(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년12월05일
(11) 등록번호 10-2473263
(24) 등록일자 2022년11월29일

(51) 국제특허분류(Int. Cl.)
H03G 3/30 (2006.01) *H03G 7/00* (2006.01)
(52) CPC특허분류
H03G 3/3089 (2013.01)
H03G 7/007 (2013.01)
(21) 출원번호 10-2021-7038455(분할)
(22) 출원일자(국제) 2014년03월17일
심사청구일자 2021년11월24일
(85) 번역문제출일자 2021년11월24일
(65) 공개번호 10-2021-0149199
(43) 공개일자 2021년12월08일
(62) 원출원 특허 10-2021-7008206
원출원일자(국제) 2014년03월17일
심사청구일자 2021년03월19일
(86) 국제출원번호 PCT/US2014/030385
(87) 국제공개번호 WO 2014/160542
국제공개일자 2014년10월02일
(30) 우선권주장
201310100422.1 2013년03월26일 중국(CN)
61/811,072 2013년04월11일 미국(US)
(56) 선행기술조사문헌
US20120257760 A1*
(뒷면에 계속)
전체 청구항 수 : 총 11 항

(73) 특허권자
돌비 레버러토리즈 라이센싱 코오포레이션
미합중국, 캘리포니아 94103, 샌프란시스코, 마켓 스트리트 1275
(72) 발명자
왕, 준
중국, 베이징 100020, 차오양 디스트릭트, 이스트 써드 링 미들 로드, 넘버 1, 월드 파이낸셜 센트리, 돌비 레버러토리즈 인터내셔널 서비스즈 (베이징) 컴퍼니 리미티드 내
루, 라이
중국, 베이징 100020, 차오양 디스트릭트, 이스트 써드 링 미들 로드, 넘버 1, 월드 파이낸셜 센트리, 돌비 레버러토리즈 인터내셔널 서비스즈 (베이징) 컴퍼니 리미티드 내
시펠트, 앨런
미국, 캘리포니아 94103-4813, 샌프란시스코, 포트레로 애비뉴 100, 돌비 레버러토리즈 인크. 내
(74) 대리인
장훈

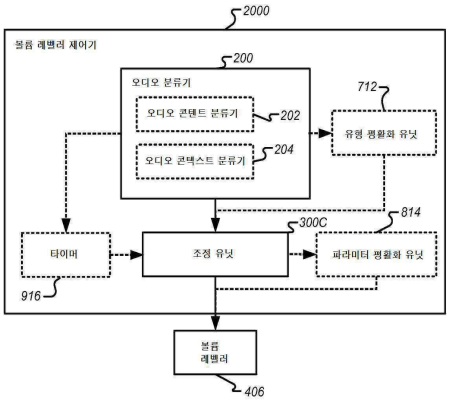
심사관 : 박정근

(54) 발명의 명칭 **볼륨 레벨러 제어기 및 제어 방법**

(57) 요 약

볼륨 레벨러 제어기 및 제어 방법이 개시된다. 일 실시예에서, 볼륨 레벨러 제어기는 오디오 신호의 콘텐트 유형을 실시간으로 식별하기 위한 오디오 콘텐트 분류기; 및 식별된 대로 상기 콘텐트 유형에 기초하여 연속적 방식으로 볼륨 레벨러를 조정하기 위한 조정 유닛을 포함한다. 상기 조정 유닛은 오디오 신호의 정보적 콘텐트 유형들과 상기 볼륨 레벨러의 동적 이득을 양으로 상관시키며, 오디오 신호의 방해 콘텐트 유형들과 볼륨 레벨러의 동적 이득을 음으로 상관시키도록 구성될 수 있다.

대 표 도 - 도20



(56) 선행기술조사문헌
US20120328115 A1*
US20080095385 A1*
US20030125933 A1*
US20070044137 A1*
*는 심사관에 의하여 인용된 문헌

명 세 서

청구범위

청구항 1

타겟 라우드니스 값에 기초한 라우드니스 정규화 방법에 있어서,

오디오 신호의 단기 또는 장기 특성에 기초하여 상기 오디오 신호의 오디오 프레임에 인가될 동적 이득 파라미터를 결정하는 단계로서, 상기 결정하는 단계는 상기 오디오 신호의 단기 특성 및 상기 타겟 라우드니스 값에 기초하여 제 1 오디오 프레임에 대한 동적 이득 파라미터를 결정하는 단계와, 상기 오디오 신호의 장기 특성 및 상기 타겟 라우드니스 값에 기초하여 제 2 오디오 프레임에 대한 동적 이득 파라미터를 결정하는 단계를 포함하는, 상기 동적 이득 파라미터를 결정하는 단계;

재생 시에 상기 타겟 라우드니스와 일치하도록 상기 제 1 오디오 프레임의 라우드니스를 조정하기 위하여 상기 제 1 오디오 프레임에 대한 상기 동적 이득 파라미터를 사용하여 상기 제 1 오디오 프레임을 스케일링하는 단계; 및

재생 시에 상기 타겟 라우드니스와 일치하도록 상기 제 2 오디오 프레임의 라우드니스를 조정하기 위하여 상기 제 2 오디오 프레임에 대한 상기 동적 이득 파라미터를 사용하여 상기 제 2 오디오 프레임을 스케일링하는 단계를 포함하는, 라우드니스 정규화 방법.

청구항 2

제 1 항에 있어서,

상기 동적 이득 파라미터는 실시간으로 식별되고 적용되는, 라우드니스 정규화 방법.

청구항 3

제 1 항에 있어서,

특정 콘텍스트 내에서 더욱 현저한 다이얼로그를 만드는 효과를 갖는 다이얼로그 강화가 적용되는, 라우드니스 정규화 방법.

청구항 4

제 1 항에 있어서,

음색 밸런스 상에서 하나 이상의 재생 레벨의 효과를 갖는 라우드니스 등화가 적용되는, 라우드니스 정규화 방법.

청구항 5

제 1 항에 있어서,

상기 동적 이득 파라미터에 파라미터 평활화가 적용되는, 라우드니스 정규화 방법.

청구항 6

타겟 라우드니스 값에 기초하여 라우드니스를 정규화하도록 구성된 오디오 처리 장치에 있어서,

적어도 하나의 프로세서; 및

컴퓨터 프로그램을 저장하는 적어도 하나의 메모리를 포함하고,

상기 적어도 하나의 메모리의 상기 컴퓨터 프로그램은 상기 적어도 하나의 프로세서가 상기 오디오 처리 장치로 하여금 적어도:

오디오 신호의 단기 또는 장기 특성에 기초하여 상기 오디오 신호의 오디오 프레임에 인가될 동적 이득 파라미터를 결정하는 것으로서, 상기 결정은 상기 오디오 신호의 단기 특성 및 상기 타겟 라우드니스 값에 기초하여

제 1 오디오 프레임에 대한 동적 이득 파라미터를 결정하는 것과, 상기 오디오 신호의 장기 특성 및 상기 타겟 라우드니스 값에 기초하여 제 2 오디오 프레임에 대한 동적 이득 파라미터를 결정하는 것을 포함하고;

재생 시에 상기 타겟 라우드니스와 일치하도록 상기 제 1 오디오 프레임의 라우드니스를 조정하기 위하여 상기 제 1 오디오 프레임에 대한 상기 동적 이득 파라미터를 사용하여 상기 제 1 오디오 프레임을 스케일링하고;

재생 시에 상기 타겟 라우드니스와 일치하도록 상기 제 2 오디오 프레임의 라우드니스를 조정하기 위하여 상기 제 2 오디오 프레임에 대한 상기 동적 이득 파라미터를 사용하여 상기 제 2 오디오 프레임을 스케일링하는, 오디오 처리 장치.

청구항 7

제 6 항에 있어서,

상기 동적 이득 파라미터는 실시간으로 식별되고 적용되는, 오디오 처리 장치.

청구항 8

제 6 항에 있어서,

특정 콘텍스트 내에서 더욱 현저한 다이얼로그를 만드는 효과를 갖는 다이얼로그 강화가 적용되는, 오디오 처리 장치.

청구항 9

제 6 항에 있어서,

음색 밸런스 상에서 하나 이상의 재생 레벨의 효과를 갖는 라우드니스 등화가 적용되는, 오디오 처리 장치.

청구항 10

제 6 항에 있어서,

상기 동적 이득 파라미터에 파라미터 평활화가 적용되는, 오디오 처리 장치.

청구항 11

머신에 의해 판독 가능한 프로그램 저장 디바이스로서, 연산들을 수행하도록 상기 머신에 의해 실행가능한 명령들의 프로그램을 유형적으로 포함하는, 상기 프로그램 저장 디바이스에 있어서, 상기 연산들은:

오디오 신호의 단기 또는 장기 특성에 기초하여 상기 오디오 신호의 오디오 프레임에 인가될 동적 이득 파라미터를 결정하는 단계로서, 상기 결정하는 단계는 상기 오디오 신호의 단기 특성 및 타겟 라우드니스 값에 기초하여 제 1 오디오 프레임에 대한 동적 이득 파라미터를 결정하는 단계와, 상기 오디오 신호의 장기 특성 및 상기 타겟 라우드니스 값에 기초하여 제 2 오디오 프레임에 대한 동적 이득 파라미터를 결정하는 단계를 포함하는, 상기 동적 이득 파라미터를 결정하는 단계;

재생 시에 상기 타겟 라우드니스와 일치하도록 상기 제 1 오디오 프레임의 라우드니스를 조정하기 위하여 상기 제 1 오디오 프레임에 대한 상기 동적 이득 파라미터를 사용하여 상기 제 1 오디오 프레임을 스케일링하는 단계; 및

재생 시에 상기 타겟 라우드니스와 일치하도록 상기 제 2 오디오 프레임의 라우드니스를 조정하기 위하여 상기 제 2 오디오 세그먼트에 대한 상기 동적 이득 파라미터를 사용하여 상기 제 2 오디오 프레임을 스케일링하는 단계를 포함하는, 머신에 의해 판독 가능한 프로그램 저장 디바이스.

