(19)대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) 。Int. Cl.[6]
H04B 1/66

(45) 공고일자 2005년01월29일
(11) 등록번호 10-0469002
(24) 등록일자 2005년01월20일

(21) 출원번호 10-1999-7008369
(22) 출원일자 1999년09월14일
번역문 제출일자 1999년09월14일
(86) 국제출원번호 PCT/FI1998/000146
(86) 국제출원출원일자 1998년02월18일
(65) 공개번호 10-2000-0076273
(43) 공개일자 2000년12월26일
(87) 국제공개번호 WO 1998/42083
(87) 국제공개일자 1998년09월24일

(81) 지정국 국내특허 : 알바니아, 아르메니아, 오스트리아, 오스트레일리아, 아제르바이잔, 보스니아-헤르체고비나, 바베이도스, 불가리아, 브라질, 벨라루스, 캐나다, 스위스, 리히텐슈타인, 중국, 쿠바, 체코, 독일, 덴마크, 에스토니아, 스페인, 핀랜드, 영국, 그루지야, 헝가리, 이스라엘, 아이슬란드, 일본, 케냐, 키르기즈, 북한, 대한민국, 카자흐스탄, 세인트루시아, 스리랑카, 라이베리아, 레소토, 리투아니아, 룩셈부르크, 라트비아, 몰도바, 마다가스카르, 마케도니아, 몽고, 말라위, 멕시코, 노르웨이, 뉴질랜드, 슬로베니아, 슬로바키아, 타지키스탄, 투르크메니스탄, 터어키, 트리니다드토바고, 우크라이나, 우간다, 미국, 우즈베키스탄, 베트남, 폴란드, 포르투칼, 루마니아, 러시아, 수단, 스웨덴, 가나, 감비아, 기네비쏘, 인도네시아, 시에라리온, 세르비아 앤 몬테네그로, 짐바브웨,

AP ARIPO특허 : 케냐, 레소토, 말라위, 수단, 스와질랜드, 우간다, 가나, 감비아,

EA 유라시아특허 : 아르메니아, 아제르바이잔, 벨라루스, 키르기즈, 카자흐스탄, 몰도바, 러시아, 타지키스탄, 투르크메니스탄,

EP 유럽특허 : 오스트리아, 벨기에, 스위스, 리히텐슈타인, 독일, 덴마크, 스페인, 프랑스, 영국, 그리스, 아일랜드, 이탈리아, 룩셈부르크, 모나코, 네덜란드, 포르투칼, 스웨덴, 핀랜드,

OA OAPI특허 : 부르키나파소, 베넹, 중앙아프리카, 콩고, 코트디브와르, 카메룬, 가봉, 기네, 말리, 모리타니, 니제르, 세네갈, 차드, 토고,

(30) 우선권주장 971108 1997년03월14일 핀랜드(FI)

(73) 특허권자 노키아 모빌 폰즈 리미티드
핀랜드 핀-02150 에스푸 카일알라덴티에 4

(72) 발명자 인린
미국캘리포니아95035밀피타스딕손랜딩로드No.A105440

(74) 대리인 이영필

심사관 : 정재우

(54) 오디오 코딩 방법 및 장치

요약

오디오 신호 코딩 방법은, 코딩될 오디오 신호 x를 수신하는 단계 및 수신된 신호를 시간에서 주파수 도메인으로 변환하는 단계를 구비한다. 양자화된 오디오 신호 $\widetilde{x}$ 는, 양자화된 오디오 신호 $\widetilde{x}$ 의 한 개 이상의 이전 시간 프레임들로부터 직접, 수신된 오디오 신호의 현시간 프레임을 예측하는데 사용될 수 있는 장기간 예측 계수의 집합 A와 함께 상기 변환된 오디오 신호 x로부터 생성된다. 예측된 오디오 신호 $\widehat{x}$ 는 예측 계수들 A를 사용하여 생성된다. 예측된 오디오 신호 $\widehat{x}$ 는 그리고나서 시간에서 주파수 도메인으로 변환되고 그 결과의 주파수 도메인 신호는 수신된 오디오 신호 x의 주파수 도메인 신호와 비교되어 다수의 주파수 서브-밴드 각각에 대한 에러 신호 E(k)를 생성한다. 에러 신호 E(k)는 그리고나서 양자화되어, 코딩된 오디오 신호를 생성하기 위해 예측 계수 A와 결합되는 양자화된 에러 신호 집합 $\widetilde{E(k)}$ 를 발생한다.

대표도

도 3

명세서

기술분야

본 발명은 오디오 코딩을 위한 방법 및 장치, 그리고 오디오 디코딩을 위한 방법 및 장치에 관한 것이다.

배경기술

디지털 형태의 데이터 전송이 전송 채널을 따라 증가되는 신호대잡음비 및 증가되는 정보 용량을 위해 제공된다는 것은 잘 알려진 사실이다. 그러나 디지털 신호를 이미 상당한 정도로 압축함으로써 채널 용량을 더 증가시키고자 하는 지속적인 욕구가 있어왔다. 오디오 신호와 관련하여, 두 개의 기본적인 압축 원리가 통상적으로 적용된다. 이들 중 첫 번째는 소스 신호의 통계적 또는 결정론적 중복을 제거함을 수반하는 한편 두 번째는 인간의 지각에 관한 한 중복을 소스 신호 성분들로부터 억압 또는 제거함을 수반한다. 최근, 후자의 원리가 고품질 오디오 어플리케이션들에서 우위를 점하게 되었고 일반적으로 오디오 신호를 그 주파수 성분들(때로는 '서브-밴드들'로 칭함)로 분리함을 수반하며, 주파수 성분들 각각은 (청취자와의) 데이터 무관계성을 없애도록 정해진 양자화 정확도를 가지고 해석 및 양자화된다. ISO(국제 표준 기구;International Standard Organization) MPEG(이동 화상 전문가 그룹;Moving Pictures Expert Group) 오디오 코딩 표준 및 그외의 오디오 코딩 표준들이 이 원리를 채용 및 추가로 정의하고 있다. 그러나, MPEG(및 그외의 표준들) 또한 '적응적 예측'으로 알려진 기술을 채용하여 데이터 레이트의 추가 감소를 제공한다. 공개번호 EP 0 396 121은, 비록 두 개의 서브-밴드들만이 거기에 사용되고 있지만, 적응 예측의 예를 가리키고 있다.
신 MPEG-2AAC 표준에 따른 인코더의 동작이 국제 표준 문서 초안 ISO/IEC DIS13818-7에 상세히 설명되어 있다. 이 신 MPEG-2 표준은 1024개의 주파수 성분들 중 672개와 함께 역방향 선형 예측을 채용한다. 신 MPEG-4 표준이 이와 유사한 요구조건을 가질 것이라는 것이 예견된다. 그러나, 그러한 많은 수의 주파수 성분들은 예측 알고리즘의 복잡성으로 인한 커다란 계산적 부담이라는 결과를 초래하고 또한 산출된 계수 및 중개 계수들을 저장하기 위한 많은 양의 메모리 사용가능 용량을 필요로한다. 이러한 종류의 역방향 적응 예측기들이 주파수 도메인에 사용될 때, 계산의 부담과 메모리의 필요성을 더 감소시키는 것이 어렵다는 것은 잘 알려져 있다. 이것은 예측기들의 수가 주파수 도메인에서 매우 크기 때문에 간단한 적응 알고리즘 조차도 여전히 엄청난 계산의 복잡성 및 메모리의 필요성을 초래하기 때문이다. 인코더에서 갱신되고 디코더로 전송되는 순방향 적응 예측기를 사용하여 이러한 문제를 피한다는 것이 알려져 있지만, 예측기들의 개수가 매우 많기 때문에 주파수 도메인의 순방향 적응 예측기들의 사용은 불가피하게 많은 양의 '부(side)' 정보를 초래한다.

