



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0028037
(43) 공개일자 2020년03월13일

(51) 국제특허분류(Int. Cl.)
G10L 21/034 (2013.01) G10L 25/18 (2013.01)
G10L 25/45 (2013.01) H03G 7/00 (2006.01)
H04B 1/66 (2006.01)
(52) CPC특허분류
G10L 21/034 (2013.01)
G10L 25/18 (2013.01)
(21) 출원번호 10-2020-7006389(분할)
(22) 출원일자(국제) 2014년04월01일
심사청구일자 2020년03월04일
(62) 원출원 특허 10-2016-7015588
원출원일자(국제) 2014년04월01일
심사청구일자 2019년03월21일
(85) 번역문제출일자 2020년03월04일
(86) 국제출원번호 PCT/US2014/032578
(87) 국제공개번호 WO 2014/165543
국제공개일자 2014년10월09일
(30) 우선권주장
61/809,028 2013년04월05일 미국(US)
61/877,167 2013년09월12일 미국(US)

(71) 출원인
돌비 레버러토리즈 라이젠싱 코오폰레이션
미합중국, 캘리포니아 94103, 샌프란시스코, 마켓
스트리트 1275
돌비 인터네셔널 에이비
네덜란드 1101 씨엔 암스트레담 주이두스트 헤리
커베르그백 1-35 3이 아폴로 빌딩
(72) 발명자
헤델린, 페르
스웨덴 에스-113 30 스톡홀름 가블레가탄 12에이
돌비 스웨덴 에이비 내
비스와스, 아리짓
독일 90429 뉘른베르크 도이치헤른스트라쎄 15-19
돌비 저머니 게엠베하 내
(뒷면에 계속)
(74) 대리인
장훈

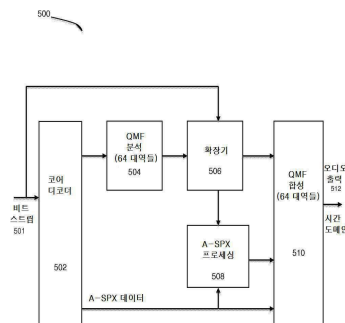
전체 청구항 수 : 총 3 항

(54) 발명의 명칭 향상된 스펙트럼 확장을 사용하여 양자화 잡음을 감소시키기 위한 압신 장치 및 방법

(57) 요약

실시예들은 오디오 코덱에서 코딩 잡음을 감소시키기 위한 압신 방법 및 시스템에 관한 것이다. 압축 프로세스는 정의된 윈도우 형태를 사용하여 초기 오디오 신호를 복수의 세그먼트들로 분할하고, 상기 초기 오디오 신호의 주파수 도메인 샘플들의 비-에너지 기반 평균을 사용하여 주파수 도메인에서 광대역 이득을 산출하고, 비교적 낮은 강도의 세그먼트들을 증폭시키며 비교적 높은 강도의 세그먼트들을 감쇠시키기 위해 개개의 이득 값들을 적용하는 압축 프로세스를 통해 초기 오디오 신호의 원래 동적 범위를 감소시킨다. 상기 압축된 오디오 신호는 그 후 비교적 높은 강도의 세그먼트들을 증폭시키며 비교적 낮은 강도의 세그먼트들을 감쇠시키기 위해 역 이득 값들을 적용하는 실질적으로 원래 동적 범위로 다시 확대된다. QMF 필터뱅크는 주파수 도메인 표현을 획득하도록 초기 오디오 신호를 분석하기 위해 사용된다.

대표도



(52) CPC특허분류

G10L 25/45 (2013.01)

H03G 7/007 (2013.01)

H04B 1/66 (2013.01)

(72) 발명자

슈그, 미하엘

독일 90429 뉘른베르크 도이치헤른스트라쎄 15-19

돌비 저머니 게엠베하 내

멜코트, 비나이

미국 캘리포니아 94103-4813 샌프란시스코 포트레
로 애비뉴 100 돌비 레버러토리즈 인코포레이티드
내

명세서

청구범위

청구항 1

다수의 채널들을 포함하는 오디오 신호를 압축하는 방법에 있어서:

상기 오디오 신호의 시간-주파수 타일 표현을 수신하는 단계로서, 상기 오디오 신호의 상기 시간-주파수 타일 표현은 상기 오디오 신호의 시간 슬롯들로의 분할을 포함하고, 각각의 시간 슬롯은 주파수 서브대역들로 분할되고, 상기 주파수 대역들은 동일하게 이격된, 상기 오디오 신호의 시간-주파수 타일 표현을 수신하는 단계; 및

상기 오디오 신호의 동적 범위를 감소시키기 위해 상기 오디오 신호의 상기 시간-주파수 타일 표현을 압축하는 단계를 포함하고;

상기 오디오 신호의 상기 시간-주파수 타일 표현을 압축하는 단계는:

그룹핑 정보에 기초하여 상기 오디오 신호의 채널들을 채널들의 별개의 서브세트들로 분할하는 단계; 및

채널들의 각각의 별개의 서브세트에 대해:

상기 오디오 신호의 상기 시간-주파수 타일 표현의 시간 슬롯에 대해 공유된 이득을 산출하는 단계와,

상기 시간 슬롯에 대한 상기 공유된 이득을 채널들의 별개의 서브세트의 각각의 채널의 각각의 주파수 서브대역에 적용하는 단계를 포함하는, 다수의 채널들을 포함하는 오디오 신호를 압축하는 방법.

청구항 2

다수의 채널들을 포함하는 오디오 신호를 압축하기 위한 장치에 있어서:

상기 오디오 신호의 시간-주파수 타일 표현을 수신하는 제 1 인터페이스로서, 상기 오디오 신호의 상기 시간-주파수 타일 표현은 상기 오디오 신호의 시간 슬롯들로의 분할을 포함하고, 각각의 시간 슬롯은 주파수 서브대역들로 분할되고, 상기 주파수 서브대역들은 동일하게 이격된, 상기 제 1 인터페이스; 및

상기 오디오 신호의 동적 범위를 감소시키기 위해 상기 오디오 신호의 상기 시간-주파수 타일 표현을 압축하는 압축기를 포함하고;

상기 오디오 신호의 상기 시간-주파수 타일 표현을 압축하는 것은:

그룹핑 정보에 기초하여 상기 오디오 신호의 채널들을 채널들의 별개의 서브세트들로 분할하는 것;

채널들의 각각의 별개의 서브세트에 대해:

상기 오디오 신호의 상기 시간-주파수 타일 표현의 시간 슬롯에 대해 공유된 이득을 산출하는 것과;

상기 시간 슬롯에 대한 상기 공유된 이득을 채널들의 별개의 서브세트의 각각의 채널의 각각의 주파수 서브대역에 적용하는 것을 포함하는, 다수의 채널들을 포함하는 오디오 신호를 압축하기 위한 장치.

청구항 3

하나 이상의 프로세서들에 의해 실행될 때, 제 1 항의 방법을 수행하는 지시들을 포함하는 비-일시적 컴퓨터 판독가능한 매체.

발명의 설명

기술 분야

[0001] 관련 출원들에 대한 상호-참조

[0002] 본 출원은 2013년 4월 5일에 출원된 미국 가 특허 출원 번호 제 61/809,028 호 및 2013년 9월 12일에 출원된 미

국 가 특허 출원 번호 제 61/877,167 호에 대한 우선권을 주장하고, 그것은 여기에 전체적으로 참조로서 통합된다.

[0003] 하나 이상의 실시예들은 일반적으로 오디오 신호 프로세싱에 관한 것이며, 보다 구체적으로 압축/확장(압신) 기술들을 사용하여 오디오 코덱들에서 코딩 잡음을 감소시키는 것에 관한 것이다.

배경 기술

[0004] 많은 인기 있는 디지털 사운드 포맷들은 저장 또는 데이터 레이트 조건들을 감소시키기 위해 데이터의 일부를 폐기하는 손실 데이터 압축 기술들을 이용한다. 손실 데이터 압축의 적용은 소스 콘텐츠(예로서, 오디오 콘텐츠)의 충실도를 감소시킬 뿐만 아니라, 그것은 또한 압축 아티팩트들의 형태로 뚜렷한 왜곡을 도입할 수 있다. 오디오 코딩 시스템들의 콘텍스트에서, 이들 사운드 아티팩트들은 코딩 잡음 또는 양자화 잡음으로 불리운다.

[0005] 디지털 오디오 시스템들은 정의된 오디오 파일 포맷 또는 스트리밍 미디어 오디오 포맷에 따라 오디오 데이터를 압축 및 압축 해제하기 위해 코덱들(코더-디코더 구성요소들)을 이용한다. 코덱들은 가능한 높은 충실도를 유지하면서 최소 수의 비트들을 가진 오디오 신호를 표현하려고 시도하는 알고리즘들을 구현한다. 통상적으로 오디오 코덱들에서 사용된 손실 압축 기술들은 인간 청각 지각의 음향 심리학적 모델 상에서 작용한다. 오디오 포맷들은 보통 시간/주파수 도메인 변환(예로서, 변경된 이산 코사인 변환 - MDCT)의 사용을 수반하고, 임의의 결보기 양자화 잡음을 포함하는 특정한 사운드들이 실제 콘텐츠에 의해 은닉되거나 마스킹되도록 주파수 마스킹 또는 시간 마스킹과 같은 마스킹 효과들을 사용한다.

[0006] 대부분의 오디오 코딩 시스템들은 프레임 기반이다. 프레임 내에서, 오디오 코덱들은 보통 그것이 적어도 들을 수 있게 되도록 주파수 도메인에서 코딩 잡음을 성형한다. 여러 개의 현재 디지털 오디오 포맷들은 프레임이 여러 개의 상이한 레벨들 또는 강도들의 사운드들을 포함할 수 있는 이러한 긴 지속 기간들의 프레임들을 이용한다. 코딩 잡음이 보통 프레임의 전진에 걸쳐 레벨이 변하지 않기 때문에, 코딩 잡음은 프레임의 낮은 강도 부분들 동안 가장 잘 들릴 수 있다. 이러한 효과는 고 강도 세그먼트 이전의 침묵(또는 저-레벨 신호)이 디코딩된 오디오 신호에서의 잡음에 의해 압도되는 전-에코 왜곡으로서 나타내어질 수 있다. 이러한 효과는 캐스터네츠 또는 다른 뚜렷한 퍼커시브 사운드 소스들과 같은, 타악기들로부터의 과도 사운드들 또는 임펄스들에서 가장 현저할 수 있다. 이러한 왜곡은 통상적으로 시간 도메인에서의 코덱의 전체 변환 윈도우에 걸쳐 확산되는 주파수 도메인에 도입된 양자화 잡음에 의해 야기된다.

