(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2019년10월14일
(11) 등록번호 10-2031954
(24) 등록일자 2019년10월07일

(51) 국제특허분류(Int. Cl.)
G10L 25/63 (2013.01) *G06T 11/00* (2006.01)
G10L 15/02 (2006.01) *G10L 15/16* (2006.01)
G10L 15/26 (2006.01)
(52) CPC특허분류
G10L 25/63 (2013.01)
G06T 11/00 (2013.01)
(21) 출원번호 10-2018-0024714
(22) 출원일자 2018년02월28일
심사청구일자 2018년02월28일
(65) 공개번호 10-2019-0103810
(43) 공개일자 2019년09월05일
(56) 선행기술조사문헌
JP2017156854 A*
(뒷면에 계속)
전체 청구항 수 : 총 13 항

(73) 특허권자
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
백성욱
서울특별시 광진구 아차산로 262, B동 1304호 (자양동, 더샵스타시티)
이미영
서울특별시 강남구 언주로85길 13, 102호 (역삼동, 경남아파트)
(뒷면에 계속)
(74) 대리인
특허법인엠에이피에스

심사관 : 이상래

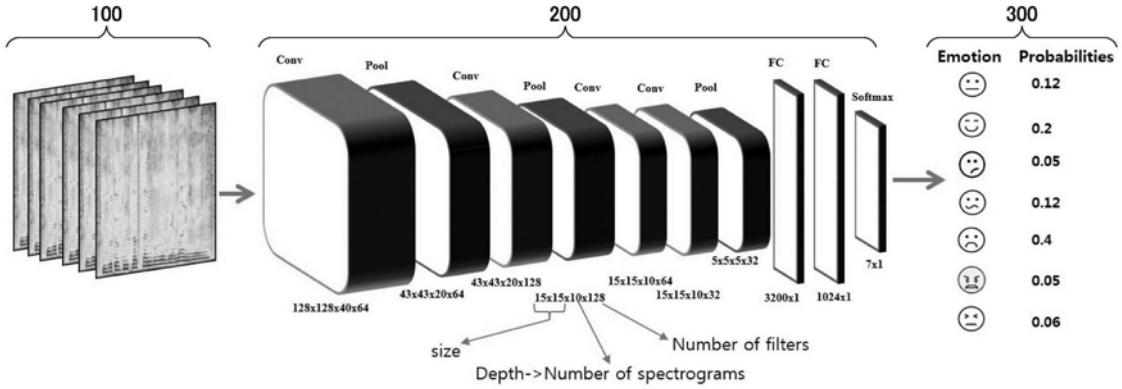
(54) 발명의 명칭 **추론 프로세스를 이용한 음성 감정 인식 장치 및 방법**

(57) 요 약

본 발명의 일 측면에 따른 음성신호에 기반한 사용자의 감정 추정 방법에 있어서, 사용자의 음성 신호를 세그먼트 단위의 음성 데이터로 변환하는 단계; 및 세그먼트 단위의 음성 데이터를 3D 컨볼루션 뉴럴 네트워크(3D Convoultion neural network)에 입력하여, 세그먼트 별로 각 음성 데이터의 특징을 추출하고, 추출된 특징값에

(뒷면에 계속)

대 표 도



기초하여 각 음성 데이터가 분류될 감정 상태 추정값을 산출하는 단계를 포함하되, 감정 상태 추정값을 산출하는 단계는 컨볼류션 뉴럴 네트워크를 통해 추출된 특징값에 기초하여 미리 설정된 복수의 감정 상태로 분류될 확률을 나타내는 제 1 신뢰도, 이전에 감정 상태가 분류된 전체 음성 데이터의 개수 대비 각 음성 데이터별 감정 분류 상태의 값에 기초하여 산출한 과거 분류 이력을 나타내는 제 2 신뢰도, 및 복수의 감정이 공존할 수 있는 확률을 나타내는 지식 베이스에 기반하여 분류 대상 세그먼트의 음성 데이터가 가질 수 있는 감정 상태 확률을 나타내는 제 3 신뢰도를 합산한 값에 기초하여 감정 상태 추정값을 산출할 수 있다.

(52) CPC특허분류

G10L 15/02 (2013.01)

G10L 15/16 (2013.01)

G10L 15/26 (2013.01)

(72) 발명자

권순일

서울특별시 강남구 압구정로 201, 77동 1402호 (압구정동, 현대아파트)

전석봉

서울특별시 마포구 월드컵북로 501, 912동 802호 (상암동, 상암월드컵파크9단지)

아마드 자밀

서울특별시 광진구 능동로 209, 율곡관 601B (군자동, 세종대학교)

무하마드 칸

서울특별시 광진구 능동로 209, 율곡관 601B (군자동, 세종대학교)

이자즈 울 하크

서울특별시 광진구 능동로 209, 율곡관 601B (군자동, 세종대학교)

박준렬

서울특별시 서초구 방배중앙로 112, A동 203호 (방배동, 방배파크빌2)

(56) 선행기술조사문헌

KR101480668 B1*

최하나 외 2명, 음성신호기반의 감정분석을 위한 특징벡터 선택. 전기학회논문지 64권 9호, 2015년 9월, pp.1363-1368

A. M. Badshah, et al. Speech Emotion Recognition from Spectrograms with Deep Convolutional Neural Network. 2017 Int'l Conf. on PlatCon. Feb. 2017, pp.1-5

N. Rahim, et al. Emotional Speech Identification using Convolutional Neural Network. The 3rd Int'l Conf. on Next Generation Computing 2017. Dec. 2017

*는 심사관에 의하여 인용된 문헌

이 발명을 지원한 국가연구개발사업

과제고유번호	1711055159
부처명	과학기술정보통신부
연구관리전문기관	정보통신기술진흥센터
연구사업명	SW컴퓨팅산업원천기술개발산업
연구과제명	음성음향 분석 기반 상황 판단 솔루션 기술 개발
기 여 율	1/1
주관기관	한국과학기술연구원(KIST)
연구기간	2017.03.01 ~ 2018.02.28

명 세 서

청구범위

청구항 1

음성신호에 기반한 사용자의 감정을 추정하는 방법에 있어서,

사용자의 음성 신호를 세그먼트 단위의 음성 데이터로 변환하는 단계; 및

상기 세그먼트 단위의 음성 데이터를 3D 컨볼루션 뉴럴 네트워크(3D Convoultion neural network)에 입력하여, 세그먼트 별로 각 음성 데이터의 특징을 추출하고, 추출된 특징값에 기초하여 각 음성 데이터가 분류될 감정 상태 추정값을 산출하는 단계를 포함하되,