발명의 설명

기 술 분 야

[0001] 관련 출원들에 대한 상호 참조

[0002] 본 출원은 2013년 3월 26일에 출원된 중국 특허 출원 번호 제201310100422.1호 및 2013년 4월 11일에 출원된 미국 가 특허 출원 번호 제61/811,072호에 대한 우선권을 주장하며, 그 각각은 여기에 전체적으로 참조로서 통합

된다.

[0003] 본 출원은 일반적으로 오디오 신호 프로세싱에 관한 것이다. 구체적으로, 본 출원의 실시예들은 다이얼로그 강화기, 서라운드 버추얼라이저, 볼륨 레벨러 및 등화기의 오디오 분류 및 프로세싱, 특히 제어를 위한 장치들 및 방법들에 관한 것이다.

배 경 기 술

[0004] 몇몇 오디오 개선 디바이스들은 오디오의 전체 품질을 개선하며 그에 따라 사용자의 경험을 강화하기 위해, 시간 도메인 또는 스펙트럼 도메인에서, 오디오 신호들을 변경하려는 경향이 있다. 다양한 오디오 개선 디바이스들은 다양한 목적들을 위해 개발되어 왔다. 오디오 개선 디바이스들의 몇몇 통상적인 예들은 다음을 포함한다:

[0005] 다이얼로그 강화기: 다이얼로그는 스토리를 이해하기 위해 영화 및 라디오 또는 TV 프로그램에서 가장 중요한 구성요소이다. 방법들은 그것들의 명료성 및 그것들의 양해도를 증가시키기 위해, 특히 감소하는 청력을 가진 연장자들을 위해 다이얼로그들을 강화하도록 개발되었다.

[0006] 서라운드 버추얼라이저: 서라운드 버추얼라이저는 서라운드(다-채널) 사운드 신호가 PC의 내부 스피커들을 통해 또는 헤드폰들을 통해 렌더링될 수 있게 한다. 즉, 스테레오 디바이스(스피커들 및 헤드폰들과 같은)를 갖고, 그것은 가상으로 서라운드 효과를 생성하며 소비자들을 위해 영화적 경험을 제공한다.

[0007] 볼륨 레벨러: 볼륨 레벨러는 재생 중인 오디오 콘텐트의 볼륨을 튜닝하는 것 및 타겟 라우드니스 값에 기초하여 타임라인에 걸쳐 그것을 거의 일정하게 유지하는 것을 목표로 한다.

[0008] 등화기: 등화기는 "톤" 또는 "음색"으로서 알려진 바와 같이, 스펙트럼 균형의 일관성을 제공하며, 사용자들로 하여금 특정한 사운드들을 강조하거나 또는 원하지 않는 사운드들을 제거하기 위해, 각각의 개개의 주파수 대역 상에서 주파수 응답 (이득)의 전체 프로파일(곡선 또는 형태)을 구성하도록 허용한다. 종래의 등화기에서, 상이한 등화기 프리셋들은 상이한 음악 장르들과 같은, 상이한 사운드들을 위해 제공될 수 있다. 프리셋이 선택되거나 또는 등화 프로파일이 설정된다면, 동일한 등화 이득들이, 등화 프로파일이 수동으로 변경될 때까지, 신호 상에 적용될 것이다. 반대로, 동적 등화기는 오디오의 스펙트럼 균형을 계속해서 모니터링하고, 그것을 원하는 톤에 비교하며, 오디오의 원래 톤을 원하는 톤으로 변환하기 위해 등화 필터를 동적으로 조정함으로써 스펙트럼 균형 일관성을 달성한다.

[0009] 일반적으로, 오디오 개선 디바이스들은 그 자신의 애플리케이션 시나리오/콘텍스트를 가진다. 즉, 오디오 개선 디바이스는, 상이한 콘텐트들이 상이한 방식들로 프로세싱될 필요가 있을 수 있기 때문에, 단지 특정한 세트의 콘텐트들에만 적절할 수 있으며 가능한 오디오 신호들 모두에 대해 적절하지는 않을 수 있다. 예를 들면, 다이얼로그 강화 방법은 보통 영화 콘텐트 상에 적용된다. 그것이 어떤 다이얼로그들도 없는 음악상에 적용된다면, 그것은 몇몇 주파수 서브-대역들을 거짓으로 부스팅하며 중 음색 변화 및 지각적 불일치성을 도입할 수 있다. 유사하게, 잡음 억제 방법이 음악 신호들 상에 적용된다면, 강한 아티팩트들이 들릴 것이다.

[0010] 그러나, 보통 오디오 개선 디바이스들의 세트를 포함하는 오디오 프로세싱 시스템에 대해, 그것의 입력은 불가피하게 가능한 유형들의 오디오 신호들 모두일 수 있다. 예를 들면, PC에 통합된, 오디오 프로세싱 시스템은 영화, 음악, VoIP 및 게임을 포함하여, 다양한 소스들로부터 오디오 콘텐트를 수신할 것이다. 따라서, 프로세싱되는 콘텐트를 식별하거나 또는 구별하는 것은 대응하는 콘텐트 상에서 보다 양호한 알고리즘들 또는 각각의 알고리즘의 보다 양호한 파라미터들을 적용하기 위해, 중요해진다.

[0011] 오디오 콘텐트를 구별하며 그에 따라 보다 양호한 파라미터들 또는 보다 양호한 오디오 개선 알고리즘들을 적용하기 위해, 종래의 시스템들은 보통 프리셋들의 세트를 사전-설계하며, 사용자들은 플레이되는 콘텐트에 대한 프리셋을 선택하도록 요청된다. 프리셋은 보통 구체적으로 영화 또는 음악 재생을 위해 설계되는 '영화' 프리셋 및 '음악' 프리셋과 같은, 적용될 오디오 개선 알고리즘들의 세트 및/또는 그것들의 최고 파라미터들을 인코딩한다.

발명의 내용

해결하려는 과제

[0012] 그러나, 수동 선택은 사용자들에게 불편하다. 사용자들은 보통 미리 정의된 프리셋들 중에서 빈번하게 스위칭하지 않으며 단지 모든 콘텐트에 대해 하나의 프리셋만을 사용하여 유지한다. 또한, 심지어 몇몇 자동 해결책들에

서, 프리셋들에서의 파라미터들 또는 알고리즘들 셋업은 보통 개별적이며(특정 콘텐트에 대하여 특정 알고리즘에 대한 턴 온 또는 오프와 같은), 그것은 콘텐트-기반 연속 방식으로 파라미터들을 조정할 수 없다.

과제의 해결 수단

[0013] 본 출원의 제 1 측면은 재생 중인 오디오 콘텐트에 기초하여 연속적 방식으로 오디오 개선 디바이스들을 자동으로 구성하는 것이다. 이러한 "자동" 모드를 갖고, 사용자들은 상이한 프리셋들을 선택할 필요 없이 그것들의 콘텐트를 간단히 즐길 수 있다. 다른 한편으로, 연속적으로 튜닝하는 것은 전이 포인트들에서 가청 아티팩트들을 회피하기 위해 더 중요하다.

[0014] 제 1 측면의 실시예에 따르면, 오디오 프로세싱 장치는 오디오 신호를 적어도 하나의 오디오 유형으로 실시간으로 분류하기 위한 오디오 분류기; 청중의 경험을 개선하기 위한 오디오 개선 디바이스; 및 상기 적어도 하나의 오디오 유형의 신뢰도 값에 기초하여 연속적 방식으로 상기 오디오 개선 디바이스의 적어도 하나의 파라미터를 조정하기 위한 조정 유닛을 포함한다.

[0015] 상기 오디오 개선 디바이스는 다이얼로그 강화기, 서라운드 버추얼라이저, 볼륨 레벨러 및 등화기 중 임의의 것일 수 있다.

[0016] 그에 상응하여, 오디오 프로세싱 방법은: 오디오 신호를 적어도 하나의 오디오 유형으로 실시간으로 분류하는 단계; 및 상기 적어도 하나의 오디오 유형의 신뢰도 값에 기초하여 연속적 방식으로 오디오 개선을 위한 적어도 하나의 파라미터를 조정하는 단계를 포함한다.

[0017] 제 1 측면의 또 다른 실시예에 따르면, 볼륨 레벨러 제어기는 오디오 신호의 콘텐트 유형을 실시간으로 식별하기 위한 오디오 콘텐트 분류기; 및 식별된 대로 상기 콘텐트 유형에 기초하여 연속적 방식으로 볼륨 레벨러를 조정하기 위한 조정 유닛을 포함한다. 상기 조정 유닛은 상기 오디오 신호의 정보적 콘텐트 유형과 상기 볼륨 레벨러의 동적 이득을 양으로 상관시키도록, 및 상기 오디오 신호의 방해 콘텐트 유형들과 상기 볼륨 레벨러의 상기 동적 이득을 음으로 상관시키도록 구성될 수 있다.

[0018] 상기 서술된 바와 같이 볼륨 레벨러 제어기를 포함한 오디오 프로세싱 장치가 또한 개시된다.

[0019] 그에 상응하여, 볼륨 레벨러 제어 방법은: 오디오 신호의 콘텐트 유형을 실시간으로 식별하는 단계; 및 상기 오디오 신호의 정보적 콘텐트 유형들과 상기 볼륨 레벨러의 동적 이득을 양으로 상관시키며, 상기 오디오 신호의 방해 콘텐트 유형들과 상기 볼륨 레벨러의 동적 이득을 음으로 상관시킴으로써, 식별된 대로 상기 콘텐트 유형에 기초하여 연속적 방식으로 볼륨 레벨러를 조정하는 단계를 포함한다.

