



(19) 대한민국특허청(KR)
 (12) 등록특허공보(B1)

(45) 공고일자 2012년07월03일
 (11) 등록번호 10-1161703
 (24) 등록일자 2012년06월26일

- (51) 국제특허분류(Int. C1.)
H04S 3/02 (2006.01) *H04S 3/00* (2006.01)
H04B 1/20 (2006.01)
- (21) 출원번호 10-2007-7002358
- (22) 출원일자(국제) 2005년07월13일
 심사청구일자 2010년07월09일
- (85) 번역문제출일자 2007년01월30일
- (65) 공개번호 10-2007-0049146
- (43) 공개일자 2007년05월10일
- (86) 국제출원번호 PCT/US2005/024630
- (87) 국제공개번호 WO 2006/019719
 국제공개일자 2006년02월23일
- (30) 우선권주장
 10/911,404 2004년08월03일 미국(US)

(56) 선행기술조사문헌

- US5862228 A
 WO2003090208 A1
 WO2004019656 A1

전체 청구항 수 : 총 18 항

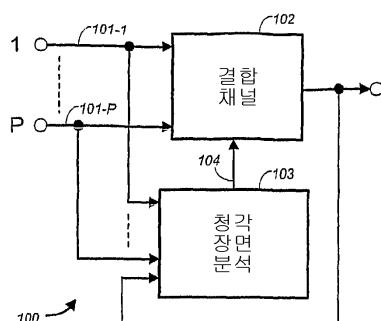
심사관 : 이정숙

(54) 발명의 명칭 청각 장면 분석을 사용한 오디오 신호의 결합

(57) 요 약

오디오 채널을 결합하는 방법은 이 오디오 채널을 결합하여 결합된 오디오 채널을 발생시키고 시간, 위상, 및 채널에 대한 진폭 또는 전력 조정 중 하나를 결합된 채널 또는 채널 둘다 및 결합된 채널에 동적으로 적용하는 것이다. 이 조정들 중 하나 이상은 채널들 중 하나 이상 및/또는 결합된 채널에서 청각 이벤트 측정에 의해 적어도 부분적으로 제어된다. 애플리케이션은 시네마 및 차량의 다채널 오디오를 제공하는 것을 포함한다. 방법뿐만 아니라 대응하는 컴퓨터 프로그램 구현방식 및 장치 구현방식이 포함된다.

대 표 도 - 도1



특허청구의 범위

청구항 1

오디오 채널을 결합하는 방법에 있어서,

결합된 오디오 채널을 생성하기 위하여 상기 오디오 채널을 결합하는 단계; 및

상기 채널, 상기 결합된 채널 또는 상기 채널과 상기 결합된 채널 모두에 대한 시간, 위상, 및 진폭 또는 전력 조정 중 하나 이상을 동적으로 적용하는 단계로서, 상기 조정 중 하나 이상은, 상기 조정이 청각 이벤트(auditory event) 동안 실질적으로 일정하게 유지되고 청각 이벤트 경계 또는 그 부근에서 변하도록 상기 채널 및/또는 상기 결합된 채널 중 하나 이상에서 적어도 부분적으로 청각 이벤트의 측정에 의해 제어되는, 상기 적용하는 단계;를 포함하며,

임계값(threshold)을 넘는 채널에서 시간에 대한 신호 특징의 변화에 반응하여 각각의 청각 이벤트 경계가 확인되어 상기 채널에 대해 한 세트의 청각 이벤트 경계가 얻어지며, 상기 청각 이벤트 경계는 하나의 청각 이벤트의 종료 및 다른 청각 이벤트의 시작을 나타내는 것을 특징으로 하는 오디오 채널 결합 방법.

청구항 2

P 오디오 채널을 Q 오디오 채널로 다운믹싱(downmixing)하는 방법으로서,

P는 Q보다 크고, 상기 Q 오디오 채널 중 적어도 하나는 제 1항의 방법에 의해 얻어지는, 다운믹싱 방법.

청구항 3

제 1항에 있어서, 상기 신호 특징은 스펙트럼 컨텐트인, 오디오 채널 결합 방법.

청구항 4

제 3항에 있어서, 채널에서 청각 이벤트 경계를 확인하는 단계는,

상기 오디오 신호를 시간 블록으로 분할하는 단계와, 각 블록의 데이터를 주파수 영역(frequency domain)으로 변환하는 단계로서, 주파수 영역으로 변환된 데이터는 계수(coefficient)로 표시되는, 변환하는 단계와, 상기 채널에서 상기 오디오 신호의 연속적인 시간 블록 사이에서 스펙트럼 컨텐트의 변화를 감지하는 단계를 포함하는, 오디오 채널 결합 방법.

청구항 5

제 4항에 있어서, 상기 채널에서 청각 이벤트 경계를 확인하는 단계는 블록의 계수와 인접 블록의 해당 계수를 비교하는 단계를 포함하는, 오디오 채널 결합 방법.

청구항 6

제 5항에 있어서, 단일 차이 측정치(single difference measure)는, 현재 및 상기 현재의 바로 이전 블록 스펙트럼들의 로그 표현된 해당 스펙트럼 값들(logarithmically-expressed corresponding spectral values)의 차이의 절대값들을 합산하고, 상기 단일 차이 측정치를 임계값과 비교하여 계산되는, 오디오 채널 결합 방법.

청구항 7

제 6항에 있어서, 청각 이벤트 경계는 상기 합산된 결과가 상기 임계값을 초과할 때 확인되는, 오디오 채널 결합 방법.

청구항 8

3개의 입력 오디오 채널 α , β , 및 δ 를 2개의 출력 오디오 채널 α "와 δ "로 다운믹싱하는 방법으로서,

상기 3개의 입력 오디오 채널은 순서대로 연속적인 공간 방향 α , β , 및 δ 를 나타내고 상기 2개의 출력 채널 α "와 β "는 비연속적인 공간 방향 α 와 δ 를 나타내는 다운믹스 방법에 있어서,

3개의 중간 채널을 생성하기 위해 방향 α 와 δ 를 나타내는 상기 2개의 입력 오디오 채널로부터 공통 신호 성분을 추출하는 단계로서, 상기 3개의 중간 채널은,

입력 채널 α 와 δ 에 공통인 신호 성분이 실질적으로 제거되는 채널 α 의 신호 성분을 포함하는 방향 α , 채널 α' 를 나타내는 채널 α 의 변형인 채널 α' ,

입력 채널 α 와 δ 에 공통인 신호 성분이 실질적으로 제거되는 채널 δ 의 신호 성분을 포함하는 방향 δ , 채널 δ' 를 나타내는 채널 δ 의 변형인 채널 δ' ,

입력 채널 α 와 δ 에 공통인 신호 성분을 포함하는 방향 β , 채널 β' 를 나타내는 새로운 채널인 채널 β' 인, 상기 추출 단계와,

출력 채널 α'' 을 생성하기 위해 중간 채널 α' , 중간 채널 β' , 입력 채널 β 를 결합하는 단계와,
출력 채널 δ'' 을 생성하기 위해 중간 채널 δ' , 중간 채널 β' , 입력 채널 β 를 결합하는 단계를
포함하는, 다운믹싱 방법.

청구항 9

제 8항에 있어서, 중간 채널 α' , β' , δ' 와 입력 채널 β 중 하나 이상, 및/또는 결합된 출력 채널 α'' 과
 δ'' 중 하나 또는 둘 모두에, 시간, 위상, 및 진폭 또는 전력 조정 중 하나 이상을 동적으로 적용하는 단계를
더 포함하는, 다운믹싱 방법.

청구항 10

제 9항에 있어서, 상기 조정 중 하나 이상은, 상기 조정이 청각 이벤트 동안 실질적으로 일정하게 유지되고
청각 이벤트 경계 또는 그 부근에서 변하도록 상기 입력 채널, 상기 중간 채널, 및/또는 상기 결합된 출력 채널
중 하나 이상의 채널에서 적어도 부분적으로 청각 이벤트의 측정에 의해 제어되고, 임계값을 넘는 채널에
서 시간에 대한 신호 특징의 변화에 반응하여 각각의 청각 이벤트 경계가 확인되어 상기 채널에 대해 한 세트
의 청각 이벤트 경계가 얻어지며, 상기 청각 이벤트 경계는 하나의 청각 이벤트의 종료 및 다른 청각 이벤트
의 시작을 나타내는 것을 특징으로 하는 다운믹싱 방법.

청구항 11

제 8항에 있어서, 상기 연속적인 공간 방향 α , β , 및 δ 는,

좌, 중앙, 및 우,

좌, 좌중간, 및 중앙,

중앙, 우중간, 및 우,

우, 우중간, 및 우 주변,

우 주변, 중앙 뒤, 및 좌 주변,

좌 주변, 좌중간, 및 좌의

방향 세트 중 한 방향인, 다운믹싱 방법.

청구항 12

제 1항 또는 제 8항 또는 제 10항 중 어느 한 항에 기재된 방법을 수행하는데 적합한 장치.

청구항 13

제 1항 또는 제 8항 또는 제 10항 중 어느 한 항에 기재된 방법을 수행하는 프로그램을 저장하는 컴퓨터 판독
가능한 기록 매체.

청구항 14

제 10항에 있어서, 상기 신호 특징은, 스펙트럼 컨텐트인, 다운믹싱 방법.

청구항 15

제 14항에 있어서, 채널에서 청각 이벤트 경계를 확인하는 단계는, 상기 오디오 신호를 시간 블록으로 분할하

는 단계와, 각 블록의 데이터를 주파수 영역으로 변환하는 단계로서, 주파수 영역으로 변환된 데이터는 계수로 표시되는, 변환하는 단계와, 상기 채널에서 상기 오디오 신호의 연속적인 시간 블록 사이에서 스펙트럼 컨텐트의 변화를 감지하는 단계를 포함하는, 다운믹싱 방법.

청구항 16

제 15항에 있어서, 상기 채널에서 청각 이벤트 경계를 확인하는 단계는 블록의 계수와 인접 블록의 해당 계수를 비교하는 단계를 포함하는, 다운믹싱 방법.