상기 감정 상태 추정값을 산출하는 단계는

상기 3D 컨볼루션 뉴럴 네트워크를 통해 추출된 특징값에 기초하여 미리 설정된 복수의 감정 상태로 분류될 확률을 나타내는 제 1 신뢰도,

이전에 감정 상태가 분류된 전체 음성 데이터의 개수 대비 각 음성 데이터별 감정 분류 상태의 값에 기초하여 산출한 과거 분류 이력을 나타내는 제 2 신뢰도, 및

복수의 감정이 공존할 수 있는 확률을 나타내는 지식 베이스에 기반하여 분류 대상 세그먼트의 음성 데이터가 가질 수 있는 감정 상태 확률을 나타내는 제 3 신뢰도를 합산한 값에 기초하여 감정 상태 추정값을 산출하되,

상기 제 3 신뢰도는 상기 지식 베이스에 미리 설정된 값으로서, 이전에 존재하였던 제 1 감정 상태와 분류 대상 세그먼트의 음성 데이터가 가질 수 있는 제 2 감정 상태가 공존할 수 있는 확률을 나타내는 것인, 음성신호에 기반한 사용자의 감정을 추정하는 방법.

청구항 2

제 1 항에 있어서,

상기 음성 데이터로 변환하는 단계는 상기 음성 신호에 대하여 FFT(Fast fourier transform)를 적용하여 2차원 그래프 형태의 스펙트로그램을 생성하는 단계 및

상기 스펙트로그램을 단위 시간 단위로 분할하여 3차원 구조로 배치하는 단계를 포함하는, 음성신호에 기반한 사용자의 감정을 추정하는 방법.

청구항 3

제 1 항에 있어서,

상기 제 1 신뢰도가 임계값 이하인 경우에는 해당 음성 데이터를 폐기하는 단계를 더 포함하는, 음성신호에 기반한 사용자의 감정을 추정하는 방법.

청구항 4

제 1 항에 있어서,

상기 감정 상태 추정값은 제 1 신뢰도에 제 1 가중치를 곱한값과 제 2 신뢰도에 제 2 가중치를 곱한값과 제 3 신뢰도에 제 3 가중치를 곱한값을 합산한 값으로서, 0 과 1 사이의 크기를 갖는 것인, 음성신호에 기반한 사용자의 감정을 추정하는 방법.

청구항 5

삭제

청구항 6

제 1 항에 있어서,

상기 제 3 신뢰도는 상기 지식 베이스에 미리 설정된 값으로서, 이전에 존재하였던 제 1 감정 상태, 이전에 존재하였던 제 2 감성 상태와 분류 대상 세그먼트의 음성 데이터가 가질 수 있는 제 3 감정 상태가 공존할 수 있는 확률을 나타내는 것인, 음성신호에 기반한 사용자의 감정을 추정하는 방법.

청구항 7

제 1 항에 있어서,

상기 각 음성 데이터는 행복(happy), 슬픔(sad), 분노(angry), 공포(fear), 혐오(disgust), 지루함(boredom) 및 중립(neutral)의 7가지 감정 상태로 분류되는 것인, 음성신호에 기반한 사용자의 감정을 추정하는 방법.

청구항 8

음성신호에 기반한 사용자의 감정을 추정하는 장치에 있어서,

사용자의 음성 신호로부터 음성 데이터를 추출하고, 상기 음성 데이터를 기초로 감정 상태 추정값을 산출하는 프로그램이 저장된 메모리 및

상기 메모리에 저장된 프로그램을 실행하여 상기 음성신호에 기반한 사용자의 감정 추정하는 프로세서를 포함하되,

상기 프로세서는 사용자의 음성 신호를 세그먼트 단위의 음성 데이터로 변환하고, 상기 세그먼트 단위의 음성 데이터를 3D 컨볼루션 뉴럴 네트워크(3D Convoultion neural network)에 입력하여, 세그먼트 별로 각 음성 데이터의 특징을 추출하고, 추출된 특징값에 기초하여 각 음성 데이터가 분류될 감정 상태 추정값을 산출하되,

상기 감정 상태 추정값을 산출하는 방법은 상기 3D 컨볼루션 뉴럴 네트워크를 통해 추출된 특징값에 기초하여 미리 설정된 복수의 감정 상태로 분류될 확률을 나타내는 제 1 신뢰도, 이전에 감정 상태가 분류된 전체 음성 데이터의 개수 대비 각 음성 데이터별 감정 분류 상태의 값에 기초하여 산출한 과거 분류 이력을 나타내는 제 2 신뢰도, 및 복수의 감정이 공존할 수 있는 확률을 나타내는 지식 베이스에 기반하여 분류 대상 세그먼트의 음성 데이터가 가질 수 있는 감정 상태 확률을 나타내는 제 3 신뢰도를 합산한 값에 기초하여 감정 상태 추정값을 산출하고,

상기 제 3 신뢰도는 상기 지식 베이스에 미리 설정된 값으로서, 이전에 존재하였던 제 1 감정 상태와 분류 대상 세그먼트의 음성 데이터가 가질 수 있는 제 2 감정 상태가 공존할 수 있는 확률을 나타내는 것인, 음성신호에 기반한 사용자의 감정을 추정하는 장치.

청구항 9

제 8 항에 있어서,

상기 음성 데이터로 변환하는 단계는 상기 음성 신호에 대하여 FFT(Fast fourier transform)를 적용하여 2차원 그래프 형태의 스펙트로그램을 생성하고,

상기 스펙트로그램을 단위 시간 단위로 분할하여 3차원 구조로 배치하는 것인, 음성신호에 기반한 사용자의 감정을 추정하는 장치.

청구항 10

제 8 항에 있어서,

상기 제 1 신뢰도가 임계값 이하인 경우에는 해당 음성 데이터를 폐기하는 것인, 음성신호에 기반한 사용자의 감정을 추정하는 장치.

청구항 11

제 8 항에 있어서,

상기 감정 상태 추정값은 제 1 신뢰도에 제 1 가중치를 곱한값과 제 2 신뢰도에 제 2 가중치를 곱한값과 제 3 신뢰도에 제 3 가중치를 곱한값을 합산한 값으로서, 0 과 1 사이의 크기를 갖는 것인, 음성신호에 기반한 사용자의 감정을 추정하는 장치.

청구항 12

제 8 항에 있어서,

상기 제 3 신뢰도는 상기 지식 베이스에 미리 설정된 값으로서, 이전에 존재하였던 제 1 감정 상태와 분류 대상 세그먼트의 음성 데이터가 가질 수 있는 제 2 감정 상태가 공존할 수 있는 확률을 나타내는 것인, 음성신호에 기반한 사용자의 감정을 추정하는 장치.

