



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0033852
(43) 공개일자 2021년03월29일

(51) 국제특허분류(Int. Cl.)
G10L 19/02 (2006.01) G06N 20/00 (2019.01)
G10L 19/16 (2013.01) G10L 19/24 (2013.01)
(52) CPC특허분류
G10L 19/0204 (2013.01)
G06N 20/00 (2019.01)
(21) 출원번호 10-2019-0115686
(22) 출원일자 2019년09월19일
심사청구일자 2019년09월19일

(71) 출원인
에스케이텔레콤 주식회사
서울특별시 중구 을지로 65 (을지로2가)
광주과학기술원
광주광역시 북구 첨단과기로 123 (오룡동)
(72) 발명자
나태영
서울특별시 중구 을지로 65 SKT-타워
이정혁
전라남도 장성군 남면 나노산단1로 58-12, 더 힐
링 201호
김홍국
광주광역시 광산구 첨단중앙로181번길 42-4 금광
아파트 101동 1101호
(74) 대리인
이철희

전체 청구항 수 : 총 13 항

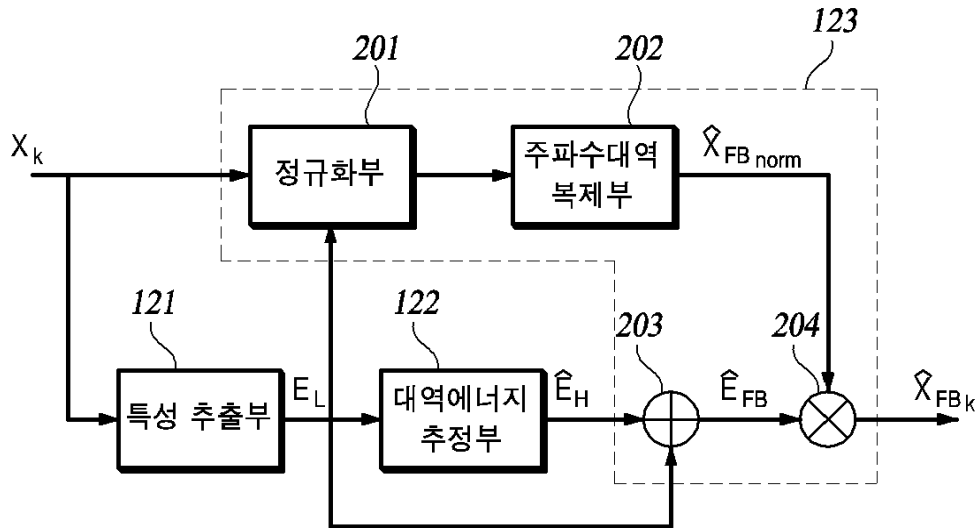
(54) 발명의 명칭 학습 모델 기반의 인공 대역 변환장치 및 방법

(57) 요약

학습 모델 기반의 인공 대역 변환장치 및 방법을 개시한다.

전송 채널의 제약으로 인하여 고압축률로 압축되어 모바일 디바이스로 전송된 저품질의 음원 데이터를 고품질 음원과 등가의 신호로 변형하기 위하여, 학습 모델 기반의 에너지 추정 방법을 이용하는 인공 대역 변환장치 및 방법을 제공하는 데 목적이 있다.

대표도 - 도2



(52) CPC특허분류

G10L 19/167 (2013.01)

G10L 19/24 (2013.01)

명세서

청구범위

청구항 1

저품질 신호에 대응하는 주파수 영역의 데이터를 획득하여 부대역(sub-band) 별로 저대역(low frequency band) 에너지를 추출하는 특성추출부;

사전에 트레이닝된 학습형 연산 모델을 포함하고, 상기 저대역 에너지를 상기 학습형 연산 모델에 입력하여 고대역(high frequency band) 에너지를 추정하는 대역에너지 추정부; 및

상기 고대역 에너지, 상기 저대역 에너지 및 상기 주파수 영역의 데이터를 이용하여 고품질 전대역(full frequency band) 데이터를 추정하는 재구성부

를 포함하는 것을 특징으로 하는 인공 대역 변환장치.

청구항 2

제1항에 있어서,

상기 특성추출부는,

상기 주파수 영역의 데이터에 대비하여 축소된 차원의 저대역 에너지를 추출하는 것을 특징으로 하는 인공 대역 변환장치.

청구항 3

제1항에 있어서,

상기 학습형 연산 모델은 딥러닝(deep learning) 모델로서, 동일 음원 신호로부터 추출된 주파수 영역의 고품질 전대역 에너지 및 주파수 영역의 저품질 저대역 에너지 각각을 타겟(target) 및 입력으로 이용하여 사전에 트레이닝되는 것을 특징으로 하는 인공 대역 변환장치.

청구항 4

제3항에 있어서,

상기 딥러닝 모델은,

복수의 전연결 레이어 중 하나의 전연결 레이어를 제외한 나머지 전연결 레이어가 입력 레이어와 출력 레이어 사이에서 서로 직렬로 연결되는 것을 특징으로 하는 인공 대역 변환장치.

청구항 5

제4항에 있어서,

상기 딥러닝 모델은,

상기 하나의 전연결 레이어가 상기 입력 레이어와 상기 출력 레이어를 연결하여 상기 고대역 에너지와 연관성이 높은 저대역 에너지의 정보를 상기 출력 레이어 측으로 전달하는 것을 특징으로 하는 인공 대역 변환장치.

