



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0041671
(43) 공개일자 2020년04월22일

(51) 국제특허분류(Int. Cl.)

G10L 17/08 (2013.01) G06F 21/32 (2013.01)

G10L 17/02 (2013.01) G10L 17/04 (2013.01)

(52) CPC특허분류

G10L 17/08 (2013.01)

G06F 21/32 (2013.01)

(21) 출원번호 10-2018-0122052

(22) 출원일자 2018년10월12일

심사청구일자 없음

(71) 출원인

삼성전자주식회사

경기도 수원시 영통구 삼성로 129 (매탄동)

(72) 발명자

권재성

경기도 수원시 영통구 도청로 65, 5404동 1503호
(이의동, 자연앤 힐스테이트)

(74) 대리인

정홍식, 김태현

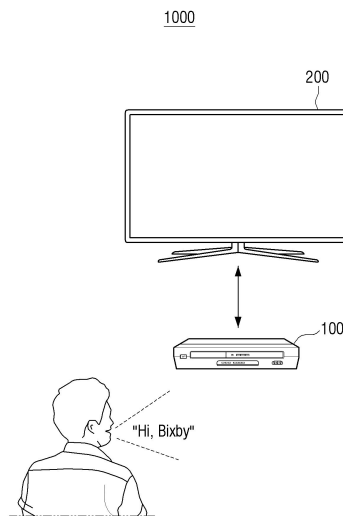
전체 청구항 수 : 총 18 항

(54) 발명의 명칭 전자 장치 및 그 제어 방법

(57) 요약

본 개시에서는 전자 장치 및 이의 제어 방법이 제공된다. 본 개시의 전자장치는 제1 사용자 음성의 음향 특징 및 문맥 정보(context information)를 포함하는 화자 모델이 저장된 메모리 및 화자 모델에 포함된 제1 사용자 음성의 음향 특징과 제2 사용자 음성의 음향 특징간 유사도를, 화자 모델에 포함된 문맥 정보와 제2 사용자 음성의 문맥 정보간 유사도에 따라 변경되는 임계 값과 비교하여, 제2 사용자 음성에 대한 인증을 수행하는 프로세서를 포함한다.

대표도 - 도1a



(52) CPC특허분류

G10L 17/02 (2013.01)

G10L 17/04 (2013.01)

명세서

청구범위

청구항 1

전자 장치에 있어서,

제1 사용자 음성의 음향 특징 및 문맥 정보(context information)를 포함하는 화자 모델을 저장하는 메모리; 및
상기 화자 모델에 포함된 제1 사용자 음성의 음향 특징과 제2 사용자 음성의 음향 특징간 유사도를, 상기 화자 모델에 포함된 문맥 정보와 제2 사용자 음성의 문맥 정보간 유사도에 따라 변경되는 임계 값과 비교하여, 상기 제2 사용자 음성에 대한 인증을 수행하는 프로세서;를 포함하는 전자 장치.

청구항 2

제1항에 있어서,

상기 임계 값은,

상기 제2 사용자 음성의 문맥 정보와 상기 화자 모델에 포함된 문맥 정보 사이의 유사도에 비례하도록 결정되는, 전자 장치.

청구항 3

제1항에 있어서,

상기 프로세서는,

상기 제2 사용자 음성의 문맥 정보와 상기 화자 모델에 포함된 문맥 정보 사이의 유사도에 기초하여 상기 임계 값을 산출하고, 상기 산출된 임계 값을, 제2 사용자 음성의 음향 특징과 상기 화자 모델에 저장된 제1 사용자 음성의 음향 특징 사이의 유사도와 비교하는, 전자 장치.

청구항 4

제3항에 있어서,

상기 프로세서는,

상기 제1 사용자 음성의 문맥 정보와 상기 제2 사용자 음성의 문맥 정보가 동일한 경우 상기 제1 사용자 음성의 음향 특징과 제2 사용자 음성의 음향 특징 사이의 유사도가 가질 수 있는 최대 값에 기초하여 상기 임계 값을 산출하는, 전자 장치.

청구항 5

제1항에 있어서,

상기 프로세서는,

상기 제2 사용자 음성의 음향 특징과 상기 화자 모델에 저장된 제1 사용자 음성의 음향 특징 사이의 유사도가 상기 임계 값 이상인 경우, 상기 제2 사용자 음성에 대한 인증이 성공된 것으로 판단하는, 전자 장치.

청구항 6

제1항에 있어서,

상기 프로세서는,

상기 제1 사용자 음성이 수신되면, 상기 제1 사용자 음성으로부터 상기 제1 사용자 음성의 음향 특징 및 문맥 정보를 획득하고, 상기 획득된 음향 특징 및 문맥 정보를 포함하는 상기 화자 모델을 상기 메모리에 저장하는, 전자 장치.

청구항 7

제6항에 있어서,

상기 프로세서는,

기정의된 복수의 문맥 정보 타입(type) 중 하나의 문맥 정보 타입을 선택하고, 상기 선택된 문맥 정보 타입을 이용하여 상기 제1 사용자 음성의 문맥 정보를 획득하고, 상기 획득된 문맥 정보를 상기 메모리에 저장하는, 전자 장치.

청구항 8

제7항에 있어서,

상기 프로세서는,

상기 제1 사용자 음성의 음향 특징과 복수의 테스트 음성 각각의 음향 특징 사이의 유사도를 판단하고,

상기 복수의 문맥 정보의 타입 별로, 상기 제1 사용자 음성의 문맥 정보와 상기 복수의 테스트 음성 각각의 문맥 정보 사이의 유사도를 판단하며,

상기 복수의 문맥 정보 타입 중, 상기 판단된 문맥 정보 사이의 유사도에 대한 상기 판단된 음향 특징의 유사도의 분포가 기설정된 조건을 만족하는 문맥 정보 타입을 선택하는, 전자 장치.

청구항 9

제7항에 있어서,

상기 프로세서는,

상기 제2 사용자 음성을 분석하고, 상기 선택된 문맥 정보 타입을 이용하여 상기 제2 사용자 음성의 문맥 정보를 획득하고, 상기 제2 사용자 음성의 문맥 정보와 상기 화자 모델에 포함된 문맥 정보 사이의 유사도에 기초하여 상기 임계 값을 결정하는, 전자 장치.

청구항 10

전자 장치의 제어 방법에 있어서,

제1 사용자 음성의 음향 특징 및 문맥 정보(context information)를 포함하는 화자 모델을 저장하는 단계;

상기 화자 모델에 포함된 제1 사용자 음성의 음향 특징과 제2 사용자 음성의 음향 특징간 유사도를, 상기 화자 모델에 포함된 문맥 정보와 제2 사용자 음성의 문맥 정보간 유사도에 따라 변경되는 임계 값과 비교하는 단계; 및

상기 제2 사용자 음성에 대한 인증을 수행하는 단계;를 포함하는 제어 방법.

청구항 11

제10항에 있어서,

상기 임계값은,

상기 제2 사용자 음성의 문맥 정보와 상기 화자 모델에 포함된 문맥 정보 사이의 유사도에 비례하도록 결정되는, 제어 방법.

청구항 12

제10항에 있어서,

상기 비교하는 단계는,

상기 제2 사용자 음성의 문맥 정보와 상기 화자 모델에 포함된 문맥 정보 사이의 유사도에 기초하여 상기 임계 값을 산출하는 단계 및,

상기 산출된 임계 값을, 제2 사용자 음성의 음향 특징과 상기 화자 모델에 저장된 제1 사용자 음성의 음향 특징

사이의 유사도와 비교하는 단계를 포함하는, 제어 방법.

청구항 13

제12항에 있어서,

상기 산출하는 단계는,

상기 제1 사용자 음성의 문맥 정보와 상기 제2 사용자 음성의 문맥 정보가 동일한 경우 상기 제1 사용자 음성의 음향 특징과 제2 사용자 음성의 음향 특징 사이의 유사도가 가질 수 있는 최대 값에 기초하여 상기 임계 값을 산출하는 단계인, 제어 방법.

청구항 14

제10항에 있어서,

상기 인증을 수행하는 단계는,

상기 제2 사용자 음성의 음향 특징과 상기 화자 모델에 저장된 제1 사용자 음성의 음향 특징 사이의 유사도가 상기 임계 값 이상인 경우, 상기 제2 사용자 음성에 대한 인증이 성공된 것으로 판단하는 단계를 포함하는, 제어 방법.

청구항 15

제10항에 있어서,

상기 제1 사용자 음성이 수신되면, 상기 제1 사용자 음성으로부터 상기 제1 사용자 음성의 음향 특징 및 문맥 정보를 획득하는 단계; 및

상기 획득된 음향 특징 및 문맥 정보를 포함하는 상기 화자 모델을 저장하는 단계;를 더 포함하는, 제어 방법.

청구항 16

제15항에 있어서,

상기 획득하는 단계는,

기정의된 복수의 문맥 정보 타입(type) 중 하나의 문맥 정보 타입을 선택하는 단계; 및

상기 선택된 문맥 정보 타입을 이용하여 상기 제1 사용자 음성의 문맥 정보를 획득하는 단계;를 포함하는, 제어 방법.

청구항 17

제16항에 있어서,

상기 선택하는 단계는,

상기 제1 사용자 음성의 음향 특징과 상기 복수의 테스트 음성 각각의 음향 특징 사이의 유사도를 판단하는 단계;

상기 복수의 문맥 정보의 타입 별로, 상기 제1 사용자 음성의 문맥 정보와 복수의 테스트 음성 각각의 문맥 정보 사이의 유사도를 판단하는 단계; 및

상기 복수의 문맥 정보 타입 중, 상기 판단된 문맥 정보 사이의 유사도에 대한 상기 판단된 음향 특징의 유사도의 분포가 기설정된 조건을 만족하는 문맥 정보 타입을 선택하는 단계;를 포함하는, 제어 방법.

청구항 18

제16항에 있어서,

상기 제2 사용자 음성을 분석하고, 상기 선택된 문맥 정보 타입을 이용하여 상기 제2 사용자 음성의 문맥 정보를 획득하는 단계; 및

상기 제2 사용자 음성의 문맥 정보와 상기 화자 모델에 포함된 문맥 정보 사이의 유사도에 기초하여 상기 임계

값을 결정하는 단계;를 더 포함하는, 제어 방법.

발명의 설명

기술 분야

[0001] 본 개시는 전자 장치 및 그 제어 방법에 관한 것으로, 구체적으로는 화자인식을 수행하는 전자 장치 및 그 제어 방법에 관한 것이다.

배경 기술

[0002] 전자 기술의 발전에 따라 무선 통신 기술이 보편화된 가운데, 해킹이나 개인정보유출을 방지하기 위한 보안 기술이 발전하고 있다. 최근에는 지문, 홍채, 얼굴, 목소리 등과 같이 사용자의 신체 일부를 이용하여 사용자를 인증하는 생체인식 기술이 등장하고 있다.

[0003] 최근 음성 인식 기술 및 인공 지능 기술의 발달로 사용자의 음성을 명령으로 입력 받아 기능을 수행하는 전자 장치가 증가함에, 음성을 통하여 인증된 사용자에게만 기능을 수행하도록 하는 전자 장치에 관한 관심 또한 증가하고 있다.

[0004] 이와 관련하여, 화자인식은 발화한 사용자의 음성에 포함된 특징을 이용하여 사용자를 인증하는 기술을 의미한다. 화자인식은 사용자가 발화하여 특정 음성을 등록하는 등록단계과 등록 이후 사용자가 발화한 음성을 등록된 특정 음성과 비교하여 화자를 인증하는 인증단계를 포함한다.

