(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0086020
(43) 공개일자 2021년07월08일

(51) 국제특허분류(Int. Cl.)
G10L 15/22 (2006.01) *G05B 19/418* (2006.01)
G10L 15/06 (2006.01) *G10L 15/26* (2006.01)
G10L 21/0208 (2013.01)
(52) CPC특허분류
G10L 15/22 (2013.01)
G05B 19/418 (2013.01)
(21) 출원번호 10-2019-0179678
(22) 출원일자 2019년12월31일
심사청구일자 2019년12월31일

(71) 출원인
한국기술교육대학교 산학협력단
충청남도 천안시 동남구 병천면 충절로 1600 (한국기술교육대학교내)
(72) 발명자
서희석
충청남도 천안시 서북구 불당19로 95 린스트라우스 105-2202
성재용
서울특별시 송파구 송파대로 42길 18 101호
전광석
경기도 성남시 분당구 장미로 139 206-1103
(74) 대리인
이철희

전체 청구항 수 : 총 10 항

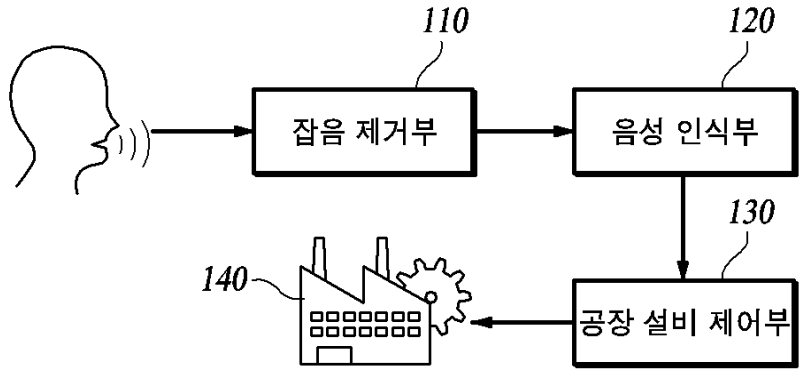
(54) 발명의 명칭 **딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 방법 및 시스템**

(57) 요 약

딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 방법 및 시스템을 개시한다.

본 실시예의 일 측면에 의하면, 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 방법에 있어서, 음원 신호를 수신하면, 심층 신경망 학습 알고리즘을 이용하여 상기 음원 신호를 음성 신호 및 잡음 신호로 분리하는 과정; 상기 음성 신호를 문자 신호로 변환하는 과정; 상기 문자 신호를 전송하는 과정; 및 상기 문자 신호를 수신한 공장 설비 제어부에서는, 상기 문자 신호를 인텐트와 엔티티로 분석하여, 상기 엔티티에 상기 인텐트에 대응하는 동작이 수행되도록 제어하는 과정을 포함하되, 상기 심층 신경망은 잡음 신호가 포함된 음원 신호에서 음성 신호만 추출하도록 학습된 신경망인 것을 특징으로 하는 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 방법을 제공한다.

대 표 도 - 도1



(52) CPC특허분류
G10L 15/063 (2013.01)
G10L 15/26 (2013.01)
G10L 21/0208 (2013.01)
G05B 2219/23373 (2013.01)

명 세 서

청구범위

청구항 1

딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 방법에 있어서,

음원 신호를 수신하면, 심층 신경망 학습 알고리즘을 이용하여 상기 음원 신호를 음성 신호 및 잡음 신호로 분리하는 과정;

상기 음성 신호를 문자 신호로 변환하는 과정;

상기 문자 신호를 전송하는 과정; 및

상기 문자 신호를 수신한 공장 설비 제어부에서는, 상기 문자 신호를 인텐트와 엔티티로 분석하여, 상기 엔티티에 상기 인텐트에 대응하는 동작이 수행되도록 제어하는 과정을 포함하되,

상기 심층 신경망은 잡음 신호가 포함된 음원 신호에서 음성 신호만 추출하도록 학습된 신경망인 것을 특징으로 하는 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 방법.

청구항 2

제 1 항에 있어서,

상기 심층 신경망은 임의의 신호 대 잡음 비(SNR: Signal-to-Noise Ratio)를 가지도록 상기 음원 신호를 합성하여 학습된 신경망인 것을 특징으로 하는 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 방법.

청구항 3

제 2 항에 있어서,

상기 음원 신호의 합성은 $SNR_{dB}=20\log_{10}\frac{A_{signal}}{A_{noise}}$ 을 이용하여 수행되는 것을 특징으로 하는 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 방법. (SNR_{dB} 는 신호 대 잡음 비, A_{signal}은 상기 음원 신호의 전압, A_{noise}는 상기 잡음 신호의 전압)

청구항 4

제 1 항에 있어서,

상기 음성 신호를 상기 문자 신호로 변환하는 과정은 클라우드 플랫폼을 이용하여 수행되는 것을 특징으로 하는 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 방법.

