



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2007년11월19일  
(11) 등록번호 10-0776803  
(24) 등록일자 2007년11월08일

(51) Int. Cl.

G10L 17/00 (2006.01) G10L 15/28 (2006.01)  
G10L 15/02 (2006.01) G10L 15/22 (2006.01)

(21) 출원번호 10-2006-0093539

(22) 출원일자 2006년09월26일

심사청구일자 2006년09월26일

(56) 선행기술조사문헌

JP07219581 A

(뒷면에 계속)

(73) 특허권자

한국전자통신연구원

대전 유성구 가정동 161번지

(72) 발명자

박근창

대전 서구 월평3동 다모아아파트 108동 1103호

김혜진

인천 서구 마전동 535-3

(뒷면에 계속)

(74) 대리인

특허법인 씨엔에스·로고스

전체 청구항 수 : 총 13 항

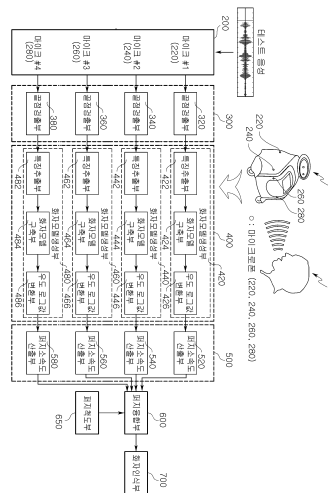
심사관 : 경연정

(54) 다채널 퍼지 융합을 통한 지능형 로봇의 화자 인식 장치 및 그 방법

(57) 요약

본 발명은 다채널 퍼지 융합을 통한 지능형 로봇의 화자 인식 장치 및 그 방법에 관한 것으로, 상기 장치는, 복수의 채널을 통해 온라인으로 화자의 음성을 취득 및 등록하는 다채널 마이크로폰, 상기 등록된 각 채널별 음성들에 대해 각각 시작점 및 끝점을 검출하여 문장을 구분하고 상기 구분된 문장에 포함된 음성의 잡음을 제거하는 음성 데이터 취득부, 상기 잡음이 제거된 각 음성 데이터에 대한 특징에 기초하여 화자 모델을 구축하고 상기 구축된 화자 모델에 대해 우도 로그값으로 변환하는 화자모델 생성부, 상기 각 채널별 우도 로그값들에 대해 융합된 퍼지 값을 산출하는 퍼지 처리부, 및 상기 산출된 퍼지 값에 기초한 융합값의 최대치를 화자로서 인식하는 화자 인식부를 포함하며, 이에 의해, 전 방향에 대해 발생하는 화자의 음성을 보다 높은 성능으로 정확하게 취득할 수 있고 잡음환경이나 원거리에서 화자 인식 성능을 높일 수 있다.

대표도 - 도1



(72) 발명자

**배경숙**

경기 안양시 만안구 석수1동 석수대림아파트  
102-1401

**지수영**

대전 유성구 공동 395-3번지 다솔아파트 102동 30  
5호

(56) 선행기술조사문헌

KR1019990001828 A

KR1020030082265 A

KR1020040010860 A

KR1020050055776 A

US6321194 B1

10-2005-0036331

특1998-040752

10-2005-0115857

---

## 특허청구의 범위

### 청구항 1

복수의 채널을 통해 온라인으로 화자의 음성을 취득 및 등록하는 다채널 마이크로폰;

상기 등록된 각 채널별 음성들에 대해 각각 시작점 및 끝점을 검출하여 문장을 구분하고, 상기 구분된 문장에 포함된 음성의 잡음을 제거하는 음성 데이터 취득부;

상기 잡음이 제거된 각 음성 데이터에 대한 특징에 기초하여 화자 모델을 구축하고, 상기 구축된 화자 모델에 대해 우도 로그값으로 변환하는 화자모델 생성부;

상기 각 채널별 우도 로그값들에 대해 융합된 퍼지 값을 산출하는 퍼지 처리부; 및

상기 산출된 퍼지 값에 기초한 융합값의 최대치를 화자로서 인식하는 화자 인식부를 포함하는 것을 특징으로 하는 지능형 로봇의 화자 인식 장치.

### 청구항 2

제 1항에 있어서,

상기 음성 데이터 취득부는 상기 각 채널별 음성들에 대해 끝점 검출 알고리즘(Endpoint detection)을 이용하여 상기 시작점 및 끝점을 검출하는 것을 특징으로 하는 지능형 로봇의 화자 인식 장치.

### 청구항 3

제 1항 또는 제2항에 있어서,

상기 음성 데이터 취득부는 위너 필터(Winer filter)를 이용하여 상기 시작점 및 끝점에 의해 구분된 문장의 음성으로부터 잡음을 제거하는 것을 특징으로 하는 지능형 로봇의 화자 인식 장치.

### 청구항 4

제 1항에 있어서,

상기 화자모델 생성부는,

상기 잡음이 제거된 각 채널의 음성에 대한 특징 정보를 추출하는 특징 추출부;

상기 특징 정보에 기초하여 상기 각 채널의 음성에 대한 화자 모델을 구축하는 화자모델 구축부; 및

상기 각 화자 모델의 음성에 대응하는 우도 로그값을 생성하는 우도 로그값 변환부를 포함하는 것을 특징으로 하는 지능형 로봇의 화자 인식 장치.

### 청구항 5

제 4항에 있어서,

상기 특징 추출부는 멜 캡스트럼(MFCC: Mel-Frequency Cepstral Coefficients)을 이용하여 상기 각 채널의 음성에 대한 특징 정보를 추출하는 것을 특징으로 하는 지능형 로봇의 화자 인식 장치.

