



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0102667
(43) 공개일자 2019년09월04일

(51) 국제특허분류(Int. Cl.)
A61B 5/16 (2006.01) A61B 5/00 (2006.01)
(52) CPC특허분류
A61B 5/165 (2013.01)
A61B 5/7264 (2013.01)
(21) 출원번호 10-2018-0023429
(22) 출원일자 2018년02월27일
심사청구일자 2018년02월27일

(71) 출원인
광주과학기술원
광주광역시 북구 첨단과기로 123 (오룡동)
(72) 발명자
신종원
광주광역시 북구 첨단과기로 123(오룡동) 광주과
과학기술원 전기전자컴퓨터공학부
김의성
광주광역시 북구 첨단과기로 123(오룡동) 광주과
과학기술원 전기전자컴퓨터공학부
(74) 대리인
김기문

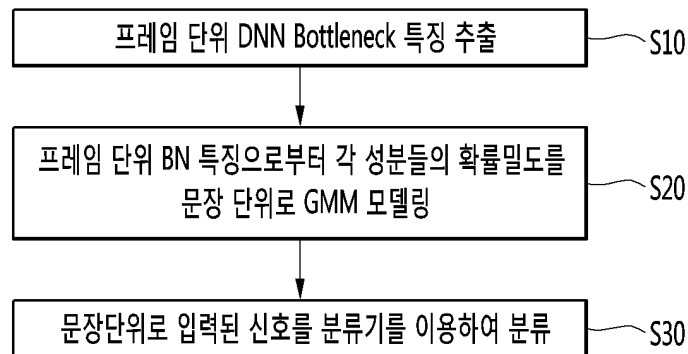
전체 청구항 수 : 총 4 항

(54) 발명의 명칭 **감정 인식 시스템 및 그 방법**

(57) 요약

감정 인식 방법이 개시된다. 본 발명의 일 실시 예에 따른 감정 인식 방법은, 프레임 단위 심층신경망 bottleneck 특징을 추출하는 단계, 상기 추출한 프레임 단위 bottleneck 특징으로부터 각 성분들의 확률밀도를 문장 단위로 가우시안 혼합 모델로 모델링하는 단계 및 문장 단위로 입력된 신호를 분류하는 단계를 포함한다.

대표도 - 도3



이 발명을 지원한 국가연구개발사업

과제고유번호 1375026594(GM11020)

부처명 문화체육관광부

연구관리전문기관 한국콘텐츠진흥원

연구사업명 문화기술연구 주관기관 지원

연구과제명 문화공연 관객들의 실시간 반응 및 공감도 모니터링과 지표화 기술 개발

기 여 율 1/1

주관기관 광주과학기술원

연구기간 2017.04.01 ~ 2018.03.31

명세서

청구범위

청구항 1

프레임 단위 심층신경망 bottleneck 특징을 추출하는 단계;

상기 추출한 프레임 단위 bottleneck 특징으로부터 각 성분들의 확률밀도를 문장 단위로 가우시안 혼합 모델로 모델링하는 단계; 및

문장 단위로 입력된 신호를 분류하는 단계를 포함하는

감정 인식 방법.

청구항 2

제 1 항에 있어서,

상기 문장 단위로 입력된 신호를 분류하는 단계는

문장 단위로 입력된 신호를 최근접 이웃 알고리즘(K-Nearest Neighbor, KNN), 서포트 벡터 머신(Support Vector Machine, SVM), 의사결정나무(Decision Tree) 또는 인공신경망(Artificial Neural Network, ANN) 중 어느 하나를 이용하여 분류하는 단계를 포함하는

감정 인식 방법.

청구항 3

제 1 항에 있어서,

상기 프레임 단위 심층신경망 bottleneck 특징을 추출하는 단계는

심층신경망의 입력 값으로 스펙트럼 기반의 Mel 주파수 캡스트럼을 넣는 단계, 출력 층에서 감정 상태 값을 넣어 사후확률을 추정하는 단계, 및 심층신경망을 역전파 방법으로 훈련하여 bottleneck 층으로부터 bottleneck 특징을 추출하는 단계를 포함하는

감정 인식 방법.

