



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년05월13일
(11) 등록번호 10-2250460
(24) 등록일자 2021년05월04일

(51) 국제특허분류(Int. Cl.)
G10L 17/04 (2013.01) G06F 21/32 (2013.01)
G10L 17/02 (2013.01) G10L 17/24 (2013.01)
(52) CPC특허분류
G10L 17/04 (2013.01)
G06F 21/32 (2013.01)
(21) 출원번호 10-2018-7005014
(22) 출원일자(국제) 2016년07월14일
심사청구일자 2019년06월19일
(85) 번역문제출일자 2018년02월20일
(65) 공개번호 10-2018-0034507
(43) 공개일자 2018년04월04일
(86) 국제출원번호 PCT/CN2016/089962
(87) 국제공개번호 WO 2017/012496
국제공개일자 2017년01월26일
(30) 우선권주장
201510438382.0 2015년07월23일 중국(CN)
(56) 선행기술조사문헌
CN103971700 A*
US20150081295 A1*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
어드밴스드 뉴 테크놀로지스 씨오., 엘티디.
케이만 군도, 그랜드 케이만 케이와이1-9008, 조지 타운, 27 하스피탈 로드, 케이만 코퍼레이트 센터
(72) 발명자
링 칭
중국 항저우 310099 완탕 로드 넘버 18 후양롱 타임즈 플라자 빌딩 비 17층 앤츠 패턴트 팀 내
(74) 대리인
김태홍, 김진희

전체 청구항 수 : 총 20 항

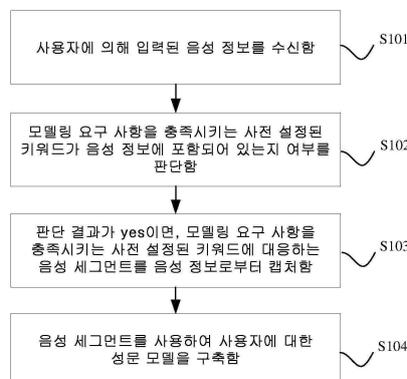
심사관 : 노지명

(54) 발명의 명칭 사용자 성문 모델을 구축하기 위한 방법, 장치 및 시스템

(57) 요약

본 출원은 사용자 성문 모델을 구축하기 위한 방법, 장치 및 시스템을 개시한다. 사용자 성문 모델을 구축하기 위한 방법은, 사용자에게 의해 입력된 음성 정보를 수신하는 단계; 모델링 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되어 있는지 여부를 판단하는 단계; 상기 판단 결과가 예스인 경우, 음성 정보로부터, 모델링 요구 사항을 충족시키는 사전 설정된 키워드에 대응하는 음성 세그먼트를 캡처하는 단계; 및 음성 세그먼트를 사용함으로써 사용자에게 대한 성문 모델을 구축하는 단계를 포함한다. 본 출원에 의해 제공되는 기술적 해결책이 적용되는 경우, 사용자의 음성이 수신될 수 있는 임의의 애플리케이션 시나리오에서, 사용자 성문 모델의 구축은 자동으로 완료될 수 있고, 따라서 사용자에게 의한 사용을 용이하게 할 수 있다.

대표도 - 도1



(52) CPC특허분류

G10L 17/02 (2013.01)

G10L 17/24 (2013.01)

명세서

청구범위

청구항 1

컴퓨터로 구현되는(computer-implemented) 방법에 있어서,

대응하는 보안 요구사항(requirement)을 가지고 특정 서비스를 구현하는 서비스 시스템의 음성 인식 시스템과 사용자가 상호 작용하는 제1 세션 동안 상기 사용자로부터 제1 음성 입력을 수신하는 단계;

하나 이상의 사전 결정된 키워드를 획득하는 단계로서, 상기 하나 이상의 사전 결정된 키워드는 적어도 최소 수의 요구되는 키워드를 포함하고, 상기 최소 수의 요구되는 키워드는 성문 인식 모델을 훈련시키는데 대한 상기 대응하는 보안 요구사항에 기초하는 것인, 상기 하나 이상의 사전 결정된 키워드를 획득하는 단계;

상기 하나 이상의 사전 결정된 키워드가 상기 제1 음성 입력에 나타나는지 여부를 결정하기 위해 상기 제1 음성 입력을 검색하는 단계;

상기 제1 음성 입력으로부터, 상기 서비스 시스템의 음성 인식 시스템과 상기 사용자가 상호 작용하는 제1 세션 동안 상기 하나 이상의 사전 결정된 키워드 중 적어도 하나를 상기 사용자가 말했다고 결정하는 단계;

상기 서비스 시스템의 음성 인식 시스템과 상기 사용자가 상호 작용하는 제1 세션 동안 상기 하나 이상의 사전 결정된 키워드 중 적어도 하나를 상기 사용자가 말했다고 결정하는 것에 응답하여, 상기 하나 이상의 사전 결정된 키워드에 각각 대응하는 하나 이상의 음성 세그먼트에 기초하여 상기 성문 인식 모델을 훈련시키는 단계;

상기 서비스 시스템의 음성 인식 시스템과 상기 사용자가 상호 작용하는 제2 세션 동안 상기 사용자로부터 제2 음성 입력을 수신하는 단계; 및

상기 제2 음성 입력을 수신하기 전에 상기 대응하는 보안 요구사항에 기초하여 상기 요구되는 키워드 전부를 획득하는 것에 응답하여, 상기 서비스 시스템의 음성 인식 시스템과 상기 사용자가 상호 작용하는 제1 세션으로부터의 상기 하나 이상의 음성 세그먼트로부터 생성된 상기 성문 인식 모델을 사용하여 상기 제2 세션 동안 수신된 제2 음성 입력에 기초하여 상기 사용자의 신원(identity)을 확인(verify)하는 단계

를 포함하는, 컴퓨터로 구현되는 방법.

청구항 2

제 1 항에 있어서, 상기 제2 음성 입력으로부터 하나 이상의 사전 결정된 키워드를 검출하는 단계를 더 포함하는, 컴퓨터로 구현되는 방법.

청구항 3

제 2 항에 있어서, 상기 제2 음성 입력으로부터 상기 하나 이상의 사전 결정된 키워드에 대응하는 하나 이상의 음성 세그먼트를 기록하는 단계를 더 포함하는, 컴퓨터로 구현되는 방법.

청구항 4

제 3 항에 있어서, 상기 제2 음성 입력으로부터 상기 하나 이상의 사전 결정된 키워드에 대응하는 하나 이상의 음성 세그먼트에 기초하여, 상기 성문 인식 모델을 업데이트하는 단계를 더 포함하는, 컴퓨터로 구현되는 방법.

청구항 5

제 1 항에 있어서, 상기 제1 음성 입력으로부터의 상기 하나 이상의 사전 결정된 키워드의 검출은 음향 모델에 기초하고, 상기 하나 이상의 음성 세그먼트는 상기 하나 이상의 사전 결정된 키워드와 동일한 하나 이상의 음향 특징을 포함하는 것인, 컴퓨터로 구현되는 방법.

청구항 6

제 1 항에 있어서, 상기 제1 음성 입력으로부터의 상기 하나 이상의 사전 결정된 키워드의 검출은 음성 인식에

기초하여 수행되는 것인, 컴퓨터로 구현되는 방법.

청구항 7

하나 이상의 명령어를 저장한 비밀시적인 컴퓨터 판독가능 저장 매체에 있어서, 상기 명령어는 동작들을 수행하도록 컴퓨터 시스템에 의해 실행가능하며, 상기 동작들은,

대응하는 보안 요구사항(requirement)을 가지고 특정 서비스를 구현하는 서비스 시스템의 음성 인식 시스템과 사용자가 상호 작용하는 제1 세션 동안 상기 사용자로부터 제1 음성 입력을 수신하는 동작;

하나 이상의 사전 결정된 키워드를 획득하는 동작으로서, 상기 하나 이상의 사전 결정된 키워드는 적어도 최소 수의 요구되는 키워드를 포함하고, 상기 최소 수의 요구되는 키워드는 성문 인식 모델을 훈련시키는데 대한 상기 대응하는 보안 요구사항에 기초하는 것인, 상기 하나 이상의 사전 결정된 키워드를 획득하는 동작;

상기 하나 이상의 사전 결정된 키워드가 상기 제1 음성 입력에 나타나는지 여부를 결정하기 위해 상기 제1 음성 입력을 검색하는 동작;

상기 제1 음성 입력으로부터, 상기 서비스 시스템의 음성 인식 시스템과 상기 사용자가 상호 작용하는 제1 세션 동안 상기 하나 이상의 사전 결정된 키워드 중 적어도 하나를 상기 사용자가 말했다고 결정하는 동작;

