

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권 기구
국제사무국

(43) 국제공개일
2019년 1월 3일 (03.01.2019)



(10) 국제공개번호

WO 2019/004592 A1

- (51) 국제특허분류:
G10L 19/038 (2013.01) G10L 21/038 (2013.01)
G10L 25/30 (2013.01)
- (21) 국제출원번호: PCT/KR2018/005647
- (22) 국제출원일: 2018년 5월 17일 (1-05.2018)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2017-0081024 2017년 6월 27일 (2-06.2017) KR
- (71) 출원인: 한양대학교 산학협력단 (IN -
DUSTRY-UNIVERSITY COOPERATION FOUN -
DATION HANYANG UNIVERSITY) [KR/KR]; 04763
서울시 성동구 왕십리로 222, Seoul (KR).
- (72) 발명자: 장준혁 (CHANG, Joon-Hyuk); 06294 서울시 강
남구 언주로 30길 26 지동 2708호, Seoul (KR). 노경진
(NOH, Kyoungjin); 04760 서울시 성동구 사근동길 65
101동 212호, Seoul (KR).
- (74) 대리인: 양성보 (YANG, Sungbo); 06099 서울시 강남구
선릉로 125길 14 삼성빌딩 2층 피앤티특허법률사무소,
Seoul (KR).
- (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국
내 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 국
내 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), 유
럽 (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).

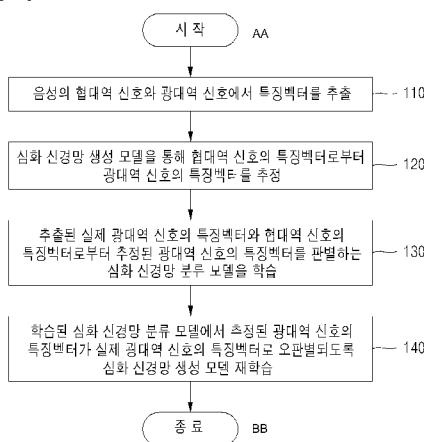
공개:

— 국제조사보고서와 함께 (조약 제21조(3))

(54) Title: GENERATIVE ADVERSARIAL NETWORK-BASED VOICE BANDWIDTH EXTENDER AND EXTENSION METHOD

(54) 발명의 명칭: 생성적 대립망 기반의 음성 대역폭 확장기 및 확장 방법

[도1]



- 110 ... Extract feature vectors from narrowband signal and wideband signal of voice.
120 ... Estimate feature vector of wideband signal from feature vector of narrowband signal through deep neural network generative model.
130 ... Learn deep neural network classification model for discriminating estimated feature vector of wideband signal from actually extracted feature vector of wideband signal and actually extracted feature vector of narrowband signal.
140 ... Re-learn deep neural network classification model so that feature vector of wideband signal estimated in learned deep neural network classification model is erroneously discriminated as actual feature vector of wideband signal.

AA ... Start

BB ... End

(57) Abstract: Proposed are a generative adversarial network-based voice bandwidth extender and extension method. A generative adversarial network-based voice bandwidth extension method, according to an embodiment, comprises the steps of: extracting feature vectors from a narrowband (NB) signal and a wideband (WB) signal of a voice; estimating the feature vector of the wideband signal from the feature vector of the narrowband signal; and learning a deep neural network classification model for discriminating the estimated feature vector of the wideband signal from the actually extracted feature vector of the wideband signal and the actually extracted feature vector of the narrowband signal.

(57) 요약서: 생성적 대립망 기반의 음성 대역폭 확장기 및 확장 방법이 제시된다. 일 실시예에 따른 생성적 대립망 기반의 음성 대역폭 확장 방법은, 음성의 협대역(Narrowband, NB) 신호와 광대역(Wideband, WB) 신호에서 특징벡터를 추출하는 단계; 상기 협대역 신호의 특징벡터로부터 광대역 신호의 특징벡터를 추정하는 단계; 및 추출된 실제 상기 광대역 신호의 특징벡터와 상기 협대역 신호의 특징벡터로부터 추정된 광대역 신호의 특징벡터를 판별하는 심화 신경망 분류 모델을 학습시 키는 단계를 포함하여 이루어질 수 있다—

명세서

발명의 명칭: 생성적 대립 망 기반의 음성 대역폭 확장기 및 확장 방법

기술분야

[1] 아래의 실시예들은 음성 대역폭 확장기 및 확장 방법에 관한 것으로, 협대역 음성 신호를 광대역 음성 신호로 확장하여 음성 통화 품질을 향상시키는 기술에 관한 것이다.

배경기술

[2] 기존의 음성 대역폭 확장기 기술은 다양한 기계학습 기법을 기반으로 하여 고주파 대역의 정보를 추정하는 방법들이 시도되어 왔으며, 특히 최근에는 심화 신경망을 이용한 방법들이 많이 시도되고 있다. 심화 신경망(Deep Neural Network, DNN)은 머신 러닝(machine learning) 기법으로서, 다수의 은닉 층(layer)과 은닉 노드들을 이용하여 입력과 출력 간의 비선형적인 관계를 효과적으로 모델링하는 것이 가능하며, 최근 음성뿐 아니라 다양한 분야에서 뛰어난 성능을 보이고 있는 기술이다.

[3] 심화 신경망을 이용한 방법은 크게 두 가지로 나눠지는데, 하나는 기존의 기계학습 분류 모델을 심화 신경망 분류 모델로 단순 대체하는 방법이고, 다른 하나는 심화 신경망 생성 모델을 이용하여 고주파 대역의 성분을 직접적으로 추정하는 방법이다.

[4] 심화 신경망 분류 모델을 사용한 경우 기존의 기계학습 분류 모델을 사용했을 때 보다는 성능이 향상되었으나 모든 음성의 특징을 몇 개의 등급으로 나눠서 추정하는 방법에서 오는 추정 성능의 한계점이 있으며, 심화 신경망 생성 모델을 사용한 경우 실제 값을 제대로 추정하지 못하고 추정 값들이 평활화(smoothing) 되는 한계점을 보여 왔다. 이러한 한계점을 원인으로, 특히 고주파 대역 스펙트럼에 대해 과추정이 발생하는 경우 듣기 싫은 잡음을 발생시키는 문제점이 있다.

[5] 한국등록특허 10-0762596 호는 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 음성 대역폭 확장기 및 확장 방법에 관하여 기술하며, 보다 구체적으로 생성적 대립 망을 이용하여 음성 신호의 고주파 대역을 추정하는 음성 대역폭 확장기 및 확장 방법을 제공한다.

[7] 실시예들은 분류 모델을 진짜 데이터인 광대역 신호와 생성 모델을 통해 고주파 대역을 생성해 낸 가짜 데이터를 구분할 수 있도록 충분히 학습시킨

후에, 분류 모델을 고정시키고 생성 모델이 실제 광대역 신호와 가까운 고주파 대역을 추정하여 분류 모델을 최대한 오분류하게 만들 수 있도록 학습하는 방법으로 생성적 대립 망을 음성 대역폭 확장 방법에 적용하여 향상된 음성 대역폭 확장 방법을 제공하는데 있다.

과제 해결 수단

- [8] 일 실시 예에 따른 생성적 대립 망 기반의 음성 대역폭 확장 방법은, 음성의 협대역(Narrowband, NB) 신호와 광대역(Wideband, WB) 신호에서 특징 백터를 추출하는 단계; 상기 협대역 신호의 특징 백터로부터 광대역 신호의 특징 백터를 추정하는 단계; 및 추출된 실제 상기 광대역 신호의 특징 백터와 상기 협대역 신호의 특징 백터로부터 추정된 광대역 신호의 특징 백터를 판별하는 심화 신경망 분류 모델을 학습시키는 단계를 포함하여 이루어질 수 있다.
- [9] 상기 협대역 신호의 특징 백터로부터 광대역 신호의 특징 백터를 추정하는 단계는, 상기 협대역 신호의 특징 백터를 입력으로 심화 신경망 생성 모델을 학습시켜 추정된 광대역 신호의 특징 백터를 출력할 수 있다.
- [10] 학습된 상기 심화 신경망 분류 모델에서 상기 추정된 광대역 신호의 특징 백터가 상기 실제 광대역 신호의 특징 백터로 오판별되도록 상기 심화 신경망 생성 모델 재학습시키는 단계를 더 포함할 수 있다.
- [11] 상기 심화 신경망 분류 모델에서 상기 추정된 광대역 신호의 특징 백터가 상기 실제 광대역 신호의 특징 백터로 오판별되도록 상기 심화 신경망 생성 모델 재학습시키는 단계는, 상기 심화 신경망 분류 모델이 상기 추정된 광대역 신호의 특징 백터가 상기 실제 광대역 신호의 특징 백터로 오판별하는 것을 목적으로 비용 함수를 설계하여 학습시킬 수 있다.
- [12] 상기 추출된 실제 상기 광대역 신호의 특징 백터와 상기 협대역 신호의 특징 백터로부터 추정된 광대역 신호의 특징 백터를 판별하는 심화 신경망 분류 모델을 학습시키는 단계는, 추출된 실제 상기 광대역 신호의 특징 백터와 상기 협대역 신호의 특징 백터로부터 추정된 광대역 신호의 특징 백터를 판별할 수 있도록 비용 함수를 설계하여 상기 심화 신경망 분류 모델을 학습시킬 수 있다.
- [13] 상기 심화 신경망 생성 모델 또는 상기 심화 신경망 분류 모델은, DNN(Deep Neural Networks), CNN(Convolutional Neural Networks), RNN(Recurrent Neural Networks) 및 LSTM(Long- Short Term Memory) 중 적어도 어느 하나일 수 있다.
- [14] 상기 심화 신경망 생성 모델은, 엔코딩(Encoding) 역할을 수행하는 컨벌루션 네이티브(Convolutive Layer) 와 디코딩(Decoding) 역할을 수행하는 디컨벌루션 네이티브(De-Convolutional Layer) 의 대칭 구조의 상기 CNN(Convolutional Neural Networks) 일 수 있다.
- [15] 상기 심화 신경망 분류 모델을 상기 실제 광대역 신호의 특징 백터와 상기 심화 신경망 생성 모델을 통해 고주파 대역을 생성해 낸 상기 추정된 광대역 신호의 특징 백터를 구분할 수 있도록 충분히 학습시킨 후, 상기 심화 신경망 분류

모델을 고정시키고 상기 심화 신경망 생성 모델이 상기 실제 광대역 신호의 특징 백터와 가까운 고주파 대역을 추정하여 상기 심화 신경망 분류

모델을 오분류하도록 학습시키는 생성적 대립 망을 통해 음성 대역폭을 확장시킬 수 있다.

