



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년04월16일
(11) 등록번호 10-2241364
(24) 등록일자 2021년04월12일

(51) 국제특허분류(Int. Cl.)
A61B 5/00 (2021.01)
(52) CPC특허분류
A61B 5/4884 (2013.01)
A61B 5/4803 (2013.01)
(21) 출원번호 10-2019-0047116
(22) 출원일자 2019년04월23일
심사청구일자 2019년04월23일
(65) 공개번호 10-2019-0135916
(43) 공개일자 2019년12월09일
(30) 우선권주장
1020180061225 2018년05월29일 대한민국(KR)
(56) 선행기술조사문헌
JP2004341493 A*
JP2007000366 A*
KR1020080056069 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
강홍구
서울특별시 서대문구 연세로 50, 연세대학교 제3공학관 505호(신촌동)
변경근
서울특별시 서대문구 연세로 50, 연세대학교 제3공학관 505호(신촌동)
(74) 대리인
민영준
(뒷면에 계속)

전체 청구항 수 : 총 7 항

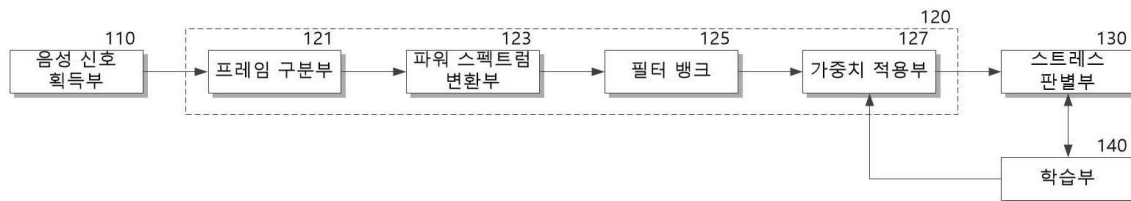
심사관 : 손준영

(54) 발명의 명칭 음성 신호를 이용한 사용자 스트레스 판별 장치 및 방법

(57) 요약

본 발명은 다수의 프레임으로 구분된 음성 신호를 파워 스펙트럼으로 변환하는 파워 스펙트럼 변환부, 각각 지정된 주파수 대역과 패턴을 갖는 다수의 멜-필터 뱅크를 포함하여, 다수 프레임 각각의 파워 스펙트럼을 각각 필터링하여 다수의 멜-필터 뱅크 에너지를 획득하는 필터 뱅크부, 미리 학습된 패턴 추정 방식에 따라 결정된 다수 (뒷면에 계속)

대표도



의 가중치를 다수의 멜-필터 뱅크 에너지 중 대응하는 멜-필터 뱅크 에너지 적용하여 프레임별 특징 벡터를 획득하는 특징 벡터 획득부, 미리 학습된 패턴 추정 방식에 따라 다수의 프레임별 특징 벡터를 순차적으로 인코딩하여 프레임 특징을 추출하되, 이전 획득된 프레임 특징을 함께 인코딩하여 프레임 특징을 획득하는 프레임 특징 추출부 및 프레임 특징 추출부에서 획득되는 프레임 특징 중 최종 프레임 특징을 인가받고, 미리 학습된 패턴 추정 방식에 따라 최종 프레임 특징으로부터 사용자의 스트레스에 대응하는 음성 특징을 추출하는 음성 특징 추출부를 포함하는 음성 신호를 이용한 사용자 스트레스 판별 장치 및 방법을 제공할 수 있다.

(52) CPC특허분류

A61B 5/7264 (2013.01)

(72) 발명자

한혜원

서울특별시 서대문구 연세로 50, 연세대학교 제3공학관 505호(신촌동)

권오성

서울특별시 서대문구 연세로 50, 연세대학교 제3공학관 505호(신촌동)

이 발명을 지원한 국가연구개발사업

과제고유번호 2016-0-00562

부처명 과학기술정보통신부

과제관리(전문)기관명 정보통신기술진흥센터

연구사업명 첨단융복합콘텐츠기술개발사업

연구과제명 상대방의 감성을 추론, 판단하여 그에 맞추어 대화하고 대응할 수 있는 감성 지능 연구개발

기 여 율 1/1

과제수행기관명 연세대학교 산학협력단

연구기간 2017.09.01 ~ 2018.06.30

공지예외적용 : 있음

명세서

청구범위

청구항 1

다수의 프레임으로 구분된 음성 신호를 파워 스펙트럼으로 변환하는 파워 스펙트럼 변환부;

각각 기지정된 주파수 대역과 패턴을 갖는 다수의 멜-필터 뱅크를 포함하여, 다수 프레임 각각의 상기 파워 스펙트럼을 각각 필터링하여 다수의 멜-필터 뱅크 에너지를 획득하는 필터 뱅크부;

미리 학습된 패턴 추정 방식에 따라 결정된 다수의 가중치를 상기 다수의 멜-필터 뱅크 에너지 중 대응하는 멜-필터 뱅크 에너지 적용하여 프레임별 특징 벡터를 획득하는 특징 벡터 획득부;

미리 학습된 패턴 추정 방식에 따라 다수의 프레임별 특징 벡터를 순차적으로 인코딩하여 프레임 특징을 추출하되, 이전 획득된 프레임 특징을 함께 인코딩하여 프레임 특징을 획득하는 프레임 특징 추출부; 및

상기 프레임 특징 추출부에서 획득되는 프레임 특징 중 최종 프레임 특징을 인가받고, 미리 학습된 패턴 추정 방식에 따라 상기 최종 프레임 특징으로부터 사용자의 스트레스에 대응하는 음성 특징을 추출하는 음성 특징 추출부;를 포함하되,

상기 음성 특징 추출부는

상기 다수의 인코더 각각에서 추출된 다수의 프레임 특징을 모두 결합하고, 미리 학습된 패턴 추정 방식에 따라 결합된 다수의 프레임 특징으로부터 상기 음성 특징을 추출하는 사용자 스트레스 판별 장치.

청구항 2

제1 항에 있어서, 상기 프레임 특징 추출부는

다수의 프레임별 특징 벡터 중 대응하는 프레임별 특징 벡터와 이전 획득된 프레임 특징을 인가받아 현재 프레임 특징을 추출하는 다수의 인코더를 포함하는 미리 학습된 인공 신경망으로 구현되는 사용자 스트레스 판별 장치.

