(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2017년04월25일
(11) 등록번호 10-1729189
(24) 등록일자 2017년04월17일

(51) 국제특허분류(Int. Cl.)
H04R 1/10 (2006.01) *G06N 3/04* (2006.01)
(52) CPC특허분류
H04R 1/1091 (2013.01)
G06N 3/04 (2013.01)
(21) 출원번호 10-2016-0030954
(22) 출원일자 2016년03월15일
심사청구일자 2016년03월15일
(56) 선행기술조사문헌
KR1020120103286 A*
JP05012242 A*
KR1020100066352 A*
KR1020130039267 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
한림대학교 산학협력단
강원도 춘천시 한림대학길 1, 한림대학교(옥천동)
주식회사 비에스엘
강원도 춘천시 신북읍 신북로 61-14, 103호, 104호(강원테크노파크 춘천벤처2공장)
(72) 발명자
김백섭
강원도 춘천시 춘주로 176-22, 105동 504호(퇴계동, 퇴계동 쌍용스윗닷홈)
(74) 대리인
특허법인 천지

전체 청구항 수 : 총 7 항 심사관 : 우만웅

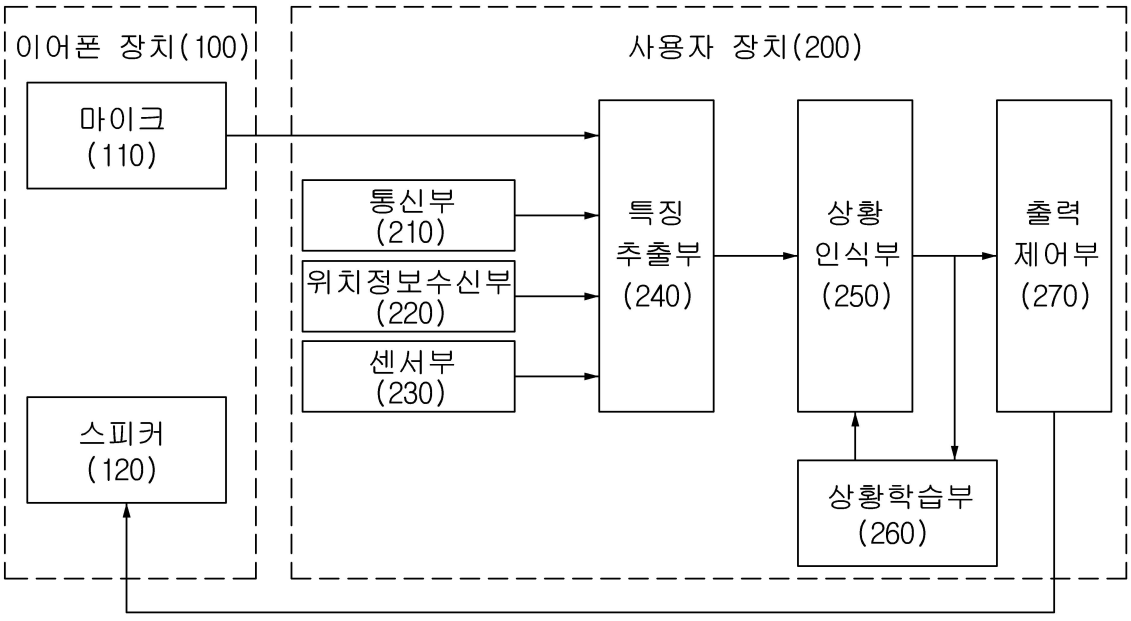
(54) 발명의 명칭 **상황 인식 기반의 이어폰 출력을 제어하기 위한 장치, 이를 위한 방법 및 이 방법이 기록된 컴퓨터 판독 가능한 기록매체**

(57) 요 약

본 발명은 상황 인식 기반의 이어폰 출력을 제어하기 위한 장치, 이를 위한 방법 및 이 방법이 기록된 컴퓨터 판독 가능한 기록매체에 관한 것으로, 이러한 본 발명은 마이크 및 스피커를 포함하는 이어폰 장치와 연결되어 상황 인식 기반의 이어폰 출력을 제어하기 위한 장치로서, 사운드 신호, 통신 신호, GPS 신호 및 센서 신호 중 적

(뒷면에 계속)

대 표 도 - 도2



어도 하나를 포함하는 특징 신호로부터 특징 파라미터를 추출하는 특징추출부와, 상기 특징 파라미터를 소정 크기의 입력 행렬로 변환한 후, 복수의 상황이 미리 학습되고 복수의 계층을 포함하는 신경 네트워크에 상기 입력 행렬을 입력하여 출력값을 산출하고, 산출된 출력값에 따라 상황을 인식하는 상황인식부와, 상기 인식의 결과에 따라 상기 스피커로 출력되는 사운드 및 사운드의 볼륨 중 적어도 하나를 제어하는 출력제어부를 포함하는 것을 특징으로 하는 이어폰 출력을 제어하기 위한 장치와, 이에 따른 방법 및 이러한 방법이 기록된 컴퓨터 판독 가능한 기록매체를 제공한다.

(52) CPC특허분류
H04R 2430/01 (2013.01)

명 세 서

청구범위

청구항 1

마이크 및 스피커를 포함하는 이어폰 장치와 연결되어 상황 인식 기반의 이어폰 출력을 제어하기 위한 장치로서,

사운드 신호, 통신 신호, GPS 신호 및 센서 신호 중 적어도 하나를 포함하는 특징 신호로부터 특징 파라미터를 추출하는 특징추출부;

상기 특징 파라미터를 소정 크기의 입력 행렬로 변환한 후, 변환된 입력 행렬을 복수의 연산을 가지는 복수의 계층을 포함하는 신경 네트워크에 입력하여 출력값을 산출하고, 산출된 출력값에 따라 상황을 인식하는 상황인식부; 및

상기 인식의 결과에 따라 상기 스피커로 출력되는 사운드 및 사운드의 볼륨 중 적어도 하나를 제어하는 출력제어부;를 포함하며,

상기 신경 네트워크는 컨볼루션 연산에 따라 산출되는 복수의 특징 지도를 가지는 컨볼루션 계층과 풀링 연산에 따라 산출되는 복수의 특징 지도를 가지는 풀링 계층이 교대로 반복되어 형성되는 특징 추출 네트워크를 포함하며, 상기 컨볼루션 연산 및 상기 풀링 연산은 소정 크기의 행렬인 커널을 이용하고, 상기 커널에 대해 적용되는 가중치는 상기 커널의 행렬 크기에 무관하게 1개인 것을 특징으로 하는 이어폰 출력을 제어하기 위한 장치.

청구항 2

제1항에 있어서,

상기 특징추출부는

시간 도메인에서 상기 사운드 신호를 중복되는 구간 없이 소정 길이의 프레임으로 구분하고, 소정 길이의 프레임 기간의 사운드 신호로부터 FTE(Frame Time Energy) 및 ZCR(Zero-Crossing Rate)를 추출하며, 상기 소정 길이의 프레임 기간의 사운드 신호를 주파수 도메인으로 변환하여 MFCC(Mel-Frequency Cepstral Coefficient) 및 SC(Spectral Centroid)를 추출하여 상기 특징 파라미터로 사용하는 것을 특징으로 하는 이어폰 출력을 제어하기 위한 장치.

청구항 3

제1항에 있어서,

상기 특징추출부는

상기 GPS 신호로부터 검색되는 위성의 개수를 획득하고, 소정 길이 프레임 동안 상기 검색되는 위성으로부터 수신되는 신호세기의 평균을 구한 후, 상기 위성의 개수 및 상기 신호세기의 평균을 상기 특징 파라미터로 사용하는 것을 특징으로 하는 이어폰 출력을 제어하기 위한 장치.

청구항 4

제1항에 있어서,

상기 특징추출부는

상기 센서 신호로부터 직교좌표계의 3축 각각의 가속도 값을 추출하여 상기 특징 파라미터로 사용하는 것을 특징으로 하는 이어폰 출력을 제어하기 위한 장치.

청구항 5

제1항에 있어서,

상기 신경 네트워크는 상기 복수의 계층 중 어느 한 계층에 포함된 복수의 연산으로부터의 출력에 가중치를 적

용한 후 다음 계층의 복수의 연산에 입력하며,

상기 신경 네트워크의 목표값을 미리 결정한 후, 상기 입력 행렬의 입력에 따른 출력값과 상기 목표값의 차이값에 따라 상기 출력값이 상기 목표값이 되도록 상기 가중치를 조정하는 상황학습부;를 더 포함하는 것을 특징으로 하는 이어폰 출력을 제어하기 위한 장치.

청구항 6

삭제

청구항 7

마이크 및 스피커를 포함하는 이어폰 장치와 연결된 사용자 장치의 상황 인식 기반의 이어폰 출력을 제어하기 위한 방법으로,

사운드 신호, 통신 신호, GPS 신호 및 센서 신호 중 적어도 하나를 포함하는 특징 신호로부터 특징 파라미터를 추출하는 단계;

상기 특징 파라미터를 소정 크기의 입력 행렬로 변환하는 단계;

복수의 상황이 미리 학습된 신경 네트워크에 상기 입력 행렬을 입력하여 출력값을 산출하고, 산출된 출력값에 따라 상황을 인식하는 단계; 및

상기 인식의 결과에 따라 상기 스피커로 출력되는 사운드 및 사운드의 볼륨 중 적어도 하나를 제어하는 단계;를 포함하며,

상기 신경 네트워크는 컨볼루션 연산에 따라 산출되는 복수의 특징 지도를 가지는 컨볼루션 계층과 풀링 연산에 따라 산출되는 복수의 특징 지도를 가지는 풀링 계층이 교대로 반복되어 형성되는 특징 추출 네트워크를 포함하며, 상기 컨볼루션 연산 및 상기 풀링 연산은 소정 크기의 행렬인 커널을 이용하고, 상기 커널에 대해 적용되는 가중치는 상기 커널의 행렬 크기에 무관하게 1개인 것을 특징으로 하는 이어폰 출력을 제어하기 위한 방법.

청구항 8

제7항에 따른 이어폰 출력을 제어하기 위한 방법이 기록된 컴퓨터 판독 가능한 기록매체.

발명의 설명

기 술 분 야

[0001] 본 발명은 이어폰 출력을 제어하기 위한 기술에 관한 것으로, 보다 상세하게는, 상황 인식을 기반으로 하여 이어폰 출력을 제어할 수 있는 장치, 이를 위한 방법 및 이 방법이 기록된 컴퓨터 판독 가능한 기록매체에 관한 것이다.

배 경 기 술

[0002] 디지털 기술혁신에 따른 현대적 문화생활의 발달로, 최근에는 PMP, MP3플레이어, DMB 수신기 또는 휴대폰 등과 같은 휴대 단말기를 이용하여 음악을 청취하거나 동영상을 시청하는 이들을 언제 어디서건 어렵지 않게 목격할 수 있다. 이러한 휴대 단말기들은, 기기의 소형화 및 경량화 추세와 더불어 공공장소에서의 에티켓문화 정착정책에 부응하기 위해, 음향을 출력하는 외부스피커가 생략된 기종들의 주류를 이루고 있다. 예컨대, 지하철ㅇ영화관ㅇ도서관 등과 같이 국민 모두가 공동으로 이용하는 공공장소에서는 타인에게 해를 줄 수 있는 소음발생의 원인 차단을 위해, 휴대폰의 벨소리를 진동으로 전환토록 권고함과 아울러 음향출력이 가능한 각종 단말기에는 소유자 자신만의 단독청취를 위한 이어폰 착용을 의무화하고 있다. 그러나, 상술한 이어폰을 귓속에 착용할 시, 소음에 따른 타인의 피해발생 우려를 방지할 수 있으면서도 음향을 집중해서 듣는데 더할 나위 없이 용이한 반면, 귓구멍을 차단함에 따라 주변인의 말소리나 생활 소음을 알아듣지 못 할 수밖에 없다. 이 때문에, 차량의 경적소리와 같은 교통상황에 따른 발생 소음의 인식하지 못해 교통사고가 빈번할 수밖에 없다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) 한국등록특허 제10-1494306호 2015년 02월 11일 등록 (명칭: 선택적으로 외부 소리 정보 청취가 가능한 이어폰 시스템)

발명의 내용

해결하려는 과제

[0004] 본 발명의 목적은 주변 상황을 인식하여 인식된 주변 상황에 따라 이어폰으로 출력을 조절할 수 있는 장치, 이를 위한 방법 및 이 방법이 기록된 컴퓨터 판독 가능한 기록매체를 제공함에 있다.

과제의 해결 수단

[0005] 상술한 바와 같은 목적을 달성하기 위한 본 발명의 바람직한 실시예에 따른 마이크 및 스피커를 포함하는 이어폰 장치와 연결되어 상황 인식 기반의 이어폰 출력을 제어하기 위한 장치는 사운드 신호, 통신 신호, GPS 신호 및 센서 신호 중 적어도 하나를 포함하는 특징 신호로부터 특징 파라미터를 추출하는 특징추출부와, 상기 특징 파라미터를 소정 크기의 입력 행렬로 변환한 후, 신경 네트워크에 상기 입력 행렬을 입력하여 출력값을 산출하고, 산출된 출력값에 따라 상황을 인식하는 상황인식부와, 상기 인식의 결과에 따라 상기 스피커로 출력되는 사운드 및 사운드의 볼륨 중 적어도 하나를 제어하는 출력제어부를 포함한다.

[0006] 상기 신경 네트워크는 복수의 상황이 미리 학습되고 각각이 복수의 연산으로 이루어진 복수의 계층을 포함한다.

