(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2018 년 6 월 21 일 (21.06.2018)

WIPO | PCT

(10) 국제공개번호
WO 2018/111038 A1

(51) 국제특허분류:
G10L 21/0216 (2013.01) *G10K 15/08* (2006.01)
G10L 25/30 (2013.01) *G10K 11/00* (2006.01)
G10K 11/18 (2006.01) *G06N 3/02* (2006.01)

(21) 국제출원번호: PCT/KR2017/014857

(22) 국제출원일: 2017 년 12 월 15 일 (15.12.2017)

(25) 출원언어: 한국어

(26) 공개언어: 한국어

(30) 우선권정보:
10-2016-0171359 2016 년 12 월 15 일 (15.12.2016) KR

(71) 출원인: 한양대학교 산학협력단 (INDUSTRY-UNIVERSITY COOPERATION FOUNDATION HANYANG UNIVERSITY) [KR/KR]; 04763 서울시 성동구 왕십리로 222, Seoul (KR).

(72) 발명자: 장준혁 (CHANG, Joon-Hyuk); 06294 서울시 강남구 언주로30길 26 지동 2708호, Seoul (KR). 이명인 (LEE, Myung In); 04760 서울시 성동구 마조로5길 12 501호, Seoul (KR).

(74) 대리인: 양성보 (YANG, Sungbo); 06099 서울시 강남구 선릉로125길 14 삼성빌딩 2층 피앤티특허법률사무소, Seoul (KR).

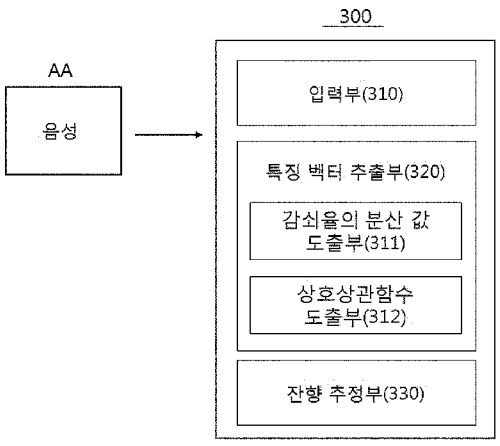
(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) **Title:** MULTICHANNEL MICROPHONE-BASED REVERBERATION TIME ESTIMATION METHOD AND DEVICE WHICH USE DEEP NEURAL NETWORK

(54) **발명의 명칭:** 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법 및 장치

[도3]



310 ... Input unit
320 ... Feature vector extraction unit
311 ... Decay rate dispersion value deriving unit
312 ... Cross correlation function deriving unit
330 ... Reverberation estimation unit
AA ... Voice

(57) **Abstract:** A multichannel microphone-based reverberation time estimation method and device which use a deep neural network (DNN) are disclosed. A multichannel microphone-based reverberation time estimation method using a DNN, according to one embodiment, comprises the steps of: receiving a voice signal through a multichannel microphone; deriving a feature vector including spatial information by using the inputted voice signal; and estimating the degree of reverberation by applying the feature vector to the DNN.

(57) **요약서:** 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법 및 장치가 제시된다. 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법은 다채널의 마이크를 통해 음성 신호를 입력 받는 단계; 입력 받은 상기 음성 신호를 이용하여 공간 정보가 포함된 특징 벡터를 도출하는 단계; 및 상기 특징 벡터를 심화 신경망(Deep Neural Network, DNN)에 적용하여 잔향의 정도를 추정하는 단계를 포함하여 이루어질 수 있다.

WO 2018/111038 A1

[다음 쪽 계속]

공개:
— 국제조사보고서와 함께 (조약 제21조(3))

명세서

발명의 명칭: 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법 및 장치

기술분야

[1] 아래의 실시예들은 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법 및 장치에 관한 것이다. 더욱 상세하게는, 다채널의 마이크를 통해 입력 받은 음성 신호를 이용하여 공간 정보를 내포할 수 있는 특징 벡터를 도출한 후 인공 신경망에 적용하여 잔향의 정도를 추정하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법 및 장치에 관한 것이다.

배경기술

[2] 대부분의 실생활 속 음향환경은 발화점에서 발생한 신호가 최단거리를 통해 전달될 뿐만 아니라 반사적인 경로를 거침으로써 잔향을 형성한다. 이러한 잔향은 음성인식, 방향 추정, 음성의 모델링, 및 위치 추정 등의 음성, 음향 신호처리 과정의 성능을 현저히 떨어뜨리는데, 이 과정에서 잔향의 정도를 정량적으로 추정하는 기술은 음성인식, 방향 추정, 음성의 모델링, 및 위치 추정 등의 알고리즘에 적용하여 적응적으로 성능을 개선시키는데 필수적이다.

[3] 종래의 잔향시간 추정 기술들은 잔향의 물리적 특징을 관찰하여 주파수 변조, 감쇠율의 분포 등과도 연관이 있음을 활용하여, 감쇠율의 분포를 2차 다항식 회귀 (polynomial regression)의 입력으로 이용해 잔향시간을 추정하는 방식이 사용되어 왔다.

[4] 단일한 특징 벡터에 의존해 잔향 추정하는 방식은 잡음에 강인하지 않은 특성을 보였으며 환경에 따라 다른 다항식을 적용해야 하는 단점이 있다. 또한 실험적으로 높은 연관성을 가지고 있는 특징 벡터를 활용해 모델링을 하더라도 다양한 잡음을 가지고 있는 실생활에서 적용되기에는 한계가 있으며, 그 정확도가 떨어진다.

[5] 한국공개특허 10-2016-0015005호는 이러한 클래스 기반 음향 모델의 변별 학습 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법 및 장치, 그리고 이를 이용한 음성 인식 장치에 관한 기술을 기재하고 있다.

발명의 상세한 설명

기술적 과제

[6] 실시예들은 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법 및 장치에 관하여 기술하며, 보다 구체적으로 공간 정보를 전혀 알지 못하는 상황에서 다채널의 마이크를 통해 입력 받은 음성 신호를 이용하여 공간 정보를 내포할 수 있는 특징 벡터를 도출한 후 인공 신경망에 적용하여 잔향의 정도를 추정하는 기술을 제공한다.

[7] 실시예들은 다채널 마이크를 이용하여 입력신호 사이의 상대적인 공간 정보를

추정에 활용할 수 있고, 공간의 잔향 특성을 잘 나타낼 수 있는 특징 벡터 간의 비선형적인 분포를 심화 신경망을 이용해 모델링하여 잔향의 정도를 추정함으로써, 정확도와 잡음에서의 강한 특성을 가지는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법 및 장치를 제공하는데 있다.

과제 해결 수단

[8] 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법은 다채널의 마이크를 통해 음성 신호를 입력 받는 단계; 입력 받은 상기 음성 신호를 이용하여 공간 정보가 포함된 특징 벡터를 도출하는 단계; 및 상기 특징 벡터를 심화 신경망(Deep Neural Network, DNN)에 적용하여 잔향의 정도를 추정하는 단계를 포함하여 이루어질 수 있다.

[9] 상기 다채널의 마이크를 통해 음성 신호를 입력 받는 단계는, 상기 다채널 마이크를 이용하여 입력되는 음성 신호 사이의 상대적인 공간 정보를 추정할 수 있다.

[10] 상기 입력 받은 상기 음성 신호를 이용하여 공간 정보가 포함된 특징 벡터를 도출하는 단계는, 입력 받은 상기 음성 신호에서 숏타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 이용해 시간 및 주파수 정보를 도출하고, 도출된 상기 시간 및 주파수 정보를 이용해 주파수 대역별로 포락선의 분포를 도출하여 감쇠율의 분산 값(Negative-side variance, NSV)을 도출하는 단계를 포함할 수 있다.

[11] 상기 감쇠율의 분산 값을 도출하는 단계는, 상기 숏타임 푸리에 변환 영역에서 로그 에너지(log-energy) 포락선을 도출하는 단계; 상기 로그 에너지 포락선에서 최소 자승 선형 피팅(least squares linear fitting)을 이용하여 기울기를 도출하는 단계; 및 양의 기울기를 가지는 것을 제외하고 음의 기울기를 가지는 잔향시간을 추정하기 위한 감쇠율의 분산 값을 도출하는 단계를 포함할 수 있다.

[12] 상기 특징 벡터를 심화 신경망에 적용하여 잔향의 정도를 추정하는 단계는, 도출된 상기 감쇠율의 분산 값을 상기 심화 신경망의 입력으로 사용하여 잔향시간을 추정할 수 있다.

[13] 상기 입력 받은 상기 음성 신호를 이용하여 공간 정보가 포함된 특징 벡터를 도출하는 단계는, 입력 받은 상기 음성 신호에서 두 개의 마이크 간의 상관관계를 나타내어 상기 다채널 마이크에서 음성의 방향의 추정하는 상호상관함수를 도출하는 단계를 포함할 수 있다.