[0005] 화자인식은 등록단계에서는 발화(등록발화)된 음성의 음향특징으로 화자 모델을 생성하고, 인증단계에서는 화자 모델에 포함된 등록단계에서 발화한 음성의 음향특징과 인증단계에서 발화(검증발화)한 음성의 음향특징을 비교하여 비교 결과가 특정 임계값 이상인 경우 등록발화의 화자와 검증발화의 화자가 동일하다고 판단한다.

[0006] 화자인식은 인증단계를 기준으로 크게 등록단계에서 등록된 음성과 동일한 음성을 발화하여 화자를 인증하는 문장중속 방식과 등록된 음성에 상관없이 자유롭게 음성을 발화하여 화자를 인증하는 문장독립 방식으로 나뉠 수 있다.

[0007] 문장중속 화자인식 방식의 경우, 전자 장치가 등록단계와 인증단계에서 동일한 발화어를 이용하기 때문에 전자 장치의 화자인식의 정확도가 높은 반면, 일정한 발화어만을 이용한다는 점에서 사용성 측면에서 활용도가 낮다.

[0008] 한편, 문장독립 화자인식 방식의 경우, 전자 장치가 사용자가 자유롭게 발화한 음성을 인식한다는 점에서 활용도가 높으나, 등록발화와 상이한 검증발화를 하는 경우가 대부분이어서, 문장중속 방식에 비하여 화자인식의 성능이 낮은 단점이 있다.

발명의 내용

해결하려는 과제

[0009] 본 개시는 상술한 문제점에서 도출된 것으로, 본 개시의 목적은 등록단계에서 등록발화와 인증단계에서의 검증발화의 유사도를 사용하여 문장독립 화자 인식을 수행하는 전자 장치 및 그 제어 방법을 제공함에 있다.

과제의 해결 수단

[0010] 본 개시의 일 실시예에 따른 전자 장치는, 제1 사용자 음성의 음향 특징 및 문맥 정보(context information)를 포함하는 화자 모델을 저장하는 메모리; 및 상기 화자 모델에 포함된 제1 사용자 음성의 음향 특징과 제2 사용자 음성의 음향 특징간 유사도를, 상기 화자 모델에 포함된 문맥 정보와 제2 사용자 음성의 문맥 정보간 유사도에 따라 변경되는 임계 값과 비교하여, 상기 제2 사용자 음성에 대한 인증을 수행하는 프로세서;를 포함한다.

[0011] 여기에서, 상기 임계 값은, 상기 제2 사용자 음성의 문맥 정보와 상기 화자 모델에 포함된 문맥 정보 사이의 유사도에 비례하도록 결정될 수 있다.

[0012] 그리고, 상기 프로세서는, 상기 제2 사용자 음성의 문맥 정보와 상기 화자 모델에 포함된 문맥 정보 사이의 유사도에 기초하여 상기 임계 값을 산출하고, 상기 산출된 임계 값을, 제2 사용자 음성의 음향 특징과 상기 화자

모델에 저장된 제1 사용자 음성의 음향 특징 사이의 유사도와 비교할 수 있다.

- [0013] 또한, 상기 프로세서는, 상기 제1 사용자 음성의 문맥 정보와 상기 제2 사용자 음성의 문맥 정보가 동일한 경우 상기 제1 사용자 음성의 음향 특징과 제2 사용자 음성의 음향 특징 사이의 유사도가 가질 수 있는 최대 값에 기초하여 상기 임계 값을 산출할 수 있다.
- [0014] 그리고, 상기 프로세서는, 상기 제2 사용자 음성의 음향 특징과 상기 화자 모델에 저장된 제1 사용자 음성의 음향 특징 사이의 유사도가 상기 임계 값 이상인 경우, 상기 제2 사용자 음성에 대한 인증이 성공된 것으로 판단할 수 있다.
- [0015] 또한, 상기 프로세서는, 상기 제1 사용자 음성이 수신되면, 상기 제1 사용자 음성으로부터 상기 제1 사용자 음성의 음향 특징 및 문맥 정보를 획득하고, 상기 획득된 음향 특징 및 문맥 정보를 포함하는 상기 화자 모델을 상기 메모리에 저장할 수 있다.
- [0016] 그리고, 상기 프로세서는, 기정의된 복수의 문맥 정보 타입(type) 중 하나의 문맥 정보 타입을 선택하고, 상기 선택된 문맥 정보 타입을 이용하여 상기 제1 사용자 음성의 문맥 정보를 획득하고, 상기 획득된 문맥 정보를 상기 메모리에 저장할 수 있다.
- [0017] 또한, 상기 프로세서는, 상기 제1 사용자 음성의 음향 특징과 복수의 테스트 음성 각각의 음향 특징 사이의 유사도를 판단하고, 상기 복수의 문맥 정보의 타입 별로, 상기 제1 사용자 음성의 문맥 정보와 상기 복수의 테스트 음성 각각의 문맥 정보 사이의 유사도를 판단하며, 상기 복수의 문맥 정보 타입 중, 상기 판단된 문맥 정보 사이의 유사도에 대한 상기 판단된 음향 특징의 유사도의 분포가 기설정된 조건을 만족하는 문맥 정보 타입을 선택할 수 있다.
- [0018] 그리고, 상기 프로세서는, 상기 제2 사용자 음성을 분석하고, 상기 선택된 문맥 정보 타입을 이용하여 상기 제2 사용자 음성의 문맥 정보를 획득하고, 상기 제2 사용자 음성의 문맥 정보와 상기 화자 모델에 포함된 문맥 정보 사이의 유사도에 기초하여 상기 임계 값을 결정할 수 있다.
- [0019] 한편, 본 개시의 일 실시예에 따른, 전자 장치의 제어 방법은, 제1 사용자 음성의 음향 특징 및 문맥 정보(context information)를 포함하는 화자 모델을 저장하는 단계; 상기 화자 모델에 포함된 제1 사용자 음성의 음향 특징과 제2 사용자 음성의 음향 특징간 유사도를, 상기 화자 모델에 포함된 문맥 정보와 제2 사용자 음성의 문맥 정보간 유사도에 따라 변경되는 임계 값과 비교하는 단계; 및 상기 제2 사용자 음성에 대한 인증을 수행하는 단계;를 포함한다.
- [0020] 이때, 상기 임계값은, 상기 제2 사용자 음성의 문맥 정보와 상기 화자 모델에 포함된 문맥 정보 사이의 유사도에 비례하도록 결정될 수 있다.
- [0021] 또한, 상기 비교하는 단계는, 상기 제2 사용자 음성의 문맥 정보와 상기 화자 모델에 포함된 문맥 정보 사이의 유사도에 기초하여 상기 임계 값을 산출하는 단계 및, 상기 산출된 임계 값을, 제2 사용자 음성의 음향 특징과 상기 화자 모델에 저장된 제1 사용자 음성의 음향 특징 사이의 유사도와 비교하는 단계를 포함할 수 있다.
- [0022] 그리고, 상기 산출하는 단계는, 상기 제1 사용자 음성의 문맥 정보와 상기 제2 사용자 음성의 문맥 정보가 동일한 경우 상기 제1 사용자 음성의 음향 특징과 제2 사용자 음성의 음향 특징 사이의 유사도가 가질 수 있는 최대 값에 기초하여 상기 임계 값을 산출하는 단계일 수 있다.
- [0023] 또한, 상기 인증을 수행하는 단계는, 상기 제2 사용자 음성의 음향 특징과 상기 화자 모델에 저장된 제1 사용자 음성의 음향 특징 사이의 유사도가 상기 임계 값 이상인 경우, 상기 제2 사용자 음성에 대한 인증이 성공된 것으로 판단하는 단계를 포함할 수 있다.
- [0024] 그리고, 상기 제1 사용자 음성이 수신되면, 상기 제1 사용자 음성으로부터 상기 제1 사용자 음성의 음향 특징 및 문맥 정보를 획득하는 단계; 및 상기 획득된 음향 특징 및 문맥 정보를 포함하는 상기 화자 모델을 저장하는 단계;를 더 포함할 수 있다.
- [0025] 또한, 상기 획득하는 단계는, 기정의된 복수의 문맥 정보 타입(type) 중 하나의 문맥 정보 타입을 선택하는 단계; 및 상기 선택된 문맥 정보 타입을 이용하여 상기 제1 사용자 음성의 문맥 정보를 획득하는 단계;를 포함할 수 있다.
- [0026] 그리고, 상기 선택하는 단계는, 상기 제1 사용자 음성의 음향 특징과 상기 복수의 테스트 음성 각각의 음향 특징 사이의 유사도를 판단하는 단계; , 상기 복수의 문맥 정보의 타입 별로, 상기 제1 사용자 음성의 문맥 정보와

복수의 테스트 음성 각각의 문맥 정보 사이의 유사도를 판단하는 단계; 및 상기 복수의 문맥 정보 타입 중, 상기 판단된 문맥 정보 사이의 유사도에 대한 상기 판단된 음향 특징의 유사도의 분포가 기설정된 조건을 만족하는 문맥 정보 타입을 선택하는 단계;를 포함할 수 있다.

- [0027] 마지막으로, 상기 제2 사용자 음성을 분석하고, 상기 선택된 문맥 정보 타입을 이용하여 상기 제2 사용자 음성의 문맥 정보를 획득하는 단계; 상기 제2 사용자 음성의 문맥 정보와 상기 화자 모델에 포함된 문맥 정보 사이의 유사도에 기초하여 상기 임계 값을 결정하는 단계;를 더 포함할 수 있다.

발명의 효과

- [0028] 상술한 바와 같은 실시예에 따라, 전자 장치는 화자 모델에 저장된 사용자 음성과 입력된 사용자 음성의 유사도에 따라 임계값을 가변적으로 변경할 수 있으므로, 화자 인식 시스템의 사용성이 증가될 수 있다.

