

(72) 발명자

이창현

서울특별시 서대문구 신촌동 134 연세대학교

정양원

서울특별시 서초구 우면동 16번지 LG전자 전자기술
원

강홍구

서울특별시 서대문구 신촌동 134 연세대학교

특허청구의 범위

청구항 1

레지듀얼 및 룬텀 예측정보를 수신하는 단계;

상기 레지듀얼에 대해 역 주파수 변환을 수행하여 합성 레지듀얼을 생성하는 단계; 및,

상기 합성 레지듀얼 및 상기 룬텀 예측정보를 근거로 룬텀 합성을 수행하여, 현재 프레임의 합성 오디오 신호를 생성하는 단계를 포함하고,

상기 룬텀 예측 정보는, 최종 게인 및 최종 지연을 포함하고,

상기 최종 지연의 범위는 0부터이고,

상기 룬텀 합성은 이전 프레임을 포함하는 프레임의 합성 오디오 신호를 근거로 수행되는 것을 특징으로 하는 오디오 신호 처리 방법.

청구항 2

레지듀얼에 대해 역 주파수 변환을 수행하여 합성 레지듀얼을 생성하는 역변환부; 및

상기 합성 레지듀얼 및 룬텀 예측정보를 근거로 룬텀 합성을 수행하여, 현재 프레임의 합성 오디오 신호를 생성하는 룬텀 합성부를 포함하고,

상기 룬텀 예측 정보는, 최종 게인 및 최종 지연을 포함하고,

상기 최종 지연의 범위는 0부터이고,

상기 룬텀 합성은 이전 프레임을 포함하는 프레임의 합성 오디오 신호를 근거로 수행되는 것을 특징으로 하는 오디오 신호 처리 장치.

청구항 3

이전 프레임의 소스 오디오 신호를 이용하여 시간 도메인 상에서 룬텀 프리딕션을 수행함으로써, 현재 프레임의 임시 레지듀얼을 생성하는 단계;

상기 임시 레지듀얼을 주파수 변환하는 단계;

상기 임시 레지듀얼을 역주파수 변환하여 이전 프레임의 합성 레지듀얼을 생성하는 단계; 및,

상기 합성 레지듀얼을 이용하여 룬텀 예측 정보를 결정하는 단계를 포함하는 것을 특징으로 하는 오디오 신호 처리 방법.

청구항 4

제 3 항에 있어서,

상기 룬텀 예측 정보를 결정하는 단계는,

상기 합성 레지듀얼을 이용하여 룬텀 합성을 수행함으로써, 이전 프레임의 합성 오디오 신호를 생성하는 단계; 및,

상기 합성 오디오 신호를 이용하여 상기 룬텀 예측 정보를 결정하는 단계를 포함하는 것을 특징으로 하는 오디오 신호 처리 방법.

청구항 5

제 4 항에 있어서,

상기 임시 레지듀얼을 생성하는 단계는, 임시 게인 및 임시 지연을 생성하는 단계를 포함하고,

상기 룬텀 합성은, 상기 임시 게인 및 상기 임시 지연을 근거로 수행되는 것을 특징으로 하는 오디오 신호 처리

방법.

청구항 6

제 5 항에 있어서,

상기 룡텁 합성은, 상기 임시 계인을 근거로 하는 하나 이상의 후보 계인, 및 상기 임시 지연을 근거로 하는 하나 이상의 후보 지연을 이용하여 수행되는 것을 특징으로 하는 오디오 신호 처리 방법.

청구항 7

제 3 항에 있어서,

상기 룡텁 예측 정보는, 최종 계인 및 최종 계인을 포함하며, 상기 소스 오디오 신호를 근거로 결정되는 것을 특징으로 하는 오디오 신호 처리 방법.

청구항 8

이전 프레임의 소스 오디오 신호를 이용하여 시간 도메인 상에서 룡텁 프리딕션을 수행함으로써, 현재 프레임의 임시 레지듀얼을 생성하는 룡텁 예측부;

상기 임시 레지듀얼을 주파수 변환하는 주파수 변환부;

상기 임시 레지듀얼을 역주파수 변환하여 이전 프레임의 합성 레지듀얼을 생성하는 역변환부; 및, 상기 합성 레지듀얼을 이용하여 룡텁 예측 정보를 결정하는 예측정보 결정부를 포함하는 것을 특징으로 하는 오디오 신호 처리 장치.

청구항 9

제 8 항에 있어서,

상기 합성 레지듀얼을 이용하여 룡텁 합성을 수행함으로써, 이전 프레임의 합성 오디오 신호를 생성하는 룡텁 합성부를 더 포함하고,

상기 예측정보 결정부는, 상기 합성 오디오 신호를 이용하여 상기 룡텁 예측 정보를 결정하는 것을 특징으로 하는 오디오 신호 처리 장치.

청구항 10

제 9 항에 있어서,

상기 룡텁 예측부는 임시 계인 및 임시 지연을 생성하고,

상기 룡텁 합성은, 상기 임시 계인 및 상기 임시 지연을 근거로 수행되는 것을 특징으로 하는 오디오 신호 처리 장치.

청구항 11

제 10 항에 있어서,

상기 룡텁 합성은, 상기 임시 계인을 근거로 하는 하나 이상의 후보 계인, 및 상기 임시 지연을 근거로 하는 하나 이상의 후보 지연을 이용하여 수행되는 것을 특징으로 하는 오디오 신호 처리 장치.

청구항 12

제 8 항에 있어서,

상기 룡텁 예측 정보는, 최종 계인 및 최종 계인을 포함하며, 상기 소스 오디오 신호를 근거로 결정되는 것을 특징으로 하는 오디오 신호 처리 장치.

청구항 13

디지털 오디오 데이터를 저장하며, 컴퓨터로 읽을 수 있는 저장 매체에 있어서,

상기 디지털 오디오 데이터는 룬텀 플래그 정보, 레지듀얼 및 룬텀 예측 정보를 포함하며,
상기 룬텀 플래그 정보는, 상기 디지털 오디오 데이터에 룬텀 예측이 적용되었는지 여부를 나타내는 정보이고,
상기 룬텀 예측 정보는, 룬텀 프리딕션 및 룬텀 합성을 통해 생성된 최종 게인 및 최종 지연을 포함하고,
상기 최종 지연의 범위는 0부터인 것을 특징으로 하는 저장 매체.

명 세 서

발명의 상세한 설명

기술 분야

<1> 본 발명은 오디오 신호를 인코딩하거나 디코딩할 수 있는 오디오 신호 처리 방법 및 장치에 관한 것이다.

배경 기술

<2> 일반적으로, 음성 신호를 압축하기 위해서, 시간 도메인 상에서 선형 예측 코딩(Linear Prediction Coding: LPC)과 같은 숏텀 프리딕션(short term prediction)을 수행한다. 그런 다음, 숏텀 프리딕션의 결과인 레지듀얼에 대해 피치(pitch)를 찾아서 룬텀 프리딕션(long term prediction)을 수행한다.

발명의 내용

해결 하고자하는 과제

<3> 종래에는, 선형 예측 코딩을 수행한 결과인 레지듀얼에 대해서 룬텀 예측을 수행하는 경우면, 음성 특성을 가진 신호에 대한 압축률이 크지만, 비음성 특성을 가진 신호에 대해서는 상대적으로 압축률이 크지 않다.

과제 해결수단

<4> 본 발명은 상기와 같은 문제점을 해결하기 위해 창안된 것으로서, 음성 특성 및 비음성 특성이 혼재된 오디오 신호에 대해, 룬텀 예측을 수행하는 오디오 신호 처리 방법 및 장치를 제공하는 데 있다.

<5> 본 발명의 또 다른 목적은, 오디오 신호에 대해 룬텀 예측을 수행하고, 그 레지듀얼을 주파수 도메인 상에서 코딩하기 위한 오디오 신호 처리 방법 및 장치를 제공하는 데 있다.

<6> 본 발명의 또 다른 목적은, 현재 프레임과 가장 유사한 프리딕션을 찾는 데 있어서, 현재 프레임의 직전 프레임을 이용하기 위한 오디오 신호 처리 방법 및 장치를 제공하는 데 있다.

<7> 본 발명의 또 다른 목적은, 디코더에서 획득할 수 없는 정보(예:소스 신호)가 아닌 획득 가능한 정보(예: 합성 레지듀얼)를 이용하여 룬텀 합성을 수행할 수 있도록 하기 위한 룬텀 예측 정보를 생성하는 오디오 신호 처리 방법 및 장치를 제공하는 데 있다.

<8> 본 발명의 또 다른 목적은, 소스 신호를 이용하여 룬텀 예측을 통해 임시적으로 룬텀 예측 정보를 생성한 후, 이 인근 범위 내에서 룬텀 합성을 통해 최종적인 룬텀 예측 정보를 결정하는 오디오 신호 처리 방법 및 장치를 제공하는 데 있다.

효 과

<9> 본 발명은 다음과 같은 효과와 이점을 제공한다.

<10> 첫째, 음성 신호뿐만 아니라 음성 특성 및 비음성 특성이 혼재된 오디오 신호에 대해 룬텀 예측을 수행함으로써, 특히, 시간 도메인 상에서 반복되는 특성을 갖는 신호에 대해, 코딩 효율을 높일 수 있다.

<11> 둘째, 현재 프레임의 프리딕션을 탐색하기 위해, 현재 프레임의 직전 프레임까지 참조함으로써, 가장 유사한 프리딕션을 획득할 수 있기 때문에, 레지듀얼의 비트레이트를 줄일 수 있다.

<12> 셋째, 획득할 수 없는 정보(예:소스 신호)가 아닌 획득 가능한 정보(예: 양자화된 레지듀얼)를 이용하여 디코더에서 룬텀 합성을 수행할 수 있기 때문에, 룬텀 합성의 복원률을 상승시킬 수 있다.

<13> 넷째, 복잡도가 비교적 낮은 프로세싱을 통해 대략적인 룬텀 예측 정보(지연, 게인)을 생성한 후, 이를 기반으

로 축소된 탐색 범위(searching range) 내에서 보다 정확하게 예측 정보를 결정함으로써, 전체적인 컴플렉시티를 감소시킬 수 있다.

