(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0107677
(43) 공개일자 2010년10월06일

(51) Int. Cl.
G10L 19/06 (2006.01) *G10L 19/02* (2006.01)
G10L 15/14 (2006.01)
(21) 출원번호 10-2009-0025891
(22) 출원일자 2009년03월26일
심사청구일자 2009년03월26일

(71) 출원인
고려대학교 산학협력단
서울 성북구 안암동5가 1
(72) 발명자
육동석
서울 용산구 서빙고동 신동아아파트 1동 102호
김동현
서울특별시 성북구 돈암동 639 돈암힐스테이트 APT 106-501
이협우
경기도 안양시 동안구 갈산동 샘마을대우아파트 101-1002
(74) 대리인
특허법인충현

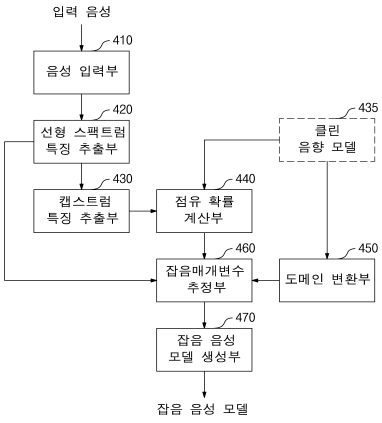
전체 청구항 수 : 총 13 항

(54) 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 방법 및 그 장치, 잡음 음성 모델을 이용한 음성 인식 방법및 그 장치

(57) 요 약

본 발명은 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 방법에 관한 것으로서 입력 음성에서 선형 스펙트럼 데이터 및 캡스트럼 데이터를 추출하는 단계; 클린 음향 모델과 상기 캡스트럼 데이터를 이용하여 가우시안 점유 확률을 추정하는 단계; 상기 가우시안 점유 확률, 상기 선형 스펙트럼 데이터 및 상기 클린 음향 모델을 선형 스펙트럼 도메인으로 변환한 음향 모델을 이용하여 잡음 매개 변수를 추정하는 단계; 및 상기 잡음 매개 변수를 이용하여 잡음 음성 모델을 생성하는 단계를 포함하는 것을 특징으로 하며, 잡음 매개변수를 추정하는 과정에서 재귀 연산을 하지 않고 닫힌 연산 추정법을 적용하여 계산 비용을 줄일 수 있고, 음향 모델 적응이나 인식 과정의 실시간성을 향상시킬 수 있다.

대 표 도 - 도4



특허청구의 범위

청구항 1

입력 음성에서 선형 스펙트럼 데이터 및 캡스트럼 데이터를 추출하는 단계;

클린 음향 모델과 상기 캡스트럼 데이터를 이용하여 가우시안 점유 확률을 추정하는 단계;

상기 가우시안 점유 확률, 상기 선형 스펙트럼 데이터 및 상기 클린 음향 모델을 선형 스펙트럼 도메인으로 변환한 음향 모델을 이용하여 잡음 매개 변수를 추정하는 단계; 및

상기 잡음 매개 변수를 이용하여 잡음 음성 모델을 생성하는 단계를 포함하는, 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 방법.

청구항 2

제 1 항에 있어서,

상기 잡음 매개 변수를 추정하는 단계는,

닫힌 연산 보조 함수를 이용하는 단계인 것을 특징으로 하는, 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 방법.

청구항 3

제 2 항에 있어서,

상기 닫힌 연산 보조 함수는,

클린 음향 모델이 λ, 잡음 음향 모델을 λ'라 하고, ϕ가 전이 확률, γ_g^t는 t 시간에 g 가우시안의 사후 확률일 때, 선형 스펙트럼 입력 데이터가 $\mathbf{o}^s$, 평균벡터가 $\boldsymbol{\mu}^s$, 분산 행렬이 $\boldsymbol{\Sigma}^s$인 경우에, 다음의 수식 (1)과 같이 표현되는 변형된 바움(Baum)의 보조 함수인 것을 특징으로 하는, 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 방법.

$$F(\lambda,\lambda') = \phi - \frac{1}{2}P(\mathbf{O}|\lambda)\sum_t\sum_g \gamma_g^t \cdot[M\log(2\pi) + \log|\boldsymbol{\Sigma}_g^s| + (\mathbf{o}^{s,t} - \boldsymbol{\mu}_g^s)^T \boldsymbol{\Sigma}_g^{s-1}(\mathbf{o}^{s,t} - \boldsymbol{\mu}_g^s)], \quad (1)$$

청구항 4

제 2 항에 있어서,

상기 닫힌 연산 보조 함수는,

클린 음향 모델이 λ, 잡음 음향 모델을 λ'라 하고, ϕ가 전이 확률, γ_g^t는 t 시간에 g 가우시안의 사후 확률일 때, 선형 스펙트럼 입력 데이터가 $\mathbf{o}^s$, 평균벡터가 $\boldsymbol{\mu}^s$, 분산 행렬이 $\boldsymbol{\Sigma}^s$이며, A가 채널 왜곡과 관계된 잡음 매개 변수이고, b가 가산 잡음과 관계된 잡음 매개 변수인 경우에, 다음의 수식 (2)와 같이 표현되는 변형된 바움(Baum)의 보조 함수인 것을 특징으로 하는, 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 방법.

$$F(\lambda,\lambda') = -\sum_t\sum_g \gamma_g^t[\log(|\boldsymbol{\Sigma}_g^s|) - 2\log(|\mathbf{A}^{-1}|) + (\mathbf{A}^{-1}\mathbf{o}^{s,t} - \mathbf{A}^{-1}\mathbf{b} - \boldsymbol{\mu}_g^s)^T \boldsymbol{\Sigma}_g^{s-1}(\mathbf{A}^{-1}\mathbf{o}^{s,t} - \mathbf{A}^{-1}\mathbf{b} - \boldsymbol{\mu}_g^s)]. \quad (2)$$

청구항 5

제 1 항에 있어서,

상기 잡음 음성 모델을 생성하는 단계는,

상기 잡음 음성 모델을 캡스트럼 도메인으로 변환하는 단계를 포함하는 것을 특징으로 하는, 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 방법.

청구항 6

발화 음성으로부터 특징 벡터를 추출하는 단계; 및

상기 특징 벡터를 잡음 음성 모델과 비교하여 최대 우도를 갖는 인식 단어를 생성하는 단계를 포함하고,

상기 잡음 음성 모델은,

클린 음향 모델과 입력 음성의 캡스트럼 데이터를 이용하여 가우시안 점유 확률을 추정하며, 상기 가우시안 점유 확률, 상기 입력 음성의 선형 스펙트럼 데이터 및 상기 클린 음향 모델을 선형 스펙트럼 도메인으로 변환한 음향 모델을 이용하여 잡음 매개 변수를 추정하는 과정을 통해 생성된 음성 모델인 것을 특징으로 하는, 잡음 음성 모델을 이용한 음성 인식 방법.

