



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0076938
(43) 공개일자 2025년05월30일

(51) 국제특허분류(Int. Cl.)

G10L 25/63 (2013.01) G06N 3/02 (2023.01)
G10L 25/18 (2013.01) G10L 25/30 (2013.01)

(52) CPC특허분류

G10L 25/63 (2013.01)
G06N 3/02 (2023.01)

(21) 출원번호 10-2023-0164192

(22) 출원일자 2023년11월23일

심사청구일자 2023년11월23일

(71) 출원인

조선대학교산학협력단

광주광역시 동구 조선대길 126(서석동, 산학협력단)

(72) 발명자

곽근창

광주광역시 서구 풍암2로 66

김태완

광주광역시 서구 천변좌로 174-5 (양동)

(74) 대리인

특허법인지원

전체 청구항 수 : 총 6 항

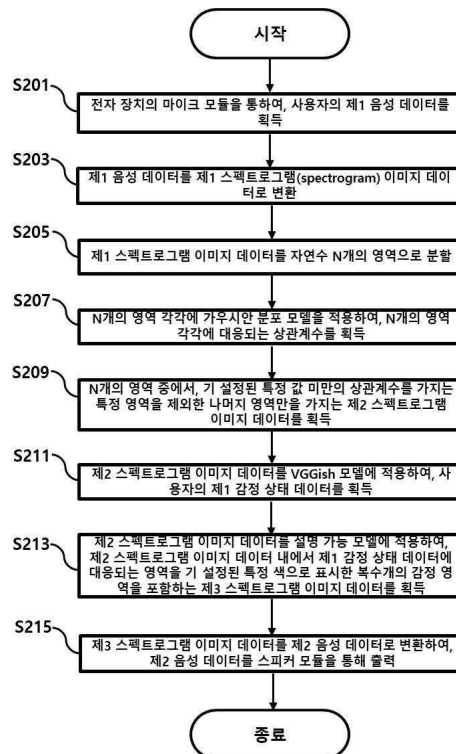
(54) 발명의 명칭 VGGish 모델 및 설명가능 인공지능을 이용한 음성감정분류 모델의 분류근거를 시청각화하는 전자 장치 및 동작 방법

(57) 요약

다양한 실시예들에 따라서, VGGish 모델 및 설명가능 인공지능을 이용한 음성감정분류 모델의 분류근거를 시청각화하는 전자 장치에 있어서, 마이크 모듈, 스피커 모듈, 및 프로세서를 포함하고, 상기 프로세서는, 상기 마이크 모듈을 통하여, 사용자의 제1 음성 데이터를 획득하고, 상기 제1 음성 데이터를 제1 스펙트로그램(spectrogram)으로 변환하고, 상기 제1 스펙트로그램 이미지를 자연수 N개의 영역으로 분할하고, N개의 영역 각각에 가우시안 분포 모델을 적용하여, N개의 영역 각각에 대응되는 상관계수를 획득하고, N개의 영역 중에서, 기 설정된 특정 값 미만의 상관계수를 가지는 특정 영역을 제외한 나머지 영역만을 가지는 제2 스펙트로그램 이미지 데이터를 획득하고, 제2 스펙트로그램 이미지 데이터를 VGGish 모델에 적용하여, 사용자의 제1 감정 상태 데이터를 획득하고, 제2 스펙트로그램 이미지 데이터를 설명 가능 모델에 적용하여, 제2 스펙트로그램 이미지 데이터 내에서 제1 감정 상태 데이터에 대응되는 영역을 기 설정된 특정 색으로 표시한 복수개의 감정 영역을 포함하는 제3 스펙트로그램 이미지 데이터를 획득하고, 제3 스펙트로그램 이미지 데이터를 제2 음성 데이터로 변환하여, 제2 음성 데이터를 스피커 모듈을 통해 출력한다.

(뒷면에 계속)

대표도 - 도2



이미지 데이터로 변환하고, 상기 제1 스펙트로그램 이미지 데이터를 자연수 N개의 영역으로 분할하고, 상기 N개의 영역 각각에 가우시안 분포 모델을 적용하여, 상기 N개의 영역 각각에 대응되는 상관계수를 획득하고, 상기 N개의 영역 중에서, 기 설정된 특정 값 미만의 상관계수를 가지는 특정 영역을 제외한 나머지 영역만을 가지는 제2 스펙트로그램 이미지 데이터를 획득하고, 상기 제2 스펙트로그램 이미지 데이터를 VGGish 모델에 적용하여, 상기 사용자의 제1 감정 상태 데이터를 획득하고, 상기 제2 스펙트로그램 이미지 데이터를 설명 가능 모델에 적용하여, 상기 제2 스펙트로그램 이미지 데이터 내에서 상기 제1 감정 상태 데이터에 대응되는 영역을 기 설정된 특정 색으로 표시한 복수개의 감정 영역을 포함하는 제3 스펙트로그램 이미지 데이터를 획득하고, 상기 제3 스펙트로그램 이미지 데이터를 제2 음성 데이터로 변환하여, 상기 제2 음성 데이터를 상기 스피커 모듈을 통해 출력할 수 있다. 그 밖에 다양한 실시예들도 가능하다.

(52) CPC특허분류

G10L 25/18 (2013.01)

G10L 25/30 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711196700
과제번호	S0102-23-1007
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신산업진흥원
연구사업명	인공지능중심산업융합집적단지조성(R&D)
연구과제명	AI 기반 다중센싱을 이용한 Brain-Body 인터페이스 기술 개발
기 여 율	1/2
과제수행기관명	조선대학교 산학협력단
연구기간	2023.01.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1345366490
과제번호	2017R1A6A1A03015496
부처명	교육부
과제관리(전문)기관명	한국연구재단
연구사업명	이공학확산연구기반구축
연구과제명	웨어러블 디바이스 기반 다중생체 신호를 이용한 사용자 인증 기술
기 여 율	1/2
과제수행기관명	조선대학교
연구기간	2023.03.01 ~ 2024.02.29

명세서

청구범위

청구항 1

VGGish 모델 및 설명가능 인공지능을 이용한 음성감정분류 모델의 분류근거를 시청각화하는 전자 장치에 있어서,

마이크 모듈, 스피커 모듈, 및

프로세서를 포함하고,

상기 프로세서는,

상기 마이크 모듈을 통하여, 사용자의 제1 음성 데이터를 획득하고,

상기 제1 음성 데이터를 제1 스펙트로그램(spectrogram) 이미지 데이터로 변환하고,

상기 제1 스펙트로그램 이미지 데이터를 자연수 N개의 영역으로 분할하고,

상기 N개의 영역 각각에 가우시안 분포 모델을 적용하여, 상기 N개의 영역 각각에 대응되는 상관계수를 획득하고,

상기 N개의 영역 중에서, 기 설정된 특정 값 미만의 상관계수를 가지는 특정 영역을 제외한 나머지 영역만을 가지는 제2 스펙트로그램 이미지 데이터를 획득하고,

상기 제2 스펙트로그램 이미지 데이터를 VGGish 모델에 적용하여, 상기 사용자의 제1 감정 상태 데이터를 획득하고,

상기 제2 스펙트로그램 이미지 데이터를 설명 가능 모델에 적용하여, 상기 제2 스펙트로그램 이미지 데이터 내에서 상기 제1 감정 상태 데이터에 대응되는 영역을 기 설정된 특정 색으로 표시한 복수개의 감정 영역을 포함하는 제3 스펙트로그램 이미지 데이터를 획득하고,

상기 제3 스펙트로그램 이미지 데이터를 제2 음성 데이터로 변환하여, 상기 제2 음성 데이터를 상기 스피커 모듈을 통해 출력하는,

전자 장치.

