(19)대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) 。Int. Cl.
G10L 15/00 (2006.01)

(45) 공고일자 2007년03월23일
(11) 등록번호 10-0698811
(24) 등록일자 2007년03월16일

(21) 출원번호	10-2001-7009889	(65) 공개번호	10-2001-0093327
(22) 출원일자	2001년08월04일	(43) 공개일자	2001년10월27일
심사청구일자	2005년02월03일		
번역문 제출일자	2001년08월04일		
(86) 국제출원번호	PCT/US2000/002903	(87) 국제공개번호	WO 2000/46791
국제출원일자	2000년02월04일	국제공개일자	2000년08월10일

(81) 지정국 국내특허 : 아랍에미리트, 알바니아, 아르메니아, 오스트리아, 오스트레일리아, 아제르바이잔, 보스니아 헤르체고비나, 바베이도스, 불가리아, 브라질, 벨라루스, 캐나다, 스위스, 중국, 코스타리카, 쿠바, 체코, 독일, 덴마크, 도미니카, 에스토니아, 스페인, 핀란드, 영국, 그라나다, 그루지야, 가나, 감비아, 크로아티아, 헝가리, 인도네시아, 이스라엘, 인도, 아이슬랜드, 일본, 케냐, 키르키즈스탐, 북한, 대한민국, 카자흐스탄, 세인트루시아, 스리랑카, 리베이라, 레소토, 리투아니아, 룩셈부르크, 라트비아, 모로코, 몰도바, 마다가스카르, 마케도니아공화국, 몽고, 말라위, 멕시코, 노르웨이, 뉴질랜드, 폴란드, 포르투칼, 루마니아, 러시아, 수단, 스웨덴, 싱가포르, 슬로베니아, 슬로바키아, 시에라리온, 타지키스탄, 투르크맨, 터어키, 트리니아드토바고, 탄자니아, 우크라이나, 우간다, 우즈베키스탄, 베트남, 세르비아 앤 몬테네그로, 남아프리카, 짐바브웨,

AP ARIPO특허 : 가나, 감비아, 케냐, 레소토, 말라위, 수단, 시에라리온, 스와질랜드, 탄자니아, 우간다, 짐바브웨,

EA 유라시아특허 : 아르메니아, 아제르바이잔, 벨라루스, 키르키즈스탐, 카자흐스탄, 몰도바, 러시아, 타지키스탄, 투르크맨,

EP 유럽특허 : 오스트리아, 벨기에, 스위스, 사이프러스, 독일, 덴마크, 스페인, 핀란드, 프랑스, 영국, 그리스, 아일랜드, 이탈리아, 룩셈부르크, 모나코, 네덜란드, 포르투칼, 스웨덴,

OA OAPI특허 : 부르키나파소, 베닌, 중앙아프리카, 콩고, 코트디브와르, 카메룬, 가봉, 기니, 기니 비사우, 말리, 모리타니, 니제르, 세네갈, 차드, 토고,

(30) 우선권주장 09/248,513 1999년02월08일 미국(US)

(73) 특허권자
퀄컴 인코포레이티드
미국 92121-1714 캘리포니아주 샌 디에고 모어하우스 드라이브 5775

(72) 발명자
비닝
미국92128캘리포니아주샌디에고브리즈웨이플레이스14209

창치엔청
미국92131캘리포니아주샌디에고사이프러스테라스플레이스11456

가루대드리하리나쓰
미국92129캘리포니아주샌디에고오비에도스트리트9435

데자코앤드류피
미국92126캘리포니아주샌디에고플란더스코브10424

(74) 대리인 특허법인코리아나

심사관 : 경연정

전체 청구항 수 : 총 40 항

(54) 음성 인식 거부 방식

(57) 요약

발음을 캡쳐하는 음성 인식 제거 구조는 상기 발음을 채택하는 단계, 상기 발음에 N-베스트 알고리즘을 적용하는 단계, 또는 상기 발음을 제거하는 단계를 포함한다. 상기 발음과 하나 이상의 다른 저장된 단어들간의 하나 이상의 비교 결과치들과 하나 이상의 가장 가까운 비교 결과치들간의 하나 이상의 차이들과 저장된 단어에 대해서 상기 발음에 대한 하나 이상의 가장 가까운 비교 결과치들간에 제 1 소정 관계가 존재하면, 상기 발음이 인정된다. 상기 하나 이상의 가장 가까운 비교 결과치들과 상기 하나 이상의 다른 비교 결과치들간의 상기 하나 이상의 차이들과 상기 하나 이상의 가장 가까운 비교 결과치들간에 제 2 소정 관계가 존재하면, 상기 발음에 N-베스트 알고리즘이 적용된다. 상기 하나 이상의 가장 가까운 비교 결과치들과 상기 하나 이상의 다른 비교 결과치들간의 상기 하나 이상의 차이들과 상기 하나 이상의 가장 가까운 비교 결과치들간에 제 3 소정 관계가 존재하면, 상기 발음이 제거된다. 하나 이상의 다른 비교 결과치들중 하나는, 상기 발음과 또 다른 저장 단어에 관한 다음으로 가장 가까운 비교 결과치가 되는 것이 바람직하다. 제 1, 제 2, 및 제 3 소정 관계는 선형 관계인 것이 바람직하다.

대표도

도 2

특허청구의 범위

청구항 1.
삭제

청구항 2.
삭제

청구항 3.
삭제

청구항 4.
삭제

청구항 5.
삭제

청구항 6.
삭제

청구항 7.
삭제

청구항 8.
삭제

청구항 9.
삭제

청구항 10.
삭제

청구항 11.
삭제

청구항 12.
삭제

청구항 13.
삭제

청구항 14.
삭제

청구항 15.
삭제

청구항 16.
삭제

청구항 17.
삭제

청구항 18.
삭제

청구항 19.
삭제

청구항 20.
삭제

청구항 21.
삭제

청구항 22.
삭제

청구항 23.
삭제

청구항 24.
삭제

청구항 25.
삭제

청구항 26.
삭제

청구항 27.
삭제

청구항 28.
삭제

청구항 29.
삭제

청구항 30.
삭제

청구항 31.
삭제

청구항 32.
삭제

청구항 33.
삭제

청구항 34.
삭제

청구항 35.
삭제

청구항 36.
삭제

청구항 37.
삭제

청구항 38.
삭제

청구항 39.
삭제

청구항 40.
삭제

청구항 41.
삭제

청구항 42.
삭제

청구항 43.
삭제

청구항 44.
삭제

청구항 45.