상기 서비스 시스템의 음성 인식 시스템과 상기 사용자가 상호 작용하는 제1 세션 동안 상기 하나 이상의 사전 결정된 키워드 중 적어도 하나를 상기 사용자가 말했다고 결정하는 것에 응답하여, 상기 하나 이상의 사전 결정된 키워드에 각각 대응하는 하나 이상의 음성 세그먼트에 기초하여 상기 성문 인식 모델을 훈련시키는 동작;

상기 서비스 시스템의 음성 인식 시스템과 상기 사용자가 상호 작용하는 제2 세션 동안 상기 사용자로부터 제2 음성 입력을 수신하는 동작; 및

상기 제2 음성 입력을 수신하기 전에 상기 대응하는 보안 요구사항에 기초하여 상기 요구되는 키워드 전부를 획득하는 것에 응답하여, 상기 서비스 시스템의 음성 인식 시스템과 상기 사용자가 상호 작용하는 제1 세션으로부터의 상기 하나 이상의 음성 세그먼트로부터 생성된 상기 성문 인식 모델을 사용하여 상기 제2 세션 동안 수신된 제2 음성 입력에 기초하여 상기 사용자의 신원(identity)을 확인(verify)하는 동작

을 포함하는 것인, 비밀시적인 컴퓨터 판독가능 저장 매체.

청구항 8

제 7 항에 있어서, 상기 동작들은,

상기 제2 음성 입력으로부터 하나 이상의 사전 결정된 키워드를 검출하는 동작을 더 포함하는 것인, 비밀시적인 컴퓨터 판독가능 저장 매체.

청구항 9

제 8 항에 있어서, 상기 동작들은,

상기 제2 음성 입력으로부터 상기 하나 이상의 사전 결정된 키워드에 대응하는 하나 이상의 음성 세그먼트를 기록하는 동작을 더 포함하는 것인, 비밀시적인 컴퓨터 판독가능 저장 매체.

청구항 10

제 9 항에 있어서, 상기 동작들은,

상기 제2 음성 입력으로부터 상기 하나 이상의 사전 결정된 키워드에 대응하는 하나 이상의 음성 세그먼트에 기초하여, 상기 성문 인식 모델을 업데이트하는 동작을 더 포함하는 것인, 비밀시적인 컴퓨터 판독가능 저장 매체.

청구항 11

제 7 항에 있어서, 상기 제1 음성 입력으로부터의 상기 하나 이상의 사전 결정된 키워드의 검출은 음향 모델에 기초하고, 상기 하나 이상의 음성 세그먼트는 상기 하나 이상의 사전 결정된 키워드와 동일한 하나 이상의 음향 특징을 포함하는 것인, 비밀시적인 컴퓨터 판독가능 저장 매체.

청구항 12

제 7 항에 있어서, 상기 제1 음성 입력으로부터의 상기 하나 이상의 사전 결정된 키워드의 검출은 음성 인식에 기초하여 수행되는 것인, 비밀시적인 컴퓨터 판독가능 저장 매체.

청구항 13

컴퓨터로 구현되는 시스템에 있어서,

하나 이상의 컴퓨터; 및

상기 하나 이상의 컴퓨터와 상호동작가능하게 연결되며, 하나 이상의 명령어를 저장한 유형의(tangible) 비밀시적인 기계 판독 가능한 저장 매체를 가지는 하나 이상의 컴퓨터 메모리 디바이스를 포함하고, 상기 하나 이상의 명령어는 상기 하나 이상의 컴퓨터에 의해 실행될 때 하나 이상의 동작을 수행하며, 상기 하나 이상의 동작은,

대응하는 보안 요구사항(requirement)을 가지고 특정 서비스를 구현하는 서비스 시스템의 음성 인식 시스템과 사용자가 상호 작용하는 제1 세션 동안 상기 사용자로부터 제1 음성 입력을 수신하는 동작;

하나 이상의 사전 결정된 키워드를 획득하는 동작으로서, 상기 하나 이상의 사전 결정된 키워드는 적어도 최소 수의 요구되는 키워드를 포함하고, 상기 최소 수의 요구되는 키워드는 성문 인식 모델을 훈련시키는데 대한 상기 대응하는 보안 요구사항에 기초하는 것인, 상기 하나 이상의 사전 결정된 키워드를 획득하는 동작;

상기 하나 이상의 사전 결정된 키워드가 상기 제1 음성 입력에 나타나는지 여부를 결정하기 위해 상기 제1 음성 입력을 검색하는 동작;

상기 제1 음성 입력으로부터, 상기 서비스 시스템의 음성 인식 시스템과 상기 사용자가 상호 작용하는 제1 세션 동안 상기 하나 이상의 사전 결정된 키워드 중 적어도 하나를 상기 사용자가 말했다고 결정하는 동작;

상기 서비스 시스템의 음성 인식 시스템과 상기 사용자가 상호 작용하는 제1 세션 동안 상기 하나 이상의 사전 결정된 키워드 중 적어도 하나를 상기 사용자가 말했다고 결정하는 것에 응답하여, 상기 하나 이상의 사전 결정된 키워드에 각각 대응하는 하나 이상의 음성 세그먼트에 기초하여 상기 성문 인식 모델을 훈련시키는 동작;

상기 서비스 시스템의 음성 인식 시스템과 상기 사용자가 상호 작용하는 제2 세션 동안 상기 사용자로부터 제2 음성 입력을 수신하는 동작; 및

상기 제2 음성 입력을 수신하기 전에 상기 대응하는 보안 요구사항에 기초하여 상기 요구되는 키워드 전부를 획득하는 것에 응답하여, 상기 서비스 시스템의 음성 인식 시스템과 상기 사용자가 상호 작용하는 제1 세션으로부터의 상기 하나 이상의 음성 세그먼트로부터 생성된 상기 성문 인식 모델을 사용하여 상기 제2 세션 동안 수신된 제2 음성 입력에 기초하여 상기 사용자의 신원(identity)을 확인(verify)하는 동작

을 포함하는 것인, 컴퓨터로 구현되는 시스템.

청구항 14

제 13 항에 있어서, 상기 하나 이상의 동작은,

상기 제2 음성 입력으로부터 하나 이상의 사전 결정된 키워드를 검출하는 동작을 더 포함하는 것인, 컴퓨터로 구현되는 시스템.

청구항 15

제 14 항에 있어서, 상기 하나 이상의 동작은,

상기 제2 음성 입력으로부터 상기 하나 이상의 사전 결정된 키워드에 대응하는 하나 이상의 음성 세그먼트를 기록하는 동작을 더 포함하는 것인, 컴퓨터로 구현되는 시스템.

청구항 16

제 15 항에 있어서, 상기 하나 이상의 동작은,

상기 제2 음성 입력으로부터 상기 하나 이상의 사전 결정된 키워드에 대응하는 하나 이상의 음성 세그먼트에 기초하여, 상기 성문 인식 모델을 업데이트하는 동작을 더 포함하는 것인, 컴퓨터로 구현되는 시스템.

청구항 17

제 13 항에 있어서, 상기 제1 음성 입력으로부터의 상기 하나 이상의 사전 결정된 키워드의 검출은 음향 모델에 기초하고, 상기 하나 이상의 음성 세그먼트는 상기 하나 이상의 사전 결정된 키워드와 동일한 하나 이상의 음향 특징을 포함하는 것인, 컴퓨터로 구현되는 시스템.

청구항 18

제 1 항에 있어서, 상기 사용자로부터 상기 제1 음성 입력을 수신하기 전에 그리고 상기 제1 음성 입력에 기초하여 상기 성문 인식 모델을 훈련시키기 전에, 상기 사용자를 패스워드, 제스처, 또는 생체 데이터로 확인하는 단계를 더 포함하는, 컴퓨터로 구현되는 방법.

청구항 19

제 7 항에 있어서, 상기 동작들은,

상기 사용자로부터 상기 제1 음성 입력을 수신하기 전에 그리고 상기 제1 음성 입력에 기초하여 상기 성문 인식 모델을 훈련시키기 전에, 상기 사용자를 패스워드, 제스처, 또는 생체 데이터로 확인하는 동작을 더 포함하는 것인, 비밀시적인 컴퓨터 관독가능 저장 매체.

청구항 20

제 13 항에 있어서, 상기 하나 이상의 동작은,

상기 사용자로부터 상기 제1 음성 입력을 수신하기 전에 그리고 상기 제1 음성 입력에 기초하여 상기 성문 인식 모델을 훈련시키기 전에, 상기 사용자를 패스워드, 제스처, 또는 생체 데이터로 확인하는 동작을 더 포함하는 것인, 컴퓨터로 구현되는 시스템.

