



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0115033
(43) 공개일자 2010년10월27일

(51) Int. Cl.

G10L 15/08 (2006.01) G10L 19/06 (2006.01)

G10L 19/02 (2006.01)

(21) 출원번호 10-2009-0033531

(22) 출원일자 2009년04월17일

심사청구일자 2009년04월17일

(71) 출원인

고려대학교 산학협력단

서울 성북구 안암동5가 1

(72) 발명자

육동석

서울 용산구 서빙고동 신동아아파트 1동 102호

유인철

서울특별시 성북구 정릉2동 동아푸른아파트 105동 403호

(74) 대리인

특허법인충현

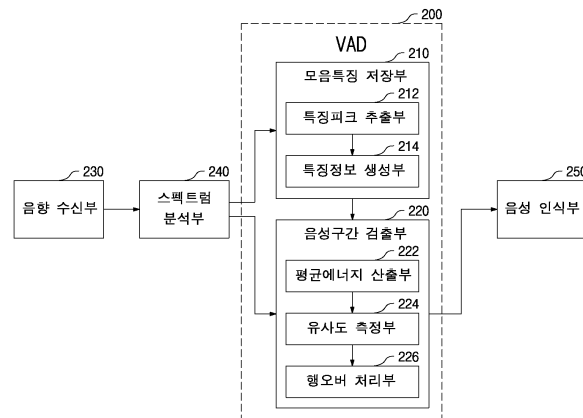
전체 청구항 수 : 총 19 항

(54) 모음 특징을 이용한 음성구간 검출 시스템 및 방법과 이에 사용되는 음향 스펙트럼 유사도 측정 방법

(57) 요약

본 발명은 음성 인식을 위한 음성구간 검출 기술에 관한 것으로서, 본 발명에 따른 모음 특징을 이용한 음성구간 검출 시스템은 모음의 스펙트럼에서 특징 피크가 위치한 피크 대역을 나타내는 모음특징정보를 저장하는 모음특징 저장부; 및 입력 음향의 스펙트럼에서 상기 저장된 모음특징정보가 나타내는 피크 대역에 대응하는 대응 대역 및 상기 대응 대역을 제외한 비대응 대역의 평균에너지를 이용하여 상기 입력 음향이 음성에 해당하는지를 판단함으로써 음성구간을 검출하는 음성구간 검출부를 포함하는 것을 특징으로 하여, 다양한 잡음 환경 및 SNR 하에서 음성구간 검출 성능을 개선함은 물론, 연산량을 감소시켜 음성 인식 시스템의 에너지 효율을 개선하는 이점을 제공한다.

대표도 - 도2



특허청구의 범위

청구항 1

모음의 스펙트럼에서 특정 피크가 위치한 피크 대역을 나타내는 모음특징정보를 저장하는 모음특징 저장부; 및

입력 음향의 스펙트럼에서 상기 저장된 모음특징정보가 나타내는 피크 대역에 대응하는 대응 대역 및 상기 대응 대역을 제외한 비대응 대역의 평균에너지를 이용하여 상기 입력 음향이 음성에 해당하는지를 판단함으로써 음성 구간을 검출하는 음성구간 검출부를 포함하는 모음 특징을 이용한 음성구간 검출 시스템.

청구항 2

제1항에 있어서,

상기 모음특징 저장부는,

상기 모음의 스펙트럼 피크들 중 미리 결정된 문턱값(threshold)보다 큰 에너지를 지니는 피크를 특정 피크로 추출하는 특징피크 추출부; 및

상기 추출된 특정 피크가 상기 모음의 스펙트럼에서 위치한 피크 대역을 나타내는 모음특징정보를 생성하는 특징정보 생성부를 포함하는 것을 특징으로 하는 모음 특징을 이용한 음성구간 검출 시스템.

청구항 3

제2항에 있어서,

상기 특징정보 생성부는, 상기 모음의 전체 스펙트럼 대역을 소정 개수의 단위 대역으로 구별하여, 상기 모음의 스펙트럼에서 상기 피크 대역에 해당하는 단위 대역을 1로 나타내고 상기 피크 대역 이외의 대역인 밸리 대역(valley band)에 해당하는 단위 대역을 0으로 나타내는 상기 모음특징정보를 생성하는 것을 특징으로 하는 모음 특징을 이용한 음성구간 검출 시스템.

청구항 4

제1항에 있어서,

상기 음성구간 검출부는,

상기 입력 음향의 스펙트럼에서, 상기 저장된 모음특징정보의 상기 피크 대역에 대응하는 대응 대역 및 상기 대응 대역을 제외한 비대응 대역의 평균에너지를 산출하는 평균에너지 산출부; 및

상기 대응 대역 및 상기 비대응 대역 간의 평균에너지 차를 이용하여 상기 모음 및 상기 입력 음향 간 스펙트럼 유사도를 측정하는 유사도 측정부를 포함하는 것을 특징으로 하는 모음 특징을 이용한 음성구간 검출 시스템.

청구항 5

제4항에 있어서,

상기 유사도 측정부는, 상기 대응 대역의 평균에너지에서 상기 비대응 대역의 평균에너지를 감산한 값을 나타내는 PVED(Peak-Valley Energy Difference) 값을 이용하여 상기 스펙트럼 유사도를 측정하는 것을 특징으로 하는 모음 특징을 이용한 음성구간 검출 시스템.

청구항 6

제5항에 있어서,

상기 음성구간 검출부는, 상기 PVED 값이 미리 결정된 임계값보다 큰 경우 상기 입력 음향을 음성으로 판단하여 음성구간으로 검출하는 것을 특징으로 하는 모음 특징을 이용한 음성구간 검출 시스템.

청구항 7

제6항에 있어서,

상기 음성구간 검출부는, 상기 저장된 모음특징정보들에 대하여 상기 PVED 값을 산출하고, 상기 산출된 PVED 값

중 최대 PVED 값이 상기 임계값보다 큰 경우 상기 입력 음향을 음성으로 판단하는 것을 특징으로 하는 모음 특징을 이용한 음성구간 검출 시스템.

청구항 8

제6항에 있어서,

상기 음성구간 검출부는, 상기 PVED 값이 상기 임계값보다 크지 않은 경우 상기 입력 음향에 대해 행오버(Hang-over) 알고리즘을 적용하는 행오버 처리부를 더 포함하고,

상기 음성구간 검출부는, 상기 행오버 처리부에 의해 상기 입력 음향이 음성으로 처리되는 경우 상기 입력 음향을 음성으로 판단하는 것을 특징으로 하는 모음 특징을 이용한 음성구간 검출 시스템.

청구항 9

음성구간 검출 시스템(Voice Activity Detection system)에서 상기 시스템에 입력되는 입력 음향들의 음성구간을 검출하는 방법에 있어서,

모음의 스펙트럼에서 특징 피크가 위치한 피크 대역을 나타내는 모음특징정보를 저장하는 모음특징 저장 단계; 및

입력 음향의 스펙트럼에서 상기 저장된 모음특징정보가 나타내는 피크 대역에 대응하는 대응 대역 및 상기 대응 대역을 제외한 비대응 대역의 평균에너지를 이용하여 상기 입력 음향이 음성에 해당하는지를 판단함으로써 음성구간을 검출하는 음성구간 검출 단계를 포함하는 모음 특징을 이용한 음성구간 검출 방법.

청구항 10

제9항에 있어서,

상기 모음특징 저장 단계는,

상기 모음의 스펙트럼 피크들 중 미리 결정된 문턱값(threshold)보다 큰 에너지를 지니는 피크를 특징 피크로 추출하는 특징피크 추출 단계; 및

상기 추출된 특징 피크가 상기 모음의 스펙트럼에서 위치한 피크 대역을 나타내는 모음특징정보를 생성하는 특징정보 생성 단계를 포함하는 것을 특징으로 하는 모음 특징을 이용한 음성구간 검출 방법.

청구항 11

제10항에 있어서,

상기 특징정보 생성 단계는, 상기 모음의 전체 스펙트럼 대역을 소정 개수의 단위 대역으로 구별하여, 상기 모음의 스펙트럼에서 상기 피크 대역에 해당하는 단위 대역을 1로 나타내고 상기 피크 대역 이외의 대역인 밸리 대역(valley band)에 해당하는 단위 대역을 0으로 나타내는 상기 모음특징정보를 생성하는 단계인 것을 특징으로 하는 모음 특징을 이용한 음성구간 검출 방법.