발명의 실시를 위한 구체적인 내용

- <14> 상기와 같은 목적을 달성하기 위하여 본 발명에 따른 오디오 신호 처리 방법은 레지듀얼 및 롬텀 예측정보를 수신하는 단계; 상기 레지듀얼에 대해 역 주파수 변환을 수행하여 합성 레지듀얼을 생성하는 단계; 및, 상기 합성 레지듀얼 및 상기 롬텀 예측정보를 근거로 롬텀 합성을 수행하여, 현재 프레임의 합성 오디오 신호를 생성하는 단계를 포함하고, 상기 롬텀 예측 정보는, 최종 게인 및 최종 지연을 포함하고, 상기 최종 지연의 범위는 0부터 이고, 상기 롬텀 합성은 이전 프레임을 포함하는 프레임의 합성 오디오 신호를 근거로 수행된다.
- <15> 본 발명의 또 다른 측면에 따르면, 레지듀얼에 대해 역 주파수 변환을 수행하여 합성 레지듀얼을 생성하는 역변환부; 및 상기 합성 레지듀얼 및 롬텀 예측정보를 근거로 롬텀 합성을 수행하여, 현재 프레임의 합성 오디오 신호를 생성하는 롬텀 합성부를 포함하고, 상기 롬텀 예측 정보는, 최종 게인 및 최종 지연을 포함하고, 상기 최종 지연의 범위는 0부터이고, 상기 롬텀 합성은 이전 프레임을 포함하는 프레임의 합성 오디오 신호를 근거로 수행되는 오디오 신호 처리 장치가 제공된다.
- <16> 본 발명의 또 다른 측면에 따르면, 이전 프레임의 소스 오디오 신호를 이용하여 시간 도메인 상에서 롬텀 프리딕션을 수행함으로써, 현재 프레임의 임시 레지듀얼을 생성하는 단계; 상기 임시 레지듀얼을 주파수 변환하는 단계; 상기 임시 레지듀얼을 역주파수 변환하여 이전 프레임의 합성 레지듀얼을 생성하는 단계; 및, 상기 합성 레지듀얼을 이용하여 롬텀 예측 정보를 결정하는 단계를 포함하는 오디오 신호 처리 방법이 제공된다.
- <17> 본 발명에 따르면, 상기 롬텀 예측 정보를 결정하는 단계는, 상기 합성 레지듀얼을 이용하여 롬텀 합성을 수행함으로써, 이전 프레임의 합성 오디오 신호를 생성하는 단계; 및, 상기 합성 오디오 신호를 이용하여 상기 롬텀 예측 정보를 결정하는 단계를 포함할 수 있다.
- <18> 본 발명에 따르면, 상기 임시 레지듀얼을 생성하는 단계는, 임시 게인 및 임시 지연을 생성하는 단계를 포함하고, 상기 롬텀 합성은, 상기 임시 게인 및 상기 임시 지연을 근거로 수행될 수 있다.
- <19> 본 발명에 따르면, 상기 롬텀 합성은, 상기 임시 게인을 근거로 하는 하나 이상의 후보 게인, 및 상기 임시 지연을 근거로 하는 하나 이상의 후보 지연을 이용하여 수행될 수 있다.
- <20> 본 발명에 따르면, 상기 롬텀 예측 정보는, 최종 게인 및 최종 게인을 포함하며, 상기 소스 오디오 신호를 근거로 결정될 수 있다.
- <21> 본 발명의 또 다른 측면에 따르면, 이전 프레임의 소스 오디오 신호를 이용하여 시간 도메인 상에서 롬텀 프리딕션을 수행함으로써, 현재 프레임의 임시 레지듀얼을 생성하는 롬텀 예측부; 상기 임시 레지듀얼을 주파수 변환하는 주파수 변환부; 상기 임시 레지듀얼을 역주파수 변환하여 이전 프레임의 합성 레지듀얼을 생성하는 역변환부; 및, 상기 합성 레지듀얼을 이용하여 롬텀 예측 정보를 결정하는 예측정보 결정부를 포함하는 오디오 신호 처리 장치가 제공된다.
- <22> 본 발명에 따르면, 상기 합성 레지듀얼을 이용하여 롬텀 합성을 수행함으로써, 이전 프레임의 합성 오디오 신호를 생성하는 롬텀 합성부를 더 포함하고, 상기 예측정보 결정부는, 상기 합성 오디오 신호를 이용하여 상기 롬텀 예측 정보를 결정할 수 있다.
- <23> 본 발명에 따르면, 상기 롬텀 예측부는 임시 게인 및 임시 지연을 생성하고, 상기 롬텀 합성은, 상기 임시 게인 및 상기 임시 지연을 근거로 수행될 수 있다.
- <24> 본 발명에 따르면, 상기 롬텀 합성은, 상기 임시 게인을 근거로 하는 하나 이상의 후보 게인, 및 상기 임시 지연을 근거로 하는 하나 이상의 후보 지연을 이용하여 수행될 수 있다.
- <25> 본 발명에 따르면, 상기 롬텀 예측 정보는, 최종 게인 및 최종 게인을 포함하며, 상기 소스 오디오 신호를 근거로 결정될 수 있다.
- <26> 본 발명의 또 다른 측면에 따르면, 디지털 오디오 데이터를 저장하며, 컴퓨터로 읽을 수 있는 저장 매체에 있어서, 상기 디지털 오디오 데이터는 롬텀 플래그 정보, 레지듀얼 및 롬텀 예측 정보를 포함하며, 상기 롬텀 플래그 정보는, 상기 디지털 오디오 데이터에 롬텀 예측이 적용되었는지 여부를 나타내는 정보이고, 상기 롬텀 예측 정보는, 롬텀 프리딕션 및 롬텀 합성을 통해 생성된 최종 게인 및 최종 지연을 포함하고, 상기 최종 지연의 범

위는 0부터인 저장 매체가 제공된다.

- <27> 이하 첨부된 도면을 참조로 본 발명의 바람직한 실시예를 상세히 설명하기로 한다. 이에 앞서, 본 명세서 및 청구범위에 사용된 용어나 단어는 통상적이거나 사전적인 의미로 한정해서 해석되어서는 아니되며, 발명자는 그 자신의 발명을 가장 최선의 방법으로 설명하기 위해 용어의 개념을 적절하게 정의할 수 있다는 원칙에 입각하여 본 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야만 한다. 따라서, 본 명세서에 기재된 실시예와 도면에 도시된 구성은 본 발명의 가장 바람직한 일 실시예에 불과할 뿐이고 본 발명의 기술적 사상을 모두 대변하는 것은 아니므로, 본 출원시점에 있어서 이들을 대체할 수 있는 다양한 균등물과 변형예들이 있을 수 있음을 이해하여야 한다.
- <28> 본 발명에서 다음 용어는 다음과 같은 기준으로 해석될 수 있고, 기재되지 않은 용어라도 하기 취지에 따라 해석될 수 있다. 코딩은 경우에 따라 인코딩 또는 디코딩으로 해석될 수 있고, 정보(information)는 값(values), 파라미터(parameter), 계수(coefficients), 성분(elements) 등을 모두 아우르는 용어로서, 경우에 따라 의미는 달리 해석될 수 있는 바, 그러나 본 발명은 이에 한정되지 아니한다.
- <29> 여기서 오디오 신호(audio signal)란, 광의로는, 비디오 신호와 구분되는 개념으로서, 재생시 청각으로 식별할 수 있는 신호를 지칭하고, 협의로는, 음성(speech) 신호와 구분되는 개념으로서, 음성 특성이 없거나 적은 신호를 의미한다. 본 발명에서의 오디오 신호는 광의로 해석되어야 하며 음성 신호와 구분되어 사용될 때 협의의 오디오 신호로 이해될 수 있다.
- <30> 한편, 프레임이란, 오디오 신호를 인코딩 또는 디코딩하기 위한 단위를 일컫는 것으로서, 특정 샘플 수나 특정 시간에 한정되지 아니한다.
- <31> 본 발명에 따른 오디오 신호 처리 방법 및 장치는, 롬 인코딩/디코딩 장치 및 방법이 될 수도 있고, 나아가 이 장치 및 방법이 적용된 오디오 신호 인코딩/디코딩 장치 및 방법이 될 수 있는 바, 이하, 롬 인코딩/디코딩 장치 및 방법에 대해서 설명하고, 이 롬 인코딩/디코딩 장치가 수행하는 롬 데이터 인코딩/디코딩 방법, 및 이 장치가 적용된 오디오 신호 인코딩/디코딩 장치 및 방법에 대해서 설명하고자 한다.
- <32> 도 1은 본 발명의 실시예에 따른 오디오 신호 처리 장치 중 롬 코딩부의 구성을 보여주는 도면이고, 도 2는 본 발명의 실시예에 따른 오디오 신호 처리 방법의 순서를 보여주는 도면이다. 도 1 및 도 2를 참조하면서 롬 코딩부가 오디오 신호를 처리하는 과정에 대해서 설명하고자 한다.
- <33> 우선, 도 1을 참조하면, 롬 인코딩 장치(100)는 롬 예측부(110), 역변환부(120), 롬 합성부(130), 예측정보 결정부(140), 및 지연부(150)를 포함하고, 주파수 변환부(210), 양자화부(220), 및 심리음향모델(230)을 더 포함할 수 있다. 여기서 롬 예측부(110)는 오픈 루프(open loop) 방식에 해당하고, 롬 합성부(130)는 클로즈드 루프(closed loop) 방식으로 명명할 수 있다. 한편, 주파수 변환부(210), 양자화부(220), 및 심리음향모델(230)은 AAC(Advanced Audio Coding) 표준에 따른 것일 수 있지만, 본 발명은 이에 한정되지 아니한다.
- <34> 도 1 및 도 2를 참조하면, 우선 롬 예측부(110)는 소스 오디오 신호 $S_t(n)$ (source audio signal)에 대해 롬 예측(Long-Term Prediction)(장구간 예측)을 수행함으로써, 임시 계인 b 및 임시 지연 d 을 생성하고 임시 레지듀얼 $r_t(n)$ 을 생성한다(S110단계). 이 단계에 대해서 설명하기 위하여 우선 도 3을 참조하면서 프레임별 신호에 대해서 설명하고자 한다. 도 3을 참조하면, 현재 프레임이 t 이고 현재 프레임의 이전 프레임이 $t-1$, 이전 프레임의 이전 프레임이 $t-2$ 이 존재하고 각각의 프레임에 대해서 오디오 신호가 $S_t(n)$, $S_{t-1}(n)$, $S_{t-2}(n)$ 존재한다. 하나의 프레임은 약 1024개의 샘플로 구성될 수 있는데, $t-2$ 번째 프레임이 $k+1$ 번째 샘플부터 $k+1024$ 번째 샘플까지인 경우, $t-1$ 번째 프레임은 $k+1025$ 번째 샘플부터 $k+2048$ 번째 샘플로, t 번째 프레임은 $k+2049$ 번째 샘플부터 $k+3072$ 번째 샘플로 구성될 수 있다.
- <35> 한편, S110 단계에서, 롬 예측이란, 주어진 시점 n 에서의 신호를 지연(delay)만큼 이전의 신호에 계인(gain)을 곱한 것으로 근사하는 것으로서, 다음 수식과 같이 정의될 수 있다.
- <36> [수식 1]
- <37>
$$r_t(n) = S_t(n) - b \cdot S_t(n-d)$$
- <38> 여기서, $S_t(n)$ 는 현재 프레임의 신호, b 는 계인, d 는 지연, $r_t(n)$ 은 레지듀얼

<39> S110 단계에서의 계인 b 및 지연 d 는 최종적인 것이 아니고 이후 단계에서 업데이트되기 때문에, S110 단계에서의 계인 및 지연을 임시 계인 및 임시 지연이라고 명명하고자 한다. 그러나, 임시 레지듀얼 $r_t(n)$ 최종 계인 및 지연으로 다시 계산되지는 않는데, 다만 주파수 변환을 통해 변환된(앨리어싱) 레지듀얼이나, 및 룬텀 합성을 통해 합성 레지듀얼이 생성될 수 있다.