청구항 2

제 1항에 있어서,

상기 제1 스펙트로그램 이미지 데이터는, 상기 제1 음성 데이터를 시간-주파수 형태의 2차원 형태의 이미지 데이터로 표현한 데이터를 포함하는,

전자 장치.

청구항 3

제 1항에 있어서,

상기 상관계수는, 상기 가우시안 분포 모델을 적용한 상기 N개의 영역 내에 대칭형태의 데이터가 분포하는 정도를 나타내는 값을 포함하는,

전자 장치.

청구항 4

제 1항에 있어서,

상기 VGGish 모델은, Shallow Layers, Middle Layers, Deep Layers 각각의 층에서 추출한 특징에 근거하여, 상기 사용자의 제1 감정 상태 데이터를 분류하는,

전자 장치.

청구항 5

제1 항에 있어서,

상기 설명 가능 모델은, Grad CAM, LIME, Occlusion Sensitivity 를 포함하는,

전자 장치.

청구항 6

VGGish 모델 및 설명가능 인공지능을 이용한 음성감정분류 모델의 분류근거를 시청각화하는 전자 장치의 동작 방법에 있어서,

상기 전자 장치의 마이크 모듈을 통하여, 사용자의 제1 음성 데이터를 획득하는 동작,

상기 제1 음성 데이터를 제1 스펙트로그램(spectrogram) 이미지 데이터로 변환하는 동작,

상기 제1 스펙트로그램 이미지 데이터를 자연수 N개의 영역으로 분할하는 동작,

상기 N개의 영역 각각에 가우시안 분포 모델을 적용하여, 상기 N개의 영역 각각에 대응되는 상관계수를 획득하는 동작,

상기 N개의 영역 중에서, 기 설정된 특정 값 미만의 상관계수를 가지는 특정 영역을 제외한 나머지 영역만을 가지는 제2 스펙트로그램 이미지 데이터를 획득하는 동작,

상기 제2 스펙트로그램 이미지 데이터를 VGGish 모델에 적용하여, 상기 사용자의 제1 감정 상태 데이터를 획득하는 동작,

상기 제2 스펙트로그램 이미지 데이터를 설명 가능 모델에 적용하여, 상기 제2 스펙트로그램 이미지 데이터 내에서 상기 제1 감정 상태 데이터에 대응되는 영역을 기 설정된 특정 색으로 표시한 복수개의 감정 영역을 포함하는 제3 스펙트로그램 이미지 데이터를 획득하는 동작 및,

상기 제3 스펙트로그램 이미지 데이터를 제2 음성 데이터로 변환하여, 상기 제2 음성 데이터를 상기 스피커 모듈을 통해 출력하는 동작을 포함하는,

전자 장치의 동작 방법.

발명의 설명

기술 분야

[0001]

본 발명은 VGGish 모델 및 설명가능 인공지능을 이용한 음성감정분류 모델의 분류근거를 시청각화하는 전자 장치 및 동작 방법에 관한 것이다. 구체적으로, 음성 데이터를 스펙트로그램 이미지로 변환하고, 이를 가우시안 분포에 적용하여 이질적인 데이터를 줄인 후에, VGGish 모델 및 설명 가능 인공지능 모델을 적용하여 사용자의 감정정보를 어떤 영역에 의해 예측 했는지 여부를 판단하여 이를 시청각화하는 전자 장치 및 동작 방법에 관한 것이다.

배경 기술

- [0003] 현재 컴퓨터와 많은 디바이스를 통해 음성과 관련된 많은 서비스를 제공하고 있고 음성은 발화자의 감정과 의사를 가장 많이 표현하고 있는 수단인 만큼 음성에 담긴 감정을 정확하게 분석하고 인식하는 것으로 사용자의 감정에 따른 개인화된 서비스에 대한 수요가 증가하고 있다.
- [0004] 본 발명에서는 이를 해결하기 위해서는 VGGish 모델 및 설명가능 인공지능을 이용한 음성감정분류 모델의 분류근거를 시정각화하는 전자 장치 및 동작 방법에 관한 발명을 제시한다.

선행기술문헌

특허문헌

- [0006] (특허문헌 0001) 국내 등록특허공보 제10-2415101호(2022.06.27)
- (특허문헌 0002) 국내 공개특허공보 제10-2023-0156482호(2023.11.14)

발명의 내용

해결하려는 과제

- [0007] 이에 본 발명이 해결하고자 하는 기술적 과제는, 전술한 문제점을 해결하기 위해 창출된 것으로, VGGish 모델 및 설명가능 인공지능을 이용한 음성감정분류 모델의 분류근거를 시정각화하는 전자 장치 및 동작 방법에 관한 것을 제공하는 것을 목적으로 한다.
- [0008] 본 발명은 효과적인 음성 감정 분류와 분석을 위해 변환된 스펙트로그램 이미지의 시간 주파수 영역 중 분류에 핵심적인 특징이 되는 영역을 시각화하고 오디오에도 적용하여 듣고 분석하는 방법에 대한 것이다 음성데이터에서 감정과 관련된 정보를 얻기 위해 시간 주파수 영역을 확인할 수 있는 스펙트로그램 이미지로 변환하고 모델이 학습하는 데 의미가 없는 데이터 침묵 잡음 버벅거림 등을 선별하고 제외하기 위해 가우시안 분포와 상관 계수를 응용한 데이터 선별 알고리즘을 적용하고 변환된 을 사용하여 감정을 분류한다.
- [0009] 본 발명이 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0011] 다양한 실시예들에 따라서, VGGish 모델 및 설명가능 인공지능을 이용한 음성감정분류 모델의 분류근거를 시정각화하는 전자 장치에 있어서, 마이크 모듈, 스피커 모듈, 및 프로세서를 포함하고, 상기 프로세서는, 상기 마이크 모듈을 통하여, 사용자의 제1 음성 데이터를 획득하고, 상기 제1 음성 데이터를 제1 스펙트로그램(spectrogram) 이미지 데이터로 변환하고, 상기 제1 스펙트로그램 이미지 데이터를 자연수 N개의 영역으로 분할하고, 상기 N개의 영역 각각에 가우시안 분포 모델을 적용하여, 상기 N개의 영역 각각에 대응되는 상관계수를 획득하고, 상기 N개의 영역 중에서, 기 설정된 특정 값 미만의 상관계수를 가지는 특정 영역을 제외한 나머지 영역만을 가지는 제2 스펙트로그램 이미지 데이터를 획득하고, 상기 제2 스펙트로그램 이미지 데이터를 VGGish 모델에 적용하여, 상기 사용자의 제1 감정 상태 데이터를 획득하고, 상기 제2 스펙트로그램 이미지 데이터를 설명가능 모델에 적용하여, 상기 제2 스펙트로그램 이미지 데이터 내에서 상기 제1 감정 상태 데이터에 대응되는 영역을 기 설정된 특정 색으로 표시한 복수개의 감정 영역을 포함하는 제3 스펙트로그램 이미지 데이터를 획득하고, 상기 제3 스펙트로그램 이미지 데이터를 제2 음성 데이터로 변환하여, 상기 제2 음성 데이터를 상기 스피커 모듈을 통해 출력할 수 있다.
- [0012] 다양한 실시예들에 따라서, VGGish 모델 및 설명가능 인공지능을 이용한 음성감정분류 모델의 분류근거를 시정각화하는 전자 장치의 동작 방법에 있어서, 상기 전자 장치의 마이크 모듈을 통하여, 사용자의 제1 음성 데이터