청구항 7

제1항 내지 제6항 중 어느 한 항의 방법을 컴퓨터 시스템에서 실행하기 위한 프로그램이 기록된, 컴퓨터 시스템이 판독할 수 있는 기록매체.

청구항 8

입력 음성에서 특징 벡터에 대한 선형 스펙트럼 데이터를 추출하는 선형 스펙트럼 특징 추출부;

상기 선형 스펙트럼 데이터를 캡스트럼 데이터로 변환하는 캡스트럼 특징 추출부;

클린 음향 모델과 상기 캡스트럼 데이터를 이용하여 가우시안 점유 확률을 추정하는 점유 확률 계산부;

상기 가우시안 점유 확률, 상기 선형 스펙트럼 데이터 및 상기 클린 음향 모델을 선형 스펙트럼 도메인으로 변환한 음향 모델을 이용하여 잡음 매개 변수를 추정하는 잡음 매개 변수 추정부; 및

상기 잡음 매개 변수를 이용하여 잡음 음성 모델을 생성하는 잡음 음성 모델 생성부를 포함하는, 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 장치.

청구항 9

제 8 항에 있어서,

상기 잡음 매개 변수 추정부는,

닫힌 연산 보조 함수를 이용하는 것을 특징으로 하는, 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 장치.

청구항 10

제 9 항에 있어서,

상기 닫힌 연산 보조 함수는,

클린 음향 모델이 λ, 잡음 음향 모델을 λ'라 하고, ϕ가 전이 확률, γ_g^t는 t 시간에 g 가우시안의 사후 확률일 때, 선형 스펙트럼 입력 데이터가 $\mathbf{0}^s$, 평균벡터가 $\boldsymbol{\mu}^s$, 분산 행렬이 $\boldsymbol{\Sigma}^s$인 경우에, 다음의 수식 (1)과 같이 표현되는 변형된 바움(Baum)의 보조 함수인 것을 특징으로 하는, 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 장치.

$$F(\lambda,\lambda') = \phi - \frac{1}{2}P(\mathbf{O}|\lambda)\sum_t\sum_g \gamma_g^t \cdot[M\log(2\pi) + \log|\boldsymbol{\Sigma}_g^s| + (\mathbf{o}^{s,t} - \boldsymbol{\mu}_g^s)^T \boldsymbol{\Sigma}_g^{s-1}(\mathbf{o}^{s,t} - \boldsymbol{\mu}_g^s)], \quad (1)$$

청구항 11

제 9 항에 있어서,

상기 닫힌 연산 보조 함수는,

클린 음향 모델이 λ, 잡음 음향 모델을 λ'라 하고, ϕ가 전이 확률, γ_g^t는 t 시간에 g 가우시안의 사후 확률일 때, 선형 스펙트럼 입력 데이터가 $\mathbf{o}^s$, 평균벡터가 $\boldsymbol{\mu}^s$, 분산 행렬이 $\boldsymbol{\Sigma}^s$이며, A가 채널 왜곡과 관계된 잡음 매개 변수이고, b가 가산 잡음과 관계된 잡음 매개 변수인 경우에, 다음의 수식 (2)와 같이 표현되는 변형된 바움(Baum)의 보조 함수인 것을 특징으로 하는, 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 장치.

$$F(\lambda,\lambda') = -\sum_t\sum_g \gamma_g^t[\log(|\boldsymbol{\Sigma}_g^s|) - 2\log(|\mathbf{A}^{-1}|) + (\mathbf{A}^{-1}\mathbf{o}^{s,t} - \mathbf{A}^{-1}\mathbf{b} - \boldsymbol{\mu}_g^s)^T\boldsymbol{\Sigma}_g^{s-1}(\mathbf{A}^{-1}\mathbf{o}^{s,t} - \mathbf{A}^{-1}\mathbf{b} - \boldsymbol{\mu}_g^s)]. \quad (2)$$

청구항 12

제 8 항에 있어서,

상기 잡음 음성 모델 생성부는,

상기 잡음 음성 모델을 캡스트럼 도메인으로 변환하는 것을 특징으로 하는, 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 장치.

청구항 13

음성 신호를 입력받는 마이크부;

상기 음성 신호로부터 특징 벡터를 추출하는 특징 벡터 추출부;

클린 음향 모델과 입력 음성의 캡스트럼 데이터를 이용하여 가우시안 점유 확률을 추정하며, 상기 가우시안 점유 확률, 상기 입력 음성의 선형 스펙트럼 데이터 및 상기 클린 음향 모델을 선형 스펙트럼 도메인으로 변환한 음향 모델을 이용하여 잡음 매개 변수를 추정하는 과정을 통해 생성된 잡음 음성 모델을 저장하는 잡음 음성 모델 저장부; 및

상기 특징 벡터를 상기 잡음 음성 모델과 비교하여 최대 우도를 갖는 인식 단어를 생성하는 음성 인식부를 포함하는, 잡음 음성 모델을 이용한 음성 인식 장치.

명 세 서

발명의 상세한 설명

기 술 분 야

[0001] 본 발명은 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 방법에 관한 것으로서, 더욱 상세하게는 잡음 매개 변수 추정의 계산 비용을 줄일 수 있고, 음향 모델 적응이나 인식 과정의 실시간성을 향상시킬 수 있는 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 방법 및 그 장치, 잡음 음성 모델을 이용한 음성 인식 방법 및 그 장치에 관한 것이다.

배 경 기 술

[0002] 음성은 가장 쉽고 자연스러운 의사 전달의 수단인 동시에 음성의 입력 및 전달 과정에 고가의 장치가 필요 없다는 장점을 가지며 인간-기계 (man-machine) 인터페이스의 응용으로 다양한 분야에서 그 효용성을 인정받고

있다. 자동 예약,음성정보 서비스, 콜센터, PC 명령장치,자동 타이프라이터 등에 적용할 수 있다.

[0003] 음성인식 기술은 인식할 수 있는 사람의 종류에 따라 특정화자만 인식할 수 있는 화자종속 기술과 불특정 다수를 대상으로 하는 화자독립 기술로 나눌 수 있다. 발음의 형태에 따라서는 고립단어, 연결단어, 연속문장, 대화체 연속문장 인식 기술 등으로 나뉘며 특정 어휘만을 검출해서 인식하는 핵심어 검출 기술이 있다. 어휘 수에 따라서는 수십,수백 단어를 다루는 소용량, 수만 어휘의 인식이 가능한 대용량 인식 기술 등으로 분류할 수 있다.

[0004] 음성 또는 소리의 검출 및 인식 시스템에 대한 다수의 방안들이 제안되어 왔고 어느 정도 성공적으로 구현되었다. 이러한 시스템들은 이론적으로 사용자의 발음(utterance)을 등록된 화자의 발음에 대해 매칭시켜 사용자 신원에 따라 장치 또는 시스템의 자원들에 대한 액세스를 허용 또는 거부하거나, 등록된 화자를 식별하거나 개별화된(customized) 커맨드 라이브러리들을 호출할 수 있다.

