(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0014812
(43) 공개일자 2014년02월06일

(51) 국제특허분류(Int. Cl.)
G10L 15/00 (2006.01) *G06K 9/20* (2006.01)
(21) 출원번호 10-2012-0081837
(22) 출원일자 2012년07월26일
심사청구일자 없음

(71) 출원인
삼성전자주식회사
경기도 수원시 영통구 삼성로 129 (매탄동)
(72) 발명자
이동열
경기 용인시 기흥구 흥덕2로118번길 25, 804동 1504호 (영덕동, 흥덕마을8단지한국아델리움)
서상범
서울 서초구 신반포로33길 66, 102동 1009호 (잠원동, 신반포청구아파트)
(74) 대리인
윤동열

전체 청구항 수 : 총 21 항

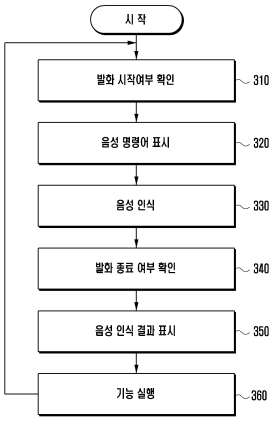
(54) 발명의 명칭 **영상 인식을 이용하여 음성 인식을 하는 방법 및 장치**

(57) 요 약

본 발명은 영상 인식을 기반으로 음성 인식의 시작 및 종료를 보다 정확하게 인식할 수 있는 음성 인식 방법 및 장치에 대한 것으로, 본 발명의 실시예에 따르는 음성 인식 방법은, 음성 인식 모드 전환 전, 제 1 비디오 데이터 또는 제 1 오디오 데이터 중 적어도 하나를 기반으로 발화(發話) 시작을 판단하는 단계; 발화 시작으로 판단한 경우, 음성 인식 모드로 전환하고 사용자의 음성 명령을 포함하는 제 2 오디오 데이터를 생성하는 단계; 및 음성 인식 모드로 전환하고 난 후의 제 2 비디오 데이터 또는 상기 제 2 오디오 데이터 중 적어도 하나를 기반으로 발화 종료를 판단하는 단계를 포함하는 것을 특징으로 한다.

본 발명에 따르면, 사용자는 별도의 제스처를 입력할 필요 없이 입술을 움직이는 것만으로 음성 인식 기능을 실행할 수 있으며, 영상 인식을 통하여 음성 인식의 시작과 종료를 보다 명확하게 판단할 수 있는 효과가 있다.

대 표 도 - 도3



특허청구의 범위

청구항 1

전자 기기가 음성 명령을 인식하는 방법에 있어서,

음성 인식 모드 전환 전의 제 1 비디오 데이터 또는 제 1 오디오 데이터 중 적어도 하나를 기반으로 발화(發話) 시작을 판단하는 단계;

발화 시작으로 판단한 경우, 음성 인식 모드로 전환하고 사용자의 음성 명령을 포함하는 제 2 오디오 데이터를 생성하는 단계; 및

음성 인식 모드로 전환 후의 제 2 비디오 데이터 또는 상기 제 2 오디오 데이터 중 적어도 하나를 기반으로 발화 종료를 판단하는 단계를 포함하는 것을 특징으로 하는 인식 방법.

청구항 2

제 1항에 있어서,

상기 발화 시작을 판단하는 단계 이후에, 화면에 표시된 아이콘을 실행하기 위한 음성 명령어를 표시하는 단계를 더 포함하고,

상기 발화 종료를 판단하는 단계 이후에, 상기 음성 명령어의 표시를 삭제하는 단계를 더 포함하는 것을 특징으로 하는 인식 방법.

청구항 3

제 2항에 있어서, 상기 발화 시작을 판단하는 단계는,

상기 제 1 비디오 데이터에서 사용자의 입술을 구별하고, 상기 입술의 움직임을 기반으로 발화 시작을 판단하는 단계인 것을 특징으로 하는 인식 방법.

청구항 4

제 3항에 있어서, 상기 발화 시작을 판단하는 단계 이후에,

상기 제 1 비디오 데이터 또는 상기 제 1 오디오 데이터를 분석한 결과 값 중 적어도 하나가 발화 시작 판단을 위하여 미리 설정된 기준값(threshold) 이하인 경우, 상기 기준 값을 완화하여 다시 발화 시작을 판단하는 단계를 더 포함하는 것을 특징으로 하는 인식 방법.

청구항 5

제 4항에 있어서, 상기 발화 시작을 판단하는 단계는,

상기 제 1 비디오 데이터에 사용자의 얼굴이 포함되지 않는 경우, 발화 시작이 아닌 것으로 판단하는 단계인 것을 특징으로 하는 인식 방법.

청구항 6

제 5항에 있어서, 상기 발화 종료를 판단하는 단계는,

상기 제 2 오디오 데이터에서 상기 사용자의 음성이 미리 설정된 시간 이상 포함되지 않는 경우, 발화 종료로

판단하는 단계인 것을 특징으로 하는 인식 방법.

청구항 7

제 6항에 있어서, 상기 발화 종료를 판단하는 단계 이후에,

상기 제 2 비디오 데이터 또는 상기 제 2 오디오 데이터를 분석한 결과 값 중 적어도 하나가 발화 종료 판단을 위하여 미리 설정된 기준값(threshold) 이하인 경우, 상기 기준 값을 완화하여 다시 발화 종료를 판단하는 단계를 더 포함하는 것을 특징으로 하는 인식 방법.

청구항 8

제 4항에 있어서, 상기 발화 종료를 판단하는 단계는,

상기 제 2 비디오 데이터에 사용자의 얼굴이 포함되지 않는 경우, 발화 종료로 판단하는 단계인 것을 특징으로 하는 인식 방법.

청구항 9

음성 명령을 인식하는 전자 기기에 있어서,

음성 입력을 수집 및 녹음하는 오디오부;

영상 입력을 수집 및 녹화하는 카메라부; 및

음성 인식 모드 전환 전 제 1 비디오 데이터 또는 제 1 오디오 데이터를 생성하도록 제어하고, 상기 1 비디오 데이터 또는 제 1 오디오 데이터 중 적어도 하나를 기반으로 발화(發話) 시작을 판단하고, 발화 시작으로 판단한 경우, 음성 인식 모드로 전환하여 사용자의 음성 명령을 포함하는 제 2 오디오 데이터 또는 제 2 비디오 데이터를 생성하도록 제어하며, 상기 제 2 비디오 데이터 또는 상기 제 2 오디오 데이터 중 적어도 하나를 기반으로 발화 종료를 판단하는 제어부를 포함하는 것을 특징으로 하는 전자기기.

청구항 10

제 9항에 있어서,

화면을 표시하는 표시부를 더 포함하고,

상기 제어부는, 상기 발화 시작으로 판단한 경우 상기 화면에 표시된 아이콘을 실행하기 위한 음성 명령어를 표시하고, 상기 발화 종료로 판단한 경우 상기 음성 명령어의 표시를 삭제하도록 상기 표시부를 제어하는 것을 특징으로 하는 전자기기.

청구항 11

제 10항에 있어서, 상기 제어부는,

상기 제 1 비디오 데이터에서 사용자의 입술을 구별하고, 상기 입술의 움직임을 기반으로 발화 시작을 판단하는 것을 특징으로 하는 전자기기.

청구항 12

제 11항에 있어서, 상기 제어부는,

상기 제 1 비디오 데이터 또는 상기 제 1 오디오 데이터를 분석한 결과 값 중 적어도 하나가 발화 시작 판단을

위하여 미리 설정된 기준값(threshold) 이하인 경우, 상기 기준 값을 완화하여 다시 발화 시작을 판단하는 것을 특징으로 하는 전자기기.

청구항 13

제 12항에 있어서, 상기 제어부는,

상기 제 1 비디오 데이터에 사용자의 얼굴이 포함되지 않는 경우, 발화 시작이 아닌 것으로 판단하는 것을 특징으로 하는 전자기기.

청구항 14

제 13항에 있어서, 상기 제어부는,

상기 제 2 오디오 데이터에서 상기 사용자의 음성이 미리 설정된 시간 이상 포함되지 않는 경우, 발화 종료로 판단하는 것을 특징으로 하는 전자기기.

청구항 15

제 14항에 있어서, 상기 제어부는,

상기 제 2 비디오 데이터 또는 상기 제 2 오디오 데이터를 분석한 결과 값 중 적어도 하나가 발화 종료 판단을 위하여 미리 설정된 기준값(threshold) 이하인 경우, 상기 기준 값을 완화하여 다시 발화 종료를 판단하는 것을 특징으로 하는 전자기기.

청구항 16

제 15항에 있어서, 상기 제어부는,

상기 제 2 비디오 데이터에 사용자의 얼굴이 포함되지 않는 경우, 발화 종료로 판단하는 것을 특징으로 하는 전자기기.

청구항 17

음성 명령을 인식하기 위한 플랫폼에 있어서,

음성 입력 또는 영상 입력을 수집 및 녹화하여, 음성 인식 모드 전환 전의 제 1 비디오 데이터 또는 제 1 오디오 데이터를 생성하고, 음성 인식 모드 전환 후의 제 2 비디오 데이터 또는 제 2 비디오 데이터를 생성하는 멀티미디어 프레임워크; 및

상기 1 비디오 데이터 또는 제 1 오디오 데이터 중 적어도 하나를 기반으로 발화(發話) 시작을 판단하고, 상기 제 2 비디오 데이터 또는 상기 제 2 오디오 데이터 중 적어도 하나를 기반으로 발화 종료를 판단하는 보이스 프레임워크를 포함하는 것을 특징으로 하는 음성 인식 플랫폼.

청구항 18

제 18항에 있어서,

화면을 표시하기 위한 사용자 인터페이스 프레임워크를 더 포함하고,

상기 사용자 인터페이스 프레임워크는, 상기 보이스 프레임 워크가 발화 시작을 판단한 경우 상기 화면에 표시된 아이콘을 실행하기 위한 음성 명령어를 표시하고, 상기 보이스 프레임 워크가 발화 종료를 판단한 경우 상기

음성 명령어의 표시를 삭제하는 것을 특징으로 하는 음성 인식 플랫폼.

청구항 19

제 18항에 있어서, 상기 보이스 프레임 워크는

상기 제 1 비디오 데이터 또는 상기 제 1 오디오 데이터를 분석한 결과 값 중 적어도 하나가 발화 시작 판단을 위하여 미리 설정된 기준값(threshold) 이하이면 상기 기준 값을 완화하여 다시 발화 시작을 판단하고, 상기 제 2 비디오 데이터 또는 상기 제 2 오디오 데이터를 분석한 결과 값 중 적어도 하나가 발화 종료 판단을 위하여 미리 설정된 기준값(threshold) 이하이면 상기 기준 값을 완화하여 다시 발화 종료를 판단하는 것을 특징으로 하는 음성 인식 플랫폼.