- [16] 다른 실시 예에 따른 생성적 대립 망 기반의 음성 대역폭 확장 장치는, 음성의 협대역(Narrowband, NB) 신호와 광대역(Wideband, WB) 신호에서 특징 백터를 추출하는 특징 백터 추출부; 상기 협대역 신호의 특징 백터를 입력으로 심화 신경망 생성 모델을 학습시켜 추정된 광대역 신호의 특징 백터를 출력하는 생성 모델링부; 및 추출된 실제 상기 광대역 신호의 특징 백터와 상기 협대역 신호의 특징 백터로부터 추정된 광대역 신호의 특징 백터를 판별하는 심화 신경망 분류 모델을 학습시키는 분류 모델링부를 포함하여 이루어질 수 있다.
- [17] 상기 생성 모델링 부는, 학습된 상기 심화 신경망 분류 모델에서 상기 추정된 광대역 신호의 특징 백터가 상기 실제 광대역 신호의 특징 백터로 오판별 되도록 상기 심화 신경망 생성 모델을 재학습시킬 수 있다.
- [18] 상기 생성 모델링 부는, 상기 심화 신경망 분류 모델이 상기 추정된 광대역 신호의 특징 백터가 상기 실제 광대역 신호의 특징 백터로 오판별하는 것을 목적으로 비용 함수를 설계하여 학습시킬 수 있다.
- [19] 상기 분류 모델링 부는, 추출된 실제 상기 광대역 신호의 특징 백터와 상기 협대역 신호의 특징 백터로부터 추정된 광대역 신호의 특징 백터를 판별할 수 있도록 비용 함수를 설계하여 상기 심화 신경망 분류 모델을 학습시킬 수 있다.
- [20] 상기 심화 신경망 생성 모델 또는 상기 심화 신경망 분류 모델은, DNN(Deep Neural Networks), CNN(Convolutional Neural Networks), RNN(Recurrent Neural Networks) 및 LSTM(Long- Short Term Memory) 중 적어도 어느 하나일 수 있다.
- [21] 상기 심화 신경망 생성 모델은, 엔코딩(Encoding) 역할을 수행하는 컨벌루셔널 레이어(Convolutional Layer) 와 디코딩(Decoding) 역할을 수행하는 디컨벌루셔널 레이어(De-Convolutional Layer) 의 대칭 구조의 상기 CNN(Convolutional Neural Networks) 일 수 있다.
- [22] 상기 심화 신경망 분류 모델을 상기 실제 광대역 신호의 특징 백터와 상기 심화 신경망 생성 모델을 통해 고주파 대역을 생성해 낸 상기 추정된 광대역 신호의 특징 백터를 구분할 수 있도록 충분히 학습시킨 후, 상기 심화 신경망 분류 모델을 고정시키고 상기 심화 신경망 생성 모델이 상기 실제 광대역 신호의 특징 백터와 가까운 고주파 대역을 추정하여 상기 심화 신경망 분류 모델을 오분류하도록 학습시키는 생성적 대립 망을 통해 음성 대역폭을 확장시킬 수 있다.
- 발명의 효과**
- [23] 실시 예들에 따르면 음성 통신 환경에서 코텍 변환이나 전송비 트율의 확장 없이 더 넓은 대역폭의 음성을 들을 수 있는 생성적 대립 망 기반의 음성 대역폭

확장기 및 확장 방법을 제공할 수 있다. 특히, 생성적 대립망을 이용하여 심화 신경망 기반 생성 모델의 추정 성능을 극대화함으로써 기존 심화 신경망 기반 생성 모델을 이용한 방법의 한계로 인해 발생되는 사람이 듣기 불편한 인위적인 소리의 합성 없이 향상된 음성 통화 품질을 기대할 수 있으며, 나아가서 음성 인식기에 적용되어 기존 협대역 코텍을 사용했을 때와 같은 전송속도로 더 높은 수준의 음성인식률을 제공할 수 있다.

도면의 간단한 설명

- [24] 도 1은 일 실시 예에 따른 생성적 대립망 기반의 음성 대역폭 확장 방법을 나타내는 흐름도이다.
- [25] 도 2는 일 실시 예에 따른 생성적 대립망 기반의 음성 대역폭 확장 장치를 나타내는 블록도이다.
- [26] 도 3은 일 실시 예에 따른 음성 대역폭 확장을 위한 심화 신경망 생성 모델의 학습 방법을 설명하기 위한 도면이다.
- [27] 도 4는 일 실시 예에 따른 음성 대역폭 확장을 위한 심화 신경망 생성 모델(DNN) 구조의 예를 나타내는 도면이다.
- [28] 도 5는 다른 실시 예에 따른 음성 대역폭 확장을 위한 심화 신경망 생성 모델(CNN) 구조의 예를 나타내는 도면이다.
- [29] 도 6은 일 실시 예에 따른 생성적 대립망 기반의 음성 대역폭 확장을 위한 심화 신경망 분류 모델의 학습 방법을 설명하기 위한 도면이다.
- [30] 도 7은 일 실시 예에 따른 생성적 대립망 기반의 음성 대역폭 확장을 위한 심화 신경망 생성 모델의 재학습 방법을 설명하기 위한 도면이다.
- 발명의 실시를 위한 최선의 형태**
- [31] 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.
- [32]
- [33] 아래의 실시예들은 음성 대역폭 확장기 및 확장 방법에 관한 것으로, 보다 구체적으로는 생성적 대립망을 이용하여 음성 신호의 고주파 대역을 추정하는 음성 대역폭 확장기 및 확장 방법에 관한 것이다.
- [34] 심화 신경망 기반 모델의 한계점을 극복하기 위하여 생성적 대립망 기법이 도입되었으며, 생성적 대립망 기법은 특히 심화 신경망 생성 모델에서 향상된 성능을 도출하였다. 생성적 대립망 기법은 분류 모델과 생성 모델을 학습시켜 서로 경쟁하게 함으로써 생성 모델의 성능을 극대화 하는 방법이다. 먼저, 분류 모델을 진짜 데이터와 생성 모델을 통해 생성해 낸 가짜 데이터를 구분할 수

있도록 충분히 학습시킨 후에, 분류 모델을 고정시키고 생성 모델이 분류 모델을 최대한 오분류하게 만들 수 있도록 학습시킬 수 있다.

- [35] 본 실시 예들에서는 FFT(Fast Fourier Transform) 및 STFT(Short Time Fourier Transform) 변환을 이용하는 경우를 예로 들어 설명하나, 이는 실시 예에 해당되며, FFT 및 STFT 이외에 ISTFT(Inverse Short Time Fourier Transform), DFT (Discrete Fourier Transform), IDFT (Inverse Discrete Fourier Transform) 변환, IFFT(Inverse Fast Fourier Transform) 변환 등이 이용될 수도 있다.
- [36]
- [37] 도 1은 일 실시 예에 따른 생성적 대립망 기반의 음성 대역폭 확장 방법을 나타내는 흐름도이다.
- [38] 도 1을 참조하면, 일 실시 예에 따른 생성적 대립망 기반의 음성 대역폭 확장 방법은, 음성의 협대역(Narrowband, NB) 신호와 광대역(Wideband, WB) 신호에서 특징 백터를 추출하는 단계(110), 협대역 신호의 특징 백터로부터 광대역 신호의 특징 백터를 추정하는 단계(120), 및 추출된 실제 광대역 신호의 특징 백터와 협대역 신호의 특징 백터로부터 추정된 광대역 신호의 특징 백터를 판별하는 심화 신경망 분류 모델을 학습시키는 단계(130)를 포함하여 이루어질 수 있다.
- [39] 여기에서, 학습된 심화 신경망 분류 모델에서 추정된 광대역 신호의 특징 백터가 실제 광대역 신호의 특징 백터로 오판별되도록 심화 신경망 생성 모델 재학습시키는 단계(140)를 더 포함하여 이루어질 수 있다.
- [40] 종래에는 심화 신경망 생성 모델을 사용한 경우 실제 값을 제대로 추정하지 못하고 추정 값들이 평활화(smoothing) 되는 한계점을 보여 왔다. 이러한 한계점을 원인으로, 특히 고주파 대역 스펙트럼에 대해 과주정이 발생하는 경우 듣기 싫은 잡음을 발생시키는 문제점이 있다.
- [41] 실제 값의 분포를 제대로 추정하지 못하는 문제점을 해결하기 위하여, 첫 번째로 광대역 신호(정답)에서 특징 백터를 추출하여 심화 신경망 분류 모델을 학습시킬 수 있다. 이 때, 특징 백터는 주파수 측의 스펙트럼 또는 스펙트럼에서 추출한 여러 가지 특징 백터가 되거나 시간 측의 광대역 신호 자체도 가능하다.
- [42] 광대역 신호를 이용하여 분류 모델을 충분히 학습(정답 학습) 시킨 후에, 다음으로는 협대역 신호와 광대역 신호를 이용하여 심화 신경망 생성 모델을 학습시켜 협대역 신호를 입력으로 추정된 광대역 신호(오답)를 만들어 낼 수 있는 심화 신경망 생성 모델을 만들어 낼 수 있다. 심화 신경망 생성 모델이 충분히 학습된 후, 이전에 협대역 신호와 광대역 신호를 이용하여 학습한 심화 신경망 분류 모델을 이번에는 협대역 신호와 추정된 광대역 신호를 이용하여 충분히 학습(오답 학습) 시킬 수 있다.
- [43] 마지막으로 정답과 오답을 충분히 학습시킨 심화 신경망 분류 모델은 고정시킨 상태로, 심화 신경망 생성 모델을 이용하여 협대역 신호의 특징 백터를 입력으로 하여 광대역 신호의 특징 백터 추정 성능을 향상 시킬 수 있다. 이 때, 광대역 신호의 특징 백터 추정 성능을 향상시키는 학습은 심화 신경망 분류 모델을