를 획득하는 동작, 상기 제1 음성 데이터를 제1 스펙트로그램(spectrogram) 이미지 데이터로 변환하는 동작, 상기 제1 스펙트로그램 이미지 데이터를 자연수 N개의 영역으로 분할하는 동작, 상기 N개의 영역 각각에 가우시안 분포 모델을 적용하여, 상기 N개의 영역 각각에 대응되는 상관계수를 획득하는 동작, 상기 N개의 영역 중에서, 기 설정된 특정 값 미만의 상관계수를 가지는 특정 영역을 제외한 나머지 영역만을 가지는 제2 스펙트로그램 이미지 데이터를 획득하는 동작, 상기 제2 스펙트로그램 이미지 데이터를 VGGish 모델에 적용하여, 상기 사용자의 제1 감정 상태 데이터를 획득하는 동작, 상기 제2 스펙트로그램 이미지 데이터를 설명 가능 모델에 적용하여, 상기 제2 스펙트로그램 이미지 데이터 내에서 상기 제1 감정 상태 데이터에 대응되는 영역을 기 설정된 특정 색으로 표시한 복수개의 감정 영역을 포함하는 제3 스펙트로그램 이미지 데이터를 획득하는 동작 및, 상기 제3 스펙트로그램 이미지 데이터를 제2 음성 데이터로 변환하여, 상기 제2 음성 데이터를 상기 스피커 모듈을 통해 출력하는 동작을 포함할 수 있다.

발명의 효과

- [0014] 음성 신호를 입력하여 감정 분류를 진행할 때 음성은 주변 잡음과 취득 장비의 성능에 쉽게 영향을 받아 잡음과 감정 분류에 관련이 없는 구간이 생겨나는 것을 효과적으로 선별하여 제거하여 컴퓨팅 자원 소모를 줄이고 기존의 모델을 변형하여 다양한 계층의 특징을 종합하여 더 높은 분류 정확도를 가지며 분류 결과를 위해 집중한 시간 주파수 영역을 시청각화하여 음성데이터를 더 쉽고 접근성 있게 분석할 수 있으므로 추가적인 응용이 가능할 것으로 기대한다.
- [0015] 본 발명의 효과들은 상술된 효과들로 제한되지 않으며, 본 발명의 기술적 특징들에 의하여 기대되는 잠정적인 효과들은 아래의 기재로부터 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [0017] 도 1은 본 발명의 다양한 실시예들에 따른 전자 장치 및 네트워크의 블록도를 도시한다.
- 도 2는 다양한 실시예들에 따른, 전자 장치에 포함된 프로세서(120) 동작의 흐름도를 나타낸다.
- 도 3은 다양한 실시예들에 따른, 스펙트로그램 이미지의 예시를 나타낸 도면이다.
- 도 4는 다양한 실시예들에 따른, 데이터 가우시안 분포에 상관 계수를 적용하는 과정을 설명하기 위한 예시 도면이다.
- 도 5는 다양한 실시예들에 따른, 다양한 계층 특징을 종합하여 분류하기 위한 변형된 VGGish 모델의 예시를 나타낸 도면이다.
- 도 6은 다양한 실시예들에 따른, 설명 가능 모델을 통하여 분류 근거에 대한 집중 영역을 시각화한 예시 도면이다.
- 도 7은 다양한 실시예들에 따른, 모델의 집중 영역을 오디오 데이터로 다시 변환하는 과정을 나타낸 예시 도면이다.
- 도 8은 다양한 실시예들에 따른, 본 발명의 전체 동작 과정을 설명하기 위한 예시 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0018] 이하, 본 문서의 다양한 실시예들이 첨부된 도면을 참조하여 기재된다. 실시예 및 이에 사용된 용어들은 본 문서에 기재된 기술을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 해당 실시예의 다양한 변경, 균등물, 및/또는 대체물을 포함하는 것으로 이해되어야 한다. 도면의 설명과 관련하여, 유사한 구성요소에 대해서는 유사한 참조 부호가 사용될 수 있다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함할 수 있다. 본 문서에서, "A 또는 B" 또는 "A 및/또는 B 중 적어도 하나" 등의 표현은 함께 나열된 항목들의 모든 가능한 조합을 포함할 수 있다. "제 1," "제 2," "첫째," 또는 "둘째," 등의 표현들은 해당 구성요소들을, 순서 또는 중요도에 상관없이 수식할 수 있고, 한 구성요소를 다른 구성요소와 구분하기 위해 사용될 뿐 해당 구성요소들을 한정하지 않는다. 어떤(예: 제 1) 구성요소가 다른(예: 제 2) 구성요소에 "(기능적으로 또는 통신적으로) 연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 상기 어떤 구성요소가 상기 다른 구성요소

에 직접적으로 연결되거나, 다른 구성요소(예: 제 3 구성요소)를 통하여 연결될 수 있다.

[0019] 본 문서에서, "~하도록 구성된(또는 설정된)(configured to)"은 상황에 따라, 예를 들면, 하드웨어적 또는 소프트웨어적으로 "~에 적합한," "~하는 능력을 가지는," "~하도록 변경된," "~하도록 만들어진," "~를 할 수 있는," 또는 "~하도록 설계된"과 상호 호환적으로(interchangeably) 사용될 수 있다. 어떤 상황에서는, "~하도록 구성된 장치"라는 표현은, 그 장치가 다른 장치 또는 부품들과 함께 "~할 수 있는" 것을 의미할 수 있다. 예를 들면, 문구 "A, B, 및 C를 수행하도록 구성된(또는 설정된) 프로세서"는 해당 동작을 수행하기 위한 전용 프로세서(예: 임베디드 프로세서), 또는 메모리 장치에 저장된 하나 이상의 소프트웨어 프로그램들을 실행함으로써, 해당 동작들을 수행할 수 있는 범용 프로세서(예: CPU 또는 application processor)를 의미할 수 있다.

[0020] 본 문서의 다양한 실시예들에 따른 전자 장치는, 예를 들면, 스마트폰, 태블릿 PC, 이동 전화기, 영상 전화기, 전자책 리더기, 데스크탑 PC, 랩탑 PC, 넷북 컴퓨터, 워크스테이션, 서버, PDA, PMP(portable multimedia player), MP3 플레이어, 의료기기, 카메라, 또는 웨어러블 장치 중 적어도 하나를 포함할 수 있다. 웨어러블 장치는 액세서리형(예: 시계, 반지, 팔찌, 발찌, 목걸이, 안경, 콘택트 렌즈, 또는 머리 착용형 장치(head-mounted-device(HMD))), 직물 또는 의류 일체형(예: 전자 의복), 신체 부착형(예: 스킨 패드 또는 문신), 또는 생체 이식형 회로 중 적어도 하나를 포함할 수 있다. 어떤 실시예들에서, 전자 장치는, 예를 들면, 텔레비전, DVD(digital video disk) 플레이어, 오디오, 냉장고, 에어컨, 청소기, 오븐, 전자레인지, 세탁기, 공기 청정기, 셋톱 박스, 홈 오토메이션 컨트롤 패널, 보안 컨트롤 패널, 미디어 박스(예: 삼성 HomeSync™, 애플TV™, 또는 구글 TV™), 게임 콘솔(예: Xbox™, PlayStation™), 전자 사전, 전자 키, 캠코더, 또는 전자 액자 중 적어도 하나를 포함할 수 있다.