발명의 상세한 설명

본 발명의 목적은 종래의 예측 방법들의 단점을 극복하거나 최소한 완화시키도록 하는 것이다.
이것과 그 이외의 목적들은, 오디오 신호의 다수의 주파수 서브-밴드들 각각 의 중복을 제거하기 위해 에러 신호를 사용하고 추가로 오디오 신호의 현재의 프레임이 한 개 이상의 이전 프레임들로부터 예측될 수 있게 하는 시간 도메인내 장기간 예측 계수들을 발생하여 오디오 신호를 코딩함으로써 성취된다.
본 발명의 첫 번째 양상에 의하면, 오디오 신호의 코딩 방법이 제공되며, 이 방법은,
코딩될 오디오 신호 x를 수신하는 단계; 수신된 오디오 신호의 타임 프레임으로부터 주파수 서브-밴드들을 발생하는 단계;
수신된 오디오 신호 x로부터 양자화된 오디오 신호 $\widetilde{x}$ 를 발생시키는 단계;

장기간(long-term) 예측 계수들의 집합 A를 발생하는 단계; 상기 양자화된 오디오 신호 $\widetilde{x}$ 의 최소한 한 개의 이전 시간 프레임으로부터 직접 시간 프레임의 각 서브-밴드들에 대한 장기간 예측 계수들의 집합 A를 사용하여 수신된 오디오 신호 x의 현재 시간 프레임을 예측하는 단계; 장기간 예측 계수들의 집합 A를 사용하여 양자화된 오디오 신호 $\widetilde{x}$ 의 예측된 오디오 신호 $\hat{x}$ 를 발생시키는 단계;

수신된 오디오 신호 x를 예측된 오디오 신호 $\hat{x}$ 와 비교하고 다수의 주파수 서브-밴드들 각각에 대해 에러 신호 E(k)를 발생하는 단계;

에러 신호 E(k)를 양자화하여 양자화된 에러 신호 $\tilde{E}(k)$ 를 발생하는 단계; 및

양자화된 에러 신호 $\tilde{E}(k)$ 와 예측 계수들 A를 결합하여 부호화된 오디오 신호를 발생시키는 단계를 구비한다.

본 발명은 시간 도메인의 순방향 적응 예측기를 사용하는 오디오 신호의 압축을 위해 제공된다. 수신된 신호의 각 시간 프레임에 있어서, 디코더로의 전송을 위해 순방향 적응 예측 계수들의 하나의 집합을 발생 및 전송하는 것만이 필요하게 된다. 이것은 각 시간 프레임의 각 주파수 서브-밴드에 대해 예측 계수들의 집합의 발생을 필요로 하는 종래의 순방향 적응 예측 기술과는 반대이다. 본 발명에 의해 얻어지는 예측 이득과 비교하여, 장기간 예측기의 부정보는 무시할만하다.

본 발명의 어떤 실시예들은 계산의 복잡성 및 메모리 필요성의 감소를 가능하게 한다. 특히, 역방향 적응 예측의 사용과 비교하여, 디코더의 예측 계수들을 재산출해야할 필요성이 없게된다. 본 발명의 어떤 실시예들은 또한 통상적인 역방향 적응 예측기보다 더 빠르게 신호 변화에 응답할 수 있다.

본 발명의 일실시예에서, 수신된 오디오 신호 x는 프레임 x_m 에서 시간 도메인으로부터 주파수 도메인으로 변환하여 주파수 서브-밴드 신호의 집합 X(k)를 제공한다. 예측된 오디오 신호 $\hat{x}$ 는 이와 유사하게 시간 도메인으로부터 주파수 도메인으로 변환하여 예측된 주파수 서브-밴드 신호들의 집합 $\hat{X}(k)$ 을 생성하며, 수신된 오디오 신호 x와 예측된 오디오 신호 $\hat{x}$ 사이의 비교는, 주파수 서브-밴드 에러 신호들 E(k)를 생성하기 위해 각각의 서브-밴드 신호들을 서로에 대해 비교함으로써 주파수 도메인에서 수행된다. 양자화된 오디오 신호 $\widetilde{x}$ 는 시간 도메인 또는 주파수 도메인에서, 예측된 신호와 양자화된 에러 신호를 더함으로써 생성된다.

본 발명의 다른 실시예에 있어서, 수신된 오디오 신호 x와 예측된 오디오 신호 $\hat{x}$ 의 비교가 시간 도메인에서 수행되어 역시 시간 도메인의 에러 신호 e를 생성한다. 이 에러 신호 e는 그리고나서 시간에서 주파수 도메인으로 전환되어 상기 다수의 주파수 서브-밴드 에러 신호들 E(k)를 생성한다.

에러 신호들의 양자화는 심리-음향학적(psycho-acoustic) 모델에 따라 수행된다.

본 발명의 두 번째 양상에 의하면, 코딩된 오디오 신호의 디코딩 방법이 제공되며, 그 방법은:

오디오 신호의 다수의 주파수 서브-밴드들 각각에 있어서 양자화된 에러 신호들 $\tilde{E}(k)$ 및, 오디오 신호의 각 시간 프레임에 있어서 재구성된 양자화 오디오 신호 $\widetilde{x}$ 의 최소한 한 개의 이전 시간 프레임으로부터 직접, 수신된 오디오 신호의 현재 시간 프레임 x_m 의 각 주파수 서브-밴드를 예측하는데 사용될 수 있는 예측 계수들의 집합 A를 구비하는 코딩된 오디오 신호를 수신하는 단계;

양자화된 에러 신호들 $\tilde{E}(k)$ 로부터 상기 재구성된 양자화 오디오 신호 $\widetilde{x}$ 를 생성하는 단계;

예측 계수들 A 및 양자화된 오디오 신호 $\widetilde{x}$ 를 사용하여 예측된 오디오 신호 $\hat{x}$ 를 생성하는 단계;

예측된 오디오 신호 $\hat{x}$ 를 시간 도메인에서 주파수 도메인으로 변환하여, 재구성된 주파수 서브-밴드 신호들의 집합 $\tilde{X}(k)$ 을 생성하기 위해 양자화된 에러 신호들 $\tilde{E}(k)$ 와 결합하도록 예측된 주파수 서브-밴드 신호들의 집합 $\hat{X}(k)$ 를 생성하는 단계; 및

재구성된 주파수 서브-밴드 신호들 $\tilde{X}(k)$ 에 대해 주파수에서 시간 도메인으로의 변환을 수행하여, 재구성된 양자화 오디오 신호 $\widetilde{x}$ 를 생성하는 단계를 포함한다.