음성 인식 시스템에서 발음을 캡쳐하는 방법으로서,

제 1 저장 단어에 상기 발음을 비교하여 제 1 스코어를 생성하는 단계;

제 2 저장 단어에 상기 발음을 비교하여 제 2 스코어를 생성하는 단계;

상기 제 1 스코어와 상기 제 2 스코어 사이의 차이를 결정하는 단계;

상기 차이에 대한 상기 제 1 스코어의 비율을 결정하는 단계; 및

상기 비율에 기초하여 상기 발음을 인식-프로세싱하는 단계를 포함하는, 발음 캡쳐 방법.

청구항 46.

제 45 항에 있어서,

상기 발음 인식-프로세싱 단계는,

상기 차이에 대한 상기 제 1 스코어의 비율이 제 1 범위값 이내이면, 상기 발음을 채택하는 단계;

상기 차이에 대한 상기 제 1 스코어의 비율이 제 2 범위값 이내이면, 상기 발음을 조회하기 위해 N-베스트 알고리즘을 적용하는 단계; 및

상기 차이에 대한 상기 제 1 스코어의 비율이 제 3 범위값 이내이면, 상기 발음을 거부하는 단계를 포함하는, 발음 캡쳐 방법.

청구항 47.

제 45 항에 있어서,

상기 차이는 상기 제 1 스코어와 상기 제 2 스코어 사이의 스코어의 변화에 대응하는, 발음 캡쳐 방법.

청구항 48.

제 45 항에 있어서,

상기 제 1 저장 단어는 음성 인식 시스템의 어휘중에서 가장 양호한 후보를 포함하고, 상기 제 2 저장 단어는 음성 인식 시스템의 어휘중에서 그 다음으로 양호한 후보를 포함하는, 발음 캡쳐 방법.

청구항 49.

제 45 항에 있어서,

상기 제 1 스코어는 가장 근접한 비교 결과를 포함하고, 상기 제 2 스코어는 그 다음으로 가장 근접한 비교 결과를 포함하는, 발음 캡쳐 방법.

청구항 50.

제 45 항에 있어서,

상기 제 1 스코어 및 상기 제 2 스코어는 선형 예측 코딩 계수를 포함하는, 발음 캡쳐 방법.

청구항 51.

제 45 항에 있어서,

상기 제 1 스코어 및 상기 제 2 스코어는 켑스트럼 계수를 포함하는, 발음 캡쳐 방법.

청구항 52.

제 45 항에 있어서,

상기 제 1 스코어 및 상기 제 2 스코어는 대역통과 필터 출력을 포함하는, 발음 캡쳐 방법.

청구항 53.

제 46 항에 있어서,

상기 제 1, 제 2, 및 제 3 범위값은 상기 제 1 스코어와 상기 차이 사이의 선형 관계를 정의하는, 발음 캡쳐 방법.

청구항 54.

제 45 항에 있어서,

상기 차이는 가장 근접한 비교 결과와 그 다음으로 가장 근접한 비교 결과 사이의 차이를 포함하는, 발음 캡쳐 방법.

청구항 55.

발음의 디지털화된 스피치 샘플로부터 스피치 파라미터를 추출하도록 구성된 음향 프로세서; 및

상기 음향 프로세서에 결합되고, 제 1 저장 단어에 상기 발음을 비교하여 제 1 스코어를 생성하고,

제 2 저장 단어에 상기 발음을 비교하여 제 2 스코어를 생성하고,

상기 제 1 스코어와 상기 제 2 스코어 사이의 차이를 결정하고,

상기 차이에 대한 상기 제 1 스코어의 비율을 결정하며, 또한

상기 관계에 기초하여 상기 발음을 인식-프로세스하도록 구성된 프로세서를 구비하는, 음성 인식 시스템.

청구항 56.

제 55 항에 있어서,

상기 프로세서는,

상기 차이에 대한 상기 제 1 스코어의 비율이 제 1 범위값 이내이면 상기 발음을 채택하고;

상기 차이에 대한 상기 제 1 스코어의 비율이 제 2 범위값 이내이면 상기 발음을 조회하기 위해 N-베스트 알고리즘을 적용하고; 또한

상기 차이에 대한 상기 제 1 스코어의 비율이 제 3 범위값 이내이면 상기 발음을 제거하도록 더 구성된, 음성 인식 시스템.

청구항 57.

제 55 항에 있어서,

상기 차이는 상기 제 1 스코어와 상기 제 2 스코어 사이의 변화에 대응하는, 음성 인식 시스템.

청구항 58.

제 55 항에 있어서,

상기 제 1 저장 단어는 상기 음성 인식 시스템의 어휘중에서 가장 양호한 후보를 포함하고, 상기 제 2 저장 단어는 상기 음성 인식 시스템의 어휘중에서 그 다음으로 가장 양호한 후보를 포함하는, 음성 인식 시스템.

청구항 59.

제 55 항에 있어서,

상기 제 1 스코어는 가장 근접한 비교 결과를 포함하고, 상기 제 2 스코어는 그 다음으로 가장 근접한 비교 결과를 포함하는, 음성 인식 시스템.

청구항 60.

제 55 항에 있어서,

상기 제 1 스코어 및 제 2 스코어는 선형 예측 코딩 계수를 포함하는, 음성 인식 시스템.