도면의 간단한 설명

- [0029] 도 1은 본 개시의 다양한 실시예에 따른, 대화 시스템을 도시한 도면,
 도 2는 본 개시의 일 실시예에 따른 전자 장치의 구성을 설명하기 위한 블록도,
 도 3은 본 개시의 일 실시예에 따른 전자 장치의 구성을 상세히 도시한 블록도,
 도 4 및 5는 본 개시의 일 실시예에 따른 전자 장치를 설명하기 위한 도면, 및
 도 6은 본 개시의 일 실시예에 따른 전자 장치의 제어 방법을 설명하기 위한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0030] 본 명세서에서 사용되는 용어에 대해 간략히 설명하고, 본 개시에 대해 구체적으로 설명하기로 한다.
- [0031] 본 개시의 실시 예에서 사용되는 용어는 본 개시에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 개시의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 개시에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 개시의 전반에 걸친 내용을 토대로 정의되어야 한다.
- [0032] 본 개시의 실시 예들은 다양한 변환을 가할 수 있고 여러 가지 실시 예를 가질 수 있는 바, 특정 실시 예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나 이는 특정한 실시 형태에 대해 범위를 한정하려는 것이 아니며, 개시된 사상 및 기술 범위에 포함되는 모든 변환, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 실시 예들을 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 요지를 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.
- [0033] 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 구성요소들은 용어들에 의해 한정되어서는 안 된다. 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다.
- [0034] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "구성되다" 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 않는 것으로 이해되어야 한다.
- [0035] 본 개시에서 "모듈" 혹은 "부"는 적어도 하나의 기능이나 동작을 수행하며, 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다. 또한, 복수의 "모듈" 혹은 복수의 "부"는 특정한 하드웨어로 구현될 필요가 있는 "모듈" 혹은 "부"를 제외하고는 적어도 하나의 모듈로 일체화되어 적어도 하나의 프로세서(미도시)로 구현될 수 있다.
- [0036] 아래에서는 첨부한 도면을 참고하여 본 개시의 실시 예에 대하여 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 개시는 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시 예에 한정되지 않는다. 그리고 도면에서 본 개시를 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

- [0037] 이하에서는 도면을 참조하면 본 개시의 다양한 실시 예들에 대하여 구체적으로 설명하도록 한다.
- [0038] 도 1은 본 개시의 다양한 실시예에 따른, 대화 시스템을 도시한 도면이다.
- [0039] 도 1(a)에 도시된 바와 같이, 대화 시스템(1000)은 전자 장치(100) 및 디스플레이 장치(200)를 포함한다.
- [0040] 전자 장치(100)는 사용자 음성을 수신하고, 수신된 사용자 음성에 대하여 음성 인식을 수행한다. 구체적으로, 전자 장치(100)는 수신된 사용자 음성에 대한 전처리를 수행하고, 사용자 음성에 대한 음성 인식을 수행하여 사용자 음성을 텍스트로 변환하고(Speech to Text, STT), 음성 인식 결과에 기초하여 사용자 음성의 의도 및 엔티티를 파악할 수 있다.
- [0041] 이를 위해, 전자 장치(100)는 자동 음성 인식(Automatic Speech Recognition, ASR) 모듈, 자연어 이해(Natural Language Understanding, NLU) 모듈, 대화 관리(Dialogue Management, DM) 모듈, 자연어 생성(Natural Language Generation, NLG) 모듈 등을 포함할 수 있다.
- [0042] 전자 장치(100)가 수신된 사용자 음성에 대하여 음성 인식을 수행하고 사용자 음성을 텍스트로 변환할 수도 있으나, 전자 장치(100)는 수신된 사용자 음성에 대한 전처리를 수행하여 사용자 음성을 음성 신호로 변환하고, 사용자 음성 신호를 서버(미도시)에 전송할 수도 있다.
- [0043] 이때, 서버(미도시)는 수신된 사용자 음성 신호에 대하여 음성 인식을 수행하여 사용자 음성을 텍스트로 변환하는 STT 기능을 수행하고, 사용자 음성에 대한 음성 인식 결과를 전자 장치(100)에 전송할 수 있다.
- [0044] 전자 장치(100)는 자연어 이해 결과에 기초하여 사용자 음성에 대한 응답을 위한 정보를 가령, 웹 서버 등으로부터 획득하고, 획득된 정보에 기초하여 사용자 음성에 대한 응답 정보를 생성할 수 있다.
- [0045] 전자 장치(100)는 생성된 응답 정보를 디스플레이 장치(200)에 전송할 수 있다.
- [0046] 디스플레이 장치(200)는 수신된 응답 정보에 기초하여 사용자 음성에 대한 응답을 제공할 수 있다.
- [0047] 예를 들어, 디스플레이 장치(200)는 TTS(Text to Speech)를 통해, 수신된 응답 정보에 포함된 텍스트를 음성으로 변환하여 디스플레이 장치(200)의 스피커를 통해 출력하거나, 또는 해당 텍스트를 포함하는 유저 인터페이스(User interface)를 디스플레이 장치(200)의 디스플레이를 통해 표시할 수 있다.
- [0048] 한편, 전자 장치(100)는 대화 시스템을 동작하기 위한 인공지능 에이전트를 저장할 수 있다. 구체적으로, 전자 장치(100)는 사용자 음성에 대한 응답으로 자연어를 생성하기 위하여 인공지능 에이전트를 이용할 수 있다. 인공지능 에이전트는 AI(Artificial Intelligence) 기반의 서비스(예를 들어, 음성 인식 서비스, 비서 서비스, 번역 서비스, 검색 서비스 등)를 제공하기 위한 전용 프로그램으로서, 기존의 범용 프로세서(예를 들어, CPU) 또는 별도의 AI 전용 프로세서(예를 들어, GPU 등)에 의해 실행될 수 있다. 특히, 인공지능 에이전트는 다양한 모듈을 제어할 수 있다.
- [0049] 이 경우, 전자 장치(100)는 자연어로 생성된 사용자 음성에 대한 응답 정보를 디스플레이 장치(200)로 전송할 수 있으며, 디스플레이 장치(200)는 응답 정보를 자연어로 출력하여 사용자에게 제공할 수 있게 된다.
- [0050] 이와 같이, 대화 시스템(1000)은 사용자 음성에 대한 응답을 제공할 수 있게 되어, 사용자는 디스플레이 장치(200)와 대화를 수행할 수 있게 된다.
- [0051] 도 1(a)에서는 전자 장치(100)와 디스플레이 장치(200)를 구분하여 도시하였으나, 도 1(b)와 같이, 전자 장치(100)와 디스플레이 장치(200)가 하나의 장치 내에서 구현될 수도 있다.
- [0052] 이 경우, 전자 장치(100)는 사용자 음성을 수신하고, 수신된 사용자 음성에 대하여 음성 인식을 수행하여 사용자 음성에 대한 응답 정보를 제공할 수 있다.
- [0053] 이를 위하여 전자 장치(100)는 사용자 음성에 대한 응답 정보를 제공하기 위하여 디스플레이, 스피커 등을 포함할 수 있다.
- [0054] 도 1(b)와 같이 전자 장치(100)와 디스플레이 장치(200)가 하나의 장치 내에서 구현된 경우, 전자 장치(100)는 서버(미도시)와 연결되어 수신된 사용자 음성을 음성 신호로 변환한 후, 음성 신호를 서버(미도시)에 전송할 수 있다. 이때, 서버(미도시)는 사용자 음성 신호를 텍스트로 변환하는 STT 기능을 수행하고 음성 인식 결과를 전자 장치(100)에 전송할 수 있다.
- [0055] 한편, 도 1 (c)와 같이, 전자 장치(100)는 서버로 구현될 수도 있다.

- [0056] 이 경우, 외부 전자 장치(300)는 사용자가 발화한 음성을 수신하고, 수신한 사용자의 음성을 음성 신호로 변환하고, 이를 전자 장치(100)에 전송할 수 있다.
- [0057] 전자 장치(100)는 외부 전자 장치(300)로부터 사용자 음성을 수신하면, 수신된 사용자 음성에 대하여 음성 인식을 수행할 수 있다. 구체적으로, 전자 장치(100)는 사용자 음성 신호를 텍스트로 변환하고, 사용자 음성에 대한 자연어 이해 처리를 수행하여, 사용자 음성의 의도 및 엔티티를 파악할 수 있다.
- [0058] 전자 장치(100)는 자연어 이해 결과에 기초하여 사용자 음성에 대한 응답을 위한 정보를 웹 서버 등과 같은 또 다른 서버로부터 획득하고, 획득한 정보에 기초하여 사용자 음성에 대한 응답 정보를 생성할 수 있다.
- [0059] 전자 장치(100)는 생성된 응답 정보를 외부 전자 장치(300)에 전송할 수 있다.
- [0060] 외부 전자 장치(300)는 디스플레이 장치(200)와 유무선으로 연결되어 전자 장치(100)로부터 수신한 응답 정보를 디스플레이 장치(200)에 전송할 수 있다. 응답 정보를 제공하는 디스플레이 장치(200)에 관한 설명은 도 1(a)와 중복되는바, 구체적인 설명은 생략한다.
- [0061] 한편, 도 1(c)에는 외부 전자 장치(300)와 디스플레이 장치(200)를 구분하여 도시하였으나, 도 1(d)에서와 같이, 외부 전자 장치(300)의 기능 및 디스플레이 장치(200)의 기능이 하나의 장치에서 구현될 수 있음은 물론이다.
- [0062] 도 1(d)에서와 같이, 전자 장치(100)가 서버로 구현되고, 외부 전자 장치(300)의 기능 및 디스플레이 장치(200)의 기능이 하나의 장치로 구현되는 경우, 디스플레이 장치(200)는 전자 장치(100)와 연결되어 사용자 음성을 수신하고, 수신된 사용자 음성을 음성 신호로 변환한 후, 변환된 음성 신호를 전자 장치(100)에 전송할 수 있다.
- [0063] 전자 장치(100)는 디스플레이 장치(200)로부터 수신한 사용자 음성에 대하여 음성 인식을 수행하고, 사용자 음성에 대한 응답 정보를 생성할 수 있다.
- [0064] 구체적으로, 전자 장치(100)는 사용자 음성 신호를 텍스트로 변환하고, 사용자 음성에 대한 자연어 이해 처리를 수행하여, 사용자 음성의 의도 및 엔티티를 파악할 수 있다. 그리고, 전자 장치(100)는 자연어 이해 결과에 기초하여 사용자 음성에 대한 응답을 위한 정보를 웹 서버 등과 같은 또 다른 서버로부터 획득하고, 획득한 정보에 기초하여 사용자 음성에 대한 응답 정보를 생성할 수 있다.
- [0065] 전자 장치(100)는 생성된 응답 정보를 디스플레이 장치(200)에 전송할 수 있으며, 디스플레이 장치(200)는 사용자에게 응답 정보를 제공하거나, 응답 정보에 따른 동작을 수행할 수 있다.
- [0066] 도 1 (a) 내지 (d)에서는 사용자가 직접 전자 장치(100)에 음성을 입력하는 것으로 도시되었으나, 사용자는 원격 제어 장치(미도시)에 사용자 음성을 입력할 수도 있다.
- [0067] 이 경우, 원격 제어 장치(미도시)는 수신한 사용자 음성(아날로그 신호)을 디지털 신호로 변환하고, 블루투스(Bluetooth) 등의 무선 통신을 통하여 변환된 사용자 음성 신호를 전자 장치(100)에 전송할 수 있다.
- [0068] 한편, 본 개시의 일 실시예에 따른 전자 장치(100)가 음성 인식 기능을 제공하기 위해서는 음성을 발화하는 화자에 대한 화자 인증이 필요하다.
- [0069] 여기에서, 화자 인증은 화자의 음성에 포함된 특징을 기저장된 음성의 특징과 비교하여 음성을 발화한 화자가 보안성 있는 전자 장치 또는 콘텐츠에 접근(access)할 권한이 있는지 식별하는 기술을 의미한다.
- [0070] 한편, 본 개시의 일 실시예에 따르면, 전자 장치(100)는 제1 사용자 음성의 음향 특징 및 문맥 정보를 포함하는 화자 모델을 생성하여 저장한다. 그리고, 제2 사용자 음성이 수신되면 전자 장치(100)는 화자 모델에 포함된 제1 사용자 음성의 음향 특징과 제2 사용자 음성의 음향 특징간 사이의 유사도를, 화자 모델에 포함된 제1 사용자 음성의 문맥 정보와 제2 사용자 음성의 문맥 정보간 유사도에 따라 변경되는 임계값과 비교하여 제2 사용자 음성에 대하여 화자 인증을 수행한다.
- [0071] 이에 따라, 본 개시의 다양한 실시예에 따르면, 제2 사용자 음성의 문맥 정보와 화자 모델에 포함된 문맥 정보의 유사도를 이용하여, 제1 및 제2 음향 특징간의 유사도에 대한 임계값을 결정한다는 점에서, 제2 사용자 음성의 문맥 정보와 화자 모델에 포함된 문맥 정보의 유사도에 따라 제2 사용자 음성의 음향 특징 유사 판단에 대한 임계값이 조정될 수 있다는 이점이 있다.
- [0072] 이하에서, 본 개시의 다양한 실시예에 따라 화자 인증을 수행하는 방법에 대해 구체적으로 설명하도록 한다.