<40> S110 단계에서는, 현재 프레임과 유사한 프리딕션을 찾기 위해 합성 신호가 아니라 소스 신호를 이용하기 때문에, 프리딕션의 탐색 범위(searching range)에 이전 프레임(t-1)을 포함시킬 수 있다. 왜냐하면 이전 프레임(t-1)의 소스 신호를 그대로 이용할 수 있기 때문이다. 또한, S110 단계의 과정을 오픈 루프(open loop) 방식 또는 룬텀 예측(Long-term Prediction)이라고 할 수 있다.

<41> 한편, 다음 표는 오픈 루프(open loop)가 수행되는 경우, 평균제곱오차(MSE), 지연(Pitch lag), 계인(Prediction gain), 결과(output), 탐색 범위(Searching range)의 일 예를 나타낸다.

<42> [표 1]

	Open loop
	$\tilde{s}(n) = bs(n-d)$
MSE	$\varepsilon_o = \sum_{n=0}^{N-1} [s(n) - \tilde{s}(n)]^2$
Pitch lag	$d_o = \arg \max_d \frac{\sum_{i=0}^{N-1} s(i)s(i-d)}{\sqrt{\sum_{i=0}^{N-1} s(i-d)^2}}$
Prediction gain	$b_o = \frac{\sum_{i=0}^{N-1} s(i)s(i-d_o)}{\sum_{i=0}^{N-1} s(i-d_o)^2}$
output	$r(n) = s(n) - b_o s(n-d_o)$
Search range	$50 \leq d_o \leq 512$ 93.75Hz ~ 960Hz

<43> <44> 상기 임시 계인은 표 1에서 나타낸 방식에 의해, 상기 임시 지연은 표 1에서 나타낸 방식에 의해 생성된 것일 수 있다. 또한, 수학적 1은 표 1에 나타낸 출력(output)과 동일하다.

<45> 한편, S110 단계에서 수행되는 내용을 버퍼 관점에서 설명하기 위해, 도 4를 참조하면, 입력 버퍼에는, 현재 프레임의 소스 신호 $S_t(n)$ 와 이전 프레임의 소스 신호 $S_{t-1}(n)$ 가 존재한다. 현재 프레임 $S_t(n)$ 과 가장 유사한 신호가 이전 프레임 신호 $S_{t-1}(n)$ 에 존재할 수 있다. 이때의 임시계인 b 및 임시지연 d 인 경우, 수학적 1과 같이 임시 레지듀얼 $r_t(n)$ 은 생성되어 출력 버퍼에 저장된다.

<46> 다시 도 1 및 도 2를 참조하면, 주파수 변환부(210)는 임시 레지듀얼 $r_t(n)$ 에 대해 시간-주파수 변환(time-to-frequency mapping)(또는 주파수 변환이라 함)을 수행하여 주파수 변환된 레지듀얼 신호 $\hat{R}_{t-1,t-2}(\omega)$

$\hat{R}_{t,t-1}(\omega)$ 를 생성한다(S120 단계). 시간-주파수 변환에는 QMF (Quadrature Mirror Filterbank), MDCT(Modified Discrete Fourier Transform) 등의 방식으로 수행될 수 있지만 본 발명은 이에 한정되지 아니한다. 이때, 스펙트럴 계수는 MDCT (Modified Discrete Transform) 변환을 통해 획득된 MDCT 계수일 수 있다. 여기서 주파수 변환된 신호는 특정 프레임에 대해서 완전한 신호가 아니기 때문에 앨리어싱된 신호라고도 지칭할 수 있다.

<47> 한편, S120 단계 및 S130단계에서 수행되는 과정을 버퍼 관점에서 설명하기 위해 도 5를 참조하고자 한다. 도 5를 참조하면, 입력 버퍼에는 t-2번째 프레임의 레지듀얼 $r_{t-2}(n)$, t-1번째 프레임의 레지듀얼 $r_{t-1}(n)$, t번째 프레임의 레지듀얼 $r_t(n)$ 이 존재함을 알 수 있다. 연속된 두 프레임의 레지듀얼에 대해 하나의 윈도우를 적용하여 주파수 변환을 수행한다. 구체적으로 t-2번째 프레임과 t-1프레임에 윈도우를 적용한 결과로서, 변환된 레지듀얼 신호 $\hat{R}_{t-1,t-2}(\omega)$ 가 생성되고, t-1번째 프레임과 t 프레임에 윈도우를 적용하여 $\hat{R}_{t,t-1}(\omega)$ 가 생성된다. 이 결과들은 추후 역변환부(120)에 입력되어 S140 단계에서 역 주파수 변환됨으로써, t-1번째 프레임의 $\hat{r}_{t-1}(n)$ 가 된다. 이에 대해서는 추후에 구체적으로 설명하고자 한다.

<48> 한편, 심리 음향 모델(130)은 입력된 오디오 신호에 대해 마스킹 효과를 적용하여 마스킹 한계치(masking threshold)를 생성한다. 마스킹(masking) 효과란, 심리 음향 이론에 의한 것으로, 크기가 큰 신호에 인접한 작은 신호들은 큰 신호에 의해서 가려지기 때문에 인간의 청각 구조가 이를 잘 인지하지 못한다는 특성을 이용하는 것이다.

<49> 양자화부(220)는 마스킹 한계치를 근거로 주파수 변환된 레지듀얼 신호를 양자화한다(S120 단계). 이 양자화된 레지듀얼 신호 $\hat{R}_{t,t-1}(\omega)$ 는 역변환부(120)로 입력되기도 하고, 룬텀 인코딩부의 출력이 되어 비트스트림을 통해 룬텀 디코딩에 전달될 수 있다.

<50> 그런 다음, 역변환부(120)는 주파수 변환된 레지듀얼에 대해 역양자화 및 역 주파수 변환(또는 주파수-시간 변환)을 수행함으로써, 이전 프레임의 합성 레지듀얼 $\hat{r}_{t-1}(n)$ 을 생성한다(S130 단계). 여기서 주파수-시간 변환(frequency to time mapping)은 IQMF (Inverse Quadrature Mirror Filterbank), IMDCT(Inverse Modified Discrete Fourier Transform) 등의 방식으로 수행될 수 있지만 본 발명은 이에 한정되지 아니한다.

<51> 다시 도 5를 참조하면, 주파수 변환 결과로 변환된 두 개의 신호 $\hat{R}_{t-1,t-2}(\omega)$, $\hat{R}_{t,t-1}(\omega)$ 가 생성되는 데, 이 신호는 t-1프레임 즉 이전 프레임에서 중첩되는 특성이 있다. 이 두 개의 신호에 대해 각각 역 변환을 수행한 후 이 둘을 가산하면 이전 프레임의 합성 레지듀얼 신호 $\hat{r}_{t-1}(n)$ 가 생성된다. 이는 룬텀 예측 단계(S110 단계)에서는 현재 프레임의 레지듀얼이 생성된 데에 비해, 주파수 변환 및 역변환을 거친 다음에는 현재 프레임이 아닌 이전 프레임에 대한 합성 레지듀얼 $\hat{r}_{t-1}(n)$ 이 생성된 것이다.

<52> 룬텀 합성부(130)는 우선, 룬텀 예측부(110)에 의해 생성된 임시 계인 b 및 임시 지연 d 을 기반으로, 후보 계인 bc 및 후보 지연 dc 을 결정한다(S140 단계). 예를 들어, 후보 계인 및 후보 지연은 수학적과 같은 범위 내에서 결정될 수 있다.

<53> [수학식 2]

<54>
$$bc = b \pm \alpha$$

<55>
$$dc = d \pm \beta$$

<56> 여기서 α , β 는 임의의 상수

<57> 후보 계인은 하나 이상의 계인의 집합으로, 후보 지연은 하나 이상의 지연의 집합으로 구성되는 데, 서칭 범위가 임시 계인 및 임시 지연을 기준으로 축소되는 것이다.

<58> 룡텀 합성부(130)는 S140 단계에서 결정된 후보 게인 b_c 및 후보 지연 d_c 과, S130 단계에서 생성된 이전 프레임의 레지듀얼 $\hat{r}_{t-1}(n)$ 를 근거로 룡텀 합성(Long-Term Synthesis)를 수행하여 이전 프레임의 합성 오디오 신호 $\hat{S}_{t-1}(n)$ 를 생성한다(S150 단계). 도 6은 룡텀 합성 과정(S150 단계) 및 예측정보 결정 과정(S160 단계)를 설명하기 위한 도면이다. 도 6을 참조하면, 입력 버퍼에는 t-2번째 프레임의 합성 오디오 신호와, S130 단계에서 생성된 t-1번째 프레임의 합성 레지듀얼 신호가 존재함을 알 수 있다. 이 두 신호를 이용하여 후보 게인 b_c 및 후보 지연 d_c 에 대해, 다음 수학적식과 같이 합성 오디오 신호를 생성한다.

<59> [수학적식 3]

<60>
$$\hat{S}_{t-1_c}(n) = \hat{r}_{t-1}(n) + b_c \cdot \hat{S}_{t-1}(n-d_c)$$

<61> 여기서, $\hat{S}_{t-1_c}(n)$ 는 후보 게인 및 후보 지연에 대한 이전 프레임의 합성 오디오 신호, $\hat{r}_{t-1}(n)$ 이전 프레임의 합성 레지듀얼, $\hat{S}_{t-1}(n)$ 는 이전 프레임의 합성 오디오 신호

<62> 한편, 다음 표는 클로즈드 루프(closed loop)가 수행되는 경우, 평균제곱오차(MSE), 지연(Pitch lag), 게인(Prediction gain), 결과(output), 탐색 범위(Searching range)의 일 예를 나타낸다.