청구항 17

제 16항에 있어서, 단일 차이 측정치는, 현재 및 상기 현재의 바로의 이전 블록 스펙트럼들의 로그 표현된 해당 스펙트럼 값들의 차이의 절대값들을 합산하고, 상기 단일 차이 측정치를 임계값과 비교하여 계산되는, 다운믹싱 방법.

청구항 18

제 17항에 있어서, 청각 이벤트 경계는 상기 합산된 결과가 상기 임계값을 초과할 때 확인되는, 다운믹싱 방법.

명세서

기술 분야

[0001]

본 발명은 채널의 오디오 채널들 중 일부가 결합되는 다중채널 오디오 신호의 채널들의 수를 변화시키는 것에 관한 것이다. 애플리케이션들은 시네마들 및 차량들에서의 다중채널 오디오의 프리젠테이션(presentation)을 포함한다. 본 발명은 방법뿐만 아니라, 대응하는 컴퓨터 프로그램 구현방식들 및 장치 구현방식들을 포함한다.

배경 기술

[0002]

최근 수십 년 동안, 다중채널 오디오 재료의 제작, 분배 및 프리젠테이션이 꾸준히 상승되고 있다. 이러한 상승은 5.1 채널 재생 시스템이 거의 유비쿼터스한 필름 산업에 의해, 그리고, 더 최근에는, 5.1 다중채널 음악을 제작하기 시작하고 있는 음악 산업에 의해 상당히 추진되었다.

[0003]

전형적으로, 이와 같은 오디오 재료는 상기 재료와 동일한 수의 채널들을 가지는 재생 시스템을 통하여 제공된다. 예를 들어, 5.1 채널 필름 사운드트랙은 5.1 채널 시네마에서 또는 5.1 채널 홈 씨어터 오디오 시스템을 통하여 제공될 수 있다. 그러나, 오디오 재료의 채널들의 수와 동일한 수의 프리젠테이션 채널을 갖지 않는 환경들에서 또는 시스템들을 통하여 다중채널 재료를 재생하는 것-예를 들어, 2개 또는 4개의 재생 채널만을 가지는 차량에서의 5.1 채널 재료를 재생하는 것, 또는 5.1 채널 시스템을 가지고 있는 시네마에서의 5.1 이상의 채널 영화 사운드트랙들을 재생하는 것이 점점 더 요구되고 있다. 이와 같은 상황들에서, 프리젠테이션을 위한 멀티채널 신호의 일부 또는 모든 채널들을 결합하거나 "다운믹싱"하는 것이 필요로된다.

[0004]

채널들을 결합하면은 가청 인공물들이 생성될 수 있다. 예를 들어, 어떤 주파수 성분들은 삭제될 수 있는 반면, 다른 주파수 성분들은 강화되거나, 점점 소리가 커진다. 가장 통상적으로, 이는 결합되고 있는 채널들 중 2개 이상에서 유사하거나 상관된 오디오 신호 성분들이 존재하는 결과이다.

발명의 상세한 설명

[0005]

본 발명의 목적은 채널들을 결합하는 결과로서 발생하는 인공물들을 최소화하거나 억제하는 것이다. 다른 목적들은 본 문서를 관독하여 이해할 때 인식될 것이다.

[0006]

채널들을 결합하는 것이 단지 채널들의 수의 감소를 위한 것이 아니라, 다른 목적들을 위해 필요로될 수 있다. 점에 유의해야 한다. 예를 들어, 다중채널 신호에서 원래 채널들 중 2개 이상의 어떤 결합인 부가적인 재생 채널을 생성하는 것이 필요로될 수 있다. 이는 그 결과가 채널들의 원래 수보다 더 많다는 점에서 "업믹싱"의 유형으로 특징지어질 수 있다. 따라서, "다운믹싱"의 상황이든 또는 "업믹싱"의 상황이든지 간에, 채널들을 결합하여 부가적인 채널을 생성하면 가청 인공물들이 야기될 수 있다.

[0007]

믹싱 또는 채널-결합 인공물들을 최소화하기 위한 공통 기술들은 예를 들어, 결합될 채널, 결과적인 결합된

채널, 또는 이들 둘 다에 시간, 위상 및 진폭(또는 전력) 조정들 중 하나 이상을 적용하는 것을 포함한다. 오디오 신호들은 본래 동적이다-즉, 이들의 특성은 시간에 걸쳐 변화한다. 따라서, 오디오 신호에 대한 이와 같은 조정들은 전형적으로 동적인 방식으로 계산되고 적용된다. 결합에 기인하는 어떤 인공물들을 제거하는 동안, 이와 같은 동적 프로세싱은 다른 인공물들을 도입할 수 있다. 이와 같은 동적 프로세싱 인공물들을 최소화하기 위하여, 본 발명은 청각 장면 분석을 사용하여, 일반적으로, 청각 장면들 또는 이벤트들 동안 동적 프로세싱 조정들이 실질적으로 일정하게 유지되고 이와 같은 조정의 변화들이 청각 장면 또는 이벤트 경계들에서 또는 그 부근에서만 허용되도록 한다.

[0008]

청각 장면 분석

[0009]

사운드들을 개별적으로 인식되는 유닛들로 분할하는 것을 종종 "청각 이벤트 분석" 또는 "청각 장면 분석"("ASA")이라 칭한다. 청각 장면 분석의 광범위한 논의는 Albert S. Bregman에 의한 책 Auditory Scene Analysis-The Perceptual Organization of Sound, Massachusetts Institute of Technology, 1991, 제4쇄, 2001, 제2 MIT 출판부 판에 설명되어 있다.

[0010]

청각 장면 분석의 양상들에 따른 (이벤트 경계들을 포함한) 청각 이벤트들을 식별하는 기술들은 2002년 12월 5일자에 WO02/097792호로서 공개되고, 미국을 지정하여 2002년 2월 2일자로 출원된 국제 출원 PCT/US02/05999에 기인하는 미국 국내 출원인 2003년 11월 20일자로 출원되고 명칭이 "Segmenting Audio Signals into Auditory Events"(대리인 문서 번호 DOL098US)인 Brett G. Crockett의 미국 특허 출원 일련 번호 제10/478,583호에 설명되어 있다. 상기 출원들은 전반적으로 본원에 참조되어 있다. 상기 Crockett 출원들의 청각 이벤트 식별 기술들을 어떤 적용분야들은 2002년 12월 5일자에 WO02/097790호로서 공개되고, 미국을 지정하여 2002년 2월 22일자로 출원된 국제 출원 PCT/US02/05329에 기인하는 미국 국내 출원인 2003년 11월 20일자로 출원되고 명칭이 "Comparing Audio Using Characterizations Based on Auditory Events"(대리인 문서 번호 DOL092US)인 Brett G. Crockett 및 Michael J. Smithers의 미국 특허 출원 일련 번호 제10/478,397호, 및 2002년 12월 5일자에 WO02/097791호로서 공개되고, 미국을 지정하여 2002년 2월 25일자로 출원된 국제 출원 PCT/US02/05806에 기인하는 미국 국내 출원인 US2004/0148159A1로서 2004년 7월 24일자로 공개된, 2003년 11월 20일자로 출원되며 명칭이 "Method for Time Aligning Audio Signals Using Characterizations Based on Auditory Events"(대리인 문서 번호 DOL09201US)인 Brett G. Crockett 및 Michael J. Smithers의 미국 특허 출원 일련 번호 제10/478,398호에 설명되어 있다. 상기 Crockett 및 Smithers 출원들 각각은 또한 전반적으로 본원에 참조되어 있다.

[0011]

상기 Crockett 및 Smithers 출원들에 설명된 기술들이 특히 본 발명의 양상들과 관련하여 유용할지라도, 청각 이벤트들 및 이벤트 경계들을 식별하는 다른 기술들이 본 발명의 양상들에서 사용될 수 있다.

[0012]

본 발명의 양상을 따르면, 오디오 채널들을 결합하는 방법은 결합된 오디오 채널을 생성하기 위하여 오디오 채널을 결합하는 단계, 및 상기 채널들, 상기 결합된 채널들, 또는 상기 채널들 및 상기 결합된 채널 둘 다에 시간, 위상, 및 진폭 또는 전력 조정들 중 하나 이상을 동적으로 적용하는 단계를 포함하며, 상기 조정들 중 하나 이상은 적어도 부분적으로 상기 채널들 및/또는 상기 결합된 채널 중 하나 이상에서 청각 이벤트들의 측정에 의해 제어된다. 상기 조정은 오디오 이벤트들 동안 실질적으로 일정하게 유지되고 오디오 이벤트 경계들에서 또는 그 부근에서 변화들을 허용하도록 제어될 수 있다.

[0013]

본 발명의 주요 목적은 결합된 오디오 신호들의 사운드 품질을 개선시키는 것이다. 이는 예를 들어, 오디오 신호들에 대한 시간, 위상 및/또는 진폭(또는 전력) 수정을 다양하게 수행하고, 적어도 부분적으로 청각 장면 분석 정보의 측정치로 이와 같은 수정들을 제어함으로써 성취될 수 있다. 본 발명의 양상들을 따르면, 오디오 신호들에 적용된 조정들은 일반적으로 오디오 이벤트 동안 비교적 일정하게 유지되고, 오디오 이벤트들 간의 전이들 또는 경계들에서 또는 그 부근에서 변화하도록 허용될 수 있다. 물론 이와 같은 조정들은 매 경계마다 맴挛 자주 발생할 필요는 없다. 이와 같은 조정들의 제어는 각 채널의 청각 이벤트 정보에 응답하여 채널마다 성취될 수 있다. 대안적으로, 이와 같은 조정들의 일부 또는 모두는 모든 채널들 또는 모든 채널보다 적은 채널에 걸쳐 결합되었던 청각 이벤트 정보에 응답하여 성취될 수 있다.

[0014]

본 발명의 다른 양상들은 상술된 방법 및 이와 같은 방법의 컴퓨터 프로그램 구현방식들과 함께 본 출원에서 설명되는 다른 방법들을 수행하는 장치들 또는 디바이스들을 포함한다. 본 발명의 또 다른 양상들은 본 문서를 관독하여 이해할 때 인식될 수 있다.