청구항 61.

제 55 항에 있어서,

상기 제 1 스코어 및 상기 제 2 스코어는 켑스트럼 계수를 포함하는, 음성 인식 시스템.

청구항 62.

제 55 항에 있어서,

상기 제 1 스코어 및 상기 제 2 스코어는 대역통과 필터 출력을 포함하는, 음성 인식 시스템.

청구항 63.

제 56 항에 있어서,

상기 제 1, 제 2, 및 제 3 범위값은 상기 제 1 스코어와 상기 차이 사이의 선형 관계를 정의하는, 음성 인식 시스템.

청구항 64.

제 55 항에 있어서,

상기 차이는 가장 근접한 비교 결과와 그 다음으로 가장 근접한 비교 결과 사이의 차이를 포함하는, 음성 인식 시스템.

청구항 65.

발음을 제 1 저장 단어에 비교하여 제 1 스코어를 생성하는 수단;

상기 발음을 제 2 저장 단어에 비교하여 제 2 스코어를 생성하는 수단;

상기 제 1 스코어와 상기 제 2 스코어 사이의 차이를 결정하는 수단;

상기 차이에 대한 상기 제 1 스코어의 비율을 결정하는 수단; 및

상기 비율에 기초하여 상기 발음을 인식-프로세싱하는 수단을 구비하는, 음성 인식 시스템.

청구항 66.

제 65 항에 있어서,

상기 차이에 대한 상기 제 1 스코어의 비율이 제 1 범위값 이내이면 상기 발음을 채택하는 수단;

상기 차이에 대한 상기 제 1 스코어의 비율이 제 2 범위값 이내이면 상기 발음을 조회하기 위해 N-베스트 알고리즘을 적용하는 수단; 및

상기 차이에 대한 상기 제 1 스코어의 비율이 제 3 범위값 이내이면 상기 발음을 거부하는 수단을 더 구비하는, 음성 인식 시스템.

청구항 67.

제 66 항에 있어서,

상기 제 1, 제 2, 및 제 3 범위값은 상기 제 1 스코어와 상기 차이 사이의 선형 관계를 정의하는, 음성 인식 시스템.

청구항 68.

제 65 항에 있어서,

상기 차이는 상기 제 1 스코어와 상기 제 2 스코어 사이의 스코어의 변화에 대응하는, 음성 인식 시스템.

청구항 69.

제 65 항에 있어서,

상기 제 1 저장 단어는 상기 음성 인식 시스템의 어휘중에서 가장 양호한 후보를 포함하고, 상기 제 2 저장 단어는 상기 음성 인식 시스템의 어휘중에서 그 다음으로 가장 양호한 후보를 포함하는, 음성 인식 시스템.

청구항 70.

제 65 항에 있어서,

상기 제 1 스코어는 가장 근접한 비교 결과를 포함하고, 상기 제 2 스코어는 그 다음으로 가장 근접한 비교 결과를 포함하는, 음성 인식 시스템.

청구항 71.

제 65 항에 있어서,

상기 제 1 스코어 및 상기 제 2 스코어는 선형 예측 코딩 계수를 포함하는, 음성 인식 시스템.

청구항 72.

제 65 항에 있어서,

상기 제 1 스코어 및 상기 제 2 스코어는 켑스트럼 계수를 포함하는, 음성 인식 시스템.

청구항 73.

제 65 항에 있어서,

상기 제 1 스코어 및 상기 제 2 스코어는 대역통과 필터 출력을 포함하는, 음성 인식 시스템.

청구항 74.

제 65 항에 있어서,

상기 차이는 가장 근접한 비교 결과와 그 다음으로 가장 근접한 비교 결과 사이의 차이를 포함하는, 음성 인식 시스템.

청구항 75.

발음의 디지털화된 스피치 샘플로부터 스피치 파라미터를 추출하는 수단; 및

상기 발음을 제 1 저장 단어에 비교하여 제 1 스코어를 생성하고,

상기 발음을 제 2 저장 단어에 비교하여 제 2 스코어를 생성하고,

상기 제 1 스코어와 상기 제 2 스코어 사이의 차이를 결정하고,

상기 차이에 대한 상기 제 1 스코어의 비율을 결정하고, 또한

상기 비율에 기초하여 상기 발음을 인식-프로세싱하는 수단을 구비하는, 음성 인식 시스템.

청구항 76.

제 75 항에 있어서,

상기 차이에 대한 상기 제 1 스코어의 비율이 제 1 범위값 이내이면 상기 발음을 채택하고,

상기 차이에 대한 상기 제 1 스코어의 비율이 제 2 범위값 이내이면 상기 발음을 조회하기 위해 N-베스트 알고리즘을 적용하고, 또한

상기 차이에 대한 상기 제 1 스코어의 비율이 제 3 범위값 이내이면 상기 발음을 거부하는 수단을 더 구비하는, 음성 인식 시스템.

청구항 77.

제 75 항에 있어서,

상기 차이는 가장 근접한 비교 결과와 그 다음으로 가장 근접한 비교 결과 사이의 차이를 포함하는, 음성 인식 시스템.

청구항 78.

제 75 항에 있어서,

상기 차이는 상기 제 1 스코어와 상기 제 2 스코어 사이의 스코어의 변화에 대응하는, 음성 인식 시스템.

청구항 79.

제 75 항에 있어서,

상기 제 1 저장 단어는 상기 음성 인식 시스템의 어휘중에서 가장 양호한 후보를 포함하고, 상기 제 2 저장 단어는 상기 음성 인식 시스템의 어휘중에서 그 다음으로 가장 양호한 후보를 포함하는, 음성 인식 시스템.

청구항 80.

제 75 항에 있어서,

상기 제 1 스코어는 가장 근접한 비교 결과를 포함하고, 상기 제 2 스코어는 그 다음으로 가장 근접한 비교 결과를 포함하는, 음성 인식 시스템.

청구항 81.