[0007] 전-에코 아티팩트들을 회피하거나 최소화하기 위한 현재의 조치들은 필터들의 사용을 포함한다. 그러나, 이러한 필터들은 위상 왜곡 및 시간적 스미어링을 도입한다. 또 다른 가능한 해결책은 보다 작은 변환 윈도우들의 사용을 포함하지만, 이러한 접근법은 주파수 분해능을 상당히 감소시킬 수 있다.

[0008] 배경 섹션에 논의된 주제는 단지 배경 섹션에서 그것의 언급의 결과로서 종래 기술인 것으로 가정되지 않아야 한다. 유사하게, 배경 섹션에서 언급되거나 배경 섹션의 주제와 연관된 문제점은 이전에 종래 기술에서 인지되어 온 것으로 가정되지 않아야 한다. 배경 섹션에서의 주제는 단지 상이한 접근법들만을 나타내며, 이것은 본질적으로 및 제 스스로 또한 발명될 수 있다.

발명의 내용

해결하려는 과제

[0009] 본 발명은 압축/확장(압신) 기술들을 사용하여 오디오 코덱들에서 코딩 잡음을 감소시키는 것에 관한 것이다.

과제의 해결 수단

[0010] 실시예들은 정의된 윈도우 형태를 사용하여 수신된 오디오 신호를 복수의 시간 세그먼트들로 분할하는 것, 상기 오디오 신호의 주파수 도메인 표현의 비-에너지 기반 평균을 사용하여 상기 주파수 도메인에서 각각의 시간 세그먼트에 대한 광대역 이득을 산출하는 것, 및 상기 확장된 오디오 신호를 획득하기 위해 각각의 시간 세그먼트에 상기 이득 값을 적용하는 것을 포함하는 프로세스를 통해 상기 오디오 신호를 확장된 동적 범위로 확장시킴으로써 상기 수신된 오디오 신호를 프로세싱하는 방법에 관한 것이다. 각각의 시간 세그먼트에 적용된 광대역 이득의 이득 값들은 비교적 높은 강도의 세그먼트들을 증폭시키며 비교적 낮은 강도의 세그먼트들을 감소시키는 효과를 갖도록 선택된다. 이러한 방법을 위해, 상기 수신된 오디오 신호는 정의된 윈도우 형태를 사용하여 원래

오디오 신호를 복수의 시간 세그먼트들로 분할하는 것, 초기 오디오 신호의 주파수 도메인 샘플들의 비-에너지 기반 평균을 사용하여 상기 주파수 도메인에서 광대역 이득을 산출하는 것, 및 상기 광대역 이득을 상기 원래 오디오 신호에 적용하는 것을 포함하는 압축 프로세스를 통해 원래 동적 범위로부터 압축된 원래 오디오 신호를 포함한다. 상기 압축 프로세스에서, 각각의 시간 세그먼트에 적용된 상기 광대역 이득의 이득 값들은 비교적 낮은 강도의 세그먼트들을 증폭시키며 비교적 높은 강도의 세그먼트들을 감쇠시키는 효과를 갖도록 선택된다. 상기 확장 프로세스는 상기 초기 오디오 신호의 동적 범위를 실질적으로 복원하도록 구성되고, 상기 확장 프로세스의 광대역 이득은 실질적으로 상기 압축 프로세스의 광대역 이득의 역일 수 있다.

[0011] 확장 프로세스에 의해 수신된 오디오 신호를 프로세싱하는 방법을 구현하는 시스템에서, 필터뱅크 구성요소는 그것의 주파수 도메인 표현을 획득하기 위해 상기 오디오 신호를 분석하기 위해 사용될 수 있으며, 복수의 시간 세그먼트들로의 분할을 위한 정의된 윈도우 형태는 필터뱅크에 대한 프로토타입 필터와 동일할 수 있다. 마찬가지로, 압축 프로세스에 의해 수신된 오디오 신호를 프로세싱하는 방법을 구현하는 시스템에서, 필터뱅크 구성요소는 그것의 주파수 도메인 표현을 획득하기 위해 원래 오디오 신호를 분석하기 위해 사용될 수 있으며, 복수의 시간 세그먼트들로의 분할을 위한 정의된 윈도우 형태는 필터뱅크에 대한 프로토타입 필터와 동일할 수 있다. 어느 경우에도 상기 필터뱅크는 QMF 뱅크 또는 단-시간 푸리에 변환 중 하나일 수 있다. 이 시스템에서, 확장 프로세스를 위한 수신된 신호는 비트스트림을 생성하는 오디오 인코더 및 상기 비트스트림을 디코딩하는 디코더에 의해 압축된 신호의 변경 후 획득된다. 상기 인코더 및 디코더는 변환-기반 오디오 코덱의 적어도 일부를 포함할 수 있다. 상기 시스템은 상기 비트스트림을 통해 수신되는 제어 정보를 프로세싱하고 상기 확장 프로세스의 활성화 상태를 결정하는 구성요소들을 추가로 포함할 수 있다.

[0012] 다음의 도면들에서 유사한 참조 부호들은 유사한 요소들을 나타내기 위해 사용된다. 다음의 도면들은 다양한 예들을 묘사하지만, 하나 이상의 구현들은 도면들에 묘사된 예들에 제한되지 않는다.

도면의 간단한 설명

[0013] 도 1은 일 실시예 하에서, 변환-기반 오디오 코덱에서 오디오 신호를 압축하고 확장시키기 위한 시스템을 도시한 도면.

도 2a는 일 실시예 하에서, 복수의 단시간 세그먼트들로 분할된 오디오 신호를 도시한 도면.

도 2b는 일 실시예 하에서, 단시간 세그먼트들의 각각에 걸친 광대역 이득의 적용 후 도 2a의 오디오 신호를 도시한 도면.

도 3a는 일 실시예 하에서, 오디오 신호를 압축하는 방법을 도시하는 흐름도.

도 3b는 일 실시예 하에서, 오디오 신호를 확장시키는 방법을 도시하는 흐름도.

도 4는 일 실시예 하에서, 오디오 신호를 압축하기 위한 시스템을 도시하는 블록도.

도 5는 일 실시예 하에서, 오디오 신호를 확장시키기 위한 시스템을 도시하는 블록도.

도 6은 일 실시예 하에서, 복수의 단시간 세그먼트들로의 오디오 신호의 분할을 도시하는 도면.

발명을 실시하기 위한 구체적인 내용

[0014] 시스템들 및 방법들이 오디오 코덱에서 양자화 잡음의 시간 잡음 성형을 달성하도록 압신 기술들의 사용을 위해 설명된다. 이러한 실시예들은 양자화 잡음의 시간적 성형을 달성하기 위해 QMF-도메인에 구현된 압신 알고리즘의 사용을 포함한다. 프로세스들은 원하는 디코더 압신 레벨의 인코더 제어, 및 스테레오 및 다-채널 압신으로의 모노포닉 적용들을 넘는 확장을 포함한다.

[0015] 여기에 설명된 하나 이상의 실시예들의 양태들은 소프트웨어 지시들을 실행하는 하나 이상의 컴퓨터들 또는 프로세싱 디바이스들을 포함하는 네트워크에 걸친 송신을 위한 오디오 신호들을 프로세싱하는 오디오 시스템에 구현될 수 있다. 설명된 실시예들 중 임의의 것은 단독으로 또는 임의의 조합으로 서로 함께 사용될 수 있다. 다양한 실시예들이 명세서에서의 하나 이상의 곳들에서 논의되거나 시사될 수 있는, 종래 기술이 가진 다양한 결점들에 의해 동기가 부여될 수 있지만, 실시예들이 반드시 이들 결점들 중 임의의 것을 다루는 것은 아니다. 다시 말해서, 상이한 실시예들은 명세서에서 논의될 수 있는 상이한 결점들을 다룰 수 있다. 몇몇 실시예들은 명세서에서 논의될 수 있는 몇몇 결점들 또는 단지 하나의 결점을 단지 부분적으로 다룰 수 있으며, 몇몇 실시예들은 이들 결점들 중 임의의 것을 다루지 않을 수 있다.

