(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년02월15일
(11) 등록번호 10-2214888
(24) 등록일자 2021년02월04일

(51) 국제특허분류(Int. Cl.)
G10L 25/84 (2013.01) *G10L 25/21* (2013.01)
(52) CPC특허분류
G10L 25/84 (2013.01)
G10L 25/21 (2013.01)
(21) 출원번호 10-2019-7013519
(22) 출원일자(국제) 2017년09월26일
심사청구일자 2019년05월13일
(85) 번역문제출일자 2019년05월10일
(65) 공개번호 10-2019-0061076
(43) 공개일자 2019년06월04일
(86) 국제출원번호 PCT/CN2017/103489
(87) 국제공개번호 WO 2018/068636
국제공개일자 2018년04월19일
(30) 우선권주장
201610890946.9 2016년10월12일 중국(CN)
(56) 선행기술조사문헌
CN103177722 A*
CN103646649 B*
KR1020110081643 A*
WO2011049516 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 코포레이트 센터
(72) 발명자
지아오 레이
중국 저지앙 311121 항저우 유 항 디스트릭트 웨스트 웬 이 로드 넘버 969 빌딩 3 알리바바 그룹 리갈 디파트먼트 5층
구안 얀추
중국 저지앙 311121 항저우 유 항 디스트릭트 웨스트 웬 이 로드 넘버 969 빌딩 3 알리바바 그룹 리갈 디파트먼트 5층
(뒷면에 계속)
(74) 대리인
김태홍, 김진희

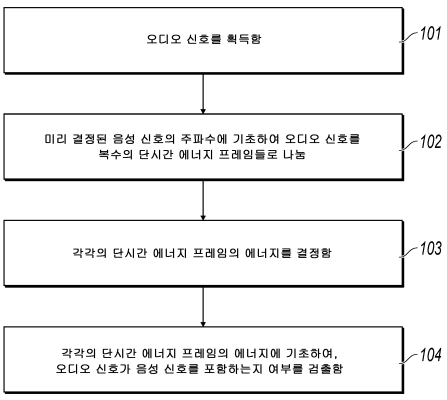
전체 청구항 수 : 총 17 항 심사관 : 노지명

(54) 발명의 명칭 **오디오 신호를 검출하기 위한 방법 및 디바이스**

(57) 요 약

본 출원은, 기존 기술에서의 음성 신호 검출 방법에서 처리 속도가 비교적 낮고 자원 소비가 비교적 높은 문제점을 해소하기 위한 음성 신호 검출 방법 및 장치를 개시한다. 방법은, 오디오 신호를 획득하는 단계, 미리 결정된 음성 신호의 주파수에 기초하여 오디오 신호를 복수의 단시간 에너지 프레임들로 나누는 단계, 각각의 단시간 에너지 프레임의 에너지를 결정하는 단계, 및 각각의 단시간 에너지 프레임의 에너지에 기초하여, 오디오 신호가 음성 신호를 포함하는지 여부를 검출하는 단계를 포함한다.

대 표 도 - 도1



(72) 발명자

젱 시아오둥

중국 저지앙 311121 항저우 유 항 디스트릭트 웨스트 웬 이 로드 넘버 969 빌딩 3 알리바바 그룹 리갈 디파트먼트 5층

린 펭

중국 저지앙 311121 항저우 유 항 디스트릭트 웨스트 웬 이 로드 넘버 969 빌딩 3 알리바바 그룹 리갈 디파트먼트 5층

명 세 서

청구범위

청구항 1

컴퓨터 구현(computer-implemented) 방법에 있어서,

사용자 단말기에 의해, 오디오 신호를 획득하는 단계;

미리 결정된 음성 신호의 최소 주파수에 대한, 상기 오디오 신호의 샘플링 레이트의 비(ratio)를 결정하는 단계;

상기 사용자 단말기에 의해, 상기 비에 기초하여 복수의 샘플들을 포함하는 최대 수의 단시간 에너지 프레임들로 상기 오디오 신호를 나누는 단계;

상기 사용자 단말기에 의해, 각각의 단시간 에너지 프레임의 에너지를 결정하는 단계; 및

상기 사용자 단말기에 의해, 상기 각각의 단시간 에너지 프레임의 에너지에 기초하여 상기 오디오 신호가 음성 신호를 포함하는지 여부를 결정하는 단계

를 포함하는, 컴퓨터 구현 방법.

청구항 2

청구항 1에 있어서, 상기 오디오 신호는 상기 샘플링 레이트로 수집되며 펄스 코드 변조(PCM; pulse code modulation) 포맷으로 이루어지는 것인, 컴퓨터 구현 방법.

청구항 3

청구항 1에 있어서, 상기 획득된 오디오 신호는 PCM 아닌(non-PCM) 포맷으로 이루어지고, 상기 방법은,

상기 오디오 신호를 나누기 전에,

상기 오디오 신호를 펄스 코드 변조(PCM) 포맷으로 변환하는 단계; 및

상기 오디오 신호의 상기 샘플링 레이트를 식별하는 단계

를 더 포함하는 것인, 컴퓨터 구현 방법.

청구항 4

청구항 1에 있어서, 상기 각각의 단시간 에너지 프레임의 에너지는 각각의 단시간 에너지 프레임 내의 각각의 샘플링 포인트와 연관된 에너지의 합이고, 상기 각각의 샘플링 포인트와 연관된 에너지는 상기 단시간 에너지 프레임 내의 샘플링 포인트에 대응하는 상기 오디오 신호의 진폭에 기초하여 결정되는 것인, 컴퓨터 구현 방법.

청구항 5

청구항 1에 있어서, 상기 오디오 신호가 음성 신호를 포함하는지 여부를 결정하는 단계는,

복수의 고에너지 프레임들을 결정하는 단계 - 상기 복수의 고에너지 프레임들의 각각의 고에너지 프레임은, 자신의 에너지가 미리 결정된 문턱값보다 더 큰 것인 단시간 에너지 프레임임 - ;

상기 오디오 신호에 포함된 단시간 에너지 프레임들의 수에 대한, 상기 복수의 고에너지 프레임들의 수의 비로 표현되는 고에너지 프레임 비를 결정하는 단계;

상기 고에너지 프레임 비가 미리 결정된 값보다 더 큰지 여부를 결정하는 단계;

상기 고에너지 프레임 비가 상기 미리 결정된 값보다 더 크다고 결정되는 경우,

상기 오디오 신호가 음성 신호를 포함한다고 결정하는 단계; 또는

상기 고에너지 프레임 비가 상기 미리 결정된 값보다 더 크지 않다고 결정되는 경우,

상기 오디오 신호가 음성 신호를 포함하지 않는다고 결정하는 단계

를 포함하는 것인, 컴퓨터 구현 방법.

청구항 6

청구항 5에 있어서, 상기 고에너지 프레임 비가 상기 미리 결정된 값보다 더 크다고 결정되고, 상기 방법은,

상기 오디오 신호에 포함된 단시간 에너지 프레임들로부터, 미리 결정된 수의 연속 단시간 에너지 프레임들이 존재하는지 여부를 결정하는 단계 - 상기 미리 결정된 수의 연속 단시간 에너지 프레임의 각각은 상기 미리 결정된 문턱값보다 더 큰 에너지를 가짐 - ;

그러한 경우, 상기 오디오 신호가 음성 신호를 포함한다고 결정하는 단계; 또는

그렇지 않은 경우, 상기 오디오 신호가 음성 신호를 포함하지 않는다고 결정하는 단계

를 더 포함하는 것인, 컴퓨터 구현 방법.

청구항 7

하나 이상의 명령어를 저장한 비일시적 컴퓨터 판독가능 매체에 있어서, 상기 하나 이상의 명령어는,

사용자 단말기에 의해, 오디오 신호를 획득하는 동작;

미리 결정된 음성 신호의 최소 주파수에 대한, 상기 오디오 신호의 샘플링 레이트의 비를 결정하는 동작;

상기 사용자 단말기에 의해, 상기 비에 기초하여 복수의 샘플들을 포함하는 최대 수의 단시간 에너지 프레임들로 상기 오디오 신호를 나누는 동작;

상기 사용자 단말기에 의해, 각각의 단시간 에너지 프레임의 에너지를 결정하는 동작; 및

상기 사용자 단말기에 의해, 상기 각각의 단시간 에너지 프레임의 에너지에 기초하여 상기 오디오 신호가 음성 신호를 포함하는지 여부를 결정하는 동작

을 포함하는 동작들을 수행하도록 컴퓨터 시스템에 의해 실행가능한 것인, 비일시적 컴퓨터 판독가능 매체.

청구항 8

청구항 7에 있어서, 상기 오디오 신호는 상기 샘플링 레이트로 수집되며 펄스 코드 변조(PCM) 포맷으로 이루어지는 것인, 비일시적 컴퓨터 판독가능 매체.