청구항 4

제 1 항에 있어서,

상기 각 성분들의 확률밀도를 문장 단위로 모델링하는 단계는,

가우시안 혼합 모델의 우도함수를 최대화하는 초기 매개변수를 추정하는 단계, 기대값 최대화 알고리즘을 이용하여 최적의 매개변수를 추정하는 단계를 포함하는

감정 인식 방법.

발명의 설명

기술 분야

[0001] 본 발명은 감정 인식 방법에 관한 것이다. 구체적으로 bottleneck 특징을 이용한 감정 인식 방법에 관한 것이다.

배경 기술

[0002] 최근 휴먼 컴퓨터 인터페이스(human-computer interface) 기술의 발전에 따라 음성인식 관련한 연구가 활발하게 진행되고 있다. 하지만, 음성을 이용해서 사람과 기계가 자연스럽게 의사소통하는 것은 여전히 한계가 있다. 이

러한 문제점을 해결하기 위해서 인간과의 상호작용 중 필요 요소인 감정을 인식하는 연구가 광범위하게 진행되고 있다.

[0003] 음성 감정인식 시스템은 크게 특징 추출 부분과 추출된 특징을 이용하여 감정을 분류하는 부분으로 나눌 수 있다. 특징을 추출하는 방법은 보통 프레임 단위와 문장 단위의 접근 방법으로 나눌 수 있다. 프레임 단위 접근 방법으로는 생성모델인 가우시안 혼합 모델(Gaussian Mixture Model, GMM) 혹은 은닉 마르코프 모델(Hidden Markov Model, HMM)로 특징을 모델링하여 각 감정상태의 분포를 생성하고 분류하는 연구가 있었다.

[0004] 반면에, 문장 단위의 접근 방식으로는 각 문장의 전체적인 특성을 반영하는 특징으로 얻기 위해서 low-level feature의 통계치를 사용하는 방법이 있었다.

선행기술문헌

비특허문헌

[0005] (비특허문헌 0001) B. Schuller, A. Batliner, S. Steidl, D. Seppi, "Recognizing Realistic Emotions and Affect in Speech: State of the Art and Lessons Leant from the First Challenge", Speech Communication, 53(10), pp. 1062-1087, 2011.

(비특허문헌 0002) F. Grezl, M. Karafiat, S. Kontar, and J. Cernocky, "Probabilistic and bottle-neck features for LVCSR of meetings," in Proceedings of ICASSP 2007, pp. 757-760.

발명의 내용

해결하려는 과제

[0006] 본 발명의 일 실시 예에 따른 감정 인식 방법은, bottleneck 특징을 추출하여 정확도가 높은 감정 인식 알고리즘을 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0007] 본 발명의 일 실시 예에 따른 감정 인식 방법은, 프레임 단위 심층신경망 bottleneck 특징을 추출하는 단계, 상기 추출한 프레임 단위 bottleneck 특징으로부터 각 성분들의 확률밀도를 문장 단위로 가우시안 혼합 모델로 모델링하는 단계 및 문장 단위로 입력된 신호를 분류하는 단계를 포함한다.

발명의 효과

[0008] 본 발명의 일 실시 예에 따른 감정 인식 방법은, bottleneck 특징을 추출하여 정확도가 높은 감정 인식 알고리즘을 제공할 수 있다.

도면의 간단한 설명

[0009] 도 1은 음성 감정인식에 사용되는 저수준(low-level) 음향특징 분류를 나타낸다.

도 2는 bottleneck 특징 추출을 위한 DNN 모델을 나타낸다.

도 3은 본 발명의 일 실시 예에 따른 감정인식 알고리즘을 나타내는 흐름도이다.