본 발명의 상기 두 번째 양상의 실시예들은 모든 가능한 양자화 에러 신호들 $\tilde{E}(k)$ 의 부분 집합만이 수신되고 이때 몇몇 부분 집합은 오디오 서브-밴드 신호들 X(k)의 전송에 의해 직접적으로 전송되는 데에 특히 적용가능하다. 신호들 $\tilde{X}(k)$ 와 X(k)는 주파수-시간 변환을 수행하기 앞서 적합하게 결합된다.

본 발명의 세 번째 양상에 의하면, 오디오 신호를 코딩하는 장치가 제공되며, 그 장치는:

코딩될 오디오 신호 x를 수신하기 위한 입력부;

수신된 오디오 신호의 시간 프레임으로부터 주파수 서브-밴드들을 생성하는 제1생성수단; 수신된 오디오 신호 x로부터 양자화된 오디오 신호 $\widetilde{x}$ 를 생성하기 위해 상기 입력부와 결합된 양자화 수단을 구비하고,

양자화된 오디오 신호 $\widetilde{x}$ 의 최소한 한 개의 이전 시간 프레임으로부터 직접적으로, 수신된 오디오 신호 x의 현재 시간 프레임 x_m 을 예측하기 위해 시간 프레임의 각 서브-밴드들에 대해 사용될, 장기간 예측 계수들의 집합 A를 발생

하도록 상기 양자화 수단에 결합된 예측 수단;
예측 계수들 A 및 양자화된 오디오 신호 $\tilde{x}$ 를 사용하여 예측된 오디오 신호 $\hat{x}$ 를 생성하고 수신된 오디오 신호 x를 예측된 오디오 신호 $\hat{x}$ 와 비교하여 다수의 주파수 서브-밴드들 각각에 대해 에러 신호 E(k)를 생성하는 제2생성 수단;

에러 신호들 E(k)를 양자화하여 양자화된 에러 신호들의 집합 $\tilde{E}(k)$ 를 생성하는 양자화 수단; 및

양자화된 에러 신호들 $\tilde{E}(k)$ 을 예측 계수들 A와 결합하여 부호화된 오디오 신호를 생성하는 결합 수단을 포함한다.
일실시예에서, 상기 생성 수단은 수신된 오디오 신호 x를 시간에서 주파수 도메인 변환하는 제1변환 수단, 예측된 오디오 신호 $\hat{x}$ 를 시간에서 주파수 도메인으로 변환하는 제2변환 수단 및, 상기 결과의 주파수 도메인의 주파수 신호들을 비교하도록 정해진 비교 수단을 포함한다.
본 발명의 다른 실시예에서, 생성 수단은 시간 도메인의 수신된 오디오 신호 x와 예측된 오디오 신호 $\hat{x}$ 를 비교하도록 정해진다.
본 발명의 네 번째 양상에 의하면, 코딩된 오디오 신호 x를 디코딩하는 장치가 제공되고, 이때 코딩된 오디오 신호가 오디오 신호의 다수의 주파수 서브-밴드들 각각에 대해 양자화된 에러 신호 $\tilde{E}(k)$ 및, 오디오 신호의 시간 프레임 중 각 주파수 서브-밴드들에 대해 사용될 예측 계수들의 집합 A를 포함하고, 상기 예측 계수 A는 재구성된 양자화 오디오 신호 $\tilde{x}$ 의 최소한 한 개의 이전 시간 프레임으로부터 직접적으로, 수신된 오디오 신호의 현 시간 프레임 x_m 을 예측하는데 사용될 수 있을 때, 상기 장치는:
코딩된 오디오 신호를 수신하기 위한 입력부;

양자화된 에러 신호들 $\tilde{E}(k)$ 로부터 상기 재구성된 양자화 오디오 신호 $\tilde{x}$ 를 생성하기 위한 생성 수단; 및
예측 계수들 A와 상기 재구성된 오디오 신호 $\tilde{x}$ 로부터 예측된 오디오 신호 $\hat{x}$ 를 생성하기 위한 신호 처리 수단을 구비하고,
이때 상기 생성 수단은, 예측된 오디오 신호 $\hat{x}$ 를 시간 도메인에서 주파수 도메인으로 변환하여 예측된 주파수 서브-밴드 신호들의 집합 $\hat{X}(k)$ 을 생성하는 제1변환 수단, 상기 예측된 주파수 서브-밴드 신호들의 집합 $\hat{X}(k)$ 과 양자화된 에러 신호들 $\tilde{E}(k)$ 을 결합하여 재구성된 주파수 서브-밴드 신호들의 집합 $\tilde{X}(k)$ 를 생성시키는 결합 수단 및, 재구성된 주파수 서브-밴드 신호들 $\tilde{X}(k)$ 에 대해 주파수-시간 도메인 변환을 수행하여 재구성된 양자화 오디오 신호 $\tilde{x}$ 를 생성하는 제2변환 수단을 포함한다.
본 발명의 보다 나은 이해를 위해, 그리고 본 발명이 어떻게 실시될 수 있는지를 보이기 위해, 예를 통해 첨부된 도면의 참조가 이제 이뤄질 것이다.

도면의 간단한 설명

도 1은 수신된 오디오 신호를 코딩하는 인코더를 개략적으로 도시한다.
도 2는 도 1의 인코더로 코딩된 오디오 신호를 디코딩하는 디코더를 개략적으로 도시한다.
도 3은 인코더의 예측기 도구를 포함하여 도 1의 인코더를 보다 상세히 도시한다.
도 4는 디코더의 예측기 도구를 포함하여 도 2의 디코더를 보다 상세히 도시한다.
도 5는 다른 예측 도구를 사용한, 도 1의 인코더의 변형을 상세히 도시한다.

실시예

도 1에서 MPEG-2 AAC 표준의 일반적 용어로 정의된 코딩 동작을 수행하는 인코더의 블록 다이어그램이 도시된다.
인코더로의 입력은 샘플링된 단상의 신호 x이고, 이 신호의 샘플 포인트들은 2N 포인트들의 시간 프레임들 또는 블록들로 그룹지어지며, 이것은 수학식 1과 같이,

수학식 1

$$\boldsymbol{x}_m=(x_m(0),x_m(1),\ldots,x_m(2N\text{-}1))^T$$

나타내지는데, 이때 m은 블록 인덱스이고 T는 역변환을 나타낸다. 샘플 포인트들

수학식 2

$$X_m=(X_m(0),X_m(1),\ldots,X_m(N-1))^T$$

서브-밴드들은 MPEG 표준으로 규정된다.
순방향 MDCT는 수학식 3과 같이,

수학식 3

$$X_m(k)=\sum_{i=0}^{2N-1} f(i)x_m(i)\cos\left(\frac{\pi}{4N}(2i+1+N)(2k+1)\right)$$

, k=0,...,N-1
나타내지는데, 이때 f(i)는 해석-합성 윈도우이고, 이것은 그것의 추가되고-겹쳐진 효과가 신호의 단위 이득을 발생하도록 대칭적인 윈도우이다.
주파수 서브-밴드 신호들 X(k)는 서브-밴드 신호들 각각의 장기간 중복을 제거하려고 시도하는 (이하에서 보다 상세히 설명될) 예측도구(2)에 번갈아 적용된다. 그 결과는 다음의 수학식 4와 같이

