(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2009년02월18일
(11) 등록번호 10-0883652
(24) 등록일자 2009년02월06일

(51) Int. Cl.
G10L 15/20 (2006.01) *G10L 15/04* (2006.01)
G10L 15/08 (2006.01) *G11B 20/10* (2006.01)
(21) 출원번호 10-2006-0073386
(22) 출원일자 2006년08월03일
심사청구일자 2006년08월03일
(65) 공개번호 10-2008-0012491
(43) 공개일자 2008년02월12일
(56) 선행기술조사문헌
KR1019970067095 A*
KR1020000056849 A*
JP2001350488 A
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
삼성전자주식회사
경기도 수원시 영통구 매탄동 416
노바우리스 테크놀러지스 리미티드
영국 지엘52 8알더블유 글로스터셔 첼튼햄 비숍스 클리브 스톡 로드 밀뱅크
(72) 발명자
장길진
경기 수원시 영통구 영통동 청명마을4단지아파트 403-1703
김정수
경기 용인시 수지구 상현동 현대7차아파트 506-901
(뒷면에 계속)
(74) 대리인
리앤목특허법인

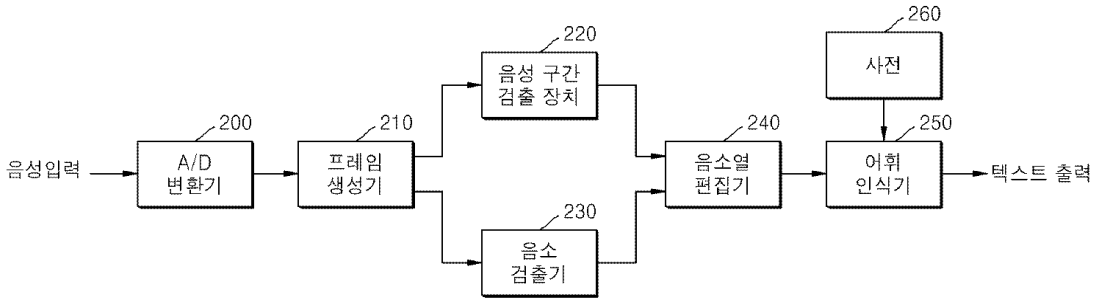
전체 청구항 수 : 총 16 항 심사관 : 경연정

(54) 음성 구간 검출 방법 및 장치, 및 이를 이용한 음성 인식시스템

(57) 요 약

본 발명은 동적 프로그래밍을 이용한 음성 구간 검출 방법 및 장치, 및 이를 이용한 음성 인식 시스템에 관한 것으로, 음성 구간 검출 방법은 음성 및 비음성을 포함하는 신호를 구성하는 프레임들 각각의 에너지의 변이를 검출하고, 검출된 결과에 기초하여 상기 프레임들 중 상기 음성만을 포함하는 프레임들의 구간에 해당하는 음성 구간을 결정함으로써, 주변에 크기가 작고 짧은 길이의 돌발 잡음이 자주 발생하는 환경, 또는 음성 입력 앞뒤에서 발생하는 작은 크기의 숨소리, 기계 마찰음, 입술소리 등에 의해 잘못 인식된 의사 음소 열을 제거하여 효과적으로 음성 구간을 검출할 수 있다.

대 표 도



(72) 발명자

브리들 존 에스

영국 지엘52 8알더블유 글로스터셔 첼튼햄 비숍스 클리브 스톡 로드 밀뱅크

헌트 멜빈 제이

영국 지엘52 8알더블유 글로스터셔 첼튼햄 비숍스 클리브 스톡로드 밀뱅크

특허청구의 범위

청구항 1

삭제

청구항 2

삭제

청구항 3

음성 및 비음성을 포함하는 신호를 구성하는 프레임들 각각의 에너지 레벨을 소정의 기준에 따라 분류하는 단계;

상기 분류된 에너지 레벨을 기초로 상기 프레임들 각각의 에너지 레벨의 변이가 일어나는 구간을 검출하는 단계; 및

상기 검출된 결과에 기초하여 상기 프레임들 중 상기 음성만을 포함하는 프레임들의 구간에 해당하는 음성 구간을 결정하는 단계를 포함하며,

상기 결정 단계는,

상기 프레임 에너지 레벨의 변이가 일어나는 구간마다 소정의 가중치를 부여하고, 상기 구간 전체의 가중치를 계산하는 과정을 반복적으로 수행하는 단계; 및

상기 계산된 전체 가중치들 중 최소가 되는 구간을 상기 신호의 음성 구간으로 결정하는 단계를 포함하는 것을 특징으로 하는 음성 구간 검출 방법.

청구항 4

제 3 항에 있어서,

상기 프레임들 각각의 에너지를 계산하는 단계를 더 포함하고,

상기 분류 단계는 상기 계산된 에너지에 따라 상기 프레임들 각각의 에너지 레벨을 분류하는 것을 특징으로 하는 음성 구간 검출 방법.

청구항 5

삭제

청구항 6

제 3 항에 있어서,

상기 변이는 인접한 프레임 에너지 레벨의 변이, 및 각각의 프레임에서 변경 전 에너지 레벨과 변경 후 에너지 레벨의 변이인 것을 특징으로 하는 음성 구간 검출 방법.

청구항 7

제 3 항에 있어서,

상기 신호의 에너지에 따라 상기 기준을 갱신하는 단계를 더 포함하는 것을 특징으로 하는 음성 구간 검출 방법.

청구항 8

제 7 항에 있어서,

상기 프레임의 에너지 레벨을 높음, 중간, 또는 낮음을 포함하는 3단계로 분류하는 것을 특징으로 하는 음성 구간 검출 방법.

청구항 9

제 3 항에 있어서,

상기 결정된 음성 구간을 상기 신호의 전체 구간으로 결합하는 단계를 더 포함하는 것을 특징으로 하는 음성 구간 검출 방법.

청구항 10

제 3 항에 있어서,

상기 비음성은,

짧은 시간에 주파수 특성이 변하는 돌발잡음을 포함하는 것을 특징으로 하는 음성 구간 검출 방법.