청구항 9

청구항 7에 있어서, 상기 획득된 오디오 신호는 PCM 아닌 포맷으로 이루어지고, 상기 동작들은,

상기 오디오 신호를 나누기 전에,

상기 오디오 신호를 펄스 코드 변조(PCM) 포맷으로 변환하는 동작; 및

상기 오디오 신호의 상기 샘플링 레이트를 식별하는 동작

을 더 포함하는 것인, 비일시적 컴퓨터 판독가능 매체.

청구항 10

청구항 7에 있어서, 상기 각각의 단시간 에너지 프레임의 에너지는 각각의 단시간 에너지 프레임 내의 각각의 샘플링 포인트와 연관된 에너지의 합이고, 상기 각각의 샘플링 포인트와 연관된 에너지는 상기 단시간 에너지 프레임 내의 샘플링 포인트에 대응하는 상기 오디오 신호의 진폭에 기초하여 결정되는 것인, 비일시적 컴퓨터 판독가능 매체.

청구항 11

청구항 7에 있어서, 상기 오디오 신호가 음성 신호를 포함하는지 여부를 결정하는 동작은,

복수의 고에너지 프레임들을 결정하는 동작 - 상기 복수의 고에너지 프레임들의 각각의 고에너지 프레임은, 자신의 에너지가 미리 결정된 문턱값보다 더 큰 것인 단시간 에너지 프레임임 - ;

상기 오디오 신호에 포함된 단시간 에너지 프레임들의 수에 대한, 상기 복수의 고에너지 프레임들의 수의 비로 표현되는 고에너지 프레임 비를 결정하는 동작;

상기 고에너지 프레임 비가 미리 결정된 값보다 더 큰지 여부를 결정하는 동작;

상기 고에너지 프레임 비가 상기 미리 결정된 값보다 더 크다고 결정되는 경우,

상기 오디오 신호가 음성 신호를 포함한다고 결정하는 동작; 또는

상기 고에너지 프레임 비가 상기 미리 결정된 값보다 더 크지 않다고 결정되는 경우,

상기 오디오 신호가 음성 신호를 포함하지 않는다고 결정하는 동작

을 포함하는 것인, 비일시적 컴퓨터 판독가능 매체.

청구항 12

청구항 11에 있어서, 상기 고에너지 프레임 비가 상기 미리 결정된 값보다 더 크다고 결정되고, 상기 동작들은,

상기 오디오 신호에 포함된 단시간 에너지 프레임들로부터, 미리 결정된 수의 연속 단시간 에너지 프레임들이 존재하는지 여부를 결정하는 동작 - 상기 미리 결정된 수의 연속 단시간 에너지 프레임의 각각은 상기 미리 결정된 문턱값보다 더 큰 에너지를 가짐 - ;

그러한 경우, 상기 오디오 신호가 음성 신호를 포함한다고 결정하는 동작; 또는

그렇지 않은 경우, 상기 오디오 신호가 음성 신호를 포함하지 않는다고 결정하는 동작

을 더 포함하는 것인, 비일시적 컴퓨터 판독가능 매체.

청구항 13

컴퓨터 구현 시스템에 있어서,

하나 이상의 컴퓨터; 및

상기 하나 이상의 컴퓨터와 상호동작가능하게 결합되며, 하나 이상의 명령어를 저장한 유형의(tangible) 비일시적 기계 판독가능 매체를 갖는 하나 이상의 컴퓨터 메모리 디바이스를 포함하고,

상기 하나 이상의 명령어는 상기 하나 이상의 컴퓨터에 의해 실행될 때,

사용자 단말기에 의해, 오디오 신호를 획득하는 동작;

미리 결정된 음성 신호의 최소 주파수에 대한, 상기 오디오 신호의 샘플링 레이트의 비를 결정하는 동작;

상기 사용자 단말기에 의해, 상기 비에 기초하여 복수의 샘플들을 포함하는 최대 수의 단시간 에너지 프레임들로 상기 오디오 신호를 나누는 동작;

상기 사용자 단말기에 의해, 각각의 단시간 에너지 프레임의 에너지를 결정하는 동작; 및

상기 사용자 단말기에 의해, 상기 각각의 단시간 에너지 프레임의 에너지에 기초하여 상기 오디오 신호가 음성 신호를 포함하는지 여부를 결정하는 동작

을 포함하는 하나 이상의 동작을 수행하는 것인, 컴퓨터 구현 시스템.

청구항 14

청구항 13에 있어서, 상기 오디오 신호는 상기 샘플링 레이트로 수집되며 펄스 코드 변조(PCM) 포맷으로 이루어지는 것인, 컴퓨터 구현 시스템.

청구항 15

청구항 13에 있어서, 상기 획득된 오디오 신호는 PCM 아닌 포맷으로 이루어지고, 상기 동작들은,

상기 오디오 신호를 나누기 전에,

상기 오디오 신호를 펄스 코드 변조(PCM) 포맷으로 변환하는 동작; 및

상기 오디오 신호의 상기 샘플링 레이트를 식별하는 동작

을 더 포함하는 것인, 컴퓨터 구현 시스템.

청구항 16

청구항 13에 있어서, 상기 각각의 단시간 에너지 프레임의 에너지는 각각의 단시간 에너지 프레임 내의 각각의 샘플링 포인트와 연관된 에너지의 합이고, 상기 각각의 샘플링 포인트와 연관된 에너지는 상기 단시간 에너지 프레임 내의 샘플링 포인트에 대응하는 상기 오디오 신호의 진폭에 기초하여 결정되는 것인, 컴퓨터 구현 시스템.

청구항 17

청구항 13에 있어서, 상기 오디오 신호가 음성 신호를 포함하는지 여부를 결정하는 동작은,

복수의 고에너지 프레임들을 결정하는 동작 - 상기 복수의 고에너지 프레임들의 각각의 고에너지 프레임은, 자신의 에너지가 미리 결정된 문턱값보다 더 큰 것인 단시간 에너지 프레임임 - ;

상기 오디오 신호에 포함된 단시간 에너지 프레임들의 수에 대한, 상기 복수의 고에너지 프레임들의 수의 비로 표현되는 고에너지 프레임 비를 결정하는 동작;

상기 고에너지 프레임 비가 미리 결정된 값보다 더 큰지 여부를 결정하는 동작;

상기 고에너지 프레임 비가 상기 미리 결정된 값보다 더 크다고 결정되는 경우,

상기 오디오 신호가 음성 신호를 포함한다고 결정하는 동작; 또는

상기 고에너지 프레임 비가 상기 미리 결정된 값보다 더 크지 않다고 결정되는 경우,

상기 오디오 신호가 음성 신호를 포함하지 않는다고 결정하는 동작

을 포함하는 것인, 컴퓨터 구현 시스템.

발명의 설명

기 술 분 야

[0001] 본 출원은 컴퓨터 기술 분야에 관한 것으로, 보다 상세하게는 음성 신호 검출 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 실제 생활에서 사람들은 종종 스마트 디바이스(예를 들어, 스마트폰 및 태블릿 컴퓨터)를 사용하여 음성 메시지를 보낸다. 그러나, 음성 메시지를 보내기 위해 스마트 디바이스를 사용할 때, 사람들은 보통 음성 메시지를 보내기 전에 스마트 디바이스의 화면 상의 시작 버튼이나 종료 버튼을 탭하여야 하며, 이 탭 동작은 사용자에게 많은 불편함을 야기한다.

[0003] 사용자가 버튼을 탭하도록 요구하는 일 없이 음성 메시지 보내기를 완료하기 위해, 스마트 디바이스는 연속으로 또는 미리 결정된 주기에 기초하여 녹음을 수행하여야 하고, 획득된 오디오 신호가 음성 신호를 포함하는지 여부를 결정하여야 한다. 획득된 오디오 신호가 음성 신호를 포함하는 경우, 스마트 디바이스는 음성 신호를 추출한 다음, 그 후에 음성 신호를 처리하여 보낸다. 그리하여, 스마트 디바이스는 음성 메시지 보내기를 완료한다.

[0004] 기존의 기술에서는, 획득된 오디오 신호가 음성 신호를 포함하는지 여부를 검출하기 위해 보통 이중 문턱값(dual-threshold) 방법, 자동상관(autocorrelation) 최대값에 기초한 검출 방법, 및 웨이브릿 변환(wavelet transformation) 기반의 검출 방법과 같은 음성 신호 검출 방법이 사용된다. 그러나, 이들 방법에서는, 오디오

정보의 주파수 특성이 보통 퓨리에 변환(Fourier Transform)과 같은 복잡한 계산을 통해 획득되고, 또한, 주파수 특성에 기초하여, 오디오 정보가 음성 신호를 포함하는지 여부가 결정된다. 따라서, 비교적 많은 양의 버퍼 데이터가 계산되어야 하고, 메모리 사용은 비교적 높으며, 그리하여 비교적 많은 양의 계산이 요구되고 처리 속도는 비교적 낮으며 전력 소비는 비교적 크다.