발명의 설명

기술 분야

- [0001] 본 출원은 2015년 7월 23일자에 출원되고, 발명의 명칭이 "사용자 성문 모델을 구축하기 위한 방법, 장치 및 시스템"인 중국 특허 출원 제201510438382.0호의 우선권을 주장하며, 그 전체가 본 명세서에 참조로 포함된다.
- [0002] 본 출원은 컴퓨터 애플리케이션 기술 분야에 관한 것으로서, 특히 사용자 성문(voiceprint) 모델을 구축하기 위한 방법, 장치 및 시스템에 관한 것이다.

배경 기술

- [0003] 성문은 생체 특징(feature) 중 하나이다. 성문 인식은 화자(speaker)의 음성 및 행동 특성에 따라 화자의 신원을 자동으로 인식하기 위한 생체 인식 방법이다. 성문 인식에 의해 제공되는 보안은 다른 생체 인식 기술(예를 들어, 지문 및 홍채)에 의해 제공되는 보안과 비슷하며, 이것은 음성 수집 장치를 구현하기 위해 다른 특수 장치가 아닌 전화/휴대폰 또는 마이크만 사용할 필요가 있다. 성문 인식 방법은 음성 언어, 방언 또는 억양과는 관련이 없으며, 어떠한 프라이버시 문제도 수반하지 않으며, 널리 적용된다. 소리 신호(sound signal)는 원격 전송 및 획득이 편리하며, 성문 인식 방법은 전기통신 및 네트워크를 기반으로 하는 신원 인식 애플리케이션에서 특히 유리하다.
- [0004] 성문 인식은 주로 두 단계를 포함한다. 제 1 단계는 사용자가 음성의 세그먼트를 시스템에 미리 기록할 필요가 있는 등록 단계이며, 시스템은 등록된 음성으로부터 추출된 성문 특징을 이용하여 사용자에 대한 성문 모델을 구축한다. 제 2 단계는 시스템 프롬프트에 따라 사용자가 지정된 음성 세그먼트를 말하는 적용 단계이며, 시스템은 사용자의 음성을 자동으로 기록하고, 성문 특징을 추출하고, 새로 추출된 성문 특징 및 사전 구축된 성문 모델에 대해 매칭 작업을 수행하며, 두 개의 매칭 정도에 따라 현재 사용자의 신원을 인식한다.
- [0005] 컴퓨터와 인터넷 기술의 발달로, 성문 인식은, 예를 들어, 스마트 단말기 애플리케이션, 자동 전화 서비스 시스템 등과 같은 사람들의 일상생활에 점차적으로 침투해왔다. 등록이 전체 성문 인식 프로세스에서 필수 단계임을 상기 설명으로부터 알 수 있다. 그러나 등록 단계는 사용자들에게 특히 중년 및 노인과 같이 새로운 기술에 민감하지 않은 특수 그룹의 사용자들에게 어느 정도 불편을 가져다줄 것이며, 심지어 성문을 성공적으로 등록할

수 있는지 여부의 문제도 있다. 또한, 일부 경우에, 시스템의 보안 또는 견고성을 향상시키기 위해, 사용자가 등록 중에 여러 음성 세그먼트를 기록하도록 요구될 수 있으며, 이것은 사용자에게 사용의 어려움을 더욱 증가시킬 것이다.

발명의 내용

- [0006] 본 출원은 성문 인식 시스템을 사용하는 어려움을 줄이기 위해 사용자 성문 모델을 구축하기 위한 방법, 장치 및 시스템을 제공한다. 기술적 해결책은 다음과 같다:
- [0007] 본 출원은 사용자 성문 모델을 구축하기 위한 방법을 제공하며, 상기 방법은:
- [0008] 사용자에게 의해 입력된 음성 정보를 수신하는 단계;
- [0009] 모델링 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되는지 여부를 판단하는 단계;
- [0010] 상기 판단 결과가 예스(yes)인 경우, 음성 정보로부터, 모델링 요구 사항을 충족시키는 사전 설정된 키워드에 대응하는 음성 세그먼트를 캡처하는 단계; 및
- [0011] 음성 세그먼트를 사용하여 사용자에게 대한 성문 모델을 구축하는 단계를 포함하는 방법.
- [0012] 본 출원은 성문 기반 사용자 신원 확인 방법을 제공하며, 상기 방법은:
- [0013] 사용자에게 의해 입력된 음성 정보를 수신하는 단계;
- [0014] 확인 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되는지 여부를 판단하는 단계;
- [0015] 상기 판단 결과가 예스인 경우, 음성 정보로부터, 확인 요구 사항을 충족시키는 사전 설정된 키워드에 대응하는 음성 세그먼트를 캡처하는 단계;
- [0016] 음성 세그먼트의 성문 특징을 추출하는 단계; 및
- [0017] 성문 특징 및 사용자의 사전 구축된 성문 모델을 사용하여 사용자의 신원을 확인하는 단계를 포함하는 방법.
- [0018] 본 출원은 사용자 성문 모델을 구축하기 위한 장치를 제공하며, 상기 장치는:
- [0019] 사용자에게 의해 입력된 음성 정보를 수신하도록 구성된 음성 정보 수신 모듈;
- [0020] 모델링 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되는지 여부를 판단하도록 구성된 모델링 키워드 판단 모듈;
- [0021] 상기 판단 결과가 예스인 경우, 음성 정보로부터, 모델링 요구 사항을 충족시키는 사전 설정된 키워드에 대응하는 음성 세그먼트를 캡처하도록 구성된 음성 세그먼트 캡처 모듈; 및
- [0022] 음성 세그먼트를 사용하여 사용자에게 대한 성문 모델을 구축하도록 구성된 성문 모델 구축 모듈을 포함하는 장치.
- [0023] 본 출원은 성문 기반 사용자 신원 확인 장치를 제공하며, 상기 장치는:
- [0024] 사용자에게 의해 입력된 음성 정보를 수신하도록 구성된 음성 정보 수신 모듈;
- [0025] 확인 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되는지 여부를 판단하도록 구성된 키워드 판단 모듈;
- [0026] 상기 판단 결과가 예스인 경우, 음성 정보로부터, 확인 요구 사항을 충족시키는 사전 설정된 키워드에 대응하는 음성 세그먼트를 캡처하도록 구성된 음성 세그먼트 캡처 모듈;
- [0027] 음성 세그먼트의 성문 특징을 추출하도록 구성된 성문 특징 추출 모듈; 및
- [0028] 성문 특징 및 사용자의 사전 구축된 성문 모델을 사용하여 사용자의 신원을 확인하도록 구성된 확인 모듈을 포함하는 장치.
- [0029] 본 출원은 성문 기반 사용자 신원 확인 시스템을 제공하고, 상기 시스템은 상술한 바와 같이 사용자 성문 모델을 구축하기 위한 장치 및 사용자 신원 확인 장치를 포함한다.
- [0030] 본 출원의 실시예들에 의해 제공되는 사용자 성문 모델을 구축하기 위한 방법을 적용함으로써, 이론적으로 사용

자 성문 모델은 사용자의 음성이 수신될 수 있는 임의의 애플리케이션 시나리오에서 자동으로 구축될 수 있다. 여기에서 음성 정보의 수신은 사용자가 다른 기능들을 구현하는데 필요한 작업일 수 있으며, 의도적으로 성문 정보 등록 작업을 실행할 필요가 없으므로 사용자에게 의한 사용을 용이하게 한다.

[0031] 따라서, 신원 확인을 요구하는 애플리케이션 시나리오에서, 사용자의 음성 정보가 획득될 수 있고 음성 정보가 이전에 성문 모델을 구축한 키워드를 포함하면, 성문을 사용하여 신원 확인을 자동으로 수행함으로써 다른 확인 방식의 복잡한 작업을 생략할 수 있다. 또한, 사용자가 확인 프로세스에서 의도적으로 확인 패스워드를 말할 필요가 없다.

[0032] 위의 일반적인 설명 및 이하의 상세한 설명은 단지 예시적이고 설명적인 것이며, 본 출원을 제한할 수 없다는 것을 이해해야 한다.

도면의 간단한 설명

[0033] 종래 기술 또는 본 출원의 실시예들의 기술적 해결책을 보다 명확하게 설명하기 위해, 이하에서는 실시예들 또는 종래 기술을 설명하기 위해 요구되는 첨부 도면을 간단히 소개한다. 명백하게, 이하에 설명되는 첨부 도면은 본 출원에 기록된 일부 실시예들에 불과하며, 당업자는 여전히 이들 첨부 도면으로부터 다른 도면을 도출할 수 있다.

