



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0032096
(43) 공개일자 2017년03월22일

(51) 국제특허분류(Int. Cl.)
G10L 15/34 (2013.01) G10L 15/08 (2006.01)
(52) CPC특허분류
G10L 15/34 (2013.01)
G10L 15/08 (2013.01)
(21) 출원번호 10-2015-0129901
(22) 출원일자 2015년09월14일
심사청구일자 없음

(71) 출원인
삼성전자주식회사
경기도 수원시 영통구 삼성로 129 (매탄동)
(72) 발명자
배재현
경기도 수원시 영통구 삼성로168번길 74, 403호 (매탄동)
(74) 대리인
정홍식, 김태현

전체 청구항 수 : 총 18 항

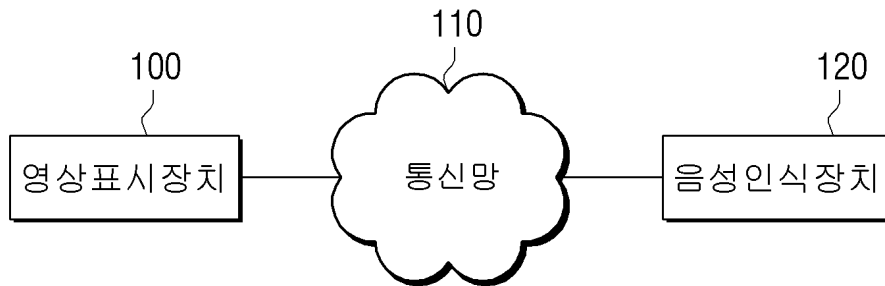
(54) 발명의 명칭 전자장치, 전자장치의 구동방법, 음성인식장치, 음성인식장치의 구동 방법 및 컴퓨터 판독가능 기록매체

(57) 요약

본 발명은 전자장치, 전자장치의 구동방법, 음성인식장치, 음성인식장치의 구동 방법 및 컴퓨터 판독가능 기록매체에 관한 것으로서, 본 발명의 실시예에 따른 음성인식시스템은 사용자가 발화한 음성 발성에 대한 음성 신호를 외부로 전송하는 전자장치, 및 전송한 음성 신호를 복수의 음성 인식기로 병렬 처리하여 얻은 인식 결과 중에서 기설정된 조건을 만족하는 인식 결과를, 전송한 음성 신호의 인식 결과로 결정하여 전자장치에 제공하는 음성인식장치를 포함할 수 있다.

대표도 - 도1

90



명세서

청구범위

청구항 1

사용자가 발화한 음성 발성에 대한 음성 신호를 선택적으로 외부로 전송하는 전자장치; 및
 상기 전송한 음성 신호를 복수의 음성 인식기로 병렬 처리하여 얻은 인식 결과 중에서 기설정된 조건을 만족하는 인식 결과를 상기 전송한 음성 신호의 인식 결과로 결정하여 상기 전자장치에 제공하는 음성인식장치;를
 포함하는 음성인식시스템.

청구항 2

사용자가 발화한 음성 발성에 대한 음성 신호를 전자장치로부터 수신하는 통신 인터페이스부; 및
 상기 수신한 음성 신호를 복수의 음성인식기로 병렬 처리하여 얻은 인식 결과 중에서 기설정된 조건을 만족하는 인식 결과를 상기 수신한 음성 신호의 인식 결과로 결정하고, 상기 결정한 인식 결과를 상기 전자장치로 전송하도록 상기 통신 인터페이스부를 제어하는 음성인식처리부;를
 포함하는 음성인식장치.

청구항 3

제2항에 있어서,
 상기 음성인식처리부는, 상기 인식 결과를 출력하는 응답 속도 및 상기 인식 결과의 신뢰 정도를 나타내는 유사도를 이용하여 상기 기설정된 조건을 만족하는지 판단하는 것을 특징으로 하는 음성인식장치.

청구항 4

제3항에 있어서,
 상기 음성인식처리부는, 상기 응답 속도가 빠른 인식 결과 중에서 상기 유사도가, 설정된 임계값보다 큰 인식 결과를 상기 전자장치에 제공하는 것을 특징으로 하는 음성인식장치.

청구항 5

제4항에 있어서,
 상기 음성인식처리부는, 상기 응답 속도가 빠른 선 순위의 인식 결과 중에서 설정된 임계값보다 작은 유사도를 갖는 복수의 인식 결과가 있으면, 설정된 시간 범위 내에서 차 순위로 제공되는 인식 결과를 참조해서 상기 전자장치로 제공하기 위한 인식 결과를 확정하는 것을 특징으로 하는 음성인식장치.

청구항 6

제5항에 있어서,
 상기 음성인식처리부는, 차 순위 인식 결과와 일치하는 선 순위 인식 결과를 선택하여 상기 전자장치로 제공하는 것을 특징으로 하는 음성인식장치.

청구항 7

제3항에 있어서,
 상기 음성인식처리부는, 설정된 시간 범위 내에 상기 복수의 음성인식기로부터 얻은 인식 결과가 없으면, 상기 전자장치로 인식 결과가 없음을 알리는 것을 특징으로 하는 음성인식장치.

청구항 8

사용자가 발화한 음성 발성에 대한 음성 신호를 전자장치로부터 수신하는 단계;

상기 수신한 음성 신호를 복수의 음성인식기로 병렬 처리하여 얻은 인식 결과 중에서 기설정된 조건을 만족하는 인식 결과를 상기 수신한 음성 신호의 인식 결과로 결정하는 단계; 및

상기 결정한 인식 결과를 상기 전자장치로 제공하는 단계;를

포함하는 음성인식장치의 구동방법.

청구항 9

제8항에 있어서,

상기 인식 결과를 결정하는 단계는,

상기 인식 결과를 출력하는 응답 속도 및 상기 인식 결과의 신뢰 정도를 나타내는 유사도를 이용하여 상기 기설정된 조건을 만족하는지 판단하는 단계를 포함하는 것을 특징으로 하는 음성인식장치의 구동방법.

청구항 10

제9항에 있어서,

상기 전자장치로 제공하는 단계는,

상기 응답 속도가 빠른 인식 결과 중에서 상기 유사도가, 설정된 임계값보다 큰 인식 결과를 상기 전자장치에 제공하는 것을 특징으로 하는 음성인식장치의 구동방법.

청구항 11

제10항에 있어서,

상기 인식 결과로 결정하는 단계는,

상기 응답 속도가 빠른 선 순위 인식 결과 중에서 설정된 임계값보다 작은 유사도를 갖는 복수의 인식 결과가 있으면, 설정된 시간 범위 내에서 차 순위로 제공되는 인식 결과를 참조해서 상기 전자장치로 제공하기 위한 인식 결과를 확정하는 것을 특징으로 하는 음성인식장치의 구동방법.

청구항 12

제11항에 있어서,

상기 전자장치로 제공하는 단계는,

상기 차 순위 인식 결과와 일치하는 선 순위 인식 결과를 선택하여 상기 전자장치로 제공하는 것을 특징으로 하는 음성인식장치의 구동방법.

청구항 13

제9항에 있어서,

설정된 시간 범위 내에 상기 복수의 음성인식기로부터 얻은 인식 결과가 없으면, 상기 전자장치로 인식 결과가 없음을 알리는 단계;를 더 포함하는 것을 특징으로 하는 음성인식장치의 구동방법.

청구항 14

음성인식장치의 구동방법을 실행하기 위한 프로그램을 포함하는 컴퓨터 판독가능 기록매체에 있어서,

상기 음성인식장치의 구동방법은,

사용자가 발화한 음성 발성에 대한 음성 신호를 전자장치로부터 수신하는 단계;

상기 수신한 음성 신호를 복수의 음성인식기로 병렬 처리하여 얻은 인식 결과 중에서 기설정된 조건을 만족하는 인식 결과를 상기 수신한 음성 신호의 인식 결과로 결정하는 단계; 및

상기 결정한 인식 결과를 상기 전자장치로 제공하는 단계;를

실행하는 컴퓨터 판독가능 기록매체.

청구항 15

사용자가 발화한 음성 발성에 대한 음성 신호를 취득하는 음성 취득부; 및

상기 취득한 음성 신호를 복수의 음성인식기로 각각 제공하여 얻은 인식 결과 중에서 기설정된 조건을 만족하는 인식 결과를 상기 수신한 음성 신호의 인식 결과로 결정하고, 상기 결정한 인식 결과에 따른 동작을 수행하는 음성인식처리부;를

포함하는 전자장치.

청구항 16

제15항에 있어서,

상기 취득한 음성 신호를 외부의 음성인식장치로 전송하는 통신 인터페이스부;를 더 포함하는 것을 특징으로 하는 전자장치.

청구항 17

사용자가 발화한 음성 발성에 대한 음성 신호를 취득하는 단계;

상기 취득한 음성 신호를 복수의 음성인식기로 병렬 처리하여 얻은 인식 결과 중에서 기설정된 조건을 만족하는 인식 결과를 상기 취득한 음성 신호의 인식 결과로 결정하는 단계; 및

상기 결정한 인식 결과에 따른 동작을 수행하는 단계;를

포함하는 전자장치의 구동방법.

청구항 18

제17항에 있어서,

상기 취득한 음성 신호를 외부의 음성인식장치로 전송하는 단계;를 더 포함하는 것을 특징으로 하는 전자장치의 구동방법.

발명의 설명

기술 분야

[0001] 본 발명은 전자장치, 전자장치의 구동방법, 음성인식장치, 음성인식장치의 구동 방법 및 컴퓨터 판독가능 기록 매체에 관한 것으로서, 더 상세하게는 가령 사용자의 음성 발성이 수신되면, 전자장치에 탑재되거나 네트워크에 연결된 복수의 음성인식기를 동시에 가동시켜, 수신된 음성 발성에 대한 인식 결과를 빠르고 정확하게 얻으려는 전자장치, 전자장치의 구동방법, 음성인식장치, 음성인식장치의 구동 방법 및 컴퓨터 판독가능 기록매체에 관련 된다.

배경 기술

[0002] 일반적으로 TV 등의 전자장치는 여러 종류의 음성인식엔진(기)을 포함할 수 있다. 예를 들어, 어떠한 음성인식 엔진은 미리 등록된 명령어를 인식할 때 동작할 수 있고, 또 어떠한 음성인식엔진은 검색 동작을 위한 음성 발성을 처리할 때 동작할 수 있다. 이러한 방식은 통상 시스템 설계자가 규정한 대로 동작하며, 종래에는 아비트레이션(arbitration)을 사용하여 가용한 여러 인식기 중 하나를 선택하여 인식 결과를 산출하였다. 여기서, 아비트레이션의 사전적 의미는 가령 여러 개의 CPU를 서로 제어하면서 동작시키는 것을 의미한다.

[0003] 종래에는 가령 검색결과가 나오기 전 네트워크 연결 유무, 인식 도메인(domain) 즉 영역 지정, 음성인식을 수행하는 장치의 유틸 리소스 등 음성인식기를 사용할 수 있는 사전 조건에 따라 동작할 인식기를 선택하였다. 예를 들어 네트워크로 연결된 음성인식기와 장치에 내장된 임베디드 인식기 중에서 선택할 경우 네트워크 연결 유무 및 통신 속도 등에 따라 사용할 음성 인식기를 선택하는 방식이 사용되었다.

[0004] 다른 방법으로는, 장치 내부에 내장된 하나 이상의 임베디드 인식기 혹은 유/무선 네트워크 등으로 연결된 하나

이상의 인식기의 결과를 모두 취합하여 최적의 결과를 선택하는 방식이 사용되었다.

- [0005] 즉 종래기술은 장치에 하나 이상의 내장형 인식기 혹은 네트워크를 이용한 인식기가 혼용될 경우, 지정된 인식도메인 혹은 인터넷에 연결되어 있는지 등의 사전 정보를 바탕으로 동작할 음성인식기를 선택하는 방식이거나, 혹은 사용할 용도 혹은 도메인에 따라 미리 어떠한 음성인식기를 사용할지 미리 정하는 방식이거나 여러 인식기의 동작 결과를 모두 수신한 후, 최적의 결과를 선택하는 방식이다.
- [0006] 그런데, 종래 방식은 사전정보와 일치하지 않은 발성이 입력될 경우 인식률이 하락할 수 있고, 또 최적의 결과를 도출하는 데에 실패할 가능성이 있다.
- [0007] 또한, 모든 음성인식기의 결과를 수신한 후, 최적의 결과를 선택하는 방식은 모든 음성인식기의 결과를 수신한 후 동작하여야 하므로, 여러 인식기 간에 결과 수신 시간이 서로 다를 경우 빠른 시간 내에 음성 발성에 대한 최종 결과를 도출할 수 없을 수 있다.