청구항 11

음성 및 비음성을 포함하는 신호를 구성하는 프레임들 각각의 에너지를 계산하는 단계;

상기 계산된 에너지에 따라 상기 프레임들 각각의 에너지 레벨을 분류하는 단계;

상기 분류된 에너지 레벨을 기초로 상기 프레임들 각각의 에너지 레벨의 변이가 일어나는 구간을 검출하는 단계; 및

상기 변이가 일어나는 구간마다 소정의 가중치를 부여하고, 상기 구간 전체의 가중치가 최소가 되는 구간을 상기 신호의 음성 구간으로 결정하는 단계를 포함하는 음성 구간 검출 방법.

청구항 12

제 11 항에 있어서,

상기 프레임들 각각의 에너지 레벨을 변경하여 생성하는 단계를 더 포함하고,

상기 검출단계는 인접한 프레임 에너지 레벨로의 변이 및 변경 전 에너지 레벨과 변경 후 에너지 레벨의 변이가 일어나는 구간을 검출하는 것을 특징으로 하는 음성 구간 검출 방법.

청구항 13

제 3 항 내지 제 4 항, 제 6 항 내지 제 12 항 중 어느 한 항에 의한 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록 매체.

청구항 14

삭제

청구항 15

삭제

청구항 16

음성 및 비음성을 포함하는 신호를 구성하는 프레임들 각각의 에너지 레벨을 소정의 기준에 따라 분류하는 에너지 레벨 분류부;

상기 분류된 에너지 레벨을 기초로 상기 프레임들 각각의 에너지 레벨의 변이가 일어나는 구간을 검출하는 변이 검출부; 및

상기 검출된 결과에 기초하여 상기 프레임들 중 상기 음성만을 포함하는 프레임들의 구간에 해당하는 음성 구간을 결정하는 결정부를 포함하며,

상기 결정부는,

상기 프레임 에너지 레벨의 변이가 일어나는 구간마다 소정의 가중치를 부여하고, 상기 구간 전체의 가중치를 계산하는 과정을 반복적으로 수행하여 상기 계산된 전체 가중치들 중 최소가 되는 구간을 상기 신호의 음성 구

간으로 결정하는 것을 특징으로 하는 음성 구간 검출 장치.

청구항 17

제 16 항에 있어서,

상기 프레임들 각각의 에너지를 계산하는 에너지 계산부; 및

상기 소정의 기준을 상기 신호의 에너지에 따라 갱신하는 에너지 레벨 갱신부를 더 포함하고,

상기 에너지 레벨 분류부는 상기 계산된 에너지에 따라 상기 프레임들 각각의 에너지 레벨을 높음, 중간, 또는 낮음을 포함하는 3단계로 분류하는 것을 특징으로 하는 음성 구간 검출 장치.

청구항 18

삭제

청구항 19

제 16 항에 있어서,

상기 결정된 음성 구간을 상기 신호의 전체 음성 구간으로 결합하는 결합부를 더 포함하는 것을 특징으로 하는 음성 구간 검출 장치.

청구항 20

마이크 등을 통해 입력된 음성 및 비음성을 포함한 아날로그 입력신호를 디지털 입력신호로 변환하는 A/D컨버터; 상기 변환된 디지털 입력신호를 입력받아 상기 디지털 입력신호에 상응하는 다수의 프레임을 생성하는 프레임 생성기; 상기 각각의 프레임을 음소 열(sequence)로써 생성하여 출력하는 음소 검출기; 및 기준 음소 열들을 저장하고 있는 사전으로부터 상기 음소 검출기에서 출력된 음소 열에 가장 근접한 음소 열을 추출하여 상기 추출된 음소 열에 상응하는 텍스트를 생성하는 어휘인식기를 포함하는 음성 인식 시스템에 있어서,

상기 프레임 생성기로부터 프레임들을 입력받아 상기 프레임들 각각의 에너지의 변이를 검출하고, 상기 검출된 결과에 기초하여 상기 프레임들 중 상기 음성만을 포함하는 프레임들의 구간에 해당하는 음성 구간을 결정하는 음성 구간 검출 장치; 및

상기 음성 구간 검출 장치로부터 입력된 음성 구간의 정보에 기초하여 상기 음소 검출기로부터 입력된 음소 열을 편집하는 음소 열 편집기를 포함하는 음성 인식 시스템.

청구항 21

제 20 항에 있어서,

상기 음성 구간 검출 장치는,

상기 결정된 음성 구간을 상기 입력신호의 전체 음성 구간으로 결합하여 상기 음소 열 편집기로 출력하는 것을 특징으로 하는 음성 인식 시스템.

청구항 22

제 20 항에 있어서,

상기 음소 열 편집기는,

상기 결정된 음성 구간 정보를 기초로 상기 음성 구간에 상응하는 음소 열을 제외한 음소 열을 제거하는 것을 특징으로 하는 음성 인식 시스템.

명 세 서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

<16> 본 발명은 음성 인식 기술에 관한 것이며, 특히 음성 인식 시스템에서 음성 구간을 검출하는 장치 및 방법에 관한 것이다.

<17> 음성 인식(Speech Recognition) 기술이란 인간의 음성을 컴퓨터가 분석해 이를 인식 또는 이해하는 기술을 말하는데, 발음에 따라 입 모양과 혀의 위치 변화로 특정한 주파수를 갖는 인간의 음성을 이용, 발성된 음성을 전기신호로 변환한 후 음성신호의 주파수 특성을 추출해 발음을 인식하는 기술이다. 최근에는 이와 같은 음성 인식 기술이 전화 다이얼링, 장난감 제어, 어학학습 또는 가전기기 제어 등과 같은 다양한 분야에 응용되고 있다.

<18> 도 1 은 종래의 음소인식 기반의 음성 인식기를 설명하기 위한 개략도이다.

<19> 도 1 을 참조하면, 상기 음성 인식기는 A/D변환기(100), 스펙트럼 분석기(110), 음소 검출기(120) 및 어휘해석기(130)를 포함한다.