도 4는 본 발명의 일 실시 예에 따른 감정인식 시스템의 성능을 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0010] 이하에서는 도면을 참조하여 본 발명의 구체적인 실시예를 상세하게 설명한다. 그러나 본 발명의 사상은 이하의 실시예에 제한되지 아니하며, 본 발명의 사상을 이해하는 당업자는 동일한 사상의 범위 내에 포함되는 다른 실시예를 구성요소의 부가, 변경, 삭제, 및 추가 등에 의해서 용이하게 제안할 수 있을 것이나, 이 또한 본 발명 사상의 범위 내에 포함된다고 할 것이다.

[0011] 첨부 도면은 발명의 사상을 이해하기 쉽게 표현하기 위하여 전체적인 구조를 설명함에 있어서는 미소한 부분은 구체적으로 표현하지 않을 수도 있고, 미소한 부분을 설명함에 있어서는 전체적인 구조는 구체적으로 반영되지

않을 수도 있다. 또한, 설치 위치 등 구체적인 부분이 다르더라도 그 작용이 동일한 경우에는 동일한 명칭을 부여함으로써, 이해의 편의를 높일 수 있도록 한다. 또한, 동일한 구성이 복수 개가 있을 때에는 어느 하나의 구성에 대해서만 설명하고 다른 구성에 대해서는 동일한 설명이 적용되는 것으로 하고 그 설명을 생략한다.

- [0012] 최근 다양한 분야에서 좋은 결과들을 보여주고 있는 딥러닝 접근은 음성 감정인식 분야에도 활발하게 적용되고 있다. 특히 고유한 방법으로 감정을 표현할 수 있는 특징을 딥러닝을 통해 생성하는 방법이 연구되고 있는데, 구체적으로 심층신경망(Deep Neural Network, DNN)을 사용하여 양방향 회귀신경망(Bi-directional LSTM, BLSTM)을 사용하여 각 감정 분류의 확률을 모델링하고 그것의 통계를 utterance 단위의 특징으로 사용한다.
- [0013] 음성 감정인식 시스템에서 중요한 다른 부분은 분류기이다. 종래에 사용된 음성 감정인식을 위한 분류기로는 K-최근접 이웃 알고리즘(K-Nearest Neighbor, KNN), 서포트 벡터 머신(Support Vector Machine, SVM), 의사결정 나무(Decision Tree), 인공신경망(Artificial Neural Network, ANN) 등이 있다.
- [0014] 이하, 본 발명에서는 음성신호에서 감정 정보의 압축된 high-level 특징을 고려하기 위해 개발된 DNN의 Bottleneck 특징을 GMM으로 모델링하고 SVM으로 해당 감정을 분류하는 알고리즘을 설명한다.
- [0015] 도 1은 음성 감정인식에 사용되는 저수준(low-level) 음향특징 분류를 나타낸다.
- [0016] 음성을 이용한 감정분류를 위해 일반적으로 사용되는 저수준 음향 특징들은 도 1에 도시된 바와 같이, 세가지로 분류될 수 있다. 음성 신호의 기본적인 특징들로는 에너지, 피치, 포먼트를 사용하며 음성 품질 특징으로는 음성 신호에서 숨소리 같은 특이한 특성을 반영하여 jitter와 shimmer 특징을 사용한다.
- [0017] 또한, 음성인식을 비롯한 음성과 관련된 다양한 분야에서 광범위하게 쓰이는 Mel 주파수 캡스트럼(Mel-frequency Cepstral Coefficients, MFCC) 등도 사용된다. 이렇게 프레임 단위로 추출된 저수준 특징들에 21가지 통계치를 적용하여 문장 단위로 정적인 특징을 생성한다. 일반적으로 사용되는 통계함수들은 maximum, minimum, means, regressions, moments, percentiles 등으로 이루어져 있다(선행기술 1 참조).
- [0018] 도 2는 bottleneck 특징 추출을 위한 DNN 모델을 나타낸다.
- [0019] Bottleneck (BN) 특징은 음성인식 분야에서 처음으로 제안되었다(선행기술 2 참조). 도 2는 Bottleneck 특징을 추출하는 DNN의 구조도를 나타낸 것이다. BN 특징은 심층신경망에서 다른 hidden 층의 hidden units의 수보다 상대적으로 적은 수의 hidden units로 구성되어 있는 BN 층을 포함하는 심층신경망으로부터 생성된다.
- [0020] BN 특징을 추출하기 위해서 심층신경망의 입력 값으로 스펙트럼 기반으로 인간의 귀가 가지는 비선형적인 주파수 특성을 반영한 MFCC를 넣고 출력 층에서는 감정상태 값을 넣어서 사후확률을 추정하는 신경망을 구축한다.
- [0021] 그리고 DNN을 역전파(back-propagation) 방법으로 훈련한 후, BN 층으로부터 BN 특징을 추출하게 된다. 이렇게 추출된 BN 특징은 심층신경망에 주어지는 입력 데이터로부터 감정 분류에 효과적인 적절한 정보를 낮은 차원의 압축된 특징으로 표현 할 수 있다는 장점을 가지고 있다.
- [0022] 프레임 단위의 BN 특징으로부터 확률밀도를 문장 단위로 GMM 모델링한다. GMM은 주어진 데이터의 확률 밀도(probability density)를 M개의 가우시안 확률밀도함수의 가중치 합(weighted sum)으로 모델링하는 방법으로 수학적 식 1과 같이 표현된다.

