어들의 조합을 합성하여, 합성 음성을 생성할 수 있다.
- [123] 음성 합성 서버(30)는 복수의 언어들 각각에 대응하는 복수의 음성 언어 그룹들을 저장하고 있을 수 있다.
- [124] 예를 들어, 음성 합성 서버(30)는 한국어로 녹음된 제1 음성 언어 그룹, 영어로, 녹음된 제2 음성 언어 그룹을 포함할 수 있다.
- [125] 음성 합성 서버(30)는 제1 언어의 텍스트 데이터를 제2 언어의 텍스트로 번역하고, 제2 음성 언어 그룹을 이용하여, 번역된 제2 언어의 텍스트에 대응하는 합성 음성을 생성할 수 있다.
- [126] 음성 합성 서버(30)는 생성된 합성 음성을 단말기(100)에 전송할 수 있다.
- [127] 음성 합성 서버(30)는 NLP 서버(20)로부터 의도 분석 정보를 수신할 수 있다.
- [128] 음성 합성 서버(30)는 의도 분석 정보에 기초하여, 사용자의 의도를 반영한, 합성 음성을 생성할 수 있다.
- [129] 일 실시 예에서, STT 서버(10), NLP 서버(20) 및 음성 합성 서버(30)는 하나의

서버로 구현될 수 있다.

- [130] 위에서, 설명된 STT 서버(10), NLP 서버(20) 및 음성 합성 서버(30) 각각의 기능은 단말기(100)에서도 수행될 수도 있다. 이를 위해, 단말기(100)는 복수의 프로세서들을 포함할 수 있다.
- [131] 도 3은 본 발명의 일 실시 예에 따라 음성 신호로부터 사용자의 발화 특징을 추출하는 과정을 설명하는 도면이다.
- [132] 도 1에 도시된, 단말기(100)는 오디오 프로세서(181)를 더 포함할 수 있다.
- [133] 오디오 프로세서(181)는 프로세서(180)와 별도의 칩으로 구현되거나, 프로세서(180)에 포함된 칩으로 구현될 수 있다.
- [134] 오디오 프로세서(181)는 음성 신호로부터, 잡음을 제거할 수 있다.
- [135] 오디오 프로세서(181)는 음성 신호를 텍스트 데이터로 변환할 수 있다. 이를 위해, 오디오 프로세서(181)는 STT 엔진을 구비할 수 있다.
- [136] 오디오 프로세서(181)는 단말기(100)의 음성 인식을 활성화시키기 위한 기동어를 인식할 수 있다. 오디오 프로세서(181)는 마이크로폰(121)을 통해 수신된 기동어를 텍스트 데이터로 변환하고, 변환된 텍스트 데이터가 기 저장된 기동어에 대응하는 텍스트 데이터인 경우, 기동어를 인식한 것으로 판단할 수 있다.
- [137] 오디오 프로세서(181)는 잡음이 제거된 음성 신호를 파워 스펙트럼으로 변환할 수 있다.
- [138] 파워 스펙트럼은 시간적으로 변동하는 음성 신호의 파형에 어떠한 주파수 성분이 어떠한 크기로 포함되어 있는지를 나타내는 파라미터일 수 있다.
- [139] 파워 스펙트럼은 음성 신호의 파형의 주파수에 따른 진폭 제곱 값의 분포를 보여준다.
- [140] 이에 대해서는 도 4를 참조하여, 설명한다.
- [141] 도 4는 본 발명의 일 실시 예에 따른 음성 신호를 파워 스펙트럼으로 변환한 예를 설명한 도면이다.
- [142] 도 4을 참조하면, 음성 신호(410)가 도시되어 있다. 음성 신호(410)는 마이크로폰(121)을 통해 수신되거나, 메모리(170)에 미리 저장된 신호일 수 있다.
- [143] 음성 신호(410)의 x축은 시간이고, y축은 진폭의 크기를 나타낼 수 있다.
- [144] 오디오 프로세서(181)는 x축이 시간 축인 음성 신호(410)를 x축이 주파수 축인 파워 스펙트럼(430)으로 변환할 수 있다.
- [145] 오디오 프로세서(181)는 고속 푸리에 변환(Fast Fourier Transform, FFT)을 이용하여, 음성 신호(410)를 파워 스펙트럼(430)으로 변환할 수 있다.
- [146] 파워 스펙트럼(430)의 x축은 주파수, y축은 진폭의 제곱 값을 나타낸다.
- [147] 다시 도 3을 설명한다.
- [148] 프로세서(180)는 오디오 프로세서(181)로부터 전달된 텍스트 데이터 또는 파워 스펙트럼(430) 중 하나 이상을 이용하여, 사용자의 발화 특징을 결정할 수 있다.
- [149] 사용자의 발화 특징은 사용자의 성별, 사용자의 음의 높낮이, 사용자의 음색,

- 사용자의 발화 주제, 사용자의 발화 속도, 사용자의 성량등을 포함할 수 있다.
- [150] 프로세서(180)는 파워 스펙트럼(430)을 이용하여, 음성 신호(410)의 주파수 및 주파수에 대응하는 진폭을 획득할 수 있다.
- [151] 프로세서(180)는 파워 스펙트럼(430)의 주파수 대역을 이용하여, 음성을 발화한 사용자의 성별을 결정할 수 있다.
- [152] 예를 들어, 프로세서(180)는 파워 스펙트럼(430)의 주파수 대역이 기 설정된 제1 주파수 대역 범위 내인 경우, 사용자의 성별을 남자로 결정할 수 있다.
- [153] 프로세서(180)는 파워 스펙트럼(430)의 주파수 대역이 기 설정된 제2 주파수 대역 범위 내인 경우, 사용자의 성별을 여자로 결정할 수 있다. 여기서, 제2 주파수 대역 범위는 제1 주파수 대역 범위보다 클 수 있다.
- [154] 프로세서(180)는 파워 스펙트럼(430)의 주파수 대역을 이용하여, 음성의 높낮이를 결정할 수 있다.
- [155] 예를 들어, 프로세서(180)는 특정 주파수 대역 범위 내에서, 진폭의 크기에 따라 음의 높낮이 정도를 결정할 수 있다.
- [156] 프로세서(180)는 파워 스펙트럼(430)의 주파수 대역을 이용하여, 사용자의 음색(tone)을 결정할 수 있다. 예를 들어, 프로세서(180)는 파워 스펙트럼(430)의 주파수 대역들 중, 진폭의 크기가 일정 크기 이상인 주파수 대역을 사용자의 주요 음역대로 결정하고, 결정된 주요 음역대를 사용자의 음색으로 결정할 수 있다.
- [157] 프로세서(180)는 변환된 텍스트 데이터로부터, 단위 시간 당 발화된 음절 수를 통해, 사용자의 발화 속도를 결정할 수 있다.
- [158] 프로세서(180)는 변환된 텍스트 데이터에 대해, Bag-Of-Word Model 기법을 이용하여, 사용자의 발화 주제를 결정할 수 있다.
- [159] Bag-Of-Word Model 기법은 문장 내 단어 빈도 수 기반으로, 주로 사용하는 단어를 추출하는 기법이다. 구체적으로, Bag-Of-Word Model 기법은 문장 내에서, 유니크한 단어를 추출하고, 추출된 각 단어의 빈도 수를 벡터로 표현하여, 발화 주제를 특징을 결정하는 기법이다.
- [160] 예를 들어, 프로세서(180)는 텍스트 데이터에 <달리기>, <체력> 등과 같은 단어가 자주 등장하면, 사용자의 발화 주제를 운동으로 분류할 수 있다.
- [161] 프로세서(180)는 공지된 텍스트 카테고리화(Text Categorization) 기법을 이용하여, 텍스트 데이터로부터 사용자의 발화 주제를 결정할 수 있다. 프로세서(180)는 텍스트 데이터로부터 키워드를 추출하여, 사용자의 발화 주제를 결정할 수 있다.
- [162] 프로세서(180)는 전체 주파수 대역에서의 진폭 정보를 고려하여 사용자의 성량을 결정할 수 있다.
- [163] 예컨대, 프로세서(180)는 파워 스펙트럼의 각 주파수 대역에서의 진폭의 평균 또는 가중치 평균을 기준으로 사용자의 성량을 결정할 수 있다.
- [164] 도 3 및 도 4에서 설명된 오디오 프로세서(181) 및 프로세서(180)의 기능은 NLP

