(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2008-0021234
(43) 공개일자 2008년03월07일

(51) Int. Cl.
G10L 15/14 (2006.01) *G10L 15/00* (2006.01)
G10L 19/06 (2006.01) *G10L 15/20* (2006.01)
(21) 출원번호 10-2006-0083523
(22) 출원일자 2006년08월31일
심사청구일자 2006년08월31일

(71) 출원인
계명대학교 산학협력단
대구 달서구 신당동 1000번지
정용주
대구 남구 이천동 상아맨션 102-809
(72) 발명자
정용주
대구 남구 이천동 상아맨션 102-809
(74) 대리인
김일환

전체 청구항 수 : 총 6 항

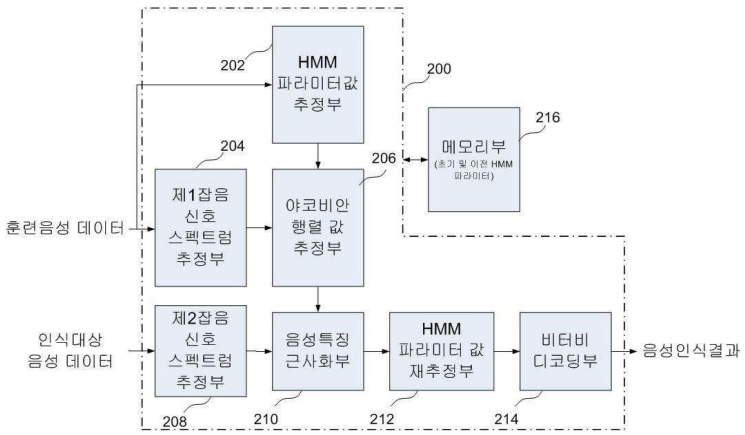
(54) 데이터 기반의 J A 방식을 이용한 음성 인식 방법 및장치

(57) 요 약

본 발명은 HMM 파라미터 값을 훈련 음성 데이터를 이용해서 바움 웰치(Baum-Welch) 방식으로 직접 훈련시킴과 아울러, 야코비안 행렬 값도 바움 웰치 방식을 이용하여 추정한 후, 상기 HMM 파라미터 값 및 야코비안 행렬 값을 이용하여 음성을 인식하는 음성인식방법 및 장치를 제공하는 것을 그 목적으로 한다.

상기한 본 발명에 따르는 음성 인식 방법은, 바움-웰치 방식에 따라 이전 HMM 파라미터 값이 훈련 음성 데이터에 대해 유사도가 증가되도록 반복 추정하여 기준 HMM 파라미터 값을 추정하는 단계; 상기 기준 HMM 파라미터 값과 상기 훈련 음성 데이터에 따라 야코비안 행렬 값을 추정하는 단계; 인식 대상인 음성데이터의 잡음 스펙트럼을 추정하는 단계; 상기 잡음 스펙트럼과 상기 야코비안 행렬 값을 이용하여 음성 특징을 근사화하는 단계; 상기 음성 특징의 근사화 결과에 따라 HMM 파라미터 값을 재추정하는 단계; 상기 재추정된 HMM 파라미터 값에 따라 상기 인식 대상인 음성 데이터를 비터비 디코딩하여 최종 인식 결과를 출력하는 단계로 구성됨을 특징으로 한다.

대표도



특허청구의 범위

청구항 1

음성 인식 방법에 있어서,

바움-웰치 방식에 따라 이전 HMM 파라미터 값이 훈련 음성 데이터에 대해 유사도가 증가되도록 반복 추정하여 기준 HMM 파라미터 값을 추정하는 단계;

상기 기준 HMM 파라미터 값과 상기 훈련 음성 데이터에 따라 야코비안 행렬 값을 추정하는 단계;

인식 대상인 음성데이터의 잡음 스펙트럼을 추정하는 단계;

상기 잡음 스펙트럼과 상기 야코비안 행렬 값을 이용하여 음성 특징을 근사화하는 단계;

상기 음성 특징의 근사화 결과에 따라 HMM 파라미터 값을 재추정하는 단계;

상기 재추정된 HMM 파라미터 값에 따라 상기 인식 대상인 음성 데이터를 비터비 디코딩하여 최종 인식 결과를 출력하는 단계

로 구성됨을 특징으로 하는 음성 인식 방법.