발명의 내용

[0005] 본 출원의 구현은, 기존 기술에서의 음성 신호 검출 방법에서 처리 속도가 비교적 낮고 자원 소비가 비교적 높다는 문제점을 해소하기 위한 음성 신호 검출 방법 및 장치를 제공한다.

[0006] 본 출원의 구현에서는 다음의 기술적 해결책이 사용된다:

[0007] 음성 신호 검출 방법이 제공되며, 상기 방법은, 오디오 신호를 획득하는 단계; 미리 결정된 음성 신호의 주파수에 기초하여 상기 오디오 신호를 복수의 단시간 에너지 프레임들로 나누는 단계; 각각의 단시간 에너지 프레임의 에너지를 결정하는 단계; 및 상기 각각의 단시간 에너지 프레임의 에너지에 기초하여, 상기 오디오 신호가 음성 신호를 포함하는지 여부를 검출하는 단계를 포함한다.

[0008] 음성 신호 장치가 제공되며, 상기 장치는, 오디오 신호를 획득하도록 구성된 획득 모듈; 미리 결정된 음성 신호의 주파수에 기초하여 상기 오디오 신호를 복수의 단시간 에너지 프레임들로 나누도록 구성된 분할 모듈; 각각의 단시간 에너지 프레임의 에너지를 결정하도록 구성된 결정 모듈; 및 상기 각각의 단시간 에너지 프레임의 에너지에 기초하여, 상기 오디오 신호가 음성 신호를 포함하는지 여부를 검출하도록 구성된 검출 모듈을 포함한다.

[0009] 본 출원의 구현에 사용되는 앞서 기재된 기술적 해결책 중의 적어도 하나는 다음의 유리한 효과들을 가져올 수 있다:

[0010] 기존의 기술에서는, 퓨리에 변환과 같은 복잡한 계산을 통해, 오디오 신호가 음성 신호를 포함하는지 여부가 결정된다. 이와 달리, 본 출원의 구현에 사용되는 음성 신호 검출 방법에서는, 퓨리에 변환과 같은 복잡한 계산이 수행될 필요가 없다. 획득된 오디오 신호는, 미리 결정된 음성 신호의 주파수에 기초하여 복수의 단시간 에너지 프레임들로 나누어지고, 각각의 단시간 에너지 프레임의 에너지가 더 결정되며, 각각의 단시간 에너지 프레임의 에너지에 기초하여, 획득된 오디오 신호가 음성 신호를 포함하는지 여부가 검출될 수 있다. 따라서 본 출원의 구현에서 제공되는 음성 신호 검출 방법에서는, 기존 기술에서의 음성 신호 검출 방법에서 처리 속도가 비교적 낮고 자원 소비가 비교적 높다는 문제점이 해소될 수 있다.

도면의 간단한 설명

[0011] 여기에 기재된 첨부 도면은 본 출원의 부가의 이해를 제공하도록 의도되고 본 출원의 일부를 구성한다. 본 출원의 예시적인 구현 및 이의 설명은 본 출원을 설명하기 위한 것이며, 본 출원의 한정을 구성하지 않는다.

도 1은 본 출원의 구현에 따른 음성 신호 검출 방법을 예시한 흐름도이다.

도 2는 본 출원의 구현에 따른 또다른 음성 신호 검출 방법을 예시한 흐름도이다.

도 3은 본 출원의 구현에 따라 미리 결정된 지속기간의 오디오 신호를 예시한 디스플레이 도면이다.

도 4는 본 출원의 구현에 따라 음성 신호 검출 장치의 구조를 예시한 개략도이다.

발명을 실시하기 위한 구체적인 내용

[0012] 본 출원의 목적, 기술적 해결책, 및 이점을 보다 명확하게 하기 위해, 다음은 본 출원의 구현 및 첨부 도면을 참조하여 본 출원의 기술적 해결책을 명확하고 완전하게 설명한다. 명백하게, 기재된 구현은 본 출원의 구현 전부가 아니라 일부일 뿐이다. 창조적 노력을 들이지 않고서 본 출원의 구현에 기초하여 당해 기술 분야에서의 통상의 지식을 가진 자에 의해 획득되는 모든 다른 구현은 본 출원의 보호 범위 내에 속할 것이다.

[0013] 본 출원의 구현에서 제공되는 기술적 해결책은 첨부 도면을 참조하여 아래에 상세하게 기재된다.

[0014] 기존 기술에서의 음성 신호 검출 방법에서 처리 속도가 비교적 낮고 자원 소비가 비교적 높다는 문제점을 해소하기 위해, 본 출원의 구현은 음성 신호 검출 방법을 제공한다.

[0015] 방법의 실행 주체는, 이동 전화, 태블릿 컴퓨터, 또는 개인용 컴퓨터(PC; Personal Computer)와 같은 사용자 단

말기일 수 있지만, 이에 한정되지 않으며, 이 사용자 단말기 상에서 실행되는 애플리케이션(APP)일 수 있고, 또는 서버와 같은 디바이스일 수 있다.

[0016] 설명을 쉽게 하기 위해, 방법의 실행 주체가 APP인 예가 방법의 구현을 기재하도록 아래에 사용된다. 방법이 APP에 의해 실행되며, 이는 단지 설명을 위한 예일 뿐이고 이 방법에 대한 한정으로서 해석되어서는 안 된다는 것을 이해할 수 있을 것이다.

[0017] 도 1은 방법의 절차의 개략도이다. 방법은 다음 단계들을 포함한다.

[0018] 단계 101: 오디오 신호를 획득한다.

[0019] 오디오 신호는 오디오 수집 디바이스를 사용함으로써 APP에 의해 수집된 오디오 신호일 수 있고, 또는 APP에 의해 수신된 오디오 신호일 수 있고, 예를 들어, 또다른 APP 또는 디바이스에 의해 전송된 오디오 신호일 수 있다. 구현은 본 출원에서 한정되지 않는다. 오디오 신호를 획득한 후에, APP은 오디오 신호를 국부적으로 저장할 수 있다.

[0020] 본 출원은 또한,오디오 신호에 대응하는 샘플링 레이트, 지속기간(duration), 포맷, 사운드 채널 등에 어떠한 제한도 두지 않는다.

[0021] APP이 본 출원의 본 구현에서 제공되는 음성 신호 검출 방법에서 오디오 신호를 획득할 수 있고 획득된 오디오 신호에 대해 음성 신호 검출을 수행할 수 있다면, APP은 챗 APP 또는 결제 APP과 같은 임의의 유형의 APP일 수 있다.

[0022] 단계 102: 미리 결정된 음성 신호의 주파수에 기초하여 오디오 신호를 복수의 단시간 에너지 프레임들로 나눈다.

[0023] 단시간 에너지 프레임은 실제로 단계 101에서 획득된 오디오 신호의 일부이다.

[0024] 구체적으로, 미리 결정된 음성 신호의 주파수에 기초하여 미리 결정된 음성 신호의 주기가 결정될 수 있고, 결정된 주기에 기초하여, 단계 101에서 획득된 오디오 신호는 자신의 대응하는 지속기간이 그 주기인 복수의 단시간 에너지 프레임들로 나누어진다. 예를 들어, 미리 결정된 음성 신호의 주기가 0.01s라고 가정하면, 단계 101에서 획득된 오디오 신호의 지속기간에 기초하여, 오디오 신호는 자신의 지속기간이 0.01s인 여러 개의 단시간 에너지 프레임들로 나누어질 수 있다. 단계 101에서 획득된 오디오 신호가 나누어질 때, 오디오 신호는 대안으로서 실제 조건 및 미리 결정된 음성 신호의 주파수에 기초하여 적어도 2개의 단시간 에너지 프레임으로 나누어질 수 있다는 것을 유의하여야 한다. 그 후의 설명을 쉽게 하기 위해, 본 출원의 본 구현에서 아래의 설명에 대하여 오디오 신호가 복수의 단시간 에너지 프레임들로 나누어지는 예가 사용된다.

[0025] 또한, 단계 101에서 APP이 오디오 수집 디바이스를 사용함으로써 오디오 신호를 수집할 때, 오디오 신호를 수집하는 것은 일반적으로, 디지털 신호를 형성하도록 실제로는 아날로그 신호인 오디오 신호, 즉 펄스 코드 변조(PCM; Pulse Code Modulation) 포맷의 오디오 신호를 특정 샘플링 레이트로 수집하는 것이기 때문에, 오디오 신호는 오디오 신호의 샘플링 레이트 및 미리 결정된 음성 신호의 주파수에 기초하여 복수의 단시간 에너지 프레임들로 더 나누어질 수 있다.