<20> 상기 A/D변환기(100)는 마이크 등을 통해 입력된 아날로그 음성입력을 디지털 신호로 변환한다. 변환된 디지털 신호는 상기 스펙트럼 분석기(110)에 입력되어, 디지털 신호의 주파수 스펙트럼 특성을 분석하여 음성 특성(Acoustic Features)만을 추출하여 음소검출기(120)에 제공한다. 음소검출기(120)는 입력된 음성신호를 미리 정의된 음소들의 열(sequence)로 출력하며, 상기 어휘해석기(130)는 음소 열을 입력받아 최종적으로 단어, 혹은 문장을 생성한다.

<21> 하지만, 음성 인식기는 입력된 음성 신호의 주파수 특성을 분석하여 음소검출기(120)에 저장되어 있는 음향 모델과 비교하여 음소를 검출하기 때문에 음성과 함께 입력되는 잡음의 영향을 전혀 고려하고 있지 못했다. 따라서, 이러한 잡음이 음소로 인식되어 음성 인식기의 성능을 저하하는 원인이 되었다.

<22> 또한, 음성 인식기의 성능 향상을 위한 공지기술로 "SPEECH PROCESSING APPARATUS AND METHOD"라는 명칭의 미국공개특허 제2004/0158465호는 입력음성에 포함되는 잡음을 고려하기 위한 잡음 모델 개시하고 있으며, 이는 필터를 사용하여 입력음성의 프레임에서 잡음을 제거하는 노이즈 마스킹(noise masking)기술이다.

<23> 하지만, 노이즈 마스킹 기법을 포함한 종래의 기술들은 시간에 따라 특성이 크게 변하지 않는 고정(stationary)잡음, 예를 들면 자동차에서 발생하는 잡음, 비행기 터빈에서 발생하는 잡음 등과 같이 주파수의 특성이 변하지 않는 잡음에 최적화되어 있기 때문에, 시간 축에서 짧은 시간에 발생하는 돌발 잡음, 예를 들면 음성 입력 앞뒤에서 발생하는 작은 크기의 숨소리, 기계 마찰음, 입술소리 등과 같은 잡음이 존재하는 상황에서는 음성과 비음성을 구분하기가 어려웠다.

<24> 또한, 음소인식기 기반의 음성 인식 기술에 있어서, 전술한 음소 입력 앞뒤에서 발생하는 돌발잡음 등을 포함한 비음성을 잘못 음소로 인식하는 결과가 자주 발생하여 음성 인식기의 성능을 저해하는 요인이었다.

발명이 이루고자 하는 기술적 과제

<25> 본원발명의 목적은 상기 종래기술의 문제점을 해결하기 위하여 안출된 것으로, 동적 프로그래밍(Dynamic Programming)을 이용하여 돌발 잡음 등이 포함된 입력 음성에서 정확한 음성 구간을 검출함으로써 짧은 길이의 돌발 잡음을 음성으로 인식하지 않도록 하는 음성 구간 검출 장치 및 방법을 제공하는 데 있다.

<26> 본원발명의 다른 목적은 상기 음성 구간 검출 장치 및 방법을 이용하여 잡음 등이 포함된 음성이 입력된 경우, 상기 음성 구간의 검출 결과를 기초로 음성을 인식함으로써, 입력된 음성을 정확하게 인식할 수 있는 음성 인식 시스템을 제공하는 데 있다.

발명의 구성 및 작용

<27> 본 발명의 일 실시예에 따른 음성 구간을 검출하는 방법은 음성 및 비음성을 포함하는 신호를 구성하는 프레임들 각각의 에너지의 변이를 검출하는 단계; 및 상기 검출된 결과에 기초하여 상기 프레임들 중 상기 음성만을 포함하는 프레임들의 구간에 해당하는 음성 구간을 결정하는 단계를 포함한다.

<28> 본 발명의 다른 실시예에 따른 음성 및 비음성을 포함하는 신호를 구성하는 프레임들 각각의 에너지를 계산하는 단계; 상기 계산된 에너지에 따라 상기 프레임들 각각의 에너지 레벨을 분류하는 단계; 상기 분류된 에너지 레벨을 기초로 상기 프레임들 각각의 에너지 레벨의 변이가 일어나는 구간을 검출하는 단계; 및 상기 변이가 일어나는 구간마다 소정의 가중치를 부여하고, 상기 구간 전체의 가중치가 최소가 되는 구간을 상기 신호의 음성 구

간으로 결정하는 단계를 포함한다.

<29> 본 발명의 또 다른 실시예에 따른 음성 구간을 검출하는 장치는 음성 및 비음성을 포함하는 신호를 구성하는 프레임들 각각의 에너지의 변이를 검출하는 변이 검출부; 및 상기 검출된 결과에 기초하여 상기 프레임들 중 상기 음성만을 포함하는 프레임들의 구간에 해당하는 음성 구간을 결정하는 결정부를 포함한다.

<30> 본 발명의 또 다른 실시예에 따른 마이크 등을 통해 입력된 음성 및 비음성을 포함한 아날로그 입력신호를 디지털 입력신호로 변환하는 A/D컨버터; 상기 변환된 디지털 입력신호를 입력받아 상기 디지털 입력신호에 상응하는 다수의 프레임을 생성하는 프레임 생성기; 상기 각각의 프레임을 음소 열(sequence)로써 생성하여 출력하는 음소 검출기; 및 기준 음소 열들을 저장하고 있는 사전으로부터 상기 음소 검출기에서 출력된 음소 열에 가장 근접한 음소 열을 추출하여 상기 추출된 음소 열에 상응하는 텍스트를 생성하는 어휘인식기를 포함하는 음성 인식 시스템은 상기 프레임 생성기로부터 프레임들을 입력받아 상기 프레임들 각각의 에너지의 변이를 검출하고, 상기 검출된 결과에 기초하여 상기 프레임들 중 상기 음성만을 포함하는 프레임들의 구간에 해당하는 음성 구간을 결정하는 결정부를 포함하는 음성 구간 검출 장치; 및 상기 음성 구간 검출 장치로부터 입력된 음성 구간의 정보에 기초하여 상기 음소 검출기로부터 입력된 음소 열을 편집하는 음소 열 편집기를 포함한다.