[0021] 도 1을 참조하여, 다양한 실시예에서의, 네트워크 환경(100) 내의 전자 장치(101)가 기재된다. 전자 장치(101)는 버스(110), 프로세서(120), 메모리(130), 입출력 인터페이스(150), 디스플레이(160), 및 통신 인터페이스(170)를 포함할 수 있다. 어떤 실시예에서는, 전자 장치(101)는, 구성요소들 중 적어도 하나를 생략하거나 다른 구성요소를 추가적으로 구비할 수 있다. 버스(110)는 구성요소들(110-170)을 서로 연결하고, 구성요소들 간의 통신(예: 제어 메시지 또는 데이터)을 전달하는 회로를 포함할 수 있다. 프로세서(120)는, 중앙처리장치, 어플리케이션 프로세서, 또는 커뮤니케이션 프로세서(communication processor(CP)) 중 하나 또는 그 이상을 포함할 수 있다. 프로세서(120)는, 예를 들면, 전자 장치(101)의 적어도 하나의 다른 구성요소들의 제어 및/또는 통신에 관한 연산이나 데이터 처리를 실행할 수 있다.

[0022] 메모리(130)는, 휘발성 및/또는 비휘발성 메모리를 포함할 수 있다. 메모리(130)는, 예를 들면, 전자 장치(101)의 적어도 하나의 다른 구성요소에 관계된 명령 또는 데이터를 저장할 수 있다. 한 실시예에 따르면, 메모리(130)는 소프트웨어 및/또는 프로그램(140)을 저장할 수 있다. 프로그램(140)은, 예를 들면, 커널(141), 미들웨어(143), 어플리케이션 프로그래밍 인터페이스(API)(145), 및/또는 어플리케이션 프로그램(또는 "어플리케이션")(147) 등을 포함할 수 있다. 커널(141), 미들웨어(143), 또는 API(145)의 적어도 일부는, 운영 시스템으로 지칭될 수 있다. 커널(141)은, 예를 들면, 다른 프로그램들(예: 미들웨어(143), API(145), 또는 어플리케이션 프로그램(147))에 구현된 동작 또는 기능을 실행하는 데 사용되는 시스템 리소스들(예: 버스(110), 프로세서(120), 또는 메모리(130) 등)을 제어 또는 관리할 수 있다. 또한, 커널(141)은 미들웨어(143), API(145), 또는 어플리케이션 프로그램(147)에서 전자 장치(101)의 개별 구성요소에 접근함으로써, 시스템 리소스들을 제어 또는 관리할 수 있는 인터페이스를 제공할 수 있다.

[0023] 미들웨어(143)는, 예를 들면, API(145) 또는 어플리케이션 프로그램(147)이 커널(141)과 통신하여 데이터를 주고받을 수 있도록 중개 역할을 수행할 수 있다. 또한, 미들웨어(143)는 어플리케이션 프로그램(147)으로부터 수신된 하나 이상의 작업 요청들을 우선 순위에 따라 처리할 수 있다. 예를 들면, 미들웨어(143)는 어플리케이션 프로그램(147) 중 적어도 하나에 전자 장치(101)의 시스템 리소스(예: 버스(110), 프로세서(120), 또는 메모리(130) 등)를 사용할 수 있는 우선 순위를 부여하고, 상기 하나 이상의 작업 요청들을 처리할 수 있다. API(145)는 어플리케이션(147)이 커널(141) 또는 미들웨어(143)에서 제공되는 기능을 제어하기 위한 인터페이스로, 예를 들면, 파일 제어, 창 제어, 영상 처리, 또는 문자 제어 등을 위한 적어도 하나의 인터페이스 또는 함수(예: 명령어)를 포함할 수 있다. 입출력 인터페이스(150)는, 예를 들면, 사용자 또는 다른 외부 기기로부터 입력된 명령 또는 데이터를 전자 장치(101)의 다른 구성요소(들)에 전달하거나, 또는 전자 장치(101)의 다른 구성요소(들)로부터 수신된 명령 또는 데이터를 사용자 또는 다른 외부 기기로부터 출력할 수 있다.

[0024] 디스플레이(160)는, 예를 들면, 액정 디스플레이(LCD), 발광 다이오드(LED) 디스플레이, 유기 발광 다이오드(OLED) 디스플레이, 또는 마이크로 전자기계 시스템 (MEMS) 디스플레이, 또는 전자종이(electronic paper) 디스플레이

플레이를 포함할 수 있다. 디스플레이(160)는, 예를 들면, 사용자에게 각종 콘텐츠(예: 텍스트, 이미지, 비디오, 아이콘, 및/또는 심볼 등)을 표시할 수 있다. 디스플레이(160)는, 터치 스크린을 포함할 수 있으며, 예를 들면, 전자 펜 또는 사용자의 신체의 일부를 이용한 터치, 제스처, 근접, 또는 호버링 입력을 수신할 수 있다. 통신 인터페이스(170)는, 예를 들면, 전자 장치(101)와 외부 장치(예: 제 1 외부 전자 장치(102), 제 2 외부 전자 장치(104), 또는 서버(106)) 간의 통신을 설정할 수 있다. 예를 들면, 통신 인터페이스(170)는 무선 통신 또는 유선 통신을 통해서 네트워크(162)에 연결되어 외부 장치(예: 제 2 외부 전자 장치(104) 또는 서버(106))와 통신할 수 있다.

[0025] 무선 통신은, 예를 들면, LTE, LTE-A(LTE Advance), CDMA(code division multiple access), WCDMA(wideband CDMA), UMTS(universal mobile telecommunications system), WiBro(Wireless Broadband), 또는 GSM(Global System for Mobile Communications) 등 중 적어도 하나를 사용하는 셀룰러 통신을 포함할 수 있다. 한 실시예에 따르면, 무선 통신은, 예를 들면, WiFi(wireless fidelity), 블루투스, 블루투스 저전력(BLE), 지그비(Zigbee), NFC(near field communication), 자력 시큐어 트랜스미션(Magnetic Secure Transmission), 라디오 프리퀀시(RF), 또는 보디 에어리어 네트워크(BAN) 중 적어도 하나를 포함할 수 있다. 한 실시예에 따르면, 무선 통신은 GNSS를 포함할 수 있다. GNSS는, 예를 들면, GPS(Global Positioning System), Glonass(Global Navigation Satellite System), Beidou Navigation Satellite System(이하 "Beidou") 또는 Galileo, the European global satellite-based navigation system일 수 있다. 이하, 본 문서에서는, "GPS"는 "GNSS"와 상호 교환적으로 사용될 수 있다. 유선 통신은, 예를 들면, USB(universal serial bus), HDMI(high definition multimedia interface), RS-232(recommended standard 232), 전력선 통신, 또는 POTS(plain old telephone service) 등 중 적어도 하나를 포함할 수 있다. 네트워크(162)는 텔레커뮤니케이션 네트워크, 예를 들면, 컴퓨터 네트워크(예: LAN 또는 WAN), 인터넷, 또는 텔레폰 네트워크 중 적어도 하나를 포함할 수 있다.

