



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0076816
(43) 공개일자 2014년06월23일

(51) 국제특허분류(Int. Cl.)

G10L 15/02 (2006.01)

(21) 출원번호 10-2012-0145284

(22) 출원일자 2012년12월13일

심사청구일자 없음

(71) 출원인

한국전자통신연구원

대전광역시 유성구 가정로 218 (가정동)

(72) 발명자

박기영

대전 유성구 지족로 362, 302동 703호 (지족동, 반석마을3단지아파트)

이윤근

대전 서구 청사서로 11, 103동 1406호 (월평동, 무지개아파트)

(74) 대리인

특허법인태평양

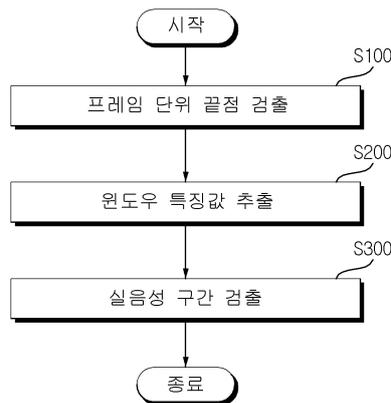
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 음성 신호의 검출 방법 및 장치

(57) 요약

본 발명은 음성 인식 시스템 등에 사용되는 음성 끝점 검출 및 특징 벡터 추출기의 성능 개선에 관한 것이다. 본 발명에 따른 음성 신호의 검출 방법은 입력 신호에 대하여 프레임 단위의 끝점 검출을 수행하여 음성 구간을 검출하는 단계; 검출된 상기 음성 구간 중 복수의 윈도우들에 대응되는 적어도 일부 구간의 신호의 특징값을 추출하는 단계; 및 상기 추출된 특징값을 미리 결정된 임계값과 비교하여 상기 음성 구간 중 실 음성 구간을 검출하는 단계를 포함한다. 본 발명에서 제시한 방법을 사용할 경우 음성 인식 시스템에서 사용하는 특징 벡터의 정규화 성능을 높일 수 있으며, 이는 잡음 환경에서의 음성 인식 성능을 높이는 역할을 한다.

대표도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호 10035252

부처명 지경부

연구사업명 산업원천기술개발사업(정보통신)

연구과제명 모바일 플랫폼 기반 대화모델 적용 자연어 음성 인터페이스 기술 개발

기여율 1/1

주관기관 ETRI

연구기간 2010.03.01 ~ 2015.02.28

특허청구의 범위

청구항 1

입력 신호에 대하여 프레임 단위의 끝점 검출을 수행하여 음성 구간을 검출하는 단계;
검출된 상기 음성 구간 중 복수의 윈도우들에 대응되는 적어도 일부 구간의 신호의 특징값을 추출하는 단계; 및
상기 추출된 특징값을 미리 결정된 임계값과 비교하여 상기 음성 구간 중 실 음성 구간을 검출하는 단계를 포함하는 음성 신호의 검출 방법.

청구항 2

제 1 항에 있어서,
상기 음성 신호의 검출 방법은 상기 실 음성 구간의 특징벡터를 정규화하는 단계를 더 포함하는 것을 특징으로 하는 음성 신호의 검출 방법.

청구항 3

제 2 항에 있어서,
상기 실 음성 구간을 검출하는 단계는 상기 추출된 특징값의 극대값을 추출하고, 상기 극대값을 상기 임계값과 비교하여 상기 실 음성 구간을 검출하는 것을 특징으로 하는 음성 신호의 검출 방법.

청구항 4

제 2 항에 있어서,
상기 실 음성 구간을 검출하는 단계는 상기 음성 구간 중 잡음 신호의 위치 및 음성 신호의 위치를 검출하여 잡음 구간을 검출하는 것을 특징으로 하는 음성 신호의 검출 방법.

청구항 5

제 4 항에 있어서,
상기 특징벡터를 정규화하는 단계는 상기 음성 구간 중 상기 잡음 구간을 제외한 구간에 대한 상기 특징벡터를 정규화하는 것을 특징으로 하는 음성 신호의 검출 방법.

청구항 6

제 1 항에 있어서,
상기 복수의 윈도우들은 상기 윈도우에 대응되는 구간의 적어도 일부가 다른 윈도우에 대응되는 구간과 중첩되는 것을 특징으로 하는 음성 신호의 검출 방법.

청구항 7

제 4 항에 있어서,
상기 정규화하는 단계는 상기 잡음 신호의 위치 또는 상기 음성 신호의 위치가 검출되지 않은 경우 검출된 상기 음성 구간에 대한 상기 특징벡터를 정규화하는 것을 특징으로 하는 음성 신호의 검출 방법.

청구항 8

제 1 항에 있어서,
상기 특징값을 추출하는 단계는 상기 윈도우에 대응되는 구간의 통계적 특성을 추출하는 것을 특징으로 하는 음성 신호의 검출 방법.