도 1은 본 출원에 따른 사용자 성문 모델을 구축하기 위한 방법의 개략적인 흐름도이다.

도 2는 본 출원에 따른 사용자 신원 확인 방법의 개략적인 흐름도이다.

도 3은 본 출원에 따른 사용자 성문 모델을 구축하기 위한 장치의 개략적인 구조도이다.

도 4는 본 출원에 따른 사용자 신원 확인 장치의 개략적인 구조도이다.

도 5는 본 출원에 따른 사용자 신원 확인 시스템의 개략적인 구조도이다.

발명을 실시하기 위한 구체적인 내용

[0034] 음성은 인간 사이의 의사 소통을 위한 중요한 매개체이다. 음성 인식과 같은 기술이 급속하게 발전함에 따라, 사용자는 이미 음성 입력 및 음성 제어와 같은, 다양한 시나리오에서 음성으로 컴퓨터와 통신할 수 있었다. 즉, 컴퓨터의 관점에서, 사용자 음성 정보를 수집할 수 있는 많은 기회가 있다. 또한, 사용자의 전화 메시지 및 실시간 전화 기록으로부터 사용자의 음성 정보를 수집할 수 있다. 이러한 실제 상황에 기초하여, 본 출원은 사용자에게 대한 어떠한 추가적인 등록 작업도 추가하지 않으면서 사용자의 음성을 수집할 수 있는 임의의 시나리오에서 사용자의 성문 모델링을 구현할 수 있는 해결책을 제공함으로써, 성문 등록 작업이 복잡하다는 종래의 문제점을 해결할 수 있다.

[0035] 성문 인식은 텍스트 종속적 성문 인식 및 텍스트 독립적 성문 인식을 포함한다. 텍스트 종속적 성문 인식 시스템은 사용자에게 등록 동안 지정된 내용에 따라 발음하도록 요구한다. 각 사용자의 성문 모델이 정확하게 확립된다. 또한, 사용자는 인식 동안 지정된 내용에 따라 발음해야 한다. 모든 사용자의 성문 모델이 "표준"이므로 원하는 인식 효과를 달성할 수 있다. 그러나 시스템은 사용자의 협조를 요구하고, 사용자의 발음이 특정 내용에 대해 확인되지 않으면, 사용자를 정확하게 인식할 수 없다. 이론적으로, 텍스트 독립적 인식 시스템은 화자의 발음 내용을 명시할 수 없다. 그러나 모델을 확립하는 것이 상대적으로 어려우며, 실제 적용 중에 이상적인 인식 효과를 달성할 수 없다.

[0036] 인식 효과를 보장하기 위해, 본 출원에서 제공된 해결책은 "텍스트 종속적" 성문 인식 기술을 위해 제안되고, 그 다음, 해결되어야 할 문제점은: 사용자에게 대한 추가적인 작업을 증가시키지 않고 사용자가 "지정된 내용"을 말하도록 하는 방법이다.

[0037] 이 해결책을 구현하는 동안, 발명자는: 특정 애플리케이션 시나리오에서, 사용자가 시나리오와 관련된 키워드를 말할 확률이 매우 높다는 것을 발견한다. 예를 들어, 사용자가 음성(예를 들어, 전화 또는 음성 메시지와 같은 방식으로)을 사용하여 제품 상담을 수행하는 경우, 사용자는 이름, 모델, 기능, 부분 및 정상 실패와 같은 제품의 키워드 중 하나 이상을 거의 불가피하게 말할 수 있다. 사용자가 음성 제어 작업을 수행할 때, 많은 음성 제어 명령들이 또한 사전 정의되고, 다양한 유사한 시나리오들이 존재한다. 이러한 실제 상황에 기초하여, 성문 인식이 적용되는 상이한 시나리오에 대해, 이러한 시나리오에 나타나는 하나 이상의 고 빈도 키워드에 대해 미리 통계를 작성할 수 있고, 고 빈도 키워드는 "지정된 내용"으로 정의될 수 있다. 이러한 방식으로, 사용자

가 음성 상호 작용 중에 이러한 키워드를 말하면 시스템은 사용자에게 대한 성문 모델링을 구현할 수 있다.

- [0038] 당업자가 본 출원의 기술적 해결책을 더 잘 이해할 수 있도록 하기 위해, 본 출원의 실시예들에서의 기술적 해결책은 본 출원의 실시예들에서의 첨부 도면을 통해 상세히 설명될 것이다. 명백하게, 설명된 실시예들은 단지 본 출원의 실시예들의 일부이며, 모든 실시예가 아니다. 본 출원의 실시예들에 기초하여, 당업자에 의해 도출된 모든 다른 실시예들은 모두 본 출원의 보호 범위 내에 있다.
- [0039] 도 1에 도시된 바와 같이, 본 출원에 따른 사용자 성문 모델을 구축하기 위한 방법의 흐름도가 도시되어 있으며, 이 방법은 다음의 단계들을 포함할 수 있다:
- [0040] S101: 사용자에게 의해 입력된 음성 정보를 수신한다;
- [0041] S102: 모델링 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되는지 여부를 판단한다;
- [0042] S103: 판단 결과가 예스라면, 모델링 요구 사항을 충족시키는 사전 설정된 키워드에 대응하는 음성 세그먼트를 음성 정보로부터 캡처한다; 및
- [0043] S104: 음성 세그먼트를 사용하여 사용자에게 대한 성문 모델을 구축한다.
- [0044] 상기 해결책은 다음에서 더 상세하게 설명될 것이다:
- [0045] 이론적으로, 본 출원의 해결책은 음성 입력, 음성 제어 및 전화 메시지와 같은 사용자의 음성이 수집될 수 있는 임의의 시나리오에서 구현될 수 있다. 당연히, 실제 적용 중에, 다음 요소를 더 고려해야 한다: 시나리오가 성문 인식 요구 사항을 가지고 있는지 여부, 현재 말하는 사용자의 신원을 결정할 수 있는지 여부, 사용자의 음성 길이(계산 복잡도를 포함), 키워드를 포함하는 음성 세그먼트를 효과적으로 수집할 수 있는지 여부(수집 효율성) 등.
- [0046] 통상적인 애플리케이션 시나리오는, 예를 들어, 대화형 음성 응답(Interactive Voice Response, IVR) 시스템에서 "한 문장으로 문제를 설명한다"이다. 시스템은 호출자 ID 또는 사용자가 수동으로 입력한 확인 정보 수단 또는 다른 방식에 따라, 사용자의 신원을 확인할 수 있다. "한 문장"의 음성 길이는 후속 처리에 적합하다. 또한, 사용자는 많은 특정 서비스 시나리오에서 일부 키워드를 말할 가능성이 크다. 예를 들어, 온라인 거래 시나리오에서, 관련 키워드로는 주문, 수령, 배송, 금액 등이 있다. 은행 및 결제와 같은 시나리오에서, 관련 키워드로는 전송, 잔액, 계좌, 패스워드 등이 있다.
- [0047] 또한, 사용자가 개인 단말기를 사용하는 경우, 운영 체제 계층 또는 다양한 특정 애플리케이션 계층에서 사용되는 다양한 음성 제어 명령 모두가 성문 모델링 중에 음성 정보 수집을 위해 사용될 수 있다. 첫째, 사용자는 단말기를 개인 소지품으로 사용할 때 다양한 확인 작업을 수행할 수 있다. 예를 들어, 휴대 전화를 잠금 해제하거나 일부 특정 애플리케이션을 시작할 때 사용자는 제스처 또는 숫자 패스워드를 입력하도록 요구될 수 있으므로, 사용자 신원 확인 요구 사항을 먼저 충족시킬 수 있다. 또한, 운영 체제 또는 특정 애플리케이션에 의해 사용되는 음성 제어 명령은 너무 길지 않으며, 그 중 대부분은 사전 정의되어 있다. 예를 들어, Android 휴대 전화의 음성 상호 작용 기능은 "OK Google" 또는 "OK Google Now" 음성 명령을 사용하여 트리거된다. 이러한 키워드 모두는 사용자 성문 모델링을 구현하는 데 사용될 수 있다.
- [0048] 계산 복잡도에 대한 요구 사항이 그다지 크지 않으면, 수동 응답 플랫폼에 실제로 액세스된 사용자 통화가 기록될 수 있고, 수동 응답 플랫폼의 음성 및 사용자의 음성은 오디오 특성에 따라 구별되어, 후속 단계에서 사용자의 음성 부분만 분석된다. 이러한 방식으로, 키워드를 포함하는 음성 세그먼트는 길고 완전한 사용자 통화로부터 보다 효과적으로 수집될 수 있다.
- [0049] 당연히, 상기의 몇몇 애플리케이션 시나리오는 단지 예시를 위해 사용된 것이고, 본 출원의 해결책에 대한 제한으로 이해되어서는 안 된다.
- [0050] S101에 따르면, 사용자의 음성 정보는 임의의 애플리케이션 시나리오에서 수신될 수 있다. 당연히, 최종 요구 사항은 성문 모델링을 구현하는 것이므로, 다른 방식(예를 들어, 사용자 ID, 확인 패스워드, 제스처, 지문 등)으로 먼저 현재 화자의 신원을 확인할 필요가 있다.
- [0051] S102에서, 모델링 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되는지 여부를 판단한다. 여기서, 하나 이상의 키워드는 실제 요구 사항에 따라 사전 설정될 수 있다. 여러 개의 키워드를 사전 설정하는 것은 다음과 같은 점에서 유리하다: 한편으로는 성공적인 모델링 확률이 향상될 수 있는 것인데 즉, 사용자가 여러 개의 키워드 중 적어도 하나를 말하면 성문 모델링을 구현할 수 있다는 것이고, 다른 한편으로는 보안 요

