



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0032935
(43) 공개일자 2020년03월27일

(51) 국제특허분류(Int. Cl.)

G10L 17/14 (2013.01) G06F 21/32 (2013.01)
G10L 25/12 (2013.01) G10L 25/24 (2013.01)
G10L 25/51 (2013.01) G10L 25/90 (2013.01)

(52) CPC특허분류

G10L 17/14 (2013.01)
G06F 21/32 (2013.01)

(21) 출원번호 10-2018-0112204

(22) 출원일자 2018년09월19일

심사청구일자 2018년09월19일

(71) 출원인

주식회사 공훈

경기도 의정부시 신흥로 234번길, 7층710호(CRC빌딩)

(72) 발명자

이태훈

경기도 의정부시 동일로416번길 31, 102동 503호
(장암동, 삼환나우빌)

송유중

경기도 의정부시 부용로 174, 108동 1605호(신곡동, 상록아이파크)

(74) 대리인

특허법인 천지

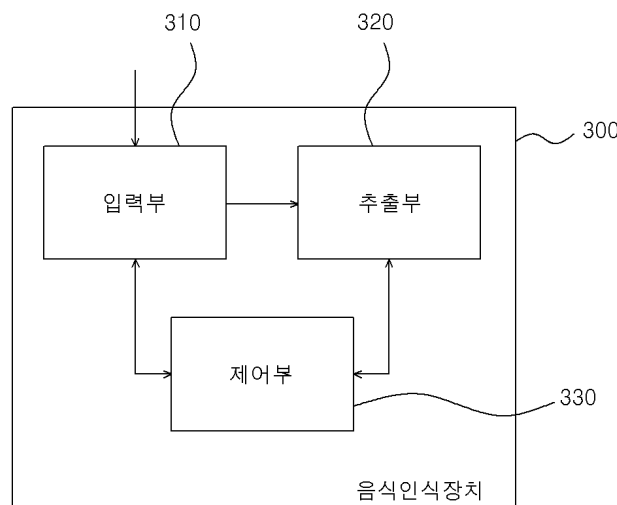
전체 청구항 수 : 총 19 항

(54) 발명의 명칭 음성인식장치 및 음성인식방법

(57) 요약

본 발명의 일 실시 예에 의한 음성인식장치는, 음성을 입력 받는 입력부; 와 상기 음성을 음소 단위로 분할하여 각 음소별로 음성특성을 추출하는 추출부; 및 상기 음성특성을 참조음성특성과 비교하고 유사도에 기초하여 상기 음성을 인증하는 제어부를 포함하되, 상기 추출부는, 사전에 입력된 복수개의 참조음성을 발화 상태를 반영한 복수개의 그룹으로 분류하고 상기 복수개의 그룹 각각으로부터 상기 참조음성특성을 추출한다.

대 표 도 - 도3



(52) CPC특허분류

G10L 17/02 (2013.01)

G10L 25/12 (2013.01)

G10L 25/15 (2013.01)

G10L 25/24 (2013.01)

G10L 25/51 (2013.01)

G10L 25/90 (2013.01)

명세서

청구범위

청구항 1

음성인식장치에 있어서,
음성을 입력 받는 입력부;
상기 음성을 음소 단위로 분할하여 각 음소별로 음성특성을 추출하는 추출부; 및
상기 음성특성을 참조음성특성과 비교하고 유사도에 기초하여 상기 음성을 인증하는 제어부를 포함하되,
상기 추출부는,
사전에 입력된 복수개의 참조음성을 발화 상태를 반영한 복수개의 그룹으로 분류하고 상기 복수개의 그룹 각각으로부터 상기 참조음성특성을 추출하는 음성인식장치.

청구항 2

제1항에 있어서,
상기 발화 상태는,
신체 상태, 감정 상태 및 주변 환경 중 적어도 하나를 포함하는 음성인식장치.

청구항 3

제1항에 있어서,
상기 추출부는,
상기 복수개의 참조음성 각각에 대하여 상기 발화 상태와 관련된 특징 벡터를 추출하고, 추출된 상기 특징 벡터의 유사도에 기초하여 상기 복수개의 참조음성을 그룹화하여 상기 복수개의 그룹으로 분류하고, 상기 복수개의 그룹 각각에 대하여 상기 참조음성특성을 추출하는 음성인식장치.

청구항 4

제1항에 있어서,
상기 음성특성과 상기 참조음성특성 각각은,
음성 주파수, 피치(pitch), 포먼트(formant), 발화시간 및 발화속도 중 적어도 하나를 포함하는 음성인식장치.

청구항 5

제1항에 있어서,
상기 추출부는,
상기 음성특성 및 상기 참조음성특성 중 적어도 하나를 추출하는 경우, 상기 복수개의 그룹 각각에 대하여 서로 다른 추출방식을 적용하는 음성인식장치.

청구항 6

제5항에 있어서,
상기 추출방식은,
PLP(Perceptual Linear Predictive Analysis), LPC(Linear Predictive Coding), MFCC(Mel-Frequency Cepstrum Coefficients) 및 켈스트럼(Cepstrum) 중 적어도 하나를 포함하는 음성인식장치.

청구항 7

제1항에 있어서,

상기 제어부는,

상기 음성이 특정화자가 특정단어를 발화한 복수개의 제1음성을 포함하는 경우, 상기 제1음성의 상기 음성특성을 상기 참조음성특성과 비교하여, 소정 값 이상의 유사도가 소정 횟수 이상이면 상기 참조음성특성을 상기 특정화자의 상기 특정단어에 대한 발화 패턴으로 결정하는 음성인식장치.

청구항 8

제1항에 있어서,

상기 제어부는,

상기 음성에 포함되는 단어 및 상기 단어의 상기 음성특성을 추출하여 데이터베이스에 포함되는 적어도 하나의 참조 단어와 비교하되, 상기 음성과 관련된 데이터를 기준으로 난수를 생성하고 소정 기준에 따라 상기 데이터베이스 상에서 상기 난수에 대응하는 상기 참조 단어를 검출하는 음성인식장치.

청구항 9

제8항에 있어서,

상기 음성과 관련된 데이터는 상기 음성의 인증이 요청된 시간이고,

상기 제어부는,

상기 음성의 인증이 요청된 시간을 기준으로 상기 난수를 생성하고, 소정 행렬 구조로 구성되는 상기 데이터베이스 상에서 상기 난수의 소정 자리 숫자에 해당하는 행과 열에 해당하는 상기 참조 단어를 검출하는 음성인식장치.

청구항 10

음성인식방법에 있어서,

음성을 입력 받는 단계;

상기 음성을 음소 단위로 분할하여 각 음소별로 음성특성을 추출하는 단계; 및

상기 음성특성을 참조음성특성과 비교하고 유사도에 기초하여 상기 음성을 인증하는 단계를 포함하되,

상기 참조음성특성은,

사전에 입력된 복수개의 참조음성을 발화 상태를 반영한 복수개의 그룹으로 분류하고 상기 복수개의 그룹 각각으로부터 추출되는 음성인식방법.

청구항 11

제10항에 있어서,

상기 발화 상태는,

신체 상태, 감정 상태 및 주변 환경 중 적어도 하나를 포함하는 음성인식방법.

청구항 12

제10항에 있어서,

상기 복수개의 참조음성 각각에 대하여 상기 발화 상태와 관련된 특징 벡터를 추출하고, 추출된 상기 특징 벡터의 유사도에 기초하여 상기 복수개의 참조음성을 그룹화하여 상기 복수개의 그룹으로 분류하고, 상기 복수개의 그룹 각각에 대하여 상기 참조음성특성을 추출하는 음성인식방법.

청구항 13

제10항에 있어서,

상기 음성특성과 상기 참조음성특성 각각은,

음성 주파수, 피치(pitch), 포먼트(formant), 발화시간 및 발화속도 중 적어도 하나를 포함하는 음성인식방법.