실시예

[0023]

본 발명의 일반화된 실시예는 오디오 채널 결합기 및 결합 방법(100)이 도시되어 있는 도1에 도시된다. 다수의 오디오 입력 채널들(P개의 입력 채널들(101-1 내지 101-P))이 채널 결합기 또는 결합 기능("결합 채널들")(102) 및 청각 장면 분석기 또는 분석 기능("청각 장면 분석")(103)에 인가된다. 결합될 2개 이상의 입력 채널들이 존재한다. 채널들(1 내지 P)은 한 세트의 입력 채널들 중 일부 또는 모두를 구성할 수 있다. 결합 채널들(102)은 자신에게 인가된 채널들을 결합한다. 이와 같은 결합이 예를 들어, 선형적, 부가적 결합일 수 있을지라도, 결합 기술은 본 발명에 중요하지 않다. 자신에 인가된 채널들을 결합하는 것 외에, 결합 채널들(102)은 또한 결합될 채널들, 결과적인 결합된 채널, 또는 결합될 채널 및 결과적인 결합된 채널 둘 다에 시간, 위상, 및 진폭 또는 전력 조정들 중 하나 이상을 동적으로 적용한다. 이와 같은 조정들은 믹싱 또는 채널-결합 인공물들을 감소시킴으로써 채널 결합의 품질을 개선시키기 위하여 행해질 수 있다. 특정 조정 기술들은 본 발명에 중요하지 않다. 결합 및 조정에 적합한 기술들의 예들은 2004년 3월 1일자로 출원되고 명칭이 "Low Bit Rate Audio Encoding and Decoding in Which Multiple Channels Are Represented by a Monophonic Channel and Auxiliary Information"(대리인 문서 번호 DOL11501)인 Mark Franklin Davis의 미국 특허 출원 일련 번호 제60/549,368호, 2004년 6월 14일자로 출원되고 명칭이 "Low Bit Rate Audio Encoding and Decoding in Which Multiple Channels are Represented by a Monophonic Channel and Auxiliary Information"(대리인 문서 번호 DOL11502)인 Mark Franklin Davis, 등의 미국 특허 출원 일련 번호 제60/579,974호, 및 2004년 7월 14일자로 출원되고 명칭이 "Low Bit Rate Audio Encoding and Decoding in Which Multiple Channels are Represented by a Monophonic Channel and Auxiliary Information"(대리인 문서 번호 DOL11503)인 Mark Franklin Davis, 등의 미국 특허 출원 일련 번호 제60/588,256호에 설명되어 있다. Davis 및 Davis, 등의 상기 3개의 가출원들 각각은 전반적으로 본원에 참조되어 있다. 청각 장면 분석(103)은 예를 들어, 상술된 출원들 중 하나 이상에 설명된 기술에 따라서 또는 어떤 다른 적절한 청각 장면 분석기 또는 분석 방법에 의하여 청각 장면 정보를 도출한다. 적어도 청각 이벤트들간의 경계들의 위치를 포함해야 하는 이와 같은 정보(104)는 결합 채널(102)에 인가된다. 상기 조정들 중 하나 이상은 적어도 부분적으로 결합될 채널 및/또는 결과적인 결합된 채널 중 하나 이상에서의 청각 이벤트들의 측정치에 의해 제어된다.

[0024]

도2는 본 발명의 양상들을 구현하는 오디오 신호 프로세서 또는 프로세싱 방법(200)의 예를 도시한다. 결합되는 다수의 오디오 채널들(1-P)로부터의 신호들(101-1 내지 101-P)은 시간 및/또는 위상 수정 디바이스 또는 방법("시간 및 위상 수정")(202) 및 도1과 관련하여 서술된 바와 같은 청각 장면 분석 디바이스 또는 방법("청각 장면 분석")(103)에 인가된다. 채널들(1 내지 P)은 한 세트의 입력 채널 중 일부 또는 모두를 구성할 수 있다. 청각 장면 분석(103)은 청각 장면 정보(104)를 도출하여 이 정보를 도3과 관련하여 후술된 바와 같이, 결합될 채널들 각각에 개별적으로 시간 및/또는 위상 수정을 적용하는 시간 및 위상 수정(202)에 인가한다. 그 후, 수정된 채널(205-1 내지 205-P)은 채널들을 결합하여 단일 출력 채널(207)을 생성하는 채널 믹싱 디바이스 또는 방법("믹스 채널들")(206)으로 인가된다. 선택적으로, 믹스 채널들(206)은 또한 이하에 더 설명되는 바와 같이, 청각 장면 분석 정보(104)에 의해 제어될 수 있다. 도1 및 2의 예들에서와 같이 본 발명의 양상들을 구현하는 오디오 신호 방법 또는 프로세싱 방법은 채널들(1 내지 P) 중 다양한 채널들을 결합하여 하나 이상의 출력 채널을 생성할 수 있다.

[0025]

청각 장면 분석(103)(도1 및 2)

[0026]

청각 장면 분석 리서치를 통해 귀가 여러 상이한 청각 큐들(auditory cue)을 사용하여 인식된 청각 이벤트의 시작 및 종료를 식별한다는 것이 제시되었다. 상술된 출원들에서 가르치는 바와 같이, 가장 강력한 큐들 중 하나는 오디오 신호의 공간 컨텐트에서의 변화이다. 각 입력 채널에 대하여, 청각 장면 분석(103)은 규정된 시간 간격들에서 각 채널(1 내지 P)의 오디오에 대한 공간적인 분석을 수행하여 신호의 주파수 표현들의 시퀀스를 생성한다. 상술된 출원들에서 설명된 방식으로, 연속적인 표현들은 임계값보다 더 큰 공간적인 컨텐트에서의 변화를 찾아내기 위하여 비교될 수 있다. 이와 같은 변화를 찾아내면 거의 하나의 청각 이벤트의 종료 및 또 다른 청각 이벤트의 시작을 나타내는 연속적인 주파수 표현들의 쌍들 간의 청각 이벤트 경계가 표시된다. 각 입력 채널에 대한 청각 이벤트 경계들의 위치는 청각 장면 분석 정보(104)의 성분들로서 출력된다. 이것이 상술된 출원들에서 설명된 방식으로 성취될 수 있을지라도, 청각 이벤트들 및 이들의 경계들은 다른 적절한 기술들에 의해 측정될 수 있다.

[0027]

청각 이벤트들은 상기 이벤트에 걸쳐 실질적으로 일정하게 유지되는 특성들을 지닌 사운드의 유닛으로 인식된다. 본 발명의 실시예들에서 사용될 수 있는 것과 같은 시간, 위상 및/또는 진폭(또는 전력) 조정이 청각 이벤트 내에서 상당히 변하면, 이와 같은 조정의 이펙트(effect)는 들릴 수 있으며, 바람직하지 않은 인공물들을 구성할 수 있다. 이벤트에 걸쳐 조정을 일정하게 유지하고 이벤트 경계들에 상당히 가까운 조정들만을 변

화시킴으로써, 청각 이벤트의 유사성은 깨어지지 않고, 변화들은 본래 이벤트 경계를 나타내는 오디오 컨텐트에서의 더 현저한 변화들 사이에서 숨겨질 것이다.

[0028] 이상적으로, 본 발명의 양상들을 따르면, 채널 결합 또는 "다운믹싱" 파라미터는 청각 이벤트 경계들에서만 변화하게 되어, 이벤트 내에서 동적 변화들이 발생하지 않도록 해야 한다. 그러나, 청각 이벤트를 검출하는 실제 시스템들은 전형적으로 디지털 도메인에서 동작하는데, 이 도메인에서, 시간-도메인의 디지털 오디오 샘플들 블록들이 주파수 도메인으로 변환되어 청각 이벤트 경계들의 분해능이 상당히 대충적인 시간 분해능을 가지게 되는데, 이 분해능은 디지털 오디오 샘플들의 블록 길이와 관련된다. 실제 이벤트 경계들에 대한 유용한 근사치들 산출하기 위하여 (블록 길이 및 주파수 분해능간의 트레이드-오프(trade-off)로) 그 분해능이 선택되면, 즉, 여러들이 청취자가 인식할 수 없을 만큼 충분히 가까운 근사 경계들을 산출하면, 본 발명을 따른 동적 다운믹싱을 위하여, 공지되지 않은 실제 경계들이 아니라, 오히려 블럭 경계들에 의해 제공되는 근사값들을 사용하는 것이 적합하다. 따라서, Crockett의 상술된 출원들에서의 예에 따라서, 이벤트 경계들은 블록 길이의 절반 내, 또는 44.1 kHz 샘플링 레이트를 사용하는 시스템에서 512 샘플 블록 길이의 예에 대해서 약 5.8 밀리초로 결정될 수 있다.

[0029] 본 발명의 양상들의 실제 구현방식에서, 각 입력 채널은 이산 시간-도메인 오디오 신호이다. 이 이산 신호는 약 10.6 밀리초의 중첩 블록들로 분할될 수 있는데, 여기서 중첩은 약 5.3 밀리초이다. 48 kHz의 오디오 샘플레이트에 대해서, 이는 256개의 샘플들이 이전 블록과 중첩하는 512개의 샘플 블록들과 등가가 된다. 각 블록은 예를 들어, 해닝 윈도우를 사용하여 윈도우링될 수 있고 예를 들어, 이산 퓨리에 변환(속도를 위한 고속 퓨리에 변환으로서 구현됨)을 사용하여 주파수 도메인으로 변환될 수 있다. 데시벨(dB) 단위의 전력이 각 스펙트럼 값에 대해 계산되고 나서, 스펙트럼은 가장 큰 dB 스펙트럼 값으로 표준화된다. 중첩하지 않거나 부분적으로 중첩하는 블록들이 계산 비용을 감소시키기 위해 사용될 수 있다. 또한, 다른 윈도우 기능들이 사용될 수 있지만, 해닝 윈도우는 본 출원에 매우 적합한 것으로 밝혀졌다.