청구항 3

삭제

청구항 4

제1 항에 있어서, 상기 사용자 스트레스 판별 장치는

학습 시에 스트레스 레벨이 레이블된 학습용 데이터를 획득하고, 획득된 학습용 데이터에 대한 음성 특징을 상기 학습용 데이터에 레이블된 스트레스 레벨과 비교하여 오차를 추출하고, 획득된 오차를 상기 프레임 특징 추출부 및 상기 음성 특징 추출부로 역전파하여 학습시키고,

학습된 상기 프레임 특징 추출부 및 상기 음성 특징 추출부를 이용하여 획득된 음성 특징과 상기 학습용 데이터에 레이블된 스트레스 레벨 사이의 오차가 기지정된 기준 오차를 초과하면, 상기 특징 벡터 획득부로 오차를 역전파하여 상기 가중치를 업데이트한 후 다시 상기 프레임 특징 추출부 및 상기 음성 특징 추출부를 학습시키는 학습부;를 더 포함하는 사용자 스트레스 판별 장치.

청구항 5

제1 항에 있어서, 상기 다수의 멜-필터 뱅크 각각은

주파수 대역별로 기지정된 멜-주파수 스케일(Mel-frequency scale)과 서로 다른 대역폭을 갖는 삼각 필터로 구현되는 사용자 스트레스 판별 장치.

청구항 6

음성 신호를 다수의 프레임으로 구분하고, 다수 프레임 각각을 파워 스펙트럼으로 변환하는 단계;

각각 기지정된 주파수 대역과 패턴을 갖는 다수의 멜-필터 बैं크를 이용하여, 다수 프레임 각각의 상기 파워 스펙트럼을 각각 필터링하여 다수의 멜-필터 बैं크 에너지를 획득하는 단계;

미리 학습된 패턴 추정 방식에 따라 결정된 다수의 가중치를 상기 다수의 멜-필터 बैं크 에너지 중 대응하는 멜-필터 बैं크 에너지 적용하여 프레임별 특징 벡터를 획득하는 단계;

미리 학습된 패턴 추정 방식에 따라 다수의 프레임별 특징 벡터를 순차적으로 인코딩하여 프레임 특징을 추출하되, 이전 획득된 프레임 특징을 함께 인코딩하여 프레임 특징을 획득하는 단계; 및

획득된 프레임 특징 중 최종 프레임 특징을 인가받고, 미리 학습된 패턴 추정 방식에 따라 상기 최종 프레임 특징으로부터 사용자의 스트레스에 대응하는 음성 특징을 추출하는 단계; 를 포함하되,

상기 음성 특징을 추출하는 단계는

상기 최종 프레임 특징을 포함하는 상기 다수의 프레임 특징을 모두 인가받아 결합하는 단계; 및

미리 학습된 패턴 추정 방식에 따라 결합된 다수의 프레임 특징으로부터 상기 음성 특징을 추출하는 단계; 를 포함하는 사용자 스트레스 판별 방법.

청구항 7

삭제

청구항 8

제6 항에 있어서, 상기 사용자 스트레스 판별 방법은

학습 단계; 를 더 포함하고,

상기 학습 단계는

스트레스 레벨이 레이블된 학습용 데이터를 획득하는 단계;

획득된 학습용 데이터에 대한 음성 특징을 상기 학습용 데이터에 레이블된 스트레스 레벨과 비교하여 오차를 추출하고, 획득된 오차를 역전파하여 상기 프레임 특징을 추출하기 위한 패턴 추정 방식 및 상기 음성 특징을 추출하기 위한 패턴 추정 방식을 업데이트 하는 단계;

업데이트된 패턴 추정 방식에 기반하여 획득된 음성 특징과 상기 학습용 데이터에 레이블된 스트레스 레벨 사이의 오차가 기지정된 기준 오차를 초과하면, 획득된 오차를 역전파하여 상기 가중치를 업데이트한 후 다시 상기 프레임 특징을 추출하기 위한 패턴 추정 방식 및 상기 음성 특징을 추출하기 위한 패턴 추정 방식을 업데이트 하는 단계; 를 더 포함하는 사용자 스트레스 판별 방법.

청구항 9

제6 항에 있어서, 상기 멜-필터 बैं크 에너지를 획득하는 단계는

각각 주파수 대역별로 기지정된 멜-주파수 스케일(Mel-frequency scale)과 서로 다른 대역폭을 갖는 삼각 필터로 구현되는 다수의 멜-필터 बैं크를 이용하여 상기 멜-필터 बैं크 에너지를 획득하는 사용자 스트레스 판별 방법.

발명의 설명

기술 분야

본 발명은 사용자 스트레스 판별 장치 및 방법에 관한 것으로, 딥러닝 기반으로 음성 신호를 이용하여 사용자의 스트레스 유/무를 판별할 수 있는 사용자 스트레스 판별 장치 및 방법에 관한 것이다.

[0001]

배경 기술

- [0002] 음성 신호를 이용하여 스트레스를 판별 하는 기술은 일반적으로 음성 신호에서 감정 혹은 스트레스 인식과 관련된 특징을 추출하고, 추출된 특징과 미리 지정된 스트레스 레이블 사이의 관계를 모델링하는 통계적 방법을 통해서 진행된다.
- [0003] 이때 특징 추출은 주로 숙련된 전문가에 의해 제안된 방법에 기반하여 수행되고 있다. 그러나 이 경우, 효율성이 크다는 장점이 있지만 훈련하고자 하는 데이터에 대하여 선택된 특징 추출 방법이 최선의 결과를 보여주는지 확신 할 수 없다는 한계가 있다.
- [0004] 또한 기존의 통계적 방식에서는 GMM(Gaussian Mixture model), SVM(Support Vector Machine) 알고리즘 등이 있다. GMM 알고리즘은 분류하고자 하는 데이터들의 확률 분포를 다수의 정규분포들의 합으로 나타냄으로써, 시험 상황에서 데이터가 주어졌을 때, 주어진 데이터가 어떤 그룹에 속하는지 분류하는 방식으로 작동한다. 그리고 SVM 알고리즘의 경우는 두 개의 그룹의 데이터가 있을 때 둘 사이를 가르는 초평면(Hyperplane)을 최적화하는 방식으로 데이터를 분류하는 방식이다.
- [0005] 그러나 이러한 통계적 방식 들은 시간 축에서의 변화를 모델링하기 어렵다는 한계가 있어, 시간 도메인에서 특징이 매우 빠르게 변화하는 음성 신호로부터 스트레스를 판별하기 어렵다는 문제가 있다.