청구항 9

입력 신호에 대하여 프레임 단위의 끝점 검출을 수행하여 음성 구간을 검출하는 끝점 검출부;

검출된 상기 음성 구간 중 복수의 윈도우들에 대응되는 적어도 일부 구간의 신호의 특징값을 추출하는 특징값 추출부; 및

상기 추출된 특징값을 미리 결정된 임계값과 비교하여 상기 음성 구간 중 실 음성 구간을 검출하는 실 음성 구간 검출부를 포함하는 음성 신호의 검출 장치.

청구항 10

제 9 항에 있어서,

상기 음성 신호의 검출 장치는 상기 실 음성 구간의 특징벡터를 정규화하는 정규화부를 더 포함하는 것을 특징으로 하는 음성 신호의 검출 장치.

청구항 11

제 10 항에 있어서,

상기 실 음성 구간 검출부는 상기 추출된 특징값의 극대값을 추출하고, 상기 극대값을 상기 임계값과 비교하여 상기 실 음성 구간을 검출하는 것을 특징으로 하는 음성 신호의 검출 장치.

청구항 12

제 10 항에 있어서,

상기 실 음성 구간 검출부는 상기 음성 구간 중 잡음 신호의 위치 및 음성 신호의 위치를 검출하여 잡음 구간을 검출하는 것을 특징으로 하는 음성 신호의 검출 장치.

청구항 13

제 12 항에 있어서,

상기 정규화부는 상기 음성 구간 중 상기 잡음 구간을 제외한 구간에 대한 상기 특징벡터를 정규화하는 것을 특징으로 하는 음성 신호의 검출 장치.

청구항 14

제 9 항에 있어서,

상기 복수의 윈도우들은 상기 윈도우에 대응되는 구간의 적어도 일부가 다른 윈도우에 대응되는 구간과 중첩되는 것을 특징으로 하는 음성 신호의 검출 장치.

청구항 15

제 12 항에 있어서,

상기 정규화부는 상기 잡음 신호의 위치 또는 상기 음성 신호의 위치가 검출되지 않은 경우 검출된 상기 음성 구간에 대한 상기 특징벡터를 정규화하는 것을 특징으로 하는 음성 신호의 검출 장치.

청구항 16

제 9 항에 있어서,

상기 특징값을 추출부는 상기 윈도우에 대응되는 구간의 통계적 특성을 추출하는 것을 특징으로 하는 음성 신호의 검출 장치.

명세서

기술분야

본 발명은 음성 인식 시스템 등에 사용되는 음성 끝점 검출 및 특징 벡터 추출기의 성능 개선에 관한 것이다.

[0001]

배경 기술

- [0002] 현재의 일반적인 음성 인식 시스템의 경우 마이크로부터 취득된 신호는 일단 끝점 검출기를 거쳐 음성 구간이라고 판정된 신호에 한해서 특징 벡터로 변환되고 변환된 특징 벡터는 디코더에 인가되어 인식 결과를 도출하게 된다. 따라서 음성 검출기 및 특징 벡터 추출기의 오동작은 음성 인식 시스템 전반에 큰 영향을 미치게 된다.
- [0003] 일반적인 음성 끝점 검출기의 경우 입력 신호의 에너지 및 통계적 모델에 따라 현재 프레임이 음성 구간인지 비음성 구간인지를 추정하고 이를 바탕으로 끝점 여부를 판정하고 된다. 이 때 정해진 환경에서는 대체로 잘 동작하나 미리 가정한 환경을 벗어나는 경우에는 오동작 확률이 높아지게 된다. 또한 실제 음성 신호의 바로 앞부분이나 바로 뒷부분에 인접한 잡음은 일반적인 끝점 검출기로는 음성 신호와의 구분이 어렵게 된다. 또한 일반적인 음성 인식기의 경우 음성 신호의 앞과 뒤에 일정 정도의 묵음 구간이 있어야 오히려 인식 결과가 바르게 나오는 경우가 많다. 음성 구간 앞뒤의 짧은 잡음은 인식기의 입장에서는 크게 문제가 되지 않으나 오히려 특징 벡터의 추출과정에서 큰 영향을 줄 수 있다. 특징 벡터의 추출과정에서는 신호의 일부분 또는 전체 구간의 통계적 특성을 이용하여 신호를 정규화하는데, 이렇게 삽입된 잡음 구간이 신호의 정규화 과정에 영향을 주기 때문이다.
- [0004] 이러한 문제점은 차량이나 사무실과 같이 일정한 환경에서 사용하는 음성 인식 시스템의 경우에는 큰 문제가 아니나, 스마트폰 등을 이용한 모바일 환경에서 음성 인식 시스템의 경우에는 큰 영향을 줄 수 있다.

발명의 내용

해결하려는 과제

- [0005] 본 발명은 상기 종래 기술의 문제를 해결하기 위하여 안출된 것으로서, 일반적인 끝점 검출기를 거친 이후의 신호에 대하여 보다 정교한 끝점 검출을 다시 한번 수행하여 특징 벡터의 정규화 과정에서 성능 향상에 도움을 줄 수 있는 방법을 제안하는 것을 목적으로 한다.
- [0006] 특징 벡터 추출기의 정규화 과정에서는 묵음 구간과 음성 구간의 정보가 필요하며 중간의 잡음 구간은 정규화 과정에서 성능 저하를 가져온다. 끝점 검출기는 음성의 시작점과 끝점을 추출하나, 일반적으로 음성의 앞부분과 뒷부분에 근접하는 잡음은 구분하기 어려운 단점이 있다.
- [0007] 본 발명에서는 일반적인 음성 끝점 검출기의 출력에 대하여 보다 정교한 끝점 검출기를 재적용하여 음성의 앞부분과 뒷부분에 인접한 잡음 구간을 별도로 추출해내고, 이 구간 정보를 이용하여 특징 추출부의 정규화 성능을 높이고자 한다.