청구항 5

제 1 항에 있어서,

상기 엔티티는 유무선 통신이 가능한 IoT(Internet of Things) 공장 설비인 것을 특징으로 하는 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 방법.

청구항 6

딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 시스템에 있어서,

음원 신호를 수신하고, 심층 신경망 학습 알고리즘을 이용하여 상기 음원 신호를 음성 신호 및 잡음 신호로 분리하는 잡음 제거부;

상기 음성 신호를 문자 신호로 변환하고, 변환된 상기 문자 신호를 전송하는 음성 인식부; 및

상기 문자 신호를 수신하고, 상기 문자 신호를 인텐트와 엔티티로 분석하여, 상기 엔티티에 상기 인텐트에 대응하는 동작이 수행되도록 제어하는 공장 설비 제어부를 포함하되,

상기 심층 신경망은 잡음 신호가 포함된 음원 신호에서 음성 신호만 추출하도록 학습된 신경망인 것을 특징으로 하는 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 시스템.

청구항 7

제 6 항에 있어서,

상기 심층 신경망은 임의의 신호 대 잡음 비(SNR: Signal-to-Noise Ratio)를 가지도록 상기 음원 신호를 합성하여 학습된 신경망인 것을 특징으로 하는 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 시스템.

청구항 8

제 7 항에 있어서,

상기 음원 신호의 합성은 $SNR_{dB}=20\log_{10}\frac{A_{signal}}{A_{noise}}$ 을 이용하여 수행되는 것을 특징으로 하는 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 시스템. (SNR_{dB} 는 신호 대 잡음 비, A_{signal}은 상기 음원 신호의 전압, A_{noise}는 상기 잡음 신호의 전압)

청구항 9

제 6 항에 있어서,

상기 음성 인식부는 클라우드 플랫폼을 이용하여 상기 음성 신호를 상기 문자 신호로 변환하는 것을 특징으로 하는 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 시스템.

청구항 10

제 6 항에 있어서,

상기 엔티티는 유무선 통신이 가능한 IoT(Internet of Things) 공장 설비인 것을 특징으로 하는 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 시스템.

발명의 설명

기 술 분 야

[0001] 본 실시예는 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 방법 및 시스템에 관한 것이다.

배 경 기 술

[0002] 이 부분에 기술된 내용은 단순히 본 발명에 대한 배경 정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.

[0003] 일반적으로 "음성"을 이용하는 것은 사람과 컴퓨터의 상호작용에 가장 이상적인 인터페이스로 여겨져 왔다. 이는 음성이 사람이 가진 가장 본질적인 커뮤니케이션 방법이기 때문이다. 하지만 과거 음성 인식(Speech recognition)의 연구는 인식률이 좋지 않아 큰 인기를 얻지 못했다. 즉, 과거에는 엄청난 양의 음성 데이터 처리를 위한 컴퓨터의 연산 능력이 부족했고 이에 따른 낮은 인식률로 인해 음성 인식 기술은 2000년대 중반까지도 일반인에게 널리 사용되지 못하였다. 이후 음성 인식 기술이 본격적으로 사용되기 시작한 시점은 2011년 10월 애플의 음성 인식 기술 "Siri"를 탑재한 아이폰 4S의 출시 이후이다.

[0004] 한편, 스마트폰으로 시작된 음성 인식 기술은 이제 단순한 음성 인식 기술을 벗어나 사물 인터넷(IoT)의 보급으로 인해 집 안에서 본격적으로 경쟁이 불붙을 전망이다. 특히 주목할 부문은 기기가 음성을 매개로 명령을 내리고 대화를 나눌 수도 있는 음성인식 AI 스피커이다. 2014년 아마존이 최초의 음성 인식 AI 스피커 Echo를 출시한 이후로 Google, Apple, Microsoft 등 글로벌 IT 대기업들이 제각각 자사 소프트웨어를 탑재한 스피커의 출

시를 결정했으며, 해외 시장 변화에 발맞춰 국내 시장도 경쟁이 가속화되고 있다.

[0005] 한편, 가정뿐만 아니라 산업 시설에서도 음성 인식을 이용하여 IoT 설비를 제어하는 것에 대한 수요가 증가하고 있다. 그런데 건설 현장이나 공장 같은 산업 시설에서는 일반 가정에서와는 비교 안 될 정도의 소음이 존재하기 때문에 일반적인 음성 인식 기술로는 음성 인식 장치의 효율이 극히 낮아질 수 있다는 문제점이 존재한다.