[0030] Crockett의 상술된 출원에 설명된 바와 같이, 현재 블록에 대한 표준화된 주파수 스펙트럼은 바로 이전 블록으로부터의 표준화된 스펙트럼과 비교되어 이들이 차이의 측정치를 획득할 수 있다. 명확하게는, 현재 및 현재의 바로 이전 스펙트럼들의 dB 스펙트럼 값들의 차이의 절대값들을 합산함으로써 단일 차이 측정치가 계산될 수 있다. 그 후, 이와 같은 차이 측정치들은 임계값과 비교될 수 있다. 차이 측정치가 임계값보다 크면, 이벤트 경계는 현재 및 현재의 바로 이전 블록 간에 표시되고, 그렇지 않으면, 이벤트 경계는 현재 및 현재의 바로 이전 블록 간에 표시되지 않는다. 이 임계값에 적절한 값은 (dB 단위의) 2500인 것으로 밝혀졌다. 따라서, 이벤트 경계들은 블록의 약 절반의 정확도 내에서 결정될 수 있다.

[0031] 이 임계값 방법은 각 서브대역이 별개의 상이한 측정치를 갖는 주파수 서브대역에 적용될 수 있다. 그러나, 본 발명의 상황에서, 시간적으로 임의의 순간에 하나의 이벤트에 초점을 맞추는 인간의 인식 능력을 고려하면 폴 대역폭 오디오를 토대로 한 단일 측정치로도 충분하다. 각 채널(1 내지 P)에 대한 청각 이벤트 경계 정보는 청각 장면 분석 정보(104)의 성분으로서 출력된다.

시간 및 위상 수정(202)(도2)

[0033] 시간 및 위상 수정(202)은 입력 채널들의 쌍들 간의 높은 상관성 및 시간 위상 차들을 기대한다. 도3은 시간 및 위상 수정(202)을 더 상세히 도시한 것이다. 후술된 바와 같이, 각 쌍의 한 채널은 기준 채널이다. 하나의 적절한 수정 검출 기술이 후술된다. 다른 적절한 수정 검출 기술들이 사용될 수 있다. 비-기준 채널 및 기준 채널 간에 높은 상관성이 존재할 때, 그 디바이스 또는 방법은 비-기준 채널의 위상 또는 시간 특성을 변경하여, 채널의 그 쌍의 결합에 기인하는 가청 채널-결합 인공물들을 감소시키거나 제거함으로써 채널들의 쌍 간의 위상 또는 시간 차들을 감소시키도록 시도한다. 이와 같은 인공물들 중 일부는 예로서 설명될 수 있다. 도5a는 백색 잡음 신호의 크기 스펙트럼을 도시한다. 도5b는 백색 잡음으로 이루어진 제1 채널의 동일한 백색 잡음 신호이지만 약 0.21 밀리초 만큼 시간적으로 지연되는 제2 신호와의 단순한 결합에 기인하는 크기 스펙트럼을 도시한다. 백색 잡음 신호의 지연되지 않은 버전 및 지연된 버전을 결합하면 통상적으로 콤 필터링(comb filtering)이라 칭하는 스펙트럼 쉐이핑(spectral shaping) 및 소거가 이루어지고, 각 입력 신호의 백색 잡음과 매우 상이한 소리가 들리게 된다.

[0034] 도3은 위상 또는 시간 지연들을 제거하는 적절한 디바이스 또는 방법(300)을 도시한 것이다. 각각의 입력 오디오 채널로부터의 신호들(101-1 내지 101-P)은 각 채널에 대한 지연-표시 신호(302)를 출력하는 지연 계산 디바이스 또는 방법("계산 지연")(301)에 인가된다. 각 채널(1 내지 P)에 대한 성분을 가질 수 있는 청각 이벤트 경계 정보(104)는 각 채널이 출력 채널들(306-1 내지 306-P)을 생성하기 위하여 지연 디바이스들 또는

기능들("지연")(305-1 내지 305-P)에 의해 각각 사용되는 갱신 지연 신호(304-1 내지 304-P)를 조건적으로 갱신하기 위하여 일시적인 메모리 디바이스 또는 방법("유지")(303)을 포함하는 디바이스 또는 프로세서에 의해 사용된다.

[0035] 계산 지연들(301)(도3)

[0036] 계산 지연들(301)은 입력 채널들의 쌍들간의 상대적인 지연을 측정한다. 바람직한 방법은 우선 입력 채널들 중에서 기준 채널을 선택하는 것이다. 이 기준은 고정되거나 시간에 걸쳐 변할 수 있다. 기준 채널이 변하도록 하면, 예를 들어, 사일런트 기준 채널의 문제가 극복된다. 기준 채널이 변하는 경우, 이 기준 채널은 예를 들어, 채널 소리의 세기(예를 들어 가장 큰 소리가 기준임)에 의해 결정될 수 있다. 상술된 바와 같이, 각 입력 채널에 대한 입력 오디오 신호들은 약 5.3 밀리초 만큼 중첩하는 약 10.6 밀리초 길이의 중첩 블록으로 분할될 수 있다. 48 kHz의 오디오 샘플 레이트에 대해서, 이는 256개의 샘플들이 이전 블록과 중첩하는 512개의 샘플 블록과 동가가 된다.

[0037] 비-기준 채널 및 기준 채널간의 지연은 임의의 적절한 교차-상관 방법을 사용하여 계산될 수 있다. 예를 들어, S_1 (길이 N_1)이 기준 채널로부터의 샘플들의 블록이고 S_2 (길이 N_2)가 비-기준 채널들 중 하나로부터 샘플들의 블록이라고 하자. 우선, 교차-상관 어레이($R_{1,2}$)를 계산하자.

수학식 1

$$R_{1,2}(l) = \sum_{n=-\infty}^{\infty} S_1(n) \cdot S_2(n-l) \quad l = 0, \pm 1, \pm 2, \dots \quad (1)$$

[0038] [0039] 교차-상관은 실행 시간을 감소시키기 위하여 표준 FFT 기반으로 한 기술들을 사용하여 수행될 수 있다. S_1 및 S_2 둘 다 길이가 유한하기 때문에, $R_{1,2}$ 의 비-제로 성분은 N_1+N_2-1 의 길이를 갖는다. $R_{1,2}$ 의 최대 요소에 대응하는 래그(lag)(1)는 S_2 와 관련된 S_1 의 지연을 나타낸다.

수학식 2

$$l_{peak} = l \text{ for } MAX[R_{1,2}(l)] \quad (2)$$

[0040] [0041] 이 래그 또는 지연은 어레이들(S_1 및 S_2)과 동일한 유닛들을 갖는다.

[0042] 현재 블록에 대한 교차-상관 결과는 1차 무한 임펄스 응답 필터를 사용하여 이전 블록으로부터의 결과와 시간-스무드되어 스무드된 교차 상관($Q_{1,2}$)을 생성한다. 다음 식은 필터 계산을 나타내며, 여기서 m 은 현재 블록을 나타내고 $m-1$ 은 이전 블록을 나타낸다.

수학식 3

$$Q_{1,2}(l, m) = \alpha \times R_{1,2}(l) + (1 - \alpha) \times Q_{1,2}(l, m-1) \quad l = 0, \pm 1, \pm 2, \dots \quad (3)$$

[0043] [0044] α 에 유용한 값은 0.1이라는 것으로 밝혀졌다. 교차-상관($R_{1,2}$)에 관해서, $Q_{1,2}$ 의 최대 요소에 대응하는 래그(1)는 S_2 와 관련된 S_1 의 지연을 나타낸다. 각각의 비-기준 채널에 대한 래그 또는 지연은 신호(302)의 신호 성분으로서 출력된다. 제로의 값이 또한 기준 채널의 지연을 나타내는 신호(302)의 성분으로서 출력될 수 있다.

[0045] 측정될 수 있는 지연 범위는 오디오 신호 블록 크기에 비례한다. 즉, 블록 크기가 클수록, 이 방법을 사용하여 측정될 수 있는 지연 범위가 더 커진다.

유지(303)(도3)

[0046] [0047] 모든 지연이 채널에 대한 ASA 정보(104)를 통해 표시될 때, 유지(303)는 302로부터의 채널에 대한 지연 값을 대응하는 출력 채널 지연 신호(304)로 복제한다. 모든 경계가 표시되지 않을 때, 유지(303)는 최종 지연 값(304)을 유지시킨다. 이 방식으로, 이벤트 경계들에서 시간 정렬 변화들이 발생되므로, 오디오 인공물들을 야기하지 않을 것이다.

[0048] 지연(305-1 내지 305-P)(도3)

[0049] 지연 신호(304)는 양 또는 음 중 하나일 수 있기 때문에, 디폴트에 의한 지연들(305-1 내지 305-P) 각각은 계산 지연들(301)에 의해 계산될 수 있는 절대 최대 지연 만큼 각 지연을 지연시키도록 구현될 수 있다. 따라서, 지연들(305-1 내지 305-P) 각각에서의 총 샘플 지연은 각각의 입력 지연 신호(304-1 내지 304-P) 더하기 디폴트 지연 량의 합이다. 이는 신호들(302 및 304)이 양 또는 음이 되도록 하는데, 여기서 음은 채널이 기준 채널과 관련된 시간에서 앞서있는 것을 나타낸다.

[0050] 입력 지연 신호들(304-1 내지 304-P) 중 어느 하나가 값을 변화시킬 때, 샘플들을 제거하거나 복제하는 것이 필요로 될 수 있다. 바람직하게는, 이는 가청 인공물을 야기하지 않는 방식으로 수행된다. 이와 같은 방법들은 중첩 및 크로스페이딩 샘플들(overlapping and crossfading sample)을 포함할 수 있다. 대안적으로, 출력 신호들(306-1 내지 306-P)이 필터뱅크(도4 참조)에 인가될 수 있기 때문에, 지연 및 필터뱅크를 결합하여 지연이 필터뱅크에 인가되는 샘플들의 정렬을 제어하도록 하는 것이 유용할 수 있다.

[0051] 대안적으로, 더 복잡한 방법은 개별적인 주파수 대역들 또는 주파수 대역들의 그룹들에서 시간 또는 위상 차들에 대해 측정하여 수정할 수 있다. 이와 같은 더 복잡한 방법에서, 계산 지연들(301) 및 지연들(305-1 내지 305-P)은 주파수 도메인에서 동작할 수 있고, 이 경우에, 지연들(305-1 내지 305-P)은 시간 도메인의 지연들 보다는 오히려, 대역들 또는 서브대역들에 대한 위상 조정들을 수행한다. 그 경우에, 신호들(306-1 내지 306-P)은 이미 주파수 도메인에 있어서, 후속 필터뱅크(401)(후술된 바와 같은 도4)가 필요하지 않게 한다.

