

명세서

발명의 명칭: 사용자 단말, 방송 장치, 이를 포함하는 방송 시스템 및 그 제어방법

기술분야

- [1] 화상 통화 콘텐츠를 실시간으로 방송함에 있어 번역 서비스를 제공하는 사용자 단말, 방송 장치, 이를 포함하는 방송 시스템 및 그 제어방법에 관한 것이다.

배경기술

- [2] IT 기술의 발달에 따라 사용자 간에 화상 통화가 자주 이루어지고 있으며 특히, 전세계 다양한 국가 사람들이 비즈니스 목적뿐만 아니라, 콘텐츠 공유, 취미 생활 공유 등을 목적으로 화상 통화 서비스를 사용하고 있다.
- [3] 다만, 모든 화상 통화 때마다 통역사를 옆에 두고 화상 통화를 하기에는 비용적으로도 시간적으로도 어려운 실정으로서, 이에 화상 통화에 대한 실시간 원문/번역 서비스를 제공하는 방법에 대한 연구가 진행 중이다.

발명의 상세한 설명

기술적 과제

- [4] 통화자 뿐만 아니라 시청자들에게 원문/번역 서비스를 실시간으로 제공함으로써 의사 교환, 의사 파악을 보다 원활하게 하며, 음성 및 텍스트 중 적어도 하나를 통해 원문/번역 서비스를 제공함으로써 시각장애인 뿐만 아니라 청각장애인도 자유롭게 의사 교환, 의사 파악을 보다 원활하게 하는데 목적이 있다.

과제 해결 수단

- [5] 일 측에 따른 방송 장치는, 통신망을 통해 채팅 방에 접속한 사용자 단말 간의 화상 통화를 지원하는 통신부; 상기 통신부를 통해 수신되는 화상 통화 관련 동영상 파일을 이용하여 영상파일과 음성파일을 생성하고, 상기 영상파일과 음성파일 중 적어도 하나를 이용하여 통화자 각각에 대한 원어정보를 추출하는 추출부; 상기 원어정보를 선택된 국가의 언어에 따라 번역한 번역정보를 생성하는 번역부; 및 상기 화상 통화 관련 동영상 파일에, 상기 원어정보 및 번역정보 중 적어도 하나가 매핑된 통번역 동영상이 상기 채팅 방에 접속한 사용자 단말 및 시청자 단말에 전송되도록 제어하는 제어부를 포함할 수 있다.
- [6] 또한 상기 원어정보는, 음성 원어정보 및 텍스트 원어정보 중 적어도 하나를 포함하고, 상기 번역정보는, 음성 번역정보 및 텍스트 번역정보 중 적어도 하나를 포함할 수 있다.
- [7] 또한 상기 추출부는, 상기 음성 파일에 대해 주파수 대역 분석 프로세스를 적용하여 통화자 각각에 대한 음성 원어정보를 추출하고, 상기 추출한 음성 원어정보에 대해 음성 인식 프로세스를 적용하여 텍스트 원어정보를 생성할 수 있다.

- [8] 또한 상기 추출부는, 상기 영상 파일에 대해 영상 처리 프로세스를 적용하여 수화 패턴을 검출하고, 상기 검출한 수화 패턴을 기반으로 텍스트 원어정보를 추출할 수 있다.
- [9] 일 측에 따른 사용자 단말은, 통신망을 통해 화상 통화 서비스를 지원하는 단말 통신부; 및 화상 통화 관련 동영상 파일에 원어정보 및 번역정보 중 적어도 하나가 매핑된 통번역 동영상을 제공하며, 적어도 하나 이상의 화상 통화 관련 설정 명령과 적어도 하나 이상의 번역 관련 설정 명령을 입력 받을 수 있는 아이콘을 제공하도록 구성된 유저인터페이스가 디스플레이 상에 표시되도록 제어하는 단말 제어부를 포함할 수 있다.
- [10] 또한 상기 적어도 하나 이상의 화상 통화 관련 설정 명령은, 화상 통화자의 발언권을 설정할 수 있는 발언권 설정 명령 및 화상 통화자 수 설정 명령, 시청자 수 설정 명령 및 텍스트 전송 명령 중 적어도 하나를 포함할 수 있다.
- [11] 또한 상기 단말 제어부는, 상기 발언권 설정 명령의 입력 여부에 따라 상기 통번역 동영상의 제공방법이 변경되거나 또는 발언권을 가진 통화자에 대한 정보가 포함된 팝업 메시지를 제공하도록 구성된 유저 인터페이스가 디스플레이 상에 표시되도록 제어할 수 있다.
- [12] 일 측에 따른 방송 장치의 제어방법은, 화상 통화 관련 동영상 파일을 수신하는 단계; 상기 화상 통화 관련 동영상 파일로부터 생성한 영상파일과 음성파일 중 적어도 하나를 이용하여 통화자 각각에 대한 원어정보를 추출하는 단계; 상기 원어정보를 선택된 국가의 언어에 따라 번역한 번역정보를 생성하는 단계; 및 상기 화상 통화 관련 동영상 파일에, 상기 원어정보 및 번역정보 중 적어도 하나가 매핑된 통번역 동영상을 채팅 창에 접속 중인 단말에 전송되도록 제어하는 단계를 포함할 수 있다.
- [13] 또한 상기 추출하는 단계는, 상기 음성 파일에 대해 주파수 대역 분석 프로세스를 적용하여 통화자 각각에 대한 음성 원어정보를 추출하는 단계; 및 상기 추출한 음성 원어정보에 대해 음성 인식 프로세스를 적용하여 텍스트 원어정보를 생성하는 단계를 포함할 수 있다.
- [14] 또한 상기 추출하는 단계는, 상기 영상 파일에 대해 영상 처리 프로세스를 적용하여 수화 패턴을 검출하고, 상기 검출한 수화 패턴을 기반으로 텍스트 원어정보를 추출하는 단계를 포함할 수 있다.