[0026] 구체적으로, 미리 결정된 음성 신호의 주파수에 대한, 오디오 신호의 샘플링 레이트의 비(ratio) m이 결정될 수 있고, 그 다음, 수집된 디지털 오디오 신호 내의 각각의 m 샘플링 포인트는 비 m에 기초하여 하나의 단시간 에너지 프레임으로 그룹핑된다. m이 양의 정수인 경우, 오디오 신호는 m에 기초하여 최대 수의 단시간 에너지 프레임들로 나누어질 수 있고, m이 양의 정수가 아닌 경우, 오디오 신호는 양의 정수로 반올림되는 m에 기초하여 최대 수의 단시간 에너지 프레임들로 나누어질 수 있다. 단계 101에서 획득된 오디오 신호에 포함된 샘플링 포인트의 수가 m의 정수 배가 아닌 경우, 오디오 신호가 최대 수의 단시간 에너지 프레임들로 나누어진 후에, 남은 샘플링 포인트는 폐기될 수 있고, 또는 남은 샘플링 포인트는 대안으로서 후속 프로세싱을 위한 단시간 에너지 프레임으로서 사용될 수 있다. M은 미리 결정된 음성 신호의 주기에서 단계 101에서 획득된 오디오 신호에 포함된 샘플링 포인트의 수를 표시하는데 사용된다.

[0027] 예를 들어, 미리 결정된 음성 신호의 주파수가 82 Hz인 경우, 단계 101에서 획득된 오디오 신호의 지속기간은 1s이고, 샘플링 레이트는 16000 Hz, m=16000/82=195.1이다. m은 여기에서 양의 정수가 아니기 때문에, 195.1은 양의 정수 195로 반올림된다. 오디오 신호의 지속기간 및 샘플링 레이트에 기초하여, 오디오 신호에 포함된 샘플링 포인트의 수가 16000이라고 결정될 수 있다. 오디오 신호에 포함된 샘플링 포인트의 수가 195의 정수 배가 아니기 때문에, 오디오 신호는 82개의 단시간 에너지 프레임들로 나누어지고, 나머지 10개의 샘플링 포인

트는 폐기될 수 있다. 각각의 단시간 에너지 프레임에 포함된 샘플링 포인트의 수는 195이다.

[0028] 단계 101에서 획득된 오디오 신호가 또다른 APP 또는 디바이스에 의해 전송된 수신 오디오 신호일 때, 오디오 신호는 이전의 방법 중의 임의의 하나를 사용함으로써 복수의 단시간 에너지 프레임들로 나누어질 수 있다. 오디오 신호의 포맷이 PCM 포맷이 아닐 수도 있다는 것을 유의하여야 한다. 단시간 에너지 프레임이 오디오 신호의 샘플링 레이트 및 미리 결정된 음성 신호의 주파수에 기초하여 이전의 방법에서 분할을 수행함으로써 획득되는 경우, 수신 오디오 신호는 PCM 포맷의 오디오 신호로 변환되어야 한다. 또한, 오디오 신호가 수신될 때, 오디오 신호의 샘플링 레이트가 식별되어야 한다. 오디오 신호의 샘플링 레이트를 식별하기 위한 방법은 기존의 기술에서의 식별 방법일 수 있다. 여기에서 세부사항은 단순화를 위해 생략된다.

[0029] 단계 103: 각각의 단시간 에너지 프레임의 에너지를 결정한다.

[0030] 본 출원의 본 구현에서, PCM 포맷의 오디오 신호가, 이전의 방법에서, 또한 PCM 포맷인 여러 개의 단시간 에너지 프레임들로 나누어질 때, 단시간 에너지 프레임 내의 각각의 샘플링 포인트에 대응하는 오디오 신호의 진폭에 기초하여 단시간 에너지 프레임의 에너지가 결정될 수 있다. 구체적으로, 각각의 샘플링 포인트의 에너지는 단시간 에너지 프레임 내의 각각의 샘플링 포인트에 대응하는 오디오 신호의 진폭에 기초하여 결정될 수 있고, 그 다음 샘플링 포인트의 에너지가 가산된다. 최종적으로 얻은 에너지의 합이 단시간 에너지 프레임의 에너지로서 사용된다.

[0031] 예를 들어, 단시간 에너지 프레임의 에너지는 다음 식; $\text{Energy}=\sum_{i}^{i+n}(A_i[t])^2$ 을 사용함으로써 결정될 수 있으며, 여기에서 i는 오디오 신호의 i번째 샘플링 포인트를 나타내고, n은 단시간 에너지 프레임에 포함된 샘플링 포인트의 수이고, $A_i[t]$는 i번째 샘플링 포인트에 대응하는 오디오 신호의 진폭이고, 단시간 에너지 프레임의 진폭의 값 범위는 -32768 내지 32767이다.

[0032] 또한, 본 출원의 본 구현에서, 계산을 단순화하고 자원을 절약하기 위해, 진폭을 32768로 나눔으로써 얻은 값이 단시간 에너지 프레임의 정규화된 진폭으로서 더 사용될 수 있다. 진폭은 오디오 신호가 수집될 때 획득된다. 단시간 에너지 프레임의 정규화된 진폭의 값 범위는 -1 내지 1이다.

[0033] 단시간 에너지 프레임이 PCM 포맷이 아닌 경우, 진폭 계산 함수는 각각의 순간에서의 단시간 에너지 프레임의 진폭에 기초하여 결정될 수 있고, 함수의 제곱에 대하여 적분이 수행되며, 최종적으로 얻은 적분 결과가 단시간 에너지 프레임의 에너지이다.

[0034] 단계 104: 각각의 단시간 에너지 프레임의 에너지에 기초하여, 오디오 신호가 음성 신호를 포함하는지 여부를 검출한다.

[0035] 구체적으로, 오디오 신호가 음성 신호를 포함하는지 여부를 결정하기 위해 다음 2가지 방법이 사용될 수 있다.

[0036] 방법 1: 모든 단시간 에너지 프레임들의 총 수에 대한, 자신의 에너지가 미리 결정된 문턱값보다 더 큰 단시간 에너지 프레임의 수의 비(이하 고에너지 프레임 비로 지칭됨)가 결정되고, 결정된 고에너지 프레임 비가 미리 결정된 비보다 더 큰지 여부가 결정된다. 그러한 경우, 오디오 신호가 음성 신호를 포함한다고 결정되고, 또는 그렇지 않은 경우, 오디오 신호가 음성 신호를 포함하지 않는다고 결정된다.

[0037] 미리 결정된 문턱값의 값 및 미리 결정된 비의 값은 실제 요구에 기초하여 설정될 수 있다. 본 출원의 본 구현에서, 미리 결정된 문턱값은 2로 설정될 수 있고, 미리 결정된 비는 20%로 설정될 수 있다. 고에너지 프레임 비가 20%보다 더 큰 경우, 오디오 신호가 음성 신호를 포함한다고 결정되고, 그렇지 않은 경우에는 오디오 신호가 음성 신호를 포함하지 않는다고 결정된다.

[0038] 본 출원의 본 구현에서, 실제 생활에서는 사람들이 말할 때 외부 환경에 일부 잡음이 있고 잡음은 일반적으로 사람의 음성보다 더 낮은 에너지를 갖기 때문에, 방법 1이 오디오 신호가 음성 신호를 포함하는지 여부를 결정하는데 사용될 수 있다. 이 경우에, 오디오 신호 세그먼트가, 자신의 에너지가 미리 결정된 문턱값보다 더 큰 단시간 에너지 프레임을 포함하고, 이 단시간 에너지 프레임이 오디오 신호 세그먼트의 특정 비를 구성하는 경우, 오디오 신호가 음성 신호를 포함한다고 결정될 수 있다.

[0039] 방법 2: 최종 검출 결과를 보다 정확하게 하기 위하여, 방법 1이, 고에너지 프레임 비를 결정하고 결정된 고에너지 프레임 비가 미리 결정된 비보다 더 큰지 여부를 결정하는데 사용될 수 있다. 그렇지 않은 경우, 오디오 신호가 음성 신호를 포함하지 않는다고 결정되고, 또는 그러한 경우, 자신의 에너지가 미리 결정된 문턱값보다

더 큰, 단시간 에너지 프레임들 내의 적어도 N개의 연속 단시간 에너지 프레임이 존재할 때, 오디오 신호가 음성 신호를 포함한다고 결정되고, 또는 자신의 에너지가 미리 결정된 문턱값보다 더 큰, 단시간 에너지 프레임들 내의 적어도 N개의 연속 단시간 에너지 프레임이 존재하지 않을 때, 오디오 신호가 음성 신호를 포함하지 않는다고 결정된다. N은 양의 정수일 수 있다. 본 출원의 본 구현에서, N은 10으로 설정될 수 있다.