청구항 6

제4항에 있어서,

상기 딥러닝 모델은,

상기 복수의 전연결 레이어의 출력에 적용되는 활성화 함수(activation function)로서 ELU(Exponential Linear Unit)를 사용하는 것을 특징으로 하는 인공 대역 변환장치.

청구항 7

제1항에 있어서,
 상기 재구성부는,
 상기 저대역 에너지 및 상기 주파수 영역의 데이터를 처리하여 평준화된 저대역 데이터를 생성하는 정규화부;
 상기 평준화된 저대역 데이터를 이용하여 평준화된 전대역 데이터를 생성하는 주파수대역 복제부;
 상기 고대역 에너지와 상기 저대역 에너지를 합성하여 전대역 에너지를 생성하는 에너지 합성부; 및
 상기 평준화된 전대역 데이터 및 상기 전대역 에너지를 처리하여 상기 고품질 전대역 데이터를 추정하는 전대역 데이터 추정부
 를 포함하는 것을 특징으로 하는 인공 대역 변환장치.

청구항 8

제7항에 있어서,
 상기 주파수대역 복제부는,
 상기 평준화된 저대역 데이터 간의 상관도에 기반하여 상기 고대역 데이터를 복원한 후, 복원된 고대역 데이터와 상기 평준화된 저대역 데이터를 합성하여 상기 평준화된 전대역 데이터를 생성하는 것을 특징으로 하는 인공 대역 변환장치.

청구항 9

제7항에 있어서,
 상기 에너지 합성부는,
 상기 대역에너지 추정부가 추정한 고대역 에너지와 상기 특성추출부에서 추출한 저대역 에너지를 합성하여 상기 전대역 에너지를 생성하는 것을 특징으로 하는 인공 대역 변환장치.

청구항 10

제7항에 있어서,
 상기 전대역데이터 추정부는,
 주파수 영역에서의 곱셈을 이용하여 상기 평준화된 전대역 데이터를 상기 전대역 에너지로 마스킹(masking)함으로써 상기 고품질 전대역 데이터를 추정하는 것을 특징으로 하는 인공 대역 변환장치.

청구항 11

인공 대역 변환장치의 인공 대역 변환방법에 있어서,
 저품질의 주파수 영역 데이터를 이용하여 부대역(sub-band) 별로 저대역(low frequency band) 에너지를 추출하는 과정;
 상기 저대역 에너지를 사전에 트레이닝된 학습 모델에 입력하여 고대역 에너지(high frequency band)를 추정하는 과정;
 상기 저대역 에너지 및 상기 주파수 영역의 데이터를 이용하여 평준화된 전대역(full frequency band) 데이터를 생성하는 과정; 및
 상기 고대역 에너지와 상기 저대역 에너지를 합성하여 전대역 에너지를 생성한 후, 상기 평준화된 전대역 데이터와 상기 전대역 에너지를 처리하여 고품질 전대역 데이터를 추정하는 과정
 을 포함하는 것을 특징으로 하는, 컴퓨터 상에 구현되는 인공 대역 변환방법.

청구항 12

제11항에 있어서,

상기 고대역 에너지를 추정하는 과정은,

동일 음원 신호로부터 추출된 주파수 영역의 고품질 전대역 에너지 및 주파수 영역의 저품질 저대역 에너지를 기반으로 사전에 트레이닝되는 상기 학습 모델을 이용하는 것을 특징으로 하는, 컴퓨터 상에 구현되는 인공 대역 변환방법.

청구항 13

제11항 및 제12항에 따른 인공 대역 변환방법의 각 단계를 실행시키기 위하여 컴퓨터로 읽을 수 있는, 비휘발성 또는 비일시적인 기록매체에 저장된 컴퓨터프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 모바일 디바이스에서 이용되는 학습 모델 기반의 인공 대역 변환장치 및 방법에 관한 것이다.

배경 기술

[0002] 이하에 기술되는 내용은 단순히 본 발명과 관련되는 배경 정보만을 제공할 뿐 종래기술을 구성하는 것이 아니다.

[0003] 인공 대역 변환은, 제한된 대역폭 혹은 하드웨어 성능 등의 원인에 의해 낮은 전송률로 수신된 저품질 신호에 대하여, 저주파수 대역(low frequency band, 이하 "저대역") 분석에 기반하여 고주파수 대역(high frequency band, 이하 "고대역")을 생성함으로써 신호 품질을 향상시키는 기술로서 ABE(Artificial Bandwidth Extension)로도 불리운다. 뮤직 음원신호 처리 분야에서 ABE는 데이터의 압축을 위한 인코더와 디코더를 통칭한 코덱(codec) 시스템에 적용되어 모바일 디바이스 상에서 작동되며, 주로 온라인 상황에서 품질 향상을 위해 추가적인 후처리(post-processing) 형태로 실행된다.

[0004] ABE를 구현한 종래의 기술로는, G.729.1에서와 같이 고대역의 정보를 따로 압축하여 모바일 디바이스로 전송하는 방식이 있다(비특허문헌 1 참조). 대역폭 확장을 위해 고대역의 정보를 이용한다는 장점은 있으나, 추가적인 비트 할당을 위한 대역을 전송 채널에 요구하고, 사용되는 코덱의 복잡도를 증가시키는 단점이 있다.