[14] 상기 특징 벡터를 심화 신경망에 적용하여 잔향의 정도를 추정하는 단계는, 도출된 상기 감쇠율의 분산 값을 상기 심화 신경망의 입력으로 사용하여 잔향시간을 추정하고, 상기 다채널 마이크를 통해 도출되는 마이크 간의 공간 정보가 포함된 상기 상호상관함수의 값을 상기 심화 신경망의 입력으로 사용하여 잔향 시간을 추정할 수 있다.

[15] 상기 심화 신경망은, 세 개의 은닉 계층을 포함하며, 각 상기 은닉 계층은 다수

개의 에폭(epoch)으로 선행학습(pre-training) 과정을 거치고 미세 조정될 수 있다.

[16] 다른 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치는 다채널의 마이크를 통해 음성 신호를 입력 받는 입력부; 입력 받은 상기 음성 신호를 이용하여 공간 정보가 포함된 특징 벡터를 도출하는 특징 벡터 추출부; 및 상기 특징 벡터를 심화 신경망(Deep Neural Network, DNN)에 적용하여 잔향의 정도를 추정하는 잔향 추정부를 포함하여 이루어질 수 있다.

[17] 상기 특징 벡터 추출부는, 입력 받은 상기 음성 신호에서 숏타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 이용해 시간 및 주파수 정보를 도출하고, 도출된 상기 시간 및 주파수 정보를 이용해 주파수 대역별로 포락선의 분포를 도출하여 감쇠율의 분산 값(Negative-side variance, NSV)을 도출하는 감쇠율의 분산 값 도출부를 포함할 수 있다.

[18] 상기 감쇠율의 분산 값 도출부는, 상기 숏타임 푸리에 변환 영역에서 로그 에너지(log-energy) 포락선을 도출하고, 상기 로그 에너지 포락선에서 최소 자승 선형 피팅(least squares linear fitting)을 이용하여 기울기를 도출하여, 양의 기울기를 가지는 것을 제외하고 음의 기울기를 가지는 잔향시간을 추정하기 위한 감쇠율의 분산 값을 도출할 수 있다.

[19] 상기 잔향 추정부는, 도출된 상기 감쇠율의 분산 값을 상기 심화 신경망의 입력으로 사용하여 잔향시간을 추정할 수 있다.

[20] 상기 특징 벡터 추출부는, 입력 받은 상기 음성 신호에서 두 개의 마이크 간의 상관관계를 나타내어 상기 다채널 마이크에서 음성의 방향의 추정하는 상호상관함수를 도출하는 상호상관함수 도출부를 포함할 수 있다.

[21] 상기 잔향 추정부는, 도출된 상기 감쇠율의 분산 값을 상기 심화 신경망의 입력으로 사용하여 잔향시간을 추정하고, 상기 다채널 마이크를 통해 도출되는 마이크 간의 공간 정보가 포함된 상기 상호상관함수의 값을 상기 심화 신경망의 입력으로 사용하여 잔향 시간을 추정할 수 있다.

[22] 상기 심화 신경망은, 세 개의 은닉 계층을 포함하며, 각 상기 은닉 계층은 다수 개의 에폭(epoch)으로 선행학습(pre-training) 과정을 거치고 미세 조정될 수 있다.

발명의 효과

[23] 실시예들에 따르면 공간 정보를 전혀 알지 못하는 상황에서 다채널의 마이크를 통해 입력 받은 음성 신호를 이용하여 공간 정보를 내포할 수 있는 특징 벡터를 도출한 후 인공 신경망에 적용하여 잔향의 정도를 추정하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법 및 장치를 제공할 수 있다.

[24] 실시예들에 따르면 다채널 마이크를 이용하여 입력신호 사이의 상대적인 공간 정보를 추정에 활용할 수 있고, 공간의 잔향 특성을 잘 나타낼 수 있는 특징 벡터 간의 비선형적인 분포를 심화 신경망을 이용해 모델링하여 잔향의 정도를 추정함으로써, 정확도와 잡음에서의 강한 특성을 가지는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법 및 장치를 제공할 수 있다.

[25] 또한, 실시예들에 따르면 모바일 기기, IoT(Internet of Thanks) 기기 등 최근 활용도가 높아진 다채널 마이크를 활용한 기기들에 적용이 용이하며 기기에서 사용되는 잔향제거, 음성의 모델링, 음성인식 등의 중요 알고리즘에 필요한 정보를 제공할 수 있다.

도면의 간단한 설명

[26] 도 1은 일 실시예에 따른 잔향환경에서 다채널 마이크를 이용한 잔향신호의 입력 방법을 설명하기 위한 도면이다.

[27] 도 2는 일 실시예에 따른 잔향환경에서 다채널 마이크를 이용한 잔향시간 추정 알고리즘을 개략적으로 나타내는 도면이다.

[28] 도 3은 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치를 나타내는 도면이다.

[29] 도 4는 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법을 나타내는 흐름도이다.

[30] 도 5는 일 실시예에 따른 스펙트럼을 통해 감쇠율의 분산 값의 특징을 설명하기 위한 도면이다.

[31] 도 6은 일 실시예에 따른 스펙트럼을 통해 상호상관함수의 특징을 설명하기 위한 도면이다.

발명의 실시를 위한 형태

[32] 이하, 첨부된 도면을 참조하여 실시예들을 설명한다. 그러나, 기술되는 실시예들은 여러 가지 다른 형태로 변형될 수 있으며, 본 발명의 범위가 이하 설명되는 실시예들에 의하여 한정되는 것은 아니다. 또한, 여러 실시예들은 당해 기술분야에서 평균적인 지식을 가진 자에게 본 발명을 더욱 완전하게 설명하기 위해서 제공되는 것이다. 도면에서 요소들의 형상 및 크기 등은 보다 명확한 설명을 위해 과장될 수 있다.

[33] 아래의 실시예들은 공간 정보를 전혀 알지 못하는 상황에서 다채널 마이크를 이용하여 입력 받은 음성 신호들로부터 상대적인 공간 정보들을 획득하고 비선형적 분포특성을 깊은 구조 기반의 머신 러닝 기법인 심화 신경망을 통하여 모델링하여 음성인식에서 잔향을 추정하는 것을 특징으로 한다.

[34]

[35] 도 1은 일 실시예에 따른 잔향환경에서 다채널 마이크를 이용한 잔향신호의 입력 방법을 설명하기 위한 도면이다.

[36] 도 1을 참조하면, 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법은 공간 정보를 전혀 알지 못하는 상황에서 다채널의 마이크를 통해 입력 받은 음성 신호만을 가지고, 공간 정보를 내포할 수 있는 특징 벡터를 도출하고 인공 신경망에 적용하여 잔향의 정도를 추정하는 기술에 관한 것이다.

[37] 공간을 통해 전달되는 과정에서 발생한 잔향은 음성인식, 방향 추정, 음성의

모델링, 및 위치 추정 등의 음성, 음향 신호처리 과정의 정확도를 현저히 떨어뜨리므로, 그 정도를 정량적으로 추정하는 것은 음성 신호처리 분야에서 늘 필수적인 요소이다.

[38] 이에 다채널 마이크를 이용하여 음성 신호를 입력 받음으로써, 입력되는 음성 신호 사이의 상대적인 공간 정보를 추정할 수 있다.

[39] 다시 말하면, 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법은 다채널 마이크를 이용하여 입력신호 사이의 상대적인 공간 정보를 추정에 활용할 수 있으며, 뿐만 아니라 공간의 잔향 특성을 잘 나타낼 수 있는 특징 벡터 간의 비선형적인 분포를 깊은 구조 기반의 머신러닝 기법인 심화 신경망을 이용해 모델링하여 잔향의 정도를 추정할 수 있다.

[40] 또한, 실시예들은 모바일 기기, IoT 기기 등 다양한 환경에서 차용되고 있는 구조의 다채널 마이크를 이용해 기존의 잔향 추정 방법의 단점을 보완하고 정확도와 잡음에서의 강인함을 개선시킨 추정 방법을 제공할 수 있다.

[41]

[42] 도 2는 일 실시예에 따른 잔향환경에서 다채널 마이크를 이용한 잔향시간 추정 알고리즘을 개략적으로 나타내는 도면이다.

[43] 도 2를 참조하면, 일 실시예에 따른 잔향환경에서 다채널 마이크를 이용한 잔향시간 추정 알고리즘(200)은 다채널의 마이크(210)를 통해 입력 받은 음성 신호를 통해 공간 정보를 내포할 수 있는 특징 벡터를 도출(220)하고, 이를 인공 신경망(230)에 적용하여 잔향의 정도를 추정할 수 있다.

