

명세서

발명의 명칭: 사용자 단말 및 그 제어방법

기술분야

- [1] 동영상에 대한 번역 서비스를 제공하는 사용자 단말 및 그 제어방법에 관한 것이다.

배경기술

- [2] IT 기술의 발달에 따라 다양한 종류의 동영상 콘텐츠 전송/공유가 사용자 간에 손쉽게 이루어지고 있으며 특히, 글로벌 트렌드에 맞추어 사용자들은 국내 동영상 콘텐츠 뿐만 아니라, 다양한 언어로 제작된 해외 동영상 콘텐츠를 전송/공유하고 있는 실정이다.
- [3] 다만, 많은 동영상 콘텐츠가 제작됨에 따라 모든 동영상 콘텐츠에 대해 번역이 이루어지고 있지는 않으며, 이에 사용자의 편의성을 증대시키고자 실시간 번역 서비스를 제공하는 방법에 대한 연구가 진행 중이다.

발명의 상세한 설명

기술적 과제

- [4] 사용자가 원하는 동영상 콘텐츠에 대한 원문 제공 서비스뿐만 아니라, 번역 제공 서비스를 실시간으로 제공함으로써 사용자가 보다 손쉽게 동영상 콘텐츠를 즐길 수 있고, 동영상 콘텐츠에 다양한 의사소통 수단이 포함되어 있더라도 전부 번역이 가능하며, 음성 및 텍스트 중 적어도 하나를 통해 번역서비스를 제공함으로써 시각장애인 뿐만 아니라 청각장애인도 자유롭게 동영상 콘텐츠를 즐길 수 있도록 하는데 목적이 있다.

과제 해결 수단

- [5] 일 측에 따른 사용자 단말은, 동영상 파일로부터 분리 생성한 영상파일 및 음성파일 중 적어도 하나를 기반으로 등장인물 각각에 대한 원어정보를 추출하는 추출부; 상기 원어정보를 선택된 언어에 따라 번역한 번역정보를 생성하는 번역부; 및 상기 원어정보 및 번역정보 중 적어도 하나를 제공하는 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 등장인물 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] 또한 상기 추출부는, 주파수 대역 분석 프로세스를 통해 상기 음성 파일에 등장하는 등장인물의 연령대 및 성별 중 적어도 하나를 판단하고, 판단 결과를 기반으로 설정한 인물정보를, 상기 원어정보에 매핑하여 저장할 수 있다.
- [10] 일 측에 따른 사용자 단말의 제어방법은, 동영상 파일로부터 분리 생성한 영상파일 및 음성파일 중 적어도 하나를 기반으로 등장인물 각각에 대한 원어정보를 추출하는 단계; 상기 원어정보를 선택된 언어에 따라 번역한 번역정보를 생성하는 단계; 및 상기 원어정보 및 번역정보 중 적어도 하나를 제공하는 단계를 포함할 수 있다.
- [11] 또한 상기 추출하는 단계는, 상기 동영상 파일 내에 포함된 의사소통 수단에 따라 영상 파일 및 음성 파일 중 적어도 하나를 기반으로 등장인물 각각에 대한 원어정보를 추출하는 단계를 포함할 수 있다.
- [12] 또한 상기 추출하는 단계는, 상기 음성 파일에 대해 주파수 대역 분석 프로세스를 적용하여 등장인물 각각에 대한 음성 원어정보를 추출하는 단계; 및 상기 추출한 음성 원어정보에 대해 음성 인식 프로세스를 적용하여 텍스트 원어정보를 생성하는 단계를 포함할 수 있다.
- [13] 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.
- [14] 또한 상기 추출하는 단계는, 주파수 대역 분석 프로세스를 통해 상기 음성 파일에 등장하는 등장인물의 연령대 및 성별 중 적어도 하나를 판단하고, 판단 결과를 기반으로 설정한 인물정보를, 상기 원어정보에 매핑하여 저장하는 단계를 더 포함할 수 있다.

발명의 효과

- [15] 일 실시예에 따른 사용자 단말 및 그 제어방법은, 사용자가 원하는 동영상 콘텐츠에 대한 원문 제공 서비스뿐만 아니라, 번역 제공 서비스를 실시간으로 제공함으로써 사용자가 보다 손쉽게 동영상 콘텐츠를 즐길 수 있다.
- [16] 다른 일 실시예에 따른 사용자 단말 및 그 제어방법은, 동영상 콘텐츠에 다양한 의사소통 수단이 포함되어 있더라도 전부 번역이 가능하며, 음성 및 텍스트 중 적어도 하나를 통해 번역서비스를 제공함으로써 시각장애인 뿐만 아니라 청각장애인도 자유롭게 동영상 콘텐츠를 즐길 수 있다.