- 서버(20), 음성 합성 서버(30) 중 어느 하나의 서버에서도 수행될 수 있다.
- [165] 예를 들어, NLP 서버(20)는 음성 신호를 이용하여, 파워 스펙트럼을 추출하고, 추출된 파워 스펙트럼을 이용하여, 사용자의 발화 특징을 결정할 수 있다.
- [166] 도 5는 본 발명의 일 실시 예에 따른 음성 합성 서버의 구성을 설명하기 위한 블록도이다.
- [167] 도 5를 참조하면, 음성 합성 서버(30)는 무선 통신부(31), 데이터 베이스(33) 및 프로세서(39)를 포함할 수 있다.
- [168] 무선 통신부(31)는 이동 단말기(100)의 무선 통신부(110)와 무선으로 통신을 수행할 수 있다.
- [169] 무선 통신부(31)는 이동 단말기(100)의 무선 통신부(110)의 구성 요소들을 포함할 수 있다.
- [170] 무선 통신부(31)는 단말기(100)로부터 텍스트 데이터 및 요청 제조 정보를 수신할 수 있다.
- [171] 데이터 베이스(33)는 합성 음성을 생성하는데 사용되는 특정 화자에 대한 음성 데이터를 저장하고 있을 수 있다.
- [172] 또한, 데이터 베이스(33)는 합성 음성을 암호화하는데 사용되는 암호화 테이블을 저장하고 있을 수 있다.
- [173] 프로세서(39)는 단말기(100)로부터 수신된 텍스트 데이터에 대응하는 합성 음성 신호를 생성할 수 있다.
- [174] 프로세서(39)는 생성된 합성 음성 신호에서, 하나 이상의 암호화 대상 구간들을 선정할 수 있다.
- [175] 프로세서(39)는 선정된 하나 이상의 암호화 대상 구간들 각각의 주파수 정보를 추출할 수 있다.
- [176] 프로세서(39)는 추출된 주파수 정보와 요청 제조 정보를 해당 암호화 대상 구간에 삽입하여, 암호화 합성 음성 신호를 생성할 수 있다.
- [177] 프로세서(39)는 생성된 암호화 합성 음성 신호를 무선 통신부(31)를 통해 단말기(100)에 전송할 수 있다.
- [178] 프로세서(39)의 상세한 기능에 대해서는 후술한다.
- [179] 도 6은 본 발명의 일 실시 예에 따른 합성 음성 암호화 시스템의 합성 음성 암호화 방법을 설명하기 위한 래더 다이어그램이다.
- [180] 도 6을 참조하면, 음성 합성 서버(30)의 무선 통신부(31)는 단말기(100)로부터 텍스트 데이터 및 요청 제조 정보를 수신한다(S601).
- [181] 일 실시 예에서, 텍스트 데이터는 합성 음성을 생성하는데 기초가 되는 데이터일 수 있다.
- [182] 요청 제조 정보는 합성 음성을 암호화하는데 사용되는 정보일 수 있다.
- [183] 요청 제조 정보는 요청 시간, 디바이스 식별자 중 하나 이상을 포함할 수 있다.
- [184] 요청 시간은 단말기(100)가 텍스트 데이터를 음성 합성 서버(30)에 전송한 시점일 수 있다. 요청 시간은 추후, 합성 음성의 암호화를 요청한 시점이

- 맞는지를 판별하는데 사용될 수 있다.
- [185] 디바이스 식별자는 텍스트 데이터를 전송한 단말기(100)를 식별하기 위한 식별자를 포함할 수 있다. 디바이스 식별자는 단말기(100)의 모델명, 고유 번호 등을 포함할 수 있다.
- [186] 디바이스 식별자는 합성 음성을 암호화한 기기가 맞는지를 판별하는데 사용될 수 있다.
- [187] 음성 합성 서버(30)의 프로세서(39)는 수신된 텍스트 데이터에 대응하는 합성 음성 신호를 생성한다(S603).
- [188] 프로세서(39)는 텍스트 데이터를 구성하는 음소들을 추출할 수 있다.
- [189] 프로세서(39)는 추출된 음소들에 대응하는 음성 신호들을 데이터 베이스(33)로부터 획득하고, 획득된 음성 신호들을 합성하여, 합성 음성 신호를 생성할 수 있다.
- [190] 음성 합성 서버(30)의 프로세서(39)는 생성된 합성 음성 신호에서, 하나 이상의 암호화 대상 구간들을 선정한다(S605).
- [191] 일 실시 예에서, 프로세서(39)는 생성된 합성 음성 신호에서, 랜덤하게 하나 이상의 암호화 대상 구간들을 선정할 수 있다.
- [192] 합성 음성 신호를 구성하는 일부 구간들에 대해서만, 암호화가 수행되는 경우, 추후, 단말기(100)에서 합성 음성 신호를 복호화하는 로드가 줄어들 수 있다.
- [193] 또 다른 실시 예에서, 프로세서(39)는 생성된 합성 음성 신호의 전체를 암호화 대상으로 선정할 수 있다.
- [194] 음성 합성 서버(30)의 프로세서(39)는 선정된 하나 이상의 암호화 대상 구간들 각각의 주파수 정보를 추출한다(S607).
- [195] 주파수 정보는 주파수 대역 및 주파수 대역에 대응하는 신호의 진폭의 제곱 값을 포함할 수 있다.
- [196] 일 실시 예에서, 프로세서(39)는 합성 음성 신호를 파워 스펙트럼으로 변환할 수 있다. 프로세서(39)는 변환된 파워 스펙트럼으로부터, 선정된 하나 이상의 암호화 대상 구간들에 대응하는 부분 파워 스펙트럼을 획득할 수 있다.
- [197] 프로세서(39)는 획득된 부분 파워 스펙트럼으로부터, 특정 Hz 또는 특정 Hz 대역에 대응하는 크기를 추출할 수 있다.
- [198] 단계 S603 내지 S607에 대해서는 이하의 도면을 참조하여 설명한다.
- [199] 도 7은 본 발명의 실시 예에 따라 텍스트 데이터를 합성 음성 신호로 변환하는 과정을 설명하는 도면이고, 도 8은 합성 음성 신호로부터 하나 이상의 암호화 대상 구간들을 선정하는 과정을 설명하는 도면이고, 도 9는 선정된 하나 이상의 암호화 대상 구간들 각각의 주파수 정보를 추출하는 과정을 설명하는 도면이다.
- [200] 도 7을 참조하면, 음성 합성 서버(30)의 프로세서(39)는 단말기(100)로부터 수신된 텍스트 데이터를 합성 음성 신호(700)로 변환할 수 있다. 합성 음성 신호(700)는 시간의 흐름에 따른 신호의 크기 변화를 보여준다.
- [201] 프로세서(39)는 특정 화자의 음색으로, 합성 음성 신호(700)를 생성할 수 있다.