[44] 더 구체적으로, 잔향환경에서 다채널 마이크를 이용한 잔향시간 추정 알고리즘(200)에서는 다채널의 마이크(210)를 통해 입력 받은 음성 신호에서 상호상관함수를 도출하고, 숏타임 푸리에 변환(Short-Time Fourier Transform, STFT)를 이용해 도출한 시간, 주파수 정보를 이용해 주파수 대역별로 포락선의 분포를 도출(220)하여, 그것을 모두 심화 신경망(Deep Neural Network, DNN)(230)의 특징 벡터로 입력할 수 있다.

[45] 여기에서, 심화 신경망(230)은 많은 데이터 속에서 복잡한 비선형 관계를 모델링하는데 좋은 성능을 보이는 것으로 알려져 있으므로, 이미 알려져 있는 감쇠율의 분포와 잔향시간의 관계뿐만 아니라, 다채널 마이크(210)를 통해 도출할 수 있는 상호상관함수를 이용할 수 있다. 이것을 이용해 마이크 간의 공간 정보 또한 잔향시간 추정에 활용할 수 있다.

[46] 여기에서, 테스트 단계(Testing) 전에 미리 학습 단계(Training)가 수행될 수 있다.

[47] 학습 단계(Training)에서는 다양한 화자, 문장, 그리고 환경을 내포하고 있는 충분한 분량의 음성 데이터베이스(Database, DB)를 이용하여 특징 벡터를 도출하고, 심화신경망을 통해 잔향시간을 추정할 수 있다.

[48] 이에, 테스트 단계(Testing)는 다양한 잡음 환경에 의해 오염된 음성 신호로부터 특징 벡터들을 도출하고, 특징 벡터들을 미리 학습된 심화 신경망을 통과시켜

잔향시간을 추정할 수 있다.

[49] 이러한 과정을 통해 모델링만을 이용해서는 도출할 수 없는 특징 벡터 간의 관계와 잔향시간의 관계를 동시에 도출해낼 수 있으며, 그 결과 기존 방법들을 상호보완 하는 결과를 기대할 수 있다. 이에 따라 기존의 추정 방식들에 비해 높은 정확도와 잡음에 강인한 특성을 가질 수 있다.

[50] 아래에서 일 실시예에 따른 잔향환경에서 다채널 마이크를 이용한 잔향시간 추정 알고리즘을 더 구체적으로 설명하기로 한다.

[51]

[52] 도 3은 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치를 나타내는 도면이다.

[53] 도 3을 참조하면, 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치(300)는 입력부(310), 특징 벡터 추출부(320), 및 잔향 추정부(330)를 포함하여 이루어질 수 있다.

[54] 입력부(310)는 다채널의 마이크를 통해 음성 신호를 입력 받을 수 있다. 여기에서 음성 신호는 오염된 음성 신호(즉, 주변 잡음에 의해 오염된 음성 신호)일 수 있다.

[55] 특징 벡터 추출부(320)는 입력 받은 음성 신호를 이용하여 공간 정보가 포함된 특징 벡터를 도출할 수 있다.

[56] 특징 벡터 추출부(330)는 감쇠율의 분산 값 도출부(311) 및 상호상관함수 도출부(312)를 포함할 수 있다.

[57] 감쇠율의 분산 값 도출부(311)는 입력 받은 음성 신호에서 숏타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 이용해 시간 및 주파수 정보를 도출하고, 도출된 시간 및 주파수 정보를 이용해 주파수 대역별로 포락선의 분포를 도출하여 감쇠율의 분산 값(Negative-side variance, NSV)을 도출할 수 있다.

[58] 더 구체적으로, 감쇠율의 분산 값 도출부(311)는 숏타임 푸리에 변환 영역에서 로그 에너지(log-energy) 포락선을 도출하고, 로그 에너지 포락선에서 최소 자승 선형 피팅(least squares linear fitting)을 이용하여 기울기를 도출하여, 양의 기울기를 가지는 것을 제외하고 음의 기울기를 가지는 잔향시간을 추정하기 위한 감쇠율의 분산 값을 도출할 수 있다.

[59] 이와 같은 방법으로 도출된 특징 벡터는 기울기만을 관찰하거나 감쇠율 자체만을 관찰했을 경우와 비교하여 잔향시간과 직접적으로 높은 관계를 보인다.

[60] 상호상관함수 도출부(312)는 입력 받은 음성 신호에서 두 개의 마이크 간의 상관관계를 나타내어 다채널 마이크에서 음성의 방향의 추정하는 상호상관함수를 도출할 수 있다.

[61] 상호상관함수 도출부(312)는 다채널 마이크를 통해 입력된 마이크 간의 두 신호의 상관관계를 나타내기 위해, 두 신호의 푸리에 변환을 이용하여

상호상관함수(GCC-PHAT) 값을 심화 신경망(DNN)의 입력으로 사용할 수 있다. 이 때 음성의 반사음과 직접음의 상관관계를 이용하기 위해 전후 소정의 프레임의 상호상관함수(GCC-PHAT) 값을 심화 신경망(DNN)의 입력으로 사용할 수 있다.

[62] 이러한 감쇠율의 분산 값 도출부(311) 및 상호상관함수 도출부(312)에서 추출된 값을 잔향 추정부의 심화 신경망의 입력으로 사용하여 잔향 정도를 추정할 수 있다.

[63] 잔향 추정부(330)는 특징 벡터를 심화 신경망(Deep Neural Network, DNN)에 적용하여 잔향의 정도를 추정할 수 있다.

[64] 잔향 추정부(330)는 도출된 감쇠율의 분산 값을 심화 신경망의 입력으로 사용하여 잔향시간을 추정할 수 있다. 또한, 잔향 추정부(330)는 도출된 마이크 간 공간 정보가 포함된 상호상관함수의 값을 심화 신경망의 입력으로 사용하여 잔향시간을 추정할 수 있다.

[65] 즉, 잔향 추정부(330)는 도출된 감쇠율의 분산 값을 심화 신경망의 입력으로 사용하여 잔향시간을 추정하고, 다채널 마이크를 통해 도출되는 마이크 간의 공간 정보를 추정함으로써 잔향 시간을 추정할 수 있다.

[66] 심화 신경망은 세 개의 은닉 계층을 포함하며, 각 은닉 계층은 다수 개의 에폭(epoch)으로 선행학습(pre-training) 과정을 거치고 미세 조정될 수 있다.

[67] 더 구체적으로, 심화 신경망은 세 개의 은닉 계층과 1500개의 노드로 구성될 수 있으며, 각각의 은닉 계층은 50개의 에폭(epoch)으로 선행학습(pre-training) 과정을 거치고, 500개의 에폭(epoch)으로 미세 조정될 수 있다.

[68] 이에 따라 높은 정확도와 잡음에 강인한 특성을 가진 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치를 제공할 수 있다.

[69] 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치(300)는 마이크 간의 공간 에일리어싱(spatial aliasing)을 고려한 밀집된 구조의 다채널 마이크에 적용이 가능하여, 소형화된 기기에 적용이 가능하다.

[70]

[71] 도 4는 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법을 나타내는 흐름도이다.

[72] 도 4를 참조하면, 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법은 다채널의 마이크를 통해 음성 신호를 입력 받는 단계(410), 입력 받은 음성 신호를 이용하여 공간 정보가 포함된 특징 벡터를 도출하는 단계(420), 및 특징 벡터를 심화 신경망(Deep Neural Network, DNN)에 적용하여 잔향의 정도를 추정하는 단계(430)를 포함하여 이루어질 수 있다.

[73] 여기에서, 입력 받은 음성 신호를 이용하여 공간 정보가 포함된 특징 벡터를 도출하는 단계(420)는 입력 받은 음성 신호에서 숏타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 이용해 시간 및 주파수 정보를 도출하고, 도출된 시간 및 주파수 정보를 이용해 주파수 대역별로 포락선의 분포를 도출하여 감쇠율의

분산 값(Negative-side variance, NSV)을 도출하는 단계를 포함할 수 있다.

[74] 또한, 입력 받은 음성 신호를 이용하여 공간 정보가 포함된 특징 벡터를 도출하는 단계(420)는 입력 받은 음성 신호에서 두 개의 마이크 간의 상관관계를 나타내어 다채널 마이크에서 음성의 방향의 추정하는 상호상관함수를 도출하는 단계를 포함할 수 있다.

[75] 아래에서 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법을 하나의 예를 들어 더 구체적으로 설명한다.

[76]

[77] 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법은 도 3에서 설명한 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치를 이용하여 더 구체적으로 설명할 수 있다. 여기에서 일 실시예에 따른 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치는 입력부, 특징 벡터 추출부, 및 잔향 추정부를 포함하여 이루어질 수 있다.

