



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2011년09월19일  
(11) 등록번호 10-1065188  
(24) 등록일자 2011년09월08일

(51) Int. Cl.  
G10L 15/14 (2006.01) G10L 15/06 (2006.01)  
(21) 출원번호 10-2009-0067685  
(22) 출원일자 2009년07월24일  
심사청구일자 2009년07월24일  
(65) 공개번호 10-2011-0010233  
(43) 공개일자 2011년02월01일  
(56) 선행기술조사문헌  
KR1020050043472 A\*  
김남수, '잡음 환경에서의 음성 인식',  
Telecommunications Review, 제13권 5호, 2003  
년10월\*  
\*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
고려대학교 산학협력단  
서울 성북구 안암동5가 1  
(72) 발명자  
육동석  
서울 용산구 서빙고동 신동아아파트 1동 102호  
이협우  
경기도 안양시 동안구 갈산동 샘마을대우아파트  
101-1002  
김동현  
서울 동작구 신대방1동 687-7  
(74) 대리인  
특허법인충현

전체 청구항 수 : 총 19 항

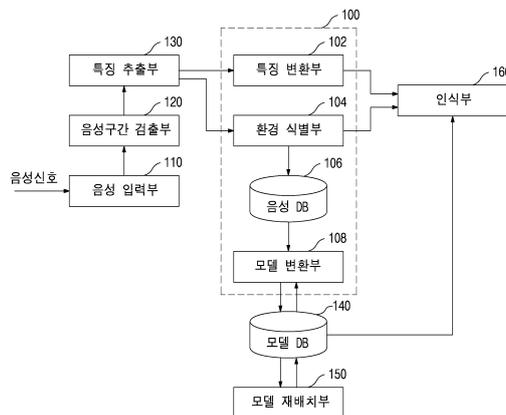
심사관 : 정성윤

(54) 진화 학습에 의한 화자 적응 장치 및 방법과 이를 이용한 음성인식 시스템

(57) 요약

본 발명은 음성인식 시스템에서의 화자 적응 기술에 관한 것으로서, 본 발명에 따른 진화 학습에 의한 화자 적응 장치는, 음성인식 시스템이 음성인식을 수행하고 있는 인식 모드에서, 미리 결정된 환경 파라미터를 이용하여 인식 대상 음성데이터의 특징벡터에 대해 특징 변환을 수행하는 특징 변환부; 상기 인식 대상 음성데이터를 저장하는 음성 데이터베이스; 및 상기 음성인식 시스템이 음성인식을 수행하지 않고 있는 대기 모드에서, 상기 음성 데이터베이스에 저장된 음성데이터를 이용하여 기존 음향모델에 대한 모델 변환을 수행하는 모델 변환부를 포함하는 것을 특징으로 하여, 사용자 편의성 및 음성인식 시스템의 인식 성능을 개선함은 물론, 화자 적응 및 환경 적응을 동시에 수행하는 이점을 제공한다.

대표도 - 도1



## 특허청구의 범위

### 청구항 1

음성인식 시스템에서 진화 학습에 의해 화자 적응을 수행하는 장치에 있어서,

상기 음성인식 시스템이 음성인식을 수행하고 있는 인식 모드에서, 미리 결정된 환경 파라미터를 이용하여 인식 대상 음성데이터의 특징벡터에 대해 특징 변환을 수행하는 특징 변환부;

상기 인식 대상 음성데이터를 저장하는 음성 데이터베이스; 및

상기 음성인식 시스템이 음성인식을 수행하지 않고 있는 대기 모드에서, 상기 음성 데이터베이스에 저장된 음성 데이터를 이용하여 기존 음향모델에 대한 모델 변환을 수행하는 모델 변환부를 포함하고,

상기 음성 인식 시스템은 수행 모드에 따라 상기 특징 변환 및 상기 모델 변환을 반복함으로써 화자 적응 및 환경 적응을 동시에 수행하는 진화 학습에 의한 화자 적응 장치.

### 청구항 2

제1항에 있어서,

상기 특징 변환부는, 최대 우도 기법(maximum likelihood method)을 이용하여 상기 환경 파라미터를 미리 결정하는 것을 특징으로 하는 진화 학습에 의한 화자 적응 장치.

### 청구항 3

제1항에 있어서,

상기 화자 적응 장치는, 상기 음성인식 시스템에서 음성구간 검출(Voice Activity Detection)을 통해 검출되는 음성구간 이외의 비음성구간에서 주변잡음의 특징을 추출하여 상기 인식 대상 음성데이터가 발생한 환경을 식별하는 환경 식별부를 더 포함하는 것을 특징으로 하는 진화 학습에 의한 화자 적응 장치.

### 청구항 4

제3항에 있어서,

상기 음성 데이터베이스는, 상기 환경 식별부에 의해 식별된 환경별로 해당 환경에서 발생한 음성데이터를 저장하는 것을 특징으로 하는 진화 학습에 의한 화자 적응 장치.

### 청구항 5

제4항에 있어서,

상기 모델 변환부는, 상기 음성 데이터베이스에 저장된 음성데이터를 이용하여 상기 식별된 환경별로 상기 기존 음향모델에 대한 모델 변환을 수행하는 것을 특징으로 하는 진화 학습에 의한 화자 적응 장치.

### 청구항 6

음성인식 시스템에서 진화 학습에 의해 화자 적응을 수행하는 방법에 있어서,

상기 음성인식 시스템이 음성인식을 수행하고 있는 인식 모드에서 미리 결정된 환경 파라미터를 이용하여 인식 대상 음성데이터의 특징벡터에 대해 특징 변환을 수행하는 특징 변환 단계;

상기 인식 대상 음성데이터를 데이터베이스에 저장하는 음성데이터 저장 단계; 및

상기 음성인식 시스템이 음성인식을 수행하지 않고 있는 대기 모드에서 상기 데이터베이스에 저장된 음성데이터를 이용하여 기존 음향모델에 대한 모델 변환을 수행하는 모델 변환 단계를 포함하고,

상기 음성 인식 시스템은 수행 모드에 따라 상기 특징 변환 및 상기 모델 변환을 반복함으로써 화자 적응 및 환경 적응을 동시에 수행하는 진화 학습에 의한 화자 적응 방법.

### 청구항 7

제6항에 있어서,

상기 특징 변환 단계는, 최대 우도 기법(maximum likelihood method)을 이용하여 상기 환경 파라미터를 미리 결정하는 단계를 포함하는 것을 특징으로 하는 진화 학습에 의한 화자 적응 방법.

**청구항 8**

제6항에 있어서,

상기 화자 적응 방법은, 상기 음성인식 시스템이 음성구간 검출(Voice Activity Detection)을 통해 검출되는 음성구간 이외의 비음성구간에서 주변잡음의 특징을 추출하여 상기 인식 대상 음성데이터가 발생한 환경을 식별하는 환경 식별 단계를 더 포함하는 것을 특징으로 하는 진화 학습에 의한 화자 적응 방법.

**청구항 9**

제8항에 있어서,

상기 음성데이터 저장 단계는, 상기 환경 식별 단계를 통해 식별된 환경별로 해당 환경에서 발생한 음성데이터를 저장하는 단계인 것을 특징으로 하는 진화 학습에 의한 화자 적응 방법.

**청구항 10**

제9항에 있어서,

상기 모델 변환 단계는, 상기 데이터베이스에 저장된 음성데이터를 이용하여 상기 식별된 환경별로 상기 기존 음향모델에 대한 모델 변환을 수행하는 단계인 것을 특징으로 하는 진화 학습에 의한 화자 적응 방법.

**청구항 11**

제6항 내지 제10항 중 어느 한 항에 따른 방법을 컴퓨터로 실행하기 위한 프로그램이 기록된 컴퓨터 판독가능 기록매체.

