

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국



(10) 국제공개번호

(43) 국제공개일
2018 년 10 월 18 일 (18.10.2018) WIPO | PCT

WO 2018/190547 A1

(51) 국제특허분류:

G10L 21/02 (2006.01) G10L 19/08 (2006.01)
G10L 19/028 (2013.01) G10L 25/30 (2013.01)
G10L 19/03 (2013.01)

(21) 국제출원번호: PCT/KR2018/003840

(22) 국제출원일: 2018 년 4 월 2 일 (02.04.2018)

(25) 출원언어: 한국어

(26) 공개언어: 한국어

(30) 우선권정보:
10-2017-0048574 2017 년 4 월 14 일 (14.04.2017) 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 G동 2708호, Seoul (KR). 서혜지 (SEO, Hyeji); 13633 경기도 성남시 분당구 미금일로 57 605동 1301호, Gyeonggi-do (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.

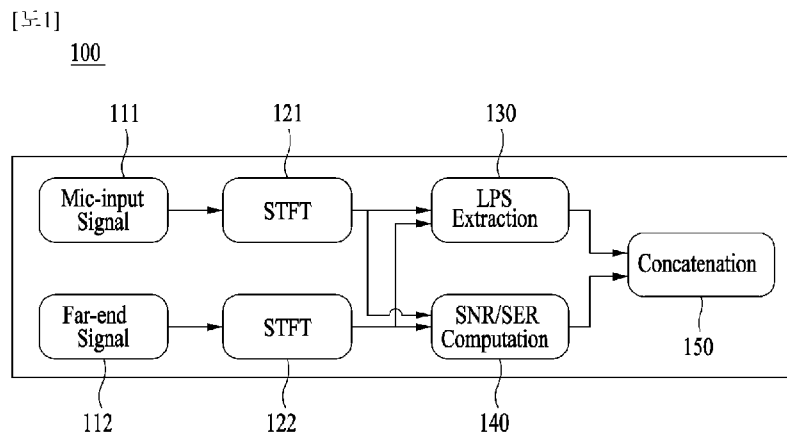
(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS,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

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

(54) Title: DEEP NEURAL NETWORK-BASED METHOD AND APPARATUS FOR COMBINED NOISE AND ECHO REMOVAL

(54) 발명의 명칭: 심화신경망 기반의 잡음 및 에코의 통합 제거 방법 및 장치



(57) Abstract: Disclosed is a deep neural network-based method and apparatus for combined noise and echo removal. The deep neural network-based method for combined noise and echo removal according to one embodiment of the present invention may comprise the steps of: extracting a feature vector from an audio signal that includes noise and echo; and acquiring a final audio signal from which both noise and echo have been removed, by using a combined noise and echo removal gain estimated by means of the feature vector and deep neural network (DNN) learning.

(57) 요약서: 심화신경망 기반의 잡음 및 에코의 통합 제거 방법 및 장치가 제시된다. 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 방법은, 잡음 및 에코가 포함된 음성 신호로부터 특징 벡터를 추출하는 단계; 및 상기 특징 벡터와 심화신경망(DNN)의 학습을 통해 추정된 잡음 및 에코의 통합 제거 이득을 이용하여 잡음 및 에코가 통합 제거된 최종 음성 신호를 획득하는 단계를 포함하여 이루어질 수 있다.



WO 2018/190547 A1

명세서

발명의 명칭: 심화신경망 기반의 잡음 및 에코의 통합 제거 방법 및 장치

기술분야

- [1] 아래의 실시예들은 음성 신호에 포함된 잡음 및 에코를 심화신경망 기반으로 통합하여 제거하는 기술에 관한 것이다.

배경기술

- [2] 음성 잡음 및 에코의 통합 제거 기술이란 음성 신호에 포함된 잡음 및 에코를 제거하는 기술이다. 일반적으로 잡음 제거기 및 에코 제거기를 독립적으로 설계한 후 직렬로 연결하여 잡음 및 에코 제거를 순차적으로 수행한다. 그러나 이러한 잡음 및 에코 제거기는 잡음 제거기 및 에코 제거기의 위치에 따라 성능의 차이가 크게 발생하게 된다. 예를 들어 잡음 제거기가 에코 제거기의 앞단에 위치할 경우 잡음 제거기의 비선형적인 연산으로 인하여 반향 제거기의 성능 저하가 발생하게 된다. 또한, 반향 제거기가 잡음 제거기의 앞단에 위치할 경우 잡음 제거기가 추정해야 할 잡음의 스펙트럼이 반향 제거 과정에서 왜곡이 생겨 잡음 추정의 성능이 저하되는 문제점이 발생하게 된다.
- [3] 이에 따라 잡음 및 에코를 한꺼번에 통합적으로 제거하는 잡음 및 에코의 통합 제거 기술이 사용될 수 있다. 종래에는 음성 신호와 잡음 및 에코 사이의 통계적 정보를 이용하는 통계 모델 기반의 잡음 및 에코의 통합 제거 기술이 주로 사용되었으나, 통계 모델 기반의 음성 향상 기술은 정상 잡음 환경과는 달리 비정상 잡음 환경에서 성능이 크게 저하되는 문제점을 가지고 있다. 예를 들어, 음성 인식에서 잡음이 존재하지 않은 깨끗한 신호를 이용하여 음성 인식 모델을 학습시킨 후 잡음이 존재하는 신호로 테스트를 수행할 경우 성능이 감소한다. 이러한 성능 감소를 해결하기 위해 잡음이 존재하는 음성을 이용하여 음성 인식 모델을 학습하는 기술이 제안되었으나, 학습된 잡음 환경에 최적화되어 학습된 잡음 환경에서 테스트하는 경우에는 우수한 성능을 보이나, 학습되지 않은 잡음 환경에서 테스트하는 경우에는 성능이 저하되는 문제점이 존재한다.
- [4] 최근 머신러닝 기법인 심화신경망(Deep Neural Network, DNN)이 다양한 음성 향상 및 음성 인식 연구에서 우수한 성능을 보이고 있다. 심화신경망은 다수의 은닉 층과 은닉 노드들을 통하여 입력 특징 벡터와 출력 특징 벡터 사이의 비선형적인 관계를 효과적으로 모델링하여 우수한 성능을 보인다.
- [5] 한국등록특허 10-0762596호는 음성 신호 전처리 시스템 및 음성 신호 특징 정보 추출 방법에 관한 것으로, 신경망 인식 방법을 이용하여 음성 신호를 전처리하는 기술을 기재하고 있다.

발명의 상세한 설명

기술적 과제

- [6] 실시예들은 심화신경망 기반의 잡음 및 에코의 통합 제거 방법 및 장치에 관하여 기술하며, 보다 구체적으로 심화신경망(DNN)을 이용하여 음성에 존재하는 잡음 및 에코를 통합적으로 제거하는 기술을 제공한다.
- [7] 실시예들은 원단 화자 신호와 마이크 입력 정보만으로는 심화신경망(DNN)이 학습이 잘되지 않는 문제점을 해결하기 위하여 음성 신호의 통계적 정보인 잡음 정보 및 에코 정보를 심화신경망(DNN)의 추가적 입력으로 사용하고, 음성 신호에서 잡음 및 에코를 통합 제거하기 위한 잡음 및 에코의 통합 제거 이득을 심화신경망(DNN)의 타겟(target) 특징 벡터로 하여 효과적으로 잡음 및 에코를 통합 제거하는 심화신경망 기반의 잡음 및 에코의 통합 제거 방법 및 장치를 제공하는데 있다.

과제 해결 수단

- [8] 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 방법은, 잡음 및 에코가 포함된 음성 신호로부터 특징 벡터를 추출하는 단계; 및 상기 특징 벡터와 심화신경망(DNN)의 학습을 통해 추정된 잡음 및 에코의 통합 제거 이득을 이용하여 잡음 및 에코가 통합 제거된 최종 음성 신호를 획득하는 단계를 포함한다.
- [9] 상기 심화신경망(DNN)을 학습시키는 단계를 더 포함하고, 상기 심화신경망(DNN)을 학습시키는 단계는, 학습 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호로부터 특징 벡터를 추출하는 단계; 및 상기 특징 벡터를 심화신경망(DNN)의 입력 특징 벡터로 이용하고, 학습된 심화신경망(DNN)의 모델인 가중치(weight) 파라미터와 바이어스(bias) 파라미터에 입력 특징 벡터를 통과시켜 잡음 및 에코의 통합 제거 이득을 추정하는 단계를 포함할 수 있다.
- [10] 상기 심화신경망(DNN)을 학습시키는 단계는, 학습 단계에서, 음성 신호의 통계적 정보인 사전 신호대에코비 및 사후 신호대에코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)를 추출된 상기 특징 벡터에 연결하여 상기 심화신경망(DNN)의 입력 특징 벡터로 이용하는 단계를 더 포함할 수 있다.
- [11] 상기 학습 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호로부터 특징 벡터를 추출하는 단계는, 상기 학습 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호를 대상으로 쇼타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 수행하여 시간 영역에서 주파수 영역의 신호로 변환하고, 변환된 주파수 영역의 신호의 로그 파워 스펙트럼(Log Power Spectrum, LPS)을 상기 특징 벡터로 추출할 수 있다.
- [12] 상기 심화신경망(DNN)을 학습시키는 단계는, 상기 심화신경망(DNN)의 회기 학습(regression)을 통하여 연속적인 최적 이득(optimal gain)을 추정하고, 평균제곱오차(Mean Squared Error, MSE)를 상기 심화신경망(DNN)의 목적

- 함수로 하여 타겟(target) 특징 벡터인 잡음 및 에코의 통합 제거 이득과 상기 심화신경망(DNN)을 통하여 추정된 상기 잡음 및 에코의 통합 제거 이득의 차이를 최소화하는 방향으로 심화신경망(DNN)의 학습이 진행될 수 있다.
- [13] 상기 잡음 및 에코가 포함된 음성 신호로부터 특징 벡터를 추출하는 단계는, 음성 향상 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호로부터 특징 벡터를 추출하고, 음성 향상 단계에서, 음성 신호의 통계적 정보인 사전 신호대예코비 및 사후 신호대예코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)를 추출된 상기 특징 벡터에 연결하여 상기 심화신경망(DNN)의 입력 특징 벡터로 이용하는 단계를 포함할 수 있다.
- [14] 상기 음성 향상 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호로부터 특징 벡터를 추출하는 단계는, 상기 음성 향상 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호를 대상으로 숏타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 수행하여 시간 영역에서 주파수 영역의 신호로 변환하고, 변환된 주파수 영역의 신호의 로그 파워 스펙트럼(Log Power Spectrum, LPS)을 상기 특징 벡터로 추출할 수 있다.
- [15] 상기 특징 벡터와 심화신경망(DNN)의 학습을 통해 추정된 잡음 및 에코의 통합 제거 이득을 이용하여 잡음 및 에코가 통합 제거된 최종 음성 신호를 획득하는 단계는, 상기 잡음 및 에코가 포함된 마이크 입력 신호의 로그 파워 스펙트럼(Log Power Spectrum, LPS)에 추정된 상기 잡음 및 에코의 통합 제거 이득을 곱하여 추정된 근단 화자 신호의 로그 파워 스펙트럼을 획득하는 단계; 및 상기 근단 화자 신호의 로그 파워 스펙트럼을 잡음이 존재하는 신호의 위상과 함께 인버스 숏타임 푸리에 변환(Inverse Short-Time Fourier Transform, ISTFT)하여 최종적으로 잡음 및 에코가 제거된 최종 근단 화자 신호의 파형을 획득하는 단계를 포함할 수 있다.
- [16] 다른 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 장치는, 잡음 및 에코가 포함된 음성 신호로부터 특징 벡터를 추출하는 특징 벡터 추출부; 및 상기 특징 벡터와 심화신경망(DNN)의 학습을 통해 추정된 잡음 및 에코의 통합 제거 이득을 이용하여 잡음 및 에코가 통합 제거된 최종 음성 신호를 획득하는 음성 신호 재구성부를 포함하여 이루어질 수 있다.
- [17] 상기 심화신경망(DNN)을 학습시키는 심화신경망 학습부를 더 포함하고, 상기 심화신경망 학습부는, 학습 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호로부터 특징 벡터를 추출하는 학습부의 특징 벡터 추출부; 및 상기 특징 벡터를 심화신경망(DNN)의 입력 특징 벡터로 이용하고, 학습된 심화신경망(DNN)의 모델인 가중치(weight) 파라미터와 바이어스(bias) 파라미터에 입력 특징 벡터를 통과시켜 잡음 및 에코의 통합 제거 이득을 추정하는 심화신경망(DNN) 모델링부를 포함할 수 있다.
- [18] 상기 학습부의 특징 벡터 추출부는, 학습 단계에서, 음성 신호의 통계적 정보인

사전 신호대예코비 및 사후 신호대예코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)를 추출된 상기 특징 벡터에 연결하여 상기 심화신경망(DNN)의 입력 특징 벡터로 이용하는 학습부의 잡음 정보 및 에코 정보부를 포함할 수 있다.