- [202] 프로세서(39)는 복수의 화자들 각각에 대응하는 음성 신호를 저장한 데이터 베이스(33)로부터 특정 화자에 대응하는 음성 신호들을 추출하여, 합성 음성 신호(700)를 생성할 수 있다.
- [203] 도 8을 참조하면, 프로세서(39)는 합성 음성 신호(700)로부터 4개의 암호화 구간들(710 내지 740)를 선정할 수 있다. 프로세서(39)는 합성 음성 신호(700)로부터 암호화할 구간 및 구간의 수를 랜덤하게 선정할 수 있다.
- [204] 도 9를 참조하면, 프로세서(39)는 선정된 암호화 대상 구간들(710 내지 740) 각각을 부분 파워 스펙트럼들(910 내지 940)으로 변환할 수 있다.
- [205] 프로세서(39)는 부분 파워 스펙트럼들(910 내지 940) 각각으로부터 각 암호화 대상 구간의 주파수 정보를 획득할 수 있다.
- [206] 주파수 정보는 주파수 대역 및 주파수 대역에 대응하는 진폭의 제곱 값을 포함할 수 있다.
- [207] 다시 도 6을 설명한다.
- [208] 음성 합성 서버(30)의 프로세서(39)는 추출된 주파수 정보와 요청 제조 정보를 해당 암호화 대상 구간에 삽입하여, 암호화 합성 음성 신호를 생성한다(S609).
- [209] 프로세서(39)는 하나 이상의 암호화 대상 구간들 각각에 대응하는 주파수 정보와 요청 제조 정보를 암호화하여, 해당 암호화 대상 구간들 각각에 삽입할 수 있다.
- [210] 프로세서(39)는 비 대칭 암호화 방식을 이용하여, 암호화 합성 음성 신호를 생성할 수 있다.
- [211] 비 대칭 암호화 방식은 주파수 정보 및 요청 제조 정보를 포함하는 요청 정보를 공개 키를 이용하여, 암호화하고, 공개 키에 대응하는 비밀 키를 이용하여, 요청 정보를 복호화하는 방식이다.
- [212] 이에 대해서는 도 10 내지 도 12를 참조하여 설명한다.
- [213] 도 10 내지 도 12는 본 발명의 일 실시 예에 따라 암호화 합성 음성 신호를 생성하는 과정을 설명하는 도면이다.
- [214] 도 10은 본 발명의 일 실시 예에 따라 암호화 합성 음성 신호를 생성하는 방법을 설명하는 흐름도이다.
- [215] 도 10을 참조하면, 음성 합성 서버(30)의 프로세서(39)는 암호화 테이블을 이용하여, 주파수 정보와 요청 제조 정보를 포함하는 요청 정보를 공개 키 값으로 변환한다(S1010).
- [216] 암호화 테이블은 하나의 글자에 매핑된 캐릭터 간의 대응 관계를 저장한 테이블일 수 있다.
- [217] 주파수 정보 및 요청 제조 정보는 텍스트이고, 텍스트를 구성하는 글자마다, 하나의 캐릭터가 매핑될 수 있다.
- [218] 프로세서(39)는 암호화 테이블을 이용하여, 요청 정보를 공개 키 값으로 변환할 수 있다.
- [219] 예를 들어, 도 11의 공개 키 테이블(1100)에 도시된 바와 같이, 제1 암호화

구간(710)에 대응하는 제1 요청 정보는 제1 공개 키 값으로 변환되고, 제2 암호화 구간(720)에 대응하는 제2 요청 정보는 제2 공개 키 값으로 변환되고, 제3 암호화 구간(730)에 대응하는 제3 요청 정보는 제3 공개 키 값으로 변환되고, 제4 암호화 구간(740)에 대응하는 제4 요청 정보는 제4 공개 키 값으로 변환될 수 있다.

- [220] 프로세서(39)는 변환된 공개 키 값을 해당 암호화 대상 구간에 삽입한다(S1030).
- [221] 프로세서(39)는 각 공개 키 값을 공개 키 신호로 변환하고, 변환된 공개 키 신호를 해당 암호화 구간에 삽입할 수 있다.
- [222] 프로세서(39)는 삽입 결과에 따라 암호화 구간들 및 비 암호화 구간들을 포함하는 암호화 합성 음성 신호를 생성한다(S1050).
- [223] 도 12에서 도시된 바와 같이, 프로세서(39)는 제1 암호화 대상 구간(710)에 제1 공개 키 값을 삽입하여, 제1 암호화 구간(1210)을 생성할 수 있다.
- [224] 마찬가지로, 프로세서(39)는 제2 암호화 대상 구간(720)에 제2 공개 키 값을 삽입하여, 제2 암호화 구간(1220)을 생성할 수 있고, 제3 암호화 대상 구간(730)에 제3 공개 키 값을 삽입하여, 제3 암호화 구간(1230)을 생성할 수 있고, 제4 암호화 대상 구간(740)에 제4 공개 키 값을 삽입하여, 제4 암호화 구간(1240)을 생성할 수 있다.
- [225] 최종적으로, 프로세서(39)는 복수의 암호화 구간들(1210 내지 1240) 및 비 암호화 구간들을 포함하는 암호화 합성 음성 신호(1200)를 생성할 수 있다.
- [226] 다시, 도 6을 설명한다.
- [227] 음성 합성 서버(30)의 무선 통신부(31)는 생성된 암호화 합성 음성 신호를 단말기(100)에 전송한다(S611).
- [228] 단말기(100)의 프로세서(180)는 음성 합성 서버(30)로부터 암호화 합성 음성 신호에 포함된 하나 이상의 암호화 구간들을 추출한다(S613).
- [229] 암호화 합성 음성 신호에 포함된 하나 이상의 암호화 구간들 각각은 공개 키 값을 포함할 수 있다.
- [230] 프로세서(180)는 각 암호화 구간에 포함된 공개 키 값을 인지하여, 암호화 합성 음성 신호로부터, 암호화 구간들을 추출할 수 있다.
- [231] 단말기(100)의 프로세서(180)는 복호화 테이블을 이용하여, 추출된 하나 이상의 암호화 구간들 각각으로부터, 복호화 요청 정보를 획득한다(S615).
- [232] 일 실시 예에서, 복호화 테이블은 공개 키 값과, 이에 대응하는 비밀 키 값 간의 대응 관계를 포함하는 테이블일 수 있다.
- [233] 프로세서(180)는 복호화 테이블을 이용하여, 각 암호화 구간을 복호화할 수 있다.
- [234] 프로세서(180)는 공개 키 값에 매칭되는 비밀 키 값을 이용하여, 각 암호화 구간을 복호화할 수 있다.
- [235] 암호화 구간이 복호화됨에 따라 요청 제조 정보 및 주파수 정보를 포함하는 복호화된 요청 정보와 암호화 대상 구간이 분리될 수 있다.