[0052] 계산 지연(301) 및 청각 장면 분석(103)과 같은 디바이스들 및 방법들 중 일부는 이벤트들 내에서 적용될 시간 또는 위상 수정들 및 이벤트 경계들의 더 정확한 추정들을 제공하기 위하여 오디오 채널들에서 미리보기될 수 있다.

[0053] 믹스 채널들(206)(도2)

[0054] 도2의 믹스 채널들(206)의 세부사항들은 다운믹싱된 출력 채널을 생성하기 위하여 입력 채널들이 결합될 수 있는 방법을 나타내는 도4의 디바이스 또는 방법(400)으로서 도시되어 있다. 채널들을 믹싱 또는 결합하는 것 외에, 이 디바이스 및 방법은 도2의 시간 및 위상 수정(203)에 의해 완전하게 수정되지 않았던 잔여 주파수 소거들에 대해 수정할 수 있다. 이것은 또한 전력 보존을 유지하는 기능을 한다. 다른 말로서, 믹스 채널들(206)은 출력 다운믹스 신호(414)(도4)의 전력이 시간 또는 위상 조정된 입력 채널들(205-1 내지 205-P)의 합과 실질적으로 동일하게 되도록 한다. 더구나, 이것은 다운믹싱된 신호의 각 주파수 대역에서의 전력이 실질적으로 개별적인 시간 또는 위상 조정된 입력 채널들의 대응하는 주파수 대역들의 전력의 합이 되도록 할 수 있다. 상기 방법은 다운믹싱된 채널로부터의 대역 전력을 입력 채널들로부터의 대역 전력들과 비교하고 나서, 각 대역에 대한 이득 수정 값을 계산함으로써 이를 성취한다. 시간 및 주파수 둘 다에 걸쳐서 이득 조정들의 변화들이 가청 인공물을 야기할 수 있기 때문에, 이득들은 다운믹싱된 신호에 인가되기 전에 스무드되는 시간 및 주파수인 것이 바람직하다. 이 디바이스 또는 방법은 채널들을 결합하는 한 가지 가능한 방법을 나타낸다. 다른 적절한 디바이스들 또는 방법들이 사용될 수 있다. 특정 결합 디바이스 또는 방법은 본 발명에 중요하지 않다.

[0055] 필터뱅크("FB")(401-1 내지 401-P)(도4)

[0056] 각 입력 채널에 대한 입력 오디오 신호들은 시간-도메인 신호들이고 상술된 바와 같이, 약 5.3 밀리초 만큼 중첩하는 대략 10.6 밀리초 길이의 중첩 블록들로 분할될 수 있다. 48kHz의 오디오 샘플 레이트에 대해서, 이는 256개의 샘플들이 이전 블록과 중첩하는 512개의 샘플 블록들과 등가가 된다. 이 샘플 블록들은 윈도우ing되어 필터뱅크들(401-1 내지 401-P)(각 입력 신호에 대한 하나의 필터뱅크)에 의해 주파수 도메인으로 변환될 수 있다. 각종 윈도우 유형들 중 임의의 한 유형이 사용될 수 있을지라도, 해닝 윈도우는 적절한 것으로 밝혀졌다. 각종 시간-도메인 대 주파수-도메인 변환기들 또는 변환 방법들 중 임의의 한 변환기 또는 변환 방법이 사용될 수 있을지라도, 적절한 변환기 또는 변환 방법은 이산 퓨리에 변환(속도를 위한 고속 퓨리에 변환으로서 구현됨)을 사용할 수 있다. 각 필터뱅크의 출력은 복소 스펙트럼 값들의 각 어레이(402-1 내지 402-P)인데, 각 복소 스펙트럼 값은 각 주파수 대역(또는 빈)에 대한 것이다.

[0057] 대역("BND") 전력(403-1 내지 403-P)(도4)

[0058] 각 채널에 대해서, 대역 전력 계산기 또는 계산 방법("BND 전력")(403-1 내지 403-P) 각각은 복소 스펙트럼 값들(402-1 내지 402-P)의 전력을 계산하고 이들을 각 전력 스펙트럼(404-1 내지 404-P)으로서 출력한다. 각 채널로부터의 전력 스펙트럼 값들은 가산 결합기 또는 결합 기능(415)에서 합산되어 새로운 결합된 전력 스펙

트럼(405)을 생성한다. 각 채널로부터의 대응하는 복소 스펙트럼 값들(402-1 내지 402-P)은 또한 부가 결합기 또는 결합 기능(416)에서 합산되어 다운믹스 복소 스펙트럼(406)을 생성한다. 다운믹스 전력 스펙트럼(406)의 전력은 또 다른 전력 계산기 또는 계산 방법("BND 전력")(403)에서 계산되고 다운믹스 전력 스펙트럼(407)으로서 출력된다.

[0059] 대역("BND") 이득(408)(도4)

[0060] 대역 이득 계산기 또는 계산 방법(대역 이득(408))은 다운믹스 전력 스펙트럼(407)으로 전력 스펙트럼(405)을 나누어 전력 이득들 또는 전력 비들의 어레이를 생성하는데, 각 전력 이득 또는 전력 비는 각 스펙트럼 값에 대한 것이다. 다운믹스 전력 스펙트럼 값이 제로이면(전력 이득이 무한이 되도록 한다), 대응하는 전력 이득은 "1"로 설정된다. 그 후, 전력 이득의 제곱근은 진폭 이득(409)의 어레이를 생성하도록 계산된다.

[0061] 제한, 시간 및 주파수 스무드(410)(도4)

[0062] 리미터 및 스무더 또는 제한 및 스무딩 기능(유한, 시간 및 주파수 스무드)(410)는 적절한 이득 제한 및 시간/주파수 스무딩을 수행한다. 바로 위에서 논의된 스펙트럼 진폭 이득은 광범위를 가질 수 있다. 최적의 결과들은 이득들이 제한된 범위 내에서 유지되는 경우 획득될 수 있다. 예를 들어, 임의의 이득이 상위 임계값보다 크다면, 이는 상위 임계값과 동일하게 설정된다. 마찬가지로, 예를 들어, 임의의 이득이 하위 임계값보다 작다면, 이는 하위 임계값과 동일하게 설정된다. 유용한 임계값들은 0.5 및 2.0($\pm 6\text{dB}$ 와 동일)이다. 그 후, 스펙트럼 이득들은 1차 무한 임펄스 응답(IIR) 필터를 사용하여 시간적으로 스무드하게 될 수 있다. 다음 식은 필터 계산을 나타내는데, 여기서 b 는 스펙트럼 대역 지수를 나타내며, B 는 총 대역 수를 나타내며, n 은 현재 블록을 나타내며, $n-1$ 은 이전 블록을 나타내며, G 는 스무드되지 않은 이득을 나타내고, G_s 는 시간적으로 스무드한 이득을 나타낸다.

수학식 4

$$G_s(b, n) = \delta(b) \times G(b) + (1 - \delta(b)) \times G_s(b, n-1) \quad b = 0, \dots, B-1 \quad (4)$$

[0063]

[0064] $\delta(b)$ 에 대한 유용한 값은 대략 200Hz 보다 아래의 대역을 제외하면 0.5라는 것이 밝혀졌다. 이 주파수보다 아래에서, $\delta(b)$ 는 대역 $b=0$ 또는 DC에서 0의 최종 값을 향하는 경향이 있다. 스무드된 이득들(G_s)이 1.0으로 초기화되면, DC에서 값은 1.0과 동일하게 유지된다. 즉, DC는 결과 이득 조정되지 않을 것이고 200Hz 보다 아래의 대역들의 이득은 나머지 스펙트럼에서 대역들보다 더 느리게 변할 것이다. 이는 더 낮은 주파수들에서 가청 변조들을 방지하는데 유용할 수 있다. 이는 200Hz보다 낮은 주파수들에서 이와 같은 주파수들의 파장이 필터뱅크에 의해 사용되는 블록 크기에 접근하거나 초과하여 이를 주파수들을 정확하게 구별하는 필터뱅크의 성능면에서 부정확성을 야기하기 때문이다. 이는 보편적이고 널리 공지된 현상이다.

[0065] 시간적으로-스무드되는 이득들은 주파수에 걸쳐서 더욱 스무드하게 되어 인접 대역들 간의 이득의 큰 변화를 방지한다. 바람직한 구현방식에서, 대역 이득들은 슬라이드하는 5개의 대역(또는 대략 470Hz) 평균을 사용하여 스무드된다. 즉, 각 빈은 스스로 평균이 되도록 개신되고 2개의 인접 대역들 둘 다는 주파수에서 위와 아래에 있도록 한다. 스펙트럼의 상부 및 하부 에지에서, 에지 값들(대역들 0 및 N-1)은 반복적으로 사용되어, 5개의 대역 평균이 여전히 수행될 수 있도록 한다.

[0066] 스무드된 대역 이득들은 신호(411)로서 출력되고 승산기 또는 승산 기능(419)에서 다운믹스 복소 스펙트럼 값들과 승산되어 수정된 다운믹스 복소 스펙트럼(412)을 생성한다. 선택적으로, 출력 신호(411)는 ASA 정보(104)의 제어 하에서 일시적 메모리 디바이스 또는 방법("홀들")(417)을 통해서 승산기 또는 승산 기능(419)에 인가될 수 있다. 유지(417)는 도3의 유지(303)와 동일한 방식으로 동작된다. 예를 들어, 이득들은 이벤트 동안 상대적으로 일정하게 유지될 수 있고 이벤트 경계들에서만 변화될 수 있다. 이 방식으로, 이벤트 동안 가능한 가청 및 큰 이득 변화들은 방지될 수 있다.

[0067] 인버스 필터뱅크(Inv FB)(413)(도4)

[0068] 승산기 또는 승산 기능(419)으로부터의 다운믹스 스펙트럼(412)은 인버스 필터뱅크 또는 필터뱅크 기능("INV FB")(413)을 통하여 출력 시간 샘플들의 블록들을 생성한다. 이 필터뱅크는 입력 필터뱅크(401)의 인버스이다. 인접 블록들은 널리 공지된 바와 같이 이전 블록들과 중첩되고 이에 부가되어 출력 시간-도메인 신호(414)를 생성한다.