- [0016] 도 1은 일 실시예 하에서, 코덱-기반 오디오 프로세싱 시스템에서 양자화 잡음을 감소시키기 위한 압신 시스템을 도시한다. 도 1은 인코더(또는 "코어 인코더")(106) 및 디코더(또는 "코어 디코더")(112)를 포함하는 오디오 코덱 주위에 구성된 오디오 신호 프로세싱 시스템을 도시한다. 상기 인코더(106)는 오디오 콘텐츠를 그것이 재생 또는 추가 프로세싱을 위해 디코더(112)에 의해 디코딩되는 네트워크(110)에 걸친 송신을 위한 데이터 스트림 또는 신호로 인코딩한다. 일 실시예에서, 코덱의 인코더(106) 및 디코더(112)는 디지털 오디오 데이터의 저장 및/또는 데이터 레이트 요건들을 감소시키기 위해 손실 압축 방법을 구현하고, 이러한 코덱은 MP3, Vorbis, Dolby Digital(AC-3), AAC, 또는 유사한 코덱으로서 구현될 수 있다. 코덱의 손실 압축 방법은 일반적으로 코덱에 의해 정의된 프레임의 전진에 걸쳐 레벨이 변하지 않는 코딩 잡음을 생성한다. 이러한 코딩 잡음은 종종 프레임의 저 강도 부분들 동안 가장 잘 들릴 수 있다. 시스템(100)은 코덱의 코어 인코더(106) 이전에 압축 전-단계 구성요소(104) 및 코어 디코더(112) 출력 상에서 동작하는 확장 후-단계 구성요소(114)를 제공함으로써 기존의 코딩 시스템들에서 지각된 코딩 잡음을 감소시키는 구성요소들을 포함한다. 상기 압축 구성요소(104)는 정의된 윈도우 형태를 사용하여 원래 오디오 입력 신호(102)를 복수의 시간 세그먼트들로 분할하고, 초기 오디오 신호의 주파수 도메인 샘플들의 비-에너지 기반 평균을 사용하여 주파수 도메인에 광대역 이득을 산출 및 적용하도록 구성되고, 여기에서 각각의 시간 세그먼트에 적용된 이득 값들은 비교적 낮은 강도의 세그먼트들을 증폭시키며 비교적 높은 강도의 세그먼트들을 감쇠시킨다. 이러한 이득 변경은 입력 오디오 신호(102)의 원래 동적 범위를 압축하거나 상당히 감소시키는 효과를 가진다. 압축된 오디오 신호는 그 후 인코더(106)에서 코딩되고, 네트워크(110)를 통해 송신되고 디코더(112)에서 디코딩된다. 디코딩된 압축 신호는 원래 입력 오디오 신호(102)의 동적 범위로 다시 압축된 오디오 신호의 동적 범위를 확장시키기 위해 각각의 시간 세그먼트에 역 이득 값들을 적용함으로써 압축 전-단계(104)의 역 동작을 수행하도록 구성되는, 확장 구성요소(114)에 입력된다. 따라서, 오디오 출력 신호(116)는 원래 동적 범위를 가진 오디오 신호를 포함하고, 코딩 잡음은 전- 및 후-단계 압신 프로세스를 통해 제거된다.
- [0017] 도 1에 도시된 바와 같이, 압축 구성요소 또는 압축 전-단계(104)는 코어 인코더(106)에 입력된 오디오 신호(102)의 동적 범위를 감소시키도록 구성된다. 입력 오디오 신호는 다수의 짧은 세그먼트들로 분할된다. 각각의 짧은 세그먼트의 크기 또는 길이는 코어 인코더(106)에 의해 사용된 프레임 크기의 부분이다. 예를 들면, 코어 코덱의 통상적인 프레임 크기는 약 40 내지 80 밀리초들일 수 있다. 이러한 경우에, 각각의 짧은 세그먼트는 약 1 내지 3 밀리초들일 수 있다. 압축 구성요소(104)는 세그먼트 단위 기반으로 입력 오디오 신호를 압축하기 위해 적절한 광대역 이득 값을 산출한다. 이것은 각각의 세그먼트에 대한 적절한 이득 값에 의해 신호의 짧은 세그먼트들을 변경함으로써 달성된다. 비교적 큰 이득 값들이 비교적 낮은 강도의 세그먼트들을 증폭시키기 위해 선택되고, 작은 이득 값들이 높은 강도의 세그먼트들을 감쇠시키기 위해 선택된다.
- [0018] 도 2a는 일 실시예 하에서, 복수의 단 시간 세그먼트들로 분할된 오디오 신호를 도시하고, 도 2b는 압축 구성요소에 의한 광대역 이득의 적용 후 동일한 오디오 신호를 도시한다. 도 2a에 도시된 바와 같이, 오디오 신호(202)는 타악기(예로서, 캐스터네즈)에 의해 생성될 수 있는 바와 같은 과도 또는 사운드 임펄스를 나타낸다. 신호는 전압(V) 대 시간(t)의 플롯에 도시된 바와 같이 진폭에서의 스파이크를 특징으로 삼는다. 일반적으로, 신호의 진폭은 음향 에너지 또는 사운드의 강도에 관련되고 임의의 시간 포인트에서 사운드의 전력의 측정을 나타낸다. 오디오 신호(202)가 프레임-기반 오디오 코덱을 통해 프로세싱될 때, 신호의 부분들은 변환(예로서, MDCT) 프레임들(204) 내에서 프로세싱된다. 통상적인 현재 디지털 오디오 시스템들은 비교적 긴 지속 기간의 프레임들을 이용하고, 따라서 선명한 과도 또는 짧은 임펄스 사운드들에 대해, 단일 프레임은 높은 강도 뿐만 아니라 낮은 강도의 사운드들을 포함할 수 있다. 따라서, 도 1에 도시된 바와 같이, 단일 MDCT 프레임(204)은 피크 전 및 후에 비교적 많은 양의 저 강도 신호 뿐만 아니라 오디오 신호의 임펄스 부분(피크)을 포함한다. 일 실시예에서, 압축 구성요소(104)는 신호를 다수의 단시간 세그먼트들(206)로 분할하고, 신호(202)의 동적 범위를 압축하기 위해 각각의 세그먼트에 광대역 이득을 적용한다. 각각의 짧은 세그먼트의 수 및 크기는 애플리케이션 요구들 및 시스템 제약들에 기초하여 선택될 수 있다. 개개의 MDCT 프레임의 크기에 대하여, 짧은 세그먼트들의 수는 범위가 12개에서 64개의 세그먼트들에 이를 수 있으며, 통상적으로 32개의 세그먼트들을 포함할 수 있지만, 실시예들은 그렇게 제한되지 않는다.
- [0019] 도 2b는 일 실시예 하에서, 단시간 세그먼트들의 각각에 걸친 광대역 이득의 적용 후 도 2a의 오디오 신호를 도시한다. 도 2b에 도시된 바와 같이, 오디오 신호(212)는 원래 신호(202)와 동일한 상대적인 형태를 갖지만, 저 강도 세그먼트들의 진폭은 이득 값들을 증폭시키는 애플리케이션에 의해 증가되고, 고 강도 세그먼트들의 진폭은 이득 값들을 감쇠시키는 애플리케이션에 의해 감소되어 왔다.
- [0020] 코어 디코더(112)의 출력은 코어 인코더(106)에 의해 도입된 양자화 잡음 더하기 감소된 동적 범위를 가진 입력

오디오 신호(예로서, 신호(212))이다. 이러한 양자화 잡음은 각각의 프레임 내에서 시간에 걸쳐 거의 균일한 레벨을 특징으로 삼는다. 확장 구성요소(114)는 원래 신호의 동적 범위를 복원하기 위해 디코딩된 신호에 따라 동작한다. 그것은 짧은 세그먼트 크기(206)에 기초하여 동일한 단시간 분해능을 사용하고 압축 구성요소(104)에서 적용된 이득들을 반전시킨다. 따라서, 확장 구성요소(114)는 원래 신호에서 낮은 강도를 가지며, 압축기에 의해 증폭되는 세그먼트들 상에 작은 이득(감쇠)을 적용하고, 원래 신호에서 높은 강도를 가지며 압축기에 의해 감쇠되는 세그먼트들 상에 큰 이득(증폭)을 적용한다. 균일한 시간 엔벨로프를 갖는, 코어 코더에 의해 압축된 양자화 잡음은 그에 따라 대략 원래 신호의 시간적 엔벨로프를 따르기 위해 후-처리 이득에 의해 동시에 성형된다. 이러한 프로세싱은 효과적으로 양자화 잡음이 조용한 흐름들 동안 덜 들리게 한다. 잡음은 고 강도의 흐름들 동안 증폭될 수 있지만, 그것은 오디오 콘텐츠 자체의 소리가 큰 신호의 마스킹 효과로 인해 덜 들리는 채로 있다.

[0021] 도 2a에 도시된 바와 같이, 압신 프로세스는 각각의 이득 값들을 갖고 개별적으로 오디오 신호의 이산 세그먼트들을 변경한다. 특정한 경우들에서, 이것은 코어 인코더(106)에서 문제점들을 야기할 수 있는 압축 구성요소의 출력에서 불연속성들을 야기할 수 있다. 마찬가지로, 확장 구성요소(114)에서 이득에서의 불연속성들은 성형된 잡음의 엔벨로프에서 불연속성들을 야기할 수 있으며, 이것은 오디오 출력(116)에서 가청 클럭들을 야기할 수 있다. 짧은 세그먼트들로의 개개의 이득 값들의 적용에 관련된 또 다른 이슈는 통상적인 오디오 신호들이 많은 개개의 소스들의 혼합이라는 사실에 기초한다. 이들 소스들의 일부는 시간에 걸쳐 변하지 않을 수 있으며, 몇몇은 과도 상태일 수 있다. 불변 신호는 일반적으로 시간에 걸쳐 그것들의 통계 파라미터들에서 일정한 반면, 과도 신호들은 일반적으로 일정하지 않다. 과도 상태들의 광대역 특징을 고려해볼 때, 이러한 혼합에서의 그것들의 핑거프린트는 보통 상위 주파수들에서 더 가시적이다. 신호의 단기 에너지(RMS)에 기초하는 이득 산출은 보다 강한 저 주파수들을 향해 바이어싱되려는 경향이 있으며 그러므로 불변 소스들에 의해 주도되고, 시간에 걸쳐 적은 변화를 보인다. 따라서, 이러한 에너지-기반 접근법은 일반적으로 코어 인코더에 의해 도입된 잡음을 성형할 때 비효과적이다.

[0022] 일 실시예에서, 시스템(100)은 개개의 이득 값들의 적용과 연관된 잠재적인 이슈들을 해결하기 위해 짧은 프로토타입 필터를 가진 필터-뱅크에서의 압축 및 확장 구성요소들에서 이득을 산출하고 적용한다. 변경된 신호(압축 구성요소(104)에서의 원래 신호, 및 확장 구성요소(114)에서의 코어 디코더(112)의 출력)는 먼저 필터-뱅크에 의해 분석되고 광대역 이득은 주파수 도메인에서 직접 적용된다. 시간 도메인에서 대응하는 효과는 물론 프로토타입 필터의 형태에 따라 이득 적용을 평활하게 하는 것이다. 이것은 상기 설명된 불연속성들의 이슈들을 해결한다. 변경된 주파수 도메인 신호는 그 후 대응하는 합성 필터-뱅크를 통해 시간 도메인으로 다시 변환된다. 필터뱅크를 갖고 신호를 분석하는 것은 그것의 스펙트럼 콘텐츠로의 액세스를 제공하고, 고 주파수들로 인한 기여를 우선적으로 신장시키는(또는 약한 임의의 스펙트럼 콘텐츠로 인한 기여를 신장시키기 위해) 이득의 산출을 허용하여, 신호에서 가장 강한 구성요소들에 의해 주도되지 않는 이득 값들을 제공한다. 이것은 상기 설명된 바와 같이, 상이한 소스들의 혼합을 포함하는 오디오 소스들과 연관된 문제점을 해결한다. 일 실시예에서, 시스템은 스펙트럼 크기들의 p-놈(norm)을 사용하여 이득을 산출하고, 여기에서 p는 통상적으로 2 미만($p < 2$)이다. 이것은 그것이 에너지($p=2$)에 기초할 때에 비교하여, 약한 스펙트럼 콘텐츠에 대한 더 많은 강화를 가능하게 한다.

[0023] 상기 서술된 바와 같이, 시스템은 이득 적용을 평활하게 하기 위해 프로토타입 필터를 포함한다. 일반적으로, 프로토타입 필터는 필터뱅크에서 기본 윈도우 형태이며, 이것은 필터뱅크들에서 상이한 서브대역 필터들에 대한 임펄스 응답들을 얻기 위해 정현 파형들에 의해 변조된다. 예를 들면, 단-시간 푸리에 변환(STFT)은 필터뱅크이며, 이러한 변환의 각각의 주파수 라인은 필터뱅크의 서브대역이다. 단-시간 푸리에 변환은, 직사각형, Hann, 카이저-베셀 도출(Kaiser-Bessel derived), 또는 몇몇 다른 형태일 수 있는, 윈도우 형태(N-샘플 윈도우)로 신호를 곱함으로써 구현된다. 윈도우 신호는 그 후 STFT를 획득하기 위해, 이산 푸리에 변환(DFT) 동작의 대상이 된다. 이 경우에 윈도우 형태는 프로토타입 필터이다. 상기 DFT는, 각각이 상이한 주파수의, 정현파 기저 함수들로 구성된다. 정현파 함수에 의해 곱하여진 윈도우 형태는 그 후 주파수에 대응하는 서브대역을 위한 필터를 제공한다. 윈도우 형태는 모든 주파수들에서 동일하므로, 그것은 "프로토타입"으로서 불리운다.