청구항 14

제10항에 있어서,

상기 음성특성 및 상기 참조음성특성 중 적어도 하나를 추출하는 경우, 상기 복수개의 그룹 각각에 대하여 서로 다른 추출방식을 적용하는 음성인식방법.

청구항 15

제14항에 있어서,

상기 추출방식은,

PLP(Perceptual Linear Predictive Analysis), LPC(Linear Predictive Coding), MFCC(Mel-Frequency Cepstrum Coefficients) 및 켈스트럼(Cepstrum) 중 적어도 하나를 포함하는 음성인식방법.

청구항 16

제10항에 있어서,

상기 음성이 특정화자가 특정단어를 발화한 복수개의 제1음성을 포함하는 경우, 상기 제1음성의 상기 음성특성을 상기 참조음성특성과 비교하여, 소정 값 이상의 유사도가 소정 횟수 이상이면 상기 참조음성특성을 상기 특정화자의 상기 특정단어에 대한 발화 패턴으로 결정하는 음성인식방법.

청구항 17

제10항에 있어서,

상기 음성에 포함되는 단어 및 상기 단어의 상기 음성특성을 추출하여 데이터베이스에 포함되는 적어도 하나의 참조 단어와 비교하되, 상기 음성과 관련된 데이터를 기준으로 난수를 생성하고 소정 기준에 따라 상기 데이터베이스 상에서 상기 난수에 대응하는 상기 참조 단어를 검출하는 음성인식방법.

청구항 18

제17항에 있어서,

상기 음성과 관련된 데이터는 상기 음성의 인증이 요청된 시간이고,

상기 음성의 인증이 요청된 시간을 기준으로 상기 난수를 생성하고, 소정 행렬 구조로 구성되는 상기 데이터베이스 상에서 상기 난수의 소정 자리 숫자에 해당하는 행과 열에 해당하는 상기 참조 단어를 검출하는 음성인식방법.

청구항 19

제10항 내지 제18항 중 어느 한 항의 방법을 구현하기 위한 프로그램이 기록된 컴퓨터로 판독 가능한 기록매체.

발명의 설명

기술 분야

본 발명은 음성인식장치 및 음성인식방법에 대한 것으로서, 보다 구체적으로 발화자의 발화 상태를 고려하여 화자를 식별하되, 발화된 단어를 음소별로 분할하여 음성 특성을 추출하고 이에 기초하여 화자를 식별하는 음성인식장치 및 음성인식방법에 관한 것이다.

배경 기술

[0001]

[0002] 최근 사용자 인증 시, 생체인식이 많이 사용되고 있다. 다양한 생체인식 중 음성인식은 기계와 사람의 인터페이스로 최적의 조건을 갖추고 있어 적용범위가 점차 넓어지고 있는 추세이다.

[0003] 그러나, 음성인식은 제3자의 모방과 녹취 및 재생에 취약하다는 문제점이 있다. 또한, 기존의 음성인식은 사용자가 발화한 모든 음성을 기준으로 공통적인 특징을 추출하고 이에 기초하여 사용자를 인식하기 때문에, 발화 상태에 따라 달라지는 음성을 정확하게 인식할 수 없다는 한계가 존재한다. 따라서, 발화자의 발음 상태나 시간 등에 따라 음성 인식률이 변화할 수 있어, 음성인식의 인식 및 인증에 있어서 정확성이 보완되어야 한다.

발명의 내용

해결하려는 과제

[0004] 본 발명은 화자의 발화 상태를 고려하여 음성 인식을 수행함으로써, 화자의 발화상태나 시간적 및 환경적 요인에 관계없이 안정적으로 화자를 식별하고 화자 식별에 대한 정확성을 높일 수 있는 음성인식장치 및 음성인식방법을 제공하는 것을 목적으로 한다.

[0005] 또한, 본 발명은 발화자의 발화상태를 고려하여 특정 단어에 대하여 발화자의 최적화된 발화패턴을 결정할 수 있는 음성인식장치 및 음성인식방법을 제공하는 것을 목적으로 한다.

[0006] 나아가, 본 발명은 발화자의 발화상태에 따른 음성특성을 지속적으로 모니터링하여 화자 개인별 특정 단어에 대한 발화 패턴을 지속적으로 업데이트함으로써, 화자의 현재 상태를 반영하여 정확하게 화자를 식별할 수 있는 음성인식장치 및 음성인식방법을 제공하는 것을 목적으로 한다.

[0007] 본 발명에서 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재에 의해 제안되는 실시 예들이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0008] 본 발명의 일 실시 예에 의한 음성인식장치에 의하면, 음성을 입력 받는 입력부; 와 상기 음성을 음소 단위로 분할하여 각 음소별로 음성특성을 추출하는 추출부; 및 상기 음성특성을 참조음성특성과 비교하고 유사도에 기초하여 상기 음성을 인증하는 제어부를 포함하되, 상기 추출부는, 사전에 입력된 복수개의 참조음성의 발화 상태를 반영한 복수개의 그룹으로 분류하고 상기 복수개의 그룹 각각으로부터 상기 참조음성특성을 추출한다.

[0009] 본 발명의 다른 실시 예에 의한 음성인식방법에 의하면, 음성을 입력 받는 단계; 와 상기 음성을 음소 단위로 분할하여 각 음소별로 음성특성을 추출하는 단계; 및 상기 음성특성을 참조음성특성과 비교하고 유사도에 기초하여 상기 음성을 인증하는 단계를 포함하되, 상기 참조음성특성은, 사전에 입력된 복수개의 참조음성을 발화 상태를 반영한 복수개의 그룹으로 분류하고 상기 복수개의 그룹 각각으로부터 추출된다.

발명의 효과

[0010] 본 발명에 따른 실시 예들에 의하면, 화자의 발화 상태를 고려하여 음성 인식을 수행함으로써, 화자의 발화상태나 시간적 및 환경적 요인에 관계없이 안정적으로 화자를 식별하고 화자 식별에 대한 정확성을 높일 수 있다.

[0011] 또한, 본 발명에 따른 실시 예들에 의하면, 발화자의 발화상태를 고려하여 특정 단어에 대하여 발화자의 최적화된 발화패턴을 결정할 수 있다.

[0012] 나아가, 본 발명에 따른 실시 예들에 의하면, 발화자의 발화상태에 따른 음성특성을 지속적으로 모니터링하여 화자 개인별 특정 단어에 대한 발화 패턴을 지속적으로 업데이트함으로써, 화자의 현재 상태를 반영하여 정확하게 화자를 식별할 수 있다.

도면의 간단한 설명

[0013] 도 1은 일반적인 화자 확인 방법을 설명하기 위한 도면이다.

도 2는 일반적인 화자 확인 시, 음성에 대한 평준화 방법을 설명하기 위한 도면이다.

도 3은 본 발명의 일 실시 예에 의한 음성인식장치의 구성을 도시한 블록도이다.

도 4는 본 발명의 일 실시 예에 의한 음성인식장치에 포함되는 제어부의 상세 구성을 도시한 블록도이다.

도 5는 본 발명의 일 실시 예에 의한 음성인식 과정을 도시한 도면이다.

도 6은 본 발명의 일 실시 예에 의한 음성인식방법과 관련하여 화자에 대한 발화 패턴을 결정하는 과정을 도시한 도면이다.

도 7은 본 발명의 일 실시 예에 의한 음성인식방법과 관련하여 화자에 대한 발화 패턴을 결정하는 상세 과정을 도시한 도면이다.