과제의 해결 수단

- [0008] 상기 기술적 과제를 해결하기 위한 음성 신호의 검출 방법은 입력 신호에 대하여 프레임 단위의 끝점 검출을 수행하여 음성 구간을 검출하는 단계; 검출된 상기 음성 구간 중 복수의 윈도우들에 대응되는 적어도 일부 구간의 신호의 특징값을 추출하는 단계; 및 상기 추출된 특징값을 미리 결정된 임계값과 비교하여 상기 음성 구간 중 실 음성 구간을 검출하는 단계를 포함한다.
- [0009] 상기 음성 신호의 검출 방법은 상기 실 음성 구간의 특징벡터를 정규화하는 단계를 더 포함하는 것이 바람직하다.
- [0010] 상기 실 음성 구간을 검출하는 단계는 상기 추출된 특징값의 극대값을 추출하고, 상기 극대값을 상기 임계값과 비교하여 상기 실 음성 구간을 검출하는 것이 바람직하다.
- [0011] 상기 실 음성 구간을 검출하는 단계는 상기 음성 구간 중 잡음 신호의 위치 및 음성 신호의 위치를 검출하여 잡음 구간을 검출하는 것이 바람직하다.
- [0012] 상기 특징벡터를 정규화하는 단계는 상기 음성 구간 중 상기 잡음 구간을 제외한 구간에 대한 상기 특징벡터를 정규화하는 것이 바람직하다.
- [0013] 상기 복수의 윈도우들은 상기 윈도우에 대응되는 구간의 적어도 일부가 다른 윈도우에 대응되는 구간과 중첩되는 것이 바람직하다.
- [0014] 상기 정규화하는 단계는 상기 잡음 신호의 위치 또는 상기 음성 신호의 위치가 검출되지 않은 경우 검출된 상기

음성 구간에 대한 상기 특징벡터를 정규화하는 것이 바람직하다.

[0015] 상기 특징값을 추출하는 단계는 상기 윈도우에 대응되는 구간의 통계적 특성을 추출하는 것이 바람직하다.

발명의 효과

[0016] 본 발명은 종래의 끝점 추출 방식에서 걸러낼 수 없는 음성 직전 및 직후의 잡음 신호가 특징 추출부의 정규화 과정에서 전체 특징 벡터에 영향을 주는 것을 막기 위하여 2차 끝점 검출기를 통하여 음성 직전 및 직후의 잡음 신호 구간을 찾아내고, 이를 정규화 구간에서 제외하는 방법에 관한 것이다. 본 발명에서 제시한 방법을 사용할 경우 음성 인식 시스템에서 사용하는 특징 벡터의 정규화 성능을 높일 수 있으며, 이는 잡음 환경에서의 음성 인식 성능을 높이는 역할을 한다.

도면의 간단한 설명

- [0017] 도 1은 본 발명의 일실시예에 따른 음성 신호의 검출 방법을 나타내는 흐름도이다.
- 도 2는 본 발명의 일실시예에 따라 검출된 음성 구간의 구성을 나타내는 도이다.
- 도 3은 도 2의 음성 구간으로부터 실 음성 구간을 검출하는 과정을 나타내는 도이다.
- 도 4는 본 발명의 일실시예에 따라 음성 신호의 검출 방법을 나타내는 흐름도이다.
- 도 5는 본 발명의 일실시예에 따른 음성 신호의 검출 방법을 나타내는 세부 흐름도이다.
- 도 6은 본 발명의 일실시예에 따른 음성 신호의 검출 방법이 적용된 시스템을 나타내는 도이다.
- 도 7은 본 발명의 일실시예에 따른 음성 신호의 검출 장치를 나타내는 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0018] 이하의 내용은 단지 발명의 원리를 예시한다. 그러므로 당업자는 비록 본 명세서에 명확히 설명되거나 도시되지 않았지만 발명의 원리를 구현하고 발명의 개념과 범위에 포함된 다양한 장치를 발명할 수 있는 것이다. 또한, 본 명세서에 열거된 모든 조건부 용어 및 실시예들은 원칙적으로, 발명의 개념이 이해되도록 하기 위한 목적으로만 명백히 의도되고, 이와같이 특별히 열거된 실시예들 및 상태들에 제한적이지 않는 것으로 이해되어야 한다.

[0019] 상술한 목적, 특징 및 장점은 첨부된 도면과 관련한 다음의 상세한 설명을 통하여 보다 분명해 질 것이며, 그에 따라 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 발명의 기술적 사상을 용이하게 실시할 수 있을 것이다. 또한, 발명을 설명함에 있어서 발명과 관련된 공지 기술에 대한 구체적인 설명이 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에 그 상세한 설명을 생략하기로 한다. 이하, 첨부된 도면을 참조하여 발명에 따른 바람직한 일실시예를 상세히 설명하기로 한다.