수학적 식 1

[0023]
$$p(x|\lambda) = \sum_{i=0}^M w_i p_i(x)$$

[0024] 여기에서, x는 D차 특징 벡터, w_i 는 $\sum_{i=1}^M w_i = 1$ 과 $w_i \geq 0$ 을 만족하는 i번째 혼합 가중치이고 $p_i(x)$ 은 하나의 가우시안 확률밀도를 표현하며 그 정의는 수학적 식 2와 같이 표현된다.

수학식 2

[0025]

$$p_i(x) = \frac{1}{(2\pi)^{D/2} |C_i|^{1/2}} \exp - \frac{1}{2} (x - \mu_i)' C_i (x - \mu_i)$$

[0026]

여기에서, μ_i 는 평균벡터, C_i 는 공분산 행렬을 갖는 D차원 가우시안 분포를 갖는다고 가정한다. λ 는 $\{w_i, \mu_i, C_i\}$, $i=1, \dots, M$ 로 GMM 파라미터들을 의미한다.

[0027]

BN 특징들의 분포를 가우시안 혼합 밀도에 맞춰주기 위해 GMM의 파라미터를 추정한다. GMM의 우도함수(likelihood function)을 최대화하는 초기 매개변수 λ 를 추정하고 기댓값 최대화(Expectation Maximization, EM) 알고리즘을 통해서 최적의 매개변수들을 추정한다. 최종적으로 EM 알고리즘을 통해서 추정된 문장 단위의 최적의 매개변수들과 저수준(low-level) 음향 특징들의 통계치를 결합하여 문장의 감정을 분류한다.

[0028]

DNN Bottleneck 특징을 입력 값으로 훈련된 GMM을 통하여 문장단위로 입력된 신호가 어떤 감정인지 SVM을 사용하여 분류한다. SVM은 서로 다른 부류(class)에 속하는 특징을 사이에서 공간(margin)을 최대화하는 결정정계(decision boundary)를 찾는 방법으로 일반화 성능이 뛰어나다. 본 발명에서는 원형기준함수(radial basis function)를 이용하여 원래의 특징을 고차원 공간의 특징으로 변환시키는 비선형 SVM을 사용하여 감정을 분류한다.

[0029]

도 3은 본 발명의 일 실시 예에 따른 감정인식 알고리즘을 나타내는 흐름도이다.

[0030]

본 발명의 일 실시 예에 따른 감정인식 시스템은 프레임 단위 DNN Bottleneck 특징을 추출한다(S10). 구체적으로 감정인식 시스템은 심층신경망의 입력 값으로 스펙트럼 기반의 MFCC를 넣고, 출력 층에서 감정 상태 값을 넣어 사후확률을 추정한다. 그리고 DNN을 역전파 방법으로 훈련하여 BN 층으로부터 BN 특징을 추출한다.