청구항 2

제1항에 있어서,

상기 야코비안 행렬 값은 바움-웰치 방식에 따라 이전 야코비안 행렬 값이 상기 훈련 음성 데이터에 대한 유사도가 증가되도록 반복 추정됨을 특징으로 하는 음성 인식 방법.

청구항 3

제1항 또는 제2항에 있어서,

상기 바움-웰치 방식에 따른 유사도 증가의 반복 수행은 이전 추정값과 현재 추정값의 변화율이 미리 정해둔 변화율이하가 될 때까지 수행됨을 특징으로 하는 음성 인식 방법.

청구항 4

음성 인식 장치에 있어서,

바움-웰치 방식에 따라 이전 HMM 파라미터 값이 훈련 음성 데이터에 대해 유사도가 증가되도록 반복 추정하여 기준 HMM 파라미터 값을 추정하는 HMM 파라미터 값 추정부;

상기 기준 HMM 파라미터 값과 상기 훈련 음성 데이터에 따라 야코비안 행렬 값을 추정하는 야코비안 행렬 값 추정부;

인식 대상인 음성 데이터의 잡음 스펙트럼을 추정하는 잡음 스펙트럼 추정부;

상기 잡음 스펙트럼과 상기 야코비안 행렬 값을 이용하여 음성 특징을 근사화하는 음성 특징 근사화부;

상기 음성 특징의 근사화 결과에 따라 HMM 파라미터 값을 재추정하는 HMM 파리미터 값 재추정부;

상기 재추정된 HMM 파라미터 값에 따라 상기 인식 대상인 음성 데이터를 비터비 디코딩하여 최종 인식 결과를 출력하는 비터비 디코딩부

로 구성됨을 특징으로 하는 음성 인식 장치.

청구항 5

제4항에 있어서,

상기 야코비안 행렬 값 추정부는 바움-웰치 방식에 따라 이전 야코비안 행렬 값이 상기 훈련 음성 데이터에 대한 유사도가 증가되도록 반복 추정함을 특징으로 하는 음성 인식 장치.

청구항 6

제4항 또는 제5항에 있어서,

상기 HMM 파라미터 값 재추정부 및 상기 야코비안 행렬 값 추정부는,

이전 추정값과 현재 추정값의 변화율이 미리 정해둔 변화율이하가 될 때까지 반복 추정을 수행함을 특징으로 하는 음성 인식 장치.

명 세 서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

<13> 본 발명은 음성 인식 기술에 관한 것으로, 더욱 상세하게는 JA를 이용한 음성인식방법 및 장치에 관한 것이다.

<14> 일반적으로 잡음이 포함된 음성 데이터로부터 유효한 음성을 인식하는 방법으로는 음질 향상 기법, 잡음에 강인한 특징 추출 기법, 인식 모델 보상 방법 등이 있다.

<15> 히든 마르코프 모델(Hidden Markov model; HMM)에 기반한 인식 모델 보상방식은 잡음이 포함된 음성 데이터(이하, 잡음음성 데이터라 칭함)로부터 추출한 잡음의 통계정보를 이용하여 HMM 파라미터 값에 대한 보상을 이행하는 것이다.

<16> 특히, 야코비안 어뎁테이션(Jacobian Adaptation;JA) 방식은 실제 환경과 비슷한 조건에서 HMM을 미리 훈련시켜 매우 효과적으로 음성을 인식할 수 있는 것으로 알려져 있다.

<17> 상기한 JA 방식은 기준 HMM 파라미터 값을 설정하고, 상기 기준 HMM의 파라미터 값을 야코비안 행렬 값을 이용하여 실제 환경의 잡음 음성에 적응되도록 수정하는 것이다. 상기한 기준 HMM의 파라미터 값은 PMC(parallel model combination)나 NOVO 등과 같은 음성과 잡음의 모델 결합 방식을 이용하여 획득된다.