발명의 효과

- [15] 일 실시예에 따른 사용자 단말, 방송 장치, 이를 포함하는 방송 시스템 및 그 제어방법은, 통화자 뿐만 아니라 시청자들에게 원문/번역 서비스를 실시간으로 제공함으로써 의사 교환, 의사 파악을 보다 원활하게 한다.
- [16] 다른 일 실시예에 따른 사용자 단말, 방송 장치, 이를 포함하는 방송 시스템 및 그 제어방법은, 음성 및 텍스트 중 적어도 하나를 통해 원문/번역 서비스를 제공함으로써 시각장애인 뿐만 아니라 청각장애인도 자유롭게 의사 교환, 의사

파악을 보다 원활하게 한다.

도면의 간단한 설명

- [17] 도 1은 일 실시예에 따른 화상통화 방송시스템의 구성을 개략적으로 도시한 도면이다.
- [18] 도 2는 일 실시예에 따른 화상통화 방송시스템의 제어 블록도를 개략적으로 도시한 도면이다.
- [19] 도 3은 일 실시예에 따른 화상통화 중일 때 디스플레이 상에 표시되는 유저 인터페이스 화면을 도시한 도면이다.
- [20] 도 4는 일 실시예에 따른 각종 설정 명령을 입력 받도록 구성된 유저 인터페이스 화면을 도시한 도면이다.
- [21] 도 5 및 도 6은 서로 다른 실시예에 따른 발언권에 따라 구성이 변경되는 유저 인터페이스 화면을 도시한 도면이다.
- [22] 도 7은 일 실시예에 따른 방송 장치의 동작 흐름도를 개략적으로 도시한 도면이다.

발명의 실시를 위한 형태

- [23] 이하에서 설명되는 사용자 단말은 각종 연산 처리가 가능한 프로세서가 내장되어 있으며, 통신 모듈이 내장되어 있어 통신망을 통해 화상 통화 서비스가 가능한 모든 기기를 포함한다.
- [24] 예를 들어, 사용자 단말은 랩탑(laptop), 데스크 탑(desk top), 테블릿 PC(tablet PC)뿐만 아니라, 스마트 폰, PDA(Personal Digital Assistant)와 같은 모바일 단말 및 사용자의 신체에 탈부착이 가능한 시계, 안경 형태의 웨어러블 단말뿐만 아니라, 스마트 TV(Television), IPTV(Internet Protocol Television) 등을 포함하며 제한은 없다. 이하에서 설명의 편의상 사용자 단말을 이용하여 화상 통화 서비스를 이용하는 자를 사용자 또는 통화자로 혼용하여 지칭하기로 한다.
- [25] 이하에서 설명되는 시청자는 화상 통화에 직접적으로 참여하기 보다는 화상 통화를 시청하고자 하는 자로서, 이하에서 설명되는 시청자 단말은 전술한 사용자 단말로서 이용 가능한 기기 전부를 포함한다. 한편, 이하에서는 사용자 단말 및 시청자 단말을 구별하여 설명할 필요가 없는 경우 단말로 지칭하기로 한다.
- [26] 또한 이하에서 설명되는 방송 장치는 통신 모듈이 내장되어 있어 통신망을 통해 화상 통화 서비스를 제공할 수 있으며, 각종 연산 처리가 가능한 프로세서가 내장되어 있는 모든 기기를 포함한다.
- [27] 예를 들어, 방송 장치는 전술한 랩탑(laptop), 데스크 탑(desk top), 테블릿 PC(tablet PC), 스마트 폰, PDA(Personal Digital Assistant)와 같은 모바일 단말 및 웨어러블 단말뿐만 아니라, 스마트 TV(Television), IPTV(Internet Protocol Television)를 통해 구현 가능하다. 이외에도, 방송 장치는 통신 모듈과 프로세서가 내장된 서버를 통해서도 구현 가능하며 제한은 없다. 이하에서는

방송 장치에 대해 보다 구체적으로 설명하도록 한다.

- [28] 이하에서는 설명의 편의를 위해 도 1에 도시된 바와 같이 스마트 폰 형태의 사용자 단말 및 시청자 단말을 예로 들고, 서버 형태의 방송 장치를 예로 들어 설명하도록 하나 전술한 바와 같이 사용자 단말, 시청자 단말 및 방송 장치의 형태가 이에 한정되는 것은 아니며 제한은 없다.
- [29] 도 1은 일 실시예에 따른 화상통화 방송시스템의 구성을 개략적으로 도시한 도면이고, 도 2는 일 실시예에 따른 화상통화 방송시스템의 제어 블록도를 개략적으로 도시한 도면이다. 또한, 도 3은 일 실시예에 따른 화상통화 중일 때 디스플레이 상에 표시되는 유저 인터페이스 화면을 도시한 도면이고, 도 4는 일 실시예에 따른 각종 설정 명령을 입력 받도록 구성된 유저 인터페이스 화면을 도시한 도면이다. 또한 도 5 및 도 6은 서로 다른 실시예에 따른 발언권에 따라 구성이 변경되는 유저 인터페이스 화면을 도시한 도면이다. 이하에서는 설명의 중복을 방지하고자 함께 설명하도록 한다.
- [30] 도 1 및 도 2를 참조하면, 방송 시스템(1)은 사용자 단말(100-1 ..., 100-n: 100)($n \geq 1$), 시청자 단말(200-1 ..., 200-n: 200)($m \geq 1$)과 사용자 단말(100)과 시청자 단말(200) 간의 연결을 지원하며, 화상 통화 관련 동영상 파일과 화상 통화 관련 동영상 파일로부터 추출한 원어정보 및 번역정보를 함께 송출함으로써 번역 서비스를 제공하는 방송 장치(300)를 포함한다. 이하에서는 방송 장치(300)에 대해 보다 구체적으로 설명하도록 한다.
- [31] 도 2를 참조하면, 방송 장치(300)는 통신망을 통해 외부 단말과 데이터를 주고 받을 뿐만 아니라 또는 외부 단말 간의 화상 통화 서비스를 지원하는 통신부(310), 통신부(310)를 통해 수신되는 화상 통화 관련 동영상 파일을 이용하여 영상 파일과 음성 파일을 생성한 다음, 이를 기반으로 원어 정보를 추출하는 추출부(320), 원어정보를 번역하여 번역정보를 생성하는 번역부(330) 및 방송 장치(300) 내 구성요소의 전반적인 동작을 제어하여 화상 통화에 대한 방송 서비스를 제공할 뿐만 아니라, 번역 서비스를 제공하는 제어부(340)를 포함할 수 있다.
- [32] 여기서, 통신부(310), 추출부(320), 번역부(330), 및 제어부(340)는 각각 별도로 구현되거나 또는 적어도 하나는 하나의 시스템 온 칩(System On Chip, SOC)으로 통합 구현될 수 있다. 다만, 방송 장치(300) 내에 시스템 온 칩이 하나만 존재하는 것은 아닐 수 있으므로, 하나의 시스템 온 칩에 집적되는 것으로 한정되는 것은 아니며 구현방법에는 제한이 없다. 이하에서는 방송 장치(300)의 구성요소에 대해 구체적으로 설명하도록 한다.
- [33] 통신부(310)는 무선통신망 또는 유선통신망을 통해 외부 기기와 각종 데이터를 주고 받을 수 있다. 여기서, 무선통신망은 데이터가 포함된 신호를 무선으로 주고 받을 수 있는 통신망을 의미한다.
- [34] 예를 들어, 통신부(310)는 3G(3Generation), 4G(4Generation), 5G(5Generation) 등과 같은 통신방식을 통해 기지국을 거쳐 단말 간에 무선 신호를 송수신할 수