[0020] 제 1 측면의 또 다른 실시예에 따르면, 등화기 제어기는 오디오 신호의 오디오 유형을 실시간으로 식별하기 위한 오디오 분류기; 및 식별된 대로 상기 오디오 유형의 신뢰도 값에 기초하여 연속적 방식으로 등화기를 조정하기 위한 조정 유닛을 포함한다.

[0021] 상기 서술된 바와 같이 등화기 제어기를 포함한 오디오 프로세싱 장치가 또한 개시된다.

[0022] 그에 상응하여, 등화기 제어 방법은: 오디오 신호의 오디오 유형을 실시간으로 식별하는 단계; 및 식별된 대로 상기 오디오 유형의 신뢰도 값에 기초하여 연속적 방식으로 등화기를 조정하는 단계를 포함한다.

[0023] 본 출원은 또한 프로세서에 의해 실행될 때, 컴퓨터 프로그램 지시들을 기록한 컴퓨터-판독 가능한 매체를 제공하며, 상기 지시들은 상기 프로세서로 하여금 상기 언급된 오디오 프로세싱 방법, 또는 볼륨 레벨러 제어 방법, 또는 등화기 제어 방법을 실행할 수 있게 한다.

[0024] 제 1 측면의 실시예들에 따르면, 다이얼로그 강화기, 서라운드 버추얼라이저, 볼륨 레벨러 및 등화기 중 하나일 수 있는, 오디오 개선 디바이스는 오디오 신호의 유형 및/또는 상기 유형의 신뢰도 값에 따라 연속적으로 조정될 수 있다.

[0025] 본 출원의 제 2 측면은 다수의 오디오 유형들을 식별하기 위해 콘텐트 식별 구성요소를 개발하는 것이며, 검출된 결과들은 연속적인 방식으로 보다 양호한 파라미터들을 찾음으로써, 다양한 오디오 개선 디바이스들의 행동들을 조종/가이드하기 위해 사용될 수 있다.

[0026] 제 2 측면의 실시예에 따르면, 오디오 분류기는: 각각이 오디오 프레임들의 시퀀스를 포함한 단기 오디오 세그먼트들로부터 단기 특징들(short-term features)을 추출하기 위한 단기 특징 추출기; 각각의 단기 특징들을 사용하여 장기 오디오 세그먼트에서의 단기 세그먼트들의 시퀀스를 단기 오디오 유형들로 분류하기 위한 단기 분

류기; 장기 특징들(long-term features)로서, 상기 장기 오디오 세그먼트에서의 상기 단기 세그먼트들의 시퀀스에 대하여 상기 단기 분류기의 결과들의 통계들을 산출하기 위한 통계 추출기; 및 상기 장기 특징들을 사용하여, 상기 장기 오디오 세그먼트를 장기 오디오 유형들로 분류하기 위한 장기 분류기를 포함한다.

[0027] 상기 서술된 바와 같이 오디오 분류기를 포함한 오디오 프로세싱 장치가 또한 개시된다.

[0028] 그에 상응하여, 오디오 분류 방법은: 각각이 오디오 프레임들의 시퀀스를 포함한 단기 오디오 세그먼트들로부터 단기 특징들을 추출하는 단계; 각각의 단기 특징들을 사용하여 장기 오디오 세그먼트에서의 단기 세그먼트들의 시퀀스를 단기 오디오 유형들로 분류하는 단계; 장기 특징들로서, 상기 장기 오디오 세그먼트에서의 단기 세그먼트들의 시퀀스에 대하여 분류 동작의 결과들의 통계들을 산출하는 단계; 및 상기 장기 특징들을 사용하여 상기 장기 오디오 세그먼트를 장기 오디오 유형들로 분류하는 단계를 포함한다.

[0029] 제 2 측면의 또 다른 실시예에 따르면, 오디오 분류기는: 오디오 신호의 단기 세그먼트의 콘텐트 유형을 식별하기 위한 오디오 콘텐트 분류기; 및 상기 오디오 콘텐트 분류기에 의해 식별된 상기 콘텐트 유형에 적어도 부분적으로 기초하여 상기 단기 세그먼트의 콘텍스트 유형을 식별하기 위한 오디오 콘텍스트 분류기를 포함한다.

[0030] 상기 서술된 바와 같이 오디오 분류기를 포함한 오디오 프로세싱 장치가 또한 개시된다.

[0031] 그에 상응하여, 오디오 분류 방법은: 오디오 신호의 단기 세그먼트의 콘텐트 유형을 식별하는 단계; 및 식별된 대로 상기 콘텐트 유형에 적어도 부분적으로 기초하여 상기 단기 세그먼트의 콘텍스트 유형을 식별하는 단계를 포함한다.

[0032] 본 출원은 또한 프로세서에 의해 실행될 때, 컴퓨터 프로그램 지시들을 기록한 컴퓨터-판독 가능한 매체를 제공하며, 상기 지시들은 상기 프로세서로 하여금 상기 언급된 오디오 분류 방법들을 실행할 수 있게 한다.

[0033] 제 2 측면의 실시예들에 따르면, 오디오 신호는 상이한 장기 유형들 또는 콘텍스트 유형들로 분류될 수 있으며, 이것은 단기 유형들 또는 콘텐트 유형들과 상이하다. 상기 오디오 신호의 유형들 및/또는 상기 유형들의 신뢰도 값은 다이얼로그 강화기, 서라운드 버추얼라이저, 볼륨 레벨러 또는 등화기와 같은, 오디오 개선 디바이스를 조정하기 위해 추가로 사용될 수 있다.

발명의 효과

[0034] 본 출원의 제 1 측면은 재생 중인 오디오 콘텐트에 기초하여 연속적 방식으로 오디오 개선 디바이스들을 자동으로 구성하는 것이다. 이러한 "자동" 모드를 갖고, 사용자들은 상이한 프리셋들을 선택할 필요 없이 그것들의 콘텐트를 간단히 즐길 수 있다. 다른 한편으로, 연속적으로 튜닝하는 것은 전이 포인트들에서 가청 아티팩트들을 회피하기 위해 더 중요하다.

[0035] 본 출원의 제 2 측면은 다수의 오디오 유형들을 식별하기 위해 콘텐트 식별 구성요소를 개발하는 것이며, 검출된 결과들은 연속적인 방식으로 보다 양호한 파라미터들을 찾음으로써, 다양한 오디오 개선 디바이스들의 행동들을 조종/가이드하기 위해 사용될 수 있다.

도면의 간단한 설명

[0036] 본 출원은 첨부한 도면들의 형태들에서, 제한에 의해서가 아닌, 예로서 예시되며, 여기에서 유사한 참조 부호들은 유사한 요소들을 나타낸다.

도 1은 출원의 실시예에 따른 오디오 프로세싱 장치를 예시한 도면.

도 2 및 도 3은 도 1에 도시된 바와 같이 실시예의 변형들을 예시한 도면들.

도 4 내지 도 6은 다수의 오디오 유형들을 식별하며 신뢰도 값의 산출을 위한 분류기들의 가능한 아키텍처를 예시한 도면들.

도 7 내지 도 9는 본 출원의 오디오 프로세싱 장치의 보다 많은 실시예들을 예시한 도면들.

도 10은 상이한 오디오 유형들 사이에서의 전이의 지연을 예시한 도면.

도 11 내지 도 14는 본 출원의 실시예들에 따른 오디오 프로세싱 방법을 예시한 흐름도들.

도 15는 본 출원의 실시예에 따른 다이얼로그 강화기 제어기를 예시한 도면.

도 16 및 도 17은 다이얼로그 강화기의 제어시 본 출원에 따른 오디오 프로세싱 방법의 사용을 예시한

흐름도들.

도 18은 본 출원의 실시예에 따른 서라운드 버추얼라이저 제어기를 예시한 도면.

도 19는 서라운드 버추얼라이저의 제어시 본 출원에 따른 오디오 프로세싱 방법의 사용을 예시한 흐름도.

도 20은 본 출원의 실시예에 따른 볼륨 레벨러 제어기를 예시한 도면.

도 21은 본 출원에 따른 볼륨 레벨러 제어기의 효과를 예시한 도면.

도 22는 본 출원의 실시예에 따른 등화기 제어기를 예시한 도면.

도 23은 원하는 스펙트럼 균형 프리셋들의 여러 개의 예들을 예시한 도면.

도 24는 본 출원의 실시예에 따른 오디오 분류기를 예시한 도면.

도 25 및 도 26은 본 출원의 오디오 분류기에 의해 사용될 몇몇 특징들을 예시한 도면들.

도 27 내지 도 29는 본 출원에 따른 오디오 분류기의 보다 많은 실시예들을 예시한 도면들.

도 30 내지 도 33은 본 출원의 실시예들에 따른 오디오 분류 방법을 예시한 흐름도들.

도 34는 본 출원의 또 다른 실시예에 따른 오디오 분류기를 예시한 도면.

도 35는 본 출원의 또 다른 실시예에 따른 오디오 분류기를 예시한 도면.

도 36은 본 출원의 오디오 분류기에서 사용된 경험적 규칙들을 예시한 도면.

도 37 및 도 38은 본 출원에 따른 오디오 분류기의 보다 많은 실시예들을 예시한 도면들.

도 39 및 도 40은 본 출원의 실시예들에 따른 오디오 분류 방법을 예시한 흐름도들.

도 41은 본 출원의 실시예들을 구현하기 위한 대표적인 시스템을 예시한 블록도.

발명을 실시하기 위한 구체적인 내용

[0037] 본 출원의 실시예들은 도면들을 참조함으로써 이하에 설명된다. 명료함의 목적을 위해, 이 기술분야의 숙련자들에게 알려져 있지만 본 출원을 이해하기 위해 필요하지 않은 이들 구성요소들 및 프로세스들에 대한 표현들 및 설명들은 도면들 및 설명에서 생략된다는 것이 주의될 것이다.