청구항 13

삭제

청구항 14

제 8 항에 있어서,

상기 각 음성 데이터는 행복(happy), 슬픔(sad), 분노(angry), 공포(fear), 혐오(disgust), 지루함(boredom) 및 중립(neutral)의 7가지 감정 상태로 분류되는 것인, 음성신호에 기반한 사용자의 감정을 추정하는 장치.

청구항 15

제 1 항 내지 제 4 항, 제 6 항 및 제 7 항 중 어느 한 항에 따른 음성신호에 기반한 사용자의 감정을 추정하는 방법을 수행하는 프로그램이 기록된 컴퓨터 판독가능 기록매체.

발명의 설명

기 술 분 야

[0001] 본 발명은 음성대화를 수행하는 화자의 감정을 인식하는 기술에 관한 것으로, 특히 심층 학습 및 추론 기술을 사용하여 화자의 감정 상태를 조사하는 것에 관한 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 사람의 음성을 인식하는 기술은 더 편리한 서비스를 원하는 사용자들의 요구에 발맞추어 빠르게 발전하고 있다. 최근에는 단순 음성의 인식에 그치지 않고, 딥러닝 기술을 이용하여 사용자의 음성에서 감정을 인식하는 기술이 발전하고 있는 상황이다.

[0003] 하지만, 음성 감정 인식은 음성 신호로부터 인간의 감정을 검출하고 인식하지만, 언어를 기반으로 감정을 인식하는 방식은 감정적 표현의 다양성으로 인해 쉽사리 파악하기 어렵고, 감정의 단서가 언어의 기반이 아닌 강한 표현력이나 음성 신호 그 자체에 있을 수 있기에 기계가 이를 판단하기에 어려운 한계가 존재한다.

발명의 내용

해결하려는 과제

[0004] 본 발명의 일 실시예는 전술한 종래 기술의 문제점을 해결하기 위한 것으로서, 사용자의 음성으로부터 사용자의 감정 상태를 인식하기 위한 기술로, 딥러닝 분석과 시간이 흐름에 따라 변하게 되는 감정 정보를 바탕으로 현재의 감정 상태를 인식하여 종래의 감정 인식 기술보다 정확성을 높이는 것을 목적으로 한다.

[0005] 다만, 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제로 한정되지 않으며, 또 다른 기술적 과제들이 존재할 수 있다.

과제의 해결 수단

[0006] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 일 측면에 따른 음성신호에 기반한 사용자의 감정 추정 방법에 있어서, 사용자의 음성 신호를 세그먼트 단위의 음성 데이터로 변환하는 단계; 및 세그먼트 단위의 음성 데이터를 3D 컨볼루션 뉴럴 네트워크(3D Convoultion neural network)에 입력하여, 세그먼트 별로 각 음성 데이터의 특징을 추출하고, 추출된 특징값에 기초하여 각 음성 데이터가 분류될 감정 상태 추정값을 산출하는 단계를 포함하되, 감정 상태 추정값을 산출하는 단계는 3D 컨볼루션 뉴럴 네트워크를 통해 추출된 특

징값에 기초하여 미리 설정된 복수의 감정 상태로 분류될 확률을 나타내는 제 1 신뢰도, 이전에 감정 상태가 분류된 전체 음성 데이터의 개수 대비 각 음성 데이터별 감정 분류 상태의 값에 기초하여 산출한 과거 분류 이력을 나타내는 제 2 신뢰도, 및 복수의 감정이 공존할 수 있는 확률을 나타내는 지식 베이스에 기반하여 분류 대상 세그먼트의 음성 데이터가 가질 수 있는 감정 상태 확률을 나타내는 제 3 신뢰도를 합산한 값에 기초하여 감정 상태 추정값을 산출할 수 있다.

[0007] 본 발명의 다른 측면에 따른 음성신호에 기반한 사용자의 감정 추정하는 장치에 있어서, 사용자의 음성 신호로부터 음성 데이터를 추출하고, 음성 데이터를 기초로 감정 상태 추정값을 산출하는 프로그램이 저장된 메모리 및 메모리에 저장된 프로그램을 실행하여 음성신호에 기반한 사용자의 감정 추정하는 프로세서를 포함하되, 프로세서는 사용자의 음성 신호를 세그먼트 단위의 음성 데이터로 변환하고, 세그먼트 단위의 음성 데이터를 3D 컨볼루션 뉴럴 네트워크(3D Convoultion neural network)에 입력하여, 세그먼트 별로 각 음성 데이터의 특징을 추출하고, 추출된 특징값에 기초하여 각 음성 데이터가 분류될 감정 상태 추정값을 산출하되, 감정 상태 추정값을 산출하는 방법은 3D 컨볼루션 뉴럴 네트워크를 통해 추출된 특징값에 기초하여 미리 설정된 복수의 감정 상태로 분류될 확률을 나타내는 제 1 신뢰도, 이전에 감정 상태가 분류된 전체 음성 데이터의 개수 대비 각 음성 데이터별 감정 분류 상태의 값에 기초하여 산출한 과거 분류 이력을 나타내는 제 2 신뢰도, 및 복수의 감정이 공존할 수 있는 확률을 나타내는 지식 베이스에 기반하여 분류 대상 세그먼트의 음성 데이터가 가질 수 있는 감정 상태 확률을 나타내는 제 3 신뢰도를 합산한 값에 기초하여 감정 상태 추정값을 산출하는 장치일 수 있다.

발명의 효과

[0008] 전술한 본 발명의 과제 해결 수단 중 어느 하나에 의하면, 사용자의 음성으로부터 사용자의 감정 상태를 인식하기 위한 기술로, 딥러닝 분석과 시간이 흐름에 따라 변하게 되는 감정 정보를 바탕으로 현재의 감정 상태를 인식하여 종래의 감정 인식 기술보다 높은 정확도를 가질 수 있게 된다.

도면의 간단한 설명

[0009] 도 1은 본 발명의 일 실시예에 따른, 추론 프로세스를 이용한 음성 감정 인식의 입출력 구조를 나타낸 도면이다.

도 2는 본 발명의 일 실시예에 따른, 연속적인 음성 세그먼트에서의 복수의 감정이 공존할 확률을 나타낸 도면이다.

도 3은 종래의 충돌하고 공존하는 감정의 상태 사이에 관계를 나타낸 전환 다이어그램이다.

도 4는 본 발명의 일 실시예에 따른, 3가지 신뢰도을 결합하는 방식과 그 예시를 나타낸 도면이다.

도 5는 본 발명의 일 실시예에 따른, 음성 세그먼트의 감정 사이의 충돌을 해결하는 방법과 그 예시 도면이다.

도 6은 본 발명의 일 실시예에 따른, 추론 프로세스를 이용한 음성 감정 인식의 방법을 나타낸 동작 흐름도이다.