청구항 20

제 19항에 있어서, 상기 보이스 프레임 워크는

상기 제 1 비디오 데이터에서 사용자의 입술을 구별하여 상기 입술의 움직임을 기반으로 발화 시작을 판단하고, 상기 제 2 오디오 데이터에서 상기 사용자의 음성이 미리 설정된 시간 이상 포함되지 않는 경, 발화 종료로 판단하는 것을 특징으로 음성 인식 플랫폼.

청구항 21

제 20항에 있어서, 상기 보이스 프레임 워크는

상기 제 1 비디오 데이터에 사용자의 얼굴이 포함되지 않는 경우 발화 시작이 아닌 것으로 판단하고, 상기 제 2 비디오 데이터에 사용자의 얼굴이 포함되지 않는 경우 발화 종료로 판단하는 것을 특징으로 하는 음성 인식 플랫폼.

명 세 서

기 술 분 야

[0001] 본 발명은 단말기에서 영상 인식을 사용하여 음성 인식을 수행하는 방법 및 장치에 관한 것이다.

[0002] 보다 구체적으로, 본 발명은 사용자의 별도의 제스처 없이 카메라를 통한 영상 인식을 통하여, 발화 시작 및 발화 종료 시점을 파악하고 음성 인식의 정확도를 높일 수 있는 방법 및 장치에 관한 것이다.

배 경 기 술

[0003] 물리적 입력을 대체하고 따라서 사용자가 움직일 필요 없이 편리하게 전자 기기들을 사용하기 위하여 음성 인식 기술이 점점 보편화되는 추세에 있다. 예를 들면 음성 인식 기술은 스마트 폰, 텔레비전, 자동차 내비게이션 등 다양한 전자 기기에서 구현될 수 있다.

[0004] 도 1은 종래의 음성 인식을 수행하는 단말기의 표시화면이다. 도 1에서 도시되는 종래의 음성 인식 기술은 사용자가 특정 프로그램을 동작시켜서 녹음 시작, 말하기, 녹음 끝, 결과 연산 등의 과정을 거칠 것을 요구한다. 그리고 도 1에서 도시되는 종래의 기술은 기기의 현재 상태에 따른 명령어를 처리하는 기술 위주가 아니라 일반적으로 이미 정의되어 있는 키워드나 일반 자유 음성 인식을 위한 구조로 구현되어 있다.

[0005] 음성 인식 기술은 입력되는 음성을 통계 분석하여 구분한다. 이때 정확한 음성 인식을 위하여 녹음된 데이터의 잡음 또는 무음 구간을 최소화하는 것이 중요하다. 하지만 음성 녹음을 수행하는 사용자의 다양한 상황을 고려할 때 음성 녹음 데이터는 발화자의 목소리 이외의 노이즈가 포함될 가능성이 높고, 음성 발화 상태를 정확하게 파악하기 어려운 문제가 존재한다.

[0006] 그리고 종래에는 음성 녹음 시작을 위해 사용자가 별도의 동작이 필요한 문제가 있다. 예를 들어 사용자가 운전

중이거나 두 손 가득 짐을 들고 있는 경우를 고려할 수 있다. 음성 인식은 이러한 상황의 사용자에게 매우 유용한 기능이다. 별도의 키 입력이나 제스처의 입력 없이도 사용자의 음성 만으로 단말기의 기능을 실행할 수 있기 때문이다. 이러한 취지를 고려할 때, 음성 녹음 시작부터 사용자의 별도의 제스처 입력이 필요하지 않는 음성 인식 기술이 요구된다.

발명의 내용

해결하려는 과제

[0007] 본 발명은 상기와 같은 문제점을 해결하기 위한 것이다. 특히 본 발명은 사용자가 손을 움직일 필요 없이 단말기의 음성인식 기능을 실행시킬 수 있는 방법 및 장치를 제공하는 것을 목적으로 한다.

[0008] 나아가 본 발명은 영상 인식을 통하여 사용자의 발화 시작 및 종료 시점을 파악하고, 이를 통하여 정확하고 빠른 음성 인식을 할 수 있는 방법 및 장치를 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0009] 상기와 같은 문제점을 해결하기 위한 음성 인식 방법은 음성 인식 모드 전환 전, 제 1 비디오 데이터 또는 제 1 오디오 데이터 중 적어도 하나를 기반으로 발화(發話) 시작을 판단하는 단계; 발화 시작으로 판단한 경우, 음성 인식 모드로 전환하고 사용자의 음성 명령을 포함하는 제 2 오디오 데이터를 생성하는 단계; 및 음성 인식 모드로 전환하고 난 후의 제 2 비디오 데이터 또는 상기 제 2 오디오 데이터 중 적어도 하나를 기반으로 발화 종료를 판단하는 단계를 포함하는 것을 특징으로 한다.

[0010] 그리고 본 발명의 음성 인식 장치는 음성 입력을 수집 및 녹음하는 오디오부; 영상 입력을 수집 및 녹화하는 카메라부; 및 음성 인식 모드 전환 전 제 1 비디오 데이터 또는 제 1 오디오 데이터를 생성하도록 제어하고, 상기 1 비디오 데이터 또는 제 1 오디오 데이터 중 적어도 하나를 기반으로 발화(發話) 시작을 판단하고, 발화 시작으로 판단한 경우, 음성 인식 모드로 전환하여 사용자의 음성 명령을 포함하는 제 2 오디오 데이터 또는 제 2 비디오 데이터를 생성하도록 제어하며, 상기 제 2 비디오 데이터 또는 상기 제 2 오디오 데이터 중 적어도 하나를 기반으로 발화 종료를 판단하는 제어부를 포함하는 것을 특징으로 한다.

[0011] 나아가 본 발명의 음성 인식 플랫폼은 음성 입력 또는 영상 입력을 수집 및 녹화하여, 음성 인식 모드 전환 전의 제 1 비디오 데이터 또는 제 1 오디오 데이터를 생성하고, 음성 인식 모드 전환 후의 제 2 비디오 데이터 또는 제 2 비디오 데이터를 생성하는 멀티미디어 프레임워크; 및 상기 1 비디오 데이터 또는 제 1 오디오 데이터 중 적어도 하나를 기반으로 발화(發話) 시작을 판단하고, 상기 제 2 비디오 데이터 또는 상기 제 2 오디오 데이터 중 적어도 하나를 기반으로 발화 종료를 판단하는 보이스 프레임워크를 포함하는 것을 특징으로 한다.

발명의 효과

[0012] 본 발명에 따르면, 사용자는 손을 사용하지 않고도 음성 인식 기능을 구동시킬 수 있으며, 연속적인 음성 인식이 가능한 효과가 있다.

[0013] 나아가 본 발명에 따르면 사용자의 발화 시점을 정확하게 파악하여 녹음 데이터 분석 양을 최소할 할 수 있어 음성 인식의 정확도와 속도를 개선할 수 있는 효과가 있다.

도면의 간단한 설명

[0014] 도 1은 종래의 음성 인식을 수행하는 단말기의 표시 화면을 도시하는 도면.

도 2는 본 발명의 실시예에 따른 단말기의 구성을 도시하는 블록도

도 3은 본 발명의 실시예에 따르는 음성 인식 과정을 도시하는 순서도.

도 4는 본 발명의 실시예에 따라 TV를 통해 음성 인식을 사용하는 예시를 도시하는 도면

도 5는 본 발명의 실시예에 따라 음성 인식을 수행하는 장치의 구성을 대략적으로 도시하는 다른 도면.

도 6은 본 발명의 실시예에 따르는, 발화 영상을 이용한 음성 인식 사용 방법의 구체적인 과정을 도시하는 도면.

도 7은 음성 명령어를 통하여 단말을 제어하는 경우, 본 발명의 실시예에 따라 음성을 사용하는 방법의 구체적

인 과정을 도시하는 도면.

도 8은 음성 인식 과정에서 본 발명에 따라, 음성 명령어를 표시하는 구체적인 그래픽 인터페이스의 예시를 도시하는 도면

도 9는 어플리케이션의 음성 명령어를 위한 위젯과 위젯에 음성 명령어를 등록할 수 있는 소스 코드의 예시를 도시하는 도면

도 10은 사용자의 입모양 판단을 통한 음성 명령어를 표시하는 화면을 구성하는 순서를 설명하기 위한 도면

도 11은 도 11은 사용자의 발화 시점을 확인하는 플레폼의 동작을 설명하기 위한 도면

도 12는 사용자의 발화 종료 시점을 확인하는 플레폼의 동작을 설명하기 위한 도면

발명을 실시하기 위한 구체적인 내용

[0015] 본 발명은 이하에 기재되는 실시예들의 설명 내용에 한정되는 것은 아니며, 본 발명의 기술적 요지를 벗어나지 않는 범위 내에서 다양한 변형이 가해질 수 있음은 자명하다. 그리고 실시예를 설명함에 있어서 본 발명이 속하는 기술 분야에 널리 알려져 있고 본 발명의 기술적 요지와 직접적으로 관련이 없는 기술 내용에 대해서는 설명을 생략한다.

[0016] 본 발명의 단말기(200)는 음성 인식 기능을 지원하는 모든 전자 기기를 포함하는 개념이다.

[0017] 예를 들면, 본 발명의 실시예에 따른 단말기(200)는 휴대 전화기, PMP(Portable Multimedia Player), 디지털 방송 플레이어, 자동차 내비게이션 , PDA(Personal Digital Assistant), 음악 파일 재생기(예컨대, MP3 플레이어), 휴대 게임 단말기, 태블릿 PC 및 스마트 폰(Smart Phone) 등의 이동 가능한 소형 기기를 포함한다. 나아가, 본 발명의 실시예에 따른 단말기(200)는 텔레비전, 냉장고, 세탁기 등과 같은 고정된 장소에 설치되어 사용되는 가전제품에 이르기까지, 음성 인식 기능을 지원하는 모든 기기를 포함한다.

[0018] 한편, 첨부된 도면에서 동일한 구성요소는 동일한 부호로 표현된다. 그리고 첨부 도면에 있어서 일부 구성요소는 과장되거나 생략되거나 개략적으로 도시될 수도 있다. 이는 본 발명의 요지와 관련이 없는 불필요한 설명을 생략함으로써 본 발명의 요지를 명확히 설명하기 위함이다. 이하 첨부된 도면을 참조하여 본 발명의 바람직한 실시 예들을 상세히 설명한다.

