

	(19) 대한민국특허청(KR) (12) 공개특허공보(A)	(11) 공개번호 10-2016-0049804 (43) 공개일자 2016년05월10일
(51) 국제특허분류(Int. Cl.) G06F 3/16 (2006.01) G10L 15/00 (2006.01)	(71) 출원인 현대모비스 주식회사 서울특별시 강남구 테헤란로 203 (역삼동)	
(21) 출원번호 10-2014-0147474 (22) 출원일자 2014년10월28일 심사청구일자 없음	(72) 발명자 권오현 경기도 용인시 기흥구 마북로240번길 17-2	
	(74) 대리인 특허법인우인	

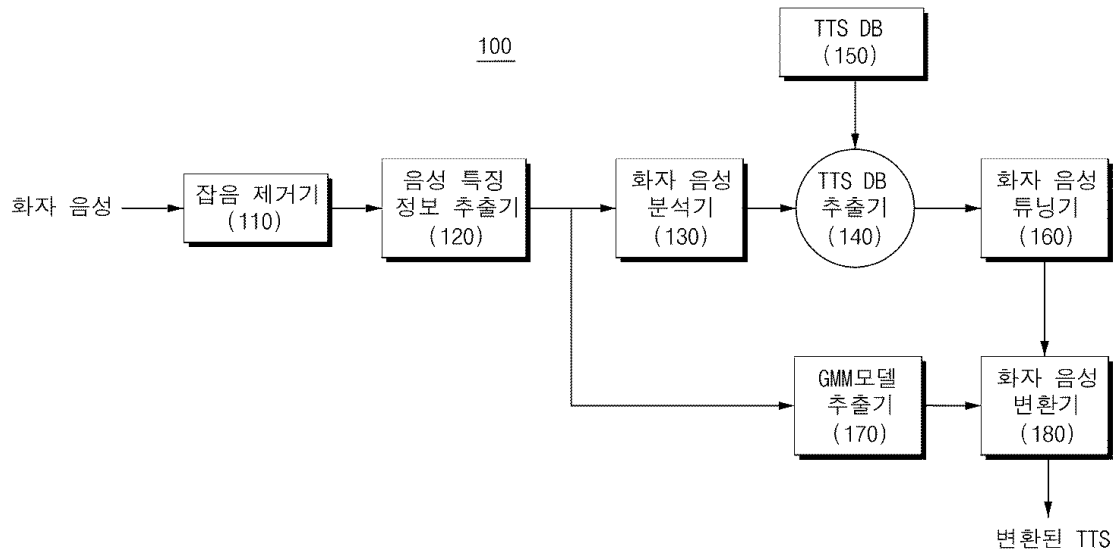
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 **사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 장치 및 방법**

(57) 요약

본 발명은 사용자의 음성으로부터 얻은 특성 정보를 기초로 TTS 서비스를 제공하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 장치 및 방법을 제안한다. 본 발명에 따른 대상 정보 음성 출력 제어 장치는 사용자의 음성 정보를 기초로 사용자의 특성 정보를 생성하는 특성 정보 생성부; 특성 정보를 기초로 텍스트 형태의 제1 대상 정보로부터 음성 형태의 제2 대상 정보를 생성하는 대상 정보 생성부; 및 제2 대상 정보를 출력하는 대상 정보 출력부를 포함한다.

대표도



명세서

청구범위

청구항 1

사용자의 음성 정보를 기초로 상기 사용자의 특성 정보를 생성하는 특성 정보 생성부;

상기 특성 정보를 기초로 텍스트 형태의 제1 대상 정보로부터 음성 형태의 제2 대상 정보를 생성하는 대상 정보 생성부; 및

상기 제2 대상 정보를 출력하는 대상 정보 출력부

를 포함하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 장치.

청구항 2

제 1 항에 있어서,

상기 특성 정보 생성부는 상기 음성 정보로부터 포먼트(Formant) 정보, 주파수(Log f0) 정보, LPC(Linear Predictive Coefficient) 정보, 스펙트럼 포락선(Spectral Envelope) 정보, 에너지 정보, 발화 속도(Pitch Period) 정보 및 로그 스펙트럼(Log Spectrum) 정보 중 적어도 하나의 정보를 추출하며, 상기 적어도 하나의 정보를 기초로 상기 특성 정보를 실시간으로 생성하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 장치.

청구항 3

제 1 항에 있어서,

상기 특성 정보 생성부는 상기 특성 정보로 상기 사용자의 성별 정보, 상기 사용자의 연령 정보, 및 상기 사용자의 감정 정보 중 적어도 하나의 정보를 실시간으로 생성하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 장치.

청구항 4

제 1 항에 있어서,

상기 특성 정보 생성부는 상기 음성 정보로부터 잡음 정보를 제거한 뒤 상기 특성 정보를 생성하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 장치.

청구항 5

제 1 항에 있어서,

상기 특성 정보 생성부는 상기 음성 정보에 상응하는 입력 정보들과 각 입력 정보의 목표 정보를 학습(training)시켜 얻은 가중치 정보를 상기 음성 정보에 적용하여 상기 특성 정보를 생성하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 장치.

청구항 6

제 5 항에 있어서,

상기 특성 정보 생성부는 ANN(Artificial Neural Network) 알고리즘, EBP(Error Back Propagation) 알고리즘 및 경사하강법(Gradient Descent Method)을 이용하여 상기 가중치 정보를 획득하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 장치.

청구항 7

제 1 항에 있어서,

상기 대상 정보 생성부는 데이터베이스에서 상기 특성 정보에 대응하는 기준 정보를 추출하며, 상기 기준 정보

를 기초로 상기 제1 대상 정보를 음성으로 변환하여 얻은 정보를 튜닝하여 상기 제2 대상 정보를 생성하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 장치.

청구항 8

제 7 항에 있어서,

상기 대상 정보 생성부는 상기 기준 정보로부터 얻은 발화 속도(Pitch Period) 정보 또는 주파수(Log f0) 정보를 기초로 상기 제1 대상 정보를 음성으로 변환하여 얻은 정보를 튜닝하여 상기 제2 대상 정보를 생성하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 장치.

청구항 9

제 7 항에 있어서,

상기 대상 정보 생성부는 상기 기준 정보와 더불어 상기 특성 정보로부터 얻은 화자 식별 정보를 기초로 상기 제2 대상 정보를 생성하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 장치.

청구항 10

제 9 항에 있어서,

상기 대상 정보 생성부는 가우시안 혼합 모델(GMM)을 기초로 상기 화자 식별 정보를 획득하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 장치.

청구항 11

사용자의 음성 정보를 기초로 상기 사용자의 특성 정보를 생성하는 단계;

상기 특성 정보를 기초로 텍스트 형태의 제1 대상 정보로부터 음성 형태의 제2 대상 정보를 생성하는 단계; 및

상기 제2 대상 정보를 출력하는 단계

를 포함하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 방법.

청구항 12

제 11 항에 있어서,

상기 특성 정보를 생성하는 단계는 상기 음성 정보로부터 포먼트(Formant) 정보, 주파수(Log f0) 정보, LPC(Linear Predictive Coefficient) 정보, 스펙트럼 포락선(Spectral Envelope) 정보, 에너지 정보, 발화 속도(Pitch Period) 정보 및 로그 스펙트럼(Log Spectrum) 정보 중 적어도 하나의 정보를 추출하며, 상기 적어도 하나의 정보를 기초로 상기 특성 정보를 실시간으로 생성하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 방법.

청구항 13

제 11 항에 있어서,

상기 특성 정보를 생성하는 단계는 상기 특성 정보로 상기 사용자의 성별 정보, 상기 사용자의 연령 정보, 및 상기 사용자의 감정 정보 중 적어도 하나의 정보를 실시간으로 생성하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 방법.