[0024] 일 실시예에서, 시스템은 필터뱅크를 위한 QMF(직교 변조 필터) 뱅크를 이용한다. 특정한 구현에서, QMF 뱅크는 64-pt 윈도우를 가질 수 있으며, 이것은 프로토타입을 형성한다. 코사인 및 사인 함수들에 의해 변조된 이러한 윈도우(64개의 동일하게 이격된 주파수들에 대응하는)는 QMF 뱅크를 위한 서브대역 필터들을 형성한다. QMF 함수의 각각의 적용 후, 윈도우는 64개의 샘플들에 의해 움직여지며, 즉 이 경우에 시간 세그먼트들 사이에서의 중첩은 $640 - 64 = 576$ 샘플들이다. 그러나, 윈도우 형태가 이 경우에 10배의 세그먼트들을 스패닝(spanning)하지

만(640 = 10*64), 윈도우의 주 로브(그것의 샘플 값들이 매우 중요한)는 길이가 약 128 샘플들이다. 따라서, 윈도우의 유효 길이는 여전히 비교적 짧다.

[0025] 일 실시예에서, 확장 구성요소(114)는 이상적으로 압축 구성요소(104)에 의해 적용된 이득들을 반전시킨다. 비트스트림을 통해 압축 구성요소에 의해 적용된 이득들을 디코더에 송신하는 것이 가능하지만, 이러한 접근법은 통상적으로 상당한 비트-레이트를 소모할 것이다. 일 실시예에서, 시스템(100)은 대신에 어떤 부가적인 비트들도 실질적으로 요구하지 않는, 그것, 즉 디코더(112)의 출력에 이용 가능한 신호로부터 직접 확장 구성요소(114)에 의해 요구된 이득들을 추정한다. 압축 및 확장 구성요소들에서의 필터뱅크는 서로의 역일인 이득들을 산출하기 위해 동일한 것으로 선택된다. 또한, 이들 필터뱅크들은 압축 구성요소(104)의 출력 및 확장 구성요소(114)로의 입력 사이에서의 임의의 유효 지연들이 필터뱅크의 스트라이드의 배수이도록 시간 동기화된다. 코어 인코더-디코더가 무손실이며, 필터뱅크가 완전한 재구성을 제공한다면, 압축 및 확장 구성요소들에서의 이득들은 서로의 정확한 역일 것이며, 따라서, 원래 신호의 정확한 재구성을 허용한다. 실제로, 그러나, 확장 구성요소(114)에 의해 적용된 이득은 단지 압축 구성요소(104)에 의해 적용된 이득의 역에 매우 가깝다.

[0026] 일 실시예에서, 압축 및 확장 구성요소들에 사용된 필터뱅크는 QMF 뱅크이다. 통상적인 사용 애플리케이션에서, 코어 오디오 프레임은 이웃 프레임과 2048의 중첩을 가진 길이가 4096 샘플들일 수 있다. 48kHz에서, 이러한 프레임은 길이가 85.3 밀리초들일 것이다. 반대로, 사용되는 QMF 뱅크는 64 샘플들(길이가 1.3 ms이다)의 스트라이드를 가질 수 있으며, 이것은 이득들에 대한 미세한 시간적 분해능을 제공한다. 뿐만 아니라, QMF는 이득 적용이 시간에 걸쳐 평활하게 달라지는 것을 보장하는 길이가 640 샘플들인 평활한 프로토타입 필터를 가진다. 이러한 QMF 필터뱅크를 가진 분석은 신호의 시간-주파수 타일 표현을 제공한다. 각각의 QMF 시간-슬롯은 스트라이드와 같으며 각각의 QMF 시간-슬롯에서 64개의 균일하게 이격된 서브대역들이 있다. 대안적으로, 단기 푸리에 변환(STFT)과 같은 다른 필터뱅크들이 이용될 수 있으며, 이러한 시간-주파수 타일 표현이 여전히 획득될 수 있다.

[0027] 일 실시예에서, 압축 구성요소(104)는 코덱 입력을 스케일링하는 전-처리 단계를 수행한다. 이러한 실시예에 대해, $S_t(k)$ 는 시간 슬롯(t) 및 주파수 빈(k)에서의 복소 값 필터 뱅크 샘플이다. 도 6은 일 실시예 하에서, 오디오 신호의 일련의 주파수들을 위한 다수의 시간 슬롯들로의 분할을 도시한다. 다이어그램(600)의 실시예에 대해, 도시된 바와 같이(반드시 일정한 비율로 그려지지는 않지만) 복수의 시간-주파수 타일들을 생성하는 64개의 주파수 빈들(k), 및 32개의 시간 슬롯들(t)이 있다. 압축 전-단계들은 $S'_t(k) = S_t(k)/g_t$ 가 되도록 코덱 입력을 스케일링한다. 이러한 식에서, $g_t = (\bar{S}_t/S_0)^{\gamma}$ 은 정규화된 슬롯 평균이다.

[0028] 상기 식에서, 표현($\bar{S}_t = \frac{1}{K} \sum_{k=1}^K |S_t(k)|$)은 평균 절대 레벨/1-놈이며 S_0 은 적절한 상수이다. 일반적인 p-놈은 다음과 같이 이러한 콘텍스트에서 정의된다:

$$\bar{S}_t = \left(\frac{1}{K} \sum_{k=1}^K |S_t(k)|^p \right)^{1/p}$$

[0029] 1-놈은 에너지(rms/2-놈)를 사용한 것보다 상당히 더 양호한 결과들을 제공할 수 있다는 것이 보여지고 있다. 지수 항(γ)의 값은 통상적으로 0 및 1 사이에서의 범위에 있으며, 1/3이도록 선택될 수 있다. 상수(S_0)는 구현 플랫폼과는 관계없이 적절한 이득 값들을 보장한다. 예를 들면, 그것은 $S_t(k)$ 값들 모두가 절대 값에서 1로 제한될 수 있는 플랫폼에서 구현될 때 1일 수 있다. 그것은 $S_t(k)$ 가 상이한 최대 절대 값을 가질 수 있는 플랫폼에서 잠재적으로 상이할 수 있다. 그것은 또한 큰 세트의 신호들에 걸친 평균 이득 값이 1에 가까움을 보장하기 위해 사용될 수 있다. 즉, 그것은 콘텐츠의 큰 코퍼스들로부터 결정된 최소 신호 값 및 최대 신호 값 사이에서의 중간 신호 값일 수 있다.

[0030] 확장 구성요소(114)에 의해 수행된 후-단계 프로세스에서, 코덱 출력은 압축 구성요소(104)에 의해 적용된 역 이득만큼 확장된다. 이것은 압축 구성요소의 필터 뱅크의 정확한 또는 거의-정확한 레플리카를 요구한다. 이 경우, $\tilde{S}_t(k)$ 는 이러한 제 2 필터 뱅크의 복소 값 샘플을 나타낸다. 확장 구성요소(114)는 $\tilde{S}'_t(k) = \tilde{S}_t(k) \cdot \tilde{g}_t$ 이도록 코덱 출력을 스케일링한다.

[0032] 상기 식에서, $\tilde{g}_t$ 은 다음으로서 주어진 정규화된 슬롯 평균이다:

$$\tilde{g}_t = (\bar{\tilde{S}}_t/S_0)^{\gamma/(1-\gamma)} \quad \text{및}$$

$$\bar{\tilde{S}}_t = (\frac{1}{K} \sum_{k=1}^K |\tilde{S}_t(k)|^p)^{1/p}$$

[0035] 일반적으로, 확장 구성요소(114)는 압축 구성요소(104)에서 사용된 바와 동일한 p-놈을 사용할 것이다. 따라서, 평균 절대 레벨이 압축 구성요소(104)에서 $\tilde{S}_t$ 를 정의하기 위해 사용된다면, $\bar{\tilde{S}}_t$ 이 또한 상기 식에서 1-놈(p=1) 을 사용하여 정의된다.

[0036] STFT 또는 복소-QMF와 같은 복소 필터뱅크(코사인 및 사인 기저 함수들 양쪽 모두로 이루어진)가 압축 및 확장 구성요소들에서 사용될 때, 복소 서브대역 샘플의 크기($|\tilde{S}_t(k)|$ 또는 $|S_t(k)|$)의 산출은 계산 집중적인 제공-근 동작을 요구한다. 이것은 다양한 방식들로, 복소 서브대역 샘플의 크기를 근사함으로써, 예를 들면, 예로서 그것의 실수 및 허수 부분들의 크기를 합산함으로써 피해질 수 있다.

[0037] 상기 식들에서, 값(K)은 필터뱅크에서의 서브대역들의 수와 같거나, 그보다 작다. 일반적으로, p-놈은 필터뱅크에서 서브대역들의 임의의 서브세트를 사용하여 산출될 수 있다. 그러나, 동일한 서브세트는 인코더(106) 및 디코더(112) 양쪽 모두에서 이용되어야 한다. 일 실시예에서, 오디오 신호의 고 주파수 부분들(예로서, 6 kHz 이상의 오디오 구성요소들)은 향상된 스펙트럼 확장(A-SPX) 톨을 갖고 코딩될 수 있다. 부가적으로, 잡음-성형을 이끌기 위해 1 kHz(또는 유사한 주파수) 이상의 신호만을 사용하는 것이 바람직할 수 있다. 이러한 경우에, 단지 1 kHz 내지 6 kHz의 범위에서의 이들 서브대역들만이 p-놈, 및 그러므로 이득 값을 산출하기 위해 사용될 수 있다. 더욱이, 이득이 서브대역들의 하나의 서브세트로부터 산출되지만, 그것은 여전히 서브대역들의 상이한, 가능하게는 더 큰 서브세트에 적용될 수 있다.

[0038] 도 1에 도시된 바와 같이, 오디오 코덱의 코어 인코더(106)에 의해 도입된 양자화 잡음을 성형하기 위한 압신 기능이 특정한 전-인코더 압축 기능들 및 후-디코더 확장 기능들을 수행하는 두 개의 별개의 구성요소들(104 및 114)에서 수행된다. 도 3a는 일 실시예 하에서, 전-인코더 압축 구성요소에서 오디오 신호를 압축하는 방법을 도시한 흐름도이며, 도 3b는 일 실시예 하에서, 후-디코더 확장 구성요소에서 오디오 신호를 확장하는 방법을 도시한 흐름도이다.