도면의 간단한 설명

- [17] 도 1은 일 실시예에 따른 사용자 단말의 외관을 개략적으로 도시한 도면이다.
- [18] 도 2는 일 실시예에 따른 사용자 단말의 제어 블록도를 개략적으로 도시한 도면이다.
- [19] 도 3은 일 실시예에 따른 디스플레이 상에 표시되는 유저 인터페이스 화면을 도시한 도면이다.

- [20] 도 4는 일 실시예에 따른 디스플레이를 통해 원어정보를 제공하는 유저 인터페이스 화면을 도시한 도면이다.
- [21] 도 5 및 도 6은 서로 다른 실시예에 따른 디스플레이를 통해 원어정보 및 번역정보 중 적어도 하나를 제공하는 유저인터페이스 화면을 도시한 도면이다.
- [22] 도 7은 일 실시예에 따른 사용자 단말의 동작 흐름도를 개략적으로 도시한 도면이다.

발명의 실시를 위한 형태

- [23] 도 1은 일 실시예에 따른 사용자 단말의 외관을 개략적으로 도시한 도면이고, 도 2는 일 실시예에 따른 사용자 단말의 제어 블록도를 개략적으로 도시한 도면이다. 또한, 도 3은 일 실시예에 따른 디스플레이 상에 표시되는 유저 인터페이스 화면을 도시한 도면이고, 도 4는 일 실시예에 따른 디스플레이를 통해 원어정보를 제공하는 유저 인터페이스 화면을 도시한 도면이다. 또한, 도 5 및 도 6은 서로 다른 실시예에 따른 디스플레이를 통해 원어정보 및 번역정보 중 적어도 하나를 제공하는 유저인터페이스 화면을 도시한 도면이다. 이하에서는 설명이 중복되는 것을 방지하기 위해 함께 설명하도록 한다.
- [24] 이하에서 설명되는 사용자 단말은 각종 연산 처리가 가능한 프로세서가 내장되어 있고, 디스플레이와 스피커가 내장되어 있어 동영상 파일을 재생할 수 있는 모든 기기를 포함한다.
- [25] 예를 들어, 사용자 단말은 랩탑(laptop), 데스크 탑(desk top), 태블릿 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. 이하에서는 설명의 편의를 위해 전술한 다양한 종류의 사용자 단말 중에서 스마트 폰 형태의 사용자 단말을 일 예로 들어 설명하도록 하나 이에 한정되는 것은 아니며 제한은 없다.
- [26] 도 1 및 도 2를 참조하면, 사용자 단말(100)은 사용자로부터 각종 명령을 입력 받는 입력부(110), 사용자에게 각종 정보를 시각적으로 제공하는 디스플레이(120), 사용자에게 각종 정보를 청각적으로 제공하는 스피커(130), 통신망을 통해 외부 기기와 각종 데이터를 주고 받는 통신부(140), 동영상 파일로부터 생성한 영상파일과 음성파일 중 적어도 하나를 이용하여 원어정보를 추출하는 추출부(150), 사용자가 요청한 언어로 원어정보를 번역하여 번역정보를 생성하는 번역부(160), 사용자 단말(100) 내 구성요소의 전반적인 동작을 제어하여 원어정보 및 번역정보 중 적어도 하나를 제공함으로써 원문/번역 서비스를 제공하는 제어부(170)를 포함할 수 있다.
- [27] 여기서, 통신부(140), 추출부(150), 번역부(160), 및 제어부(170)는 각각 별도로 구현되거나 또는 통신부(140), 추출부(150), 번역부(160), 및 제어부(170) 중 적어도 하나는 하나의 시스템 온 칩(System On Chip, SOC)으로 통합 구현될 수

있는 등 구현방법에는 제한이 없다. 다만, 사용자 단말(100) 내에 시스템 온 칩이 하나만 존재하는 것은 아닐 수 있으므로, 하나의 시스템 온 칩에 집적되는 것으로 한정되는 것은 아니며 구현방법에는 제한이 없다. 이하에서는 사용자 단말(100)의 각 구성요소에 대해 설명하도록 한다.