[0026] 제 1 및 제 2 외부 전자 장치(102, 104) 각각은 전자 장치(101)와 동일한 또는 다른 종류의 장치일 수 있다. 다양한 실시예에 따르면, 전자 장치(101)에서 실행되는 동작들의 전부 또는 일부는 다른 하나 또는 복수의 전자 장치(예: 전자 장치(102, 104), 또는 서버(106))에서 실행될 수 있다. 한 실시예에 따르면, 전자 장치(101)가 어떤 기능이나 서비스를 자동으로 또는 요청에 의하여 수행해야 할 경우에, 전자 장치(101)는 기능 또는 서비스를 자체적으로 실행시키는 대신에 또는 추가적으로, 그와 연관된 적어도 일부 기능을 다른 장치(예: 전자 장치(102, 104), 또는 서버(106))에게 요청할 수 있다. 다른 전자 장치(예: 전자 장치(102, 104), 또는 서버(106))는 요청된 기능 또는 추가 기능을 실행하고, 그 결과를 전자 장치(101)로 전달할 수 있다. 전자 장치(101)는 수신된 결과를 그대로 또는 추가적으로 처리하여 요청된 기능이나 서비스를 제공할 수 있다. 이를 위하여, 예를 들면, 클라우드 컴퓨팅, 분산 컴퓨팅, 또는 클라이언트-서버 컴퓨팅 기술이 이용될 수 있다.

[0027] 도 2는 다양한 실시예들에 따른, 전자 장치에 포함된 프로세서(120) 동작의 흐름도를 나타낸다.

[0028] 도 3은 다양한 실시예들에 따른, 스펙트로그램 이미지의 예시를 나타낸 도면이다.

[0029] 도 4는 다양한 실시예들에 따른, 데이터 가우시안 분포에 상관 계수를 적용하는 과정을 설명하기 위한 예시 도면이다.

[0030] 도 5는 다양한 실시예들에 따른, 다양한 계측 특징을 종합하여 분류하기 위한 변형된 VGGish 모델의 예시를 나타낸 도면이다.

[0031] 도 6은 다양한 실시예들에 따른, 설명 가능 모델을 통하여 분류 근거에 대한 집중 영역을 시각화한 예시 도면이다.

[0032] 도 7은 다양한 실시예들에 따른, 모델의 집중 영역을 오디오 데이터로 다시 변환하는 과정을 나타낸 예시 도면이다.

[0033] 도 8은 다양한 실시예들에 따른, 본 발명의 전체 동작 과정을 설명하기 위한 예시 도면이다.

[0035] 201 동작에서, 다양한 실시예들에 따르면, 전자 장치(101)의 프로세서(120)는, 마이크 모듈(미도시)을 통하여, 사용자의 제1 음성 데이터를 획득할 수 있다. 여기서, 마이크 모듈은 사용자의 음성을 인식하여 1차원 신호 형태의 데이터로 변환할 수 있는 모든 종류의 장치를 포함할 수 있다.

[0036] 본 발명에서 201 동작은 사용자의 감정 인식을 위해 사용자가 언급한 단어나 문장등을 마이크 모듈로 인식하여,

1차원 신호 형태의 데이터로 변환하는 단계를 의미 할 수 있다.

- [0038] 203 동작에서, 다양한 실시예들에 따르면, 전자 장치(101)의 프로세서(120)는, 제1 음성 데이터를 제1 스펙트로그램(spectrogram) 이미지 데이터로 변환 할 수 있다.
- [0039] 일 실시예에 따라, 제1 스펙트로그램 이미지 데이터는, 제1 음성 데이터로 시간-주파수 형태의 2차원 형태의 이미지 데이터로 표현한 데이터를 포함 할 수 있다. 구체적으로, 스펙트로그램 이미지란, 도 3의 형태와 같은 이미지로, 소리나 파동을 시각화하기 위하여 파형과 스펙트럼의 특징이 조합된 이미지를 의미 할 수 있다. 또한 스펙트로그램 이미지는 시간축과 주파수 축의 변화에 따라 진폭의 차이를 인쇄 농도나 표시 색상의 차이로 나타낼 수 있으며, 주로 Praat, MATLAB, Python 프로그램을 통하여 변환하거나 획득 할 수 있다.
- [0041] 205 동작에서, 다양한 실시예들에 따르면, 전자 장치(101)의 프로세서(120)는, 제1 스펙트로그램 이미지 데이터를 자연수 N개의 영역으로 분할 할 수 있다. 예를 들어, 도3과 같은 스펙트로그램 이미지 데이터를 도 4의 "분할 음성 데이터"부분과 같이 사용자가 지정하거나, 임의로 선택된 영역의 단위로 분할(주로 x축을 분할) 분할하는 것을 의미 할 수 있다. 본 발명에서는 후술할 207 동작에서 나올 가우시안 분포 모델을 적용하기 위해 각 영역별 이미지 데이터의 영역을 분할하는 과정을 의미 할 수 있다.
- [0043] 207 동작에서, 다양한 실시예들에 따르면, 전자 장치(101)의 프로세서(120)는, N개의 영역 각각에 가우시안 분포 모델을 적용하여, N개의 영역 각각에 대응되는 상관계수를 획득 할 수 있다.
- [0044] 일 실시예에 따르면, 상관계수는 가우시안 분포 모델을 적용한 N개의 영역내에 대칭형태의 데이터가 분포하는 정도를 나타내는 값을 포함할 수 있다.
- [0046] 209 동작에서, 다양한 실시예들에 따르면, 전자 장치(101)의 프로세서(120)는, N개의 영역 중에서, 기 설정된 특정 값 미만의 상관계수를 가지는 특정 영역을 제외한 나머지 영역만을 가지는 제2 스펙트로그램 이미지 데이터를 획득 할 수 있다. 예를들어, 도 4를 참고하여 앞선 207 동작에서 분류한 제1 스펙트로그램 이미지 영역 각각에 가우시안 분포 모델을 적용하여 각 영역의 상관 계수를 획득 하고, 이렇게 획득한 상관 계수가 기 설정된 특정 값(예:0.6) 미만 일 영역을 제외한 나머지 영역만으로 구성된 제2 스펙트로그램 이미지 데이터를 획득 할 수 있다.
- [0048] 211 동작에서, 다양한 실시예들에 따르면, 전자 장치(101)의 프로세서(120)는, 제2 스펙트로그램 이미지 데이터를 VGGish 모델에 적용하여, 사용자의 제1 감정 상태 데이터를 획득 할 수 있다. 여기서 VGGish 모델은 스펙트로그램 이미지 데이터를 통해 사용자의 감정 상태를 분류하는 모델로, 본 발명에서의 VGGish 모델은, 도 5를 참고하여 Shallow Layers, Middle Layers, Deep Layers 각각의 층에서 추출한 특징에 근거하여, 상기 사용자의 제1 감정 상태 데이터를 분류하는 변형된 VGGish 모델을 사용 할 수 있다.
- [0050] 213 동작에서, 다양한 실시예들에 따르면, 전자 장치(101)의 프로세서(120)는, 제2 스펙트로그램 이미지 데이터를 설명 가능 모델에 적용하여, 제2 스펙트로그램 이미지 데이터 내에서 제1 감정 상태 데이터에 대응되는 영역을 기 설정된 특정 색으로 표시한 복수개의 감정 영역을 포함하는 제3 스펙트로그램 이미지 데이터를 획득 할 수 있다.
- [0051] 일 실시예에 따라, 설명 가능 모델은, Grad CAM, LIME, Occlusion Sensitivity 를 포함할 수 있다. 여기서 말하는 설명 가능 모델이란, 인공 지능 분류 모델이 어떤 근거 또는 기준에 의해 분류하였는지 시각화하여 표시할 수 있도록 하는 모델로, 도 6과 같이 스펙트로그램 이미지를 설명가능 모델에 적용하여 분류한 근거에 해당 하는 영역을 강조(예:붉은색표시)하여 나타낼 수 있다.
- [0053] 215 동작에서, 다양한 실시예들에 따르면, 전자 장치(101)의 프로세서(120)는, 제3 스펙트로그램 이미지 데이터