<31> 이하, 첨부한 도면들을 참조하여 본 발명의 바람직한 실시예들을 상세히 설명한다.

<32> 도 2 는 본 발명의 일 실시예에 따른 음성 구간 검출장치가 사용된 음성 인식 시스템을 설명하기 위한 블록도이다.

<33> 도 2 를 참조하면, 음성 인식 시스템은 A/D변환기(200), 프레임 생성기(210), 음성 구간 검출 장치(220), 음소 검출기(230), 음소 열 편집기(240), 어휘 인식기(250) 및 사전(260)을 포함한다.

<34> A/D 변환기(200)는 마이크 등을 통해 입력된 아날로그 음성신호(잡음이 포함된)를 디지털 신호로 변환한다. 프레임 생성기(210)는 입력 신호를 일정 크기의 짧은 구간인 프레임으로 생성한다. 여기서 생성된 프레임은 음성 구간 검출 장치(220)와 음소 검출기(230)에 각각 입력된다. 음성 구간 검출 장치(220)는 입력된 프레임의 에너지를 계산하고, 계산된 에너지에 따라 상기 프레임의 에너지 레벨을 분류하여, 동적 프로그래밍을 이용하여 음성 구간을 결정한다. 음소 검출기(230)는 입력된 프레임으로부터 소정의 음향 모델(Acoustic Model)을 기초로 음성의 최소 단위인 음소를 검출한다. 음성 구간 검출 장치(220)의 출력인 음성 구간에 관한 정보와 음소 검출기(230)의 출력인 음소 열에 관한 정보가 함께 음소 열 편집기(230)에 제공된다.

<35> 음소 열 편집기(240)는 음성 구간 검출기(220)로부터 결정된 음성 구간을 경계 부분에 먼저 걸치는 음소 열을 포함하도록 확장시킨다. 다음으로, 확장된 음성 구간에 포함되지 않는 모든 음소 열을 제거하고, 앞 부분과 뒷부분에 적당길이의 묵음 구간을 추가하고, 인접한 묵음 구간들을 통합한다.

<36> 이러한 과정을 도 5 를 참조하여 설명하면, 확장된 구간은 모든 [d] 구간과 [sil] (묵음) 구간의 일부를 포함한다. 따라서, [sil]은 유지가 된다. 초기 [sil] 구간은 [!ENTER]로 마크된다. 음소 열 편집기(240)를 통해 잡음 등이 원인이 되어 생긴 잘못된 음소 인식 결과를 제거한 후, 정화된 음소 열이 어휘 인식기(250)에 입력된다. 어휘 인식기(250)는 음소 열과 어휘에 관한 정보를 저장하고 있는 데이터베이스인 사전(260)으로부터 음소 열에 상응하는 어휘를 추출하여 분석하고, 음소 열에 상응하는 텍스트를 생성하여 출력 수단(미도시)에 제공한다.

<37> 도 3 은 도 2 에 도시된 음성 구간 검출 장치를 설명하기 위한 상세 블록도이다.

<38> 도 3 을 참조하면, 상기 음성 구간 검출 장치(300)는 에너지 레벨 관리부(310), 정합부(320) 및 결합부(330)를 포함한다.

<39> 에네지 레벨 관리부(310)는 잡음 등을 포함한 음성 신호(이하, 입력신호라 칭함)를 일정한 크기의 프레임으로 만드는 프레임 생성기로부터 다수의 프레임을 입력받는다. 여기서, 프레임은 짧은 크기의 음성 세그먼트를 의미한다. 에너지 레벨 관리부(310)는 다수의 프레임의 에너지를 계산하여 상위(음성) 에너지 레벨 및 하위(잡음) 에너지 레벨의 열(sequence)을 생성한다. 또한, 계산된 에너지에 따라 각각의 프레임의 에너지 레벨을 분류한다.

<40> 에너지 레벨 관리부(310)는 에너지 레벨 계산부(311), 에너지 레벨 분류부(312) 및 에너지 레벨 갱신부(313)를 포함한다. 에너지 레벨 계산부(311)는 각각의 프레임의 에너지를 계산하는데, 여기서 프레임 에너지는 각각의 프레임이 갖는 에너지를 의미하며, 일반적으로 실제 사람의 음성은 높은 에너지 특성을 나타내며, 잡음 등을 포함한 비음성은 낮은 에너지 특성을 나타낸다. 하지만, 낮은 에너지 특성을 갖는 비음성 중에서도 짧은 시간에

에너지가 급격히 높은 부분, 예를 들면 짧은 순간에 주파수 특성이 크게 변하는 돌발잡음은 비교적 높은 에너지 특성을 보인다. 따라서, 본 발명은 이러한 높은 에너지 특성을 보이는 비음성이 실제 사람의 음성으로 인식될 수 있는 가능성을 배제하기 위한 것이다. 이러한 음성과 비음성을 포함한 신호로 구성된 프레임의 에너지를 계산하는 것은 음성 인식 기술에서 일반적으로 사용되는 에너지 계산법에 의해 계산되어 질 수 있다. 에너지 레벨 계산부(311)는 계산된 다수의 프레임의 에너지를 에너지 레벨 분류부(312)에 제공한다. 에너지 레벨 분류부(312)는 각각의 프레임을 계산된 에너지 레벨에 따라 높음(2), 중간(1), 또는 낮음(0)의 3단계로 분류한다. 에너지 레벨 갱신부(313)는 높음, 중간, 또는 낮음을 설정하는 기준을 입력신호의 에너지에 따라 업데이트 한다. 즉, 현재 입력되는 프레임의 에너지 레벨에 관한 정보를 다시 에너지 레벨 분류부(212)에 피드백하여 새로운 기준값을 적용하여 계속 입력되는 프레임들 각각의 에너지를 높음, 중간, 또는 낮음의 상대적 값으로 분류한다. 이는 분류 기준값을 입력 신호의 에너지값을 반영하도록 하여 비교적 높은 에너지값을 가지는 비음성과 더 높은 에너지값을 가지는 음성을 더 정확하게 분류하기 위한 것이다.