- [28] 먼저 도 1 및 도 2를 참조하면, 사용자 단말(100)에는 사용자로부터 각종 제어 명령을 입력 받는 입력부(110)가 마련될 수 있다. 예를 들어, 입력부(110)는 도 1에 도시된 바와 같이 하드 키 타입으로 사용자 단말(100)의 일 면에 마련될 수 있다. 이외에도, 디스플레이(120)가 터치 스크린 타입으로 구현되는 경우 디스플레이(120)는 입력부(110)의 기능을 대신 수행할 수 있다.
- [29] 입력부(110)는 사용자로부터 각종 제어 명령을 입력 받을 수 있다. 예를 들어, 입력부(110)는 사용자로부터 동영상 재생 명령을 입력 받을 수 있을 뿐만 아니라, 사용자로부터 번역을 원하는 언어에 대한 설정 명령 및 원문 추출 명령 및 번역 서비스 실행 명령 등을 입력 받을 수 있다. 이외에도, 입력부(110)는 원어정보 및 번역정보에 대한 저장 명령을 입력 받을 수 있는 등 다양한 제어 명령을 입력 받을 수 있으며, 제어부(170)는 입력 받은 제어 명령에 따라 사용자 단말(100) 내 구성요소의 동작을 제어할 수 있다. 원어정보 및 번역정보에 대한 구체적인 설명은 후술하도록 한다.
- [30] 도 1 및 도 2를 참조하면, 사용자 단말(100)에는 사용자에게 각종 정보를 시각적으로 제공하는 디스플레이(120)가 마련될 수 있다. 디스플레이(120)는 도 1에 도시된 바와 같이 사용자 단말(100)의 일 면에 마련될 수 있는데 이에 한정되는 것은 아니고 제한은 없다.
- [31] 일 실시예에 따르면, 디스플레이(120)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, 전술한 바와 같이 디스플레이(120)가 터치 스크린 패널(Touch Screen Panel, TSP) 타입으로 구현된 경우에는 입력부(110)의 기능을 대신할 수도 있다.
- [32] 터치 스크린 패널 타입으로 구현된 경우, 디스플레이(120)는 사용자가 요청한 동영상을 표시할 수 있을 뿐만 아니라, 디스플레이(120) 상에 표시되는 유저 인터페이스를 통해 각종 제어 명령을 입력 받을 수 있다.
- [33] 이하에서 설명되는 유저 인터페이스는 사용자와 사용자 단말(100) 간의 각종 정보, 명령의 교환 동작이 보다 편리하게 수행되도록 디스플레이(120) 상에 표시되는 화면을 그래픽으로 구현한 그래픽 유저 인터페이스일 수 있다.
- [34] 예를 들어, 그래픽 유저 인터페이스는, 디스플레이(120)를 통해 표시되는 화면 상에서 특정 영역에는 사용자로부터 각종 제어 명령을 손쉽게 입력 받기 위한 아이콘, 버튼 등이 표시되고, 다른 영역에는 적어도 하나의 위젯을 통해 각종 정보가 표시되도록 구현될 수 있는 등 제한은 없다.
- [35] 도 3을 참조하면, 디스플레이(120) 상에는 동영상 재생 명령을 입력 받을 수 있는 아이콘(I1), 번역 명령을 입력 받을 수 있는 아이콘(I2) 및 전술한 명령

이외에 각종 설정 명령을 입력 받을 수 있는 아이콘(I3)이 포함된 그래픽 유저 인터페이스가 표시될 수 있다.