- [0073] 도 2는 본 개시의 일 실시 예에 따른 전자 장치의 구성을 설명하기 위한 블록도이다.
- [0074] 도 2를 참조하면, 전자 장치(100)는 메모리(110) 및 프로세서(120)를 포함할 수 있다.
- [0075] 메모리(110)는 전자 장치(100)의 동작에 필요한 각종 프로그램 및 데이터 등을 저장할 수 있다.
- [0076] 구체적으로, 메모리(110)는 제1 사용자 음성의 음향 특징 및 문맥 정보(context information)을 포함하는 화자 모델을 저장할 수 있다.
- [0077] 여기에서, 제1 사용자 음성의 음향 특징은 제1 사용자 음성의 주파수 별 분포를 나타내는 spectrum 특징 또는 제1 사용자 음성의 높낮이, 크기, 속도, 리듬, 강세 등을 나타내는 prosody 특징을 포함할 수 있다.
- [0078] 한편, 제1 사용자 음성의 문맥 정보는 제1 사용자 음성에 포함된 음소의 위치 정보, 음절의 위치 정보, 단어의 위치 정보, 음소 정보(Uniphone, Triphone, Quinphone), 음소의 개수 정보, 음절 정보, 단어 정보 중 적어도 하나를 포함할 수 있다.
- [0079] 한편, 상술한 문맥 정보는 일 실시예이며, 시스템에 따라 또 다른 문맥 정보가 추가될 수 있음은 물론이다.
- [0080] 프로세서(120)는 메모리(110)와 전기적으로 연결되어 전자 장치(100)의 전반적인 동작 및 기능을 제어할 수 있다.
- [0081] 프로세서(120)는 수신된 사용자 음성을 처리하여 음성 신호를 생성하도록 마이크를 제어할 수 있다. 구체적으로, 프로세서(120)는 사용자 음성에 대하여 아날로그/디지털 변환, 필터링, 압축, 암호화 등의 신호 처리를 수행하여 음성 신호를 생성하도록 마이크를 제어할 수 있다.
- [0082] 프로세서(120)는 음성 신호로부터 음향 특징을 획득할 수 있다. 구체적으로, 프로세서(120)는 음성 신호로부터 사용자 음성의 주파수, 높낮이, 크기, 속도, 리듬, 강세 등 음향 특징을 획득할 수 있다.
- [0083] 프로세서(120)는 음성 신호로부터 문맥 정보를 획득할 수도 있다. 프로세서(120)는 음성 신호에 대한 음성 인식을 수행하여 문맥 정보를 획득할 수 있다. 프로세서(120)는 ASR 엔진을 이용하여 사용자의 음성 신호를 텍스트 데이터로 변환하고, NLU 엔진을 이용하여 사용자 음성에 대한 형태소 분석, 구문 분석, 의미이행 분석 등의 자연어 처리를 수행할 수 있다.
- [0084] 한편, 프로세서(120)는 사용자 음성 신호를 텍스트 데이터로 변환하는 대신에 사용자 음성 신호를 서버(미도시)에 전송할 수 있다. 이 경우 서버(미도시)는 STT(Speech to Text)기능만을 수행하는 STT서버이거나, 사용자 음성 신호에 대한 STT 기능을 수행할 뿐만 아니라 자연어 처리까지 수행하는 서버일 수 있다. 서버(미도시)가 STT 서버인 경우, 서버(미도시)는 사용자 음성 신호를 텍스트로 변환한 후에 변환된 사용자 음성에 대한 자연어 처리 등을 수행하기 위하여 이를 다른 서버에 또 다시 전송할 수도 있다.
- [0085] 프로세서(120)는 음성 인식을 수행하기 전에 음성 신호로부터 음향 특징을 획득하고, 음성 데이터에 대한 음성 인식을 수행할 수 있다. 그러나, 반드시 이에 한정되는 것은 아니며, 프로세서(120)는 음성 신호로부터 음향 특징을 획득하는 동시에, 음성 인식 엔진, 자연어 처리 엔진 등을 이용하여 문맥 정보를 획득할 수 있다.
- [0086] 프로세서(120)는 메모리(110)에 저장된 화자 모델(111)을 이용하여 화자에 대한 인증을 수행할 수 있다.
- [0087] 화자에 대한 인증을 위해, 먼저, 프로세서(120)는 화자를 등록할 수 있다. 이 경우, 화자 인증은 사용자 음성의 음향 특성을 전자 장치(100)에 기저장된 화자 모델에 포함된 음향 특성과 비교하여 수행된다는 점에서, 화자를 등록하는 것은 화자 모델을 생성하는 것으로 볼 수 있다.
- [0088] 프로세서(120)는 제1 사용자 음성이 수신되면, 제1 사용자 음성으로부터 제1 사용자 음성에 대한 음향 특징 및 문맥 정보를 획득하고, 획득된 음향 특징에 및 문맥 정보를 포함하는 화자 모델을 메모리(110)에 저장할 수 있다.
- [0089] 이하에서는 도 4와 함께, 화자 모델을 생성하는 방법에 대해 설명하도록 한다.
- [0090] 먼저, 프로세서(120)는 제1 사용자 음성이 수신되면(410), 제1 사용자 음성에 대한 음향 특징을 획득할 수 있다.
- [0091] 구체적으로, 프로세서(120)는 제1 사용자 음성을 시간 도메인 및 주파수 도메인에서 분석하여, 제1 사용자 음성의 스펙트럼, 크기, 속도, 리듬, 강세 등과 같은 제1 사용자 음성에 대한 음향 특징을 획득할 수 있다(450).
- [0092] 또한, 프로세서(120)는 제1 사용자 음성이 수신되면(410), 제1 사용자 음성에 대한 문맥 정보를 획득할 수

있다.

- [0093] 구체적으로, 프로세서(120)는 메모리(110)에 저장된 음성 인식 모듈을 이용하여 제1 사용자 음성에 대해 음성 인식을 수행할 수 있다(420). 즉, 프로세서(120)는 제1 사용자 음성에 대한 음성 인식을 수행하여, 제1 사용자 음성을 텍스트로(Speech-To-Text, STT) 변환할 수 있다.
- [0094] 그리고, 프로세서(120)는 자연어 처리 모듈을 이용하여 제1 사용자 음성의 음성 인식 결과를 자연어 처리할 수 있다(430). 구체적으로, 프로세서(120)는 자연어 처리 과정을 통해, 텍스트로 변환된 제1 사용자 음성에 대해 형태소 분석, 구문 분석, 의미 이해 분석 등을 할 수 있다.
- [0095] 그리고, 프로세서(120)는 자연어 처리된 제1 사용자 음성을 이용하여 제1 사용자 음성으로부터 문맥 정보를 획득할 수 있다(440).
- [0096] 예를 들어, 제1 사용자 음성의 문맥 정보는 제1 사용자 음성의 음소 정보, 음소의 위치 정보, 음소의 개수 정보, 음절의 위치 정보 및 단어의 위치 정보 중 적어도 하나를 포함할 수 있다.
- [0097] 여기에서, 음소의 위치 정보는 음소가 단어나 문장을 기준으로 위치하는 순서를 나타내고, 음절의 위치 정보는 음절이 단어나 문장을 기준으로 위치하는 순서를 나타내며, 단어의 위치 정보는 단어가 문장을 기준으로 위치하는 순서를 나타낸다. 그리고, 음소의 개수 정보는 단어나 문장에 존재하는 음소의 개수를 나타낸다. 또한, 음소 정보는 현재 음소 정보를 나타내는 Uniphone 정보, 현재 음소 및 현재 음소를 기준으로 앞, 뒤에 존재하는 음소에 대한 정보(즉, 이전 음소 - 현재 음소 - 다음 음소에 대한 정보)를 나타내는 Triphone 정보 및 현재 음소 및 현재 음소를 기준으로 앞, 뒤에 2 개씩 연속적으로 존재하는 음소에 대한 정보(즉, 이전이전 음소 - 이전 음소 - 현재 음소 - 다음 음소 - 다음다음 음소)를 나타내는 Quinphone 정보를 포함할 수 있다.
- [0098] 예를 들어, 사용자 음성이 "sun"인 경우를 가정한다. 이 경우, Uniphone 정보는 현재 음소를 나타낸다는 점에서, Uniphone(s) = "s", Uniphone(u) = "u", Uniphone(n) = "n"이 될 수 있다. 또한, Triphone 정보는 현재 음소 및 현재 음소를 기준으로 앞, 뒤에 존재하는 음소를 나타낸다는 점에서, Triphone(s) = "- s u", Triphone(u) = "s u n", Triphone(n) = "u n -"이 될 수 있다. 그리고, Quinphone 정보는 현재 음소 및 현재 음소를 기준으로 앞, 뒤에 2 개씩 연속적으로 존재하는 음소를 나타낸다는 점에서, Quinphone(s) = "- - s u n", Quinphone(u) = "- s u n -", Quinphone(n) = "s u n -"이 될 수 있다.
- [0099] 이때, 프로세서(120)는 기정의된 복수의 문맥 정보 타입 중, 특정한 타입의 문맥 정보를 선택하고, 제1 사용자 음성으로부터, 선택된 타입의 문맥 정보를 획득할 수 있다.
- [0100] 즉, 전술한 바와 같이, 문맥 정보는 음소 정보, 음소의 위치 정보, 음소의 개수 정보, 음절의 위치 정보 및 단어의 위치 정보 등 다양한 타입으로 구성될 수 있다. 이 경우, 프로세서(120)는 기정의된 복수의 문맥 정보 타입 중 하나의 문맥 정보를 선택하고, 선택된 문맥 정보 타입을 이용하여 제1 사용자 음성의 문맥 정보를 획득할 수 있다.
- [0101] 이를 위해, 프로세서(120)는 제1 사용자 음성의 음향 특징과 복수의 테스트 음성 각각의 음향 특징 사이의 유사도 및 기정의된 복수의 문맥 정보의 타입 별로, 제1 사용자 음성의 문맥 정보와 복수의 테스트 음성 각각의 문맥 정보 사이의 유사도를 판단할 수 있다.
- [0102] 여기에서, 복수의 테스트 음성은 제1 사용자 음성을 발화한 화자로부터 획득할 수도 있으며, 복수의 사용자 음성을 포함하는 데이터 베이스를 이용하여 획득할 수도 있다.
- [0103] 먼저, 프로세서(120)는 제1 사용자 음성의 음향 특징과 복수의 테스트 음성 각각의 음향 특징 사이의 유사도를 판단할 수 있다.
- [0104] 구체적으로, 프로세서(120)는 복수의 테스트 음성 각각에 포함된 복수의 테스트 음성의 spectrum 특징 또는 prosody 특징을 제1 사용자 음성의 spectrum 특징 또는 prosody 특징과 비교하여, 제1 사용자 음성에 대한 복수의 테스트 음성 각각의 스펙트럼, 높낮이, 크기, 속도, 리듬, 강세 등의 유사한 정도를 판단하고, 유사한 정도에 대응되는 스코어를 산출할 수 있다.
- [0105] 또한, 프로세서(120)는 기정의된 복수의 문맥 정보의 타입 별로, 제1 사용자 음성의 문맥 정보와 복수의 테스트 음성 각각의 문맥 정보 사이의 유사도를 판단할 수 있다.
- [0106] 이를 위해, 프로세서(120)는 기 정의된 복수의 문맥 정보의 타입 별로, 제1 사용자 음성 및 복수의 테스트 음성 각각으로부터 각 타입에 대한 문맥 정보를 획득할 수 있다.