<63> [표 2]

	Closed loop
	$\hat{s}(n) = \hat{r}(n) + b\hat{s}(n-d)$
MSE	$\varepsilon_c = \sum_{n=0}^{N-1} [(s(n) - \hat{r}(n)) - b\hat{s}(n-d)]^2$
Pitch lag	$d_o - C \leq d_c \leq d_o + C$ $d_c = \arg \min_d \{ \varepsilon_c \}$
Prediction gain	$s'(n) = s(n) - \hat{r}(n)$ $b_c = \frac{\sum_{i=0}^{N-1} s'(i)s'(i-d_c)}{\sum_{i=0}^{N-1} s'(i-d_c)^2}$
output	$\hat{s}(n) = \hat{r}(n) + b_c \hat{s}(n-d_c)$
Search range	$d_o - C \leq d_c \leq d_o + C$ C = 10 (samples)

<64>

<65> 또한, 수학적식 4는 표 2에 나타낸 출력(output)과 동일할 수 있다. 한편, 탐색 범위는 표 2와 같이 결정되기 보다는, S110 단계의 임시 게인 및 임시 지연을 근거로 한 후보 지연 및 후보 게인에 따라 결정된다.

<66> 한편 지연부(150)는 현재 프레임에 대한 소스 신호 $S_t(n)$ 를 지연시켜서 다음 프레임을 프로세싱할 때 이전 프레임의 소스 신호 $S_{t-1}(n)$ 를 예측정보 결정부(140)에 입력한다.

<67> 예측정보 결정부(140)는, 지연부(150)로부터 수신한 이전 프레임의 소스 신호 $S_{t-1}(n)$ 과 S150 단계에서 생성한 이전 프레임의 합성 오디오 신호 $\hat{S}_{t-1,c}(n)$ 를 비교함으로써, 가장 적절한 룬텀 예측정보 즉, 최종 개인 bo 및 최종 지연 do 를 결정한다(S160 단계). 이때, 다음과 같은 수학적식에 따라서 최종 개인 및 최종 지연이 결정될 수 있다.

<68> [수학적식 4]

<69>
$$\{bo, do\} = \arg \min \{S_{t-1}(n) - \hat{S}_{t-1,c}(n)\}$$

<70> $S_{t-1}(n)$ 이전 프레임의 소스 신호, $\hat{S}_{t-1,c}(n)$ 는 후보 개인 및 후보 개인에 대한 이전 프레임의 합성 오디오 신호, bo 는 최종 개인, do 는 최종 지연.

<71> 상기 수학적식 4는 앞서 표 2에 나타난 평균제곱오차(MSE)에 따른 것일 수 있다.

<72> S160 단계에서 생성된 최종 개인 및 최종 지연은, 디코더에서 획득할 수 있는 정보를 근거로, 현재 프레임과 가장 유사한 신호를 t-1번째 프레임(즉, 이전 프레임)을 포함한 프레임 중에서 찾은 결과이다. 최종 지연은 t-1번째 프레임까지 포함한 이전 프레임을 포함한 프레임 중에서 찾은 결과이기 때문에, 최종 지연의 범위는 N(프레임 길이)부터가 아니라 0부터이다. N부터인 경우, N을 뺀 나머지 값만을 전송할 수 있지만, 0부터인 경우, N을 뺀 나머지 값이 아니라 그 값 그대로 전송될 수 있다.

<73> 한편, 예측정보 결정부는 최종 개인 및 최종 지연뿐만 아니라, 룬텀 예측(/합성)이 적용되었는지 여부를 나타내는 룬텀 플래그 정보를 더 생성할 수 있다.

<74> 도 7은 본 발명의 실시예에 따른 오디오 신호 처리 장치 중 룬텀 디코딩 장치의 구성을 보여주는 도면이다. 도 7을 참조하면, 룬텀 디코딩 장치(300)는 룬텀 합성부(330)를 포함하고, 역양자화부(310) 및 역변환부(320)를 더 포함할 수 있다. 한편 역양자화부(310) 및 역변환부(320)는 AAC(Advanced Audio Coding) 표준에 따른 것일 수 있지만, 본 발명은 이에 한정되지 않는다.

<75> 우선, 역양자화부(310)는 비트스트림으로부터 레지듀얼 $\hat{R}_{t,t-1}(\omega)$ 를 추출하여 이를 역양자화한다. 여기서 레지듀얼은 앞서 설명한 주파수 변환된 레지듀얼 또는 앨리어싱된 레지듀얼일 수 있다.

<76> 그런 다음 역변환부(320)는 주파수 변환된 레지듀얼 $\hat{R}_{t,t-1}(\omega)$ 에 대해 역 주파수 변환(또는 주파수-시간 변환)을 수행하여, 현재 프레임의 레지듀얼 $\hat{r}_t(n)$ 을 생성한다. 여기서 주파수-시간 변환(frequency to time mapping)은 IQMF (Inverse Quadrature Mirror Filterbank), IMDCT(Inverse Modified Discrete Fourier Transform) 등의 방식으로 수행될 수 있지만 본 발명은 이에 한정되지 아니한다.

<77> 여기서 획득된 레지듀얼 $\hat{r}_t(n)$ 은 앞서 룬텀 인코딩 장치에서 앨리어싱된 레지듀얼들을 기반으로 생성한 합성 레지듀얼 $\hat{r}_t(n)$ 일 수 있다. 룬텀 디코딩 장치에서의 역양자화 과정 및 룬텀 합성 과정이 도 8에 도시되어 있다. 도 8를 참조하면, 입력버퍼에는 t-1번째 프레임과 t번째 프레임에 대한 레지듀얼 $\hat{R}_{t,t-1}(\omega)$ 와, t번째 프레임과 t+1번째 프레임에 대한 레지듀얼 $\hat{R}_{t,t+1}(\omega)$ 이 존재함을 알 수 있다. 한편, 입력 버퍼에 존재하는 신호에 대해 역 주파수 변환을 수행함으로써 생성된 (합성) 레지듀얼 $\hat{r}_t(n)$ 이 출력버퍼에 존재한다.

<78> 다시 도 7을 참조하면, 룬텀 합성부(330)는 우선, 룬텀 예측이 적용되었는지 여부를 나타내는 룬텀 플래그 정보를 수신하고, 이를 근거로 룬텀 합성을 수행할지 여부를 결정한다. 룬텀 합성을 수행하는 경우, 레지듀얼 $\hat{r}_t(n)$ 과 룬텀 예측 정보(bo, do)를 이용하여 룬텀 합성을 수행함으로써, 현재 프레임에 대한 합성 오디오

신호 $\hat{S}_t(n)$ 를 생성한다. 여기서 룡텁 합성은 다음 수학과 같이 수행될 수 있다.

<79> [수학식 6]

$$\hat{S}_t(n) = \hat{r}_t(n) + b_0 \cdot \hat{S}_t(n-d_0)$$

<81> 여기서, $\hat{r}_t(n)$ 은 (합성) 레지듀얼, b_0 는 최종 게인, d_0 은 최종 지연, $\hat{S}_t(n)$ 은 현재 프레임의 합성 오디오 신호

<82> 이 룡텁 합성 과정은 앞서 도 1과 함께 설명한 룡텁 인코딩 장치에서의 룡텁 합성부(130)가 수행하는 과정과 유사하지만, 다만, 인코딩 장치의 룡텁 합성부(130)는 룡텁 예측 정보(후보 게인 및 후보 지연)을 근거로 룡텁 합성을 수행하는 데 비해, 룡텁 디코딩 장치의 룡텁 합성부(330)는 비트스트림을 통해 전송된 최종 게인 및 최종 지연에 대해 룡텁 합성을 수행한다는 점에서 차이가 있다. 앞서 언급한 바 대로, 최종 지연 d_0 은 t-1번째 프레임까지 포함한 이전 프레임을 포함한 프레임 중에서 찾은 결과이기 때문에, 최종 지연의 범위는 N(프레임 길이)부터가 아니라 0부터이다. N부터인 경우, N을 뺀 나머지 값만을 전송할 수 있지만, 0부터인 경우, N을 뺀 나머지 값이 아니라 그 값 그대로 전송될 수 있다. 최종 지연에서 특정 값을 뺀 값이 아니라 최종 지연 값 그대로 전송된 경우, 상기 수학식 6과 같이 룡텁 합성 과정에서, 최종 지연 d_0 이외에 다른 값(예: N)을 적용하지 않는다.

<83> 룡텁 디코딩 장치는 위와 같은 과정을 통해 룡텁 예측 정보 및 이전 프레임의 오디오 신호를 이용하여 현재 프레임의 오디오 신호를 복원한다.

<84> 도 9은 본 발명의 실시예에 따른 오디오 신호 처리 장치의 제1 예 (인코딩 장치)의 구성을 보여주는 도면이다. 도 9를 참조하면, 오디오 신호 인코딩 장치(400)는 복수채널 인코더(410), 대역확장 인코딩 장치(420), 오디오 신호 인코더(440), 음성 신호 인코더(450), 및 멀티 플렉서(460)를 포함할 수 있다. 물론, 본 발명의 실시예에 따른 룡텁 인코딩부(430)를 더 포함할 수 있다.

<85> 복수채널 인코더(410)는 복수의 채널 신호(둘 이상의 채널 신호)(이하, 멀티채널 신호)를 입력받아서, 다운믹스를 수행함으로써 모노 또는 스테레오의 다운믹스 신호를 생성하고, 다운믹스 신호를 멀티채널 신호로 업믹스하기 위해 필요한 공간 정보를 생성한다. 여기서 공간 정보는, 채널 레벨 차이 정보, 채널간 상관정보, 채널 예측 계수, 및 다운믹스 게인 정보 등을 포함할 수 있다. 만약, 오디오 신호 인코딩 장치(400)가 모노 신호를 수신할 경우, 복수 채널 인코더(410)는 모노 신호에 대해서 다운믹스하지 않고 바이패스할 수도 있음은 물론이다.

<86> 대역 확장 인코딩 장치(420)는 다운믹스 신호의 일부 대역(예: 고주파 대역)의 스펙트럴 데이터를 제외하고, 이 제외된 데이터를 복원하기 위한 대역확장정보를 생성할 수 있다.

<87> 룡텁 인코딩부(430)은 입력 신호에 대해서 룡텁 예측을 수행함으로써, 룡텁 예측 정보 b_0 , d_0 를 생성한다. 한편, 도 1과 함께 설명된 구성요소 중에서 일부 구성(200)(주파수 변환부(210), 양자화부(220), 심리음향모델(230))는 추후 설명한 오디오 신호 인코더(440) 또는 음성 신호 인코더(450)내에 포함되는 것일 수 있다. 따라서, 상기 일부 구성(200)이 제외된 룡텁 인코딩부(430)는 오디오 신호 인코더(440) 또는 음성 신호 인코더(450)에 임시 레지듀얼 $r_t(n)$ 을 전달하고, 주파수 변환된 레지듀얼 $\hat{R}_{t,t-1}(\omega)$ 을 수신한다.