[0019] 도 2은 본 발명의 실시예에 따른 단말기(200)의 구성을 개략적으로 나타낸 블록도이다.

[0020] 도 2을 참조하면, 본 발명의 단말기(200)는 무선통신부(210), 키 입력부(220), 카메라부(230), 오디오부(240), 터치스크린(250), 저장부(260), 및 제어부(270)를 포함할 수 있다.

[0021] 무선통신부(210)는 음성 통화를 위한 통신 채널 형성 및 화상 통화를 위한 통신 채널 형성, 영상이나 메시지 등의 데이터 전송을 위한 통신 채널(이하, 데이터 통신 채널) 형성 등을 제어부(270)의 제어 하에 수행한다.

[0022] 키 입력부(220)는 숫자 또는 문자 정보를 입력받고 각종 기능들을 설정하기 위한 다수의 입력키 및 기능키들을 포함한다. 기능키들은 특정 기능을 수행하도록 설정된 방향키, 사이드 키 및 단축키 등을 포함할 수 있다. 또한 키 입력부(220)는 사용자 설정 및 단말기(200)의 기능 제어와 관련된 키 신호를 생성하고 제어부(270)로 전달한다. 키 입력부(220)는 단말기(200)의 터치스크린(250)이 풀 터치스크린 형태로 지원되는 경우 단말기(100)의 케이스 측면에 형성되는 사이드 키, 홈 키 및 기타 기능 키들을 포함할 수 있다.

[0023] 특히 본 발명의 키 입력부(220)는 음성 인식 기능을 수행하도록 설정된 음성 인식 기능키를 포함할 수 있다. 그리고 키 입력부(220)는 음성 인식 기능키에서 생성되는 음성 인식 기능 키 이벤트를 제어부(270)로 전달할 수 있다. 제어부(270)는 음성 인식 기능키의 상기 요청 신호에 따라 음성 인식 모드의 시종을 결정할 수 있다.

[0024] 카메라부(230)는 피사체에 대한 촬영을 통하여 수집 영상을 제공한다. 이때 상기 카메라부(230)는 터치스크린(250) 또는 키 입력부(220)에서 발생하는 신호에 따라 활성화되어 영상을 수집할 수 있다. 카메라부(230)는 광학적 신호를 전기적 신호로 변환하는 카메라 센서, 아날로그 비디오 신호를 디지털 비디오 신호로 변환하는 영상처리장치(Image Signal Processor) 및 상기 영상처리장치에서 출력되는 비디오 신호를 터치스크린(250)에 표시하기 위해 상기 비디오 신호를 영상 처리(크기조정(Scaling), 잡음 제거, RGB 신호 변환 등)하는 신호처리장치(Digital Signal Processor) 등을 포함할 수 있다. 여기서 상기 카메라 센서는 CCD(Charge-coupled Device)

센서 또는 CMOS(Complementary Metal-Oxide Semiconductor) 센서 등이 사용될 수 있고, 상기 신호처리장치는 구성이 생략되고 DSP로 구현될 수 있다.

[0025] 특히 본 발명의 카메라부(230)는 제어부(270)의 제어 하에 단말기(200)의 대기 모드에서도 카메라를 구동하여 사용자 영상을 녹화한다. 그리고 카메라부(230)는 녹화한 비디오를 제어부(270)로 전달하여 사용자의 얼굴을 인식할 수 있는 영상 인식 데이터를 제공한다. 또한 본 발명의 비디오부(230)는 사용자의 설정에 따라 단말기(200)의 대기 모드에서도 활성화될 수 있으나, 사용자의 별도의 입력을 받아 활성화될 수 도 있다.

[0026] 오디오부(240)는 통화 시 송수신 되는 오디오 데이터, 수신된 메시지에 포함된 오디오 데이터, 저장부(260)에 저장된 오디오 파일 재생에 따른 오디오 데이터 등을 재생하기 위한 스피커(SPK)와, 통화 시 사용자의 음성 또는 기타 오디오 신호를 수집하기 위한 마이크(MIC)를 포함한다.

[0027] 특히 본 발명의 오디오부(240)는 제어부(270)의 제어 하에 음성 인식 모드에서 마이크(MIC)를 구동하여 마이크(MIC)를 통해 수집되는 사용자 음성을 녹음한다. 그리고 오디오부(240)는 녹음된 음성을 제어부(270)로 전달하여 녹음된 음성에 대한 음성 인식이 수행될 수 있도록 지원한다. 또한 본 발명의 오디오부(240)는 음성 인식 모드가 개시되거나 종료되면 그에 대응하는 효과음을 출력할 수 있다. 이러한 효과음들은 사용자 설정에 따라 생략될 수 있다.

[0028] 터치스크린(250)은 터치 패널(253)과 표시부(256)를 포함한다. 이러한 터치스크린(250)은 표시부(256) 전면에 터치 패널(253)이 배치되는 구조를 가질 수 있다. 터치스크린(250)의 크기는 터치 패널(253)의 크기로 결정될 수 있다. 그리고 터치스크린(250)은 사용자 기능 실행에 따른 화면을 표시하고 사용자 기능 제어 관련 터치 이벤트를 감지할 수 있다.

[0029] 터치 패널(253)은 표시부(256)의 상하부 중 적어도 한 곳에 배치되고, 터치 패널(253)을 구성하는 센서는 매트릭스 형태로 배열된다. 이에 따라 터치 패널(253)은 터치 패널(253) 상의 터치 물체의 접촉 또는 접근 거리에 따라 터치 이벤트를 생성하고, 생성된 터치 이벤트를 제어부(270)에 전달할 수 있다. 여기서 터치 이벤트는 터치 종류와 위치 정보를 포함한다.

[0030] 표시부(256)는 단말기(100)의 각종 메뉴를 비롯하여 사용자가 입력한 정보 또는 사용자에게 제공되는 정보를 표시한다. 즉 표시부(256)는 단말기(200) 이용에 따른 다양한 사용자 기능의 실행 화면을 제공할 수 있다. 이러한 표시부(256)는 액정 표시 장치(Liquid Crystal Display), OLED(Organic Light Emitted Diode) 등으로 형성될 수 있다. 그리고 표시부(256)는 터치 패널(253) 상부 또는 하부에 배치될 수 있다.

[0031] 특히 본 발명의 표시부(256)는 제어부(270)가 현재 사용자의 상태를 발화 시작으로 판단한 경우, 제어부(270)의 제어 아래 음성 명령어를 팝업으로 표시할 수 있다. 그리고 본 발명의 실시예를 따르는 표시부(256)는 제어부(270)가 음성 인식을 성공한 경우, 인식한 음성 명령어를 팝업으로 표시할 수 있다. 표시부(256)를 통하여 음성 명령어가 표시되는 구체적인 설명은 첨부된 도면과 함께 후술된다.

[0032] 저장부(260)는 본 발명의 실시예에 따른 기능 동작에 필요한 적어도 하나의 어플리케이션, 사용자에 의해 생성되는 사용자 데이터, 네트워크와 송수신하는 메시지 및 어플리케이션 실행에 따른 데이터 등을 저장한다. 이러한 저장부(160)는 크게 프로그램 영역과 데이터 영역을 포함할 수 있다.

[0033] 프로그램 영역은 단말기(200)의 부팅 및 상술한 각 구성의 운용을 위한 운영체제(OS, Operating System)와 다운로드 및 설치된 어플리케이션들 등을 저장할 수 있다. 특히, 본 발명의 프로그램 영역은 영상 인식 운용 프로그램과 음성 인식 운용 프로그램을 더 저장할 수 있다.

[0034] 영상 인식 운용 프로그램은 본 발명의 실시예에 따라 수집한 비디오 데이터를 분석하여 사용자의 발화 시작 시점과 발화 종료 시점을 판단할 수 있다. 이러한 영상 운용 프로그램은 단말기의 대기 상태에서도 구동될 수 있으며, 사용자의 별도의 입력을 수신하여 구동될 수 있다.

[0035] 음성 인식 운용 프로그램은 본 발명의 실시예에 따라 음성 인식을 이용하여 단말기의 기능이 수행되도록 지원한다. 특히 본 발명의 실시예에 따르는 음성 인식 운용 프로그램은 상기 영상 인식 운용 프로그램과 함께, 사용자의 발화 시작 시점과 발화 종료 시점을 판단하기 위하여 사용된다. 나아가 음성 인식 운용 프로그램은 음성 인식 결과에 기초하여 미리 설정된 음성 명령어들 중 하나가 음성 입력된 것으로 판단되면 해당 기능을 실행하는 루틴을 포함한다.

[0036] 데이터 영역은 단말기(200)의 사용에 따라 발생하는 데이터가 저장되는 영역이다. 특히 본 발명의 데이터 영상 인식 운용 프로그램 및 음성 인식 운용 프로그램이 실행되는 중에 사용되거나 생성되는 데이터를 저장한다 그리

고 데이터 영역은 음성 인식 운용 프로그램(163)과 연계하여 음성 인식을 위한 각종 통계 모형 및 음성 인식 결과 등을 저장할 수 있다.

[0037] 제어부(270)는 단말기(200)의 전반적인 동작을 제어한다. 특히 본 발명의 제어부(270)는 영상 인식을 통하여 발화 시작 및 종료 시점을 판단하고, 사용자가 입력한 음성 명령어를 판단하며, 상기 음성 명령어와 연결된 단말기의 기능을 수행하는 일련의 과정을 제어한다. 제어부(270)의 세부 모듈 및 세부 모듈의 기능에 대한 상세한 설명은 첨부된 도면과 함께 후술된다.

[0038] 도 3는 본 발명의 실시예에 따르는 음성 인식 과정을 도시하는 순서도이다.

[0039] 본 발명의 실시예에 따르는 음성 인식 절차는 카메라부(230)만 활성화된 단말기의 대기 모드에서 출발할 수 있다. 그리고 단말기의 활성 모드에서 출발할 수도 있는데, 이 때 표시부(256)는 사용자 기능 실행 화면을 표시한다. 사용자 기능 실행 화면은 적어도 하나의 이미지로 구성된 위젯들, 예컨대 아이콘, 썸네일(thumbnail) 또는 문자 등으로 구성될 수 있다. 이러한 이미지 구성요소들 각각은 특정 기능과 연계되어있다. 따라서 음성 명령어가 인식되면, 제어부(270)는 상기 특정 이미지 구성요소와 연계된 기능이 실행되도록 제어할 수 있다.

[0040] 단계 310 내지 단계 320는 제어부(270)가 발화 시작 여부를 확인하여 음성 인식 모드로 전환할지 여부를 판단하는 단계이다.