[0007] 상기 특징추출부는 시간 도메인에서 상기 사운드 신호를 중복되는 구간 없이 소정 길이의 프레임으로 구분하고, 소정 길이의 프레임 기간의 사운드 신호로부터 FTE(Frame Time Energy) 및 ZCR(Zero-Crossing Rate)를 추출하며, 상기 소정 길이의 프레임 기간의 사운드 신호를 주파수 도메인으로 변환하여 MFCC(Mel-Frequency Cepstral Coefficient) 및 SC(Spectral Centroid)를 추출하여 상기 특징 파라미터로 사용하는 것을 특징으로 한다.

[0008] 상기 특징추출부는 시간 도메인에서 상기 사운드 신호를 중복되는 구간 없이 소정 길이의 프레임으로 구분하고, 구분된 프레임 동안의 에너지인 FTE(Frame Time Energy)를 구하여, 구해진 FTE를 상기 특징 파라미터로 사용하는 것을 특징으로 한다.

[0009] 상기 특징추출부는 상기 사운드 신호로부터 추출되는 특징 파라미터는 시간 도메인에서 상기 사운드 신호를 중복되는 구간 없이 소정 길이의 프레임으로 구분하고, 소정 길이의 프레임 기간 동안 사운드 신호가 진폭축의 0을 지나는 빈도를 나타내는 ZCR(Zero-Crossing Rate)을 구하여, 구해진 ZCR을 상기 특징 파라미터로 사용하는 것을 특징으로 한다.

[0010] 상기 특징추출부는 시간 도메인에서 상기 사운드 신호를 중복되는 구간 없이 소정 길이의 프레임으로 구분하고, 소정 길이 프레임 단위의 사운드 신호를 주파수 도메인으로 변환하며, 주파수 도메인으로 변환된 사운드 신호를 멜-필터뱅크(Mel-filterbank)를 적용하여 멜-파워(Mel-power)를 구하고, 상기 멜-파워를 로그(log) 연산하며, 로그 연산된 멜 파워를 DCT(Discrete Cosine Transform) 연산하여 MFCC(Mel-Frequency cepstral coefficients)를 구하고, 구해진 MFCC를 특징 파라미터로 사용하는 것을 특징으로 한다.

[0011] 상기 특징추출부는 상기 사운드 신호를 시간 도메인에서 중복되는 구간 없이 소정 길이의 프레임으로 구분하고, 소정 길이 프레임 단위의 사운드 신호를 주파수 도메인으로 변환하며, 주파수 도메인의 사운드 신호에서 각 주파수의 진폭을 구한 후, 중심 주파수의 위치를 나타내는 SC를 구하여 특징 파라미터로 사용하는 것을 특징으로 한다.

[0012] 상기 특징추출부는 상기 GPS 신호로부터 검색되는 위성의 개수를 획득하고, 소정 길이 프레임 동안 상기 검색되는 위성으로부터 수신되는 신호세기의 평균을 구한 후, 상기 위성의 개수 및 상기 신호세기의 평균을 상기 특징 파라미터로 사용하는 것을 특징으로 한다.

[0013] 상기 특징추출부는 상기 센서 신호로부터 직교좌표계의 3축 각각의 가속도 값을 추출하여 상기 특징 파라미터로 사용하는 것을 특징으로 한다.

[0014] 상기 신경 네트워크는 상기 복수의 계층의 출력이 가중치(w)가 곱해져서 다음 계층의 입력이 되며, 상기 신경 네트워크의 목표값을 미리 결정한 후, 상기 입력 행렬의 입력에 따른 출력값과 상기 목표값의 차이값에 따라 상기 출력값이 상기 목표값이 되도록 상기 가중치를 조정하는 상황학습부를 더 포함하는 것을 특징으로 한다.

[0015] 상기 신경 네트워크에서 상기 복수의 계층 중 어느 한 계층에 포함된 복수의 연산의 출력은 가중치(w)가 적용되어 다음 계층의 복수의 연산의 입력이 된다. 이에 따라, 상기 이어폰 출력을 제어하기 위한 장치는 상기 신경 네트워크의 목표값을 미리 결정한 후, 상기 입력 행렬의 입력에 따른 출력값과 상기 목표값의 차이값에 따라 상기 출력값이 상기 목표값이 되도록 상기 가중치를 조정하는 상황학습부를 더 포함한다.

[0016] 상기 신경 네트워크는 컨볼루션 연산에 따라 산출되는 복수의 특징 지도를 가지는 컨볼루션 계층과 풀링 연산에 따라 산출되는 복수의 특징 지도를 가지는 풀링 계층이 교대로 반복되어 형성되는 특징 추출 네트워크를 포함하며, 상기 컨볼루션 연산에 사용되는 커널 및 상기 풀링 연산에 사용되는 커널 중 어느 하나의 커널에 적용되는 가중치는 상기 커널의 행렬의 크기에 무관하게 1개인 것을 특징으로 한다.

[0017] 상술한 바와 같은 목적을 달성하기 위한 본 발명의 바람직한 실시예에 따른 마이크 및 스피커를 포함하는 이어폰 장치와 연결된 사용자 장치의 상황 인식 기반의 이어폰 출력을 제어하기 위한 방법은 사운드 신호, 통신 신호, GPS 신호 및 센서 신호 중 적어도 하나를 포함하는 특징 신호로부터 특징 파라미터를 추출하는 단계와, 상기 특징 파라미터를 소정 크기의 입력 행렬로 변환하는 단계와, 상기 입력 행렬을 복수의 상황이 미리 학습되고 복수의 계층을 포함하는 신경 네트워크에 입력하여 출력값을 산출하고, 산출된 출력값에 따라 상황을 인식하는 단계와, 상기 인식의 결과에 따라 상기 스피커로 출력되는 사운드 및 사운드의 볼륨 중 적어도 하나를 제어하는 단계를 포함한다.

[0018] 또한, 전술한 본 발명의 바람직한 실시예에 따른 이어폰 출력을 제어하기 위한 방법이 기록된 컴퓨터 판독 가능한 기록매체를 제공한다.

발명의 효과

[0019] 본 발명에 따르면, 다양한 신호, 즉, 사운드 신호, 통신 신호, GPS 신호 및 센서 신호로부터 추출되는 특징 파라미터를 이용하여 상황을 학습하고, 인식하기 때문에 인식율이 높다. 또한, 인식된 상황에 의해 이어폰의 출력을 제어할 수 있어 사용자는 실외에서 발생할 수 있는 주변의 위험한 상황에 대비할 수 있다.

도면의 간단한 설명

[0020] 도 1은 본 발명의 실시예에 따른 상황 인식 기반의 이어폰 출력을 제어하기 위한 장치의 외형을 설명하기 위한 도면이다.

도 2는 본 발명의 실시예에 따른 상황 인식 기반의 이어폰 출력을 제어하기 위한 장치의 구성을 설명하기 위한 블록도이다.

도 3 내지 도 8은 본 발명의 실시예에 따른 마이크로부터 출력되는 사운드 신호로부터 추출되는 특징 파라미터를 설명하기 위한 도면이다.

도 9는 본 발명의 실시예에 따른 GPS 신호로부터 추출되는 특징 파라미터를 설명하기 위한 도면이다.

도 10은 본 발명의 실시예에 따른 센서 신호로부터 추출되는 특징 파라미터를 설명하기 위한 도면이다.

도 11 내지 도 14는 본 발명의 실시예에 따른 신경 네트워크를 설명하기 위한 도면이다.

도 15는 본 발명의 실시예에 따른 학습 방법을 설명하기 위한 흐름도이다.

도 16은 본 발명의 실시예에 따른 상황 인식 기반의 이어폰 출력을 제어하기 위한 방법을 설명하기 위한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0021] 본 발명의 상세한 설명에 앞서, 이하에서 설명되는 본 명세서 및 청구범위에 사용된 용어나 단어는 통상적이거나 사전적인 의미로 한정해서 해석되어서는 아니 되며, 발명자는 그 자신의 발명을 가장 최선의 방법으로 설명하기 위해 용어의 개념으로 적절하게 정의할 수 있다는 원칙에 입각하여 본 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야만 한다. 따라서 본 명세서에 기재된 실시예와 도면에 도시된 구성은 본 발명의 가장 바람직한 실시예에 불과할 뿐, 본 발명의 기술적 사상을 모두 대변하는 것은 아니므로, 본 출원시점에 있어서 이들을 대체할 수 있는 다양한 균등물과 변형 예들이 있을 수 있음을 이해하여야 한다.

[0022] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예들을 상세히 설명한다. 이때, 첨부된 도면에서 동일한

구성 요소는 가능한 동일한 부호로 나타내고 있음을 유의해야 한다. 또한, 본 발명의 요지를 흐리게 할 수 있는 공지 기능 및 구성에 대한 상세한 설명은 생략할 것이다. 마찬가지의 이유로 첨부 도면에 있어서 일부 구성요소는 과장되거나 생략되거나 또는 개략적으로 도시되었으며, 각 구성요소의 크기는 실제 크기를 전적으로 반영하는 것이 아니다.

[0023] 먼저, 본 발명의 실시예에 따른 상황 인식 기반의 이어폰 출력을 제어하기 위한 장치에 대해서 설명하기로 한다. 도 1은 본 발명의 실시예에 따른 상황 인식 기반의 이어폰 출력을 제어하기 위한 장치의 외형을 설명하기 위한 도면이며, 도 2는 본 발명의 실시예에 따른 상황 인식 기반의 이어폰 출력을 제어하기 위한 장치의 구성을 설명하기 위한 블록도이다.

[0024] 도 1에 보인 바와 같이, 본 발명의 실시예에 따른 상황 인식 기반의 이어폰 출력을 제어하기 위한 장치는 이어폰 장치(100) 및 사용자 장치(200)를 포함한다. 이어폰 장치(100)와 사용자 장치(200)는 유선 혹은 무선으로 연결될 수 있다. 특히, 이어폰 장치(100)는 사운드를 출력하는 스피커(110)와 스피커(110)에 대향하여 외향의 사운드를 입력받는 마이크(120)를 포함한다.

[0025] 사용자 장치(200)는 통신부(210), 위치정보수신부(210), 센서부(220), 특징추출부(240), 상황인식부(250), 상황학습부(260) 및 출력제어부(270)를 포함한다.

[0026] 통신부(210)는 프로파일 서버(100)와 통신을 위한 것으로, 광대역이동통신 혹은 무선근거리통신 방식을 이용하여 통신할 수 있다. 이러한 통신부(210)는 광대역이동통신 혹은 무선근거리통신 연결 방식의 통신 기능이 모두 구현되어 있는 하나의 모듈로 구현되어, 전술한 2가지 기능 중 적어도 하나를 선택하여 동작하는 것으로 설명될 것이다. 하지만, 통신부(210)는 광대역이동통신, 무선근거리통신 연결 방식 각각의 통신 기능을 별도로 수행하는 복수의 모듈로 구현될 수도 있다. 광대역이동통신은 WCDMA, LTE, LTE-A 등의 표준에 따라 기지국을 통해 네트워크에 접속하는 통신 방식을 예시할 수 있고, 무선근거리통신은 Wi-Fi(wireless fidelity)를 이용하는 WLAN(Wireless Local Area Network) 방식에 따라 접속포인트(AP: Access Point)를 통해 네트워크에 접속하는 통신을 예시할 수 있다. 통신부(210)는 송신되는 신호의 주파수를 상승 변환 및 증폭하는 RF송신기와, 수신되는 신호를 저 잡음 증폭하고 주파수를 하강 변환하는 RF수신기 등으로 구성될 수 있다. 또한, 통신부(210)는 무선 채널을 통해 데이터를 포함하는 무선 신호를 수신하여 제어부(280)로 전달할 수 있다. 또한, 제어부(280)로부터 전달 받은 데이터를 무선 신호로 변환하여 무선 채널을 통해 전송할 수 있다.

[0027] 위치정보수신부(220)는 위치 정보를 수신하기 위한 것이다. 예컨대, 위치정보수신부(220)는 지속적으로 GPS 위성(도시되지 않음)으로부터 GPS 신호를 수신하여, 수신된 GPS 신호로부터 위치 정보를 추출한다. 그런 다음, 위치정보수신부(220)는 위치 정보를 제어부(280)로 전달할 수 있다. 이러한 위치 정보는 위도, 경도, 고도 등의 좌표가 될 수 있다.

[0028] 센서부(230)는 사용자 장치(200)의 움직임의 변위, 속도 등을 감지하기 위한 것이다. 센서부(230)는 사용자 장치(200)의 움직임의 변위를 감지하여, 감지된 움직임의 변위를 좌표 데이터를 통해 제어부(280)에 제공할 수 있다. 센서부(230)가 제공하는 좌표는 임의의 기준 위치로부터 변위를 3차원 직교좌표계(x, y, z) 혹은 위도, 경도 및 고도로 제공할 수 있다. 추가로, 센서부(230)가 제공하는 좌표는 요(yaw), 피치(pitch), 롤(roll)의 변위를 제공할 수 있다. 더욱이, 센서부(230)는 이러한 움직임의 속도를 감지하여, 그 속도를 제어부(280)에 제공할 수 있다. 센서부(230)는 하나 이상의 센서를 통해 구현될 수 있으며, 이러한 센서는 가속도 센서(accelerometer), 자이로 센서(gyroscope), 지자기 센서(magnetometer) 등이 될 수 있다.