### 청구항 6

제 4항에 있어서,

상기 화자모델 구축부는 가우시안 혼합모델(GMM: Gaussian Mixture Model)을 이용하여 상기 각 채널의 음성에 대한 화자 모델을 구축하는 것을 특징으로 하는 지능형 로봇의 화자 인식 장치.

### 청구항 7

제 1항에 있어서,

상기 퍼지 처리부는,

상기 각 채널별 우도 로그값들에 대해 각각 퍼지 소속도를 산출하는 퍼지 소속도 산출부; 및

상기 산출된 퍼지소속도 값들을 퍼지 융합하는 퍼지 융합부를 포함하는 것을 특징으로 하는 지능형 로봇의 화자 인식 장치.

**청구항 8**

제 7항에 있어서,

상기 퍼지 소속도 산출부는 시그모이드(sigmoid) 소속함수를 이용하여 상기 퍼지 소속도를 산출하는 것을 특징으로 하는 지능형 로봇의 화자 인식 장치.

**청구항 9**

제 7항에 있어서,

상기 퍼지 융합부는 퍼지 적분을 이용하여 상기 퍼지소속도 값들을 퍼지 융합하는 것을 특징으로 하는 지능형 로봇의 화자 인식 장치.

**청구항 10**

다채널 마이크폰을 통해 온라인으로 화자의 음성을 취득 및 등록하는 단계;

상기 등록된 각 채널별 음성들에 대해 화자 모델을 구축하는 단계; 및

상기 각 채널의 화자 모델에 대응하는 우도 로그값의 퍼지 소속도를 융합하여, 이에 대한 최대치를 화자로서 인식하는 단계를 포함하는 것을 특징으로 하는 지능형 로봇의 화자 인식 방법.

**청구항 11**

제 10항에 있어서,

상기 화자 음성 데이터 등록 단계는,

상기 다채널 마이크폰으로부터 온라인으로 출력되는 화자의 음성을 취득하는 단계;

상기 각 채널별 음성에 대해 시작점 및 끝점을 검출하는 단계; 및

상기 시작점과 끝점으로 구분된 음성의 문장들에 대해 잡음 요소를 제거하는 단계를 포함하는 것을 특징으로 하는 지능형 로봇의 화자 인식 방법.

**청구항 12**

제 10항 또는 제 11항에 있어서,

상기 화자 모델 구축 단계는,

상기 등록된 각 채널별 음성들로부터 특징 값들을 추출하는 단계;

상기 특징 값들에 기초하여 상기 채널별 음성에 대한 화자 모델을 구축하는 단계; 및

상기 구축된 화자 모델에 대한 음성에 대응하는 우도 로그값을 생성하는 단계를 포함하는 것을 특징으로 하는 지능형 로봇의 화자 인식 방법.

**청구항 13**

제 12항에 있어서,

상기 화자 인식 단계는,

상기 각 채널별 우도 로그값에 기초하여 퍼지 소속도를 산출하는 단계;

상기 산출된 퍼지 소속도 값들을 퍼지 융합하는 단계; 및

상기 퍼지 융합된 값의 최대치를 상기 화자로 인식하는 단계를 포함하는 것을 특징으로 하는 지능형 로봇의 화

자 인식 방법.

## 명세서

### 발명의 상세한 설명

#### 발명의 목적

##### 발명이 속하는 기술 및 그 분야의 종래기술

- <6> 본 발명은 지능형 로봇의 화자 인식 장치 및 방법에 관한 것으로서, 보다 상세하게는, 마이크로폰을 통해 발성자의 음성을 보다 선명한 감도로 독취하여 화자가 누구인지를 보다 정확하게 판별할 수 있는 지능형 로봇의 화자 인식 장치 및 방법에 관한 것이다.
- <7> 최근 들어, 삶 속에서 사용자의 편의를 도모하기 위해 사용자가 의도하는 해당 작업을 돕는 로봇이 개발되고 있다. 특히, 사용자와 로봇 간에 상호작용을 통해 로봇이 지능적으로 판단하고 그 결과에 따른 동작을 수행할 수 있는 지능형 로봇이 개발되고 있다.
- <8> 이러한 지능형 로봇 기술을 구현하기 위해서는 로봇이 사용자의 명령이나 동작을 인식하는 기술이 요구된다. 이 중에서 사용자 인식기술의 핵심기술로서 부각되고 있는 것이 사용자의 얼굴 인식과 화자(발성자) 인식 기술이다.
- <9> 현재까지, 로봇 환경에서 화자인식기술은 얼굴인식기술과 달리 많은 연구와 개발이 이루어지고 있지 않은 실정이다. 다만, 단일 채널을 가진 마이크로폰에 의해 화자의 음성을 독취하고, 이를 기초로 화자를 인식하는 기술이 일부 기술 분야에서 행해지고 있을 뿐이다. 그러나 이는 화자 인식을 위한 감도 및 정확도가 떨어지는 문제점이 있다.
- <10> 상기와 같은 화자인식방법은 일반적으로 보안 및 전자상거래 분야에서 주로 사용되고 있다. 주로 사용되는 화자인식방법은 문장종속 화자인식, 문장제시형 화자인식, 및 문장독립형 화자인식 방법을 예로 들 수 있다. 이 중에서 로봇환경에서는 어떠한 문장을 발성해도 화자를 인식하기 위해서는, 문장 독립형 화자인식 방법이 필요하다.
- <11> 또한, 로봇 환경에서의 화자 인식 방법은 로봇이 위치하는 전(모든) 방향에서 발성자가 대화나 명령을 발성할 경우, 로봇이 화자인식을 수행해야 한다. 또한 로봇은 근거리뿐만 아니라 원거리에서 발성되는 음성도 독취하여 발성자가 누구인지 알아내는 것이 필요하다. 뿐만 아니라, 지능형 로봇은 다양한 환경에서 존재하는 잡음환경 요소를 식별하여 오동작을 방지할 수 있는 기술이 요구된다.
- <12> 이와 같이, 종래의 단일 채널 마이크로폰이 부착된 로봇은 잡음환경, 모든 방향, 및 원거리 환경에서 화자 인식 성능이 떨어지게 되는 문제점이 있다.