[0041] 단계 310에서 카메라부(230)는 비디오 데이터를 수집하고, 오디오부(240)는 녹음을 통하여 오디오 데이터를 수집할 수 있다. 제어부(270)는 이러한 비디오 데이터와 오디오 데이터를 함께 분석하여 발화 시작 여부를 판단할 수 있다. 나아가 본 발명의 다른 실시예에 따르면 오디오부(240)의 구동 없이 카메라부(230)만 구동하여 비디오 데이터를 수집할 수 있으며, 제어부(270)는 비디오 데이터에 녹화된 사용자의 입 모양이 열린 것으로 판단한 경우, 이를 음성 인식 기능을 시작하기 위한 발화 시작으로 판단할 수 있다.

[0042] 보다 구체적으로 본 발명의 실시예에 따라 발화 시작 시점을 확인하는 방법은, 첫째로 녹화된 비디오 데이터만으로 판단하는 방법이 있다. 즉, 녹화된 영상을 분석하여 사용자의 입모양을 찾아내고, 사용자의 입모양을 분석하여 사용자가 발화하는 시점을 찾아내는 것이다.

[0043] 둘째로 녹화된 영상과 녹음된 음성을 동시에 분석하여 발화 시점을 찾는 방법이 있다. 이러한 방법을 따르면, 실제 사용자가 발화하지 않고 입을 움직이거나 소리가 작을 경우 등을 감안할 수 있으며, 입 모양과 녹음된 음성의 크기 및 사람의 목소리인지를 파악하여 발화 시점을 보다 정확하게 파악할 수 있는 장점이 있다.

[0044] 두번째 방법을 따르는 경우, 녹화된 영상과 녹음된 음성을 분석한 결과가 모두 발화시작으로 판단되는 경우, 제어부(270)는 발화 시작으로 판단한다. 그러나 영상을 분석한 결과만 발화시작으로 판단되는 경우, 제어부(270)는 음성 녹음 데이터 판단 기준(Threshold)를 완화하여 다시 판단하고, 유효한 경우 발화 시작으로 판단하고 유효하지 않으면 발화 시작으로 판단하지 않는다.

[0045] 한편 음성을 분석한 결과만 발화시작으로 판단되는 경우, 예를 들어 영상을 통한 인식에서 사용자의 얼굴 자체가 나타나지 않을 때 제어부(270)는 발화 시작으로 판단하지 않는다. 그러나 얼굴 자체는 인식된다면 제어부(270)는 입모양의 움직임의 판단 기준을 완화하여 다시 확인하고 유효하면 발화 시작으로 판단하고 유효하지 않으면 발화 시작으로 판단하지 않는다.

[0046] 단계 320에서 제어부(270)가 사용자가 음성 인식을 위한 발화를 시작한 것으로 판단한 경우, 제어부(270)는 표시부(256)를 제어하여 현재 표시된 위젯들과 연결된 음성 명령어를 표시할 수 있다. 그리고 제어부(270)는 단계 310에서 오디오부(240)가 구동되어 있지 않던 경우, 단계 320에서 오디오부(240)를 구동시켜 음성 인식 모드로 전환할 수 있다.

[0047] 본 발명의 일 실시예에 따르면, 단말기(200)가 음성 인식 모드로 전환하면 제어부(170)는 현재의 실행 화면과 관련된 음성 명령어들을 표시부(256)에 표시할 수 있다. 일반적인 dictation 기능이 아닌 음성 명령을 통해 기기를 제어하기 위한 음성 명령인 경우 사용자는 화면에 실행할 명령어를 볼 수 있어야 하기 때문이다. 예를 들어 화면의 아이콘을 터치하는 기능을 음성을 통해 수행하고자 하는 경우, 사용자에게 실행 가능한 명령어를 표시하고 사용자는 실행하고자 하는 명령어를 발화할 수 있게 된다.

[0048] 예를 들면 사용자가 입을 벌리는 경우, 제어부(270)는 단계 310에서 발화 시작으로 판단하고, 단계 320에서 음성 명령어를 팝업 창에 표시하며 레코더를 통해 녹음을 시작할 수 있다. 그러나 음성 인식 모드로 전환하기 전

에 사용자가 입을 닫으면 명령어를 표시하지 않고 레코더 녹음을 중단할 수 있다.

[0049] 위의 예에서 제어부(270)는 레코더를 통해 녹음되는 데이터를 판단하여 음성이 유효한지 판단하며, 음성이 유효한 경우 녹음되는 데이터를 음성 인식을 시작하게 된다. 이후 발화 종료 시점을 확인하는 단계 340 내지 단계 350을 통해 발화 조율가 확인되면 녹음을 중단하고 실제 녹음되는 데이터의 무음 구간을 분석하여 종합적으로 음성 인식 결과를 판단하는 과정을 따르게 된다. 이러한 구체적인 설명은 각 단계에 대한 설명에서 보다 상세히 후술된다.

[0050] 단계 320에서 제어부(270)는 홈 화면에서 특정 아이콘, 예컨대 캘린더 아이콘이 표시된 구역에서 일정 거리 이내에 'Calendar'라는 음성 명령어가 표시되도록 표시부(256)를 제어할 수 있다. 또한 제어부(270)는 'Help'라는 음성 명령어가 실행 화면의 여백 어느 곳에든지 출력되도록 지원할 수 있다. 이러한 기능은 음성 인식을 위한 데이터 처리량을 최소화하기 위한 것으로, 음성 인식 기능을 특정 음성 명령어들로 한정하여 정확하고 빠른 음성 인식 서비스를 제공하기 위한 것이다. 음성 명령어를 표시하는 과정을 보다 구체적으로 설명한다.

[0051] 특히 제어부(270)는 음성 명령어의 표시 위치를 결정할 때 음성 명령어와 연계된 기능과 연관된 이미지 구성요소가 존재하면 상기 이미지 구성요소 부근으로 음성 명령어 표시 위치를 선택할 수 있다. 이때 제어부(270)는 실행 화면의 맵을 참조하여 상기 이미지 구성요소의 존재 여부를 판단할 수 있다. 그리고 제어부(270)는 실행 화면의 맵을 참조하여 음성 명령어 표시 위치를 선택할 수 있다.

[0052] 예를 들어 제어부(270)는 동영상 실행 화면에서, 동영상 플레이 기능과 연결된 아이콘이 표시된 위치에 음성 명령어가 표시되도록 지원할 수 있다. 즉, 일시 정지 아이콘 근처에 일시정지 음성 명령어가 표시되도록 지원할 수 있다.

[0053] 이에 더하여 제어부(270)는 음성 명령어가 미리 설정된 표시 방식에 따라 출력되도록 지원할 수 있다. 예를 들면 음성 명령어가 툴 팁(tool-tip) 또는 말풍선과 같은 형태로 실행 화면에 오버레이(overlay)되어 표시되도록 지원할 수 있다. 그리고 다수 개의 기능들이 하나의 이미지 구성요소와 연계되어 복수의 음성 명령어들이 존재하는 경우, 제어부(270)는 음성 명령어들이 리스트 형태로 표시되도록 지원할 수 있다. 이에 더하여 제어부(270)는 음성 명령어들을 종합하여 하나의 팝업 창(pop-up window)이나 별도의 화면에 표시하도록 표시부(256)를 제어할 수 있다. 이러한 음성 명령어를 표시하는 구체적인 사용자 인터페이스는 첨부된 도 8과 함께 후술된다.

[0054] 단계 330은 제어부(270)가 오디오부(240)를 통하여 음성을 인식하는 단계이다.

[0055] 이 때, 음성 인식 중에 사용자의 얼굴이 영상 녹화 영역을 벗어나는 경우가 존재한다. 사용자가 녹음 중에 영상 녹화 영역을 벗어나는 경우, 제어부(270)는 음성 분석만으로 발화 종료 구간을 확인할 수 있다. 그리고 제어부(270)는 표시부(256) 또는 오디오부(240)의 스피커를 통하여 사용자에게 이러한 사실을 알릴 수 있다.

[0056] 한편, 사용자가 발화 시작 전에 녹화 영역을 벗어나는 경우가 있다. 이 경우에는 사용자가 기기를 사용하지 않는다고 가정할 수 있기 때문에 제어부(270)는 사용자의 얼굴이 녹화 영역에 들어올 때까지 음성인식을 시작하지 않는다. 그리고 제어부(270)는 표시부(256) 또는 오디오부(240)의 스피커를 통하여 사용자에게 이러한 사실을 알릴 수 있다.

[0057] 일반적인 경우, 제어부(270)는 녹음된 음성 데이터에서 음소들을 구분하고 음소들이 구성하는 단어(또는 단어열)를 파악한다. 특히 본 발명의 실시예에 따른 제어부(270)는 단계 320에 따라 실행화면에 미리 표시된 음성 명령어들을 기준으로 음성 인식을 수행할 수 있다.

[0058] 예를 들어 제어부(270)는 음소 또는 단어 단위로 음성을 인식하고, 표시부(256)에 표시된 음성 명령어들의 음소 또는 단어와 비교하여 음성 인식 시간을 단축할 수 있다. 이는 외부 서버에게 음성 인식을 위탁하지 않고, 단말기(100) 자체적으로 음성 인식을 수행하는 경우이다.

[0059] 그러나 본 발명은 이에 제한되지 않으며, 지능형 음성 인식을 지원하는 종래 기술에 따라 구현될 수 있다. 예를 들면 사용자로부터 오늘 일정을 확인해 달라는 음성 명령을 수신한 경우, 제어부(270)는 단계 330에서 사용자의 자연어를 분석하여 해당 명령을 판단할 수 있다.

[0060] 한편, 음성 명령을 포함하는 오디오 데이터가 잡음, 오류, 인식 범위를 벗어난 것으로 판단되면, 제어부(270)는 사용자에게 음성 인식 중단을 고지하고, 단계 310로 되돌아갈 수 있다. 이는 실행화면에 음성 명령을 미리 표시한 경우도 포함된다.

[0061] 단계 340 내지 단계 350은 제어부(270)가 발화 종료 여부를 확인하여 음성 인식 모드를 종료할지 여부를 판단하는 단계이다.

[0062] 단계 340에서 카메라부(230)는 비디오 데이터를 수집하고, 오디오부(240)는 녹음을 통하여 오디오 데이터를 수집할 수 있다. 제어부(270)는 이러한 비디오 데이터와 오디오 데이터를 함께 분석하여 발화 종료 여부를 판단할 수 있다. 나아가 본 발명의 다른 실시예에 따르면 카메라부(230)의 구동 없이 오디오부(240)만 구동하여 오디오 데이터를 수집할 수 있으며, 제어부(270)는 오디오 데이터에서 사용자의 음성이 미리 설정된 시간 이상 입력되지 않은 경우 음성 인식 기능을 종료하기 위한 발화 종료로 판단할 수 있다.