- [236] 즉, 프로세서(180)는 복호화에 따라, 복호화된 합성 음성 신호 및 복호화된 요청 정보를 획득할 수 있다.
- [237] 도 13은 본 발명의 일 실시 예에 따른 복호화 테이블을 설명하는 도면이다.
- [238] 도 13을 참조하면, 공개 키 값과 비밀 키 값이 서로 매핑되어 있는 복호화 테이블(1300)이 도시되어 있다.
- [239] 프로세서(180)는 하나 이상의 암호화 구간들 각각의 공개 키 값과, 그에 대응하는 비밀 키 값을 이용하여, 각 암호화 구간을 복호화할 수 있다.
- [240] 예를 들어, 제1 공개 키 값에는 제1 비밀 키 값이 매핑되어 있을 수 있다.
- [241] 프로세서(180)는 제1 암호화 구간에 대응하는 제1 공개 키 값을 추출하고, 복호화 테이블(1300)로부터 제1 공개 키 값에 대응하는 제1 비밀 키 값을 획득할 수 있다.
- [242] 프로세서(180)는 제1 비밀 키 값을 이용하여, 각 암호화 구간을 복호화할 수 있다.
- [243] 다시, 도 6을 설명한다.
- [244] 단말기(100)의 프로세서(180)는 획득된 복호화 요청 정보와 메모리(170)에 미리 저장된 요청 정보가 동일한지를 판단한다(S617).
- [245] 프로세서(180)는 각 암호화 구간에 대해 복호화된 요청 정보와 미리 저장된 요청 정보 간에 동일한지를 판단할 수 있다.
- [246] 즉, 프로세서(180)는 미리 저장된 요청 정보에 포함된 단말기(100)가 텍스트 데이터를 합성 음성 서버(30)에 전송한 요청 시간과 복호화된 요청 시간과 동일한지 판단할 수 있다.
- [247] 또한, 프로세서(180)는 미리 저장된 요청 정보에 포함된 단말기(100)의 식별자와 복호화된 식별자가 동일한지를 판단할 수 있다.
- [248] 또한, 프로세서(180)는 미리 저장된 요청 정보에 포함된 주파수 정보와, 복호화된 주파수 정보가 동일한지를 판단할 수 있다.
- [249] 단말기(100)의 프로세서(180)는 각 암호화 구간에 대해 복호화된 요청 정보와 미리 저장된 요청 정보가 동일한 경우, 복호화된 합성 음성 신호를 음향 출력부(152)를 통해 출력한다(S619).
- [250] 프로세서(180)는 복수의 암호화 구간들 각각에 대해, 요청 시간, 단말기(100)의 식별자, 주파수 정보가 모두 동일한 경우, 합성 음성이 위조되지 않은 것으로 판단할 수 있다. 즉, 합성 음성의 인증이 성공한 것으로 판단할 수 있다.
- [251] 프로세서(180)는 암호화 합성 음성 신호를 음향 출력부(152)를 통해 출력할 수 있다. 동시에, 프로세서(180)는 합성 음성 신호의 인증이 성공되었음을 나타내는 알림을 디스플레이부(151) 또는 음향 출력부(152)를 통해 출력할 수도 있다.
- [252] 단말기(100)의 프로세서(180)는 복수의 암호화 구간들 중 어느 하나의 암호화 구간에 대해 복호화된 요청 정보와 미리 저장된 요청 정보가 동일하지 않은 경우, 합성 음성의 인증이 실패했음을 나타내는 인증 실패 알림을 음향 출력부(152)를 통해 출력한다(S621).

- [253] 프로세서(180)는 복수의 암호화 구간들 중 어느 하나의 암호화 구간에 대해 복호화된 요청 정보와 미리 저장된 요청 정보가 동일하지 않은 경우, 합성 음성을 출력하지 않을 수 있다.
- [254] 즉, 프로세서(180)는 복수의 암호화 구간들 중 어느 하나의 암호화 구간에 대해 복호화된 요청 정보와 미리 저장된 요청 정보가 동일하지 않은 경우, 합성 음성이 위조된 것으로 결정할 수 있다.
- [255] 프로세서(180)는 복수의 암호화 구간들 중 어느 하나의 암호화 구간에 대해 복호화된 요청 제조 정보가 미리 저장된 요청 제조 정보와 일치하지 않은 경우, 합성 음성이 위조된 것으로 결정할 수 있다.
- [256] 프로세서(180)는 복수의 암호화 구간들 중 어느 하나의 암호화 구간에 대해 복호화된 요청 시간이 미리 저장된 요청 시간과 일치하지 않은 경우, 합성 음성이 위조된 것으로 결정할 수 있다.
- [257] 단말기(100)는 합성 음성이 위조된 것으로 판단한 경우, 합성 음성을 출력하지 않아, 의도치 않은 합성 음성이 출력되어, 도용되는 상황이 미연에 방지될 수 있다.
- [258] 본 발명의 실시 예에 따르면, 합성 음성에 암호화된 정보의 일치 여부를 판단하여, 합성 음성을 인증함에 따라, 합성 음성을 이용한 보이스 피싱 범죄를 방지할 수 있다.
- [259] 한편, 도 6의 단계 S603 내지 S609는 단말기(100)의 프로세서(180)가 수행할 수도 있다.
- [260] 이 경우, 단계 S601 및 S603은 생략될 수 있다.
- [261] 또한, 본 발명의 실시 예에 따르면, 상업적으로 합성 음성을 무단으로 도용하는 업체를 감지하여, 합성 음성의 악용이 방지될 수 있다.
- [262] 전술한 본 발명은, 프로그램이 기록된 매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 매체는, 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 매체의 예로는, HDD(Hard Disk Drive), SSD(Solid State Disk), SDD(Silicon Disk Drive), ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광 데이터 저장 장치 등이 있다. 또한, 상기 컴퓨터는 단말기의 프로세서(180)를 포함할 수도 있다.
- [263]
- [264]

청구범위

- [청구항 1] 음성 합성 서버에 있어서,
단말기로부터 텍스트 데이터 및 요청 제조 정보를 수신하는 무선 통신부;
및
상기 텍스트 데이터에 대응하는 합성 음성 신호를 생성하고, 생성된 합성 음성 신호로부터, 주파수 정보를 추출하고, 추출된 주파수 정보와 상기 요청 제조 정보를 포함하는 요청 정보를 상기 합성 음성 신호에 삽입하여, 암호화 합성 음성 신호를 생성하는 프로세서를 포함하는 음성 합성 서버.
- [청구항 2] 제1항에 있어서,
상기 프로세서는
상기 생성된 합성 음성 신호에서, 하나 이상의 암호화 대상 구간들을 선정하고, 선정된 하나 이상의 암호화 대상 구간들의 주파수 정보를 추출하고, 추출된 주파수 정보와 상기 요청 제조 정보를 포함하는 상기 요청 정보를 선정된 하나 이상의 암호화 대상 구간들에 삽입하는 음성 합성 서버.
- [청구항 3] 제2항에 있어서,
상기 요청 제조 정보는
상기 단말기가 텍스트 데이터를 상기 음성 합성 서버에 전송한 시간을 나타내는 요청 시간 및 상기 단말기의 식별자를 포함하는 음성 합성 서버.
- [청구항 4] 제3항에 있어서,
상기 프로세서는
상기 요청 정보를 공개 키 값으로 변환하고, 변환된 공개 키 값을 상기 각 암호화 대상 구간들에 삽입하여, 상기 암호화 합성 음성 신호를 생성하는 음성 합성 서버.
- [청구항 5] 제1항에 있어서,
상기 프로세서는
상기 암호화 합성 음성 신호를 상기 무선 통신부를 통해, 상기 단말기로 전송하는 음성 합성 서버.
- [청구항 6] 단말기에 있어서,
메모리;
음향 출력부;
음성 합성 서버로부터 암호화 합성 음성 신호를 수신하는 무선 통신부; 및
상기 암호화 합성 음성 신호를 비밀 키 값을 이용하여, 복호화하여, 복호화된 합성 음성 신호 및 복호화된 요청 정보를 획득하고, 복호화된

요청 정보와 상기 메모리에 저장된 요청 정보를 비교하고, 비교 결과에 따라 상기 복호화된 합성 음성 신호의 위조 여부를 결정하는 프로세서를 포함하는

단말기.