발명의 내용

해결하려는 과제

[0006] 본 실시예는, 음성 인식 기술을 이용하여 공장 기계를 제어하되, 음성 인식시 딥러닝 기술을 이용하여 잡음을 제거함으로써 음성 인식에 대한 효율을 극대화할 수 있는, 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 방법 및 시스템을 제공하는 데 주된 목적이 있다.

과제의 해결 수단

[0007] 본 실시예의 일 측면에 의하면, 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 방법에 있어서, 음원 신호를 수신하면, 심층 신경망 학습 알고리즘을 이용하여 상기 음원 신호를 음성 신호 및 잡음 신호로 분리하는 과정; 상기 음성 신호를 문자 신호로 변환하는 과정; 상기 문자 신호를 전송하는 과정; 및 상기 문자 신호를 수신한 공장 설비 제어부에서는, 상기 문자 신호를 인텐트와 엔티티로 분석하여, 상기 엔티티에 상기 인텐트에 대응하는 동작이 수행되도록 제어하는 과정을 포함하되, 상기 심층 신경망은 잡음 신호가 포함된 음원 신호에서 음성 신호만 추출하도록 학습된 신경망인 것을 특징으로 하는 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 방법을 제공한다.

[0008] 본 실시예의 다른 측면에 의하면, 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 시스템에 있어서, 음원 신호를 수신하고, 심층 신경망 학습 알고리즘을 이용하여 상기 음원 신호를 음성 신호 및 잡음 신호로 분리하는 잡음 제거부; 상기 음성 신호를 문자 신호로 변환하고, 변환된 상기 문자 신호를 전송하는 음성 인식부; 및 상기 문자 신호를 수신하고, 상기 문자 신호를 인텐트와 엔티티로 분석하여, 상기 엔티티에 상기 인텐트에 대응하는 동작이 수행되도록 제어하는 공장 설비 제어부를 포함하되, 상기 심층 신경망은 잡음 신호가 포함된 음원 신호에서 음성 신호만 추출하도록 학습된 신경망인 것을 특징으로 하는 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 시스템을 제공한다.

발명의 효과

[0009] 이상에서 설명한 바와 같이 본 실시예에 의하면, 음성 인식 기술을 이용하여 공장 기계를 제어하되, 음성 인식시 딥러닝 기술을 이용하여 잡음을 제거함으로써 음성 인식에 대한 효율을 극대화할 수 있다는 효과가 있다.

도면의 간단한 설명

[0010] 도 1은 본 실시예에 따른 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 시스템을 나타낸 블록도이다.

도 2는 음성과 잡음이 섞여 있는 입력 파형을 음성 파형과 잡음 파형으로 각각 분리해내는 개념을 설명하기 위한 도면이다.

도 3은 심층 신경망을 이용하여 잡음과 음성이 섞여 있는 음원 파형에서 음성 파형만 추출하는 개념을 설명하기 위한 도면이다.

도 4는 딥러닝 음성 인식에 필요한 Python 프로그래밍의 예시를 나타낸 도면이다.

도 5는 동작 테이블의 예시를 나타낸 도면이다.

도 6은 본 실시예에 따른 음성 인식 AI 스피커의 음성 인식 API 구성에 대한 예시를 나타낸 도면이다.

도 7은 본 발명의 실시예에 따른 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 과정을 나타내는 순서도이다.

도 8은 본 발명의 실시예에 따른 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 과정의 예시를 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0011] 이하, 본 발명의 일부 실시예들을 예시적인 도면을 통해 상세하게 설명한다. 각 도면의 구성요소들에 참조부호를 부가함에 있어서, 동일한 구성요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부호를 가지도록 하고 있음에 유의해야 한다. 또한, 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.

[0012] 또한, 본 발명의 구성 요소를 설명하는 데 있어서, 제 1, 제 2, A, B, (a), (b) 등의 용어를 사용할 수 있다. 이러한 용어는 그 구성 요소를 다른 구성 요소와 구별하기 위한 것일 뿐, 그 용어에 의해 해당 구성 요소의 본질이나 차례 또는 순서 등이 한정되지 않는다. 명세서 전체에서, 어떤 부분이 어떤 구성요소를 '포함', '구비'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', '모듈' 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0013] 도 1은 본 실시예에 따른 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 시스템을 나타낸 블록도이다.

[0014] 도 1을 참조하면, 본 실시예에 따른 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 시스템은 잡음 제거부(110), 음성 인식부(120), 및 공장 설비 제어부(130) 등을 포함할 수 있다.