[0029] 특징추출부(240)는 본 발명의 실시예에 따른 마이크(110), 통신부(210), 위치정보수신부(220), 센서부(230)로부터 수집되는 특징 신호로부터 상황 인식을 위한 학습에 필요한 특징 파라미터를 추출하는 기능을 수행한다. 일례로, 마이크(110)로부터 수신되는 사운드 신호의 시간 도메인으로부터 FTE(Frame Time Energy) 및 ZCR(Zero-crossing Rate)을 추출하고, 주파수 도메인으로부터 MFCC(Mel-Frequency Cepstral Coefficient) 및 SC(Spectral Centroid)를 추출한다. 또한, 통신부(210)를 통해 검색되는 AP(Access Point)의 개수 및 각 AP로부터의 신호 세기의 평균을 추출한다. 그리고 위치정보수신부(220)를 통해 검색되는 위성의 개수와 각 위성으로부터의 신호 세기의 평균을 추출한다. 또한, 센서부(230)로부터 센서 데이터를 추출한다. 예컨대, 가속도 센서의 경우, 직교좌표계 상의 3개의 축(x, y, z)의 가속도 데이터를 특징 파라미터로 추출한다.

[0030] 상황인식부(250)는 입력되는 특징 파라미터를 분석하여 상황을 인식하기 위한 것이다. 예컨대, 상황인식부(250)는 사용자의 위치가 실내인지 혹은 실외인지 여부 그리고 사용자에게 위급상황이 발생했는지 여부를 인식할 수 있다. 이러한 상황 인식은 딥 러닝(deep learning) 기법을 통해 이루어진다. 이에 따라, 상황인식부(250)는

신경 네트워크를 통해 구현될 수 있다. 이러한 신경 네트워크는 컨볼루션 신경 네트워크(CNN: convolutional neural network)가 될 수 있다. 따라서 상황인식부(250)는 상황학습부(260)에 의해 학습 데이터인 특징 파라미터를 사용하여 복수의 상황이 미리 학습된 신경 네트워크를 이용하여 복수의 상황을 인식할 수 있다. 구체적으로 상황인식부(250)는 특징 파라미터를 소정 크기의 입력 행렬로 변환한다. 그런 다음, 상황인식부(250)는 변환된 입력 행렬을 신경 네트워크에 입력하여 출력값을 구하고, 산출된 출력값에 따라 상황을 인식한다.

[0031] 상황학습부(260)는 특징 파라미터를 학습 데이터로 사용하여 신경 네트워크가 복수의 상황을 학습하도록 한다. 신경 네트워크는 복수의 연산을 포함하는 복수의 계층을 포함하며, 복수의 계층 중 어느 하나의 계층에 포함된 복수의 연산으로부터의 복수의 출력 각각은 가중치(w)가 적용되어 다음 계층의 입력이 된다. 상황학습부(260)는 각 상황에 따라 학습 데이터(특징 파라미터)를 신경 네트워크에 입력하는 것을 가정할 때 목표값을 미리 설정하고, 학습 데이터(특징 파라미터)를 신경 네트워크에 실제 입력 했을 때, 출력값과 미리 설정된 목표값을 비교하여, 출력값이 목표값이 되도록 가중치(w)를 조정한다. 이와 같이, 가중치(w)를 조정하는 과정을 '학습'이라고 칭한다. 상황학습부(260)는 학습에 따라 조정된 가중치(w)를 상황인식부(250)에 제공하여 상황인식부(250)가 인식을 수행하도록 한다.

[0032] 이러한 상황학습부(260)는 도시된 바와 같이, 본 발명의 실시예에 따른 사용자 장치(200)의 구성이고, 특징추출부(240)가 마이크(110), 통신부(210), 위치정보수신부(220), 센서부(230)로부터 추출한 특징 파라미터를 이용하여 학습을 수행하는 것이 바람직하다. 하지만, 상황학습부(260)는 예컨대, 랩탑 컴퓨터 등과 같은 사용자 장치(200)가 아닌 다른 장치에서 구현되고, 사용자 장치(200)와 동일한 특징 파라미터를 이용하여 학습을 수행할 수도 있다. 전술한 상황인식부(250) 및 상황학습부(260)의 동작은 아래에서 더 상세하게 설명될 것이다.

[0033] 출력제어부(270)는 상황인식부(250)의 인식 결과에 따라 이어폰 장치(100)의 스피커(120)의 출력을 제어한다. 일례로, 사용자는 음악을 청취하고 있는 상황이라고 가정한다. 이때, 상황 인식 결과, 실외인 경우, 주변 환경 소음을 듣지 못해서 위험한 상황이 발생할 수 있는 개연성이 실내 보다 높다. 이에 따라, 출력제어부(270)는 마이크(110)를 통해 입력되는 주변 환경 소음 등을 포함하는 사운드를 재생되는 음악의 사운드와 함께 출력할 수 있다. 다른 예에 따르면, 사용자는 마찬가지로 음악을 청취하고 있는 상황이라고 가정한다. 이때, 상황 인식 결과, 실내인 경우, 스피커(120)로부터 출력되는 음악의 사운드는 이어폰임에도 불구하고 주변 사람들에게 실례가 될 수 있다. 따라서 실내의 경우, 출력제어부(270)는 마이크(110)로 입력되는 주변 환경 소음 등의 사운드 신호의 크기를 측정하여 크기가 낮을수록 스피커(120)를 통해 출력되는 음악 사운드의 볼륨을 낮게 조절한다. 또 다른 예에 따르면, 사용자는 마찬가지로 음악을 청취하고 있는 상황이라고 가정한다. 이때, 혹은, 상황인식 결과, 위급 상황인 경우, 출력제어부(270)는 음악을 끄고, 위급 상황임을 알리는 음성, 경고음 등을 스피커(120)를 통해 출력한다.

[0034] 본 발명은 마이크(110)로부터의 사운드 신호, 통신부(210)로부터의 통신 신호, 위치정보수신부(220)로부터의 GPS 신호 그리고 센서부(230)로부터의 센서 신호 중 적어도 하나를 포함하는 특징 신호를 수집하고, 이로부터 특징 파라미터를 추출한다. 이러한 특징 신호 및 특징 파라미터에 대해서 상세하게 설명하기로 한다.

[0035] 1. 사운드 신호

[0036] 도 3 내지 도 8은 본 발명의 실시예에 따른 마이크로부터 출력되는 사운드 신호로부터 추출되는 특징 파라미터를 설명하기 위한 도면이다. 특징추출부(240)는 마이크(110)로부터 사운드 신호가 입력되면, 소정 길이의 프레임 단위로, 시간 도메인에서 FTE(Frame Time Energy) 및 ZCR(Zero-Crossing Rate)을 특징 파라미터로 추출한다. 또한, 특징추출부(240)는 사운드 신호를 푸리에 변환하여 주파수 도메인으로 변환한 후, 주파수 도메인에서 MFCC(Mel-Frequency Cepstral Coefficient) 및 SC(Spectral Centroid)를 특징 파라미터로 추출한다. 이러한 사운드 신호의 각 특징 파라미터에 대해서 설명하기로 한다.

[0037] 먼저, STE(Short-Time Energy)에 대해서 설명하기로 한다. STE는 사운드 신호에서 구간별 에너지를 측정하여 표현하는 방법이다.

[0038] 다음의 수학식 1은 사운드 신호에서 에너지를 구하기 위한 것이다.

수학식 1

$$E = \frac{1}{N}\sum_{m} x(m)^2$$

[0039]

[0040] 수학식 1의 경우 사운드 신호 전체에 대해 제곱하므로, 시간에 따른 에너지변화를 알 수 없다. 따라서 다음의 수학식 2와 같이 윈도우 함수를 곱하여 일정 시간동안의 에너지를 산출한다. 즉, 수학식 2에 따라 STE(Short-Time Energy)가 산출된다.

수학식 2

$$E_n = \frac{1}{N}\sum_{m} [x(m)w(n-m)]^2$$

[0041]

[0042] 여기서, $x(m)$은 입력되는 사운드 신호, n은 STE에서의 위치, $w(m)$은 윈도우의 길이 N을 뜻한다. 이와 같이, STE(Short-Time Energy)는 신호의 흐름에 따른 지속적인 변화량을 알기 위해서 윈도우 함수를 사용하여 처리된다. 하지만, 본 발명의 실시예에 따르면, 딥 러닝, 바람직하게는, CNN에 따른 학습을 위하여, 전체 사운드 신호를 중복되는 구간 없이 소정 길이의 프레임 단위로 나누고, 소정 길이의 프레임 동안의 에너지를 나타내는 FTE(Frame Time Energy)를 구하여 사용한다. 윈도우 연속적인 표현력이 낮아지지만, 전체 연산량이 현저히 낮아지기 때문에 연산 속도가 향상되는 이점이 있다. 도 3에서 입력디는 사운드 신호와 STE 및 200ms의 프레임 단위로 표시한 FTE가 도시되었다. 여기서, 에너지값은 식별을 쉽게 하도록 단위를 낮추어 표현할 수 있다.

[0043] ZCR(Zero-Crossing Rate)는 사운드 신호에서 부호가 바뀌는 비율을 나타낸다. 다시 말하면, ZCR(Zero-Crossing Rate)은 사운드 신호가 진폭축의 0을 지나는 빈도를 나타낸다. 도 4에 사운드 신호에서 STZCR(Short Time Zero-Crossing Rate)을 표현한 그래프가 도시되었다. 사운드 신호의 크기는 신경 쓰지 않고 부호의 변화만 보기 때문에 ZCR은 다음의 수학식 3을 통해 구해진다.

수학식 3

$$zcr = \frac{1}{T-1}\sum_{t=1}^{T-1} \Pi\{s_t s_{t-1} < 0\}$$

[0044]

[0045] 여기서, s는 사운드 신호를 뜻하며, T는 사운드 신호의 길이를 나타낸다. 또한, $\Pi\{A\}$는 A가 참일 경우 1을 반환하며, 거짓일 경우 0을 반환하는 함수이다. STE의 경우와 마찬가지로, CNN에 따른 학습을 위하여, 입력되는 전체 사운드 신호를 중복되는 구간 없이 소정 길이의 프레임으로 나누고, 소정 길이의 프레임 기간 동안 사운드 신호가 진폭축의 0을 지나는 빈도를 나타내는 ZCR을 산출하여 이용한다.

[0046] MFCC(Mel-Frequency cepstral coefficients)는 복수의 MFC(Mel-Frequency cepstrum)들을 모아서 만들어 진다. MFC는 단구간 신호의 파워스펙트럼을 표현하는 방법 중 하나로서, 비선형 멜 스케일(Mel-scale)을 사용해 주파수로부터 얻은 파워스펙트럼을 로그 값을 취하고 코사인 변환을 통해 나타낸 값이다. 여기서, 캡스트럼(Cepstrum)은 신호를 푸리에변환(Fourier Transform)하여 스펙트럼(Spectrum)으로 변환한 후, 로그 연산을 취한 후, 다시 역푸리에변환(Inverse Fourier Transform)을 취한 값을 일컫는다. 일반적인 캡스트럼과 MFCC 캡스트럼의 다른 점은, 주파수 밴드가 일정하게 나누어져 있는 일반적인 캡스트럼과 달리, MFCC의 경우 멜 스케일(Mel-scale)에서 주파수 밴드가 일정하게 나누어진다. 여기서, 멜 스케일(Mel-scale)은 사람의 청각에 관련하여 주파수를 새로 변환한 방식이다. 사람의 청각에서 주파수 특성은 1000Hz보다 낮은 주파수에서는 선형적인 형태를 나타내고, 1,000Hz보다 높은 주파수에서는 로그 스케일의 형태로 나타난다. 이러한 특성을 기반으로 임계주

파수인 1,000Hz를 1,000Mel에 맞추어 변형하여 다음의 수학식 4와 같이 표현된다.

수학식 4

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

[0047]

[0048] 이와 같이 표현되는 멜-스케일(Mel-scale) 식을 사용하여 MFCC를 구하는 과정이 도 5에 도시되었다.

[0049] MFCC를 구하기 위해, 먼저, 특징추출부(240)는 입력되는 사운드 신호를 시간 도메인에서 중복되는 구간 없이 소정 길이의 프레임으로 구분하고, 소정 길이 프레임 단위의 사운드 신호를 주파수 도메인으로 변환한다. 이때, 특징추출부(240)는 고속푸리에변환(FFT: Fast Fourier Transform)을 이용하여 소정 길이 프레임 단위의 사운드 신호를 주파수 도메인으로 변환한다.

[0050] 두 번째로, 특징추출부(240)는 주파수 도메인으로 변환된 사운드 신호를 멜-스케일을 사용하는 멜 필터뱅크에 통과시켜 멜-파워를 구한다. 도 5의 멜 필터뱅크의 그래프(50)를 살펴보면, x축은 주파수를 나타내며, y축은 낮은 주파수부터 점점 높은 주파수의 멜-파워(Mel-power)를 숫자로 나타낸 것이다. 그래프에 나타난 것과 같이 주파수가 낮은 곳에서의 필터는 좁은 주파수 범위만을 포함하지만, 주파수가 높은 곳에서는 주파수 범위도 넓어진다.

[0051] 도 6에 사운드 신호를 주파수 도메인으로 변환하고, 멜 필터뱅크를 이용한 연산하는 과정을 설명하기 위한 그래프가 도시되었다. 도 6의 (a)는 입력되는 사운드 신호 전체에서 소정 크기의 단 구간 신호, 즉, 소정 길이의 프레임 단위로 구분한 신호를 나타낸다. 도 6의 (b)는 소정 크기의 단 구간 신호를 FFT(Fast Fourier Transform)을 통하여 연산한 주파수별 세기이다. FFT 연산에 따라 출력되는 값은 허수부(Im)와 실수부(Re)로 구분되기 때문에 절대 값을 취하여 주파수별 세기를 표현할 수 있다. 도 6의 (c)는 멜 필터뱅크(Mel-filterbank)의 복수의 필터 중 4번째 filter만을 표현한 것이다. 마지막으로 도 6의 (d)는 (b)의 주파수별 세기와 (c)의 필터를 곱연산의 결과이다. 실제 멜 파워(Mel-power)로 표현 할 경우, (d)에서 표현된 값들을 모두 더하면 해당 필터를 통한 멜 파워(Mel-power)가 산출된다.