발명의 내용

- [0008] 본 발명의 실시예는 가령 사용자의 음성 발성이 수신되면, 전자장치에 탑재되거나 네트워크로 연결된 복수의 음성인식기를 동시에 가동시켜, 수신된 음성 발성에 대한 인식 결과를 빠르고 정확하게 얻으려는 전자장치, 전자장치의 구동방법, 음성인식장치, 음성인식장치의 구동 방법 및 컴퓨터 판독가능 기록매체를 제공함에 그 목적이 있다.
- [0009] 본 발명의 실시예에 따른 음성인식시스템은 사용자가 발화한 음성 발성에 대한 음성 신호를 선택적으로 외부로 전송하는 전자장치, 및 상기 전송한 음성 신호를 복수의 음성 인식기로 병렬 처리하여 얻은 인식 결과 중에서 기설정된 조건을 만족하는 인식 결과를 상기 전송된 음성 신호의 인식 결과로 결정하여 상기 전자장치에 제공하는 음성인식장치를 포함한다.
- [0010] 또한, 본 발명의 실시예에 따른 음성인식장치는 사용자가 발화한 음성 발성에 대한 음성 신호를 전자장치로부터 수신하는 통신 인터페이스부, 및 상기 수신한 음성 신호를 복수의 음성인식기로 병렬 처리하여 얻은 인식 결과 중에서 기설정된 조건을 만족하는 인식 결과를 상기 수신한 음성 신호의 인식 결과로 결정하고, 상기 결정한 인식 결과를 상기 전자장치로 전송하도록 상기 통신 인터페이스부를 제어하는 음성인식처리부를 포함한다.
- [0011] 상기 음성인식처리부는, 상기 인식 결과를 출력하는 응답 속도 및 상기 인식 결과의 신뢰 정도를 나타내는 유사도를 이용하여 상기 기설정된 조건을 만족하는지 판단할 수 있다.
- [0012] 상기 음성인식처리부는, 상기 응답 속도가 빠른 인식 결과 중에서 상기 유사도가, 설정된 임계값보다 큰 인식 결과를 상기 전자장치에 제공할 수 있다.
- [0013] 상기 음성인식처리부는, 상기 응답 속도가 빠른 선 순위의 인식 결과 중에서 설정된 임계값보다 작은 유사도를 갖는 복수의 인식 결과가 있으면, 설정된 시간 범위 내에서 차 순위로 제공되는 인식 결과를 참조해서 상기 전자장치로 제공하기 위한 인식 결과를 확정할 수 있다.
- [0014] 상기 음성인식처리부는, 차 순위 인식 결과와 일치하는 선 순위 인식 결과를 선택하여 상기 전자장치로 제공할 수 있다.
- [0015] 상기 음성인식처리부는, 설정된 시간 범위 내에 상기 복수의 음성인식기로부터 얻는 인식 결과가 없으면, 상기 전자장치로 인식 결과가 없음을 알릴 수 있다.
- [0016] 본 발명의 실시예에 따른 음성인식장치의 구동방법은 사용자가 발화한 음성 발성에 대한 음성 신호를 전자장치로부터 수신하는 단계, 상기 수신한 음성 신호를 복수의 음성인식기로 병렬 처리하여 얻은 인식 결과 중에서 기설정된 조건을 만족하는 인식 결과를 상기 수신한 음성 신호의 인식 결과로 결정하는 단계, 및 상기 결정한 인식 결과를 상기 전자장치로 제공하는 단계를 포함한다.
- [0017] 상기 인식 결과를 결정하는 단계는, 상기 인식 결과를 출력하는 응답 속도 및 상기 인식 결과의 신뢰 정도를 나타내는 유사도를 이용하여 상기 기설정된 조건을 만족하는지 판단하는 단계를 포함할 수 있다.
- [0018] 상기 전자장치로 제공하는 단계는, 상기 응답 속도가 빠른 인식 결과 중에서 상기 유사도가, 설정된 임계값보다 큰 인식 결과를 상기 전자장치에 제공할 수 있다.
- [0019] 상기 인식 결과를 결정하는 단계는, 상기 응답 속도가 빠른 선 순위 인식 결과 중에서 설정된 임계값보다 작은 유사도를 갖는 복수의 인식 결과가 있으면, 설정된 시간 범위 내에서 차 순위로 제공되는 인식 결과를 참조해서

상기 전자장치로 제공하기 위한 인식 결과를 확정할 수 있다.

- [0020] 상기 전자장치로 제공하는 단계는, 상기 차 순위 인식 결과와 일치하는 선 순위 인식 결과를 선택하여 상기 전자장치로 제공할 수 있다.
- [0021] 상기 음성인식장치의 구동방법은 설정된 시간 범위 내에 상기 복수의 음성인식기로부터 얻는 인식 결과가 없으면, 상기 전자장치로 인식 결과가 없음을 알리는 단계를 더 포함할 수 있다.
- [0022] 한편, 본 발명의 실시예에 따른 컴퓨터 판독가능 기록매체는 음성인식장치의 구동방법을 실행하기 위한 프로그램을 포함하는 컴퓨터 판독가능 기록매체에 있어서, 상기 음성인식장치의 구동방법은, 사용자가 발화한 음성 발성에 대한 음성 신호를 전자장치로부터 수신하는 단계, 상기 수신한 음성 신호를 복수의 음성인식기로 병렬 처리하여 얻은 인식 결과 중에서 기설정된 조건을 만족하는 인식 결과를 상기 수신한 음성 신호의 인식 결과로 결정하는 단계, 및 상기 결정한 인식 결과를 상기 전자장치로 제공하는 단계를 실행한다.
- [0023] 본 발명의 실시예에 따른 전자장치는 사용자가 발화한 음성 발성에 대한 음성 신호를 취득하는 음성 취득부, 및 상기 취득한 음성 신호를 복수의 음성인식기로 각각 제공하여 얻은 인식 결과 중에서 기설정된 조건을 만족하는 인식 결과를 상기 수신한 음성 신호의 인식 결과로 결정하고, 상기 결정한 인식 결과에 따른 동작을 수행하는 음성인식처리부를 포함한다.
- [0024] 상기 전자장치는, 상기 취득한 음성 발성에 대한 음성 신호를 외부의 음성인식장치로 전송하는 통신 인터페이스부를 더 포함할 수 있다.
- [0025] 또한, 본 발명의 실시예에 따른 전자장치의 구동방법은, 사용자가 발화한 음성 발성에 대한 음성 신호를 취득하는 단계, 상기 취득한 음성 신호를 복수의 음성인식기로 병렬 처리하여 얻은 인식 결과 중에서 기설정된 조건을 만족하는 인식 결과를 상기 취득한 음성 신호의 인식 결과로 결정하는 단계, 및 상기 결정한 인식 결과에 따른 동작을 수행하는 단계를 포함한다.
- [0026] 상기 전자장치의 구동방법은 상기 취득한 음성 신호를 외부의 음성인식장치로 전송하는 단계를 더 포함할 수 있다.

도면의 간단한 설명

- [0027] 도 1은 본 발명의 제1 실시예에 따른 음성인식시스템을 나타내는 도면,
- 도 2는 본 발명의 제2 실시예에 따른 음성인식시스템을 나타내는 도면,
- 도 3은 도 1 및 도 2의 영상표시장치의 세부 구조를 예시하여 나타낸 블록다이어그램,
- 도 4는 도 1 및 도 2의 영상표시장치의 다른 세부 구조를 예시하여 나타낸 블록다이어그램,
- 도 5는 도 1 및 도 2의 영상표시장치의 또 다른 세부 구조를 예시하여 나타낸 블록다이어그램,
- 도 6은 도 5의 제어부의 구조를 예시하여 나타낸 도면,
- 도 7은 도 3 내지 도 5의 음성인식처리부 및 음성인식실행부의 세부 구조를 예시하여 나타낸 블록다이어그램,
- 도 8은 도 1 및 도 2의 음성인식장치의 세부 구조를 예시하여 나타낸 블록다이어그램,
- 도 9는 도 1 및 도 2의 음성인식장치의 다른 세부 구조를 예시하여 나타낸 블록다이어그램,
- 도 10은 도 8 및 도 9의 음성인식처리부 및 음성인식실행부의 세부 구조를 예시하여 나타낸 블록다이어그램,
- 도 11은 도 1의 시스템에서의 음성인식과정을 예시하여 나타낸 도면,
- 도 12는 도 1의 시스템에서의 다른 음성인식과정을 예시하여 나타낸 도면,
- 도 13은 도 2의 시스템에서의 음성인식과정을 예시하여 나타낸 도면,
- 도 14는 본 발명의 실시예에 따른 영상표시장치의 구동 과정을 나타내는 흐름도,
- 도 15는 본 발명의 제1 실시예에 따른 음성인식장치의 구동 과정을 나타내는 흐름도, 그리고
- 도 16은 본 발명의 제2 실시예에 따른 음성인식장치의 구동 과정을 나타내는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0028] 이하, 도면을 참조하여 본 발명의 실시예에 대하여 상세히 설명한다.
- [0029] 도 1은 본 발명의 제1 실시예에 따른 음성인식시스템을 나타내는 도면이다.
- [0030] 도 1에 도시된 바와 같이, 본 발명의 제1 실시예에 따른 음성인식시스템(90)은 영상표시장치(100), 통신망(110) 및 음성인식장치(120)의 일부 또는 전부를 포함할 수 있다.
- [0031] 여기서, "일부 또는 전부를 포함한다"는 것은 시스템(90)에서 통신망(110)이 생략되어 영상표시장치(100)와 음성인식장치(120)가 다이렉트 통신(ex. P2P)을 수행하거나, 영상표시장치(100)가 스탠드 얼론(stand alone) 형태로 통신망(110)이나 음성인식장치(120)와 연동 없이 자체적으로 음성인식동작을 수행할 수 있는 것 등을 의미하는 것으로서, 발명의 충분한 이해를 돕기 위하여 전부 포함하는 것으로 설명한다.
- [0032] 영상표시장치(100)는 휴대폰, 랩탑 컴퓨터, 데스크탑 컴퓨터, 태블릿 PC, PDP, MP3 및 TV 등과 같이 영상표시가 가능한 장치를 포함한다. 또한, 본 발명의 실시예에 따른 영상표시장치(100)는 클라우드 단말기의 하나일 수 있다. 다시 말해, 사용자가 영상표시장치(100)의 특정 기능을 실행시키거나 동작을 수행시키기 위해 단어나 문장 형태의 음성 발성(혹은 사용자 명령)을 발화하는 경우, 이러한 음성 발성(혹은 음성 발화)을 취득하여 오디오 데이터(혹은 음성 신호)의 형태로 통신망(110)을 경유해 음성인식장치(120)로 제공할 수 있다. 이후, 영상표시장치(100)는 음성인식장치(120)로부터 음성 발성에 대한 인식 결과를 수신하여 수신한 인식 결과에 근거해 특정 기능 또는 동작을 수행한다. 여기서, "특정 기능을 실행하거나 동작을 수행한다"는 것은 화면에 표시된 애플리케이션(이하, '어플'이라 함)을 실행시키거나, 영상표시장치(100)의 전원 오프, 채널 전환, 볼륨 조정 등과 같은 동작을 수행하는 것을 의미한다. 이의 과정에서, 영상표시장치(100)는 기설정된 UI 창을 화면에 팝업(pop-up)하여 어플의 실행을 사용자에게 알릴 수도 있을 것이다.
- [0033] 본 발명의 실시예에 따른 영상표시장치(100)는 클라우드 단말기와 같이 동작하기 위하여, 내부에 탑재된(embedded) 음성인식엔진 즉 음성인식기를 갖지 않을 수 있다. 여기서, 음성인식엔진은 음성인식기를 포함하는 상위 개념일 수 있다. 영상표시장치(100)는 사용자의 음성 발성을 취득한 후 오디오 데이터의 형태로 음성인식장치(120)에 제공할 수 있다. 만약 영상표시장치(100)가 음성인식기를 포함하는 경우라면, 영상표시장치(100)는 음성인식장치(120)보다 저사양의 음성인식기를 탑재할 수 있거나 동등 수준의 음성인식기를 탑재할 수도 있다. 예를 들어, 동등 수준의 음성인식기를 탑재한다면, 통상의 음성인식처리는 자체적으로 처리할 수 있겠지만, 그렇다 해도 영상표시장치(100)의 내부 부하가 있는 경우에는 외부의 음성인식장치(120)로 음성인식 동작을 요청할 수 있을 것이다.
- [0034] 이와 같이 영상표시장치(100)는 내부에 음성인식기를 탑재하는 경우, 자신이 처리할 지 외부의 음성인식장치(120)에서 처리하도록 할지를 판단할 수 있다. 가령 저사양의 음성인식기를 탑재하는 경우라면, 영상표시장치(100)는 수신된 음성 발성의 발화 길이를 확인할 수 있다. 따라서, 발화 길이가 짧은 음성 발성은 내부에 탑재된 음성인식기를 통해 인식 결과를 생성할 수 있다. 그리고, 인식된 인식 결과를 이용하여 볼륨 조정이나 채널 변환과 같은 동작을 수행하거나, 외부의 검색 서버 등으로 제공하여 검색 결과를 요청할 수 있다.
- [0035] 만약 음성인식장치(120)와 동등 수준의 음성인식기를 포함하는 경우에, 영상표시장치(100)는 내부 동작 상태나 네트워크 상태를 판단하여 적절히 음성인식 동작을 수행할 수 있다. 예를 들어, 내부적으로 처리해야 하는 부담이 클 경우, 즉 내부의 하드웨어나 소프트웨어 자원을 이용하여 음성인식동작을 수행하기에는 자원의 부하가 있는 경우에는 음성인식장치(120)로 수신된 음성명령어의 오디오 데이터를 전송한다. 반면, 통신망(110)의 네트워크 상태가 좋지 않다고 판단되는 경우에는 자체적으로 음성인식동작을 처리할 수 있으며, 부하가 있어 부담이 크다 하더라도 이를 감내어 동작을 수행할 수 있다.
- [0036] 이와 같이, 영상표시장치(100)가 내부적으로 음성 발성을 처리할지 아니면 외부의 음성인식장치(120)에서 처리할지를 결정하여, 내부 자원을 이용하여 처리하기로 결정하였다면, 영상표시장치(100)는 수신된 음성 발성에 대한 인식 결과를 얻기 위하여 내부에 탑재되어 있는 복수의 음성인식기를 동시에 가동시킬 수 있다. 다시 말해, 영상표시장치(100)는 각각의 용도에 부합한 음성인식기를 가질 수 있는데, 예를 들어, 검색 서버로 검색을 요청하는 경우라면 *-Voice와 같은 음성인식기를 실행시킬 수 있고, 음성 인식을 시작하는 발화 시작어인 '하이 티비'와 같은 '트리거 워드'를 인식하기 위한 음성인식기를 실행시키는 등 이와 같은 다양한 음성인식기를 포함할 수 있다.
- [0037] 음성 인식기를 다시 구분해 보면, 채널 변환은 튜너를 제어하는 것에 관련될 것이고, 볼륨 조정은 스피커의 음

량을 조절할 것에 관련될 것이므로 이는 곧 기본 기능 또는 하드웨어 자원을 제어하기 위한 음성인식기에 해당된다. 반면, '하이 티비' 등은 특정 어플과 같은 부가 기능 또는 소프트웨어 자원을 실행하기 위한 음성인식기에 해당될 것이다. 또는 복수의 음성 인식기는 미리 정해진 단어 후보 중에서 인식하는 인식기와 미리 정해지지 않은 단어 혹은 문장 등을 인식하는 인식기 등을 포함할 수 있을 것이다.