[0015] 본 실시예에 따른 잡음 제거부(110)는 음원 신호를 마이크로부터 실시간으로 수신하고, 심층 신경망 학습 알고리즘을 이용하여 음원 신호를 음성 신호 및 잡음 신호로 분리한다. 또한, 본 실시예에서는 잡음 제거부(110)가 마이크로부터 음원 신호를 획득한다고 명시하였지만 반드시 이에 한정되지 않고, 잡음 제거부(110) 자체에 음원 신호를 수집하기 위한 모듈이 구비될 수도 있다.

[0016] 여기서 음원 분리란 여러 개의 음원이 섞여 있는 입력 파형을 개별 음원의 파형으로 분리하는 것을 말한다. 다른 표현으로, 음성 강조 또는 잡음 제거라고 할 수 있으며, 음성과 잡음이 섞여 있는 입력 파형을 도 2와 같이 음성 파형과 잡음 파형으로 각각 분리해내는 것을 음원 분리라 한다.

[0017] 본 실시예에 따른 잡음 제거부(110)에서 이용하는 심층 신경망은 잡음 신호가 포함된 음원 신호에서 음성 신호만 추출하도록 학습된 신경망이다. 구체적으로, 본 실시예에 따른 심층 신경망은 임의의 신호 대 잡음 비(SNR: Signal-to-Noise Ratio)를 가지도록 음원 신호를 합성하여 학습된 신경망이다. 딥 러닝(Deep Learning)으로 음원 분리를 구현하려면 학습용 데이터 세트를 준비해야 한다. 훈련 데이터(Training Data)에는 음성과 잡음이 섞여 있는 파형이 필요하고, 또한 교사 데이터(Teacher Data)에는 음성만 있는 파형이 필요하다. 도 3을 참조하면, 이 데이터들을 이용해 본 실시예에 따른 심층 신경망(300)은 잡음이 섞여 있는 파형에서 음성만 추출하도록 학습되어, 본 실시예에 따른 잡음 제거부(110)에서는 이렇게 학습된 심층 신경망(300)을 이용하여 잡음과 음성이 섞여 있는 음원 파형(310)에서 음성 파형(320)만 추출할 수 있게 된다.

[0018] 다만, 이미 나와 있는 데이터 세트로는 공장 가동시의 응급 상황에 대처하는 기기 개발에는 부족하다. 따라서 본 실시예에서는 임의의 신호 대 잡음 비를 가진 파형을 합성하여 '음성과 잡음'이 혼합된 데이터 세트를 만들어, 개발 요건에 맞는 데이터 세트를 필요에 따라 생성하여 학습에 사용한다. 본 실시예에서는 Python으로 임의의 신호 대 잡음 비를 가지는 음성 파형을 제작하여 학습에 활용할 수 있으며, 이렇게 합성된 다양한 음성으로 심층 신경망(300)을 학습(딥러닝)시켜 성능을 개선한다. 도 4는 딥러닝 음성 인식에 필요한 Python 프로그래밍의 예시를 나타낸 도면이다.

[0019] 또한, 본 실시예에서 심층 신경망(300)을 학습시킬 때 이용될 음원 신호의 합성은 다음 수학식 1을 이용하여 수행된다.

수학식 1

[0020] $$SNR_{dB}=20\log_{10}\frac{A_{signal}}{A_{noise}}$$

[0021] 여기서, SNR_{dB}는 신호 대 잡음 비, A_{signal}은 음원 신호의 전압, A_{noise}는 잡음 신호의 전압을 나타낸다.

[0022] 본 실시예에 따른 음성 인식부(120)는 잡음 제거부(110)에서 분리시킨 음성 신호를 문자 신호로 변환하고, 변환된 문자 신호를 전송한다. 또한 본 실시예에서의 음성 인식부(120)는 클라우드 플랫폼을 이용하여 음성 신호를 문자 신호로 변환할 수 있다. 또 다른 실시예로 본 실시예에 따른 음성 인식부(120)는 Speech-to-Text(STT)를 포함하는 구성으로서, 잡음 제거부(110)에서 분리시킨 음성 신호에 대해 STT를 거쳐 자연어 처리/이해 및 응답을 담당하는 Dialog Manager 및 자연어 발화 여부를 확인하는 Voice Activity Detector(VAD)로 병렬 전달되도록 할 수 있다.