[0005] 한편, 음성 인식에 사용되는 HMM (Hidden Markov Models)은 음의 상태가 한 상태에서 다음 상태로 바뀌는 것을 천이 확률로 표현한다. HMM은 음성 신호의 시간적인 통계적 특성을 이용하여 훈련 데이터로부터 이들을 대표하는 모델을 구성한 후 실제 음성 신호와 유사도가 높은 확률 모델을 인식 결과로 채택하는 방법이다. 이 방법은 단독음이나 연결음, 연속음 인식에까지 구현이 용이하며 좋은 인식 성능을 나타내어 여러 가지 응용 분야에 많이 이용되고 있다. 실제로 음성 인식 기술이 대중화된 계기는 HMM의 등장이라고 할 수 있다.

[0006] HMM은 수학적인 배경에서 개발된 알고리즘으로 전통적인 확률분포를 이용하며, 시간 정보와 잘 연동되기 때문에 화자 독립, 대화체 음성 인식 등 많은 장점을 갖고 있다. 또한 대어휘에서 DTW(Dynamic Time Warping)보다는 계산량이 적은 장점을 갖고 있다.그러나 학습 데이터가 부족할 경우, 모델간의 변별력이 부족하고 음성 신호간의 연관성을 무시하는 경향이 있다.

[0007] 음성 인식 도메인에서의 음향 모델 적응 알고리즘은 환경 잡음 등의 잡음 데이터에 대한 분석 기법을 제대로 적용할 수 없다는 단점이 있다. 왜냐하면 입력된 음성 데이터는 특징 추출 과정에서 가산 잡음과 채널 왜곡을 분석할 수 있는 선형 스펙트럼 영역에서 켑스트럼 도메인의 데이터로 변환되기 때문이다. 이를 극복하기 위해 선형 스펙트럼 도메인에서 잡음을 모델링하여 음향 모델과 결합시키는 방법이 제안 되었는데, 이 방법은 음성이 없는 구간의 잡음만 별도로 정확히 모델링해야 하기 때문에 음성 데이터 다루기 이전의 사전 연산 과정을 필요로 한다. 그리고 선형 스펙트럼에서 잡음 매개변수를 추정한 뒤에 향상된 음향 모델을 캡스트럼 영역으로 변환시키는 기법인 ML-LST (Maximum Likelihood Linear Spectral Transformation) 기법이 제안 되었는데, 이 방법은 매개변수를 추정하기 위한 반복 재귀연산을 수행하기 때문에 계산 비용이 과다한 문제점이 있다.

발명의 내용

해결 하고자하는 과제

[0008] 따라서, 본 발명이 해결하고자 하는 첫 번째 과제는 음향 모델 적응 과정의 계산 비용을 줄이고 실시간성을 향상시킬 수 있는 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 방법을 제공하는 것이다.

[0009] 본 발명이 해결하고자 하는 두 번째 과제는 적응 양의 데이터를 이용하여 음성 인식의 실시간성을 향상시킬 수 있는 잡음 음성 모델을 이용한 음성 인식 방법을 제공하는 것이다.

[0010] 본 발명이 해결하고자 하는 세 번째 과제는 음향 모델 적응 과정의 계산 비용을 줄이고 실시간성을 향상시킬 수 있는 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 장치를 제공하는 것이다.

[0011] 본 발명이 해결하고자 하는 네 번째 과제는 적응 양의 데이터를 이용하여 음성 인식의 실시간성을 향상시킬 수 있는 잡음 음성 모델을 이용한 음성 인식 장치를 제공하는 것이다.

과제 해결수단

[0012] 본 발명은 상기 첫 번째 과제를 달성하기 위하여, 입력 음성에서 선형 스펙트럼 데이터 및 캡스트럼 데이터를 추출하는 단계; 클린 음향 모델과 상기 캡스트럼 데이터를 이용하여 가우시안 점유 확률을 추정하는 단계; 상기 가우시안 점유 확률, 상기 선형 스펙트럼 데이터 및 상기 클린 음향 모델을 선형 스펙트럼 도메인으로 변환한 음향 모델을 이용하여 잡음 매개 변수를 추정하는 단계; 및 상기 잡음 매개 변수를 이용하여 잡음 음성 모델을 생성하는 단계를 포함하는, 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 방법을 제공한다.

[0013] 본 발명의 일 실시 예에 의하면, 상기 잡음 매개 변수를 추정하는 단계는 닫힌 연산 보조 함수를 이용하는 단계

일 수 있다.

[0014] 본 발명의 다른 실시 예에 의하면, 상기 잡음 음성 모델을 생성하는 단계는 상기 잡음 음성 모델을 캡스트럼 도메인으로 변환하는 단계를 포함할 수 있다.

[0015] 본 발명은 상기 두 번째 과제를 달성하기 위하여, 발화 음성으로부터 특징 벡터를 추출하는 단계; 및 상기 특징 벡터를 잡음 음성 모델과 비교하여 최대 우도를 갖는 인식 단어를 생성하는 단계를 포함하는 잡음 음성 모델을 이용한 음성 인식 방법을 제공한다. 여기서, 상기 잡음 음성 모델은 클린 음향 모델과 입력 음성의 캡스트럼 데이터를 이용하여 가우시안 점유 확률을 추정하며, 상기 가우시안 점유 확률, 상기 입력 음성의 선형 스펙트럼 데이터 및 상기 클린 음향 모델을 선형 스펙트럼 도메인으로 변환한 음향 모델을 이용하여 잡음 매개 변수를 추정하는 과정을 통해 생성된 음성 모델이다.

[0016] 본 발명은 상기 세 번째 과제를 달성하기 위하여, 입력 음성에서 특징 벡터에 대한 선형 스펙트럼 데이터를 추출하는 선형 스펙트럼 특징 추출부; 상기 선형 스펙트럼 데이터를 캡스트럼 데이터로 변환하는 캡스트럼 특징 추출부; 클린 음향 모델과 상기 캡스트럼 데이터를 이용하여 가우시안 점유 확률을 추정하는 점유 확률 계산부; 상기 가우시안 점유 확률, 상기 선형 스펙트럼 데이터 및 상기 클린 음향 모델을 선형 스펙트럼 도메인으로 변환한 음향 모델을 이용하여 잡음 매개 변수를 추정하는 잡음 매개 변수 추정부; 및 상기 잡음 매개 변수를 이용하여 잡음 음성 모델을 생성하는 잡음 음성 모델 생성부를 포함하는, 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 장치를 제공한다.

[0017] 본 발명의 일 실시 예에 의하면, 상기 잡음 매개 변수 추정부는 닫힌 연산 보조 함수를 이용할 수 있다.