[0107] 그리고, 프로세서(120)는 제1 사용자 음성 및 복수의 테스트 음성 각각으로부터 획득된 문맥 정보를 서로 비교하여, 이들 문맥 정보 사이의 유사도를 판단할 수 있다.

[0108] 이 경우, 프로세서(120)는 하기의 수학적 식 1에 기초하여 유사도를 판단할 수 있다.

수학적 식 1

$$S(A,B) = \frac{|A \cap B|}{|B|}$$

[0110] 여기에서, A는 제1 사용자 음성의 문맥 정보를 의미하고, B는 테스트 음성의 문맥 정보를 의미한다.

[0111] 가령, 제1 사용자 음성은 'sun'이고, 테스트 음성은 'sing'이며, 문맥 정보 타입은, 도 2에서 상술한 복수의 문맥 정보 타입 중 현재 음소 정보 즉, uniphone 정보를 사용한다고 가정한다.

[0112] 이 경우, 음소는 발음을 기준으로 판단한다는 점에서, 'sun'의 uniphone은 "s","u","n"이고, 'sing'의 uniphone은 "s","i","ng"이 될 수 있으며, 이들을 binary로 표기하면 하기의 표 1과 같이 나타낼 수 있다.

표 1

	sun	sing
uniphone (s)	1	1
uniphone (u)	1	0
uniphone (n)	1	0
uniphone (i)	0	1
uniphone (ng)	0	1

[0114] 구체적으로, 음소 s는 'sun' 및 'sing'에 포함되므로, A의 uniphone(s) 및 B의 uniphone(s)의 값은 모두 1이다. 음소 u, n은 'sun'에 포함되나, 'sing'에 포함되지 않으므로, A의 uniphone(u), uniphone(n)은 1, B의 uniphone (u), uniphone(n)의 값은 0이다. 또한, 음소 i, ng는 'sing'에 포함되나, 'sun'에 포함되지 않으므로, A의 uniphone(i), uniphone(ng)은 0, B의 uniphone(i), uniphone(ng)의 값은 1이다.

[0115] 여기에서 음소 s가 공통된다는 점에서 $|A \cap B| = 1$ 이고, 테스트 음성 B의 음소가 3개라는 점에서 $|B| = 3$ 이다.

[0116] 이에 따라, 수학적 식 1에서, $S(A,B) = 1/3$ 인 것을 알 수 있다.

[0117] 또 다른 실시 예로, 문맥 정보 타입으로 triphone을 이용하는 경우를 가정한다. 이 경우, 제1 사용자 음성 'sun' 및 테스트 음성 'sing'의 triphone에 대한 정보는 하기의 표 2와 같이 나타낼 수 있다.

표 2

	sun	sing
triphone (s)	- s n	- s i
triphone (u)	s u n	0
triphone (n)	u n -	0
triphone (i)	0	s i n
triphone (ng)	0	i ng -

[0119] 즉, triphone(x)의 경우 제1 및 테스트 음성의 '음소 x의 이전 음소 - 현재 음소 x - 음소 x의 다음 음소' 중 적어도 2개 이상의 음소가 동일한 경우, 제1 사용자 음성 및 테스트 음성의 triphone(x)를 1로 표시할 수 있다.

[0120] 가령, 표 2에서 triphone(x)에서 제1 사용자 음성 및 테스트 음성의 '음소 x의 이전 음소 - 현재 음소 x - 음소 x'의 다음 음소 중 적어도 2개 이상이 동일한 경우가 없으므로, 제1 사용자 음성 및 테스트 음성의 유사도 값은 0이 될 수 있다.

- [0121] 한편, 이는 일 실시예이며, 유사도 판단은 경우에 따라 다를 수 있다. 가령 triphone(x)의 경우 제1 사용자 음성 및 테스트 음성의 '음소 x의 이전 음소 - 현재 음소 x - 음소 x의 다음 음소' 중 적어도 1개 이상의 음소가 동일한 경우, 제1 사용자 음성 및 테스트 음성의 triphone(x)를 1로 표시할 수 있을 수도 있다.
- [0122] 이 경우, 상기 제1 사용자 음성의 triphone(s) = "- s n" 와 테스트 음성의 triphone(s) = "- s i" 에서 음소 s가 동일하므로, 제1 사용자 음성 및 테스트 음성의 triphone(s)는 binary 표현으로 1로 나타낼 수 있다. 이 경우, 제 1 사용자 음성 및 테스트 음성의 유사도 $S(A, B) = 1/3$ 이 될 수 있다.
- [0123] 한편, 제1 사용자 음성의 문맥 정보와 테스트 음성의 문맥 정보 사이의 유사도를 판단하는 식은 상기 수학적 1에 국한된 것만은 아니며, 다양한 실시 예에 따라 다양한 판단 방법이 사용될 수 있다. 가령 자연어 처리 방식에서 사용되는 jaccard similarity measure, cosine similarity measure 등의 방법이 유사도 판단에 사용될 수도 있다.
- [0124] 한편, 프로세서(120)는 기 정의된 복수의 문맥 정보 타입 별로, 테스트 음향 특징 유사도에 대한 테스트 문맥 정보 유사도의 분포가 기설정된 조건을 만족하는 문맥 정보 타입을 선택할 수 있다.
- [0125] 예를 들어, 프로세서(120)는 테스트 음향 특징 유사도에 대한 테스트 문맥 정보 유사도의 분포가 선형성을 나타내는 문맥 정보 타입을 선택할 수 있다.
- [0126] 예를 들어, 제1 사용자 음성은 'sun' 이고, 테스트 음성은 'drum', 'sing' 및 'son' 라고 하자. 이 경우, 음향 특징의 유사도와 관련하여 'sun' 음향 특징과 'drum', 'sing' 및 'son' 음향 특징 사이의 유사도는 각각 0.3, 0.5, 0.8 이라고 가정한다.
- [0127] 그리고, A 타입의 문맥 정보를 사용할 경우 'sun'의 문맥 정보와 'drum', 'sing' 및 'son' 문맥 정보 사이의 유사도는 0.2, 0.8, 0.6 이라고 가정하고, B 타입의 문맥 정보를 사용할 경우 'sun'의 문맥 정보와 'drum', 'sing' 및 'son' 문맥 정보 사이의 유사도는 0.2, 0.6, 0.8 이라고 가정한다.
- [0128] 이때, 테스트 음향 특징 유사도에 대한 테스트 문맥 정보 유사도의 분포가 선형성을 나타내는 문맥 정보 타입은 B 타입의 문맥 정보이며, 프로세서(120)는 B 타입의 문맥 정보를 선택할 수 있다.
- [0129] 이에 따라, 프로세서(120)는 제1 사용자 음성으로부터 문맥 정보를 획득할 수 있다(440).
- [0130] 이후, 프로세서(120)는 획득한 문맥 정보 및 음향 특징을 포함하는 화자 모델(111)을 생성할 수 있다.
- [0131] 화자 모델(111)이 생성되어, 메모리(110)에 저장되면 화자 등록 단계가 완료된다.
- [0132] 한편, 프로세서(120)는 화자 모델을 이용하여, 화자에 대한 인증을 수행할 수 있다.
- [0133] 도 5는 본 개시의 일 실시예에 따른 전자 장치를 설명하기 위한 도면으로, 구체적으로, 화자 인증을 수행하는 전자 장치(100)를 설명하기 위한 도면이다.
- [0134] 프로세서(120)는 제2 사용자 음성이 수신되면(510), 제 2 사용자 음성의 음향 특징을 이용하여 제 2 사용자 음성의 화자 모델에 대한 스코어(score)를 계산할 수 있다(540).
- [0135] 여기에서 제2 사용자 음성의 화자 모델에 대한 스코어는 제2 사용자 음성의 음향 특징과 화자 모델에 저장된 제 1 사용자 음성의 음향 특징 사이의 유사도를 나타내는 값일 수 있다. 이와 관련하여, 제2 사용자 음성의 음향 특징과 화자 모델에 저장된 제1 사용자 음성의 음향 특징 사이의 유사도 판단 방법은, 상기 도 4에서 상술한 제 1 사용자 음성의 음향 특징과 복수의 테스트 음성의 음향 특징 사이의 유사도 판단 방법과 동일한 바, 구체적인 설명은 생략하기로 한다.
- [0136] 그리고, 프로세서(120)는 제2 사용자 음성의 음향 특징과 화자 모델(111)에 저장된 제1 사용자 음성의 음향 특징 사이의 유사도를 임계값과 비교할 수 있다.
- [0137] 이때, 임계값은 제2 사용자 음성의 문맥 정보와 화자 모델(111)에 포함된 문맥 정보 사이의 유사도에 기초하여 산출될 수 있다(550).
- [0138] 한편, 이하에서는 임계 값을 산출하는 방법에 대해 구체적으로 설명하도록 한다.
- [0139] 먼저, 임계값을 산출하기 위하여, 프로세서(120)는 메모리(110)에 저장된 음성 인식 모듈을 이용하여 제2 사용자 음성에 대해 음성 인식을 수행할 수 있다(520). 즉, 프로세서(120)는 제2 사용자 음성에 대한 음성 인식을 수행하여, 제1 사용자 음성을 텍스트로(Speech-To-Text, STT) 변환할 수 있다.

- [0140] 그리고, 프로세서(120)는 음성 인식 모듈을 이용하여 제2 사용자 음성의 음성 인식 결과를 자연어 처리할 수 있다(530). 구체적으로, 프로세서(120)는 자연어 처리 과정을 통해, 텍스트로 변환된 제1 사용자 음성에 대해 형태소 분석, 구문 분석, 의미 이해 분석 등을 할 수 있다.
- [0141] 프로세서(120)는 자연어 처리된 제2 사용자 음성을 이용하여 제2 사용자 음성의 문맥 정보를 생성할 수 있다.
- [0142] 이때, 프로세서(120)는 화자 모델(111)에 포함된 제1 사용자 음성의 문맥 정보를 고려하여, 제2 사용자 음성의 문맥 정보를 생성할 수 있다. 가령, 화자 모델(111)에 포함된 제1 사용자 음성의 문맥 정보가 Uniphone 타입의 문맥 정보를 이용하여 생성된 것이라면, 프로세서(120)는 화자 모델에 저장된 제1 사용자 음성의 문맥 정보를 고려하여, 제2 사용자 음성의 문맥 정보를 Uniphone 타입의 문맥 정보로 생성할 수 있다.
- [0143] 그 후, 프로세서(120)는 제2 사용자 문맥 정보와 화자 모델에 저장된 제1 사용자 문맥 정보의 유사도를 판단할 수 있다. 제2 사용자 음성의 문맥 정보와 화자 모델에 저장된 제1 사용자 음성의 문맥 정보 사이의 유사도 판단 방법은 상기 도 4에서 상술한 제1 사용자 음성의 문맥 정보와 복수의 테스트 음성의 문맥 정보 사이의 유사도 판단 방법과 동일한 바, 구체적인 설명은 생략하기로 한다.
- [0144] 한편, 프로세서(120)는 제2 사용자 음성의 문맥 정보와 화자 모델(111)에 포함된 문맥 정보 사이의 유사도에 기초하여, 임계값을 산출할 수 있다.
- [0145] 구체적으로, 프로세서(120)는 제2 사용자 음성의 문맥 정보와 화자 모델에 포함된 문맥 정보 사이의 유사도, 및 제1 사용자 음성의 문맥 정보와 제2 사용자 음성의 문맥 정보가 동일한 경우 제1 및 제2 음향 특징 사이의 유사도가 가질 수 있는 최대 값에 기초하여 임계값을 산출할 수 있다.
- [0146] 이 경우, 임계값은 제2 사용자 음성의 문맥 정보와 화자 모델에 포함된 문맥 정보 사이의 유사도에 비례하도록 결정될 수 있다.
- [0147] 가령, 프로세서(120)는 아래의 식을 이용하여 임계값을 산출할 수 있다.