[0038] 본 발명의 실시예에 따른 영상표시장치(100)는 내부에 탑재된 복수의 음성인식기를 동시에 가동시켜, 가장 먼저 응답이 있는 즉 인식 결과를 내어주는 음성인식기의 인식 결과를, 취득한 음성 발성의 인식 결과로 사용할지를 판단하게 된다. 보통 각각의 음성인식기가, 수신된 음성 발성에 대한 인식 결과를 내어줄 때, 해당 인식 결과의 정확도 즉 신뢰 정도가 얼마나 되는지에 관련된 유사도(또는 유사도 점수)를 함께 내어주게 되는데, 영상표시장치(100)는 응답 속도가 빠른 인식 결과 중에서 유사도 점수가 높은 인식 결과를 사용자가 발화한 음성 발성에 대한 인식 결과로 확정하게 되는 것이다. 그리고 이를 통해 사용자가 의도한 동작을 수행하게 된다. 따라서, 다른 음성인식기보다 가장 빠르게 인식 결과를 내어준 음성인식기의 인식 결과의 유사도 점수가 높다면, 나머지 인식 결과는 버리고 해당 인식 결과만을 사용하면 된다.

[0039] 다만, 영상표시장치(100)는 보다 정확도가 높은 인식 결과를 도출하기 위하여, 기설정된 조건을 만족하는지를 더 판단할 수 있다. 예를 들어, 사용자가 발화할 때 영상표시장치(100)를 즉각적으로 동작시키기 위하여 기설정된 시간 범위 내에 있는 인식 결과만을 사용할 수 있다. 그리고, 주어진 시간 내의 인식 결과 중에서 유사도 점수는 설정된 임계값을 넘어야 한다. 따라서, 임계값을 초과하는 인식 결과는 무조건 사용자가 의도한 동작의 인식 결과로 반영할 수 있다. 물론 하나의 인식기는 복수 개의 인식 결과를 동시에 내어줄 수 있기 때문에 임계값을 초과하는 인식 결과가 복수 개 존재할 수도 있다. 이의 경우에는 유사도 점수가 높은 인식 결과를 최종 인식 결과로 확정할 수 있지만, 유사도 점수의 차이가 그다지 크지 않다면, 이의 경우에는 다른 추가 정보를 활용할 수 있다.

[0040] 예를 들어, 주어진 시간 범위 내에만 있다면, 차 순위의 인식 결과를 더 확인해 보는 것이다. 확인 결과, 차 순위 인식 결과와 일치하는 선 순위의 인식 결과 있다면 해당 인식 결과를 최종적으로 확정한다. 그러나 차 순위 인식 결과와 일치하는 선 순위 인식 결과가 없다면, 이의 경우에는 차 순위에 있는 복수의 인식 결과에 대한 유사도 점수 차가 설정된 임계 차값을 벗어나지 않는 한, 복수의 인식 결과들 중에서 유사도 점수가 가장 큰 인식 결과를 최종적으로 확정할 수도 있다. 이와 관련해서는 이후에 자세히 살펴보기로 한다.

[0041] 또한, 주변의 음성인식장치(120)에서 제공하는 인식 결과를 참조하여 공통으로 존재하는 인식 결과를 최종적으로 확정할 수도 있다. 만약 유사도 점수가 임계값보다 높은 인식 결과가 없고 임계값보다 낮은 인식 결과만 존재한다면, 이의 경우에도 비교적 높은 유사도 범위에 있는 인식 결과를 활용하되, 이의 경우에도 주변의 음성인식장치(120)에서 제공하는 인식 결과를 참조할 수 있다.

[0042] 이와 같이, 영상표시장치는 음성인식이라는 목적 즉 용도는 같지만, 음성인식의 범주가 서로 다른 복수의 음성인식기를 동시에 가동 즉 병렬 처리함으로써 응답 속도가 빠른 음성인식기의 인식 결과를 이용할 수 있어 음성인식동작이 빠르게 이루어지며, 또 취득한 일정 시간 내의 인식 결과 중에서 기설정된 조건을 만족하는 인식 결과를 최종적으로 확정하기 때문에 그만큼 정확도도 증가시킬 수 있다.

[0043] 본 발명의 실시예에서는 복수의 음성인식기를 동시에 가동시키는 것을 "병렬 처리"라 명명하였다. "병렬 처리(parallel processing)"라 함은 서로 다른 입력(부)과 출력(부)에 대하여 복수의 음성 인식기가 서로 병렬 연결된 것을 의미하며, 따라서 음성 발성, 더 정확하게는 음성 발성에 대한 오디오 데이터가 입력되는 입력 경로와 인식 결과가 출력되는 출력 경로는 분명 다르다. 이러한 점에서, 입력과 출력이 하나인 '분산 처리(distribution processing)'와는 분명한 차이가 있다. 여기서 "분산 처리"는 음성 발성이 동시에 입력되지 않는 것을 의미한다.

[0044] 통신망(110)은 유무선 통신망을 모두 포함한다. 여기서 유선망은 케이블망이나 공중 전화망(PSTN)과 같은 인터넷망을 포함하는 것이고, 무선 통신망은 CDMA, WCDMA, GSM, EPC(Evolved Packet Core), LTE(Long Term Evolution), 와이브로 망 등을 포함하는 의미이다. 물론 본 발명의 실시예에 따른 통신망(110)은 이에 한정되는 것이 아니며, 향후 구현될 차세대 이동통신 시스템의 접속망으로서 가령 클라우드 컴퓨팅 환경하의 클라우드 컴퓨팅망 등에 사용될 수 있다. 가령, 통신망(110)이 유선 통신망인 경우 통신망(110) 내의 액세스포인트는 전화국의 교환국 등에 접속할 수 있지만, 무선 통신망인 경우에는 통신사에서 운영하는 SGSN 또는 GGSN(Gateway GPRS Support Node)에 접속하여 데이터를 처리하거나, BTS(Base Station Transmission), NodeB, e-NodeB 등의 다양한 중계기에 접속하여 데이터를 처리할 수 있다.

- [0045] 통신망(110)은 액세스포인트를 포함할 수 있다. 액세스포인트는 건물 내에 많이 설치되는 펨토(femto) 또는 피코(pico) 기지국과 같은 소형 기지국을 포함한다. 여기서, 펨토 또는 피코 기지국은 소형 기지국의 분류상 영상 표시장치(100)를 최대 몇 대까지 접속할 수 있느냐에 따라 구분된다. 물론 액세스포인트는 영상표시장치(100)와 지그비 및 와이파이(Wi-Fi) 등의 근거리 통신을 수행하기 위한 근거리 통신 모듈을 포함한다. 액세스포인트는 무선통신을 위하여 TCP/IP 혹은 RTSP(Real-Time Streaming Protocol)를 이용할 수 있다. 여기서, 근거리 통신은 와이파이 이외에 블루투스, 지그비, 적외선(IrDA), UHF(Ultra High Frequency) 및 VHF(Very High Frequency)와 같은 RF(Radio Frequency) 및 초광대역 통신(UWB) 등의 다양한 규격으로 수행될 수 있다. 이에 따라 액세스포인트는 데이터 패킷의 위치를 추출하고, 추출된 위치에 대한 최상의 통신 경로를 지정하며, 지정된 통신 경로를 따라 데이터 패킷을 다음 장치, 예컨대 영상표시장치(100)로 전달할 수 있다. 액세스포인트는 일반적인 네트워크 환경에서 여러 회선을 공유할 수 있으며, 예컨대 라우터(router), 리피터(repeater) 및 중계기 등이 포함될 수 있다.
- [0046] 음성인식장치(120)는 음성인식서버를 포함하며, 일종의 클라우드 서버로서 동작할 수 있다. 다시 말해, 음성인식장치(120)는 음성 인식과 관련한 모든(또는 일부의) HW 자원이나 SW 자원을 구비함으로써 최소한의 자원을 가진 영상표시장치(100)로부터 수신된 음성 발성에 대한 인식 결과를 생성하여 제공할 수 있다. 물론 본 발명의 실시예에 따른 음성인식장치(120)는 클라우드 서버에 한정되지는 않는다. 예를 들어, 통신망(110)이 생략 구성되어 영상표시장치(100)가 음성인식장치(120)와 직접 통신을 수행하는 경우, 음성인식장치(120)는 외부 장치 즉 액세스포인트이거나 데스크탑 컴퓨터와 같은 주변 장치가 될 수도 있다. 또는 영상표시장치(100)에서 제공한 음향 신호, 더 정확하게는 오디오 데이터에 대한 인식 결과만 제공해 줄 수 있다면 어떠한 형태의 장치이여도 무관하다. 이러한 점에서 음성인식장치(120)는 인식결과 제공장치가 될 수도 있을 것이다.
- [0047] 본 발명의 실시예에 따른 음성인식장치(120)는 사용자가 발화한 음성 발성에 대한 오디오 데이터를 영상표시장치(100)로부터 수신하는 경우, 이에 대한 인식 결과를 도출할 수 있다. 만약, 사용자가 스포츠 스타의 이름을 발화하여 검색을 요청한 경우라면, 검색어에 해당되는 음성 발성의 인식 결과에 근거하여 검색된 검색 결과를 영상표시장치(100)로 제공할 수도 있다. 반면, 영상표시장치(100)의 하드웨어(ex. 튜너)나 소프트웨어(ex. 어플)를 동작시키기 위한 음성 발성을 발화하였다면 이에 대한 인식 결과를 제공하게 될 것이다.
- [0048] 이의 과정에서, 음성인식장치(120)는 앞서 영상표시장치(100)를 설명하면서 충분히 살펴본 바와 같이, 음성인식 동작을 수행하되, 서로 다른 범주의 음성인식 동작을 수행하는 복수의 음성인식기를 동시에 가동하여 사용자가 의도한 최적의 인식 결과를 도출하게 된다. 예를 들어, 사용자가 "오늘 날씨 어때?"라고 발화하였다고 가정하자. 이의 경우, 영상표시장치(100)는 이에 대한 오디오 데이터를 음성인식장치(120)로 제공할 수 있을 것이다.
- [0049] 그러면, 음성인식장치(120)는 사용자가 발화한 음성 발성에 대한 오디오 데이터를 복수의 음성인식기로 입력시킨다. 이의 경우, 어떠한 음성인식기는 '오늘 날씨 어때'라는 텍스트 기반의 정확한 인식 결과를 내어 줄 수 있다. 그리고, 이에 대한 유사도 점수를 함께 출력할 것이다. 반면 어떠한 음성인식기는 '오늘 날씨 어때'라는 입력에 대하여 'MBC'나 'SBS'와 같은 인식 결과를 내어 주고, 이에 대한 유사도 점수를 함께 출력한다. 이의 경우, 음성인식장치(120)는 응답 속도가 빠른, 즉 가장 먼저 인식 결과를 내어준 음성인식기의 인식 결과를 확인(혹은 분석)한다. 이를 위하여, 음성인식장치(120)의 인식 결과에 연계되어 있는 유사도 점수 등을 확인할 수 있다. 예를 들어, 먼저 출력된 'MBC'나 'SBS'의 인식 결과가 유사도 점수가 낮은 경우, 음성인식장치(120)는 사용자에게 빠르게 응답하기 위하여 설정된 시간 범위 내에 있는 '오늘 날씨 어때'를 출력한 음성인식기의 인식 결과와, 해당 인식 결과의 유사도 점수를 확인함으로써 사용자가 의도한 동작에 대한 최적의 인식 결과를 찾는다. 이에 따라, 음성인식장치(120)는 사용자 질의에 대한 답변을 영상표시장치(100)로 제공해 줄 수 있다.
- [0050] 본 발명의 실시예에서는 이와 같은 최적의 인식 결과를 찾아내기 위하여, 가장 응답이 빠른 인식 결과를 최우선으로 고려하되, 정확도를 높이기 위하여 일정한 시간 범위 내에 있는 인식 결과들의 유사도 점수를 확인하는 등의 과정을 수행할 수 있는데, 이와 관련한 기타 자세한 내용은 앞서 영상표시장치(100)를 설명하면서 충분히 설명하였으므로 더 이상의 설명은 생략하도록 한다.
- [0051] 상기한 바와 같이, 영상표시장치(100) 또는 음성인식장치(120)는 사용자가 발화한 음성 발성에 대한 최적의 인식결과를 도출하기 위하여, 음성인식에 관련된 내부의 모든 자원을 동시에 가동시키되, 적어도 하나의 인식 결과 중 특정 조건을 만족시키는 인식 결과를 도출함으로써 응답 속도와 정확도를 동시에 증가시킬 수 있게 된다.
- [0052] 다시 말해, 종래와 같이 인식기를 동작하기 전 미리 동작시킬 인식기를 선택함으로써 올바른 결과를 얻지 못하는 것을 방지할 수 있고, 또 빠른 응답을 보이는 인식기의 결과만을 비교적 정확하게 사용자에게 응답할 수 있

어 인식기의 동작 환경에 따라 느린 응답 속도를 보이는 인식기들의 인식 결과를 기다려 비교할 필요가 없게 된다. 즉 본 발명의 실시예에서와 같이 여러 인식기를 동시에 사용하는 경우, 정확하면서도 빠른 응답 즉 인식 결과를 선택할 수 있어 인식 정확도와 빠른 응답 속도를 기대할 수 있을 것이다.