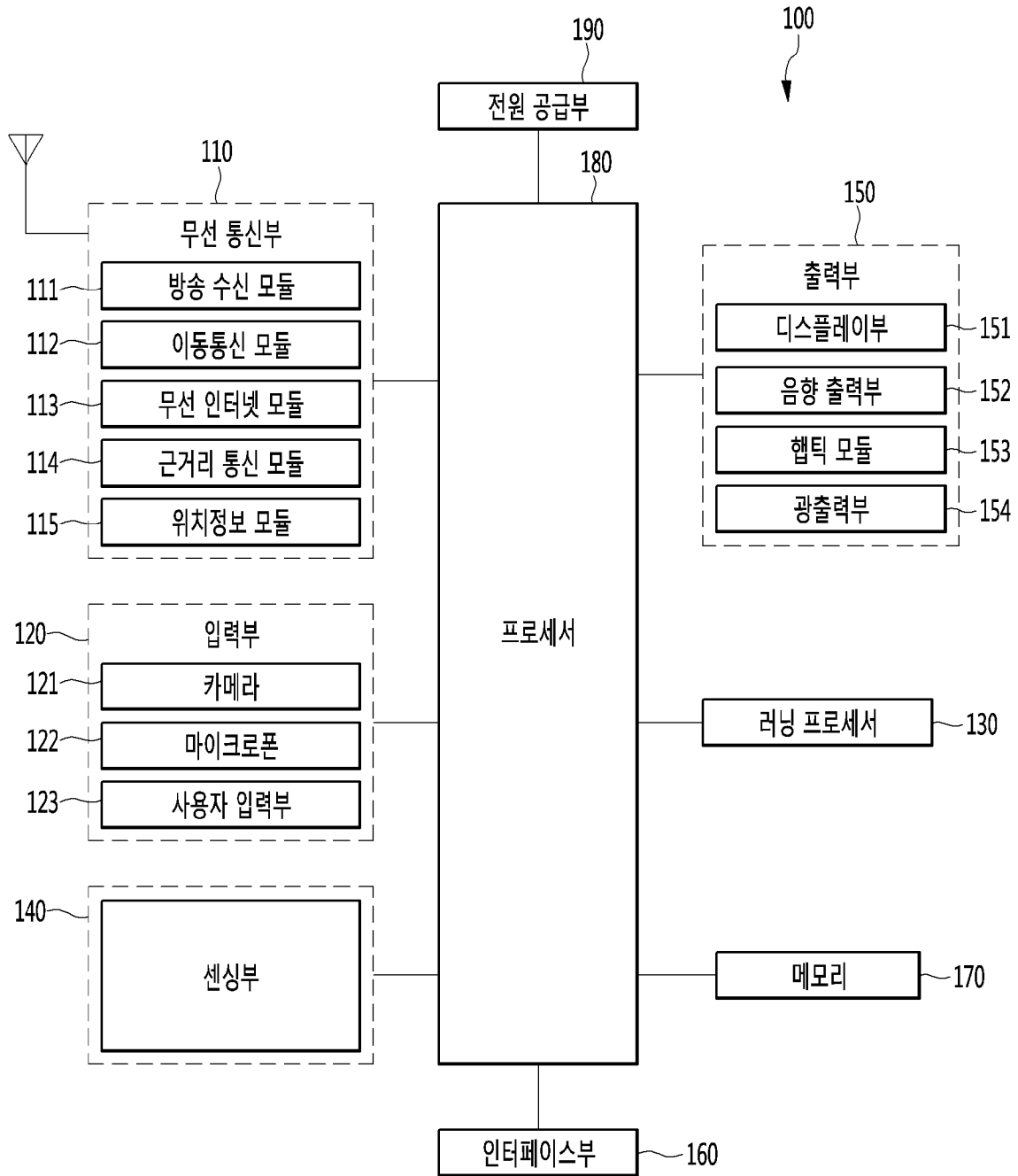
- [청구항 7] 제6항에 있어서,
상기 프로세서는
상기 암호화 합성 음성 신호에 포함된 하나 이상의 암호화 구간들을 추출하고,
상기 비밀 키 값을 이용하여, 추출된 하나 이상의 암호화 구간들 각각으로부터, 상기 복호화된 요청 정보를 획득하고,
상기 복호화된 요청 정보가 상기 메모리에 저장된 요청 정보와 일치하는 경우, 상기 복호화된 합성 음성 신호를 위조되지 않은 것으로 판단하는
단말기.
- [청구항 8] 제7항에 있어서,
상기 프로세서는
상기 복호화된 합성 음성 신호를 위조되지 않은 것으로 판단된 경우, 상기 음향 출력부를 통해 상기 복호화된 합성 음성 신호를 출력하는
단말기.
- [청구항 9] 제7항에 있어서,
상기 복호화된 요청 정보가 상기 메모리에 저장된 요청 정보와 일치하지 않는 경우, 상기 복호화된 합성 음성 신호를 위조된 것으로 판단하는
단말기.
- [청구항 10] 제6항에 있어서,
상기 저장된 요청 정보는
상기 단말기가 텍스트 데이터를 상기 음성 합성 서버에 전송한 시간을 나타내는 요청 시간 및 상기 단말기의 식별자를 포함하는
단말기.
- [청구항 11] 합성 음성 신호의 암호화 방법을 수행하는 프로그램을 기록한 기록 매체에 있어서,
상기 합성 음성 신호의 암호화 방법은
텍스트 데이터 및 요청 제조 정보를 획득하는 단계와,
상기 텍스트 데이터에 대응하는 합성 음성 신호를 생성하는 단계와,
생성된 합성 음성 신호로부터, 주파수 정보를 추출하는 단계와
추출된 주파수 정보와 상기 요청 제조 정보를 포함하는 요청 정보를 상기 합성 음성 신호에 삽입하여, 암호화 합성 음성 신호를 생성단계를 포함하는
기록 매체.
- [청구항 12] 제11항에 있어서,

상기 삽입하는 단계는
 상기 생성된 합성 음성 신호에서, 하나 이상의 암호화 대상 구간들을
 선정하는 단계와
 선정된 하나 이상의 암호화 대상 구간들의 주파수 정보를 추출하는
 단계와
 추출된 주파수 정보와 상기 요청 제조 정보를 포함하는 상기 요청 정보를
 선정된 하나 이상의 암호화 대상 구간들에 삽입하는 단계를 포함하는
 기록 매체.

[청구항 13] 제12항에 있어서,
 상기 요청 제조 정보는
 상기 단말기가 텍스트 데이터를 상기 음성 합성 서버에 전송한 시간을
 나타내는 요청 시간 및 상기 단말기의 식별자를 포함하는
 기록 매체.

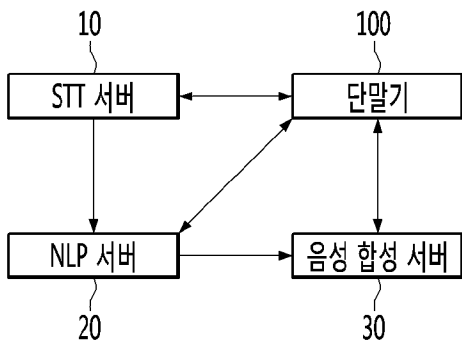
[청구항 14] 제13항에 있어서,
 상기 암호화 합성 음성 신호를 생성하는 단계는
 상기 요청 정보를 공개 키 값으로 변환하는 단계와
 변환된 공개 키 값을 상기 각 암호화 대상 구간들에 삽입하여, 상기
 암호화 합성 음성 신호를 생성하는 단계를 포함하는
 기록 매체.