##### 발명이 이루고자 하는 기술적 과제

- <13> 상기와 같은 문제점을 해결하기 위한 본 발명의 제1 목적은, 사용자와 로봇 간에 상호작용을 위해 임의의 잡음 환경에서 발성자가 말하는 문장에 상관없이 화자가 누구인지를 보다 정확하게 인식할 수 있는 지능형 로봇의 화자 인식 장치 및 방법을 제공하는 데 있다.
- <14> 본 발명의 제2 목적은, 로봇이 위치하는 모든 방향으로부터 발성되는 음성을 독취하여 보다 정확하게 화자를 인식할 수 있는 지능형 로봇의 화자 인식 장치 및 방법을 제공하는 데 있다.
- <15> 본 발명의 제3 목적은, 로봇이 위치하는 곳으로부터 원거리에서 발성되는 발성자의 음성을 독취하여 보다 정확하게 화자를 인식 할 수 있는 지능형 로봇의 화자 인식 장치 및 방법을 제공하는 데 있다.

#### 발명의 구성 및 작용

- <16> 상기와 같은 목적을 달성하기 위한 본 발명의 실시예에 따른 지능형 로봇의 화자 인식 장치는, 복수의 채널을 통해 온라인으로 화자의 음성을 취득 및 등록하는 다채널 마이크로폰; 상기 등록된 각 채널별 음성들에 대해 각각 시작점 및 끝점을 검출하여 문장을 구분하고, 상기 구분된 문장에 포함된 음성의 잡음을 제거하는 음성 데이터 취득부; 상기 잡음이 제거된 각 음성 데이터에 대한 특징에 기초하여 화자 모델을 구축하고, 상기 구축된 화

자 모델에 대해 우도 로그값으로 변환하는 화자모델 생성부; 상기 각 채널별 우도 로그값들에 대해 융합된 퍼지 값을 산출하는 퍼지 처리부; 및 상기 산출된 퍼지 값에 기초한 융합값의 최대치를 화자로서 인식하는 화자 인식부를 포함한다.

- <17> 상기 음성 데이터 취득부는 상기 각 채널별 음성들에 대해 끝점 검출 알고리즘(Endpoint detection)을 이용하여 상기 시작점 및 끝점을 검출한다. 상기 음성 데이터 취득부는 위너 필터(Winer filter)를 이용하여 상기 시작점 및 끝점에 의해 구분된 문장의 음성으로부터 잡음을 제거한다.
- <18> 또한 상기 화자모델 생성부는, 상기 잡음이 제거된 각 채널의 음성에 대한 특징 정보를 추출하는 특징 추출부; 상기 특징 정보에 기초하여 상기 각 채널의 음성에 대한 화자 모델을 구축하는 화자모델 구축부; 및 상기 각 화자 모델의 음성에 대응하는 우도 로그값을 생성하는 우도 로그값 변환부를 포함한다. 이때, 상기 특징 추출부는 멜 캡스트럼(MFCC: Mel-Frequency Cepstral Coefficients)을 이용하여 상기 각 채널의 음성에 대한 특징 정보를 추출한다. 상기 화자모델 구축부는 가우시안 혼합모델(GMM: Gaussian Mixture Model)을 이용하여 상기 각 채널의 음성에 대한 화자 모델을 구축한다.
- <19> 상기 퍼지 처리부는, 상기 각 채널별 우도 로그값들에 대해 각각 퍼지 소속도를 산출하는 퍼지 소속도 산출부; 및 상기 산출된 퍼지소속도 값들을 퍼지 융합하는 퍼지 융합부를 포함한다. 이때, 상기 퍼지 소속도 산출부는 시그모이드(sigmoid) 소속함수를 이용하여 상기 퍼지 소속도를 산출한다. 상기 퍼지 융합부는 퍼지 적분을 이용하여 상기 퍼지소속도 값들을 퍼지 융합한다.
- <20> 한편, 상기와 같은 목적을 달성하기 위한 본 발명의 실시예에 따른 지능형 로봇의 화자 인식 방법은, 다채널 마이크로폰을 통해 온라인으로 화자의 음성을 취득 및 등록하는 단계; 상기 등록된 각 채널별 음성들에 대해 화자 모델을 구축하는 단계; 및 상기 각 채널의 화자 모델에 대응하는 우도 로그값의 퍼지 소속도를 융합하여, 이에 대한 최대치를 화자로서 인식하는 단계를 포함한다.
- <21> 바람직하게는, 상기 화자 음성 데이터 등록 단계는, 상기 다채널 마이크로폰으로부터 온라인으로 출력되는 화자의 음성을 취득하는 단계; 상기 각 채널별 음성에 대해 시작점 및 끝점을 검출하는 단계; 및 상기 시작점과 끝점으로 구분된 음성의 문장들에 대해 잡음 요소를 제거하는 단계를 포함한다.
- <22> 상기 화자 모델 구축 단계는, 상기 등록된 각 채널별 음성들로부터 특징 값들을 추출하는 단계; 상기 특징 값들에 기초하여 상기 채널별 음성에 대한 화자 모델을 구축하는 단계; 및 상기 구축된 화자 모델에 대한 음성에 대응하는 우도 로그값을 생성하는 단계를 포함한다.
- <23> 상기 화자 인식 단계는, 상기 각 채널별 우도 로그값에 기초하여 퍼지 소속도를 산출하는 단계; 상기 산출된 퍼지 소속도 값들을 퍼지 융합하는 단계; 및 상기 퍼지 융합된 값의 최대치를 상기 화자로 인식하는 단계를 포함한다.
- <24> 본 발명에 따르면, 지능형 로봇에서 음성 기반 인간-로봇 상호작용을 수행하기 위해 다채널의 마이크로폰을 통해 온라인 화자등록과 로봇의 모든 방향에서 음성 인식이 가능하도록 구현함으로써, 전 방향에 대해 발생되는 화자의 음성을 보다 높은 성능으로 정확하게 취득할 수 있다. 또한 본 발명은 다채널 기반 온라인 화자등록/인식/퍼지 융합을 통해 보다 정확하게 화자를 인식할 수 있도록 함으로써, 잡음환경이나 원거리에서 화자 인식 성능의 저하를 최소화할 수 있다.
- <25> 이하, 본 발명의 바람직한 실시예들을 첨부한 도면을 참조하여 상세히 설명한다. 도면들 중 동일한 구성요소들은 가능한 한 어느 곳에서든지 동일한 부호들로 나타내고 있음에 유의해야 한다. 또한 본 발명의 요지를 불필요하게 흐릴 수 있는 공지 기능 및 구성에 대한 상세한 설명은 생략한다.
- <26> 본 발명은 마이크로폰을 통해 발생자의 음성을 보다 선명한 감도로 독취하여 화자가 누구인지를 보다 정확하게 판별할 수 있도록 하기 위해, 다채널 마이크로폰을 통해 발생자의 음성을 독취하고 독취한 다채널의 음성을 퍼지 융합하여 보다 높은 감도에서 화자를 보다 정확하게 판별할 수 있는 지능형 로봇의 화자 인식 방법을 제안한다. 이와 같이, 로봇의 모든 방향에서 음성 인식이 가능하기 위해 본 발명에서는 로봇 주위에 여러 개의 마이크로폰을 부착하고, 잡음환경 및 원거리에서도 보다 정확한 화자 인식이 가능하도록 다채널 마이크로폰의 각 채널로부터 독취한 음성 데이터에 대해 퍼지 적분을 통해 퍼지 소속 값을 융합하는 기법을 개시한다.
- <27> 도 1은 본 발명의 바람직한 실시예에 따른 다채널 기반 퍼지 융합을 이용한 지능형 로봇의 화자 인식 장치를 도시한 블록도이다.
- <28> 도시된 바와 같이, 지능형 로봇(100)에는 화자(10)로부터 발생되는 음성을 전 방향에 대해 독취가 가능하도록