- 속이고 광대역 신호(정답)로 분류하게 만드는 방향으로 진행될 수 있다.
- [44] 심화 신경망 분류 모델을 속일 수 있도록 충분히 학습된 심화 신경망 생성 모델 기반의 고주파 대역 특징 백터 추정 성능은 종래의 심화 신경망 생성 모델보다 향상된 추정 성능을 보이게 될 수 있다.
- [45] 이와 같이, 실시 예들은 분류 모델을 진짜 데이터인 광대역 신호와 생성 모델을 통해 고주파 대역을 생성해 낸 가짜 데이터를 구분할 수 있도록 충분히 학습시킨 후에, 분류 모델을 고정시키고 생성 모델이 실제 광대역 신호와 가까운 고주파 대역을 추정하여 분류 모델을 최대한 오분류하게 만들 수 있도록 학습하는 방법으로 생성적 대립 망을 음성 대역폭 확장 방법에 적용하여 향상된 음성 대역폭 확장 방법을 제공할 수 있다.
- [46] 아래에서는 생성적 대립 망 기반의 음성 대역폭 확장 방법에 대해 하나의 예를 들어 보다 구체적으로 설명한다.
- [47]
- [48] 도 2는 일 실시 예에 따른 생성적 대립 망 기반의 음성 대역폭 확장 장치를 나타내는 블록도이다.
- [49] 도 1에서 각 단계들 (110-140)은 도 2의 구성요소인 특징 백터 추출부 (210), 생성 모델링부 (220) 및 분류 모델링부 (230)에 의해 수행될 수 있다.
- [50] 도 2에 따르면, 일 실시 예에 따른 생성적 대립 망 기반의 음성 대역폭 확장 장치(200)는 특징 백터 추출부 (210), 생성 모델링부 (220) 및 분류 모델링부 (230)를 포함할 수 있다.
- [51] 단계(110)에서, 특징 백터 추출부 (210)는 음성의 협대역(Narrowband, NB) 신호와 광대역(Wideband, WB) 신호에서 특징 백터를 추출할 수 있다.
- [52] 이 때, 광대역 신호를 다운 샘플링(Down Sampling) 하여 협대역 신호를 생성할 수 있으며, 실제 통신 환경에서의 성능 저하를 반영하기 위해 협대역 코텍(Adaptive Multi-Rate(AMR) 또는 Adaptive Multi-Rate Narrowband(AMR-NB))을 이용하여 협대역 신호를 변형시킬 수도 있다.
- [53] 단계(120)에서, 생성 모델링부 (220)는 협대역 신호의 특징 백터로부터 광대역 신호의 특징 백터를 추정할 수 있다. 생성 모델링부 (220)는 협대역 신호의 특징 백터를 입력으로 심화 신경망 생성 모델을 학습시켜 추정된 광대역 신호의 특징 백터를 출력할 수 있다.
- [54] 여기에서, 심화 신경망 생성 모델은 DNN(Deep Neural Networks), CNN(Convolutional Neural Networks), RNN(Recurrent Neural Networks) 및 LSTM(Long-Short Term Memory) 중 적어도 어느 하나일 수 있다. 예컨대, 심화 신경망 생성 모델은 엔코딩(Encoding) 역할을 수행하는 컨벌루셔널 레이어(Convolutional Layer)와 디코딩(Decoding) 역할을 수행하는 디컨벌루셔널 레이어(De-Convolutional Layer)의 대칭 구조의 CNN(Convolutional Neural Networks)일 수 있다.
- [55] 단계(130)에서, 분류 모델링부 (230)는 추출된 실제 광대역 신호의 특징 백터와

협대역 신호의 특징 백터로부터 추정된 광대역 신호의 특징 백터를 판별하는 심화 신경망 분류 모델을 학습시킬 수 있다.

- [56] 여기에서, 심화 신경망 분류 모델은 DNN(Deep Neural Networks), CNN(Convolutional Neural Networks), RNN(Recurrent Neural Networks) 및 LSTM(Long-Short Term Memory) 중 적어도 어느 하나일 수 있다.
- [57] 분류 모델링부 (230)는 추출된 실제 광대역 신호의 특징 백터와 협대역 신호의 특징 백터로부터 추정된 광대역 신호의 특징 백터를 판별할 수 있도록 비용 함수를 설계하여 심화 신경망 분류 모델을 학습시킬 수 있다.
- [58] 단계(140)에서, 생성 모델링부 (220)는 학습된 심화 신경망 분류 모델에서 추정된 광대역 신호의 특징 백터가 실제 광대역 신호의 특징 백터로 오판별되도록 심화 신경망 생성 모델 재학습시킬 수 있다.
- [59] 여기에서, 생성 모델링부 (220)는 심화 신경망 분류 모델이 추정된 광대역 신호의 특징 백터가 실제 광대역 신호의 특징 백터로 오판별하는 것을 목적으로 비용 함수를 설계하여 학습시킬 수 있다.
- [60] 실시 예들에 따르면 심화 신경망 분류 모델을 실제 광대역 신호의 특징 백터와 심화 신경망 생성 모델을 통해 고주파 대역을 생성해 낸 추정된 광대역 신호의 특징 백터를 구분할 수 있도록 충분히 학습시킨 후, 심화 신경망 분류 모델을 고정시키고 심화 신경망 생성 모델이 실제 광대역 신호의 특징 백터와 가까운 고주파 대역을 추정하여 심화 신경망 분류 모델을 오분류 하도록 학습시키는 생성적 대립 망을 통해 음성 대역폭을 확장시킬 수 있다.
- [61]
- [62] 본 실시 예에서 제안하는 생성적 대립 망의 학습 방법은 다음과 같다. 먼저 협대역 신호의 특징 백터를 이용하여 광대역 신호를 추정할 수 있는 심화 신경망 생성 모델을 학습할 수 있다. 이 때, 특징 백터는 일반적으로 주파수 측의 특징 백터를 사용하나, 시간 측의 신호를 사용하는 것도 가능하다. 다음으로 실제 광대역 신호와 심화 신경망 생성 모델을 이용하여 추정된 광대역 신호를 판별할 수 있는 심화 신경망 분류 모델을 충분히 학습할 수 있다. 심화 신경망 분류 모델이 실제 광대역 신호와 추정된 광대역 신호를 판별 가능하도록 충분히 학습되고 나면, 심화 신경망 생성 모델을 이용하여 추정된 광대역 신호가 심화 신경망 분류 모델을 속이고 실제 광대역 신호로 판별될 수 있도록 심화 신경망 생성 모델을 학습시킬 수 있다.
- [63] 이하에서는 생성적 대립 망 기반의 음성 대역폭 확장 기술에 대해 보다 상세히 설명하기로 한다.
- [64] 먼저, 특징 백터 추출부 (210)는 음성의 협대역(Narrowband, NB) 신호와 광대역(Wideband, WB) 신호에서 특징 백터를 추출할 수 있다. 협대역 신호는 광대역 신호를 다운 샘플링(Down Sampling) 하여 생성하며, 실제 통신 환경에서의 코덱으로 인한 성능 저하를 반영하기 위해 협대역 코덱을 이용하여 성능 저하를 시킬 수 있다. 예컨대, 실제 통신 환경에서의 성능 저하를 반영하기 위해 협대역