- [0069] 서술된 장치들은 포워드 필터뱅크(401)에서 원도우를 단위 신호가 시스템을 통해서 유지되도록 승산되는 2개의 원도우들(포워드에서 사용되는 한 원도우 및 인버스 필터뱅크에서 사용되는 한 원도우)로 분리시키는 관행을 배제하지 않는다.
- [0070] **다운믹싱 애플리케이션들**
- [0071] 본 발명의 양상을 따른 다운믹싱의 한 가지 애플리케이션은 자동차에서의 5.1 채널 컨텐트의 재생이다. 자동차들은 이와 같은 시스템의 좌, 우, 좌 서라운드 및 우 서라운드 채널들에 거의 대응하는 5.1 채널 컨텐트의 4개의 채널들만을 재생할 수 있다. 각 채널은 특정 채널과 관련된 지향성 정보를 재생하는데 적합한 것으로 간주되는 장소들에 위치되는 하나 이상의 확성기들로 향한다. 그러나, 자동차들은 통상적으로 이와 같은 5.1 재생 시스템에서 중앙 채널의 재생을 위하여 중앙 확성기 위치를 갖지 않는다. 이 상황을 수용하기 위하여, 중앙 채널 신호를 (예를 들어, 3dB 또는 6dB 만큼) 감쇠시키고 이를 좌 및 우 채널 신호들 각각과 결합하여 가상 중앙 채널을 제공한다. 그러나, 이와 같은 간단한 결합은 상술된 인공물을 야기한다.
- [0072] 이와 같은 단순한 결합을 적용하는 대신에, 본 발명의 양상들을 따른 채널 결합 또는 다운믹싱이 적용될 수 있다. 예를 들어, 도1의 장치 또는 도2의 장치는 2회 적용될 수 있는데, 한번은 좌 및 중앙 신호들을 결합하기 위한 것이고 다른 한번은 중앙 및 우 신호들을 결합하기 위한 것이다. 그러나, 중앙 채널 신호를 각 좌 채널 및 우 채널들과 결합시키기 전에, 이 중앙 채널 신호를 예를 들어, 3dB 또는 6dB(6dB는 자동차 내부의 니어-필드 공간에서 3dB보다 더 적합할 수 있다) 만큼 감쇠시켜 중앙 채널 신호로부터 출력 음향 전력이 전용 중앙 채널 스피커를 통해서 제공되는 경우와 거의 동일하게 되도록 하는데 유용할 수 있다. 더구나, 중앙 신호를 좌 채널 및 우 채널 신호들 각각과 결합시킬 때 이 중앙 신호를 기준 채널로서 나타내어 중앙 채널 신호가 인가되는 시간 및 위상 수정(103)이 중앙 채널의 시간 정렬 또는 위상을 변경시키는 것이 아니라 단지 좌 채널 및 우 채널 신호들의 시간 정렬 또는 위상만을 변경시키도록 하는데 유용할 수 있다. 결국, 중앙 채널 신호는 2개의 합산들 (즉, 좌 채널 더하기 중앙 채널 신호들 합산과 우 채널 더하기 중앙 채널 신호들 합산) 각각에서 다르게 조정되지 않으므로, 가상 중앙 채널 영상이 안정하게 유지되도록 한다.
- [0073] 이와 반대의 사항이 또한 적용될 수 있다. 즉, 시간 또는 위상은 중앙 채널만을 조정하여, 또다시 가상 중앙 채널 영상이 안정하게 유지되도록 한다.
- [0074] 본 발명의 양상들을 따른 다운믹싱의 또 다른 애플리케이션은 시네마에서 다중채널 오디오의 재생하는 것이다. 차세대 디지털 시네마 시스템을 위하여 개발 중인 표준들은 16개의 오디오 채널까지 및 그 이상으로의 전달을 필요로 한다. 대다수의 설치된 시네마 시스템들은 5.1 재생 또는 "프리젠테이션" 채널들(널리 공지된 바와 같이, "0.1"은 저 주파수 "이펙트" 채널을 표시한다)을 제공한다. 그러므로, 재생 시스템들이 상당한 비용을 들여 업그레이드될 때까지, 5.1보다 많은 채널들을 지닌 컨텐트를 5.1 채널들로 다운믹싱할 필요가 있다. 이와 같은 채널들의 다운믹싱 또는 결합은 상술된 바와 같은 인공물들을 야기한다.
- [0075] 그러므로, P 채널들이 Q 채널들(여기서, P>Q)로 다운믹싱되어야 한다면, 본 발명의 양상들을 따른 다운믹싱은 출력 채널들의 일부 또는 전부가 P 입력 채널들 중 각 채널의 2개 이상을 결합하는 Q 출력 채널들 중 하나 이상을 획득하도록 적용될 수 있다. 입력 채널이 하나 이상의 출력 채널에 결합되면, 도2에서 시간 및 위상 수정(202)이 결합되는 각 출력 채널에 대해서 이와 같은 입력 채널의 시간 정렬 또는 위상을 다르게 변경하지 않도록 이와 같은 채널을 기준 채널로서 나타내는데 유용할 수 있다.
- [0076] **대안들**
- [0077] 본원에 설명된 바와 같은 시간 또는 위상 정렬은 다운믹싱 동안 주파수들의 완전한 또는 부분적인 소거를 최소화하도록 작용한다. 입력 채널이 하나 이상의 출력 채널에 결합될 때, 이 채널은 기준 채널로서 표시되어 다수의 출력 채널들과 믹싱될 때, 시간 또는 위상을 다르게 조정하지 않도록 한다는 것이 앞서 설명되었다. 이는 다른 채널들이 실질적으로 동일한 컨텐트를 갖지 않을 때 양호하게 작동한다. 그러나, 2개 이상의 다른 채널들은 동일하거나 실질적으로 동일한 컨텐트를 갖는 상황들이 야기될 수 있다. 이와 같은 채널들이 하나 이상의 출력 채널에 결합되면, 결과적인 출력 채널들을 청취할 때, 공통 컨텐트는 이들 출력 채널들을 수용하는 확성기들의 물리적 위치들 간의 어느 곳에 있는 방향 내 공간에서 가상 영상으로서 인식된다. 이 문제는 실질적으로 동일한 컨텐트를 갖는 이들 2개 이상의 입력 채널들이 다른 채널들과 결합되기 전에 독립적으로 위상 조정되어 출력 채널들을 생성할 때 야기된다. 독립적인 위상 조정은 부정확한 가상 영상 위치 및/또는 결정되지 않은 영상 위치 둘 다를 야기할 수 있는데, 이들 둘 다는 들을 수 있지만 부자연스러운 것으로서 인식될 수 있다.

[0078] 실질적으로 유사한 컨텐트를 갖는 입력 채널들을 찾고 동일하거나 유사한 방식으로 이와 같은 채널들을 시간 또는 위상 조정하도록 시도하여 이들의 가상 영상 위치를 변경시키지 않도록 하는 시스템을 고안할 수 있다. 그러나, 이와 같은 시스템은 특히, 입력 채널들의 수가 실질적으로 출력 채널들의 수보다 크기 때문에, 매우 복잡하다. 실질적으로 유사한 컨텐트가 하나 이상의 입력 채널에서 빈번하게 발생되는 시스템들에서, 이는 위상 조정을 할 필요가 없고 단지 전력 수정만을 수행하도록 더 간단해질 수 있다.

[0079] 이 조정 문제는 중앙 채널 신호가 좌 및 우 확성기들 각각을 통해서 재생을 위한 좌 및 우 채널들 각각과 결합되는 상술된 자동차 애플레이션에서 더 설명될 수 있다. 5.1 채널 재료에서, 좌 및 우 입력 채널들은 종종 다수의 신호들(예를 들어, 인스트루먼트들, 보컬들, 다이얼로그 및/또는 이펙트들)을 포함하는데, 이들 중 일부는 상이하고 이들 중 일부는 동일하다. 중앙 채널이 좌 및 우 채널들 각각과 믹싱될 때, 중앙 채널은 기준 채널로서 표시되고 시간 또는 위상 조정되지 않는다. 좌 채널은 시간 또는 위상 조정되어, 중앙 채널과 결합될 때 최소 위상 소거를 발생시키고, 유사하게 우 채널은 시간 또는 위상 조정되어 중앙 채널과 결합될 때 최소 위상 소거를 발생시킨다. 좌 및 우 채널들이 독립적으로 시간 또는 위상 조정되기 때문에, 좌 및 우 채널들에 공통되는 신호들은 더 이상 좌 및 우 확성기들의 물리적 위치들 간에서 가상 영상을 갖지 않을 수 있다. 게다가, 가상 영상은 임의의 한 방향으로 국부화되는 것이 아니라, 청취 공간에 걸쳐서 확산될 수 있는데, 이는 부자연스럽고 바람직하지 않은 이펙트이다.