제 75 항에 있어서,

상기 제 1 스코어 및 상기 제 2 스코어는 선형 예측 코딩 계수를 포함하는, 음성 인식 시스템.

청구항 82.

제 75 항에 있어서,

상기 제 1 스코어 및 상기 제 2 스코어는 켑스트럼 계수를 포함하는, 음성 인식 시스템.

청구항 83.

제 75 항에 있어서,

상기 제 1 스코어 및 제 2 스코어는 대역통과 필터 출력을 포함하는, 음성 인식 시스템.

청구항 84.

제 76 항에 있어서,

상기 제 1, 제 2, 및 제 3 범위값은 상기 제 1 스코어와 상기 차이 사이의 선형 관계를 정의하는, 음성 인식 시스템.

명세서

발명의 배경

Ⅰ. 발명의 분야

본 발명은, 일반적으로 통신 분야에 관한 것이고, 특히 음성 인식 시스템에 관한 것이다.

Ⅱ. 배경

음성 인식 (voice recognition, VR) 은 기기에 시뮬레이션된 지능을 부여하여 유저 또는 유저-음성 명령을 인식하고, 기기와 휴먼 인터페이스를 용이하게 하는 가장 중요한 기술들중 하나를 나타낸다. 또한, VR 은 휴먼 스피치를 이해하기 위한 핵심 기술을 나타낸다. 음향 스피치 신호로부터 언어적인 메시지를 재생시키는 기술을 채용하는 시스템들을 음성 인식기라 한다. 통상, 음성 인식기는 인커밍 로우 스피치 (incoming raw speech) 의 VR 을 달성하는데 필요한, 정보전달 특성, 또는 벡터의 일 시퀀스를 추출하는 음향 프로세서, 및 그 특성, 또는 벡터의 시퀀스를 디코드하여 입력된 발음 (utterence) 에 대응하는 언어적 단어의 시퀀스 등의 의미있고 원하는 출력 포맷을 하는 단어 디코더를 포함한다. 소정 시스템의 성능을 증대하기 위하여, 그 시스템이 유효한 파라미터들을 구비하기 위한 트레이닝이 필요하다. 즉, 최적으로 기능시키기 전에 시스템을 학습시킬 필요가 있다.

음향 프로세서는 음성 인식기내의 프론트 엔드 (front-end) 스피치 분석 서브 시스템을 나타낸다. 입력 스피치 신호에 응답하여, 음향 프로세서는 시변 스피치 신호의 특징을 나타내는 적절한 표시를 제공한다. 음향 프로세서는 배경잡음, 채널 왜곡, 화자특성, 및 화법과 같은 부적절한 정보를 제거해야 한다. 효율적인 음향 프로세싱은 음성 인식기에 증대된 음향 식별력을 제공한다. 이 목적을 위해, 분석하기에 유용한 특성은 단기 스펙트럼 엔벨로프 (short time spectral envelope) 이다. 단시간 스펙트럼 엔벨로프의 특징을 나타내기 위한 두 가지의 보편적으로 사용되는 스펙트럼 분석 기술로는, 선형 예측 코딩 (LPC) 및 필터뱅크 기반의 스펙트럼 모델링 (filter-bank-based spectral modeling) 이 있다. 예시적인 LPC 기술들이 본 발명의 양수인에게 양도되고 참고로 여기에 포함된 미국특허 제 5,414,796 호, 및 역시 참고로 여기에 포함된 L.B. Rabiner & R.W. Schafer, *Digital Processing of Speech signals* 396-453(1978) 에 개시되어 있다.

VR (또는 통상 스피치 인식이라고도 함) 의 사용이 안전상의 이유로 점점 더 중요해지고 있다. 예를 들어, VR 은 무선전화 키패드상의 버튼을 누르는 수동적인 임무를 대체하기 위하여 사용될 수 있다. 이는 특히, 유저가 운전을 하면서 전화호출을 개시하는 경우에 중요하다. VR 없이 전화를 사용하는 경우, 운전자는 운전대로부터 한 손을 들어야 하고, 호출을 다이얼하기 위해 버튼을 누르는 동안 전화기 키패드를 주시해야 한다. 이 행동들은 자동차 사고의 가능성을 높인다. 스피치 인에이블 전화 (즉, 스피치 인식용으로 설계된 전화) 는 운전자로 하여금 지속적으로 도로를 주시하면서 전화 호출을 할 수 있도록 한다. 그리고 부가적으로, 핸즈프리 자동차-키트 시스템(hands-free car-kit system)은 운전자로 하여금 호출 개시동안 운전대상에 양 손을 유지할 수 있도록 한다.

스피치 인식 장치들은 화자종속 (speaker-dependent) 또는 화자독립 (speaker-independent) 장치들로 분류된다. 화자독립 장치들은 어떠한 유저들로부터도 음성 명령들을 받을 수 있다. 더욱 보편적인 화자종속 장치들은 특정의 유저들로부터의 명령을 인식하도록 트레이닝된다. 화자종속 VR 장치는 트레이닝 단계및 인식 단계의 두 단계로 동작하는 것이 보통이다. 트레이닝 단계에서, VR 시스템은 유저로 하여금 시스템의 어휘 (vocabulary) 에 있는 각 단어들을 한 번 또는 두 번 말하도록 하여, 그 특정의 단어나 어구에 대한 유저 스피치의 특성을 시스템이 학습하도록 한다. 다른 방법으로는, 음성 (phonic) VR 장치의 경우, 언어의 모든 음소를 커버하도록 특별히 스크립트된 하나 이상의 간단한 기사를 읽음으로써 트레이닝이 성취된다. 핸즈프리 자동차 키트에 대한 예시적인 어휘는 키패드상의 숫자들; 키워드 "call", "send", "dial", "cancel", "clear", "add", "delete", "history", "program", "yes", 및 "no" 및 소정 수의 통상적으로 호출되는 동료, 친구, 또는 가족 구성원들의 이름들을 포함할 수 있다. 일단 트레이닝이 완료되면, 인식 단계에서, 유저가 트레이닝된 키워드를 말하여 호출을 개시할 수 있다. 예를 들면, "존"이라는 이름이 트레이닝된 이름들중 하나라면, 유저는 "존 호출"이라는 어구를 말함으로써 존에게로의 호출을 개시할 수 있을 것이다. VR 시스템은, 단어 "호출" 및 "존"을 인식할 것이고, 유저가 이전에 존의 전화 번호로서 입력했던 번호를 다이얼할 것이다.