청구항 12

제9항에 있어서,

상기 음성구간 검출 단계는,

상기 입력 음향의 스펙트럼에서, 상기 저장된 모음특징정보의 상기 피크 대역에 대응하는 대응 대역 및 상기 대응 대역을 제외한 비대응 대역의 평균에너지를 산출하는 평균에너지 산출 단계; 및

상기 대응 대역 및 상기 비대응 대역 간의 평균에너지 차를 이용하여 상기 모음 및 상기 입력 음향 간 스펙트럼 유사도를 측정하는 유사도 측정 단계를 포함하는 것을 특징으로 하는 모음 특징을 이용한 음성구간 검출 방법.

청구항 13

제12항에 있어서,

상기 유사도 측정 단계는, 상기 대응 대역의 평균에너지에서 상기 비대응 대역의 평균에너지를 감산한 값을 나

타내는 PVED(Peak-Valley Energy Difference) 값을 이용하여 상기 스펙트럼 유사도를 측정하는 단계인 것을 특징으로 하는 모음 특징을 이용한 음성구간 검출 방법.

청구항 14

제13항에 있어서,

상기 음성구간 검출 단계는, 상기 PVED 값이 미리 결정된 임계값보다 큰 경우 상기 입력 음향을 음성으로 판단하여 음성구간으로 검출하는 단계인 것을 특징으로 하는 모음 특징을 이용한 음성구간 검출 방법.

청구항 15

제14항에 있어서,

상기 음성구간 검출 단계는, 상기 저장된 모음특징정보들에 대하여 상기 PVED 값을 산출하고, 상기 산출된 PVED 값 중 최대 PVED 값이 상기 임계값보다 큰 경우 상기 입력 음향을 음성으로 판단하는 단계인 것을 특징으로 하는 모음 특징을 이용한 음성구간 검출 방법.

청구항 16

제14항에 있어서,

상기 음성구간 검출 단계는, 상기 PVED 값이 상기 임계값보다 크지 않은 경우 상기 입력 음향에 대해 행오버(Hang-over) 알고리즘을 적용하는 행오버 처리 단계를 더 포함하고,

상기 음성구간 검출 단계는, 상기 행오버 처리 단계에 의해 상기 입력 음향이 음성으로 처리되는 경우 상기 입력 음향을 음성으로 판단하는 단계인 것을 특징으로 하는 모음 특징을 이용한 음성구간 검출 방법.

청구항 17

음성구간 검출 시스템(Voice Activity Detection system)에서 사전 학습된 모음 및 상기 시스템에 입력되는 입력 음향 간 주파수 스펙트럼의 유사도(similarity)를 측정하는 방법에 있어서,

상기 입력 음향의 스펙트럼에서, 상기 모음의 스펙트럼 상 특징 피크가 위치한 피크 대역에 대응하는 대응 대역 및 상기 대응 대역을 제외한 비대응 대역의 평균에너지를 산출하는 단계; 및

상기 대응 대역 및 상기 비대응 대역 간의 평균에너지 차를 이용하여 상기 모음 및 상기 입력 음향 간 스펙트럼 유사도를 측정하는 단계를 포함하는 음향 스펙트럼 유사도 측정 방법.

청구항 18

제17항에 있어서,

상기 유사도 측정 단계는, 상기 대응 대역의 평균에너지에서 상기 비대응 대역의 평균에너지를 감산한 값을 나타내는 PVED(Peak-Valley Energy Difference) 값을 이용하여 상기 스펙트럼 유사도를 측정하는 단계인 것을 특징으로 하는 음향 스펙트럼 유사도 측정 방법.

청구항 19

제9항 내지 제18항 중 어느 한 항에 따른 방법을 컴퓨터로 실행하기 위한 프로그램이 기록된 컴퓨터 판독가능 기록매체.

명세서

발명의 상세한 설명

기술 분야

본 발명은 음성 인식을 위한 음성구간 검출 기술에 관한 것으로서, 더욱 상세하게는 모음의 특징적 스펙트럼 피크를 이용하여 다양한 잡음 환경 및 SNR 하에서 음성구간 검출 성능을 개선함은 물론, 연산량을 감소시켜 음성 인식 시스템의 에너지 효율을 개선하는 음성구간 검출 시스템 및 방법과 이에 사용되는 음향 스펙트럼 유사도

[0001]

측정 방법에 관한 것이다.

배경 기술

- [0002] 음성 인식(voice recognition)이란, 자동적인 음성 인식 시스템을 통해 인간으로부터 발화된 음성으로부터 언어적 의미, 내용을 식별하는 것으로서, 구체적으로는 음성 파형을 입력하여 단어나 단어열을 식별하고 의미를 추출하는 처리 과정을 말한다. 이러한 음성 인식 기술은 가정용 전자기기, 휴대폰, 보안 및 인증 등 다양한 분야에 적용되고 있으며, 그 적용분야 및 수요는 더욱 급속히 증가할 전망이다.
- [0003] 한편, 음성을 수행하는 음성 인식 시스템(voice recognition system)은 크게 단어 음성 인식 시스템(word recognition system), 연속 음성 인식 시스템(continuous speech recognition system) 및 화자 인식 시스템(speaker recognition system)으로 분류할 수 있으며, 상기 음성 인식 시스템들에 적용되는 음성 인식 알고리즘들은 일반적으로 음성구간 검출 과정, 특징(feature) 추출 과정, 그리고 매칭(matching) 과정을 포함하고 있다.
- [0004] 상기 음성 인식 과정들 중 상기 음성구간 검출(Voice Activity Detection: VAD) 과정은 발생한 특정 소리가 인간의 목소리에 해당하는지를 판별하는 기술로서, 주로 음성 인식의 전처리 단계로 적용되는 기술이다. 이러한 음성구간 검출 기술에 있어서, 연구·개발시의 실험실 환경은 물론 다양한 잡음이 존재하는 실제 생활 환경에서 높은 인식을 내지 정확도를 유지할 수 있는지 여부는 매우 중요한 문제이다.
- [0005] 음성구간 검출에 관한 기존의 기술들로는, 음성 및 음성 이외의 소리에 관한 학습 데이터를 이용하여 모델을 학습시킨 후 음성구간을 검출하는 기계학습 방식과, 음성의 특성과 깊이 관련된 특징(zero crossing rate, spectral entropy 등)을 모델링하고 해당 특징의 출현 여부를 탐색하여 음성구간을 검출하는 방식이 소개된 바 있다.
- [0006] 그러나, 기존의 기계학습 방식의 경우, 해당 모델이 음성 이외의 모든 소리를 학습하는 것은 불가능하며, 그로 인해 학습 되지 않은 소리에 대해 음성구간 검출 성능이 크게 떨어질 수밖에 없다는 문제점이 있다.
- [0007] 또한, 기존의 음성 특징을 모델링하는 방식의 경우, 음성 이외의 많은 잡음들도 음성과 유사한 특징을 나타내며, 그로 인해 음성이 아닌 잡음을 음성으로 오인식하는 문제점이 있다.
- [0008] 요컨대, 기존 기술들은 음성 전체를 대상으로 특징을 찾아내어 이를 이용한다는 점에서 음성구간 검출 성능을 저해하는 근본적인 문제점이 있다. 그 이유는, 인간의 음성은 폐, 입술, 혀, 비강 등 다양한 조음기관의 상호작용을 거쳐 발생하는 매우 복잡한 소리 신호에 대하여 음성 전체를 아우르는 단일한 특징을 찾는 것이 매우 곤란하고, 음성 신호와 음성 이외의 소리 신호를 구분하는 명백한 특징도 밝혀진 바 없기 때문이다.
- [0009] 한편, 기존의 음성구간 검출 방식들의 경우 제한된 실험실 환경에서는 높은 인식을 보이는 반면, 실제 생활 환경, 예컨대 집 또는 사무실 등에서는 인식률의 급격한 저하를 보이는 다른 주된 요인으로 크게 두 가지를 들 수 있다.
- [0010] 한 가지 요인은, 배경 잡음(background noise) 및 실내 잔향(room reverberation) 등에 기인한 신호 손상(signal corruption)이다. 이러한 신호 손상은 동일한 음성 신호인 경우라도 음성 인식 시스템의 학습시 사용되는 깨끗한 음성 신호와 실제 환경에서 입력되는 잡음이 포함된 음성 신호 간에 매칭을 곤란하게 한다.
- [0011] 그러나, 상기 신호 손상 문제를 해결하려는 기존의 연구들은, 단지 특정 조건들 하에서만 상기 신호 손상의 영향을 어느 정도 감소시킬 수 있을 뿐, 실제 생활 사용되는 기기들에 적용되었을 때 일관되고 신뢰성 있는 성능을 보여줄 수 없었다. 그 이유는, 실제 생활 환경에서 발생할 수 있는 배경 잡음은 매우 다양하여 모든 배경 잡음을 예상할 수 없을 뿐 아니라, 배경 잡음의 특성도 시간에 따라 변화한다는 점에 기인한다. 따라서, S. F. Boll이, "Suppression of acoustic noise in speech using spectral subtraction"(IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-27(2), pp. 113-120, 1979.)에서 소개한 바와 같이 배경 잡음을 추정하는 알고리즘들을 적용하여 신호 손상 문제를 해결하는 것은 한계가 있다.
- [0012] 다른 한 가지 요인은, 음성 인식 시스템으로 하여금 학습하지 않은 음향 신호를 식별하여 음성 신호 이외의 음향 신호를 이미 학습한 음성 신호로 인식하지 않도록 하는 오픈-셋(open-set) 문제이다.
- [0013] 그러나, 상기 오픈-셋 문제를 해결하려는 기존의 연구들은, 실제 생활 환경에서 신뢰할 만한 성능을 달성할 수 없었다. 그 이유는, 앞서 설명한 바와 같이 실제 생활 환경에서 사전에 방대한 음향들의 특징들을 모두 학습하는 것은 사실상 불가능하기 때문이다.
- [0014] 요컨대, 기존 기술은 다양한 잡음이 발생하는 실제 생활 환경에서 신호 손상 및 오픈-셋 문제를 해결하여 안정