[도 1]

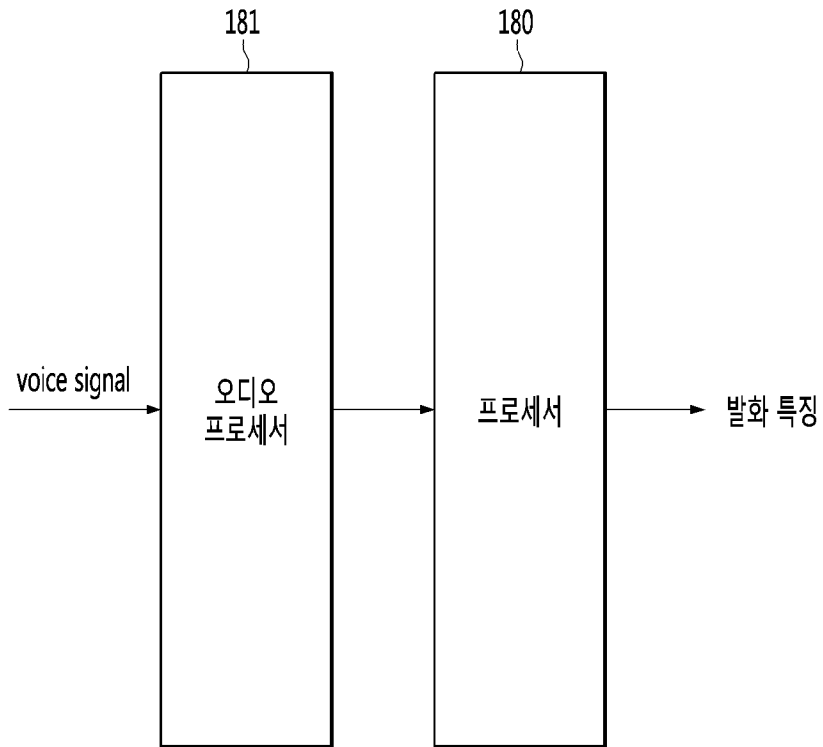


[도 2]

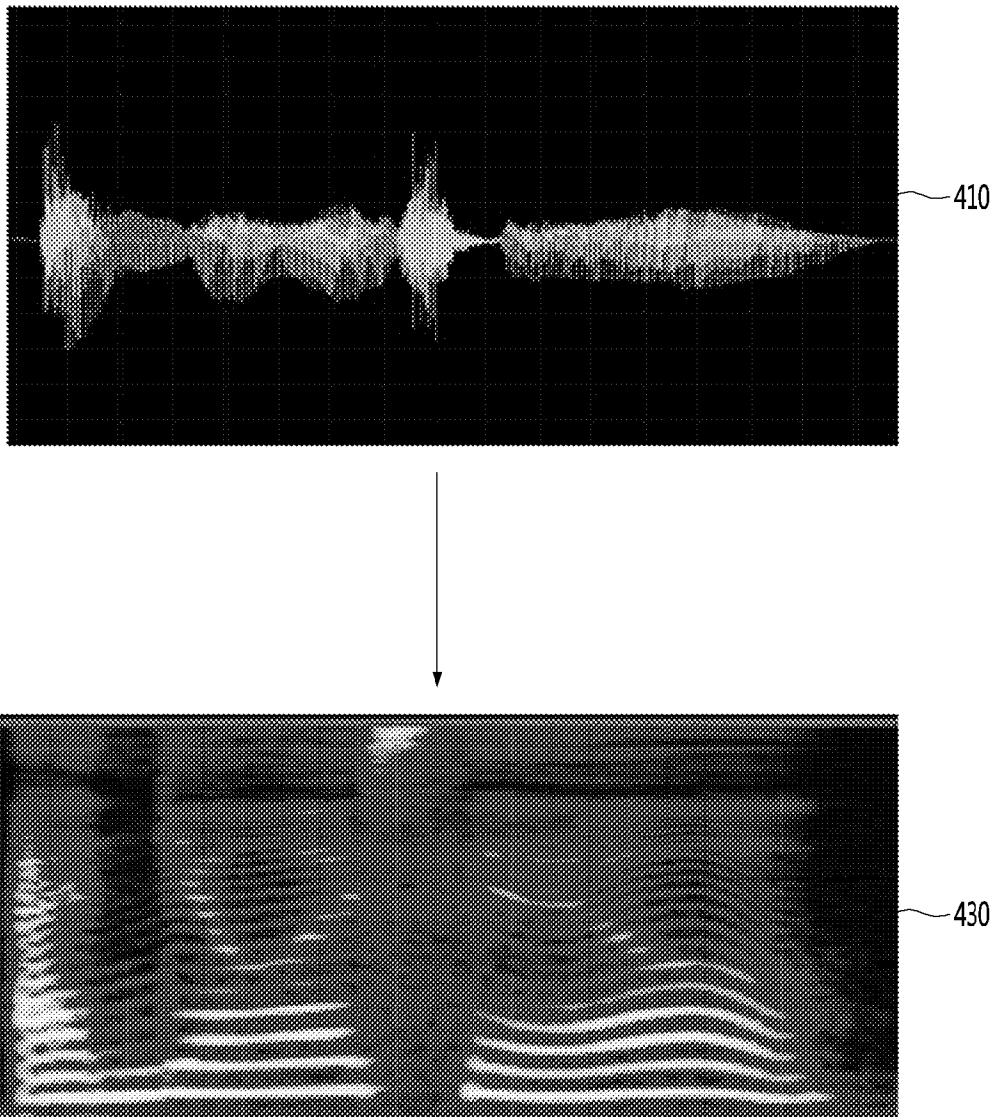
1



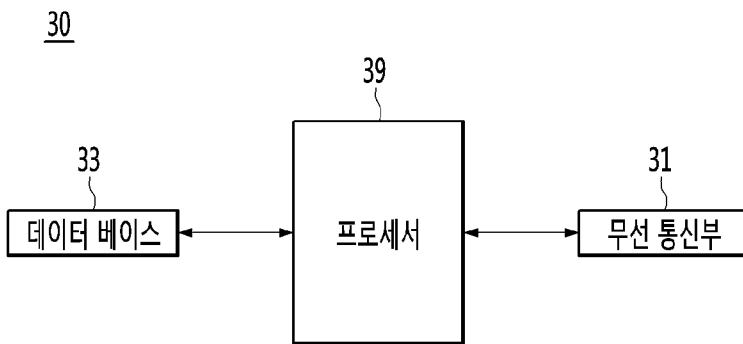
[도3]



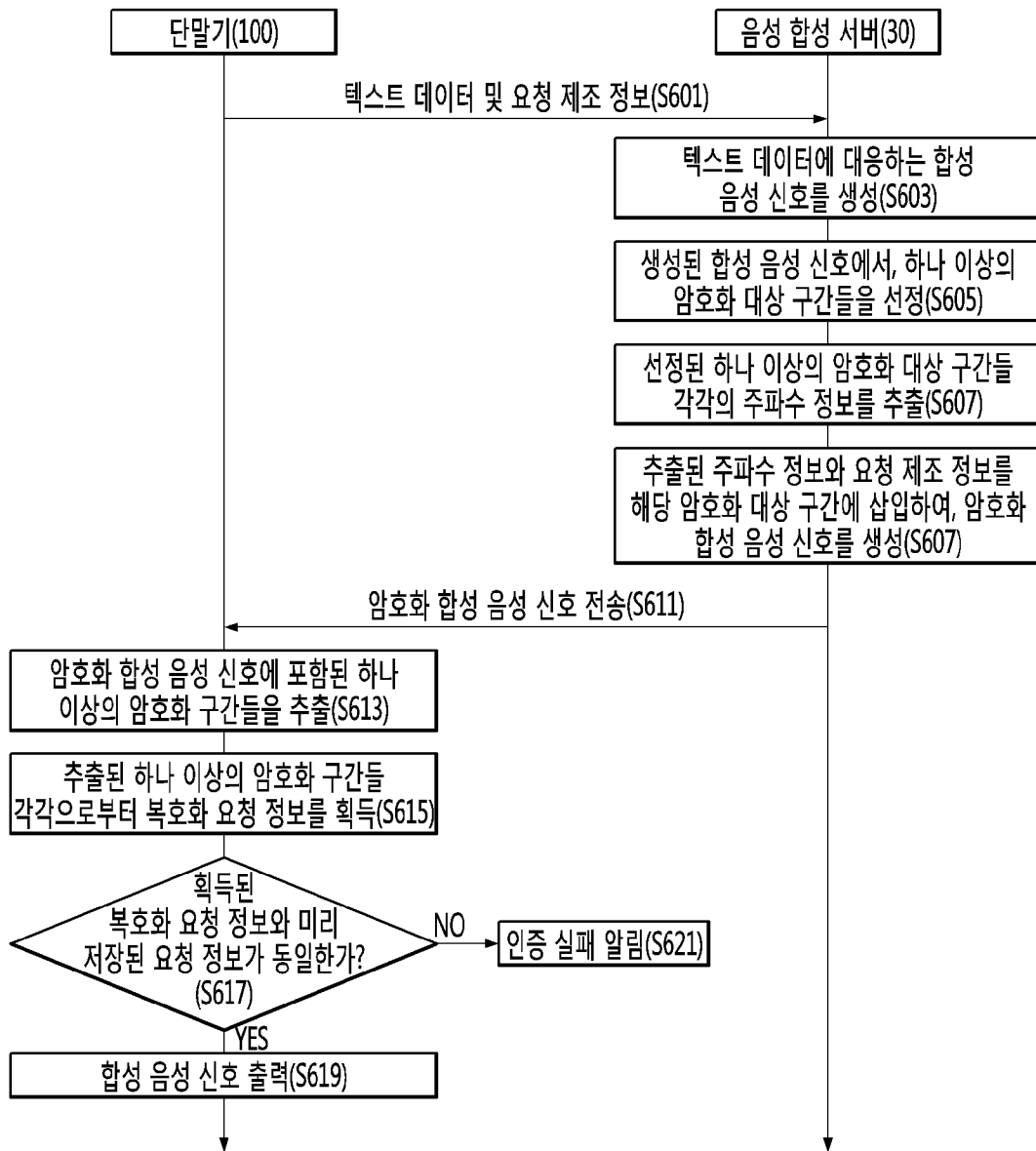
[도4]