도 8은 본 발명의 일 실시 예에 의한 음성인식방법을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0014] 이하에서는 본 발명의 구체적인 실시예를 도면을 참조하여 상세히 설명한다. 그러나 본 발명의 기술적 사상이 이하에서 기술되는 실시예들에 의하여 제한되는 것은 아니며, 또 다른 구성요소의 추가, 변경 및 삭제 등에 의해서 퇴보적인 다른 발명이나 본 발명의 기술적 사상의 범위 내에 포함되는 다른 실시예들을 용이하게 제안할 수 있다.
- [0015] 본 발명에서 사용되는 용어는 가능한 한 현재 해당 기술과 관련하여 널리 사용되는 일반적인 용어를 선택하였으나, 특별한 경우에는 출원인이 임의로 선정한 용어도 있으며 이 경우 해당되는 발명의 설명 부분에서 그 의미를 상세히 기재하였다. 그러므로, 단순한 용어의 명칭이 아닌 용어가 가지는 의미로서 본 발명을 파악하여야 함을 미리 밝혀둔다. 이하에서 기술하는 설명에 있어서, 단어 '포함하는'은 열거된 것과 다른 구성요소들 또는 단계들의 존재를 배제하지 않는다.
- [0016] 도 1은 일반적인 화자 확인 방법을 설명하기 위한 도면이다.
- [0017] 도 1에서 확인 대상은 특정 화자 A이다. 이 경우, 특정 화자 A에 대한 음성 모델을 설정하고, 이후 불특정 화자 B의 음성이 입력되면, B가 A와 동일인인지 여부를 판단하여 화자를 식별할 수 있다.
- [0018] 이를 위해, 먼저, 확인 대상인 특정 화자 A에 대한 음성 모델 수립 과정을 통하여 특정 화자 A의 음성 특징을 사전에 데이터베이스화 할 수 있다. 구체적으로, 도 1에 도시된 바와 같이, 특정 화자 A로부터 복수의 음성 샘플을 획득하여, 각각의 음성에 대한 주파수, 피치 등과 같은 음성 특성 파라미터를 추출한 후, 중첩되는 부분을 기준으로 음성에 대한 평준화를 진행한다. 평준화 이후 특정 화자 A에 대한 음성 모델을 수립한다.
- [0019] 이후, 불특정 화자 B의 음성이 입력되면, 입력된 음성으로부터 음성 특성 파라미터를 추출하고, 이를 기초로 형성된 데이터를 확인 대상인 특정 화자 A의 음성 모델과 비교한다. 이 경우, 유사도가 기 설정된 임계값 이상이면, 불특정 화자는 특정 화자와 동일인인 것으로 판단한다.
- [0020] 일반적인 화자 확인은, 사용자가 지정한 문장 또는 단어를 사용하는 단어(이하, 문맥이라고도 함) 고정형 화자 확인과 사용자의 발음 내용에 제한이 없는 단어 자유형 화자 확인으로 분류될 수 있다. 단어 고정형 화자 확인의 경우, 효율성은 좋으나, 정해진 단어의 노출 위험과 사용자를 사칭한 녹음 등의 불법적인 방법이 사용될 수 있어 보안성에 취약하다. 반면, 단어 자유형 화자 확인의 경우, 화자 확인을 위해 많은 훈련 데이터가 필요하여 시간 및 자원 활용도의 측면에서 효율성이 낮다. 이러한 단어 고정형 화자 확인과 단어 자유형 화자 확인의 장점은 살리고 단점은 보완하고자, 단어 제시형 화자 확인이 제안되었다.
- [0021] 단어 제시형 화자 확인은, 화자 확인이 필요한 경우 시스템에서 화자에게 매번 다른 단어나 문장을 발음할 것을 요구하고, 발음된 단어나 문장에 대한 음성 인식을 수행하여 요구된 단어 또는 문장과 일치되는지를 1차적으로 확인한 후, 발음된 단어 또는 문장의 발음정보에서 화자 고유의 음성 특성값을 추출하여 사전 정의된 화자의 음성 특성값과 비교하여 화자를 확인한다.
- [0022] 이러한 절차에 의하여, 사용자가 사전에 지정한 단어 또는 문장을 기억하지 못하거나 사용자를 사칭하는 녹음 등의 위험이 줄어들고, 성능면에서는 단어 고정형 화자 확인과 동일한 효율을 확보할 수 있다.
- [0023] 그러나, 단어 제시형 화자 확인의 경우, 임의로 단어를 생성하는 과정이 화자 음성 모델을 기반으로 하기 때문에 화자의 음성의 원천적인 입력과는 근본적인 차이가 있을 수 있고, 음성 모델을 형성하는 과정에서 평준화 오류가 발생할 수도 있다.
- [0024] 도 2는 일반적인 화자 확인 시, 음성에 대한 평준화 방법을 설명하기 위한 도면이다.
- [0025] 사용자의 음성은 연속적 과정을 가지며, 샘플링 과정을 거쳐 디지털화될 수 있다. 일반적으로 화자 확인을 수행하는 장치는 화자 확인(식별 또는 인증)을 하기 위한 준거 데이터를 생성하는 경우, 다수의 음성 데이터를 샘플

링하여 디지털화하고 이에 대해 공통된 데이터를 추출함으로써 평준화된 데이터(210)를 생성한다. 이렇게 평준화된 데이터(210)를 기준으로 LPC(linear predictive coding), MFCC(Mel-Frequency Cepstral Coefficients) 등을 사용하여 음성에 대한 특성값을 추출한 후 이를 기초로 화자 확인을 위한 준거 데이터를 생성한다.

[0026] 그러나, 화자가 발화하는 음성은 발화자의 상태나 환경에 따라 변화할 수 있다. 즉, 화자가 동일한 단어를 발화하더라도, 화자의 감정이나 건강상태(예를 들어, 감기나 발병 등), 시간(예를 들어, 발화자의 노화 등), 발화시의 주변 환경(예를 들어, 소음 등) 등에 의해 화자의 음성 특성(예를 들어, 주파수나 피치 등의 음성 톤(tone))이 일시적으로 또는 일정 기간 동안 변화될 수 있다. 따라서, 도 2에 도시된 바와 같이 평준화된 데이터를 기준으로 음성 모델을 수립하게 되면, 발화자의 상태나 생활 환경 등에 따른 일상적인 특성값을 왜곡하여 정확한 화자 확인을 하지 못하게 된다.

[0027] 도 3은 본 발명의 일 실시 예에 의한 음성인식장치의 구성을 도시한 블록도이다.

[0028] 본 발명의 일 실시 예에 의한 음성인식장치(300)는 입력부(310)와 추출부(320) 및 제어부(330)를 포함할 수 있다.

[0029] 입력부(310)는 화자가 발화한 음성을 입력 받을 수 있다. 여기서, 화자는 특정 화자 및 불특정 화자를 포함할 수 있다. 음성은 음성인식장치(300)에 구비된 입력 수단(예를 들어, 음성 마이크 등)에 의해 직접 입력되거나, 유무선 통신에 의해 다른 장치로부터 입력될 수 있다.

[0030] 추출부(320)는 음성에 포함된 단어 및 단어에 대한 음성 특성을 추출할 수 있다. 여기서, 음성 특성은 음성의 주파수, 피치(pitch), 포먼트(formant), 발화시간 및 발화속도 중 적어도 하나를 포함할 수 있다.

[0031] 피치(pitch)는 음의 높이이다.

[0032] 포먼트(formant)는 음성에 존재하는 복수개의 공진 대역일 수 있다. 구체적으로, 음성(유성음)은 성대 진동의 기본 주파수 성분과 그 고조파 성분으로 구성된다. 진동 근원체는 모두 특유한 진동 특성(예를 들어, 공진특성)을 가지고 있다. 사람의 조음 기관(예를 들어, 성대 등)도 조음에 따라 변하는 순간에서의 공진 특성이 있으며, 성대파가 이러한 공진 특성에 따라 여파되어 소리로서 표현될 수 있다. 특정음(예를 들어, 모음)의 주파수 스펙트럼을 살펴보면 공진 특성 발현 시, 그 공진대역이 복수 개 이상 존재함을 확인할 수 있다. 이러한 복수개의 공진 주파수대역을 포먼트로 정의한다.