- [36] 제어부(170)는 제어 신호를 통해 디스플레이(120) 상에 도 3에 도시된 바와 같은 그래픽 유저 인터페이스가 표시되도록 제어할 수 있다. 유저 인터페이스를 구성하는 위젯, 아이콘 등의 표시방법, 배치방법 등은 알고리즘 또는 프로그램 형태의 데이터로 구현되어, 사용자 단말(100) 내 메모리에 미리 저장될 수 있으며, 제어부(170)는 미리 저장된 데이터를 이용하여 제어 신호를 생성하고, 생성한 제어 신호를 통해 그래픽 유저 인터페이스가 표시되도록 제어할 수 있다. 제어부(170)에 대한 구체적인 설명은 후술하도록 한다.
- [37] 한편 도 2를 참조하면, 사용자 단말(100)에는 각종 사운드를 출력할 수 있는 스피커(130)가 마련될 수 있다. 스피커(130)는 사용자 단말(100)의 일 면에 마련되어, 동영상 파일에 포함된 각종 사운드를 출력할 수 있다. 스피커(130)는 기 공지된 다양한 종류의 사운드 출력장치를 통해 구현될 수 있으며, 제한은 없다.
- [38] 사용자 단말(100)에는 통신망을 통해 외부 기기와 각종 데이터를 주고 받는 통신부(140)가 마련될 수 있다.
- [39] 통신부(140)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.
- [40] 예를 들어, 통신부(140)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.
- [41] 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.
- [42] 통신부(140)는 통신망을 통해 외부에 위치한 서버로부터 동영상을 다운로드 받을 수 있을 뿐만 아니라, 동영상 내에 포함된 국가의 언어를 기반으로 번역된 정보를 동영상과 함께 외부 단말에 전송할 수도 있는 등 송수신 가능한 데이터에는 제한이 없다.
- [43] 도 2를 참조하면, 사용자 단말(100)에는 추출부(150)가 마련될 수 있다.
- [44] 번역 서비스를 제공하기 위해서는 먼저 원어의 인식이 요구된다. 이에 따라, 추출부(150)는 동영상 파일을 영상 파일과 음성 파일로 분리 생성한 다음, 영상

- 파일 및 음성 파일 중 적어도 하나로부터 원어정보를 추출할 수 있다.
- [45] 이하에서 설명되는 원어정보는 동영상 내에 포함된 음성, 수화 등과 같은 의사소통 수단으로부터 추출된 정보를 의미하며 원어정보는 음성 또는 텍스트 형태로 추출될 수 있다. 이하에서는 설명의 편의상 음성으로 구성된 원어정보를 음성 원어정보라 하기로 하고, 텍스트로 구성된 원어정보를 텍스트 원어정보라 하기로 한다. 예를 들어, 동영상에서 등장하는 인물이 영어로 'Hello'라는 음성을 발화하는 경우, 음성 원어정보는 등장인물이 발화한 음성 'Hello'이며, 텍스트 원어정보는 음성 인식 프로세스를 통해 음성 'Hello'를 인식한 다음, 인식 결과를 기반으로 변환한 'Hello' 텍스트 자체를 의미한다.
- [46] 한편 의사소통 수단, 예를 들어 의사소통 수단이 음성인지 또는 수화인지에 따라 원어정보를 추출하는 방법은 다를 수 있다. 이하에서는 먼저 등장인물의 음성이 담긴 음성 파일로부터 음성 원어정보를 추출하는 방법에 대해 설명하도록 한다.
- [47] 음성 파일 내에는 다양한 등장인물의 음성이 담겨 있을 수 있으며, 이러한 다양한 음성이 동시에 출력되면 식별이 어려울 수 있으며, 이로 인해 번역 정확도 또한 낮아질 수 있다. 이에 따라, 추출부(150)는 음성 파일에 대해 주파수 대역 분석 프로세스를 적용하여 등장인물 각각에 대한 음성 원어정보를 추출할 수 있다.
- [48] 음성은 성별, 연령대, 발음 톤, 발음 세기 등에 따라 개개인 마다 다를 수 있으며, 주파수 대역을 분석하면 해당 특성들을 파악함으로써 음성 별로 개별적인 식별이 가능할 수 있다. 이에 따라, 추출부(150)는 음성 파일의 주파수 대역을 분석하고, 분석 결과를 기반으로 동영상 내에 등장하는 등장인물 각각에 대한 음성을 분리함으로써 음성 원어정보를 추출할 수 있다.
- [49] 추출부(150)는 음성 원어정보에 대해 음성인식 프로세스를 적용함으로써 음성을 텍스트로 변환한 텍스트 원어정보를 생성할 수 있다. 추출부(150)는 음성 원어정보 및 텍스트 원어정보를 등장인물 별로 나누어 저장할 수 있다.
- [50] 주파수 대역 분석 프로세스를 통해 등장인물 별로 음성 원어정보를 추출하는 방법 및 음성인식 프로세스를 통해 음성 원어정보로부터 텍스트 원어정보를 생성하는 방법 등은 알고리즘 또는 프로그램 형태의 데이터로 구현되어 사용자 단말(100) 내에 기 저장될 수 있으며, 추출부(150)는 기 저장된 데이터를 이용하여 원어정보를 분리 생성할 수 있다.
- [51] 한편, 동영상 내에 등장하는 인물이 수화를 사용할 수도 있다. 이 경우 음성 파일로부터 음성 원어정보를 추출한 다음 음성 원어정보로부터 텍스트 원어정보를 생성하는 전술한 방법과 달리, 추출부(150)는 영상 파일로부터 바로 텍스트 원어정보를 추출할 수 있다. 이하에서는 영상 파일로부터 텍스트 원어정보를 추출하는 방법에 대해 설명하도록 한다.
- [52] 추출부(150)는 영상 파일에 대해 영상 처리 프로세스를 적용하여 수화 패턴을 검출하고, 검출된 수화 패턴을 기반으로 텍스트 원어정보를 생성할 수 있다.