[0038] 이 기술분야의 숙련자에 의해 이해될 바와 같이, 본 출원의 측면들은 시스템, 디바이스(예로서, 셀룰러 전화, 휴대용 미디어 플레이어, 개인용 컴퓨터, 서버, 텔레비전 셋-탑 박스, 또는 디지털 비디오 레코더, 또는 임의의 다른 미디어 플레이어), 방법 또는 컴퓨터 프로그램 제품으로서 구체화될 수 있다. 따라서, 본 출원의 측면들은 하드웨어 실시예, 소프트웨어 실시예(펌웨어, 상주 소프트웨어, 마이크로코드들 등을 포함) 또는 모두가 일반적으로 여기에서 "회로", "모듈" 또는 "시스템"으로서 불리울 수 있는 소프트웨어 및 하드웨어 측면들 양쪽 모두를 조합한 실시예의 형태를 취할 수 있다. 더욱이, 본 출원의 측면들은 컴퓨터 판독 가능한 프로그램 코드를 구체화한 하나 이상의 컴퓨터 판독 가능한 매체들에서 구체화된 컴퓨터 프로그램 제품의 형태를 취할 수 있다.

[0039] 하나 이상의 컴퓨터 판독 가능한 매체들의 임의의 조합이 이용될 수 있다. 컴퓨터 판독 가능한 매체는 컴퓨터 판독 가능한 신호 매체 또는 컴퓨터 판독 가능한 저장 매체일 수 있다. 컴퓨터 판독 가능한 저장 매체는, 예를 들면, 이에 제한되지 않지만, 전자, 자기, 광학, 전자기, 적외선, 또는 반도체 시스템, 장치, 또는 디바이스, 또는 앞서 말한 것의 임의의 적절한 조합일 수 있다. 컴퓨터 판독 가능한 저장 매체의 보다 구체적인 예들(비-철저한 리스트)은 다음을 포함할 것이다: 하나 이상의 와이어들을 가진 전기적 연결, 휴대용 컴퓨터 디스켓, 하드 디스크, 랜덤 액세스 메모리(RAM), 판독-전용 메모리(ROM), 삭제 가능한 프로그램 가능한 판독-전용 메모리(EPROM 또는 플래시 메모리), 광 섬유, 휴대용 콤팩트 디스크 판독-전용 메모리(CD-ROM), 광학 저장 디바이스, 자기 저장 디바이스, 또는 앞서 말한 것의 임의의 적절한 조합. 본 문서의 콘텍스트에서, 컴퓨터 판독 가능한 저장 매체는 지시 실행 시스템, 장치, 또는 디바이스에 의한 사용을 위해 또는 그것과 관련되어 프로그램을 포함하거나 또는 저장할 수 있는 임의의 유형의 매체일 수 있다.

[0040] 컴퓨터 판독 가능한 신호 매체는 예를 들면, 기저대역에서 또는 캐리어 파의 일부로서, 그 안에 구체화된 컴퓨터 판독 가능한 프로그램 코드를 가진 전파된 데이터 신호를 포함할 수 있다. 이러한 전파된 신호는 이에 제한되지 않지만, 전-자기 또는 광학 신호, 또는 그것의 임의의 적절한 조합을 포함하여, 다양한 형태들 중 임의의

것을 취할 수 있다.

[0041] 컴퓨터 판독 가능한 신호 매체는 컴퓨터 판독 가능한 저장 매체가 아니며 지시 실행 시스템, 장치, 또는 디바이스에 의한 사용을 위해 또는 그것과 관련되어 프로그램을 전달, 전파, 또는 수송할 수 있는 임의의 컴퓨터 판독 가능한 매체일 수 있다.

[0042] 컴퓨터 판독 가능한 매체상에 구체화된 프로그램 코드는 이에 제한되지 않지만, 무선, 유선 라인, 광 섬유 케이블, RF 등, 또는 앞서 말한 것의 임의의 적절한 조합을 포함한, 임의의 적절한 매체를 사용하여 송신될 수 있다.

[0043] 본 출원의 측면들에 대한 동작들을 실행하기 위한 컴퓨터 프로그램 코드는 Java, Smalltalk, C++ 등과 같은 객체 지향 프로그래밍 언어 및 "C" 프로그래밍 언어 또는 유사한 프로그래밍 언어들과 같은, 종래의 절차적 프로그래밍 언어들을 포함하여, 하나 이상의 프로그래밍 언어들의 임의의 조합으로 기록될 수 있다. 프로그램 코드는 전체적으로 독립형 소프트웨어 패키지로서 사용자의 컴퓨터상에서 또는 부분적으로 사용자의 컴퓨터상에서 및 부분적으로 원격 컴퓨터상에서 또는 전체적으로 원격 컴퓨터 또는 서버상에서 실행할 수 있다. 후자의 시나리오에서, 원격 컴퓨터는 근거리 네트워크(LAN) 또는 광역 네트워크(WAN)를 포함하여, 임의의 유형의 네트워크를 통해 사용자의 컴퓨터에 연결될 수 있거나, 또는 연결은 외부 컴퓨터에 대해 이루어질 수 있다(예를 들면, 인터넷 서비스 제공자를 사용하여 인터넷을 통해).

[0044] 본 출원의 측면들은 출원의 실시예들에 따른 방법들, 장치(시스템들) 및 컴퓨터 프로그램 제품들의 흐름도 예시들 및/또는 블록도들을 참조하여 이하에 설명된다. 흐름도 예시들 및/또는 블록도들의 각각의 블록, 및 흐름도 예시들 및/또는 블록도들에서의 블록들의 조합들은 컴퓨터 프로그램 지시들에 의해 구현될 수 있다는 것이 이해될 것이다. 이들 컴퓨터 프로그램 지시들은 기계를 생성하기 위해 범용 컴퓨터, 특수 목적 컴퓨터, 또는 다른 프로그램 가능한 데이터 프로세싱 장치의 프로세서에 제공될 수 있으며, 따라서 컴퓨터 또는 다른 프로그램 가능한 데이터 프로세싱 장치의 프로세서를 통해 실행하는 지시들이 흐름도 및/또는 블록도 블록 또는 블록들에서 특정된 기능들/동작들을 구현하기 위한 수단을 생성한다.

[0045] 이들 컴퓨터 프로그램 지시들은 또한 특정한 방식으로 기능하도록 컴퓨터, 다른 프로그램 가능한 데이터 프로세싱 장치, 또는 다른 디바이스들에 지시할 수 있는 컴퓨터 판독 가능한 매체에 저장될 수 있으며, 따라서 컴퓨터 판독 가능한 매체에 저장된 지시들은 흐름도 및/또는 블록도 블록 또는 블록들에 특정된 기능/동작을 구현하는 지시들을 포함한 제조 물품을 생성한다.

[0046] 컴퓨터 프로그램 지시들은 또한 동작적 동작들의 시리즈들이 컴퓨터 또는 다른 프로그램 가능한 장치상에서 실행하는 지시들이 흐름도 및/또는 블록도 블록 또는 블록들에서 특정된 기능들/동작들을 구현하기 위한 프로세스들을 제공하도록 컴퓨터 구현된 프로세스를 생성하기 위해 컴퓨터, 다른 프로그램 가능한 장치 또는 다른 디바이스들 상에서 수행되게 하기 위해 컴퓨터, 다른 프로그램 가능한 데이터 프로세싱 장치, 또는 다른 디바이스들로 로딩될 수 있다.

[0047] 본 출원의 실시예들이 이하에서 상세히 설명될 것이다. 명료함을 위해, 설명은 다음의 아키텍처로 조직화된다:

[0048] 파트 1: 오디오 프로세싱 장치 및 방법들

[0049] 섹션 1.1 오디오 유형들

[0050] 섹션 1.2 오디오 유형들의 신뢰도 값들 및 분류기들의 아키텍처

[0051] 섹션 1.3 오디오 유형들의 신뢰도 값들의 평활화

[0052] 섹션 1.4 파라미터 조정

[0053] 섹션 1.5 파라미터 평활화

[0054] 섹션 1.6 오디오 유형들의 전이

[0055] 섹션 1.7 실시예들 및 애플리케이션 시나리오들의 조합

[0056] 섹션 1.8 오디오 프로세싱 방법

[0057] 파트 2: 다이얼로그 강화기 제어기 및 제어 방법

[0058] 섹션 2.1 다이얼로그 강화의 레벨

[0059] 섹션 2.2 강화될 주파수 대역들을 결정하기 위한 임계치들

[0060] 섹션 2.3 배경 레벨로의 조정

[0061] 섹션 2.4 실시예들 및 애플리케이션 시나리오들의 조합

[0062] 섹션 2.5 다이얼로그 강화기 제어 방법

[0063] 파트 3: 서라운드 버추얼라이저 제어기 및 제어 방법

[0064] 섹션 3.1 서라운드 부스트 양

[0065] 섹션 3.2 시작 주파수

[0066] 섹션 3.3 실시예들 및 애플리케이션 시나리오들의 조합

[0067] 섹션 3.4 서라운드 버추얼라이저 제어 방법

[0068] 파트 4: 볼륨 레벨러 제어기 및 제어 방법

[0069] 섹션 4.1 정보적 및 방해 콘텐트 유형들

[0070] 섹션 4.2 상이한 콘텍스트들에서의 콘텐트 유형들

[0071] 섹션 4.3 콘텍스트 유형들

[0072] 섹션 4.4 실시예들 및 애플리케이션 시나리오들의 조합

[0073] 섹션 4.5 볼륨 레벨러 제어 방법

[0074] 파트 5: 등화기 제어기 및 제어 방법

[0075] 섹션 5.1 콘텐트 유형에 기초한 제어

[0076] 섹션 5.2 음악에서 우세 소스들의 우도

[0077] 섹션 5.3 등화기 프리셋들

[0078] 섹션 5.4 콘텍스트 유형에 기초한 제어

[0079] 섹션 5.5 실시예들 및 애플리케이션 시나리오들의 조합

[0080] 섹션 5.6 등화기 제어 방법

[0081] 파트 6: 오디오 분류기들 및 분류 방법들

[0082] 섹션 6.1 콘텐트 유형 분류에 기초한 콘텍스트 분류기

[0083] 섹션 6.2 장기 특징들의 추출

[0084] 섹션 6.3 단기 특징들의 추출

[0085] 섹션 6.4 실시예들 및 애플리케이션 시나리오들의 조합

[0086] 섹션 6.5 오디오 분류 방법들

[0087] 파트 7: VoIP 분류기들 및 분류 방법들

[0088] 섹션 7.1 단기 세그먼트에 기초한 콘텍스트 분류

[0089] 섹션 7.2 VoIP 스피치 및 VoIP 잡음을 사용한 분류

[0090] 섹션 7.3 평활화 변동

[0091] 섹션 7.4 실시예들 및 애플리케이션 시나리오들의 조합

[0092] 섹션 7.5 VoIP 분류 방법들

[0093] <u>파트 1: 오디오 프로세싱 장치 및 방법들</u>

[0094] 도 1은 재생 중인 오디오 콘텐트에 기초하여 개선된 파라미터들을 가진 적어도 하나의 오디오 개선 디바이스

(400)의 자동 구성을 지원하는 콘텐트-적응적 오디오 프로세싱 장치(100)의 일반적인 프레임워크를 예시한다. 그것은 3개의 주요 구성요소들: 오디오 분류기(200), 조정 유닛(300) 및 오디오 개선 디바이스(400)를 포함한다.

[0095] 오디오 분류기(200)는 오디오 신호를 적어도 하나의 오디오 유형으로 실시간으로 분류하기 위한 것이다. 그것은 재생 중인 콘텐트의 오디오 유형들을 자동으로 식별한다. 신호 프로세싱, 기계 학습 및 패턴 인식을 통해서와 같은, 임의의 오디오 분류 기술들이 오디오 콘텐트를 식별하기 위해 적용될 수 있다. 미리-정의된 타겟 오디오 유형들의 세트에 관한 오디오 콘텐트의 확률들을 나타내는, 신뢰도 값들이 일반적으로 동시에 추정된다.

[0096] 오디오 개선 디바이스(400)는 오디오 신호 상에서 프로세싱을 수행함으로써 청중의 경험을 개선하기 위한 것이며, 나중에 상세히 논의될 것이다.

[0097] 조정 유닛(300)은 적어도 하나의 오디오 유형의 신뢰도 값에 기초하여 연속적 방식으로 오디오 개선 디바이스의 적어도 하나의 파라미터를 조정하기 위한 것이다. 그것은 오디오 개선 디바이스(400)의 행동을 조종하도록 설계된다. 그것은 오디오 분류기(200)로부터 획득된 결과들에 기초하여 대응하는 오디오 개선 디바이스의 가장 적절한 파라미터들을 추정한다.

[0098] 다양한 오디오 개선 디바이스들이 이러한 장치에 적용될 수 있다. 도 2는 다이얼로그 강화기(DE)(402), 서라운드 버추얼라이저(SV)(404), 볼륨 레벨러(VL)(406) 및 등화기(EQ)(408)를 포함한, 4개의 오디오 개선 디바이스들을 포함한 예시적인 시스템을 도시한다. 각각의 오디오 개선 디바이스는 오디오 분류기(200)에서 획득된 결과들(오디오 유형들 및/또는 신뢰도 값들)에 기초하여, 연속적인 방식으로 자동으로 조정될 수 있다.

[0099] 물론, 오디오 프로세싱 장치는 반드시 오디오 개선 디바이스들의 종류들 모두를 포함하는 것은 아닐 수 있으며, 단지 그것들 중 하나 이상만을 포함할 수 있다. 다른 한편으로, 오디오 개선 디바이스들은 본 개시에서 주어진 이들 디바이스들에 제한되지 않으며 또한 본 출원의 범위 내에 있는 보다 많은 종류들의 오디오 개선 디바이스들을 포함할 수 있다. 더욱이, 다이얼로그 강화기(DE)(402), 서라운드 버추얼라이저(SV)(404), 볼륨 레벨러(VL)(406) 및 등화기(EQ)(408)를 포함한, 본 개시에 논의된 이들 오디오 개선 디바이스들의 명칭들은 제한을 구성하지 않을 것이며, 그것들 각각은 동일한 또는 유사한 기능들을 실현한 임의의 다른 디바이스들을 커버하는 것으로서 해석될 것이다.

[0100] 1.1 오디오 유형들

[0101] 다양한 종류들의 오디오 개선 디바이스를 적절하게 제어하기 위해, 본 출원은, 종래 기술에서의 이들 오디오 유형들이 또한 여기에서 적용 가능할지라도, 오디오 유형들의 새로운 아키텍처를 추가로 제공한다.

[0102] 구체적으로, 오디오 신호들에서 기본 구성요소들을 표현한 저-레벨 오디오 요소들 및 실생활 사용자 환경 애플리케이션들에서 가장 인기 있는 오디오 콘텐트들을 표현한 고-레벨 오디오 장르들을 포함한, 상이한 의미 레벨들로부터의 오디오 유형들이 모델링된다. 전자는 또한 "콘텐트 유형"으로서 지명될 수 있다. 기본 오디오 콘텐트 유형들은 스피치, 음악(노래를 포함한), 배경 사운드들(또는 사운드 효과들) 및 잡음을 포함할 수 있다.

[0103] 스피치 및 음악의 의미는 자명하다. 본 출원에서의 잡음은 의미적 잡음이 아닌, 물리적 잡음을 의미한다. 본 출원에서의 물리적 잡음은 예를 들면, 공기 조절기들로부터의 잡음들, 및 신호 송신 경로로 인한 핑크 잡음들과 같은, 기술적 이유들로부터 발생한 이들 잡음들을 포함할 수 있다. 반대로, 본 출원에서 "배경 사운드들"은 청취자의 관심의 핵심 타겟 주위에서 일어나는 청각 이벤트들일 수 있는 이들 사운드 효과들이다. 예를 들면, 전화 호출에서의 오디오 신호에서, 화자의 음성 외에, 전화 호출과 관계없는 몇몇 다른 사람들의 음성들, 키보드들의 사운드들, 발소리들의 사운드들 등과 같은, 의도되지 않은 몇몇 다른 사운드들이 있을 수 있다. 이들 원치 않는 사운드들은 잡음이 아닌 "배경 사운드들"로서 불리운다. 다시 말해서, 우리는 타겟(또는 청취자의 관심의 핵심 타겟)이 아니거나 또는 심지어 원치 않는 이들 사운드들로서 "배경 사운드들"을 정의할 수 있지만, 여전히 몇몇 의미론적 의미를 가진다; 반면 "잡음"은 타겟 사운드들 및 배경 사운드들을 제외한 이들 원치않는 사운드들로서 정의될 수 있다.

[0104] 때때로 배경 사운드들은 실제로 "원치 않지만" 의도적으로 생성되며, 영화에서의 이들 배경 사운드들, TV 프로그램 또는 라디오 방송 프로그램과 같은, 몇몇 유용한 정보를 운반한다. 따라서, 때때로 그것은 또한 "사운드 효과들"로서 불리울 수 있다. 이후 본 개시에서, 단지 "배경 사운드들"은 간결성을 위해 사용되며 그것은 "배경"으로서 추가로 축약될 수 있다.

[0105] 뿐만 아니라, 음악은 우세 소스들이 없는 음악 및 우세 소스들을 가진 음악으로서 추가로 분류될 수 있다. 음악

조각에서의 다른 소스들보다 훨씬 더 강한 소스(음성 또는 악기)가 있다면, 그것은 "우세 소스를 가진 음악"으로서 불리우며; 그렇지 않다면, 그것은 "우세 소스가 없는 음악"으로서 불리운다. 예를 들면, 노래하는 음성 및 다양한 악기들을 갖고 동반된 다음(polyphonic) 음악에서, 그것이 화성적으로 균형이 이루어졌거나 또는 여러 개의 가장 핵심적인 소스들의 에너지가 서로 비교 가능하다면, 그것은 우세 소스가 없는 음악인 것으로 고려되며; 반대로, 소스가 다른 것들이 훨씬 더 조용한 동안 훨씬 더 소리가 크다면, 그것은 우세 소스를 포함한 것으로 고려된다. 또 다른 예로서, 단수형 또는 개별적인 악기 톤들은 "우세 소스를 가진 음악"이다.

[0106] 음악은 상이한 표준들에 기초하여 상이한 유형들로서 추가로 분류될 수 있다. 그것은 이에 제한되지 않지만, 록, 재즈, 랩 및 포크송과 같은, 음악의 장르들에 기초하여 분류될 수 있다. 그것은 또한 성악 및 기악과 같은, 악기들에 기초하여 분류될 수 있다. 기악은 피아노 음악 및 기타 음악과 같은, 상이한 악기들을 갖고 플레이된 다양한 음악을 포함할 수 있다. 다른 예시적인 표준들은 음악의 리듬, 템포, 음색 및/또는 임의의 다른 음악적 속성들을 포함하며, 따라서 음악은 이들 속성들의 유사성에 기초하여 함께 그룹핑될 수 있다. 예를 들면, 음색에 따르면, 성악은 테너, 바리톤, 베이스, 소프라노, 메조 소프라노 및 알토와 같이 분류될 수 있다.

[0107] 오디오 신호의 콘텐트 유형은 복수의 프레임들로 구성된 것과 같이, 단기 오디오 세그먼트들에 대하여 분류될 수 있다. 일반적으로 오디오 프레임은 20ms와 같은, 다수의 밀리초들의 길이이며, 오디오 분류기에 의해 분류될 단기 오디오 세그먼트의 길이는 수백 밀리초들에서 1초와 같은 수 초들까지의 길이를 가질 수 있다.