<88> 오디오 신호 인코더(440)는 다운믹스 신호의 특정 프레임 또는 특정 세그먼트가 큰 오디오 특성을 갖는 경우, 오디오 코딩 방식(audio coding scheme)에 따라 다운믹스 신호를 인코딩한다. 여기서 오디오 코딩 방식은 AAC (Advanced Audio Coding) 표준 또는 HE-AAC (High Efficiency Advanced Audio Coding) 표준에 따른 것일 수 있으나, 본 발명은 이에 한정되지 아니한다. 한편, 오디오 신호 인코더(440)는, MDCT(Modified Discrete Transform) 인코더에 해당할 수 있다.

<89> 한편, 오디오 신호 인코더(440)는 앞서 설명한 바와 같이, 도 1과 함께 설명된 주파수 변환부(210), 양자화부(220), 심리음향모델(230)를 포함할 수 있다. 따라서 오디오 신호 인코더(440)는 룡텁 인코딩부 (430)로부터 임시 레지듀얼 $r_t(n)$ 을 수신하고, 주파수 변환된 레지듀얼 $\hat{R}_{t,t-1}(\omega)$ 을 생성하여 룡텁 인코딩부(430)에 전달

한다. 여기서 주파수 변환된 레지듀얼 $\hat{R}_{t,t-1}(\omega)$ 에 대하여 양자화를 수행한 결과인 스펙트럴 데이터 및 스케일 팩터는 멀티플렉서(460)로 전달될 수 있다.

- <90> 음성 신호 인코더(450)는 다운믹스 신호의 특정 프레임 또는 특정 세그먼트가 큰 음성 특성을 갖는 경우, 음성 코딩 방식(speech coding scheme)에 따라서 다운믹스 신호를 인코딩한다. 여기서 음성 코딩 방식은 AMR-WB(Adaptive multi-rate Wide-Band) 표준에 따른 것일 수 있으나, 본 발명은 이에 한정되지 아니한다. 한편, 음성 신호 인코더(450)는 선형 예측 부호화(LPC: Linear Prediction Coding) 방식을 더 이용할 수 있다. 하모닉 신호가 시간축 상에서 높은 중복성을 가지는 경우, 과거 신호로부터 현재 신호를 예측하는 선형 예측에 의해 모델링될 수 있는데, 이 경우 선형 예측 부호화 방식을 채택하면 부호화 효율을 높일 수 있다. 한편, 음성 신호 인코더(450)는 타임 도메인 인코더에 해당할 수 있다.
- <91> 멀티플렉서(460)는 공간정보, 대역확장 정보, 룬텀 예측 정보 및 스펙트럴 데이터 등을 다중화하여 오디오 신호 비트스트림을 생성한다.
- <92> 도 10은 본 발명의 실시예에 따른 오디오 신호 처리 장치의 제2 예(디코딩 장치)의 구성을 보여주는 도면이다. 도 10을 참조하면, 오디오 신호 디코딩 장치(500)는 디멀티플렉서(510), 오디오 신호 디코더(520), 음성 신호 디코더(530), 대역확장 디코딩 장치(550), 및 복수채널 디코더(560)를 포함한다. 본 발명에 따른 룬텀 디코딩부(540)를 더 포함한다.
- <93> 디멀티플렉서(510)는 오디오신호 비트스트림으로부터 스펙트럴 데이터, 대역확장정보, 룬텀 예측 정보 및 공간정보 등을 추출한다.
- <94> 오디오 신호 디코더(520)는, 다운믹스 신호에 해당하는 스펙트럴 데이터가 오디오 특성이 큰 경우, 오디오 코딩 방식으로 스펙트럴 데이터를 디코딩한다. 여기서 오디오 코딩 방식은 앞서 설명한 바와 같이, AAC 표준, HE-AAC 표준에 따를 수 있다. 한편 오디오 신호 디코더(520)는 앞서 도 7과 함께 설명된 역양자화부(310), 역변환부(320)를 포함할 수 있다. 따라서 오디오 신호 디코더(520)는 비트스트림을 통해 전송된 스펙트럴 데이터 및 스케일 팩터에 대해 역양자화를 수행하여 주파수 변환된 레지듀얼을 복원한다. 그런 다음 주파수 변환된 레지듀얼에 대해 역 주파수 변환을 수행하여 (역변환된) 레지듀얼을 생성한다.
- <95> 음성 신호 디코더(530)는 상기 스펙트럴 데이터가 음성 특성이 큰 경우, 음성 코딩 방식으로 다운믹스 신호를 디코딩한다. 음성 코딩 방식은, 앞서 설명한 바와 같이, AMR-WB 표준에 따를 수 있지만, 본 발명은 이에 한정되지 아니한다.
- <96> 룬텀 디코딩부(540)는 룬텀 예측 정보 및 (역변환된) 레지듀얼 신호를 이용하여 룬텀 합성을 수행함으로써 합성 오디오 신호를 복원한다. 룬텀 디코딩부(320)는 앞서 도 7과 함께 설명된 룬텀 합성부(330)를 포함할 수 있다.
- <97> 대역 확장 디코딩 장치(550)는 대역확장정보 비트스트림을 디코딩하고, 이 정보를 이용하여 오디오 신호 (또는 스펙트럴 데이터) 중 일부 또는 전부로부터 다른 대역(예: 고주파대역)의 오디오 신호 (또는 스펙트럴 데이터)를 생성한다.
- <98> 복수채널 디코더(560)은 디코딩된 오디오 신호가 다운믹스인 경우, 공간정보를 이용하여 멀티채널 신호(스테레오 신호 포함)의 출력 채널 신호를 생성한다.
- <99> 본 발명에 따른 룬텀 인코딩 장치 또는 룬텀 디코딩 장치는 다양한 제품에 포함되어 이용될 수 있다. 이러한 제품은 크게 스탠드 얼론(stand alone) 군과 포터블(portable) 군으로 나눌 수 있는데, 스탠드 얼론군은 티비, 모니터, 셋탑 박스 등을 포함할 수 있고, 포터블군은 PMP, 휴대폰, 네비게이션 등을 포함할 수 있다.
- <100> 도 11은 본 발명의 실시예에 따른 룬텀 코딩(인코딩 및/또는 디코딩) 장치가 구현된 제품의 개략적인 구성을 보여주는 도면이다. 도 11은 본 발명의 실시예에 따른 룬텀 코딩(인코딩부 및/또는 디코딩부)장치가 구현된 제품들의 관계를 보여주는 도면이다.
- <101> 우선 도 11을 참조하면, 유무선 통신부(610)는 유무선 통신 방식을 통해서 비트스트림을 수신한다. 구체적으로 유무선 통신부(610)는 유선통신부(610A), 적외선통신부(610B), 블루투스부(610C), 무선랜통신부(610D) 중 하나 이상을 포함할 수 있다.
- <102> 사용자 인증부는(620)는 사용자 정보를 입력 받아서 사용자 인증을 수행하는 것으로서 지문인식부(620A), 홍채인식부(620B), 얼굴인식부(620C), 및 음성인식부(620D) 중 하나 이상을 포함할 수 있는데, 각각 지문, 홍채정보, 얼굴 윤곽 정보, 음성 정보를 입력받아서, 사용자 정보로 변환하고, 사용자 정보 및 기존 등록되어

있는 사용자 데이터와의 일치여부를 판단하여 사용자 인증을 수행할 수 있다.

- <103> 입력부(630)는 사용자가 여러 종류의 명령을 입력하기 위한 입력장치로서, 키패드부(630A), 터치패드부(630B), 리모컨부(630C) 중 하나 이상을 포함할 수 있지만, 본 발명은 이에 한정되지 아니한다. 신호 코딩부(640)는 룬텀 코딩 장치(645)(룬텀 인코딩 장치 및/또는 룬텀 디코딩 장치)를 포함하는데, 룬텀 인코딩 장치(645)는 앞서 도 1과 함께 설명된 룬텀 인코딩 인코딩 장치 중 적어도 룬텀 예측부, 역변환부, 룬텀 합성부, 예측정보 결정부를 포함하는 장치로서, 소스 오디오 신호에 대해 룬텀 예측을 수행하여 임시 게인 및 임시 지연을 생성하고, 룬텀 합성 및 예측정보 결정을 수행함으로써 최종 게인 및 최종 지연을 생성한다. 한편, 룬텀 디코딩 장치(미도시)는 앞서 도 7과 함께 설명된 룬텀 디코딩 장치 중 적어도 룬텀 합성부를 포함하는 장치로서, 룬텀 레지듀얼 및 최종 룬텀 예측정보를 근거로 룬텀 합성을 수행함으로써, 합성 오디오 신호를 생성한다.
- <104> 신호 코딩부(650)는 입력 신호를 양자화하여 인코딩하여 비트스트림을 생성하거나, 수신된 비트스트림 및 스펙트럴 데이터를 이용하여 신호를 디코딩하여 출력신호를 생성한다.
- <105> 제어부(650)는 입력장치들로부터 입력 신호를 수신하고, 신호 코딩부(640)와 출력부(660)의 모든 프로세스를 제어한다. 출력부(660)는 신호 코딩부(640)에 의해 생성된 출력 신호 등이 출력되는 구성요소로서, 스피커부(660A) 및 디스플레이부(660B)를 포함할 수 있다. 출력 신호가 오디오 신호일 때 출력 신호는 스피커로 출력되고, 비디오 신호일 때 출력 신호는 디스플레이를 통해 출력된다.
- <106> 도 12는 도 11에서 도시된 제품에 해당하는 단말 및 서버와의 관계를 도시한 것으로서, 도 12의 (A)를 참조하면, 제1 단말(600.1) 및 제2 단말(600.2)이 각 단말들은 유무선 통신부를 통해서 데이터 내지 비트스트림을 양방향으로 통신할 수 있음을 알 수 있다. 도 12의 (B)를 참조하면, 서버(700) 및 제1 단말(600.1) 또한 서로 유무선 통신을 수행할 수 있음을 알 수 있다.
- <107> 본 발명에 따른 오디오 신호 처리 방법은 컴퓨터에서 실행되기 위한 프로그램으로 제작되어 컴퓨터가 읽을 수 있는 기록 매체에 저장될 수 있으며, 본 발명에 따른 데이터 구조를 가지는 멀티미디어 데이터도 컴퓨터가 읽을 수 있는 기록 매체에 저장될 수 있다. 상기 컴퓨터가 읽을 수 있는 기록 매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 저장 장치를 포함한다. 컴퓨터가 읽을 수 있는 기록 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있으며, 또한 캐리어 웨이브(예를 들어 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다. 또한, 상기 인코딩 방법에 의해 생성된 비트스트림은 컴퓨터가 읽을 수 있는 기록 매체에 저장되거나, 유/무선 통신망을 이용해 전송될 수 있다.
- <108> 이상과 같이, 본 발명은 비록 한정된 실시예와 도면에 의해 설명되었으나, 본 발명은 이것에 의해 한정되지 않으며 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 본 발명의 기술사상과 아래에 기재될 특허 청구범위의 균등범위 내에서 다양한 수정 및 변형이 가능함은 물론이다.