**청구항 12**

진화 학습에 의해 화자 적응을 수행하여 음성을 인식하는 음성인식 시스템에 있어서,

상기 음성인식 시스템이 음성인식을 수행하고 있는 인식 모드에서, 인식 대상 음성데이터의 특징벡터에 대해 미리 결정된 환경 파라미터를 이용하여 특징 변환을 수행하는 특징 변환부;

상기 인식 대상 음성데이터를 저장하는 음성 데이터베이스;

상기 음성인식 시스템이 음성인식을 수행하지 않고 있는 대기 모드에서, 상기 음성 데이터베이스에 저장된 음성 데이터를 이용하여 기존 음향모델에 대한 모델 변환을 수행하는 모델 변환부; 및

상기 인식 모드에서, 상기 특징 변환부에 의해 변환된 특징벡터 및 상기 모델 변환부에 의해 변환된 음향모델을 이용하여 음성인식을 수행하는 인식부를 포함하고,

상기 음성 인식 시스템은 수행 모드에 따라 상기 특징 변환 및 상기 모델 변환을 반복함으로써 화자 적응 및 환경 적응을 동시에 수행하는 진화 학습에 의한 음성인식 시스템.

**청구항 13**

제12항에 있어서,

상기 특징 변환부는, 최대 우도 기법(maximum likelihood method)을 이용하여 상기 환경 파라미터를 미리 결정하는 것을 특징으로 하는 진화 학습에 의한 음성인식 시스템.

**청구항 14**

제12항에 있어서,

상기 음성인식 시스템은, 음성구간 검출(Voice Activity Detection)을 통해 검출되는 음성구간 이외의 비음성구간에서 주변잡음의 특징을 추출하여 상기 인식 대상 음성데이터가 발생한 환경을 식별하는 환경 식별부를 더 포

함하는 것을 특징으로 하는 진화 학습에 의한 음성인식 시스템.

**청구항 15**

제14항에 있어서,

상기 음성 데이터베이스는, 상기 환경 식별부에 의해 식별된 환경별로 해당 환경에서 발생한 음성데이터를 저장하는 것을 특징으로 하는 진화 학습에 의한 음성인식 시스템.

**청구항 16**

제15항에 있어서,

상기 모델 변환부는, 상기 음성 데이터베이스에 저장된 음성데이터를 이용하여 상기 식별된 환경별로 상기 기존 음향모델에 대한 모델 변환을 수행하는 것을 특징으로 하는 진화 학습에 의한 음성인식 시스템.

**청구항 17**

제16항에 있어서,

상기 음성인식 시스템은, 상기 모델 변환부에 의해 상기 식별된 환경별로 모델 변환된 환경별 음향모델을 저장하는 모델 데이터베이스를 더 포함하는 것을 특징으로 하는 진화 학습에 의한 음성인식 시스템.

**청구항 18**

제17항에 있어서,

상기 음성인식 시스템은, 미리 결정된 임계치에 의해 상기 모델 데이터베이스에 저장되는 음향모델의 개수를 제한하는 모델 재배치부를 더 포함하는 것을 특징으로 하는 진화 학습에 의한 음성인식 시스템.

**청구항 19**

제17항에 있어서,

상기 인식부는, 상기 모델 데이터베이스에 저장된 음향모델 중 상기 인식 대상 음성데이터가 발생한 환경에 대응하는 음향모델을 이용하여 음성인식을 수행하는 것을 특징으로 하는 진화 학습에 의한 음성인식 시스템.

**명세서**

**발명의 상세한 설명**

**기술분야**

[0001] 본 발명은 음성인식 시스템에서의 화자 적응 기술에 관한 것으로서, 더욱 상세하게는, 진화 학습에 의한 비지도 적응 방식을 통해 특징 변환 및 모델 변환을 모두 수행함으로써, 사용자 편의성 및 음성인식 시스템의 인식 성능을 개선하는 물론, 화자 적응 및 환경 적응을 동시에 수행하는 화자 적응 장치 및 방법과 이를 이용한 음성인식 시스템에 관한 것이다.

**배경기술**

[0002] 최근 음성인식 기술은 가정용 전자기기, 휴대폰, 보안 및 인증 등 다양한 분야에 적용되고 있으며, 그 적용분야 및 수요가 급속히 증가하고 있는 추세이다.

[0003] 여기서 음성인식(speech or voice recognition)이란, 자동적인 음성인식 시스템을 통해 인간으로부터 발화된 음성으로부터 언어적 의미, 내용을 식별하는 것으로서, 구체적으로는 음성 파형을 입력하여 단어나 단어열을 식별하고 처리하는 과정을 말한다.

[0004] 이러한 음성인식을 수행하는 음성인식 시스템(speech recognition system)들은 일반적으로, 단어 음성인식 시스템(word recognition system), 연속 음성인식 시스템(continuous speech recognition system) 및 화자 인식 시스템(speaker recognition system) 등으로 분류할 수 있으며, 상기 음성인식 시스템들에 공통적으로 적용되는 음성인식 알고리즘들은 음성구간 검출(Voice Activity Detection) 과정, 특징(feature) 추출 과정, 그리고 매

칭(matching) 과정을 포함하고 있다.

- [0005] 이와 같이, 음성인식 기술은 일종의 패턴인식 기술로서, 상기 매칭 과정을 통해 최종적으로 입력 음성데이터의 의미를 인식해 내기 위해서는, “훈련(training) 과정” 및 “테스트(test) 과정”을 필요로 한다.
- [0006] 상기 훈련 과정은 화자(speaker)로부터 특정 음성 샘플들을 수집하여 음향모델을 생성하는 과정이며, 상기 테스트 과정은 입력 음성데이터로부터 추출된 일련의 특징벡터들을 상기 생성된 음향모델의 패턴과 매칭시키는 과정이다. 따라서, 음성인식 시스템의 인식성능은, 상기 훈련 과정을 통해 생성된 음향모델이 테스트용 입력 음성 샘플들과의 패턴 매칭에 사용되는 경우 얼마만큼의 높은 정확도 내지 신뢰도를 나타낼 수 있는가에 달려있다.
- [0007] 한편, 상기한 음성인식 시스템에서 사용되는 음향모델에는, 화자 독립 음향모델과 화자 종속 음향모델이 있다. 화자 독립 음향모델은 다수의 화자로부터 발화된 음성데이터를 사용하여 훈련된 일반적 음향모델인 반면, 화자 종속 음향모델은 특정 화자로부터 발화된 음성데이터를 사용하여 훈련된 고유 음향모델이다.
- [0008] 화자 독립 음향모델의 경우, 광범위한 불특정 화자들에 대해서는 평균적으로 최상의 인식 성능을 나타낼 수 있도록 설계된다. 그러나, 상기 화자 독립 음향모델은 단일한 특정 화자에 대해서는 최상의 인식 성능을 발휘할 수 없는 문제점이 있다. 따라서, 일반적으로 단일한 특정 화자만 사용하는 핸드폰, PDA 등 개인용 모바일 기기에 적용되는 음성인식 시스템에서는 상기 특정 화자의 음성을 이용하여 화자 독립 음향모델을 화자 종속 음향모델로 변환시키는 화자 적응 방식이 필요하다.
- [0009] 기존의 화자 적응 방식으로는 지도 적응(Supervised Adaptation) 방식과 비지도 적응(Unsupervised Adaptation) 방식이 있다.
- [0010] 그러나, 상기 지도 적응 방식은, 실제 음성인식 서비스 환경에서 음성인식 수행 전 사용자로 하여금 특정 문장을 발성하도록 하거나 발화한 문장의 전사데이터(label 또는 transcription)를 입력받아야 하므로 사용자 편의성을 저해한다는 문제점이 있다.
- [0011] 또한, 상기 비지도 적응 방식은, 화자 종속 음향모델을 생성하기 위해 상당한 적응시간을 요구하며, 특히 적응을 위한 전사데이터를 음성인식 시스템 자체에서 예측해야 하는 결과, 부정확한 전사데이터로 인한 음성인식 성능 저하를 초래하게 되는 문제점이 있다.
- [0012] 한편, 실제로 음성인식을 수행하는 경우에는 잡음이 섞인 음성이 사용되므로 학습과 음성인식 테스트시에 환경이 일치하지 않아 음성인식률이 저하한다. 그러므로 음성인식 시스템에서 학습환경과 음성인식 환경을 일치시켜 잡음환경하에서 음성인식 성능을 높이는 기술이 개발되고 있다.
- [0013] 그러나, 기존의 화자 적응 기술은, 실제 잡음 환경에서 화자 적응 및 환경 적응을 동시에 효율적으로 수행할 수 있는 화자 적응 기술을 제시하지 못하고 있다는 문제점이 있다.