[0018] 본 발명의 다른 실시 예에 의하면, 상기 잡음 음성 모델 생성부는 상기 잡음 음성 모델을 캡스트럼 도메인으로 변환할 수 있다.

[0019] 본 발명은 상기 네 번째 과제를 달성하기 위하여, 음성 신호를 입력받는 마이크부; 상기 음성 신호로부터 특징 벡터를 추출하는 특징 벡터 추출부; 클린 음향 모델과 입력 음성의 캡스트럼 데이터를 이용하여 가우시안 점유 확률을 추정하며, 상기 가우시안 점유 확률, 상기 입력 음성의 선형 스펙트럼 데이터 및 상기 클린 음향 모델을 선형 스펙트럼 도메인으로 변환한 음향 모델을 이용하여 잡음 매개 변수를 추정하는 과정을 통해 생성된 잡음 음성 모델을 저장하는 잡음 음성 모델 저장부; 및 상기 특징 벡터를 상기 잡음 음성 모델과 비교하여 최대 우도를 갖는 인식 단어를 생성하는 음성 인식부를 포함하는, 잡음 음성 모델을 이용한 음성 인식 장치를 제공한다.

효 과

[0020] 본 발명에 의하면, 잡음 매개변수를 추정하는 과정에서 재귀 연산을 하지 않고 닫힌 연산 추정법을 적용하여 계산 비용을 줄일 수 있고, 음향 모델 적응이나 인식 과정의 실시간성을 향상시킬 수 있다.

발명의 실시를 위한 구체적인 내용

[0021] 이하에서는 도면을 참조하여 본 발명의 바람직한 실시 예를 설명하기로 한다. 그러나, 다음에 예시하는 본 발명의 실시 예는 여러 가지 다른 형태로 변형될 수 있으며, 본 발명의 범위가 다음에 상술하는 실시 예에 한정되는 것은 아니다.

[0022] 도 1은 음성 인식 시스템의 일 예를 도시한 것이다.

[0023] 음성 인식 시스템은 크게 전처리부와 인식부로 나눌 수 있다. 전처리부에서는 사용자가 발성한 음성에 대한 음성 분석 과정을 거치면서 인식에 필요한 특징 벡터를 추출한다. 패턴 인식 과정에서 음성 데이터베이스로부터 훈련한 기준 패턴과의 비교를 통해서 인식 결과를 얻게 된다. 보다 복잡한 구조의 음성을 인식할 때에는 언어모델을 이용한 언어 처리 과정을 통해 최종 인식 결과를 출력한다.

[0024] 도 2는 본 발명에 이용되는 HMM에서 상태 천이의 예를 도시한 것이다.

[0025] 여기서, 시간 t에서 관측될 수 있는 심볼은 $V=\{v_1, v_2, \cdots, v_m\}$ 와 같이 표현되고, 상태 천이 확률 분포는 $A=\{a_{ij}\}$, 각각의 상태에서 관측되는 심볼의 확률 분포는 $B=\{b_j(k)\}$ 이다. 초기 상태 분포는 $\pi=\{\pi_i\}$ 이고, HMM의 파라미터는 $\lambda=\{A,B,\pi\}$ 이다.

[0026] 도 2에는 HMM의 상태가 3가지인 경우를 도시되어 있는데, a_{ij}는 상태 i에서 j로 천이될 확률을 나타내며, a_{ik}는 상태 i에서 심볼 k가 관측될 확률을 나타낸다.

[0027] HMM을 이용하여 음성인식을 하고자 할 때 다음과 같은 세가지가 중요하다. 첫 번째 는, 모델 λ가 주어 졌을 때, 관측 열 O가 λ에서 발생할 확률을 계산하는 것이다. 두 번째는, 관측 열 O가 주어졌을 때, 가장 확률이 높은 상태 열 X를 찾아 내는 것이다. 마지막은 모델 λ와 관측 열 O가 주어졌을 때를 최대로 하는 P(O|λ)모델을 추정하는 것이다.

[0028] 도 3은 깨끗한 음성이 잡음 섞인 음성으로 변화되는 흐름도이다.

[0029] 잡음과 음성이 서로 독립적이라고 가정하면, 가산 잡음은 선형 스펙트럼 영역에서 합산으로 영향을 주고, 채널 왜곡과 같은 컨볼루션 (convolutional) 잡음은 선형 스펙트럼에서 곱셈 연산과 같이 영향을 준다.

[0030] 도 4는 본 발명의 일 실시 예에 따른 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 장치의 블록도이다.

[0031] 본 발명은 기존 적응 알고리즘의 단점으로 지적했던 잡음 데이터에 대한 분석을 적용할 수 있고 연산 비용을 줄일 수 있는 알고리즘 제공한다. 기존에 ML-LST 적응 알고리즘에서 잡음 매개변수를 선형 스펙트럼에서 재귀 연산으로 반복 추정이 필요했던 방법을 닫힌 연산(closed-form) 방법으로 변경한다.

[0032] 음성 입력부(410)는 인식하기 위한 음성을 입력다. 음향 모델 차원에서 입력되는 음성은 한 두 단어에서 여러 문장에 이르는 잡음 데이터에 해당한다.

[0033] 선형 스펙트럼 특징 추출부(420)는 입력 음성에서 특징을 추출하여 선형 스펙트럼 영역 데이터를 생성한다. 선형 스펙트럼 특징 추출부(420)는 입력된 음성 신호로부터 인식에 유효한 특징 파라미터를 뽑아낸다. 동일한 단어를 여러 사람이 발음하였을 경우 단어의 의미가 동일하더라도 음성 파형은 동일하지 않으며, 동일한 사람이 동일한 단어를 동일한 시간에 연속으로 발음하였다고 하여도 음성 파형은 동일하지 않다. 이와 같은 현상의 이유는 음성 파형에서는 음성의 의미 정보 이외에도 화자의 음색, 감정 상태 등과 같은 정보도 포함하고 있기 때문이다. 그러므로 음성의 특징 추출이란 음성으로부터 의미 정보를 나타내어주는 특징을 추출하는 것으로 일종의 음성 압축 부분이며 한편으로 인간의 발성기관을 모델링하는 부분이라고 생각할 수 있다.

[0034] 캡스트럼 특징 추출부(430)는 특징을 반영한 선형 스펙트럼 데이터를 캡스트럼 영역 데이터로 변환한다.

[0035] 점유 확률 계산부(440)는 클린 음향 모델(435)과 포워드-백워드 (Forward-Backward) 알고리즘을 이용하여 가우시안 점유 확률을 계산한다. 보다 구체적으로, 음향모델의 평균값은 가우시안 점유 확률로부터 구할 수 있고, 잡음 매개변수를 구하기 위해서는 가우시안 점유 확률이 필요하다.

[0036] 도메인 변환부(450)는 캡스트럼 영역의 클린 음향 모델(435)을 선형 스펙트럼 영역 음향 모델로 변환한다.