[0005] 다른 기술로는 HMM(Hidden Markov Model) 및 GMM(Gaussian Mixture Model) 등 패턴인식(pattern recognition) 기반의 추정(estimation) 방식이 있다. 그러나 패턴인식 기반의 추정 방식은 목표로 하는 성능을 달성하기 위하여 모델 학습에 긴 시간을 요구하고, 학습 결과를 적용한 추론 과정 또한 복잡하다. 최근에 SoC(System on Chip) 기술의 발달에 따라 종전에 비해 추론 과정에 대한 실시간 처리 여건이 좋아졌음에도 불구하고 패턴인식 기반의 추정 방식은 모바일 디바이스에서 실시간 처리가 어렵다는 단점이 있다.

[0006] 따라서 전송 채널에 부담을 주지 않은 채로, 모바일 디바이스 상에서 실시간 처리가 가능하면서도, SNR(Signal to Noise Ratio)로 대표되는 성능은 향상된 음원 신호처리를 위한 인공 대역 변환장치 및 방법이 요구된다.

선행기술문헌

비특허문헌

[0007] (비특허문헌 0001) 비특허문헌 1: ITU-T (January 2007). "G.729: Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)"

발명의 내용

해결하려는 과제

[0008] 본 개시는, 전송 채널의 제약으로 인하여 고압축률로 압축되어 모바일 디바이스로 전송된 저품질의 음원 데이터

를 고품질 음원과 등가의 신호로 변형하기 위하여, 학습 모델 기반의 에너지 추정 방법을 이용하는 인공 대역 변환장치 및 방법을 제공하는 데 주된 목적이 있다.

과제의 해결 수단

- [0009] 본 발명의 실시예에 의하면, 저품질 신호를 표현한 주파수 영역의 데이터를 획득하여 부대역(sub-band) 별로 저대역(low frequency band) 에너지를 추출하는 특성추출부; 사전에 트레이닝된 학습형 연산 모델을 포함하고, 상기 저대역 에너지를 상기 학습형 연산 모델에 입력하여 고대역(high frequency band) 에너지를 추정하는 대역에너지 추정부; 및 상기 고대역 에너지, 상기 저대역 에너지 및 상기 주파수 영역의 데이터를 이용하여 전대역(full frequency band) 주파수 데이터를 추정하는 재구성부를 포함하는 것을 특징으로 하는 인공 대역 변환장치를 제공한다.
- [0010] 본 발명의 다른 실시예에 의하면, 인공 대역 변환장치의 인공 대역 변환방법에 있어서, 주파수 영역의 데이터를 이용하여 부대역(sub-band) 별로 저대역(low frequency band) 에너지를 추출하는 과정; 상기 저대역 에너지를 사전에 트레이닝된 학습 모델에 입력하여 고대역 에너지(high frequency band)를 추정하는 과정; 상기 저대역 에너지 및 상기 주파수 영역의 데이터를 이용하여 평균화된 전대역(full frequency band) 데이터를 생성하는 과정; 및 상기 고대역 에너지와 상기 저대역 에너지를 합성하여 전대역 에너지를 생성한 후, 상기 평균화된 전대역 데이터와 상기 전대역 에너지를 처리하여 전대역 주파수 데이터를 추정하는 과정을 포함하는 것을 특징으로 하는, 컴퓨터 상에 구현되는 인공 대역 변환방법을 제공한다.
- [0011] 본 발명의 다른 실시예에 의하면, 인공 대역 변환방법의 각 단계를 실행시키기 위하여 컴퓨터로 읽을 수 있는, 비휘발성 또는 비일시적인 기록매체에 저장된 컴퓨터프로그램을 제공한다.

발명의 효과

- [0012] 이상에서 설명한 바와 같이 본 실시예에 의하면, 전송 채널의 제약으로 인하여 고압축률로 압축되어 전송된 저품질의 음원 데이터를 고품질 음원 신호로 변형하기 위하여, 학습 모델 기반의 에너지 추정 방법을 이용하는 인공 대역 변환장치 및 방법을 제공함으로써, 모바일 디바이스에서 고품질 음원과 등가의 음질을 서비스하여 사용자의 만족도를 높이는 효과가 있다.

도면의 간단한 설명

- [0013] 도 1은 본 발명의 일 실시예에 따른, 모바일 디바이스 상에 구현되는 인공 대역 변환기의 구성도이다.
- 도 2는 본 발명의 일 실시예에 따른 대역폭 확장부의 구성도이다.
- 도 3은 본 발명의 일 실시예에 따른 고대역 에너지 추정을 위한 학습 모델의 구조도이다.
- 도 4는 본 발명의 일 실시예에 따른 인공 대역 변환방법에 대한 순서도이다.

발명을 실시하기 위한 구체적인 내용

- [0014] 이하, 본 발명의 실시예들을 예시적인 도면을 참조하여 상세하게 설명한다. 각 도면의 구성요소들에 참조부호를 부가함에 있어서, 동일한 구성요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부호를 가지도록 하고 있음에 유의해야 한다. 또한, 본 실시예들을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 실시예들의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.
- [0015] 또한, 본 실시예들의 구성 요소를 설명하는 데 있어서, 제 1, 제 2, A, B, (a), (b) 등의 용어를 사용할 수 있다. 이러한 용어는 그 구성 요소를 다른 구성 요소와 구별하기 위한 것일 뿐, 그 용어에 의해 해당 구성 요소의 본질이나 차례 또는 순서 등이 한정되지 않는다. 명세서 전체에서, 어떤 부분이 어떤 구성요소를 '포함', '구비'한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 '...부', '모듈' 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.
- [0016] 첨부된 도면과 함께 이하에 개시될 상세한 설명은 본 발명의 예시적인 실시형태를 설명하고자 하는 것이며, 본 발명이 실시될 수 있는 유일한 실시형태를 나타내고자 하는 것이 아니다.
- [0017] 이하의 설명에서 음원의 품질은 음원의 압출률(kbps로 표시)과 반비례하는 것으로 가정한다. 즉 압출률이 높을

수록 음원의 품질은 낮은 것으로 가정한다. 또한 낮은 품질의 음원은 높은 품질의 음원에 대하여 상대적으로 저대역(low frequency band) 데이터인 것으로 가정한다.