**발명의 내용**

**해결 하고자하는 과제**

- [0014] 따라서, 본 발명이 해결하고자 하는 첫 번째 기술적 과제는, 진화 학습에 의한 비지도 적응 방식을 통해 특징 변환 및 모델 변환을 모두 수행함으로써, 사용자 편의성 및 음성인식 시스템의 인식 성능을 개선함은 물론, 화자 적응 및 환경 적응을 동시에 수행하는 화자 적응 장치를 제공하는 것이다.
- [0015] 본 발명이 해결하고자 하는 두 번째 기술적 과제는, 진화 학습에 의한 비지도 적응 방식을 통해 특징 변환 및 모델 변환을 모두 수행함으로써, 사용자 편의성 및 음성인식 시스템의 인식 성능을 개선함은 물론, 화자 적응 및 환경 적응을 동시에 수행하는 화자 적응 방법을 제공하는 것이다.
- [0016] 본 발명이 해결하고자 하는 세 번째 기술적 과제는, 상기 화자 적응 장치 내지 상기 화자 적응 방법을 이용한 음성인식 시스템을 제공하는 것이다.

**과제 해결수단**

- [0017] 상기와 같은 첫 번째 기술적 과제를 해결하기 위하여 본 발명은, 음성인식 시스템에서 진화 학습에 의해 화자 적응을 수행하는 장치에 있어서, 상기 음성인식 시스템이 음성인식을 수행하고 있는 인식 모드에서, 미리 결정된 환경 파라미터를 이용하여 인식 대상 음성데이터의 특징벡터에 대해 특징 변환을 수행하는 특징 변환부; 상기 인식 대상 음성데이터를 저장하는 음성 데이터베이스; 및 상기 음성인식 시스템이 음성인식을 수행하지 않고

있는 대기 모드에서, 상기 음성 데이터베이스에 저장된 음성데이터를 이용하여 기존 음향모델에 대한 모델 변환을 수행하는 모델 변환부를 포함하는 진화 학습에 의한 화자 적응 장치를 제공한다.

- [0018] 일 실시예에 있어서, 상기 특징 변환부는, 최대 우도 기법(maximum likelihood method)을 이용하여 상기 환경 파라미터를 미리 결정한다.
- [0019] 일 실시예에 있어서, 상기 화자 적응 장치는, 상기 음성인식 시스템에서 음성구간 검출(Voice Activity Detection)을 통해 검출되는 음성구간 이외의 비음성구간에서 주변잡음의 특징을 추출하여 상기 인식 대상 음성 데이터가 발생한 환경을 식별하는 환경 식별부를 더 포함한다.
- [0020] 일 실시예에 있어서, 상기 음성 데이터베이스는, 상기 환경 식별부에 의해 식별된 환경별로 해당 환경에서 발생한 음성데이터를 저장한다.
- [0021] 일 실시예에 있어서, 상기 모델 변환부는, 상기 음성 데이터베이스에 저장된 음성데이터를 이용하여 상기 식별된 환경별로 상기 기존 음향모델에 대한 모델 변환을 수행한다.
- [0022] 상기와 같은 두 번째 기술적 과제를 해결하기 위하여 본 발명은, 음성인식 시스템에서 진화 학습에 의해 화자 적응을 수행하는 방법에 있어서, 상기 음성인식 시스템이 음성인식을 수행하고 있는 인식 모드에서 미리 결정된 환경 파라미터를 이용하여 인식 대상 음성데이터의 특징벡터에 대해 특징 변환을 수행하는 특징 변환 단계; 상기 인식 대상 음성데이터를 데이터베이스에 저장하는 음성데이터 저장 단계; 및 상기 음성인식 시스템이 음성인식을 수행하지 않고 있는 대기 모드에서 상기 데이터베이스에 저장된 음성데이터를 이용하여 기존 음향모델에 대한 모델 변환을 수행하는 모델 변환 단계를 포함하는 진화 학습에 의한 화자 적응 방법을 제공한다.
- [0023] 상기와 같은 세 번째 기술적 과제를 해결하기 위하여 본 발명은, 진화 학습에 의해 화자 적응을 수행하여 음성을 인식하는 음성인식 시스템에 있어서, 상기 음성인식 시스템이 음성인식을 수행하고 있는 인식 모드에서, 인식 대상 음성데이터의 특징벡터에 대해 미리 결정된 환경 파라미터를 이용하여 특징 변환을 수행하는 특징 변환부; 상기 인식 대상 음성데이터를 저장하는 음성 데이터베이스; 상기 음성인식 시스템이 음성인식을 수행하지 않고 있는 대기 모드에서, 상기 음성 데이터베이스에 저장된 음성데이터를 이용하여 기존 음향모델에 대한 모델 변환을 수행하는 모델 변환부; 및 상기 인식 모드에서, 상기 특징 변환부에 의해 변환된 특징벡터 및 상기 모델 변환부에 의해 변환된 음향모델을 이용하여 음성인식을 수행하는 인식부를 포함하는 진화 학습에 의한 음성인식 시스템을 제공한다.
- [0024] 일 실시예에 있어서, 상기 특징 변환부는, 최대 우도 기법(maximum likelihood method)을 이용하여 상기 환경 파라미터를 미리 결정한다.
- [0025] 일 실시예에 있어서, 상기 음성인식 시스템은, 음성구간 검출(Voice Activity Detection)을 통해 검출되는 음성구간 이외의 비음성구간에서 주변잡음의 특징을 추출하여 상기 인식 대상 음성데이터가 발생한 환경을 식별하는 환경 식별부를 더 포함한다.
- [0026] 일 실시예에 있어서, 상기 음성 데이터베이스는, 상기 환경 식별부에 의해 식별된 환경별로 해당 환경에서 발생한 음성데이터를 저장한다.
- [0027] 일 실시예에 있어서, 상기 모델 변환부는, 상기 음성 데이터베이스에 저장된 음성데이터를 이용하여 상기 식별된 환경별로 상기 기존 음향모델에 대한 모델 변환을 수행한다.
- [0028] 일 실시예에 있어서, 상기 음성인식 시스템은, 상기 모델 변환부에 의해 상기 식별된 환경별로 모델 변환된 환경별 음향모델을 저장하는 모델 데이터베이스를 더 포함한다.
- [0029] 일 실시예에 있어서, 상기 음성인식 시스템은, 미리 결정된 임계치에 의해 상기 모델 데이터베이스에 저장되는 음향모델의 개수를 제한하는 모델 재배치부를 더 포함한다.
- [0030] 일 실시예에 있어서, 상기 인식부는, 상기 모델 데이터베이스에 저장된 음향모델 중 상기 인식 대상 음성데이터가 발생한 환경에 대응하는 음향모델을 이용하여 음성인식을 수행한다.

**효 과**

- [0031] 본 발명은, 음성인식 시스템의 화자 적응을 위해 진화 학습에 의한 비지도 적응 방식을 적용함으로써 사용자 편의성을 도모하는 이점을 제공한다.
- [0032] 또한, 음성인식 시스템의 동작 상태에 따라 특징 변환 및 모델 변환을 적절하게 수행함으로써, 화자 적응 방식

의 효율성 및 음성인식 시스템의 인식 성능을 개선하는 이점을 제공한다.

[0033] 나아가, 음성인식 환경별로 화자 종속 음향모델을 생성함으로써, 화자 적응은 물론 환경 적응까지 동시에 수행하는 이점을 제공한다.

**발명의 실시를 위한 구체적인 내용**

[0034] 본 발명에 관한 구체적인 내용의 설명에 앞서 이해의 편의를 위해 본 발명의 기술적 개요를 우선 설명한다.

[0035] 화자 내지 환경 적응 기법의 기본적인 목적은, 음성인식 시스템의 음향모델 생성시 사용되는 불특정 다수 화자의 잡음없는 훈련 데이터와 실제 테스트시 사용되는 특정 화자의 잡음 섞인 음성데이터 간의 불일치로 인한 음성인식 성능 저하를 방지하고자 하는 것이다.

[0036] 이러한 적응 기법은, 적응 변환을 수행하는 대상에 따라 “특징 변환”에 의한 적응 기법과 “모델 변환”에 의한 적응 기법으로 분류할 수 있다.