[0033] 일 실시 예에 의하면, 추출부(320)는 음성을 음소 단위로 분할하고, 각 음소 별로 음성 특성을 추출할 수 있다. 구체적으로, 추출부(320)는 발화 음성으로부터 단어나 음절 등을 추출하고, 추출된 단어나 음절 등을 음소 단위로 분할할 수 있다. 여기서, 음소는 더 이상 작게 나눌 수 없는, 의미의 구별을 나타내는 음운론상의 최소 단위로 정의될 수 있다. 예를 들어, 발화자가 발화한 "출입문 개방"이라는 문맥이 입력되는 경우, 추출부(320)는 문맥으로부터 "출입문"이라는 단어를 추출하고, 이를 ㅈ, ㅌ, ㄹ, ㅇ, ㅣ, ㅂ, ㅍ, ㅊ, ㄴ 으로 분할하고 이로부터 음성 특성을 추출할 수 있다. 이에 의하면, 본 발명은 음소로부터 음성 특성을 추출함으로써, 음성 인식을 수행하기 위한 음성 특성을 문맥을 구성하는 최소 단위로부터 추출하게 되어, 보다 정확성을 음성 인식을 수행할 수 있다.

[0034] 추출부(320)는 발화자의 발화 상태를 고려하여 참조음성특성을 추출할 수 있다. 여기서, 참조음성특성은 입력된 음성을 인증하기 위하여 사전에 데이터베이스화 된 음성특성 일 수 있다. 참조음성특성은, 음성 주파수, 피치(pitch), 포먼트(formant), 발화시간 및 발화속도 중 적어도 하나를 포함할 수 있다. 발화 상태는, 발화자의 신체 상태(아플 때, 정상일 때 등), 감정 상태(기쁠 때, 슬플 때, 우울할 때, 화났을 때 등), 시간 변화(발화자의 나이, 발화자의 노화 등) 및 주변 환경(조용할 때, 소음이 많을 때 등) 중 적어도 하나를 포함할 수 있다.

[0035] 구체적으로, 추출부(320)는 사전에 입력된 복수개의 참조음성을 발화 상태를 반영한 복수개의 그룹으로 분류하고, 복수개의 그룹 각각으로부터 참조음성특성을 추출할 수 있다. 여기서, 참조음성은, 음성 인증을 위한 데이터베이스를 구축하기 위하여 사전에 입력 및 학습되는 음성일 수 있다

[0036] 일 실시 예에 의하면, 추출부(320)는 복수개의 참조음성 각각에 대하여 발화 상태와 관련된 특징 벡터를 추출하고, 추출된 특징 벡터의 유사도에 기초하여 복수개의 참조음성을 그룹화하여 복수개의 그룹으로 분류한 후, 복수개의 그룹 각각에 대하여 참조음성특성을 추출할 수 있다.

[0037] 이 경우, 추출부(320)는 음성특성 및 참조음성특성 중 적어도 하나를 추출하는 경우, 복수개의 그룹 각각에 대하여 서로 다른 추출방식을 적용할 수 있다. 여기서, 추출방식은, PLP(Perceptual Linear Predictive Analysis), LPC(Linear Predictive Coding), MFCC(Mel-Frequency Cepstrum Coefficients) 및 캡스트럼

(Cepstrum) 중 적어도 하나를 포함할 수 있다.

- [0038] 제어부(330)는 음성특성을 참조음성특성과 비교하고 유사도에 기초하여 음성을 인증할 수 있다.
- [0039] 제어부(330)는 특정화자에 대한 발화 패턴을 결정할 수 있다. 구체적으로, 제어부(330)는 음성이 특정화자가 특정단어를 발화한 복수개의 제1음성을 포함하는 경우, 제1음성의 음성특성을 참조음성특성과 비교하여, 소정 값 이상의 유사도가 소정 횟수 이상이면 해당 참조음성특성을 특정화자의 특정단어에 대한 발화 패턴으로 결정할 수 있다.
- [0040] 제어부(330)는 음성에 포함되는 단어 및 해당 단어의 음성특성을 추출하여 데이터베이스에 포함되는 적어도 하나의 참조 단어와 비교하되, 음성과 관련된 데이터를 기준으로 난수를 생성하고 소정 기준에 따라 데이터베이스 상에서 난수에 대응하는 참조 단어를 검출할 수 있다.
- [0041] 여기서, 음성과 관련된 데이터는 다양하게 설정될 수 있다. 예를 들어, 음성과 관련된 데이터는 음성의 인증이 요청된 시간일 수 있다. 이 경우, 제어부(330)는 음성의 인증이 요청된 시간을 기준으로 난수를 생성하고, 소정 행렬 구조로 구성되는 데이터베이스 상에서 난수의 소정 자리 숫자에 해당하는 행과 열에 해당하는 참조 단어를 검출할 수 있다. 이에 의하여, 음성 인식에 사용되는 단어가 데이터베이스 상에서 무작위적으로 추출되게 되어, 음성 인식의 보안성이 향상될 수 있다.
- [0042] 도 4는 본 발명의 일 실시 예에 의한 음성인식장치에 포함되는 제어부의 상세 구성을 도시한 블록도이다.
- [0043] 음성인식장치(300)에 포함되는 제어부(330)는 검색부(410), 비교부(420), 유사도 추정부(430), 발화 패턴 결정부(440) 및 화자 식별부(450) 등을 포함하는 연산처리블록들로 구성될 수 있다.
- [0044] 검색부(410)는 사전에 구축된 데이터베이스(DB)(460)에서 단어를 검색할 수 있다.
- [0045] 구체적으로, 검색부(410)는 음성과 관련된 데이터를 기준으로 난수를 생성하고, 소정 기준에 따라 데이터베이스(460)에서 생성된 난수에 대응하는 단어를 검출할 수 있다.
- [0046] 일 실시 예에 의하면, 검색부(410)는 행렬 구조로 구성되는 데이터베이스(460)에서 난수의 소정 자리 숫자에 해당하는 행과 열에 해당하는 참조 단어를 검출할 수 있다. 예를 들어, 검색부(410)는 생성된 난수의 자릿수를 구분하여 기 설정된 자릿수에 대응되는 2자리 숫자를 결정하고, 10 X 10 행렬 구조의 단어 데이터에서 결정된 2자리 숫자와 대응되는 행과 열의 단어를 검출할 수 있다. 만일, 생성된 난수가 987653이라고 가정하면, 2자리 숫자씩 분류했을 때 맨앞의 2자리 숫자는 98이고, 중간 2자리 숫자는 76이며, 맨끝의 2자리 숫자는 53이 된다. 이 때 맨앞의 2자리 숫자를 설정된 자릿수에 대응되는 2자리 숫자로 가정하면 98이라는 2자리 숫자를 결정할 수 있고, 10 X 10 행렬 구조의 단어 데이터에서 9행 8열에 해당하는 단어를 선택할 수 있다. 한편, 상기에서는 설명의 편의를 위하여 데이터베이스에 저장된 단어가 10 X 10 행렬 구조를 가지는 것을 가정하여 설명하지만, 본 발명이 이에 한정되는 것은 아니다. 본 발명의 데이터베이스에 저장되는 단어 데이터는 다양한 형태의 행렬 구조로 구성될 수 있다.
- [0047] 비교부(420)는 음성에 포함된 단어 및 음성특성을 데이터베이스(460)와 비교할 수 있다.
- [0048] 구체적으로, 비교부(420)는 데이터베이스(460)에 단어가 존재하면, 발화된 단어의 음성특성과 데이터베이스(460)상에 저장되어 있는 적어도 하나의 참조음성특성을 비교하여, 유사한지 여부를 판단할 수 있다. 또한, 비교부(420)는 데이터베이스(460)에 단어가 존재하지 않으면, 단어 및 단어에 대한 음성특성을 데이터베이스(460)에 추가할 수 있다.
- [0049] 유사도 추정부(430)는 발화된 단어의 음성특성과 각각의 참조음성특성과의 비교에 따른 유사도를 추정할 수 있다. 이 경우, 유사도 추정부(430)는 추정된 유사도가 제1기준값 미만인 경우에는 신규 참조음성특성으로 생성하여 데이터베이스(460)에 저장하고, 제1기준값 이상인 경우에는 해당 유사도를 갖는 참조음성특성의 매칭 횟수를 증가하여 카운팅할 수 있다.
- [0050] 발화 패턴 결정부(440)는 추정된 유사도에 상응하는 음성특성이 수신되는 횟수에 기초하여 화자의 단어에 대한 발화 패턴을 결정할 수 있다. 구체적으로, 카운팅된 매칭 횟수가 제2기준값 미만인 경우에는 화자로부터 발화된 신규의 음성이 수신되어 유사도가 추정되는 과정이 반복적으로 수행되도록 하고, 제2기준값 이상인 경우에는 화자의 단어에 대한 발화 패턴으로 결정할 수 있다.
- [0051] 이에 의하면, 제2기준값 이상의 카운팅된 매칭 횟수를 갖는 유사도에 상응하는 음성특성에 기초하여 화자의 음성 모델이 수립됨으로써 발화 패턴이 결정되고, 화자 식별부(450)에서는 발화된 음성에 대하여 결정된 발화 패