- [19] 상기 학습부의 특징 벡터 추출부는, 상기 학습 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호를 대상으로 숏타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 수행하여 시간 영역에서 주파수 영역의 신호로 변환하고, 변환된 주파수 영역의 신호의 로그 파워 스펙트럼(Log Power Spectrum, LPS)을 상기 특징 벡터로 추출하는 학습부의 스펙트럼 추출부를 포함할 수 있다.
- [20] 상기 심화신경망(DNN) 모델링부는, 상기 심화신경망(DNN)의 회기 학습(regression)을 통하여 연속적인 최적 이득(optimal gain)을 추정하고, 평균제곱오차(Mean Squared Error, MSE)를 상기 심화신경망(DNN)의 목적 함수로 하여 타겟(target) 특징 벡터인 잡음 및 에코의 통합 제거 이득과 상기 심화신경망(DNN)을 통하여 추정된 상기 잡음 및 에코의 통합 제거 이득의 차이를 최소화하는 방향으로 심화신경망(DNN)의 학습이 진행될 수 있다.
- [21] 상기 특징 벡터 추출부는, 상기 음성 향상 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호를 대상으로 숏타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 수행하여 시간 영역에서 주파수 영역의 신호로 변환하고, 변환된 주파수 영역의 신호의 로그 파워 스펙트럼(Log Power Spectrum, LPS)을 상기 특징 벡터로 추출하는 스펙트럼 추출부; 및 음성 향상 단계에서, 음성 신호의 통계적 정보인 사전 신호대예코비 및 사후 신호대예코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)를 추출된 상기 특징 벡터에 연결하여 상기 심화신경망(DNN)의 입력 특징 벡터로 이용하는 잡음 및 에코 정보부를 포함할 수 있다.
- [22] 상기 음성 신호 재구성부는, 상기 잡음 및 에코가 포함된 마이크 입력 신호의 로그 파워 스펙트럼(Log Power Spectrum, LPS)에 추정된 상기 잡음 및 에코의 통합 제거 이득을 곱하여 추정된 근단 화자 신호의 로그 파워 스펙트럼을 획득하고, 상기 근단 화자 신호의 로그 파워 스펙트럼을 잡음이 존재하는 신호의 위상과 함께 인버스 숏타임 푸리에 변환(Inverse Short-Time Fourier Transform, ISTFT)하여 최종적으로 잡음 및 에코가 제거된 최종 근단 화자 신호의 파형을 획득할 수 있다.

발명의 효과

- [23] 실시예들에 따르면 원단 화자 신호와 마이크 입력 정보만으로는 심화신경망(DNN)이 학습이 잘되지 않는 문제점을 해결하기 위하여 음성 신호의 통계적 정보인 잡음 정보 및 에코 정보를 심화신경망(DNN)의 추가적 입력으로

사용하고, 음성 신호에서 잡음 및 에코를 통합 제거하기 위한 잡음 및 에코의 통합 제거 이득을 심화신경망(DNN)의 타겟(target) 특징 벡터로 하여 효과적으로 잡음 및 에코를 통합 제거하는 심화신경망 기반의 잡음 및 에코의 통합 제거 방법 및 장치를 제공할 수 있다.

도면의 간단한 설명

- [24] 도 1은 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 장치의 특징 벡터 추출을 설명하기 위한 도면이다.
- [25] 도 2는 일 실시예에 따른 심화신경망(DNN) 기반의 잡음 및 에코의 통합 제거 장치의 최적 이득을 설명하기 위한 도면이다.
- [26] 도 3은 일 실시예에 따른 학습된 잡음 환경에서 잡음 및 에코 제거된 신호의 파형 비교를 나타낸다.
- [27] 도 4는 일 실시예에 따른 학습된 잡음 환경과 비교를 위해 학습되지 않은 잡음 환경에서 잡음 및 에코 제거된 신호의 파형 비교를 나타낸다.
- [28] 도 5는 일 실시예에 따른 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 장치를 설명하기 위한 도면이다.
- [29] 도 6은 일 실시예에 따른 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 방법을 나타내는 흐름도이다.
- [30] 도 7은 일 실시예에 따른 심화신경망의 학습 방법을 나타내는 흐름도이다.
- [31] 도 8은 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 장치를 나타내는 블록도이다.

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]
- [34] 음성 향상 기술은 마이크로폰으로 입력된 잡음 및 에코가 존재하는 음성 신호의 잡음 및 에코를 제거하여 깨끗한 음성을 추정하는 기술로, 음성 인식과 음성 통신과 같은 음성 어플리케이션에 필수적인 기술이다. 예를 들어 음성 인식에서 잡음 및 에코가 존재하지 않은 깨끗한 신호로 음성 인식 모델을 학습시킨 후 잡음이 존재하는 신호로 테스트를 할 경우 성능이 감소하게 된다. 이를 해결하기 위하여 음성 인식 수행 전에 잡음 및 에코를 제거하는 음성 향상 기술을 도입하여 음성 인식의 성능을 높일 수 있다. 또한, 음성 향상 기술은 음성 통신에서 잡음 및 에코를 제거하여 선명하고 명확하게 음성을 전달하여 통화 품질을 높이기 위해서도 사용될 수 있다.

- [35] 아래의 실시예들은 심화신경망(DNN)을 이용하여 음성에 존재하는 잡음 및 에코를 통합적으로 제거할 수 있다. 더 구체적으로 실시예들은 원단 화자 신호와 마이크 입력 정보만으로는 심화신경망(DNN)이 학습이 잘되지 않는 문제점을 해결하기 위하여 음성 신호의 통계적 정보인 사전 신호대예코비 및 사후 신호대예코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)를 심화신경망(DNN)의 추가적 입력으로 사용하고, 음성 신호에서 잡음 및 에코를 통합 제거하기 위한 잡음 및 에코의 통합 제거 이득을 심화신경망(DNN)의 타겟(target) 특징 벡터로 하여 효과적으로 잡음 및 에코를 통합 제거할 수 있다.
- [36]본 실시예들에서는 STFT(Short Time Fourier Transform) 및 ISTFT(Inverse Short Time Fourier Transform) 변환을 이용하는 경우를 예로 들어 설명하나, 이는 실시예에 해당되며, STFT, ISTFT 이외에 DFT(Discrete Fourier Transform), IDFT(Inverse Discrete Fourier Transform) 변환, FFT(Fast Fourier Transform), IFFT(Inverse Fast Fourier Transform) 변환 등이 이용될 수도 있다.
- [37]이하에서는 음성 신호에 포함된 잡음 및 에코를 심화신경망 기반으로 통합하여 제거하는 기술에 대해 보다 상세히 설명하기로 한다.
- [38]
- [39]도 1은 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 장치의 특징 벡터 추출을 설명하기 위한 도면이다.
- [40]도 1을 참조하면, 심화신경망 기반의 잡음 및 에코의 통합 제거를 위한 특징 벡터 추출 장치(100)는 잡음 및 에코의 통합 제거를 위한 특징 벡터 추출을 위해 마이크 입력 신호(111) 및 원단 화자 신호(far-end signal)(112)를 입력 받을 수 있다.
- [41]마이크 입력 신호(111)는 쏫타임 푸리에 변환(STFT)(121)을 통하여 변환되어 주파수 영역으로 표현될 수 있다. 이에 따라 입력 특징 벡터의 추출을 위하여 주파수 영역 신호의 로그 파워 스펙트럼(Log Power Spectrum, LPS)을 추출할 수 있다. 즉, 마이크 입력 신호(111)의 로그 파워 스펙트럼(log power spectrum)을 추출(130)하여 심화신경망(DNN)의 입력 특징 벡터로 사용할 수 있다.
- [42]그리고, 원단 화자 신호(112)는 쏫타임 푸리에 변환(STFT)(122)을 통하여 변환되어 주파수 영역으로 표현될 수 있다. 이에 따라 입력 특징 벡터의 추출을 위하여 주파수 영역 신호의 로그 파워 스펙트럼(log power spectrum)을 추출할 수 있다. 즉, 원단 화자 신호(112)의 로그 파워 스펙트럼(log power spectrum)을 추출(130)하여 심화신경망(DNN)의 입력 특징 벡터로 사용할 수 있다.
- [43]심화신경망(DNN)의 입력으로는 마이크 입력 신호의 로그 파워 스펙트럼과 원단 화자 신호(far-end signal)의 로그 파워 스펙트럼뿐만 아니라, 성능 향상을 위한 추가적 에코 정보(반향 정보) 및 잡음 정보(140)를 이어 붙여 사용(150)할 수 있다.
- [44]

- [45] 아래의 실시예들은 잡음 및 에코의 통합 제거 알고리즘에 DNN을 도입하여 최적의 잡음 및 에코의 통합 제거 이득을 추정할 수 있다. 또한, 심화신경망(DNN)의 입력으로 추가적 잡음 및 에코 정보인 사전 신호대예코비 및 사후 신호대예코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)를 주어 심화신경망(DNN) 기반의 잡음 및 에코 제거의 성능을 높일 수 있다.
- [46]먼저, 특징 벡터를 추출하는 방법에 대해 수학식 1 내지 6을 참조하여 보다 구체적으로 설명한다.
- [47]잡음 및 에코가 포함된 마이크 입력 신호(microphone signal power) $y(t)$ 는 근단 화자 신호(near-end speech power) $x(t)$ 에 잡음 신호(noise power) $d(t)$ 및 에코 신호(echo power) $e(t)$ 가 더해져 생성될 수 있으며, 다음 식과 같이 나타낼 수 있다. 여기에서 잡음 및 에코가 포함된 마이크 입력 신호는 잡음과 에코가 동시에 존재하는 마이크 입력 신호를 의미한다.
- [48][수학식 1]
- [49]
$$y(t) = x(t) + d(t) + e(t)$$
- [50]수학식 1과 같이 잡음 및 에코가 포함된 마이크 입력 신호 $y(t)$ 는 쏫타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 통하여 변환되어 시간 영역(time domain)에서 주파수 영역(frequency domain)으로 표현될 수 있으며, 다음 식과 같이 나타낼 수 있다.
- [51][수학식 2]
- [52]
$$Y(l, m) = X(l, m) + D(l, m) + E(l, m)$$
- [53]여기서, $Y(l, m)$ 은 잡음 및 에코가 포함된 마이크 입력 신호, $X(l, m)$ 은 근단 화자 신호, $D(l, m)$ 은 잡음 신호, $E(l, m)$ 은 에코 신호의 쏫타임 푸리에 변환(STFT) 계수를 나타낼 수 있다. 그리고 l 은 주파수 색인을 나타내고 m 은 프레임 색인을 나타낼 수 있다.
- [54]수학식 2에서 주파수 영역으로 표현된 잡음 및 에코가 포함된 마이크 입력 신호 $Y(l, m)$ 를 스펙트럼과 위상으로 나타낼 수 있다. 즉, 주파수 영역에서 잡음 및 에코가 포함된 마이크 입력 신호 $Y(l, m)$ 는 다음 식과 같이 스펙트럼과 위상으로 표현될 수 있다.
- [55][수학식 3]
- [56]
$$\begin{aligned} Y(l, m) &= \sum_{t=0}^{N-1} y(mL+t)w(t)e^{-j\frac{2\pi l}{N}t} \\ &= |Y(l, m)|e^{-j\theta Y(l, m)} \end{aligned}$$
- [57]여기서 θ 는 위상 값을 나타내며, L 은 프레임 스텝(frame step)을 나타낼 수 있다.
- [58]잡음 및 에코가 포함된 마이크 입력 신호의 위상은 음성 합성 단계에서 신호를 시간 영역으로 복원할 때 사용될 수 있다. 입력 특징 벡터의 추출을 위하여