- [0053] 한편, 지금까지는 음성인식장치(120)가 영상표시장치(100)에 연계하여 동작하는 것을 기술하였지만, 본 발명의 실시예에 따르면 음성인식을 지원하는 모든 기기, 예를 들어, 도어(door) 시스템이나, 자동차 등 모든 기기에서 사용될 수 있으며, 이때에도 임베디드 및 서버 인식기 모두에서 활용될 수 있다. 여기서, 임베디드란 서버의 연계없이 영상표시장치(100)와 같은 개별 장치에서 위의 음성인식이 이루어질 수 있는 것을 의미한다. 따라서 본 발명의 실시예에서는 상기의 기기들을 통칭하여 '전자장치' 또는 '사용자 장치'라 명명할 수 있을 것이다.
- [0054] 도 2는 본 발명의 제2 실시예에 따른 음성인식시스템을 나타내는 도면이다.
- [0055] 도 2에 도시된 바와 같이, 본 발명의 제2 실시예에 따른 음성인식시스템(190)은 영상표시장치(200), 통신망(210) 및 복수의 음성인식장치(220)의 일부 또는 전부를 포함하며, 여기서 일부 또는 전부를 포함한다는 것은 앞서의 의미와 동일하다.
- [0056] 도 2의 음성인식시스템(190)을 도 1의 음성인식시스템(90)과 비교해 보면, 도 2의 음성인식장치 1(220-1)은 메인 장치로서 동작하여, 주변, 더 정확하게는 외부의 음성인식장치 2(220-2)로부터 사용자가 발화한 음성 발성에 대한 인식 결과를 제공받을 수 있다.
- [0057] 예를 들어, 사용자가 영상표시장치(200)로 음성 발성을 발화하면, 취득된 음성 발성의 오디오 데이터가 음성인식장치 1(220-1)과 음성인식장치 2(220-2)에 동시에 제공되게 된다. 물론 이때, 음성인식장치 1(220-1)과 음성인식장치 2(220-2)는 음성인식을 위해 동일 범주에 있는 음성인식기를 갖는 것이 바람직하다.
- [0058] 이에 따라, 음성인식장치 1(220-1)은 앞서 도 1에서 충분히 살펴본 바와 같이, 음성인식장치(120)와 동일한 동작을 수행하게 된다. 통상 하나의 인식기는 하나의 인식 결과가 아닌 유사도 점수가 비슷한 범위에서 복수의 인식 결과를 내어 줄 수 있다. 이의 경우, 유사도 점수가 비슷하기 때문에 어떠한 인식 결과가 사용자가 발화한 음성 발성에 부합한 인식 결과인지 확정하는 것이 어려울 수 있다. 이러한 점을 고려하여, 음성인식장치 1(220-1)은 음성인식장치 2(220-2)에서 제공한 인식 결과를 참조하여, 동일한 이름(또는 명칭)에 해당되는 인식 결과를 선택함으로써 정확도를 좀 더 높일 수 있게 될 것이다.
- [0059] 또한, 복수의 음성인식장치(220)가 서로 연동할 때, 음성인식장치 2(220-2)는 음성인식장치 1(220-1)부터 인식 결과의 요청이 있을 때, 인식 결과를 제공할 수 있지만, 별도의 요청이 없다 하더라도, 생성되는 순서대로 인식 결과를 제공하는 것이 얼마든지 가능하고, 이는 시스템 설계자에 의해 얼마든지 다양하게 변경될 수 있는 것이므로, 본 발명의 실시예에서는 어떠한 방식으로 연동하는지에 특별히 한정하지는 않을 것이다.
- [0060] 그 이외에 기타 영상표시장치(200), 통신망(210) 및 복수의 음성인식장치(220)와 관련해서는 도 1의 영상표시장치(100), 통신망(110) 및 음성인식장치(120)와 크게 다르지 않으므로 그 내용들로 대신하고자 한다.
- [0061] 도 3은 도 1 및 도 2의 영상표시장치의 세부 구조를 예시하여 나타낸 블록다이어그램이다.
- [0062] 설명의 편의상 도 3을 도 1과 함께 참조하면, 본 발명의 실시예에 따른 영상표시장치(100)는 음성 취득부(300) 및 음성인식처리부(310)의 일부 또는 전부를 포함한다.
- [0063] 여기서, "일부 또는 전부를 포함한다"는 것은 음성 취득부(300)와 같은 구성요소가 생략되어 영상표시장치(100)가 구성되거나, 음성 취득부(300)가 음성인식처리부(310)에 통합되어 구성될 수 있는 것 등을 의미하는 것으로서, 발명의 충분한 이해를 돕기 위하여 전부 포함하는 것으로 설명한다.
- [0064] 음성 취득부(300)는 사용자가 발화한 음성 발성을 취득하는 마이크로폰을 포함할 수 있다. 이의 경우는 영상표시장치(100)가 마이크로폰을 탑재한 경우에 해당된다. 그러나, 마이크로폰은 독립된 장치로서, 영상표시장치(100)의 외부에서 연결하여 사용하는 것도 얼마든지 가능하다. 이의 경우, 마이크로폰은 음성 취득부(300)에 연결될 수 있다. 따라서, 음성 취득부(300)는 연결부 즉 커넥터가 될 수 있으며, 이의 경우 음성 취득부(300)는 음성 발성을 수신하므로 취득한다고 볼 수 있다.
- [0065] 또한, 음성인식처리부(310)는 본 발명의 실시예에 따라 취득 또는 수신한 음성 발성을 복수의 음성인식기로 병렬 처리하여, 빠르면서 정확한 인식 결과를 확정한다. 사실, 도 3과 관련해서도 앞서 충분히 설명한 바 있지만, 영상표시장치(100)가 스탠드 얼론 형태로 동작할 때의 구조를 나타낸다. 예를 들어, 음성인식처리부(310)는 사용자가 발화한 음성 발성에 대한 최적의 인식 결과를 도출하여 내부의 메모리 또는 레지스트리에 저장할 수 있

기 때문이다. 여기서, 메모리는 HW 구성이라면, 레지스트리는 SW 구성을 의미한다.

- [0066] 이와 같이 저장된 인식 결과는, 이후에 시스템 설계자에 의해 분석되어 음성인식기를 교체해야 할지를 결정할 때에 사용될 수도 있을 것이다.
- [0067] 또한 음성인식처리부(310)는 인식 결과가 최종적으로 도출되었다고 판단되면, 음성 취득부(300)의 동작을 턴오프시킬 수도 있을 것이다.
- [0068] 이러한 점을 제외하면, 음성인식처리부(310)는 도 1의 영상표시장치(100)나 음성인식장치(120)를 통해 충분히 설명되었으므로 더 이상의 설명은 생략하도록 한다. 다만, 기타 추가되는 내용은 이후에 좀 더 다루기로 한다.
- [0069] 도 4는 도 1 및 도 2의 영상표시장치의 다른 세부 구조를 예시하여 나타낸 블록다이어그램이다.
- [0070] 설명의 편의상 도 4를 도 1과 함께 참조하면, 본 발명의 다른 실시예에 따른 영상표시장치(100')는 통신 인터페이스부(400), 음성인식처리부(410), 동작 수행부(420) 및 저장부(430)의 일부 또는 전부를 포함할 수 있다.
- [0071] 여기서, "일부 또는 전부를 포함한다"는 것은 통신인터페이스부(400) 및/또는 저장부(430)와 같은 일부 구성요소가 생략되어 구성되거나, 저장부(430)와 같은 일부 구성요소가 음성인식처리부(410)와 같은 다른 구성요소에 통합하여 구성될 수 있는 것 등을 의미하는 것으로서, 발명의 충분한 이해를 돕기 위하여 전부 포함하는 것으로 설명한다.
- [0072] 도 4의 구조는, 영상표시장치(100')가 내부에 음성인식기를 탑재하고 있고, 경우에 따라, 통신 인터페이스부(400)를 통해 외부의 음성인식장치, 가령 도 1의 음성인식장치(120)로 음성 발성을 전송하여 이에 대한 인식결과 또는 검색결과를 수신하는 것에 적합할 수 있다.
- [0073] 다시 말해, 통신 인터페이스부(400)는 가령 외부의 마이크로폰을 통해 수신된 사용자의 음성 발성을 음성인식처리부(410)로 전달할 수 있다. 이때, 통신 인터페이스부(400)는 외부의 마이크로폰으로부터 음성 발성을 유선 또는 무선으로 수신할 수 있을 것이다.
- [0074] 이어 음성인식처리부(410)는 수신된 음성 발성에 대하여 자신이 처리할지, 또는 도 1의 음성인식장치(120)로 인식결과를 요청할지를 결정할 수 있다. 이를 위하여, 음성인식처리부(410)는 먼저 음성 발성의 발화 길이를 확인한다. 만약, 발화한 음성 발성의 시작과 끝으로 판단된 구간의 시간이 설정된 시간 범위 내에 있다면, 내부 음성인식기들을 이용해 음성 발성의 오디오 데이터를 처리할 수 있다. 반면 설정된 시간 범위를 벗어나면 통신 인터페이스부(400)를 통해 음성인식장치(120)로 음성 발성의 오디오 데이터를 전송할 수 있다.
- [0075] 또한, 외부의 음성인식장치(120)로 음성 발성의 오디오 데이터를 전송하기에 앞서, 음성인식처리부(410)는 네트워크 상태를 점검할 수 있다. 만약 도 1의 통신망(110)의 상태가 불안정하고, 부하가 심하다고 판단되면, 사용자에게 음성인식이 어려움을 동작 수행부(420)를 통해 알릴 수 있다. 이를 위하여, 음성인식처리부(410)는 동작 수행부(420)를 통해 사용자에게 메시지를 출력하거나 음성을 출력해 줄 수 있을 것이다.
- [0076] 나아가, 음성인식처리부(410)는 내부에서 처리할 것으로 결정하였다면, 내부적으로 처리하기에 자원의 부담 즉 부하(load)가 있는지를 점검할 수 있다. 만약 부하가 심하다고 판단되면, 설정된 시간 범위 내에 있는 음성 발성이라 하더라도 외부의 음성인식장치(120)로 전송할 수 있을 것이다.
- [0077] 만약 내부적으로 처리하기에 큰 문제가 없다고 판단되면, 음성인식처리부(410)는 서로 다른 범주에 속하는 다양한 음성인식기를 동시에 가동시켜 수신된 음성 발성의 오디오 데이터를 분석하여 인식 결과를 출력하도록 한다. 이와 관련해서는 앞서 충분히 설명하였으므로 더 이상의 설명은 생략한다.
- [0078] 동작 수행부(420)는 튜너나 음향 출력부 및/또는 디스플레이부를 포함할 수 있다. 예를 들어, 음성인식처리부(410)는 사용자가 발화한 음성 발성이 '채널 변경'이었다면 튜너를 조절할 것이다. 반면 사용자가 발화한 음성 발성이 '볼륨 조정'에 관련된다면, 가령 "볼륨 올려"라고 발화하였다면, 음향 출력부로 출력되는 음량의 레벨을 올릴 것이다. 이를 위하여, 증폭기에서 출력되는 음량의 레벨을 증폭시킬 수 있다. 나아가, 사용자가 검색 동작을 위해서 "김*아"라고 발화하였다면, 내부의 고정발화엔진인 '*-Voice'를 실행시키고, 이를 사용자에게 알리기 위해 화면에 어플의 실행을 보여줄 수 있을 것이다.
- [0079] 이와 같이 본 발명의 실시예에 따른 동작 수행부(420)의 동작은 다양한 예가 가능하므로, 위의 내용에 특별히 한정하지는 않을 것이다.
- [0080] 저장부(430)는 롬(ROM) 또는 램(RAM), 하드디스크드라이브(HDD)와 같은 하드웨어 자원이 바람직하다. 저장부

(430)는 음성인식처리부(410)에서 처리되는 데이터를 임시 저장할 수 있으며, 음성인식처리부(410)에서 최적의 인식 결과를 도출하기 위해 필요한 다양한 정보를 저장할 수 있다. 일례로, 저장부(430)는 인식 결과의 유사도 점수를 기준값과 비교하기 위한 기준값 즉 임계값 관련 정보 등 다양한 정보를 저장할 수 있을 것이다.