턴을 기준으로 화자가 누구인지 여부를 식별할 수 있다.

- [0052] 화자 식별부(450)는 결정된 발화 패턴에 기초하여 화자를 식별할 수 있다.
- [0053] 한편, 제어부(330)는 사전에 데이터베이스(460)를 구축하고 이를 관리할 수 있다. 이 경우, 데이터베이스(460)는 행렬 구조로 구성되어, 사용자와 음성특성 및 발화 매칭 횟수 정보 등의 인자에 기초하여 분류 및 정렬될 수 있다.
- [0054] 도 4를 참조하면, 예를 들어, 제1화자가 기업이라는 단어를 발화하면, 제1화자에 대한 식별자인 태그 정보(U000)가 할당된다. 이 경우, 발화된 기업이라는 단어에 대한 음성특성(벡터 특성 정보 V_Info000)와, 발화 매칭 횟수 정보(2)가 태그 정보(U000)와 연동되어 데이터베이스(460)상에 저장되어 관리될 수 있다.
- [0055] 또한, 제1화자가 은행이라는 단어를 발화하면, 제1화자에 대한 식별자인 태그 정보(U000)에 연동되어, 발화된 은행이라는 음성에 대한 음성 정보(V_Info003)가 발화 매칭 횟수 정보(7)와 함께 저장되어 관리될 수 있다.
- [0056] 한편, 이와 동일한 방식으로, 제2화자의 태그 정보는 U011로 할당되고, 이에 연동하여 발화된 단어와 음성특성 및 매칭 횟수 정보가 저장되어 관리될 수 있다.
- [0057] 도 5는 본 발명의 일 실시 예에 의한 음성인식 과정을 도시한 도면이다.
- [0058] 음성을 입력 받는다(S501).
- [0059] 음성인식장치(300)의 입력부(310)는 화자로부터 음성을 입력 받을 수 있다. 여기서, 화자는 특정 화자 및 불특정 화자를 포함할 수 있다. 음성은 음성인식장치(300)에 구비된 입력 수단(예를 들어, 음성 마이크 등)에 의해 직접 입력되거나, 유무선 통신에 의해 다른 장치로부터 입력될 수 있다.
- [0060] 음성을 음소 단위로 분할하고, 각 음소별로 음성특성을 추출한다(S502).
- [0061] 이를 위해, 음성인식장치(300)의 추출부(320)는 음성에 포함된 단어를 추출하고, 단어를 음소 단위로 분할하여 각 음소별로 음성특성을 추출할 수 있다. 여기서, 음성 특성은 음성의 주파수, 피치(pitch), 포먼트(formant), 발화시간 및 발화속도 중 적어도 하나를 포함할 수 있다.
- [0062] 음성특성을 참조음성특성과 비교하고, 유사도에 기초하여 음성을 인증한다(S503).
- [0063] 구체적으로, 음성인식장치(300)의 제어부(330)는 사전에 입력된 복수개의 참조음성을 발화 상태를 반영한 복수개의 그룹으로 분류하고, 복수개의 그룹 각각으로부터 참조음성특성을 추출할 수 있다. 이 경우, 제어부(330)는 입력된 음성의 음성특성을 참조음성특성과 비교하고 유사도에 기초하여 음성을 인증할 수 있다.
- [0064] 여기서, 발화 상태는, 발화자의 신체 상태(아플 때, 정상일 때 등), 감정 상태(기쁠 때, 슬플 때, 우울할 때, 화났을 때 등), 시간 변화(발화자의 나이, 발화자의 노화 등) 및 주변 환경(조용할 때, 소음이 많을 때 등) 중 적어도 하나를 포함할 수 있다. 또한, 참조음성특성은 음성 주파수, 피치(pitch), 포먼트(formant), 발화시간 및 발화속도 중 적어도 하나를 포함할 수 있다.
- [0065] 일 실시 예에 의하면, 제어부(330)는 복수개의 참조음성 각각에 대하여 발화 상태와 관련된 특징 벡터를 추출하고, 추출된 특징 벡터의 유사도에 기초하여 복수개의 참조음성을 그룹화하여 복수개의 그룹으로 분류하고, 복수개의 그룹 각각에 대하여 참조음성특성을 추출할 수 있다.
- [0066] 제어부(330)는 음성특성 및 참조음성특성 중 적어도 하나를 추출하는 경우, 복수개의 그룹 각각에 대하여 서로 다른 추출방식을 적용할 수 있다. 여기서, 추출방식은, PLP(Perceptual Linear Predictive Analysis), LPC(Linear Predictive Coding), MFCC(Mel-Frequency Cepstrum Coefficients) 및 켈스트럼(Cepstrum) 중 적어도 하나를 포함할 수 있다.
- [0067] 도 6은 본 발명의 일 실시 예에 의한 음성인식방법과 관련하여 화자에 대한 발화 패턴을 결정하는 과정을 도시한 도면이다.
- [0068] 화자로부터 발화된 음성을 수신한다(S601).
- [0069] 음성인식장치(300)의 입력부(310)는 화자로부터 음성을 입력 받을 수 있다.
- [0070] 음성으로부터 음성특성을 추출한다(S602).
- [0071] 일 실시 예에 의하면, 음성인식장치(300)의 추출부(320)는 음성에 포함된 단어를 추출하고, 단어를 음소 단위로

분할하여 각 음소별로 음성특성을 추출할 수 있다.