있으며, 이외에도 무선 랜(Wireless LAN), 와이파이(Wi-Fi), 블루투스(Bluetooth), 지그비(Zigbee), WFD(Wi-Fi Direct), UWB(Ultra wideband), 적외선 통신(IrDA; Infrared Data Association), BLE (Bluetooth Low Energy), NFC(Near Field Communication) 등과 같은 통신방식을 통해 소정 거리 이내의 단말과 데이터가 포함된 무선 신호를 송수신할 수 있다.

- [35] 또한, 유선통신망은 데이터가 포함된 신호를 유선으로 주고 받을 수 있는 통신망을 의미한다. 예를 들어, 유선통신망은 PCI(Peripheral Component Interconnect), PCI-express, USB(Universe Serial Bus) 등을 포함하나, 이에 한정되는 것은 아니다. 이하에서는 설명되는 통신망은 무선통신망과 유선통신망을 전부 포함한다.
- [36] 통신부(310)는 화상 통화 서비스를 제공하기 위해 통신망을 통해 사용자 단말(200) 간에 연결되도록 할 수 있으며, 화상 통화를 시청할 수 있게끔 시청자 단말(300)을 연결할 수도 있다.
- [37] 예를 들어, 화상 통화를 실시간 스트리밍하고자 사용자들이 모여 채팅 방을 개설한 경우, 해당 채팅 방에는 시청자들이 접속할 수 있다. 이 경우, 통신부(310)는 통신망을 통해 사용자 간에 화상 통화가 원활히 이루어질 수 있도록 할 뿐만 아니라, 화상 통화 콘텐츠를 시청자들에게 전송함으로써 실시간 화상 통화 방송서비스가 이루어지게끔 한다.
- [38] 구체적인 예로, 제어부(340)는 통신부(310)를 통해 사용자 단말(200)로부터 수신한 채팅 방 생성 요청에 따라 채팅 방을 생성한 다음, 채팅 방에 접속한 시청자 단말(300)에서도 화상 통화를 시청할 수 있게끔 통신부(310)를 제어할 수도 있다. 제어부(340)에 관한 구체적인 설명은 후술하도록 한다.
- [39] 도 2를 참조하면, 방송 장치(300)에는 추출부(320)가 마련될 수 있다. 추출부(320)는 통신부(310)를 통해 수신되는 화상 통화 관련 동영상 파일을 이용하여 영상 파일과 음성 파일을 생성할 수 있다. 화상 통화 관련 동영상 파일은 화상 통화 중에 사용자 단말(200)로부터 수집되는 데이터로서 시각적인 정보를 제공하는 영상정보와 청각적인 정보를 제공하는 음성정보가 포함될 수 있다. 예를 들어, 화상 통화 관련 동영상 파일은 사용자 단말(200)에 내장된 카메라 및 마이크 중 적어도 하나를 이용하여 통화자의 의사소통을 저장한 파일을 의미할 수 있다.
- [40] 화상 통화 중에 나오는 모든 언어에 대해 번역 서비스를 제공하기 위해서는 먼저 원어의 인식이 요구된다. 이에 따라, 추출부(320)는 화상 통화 관련 동영상 파일을 영상 파일과 음성 파일로 분리 생성한 다음, 영상 파일 및 음성 파일 중 적어도 하나로부터 원어정보를 추출할 수 있다.
- [41] 이하에서 설명되는 원어정보는 화상 통화 관련 동영상 내에 포함된 음성, 수화 등과 같은 의사소통 수단으로부터 추출된 정보로서, 원어정보는 음성 또는 텍스트로 추출될 수 있다.
- [42] 이하에서는 설명의 편의상 음성으로 구성된 원어정보를 음성 원어정보라

하기로 하고, 텍스트로 구성된 원어정보를 텍스트 원어정보라하기로 한다. 예를 들어 화상 통화 관련 동영상에 나온 인물(통화자)가 영어로 'Hello'라는 음성을 발화한 경우, 음성 원어정보는 통화자가 발화한 음성 'Hello'이며, 텍스트 원어정보는 'Hello' 텍스트 자체를 의미한다. 이하에서는 먼저 음성 파일로부터 음성 원어정보를 추출하는 방법에 대해 설명하도록 한다.