영상 처리 프로세스의 적용 여부는 자동 또는 수동으로 설정될 수 있다. 예를 들어, 입력부(110) 또는 디스플레이(120)를 통해 사용자로부터 수화번역 요청명령을 입력 받으면 추출부(150)가 영상 처리 프로세스를 통해 수화 패턴을 검출할 수 있다. 또 다른 예로, 추출부(150)는 자동으로 영상 파일에 대해 영상 처리 프로세스를 적용할 수 있는 등 제한은 없다.

- [53] 영상 처리 프로세스를 통해 수화 패턴에 검출하는 방법은 알고리즘 또는 프로그램 형태의 데이터로 구현되어 사용자 단말(100) 내에 기 저장될 수 있으며, 추출부(150)는 기 저장된 데이터를 이용하여 영상 파일 상에 포함된 수화 패턴을 검출하고, 검출한 수화 패턴으로부터 텍스트 원어정보를 생성할 수 있다.
- [54] 추출부(150)는 원어정보를 인물정보와 매핑하여 저장할 수 있다. 인물정보는 미리 설정된 방법에 따라 임의로 설정되거나 또는 동영상 파일로부터 검출되는 등장인물의 특성에 따라 적응적으로 설정될 수도 있다.
- [55] 예를 들어, 추출부(150)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.
- [56] 일 실시예로, 주파수 대역 분석 프로세스를 통해 음성을 분석한 결과 제1 등장인물은 20대 남자인 것으로 파악되고, 제2 등장인물은 40대 여자인 것으로 파악되면, 추출부(150)는 제1 등장인물에 대한 원어정보에 대해서는 인물정보를 '민수'로, 제2 등장인물에 대한 원어정보에 대해서는 인물정보를 '미자'라 설정하여 매핑할 수 있다.
- [57] 또 다른 예로, 제어부(170)는 텍스트 원어정보에서 검출된 인물명을 인물정보로 설정할 수 있는 등 인물정보 설정방법에는 제한이 없다.
- [58] 제어부(170)는 디스플레이(120) 및 스피커(130)를 통해 원어정보를 제공할 때 매핑한 인물정보를 함께 표시할 수 있으며, 번역정보를 제공할 때에도 매핑한 인물정보를 함께 표시할 수 있다. 예를 들어, 제어부(170)는 도 6에 도시된 바와 같이 자체적으로 설정한 인물정보와, 원어정보 및 번역정보가 함께 제공되도록 구성된 유저인터페이스가 디스플레이(120) 상에 표시되도록 제어할 수 있다.
- [59] 한편, 매핑된 인물정보는 사용자가 변경할 수 있으며, 매핑되는 인물정보가 전술한 바와 한정되는 것은 아니다. 예를 들어, 사용자는 입력부(110) 및 터치 스크린 타입으로 구현된 디스플레이(120)를 통해 원하는 인물정보를 설정할 수 있는 등 제한은 없다.
- [60] 도 2를 참조하면, 사용자 단말(100)에는 번역부(160)가 마련될 수 있다. 번역부(160)는 원어정보를 사용자가 원하는 언어로 번역하여 번역정보를 생성할 수 있다. 사용자로부터 입력 받은 국가의 언어로 원어정보를 번역함에 있어, 번역부(160)는 번역 결과를 텍스트로 생성할 수도 있고 음성으로 생성할 수도 있다. 이하에서는 설명의 편의상 원어정보가 다른 국가의 언어로 번역된 정보를 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.