[0031]

감정인식 시스템은 추출한 프레임 단위 BN 특징으로부터 각 성분들의 확률밀도를 문장 단위로 GMM 모델링한다(S20). 구체적으로, 감정인식 시스템은 GMM 파라미터를 추정하는데, 이때 감정인식 시스템은 GMM의 우도함수를 최대화하는 초기 매개변수를 추정하고, 기대값 최대화 알고리즘을 이용하여 최적의 매개변수들을 추정한다.

[0032]

감정인식 시스템은 문장단위로 입력된 신호를 분류기를 사용하여 분류한다(S30). 여기에서 사용되는 분류기는 최근접 이웃 알고리즘(K-Nearest Neighbor, KNN), 서포트 벡터 머신(Support Vector Machine, SVM), 의사결정나무(Decision Tree) 또는 인공신경망(Artificial Neural Network, ANN) 중 어느 하나일 수 있다.

[0033]

도 4는 본 발명의 일 실시 예에 따른 감정 인식 시스템의 전체적인 구조를 나타낸다.

[0034]

도 4에 도시된 바와 같이, 본 발명의 일 실시 예에 따른 감정 인식 시스템은 저수준 음향 특징 추출부, 심층 신경망 학습부, bottleneck 추정부, GMM 추정부, statistical function 적용부 및 SVM 학습부를 포함할 수 있다.

[0035]

도 4에 도시된 바와 같이, 본 발명의 일 실시 예에 따른 감정 인식 시스템은 훈련 단계와 테스트 단계로 나누어 동작할 수 있다. 훈련 단계에서는 음성 훈련 데이터 베이스가 저수준 음향 특징 추출부에 입력된다. 그리고, 심층 신경망 학습부, bottleneck 추정부, GMM 추정부, SVM 학습부를 거쳐 훈련이 완료된다. 각 모듈의 상세 구조는 상술한바 여기에서는 상세한 설명을 생략한다.

[0036]

훈련이 완료되면, 테스트 단계에서 실제 음성 데이터 베이스가 저수준 음향 특징 추출부에 입력되고, 동일한 단계를 거쳐 학습된 SVM을 통해 최종 감정 판단이 수행된다.

[0037]

도 5는 본 발명의 일 실시 예에 따른 감정인식 시스템의 성능을 나타내는 도면이다.

[0038]

도 5에 도시된 성능은 기존의 알고리즘과 본 발명의 알고리즘의 성능을 나타내는 것으로, 성능평가의 기준으로 가중 재현율(WAR)과 비가중 재현율(UAR)을 사용하였다.

[0039]

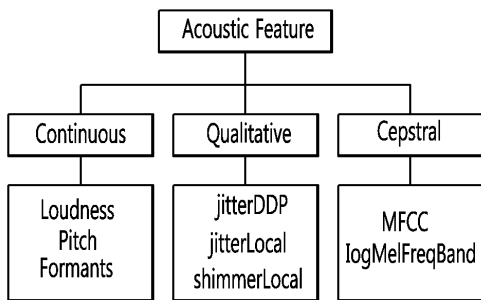
도 5에 도시된 바와 같이, 본 발명에서 제안하는 시스템(proposed system)의 감정인식 알고리즘이 기존의 알고리즘에 비해 우수한 성능을 나타냄을 확인할 수 있다.

[0040] 전술한 본 발명은, 프로그램이 기록된 매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 매체는, 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 매체의 예로는, HDD(Hard Disk Drive), SSD(Solid State Disk), SDD(Silicon Disk Drive), ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광 데이터 저장 장치 등이 있으며, 또한 캐리어 웨이브(예를 들어, 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다.