청구항 14

제 11 항에 있어서,

상기 제2 대상 정보를 생성하는 단계는 데이터베이스에서 상기 특성 정보에 대응하는 기준 정보를 추출하며, 상기 기준 정보를 기초로 상기 제1 대상 정보를 음성으로 변환하여 얻은 정보를 튜닝하여 상기 제2 대상 정보를 생성하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 방법.

청구항 15

제 14 항에 있어서,

상기 제2 대상 정보를 생성하는 단계는 상기 기준 정보와 더불어 상기 특성 정보로부터 얻은 화자 식별 정보를 기초로 상기 제2 대상 정보를 생성하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 방법.

발명의 설명

기술 분야

[0001] 본 발명은 대상 정보를 음성으로 출력하는 제어 장치 및 방법에 관한 것이다. 보다 상세하게는, 차량에서 대상 정보를 음성으로 출력하는 제어 장치 및 방법에 관한 것이다.

배경 기술

[0002] 일반적으로 TTS(Text To Speech)는 문자 또는 기호를 음성으로 변환하여 들려주는 기술이다. TTS는 음소에 대한 발음 데이터베이스를 구축하고 이를 연결하여 연속된 음성을 생성하는데, 이때 음성의 크기, 길이 높낮이 등을 조절하여 자연스러운 음성을 합성하는 것이 관건이다.

[0003] 즉, TTS는 문자열(문장)을 음성으로 변환하는 문자-음성 변환 장치로서, 크게 언어 처리, 운율 생성, 파형 합성 등의 3단계로 나누어 지는데, 텍스트가 입력되면 언어 처리 과정에서 입력된 문서의 문법적 구조를 분석하고, 분석된 문서 구조에 의해 사람이 읽는 것과 같은 운율을 생성하고, 생성된 운율에 따라 저장된 음성 DB의 기본 단위를 모아 합성음을 생성한다.

[0004] TTS는 대상 어휘에 제한이 없으며, 일반적인 문자 형태의 정보를 음성으로 변환하는 것이므로, 시스템의 구현시 음성학, 음성 분석, 음성 합성 및 음성인식 기술 등이 접목되어 보다 자연스럽고 다양한 음성이 출력된다.

[0005] 그러나 이러한 종래의 TTS를 제공하는 단말은 문자 메시지 등의 음성을 출력하는 경우 상대방이 누구인지 관계 없이 기설정된 항상 동일한 음성으로 출력하기 때문에 다양한 사용자의 욕구를 만족시키지 못하는 문제점이 있었다.

[0006] 한국공개특허 제2011-0032256호는 TTS 안내 방송 장치에 대하여 제안하고 있다. 그러나 이 장치는 지정된 텍스트를 음성으로 단순 변환하는 장치에 불과하기 때문에 상기한 문제점을 해결할 수 없다.

발명의 내용

해결하려는 과제

[0007] 본 발명은 상기한 문제점을 해결하기 위해 안출된 것으로서, 사용자의 음성으로부터 얻은 특성 정보를 기초로 TTS(Text To Speech) 서비스를 제공하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 장치 및 방법을 제안하는 것을 목적으로 한다.

[0008] 그러나 본 발명의 목적은 상기에 언급된 사항으로 제한되지 않으며, 언급되지 않은 또 다른 목적들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0009] 본 발명은 상기한 목적을 달성하기 위해 안출된 것으로서, 사용자의 음성 정보를 기초로 상기 사용자의 특성 정보를 생성하는 특성 정보 생성부; 상기 특성 정보를 기초로 텍스트 형태의 제1 대상 정보로부터 음성 형태의 제2 대상 정보를 생성하는 대상 정보 생성부; 및 상기 제2 대상 정보를 출력하는 대상 정보 출력부를 포함하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 장치를 제안한다.

[0010] 바람직하게는, 상기 특성 정보 생성부는 상기 음성 정보로부터 포먼트(Formant) 정보, 주파수(Log f0) 정보, LPC(Linear Predictive Coefficient) 정보, 스펙트럼 포락선(Spectral Envelope) 정보, 에너지 정보, 발화 속도(Pitch Period) 정보 및 로그 스펙트럼(Log Spectrum) 정보 중 적어도 하나의 정보를 추출하며, 상기 적어도 하나의 정보를 기초로 상기 특성 정보를 실시간으로 생성한다.

[0011] 바람직하게는, 상기 특성 정보 생성부는 상기 특성 정보로 상기 사용자의 성별 정보, 상기 사용자의 연령 정보, 및 상기 사용자의 감정 정보 중 적어도 하나의 정보를 실시간으로 생성한다.

- [0012] 바람직하게는, 상기 특성 정보 생성부는 상기 음성 정보로부터 잡음 정보를 제거한 뒤 상기 특성 정보를 생성한다.
- [0013] 바람직하게는, 상기 특성 정보 생성부는 상기 음성 정보에 상응하는 입력 정보들과 각 입력 정보의 목표 정보를 학습(training)시켜 얻은 가중치 정보를 상기 음성 정보에 적용하여 상기 특성 정보를 생성한다.
- [0014] 바람직하게는, 상기 특성 정보 생성부는 ANN(Artificial Neural Network) 알고리즘, EBP(Error Back Propagation) 알고리즘 및 경사하강법(Gradient Descent Method)을 이용하여 상기 가중치 정보를 획득한다.
- [0015] 바람직하게는, 상기 대상 정보 생성부는 데이터베이스에서 상기 특성 정보에 대응하는 기준 정보를 추출하며, 상기 기준 정보를 기초로 상기 제1 대상 정보를 음성으로 변환하여 얻은 정보를 튜닝하여 상기 제2 대상 정보를 생성한다.
- [0016] 바람직하게는, 상기 대상 정보 생성부는 상기 기준 정보로부터 얻은 발화 속도(Pitch Period) 정보 또는 주파수(Log f0) 정보를 기초로 상기 제1 대상 정보를 음성으로 변환하여 얻은 정보를 튜닝하여 상기 제2 대상 정보를 생성한다.
- [0017] 바람직하게는, 상기 대상 정보 생성부는 상기 기준 정보와 더불어 상기 특성 정보로부터 얻은 화자 식별 정보를 기초로 상기 제2 대상 정보를 생성한다.
- [0018] 바람직하게는, 상기 대상 정보 생성부는 가우시안 혼합 모델(GMM)을 기초로 상기 화자 식별 정보를 획득한다.
- [0019] 또한 본 발명은 사용자의 음성 정보를 기초로 상기 사용자의 특성 정보를 생성하는 단계; 상기 특성 정보를 기초로 텍스트 형태의 제1 대상 정보로부터 음성 형태의 제2 대상 정보를 생성하는 단계; 및 상기 제2 대상 정보를 출력하는 단계를 포함하는 것을 특징으로 하는 사용자의 음성 특성을 이용한 대상 정보 음성 출력 제어 방법을 제안한다.
- [0020] 바람직하게는, 상기 특성 정보를 생성하는 단계는 상기 음성 정보로부터 포먼트(Formant) 정보, 주파수(Log f0) 정보, LPC(Linear Predictive Coefficient) 정보, 스펙트럼 포락선(Spectral Envelope) 정보, 에너지 정보, 발화 속도(Pitch Period) 정보 및 로그 스펙트럼(Log Spectrum) 정보 중 적어도 하나의 정보를 추출하며, 상기 적어도 하나의 정보를 기초로 상기 특성 정보를 실시간으로 생성한다.
- [0021] 바람직하게는, 상기 특성 정보를 생성하는 단계는 상기 특성 정보로 상기 사용자의 성별 정보, 상기 사용자의 연령 정보, 및 상기 사용자의 감정 정보 중 적어도 하나의 정보를 실시간으로 생성한다.
- [0022] 바람직하게는, 상기 특성 정보를 생성하는 단계는 상기 음성 정보로부터 잡음 정보를 제거한 뒤 상기 특성 정보를 생성한다.
- [0023] 바람직하게는, 상기 특성 정보를 생성하는 단계는 상기 음성 정보에 상응하는 입력 정보들과 각 입력 정보의 목표 정보를 학습(training)시켜 얻은 가중치 정보를 상기 음성 정보에 적용하여 상기 특성 정보를 생성한다.
- [0024] 바람직하게는, 상기 특성 정보를 생성하는 단계는 ANN(Artificial Neural Network) 알고리즘, EBP(Error Back Propagation) 알고리즘 및 경사하강법(Gradient Descent Method)을 이용하여 상기 가중치 정보를 획득한다.
- [0025] 바람직하게는, 상기 제2 대상 정보를 생성하는 단계는 데이터베이스에서 상기 특성 정보에 대응하는 기준 정보를 추출하며, 상기 기준 정보를 기초로 상기 제1 대상 정보를 음성으로 변환하여 얻은 정보를 튜닝하여 상기 제2 대상 정보를 생성한다.
- [0026] 바람직하게는, 상기 제2 대상 정보를 생성하는 단계는 상기 기준 정보로부터 얻은 발화 속도(Pitch Period) 정보 또는 주파수(Log f0) 정보를 기초로 상기 제1 대상 정보를 음성으로 변환하여 얻은 정보를 튜닝하여 상기 제2 대상 정보를 생성한다.
- [0027] 바람직하게는, 상기 제2 대상 정보를 생성하는 단계는 상기 기준 정보와 더불어 상기 특성 정보로부터 얻은 화자 식별 정보를 기초로 상기 제2 대상 정보를 생성한다.
- [0028] 바람직하게는, 상기 제2 대상 정보를 생성하는 단계는 가우시안 혼합 모델(GMM)을 기초로 상기 화자 식별 정보를 획득한다.