도 7은 본 발명의 일 실시예에 따른, 3D 컨볼루션 뉴럴 네트워크를 통해 반환된 신뢰도 확인을 통해 신뢰 레벨의 높낮음을 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0010] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 본 발명을 명확하게 설명하기 위해 도면에서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다. 또한, 도면을 참고하여 설명하면서, 같은 명칭으로 나타낸 구성일지라도 도면에 따라 도면 번호가 달라질 수 있고, 도면 번호는 설명의 편의를 위해 기재된 것에 불과하고 해당 도면 번호에 의해 각 구성의 개념, 특징, 기능 또는 효과가 제한 해석되는 것은 아니다.

[0011] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한, 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 의미하며, 하나 또는 그 이상의 다른 특징이나 숫자, 단계,

동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0012] 본 명세서에 있어서 '부(部)' 또는 '모듈'이란, 하드웨어 또는 소프트웨어에 의해 실현되는 유닛(unit),양방을 이용하여 실현되는 유닛을 포함하며, 하나의 유닛이 둘 이상의 하드웨어를 이용하여 실현되어도 되고, 둘 이상의 유닛이 하나의 하드웨어에 의해 실현되어도 된다.

[0013] 도 1은 본 발명의 일 실시예에 따른, 추론 프로세스를 이용한 음성 감정 인식을 위한 입출력 구조를 나타낸 도면이다.

[0014] 도 1을 참조하면, 추론 프로세스를 이용한 음성 감정 인식을 위한 입출력 구조는 사용자의 음성 신호(100)와 딥러닝인 3D 컨볼루션 뉴럴 네트워크(3D Convoultion neural network, 200), 감정 상태 추정모듈(300)이 포함될 수 있다.

[0015] 사용자의 음성이 어떠한 음성 녹음 장치를 통해 본 발명의 시스템에 수신되면, 음성 신호(100)는 세그먼트 단위로 전환되어 인풋값은 고속퓨리에변환(fast fourier transform)을 사용하여 스펙트로그램으로 변환된다.

[0016] 이때 스펙트로그램은 특정 파형에 존재하는 다양한 주파수에서 시간에 다른 신호가 가지는 파형의 강도 또는 크기에 대한 시각적인 인식 값으로, 스펙트럼 그래프으로 표현할 수 있는데, 2차원 표현에서 X축에는 시간이, Y축에는 주파수 강도를 나타낸다.

[0017] 또한, 주파수의 진폭은 해당 시간 간격의 특정 포인트에서 강도 또는 색상 값을 가지고, 이를 기초로 음성 신호를 작은 조각으로 나누어 계산한다. 그 후 고속퓨리에변환 알고리즘을 각 조각에 적용하여 주파수 스펙트럼의 크기를 계산한다. 고속퓨리에변환은 특정한 시간에 진폭에 해당하는 각 조각의 열(row) 값을 반환하고, 스펙트로그램 이미지를 형성하기 위해 나란히 열게 된다.

[0018] 생성된 스펙트로그램은 특징 추출 및 분류를 위해 3D 컨볼루션 뉴럴 네트워크(200)로 전달되는데, 이때 3D 컨볼루션 뉴럴 네트워크(200)의 마지막 층은 행복(happy), 슬픔(sad), 분노(angry), 공포(fear), 혐오(disgust), 지루함(boredom) 및 중립(neutral)의 7가지 감정 분류로 구성된다.

[0019] 따라서 음성 세그먼트에 대한 감정은 7개의 확률을 가지는 신뢰 레벨이 결정되면, 이를 통해 낮은 신뢰도 값을 처리하기 위해 각 확률 값의 상, 중 하의 신뢰도로 메커니즘을 분류한다. 신뢰도가 높으면 다음 단계에서는 충돌하는 감정을 추가적으로 검사하고, 작은 차이와 상충되는 감정을 가지는 신뢰도 값은 현재 화자의 공존 감정과 감정 기록에 대한 지식을 기반으로 구성된 본 발명에서 제안된 증거, 검증 및 추론을 통해 해결하게 된다.

[0020] 이때, 각 확률 값의 상, 중, 하의 신뢰도 메커니즘의 분류하는 상세한 방법은 도 7을 통해 설명하도록 한다.

[0021] 감정 상태 추정모듈(300)에서 감정 상태 추정값은 음성의 분석을 통해 앞서 설명한 7가지의 감정 분류에 현재의 음성이 얼마나 충족하는가에 대한 것으로 행복(happy), 슬픔(sad), 분노(angry), 공포(fear), 혐오(disgust), 지루함(boredom) 및 중립(neutral)이 각각의 확률 점수로 표현될 수 있고, 필요에 따라 7개 이상의 감정이나 그 이하의 감정이 사용될 수 있다.

[0022] 감정 상태 추정모듈(300)에서 감정 상태 추정값이 산출되는 과정을 자세하게 설명하면, 3D 컨볼루션 뉴럴 네트워크(200)를 통해 각 세그먼트의 단위마다 음성 데이터의 특징을 추출하고, 추출된 특징값에 기초하여 미리 설정된 복수의 감정 상태를 분류하는 제 1 신뢰도와 제 1 신뢰도를 산출하기 직전인 과거를 기준으로 전체 음성 데이터 대비 과거 분류 이력을 나타내는 제 2 신뢰도, 앞서 서술한 7가지 감정이 공존할 수 있는 확률을 지식 베이스에 기반하여 분류 대상 세그먼트의 음성 데이터를 가질 수 있는 감정의 상태 확률인 제 3 신뢰도를 결합하여 감정 상태 추정값을 산출하게 된다.

[0023] 도 2는 본 발명의 일 실시예에 따른, 연속적인 음성 세그먼트에서의 복수의 감정이 공존할 시, 감정마다의 관계를 확률로 나타낸 도면이다.

[0024] 도 2를 살펴보면, 연속적인 음성 세그먼트에서의 복수의 감정이 공존할 확률은 각각 음성 세그먼트에 대한 시퀀스가 2개 혹은 3개의 감정이 동시에 표현된 세트를 확인할 수 있다.

[0025] 최종 감정 클래스는 해결된 갈등과 중간 감정의 집계 예측을 통해 높은 수준의 감정 및 충돌하지 않은 감정을 결합하여 예측하게 된다. 본 발명의 일 실시예로 각 시퀀스의 감정 개수를 측정하고, 가장 반복적인 감정을 선택한 후 시퀀스의 특정 감정 개수를 추가하여 이를 총 감정 개수로 나눈다. 이때, 높은 값을 가진 감정이 관찰 중인 음성 신호로 선택되고, 신뢰도가 높은 스펙트로그램과 이들의 주석은 정확도를 향상시키고 화자의 감정적

기록을 업데이트하기 위해 데이터베이스에 누적 저장된다.