수학식 2

$$Th(S(A,B)) = W \times S(A,B) \times Max_score$$

- [0148]
- [0149] 여기에서 $Th(S(A,B))$ 는 제1 사용자 음성의 문맥 정보와 제2 사용자 음성의 문맥 정보의 유사도에 대한 임계값을 나타내며, $S(A,B)$ 는 제1 사용자 문맥 정보와 제2 사용자 문맥 정보의 유사도를 나타낸다. Max_score 는 화자 모델에 포함된 제1 사용자 음성의 문맥 정보와 제2 사용자 음성의 문맥 정보가 동일한 경우 제1 및 제2 사용자 음성의 음향 특징 사이의 유사도가 가질 수 있는 최대 값에 해당한다. 또한, w 는 기설정된 가중치에 해당한다.
- [0150] 예를 들어, 제1 사용자 음성이 'sun'이고, 제2 사용자 음성이 'gun'인 경우, 제2 사용자 음성의 문맥 정보 및 화자 모델에 포함된 제1 사용자 음성의 문맥 정보는 binary로 아래와 같이 표현될 수 있다.

표 3

	sun	sing
uniphone (s)	1	0
uniphone (u)	1	1
uniphone (n)	1	1
uniphone (g)	0	1

- [0151]
- [0152] 이때, 수학식(2)를 사용하여 제2 사용자 음성의 문맥 정보 및 화자 모델에 포함된 제1 사용자 음성의 문맥 정보의 유사도 $S(A,B)$ 는 2/3로 나타낼 수 있다.
- [0153] 그리고, w 및 Max_score 는 기설정된 값이라는 점에서, 임계값 ($Th(S(A,B))$)은 제2 사용자 음성의 문맥 정보와 화자 모델에 포함된 제1 사용자 문맥 정보 사이의 유사도($S(A,B)$)에 비례하도록 결정되는 것을 알 수 있다. 즉, 제2 사용자 문맥 정보와 화자 모델에 포함된 문맥 정보간의 유사도가 낮아지면, 임계값 또한 낮아질 수 있다.
- [0154] 한편 상기 수학식 2는 본 개시의 일 실시 예에 따른 수식이며, 경우에 따라서는 임계값은 제2 사용자 음성의 문맥 정보와 화자 모델에 포함된 문맥 정보 사이의 유사도에 반비례하도록 결정될 수도 있다. 즉, 아래와 같은 수

식을 적용하여 임계값을 산출할 수도 있다.

수학식 3

$$Th(S(A,B)) = W \times \frac{1}{S(A,B)} \times Max_score$$

[0155]

[0156]

수학식 2와 마찬가지로, $Th(S(A,B))$ 는 제1 사용자 음성의 문맥 정보와 제2 사용자 음성의 문맥 정보의 유사도에 대한 임계값을 나타내며, $S(A,B)$ 는 제1 사용자 문맥 정보와 제2 사용자 문맥 정보의 유사도, Max_score 는 화자 모델에 포함된 제1 사용자 음성의 문맥 정보와 제2 사용자 음성의 문맥 정보가 동일한 경우 제1 및 제2 사용자 음성의 음향 특징 사이의 유사도가 가질 수 있는 최대 값을 나타낸다.

[0157]

예를 들어, 제1 사용자 음성이 'sun'이고, 제2 사용자 음성이 'gun'인 경우, 상기 표 2 및 수학식 2를 사용하여 제2 사용자 음성의 문맥 정보 및 화자 모델에 포함된 제1 사용자 음성의 문맥 정보의 유사도 $S(A,B)$ 는 2/3로 나타낼 수 있다.

[0158]

그리고, w 및 Max_score 는 기설정된 값이라는 점에서, 임계값 ($Th(S(A,B))$)은 제2 사용자 음성의 문맥 정보와 화자 모델에 포함된 제1 사용자 문맥 정보 사이의 유사도($S(A,B)$)에 반비례하도록 결정되는 것을 알 수 있다. 즉, 제2 사용자 문맥 정보와 화자 모델에 포함된 문맥 정보간의 유사도가 낮아지면, 반대로 임계값은 높아질 수 있다.

[0159]

한편, 상기 수학식 2 및 수학식 3의 실시예는 본 개시의 일 실시예이며, 프로세서(120)는, 화자 모델(111)에 포함된 제1 사용자 음성의 문맥 정보 및 제2 사용자 음성의 문맥 정보 간의 유사도가 기설정된 값(제1 값) 이하인 경우, 임계값을 기설정된 값(제2 값)으로 설정할 수도 있다.

[0160]

가령, 프로세서(120)는, 화자 모델(111)에 포함된 제1 사용자 음성의 문맥 정보 및 제2 사용자 음성의 문맥 정보 간의 유사도가 0.5 이하인 경우, 임계값을 1로 설정할 수도 있다.

[0161]

반대로, 프로세서(120)는 화자 모델(111)에 포함된 제1 사용자 음성의 문맥 정보 및 제2 사용자 음성의 문맥 정보 간의 유사도가 기설정된 값(제1 값) 이상인 경우, 임계값을 기설정된 값(제2 값)으로 설정할 수도 있다.

[0162]

가령, 프로세서(120)는 화자 모델(111)에 포함된 제1 사용자 음성의 문맥 정보 및 제2 사용자 음성의 문맥 정보 간의 유사도가 0.9 이상인 경우, 임계값을 0.5로 설정할 수도 있다.

[0163]

프로세서(120)는 이와 같이 산출된 임계값을 제 2 사용자 음성의 음향 특징과 화자 모델에 저장된 제1 사용자 음성의 음향 특징 사이의 유사도와 비교하여 제2 사용자 음성을 발화한 화자에 대한 인증을 수행할 수 있다 (560).

[0164]

구체적으로, 프로세서(120)는 제2 사용자 음성의 음향 특징과 화자 모델(111)에 저장된 제1 사용자 음성의 음향 특징 사이의 유사도가 임계 값 이상인 경우, 화자에 대한 인증이 성공된 것으로 판단할 수 있다.

[0165]

가령, 제1 사용자 음성(A)이 'sun'이고, 제2 사용자 음성(B)이 'gun'이며, $S(A,B) = 2/3 \approx 0.66$ 라고 가정하자.

[0166]

이 경우 임계값 $Th(S(A,B))$ 는 상기 수학식 2, 수학식 3 또는 설정 방법에 따라 서로 다른 값을 가지도록 산출될 수 있으며, 프로세서(120)는 제2 사용자 음성의 음향 특징과 화자 모델의 음향 특징 사이의 유사도가 산출된 임계값 이상의 값을 가지는 경우에, 제2 사용자 음성을 발화한 화자에 대한 인증이 성공된 것으로 판단할 수 있다.

[0167]

인증이 성공된 것으로 판단된 경우, 프로세서(120)는 제2 사용자 음성을 발화한 화자의 접근(access)을 허용하고 보안성 있는 콘텐츠를 제공할 수 있다.

[0168]

반면, 제2 사용자 음성의 음향 특징과 화자 모델의 음향 특징 사이의 유사도가 산출된 임계값 미만의 값을 가져 인증이 성공된 것으로 판단되지 않은 경우, 프로세서(120)는 제2 사용자 음성을 발화한 화자의 접근(access)을 제한하고 보안성 있는 콘텐츠를 제공하지 않을 수 있다.

[0169]

도 3은 본 개시의 일 실시예에 따른 전자 장치의 구성을 상세히 도시한 블록도이다. 도 3에 도시된 바와 같이, 전자 장치(100)는 메모리(110), 통신부(130), 디스플레이(140), 마이크(150), 오디오출력부(160), 사용자입력부

(170) 및 프로세서(120)를 포함할 수 있다.

- [0170] 메모리(110)는 전자 장치(100)의 동작에 필요한 각종 프로그램 및 데이터를 저장할 수 있다. 메모리(110)는 비휘발성 메모리, 휘발성 메모리, 플래시메모리(flash-memory), 하드디스크 드라이브(HDD) 또는 솔리드 스테이트 드라이브(SSD) 등으로 구현될 수 있다. 메모리(110)는 프로세서(120)에 의해 액세스되며, 프로세서(120)에 의한 데이터의 독취/기록/수정/삭제/갱신 등이 수행될 수 있다. 본 문서에서 메모리라는 용어는 메모리(110), 프로세서(120) 내 롬(미도시), 램(미도시) 또는 전자 장치(100)에 장착되는 메모리 카드(미도시)(예를 들어, micro SD 카드, 메모리 스틱)를 포함할 수 있다.
- [0171] 또한, 메모리(110)는 제1 사용자 음성의 음향 특징 및 문맥 정보를 포함하는 화자 모델을 저장할 수도 있다.
- [0172] 그리고, 메모리(110)는 제1 및 제2 사용자 음성을 음성 인식하고 자연어 처리 하기 위한 음성인식모듈(112) 및 제1 및 제2 사용자 음성의 음향 특징을 추출하기 위한 음향 특징 추출 모듈(113)을 저장할 수도 있다.
- [0173] 통신부(130)는 전자 장치(100)가 외부 전자 장치(200)와 통신을 수행하기 위한 구성요소이다. 통신부(130)를 통하여 전자 장치(100)는 외부 장치(200)로부터 사용자 음성 신호를 수신할 수 있다.
- [0174] 이를 위해, 통신부(130)는 유선 통신 모듈(미도시), 근거리 무선 통신 모듈(미도시), 무선 통신 모듈(미도시) 등과 같은 다양한 통신 모듈을 포함할 수 있다.
- [0175] 여기에서, 유선 통신 모듈은 유선 이더넷(Ethernet)과 같이 유선 통신 방식에 따라 외부 장치(미도시)와 통신을 수행하기 위한 모듈이다. 그리고, 근거리 무선 통신 모듈이란 블루투스(Bluetooth, BT), BLE(Bluetooth Low Energy), ZigBee 방식 등과 같은 근거리 무선 통신 방식에 따라 근거리에서 위치한 외부 장치(미도시)와 통신을 수행하기 위한 모듈이다. 또한, 무선 통신 모듈이란 WiFi, IEEE 등과 같은 무선 통신 프로토콜에 따라 외부 네트워크에 연결되어 외부 장치(미도시) 및 음성 인식 서버(미도시)와 통신을 수행하는 모듈이다. 이 밖에 무선 통신 모듈은 3G(3rd Generation), 3GPP(3rd Generation Partnership Project), LTE(Long Term Evolution), LTE-A(LTE Advanced), 5세대 네트워크(5G Networks) 등과 같은 다양한 이동 통신 규격에 따라 이동 통신망에 접속하여 통신을 수행하는 이동 통신 모듈을 더 포함할 수도 있다.
- [0176] 한편, 통신부(130)는 외부 장치(200)뿐만 아니라, 외부 전자 장치(200)를 제어하기 위한 원격 제어 장치와 통신을 수행할 수도 있다. 이때, 외부 제어 장치(200)는 리모컨, 포인팅 디바이스 뿐만 아니라 어플리케이션을 설치하여 외부 장치를 제어할 수 있는 스마트폰, 스마트워치, 태블릿PC, PC, 노트북 등이 될 수도 있다.
- [0177] 예를 들어, 통신부(130)는 블루투스 또는 BLE 등과 같은 근거리 통신 방식에 따라 원격 제어 장치(미도시)와 통신을 수행하여, 다양한 데이터를 송수신할 수 있다.
- [0178] 한편, 또 다른 실시예로, 전자 장치(100)는 외부 장치(300)와 연결되어 비디오/오디오 신호를 주고 받기 위한 인터페이스부(미도시)를 포함할 수도 있다. 인터페이스부는 DP(Display Port), HDMI(high definition multimedia interface), HDMI-CEC(consumer electronics control), RGB(Red Green Blue) cable, USB(Universal Serial Bus), DVI(Digital Visual Interface), 썬더볼트(Thunderbolt), 컴포넌트(Component) 등의 규격에 따른 신호/데이터를 송/수신할 수 있으며, 이들 각각의 규격에 대응하는 적어도 하나 이상의 커넥터 또는 단자를 포함한다.
- [0179] 디스플레이(140)는 영상 처리부(미도시)에서 처리한 영상 데이터를 디스플레이 영역(또는, 디스플레이)에 디스플레이할 수 있다. 디스플레이 영역은 전자 장치(100)의 하우징의 일면에 노출된 디스플레이(140)의 적어도 일부를 의미할 수 있다. 디스플레이(140)의 적어도 일부는 플렉서블 디스플레이(flexible display)의 형태로 전자 장치(100)의 전면 영역 및, 측면 영역 및 후면 영역 중 적어도 하나에 결합될 수도 있다. 플렉서블 디스플레이는 종이처럼 얇고 유연한 기판을 통해 손상 없이 휘거나 구부리거나 말 수 있는 것을 특징으로 할 수 있다.
- [0180] 마이크(150)는 제1 및 제2 사용자 음성을 획득할 수 있다. 구체적으로, 마이크(150)는 마이크로폰(Microphone)을 이용하여 외부의 사용자 음성을 수신하고, 이를 전기적인 음성 데이터로 처리할 수 있다. 이 경우, 마이크(150)는 처리된 음성 데이터를 프로세서(120)에 전달할 수 있다. 마이크(150)는 전자 장치(100) 내부에 구비될 수 있으나, 이는 일 실시예에 불과할 뿐, 전자 장치(100)와 전기적으로 연결되어 외부에 구비될 수 있다. 오디오 출력부(160)는 오디오 처리부(미도시)에 의해 디코딩이나 증폭, 노이즈 필터링과 같은 다양한 처리 작업이 수행된 각종 오디오 데이터뿐만 아니라 각종 알람 음이나 음성 메시지를 출력하는 구성이다. 특히, 오디오를 출력하는 구성은 스피커(160)로 구현될 수 있으나, 이는 일 실시예에 불과할 뿐, 오디오 데이터를 출력할 수 있는 출력 단자로 구현될 수 있다.