[0052] 세 번째로, 멜-필터뱅크(Mel-filterbank)를 통해 구한 멜-파워(Mel-power)에 로그(log) 연산을 수행한다.

[0053] 네 번째로, 멜 파워에 로그 연산을 수행한 신호에 대해 DCT(Discrete Cosine Transform) 연산을 수행하여 MFCC(Mel-Frequency cepstral coefficients)를 구한다. DCT 연산은 추가적인 정보 없이 주어진 값만을 가지고 주파수 영역으로 나타내어 압축하는 방법이다. 이는 로그 연산된 멜-파워(Mel-power)로부터 특징값을 압축하여 나타내는데 사용된다. 도 7의 그래프는 20개의 로그 연산을 한 멜 파워를 DCT 연산의 입력으로 하였을 때, 그 DCT 연산의 13개의 값을 보인다.

[0054] SC(Spectral Centroid)는 스펙트럼의 무게중심을 나타내는 특징 파라미터이다. 이는 일반적으로 소리의 밝기를 나타낸다. 푸리에 변환(Fourier transform)으로부터 계산된 주파수 영역에서 무게 중심을 나타낸다. 도 8에 사운드 신호와 주파수 영역에서의 SC를 같이 표시한 그래프가 도시되었다. SC는 사운드 신호의 주파수별 진폭을 사용하여 다음의 수학식 5에 따라 산출될 수 있다.

수학식 5

$$SpectralCentroid = \frac{\sum_{n=1}^{N} f(n)x(n)}{\sum_{n=1}^{N} x(n)}$$

[0055]

[0056] x(n)은 주파수 영역에서 주파수 값을 나타내며, f(n)의 해당 주파수의 진폭을 나타낸다. 이러한 연산에 따라 각

주파수의 진폭을 참조하여 중심 주파수의 위치를 알려준다. 본 발명에 따르면, SC(Spectral Centroid)를 구하기 위해, 특징추출부(240)는 입력되는 사운드 신호를 시간 도메인에서 중복되는 구간 없이 소정 길이의 프레임으로 구분하고, 소정 길이 프레임 단위의 사운드 신호를 고속푸리에변환(FFT: Fast Fourier Transform)을 이용하여 주파수 도메인으로 변환한다. 그런 다음, 특징추출부(240)는 각 주파수의 진폭을 구한 후, 중심 주파수의 위치를 나타내는 SC를 구한다.

[0057] 2. 통신 신호

[0058] 본 발명의 실시예에서 통신 신호는 통신부(210)가 AP로부터 수신되는 무선주파수 신호를 의미한다. 통신부(210)는 주변의 AP의 수 및 그 AP로부터 수신되는 신호 세기를 측정할 수 있다. 이에 따라, 특징추출부(240)는 통신부(210)로부터 소정 프레임 단위로 검색되는 AP의 수 및 그 AP로부터 수신되는 신호 세기를 수신하여, 검색된 AP의 수 및 검색된 모든 AP로부터 수신되는 신호세기의 평균을 특징 파라미터로 추출한다.

[0059] 3. GPS 신호

[0060] 도 9는 본 발명의 실시예에 따른 GPS 신호로부터 추출되는 특징 파라미터를 설명하기 위한 도면이다. GPS 신호는 위성을 통해 정보를 수신 받기 때문에 다양한 정보가 제공된다. GPS를 통해 수신 받는 정보는 형식에 따라서 다르게 표현될 수 있다. 예컨대, 형식의 종류는 GPGGA(Global Positioning System Fixed Data.)와 GPRMC (Recommended Minimum Specific GNSS data), GPGSA(GNSS DOP and active satellites), GPGSV (GPS Satellites in View) 등이 있다. 이는 NMEA(The National Marine Electronics Association)라는 규격을 통해 문장으로 표현이 된다. NMEA란 시간, 위치 등의 정보 송수신을 위한 규격으로서, 자이로센서, GPS, 나침반 등에서 통신을 사용하여 데이터 전송이 발생할 경우 주로 사용된다.

[0061] 각 형식별로 표현되는 데이터가 다른데, GPGGA의 경우 시간과 위도, 경도, 고도 등의 정보를 볼 수 있으며, GPGSV는 위치정보수신부(220)에서 수신 할 수 있는 모든 위성의 정보를 표시하며, GPRMC는 최소한의 정보만을 제공한다.

[0062] 도 9에 NMEA에 따라 표현된 정보가 도시되었다. 도면 부호 91은 표현 형식을 나타내며, 도면 부호 93은 그 표현 형식(91)에 따른 정보를 나타낸다. 도면부호 95는 GPGSV 형식으로 제공되는 정보를 나타낸다. 첫 번째, GPGSV 문장은 '3, 1, 12, 02, 64, 331, 38, 04, 61, 070, 35, 11, 285, 29, 10, 41, 249, 35*7C'라고 기재되어 있으며, 이에 대해 설명하면 다음과 같다. 가장 처음 기재된 '3'은 GPGSV가 전체 몇 개의 문장(Sentence)인지를 나타낸다. 이를 통해 총 3개의 문장임을 알 수 있다. '3' 다음에 기재된 '1'은 GPGSV 문장 중 몇 번째의 문장인지를 나타낸다. 이를 통해 첫 번째 문장임을 알 수 있다. '3' 다음의 '12'는 현재 수신할 수 있는 모든 위성의 개수를 나타낸다. '12' 다음에 기재된 '02, 64, 331, 38'은 2번 위성이고, 사용자 장치(100)의 위치에서 64도의 상향각에 위치하며, 사용자 장치(100)의 위치에서 331도의 방위각(Azimuth)에 위치하고, 신호대잡음비(SNR: Signal to Noise Ratio)는 38임을 나타낸다. 또한, '7C'는 체크섬(Checksum)이다. $'와 '*'사이의 모든 데이터를 'exclusive or' 연산하여 체크섬을 만든다. 하나의 GPGSV 문장에는 4개의 위성 정보가 기술될 수 있다.

[0063] 전술한 GPGSV와 비교하였을 때 GPGGA를 이용하는 경우, 현재 위치 정보를 그대로 나타내어 직관적이다. 하지만, 이러한 정보는 정보의 적어도 일부라도 수신되지 않는 경우, 문제가 발생할 수 있다. 통상적으로 GPGGA는 한 문장이 전송되며, GPGSV는 두 문장 이상이 전송된다. 더욱이, 따라서 총 위성의 수는 2번 이상 전송되며, 총 위성의 수를 추정할 수 있는 GPGSV 문장의 총 수도 2번 이상 전송된다. 이에 따라, 총 위성의 수와 관련된 부분을 수신하지 못한 경우, GPGSV 문장의 총 수에 따라 문장에 4개로 총 위성의 수를 추정할 수 있다. 이러한 이유로 본 발명에 따르면, GPGSV에서 표현되는 정보인 검색된 위성의 개수와 각 위성으로부터의 신호 세기만을 사용한다. 위성의 개수는 그대로 사용하며, 위성의 신호세기는 수신된 위성의 신호세기를 평균하여 사용한다. 즉, 특징추출부(240)는 GPS 신호로부터 검색되는 위성의 개수를 획득하고, 소정 길이 프레임 동안 검색되는 위성으로부터 수신되는 신호세기의 평균을 구한 후, 위성의 개수 및 신호세기의 평균을 특징 파라미터로 추출한다.

[0064] 4. 센서 신호

[0065] 도 10은 본 발명의 실시예에 따른 센서 신호로부터 추출되는 특징 파라미터를 설명하기 위한 도면이다. 센서부(230)는 가속도 센서를 포함한다. 본 발명의 실시예에서 가속도 센서는 3축, X축, Y축 및 Z축의 가속도를 감지하는 3축 가속도 센서이다. 이러한 가속도 센서는 3축의 가속도를 감지하여 센서 신호를 출력하며, 가속도 센서가 출력한 센서 신호를 가속도 데이터라고 칭하기로 한다. 가속도 데이터는 3개의 축을 사용하여 표현되며, 특징 파라미터로 사용된다.

[0066] 도 10에 사용자 장치(200)의 각 축의 방향(X축 방향, Y축 방향 및 Z축 방향)이 도시되었다. 도 10과 같이, 사용

자 장치(200)가 정지된 상태에서 센서부(230)의 가속도 센서는 Z축 방향의 중력 가속도만을 출력하고, X축 방향 및 Y축 방향의 출력은 0이다. 그리고 사용자 장치(200)가 외력에 의해 움직이는 경우, 가속도 센서는 3축의 가속도를 감지하여 센서 신호를 출력한다. 센서 신호의 각 축의 값은 0을 기준으로 음수, 양수 값이 존재한다. 가속도 데이터는 센서 신호의 출력되는 값 그대로 사용된다.

[0067] 본 발명의 실시예는 전술한 바와 같이 구해진 특징 파라미터를 정규화하고, 정규화된 특징 파라미터를 신경 네트워크의 입력으로 하여 상황을 학습한다. 그리고 이러한 학습에 따라 특징 파라미터가 신경 네트워크에 입력되면, 그 상황을 인식할 수 있다. 그러면, 본 발명의 실시예에 따른 학습 및 인식을 위한 신경 네트워크에 대해서 설명하기로 한다. 도 11 내지 도 14는 본 발명의 실시예에 따른 신경 네트워크를 설명하기 위한 도면이다. 도 11을 참조하면, 본 발명의 실시예에 따른 상황인식부(250)는 도시된 바와 같은 신경 네트워크를 포함한다.

[0068] 특징 추출부(240)는 전술한 바와 같이, 마이크(110), 통신부(210), 위치정보수신부(220) 및 센서부(230)를 통해 특징 파라미터를 추출하여, 상황인식부(250)에 입력할 수 있다. 상황인식부(250)는 특징 추출부(240)로부터 수신되는 특징 파라미터를 기 설정된 크기를 가지는 행렬로 변환하여 신경 네트워크의 입력으로 사용한다. 예컨대, 도시된 바와 같이, 특징 파라미터는 32 × 32 크기의 입력 행렬로 변환되어 신경 네트워크의 입력으로 사용된다.

[0069] 본 발명의 실시예에 따른 신경 네트워크는 1개의 입력과 3개의 출력을 가진다. 1개의 입력은 특징 파라미터로 이루어진 입력 행렬을 통해 이루어지며, 3개의 출력은 3개의 노드(혹은 sigmoid: N1, N2, N3)를 통해 이루어진다. 3개의 출력 노드(N1, N2, N3)의 출력은 상황을 나타내며, 제1 노드(N1)의 출력은 현재 위치가 실내일 확률, 제2 노드(N2)의 출력은 현재 위치가 실외일 확률, 제3 노드(N1)의 출력은 현재 위험 상황일 확률을 나타낼 수 있다.

[0070] 1개의 입력 행렬은 특징 추출 네트워크(251) 및 분류 네트워크(253)를 거쳐 3개의 출력이 된다. 특징 추출 네트워크(251)는 컨볼루션 계층(convolution layer) 및 풀링 계층(pooling layer)이 교대로 반복하여 구성된다. 이 실시예에서 특징 추출 네트워크(251)는 제1 컨볼루션 계층(C1), 제1 풀링 계층(S1), 제2 컨볼루션 계층(C2) 및 제2 풀링 계층(S2)를 포함한다. 분류 네트워크(253)는 적어도 하나의 완전연결 계층(Full Connection Layer: F1)으로 이루어진다. 이 실시예서 분류 네트워크(253)는 하나의 완전연결 계층(F1)으로 이루어진다.

[0071] 컨볼루션 계층(C1, C2)과 풀링 계층(S1, S2) 각각은 복수의 특징 지도(Feature Map)로 이루어지며, 이러한 특징 지도 각각은 소정 크기의 행렬이다. 특징 지도를 이루는 행렬의 각 성분은 소정 크기의 행렬인 커널(혹은 필터)을 이용한 컨볼루션 연산(convolution) 혹은 풀링 연산(pooling 혹은 subsampling)을 통해 산출된다. 즉, 컨볼루션 계층(C1, C2)의 특징 지도를 이루는 행렬의 각 성분은 소정 크기의 행렬인 커널(혹은 필터)을 이용한 컨볼루션 연산(convolution)을 통해 산출되며, 즉, 풀링 계층(S1, S2)의 특징 지도를 이루는 행렬의 각 성분은 소정 크기의 행렬인 커널(혹은 필터)을 이용한 풀링 연산(pooling 혹은 subsampling)을 통해 산출된다.