산업이용 가능성

- <109> 본 발명은 오디오 신호를 인코딩하고 디코딩하는 데 적용될 수 있다.

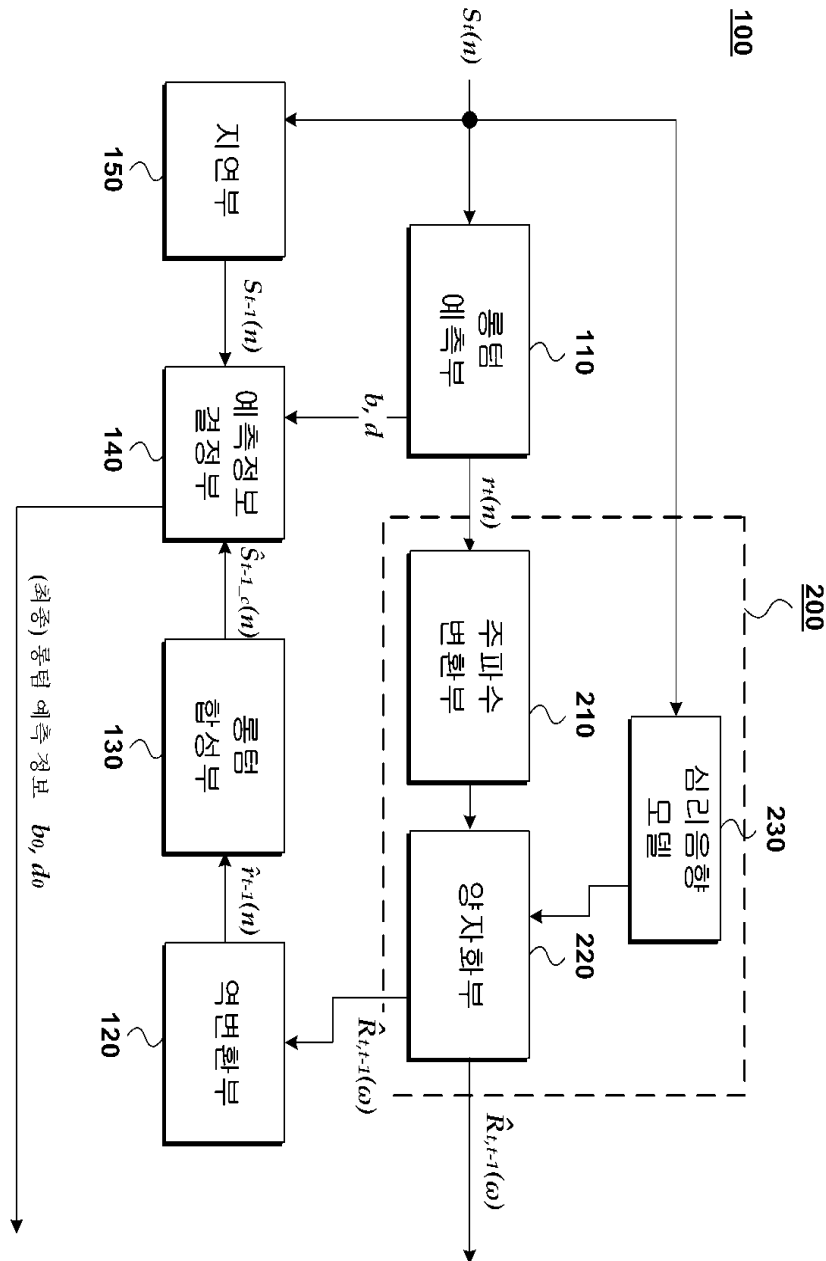
도면의 간단한 설명

- <110> 도 1은 본 발명의 실시예에 따른 오디오 신호 처리 장치 중 룬텀 인코딩 장치의 구성도.
- <111> 도 2는 본 발명의 실시예에 따른 오디오 신호 처리 방법의 순서도.
- <112> 도 3은 프레임별 소스 시그널에 대한 개념을 설명하기 위한 도면.
- <113> 도 4는 룬텀 예측 과정(S110 단계)을 설명하기 위한 도면.
- <114> 도 5는 주파수 변환 과정(S120 단계) 및 역 주파수 변환 과정(S130 단계)을 설명하기 위한 도면.
- <115> 도 6은 룬텀 합성 과정(S150 단계) 및 예측정보 결정 과정(S160 단계)을 설명하기 위한 도면.
- <116> 도 7은 본 발명의 실시예에 따른 오디오 신호 처리 장치 중 룬텀 디코딩 장치의 구성도.
- <117> 도 8는 룬텀 디코딩 장치에서의 역양자화 과정 및 룬텀 합성 과정을 설명하기 위한 도면.
- <118> 도 9는 본 발명의 실시예에 따른 오디오 신호 처리 장치의 제1 예의 구성도(인코딩 장치).
- <119> 도 10은 본 발명의 실시예에 따른 오디오 신호 처리 장치의 제2 예의 구성도(디코딩 장치).

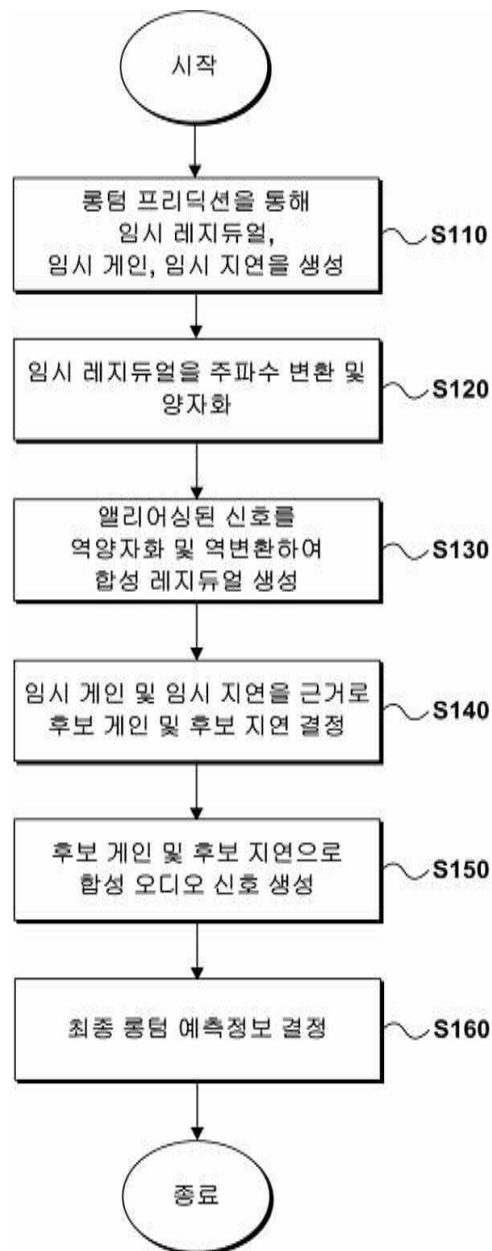
- <120> 도 11은 본 발명의 실시예에 따른 룬텀 코딩(인코딩 및/또는 디코딩) 장치가 구현된 제품의 개략적인 구성도.
- <121> 도 12는 본 발명의 실시예에 따른 룬텀 코딩(인코딩 및/또는 디코딩) 장치가 구현된 제품들의 관계도.