- [0081] 도 5는 도 1 및 도 2에 도시된 영상표시장치의 또 다른 세부 구조를 예시하여 나타낸 블록다이어그램이고, 도 6은 도 5에 도시된 제어부의 구조를 예시하여 나타낸 도면이다.
- [0082] 설명의 편의상 도 5를 도 1과 함께 참조하면, 본 발명의 또 다른 실시예에 따른 영상표시장치(100')는 통신 인터페이스부(500), 음성 취득부(510), 제어부(520), 동작 수행부(530), 음성인식실행부(540) 및 저장부(550)의 일부 또는 전부를 포함하며, 여기서 일부 또는 전부를 포함한다는 것은 앞서의 의미와 동일하다.
- [0083] 도 5의 구조는, 도 4의 구조를 변형한 것에 해당된다. 물론 마이크로폰과 같은 음성 취득부(510)를 내부에 탑재한다는 점에서는 다르지만, 도 4의 음성인식처리부(310'/410')는 도 5에서와 같이, 하드웨어적으로 제어부(520)와 음성인식실행부(540)로 구분할 수 있다는 데에 더욱 차이가 있다.
- [0084] 물론 제어부(520)는 도 6에 예시한 바와 같이, 프로세서(600)와 메모리(610)를 포함할 수도 있다. 따라서, 제어부(520)는 도 6에서와 같이 메모리(610)를 포함하느냐 포함하지 않느냐에 따라 동작이 다소 다를 수 있다.
- [0085] 예를 들어, 사용자가 발화한 음성 발성이 수신되면, 제어부(520)는 음성인식실행부(540)를 실행시킨 후 음성 발성을 전달한다. 그러면, 음성인식실행부(540)는 수신된 음성 발성을 복수의 음성인식기로 병렬 처리하여, 수신된 음성 발성에 대한 최적의 인식 결과를 도출하여 제어부(520)에 제공해 줄 수 있다. 그러면, 제어부(520)는 해당 인식 결과를 근거로 다양한 동작을 수행하게 된다. 이러한 점에서 음성인식실행부(540)는 도 4의 음성인식처리부(410)와 크게 다르지 않지만, 음성인식처리부(410)는 소프트웨어적으로 제어 기능을 더 수행할 수 있다는 점에서 차이가 있다.
- [0086] 만약, 제어부(520)가 도 6에서와 같은 구조를 갖는 경우, 영상표시장치(100')는 시스템 즉 장치의 초기 구동시 음성인식실행부(540)에 저장된 음성인식기(엔진) 관련 프로그램을 도 6의 메모리(610)에 로딩하여 저장시킨다. 그리고, 프로세서(600)는 음성 발성이 수신되면, 메모리(610)에 저장된 프로그램을 실행시켜 즉 복수의 음성인식기를 병렬 처리하여 최적의 인식 결과를 도출한다. 이와 같이 동작하는 경우, 위의 경우에 비하여 그만큼 데이터 처리가 빠른 장점이 있다.
- [0087] 이러한 점을 제외하면, 도 5의 통신 인터페이스부(500), 제어부(520), 동작 수행부(530), 음성인식실행부(540) 및 저장부(550)는 도 4의 통신 인터페이스부(400), 음성인식처리부(410), 동작 수행부(420) 및 저장부(430)와 크게 다르지 않으므로 그 내용들로 대신하고자 한다.
- [0088] 도 7은 도 3 내지 도 5에 도시된 음성인식처리부 및 음성인식실행부의 세부 구조를 예시하여 나타낸 블록다이어그램이다.
- [0089] 설명의 편의상 도 7을 도 5와 함께 참조하면, 본 발명의 실시예에 따른 음성인식실행부(540)는 음성 입력부(모듈)(700), 아비트레이션부(모듈)(710), 복수의 음성인식기(720) 및 인식결과 처리부(모듈)(730)의 일부 또는 전부를 포함할 수 있다.
- [0090] 여기서, 일부 또는 전부를 포함한다는 것은 음성 입력부(700)나 인식결과 처리부(730)와 같은 구성요소가 생략되거나 아비트레이션부(710)와 같은 다른 구성요소에 통합될 수 있는 것 등을 의미하는 것으로서, 발명의 충분한 이해를 돕기 위하여 전부 포함하는 것으로 설명한다.
- [0091] 또한, 본 발명의 실시예에 따라 "부"는 HW를 의미하고, "모듈"은 소프트웨어를 의미할 수 있지만, 소프트웨어는 하드웨어로 구성되는 것(ex. 메모리와 레지스터)도 얼마든지 가능하므로 HW나 SW에 특별히 한정하지는 않을 것이다.
- [0092] 음성 입력부(700)는 사용자가 발화한 음성 발성을 음성인식엔진(혹은 시스템)에 넣어주는 역할을 담당한다. 이는 다시 말해, 도 5의 제어부(520)와 음성인식엔진을 포함하는 음성인식실행부(540) 사이에 인터페이스 동작을 수행한다고 볼 수 있을 것이다.
- [0093] 아비트레이션부(710)는 먼저 수신된 음성 발성에 대한 발화 길이를 확인할 수 있다 만약 발화 길이가 설정된 시간 범위를 초과한다면, 이를 인식결과 처리부(730)를 통해 제어부(520)에 알릴 수 있다. 물론 이러한 발화 길이의 확인은 시스템 설계자에 따라 선택적으로 실행될 수 있는 것이므로, 이에 특별히 한정하지는 않을 것이다. 또한, 이러한 동작은 제어부(520)에서도 이루어질 수 있기 때문이다. 가령, 제어부(520)에서 이루어진다면, 제

어부(520)는 그 결과에 따라 음성인식실행부(540)를 실행시킬 수 있을 것이다.

- [0094] 이러한 점에서 본다면, 본 발명의 실시예에 따른 음성인식실행부(540)는 도 10에서와 같은 구조를 갖는 것이 바람직하지만, 이와 관련해서는 이후에 충분히 설명할 것이므로, 자세한 설명은 이후의 도 10에서 살펴보기로 한다.
- [0095] 다만, 만약 음성인식실행부(540)가 발화 길이를 확인해야 하는 경우라면, 도 7에서와 같은 구조로 변경되는 것이 바람직하다.
- [0096] 이러한 관점에서 가령, 아비트레이션부(710)가 수신된 음성 발성에 대하여 자체적으로 처리하기로 결정하였다면, 아비트레이션부(710)는 수신된 음성 발성을 복수의 음성인식기(720)로 동시에 입력할 수 있다. 이의 경우에는, 엄격히 말해서, 앞서 언급한 바 있는 "병렬 처리"와 정확히 부합하지 않을 수 있지만, 복수의 음성인식기가 하나의 아비트레이션부(710)에 연결되어, 동시에 음성 발성을 입력받을 수 있다는 점에서, 통상의 "분산 처리"와는 분명한 차이가 있을 것이다. 가령, "분산 처리"는 제어부와 제어부의 동작에 해당되기 때문이다.
- [0097] 아비트레이션부(710)는 복수의 음성인식(720)에서 출력되는 인식 결과로서 인식 텍스트와 유사도 점수, 응답 시간 등을 이용하여 어느 음성 인식기의 인식 결과를, 즉 인식 텍스트를 사용자가 발화한 음성 발성에 부합한 최적의 인식 결과로 사용할지를 결정한다. 다시 말해, 아비트레이션부(710)는 응답 속도가 빠른 인식 결과에 대하여 우선적으로 유사도 점수를 확인하되, 유사도 점수가 기준에 못미치는 경우에는 차 순위의 응답 속도를 보이는 인식 결과의 유사도 점수를 확인하는 방식으로 최적의 인식 결과를 찾게 된다.
- [0098] 복수의 음성인식기(720)는 음성 인식을 위한, 즉 입력된 음성 발성의 오디오 데이터를 분석하여 텍스트로 변환하고, 이때 유사도와 같은 인식 점수 등의 인식 결과를 출력한다는 점에서 공통의 목적을 갖지만, 각각의 음성 인식기(720-1, ..., 720-n)는 서로 다른 범주의 음성인식을 수행하게 된다. 예를 들어, 어떤 음성인식기는 영상 표시장치(100)의 채널 또는 볼륨 조정과 같은 하드웨어 자원을 제어하기에 필요한 인식 결과를 내어주는 반면, 어떤 음성인식기는 어플의 실행이나 검색 등에 관련된 음성 명령어를 처리하여 인식 결과를 내어 줄 수 있다.
- [0099] 이러한 점에서, 복수의 음성인식기(720)로 사용자의 음성명령어가 동시에 입력되더라도, 인식 결과를 출력하는 응답 속도는 제각각 다르기 마련이다. 다만, 본 발명의 실시예에서는 가장 빠르게 얻는 인식 결과가 가장 정확한 인식 결과라 반드시 판단할 수는 없기 때문에 사용자가 거부감을 느끼지 않을 정도의 응답 시간 내에서, 유사도 점수가 가급적 기준 임계값보다 높은 인식 결과, 더 정확하게는 인식 텍스트를 도출함으로써 정확도를 좀 더 높일 수 있게 된다.
- [0100] 인식결과 처리부(730)는 아비트레이션부(710)에서 제공하는 최적의 인식 결과를 제공받아 도 5의 제어부(520)로 제공할 수 있다.
- [0101] 다시 정리해 보면, 사용자가 발화한 음성 발성은 유, 무선으로 영상표시장치(100)와 연결된 마이크폰(혹은 마이크) 등의 집음 장치로 입력되어 영상표시장치(100) 또는 네트워크 등을 통하여 음성인식장치(120)의 하나 혹은 그 이상의 음성 인식기로 입력이 되며, 음성인식기는 입력된 오디오 데이터를 기반으로 인식 결과를 출력한다. 음성인식기는 위에서와 같은 일련의 과정을 거쳐 인식 결과에 대한 신뢰 정도를 특정한 점수의 형태로 출력한다. <표 1>은 인식 결과의 출력을 예시하여 나타낸 것으로서, 음성인식기는 <표 1>에서와 같이 인식 텍스트와 유사도 점수를 인식 결과로서 출력할 수 있다. 이때 각 인식 결과는 인식 범주(domain)가 서로 상이할 수 있을 것이다.

표 1

No.	Result Text	Confidence Score	Domain
1	Volume up	5300	Control Command
2	Volume down	4200	Control Command
3	Facebook	3200	Application

- [0102]
- [0103] 앞서 언급한 대로 여러 개의 음성인식기 각각에서 인식 과정을 수행하기 위한 시간 즉 응답시간은 다르다. 본 발명의 실시예에서는 이러한 인식 텍스트, 유사도 점수, 응답 시간, 나아가 발화 길이 등을 더 고려함으로써 어

떠한 음성인식기의 인식 결과를 선택할지를 결정할 수 있게 되는 것이다.

- [0104] 도 8은 도 1 및 도 2에 도시된 음성인식장치의 세부 구조를 예시하여 나타낸 블록다이어그램이다.
- [0105] 설명의 편의상 도 1을 도 8과 함께 참조하면, 본 발명의 실시예에 따른 음성인식장치(120)는 통신 인터페이스부(800) 및 음성인식처리부(810)를 포함할 수 있다.
- [0106] 통신 인터페이스부(800)는 음성인식처리부(810)의 제어 하에 영상표시장치(100)와 통신을 수행한다. 이의 과정에서 통신 인터페이스부(800)는 영상표시장치(100)에서 제공하는 사용자의 음성 발성을 수신하여 음성인식처리부(810)로 전달한다. 그리고, 통신 인터페이스부(800)는 음성인식처리부(810)로 제공한 음성 발성에 대한 최적의 인식 결과를 음성인식처리부(810)로부터 제공받아 영상표시장치(100)로 전송한다.
- [0107] 음성인식처리부(810)는 도 3 내지 도 5에 도시된 영상표시장치(100)의 음성인식처리부(310, 410) 및 음성인식실행부(540)를 통해 충분히 설명하였으므로 그 내용들로 대신하며 더 이상의 설명은 생략하도록 한다.
- [0108] 도 9는 도 1 및 도 2에 도시된 음성인식장치의 다른 세부 구조를 예시하여 나타낸 블록다이어그램이다.
- [0109] 설명의 편의상 도 9를 도 1과 함께 참조하면, 본 발명의 다른 실시예에 따른 음성인식장치(120')는 통신 인터페이스부(900), 제어부(910), 음성인식실행부(920) 및 저장부(930)의 일부 또는 전부를 포함하며, 여기서 일부 또는 전부를 포함한다는 것은 앞서의 의미와 동일하다.
- [0110] 도 9의 음성인식장치(120')를 도 8의 음성인식장치(120)와 비교해 보면, 도 8에 도시된 음성인식장치(120)의 음성인식처리부(810)는 도 9에서와 같이 제어부(910)와 음성인식실행부(920)로 분리되어 구성될 수 있으며, 이때 제어부(910)는 도 6에서와 같이 프로세서(600)와 메모리(610)를 포함할 수도 있다. 이와 관련해서도 앞서 도 3 내지 도 6에서 영상표시장치(100, 100', 100'')의 구조를 설명하면서 충분히 설명하였으므로 그 내용들로 대신하고자 한다.
- [0111] 도 10은 도 8 및 도 9에 도시된 음성인식처리부 및 음성인식실행부의 세부 구조를 예시하여 나타낸 블록다이어그램이다.
- [0112] 설명의 편의상 도 10을 도 9와 함께 참조하면, 본 발명의 실시예에 따른 음성인식실행부(920)는 음성 입력부(모듈)(1000), 복수의 음성인식기(1010), 아비트레이션부(모듈)(1020) 및 인식결과 처리부(모듈)(1030)의 일부 또는 전부를 포함하며, 여기서 일부 또는 전부를 포함한다는 의미는 앞서와 동일하다.
- [0113] 음성 입력부(1000)는 수신된 음성 발성의 오디오 데이터를 복수의 음성인식기(1010)로 각각 (동시에) 제공한다. 따라서, 음성 입력부(1000)는 복수의 음성인식기(1010)에 대한 입력측이 된다.
- [0114] 복수의 음성인식기(1010)는 수신된 음성 발성에 대한 각각의 인식 결과를 아비트레이션부(1020)로 제공한다. 복수의 음성인식기(1010)와 관련해서는 앞서 도 7 등을 참조하여 충분히 설명하였으므로 더 이상의 설명은 생략한다.
- [0115] 또한 아비트레이션부(1020)는 복수의 음성인식기(1010)에서 제공된 인식 결과에서 사용자가 발화한 음성 발성에 대한 최적의 인식 결과를 도출한다. 이와 관련해서도 앞서 충분히 설명하였으므로 더 이상의 설명은 생략한다. 다만, 아비트레이션부(1020)는 복수의 음성인식기(1010)의 출력측이 된다.
- [0116] 도 10의 구조는 본 발명의 실시예에 따른 바람직한 구조를 나타낸다. 다시 말해, 본 발명의 실시예에서 언급한 "병렬 처리"의 의미에 좀 더 부합하기 때문이다. 도 10에서 볼 때, 복수의 음성인식기(1010)는 입력측 즉 음성 입력부(1000)와 출력측 즉 아비트레이션부(1020)에 근거해 볼 때, 각각의 음성인식기(1010-1, ..., 1010-N)가 서로 병렬 연결되어 있다. 입력측이 공통 연결되고, 출력측이 공통 연결된 구조를 띠고 있는 것을 확인할 수 있다.
- [0117] 이러한 점을 제외하면, 도 10의 음성 입력부(모듈)(1000), 복수의 음성인식기(1010), 아비트레이션부(모듈)(1020) 및 인식결과 처리부(모듈)(1030)는 도 7의 음성 입력부(모듈)(700), 아비트레이션부(모듈)(710), 복수의 음성인식기(720) 및 결과 처리부(모듈)(730)와 크게 다르지 않으므로 그 내용들로 대신하고자 한다.
- [0118] 한편, 앞서 언급한 바와 같이, 도 1의 영상표시장치(100)는 사용자가 발화한 음성 발성에 대한 발화 길이를 확인하지 않고, 내부에 탑재된 복수의 음성인식기를 사용하여 음성 인식 동작을 수행하는 경우에는 도 10에서와 같은 구조를 가질 수 있으므로, 본 발명의 실시예에서는 도 10의 구조를 음성인식장치(120)에 특별히 한정하지는 않고, 도 1의 영상표시장치(100)에도 적용할 수 있을 것이다.