[0063] 보다 구체적으로 본 발명의 실시예에 따라 발화 종료 시점을 확인하는 방법은, 녹화된 영상과 녹음된 음성을 모두 활용하여 판단하는 방법을 따를 수 있다. 이 때, 영상과 녹음된 음성을 분석한 결과가 모두 발화 종료로 판단되는 경우, 제어부(270)는 발화 종료로 판단할 것이다. 그러나 영상을 분석한 결과만 발화 종료로 판단되는 경우, 제어부(270)는 음성 녹음 데이터 판단 기준(Threshold)를 완화하여 다시 판단하고, 유효한 경우 발화 종료로 판단하고 유효하지 않으면 발화 종료로 판단하지 않는다.

[0064] 한편 음성을 분석한 결과만 발화 종료로 판단되는 경우, 예를 들어 영상을 통한 인식에서 사용자의 얼굴 자체가 나타나지 않을 때 제어부(270)는 발화 종료로 판단한다. 그러나 얼굴 자체는 인식된다면 제어부(270)는 입모양의 움직임의 판단 기준을 완화하여 다시 확인하고 유효하면 발화 종료로 판단하고 유효하지 않으면 발화 종료로 판단하지 않는다.

[0065] 단계 340에서 제어부(270)가 사용자가 음성 인식을 위한 발화를 종료한 것으로 판단한 경우, 제어부(270)는 표시부(256)를 제어하여 단계 350에서 음성 인식의 결과를 표시할 수 있다.

[0066] 예를 들면 사용자로부터 근처 맛집을 검색해 달라는 음성 명령을 단계 330에서 수신한 경우, 제어부(270)는 단계 350에서 <근처 맛집 검색 중>이라는 음성 인식 결과를 표시부(256)를 통하여 표시할 수 있다.

[0067] 이후 단계 360에서 제어부(270)는 인식된 음성 명령에 대한 기능을 실행하며, 음성 인식 기능을 완료한다. 그리고 제어부(270)는 단계 310으로 되돌아 가서 다시 음성 인식의 과정을 수행할 수 있다.

[0068] 한편 본 발명의 실시예에 따르면, 속도 및 메모리 개선을 위하여 단계 320에서 발화 시작이 판단된 후에는 제어부(270)는 영상을 통한 분석은 수행하지 않고 발화 종료 여부는 음성만으로도 분석할 수 있다. 그리고 단계 360을 거쳐 단계 310으로 되돌아간 경우에는 다시 영상을 통한 분석을 시작할 수 있다.

[0069] 도 4는 본 발명의 실시예에 따라 TV를 통해 음성 인식을 사용하는 예시를 도시하는 도면이다.

[0070] 음성을 통해 기기를 제어하고자 하는 경우, 음성 인식의 범위를 설정하는 것은 쉽지 않다. 그러나 본 발명의 실시예를 활용하면 사용자는 한번에 한가지의 ACTION을 인식 시켜서 단말기를 동작할 수 있으며 쉽고 연속적으로 음성 명령을 적용할 수 있다.

[0071] 예를 들어 도 4에서 도시된 것과 같이, 사용자가 스마트 TV를 사용하는 경우, 일반적인 리모컨의 버튼 입력 대신에 음성으로 명령하여 스마트 TV의 특정 기능을 수행시킬 수 있다. 또한 백그라운드에서 동작하는 어플리케이션을 foreground로 가져오거나 세팅 메뉴의 옵션을 쉽게 조절할 수 있는 효과가 있다.

[0072] 도 5는 본 발명의 실시예에 따라 음성 인식을 수행하는 장치의 구성을 대략적으로 도시하는 도면이다. 도 5에서 도시된 바와 같이 본 발명의 실시예를 따르면, 카메라와 마이크를 포함하는 휴대 단말기에서 영상 인식을 통한 음성 인식 기능을 구현할 수 있다. 그리고 도 5에서 도시된 바와 같이 본 발명의 실시예를 따르는 음성 인식 장치는 카메라부(230)와 오디오부(240)가 포함하는 마이크를 구비한다. 그리고 제어부(270)는 카메라가 녹화한 영상 및 마이크가 녹음한 음성을 처리하여 이를 분석하는 구성 요소를 포함한다.

[0073] 도 6은 본 발명의 실시예에 따르는, 발화 영상을 이용한 음성 인식 사용 방법의 구체적인 과정을 도시하는 도면이다. 앞서 도 3에 대한 설명에서 본 발명의 구현 방법에 대해 상세히 설명하였으므로 도 6에 대한 설명은 본 발명의 대표 실시예에 대해 간단히 설명하도록 한다.

[0074] 먼저, 제어부(270)는 카메라(230)를 통한 영상을 분석하는데, 구체적으로는 사용자의 입 모양을 분석하여 발화

시점을 찾아낼 수 있다. 그리고 제어부(270)는 사용자가 발화한 것으로 판단하면 레코더(240)를 동작하여 녹음을 시작하며, 녹음된 데이터를 이용하여 음성 인식을 수행할 수 있다.

[0075] 이후, 제어부(270)는 카메라를 통한 영상과 녹음되는 음성 데이터를 분석하여 발화가 종료되는 시점을 판단할 수 있다. 그리고 발화가 종료된 것으로 판단되면 녹음을 중지하고 음성 인식을 수행한다.

[0076] 이 때, 제어부(270)가 음성 인식 결과 파악에 성공하면 음성 인식 결과에 해당하는 동작을 수행하게 된다. 한편, 음성 인식 결과 파악에 실패하거나 음성 인식 결과 파악에 성공하고 해당 동작 수행까지 완료한 경우, 그리고 다시 제어부(270)는 처음으로 되돌아가서 사용자의 발화시작을 대기하게 된다.

[0077] 도 7은 음성 명령어를 통하여 단말을 제어하는 경우, 본 발명의 실시예에 따라 음성을 사용하는 방법의 구체적인 과정을 도시하는 도면이다. 앞서 도 3에 대한 설명에서 본 발명의 구현 방법에 대해 상세히 설명하였으므로 도 7에 대한 설명은 본 발명의 대표 실시예에 대해 간단히 설명하도록 한다.

[0078] 먼저, 제어부(270)는 카메라(230)를 통한 영상을 분석하는데, 구체적으로는 사용자의 입 모양을 분석하여 발화 시점을 찾아낼 수 있다. 이 때, 제어부(270)는 사용자가 발화 시작 전에 입을 벌렸다고 판단하는 경우, 음성 명령어를 표시부(256)에 표시할 수 있다. 동시에 레코더(240)를 동작하여 녹음을 시작하며, 녹음된 데이터를 이용하여 음성 인식을 수행할 수 있다.

[0079] 이후, 제어부(270)는 카메라를 통한 영상과 녹음되는 음성 데이터를 분석하여 발화가 종료되는 시점을 판단할 수 있다. 그리고 발화가 종료된 것으로 판단되면 녹음을 중지하고 표시부(256)의 음성 명령어 표시를 종료한다. 그리고 음성 인식을 수행한다.

[0080] 이 때, 제어부(270)가 음성 인식 결과 파악에 성공하면 음성 인식 결과에 해당하는 동작을 수행하게 된다. 한편, 음성 인식 결과 파악에 실패하거나 음성 인식 결과 파악에 성공하고 해당 동작 수행까지 완료한 경우, 그리고 다시 제어부(270)는 처음으로 되돌아가서 사용자의 발화시작을 대기하게 된다.

[0081] 도 8은 음성 인식 과정에서 본 발명에 따라, 음성 명령어를 표시하는 구체적인 그래픽 인터페이스의 예시를 도시하는 도면이다.

[0082] 도 8에서 도시된 바와 같이 각종 어플리케이션에 대한 아이콘이 표시된 일반 화면이 표시된 상태에서 음성 인식이 수행될 수 있다. 본 발명의 실시예에 따라 제어부(270)가 사용자의 상태를 발화 시작으로 판단한 경우, 제어부(270)는 도 8에 도시된 바와 같이 표시부(256)에 어플리케이션을 실행하기 위한 음성 명령어를 아이콘과 함께 표시할 수 있다.

[0083] 도 9는 어플리케이션의 음성 명령어를 위한 위젯과 위젯에 음성 명령어를 등록할 수 있는 소스 코드의 예시를 도시하는 도면이다.

[0084] 도 9에서 도시된 바와 같이 표시부(256)에 표시된 복수의 위젯은 각각 어플리케이션과 연결되어 있다. 그리고 도 9에서 도시된 바와 같이, <CREATE> 라는 음성을 <CREATE NEW MESSAGE> 위젯과 연결된 어플리케이션을 실행하기 위한 음성 명령어로 등록할 수 있다.

[0085] 도 10은 사용자의 입모양 판단을 통한 음성 명령어를 표시하는 화면을 구성하는 순서를 설명하기 위한 도면이다.

[0086] 먼저, 번호 1번에서 카메라(230)를 통해 레코드 모듈은 녹화를 시작한다. 그리고 번호 2번 내지 3번에서 카메라부(230)는 녹화된 영상을 얼굴 인식 엔진에 전달하고, 얼굴 인식 엔진은 영상을 분석한다. 이후 번호 4번에서 얼굴 인식 엔진이 사용자의 입 모양이 열린 것을 판단하면, 사용자 인터페이스 프레임워크에 음성 명령어 구성을 요청하게 된다.

[0087] 이후 번호 5번에서 사용자 인터페이스 프레임워크는 어플리케이션의 위젯들의 명령어를 수집하여 구성할 수 있다. 그리고 번호 6번에서 수집된 명령어를 보이스 프레임워크에 전달하고, 명령어를 음성 인식의 후보군으로 쓸

수 있도록 한다.

[0088] 번호 7번에서 보이스 프레임워크는 음성 인식 엔진에 해당 내용을 전달하여 음성 인식을 시작할 수 있도록 준비하고, 번호 8번에서 사용자 인터페이스 프레임워크는 구성된 명령어가 문제가 없는 것으로 보이스 프레임워크가 확인하면, 사용자가 인식할 수 있도록 디스플레이를 통해 표시한다.

[0089] 도 11은 사용자의 발화 시점을 확인하는 플랫폼의 동작을 설명하기 위한 도면이다.

[0090] 먼저 번호 1번에서 카메라 또는 마이크를 통해 레코더 모듈은 녹화와 녹음을 시작한다. 다만, 녹화 영상만으로 발화시점을 판단하도록 미리 설정된 경우에는 녹음 기능은 구동하지 않는다.