[0039] 도 3a에 도시된 바와 같이, 프로세스(300)는 압축 구성요소가 입력 오디오 신호를 수신하는 것으로 시작한다(302). 이러한 구성요소는 그 후 오디오 신호를 단시간-세그먼트들로 분할하고(304) 상기 짧은 세그먼트들의 각각에 광대역 이득 값들을 적용함으로써 감소된 동적 범위로 오디오 신호를 압축한다(306). 압축 구성요소는 또한 상기 설명된 바와 같이, 인접한 세그먼트들에 상이한 이득 값들을 적용함으로써 야기된 임의의 불연속성들을 감소시키거나 제거하기 위해 특정한 프로토타입 필터링 및 QMF 필터뱅크 구성요소들을 구현한다(308). 오디오 콘텐츠의 유형 또는 오디오 콘텐츠의 특정한 특성들에 기초하여서와 같은, 특정한 경우들에서, 오디오 코덱의 인코딩/디코딩 스테이지들 전 및 후에 오디오 신호의 압축 및 확장은 출력 오디오 품질을 강화하기보다는 저하시킬 수 있다. 이러한 인스턴스들에서, 압신 프로세스는 상이한 압신(압축/확장) 레벨들을 리턴하기 위해 턴 오프되거나 변경될 수 있다. 따라서, 압축 구성요소는 다른 변수들 중에서, 특정 신호 입력 및 오디오 재생 환경을 위해 요구된 압신의 최적의 레벨 및/또는 압신 기능의 적절성을 결정한다(310). 이러한 결정 단계(310)는 오디오 신호의 분할(304) 또는 오디오 신호의 압축(306) 이전에서와 같은, 프로세스(300)의 임의의 실제 포인트에서 발생할 수 있다. 압신이 적절한 것으로 간주된다면, 이득들이 적용되고(306), 인코더는 그 후 코덱의 데이터 포맷에 따른 디코더로의 송신을 위해 신호를 인코딩한다(312). 활성화 데이터, 동기화 데이터, 압신 레벨 데이터, 및 다른 유사한 제어 데이터와 같은, 특정한 압신 제어 데이터가 확장 구성요소에 의한 프로세싱을 위한 비트스트림의 일부로서 송신될 수 있다.

[0040] 도 3b는 일 실시예 하에서, 후-디코더 확장 구성요소에서 오디오 신호를 확장시키는 방법을 도시한 흐름도이다. 프로세스(350)에 도시된 바와 같이, 코덱의 디코더 스테이지는 인코더 스테이지로부터 오디오 신호를 인코딩한 비트스트림을 수신한다(352). 디코더는 그 후 코덱 데이터 포맷에 따라 인코딩된 신호를 디코딩한다(353). 확장 구성요소는 그 후 비트스트림을 프로세싱하고 확장을 스위치 오프하거나 제어 데이터에 기초하여 확장 파라미터들을 변경하기 위해 임의의 인코딩된 제어 데이터를 적용한다(354). 확장 구성요소는 적절한 윈도우 형태를 사

용하여 오디오 신호를 시간 세그먼트들로 분할한다(356). 일 실시예에서, 시간 세그먼트들은 압축 구성요소에 의해 사용된 동일한 시간 세그먼트들에 대응한다. 확장 구성요소는 그 후 주파수 도메인에서 각각의 세그먼트에 대한 적절한 이득 값들을 산출하고(358) 원래의 동적 범위, 또는 임의의 다른 적절한 동적 범위로 다시 오디오 신호의 동적 범위를 확장시키기 위해 각각의 시간 세그먼트에 이득 값들을 적용한다(360).

[0041] 압신 제어

[0042] 시스템(100)의 압신기를 포함하는 압축 및 확장 구성요소들은 오디오 신호 프로세싱 동안 단지 특정한 시간에, 또는 단지 특정한 유형들의 오디오 콘텐츠에 대해 전 및 후-처리 단계들을 적용하도록 구성될 수 있다. 예를 들면, 압신은 스피치 및 음악 과도 신호들을 위한 이득들을 보일 수 있다. 그러나, 불변 신호들과 같은, 다른 신호들에 대해, 압신은 신호 품질을 저하시킬 수 있다. 따라서, 도 3a에 도시된 바와 같이, 압신 제어 메커니즘이 블록(310)으로서 제공되고, 제어 데이터는 압신 동작을 조정하기 위해 압축 구성요소(104)에서 확장 구성요소(114)로 송신된다. 이러한 제어 메커니즘의 가장 간단한 형태는 압신의 적용이 오디오 품질을 저하시키는 오디오 샘플들의 블록들에 대한 압신 기능을 스위치 오프하는 것이다. 일 실시예에서, 압축기 및 확장은 동일한 QMF 시간 슬롯에서 스위치 온/오프될 수 있도록 압신 온/오프 결정이 인코더에서 검출되고 비트스트림 요소로서 디코더에 송신된다.

[0043] 두 개의 상태들 사이에서의 스위칭은 보통 적용된 이득에서의 불연속성으로 이어질 것이며, 가청 스위칭 아티팩트들 또는 클릭들을 야기한다. 실시예들은 이들 아티팩트들을 감소시키거나 제거하기 위해 메커니즘들을 포함한다. 제 1 실시예에서, 시스템은 단지 이득이 1에 가까운 프레임들에서만 압신 기능의 스위칭 온 및 오프를 허용한다. 이 경우에, 단지 압신 기능의 스위칭 온/오프 사이에서의 작은 불연속성만이 있다. 제 2 실시예에서, 온 및 오프 모드 사이에 있는, 제 3 약 압신 모드는 온 및 오프 프레임들 사이에서의 오디오 프레임에 적용되고, 비트스트림으로 시그널링된다. 약 압신 모드는 지수 γ 을 압신 동안 그것의 디폴트 값에서 0으로 느리게 전이시키며, 이것은 무 압신의 등가이다. 중간 약 압신 모드에 대한 대안으로서, 시스템은 오디오 샘플들의 블록에 걸쳐 압신 기능을 갑자기 스위칭 오프하는 대신에 압신-외 모드로 평활하게 페이딩하는 시작-프레임들 및 정지-프레임들을 구현할 수 있다. 추가 실시예에서, 시스템은 간단히 압신을 스위칭 오프하기보다는 평균 이득을 적용하도록 구성된다. 특정한 경우들에서, 계조-불변 신호들의 오디오 품질은 일정한 이득 인자가 압신 오프 상황에서 1.0의 일정한 이득 인자보다 인접한 압신-온 프레임들의 이득 인자들과 더 크게 비슷한 오디오 프레임에 적용된다면 증가될 수 있다. 이러한 이득 인자는 하나의 프레임에 걸쳐 모든 압신 이득들을 평균함으로써 산출될 수 있다. 일정한 평균 압신 이득을 포함하는 프레임은 그에 따라 비트스트림으로 시그널링된다.

[0044] 실시예들이 모노포닉 오디오 채널의 컨텍스트에서 설명될지라도, 간단한 확장에서 다수의 채널들이 각각의 채널 상에서 개별적으로 접근법을 반복함으로써 핸들링될 수 있다는 것이 주의되어야 한다. 그러나, 둘 이상의 채널들을 포함하는 오디오 신호들은 도 1의 압신 시스템의 실시예들에 의해 처리되는 특정한 추가적인 복잡도들을 준다. 압신 전략은 채널들 사이에서의 유사성에 의존해야 한다.

[0045] 예를 들면, 스테레오-패닝 과도 신호들의 경우에, 개개의 채널들의 독립적인 압신이 가청 이미지 아티팩트들을 야기할 수 있다는 것이 관찰되어 왔다. 일 실시예에서, 시스템은 양쪽 채널들의 서브대역 샘플들로부터 각각의 시간-세그먼트에 대한 단일 이득 값을 결정하고 두 개의 신호들을 압축/확장시키기 위해 동일한 이득 값을 사용한다. 이러한 접근법은 일반적으로 두 개의 채널들이 매우 유사한 신호들을 가질 때마다 적절하고, 여기에서 유사성은 예를 들면 교차 상관을 사용하여 정의된다. 검출기는 채널들 사이에서의 유사성을 산출하고 채널들의 개개의 압신을 사용하는 것 또는 채널들을 공동으로 압신하는 것 사이에서 스위칭한다. 더 많은 채널들로의 확장들은 유사성 기준들을 사용하여 채널들을 채널들의 그룹들로 분할하고 그룹들에 대한 공동 압신을 적용할 것이다. 이러한 그룹핑 정보는 그 후 비트스트림을 통해 송신될 수 있다.

[0046] 시스템 구현

[0047] 도 4는 일 실시예 하에서, 코덱의 인코더 스테이지와 함께 오디오 신호를 압축하기 위한 시스템을 도시한 블록도이다. 도 4는 도 3a에 도시된 코덱-기반 시스템에서의 사용을 위한 압축 방법의 적어도 일부를 구현하는 하드웨어 회로 또는 시스템을 도시한다. 시스템(400)에 도시된 바와 같이, 시간 도메인에서 입력 오디오 신호(401)는 QMF 필터뱅크(402)로 입력된다. 이러한 필터뱅크는 각각의 대역통과 필터가 원래 신호의 주파수 서브-대역을 운반하는 다수의 구성요소들로 입력 신호를 분리하는 분석 동작을 수행한다. 신호의 재구성은 QMF 필터뱅크(410)에 의해 수행된 합성 동작에서 수행된다. 도 4의 예시적인 실시예에서, 분석 및 합성 필터뱅크들 양쪽 모두는 64개의 대역들을 핸들링한다. 코어 인코더(412)는 합성 필터뱅크(410)로부터 오디오 신호를 수신하고 적절한 디지털 포맷(예로서, MP3, AAC 등)으로 오디오 신호를 인코딩함으로써 비트스트림(414)을 생성한다.