[0037] 상기 특징 변환 적응 기법은, 특징벡터 보상(feature compensation)을 통해 실제 음성데이터에서 잡음을 제거하고 깨끗한 음성데이터로 매핑하여 음성인식을 수행하는 기법으로서, 비교적 적은 연산량을 통해 순간적인 환경 변화에 신속하게 적응할 수 있는 장점이 있다.

[0038] 그러나, 상기 특징 변환 적응 기법은, 상기 모델 변환 적응 기법에 비해 인식성능의 향상 정도가 제한된다는 단점이 있다.

[0039] 상기 모델 변환 적응 기법은, 적응 데이터를 이용한 모델 보상(model compensation)을 통해 기존에 학습된 음향 모델을 실제 음성인식 환경에 종속된 음향모델로 적응시켜 음성인식을 수행하는 기법으로서, 인식성능 향상률이 크다는 장점이 있다.

[0040] 그러나, 상기 모델 변환 적응 기법은, 특정 적응 데이터와 관련된 모델 변수만 각각 변환하기 때문에 상대적으로 많은 양의 적응 데이터 및 연산시간을 필요로 하는 단점이 있다.

[0041] 따라서, 본 발명은 음성인식에 있어서 상기 양 기법을 효과적으로 모두 적용함과 동시에 각각의 단점을 보완하는 새로운 화자 적응 기법을 제공한다. 즉, 음성인식 시스템이 음성인식을 수행하고 있는 인식 모드에서, 신속한 환경 적응이 가능한 특징 변환을 수행하도록 하고(온라인 적응), 상기 음성인식 시스템이 음성인식을 수행하지 않고 있는 대기 모드에서, 미리 저장해 둔 충분한 양의 음성데이터를 적응 데이터로 이용하여 모델 변환을 수행하도록 한다(오프라인 적응). 이와 같이, 본 발명은, 화자 적응을 위한 진화 학습(evolutional learning) 과정으로서 특징 변환 및 모델 변환을 반복적으로 수행하도록 하여 축적된 적응 데이터의 양과 무관하게 안정적이고 높은 인식성능을 보장한다.

[0042] 이하, 본 발명의 기술적 과제의 해결 방안을 명확화하기 위해 첨부도면을 참조하여 본 발명의 바람직한 실시예를 상세하게 설명한다. 다만, 본 발명을 설명함에 있어서 관련 공지기술에 관한 설명이 오히려 본 발명의 요지를 불명료하게 할 수 있다고 판단되는 경우 그에 관한 설명을 생략하기로 한다. 또한, 후술하는 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.

[0043] 도 1에는 본 발명에 따른 진화 학습에 의한 화자 적응 장치(100)가 적용된 음성인식 시스템의 일례가 블록도로 도시되어 있다.

[0044] 도 2에는 본 발명에 따른 진화 학습에 의한 화자 적응 방법의 일례가 흐름도로 도시되어 있다.

[0045] 도 1 및 도 2를 참조하면, 본 발명에 따른 화자 적응 장치(100)는, 특징 변환부(102), 음성 데이터베이스(106) 및 모델 변환부(108)를 포함하며, 환경 식별부(104)를 더 포함할 수 있다.

[0046] 우선, 상기 음성인식 시스템은, 초기에 훈련(training) 단계를 통해 화자 및 환경 독립 음향모델을 생성하여 저장하고 있다(S210). 이러한 화자 및 환경 독립 음향모델을 특정 화자 및 환경에 적합한 모델로 재추정하는 것이 바로 적응 기법이다.

[0047] 도 3에는 화자 적응 기법의 기본원리가 도시되어 있다.

[0048] 도 3을 참조하면, 화자 적응 기법은 화자 독립 음향모델을 추정하는 과정(300)과 이를 통해 추정된 화자 독립 음향모델을 화자 종속 음향모델로 재추정하는 과정(310)으로 이루어진다. 즉, 불특정 다수의 화자로부터 얻어진

많은 양의 음성데이터에서 MFCC(Mel-Frequency Cepstral Coefficient)와 같은 다차원의 특징벡터를 추출하고, 추출된 특징벡터들에 대해 EM(Expectation-Maximization) 기법 등을 이용하여 HMM(Hidden Markov Model)을 추정한다. 이렇게 추정된 HMM은 불특정 다수의 화자에 대한 음향모델로서 화자 독립적인 특성을 지닌다. 한편, 상기 HMM의 훈련에 사용된 음성데이터가 다양한 여러 환경(다양한 종류의 음향기기, 잡음환경 등)에서 수집된 것이라면 상기 HMM은 동시에 환경 독립적인 특성도 지니게 된다.

[0049] 이와 같이 훈련된 화자 독립 음향모델을 이용하여 특정 사용자가 음성인식을 수행하면, 상기 특정 사용자의 음성 특성을 반영하지 못하기 때문에 만족할만한 인식성능을 나타낼 수 없다. 따라서, 핸드폰, PDA 등과 같은 개인용 모바일 기기에서 음성인식 기반 기능을 원활하게 사용하기 위해서는 특정 사용자의 음향 특성에 최적화된 음향모델로 재추정(적응)할 필요가 있다. 아래에서 다시 설명하겠지만, 본 발명은 새로운 적응 방식을 도입하여 이미 추정된 화자 및 환경 독립 음향모델과 특정 화자(사용자)로부터 획득되는 상대적으로 적은 양의 음성데이터를 이용하여 화자 및 환경 종속 음향모델을 추정한다.

[0050] 상기 특징 변환부(102)는, 상기 음성인식 시스템이 음성인식을 수행하고 있는 인식 모드에서, 인식 대상 음성데이터 입력되면 미리 결정된 환경 파라미터를 이용하여 상기 인식 대상 음성데이터의 특징벡터에 대해 특징 변환을 수행한다(S220, S230). 즉, 음성인식 시스템이 적용된 핸드폰, PDA, 기타 모바일 기기 등에서 사용자의 음성 명령을 받아들이는 작업을 수행하는 동안 신속한 환경 적응이 가능한 특징 변환을 수행한다(온라인 적응).

[0051] 더욱 구체적으로 설명하면, 특징 변환이란, 입력된 음성데이터에서 추출한 특징벡터들을 미리 예측한 환경 파라미터들을 이용하여 변환하는 방법이다. 즉, 특징 변환 과정은 기본적으로 수학적 식 1과 같이 나타낼 수 있다.

**수학적 식 1**

[0052] 
$$X' = A \cdot X + B$$

[0053] 상기 수학적 식 1에서,  $X$ 는 입력 음성데이터의 특징벡터,  $A$  및  $B$ 는 미리 결정된 환경 파라미터를 나타낸다. 결국, 특징 변환은 미리 결정된 환경 파라미터  $A$ ,  $B$ 를 이용하여 입력 음성데이터로부터 추출한 특징벡터  $X$ 만을 변화시킴으로써 음성인식의 성능을 높이는 것이다. 상기 환경 파라미터  $A$  및  $B$ 는 입력 음성데이터의 인식에 앞서 입력 음성 앞부분의 짧은 구간에서 획득되는 음성 데이터 등과 최대 우도 기법(maximum likelihood method)을 이용하여 예측할 수 있고, 이렇게 미리 결정된 환경 파라미터들을 이용하여 신속하게 특징 변환을 수행함으로써 환경 변화에 따라 실시간적인 적응을 수행할 수 있게 된다. 더욱 구체적으로 설명하면, 상기 환경 파라미터  $A$ ,  $B$ 는 입력된 음성과 기존에 존재하는 음향모델 간의 차이를 줄이기 위해 사용되는 일종의 변환 파라미터에 해당한다. 상기 환경 파라미터는 기존의 음향모델, 즉 상기 특징 변환 전의 기존의 음향모델과 입력된 음성 중 앞부분의 짧은 시간 동안 획득되는 데이터를 이용하여 예측할 수 있다. 예컨대, 특정 음향모델이 주어진 경우 음성이 발생할 수 있는 우도(likelihood)를 Baum's auxiliary 함수를 이용하여 최대화하는 기법 등을 사용하여 상기 환경 파라미터를 예측할 수 있고, 상기 예측된 상기 환경 파라미터를 상기 입력 음성데이터의 특징벡터에 적용하여 실제 음성인식 환경에 조금 더 가까운 상태의 특징벡터로 빠르게 변환함으로써 음성인식 성능을 보완할 수 있다.