된 인식률을 제공할 수 있는 직접적인 해결책을 제시하지 못하고 있다는 문제점이 있다. 또한, 기존 기술은 음성구간 검출시 정보 처리 및 연산량을 감소시켜 오늘날 급속도로 발전하고 있는 이동 통신 시스템, USN(Ubiquitous Sensor Network) 등 자원 제약적 환경에 용이하게 적용할 수 있는 음성 인식 기술을 제시하지 못하고 있다는 문제점이 있다.

발명의 내용

해결 하고자 하는 과제

- [0015] 따라서, 본 발명이 해결하고자 하는 첫 번째 기술적 과제는, 모음의 특징적 스펙트럼 피크를 이용하여 다양한 잡음 환경 및 SNR 하에서 음성구간 검출 성능을 개선함은 물론, 연산량을 감소시켜 음성 인식 시스템의 에너지 효율을 개선하는 음성구간 검출 시스템을 제공하는 것이다.
- [0016] 본 발명이 해결하고자 하는 두 번째 기술적 과제는, 모음의 특징적 스펙트럼 피크를 이용하여 다양한 잡음 환경 및 SNR 하에서 음성구간 검출 성능을 개선함은 물론, 연산량을 감소시켜 음성 인식 시스템의 에너지 효율을 개선하는 음성구간 검출 방법을 제공하는 것이다.
- [0017] 본 발명이 해결하고자 하는 세 번째 기술적 과제는, 상기 음성구간 검출 시스템 및 방법에 사용되는 음향 스펙트럼 유사도 측정 방법을 제공하는 것이다.

과제 해결수단

- [0018] 상기와 같은 첫 번째 기술적 과제를 해결하기 위하여 본 발명은, 모음의 스펙트럼에서 특징 피크가 위치한 피크 대역을 나타내는 모음특징정보를 저장하는 모음특징 저장부; 및 입력 음향의 스펙트럼에서 상기 저장된 모음특징정보가 나타내는 피크 대역에 대응하는 대응 대역 및 상기 대응 대역을 제외한 비대응 대역의 평균에너지를 이용하여 상기 입력 음향이 음성에 해당하는지를 판단함으로써 음성구간을 검출하는 음성구간 검출부를 포함하는 모음 특징을 이용한 음성구간 검출 시스템을 제공한다.
- [0019] 일 실시예에 있어서, 상기 모음특징 저장부는, 상기 모음의 스펙트럼 피크들 중 미리 결정된 문턱값(threshold)보다 큰 에너지를 지니는 피크를 특징 피크로 추출하는 특징피크 추출부; 및 상기 추출된 특징 피크가 상기 모음의 스펙트럼에서 위치한 피크 대역을 나타내는 모음특징정보를 생성하는 특징정보 생성부를 포함한다.
- [0020] 일 실시예에 있어서, 상기 특징정보 생성부는, 상기 모음의 전체 스펙트럼 대역을 소정 개수의 단위 대역으로 구별하여, 상기 모음의 스펙트럼에서 상기 피크 대역에 해당하는 단위 대역을 1로 나타내고 상기 피크 대역 이외의 대역인 밸리 대역(valley band)에 해당하는 단위 대역을 0으로 나타내는 상기 모음특징정보를 생성한다.
- [0021] 일 실시예에 있어서, 상기 음성구간 검출부는, 상기 입력 음향의 스펙트럼에서, 상기 저장된 모음특징정보의 상기 피크 대역에 대응하는 대응 대역 및 상기 대응 대역을 제외한 비대응 대역의 평균에너지를 산출하는 평균에너지 산출부; 및 상기 대응 대역 및 상기 비대응 대역 간의 평균에너지 차를 이용하여 상기 모음 및 상기 입력 음향 간 스펙트럼 유사도를 측정하는 유사도 측정부를 포함한다.
- [0022] 일 실시예에 있어서, 상기 유사도 측정부는, 상기 대응 대역의 평균에너지에서 상기 비대응 대역의 평균에너지를 감산한 값을 나타내는 PVED(Peak-Valley Energy Difference) 값을 이용하여 상기 스펙트럼 유사도를 측정한다.
- [0023] 일 실시예에 있어서, 상기 음성구간 검출부는, 상기 PVED 값이 미리 결정된 임계값보다 큰 경우 상기 입력 음향을 음성으로 판단하여 음성구간으로 검출한다.
- [0024] 일 실시예에 있어서, 상기 음성구간 검출부는, 상기 저장된 모음특징정보들에 대하여 상기 PVED 값을 산출하고, 상기 산출된 PVED 값 중 최대 PVED 값이 상기 임계값보다 큰 경우 상기 입력 음향을 음성으로 판단한다.
- [0025] 일 실시예에 있어서, 상기 음성구간 검출부는, 상기 PVED 값이 상기 임계값보다 크지 않은 경우 상기 입력 음향에 대해 행오버(Hang-over) 알고리즘을 적용하는 행오버 처리부를 더 포함하고, 상기 음성구간 검출부는, 상기 행오버 처리부에 의해 상기 입력 음향이 음성으로 처리되는 경우 상기 입력 음향을 음성으로 판단한다.
- [0026] 상기와 같은 두 번째 기술적 과제를 해결하기 위하여 본 발명은, 음성구간 검출 시스템(Voice Activity Detection system)에서 상기 시스템에 입력되는 입력 음향들의 음성구간을 검출하는 방법에 있어서, 모음의 스펙트럼에서 특징 피크가 위치한 피크 대역을 나타내는 모음특징정보를 저장하는 모음특징 저장 단계; 및 입력

음향의 스펙트럼에서 상기 저장된 모음특징정보가 나타내는 피크 대역에 대응하는 대응 대역 및 상기 대응 대역을 제외한 비대응 대역의 평균에너지를 이용하여 상기 입력 음향이 음성에 해당하는지를 판단함으로써 음성구간을 검출하는 음성구간 검출 단계를 포함하는 모음 특징을 이용한 음성구간 검출 방법을 제공한다.