발명의 효과

- [0029] 본 발명은 사용자의 음성으로부터 얻은 특성 정보를 기초로 TTS(Text To Speech) 서비스를 제공함으로써 다음

효과를 얻을 수 있다.

[0030] 첫째, 일방적인 방식에서 벗어나 양방향으로 소통하여 자연스러운 음성인식 시스템을 구현할 수 있다.

[0031] 둘째, 시스템에서 운전자의 성별, 연령, 성향 등에 맞춰진 TTS 서비스를 제공함으로써 차량의 음성인식 시스템에 기계적이지 않고 친숙함과 알아듣기 쉬운 음성을 제공할 수 있다.

도면의 간단한 설명

[0032] 도 1은 본 발명의 일실시예에 따른 차량용 음성 안내 제공 시스템의 내부 구성을 도시한 개념도이다.

도 2와 도 3은 도 1에 도시된 차량용 음성 안내 제공 시스템을 구성하는 화자 음성 분석기를 설명하기 위한 참고도이다.

도 4는 본 발명의 일실시예에 따른 차량용 음성 안내 제공 시스템의 작동 방법을 도시한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0033] 이하, 본 발명의 바람직한 실시예를 첨부된 도면들을 참조하여 상세히 설명한다. 우선 각 도면의 구성요소들에 참조 부호를 부가함에 있어서, 동일한 구성요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부호를 가지도록 하고 있음에 유의해야 한다. 또한, 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다. 또한, 이하에서 본 발명의 바람직한 실시예를 설명할 것이나, 본 발명의 기술적 사상은 이에 한정하거나 제한되지 않고 당업자에 의해 변형되어 다양하게 실시될 수 있음은 물론이다.

[0034] 본 발명은 차량 내 운전자 개인의 음성 특징을 분석하여 보다 더 자연스럽게 친숙한 음성 안내 서비스를 제공하는 것을 목적으로 한다.

[0035] 도 1은 본 발명의 일실시예에 따른 차량용 음성 안내 제공 시스템의 내부 구성을 도시한 개념도이다.

[0036] 차량용 음성 안내 제공 시스템(100)은 운전자의 음성을 이용하여 현재 운전자의 음성과 유사한 패턴으로 음성 안내를 제공해 주는 시스템으로서, 도 1에 도시한 바와 같이 잡음 제거기(110), 음성 특징 정보 추출기(120), 화자 음성 분석기(130), TTS DB 추출기(140), TTS DB(150), 화자 음성 튜닝기(160), GMM 모델 추출기(170) 및 화자 음성 변환기(180)를 포함한다.

[0037] 일반적으로 차량 내 내비게이션 안내 음성이나 음성인식 프롬프트 음성의 경우 이미 생산시 고정된 특정 TTS DB를 사용하고 있다. 그로 인해 나이별, 성별, 운전자의 성향별 음성 안내에 대한 소비자의 요구(Needs)를 적절하게 충족시키지 못하고 있는 상황이다. 예를 들자면, 나이는 노년층에게 발랄한 20대의 빠른 음성은 자칫 알아듣기 어려울 수 있으며, 젊은층에게 온화한 50대의 느린 음성은 지루하고 개성이 없어 보일 수 있다.

[0038] 본 발명에 따른 차량용 음성 안내 제공 시스템(100)은 젊은층과 중년층과 노년층 및 남성과 여성, 그리고 성향이 활발하거나 온화한 운전자에게 기계적인 TTS 안내 음성이 아닌 친숙하며 알아듣기 쉬운 음성의 품질을 제공하는 것을 목적으로 한다.

[0039] 또한 차량용 음성 안내 제공 시스템(100)은 양방향 소통 방식으로 기술이 변화함에 따라 음성인식이라는 화자 식별 기능을 이용해 운전자를 구별해서 운전자에게 적합한 기능을 먼저 제안해 인공지능 추세에 맞춰갈 수 있도록 하는 것을 목적으로 한다.

[0040] 이하 도 1을 참조하여 보다 자세하게 설명한다.

[0041] 잡음 제거기(110)는 화자의 음성 정보가 입력되면 이 음성 정보로부터 잡음 성분을 제거하는 기능을 수행한다. 잡음 제거기(110)는 차량 내의 노이즈를 제거하여 보다 더 명확한 운전자의 음성을 취득한다.

[0042] 음성 특징 정보 추출기(120)는 잡음 성분이 제거된 음성 정보로부터 화자의 음성 특징 정보를 추출하는 기능을 수행한다. 음성 특징 정보 추출기(120)는 화자의 나이, 성별, 성향 등을 분석하기 위해 각 개인별 음성의 특징 정보를 추출한다.

[0043] 음성 특징 정보 추출기(120)는 음성 정보로부터 포먼트(Formant) 정보, 주파수(Log f0) 정보, LPC(Linear Predictive Coefficient) 정보, 스펙트럼 포락선(Spectral Envelope) 정보, 에너지(Energy) 정보, 발화 속도(Pitch Period) 정보, 로그 스펙트럼(Log Spectrum) 정보 등의 음성 특징 정보를 추출한다.