수학식 4

E_m (k)= (E_m (0), E_m (1),...,E_m (N-1)) ^ T

이들은 각각의 서브-밴드의 장기간 변화를 가리키는 주파수 서브-밴드 에러 신호들의 집합 및 각 프레임에 대한 순방향 적응 예측 계수들의 집합 A이다.
서브-밴드 에러 신호들 E(k)는 심리음향학적 모델에 의해 결정된 다수의 비트로 각 신호를 양자화하는 양자화기(3)로 제공된다. 이 모델은 제어기(4)에 의해 제공된다. 논의한 바와 같이, 심리 음향학적 모델은 인간의 청각 시스템의 차폐 반응을 모델링하는데 사용된다. 양자화된 에러 신호들 $\tilde{E}(k)$ 과 예측 계수들 A는 그리고나서 전송 채널(6)을 통한 전송을 위해 비트 스트림 멀티플렉서(5)에서 결합된다.
도 2는 도 1의 인코더로 코딩된 오디오 신호를 디코딩하기 위한 디코더의 일반적 구성을 도시하고 있다. 비트-스트림 디멀티플렉서(7)는 먼저 양자화 에러 신호 $\tilde{E}(k)$ 로부터 예측 계수들 A를 분리하고 그 에러 신호들을 별개의 서브-밴드 신호들로 분리한다. 예측 계수들 A 및 양자화된 에러 서브-밴드 신호들 $\tilde{E}(k)$ 는, 인코더에서 수행된 예측 과정을 역으로 하는 예측 도구(8)로 제공된다, 즉 예측 도구는 인코더에서 추출된 중복을 재삽입하여, 재구성된 양자화 서브-밴드 신호들 $\tilde{X}(k)$ 를 생성한다. 필터 뱅크 도구(9)는 그리고나서 수신된 버전 $\tilde{X}(k)$ 에 대해 역변환함으로써 시간 도메인 신호 $\tilde{x}$ 를 복원하며, 이는 다음의 수학식 5에 의해 나타내진다.

수학식 5

$$\tilde{x}_m(i)=\tilde{u}_{m-1}(i+N)+\tilde{u}_{m(i)}$$
$$i=0,\ldots,2N-1$$

이때, $\tilde{u}_k(i), i=0,\ldots,2N-1$ 는 $\tilde{X}$ 의 역변환

$$\tilde{u}_m(i)=f(i)\sum_{k=0}^{N-1}\tilde{X}_m(k)\cos\left(\frac{\pi}{4N}(2i+1+N)(2k+1)\right),$$
$$i=0,\ldots,2N-1$$

이고, 이것은 원래의 오디오 신호 x에 근접한다.
도 3은 도 1의 인코더의 예측 방법을 보다 상세히 도시한 것이다. 양자화된 주파수 서브-밴드 신호들의 집합 $\tilde{X}(k)$ 가, 신호 처리부(10)에 의해, 양자화된 주파수 서브-밴드 에러 신호들 E(k)를 사용하여 생성된다. 재구성된 양자화 서브-밴드 신호들 $\tilde{X}(k)$ 은 필터 뱅크(11)로 번갈아 제공되고, 필터 뱅크(11)는 역변형 이산 코사인 변환(inverse modification discrete cosine transform;IMDCT)을 그 신호들에 적용하여, 양자화된 시간 도메인 신호 $\tilde{x}$ 를 생성한다. 양자되된 시간 도메인 신호 $\tilde{x}$ 는 그리고나서 오디오 입력 신호 x를 수신하기도 하는 장기간 예측기 수단(12)으로 제공된다. 예측기 수단(12)은 장기간(long term;LT) 예측기를 사용하여, 이전에 양자화된 데이터를 기반으로, 현 프레임 m+1에 존재하는 오디오 신호의 중복 요소를 제거한다. 이 예측기의 전송 함수 P는 다음의 수학식 6과 같이,

수학식 6

$$P(z)=\sum_{k=-m_1}^{m_2} b_k z^{-(\alpha+k)}$$

로 나타내며, 이때 α(alpha)는 1 내지 1024 샘플들의 범위내 긴 딜레이를 나타내고 b_k는 예측 계수들을 나타낸다. $m_1 = m_2 = 0$일 때, 예측기는 한 개의 탭(tap)이고, $m_1 = m_2 = 1$일 때 예측기는 세 개의 탭이다.
패러미터 alpha와 b_k는 2N 샘플들의 기간에 대해 LT 예측한 후 평균 제곱된 에러를 최소화시킴으로써 결정된다. 한 개의 탭 예측기에 있어서, LT 예측 오차 r(i)는 수학식 7과 같이 주어진다.

수학식 7

$$r(i)=x(i)-b\tilde{x}(i-2N+1-\alpha)$$

여기서 x는 시간 도메인 오디오 신호이고 $\tilde{x}$는 시간 도메인 양자화 신호이다. 평균 제곱된 오차 R은 다음의 수학식 8과 같이 주어진다.

수학식 8

$$R=\sum_{i=0}^{2N-1} r^2(i)=\sum_{i=0}^{2N-1}(x(i)-b\tilde{x}(i-2N+1-\alpha))^2$$

$\partial R/\partial b=0$ 이라고 할 때, 수학식 9가 도출된다.

수학식 9

$$b=\frac{\sum_{i=0}^{2N-1} x(i)\tilde{x}(i-2N+1-\alpha)}{\sum_{i=0}^{2N-1}(\tilde{x}(i-2N-\alpha))^2}$$

b를 수학식 8에 대입하면, 다음의 수학식 10이 주어진다.

수학식 10

$$R=\sum_{i=0}^{2N-1} x^2(i)-\frac{(\sum_{i=0}^{2N-1} x(i)\tilde{x}(i-2N+1-\alpha))^2}{\sum_{i=0}^{2N-1}(\tilde{x}(n-2N+1-\alpha))^2}$$

R을 최소화하는 것은 수학식 10의 우변의 두 번째 항을 최대화시킨다는 것을 의미한다. 이 항은 그것의 명시된 범위를 통해 모든 가능한 alpha의 값들에 대해 계산되어지고, 이 항을 최대화시키는 alpha의 값이 선택된다. OMEGA(Ω)로서 나타낼 수 있는 수학식 10의 분모의 에너지는, 에너지를 새로 재산출하는 대신, 다음의 수학식 11을 사용하여 딜레이 alpha-1 에서 alpha까지 쉽게 갱신될 수 있다.