[0020] 도 1은 본 발명의 일실시예에 따른 음성 신호의 검출 방법을 나타내는 흐름도이다. 도 1을 참조하면, 본 실시예에 따른 음성 신호의 검출 방법은 프레임 단위 끝점 검출 단계(S100), 윈도우 특징값 추출 단계(S200), 실 음성 구간 검출 단계(S300)를 포함한다.

[0021] 프레임 단위 끝점 검출 단계(S100)는 입력 신호의 음성 구간을 검출하기 위하여 프레임 단위의 끝점 검출을 수행한다.

[0022] 일반적인 끝점 추출기는 주로 음성 인식기의 앞단에 적용되어 입력 신호를 분석하여 음성의 시작점 및 끝점을 판정한다. 이를 위하여 여러 가지 알고리즘을 적용하고 있으나, 일반적으로 개별 프레임별 음성/비음성 판정 및 이를 바탕으로한 결정 로직 알고리즘을 이용한다.

[0023] 즉 입력 신호가 들어오면 현재 프레임이 음성 신호에 해당하는지 그렇지 않은지를 판단한다. 이를 위해서는 음성 및 비음성 신호의 확률 모델을 이용하여 음성일 확률 또는 음성이 아닐 확률을 계산하거나, 신호의 에너지를 문턱값과 비교하여 음성 신호임을 판정하거나, 그 밖의 복잡한 판정 네트워크를 이용하여 음성일 확률을 계산하거나, 음성 구간 또는 비음성 구간이라고 확정하여 결정을 하게 된다.

[0024] 판정 로직에서는 시간에 따른 프레임별 결과를 분석하여 예를 들어 일정 개수 이상의 음성 프레임이 연속하여 입력되면 음성의 시작점이 검출되었다고 판정하거나, 반대로 일정 개수 이상의 비음성 프레임이 연속하여 입력되면 음성의 끝점이 검출되었다고 판정하게 된다. 최종적인 음성 끝점 검출기는 연속된 음성 신호에 대하여 현

재 프레임이 음성 구간의 시작점 또는 끝점이라는 판단을 출력으로 내게 된다.

- [0025] 따라서 본 실시예에서 프레임 단위 끝점 검출 단계(S100)는 1차적으로 입력 신호 중 음성 구간을 검출하기 위한 끝점을 검출하는 것으로서, 이때의 끝점 검출은 확률 모델을 이용하여 음성일 확률 또는 음성이 아닐 확률을 계산하는 등의 일반적인 끝점 검출 방법을 이용한다.
- [0026] 또한 본 실시예에서 프레임이란 입력 신호를 일반적인 10msec를 길이의 신호를 세분화하는 단위로서, 이는 인식하고자 하는 음성의 특성, 예를 들어 음성의 발화 상황이나 화자, 또는 음성의 언어에 따라 변경될 수 있다.
- [0027] 다만 상술한 방식의 끝점 추출 방법은 실제 음성 구간에 인접하여 잡음이 있는 경우 이를 구분하는 것이 어렵다. 이렇게 삽입된 잡음 구간은 인식에는 큰 영향을 주지 않을 수 있으나, 특징벡터의 정규화 과정에서는 묵음 구간 및 음성 구간의 신호 특성을 왜곡시켜서 정규화 성능을 떨어뜨리게 된다.
- [0028] 따라서, 본 발명에서는 일반적인 끝점 검출기를 거친 이후의 신호에 대하여 보다 정교한 끝점 검출을 다시 한번 수행하여 특징 벡터의 정규화 과정에서 성능 향상에 도움을 줄 수 있는 방법을 제안한다.
- [0029] 이하 보다 정교한 끝점 검출을 수행하기 위한 윈도우 특징값 추출 단계(S200)에 대하여 설명한다.
- [0030] 윈도우 특징값 추출 단계(S200)는 검출된 음성 구간 중 일부 구간의 특징 값을 추출하기 위한 복수의 윈도우들의 특징값을 추출한다. 즉 본 실시예에서 윈도우 특징값 추출 단계(S200)는 1차적인 프레임 단위 끝점 검출 단계(S100)를 통해 검출된 음성 구간에 대하여 2차적인 실 음성 구간 검출을 위한 단계이다.
- [0031] 본 실시예에서 윈도우란 상술한 프레임과 같이 입력 신호의 특성을 판단하기 위하여 입력 신호를 특정 길이로 세분화하는 기준이 된다. 다만 본 실시예에서 윈도우의 길이는 상술한 끝점 검출 단계에서 프레임 단위의 끝점 검출 방식을 쓰는 경우에는 프레임 길이와 동일한 길이인 경우 새로운 정보를 얻을 수 없으므로, 일반적인 프레임 길이인 10msec보다 긴 구간의 길이인 것이 바람직하다.
- [0032] 또한 프레임의 길이가 인식하고자 하는 음성의 특성, 예를 들어 음성의 발화 상황이나 화자, 또는 음성의 언어에 따라 변경될 수 있는 바 본 실시예에서의 윈도우의 길이 역시 음성의 성질에 따라 변경될 수 있다.
- [0033] 또한 본 실시예에서 음성 구간 중 일부 구간은 상술한 윈도우에 대응되는 구간으로, 음성 구간의 통계적 특성을 구하기 위한 기준이 된다. 따라서 본 실시예에서 윈도우 특징값 추출 단계(S200)는 결정된 윈도우 길이에 따른 구간에 대한 전체 구간의 통계적 특성을 이용하여 신호의 변화 정도를 측정한다.
- [0034] 즉, 도 2와 같이 프레임 단위 끝점 검출 단계(S100)를 통해 음성 구간(20)이 검출되었다고 할 때, 윈도우 특징값 추출 단계(S200)는 도 3과 같이 복수의 윈도우에 대응되는 구간의 적어도 일부가 다른 윈도우에 대응되는 구간과 중첩되고 소정의 길이를 갖는 윈도우(24)를 음성 구간에 적용하여 해당 윈도우에 대응되는 음성 구간의 통계적 특성을 구한다.
- [0035] 이 때 각 윈도우(24) 간의 중첩되는 구간이 늘어날수록 보다 정밀한 실 음성 구간의 검출이 가능하며, 줄어들수록 검출의 속도는 빨라 질 수 있는 바, 인식하고자 하는 음성의 성질 또는 그 시스템 환경에 따라 각 윈도우들의 간격을 조절하여 중첩되는 구간의 길이를 결정하는 것이 바람직하다.
- [0036] 이하 윈도우에 대응되는 음성 구간의 통계적 특성에 대하여 설명하면, 본 실시예에 따른 통계적 특징값은 수학적 식 1로 계산될 수 있다.