<41> 정합부(320)는 에너지 레벨 관리부(310)로부터 프레임들을 입력받아, 분류된 에너지 레벨을 기초로 동적 프로그래밍(Dynamic Programming)을 이용하여 입력신호의 음성 구간을 결정한다. 일반적으로, 동적 프로그래밍은 주어진 문제를 여러 부분 문제로 분할하여 순환 수행함으로써 최종 해답에 접근하는 방법론을 의미한다. 이러한 개념을 본 발명의 바람직한 실시예와 함께 설명하면, 최종적으로 구하고자하는 것은 음성과 비음성을 포함하는 다수의 프레임 에너지 열에서 (비음성이지만 음성이라고 오인식할 수 있는) 짧은 에너지 튐을 무시하고 실제 음성이 시작하는 구간을 검출하는 것이다. 따라서 각각의 프레임과 각각의 프레임의 에너지 레벨을 변수로 하여 이들 변수들을 계산하여 일정하게 분류하는 과정을 통해 부분 문제들, 즉 각각의 프레임 구간이 음성 및 비음성인지를 구하게 된다. 이러한 부분 구간의 음성 및 비음성 여부의 판단을 기초로 최종 음성 구간을 검출하게 된다.

<42> 본 발명의 바람직한 실시예에 따른 동적 프로그래밍의 절차를 설명하면, 현재 경로에서 변경 가능한 모든 경로를 생성한다. 여기서 경로는 입력신호의 각각의 프레임의 에너지 레벨의 추이를 의미한다. 이어, 각각의 변경된 경로에서 프레임 에너지 레벨의 변이에 대해 불이익을 주고, 원래 입력에서 에너지 레벨이 바뀌었을 경우에 불이익을 준다. 이어, 전체 불이익의 합이 최소가 되는 입력을 선택하고 로컬 음성 구간으로 결정한다. 여기서 로컬(local)의 의미는 동적 프로그래밍 적용하여 구해진 일정 범위의 구간을 의미한다. 따라서, 전술한 과정을 통해 구해진 로컬 음성 구간은 전체 입력 신호의 프레임 구간을 일정한 구간들로 나누고, 동적 프로그래밍을 적용하여 음성인지 비음성인지를 판단하게 되는 구간을 의미한다.

<43> 이러한 동적 프로그래밍의 구체적인 알고리즘을 도 6 을 참조하여 설명한다.

<44> 먼저, 가로축 상의 숫자(1-14)는 입력 신호의 프레임 넘버를 나타내고, 세로축은 입력 신호의 에너지 레벨을 0(낮음), 1(중간), 또는 2(높음)로 분류한 것을 나타낸다. 실선 화살표는 음성/비음성 변이에 의한 불이익을 나타내고, 점선 화살표는 원래의 경로를 바꾸어 주는 것에 의한 불이익을 나타낸다. 큰 원은 에너지 레벨 측정에 의한 최초 에너지 레벨, 즉 입력을 나타내고, 작은 원은 동적 프로그래밍으로 검출된 불이익이 최소로 되는 검출된 음성/비음성 구간 경로, 즉 출력을 나타낸다.

<45> 도면부호 600의 구간, 즉 프레임 넘버 2-6까지의 에너지 레벨 변이를 보면, 최초 입력은 2-3번 프레임에서 0-1, 3-4번 프레임에서 1-2, 4-5번 프레임에서 2-1, 5-6번 프레임에서 1-0으로 에너지 변이가 발생한다. 5개의 프레임에서 총 4번의 불이익을 주고 있으므로, 이는 동적 프로그래밍에 의해 3,4,5번 프레임의 에너지 레벨을 0으로 만듦으로써 불이익을 3번으로 줄이게 된다.

<46> 마찬가지로, 도면부호 610의 구간, 즉 프레임 넘버 10-14까지의 에너지 레벨 변이를 보면, 최초 입력은 10-11번 프레임에서 0-2로 에너지 변이가 발생하고, 11-14번 프레임에서는 2로 에너지 변이가 발생하지 않았다. 결국, 5개의 프레임에서 총 1번의 불이익을 주고 있으므로, 이는 그대로 두는 것이 비용함수(cost function) 측면에서 유리하다. 여기서 비용함수는 프로그래밍 측면에서 시간 복잡도, 즉 여러 가지 경로를 통해 해를 구하는데 들이는 시간의 정도를 의미한다. 본 발명의 바람직한 실시예에서 비용함수를 결정하는 인자는 프레임 수×에너지 레벨(0,1,2)이다. 이는 동적 알고리즘을 수행함에 있어 유리함을 주는데, 왜냐하면 에너지 레벨을 (0,1)로 할 경우 비용함수를 줄일 수는 있지만, 음성 및 비음성을 검출할 수 있는 정밀도가 떨어지고, 에너지 레벨을 (0,1,2,3..)로 더욱 세분화한다면 정밀도는 높아지지만 그만큼 비용함수가 커진다. 즉 효율이 떨어지게 된다. 따라서, 본 발명의 바람직한 실시예에서 선택한 높음(2), 중간(1), 또는 낮음(0)을 포함하는 3단계 분류는 본 발명을 실시하는데 최적의 실시예이다. 하지만, 3단계 분류는 본 발명의 바람직한 실시예일뿐, 3단계 이상의 분류 또는 그 이하의 분류도 본 발명의 범위에 포함되는 것임은 당업자에게 자명한 사항이다.

<47> 도면부호 600 및 610의 결과는 전체 불이익의 합이 최소가 되는 경로를 선택한다는 동적 프로그래밍 알고리즘에 의한 것이다. 이상 바람직한 실시예에 따른 동적 프로그래밍 알고리즘의 특징을 요약하면, 가능한 경로들이 있고 각각의 경로들의 비용함수(cost function)가 있을 때 이를 최소화하는 경로를 찾아 최적의 솔루션을 구하는 것이다.

<48> 다시, 도 3 을 참조하여 상기 정합부(320)는 음성/비음성 변이, 즉 각각의 프레임에서 에너지 레벨의 변이가 일어나는 부분에 적당한 불이익 값(Penalty)을 준다. 여기서 불이익 값은 동적 프로그래밍 알고리즘을 수행하기 위해 부여되는 일종의 가중치를 의미한다. 이러한 과정을 일정한 범위의 구간마다 프레임 에너지 레벨의 변경하여 반복적으로 수행한다. 이어, 일정한 프레임 구간의 전체 불이익 값의 합이 최소가 되는 음성 및 비음성 변이를 구한다. 따라서, 음성 및 비음성 구간의 개수가 최소가 되는 변이를 구하게 된다.