VR 시스템의 처리 능력은 유저가 인식 작업을 성공적으로 행한 경우의 백분율로 정의될 수 있다. 인식 작업은 통상 다단계로 이루어진다. 예를 들면, 무선 전화기로 음성 다이얼링할 때, 처리 능력은 유저가 VR 시스템으로 성공적으로 전화 통화를 완료하는 경우의 평균 백분율을 가리킨다. VR로 성공적인 전화 통화를 이루기 위해 필요한 단계들의 수는 한 통화 이상으로 가변적이 된다. 일반적으로, VR 시스템의 처리 능력은 주로 2가지 요소 즉, VR 시스템의 인식 정확도와 인간-기계 인터페이스에 의존한다. VR 시스템 성능의 유저 주관적인 식별은 처리 능력에 기반한다. 그러므로, 높은 인식 정확도를 가진 VR 시스템과 처리 능력을 증가시키기 위한 인텔리전트 인간-기계 인터페이스에 대한 요구가 있게 되었다.

발명의 개요

본 발명은, 높은 인식 정확도와 처리 능력을 증가시키기 위한 인텔리전트 인간-기계 인터페이스를 가진 VR 시스템을 지향한다. 따라서, 본 발명의 일 태양에서, 음성 인식 시스템에서 발음을 캡쳐하는 방법은, 발음과 하나 이상의 다른 저장된 단어 사이의 하나 이상의 다른 비교 결과와 하나 이상의 비교 결과 사이의 하나 이상의 차이와 저장된 단어에 대해서 발음에 관한 하나 이상의 비교 결과 사이에 제 1 소정 관계가 존재하면 발음을 채택하는 단계; 하나 이상의 비교 결과와 하나 이상

의 다른 비교 결과 사이에 하나 이상의 차이와 하나 이상의 비교 결과 사이에 제 2 소정 관계가 존재하면 발음에 N-베스트 알고리즘을 인가하는 단계; 하나 이상의 비교 결과와 하나 이상의 다른 비교 결과 사이에 하나 이상의 차이와 하나 이상의 비교 결과 사이에 제 3 소정 관계가 존재하면 발음을 거부하는 단계를 포함하는 것이 바람직하다.

본 발명의 또 다른 측면에서, 음성 인식 시스템은 발음의 디지탈화된 스피치 샘플들로부터 스피치 파라미터들을 추출하도록 구성된 음향 프로세서와, 이 음향 프로세서에 결합되고 (1) 발음과 하나 이상의 다른 저장된 단어 사이의 하나 이상의 다른 비교 결과와 하나 이상의 비교 결과 사이의 하나 이상의 차이와 저장된 단어에 대한 발음에 있어서의 하나 이상의 비교 결과 사이에 제 1 소정 관계가 존재하면 발음을 채택하고, (2) 하나 이상의 비교 결과와 하나 이상의 다른 비교 결과 사이의 하나 이상의 차이와 하나 이상의 비교 결과 사이에 제 2 소정 관계가 존재하면 N-베스트 알고리즘을 발음에 인가하고, (3) 하나 이상의 비교 결과와 하나 이상의 다른 비교 결과 사이의 하나 이상의 차이와 하나 이상의 비교 결과 사이에 제 3 의 소정 관계가 존재하면 발음을 거부하도록 구성된 프로세서를 포함한다.

본 발명의 또 다른 측면에서, 음성 인식 시스템은 발음과 하나 이상의 다른 저장된 단어 사이의 하나 이상의 다른 비교 결과와 하나 이상의 비교 결과 사이의 하나 이상의 차이와 저장된 단어에 대한 발음에 있어서의 하나 이상의 비교 결과 사이에 제 1 소정 관계가 존재하면 발음을 채택하는 수단, 하나 이상의 비교 결과와 하나 이상의 다른 비교 결과 사이의 하나 이상의 차이와 하나 이상의 비교 결과 사이에 제 2 소정 관계가 존재하면 N-베스트 알고리즘을 발음에 인가하는 수단, 하나 이상의 비교 결과와 하나 이상의 다른 비교 결과 사이의 하나 이상의 차이와 하나 이상의 비교 결과 사이에 제 3 의 소정 관계가 존재하면 발음을 거부하는 수단을 포함하는 것이 바람직하다.

본 발명의 또 다른 측면에서, 음성 인식 시스템은 발음의 디지탈화된 스피치 샘플들로부터 스피치 파라미터들을 추출하는 수단과, (1) 발음과 하나 이상의 다른 저장된 단어 사이의 하나 이상의 다른 비교 결과와 하나 이상의 비교 결과 사이의 하나 이상의 차이와 저장된 단어에 대한 발음에 있어서의 하나 이상의 비교 결과 사이에 제 1 소정 관계가 존재하면 발음을 채택하고, (2) 하나 이상의 비교 결과와 하나 이상의 다른 비교 결과 사이의 하나 이상의 차이와 하나 이상의 비교 결과 사이에 제 2 소정 관계가 존재하면 N-베스트 알고리즘을 발음에 인가하고, (3) 하나 이상의 비교 결과와 하나 이상의 다른 비교 결과 사이의 하나 이상의 차이와 하나 이상의 비교 결과 사이에 제 3 의 소정 관계가 존재하면 발음을 거부하는 수단을 포함하는 것이 바람직하다.

도면의 간단한 설명

도 1 은 음성 인식 시스템의 블록도이다.