- [0027] 상기와 같은 세 번째 기술적 과제를 해결하기 위하여 본 발명은, 음성구간 검출 시스템(Voice Activity Detection system)에서 사전 학습된 모음 및 상기 시스템에 입력되는 입력 음향 간 주파수 스펙트럼의 유사도(similarity)를 측정하는 방법에 있어서, 상기 입력 음향의 스펙트럼에서, 상기 모음의 스펙트럼 상 특징 피크가 위치한 피크 대역에 대응하는 대응 대역 및 상기 대응 대역을 제외한 비대응 대역의 평균에너지를 산출하는 단계; 및 상기 대응 대역 및 상기 비대응 대역 간의 평균에너지 차를 이용하여 상기 모음 및 상기 입력 음향 간 스펙트럼 유사도를 측정하는 단계를 포함하는 음향 스펙트럼 유사도 측정 방법을 제공한다.

효 과

- [0028] 본 발명은, 잡음에 의해 손상된 입력 음향 신호에 대한 특징 추출 과정 없이 학습된 모음의 특징적 스펙트럼 피크를 이용함으로써 다양한 잡음 환경 및 SNR(Signal-to-Noise Ratio) 하에서 음성구간 검출 성능을 개선하는 이점을 제공한다.
- [0029] 또한, 음성구간 검출 과정에서 정보 처리 및 연산량을 감소시켜 오늘날 급속도로 발전하고 있는 이동단말기기, USN의 센서노드 등 자원 제약적 환경에 용이하게 적용할 수 있다는 이점을 제공한다.
- [0030] 나아가, 음성 인식의 전처리 단계로 적용함으로써 비음성 구간에서 음성 인식 시스템으로 하여금 불필요한 동작을 제거하도록 하여 에너지 효율을 더욱 개선할 수 있다는 이점을 제공한다.

발명의 실시를 위한 구체적인 내용

- [0031] 본 발명에 관한 구체적인 내용의 설명에 앞서 이해의 편의를 위해 본 발명이 해결하고자 하는 기술적 과제의 개요를 우선 제시한다.
- [0032] 도 1에는 본 발명에 따른 모음 특징을 이용한 음성구간 검출 방법의 기본 원리가 도시되어 있다.
- [0033] 도 1을 참조하면, 본 발명은 기존 기술에서 음성 전체에 관한 특징 추출이 곤란하다는 문제와 신호 손상(signal corruption) 및 오픈-셋(open-set) 문제를 해결하기 위해, 음성 인식 과정에 있어서 잡음 왜곡이 발생한 입력 음향 신호로부터 특징을 추출하는 과정 없이, 학습 단계(110)를 통해 사전 학습된 잡음없는 모음 신호(112)로부터 스펙트럼 특징(114)을 추출 및 저장하고, 검출 단계(120)에서 상기 저장된 스펙트럼 특징(114)을 이용하여 상기 모음 신호(112) 및 잡음 왜곡이 발생한 입력 음향 신호(122) 간의 매칭을 수행하도록 한다. 그 이유는, 기존 기술의 신호 손상 및 오픈-셋 문제가 바로 잡음에 의해 왜곡된 입력 음향 신호(122)로부터 추출된 특징(124)을 이용하여 소정의 매칭을 수행함으로써 초래되었기 때문이다.
- [0034] 따라서, 본 발명은 모음의 스펙트럼 특징만을 추출하여 모음 신호 및 입력 음향 신호 간 매칭을 수행하며, 이를 위해 PVED(Peak-Valley Energy Difference)라는 새로운 유사도(similarity) 측정 기준을 도입한다.
- [0035] 이하, 본 발명의 기술적 과제의 해결 방안을 명확화하기 위해 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 상세하게 설명한다. 다만, 본 발명을 설명함에 있어서 관련된 공지기술 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불명료하게 할 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다. 또한, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자 등의 의도 또는 관례 등에 따라 달라질 수 있을 것이다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.
- [0036] 도 2에는 본 발명에 따른 모음 특징을 이용한 음성구간 검출 시스템의 일례가 블록도로 도시되어 있다.
- [0037] 도 3에는 본 발명에 따른 모음 특징을 이용한 음성구간 검출 방법의 일례가 흐름도로 도시되어 있다.
- [0038] 도 2 및 도 3을 참조하면, 우선 본 발명에 따른 음성구간 검출 시스템(200)은, 모음특징 저장부(210) 및 음성구간 검출부(220)를 포함하며, 음향 수신부(230) 및 스펙트럼 분석부(240)를 더 포함할 수 있다.
- [0039] 상기 모음특징 저장부(210)는, 학습 단계를 통해 모음의 주파수 스펙트럼에서 특징 피크(characteristic peak)가 위치한 피크 대역(peak band)을 나타내는 모음특징정보를 저장한다(S310, S320).
- [0040] 더욱 구체적으로 설명하면, 상기 모음특징 저장부(210)는 인간의 음성을 구성하는 모음에 대한 학습 단계를 수행하기 위해 특징피크 추출부(212) 및 특징정보 생성부(214)를 포함할 수 있다.

- [0041] 상기 특징피크 추출부(212)는, 상기 모음의 스펙트럼에서 상기 특징 피크를 추출한다(S310). 이 경우, 다양한 형태의 추출 방식을 적용하여 상기 특징 피크를 추출할 수 있다. 일 실시예에 있어서, 연산의 단순화를 위해 상기 특징피크 추출부(212)는 상기 모음의 스펙트럼 피크들 중 미리 결정된 문턱값(threshold)보다 큰 에너지를 지니는 피크를 특징 피크로 추출할 수 있다.
- [0042] 인간의 음성(voice 또는 speech)은 크게 자음(consonant)과 모음(vowel)으로 구성된다. 자음의 경우 지속시간이 매우 짧으며 항상 모음과 함께 발생되는 특징이 있다. 반면, 모음의 경우 상대적으로 지속시간이 길고 각각의 모음은 스펙트럼 상에 포먼트(formant)라 알려진 고유한 특징 피크들을 지니고 있다.
- [0043] 도 4에는 'a' 모음의 주파수 스펙트럼이 도시되어 있다.
- [0044] 도 4에 도시된 바와 같이, 모음의 경우 각각 고유한 특징 피크를 지니고 있음을 알 수 있다. 모음 스펙트럼 상에 특징 피크들이 존재하는 피크 대역은 주변 대역에 비해 현저히 높은 에너지를 지니기 때문에 잡음에 의한 왜곡에도 쉽게 사라지지 않는다. 따라서, 인간의 음성 중 모음 부분은 다양한 잡음이 발생하는 실제 생활 환경에서도 검출이 용이하며, 상기 모음 부분만 검출할 수 있으면 모음 앞의 수십~수백 ms를 함께 음성으로 간주하는 방식을 이용하여 자음 및 모음 전체를 검출할 수 있게 된다. 상기 특징피크 추출부(212)는 이러한 모음의 특징 피크들을 추출하는 것이다. 물론, 상술한 바와 같이, 단순한 에너지 기반 피크 추출 방식뿐만 아니라, 더욱 복잡한 알고리즘들이 사용될 수도 있다.
- [0045] 상기 특징정보 생성부(214)는, 상기 추출된 특징 피크가 상기 모음의 스펙트럼에서 위치한 피크 대역을 나타내는 모음특징정보를 생성한다(S320).
- [0046] 도 5에는 상기 모음특징정보 생성 과정이 도시되어 있다.
- [0047] 도 5를 참조하면, 상기 특징피크 추출부(212)가 예컨대 특정 모음의 스펙트럼(510)에서 4개의 특징 피크를 추출한 경우, 상기 특징정보 생성부(214)는 상기 4개의 특징 피크들이 위치한 4개의 피크 대역들을 나타내는 모음특징정보(520)를 생성한다.
- [0048] 일 실시예에 있어서, 상기 특징정보 생성부(214)는, 상기 모음의 전체 스펙트럼 대역을 소정 개수의 단위 대역으로 구별하여, 상기 모음의 스펙트럼에서 상기 피크 대역에 해당하는 단위 대역을 1로 나타내고 상기 피크 대역 이외의 대역인 밸리 대역(valley band)에 해당하는 단위 대역을 0으로 나타냄으로써 상기 모음특징정보를 생성한다. 예컨대, 상기 모음특징정보(520)는 피크 특징 벡터(peak signature vector) 형태로 생성될 수 있다. 즉, 모음 신호에 관한 학습 데이터에 DFT(Discrete Fourier Transform)를 적용하여 얻어진 N 차원(N dimension)의 평균 스펙트럼으로부터 상기 특징피크 추출부(212)가 상기 문턱값 이상의 특징 피크를 추출하면, 상기 특징정보 생성부(214)는 상기 특징 피크가 위치한 피크 대역(peak band)에 해당하는 각각의 차원에 1을 할당하고, 상기 피크 대역 이외의 대역인 밸리 대역(valley band)에 해당하는 각각의 차원에 0을 할당하여 N 차원의 피크 특징 벡터(peak signature vector)를 생성할 수 있다.
- [0049] 아래에서 다시 설명하겠지만, 이와 같이 상기 모음특징정보를 이진값으로 생성하여 저장함으로써 상기 모음특징 저장부(210)의 데이터 저장량을 감소시킬 수 있음은 물론, 본 발명에 따른 음향 스펙트럼 유사도 측정 방법을 통해 음성구간 검출시 연산량을 감소시키고 에너지 효율을 개선할 수 있게 된다.
- [0050] 상술한 바와 같이, 상기 모음특징 저장부(210)는 상기 모음특징정보를 저장함으로써 상기 모음을 학습한다. 일 실시예에 있어서, 상기 모음특징 저장부(210)에 의한 학습 단계(S310 내지 S320)는 생산자 내지 제조자의 의도 등에 따라 음성 인식 시스템의 생산 내지 제작시에 수행될 수 있다. 이 경우, 상기 특징피크 추출부(212) 및 특징정보 생성부(214)의 동작을 외부 장치에서 수행하고 상기 모음특징 저장부(210)는 상기 모음특징정보만을 저장하도록 구성할 수 있다. 또한, 일 실시예에 있어서, 음성 인식 시스템의 사용자 또는 운용자에 따라 커스터마이징(customizing)이 가능하도록 구성할 수 있다. 즉, 음성 인식 전, 사용자 또는 운용자가 잡음없는 환경에서 상기 음성구간 검출 시스템(200)으로 하여금 개별적으로 모음을 학습하도록 구성할 수 있다. 이 경우, 상기 음향 수신부(230)가 웨이브폼(waveform)의 음성 신호를 수신하여 상기 스펙트럼 분석부(240)로 전달하면, 상기 스펙트럼 분석부(240)는 주파수 분석을 통해 상기 음성 신호 중 모음 신호를 스펙트럼 벡터로 변경하여 상기 모음특징 저장부(210)로 전달하도록 구성할 수 있다.
- [0051] 그 다음, 상기 음성구간 검출부(220)는, 입력 음향에 대한 음성구간 검출(Voice Activity Detection; VAD) 단계에서 입력 음향의 스펙트럼 상 상기 저장된 모음특징정보가 나타내는 피크 대역에 대응하는 대응 대역(relevant band) 및 상기 대응 대역을 제외한 비대응 대역(irrelevant band)의 평균에너지(average energy)를