측면에 다수개의 마이크로폰(200)이 배치된다. 본 실시예에서는 지능형 로봇(100)의 측면에 4개의 마이크로폰(220,240,260,280)이 배치되어 있음을 알 수 있다. 이에 따라, 화자(10)의 음성이 어느 방향으로부터 발생되더라도 소정 간격을 두고 서로 다른 방향을 향하여 배치된 복수의 다채널 마이크로폰(220,240,260,280)을 통해 발생하는 음성을 보다 용이하게 취득할 수 있다. 이에 따라, 다채널의 마이크로폰(200)은 테스트 음성을 통해 온라인으로 화자(10)의 음성을 취득한다.

- <29> 본 실시예에서 네 개의 마이크로폰(220,240,260,280)이 로봇에 부착되어 있는 경우, 이에 대응하여 각각 끝점 검출부(320,340,360,380)가 마련된다. 이에 따라, 다채널의 마이크로폰(200)에 대응하여 마련되는 끝점 검출부(300)는 다채널의 마이크로폰(200)으로부터 각각 취득되어 출력되는 음성에 대해 시작점 및 끝점을 검출하고, 각각 구분되는 문장들에 대한 잡음을 제거한다.
- <30> 본 실시예에서는 복수의 끝점 검출부(300)에 대응하여 각 채널별로 음성에 대한 화자 모델을 구축하기 위한 화자모델 생성부(420,440,460,480)가 각각 구비된다.
- <31> 여기서, 화자모델 생성부(420,440,460,480)는 각각 특징추출부(422,442,462,482), 화자모델 구축부(424,444,464,484), 및 우도 로그값 변환부(426,446,466,486)를 포함하여 구성된다.
- <32> 이에 따라, 특징추출부(422,442,462,482)는 각각 끝점 검출부(300)에서 잡음이 제거된 각 채널의 음성데이터로부터 특징 정보를 추출한다. 화자모델 구축부(424,444,464,484)는 각 채널별 음성에 대한 화자 모델을 구축한다. 우도 로그값 변환부(426,446,466,486)는 각 화자 모델의 음성에 대해 우도 로그값으로 변환한다.
- <33> 도면에서와 같이, 화자모델 생성부(400)의 각 부(420,440,460,480)에 대응하여 퍼지 소속도 산출부(520,540,560,580)가 구비된다. 퍼지 소속도 산출부(520,540,560,580)는 각각 화자모델 생성부(420,440,460,480)로부터 변환된 우도 로그값에 기초하여 퍼지 소속도를 산출한다. 퍼지 소속도 산출부(520,540,560,580)의 출력단에 연결되는 퍼지 융합부(600)는 퍼지 소속도 산출부(520,540,560,580)로부터 산출된 각 채널의 모델 음성에 대한 우도 로그값들을 퍼지 융합한다.
- <34> 이에 따라, 퍼지 융합부(600)의 출력단에 연결되는 화자 인식부(700)는 퍼지 융합값이 최대치인 음성을 발성한 사용자를, 화자로서 인식한다.
- <35> 도 2는 본 발명의 바람직한 실시예에 따른 다채널 기반 퍼지 융합을 이용한 지능형 로봇의 화자 인식 방법을 도시한 흐름도이다.
- <36> 먼저, 끝점 검출부(300)는 여러 문장을 따라 읽게 함으로서 화자로부터 온라인상에서 발생하는 음성을, 다채널 마이크로폰(200)을 통해 취득하여 온라인으로 화자 음성 데이터로서 등록한다(S100).
- <37> 온라인으로 화자 음성 데이터가 등록되면, 화자 모델 생성부(400)는 각 채널별로 음성 데이터에 대한 화자 모델을 구축한다(S200).
- <38> 화자 모델이 구축되면, 퍼지 소속도 산출부(500) 및 퍼지 융합부(600)는 각 채널의 화자 모델에 대응하는 우도 로그값을 산출하고 산출된 각 퍼지 소속도 값들을 융합한다(S300). 이에 따라 화자 인식부(700)는 퍼지 융합된 값의 최대치에 대응하는 음성을 발성한 사용자를 화자로 인식한다.
- <39> 도 3은 본 발명의 실시예에 따른 도 2의 화자 음성 데이터 등록 단계(S100)를 보다 상세하게 도시한 흐름도이다.
- <40> 먼저, 끝점 검출부(300)는 사용자가 여러 문장을 읽어서 다채널 마이크로폰(200)으로부터 온라인으로 출력되는 화자의 음성을 취득한다(S120). 즉, 기존에는 오프라인 상에서 사용자의 음성을 미리 등록해 놓고 각 화자의 모델을 구축하는 반면, 본 발명에서의 화자 등록은 다채널 마이크로폰(200)을 통해 각각 이루어지며 온라인상에서 각 화자의 모델을 구축하는 것을 특징으로 한다.
- <41> 또한 끝점 검출부(300)는 각 채널별로 등록된 음성에 대해, 끝점검출 알고리즘(Endpoint detection)을 이용하여 음성의 시작점 및 끝점을 검출한다(S140). 뿐만 아니라, 끝점 검출부(300)는 시작점과 끝점으로 구분된 음성의 문장들에 대해 윈너필터(Winer filter)를 이용하여 잡음 요소를 각각 제거한다(S160).
- <42> 도 4는 본 발명의 실시예에 따른 도 2의 화자 모델 구축 단계(S200)를 보다 상세하게 도시한 흐름도이다.
- <43> 먼저, 특징 추출부(422,442,462,482)는 각 채널별로 등록된 음성 데이터로부터 스펙트럼 기반 청각특성을 적용한 멜 캡스트럼(MFCC: Mel-Frequency Cepstral Coefficients)을 이용하여 특징 값들을 추출한다(S220).