[78] 단계(410)에서, 입력부는 다채널의 마이크를 통해 음성 신호를 입력 받을 수 있다. 이에, 다채널 마이크를 이용하여 입력되는 음성 신호 사이의 상대적인 공간 정보를 추정할 수 있다.

[79] 단계(420)에서, 특징 벡터 추출부는 입력 받은 음성 신호를 이용하여 공간 정보가 포함된 특징 벡터를 도출할 수 있다.

[80] 이러한 특징 벡터 추출부는 감쇠율의 분산 값 도출부 및 상호상관함수 도출부를 포함할 수 있다.

[81] 감쇠율의 분산 값 도출부는 특징 벡터 추출부는 입력 받은 음성 신호에서 숏타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 이용해 시간 및 주파수 정보를 도출하고, 도출된 시간 및 주파수 정보를 이용해 주파수 대역별로 포락선의 분포를 도출하여 감쇠율의 분산 값(Negative-side variance, NSV)을 도출할 수 있다.

[82] 더 구체적으로, 감쇠율의 분산 값 도출부는 입력 받은 음성 신호로부터 숏타임 푸리에 변환 영역에서 로그 에너지(log-energy) 포락선을 도출하는 단계, 로그 에너지 포락선에서 최소 자승 선형 피팅(least squares linear fitting)을 이용하여 기울기를 도출하는 단계, 및 양의 기울기를 가지는 것을 제외하고 음의 기울기를 가지는 잔향시간을 추정하기 위한 감쇠율의 분산 값을 도출하는 단계를 통해 감쇠율의 분산 값을 도출할 수 있다.

[83] 특징 벡터를 심화 신경망에 적용하여 잔향의 정도를 추정하는 단계는, 도출된 감쇠율의 분산 값을 심화 신경망의 입력으로 사용하여 잔향시간을 추정할 수 있다.

[84] 그리고 상호상관함수 도출부는 입력 받은 음성 신호에서 두 개의 마이크 간의 상관관계를 나타내어 다채널 마이크에서 음성의 방향의 추정하는 상호상관함수를 도출할 수 있다.

[85] 단계(430)에서, 잔향 추정부는 특징 벡터를 심화 신경망(Deep Neural Network,

DNN)에 적용하여 잔향의 정도를 추정할 수 있다.

[86] 잔향 추정부는 도출된 감쇠율의 분산 값을 심화 신경망의 입력으로 사용하여 잔향시간을 추정하는 동시에, 다채널 마이크를 통해 도출되는 상호상관함수의 값을 심화 신경망의 입력으로 사용하여 마이크 간의 공간 정보를 추정할 수 있다.

[87] 여기에서 심화 신경망은 세 개의 은닉 계층을 포함하며, 각 은닉 계층은 다수 개의 에폭(epoch)으로 선행학습(pre-training) 과정을 거치고 미세 조정될 수 있다.

[88]

[89] 아래에서 특징 벡터의 추출 방법에 대해 더 구체적으로 설명한다.

[90] 특징 벡터는 감쇠율의 분산 값(Negative-side variance, NSV) 및 상호상관함수(Generalized cross-correlation with phase transform, GCC-PHAT 또는 GCC)가 사용될 수 있다.

[91] 감쇠율의 분산 값(Negative-side variance, NSV)은 숏타임 푸리에 변환(Short-Time Fourier Transform, STFT) 영역에서 로그 에너지(log-energy) 포락선을 도출하고, 최소 자승 선형 피팅(least squares linear fitting)을 이용해 기울기를 도출한 뒤, 그 중에서 양의 기울기를 가지는 것을 제외하고 음의 기울기를 가지는 감쇠율의 분산 값을 도출할 수 있다.

[92] 이와 같은 방법으로 도출된 특징 벡터는 기울기만을 관찰하거나 감쇠율 자체만을 관찰했을 경우와 비교하여 잔향시간과 직접적으로 높은 관계를 보인다. 이러한 특징은 아래의 실험 결과를 통해 확인할 수 있다.

[93] 도 5는 일 실시예에 따른 스펙트럼을 통해 감쇠율의 분산 값의 특징을 설명하기 위한 도면이다.

[94] 도 5를 참조하면, 일 실시예에 따른 심화 신경망(DNN)을 이용한 다채널 마이크 기반의 잔향시간 추정 방법의 적용한 실험에서 획득한 스펙트럼을 통해 특징 벡터 중 감쇠율의 분산 값(NSV)의 특징을 나타낸다. 이 때, T_{60} = 0.20, 0.35, 0.50, 0.65, 0.80, 1.10 s 에 대한 스펙트럼을 통해 감쇠율의 분산 값(NSV)의 특징을 확인할 수 있다.

[95] 여기에서 실험에 사용된 수치는 Sampling rate는 8,000 Hz, 숏타임 푸리에 변환(STFT) 영역에서의 관찰을 위해 사용한 Hamming window의 크기는 256, FFT 크기는 128, 및 최소 자승 선형 피팅(least squares linear fitting)은 숏타임 푸리에 변환(STFT) 영역에서 30개의 순차적 프레임을 이용해 기울기를 도출하며, 이는 25%씩 오버랩(overlap)된 상태로 도출할 수 있다.

[96] 또한, 상호상관함수(Generalized cross-correlation with phase transform, GCC-PHAT 또는 GCC)는 두 신호 간의 상관관계를 나타내는 지표로, 다채널 마이크에서 음성의 방향의 추정하는데 자주 사용되는 특징 벡터다.

[97] 두 신호 $x_1(n)$, $x_2(n)$가 정의되어 있을 때 상호상관함수(GCC-PHAT)를 다음 식과 같이 나타낼 수 있다.

[98] [수학식 1]

[99] $G_{PHAT}(f)=X_1(f)[X_2(f)^*]/|X_1(f)[X_2(f)^*]|$

[100] 여기에서, $X_1(f)$와 $X_2(f)$는 각 두 신호의 푸리에 변환을 나타내고, $[\,]^*$는 복소 결합(complex conjugate)을 나타낼 수 있다.

[101] $G_{PHAT}(f)$에 역푸리에 변환을 적용하는 경우, 상호상관관계 정보를 내포한 $R_{PHAT}(d)$를 도출할 수 있다. 이 특징 벡터의 일반적인 활용에서는 방향 추정을 위해서 도출한 $R_{PHAT}(d)$에서 최대값의 인덱스만을 도출해 수식에 적용할 수 있다.

[102] 여기에서는 음성의 반사음과 직접음의 상관관계를 이용하기 위해 전후 10 프레임씩, 총 21 프레임의 상호상관함수(GCC-PHAT) 값을 심화 신경망(DNN)의 입력으로 사용할 수 있다.

[103] 또한 특징 벡터는 감쇠율의 분산 값(NSV)와 같은 동조(sync)로 도출할 수 있다. 즉, 사용한 Hamming window의 크기는 256이고, FFT 크기는 128이며, 25%씩 오버랩(overlap)된 상태로 도출할 수 있다.

[104] 도 6은 일 실시예에 따른 스펙트럼을 통해 상호상관함수의 특징을 설명하기 위한 도면이다.

[105] 도 6을 참조하면, 일 실시예에 따른 심화 신경망(DNN)을 이용한 다채널 마이크 기반의 잔향시간 추정 방법의 적용한 실험에서 획득한 스펙트럼을 통해 특징 벡터 중 상호상관함수(GCC-PHAT)의 특징을 나타낸다. 이 때, T_{60} = 0.20, 0.35, 0.50, 0.65, 0.80, 1.10 s 에 대한 스펙트럼을 통해 상호상관함수(GCC-PHAT)의 특징을 확인할 수 있다.

[106] 다양한 화자, 문장, 그리고 환경을 내포하고 있는 충분한 분량의 음성 데이터베이스(Database, DB)를 이용해 특징 벡터를 도출하여, 감쇠율의 분산 값(NSV)은 4개의 채널에 대해 각각 127개의 특징 벡터 (4 x 127 = 50), 상호상관함수(GCC-PHAT)는 4개의 마이크에서 존재할 수 있는 마이크 쌍의 수(3!)에 각각 21개의 특징 벡터 (6 x 21 = 12)로 구성될 수 있다. 이에 따라 입력 벡터는 총 634개의 차원을 가질 수 있다.