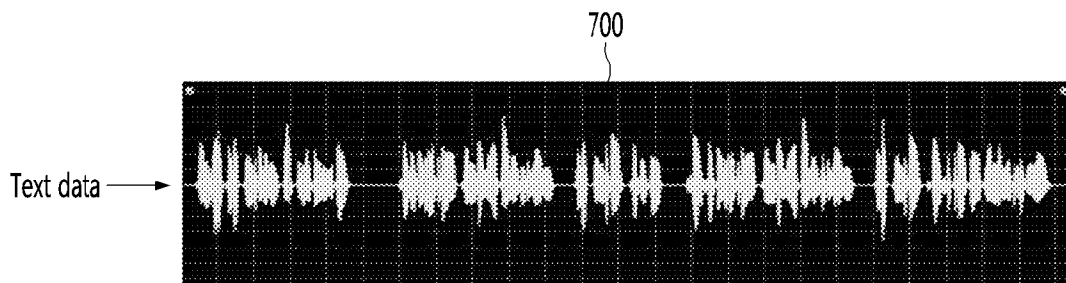
[도5]



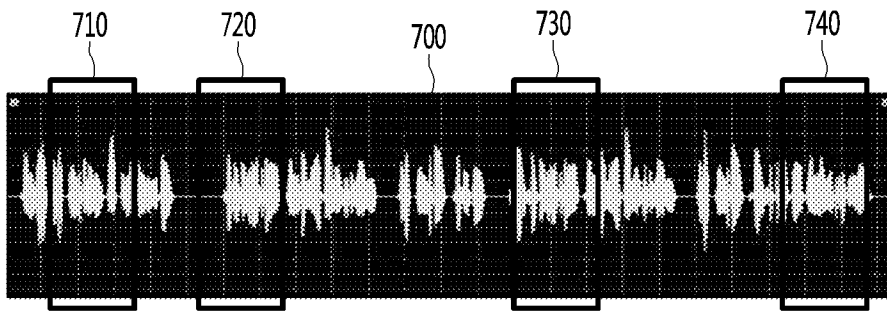
[도6]



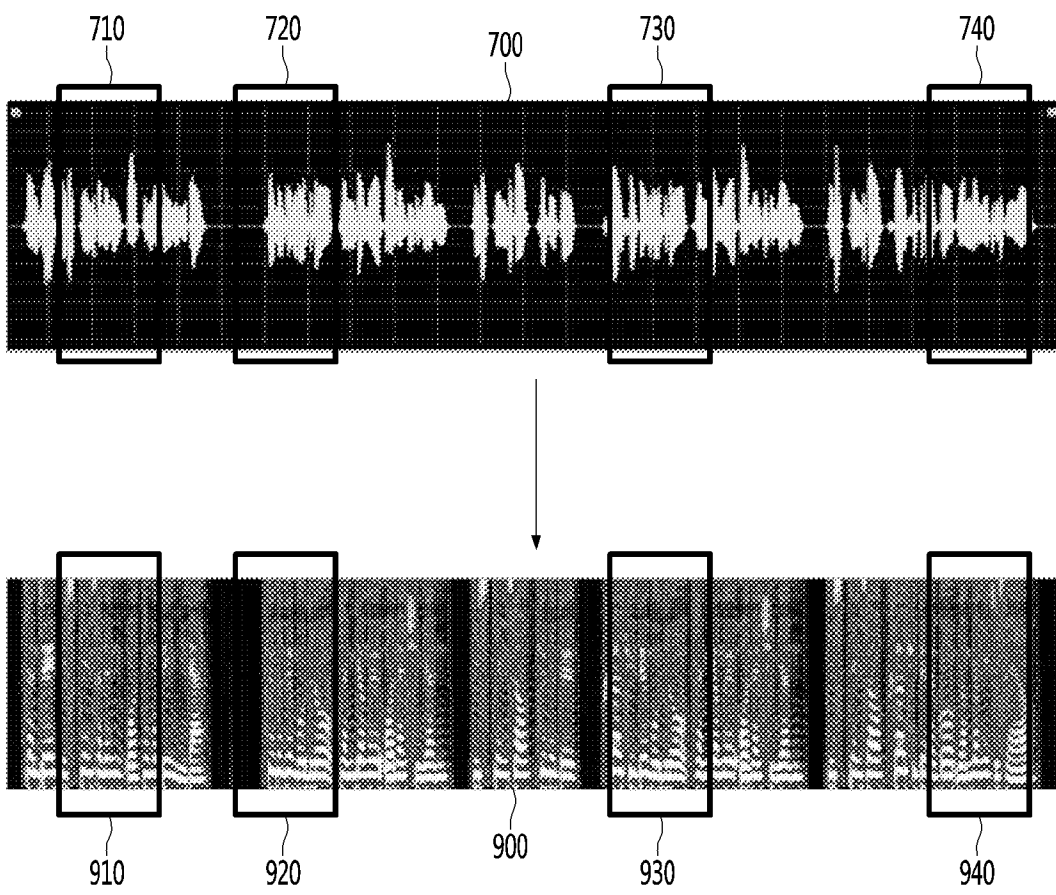
[도7]



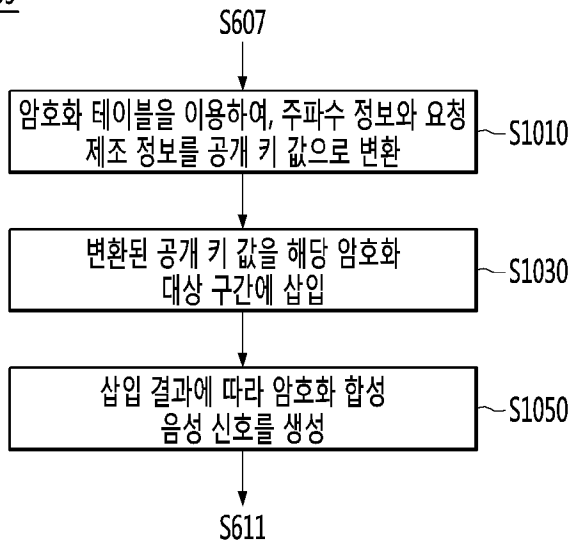
[도8]