<44> 화자 모델 구축부(424,444,464,484)는 화자별로 가우시안 혼합모델(GMM: Gaussian Mixture Model)을 이용하여 채널별 음성에 대한 화자 모델을 구축한다(S240). 가우시안 혼합모델(GMM)은 아래 수학적 식 1과 같이 나타낼 수 있다.

**수학적 식 1**

<45> 
$$p(\vec{x}|\lambda_s) = \sum_{i=1}^M w_i b_i(\vec{x})$$

<46> 
$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2} (\vec{x}-\mu_i)^T (\Sigma_i)^{-1} (\vec{x}-\mu_i)\right)$$

<47> 여기서  $w_i$ 는 혼합 가중치이고,  $b_i$ 는 가우시안 혼합모델을 통해 얻어진 확률 값이다.

<48> 여기서 밀도는 평균벡터와 공분산 행렬에 의해 파라미터화된 M개의 가우시안 혼합모델의 가중치된 선형적인 결합이다. 화자 모델 구축부(424,444,464,484)는 임의의 화자로부터 온라인 등록된 음성이 주어졌을 때, 가우시안 혼합모델의 파라미터를 추정한다. 본 실시예에서는 이를 위해 최대 우도 추정방법(maximum likelihood estimation)을 이용한다.

<49> 한편, 우도 로그값 변환부(426,446,466,486)는 각 화자 모델의 음성에 대해 우도의 로그 값으로 변환한다(S260). T개의 프레임으로 구성된 한 음성으로부터 얻어진 확률에 대해서, 가우시안 혼합모델의 우도 값은 아래 수학적 식 2와 같이 나타낼 수 있다.

**수학적 식 2**

<50> 
$$p(X|\lambda_s) = \prod_{t=1}^T p(\vec{x}_t|\lambda_s)$$

<51> 여기서 화자 모델의 파라미터는 가중치, 평균, 공분산으로 구성된  $\lambda_s = \{w_i, \mu_i, \Sigma_i\}$ ,  $i=1,2,\dots,M$  이다. 본 실시예에서 최대 우도 파라미터 추정은 최대 기대치(Expectation-Maximization: EM) 알고리즘을 이용함으로써 얻어질 수 있다. 수학적 식 2의 우도 값은 편리성을 위해 로그(log) 값으로 변환한다.

<52> 도 5는 본 발명의 실시예에 따른 도 2의 퍼지 융합 및 화자 인식 단계(S300)를 보다 상세하게 도시한 흐름도이다.

<53> 먼저, 퍼지 소속도 산출부(520,540,560,580)는 각각 채널별로 변환된 우도의 로그값에 기초하여 퍼지 소속도를 산출한다(S320). 이때 퍼지 소속도 산출부(520,540,560,580)는 시그모이드(sigmoid) 소속함수를 이용하여 아래 수학적 식 3과 같은 퍼지 소속도 값을 산출할 수 있다.

**수학적 식 3**

<54> 
$$S_k = \frac{1}{(1 + \exp^{-a(\log p(X|\lambda)_k - c)})}$$
,  $k=1,2,\dots,N$

<55> 여기서, a는 경사값을, c는 중심값을 나타낸다. 이들 값(a,c)은 온라인상에서 등록된 학습 음성데이터로부터 얻어진 특징 값의 통계에 의해서 얻어낼 수 있다.