수학식 11

$$\Omega_\alpha=\Omega_{\alpha-1}+\tilde{x}^2(-\alpha)-\tilde{x}^2(-\alpha+N)$$

하나의 탭 LT 예측기가 사용되면, 수학식 9는 예측 계수 b_j를 산출하는데 사용된다. j개의 탭 예측기에 있어서, LT 예측 딜레이 alpha는 수학식 10의 두 번째 항을 최대화시킴으로써 먼저 정해지고, 그리고나서 j×j 개의 방정식들의 집합을 풀어 j 예측 계수들을 산출한다.
LT 예측 패러미터들 A는 딜레이 alpha와 예측 계수 b_j이다. 딜레이는 사용된 범위에 따라 9 내지 11 비트로 양자화된다. 1 내지 1024 범위에 있는 1024개의 가능한 값들에 가장 일반적으로 10 비트가 활용된다. 비트 수를 감소시키기 위해, LT 예측 딜레이들은 5비트의 프레임에서 조차 델타 코딩될 수 있다. 실험에는 3 내지 6 비트를 갖는 이득을

양자화하는 것이 충분하다는 것을 보인다. 균일하지 않은 이득의 분포로 인해, 비균일한 양자화가 사용되어져야 한다. 상술한 방법에서, LT 합성 필터 1/P(z)의 안정성이 항상 보장되는 것은 아니다. 1-탭 예측기에 있어서, 안정성 조건은 $|b| \leq 1$ 이다. 따라서, 안정화는 $|b|>1$ 일 때마다 $|b|$ =1로 정함으로써 쉽게 달성될 수 있다. 3-탭 예측기에 대해서, 1987년 7월 IEEE Trans. ASSP, 35권, 7번에 수록된 알 피 라마샨드란과 피 케이발의 '음성 코더의 피치 필터의 안정성 및 성능 분석'에 설명된 바와 같은 다른 안정화 과정이 사용될 수 있다. 그러나, LT 합성 필터의 불안정성은 재구성된 신호의 품질에 그렇게 해롭지는 않다. 불안정한 필터는 (에너지를 증가시키는) 많은 프레임들을 고수할 것이지만, 결국 출력이 시간과 함께 계속 증가하지 않도록 하기 위해 안정성의 기간과 만나게 된다.
LT 예측기 계수들이 정해진 후, (m+1)번째 프레임에 대해 예측된 신호가 수학식 12와 같이 결정될 수 있다.

수학식 12

$$\hat{x}(i)=\sum_{j=-m_1}^{m_2} b_j \tilde{x}(i-2N+1-j-\alpha), pha+N)$$
$$i=mN+1, mN+2, \ldots, (m+1)N$$

그리고나서, 예측된 시간 도메인 신호 $\hat{x}$ 는, 필터 뱅크(13)으로 제공되고, 필터 뱅크(13)는 신호에 MDCT를 적용하여, (m+1)번째 프레임에 대한 예측된 스펙트럼의 계수 $\hat{X}_{m+1}(k)$ 를 생성한다. 그리고나서, 예측된 스펙트럼의 계수들인 $\hat{X}(k)$ 는, 감산기(14)에서 스펙트럼의 계수들인 X(k)로부터 감산된다.
예측이 코딩 이득(coding gain)을 초래했을 경우로 제한하고, 그 예측이 이용되는 것을 보증하기 위해서는, 적합한 예측기 제어가 필요함과 동시에 소량의 예측기 제어 정보를 디코더로 전송하지 않으면 안된다. 이 기능은 감산기(14)에서 수행된다. 예측기 제어 방식은 MPEG-2 Advanced Audio Coding (AAC)에 사용되고 있는 역방향 적응 예측기 제어 방식과 동일하다. 부정보로서 전송되는 각 프레임에 대한 예측기 제어 정보는, 다음의 두 단계로 결정된다. 먼저, 각 눈금계수(scalefactor) 밴드에 대해 예측이 코딩 이득을 낼 것인지 아닌지를 결정하고, 만일 예측이 코딩 이득을 가져온다면, 그 눈금계수(scalefactor) 밴드에 대한 **사용_예측치 (predictor_used)** 비트는 1로 정해진다. 다음, 모든 눈금계수 밴드들에 대해 상기 동작이 완료된 후, 이 프레임의 예측기에 의해 전체 코딩 이득이 예측 부정보에 대해 최소한 추가 비트 필요성을 보상하는지의 여부를 결정한다. 만약 전체 코딩 이득이 추가 비트 필요성을 보상한다면, **현재_예측치_데이터 (predictor_data_present)** 비트는 1로 정해지고 그 예측기 리셋을 위해 필요로되는 것을 포함하는 완전한 부정보가 전송되어 예측 에러 값이 양자화기(3)로 주어진다. 만일, 전체 코딩 이득이 예측 부정보에 대해 최소한 추가 비트 필요성을 보상하지 않는다면, **현재_예측치_데이터(predictor_data_present)** 비트는 0으로 정해지고 **사용_예측치(predictor_used)** 비트들은 모두 0으로 리셋되며 양자화기(3)로의 송신은 되지 않는다. 이런 경우, 스펙트럼의 성분 값이 양자화기(3)로 주어진다. 상술한 바와 같이, 예측기 제어는 먼저 한 눈금계수 밴드의 모든 예측기들에 대해 동작하고, 그리고나서 모든 눈금 계수 밴드들에 대해 제2단계를 따른다.
LT 예측의 목표가 가장 큰 전체 예측 게인을 얻는 것이라는 것이 명백할 것이다. G_i 가 l번째 주파수 서브-밴드의 예측 이득을 나타내도록 하자. 주어진 프레 임의 전체 예측 이득은 다음의 수학식 13과 같이 산출될 수 있다.

수학식 13

$$G=\sum_{l=1\&(G_1>0)}^{N} G_l$$

만약 이득이 예측기 부정보에 대한 추가의 비트 필요성을 보상한다면, 즉 G>T (dB)이면, 완전한 부정보가 전송되고 양의 이득을 발생하는 예측기들이 스위치 온된다. 그렇지 않으면, 예측기들은 사용되지 않는다.
상술한 방법에 의해 획득된 LP 패러미터들은 이득을 최대화하는 것과 직접적으로 관련되지 않는다. 그러나, 각 블록 및 선택된 범위(이 예에서 1 내지 1024)내의 각 딜레이에 대한 이득을 산출하고, 가장큰 전체 예측 이득을 발생하는 딜레이를 선택함으로써, 예측 과정이 최적화된다. 선택된 딜레이 alpha 및 상응하는 계수들 b가 양자화된 에러 서브-밴드 신호들과 함께 부정보로서 전송된다. 인코더에서 계산의 복잡성이 증가하는 한편, 디코더에서는 복잡성의 증가가 전혀 초래되지 않는다.
도 4는 도 2의 디코더를 보다 상세히 도시한다. 부호화된 오디오 신호가 상술한 바와 같은 비트 스트림 디멀티플렉서(7)에 의해 전송 채널(6)로부터 수신된다. 비트스트림 디멀티플렉서(7)는 예측 계수들 A와 양자화된 에러 신호들 $\hat{E}(k)$ 를 분리하고 이들을 예측 수단(8)으로 공급한다. 이 수단은 주파수 도메인의 양자화된 에러 신호들 $\hat{E}(k)$ 과 예측된 오디오 신호 $\hat{X}(k)$ 를 결합하여 역시 주파수 도메인의 재구성된 오디오 신호 $\tilde{X}(k)$ 를 생성하는 결합기(24)를 구비한다. 필터 뱅크(9)는 재구성된 신호 $\tilde{X}(k)$ 를 주파수 도메인에서 타임 도메인으로 변환하여 재구성된 시간 도메인 오디오 신호 $\tilde{x}$ 를 생성한다. 이 신호는 예측 계수들 A도 또한 수신하는 장기간 예측 수단(26)으로 번갈아 피드백