- [0072] 음성특성을 참조음성특성과 비교하고, 유사도를 추정한다(S603).
- [0073] 이 경우, 음성인식장치(300)의 제어부(330)는 발화된 단어의 음성특성과 각각의 참조음성특성과의 비교에 따른 유사도를 추정할 수 있다.
- [0074] 추정된 유사도에 상응하는 음성특성이 수신되는 횟수에 기초하여 화자의 발화 패턴을 결정한다(S604).
- [0075] 구체적으로, 음성인식장치(300)의 제어부(330)는 음성특성을 참조음성특성과 비교하여, 소정 값 이상의 유사도가 소정 횟수 이상이면 참조음성특성을 화자의 단어에 대한 발화 패턴으로 결정할 수 있다.
- [0076] 결정된 발화 패턴에 기초하여 화자를 식별한다(S605).
- [0077] 도 7은 본 발명의 일 실시 예에 의한 음성인식방법과 관련하여 화자에 대한 발화 패턴을 결정하는 상세 과정을 도시한 도면이다.
- [0078] 화자로부터 발화된 음성을 수신한다(S701).
- [0079] 음성인식장치(300)의 입력부(310)는 화자로부터 음성을 입력 받을 수 있다.
- [0080] 음성으로부터 단어 및 음성특성을 추출한다(S702).
- [0081] 일 실시 예에 의하면, 음성인식장치(300)의 추출부(320)는 음성에 포함된 단어를 추출하고, 단어를 음소 단위로 분할하여 각 음소별로 음성특성을 추출할 수 있다.
- [0082] 데이터베이스에서 단어 및 참조음성특성을 검색한다(S703).
- [0083] 구체적으로, 음성인식장치(300)의 제어부(330)는 음성과 관련된 데이터를 기준으로 단수를 생성하고, 소정 기준에 따라 데이터베이스(460)상에서 생성된 단수에 대응하는 단어를 검출할 수 있다. 이 경우, 제어부(330)는 단어에 대한 참조음성특성을 함께 검색할 수 있다.
- [0084] 만일, 데이터베이스(460)에 단어가 존재하면, 제어부(330)는 발화된 단어의 음성특성과 데이터베이스(460)에 저장되어 있는 대응하는 단어의 참조음성특성을 비교할 수 있다.
- [0085] 반면, 데이터베이스(460)에 단어가 존재하지 않으면, 제어부(330)는 단어 및 단어에 대한 음성특성을 데이터베이스(460)에 추가할 수 있다. 추가된 음성특성은 참조음성특성으로서, 이후 화자에 의한 음성이 입력되는 경우 입력된 음성의 음성특성과의 비교를 위한 기준 데이터로 활용될 수 있다.
- [0086] 입력된 단어의 음성특성과 데이터베이스 단어의 참조음성특성을 비교한다(S704).
- [0087] 이 경우, 음성인식장치(300)의 제어부(330)는 입력된 단어의 음성특성과 이에 대응하는 데이터베이스(460)에 저장된 단어의 참조음성특성과의 비교에 따른 유사도를 추정할 수 있다.
- [0088] 유사도가 제1기준값 이상인지 판단한다(S705).
- [0089] 여기서, 제1기준값은 사용자에게 의해 임의로 설정될 수 있다. 예를 들어, 제1기준값은 70%(또는 0.7)일 수 있다.
- [0090] 만일, 유사도가 제1기준값 미만이면(S705-No), 음성인식장치(300)의 제어부(330)는 신규 참조음성특성을 생성하여 데이터베이스(460)에 저장한다(S707). 이 경우, 판단된 유사도가 참조음성특성과 연동되어 데이터베이스(460)에 함께 저장될 수 있다. 동일한 화자가 동일한 단어를 발화하더라도, 화자의 상태나 환경 조건들에 따라 음성특성이 변화할 수 있다. 따라서, 이와 같이 유사도에 따라 신규 참조음성특성을 생성하여 데이터베이스(460)를 지속적으로 업데이트하고, 화자 개인별 특정 단어에 대한 발화 패턴을 지속적으로 추적 및 관리함으로써, 화자 인식의 정확성을 향상시킬 수 있다.
- [0091] 반면, 유사도가 제1기준값 이상이면(S705-Yes), 음성인식장치(300)의 제어부(330)는 해당 유사도를 갖는 참조음성특성의 매칭 횟수를 증가시킨다(S706).
- [0092] 카운팅된 매칭 횟수가 제2기준값 이상인지 판단한다(S708).
- [0093] 일 실시 예에 의하면, 제2기준값은 5에서 10사이의 범위에 포함되는 값일 수 있다.
- [0094] 만일, 매칭 횟수가 제2기준값 이상이면(S708-Yes), 음성인식장치(300)의 제어부(330)는 화자의 해당 단어에 대한 발화 패턴으로 결정한다(S709). 구체적으로, 해당 참조음성정보를 화자의 단어에 대한 발화 패턴으로써 결정

할 수 있다. 이후, 결정된 발화 패턴에 기초하여 화자를 식별한다(S710). 화자에 의하여 동일한 단어가 동일하거나 유사도가 높은 음성특성을 갖도록 반복적으로 발화된다면, 화자는 현재의 이러한 발화 패턴으로 또 다시 발화할 가능성이 매우 높다. 따라서, 화자의 발화 패턴에 대한 빈도를 카운팅하여 화자 인식(식별)에 사용함으로써 인식의 정확성과 신뢰성을 향상시킬 수 있다.

[0095] 반면, 매칭 횟수가 제2기준값 미만이면(S708-No), 음성인식장치(300)의 제어부(330)는 화자로부터 발화된 신규 음성이 수신되어 유사도가 추정되는 전술한 과정들이 반복적으로 수행되게 할 수 있다. 즉, 카운팅된 매칭 횟수가 일정 수준을 넘겨야 화자의 반복적인 현재의 발화 패턴임을 신뢰성 있게 추정할 수 있기 때문에, 소정의 수준(예컨대, 제2기준값 이상)이 되도록 화자로부터 신규 음성을 수신 받아 전술한 단계들을 반복적으로 수행한다.

[0096] 이에 의하면, 제1기준값 및 제2기준값 이상인 참조음성특성은 확인 대상 화자의 발화 패턴으로 결정될 수 있고, 음성이 입력되면 이에 의해 결정된 발화 패턴에 따라 해당 음성을 발화한 화자가 확인 대상 화자와 동일 인물인지 타인인지 여부를 신속하게 식별할 수 있다.

[0097] 도 8은 본 발명의 일 실시 예에 의한 음성인식방법을 설명하기 위한 도면이다.

[0098] 본 발명의 일 실시 예에 의한 음성인식방법에 의하면, 발화자의 발화상태를 고려하여 음성인식을 수행할 수 있다. 여기서, 발화 상태는, 발화자의 신체 상태(아플 때, 정상일 때 등), 감정 상태(기쁠 때, 슬플 때, 우울할 때, 화났을 때 등), 시간 변화(발화자의 나이, 발화자의 노화 등) 및 주변 환경(조용할 때, 소음이 많을 때 등) 중 적어도 하나를 포함할 수 있다.

[0099] 도 8에 도시된 바와 같이, 음성인식장치(300)는 화자의 일상적인 발화 패턴과 발화 상태를 모르는 상태에서, 발화자가 발화하는 음성에 대하여 각 개별 단어 수준에서 음성특성(정보) 및 참조음성특성(정보)에 대한 데이터베이스를 생성 및 구축한다. 이 경우, 음성인식장치(300)는 음성으로부터 발화 상태와 관련된 특징 벡터를 추출하고, 추출된 특징 벡터의 유사도에 기초하여 복수개의 음성을 복수개의 그룹으로 분류하여 각각의 그룹에 대하여 참조음성특성을 추출할 수 있다. 이에 의하여, 발화자의 상태를 고려한 참조음성특성을 생성하고, 이를 참조하여 음성인식을 수행할 수 있다.

[0100] 본 명세서에서 전술한 수치값들은 이해를 돕기 위하여 설명상의 편의를 위하여 제시된 예로써, 본 발명이 이러한 값들에 제한되는 것은 아니다.

[0101] 한편, 전술한 방법은 컴퓨터에서 실행될 수 있는 프로그램으로 작성 가능하고, 컴퓨터 판독 가능 매체를 이용하여 상기 프로그램을 동작시키는 범용 디지털 컴퓨터에서 구현될 수 있다. 또한, 상술한 방법에서 사용된 데이터의 구조는 컴퓨터 판독 가능 매체에 여러 수단을 통하여 기록될 수 있다. 본 발명의 다양한 방법들을 수행하기 위한 실행 가능한 컴퓨터 프로그램이나 코드를 기록하는 기록 매체는, 반송파(carrier waves)나 신호들과 같이 일시적인 대상들은 포함하지 않는 것으로 이해되어야 한다. 상기 컴퓨터 판독 가능 매체는 마그네틱 저장매체(예를 들면, 롬, 플로피 디스크, 하드 디스크 등), 광학적 판독 매체(예를 들면, 시디 롬, DVD 등)와 같은 저장매체를 포함할 수 있다.