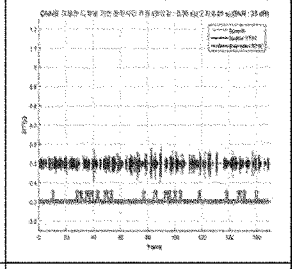
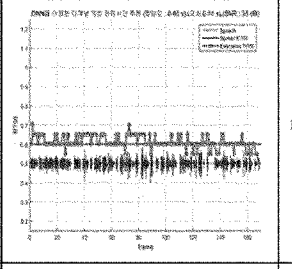
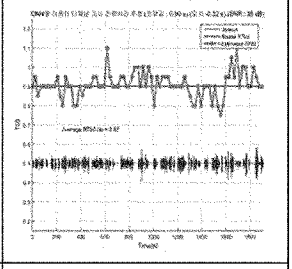
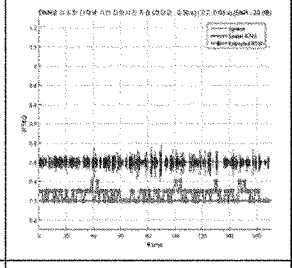
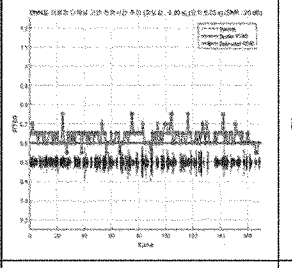
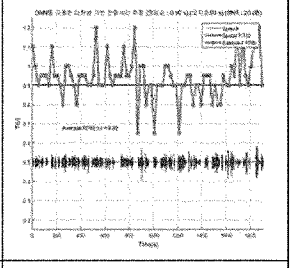
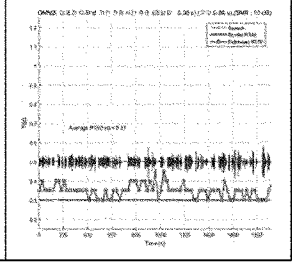
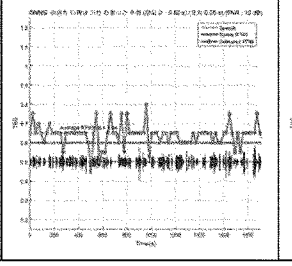
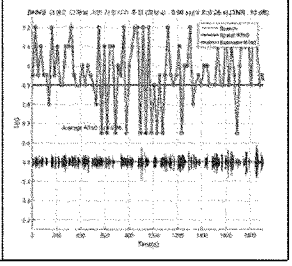
[107] 예를 들어, 심화 신경망(DNN)의 구성은 세 개의 은닉 계층과 1500개의 노드로 구성될 수 있으며, 각각의 은닉 계층은 50개의 에폭(epoch)으로 선행학습(pre-training) 과정을 거치고, 500개의 에폭(epoch)으로 미세 조정될 수 있다.

[108] 이와 같은 수치는 실험적인 과정을 통해 입력 데이터 자체에 과학습되는 것을 방지하는 동시에 잔향시간의 특징을 잘 나타낼 수 있는 범위 내에서 도출된 것이다.

[109]

[110] 표 1은 일 실시예에 따른 잔향환경에서 다채널 마이크를 이용한 잔향시간 추정의 예를 나타낸다.

[111] [표1]

잔향시간 / Noise SNR	RT60 = 0.30 s	RT60 = 0.60 s	RT60 = 0.90 s
Babble (35dB)			
Babble (20dB)			
Babble (10dB)			

[112] 표 1을 참조하면, 학습된 적이 없는 잔향을 포함한 음성의 잔향시간을 다양한 잡음 환경에서 추정할 수 있다.

[113] 이는 기존의 기법들보다 열악한 잔향환경 및 잡음환경에서 높은 성능을 보인다.

[114] 표 2는 일 실시예에 따른 잔향환경에서 다채널 마이크를 이용한 잔향시간 추정 성능과 기존 기술의 성능을 비교하여 나타낸다.

[115] [표2]

알고리즘 / 잡음 유형	ML			SDDSA-G			DNN-MCF		
	Bias	MSE	ρ	Bias	MSE	ρ	Bias	MSE	ρ
공항	-0.1680	0.0467	0.9277	0.1013	0.0327	0.8722	-0.1363	0.0314	0.9659
베블	-0.1688	0.0486	0.9218	0.0943	0.0319	0.8674	-0.1192	0.0261	0.9685
전시장	-0.1758	0.0524	0.9167	0.0966	0.0328	0.8660	-0.0851	0.0179	0.9703
공장	-0.2961	0.1286	0.8031	0.0078	0.0331	0.8057	0.0020	0.0112	0.9540
레스토랑	-0.1731	0.0512	0.9171	0.1095	0.0369	0.8561	-0.1195	0.0269	0.9669
거리	-0.1604	0.0441	0.9284	0.1229	0.0401	0.8548	-0.1170	0.0256	0.9670

[116] 표 2를 참조하면, 일 실시예에 따른 잔향환경에서 다채널 마이크를 이용한 잔향시간 추정 성능과 기존 기술의 성능을 비교하여 나타내는 것으로, 이 때 SNR은 35 dB이다.

[117] 표 3은 일 실시예에 따른 잔향환경에서 다채널 마이크를 이용한 잔향시간 추정 성능과 기존 기술의 성능을 비교하여 나타낸다.

[118] [표3]

알고리즘 / 잡음 유형	ML			SDDSA-G			DNN-MCF		
	Bias	MSE	ρ	Bias	MSE	ρ	Bias	MSE	ρ
공항	-0.3484	0.1948	0.6338	-0.1393	0.0444	0.8537	0.0872	0.0134	0.9707
베블	-0.3793	0.2258	0.5710	-0.1747	0.0602	0.8233	0.0893	0.0145	0.9700
전시장	-0.3829	0.2268	0.5724	-0.1763	0.0683	0.7700	0.2158	0.0811	0.8127
공장	-0.3115	0.1527	0.7408	-0.0994	0.0388	0.8282	0.1515	0.0351	0.9320
레스토랑	-0.3714	0.2184	0.5945	-0.1524	0.0498	0.8456	0.0594	0.0115	0.9680
거리	-0.2851	0.1306	0.7736	-0.1108	0.0397	0.8372	0.1022	0.0200	0.9525

[119] 표 3을 참조하면, 일 실시예에 따른 잔향환경에서 다채널 마이크를 이용한 잔향시간 추정 성능과 기존 기술의 성능을 비교하여 나타내는 것으로, 이 때 SNR은 10 dB이다.

[120] 예를 들어 30분 가량의 영어와 한글로 이루어진 잔향을 포함한 음성의 잔향시간을 다양한 잡음 환경에서 추정한다.

[121] 성능에 대한 평가는 각 조건에 따른 편향 값(bias), 평균제곱오차(MSE), 및 피어슨 상관관계 상수(ρ)를 도출하여 비교할 수 있다.

[122] 여기에서, 편향 값은 정답 값으로부터 추정 값이 가지는 오차의 평균을 나타내며, 평균제곱오차는 그 오차 제곱의 평균을 나타내고, 피어슨 상관관계 상수는 정답 값과 추정 값의 상관관계를 나타내는 값으로, 1에 가까울수록 좋은 성능을 의미할 수 있다.

[123] 실시예들에 따르면 기존의 추정 방식들에 비해 높은 정확도와 잡음에 강인한 특성을 가질 수 있는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법 및 장치를 제공할 수 있다.

[124] 기존의 다채널 기반의 추정 기술들은 소형화된 IoT 기기에 적용될 수 있는 구조가 아닌 반면, 실시예들에 따르면 마이크 간의 공간 에일리어싱(spatial aliasing)을 고려한 밀집된 구조의 다채널 마이크에 적용이 가능한 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법 및 장치를 제공할 수 있다.

[125] 이에 따라 모바일 기기, IoT 기기 등 최근 활용도가 높아진 다채널 마이크를 활용한 기기들에서 사용되는 잔향제거, 음성의 모델링, 음성인식 등의 중요 알고리즘에 필요한 정보를 제공할 수 있다. 또한, 잡음과 잔향으로 인한 열악한 환경에서의 음성 인식률을 높일 수 있는데 기여할 수 있다.

[126]

[127] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 컨트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPA(field programmable array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬

프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[128] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[129] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 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 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[130] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

[131] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

청구범위

[청구항 1] 다채널의 마이크를 통해 음성 신호를 입력 받는 단계;
입력 받은 상기 음성 신호를 이용하여 공간 정보가 포함된 특징 벡터를 도출하는 단계; 및
상기 특징 벡터를 심화 신경망(Deep Neural Network, DNN)에 적용하여 잔향의 정도를 추정하는 단계
를 포함하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법.

[청구항 2] 제1항에 있어서,
상기 다채널의 마이크를 통해 음성 신호를 입력 받는 단계는,
상기 다채널 마이크를 이용하여 입력되는 음성 신호 사이의 상대적인 공간 정보를 추정하는 것
을 특징으로 하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법.

[청구항 3] 제1항에 있어서,
상기 입력 받은 상기 음성 신호를 이용하여 공간 정보가 포함된 특징 벡터를 도출하는 단계는,
입력 받은 상기 음성 신호에서 숏타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 이용해 시간 및 주파수 정보를 도출하고, 도출된 상기 시간 및 주파수 정보를 이용해 주파수 대역별로 포락선의 분포를 도출하여 감쇠율의 분산 값(Negative-side variance, NSV)을 도출하는 단계
를 포함하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법.