[도9]



[도 10]

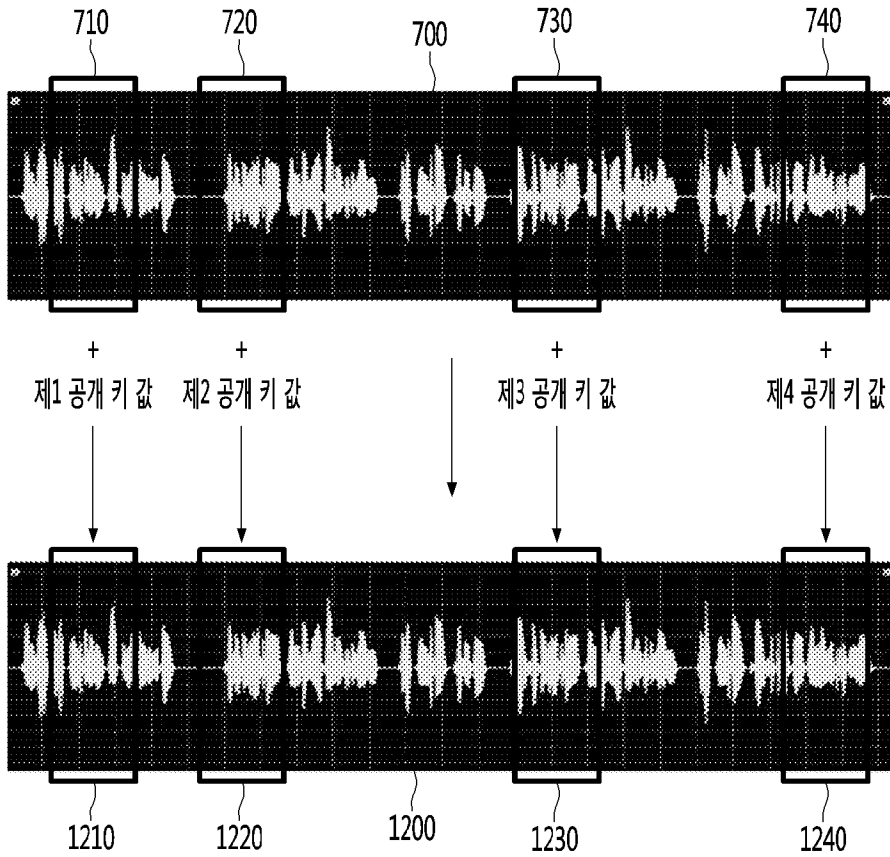
S609

[도 11]

1100

요청 정보	공개 키 값
제1 요청 정보	제1 공개 키 값(Kasjd983s9dasd9)
제2 요청 정보	제2 공개 키 값(Fghurei291asjf0)
제3 요청 정보	제3 공개 키 값(Dfndjtbe84dgw3u)
제4 요청 정보	제4 공개 키 값(Sdgfhn34h83nsol)

[도 12]



[도 13]

1300

공개 키 값	비밀 키 값
제1 공개 키 값(Kasjd983s9dasd9)	제1 비밀 키 값(eosdachorus)
제2 공개 키 값(Fghurei291asjf0)	제2 비밀 키 값(eosdacchint)
제3 공개 키 값(Dfndjtbe84dgw3u)	제3 비밀 키 값(eospublito)
제4 공개 키 값(Sdgfhn34h83nsol)	제4 비밀 키 값(eosscatterr)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2019/000118

A. CLASSIFICATION OF SUBJECT MATTER

G10L 13/047(2013.01)i, G10L 13/08(2006.01)i, G10L 15/30(2013.01)i, H04L 9/08(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 13/047; G06F 21/60; G07B 15/00; G10L 13/06; G10L 13/08; G10L 15/06; G10L 19/018; G10L 21/013; H04B 1/40; G10L 15/30; H04L 9/08

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 aboveElectronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS (KIPO internal) & Keywords: speech, synthesis, counterfeit, frequency, encrypt

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KR 10-2014-0098995 A (KIM, Hyoung Seok) 11 August 2014 See paragraphs [0037]-[0050]; claim 1; and figures 2-3.	1-14
Y	JP 2011-028130 A (PANASONIC ELECTRIC WORKS CO., LTD.) 10 February 2011 See paragraphs [0027], [0035]; claim 1; and figure 1.	1-14
A	US 2018-0342256 A1 (MODULATE, LLC.) 29 November 2018 See claim 1; and figures 1, 4-5C.	1-14
A	KR 10-0971443 B1 (YOO, Jang-ho) 23 July 2010 See claim 1.	1-14
A	JP 2002-183779 A (DAINIPPON PRINTING CO., LTD.) 28 June 2002 See paragraphs [0022]-[0026]; and claim 2.	1-14

 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

"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

27 SEPTEMBER 2019 (27.09.2019)

Date of mailing of the international search report

01 OCTOBER 2019 (01.10.2019)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex Daejeon Building 4, 189, Cheongsa-ro, Seo-gu,
Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2019/000118

Patent document cited in search report	Publication date	Patent family member	Publication date
KR 10-2014-0098995 A	11/08/2014	KR 10-1440898 B1	17/09/2014
JP 2011-028130 A	10/02/2011	None	
US 2018-0342256 A1	29/11/2018	US 2018-0342257 A1 US 2018-0342258 A1 WO 2018-218081 A1	29/11/2018 29/11/2018 29/11/2018
KR 10-0971443 B1	23/07/2010	KR 10-2009-0126354 A	09/12/2009
JP 2002-183779 A	28/06/2002	None	

A. 발명이 속하는 기술분류(국제특허분류(IPC))
G10L 13/047(2013.01)i, G10L 13/08(2006.01)i, G10L 15/30(2013.01)i, H04L 9/08(2006.01)i

B. 조사된 분야
 조사된 최소문헌(국제특허분류를 기재)
 G10L 13/047; G06F 21/60; G07B 15/00; G10L 13/06; G10L 13/08; G10L 15/06; G10L 19/018; G10L 21/013; H04B 1/40; G10L 15/30; H04L 9/08

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: 음성(speech), 합성(synthesis), 위조(counterfeit), 주파수(frequency), 암호화(encrypt)

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	KR 10-2014-0098995 A (김형석) 2014.08.11 단락 [0037]-[0050]; 청구항 1; 및 도면 2-3 참조.	1-14
Y	JP 2011-028130 A (PANASONIC ELECTRIC WORKS CO., LTD.) 2011.02.10 단락 [0027], [0035]; 청구항 1; 및 도면 1 참조.	1-14
A	US 2018-0342256 A1 (MODULATE, LLC) 2018.11.29 청구항 1; 및 도면 1, 4-5C 참조.	1-14
A	KR 10-0971443 B1 (유장호) 2010.07.23 청구항 1 참조.	1-14
A	JP 2002-183779 A (DAINIPPON PRINTING CO., LTD.) 2002.06.28 단락 [0022]-[0026]; 및 청구항 2 참조.	1-14

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 이후 “X”에 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 2019년 09월 27일 (27.09.2019)	국제조사보고서 발송일 2019년 10월 01일 (01.10.2019)
--	---

ISA/KR의 명칭 및 우편주소 대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578	심사관 권성호 전화번호 +82-42-481-3547
---	------------------------------------



국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2014-0098995 A	2014/08/11	KR 10-1440898 B1	2014/09/17
JP 2011-028130 A	2011/02/10	없음	
US 2018-0342256 A1	2018/11/29	US 2018-0342257 A1 US 2018-0342258 A1 WO 2018-218081 A1	2018/11/29 2018/11/29 2018/11/29
KR 10-0971443 B1	2010/07/23	KR 10-2009-0126354 A	2009/12/09
JP 2002-183779 A	2002/06/28	없음	