- [0119] 도 11은 도 1의 시스템에서의 음성인식과정을 예시하여 나타낸 도면이다.
- [0120] 도 11에 도시된 바와 같이, 영상표시장치(100)는 사용자가 발화한 음성 발성을 수신한다(S1100). 이를 위하여, 영상표시장치(100)의 내부에 구비된 마이크로폰을 사용할 수 있으며, 외부에서 영상표시장치(100)에 연결된 마이크 즉 집음장치로부터 수신하는 것도 가능할 수 있다.
- [0121] 이어 영상표시장치(100)는 수신된 음성 발성을 음성인식장치(120)로 전송한다(S1110). 도 11의 경우에는 영상표시장치(100)가 내부에 음성인식기를 구비하지 않는 경우에 해당될 수 있기 때문에, S1110 단계를 수행하는 것이 바람직하다.
- [0122] 한편, 음성인식장치(120)는 음성 발성이 수신되면, 내부에 구비된 복수의 음성인식기로 병렬 처리하여 얻은 인식 결과가 기설정된 조건을 만족할 때, 최적의 인식 결과로 확정한다(S1120, S1130). 이에 대하여는 앞서 충분히 설명하였다.
- [0123] 이후 음성인식장치(120)는 최적의 인식 결과를 영상표시장치(100)로 제공하게 된다(S1140).
- [0124] 그러면, 영상표시장치(100)는 수신한 인식 결과에 따른 동작을 수행한다(S1150). 여기서, "인식 결과에 따른 동작을 수행한다"라는 것은 볼륨 조정이나 채널 변경, 어플 실행 등과 같은 동작을 의미한다.
- [0125] 동작 수행과 관련해 좀 더 살펴보면, 가령 영상표시장치(100)는 음성인식장치(120)로부터 인식 결과로서 인식 텍스트를 수신한다. 그러면, 영상표시장치(100)는 수신된 인식 텍스트와 일치하는 텍스트 즉 기설정된 동작 정보가 있는지를 검색할 수 있다. 만약 일치하는 텍스트가 검색되면, 영상표시장치(100)는 검색된 텍스트에 매칭된 이진(binary) 정보에 근거하여 영상표시장치(100)를 동작시키게 된다. 여기서, 이진 정보는 영상표시장치(100)가 인식 가능한 기계어에 해당된다.
- [0126] 도 12는 도 1의 시스템에서의 다른 음성인식과정을 예시하여 나타낸 도면이다.
- [0127] 도 12에 도시된 바와 같이, 영상표시장치(100)는 내부에 음성인식동작을 수행하기 위한 음성인식기를 포함하는 경우, 수신한 음성 발성을 어디에서 처리할지를 먼저 판단할 수 있다(S1200). 이러한 판단 동작은 음성인식엔진에서 판단하는 것이 가능할 수 있지만, 별도의 프로그램을 통해 판단될 수 있는 등 다양하게 변형될 수 있기 때문에 음성인식엔진이 판단하는 것에 특별히 한정하지는 않을 것이다.
- [0128] 영상표시장치(100)는 먼저 음성 발성의 발화 길이를 확인할 수 있다. 예를 들어, 설정된 시간 길이는 1초이지만, 수신된 음성 발성의 발화 길이가 3초로 확인되면, 수신된 음성 발성을 음성인식장치(120)로 전송할 수 있다(S1210).
- [0129] 물론 이의 과정에서, 영상표시장치(100)는 발화 길이가 1초를 초과하지 않는 경우라 하더라도, 내부 자원에 부하가 발생하는 경우에는 음성인식장치(120)로 수신된 음성 발성을 전송할 수 있다.
- [0130] 또한 영상표시장치(100)는 수신된 음성 발성을 음성인식장치(120)로 전송하려는 시점에서, 네트워크의 상태를 판단한 경우, 네트워크가 불안정한 상태일 때는 사용자에게 해당 처리가 수월하지 않음을 알릴 수 있을 것이다.
- [0131] 이러한 점을 제외하면, 도 12의 S1230 내지 S1260 단계는, 도 11의 S1120 내지 S1150 단계와 크게 다르지 않으므로 그 내용들로 대신하고자 한다.
- [0132] 도 13은 도 2의 시스템에서의 음성인식과정을 예시하여 나타낸 도면이다.
- [0133] 도 13의 경우에는, 영상표시장치(200)가 내부에 음성인식엔진을 탑재하느냐에 상관없이 수신된 음성 발성을 복수의 음성인식장치(220-1, 220-2)로 전송하는 경우를 상정한 것이다.
- [0134] 영상표시장치(200)는 수신된 음성 발성을 복수의 음성인식장치(220-1, 220-2)로 동시에 전송할 수 있다(S1310). 본 발명의 실시예에 따라 음성인식장치(220-1)가 메인 장치로서 동작하는 것이 바람직하다. 여기서, 메인 장치란 영상표시장치(200)가, 전송한 음성 발성에 대한 최적의 인식 결과를 제공받는 장치로 정의될 수 있다.
- [0135] 이에 근거해 볼 때, 도 13의 음성인식장치 1(220-1)은 도 11의 S1120 내지 S1150 단계를 수행할 수 있다. 다만, 도 13의 음성인식장치 1(220-1)은 후보군에 해당되는 인식결과가 복수 개 존재할 때, 음성인식장치 2(220-2)에서 제공하는 인식 결과를 참조하여 최적의 인식 결과를 도출할 수도 있다. 가령 하나의 음성인식기는 복수 개의 인식 결과를 내어 줄 수 있는데, 이러한 인식 결과들의 유사도 점수는 서로 유사할 수도 있다. 따라서, 음성인식장치 1(220-1)은 유사도 점수가 유사하여 최적의 인식 결과를 도출하기에 어려움이 있다고 판단될 때, 음성인식장치 2(220-2)에서 제공하는 인식 결과를 참조하여 최종 판단을 내릴 수 있게 된다.

- [0136] 이러한 점을 제외하면, 도 13에서의 동작 과정은 도 11에서와 크게 다르지 않으므로 그 내용들로 대신하고자 한다.
- [0137] 도 14는 본 발명의 실시예에 따른 영상표시장치의 구동 과정을 나타내는 흐름도이다.
- [0138] 설명의 편의상 도 14를 도 1과 함께 참조하면, 본 발명의 실시예에 따른 영상표시장치(100)는 사용자가 발화한 음성 발성을 취득한다(S1400).
- [0139] 그리고, 취득한 음성 발성을 복수의 음성 인식기로 각각 제공하여, 즉 병렬 처리하여 얻은 인식 결과 중 기설정된 조건을 만족하는 인식 결과를 취득한 음성 발성의 인식 결과로 결정 즉 확정한다(S1410).
- [0140] 이는 물론, 영상표시장치(100)가 여러 정확을 고려하여 내부의 음성인식엔진을 이용하여 음성인식동작을 수행하기로 판단한 경우에 해당된다.
- [0141] 이어 영상표시장치(100)는 결정한 인식 결과에 따른 동작을 수행한다(S1420). 이를 위하여, 영상표시장치(100)는 채널 변경, 음량 조절, 검색 동작, 어플 실행 동작 등을 수행할 수 있을 것이다.
- [0142] 도 15는 본 발명의 제1 실시예에 따른 음성인식장치의 구동 과정을 나타내는 흐름도이다.
- [0143] 설명의 편의상 도 15를 도 1과 함께 참조하면, 본 발명의 실시예에 따른 음성인식장치(120)는 사용자가 발화한 음성 발성을 영상표시장치(100)로부터 수신한다(S1500).
- [0144] 이어, 수신한 음성 발성을 복수의 음성 인식기로 병렬 처리하여 얻은 인식 결과 중 기설정된 조건을 만족하는 인식 결과를 수신한 음성 발성의 인식 결과로 확정한다(S1510).
- [0145] 그리고, 최종적으로 결정된 즉 확정된 인식 결과를 영상표시장치(100)로 전송하게 된다(S1520). 이의 과정에서 만약 음성인식장치(120)가 인식 결과에 매칭되는 검색 결과를 제공해야 하는 경우에는 검색 결과를 제공할 수 있다. 예를 들어, 사용자가 스포츠 스타의 이름을 음성 발성으로 발화한 경우라면, 음성인식장치(120)는 스포츠 스타에 대한 인식 결과를 1차적으로 얻고, 인식 결과를 기반으로 하여 검색을 수행하여 검색 결과를 최종적으로 제공할 수도 있다. 검색 결과로는 스타의 고향, 출신학교 등 다양한 정보가 포함될 수 있을 것이다.
- [0146] 도 16은 본 발명의 제2 실시예에 따른 음성인식장치의 구동 과정을 나타내는 흐름도이다.
- [0147] 구체적인 설명에 앞서, 표기에 대하여 간략하게 살펴보기로 한다. 인식 결과가 나온 순서에 따라 제1 ~ 제n 인식기라고 하면, Result_1_ASR1은 제1 인식기의 1순위 후보결과를, Result_1_ASR2는 제2 인식기의 1순위 인식 후보 결과를, Score_1_ASR1은 Result_1_ASR1의 인식 점수(혹은 유사도 점수)를, Result_i_ASR1은 제1 인식기의 임계값(THD_ASR1)보다 큰 점수를 갖는 여러 인식결과 후보 중 i 순위 인식결과를, DScore_1_2_ASR1은 제1 인식기의 1순위 결과 후보와 2순위 후보 결과와의 점수 차이를, DScore_1_2_ASR2는 제2 인식기의 1순위 결과 후보와 2순위 후보 결과와의 점수 차이를, THD_ASR1은 제1 인식기의 인식 여부를 판단하기 위한 점수의 임계치를, THD_ASR2은 제2 인식기의 인식 여부를 판단하기 위한 점수의 임계치를, THD_diff_ASR1은 제1 인식기의 인식결과 점수의 차이에 대한 임계치를, THD_diff_ASR2는 제2 인식기의 인식결과 점수의 차이에 대한 임계치를 각각 나타낸다. 또한, THD_time은 음성인식 결과를 기다리는 최대 시간을 의미한다. 즉 설정된 시간 범위를 나타내는 임계값이다.
- [0148] 설명의 편의상 도 16을 도 1과 함께 참조하면, 본 발명의 실시예에 따른 음성인식장치(120)는 음성 발성이 입력되면, 제1 인식기 ~ 제n 인식기를 동시에 가동시킨다(S1601).
- [0149] 여러 인식기 중 제일 빨리 응답을 출력한 인식기부터의 인식 결과를 각각 ASR1, ASR2, ..., ASRn 이라고 가정할 때, 음성인식장치(120)는 응답 속도가 빠른 순으로 인식 결과를 취득하게 된다(S1603).
- [0150] 물론 이의 과정에서 음성인식장치(120)는 설정된 시간 범위(THD_time) 내에 응답된 인식 결과가 있는지 판단하고(S1605), 판단 결과 없으면 사용자에게 응답 결과가 없음을 알린다(S1607).
- [0151] 만약 최초의 인식 결과(ASR1)인 1순위 후보의 점수 즉 유사도 점수를 기준 임계값(THD_ASR1)과 비교한다(S1609). 이때, 기준이 되는 임계값은 복수 개일 수 있다. 다시 말해, 최상의 기준값이 있어 이를 초과하면 바로 인식 결과로 반영할 수 있고, 최하의 기준값이 있어, 이는 무조건 인식 결과로 반영하지 않을 수 있다. 그리고 그 중간 레벨의 기준값은 인식 결과로 반영하지 말지를 좀 더 고려하기 위해 필요할 수 있다.
- [0152] 이러한 점에서, 비교 결과, 기준 임계값 가령 최상의 기준값보다 작으면, 음성인식장치(120)는 해당 인식 결과를 버리고, 주어진 시간 범위 내에서 다른 인식 결과를 기다리게 된다(S1611).