[0072] 도 12를 참조로 컨벌루션 연산에 대해서 설명한다. 도시된 바와 같이, 입력이 7×7 크기의 행렬이며, 커널은 3×3 크기의 행렬이라고 가정한다. 커널을 이루는 행렬의 각 성분은 가중치(w)를 가질 수 있다. 커널을 이루는 행렬의 각 성분에 대한 가중치는 모두 1이라고 가정한다. 입력은 입력 행렬 혹은 이전 계층의 특징 지도가 될 수 있다. 7×7 크기의 행렬인 입력과 3×3 크기의 행렬인 커널을 컨볼루션하면 5×5 크기의 특징 지도를 생성할 수 있다. 즉, 커널의 크기가 3×3이기 때문에 입력의 (1, 1) 성분에 대한 컨볼루션 연산은 수행할 수 없다. 컨볼루션 연산은 커널이 한칸의 행 혹은 열 만큼 시프트하며, 대응하는 성분에 대해 연산한다. 이에 따라, 특징 지도에서 입력 행렬의 가장 자리의 성분은 버려진다. 이러한 컨볼루션 연산은 대응하는 성분의 곱과 그 곱들의 합으로 이루어진다. 일례로, 입력(7×7 크기의 행렬)의 (2, 2) 성분에 대한 커널(3×3 크기의 행렬)을 컨벌루션 연산하면, (0×4)(×1) + (0×0)(×1) + (0×0)(×1) + (0×0)(×1) + (1×0)(×1) + (1×0)(×1) + (0×0)(×1) + (1×0)(×1) + (2×(-4))(×1) = -8이 된다. 이에 따라, 5×5 크기의 특징 지도의 (1, 1) 성분은 -8이 된다. 다른 예로, 입력(7×7 크기의 행렬)의 (3, 4) 성분에 대한 커널(3×3 크기의 행렬)을 컨벌루션 하면, (1×4)(×1) + (1×0)(×1) + (1×0)(×1) + (2×0)(×1) + (2×0)(×1) + (1×0)(×1) + (2×0)(×1) + (2×0)(×1) + (2×(-4))(×1) = -4가 된다. 이에 따라, 5×5 크기의 특징 지도의 (2, 3) 성분은 -4가 된다. 앞선 예에서 (×1)은 가중치를 곱한 것이다.

[0073] 도 13을 참조로 풀링 연산 혹은 서브 샘플링 연산에 대해서 설명한다. 도시된 바와 같이, 입력은 8×8 크기의 행렬이며, 커널은 2×2 크기의 행렬이라고 가정한다. 또한, 각 커널을 이루는 행렬의 각 성분은 가중치 1을 가진다고 가정한다.

[0074] 입력은 이전 계층의 특징 지도가 될 수 있다. 8×8 크기의 행렬인 입력을 2×2 크기의 행렬인 커널을 이용하여 서브샘플링(혹은 풀링)하면, 4×4 크기의 특징 지도가 생성된다. 풀링 연산은 커널이 커널의 크기 만큼 시프트하여 연산을 수행한다. 이에 따라, 출력되는 특징 지도는 입력 크기에서 커널의 크기에 따라 줄어든다. 예컨대, 첫 번째 연산은 8×8 크기 행렬의 (1, 1), (1, 2), (2, 1), (2, 2) 성분과 2×2 크기 행렬인 커널과의 연산을 통해 이루어지며, 이 연산의 결과는 4×4의 특징 지도의 (1, 1) 성분의 값이 된다. 두 번째 연산은 8×8 크기 행렬의 (1, 3), (1, 4), (2, 3), (2, 4) 성분과 2×2 크기 행렬인 커널과의 연산을 통해 이루어지며, 이 연산의 결과는 4×4의 특징 지도의 (1, 2) 성분의 값이 된다. 이와 같이, 커널이 2×2 크기에 따라 2행 혹은 2열씩 시프트되어 풀링(pooling 혹은 subsampling) 연산이 수행되기 때문에 8×8 크기의 입력은 2×2 크기의 커널에 따라 줄어들어 4×4 크기의 특징 지도가 된다. 본 발명의 실시예에 따른 커널의 풀링 연산은 최대값을 선택하는 연산이며, 이에 따라, 각 연산에 사용되는 4개의 성분의 값 중 최대값이 선택된다. 예컨대, 입력의 (1, 1), (1, 2), (2, 1) 및 (2, 2) 성분을 커널을 통해 풀링 연산하면, (1, 1), (1, 2), (2, 1) 및 (2, 2) 성분 각각의 값은 각각 0 (×1) = 0, 0 (×1) = 0, 0 (×1) = 0, 0 (×1) = 0이기 때문에 최대값 Max(0, 0, 0, 0) = 0이 선택된다(여기서, (×1)은 가중치). 다른 예로, 입력의 (1, 5), (1, 6), (2, 5) 및 (2, 6) 성분을 커널을 통해 풀링 연산하면, (1, 5), (1, 6), (2, 5) 및 (2, 6) 성분 각각의 값은 1 (×1) = 1, 1 (×1) = 1, 0 (×1) = 0, 0 (×1) = 0이기 때문에 최대값 Max(1, 1, 0, 0) = 1이 선택된다(여기서, (×1)은 가중치).

[0075] 도 12의 예에서, 하나의 커널은 그 커널을 이루는 행렬의 크기에 상응하는 가중치를 가지는 것으로 설명하였다. 하지만, 본 발명의 대안적인 실시예에 따르면, 학습 및 인식을 위한 연산량을 줄이고, 속도를 증가시키기 위하여 소정 크기의 행렬인 커널에 적용되는 가중치를 해당 커널의 행렬의 크기에 무관하게 1개로 설정한다. 즉, 도 12에서 3×3 행렬인 커널은 3×3개의 가중치를 가지지만, 본 발명의 대안적인 실시예에 따르면, 이러한 가중치는 1개로 설정할 수 있다. 이와 같이, 1개의 커널이 1개의 가중치를 가질 때, 가중치는 1이라고 가정하면, 연산은 다음과 같다. 입력(7×7 크기의 행렬)의 (2, 2) 성분에 대한 커널(3×3 크기의 행렬)을 컨벌루션 연산하면, (0×4) + (0×0) + (0×0) + (0×0) + (1×0) + (1×0) + (0×0) + (1×0) + (2×(-4)) = -8 (×1) = -8이 된다. 여기서, (×1)은 가중치를 곱한 것이다. 이전 예에 따르면, "(0×4)(×1) + (0×0)(×1) + (0×0)(×1) + (0×0)(×1) + (1×0)(×1) + (1×0)(×1) + (0×0)(×1) + (1×0)(×1) + (2×(-4))(×1)"와 같이, 가중치를 곱하는 연산 '(×1)'이 9번 이루어지지만, 본 발명의 대안적인 실시예는 1번만 수행하면 된다.

[0076] 앞서 설명된 바와 마찬가지로, 본 발명의 대안적인 실시예에 따르면, 학습 및 인식을 위한 연산량을 줄이고, 속도를 증가시키기 위하여 소정 크기의 행렬인 커널에 적용되는 가중치를 해당 커널의 행렬의 크기에 무관하게 1개로 설정할 수 있다. 예컨대, 도 13에서 2×2 행렬인 커널은 2×2개의 가중치를 가지지만, 본 발명의 대안적인 실시예에 따르면, 이러한 가중치는 1개로 설정할 수 있다. 이와 같이, 1개의 커널이 1개의 가중치를 가질 때, 가중치는 1이라고 가정하면, 연산은 다음과 같다. 예컨대, 입력의 (1, 5), (1, 6), (2, 5) 및 (2, 6) 성분을 커널을 통해 풀링 연산하면, (1, 5), (1, 6), (2, 5) 및 (2, 6) 성분 각각의 값은 1, 1, 0, 0이며, 최대값 Max(1, 1, 0, 0) = 1이 선택되고, 가중치를 적용하여 1 (×1) = 1이 된다. 이전 예에 따르면, (1, 5), (1, 6), (2, 5) 및 (2, 6) 성분 각각의 값은 1 (×1) = 1, 1 (×1) = 1, 0 (×1) = 0, 0 (×1) = 0이고, 이중 최대값은 Max(1, 1, 0, 0) = 1이 된다. 보인 바와 같이, 가중치를 곱하는 연산 '(×1)'이 4번 이루어지지만, 본 발명의 대안적인 실시예는 1번만 수행하면 된다.

[0077] 전술한 바와 같이, 1개의 커널에 1개의 가중치를 할당하는 경우, 도 11에 보인 신경 네트워크에서 입력 행렬(IM)과 제1 컨볼루션 계층(C1) 사이의 가중치 연산은 5 × 5 × 6번에서 1 × 6번으로 줄어들고, 제1 컨볼루션 계층(C1)과 제1 풀링 계층(S1) 사이의 가중치 연산은 2 × 2번에서 1번으로 줄어들며, 제1 풀링 계층(S1)과 제2 컨볼루션 계층(C2) 사이의 가중치 연산은 6 × 5 × 5 × 16번에서 6 × 1 × 16번으로 줄어들고, 제2 컨볼루션 계층(C2)과 제2 풀링 계층(S2) 사이의 가중치 연산은 2 × 2번에서 1번으로 줄어들고, 제2 풀링 계층(S2)과 완전연결 계층(F1) 사이의 가중치 연산은 16 × 5 × 5 × 120번에서 16 × 1 × 120번으로 줄어든다. 이와 같이, 1개의 커널에 1개의 가중치를 할당하는 경우, 가중치 연산의 연산량이 줄어들고, 연산 속도가 향상된다. 더욱이, 가중치의 수가 줄어들었기 때문에 학습에 따른 연산 복잡도 또한 줄어든다. 이에 따라, 학습 및 인식을 위한 연산량이 줄어들고 연산 속도가 향상된다.

[0078] 전술한 컨볼루션 연산 및 풀링 연산을 수행하면, 특징 지도의 수는 증가하고, 성분의 수는 감소할 수 있다. 이에 따라, 모든 특징 지도가 하나의 성분만을 가질 수 있다. 이러한 상태를 완전 연결(full connection) 상태라고 하며, 완전연결 계층(F1)은 완전 연결 상태인 계층을 의미한다. 본 발명의 실시예에서 완전 연결 계층은 하나인 것으로 설명되지만, 복수의 완전연결 계층이 있을 수 있다.

[0079] 완전연결 계층은 복수의 노드(혹은 sigmoid): X001 내지 X120)로 이루어진다. 각 노드는 이전 계층의 출력을 입

력으로 받아들여 연산을 수행하고, 가중치를 적용한 후, 다음 계층으로 출력한다. 예컨대, 도 14를 참조로 어느 하나의 퍼셉트론(P)의 3개의 입력에 따른 하나의 출력에 대해서 설명한다. 3개의 입력 in1, in2, in3에 가중치 w1, w2, 및 w3을 곱하고, 모두 합산한 후, 합산된 값을 전달 함수에 대입하여 출력을 산출한다. 예컨대, 입력 in1, in2 및 in3은 각각 0.5, -0.3 및 0이고, 가중치 w1, w2 및 w3은 각각 4, 5 및 2라고 가정한다. 한편, 본 발명의 신경 네트워크에서 전달함수는 다음의 수학식 6과 같은 시그모이드(sigmoid) 함수를 이용하는 것이 바람직하다.

수학식 6

[0080]
$$\sigma(z) \equiv \frac{1}{1+e^{-z}}$$

[0081] 하지만, 설명의 편의를 위하여 전달함수는 'sgn()'이라고 가정하면, 다음과 같이 출력 값이 산출된다.

[0082] in1 × w1 = 0.5 × 4 = 2

[0083] in2 × w2 = - 0.3 × 5 = -1.5

[0084] in3 × w3 = 0 × 2 = 0

[0085] 2 + (-1.5) + 0 = 0.5

[0086] sgn(0.5) = 1

[0087] 이와 같이, 완전연결 계층은 이전 계층의 입력을 이용하여 가중치를 이용한 연산을 수행한 후, 그 출력을 다음 계층으로 전달한다. 이러한 전달 과정을 적어도 한번 거친 후, 최종 출력을 얻을 수 있다. 본 발명의 실시예에서 완전 연결 계층(F1)은 1개이기 때문에 한 번의 전달 과정을 거쳐 3개의 출력을 얻는다.

[0088] 한편, 도 11을 다시 참조하여 본 발명에 따른 신경 네트워크의 연산 절차에 따른 상황 인식 방법에 대해서 설명하기로 한다.

[0089] 입력 행렬은 32×32 크기의 행렬이며, 커널은 5×5 크기의 행렬을 이용한다. 전술한 바와 같이, 특징 파라미터는 32 × 32 크기의 입력 행렬로 변환되어 신경 네트워크의 입력으로 사용된다.

[0090] 입력 행렬에 대해 커널을 이용한 컨벌루션 연산(convolution)을 수행하여, 각각이 28×28 크기의 행렬인 6개의 특징 지도로 이루어진 제1 컨벌루션 계층(C1)을 구성한다.

[0091] 그런 다음, 제1 컨벌루션 계층(C1)의 복수의 특징 지도에 대해 2×2 크기의 행렬인 커널을 이용한 풀링 연산(subsampling)을 수행하여 각각이 14×14 크기의 행렬인 6개의 특징 지도로 이루어진 제1 폴링 계층(S1)을 구성한다.

[0092] 이어서, 제1 폴링 계층(S1)의 복수의 특징 지도에 대해 5×5 크기의 행렬인 커널을 이용한 컨벌루션 연산(convolution)을 수행하여, 각각이 10×10 크기의 행렬인 16개의 특징 지도로 이루어진 제2 컨벌루션 계층(C2)을 구성한다.

[0093] 다음으로, 제2 컨벌루션 계층(C2)의 복수의 특징 지도에 대해 2×2 크기의 행렬인 커널을 이용한 풀링 연산(subsampling)을 수행하여 각각이 5×5 크기의 행렬인 16개의 특징 지도로 이루어진 제2 폴링 계층(S2)을 구성한다.

[0094] 그런 다음, 제2 폴링 계층(S2)의 복수의 특징 지도에 대해 5×5 크기의 행렬인 커널을 이용한 컨벌루션 연산(convolution)을 수행하여, 120개의 노드(X001 내지 X120)로 이루어진 완전 연결 계층(F1)을 구성한다.