구 사항이 높은 일부 애플리케이션 시나리오에서 확인을 위해 여러 개의 성문 세그먼트를 함께 사용하는 것이 요구될 수 있다는 것이다.

[0052] 이 단계에서, 사전 설정된 키워드가 음성 정보에 포함되는지 여부는, 음성 인식 관련 기술을 사용하여 결정될 필요가 있다. 음성 인식 시스템의 모델은 일반적으로 음향 모델 및 언어 모델을 포함하는 두 개의 부분을 포함하며, 각각은 음성으로부터 음절로의 확률의 계산 및 음절로부터 단어로의 확률의 계산에 대응한다. 본 출원의 해결책의 실제 요구 사항에 따르면, 음성 정보에 대한 인식은 다음 두 가지 관점에서 구현될 수 있다:

[0053] a) 음향 특징 사용:

[0054] 인간 언어는 모두 일련의 연속적인 음절로 구성되며, 연속적인 소리 정보는 음절 흐름으로 지칭된다. 음향의 관점에서, 음절 흐름은 몇 개의 독립적인 음절로 분할될 수 있다. 또한, 각각의 발음 음절에는 고정된 오디오 특징이 있다. 음성 인식 기술은 두 가지 측면에 기반을 두고 음향 모델을 확립하여 음절 흐름에 대해 단어 세그먼트화를 구현한다. 그러나 본 출원에서, 기술은 사용자 음성 정보에 대한 인식을 구현하기 위해 직접 사용될 수 있으며, 특정 해결책은 다음의 단계를 포함한다:

[0055] 적어도 하나의 음성 세그먼트를 획득하기 위해 음향 모델을 사용하여 음성 정보를 분할하는 단계; 및

[0056] 획득된 음성 세그먼트 또는 이들의 조합이 모델링 요구 사항을 충족시키는 사전 설정된 키워드의 오디오 특징과 매칭하는지 여부를 판단하고, 만약 예스라면, 모델링 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되는 것으로 결정하는 단계.

[0057] 예를 들어, 사전 설정된 키워드는 전송, 잔액, 계좌 및 패스워드를 포함하며, 시스템은 네 개의 키워드의 오디오 데이터 또는 오디오 특징 데이터를 사전 저장해야 한다. 사용자의 수신된 음성이 "잔액 조회"인 경우, 몇몇 음성 세그먼트(여기서는 논의되지 않은 상이한 분할 알고리즘에 따라 특정 분할 결과는 상이할 수 있음)가 먼저 음향 모델로 분할되어 획득될 수 있으며, 비교시 이들 세그먼트 또는 세그먼트의 조합이 "잔액" 오디오 특징과 매칭하는 부분을 포함한다는 것을 알 수 있다. 그러므로 모델링 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되는 것으로 결정할 수 있다.

[0058] b) 언어 특징 사용:

[0059] 텍스트 단어 세그먼트화는 자연 언어 처리 기술 분야에 속한다. 순수 음향 레벨에서의 단어 세그먼트화와 비교할 때, 텍스트 단어 세그먼트화 기술은 보다 성숙되어 더 양호한 효과를 갖는다. 또한, 동일한 발음이 여러 종류의 텍스트에 대응할 수 있으므로, 매칭 동안 텍스트 내용을 직접 비교함으로써 더 양호한 효과가 달성될 수 있음을 이해할 수 있다. 본 출원에서, 사용자 음성 정보의 인식은 음성 인식 기술과 텍스트 단어 세그먼트화 기술의 조합을 직접 사용하여 구현될 수 있다. 특정 해결책은 다음 단계들을 포함한다:

[0060] 대응하는 텍스트 정보를 획득하기 위해 음성 정보에 대해 음성 인식을 수행하는 단계;

[0061] 적어도 하나의 텍스트 세그먼트를 획득하기 위해 언어 모델을 사용하여 텍스트 정보를 분할하는 단계; 및

[0062] 획득된 텍스트 세그먼트 또는 이들의 조합이 모델링 요구 사항을 충족시키는 사전 설정된 키워드의 텍스트 내용과 일치하는지 여부를 판단하고, 만약 예스라면, 모델링 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되는 것으로 결정하는 단계.

[0063] 상기 두 가지 해결책에 따르면, 당업자는 실제 요구 사항에 따라 유연하게 선택을 할 수 있다. 예를 들어, 상이한 언어의 특성에 따르면, 서양 언어(예를 들어, 영어 및 프랑스어)의 각 단어는 비교적 완전한 의미를 갖지만, 중국어의 각 문자에는 이러한 특성이 없다. 그러므로 음향 특징만을 사용하는 처리 방식은 현재 서양 언어를 처리하는 데 더 적합하다. 중국어는 언어 특징을 참조하여 처리되도록 제안된다. 당연히, 본 출원은 음성 인식 알고리즘 또는 단어 세그먼트화 알고리즘에 중점을 둔 것이 아니라, 임의의 기술적 수단이 대응하는 요구 사항을 만족시킬 수 있는 한 본 출원에 적용될 수 있다는 것을 알아야 한다.

[0064] S102에서 모델링 요구 사항을 충족시키는 사전 설정된 키워드가 사용자 음성 정보에 포함되어있는 것으로 확인되면, S103에서 대응하는 음성 세그먼트는 음성 정보로부터 추가로 캡처될 것이다. S102에서 음성 세그먼트에 기초하여 판단이 수행되면, 음성 세그먼트는 음성 세그먼트의 매칭 결과에 따라 직접 캡처된다. S102에서 텍스트 세그먼트에 기초하여 판단이 수행되면, 음성 세그먼트는 텍스트 세그먼트의 매칭 결과로부터 역으로 추론함으로써 획득된 대응하는 음성 세그먼트의 위치에 따라 캡처된다. 예를 들어, 사용자가 "나는 계좌의 잔액을 확인하고 싶다"라고 말했고, 여기서 "계좌" 및 "잔액"은 모델링 요구 사항을 충족시키는 사전 설정된 키워드이다.

전체 음성 정보에서 두 단어의 상대 위치는 S102의 처리 절차에 따라 결정될 수 있으며, 그 다음 두 단어 "계좌" 및 "잔액"의 오디오 세그먼트가 캡처된다.