- [0153] 이의 과정에서 수신된 인식 결과가 기준 임계값을 초과하는데, 유사도 점수가 유사한 복수의 인식 결과(DScore 1, 2_ASR 1)가 검색되었다면, 음성인식장치(120)는 두 개의 인식 결과 사이에 유사도 점수 차(THD_diff_ASR1)를 비교한다(S1613). 여기서, ASR 1은 제1 인식기를 의미하며, 따라서 제1 인식기에서 복수의 인식 결과가 출력되었다는 것으로 이해해도 좋다.
- [0154] 만약, 유사도 점수 차가 크면, 음성인식장치(120)는 유사도 점수가 높은 인식 결과를 최적의 인식 결과로 사용하게 된다(S1615).
- [0155] 만약, 유사도 점수가 서로 비슷하여 최종적으로 결정을 내리기가 어려우면, 음성인식장치(120)는 차 순위로 수신된 인식기의 인식 결과를 참조해서 최적의 인식 결과를 확정할 수 있다(S1617 ~ S1639).
- [0156] 이에 대하여 좀 더 살펴보면, 음성인식장치(120)는 제2 음성인식기의 인식 결과(ASR2)를 기다린다(S1617).
- [0157] 총 대기시간보다 길거나 같을 경우 최초의 인식 결과(Result_1_ASR1)를 사용하고 종료한다(S1619, S1621).
- [0158] 또한, 제2 음성인식기의 인식 결과(ASR2)의 1순위 후보 점수가 기준 임계값(THD_ASR2)보다 작은 경우 현재 음성에 대해 제2 음성인식기의 인식 결과(ASR2)를 제외하고 다음 인식결과를 보내는 인식기를 제2 음성인식기의 인식결과(ASR2)로 판단하여, S1617 단계로 복귀할 수 있다.
- [0159] 제2 음성인식기의 인식결과(ASR2)의 1순위 후보점수가 기준 임계값(THD_ASR2)보다 크거나 같은 경우(S1623), 제1 음성인식기의 인식결과(Result_i_ASR1)와 제2 음성인식기의 인식결과(Result_1_ASR2)가 같다면 제1 음성인식기의 인식 결과(Result_i_ASR1)를 사용하고 종료한다(S1627, S1629).
- [0160] 제2 음성인식기의 인식결과(ASR2)의 1순위 후보점수가 기준 임계값(THD_ASR2)보다 크거나 같지만, 제1 음성인식기의 인식결과(Result_i_ASR1)와 제2 음성인식기의 인식결과(Result_1_ASR2)가 같지 않다면, 복수의 인식 결과의 유사도 점수(DScore1_2_ASR2)와 유사도 점수 차(THD_diff_ASR2)를 비교하여, 유사도 점수(DScore1_2_ASR2)가 더 크거나 같다면 제2 음성인식기의 인식결과(Result_1_ASR2)를 사용하고 종료한다(S1631, S1633).
- [0161] 제2 음성인식기의 인식 결과(ASR2)의 1순위 후보점수가 기준 임계값(THD_ASR2)보다 크거나 같지만, 제1 음성인식기의 후보 인식 결과(Result_i_ASR1)와 제2 음성인식기의 후보 인식결과(Result_1_ASR2)가 같지 않고, 유사도 점수 차(DScore1_2_ASR2)가 임계값(THD_diff_ASR2)보다 작다면, 제1 음성인식기의 인식결과에 대한 유사도 점수 차(DScore1_2_ASR1)와 제2 음성인식기의 인식결과에 대한 유사도 점수 차(DScore1_2_ASR2)를 비교하여 더 큰 차이를 보이는 인식결과(ASR)의 제1 결과를 사용하게 된다(S1631 ~ S1639).
- [0162] 다시 정리해보면, 음성인식장치(120)는 가장 빠르게 응답한 인식기로부터 복수의 인식결과를 수신하였는데, 둘 사이의 유사도 점수가 비슷하고, 점수 차가 임계값보다 작은 경우가 발생할 수 있다(S1613).
- [0163] 이의 경우에는, 주어진 시간 범위 내에서 다음 인식기에서 출력해주는 인식결과를 기다리게 된다(S1617 ~ S1619).
- [0164] 물론 이때, 주어진 시간 범위 내에 있으면서 다음 인식기에서 내어 준 인식 결과는 기준 임계값보다 커야 한다(S1623). 그래야만 둘 사이에 비교 대상이 될 수 있다.
- [0165] 비교 결과, 인식 결과들이 서로 일치하지 않는 경우가 발생할 수 있다(S1627).
- [0166] 이의 경우, 음성인식장치(120)는 차 순위로 얻은 복수의 인식 결과들에 대한 유사도 점수 차가 설정된 임계값보다 큰지를 판단한다(S1631).
- [0167] 만약 유사도 점수 차를 비교해 본 결과, 설정된 임계값보다 역시 크지 않다면, 음성인식장치(120)는 선 순위의 인식 결과에서 유사도 점수가 큰 인식 결과와 후순위의 인식 결과에서 유사도 점수가 큰 인식 결과 중 어느 것이 유사도 점수가 큰지를 판단하여 최적의 인식 결과를 결정할 수 있다(S1635 ~ S1639).
- [0168] 이는 어디까지나 주어진 시간 범위 내에 있는 최초의 인식결과를 내어준 인식기와 차 순위로 인식결과를 내어준 인식기의 인식 결과들을 이용하여 설명한 것이다. 따라서, 시간 범위 내에만 포함된다면, 음성인식장치(120)는 S1631 단계에서 제3 음성인식기의 인식 결과(ASR3)를 기다릴 수도 있을 것이다.
- [0169] 이에 본 발명의 실시예에서는, 두 개의 인식기로 제공되는 인식결과를 이용하는 것에 특별히 한정하지는 않을 것이다.
- [0170] 한편, 본 발명의 실시 예를 구성하는 모든 구성 요소들이 하나로 결합하거나 결합하여 동작하는 것으로 설명되

었다고 해서, 본 발명이 반드시 이러한 실시 예에 한정되는 것은 아니다. 즉, 본 발명의 목적 범위 안에서라면, 그 모든 구성 요소들이 하나 이상으로 선택적으로 결합하여 동작할 수도 있다. 또한, 그 모든 구성요소들이 각각 하나의 독립적인 하드웨어로 구현될 수 있지만, 각 구성 요소들의 그 일부 또는 전부가 선택적으로 조합되어 하나 또는 복수 개의 하드웨어에서 조합된 일부 또는 전부의 기능을 수행하는 프로그램 모듈을 갖는 컴퓨터 프로그램으로서 구현될 수도 있다. 그 컴퓨터 프로그램을 구성하는 코드들 및 코드 세그먼트들은 본 발명의 기술 분야의 당업자에 의해 용이하게 추론될 수 있을 것이다. 이러한 컴퓨터 프로그램은 컴퓨터가 읽을 수 있는 비일시적 저장매체(non-transitory computer readable media)에 저장되어 컴퓨터에 의하여 읽혀지고 실행됨으로써, 본 발명의 실시 예를 구현할 수 있다.

[0171] 여기서 비일시적 판독 가능 기록매체란, 레지스터, 캐시(cache), 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라, 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 구체적으로, 상술한 프로그램들은 CD, DVD, 하드 디스크, 블루레이 디스크, USB, 메모리 카드, ROM 등과 같은 비일시적 판독가능 기록매체에 저장되어 제공될 수 있다.

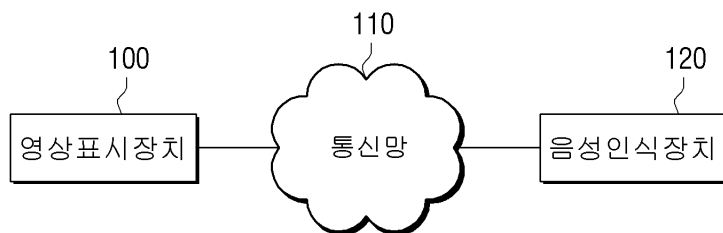
[0172] 이상에서는 본 발명의 바람직한 실시 예에 대하여 도시하고 설명하였지만, 본 발명은 상술한 특징의 실시 예에 한정되지 아니하며, 청구범위에 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 발명의 기술적 사상이나 전망으로부터 개별적으로 이해되어서는 안 될 것이다.

부호의 설명

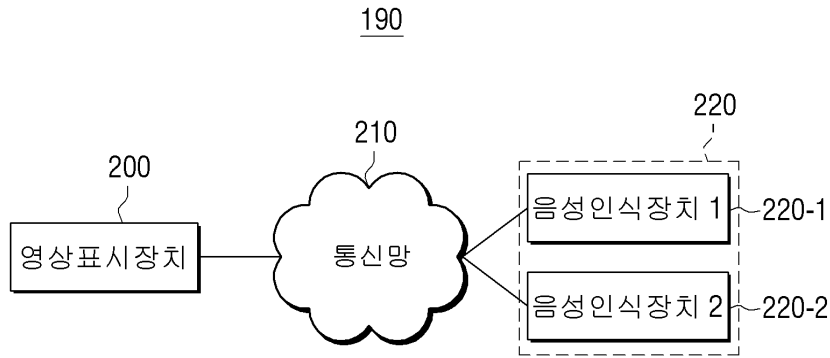
- [0173] 100, 200: 영상표시장치 110, 210: 통신망
- 120, 220: 음성인식장치 300, 510: 음성 취득부
- 310, 410, 810: 음성인식처리부 400, 500, 800, 900: 통신 인터페이스부
- 420, 530: 동작 수행부 430, 550, 930: 저장부
- 540, 920: 음성인식실행부 600: 프로세서
- 610: 메모리 700, 1000: 음성 입력부
- 720, 1010: 음성인식기 710, 1020: 아비트레이션부
- 730, 1030: 인식결과 처리부