- 코 텍(Adaptive Multi-Rate(AMR)) 또는 Adaptive Multi-Rate Narrowband(AMR-NB)) 을 이용하여 협대역 신호를 변형시킬 수도 있다.
- [65] 협대역 신호와 광대역 신호에서 추출할 특징 백터는 시간 축의 음성 신호 자체가 될 수도 있고, 주파수 축의 로그 파워 스펙트럼(Log Power Spectrum, LPS) 이 될 수도 있다.
- [66] 시간 축의 음성 신호를 특징 백터로 사용할 경우, 협대역 음성 신호를 광대역 신호로 업샘플링(Up Sampling) 한 후에 일정한 길이의 프레임으로 나눠서 사용할 수 있다. 예를 들어, 1초 길이의 프레임을 0.5초 길이로 겹치게 나눠서 추출한 후 특징 백터로 사용할 수 있다.
- [67] 그리고 주파수 축의 로그 파워 스펙트럼을 특징 백터로 사용할 경우, 협대역 음성 신호를 광대역 신호로 업샘플링 한 후에 일정한 길이의 프레임으로 나눠서 STFT(Short Time Fourier Transform) 을 이용하여 주파수 성분의 값을 구할 수 있다.
- [68] 고주파 대역 확장 방법에 주파수 축의 값을 사용할 경우, 주파수 성분 값을 크기와 위상으로 분리하여 크기 값만을 특징 백터로 사용할 수 있다. 위상 값은 음성 신호의 음질에 크게 영향을 주지 않기 때문에, 위상 값의 일반적인 특성에 따라 협대역의 위상 값을 영점 대칭하여 광대역의 위상 값으로 추정해 사용할 수 있다. 특징 백터로 사용되는 크기 값은 제곱한 후 로그를 취하여 최종적으로 특징 백터로 사용하게 될 수 있다. 이는 사람의 청각적 특성을 잘 반영할 수 있는 범위로 값을 변환해 주기 위함이기도 하며, 심화 신경망 생성 모델의 추정 성능을 향상시키기 위함이기도 하다. 일반적으로 심화 신경망 생성 모델은 값의 범위가 너무 넓을 경우 추정이 잘 되지 않는다.
- [69] 도 3은 일 실시 예에 따른 음성 대역폭 확장을 위한 심화 신경망 생성 모델의 학습 방법을 설명하기 위한 도면이다.
- [70] 도 3을 참조하면, 생성 모델링부(220)는 협대역 신호의 특징 백터(301)를 입력으로 광대역 신호의 특징 백터(302)를 추정하는 심화 신경망 생성 모델(310)을 학습시킬 수 있다. 즉, 생성 모델링부(220)는 협대역 신호의 특징 백터(301)를 심화 신경망 생성 모델(310)에 입력하여 광대역 신호의 특징 백터(302)를 추정할 수 있다. 여기에서, 추정된 광대역 신호의 특징 백터(302)는 음성 대역폭 확장(Bandwidth Extension, BWE) 이 처리된 신호이다.
- [71] 이 때, 심화 신경망 생성 모델(310)은 DNN(Deep Neural Networks), CNN(Convolutional Neural Networks), RNN(Recurrent Neural Networks) 및 LSTM(Long-Short Term Memory) 중 어느 것으로도 설계가 가능하다.
- [72] 심화 신경망 생성 모델(310)의 비용 함수는 평균제곱오차(Mean Squared Error, MSE) 를 사용하여, 실제 값과 추정 값의 제곱의 평균이 작아지는 방향으로 학습할 수 있도록 설계할 수 있다.
- [73] 도 4는 일 실시 예에 따른 음성 대역폭 확장을 위한 심화 신경망 생성

모델(DNN) 구조의 예를 나타내는 도면이다.

- [74] 도 4를 참조하면, 협대역 신호의 특징 백터(401)를 입력으로 광대역 신호의 특징 백터(402)를 추정하는 심화 신경망 생성 모델(410)을 학습할 수 있다. 여기에서 심화 신경망 생성 모델(410)은 Deep Neural Networks (DNN) 이 될 수 있다.
- [75] 추정된 광대역 신호의 특징 백터(402)는 음성 대역폭 확장(Bandwidth Extension, BWE)이 처리된 신호이고, 403은 실제 광대역 신호를 의미한다.
- [76] 도 5는 다른 실시예에 따른 음성 대역폭 확장을 위한 심화 신경망 생성 모델(CNN) 구조의 예를 나타내는 도면이다.
- [77] 도 5를 참조하면, 협대역 신호의 특징 백터(501)를 입력으로 광대역 신호의 특징 백터(502)를 추정하는 심화 신경망 생성 모델(510)을 학습할 수 있다. 여기에서 심화 신경망 생성 모델(510)은 Convolutional Neural Networks (CNN) 이 될 수 있다.
- [78] 다만, CNN으로 심화 신경망 생성 모델(510)을 설계할 경우, 일반적으로 Convolutional Layer, Pooling Layer, Fully Connected Layer로 구성되는 CNN 분류 모델과 달리 컨벌루셔널 레이어(Convolutional Layer)(5 11)와 디컨벌루셔널 레이어(De-Convolutional Layer)(5 12)의 대칭 구조를 사용할 수 있다. 여기에서 컨벌루셔널 레이어(Convolutional Layer)(5 11)는 엔코딩(Encoding) 역할을 수행하며, 디컨벌루셔널 레이어(De-Convolutional Layer)(5 12)는 디코딩(Decoding) 역할을 수행할 수 있다.
- [79] 도 6은 일 실시예에 따른 생성적 대립망 기반의 음성 대역폭 확장을 위한 심화 신경망 분류 모델의 학습 방법을 설명하기 위한 도면이다.
- [80] 도 6을 참조하면, 분류 모델링부(230)는 실제 광대역 신호의 특징 백터(601)와 심화 신경망 생성 모델(610)을 이용하여 추정된 광대역 신호의 특징 백터(603)를 잘 판별할 수 있는 심화 신경망 분류 모델(620)을 학습시킬 수 있다. 추정된 광대역 신호의 특징 백터(603)는 협대역 신호의 특징 백터(602)로부터 음성 대역폭 확장(Bandwidth Extension, BWE)이 처리된 신호이다.
- [81] 여기에서, 심화 신경망 생성 모델(610) 또는 심화 신경망 분류 모델(620)은 DNN(Deep Neural Networks), CNN(Convolutional Neural Networks), RNN(Recurrent Neural Networks) 및 LSTM(Long-Short Term Memory) 중 어느 것으로도 설계가 가능하다.
- [82] 비용 함수는 아래 식과 같이 실제 광대역 신호 x 의 분류 결과인 $D(x)$ 를 1로, 추정된 광대역 신호 $G(z)$ 의 분류 결과인 $D(G(z))$ 를 0으로 판별하도록 설계할 수 있다.
- [83] [식 1]
- [84]
$$\min_D V_{LSGAN}(D) = \frac{1}{2} \mathbb{E}_{x \sim p_{t,a}(x)} [(D(x) - 1)^2] + \frac{1}{2} \mathbb{E}_{z \sim p_z(z)} [D(G(z)) - 2]^2$$
- [85] 여기에서, ζ 는 협대역 신호의 특징 백터이고, χ 는 실제 광대역 신호의

특징 백터일 수 있다. 그리고 $D(x)$ 는 심화 신경망 분류 모델의 결과를 나타내고, $G(z)$ 는 심화 신경망 생성 모델의 결과를 나타낼 수 있다. 또한, $x \sim p(x)$, $z \sim p(z)$ 는 데이터의 샘플링을 의미하며 E 는 기댓값을 나타낼 수 있다. LSGAN 은 Least-Square Generative Adversarial Networks 를 의미하는 것으로, 심화 신경망 분류 모델이 위의 비용 함수를 최소화하도록 모델을 학습시키는 과정이다.

[86] 도 7은 일 실시 예에 따른 생성적 대립 망 기반의 음성 대역폭 확장을 위한 심화 신경망 생성 모델의 재학습 방법을 설명하기 위한 도면이다.

[87] 도 7을 참조하면, 생성 모델링부 (220) 는 학습된 심화 신경망 분류 모델(720)에서 추정된 광대역 신호의 특징 백터(702) 가 실제 광대역 신호로 오판별되도록 심화 신경망 생성 모델(710)을 추가 학습시킬 수 있다. 즉, 생성 모델링부 (220) 는 학습된 심화 신경망 분류 모델(720)에서 추정된 광대역 신호의 특징 백터(702) 가 실제 광대역 신호로 오판별 되도록 심화 신경망 생성 모델(710)을 재학습시킬 수 있다.

[88] 이 때, 생성 모델링 부(220) 는 심화 신경망 분류 모델(720) 은 업데이트 하지 않고 고정된 상태이며, 오직 심화 신경망 생성 모델(710) 만 심화 신경망 분류 모델(720) 을 속일 수 있는 방향으로 학습시킬 수 있다.

[89] 비용 함수는 아래와 같이 $G(z)$ 의 분류 결과인 $D(G(z))$ 가 1에 가까운 값을 가지도록 설계할 수 있다.

[90] [식 ¾]

$$\min_G V_{LSGAN}(G) = \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - 1)^2]$$

[92] 여기에서, ζ 는 협대역 신호의 특징 백터이고, χ 는 실제 광대역 신호의 특징 백터일 수 있다. 그리고 $D(x)$ 는 심화 신경망 분류 모델의 결과를 나타내고, $G(z)$ 는 심화 신경망 생성 모델의 결과를 나타낼 수 있다. 또한, $x \sim p(x)$, $z \sim p(z)$ 는 데이터의 샘플링을 의미하며 E 는 기댓값을 나타낼 수 있다. LSGAN 은 Least-Square Generative Adversarial Networks 를 의미하는 것으로, 심화 신경망 생성 모델의 결과가 심화 신경망 분류 모델에서 실제 광대역 신호로 판별되어 위의 비용 함수가 최소화되도록 모델을 학습시키는 과정이다.

[93]

[94] 아래에서는 고주파 대역 확장 방법의 평가 방법을 설명한다.

[95] 고주파 대역 확장 방법의 주요 평가 방법으로는 MOS(Mean Opinion Score) 가 있다. 실제 목표 신호인 광대역 신호를 비롯하여, 광대역 코텍 기반의 광대역 음성 신호, 협대역 코텍 기반의 협대역 음성 신호, 레퍼런스(Reference) 고주파 대역 확장 방법, 제안하는 고주파 대역 확장 방법을 듣고 평가하는 방식이다. MOS 평가는 크게 ACR(Absolute Category Rating), DCR(Degradation Category Rating) 방식으로 나눌 수 있는데, ACR 은 소리를 한번 듣고 1~5 점의 점수로 소리가 얼마나 좋게 들리는지 평가하는 방법이며, DCR 은 광대역 음성 신호 (5점)를 한 번 듣고, 평가할 소리를 듣고 광대역 신호에 비해 소리가 얼마나 좋지

않은지 1~5 점의 점수로 평가하는 방법이다.