- [0065] S104에서, S103에서 캡처된 오디오 세그먼트를 사용하여 성문 모델링이 수행된다. 특정 모델링 방법은 본 출원과 관련이 없으므로, 여기서는 상세히 설명하지 않는다. 마지막으로, 모델링 결과는 S101에서 결정된 화자 식별자와 연관되어 저장되어 사용자의 성문 모델 라이브러리를 형성한다.
- [0066] 상기 방법을 적용함으로써, 사용자가 다른 음성 기능을 사용할 때, 성문 모델링은 "부수적으로" 수행될 수 있으며, 따라서 사용자에게 의한 사용을 용이하게 한다. 또한 "키워드"를 사용하는 것은, "텍스트 종속적" 모델링 요구 사항을 충족시키고 높은 모델링 성공률을 보장할 수 있다.
- [0067] 또한, 상기 방법은 처음으로 성문 정보의 등록에 적용되는 것 외에도 성문 모델의 업데이트에도 또한 적용될 수 있다. 특정 업데이트 전략은 다음 두 가지 관점에서 고려될 수 있다:
- [0068] 1) 모델 개선:
- [0069] 성문 모델을 구축하는 상기 절차는 여러 개의 기본 키워드를 지원하며, 사용자는 한 번의 음성 상호 작용 동안 반드시 모든 키워드를 말할 수 있는 것은 아니다. 그러나 사용자의 음성 상호 작용은 정상적인 행동이므로, 처음으로 성문 모델의 구축이 완료된 이후에, 사용자와의 후속 상호 작용 절차에서 사용자의 음성 정보를 다시 수집할 수 있다. 새롭게 수집된 음성 정보가 이전 모델에 포함되지 않은 키워드를 포함하면, 새로운 키워드가 기존의 성문 모델을 보완하기 위해 사용될 수 있다.
- [0070] 구체적으로, 특정 사용자에게 대해, 사용자의 성문 모델이 현재 존재한다고 가정하면, S102에서, "모델링 요구 사항을 충족시키는 사전 설정된 키워드"는 "사용자의 현재 성문 모델에 포함되지 않은 사전 설정된 키워드"로 또한 정의될 수 있다.
- [0071] 예를 들어, 시스템에 의해 지정된 키워드는: "전송", "잔액", "계좌" 및 "패스워드"를 포함한다. 사용자 X가 처음으로 성문 정보를 등록할 때, "계좌" 및 "잔액"이라는 두 개의 키워드에 대한 성문 모델이 각각 사용자에게 대해 확립되었다. 그 후, 사용자 X의 음성 정보 "나는 계좌의 패스워드를 변경하고 싶다"가 추가로 수집된다. S102에 따르면, "계좌" 및 "패스워드"가 모두 사전 설정된 키워드에 속하는 것으로 결정될 수 있으며, 여기서 "패스워드"는 사용자 X의 현재의 성문 모델에 포함되지 않은 사전 설정된 키워드이다. 그러므로 성문 모델링은 그 뒤에 "패스워드"를 사용하여 수행되고, 모델링 결과는 사용자 X의 성문 모델 라이브러리에 추가된다.
- [0072] 이 방법을 적용함으로써, 다양한 특정 요구 사항을 충족하기 위해 사용자의 성문 모델이 지속적으로 개선될 수 있다. 더욱이, 시스템이 보안 성능에 대한 더 높은 요구 사항을 갖는다면, 키워드가 적용되기 전에 사용자의 성문 모델에 포함된 키워드의 수가 특정 수보다 커야한다는 것이 명시될 수 있다. 음성 정보 수집을 한 번만 수행하여 사용자의 성문 모델링을 구현할 수 없는 경우, 본 실시예의 방법이 채택되어 사용자의 음성 정보를 여러 번 수집함으로써 사용자의 성문 모델링을 완료할 수 있다.
- [0073] 2) 모델 훈련:
- [0074] 생체 인식 기술 및 기계 학습 기술은 밀접한 관련이 있다. 많은 애플리케이션 시나리오에서, 모델은 일반적으로 샘플을 여러 번 수집함으로써 훈련되어, 인식 정확도 및 내결함성과 같은 여러 측면에서 성능이 개선된다. 본 출원의 해결책을 위해, 동일한 키워드에 대해 동일한 사용자의 음성 정보를 여러 번 수집함으로써 성문 모델의 훈련이 또한 달성될 수 있다.
- [0075] 여기의 요구 사항은 1)의 요구 사항과 반대인 것을 볼 수 있다. 구체적으로, 동일한 키워드에 대해 성문 샘플의 여러 번 수집을 구현하기 위해, S102에서, "모델링 요구 사항을 충족시키는 사전 설정된 키워드"는 "사용자의 현재 성문 모델에 포함되어 있는 사전 설정된 키워드"로 추가로 정의되어야 한다.
- [0076] 예를 들어, 시스템에 의해 지정된 키워드는: "전송", "잔액", "계좌" 및 "패스워드"를 포함한다. 사용자 X가 처음으로 성문 정보를 등록할 때, "계좌" 및 "잔액"이라는 두 개의 키워드에 대한 성문 모델이 각각 사용자에게 대해 확립되었다. 후속 음성 정보 수집 프로세스에서, "계좌" 및 "잔액"이라는 두 개의 키워드에만 집중할 수 있어, 두 개의 키워드의 성문 모델은 지속적으로 훈련된다.
- [0077] 실제 애플리케이션에서, 사용자의 음성 정보를 무제한으로 수집하는 것을 피하기 위해 몇몇 제한 조건이 또한 설정될 수 있다. 예를 들어, 키워드에 대해 수집된 성문 샘플의 수가 임계값에 도달하면, 그 뒤에 그 키워드에 대한 성문 샘플이 수집되지 않을 것이다.

- [0078] 또한, 특정 애플리케이션 요구 사항 및 선택된 훈련 알고리즘에 따라, 여러 번 수집된 샘플은 평균화 및 동시 보유와 같은 방법을 사용하여 처리될 수 있으며, 이는 본 출원에서 제한될 필요는 없다.
- [0079] 상기 두 가지 업데이트 전략은 독립적으로 또는 조합하여 구현될 수 있다. 실제로, 거시적인 관점에서, "모델링 요구 사항을 충족시키는 사전 설정된 키워드"가 정의되지 않더라도 두 기능의 구현에는 영향을 미치지 않을 것이다. 즉, S102가 수행될 때마다, 이 음성 정보에 포함된 모든 키워드가 가능한 한 많이 획득된다. 이전 모델에는 포함되지 않은 키워드에 대해 새로운 성문 모델이 확립되고 이전 모델에 포함된 키워드는 이전 모델을 학습하는 데 사용된다. 또한, 상기 두 기능 중 하나가 구현을 위해 선택될 수 있다. 본 출원의 해결책을 적용하는 것은 사용자에게 어떠한 문제도 일으키지 않고 학습을 가능하게 하여 점차적으로 사용자의 성문 모델을 완성하는 것이 중요하다.
- [0080] 본 출원에 의해 제공되는 사용자 성문 모델을 구축하기 위한 방법이 위에 소개되었다. 유사한 개념에 기초하여, 본 출원은 또한 성문 기반 사용자 신원 확인 방법을 제공한다. 도 2를 참조하면, 상기 방법은 다음 단계들을 포함할 수 있다:
- [0081] S201: 사용자에게 의해 입력된 음성 정보를 수신한다.
- [0082] S202: 확인 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되는지 여부를 판단한다.
- [0083] S203: 판단 결과가 예스라면, 확인 요구 사항을 충족시키는 사전 설정된 키워드에 대응하는 음성 세그먼트를 음성 정보로부터 캡처한다.
- [0084] S204: 음성 세그먼트의 성문 특징을 추출한다.
- [0085] S205: 성문 특징 및 사용자의 사전 구축된 성문 모델을 사용하여 사용자의 신원을 확인한다.
- [0086] S201 내지 S203은 기술적 구현의 관점에서 기본적으로 S101 내지 S103과 동일하고, 차이점은 단지 상이한 애플리케이션 시나리오에 있다. S204 내지 S205는 통상적인 성문 확인 단계이며, 상세하게 소개되지 않을 것이다. 본 출원의 해결책과 종래의 성문 확인 해결책 사이의 차이는 다음에서 강조된다.
- [0087] 본 출원의 배경 기술의 소개에 따르면, 성문 인식은 주로 등록 및 적용을 포함하는 두 가지 단계를 포함하고, 등록 단계는 적용 섹션에 필요한 데이터 준비를 제공한다는 것을 알 수 있다. 본 출원의 방법을 적용함으로써 구현되는 성문 모델링은 주로 다양한 애플리케이션 요구 사항에 적용 가능하다. 또한, 본 출원의 해결책의 특성에 따라, 일부 특정 애플리케이션 요구 사항이 또한 충족될 수 있다.
- [0088] 엄밀히 말하면, 성문 인식을 포함한 생체 특징 인식 기술의 애플리케이션 시나리오는 두 가지 범주: 확인 및 식별로 더 세분될 수 있다. 확인은 사용자가 지문 잠금 장치, 성문 잠금 장치 등과 같이 자신이 선언한 신원인지 여부를 인증하는 것을 말한다. 식별은 사용자의 신원을 결정하는 것을 말하며, 예를 들어 공공 보안 기관은 지문 또는 범죄 현장 기록을 조사하여 범죄자를 결정한다. 확인은 식별보다 요구 사항이 적다는 것을 쉽게 알 수 있다.
- [0089] 본 출원에서 제공되는 기술적 해결책은 "확인" 요구 사항에 기초하여 제안되며, 통상적인 성문 확인과 비교하여 적어도 다음의 특성을 갖는다.
- [0090] 첫째, 본 출원에 의해 제공되는 해결책은 "텍스트 종속적" 성문 기술을 기반으로 하지만, 해결책이 적용될 때, 사용자는 의도적으로 "특정 내용"을 말할 필요가 없고, 시스템도 또한 의도적으로 사용자를 자극하거나 안내할 필요가 없다. 이것은 사용자가 다른 작업을 수행할 때 상기 요구 사항이 모두 부수적으로 충족될 수 있기 때문이다.
- [0091] 예를 들어, 은행의 셀프 서비스 음성 서비스 시스템에서, 사용자가 서비스 번호로 전화를 걸고 셀프 서비스 음성 프롬프트에 따라 "한 문장으로 문제를 설명하기" 섹션에서 "잔액 조회"를 말하면, 시스템은 정보에 따라 음성 인식, 의미 인식, 요구 사항 분석 및 기타 처리를 수행할 수 있으며, 서비스가 개인 프라이버시와 관련이 있으며 사용자의 신원을 추가 인증해야한다고 결정할 수 있다. 종래의 처리 방식은, "당신의 계좌 패스워드를 입력하고 # 키를 눌러 종료하십시오"라는 프롬프트가 사용자에게 제공되고, 그러면 사용자는 프롬프트에 따라 동작하여 확인을 완료하는 것일 수 있다. 본 출원에서 제공되는 해결책을 사용함으로써, 키워드 "잔액"에 대한 모델링이 사용자에게 대해 이전에 완료된 경우, 사용자가 "잔액 조회"를 말한 후에, 확인은 성문을 사용하여 구현될 수 있고, 사용자가 패스워드를 입력할 필요가 없다. 의도적으로 "특정 내용"을 말할 필요가 없는 경우 성문 확인의 편리함을 누릴 수 있다.