<18> 상기한 바와 같이 종래에는 모델 결합 방식을 이용함으로서, 원래의 깨끗한 음성 데이터에 대한 HMM의 평균 벡터 값을 이용하여 야코비안 행렬 값을 쉽게 추정할 수 있는 장점이 있다.

<19> 상술한 종래 기술에 따른 JA 방식의 음성 인식 장치를 도 1을 참조하여 설명한다.

<20> 상기 음성 인식 장치는 크게 음성 인식부(100)와 메모리부(110)로 구성된다.

<21> 상기 메모리부(110)는 깨끗한 음성 데이터에 따른 기준 파라미터 값을 저장하여 음성 인식부(100)의 각부로 제공한다.

<22> 상기 음성 인식부(100)는 잡음신호 스펙트럼 추정부(102), 야코비안 행렬 값 산출부(104), HMM 파라미터 값 조정부(106), 비터비 복호부(108)로 구성된다.

<23> 상기 스펙트럼 추정부(102)는 잡음음성 데이터를 입력받아 잡음신호를 추출하여 잡음신호의 스펙트럼을 추정한다.

<24> 상기 잡음신호의 스펙트럼 추정정보는 야코비안 행렬 값 산출부(104)에 입력되며, 상기 야코비안 행렬 값 산출부(104)는 메모리부(110)에 저장된 기준 HMM 파라미터 값을 메모리부(110)로부터 독취하여, 상기 기준 HMM 파라미터 값과 상기 잡음신호의 스펙트럼 정보를 해석적으로 결합하여 야코비안 행렬 값을 산출한다.

<25> 상기 산출된 야코비안 행렬 값과 상기 기준 HMM 파라미터 값은 HMM 파라미터 값 조정부(106)에 입력된다.

<26> 상기 HMM 파라미터 값 조정부(106)는 상기 야코비안 행렬 값에 따라 상기 기준 HMM 파라미터 값을 수정하여, 비터비 디코딩(Viterbi decoding)부(108)에 제공한다.

<27> 상기 비터비 디코딩부(108)는 상기 수정된 HMM 파라미터 값을 이용하여 잡음음성신호에 대해 비터비 디코딩을 이행하여 최종 음성인식결과를 생성한다.

<28> 그러나 상기한 바와 같은 종래 인식 장치는 해석적 모델 결합방식을 이용하여 상기 HMM 파라미터 값을 획득하므로, 실제 환경에 적합하게 훈련될 수 없으므로, 실사용시에는 그 인식 성능을 보장할 수 없는 문제가 있었다.

발명이 이루고자 하는 기술적 과제

<29> 본 발명은 상술한 종래 기술의 문제를 해소하기 위한 것으로, HMM 파라미터 값을 훈련 음성 데이터를 이용해서 바움 웰치(Baum-Welch) 방식으로 직접 훈련시킴과 아울러, 야코비안 행렬 값도 바움 웰치 방식을 이용하여 추정한 후, 상기 HMM 파라미터 값 및 야코비안 행렬 값을 이용하여 음성을 인식하는 음성인식방법 및 장치를 제공하는 것을 그 목적으로 한다.

발명의 구성 및 작용

<30> 상기한 목적을 달성하기 위한 본 발명에 따르는 음성 인식 방법은, 바움-웰치 방식에 따라 이전 HMM 파라미터 값이 훈련 음성 데이터에 대해 유사도가 증가되도록 반복 추정하여 기준 HMM 파라미터 값을 추정하는 단계; 상기 기준 HMM 파라미터 값과 상기 훈련 음성 데이터에 따라 야코비안 행렬 값을 추정하는 단계; 인식 대상인 음성데이터의 잡음 스펙트럼을 추정하는 단계; 상기 잡음 스펙트럼과 상기 야코비안 행렬 값을 이용하여 음성 특징을 근사화하는 단계; 상기 음성 특징의 근사화 결과에 따라 HMM 파라미터 값을 재추정하는 단계; 상기 재추정된 HMM 파라미터 값에 따라 상기 인식 대상인 음성 데이터를 비터비 디코딩하여 최종 인식 결과를 출력하는 단계로 구성됨을 특징으로 한다.