- [0181] 사용자 입력부(170)는 다양한 사용자 입력을 수신하여 프로세서(120)로 전달할 수 있다. 사용자 입력부(170)는, 예를 들면, 터치 패널, (디지털) 펜 센서 또는 키를 포함할 수 있다. 터치 패널은, 예를 들면, 정전식, 감압식, 적외선 방식, 또는 초음파 방식 중 적어도 하나의 방식을 사용할 수 있다. 또한, 터치 패널은 제어 회로를 더 포함할 수도 있다. 터치 패널은 택타일 레이어(tactile layer)를 더 포함하여, 사용자에게 촉각 반응을 제공할 수 있다. (디지털) 펜 센서는, 예를 들면, 터치 패널의 일부이거나, 별도의 인식용 슈트를 포함할 수 있다. 키는, 예를 들면, 물리적인 버튼, 광학식 키, 또는 키패드를 포함할 수 있다.
- [0182] 프로세서(120)(또는, 제어부)는 메모리(1140)에 저장된 각종 프로그램을 이용하여 전자 장치(100)의 전반적인 동작을 제어할 수 있다.
- [0183] 프로세서(1170)는 RAM(미도시), ROM(미도시), 그래픽 처리부(미도시), 메인 CPU(미도시), 제1 내지 n 인터페이스(미도시), 버스(미도시)로 구성될 수 있다. 이때, RAM(미도시), ROM(미도시), 그래픽 처리부(미도시), 메인 CPU(미도시), 제1 내지 n 인터페이스(미도시) 등은 버스(미도시)를 통해 서로 연결될 수 있다.
- [0184] 한편, 전자 장치(100)의 구현 예에 따라 도 3에 도시된 전자 장치(100)의 구성요소 중 적어도 일부는 생략될 수 있다.
- [0185] 예를 들어, 전자 장치(100)가 도 1a 및 도 1b에 도시된 바와 같이 서버로 구현되는 경우, 외부 전자 장치에서 사용자 음성을 수신하고, 사용자 음성에 대한 응답 역시 외부 전자 장치에서 출력된다는 점에서, 도 3의 디스플레이(140), 마이크(150) 및 오디오 출력부(160)는 생략될 수 있다.
- [0186] 도 6은 본 개시의 일 실시예에 따른, 전자 장치의 제어 방법을 설명하기 위한 흐름도이다.
- [0187] 우선, 전자 장치(100)는 제1 사용자 음성의 음향 특징 및 문맥 정보(context information)를 포함하는 화자 모델을 저장한다(S610).
- [0188] 그리고, 전자 장치(100)는 화자 모델에 포함된 제1 사용자 음성의 음향 특징과 제2 사용자 음성의 음향 특징간 유사도를, 화자 모델에 포함된 문맥 정보와 제2 사용자 음성의 문맥 정보간 유사도에 따라 변경되는 임계 값과 비교한다(S620). 그 후, 제2 사용자 음성에 대한 인증을 수행한다(S630).
- [0189] 여기에서, 임계값은 제2 사용자 음성의 문맥 정보와 화자 모델에 포함된 문맥 정보 사이의 유사도에 비례하도록 결정될 수 있다.
- [0190] 한편, 전자 장치(100)는 제2 사용자 음성의 문맥 정보와 화자 모델에 포함된 문맥 정보 사이의 유사도에 기초하여 임계 값을 산출하고, 산출된 임계 값을, 제2 사용자 음성의 음향 특징과 화자 모델에 저장된 제1 사용자 음성의 음향 특징 사이의 유사도와 비교할 수 있다.
- [0191] 이때, 전자 장치(100)는 제1 사용자 음성의 문맥 정보와 제2 사용자 음성의 문맥 정보가 동일한 경우 제1 및 제2 음향 특징 사이의 유사도가 가질 수 있는 최대 값에 기초하여 임계 값을 산출할 수 있다.
- [0192] 그리고, 전자 장치(100)는 제2 사용자 음성의 음향 특징과 화자 모델에 저장된 제1 사용자 음성의 음향 특징 사이의 유사도가 임계 값 이상인 경우, 제2 사용자 음성에 대한 인증이 성공된 것으로 판단할 수 있다.
- [0193] 한편, 전자 장치(100)는 제1 사용자 음성이 수신되면, 제1 사용자 음성으로부터 제1 사용자 음성의 음향 특징 및 문맥 정보를 획득하고, 획득된 음향 특징 및 문맥 정보를 포함하는 화자 모델을 저장할 수 있다.
- [0194] 이때, 전자 장치(100)는 기정의된 복수의 문맥 정보 타입(type) 중 하나의 문맥 정보 타입을 선택할 수 있다.
- [0195] 구체적으로, 전자 장치(100)는 제1 사용자 음성의 음향 특징과 복수의 테스트 음성 각각의 음향 특징 사이의 유사도를 판단하고, 복수의 문맥 정보의 타입 별로, 제1 사용자 음성의 문맥 정보와 복수의 테스트 음성 각각의 문맥 정보 사이의 유사도를 판단할 수 있으며, 복수의 문맥 정보 타입 중, 판단된 문맥 정보 사이의 유사도에 대한 판단된 음향 특징의 유사도의 분포가 기설정된 조건을 만족하는 문맥 정보 타입을 선택할 수 있다.
- [0196] 그리고, 전자 장치(100)는 선택된 문맥 정보 타입을 이용하여 제1 사용자 음성의 문맥 정보를 획득할 수 있다.
- [0197] 또한, 전자 장치(100)는 제2 사용자 음성을 분석하고, 선택된 문맥 정보 타입을 이용하여 제2 사용자 음성의 문맥 정보를 획득하고, 제2 사용자 음성의 문맥 정보와 화자 모델에 포함된 문맥 정보 사이의 유사도에 기초하여 임계 값을 결정할 수 있다.
- [0198] 한편, 상술한 바와 같은 제어 방법은 컴퓨터에서 실행될 수 있는 실행가능한 알고리즘을 포함하는 프로그램으로

구현될 수 있고, 프로그램은 비일시적 판독 가능 매체(non-transitory computer readable medium)에 저장되어 제공될 수 있다.

[0199] 비일시적 판독 가능 매체란 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 구체적으로는, 상술한 다양한 애플리케이션 또는 프로그램들은 CD, DVD, 하드 디스크, 블루레이 디스크, USB, 메모리카드, ROM 등과 같은 비일시적 판독 가능 매체에 저장되어 제공될 수 있다.

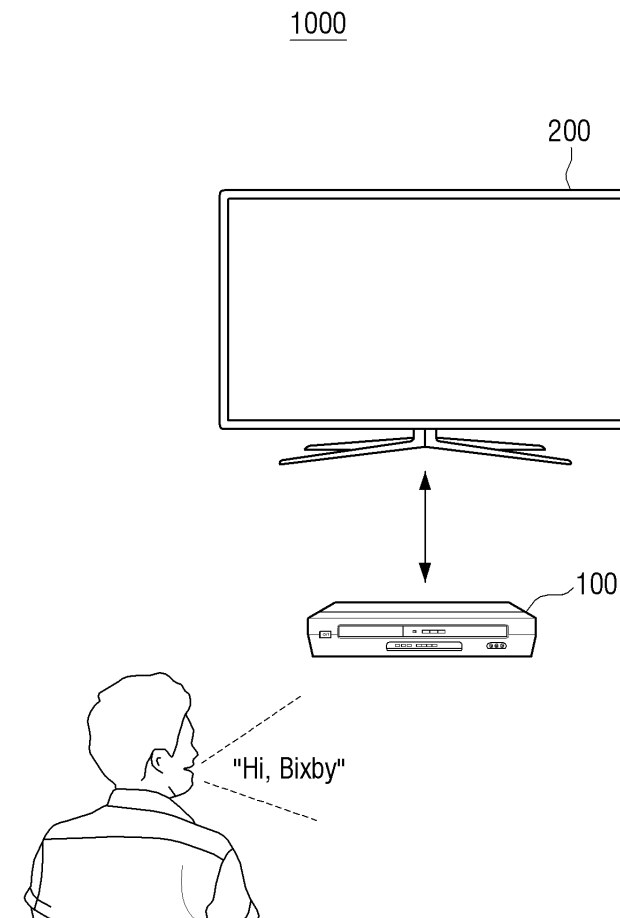
[0200] 이상에서는 본 개시의 바람직한 실시예에 대해서 도시하고, 설명하였으나, 본 개시는 상술한 특정의 실시예에 한정되지 아니하며, 청구범위에서 청구하는 본 개시의 요지를 벗어남이 없이 당해 개시가 속하는 기술분야에서 통상의 지식을 가진 자라면 누구든지 다양한 변형 실시가 가능한 것은 물론이고, 그와 같은 변경은 청구범위 기재의 범위 내에 있게 된다.