- [0044] 화자 음성 분석기(130)는 음성 특징 정보 추출기(120)에 의해 추출된 음성 특징 정보를 이용하여 화자의 나이, 성별, 성향 등을 분류(Classification)하는 기능을 수행한다. 화자 음성 분석기(130)는 성별을 구분할 때 Log f0 정보를 사용할 수 있는데, Log f0 평균값이 120Hz ~ 240Hz이면 여성으로 판단할 수 있으며, Log f0 평균값이 0Hz ~ 120Hz이면 남성으로 판단할 수 있다.
- [0045] 화자 음성 분석기(130)는 음성 특징 정보 추출기(120)에 의해 개인별 음성 특징 정보가 추출되면 인공 신경망(ANN; Artificial Neural Network) 알고리즘을 이용하여 모델링(Modeing)을 수행함으로써 일반화된 나이별, 성별, 성향별 등으로 분석된 인공 신경망 알고리즘의 가중치(Weight) 정보를 추출할 수 있다. 화자 음성 분석기(130)는 이렇게 추출된 일반화된 가중치 정보(즉, 인공 신경망 알고리즘을 이용한 모델링 결과 데이터)를 토대로 실시간으로 입력되는 운전자의 음성 특징 정보를 추출하여 화자의 나이, 성별, 성향 등을 추정할 수 있다.
- [0046] 화자 음성 분석기(130)는 화자의 나이, 성별, 성향 등을 추정하기 위해 인공 신경망 알고리즘으로 나이 분석용 신경망(Neural Network), 성별 분석용 신경망, 성향 분석용 신경망 등을 이용할 수 있다.
- [0047] 이하 도 2와 도 3을 참조하여 화자 음성 분석기(130)에 대하여 부연 설명한다.
- [0048] 도 2와 도 3은 도 1에 도시된 차량용 음성 안내 제공 시스템을 구성하는 화자 음성 분석기를 설명하기 위한 참고도이다.
- [0049] 인공 신경망(ANN; Artificial Neural Network) 알고리즘은 인간의 두뇌 작용을 신경 세포들 간의 연결 관계로 모델링하고 구분하는 알고리즘을 말한다. 본 실시예에서는 화자 음성 분석기(130)가 다음 두 단계를 순차적으로 수행하여 인공 신경망 알고리즘을 구현한다. 도 2는 본 발명에 적용될 인공 신경망 알고리즘에서 인공 신경 회로망의 뉴런(처리 요소)의 구조를 설명하기 위한 참고도이다.
- [0050] 1. 학습 단계(Training, Modeling)
- [0051] 화자 음성 분석기(130)는 학습 단계에서 많은 양의 입력 벡터와 목표 벡터를 주어진 신경망 네트워크에 입력하여 패턴 분류를 시키고, 이에 따라 최적화된 연결 가중치(Weight; 220)를 획득한다.
- [0052] 2. 판별(Classification)
- [0053] 화자 음성 분석기(130)는 판별 단계에서 학습된 가중치(220)와 입력 벡터(210) 간 연산식(230)을 통해 출력값(240)을 산출한다. 화자 음성 분석기(130)는 가중치(220)와 입력 벡터(210) 간 차이값을 계산하여 가장 근사한 출력(Output)을 최종적인 결과로 판별하여 산출할 수 있다. 연산식(230)에서 θ 는 임계값을 의미한다.
- [0054] 화자 음성 분석기(130)는 인공 신경망 알고리즘을 이용하여 화자의 음성 특징 정보로부터 화자의 나이, 성별, 성향 등을 분석할 때 다층 퍼셉트론(Multi-Layer Perceptron)을 적용할 수 있으며, 특히 오류 역전파(EBP; Error Back Propagation) 알고리즘을 적용할 수 있다. 이하 도 3을 참조하여 보다 자세하게 설명한다. 도 3은 본 발명에 적용될 EBP 알고리즘의 구조를 도시한 참고도이다.
- [0055] 종래 음성과 관련된 퍼셉트론 이론은 음성을 인식하거나(음성을 입력받으면 어떤 내용의 음성인지를 판단), 사람의 감정을 판별하는 용도로 쓰여 왔다.
- [0056] 다층 퍼셉트론(multilayer perceptron)은 입력층과 출력층 사이에 하나 이상의 중간층이 존재하는 신경망이다. 네트워크는 입력층, 은닉층, 출력층 방향으로 연결되어 있으며, 각 층 내의 연결과 출력층에서 입력층으로의 직접적인 연결은 존재하지 않는 전방향(Feedforward) 네트워크이다.
- [0057] 이러한 다층 퍼셉트론을 화자 음성 분석기(130)에 적용하기 위해 본 발명에서는 EBP 알고리즘을 채택한다.
- [0058] EBP 알고리즘은 입력층과 출력층 사이에 하나 이상의 은닉층을 가지며, 수학적 1에서 보는 바와 같이 일반화된 델타 규칙을 이용하여 원하는 목표값(D_{pj})과 실제 출력값(O_{pj}) 사이의 오차제곱합으로 정의된 Cost function 값을 경사 하강 추적법(gradient-descent method)에 의해 최소화하는 방향으로 학습을 진행시켜서 원하는 가중치 값을 얻는 방법을 말한다.

수학식 1

$$E = \sum_p E_p, \quad (E_p = \frac{1}{2} \sum_j (D_{pj} - O_{pj})^2)$$

[0059]

[0060]

상기에서, p 는 p 번째 학습 패턴을 의미하며, E_p 는 p 번째 패턴에 대한 오차를 의미한다. 그리고 D_{pj} 는 p 번째 패턴에 대한 j 번째 요소를 의미하며, O_{pj} 는 실제 출력의 j 번째 요소를 의미한다.

[0061]

화자 음성 분석기(130)는 이상 설명한 EBP 알고리즘을 이용함으로써 은닉층의 학습을 위해 출력층에서 발생한 오류를 이용하여 은닉층 오차를 계산하고, 다시 이 값을 입력층으로 역으로 전파시켜서 출력층의 오차가 원하는 수준이 될 때까지 반복시켜서 최적화된 가중치 값을 얻을 수 있다.

[0062]

화자 음성 분석기(130)는 EBP 알고리즘을 이용하여 학습(Training) 단계를 다음 절차에 따라 수행할 수 있다.

[0063]

먼저 제1 단계에서 가중치(Weight)와 임계치를 초기화시킨다.

[0064]

이후 제2 단계에서 입력 벡터(Input Vector)와 목표 벡터(Target Vector)를 제시한다.

[0065]

이후 제3 단계에서 제시된 입력 벡터를 이용하여 은닉층(Hidden Layer) j 번째 뉴런으로의 입력값을 계산한다. 이때 수학식 2가 이용될 수 있다.

수학식 2

$$net_{pj} = \sum_{i=0}^{N-1} W_{ji} X_{pi} - \theta_j$$

[0066]

[0067]

상기에서, net_{pj} 는 은닉층 j 번째 뉴런으로의 입력값을 의미한다. W_{ji} 는 j 번째 뉴런에서 i 번째 뉴런으로의 연결 가중치를 의미하며, X_{pi} 는 입력 벡터를 의미한다. 그리고 θ_j 는 임계값을 의미한다.

[0068]

이후 제4 단계에서 시그모이드(Sigmoid) 함수를 이용하여 은닉층의 출력(O_{pj})을 계산한다.

[0069]

이후 제5 단계에서 은닉층의 출력을 이용하여 출력층 뉴런 k 로의 입력값을 계산한다. 이때 수학식 3이 이용될 수 있다.

수학식 3

$$net_{pk} = \sum_{j=0}^{L-1} W_{kj} O_{pj} - \theta_k$$

[0070]

[0071]

상기에서, net_{pk} 는 출력층 뉴런 k 로의 입력값을 의미한다.

[0072]

이후 제6 단계에서 시그모이드 함수($f'()$)를 사용하여 출력층의 출력(O_{pk})을 계산한다.

[0073]

이후 제7 단계에서 입력 패턴의 목표 출력과 실제 출력 사이의 오차를 계산하고 출력층 오차합을 학습 패턴의 오차로 누적시킨다. 이때 수학식 4가 이용될 수 있다.

수학식 4

$$\delta_{pk} = (d_{pk} - O_{pk}) f'_k(net_{pk}) = (d_{pk} - O_{pk}) O_{pk} (1 - O_{pk})$$

$$E = E + E_p, \quad (E_p = \sum_{k=1}^{M-1} \delta_{pk}^2)$$

[0074]

[0075]

상기에서 d_{pk} 는 입력 패턴의 목표 출력을 의미하며, O_{pk} 는 입력 패턴의 실제 출력을 의미한다. 그리고 δ_{pk} 는 목표 출력과 실제 출력 사이의 오차를 의미한다. E 는 출력층 오차합을 의미하며, E_p 는 학습 패턴의 오차를 의미한다.