- [0092] 다음으로, 사용자가 말하는 특정 내용이 불확실하지만, "키워드"의 특성에 따라, 사용자가 실제로 이러한 키워드를 말할 확률이 높으며, 키워드의 수가 증가함에 따라 이 확률은 명백하게 향상될 수 있다. 성문 확인이 실패하더라도, 통상적인 확인 방식이 추가로 사용될 수 있다. 예를 들어, 상기 예에서 패스워드의 수동 입력이 사용될 수 있으며, 이것은 사용자에게 추가적인 문제를 일으키지 않을 것이다.
- [0093] 또한, S203에서 "확인 요구 사항을 충족시키는 사전 설정된 키워드"를 정의함으로써, 시스템 성능은 더욱 개선될 수 있거나 특정 요구 사항은 더욱 충족될 수 있다. 예를 들어, 시스템이 성문 추출 또는 매칭과 같은 후속적으로 무효한 처리를 수행하는 것을 방지하기 위해, "확인 요구 사항을 충족시키는 사전 설정된 키워드"는 사용자의 현재 성문 모델에 포함되어 있는 사전 설정된 키워드로서 정의될 수 있다. 또한, 지불 거래와 같은 높은 보안 요구 사항을 갖는 애플리케이션 시나리오의 경우, 사용자 음성에 포함된 키워드의 수가 또한 "확인 요구 사항을 충족시키는 사전 설정된 키워드"에 정의될 수도 있다. 키워드 수 요구 사항이 충족되지 않으면, 이 방식은 충분히 안전하지 않은 것으로 간주되며, 이 시점에서 성문 확인 방식은 포기되고 다른 보안 방식이 대신 사용될 수 있다.
- [0094] 마지막으로, 본 출원에 의해 제공되는 성문 모델링 방법 및 성문 확인 방법은 동일한 원리에 기초하고, 따라서, 두 가지 해결책이 다양한 방식으로 통합될 수 있다는 것을 이해할 수 있다. 예를 들어, 확인 절차 후에, 이 확인에 사용된 사용자의 음성 정보는 여전히 성문 모델을 완성하는 데 사용될 수 있다. 통상적인 적용에서, 먼저 이 확인에 사용된 음성 정보가 사용자의 현재 성문 모델에 포함되지 않은 사전 설정된 키워드를 포함하는지 여부를 판단하고, 만약 예스라면, 키워드에 대응하는 음성 세그먼트를 음성 정보로부터 캡처하고, 음성 세그먼트를 사용하여 사용자의 성문 모델을 업데이트한다. 특정 업데이트 방식은 이전의 실시예를 참조하여 획득될 수 있으며, 여기에서 반복되지 않을 것이다. 또한, 기술 절차의 관점에서, 신원 확인이 성공했는지 여부는 성문 모델 업데이트에 영향을 미치지 않는다. 당연히, 보안의 개선이 고려된다면, 사용자 신원 확인이 성공할 때, 성문 모델을 업데이트하는 작업이 추가로 트리거될 수 있다. 또한, "신원 확인 성공"은 성문 신원 확인으로 한정되지 않는다.
- [0095] 요약하면, 종래 기술과 비교하여, 본 출원은 주로 다음과 같은 점에서 유리하다: 사용자 성문 정보의 등록 및 확인이 사용자가 다른 음성 기반 작업을 수행할 때 부수적으로 구현될 수 있어 사용자에게 의한 사용을 용이하게 한다. 또한, 특정 애플리케이션 시나리오에 대한 특정 키워드를 설정함으로써, "텍스트 종속적"의 장점을 유지할 수 있으며, 원래의 불확실한 애플리케이션 시나리오에서 더 높은 성공률이 또한 보장될 수 있다.
- [0096] 상기 방법 실시예에 대응하여, 본 출원은 사용자 성문 모델을 구축하기 위한 장치를 더 제공한다. 도 3를 참조하면, 상기 장치는:
- [0097] 사용자에게 의해 입력된 음성 정보를 수신하도록 구성된 음성 정보 수신 모듈(110);
- [0098] 모델링 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되는지 여부를 판단하도록 구성된 모델링 키워드 판단 모듈(120);
- [0099] 상기 판단 결과가 예스인 경우, 음성 정보로부터, 모델링 요구 사항을 충족시키는 사전 설정된 키워드에 대응하는 음성 세그먼트를 캡처하도록 구성된 음성 세그먼트 캡처 모듈(130); 및
- [0100] 음성 세그먼트를 사용하여 사용자에게 대한 성문 모델을 구축하도록 구성된 성문 모델 구축 모듈(140)을 포함할 수 있다.
- [0101] 본 출원의 특정 구현예에 따르면, 모델링 키워드 판단 모듈(120)은 구체적으로:
- [0102] 사용자의 성문 모델이 현재 이미 존재하는 경우, 성문 모델에 포함되지 않은 사전 설정된 키워드가 음성 정보에 포함되는지 여부를 판단하고;
- [0103] 또는
- [0104] 사용자의 성문 모델이 현재 이미 존재하는 경우, 성문 모델에 이미 포함되어 있는 사전 설정된 키워드가 음성 정보에 포함되는지 여부를 판단하도록 구성될 수 있다.
- [0105] 본 출원의 특정 구현예에 따르면, 성문 모델 구축 모듈(140)은 구체적으로:
- [0106] 사용자의 성문 모델이 현재 이미 존재하는 경우, 음성 세그먼트를 사용하여 성문 모델을 업데이트하도록 구성될 수 있다.