[0037] 잡음 매개 변수 추정부(460)는 선형 스펙트럼 특징 추출부(420)의 선형 스펙트럼 데이터, 점유 확률 계산부(440)의 가우시안 점유 확률, 도메인 변환부(450)의 선형 스펙트럼 음향 모델을 이용하여 잡음 매개변수(또는 변환 매개 변수)를 추정한다.

[0038] 잡음 음성 모델 생성부(470)는 잡음 매개 변수 추정부(460)에서 추정된 잡음 매개 변수와 선형 스펙트럼 음향 모델과 결합시켜 잡음 섞인 음향 모델을 생성한다.

[0039] 본 발명의 일 실시 예에 따른 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 장치는 잡음 음성 모델 생성부(470)에서 생성된 잡음 음성 모델을 캡스트럽 영역으로 변환하는 도메인 변환 수단을 더 구비할 수도 있다.

[0040] 한편, 캡스트럼 특징 추출부(430)에서 입력 받은 특징 관측벡터로부터 HMM 음향모델의 스테이트들에 대한 가우시안 점유 확률은 다음과 같은 알고리즘을 통해 구해진다.

[0041] 먼저, 포워드 알고리즘을 통한 포워드 확률계산에 대해 설명한다.

[0042] λ 음향모델이 주어진 경우, 처음부터 특정 t 시간까지 관측벡터 o_t가 입력되었을 때 q 스테이트열의 t번째인 i 스테이트를 지나갈 확률 $\alpha_t(i)$는 수학식 1과 같이 표현된다.

수학식 1

[0043] $$\alpha_t(i) = P(o_1 o_2 o_3 \cdots o_t, q_t = i \mid \lambda)$$

[0044] 초기값은 수학식 2와 같이 표현된다.

수학식 2

[0045] $$\alpha_1(i) = a_{0i} b_i(o_1) \quad for \ (1 < i < N)$$

[0046] 여기서 a_{0i} 는 초기 0 스테이트에서 i 스테이트까지의 전이확률이다. $b_i(o_1)$는 i 스테이트에서 관측벡터 o_1 이 발생할 확률이다.

[0047] 따라서, 매 t+1 시간에 j 스테이트에 대한 포워드 확률은 수학식 3과 같이 표현된다.

수학식 3

[0048] $$\alpha_{t+1}(j) = [\sum_{i=1}^{N} \alpha_t(i) \cdot a_{ij}] \cdot b_j(o_{t+1}) \quad for \ (1 < j < N) \quad for \ (1 < t < T-1)$$

[0049] 결론적으로, 1부터 N까지 모든 스테이트에 대해 각각의 스테이트를 마지막으로 끝내는 모든 관측열의 확률은 수학식 4와 같이 표현된다.

수학식 4

[0050] $$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

[0051] 다음으로 백워드 알고리즘을 통한 확률 계산을 설명한다.

[0052] 먼저, 주어진 음향모델 λ 에서 특정 t 시간에 q_t 인 i 스테이트를 거쳐서 마지막 T시간까지의 관측 벡터열이 나올 확률 $\beta_t(i)$ 은 수학식 5와 같이 표현된다.

수학식 5

[0053] $$\beta_t(i) = P(o_{t+1} o_{t+2} \cdots o_T \mid q_t = i, \lambda)$$

[0054] 초기값 $\beta_T(i)$ 는 마지막 T 시간에 i 스테이트에 있을 확률로서, 수학식 6과 같이 표현된다.

수학식 6

[0055] $$\beta_T(i) = 1 \quad for \ (1 < i < N)$$

[0056] 따라서, t시간부터 마지막 T시간까지 i 스테이트에서의 backward 확률 $\beta_t(i)$ 는 수학식 7과 같이 표현된다.

수학식 7

[0057] $$\beta_t(i) = \sum_{j=1}^{N} a_{ij} \cdot b_j(o_{t+1}) \cdot \beta_{t+1}(j) \quad for \ (T-1 \geq t \geq 1) \quad for \ (1 \leq i \leq N)$$

[0058] 결론적으로, 처음 시간 1에서부터 스테이트 j를 거쳐서 T시간까지 관측 벡터열이 끝날 확률 $P(O \mid \lambda)$는 수학식 8과 같이 표현된다.

수학식 8

[0059] $$P(O|\lambda) = \sum_{j=1}^{N} a_{0j} \cdot b_j(o_1) \cdot \beta_1(j)$$

[0060] 이하에서는, 백워드-포워드 확률을 통한 점유 확률 계산에 대해 설명한다.

[0061] 먼저, 음향모델 λ 와 관측 벡터열 O 이 주어졌을 때 t시간에 스테이트 i를 지나갈 점유 확률 $\gamma_t(i)$ 는 수학식 9와 같이 표현된다.

수학식 9

[0062] $$\gamma_t(i) = P(q_t = i | O, \lambda) = P(q_t = i | o_1 o_2 \cdots o_T, \lambda)$$

[0063] 수학식 9는 수학식 10과 같이 표현될 수 있다.

수학식 10

[0064] $$= \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)},$$

[0065] 여기서, $\alpha_t(i) = [\sum_{i=1}^{N} \alpha_{t-1}(i) \cdot a_{ij}] \cdot b_i(o_t)$, $\beta_t(i) = \sum_{j=1}^{N} a_{ij} \cdot b_j(o_{t+1}) \cdot \beta_{t+1}(j)$ 를 나타낸다. 또한, $\alpha_t(i)$ 는 처음시간부터 t시간까지 i 스테이트에 대한 포워드 확률값이고, $\beta_t(i)$ 는 i 스테이트를 거쳐서 t 시간 이후부터 마지막 T시간까지의 백워드 확률값이다.

[0066] 이와 같은 알고리즘을 이용하여 입력된 특징 관측벡터로부터 HMM 음향모델의 각 스테이트에 대한 점유 확률을 구할 수 있다.

[0067] 도 5는 캡스트럼 영역의 음향 모델과 선형 스펙트럼 영역의 음향 모델이 서로 변환되는 과정을 도시한 것이다.

[0068] 도 6은 도 5의 변환 과정을 상세히 도시한 것이다.

[0069] 켑스트럼 도메인의 HMM은 이산 코사인 역변환 (Inverse Discrete Cosine Transformation; IDCT)과 지수 연산을 거쳐 선형 스펙트럼 도메인으로 변환된다. 한편, 선형 스펙 트럼 도메인에서 생성된 잡음 음성 모델은 로그 연산과 이산 코사인 변환 (Discrete Cosine Transformation; DCT)을 거치면서 켑스트럼 도메인의 HMM으로 변환된다.

[0070] 이하에서는 본 발명에서 잡음 매개 변수를 추정하기 위해 사용되는 변형된 바움(Baum)의 보조 함수에 대해 설명한다.