[0076]

이후 제8 단계에서 출력층 오차값(δ_{pk}), 은닉층과 출력층의 가중치값(W_{kj}) 등을 이용하여 은닉층의 오차(δ_{pj})를 계산한다. 이때 수학식 5가 이용될 수 있다.

수학식 5

$$\delta_{pj} = f'_j(net_{pj}) \sum_{k=0}^{M-1} \delta_{pk} W_{kj} = \sum_{k=0}^{M-1} \delta_{pk} W_{kj} O_{pj} (1 - O_{pj})$$

[0077]

[0078]

이후 제9 단계에서 제4 단계와 제7 단계에서 구한 은닉층 뉴런 j 의 출력값(O_{pj})과 출력층의 오차값(δ_{pk})을 이용하여 출력층의 가중치(W_{kj})를 갱신한다. 이때 임계치도 조정한다. 이때 수학식 6이 이용될 수 있다.

수학식 6

$$W_{kj}(t+1) = W_{kj}(t) + \eta \delta_{pk} O_{pj}$$

$$\theta_k(t+1) = \theta_k(t) + \beta \delta_{pk}$$

[0079]

[0080]

상기에서 η 과 β 는 이득값을 의미하며, t 는 시각을 의미한다.

[0081]

이후 제10 단계에서 출력층에서와 마찬가지로 입력층과 은닉층의 가중치 값과 임계치 값을 갱신한다. 이때 수학식 7이 이용될 수 있다.

수학식 7

$$W_{ji}(t+1) = W_{ji}(t) + \eta \delta_{pj} X_{pi}$$

$$\theta_j(t+1) = \theta_j(t) + \beta \delta_{pj}$$

[0082]

[0083]

이후 제11 단계에서 모든 학습 패턴에 대하여 전부 학습할 때까지 2단계로 분기하여 반복 수행한다.

[0084]

이후 제12 단계에서 출력층의 오차합 E 가 허용값 이하이거나 최대 반복 횟수보다 크면 종료하며, 그렇지 않으면 제2 단계로 가서 이후 절차를 반복한다.

[0085]

한편 화자 음성 분석기(130)는 화자가 복수일 때 다층 퍼셉트론(multilayer perceptron)을 이용할 때 각 화자의 음성 특징 정보로부터 각 화자의 나이, 성별, 성향 등을 분석하는 것도 가능하다. 이하 이에 대해 설명한다.

[0086]

일반적인 노이즈 필터링 방법에 따르면, 음성인식 마이크 오픈 후 일정 시간 후에 음성인식 발화를 함으로써 음

성인식 전에 마이크로 들어오는 신호를 차량 내 노이즈라고 판단하고 신호에서 그 노이즈만 필터링시킨다.

- [0087] 그런데 차량 내에 운전자 방향으로 지향성 마이크가 달려 있지만 음성 발화 전의 잠깐의 시간동안 입력된 신호를 노이즈로 판단하기 때문에, 만일 음성인식 발화 시점에 운전자 외에 다른 좌석에서 발화시 음성이 섞이게 되어 음성인식률이 떨어지는 문제점이 있다.
- [0088] 그래서 본 발명에서는 차량 내 4개의 좌석 영역에 지향성 마이크를 각각 설치하고, 운전자 영역의 마이크의 입력 신호를 기준으로 다른 영역들의 마이크 신호를 노이즈로 판별하고 필터링한다. 신호를 처리하는 과정에서 실시간으로 운전자 영역의 운전자의 특징을 판별하여 멀티미디어 기기에서 운전자에게 적합한 정보를 제공하도록 한다.
- [0089] 이하에서 보다 자세하게 설명하며, 이하 설명에서는 운전석을 A 영역으로 정의하고, 조수석을 B 영역으로 정의하며, 운전석의 뒤쪽과 조수석의 뒤쪽을 각각 C 영역과 D 영역으로 정의한다.
- [0090] 운전자가 음성인식 기능을 시작시, A, B, C, D 영역의 마이크들이 동시에 오픈되면서 마이크로 4 영역의 음성 신호를 받는다. 사람의 음성이 아닌 차량 노이즈는 4 영역의 마이크에 입력되는 값이 거의 동일하므로 차량 노이즈 값을 A에서 필터링한다. 그리고 4 영역의 음성 목소리를 분석한다. 우선 4 영역의 성별을 나타내는 음성 벡터값을 분석하고, A 영역을 기준으로 B, C, D 영역에서 A 영역과 다른 성별을 나타내는 벡터값이 추출되면 A 영역에서 그 벡터값에 해당하는 신호를 필터링한다. 성별 분석이 완료되면 동일한 방법으로 연령, 기분/컨디션 등에 대해 분석한다.
- [0091] A 영역에서 운전자의 음성 신호가 가장 크겠지만, B, C, D 영역의 음성 신호가 있을 경우 A 영역에서 완벽하게 운전자의 목소리만 추출하기 어렵기 때문에 이 방법을 사용한다.
- [0092] 이때에는 상관관계(CORRELATION), ICA 기술, BEAM FORMING 기술 외의 다른 알고리즘을 사용하여 신호가 독립적인지 유사성을 띄고 있는지를 판별할 수 있다.
- [0093] 4개의 마이크를 통해 필터링을 하면서 화자의 개별 특성을 파악할 수 있고, 개별 특성을 파악한 정보를 이용한 노이즈 필터링으로 인식률을 높일 수 있다.
- [0094] 차량의 경우는 일반적으로 4개의 좌석이 지정되어 있고, 차량 내 음성인식 시스템은 보통 운전자가 사용하는데, 운전자의 음성인식 시스템 사용 중 나머지 좌석의 탑승자가 발화시 여러 명의 음성이 더해지므로 인식 시스템에서 운전자의 명령을 인식하기가 어렵다. 현재 일반적으로 쓰이는 음성인식 시스템에서는 음성인식 구간 앞에 음성이 없는 구간을 설정해 그 구간을 노이즈로 인식하고, 음성이 들어오는 구간에서 노이즈를 필터링하는 구조이다.
- [0095] 본 발명은 퍼셉트론 이론을 이용해 음성의 특징을 추출해 발화자의 특성을 식별하고, 그 데이터로 발화자에게 적합한 정보를 실시간으로 제공하는 기술이다. 퍼셉트론을 이용하면 ①화자의 특성에 따라 맞춤형 정보를 제공하거나, ②발화자 위치를 인식하고 그 위치에 발화자가 원하는 기능을 제공할 수 있다. 이하 ①과 ②에 대해 보다 자세하게 설명한다.
- [0096] 1. 화자 특성에 따른 맞춤형 정보 제공
- [0097] 다층 퍼셉트론을 이용해 시스템을 구성하면, 여러 명의 음성이 더해지더라도 운전자의 음성을 추출하는 것이 가능해진다. 이 방법은 운전자에 국한되지 않고 나머지 사람의 인식도 가능하다. 예시로 A 영역의 음성 특성만 추출하고, 나머지 B, C, D 영역의 음성 신호는 무시하는 경우이다.
- [0098] 퍼셉트론의 경우 미리 많은 DB를 바탕으로 BACK PROPAGATION 기법을 이용해 트레이닝된 알고리즘이 형성된 상태가 대전제이다.
- [0099] 퍼셉트론 모델링은 예로 20대 컨디션이 좋은 서울 여성의 수많은 음성을 분석해 특성(포먼트, 기본 주파수, 에너지값, LPC 값 등)을 추출해 인풋에 넣고, OUTPUT 타겟을 20대 컨디션이 좋은 서울 여성으로 하면 퍼셉트론 구조 내부적으로 BACK PROPAGATION 과정을 거쳐 적절한 WEIGHT 값이 결정된다. 이렇게 다양한 특성의 사람들을 트레이닝시키면 어떤 음성이 들어가더라도 트레이닝된 구조 안에서 특징을 찾아갈 수 있다. LPC 값은 선형 예측 부호화 값으로 인간 발성 모델에 근거한 음성 부호화 방식 중 하나로 26차원의 벡터를 갖는다.
- [0100] 특정 타겟의 수만은 음성의 formant, 기본 주파수, LPC 모델의 26차원 벡터값을 입력했을 때 역전개 과정을 거쳐 적절한 가중치 값들이 정하는 작업을 여러 타겟으로 반복한다(20대 컨디션 좋은 서울 여성, 30대 컨디션 안 좋은 경상도 지역의 남성...).