도 2 는 거부, N-베스트, 및 채택 영역을 예시하는 VR 시스템의 거부 방식에 있어서의 스코어 대 스코어의 변화 그래프이다.

바람직한 실시예의 상세한 설명

도 1 에 예시된 바와 같은 일 실시예에 의하면, 음성 인식 시스템 (10) 은 아날로그/디지털 변환기(A/D)(12), 음향 프로세서 (14), VR 템플릿 데이터 베이스 (VR template database; 16), 패턴 비교 로직 (18), 및 결정 로직 (decision; 20) 을 포함한다. VR 시스템 (10) 은 예를 들면, 무선전화 또는 핸즈프리 자동차 키트에 설치할 수 있다.

VR 시스템 (10) 이 음성 인식 단계에 있는 경우, 사람 (미도시) 은 한 단어 또는 어구를 말하여 스피치 신호를 생성한다. 스피치 신호는 종래의 트랜스듀서 (역시 미도시) 에 의해 전기적 스피치 신호 s(t) 로 변환된다. 스피치 신호 s(t) 는 A/D (12) 로 제공되고, 여기서는 예를 들면 펄스 코드 변조 (pulse coded modulation, PCM) 와 같은 공지된 샘플링법에 따라 스피치 신호 s(t) 를 디지털 스피치 샘플 s(n) 로 변환한다.

스피치 샘플 s(n) 은 파라미터 결정을 위해 음향 프로세서 (14) 로 제공된다. 음향 프로세서 (14) 는 입력 스피치 신호 s(t) 의 특성을 모델링한 한 세트의 추출된 파라미터들을 생성한다. 이 파라미터들은, 예를 들면 스피치 코더 인코딩을 포함하고, 이전에 언급한 미국특허 제 5,414,796 호에 개시된 바와 같이, 고속 푸리에 변환 (FFT) 기반의 켑스트럼 계수들을 사용하는 어떠한 다수의 공지된 스피치 파라미터 결정 기술들에 따라서 결정할 수 있다. 음향 프로세서 (14) 는 디지털 신호 프로세서 (DSP) 로서 구현할 수도 있다. DSP는 스피치 코더를 포함할 수 있다. 다른 방법으로, 음향 프로세서 (14) 를 스피치 코더로 구현할 수 있다.

VR 시스템 (10) 의 트레이닝 동안, 파라미터 결정을 수행하며, 이때 VR 시스템 (10) 의 모든 어휘 단어에 관한 한 세트의 템플릿이 영구 저장을 위하여 VR 템플릿 데이터베이스 (16) 로 라우팅된다. VR 템플릿 데이터베이스 (16) 는 예를 들면, 플래시 메모리와 같은 임의의 종래의 비휘발성 저장매체의 형태로 구현하는 것이 바람직하다. 이것은 VR 시스템 (10) 으로의 파워가 턴오프되는 경우 템플릿을 VR 템플릿 데이터베이스 (16) 에 유지할 수 있도록 한다.

파라미터의 세트는 패턴비교 로직 (18) 으로 제공된다. 패턴비교 로직 (18) 은 발음의 시작점 및 종료점을 검출하고, 동적 음향 특성(예를 들면, 시간 도함수(time derivatives), 2차 시간 도함수(second time derivatives) 등)을 계산하고, 적절한 프레임들을 선택하여 음향 특성들을 압축하며, 정적 및 동적 음향 특성들을 양자화하는 것이 바람직하다. 종료점 검출, 동적 음향 특성 미분(dynamic acoustic feature derivation), 패턴 압축, 및 패턴 양자화의 공지된 방법은 예를 들면, 참고로 여기서 포함된, Lawrence Rabiner & Biing-Hwang Juang, *Fundamentals of Speech Recognition (1993)*에 개시되어 있다.

패턴비교 로직 (18) 은 파라미터 세트를 VR 템플릿 데이터베이스 (16) 에 저장된 모든 템플릿들과 비교한다. 파라미터 세트와 VR 템플릿 데이터베이스 (16) 에 저장된 모든 템플릿 사이의 거리들, 또는 비교결과들을 결정 로직 (20) 에 제공한다. 결정 로직 (20) 은 (1) VR 템플릿 데이터베이스 (16) 로부터 그 파라미터 세트와 가장 근접하게 매칭하는 템플릿을 선택하거나, (2) 소정의 매칭 스페시홀드 내에서 N개의 가장 근접한 매칭들을 선택하는 "N-베스트" 선택 알고리즘을 적용하거나; 또는 (3) 파라미터 세트를 거부할 수 있다. N-베스트 알고리즘이 사용되면, 그 후 사람은 어느 것을 선택할 지에 대하여 질문을 받는다. 결정 로직 (20) 의 출력은 어휘에서 어느 단어가 말해졌는지에 대한 결정이다. 예를 들면, N-베스트 상황에서 유저가 "존 앤더스"라고 말하면, VR 시스템 (10) 은 "존 앤드류라고 말했습니까?"라고 대답하고 그 다음 유저는 "존 앤더스"라고 답할 것이다. 그러면, VR 시스템 (10) 은 "존 앤더스라고 말했습니까?"라고 말한다. 그 다음 유저가 "예"라고 대답하면 그 시점에서, VR 시스템 (10) 이 전화 통화의 다이얼링을 개시하게 된다.

패턴 비교 로직 (18) 과 결정 로직 (20) 은 마이크로프로세서로서 구현하는 것이 바람직할 수 있다. 또한, 패턴 비교 로직 (18) 과 결정 로직 (20) 은 프로세서, 제어기 또는 상태 기계와 같은 종래의 어떤 형태로도 구현될 수 있다. VR 시스템 (10) 은 예를 들면, 주문형 집적 회로(ASIC)일 수 있다. VR 시스템 (10) 의 인식 정확도는 VR 시스템 (10) 이 어휘에서 말해진 단어 또는 어구들을 얼마나 정확히 인식하는지에 대한 기준 (measure) 이다. 예를 들면, 95%의 인식 정확도는, VR 시스템 (10) 이 어휘에서 단어를 100회 중 95회 정확히 인식하는 것을 나타낸다.