<56> 한편, 퍼지 융합부(600)는 퍼지 소속도 산출부(520,540,560,580)에서 각각 산출한 퍼지 소속도 값들을 퍼지 융합한다(S340). 이때 퍼지 융합부(600)는 퍼지 적분을 이용하여 퍼지 소속도 산출부(520,540,560,580)에서 산출된 퍼지 소속도 값을 융합할 수 있다. 본 실시예에서 이용되는 퍼지 적분을 수식으로 전개하여 설명하면 다음과 같다.

<57> 집합함수  $g : P(S) \rightarrow [0,1]$ 은 아래 수학적 식 4를 만족하면 퍼지 척도라고 한다.



**수학식 4**

<58>  $g(\psi)=0, g(S)=1$

<59>  $g(A) \leq g(B)$ . if  $A \subset B$  and  $A, B \in P(S)$

<60> 이와 같은 정의로부터  $g^\lambda$  퍼지 척도는 임의의  $\lambda > -1$  에 대해서 아래 수학식 5의 성질을 만족한다.

**수학식 5**

<61>  $g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B)$ ,  $A, B \subset Y$  and  $A \cap B = \psi$

<62> 경계조건  $g(S)=1$ 이기 때문에,  $\lambda$ 는 아래 수학식 6의 다항식을 해석함으로써 결정되어진다.

**수학식 6**

<63>  $\lambda + 1 = \prod_{k=1}^N (1 + \lambda g^k)$

<64> 따라서, 최종적으로 퍼지 적분은 아래 수학식 7에 의해서 산출될 수 있다.

**수학식 7**

<65>  $\int_S h(S) \circ g(\cdot) = \sum_{k=1}^N (h(S_k) - h(S_{k+1}))g(A_k)$ ,  $h(S_{N+1})=0$

<66> 여기서  $h(\cdot)$ 의 값은  $h(S_1) \geq h(S_2) \geq \dots \geq h(S_N)$ 의 형태로 순서화되고,  $g(A_k)$ 의 값은  $1 < k \leq N$ 에 대해서 아래 수학식 8에 의해 반복적으로 결정되어진다.

**수학식 8**

<67>  $g(A_1) = g(S_1) = g^1$ ,  $g(A_k) = g^k + g(A_{k-1}) + \lambda g^k g(A_{k-1})$

<68> 이에 따라, 화자 인식부(700)는 퍼지 융합을 통해 인식된 결과에 따라 수학식 8에 의해 얻어진 N개의 다채널의 퍼지 융합 값들 중에서 최대치를 구함으로써 화자로서 인식한다(S360).

<69> 이상에서는 본 발명에서 특정의 바람직한 실시예에 대하여 도시하고 또한 설명하였다. 그러나 본 발명은 상술한 실시예에 한정되지 아니하며, 특허 청구의 범위에서 첨부하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 누구든지 다양한 변형 및 균등한 다 실시가 가능할 것이다. 따라서 본 발명의 진정한 기술적 보호범위는 첨부한 특허청구범위에 의해서만 정해져야 할 것이다.

**발명의 효과**

<70> 상기와 같은 본 발명에 따르면, 지능형 로봇에서 음성 기반 인간-로봇 상호작용을 수행하기 위해 다채널의 마이크로폰을 통해 온라인 화자등록과 로봇의 모든 방향에서 음성 인식이 가능하도록 구현함으로써, 전 방향에 대해 발생되는 화자의 음성을 보다 높은 성능으로 정확하게 취득할 수 있다.

<71> 또한 본 발명은 다채널 기반 온라인 화자등록/인식/퍼지 융합을 통해 보다 정확하게 화자를 인식할 수 있도록 함으로써, 잡음환경이나 원거리에서 화자 인식 성능의 저하를 최소화할 수 있다.