- [43] 음성 파일 내에는 다양한 사용자의 음성이 담겨 있을 수 있으며, 이러한 다양한 음성이 동시에 출력되면 식별이 어려울 수 있고, 이로 인해 번역 정확도 또한 낮아질 수 있다. 이에 따라, 추출부(320)는 음성 파일에 대해 주파수 대역 분석 프로세스를 적용하여 사용자(통화자) 각각에 대한 음성 원어정보를 추출할 수 있다.
- [44] 음성은 성별, 연령대, 발음 톤, 발음 세기 등에 따라 개개인마다 다를 수 있으며, 주파수 대역을 분석하면 해당 특성들을 파악함으로써 음성 별로 개별적인 식별이 가능할 수 있다. 이에 따라, 추출부(320)는 음성 파일의 주파수 대역을 분석하고, 분석 결과를 기반으로 화상 통화 중에 등장하는 통화자 각각에 대한 음성을 분리함으로써 음성 원어정보를 추출할 수 있다.
- [45] 추출부(320)는 음성 원어정보에 대해 음성인식 프로세스를 적용함으로써 음성을 텍스트로 변환한 텍스트 원어정보를 생성할 수 있다. 추출부(150)는 음성 원어정보 및 텍스트 원어정보를 통화자 별로 나누어 저장할 수 있다.
- [46] 주파수 대역 분석 프로세스를 통해 사용자 각각에 대한 음성 원어정보를 추출하는 방법 및 음성인식 프로세스를 통해 음성 원어정보로부터 텍스트 원어정보를 생성하는 방법 등은 알고리즘 또는 프로그램 형태의 데이터로 구현되어 방송 장치(200) 내에 기 저장될 수 있으며, 추출부(320)는 기 저장된 데이터를 이용하여 원어정보를 분리 생성할 수 있다.
- [47] 한편, 화상 통화 중에 특정 통화자는 수화를 사용할 수도 있다. 이 경우 음성 파일로부터 음성 원어정보를 추출한 다음 음성 원어정보로부터 텍스트 원어정보를 생성하는 진술한 방법과 달리, 추출부(320)는 영상 파일로부터 바로 텍스트 원어정보를 추출할 수 있다. 이하에서는 영상 파일로부터 텍스트 원어정보를 추출하는 방법에 대해 설명하도록 한다.
- [48] 추출부(320)는 영상 파일에 대해 영상 처리 프로세스를 적용하여 수화 패턴을 검출하고, 검출된 수화 패턴을 기반으로 텍스트 원어정보를 생성할 수 있다.
- [49] 영상 처리 프로세스의 적용 여부는 자동 또는 수동으로 설정될 수 있다. 예를 들어 통신부(310)를 통해 사용자 단말(100)로부터 수화번역 요청명령을 입력 받은 경우, 추출부(320)가 영상 처리 프로세스를 통해 수화 패턴을 검출할 수 있다. 또 다른 예로, 추출부(320)는 자동으로 영상 파일에 대해 영상 처리 프로세스를 적용하여 영상 파일 상에 수화 패턴이 존재하는지 여부를 판단할 수 있는 등 제한은 없다.
- [50] 영상 처리 프로세스를 통해 수화 패턴에 검출하는 방법은 알고리즘 또는 프로그램 형태의 데이터로 구현되어 방송 장치(300) 내에 기 저장될 수 있으며,

추출부(320)는 기 저장된 데이터를 이용하여 영상 파일 상에 포함된 수화 패턴을 검출하고, 검출한 수화 패턴으로부터 텍스트 원어정보를 생성할 수 있다.

- [51] 추출부(320)는 원어정보를 특정 인물정보와 매핑하여 저장할 수 있다.
- [52] 예를 들어, 추출부(320)는 특정 음성을 송신한 사용자 단말(100)을 식별한 다음, 해당 사용자 단말(100)에 대해 기 설정된 ID 또는 사용자(통화자)가 기 설정한 닉네임 등을 원어정보에 매핑함으로써, 복수의 사용자가 동시에 음성을 발화하더라도 어떠한 사용자가 어떠한 발언을 했는지 시청자가 정확히 파악할 수 있게끔 한다.
- [53] 또 다른 예로, 하나의 화상 통화 관련 동영상 파일 내에 복수의 통화자가 포함된 경우, 추출부(320)는 미리 설정된 방법에 따라 또는 화상 통화 관련 동영상 파일로부터 검출되는 통화자의 특성에 따라 적응적으로 인물정보를 설정할 수도 있다. 일 실시예로, 추출부(320)는 주파수 대역 분석 프로세스를 통해 음성을 발화한 등장인물의 성별, 나이대 등을 파악할 수 있으며, 파악 결과를 기반으로 가장 적합한 것으로 판단되는 등장인물의 이름을 임의로 설정하여 매핑할 수 있다.
- [54] 제어부(340)는 통신부(310)를 제어하여 사용자 단말(100) 및 시청자 단말(200)에 인물정보를 매핑한 원어정보 및 번역정보를 송출할 수 있으며, 이에 사용자 및 시청자들은 보다 손쉽게 발언자가 누구인지 식별할 수 있다. 제어부(340)에 관한 구체적인 설명은 후술하도록 한다.
- [55] 도 2를 참조하면, 번역 장치(300)에는 번역부(330)가 마련될 수 있다. 번역부(330)는 원어정보를 사용자 또는 시청자가 원하는 언어로 번역하여 번역정보를 생성할 수 있다. 사용자 또는 시청자로부터 입력 받은 언어로 번역정보를 생성함에 있어, 번역부(330)는 번역 결과를 텍스트로 생성할 수도 있고 음성으로 생성할 수도 있다. 실시예에 따른 방송 시스템(1)은 원어정보 및 번역정보 각각을 음성 또는 텍스트로 제공함으로써 청각장애인과 시각장애인도 화상 통화 서비스를 이용할 수 있을 뿐만 아니라, 시청까지 가능하게 하는 장점이 있다.
- [56] 이하에서는 설명의 편의상 원어정보를 사용자 또는 시청자가 요청한 언어로 번역한 것을 번역정보라 하기로 하며, 번역정보 또한 원어정보와 같이 음성 또는 텍스트 형태로 구성될 수 있다. 이때 텍스트로 구성된 번역정보에 대해서는 텍스트 번역정보라 하고, 음성으로 구성된 번역정보에 대해서는 음성 번역정보라 하기로 한다.
- [57] 음성 번역정보는 특정 음성으로 더빙된 음성 정보로서, 번역부(330)는 미리 설정된 음성 또는 사용자가 설정한 톤으로 더빙한 음성 번역정보를 생성할 수 있다. 사용자 마다 청취를 원하는 톤은 다를 수 있다. 예를 들어, 특정 시청자는 남자 톤의 음성 번역정보를 원할 수 있고 다른 시청자는 여자 톤의 음성 번역정보를 원할 수 있다. 이에 따라, 번역부(330)는 시청자들의 시청을 보다 편안하게 할 수 있도록 다양한 톤으로 음성 번역정보를 생성할 수 있다. 또는