선행기술문헌

특허문헌

- [0006] (특허문헌 0001) 한국 공개 특허 제10-2017-0117019호 (2017.10.20 공개)

발명의 내용

해결하려는 과제

- [0007] 본 발명의 목적은 음성 신호로부터 최적화된 특징 벡터를 추출하여 사용자의 스트레스를 정확하게 판단할 수 있는 음성 신호를 이용한 사용자 스트레스 판별 장치 및 방법을 제공하는데 있다.

과제의 해결 수단

- [0008] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 음성 신호를 이용한 사용자 스트레스 판별 장치는 다수의 프레임으로 구분된 음성 신호를 파워 스펙트럼으로 변환하는 파워 스펙트럼 변환부; 각각 기지정된 주파수 대역과 패터를 갖는 다수의 멜-필터 뱅크를 포함하여, 다수 프레임 각각의 상기 파워 스펙트럼을 각각 필터링하여 다수의 멜-필터 뱅크 에너지를 획득하는 필터 뱅크부; 미리 학습된 패턴 추정 방식에 따라 결정된 다수의 가중치를 상기 다수의 멜-필터 뱅크 에너지 중 대응하는 멜-필터 뱅크 에너지 적용하여 프레임별 특징 벡터를 획득하는 특징 벡터 획득부; 미리 학습된 패턴 추정 방식에 따라 다수의 프레임별 특징 벡터를 순차적으로 인코딩하여 프레임 특징을 추출하되, 이전 획득된 프레임 특징을 함께 인코딩하여 프레임 특징을 획득하는 프레임 특징 추출부; 및 상기 프레임 특징 추출부에서 획득되는 프레임 특징 중 최종 프레임 특징을 인가받고, 미리 학습된 패턴 추정 방식에 따라 상기 최종 프레임 특징으로부터 사용자의 스트레스에 대응하는 음성 특징을 추출하는 음성 특징 추출부; 를 포함한다.
- [0009] 상기 프레임 특징 추출부는 상기 프레임 특징 추출부는 다수의 프레임별 특징 벡터 중 대응하는 프레임별 특징 벡터와 이전 획득된 프레임 특징을 인가받아 현재 프레임 특징을 추출하는 다수의 인코더를 포함하는 미리 학습된 인공 신경망으로 구현될 수 있다.
- [0010] 상기 음성 특징 추출부는 상기 다수의 인코더 각각에서 추출된 다수의 프레임 특징을 결합하고, 미리 학습된 패턴 추정 방식에 따라 결합된 다수의 프레임 특징으로부터 상기 음성 특징을 추출할 수 있다.
- [0011] 상기 사용자 스트레스 판별 장치는 학습 시에 스트레스 레벨이 레이블된 학습용 데이터를 획득하고, 획득된 학습용 데이터에 대한 음성 특징을 레이블된 스트레스 레벨과 비교하여 오차를 추출하고, 획득된 오차를 상기 프레임 특징 추출부 및 상기 음성 특징 추출부로 역전파하여 학습시키고, 학습된 상기 프레임 특징 추출부 및 상기 음성 특징 추출부를 이용하여 획득된 음성 특징과 레이블된 스트레스 레벨 사이의 오차가 기지정된 기준 오

차를 초과하면, 상기 특징 벡터 획득부로 오차를 역전파하여 상기 가중치를 업데이트한 후 다시 상기 프레임 특징 추출부 및 상기 음성 특징 추출부를 학습시키는 학습부; 를 더 포함할 수 있다.

[0012] 상기 다수의 멜-필터 बैं크 각각은 주파수 대역별로 기지정된 멜-주파수 스케일(Mel-frequency scale)과 서로 다른 대역폭을 갖는 삼각 필터로 구현될 수 있다.

[0013] 상기 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 음성 신호를 이용한 사용자 스트레스 판별 방법은 음성 신호를 다수의 프레임으로 구분하고, 다수 프레임 각각을 파워 스펙트럼으로 변환하는 단계; 각각 기지정된 주파수 대역과 패턴을 갖는 다수의 필터를 갖는 필터 बैं크를 이용하여, 프레임별로 구분된 상기 파워 스펙트럼을 필터링하여 다수 프레임 각각에서의 멜-필터 बैं크 에너지를 획득하는 단계; 다수의 멜-필터 बैं크 에너지에 미리 학습된 패턴 추정 방식에 의해 프레임별로 결정된 가중치를 적용하여 특징 벡터를 획득하는 단계; 미리 학습된 패턴 추정 방식에 따라 상기 특징 벡터에서 가중치가 적용된 다수의 멜-필터 बैं크 에너지 각각을 순차적으로 인코딩하여 프레임 특징을 추출하되, 이전 획득된 프레임 특징을 함께 인코딩하여 프레임 특징을 획득하는 단계; 및 획득된 프레임 특징 중 최종 프레임 특징을 인가받고, 미리 학습된 패턴 추정 방식에 따라 상기 최종 프레임 특징으로부터 사용자의 스트레스에 대응하는 음성 특징을 추출하는 단계; 를 포함한다.

발명의 효과

[0014] 따라서, 본 발명의 실시예에 따른 음성 신호를 이용한 사용자 스트레스 판별 장치 및 방법은 음성 신호로부터 최적화된 특징 벡터를 추출하고, 추출된 특징 벡터에서 이전 정보를 반영할 수 있는 인공신경망을 이용하여 스트레스 판별 정확도를 크게 높일 수 있다. 또한 음성에서 스트레스에 연관이 되는 파라미터를 정확하게 모델링할 수 있다.

도면의 간단한 설명

[0015] 도 1은 본 발명의 일 실시예에 따른 음성 신호를 이용한 사용자 스트레스 판별 장치의 개략적 구조를 나타낸다.