- [0101] 이 트레이닝 과정을 거치면 어떤 음성이 입력되더라도 그 음성의 특징 벡터들을 모델링한 퍼셉트론 구조에 입력하면 발화자의 특성을 알 수 있다.
- [0102] 좌석 선택의 기준은 PTT로 한다. PTT 버튼이 4개가 있다면 위치에 따라 해당 PTT 입력된 자리에 위치한 마이크에 입력된 음성이 분석해야 될 음성으로 판단하고 나머지는 노이즈로 판단해 필터링한다. 필터링된 음성으로 인식을 하여 발화자에게 최적의 정보를 제공하는데, 예를 들어 멀티 제품에 발화자가 명령할 경우, 주변 음식점을 찾는다면 발화자의 특성에 적합한 주변 음식점을 먼저 찾아주게 된다.
- [0103] 이상 설명한 내용을 정리하면 다음과 같은 특징 도출이 가능하다.
- [0104] 먼저, PTT 위치를 판별하며 음성 신호별 특성에 따른 벡터를 추출한다.
- [0105] 이후, 다층 퍼셉트론 구조에 4가지 신호의 특성 벡터를 입력한다.
- [0106] 이후, 각각 음성 신호마다의 특성을 추출한다.
- [0107] 이후, 기준 음성(A)과 다른 특성을 가질 경우 A 마이크 신호에서 다른 특성값을 노이즈로 판단하고 필터링한다.
- [0108] 이후, A 영역의 음성만 추출된 데이터로 음성 인식을 수행하며, 음성이 어떤 의미인지 판별한다.
- [0109] 이후, A 영역의 발화자의 명령어에 대해 최적화된 정보를 제공한다.
- [0110] 2. 발화자 위치를 인식하고 그 위치에 발화자가 원하는 기능을 제공
- [0111] 좌석 선택의 기준은 PTT로 한다. PTT 버튼이 4개가 있다면 위치에 따라 해당 PTT 입력된 자리에 위치한 마이크에 입력된 음성이 분석해야 될 음성으로 판단하고 나머지는 노이즈로 판단해 필터링한다. 예를 들어 공조의 경우 D 영역에 앉은 사람이 에어컨 온도 관련 명령을 할 경우 D 영역의 공조 장치에만 명령에 따라 공조 레벨이 변화게도 할 수 있다.
- [0112] 다시 도 1을 참조하여 설명한다.
- [0113] TTS DB(150)는 나이에 관련된 기준 특징 정보(10대, 20대, 30대, 40대, 50대, 60대, 70대 이상 등), 성별에 관련된 기준 특징 정보(남성, 여성 등), 성향에 관련된 기준 특징 정보(온화, 활발 등) 등을 저장하는 데이터베이스이다.
- [0114] TTS DB 추출기(140)는 화자 음성 분석기(130)에 의해 발견된 화자의 나이, 성별, 성향 등에 대응하는 정보를 TTS DB(150)로부터 검출하는 기능을 수행한다.
- [0115] 화자 음성 튜닝기(160)는 TTS DB(150)로부터 검출된 정보를 기초로 TTS 서비스를 위해 출력될 음성을 튜닝(tuning)하는 기능을 수행한다. 화자 음성 튜닝기(160)는 운전자의 음성으로부터 얻은 발화 속도 정보(Pitch Period), 주파수의 고저(高低)에 대한 정보(Log f0) 등을 출력하려는 음성에 적용하여 튜닝할 수 있다.
- [0116] GMM(Gaussian Mixture Model) 모델 추출기(170)는 음성 특징 정보 추출기(120)에 의해 추출된 화자의 음성 특징 정보를 기초로 가우시안 혼합 모델을 생성하는 기능을 수행한다.
- [0117] 화자 음성 변환기(180)는 화자 음성 튜닝기(160)에 의해 튜닝된 음성에 가우시안 혼합 모델을 적용하여 음성을 추가적으로 변환하는 기능을 수행한다. 본 발명에서는 화자 음성 튜닝기(160)에 의해 튜닝된 음성을 TTS 서비스를 위한 음성으로 제공할 수 있다. 그러나 본 발명은 이에 한정하지 않고 실시간으로 화자의 음성 특성이 적절하게 변환될 수 있도록 GMM(Gaussian Mixture Model)을 통해 화자의 음성을 추가적으로 변환하는 것도 가능하다.
- [0118] 이하 가우시안 혼합 모델을 이용한 화자 음성 변환기(180)에 대하여 부연 설명한다.
- [0119] $x \in \mathbb{R}^n$ 이라는 특정 랜덤 벡터의 가우시안 혼합 밀도(Gaussian Mixture Density)는 수학식 8과 같이 나타낼 수 있다.

수학식 8

$$p(x | \lambda) = \sum_{i=0}^Q \alpha_i b_i(x), \quad \sum_{i=0}^Q \alpha_i = 1, \quad \alpha_i \geq 0$$

상기에서 $p()$ 는 성분 파라미터로 평균과 분산을 가진 가우시안 함수를 의미한다. Q 는 단일 가우시안 밀도(Gaussian Density)의 총 갯수를 의미하며, α_i 는 단일 가우시안 밀도의 가중치를 의미한다.

여기서 $b_i(x)$ 를 단일 가우시안 밀도로 나타내면 수학식 9와 같이 정의된다.

수학식 9

$$b_i(x) = \frac{1}{(2\pi)^{n/2} |C_i|^{1/2}} \exp \left[-\frac{1}{2} (x - \mu_i)^T C_i^{-1} (x - \mu_i) \right]$$

μ_i : $n \times 1$ mean vector, C_i : $n \times n$ covariance matrix

그러므로 완성된 가우시안 혼합 밀도(Gaussian Mixture Density)는 다음 3가지 변수로 구성된다.

$$\lambda = \{ \alpha_i, \mu_i, C_i \}, \quad i = 1, \dots, Q$$

$x \in \mathbb{R}^n$ 를 TTS DB 추출기(140)에 의해 선별된 음성으로 정의하고 $y \in \mathbb{R}^n$ 를 운전자의 음성으로 정의하면, $z = (x, y)^T$ 는 TTS DB 추출기(140)에 의해 선별된 음성과 운전자 음성 간의 결합 밀도(joint density) 음성으로 정의할 수 있다. 이를 수학식으로 나타내면 다음과 같다.

수학식 10

$$p(z | \lambda) = \sum_{i=1}^Q \frac{\alpha_i}{(2\pi)^n |C_i|^{1/2}} \exp \left[-\frac{1}{2} (z - \mu_i)^T C_i^{-1} (z - \mu_i) \right]$$

$$\sum_{i=1}^Q \alpha_i = 1, \quad \alpha_i \geq 0$$

따라서 화자 음성 변환기(180)는 수학식 11과 같이 평균 제곱 오차(Mean Square Error)를 최소화하는 맵핑(Mapping) 함수 $F(x)$ 를 발견하는 것이다.

수학식 11

$$\varepsilon_{mse} = E \left[\|y - F(x)\|^2 \right]$$

$E[\dots]$ 는 기대값(expectation)을 의미하며, $F(x)$ 는 추정된(estimated) 음성의 스펙트럴 벡터(Spectral Vector)를 의미한다.