- 번역부(330)는 발화자의 음성을 분석한 결과를 토대로 발화자의 음성과 유사한 음성 톤으로 음성 번역정보를 생성할 수 있는 등 제한은 없다.
- [58] 번역 방법 및 번역시 사용되는 음성 톤 설정 방법은 알고리즘 또는 프로그램 형태의 데이터는 방송 장치(300) 내에 기 저장될 수 있으며, 번역부(330)는 기 저장된 데이터를 이용하여 번역을 수행할 수 있다.
- [59] 도 2를 참조하면, 방송 장치(300)에는 방송 장치(300)내 구성요소들의 전반적인 동작을 제어하는 제어부(340)가 마련될 수 있다.
- [60] 제어부(340)는 각종 연산의 처리가 가능한 MCU(Micro Control Unit)와 같은 프로세서와, 방송 장치(300)의 동작을 제어하기 위한 제어 프로그램 또는 제어 데이터를 기억하거나 또는 프로세서가 출력하는 제어 명령 데이터 또는 영상 데이터를 임시로 기억하는 메모리로 구현될 수 있다.
- [61] 이때, 프로세서 및 메모리는 방송 장치(300)에 내장된 시스템 온 칩(System On Chip, SOC)에 집적될 수 있다. 다만, 방송 장치(300)에 내장된 시스템 온 칩이 하나만 존재하는 것은 아닐 수 있으므로, 하나의 시스템 온 칩에 집적되는 것으로 제한되는 것은 아니다.
- [62] 메모리는 S램, D램 등의 휘발성 메모리(임시 저장 메모리라 지칭되기도 함)과, 플래시 메모리, 롬(Read Only Memory), 이퍼롬(Erasable Programmable Read Only Memory: EPROM), 이이퍼롬(Electrically Erasable Programmable Read Only Memory: EEPROM) 등의 비휘발성 메모리를 포함할 수 있다. 다만, 이에 한정되는 것은 아니며, 당업계에 알려져 있는 임의의 다른 형태로 구현될 수도 있다.
- [63] 일 실시예로, 비휘발성 메모리에는 방송 장치(300)의 동작 제어하기 위한 제어 프로그램 및 제어 데이터가 저장될 수 있으며, 휘발성 메모리에는 비휘발성 메모리로부터 제어 프로그램 및 제어 데이터를 불러와 임시로 저장되거나, 프로세서가 출력하는 제어 명령 데이터 등이 임시로 저장될 수 있는 등 제한은 없다.
- [64] 제어부(340)는 메모리에 저장된 데이터를 기반으로 제어신호를 생성하고, 생성한 제어신호를 통해 방송 장치(300) 내 구성요소의 전반적인 동작을 제어할 수 있다.
- [65] 예를 들어, 제어부(340)는 제어신호를 통해 통신부(310)를 제어하여 화상 통화를 지원할 수 있다. 또한, 제어부(340)는 제어 신호를 통해 추출부(320)가 화상 통화 관련 파일, 예를 들어 동영상 파일로부터 영상 파일과 음성 파일을 생성하고, 영상 파일과 음성 파일 중 적어도 하나로부터 원어정보를 추출하도록 제어할 수 있다.
- [66] 제어부(340)는 통신부(310)를 제어하여 화상 통화 관련 동영상 파일에, 원어정보 및 번역정보 중 적어도 하나를 매핑한 통번역 동영상을 화상 통화 중인 다른 사용자 단말과 채팅 방에 접속 중인 시청자 단말(200), 즉 채팅 방에 접속 중인 단말에 전송함으로써 다양한 국가의 통화자, 시청자들 간에 의사소통이

원활히 이루어지게끔 할 수 있다.

- [67] 전술한 바와 같이, 통번역 동영상에는 원어정보 또는 번역정보 만이 매핑되어 있을 수도 있고, 원어정보 및 번역정보가 함께 매핑되어 있을 수 있다.
- [68] 예를 들어 통번역 동영상 내에 텍스트 원어정보 및 텍스트 번역정보 만이 매핑되어 있는 경우, 통번역 동영상에는 통화자가 발화할 때마다 해당 발화에 관한 텍스트 원어정보와 텍스트 번역정보가 자막으로서 포함될 수 있다. 또 다른 예로 통번역 동영상 내에 음성 번역정보 및 텍스트 번역정보가 매핑되어 있는 경우 통번역 동영상에는 통화자가 발화할 때 마다 특정 국가의 언어로 번역된 음성 번역정보가 더빙되어 포함될 수 있으며, 텍스트 번역정보가 자막으로서 포함될 수 있다.
- [69] 한편, 제어부(340)는 통신부(310)를 통해 사용자 단말(200)로부터 수신한 설정 명령 또는 미리 설정된 방법을 기반으로 화상 통화 서비스 및 번역 서비스를 제공하는 방법을 변경할 수 있다.
- [70] 예를 들어 통신부(310)를 통해 사용자 단말(100)로부터 화상 통화자 수 설정 명령 또는 시청자 수 설정 명령을 수신한 경우, 제어부(340)는 해당 명령에 맞추어 채팅 방으로의 사용자 단말(100) 및 시청자 단말(200)의 접속을 제한할 수 있다.
- [71] 또 다른 예로, 통신부(310)를 통해 사용자 단말(100) 또는 시청자 단말(200)로부터 별도의 텍스트 데이터 또는 이미지 데이터가 수신되면, 제어부(340)는 수신한 텍스트 데이터 또는 이미지 데이터를 원어/번역정보와 함께 송출함으로써 사용자 및 시청자 간에 의견 교환이 보다 확실히 이루어지게끔 할 수 있다.
- [72] 또 다른 예로, 통신부(310)를 통해 사용자 단말(100)로부터 발언권 설정 명령, 예를 들어 발언 제한 명령 또는 발언 순서에 대한 명령이 수신되면, 제어부(340)는 해당 명령에 맞추어 복수의 사용자 단말(100) 중 발언권이 있는 사용자 단말에 대한 통번역 동영상 만을 전송할 수 있다. 또는, 제어부(340)는 해당 명령에 맞추어 발언권에 대한 내용이 포함된 팝업 메시지를 통번역 동영상과 함께 전송할 수 있는 등 구현 방법에 제한은 없다.
- [73] 사용자 단말(100) 및 시청자 단말(200)에는 후술할 바와 같이 화상 통화 서비스 및 번역 서비스를 지원하며 전술한 서비스를 지원함에 있어 사용자 및 시청자 개개인의 성향에 맞추어 다양한 설정을 가능하게 하는 어플리케이션이 미리 저장될 수 있으며, 사용자 및 시청자는 해당 어플리케이션을 이용하여 다양한 설정이 가능하다. 이하에서는 사용자 단말(100)에 대해 설명하도록 한다.
- [74] 도 2를 참조하면, 사용자 단말(100)은 사용자에게 각종 정보를 시각적으로 제공하는 디스플레이(110), 사용자에게 각종 정보를 청각적으로 제공하는 스피커(120), 통신망을 통해 외부 기기와 각종 데이터를 주고 받는 단말 통신부(130), 사용자 단말(100) 내 구성요소의 전반적인 동작을 제어하여 화상 통화 서비스를 지원하는 단말 제어부(140)를 포함할 수 있다.