된다. 장기간 예측 수단(26)은 현재 프레임에 대한 예측 계수들을 사용하여 이전의 재구성된 시간 프레임들로부터 예측된 현 시간 프레임을 생성시킨다. 필터 뱅크(25)는 예측된 신호 $\hat{x}$ 를 변환한다.
인코더로부터 전송된 예측기 제어 정보는 디코더에서 사용되어 디코딩 동작을 제어할 수 있다는 것을 이해할 수 있을 것이다. 특히, **사용된_예측기(predictor_use)** 비트는 결합기(24)에서 사용되어 예측이 어떤 주어진 주파수 대역에서 채용되었는지 아닌지를 정할 수 있다.
도 5에서는 도 1의 오디오 신호 인코더에 대한 다른 구성이 도시되어 있으며, 여기서 코딩될 오디오 신호 x는 비교기(15)에 의해 시간 도메인의 예측된 신호 $\hat{x}$ 와 비교되어 주파수 서브-밴드 에러 신호들의 집합 E(k)를 생성한다. 이들 신호들은 양자화기(17)에 의해 양자화되어 양자화된 에러 신호들의 집합 $\hat{E}(k)$ 를 생성한다.

그러면 제2필터 뱅크(18)는 양자화된 에러 신호들 $\hat{E}(k)$ 을 신호 $\tilde{e}$ 의 결과를 낳는 시간 도메인으로 전환하는데 사용된다. 이 시간 도메인의 양자화된 에러 신호 $\tilde{e}$ 는 그리고나서 신호 처리부(19)에서 예측된 시간 도메인 오디오 신호 $\hat{x}$ 와 결합되어 양자화된 오디오 신호 $\tilde{x}$ 를 생성한다. 신호 처리부 즉, 예측 수단(19)은 도 3의 인코더에서의 수단(120과 동일한 동작을 수행하여, 예측된 오디오 신호 $\hat{x}$ 와 예측 계수들 A를 생성한다. 예측 계수들 및 양자화된 에러 신호들은 전송 채널(22)을 통한 전송을 위해 비트 스트림 멀티플렉서(21)에서 결합된다. 상술한 바와 같이, 에러 신호들은 제어부(23)에 의한 심리 음향학적 모델에 따라 양자화된다.
상술한 오디오 코딩 알고리즘은 낮은 비트 레이트에서 오디오 신호 압축을 허용한다. 그 기술은 장기간(LT) 예측을 기반으로 한다. 종래의 역방향 적응 예측 기술과 비교하여, 여기서 설명된 기술은 단일 악기 음악 신호 및 스피치 신호에 대해 보다 높은 예측 이득을 가져다주는 한편 단지 낮은 계산적 복잡성을 필요로 한다.

(57) 청구의 범위

청구항 1.

코딩될 오디오 신호 x를 수신하는 단계;
수신된 오디오 신호의 시간 프레임으로부터 주파수 서브-밴드(sub-band)를 생성하는 단계(1);
수신된 오디오 신호 x로부터 양자화된 오디오 신호 $\tilde{x}$ 를 생성하는 단계(9, 11);
장기간(long-term) 예측 계수들의 집합 A를 생성하는 단계(2);
상기 양자화된 오디오 신호 $\tilde{x}$ 의 최소한 한 개의 이전 시간 프레임으로부터 직접적으로 시간 프레임의 각 서브-밴드들을 위한 장기간 예측 계수들의 집합 A를 이용함으로써 수신된 오디오 신호의 현재 시간 프레임을 예측하는 단계(8);
상기 장기간 예측 계수들의 집합 A를 이용하여, 양자화된 오디오 신호 $\tilde{x}$ 의 예측된 오디오 신호 $\hat{x}$ 를 생성하는 단계(12);
상기 수신된 오디오 신호 x와 상기 예측된 오디오 신호 $\hat{x}$ 를 비교하여 복수개의 주파수 서브-밴드들 각각에 대한 에러 신호 E(k)를 생성하는 단계;(14)
에러 신호들 E(k)를 양자화하여 양자화된 에러 신호 $\hat{E}(k)$ 를 생성하는 단계(3); 및
양자화 에러 신호들 $\hat{E}(k)$ 과 예측 계수들 A를 결합하여 코딩된 오디오 신호를 생성하는 단계(5)를 구비함을 특징으로 하는 오디오 신호를 코딩하는 방법.

청구항 2.

제 1항에 있어서,
프레임들 x_m 의 상기 수신된 오디오 신호 x를 시간 도메인에서 주파수 도메인으로 변환하여 주파수 서브-밴드(sub-band) 신호들의 집합 X(k)을 제공하는 단계(1)와, 예측 오디오 신호 $\hat{x}$ 를 시간 도메인에서 주파수 도메인으로 변환하여 예측 주파수 서브-밴드 신호들의 집합 $\hat{X}(k)$ 을 생성하는 단계(13), 여기서 수신된 오디오 신호 x와 예측 오디오 신호 $\hat{x}$ 사이의 비교가 주파수 도메인에서 수행될 때, 해당 서브-밴드 신호들을 서로에 대해 비교함으로써 주파수 서브-밴드 에러 신호들 E(k)을 생성하는 단계(14)를 구비함을 특징으로 하는 오디오 신호를 코딩하는 방법.

청구항 3.

제 1항에 있어서,
시간 도메인의, 수신된 오디오 신호 x와 예측 오디오 신호 $\hat{x}$ 사이의 비교를 수행하여 역시 시간 도메인의 에러 신호 e를 생성하는 단계(15) 및 상기 에러 신호 e를 시간 도메인에서 주파수 도메인으로 전환하여 상기 다수의 주파수 서브-밴드 에러 신호들 E(k)을 생성하는 단계(16)를 구비함을 특징으로 하는 오디오 신호를 코딩하는 방법.

청구항 4.

오디오 신호의 다수의 주파수 서브-밴드들 각각에 대해 양자화된 에러 신호 $\tilde{E}(k)$ 및 재구성된 양자화된 오디오 신호 $\tilde{x}$ 의 최소한 한 개의 이전 시간 프레임으로부터 직접적으로, 수신된 오디오 신호의 현 시간 프레임 x_m 의 각 주파수 서브-밴드를 예측하는데 사용될 수 있는 예측 계수들의 집합 A를 구비한, 코딩된 오디오 신호를 수신하는 단계;

상기 재구성된 양자화된 오디오 신호 $\tilde{x}$ 를 양자화된 에러 신호들 $\tilde{E}(k)$ 로부터 생성하는 단계(24, 25, 9);

예측 계수들 A와 양자화된 오디오 신호 $\tilde{x}$ 를 사용하여 예측된 오디오 신호 $\hat{x}$ 를 생성하는 단계(26);

예측된 오디오 신호 $\hat{x}$ 를 시간 도메인에서 주파수 도메인으로 변환하여, 재구성된 주파수 서브-밴드 신호들의 집합 $\tilde{X}(k)$ 을 생성(24)하도록 양자화된 에러 신호들 $\tilde{E}(k)$ 과 결합하기 위한, 예측 주파수 서브-밴드 신호들의 집합 $\hat{X}(k)$ 을 생성하는 단계(25); 및

상기 재구성된 주파수 서브-밴드 신호들 $\tilde{X}(k)$ 에 대해 주파수에서 시간 도메인으로의 변환을 수행하여 재구성된 양자화된 오디오 신호 $\tilde{x}$ 를 생성하는 단계(9)를 포함함을 특징으로 하는 코딩된 오디오 신호를 디코딩하는 방법.