[0091] 번호 2번에서 레코더 모듈은 녹화 및 녹음 데이터를 보이스 프레임워크로 전달하고, 번호 3번에서 녹화 데이터는 얼굴 인식 엔진으로 전달되고 녹음 데이터는 음성 인식 엔진으로 전달되어 발화 시작 여부를 판단하게 된다. 즉, 번호 4번에서 제어부(270)는 각 엔진의 판단을 바탕으로 발화시작 여부를 판단할 수 있다.

[0092] 발화 시작으로 판단된 경우, 번호 5번에서 제어부(270)는 음성 인식 엔진에게 음성 인식 시작을 알리며 녹음된 데이터를 지속적으로 전달한다. 동시에 번호 6번에서 사용자 인터페이스 프레임워크에게 현재 화면에 표시된 명령어를 제거할 것을 요청할 수 있다. 이후 번호 7번에서 사용자 인터페이스 프레임워크는 화면에 표시된 명령어를 제거하게 된다.

[0093] 도 12는 사용자의 발화 종료 시점을 확인하는 플레폼의 동작을 설명하기 위한 도면이다.

[0094] 먼저 번호 1번의 경우, 발화 종료 확인 시점에는 카메라, 마이크를 통해 레코더 모듈이 녹화와 녹음을 진행 중인 상태이다. 다만, 녹음된 음성 데이터 만으로 발화 종료 시점을 판단하는 경우, 카메라를 통화 녹화 기능을 구동하지 않을 수 있다.

[0095] 번호 2번에서 녹화된 데이터와 녹음된 데이터는 보이스 프레임 워크로 전달되며, 번호 3번에서 이는 각각 얼굴 인식 엔진과 음성 인식 엔진으로 전달되어 발화 종료 여부를 판단하는데 사용된다.

[0096] 이후 번호 4번에서 제어부(270)는 각 엔진의 판단을 바탕으로 발화 종료 여부를 판단할 수 있다. 번호 5번에서 발화 종료로 판단되면 음성 인식 엔진을 제어하여 녹음을 종료하고, 녹음된 데이터로부터 음성 인식 최종 결과를 판단하게 된다. 즉, 번호 6번에서 음성 인식 엔진을 통해 최종 인식 결과를 수령하게 된다.

[0097] 이후 번호 7번에서 제어부(270)는 최종 음성 인식 결과에 해당하는 ACTION을 파악하여 해당 모듈에 ACTION 수행을 요청하게 된다.

[0098] 한편, 본 발명은 복수의 사용자가 전제되는 경우에도 구현될 수 있다. 예를 들어, 카메라 모듈, 레코더 모듈, 프로젝터 또는 모니터 등의 디스플레이 모듈 및 본 발명의 실시예를 따르는 음성 인식 및 영상 인식 모듈을 구비하는 영상회의 시스템이 있을 수 있다.

[0099] 영상회의 시스템에서는 복수의 사용자가 원격으로 영상 회의를 참가하고, 복수의 사용자의 얼굴이 동시에 모니터에 표시된다. 영상 회의가 진행되면서, 복수의 사용자 중 적어도 하나 이상의 발화자가 존재하는 경우, 본 발명의 실시예에 따르면 복수의 영상 회의 참가자 들의 발화 시작 시점 및 종료 시점을 보다 명확하게 판단할 수 있다.

[0100] 예를 들어, 본 발명의 실시예에 따르면 A,B,C,D의 사용자가 영상 회의를 진행하면서 A의 입이 열리게 되면 A의 발화가 시작된 것으로 판단할 수 있다. 나아가, 본 발명의 실시예에 따르면 녹화된 영상과 녹음된 음성을 동시에 분석하여, A의 발화가 종료하고 B의 발화가 시작된 것으로 판단할 수 있다. 그리고 영상 회의라는 특수성을 감안하여, 본 발명의 다른 실시예에 따르면 C가 손을 드는 제스처를 취하면 녹화된 영상을 분석하여 C의 발화가 시작된 것으로 판단하고, 손을 내리는 제스처를 취하면 C의 발화가 종료된 것으로 판단할 수 있다.

[0101] 이러한 판단을 통하여 영상회의 시스템은 발화자의 얼굴이 모니터에 가장 크게 표시되도록 영상 회의 사용자 인터페이스를 설계할 수 있으며, 나아가 발화자만이 영상회의 시스템을 제어하기 위한 음성 인식 기능을 사용하도록 할 수 있다.

[0102] 본 명세서와 도면에 개시된 본 발명의 실시 예들은 본 발명의 기술 내용을 쉽게 설명하고 본 발명의 이해를 돕기 위해 특정 예를 제시한 것뿐이며, 본 발명의 범위를 한정하고자 하는 것은 아니다. 여기에 개시된 실시 예들 이외에도 본 발명의 기술적 사상에 바탕을 둔 다른 변형 예들이 실시 가능하다는 것은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 자명한 것이다.

부호의 설명

[0103] 210 : 통신부

220 : 키입력부

230 : 카메라부

240 : 오디오부

250 : 터치 스크린부

253 : 터치 패널

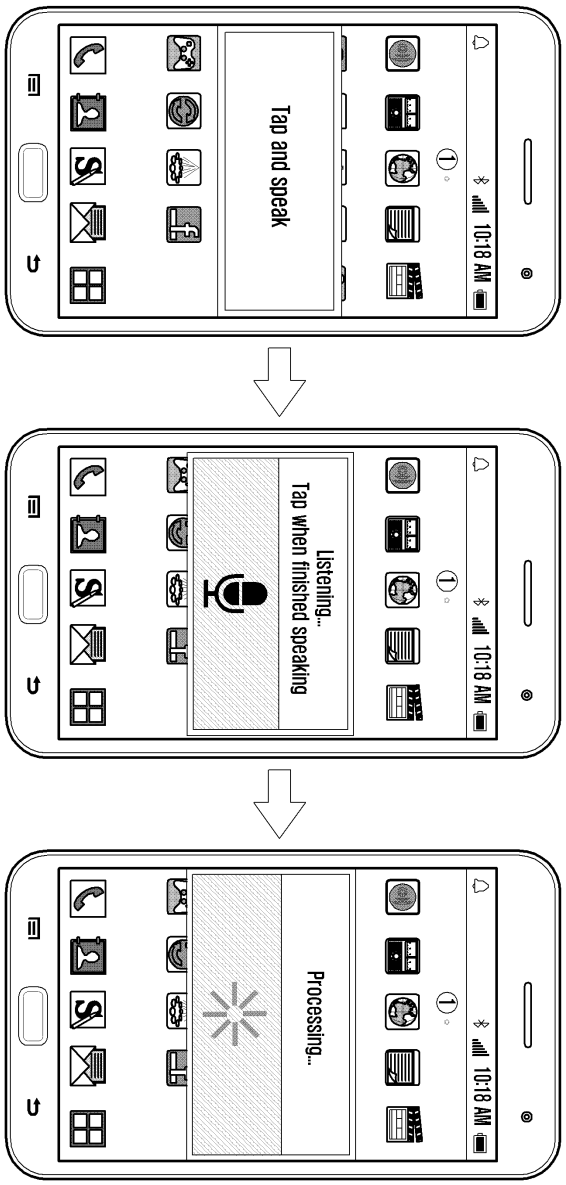
256 : 표시부

260 : 저장부

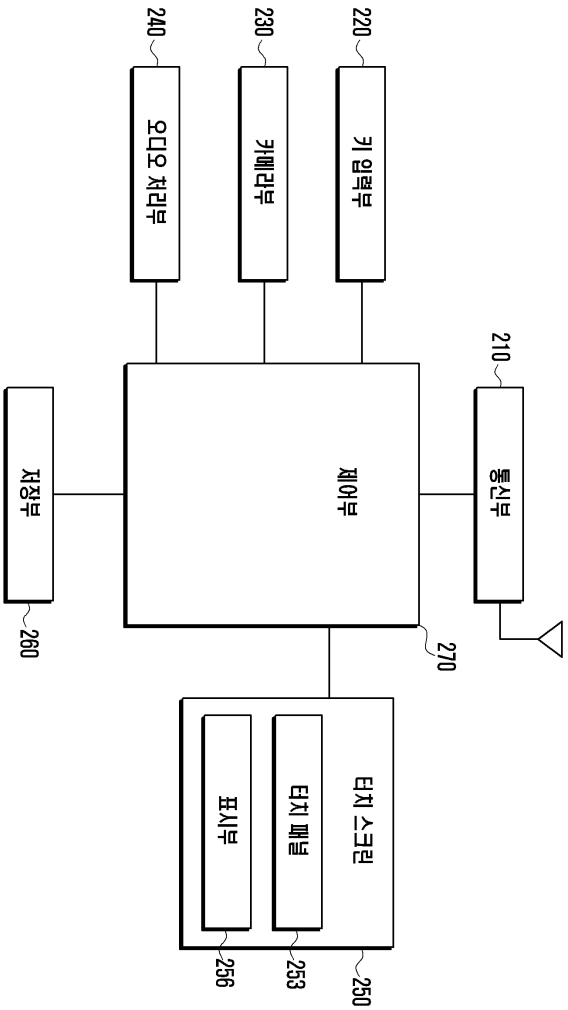
270 : 제어부

도면

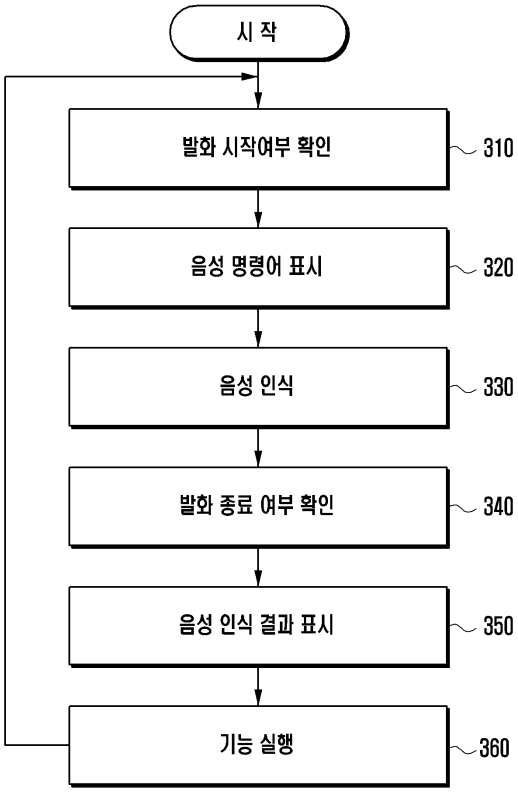
도면1


Tap and speak
10:18 AM
Listening...
Tap when finished speaking
10:18 AM
Processing...
10:18 AM