[0080] 조정 문제에 대한 해결책은 이와 같은 입력 채널들로부터 하나 이상의 입력 채널에 공통되는 신호들을 추출하고 이들을 새롭고 분리된 입력 채널들에 배치하는 것이다. 이는 다운믹싱될 전체 입력 채널들(P)의 수를 증가시키지만, 이는 출력 다운믹싱된 채널들에서 의사 및 바람직하지 않은 가상 영상 왜곡을 감소시킨다. 2개로 다운믹싱되는 3개의 채널들의 경우에 대한 자동차의 예의 디바이스 또는 방법(600)이 도6에 도시되어 있다. 좌 및 우 입력 채널들에 공통되는 신호들은 능동 매트릭스 디코더와 같은 임의의 적절한 채널 승산기 또는 승산 방법("무상관 채널들")(601) 또는 공통 신호 성분들을 추출하는 다른 유형의 채널 승산기를 사용하여 좌 및 우 채널들로부터 또 다른 새로운 채널로 추출된다. 이와 같은 디바이스는 무상관기 또는 무상관 기능의 유형을 특징으로 할 수 있다. Dolby Surround Pro Logic II로서 공지된 한 가지 적절한 능동 매트릭스는 "2000년 3월 22일자로 출원되고 명칭이 "Method for deriving at least three audio signals from two input audio signals"(대리인 문서 번호 DOL07201)인 James W.Fosgate의 미국 출원 일련 번호 제09/532,711호 및 2002년 7월 7일에 WO02/19768로서 공개되고, 미국을 지정하여 2001년 8월 30일자로 출원되며, 국제 출원 PCT/US01/27006에 기인하는 미국 국내 출원인 2004년 7월 1일에 US2004/0125960 A1으로서 공개된, 2003년 2월 25일자로 출원되고 명칭이 "Method for apparatus for audio matrix decoding"(대리인 문서 번호 DOL07203U S)인 James W. Fosgate, 등의 미국 특허 출원 일련 번호 제10/362,786호에 설명되어 있다. 상기 Fosgate와 Fosgate, 등의 출원들은 전반적으로 본원에 참조되어 있다. 사용될 수 있는 또 다른 유형의 적절한 채널 승산기 및 무상관기는 2003년 8월 7일에 WO02/063925호로서 공개되고, 미국을 지정하여 2002년 2월 7일자로 출원되며, 국제 출원 PCT/US02/03619호에 기인하는 미국 국내 출원인 2004년 4월 1일에 US2004/0062401로서 공개된, 2003년 5월 8일자로 출원되고 명칭이 "Audio Channel Translation"(대리인의 문서 번호 DOL088US)인 Mark Franklin Davis의 미국 특허 출원 일련 번호 제10/467,213호, 및 2004년 4월 4일에 WO2004/019656호로서 공개되며 미국을 지정하여 2003년 6월 8일자로 출원된 국제 출원 PCT/US03/24570호(대리인의 문서 번호 DOL08801PCT)에 설명되어 있다. Davis 출원들 각각은 전반적으로 본원에 참조되어 있다. 또 다른 적절한 채널 승산/무상관 기술은 독일 뮌헨에서의 2002년 5월 10일에서 13일까지의 112번째 컨벤션에서 제공된 오디오 엔지니어링 사회 컨벤션 논문 5529에서 Mitianoudis and Davies에 의한 "Intelligent Audio Source Separation using Independent Component Analysis"에 설명되어 있다. 상기 논문은 또한 전반적으로 본원에 참조되어 있다. 이 결과는 4개의 채널들, 즉 새로운 채널(C_D), 원래 중앙 채널(C) 및 변경된 좌 및 우 채널들(L_D 및 R_D)이 된다.

[0081] 도2의 장치를 토대로 하지만 2개의 출력 채널들을 갖는 이 디바이스 또는 방법(602)은 4개의 채널들을 결합하여 좌 및 우 재생 채널들(L_P 및 R_P)을 생성한다. 변경된 채널들(L_D 및 R_D) 각각은 단지 하나의 재생 채널(L_P 및 R_P) 각각에 믹싱된다. 이들은 실질적으로 임의의 상관된 컨텐트를 포함하지 않기 때문에, 공통 성분(C_D)을 추출하는 변경된 채널들(L_D 및 R_D)은 입력 채널들(L 및 R)에 존재하는 어떠한 가상 중앙 영상들에 영향을 주지 않고 시간 또는 위상 조정될 수 있다. 시간 및/또는 위상 조정을 수행하기 위하여, 채널(C_D)과 같은 채널들 중 한 채널은 기준 채널로서 표시된다. 그 후, 다른 채널들(L_D , R_D 및 C)은 기준 채널에 대해서 시간 및/또는 위상 조정된다. 대안적으로, L_D 및 R_D 채널들이 C채널과 상관되지 않고 이들이 방법(601)에 의해 C_D 채널로부터

터 무상관되기 때문에, 이들은 어떤 시간 또는 위상 조정 없이 채널들을 믹싱하도록 통과될 수 있다. 원래 채널(C) 및 도출된 중앙 채널(C_D) 둘 다는 재생 채널들(L_p 및 R_p)을 발생시키도록 디바이스 또는 방법(602)의 믹스 채널들 부분에 각 중간 채널들(L_D 및 R_D)과 믹싱될 수 있다. C 및 C_D 의 동일한 비례가 만족한 결과들을 발생시키도록 밝혀졌지만, 정확한 비례는 중요하지 않고 동일한 것 이외일 수 있다. 결과적으로, C_D 및 C에 적용되는 시간 위상 조정은 둘 다 재생 채널들에서 나타나므로, 가상 중앙 영상들의 방향을 유지할 것이다. 일부 감쇠(예를 들어, 3dB)는 중앙 채널들 각각 상에 필요로 될 수 있는데, 그 이유는 이들 채널들이 하나의 스피커가 아니라, 2개의 스피커들을 통해서 재생되기 때문이다. 또한 출력 채널들에 믹싱되는 중앙 채널들(C 및 C_D) 각각의 량은 청취자에 의해 제어될 수 있다. 예를 들어, 청취자는 원래 중앙 채널(C) 전부를 원할 수 있지만, 도출된 중앙 채널(C_D)상의 일부 감쇠를 원할 수 있다.

[0082] 이 해결책은 또한 시네마 오디오의 예로서 설명될 수 있다. 도7a 및 도7b는 2 세트의 오디오 채널들의 룸 또는 공간 위치를 도시한다. 도7a는 다중채널 오디오 신호에 존재하는 바와 같은 채널들의 대략적인 공간 위치들을 도시하는데, 그렇지 않다면 "컨텐트 채널들"로서 표시된다. 도7b는 5개의 채널 오디오 재료를 재생하도록 장착된 시네마에서 재생될 수 있는 "재생 채널들"로서 표시된 채널들의 대략적인 위치들을 도시한다. 컨텐트 채널들의 일부는 대응하는 재생 채널 위치들, 즉 L, C, R, R_s 및 L_s 채널들을 갖는다. 다른 컨텐트 채널들은 대응하는 재생 채널 위치들을 갖지 않으므로, 재생 채널들 중 하나 이상의 채널과 믹싱되어야만 한다. 전형적인 방법은 이와 같은 컨텐트 채널들을 가장근접한 2개의 재생 채널들과 결합시키는 것이다.

[0083] 상술된 바와 같이, 간단한 가산 결합은 가청 인공물을 야기할 수 있다. 또한 언급된 바와 같이, 도1 및 도2와 관련하여 서술된 바와 같은 결합은 또한 실질적으로 공통 컨텐트를 갖는 채널들이 다르게 위상 또는 시간 조정될 때 가상 활상 인공물들을 야기시킬 수 있다. 해결책은 이와 같은 입력 채널들로부터 하나 이상의 입력 채널에 공통되는 추출 신호들을 포함하고 이들을 새롭고 분리된 채널들에 배치하는 것이다.

[0084] 도7c는 디바이스 또는 방법("무상관 채널들")(701)을 사용하여 입력 또는 컨텐트 채널들의 일부 결합에 공통되는 정보를 추출함으로써 5개의 부가적인 채널들(Q_1 내지 Q_5)이 생성되는 경우에 대한 디바이스 또는 방법(700)을 도시한다. 디바이스 또는 방법(701)은 "무상관 채널들" 디바이스 또는 기능(601)에 사용하기 위하여 상술된 바와 같은 적절한 채널 승산/무상관 기술을 사용할 수 있다. 이들 부가적인 중간 채널들의 실제 수 및 공간 위치는 컨텐트 채널들에 포함되는 오디오 신호들의 변화들에 따라서 변할 수 있다. 도2의 배열을 토대로 하지만 5개의 출력 채널들을 갖는 이 디바이스 또는 방법(702)은 무상관 채널들(701)로부터 중간 채널들을 결합시켜 5개의 재생 채널들을 생성한다.

[0085] 시간 및 위상 수정을 위하여, C채널과 같은 중간 채널들 중 하나는 기준 채널로서 표시될 수 있고 모든 다른 중간 채널들은 이 기준에 대해서 시간 및 위상 조정될 수 있다. 대안적으로, 기준 채널로서 채널들 중 하나 이상을 표시하여 중간 채널들의 총수보다 적은 채널들의 그룹들에서 시간 또는 위상 수정을 수행하는 것이 유용할 수 있다. 예를 들어, 채널(Q_1)이 컨텐트 채널들(L 및 C)에서 추출되는 공통 신호들을 표시하고 Q_1 및 L_c 가 중간 채널들(L 및 C)와 결합되어 재생 채널들(L 및 C)를 생성하면, 채널(L_c)은 기준 채널로서 표시될 수 있다. 그 후, 중간 채널들(L, C 및 Q_1)은 기준 중간 채널(L_c)에 대해서 시간 또는 위상 조정된다. 중간 채널들의 각 더 작은 그룹은 모든 중간 채널들이 시간 및 위상 수정 방법에 의해 고려될 때까지 연속적으로 조정되는 시간 또는 위상이다.

[0086] 재생 채널들을 생성시, 디바이스 또는 방법(702)은 컨텐트 채널들의 공간 위치들의 선형적 지식을 추정할 수 있다. 부가적인 중간 채널들의 수 및 공간 위치에 관한 정보는 추정되거나 경로(703)를 통해서 무상관 디바이스 또는 방법(701)으로부터 디바이스 또는 방법(702)으로 통과될 수 있다. 이는 방법 또는 디바이스(702)가 부가적인 중간 채널들을 예를 들어 가장근접한 2개의 재생 채널들에 결합하도록 하여 이들 부가적인 채널들의 가상 영상 방향이 유지되도록 한다.

구현방식

[0088] 본 발명은 하드웨어 또는 소프트웨어로 또는 이들의 조합(예를 들어, 프로그램가능한 논리 어레이들)로 구현될 수 있다. 달리 규정되지 않은 한, 본 발명의 부분으로서 포함되는 이 알고리즘들은 임의의 특정 컴퓨터 또는 다른 장치에 원래 관련되지 않는다. 특히, 각종 범용 기계들은 본원에 기술 내용에 따라서 기록되는 프로그램들에 의해 사용될 수 있는데, 또는 이는 필요로 되는 방법 단계들을 수행하도록 더욱 특수화된 장치(예를 들어, 집적 회로들)를 구성하는데 더 편리할 수 있다. 따라서, 본 발명은 적어도 하나의 프로세서, 적어도 하

나의 데이터 저장 시스템(회발성 및 비회발성 메모리 및/또는 저장 소자들을 포함), 적어도 하나의 입력 디바이스 또는 포트 및 적어도 하나의 출력 디바이스 또는 포트를 각각 포함하는 하나 이상의 프로그램가능한 컴퓨터 시스템들 상에서 실행되는 하나 이상의 컴퓨터 프로그램들로 구현될 수 있다. 프로그램 코드가 적용되어 본원에 서술된 기능들을 수행하도록 데이터를 입력하여 출력 정보를 발생시킨다. 출력 정보는 공지된 방식으로 하나 이상의 출력 디바이스들에 인가된다.