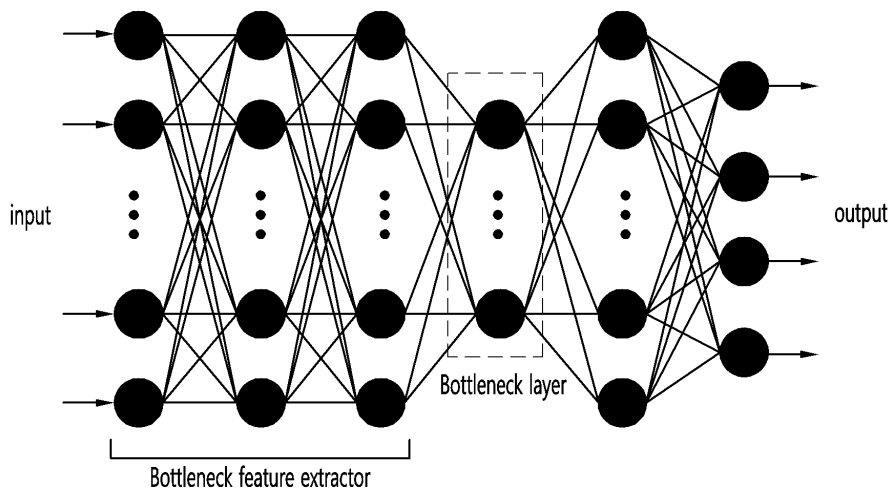
[0041] 따라서, 상기의 상세한 설명은 모든 면에서 제한적으로 해석되어서는 아니되고 예시적인 것으로 고려되어야 한다. 본 발명의 범위는 첨부된 청구항의 합리적 해석에 의해 결정되어야 하고, 본 발명의 등가적 범위 내에서의 모든 변경은 본 발명의 범위에 포함된다.

도면

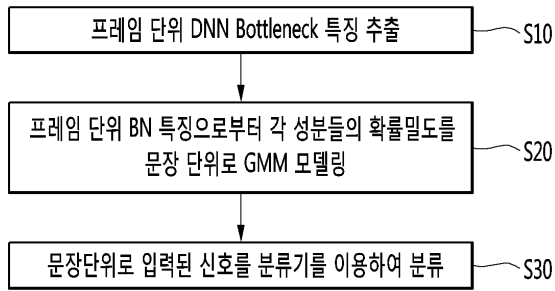
도면1



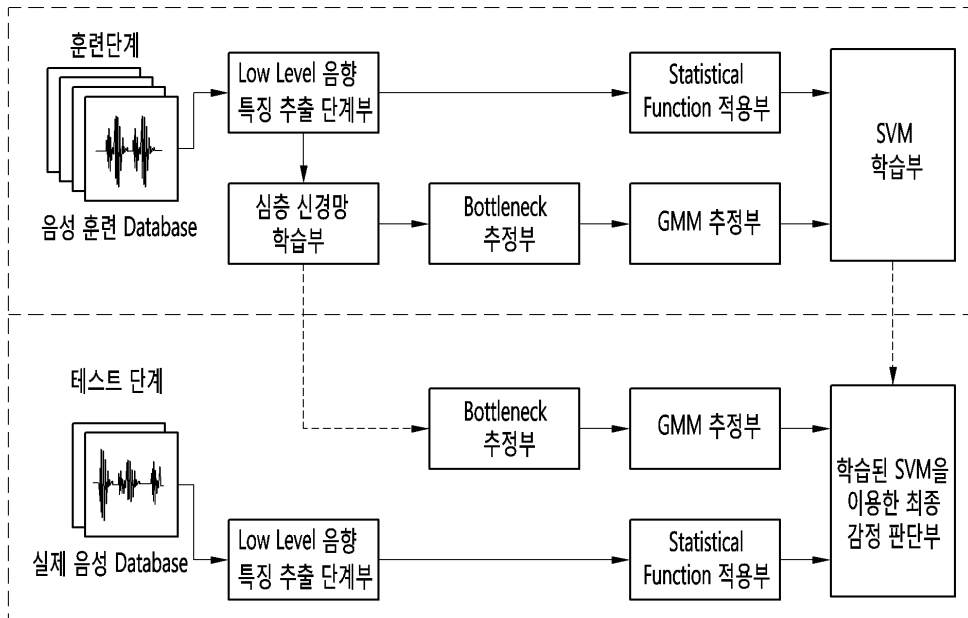
도면2



도면3



도면4



도면5