- [61] 음성 번역정보는 특정 음성으로 더빙된 음성 정보로서, 번역부(330)는 미리 설정된 음성 또는 사용자가 설정한 톤으로 더빙한 음성 번역정보를 생성할 수 있다. 사용자마다 청취를 원하는 톤은 다를 수 있다. 예를 들어, 특정 사용자는 남자 톤의 음성 번역정보를 원할 수 있고 다른 사용자는 여자 톤의 음성 번역정보를 원할 수 있다. 또는 번역부(160)는 전술한 주파수 대역 분석 프로세스를 통해 식별된 등장인물의 성별에 맞추어 적응적으로 톤을 설정할 수도 있다.
- [62] 번역 방법 및 번역시 사용되는 음성 톤 설정 방법은 알고리즘 또는 프로그램 형태의 데이터는 사용자 단말(100) 내에 기 저장될 수 있으며, 번역부(160)는 기 저장된 데이터를 이용하여 번역을 수행할 수 있다.
- [63] 도 2를 참조하면, 사용자 단말(100)에는 사용자 단말(100)의 전반적인 동작을 제어하는 제어부(170)가 마련될 수 있다.
- [64] 제어부(170)는 각종 연산의 처리가 가능한 MCU(Micro Control Unit)와 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.
- [65] 이때, 프로세서 및 메모리는 사용자 단말(100)에 내장된 시스템 온 칩(System On Chip, SOC)에 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.
- [66] 메모리는 S램, D램 등의 휘발성 메모리(임시 저장 메모리라 지칭되기도 함)과, 플래시 메모리, 롬(Read Only Memory), 이피롬(Erasable Programmable Read Only Memory: EPROM), 이이피롬(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.
- [67] 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.
- [68] 제어부(170)는 메모리에 저장된 데이터를 기반으로 제어신호를 생성하고, 생성한 제어신호를 통해 사용자 단말(100) 내 구성요소의 전반적인 동작을 제어할 수 있다.
- [69] 제어부(170)는 제어신호를 통해 디스플레이(120) 상에 다양한 정보가 표시되도록 제어할 수 있다. 예를 들어, 제어부(170)는 제어신호를 통해

- 디스플레이(120) 상에 사용자가 재생을 요청한 동영상을 재생할 수 있다. 일 실시예로 사용자가 도 3에 도시된 아이콘(I2)을 터치하면, 제어부(170)는 사용자 단말(100)의 구성요소를 제어하여 사용자가 설정한 국가의 언어로 번역된 텍스트 번역정보 및 음성 번역정보 중 적어도 하나를 제공할 수 있다.
- [70] 예를 들어, 제어부(170)는 디스플레이(120) 상에 동영상과 함께 텍스트 번역정보가 표시되도록 제어할 수 있으며, 제어부(170)는 스피커(130)를 통해 음성 번역정보가 송출되도록 제어할 수도 있다.
- [71] 제어부(170)가 원어정보 및 번역정보를 제공하는 방법은 다양할 수 있다. 예를 들어, 제어부(170)는 도 4에 도시된 바와 같이 동영상에 텍스트 원어정보를 자막으로 매핑시킨 다음, 디스플레이(120) 상에 표시되도록 제어할 수 있다.
- [72] 또 다른 예로, 제어부(170)는 도 5에 도시된 바와 같이 동영상에 텍스트 원어정보 및 텍스트 번역정보를 자막으로 매핑시킨 다음, 디스플레이(120) 상에 함께 표시되도록 제어할 수 있다. 이외에도, 제어부(170)는 텍스트 원어정보가 먼저 표시된 다음, 미리 설정된 간격 뒤에 텍스트 번역정보가 자막으로 표시되도록 제어할 수도 있다.
- [73] 또 다른 예로 제어부(170)는 동영상에서 등장인물이 발화할 때 마다 음성 원어정보가 스피커(130)를 통해 출력된 다음, 미리 설정된 간격 뒤에 특정 음성으로 더빙된 음성 번역정보가 출력되도록 제어할 수 있으며, 이때 음성 원어정보와 음성 번역정보의 출력 크기를 다르게 제어할 수도 있는 등 원문/번역 서비스를 제공하는 방법에는 제한이 없다.
- [74] 전술한 동영상 파일을 영상 파일과 음성 파일로 분리 생성하는 프로세스, 영상 파일과 음성 파일로부터 원어정보를 추출하는 프로세스, 원어정보로부터 번역정보를 생성하는 프로세스는 사용자 단말(100)에서 직접 수행할 수도 있으나 연산 처리 과부하를 방지하고자 외부에 마련된 장치에서 별도로 수행될 수도 있다. 이 경우, 외부에 마련된 장치는 사용자 단말(100)로부터 번역 명령을 전달 받으면, 전술한 프로세스를 수행한 다음, 결과물을 사용자 단말(100)에 송신할 수 있는 등 제한은 없다.
- [75] 이하에서는 동영상에 대한 번역 서비스를 지원하는 사용자 단말의 동작에 대해서 간단하게 설명하도록 한다.
- [76] 도 7는 일 실시예에 따른 사용자 단말의 동작 흐름도를 개략적으로 도시한 도면이다.
- [77] 도 7을 참조하면, 사용자 단말은 동영상 파일을 영상 파일과 음성 파일로 분리 생성할 수 있다(700). 여기서, 동영상 파일은 사용자 단말에 기 저장되어 있는 파일일 수도 있고 통신망을 통해 실시간으로 스트리밍 중인 파일일 수도 있는 등 제한은 없다.
- [78] 예를 들어, 사용자 단말은 내장된 메모리에 저장된 동영상 파일을 읽어들이고, 이를 기반으로 영상 파일과 음성 파일을 생성할 수 있으며 또 다른 예로, 사용자 단말은 통신망을 통해 실시간으로 동영상 파일 데이터를 수신하고, 이를