[0026] 도 3은 종래의 충돌하고 공존하는 감정의 상태 사이에 관계를 나타낸 상태 전환 다이어그램이다.

[0027] 도 3를 살펴보면, 상태 전환 다이어그램은 사람이 현재의 감정에서 다른 감정으로 변할 수 있는 관계를 확률로 제 3 신뢰도로 정의될 수 있다.

[0028] 이러한 확률은 심리학적 연구와 전문가의 의견에 기초하여 할당될 수 있고, 서로 갈등하거나 반대의 감정은 공존하는 감정에 비해 낮은 확률 점수를 가지게 된다.

[0029] 도 3에서는 7가지의 감정을 점선과 실선으로 연결하였는데, 여기서 서로 충돌하는 감정은 점선을, 공존할 수 있는 감정을 실선으로 표기하였다. 예를들어 사람의 감정 상태가 "행복"에서 "분노"상태로 변할 수 있는 확률은 0.12(충돌 감정)값을 가지고, "행복"에서 "중립"상태는 0.6(공존 감정)의 확률값을 가질 수 있다.

[0030] 도 4는 본 발명의 일 실시예에 따른, 3가지 신뢰도을 합산하여 감정 상태 추정모듈(300)에서 감정 상태 추정값을 산출하는 방식과 그 예시를 나타낸 도면이다.

[0031] 도 4를 살펴보면, 3가지 신뢰도를 이용하여 감정상태를 결정하기 위한 공식(a)와 공식에 사용되는 가중치 값(b)를 활용하여 연산작업이 수행된다.

[0032] 공식(a)에서 S_1, S_2, S_3은 각각 제 1 신뢰도, 제 2 신뢰도, 제 3 신뢰도의 점수를 뜻하고, 각각의 신뢰도 값에 주어지는 가중치 (b)인 r_1, r_2, r_3,을 통해 감정 상태 추정모듈(300)에서 감정 상태 추정값을 도출할 수 있다.

[0033] 이때, 가중치 (b)는 심리학적 연구 및 누적된 데이터를 기초로 산출되고 가변형 수치가 될 수 있다.

[0034] 그림 (c)를 통해 예를 들어 설명하면, 신뢰도 값은 아래의 표1와 같고 가중치 (b)를 활용하게 되면 아래의 공식을 따르게 된다.

표 1

[0035]

	제 1 신뢰도	제 2 신뢰도	제 3 신뢰도
행복	0.5	0.6	0.99
중립	0.58	0	0.79

[0036] 각각의 감정이 가지는 감정 상태 추정값은 다음과 같이 산출될 수 있다.

[0037] 케이스 1 : F(happy) = (0.15*0.5+0.5*0.6+0.35*0.99) = 0.7215

[0038] 케이스 2 : F(neutral) = (0.15*0.58+0.5*0+0.35*0.79) = 0.3635

[0039] 따라서 위의 공식에 따라 사용자는 현재 행복이라는 감정을 가질 확률이 0.7215, 중립이라는 감정을 가질 확률이 0.3635일 수 있다. 이때 도 4에서는 행복 및 중립이라는 2가지 감정을 통해 예시를 제시하였으나, 실제 발명에서는 7가지의 감정에 대한 확률을 모두 계산하게 된다.

[0040] 도 5는 본 발명의 일 실시예에 따른, 음성 세그먼트에서 감정을 추출하는 과정을 나타낸 예시 도면이다.

[0041] 먼저, 도 5의 그림 (a)를 살펴보면 이전의 N개의 샘플에서 고유한 감정의 숫자가 계산될 수 있다. 이전 감정의 예시는 행복/행복/중립/행복/행복의 값을 가지고, 각각 E1: 행복(happy) = 4, E2: 중립(neutral) = 1로 N = 5개의 값을 가진다.

[0042] 또한, 예시로 주어진 현재 추정하는 감정(즉, 본 발명에서는 이를 제 1신로도로 정의 한다.)은 슬픔=0.34, 행복=0.31을 제시하고 있다. 실제로 본 발명에서는 7가지의 감정(행복(happy), 슬픔(sad), 분노(angry), 공포(fear), 혐오(disgust), 지루함(boredom) 및 중립(neutral))에 대한 모든 확률을 계산하게 된다.

[0043] 그림 (b)는 그림 (a)에서 제시된 파라미터값을 이용하여, 사용자의 감정 충돌을 해결하고, 감정 상태 추정모듈(300)에서 감정 상태 추정값을 산출하는 방법을 타나낸 예시이다.

[0044] 먼저 복수의 이전의 감정에 대한 확률값은

[0045] emotion_x score = no. of occurrences of emotion_x in N / N

[0046] 의 공식을 통해 계산할 수 있다.

[0047] 그림 (a)에서 주어진 값은 각각 E1: 행복=4, E2: 중립= 1, N = 5이기에

[0048] 케이스 1 : 행복(happy) = 4/5 = 0.8

[0049] 케이스 2 : 중립(neutral) = 1/5 = 0.2

[0050] 라는 계산을 통해 높은 값인 행복(happy) = 0.8을 사용자가 이전에 가지는 감정은 행복이라고 결정할 수 있다.

[0051] 다음으로 데이터베이스에 기 저장된 연속적인 음성 세그먼트에서의 복수의 감정이 공존할 확률을 통해 제 3 신뢰도 값을 도출하게 된다.

[0052] 그림 (b)의 표를 참조하면 현재의 감정이 슬픔일 경우를 케이스1로, 행복일 경우를 케이스 2로 가정하고, 이전의 감정을 각각 행복(happy) - 중립(neutral) 순으로 가정하면,

[0053] 케이스 1은 행복(happy) - 중립(neutral) - 슬픔(sad)의 순서를 가지게 되고 매핑되는 제 3 신뢰도 값은 0.2가 된다. 또한, 케이스2는 행복(happy) - 중립(neutral) - 행복(happy) 의 순서를 가지게되고 매핑되는 제 3 신뢰도는 0.7이 된다.

[0054] 이를 도4에서 제시한 방법을 감정 상태 추정값은 각각 슬픔(sad)이 0.121, 행복(happy)이 0.691이 산출되고, 시스템은 사용자가 느끼는 감정이 행복이라고 결론짓게 된다.

[0055] 도 6은 본 발명의 일 실시예에 따른, 추론 프로세스를 이용한 음성 감정 인식의 방법을 나타낸 동작 흐름도이다.

[0056] 도 6을 참조하면, 추론 프로세스를 이용한 음성 감정 인식의 방법은 사용자로부터 음성 신호(100)를 수신하는 단계를 가진다(S610).

[0057] 이때, 사용자의 음성 신호(100)는 다양한 음성 녹음장치를 통해 실시간으로 입력될 수 있다.

[0058] 단계(S610)에서 수신한 음성 신호(100)는 세그먼트 단위의 음성 데이터로 변환한다(S620).