이용하여 상기 입력 음향이 음성에 해당하는지를 판단함으로써 음성구간을 검출한다(S330 내지 S370).

- [0052] 더욱 구체적으로 설명하면, 상기 음성구간 검출부(220)는 상기 입력 음향에 대한 VAD 단계를 수행하기 위해, 평균에너지 산출부(222) 및 유사도 측정부(224)를 포함하며, 행오버 처리부(226)를 더 포함할 수 있다.
- [0053] 우선, 상기 음향 수신부(230)가 웨이브폼(waveform)의 입력음향 신호를 수신하여 상기 스펙트럼 분석부(240)로 전달하면, 상기 스펙트럼 분석부(240)는 주파수 분석 과정을 통해 상기 입력음향 신호를 스펙트럼 벡터로 변경한다(S330). 상기 스펙트럼 분석부(240)는 DFT(Discrete Fourier Transform) 등을 이용하여 상기 주파수 분석을 수행할 수 있다.
- [0054] 상기 평균에너지 산출부(222)는, 상기 대응 대역 및 상기 비대응 대역의 단위 대역 당 평균에너지를 산출한다(S340).
- [0055] 그 다음, 상기 유사도 측정부(224)는, 상기 대응 대역 및 상기 비대응 대역 간의 평균에너지 차를 이용하여 상기 모음 및 상기 입력 음향 간 스펙트럼 유사도(similarity)를 측정한다(S350).
- [0056] 상술한 바와 같이, 인간의 음성 중 모음들은 특징적인 스펙트럼 피크를 지닌다. 따라서, 각 모음의 스펙트럼의 대역은 상기 특징 피크가 위치한 피크 대역(peak band)과 상기 피크 대역 이외의 대역인 밸리 대역(valley band)으로 구분될 수 있다.
- [0057] 특히 주목할 점은, 상기 밸리 대역은 상기 피크 대역에 비해 상당히 넓다는 것이다. 즉, 상기 피크 대역은 좁은 대역에 음향 에너지가 집중되어 있는 대역에 해당하는 반면, 상기 밸리 대역은 매우 넓은 대역에 음향 에너지가 산재되어 있는 대역에 해당한다. 그 결과, 상기 피크 대역의 평균에너지는 상대적으로 큰 값을 나타내고, 상기 밸리 대역의 평균에너지는 상대적으로 작은 값을 나타낸다.
- [0058] 마찬가지로, 상기 입력 음향이 상기 모음을 포함하는 음성에 해당한다면, 잡음의 존재에 무관하게 상기 입력 음향 스펙트럼에서 상기 대응 대역의 평균에너지는 상대적으로 큰 값을 나타내고, 상기 비대응 대역의 평균에너지는 작은 값을 나타내게 된다.
- [0059] 따라서, 본 발명은 상기 대응 대역 및 상기 비대응 대역의 평균에너지를 이용하여 상기 모음 및 상기 입력 음향 간 유사도를 측정하고자 하는 것이다.
- [0060] 일 실시예에 있어서, 상기 유사도 측정부(224)는 상기 대응 대역의 평균에너지에서 상기 비대응 대역의 평균에너지를 감산한 값을 나타내는 PVED(Peak-Valley Energy Difference) 값을 이용하여 상기 스펙트럼 유사도를 측정할 수 있다.
- [0061] 상기 PVED는 상기 모음의 모음특징정보인 피크 특징 벡터(S) 및 상기 입력 음향의 스펙트럼(X)만을 이용하여 유사도를 측정할 수 있도록 한다.
- [0062] 상기 PVED 값은 수학식 1과 같이 나타낼 수 있다.

수학식 1

$$PVED(X, S) = [\text{대응대역 평균에너지}] - [\text{비대응대역 평균에너지}]$$

상기 수학식 1은 구체적으로 수학식 2와 같이 산출될 수 있다.

수학식 2

$$PVED(X, S) = \left[\frac{\sum_{k=0}^{N-1} (X[k] \times S[k])}{\sum_{k=0}^{N-1} S[k]} \right] - \left[\frac{\sum_{k=0}^{N-1} (X[k] \times (1 - S[k]))}{\sum_{k=0}^{N-1} (1 - S[k])} \right]$$

상기 수학식 2에서, N은 상기 S 및 X의 차원을 나타내고, X[k]는 상기 입력 음향 스펙트럼(X)의 k 차원 에너지를 나타내고, S[k]는 상기 피크 특징 벡터(S)의 k 차원 이진값을 나타낸다. 즉, S[k]는 상기 모음의 상기 피크 대역에 대해 이진값 1을, 그리고 상기 밸리 지역에 대해 이진값 0을 지닌다.

상기 음성구간 검출부(220)는, 상기 PVED 값이 미리 결정된 임계값보다 큰 경우(S352), 상기 입력 음향을 상기

모음을 포함하는 음성으로 판단하여 상기 입력음향을 음성구간으로 검출한다(S360). 그러나, 상기 PVED 값이 상기 임계값 이하인 경우(S352), 상기 입력 음향을 잡음 또는 비음성 음향으로 판단한다(S362).

[0068] 일 실시예에 있어서, 상기 음성구간 검출부(220)는 상기 PVED 값이 상기 임계값 이하인 경우 상기 행오버 처리부(226)를 통해 상기 입력 음향에 대하여 행오버(Hang-over) 알고리즘을 적용한다(S354).