일 실시예에서는 도 2 에 예시된 바와 같이, 스코어 대 스코어 변화 그래프의 채택, N-베스트, 및 거부 영역으로 분할된다. 이러한 영역들은 참고로 여기에 포함된, Richard O. Duda & Peter E. Hart의 *Pattern Classification and Scene Analysis (1973)*에 개시된 공지된 선형 구별 분석 기술에 따라 라인들에 의해 분리된다. VR 시스템 (10) 으로의 각 발음 입력이, 상술한 바와 같이 패턴 비교 로직 (18) 에 의해 VR 템플릿 데이터베이스 (16) 에 저장된 모든 템플릿에 관한 비교 결과나 템플릿으로부터의 거리에 할당된다. 이들 거리들 또는 "스코어들"은 다중 프레임들에 걸쳐 합쳐진 N-차원 벡터 공간에서의 벡터들간의 유클리드의 거리가 되는 것이 바람직하다. 벡터 공간이 24차원 벡터 공간인 일 실시예에서, 스코어는 20개의 프레임들에 걸쳐 누적되고, 이 스코어는 정수인 거리이다. 당업자는 이러한 스코어가 분수 또는 다른 값으로서도 동등하게 표현될 수 있음을 알 수 있을 것이다. 또, 당업자는 스코어들이 예를 들면 확률 측정, 가능성 측정 등과 같이, 다른 거리들도 유클리드 거리로 대용될 수 있음을 알 수 있을 것이다.

발음이 주어지고, VR 템플릿 데이터 베이스 (16) 로부터의 VR 템플릿이 주어지면, 스코어가 낮을수록 (즉, 발음과 VR 템플릿간의 거리가 더 작을수록), 발음과 VR 템플릿간의 매칭이 더 가까워진다. 각 발음에 있어서, 결정 로직 (20) 은 스코어와 VR 템플릿 데이터 베이스 (16) 에서 두번째로 가장 근접한 매칭 (즉, 두번째로 가장 낮은 스코어) 과 연관된 스코어간의 차이에 관련해서 VR 템플릿 데이터베이스 (16) 에서의 가장 근접한 매칭과 연관된 스코어를 분석한다. 도 2 의 그래프에 표시된 바와 같이, "스코어"는 "스코어 변화"에 대해서 그려져 있고 3개의 영역들이 정의된다. 거부 영역은 스코어가 상대적으로 높고, 스코어와 그 다음으로 가장 낮은 스코어간의 차이가 상대적으로 작은 영역을 나타낸다. 발음이 제거 영역 내에 있으면, 결정 로직 (20) 은 발음을 거부한다. 채택 영역은 스코어가 상대적으로 낮고, 스코어와 그 다음으로 가장 낮은 스코어간의 차이가 상대적으로 큰 영역을 나타낸다. 발음이 채택 영역 내에 있다면, 결정 로직 (20) 은 발음을 채택한다. N-베스트 영역은 거부 영역과 채택 영역 사이에 있다. N-베스트 영역은 스코어가 거부 영역에서의 스코어보다 작거나, 스코어와 그 다음으로 가장 낮은 스코어간의 차이가 거부 영역에서의 스코어 차이보다 큰 영역을 나타낸다. 또한, N-베스트 영역은 N-베스트 영역에서의 스코어 차이가 스코어 값 변화의 소정의 한계치보다 크다는 가정하에, 스코어가 채택 영역에서의 스코어보다 크거나 스코어와 그 다음으로 가장 낮은 스코어간의 차이가 채택 영역에서의 스코어 차이보다 작은 영역을 나타낸다. N-베스트 영역 내에 발음이 있게 되면, 결정 로직 (20) 은 상술한 바와 같이, 발음에 N-베스트 알고리즘을 적용한다.

도 2 를 참조하여 설명한 실시예에서, 제 1 라인 세그먼트는 N-베스트 영역으로부터 거부 영역을 분리시킨다. 제 1 라인 세그먼트는 소정의 한계 스코어값에서 "스코어"축과 교차한다. 또, 제 1 라인 세그먼트의 기울기도 미리 정해진다. 제 2 라인 세그먼트는 채택 영역으로부터 N-베스트 영역을 분리시킨다. 제 2 라인 세그먼트의 기울기는, 제 1 라인 세그먼트의 기울기와 동일하도록 미리 정해져서, 제 1 및 제 2 라인 세그먼트들은 평행하다. 제 3 라인 세그먼트는 "스코어의 변화" 축상에서의 소정의 한계 변화값으로부터 수직으로 연장하여 제 2 라인 세그먼트의 종료점과 만난다. 당업자들은 제 1 및 제 2 라인 세그먼트들은 평행할 필요가 없고, 어떠한 임의 할당 기울기도 가질 수 있음을 알 수 있을 것이다. 또한, 제 3 라인 세그먼트는 사용될 필요가 없다.

한계 스코어값이 375인 실시예에서, 한계 변화값이 28이고, 제 2 라인 세그먼트의 종료점이 연장되었다면, 제 2 라인 세그먼트는 값이 250인 곳에서 "스코어"축과 교차하여 제 1 및 제 2 라인 세그먼트의 기울기는 각각 1 이다. 만약, 스코어값이 스코어값의 변화 + 375 보다 크다면, 발음이 거부된다. 그렇지 않고, 스코어값이 스코어값의 변화 + 250 보다 크거나 스코어 값의 변화가 28보다 작다면, N-베스트 알고리즘이 발음에 적용된다. 그렇지 않으면, 발음은 채택된다.