[0059] 음성 신호(100)는 고속퓨리에변환을 통해 세그먼트 단위를 가지는 2차원 그래프의 형태인 스펙트로그램이 생성되고, 이를 시간단위로 분할하여 3차원 구조로 재배치하게 된다.

[0060] 다음으로 변환된 데이터를 3D 컨볼루션 뉴럴 네트워크(200)를 통해 특징값을 추출하게 된다(S630).

[0061] 이때 특징값은 현재 음성에서 추출할 수 있는 제 1 신뢰도와 기 추출되어 판단된 감정값인 제 2 신뢰도를 산출하게 된다.

[0062] 마지막으로 특징값에서 각 음성 데이터의 구간에 매핑하는 감정 상태 추정값을 산출하여 단계를 마무리한다(S640).

[0063] 도 5에서 제시된 방법을 통해, 제 1 신뢰도, 제 2 신뢰도, 제 3 신뢰도를 이용하여 각 감정에 대한 감정 상태 추정값을 산출하고, 산출한 값 중 가장 높은 수치를 지니는 감정을 선택하여 사용자에게 제공하게 된다.

[0064] 도 7은 본 발명의 일 실시예에 따른, 3D 컨볼루션 뉴럴 네트워크(200)를 통해 반환된 신뢰도 확인을 통해 신뢰레벨의 높낮음을 나타낸 도면이다.

[0065] 예를 들어 도 7을 설명하면, 높은 수준과 중간 수준의 기 설정된 임계값을 가정하고, 처음 조건에서 도출된 신뢰도가 높은 수준으로 설정된 임계값 보다 높으면 신뢰 수준을 "높은 신뢰도"라고 정한 후 충돌하는 감정으로 정의 후 처리한다.

[0066] 만약 높은 수준의 임계값 보다 낮으면, 중간 수준으로 설정된 임계값과 비교하고, 해당 임계값보다 높으면 "중간 신뢰도"라고 정의하고, 그 이하이면 "낮은 신뢰도"로 분류하여 해당 값을 폐기한다.

[0067] 이때, 신뢰 점수가 "중간 신뢰도"라면, 도 2에 제시된 과정을 거치게 된다.

[0068] 이상에서 설명한 본 발명의 실시예에 따른 최적의 학습 모델 선택 방법은, 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행 가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 이러한 기록 매체는 컴퓨터 판독 가능 매체를 포함하며, 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판

독가능 매체는 컴퓨터 저장 매체를 포함하며, 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 포함한다.

[0069] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 조사 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

[0070] 또한, 본 발명의 방법 및 시스템은 특정 실시예와 관련하여 설명되었지만, 그것들의 구성 요소 또는 동작의 일부 또는 전부는 범용 하드웨어 아키텍처를 갖는 컴퓨터 시스템을 사용하여 구현될 수도 있다.

[0071] 본 발명의 범위는 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

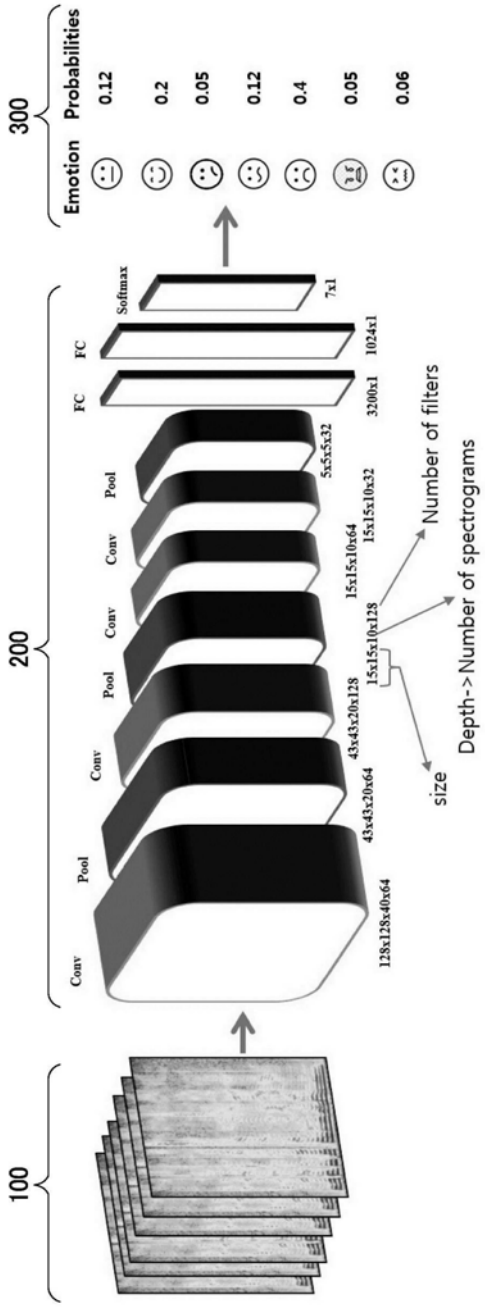
부호의 설명

[0072] 100: 음성 신호 200: 3D 컨볼루션 뉴럴 네트워크

300: 감정 상태 추정모듈

도면

도면1


100
200
300
Emotion
Probabilities
0.12
0.2
0.05
0.12
0.4
0.05
0.06
Softmax
7x1
FC
1024x1
FC
3200x1
Pool
5x5x32
Conv
15x15x10x32
Conv
15x15x10x64
Pool
15x15x10x128
Conv
43x43x20x64
Pool
43x43x20x128
Conv
128x128x40x64
Number of filters
Depth->Number of spectrograms
size