[0095] 완전 연결 계층(F1)의 복수의 노드의 출력은 출력 계층에 입력된다. 즉, 120개의 노드(X001 내지 X120)의 출력 각각은 가중치와 곱해져 출력 계층의 3개의 노드(IN, OU, DS)의 입력이 된다.

[0096] 이러한 출력 계층의 3개의 노드(IN, OU, DS)의 출력은 각각 실내(IN: indoor)인 경우, 실외(OU: outdoor)인 경우 그리고 위험 상황(DS: dangerous situation)인 경우의 확률을 나타낸다.

[0097] 따라서 상황인식부(250)는 신경 네트워크에 특징 파라미터가 행렬로 변환되어 입력되면 실내인 경우, 실외인 경

우 그리고 위험 상황인 경우의 확률이 출력되며, 가중 높은 확률이 출력되는 상황을 현재의 상황으로 인식한다. 즉, 본 발명에 따른 인식, 상황 인식은 입력 파라미터를 입력하였을 때, 신경 네트워크의 연산을 수행하여 출력되는 출력(확률)값을 통해 이루어진다. 이러한 인식은 도 11의 왼쪽에서 오른쪽 방향의 화살표와 같은 순서로 이루어진다.

[0098] 한편, 도 11을 참조로 본 발명의 실시예에 따른 학습 방법에 대해서 설명하기로 한다. 전술한 바와 같이, 신경 네트워크의 어느 하나의 계층의 출력은 다른 하나의 계층의 입력이 된다. 이때, 어느 하나의 계층의 출력은 가중치가 곱해진 후, 다음 계층의 입력이 될 수 있다. 예컨대, 전술한 바와 같이, 32×32 크기의 행렬인 입력 행렬에 대해 5×5 크기의 행렬인 커널을 이용한 컨벌루션 연산을 수행하여, 각각이 28×28 크기의 행렬인 6개의 특징 지도로 이루어진 제1 컨벌루션 계층(C1)을 구성한다. 이때, 커널의 크기 5×5, 입력 행렬의 수 1 그리고 특징 지도의 수 6을 고려하면, 1 × 5 × 5 × 6개의 가중치가 존재한다. 다른 예로, 16개의 특징 지도를 가지는 제2 풀링 계층(S2)의 출력은 120개의 노드를 가지는 완전연결 계층(F1)의 입력이 된다. 제2 풀링 계층(S2)은 5×5 행렬로 이루어진 커널을 통해 컨볼루션 연산의 결과가 출력된다. 또한, 하나의 노드 당 16개의 입력이 있으며, 120개의 노드가 존재하기 때문에 5 × 16 × 120개의 가중치가 존재한다. 또 다른 예로, 120개의 노드를 가지는 완전연결 계층(F1)의 출력은 3개의 노드(IN, OU, DS)를 가지는 출력 계층에 입력된다. 이에 따라, 120 × 3개의 가중치가 존재한다.

[0099] 본 발명의 실시예에 따르면 특정 상황을 결정하고, 특정 상황에서 출력을 미리 결정하고, 결정된 상황에서 특징 파라미터를 추출하여 추출된 파라미터를 통해 출력된 출력값이 해당 상황에 상응하는 값이 되도록 신경 네트워크의 가중치를 수정하는 것을 학습이라고 한다. 예컨대, 사용자는 명확히 실외에 있다고 가정한다. 이때, 실외에서 측정된 특징 파라미터를 신경 네트워크에 입력했을 때, 그 출력에 따르면 실내인 확률이 가장 높게 출력되었다면, 인식 결과의 오류이다. 따라서 현재 상황인 실외의 경우인 확률이 가장 높게 출력되고 실내인 확률이 낮게 출력되도록 가중치를 수정하는 것을 학습이라고 한다. 이러한 학습은 도 11의 오른쪽에서 왼쪽 방향의 화살표와 같은 순서로 이루어진다. 즉, 출력으로부터 입력 방향으로 이루어지며, 이는 인식의 역순이 된다.

[0100] 전술한 본 발명의 실시예에 따른 학습 방법에 대해서 흐름도를 참조하여 설명하기로 한다. 도 15는 본 발명의 실시예에 따른 학습 방법을 설명하기 위한 흐름도이다.

[0101] 도 15에서 사용자는 사용자 장치(200)가 실내 환경임을 인식할 수 있도록 학습하기 위하여 사용자 장치(200)를 실내에 위치시킨 상태라고 가정한다. 그리고 사용자 장치(200) 현재의 목표값이 실내임을 입력한다. 이에 따라, 상황학습부(260)는 S110 단계에서 목표값을 설정한다. 이때, 실내인 상황을 인식하도록 학습하는 것이기 때문에 목표값이 실내 환경이 되도록 설정하였다고 가정한다. 즉, 도 11에 도시된 제1 노드(N1), 제2 노드(N2) 및 제3 노드(N3) 중 제1 노드(N1)의 출력이 가장 높은 값을 가지도록 제1 노드(N1), 제2 노드(N2) 및 제3 노드(N3)의 목표값을 설정하였다고 가정한다.

[0102] 다음으로, 특징추출부(240)는 S120 단계에서 미리 설정된 소정 길이의 프레임 단위로 마이크(110), 통신부(210), 위치정보수신부(220) 및 센서부(230)로부터 사운드 신호, 통신 신호, GPS 신호 및 센서 신호 중 적어도 하나를 포함하는 특징 신호를 수집하고, S130 단계에서 특징 신호로부터 특징 파라미터를 추출하여 상황인식부(250)에 입력한다.

[0103] 이어서, 상황인식부(250)는 S140 단계에서 특징 파라미터를 소정 크기의 입력 행렬로 변환한다. 그런 다음, 상황인식부(250)는 S150 단계에서 입력 행렬을 신경 네트워크에 입력하여 신경 네트워크의 가중치가 부여되는 연산에 따라 출력값을 산출하고, 산출된 출력값을 상황학습부(260)에 입력한다. 예컨대, 도 11에 보인 바와 같은 신경 네트워크에 입력 행렬(IM)을 입력으로 하면 제1 컨볼루션 계층(C1), 제1 풀링 계층(S1), 제1 컨볼루션 계층(C2), 제1 풀링 계층(S2), 완전 연결 계층(F1) 및 출력 계층(O1)에 의해 연산이 이루어진다. 특히, 신경 네트워크는 입력 행렬(IM) 및 복수의 계층(C1, S1, C2, S2, F1)의 출력에 가중치(w)가 곱해져서 다음 계층(C1, S1, C2, S2, F1, O1)의 입력이 된다. 이러한 연산에 따라 3개의 출력값(실내, 실외, 위급상황)을 얻을 수 있다.

[0104] 다음으로, 상황학습부(260)는 출력값이 입력되면, S160 단계에서 목표값과 비교하여 차이값을 산출한다. 그런 다음, 상황학습부(260)는 S170 단계에서 차이값을 출력으로부터 입력으로 역전파하여 출력값이 목표값이 되도록 가중치를 수정한다.

[0105] 이와 같이, 본 발명의 실시예에 따른 학습은 답(목표값)을 정해 놓고, 입력값에 따라 출력값이 목표값이 되도록 신경 네트워크의 연산에 사용되는 파라미터인 가중치를 수정하는 것이다. 이러한 도 15의 S110 단계 내지 S170 단계의 학습 절차는 충분히 학습된 것으로 판단될 때까지 다양한 상황에서 다양한 입력을 통해 반복해서 수행된

다.

[0106] 이제, 3가지 상황, 즉, 실내, 실외, 위험상황에 대해 충분히 학습된 것으로 가정하고, 본 발명의 실시예에 따른 상황 인식 기반의 이어폰 출력을 제어하기 위한 방법에 대해서 설명하기로 한다. 도 16은 본 발명의 실시예에 따른 상황 인식 기반의 이어폰 출력을 제어하기 위한 방법을 설명하기 위한 흐름도이다.

[0107] 도 16를 참조하면, 특징추출부(240)는 S210 단계에서 미리 설정된 소정 길이의 프레임 단위로 마이크(110), 통신부(210), 위치정보수신부(220) 및 센서부(230)로부터 사운드 신호, 통신 신호, GPS 신호 및 센서 신호 중 적어도 하나를 포함하는 특징 신호를 수집한다. 그런 다음, 특징추출부(240)는 S220 단계에서 특징 신호로부터 특징 파라미터를 추출하여 상황인식부(250)에 입력한다. 여기서, 특징 신호는 소정 프레임 간격으로 수집되고, 이러한 특징 신호는 특징 파라미터로 변환된다. 이에 따라, 특징 파라미터 또한 소정 프레임 시간 동안의 상황을 나타낸다. 이어서, 상황인식부(250)는 S230 단계에서 특징 파라미터를 소정 크기의 입력 행렬로 변환한다. 도 11의 예에서는 32×32 크기의 행렬이 사용되었다.

[0108] 그런 다음, 상황인식부(250)는 S240 단계에서 입력 행렬을 신경 네트워크에 입력하고, 그 신경 네트워크의 연산에 따라 출력값을 산출한다. 도 11에서 설명된 바와 같이, 신경 네트워크는 입력 행렬(IM) 및 복수의 계층(C1, S1, C2, S2, F1)의 출력이 가중치(w)가 곱해져서 다음 계층(C1, S1, C2, S2, F1, O1)의 입력이 된다. 이러한 가중치는 미리 학습에 의해 정해진다. 도 11에 보인 바와 같은 신경 네트워크에 입력 행렬(IM)을 입력으로 하면 제1 컨볼루션 계층(C1), 제1 풀링 계층(S1), 제1 컨볼루션 계층(C2), 제1 풀링 계층(S2), 완전 연결 계층(F1) 및 출력 계층(O1)에 의해 연산이 이루어진다. 특히, 신경 네트워크는 입력 행렬(IM) 및 복수의 계층(C1, S1, C2, S2, F1)의 출력이 가중치(w)가 곱해져서 다음 계층(C1, S1, C2, S2, F1, O1)의 입력이 된다. 이러한 연산에 따라 3개의 출력값(실내, 실외, 위급상황)을 얻을 수 있다.

[0109] 다음으로, 상황인식부(250)는 S250 단계에서 출력값에 따라 상황을 인식한다. 즉, 출력값에 따라 현재 실내에 위치해 있는지, 실외에 위치해 있는지 혹은 위급 상황인지를 인식한다. 예컨대, 도 11에서 제2 노드(N2)의 출력값이 가장 높은 경우, 현재 위치가 실외임을 인식할 수 있다. 이러한 인식 결과는 출력제어부(270)에 전달된다.

[0110] 출력제어부(270)는 S260 단계에서 상황인식부(250)의 인식 결과에 따라 이어폰 장치(100)의 스피커(120)로 출력되는 사운드 및 사운드의 볼륨 중 적어도 하나를 제어한다. 일례로, 상황 인식 결과, 실외인 경우, 출력제어부(270)는 마이크(110)를 통해 입력되는 주변 환경 소음 등을 재생되는 음악의 사운드와 함께 출력할 수 있다. 다른 예에 따르면, 상황 인식 결과, 실내인 경우, 출력제어부(270)는 마이크(110)로 입력되는 주변 환경 소음 등의 사운드 신호의 크기를 측정하고, 측정된 사운드 신호의 크기에 따라 스피커(120)를 통해 출력되는 음악 사운드의 볼륨을 조절한다. 즉, 실내가 조용한 경우, 볼륨을 낮추고, 실내가 시끄러운 경우, 볼륨을 높일 수 있다. 또 다른 예에 따르면, 상황인식 결과, 위급 상황인 경우, 출력제어부(270)는 음악을 끄고, 위급 상황임을 알리는 음성, 경고음 등을 스피커(120)를 통해 출력한다.

[0111] 한편, 앞서 도 3 내지 도 16을 통해 설명된 본 발명의 실시예에 따른 방법은 다양한 컴퓨터수단을 통하여 판독 가능한 프로그램 형태로 구현되어 컴퓨터로 판독 가능한 기록매체에 기록될 수 있다. 여기서, 기록매체는 프로그램 명령, 데이터 파일, 데이터구조 등을 단독으로 또는 조합하여 포함할 수 있다. 기록매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 예컨대 기록매체는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광 기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함한다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 와이어뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 와이어를 포함할 수 있다. 이러한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0112] 이상 본 발명을 몇 가지 바람직한 실시예를 사용하여 설명하였으나, 이들 실시예는 예시적인 것이며 한정적인 것이 아니다. 이와 같이, 본 발명이 속하는 기술분야에서 통상의 지식을 지닌 자라면 본 발명의 사상과 첨부된 특허청구범위에 제시된 권리범위에서 벗어나지 않으면서 균등론에 따라 다양한 변화와 수정을 가할 수 있음을 이해할 것이다.