[0054] 물론, 실제 구현에 있어서 상기 특징 변환을 위해 다양한 방식들이 적용될 수 있다. 예컨대, 특징벡터의 시간적인 특성을 고려한 cepstral mean subtraction), 평균-분산 정규화(mean-variance normalization: On real-time mean-variance normalization of speech recognition features, P.Pujol, D.Macho and C.Nadeu, ICASSP, 2006, pp.773-776 참조), RASTA 알고리즘(Relative SpecTrAl algorithm: Data-driven RASTA filters in reverberation, M.L.Shire et al, ICASSP, 2000, pp. 1627-1630 참조), 히스토그램 정규화(histogram normalization: Quantile based histogram equalization for noise robust large vocabulary speech recognition, F.Hilger and H.Ney, IEEE Trans. Audio, Speech, Language Processing, vol.14, no.3, pp.845-854 참조), 델타 특징 증강 알고리즘(augmenting delta feature: On the use of high order derivatives for high performance alphabet recognition, J. diMartino, ICASSP, 2002, pp.953-956) 등이 적용될 수 있다. 또한, 특징벡터들을 선형적으로 변환하는 기술로서, LDA(Linear Discriminant Analysis) 및 PCA(Principal Component Analysis: Optimization of temporal filters for constructing robust features in speech recognition, Jeih-Weih Hung et. al, IEEE Trans. Audio, Speech, and Language Processing, vol.14, No.3, 2006, pp.808-832 참조) 등이 적용될 수 있다. 또한, 비선형 신경망을 사용하는 방법으로서, TRAP(TempoRAI Patterns: Temporal patterns in ASR of noisy speech, H.Hermansky and S.Sharma, ICASSP, 1999, pp.289-292 참조), ASAT(Automatic Speech Attribute Transcription: A study on knowledge source integration for candidate rescoring in automatic speech recognition, Jinyu Li, Yu Tsao and Chin-Hui

Lee, ICASSP, 2005, pp.837-840 참조) 등이 적용될 수 있다.

- [0055] 그 다음, 상기 음성 데이터베이스(106)는, 상기 인식 대상 음성데이터를 음향모델 적응에 사용될 적응 데이터로서 저장한다(S240). 이때, 핸드폰, 로봇 등과 같이 자원 제약적인 모바일 환경에서는 모든 음성데이터를 저장하기보다 선입선출 방식 등을 통해 제한된 양의 음성데이터만을 저장할 수 있다.
- [0056] 그 다음, 상기 모델 변환부(108)는, 상기 음성인식 시스템이 음성인식을 수행하지 않고 있는 대기 모드에서(S250), 상기 음성 데이터베이스(106)에 저장된 음성데이터를 이용하여 기존 음향모델에 대한 모델 변환을 수행한다(S260). 즉, 음성인식 시스템이 적용된 기기의 데이터 연산량이 많지 않은 경우(예컨대, 야간이나 충전 중인 경우 등), 상기 인식 모드 상태에 있는 동안 상기 음성 데이터베이스(106)에 누적된 충분한 양의 적응 데이터를 이용하여 모델 변환을 수행한다(오프라인 적응). 상기 모델 변환이란 음향모델인 HMM의 파라미터들을 직접 변환하는 기법을 말한다. 환언하면, 음향모델은 평균 벡터(mean vector), 공분산 행렬(covariance matrix) 및 가중치(weight)로 구성되는데 상기 모델 변환은 이것들을 전부 또는 선택적으로 변화시키는 것이다. 이를 위해 상기 음성 데이터베이스(106)에 저장된 음성데이터를 이용하여 파라미터들을 예측하고 선형회귀트리(linear regression tree) 등을 이용하여 모델 파라미터를 변환한다. 이러한 과정은, 모바일 환경 등에서 시스템이 음향 모델 훈련을 위한 transcription을 예측하여 훈련하도록 하는 비지도 적응을 수행하는 과정에 해당한다.
- [0057] 이와 같이, 상기 음성인식 시스템의 대기 모드에서는, 상대적으로 연산량이 많지만 인식성능을 최대한 높일 수 있는 모델 변환을 수행한다. 이때, 상기 모델 변환부(108)가 모델 변환시 사용하는 저장된 음성데이터의 특징은, 화자 독립 음향모델 추정시와 마찬가지로 MFCC(Mel-Frequency Cepstral Coefficient)를 사용할 수 있지만, 화자 적응 과정의 특성을 고려하여 일반적인 EM(Expectation-Maximization) 기법 이외에 MLLR(Maximum Likelihood Linear Regression) 또는 MAP(Maximum A Posterior) 기법을 사용할 수 있다.
- [0058] 상기 모델 변환부(108)는, 상기 음성 데이터베이스(106)에 새롭게 저장된 음성데이터가 존재하는 경우, 상기 새롭게 저장된 음성데이터를 이용하여 최종 음향모델에 대한 모델 변환을 수행한다(S270). 한편, 상기 음성인식 시스템이 다시 인식 모드가 되는 경우(S250), 상술한 S220 내지 S240 단계들을 반복 수행한다.
- [0059] 아래에서 다시 설명하겠지만, 본 발명의 일 실시예에 있어서, 상기 인식 대상 음성데이터는 사용자의 발화 환경(예컨대, 지하철, 실내, 거리 등)에 따라 다른 특성을 지니게 되므로, 상기 화자 적응 장치(100)는, 상기 음성인식 시스템에서 음성구간 검출(Voice Activity Detection)을 통해 검출되는 음성구간 이외의 비음성구간에서 주변잡음의 특징을 추출하여 상기 인식 대상 음성데이터가 발생한 환경을 식별하는 환경 식별부(104)를 더 포함할 수 있다. 이 경우, 상기 음성 데이터베이스(106)는, 상기 환경 식별부(104)에 의해 식별된 환경별로 해당 환경에서 발생한 음성데이터를 저장한다. 상기 모델 변환부(108)는, 상기 음성 데이터베이스(106)에 저장된 음성 데이터를 이용하여 상기 식별된 환경별로 상기 기존 음향모델에 대한 모델 변환을 수행할 수 있다. 그 결과 생성되는 각각의 환경별 음향모델들은, 상기 저장된 음성데이터들이 모두 특정 사용자로부터 획득된 것인 점에서 화자 종속적 특징을 지니게 되며, 각각의 환경별로 음향모델을 훈련하는 점에서 환경 종속적 특징을 지니게 된다.
- [0060] 결과적으로, 본 발명에 따른 화자 적응 장치는, 특징 변환 적응과 모델 변환 적응을 반복적으로 수행하는 진화 학습(evolutional learning)을 통해 화자 적응 및 환경 적응을 동시에 수행하게 된다.
- [0061] 도 4에는 본 발명에 따른 진화 학습에 의한 음성인식 방법의 일례가 흐름도로 도시되어 있다.
- [0062] 도 1 및 도 4를 참조하면, 본 발명에 따른 음성인식 시스템은, 상기 화자 적응 장치(100)를 포함하며, 음성 입력부(110), 음성구간 검출부(120), 특징 추출부(130), 모델 데이터베이스(140), 모델 재배치부(150) 및 인식부(160)를 더 포함한다.
- [0063] 우선, 상기 음성 입력부(110)는, 아날로그 형태의 음성신호를 입력받는다(S400). 또한, 상기 음성 입력부(110)는, 반-에일리어싱 필터(anti-aliasing filter)를 통한 필터링, ADC(Analog to Digital Converter)를 통한 변환을 거쳐 상기 음성신호를 디지털 형태의 음성데이터로 변환하는 전처리 과정을 수행할 수 있다.
- [0064] 그 다음, 상기 음성구간 검출부(120)는, 상기 음성데이터가 인간의 음성에 해당하는 것인지 판별하기 위해 음성구간 검출(Voice Activity Detection: VAD)을 수행한다(S402). 이때, 상기 음성구간 검출을 수행하기 위해 상기 음성구간 검출부(120)는, 음성 및 음성 이외의 소리에 관한 학습 데이터를 이용하여 모델을 학습시킨 후 음성구간을 검출하는 기계학습 방식을 사용하거나, 또는 음성의 특성과 깊이 관련된 특징(zero crossing rate, spectral entropy 등)을 모델링하고 해당 특징의 출현 여부를 탐색하여 음성구간을 검출하는 방식 등을 사용할

수 있다.