도면2

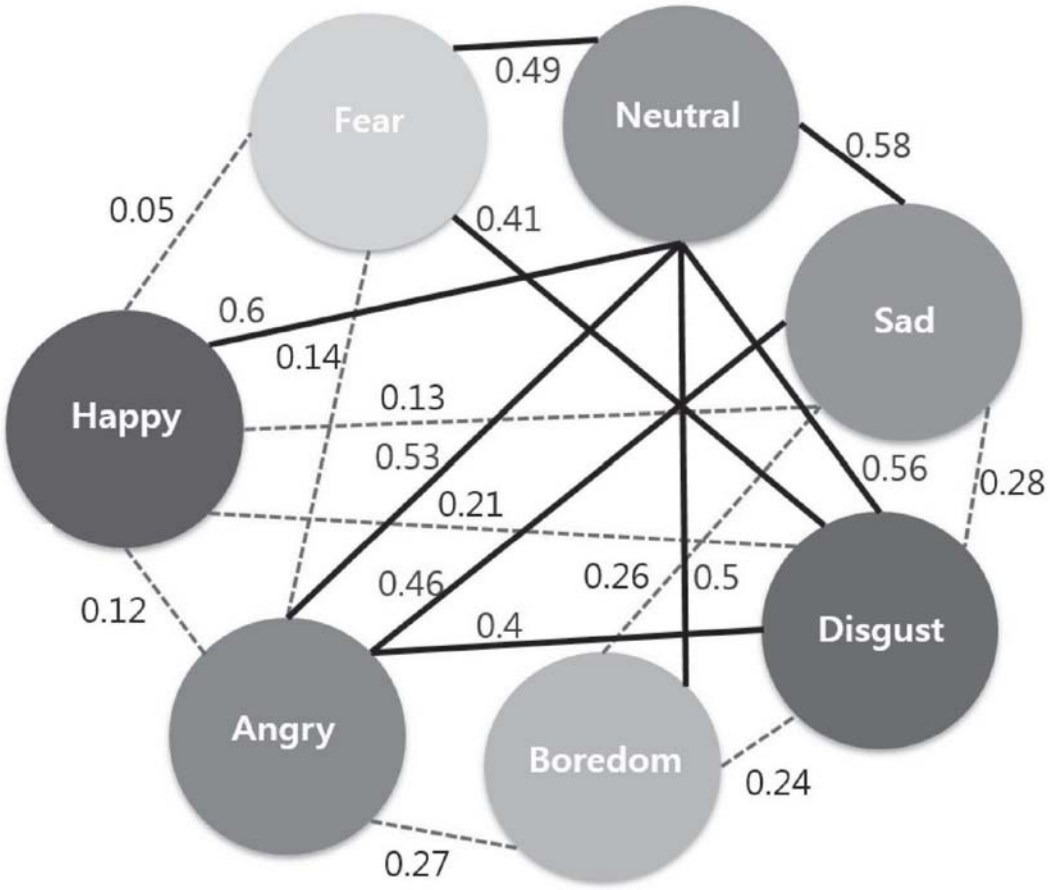
Emotion 1	Emotion 2	Probability
Happy	Happy	0.98
Happy	Neutral	0.6
Happy	Sad	0.05
Sad	Sad	0.98
Sad	Angry	0.46
Angry	Angry	0.98
Angry	Fear	0.14
Fear	Neutral	0.5
Boredom	Disgust	0.24
Disgust	Neutral	0.58
.....	.....	.....

(a)

Emotion 1	Emotion 3	Emotion 2	Probability
Happy	Happy	Happy	0.98
Happy	Neutral	Happy	0.83
Happy	Neutral	Neutral	0.8
Neutral	Sad	Sad	0.73
Fear	Neutral	Fear	0.64
Sad	Boredom	Neutral	0.41
Fear	Fear	Angry	0.2
Sad	Fear	Boredom	0.3
Boredom	Happy	Neutral	0.1
Happy	Sad	Happy	0.1
.....	.....	.....	.....

(b)

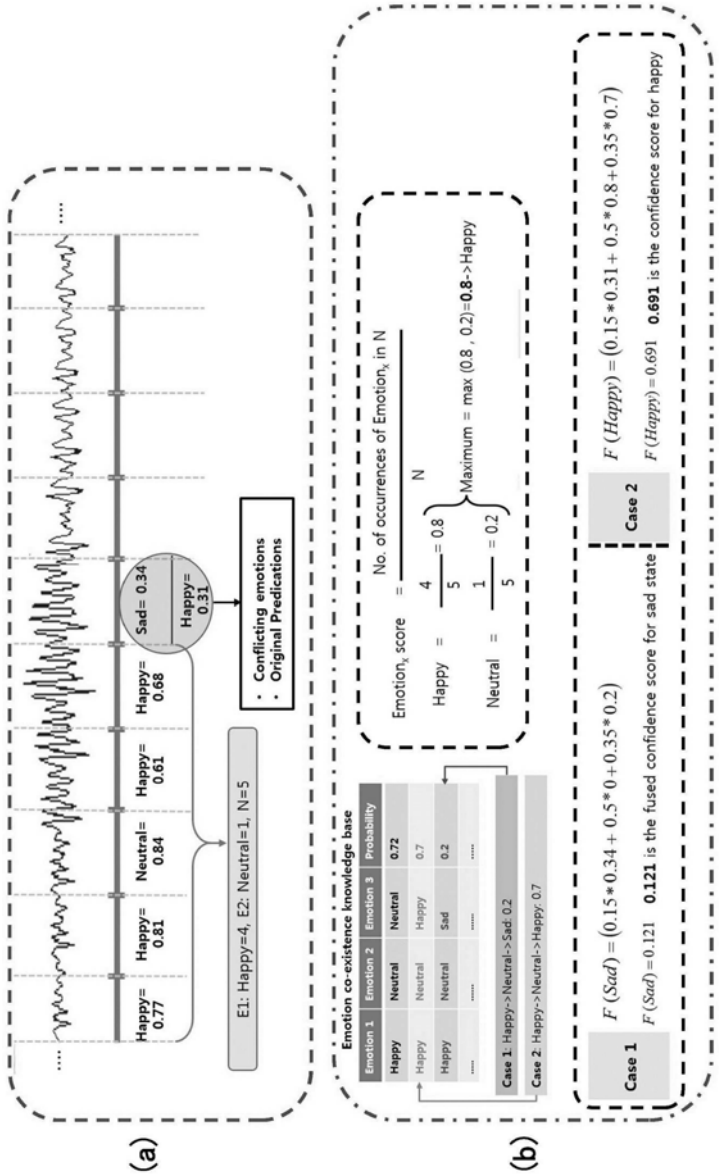
도면3


Fear
Neutral
Sad
Happy
Disgust
Angry
Boredom
0.49
0.58
0.05
0.41
0.6
0.14
0.13
0.53
0.56
0.28
0.21
0.46
0.26
0.5
0.12
0.4
0.24
0.27