[96] 표 1은 일 실시 예에 따른 음성 대역폭 확장 방법 주관적 평가 방법(MOS)의 예를 나타낸다.

[97] [표 1]

	WB original	AMR-WR@ 12.651 kbps	AMR-NB® 12.2kbps	Ref-BWE	BWE-GAN
Clean	4.96	3.94	2.35	3.15	3.36
Babble (snr-25dB)	4.88	3.87	2.30	2.89	3.11

[98] 표 1에서, MOS 측정 결과 조용한 환경(Clean)에서는 일 실시 예에 따른 기술(BWE-GAN)이 종래 고주파 대역 확장 기술(Ref-BWE)보다 약 0.21 점, 노이즈 환경(Babble)에서는 약 0.22 점이 상승하였으며, 전체 데이터상으로는 상승하는 것을 확인할 수 있다.

[99] 고주파 대역 확장 방법의 보조 평가 방법으로는 PESQ(Perceptual Evaluation of Speech Quality)가 있다. 이 방법은 객관적 평가 방법으로, 알고리즘에 음성 신호를 입력하여 점수로 평가하는 방법이다. 예컨대, 알고리즘에 음성 신호를 입력하여 -0.5~4.5의 점수로 평가할 수 있다. 객관적 평가 방법으로 얻은 점수는 고주파 대역 확장 방법에 있어서 점수와 음질이 비례하지 않는 경향을 보여 보조적 인 평가 방법으로 사용되고 있으며, PESQ 이외에 Segmental SNR 등이 사용될 수 있다.

[100] 표 2는 일 실시 예에 따른 음성 대역폭 확장 방법 객관적 평가 방법(PESQ)의 예를 나타낸다.

[101] [표 2]

	AMR-WR@ 12.651 kbps	扁 -NB@ 12.2kbps	Ref-BWE	BWE-GAN
Clean	3.95	2.68	2.98	3.22
Babble (snr-25dB)	3.72	2.63	2.94	3.07

[102] 표 2에서, PESQ 측정 결과 조용한 환경(Clean)에서는 일 실시 예에 따른 기술(BWE-GAN)이 종래 고주파 대역 확장 기술(Ref-BWE)보다 약 0.24 점, 노이즈 환경(Babble)에서는 약 0.13 점이 상승하였으며, 전체 데이터상으로는 상승하는 것을 확인할 수 있다.

[103] 이와 같이, 실시 예들에 따르면 음성 대역폭 확장기 및 확장 방법을 사용하여,

음성 통신 환경에서 코텍 변환이나 전송비 트율의 확장 없이 더 넓은 대역폭의 음성을 들을 수 있다. 특히, 생성적 대립 망을 이용하여 심화 신경망 기반 생성 모델의 추정 성능을 극대화 함으로써 기존 심화 신경망 기반 생성 모델을 이용한 방법의 한계로 인해 발생되는 사람이 듣기 불편한 인위적인 소리의 합성 없이 향상된 음성 통화 품질을 기대할 수 있으며, 나아가서 음성 인식기에 적용되어 기존 협대역 코텍을 사용했을 때와 같은 전송속도로 더 높은 수준의 음성 인식률을 제공할 수 있다.

- [104] 이러한 본 실시예에 따른 기술은 유무선 음성통신 단말기, 인터넷 망 음성통화 서비스, 모바일 인터넷 망 음성통화 서비스, 음성인식 어플리케이션, 스마트폰 블루투스 통신 시스템, 음성 향상 시스템, 무선 통신 시스템 등에 적용 가능하다.
- [105] 또한, 기존에 협대역 통신망을 사용하는 유무선전화기, 스마트폰, VoIP 등의 통신 시스템에 적용되어 통화 음성 품질을 향상 시킬 수 있으며, 또한 서버 기반의 음성인식기에서 음성 대역폭의 확장을 통한 음성인식률 향상이 가능하다. 더욱 이, 협대역 통신망 기반의 유무선전화기, 스마트폰, VoIP, 블루투스 음성 송수신 시스템, 음성인식기가 필요한 스마트폰 어플리케이션에 적용 가능하다.
- [106]
- [107] 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 운영체제(OS) 및 상기 운영체제 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.
- [108] 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 분산되어 어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[109] 실시 예에 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 실시 예를 위하여 특별히 설계되고 구성된 것들 이거나 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, DVD 와 같은 광기록 매체(optical media), 플롭티 컬 디스크(floptical disk) 와 같은 자기-광 매체(magneto-optical media), 및 럼(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 코드뿐 만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[110] 이상과 같이 실시 예들이 비록 한정된 실시 예와 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.

[111] 그러므로, 다른 구현들, 다른 실시 예들 및 특허청구 범위와 균등한 것들도 후술하는 특허청구 범위의 범위에 속한다.

청구 범위

- [청구 항 1] 음성의 협대역(Narrowband, NB) 신호와 광대역(Wideband, WB) 신호에서 특징 백터를 추출하는 단계;
상기 협대역 신호의 특징 백터로부터 광대역 신호의 특징 백터를 추정하는 단계; 및
추출된 실제 상기 광대역 신호의 특징 백터와 상기 협대역 신호의 특징 백터로부터 추정된 광대역 신호의 특징 백터를 판별하는 심화 신경망 분류 모델을 학습시키는 단계
를 포함하는 생성적 대립망 기반의 음성 대역폭 확장 방법.
- [청구 항 2] 제1항에 있어서,
상기 협대역 신호의 특징 백터로부터 광대역 신호의 특징 백터를 추정하는 단계는,
상기 협대역 신호의 특징 백터를 입력으로 심화 신경망 생성 모델을 학습시켜 추정된 광대역 신호의 특징 백터를 출력하는 것
을 특징으로 하는 생성적 대립망 기반의 음성 대역폭 확장 방법.
- [청구 항 3] 제2항에 있어서,
학습된 상기 심화 신경망 분류 모델에서 상기 추정된 광대역 신호의 특징 백터가 상기 실제 광대역 신호의 특징 백터로 오판별되도록 상기 심화 신경망 생성 모델 재학습시키는 단계
를 더 포함하는 생성적 대립망 기반의 음성 대역폭 확장 방법.
- [청구 항 4] 제3항에 있어서,
상기 심화 신경망 분류 모델에서 상기 추정된 광대역 신호의 특징 백터가 상기 실제 광대역 신호의 특징 백터로 오판별되도록 상기 심화 신경망 생성 모델 재학습시키는 단계는,
상기 심화 신경망 분류 모델이 상기 추정된 광대역 신호의 특징 백터가 상기 실제 광대역 신호의 특징 백터로 오판별하는 것을 목적으로 비용 함수를 설계하여 학습시키는 것
을 특징으로 하는 생성적 대립망 기반의 음성 대역폭 확장 방법.
- [청구 항 5] 제1항에 있어서,
상기 추출된 실제 상기 광대역 신호의 특징 백터와 상기 협대역 신호의 특징 백터로부터 추정된 광대역 신호의 특징 백터를 판별하는 심화 신경망 분류 모델을 학습시키는 단계는,
추출된 실제 상기 광대역 신호의 특징 백터와 상기 협대역 신호의 특징 백터로부터 추정된 광대역 신호의 특징 백터를 판별할 수 있도록 비용 함수를 설계하여 상기 심화 신경망 분류 모델을 학습시키는 것
을 특징으로 하는 생성적 대립망 기반의 음성 대역폭 확장 방법.
- [청구 항 6] 제1항 또는 제2항에 있어서,

상기 심화 신경망 생성 모델 또는 상기 심화 신경망 분류 모델은, DNN(Deep Neural Networks), CNN(Convolutional Neural Networks), RNN(Recurrent Neural Networks) 및 LSTM(Long-Short Term Memory) 중 적어도 어느 하나인 것

을 특징으로 하는 생성적 대립 망 기반의 음성 대역폭 확장 방법.

[청구항 7]

제6항에 있어서,

상기 심화 신경망 생성 모델은,
엔코딩(Encoding) 역할을 수행하는 컨벌루셔널 레이어(Convolutional Layer) 와 디코딩(Decoding) 역할을 수행하는 디컨벌루셔널 레이어(De-Convolutional Layer) 의 대칭 구조의 상기 CNN(Convolutional Neural Networks) 인 것

을 특징으로 하는 생성적 대립 망 기반의 음성 대역폭 확장 방법.

[청구항 8]

상기 심화 신경망 분류 모델을 상기 실제 광대역 신호의 특징 백터와 상기 심화 신경망 생성 모델을 통해 고주파 대역을 생성해 낸 상기 추정된 광대역 신호의 특징 백터를 구분할 수 있도록 충분히 학습시킨 후, 상기 심화 신경망 분류 모델을 고정시키고 상기 심화 신경망 생성 모델이 상기 실제 광대역 신호의 특징 백터와 가까운 고주파 대역을 추정하여 상기 심화 신경망 분류 모델을 오분류하도록 학습시키는 생성적 대립 망을 통해 음성 대역폭을 확장시키는 것

을 특징으로 하는 생성적 대립 망 기반의 음성 대역폭 확장 방법.

[청구항 9]

음성의 협대역(Narrowband, NB) 신호와 광대역(Wideband, WB) 신호에서 특징 백터를 추출하는 특징 백터 추출부;
상기 협대역 신호의 특징 백터를 입력으로 심화 신경망 생성 모델을 학습시켜 추정된 광대역 신호의 특징 백터를 출력하는 생성 모델링부; 및 추출된 실제 상기 광대역 신호의 특징 백터와 상기 협대역 신호의 특징 백터로부터 추정된 광대역 신호의 특징 백터를 판별하는 심화 신경망 분류 모델을 학습시키는 분류 모델링부
를 포함하는 생성적 대립 망 기반의 음성 대역폭 확장 장치.