주파수 영역 신호의 로그 파워 스펙트럼(Log Power Spectrum 또는 Log Power Spectra, LPS)을 추출할 수 있다. 그러면, 수학식 3의 음성 스펙트럼 특징 벡터에 로그 연산 값을 취하여 얻어진 로그 파워 스펙트럼(log power spectrum)은 입력 특징 벡터로 사용될 수 있다. 예컨대, 크기($|Y(l,m)|$)에 로그(log)를 취함에 따라 로그 파워 스펙트럼(log power spectrum)이 계산될 수 있다.

- [59] 심화신경망(DNN)의 입력으로는 마이크 입력 신호의 로그 파워 스펙트럼과 원단 화자 신호(far-end signal)의 로그 파워 스펙트럼뿐만 아니라, 성능 향상을 위한 추가적 에코 정보(반향 정보) 및 잡음 정보를 이어 붙여 사용할 수 있다. 추가적 에코 정보로는 사전 신호대에코비 및 사후 신호대에코비(a priori Signal-to-Echo Ratio and a posteriori Signal-to-Echo Ratio, a priori and a posteriori SERs)을 사용할 수 있고, 추가적 잡음 정보로는 사전 신호대잡음비 및 사후 신호대잡음비(a priori Signal-to-Noise Ratio and a posteriori Signal-to-Noise Ratio, a priori and a posteriori SNRs)를 사용할 수 있다. 추가적 에코 정보 및 잡음 정보를 나타내는 사전 신호대에코비(a priori SER), 사후 신호대에코비(a posteriori SER), 사전 신호대잡음비(a priori SNR), 및 사후 신호대잡음비(a posteriori SNR)는 순차적으로 다음 식과 같이 표현될 수 있다.

[60] [수학식 4]

[61]

$$\xi_e(i, k) \equiv \frac{|\hat{E}(i, k)|^2}{\lambda_E}$$

$$\gamma_e(i, k) \equiv \frac{|Y(i, k)|^2}{\lambda_E}$$

$$\xi_n(i, k) \equiv \frac{|\tilde{E}(i, k)|^2}{\lambda_D}$$

$$\gamma_n(i, k) \equiv \frac{|Y(i, k)|^2}{\lambda_D}$$

- [62] 여기서 $Y(i, k)$ 는 마이크 입력 신호, $\hat{E}(i, k)$ 는 음향학적 반향이 제거된 신호, $\tilde{E}(i, k)$ 는 배경 잡음이 제거된 신호, λ_E 는 반향 신호의 전력, 그리고 λ_D 는 잡음 신호의 전력을 나타낼 수 있다.

- [63] 표 1은 추가적 잡음 정보 및 에코 정보를 나타낸다.

[64] [표 1]

A posteriori SER	A priori SER
$\frac{\text{Microphone signal power}}{\text{Echo power}}$	$\frac{\text{Near - end speech power}}{\text{Echo power}}$
$\gamma(i, k) = \frac{ Y(i, k) ^2}{\lambda_d(i, k)}$	$\zeta(i, k) = \frac{\lambda_s(i, k)}{\lambda_d(i, k)}$
A posteriori SNR	A priori SNR
$\frac{\text{Microphone signal power}}{\text{Noise power}}$	$\frac{\text{Enhanced power}}{\text{Noise power}}$
$\gamma(i, k) = \frac{ Y(i, k) ^2}{\lambda_n(i, k)}$	$\zeta(i, k) = \frac{\lambda_s(i, k)}{\lambda_n(i, k)}$

[65] 표 1과 같이, 추가적 잡음 및 에코 정보인 사전 신호대에코비 및 사후 신호대에코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)를 나타낼 수 있다. 이러한 추가적 잡음 및 에코 정보인 사전 신호대에코비 및 사후 신호대에코비(a priori and a posteriori SER)와 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)는 기존의 통계적 기반의 통합 제거로부터 추정할 수 있다.

[66] 예컨대 사후 신호대잡음비(a posteriori SNR)는 음성이 존재하지 않는 구간에서 잡음 신호의 분산 값을 이용하여 업데이트 될 수 있고, 사전 신호대잡음비(a priori SNR)는 음성이 존재하는 구간에서 결정 지향(decision-directed) 방법을 통하여 구할 수 있다. 마찬가지로 사후 신호대에코비(a posteriori SER)는 음성이 존재하지 않는 구간에서 에코 신호의 분산 값을 이용하여 업데이트 될 수 있고, 사전 신호대에코비(a priori SER)는 음성이 존재하는 구간에서 결정 지향(decision-directed) 방법을 통하여 구할 수 있다.

[67] 이에 따라 마이크 입력 신호의 로그 파워 스펙트럼(log power spectrum), 원단 화자 신호의 로그 파워 스펙트럼(log power spectrum), 추가적 에코 정보 및 잡음 정보인 사전 신호대에코비 및 사후 신호대에코비(a priori and a posteriori SERs)와 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNRs)를 이어 붙여 총 6682 개를 심화신경망(DNN)의 입력 특징 벡터로 사용할 수 있다.

[68] 심화신경망(DNN)의 타겟(target) 특징 벡터로는 음향학적 반향(에코) 및 배경 잡음 통합 제거를 위한 최적 이득(optimal gain) 값을 추정할 수 있다. 최적 이득(optimal gain) G_{opt} 은 다음 식과 같이 표현될 수 있다.

[69] [수학식 5]

[70]

$$G_{opt}(i, k) = \max \left\{ \min \left(1, \frac{|S(i, k)|}{|Y(i, k)|} \right), G_{min} \right\}$$

- [71] 여기에서 $S(n, i)$ 는 깨끗한 근단 화자 신호의 쏘타임 푸리에 변환(STFT) 계수를 의미하고, $Y(n, i)$ 는 마이크 입력 신호의 쏘타임 푸리에 변환(STFT) 계수를 의미할 수 있다. 이 때, $G_{\min}$ 의 값은 10_{-4} 으로 설정될 수 있다.
- [72] 심화신경망(DNN) 기반의 잡음 및 에코의 통합 제거를 위해 먼저 심화신경망(DNN)의 학습을 수행할 수 있다.
- [73] 심화신경망(DNN) 기반의 잡음 및 에코의 통합 제거 방법은 마이크 입력 신호의 스펙트럼, 원단 화자 신호의 스펙트럼, 및 추가적 잡음 정보 및 에코 정보를 입력 특징 벡터로 하여 심화신경망(DNN)을 학습함으로써 잡음 및 에코의 통합 제거를 위한 최적의 이득(gain)을 추정할 수 있다.
- [74] 심화신경망(DNN)은 심화신경망(DNN)의 회기 학습(regression)을 통하여 연속적인 최적 이득(optimal gain) 값을 추정할 수 있다. 평균제곱오차(Mean Squared Error, MSE)를 심화신경망(DNN)의 목적 함수로 하여, 타겟(target) 특징 벡터인 잡음 및 에코의 통합 제거 이득과 심화신경망(DNN)을 통하여 추정된 잡음 및 에코의 통합 제거 이득의 차이를 최소화하는 방향으로 심화신경망(DNN)의 학습이 진행될 수 있다. 이 때, 평균제곱오차(MSE)는 다음 식과 같이 나타낼 수 있다.
- [75] [수학식 6]
- [76]
- $$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$
- [77] 여기에서, n 은 전체 학습 데이터 샘플의 개수(training data sample)를 나타내고, Y_i 는 심화신경망(DNN)의 타겟 특징 벡터인 잡음 및 에코가 존재하지 않는 깨끗한 음성 신호의 로그 파워 스펙트럼을 나타내고, $\hat{Y}_i$ 는 심화신경망(DNN)을 통하여 음성 향상되어 잡음 및 에코의 통합 제거된 신호의 로그 파워 스펙트럼을 나타낼 수 있다.
- [78] 심화신경망(DNN)은 다수의 은닉 층과 은닉 노드를 통하여 비선형적 특성의 모델링이 가능하다. 심화신경망(DNN)의 학습이 완료되면, 심화신경망(DNN)을 통하여 학습된 모델인 가중치(weight) 파라미터와 바이어스(bias) 파라미터를 잡음 및 에코의 통합 제거 단계에 사용할 수 있다.
- [79] 심화신경망(DNN)의 학습이 완료되면, 잡음 및 에코의 통합 제거 단계(테스트 단계)가 수행될 수 있다.
- [80] 잡음 및 에코의 통합 제거 단계에서, 잡음 및 에코를 통합 제거하기 위해 잡음 및 에코를 통합 제거할 신호의 마이크 입력 신호와 원단 화자 신호에서 입력 특징 벡터를 추출할 수 있다. 이 때, 마이크 입력 신호와 원단 화자 신호의 로그 파워 스펙트럼과 추가적 잡음 정보 및 에코 정보인 사전 신호대에코비 및 사후 신호대에코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)를 이어 붙여 입력 특징 벡터를 생성할 수 있다.