수학식 12

$$F(x) = E[y|x] = \sum_{i=1}^Q h_i(x) \left[\mu_i^y + C_i^{yx} C_i^{xx-1} (x - \mu_i^x) \right]$$

$$h_i(x) = \frac{\frac{\alpha_i}{(2\pi)^{n/2} |C_i^{xx}|^{1/2}} \exp \left[-\frac{1}{2} (x - \mu_i^x)^T C_i^{xx-1} (x - \mu_i^x) \right]}{\sum_{j=1}^Q \frac{\alpha_j}{(2\pi)^{n/2} |C_j^{xx}|^{1/2}} \exp \left[-\frac{1}{2} (x - \mu_j^x)^T C_j^{xx-1} (x - \mu_j^x) \right]}$$

$$C_i = \begin{bmatrix} C_i^{xx} & C_i^{xy} \\ C_i^{yx} & C_i^{yy} \end{bmatrix}, \quad \mu_i = \begin{bmatrix} \mu_i^x \\ \mu_i^y \end{bmatrix}$$

[0131]

[0132] 다음으로 도 1 내지 도 3을 참조하여 설명한 차량용 음성 안내 제공 시스템(100)의 작동 방법에 대하여 설명한다. 도 4는 본 발명의 일실시예에 따른 차량용 음성 안내 제공 시스템의 작동 방법을 도시한 흐름도이다.

[0133] 운전자가 특정 명령어를 발화하면(S405), 음성 특징 정보 추출기(120)가 화자의 음성으로부터 특징 정보를 추출한다(S410).

[0134] 이후 화자 음성 분석기(130)가 특징 정보로부터 실시간으로 성별, 연령, 성향 등을 분석한다(S415).

[0135] 이후 TTS DB 추출기(140)가 TTS DB(150)에서 각 분석 결과에 대응하는 정보를 선택한다(S420).

[0136] 이후 화자 음성 튜닝기(160)가 TTS DB 추출기(140)에 의해 선택된 정보들을 기초로 음성 변환된 정보를 튜닝한다(S425).

[0137] 이후 화자 음성 변환기(190)가 화자의 음성으로부터 얻은 GMM 모델을 기초로 튜닝된 음성을 운전자의 실제 음성에 가깝도록 변환한다(S430).

[0138] 이후 TTS 출력부(미도시)가 화자 음성 변환기(190)에 의해 변환된 음성을 출력한다(S435).

[0139] 이상 도 1 내지 도 4를 참조하여 본 발명의 일 실시 형태에 대하여 설명하였다. 이하에서는 이러한 일 실시 형태로부터 추론 가능한 본 발명의 바람직한 형태에 대하여 설명한다.

[0140] 본 발명의 바람직한 실시예에 따른 대상 정보 음성 출력 제어 장치는 특성 정보 생성부, 대상 정보 생성부, 대상 정보 출력부, 전원부 및 주제어부를 포함한다.

[0141] 전원부는 대상 정보 음성 출력 제어 장치를 구성하는 각 구성에 전원을 공급하는 기능을 수행한다. 주제어부는 대상 정보 음성 출력 제어 장치를 구성하는 각 구성의 전체 작동을 제어하는 기능을 수행한다. 대상 정보 음성 출력 제어 장치가 차량에 적용되는 것임을 참작할 때 전원부와 주제어부는 본 실시예에서 구비되지 않아도 무방하다.

[0142] 특성 정보 생성부는 사용자의 음성 정보를 기초로 사용자의 특성 정보를 생성하는 기능을 수행한다. 특성 정보 생성부는 도 1의 음성 특징 정보 추출기(120)에 대응하는 개념이다.

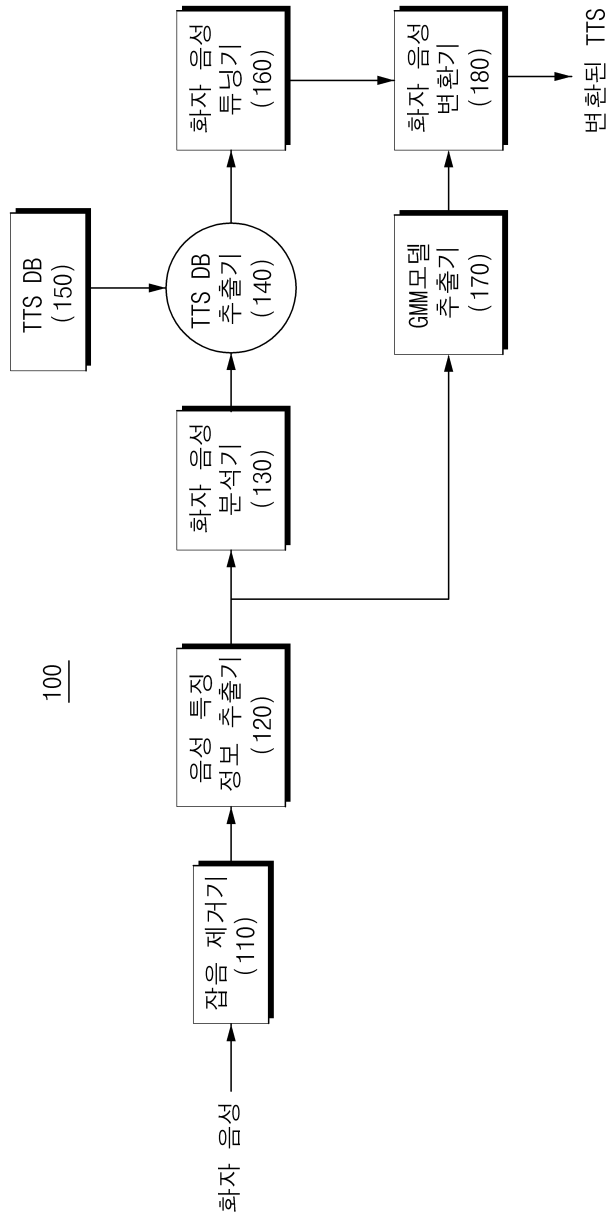
[0143] 특성 정보 생성부는 음성 정보로부터 포먼트(Formant) 정보, 주파수(Log f0) 정보, LPC(Linear Predictive Coefficient) 정보, 스펙트럼 포락선(Spectral Envelope) 정보, 에너지 정보, 발화 속도(Pitch Period) 정보 및 로그 스펙트럼(Log Spectrum) 정보 중 적어도 하나의 정보를 추출하며, 적어도 하나의 정보를 기초로 특성 정보를 실시간으로 생성할 수 있다.

[0144] 특성 정보 생성부는 특성 정보로 사용자의 성별 정보, 사용자의 연령 정보, 및 사용자의 감정 정보 중 적어도 하나의 정보를 실시간으로 생성할 수 있다. 이러한 특성 정보 생성부는 도 1의 음성 특징 정보 추출기(120)와 화자 음성 분석기(130)의 결합 구성에 대응하는 개념이다.