부호의 설명

[0113] 100: 이어폰 장치 110: 마이크

120: 스피커 200: 사용자 장치

210: 통신부 220: 위치정보수신부

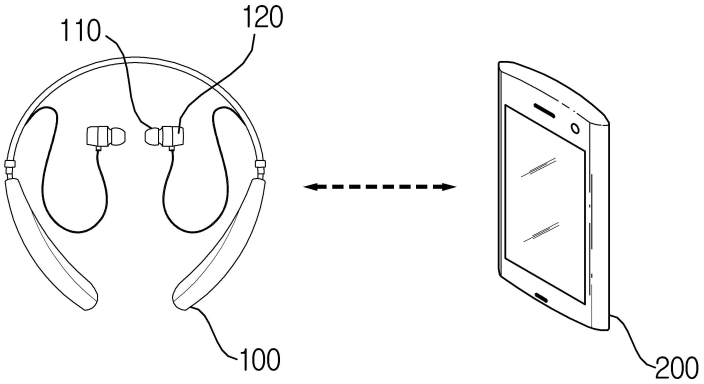
230: 센서부 240: 특징추출부

250: 상황인식부 260: 상황학스부

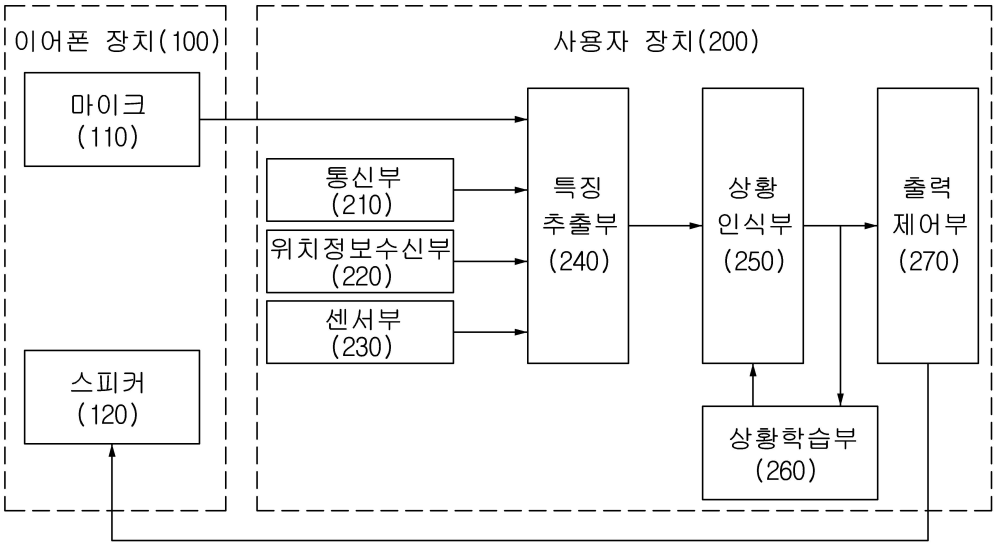
270: 출력제어부

도면

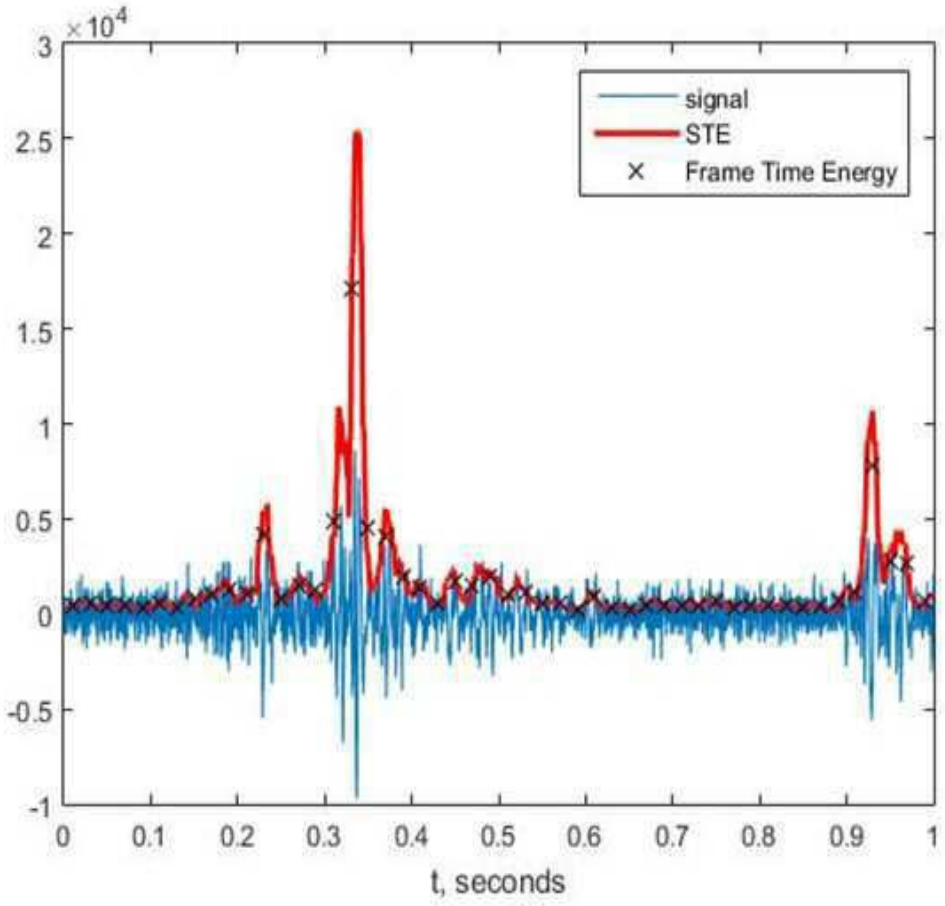
도면1



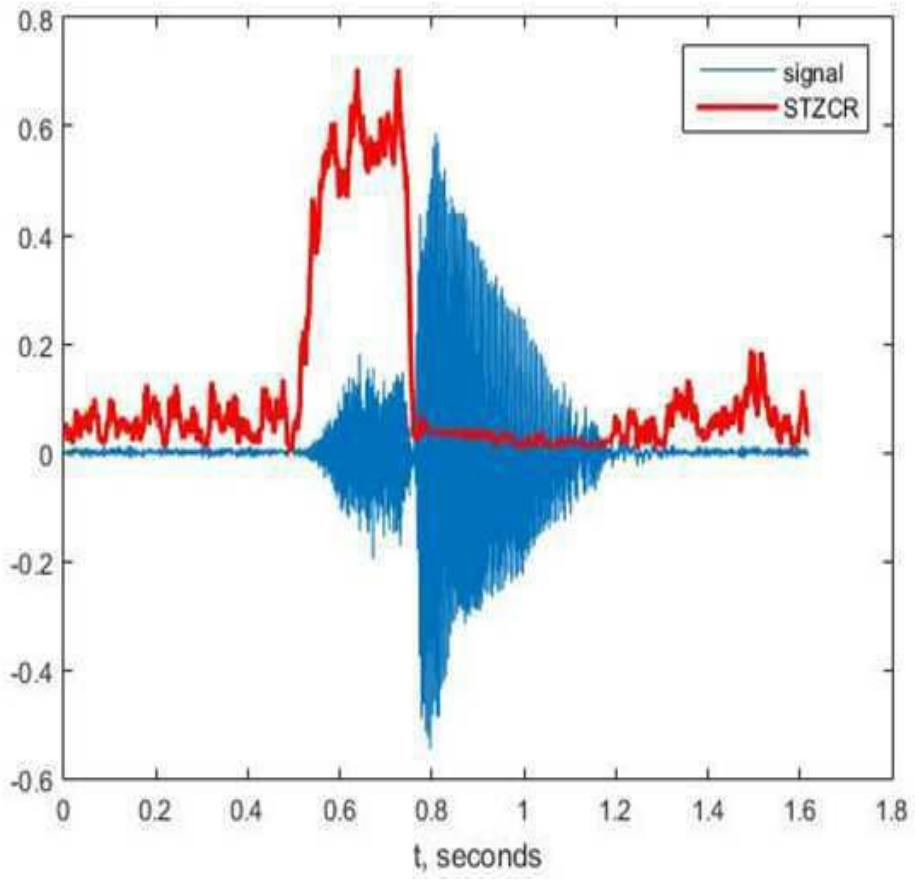
도면2



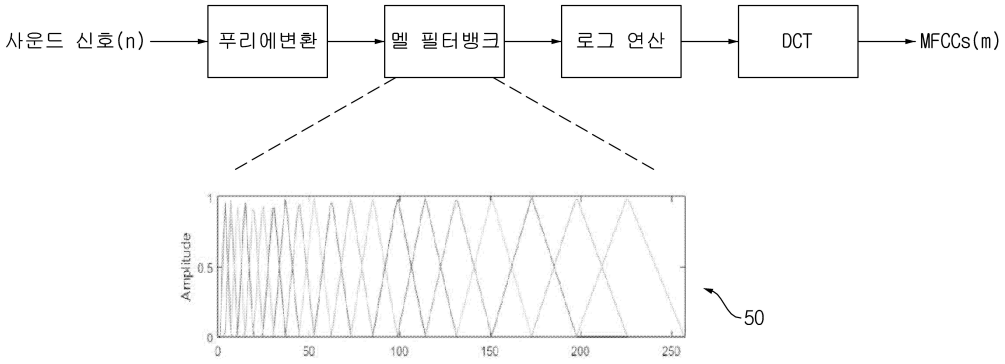
도면3


×10^4
3
2.5
2
1.5
1
0.5
0
-0.5
-1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
t, seconds
signal
STE
Frame Time Energy

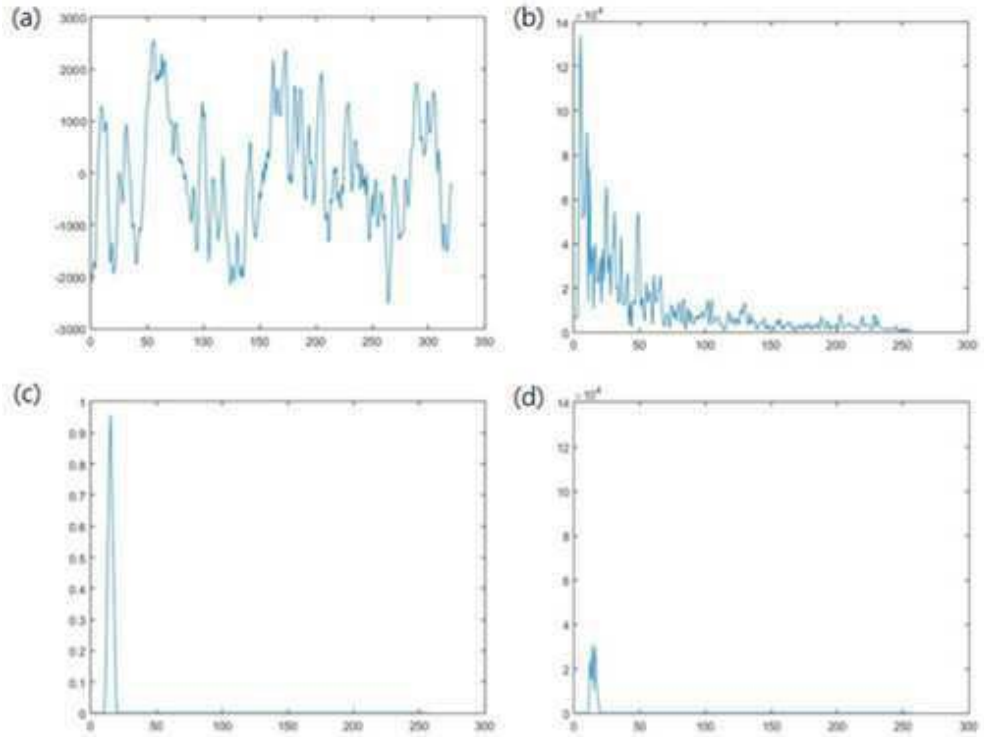
도면4


signal
STZCR
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
t, seconds

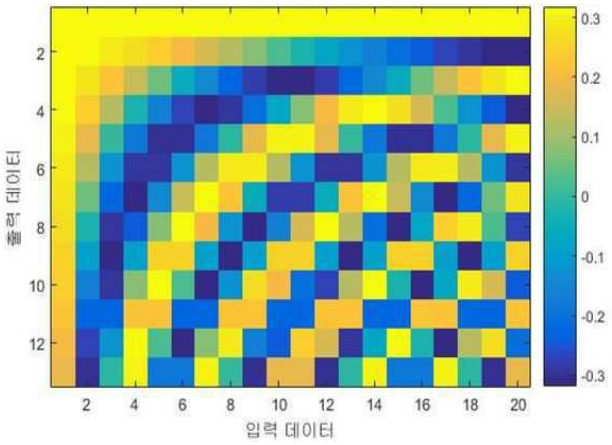
도면5


사운드 신호(n)
푸리에변환
멜 필터뱅크
로그 연산
DCT
MFCCs(m)
Amplitude
1
0.5
0
0
50
100
150
200
250
50