[0069] 상기 행오버 알고리즘이란, 일반적 음성 인식 시스템(voice recognition system)에서 발화시 발생하는 음성 중간의 묵음(intra-speech silence) 구간을 음성으로 함께 처리하는 과정을 말한다. 즉, 인간의 음성은 귀를 통해 듣는 경우 연속적으로 끊임없이 발생하는 것으로 들리지만, 상기 음성을 밀리센드(ms) 단위로 분석하면 음성 중간 중간에 몇십~몇백 ms 이하의 소리없는 짧은 휴지기간이 존재한다. 따라서, 상기 행오버 알고리즘은 상기 음성 중간에 등장하는 묵음을 음성으로 함께 처리하는 것이다.

[0070] 본 발명에서도 상기 행오버 알고리즘을 적용한다. 즉, 상기 음성구간 검출부(220)는 상기 행오버 처리부(226)에 의한 상기 행오버 알고리즘 적용 후(S354), 상기 입력 음향이 음성으로 처리되는 구간에 속하는 것이면(S356), 상기 입력 음향을 음성으로 판단하여 음성구간으로서 검출한다(S360).

[0071] 도 6에는 잡음 왜곡이 발생한 입력 음향의 스펙트럼이 도시되어 있다.

[0072] 도 6에 도시된 바와 같이, 채널 왜곡(channel distortion), 예컨대 잔향(reverberation)과 쏏 버스트(short burst) 잡음 등 배경 잡음들로 인해 모음 소리의 특징적 스펙트럼 피크들이 일부 묻혀버리는 경우는 물론, 배경 잡음에 불구하고 입력 음향 스펙트럼 상에 여전히 모음 소리의 특징적인 스펙트럼 피크들이 존재하는 경우라도, 기존 기술은 상기 입력 음향 스펙트럼으로부터 특징을 추출하여 모음 스펙트럼과의 유사성을 판단하기 때문에 오인식을 초래하게 된다. 그 이유는, 기존 기술에서 좁은 대역에 존재하는 피크들의 유사성 여부는 상대적으로 매우 넓은 대역에 걸쳐 존재하는 밸리들의 유사성 여부에 따라 무의미해 질 수 있기 때문이다.

[0073] 도 7a 내지 도 7c에는 도 5의 모음특징정보를 이용하여 산출한 입력 음향 스펙트럼의 대응 대역 및 비대응 대역의 상대적 평균에너지가 히스토그램으로 도시되어 있다.

[0074] 도 7a 및 도 7b에 도시된 바와 같이, 입력 음향에 배경 잡음이 존재하는 경우뿐만 아니라 추가적 피크들이 존재하는 경우에도 상기 입력 음향이 상기 모음을 포함하는 음성이라면, 상기 입력 음향의 스펙트럼에서 상기 대응 대역의 평균에너지(black bar)가 상기 비대응 대역의 평균에너지(white bar)에 비해 상대적으로 큰 값을 지닐 수 있다. 그러나, 도 7c에 도시된 바와 같이, 상기 입력 음향이 잡음에 불과하다면 상기 입력 음향의 스펙트럼에서 상기 대응 대역 및 상기 비대응 대역 간에 평균에너지 차가 크지 않음을 알 수 있다.

[0075] 따라서, 본 발명은 에너지 기반의 유사도 측정 방식, 즉 상기 PVED 기반의 유사도 측정 방식을 통해 오인식 발생의 원인이 되는 입력 음향 신호로부터의 특징 추출 과정을 제거하고, 모음 및 입력 음향 간 유사도 측정에 있어서 특정 관련 대역에 집중할 수 있도록 한다. 그 결과, 본 발명은 잡음 또는 비음성 음향이 음성으로 오인식 되는 경우를 방지할 수 있다. 이러한 결과는, 오직 모음의 특징 피크들이 입력 음향의 스펙트럼 상에 나타나는 경우에만 상기 PVED 값이 상대적으로 크게 나타나기 때문이다.

[0076] 일 실시예에 있어서, 상기 모음특징 저장부(210)에 저장된 모음특징정보들이 복수인 경우 상기 음성구간 검출부(220)는, 각각의 모음특징정보에 대한 상기 PVED 값을 산출하고, 상기 산출된 PVED 값 중 최대 PVED 값이 상기 임계값보다 큰 경우, 상기 입력 음향을 상기 최대 PVED 값에 대응하는 모음이 존재하는 음성으로 판단할 수 있다. 즉, 상기 학습 단계를 위한 모음 학습 데이터가 200개이었다면, 상기 학습 단계에서 200개의 모음특징정보, 예컨대 이진값들로 이루어진 피크 특징 벡터(peak signature vector)들이 생성되었을 것이고, 상기 모음특징정보들 각각에 대해서 입력 음향 신호의 PVED 값을 측정하여 최대 PVED 값을 상기 임계값과 비교하는 것이다.

[0077] 그 다음, 상기 음성구간 검출 시스템에 입력되는 후속 입력 음향이 존재하는 경우, 상술한 음성구간 검출 과정을 반복한다(S370).

[0078] 상기 음성구간 검출 시스템(200)에 의해 입력 음향들로부터 음성구간이 검출되면, 음성 인식부(250)는 기존 기술들을 이용하여 음성 분석, 음소 인식, 단어 인식, 문장 해석, 및 의미 추출 과정 등을 수행할 수 있다.

[0079] 한편, 본 발명은 컴퓨터로 판독할 수 있는 기록매체에 컴퓨터가 읽어들일 수 있는 프로그램 코드로서 구현하는 것이 가능하다. 본 발명이 소프트웨어를 통해 실행될 때, 본 발명의 구성 수단들은 필요한 작업을 실행하는 코드 세그먼트들이다. 프로그램 또는 코드 세그먼트들은 프로세서 판독 가능 매체에 저장되거나 전송 매체 또는 통신망에서 반송파와 결합된 컴퓨터 데이터 신호에 의하여 전송될 수 있다.