[0071] 먼저, 잡음이 섞인 선형 스펙트럼 평균 $\tilde{\boldsymbol{\mu}}^s$ 를 다음의 수학식 11과 같이 표현할 수 있다.

수학식 11

[0072] $$\tilde{\boldsymbol{\mu}}^s = \mathbf{A} \cdot \boldsymbol{\mu}^s + \mathbf{b}$$

[0073] 여기서 A는 채널 왜곡과 관계된 잡음 매개 변수이고, b는 가산 잡음과 관계된 잡음 매개 변수이다. 이때 A는 대각행렬이고 b는 벡터인데 $\boldsymbol{\mu}^s$ 는 선형 스펙트럼 영역의 평균값 벡터이다. 음성 인식 시스템은 일반적으로 선형 스펙트럼 영역 보다 캡스트럼 영역에서 표현된 음향 모델을 이용하기 때문에 수학식 11을 사용하기 전에 도 5와 같이 캡스트럼 영역의 음향 모델을 선형 스펙트럼 영역의 음향 모델로 만들어야 한다. 그러나 닫힌 연산을 하는 매개변수 추정식을 만들기 위해서는 영역 변환 연산에서 생기는 복잡한 연산과정을 극복해야 한다. 그래서 캡스트럼 영역에서 최우도(maximum likelihood)를 지향하는 바움의 보조 함수를 사용하는 대신에 입력된 데이터와 음향 모델 사이의 거리 값을 줄이는 변형된 목적 함수를 이용한다. 이를 이용하면, 재귀연산에 의하지 않고 직접적으로 잡음 매개변수를 구할 수 있고, 보조 함수에 대한 복잡한 영역 변환을 피할 수 있다. 바움의 보조 함

수를 근사한 함수는 다음의 수학식 12와 같다.

수학식 12

$$F(\lambda,\lambda') = \phi - \frac{1}{2}P(\mathbf{O}\,|\,\lambda)\sum_t\sum_g \gamma_g^t \cdot[M\log(2\pi) + \log|\,\boldsymbol{\Sigma}_g^s\,| + (\mathbf{o}^{s,t} - \boldsymbol{\mu}_g^s)^T \boldsymbol{\Sigma}_g^{s-1}(\mathbf{o}^{s,t} - \boldsymbol{\mu}_g^s)],$$

[0074]

[0075] 여기서 λ과 λ'는 각각 클린 음향 모델과 잡음 음성 모델을 나타낸다. ϕ는 전이 확률이고, γ_g^t는 t 시간에 g 가우시안의 사후 확률이다. 변형된 바움의 보조 함수는 캡스트럼 영역의 입력 데이터와 음향모델과 달리 선형 스펙트럼 입력 데이터 $\mathbf{o}^s$와 평균벡터 $\boldsymbol{\mu}^s$, 그리고 분산 행렬 $\boldsymbol{\Sigma}^s$를 사용한다. 이 수학식이 최적화될 때 선형 스펙트럼 영역에서 잡음 섞인 입력 데이터와 잡음 없는 음향 모델 사이의 거리는 종래 바움의 보조 함수의 값을 증가시키는 방향으로 줄어들게 된다. 위 식을 이용하여 잡음 매개변수를 구하기 위해서 평균값은 수학식 11처럼 변환되고, 잡음 섞인 선형 스펙트럼 분산값 $\tilde{\boldsymbol{\Sigma}}^s$는 다음과 같이 변환된다.

수학식 13

$$\tilde{\boldsymbol{\Sigma}}^s = \mathbf{A}\boldsymbol{\Sigma}^s\mathbf{A}^T$$

[0076]

[0077] 변형된 바움의 보조 함수는 결과에 영향을 주지 않는 상수 값들을 없애 다음 수학식처럼 단순화할 수 있다.

수학식 14

$$F(\lambda,\lambda') = -\sum_t\sum_g \gamma_g^t[\log(|\mathbf{A}\boldsymbol{\Sigma}_g^s\mathbf{A}^T|) + (\mathbf{o}^{s,t} - (\mathbf{A}\boldsymbol{\mu}_g^s + \mathbf{b}))^T(\mathbf{A}\boldsymbol{\Sigma}_g^s\mathbf{A}^T)^{-1}(\mathbf{o}^{s,t} - (\mathbf{A}\boldsymbol{\mu}_g^s + \mathbf{b}))]$$

[0078]

[0079]

[0080] 그리고 잡음 매개 변수가 포함된 식은 다음과 같이 전개 할 수 있다.

수학식 15

$$F(\lambda,\lambda') = -\sum_t\sum_g \gamma_g^t[\log(|\boldsymbol{\Sigma}_g^s|) - 2\log(|\mathbf{A}^{-1}|) + (\mathbf{A}^{-1}\mathbf{o}^{s,t} - \mathbf{A}^{-1}\mathbf{b} - \boldsymbol{\mu}_g^s)^T\boldsymbol{\Sigma}_g^{s-1}(\mathbf{A}^{-1}\mathbf{o}^{s,t} - \mathbf{A}^{-1}\mathbf{b} - \boldsymbol{\mu}_g^s)].$$

[0081]

[0082]

[0083] 최우도를 증가시키는 방향으로 최적의 잡음 매개변수를 얻기 위해 위 식은 $\mathbf{A}^{-1}$의 K번째 성분인 A_{kk}^{-1}으로 편미분하여 다음 식과 같이 표현할 수 있다.

수학식 16

$$\frac{dF(\lambda,\lambda')}{dA_{kk}^{-1}} = \sum_t\sum_g \gamma_g^t\left[\frac{-1}{A_{kk}^{-1}} + \frac{(A_{kk}^{-1}o_k^{s,t} - A_{kk}^{-1}b_k - \mu_{g,k}^s)(o_k^{s,t} - b_k)}{\Sigma_{g,kk}^s}\right]$$

$$= 0$$

[0084]

[0085] 또한 수학식 15는 잡음 매개변수 b_k에 대해 편미분하여 다음과 같이 전개할 수 있다.

수학식 17

$$\frac{dF(\lambda,\lambda')}{db_k} = \sum_t \sum_g \gamma_g^t \left[\frac{(A_{kk}^{-1} o_k^{s,t} - A_{kk}^{-1} b_k - \mu_{g,k}^s) A_{kk}^{-1}}{\Sigma_{g,kk}^s} \right] = 0$$

[0086]

[0087] 각각 A_{kk}^{-1}와 b_k에 대해 수학식 16과 수학식 17을 동시에 풀면 잡음 매개변수는 다음 수학식 18과 수학식 19처럼 구할 수 있다.