<49> 정합부(320)는 생성부(321), 변이 검출부(322) 및 결정부(323)를 포함한다. 생성부(321)는 동적 프로그래밍 알고리즘을 시작하기 위해서, 현재의 입력신호의 각각의 프레임의 에너지 레벨에서 변경 가능한, 즉 각각의 프레임 에너지 레벨을 0부터 2까지의 에너지 레벨 중 3가지 경우의 수로 변경하여 생성한다. 변이 검출부(322)는 현재 입력신호의 각각의 프레임에서 인접한 프레임 간의 에너지 레벨 변이, 예를 들면 1번 프레임에서 2번 프레임 간의 에너지값이 변화된 구간을 검출한다. 또한, 최초의 입력신호의 프레임의 에너지 레벨에서 생성부(321)로부터 변경된 에너지 레벨이 변화된 구간을 검출한다. 결정부(323)는 변이 검출부(322)를 통해 검출된 프레임 구간에 대하여 불이익 값을 준다. 여기서 불이익 값이 부여되는 구간은 인접한 프레임 간의 에너지 레벨 변이 구간과 변경 전후의 에너지 레벨의 차이가 있는 구간이다. 입력신호의 전체 구간에 대해 상술한 과정을 반복적으로 수행하여 불이익 값을 계산하고, 이어, 전체 불이익 값이 최소가 되는 구간을 음성 구간으로 결정한다. 여기서, 음성 구간은 음성 및 비음성을 포함하는 신호를 구비하는 프레임들 중에서 음성만을 포함하는 프레임 구간을 의미한다.

<50> 결합부(330)는 정합부(320)에서 결정된 음성 구간을 전체 입력 신호의 음성 구간으로 결합한다.

<51> 도 4 은 본 발명의 또 다른 실시예에 따른 음성 구간 검출 방법을 설명하기 위한 흐름도이다.

<52> 도 4 를 참조하면, 음성 구간 검출장치는 입력 신호를 일정한 크기의 프레임으로 만드는 프레임 생성기로부터 프레임들을 제공받는다. 여기서 프레임은 작은 크기의 입력 신호 열로써 이루어진다. 400 단계에서, 도 3에 도시된 에너지 레벨 갱신부(313)는 에너지 레벨 음성 구간을 나타내는 상부 레벨 트랙과 잡음 구간을 나타내는 하부 레벨 트랙, 즉 음성과 잡음을 결정하는 구분값들을 현재 프레임의 에너지 레벨로 업데이트 하는데, 이는 높음, 중간, 또는 낮음으로 구분되는 구분값들을 미리 정하지 않고 입력되는 신호의 에너지값에 따라 정한다. 402 단계에서, 에너지 레벨 분류부(312)는 현재 프레임의 에너지 레벨 값을 업데이트된 상하부 레벨 트랙에 기초하여 에너지 레벨을 분류한다. 404 단계에서, 상기 정합부(320)는 동적 프로그래밍을 이용하여 분류된 에너지 레벨을 기초로 로컬(local) 음성 구간들을 결정한다. 여기서 동적 프로그래밍은 각각의 프레임의 에너지 레벨을 기초로 음성/비음성 변이가 일어나는 구간에 대하여 불이익 값을 주고, 전체 불이익 값이 최소가 되도록 음성/비음성 구간을 결정하는 과정으로 이루어진다. 406 단계에서, 결합부(330)는 단계 404에서 검출된 로컬 음성 구간들을 입력 신호의 전체 음성 구간으로 결합한다. 이어 전체 음성 구간의 위치 정보를 음소 열 편집기에 전달한다.

<53> 도 5 는 본 발명의 또 다른 실시예에 따른 음성 구간 검출 방법을 이용한 음성 구간 검출 결과를 설명하기 위한 도면이다.

<54> 본 실시예에서 입력음성은 "Joining you by Alanis Morisette" 이며, "<곡명> by <가수>"의 형태로 입력된다.

<55> 도 5 를 참조하면, 도면의 상부는 입력신호의 측정된 에너지(C0)이며, 상위 에너지 변화(음성에너지)와 하위 에너지 변화(배경잡음)로 표시된다. 도면의 가운데 부분은 입력신호의 주파수 변화를 도시하고 있다. 도면의 맨 아래 두 줄은 각각 음소 검출기의 음소 열 출력과 제안된 발명으로 정화된 음소 열이다. 에너지는 각각 high, medium, low (2, 1, 0)로 마크되었으며 짧은 길이의 마크 1과 2는 본 실시예에 따른 음성 구간 검출 장치에 의해 무시되었다. 결과적으로 음성입력은 1.18초, 즉 [d] 구간에서 시작된다.

<56> 이상 본 발명의 바람직한 실시예들을 기초로 설명되었지만, 당업자들은 본 발명이 속하는 기술분야의 기술적 사상이나 필수적 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적인 것이 아닌 것으로서 이해되어야 한다. 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 한정되며, 특허청구범위의 의미 및 범위 그리고 그 등가 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는

것으로 해석되어야 한다.

발명의 효과

<57> 본 발명에 따른 음성 구간 검출 방법 및 장치를 사용함으로써, 주변에 크기가 작고 짧은 길이의 돌발 잡음이 자주 발생하는 환경, 또는 음성 입력 앞뒤에서 발생하는 작은 크기의 숨소리, 기계 마찰음, 입술 소리 등의 의사 음소열(spurious phone-sequence)을 실제 사용자의 음성으로 인식하지 않음으로써, 효과적으로 음성 구간을 검출할 수 있다.

<58> 또한, 본 발명에 따른 음성 구간 검출 방법 및 장치는 입력신호의 에너지만을 분석함으로써 간단하게 구현할 수 있는 장점이 있다.