- [0065] 그 다음, 상기 특징 추출부(130)는, 상기 음성구간의 인식 대상 음성데이터를 입력받아 상기 인식 대상 음성데이터에서 특징벡터를 추출한다(S404). 이때, 상기 추출된 특징벡터는 상기 인식 대상 음성데이터에서 음성인식에 필요한 성분만을 압축하여 가지고 있는 형태로 시간에 따른 주파수 정보를 지니는 것이 일반적이다. 상기 특징벡터는 MFCC(Mel-Frequency Cepstral Coefficients), LPCC(Linear Prediction Cepstral Coefficients), EIH(Ensemble Interaval Histogram) 등을 포함하며, 본 발명의 일 실시예에 있어서 MFCC를 특징벡터로 사용한다. 또한, 상기 특징 추출부(130)는, 상기 인식 대상 음성데이터에서 특징벡터를 추출하기 위해 여러 가지 전처리 과정, 예컨대 프레임 단위 구성, 해밍 윈도우, 푸리에 변환, 필터 뱅크, 켈스트럼 변환 등의 처리를 수행할 수 있다.
- [0066] 그 다음, 상기 화자 적응 장치(100)의 상기 특징 변환부(102)는, 상술한 바와 같이, 상기 음성인식 시스템이 음성인식을 수행하고 있는 인식 모드에서 미리 결정된 환경 파라미터를 이용하여 상기 추출된 특징벡터에 대해 특징 변환을 수행한다(S406). 그러면, 상기 인식부(160)가 상기 변환된 특징벡터 및 음향모델을 이용하여 음성인식을 수행하게 된다.
- [0067] 한편, 상기 화자 적응 장치(100)의 상기 환경 식별부(104)는, 상기 음성구간 검출부(120)에 의해 검출된 음성구간 이외의 비음성구간에서 주변잡음의 특징을 추출하여 상기 음성데이터가 발생한 환경을 식별한다(S408). 이때, 상기 환경 식별부(104)는, 상기 특징 추출부(130)와 동일한 원리로 비음성구간에서 주변잡음의 특징을 추출할 수 있으며, 또한 별도의 주변잡음 추정 알고리즘들(예컨대, S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-27(2), pp. 113-120, 1979. 참조)을 활용하여 주변잡음의 특징을 추출할 수도 있다.
- [0068] 그 다음, 상기 인식부(160)는, 상기 인식 모드에서 상기 특징 변환부(102)에 의해 변환된 특징벡터와 상기 화자 적응 장치(100)의 상기 모델 변환부(108)에 의해 변환된 환경 종속적 음향모델들 중에서 상기 식별된 환경에 대응하는 최적의 음향모델을 이용하여 상기 인식 대상 음성데이터에 대한 음성인식을 수행한다(S410).
- [0069] 그 다음, 상기 화자 적응 장치(100)의 상기 음성 데이터베이스(106)는, 상기 환경 식별부(104)에 의해 식별된 환경별로 해당 환경에서 발생한 음성데이터를 음향모델 적응에 사용될 적응 데이터로서 저장한다(S412).
- [0070] 한편, 상기 음성인식 시스템이 계속 인식 모드 상태인 경우(S414), 상술한 S400 내지 S412 단계들을 반복하여 수행한다.
- [0071] 반면, 상기 음성인식 시스템이 음성인식을 수행하지 않고 있는 대기 모드 상태인 경우(S414), 상기 화자 적응 장치(100)의 상기 모델 변환부(108)는, 상기 음성 데이터베이스(106)에 저장된 음성데이터를 이용하여 기존 음향모델에 대한 모델 변환을 수행한다(S416). 특히, 상기 모델 변환부(108)는, 상기 음성 데이터베이스(106)부에 저장된 음성데이터를 이용하여 상기 식별된 환경별로 상기 기존 음향모델에 대한 모델 변환을 각각 수행한다. 즉, 상기 식별된 환경들 각각에 종속하는 음향모델을 생성한다.
- [0072] 그 다음, 상기 모델 데이터베이스(140)는, 상기 모델 변환부(108)에 의해 상기 식별된 환경별로 모델 변환된 환경별 음향모델을 저장한다(S418). 즉, 상기 모델 데이터베이스(140)는 최초에 화자 독립 또는 화자 및 환경 독립 음향모델을 저장하고 있을 수 있다. 그러나, 상술한 진화 학습 과정들을 거치면서 상기 모델 데이터베이스(140)는 화자 및 환경 종속적인 음향모델들을 저장하게 된다.
- [0073] 도 5에는 진화 학습을 통해 화자 및 환경 종속 음향모델을 생성하는 과정이 도시되어 있다.
- [0074] 도 5를 참조하면, 상기 음성인식 시스템에는 최초에 화자 및 환경 독립 음향모델(500)이 생성 및 저장되어 있다. 상기 화자 및 환경 독립 음향모델은 상기 음성 데이터베이스(106)에 저장된 환경별 음성데이터(510)를 이용하여 각각의 환경별로 모델 변환을 수행함으로써 상기 각각의 환경에 대응하는 환경 종속 음향모델들(520)을 생성한다. 그리고, 상기 음성 데이터베이스(106)에 새로운 환경에서의 음성데이터(530)가 저장되는 경우, 상기 환경 종속 음향모델들(520) 각각에 대하여 상기 새로운 환경별로 해당 음성데이터를 이용하여 모델 변환을 수행하고 새로운 환경 종속 음향모델들(540)을 생성한다.
- [0075] 일 실시예에 있어서, 상기 음성 데이터베이스(106)에 새롭게 저장된 음성데이터가 이미 환경 종속 음향모델이 생성된 환경에서 발생한 것이라면, 상기 새롭게 저장된 음성데이터를 이용하여 상기 이미 생성된 환경 종속 음향모델에 대한 모델 변환을 다시 수행함으로써 특정 환경에 진화적으로 적응하도록 할 수 있다. 또한, 상기 새롭게 저장된 음성데이터가 새로운 환경에서 발생한 것이라면, 상술한 바와 같이 상기 새롭게 저장된 음성데이터를

이용하여 상기 이미 생성된 환경 종속 음향모델들 전부에 대해 모델 변환을 수행하고, 그 결과 생성되는 환경 종속 모델들의 우도(likelihood)를 이용하여 상기 새로운 환경에 종속된 음향 모델을 생성할 수 있다.