[0023] 본 실시예에 따른 공장 설비 제어부(130)는 음성 인식부(120)로부터 문자 신호를 수신하고, 수신한 문자 신호를 인텐트와 엔티티로 분석하여, 분석된 엔티티에 분석된 인텐트에 대응하는 동작이 수행되도록 제어한다. 본 실시예에서의 엔티티는 유무선 통신이 가능한 IoT(Internet of Things) 공장 설비일 수 있다. 예를 들어 "1번 기기 전원 켜 주세요" 또는 "2번 기기 전등을 켜 주세요" 라는 문장이 있는 경우, 인텐트는 어떤 것을 켠다는 의미로 정의할 수 있으며, 엔티티는 1번 기기의 전원 및 2번 기기의 전등으로 정의할 수 있다.

[0024] 본 실시예에 따른 AI 스피커에서 음성을 IoT 디바이스에 맵핑하기 위해서는 다음의 규칙을 사용할 수 있다.

[0025] (1) 인텐트로 음악을 들려줄지 아니면 IoT 디바이스 등을 제어해야 할지 판단한다.

[0026] (2) 엔티티가 되는 대상은 무엇인지 어떤 동작을 해야 하는지를 판단한다.

[0027] (3) 엔티티 정보를 리모컨 키 코드와 맵핑시킨다.

[0028] 본 실시예에서는 상술한 과정을 통해 해석된 인텐트에 따라 IoT 디바이스 제어 등의 기능을 수행하는 액션을 호출한다. 이때, 제어를 위해 3가지의 엔티티를 정의한다.

[0029] (1) @action 엔티티: 어떤 동작(turnon, turnoff, play 등)을 해야 하는지를 나타낸다.

[0030] (2) @target 엔티티: 동작 대상(IoT 디바이스 등)이 무엇인지를 나타낸다.

[0031] (3) @channel 엔티티: 케이블 디바이스의 채널을 나타낸다. (의미상으로는 @target과 동일함)

[0032] 이러한 동작 테이블의 예시를 나타낸 것이 도 5이다.

[0033] 전술한 본 실시예에 따른 잡음 제거부(110), 음성 인식부(120), 및 공장 설비 제어부(130)의 일부 구성을 음성 인식 AI 스피커를 통해 구현될 수 있다.

[0034] 도 6은 본 실시예에 따른 음성 인식 AI 스피커의 음성 인식 API 구성에 대한 예시를 나타낸 도면이다.

[0035] 도 6에 도시한 바와 같이, 본 실시예에 따른 음성 인식 AI 스피커(600)는 Speech-to-Text(STT)를 포함하여, 잡음 신호를 분리시킨 음성 신호에 대해 STT를 거쳐 자연어 처리/이해 및 응답을 담당하는 Dialog Manager 및 자연어 발화 여부를 확인하는 Voice Activity Detector(VAD)로 병렬 전달되도록 할 수 있다. 또한, 다른 실시예에서의 음성 인식 AI 스피커(600)는 클라우드 플랫폼(610)을 이용하여 음성 신호를 문자 신호로 변환할 수 있다.

[0036] 또한, 본 실시예에 따른 음성 인식 AI 스피커(600)는 MQTT(Message Queuing Telemetry Transport)를 이용하여 IoT 디바이스와 통신을 하여, 공장 전원 시스템의 파워를 관리한다든지 공장 서비스 제어를 위한 리모콘 제어를 할 수 있다.

[0037] 도 7은 본 발명의 실시예에 따른 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 과정을 나타내는 순서도이고, 도 8은 본 발명의 실시예에 따른 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 과정의 예시를 나타낸 도면이다.

[0038] 우선, 사용자로부터 음원 신호를 수신하면, 본 실시예에 따른 잡음 제거부(110)에서는 심층 신경망 학습 알고리즘을 이용하여 수신된 음원 신호를 음성 신호 및 잡음 신호로 분리한다(S700). 여기서, 심층 신경망은 잡음 신호가 포함된 음원 신호에서 음성 신호만 추출하도록 학습된 신경망으로, 이는 임의의 신호 대 잡음 비(SNR: Signal-to-Noise Ratio)를 가지도록 음원 신호를 합성하여 학습된 신경망이 될 수 있다.

[0039] 다음으로, 본 실시예에 따른 음성 인식부(120)에서는 분리된 음성 신호를 문자 신호로 변환한다(S710). 도 8을 참조하면, 사용자가 "1번 정지"(800)라고 이야기하면, 본 실시예에 따른 음성 인식부(120)에서는 STT(810)를 이용하여 이를 "1번 정지"(820)라는 문자 신호로 변환한다. 또한, 다른 실시예에서는, 음성 인식부(120)에서는

분리된 음성 신호를 인터넷(830)을 통해 클라우드 플랫폼(840)으로 전송하여 이를 문자 신호로 변환할 수 있다.