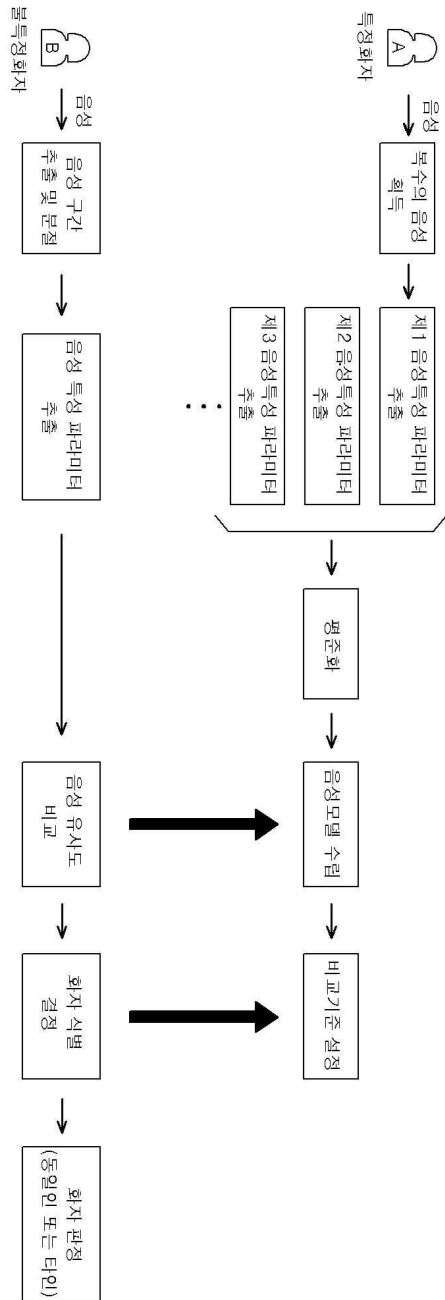
[0102] 이상에서 실시예를 중심으로 설명하였으나 이는 단지 예시일 뿐 본 발명을 한정하는 것이 아니며, 본 발명이 속하는 분야의 통상의 지식을 가진 자라면 본 실시예의 본질적인 특성을 벗어나지 않는 범위 내에서 이상에 예시되지 않은 여러 가지의 변형과 응용이 가능함을 알 수 있을 것이다. 예를 들어, 실시예에 구체적으로 나타난 각 구성 요소는 변형하여 실시할 수 있다. 그리고 이러한 변형과 응용에 관계된 차이점들은 첨부된 청구 범위에서 규정하는 본 발명의 범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

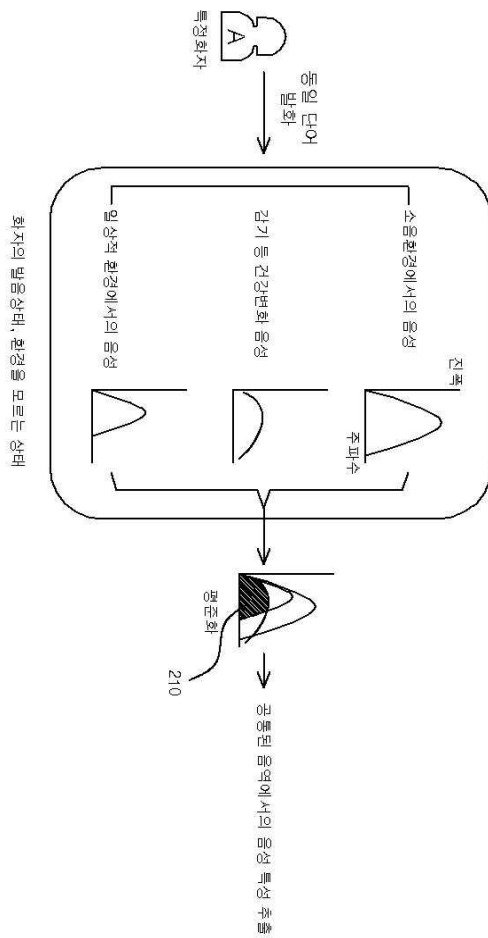
[0103]	300: 음성인식장치	310: 입력부
	320: 추출부	330: 제어부
	410: 검색부	420: 비교부
	430: 유사도 추정부	440: 발화 패턴 결정부
	450: 화자 식별부	460: 데이터베이스