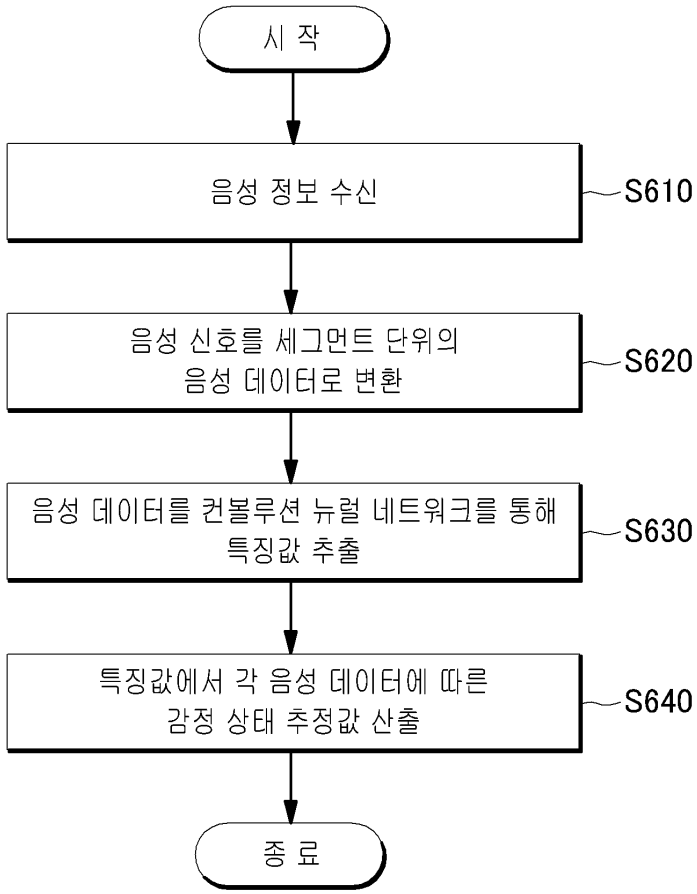
도면4

Fusion (Emotion State) = (r1S1 + r2S2 + r3S3)
(a)
Reliability weights
Original Prediction
Speaker's History
Emotion Co-existence
r1
r2
r3
0.15
0.5
0.35
(b)
Source 1
Original Prediction
Score 1 = 0.5-> Happy
Score 2 = 0.58-> Neutral
Source 2
Speaker's History
Score 1 = 0.6-> Happy
Score 2 = 0-> Neutral
Source 3
Emotion Co-existence
Score 1= 0.99->Happy
Score 2= 0.79->Neutral
F (Happy) = (0.15 * 0.5 + 0.5 * 0.6 + 0.35 * 0.99)
F (Happy) = 0.7215
F (Neutral) = (0.15 * 0.58 + 0.5 * 0 + 0.35 * 0.79)
F (Neutral) = 0.3635
(c)

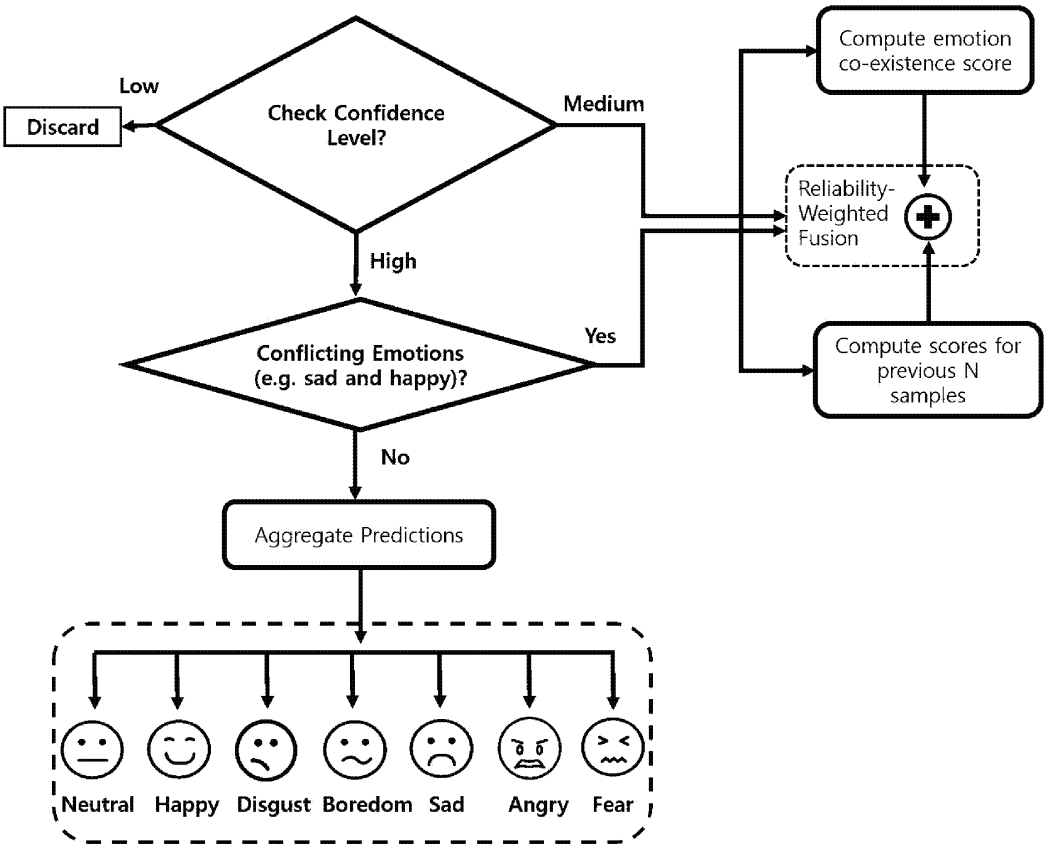
도면5


(a)
Happy= 0.77
Happy= 0.81
Neutral= 0.84
Happy= 0.61
Happy= 0.68
Sad= 0.34
Happy= 0.31
E1: Happy=4, E2: Neutral=1, N=5
· Conflicting emotions
· Original Predications
(b)
Emotion co-existence knowledge base
Emotion 1 | Emotion 2 | Emotion 3 | Probability
Happy | Neutral | Neutral | 0.72
Happy | Neutral | Happy | 0.7
Happy | Neutral | Sad | 0.2
Case 1: Happy->Neutral->Sad: 0.2
Case 2: Happy->Neutral->Happy: 0.7
Emotion_x score = No. of occurrences of Emotion_x in N / N
Happy = 4/5 = 0.8
Neutral = 1/5 = 0.2
Maximum = max (0.8 , 0.2)=0.8->Happy
Case 1
F(Sad) = (0.15*0.34 + 0.5*0 + 0.35*0.2)
F(Sad) = 0.121 0.121 is the fused confidence score for sad state
Case 2
F(Happy) = (0.15*0.31 + 0.5*0.8 + 0.35*0.7)
F(Happy) = 0.691 0.691 is the confidence score for happy

도면6


시 작
음성 정보 수신
S610
음성 신호를 세그먼트 단위의
음성 데이터로 변환
S620
음성 데이터를 컨볼루션 뉴럴 네트워크를 통해
특징값 추출
S630
특징값에서 각 음성 데이터에 따른
감정 상태 추정값 산출
S640
종 료

도면7


Low
Discard
Check Confidence Level?
Medium
High
Conflicting Emotions (e.g. sad and happy)?
Yes
No
Aggregate Predictions
Compute emotion co-existence score
Reliability-Weighted Fusion
Compute scores for previous N samples
Neutral
Happy
Disgust
Boredom
Sad
Angry
Fear