- [0048] 시스템(400)은 오디오 신호가 분할되는 짧은 세그먼트들의 각각에 이득 값을 적용하는 압축기(406)를 포함한다. 이것은 도 2b에 도시된 바와 같이, 압축된 동적 범위 오디오 신호를 생성한다. 압신 제어 유닛(404)은 얼마나 많은 압축이 신호의 유형(예로서, 스피치), 또는 신호의 특성들(예로서, 불변 대 과도), 또는 다른 관련 파라미터들에 기초하여 적용되어야 하는지 여부를 결정하기 위해 오디오 신호를 분석한다. 제어 유닛(404)은 오디오 신호의 시간적 피크니스(peakness) 특성을 검출하기 위해 검출 메커니즘을 포함할 수 있다. 오디오 신호 및 특정한 사전-정의된 기준들의 검출된 특성에 기초하여, 제어 유닛(404)은 압축 기능을 턴 오프하거나 짧은 세그먼트들에 적용된 이득 값들을 변경하기 위해 압축기(406)에 적절한 제어 신호들을 전송한다.
- [0049] 압신 외에, 많은 다른 코딩 톨들이 또한 QMF 도메인에서 동작할 수 있다. 하나의 이러한 톨은 A-SPX(향상된 스펙트럼 확장)이며, 이것은 도 4의 블록(408)에서 도시된다. A-SPX는 각각적으로 덜 중요한 주파수들이 더 중요한 주파수들보다 더 개략적인 코딩 기법을 갖고 코딩되도록 허용하기 위해 사용되는 기술이다. 예를 들면, 디코더 단에서의 A-SPX에서, 하위 주파수로부터의 QMF 서브대역 샘플들은 상위 주파수들에서 복제될 수 있으며, 고 주파수 대역에서의 스펙트럼 엔벨로프는 그 후 인코더에서 디코더로 송신된 부 정보를 사용하여 성형된다.
- [0050] 압신 및 A-SPX 양쪽 모두가 QMF 도메인에서 수행되는 시스템에서, 인코더에서, 상위 주파수들에 대한 A-SPX 엔벨로프 데이터는 도 4에 도시된 바와 같이 아직 압축되지 않은 서브대역 샘플로부터 추출될 수 있으며, 압축은 단지 코어 인코더(412)에 의해 인코딩된 신호의 주파수 범위에 대응하는 하위 주파수 QMF 샘플들에만 적용될 수 있다. 도 5의 디코더(502)에서, 디코딩된 신호의 QMF 분석(504) 후, 확장 프로세스(506)가 먼저 적용되고, A-SPX 동작(508)이 그 다음에 하위 주파수들에서 확장된 신호로부터 보다 높은 서브대역 샘플들을 재생한다.
- [0051] 이러한 예시적인 구현에서, 인코더에서의 QMF 합성 필터뱅크(410) 및 디코더(504)에서의 QMF 분석 필터뱅크는 함께 640-64+1 샘플 지연(~9 QMF 슬롯들)을 도입한다. 이 예에서 코어 코덱 지연은 3200개의 샘플들(50 QMF 슬롯들)이며, 따라서 총 지연은 59 슬롯들이다. 이러한 지연은 제어 데이터를 비트스트림에 내장하고 그것을 디코더에서 사용함으로써 설명되고, 따라서 인코더 압축기 및 디코더 확장이 동작들 양쪽 모두가 동시에 이루어진다.
- [0052] 대안적으로, 인코더에서, 압축은 원래 신호의 전체 대역폭 상에서 적용될 수 있다. A-SPX 엔벨로프 데이터는 그 다음에 압축된 서브대역 샘플들로부터 추출될 수 있다. 이러한 경우에, 디코더는, QMF 분석 후에, 먼저 전체 대역폭 압축 신호를 먼저 재구성하기 위해 A-SPX 톨을 구동한다. 확장 스테이지는 그 후 그것의 원래 동적 범위를 가진 신호를 복원하기 위해 적용된다.
- [0053] QMF 도메인에서 동작할 수 있는 또 다른 톨은 도 4에서의 향상된 결합(AC) 톨(도시되지 않음)일 수 있다. 향상된 결합 시스템에서, 두 개의 채널들이 스테레오 출력을 재구성하기 위해 디코더에서의 QMF 도메인에서 적용될 수 있는 부가적인 파라메트릭 공간 정보를 갖고 모노 다운믹스로서 인코딩된다. AC 및 압신이 서로 함께 사용될 때, AC 톨은 인코더에서 압축 스테이지(406) 후 배치될 수 있으며, 이 경우에 그것은 디코더에서 확장 스테이지(506) 전에 적용될 것이다. 대안적으로, AC 부-정보는 이 경우에 AC 톨이 디코더에서 확장 스테이지(506) 후 동작할 압축되지 않은 스테레오 신호로부터 추출될 수 있다. AC가 특정한 주파수 위에서 사용되고 이산 스테레오가 이러한 주파수 아래에서 사용되거나; 대안적으로 이산 스테레오가 특정한 주파수 위에서 사용되고 AC가 이러한 주파수 아래에서 사용되는 하이브리드 AC 모드가 또한 지원될 수 있다.
- [0054] 도 3a 및 도 3b에 도시된 바와 같이, 코덱의 인코더 스테이지 및 디코더 스테이지 사이에서 송신된 비트스트림은 특정한 제어 데이터를 포함한다. 이러한 제어 데이터는 시스템이 상이한 압신 모드들 사이에서 스위칭하도록 허용하는 부-정보를 구성한다. 스위칭 제어 데이터(압신을 스위칭 온/오프하기 위한) 더하기 잠재적으로 몇몇 중간 상태들이 채널마다 약 1 또는 2 비트들을 부가할 수 있다. 다른 제어 데이터는 이산 스테레오 또는 다채널 구성의 채널들 모두가 공통 압신 이득 인자들을 사용하는지 또는 그것들이 각각의 채널에 대해 독립적으로 산출되어야 하는지를 결정하기 위해 신호를 포함할 수 있다. 이러한 데이터는 단지 채널당 단일 추가 비트를 요구할 수 있다. 다른 유사한 제어 데이터 요소들 및 그것들의 적절한 비트 가중들이 시스템 요건들 및 제약들에 의존하여 사용될 수 있다.
- [0055] 검출 메커니즘
- [0056] 일 실시예에서, 압신 제어 메커니즘은 QMF-도메인에서 압신의 제어를 제공하기 위해 압축 구성요소(104)의 부분으로서 포함된다. 압신 제어는 오디오 신호 유형과 같은, 다수의 인자들에 기초하여 구성될 수 있다. 예를 들면, 대부분의 애플리케이션들에서, 압신은 스피치 신호들 및 과도 신호들 또는 시간적으로 절정의 신호들의 클래스 내에서의 임의의 다른 신호들에 대해 턴 온되어야 한다. 시스템은 압신기 기능을 위한 적절한 제어 신호

를 발생시키도록 돕기 위해 신호의 피크니스를 검출하기 위해 검출 메커니즘을 포함한다.

[0057] 일 실시예에서, 시간적 피크니스($TP(k)_{frame}$)에 대한 측정은 주어진 코어 코덱에 대해 주파수 빈(k)에 걸쳐 계산되고, 다음의 공식을 사용하여 산출된다:

$$TP(k)_{frame} = \frac{\sqrt{\frac{1}{T} \sum_{t=0}^{T-1} S_t(k)^4}}{\sqrt{\frac{1}{T} \sum_{t=0}^{T-1} S_t(k)^2}},$$

[0058]

[0059] 상기 식에서, $S_t(k)$ 는 서브-대역 신호이며, T 는 하나의 코어 인코더 프레임에 대응하는 QMF 슬롯들의 수이다. 일 예시적인 구현에서, T 의 값은 32일 수 있다. 대역당 계산된 시간적 피크니스는 사운드 콘텐츠를 일반적인 두 개의 카테고리들로 분류하기 위해 사용될 수 있다: 불변 음악 신호들, 및 음악 과도 신호들 또는 스피치 신호들. $TP(k)_{frame}$ 의 값이 정의된 값(예로서, 1.2)보다 작다면, 프레임의 상기 서브대역에서의 신호는 불변 음악 신호일 가능성이 높다. $TP(k)_{frame}$ 의 값이 이러한 값보다 크다면, 신호는 음악 과도 신호들 또는 스피치 신호들일 가능성이 높다. 값이 훨씬 더 높은 임계 값(예로서, 1.6)보다 크다면, 신호는 완전한 음악 과도 신호, 예로서 캐스터네즈일 가능성이 매우 높다. 더욱이, 자연스럽게 발생한 신호들에 대해, 상이한 대역들에서 획득된 시간적 피크니스의 값들이 더 많이 또는 덜 유사하다는 것이 관찰되었으며, 이러한 특성은 시간적 피크니스 값이 산출되는 서브대역들의 수를 감소시키기 위해 이용될 수 있다. 이러한 관찰에 기초하여, 시스템은 다음의 두 개 중 하나를 구현할 수 있다.

[0060]

제 1 실시예에서, 검출기는 다음의 프로세스를 실행한다. 제 1 단계로서, 그것은 1.6보다 큰 시간적 피크니스를 가진 대역들의 수를 계산한다. 제 2 단계로서, 그것은 그 후 그것이 1.6보다 작은 대역들의 시간적 피크니스 값들의 평균을 계산한다. 제 1 단계에서 발견된 대역들의 수가 51보다 크다면, 또는 제 2 단계에서 결정된 평균 값이 1.45보다 크다면, 신호는 음악 과도 신호인 것으로 결정되고 그러므로 압신은 스위칭 온되어야 한다. 그렇지 않다면, 그것은 압신이 스위칭 온되지 않아야 하는 신호인 것으로 결정된다. 이러한 검출기는 스피치 신호들을 위한 시간의 대부분을 스위칭 오프할 것이다. 몇몇 실시예들에서, 스피치 신호들은 보통 별개의 스피치 코더에 의해 코딩될 것이며, 따라서 이것은 일반적으로 문제가 아니다. 그러나, 특정한 경우들에서, 또한 스피치를 위한 압신 기능을 스위칭 온하는 것이 요구될 수 있다. 이 경우에, 제 2 유형의 검출기가 바람직할 수 있다.

[0061]

일 실시예에서, 제 2 유형의 검출기는 다음의 프로세스를 실행한다. 제 1 단계로서, 그것은 1.2보다 큰 시간적 피크니스를 가진 대역들의 수를 계산한다. 제 2 단계에서, 그것은 그 후 그것이 1.2보다 작은 대역들의 시간적 피크니스 값들의 평균을 계산한다. 그것은 그 후 다음의 규칙을 적용한다: 제 1 단계의 결과가 55보다 크다면: 압신을 턴 온하고, 제 1 단계의 결과가 15보다 작다면: 압신을 턴 오프하고; 제 1 단계의 결과가 15 및 55 사이에 있고 제 2 단계의 결과가 1.16보다 크다면: 압신을 턴 온하고; 제 1 단계의 결과가 15 및 55 사이에 있고 제 2 단계의 결과가 1.16보다 작다면: 압신을 턴 오프한다. 두 개의 유형들의 검출기들은 검출기 알고리즘을 위한 많은 가능한 해결책들 중 단지 2개의 예들만을 설명하였으며, 다른 유사한 알고리즘들이 또한 또는 대안적으로 사용될 수 있다는 것이 주의되어야 한다.

[0062]

도 4의 요소(404)에 의해 제공된 압신 제어 기능은 압신이 특정한 동작 모드들에 기초하여 사용되거나 사용되지 않도록 허용하기 위해 임의의 적절한 방식으로 구현될 수 있다. 예를 들면, 압신은 일반적으로 서라운드 사운드 시스템의 LFE(저 주파수 효과들) 채널 상에서 사용되지 않으며, 구현된 어떤 A-SPX(즉, QMF 없음) 기능도 없을 때 또한 사용되지 않는다. 일 실시예에서, 압신 제어 기능은 압신 제어 요소(404)와 같은, 회로 또는 프로세서-기반 요소들에 의해 실행된 프로그램에 의해 제공될 수 있다. 일 실시예 하에서, 압신 제어를 구현할 수 있는 프로그램 세그먼트의 몇몇 예시적인 구문이 이어진다.

```

Companding_control(nCh)
{
    sync_flag=0;
    if (nCh>1){
        sync_flag
    }
    b_needAvg=0
    ch_count=sync_flag?1:nCh
    for (ch=0; ch<ch_count; ch++){
        b_compend_on[ch]
        if (!b_compend_on[ch]){
            b_needAvg=1;
        }
    }
    if (b_needAvg){
        b_compend_avg;
    }
}

```

[0063]

[0064]

sync_flag, b_compend_on[ch], 및 b_compend_avg 플래그들 또는 프로그램 요소들은 길이가 약 1-비트, 또는 시스템 제약들 및 조건들에 의존한 임의의 다른 길이일 수 있다. 상기 예시된 프로그램 코드는 압신 제어 기능을 구현하는 하나의 방식의 예이며, 다른 프로그램들 또는 하드웨어 구성요소들이 몇몇 실시예들에 따라 압신 제어를 구현하기 위해 사용될 수 있다는 것이 주의되어야 한다.