도면

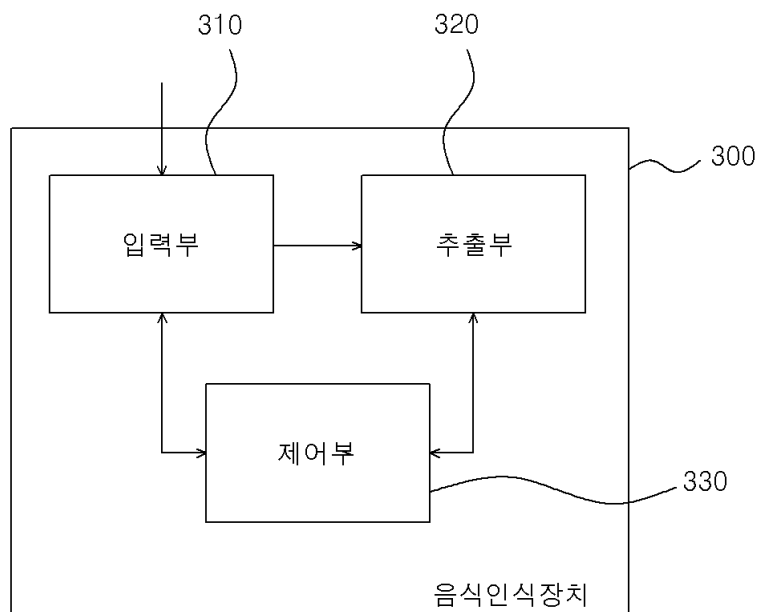
도면1



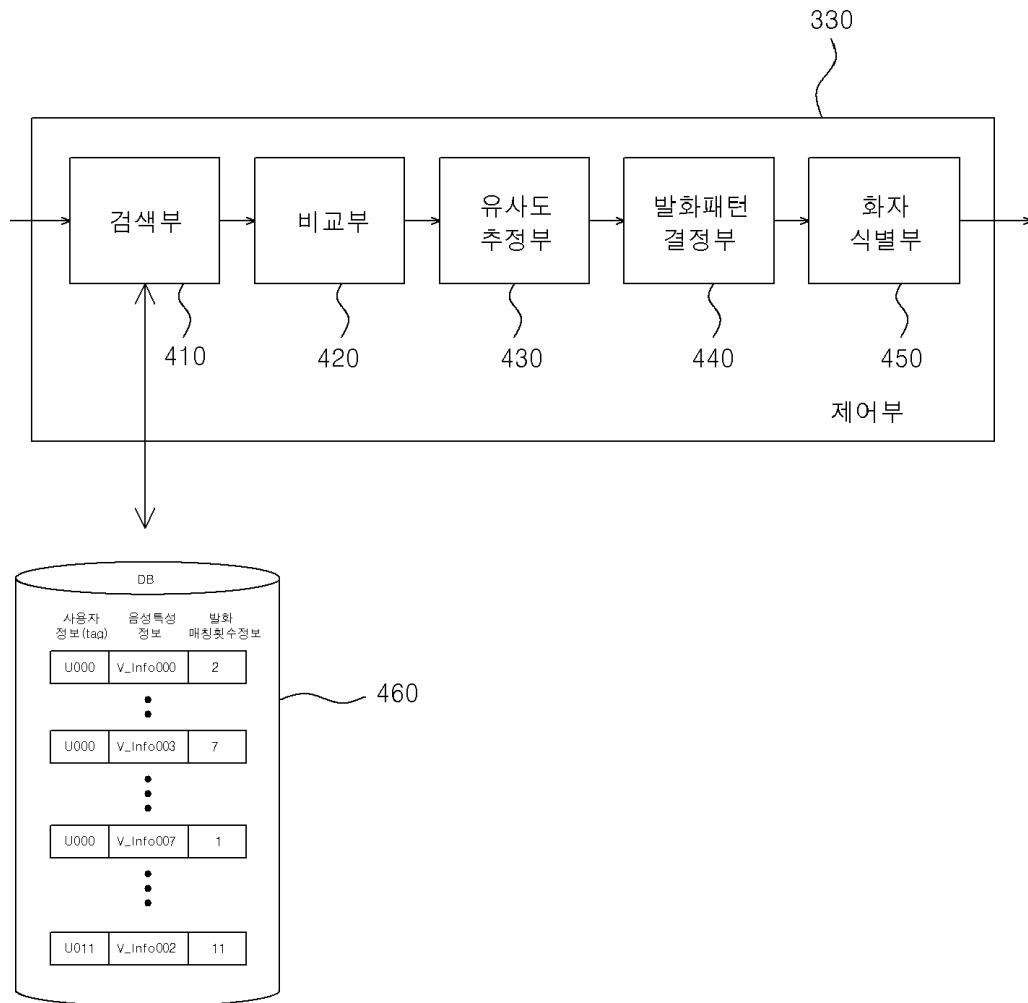
도면2



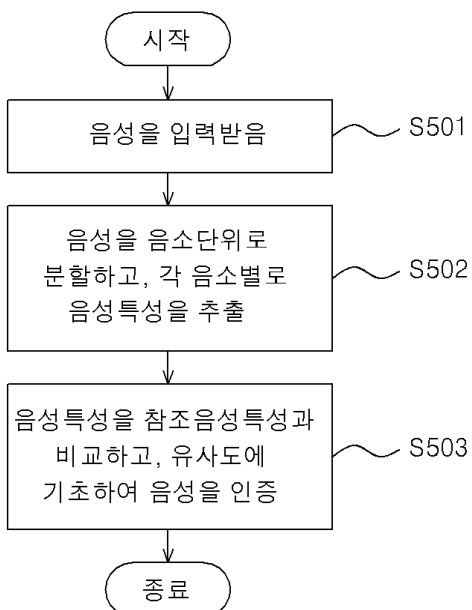
도면3



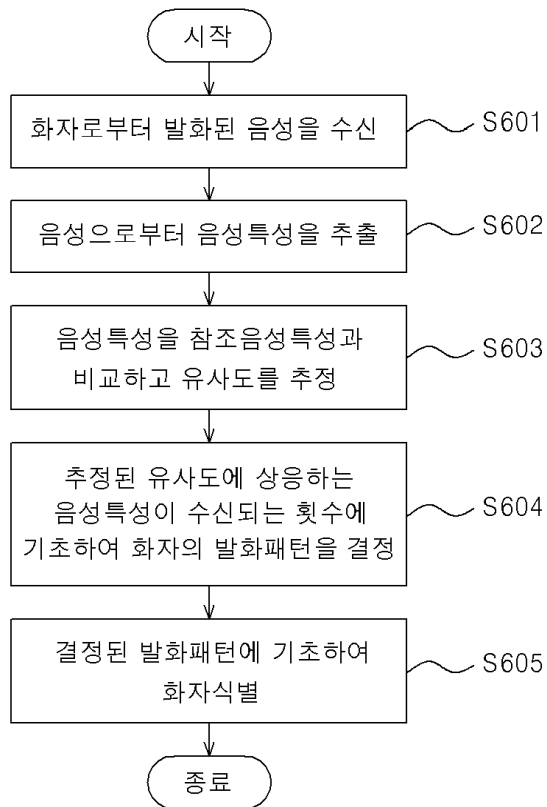
도면4



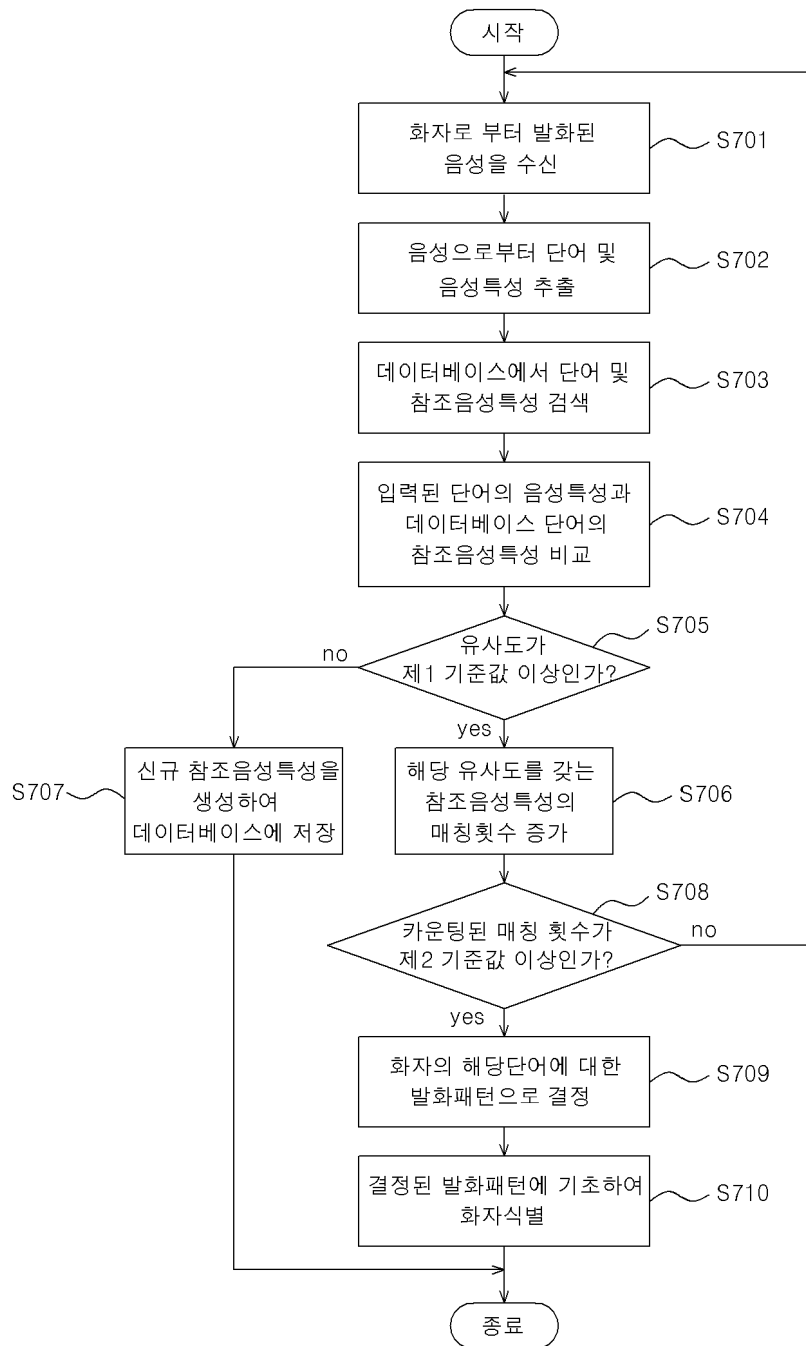
도면5



도면6



도면7



도면8