- [75] 여기서, 단말 통신부(130), 단말 제어부(140)는 각각 별도로 구현되거나 또는 하나의 시스템 온 칩(System On Chip, SOC)으로 통합 구현될 수 있는 등 구현방법에는 제한이 없다. 이하에서는 사용자 단말(100)의 각 구성요소에 대해 설명하도록 한다.
- [76] 사용자 단말(100)에는 사용자에게 각종 정보를 시각적으로 제공하는 디스플레이(110)가 마련될 수 있다. 일 실시예에 따르면, 디스플레이(110)는 LCD(Liquid Crystal Display), LED(Light Emitting Diode), PDP(Plasma Display Panel), OLED(Organic Light Emitting Diode), CRT(Cathode Ray Tube) 등으로 구현될 수 있으나, 이에 한하지 않으며 제한은 없다. 한편, 디스플레이(110)가 터치 스크린 패널(Touch Screen Panel, TSP) 타입으로 구현된 경우에는 사용자는 디스플레이(110)의 특정 영역을 터치함으로써 각종 설정 명령을 입력할 수 있다.
- [77] 디스플레이(110)는 화상 통화와 관련한 동영상을 표시할 수 있을 뿐만 아니라, 디스플레이(110) 상에 표시된 유저 인터페이스를 통해 각종 제어 명령을 입력 받을 수도 있다.
- [78] 이하에서 설명되는 유저 인터페이스는 사용자와 사용자 단말(100) 간의 각종 정보, 명령의 교환 동작이 보다 편리하게 수행되도록 디스플레이(110) 상에 표시되는 화면을 그래픽으로 구현한 그래픽 유저 인터페이스일 수 있다.
- [79] 예를 들어, 그래픽 유저 인터페이스는, 디스플레이(110)를 통해 표시되는 화면 상에서 일부 영역에는 사용자로부터 각종 제어 명령을 손쉽게 입력 받기 위한 아이콘, 버튼 등이 표시되고, 또한 다른 일부 영역에는 적어도 하나의 위젯을 통해 각종 정보가 표시되도록 구현될 수 있는 등 제한은 없다.
- [80] 예를 들어 디스플레이(110) 상에는 도 3에 도시된 바와 같이 화상 통화 중인 다른 4명의 사용자에게 대한 동영상이 일정 영역에 분할 표시되도록 구성되어 있고, 번역 명령을 입력할 수 있는 아이콘(I1), 화상 통화 서비스 상태에 대한 정보를 제공하는 이모티콘(I2), 접속 중인 시청자 수를 알려주는 이모티콘(I3), 각종 설정 명령을 입력할 수 있는 아이콘(I4)이 포함되도록 구성된 그래픽 유저 인터페이스가 표시될 수 있다.
- [81] 단말 제어부(140)는 제어 신호를 통해 디스플레이(110) 상에 도 3에 도시된 바와 같은 그래픽 유저 인터페이스가 표시되도록 제어할 수 있다. 유저 인터페이스를 구성하는 위젯, 아이콘, 이모티콘 등의 표시방법, 배치방법 등은 알고리즘 또는 프로그램 형태의 데이터로 구현되어, 사용자 단말(100) 내 메모리에 또는 방송 장치(300) 내 메모리에 미리 저장될 수 있으며, 단말 제어부(140)는 미리 저장된 데이터를 이용하여 제어 신호를 생성하고, 생성한 제어 신호를 통해 그래픽 유저 인터페이스가 표시되도록 제어할 수 있다. 단말 제어부(140)에 대한 구체적인 설명은 후술하도록 한다.
- [82] 한편 도 2를 참조하면, 사용자 단말(100)에는 각종 사운드를 출력할 수 있는 스피커(120)가 마련될 수 있다. 스피커(120)는 사용자 단말(100)의 일 면에 마련되어, 화상 통화와 관련한 동영상 파일에 포함된 각종 사운드를 출력할 수