[0065]

지금까지 설명된 실시예들은 코덱에서의 인코더에 의해 도입된 양자화 잡음을 감소시키기 위한 압신 프로세스를 포함할지라도, 이러한 압신 프로세스의 양태들은 인코더 및 디코더(코덱) 스테이지들을 포함하지 않는 신호 프로세싱 시스템들에 또한 적용될 수 있다는 것이 주의되어야 한다. 더욱이, 압신 프로세스가 코덱과 함께 사용되는 경우에, 코덱은 변환-기반 또는 비 변환-기반일 수 있다.

[0066]

여기에 설명된 시스템들의 양태들은 디지털 또는 디지털화된 오디오 파일들을 프로세싱하기 위한 적절한 컴퓨터-기반 사운드 프로세싱 네트워크 환경에서 구현될 수 있다. 적응적 오디오 시스템의 부분들은 컴퓨터들 중에서 송신된 데이터를 버퍼링하고 라우팅하도록 작용하는 하나 이상의 라우터들(도시되지 않음)을 포함하여, 임의의 원하는 수의 개개의 기계들을 포함하는 하나 이상의 네트워크들을 포함할 수 있다. 이러한 네트워크는 다양한 상이한 네트워크 프로토콜들 상에서 구성될 수 있으며, 인터넷, 광역 네트워크(WAN), 근거리 네트워크(LAN), 또는 그것의 임의의 조합일 수 있다.

[0067]

구성요소들, 블록들, 프로세스들 또는 다른 기능적 구성요소들 중 하나 이상이 시스템의 프로세서-기반 컴퓨팅 디바이스의 실행을 제어하는 컴퓨터 프로그램을 통해 구현될 수 있다. 여기에 개시된 다양한 기능들이 그것들의 행동, 레지스터 전달, 로직 구성요소, 및/또는 다른 특성들에 대하여, 하드웨어, 펌웨어의 임의의 수의 조합들을 사용하여 및/또는 다양한 기계-관독가능한 또는 컴퓨터-관독가능한 미디어에 구체화된 데이터 및/또는 지시들로서 설명될 수 있다는 것이 또한 주의되어야 한다. 이러한 포맷팅된 데이터 및/또는 지시들이 구체화될 수 있는 컴퓨터-관독가능한 미디어는, 이에 제한되지 않지만, 광학적, 자기 또는 반도체 저장 미디어와 같은, 다양한 형태들에서의 물리적(비-일시적), 비-휘발성 저장 미디어를 포함한다.

[0068]

달리 문맥이 명확하게 요구하지 않는다면, 설명 및 청구항들 전체에 걸쳐, 단어들("포함하다", "포함하는" 등)은 배타적 또는 철저한 의미와는 대조적으로 포괄적인 의미로; 즉 "이에 제한되지 않지만, ~을 포함하는"의 의미로 해석될 것이다. 단수형 또는 복수형 숫자를 사용한 단어들은 또한 각각 복수 또는 단수 숫자를 포함한다. 부가적으로, 단어들("여기에서", "아래에", "상기", "이하에") 및 유사한 의미의 단어들이 본 출원의 임의의 특정한 부분들이 아닌, 전체로서 본 출원을 나타낸다. 단어("또는")가 둘 이상의 아이템들의 리스트를 참조하여 사용될 때, 상기 단어는 단어의 다음의 해석들의 모두를 커버한다: 리스트에서의 아이템들 중 임의의 것, 리스트에서의 아이템들의 모두 및 리스트에서의 아이템들의 임의의 조합.

[0069]

하나 이상의 구현들이 예로서 및 특정한 실시예들에 대하여 설명되었지만, 하나 이상의 구현들은 개시된 실시예들에 제한되지 않는다는 것이 이해될 것이다. 그와는 대조적으로, 그것은 당업자들에게 명백할 바와 같이 다양

한 수정들 및 유사한 배열들을 커버하도록 의도된다. 그러므로, 첨부된 청구항들의 범위는 모든 이러한 변경들 및 유사한 배열들을 포함하도록 가장 광범위한 해석에 부합되어야 한다.

부호의 설명

[0070]