수학적 식 1

[0037]
$$F_k = f(s_{k_i}, s_{k_i + 1}, s_{k_i + 2}, \dots, s_{k_i + N - 1})$$

- [0038] 여기서 Fk는 k번째 윈도우의 통계적 특징값, s는 신호값, ki는 k번째 윈도우의 시작점, N은 윈도우의 길이를 의미한다.
- [0039] f()는 입력 신호로부터 통계적 특성을 추출하는 함수로 신호의 평균값, 분산값, 엔트로피, Bayes Information Criterion등 다양한 방법이 사용될 수 있다.
- [0040] 이하 실 음성 구간 검출 단계(S300)에 대하여 설명하면, 본 실시예에서 실 음성 구간 검출 단계(S300)는 추출된

특징값을 미리 결정된 임계값과 비교하여 음성 구간 중 실 음성 구간을 검출한다.

- [0041] 본 실시예에서 실 음성 구간이란 프레임 단위 끝점 검출 단계(S100)에서 검출된 음성 구간에 포함될 수 있는 잡음 구간을 제외한 실제의 음성 신호로 구성된 구간으로서, 도 2를 참조하면 음성 구간(20) 중 실 음성 구간의 시작 또는 끝에 인접하여 포함되는 잡음 구간(23)을 제외한 구간이 실 음성 구간(22)이다.
- [0042] 즉 실제 음성 신호의 바로 앞부분이나 바로 뒷부분에 인접한 잡음은 일반적인 끝점 검출기로는 음성 신호와의 구분이 어렵게 된다. 또한 일반적인 음성 인식기의 경우 음성 신호의 앞과 뒤에 일정 정도의 묵음 구간이 있어야 오히려 인식 결과가 바르게 나오는 경우가 많다.
- [0043] 즉 실 음성 구간 앞뒤의 짧은 잡음은 인식기의 입장에서는 크게 문제가 되지 않으나 오히려 특징 벡터의 추출과정에서 큰 영향을 줄 수 있는 바, 본 실시예에서 실 음성 구간 검출 단계(S300)는 이러한 인접 잡음 구간의 영향을 제한하기 위한 동작을 수행한다.
- [0044] 이하 도 5를 참조하여 보다 상세히 설명한다. 도 5를 참조하면, 본 실시예에 따른 실 음성 구간 검출 단계(S300)는 윈도우 특징값의 극대값 추출 단계(S310), 극대값과 임계값 비교 단계(S320)를 포함한다.
- [0045] 윈도우 특징값의 극대값 추출 단계(S310)는 추출된 특징값의 극대값을 추출한다.
- [0046] 수학식 1에 의해서 계산된 특징 값은 윈도우를 1개 또는 1개 이상의 샘플만큼 이동시켜가면서 연속된 값을 계산할 수 있다. 이 값이 극대값이 되는 점이 신호의 특성이 크게 바뀌는 시점으로 본 실시예에서 윈도우 특징값의 극대값 추출 단계(S310)는 극대값을 통해 신호의 특성이 크게 바뀌는 시점을 찾는다.
- [0047] 다음 극대값과 임계값 비교 단계(S320)는 추출된 극대값을 임계값과 비교하여 실 음성 구간을 검출한다. 이 값을 임계값과 비교하여 음성 신호 앞뒤에 인접한 잡음 신호의 위치 및 음성 신호의 위치를 찾을 수 있다.
- [0048] 따라서 본 실시예에서 실 음성 구간 검출 단계(S300)는 극대값과 임계값 비교 단계(S320)에서 잡음 신호 및 음성 신호의 위치를 통해 두 신호의 시작점을 찾고 잡음 신호의 시작점부터 음성 신호의 시작점으로 구성되는 잡음 구간을 검출한다.
- [0049] 이하 본 실시예에 따른 음성 신호의 검출 방법은 도 4와 같이 실 음성 구간 정규화 단계(S400)를 더 포함할 수 있다.
- [0050] 실 음성 구간 정규화 단계(S400)는 실 음성 구간의 특징벡터를 정규화한다. 특징 벡터의 정규화 방법으로는 대표적인 것은 Cepstral Mean Subtraction (CMS) 방법이다. CMS 방법은 마이크 및 화자의 발성 중 인식에 불필요한 특성 등을 정규화해주는 방법으로 캡스트럼 영역의 특징벡터를 구한 후 시간 축으로 전체 특징벡터의 평균을 구하여 이 평균값을 전체 벡터에서 빼는 방법이다.
- [0051] 상술한 바와 같이 실제 음성 구간에 인접하여 잡음이 있는 경우 이를 구분하는 것이 어렵고, 이렇게 삽입된 잡음 구간은 인식에는 큰 영향을 주지 않을 수 있으나, 특징벡터의 정규화 과정에서는 묵음 구간 및 음성 구간의 신호 특성을 왜곡시켜서 정규화 성능을 떨어뜨리게 된다.
- [0052] 따라서, 본 실시예에서 정규화 단계는 실 음성 구간 검출 단계(S300)에서 음성 신호 및 잡음 신호의 시작점이 찾아진 경우 찾아진 잡음 신호의 시작점을 t_n , 음성 신호의 시작점을 t_s 라고 하자.
- [0053] 이때, 실 음성 구간 정규화 단계(S400)에서 특징 벡터의 정규화는 프레임 단위 끝점 검출 단계(S100)에서 검출된 끝점에 따른 음성구간 L 중, t_n 에서 t_s 의 구간을 제외하고 이루어진다. 예를 들어 신호 전체의 평균을 구할 경우 본 방법을 적용하지 않은 경우에는 전체 구간의 평균을 구하지만, 본 방법을 적용한 경우에는 아래 수학식 2 와 같이 구해진다.