[청구항 4] 제3항에 있어서,
상기 감쇠율의 분산 값을 도출하는 단계는,
상기 숏타임 푸리에 변환 영역에서 로그 에너지(log-energy) 포락선을 도출하는 단계;
상기 로그 에너지 포락선에서 최소 자승 선형 피팅(least squares linear fitting)을 이용하여 기울기를 도출하는 단계; 및
양의 기울기를 가지는 것을 제외하고 음의 기울기를 가지는 잔향시간을 추정하기 위한 감쇠율의 분산 값을 도출하는 단계
를 포함하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법.

[청구항 5] 제3항에 있어서,
상기 특징 벡터를 심화 신경망에 적용하여 잔향의 정도를 추정하는 단계는,
도출된 상기 감쇠율의 분산 값을 상기 심화 신경망의 입력으로 사용하여

잔향시간을 추정하는 것
을 특징으로 하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법.

[청구항 6] 제3항에 있어서,
상기 입력 받은 상기 음성 신호를 이용하여 공간 정보가 포함된 특징 벡터를 도출하는 단계는,
입력 받은 상기 음성 신호에서 두 개의 마이크 간의 상관관계를 나타내어 상기 다채널 마이크에서 음성의 방향의 추정하는 상호상관함수를 도출하는 단계
를 포함하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법.

[청구항 7] 제6항에 있어서,
상기 특징 벡터를 심화 신경망에 적용하여 잔향의 정도를 추정하는 단계는,
도출된 상기 감쇠율의 분산 값을 상기 심화 신경망의 입력으로 사용하여 잔향시간을 추정하고, 상기 다채널 마이크를 통해 도출되는 마이크 간의 공간 정보가 포함된 상기 상호상관함수의 값을 상기 심화 신경망의 입력으로 사용하여 잔향 시간을 추정하는 것
을 특징으로 하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법.

[청구항 8] 제1항에 있어서,
상기 심화 신경망은,
세 개의 은닉 계층을 포함하며, 각 상기 은닉 계층은 다수 개의 에폭(epoch)으로 선행학습(pre-training) 과정을 거치고 미세 조정되는 것
을 특징으로 하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 방법.

[청구항 9] 다채널의 마이크를 통해 음성 신호를 입력 받는 입력부;
입력 받은 상기 음성 신호를 이용하여 공간 정보가 포함된 특징 벡터를 도출하는 특징 벡터 추출부; 및
상기 특징 벡터를 심화 신경망(Deep Neural Network, DNN)에 적용하여 잔향의 정도를 추정하는 잔향 추정부
를 포함하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치.

[청구항 10] 제9항에 있어서,
상기 특징 벡터 추출부는,
입력 받은 상기 음성 신호에서 숏타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 이용해 시간 및 주파수 정보를 도출하고, 도출된 상기 시간 및 주파수 정보를 이용해 주파수 대역별로 포락선의 분포를

도출하여 감쇠율의 분산 값(Negative-side variance, NSV)을 도출하는 감쇠율의 분산 값 도출부
를 포함하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치.

[청구항 11] 제10항에 있어서,
상기 감쇠율의 분산 값 도출부는,
상기 숏타임 푸리에 변환 영역에서 로그 에너지(log-energy) 포락선을 도출하고, 상기 로그 에너지 포락선에서 최소 자승 선형 피팅(least squares linear fitting)을 이용하여 기울기를 도출하여, 양의 기울기를 가지는 것을 제외하고 음의 기울기를 가지는 잔향시간을 추정하기 위한 감쇠율의 분산 값을 도출하는 것
을 특징으로 하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치.

[청구항 12] 제10항에 있어서,
상기 잔향 추정부는,
도출된 상기 감쇠율의 분산 값을 상기 심화 신경망의 입력으로 사용하여 잔향시간을 추정하는 것
을 특징으로 하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치.

[청구항 13] 제10항에 있어서,
상기 특징 벡터 추출부는,
입력 받은 상기 음성 신호에서 두 개의 마이크 간의 상관관계를 나타내어 상기 다채널 마이크에서 음성의 방향의 추정하는 상호상관함수를 도출하는 상호상관함수 도출부
를 포함하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치.

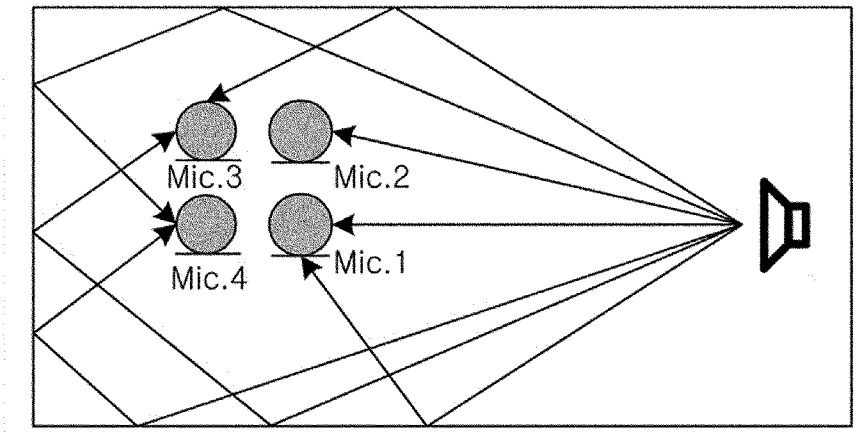
[청구항 14] 제13항에 있어서,
상기 잔향 추정부는,
도출된 상기 감쇠율의 분산 값을 상기 심화 신경망의 입력으로 사용하여 잔향시간을 추정하고, 상기 다채널 마이크를 통해 도출되는 마이크 간의 공간 정보가 포함된 상기 상호상관함수의 값을 상기 심화 신경망의 입력으로 사용하여 잔향 시간을 추정하는 것
을 특징으로 하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치.

[청구항 15] 제9항에 있어서,
상기 심화 신경망은,
세 개의 은닉 계층을 포함하며, 각 상기 은닉 계층은 다수 개의 에폭(epoch)으로 선행학습(pre-training) 과정을 거치고 미세 조정되는 것

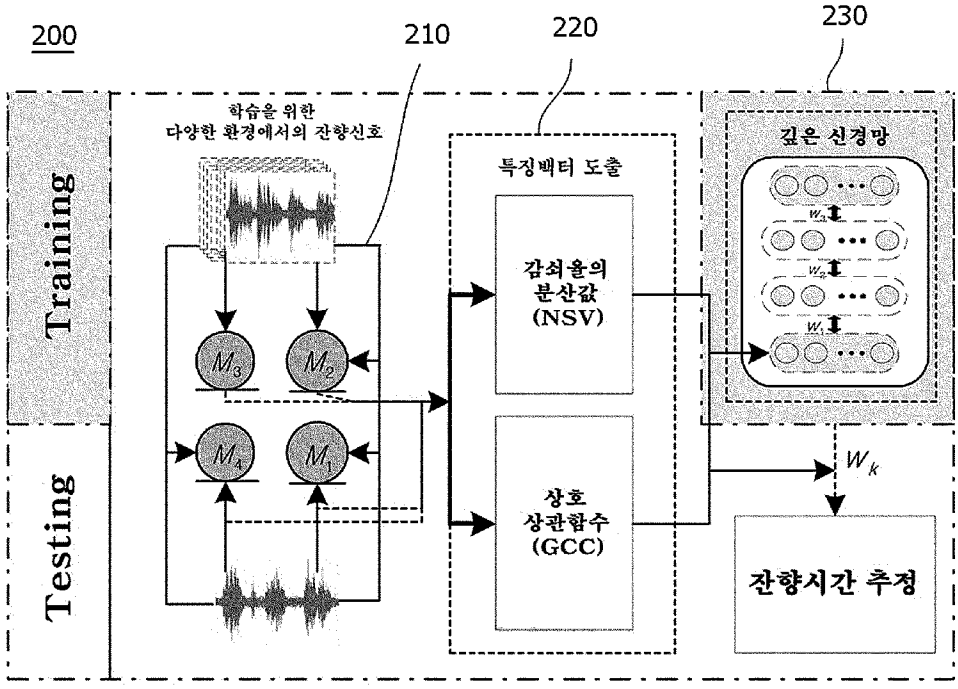
을 특징으로 하는 심화 신경망을 이용한 다채널 마이크 기반의 잔향시간 추정 장치.