도 2 및 도 3은 도 2의 특징 벡터 추출부의 동작을 설명하기 위한 도면이다.

도 4는 도 1의 특징 추출부의 상세 구성을 나타낸다.

도 5는 본 발명의 일 실시예에 따른 음성 신호를 이용한 사용자 스트레스 판별 방법을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0016] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.

[0017] 이하, 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써, 본 발명을 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 설명하는 실시예에 한정되는 것이 아니다. 그리고, 본 발명을 명확하게 설명하기 위하여 설명과 관계없는 부분은 생략되며, 도면의 동일한 참조부호는 동일한 부재임을 나타낸다.

[0018] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라, 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈", "블록" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0019] 도 1은 본 발명의 일 실시예에 따른 음성 신호를 이용한 사용자 스트레스 판별 장치의 개략적 구조를 나타내고, 도 2 및 도 3은 도 2의 특징 벡터 추출부의 동작을 설명하기 위한 도면이며, 도 4는 도 1의 특징 추출부의 상세 구성을 나타낸다.

[0020] 도 1을 참조하면, 본 실시예에 따른 음성 신호를 이용한 사용자 스트레스 판별 장치는 음성 신호 획득부(110), 특징 벡터 획득부(120) 및 스트레스 판별부(130)를 포함한다.

[0021] 음성 신호 획득부(110)는 스트레스 판별 대상이 되는 사용자의 음성 신호를 획득한다. 음성 신호 획득부(110)는 마이크 등의 장치로 구현되어 음성 신호를 획득할 수 있으나, 유/무선 통신을 통해 음성 신호를 획득하는 통신 장치 또는 음성 신호가 미리 저장된 저장 장치 등과 같이 다양한 장치로 구현될 수 있다.

[0022] 특징 벡터 획득부(120)는 음성 신호 획득부(110)에서 획득된 음성 신호를 인가받아 기지정된 방식으로 특징 벡

터를 획득한다. 특징 벡터 획득부(120)는 시간 도메인의 음성 신호를 기기정된 길이(일예로 5ms)의 프레임 단위로 구분하고, 구분된 각 프레임에 대해 주파수 대역별 에너지를 분석하여 음성 신호에 대한 특징 벡터를 획득한다.

[0023] 특징 벡터 획득부(120)는 일예로 음성 신호로부터 MFCC(Mel-frequency cepstral coefficients) 기법에 기초하여 프레임별 멜-필터 뱅크 에너지(Mel-filter bank energy)를 추출하여 특징 벡터를 획득할 수 있다.

[0024] MFCC(Mel-frequency cepstral coefficients) 기법은 기존에 음성 신호로부터 특징 벡터를 획득하는 대표적인 기법으로 사람의 청각 특성을 고려하여 각각 대응하는 크기의 주파수 대역(Mel-frequency scale 이라함)을 갖도록 미리 설정된 다수의 멜-필터 뱅크(Mel-filter bank)를 이용하여 음성 신호를 여러 주파수 대역으로 나눈 뒤 각 멜-필터 뱅크에서 필터링된 신호의 에너지를 획득하고, 획득된 에너지를 여러 통계적 기법으로 후처리하여 특징 벡터를 획득하는 기법이다.

[0025] MFCC 기법은 다수의 멜-필터 뱅크에서 필터링되는 주파수 범위에 따라 음성 신호가 서로 다른 의미를 가지도록 하는 추출 방법으로 음성 인식, 음성 합성등 다양한 분야에서 사용되고 있다.

[0026] 다만 본 실시예에서 특징 벡터 획득부(120)는 다수의 멜-필터 뱅크에서 필터링된 다수의 주파수 대역별 에너지 자체를 특징 벡터로 활용한다. 즉 통계적 기법에 따른 별도의 후처리를 수행하지 않는다. 이는 후술하는 스트레스 판별부(130)가 가능한 원래의 음성 신호 자체의 특징을 추출하여 스트레스를 판별할 수 있도록 하기 위함이다.

[0027] 다만 본 실시예에서 특징 벡터 획득부(120)는 사용자의 스트레스를 정확하게 판별할 수 있도록 프레임별로 구분된 주파수 대역별 에너지에 대해 대응하는 가중치(w)를 적용하여 특징 벡터를 획득한다. 이때 주파수 대역별 에너지 각각에 대한 다수의 가중치(w)는 스트레스 판별 장치의 학습 과정 동안 업데이트되어 최적화 될 수 있다.

[0028] 특징 벡터 획득부(120)는 프레임 구분부(121), 파워 스펙트럼 변환부(123), 필터 뱅크부(125) 및 가중치 적용부(127)를 포함할 수 있다.

[0029] 프레임 구분부(121)는 음성 신호 획득부(110)로 음성 신호를 인가받아 기기정된 시간 단위(일예로 5ms)의 다수의 프레임으로 구분한다.

[0030] 파워 스펙트럼 변환부(123)는 다수의 프레임으로 구분된 음성 신호를 파워 스펙트럼으로 변환한다. 도 2의 (a)는 프레임별로 변환된 파워 스펙트럼의 일예를 나타낸다. 여기서 파워 스펙트럼 변환부(123)는 일예로 다수 프레임 각각에 대해 고속 푸리에 변환(Fast Fourier Transform: FFT)을 수행하여 파워 스펙트럼을 획득할 수 있다.

[0031] 필터 뱅크부(125)는 각각 지정된 주파수 대역을 필터링하도록 구성된 다수의 멜-필터 뱅크를 포함하고, 다수의 멜-필터 뱅크 각각이 파워 스펙트럼을 필터링하여 각 프레임별, 주파수 대역별 에너지를 나타내는 멜-필터 뱅크 에너지를 획득한다. 여기서 다수의 멜-필터 뱅크는 도 2의 (b)에 도시된 바와 같이, 주파수 대역별로 기지정된 멜-주파수 스케일(Mel-frequency scale)로 서로 다른 대역폭을 갖는 삼각 필터로 구현될 수 있으며, 다수의 프레임 각각에 대해 다수의 멜-필터 뱅크가 필터링을 수행하여 도 2의 (c)에 도시된 바와 같이 프레임별, 주파수 대역별 멜-필터 뱅크 에너지를 추출할 수 있다.