도면

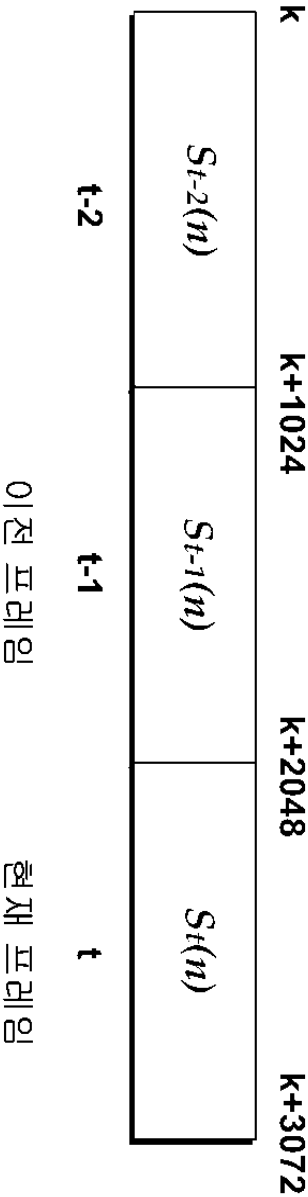
도면1



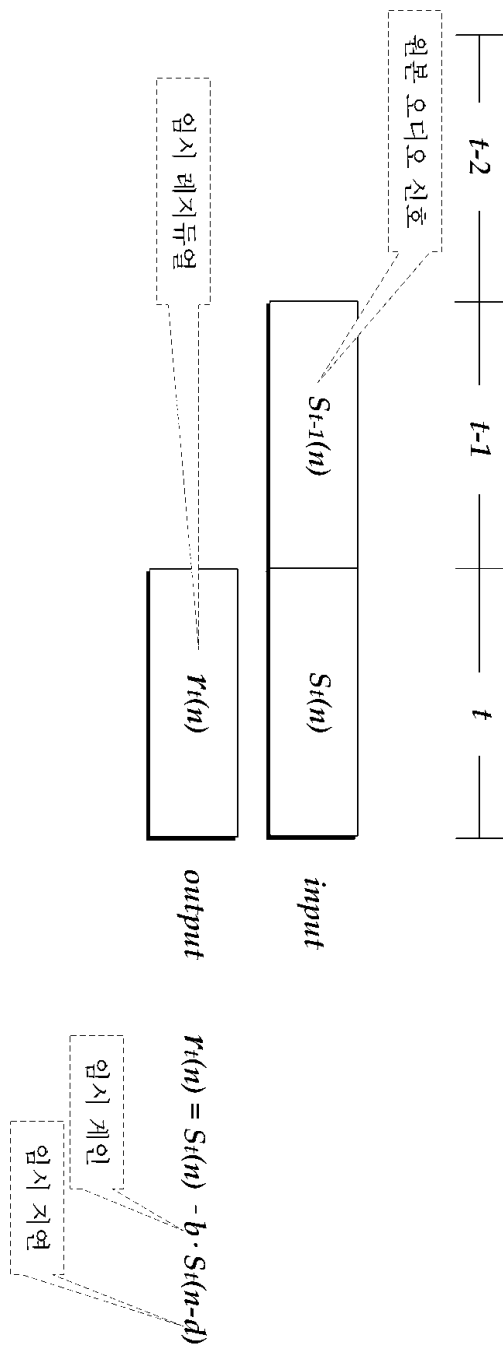
도면2



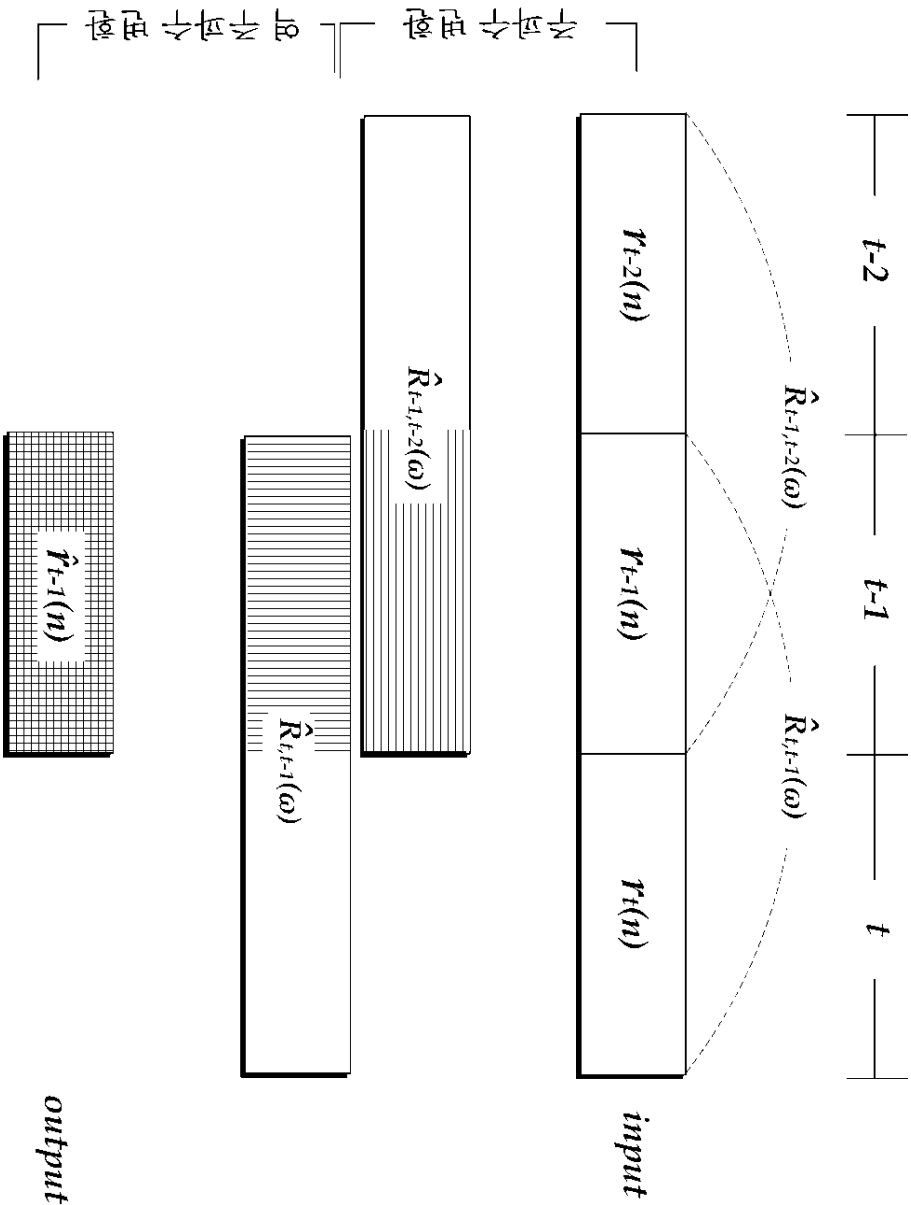
도면3



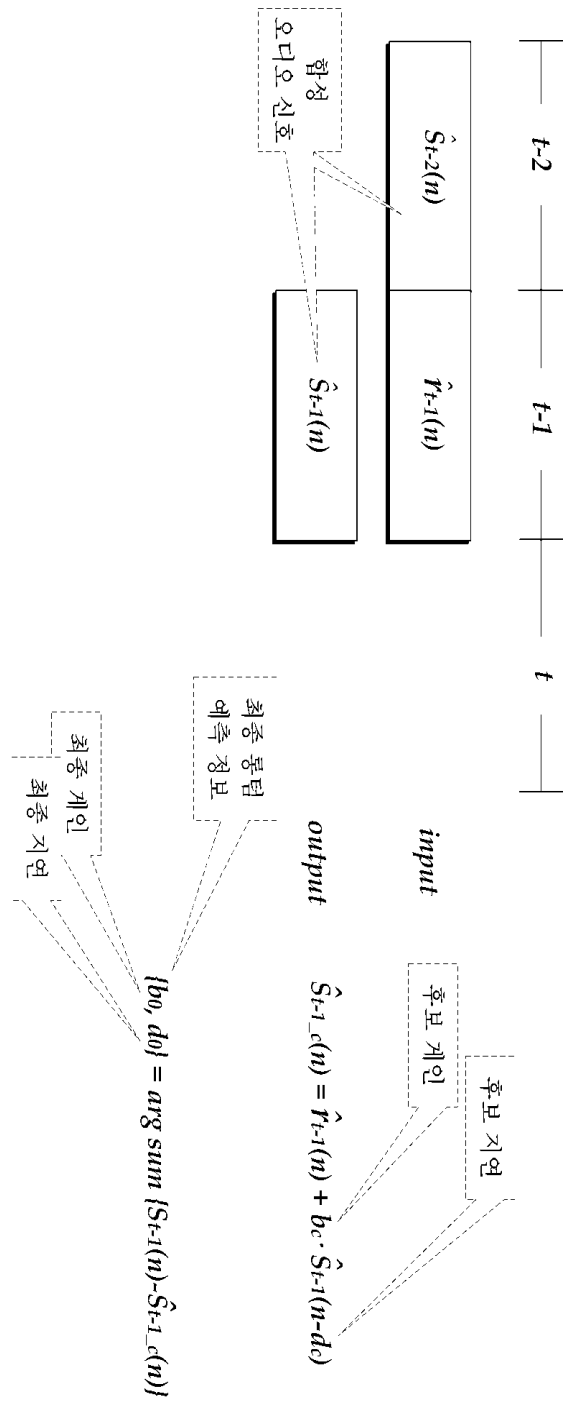
도면4



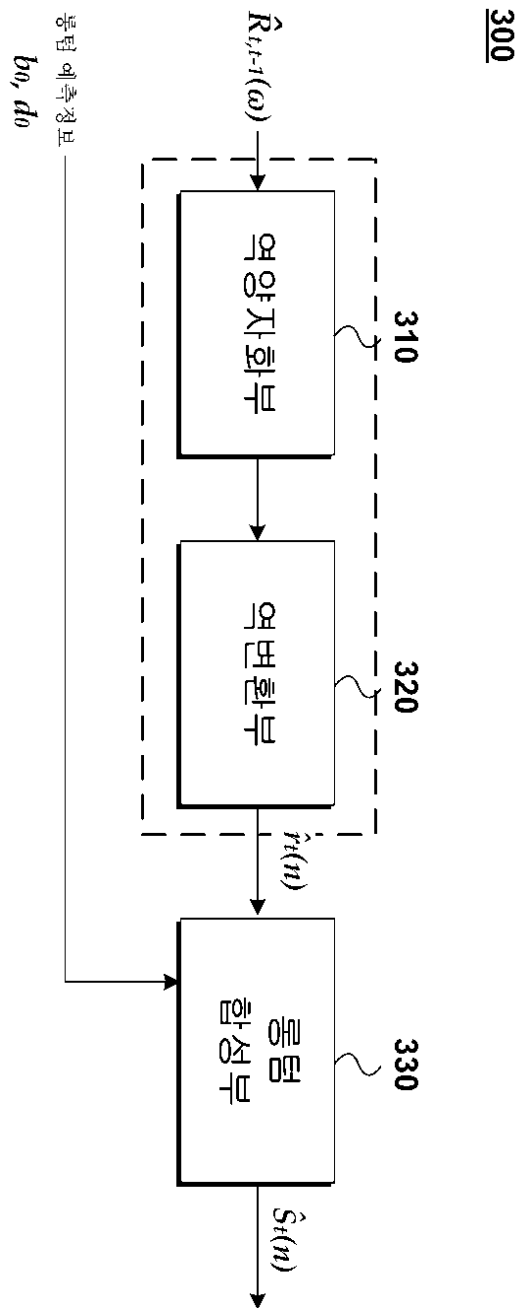