부호의 설명

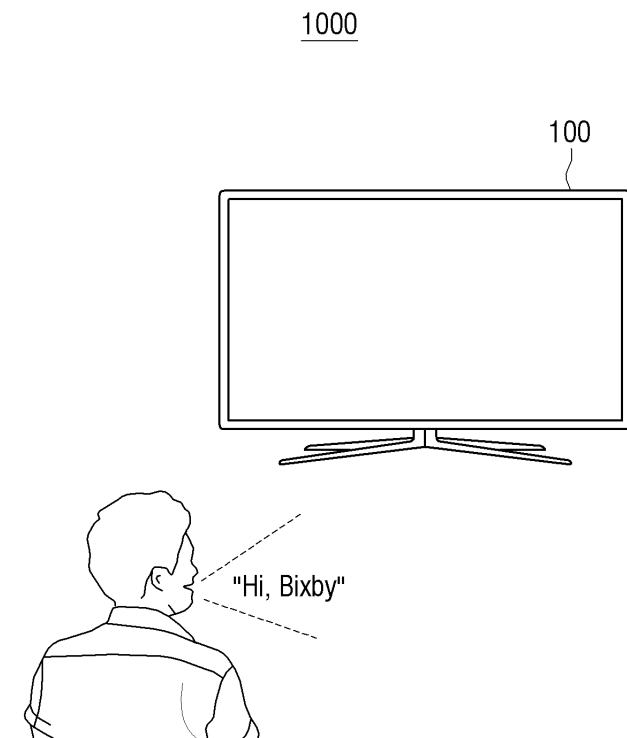
[0203] 100: 전자 장치 110: 메모리
120: 프로세서

도면

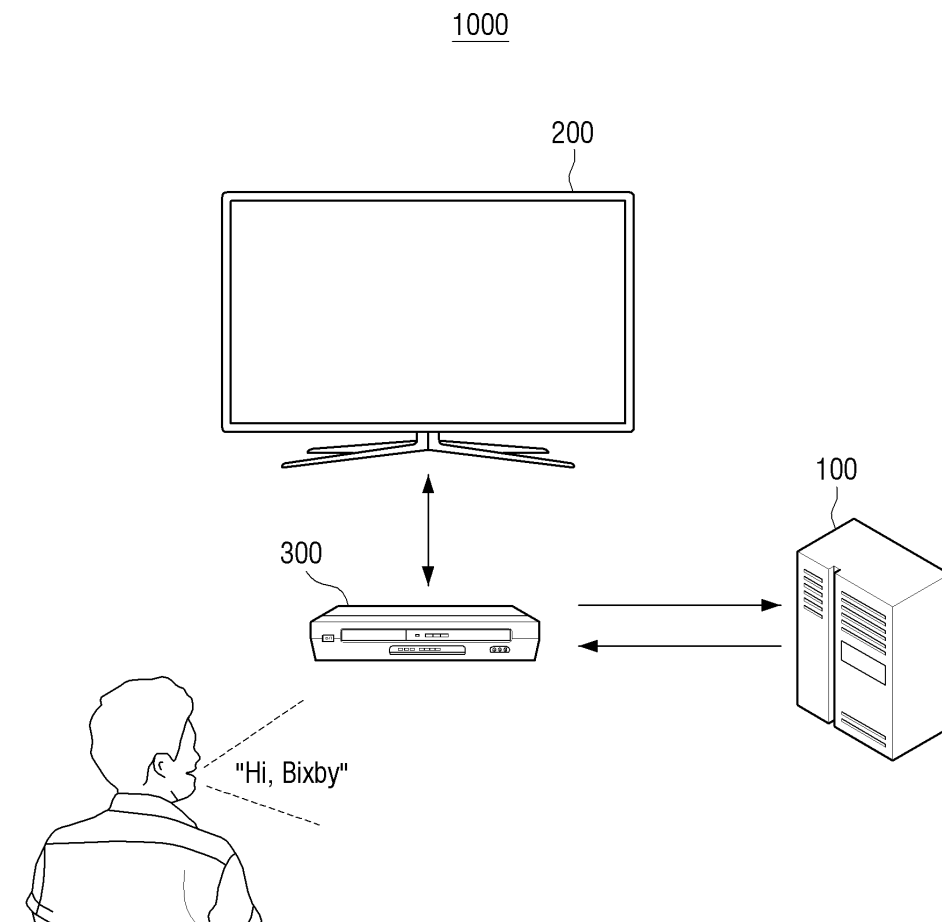
도면1a



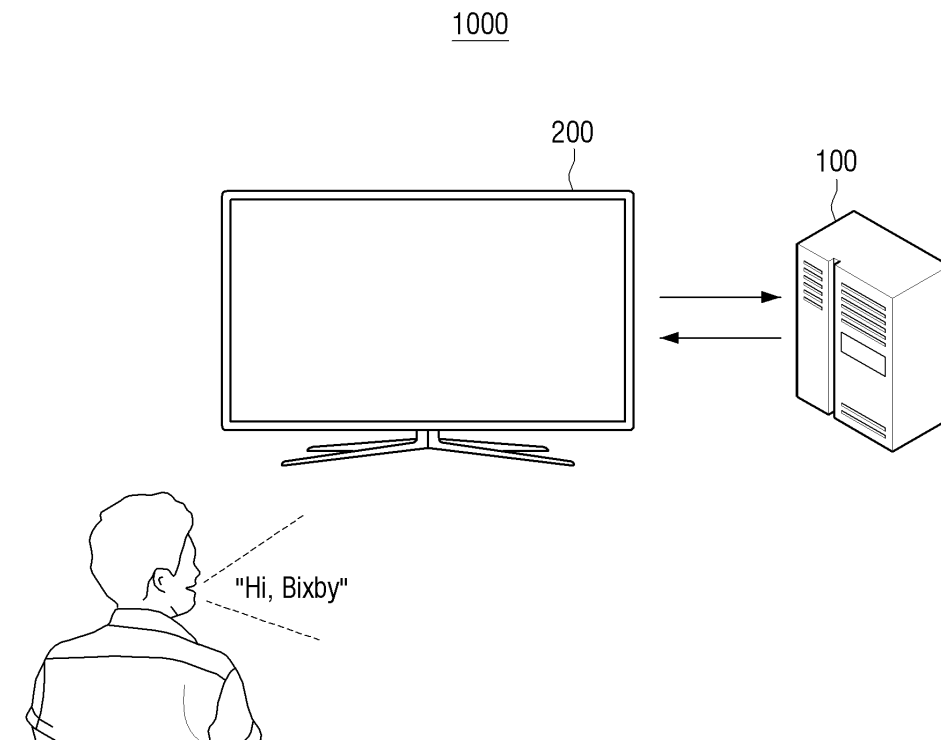
도면1b



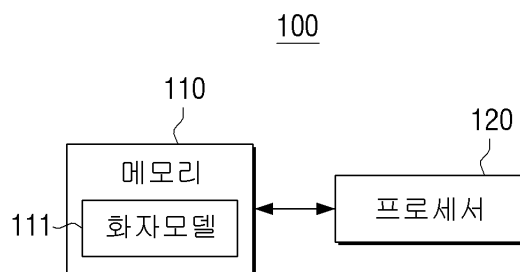
도면1c



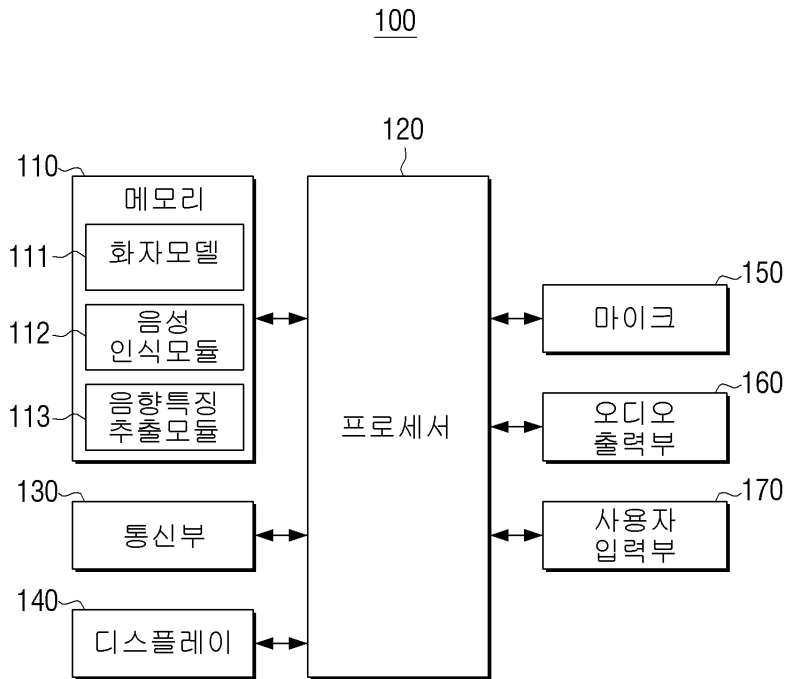
도면1d



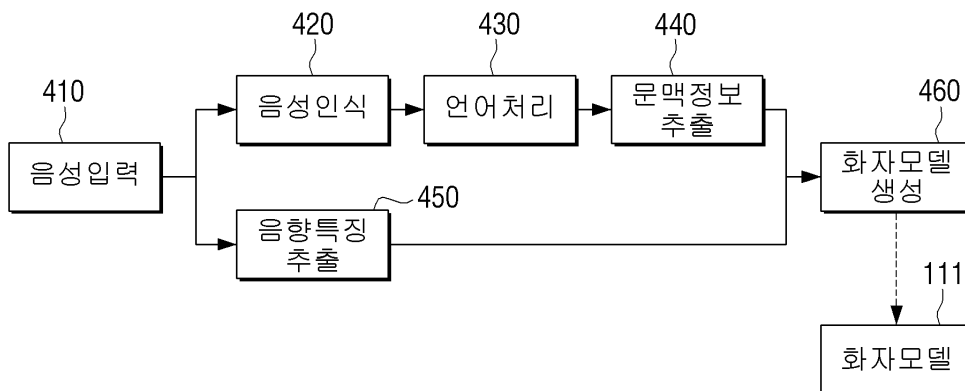
도면2



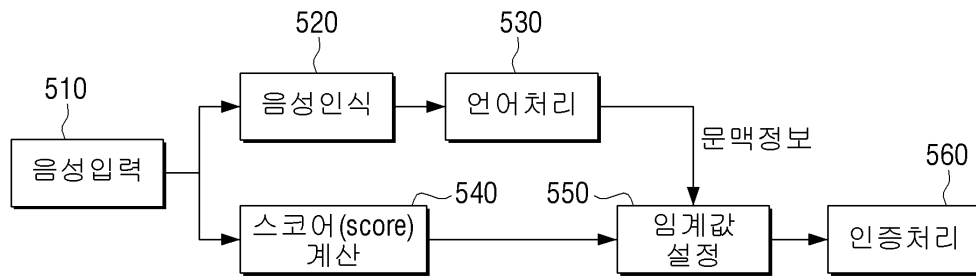
도면3



도면4



도면5



도면6