- [0018] 도 1은 본 발명의 일 실시예에 따른, 모바일 디바이스 상에 구현되는 인공 대역 변환기의 구성도이다.
- [0019] 도 1에 예시된 인공 대역 변환기(110)는 입력부(111), 대역폭 확장부(112) 및 출력부(113)의 전부 또는 일부를 포함한다. 여기서, 본 실시예에 따른 인공 대역 변환기(110)에 포함되는 구성요소가 반드시 이에 한정하는 것은 아니다. 예컨대, 인공 대역 변환기(110) 상에 학습 모델의 트레이닝을 위한 트레이닝부(미도시)를 추가로 구비하거나, 외부의 트레이닝부와 연동되는 형태로 구현될 수 있다.
- [0020] 입력부(111)는 인공 대역폭 확장(Artificial Bandwidth Extension: ABE)을 수행하는 과정에서 필요한 데이터를 획득하여 ABE에 적절한 형태로 변환한다.
- [0021] 예컨대, 본 실시예에 따른 입력부(111)는 전송된 고압축률의 음원을 스트리밍(streaming) 서비스 시스템으로부터 수신한 후, 주파수 영역(frequency domain)의 데이터로 변환(transformation)한다. 변환 방법으로는 MDCT(Modified Discrete Cosine Transform), FFT(Fast Fourier Transform), 또는 셉스트럼(cepstrum) 변환 등을 사용할 수 있으나, 반드시 이에 한정하는 것은 아니다.
- [0022] 본 실시예에서, 입력부(111)는 고압축률의 음원 스트리밍(streaming) 데이터를 수신한 후, 우선 코덱(codec)을 이용하여 시간 영역(time domain)의 데이터를 합성(synthesis)한다. 시간 영역 데이터의 합성 시에 샘플링 주파수(sampling frequency)는 44.1 KHz를 사용하는 것으로 가정한다. 다음, 시간 영역 상의 음원 데이터에 대하여, 수행 구간을 겹쳐가면서(overlapping and sliding) MDCT를 진행하여 주파수 영역 상의 데이터 X_k 를 생성한다.
- [0023] 예컨대, 본 실시예에서, 고압축률의 음원은 64 kbps, 96 kbps 및 128 kbps 중의 하나로 압축된 음원을 의미하나 반드시 이에 한정하는 것은 아니다. ABE의 목표가 되는 고품질 음원은 320 kbps 음원 혹은 그 등가 품질의 음원을 의미하나 반드시 이에 한정하는 것을 아니며, 고압축률의 음원보다 낮은 압축률을 가지며 스트리밍 서비스에서 제공받을 수 있는 최상의 음질을 갖는 음원 중의 하나일 수 있다.
- [0024] 또한, 본 발명의 다른 실시예에 따른 입력부(111)는, 시간 영역 상의 데이터 형태로 존재하는 저품질 음원 데이터를 획득할 수 있다.
- [0025] 대역폭 확장부(112)는 저품질 음원에 대한 인공 대역 변환을 실행한다.
- [0026] 본 실시예에 따른 대역폭 확장부(112)는, 특성추출부(121), 대역에너지 추정부(122) 및 재구성부(123)를 포함한다. 대역폭 확장부(112)는 입력부(111)에 의하여 변환된 주파수 영역의 데이터를 입력으로 받아들인 후, ABE를 실행하여 주파수 영역 상의 데이터로 변환한다. 대역폭 확장부(112)의 구조 및 동작은 추후에 설명하기로 한다.
- [0027] 출력부(113)는 변환된 음원을 인공 대역 변환기(110)의 사용자에게 청각적인 형태(audible form)로 제공한다.
- [0028] 본 실시예에 따른 출력부(113)는, 대역폭 확장부(112)로부터 주파수 영역에서 변형된 데이터를 수신한 후, 합성 과정을 통하여 시간 영역에서의 확장된 음원으로 변환한다. 합성 과정은 IMDCT(Inverse MDCT), IFFT(Inverse FFT), 또는 역셉스트럼(inverse cepstrum) 변환과 같은, 주파수 영역의 데이터를 시간 영역의 데이터로 변환하는 과정을 포함할 수 있다. 최종적으로 시간 영역 상의 확장된 데이터를 청각적인 형태로 출력한다.
- [0029] 본 실시예에 따른 인공 대역 변환기(110)는 모바일 디바이스(100) 상에 구현될 수 있다. 모바일 디바이스(100)는 전송대역의 제한을 받는 무선 또는 유선 전송로를 이용하여 음원 스트리밍서비스를 제공받을 수 있는 어느 기기든 될 수 있다.
- [0030] 도 2는 본 발명의 일 실시예에 따른 대역폭 확장부의 구성도이다.
- [0031] 대역폭 확장부(112)는 저품질 음원에 대한 인공 대역 변환을 실행한다. 대역폭 확장부(112)는 특성추출부(121), 대역에너지 추정부(122) 및 재구성부(123)의 전부 또는 일부를 포함한다. 재구성부(123)는 정규화부(201), 주파수대역 복제부(202, Spectral Band Replication: SBR), 에너지 합성부(203) 및 전대역데이터 추정부(204)를 포함하나, 반드시 이에 한정하는 것은 아니다. 도 2의 도시는 본 실시예에 따른 예시적인 구성이며, 주파수 영역에서의 데이터 처리 기법에 의거하여 다른 구성요소 또는 구성요소 간의 다른 연결을 포함하는 구현이 가능하다.