<31> 이하, 본 발명의 바람직한 실시예에 따른 데이터 기반의 JA 방식을 이용한 음성 인식 방법을 첨부된 도면을 참조하여 상세히 설명하기로 한다.

<32> 도 2는 본 발명에 따른 데이터 기반의 JA 방식을 이용한 음성 인식 장치의 구성도를 도시한 것이며, 이하 도 1에서 전술된 구성과 동일한 구성에 대해서는 그 상세한 설명을 생략한다.

<33> 상기 음성 인식 장치는 음성 인식부(200)와 메모리부(216)로 구성된다.

<34> 상기 메모리부(216)에는 초기 HMM 파라미터 값 및 이전에 추정된 HMM 파라미터 값을 저장하여 상기 음성 인식부(200)의 각 부에 제공한다.

<35> 상기 음성 인식부(200)의 HMM 파라미터 값 추정부(202)는 훈련 음성 데이터를 입력받아 메모리부(216)에 저장된 초기 HMM 파라미터 또는 이전 HMM 파라미터 값을 독출하고, 바움-웰치 방식에 따라 이전 HMM 파라미터 값이 훈련 음성 데이터에 대해 유사도가 증가되도록 반복 추정하여 기준 HMM 파라미터 값을 추정하여 야코비안 행렬값 추정부(206)에 제공한다.

<36> 제1잡음신호 스펙트럼 추정부(204)는 훈련 음성 데이터로부터 잡음신호 스펙트럼을 추정하여 야코비안 행렬값 추정부(206)는 이전 야코비안 행렬값이 잡음 신호 스펙트럼과 기준 HMM 파라미터 값에 대해 유사도가 증가하도록 반복 추정함으로써, 최종 야코비안 행렬값을 추정한다. 상기 추정된 야코비안 행렬값은 음성 특징 근사화부(210)에 제공된다.

<37> 제2잡음신호 스펙트럼 추정부(208)는 인식 대상이 되는 음성 데이터를 입력받아 잡음신호 스펙트럼을 추정하고, 그 결과를 음성특징 근사화부(210)에 제공한다.

<38> 상기 음성특징 근사화부(210)는 추정된 잡음신호 스펙트럼 값과 야코비안 행렬값을 이용하여 테일러(TAYLOR) 전개식을 이용하여 음성특징의 근사화를 이행한다.

<39> 상기 음성특징의 근사화 결과는 HMM 파라미터 재추정부(212)에 제공된다.

<40> 상기 HMM 파라미터 재추정부(212)는 음성 특징의 근사화를 나타내는 수식에서 잡음 음성을 제외한 나머지 변수들에게 시간에 따른 평균을 취함으로써 HMM 파라미터를 재추정한다. 상기 재추정된 HMM 파라미터는 비터비 디코딩부(214)에 제공된다.

<41> 상기 비터비 디코딩부(214)는 상기 재추정된 HMM 파라미터를 이용하여 음성 데이터에 대해 비터비 디코딩을 이행하여 음성인식결과를 출력한다.

<42> 상기한 바와 같이 본 발명은 기준 HMM 파라미터 값과 야코비안 행렬 값을 바움-웰치 방식을 이용하여 직접 추정한다.

<43> 상기한 바움-웰치 방식은 HMM 훈련 과정에서 주어진 음성 신호로부터 음성 신호에 해당하는 HMM 파라미터 값을

추정하는 것으로, 주어진 훈련 음성신호를 이용하여 반복적으로 새로운 HMM 파라미터 값을 추정하는 것이다.

<44> 이때 추정된 HMM의 파라미터 값은 주어진 음성 신호에 대한 유사도(likelihood) 값이 이전에 추정된 HMM 파라미터 값에 비해서 증가되도록 구성된다. 이러한 계속적인 유사도 증가를 반복적으로 수행하는 바움-웰치 과정을 통해서 추정된 HMM 파라미터 값이 주어진 음성신호에 대해서 최대의 유사도 값을 갖도록 할 수 있다.