[0040] 이러한 과정을 통해 변환된 문자 신호는 무선 모듈(850)을 통해 Wifi 등의 무선 통신을 이용하여 IoT 공장 설비로 전송된다(S720).

[0041] 음성 인식부(120)로부터 문자 신호를 수신한 공장 설비 제어부(130)에서는, 수신한 문자 신호를 인텐트와 엔티티로 분석하여, 엔티티에 인텐트에 대응하는 동작이 수행되도록 제어한다(S730). 도 8을 참조하면, "1번 정지"라는 텍스트를 수신한 공장 기계 정지 모듈(860)에서는 제어 회로를 통해 1번 기기의 모터를 제어하여 1번 기기가 정지되도록 한다.

[0042] 도 7에서는 과정 S700 내지 과정 S730을 순차적으로 실행하는 것으로 기재하고 있으나, 이는 본 발명의 일 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것이다. 다시 말해, 본 발명의 일 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 일 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 7에 기재된 순서를 변경하여 실행하거나 과정 S700 내지 과정 S730 중 하나 이상의 과정을 병렬적으로 실행하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이므로, 도 7은 시계열적인 순서로 한정되는 것은 아니다.

[0043] 한편, 전술한 바와 같이 도 1에 기재된 본 실시예에 따른 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 시스템의 동작은 프로그램으로 구현되고 컴퓨터로 읽을 수 있는 기록매체에 기록될 수 있다. 본 실시예에 따른 딥러닝에 기반한 음성 인식을 이용하여 공장 설비를 제어하는 시스템의 동작을 구현하기 위한 프로그램이 기록되고 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 이러한 컴퓨터가 읽을 수 있는 기록매체는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등의 비일시적인(non-transitory) 매체일 수 있으며, 또한 캐리어 웨이브(예를 들어, 인터넷을 통한 전송) 및 데이터 전송 매체(data transmission medium)와 같은 일시적인(transitory) 매체를 더 포함할 수도 있다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수도 있다.

[0044] 이상의 설명은 본 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 실시예들은 본 실시예의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