를 제2 음성 데이터로 변환하여, 제2 음성 데이터를 스피커 모듈을 통해 출력 할 수 있다. 구체적으로, 도 6과 같이 213 동작에서 설명 가능 모델에 의해 제2 스펙트로그램 이미지 중에서 어떤 영역에 근거하여 제1 감정 상태 데이터를 분류하였는지 표시할 수 있고, 이를 도 7과 같이 해당 표시 영역에 대해서만 전술한 201 동작의 반대 과정을 통해 다시 음성 데이터로 변환하여, 사용자의 음성 입력 데이터 중에서 어떤 부분에 의해서 감정 상태를 분류하였는지를 명확하게 나타낼 수 있다.

[0055] 도 8을 참고하여, 전술한 201 내지 215 동작을 한번에 나타낸 모습을 확인 할 수 있다. 201 동작에서 "와 무조건 콜이지, 나 스키장 너무 좋아" 라는 음성 데이터를 입력 받아 203 동작에서 이를 제1 스펙트로그램 이미지 데이터로 변환하고, 이를 205 내지 209 동작을 통하여 사용자의 말 이외에 노이즈나 빈 곳에 해당하는 음성 데이터를 제거한 제2 스펙트로그램 이미지 데이터를 211 동작에서 변형된 VGGish 모델을 통하여 현재 감정상태 중 "기쁨(Happy)"에 해당하는 감정 상태 데이터를 분류하고, 213 동작을 통하여 제2 스펙트로그램 이미지 데이터에서 어떤 영역에 근거하여 "기쁨(Happy)"라는 감정상태 데이터를 획득하였는지를 표시하고, 이를 215 동작을 통하여 다시 음성 데이터로 변형하여, "무조건 콜"과 "너무 좋아"라는 부분만 출력하여, 해당 영역을 통하여 "기쁨(Happy)"이라는 감정 상태 데이터를 획득하였음을 사용자가 알 수 있게 할 수 있다.

[0057] VGGish 모델 및 설명가능 인공지능을 이용한 음성감정분류 모델의 분류근거를 시청각화하는 전자 장치는 마이크 모듈, 스피커 모듈, 및 프로세서를 포함하고, 상기 프로세서는, 상기 마이크 모듈을 통하여, 사용자의 제1 음성 데이터를 획득하고, 상기 제1 음성 데이터를 제1 스펙트로그램(spectrogram) 이미지 데이터로 변환하고, 상기 제1 스펙트로그램 이미지 데이터를 자연수 N개의 영역으로 분할하고, 상기 N개의 영역 각각에 가우시안 분포 모델을 적용하여, 상기 N개의 영역 각각에 대응되는 상관계수를 획득하고, 상기 N개의 영역 중에서, 기 설정된 특정 값 미만의 상관계수를 가지는 특정 영역을 제외한 나머지 영역만을 가지는 제2 스펙트로그램 이미지 데이터를 획득하고, 상기 제2 스펙트로그램 이미지 데이터를 VGGish 모델에 적용하여, 상기 사용자의 제1 감정 상태 데이터를 획득하고, 상기 제2 스펙트로그램 이미지 데이터를 설명 가능 모델에 적용하여, 상기 제2 스펙트로그램 이미지 데이터 내에서 상기 제1 감정 상태 데이터에 대응되는 영역을 기 설정된 특정 색으로 표시한 복수개의 감정 영역을 포함하는 제3 스펙트로그램 이미지 데이터를 획득하고, 상기 제3 스펙트로그램 이미지 데이터를 제2 음성 데이터로 변환하여, 상기 제2 음성 데이터를 상기 스피커 모듈을 통해 출력할 수 있다.

[0058] 다양한 실시예들에 따르면, 상기 제1 스펙트로그램 이미지 데이터는, 상기 제1 음성 데이터를 시간-주파수 형태의 2차원 형태의 이미지 데이터로 표현한 데이터를 포함할 수 있다.

[0059] 다양한 실시예들에 따르면, 상기 상관계수는, 상기 가우시안 분포 모델을 적용한 상기 N개의 영역 내에 대칭형태의 데이터가 분포하는 정도를 나타내는 값을 포함할 수 있다.

[0060] 다양한 실시예들에 따르면, 상기 VGGish 모델은, Shallow Layers, Middle Layers, Deep Layers 각각의 층에서 추출한 특징에 근거하여, 상기 사용자의 제1 감정 상태 데이터를 분류할 수 있다.

[0061] 다양한 실시예들에 따르면, 상기 설명 가능 모델은, Grad CAM, LIME, Occlusion Sensitivity 를 포함할 수 있다.

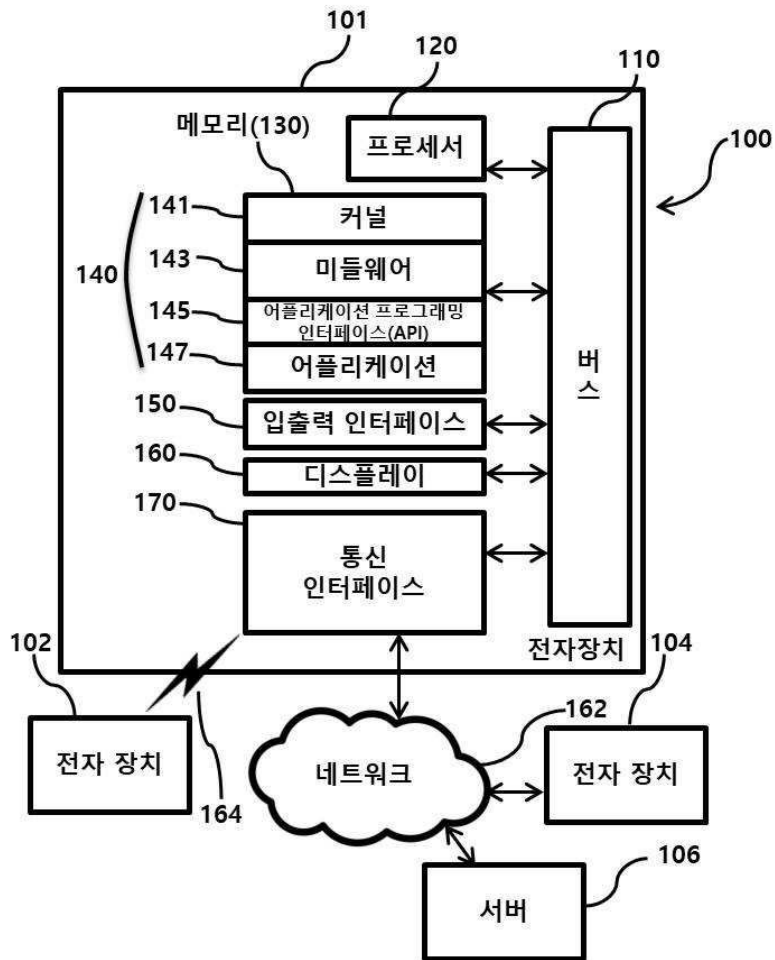
[0062] 다양한 실시예들에 따르면, VGGish 모델 및 설명가능 인공지능을 이용한 음성감정분류 모델의 분류근거를 시청각화하는 전자 장치의 동작 방법은, 상기 전자 장치의 마이크 모듈을 통하여, 사용자의 제1 음성 데이터를 획득하는 동작, 상기 제1 음성 데이터를 제1 스펙트로그램(spectrogram) 이미지 데이터로 변환하는 동작, 상기 제1 스펙트로그램 이미지 데이터를 자연수 N개의 영역으로 분할하는 동작, 상기 N개의 영역 각각에 가우시안 분포 모델을 적용하여, 상기 N개의 영역 각각에 대응되는 상관계수를 획득하는 동작, 상기 N개의 영역 중에서, 기 설정된 특정 값 미만의 상관계수를 가지는 특정 영역을 제외한 나머지 영역만을 가지는 제2 스펙트로그램 이미지 데이터를 획득하는 동작, 상기 제2 스펙트로그램 이미지 데이터를 VGGish 모델에 적용하여, 상기 사용자의 제1 감정 상태 데이터를 획득하는 동작, 상기 제2 스펙트로그램 이미지 데이터를 설명 가능 모델에 적용하여, 상기 제2 스펙트로그램 이미지 데이터 내에서 상기 제1 감정 상태 데이터에 대응되는 영역을 기 설정된 특정 색으로 표시한 복수개의 감정 영역을 포함하는 제3 스펙트로그램 이미지 데이터를 획득하는 동작 및, 상기 제3 스펙트로그램 이미지 데이터를 제2 음성 데이터로 변환하여, 상기 제2 음성 데이터를 상기 스피커 모듈을 통해 출력하는 동작을 포함할 수 있다.