<45> 이러한 반복적 추정 과정을 위해서는 초기 HMM 파라미터 값이 주어져야 하며, 상기 반복과정을 끝내기 위해서는 새로이 추정된 HMM 파라미터 값과 이전 추정치간의 변화율을 매 반복 과정마다 검사하여, 이 변화율이 주어진 기준치보다 작을 경우 반복과정을 종료하며 마지막으로 추정된 HMM 파라미터 값이 최종 HMM 파라미터 값이 된다.

<46> 상기 추정되는 기준 HMM 파라미터 값에는 HMM의 각 상태(state)의 평균값과 분산값 등이 포함되며, 본 발명에서는 HMM의 각 상태마다의 야코비안 행렬값도 바움-웰치 방식을 이용하여 추정한다.

<47> 여기서, 본 발명에 따른 바움-웰치 방식의 HMM 추정방법을 도 3의 흐름도를 참조하여 설명한다.

<48> 먼저, 훈련 음성 데이터가 입력되면(300단계), HMM 파라미터 값 추정부(202)는 상기 훈련 음성 데이터에 따라 HMM 파라미터 값을 재추정한다(302단계).

<49> 이후 HMM 파라미터 값 추정부(202)는 메모리부(216)에 저장된 초기 HMM 파라미터 값 또는 이전 HMM 파라미터 값과의 변화율을 산출한다(304단계).

<50> 이후, HMM 파라미터 값 추정부(202)는 상기 산출된 변화율이 미리 정해둔 소정치 이하인지를 체크하고(306단계), 상기 산출된 변화율이 상기 소정치 이상이면 상기 302단계로 복귀하여 HMM 파라미터 값을 다시 추정한다. 이러한 과정을 반복하면서 상기 HMM 파라미터 값 추정부(202)는 훈련 음성 데이터와 근사화된 HMM 파라미터 값을 추정하게 된다.

<51> 상기한 바와 다르게 상기 산출된 변화율이 상기 소정치 이하이면, 상기 HMM 파라미터 값 추정부(202)는 HMM 파라미터 값을 기준 HMM 파라미터 값으로 설정하고 프로세스를 종료한다(308단계).

발명의 효과

<52> 상기한 바와 같이 본 발명은 잡음 음성 인식을 위해서 HMM 파라미터 값을 재추정할 경우에 기준 HMM 파라미터 값과 야코비안 행렬 값을 데이터 기반의 바움-웰치 방식을 이용하여 직접 추정함으로써 보다 신뢰성 있는 통계정보를 얻을 수 있다.

<53> 좀더 설명하면, JA 방식에서는 기준 HMM 파라미터 값을 수정하여 인식시에 사용되는 HMM 파라미터 값을 얻게 되는데, 이때 중요한 것은 신뢰성있는 기준 HMM 파라미터 값을 얻는 것이다.

<54> 그런데, 깨끗한 음성에 대한 HMM 파라미터 값과 잡음 신호의 스펙트럼 값을 해석적으로 결합하여 얻은 HMM 파라미터 값 보다는 잡음음성을 이용하여 바움-웰치방식으로 직접 추정하여 얻은 HMM 파라미터 값이 인식 성능을 향상시킬 수 있으며, 이는 컴퓨터 시뮬레이션 결과를 통해서 확인할 수 있었다.

<55> 이에 따라 본 발명은 보다 우수한 성능을 가지는 기준 HMM 파라미터의 이용을 가능하게 함으로써, 결과적으로 잡음 음성 인식 성능의 향상을 야기할 수 있다.

<56> 이상과 같이 본 발명은 비록 한정된 실시예와 도면에 의해 설명되었으나, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 이는 본 발명이 속하는 분야에서 통상의 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다.

<57> 따라서, 본 발명 사상은 아래에 기재된 특허청구범위에 의해서만 파악되어야 하고, 이의 균등 또는 등가적 변형 모두는 본 발명 사상의 범주에 속한다고 할 것이다.