[0108] 콘텐트-적응적 방식으로 오디오 개선 디바이스를 제어하기 위해, 오디오 신호는 실시간으로 분류될 수 있다. 상기 서술된 콘텐트 유형에 대해, 현재 단기 오디오 세그먼트의 콘텐트 유형은 현재 오디오 신호의 콘텐트 유형을 나타낸다. 단기 오디오 세그먼트의 길이가 너무 길지 않기 때문에, 오디오 신호는 잇따라서 비-중첩된 단기 오디오 세그먼트들로서 분할될 수 있다. 그러나, 단기 오디오 세그먼트들은 또한 오디오 신호의 타임 라인을 따라 연속적으로/반-연속적으로 샘플링될 수 있다. 즉, 단기 오디오 세그먼트들은 하나 이상의 프레임들의 스텝 크기에서 오디오 신호의 타임 라인을 따라 이동하는 미리 결정된 길이(단기 오디오 세그먼트의 의도된 길이)를 가진 윈도우를 갖고 샘플링될 수 있다.

[0109] 고-레벨 오디오 장르들은 또한, 그것이 오디오 신호의 장기 유형을 표시하기 때문에, "콘텍스트 유형"으로서 지명될 수 있으며, 상기 서술된 바와 같이 콘텐트 유형들로 분류될 수 있는, 즉각적인 사운드 이벤트의 환경 또는 콘텍스트로서 간주될 수 있다. 본 출원에 따르면, 콘텍스트 유형은 영화-형 미디어, 음악(노래를 포함한), 게임 및 VoIP(인터넷 프로토콜 상에서의 음성)와 같은, 가장 인기 있는 오디오 애플리케이션들을 포함할 수 있다.

[0110] 음악, 게임 및 VoIP의 의미는 자명하다. 영화-형 미디어는 영화, TV 프로그램, 라디오 방송 프로그램 또는 앞서 언급한 것과 유사한 임의의 다른 오디오 미디어를 포함할 수 있다. 영화-형 미디어의 주요 특성은 가능한 스피치들, 음악 및 다양한 종류들의 배경 사운드들(사운드 효과들)의 혼합이다.

[0111] 콘텐트 유형 및 콘텍스트 유형 양쪽 모두가 음악(노래를 포함한)을 포함한다는 것이 주의될 수 있다. 이후 본 출원에서, 우리는 그것들을 각각 구별하기 위해 표현들("단기 음악" 및 "장기 음악")을 사용한다.

[0112] 본 출원의 몇몇 실시예들에 대해, 몇몇 다른 콘텍스트 유형 아키텍처들이 또한 제안된다.

[0113] 예를 들면, 오디오 신호는 고-품질 오디오(영화-형 미디어 및 음악 CD) 또는 저-품질 오디오(VoIP, 저 비트 레이트 온라인 스트리밍 오디오 및 사용자 생성 콘텐트)로서 분류될 수 있으며, 이것은 총괄하여 "오디오 품질 유형들"로서 불리울 수 있다.

[0114] 또 다른 예로서, 오디오 신호는 VoIP 또는 비-VoIP로서 분류될 수 있으며, 이것은 상기 언급된 4-콘텍스트 유형 아키텍처(VoIP, 영화-형 미디어, (장기) 음악 및 게임)의 변환으로서 간주될 수 있다. VoIP 또는 비-VoIP의 콘텍스트와 관련되어, 오디오 신호는 VoIP 스피치, 비-VoIP 스피치, VoIP 잡음 및 비-VoIP 잡음과 같은, VoIP-관련 오디오 콘텐트 유형들로서 분류될 수 있다. VoIP 오디오 콘텐트 유형들의 아키텍처는 VoIP 콘텍스트가 보통 볼륨 레벨러(오디오 개선 디바이스의 일 종류)의 가장 도전적인 애플리케이션 시나리오이므로 VoIP 및 비-VoIP 콘텍스트들을 구별하는데 특히 유용하다.

[0115] 일반적으로 오디오 신호의 콘텍스트 유형은 단기 오디오 세그먼트들보다 긴 장기 오디오 세그먼트들에 대하여 분류될 수 있다. 장기 오디오 세그먼트는 단기 오디오 세그먼트에서의 프레임들의 수보다 많은 수에서의 복수의 프레임들로 구성된다. 장기 오디오 세그먼트는 또한 복수의 단기 오디오 세그먼트들로 구성될 수 있다. 일반적으로 장기 오디오 세그먼트는 수 초들 내지 수십 초들, 즉 10초들과 같은 대략 초들에서의 길이를 가질 수 있다.

[0116] 유사하게, 적응적 방식으로 오디오 개선 디바이스를 제어하기 위해, 오디오 신호는 실시간으로 콘텍스트 유형들로 분류될 수 있다. 유사하게, 현재 장기 오디오 세그먼트의 콘텍스트 유형은 현재 오디오 신호의 콘텍스트 유형을 나타낸다. 장기 오디오 세그먼트의 길이가 비교적 길기 때문에, 오디오 신호는 그것의 콘텍스트 유형의 갑작스런 변화 및 그에 따라 오디오 개선 디바이스(들)의 작동 파라미터들의 갑작스런 변화를 회피하기 위해 오디오 신호의 타임 라인을 따라 연속적으로/반-연속적으로 샘플링될 수 있다. 즉, 장기 오디오 세그먼트들은 하나 이상의 프레임들, 또는 하나 이상의 단기 세그먼트들의 스텝 크기에서 오디오 신호의 타임 라인을 따라 이동하는 미리 결정된 길이(장기 오디오 세그먼트의 의도된 길이)를 가진 윈도우를 갖고 샘플링될 수 있다.

[0117] 콘텐트 유형 및 콘텍스트 유형 양쪽 모두가 상기 설명되었다. 본 출원의 실시예들에서, 조정 유닛(300)은 다양한 콘텐트 유형들 중 적어도 하나, 및/또는 다양한 콘텍스트 유형들 중 적어도 하나에 기초하여 오디오 개선 디바이스(들)의 적어도 하나의 파라미터를 조정할 수 있다. 그러므로, 도 3에 도시된 바와 같이, 도 1에 도시된 실시예의 변형에서, 오디오 분류기(200)는 오디오 콘텐트 분류기(202) 또는 오디오 콘텍스트 분류기(204), 또는 양쪽 모두를 포함할 수 있다.

[0118] 상이한 표준들(콘텍스트 유형들에 대한 것과 같은)에 기초한 상이한 오디오 유형들, 뿐만 아니라 상이한 계층 레벨들(콘텐트 유형들에 대한 것과 같은) 상에서의 상이한 오디오 유형들이 상기 언급되었다. 그러나, 표준들 및 계층 레벨들은 단지 여기에서 설명의 편리함을 위한 것이며 분명히 비 제한적이다. 다시 말해서, 본 출원에서, 나중에 설명될 바와 같이, 상기 언급된 오디오 유형들 중 임의의 둘 이상은 오디오 분류기(200)에 의해 동시에 식별될 수 있으며 조정 유닛(300)에 의해 동시에 고려될 수 있다. 다시 말해서, 상이한 계층 레벨들에서의 모든 오디오 유형들은 병렬, 또는 동일한 레벨에 있을 수 있다.

[0119] 1.2 오디오 유형들의 신뢰도 값들 및 분류기들의 아키텍처

[0120] 오디오 분류기(200)는 경-판정 결과들을 출력할 수 있거나, 또는 조정 유닛(300)은 오디오 분류기(200)의 결과들을 경판정 결과들로서 간주할 수 있다. 경판정에 대해서조차, 다수의 오디오 유형들은 오디오 세그먼트에 할당될 수 있다. 예를 들면, 오디오 세그먼트는 그것이 스피치 및 단기 음악의 혼합 신호일 수 있기 때문에 '스피치' 및 '단기 음악' 양쪽 모두에 의해 라벨링될 수 있다. 획득된 라벨들은 오디오 개선 디바이스(들)(400)를 조종하기 위해 직접 사용될 수 있다. 단순한 예는 스피치가 존재할 때 다이얼로그 강화기(402)를 가능하게 하며 스피치가 존재하지 않을 때 그것을 턴 오프하는 것이다. 그러나, 이러한 경판정 방법은 신중한 평활화 기법(나중에 논의될)이 없다면, 하나의 오디오 유형에서 또 다른 것으로 전이 포인트들에서의 몇몇 부자연스러움을 도입할 수 있다.

[0121] 보다 많은 유연성을 가지며 연속적 방식으로 오디오 개선 디바이스들의 파라미터들을 튜닝하기 위해, 각각의 타겟 오디오 유형의 신뢰도 값이 추정될 수 있다(연 판정). 신뢰도 값은 0에서 1까지의 값들을 갖고, 식별될 오디오 콘텐트 및 타겟 오디오 유형 사이에서의 매칭된 레벨을 나타낸다.

[0122] 이전 서술된 바와 같이, 많은 분류 기술들은 신뢰도 값들을 직접 출력할 수 있다. 신뢰도 값은 또한 다양한 방법들로부터 산출될 수 있으며, 이것은 분류기의 일부로서 간주될 수 있다. 예를 들면, 오디오 모델들이 가우시안 혼합 모델들(Gaussian Mixture Models; GMM)과 같은 몇몇 확률적 모델링 기술들에 의해 트레이닝된다면, 사후 확률은 다음과 같이, 신뢰도 값을 표현하기 위해 사용될 수 있다:

[0123]
$$p(c_i \mid x) = \frac{p(x \mid c_i)}{\sum_{i=1}^{N} p(x \mid c_i)} \quad (1)$$

[0124] 여기에서 x는 오디오 세그먼트의 조각이고, c_i는 타겟 오디오 유형이고, N은 타겟 오디오 유형들의 수이고, $p(x|c_i)$는 오디오 세그먼트(x)가 오디오 유형(c_i)인 가능성이며, $p(c_i|x)$는 대응하는 사후 확률이다.