- [0076] 그 다음, 상기 모델 재배치부(150)는, 미리 결정된 임계치에 의해 상기 모델 데이터베이스(140)에 저장되는 음향모델의 개수를 제한한다(S420). 즉, 자원 제약적인 모바일 기기 등에 있어서 환경 종속 음향모델들을 무제한 생성하여 모두 저장 및 사용하는 것이 아니라 저장용량 내지 연산량을 고려하여 저장 및 사용하는 음향모델의 개수를 제한할 수 있다.
- [0077] 상기 모델 재배치부(150)는, 상기 임계치 이하로 음향모델의 개수를 제한하기 위해 다양한 방식들을 적용할 수 있다. 예컨대, 일 실시예에 있어서, 상기 모델 데이터베이스(140)에 저장되는 음향모델의 개수가 상기 임계치를 초과하는 경우, 상기 모델 재배치부(150)는 새롭게 생성되는 음향모델이 단순히 상기 모델 데이터베이스(140)에 저장되지 않도록 제한할 수 있다. 다른 일 실시예에 있어서, 상기 모델 재배치부(150)는 상기 생성된 음향모델들의 우도(likelihood)를 이용하거나 환경별 저장된 데이터 양을 고려하여 각 음향모델들을 가중치합(weighted sum) 또는 선형결합(linear combination) 등을 통해 상기 임계치 이하로 음향모델의 개수를 감소시킬 수 있다. 또한, 다른 일 실시예에 있어서, 상기 모델 재배치부(150)는, 상기 환경 식별부(104)로 하여금 음성데이터가 발생하는 환경들을 소정 개수의 유형으로만 분류하여 식별하도록 하여, 상기 모델 변환부(108)가 생성하는 음향모델의 개수를 상기 임계치 이하로 조절할 수 있다.
- [0078] 그 다음, 상기 음성 데이터베이스(106)에 새롭게 저장된 음성데이터가 존재하는 경우(S422), 상기 음성인식 시스템은, 상술한 S416 내지 S420 단계들을 반복하게 된다. 다만, 상기 음성인식 시스템이 음성인식을 수행하는 인식 모드 상태로 전환되는 경우, 상술한 S400 내지 S412 단계들을 반복하게 된다.
- [0079] 앞서 설명한 바와 같이, 각각의 환경별 음향모델들은, 상기 저장된 음성데이터들이 모두 특정 사용자로부터 획득된 것인 점에서 화자 종속적 특징을 지니게 되며, 각각의 환경별로 음향모델을 훈련하는 점에서 환경 종속적 특징을 지니게 된다.
- [0080] 결과적으로, 본 발명에 따른 음성인식 시스템은, 특징 변환 적응과 모델 변환 적응을 반복적으로 수행하는 진화 학습(evolutional learning)을 통해 화자 적응 및 환경 적응을 동시에 수행하게 된다.
- [0081] 한편, 상기 인식부(160)는, 인식 대상 음성데이터에 대한 음성인식 수행시, 상기 환경 식별부(104)에 의해 상기 인식 대상 음성데이터의 발생 환경을 식별하고, 상기 모델 데이터베이스(140)에서 상기 식별된 환경에 대응하는 음향모델을 검색 및 이용한다. 즉, 상기 인식부(160)는, 상기 음성인식 시스템의 인식 모드에서 상기 특징 변환부(102)에 의해 변환된 특징벡터와 상기 모델 변환부(108)에 의해 변환된 화자 및 환경 종속 음향모델을 이용하여 음성인식을 수행하게 된다.
- [0082] 일 실시예에 있어서, 상기 인식부(160)는, 인식 대상 음성데이터에 대한 음성인식 수행시, 상기 모델 데이터베이스(140)에 저장된 음향모델들과 우도(likelihood)를 계산하여 가장 큰 값을 지니는 음향모델을, 상기 인식 대상 음성데이터의 발생 환경(즉, 현재 음성인식 환경)에 종속된 것으로 판단하고, 해당 음향모델을 이용하여 음성인식을 수행할 수 있다.
- [0083] 한편, 본 발명은, 컴퓨터로 판독할 수 있는 기록매체에 컴퓨터가 읽어들일 수 있는 프로그램 코드로 구현하는 것이 가능하다. 본 발명이 소프트웨어를 통해 실행될 때, 본 발명의 구성 수단들은 필요한 작업을 실행하는 코드 세그먼트들이다. 또한, 프로그램 또는 코드 세그먼트들은 컴퓨터의 프로세서 판독가능 매체에 저장되거나 전송 매체 또는 통신망을 통해 반송파와 결합된 컴퓨터 데이터 신호로 전송될 수 있다.
- [0084] 컴퓨터 판독가능 기록매체에는 컴퓨터 시스템이 읽어들일 수 있는 데이터를 저장하는 모든 종류의 기록장치가 포함된다. 예컨대, 컴퓨터 판독가능 기록매체에는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광데이터 저장장치 등이 포함될 수 있다. 또한, 컴퓨터 판독가능 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어 컴퓨터가 읽어들일 수 있는 코드를 분산방식으로 저장하고 실행되도록 할 수 있다.
- [0085] 상술한 바와 같이, 본 발명은 음성인식 시스템의 화자 적응을 위해 진화 학습에 의한 비지도 적응 방식을 적용함으로써 사용자 편의성을 도모하는 이점을 제공한다. 또한, 음성인식 시스템의 동작 상태에 따라 특징 변환 및 모델 변환을 적절하게 수행함으로써, 화자 적응 방식의 효율성 및 음성인식 시스템의 인식 성능을 개선하는 이점을 제공한다. 나아가, 음성인식 환경별로 화자 종속 음향모델을 생성함으로써, 화자 적응은 물론 환경 적응까지 동시에 수행하는 이점을 제공한다.
- [0086] 지금까지 본 발명에 대해 실시예들을 참고하여 설명하였다. 그러나 당업자라면 본 발명의 본질적인 기술적 사상

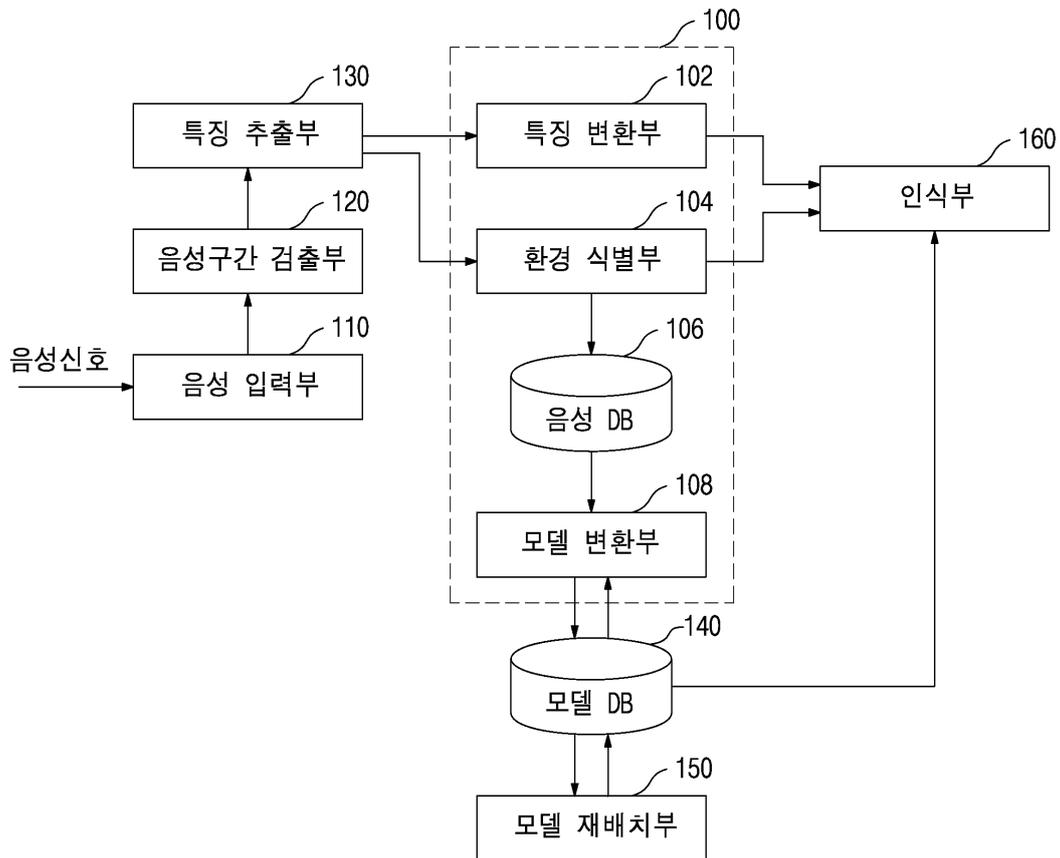
으로부터 벗어나지 않는 범위에서 본 발명이 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 즉, 본 발명의 진정한 기술적 범위는 첨부된 특허청구범위에 나타나 있으며, 그와 균등범위 내에 있는 모든 차이점은 본 발명에 포함되는 것으로 해석되어야 할 것이다.