[청구항 10]

제9항에 있어서,

상기 생성 모델링 부는,
학습된 상기 심화 신경망 분류 모델에서 상기 추정된 광대역 신호의 특징 백터가 상기 실제 광대역 신호의 특징 백터로 오판별되도록 상기 심화 신경망 생성 모델 재학습시키는 것

을 특징으로 하는 생성적 대립 망 기반의 음성 대역폭 확장 장치.

[청구항 11]

제10항에 있어서,

상기 생성 모델링 부는,

상기 심화 신경망 분류 모델이 상기 추정된 광대역 신호의 특징 백터가

상기 실제 광대역 신호의 특징 백터로 오판별하는 것을 목적으로
비용함수를 설계하여 학습시키는 것
을 특징으로 하는 생성적 대립 망 기반의 음성 대역폭 확장 장치.

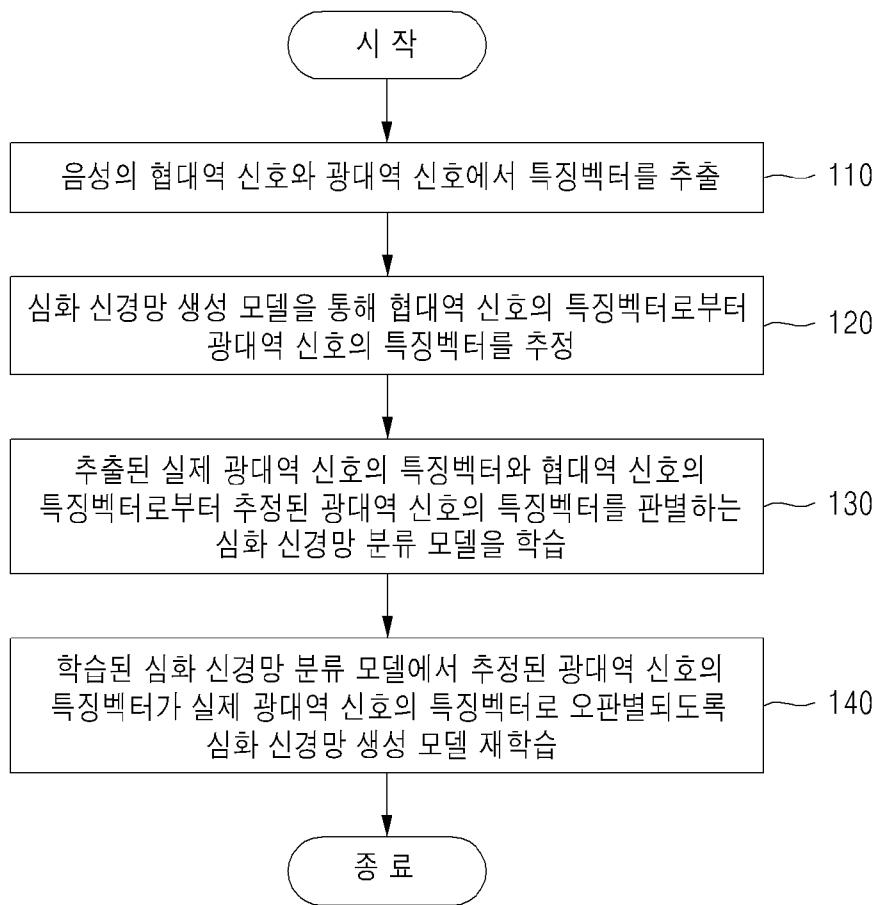
[청구항 12] 제 10항에 있어서,
상기 분류 모델링 부는,
추출된 실제 상기 광대역 신호의 특징 백터와 상기 협대역 신호의
특징 백터로부터 추정된 광대역 신호의 특징 백터를 판별할 수 있도록
비용함수를 설계하여 상기 심화 신경망 분류 모델을 학습시키는 것
을 특징으로 하는 생성적 대립 망 기반의 음성 대역폭 확장 장치.

[청구항 13] 제 10항에 있어서,
상기 심화 신경망 생성 모델 또는 상기 심화 신경망 분류 모델은,
DNN(Deep Neural Networks), CNN(Convolutional Neural Networks),
RNN(Recurrent Neural Networks) 및 LSTM(Long-Short Term Memory) 중
적어도 어느 하나인 것
을 특징으로 하는 생성적 대립 망 기반의 음성 대역폭 확장 장치.

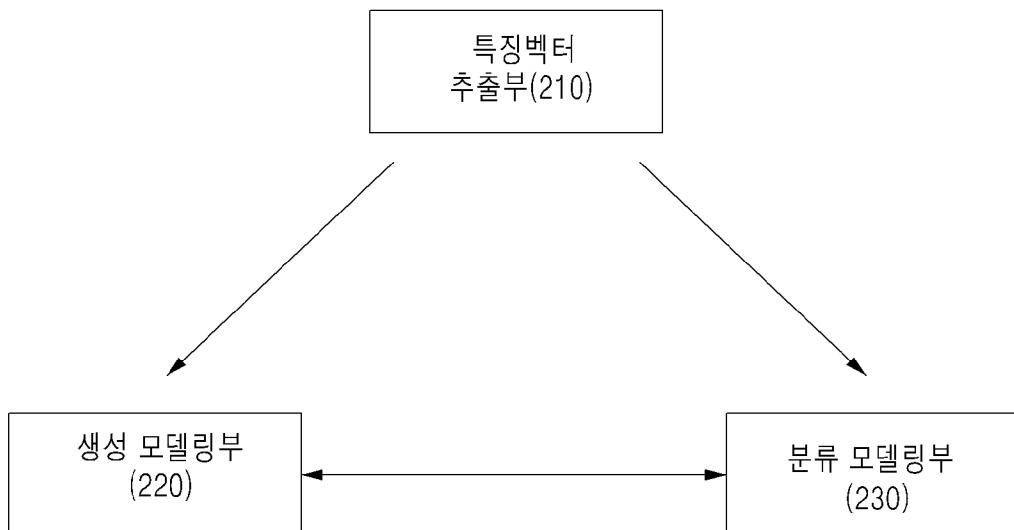
[청구항 14] 제 13항에 있어서,
상기 심화 신경망 생성 모델은,
엔코딩(Encoding) 역할을 수행하는 컨벌루셔널 레이어(Convolutional
Layer) 와 디코딩(Decoding) 역할을 수행하는 디컨벌루셔널
레이어(De-Convolutional Layer) 의 대칭 구조의 상기 CNN(Convolutional
Neural Networks) 인 것
을 특징으로 하는 생성적 대립 망 기반의 음성 대역폭 확장 장치.

[청구항 15] 제 10항에 있어서,
상기 심화 신경망 분류 모델을 상기 실제 광대역 신호의 특징 백터와 상기
심화 신경망 생성 모델을 통해 고주파 대역을 생성해 낸 상기 추정된
광대역 신호의 특징 백터를 구분할 수 있도록 충분히 학습시킨 후, 상기
심화 신경망 분류 모델을 고정시키고 상기 심화 신경망 생성 모델이 상기
실제 광대역 신호의 특징 백터와 가까운 고주파 대역을 추정하여 상기
심화 신경망 분류 모델을 오분류하도록 학습시키는 생성적 대립 망을
통해 음성 대역폭을 확장시키는 것
을 특징으로 하는 생성적 대립 망 기반의 음성 대역폭 확장 장치.

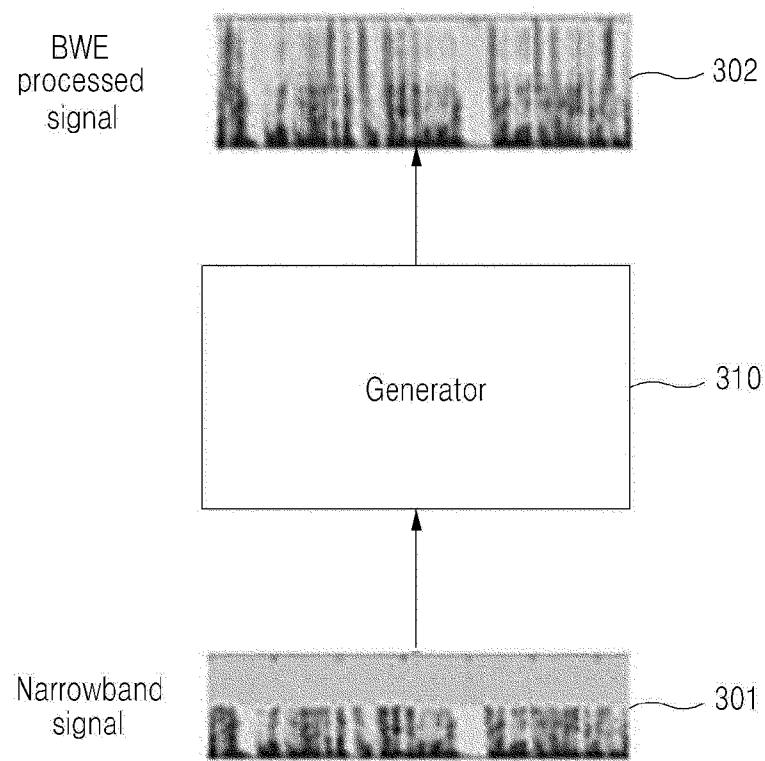
[도1]