청구항 5.

코딩될 오디오 신호 x를 수신하는 입력부;

수신된 오디오 신호의 타임 프레임으로부터 주파수 서브-밴드들을 생성하는 제1생성수단;

수신된 오디오 신호 x로부터 양자화된 오디오 신호 $\tilde{x}$ 를 생성하기 위해 상기 입력부와 결합된 처리 수단(2, 3; 15 내지 19);

양자화된 오디오 신호 $\tilde{x}$ 의 최소한 한 개의 이전 시간 프레임으로부터 직접적으로, 수신된 오디오 신호 x 의 현재 시간 프레임 x_m 을 예측하기 위한 시간 프레임의 각 서브-밴드들에 대해 사용될 장기간 예측 계수들의 집합 A를 생성하도록 상기 처리 수단(3)에 결합된 예측 수단(12;19);

장기간 예측 계수들의 집합 A 및 양자화된 오디오 신호 $\tilde{x}$ 를 사용하여 예측 오디오 신호 $\hat{x}$ 를 생성하고, 수신된 오디오 신호 x를 예측 오디오 신호 $\hat{x}$ 와 비교하여 다수의 주파수 서브-밴드들 각각에 대한 에러 신호 E(k)를 생성하는 제2생성수단(10 내지 14; 20, 15);

에러 신호들 E(k)을 양자화하여 양자화 에러 신호들의 집합 $\tilde{E}(k)$ 을 생성하는 양자화 수단(3; 17); 및

양자화된 에러 신호들 $\tilde{E}(k)$ 을 예측 계수들 A과 결합하여 코딩된 오디오 신호를 생성하는 결합 수단(5; 21)을 포함함을 특징으로 하는 오디오 신호 코딩 장치.

청구항 6.

제 5항에 있어서,

상기 제2생성수단은 수신된 오디오 신호 x를 시간에서 주파수 도메인으로 변환하는 제1변환수단(1)과 상기 예측 오디오 신호 $\hat{x}$ 를 시간에서 주파수 도메인으로 변환하는 제2변환 수단(13) 및 상기 결과의 주파수 도메인 신호들을 주파수 도메인에서 비교하도록 정해진 비교 수단(14)을 구비함을 특징으로 하는 오디오 신호 코딩 장치.

청구항 7.

제 5항에 있어서,

상기 제2생성수단은 시간 도메인의 수신된 오디오 신호 x와 예측 오디오 신호 $\hat{x}$ 를 비교하도록 정해짐을 특징으로 하는 오디오 신호 코딩 장치.

청구항 8.

코딩된 오디오 신호가 오디오 신호의 여러 주파수 서브-밴드들 각각에 대해 양자화된 에러 신호 $\tilde{E}(k)$ 및 오디오 신호의 시간 프레임 중 각 주파수 서브-밴드들에 대해 사용될 예측 계수들의 집합 A를 구비하고, 이때 예측 계수들 A가 , 재구성된 양자화 오디오 신호 $\tilde{x}$ 의 최소한 한 개의 이전 시간 프레임으로부터 직접적으로, 수신된 오디오 신호의 현시간 프레임 x_m 을 예측하는데 사용될 수 있을 때, 코딩된 오디오 신호를 디코딩하는 장치에 있어서,

코딩된 오디오 신호를 수신하는 입력부;

양자화된 에러 신호들 $\tilde{E}(k)$ 로부터 상기 재구성된 양자화된 오디오 신호 $\tilde{x}$ 를 생성하는 생성 수단(24, 25, 9); 및

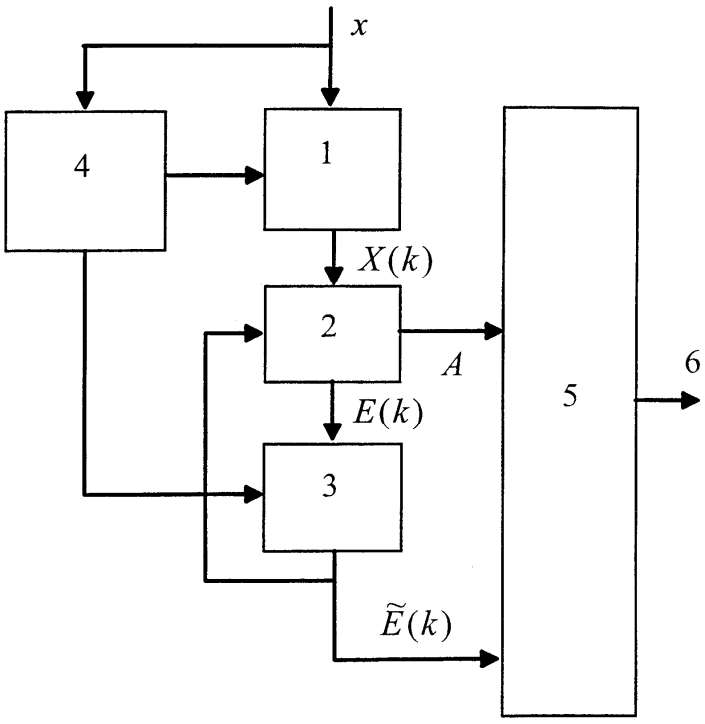
예측 계수들 A와 상기 재구성된 오디오 신호 $\tilde{x}$ 로부터 예측된 오디오 신호 $\hat{x}$ 를 생성하는 신호 처리 수단(26)을 구비하고,

이때 상기 생성수단은 상기 예측된 오디오 신호 $\hat{x}$ 를 시간 도메인에서 주파수 도메인으로 변환하여 예측된 주파수 서브-밴드 신호들의 집합 $\hat{X}(k)$ 을 생성하는 제1변환수단(25), 상기 예측된 주파수 서브-밴드 신호들의 집합 $\hat{X}(k)$ 을 양자화된 에러 신호들 $\tilde{E}(k)$ 과 결합하여 재구성된 주파수 서브-밴드 신호들의 집합 $\tilde{X}(k)$ 을 생성하는 결합 수단(24) 및 재구성된 주파수 서브-밴드 신호들 $\tilde{X}(k)$ 에 대해 주파수에서 시간 도메인으로의 변환을 수행하여 재구성된

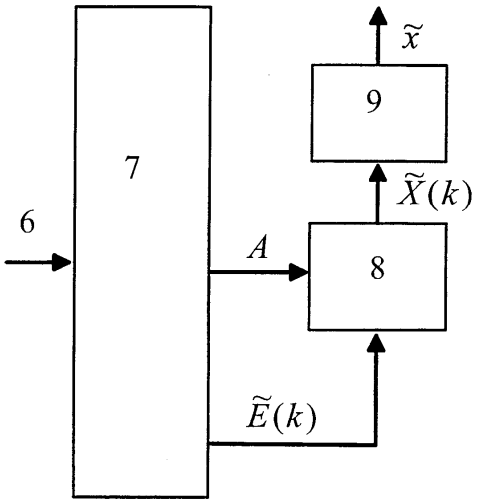
양자화된 오디오 신호 $\widetilde{x}$ 를 생성하는 제2변환수단을 포함함을 특징으로 하는 디코딩 장치.