[0032] 다수의 멜-필터 뱅크 각각의 필터링 함수는 $H_m(k)$ 로 표현될 수 있으며, 수학식 1에 따라 프레임별 파워 스펙트럼을 필터링 할 수 있다.

수학식 1

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \leq k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases}$$

[0033]

[0034] 여기서 k 는 파워 스펙트럼의 인덱스를 나타내고, m 은 멜-필터 बैं크 인덱스를 나타내며, $f(\cdot)$ 는 멜-필터 बैं크 인덱스(m)의 주파수 대역을 나타낸다.

[0035] 그리고 멜-필터 बैं크 에너지(e_m)는 프레임별로 다수의 멜-필터 बैं크 각각에서 필터링된 파워 스펙트럼의 합으로 수학식 2와 같이 획득된다.

수학식 2

$$e_m = \sum_{k=1}^K H_m(k) * s(k)$$

[0036]

[0037] 여기서 $s(k)$ 는 임의의 프레임에서 k 번째 인덱스의 파워스펙트럼을 나타낸다.

[0038] 즉 임의의 프레임에서의 m 번째 멜-필터에 의해서 필터링을 통해 얻어진 멜-뱅크 에너지(e_m)는 수학식 2에 따라 계산된다. 이를 통해 하나의 프레임에서 멜-필터 बैं크 에너지(e_m)를 계산할 수 있다.

[0039] 가중치 적용부(127)는 미리 학습된 인공 신경망으로 구현되어 필터 बैं크부(125)에서 획득된 다수의 멜-필터 बैं크 에너지(e_m)에 대응하는 가중치(w_m)를 적용한다. 이때 가중치 적용부(127)는 미리 학습된 패턴에 따라 각 멜-필터 बैं크 별로 서로 다른 가중치(w_m)를 적용할 수 있다. 그리고 가중치 적용부(127)는 프레임별로 가중치(w_m)가 적용된 멜-필터 बैं크 에너지($w_m * e_m$)로 구성되는 특징 벡터를 획득한다.

[0040] 여기서 가중치(w_m)는 멜-필터 बैं크에서 필터링된 다수의 주파수 대역별 에너지에 가중되도록 적용될 수도 있으나, 멜-필터 बैं크의 다수의 멜-필터의 이득(gain)으로서 적용될 수 있다. 즉 다수의 멜-필터 बैं크 각각에 직접 가중치(w_m)를 적용하여, 가중치(w_m)가 적용된 주파수 대역별 에너지($e_m * w_m$)를 획득하고, 획득된 주파수 대역별 멜-필터 बैं크 에너지를 합하도록 구성될 수 있다.

[0041] 도 3은 특징 벡터 획득부(120)가 특징 벡터를 획득하는 개념을 나타낸 도면으로, 도 3에서 (a)는 프레임 구분부(121)에서 구분되고 파워 스펙트럼 변환부(123)에서 변환된 하나의 프레임에 대한 다수의 파워 스펙트럼($s(k)$)을 나타낸다. (b)는 다수의 멜-필터 बैं크 각각의 필터링 합수($H_m(k)$)를 나타내고, (c)는 다수의 파워 스펙트럼($s(k)$)이 다수의 멜-필터 बैं크 각각에 의해 필터링된 멜-필터 बैं크 에너지(e_m)를 나타내며, (d)는 각 멜-필터 बैं크 에너지(e_m)에 가중치(w_m)를 적용하는 과정을 나타낸다. 마지막으로 (e)는 각 프레임에서 다수의 멜-필터 बैं크 각각에 대응하는 가중치(w_m)가 적용된 가중 멜-필터 बैं크 에너지($w_m * e_m$)로 구성되는 특징 벡터(v_f)를 나타낸다. 도 3에 도시된 바와 같이, 멜-필터 बैं크가 M 개일 때, 프레임별 특징 벡터(v_f)는 M 개의 가중 멜-필터 बैं크 에너지($w_{e1} \sim w_{eM}$)를 포함하여 구성된다.

[0042] 한편, 스트레스 판별부(130)는 패턴 추정 방식이 미리 학습된 인공 신경망으로 구현되고, 다수 프레임에서 각각 획득된 다수의 프레임별 특징 벡터($v_1 \sim v_F$)를 인가받고, 인가된 다수의 프레임별 특징 벡터($v_1 \sim v_F$)로부터 스트레스에 관련된 특징($f_1 \sim f_F$)을 추출하여 사용자의 스트레스를 판별한다. 스트레스 판별부(130)는 일례로 인공 신경망 중 이전 추출된 특징을 현재 특징 추출 시에 반영하는 인공 신경망인 RNN(Recurrent neural network) 또는 LSTM(Long Short Term Memory)으로 구현될 수 있다.

[0043] 도 4를 참조하면, 스트레스 판별부(130)는 프레임 특징 추출부(131) 및 음성 특징 추출부(133)를 포함한다.

[0044] 프레임 특징 추출부(131)는 다수의 인코더($EN_1 \sim EN_F$)를 포함하는 인공 신경망으로 구현되어 다수의 프레임별 특징 벡터($v_1 \sim v_F$) 중 대응하는 프레임별 특징 벡터를 각각 인코딩하여 프레임 특징($f_1 \sim f_F$)을 추출한다. 도 4에서는 프레임 특징 추출부(131)가 LSTM으로 구현되는 경우를 도시하였으며, 이 경우 다수의 인코더($EN_1 \sim EN_F$)는 LSTM 셀로 볼 수 있다.

[0045] 다수의 인코더($EN_1 \sim EN_F$) 각각은 추출된 프레임 특징($f_1 \sim f_F$)을 다음 단의 인코더($EN_1 \sim EN_F$)로 전달하고, 최종

단의 인코더(EN_F)는 추출된 최종 프레임 특징(f_F)을 음성 특징 추출부(133)로 전달한다.