- [81] 심화신경망(DNN)의 학습 단계에서 학습된 심화신경망(DNN)의 모델인 가중치(weight) 파라미터와 바이어스(bias) 파라미터들에 입력 특징 벡터를 통과시킴으로써, 잡음 및 에코의 통합 제거 이득을 얻을 수 있다. 마이크 입력 신호의 스펙트럼에 잡음 및 에코의 통합 제거 이득을 곱하여 추정된 근단 화자 신호의 스펙트럼을 얻을 수 있다. 추정된 깨끗한 근단 화자 신호의 로그 파워 스펙트럼을 잡음이 존재하는 신호의 위상과 함께 인버스 쇼타임 푸리에 변환(Inverse Short-Time Fourier Transform, ISTFT)하여 최종 추정된 깨끗한 근단 화자 신호의 파형을 얻을 수 있다. 사람의 청각은 위상에 큰 영향을 받지 않으므로 위상은 잡음이 포함된 음성 신호의 위상을 그대로 사용할 수 있다.
- [82]
- [83] 도 2는 일 실시예에 따른 심화신경망(DNN) 기반의 잡음 및 에코의 통합 제거 장치의 최적 이득을 설명하기 위한 도면이다.
- [84] 아래에서는 도 2를 참조하여, 학습된 잡음 환경과 학습되지 않은 잡음 환경에서 신호 처리된 신호가 어떠한 차이가 있는지를 비교하여 설명하기로 한다. 여기에서는 시간 영역의 음성 파형과 객관적 음성 평가 방식인 PESQ(Perceptual Evaluation of Speech Quality) 및 에코가 제거된 정도를 측정하는 ERLE(Echo Return Loss Enhancement)에 의하여 성능을 비교할 수 있다.
- [85] 성능 비교 실험을 위해 4620개의 TIMIT 음성 데이터 베이스에 Room Impulse Response Generator를 이용하여 잔향 환경 음성을 생성할 수 있다. 시뮬레이션 공간을 설정하고, 잔향 시간을 0.1s, 0.2s, 0.3s로 하여 잔향 환경을 설정할 수 있다. 마이크 입력 신호는 원단 화자 신호와 근단 화자 신호를 결합하여 얻으며, 마이크 입력 신호에서 음향학적 반향 신호는 근단 화자 신호보다 평균 3.5 dB 낮게 설정될 수 있다. 잡음 환경 설정을 위하여 학습 데이터 셋에는 NOISEX-92의 Babble, Factory, HFchannel, F16의 4가지 잡음이 부가되고, 테스트 데이터 셋에는 Babble, Factory, HFchannel, F16, Destroyer-engine, Buccaneer, Pink, White의 8가지 잡음이 10 dB의 신호대잡음비(Signal-to-Noise Ratio, SNR)로 부가될 수 있다.
- [86] 마이크 입력 신호와 원단 화자 신호, 심화신경망(DNN)의 타겟(target) 특징 벡터인 깨끗한 근단 화자 신호를 프레임 사이즈 320 샘플(size 320 samples)(20ms), 지연 80 샘플(delay 80 samples)로 하여, 제로 패딩(zero padding) 후 512-point 쇼타임 푸리에 변환(STFT) 하여 주파수 영역의 신호를 얻을 수 있다. 여기에서 주파수 영역 신호의 로그 파워 스펙트럼(Log Power Spectrum, LPS)을 특징 벡터로 사용하였다. 즉, 마이크 입력 신호의 로그 파워 스펙트럼(log power spectrum)(211)과 원단 화자 신호의 로그 파워 스펙트럼(log power spectrum)(212)을 심화신경망(DNN)의 입력 특징 벡터로 사용할 수 있다.
- [87] 또한, 음성 통계 모델 기반의 잡음 제거 알고리즘 및 음성 통계 모델 기반의 반향 제거 단계에서 추가적인 에코 정보 및 잡음 정보(213)인 사전 신호대에코비 및 사후 신호대에코비(a priori and a posteriori SERs), 사전 신호대잡음비 및 사후

신호대잡음비(a priori and a posteriori SNRs)를 계산하였다. 이 때 마이크 입력 신호의 로그 파워 스펙트럼(log power spectrum)의 현재 프레임과 전, 후 5 프레임, 원단 화자 신호의 로그 파워 스펙트럼(log power spectrum)의 전, 후 5 프레임 및 추가적 잡음 및 에코 정보(213)인 사전 신호대에코비 및 사후 신호대에코비(a priori and a posteriori SERs), 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNRs)를 이어 붙여 총 6682 dimensions을 심화신경망(DNN)의 입력 특징 벡터로 사용하였다.

- [88] 또한, 심화신경망(DNN)의 타겟(target) 특징 벡터로는 음향학적 반향 및 배경 잡음 통합 제거를 위한 최적 이득(optimal gain)(230)으로 총 257 dimensions이 사용되었다. 심화신경망(DNN)(220)은 복수의 층으로 구성될 수 있으며, 예컨대 3개의 층(222, 223, 224)으로 각 층은 1024 은닉 유닛(hidden units)으로 구성되었으며, 심화신경망(DNN)(220)의 활성화 함수로는 시그모이드(sigmoid) 활성화 함수가 사용되었다. 심화신경망(DNN)(220)의 회기 학습은 MMSE(Minimum Mean Squared Error)를 목적함수로 mini-batch의 크기를 100으로 하여 SGD(Stochastic Gradient Descent)를 통한 BP(Back-Propagation)로 학습이 진행되었다. 모든 학습은 epoch 50까지 진행되었으며 epoch 20까지는 일정한 학습율(learning rate)로 진행시킨 후, epoch 21부터 0.9의 비율로 학습율을 감소시켰다. 잡음 및 에코의 통합 제거 알고리즘에 심화신경망(DNN)을 도입하여 최적의 잡음 및 에코의 통합 제거 이득을 추정할 수 있다.
- [89] 실험 결과, 실시예에 따른 심화신경망(DNN) 기반의 잡음 및 에코의 통합 제거 기술은 다수의 은닉 층과 은닉 노드들을 통하여 고차원 벡터의 비선형적인 성질의 모델링에 강인한 심화신경망(DNN)을 통하여 학습된 잡음 환경 및 학습되지 않은 잡음 환경에서 모두 기존의 통계모델 기반의 잡음 및 에코의 통합 제거 기술보다 우수한 성능을 보였다.
- [90]
- [91] 도 3은 일 실시예에 따른 학습된 잡음 환경에서 잡음 및 에코 제거된 신호의 파형 비교를 나타낸다. 그리고 도 4는 일 실시예에 따른 학습된 잡음 환경과 비교를 위해 학습되지 않은 잡음 환경에서 잡음 및 에코 제거된 신호의 파형 비교를 나타낸다.
- [92] 도 3은 학습된 잡음 환경인 factory 10 dB (RT60: 0.1s)에서 잡음 및 에코 제거된 신호의 파형 비교를 나타내고, 도 4는 학습되지 않은 잡음 환경인 machinegun 10 dB (RT60: 0.1s)에서 잡음 및 에코 제거된 신호의 파형 비교를 나타낸다. 도 3과 도 4는 학습된 잡음 환경인 factory 10 dB와 학습되지 않은 잡음 환경인 machinegun 10 dB (RT60: 0.1s)에서 잡음 및 에코 제거된 신호의 파형을 보여준다.
- [93] 실시예에 따른 심화신경망(DNN) 기반의 잡음 및 에코의 통합 제거 방법은 기존의 통계 모델보다 원단 화자 구간에서 더 높은 에코 제거 성능을 보였으며, 근단 화자 구간에서 근단 화자 신호가 더 잘 보존되는 것을 확인할 수 있다.

[94] 표 2는 원단 화자 구간에서의 ERLE 비교를 나타낸다.

[95] [표2]

		EURASIP 2012 ver.	DNN ver.
Seen noise	Babble	18.0	46.7
	Factory	18.1	46.3
	Hfchannel	24.3	47.0
	F16	19.9	47.0
Unseen noise	Volvo	12.6	46.7
	Leopard	16.1	47.0
	Buccaneer	20.1	46.5
	Destroyer engine	22.5	46.7
	White	25.6	41.1
	Machine gun	18.4	39.6
	Pink	19.4	45.6

[96] 그리고 표 3은 근단 화자 구간에서의 PESQ 비교를 나타낸다.

[97] [표3]

		Mic input	EURASIP 2012 ver.	DNN ver.
Seen noise	Babble	1.87	2.08	2.26
	Factory	1.68	1.91	2.27
	Hfchannel	1.47	1.95	2.32
	F16	1.79	1.98	2.31
Unseen noise	Volvo	2.94	3.13	3.17
	Leopard	2.52	2.71	2.74
	Buccaneer	1.67	2.04	2.30
	Destroyer engine	1.77	2.03	2.29
	White	1.78	2.33	2.43
	Machine gun	2.52	2.54	2.68
	Pink	1.70	2.08	2.39

[98] 표 2와 표 3에 따르면 원단 화자 구간에서 에코가 제거된 정도를 측정하는 ERLE 값과 근단 화자 구간에서의 음질 측정 기준인 PESQ를 비교할 수 있다. 심화신경망(DNN) 기반의 잡음 및 에코 제거 기술이 기존의 통계 모델 기반보다 높은 에코 제거 성능을 보였으며, 근단 화자 신호의 음질이 더 잘 보존되어 높은 PESQ 결과를 보이는 것을 확인할 수 있다.

- [99] 최근 다양한 IoT 기기에서 심화신경망(DNN) 기반의 음성 인식을 수행하고 있다. 음성 인식 수행 전 잡음 및 에코를 제거하여 음성 인식의 성능을 높이기 위하여 심화신경망(DNN) 기반의 잡음 및 에코 제거를 수행하여 음성 인식의 성능을 높일 뿐만 아니라 다양한 음성 통신 기술에 적용되어 음질을 향상시킬 수 있을 것이다.
- [100]
- [101] 도 5는 일 실시예에 따른 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 장치를 설명하기 위한 도면이다.
- [102] 도 5를 참조하면, 본 실시예에서 심화신경망(DNN)을 기반으로 하는 잡음 및 에코의 통합 제거 기술은 학습 단계(510)와 음성 향상 단계(520)로 구분될 수 있다.
- [103] 먼저, 학습 단계(510)에서는 잡음과 에코가 존재하는 마이크 입력 신호(511) 및 원단 화자 신호(512)의 학습 데이터 셋을 통하여 특징 벡터를 추출할 수 있다(513, 514). 여기에서, 특징 벡터는 마이크 입력 신호(511) 및 원단 화자 신호(512)의 로그 파워 스펙트럼(Log Power Spectrum, LPS)과 이들의 통계적 정보인 SER 및 SNR이 사용될 수 있다. 이를 바탕으로 심화신경망(DNN)을 통하여 잡음 및 에코의 통합 제거 이득을 추정하도록 학습이 진행될 수 있다(515).
- [104] 다음으로, 음성 향상 단계(520)가 진행될 수 있다. 음성 향상 단계(520)도 학습 단계(510)와 마찬가지로 특징 벡터의 추출로부터 시작할 수 있다(523, 524, 525). 잡음과 에코가 포함된 마이크 입력 신호(521) 및 원단 화자 신호(522)의 학습 데이터 셋을 통하여 로그 파워 스펙트럼(Log Power Spectrum, LPS)과 이들의 통계적 정보인 SER 및 SNR을 추출한 뒤, 이를 이어 붙여 입력 특징 벡터로 사용할 수 있다. 또한, 잡음과 에코가 포함된 마이크 입력 신호(521) 및 원단 화자 신호(522)의 학습 데이터 셋을 통하여 위상을 추출할 수 있다.
- [105] 입력 특징 벡터를 심화신경망(DNN)의 학습 과정에서 학습된 모델인 가중치(weight) 파라미터와 바이어스(bias) 파라미터에 통과시켜 최종적으로 추정된 잡음 및 에코의 통합 제거 이득을 얻을 수 있다(526).
- [106] 그리고 마이크 입력 신호의 로그 파워 스펙트럼에 추정된 잡음 및 에코의 통합 제거 이득을 곱하여 추정된 깨끗한 근단 화자 신호의 로그 파워 스펙트럼을 마이크 입력 신호의 위상과 함께 음성을 복원하는 과정(527)을 거쳐 최종적으로 잡음 및 에코가 제거된 깨끗한 근단 화자 신호의 파형을 얻을 수 있다(528).
- [107] 아래에서 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 장치에 대해 하나의 예를 들어 더 구체적으로 설명하기로 한다.
- [108]
- [109] 도 6은 일 실시예에 따른 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 방법을 나타내는 흐름도이다. 그리고 도 7은 일 실시예에 따른 심화신경망의 학습 방법을 나타내는 흐름도이다.