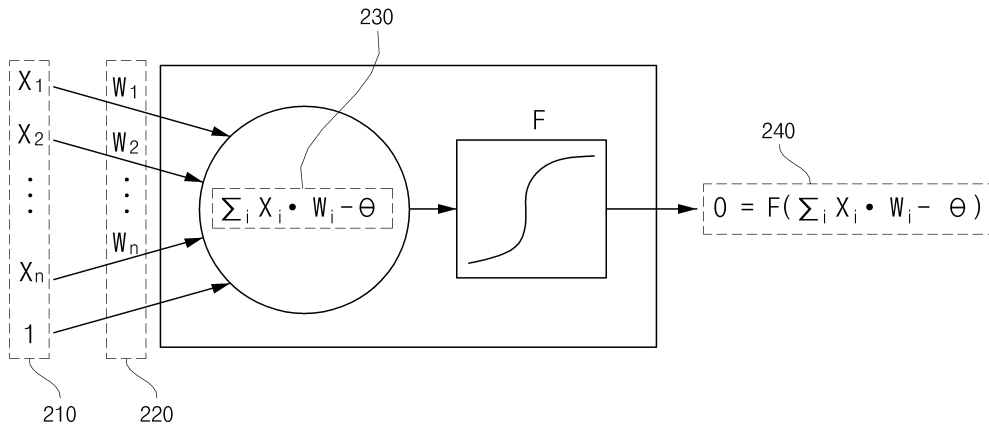
- [0145] 특성 정보 생성부는 음성 정보로부터 잡음 정보를 제거한 뒤 특성 정보를 생성할 수 있다. 이러한 특성 정보 생성부는 도 1의 잡음 제거기(110)와 음성 특징 정보 추출기(120)의 결합 구성에 대응하는 개념이다.
- [0146] 특성 정보 생성부는 음성 정보에 상응하는 입력 정보들과 각 입력 정보의 목표 정보를 학습(training)시켜 얻은 가중치 정보를 음성 정보에 적용하여 특성 정보를 생성할 수 있다.
- [0147] 특성 정보 생성부는 ANN(Artificial Neural Network) 알고리즘, EBP(Error Back Propagation) 알고리즘 및 경사하강법(Gradient Descent Method)을 이용하여 가중치 정보를 획득할 수 있다.
- [0148] 대상 정보 생성부는 특성 정보를 기초로 텍스트 형태의 제1 대상 정보로부터 음성 형태의 제2 대상 정보를 생성하는 기능을 수행한다.
- [0149] 대상 정보 생성부는 데이터베이스에서 특성 정보에 대응하는 기준 정보를 추출하며, 이 기준 정보를 기초로 제1 대상 정보를 음성으로 변환하여 얻은 정보를 튜닝하여 제2 대상 정보를 생성할 수 있다. 이러한 대상 정보 생성부는 도 1의 TTS DB(150), TTS DB 추출기(140) 및 화자 음성 튜닝기(160)의 결합 구성에 대응하는 개념이다.
- [0150] 대상 정보 생성부는 기준 정보로부터 얻은 발화 속도(Pitch Period) 정보 또는 주파수(Log f0) 정보를 기초로 제1 대상 정보를 음성으로 변환하여 얻은 정보를 튜닝하여 제2 대상 정보를 생성할 수 있다.
- [0151] 대상 정보 생성부는 기준 정보와 더불어 특성 정보로부터 얻은 화자 식별 정보를 기초로 제2 대상 정보를 생성할 수 있다. 이러한 대상 정보 생성부는 TTS DB(150), TTS DB 추출기(140), 화자 음성 튜닝기(160), GMM 모델 추출기(170) 및 화자 음성 변환기(180)의 결합 구성에 대응하는 개념이다.
- [0152] 대상 정보 생성부는 가우시안 혼합 모델(GMM)을 기초로 화자 식별 정보를 획득할 수 있다.
- [0153] 다음으로 대상 정보 음성 출력 제어 장치의 작동 방법에 대하여 설명한다.
- [0154] 먼저 특성 정보 생성부가 사용자의 음성 정보를 기초로 사용자의 특성 정보를 생성한다.
- [0155] 이후 대상 정보 생성부가 특성 정보를 기초로 텍스트 형태의 제1 대상 정보로부터 음성 형태의 제2 대상 정보를 생성한다.
- [0156] 이후 대상 정보 출력부가 제2 대상 정보를 출력한다.
- [0157] 이상에서 설명한 본 발명의 실시예를 구성하는 모든 구성요소들이 하나로 결합하거나 결합하여 동작하는 것으로 기재되어 있다고 해서, 본 발명이 반드시 이러한 실시예에 한정되는 것은 아니다. 즉, 본 발명의 목적 범위 안에서라면, 그 모든 구성요소들이 하나 이상으로 선택적으로 결합하여 동작할 수도 있다. 또한, 그 모든 구성요소들이 각각 하나의 독립적인 하드웨어로 구현될 수 있지만, 각 구성요소들의 그 일부 또는 전부가 선택적으로 조합되어 하나 또는 복수개의 하드웨어에서 조합된 일부 또는 전부의 기능을 수행하는 프로그램 모듈을 갖는 컴퓨터 프로그램으로서 구현될 수도 있다. 또한, 이와 같은 컴퓨터 프로그램은 USB 메모리, CD 디스크, 플래쉬 메모리 등과 같은 컴퓨터가 읽을 수 있는 기록매체(Computer Readable Media)에 저장되어 컴퓨터에 의하여 읽혀지고 실행됨으로써, 본 발명의 실시예를 구현할 수 있다. 컴퓨터 프로그램의 기록매체로서는 자기 기록매체, 광 기록매체, 캐리어 웨이브 매체 등이 포함될 수 있다.
- [0158] 또한, 기술적이거나 과학적인 용어를 포함한 모든 용어들은, 상세한 설명에서 다르게 정의되지 않는 한, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 갖는다. 사전에 정의된 용어와 같이 일반적으로 사용되는 용어들은 관련 기술의 문맥상의 의미와 일치하는 것으로 해석되어야 하며, 본 발명에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0159] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위 내에서 다양한 수정, 변경 및 치환이 가능할 것이다. 따라서, 본 발명에 개시된 실시예 및 첨부된 도면들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예 및 첨부된 도면에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구 범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리 범위에 포함되는 것으로 해석되어야 할 것이다.

도면

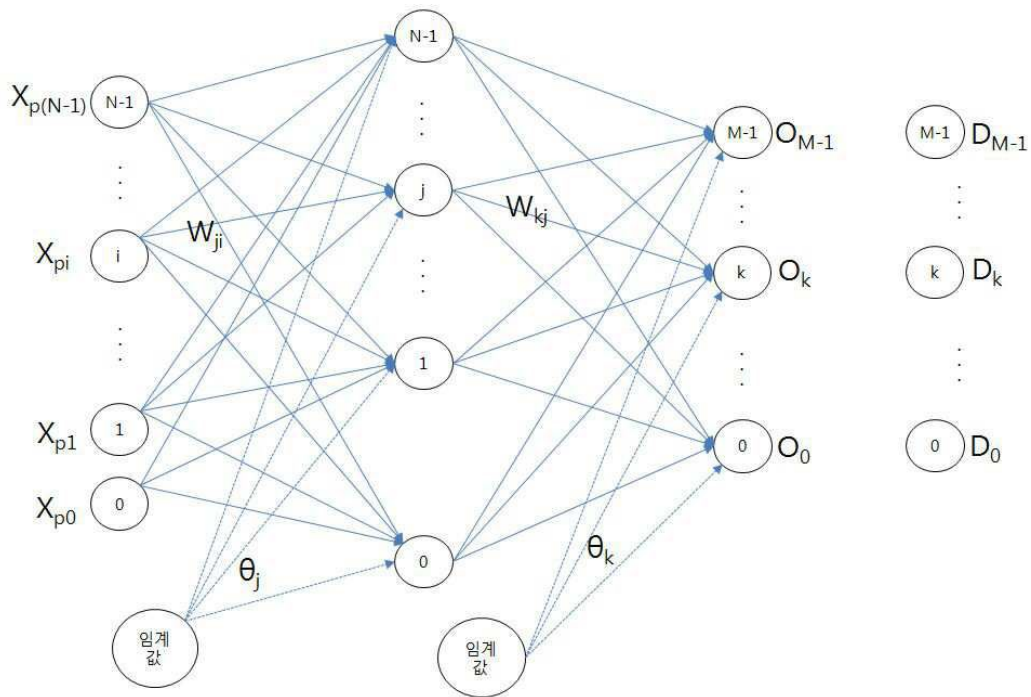
도면1



도면2



도면3



도면4