[0040] 구체적으로, 방법 1에 기초하여, 방법 2에서는, 오디오 신호가 음성 신호를 포함하는지 여부를 결정하기 위해 다음의 요건이 추가된다: 자신의 에너지가 미리 결정된 문턱값보다 더 큰, 단시간 에너지 프레임들 내의 적어도 N개의 연속 단시간 에너지 프레임이 존재하는지 여부가 결정된다. 그러한 경우, 잡음은 효과적으로 감소될 수 있다. 실제 생활에서, 잡음은 사람들의 음성보다 더 작은 에너지를 갖고 오디오 신호는 랜덤이며, 방법 2에서, 오디오 신호가 과도한 잡음을 포함하는 경우가 효과적으로 배제될 수 있고, 외부 환경에서의 잡음의 영향이 감소되어, 잡음 감소 기능을 달성한다.

[0041] 본 출원의 본 구현에서 제공되는 음성 신호 검출 방법은 모노(mono) 오디오 신호, 바이노럴(binaural) 오디오 신호, 멀티채널 오디오 신호 등의 검출에 적용될 수 있다는 것을 유의하여야 한다. 하나의 사운드 채널을 사용함으로써 수집된 오디오 신호는 모노 오디오 신호이고, 2개의 사운드 채널을 사용함으로써 수집된 오디오 신호는 바이노럴 오디오 신호이고, 복수의 사운드 채널을 사용함으로써 수집된 오디오 신호는 멀티채널 오디오 신호이다.

[0042] 바이노럴 오디오 신호 및 멀티채널 오디오 신호가 도 1에 도시된 방법에서 검출될 때, 각각의 채널의 획득된 오디오 신호는 단계 101 내지 단계 104에서 언급된 동작들을 수행함으로써 검출될 수 있고, 최종적으로, 각각의 채널의 오디오 신호의 검출 결과에 기초하여, 획득된 오디오 신호가 음성 신호를 포함하는지 여부가 결정된다.

[0043] 구체적으로, 단계 101에서 획득된 오디오 신호가 모노 오디오 신호인 경우, 단계 101 내지 단계 104에서 언급된 동작들은 오디오 신호에 대해 직접 수행될 수 있고, 검출 결과가 최종 검출 결과로서 사용된다.

[0044] 단계 101에서 획득된 오디오 신호가 모노 오디오 신호 대신 바이노럴 오디오 신호 또는 멀티채널 오디오 신호인 경우, 각각의 채널의 오디오 신호는 단계 101 내지 단계 104에서 언급된 동작들을 수행함으로써 처리될 수 있다. 각각의 채널의 오디오 신호가 음성 신호를 포함하지 않는다는 것이 검출되는 경우, 단계 101에서 획득된 오디오 신호가 음성 신호를 포함하지 않는다고 결정된다. 적어도 하나의 채널의 오디오 신호가 음성 신호를 포함한다는 것이 검출되는 경우, 단계 101에서 획득된 오디오 신호가 음성 신호를 포함한다고 결정된다.

[0045] 또한, 단계 102에서 언급된 미리 결정된 음성 신호의 주파수는 임의의 음성의 주파수일 수 있다. 구현은 본 출원에서 한정되지 않는다. 실제로, 실제 경우에 기초하여, 미리 결정된 음성 신호의 상이한 주파수가 단계 101에서 획득된 상이한 오디오 신호에 대하여 설정될 수 있다. 분할을 통해 최종적으로 획득되는 단시간 에너지 프레임이 다음 요건을 충족한다면, 미리 결정된 음성 신호의 주파수는 고음의 음성 주파수 또는 저음의 음성 주파수와 같은 임의의 음성 신호의 주파수일 수 있다는 것을 유의하여야 한다: 단시간 에너지 프레임에 대응하는 지속기간이 단계 101에서 획득된 오디오 신호에 대응하는 주기보다 더 작지 않다. 더 나은 검출 효과를 보장하고 가능한 많은 자원을 절약하며 처리 속도를 개선하기 위하여, 본 출원의 본 구현에서, 미리 결정된 음성 신호의 주파수는 최소 사람 음성 주파수, 즉 82 Hz로 설정될 수 있다. 주기는 주파수의 역수이기 때문에, 미리 결정된 음성 신호의 주파수가 최소 사람 음성 주파수인 경우, 미리 결정된 음성 신호의 주기는 최대 사람 음성 주기이다. 따라서, 단계 101에서 획득된 오디오 신호의 주기에 관계없이, 단시간 에너지 프레임에 대응하는 지속기간은 이전에 획득된 오디오 신호의 주기보다 더 작지 않다.

[0046] 본 출원의 본 구현에서, 여기에서 설명되는 검출 방법은 오디오 신호가 사람 음성의 특징에 기초한 음성 신호를 포함하는지 여부를 결정하는데 사용되기 때문에, 단시간 에너지 프레임에 대응하는 지속기간이 단계 101에서 획득된 오디오 신호의 주기보다 더 작아야 한다는 것을 유의하여야 한다. 잡음과 비교하여, 사람의 음성은 더 높은 에너지를 가지며 보다 안정적이고 연속적이다. 단시간 에너지 프레임에 대응하는 지속기간이 단계 101에서 획득된 오디오 신호의 주기보다 더 작은 경우, 단시간 에너지 프레임에 대응하는 파형은 완전한 주기의 파형을 포함하지 않고, 단시간 에너지 프레임의 지속기간은 비교적 짧다. 이 경우에, 고에너지 프레임 비가 미리 결정된 비보다 더 크고, 그의 에너지가 미리 결정된 문턱값보다 더 큰 단시간 에너지 프레임들 내의 적어도 N개의 연속 단시간 에너지 프레임이 존재하더라도, 오디오 신호가 사운드 신호를 포함한다는 것을 나타내기만 하며, 사운드 신호가 음성 신호라는 것을 나타내지 않는다. 따라서, 본 출원의 본 구현에서, 단계 101에서 획득된 오디오 신호의 지속기간은 최대 사람 음성 주기보다 더 커야 한다.

[0047] 또한, 본 출원의 본 구현에서 제공되는 음성 신호 검출 방법은, 음성 메시지의 송신이 사용자의 어떠한 탭 동작

없이 챗 APP을 사용함으로써 완료될 수 있는 응용 시나리오에 특히 적용 가능하다. 시나리오에 기초하여, 다음은 본 출원의 본 구현에서 제공되는 음성 신호 검출 방법을 상세하게 기재한다. 이 시나리오에서 도 2는 방법의 절차의 개략도이다. 방법은 다음 단계들을 포함한다.

[0048] 단계 201: 실시간으로 오디오 신호를 수집한다.

[0049] 사용자는 사용자가 챗 APP을 시작한 후에 어떠한 탭 동작도 없이도 챗 APP이 음성 메시지 보내기를 완료할 것으로 예상할 수 있다. 이 경우에, APP은 사용자의 음성 생략을 감소시키기 위해 실시간으로 오디오 신호를 수집하도록 외부 환경을 연속으로 녹음한다. 또한, 오디오 신호를 수집한 후에, APP은 실시간으로 오디오 신호를 국부적으로 저장할 수 있다. 사용자가 APP을 정지한 후에, APP은 녹음을 정지한다.

[0050] 단계 202: 실시간으로 수집된 오디오 신호로부터 미리 결정된 지속기간을 갖는 오디오 신호를 클리핑(clipping)한다.

[0051] APP이 실시간으로 음성 신호를 검출하는 대신 계속해서 녹음하는 경우, 음성 메시지는 실시간으로 보내지지 않는다. 따라서, APP은 단계 201에서 수집된 오디오 신호로부터 미리 결정된 지속기간을 갖는 오디오 신호를 실시간으로 클리핑할 수 있고, 미리 결정된 지속기간을 갖는 오디오 신호에 대해 후속 검출을 수행할 수 있다.

[0052] 미리 결정된 지속기간으로 현재 클리핑된 오디오 신호가 현재 오디오 신호로 지칭될 수 있고, 미리 결정된 지속기간으로 마지막 클리핑된 오디오 신호가 마지막 획득된 오디오 신호로 지칭될 수 있다.

[0053] 단계 203: 미리 결정된 음성 신호의 주파수에 기초하여 미리 결정된 지속기간 내의 오디오 신호를 복수의 단시간 에너지 프레임들로 나눈다.

[0054] 단계 204: 각각의 단시간 에너지 프레임의 에너지를 결정한다.

[0055] 단계 205: 각각의 단시간 에너지 프레임의 에너지에 기초하여, 미리 결정된 지속기간 내의 오디오 신호가 음성 신호를 포함하는지 여부를 검출한다.

[0056] 현재 오디오 신호가 음성 신호를 포함한다는 것이 검출되는 경우, 마지막 획득된 오디오 신호가 음성 신호를 포함하는지 여부가 결정된다. 마지막 획득된 오디오 신호가 음성 신호를 포함하지 않는다고 결정되는 경우, 현재 오디오 신호의 시작점은 음성 신호의 시작점으로서 결정될 수 있고, 또는 마지막 획득된 오디오 신호가 음성 신호를 포함한다고 결정되는 경우, 현재 오디오 신호의 시작점은 음성 신호의 시작점이 아니다.