도면의 간단한 설명

<1> 도 1은 종래 기술에 따른 JA를 이용한 음성인식장치의 구성도.

<2> 도 2는 본 발명의 바람직한 실시예에 따른 JA를 이용한 음성인식장치의 구성도.

<3> 도 3은 본 발명의 바람직한 실시예에 따른 바움-웰치 방식의 HMM 파라미터 값 추정방법의 흐름도.

<4> <도면의 주요 부분에 대한 부호의 설명>

<5> 200 : 음성인식부

<6> 202 : HMM 파라미터 값 추정부

<7> 204 : 제1잡음신호 스펙트럼 추정부

<8> 206 : 야코비안 행렬값 추정부

<9> 208 : 제2잡음신호 스펙트럼 추정부

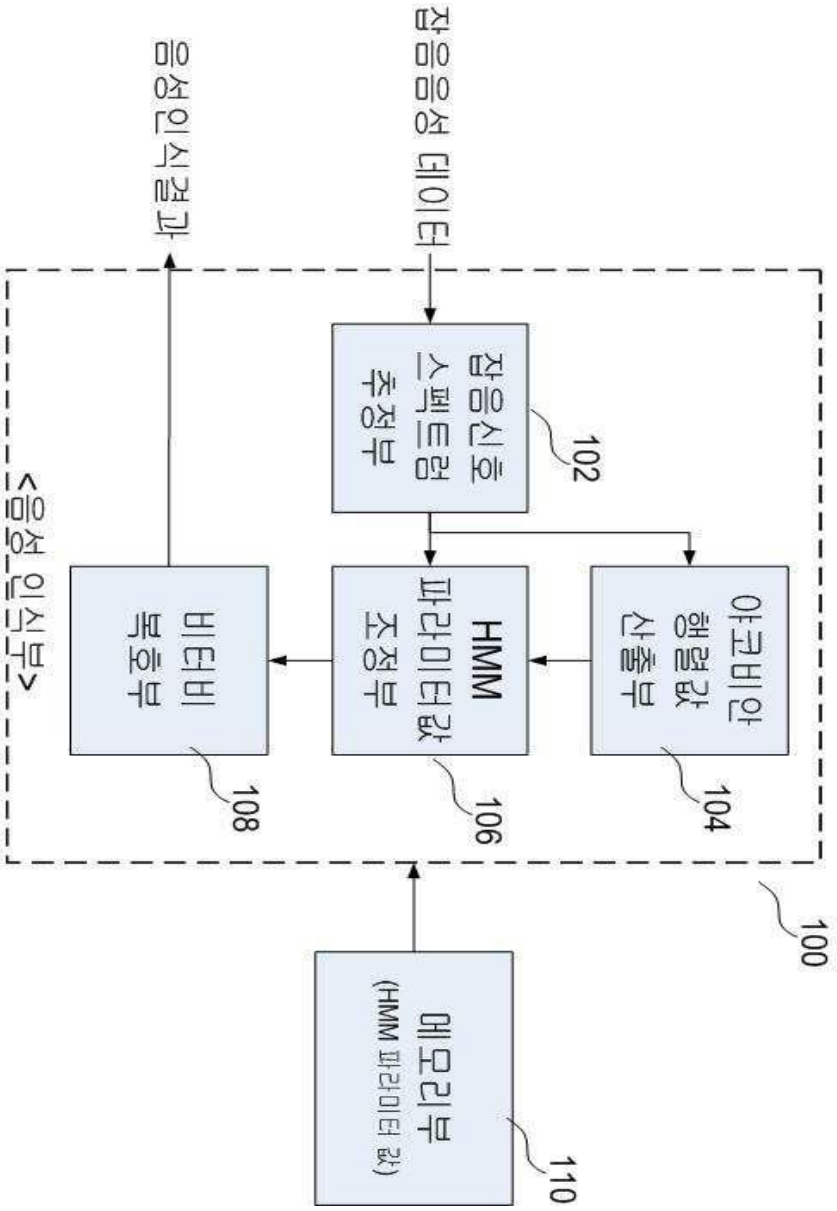
<10> 210 : 음성특징 근사화부

<11> 212 : HMM 파라미터 값 재추정부

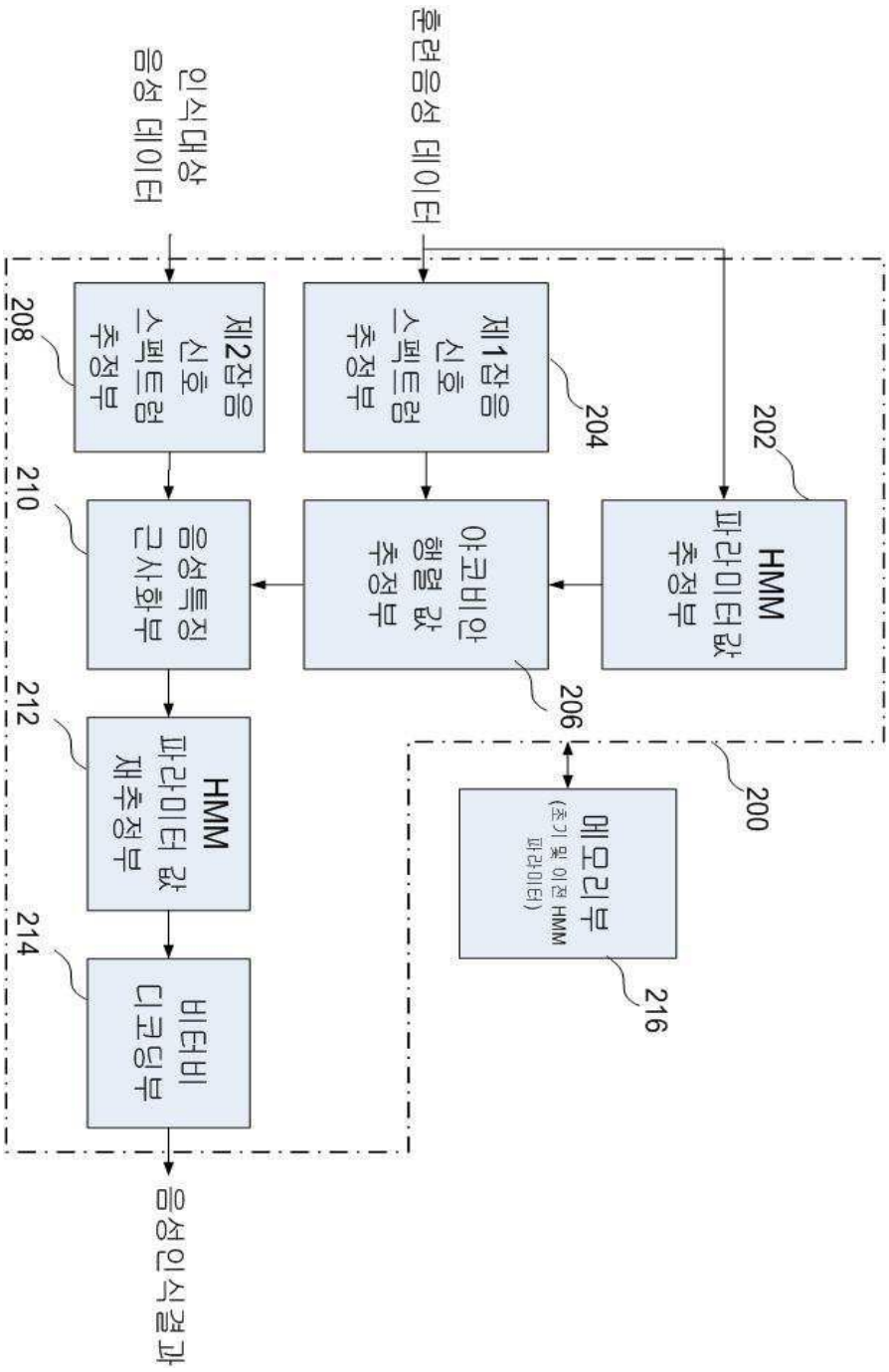
<12> 214 : 비터비 디코딩부

도면

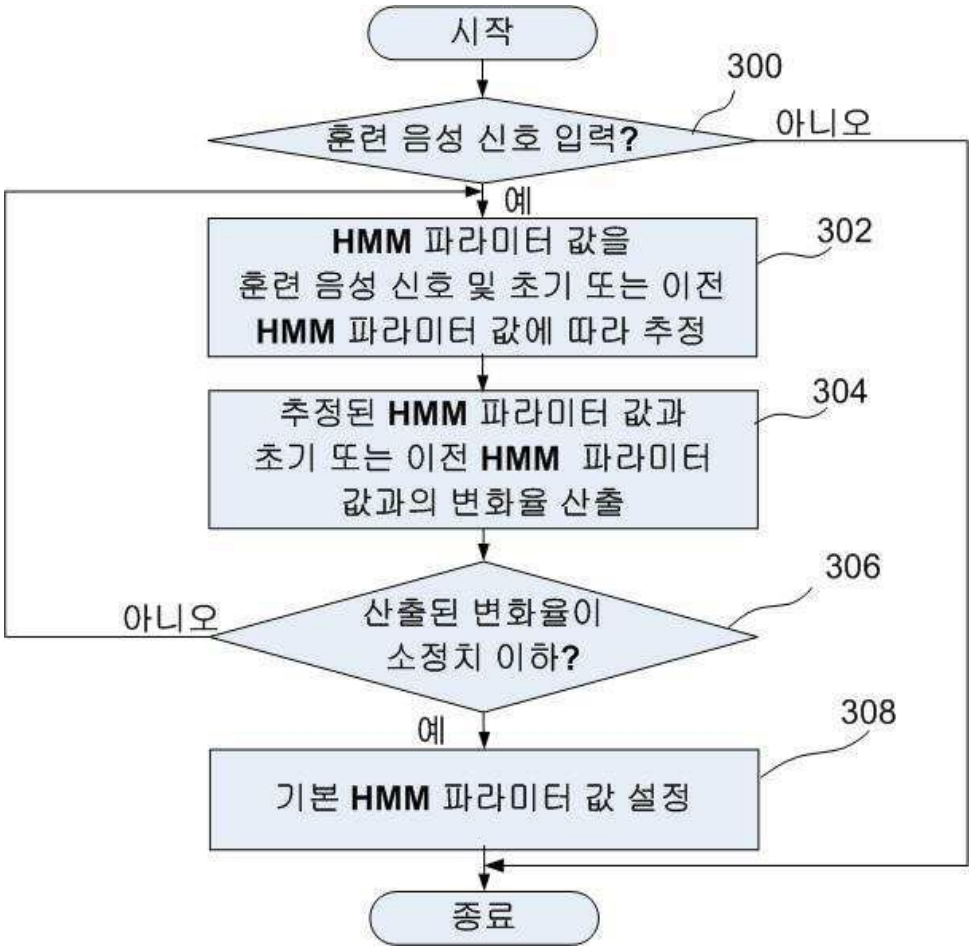
도면1



도면2


훈련음성 데이터
인식대상 음성 데이터
제1잡음 신호 스펙트럼 추정부
204
제2잡음 신호 스펙트럼 추정부
208
HMM 파라미터값 추정부
202
야코비안 행렬 값 추정부
206
음성특징 근사화부
210
HMM 파라미터 값 재추정부
212
비터비 디코딩부
214
메모리부
(초기 및 이전 HMM 파라미터)
216
200
음성인식결과

도면3


시작
300
훈련 음성 신호 입력?
아니오
예
HMM 파라미터 값을
훈련 음성 신호 및 초기 또는 이전
HMM 파라미터 값에 따라 추정
302
추정된 HMM 파라미터 값과
초기 또는 이전 HMM 파라미터
값과의 변화율 산출
304
306
아니오
산출된 변화율이
소정치 이하?
예
308
기본 HMM 파라미터 값 설정
종료