<59> 또한, 본 발명에 따른 음성 인식 시스템을 사용함으로써, 음소인식 결과에 크게 의지하는 음소 인식 기반의 음성 인식기에서 빈번히 나타날 수 있는 음소 열을 잘못 인식하는 문제를 해결하는 것이 가능하다.

도면의 간단한 설명

<1> 도 1 은 종래의 음소인식 기반의 음성 인식기를 설명하기 위한 블록도이다.

<2> 도 2 는 본 발명의 일 실시예에 따른 음성 구간 검출장치가 사용된 음성 인식 시스템을 설명하기 위한 블록도이다.

<3> 도 3 은 도 2 에 도시된 음성 구간 검출 장치를 위한 상세 블록도이다.

<4> 도 4 는 본 발명의 다른 실시예에 따른 음성 구간 검출 방법을 설명하기 위한 흐름도이다.

<5> 도 5 는 본 발명의 또 다른 실시예에 따른 음성 구간 검출 방법을 이용한 음성 구간 검출 결과를 설명하기 위한 도면이다.

<6> 도 6 은 본 발명의 또 다른 실시예에 따른 동적 프로그래밍(Dynamic Programming) 알고리즘을 설명하기 위한 도면이다.

<7> <도면의 주요 부분에 대한 부호의 설명>

<8> 210: 프레임 생성기 220: 음성 구간 검출 장치

<9> 230: 음소 검출기 240: 음소 열 편집기

<10> 250: 어휘 인식기 300: 음성 구간 검출장치

<11> 310: 에너지 레벨 관리부 311: 에너지 레벨 계산부

<12> 312: 에너지 레벨 분류부 313: 에너지 레벨 갱신부

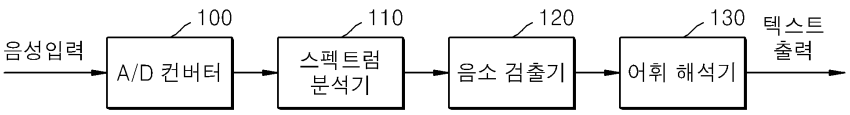
<13> 320: 정합부 321: 생성부

<14> 322: 변이 검출부 323: 결정부

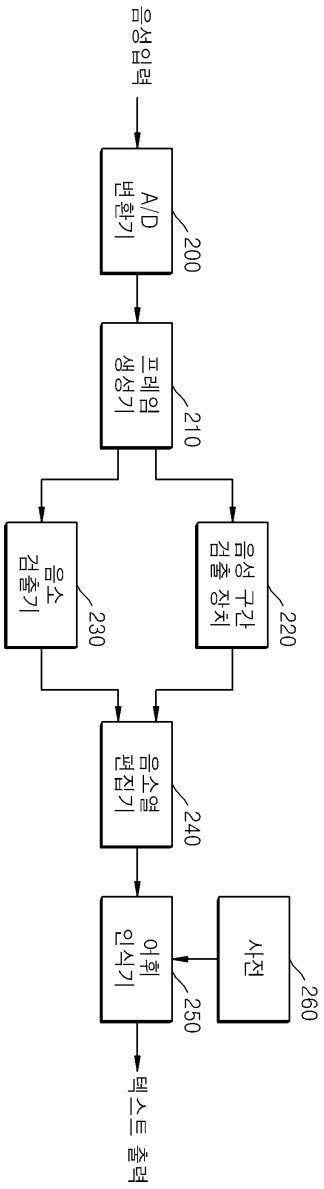
<15> 330: 결합부

도면

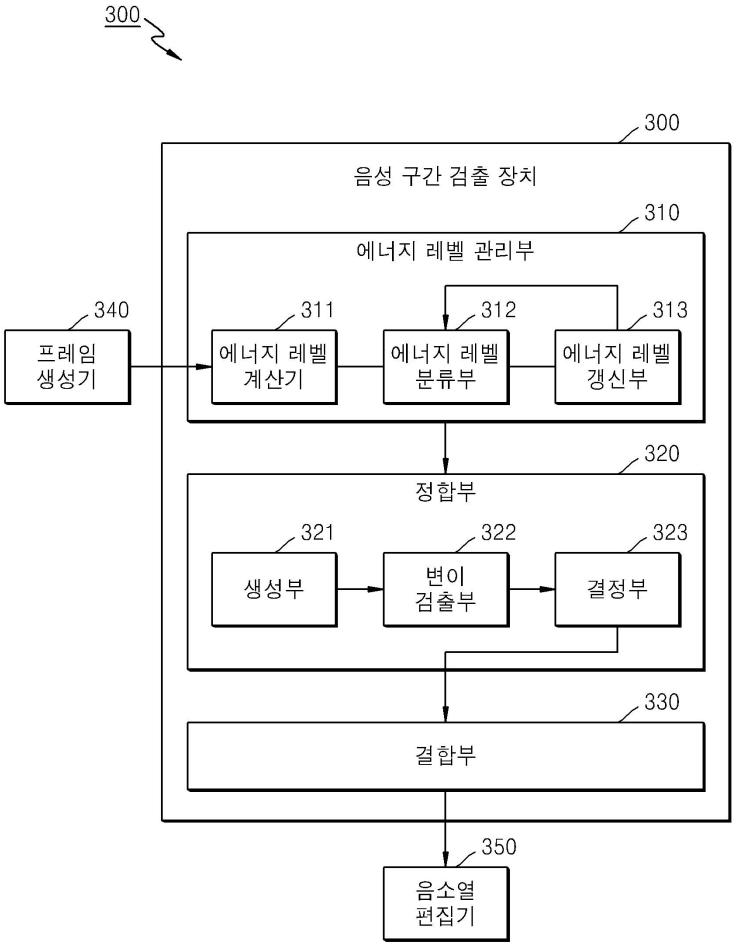
도면1



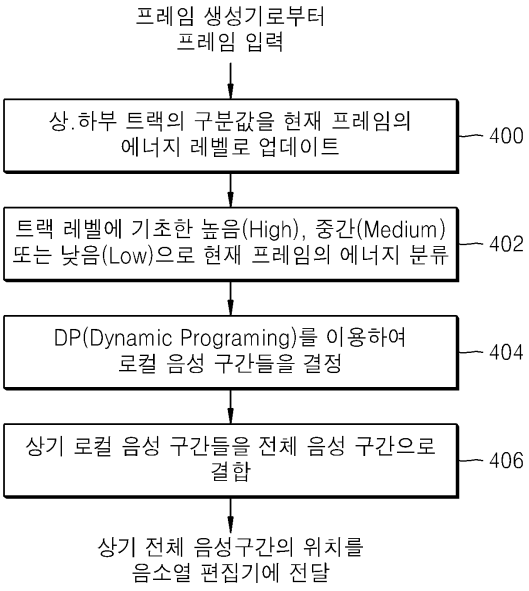
도면2


음성입력
A/D 변환기
200
프레임 생성기
210
음소 검출기
230
음성 구간 검출 장치
220
음소열 편집기
240
어휘 인식기
250
사전
260
텍스트 출력