[0032] 특성추출부(121)는, 입력부(111)에 의하여 변환되어 주파수 영역의 데이터 X_k 로 표현된 음원 신호를 입력으로 받아들이며 부대역(sub-band) 별로 저대역(low frequency band) 에너지를 추출한다. 추출하는 방법은 수학적 식 1에 따른다. 수학적 식 1에 따르면 하나의 부대역 에너지를 구하기 위해 SB의 2 배수에 해당하는 주파수 영역의 데이터를 이용하고, 각 부대역은 서로 50 %씩 겹쳐져(overlapping) 있다. 전술한 바와 같은 부대역 에너지를 구하는 방법은 하나의 예시이며, SB 및 수학적 식 1을 변형하여 다양한 방법으로 부대역 에너지를 구하는 것이 가능하다.

수학적 식 1

$$E(b) = \sqrt{\sum_{k=SB \cdot b+1}^{SB \cdot (b+2)} X_k^2}, \quad b = 0, 1, \dots, N_L$$

[0033]

[0034] $E(b)$ 를 구성요소로 포함하는 저대역 에너지 벡터 E_L 은 대역에너지 추정부(122)에 입력으로 전달된다.

따라서 E_L 은 N_L 개의 부대역 에너지를 포함하고, N_L 은 확장 대상 음원의 압축률에 의존한다.

[0035] 전술한 바와 같이, 본 실시예에서는 주파수 영역의 데이터 X_k 를 처리하여 축소된 차원의 데이터인 저대역 에너지 벡터를 대역에너지 추정부(122)에 제공함으로써, 대역에너지 추정부(122) 구현 시 복잡도를 감소시키고, 감소된 복잡도에 따라 모바일 디바이스(100) 상에서의 실시간 동작 가능성 및 실시간 동작 성능을 높일 수 있다.

[0036] 대역에너지 추정부(122)는 저대역 에너지 E_L 을 입력으로 받아들이며 고대역(high frequency band) 에너지

\hat{E}_H 를 추정한다. 대역에너지 추정부(122)는 저대역 에너지를 고대역 에너지 추정에 직접적으로 이용하기 때문에 주파수 영역 상의 음원 특성을 더 강조하는 고대역 에너지 추정이 가능하다. 대역에너지 추정부(122)는 고대역 에너지 추정을 위하여 트레이닝부에 의하여 기 학습된 신경회로망(Neural Network) 기반의 학습 모델을 이용한다. 학습 모델의 구조 및 학습 모델의 트레이닝 과정은 추후에 설명하기로 한다.

[0037] 재구성부(123)는 대역에너지 추정부(122)가 추정한 고대역 에너지, 특성추출부(121)가 추출한 저대역 에너지, 및 주파수 영역의 데이터 X_k 를 이용하여 전대역(full frequency band) 주파수 데이터를 추정한다.

[0038] 정규화부(201)는 부대역 별로 추출된 저대역 에너지 및 주파수 영역의 데이터를 처리하여 평준화된 저대역 데이터를 생성한다. 평준화를 적용하는 이유는, 주파수대역 복제부(202)에서 상관도(correlation)를 구할 때 발생하는 편차를 축소할 수 있기 때문이다.

[0039] 주파수대역 복제부(202)는 평준화된 저대역 데이터를 이용하여 고대역 데이터를 복원한다. 복원 과정에서는 부대역 순서대로 복원이 이루어지는데, 현재까지 복원된 부대역의 다음(고주파수 방향) 부대역이 복원 대상이 된다. 복원 대상 부대역의 바로 앞에 위치하는 부대역과 가장 상관도가 높은 저주파수 데이터 영역을 찾은 후, 찾아진 영역의 데이터를 복원 대상 부대역에 복사한다. 여기서, 바로 앞에 위치하는 부대역은, 복원 과정의 첫 단계에서는 저대역 내의 마지막 부대역이고, 이후 단계에서는 직전에 복원된 부대역이다. 주파수대역 복제부(202)는 복원된 고대역 데이터와 평준화된 저대역 데이터를 합성하여 평준화된 전대역 데이터 \hat{X}_{FBnorm} 를 생성한다.

[0040] 한편, 에너지 합성부(203)는 대역에너지 추정부(122)가 추정한 고대역 에너지 \hat{E}_H 와 특성추출부(121)에서

추출한 저대역 에너지 E_L 을 합성하여 전대역 에너지 \tilde{E}_{FB} 를 생성한다.

[0041] 전대역데이터 추정부(204)는 평균화된 전대역 데이터 \hat{X}_{FBnorm} 과 전대역 에너지 \tilde{E}_{FB} 를 곱하여 전대역

주파수 데이터 \hat{X}_{FBk} 를 추정한다. 주파수 영역에서의 곱셈을 수행하므로, 시간 영역에서의 컨볼루션(convolution) 과정을 수행하는 것과 동일하다. 따라서 곱셈 과정은 평균화된 주파수 데이터를 부대역 별 에너지에 해당하는 계수를 갖는 필터를 이용하여 필터링하는 것과 동일한 과정이며, 달리 말하면 평균화된 전대역 데이터를 전대역 에너지로 마스킹(masking)하는 과정이다.