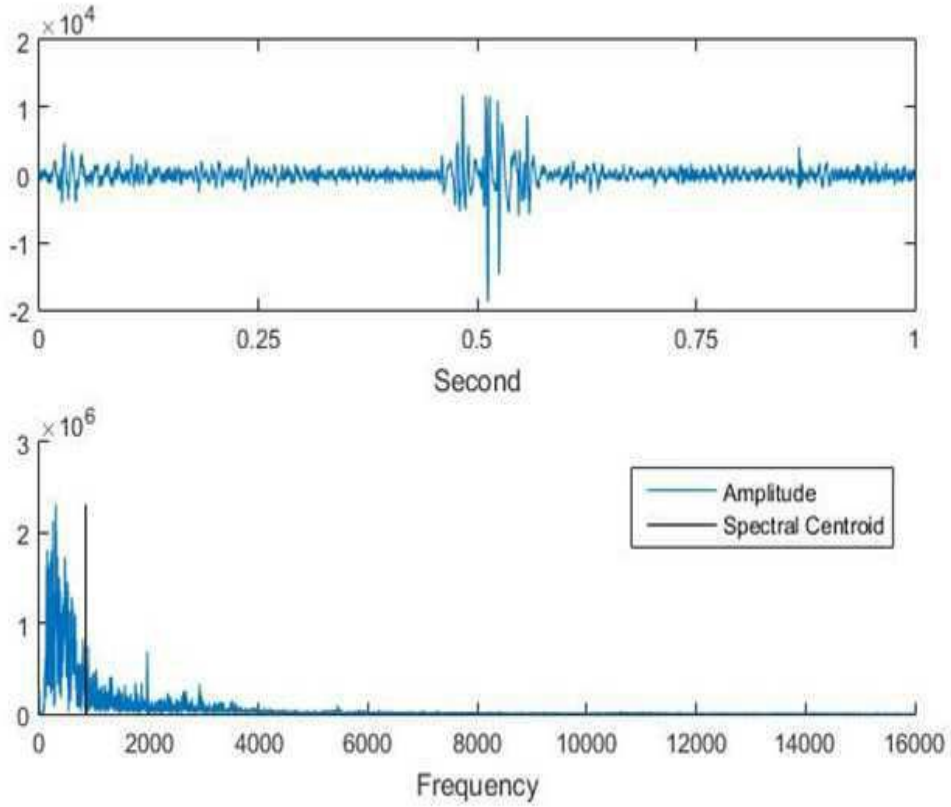
도면6



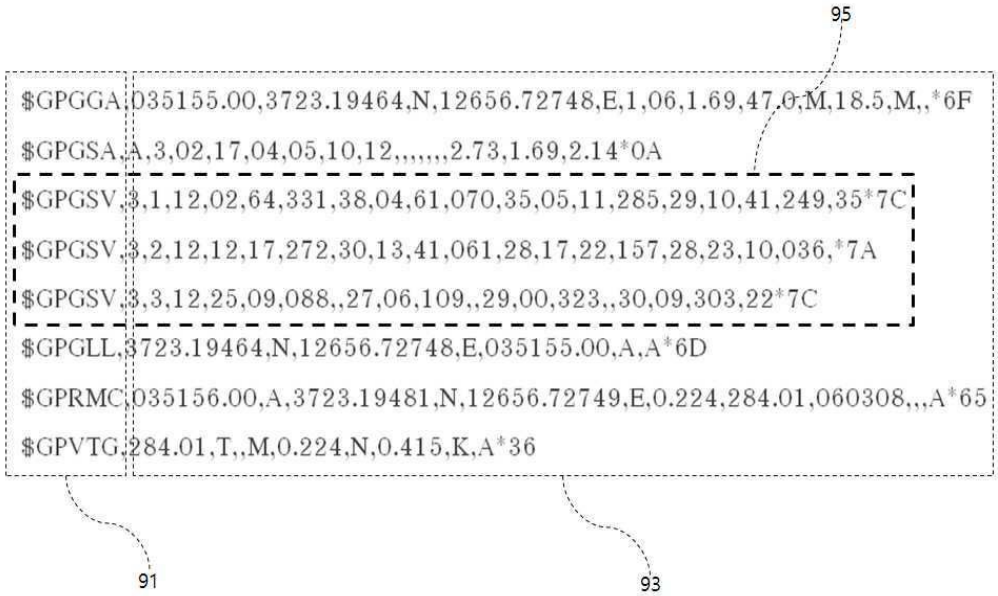
도면7



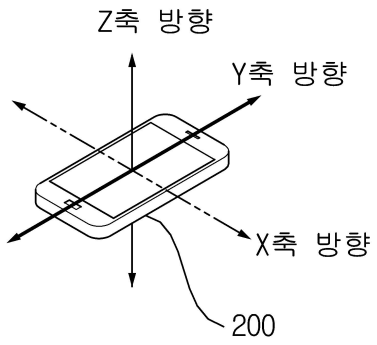
도면8


×10^4
2
1
0
-1
-2
0
0.25
0.5
0.75
1
Second
×10^6
3
2
1
0
0
2000
4000
6000
8000
10000
12000
14000
16000
Frequency
Amplitude
Spectral Centroid

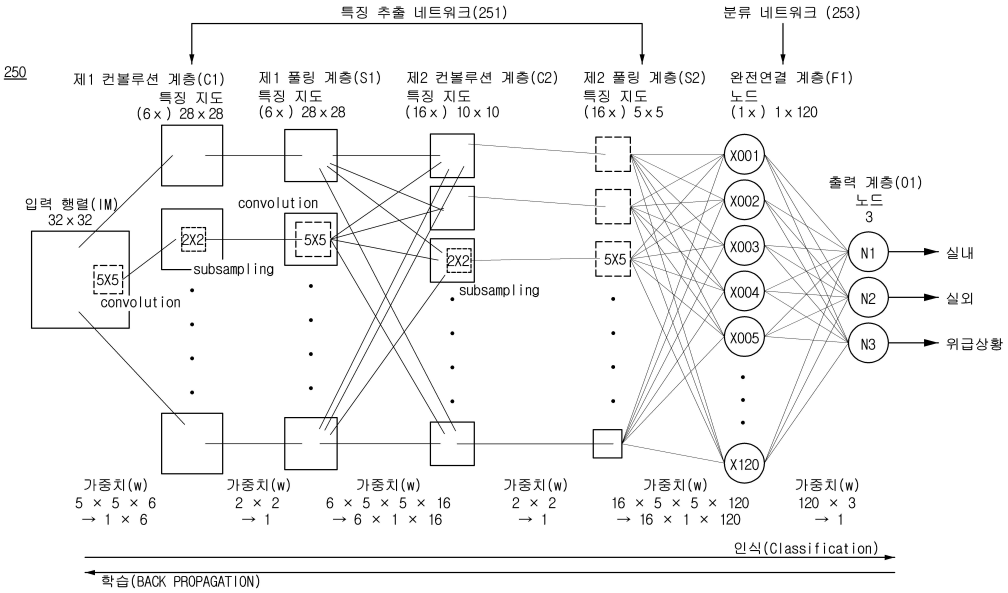
도면9


95
$GPGGA,035155.00,3723.19464,N,12656.72748,E,1,06,1.69,47.0,M,18.5,M,,*6F
$GPGSA,A,3,02,17,04,05,10,12,,,,,,,2.73,1.69,2.14*0A
$GPGSV,3,1,12,02,64,331,38,04,61,070,35,05,11,285,29,10,41,249,35*7C
$GPGSV,3,2,12,12,17,272,30,13,41,061,28,17,22,157,28,23,10,036,*7A
$GPGSV,3,3,12,25,09,088,,27,06,109,,29,00,323,,30,09,303,22*7C
$GPGLL,3723.19464,N,12656.72748,E,035155.00,A,A*6D
$GPRMC,035156.00,A,3723.19481,N,12656.72749,E,0.224,284.01,060308,,,A*65
$GPVTG,284.01,T,,M,0.224,N,0.415,K,A*36
91
93

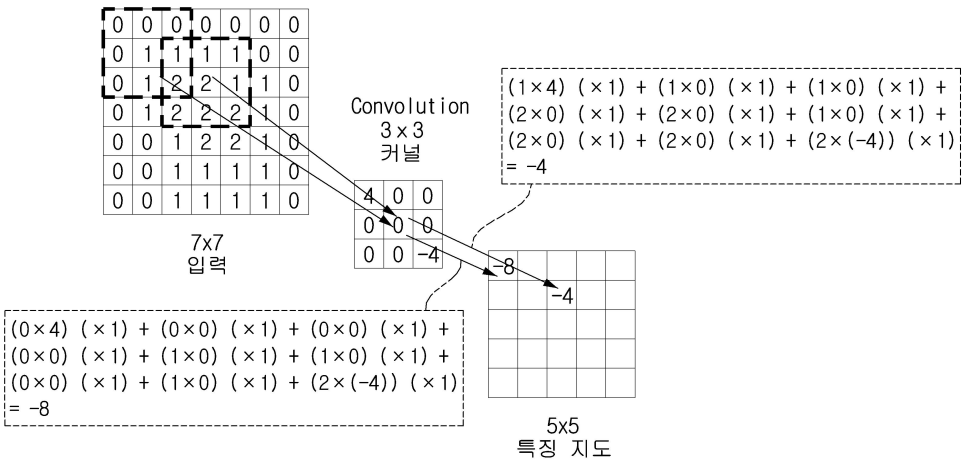
도면10



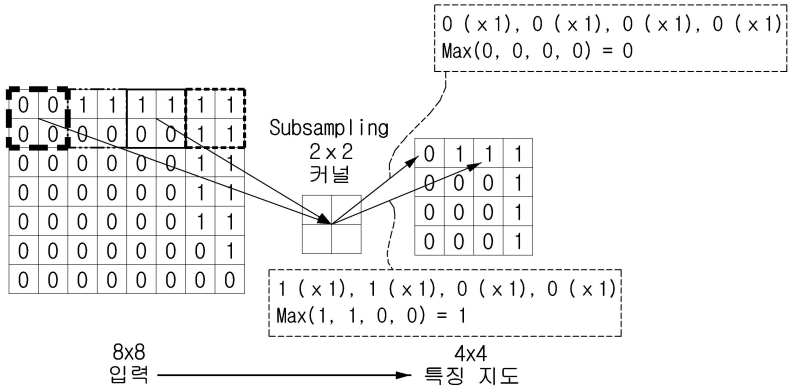
도면11



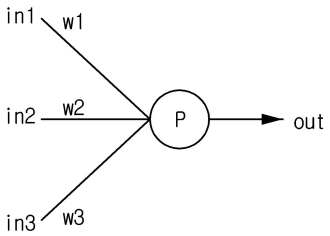
도면12



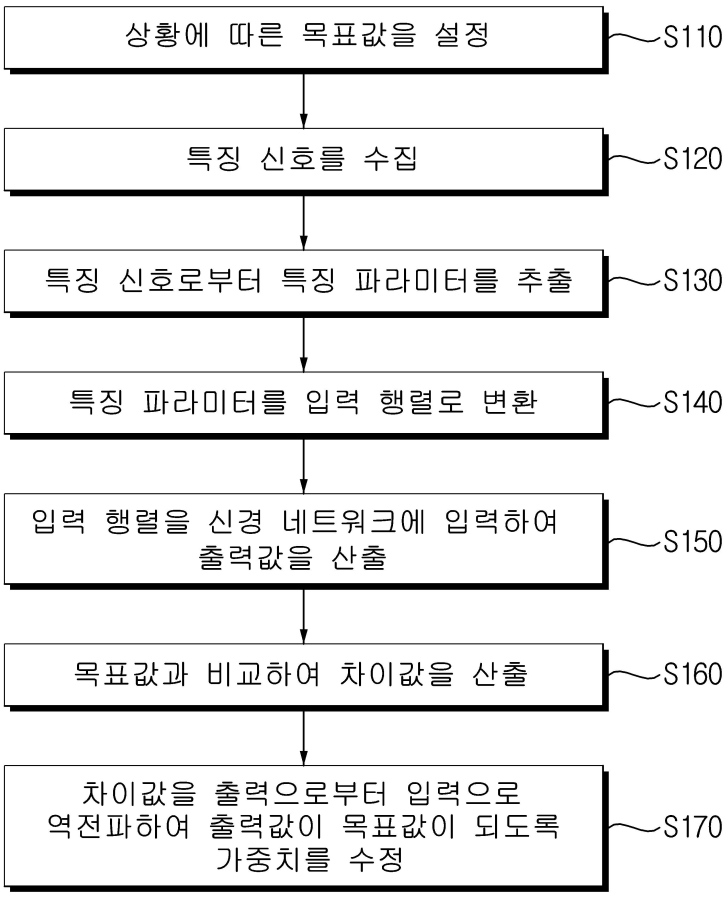
도면13



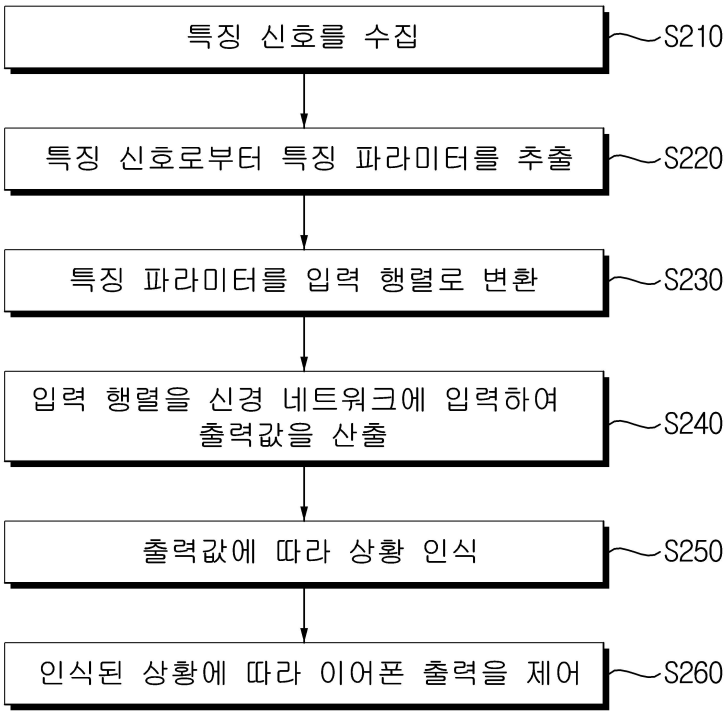
도면14



도면15



도면16


특징 신호를 수집
S210
특징 신호로부터 특징 파라미터를 추출
S220
특징 파라미터를 입력 행렬로 변환
S230
입력 행렬을 신경 네트워크에 입력하여 출력값을 산출
S240
출력값에 따라 상황 인식
S250
인식된 상황에 따라 이어폰 출력을 제어
S260