도면5



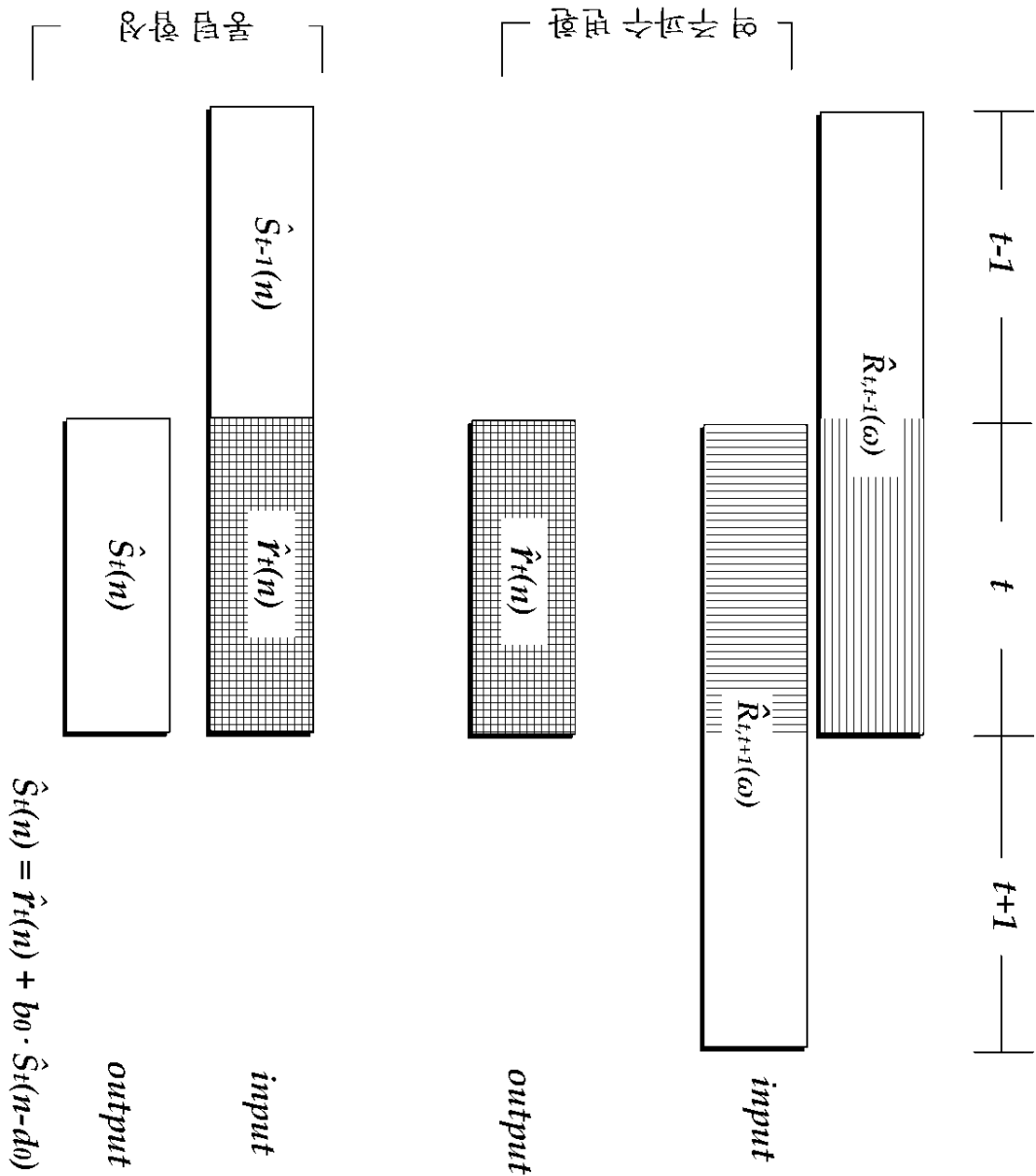
도면6



도면7

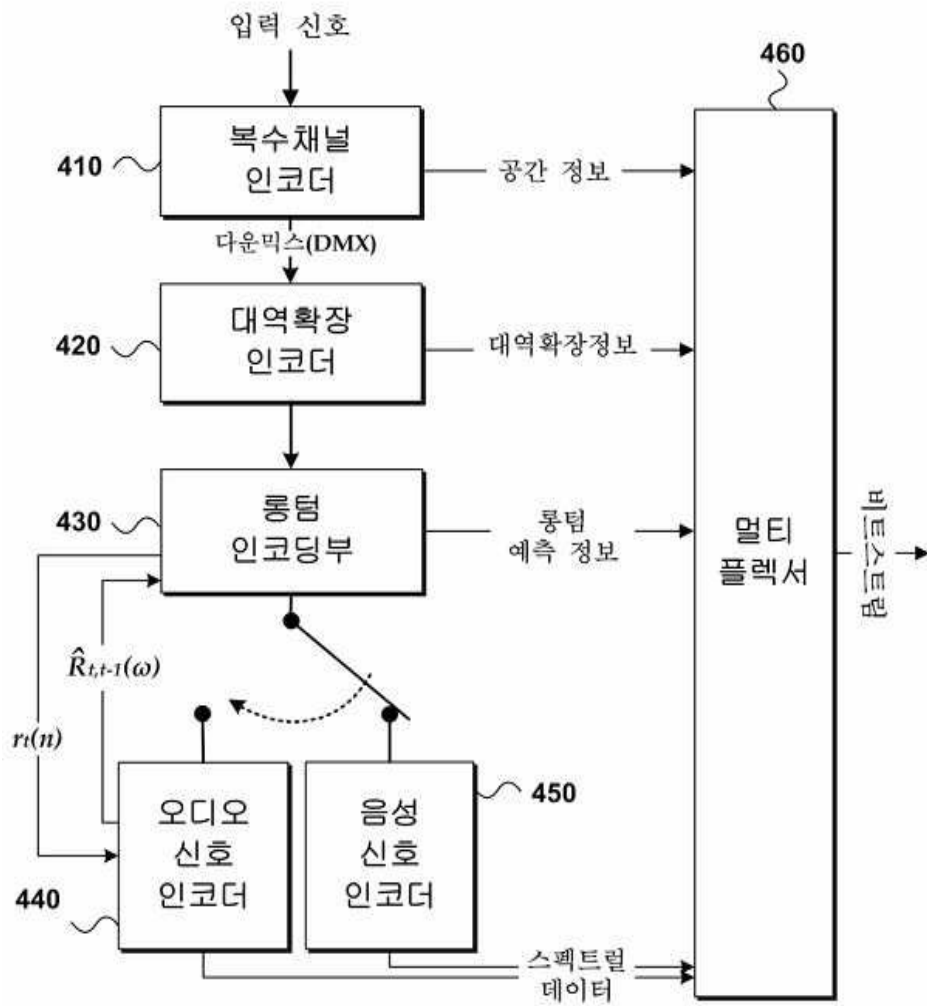


도면8

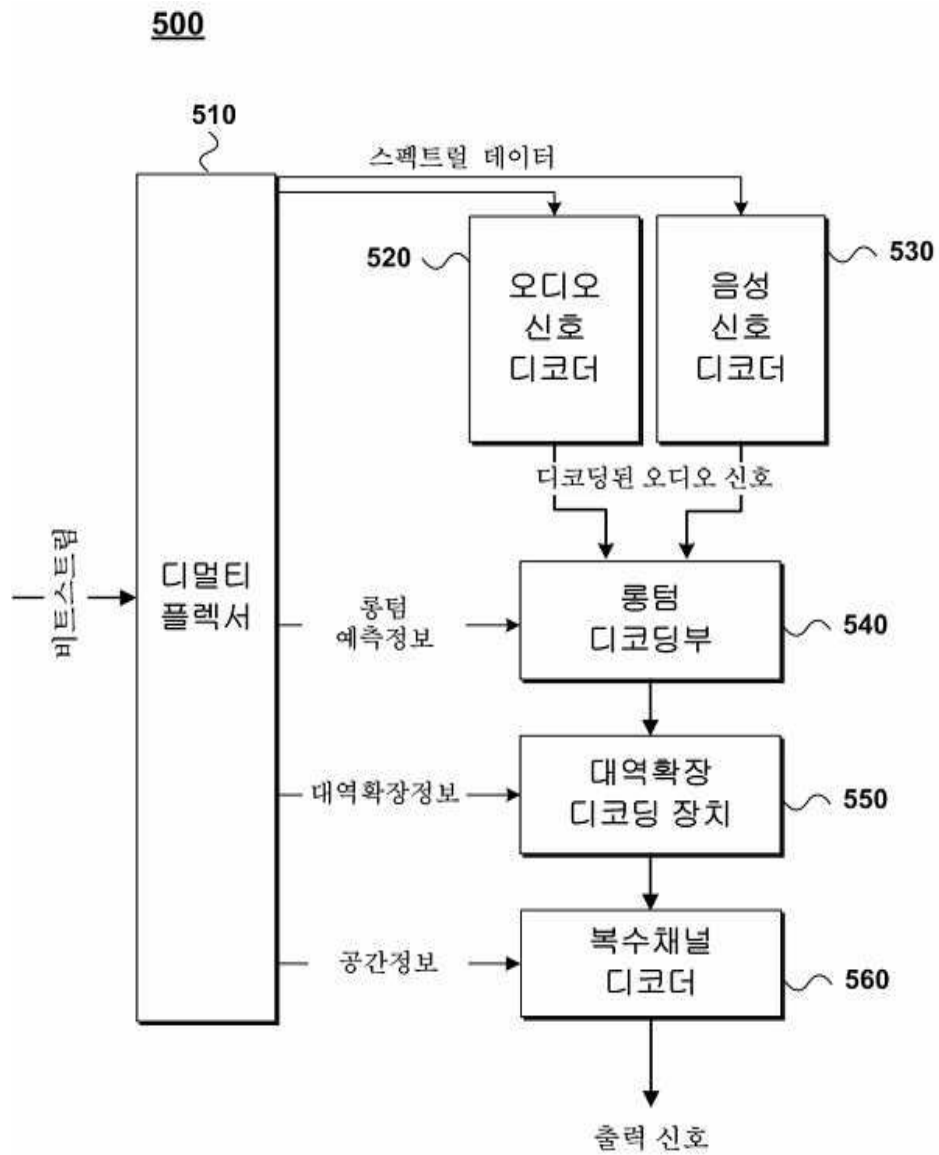


도면9

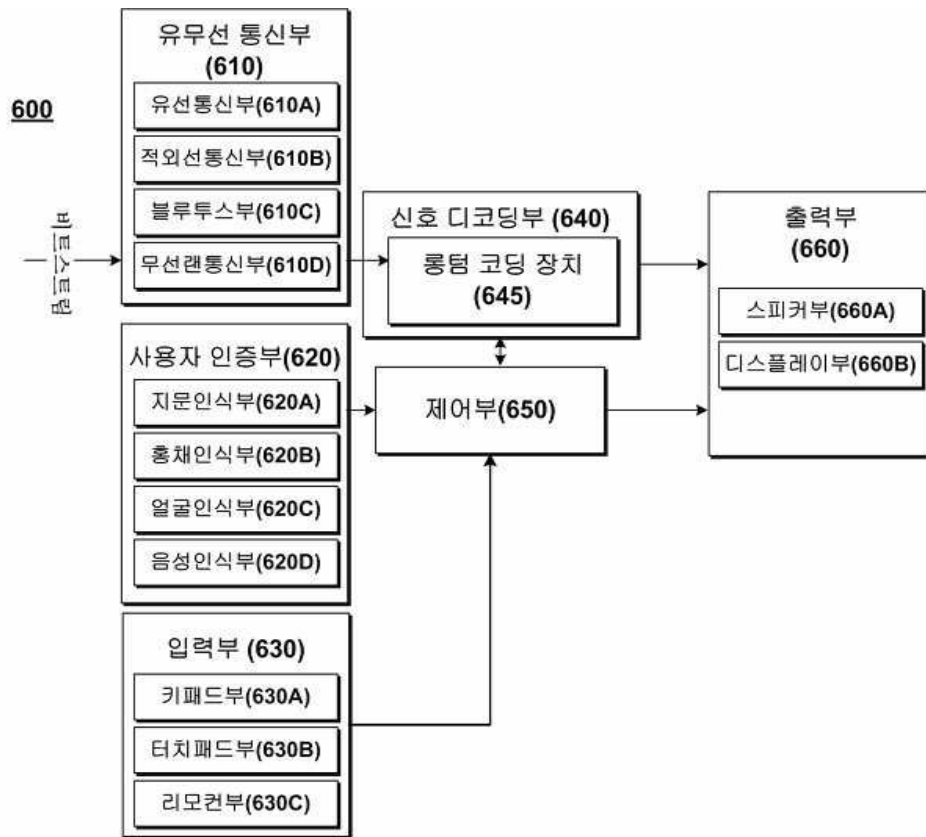
400



도면10



도면11



도면12