**도면의 간단한 설명**

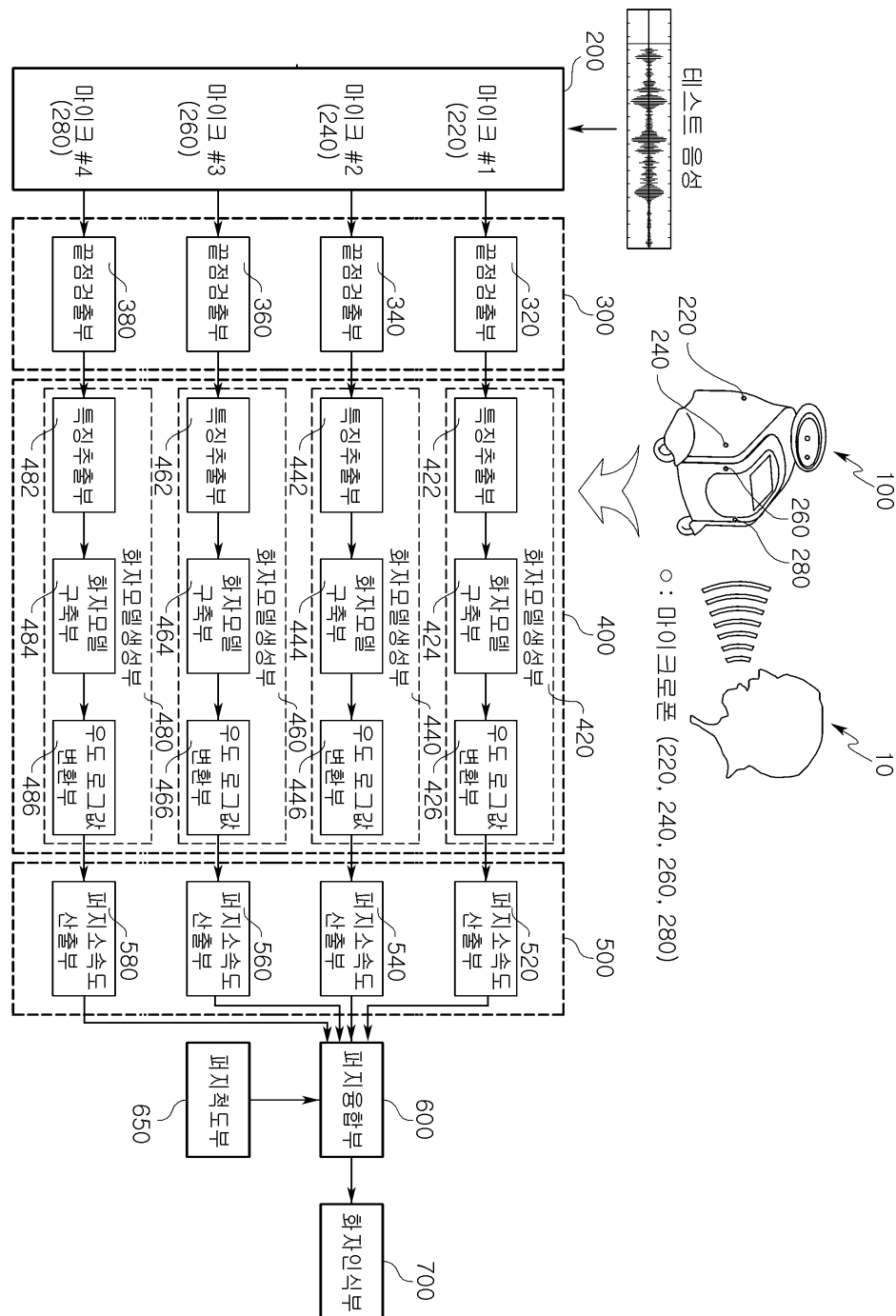
<1> 도 1은 본 발명의 바람직한 실시예에 따른 다채널 기반 퍼지 융합을 이용한 지능형 로봇의 화자 인식 장치를 도

시한 블록도,

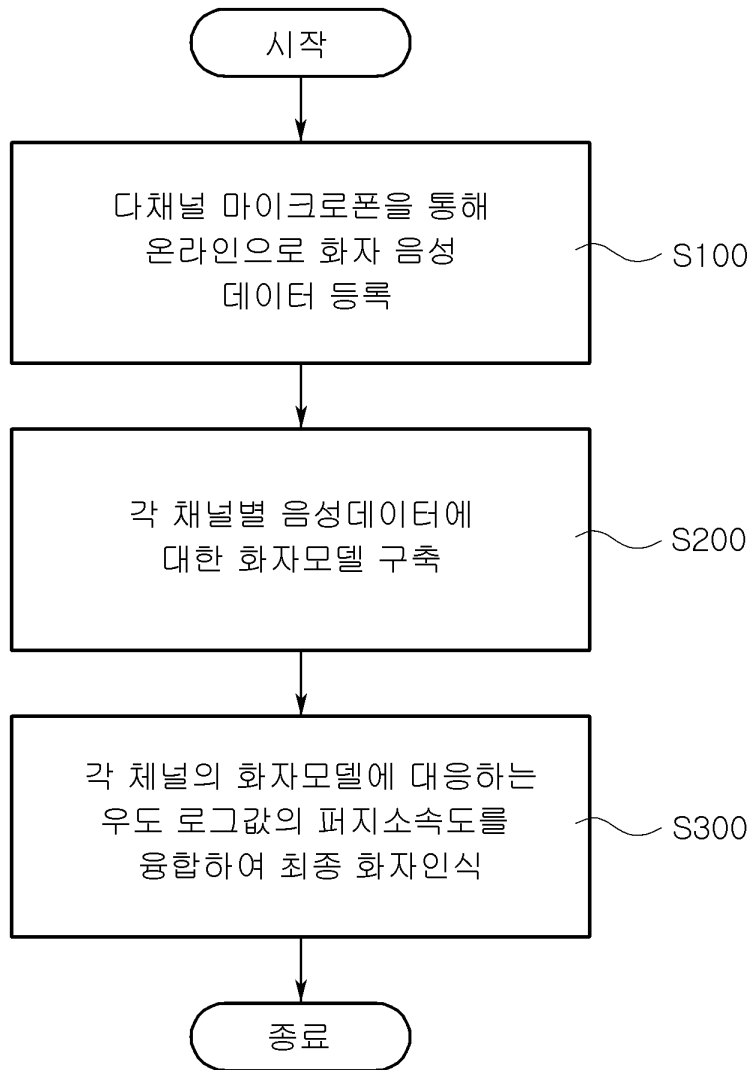
- <2> 도 2는 본 발명의 바람직한 실시예에 따른 다채널 기반 퍼지 융합을 이용한 지능형 로봇의 화자 인식 방법을 도시한 흐름도,
- <3> 도 3은 본 발명의 실시예에 따른 도 2의 화자 음성 데이터 등록 단계를 보다 상세하게 도시한 흐름도,
- <4> 도 4는 본 발명의 실시예에 따른 도 2의 화자 모델 구축 단계를 보다 상세하게 도시한 흐름도, 그리고
- <5> 도 5는 본 발명의 실시예에 따른 도 2의 퍼지 융합 및 화자 인식 단계를 보다 상세하게 도시한 흐름도이다.