도면

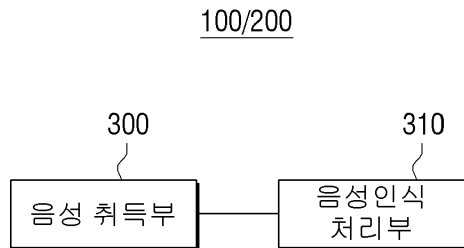
도면1



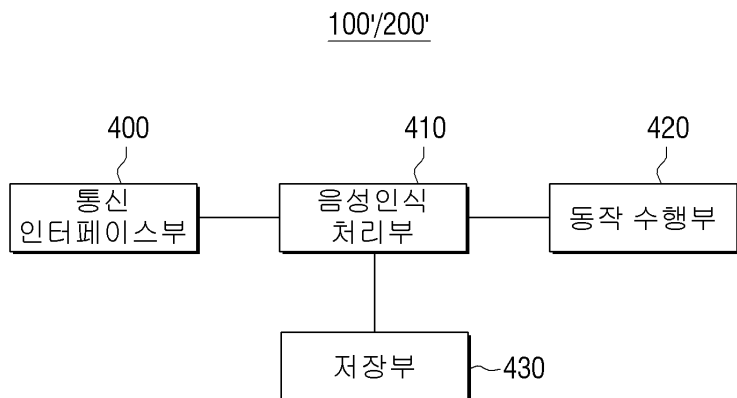
도면2



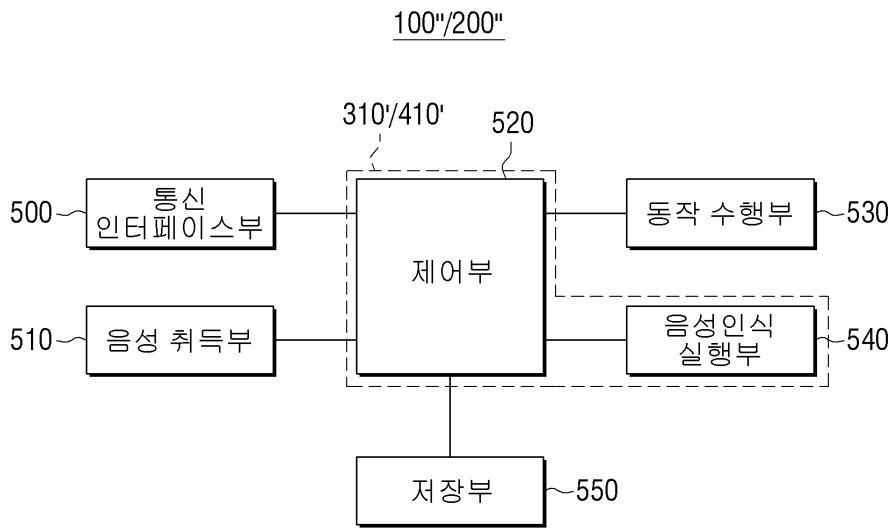
도면3



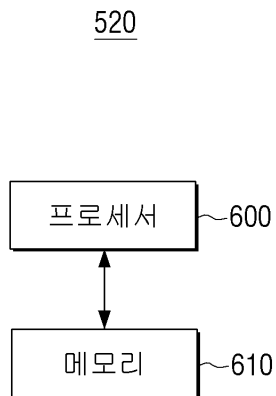
도면4



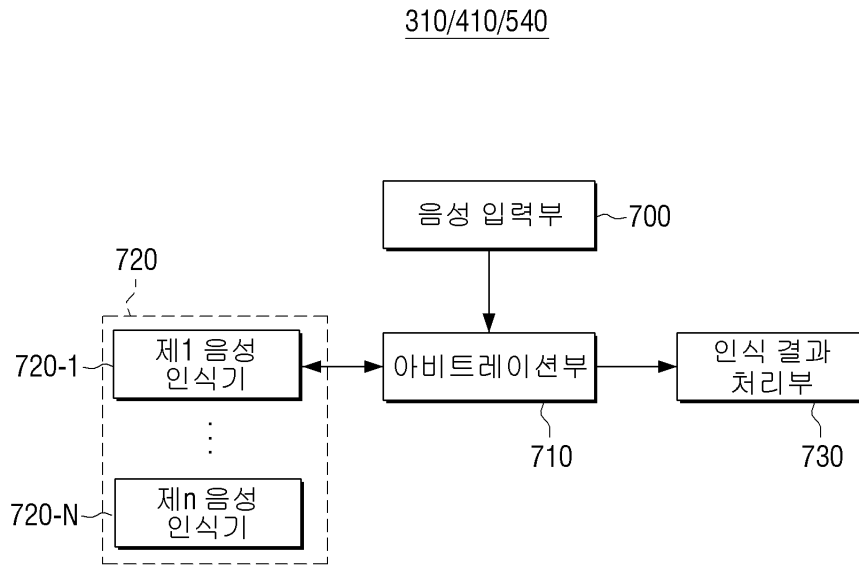
도면5



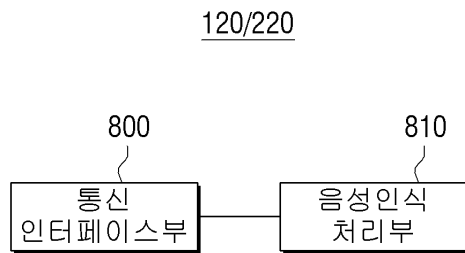
도면6



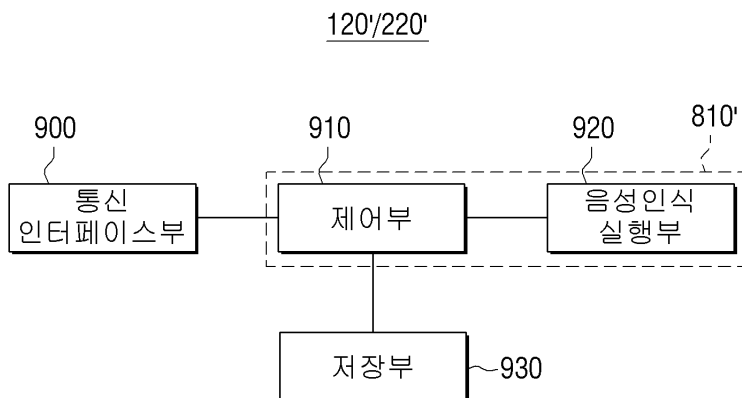
도면7



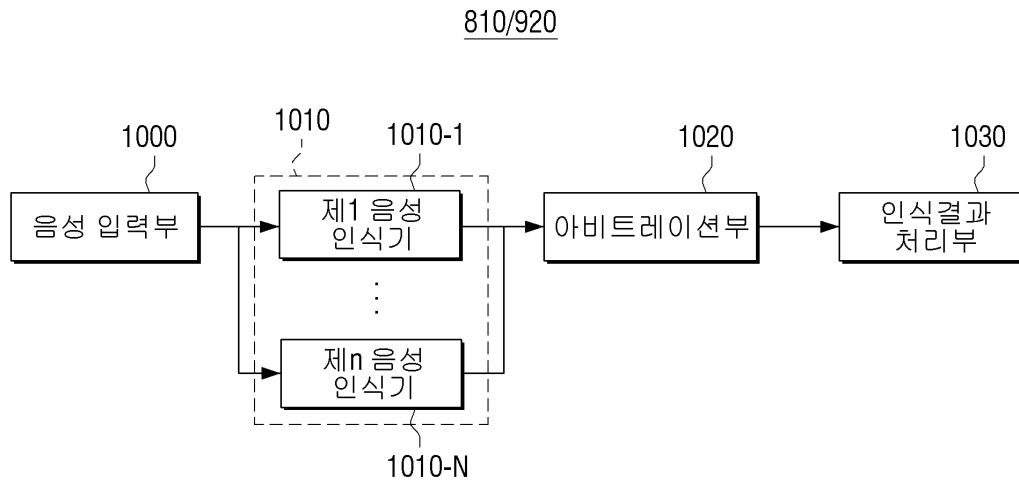
도면8



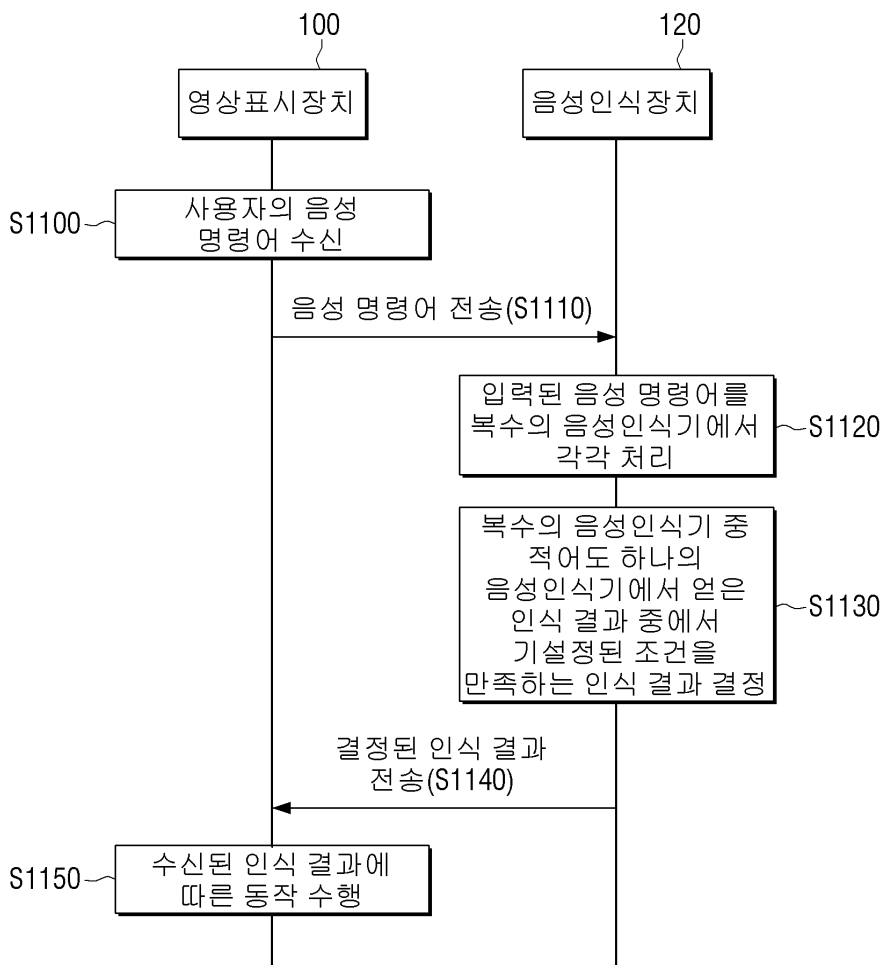
도면9



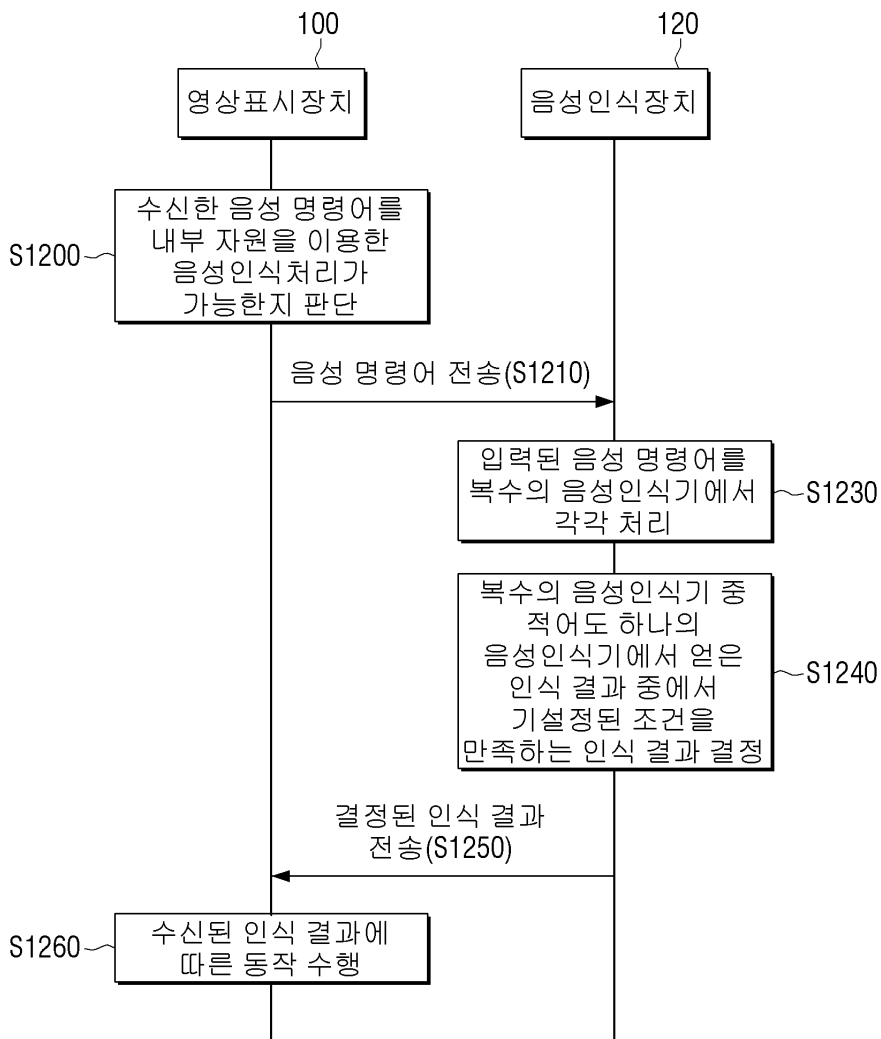
도면10



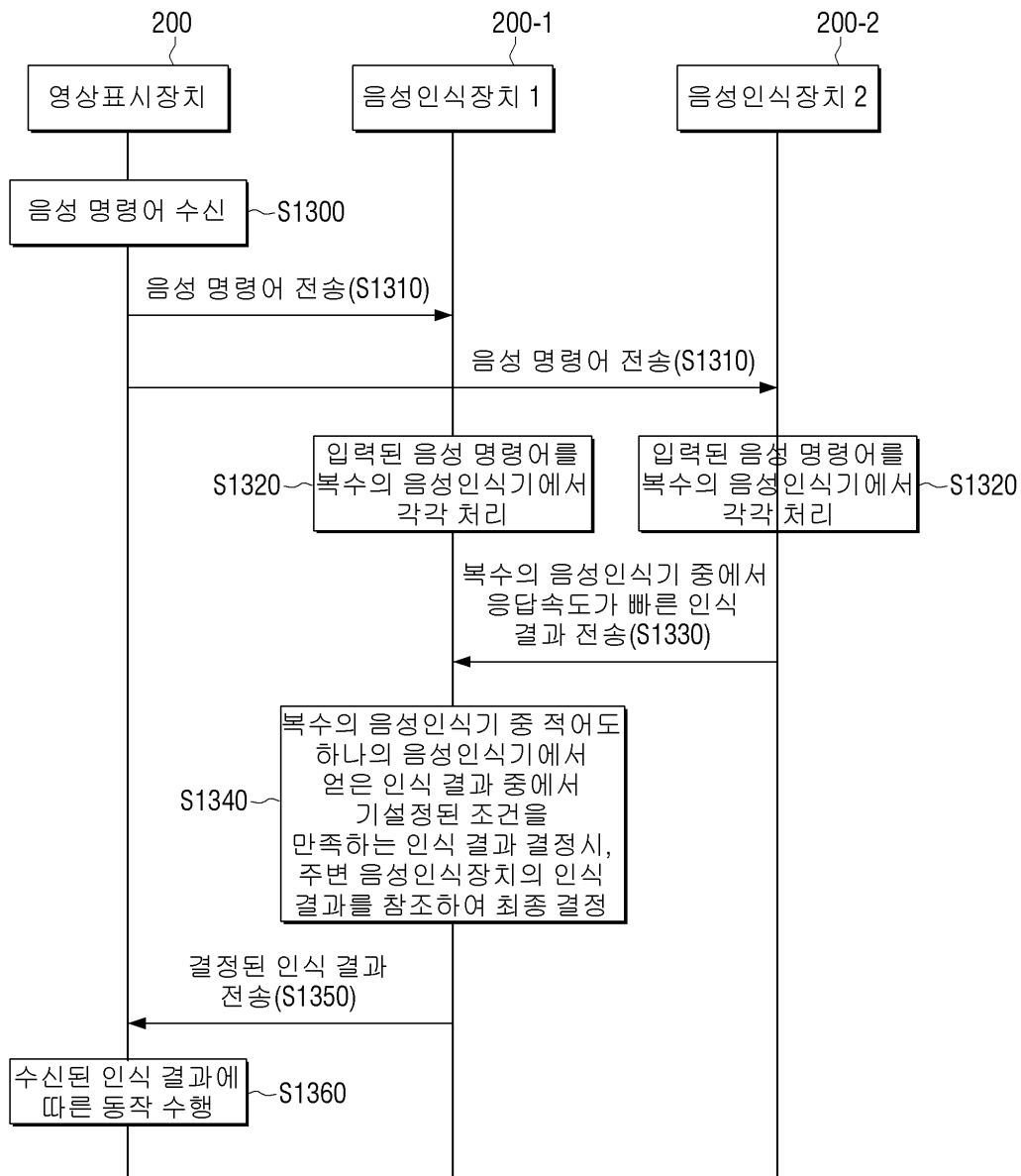
도면11



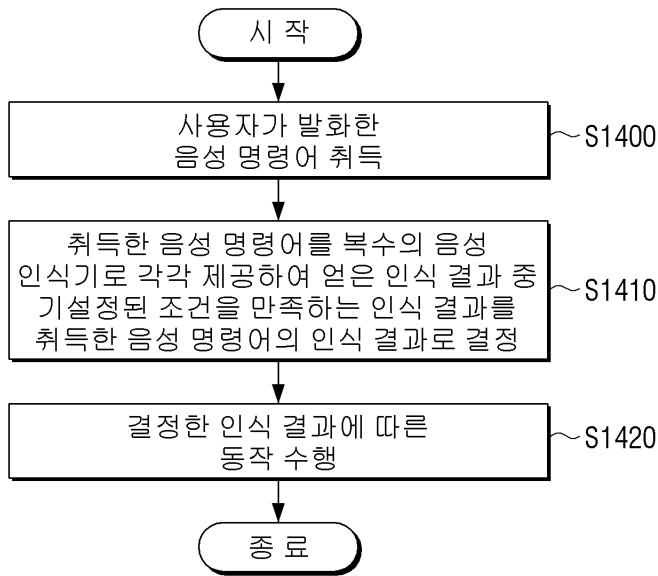
도면12



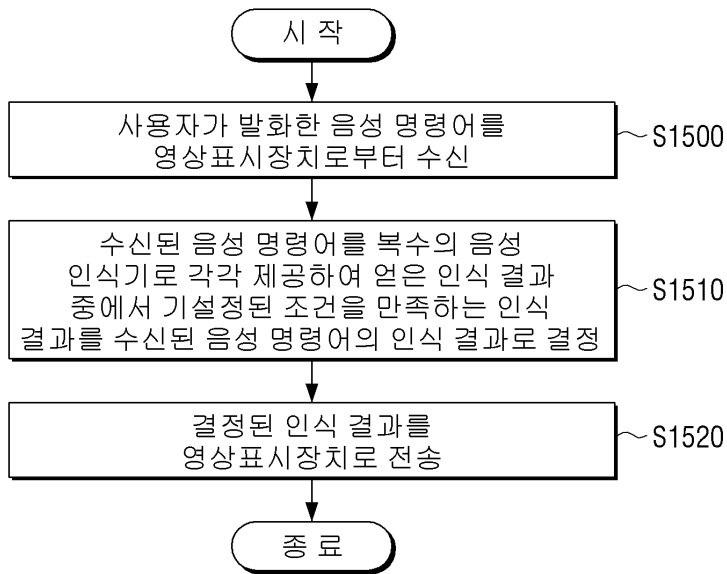
도면13



도면14



도면15



도면16