[0057] 현재 오디오 신호가 음성 신호를 포함하지 않는다는 것이 검출되는 경우, 마지막 획득된 오디오 신호가 음성 신호를 포함하는지 여부가 결정된다. 마지막 획득된 오디오 신호가 음성 신호를 포함한다고 결정되는 경우, 마지막 획득된 오디오 신호의 종료점은 음성 신호의 종료점으로서 결정될 수 있고, 또는 마지막 획득된 오디오 신호가 음성 신호를 포함하지 않는다고 결정되는 경우, 현재 오디오 신호의 종료점도 마지막 획득된 오디오 신호의 종료점도 음성 신호의 종료점이 아니다.

[0058] 예를 들어, 도 3에 도시된 바와 같이, A, B, C 및 D는 미리 결정된 지속기간을 갖는 4개의 인접한 오디오 신호이다. A와 D는 음성 신호를 포함하지 않고, B와 C는 음성 신호를 포함한다. 이 경우에, B의 시작점이 음성 신호의 시작점으로서 결정될 수 있고, C의 종료점이 음성 신호의 종료점으로서 결정될 수 있다.

[0059] 종종 현재 오디오 신호가 사용자의 문장의 시작 부분이나 종료 부분이 되고, 오디오 신호는 몇몇 음성 신호를 포함한다. 이 경우에, APP은 오디오 신호가 음성 신호를 포함하지 않는다고 잘못 결정할 수 있다. 잘못된 결정으로 인한 사용자 음성의 생략을 감소시키기 위해, 현재 오디오 신호가 음성 신호를 포함한다는 것이 검출된 후에, 마지막 획득된 오디오 신호가 음성 신호를 포함하는지 여부가 결정될 수 있고, 마지막 획득된 오디오 신호가 음성 신호를 포함하지 않는다고 결정되는 경우, 마지막 획득된 오디오 신호의 시작점은 음성 신호의 시작점으로서 결정될 수 있다. 또한, 현재 오디오 신호가 음성 신호를 포함하지 않는다는 것이 검출된 후에, 마지막 획득된 오디오 신호가 음성 신호를 포함하는지 여부가 결정될 수 있고, 마지막 획득된 오디오 신호가 음성 신호를 포함한다고 결정되는 경우, 현재 오디오 신호의 종료점이 음성 신호의 종료점으로서 결정될 수 있다. 앞의 예에서, A의 시작점이 음성 신호의 시작점으로서 결정될 수 있고, D의 종료점이 음성 신호의 종료점으로서 결정될 수 있다.

[0060] 현재 오디오 신호가 음성 신호를 포함한다는 것을 검출한 후에, APP은 음성 식별 장치에 오디오 신호를 보낼 수 있고, 그리하여 음성 식별 장치는 음성 결과를 획득하도록 오디오 신호에 대해 음성 프로세싱을 수행할 수 있다. 그 다음, 음성 식별 장치는 후속 프로세싱 장치에 오디오 신호를 보내고, 최종적으로 오디오 신호는 음

성 메시지 형태로 보내진다. 보내진 음성 메시지에서의 사용자 음성이 완전한 문장임을 보장하기 위해, 음성 신호의 결정된 시작점과 결정된 종료점 사이의 모든 오디오 신호를 음성 식별 장치에 보낸 후에, APP은, 사용자가 현재 말한 이 문장이 완료된 것을 음성 식별 장치에 알리도록 오디오 정지 신호를 음성 식별 장치에 보낼 수 있으며, 그리하여 음성 식별 장치는 후속 프로세싱 장치에 모든 오디오 신호를 보낸다. 최종적으로, 오디오 신호는 음성 메시지의 형태로 보내진다.

[0061] 또한, 정확한 결정을 보장하기 위해, 현재 오디오 신호가 획득된 후에, 미리 결정된 기간을 갖는 서브신호가 마지막 획득된 오디오 신호로부터 더 클리핑될 수 있고, 현재 오디오 신호와 클리핑된 서브 신호는, 획득된 오디오 신호(이하, 연결된 오디오 신호로 지칭됨)로 쓰이도록 연결된다(concatenated). 또한, 연결된 오디오 신호에 대해 후속 음성 신호 검출이 수행된다.

[0062] 서브신호는 현재 오디오 신호 전에 연결될 수 있다. 미리 결정된 기간은, 마지막 획득된 오디오 신호의 테일(tail) 기간일 수 있고, 그 기간에 대응하는 지속기간이 임의의 지속기간일 수 있다. 최종 검출 결과가 더 정확함을 보장하기 위하여, 본 출원의 본 구현에서, 미리 결정된 기간에 대응하는 지속기간은, 미리 결정된 비 그리고 연결된 오디오 신호에 대응하는 지속기간의 곱보다 더 크지 않은 값으로 설정될 수 있다.

[0063] 연결된 오디오 신호가 음성 신호를 포함한다는 것이 검출되는 경우, 마지막 획득된 연결된 오디오 신호가 음성 신호를 포함하는지 여부가 결정될 수 있다. 마지막 획득된 연결된 오디오 신호가 음성 신호를 포함하지 않는다고 결정되는 경우, 연결된 오디오 신호의 시작점은 음성 신호의 시작점으로서 사용될 수 있다. 연결된 오디오 신호가 음성 신호를 포함하지 않는다는 것이 검출되는 경우, 마지막 획득된 연결된 오디오 신호가 음성 신호를 포함하는지 여부가 결정될 수 있다. 마지막 획득된 연결된 오디오 신호가 음성 신호를 포함한다고 결정되는 경우, 연결된 오디오 신호의 종료점은 음성 신호의 종료점으로서 사용될 수 있다.

[0064] 본 출원의 본 구현에서, 연속 녹음에 추가적으로, APP은 녹음을 주기적으로 수행할 수 있다. 구현은 본 출원의 본 구현에서 한정되지 않는다.

[0065] 본 출원의 본 구현에서 제공되는 음성 신호 검출 방법은 음성 신호 검출 장치를 사용함으로써 더 구현될 수 있다. 장치의 개략 구조도가 도 4에 도시되어 있다. 음성 신호 검출 장치는 주로, 오디오 신호를 획득하도록 구성된 획득 모듈(41); 오디오 신호를 미리 결정된 음성 신호의 주파수에 기초하여 복수의 단시간 에너지 프레임들로 나누도록 구성된 분할 모듈(42); 각각의 단시간 에너지 프레임의 에너지를 결정하도록 구성된 결정 모듈(43); 및 각각의 단시간 에너지 프레임의 에너지에 기초하여, 오디오 신호가 음성 신호를 포함하는지 여부를 검출하도록 구성된 검출 모듈(44)을 포함한다.

[0066] 구현에서, 획득 모듈(41)은, 현재 오디오 신호를 획득하고, 마지막 획득된 오디오 신호로부터 미리 결정된 기간을 갖는 서브신호를 클리핑하고, 획득된 오디오 신호로 쓰이게 현재 오디오 신호와 클리핑된 서브신호를 연결하도록 구성된다.

[0067] 구현에서, 분할 모듈(42)은, 미리 결정된 음성 신호의 주파수에 기초하여 미리 결정된 음성 신호의 주기를 결정하고, 결정된 주기에 기초하여 오디오 신호를 그의 대응하는 지속기간이 그 주기인 복수의 단시간 에너지 프레임들로 분할하도록 구성된다.

[0068] 구현에서, 검출 모듈(44)은, 모든 단시간 에너지 프레임의 총 수에 대한, 그의 에너지가 미리 결정된 문턱값보다 더 큰 단시간 에너지 프레임의 수의 비를 결정하고, 비가 미리 결정된 비보다 더 큰지 여부를 결정하고, 그러한 경우 오디오 신호가 음성 신호를 포함한다고 결정하고, 또는 그렇지 않은 경우 오디오 신호가 음성 신호를 포함하지 않는다고 결정하도록 구성된다.

[0069] 구현에서, 검출 모듈(44)은, 모든 단시간 에너지 프레임의 총 수에 대한, 그의 에너지가 미리 결정된 문턱값보다 더 큰 단시간 에너지 프레임의 수의 비를 결정하고, 비가 미리 결정된 비보다 더 큰지 여부를 결정하고, 그렇지 않은 경우 오디오 신호가 음성 신호를 포함하지 않는다고 결정하고, 또는 그러한 경우 그의 에너지가 미리 결정된 문턱값보다 더 큰 단시간 에너지 프레임들 내의 적어도 N개의 연속 단시간 에너지 프레임이 존재할 때 오디오 신호가 음성 신호를 포함한다고 결정하고, 또는 그의 에너지가 미리 결정된 문턱값보다 더 큰 단시간 에너지 프레임 내의 적어도 N개의 연속 단시간 에너지 프레임이 존재하지 않을 때 오디오 신호가 음성 신호를 포함하지 않는다고 결정하도록 구성된다.