도면

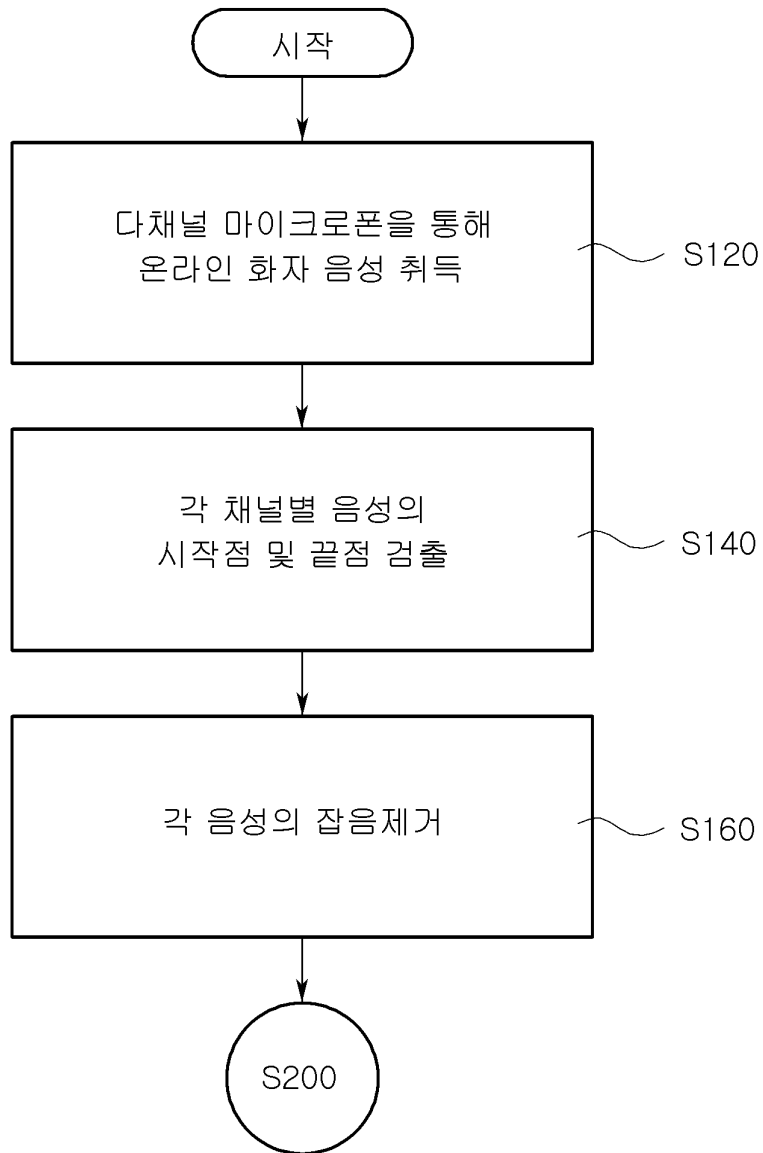
도면1



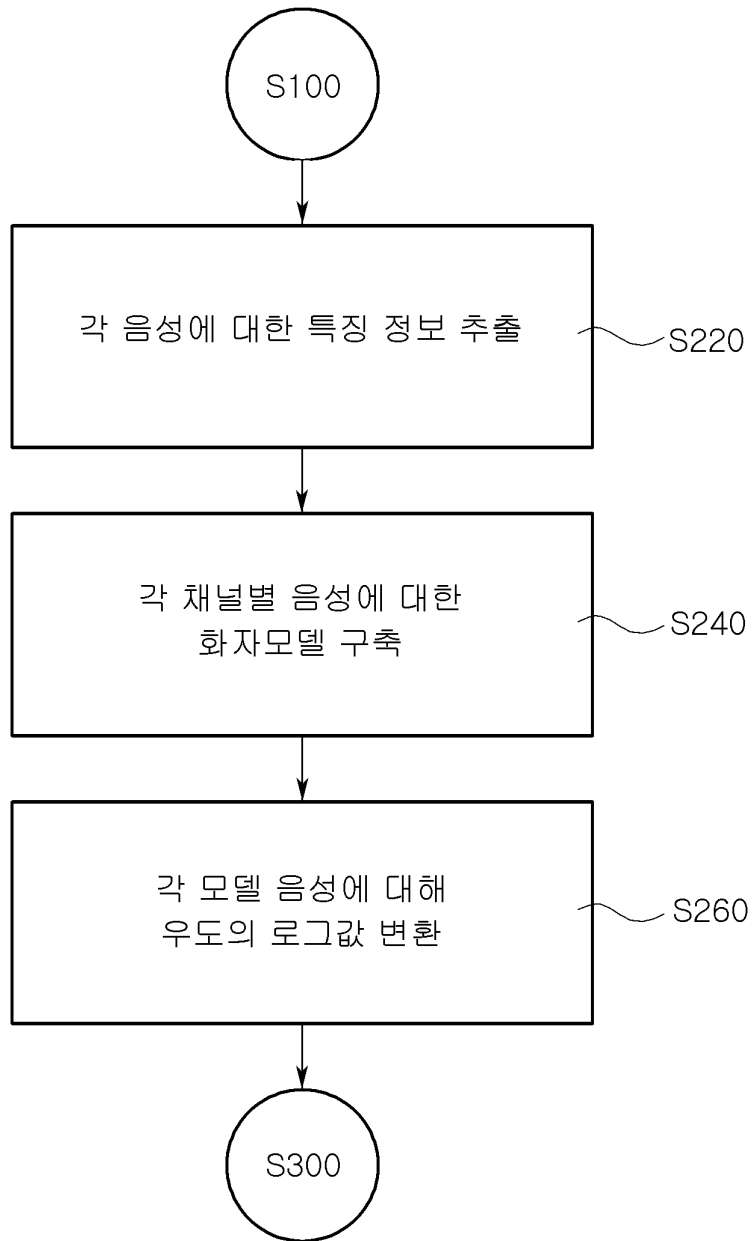
도면2



도면3



도면4



도면5