100, 400: 시스템 102: 원래 오디오 입력 신호

104: 압축 구성요소 106: 인코더

110: 네트워크 112, 502: 디코더

114: 확장 구성요소 116: 오디오 출력 신호

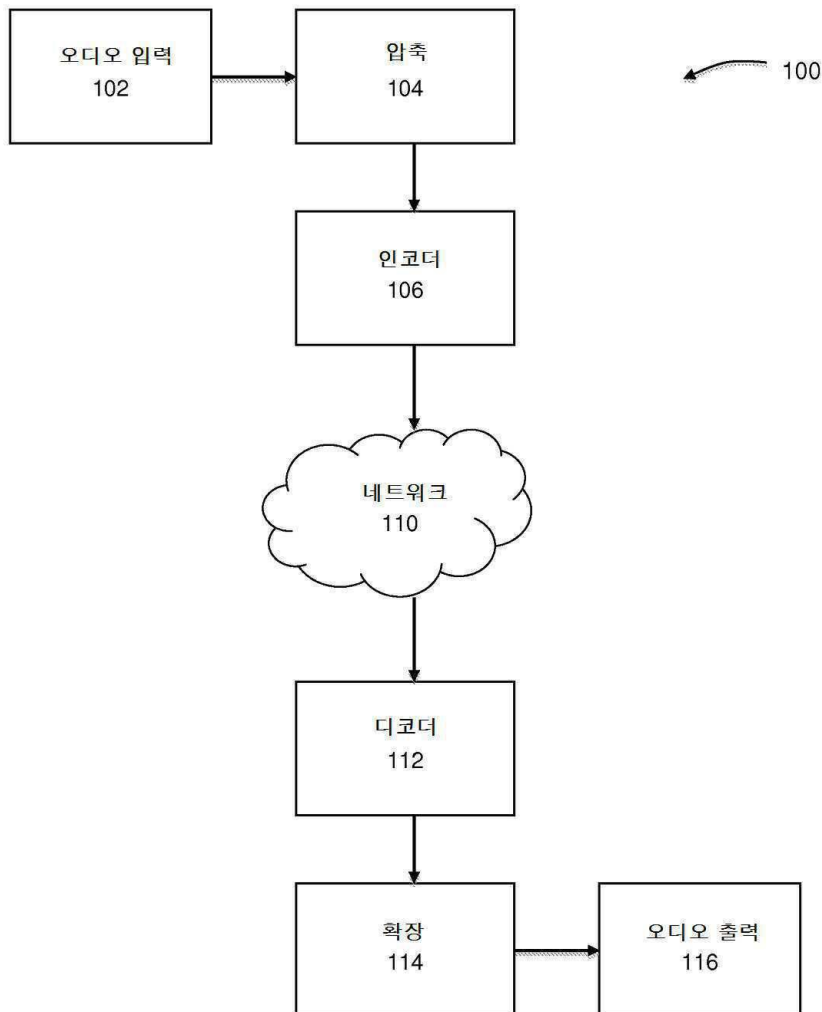
401: 입력 오디오 신호 402: QMF 필터뱅크

404: 압신 제어 유닛 406: 압축기

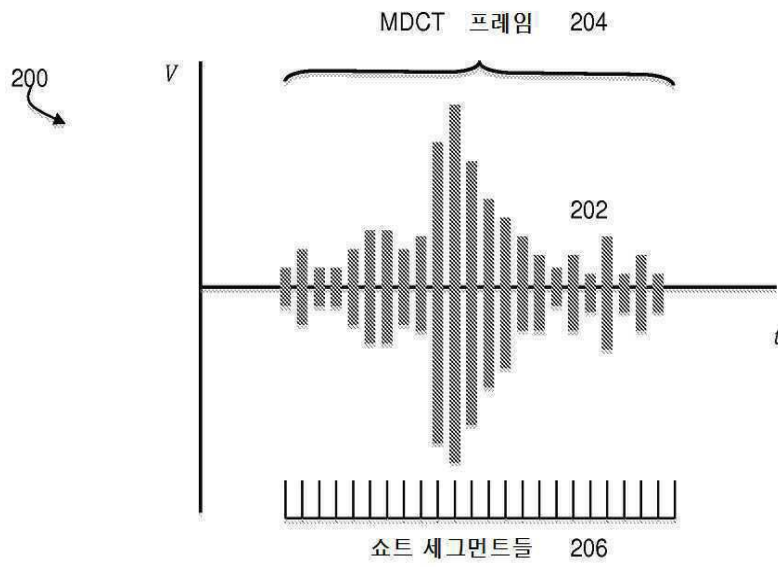
410: 합성 필터뱅크 412: 코어 인코더

도면

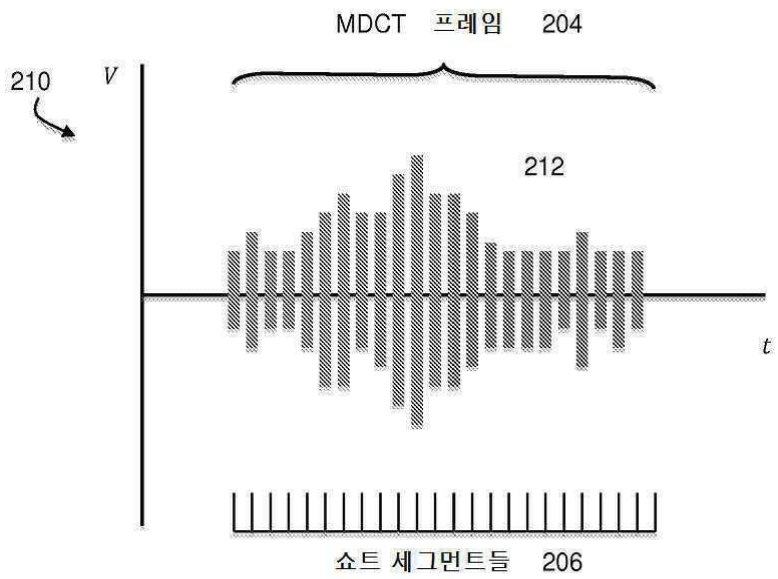
도면1



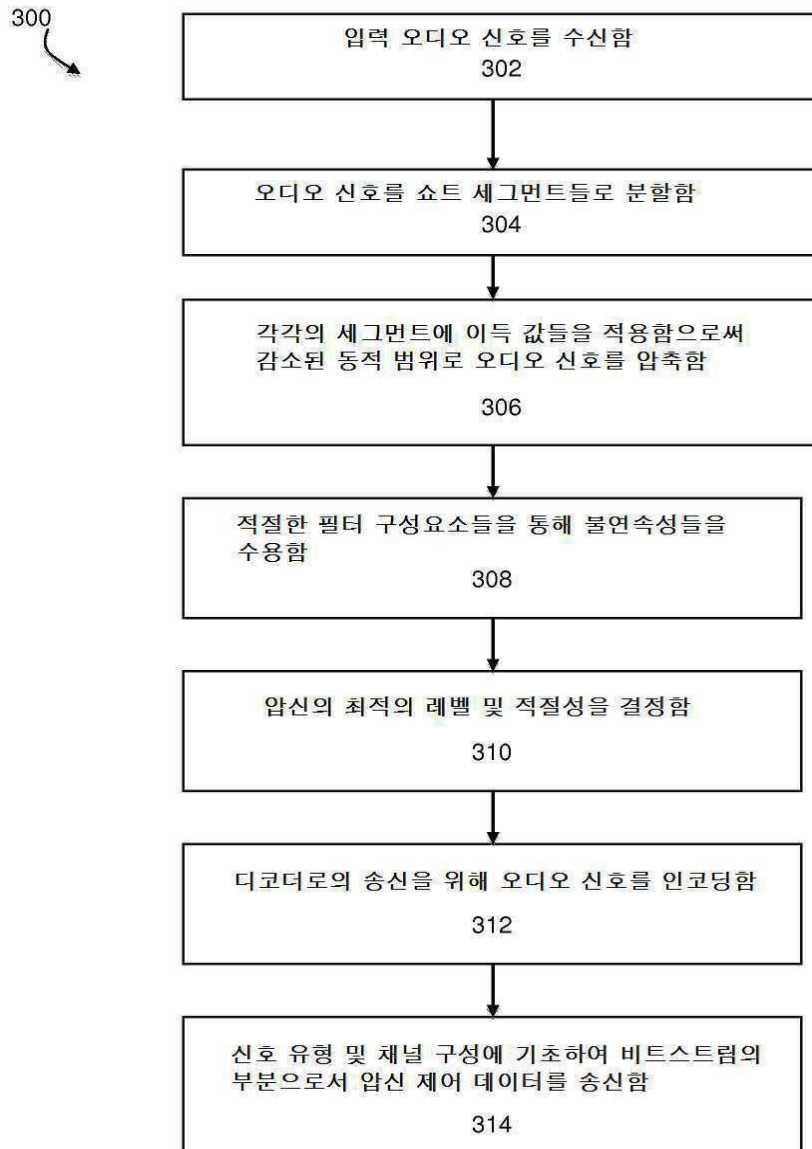
도면2a



도면2b

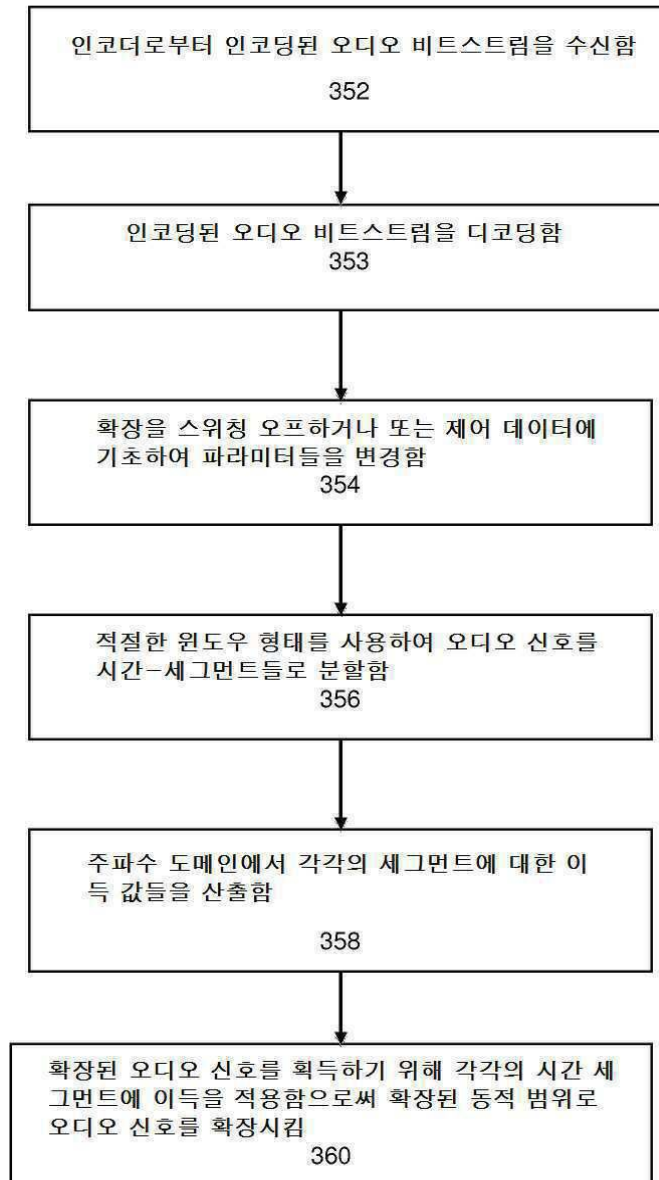


도면3a

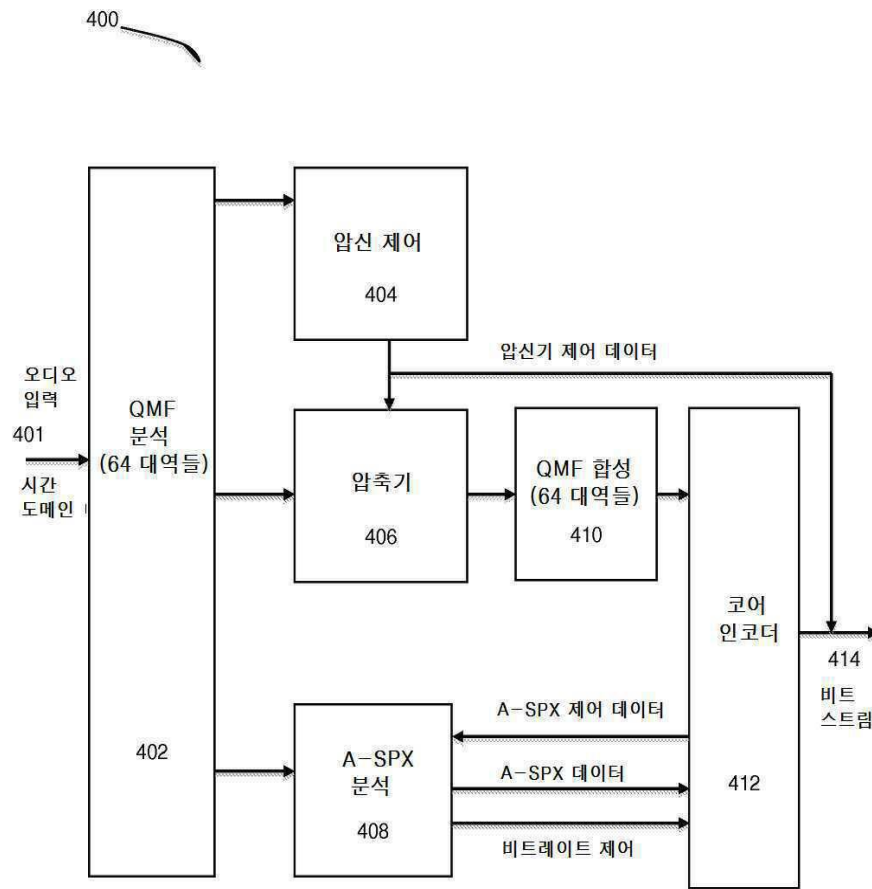


도면3b

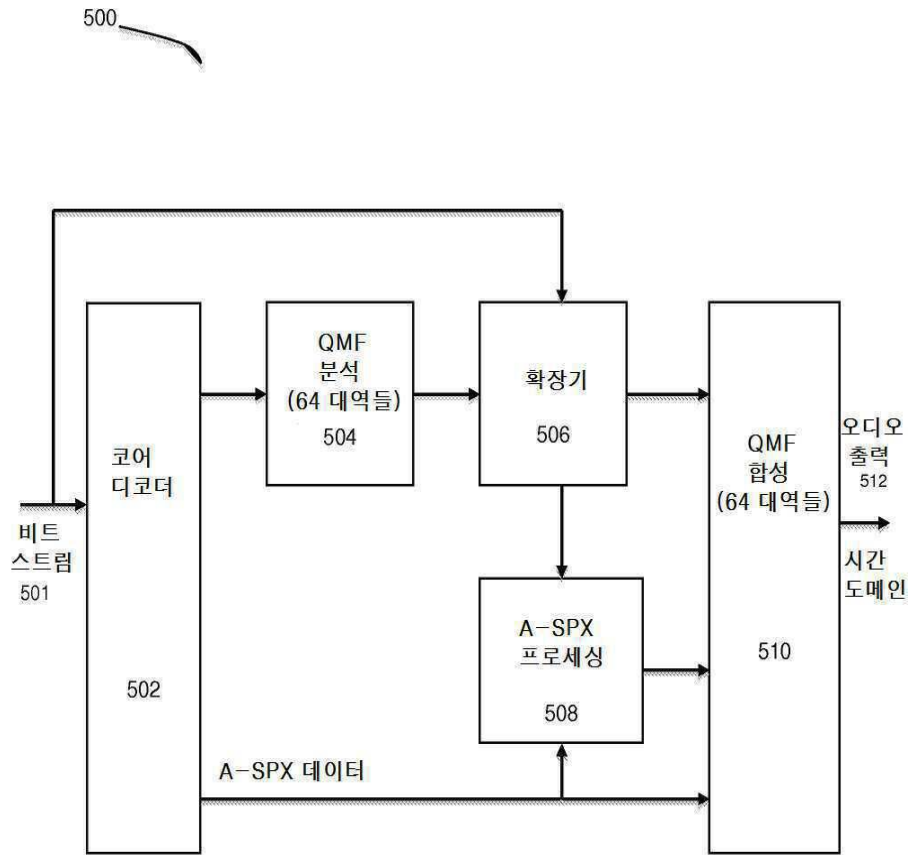
350

도면4



도면5



도면6

600