[0064] 본 문서에서 사용된 용어 "모듈" 또는 "~부"은 하드웨어, 소프트웨어 또는 펌웨어로 구성된 유닛을 포함하며, 예를 들면, 로직, 논리 블록, 부품, 또는 회로 등의 용어와 상호 호환적으로 사용될 수 있다. "모듈" 또는 "~부"는, 일체로 구성된 부품 또는 하나 또는 그 이상의 기능을 수행하는 최소 단위 또는 그 일부가 될 수 있다. "모듈" 또는 "~부"는 기계적으로 또는 전자적으로 구현될 수 있으며, 예를 들면, 어떤 동작들을 수행하는, 알려진거나 앞으로 개발될, ASIC(application-specific integrated circuit) 칩, FPGAs(field-programmable gate arrays), 또는 프로그램 가능 논리 장치를 포함할 수 있고, 프로세서(120)에 의하여 실행될 수 있다. 다양한 실시예에 따른 장치(예: 모듈들 또는 그 기능들) 또는 방법(예: 동작들)의 적어도 일부는 프로그램 모듈의 형태로 컴퓨터로 판독 가능한 저장 매체(예: 메모리(130))에 저장된 명령어로 구현될 수 있다. 상기 명령어가 프로세서(예: 프로세서(120))에 의해 실행될 경우, 프로세서가 상기 명령어에 해당하는 기능을 수행할 수 있다. 컴퓨터로 판독 가능한 기록 매체는, 하드디스크, 플로피디스크, 마그네틱 매체(예: 자기테이프), 광기록 매체(예: CD-ROM, DVD, 자기-광 매체 (예: 플롭티컬 디스크)), 내장 메모리 등을 포함할 수 있다. 명령어는 컴파일러에 의해 만들어지는 코드 또는 인터프리터에 의해 실행될 수 있는 코드를 포함할 수 있다. 다양한 실시예에 따른 모듈 또는 프로그램 모듈은 전술한 구성요소들 중 적어도 하나 이상을 포함하거나, 일부가 생략되거나, 또는 다른 구성요소를 더 포함할 수 있다. 다양한 실시예에 따른, 모듈, 프로그램 모듈 또는 다른 구성요소에 의해 수행되는 동작들은 순차적, 병렬적, 반복적 또는 휴리스틱하게 실행되거나, 적어도 일부 동작이 다른 순서로 실행되거나, 생략되거나, 또는 다른 동작이 추가될 수 있다.

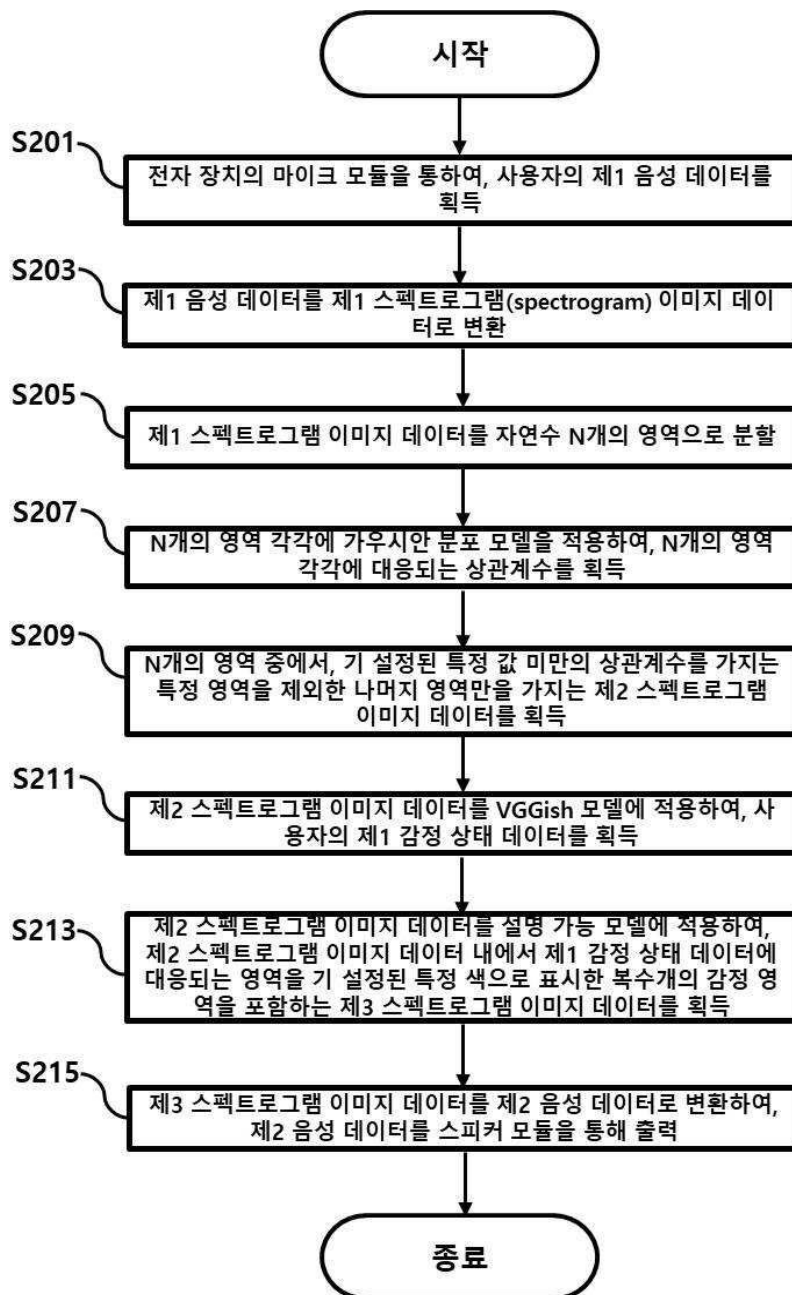
[0065] 그리고 본 문서에 개시된 실시예는 개시된, 기술 내용의 설명 및 이해를 위해 제시된 것이며, 본 개시의 범위를 한정하는 것은 아니다. 따라서, 본 개시의 범위는, 본 개시의 기술적 사상에 근거한 모든 변경 또는 다양한 다른 실시예를 포함하는 것으로 해석되어야 한다.