도 2 를 참조하여 설명한 실시예에서, 선형 구별 분석용으로 2차원이 사용된다. 차원 "스코어"는 다중 대역통과 필터(미도시)의 출력들로부터 유도된 것처럼, 주어진 발음과 주어진 VR 템플릿 사이의 거리를 나타낸다. 차원 "스코어의 변화"는 가장 낮은 스코어 즉, 가장 근접하게 매칭된 스코어와 그 다음으로 가장 낮은 스코어 즉, 그 다음으로 가장 근접하게 매칭된 발음에 대한 스코어간의 차이를 나타낸다. 또 다른 실시예에서, 차원 "스코어"는 발음의 켑스트럼 계수로부터 유도되는 것처럼 주어진 발음과 주어진 VR 템플릿 사이의 거리를 나타낸다. 또다른 실시예에서, 차원 "스코어"는 발음의 선형 예측 코딩 (LPC) 계수로부터 유도되는 바와 같이, 주어진 발음과 주어진 VR 템플릿 사이의 거리를 나타낸다. 발음의 LPC 계수와 켑스트럼 계수를 유도하는 기술이 상술한 미국 특허 5,414,796호에 기재되어 있다.

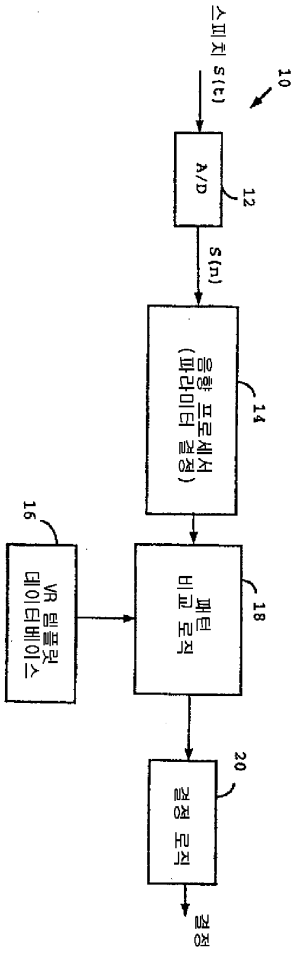
다른 실시예에서는, 선형 구별 분석이 2차원으로 한정되지 않는다. 따라서, 제 1 스코어 기반 대역통과 필터 출력들, 제 2 스코어 기반 켑스트럼 계수들, 및 스코어의 변화가 서로 연관되어 분석된다. 또는, 제 1 스코어 기반 대역통과 필터 출력들, 제 2 스코어 기반 켑스트럼 계수들, 제 3 스코어 기반 LPC 계수들, 및 스코어의 변화가 서로 연관되어 분석된다. 당업자라면 "스코어" 에 관한 차원의 갯수가 어떤 특정한 수로 한정될 필요가 없다는 것을 즉시 알 수 있을 것이다. 당업자는 스코어 차원의 갯수가 VR 시스템의 어휘에서의 단어수에 의해서만 한정되는 것을 알 수 있을 것이다. 또한, 당업자는 사용된 스코어의 타입들이 스코어의 어떤 특정 타입에 한정될 필요는 없지만, 관련 분야에 공지된 어떠한 스코어링 방법도 포함할 수 있음을 알 수 있을 것이다. 또, "스코어의 변화"에 관한 차원의 갯수가 1 또는 어떤 특정 숫자에 한정될 필요가 없다는 것도 당업자에 의해 즉시 알 수 있는 것이다. 예를 들면, 일 실시예에서 가장 근접한 매칭과 다음으로 가장 근접한 매칭간의 스코어 변화와 연관하여 스코어가 분석되고, 또한 이러한 스코어는 가장 근접한 매칭과 세번째로 근접한 매칭간의 스코어 변화와 연관되어 분석된다. 당업자는 스코어 변화 차원의 갯수가 VR 시스템의 어휘에서의 단어수에만 한정된다는 것을 알 수 있을 것이다.

이상과 같이, 선형 구별 분석에 기초한 신규하고도 향상된 음성 인식 거부 기술이 설명되었다. 당업자는, 여기서 개시된 실시예들을 통하여 설명된 다양한 로직 블록들 및 알고리즘 단계들의 일예는 디지털 신호 프로세서(DSP), 주문형 특정 집적 회로(ASIC), 이산 게이트 또는 트랜지스터 로직, 예를 들어 레지스터 및 FIFO 같은 이산 하드웨어 구성요소들, 한 세트의 펌웨어 명령들을 실행하는 프로세서, 또는 임의의 종래의 프로그램 가능한 소프트웨어 모듈 및 프로세서로 구현 또는 수행될 수 있음을 이해할 것이다. 프로세서는 마이크로프로세서인 것이 바람직하지만, 다른 방법으로는, 프로세서가 어떤 종래의 프로세서, 제어기, 마이크로 제어기, 또는 상태기기일 수 있다. 소프트웨어 모듈은 RAM 메모리, 플래시 메모리, 레지스터 또는 임의의 다른 형태의 공지된 기록 가능한 저장매체에 내장할 수 있다. 또한, 당업자는, 상술한 설명에 걸쳐 참조될 수 있는 데이터, 지시, 명령, 정보, 신호, 비트, 심볼 및 칩은, 전압, 전류, 전자기파, 자계 필드 또는 입자, 광학적 필드 또는 입자, 또는 그들의 임의의 조합으로 표시되는 것이 바람직하다는 것 또한 이해할 것이다.

이상, 본 발명의 바람직한 실시예들이 도시되고 설명하였다. 그러나, 본 발명의 정신이나 범주에서 일탈함이 없이 여기에서 설명된 실시예들에 대한 수많은 변형들이 가능하다는 것이 당업자에게는 명백할 것이다. 따라서, 본 발명은 이하의 청구범위에 의해서만 한정된다.

도면

도면1


스피치 s(t)
10
A/D
12
s(n)
음향 프로세서
(파라미터 결정)
14
16
VR 템플릿
데이터베이스
패턴
비교 로직
18
결정 로직
20
결정

도면2