수학식 18

$$A_{kk}^{-1} = \frac{\sum_t \sum_g \frac{\gamma_g^t}{\Sigma_{g,kk}^s} \left(o_k^{s,t} - \frac{Gvo}{Gv} \right) \left(\frac{2}{Gv} + \mu_{g,k}^s \right)}{2 \sum_t \sum_g \frac{\gamma_g^t}{\Sigma_{g,kk}^s} \left(o_k^{s,t} - \frac{Gvo}{Gv} \right)^2}$$

[0088]

$$+ \frac{\sqrt{\left[\sum_t \sum_g \frac{\gamma_g^t}{\Sigma_{g,kk}^s} \left(o_k^{s,t} - \frac{Gvo}{Gv} \right) \left(\frac{2}{Gv} + \mu_{g,k}^s \right) \right]^2 + 4 \sum_t \sum_g \gamma_g^t \cdot \sum_t \sum_g \frac{\gamma_g^t}{\Sigma_{g,kk}^s} \left(o_k^{s,t} - \frac{Gvo}{Gv} \right)^2}}{2 \sum_t \sum_g \frac{\gamma_g^t}{\Sigma_{g,kk}^s} \left(o_k^{s,t} - \frac{Gvo}{Gv} \right)^2}$$

[0089]

수학식 19

$$b_k = \frac{A_{kk}^{-1} \cdot Gvo - \sum_t \sum_g \frac{\gamma_g^t}{\Sigma_{g,kk}^s} \mu_{g,k}^s}{A_{kk}^{-1} \cdot Gv}$$

[0090]

[0091] 여기서 표현된 Gvo와 Gv는 각가 수학식 20과 수학식21과 같다.

수학식 20

$$Gvo = \sum_t \sum_g \frac{\gamma_g^t}{\Sigma_{g,kk}^s} o_k^{s,t}$$

[0092]

수학식 21

$$Gv = \sum_t \sum_g \frac{\gamma_g^t}{\Sigma_{g,kk}^s}$$

[0093]

[0094] 이렇게 구해진 잡음 매개변수를 수학식 11에 적용하여 잡음 섞인 음향 모델을 만든 후에 도 5와 같이 캡스트럼 영역으로 변환하면 음성 인식에 이용할 수 있다. 본 발명은 이와 같이 음성과 함께 섞인 잡음에서 잡음 매개변수를 추정하는 방법을 이용하므로, 따로 잡음 구간만 선별하거나 잡음 특성을 추출하는 과정을 필요로하지 않는다. 또한, 잡음 모델을 따로 만드는 절차를 거치지 않고 추정된 잡음 매개변수를 이용하여 잡음 음성 모델을 생성할 수 있다.

[0095] 이하에서는 잡음 음성 모델을 이용하여 음성을 인식하는 방법에 대해 설명한다.

[0096] 도 7은 본 발명의 일 실시 예에 따른 잡음 음성 모델을 이용한 음성 인식 장치의 블록도이다.

[0097] 마이크부(781)는 화자로부터 발화된 음성 신호를 입력받는다.

[0098] 특징 벡터 추출부(782)는 음성 신호로부터 특징 벡터를 추출한다.

[0099] 잡음 음성 모델 저장부(790)는 변형된 바움의 보조함수를 이용하여 구해진 잡음 매개 변수를 이용하여 생성된 잡음 음성 모델을 저장한다. 잡음 음성 모델은 상술한 바와 같이, 클린 음향 모델과 입력 음성의 캡스트럼 데이터를 이용하여 가우시안 점유 확률을 추정하며, 상기 가우시안 점유 확률, 상기 입력 음성의 선형 스펙트럼 데이터 및 상기 클린 음향 모델을 선형 스펙트럼 도메인으로 변환한 음향 모델을 이용하여 잡음 매개 변수를 추정하는 과정을 통해 생성된다.

[0100] 음성 인식부(783)는 특징 벡터를 잡음 음성 모델과 비교하여 최대 우도를 갖는 인식 단어나 문장을 산출한다.

[0101] 본 발명에서는 바움의 보조함수를 변형하여 실시간 적응이 가능하게 하였다. 그리고 캡스트럼 영역보다 잡음에 대한 분석이 용이한 선형 스펙트럼 영역에서 적은 수의 매개변수를 다루기 때문에 적은 양의 적응 데이터를 이용하여 실시간 음향 모델 적응을 할 수 있다.

[0102] 표 1은 여러 잡음 배경에서 기존 MLLR 알고리즘과 비교한 성능표이다.

표 1

[0103]

잡음배경의 데이터	비 적응	MLLR	본 발명
Airport	54.7	35.2	33.1
Babble	57.9	38.1	37.3
Car	53.5	26.9	20.5
평균	55.4	33.4	30.3

[0104] 본 발명에 의하면, 캡스트럼 영역에서 적응하는 알고리즘 MLLR보다 성능은 좋고, 반복 재귀연산 알고리즘을 닫힌 연산으로 바꿔 ML-LST보다 계산 비용을 줄이는 효과를 볼 수 있다.

[0105] 본 발명은 소프트웨어를 통해 실행될 수 있다. 바람직하게는, 본 발명의 일 실시 예에 따른 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 방법이나 잡음 음성 모델을 이용한 음성 인식 방법을 컴퓨터에서 실행시키기 위한 프로그램을 컴퓨터로 읽을 수 있는 기록매체에 기록하여 제공할 수 있다. 소프트웨어로 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.

[0106] 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다. 컴퓨터가 읽을 수 있는 기록 장치의 예로는 ROM, RAM, CD-ROM, DVD±ROM, DVD-RAM, 자기 테이프, 플로피 디스크, 하드 디스크(hard disk), 광데이터 저장장치 등이 있다. 또한, 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 장치에 분산되어 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.

[0107] 본 발명은 도면에 도시된 일 실시 예를 참고로 하여 설명하였으나 이는 예시적인 것에 불과하며 당해 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 실시 예의 변형이 가능하다는 점을 이해할 것이다. 그리고, 이와 같은 변형은 본 발명의 기술적 보호범위 내에 있다고 보아야 한다. 따라서, 본 발명의 진정한 기술적 보호범위는 첨부된 특허청구범위의 기술적 사상에 의해서 정해져야 할 것이다.

산업이용 가능성

[0108] 본 발명은 잡음 매개 변수 추정의 계산 비용을 줄일 수 있고, 음향 모델 적응이나 인식 과정의 실시간성을 향상시킬 수 있는 모델 적응 및 음성 인식에 관한 것으로, 음성 정보 처리 분야 중 음향모델 적응 기술 관련 장치 및 소프트웨어, 음성 인식 장치 및 소프트웨어에 적용될 수 있다.

도면의 간단한 설명

[0109] 도 1은 음성 인식 시스템의 일 예를 도시한 것이다.

[0110] 도 2는 본 발명에 이용되는 HMM에서 상태 천이의 예를 도시한 것이다.

[0111] 도 3은 깨끗한 음성이 잡음 섞인 음성으로 변화되는 흐름도이다.

[0112] 도 4는 본 발명의 일 실시 예에 따른 최대 우도 선형 스펙트럴 변환을 이용한 음향 모델 적응 장치의 블록도이다.