[0045] 110: 잡음 제거부 120: 음성 인식부

130: 공장 설비 제어부 140: 공장 IoT 기기

300: 심층 신경망 600: 인공 지능 스피커

610: 클라우드 플랫폼 620: IoT 디바이스

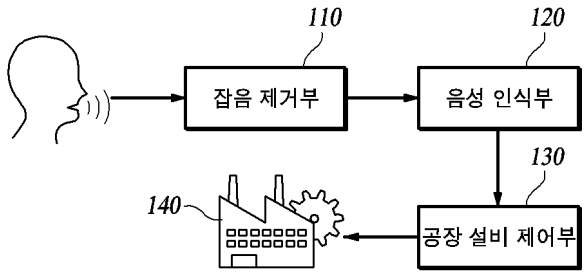
810: STT 830: 인터넷

840: 클라우드 플랫폼 850: 무선 모듈

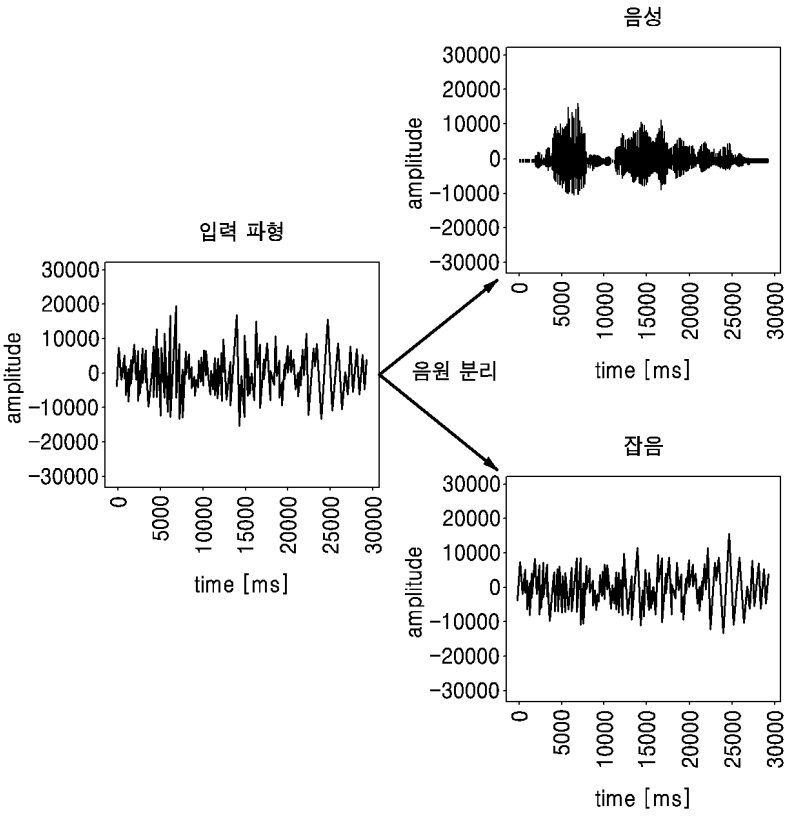
860: 공장 기계 정지 모듈

도면

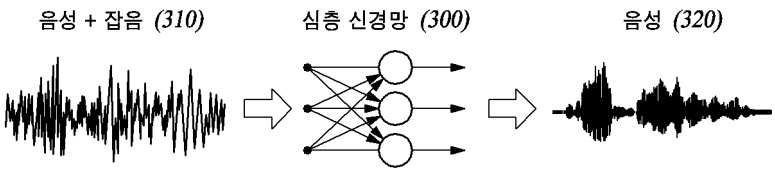
도면1



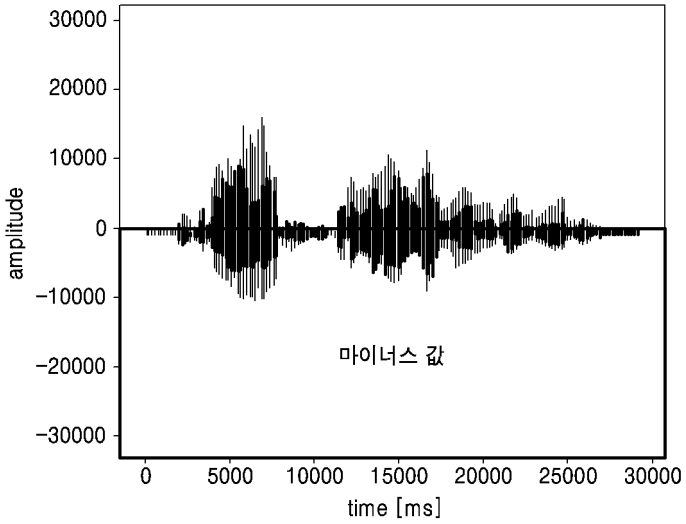
도면2



도면3



도면4



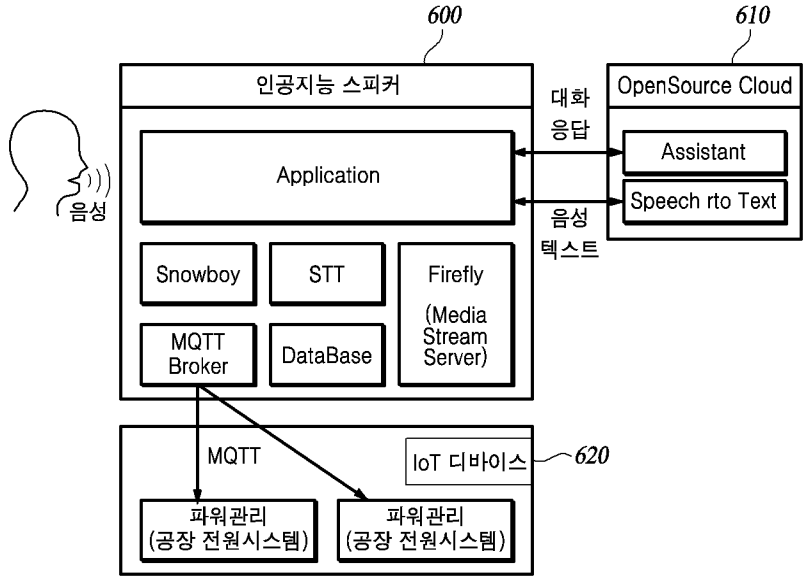
Python으로 임의 Signal-to-Noise ratio의 음성 파형 제작

```
def get_args():
parser = argparse.ArgumentParser()
parser.add_argument('--clean_file', type=str, required=True)
parser.add_argument('--noise_file', type=str, required=True)
parser.add_argument('--output_clean_file', type=str, default='')
parser.add_argument('--output_noise_file', type=str, default='')
parser.add_argument('--output_noisy_file', type=str, default='', required=True)
parser.add_argument('--snr', type=float, default='', required=True)
args = parser.parse_args()
return args
if __name__ == '__main__':
args = get_args()
clean_file = args.clean_file
noise_file = args.noise_file
snr = args.snr
```