WO 2018/111038 PCT/KR2017/014857

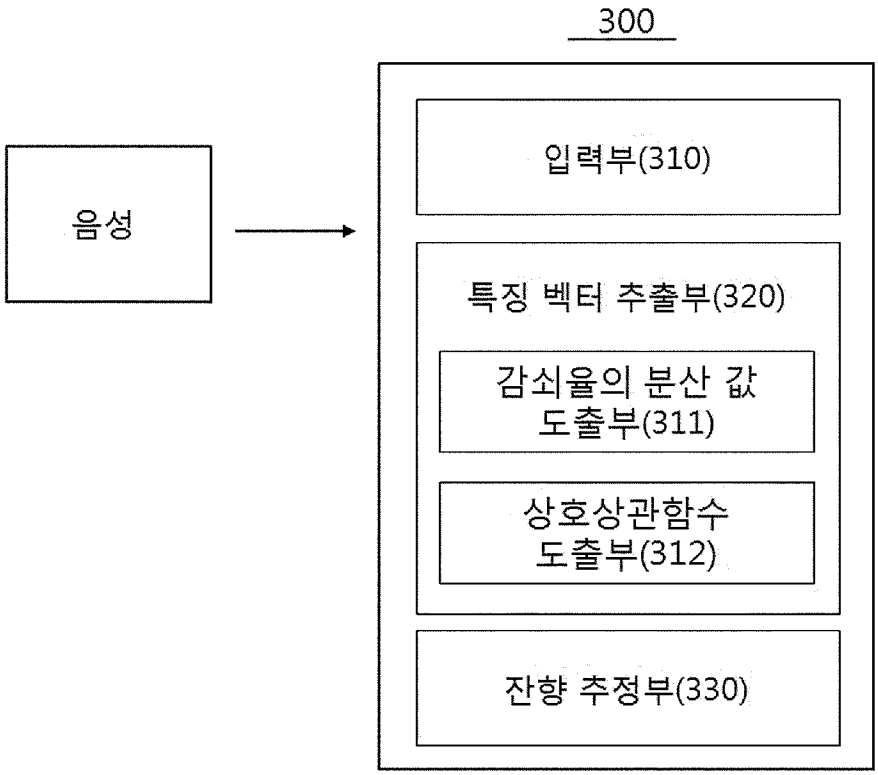
[도1]



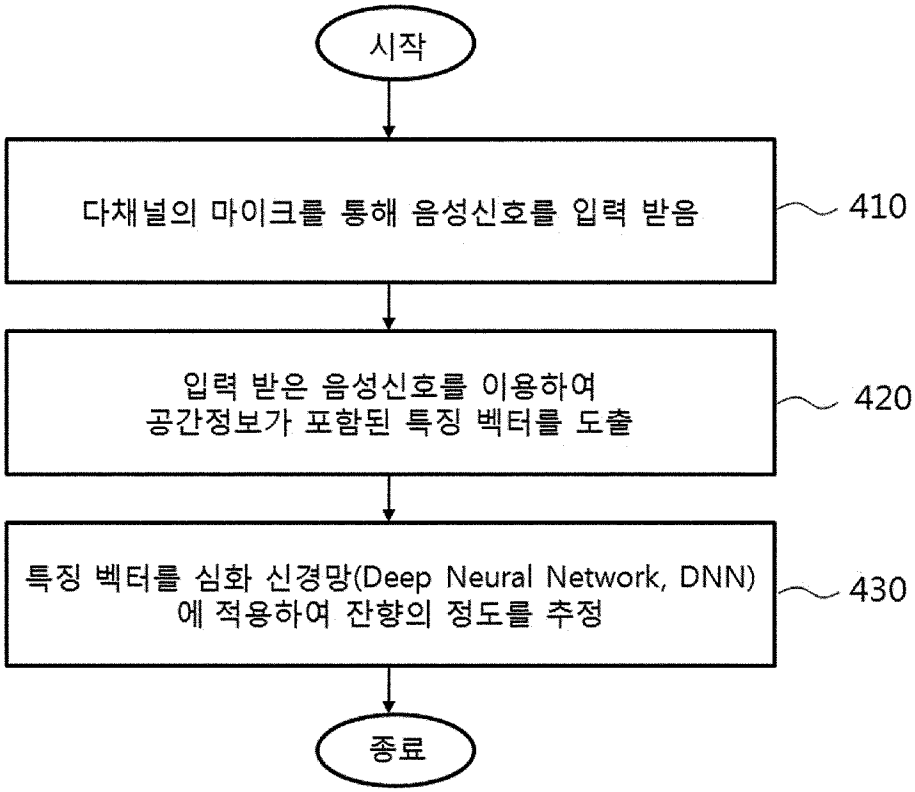
[도2]



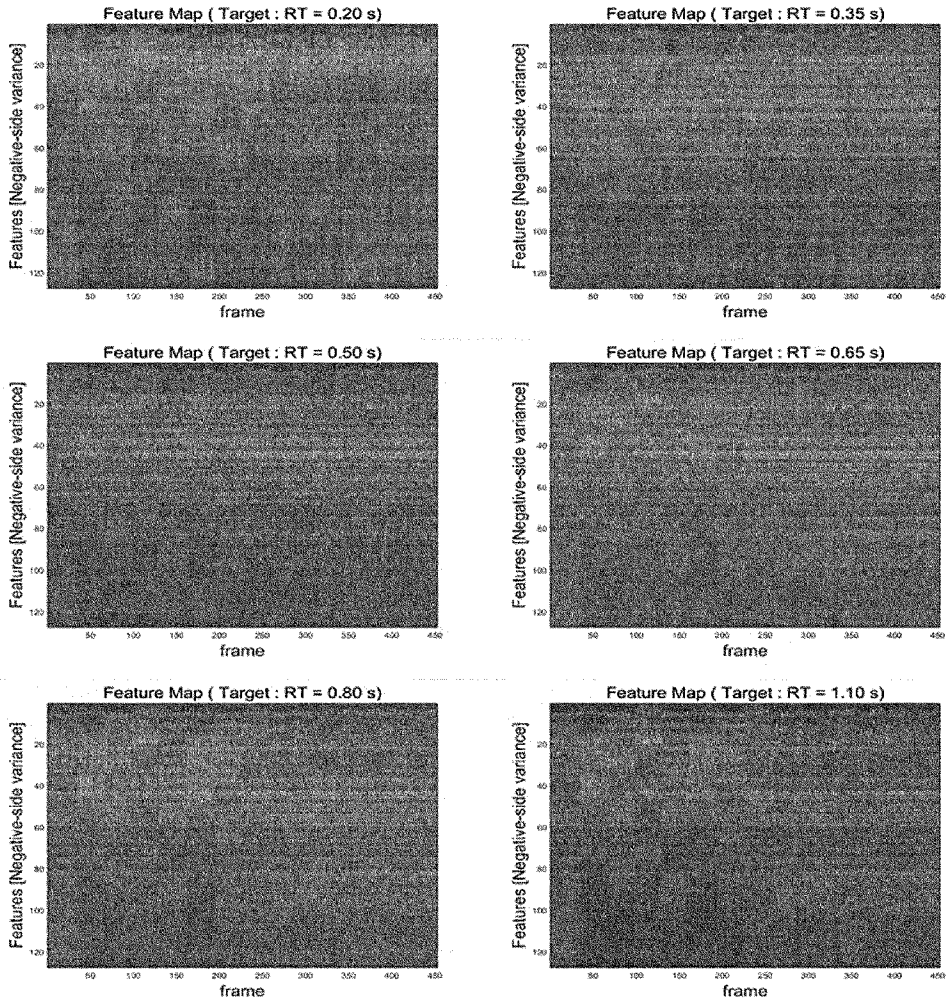
[도3]


300
음성
입력부(310)
특징 벡터 추출부(320)
감쇠율의 분산 값 도출부(311)
상호상관함수 도출부(312)
잔향 추정부(330)