- [110] 도 6 및 도 7을 참조하면, 일 실시예에 따른 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 방법은 잡음 및 에코가 포함된 음성 신호로부터 특징 벡터를 추출하는 단계(620), 및 특징 벡터와 심화신경망(DNN)의 학습을 통해 추정된 잡음 및 에코의 통합 제거 이득을 이용하여 잡음 및 에코가 통합 제거된 최종 음성 신호를 획득하는 단계(630)를 포함하여 이루어질 수 있다.
- [111] 심화신경망(DNN)을 학습시키는 단계(610)를 더 포함하고, 심화신경망(DNN)을 학습시키는 단계(610)는, 학습 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호로부터 특징 벡터를 추출하는 단계(611), 및 특징 벡터를 심화신경망(DNN)의 입력 특징 벡터로 이용하고, 학습된 심화신경망(DNN)의 모델인 가중치(weight) 파라미터와 바이어스(bias) 파라미터에 입력 특징 벡터를 통과시켜 잡음 및 에코의 통합 제거 이득을 추정하는 단계(613)를 포함할 수 있다.
- [112] 심화신경망(DNN)을 학습시키는 단계는, 학습 단계에서, 음성 신호의 통계적 정보인 사전 신호대에코비 및 사후 신호대에코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)를 추출된 특징 벡터에 연결하여 심화신경망(DNN)의 입력 특징 벡터로 이용하는 단계(612)를 더 포함할 수 있다.
- [113] 실시예들에 따르면 원단 화자 신호와 마이크 입력 정보만으로는 심화신경망(DNN)이 학습이 잘되지 않는 문제점을 해결하기 위하여 음성 신호의 통계적 정보인 잡음 정보 및 에코 정보를 심화신경망(DNN)의 추가적 입력으로 사용하고, 음성 신호에서 잡음 및 에코를 통합 제거하기 위한 잡음 및 에코의 통합 제거 이득을 심화신경망(DNN)의 타겟(target) 특징 벡터로 하여 효과적으로 잡음 및 에코를 통합 제거할 수 있다.
- [114]
- [115] 도 8은 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 장치를 나타내는 블록도이다.
- [116] 도 8을 참조하면, 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 장치를 이용하여 도 6 및 도 7의 일 실시예에 따른 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 방법을 보다 구체적으로 설명할 수 있다.
- [117] 일 실시예에 따른 심화신경망 기반의 잡음 및 에코의 통합 제거 장치는 심화신경망(DNN) 학습부(810), 특징 벡터 추출부(820), 그리고 음성 신호 재구성부(830)를 포함하여 이루어질 수 있다. 여기에서 심화신경망(DNN) 학습부(810)는 학습부의 특징 벡터 추출부(811) 및 심화신경망(DNN) 모델링부(812)를 포함하여 이루어질 수 있고, 학습부의 특징 벡터 추출부(811)는 학습부의 스펙트럼 추출부와 학습부의 잡음 및 에코 정보부를 포함하여 이루어질 수 있다. 또한, 특징 벡터 추출부(820)는 스펙트럼 추출부와 잡음 및

에코 정보부를 포함하여 이루어질 수 있다.

- [118] 단계(610)에서, 심화신경망(DNN) 학습부(810)는 심화신경망(DNN)을 학습시킬 수 있다.
- [119] 더 구체적으로 단계(611)에 있어서, 학습부의 특징 벡터 추출부(811)는 학습 단계에서 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호로부터 특징 벡터를 추출할 수 있다.
- [120] 그리고 단계(612)에 있어서, 학습부의 특징 벡터 추출부(811)는 학습 단계에서 음성 신호의 통계적 정보인 사전 신호대에코비 및 사후 신호대에코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)를 추출된 특징 벡터에 연결하여 심화신경망(DNN)의 입력 특징 벡터로 이용할 수 있다.
- [121] 여기에서 학습부의 특징 벡터 추출부(811)는 학습부의 스펙트럼 추출부와 학습부의 잡음 및 에코 정보부를 포함하여 이루어질 수 있다.
- [122] 학습부의 스펙트럼 추출부는 학습 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호를 대상으로 숏타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 수행하여 시간 영역에서 주파수 영역의 신호로 변환하고, 변환된 주파수 영역의 신호의 로그 파워 스펙트럼(Log Power Spectrum, LPS)을 특징 벡터로 추출할 수 있다.
- [123] 그리고 학습부의 잡음 및 에코 정보부는 학습 단계에서 음성 신호의 통계적 정보인 사전 신호대에코비 및 사후 신호대에코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)를 추출된 특징 벡터에 연결하여 심화신경망(DNN)의 입력 특징 벡터로 이용할 수 있다.
- [124] 단계(613)에서, 심화신경망(DNN) 모델링부(812)는 특징 벡터를 심화신경망(DNN)의 입력 특징 벡터로 이용하고, 학습된 심화신경망(DNN)의 모델인 가중치(weight) 파라미터와 바이어스(bias) 파라미터에 입력 특징 벡터를 통과시켜 잡음 및 에코의 통합 제거 이득을 추정할 수 있다.
- [125] 이러한 심화신경망(DNN) 모델링부는 심화신경망(DNN)의 회기 학습(regression)을 통하여 연속적인 최적 이득(optimal gain)을 추정하고, 평균제곱오차(Mean Squared Error, MSE)를 심화신경망(DNN)의 목적 함수로 하여 타겟(target) 특징 벡터인 잡음 및 에코의 통합 제거 이득과 심화신경망(DNN)을 통하여 추정된 잡음 및 에코의 통합 제거 이득의 차이를 최소화하는 방향으로 심화신경망(DNN)의 학습이 진행될 수 있다.
- [126] 단계(620)에서, 특징 벡터 추출부(820)는 잡음 및 에코가 포함된 음성 신호로부터 특징 벡터를 추출할 수 있다.
- [127] 여기에서 특징 벡터 추출부(820)는 스펙트럼 추출부와 잡음 및 에코 정보부를 포함하여 이루어질 수 있다.
- [128] 스펙트럼 추출부는 음성 향상 단계에서 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호를 대상으로 숏타임 푸리에 변환(Short-Time Fourier

Transform, STFT)을 수행하여 시간 영역에서 주파수 영역의 신호로 변환하고, 변환된 주파수 영역의 신호의 로그 파워 스펙트럼(Log Power Spectrum, LPS)을 특징 벡터로 추출할 수 있다.

- [129] 그리고 잡음 및 에코 정보부는 음성 향상 단계에서 음성 신호의 통계적 정보인 사전 신호대예코비 및 사후 신호대예코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)를 추출된 특징 벡터에 연결하여 심화신경망(DNN)의 입력 특징 벡터로 이용할 수 있다.
- [130] 단계(630)에서, 음성 신호 재구성부(830)는 특징 벡터와 심화신경망(DNN)의 학습을 통해 추정된 잡음 및 에코의 통합 제거 이득을 이용하여 잡음 및 에코가 통합 제거된 최종 음성 신호를 획득할 수 있다.
- [131] 이러한 음성 신호 재구성부(830)는 잡음 및 에코가 포함된 마이크 입력 신호의 로그 파워 스펙트럼(Log Power Spectrum, LPS)에 추정된 잡음 및 에코의 통합 제거 이득을 곱하여 추정된 근단 화자 신호의 로그 파워 스펙트럼을 획득하고, 근단 화자 신호의 로그 파워 스펙트럼을 잡음이 존재하는 신호의 위상과 함께 인버스 쇼트타임 푸리에 변환(Inverse Short-Time Fourier Transform, ISTFT)하여 최종적으로 잡음 및 에코가 제거된 최종 근단 화자 신호의 파형을 획득할 수 있다. 즉, 잡음 및 에코가 통합 제거된 최종 음성 신호를 획득할 수 있다.
- [132] 이상과 같이 실시예들은 음성 향상 기술로 음성 인식과 음성 통신 기술을 수행하기 이전에 잡음 및 에코를 제거하여 보다 우수한 성능을 도출할 수 있다. 또한 휴대폰 단말기나 카카오톡의 보이스톡 등에서 음성 통화 품질을 높이기 위해 적용될 수 있다.
- [133] 이러한 음성 향상 기술은 잡음 및 에코가 존재하는 환경에서 음성 인식의 성능을 향상시키는데 사용될 수 있다. 최근 다양한 사물인터넷(Internet of Things, IoT) 기기에서 음성 인식이 수행되는데 이는 조용한 환경에서만 수행되는 것이 아니라 주변 잡음이 존재하는 환경에서 수행될 수 있으며, IoT 기기의 스피커에서 소리가 나올 때 이 소리가 다시 들어가 에코를 발생시킬 수 있다. 따라서 음성 인식 수행 전 잡음 및 에코를 제거하여 IoT 기기에서 수행되는 음성 인식의 성능을 높일 수 있다. 또한, 본 실시예들은 우수한 품질의 음성 향상 신호를 제공하므로 다양한 음성 통신 기술에 적용되어 깨끗한 품질의 음성을 제공할 수 있다.
- [134]
- [135] 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 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.

- [136] 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.
- [137] 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.
- [138] 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.

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

청구범위

- [청구항 1] 잡음 및 에코가 포함된 음성 신호로부터 특징 벡터를 추출하는 단계; 및 상기 특징 벡터와 심화신경망(DNN)의 학습을 통해 추정된 잡음 및 에코의 통합 제거 이득을 이용하여 잡음 및 에코가 통합 제거된 최종 음성 신호를 획득하는 단계를 포함하는 심화신경망 기반의 잡음 및 에코의 통합 제거 방법.
- [청구항 2] 제1항에 있어서, 상기 심화신경망(DNN)을 학습시키는 단계를 더 포함하고, 상기 심화신경망(DNN)을 학습시키는 단계는, 학습 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호로부터 특징 벡터를 추출하는 단계; 및 상기 특징 벡터를 심화신경망(DNN)의 입력 특징 벡터로 이용하고, 학습된 심화신경망(DNN)의 모델인 가중치(weight) 파라미터와 바이어스(bias) 파라미터에 입력 특징 벡터를 통과시켜 잡음 및 에코의 통합 제거 이득을 추정하는 단계를 포함하는 심화신경망 기반의 잡음 및 에코의 통합 제거 방법.
- [청구항 3] 제2항에 있어서, 상기 심화신경망(DNN)을 학습시키는 단계는, 학습 단계에서, 음성 신호의 통계적 정보인 사전 신호대에코비 및 사후 신호대에코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)를 추출된 상기 특징 벡터에 연결하여 상기 심화신경망(DNN)의 입력 특징 벡터로 이용하는 단계를 더 포함하는 심화신경망 기반의 잡음 및 에코의 통합 제거 방법.
- [청구항 4] 제2항에 있어서, 상기 학습 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호로부터 특징 벡터를 추출하는 단계는, 상기 학습 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호를 대상으로 쇼타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 수행하여 시간 영역에서 주파수 영역의 신호로 변환하고, 변환된 주파수 영역의 신호의 로그 파워 스펙트럼(Log Power Spectrum, LPS)을 상기 특징 벡터로 추출하는 것을 특징으로 하는 심화신경망 기반의 잡음 및 에코의 통합 제거 방법.
- [청구항 5] 제2항에 있어서, 상기 심화신경망(DNN)을 학습시키는 단계는, 상기 심화신경망(DNN)의 회기 학습(regression)을 통하여 연속적인 최적 이득(optimal gain)을 추정하고, 평균제곱오차(Mean Squared Error, MSE)를 상기 심화신경망(DNN)의 목적 함수로 하여 타겟(target) 특징 벡터인 잡음

및 에코의 통합 제거 이득과 상기 심화신경망(DNN)을 통하여 추정된
상기 잡음 및 에코의 통합 제거 이득의 차이를 최소화하는 방향으로
심화신경망(DNN)의 학습이 진행되는 것

을 특징으로 하는 심화신경망 기반의 잡음 및 에코의 통합 제거 방법.