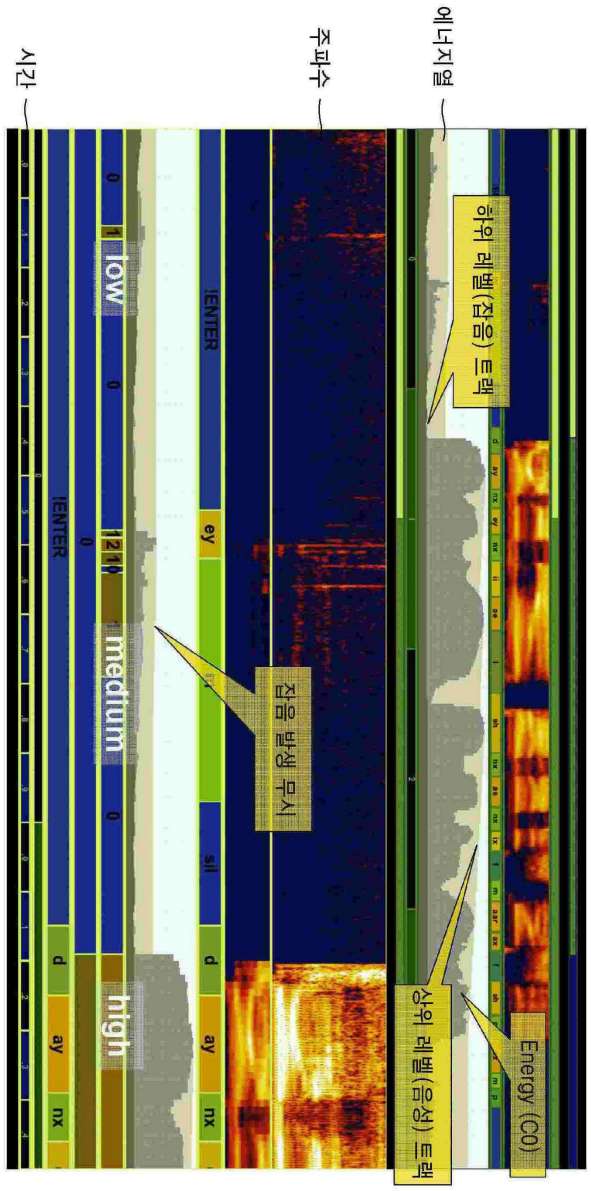
도면3


300
300
음성 구간 검출 장치
310
에너지 레벨 관리부
340
프레임
생성기
311
에너지 레벨
계산기
312
에너지 레벨
분류부
313
에너지 레벨
갱신부
320
정합부
321
생성부
322
변이
검출부
323
결정부
330
결합부
350
음소열
편집기

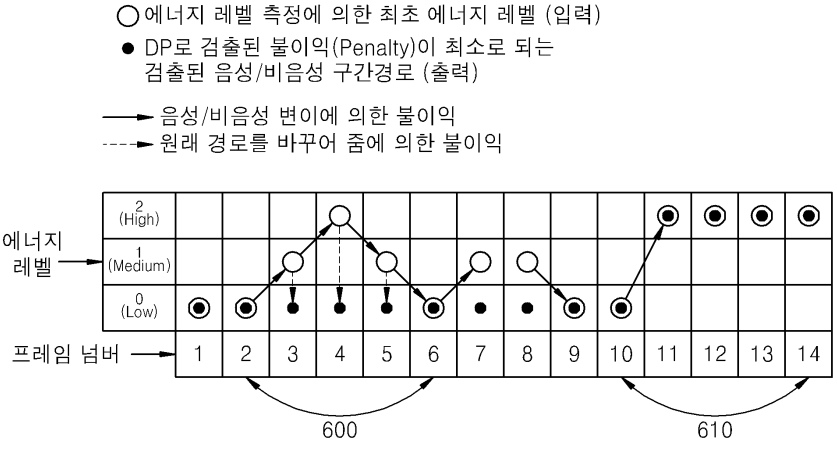
도면4


프레임 생성기로부터
프레임 입력
상.하부 트랙의 구분값을 현재 프레임의
에너지 레벨로 업데이트
400
트랙 레벨에 기초한 높음(High), 중간(Medium)
또는 낮음(Low)으로 현재 프레임의 에너지 분류
402
DP(Dynamic Programing)를 이용하여
로컬 음성 구간들을 결정
404
상기 로컬 음성 구간들을 전체 음성 구간으로
결합
406
상기 전체 음성구간의 위치를
음소열 편집기에 전달

도면5


시간
주파수
에너지 레벨
하위 레벨(잡음) 트랙
상위 레벨(음성) 트랙
잡음 발생 구간
Energy (C0)
!ENTER
low
medium
high
ey
sil
d
ay
nx

도면6


○ 에너지 레벨 측정에 의한 최초 에너지 레벨 (입력)
● DP로 검출된 불이익(Penalty)이 최소로 되는 검출된 음성/비음성 구간경로 (출력)
음성/비음성 변이에 의한 불이익
원래 경로를 바꾸어 줌에 의한 불이익
에너지 레벨
2 (High)
1 (Medium)
0 (Low)
프레임 넘버
1 2 3 4 5 6 7 8 9 10 11 12 13 14
600
610