- [0080] 컴퓨터가 판독할 수 있는 기록매체에는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 판독할 수 있는 기록매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광데이터 저장장치 등이 있다. 또한, 컴퓨터가 판독할 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어 분산방식으로 컴퓨터가 읽어들일 수 있는 코드가 저장되고 실행될 수 있다.
- [0081] 이하, 본 발명의 현저한 효과를 검증한다.
- [0082] 본 발명의 성능 실험을 위해 TIMIT, NTIMIT(telephone), FFMTIMIT(far field microphone), 및 CTIMIT(cellular phone) 코퍼스(corpus)들을 사용하였다. 각각의 음향들은 10ms 간격을 두어 32ms 블록들로 세그먼트(segment)화하고, 각각의 블록에 해밍 윈도우(Hamming window)를 통한 푸리에 분석을 적용하였다.
- [0083] 또한, 본 발명의 성능을 비교 평가하기 위해 동일 테스트 데이터에 대하여 기존의 VAD 방식들, 즉 "A Silence Compression Scheme for G.729 Optimized for Terminals Conforming to ITU-T V.70"(ITU-T Rec. G.729 Annex B, 1996.)에 소개된 방식(G.729), "Digital Cellular Telecommunications System(Phase 2+); Voice Activity Detector(VAD) for Adaptive Multi Rate(AMR) Speech Traffic Channels"(GSM 06.94 v7.1.1(ETSI EN 301 708), 1999.)에 소개된 두 방식(AMR1 및 AMR2), A. Davis 외 2인이 "A statistical voice activity detection using low-variance spectrum estimation and an adaptive threshold"(*IEEE Transactions on Audio, Speech, and Language Processing*, vol.14, no.2, pp.412-424, 2006.)에서 소개한 방식(LVS), 그리고 J. L. Shen 외 2인이 "Robust entropy-based endpoint detection for speech recognition in noisy environments"(*Proceeding of International Conference on Spoken Language Processing*, paper 0232, 1998.)에서 소개한 방식(SE)을 사용하였다.
- [0084] 도 8에는 상기 4가지 코퍼스(corpus)에 대한 본 발명의 음성구간 검출 성능 실험결과가 도시되어 있다.
- [0085] 도 8에 도시된 바와 같이, 본 발명(PVED)의 음성구간 검출 성능이 평균적으로는 물론, 각각의 코퍼스에 대해서도 기존 방식들보다 현저히 우수함을 알 수 있다.
- [0086] 도 9에는 다양한 SNR 레벨에 따른 본 발명의 음성구간 검출 성능 실험결과가 도시되어 있다.
- [0087] 도 10에는 다양한 배경 잡음에 따른 본 발명의 음성구간 검출 성능 실험결과가 도시되어 있다.
- [0088] 다양한 SNR 레벨 및 배경 잡음에 따른 본 발명의 검출 성능을 측정하기 위해, 0dB에서 30dB까지 5dB 단위로 7가지 SNR 레벨의 배경 잡음들을 생성하고 전체 TIMIT 코퍼스와 혼합하여 테스트 데이터를 생성하였다. 상기 배경 잡음으로는 10가지 배경 잡음, 즉 화이트 노이즈(white noise), 핑크 노이즈(pink noise), 그리고 오로라-2 코퍼스(Aurora-2 corpus)에서의 8개 잡음(공항, 전람회, 식당, 거리, 지하철, 자동차, 기차, 및 다중 누화(babble) 잡음)을 사용하였다. 즉, $441000\text{개}(441000=\text{TIMIT}(6300)\times\text{잡음종류}(10)\times\text{SNR레벨}(7))$ 의 테스트 데이터를 사용하였다.
- [0089] 도 9 및 도 10에 도시된 바와 같이, 본 발명(PVED)의 음성구간 검출 성능이 평균적으로는 물론, 각각의 SNR 레벨 및 배경 잡음에 대해서도 기존 방식들보다 현저히 우수함을 알 수 있다.
- [0090] 상술한 바와 같이, 본 발명은 잡음에 의해 손상된 입력 음향 신호에 대한 특징 추출 과정 없이 학습된 모음의 특징적 스펙트럼 피크를 이용함으로써 다양한 잡음 환경 및 SNR 하에서 음성구간 검출 성능을 개선하는 이점을 제공한다. 또한, 음성구간 검출 과정에서 정보 처리 및 연산량을 감소시켜 오늘날 급속도로 발전하고 있는 이동 단말기기, USN의 센서노드 등 자원 제약적 환경에 용이하게 적용할 수 있다는 이점을 제공한다. 나아가, 음성 인식의 전처리 단계로 적용함으로써 비음성 구간에서 음성 인식 시스템으로 하여금 불필요한 동작을 제거하도록 하여 에너지 효율을 더욱 개선할 수 있다는 이점을 제공한다.
- [0091] 지금까지 본 발명에 대해 실시예들을 참고하여 설명하였다. 그러나 당업자라면 본 발명의 본질적인 기술적 사상으로부터 벗어나지 않는 범위에서 본 발명이 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 즉, 본 발명의 진정한 기술적 범위는 첨부된 특허청구범위에 나타나 있으며, 그와 균등범위 내에 있는 모든 차이점은 본 발명에 포함되는 것으로 해석되어야 할 것이다.

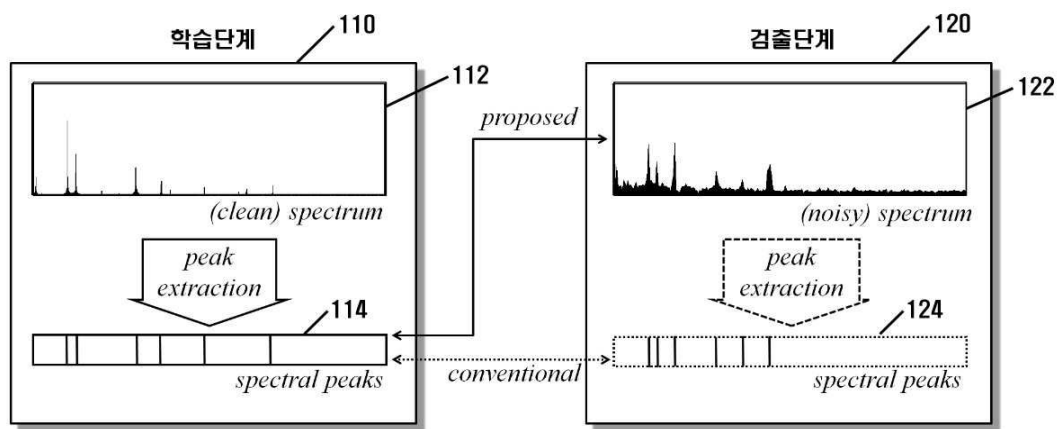
도면의 간단한 설명

- [0092] 도 1은 본 발명에 따른 모음 특징을 이용한 음성구간 검출 방법의 기본 원리를 나타낸 도면이다.

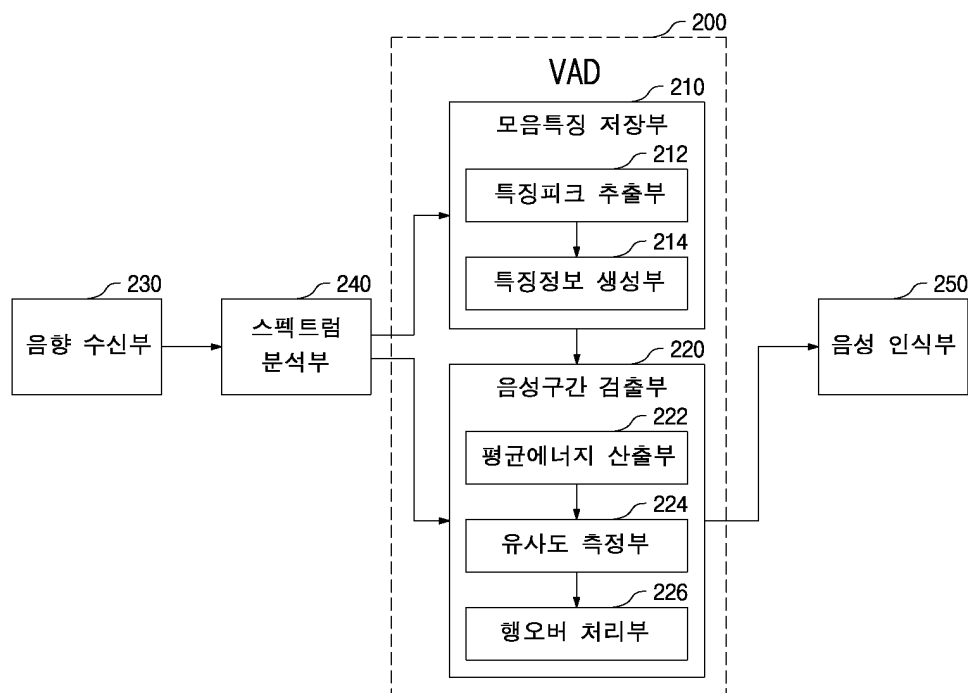
- [0093] 도 2은 본 발명에 따른 모음 특징을 이용한 음성구간 검출 시스템의 일례를 나타낸 블록도이다.
- [0094] 도 3은 본 발명에 따른 모음 특징을 이용한 음성구간 검출 방법의 일례를 나타낸 흐름도이다.
- [0095] 도 4는 'a' 모음의 주파수 스펙트럼을 나타낸 도면이다.
- [0096] 도 5는 본 발명의 모음특징정보 생성 과정을 나타낸 도면이다.
- [0097] 도 6은 잡음 왜곡이 발생한 입력 음향의 스펙트럼을 나타낸 도면이다.
- [0098] 도 7a 내지 도 7c는 도 5의 모음특징정보를 이용하여 산출한 입력 음향 스펙트럼의 대응 대역 및 비대응 대역의 상대적 평균에너지를 히스토그램으로 나타낸 도면이다.
- [0099] 도 8 내지 도 10은 본 발명의 음성구간 검출 성능 실험결과를 나타낸 도면이다.