[청구항 6]

제1항에 있어서,

상기 잡음 및 에코가 포함된 음성 신호로부터 특징 벡터를 추출하는
단계는,

음성 향상 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단
화자 신호로부터 특징 벡터를 추출하는 단계; 및

음성 향상 단계에서, 음성 신호의 통계적 정보인 사전 신호대에코비 및
사후 신호대에코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및
사후 신호대잡음비(a priori and a posteriori SNR)를 추출된 상기 특징
벡터에 연결하여 상기 심화신경망(DNN)의 입력 특징 벡터로 이용하는
단계

를 포함하는 심화신경망 기반의 잡음 및 에코의 통합 제거 방법.

[청구항 7]

제6항에 있어서,

상기 음성 향상 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와
원단 화자 신호로부터 특징 벡터를 추출하는 단계는,

상기 음성 향상 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와
원단 화자 신호를 대상으로 숏타임 푸리에 변환(Short-Time Fourier
Transform, STFT)을 수행하여 시간 영역에서 주파수 영역의 신호로
변환하고, 변환된 주파수 영역의 신호의 로그 파워 스펙트럼(Log Power
Spectrum, LPS)을 상기 특징 벡터로 추출하는 것

을 특징으로 하는 심화신경망 기반의 잡음 및 에코의 통합 제거 방법.

[청구항 8]

제7항에 있어서,

상기 특징 벡터와 심화신경망(DNN)의 학습을 통해 추정된 잡음 및
에코의 통합 제거 이득을 이용하여 잡음 및 에코가 통합 제거된 최종 음성
신호를 획득하는 단계는,

상기 잡음 및 에코가 포함된 마이크 입력 신호의 로그 파워 스펙트럼(Log
Power Spectrum, LPS)에 추정된 상기 잡음 및 에코의 통합 제거 이득을
곱하여 근단 화자 신호의 로그 파워 스펙트럼을 획득하는 단계; 및
상기 근단 화자 신호의 로그 파워 스펙트럼을 잡음 및 에코가 존재하는
신호의 위상과 함께 인버스 숏타임 푸리에 변환(Inverse Short-Time Fourier
Transform, ISTFT)하여 최종적으로 잡음 및 에코가 제거된 최종 근단 화자
신호의 파형을 획득하는 단계

를 포함하는 심화신경망 기반의 잡음 및 에코의 통합 제거 방법.

[청구항 9]

잡음 및 에코가 포함된 음성 신호로부터 특징 벡터를 추출하는 특징 벡터
추출부; 및

상기 특징 벡터와 심화신경망(DNN)의 학습을 통해 추정된 잡음 및 에코의 통합 제거 이득을 이용하여 잡음 및 에코가 통합 제거된 최종 음성 신호를 획득하는 음성 신호 재구성부를 포함하는 심화신경망 기반의 잡음 및 에코의 통합 제거 장치.

[청구항 10] 제9항에 있어서,
상기 심화신경망(DNN)을 학습시키는 심화신경망 학습부를 더 포함하고,
상기 심화신경망 학습부는,
학습 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호로부터 특징 벡터를 추출하는 학습부의 특징 벡터 추출부; 및
상기 특징 벡터를 심화신경망(DNN)의 입력 특징 벡터로 이용하고,
학습된 심화신경망(DNN)의 모델인 가중치(weight) 파라미터와 바이어스(bias) 파라미터에 입력 특징 벡터를 통과시켜 잡음 및 에코의 통합 제거 이득을 추정하는 심화신경망(DNN) 모델링부를 포함하는 심화신경망 기반의 잡음 및 에코의 통합 제거 장치.

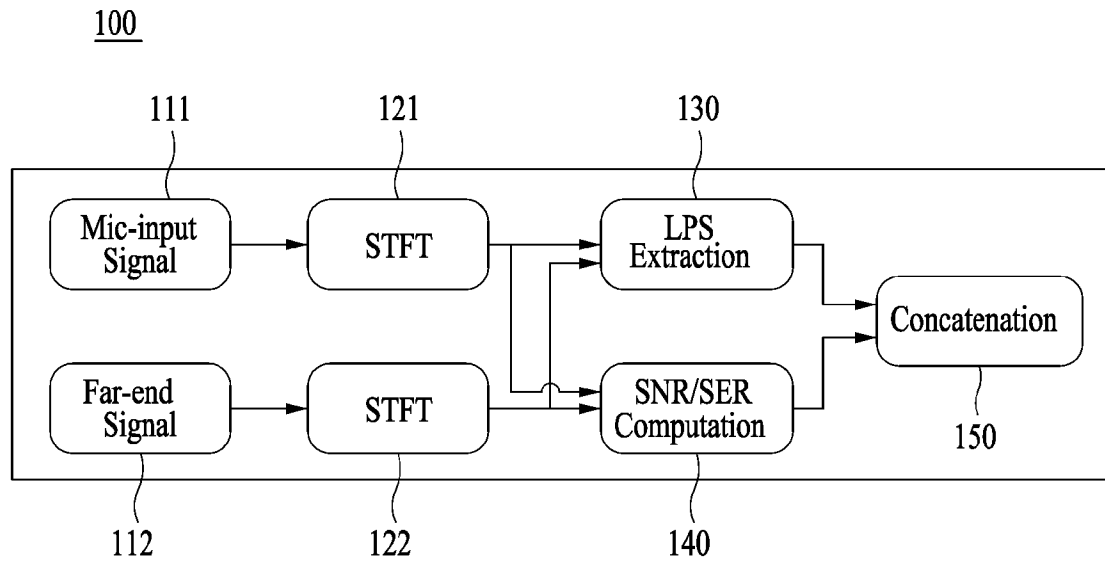
[청구항 11] 제10항에 있어서,
상기 학습부의 특징 벡터 추출부는,
학습 단계에서, 음성 신호의 통계적 정보인 사전 신호대예코비 및 사후 신호대예코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및 사후 신호대잡음비(a priori and a posteriori SNR)를 추출된 상기 특징 벡터에 연결하여 상기 심화신경망(DNN)의 입력 특징 벡터로 이용하는 학습부의 잡음 및 에코 정보부를 포함하는 심화신경망 기반의 잡음 및 에코의 통합 제거 장치.

[청구항 12] 제10항에 있어서,
상기 학습부의 특징 벡터 추출부는,
상기 학습 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와 원단 화자 신호를 대상으로 숏타임 푸리에 변환(Short-Time Fourier Transform, STFT)을 수행하여 시간 영역에서 주파수 영역의 신호로 변환하고,
변환된 주파수 영역의 신호의 로그 파워 스펙트럼(Log Power Spectrum, LPS)을 상기 특징 벡터로 추출하는 학습부의 스펙트럼 추출부를 포함하는 심화신경망 기반의 잡음 및 에코의 통합 제거 장치.

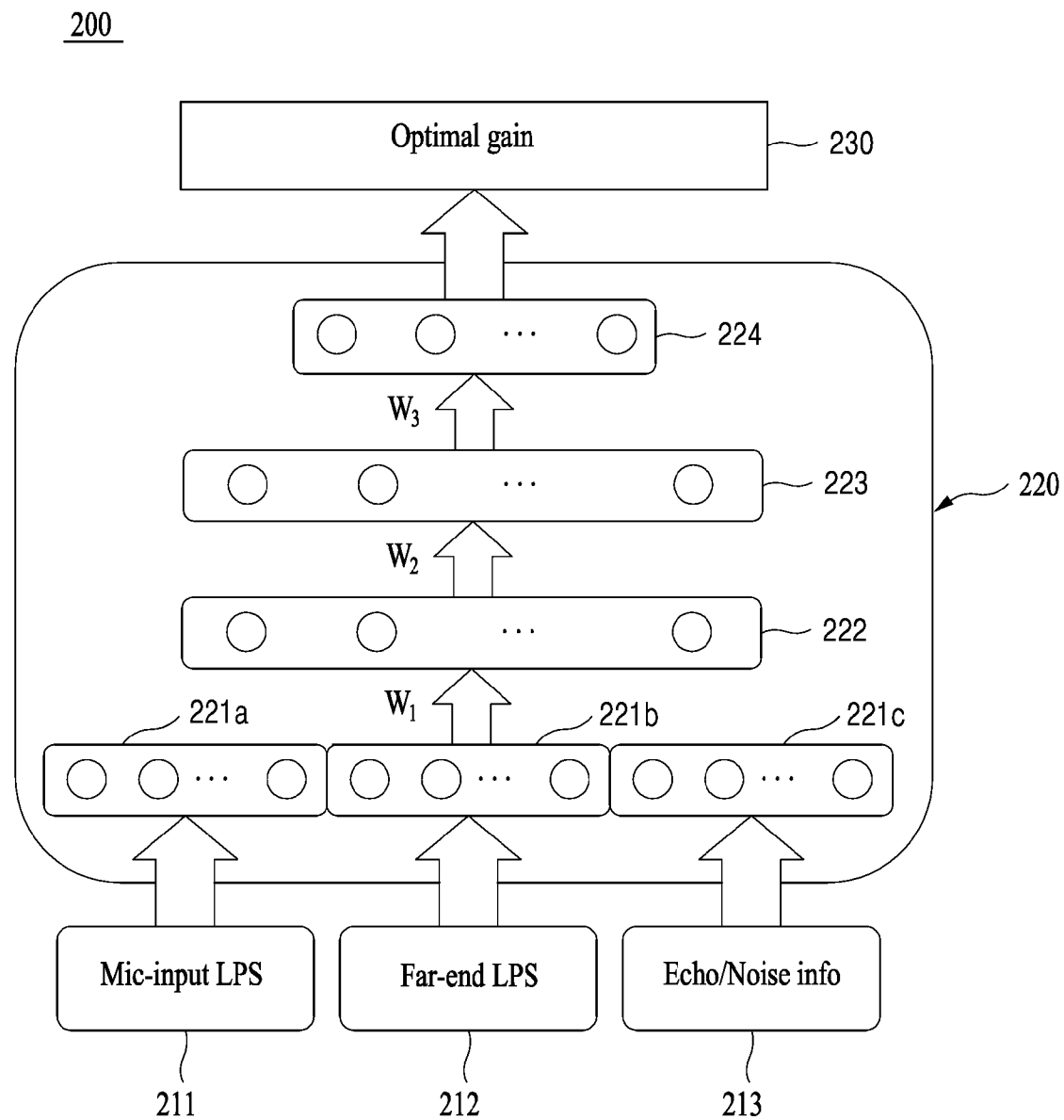
[청구항 13] 제10항에 있어서,
상기 심화신경망(DNN) 모델링부는,
상기 심화신경망(DNN)의 회기 학습(regression)을 통하여 연속적인 최적 이득(optimal gain)을 추정하고, 평균제곱오차(Mean Squared Error, MSE)를 상기 심화신경망(DNN)의 목적 함수로 하여 타겟(target) 특징 벡터인 잡음 및 에코의 통합 제거 이득과 상기 심화신경망(DNN)을 통하여 추정된 상기 잡음 및 에코의 통합 제거 이득의 차이를 최소화하는 방향으로 심화신경망(DNN)의 학습이 진행되는 것

- 을 특징으로 하는 심화신경망 기반의 잡음 및 에코의 통합 제거 장치.
- [청구항 14] 제9항에 있어서,
 상기 특징 벡터 추출부는,
 상기 음성 향상 단계에서, 잡음 및 에코가 포함된 마이크 입력 신호와
 원단 화자 신호를 대상으로 숏타임 푸리에 변환(Short-Time Fourier
 Transform, STFT)을 수행하여 시간 영역에서 주파수 영역의 신호로
 변환하고, 변환된 주파수 영역의 신호의 로그 파워 스펙트럼(Log Power
 Spectrum, LPS)을 상기 특징 벡터로 추출하는 스펙트럼 추출부; 및
 음성 향상 단계에서, 음성 신호의 통계적 정보인 사전 신호대예코비 및
 사후 신호대예코비(a priori and a posteriori SER) 및 사전 신호대잡음비 및
 사후 신호대잡음비(a priori and a posteriori SNR)를 추출된 상기 특징
 벡터에 연결하여 상기 심화신경망(DNN)의 입력 특징 벡터로 이용하는
 잡음 및 에코 정보부
 를 포함하는 심화신경망 기반의 잡음 및 에코의 통합 제거 장치.
- [청구항 15] 제14항에 있어서,
 상기 음성 신호 재구성부는,
 상기 잡음 및 에코가 포함된 마이크 입력 신호의 로그 파워 스펙트럼(Log
 Power Spectrum, LPS)에 추정된 상기 잡음 및 에코의 통합 제거 이득을
 곱하여 근단 화자 신호의 로그 파워 스펙트럼을 획득하고, 상기 근단 화자
 신호의 로그 파워 스펙트럼을 잡음 및 에코가 존재하는 신호의 위상과
 함께 인버스 숏타임 푸리에 변환(Inverse Short-Time Fourier Transform,
 ISTFT)하여 최종적으로 잡음 및 에코가 제거된 최종 근단 화자 신호의
 파형을 획득하는 것
 을 특징으로 하는 심화신경망 기반의 잡음 및 에코의 통합 제거 장치.