[0070] 기존의 기술에서는, 퓨리에 변환과 같은 복잡한 계산을 통해 오디오 신호가 음성 신호를 포함하는지 여부가 결정된다. 이와 달리, 본 출원의 구현에 사용되는 음성 신호 검출 방법에서는, 퓨리에 변환과 같은 복잡한 계산이 수행될 필요가 없다. 획득된 오디오 신호는, 미리 결정된 음성 신호의 주파수에 기초하여 복수의 단시간 에

너지 프레임들로 나누어지고, 각각의 단시간 에너지 프레임의 에너지가 더 결정되며, 각각의 단시간 에너지 프레임의 에너지에 기초하여, 획득된 오디오 신호가 음성 신호를 포함하는지 여부가 검출될 수 있다. 따라서 본 출원의 구현에서 제공되는 음성 신호 검출 방법에서는, 기존 기술에서의 음성 신호 검출 방법에서 처리 속도가 비교적 낮고 자원 소비가 비교적 높다는 문제점이 해소될 수 있다.

[0071] 본 개시는 본 개시의 구현에 기초한 방법, 디바이스(시스템), 및 컴퓨터 프로그램 제품의 흐름도 및/또는 블록도를 참조하여 기재되어 있다. 컴퓨터 프로그램 명령어는 흐름도 및/또는 블록도에서의 각각의 프로세스 및/또는 각각의 블록 및 흐름도 및/또는 블록도에서의 프로세스 및/또는 블록의 조합을 구현하도록 사용될 수 있다는 것을 유의하여야 한다. 이들 컴퓨터 프로그램 명령어는, 기계를 발생시키도록 범용 컴퓨터, 전용 컴퓨터, 내장 프로세서, 또는 또다른 프로그램가능 데이터 프로세싱 디바이스의 프로세서에 대하여 제공될 수 있으며, 그리하여 컴퓨터 또는 또다른 프로그램가능 데이터 프로세싱 디바이스의 프로세서에 의해 실행된 명령어는 흐름도에서의 하나 이상의 프로세스 및/또는 블록도에서의 하나 이상의 블록에서의 지정된 기능을 구현하기 위한 디바이스를 발생시킨다.

[0072] 이들 컴퓨터 프로그램 명령어는, 컴퓨터 판독가능한 메모리에 저장된 명령어가 명령 디바이스를 포함하는 인공물을 발생시키는 방식으로 작업하게끔 컴퓨터 또는 또다른 프로그램가능 데이터 프로세싱 디바이스에 명령할 수 있는 컴퓨터 판독가능한 메모리에 저장될 수 있다. 명령 디바이스는 흐름도에서의 하나 이상의 프로세스 및/또는 블록도에서의 하나 이상의 블록에서의 지정된 기능을 구현한다.

[0073] 이들 컴퓨터 프로그램 명령어는 컴퓨터 또는 또다른 프로그램가능 데이터 프로세싱 디바이스로 로딩될 수 있으며, 그리하여 일련의 동작들 및 단계들이 컴퓨터 또는 또다른 프로그램가능 디바이스 상에서 수행됨으로써, 컴퓨터 구현 프로세싱을 발생시킨다. 따라서, 컴퓨터 또는 또다른 프로그램가능 디바이스 상에서 실행되는 명령어는 흐름도에서의 하나 이상의 프로세스 및/또는 블록도에서의 하나 이상의 블록에서의 지정된 기능을 구현하기 위한 단계를 제공한다.

[0074] 통상의 구성에서, 계산 디바이스는 하나 이상의 중앙 처리 유닛(CPU; central processing unit), 하나 이상의 입력/출력 인터페이스, 하나 이상의 네트워크 인터페이스, 및 하나 이상의 메모리를 포함한다.

[0075] 메모리는 비영구적 메모리, 랜덤 액세스 메모리(RAM; random access memory), 비휘발성 메모리, 및/또는 컴퓨터 판독가능한 매체에 있는 또다른 형태, 예를 들어 판독 전용 메모리(ROM; read-only memory) 또는 플래시 메모리(flash RAM)를 포함할 수 있다. 메모리는 컴퓨터 판독가능한 매체의 예이다.

[0076] 컴퓨터 판독가능한 매체는, 임의의 방법 또는 기술을 사용함으로써 정보를 저장할 수 있는 영구적, 비영구적, 이동식 및 비이동식 매체를 포함한다. 정보는 컴퓨터 판독가능한 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터일 수 있다. 컴퓨터 저장 매체의 예는, PRAM(phase-change random access memory), SRAM(static random access memory), DRAM(dynamic random access memory), 또다른 유형의 RAM, ROM, EEPROM(electrically erasable programmable read-only memory), 플래시 메모리 또는 또다른 메모리 기술, CD-ROM(compact disk read-only memory), DVD(digital versatile disc) 또는 또다른 광 스토리지, 카세트 자기 테이프, 자기 테이프/자기 디스크 저장장치, 또다른 자기 저장 디바이스, 또는 임의의 기타 비전송 매체를 포함하지만, 이에 한정되지 않는 것은 아니다. 컴퓨터 저장 매체는 계산 디바이스가 액세스할 수 있는 정보를 저장하도록 구성될 수 있다. 본 명세서에서의 정의에 기초하여, 컴퓨터 판독가능한 매체는 일시적 컴퓨터 판독가능한 매체(일시적 매체), 예컨대 변조된 데이터 신호 및 캐리어를 포함하지 않는다.

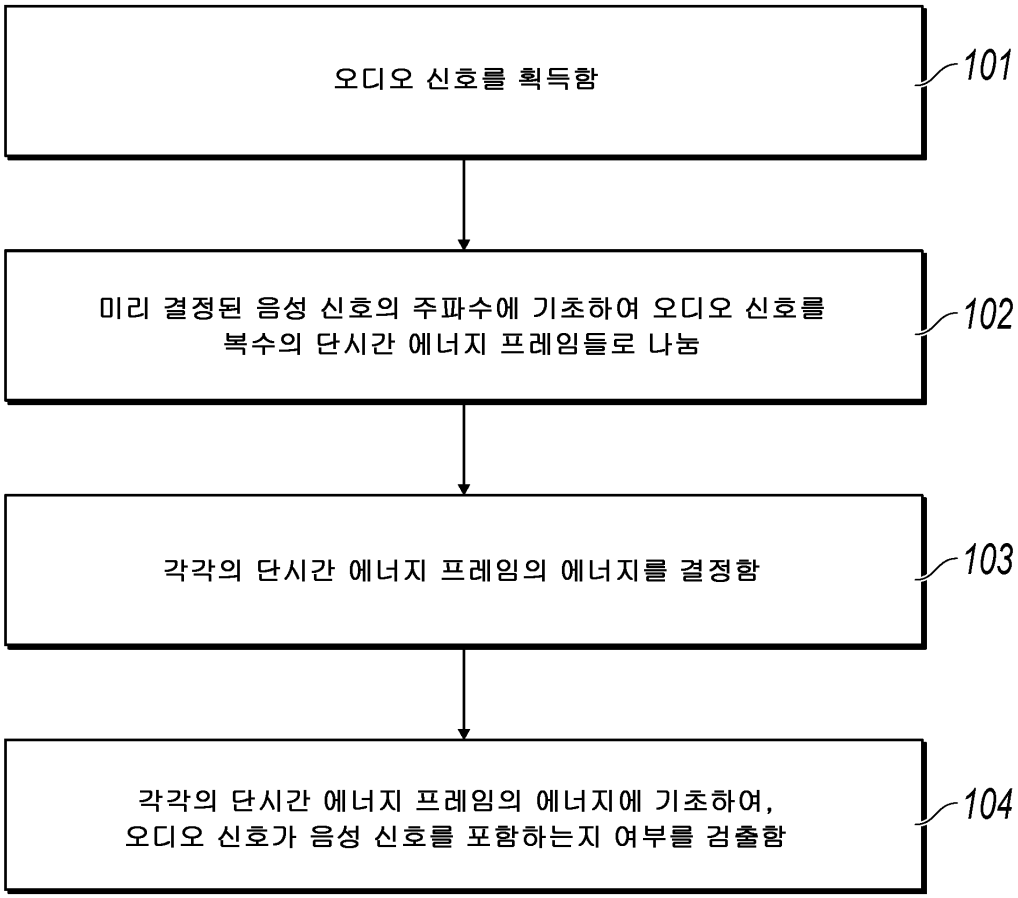
[0077] 용어 "포함한다", "함유한다" 또는 이의 기타 변형어는 비배타적인 포함(non-exclusive inclusion)을 커버하도록 의도되며, 그리하여 일련의 요소들을 포함하는 프로세스, 방법, 물품, 또는 디바이스가 이 요소들을 포함할 뿐만 아니라, 명시적으로 열거되지 않은 다른 요소도 포함하며, 또는 이러한 프로세스, 방법, 물품, 또는 디바이스에 고유한 요소를 더 포함한다는 것을 더 유의하여야 한다. "...를 포함한다"에 의해 기재된 요소는, 더 이상의 제약 없이, 그 요소를 포함하는 프로세스, 방법, 물품, 또는 디바이스에서의 추가의 동일 요소의 존재를 배제하지 않는다.