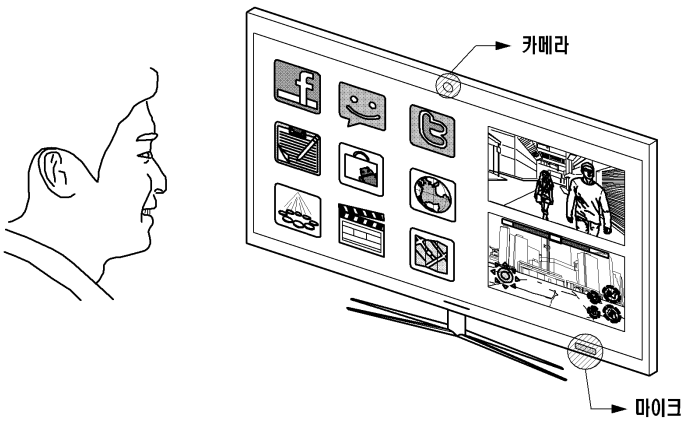
도면2


240
230
220
오디오 처리부
카메라부
키 입력부
210
저장부
제어부
통신부
260
270
터치 스크린
표시부
터치 패널
256
253
250

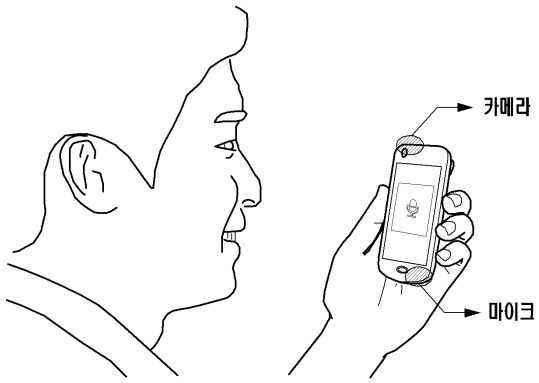
도면3

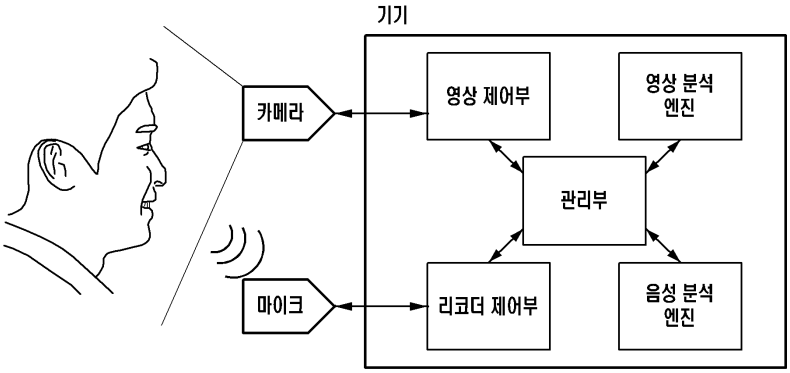

시 작
발화 시작여부 확인
310
음성 명령어 표시
320
음성 인식
330
발화 종료 여부 확인
340
음성 인식 결과 표시
350
기능 실행
360

도면4

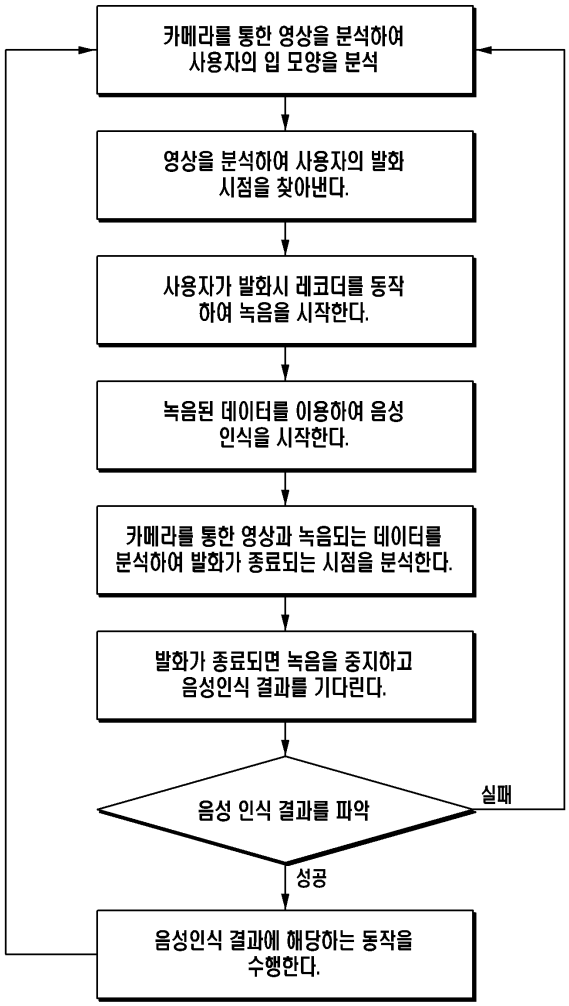

카메라
마이크

도면5

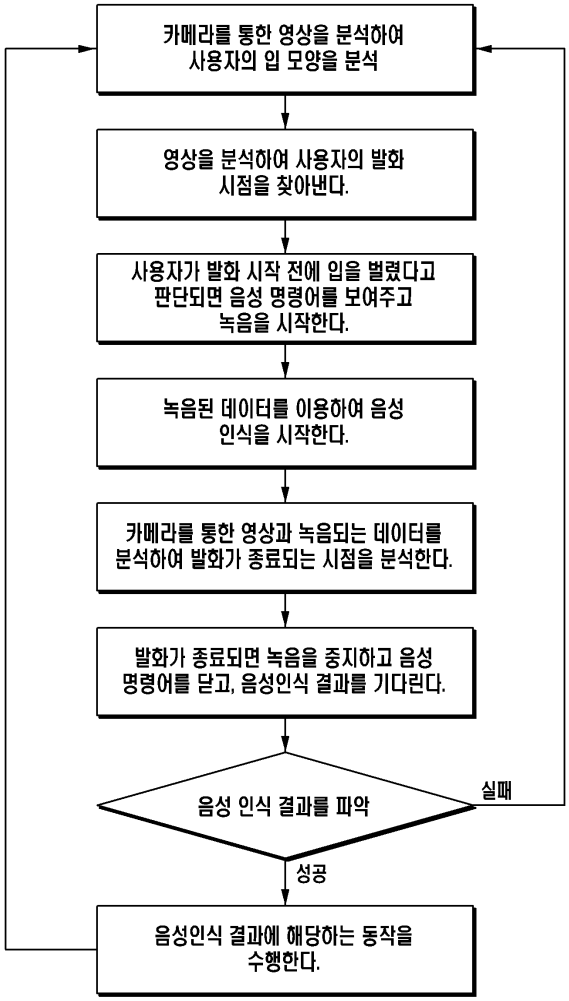




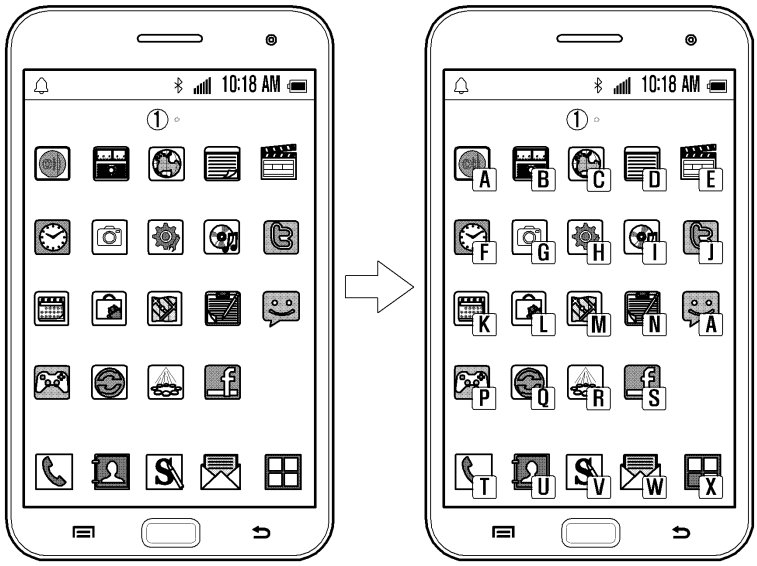
도면6


카메라를 통한 영상을 분석하여
사용자의 입 모양을 분석
영상을 분석하여 사용자의 발화
시점을 찾아낸다.
사용자가 발화시 레코더를 동작
하여 녹음을 시작한다.
녹음된 데이터를 이용하여 음성
인식을 시작한다.
카메라를 통한 영상과 녹음되는 데이터를
분석하여 발화가 종료되는 시점을 분석한다.
발화가 종료되면 녹음을 중지하고
음성인식 결과를 기다린다.
음성 인식 결과를 파악
실패
성공
음성인식 결과에 해당하는 동작을
수행한다.

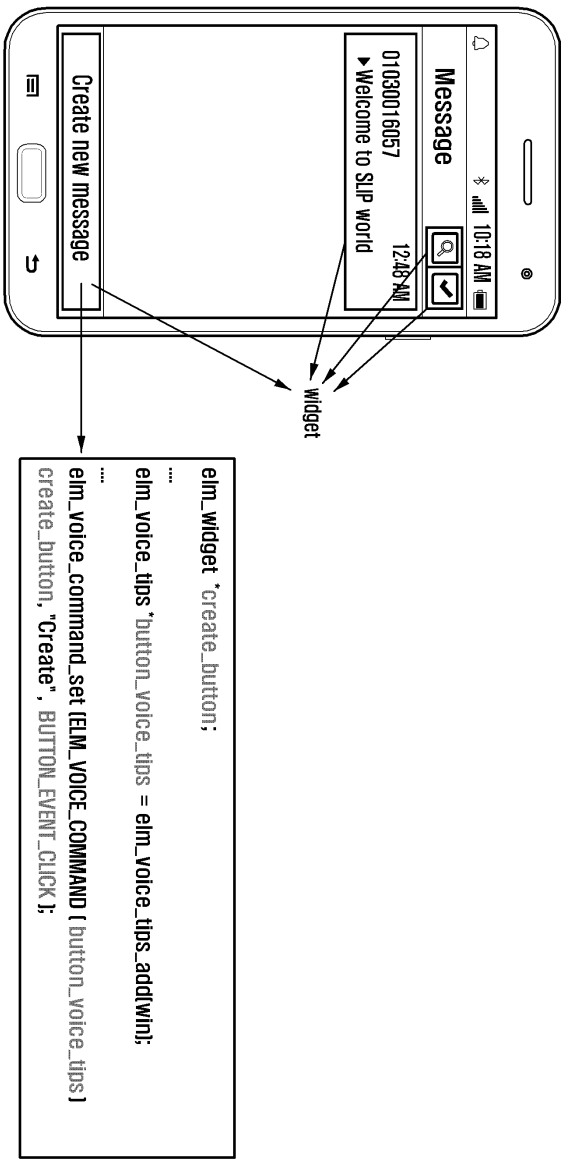
도면7


카메라를 통한 영상을 분석하여
사용자의 입 모양을 분석
영상을 분석하여 사용자의 발화
시점을 찾아낸다.
사용자가 발화 시작 전에 입을 벌렸다고
판단되면 음성 명령어를 보여주고
녹음을 시작한다.
녹음된 데이터를 이용하여 음성
인식을 시작한다.
카메라를 통한 영상과 녹음되는 데이터를
분석하여 발화가 종료되는 시점을 분석한다.
발화가 종료되면 녹음을 중지하고 음성
명령어를 닫고, 음성인식 결과를 기다린다.
음성 인식 결과를 파악
실패
성공
음성인식 결과에 해당하는 동작을
수행한다.