**도면의 간단한 설명**

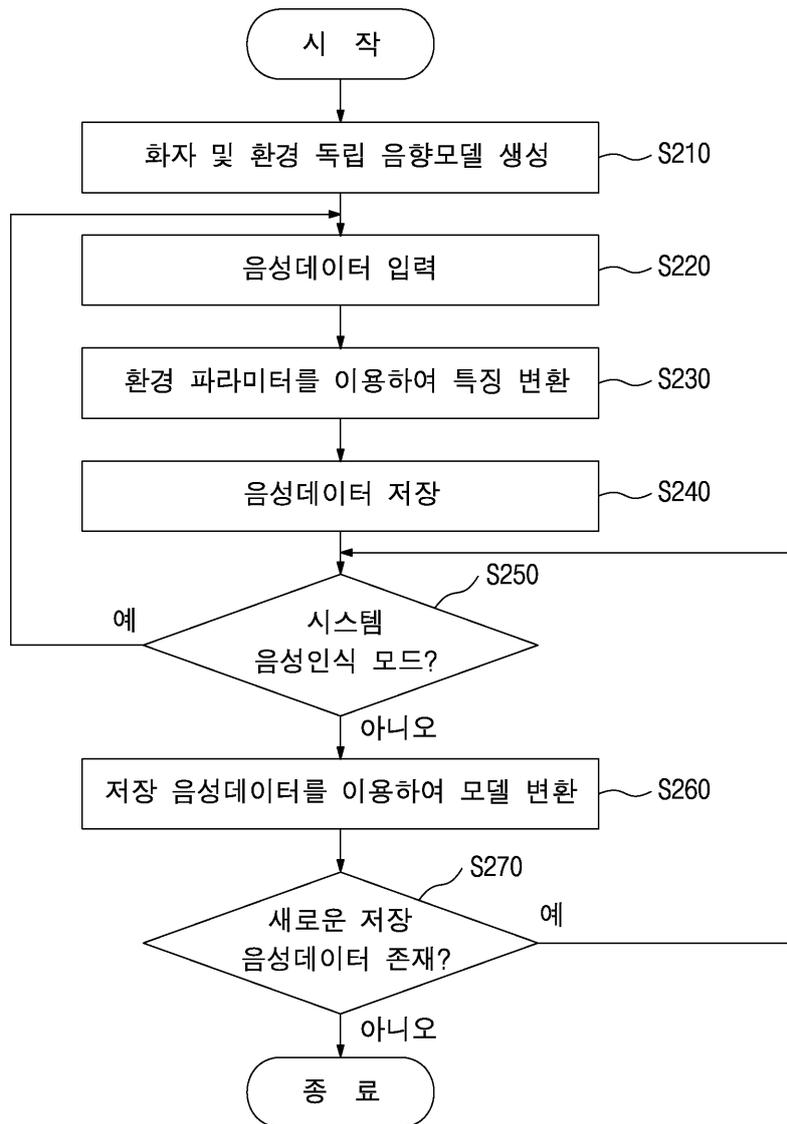
- [0087] 도 1은 본 발명에 따른 진화 학습에 의한 화자 적응 장치가 적용된 음성인식 시스템의 일례를 나타낸 블록도이다.
- [0088] 도 2는 본 발명에 따른 진화 학습에 의한 화자 적응 방법의 일례를 나타낸 흐름도이다.
- [0089] 도 3은 화자 적응 기법의 기본원리를 나타낸 도면이다.
- [0090] 도 4는 본 발명에 따른 진화 학습에 의한 음성인식 방법의 일례를 나타낸 흐름도이다.
- [0091] 도 5는 진화 학습을 통해 화자 및 환경 종속 음향모델을 생성하는 과정을 나타낸 도면이다.

**도면**

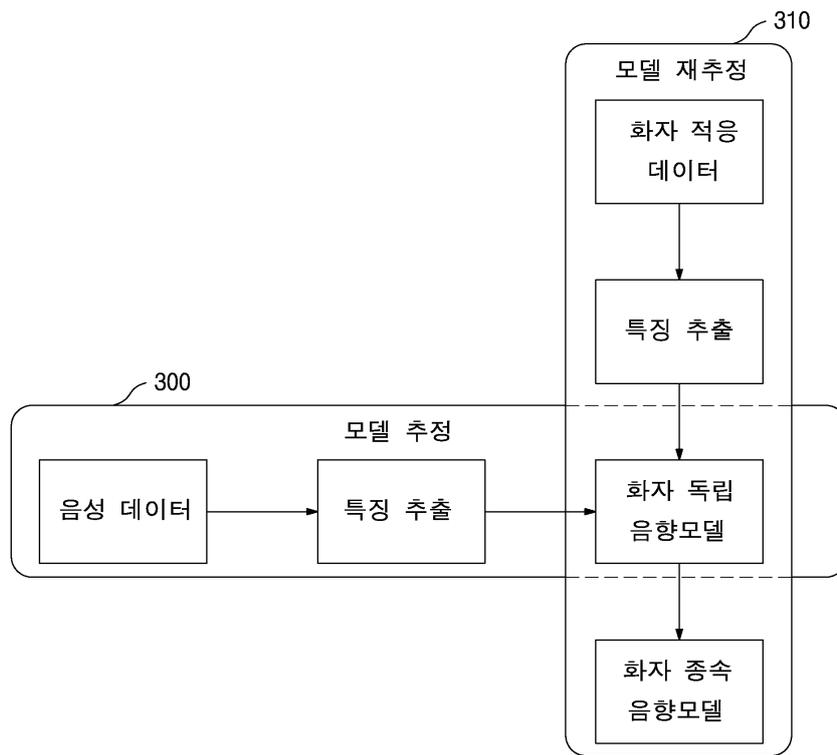
**도면1**



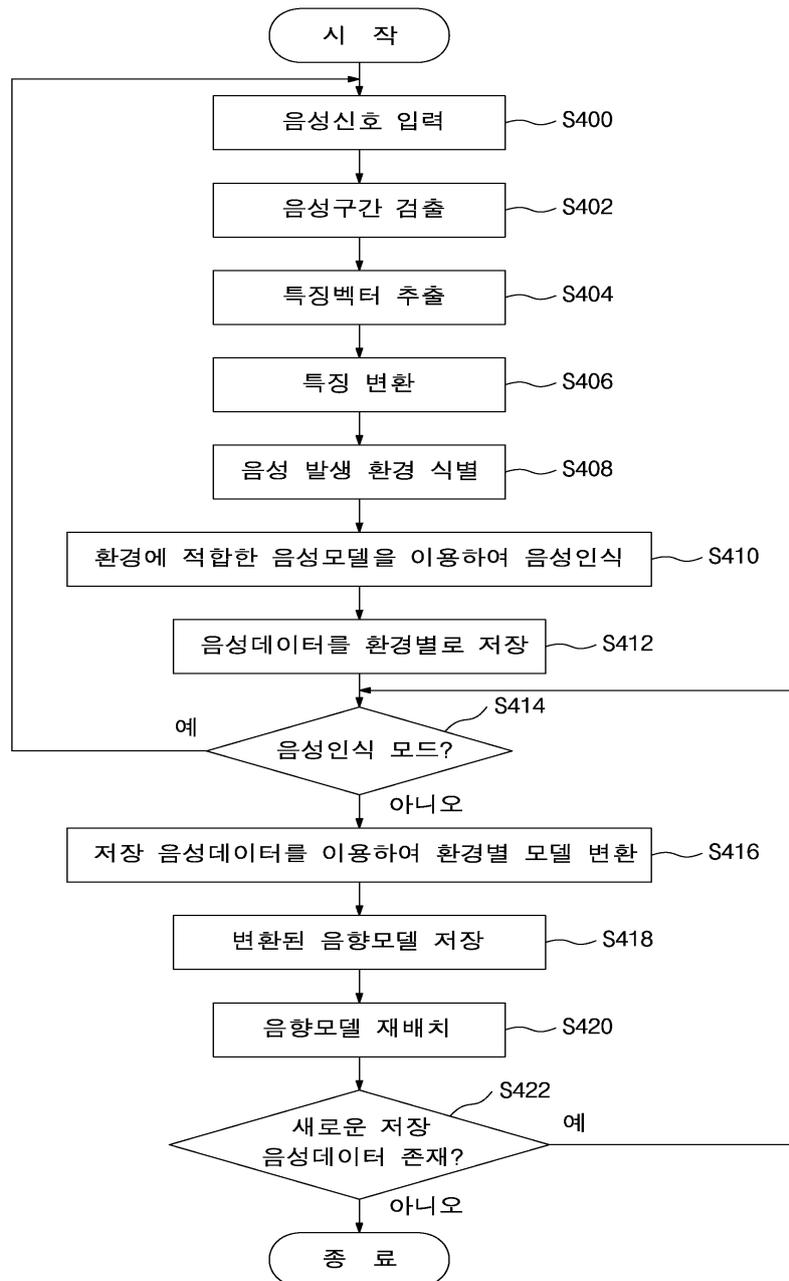
도면2



도면3



도면4



도면5