[0125] 다른 한편으로, 오디오 모델들이 지지 벡터 기계(Support Vector Machine) 및 adaBoost처럼 몇몇 판별적 방법으로부터 트레이닝된다면, 단지 스코어들(실제 값들)은 모델 비교로부터 획득된다. 이들 경우들에서, 시그모이드 함수는 보통 획득된 스코어(이론적으로 -∞에서 ∞까지)를 예상된 신뢰도(0에서 1까지)에 매핑시키기 위해 사용된다:

$$conf = \frac{1}{1+e^{Ay+B}} \quad (2)$$

[0126]

[0127] 여기에서 y는 SVM 또는 adaBoost로부터의 출력 스코어이고, A 및 B는 몇몇 잘-알려진 기술들을 사용함으로써 설정된 트레이닝 데이터로부터 추정되도록 요구된 두 개의 파라미터들이다.

[0128] 본 출원의 몇몇 실시예들에 대해, 조정 유닛(300)은 둘 이상의 콘텐트 유형들 및/또는 둘 이상의 콘텍스트 유형들을 사용할 수 있다. 그 후, 오디오 콘텐트 분류기(202)는 둘 이상의 콘텐트 유형들을 식별하는 것을 요구하며 및/또는 오디오 콘텍스트 분류기(204)는 둘 이상의 콘텍스트 유형들을 식별하는 것을 요구한다. 이러한 상황에서, 오디오 콘텐트 분류기(202) 또는 오디오 콘텍스트 분류기(204)는 특정한 아키텍처에서 조직된 분류기들의 그룹일 수 있다.

[0129] 예를 들면, 조정 유닛(300)이 4개의 종류들의 콘텍스트 유형들, 즉 영화-형 미디어, 장기 음악, 게임 및 VoIP 모두를 요구한다면, 오디오 콘텍스트 분류기(204)는 다음의 상이한 아키텍처들을 가질 수 있다:

[0130] 첫 번째로, 오디오 콘텍스트 분류기(204)는 도 4에 도시된 바와 같이 조직된 6개의 1-대-1 이진 분류기들(각각의 분류기는 하나의 타겟 오디오 유형을 또 다른 타겟 오디오 유형으로부터 식별한다), 도 5에 도시된 바와 같이 조직된 3개의 1-대-나머지 이진 분류기들(각각의 분류기는 다른 것들로부터 타겟 오디오 유형을 식별한다), 및 도 6에 도시된 바와 같이 조직된 4개의 1-대-나머지 분류기들을 포함할 수 있다. 결정 방향성 비순환 그래프(Decision Directed Acyclic Graph; DDAG) 아키텍처와 같은 다른 아키텍처들이 또한 있다. 도 4 내지 도 6 및 이하의 대응하는 설명에서, "영화-형 미디어" 대신에 "영화"가 간결성을 위해 사용된다는 것을 주의하자.

[0131] 각각의 이진 분류기는 그것의 출력에 대한 신뢰도 스코어(H(x))를 줄 것이다(x는 오디오 세그먼트를 나타낸다). 각각의 이진 분류기의 출력들이 획득된 후, 우리는 식별된 콘텍스트 유형들의 최종 신뢰도 값들에 그것들을 매핑시키도록 요구한다.

[0132] 일반적으로, 오디오 신호가 M개의 콘텍스트 유형들로 분류될 것이라고 가정하자(M은 양수이다). 종래의 1-대-1 아키텍처는 각각의 하나가 두 개의 클래스들로부터의 데이터에 대해 트레이닝되는 M(M-1)/2 분류기들을 구성하고, 그 후 각각의 1-대-1 분류기는 그것의 선호된 클래스에 대해 하나를 투표하며, 최종 결과는 M(M-1)/2 분류기들의 분류들 중에서 대부분의 표들이 가진 클래스이다. 종래의 1-대-1 아키텍처와 비교하면, 도 4에서의 계층적 아키텍처는 또한 M(M-1)/2 분류기들을 구성하도록 요구한다. 그러나, 검사 반복들은, 세그먼트(x)가 각각의 계층 레벨에서 대응하는 클래스에 있거나 또는 있지 않은 것으로 결정되며 전체 레벨 카운트가 M-1이므로, M-1로 단축될 수 있다. 다양한 콘텍스트 유형들에 대한 최종 신뢰도 값들은 이진 분류 신뢰도($H_k(x)$)로부터 산출될 수 있으며, 예를 들면(k=1, 2, ..., 6, 상이한 콘텍스트 유형들을 표현하는):

$$C_{MOVIE} = (1 - H_1(x)) \cdot (1 - H_3(x)) \cdot (1 - H_6(x))$$

$$C_{VOIP} = H_1(x) \cdot H_2(x) \cdot H_4(x)$$

$$C_{MUSIC} = H_1(x) \cdot (1 - H_2(x)) \cdot (1 - H_5(x)) + H_3(x) \cdot (1 - H_1(x)) \cdot (1 - H_5(x)) + H_6(x) \cdot (1 - H_1(x)) \cdot (1 - H_3(x))$$

$$C_{GAME} = H_1(x) \cdot H_2(x) \cdot (1 - H_4(x)) + H_1(x) \cdot H_5(x) \cdot (1 - H_2(x)) + H_3(x) \cdot H_5(x) \cdot (1 - H_1(x))$$

[0133]

[0134] 도 5에 도시된 아키텍처에서, 이진 분류 결과들($H_k(x)$)로부터 최종 신뢰도 값들로의 매핑 함수는 다음의 예로서 정의될 수 있다:

$$C_{MOVIE} = H_1(x)$$
$$C_{MUSIC} = H_2(x) \cdot \big(1 - H_1(x)\big)$$
$$C_{VOIP} = H_3(x) \cdot \big(1 - H_2(x)\big) \cdot \big(1 - H_1(x)\big)$$
$$C_{GAME} = \big(1 - H_3(x)\big) \cdot \big(1 - H_2(x)\big) \cdot \big(1 - H_1(x)\big)$$

[0135]

[0136] 도 6에 예시된 아키텍처에서, 최종 신뢰도 값들은 대응하는 이진 분류 결과들($H_k(x)$)과 같을 수 있거나, 또는 모든 클래스들에 대한 신뢰도 값들의 합이 1이도록 요구된다면, 최종 신뢰도 값들은 추정된 $H_k(x)$에 기초하여 간단히 정규화될 수 있다:

$$C_{MOVIE} = H_1(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$
$$C_{MUSIC} = H_2(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$
$$C_{VOIP} = H_3(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$
$$C_{GAME} = H_4(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$

[0137]

[0138] 최대 신뢰도 값들을 가진 하나 이상은 최종 식별된 클래스인 것으로 결정될 수 있다.

[0139] 도 4 내지 도 6에 도시된 아키텍처들에서, 상이한 이진 분류기들의 시퀀스는 반드시 도시된 바와 같지는 않으며, 다른 시퀀스들일 수 있고, 이것은 다양한 출원들의 상이한 요건들에 따라 수동 할당 또는 자동 학습에 의해 선택될 수 있다는 것이 주의되어야 한다.

[0140] 상기 설명들은 오디오 콘텍스트 분류기들(204)에 관한 것이다. 오디오 콘텐트 분류기(202)에 대해, 상황이 유사하다.

[0141] 대안적으로, 오디오 콘텐트 분류기(202) 또는 오디오 콘텍스트 분류기(204)는 콘텐트 유형들/콘텍스트 유형들 모두를 동시에 식별하는 하나의 단일 분류기로서 구현될 수 있으며, 대응하는 신뢰도 값들을 동시에 제공할 수 있다. 이를 하기 위한 많은 기존의 기술들이 있다.

[0142] 신뢰도 값을 사용하여, 오디오 분류기(200)의 출력은 벡터로서 표현될 수 있으며, 각각의 치수는 각각의 타겟 오디오 유형의 신뢰도 값을 표현한다. 예를 들면, 타겟 오디오 유형들이 순차적으로 (스피치, 단기 음악, 잡음, 배경)이면, 예시적인 출력 결과는 (0.9, 0.5, 0.0, 0.0)일 수 있으며, 그것은 오디오 콘텐트가 스피치임을 90% 확신하며, 오디오가 음악임을 50% 확신한다는 것을 표시한다. 오디오 신호가 스피치 및 단기 음악의 혼합 신호일 수 있음을 의미하는, 출력 벡터에서의 치수들 모두의 합이 1일 필요가 없다는(예를 들면, 도 6으로부터의 결과들은 정규화될 필요가 없다) 것이 주의된다.

[0143] 나중에 파트 6 및 파트 7에서, 오디오 콘텍스트 분류 및 오디오 콘텐트 분류의 신규 구현이 상세히 논의될 것이다.

[0144] 1.3 오디오 유형들의 신뢰도 값들의 평활화

[0145] 선택적으로, 각각의 오디오 세그먼트가 미리 정의된 오디오 유형들로 분류된 후, 부가적인 단계는 하나의 유형에서 또 다른 것으로의 갑작스런 점프를 회피하기 위해 및 오디오 개선 디바이스들에서의 파라미터들의 보다 평활한 추정을 하기 위해 타임 라인을 따라 분류 결과들을 평활화하는 것이다. 예를 들면, 긴 발췌는 VoIP로서 분류된 단지 하나의 세그먼트를 제외하고 영화-형 미디어로서 분류되며, 그 후 갑작스런 VoIP 결정은 평활화에 의해 영화-형 미디어로 수정될 수 있다.

[0146] 그러므로, 도 7에 도시된 바와 같은 실시예의 변형에서, 유형 평활화 유닛(712)은 각각의 오디오 유형에 대해, 현재 시간에 오디오 신호의 신뢰도 값을 평활화하기 위해 추가로 제공된다.