[0042] 도 3은 본 발명의 일 실시예에 따른 고대역 에너지 추정을 위한 학습 모델의 구조도이다.

[0043] 도 3에 표시된 학습 모델은 딥러닝(deep learning) 모델을 기반으로 한다. 이하 도 3을 참조하여, 본 실시예에 따른 딥러닝 모델의 구조 및 학습 과정을 설명한다.

[0044] 딥러닝 모델은 입력 레이어(input layer), 3 개의 전연결 레이어(fully-connected layer 또는 dense layer, 이하 제1 전연결 레이어, 제2 전연결 레이어 및 제3 전연결 레이어) 및 출력 레이어(output layer)를 포함하나, 반드시 이에 한정하는 것은 아니며, 트레이닝에 소요되는 시간, 대역에너지 추정부(122)의 성능 및 실시한 구현 가능성 등에 따라 각 레이어를 구성하는 노드의 개수, 노드의 출력에 대한 활성화 함수(activation function) 및 전연결 레이어의 개수 등에 대한 구조 변형이 가능하다.

[0045] 본 실시예에 따른 딥러닝 모델은 64k bps, 96 kbps 및 128 kbps의 음원 압축률에 따라, 각각 정해진 11, 16 및 23 개의 부대역 에너지 입력에 대해 31 개의 부대역 에너지를 출력한다. 따라서 입력 레이어는 11, 16 및 23 개

의 부대역 에너지 입력에 해당하는 수(N_L)만큼의 노드(node)로 구성되는 한편, 출력 레이어는 31 개의 부대역 에너지에 해당하는 31 개(N_F)의 노드를 포함한다. 전연결 레이어는 연결되는 이전 레이어와 전연결(fully-connected)된다. 따라서, 도 3의 도식에서 입력 레이어는 이어지는 제1 및 제3 전연결 레이어와 전연결되나, 제2 및 제3 전연결 레이어의 출력은 뒷단의 출력 레이어에서 단순히 합성된다.

[0046] 제1 전연결 레이어는 은닉층(hidden layer)으로 작용하고, 64개의 노드를 포함하며 31개의 노드를 포함하는 제2 전연결 레이어와 전연결된다. 제1 및 제2 전연결 레이어는 서로 직렬로 연결되어 고대역 에너지를 추정한다. 한편, 제3 전연결 레이어는 출력 레이어와 같은 수의 노드를 포함하고 제1 및 제2 전연결 레이어와는 병렬로 입력 레이어와 출력 레이어를 연결한다. 따라서, 제3 전연결 레이어는 숏컷(shortcut) 경로를 형성하여 고대역 에너지와 연관성이 높은 저대역 에너지의 정보를 출력 측으로 전달함으로써 고대역 에너지를 재귀적으로(recursively) 추정하는 것을 보조한다.

[0047] 앞에서 기술한 바와 같은 딥러닝 모델의 동작은 수학식 2로 표현될 수 있다.

수학식 2

$$z_j = W_{i,j}x + b_{i,j}, \quad j=1, 3$$

$$z_2 = W_{1,2}z_1 + b_{1,2}$$

$$\hat{E}_H = z_2 + z_3$$

[0048]

[0049] 여기서 입력 벡터 x 는 저대역 에너지 E_L 이고, $W_{i,j}$ 는 i 레이어와 j 레이어 사이의 전연결 가중치 행렬(weight matrix), $b_{i,j}$ 는 i 레이어와 j 레이어 사이의 편향치 벡터(bias vector), z_k 는 k 레이어의 출력 벡터를 나타낸다.