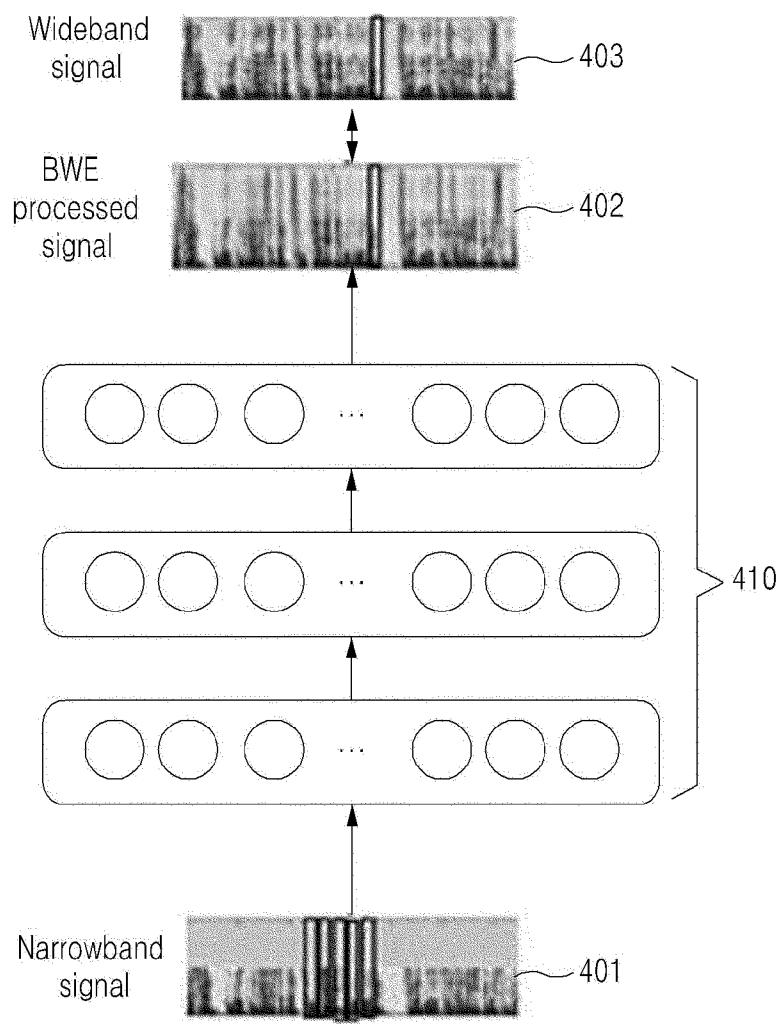
[도2]

200

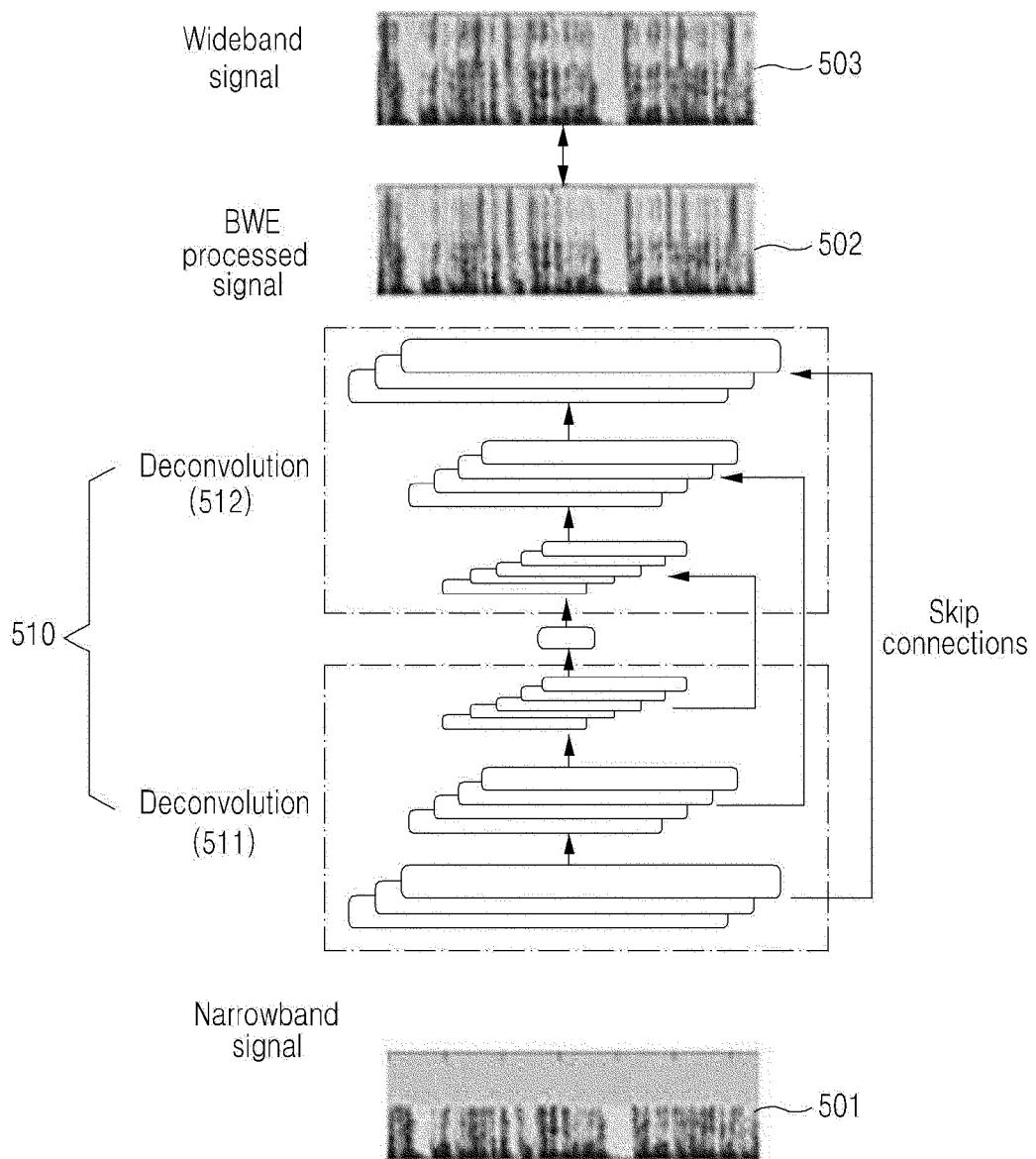
[도3]



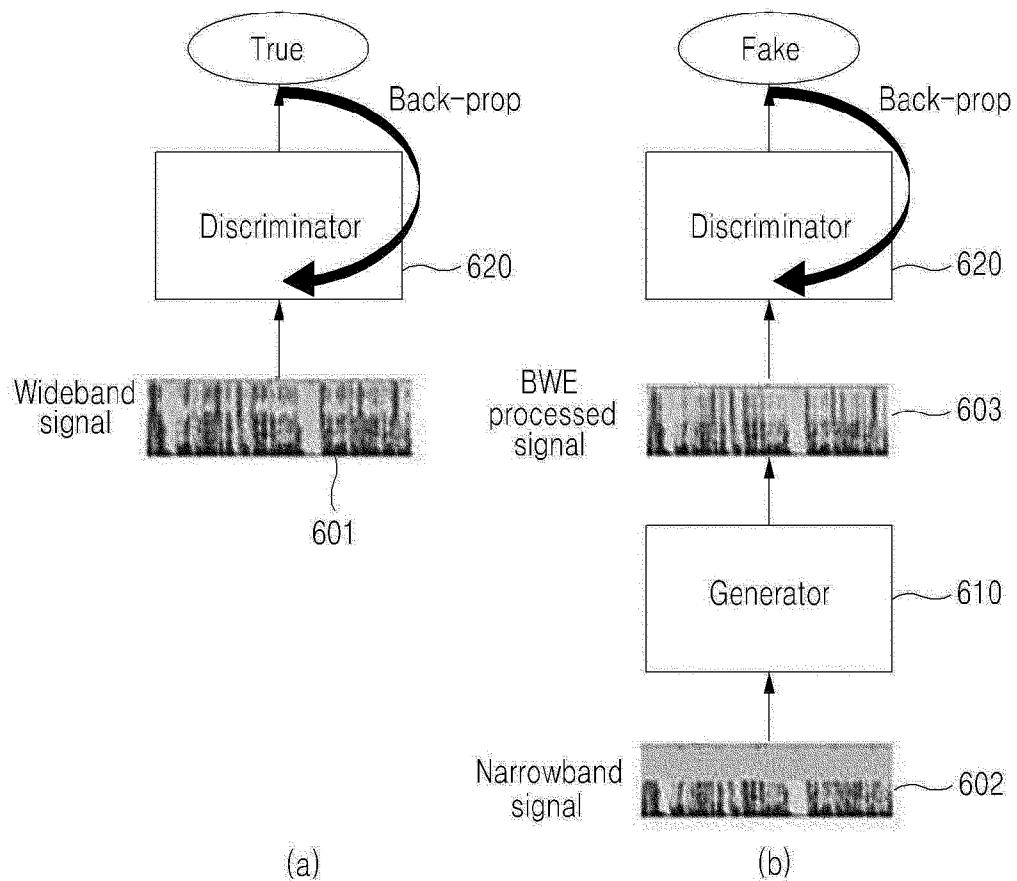
[도4]