도면

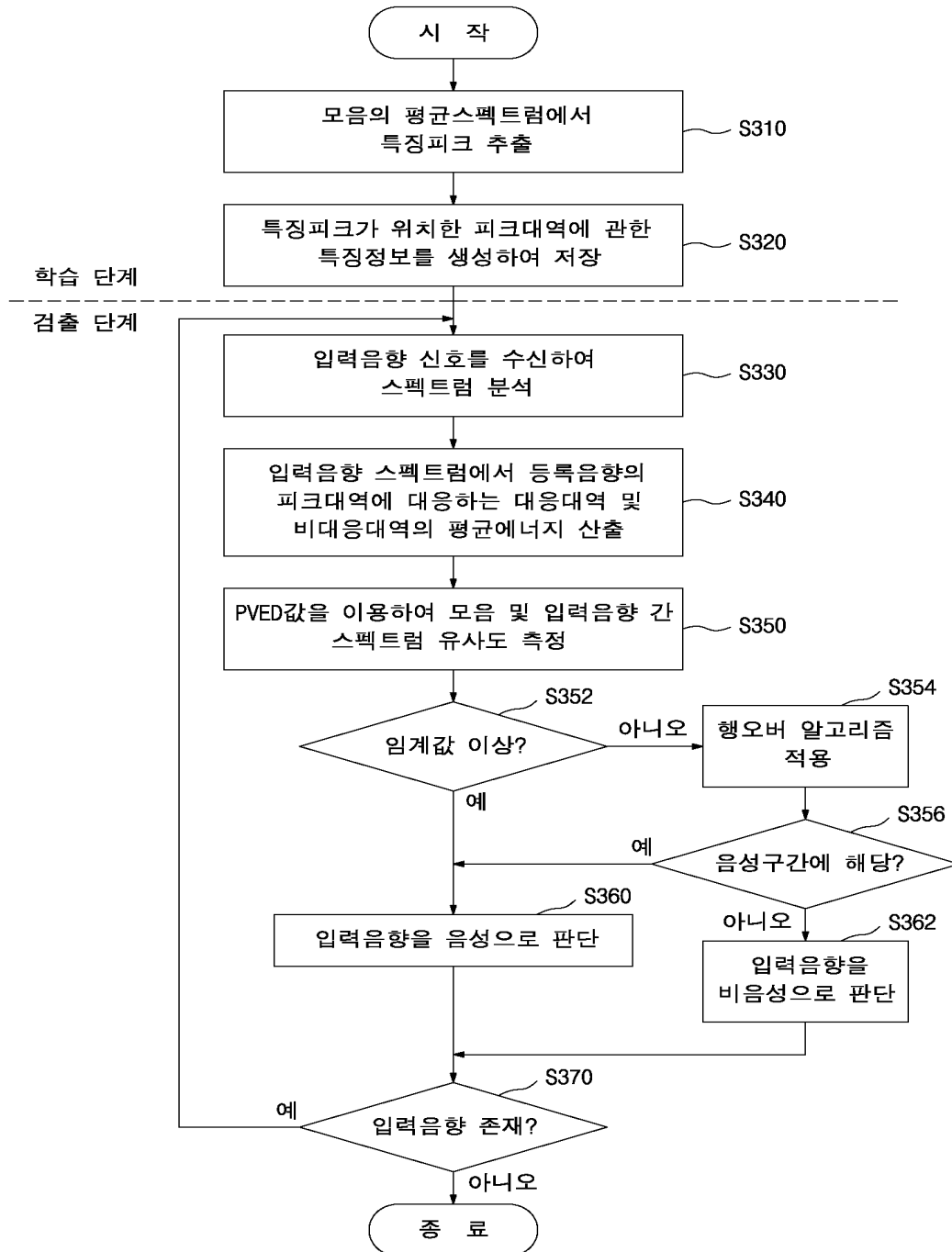
도면1



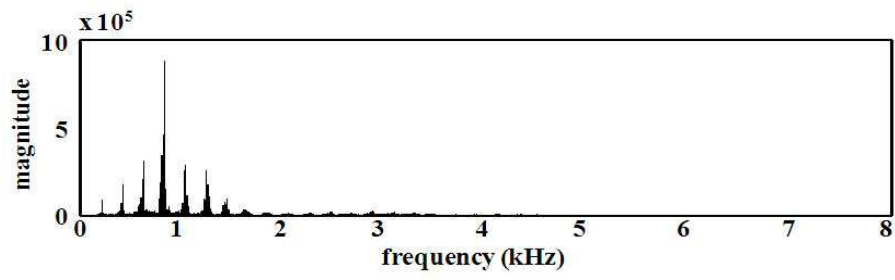
도면2



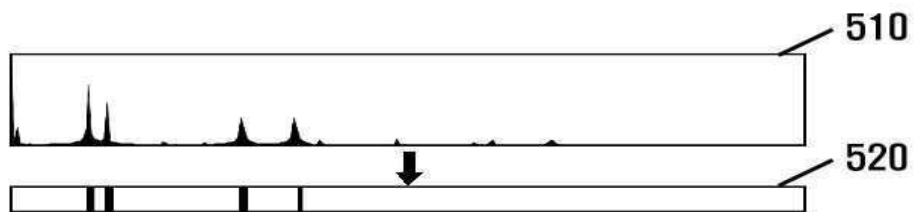
도면3



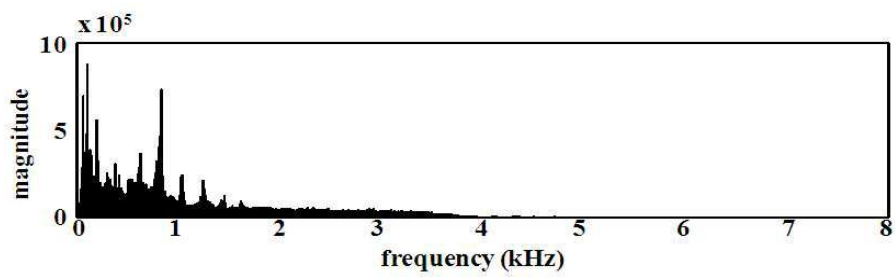
도면4



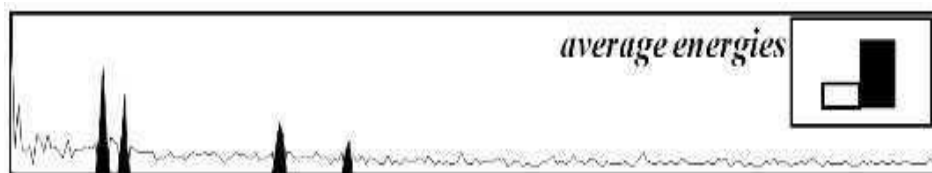
도면5



도면6



도면7a



도면7b



도면7c



도면8

	VAD accuracy (%) of various corpora				
	TIMIT	NTIMIT	FFMTIMIT	CTIMIT	Avg
G729	87.3	85.8	74.2	62.6	77.5
AMR1	89.9	88.1	89.4	72.8	85.1
AMR2	80.1	84.7	82.4	77.5	81.2
LVS	72.7	88.4	86.0	80.5	81.9
SE	83.3	17.9	89.3	37.6	57.0
PVED	91.9	91.8	92.1	84.2	90.0

도면9

	VAD accuracy (%) by SNR (dB)							
	0	5	10	15	20	25	30	Avg
G729	55.0	60.3	64.3	67.2	69.0	69.8	70.3	65.1
AMR1	57.2	57.1	58.9	63.6	72.4	82.9	88.8	68.7
AMR2	79.7	79.4	79.2	79.5	81.3	84.4	88.9	81.8
LVS	68.2	81.3	90.6	94.8	96.0	95.8	95.0	88.8
SE	65.5	72.2	76.4	79.0	80.8	81.8	82.4	76.9
PVED	89.8	93.3	95.0	95.8	96.0	96.0	96.0	94.5

도면10

	VAD accuracy (%) by noise types										
	Airport	Babble	car	exhibition	pink	restaurant	street	subway	train	white	Avg
G.729	62.5	68.4	65.8	66.5	65.7	63.2	58.	67.9	69.7	63.6	65.1
AMR1	68.9	62.0	73.5	69.2	75.1	63.7	62.0	61.0	80.7	70.7	68.7
AMR2	73.9	65.7	93.0	89.8	94.3	63.7	83.0	66.6	93.1	94.6	81.8
LVS	89.8	91.5	90.1	86.5	87.8	92.3	81.9	88.0	94.1	86.1	88.8
SE	73.1	75.6	81.9	91.3	87.7	73.6	60.1	90.6	43.0	91.8	76.9
PVED	95.8	92.5	95.8	95.5	95.9	93.8	88.5	95.7	95.9	95.9	94.5