[0089] 각 이와 같은 프로그램은 컴퓨터 시스템과 통신하도록 임의의 원하는 컴퓨터 언어(기계, 어셈블리, 또는 고래벨 절차적, 논리적 또는 대상 지향된 프로그래밍 언어를 포함)로 구현될 수 있다. 어쨌든, 언어는 컴파일되고 해석된 언어일 수 있다.

[0090] 각 이와 같은 컴퓨터 프로그램은 범용 또는 특수용 프로그램가능한 컴퓨터에 의해 관독될 수 있는 저장 매체 또는 디바이스(예를 들어, 고상 메모리 또는 매체, 또는 자기 또는 광학 매체)상에 저장되거나 다운로딩되어, 저장 매체 또는 디바이스가 본원에 서술된 절차들을 수행하도록 컴퓨터 시스템에 의해 관독될 때 컴퓨터를 구성하고 동작시킨다. 본 발명의 시스템은 또한 컴퓨터 프로그램으로 구성되는 컴퓨터 관독가능한 저장 매체로서 구현되는 것으로 간주될 수 있는데, 이와 같이 구성된 저장 매체는 컴퓨터 시스템이 특정 및 규정된 방식으로 동작되도록 하여 본원에 서술된 기능들을 수행한다.

[0091] 본 발명의 다수의 실시예들이 서술된다. 그럼에도 불구하고, 각종 변경들이 본 발명의 원리 및 범위를 벗어남이 없이 이루어질 수 있다는 것을 이해할 것이다. 예를 들어, 상술된 단계들 중 일부는 독립적으로 순서화될 수 있음으로, 서술된 바와 다른 순서로 수행될 수 있다. 따라서, 다른 실시예들이 이하의 청구범위 내에 있다.

도면의 간단한 설명

[0015] 도1은 본 발명의 일반화된 실시예의 기능적인 개략적 블록도이다.

[0016] 도2는 본 발명의 양상들을 구현하는 오디오 신호 방법 또는 프로세싱 방법의 기능적인 개략적 블록도이다.

[0017] 도3은 도2의 시간 및 위상 수정(202)을 더 상세히 도시한 기능적인 개략적 블록도이다.

[0018] 도4는 도2의 믹스 채널들(206)을 더 상세히 도시한 기능적인 개략적 블록도이다.

[0019] 도5a는 백색 잡음 신호의 크기 스펙트럼을 도시한 이상적인 응답이다. 도5b는 백색 잡음으로 이루어진 제1 채널의 동일한 백색 잡음 신호이지만 약 밀리초의 프랙션(fraction) 만큼 시간적으로 지연되는 제2 신호와의 단순한 결합에 기인하는 진폭 스펙트럼을 도시한 이상적인 응답이다. 도5a 및 5b 둘 다에서, 수평 축은 Hz의 주파수이며, 수직 축은 데시벨(dB)의 상대 레벨이다.

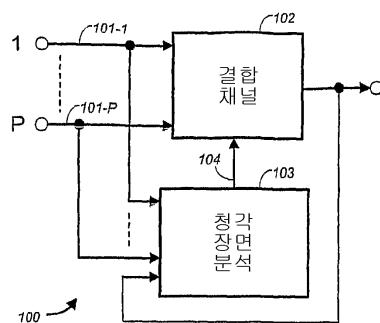
[0020] 도6은 본 발명의 양상을 따른 3 채널 대 2 채널 다운믹스의 기능적인 개략적 블록도이다.

[0021] 도7a 및 7b는 시네마 객석과 같은 룸(room) 내의 두 세트의 오디오 채널들의 공간적인 위치를 도시한 이상적인 표현들이다. 7a는 다중채널 오디오 신호의 "컨텐트" 채널들의 근사한 공간적인 위치들을 도시하는 반면, 도7b는 5-채널 오디오 재료를 재생할 준비가 된 극장 내의 "재생"의 근사한 공간적인 위치들을 도시한다.

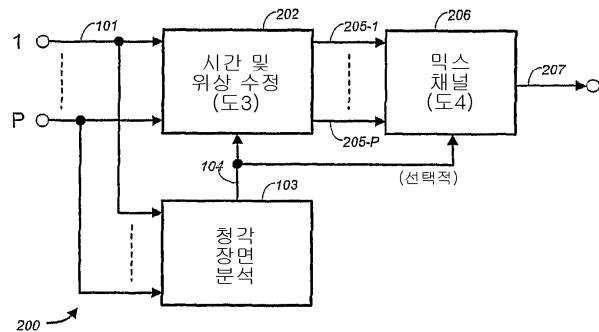
[0022] 도7c는 본 발명의 양상들을 따른 10 채널 대 5 채널 다운믹스의 기능적인 개략적 블록도이다.

도면

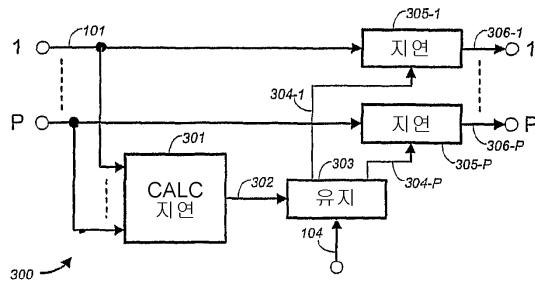
도면1



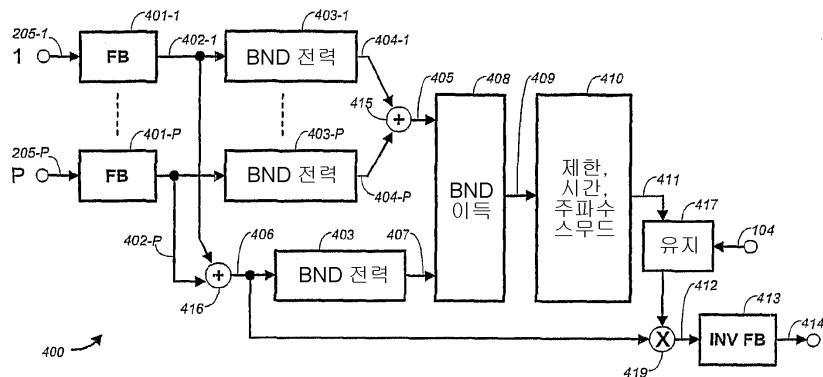
도면2



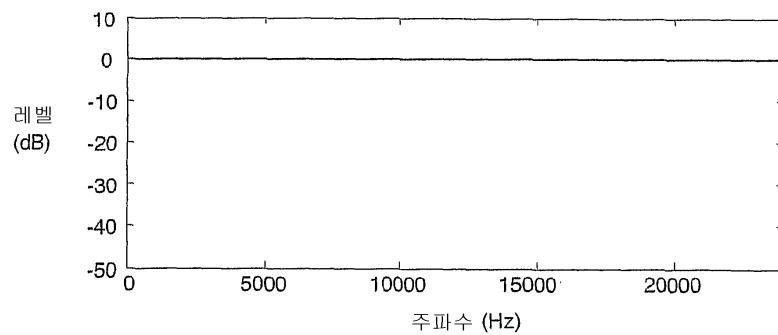
도면3



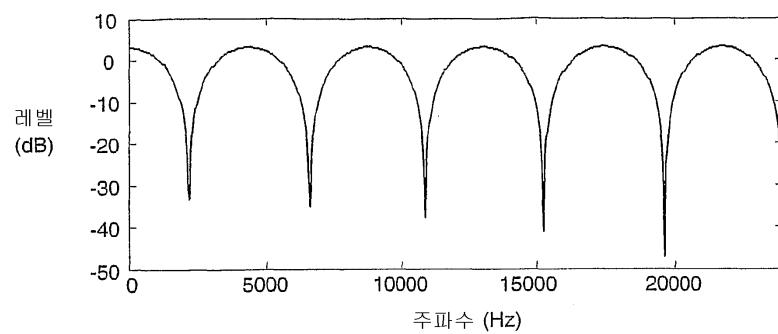
도면4



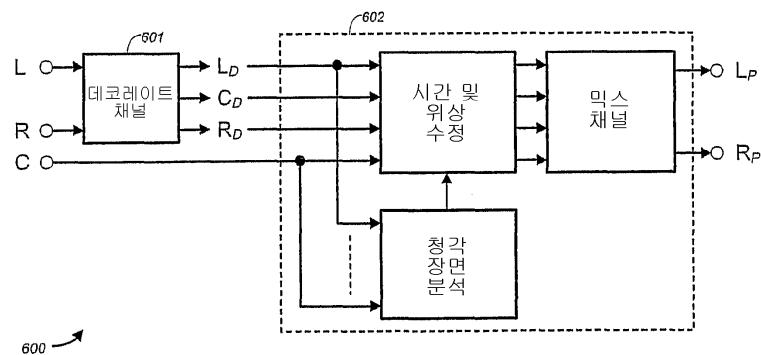
도면5a



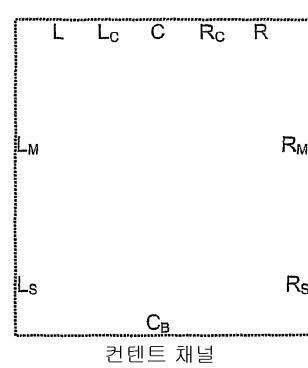
도면5b



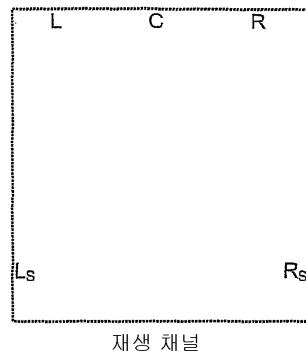
도면6



도면7a



도면7b



도면7c