- 있다. 스피커(120)는 기 공지된 다양한 종류의 사운드 출력장치를 통해 구현될 수 있으며, 제한은 없다.
- [83] 사용자 단말(100)에는 통신망을 통해 외부 기기와 각종 데이터를 주고 받는 단말 통신부(130)가 마련될 수 있다.
- [84] 단말 통신부(130)는 무선통신망 또는 유선통신망을 통해 외부 기기와 각종 데이터를 주고 받을 수 있다. 여기서 무선통신망 및 유선통신망에 대한 구체적인 설명은 전술하였는 바 생략하도록 한다.
- [85] 단말 통신부(130)는 통신망을 통해 장치(300)와 연결되어 채팅 방을 개설할 수 있으며, 채팅 방에 접속한 다른 사용자 단말과 화상 통화에 관한 동영상 파일을 실시간으로 주고 받아 화상 통화 서비스를 제공할 뿐만 아니라, 채팅 방에 접속한 시청자 단말(300)에도 화상 통화에 관한 동영상 파일을 송신함으로써 방송 서비스를 제공할 수 있다.
- [86] 도 2를 참조하면, 사용자 단말(100)에는 사용자 단말(100)의 전반적인 동작을 제어하는 단말 제어부(140)가 마련될 수 있다.
- [87] 단말 제어부(140)는 각종 연산의 처리가 가능한 MCU와 같은 프로세서와, 사용자 단말(100)의 동작을 제어하기 위한 제어 프로그램 또는 제어 데이터를 기억하거나 또는 프로세서가 출력하는 제어 명령 데이터 또는 영상 데이터를 임시로 기억하는 메모리로 구현될 수 있다.
- [88] 이때, 프로세서 및 메모리는 사용자 단말(100)에 내장된 시스템 온 칩에 집적될 수 있다. 다만, 사용자 단말(100)에 내장된 시스템 온 칩이 하나만 존재하는 것은 아닐 수 있으므로, 하나의 시스템 온 칩에 집적되는 것으로 제한되는 것은 아니다.
- [89] 메모리는 S램, D램 등의 휘발성 메모리(임시 저장 메모리라 지칭되기도 함)과, 플래시 메모리, 롬, 이퍼롬, 이이퍼롬 등의 비휘발성 메모리를 포함할 수 있다. 다만, 이에 한정되는 것은 아니며, 당업계에 알려져 있는 임의의 다른 형태로 구현될 수도 있다.
- [90] 일 실시예로, 비휘발성 메모리에는 사용자 단말(100)의 동작 제어하기 위한 제어 프로그램 및 제어 데이터가 저장될 수 있으며, 휘발성 메모리에는 비휘발성 메모리로부터 제어 프로그램 및 제어 데이터를 불러와 임시로 저장되거나, 프로세서가 출력하는 제어 명령 데이터 등이 임시로 저장될 수 있는 등 제한은 없다.
- [91] 단말 제어부(140)는 메모리에 저장된 데이터를 기반으로 제어신호를 생성하고, 생성한 제어신호를 통해 사용자 단말(100) 내 구성요소의 전반적인 동작을 제어할 수 있다.
- [92] 예를 들어, 단말 제어부(140)는 제어신호를 통해 디스플레이(110) 상에 다양한 정보가 표시되도록 제어할 수 있다. 단말 통신부(130)를 통해 4명의 사용자로부터 영상 파일과 원어정보 및 번역정보 중 적어도 하나가 매핑된 동영상 파일을 각각 수신하면, 단말 제어부(140)는 도 3에 도시된 바와 같이

디스플레이 상에 4개의 화면으로 분할하여 사용자 각각에 대한 동영상 파일이 표시되도록 제어할 수 있다.

- [93] 또한, 단말 제어부(140)는 화상 통화 서비스에 대한 각종 설정 명령을 입력 받을 수 있는 유저 인터페이스를 디스플레이(110) 상에 표시되도록 제어할 수 있으며, 해당 유저 인터페이스를 통해 입력 받은 설정 명령을 기반으로 유저 인터페이스 구성을 변경할 수 있다.
- [94] 예를 들어 사용자가 도 3에 도시된 아이콘(I4)를 클릭한 경우, 단말 제어부(140)는 디스플레이(110) 상에 화상 통화 관련 동영상에 표시되는 영역을 도 4에 도시된 바와 같이 축소되고, 사용자로부터 각종 설정 명령을 입력 받을 수 있는 아이콘이 나타나도록 구성된 유저 인터페이스가 표시되도록 제어할 수 있다. 구체적으로 도 4를 참조하면, 단말 제어부(140)는 화상 통화자 초대 명령, 시청자 초대 명령, 번역어 선택 명령, 발언권 설정 명령, 채팅창 활성화 명령, 자막 설정 명령, 통화자 수 설정 명령, 시청자 수 설정 명령, 기타 설정 등을 입력 받을 수 있는 아이콘이 포함된 유저 인터페이스가 디스플레이(110) 상에 표시되도록 제어할 수 있으며, 입력 가능한 설정 명령이 전술한 예로 한정되는 것은 아니다.
- [95] 일 실시예로 사용자가 화상 통화자 초대 아이콘을 클릭하여 다른 사용자를 초대하는 경우, 단말 제어부(140)는 초대된 사용자 수에 맞추어 화상 통화 관련 동영상에 표시되는 영역을 추가 분할할 수 있다.
- [96] 다른 일 실시예로 사용자가 발언권 설정 아이콘을 클릭하는 경우, 단말 제어부(140)는 다양한 방법을 통해 발언권을 가진 사용자에 대한 동영상에 강조되도록 표시할 수 있다.
- [97] 예를 들어 단말 제어부(140)는 도 5에 도시된 바와 같이 발언권을 가진 사용자에 대한 통번역 동영상에 다른 사용자에 대한 동영상 보다 크게 설정되도록 구현된 유저 인터페이스가 디스플레이(110) 상에 표시되도록 제어할 수 있다. 또 다른 예로, 단말 제어부(140)는 도 6에 도시된 바와 같이 발언권을 가진 사용자에 대한 통번역 동영상만 디스플레이(110) 상에 표시되도록 제어할 수도 있다.
- [98] 이외에도, 단말 제어부(140)는 다양한 방법을 통해 발언권을 가진 사용자에 대한 동영상과 발언권을 가지지 않은 사용자에 대한 동영상이 다르게 표시되도록 제어할 수 있는 등 제한은 없다.
- [99] 전술한 유저 인터페이스를 구성하는 방법의 경우 프로그램 또는 알고리즘 형태의 데이터로 구현되어 사용자 단말(100) 내에 미리 저장되거나 또는 방송 장치(300) 내에 미리 저장될 수 있다. 방송 장치(300) 내에 미리 저장된 경우, 단말 제어부(140)는 단말 통신부(110)를 통해 방송 장치(300)로부터 위 데이터를 수신한 다음, 이를 기반으로 디스플레이(110) 상에 유저 인터페이스가 표시되도록 제어할 수 있다.
- [100] 시청자 단말(200)의 경우 사용자 단말(100)과 구성이 동일하므로 이에 대한

구체적인 설명은 생략하도록 한다. 한편, 시청자 단말(200)과 사용자 단말(100)의 디스플레이 상에 표시되는 유저 인터페이스는 같거나 또는 다를 수 있다. 예를 들어 시청자 단말(200)의 시청자는 화상 통화에 참여할 수 없기 때문에 화상 통화자 초대 명령을 입력할 수 있는 아이콘은 유저 인터페이스 상에서 제외될 수 있다.