- [0107] 본 출원의 특정 구현예에 따르면, 모델링 키워드 판단 모듈(120)은 구체적으로:
- [0108] 적어도 하나의 음성 세그먼트를 획득하기 위해 음향 모델을 사용하여 음성 정보를 분할하고;
- [0109] 획득된 음성 세그먼트 또는 이들의 조합이 모델링 요구 사항을 충족시키는 사전 설정된 키워드의 오디오 특징과 매칭하는지 여부를 판단하고, 만약 예스라면, 모델링 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되는 것으로 결정하도록 구성될 수 있다.
- [0110] 또는
- [0111] 모델링 키워드 판단 모듈(120)은 또한 구체적으로:
- [0112] 대응하는 텍스트 정보를 획득하기 위해 음성 정보에 대해 음성 인식을 수행하고;
- [0113] 적어도 하나의 텍스트 세그먼트를 획득하기 위해 언어 모델을 사용하여 텍스트 정보를 분할하며;
- [0114] 획득된 텍스트 세그먼트 또는 이들의 조합이 모델링 요구 사항을 충족시키는 사전 설정된 키워드의 텍스트 내용과 일치하는지 여부를 판단하고, 만약 예스라면, 모델링 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되는 것으로 결정하도록 구성될 수 있다.
- [0115] 상기 방법 실시예에 대응하여, 본 출원은 또한 성문 기반 사용자 신원 확인 장치를 제공한다. 도 4를 참조하면, 상기 장치는:
- [0116] 사용자에게 의해 입력된 음성 정보를 수신하도록 구성된 음성 정보 수신 모듈(210);
- [0117] 확인 요구 사항을 충족시키는 사전 설정된 키워드가 음성 정보에 포함되는지 여부를 판단하도록 구성된 확인 키워드 판단 모듈(220);
- [0118] 상기 판단 결과가 예스인 경우, 음성 정보로부터, 확인 요구 사항을 충족시키는 사전 설정된 키워드에 대응하는 음성 세그먼트를 캡처하도록 구성된 음성 세그먼트 캡처 모듈(230);
- [0119] 음성 세그먼트의 성문 특징을 추출하도록 구성된 성문 특징 추출 모듈(250); 및
- [0120] 성문 특징 및 사용자의 사전 구축된 성문 모델을 사용하여 사용자의 신원을 확인하도록 구성된 확인 모듈(260)을 포함할 수 있다.
- [0121] 본 출원은 또한 성문 기반 사용자 신원 확인 시스템을 제공하고, 상기 시스템은 상술한 바와 같이 사용자 성문 모델을 구축하기 위한 장치 및 사용자 신원 확인 장치를 포함할 수 있다.
- [0122] 사실, 상기 두 장치에서 일부 모듈에 의해 구현된 기능은 완전히 동일하다. 예를 들면:
- [0123] 음성 정보 수신 모듈(110) 및 음성 정보 수신 모듈(210);
- [0124] 음성 세그먼트 캡처 모듈(130) 및 음성 세그먼트 캡처 모듈(230)이 있고;
- [0125] 확인 키워드 판단 모듈(120)과 확인 키워드 판단 모듈(220)의 차이는 단지 판단이 수행되는 상이한 규칙에 있다. 따라서, 실제 적용 동안, 이들 동일하거나 유사한 모듈은 도 5에 도시된 바와 같이 시스템에서 완전히 다중화될 수 있다.
- [0126] 음성 정보 수신 모듈(310)은 음성 정보 수신 모듈(110)과 음성 정보 수신 모듈(210)의 기능을 통합한다.
- [0127] 키워드 판단 모듈(320)은 확인 키워드 판단 모듈(120)과 확인 키워드 판단 모듈(220)의 기능을 통합한다.
- [0128] 음성 세그먼트 캡처 모듈(330)은 음성 세그먼트 캡처 모듈(130)과 음성 세그먼트 캡처 모듈(230)의 기능을 통합한다.
- [0129] 성문 모델 구축 모듈(340), 성문 특징 추출 모듈(350) 및 확인 모듈(360)의 기능은, 상기 장치의 동일 명칭의 모듈의 기능과 각각 동일하다.
- [0130] 상기 장치 내의 모든 유닛의 기능 및 효과의 구현 프로세스는, 구체적으로 상기 방법에서의 대응하는 단계의 구현 프로세스를 지칭할 수 있으며, 여기에서 반복되지 않는다.
- [0131] 구현 방식에 대한 앞서 말한 설명에 기초하여, 당업자는 본 출원이 필요한 범용 하드웨어 플랫폼과 결합하여 소프트웨어에 의해 구현될 수 있음을 명확히 이해할 수 있다. 이러한 이해에 기초하여, 본 출원의 기술적 해결책 또는 종래 기술에 기여하는 부분은 소프트웨어 제품의 형태로 구현될 수 있다. 소프트웨어 제품은 ROM/RAM, 자

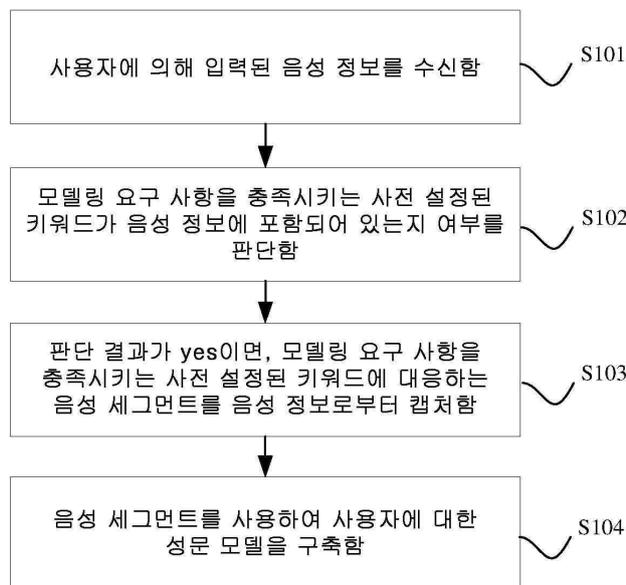
기 디스크 또는 광 디스크와 같은 저장 매체에 저장될 수 있으며, 컴퓨터 디바이스(개인용 컴퓨터, 서버, 네트워크 디바이스 등일 수 있음)가 본 출원의 모든 실시예 또는 실시예들의 특정 부분에서의 방법을 실행할 수 있도록 하는 몇몇 명령어를 포함할 수 있다.

[0132] 본 명세서의 다양한 실시예들은 점진적인 방식으로 설명된다. 실시예들 사이에서 동일하거나 유사한 부분은 서로 참조될 수 있다. 각각의 실시예에서, 다른 실시예들과의 상이한 부분이 집중되고 기술된다. 특히, 장치 또는 시스템 실시예는 기본적으로 방법 실시예와 유사하므로 간단히 설명되고, 관련 부분에 대해서는 방법 실시예에서의 부분에 대한 설명을 참조할 수 있다. 상술한 장치 또는 시스템 실시예는 단지 예시적인 것으로서, 분리된 부분으로 나타난 모듈은 물리적으로 분리될 수도 있고 분리되지 않을 수도 있다. 본 출원의 구현 동안, 모듈들의 기능들은 소프트웨어 및/또는 하드웨어의 동일한 하나 이상의 조각들에 구현될 수 있다. 이 실시예의 해결책의 목적은 실제 요구 사항에 따라 모듈의 일부 또는 전부를 선택함으로써 구현될 수 있다. 당업자는 창조적인 노력 없이 본 출원을 이해하고 구현할 수 있다.

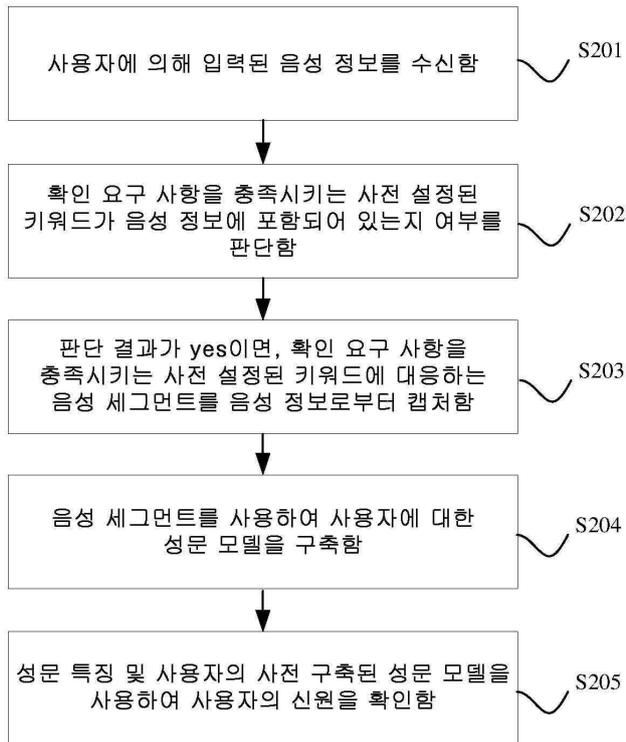
[0133] 본 출원의 특정 구현예가 위에 설명되었다. 당업자라면, 본 출원의 원리를 벗어나지 않고 여러 가지 개선 및 수정을 이룰 수 있으며, 이러한 개선 및 수정은 또한 본 출원의 보호 범위 내에 속하는 것으로 고려되어야 한다는 것을 알아야 한다.

도면

도면1



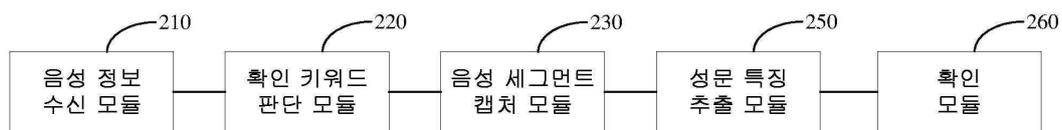
도면2



도면3



도면4



도면5