- [0050] 이하 트레이닝부(미도시)에 의한 딥러닝 모델의 트레이닝 과정을 설명한다. 먼저 고음질 음원(예컨대, 320 kbps로 압축된 음원)에 본 발명에 따른 전처리 과정을 적용하여 타겟(target) 데이터 E_{FB} 를 생성한다. 여기서 전처리 과정은 압축된 음원을 시간 영역 상의 데이터로 합성하는 과정, 시간 영역의 데이터를 주파수 영역의 데이터로 변환하는 과정, 및 주파수 영역의 데이터를 이용하여 부대역 별 대역 에너지를 추출하는 과정 등을 포함한다.
- [0051] 다음, 고음질 음원에 해당하는 저품질의 학습용 음원(예컨대, 64 kbps, 96 kbps 또는 128 kbps로 압축된 음원)의 저대역 에너지 E_L 을 대역에너지 추정부(122)에 입력하여 전대역 에너지 출력 \hat{E}_{FB} 를 산정한다. 트레이닝부는 타겟 데이터 E_{FB} 와 딥러닝 모델의 출력 \hat{E}_{FB} 간의 거리 메트릭에 기반하여 딥러닝 모델의 파라미터를 업데이트한다. 여기서 거리 메트릭은 L1 및 L2 메트릭 등, 두 비교 대상 간의 메트릭 거리 차이를 표현할 수 있는 것이면 어느 것이든 이용 가능하다.
- [0052] 딥러닝 모델에 대한 트레이닝 효율을 높이기 위해 각 노드의 값을 표현하는 에너지는 로그함수를 적용하여 사용한다. 에너지에 로그함수가 적용되었으므로 음수 값이 나타날 수 있다. 따라서, 트레이닝 과정에서 음수 값의 영향을 적절하게 반영하기 위해 도 3에 도시된 바대로 활성화 함수 ELU(Exponential Linear Unit)를 전연결 레이어의 출력에 적용한다.
- [0053] 도 2 및 수학적 식 2에 따르면, 딥러닝 모델의 출력이 전대역 에너지가 아니라 고대역 에너지 \hat{E}_H 로 표기되어 있다. 그 이유는 트레이닝 과정이 아닌 대역폭확대부(112)의 일부로 동작 시, 딥러닝 모델이 고대역 에너지를 추정하는 것을 표현하기 위함이다. 전술한 바와 같이 트레이닝 과정에서는 전대역에 대한 타겟 데이터가 존재하므로 전대역 에너지를 추정하도록 트레이닝하는 것이 가능하다. 그러나 대역폭확대부(112)의 일부로 동작 시, 딥러닝 모델의 입력으로 저대역 에너지가 추출되므로, 추출된 저대역 에너지를 재활용하는 것이 합리적이다. 따라서 딥러닝 모델이 대역폭확대부(112)의 일부로 동작 시에는 딥러닝 모델이 구한 전대역 에너지 중, 고대역 에너지에 해당하는 부분만을 사용한다.
- [0054] 본 실시예에 따른 인공 대역 변환기(110)가 탑재되는 모바일 디바이스(100)는 프로그램가능 컴퓨터일 수 있으며, 서버(미도시)와 연결이 가능한 적어도 한 개의 통신 인터페이스를 포함한다.
- [0055] 전술한 바와 같은 학습 모델에 대한 트레이닝은 서버에서 진행될 수 있다. 모바일 디바이스(100) 상에 탑재된 학습 모델과 동일한 구조의 딥러닝 모델에 대하여 서버의 트레이닝부는 트레이닝을 수행할 수 있다. 모바일 디바이스(100)와 연결되는 통신 인터페이스를 이용하여 서버는 트레이닝된 파라미터를 모바일 디바이스(100)로 전달하고, 전달받은 파라미터를 이용하여 인공 대역 변환기(110)는 학습 모델의 파라미터를 업데이트할 수 있다. 또한 모바일 디바이스(100) 출하 시점 또는 인공 대역 변환기(110)가 모바일 디바이스(100)에 탑재되는 시점에, 학습 모델의 파라미터가 설정될 수 있다.
- [0056] 도 4는 본 발명의 일 실시예에 따른 인공 대역 변환방법에 대한 순서도이다.
- [0057] 도 4에 도시된 순서도에서는, 대역폭확대부(112)가 수행하는 인공 대역 변환방법 위주로 표현하기 위해 도 1에 도시된 바에 따른 입력부(111) 및 출력부(113)가 수행하는 과정을 생략하였다. 또한 순서도의 각 과정에 기술함에 있어, 각 과정을 수행하는 대역폭확대부(112)의 구성요소에 대한 기술도, 도 2에 대한 설명에서 다루었으므로 더 이상의 상세한 설명은 생략한다.
- [0058] 먼저, 주파수 영역의 데이터를 이용하여 부대역 별로 저대역 에너지를 추출하고(S401), 저대역 에너지를 대역에너지 추정부(122)에 입력하여 고대역 에너지를 추정한 후(S402), 저대역 에너지 및 주파수 영역의 데이터를 처리하여 평균화된 저대역 데이터를 생성한다(S403).
- [0059] 다음, 평균화된 저대역 데이터를 이용하여 고대역 데이터를 복원한 후, 복원된 고대역 데이터와 평균화된 저대역 데이터를 합성하여 평균화된 전대역 데이터를 생성한다(S404).
- [0060] 다음, 대역에너지 추정부(122)가 추정한 고대역 에너지와 특성추출부(121)에서 추출한 저대역 에너지를 합성하여 전대역 에너지를 생성한 후, 평균화된 전대역 데이터와 전대역 에너지를 곱하여 전대역 주파수 데이터를 추정한다(S405).