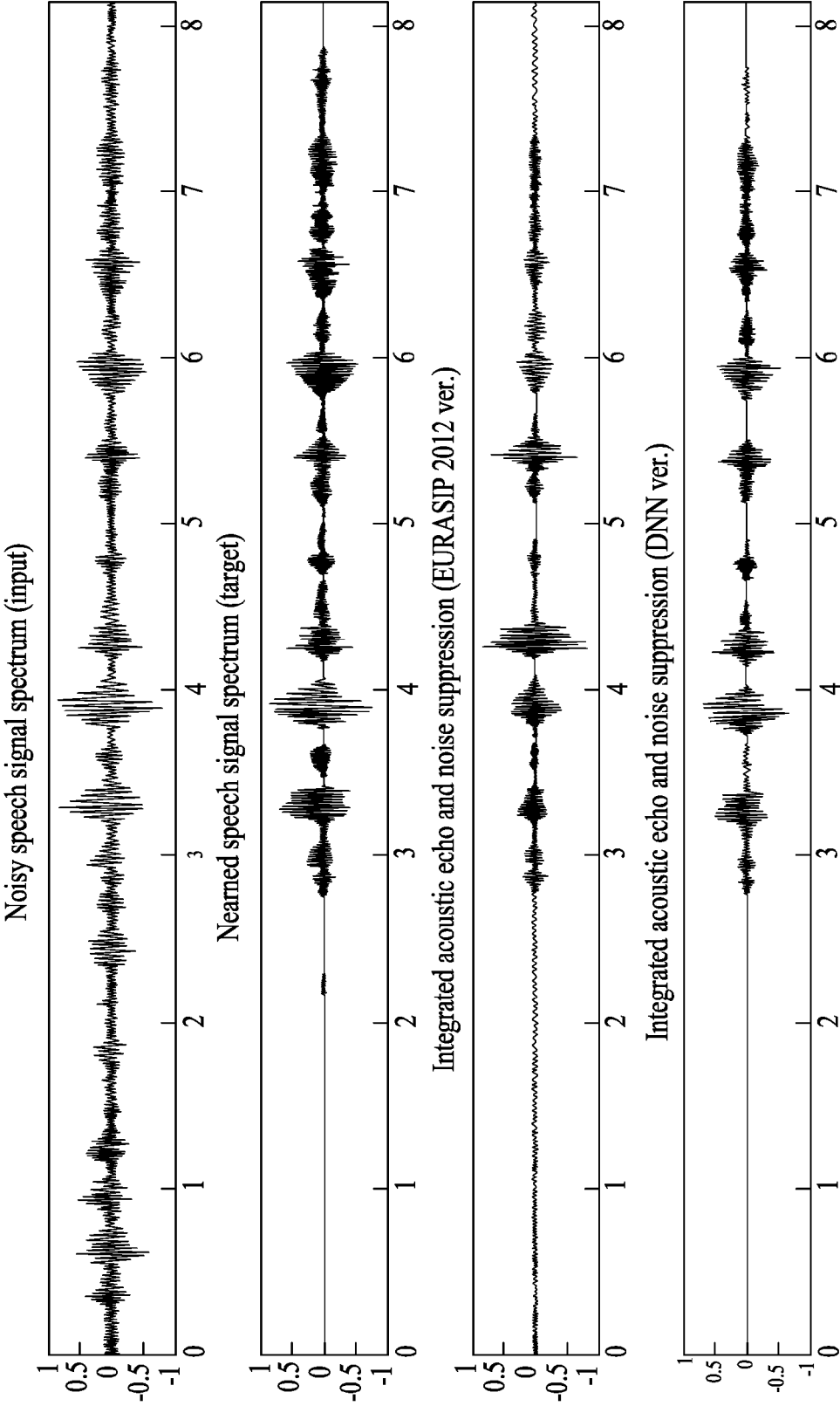
[도1]



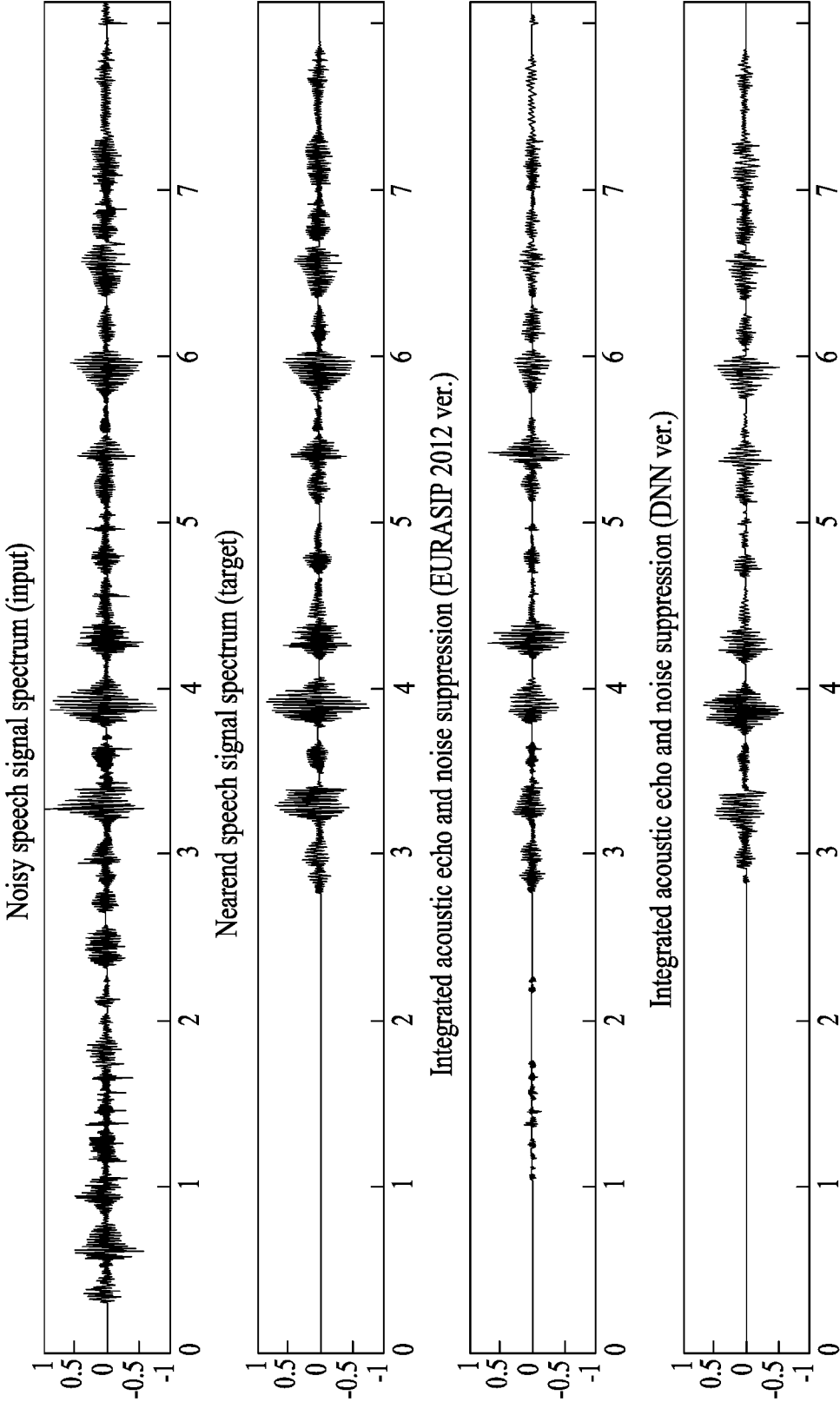
[도2]



[도3]

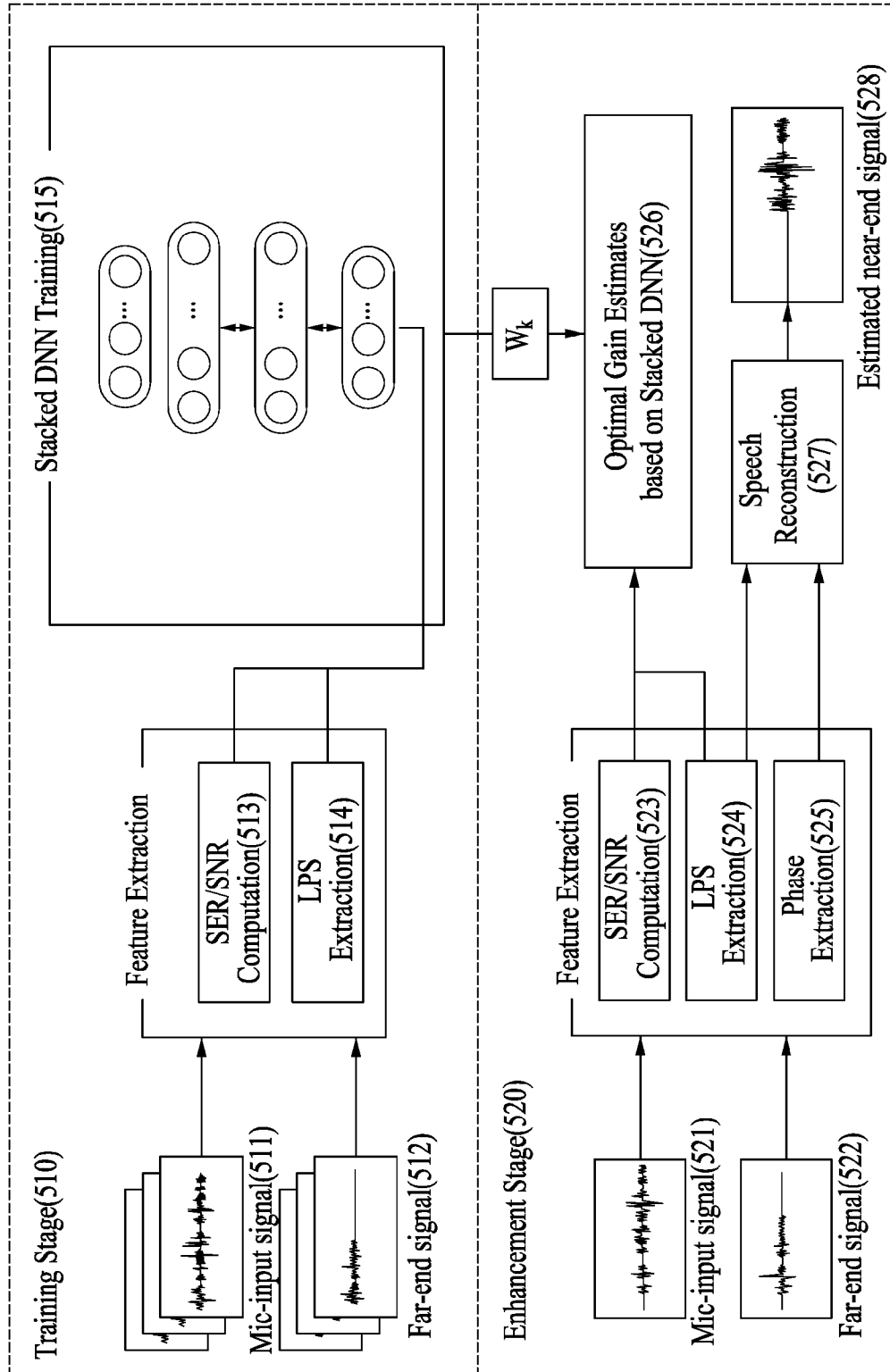


[도4]