- [0046] 여기서 다수의 인코더($EN_1 \sim EN_{F-1}$)가 추출된 프레임 특징($f_1 \sim f_{F-1}$)을 다음단의 인코더로 전달하는 것은 이전 추출된 프레임 특징($f_1 \sim f_{F-1}$)이 다음 프레임 특징($f_2 \sim f_F$) 추출 시에 함께 고려될 수 있도록 하기 위함이다. 이는 기지정된 시간 단위로 구분된 다수의 프레임 사이의 특징이 서로 연관되도록 하여 시간 정보가 반영될 수 있도록 하기 위함이다. 따라서 최종 프레임 특징(f_F)에는 이전 프레임 특징($f_1 \sim f_{F-1}$)의 특성이 누적되어 반영된 것으로 볼 수 있다.
- [0047] 음성 특징 추출부(133) 또한 미리 학습된 인공 신경망으로 구현되어 최종프레임 특징(f_F)으로부터 사용자 스트레스에 대응하는 음성 특징(s)을 추출한다.
- [0048] 경우에 따라서, 음성 특징 추출부(133)는 다수의 인코더($EN_1 \sim EN_F$)에서 추출되는 프레임 특징($f_1 \sim f_F$)을 모두 인가받아 결합(concatenate)하여 음성 특징(s)을 추출하도록 구성될 수도 있다.
- [0049] 또한 프레임 특징 추출부(131)의 다수의 인코더($EN_1 \sim EN_F$)는 프레임 특징($f_1 \sim f_F$)을 추출할 때, 히든 특징($h_1 \sim h_F$)을 추가로 추출하고, 추출된 히든 특징을 다음 단의 인코더로 전달하여 다음 프레임 특징($f_1 \sim f_F$) 추출 시에 이전 추출된 히든 특징($h_1 \sim h_F$)이 함께 고려되도록 할 수도 있다.
- [0050] 음성 특징 추출부(133)는 사용자의 스트레스에 대응하는 음성 특징(s)을 스트레스 유무로 추출할 수도 있으나, 기지정된 단위로 구분된 스트레스 레벨로 추출할 수도 있다.
- [0051] 본 실시예에 따른 스트레스 판별 장치에서 인공 신경망으로 구현되는 스트레스 판별부(130)와 가중치 적용부(127)는 상기한 바와 같이 미리 학습되어 최적화되어야 한다. 이에 스트레스 판별 장치는 스트레스 판별부(130)와 가중치 적용부(127)의 학습을 위하여 학습부(140)를 더 포함할 수 있다.
- [0052] 학습부(140)는 스트레스 판별 장치의 학습 과정에서 스트레스 여부 또는 스트레스 레벨이 레이블된 음성 신호인 다수의 학습용 데이터를 획득하고, 획득된 학습용 데이터를 음성 신호 획득부(110)에 제공할 수 있다. 그리고 스트레스 판별부(130)로부터 스트레스 판별 결과를 인가받아 학습용 데이터에 레이블된 스트레스 여부 또는 스트레스 레벨과 판별 결과를 비교하여 오차를 판별하고, 판별된 오차를 스트레스 판별부(130)로 역전파하여 스트레스 판별부(130)를 학습시킬 수 있다. 이하에서는 설명의 편의를 위하여 학습용 데이터에 스트레스 레벨이 레이블 된 것으로 가정하여 설명하지만, 상기한 바와 같이 스트레스 여부가 레이블된 학습 데이터가 이용되어도 무방하다.
- [0053] 학습부(140)는 다수의 학습용 데이터를 이용하여 스트레스 판별부(130)를 반복학습 시키며, 이때, 특징 벡터 획득부(120)의 다수의 맬-필터 뱅크 각각에 대응하는 다수의 가중치($w_1 \sim w_M$)는 프레임별로 이전 획득된 값으로 고정된다. 그리고 학습부(140)는 현재 가중치들($w_1 \sim w_M$)에 기반하여 다수의 학습용 데이터에 대한 스트레스 판별 결과에 따른 오차를 스트레스 판별부(130)로 역전파하여 스트레스 판별부(130)를 학습시킨다. 그리고 학습된 스트레스 판별부(130)에 기반하여 다시 다수의 학습용 데이터에 대한 스트레스 판별 결과에 따른 오차를 획득하고, 획득된 오차를 가중치 적용부(127)로 전달하여, 다수의 가중치($w_1 \sim w_M$)를 업데이트한다.
- [0054] 가중치 업데이트부(150)에 의해 다수의 맬-필터 뱅크에 대응하는 다수의 가중치($w_1 \sim w_M$)가 업데이트되면, 업데이트된 가중치($w_1 \sim w_M$)에 따라 다시 다수의 학습용 데이터를 이용하여 스트레스 판별부(130)를 학습시키는 과정을 반복한다. 즉 가중치($w_1 \sim w_M$)가 업데이트 될 때마다, 스트레스 판별부(130)를 다시 학습시킨다.
- [0055] 학습부(140)는 학습 횟수가 미리 지정된 기준 횟수에 도달하거나, 오차가 기지정된 기준 오차 이내이면, 스트레스 판별 장치의 학습을 종료한다.
- [0056] 도 5는 본 발명의 일 실시예에 따른 음성 신호를 이용한 사용자 스트레스 판별 방법을 나타낸다.
- [0057] 도 1 내지 도 4를 참조하여, 본 실시예에 따른 음성 신호를 이용한 사용자 스트레스 판별 방법을 설명하면, 우선 스트레스 판별 장치를 학습 시키기 위한 다수의 학습 데이터를 획득한다(S11). 여기서 학습 데이터는 스트레스 여부 또는 스트레스 레벨이 레이블된 음성 신호이다.
- [0058] 그리고 획득된 학습 데이터를 학습되지 않은 스트레스 판별 장치에 입력하여 스트레스 판별 결과를 획득한다

(S12). 학습용 데이터에 대한 스트레스 판별 결과가 획득되면, 획득된 스트레스 판별 결과와 학습용 데이터에 레이블된 스트레스 레벨을 비교하여 오차를 판별하고, 판별된 오차를 스트레스 판별 장치의 스트레스 판별부(130)로 역전파하여 스트레스 판별부를 학습시킨다(S13).

[0059] 그리고 획득된 전체 학습 데이터를 이용하여 스트레스 판별부(130)를 학습시켰는지 여부를 판별한다(S14). 만일 전체 학습 데이터에 대한 학습이 수행되지 않은 것으로 판단되면, 다시 다른 학습 데이터를 스트레스 판별 장치에 입력하여 스트레스 판별 결과를 획득한다(S12). 그러나 획득된 전체 학습 데이터에 대한 학습이 수행된 것으로 판단되면, 판별된 오차가 기지정된 기준 오차 이하인지 판별한다(S15).