도면

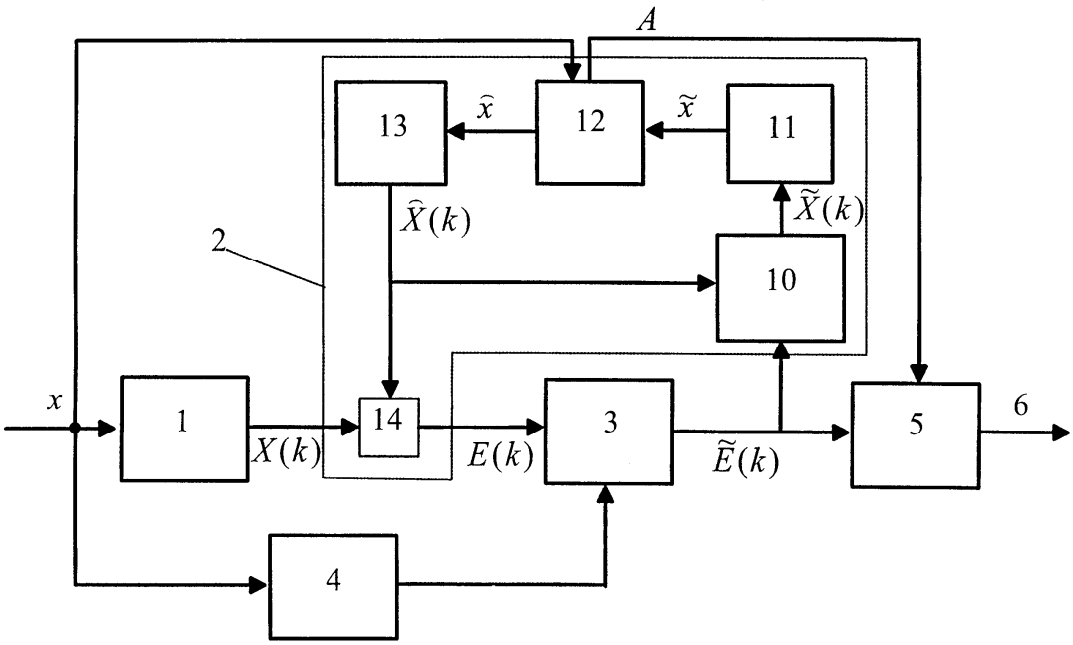
도면1



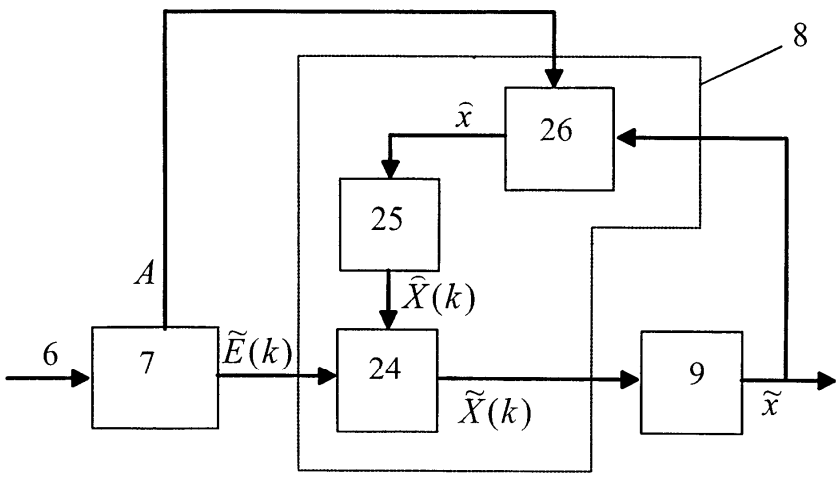
도면2



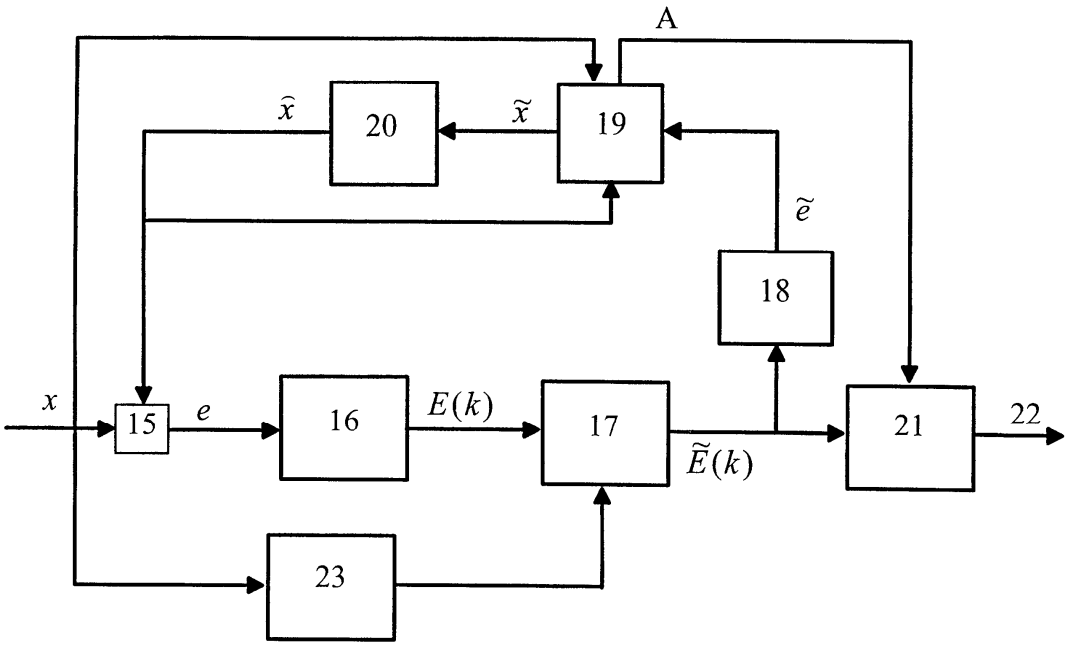
도면3


A
13
12
11
$\hat{x}$
$\tilde{x}$
$\hat{X}(k)$
$\tilde{X}(k)$
2
10
x
1
14
3
5
6
$X(k)$
$E(k)$
$\tilde{E}(k)$
4

도면4


8
26
$\hat{x}$
25
A
$\hat{X}(k)$
6
7
$\tilde{E}(k)$
24
9
$\tilde{X}(k)$
$\tilde{x}$

도면5


A
$\hat{x}$
20
$\tilde{x}$
19
$\tilde{e}$
18
x
15
e
16
$E(k)$
17
21
22
$\tilde{E}(k)$
23