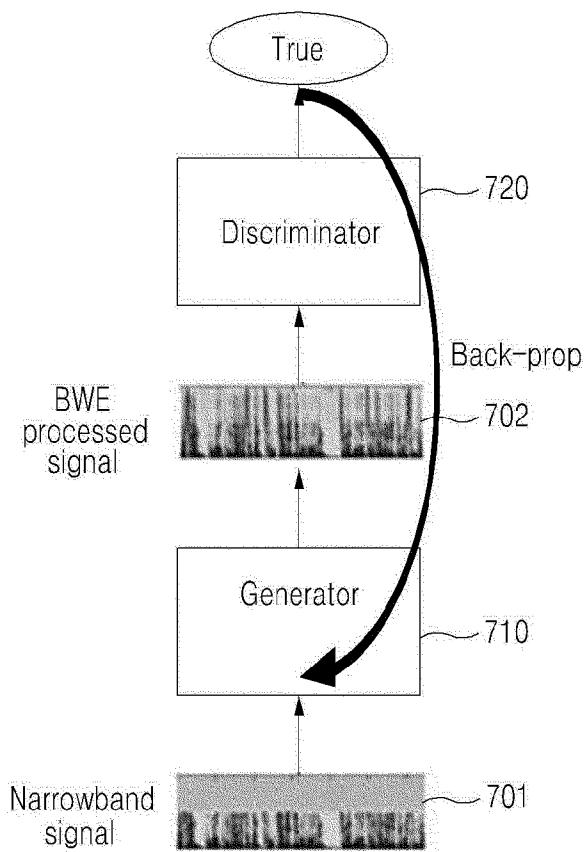
[도5]



[도6]



[도7]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2018/008647

A. CLASSIFICATION OF SUBJECT MATTER

G10L 19/03 S(2013.6) i, G10L 25/3 Q(2013.0) i, G10L 21/038(2013.6) 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)

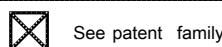
G10L 19/038; G10L 19/07; G10L 13/00; G10L 19/00; G10L 19/04; G10L 25/30; G10L 21/038

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: broadband, neural network, voice, bandwidth expansion, estimation, feature vector

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KR 10-2007-0085982 A (MATSUSHITA ELECTRIC IND. CO., LTD.) 27 August 2007 See paragraphs [005-列] 60], [0080]; claim 17; and figure 1.	1,2,6-7,9
A		3-5,8, 10-15
Y	NOH, Kyoung Jin et al., "EnseabSe of Deep Neural Networks for Artificial Noise Robust Speech Bandwidth Extension", Proceedings of the 2017 Korean Institute of Communications and Information Sciences Summer Conference, The Korean Institute of Communications and Information Sciences, pages 792-793, 2-23 June 2017, Retrieved from the Internet: <URL: http://www.dbpia.co.kr/Journal/ArticleDetail/NODE07218463 >. See pages 792-793, section 13.	1,2,6-7,9
A	JP 2004-252477 A (MITSUBISHI ELECTRIC CORP.) 09 September 2004 See paragraphs [0040]-[0051] and figure 1.	i-15
A	ZHANG, Xiao -Lei et al., "Deep Belief Networks Based Voice Activity Detection", IEEE Transactions on Audio, Speech, and Language Processing, Vol 21, No. 4, pages 697-710, April 2013, Retrieved from the Internet: <URL: http://ieeexplore.ieee.org/document/6362156 >. See pages 697-700, sections 1-2.	i-15
A	JP 2003-526123 A (NOKIA CORP.) 02 September 2003 See paragraphs [0020]-[0023] and figures 3a-3b.	i-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	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"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
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

29 AUGUST 2018 (29.08.2018)

Date of mailing of the international search report

29 AUGUST 2018 (29.08.2018)

Name and mailing address of the ISA/KR



Korean Intellectual Property Office
 Government Complex Daejeon Building 4, 189, Cheongsa-ro, Seo-gu,
 Daejeon, 35208, 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/KR2018/005647

Patent document cited in search report	Publication date	Patent family member	Publication date
KR 10-2007-0085982 A	27/08/2007	AT 520 4 T BR P105 158 14 A C1 10 1076853 A CN 10 1076853 B EP 18 189 13 A1 EP 18 189 13 B1 JP 4903053 B2 RU 2007 121739 A US 2009—0292537 A1 US 8229749 B2 WO 2006-062202 A1	15/08/2011 05/08/2008 21/11/2007 13/10/2010 15/08/2007 10/08/2011 21/03/2012 20/12/2008 26/11/2009 24/07/2012 15/06/2006
JP 2004—252477 A	09/09/2004	JP 3598 111 B2	08/12/2004
JP 2003-526123 A	02/09/2003	AT 343835 T AU 200 1-42539 A1 AU 4253901 A BR 0109043 A BR P10109043 B1 CA 02399253 A1 CA 2399253 C CN 1193344 C CN 141656 1 A Df. 60 124079 T2 EP 1264303 A1 EP 1264303 B1 ES 2274873 T3 F1 1195 6 B F1 20000524 AO JP 2007—156506 A JP 47779 18 B2 KR 10—0535778 B1 PT 1264303 E US 200 1—0027390 A1 US 7483830 B2 WO 01—67437 A1 ZA 200205089 A	15/11/2006 17/09/2001 17/09/2001 03/06/2003 06/06/2017 13/09/2001 23/11/2010 16/03/2005 07/05/2003 08/03/2007 11/12/2002 25/10/2006 01/06/2007 31/12/2008 07/03/2000 21/06/2007 21/09/2011 12/12/2005 31/01/2007 04/10/2001 27/01/2009 13/09/2001 30/04/2003

A. 발 명 이 속 하는 기술분류 (국 제 특 허 분 류(IPC))

GIOL 19/038(2013.01)i, GIOL 25/30(2013.01)i, GIOL 21/038(2013.01)i

B. 조사된 분야

조사된 최소문헌 (국제 특허 분류를 기재)

G10L 19/038 ; G10L 19/07 ; G10L 13/00 ; G10L 19/00 ; G10L 19/04 ; G10L 25/30 ; G10L 21/038

조사된 기술 분야에 속하는 최소문헌 이외의 문헌

한국등록 실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC
 일본등록 실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: 광대역, 신경망, 음성, 대역폭 확장, 추정, 특징벡터

C. 관련 문헌

카 테고 리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y A	KR 10-2007-0085982 A (마츠시타 덴끼 산교 가부시키가이사) 2007. 08.27 단락 [0054]-[0060], [0080] ; 청구항 17; 및 도면 1 참조.	1, 2, 6-7, 9 3-5, 8, 10-15
Y	노경진 등, '인공 잡음에 강인한 음성 대역폭 확장을 위한 심화 신경망 양상' 불., 2017년도 한국통신학회 학계종 학술 대회 논문집, 한국통신학회, 페이지 792-793, 2017. 06. 21-23 <URL: http://w.ripia.co.kr/Journal/ArticleDetail.asp?NODE07218463>에서 검색됨. 페이지 792-793, 섹션 1-1 참조.	1, 2, 6-7, 9
A	JP 2004-252477 A (MITSUBISHI ELECTRIC CORP.) 2004. 09. 09 단락 [0040]-[0051] 및 도면 1 참조.	1-15
A	XIAO-LEI ZHANG 등, 'Deep Belief Networks Based Voice Activity Detection', IEEE Transactions on Audio, Speech, and Language Processing, Vol. 21, No. 4, 페이지 697-710, 2013. 04 <URL: http://ieeexplore.ieee.org/document/63621866>에서 검색됨. 페이지 697-700, 섹션 1-2 참조.	1-15
A	JP 2003-526123 A (NOKIA CORP.) 2003. 09. 02 단락 [0020]-[0023] 및 도면 3a-3b 참조.	1-15

추가 문헌이 C(계속)에 기재되어 있습니다.

* 대응 특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

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

"1" 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

"&" 동일한 대응특허 문헌에 속하는 문헌

"P" 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

국제조사의 실제 완료일

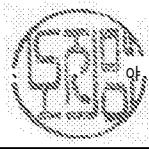
2018년 08월 29일 (29.08.2018)

국제조사 보고서 발송일

2018년 08월 29일 (29.08.2018)

ISA/KR 의 명 칭 ④ ④ 주 소
대한민국 특허청
(35208) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)
팩스 번 호 +82-42-481-8578

심사관
노지명
전화번호 +82-42-481-8528



국 제 조 사 보 고 서
대용 특허에 관한 정보

국제출 원번호
PCT/KR2018/005647

국 제 조 사 보 고 서 에서
인용된 특허문헌

공 개 일

대용 특 허문헌

공 개 일

KR 10-2007-0085982 A	2007/08/27	AT 520124 T BR PI0515814 A CN 101076853 A CN 101076853 B EP 1818913 AI EP 1818913 BI JP 4903053 B2 RU 2007121739 A US 2009-0292537 AI us 8229749 B2 wo 2006-062202 AI	2011/08/15 2008/08/05 2007/11/21 2010/10/13 2007/08/15 2011/08/10 2012/03/21 2008/12/20 2009/11/26 2012/07/24 2006/06/15
JP 2004-252477 A	2004/09/09	JP 3598111 B2	2004/12/08
JP 2003-526123 A	2003/09/02	AT 343835 T AII 2001-42539 AI AII 4253901 A BR 0109043 A BR PI0109043 BI CA 02399253 AI CA 2399253 C CN 1193344 C CN 1416561 A DE 60124079 T2 EP 1264303 AI EP 1264303 BI ES 2274873 T3 FI 119576 B FI 20000524 AO JP 2007-156506 A JP 4777918 B2 KR 10-0535778 BI PT 1264303 E US 2001-0027390 AI us 7483830 B2 wo 01-67437 AI ZA 200205089 A	2006/11/15 2001/09/17 2001/09/17 2003/06/03 2017/06/06 2001/09/13 2010/11/23 2005/03/16 2003/05/07 2007/03/08 2002/12/11 2006/10/25 2007/06/01 2008/12/31 2000/03/07 2007/06/21 2011/09/21 2005/12/12 2007/01/31 2001/10/04 2009/01/27 2001/09/13 2003/04/30