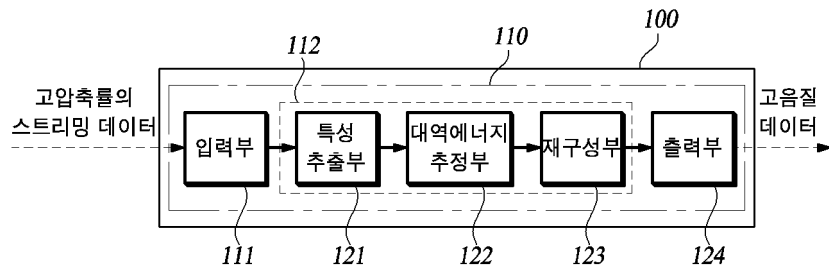
- [0061] 이하 본 실시예에 따른 인공 대역 변환장치의 성능을 평가한 결과를 설명한다. 평가 및 학습에는 고전음악, 케이팝(K-pop) 및 팝음악 등을 압축한 음원을 사용하였다. 전문적인 트레이닝 방법을 적용하여 본 실시예에 따른 대역에너지 추정부(122)를 트레이닝시킨 후, 동일한 평가용 음원을 이용하여 일반적인 HMM 기반의 추정 방식과 SNR 측정 결과를 비교하였다. 측정된 SNR은 20.8 dB로서, HMM 기반의 추정 방식에서 측정된 17.5 dB와 대비하여 더 우수한 성능을 보였다.
- [0062] 이상에서 설명한 바와 같이 본 실시예에 의하면, 전송 채널의 제약으로 인하여 고압축률로 압축되어 전송된 저품질의 음원 데이터를 고품질 음원 신호로 변형하기 위하여, 학습 모델 기반의 에너지 추정 방법을 이용하는 인공 대역 변환장치 및 방법을 제공함으로써, 모바일 디바이스에서 고품질 음원과 동가의 음질을 서비스하여 사용자의 만족도를 높이는 효과가 있다.
- [0063] 본 실시예에 따른 각 순서도에서는 각각의 과정을 순차적으로 실행하는 것으로 기재하고 있으나, 반드시 이에 한정되는 것은 아니다. 다시 말해, 순서도에 기재된 과정을 변경하여 실행하거나 하나 이상의 과정을 병렬적으로 실행하는 것이 적용 가능할 것이므로, 순서도는 시계열적인 순서로 한정되는 것은 아니다.
- [0064] 본 명세서에 설명되는 시스템들 및 기법들의 다양한 구현예들은, 디지털 전자 회로, 집적 회로, FPGA(field programmable gate array), ASIC(application specific integrated circuit), 컴퓨터 하드웨어, 펌웨어, 소프트웨어, 및/또는 이들의 조합으로 실현될 수 있다. 이러한 다양한 구현예들은 프로그래밍가능 시스템 상에서 실행가능한 하나 이상의 컴퓨터 프로그램들로 구현되는 것을 포함할 수 있다. 프로그래밍가능 시스템은, 저장 시스템, 적어도 하나의 입력 디바이스, 그리고 적어도 하나의 출력 디바이스로부터 데이터 및 명령들을 수신하고 이들에게 데이터 및 명령들을 전송하도록 결합되는 적어도 하나의 프로그래밍가능 프로세서(이것은 특수 목적 프로세서일 수 있거나 혹은 범용 프로세서일 수 있음)를 포함한다. 컴퓨터 프로그램들(이것은 또한 프로그램들, 소프트웨어, 소프트웨어 애플리케이션들 혹은 코드로서 알려져 있음)은 프로그래밍가능 프로세서에 대한 명령어들을 포함하며 "컴퓨터-판독가능 매체"에 저장된다.
- [0065] 컴퓨터-판독가능 매체는, 명령어들 및/또는 데이터를 프로그래밍가능 프로세서에게 제공하기 위해 사용되는, 임의의 컴퓨터 프로그램 제품, 장치, 및/또는 디바이스(예를 들어, CD-ROM, ROM, 메모리 카드, 하드 디스크, 광자기 디스크, 스토리지 디바이스 등의 비휘발성 또는 비일시적인 기록매체)를 나타낸다.
- [0066] 본 명세서에 설명되는 시스템들 및 기법들의 다양한 구현예들은, 프로그램가능 컴퓨터에 의하여 구현될 수 있다. 여기서, 컴퓨터는 프로그램가능 프로세서, 데이터 저장 시스템(휘발성 메모리, 비휘발성 메모리, 또는 다른 종류의 저장 시스템이거나 이들의 조합을 포함함) 및 적어도 한 개의 커뮤니케이션 인터페이스를 포함한다. 예컨대, 프로그램가능 컴퓨터는 서버, 네트워크 기기, 셋탑 박스, 내장형 장치, 컴퓨터 확장 모듈, 개인용 컴퓨터, 랩탑, PDA(Personal Data Assistant), 클라우드 컴퓨팅 시스템 또는 모바일 장치 중 하나일 수 있다.
- [0067] 이상의 설명은 본 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 실시예들은 본 실시예의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

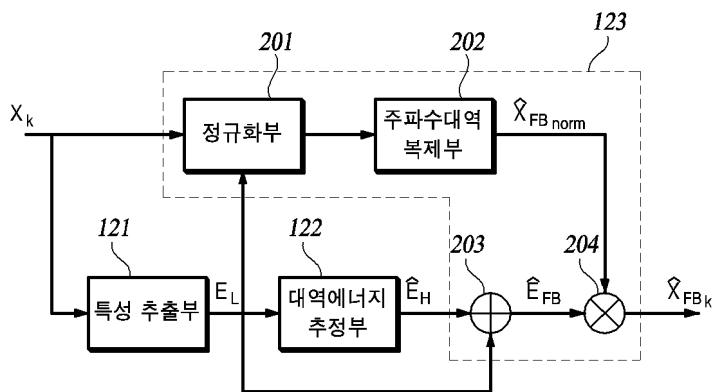
- [0068] 100: 모바일 디바이스 110: 인공 대역 변환기
- 111: 입력부 112: 대역폭 확장부
- 113: 출력부 121: 특성추출부
- 122: 대역에너지 추정부 123: 재구성부

도면

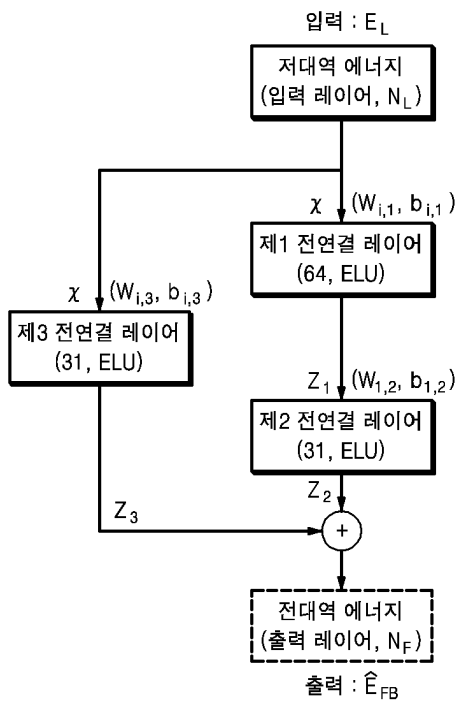
도면1



도면2



도면3



도면4