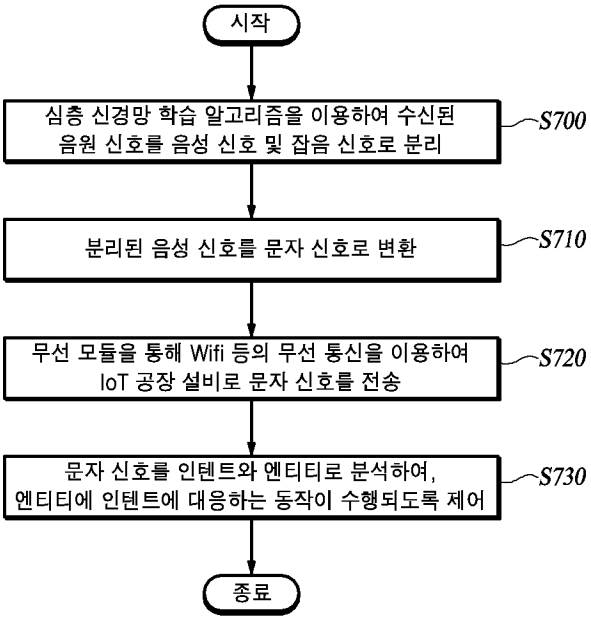
도면5

action	target	channel	
down	volumn	CH_9	CH_13
turnoff	music	CH_7	CH_170
turnon	aircon	CH_23	CH_16
play	tv	CH_5	CH_17
up		CH_53	CH_15
		CH_55	CH_11

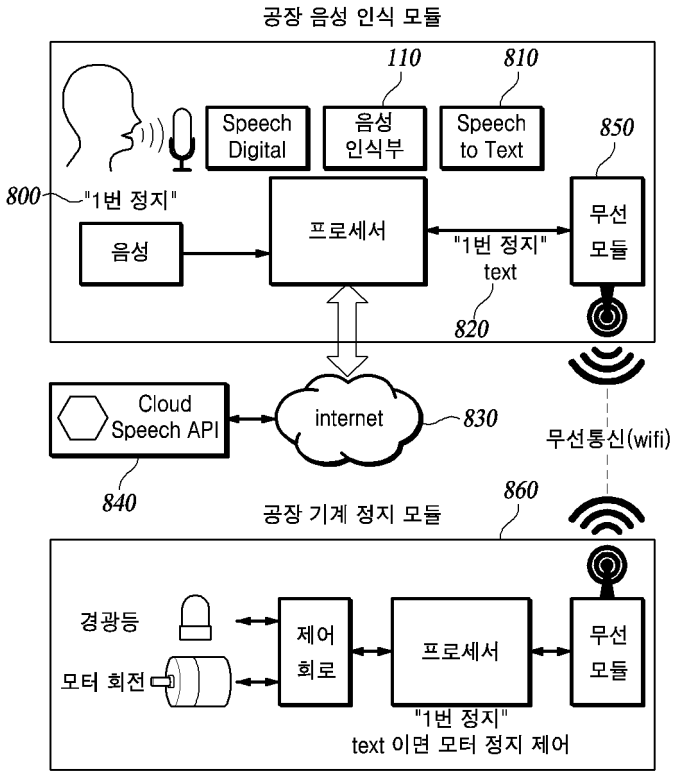
도면6


600
610
인공지능 스피커
OpenSource Cloud
대화
응답
음성
Application
Assistant
Speech rto Text
음성
텍스트
Snowboy
STT
Firefly
(Media
Stream
Server)
MQTT
Broker
DataBase
MQTT
IoT 디바이스
620
파워관리
(공장 전원시스템)
파워관리
(공장 전원시스템)

도면7


시작
심층 신경망 학습 알고리즘을 이용하여 수신된
음원 신호를 음성 신호 및 잡음 신호로 분리
S700
분리된 음성 신호를 문자 신호로 변환
S710
무선 모듈을 통해 Wifi 등의 무선 통신을 이용하여
IoT 공장 설비로 문자 신호를 전송
S720
문자 신호를 인텐트와 엔티티로 분석하여,
엔티티에 인텐트에 대응하는 동작이 수행되도록 제어
S730
종료

도면8


공장 음성 인식 모듈
110
810
Speech Digital
음성 인식부
Speech to Text
850
800
"1번 정지"
음성
프로세서
"1번 정지" text
무선 모듈
820
Cloud Speech API
internet
830
무선통신(wifi)
840
860
공장 기계 정지 모듈
경광등
모터 회전
제어 회로
프로세서
무선 모듈
"1번 정지" text 이면 모터 정지 제어