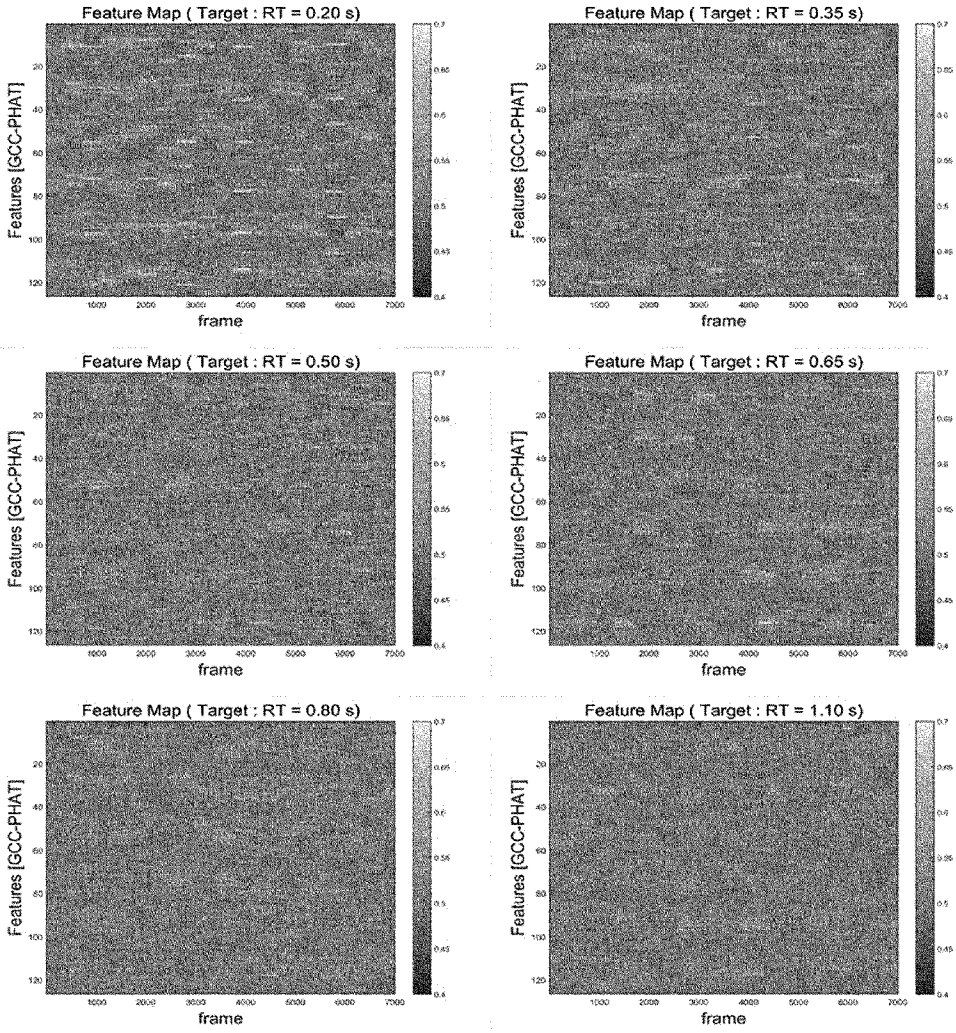
[도4]


시작
다채널의 마이크를 통해 음성신호를 입력 받음
410
입력 받은 음성신호를 이용하여 공간정보가 포함된 특징 벡터를 도출
420
특징 벡터를 심화 신경망(Deep Neural Network, DNN)에 적용하여 잔향의 정도를 추정
430
종료

[도5]


Feature Map (Target : RT = 0.20 s)
Feature Map (Target : RT = 0.35 s)
Feature Map (Target : RT = 0.50 s)
Feature Map (Target : RT = 0.65 s)
Feature Map (Target : RT = 0.80 s)
Feature Map (Target : RT = 1.10 s)
Features [Negative-side variance]
frame

[도6]


Feature Map (Target : RT = 0.20 s)
Feature Map (Target : RT = 0.35 s)
Feature Map (Target : RT = 0.50 s)
Feature Map (Target : RT = 0.65 s)
Feature Map (Target : RT = 0.80 s)
Feature Map (Target : RT = 1.10 s)
Features [GCC-PHAT]
frame

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2017/014857

A. CLASSIFICATION OF SUBJECT MATTER

G10L 21/0216(2013.01)i, G10L 25/30(2013.01)i, G10K 11/18(2006.01)i, G10K 15/08(2006.01)i, G10K 11/00(2006.01)i, G06N 3/02(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L 21/0216; G10L 25/30; G10L 25/78; G10K 11/18; G10K 15/08; G10K 11/00; G06N 3/02

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean Utility models and applications for Utility models: IPC as above
Japanese Utility models and applications for Utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & Keywords: multi-channel, microphone, voice signal, feature vector, DNN, reverberation duration, estimation

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	HIOKA, Yusuke et al., "ESTIMATING DIRECT-TO-REVERBERANT RATIO MAPPED FROM POWER SPECTRAL DENSITY USING DEEP NEURAL NETWORK.", 2016 IEEE International Conference., 20 March 2016., pages 26-30. See pages 26-27; and figure 1.	1-2,8-9,15
A		3-7,10-14
Y	GAUBITCH, N. D. et al., "PERFORMANCE COMPARISON OF ALGORITHMS FOR BLIND REVERBERATION TIME ESTIMATION FROM SPEECH.", International Workshop on Acoustic Signal Enhancement 2012., 04 September 2012., pages 1-4. See pages 1-2; and figure 1.	1-2,8-9,15
A	LEE, Myung In et al., "Intelligent Estimation Technology of Acoustic Environmental Parameter", The Korean Institute of Communications and Information Sciences (KICS), September 2016., pages 30-34. See pages 32-33; and figure 5.	1-15
A	KR 10-2016-0073874 A (SEOUL NATIONAL UNIVERSITY R&DB FOUNDATION et al.) 27 June 2016 See paragraph [0148]; and figure 4.	1-15

☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 MARCH 2018 (20.03.2018)	**21 MARCH 2018 (21.03.2018)**

Name and mailing address of the ISA/KR
Korean Intellectual Property Office
Government Complex-Daejeon, 189 Seonsa-ro, Daejeon 302-701, Republic of Korea
Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2017/014857

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LEE, Myungin et al., "BLIND ESTIMATION OF REVERBERATION TIME USING DEEP NEURAL NETWORK"., 2016 5th IEEE International Conference on Network Infrastructure and Digital Content, 23 September 2016., pages 1-4. See pages 1-3; and figures 1-2. ※ The above document is a document declaring exceptions to lack of novelty when an applicant files an earlier application whose priority is claimed.	1-15

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/KR2017/014857

Patent document cited in search report	Publication date	Patent family member	Publication date
KR 10-2016-0073874 A	27/06/2016	KR 10-1640188 B1	15/07/2016

국 제 조 사 보 고 서

국제출원번호
PCT/KR2017/014857

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G10L 21/0216(2013.01)i, G10L 25/30(2013.01)i, G10K 11/18(2006.01)i, G10K 15/08(2006.01)i, G10K 11/00(2006.01)i, G06N 3/02(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)
G10L 21/0216; G10L 25/30; G10L 25/78; G10K 11/18; G10K 15/08; G10K 11/00; G06N 3/02

조사된 기술분야에 속하는 최소문헌 이외의 문헌
한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC
일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))
eKOMPASS(특허청 내부 검색시스템) & 키워드: 다채널, 마이크, 음성 신호, 특징 벡터, DNN, 잔향시간, 추정

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	YUSUKE HIOKA 등. 'ESTIMATING DIRECT-TO-REVERBERANT RATIO MAPPED FROM POWER SPECTRAL DENSITY USING DEEP NEURAL NETWORK'. 2016 IEEE International Conference. 2016.03.20., 페이지 26-30. 페이지 26-27; 및 도면 1 참조.	1-2,8-9,15
A		3-7,10-14
Y	N. D. GAUBITCH 등. 'PERFORMANCE COMPARISON OF ALGORITHMS FOR BLIND REVERBERATION TIME ESTIMATION FROM SPEECH'. International Workshop on Acoustic Signal Enhancement 2012. 2012.09.04., 페이지 1-4. 페이지 1-2; 및 도면 1 참조.	1-2,8-9,15
A	이명인 등. '지능형 음향환경파라미터추정기술'. 한국통신학회, 2016.09. 페이지 30-34. 페이지 32-33; 및 도면 5 참조.	1-15
A	KR 10-2016-0073874 A (서울대학교산학협력단 등) 2016.06.27 단락 [0148]; 및 도면 4 참조.	1-15

☒ 추가 문헌이 C(계속)에 기재되어 있습니다.

☒ 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

"A" 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌

"E" 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌

"L" 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌

"O" 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

"P" 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

"T" 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌

"X" 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.

"Y" 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.

"&" 동일한 대응특허문헌에 속하는 문헌

국제조사의 실제 완료일	국제조사보고서 발송일
2018년 03월 20일 (20.03.2018)	2018년 03월 21일 (21.03.2018)

ISA/KR의 명칭 및 우편주소
대한민국 특허청
(35208) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)
팩스 번호 +82-42-481-8578

심사관
진상범
전화번호 +82-42-481-8398

국제조사보고서

국제출원번호
PCT/KR2017/014857

C (계속). 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	MYUNGIN LEE 등. 'BLIND ESTIMATION OF REVERBERATION TIME USING DEEP NEURAL NETWORK'. 2016 5th IEEE International Conference on Network Infrastructure and Digital Content, 2016.09.23., 페이지 1-4. 페이지 1-3; 및 도면 1-2 참조. ※ 위 문헌은 출원인이 우선권 주장의 기초가 된 선출원 출원시 공지예외선언한 문헌임.	1-15

국제조사보고서
대응특허에 관한 정보

국제출원번호
PCT/KR2017/014857

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2016-0073874 A	2016/06/27	KR 10-1640188 B1	2016/07/15