수학식 2

$$mean = \frac{\sum_{t=1}^{t_N} s_t + \sum_{t=t_s}^L s_t}{L - (t_s - t_N)}$$

[0054]

[0055] 한편 본 실시예에서 실 음성 구간 정규화 단계(S400)는 실 음성 구간 검출 단계(S300)에서 잡음 구간의 검출 여부를 판단(S410)하고 이에 따라 정규화 범위를 결정하는 것이 바람직하다.

[0056] 따라서 실 음성 구간 검출 단계(S300)에서 음성 신호 및 잡음 신호의 시작점이 찾아진 경우 특징 벡터의 정규화는 잡음 구간을 제외하고 이루어지는 것(S420)이 바람직하나, 잡음 신호의 위치 또는 음성 신호의 위치가 검출되지 않은 경우 검출된 프레임 단위 끝점 검출 단계(S100)에서 검출된 음성 구간에 대한 특징벡터를 정규화하는 것(S430)이 바람직하다.

[0057] 이하 도 6 및 도 7을 참고하여 본 실시예에 따른 음성 신호의 검출 방법이 적용된 음성 신호 검출 시스템에 대하여 설명한다.

[0058] 도 6을 참고하면, 본 실시예에 따른 음성 신호 검출 시스템은, 음성 신호원(10), 사용자 단말기(100), 음성 신호 검출 서버(200), 데이터 베이스(300)를 포함한다.

[0059] 음성 신호원(10)은 음성 신호를 사용자 단말기(100)에 제공하는 사용자를 포함한다.

[0060] 음성 신호 검출 서버(200)는 사용자 단말기(100)로부터 입력 받은 신호를 통해 상술한 음성 신호의 검출 방법을 수행하고, 그 결과를 사용자 단말기(100)에 전송한다.

[0061] 데이터 베이스(300)는 음성 신호 검출 서버에서 음성 및 비음성 신호의 확률 모델을 이용하여 음성일 확률 또는 음성이 아닐 확률을 계산하기 위한, 또는 신호의 에너지를 문턱값과 비교하여 음성 신호임을 판정하기 위한 데이터들을 저장 및 관리한다.

[0062] 이하 도 7을 참조하여 음성 신호 검출 서버(200)에 대하여 상세히 설명하면, 본 실시예에 따른 음성 신호 검출 서버(200)는 끝점 검출부(210), 특징값 추출부(220), 실 음성 구간 검출부(230), 정규화부(240), 제어부(250), 통신부(260)를 포함한다.

[0063] 끝점 검출부(210)는 사용자 단말기를 통해 입력된 입력 신호의 음성 구간을 검출하기 위하여 프레임 단위의 끝점 검출을 수행한다.