[0078] 당해 기술분야에서의 숙련자는 본 출원의 구현이 방법, 시스템, 또는 컴퓨터 프로그램 제품으로서 제공될 수 있다는 것을 이해하여야 한다. 따라서, 본 출원은 하드웨어 전용 구현, 소프트웨어 전용 구현, 또는 소프트웨어와 하드웨어 조합으로의 구현의 형태를 사용할 수 있다. 또한, 본 출원은 컴퓨터 사용가능한 프로그램 코드를 포함하는 하나 이상의 컴퓨터 사용가능한 저장 매체(디스크 메모리, CD-ROM, 및 광학 메모리 등을 포함하지만, 이에 한정되지 않음) 상에 구현되는 컴퓨터 프로그램 제품의 형태를 사용할 수 있다.

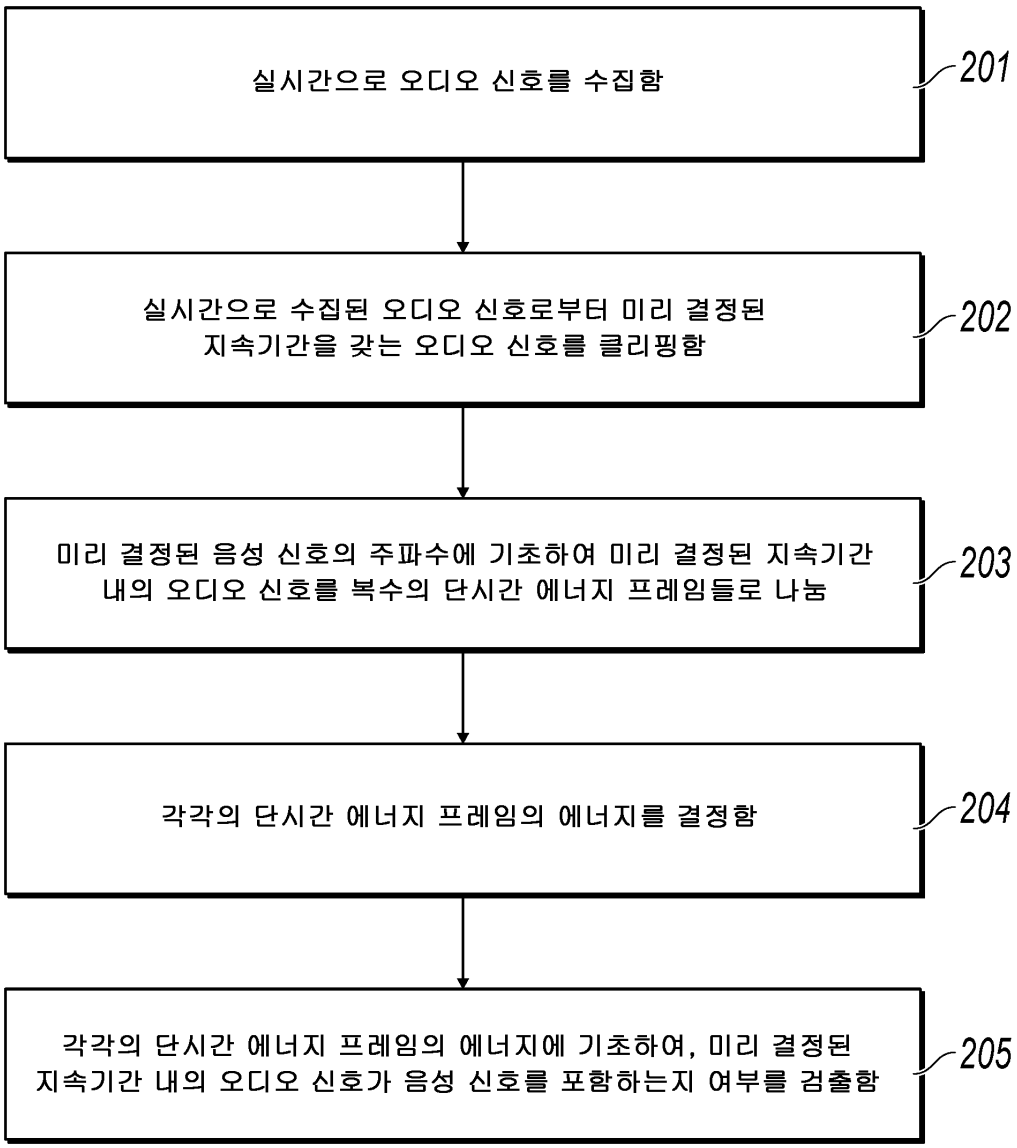
[0079] 앞의 구현은 본 출원의 구현일 뿐이고, 본 출원을 한정하도록 의도되지 않는다. 당해 기술분야에서의 숙련자는 본 출원에 대해 다양한 수정 및 변경을 행할 수 있다. 본 출원의 사상 및 원리에서 벗어나지 않고서 행해지는 임의의 수정, 등가의 교체, 또는 개선은 본 출원의 청구항의 범위 내에 속할 것이다.

도면

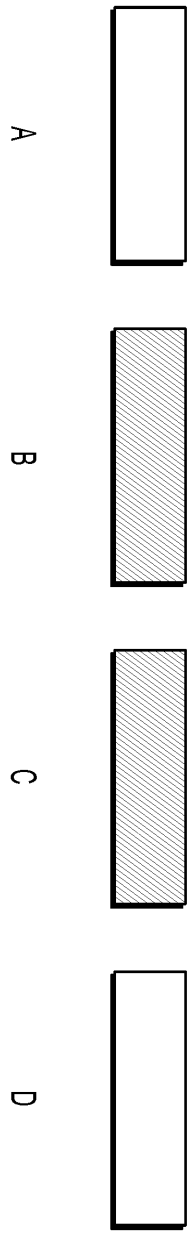
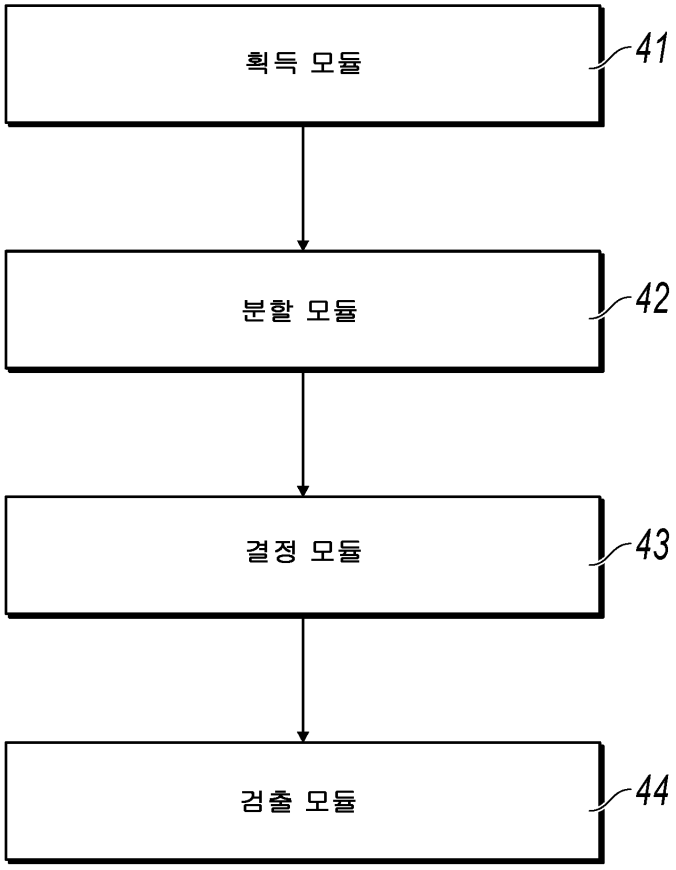
도면1



도면2


실시간으로 오디오 신호를 수집함
201
실시간으로 수집된 오디오 신호로부터 미리 결정된 지속기간을 갖는 오디오 신호를 클리핑함
202
미리 결정된 음성 신호의 주파수에 기초하여 미리 결정된 지속기간 내의 오디오 신호를 복수의 단시간 에너지 프레임들로 나눔
203
각각의 단시간 에너지 프레임의 에너지를 결정함
204
각각의 단시간 에너지 프레임의 에너지에 기초하여, 미리 결정된 지속기간 내의 오디오 신호가 음성 신호를 포함하는지 여부를 검출함
205

도면3



도면4