기반으로 영상 파일과 음성 파일을 생성할 수 있다.

- [79] 사용자 단말은 영상 파일 및 음성 파일 중 적어도 하나를 이용하여 원어정보를 추출할 수 있다(710).
- [80] 여기서, 원어정보라 함은 원본 동영상 파일 내에 포함된 의사소통 수단을 음성 및 텍스트 중 적어도 하나의 형태로 나타낸 정보로서, 특정 국가의 언어로 번역하기 전의 정보에 해당한다.
- [81] 사용자 단말은 동영상 내에 등장하는 인물이 사용하는 의사소통 수단에 따라 영상 파일 및 음성 파일 중 전부를 이용하거나 또는 하나 만을 이용하여 원어정보를 추출할 수 있다.
- [82] 예를 들어 동영상 내에 등장하는 인물 중 어느 하나가 음성을 이용하여 대화를 함과 동시에 다른 인물은 수화를 이용하여 대화를 하고 있는 경우, 사용자 단말은 영상 파일로부터 수화 패턴을, 음성 파일로부터 음성을 식별하여 원어정보를 추출할 수 있다.
- [83] 또 다른 예로 동영상 내에 등장하는 인물들이 음성 만을 이용하여 대화 중인 경우 사용자 단말은 음성 파일 만을 이용하여 원어정보를 추출할 수 있고, 또 다른 예로 동영상 내에 등장하는 인물들이 수화 만을 이용하여 대화 중인 경우, 사용자 단말은 영상 파일 만을 이용하여 원어정보를 추출할 수 있다.
- [84] 사용자 단말은 원어정보를 이용하여 번역정보를 생성할 수 있다(720).
- [85] 이때, 사용자 단말은 자체적으로 원어정보를 번역하여 번역정보를 생성할 수도 있고, 연산 과부하를 방지하고자 실시예에 따른 번역 서비스를 제공하는 외부 서버에 원어정보를 전송하고, 번역정보를 수신하여 제공할 수도 있는 등 구현형태에는 제한이 없다.
- [86] 이외에도, 사용자 단말은 동영상 파일에, 원어정보 및 번역정보를 매핑시킨 다음, 통신망을 통해 외부 단말과 공유함으로써, 다른 사용자들과 함께 콘텐츠를 즐길 수 있다.
- [87] 사용자 단말은 원어정보 및 번역정보 중 적어도 하나를 동영상과 함께 제공할 수 있으며, 제공방법에는 전술한 바와 같이 제한이 없다. 실시예에 따른 사용자 단말은 다양한 국가의 언어로 제작된 동영상 콘텐츠를 사용자가 보다 손쉽게 즐길 수 있게 함과 동시에, 효과적인 언어 교육이 가능하다는 장점이 있다.
- [88] 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.
- [89] 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.

- [90] 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.
- [91] 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.
- [92] [부호의 설명]
- [93] 100: 사용자 단말
- [94] 110: 입력부
- [95] 120: 디스플레이