도면

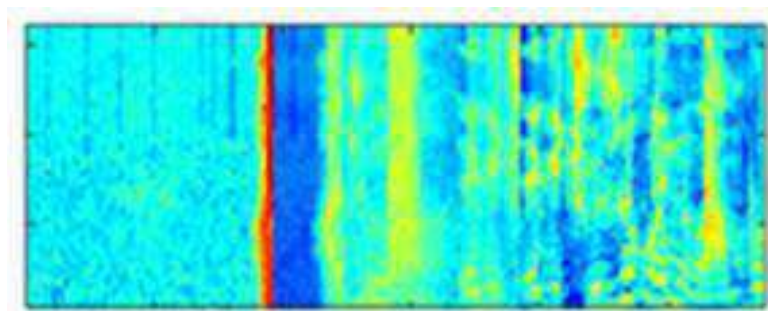
도면1



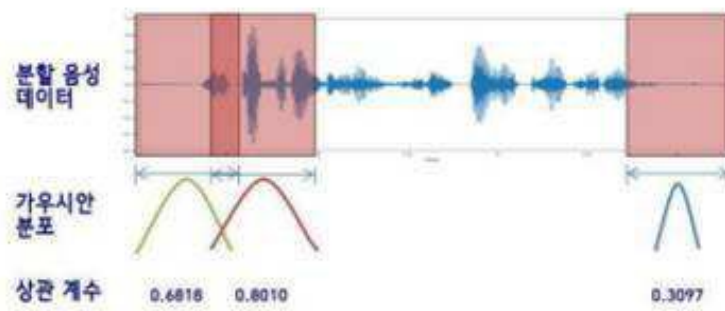
도면2



도면3

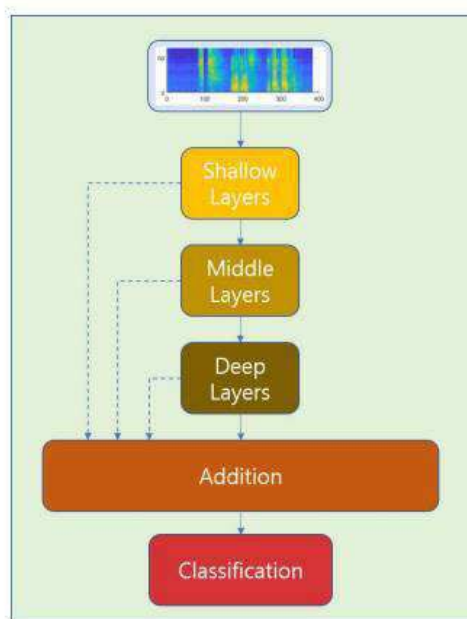


도면4



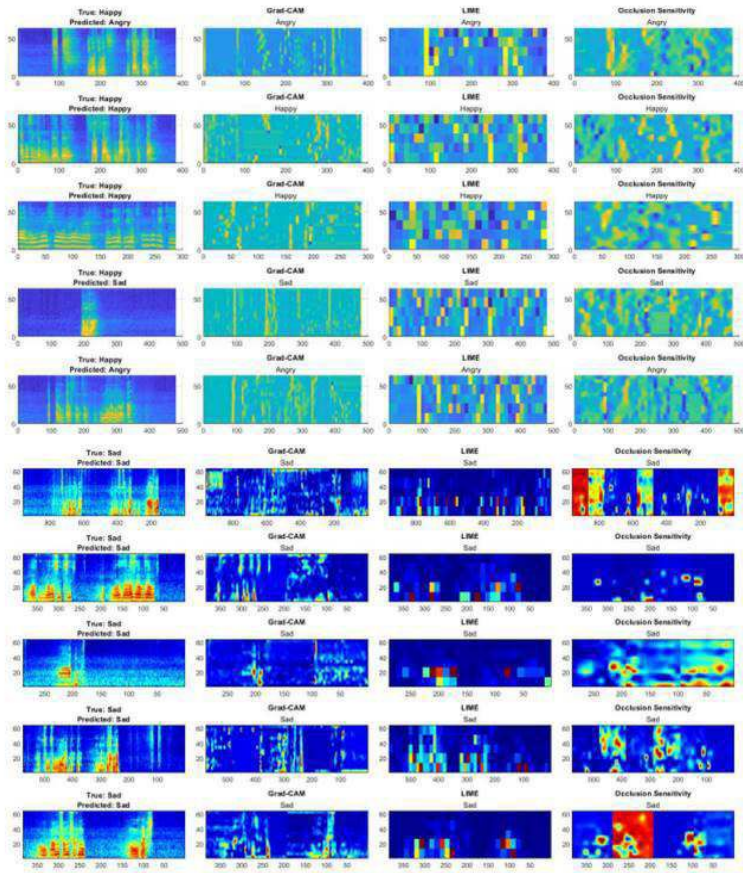
[데이터 가우시안 분포에 상관 계수 적용 과정]

도면5



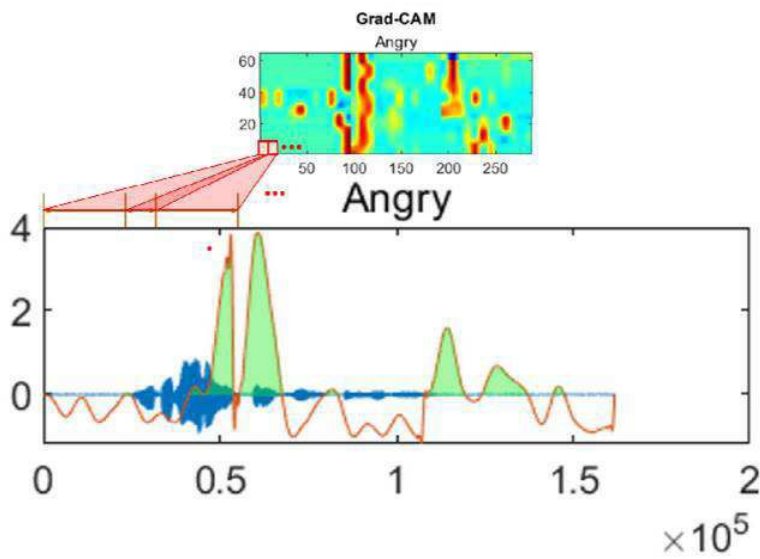
[다양한 계층 특징을 종합하여 분류하기 위한 변형된 VGGish]

도면6



[집중 영역 시각화 (GradCAM, LIME, Occlusion Sensitivity)]

도면7



[모델의 집중 영역을 오디오 데이터에 적용하는 과정]

도면8