[0060] 만일 오차가 기준 오차를 초과하면, 오차를 특징 벡터 획득부(120)의 가중치 적용부(127)로 역전파하여 다수의 가중치($w_1 \sim w_m$)를 업데이트한다(S16). 그리고 다시 전체 학습 데이터를 이용하여 스트레스 판별부(130)를 학습시킨다.

[0061] 반면, 오차가 기준 오차 이내이면, 스트레스 판별 장치의 학습을 종료하고, 스트레스를 판별하고자 하는 사용자의 음성 신호를 획득한다(S17). 그리고 획득된 음성 신호를 기지정된 시간 단위의 다수의 프레임으로 구분한다(S18). 이후 다수의 프레임 각각을 파워 스펙트럼으로 변환한다(S19).

[0062] 한편 특징 벡터 획득부(120)의 다수의 멜-필터 बैं크를 이용하여 프레임별로 변환된 다수의 파워 스펙트럼을 기지정된 멜-주파수 스케일과 주파수 대역폭에 따라 필터링하여 다수의 멜-필터 बैं크 에너지(e_m)를 획득한다(S20). 그리고 획득된 다수의 멜-필터 बैं크 에너지(e_m)에 학습을 통해 획득된 가중치(w_m)를 적용하여 다수의 가중 멜-필터 बैं크 에너지(we_m)를 획득하고, 획득된 다수의 가중 멜-필터 बैं크 에너지(we_m)로 구성되는 프레임별 특징 벡터($v_1 \sim v_F$)를 획득한다(S21).

[0063] 프레임별 특징 벡터($v_1 \sim v_F$)가 획득되면, 미리 학습된 패턴 인식 방식에 따라 다수의 프레임 각각에 대해 획득된 프레임별 특징 벡터($v_1 \sim v_F$)에서 프레임 특징($f_1 \sim f_F$)을 추출한다(S22). 이때, 다수의 프레임 특징($f_1 \sim f_F$)은 이전 추출된 프레임 특징($f_1 \sim f_{F-1}$)을 함께 반영하여 추출할 수 있다. 그리고 다수의 프레임 특징($f_1 \sim f_{F-1}$)이 추출되면, 추출된 다수의 프레임 특징($f_1 \sim f_F$) 중 최종 프레임 특징(f_F)으로부터 미리 학습된 패턴 인식 방식에 따라 음성 특징(s)을 추출하여 사용자의 스트레스를 판별한다(S23).

[0064] 결과적으로 본 실시예에 따른 음성 신호를 이용한 사용자 스트레스 판별 장치 및 방법은 음성 신호로부터 구분되는 다수 프레임 각각에서 가중치가 적용된 다수의 가중 멜-필터 बैं크 에너지(we_m)로 구성된 프레임별 특징 벡터($v_1 \sim v_F$)를 추출하고, 다수의 프레임별 특징 벡터($v_1 \sim v_F$)에 대해 이전 추출된 프레임 특징($f_1 \sim f_{F-1}$)이 고려된 프레임 특징($f_1 \sim f_F$)을 추출하여 사용자의 스트레스에 대응하는 음성 특징(s)을 추출한다. 따라서 사용자의 스트레스를 정확하게 판단할 수 있다.

[0065] 본 발명에 따른 방법은 컴퓨터에서 실행시키기 위한 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 여기서 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스 될 수 있는 임의의 가용 매체일 수 있고, 또한 컴퓨터 저장 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함하며, ROM(판독 전용 메모리), RAM(랜덤 액세스 메모리), CD(컴팩트 디스크)-ROM, DVD(디지털 비디오 디스크)-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등을 포함할 수 있다.

[0066] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다.

[0067] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

[0068]	110: 음성 신호 획득부	120: 특징 벡터 획득부
	130: 스트레스 판별부	140: 학습부
	121: 프레임 구분부	123: 파워 스펙트럼 변환부