[0064] 특징값 추출부(220)는 끝점 검출부(210)에서 검출된 음성 구간 중 일부 구간의 특징 값을 추출하기 위한 복수의 윈도우들의 특징값을 추출한다.

[0065] 실 음성 구간 검출부(230)는 추출된 특징값을 미리 결정된 임계값과 비교하여 음성 구간 중 실 음성 구간을 검출한다.

[0066] 또한 정규화부(240)는 실 음성 구간의 특징벡터를 정규화하거나, 잡음 신호의 위치 또는 음성 신호의 위치가 검출되지 않은 경우 검출된 음성 구간에 대한 특징벡터를 정규화한다.

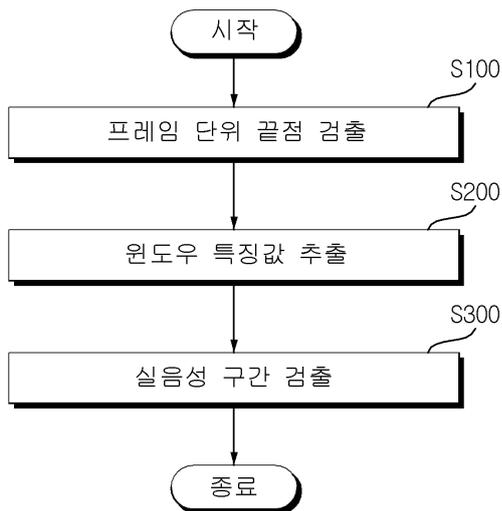
[0067] 통신부(260)는 사용자 단말기를 통해 신호를 입력 받고 검출된 음성 구간의 결과를 사용자 단말기에게 다시 제공한다. 나아가 제어부(250)는 상술한 끝점 검출부(210), 특징값 추출부(220), 실 음성 구간 검출부(230), 정규화부(240), 통신부(260)들의 작동을 제어한다.

[0068] 이상의 끝점 검출부(210), 특징값 추출부(220), 실 음성 구간 검출부(230), 정규화부(240)의 세부적인 동작은 상술한 음성 신호의 검출 방법에 대응되는 것으로 이에 대한 상세한 설명은 중복되므로 생략한다.

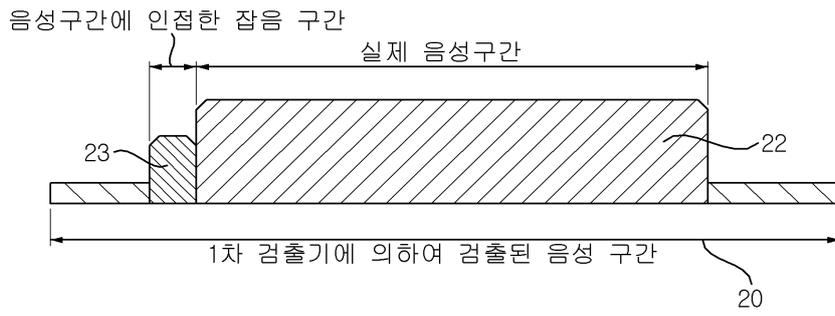
- [0069] 나아가 상술한 음성 신호 검출 서버(200) 서버로 구현되는 것 외에 사용자 단말기의 성능에 따라 단말기(100) 내부에 구현되는 것도 가능하며 이때의 통신부(260) 내부의 데이터 송수신 모듈에 해당하는 것이 바람직하다.
- [0070] 이상의 본 실시예에 따른 음성 신호의 검출 방법을 통하면 성 인식 시스템에서 사용하는 특징 벡터의 정규화 성능을 높일 수 있으며, 이는 잡음 환경에서의 음성 인식 성능을 높일 수 있다.
- [0071] 한편 본 발명의 음성 신호의 검출 방법은 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록 매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다.
- [0072] 컴퓨터가 읽을 수 있는 기록 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등이 있으며, 또한 컴퓨터가 읽을 수 있는 기록 매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다. 그리고 본 발명을 구현하기 위한 기능적인 (functional) 프로그램, 코드 및 코드 세그먼트 들은 본 발명이 속하는 기술 분야의 프로그래머들에 의하여 용이하게 추론될 수 있다.
- [0073] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위 내에서 다양한 수정, 변경 및 치환이 가능할 것이다.
- [0074] 따라서, 본 발명에 개시된 실시예 및 첨부된 도면들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예 및 첨부된 도면에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구 범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리 범위에 포함되는 것으로 해석되어야 할 것이다.