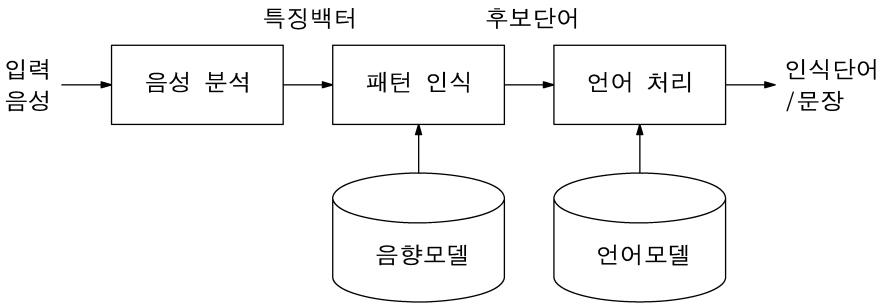
[0113] 도 5는 캡스트럼 영역의 음향 모델과 선형 스펙트럼 영역의 음향 모델이 서로 변환되는 과정을 도시한 것이다.

[0114] 도 6은 도 5의 변환 과정을 상세히 도시한 것이다.

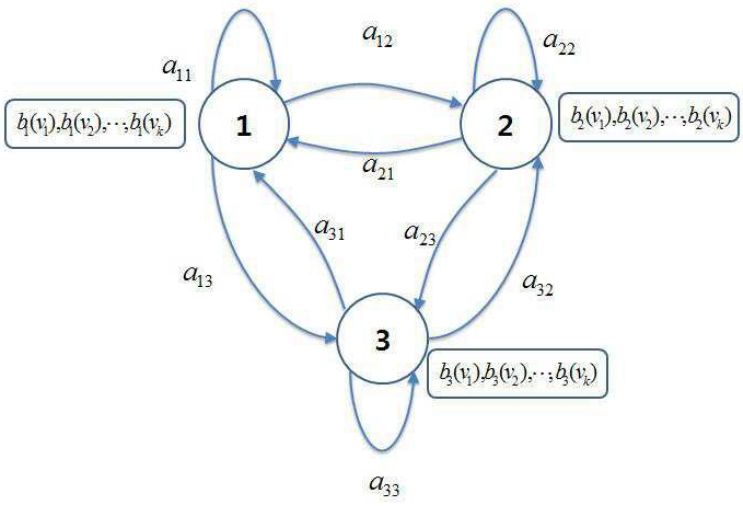
[0115] 도 7은 본 발명의 일 실시 예에 따른 잡음 음성 모델을 이용한 음성 인식 장치의 블록도이다.

도면

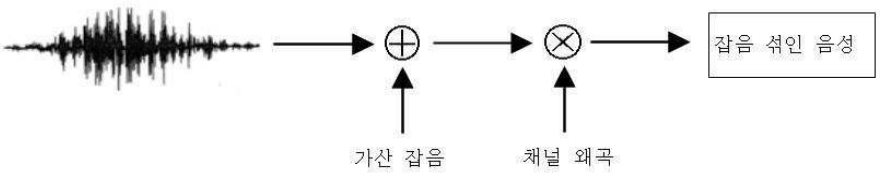
도면1



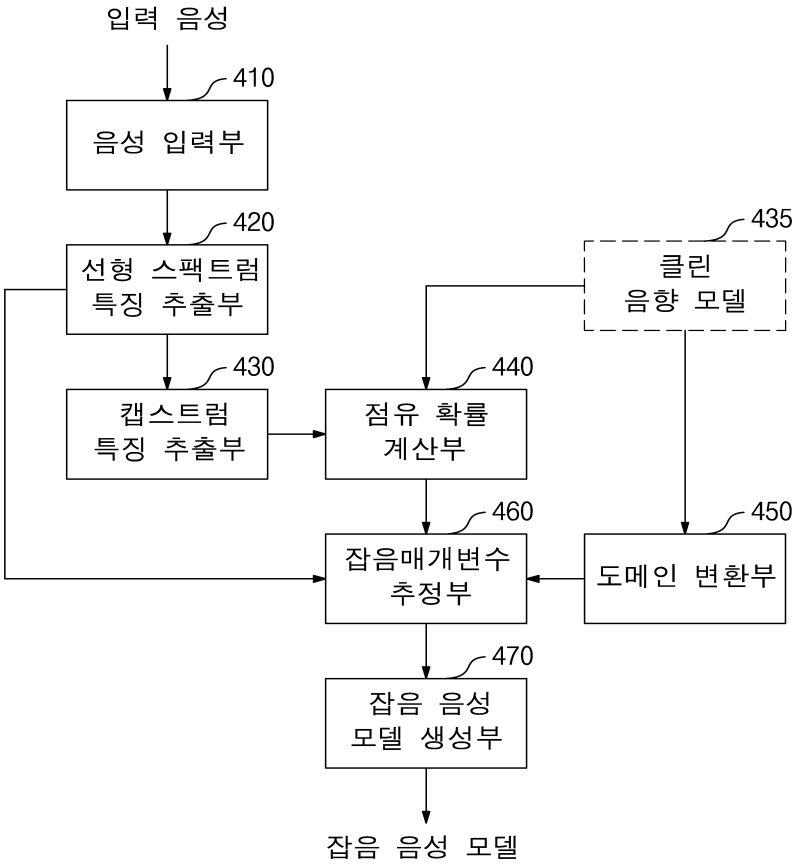
도면2



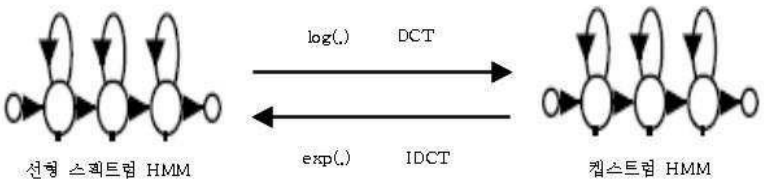
도면3



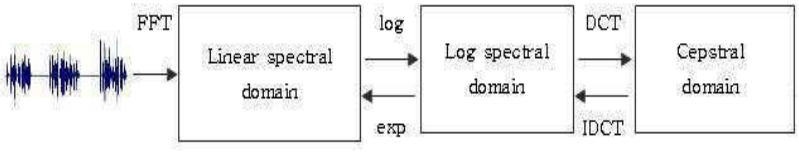
도면4


입력 음성
410
음성 입력부
420
선형 스팩트럼
특징 추출부
435
클린
음향 모델
430
캡스트럼
특징 추출부
440
점유 확률
계산부
460
잡음매개변수
추정부
450
도메인 변환부
470
잡음 음성
모델 생성부
잡음 음성 모델

도면5


log(.)
DCT
exp(.)
IDCT
선형 스펙트럼 HMM
켑스트럼 HMM

도면6


FFT
Linear spectral
domain
log
exp
Log spectral
domain
DCT
IDCT
Cepstral
domain

도면7