- [101] 이외에도 시청자 단말(200) 상에서 구현되는 유저 인터페이스와 사용자 단말(100) 상에 구현되는 유저 인터페이스는 사용자 또는 시청자의 편의를 고려하여 다르게 구성될 수 있으며 제한은 없다. 이하에서는 방송 장치의 동작에 대해서 간단하게 설명하도록 한다.
- [102] 도 7은 일 실시예에 따른 방송 장치의 동작 흐름도를 개략적으로 도시한 도면이다.
- [103] 방송 장치는 사용자 단말과 시청자 단말 간을 연결하여 화상 통화 서비스를 제공할 수 있다. 이에, 방송 장치는 화상 통화 서비스를 제공 중에 있어, 화상 통화 중인 사용자 단말로부터 화상 통화 데이터를 수집할 수 있다. 화상 통화 데이터는 사용자 단말에 내장된 카메라 및 마이크 중 적어도 하나를 이용하여 생성된 데이터로서, 전송한 카메라 및 마이크 중 적어도 하나를 이용하여 사용자의 의사소통이 저장된 데이터를 의미할 수 있다.
- [104] 방송 장치는 화상 통화 관련 동영상으로부터 영상 파일과 음성 파일을 각각 분리 생성할 수 있으며(700), 생성한 영상 파일 및 음성 파일 중 적어도 하나를 이용하여 사용자 각각에 대한 원어정보를 추출할 수 있다(710).
- [105] 여기서, 원어정보라 함은 화상 통화 관련 동영상 내에 포함된 의사소통 수단을 음성 및 텍스트 중 적어도 하나의 형태로 나타낸 정보로서, 특정 국가의 언어로 번역하기 전의 정보에 해당한다.
- [106] 방송 장치는 화상 통화 관련 동영상 내에 등장하는 통화자가 사용하는 의사소통 수단에 따라 영상 파일 및 음성 파일 중 전부를 이용하거나 또는 하나만을 이용하여 원어정보를 추출할 수 있다.
- [107] 예를 들어 화상 통화 관련 동영상 내에 등장하는 통화자 중 어느 하나가 음성을 이용하여 화상 통화를 함과 동시에 다른 통화자는 수화를 이용하여 화상 통화를 하고 있는 경우, 방송 장치는 영상 파일로부터 수화 패턴을, 음성 파일로부터 음성을 식별하여 원어정보를 추출할 수 있다.
- [108] 또 다른 예로 통화자들이 음성만을 이용하여 화상 통화 중인 경우 방송 장치는 음성 파일만을 이용하여 원어정보를 추출할 수 있고, 또 다른 예로 통화자들이 수화만을 이용하여 대화 중인 경우, 방송 장치는 영상 파일만을 이용하여 원어정보를 추출할 수 있다.
- [109] 방송 장치는 원어정보로부터 통화자 또는 시청자의 요청에 따라 개별적으로 번역정보를 생성할 수 있으며(720), 채팅 방에 접속 중인 단말, 사용자 단말 및 시청자 단말 전부에 원어정보 및 번역정보 중 적어도 하나가 매핑된 통번역 동영상을 전송할 수 있다.

- [110] 방송 장치는 자체적으로 원어정보를 번역하여 번역정보를 생성할 수도 있고, 연산 과부하를 방지하고자 번역 프로세스를 처리하는 외부 서버에 원어정보를 전송하고, 번역정보를 수신하여 제공할 수도 있는 등 구현형태에는 제한이 없다.
- [111] 방송 장치는 원어 정보 및 번역 정보 중 적어도 하나를 전송할 수 있다(730). 이때, 방송 장치는 화상 통화 관련 동영상에 원어 정보 및 번역 정보 중 적어도 하나가 매핑된 통번역 동영상을 전송함으로써 통화자 간의 의사 소통이 원활히 이루어질 수 있을 뿐만 아니라, 시청자들 또한 통화자들의 의견을 정확히 파악하게끔 한다.
- [112] 또한 전송한 바와 같이 실시예에 따른 유저 인터페이스는 텍스트 전송 기능을 지원하여, 통화자 또는 시청자들이 자신의 의견을 텍스트로 전송하게끔 함으로써 의사 소통이 보다 원활히 이루어지도록 하며, 이외에도 발언권 설정 기능을 지원하여 원활하게 의견 교환이 이루어지게끔 도울 수 있다.
- [113] 명세서에 기재된 실시예와 도면에 도시된 구성은 개시된 발명의 바람직한 일 예에 불과할 뿐이며, 본 출원의 출원시점에 있어서 본 명세서의 실시예와 도면을 대체할 수 있는 다양한 변형 예들이 있을 수 있다.
- [114] 또한, 본 명세서에서 사용한 용어는 실시예를 설명하기 위해 사용된 것으로, 개시된 발명을 제한 및/또는 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는다.
- [115] 또한, 본 명세서에서 사용한 "제1", "제2" 등과 같이 서수를 포함하는 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되지는 않으며, 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. "및/또는"이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.
- [116] 또한, 본 명세서 전체에서 사용되는 "~부(unit)", "~기", "~블록(block)", "~부재(member)", "~모듈(module)" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미할 수 있다. 예를 들어, 소프트웨어, FPGA 또는 ASIC과 같은 하드웨어를 의미할 수 있다. 그러나, "~부", "~기", "~블록", "~부재", "~모듈" 등이 소프트웨어 또는 하드웨어에 한정되는 의미는 아니며, "~부", "~기", "~블록", "~부재", "~모듈" 등은 접근할 수 있는 저장 매체에 저장되고 하나 또는 그 이상의 프로세서에 의하여 수행되는 구성일 수 있다.
- [117] [부호의 설명]

WO 2021/118180**PCT/KR2020/017734**

- [118] 1: 방송 시스템
- [119] 100: 사용자 단말
- [120] 200: 시청자 단말
- [121] 300: 방송 장치