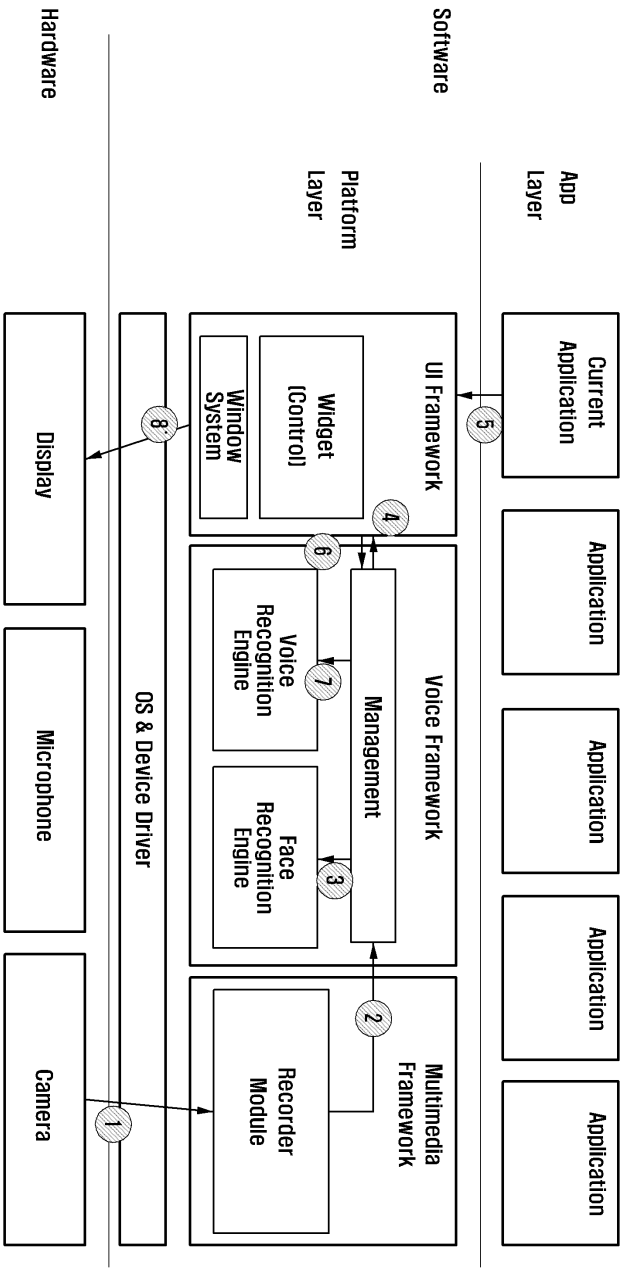
도면8


10:18 AM
10:18 AM

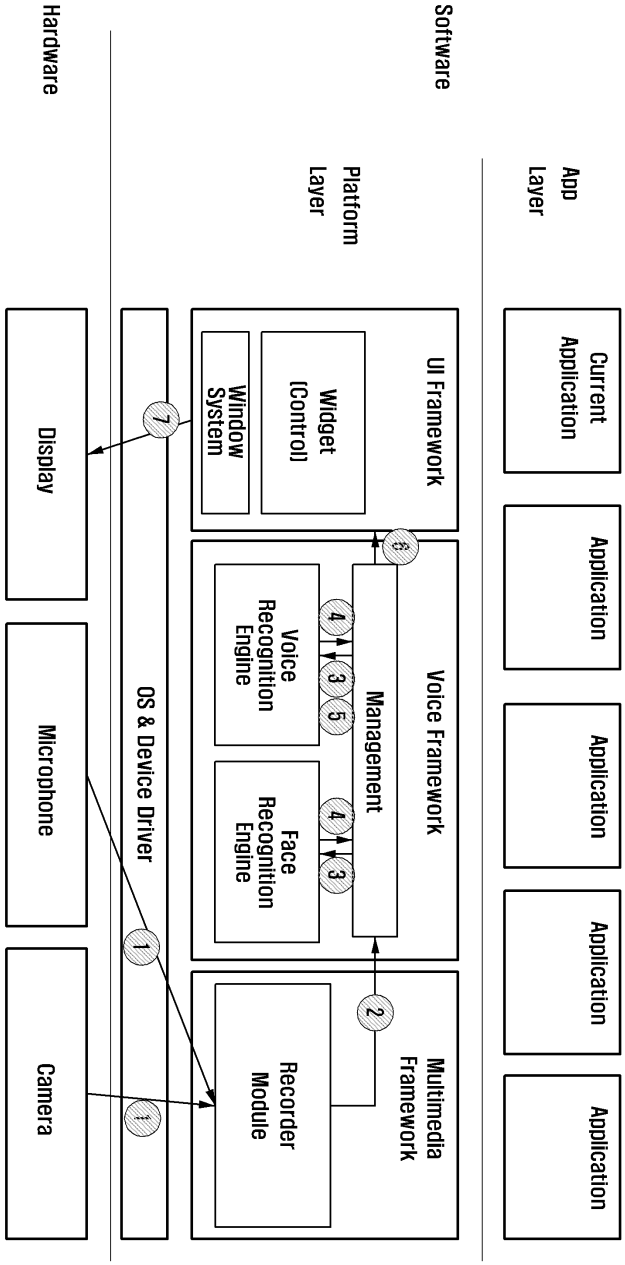
도면9


Message
10:18 AM
01030016057
▶Welcome to SLP world
12:48 AM
Create new message
widget
elm_widget *create_button;
....
elm_voice_tips *button_voice_tips = elm_voice_tips_add(win);
....
elm_voice_command_set (ELM_VOICE_COMMAND [button_voice_tips]
create_button, "Create", BUTTON_EVENT_CLICK);

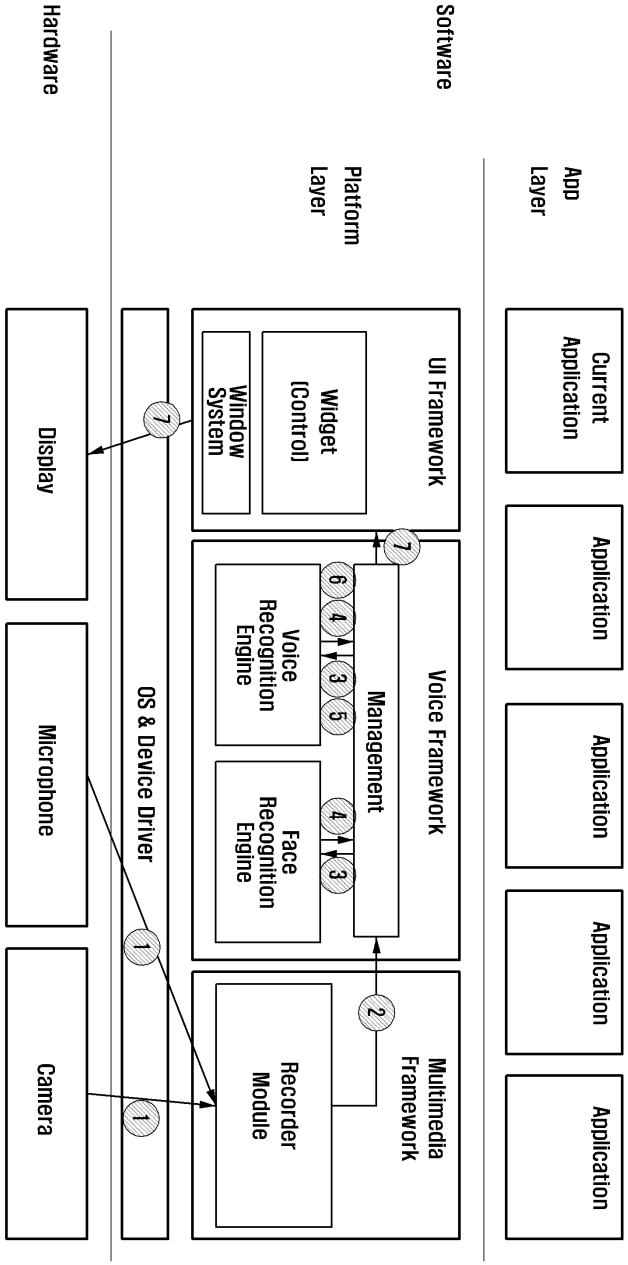
도면10


Hardware
Software
App Layer
Platform Layer
Current Application
Application
Application
Application
Application
UI Framework
Widget (Control)
Window System
Voice Framework
Management
Voice Recognition Engine
Face Recognition Engine
Multimedia Framework
Recorder Module
OS & Device Driver
Display
Microphone
Camera
1
2
3
4
5
6
7
8

도면11


Hardware
Software
App Layer
Platform Layer
Current Application
Application
Application
Application
Application
UI Framework
Widget [Control]
Window System
Voice Framework
Management
Voice Recognition Engine
Face Recognition Engine
Multimedia Framework
Recorder Module
OS & Device Driver
Display
Microphone
Camera
1
2
3
4
5
6
7

도면12


App Layer
Platform Layer
Software
Hardware
Current Application
Application
Application
Application
Application
UI Framework
Widget [Control]
Window System
Voice Framework
Management
Voice Recognition Engine
Face Recognition Engine
Multimedia Framework
Recorder Module
OS & Device Driver
Display
Microphone
Camera