도면

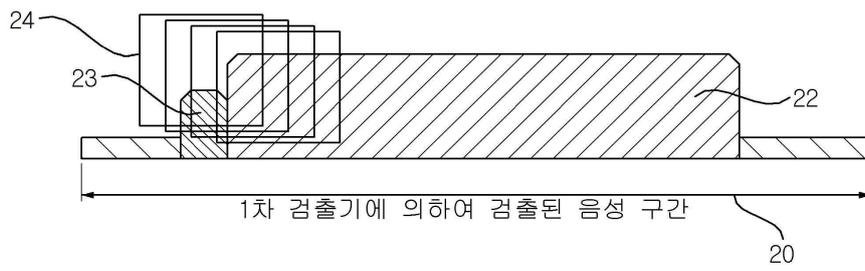
도면1



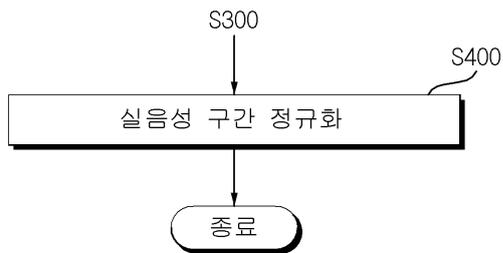
도면2



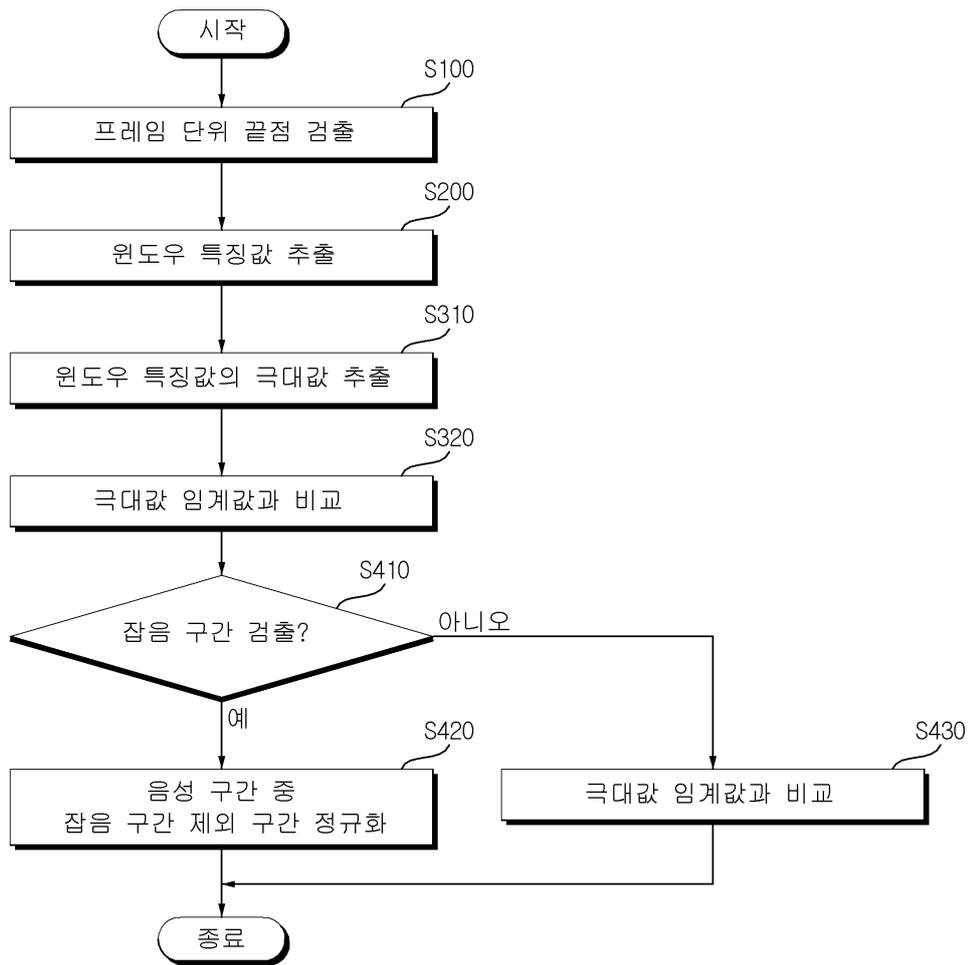
도면3



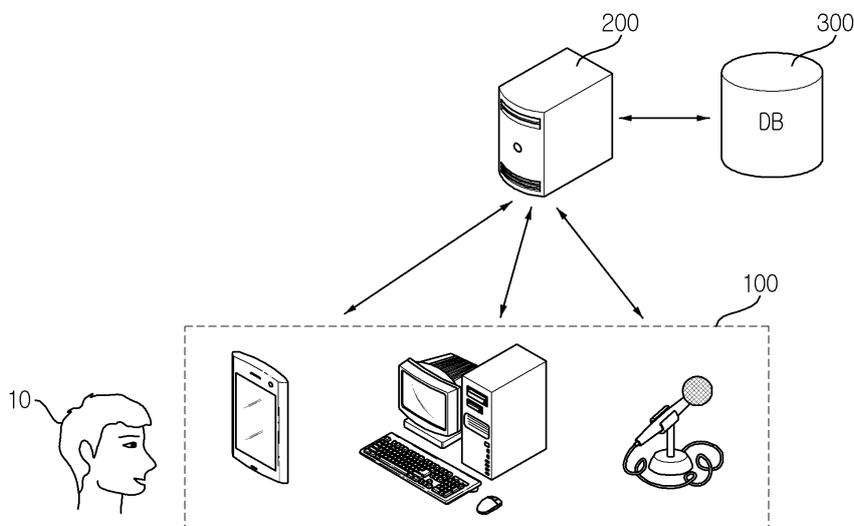
도면4



도면5



도면6



도면7