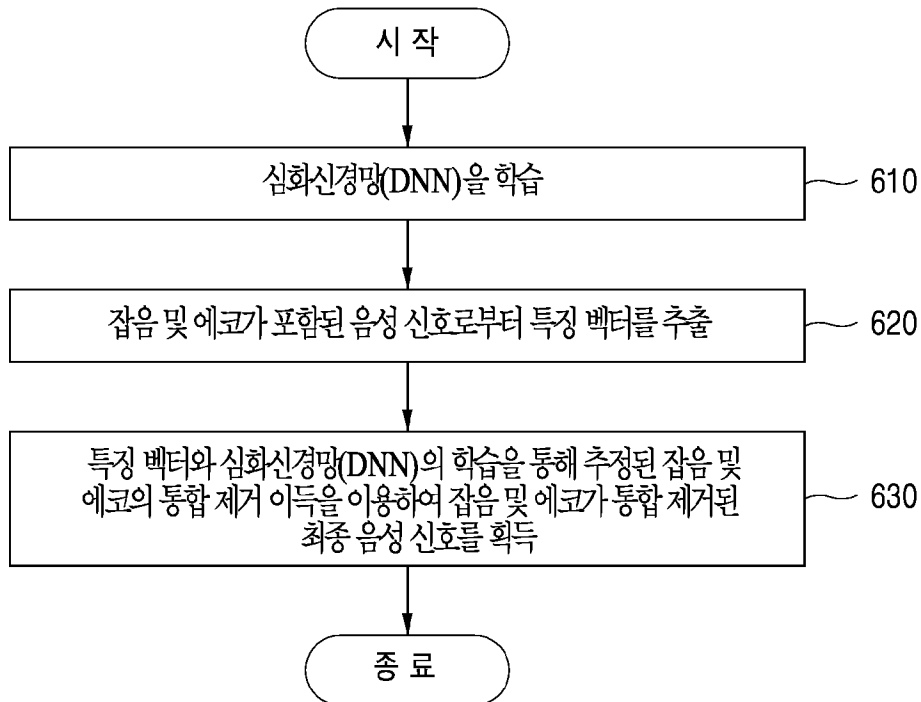


[도5]

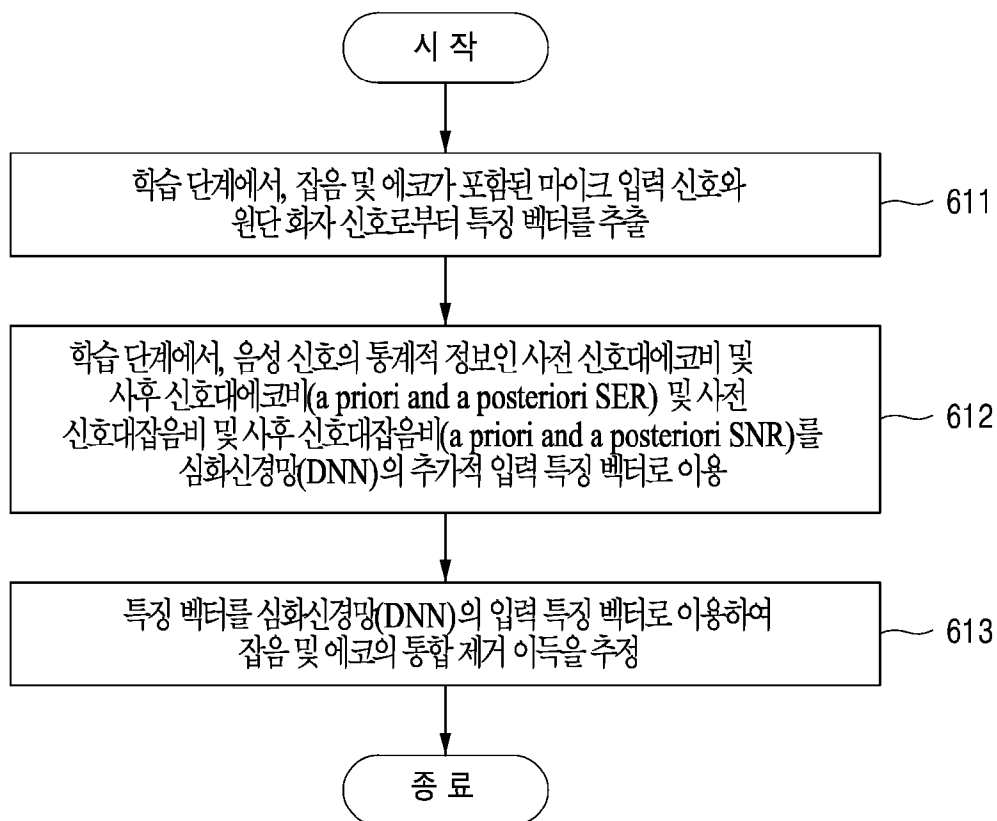
500



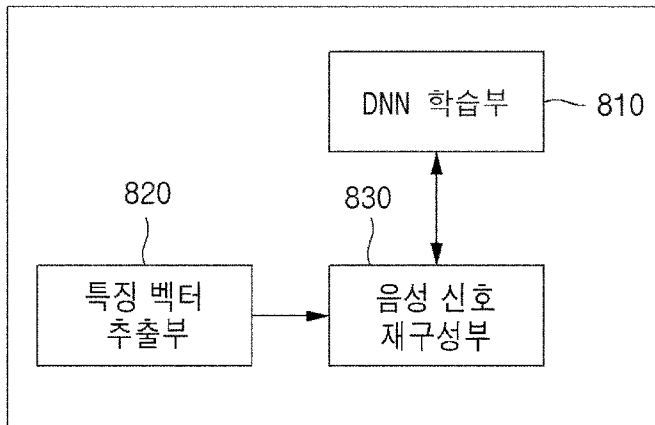
[도6]



[도7]



[도8]

800

INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2018/003840**A. CLASSIFICATION OF SUBJECT MATTER***G10L 21/02(2006.01)i, G10L 19/028(2013.01)i, G10L 19/03(2013.01)i, G10L 19/08(2006.01)i, G10L 25/30(2013.01)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 21/02; G10L 15/16; G10L 15/20; G10L 21/0208; G10L 15/06; G10L 19/028; G10L 19/03; G10L 19/08; G10L 25/30

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: sound treatment, echo, noise, deep neural network (DNN, Deep Neural Network), learning

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2017-0061978 A1 (CAMPBELL, Shannon) 02 March 2017 See paragraphs [0023], [0037], [0040]-[0057]; and figures 2-6.	1-15
Y	JP 2013-037174 A (NIPPON TELEGR & TELEPH CORP. <NTT>) 21 February 2013 See paragraphs [0010], [0025]; and claim 1.	1-15
Y	US 2014-0257803 A1 (MICROSOFT CORPORATION) 11 September 2014 See paragraphs [0037]-[0038], [0054].	2-5, 10-13
Y	US 2011-0238416 A1 (SELTZER, Michael Lewis et al.) 29 September 2011 See paragraph [0035].	3, 6-8, 11, 14-15
A	JP 2014-157323 A (NIPPON TELEGR & TELEPH CORP. <NTT>) 28 August 2014 See paragraphs [0037]-[0042] and claims 1-5.	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

06 JULY 2018 (06.07.2018)

Date of mailing of the international search report

06 JULY 2018 (06.07.2018)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 189 Sconsa-ro, Daejeon 302-701,
Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2018/003840

Patent document cited in search report	Publication date	Patent family member	Publication date
US 2017-0061978 A1	02/03/2017	NONE	
JP 2013-037174 A	21/02/2013	JP 5634959 B2	03/12/2014
US 2014-0257803 A1	11/09/2014	CN 105122279 A	02/12/2015
		EP 2965268 A2	13/01/2016
		US 9177550 B2	03/11/2015
		WO 2014-137952 A2	12/09/2014
		WO 2014-137952 A3	05/02/2015
US 2011-0238416 A1	29/09/2011	US 8700394 B2	15/04/2014
JP 2014-157323 A	28/08/2014	JP 5982297 B2	31/08/2016

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G10L 21/02(2006.01)i, G10L 19/028(2013.01)i, G10L 19/03(2013.01)i, G10L 19/08(2006.01)i, G10L 25/30(2013.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G10L 21/02; G10L 15/16; G10L 15/20; G10L 21/0208; G10L 15/06; G10L 19/028; G10L 19/03; G10L 19/08; G10L 25/30

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: 음향 처리, 에코, 잡음, 심화신경망(DNN, Deep Neural Network), 학습

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	US 2017-0061978 A1 (SHANNON CAMPBELL) 2017.03.02 단락 [0023], [0037], [0040]-[0057]; 및 도면 2-6 참조.	1-15
Y	JP 2013-037174 A (NIPPON TELEGR & TELEPH CORP. <NTT>) 2013.02.21 단락 [0010], [0025]; 및 청구항 1 참조.	1-15
Y	US 2014-0257803 A1 (MICROSOFT CORPORATION) 2014.09.11 단락 [0037]-[0038], [0054] 참조.	2-5, 10-13
Y	US 2011-0238416 A1 (MICHAEL LEWIS SELTZER 등) 2011.09.29 단락 [0035] 참조.	3, 6-8, 11, 14-15
A	JP 2014-157323 A (NIPPON TELEGR & TELEPH CORP. <NTT>) 2014.08.28 단락 [0037]-[0042] 및 청구항 1-5 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

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

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

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

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

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

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

“P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

“&” 동일한 대응특허문헌에 속하는 문헌

국제조사의 실제 완료일

2018년 07월 06일 (06.07.2018)

국제조사보고서 발송일

2018년 07월 06일 (06.07.2018)

ISA/KR의 명칭 및 우편주소

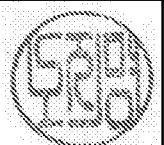
대한민국 특허청
(35208) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)

팩스 번호 +82-42-481-8578

심사관

노지명

전화번호 +82-42-481-8528



국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
US 2017-0061978 A1	2017/03/02	없음	
JP 2013-037174 A	2013/02/21	JP 5634959 B2	2014/12/03
US 2014-0257803 A1	2014/09/11	CN 105122279 A	2015/12/02
		EP 2965268 A2	2016/01/13
		US 9177550 B2	2015/11/03
		WO 2014-137952 A2	2014/09/12
		WO 2014-137952 A3	2015/02/05
US 2011-0238416 A1	2011/09/29	US 8700394 B2	2014/04/15
JP 2014-157323 A	2014/08/28	JP 5982297 B2	2016/08/31