125: 필터 뱅크부

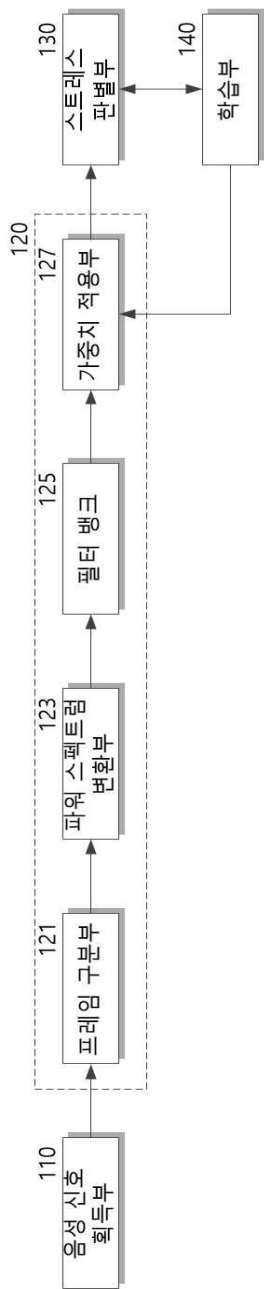
127: 가중치 적용부

131: 프레임 특징 추출부

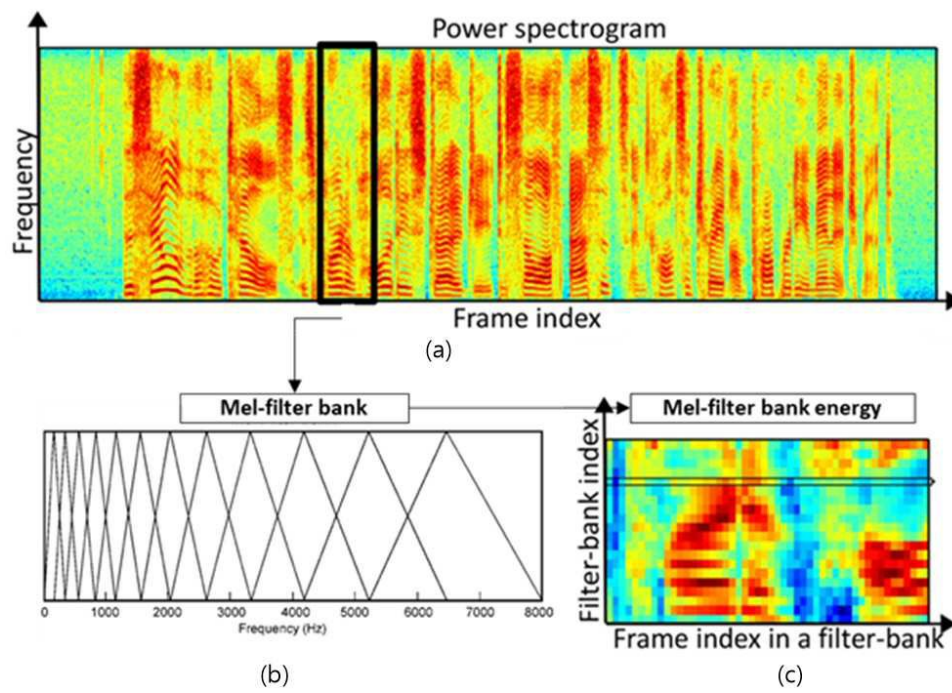
133: 음성 특징 추출부

도면

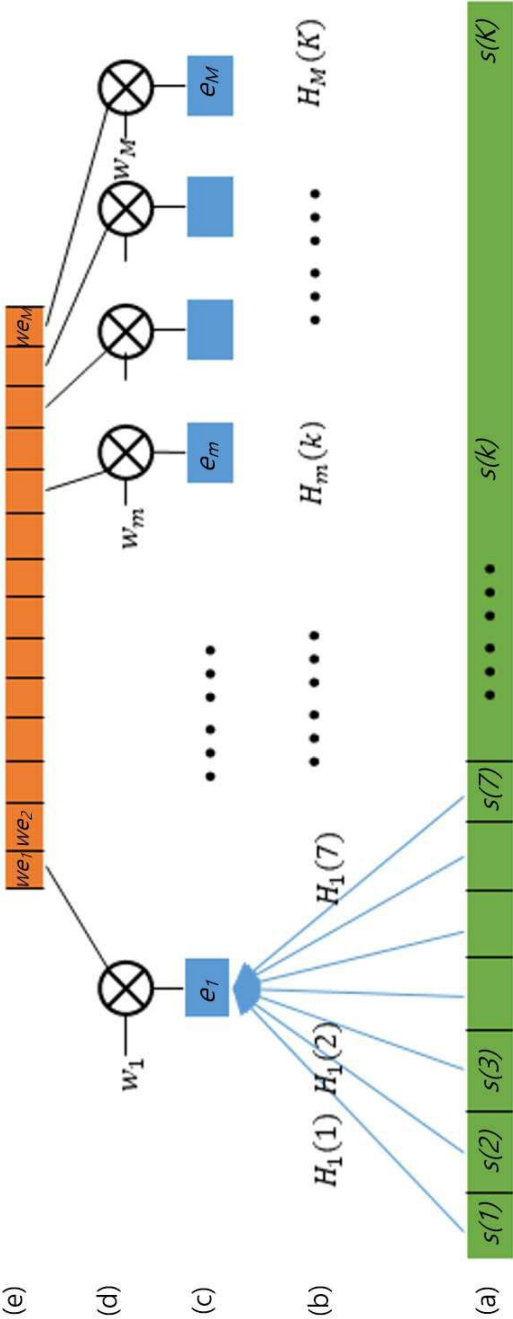
도면1



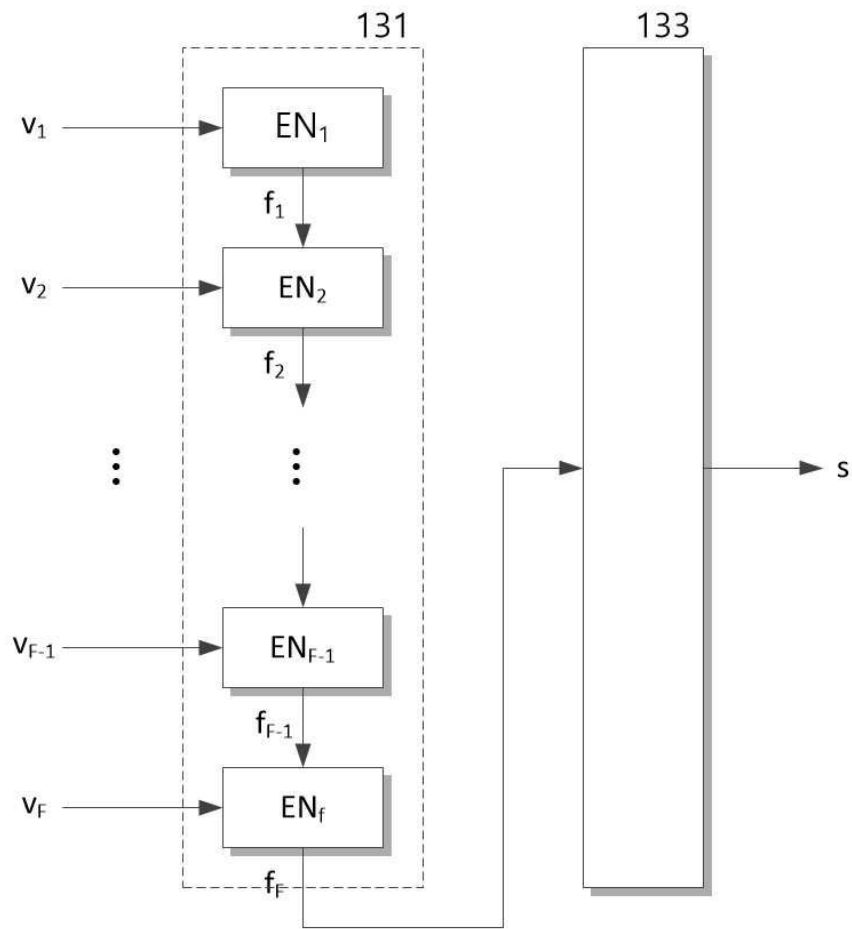
도면2



도면3



도면4



도면5