[0147] 공통 평활화 방법은 다음과 같이, 현재의 실제 신뢰도 값 및 지난 시간의 평활화된 신뢰도 값의 가중 합을 산출하는 것과 같은, 가중 평균에 기초한다:

[0148] $$smoothConf(t) = \beta \cdot smoothConf(t-1) + (1-\beta) \cdot conf(t) \quad (3)$$

[0149] 여기에서 t는 현재 시간(현재 오디오 세그먼트)을 나타내고, t-1은 지난 시간(지난 오디오 세그먼트)을 나타내고, β는 가중치이고, conf 및 smoothConf는 각각 평활화 전 및 후의 신뢰도 값들이다.

[0150] 신뢰도 값들의 관점으로부터, 분류기들의 경판정으로부터의 결과들은 또한 신뢰도 값들을 갖고 표현될 수 있으며, 상기 값들은 0 또는 1이다. 즉, 타겟 오디오 유형이 선택되고 오디오 세그먼트에 할당된다면, 대응하는 신뢰도는 1이며; 그렇지 않다면 신뢰도는 0이다. 그러므로, 오디오 분류기(200)가 신뢰도 값을 제공하지 않으며 단지 오디오 유형에 대한 경판정을 제공할지라도, 조정 유닛(300)의 연속적 조정은 유형 평활화 유닛(712)의 평활화 동작을 통해 여전히 가능하다.

[0151] 평활화 알고리즘은 상이한 경우들에 대해 상이한 평활화 가중치를 사용함으로써 '비대칭'일 수 있다. 예를 들면, 가중 합을 산출하기 위한 가중치들은 오디오 신호의 오디오 유형의 신뢰도 값에 기초하여 적응적으로 변경될 수 있다. 현재 세그먼트의 신뢰도 값은 더 크며, 그것의 가중도 더 크다.

[0152] 또 다른 관점으로부터, 가중 합을 산출하기 위한 가중치들은 특히 오디오 개선 디바이스(들)가 하나의 단일 콘텐트 유형의 존재 또는 부재에 기초하는 대신에, 오디오 분류기(200)에 의해 식별된 대로 다수의 콘텐트 유형들에 기초하여 조정될 때, 일 오디오 유형에서 또 다른 오디오 유형으로의 상이한 전이 쌍들에 기초하여 적응적으로 변경될 수 있다. 예를 들면, 특정한 콘텍스트에서 보다 빈번하게 나타나는 오디오 유형으로부터 콘텍스트에서 그렇게 빈번하게 나타나지 않는 또 다른 오디오 유형으로의 전이에 대해, 후자의 신뢰도 값은, 그것이 단지 간헐적인 중단일 수 있기 때문에, 그것이 너무 빠르게 증가하지 않도록 평활화될 수 있다.

[0153] 또 다른 인자는 변화 레이트를 포함하여, 변화하는(증가하거나 또는 감소하는) 경향이다. 우리가 오디오 유형이 존재하게 될 때(즉, 그것의 신뢰도 값이 증가할 때) 대기 시간에 대해 더 많은 관심을 갖는다고 가정할 때, 우리는 다음의 방식으로 평활화 알고리즘을 설계할 수 있다:

[0154]
$$smoothConf(t)=\begin{cases} conf(t) & conf(t)\geq smoothConf(t-1) \\ \beta\cdot smoothConf(t-1)+(1-\beta)\cdot conf(t) & \text{그렇지 않다면} \end{cases} \quad (4)$$

[0155] 상기 공식은 신뢰도 값이 증가할 때 평활화된 신뢰도 값이 현재 상태에 빠르게 응답하도록 허용하며 신뢰도 값이 감소할 때 느리게 제거하는 것을 허용한다. 평활화 함수들의 변형들은 유사한 방식으로 쉽게 설계될 수 있다. 예를 들면, 공식(4)은 conf(t)>=smoothConf(t-1)일 때 conf(t)의 가중치가 더 커지도록 수정될 수 있다. 사실상, 공식(4)에서, 그것은 β=0이며 conf(t)의 가중치는 가장 크게, 즉 1이 되는 것으로 간주될 수 있다.

[0156] 상이한 관점으로부터, 특정한 오디오 유형의 변화하는 경향을 고려하는 것은 단지 오디오 유형들의 상이한 전이 쌍들을 고려하는 특정 예이다. 예를 들면, 유형 A의 신뢰도 값의 증가는 비-A에서 A로의 전이로서 간주될 수 있으며, 유형 A의 신뢰도 값의 감소는 A의 비-A로의 전이로서 간주될 수 있다.

[0157] 1.4 파라미터 조정

[0158] 조정 유닛(300)은 오디오 분류기(200)로부터의 획득된 결과들에 기초하여 오디오 개선 디바이스(들)(400)에 대한 적절한 파라미터들을 추정하거나 또는 조정하도록 설계된다. 상이한 조정 알고리즘들이 콘텐트 유형 또는 콘텍스트 유형을 사용함으로써 상이한 오디오 개선 디바이스들에 대해, 또는 양쪽 모두 합동 결정을 위해 설계될 수 있다. 예를 들면, 영화-형 미디어 및 장기 음악과 같은 콘텍스트 유형 정보를 갖고, 이전에 언급된 바와 같은 프리셋들이 자동으로 선택되며 대응하는 콘텐트 상에 적용될 수 있다. 이용 가능한 콘텐트 유형 정보를 갖고, 각각의 오디오 개선 디바이스의 파라미터들은 후속 파트들에 도시된 바와 같이, 보다 정교한 방식으로 튜닝될 수 있다. 콘텐트 유형 정보 및 콘텍스트 정보는 장기 및 단기 정보의 균형을 잡기 위해 조정 유닛(300)에서 추가로 공동으로 사용될 수 있다. 특정 오디오 개선 디바이스에 대한 특정 조정 알고리즘은 별개의 조정 유닛으로서 간주될 수 있거나, 또는 상이한 조정 알고리즘들은 통괄하여 통합된 조정 유닛으로서 간주될 수 있다.

[0159] 즉, 조정 유닛(300)은 적어도 하나의 콘텐트 유형의 신뢰도 값 및/또는 적어도 하나의 콘텍스트 유형의 신뢰도 값에 기초하여 오디오 개선 디바이스의 적어도 하나의 파라미터를 조정하도록 구성될 수 있다. 특정 오디오 개선 디바이스에 대해, 오디오 유형들 중 몇몇은 정보적이며, 오디오 유형들 중 몇몇은 방해적이다. 따라서, 특정 오디오 개선 디바이스의 파라미터들은 정보적 오디오 유형(들) 또는 방해 오디오 유형(들)의 신뢰도 값(들)에 양으로 또는 음으로 상관시킬 수 있다. 여기에서 "양으로 상관시키는"은 선형적 방식으로 또는 비-선형적 방식으로, 오디오 유형의 신뢰도 값의 증가 또는 감소에 따라 파라미터가 증가하거나 또는 감소하는 것을 의미한다. "음으로 상관시키는"은 선형적 방식으로 또는 비-선형적 방식으로, 각각, 오디오 유형의 신뢰도 값의 감소 또는

증가에 따라 파라미터가 증가하거나 또는 감소하는 것을 의미한다.

[0160] 여기에서, 신뢰도 값의 감소 및 증가는 양의 또는 음의 상관에 의해 조정될 파라미터들에 직접 "전송"된다. 수학에서, 이러한 상관 또는 "전송"은 선형 비례 또는 반비례, 플러스 또는 마이너스(가산 또는 감산) 동작, 곱하기 또는 나누기 동작 또는 비-선형 함수로서 구체화될 수 있다. 모든 이들 형태들의 상관은 "전달 함수"로서 불리울 수 있다. 신뢰도 값의 증가 또는 감소를 결정하기 위해, 우리는 또한 지난 신뢰도 값 또는 복수의 이력 신뢰도 값들, 또는 그것들의 수학적 변환들과 현재 신뢰도 값 또는 그것의 수학적 변환을 비교할 수 있다. 본 출원의 콘텍스트에서, 용어 "비교"는 삭감 동작을 통한 비교 또는 나누기 동작을 통한 비교를 의미한다. 우리는 차이가 0 이상인지 또는 비가 1 이상인지를 결정함으로써 증가 또는 감소를 결정할 수 있다.

[0161] 특정 구현들에서, 우리는 적절한 알고리즘(전달 함수와 같은)을 통해 신뢰도 값들 또는 그것들의 비들 또는 차이들과 파라미터들을 직접 연관시킬 수 있으며, "외부 관찰자"가 특정 신뢰도 값 및/또는 특정 파라미터가 증가하거나 또는 감소하였는지를 명확하게 아는 것은 필요하지 않다. 몇몇 특정 예들이 특정 오디오 개선 디바이스들에 대해 후속 파트들(2 내지 5)에서 주어질 것이다.

[0162] 이전 섹션에서 서술된 바와 같이, 동일한 오디오 세그먼트에 대하여, 분류기(200)는, 오디오 세그먼트가 음악 및 스피치 및 배경 사운드들과 같은, 다수의 구성요소들을 동시에 포함할 수 있으므로, 신뢰도 값들이 반드시 1에 이르는 것이 아닐 수 있는, 각각의 신뢰도 값들을 가진 다수의 오디오 유형들을 식별할 수 있다. 이러한 상황에서, 오디오 개선 디바이스들의 파라미터들은 상이한 오디오 유형들 사이에서 균형이 이루어질 것이다. 예를 들면, 조정 유닛(300)은 적어도 하나의 오디오 유형의 중요성에 기초하여 적어도 하나의 오디오 유형의 신뢰도 값들을 가중시키는 것을 통해 다수의 오디오 유형들 중 적어도 몇몇을 고려하도록 구성될 수 있다. 특정 오디오 유형이 더 중요할수록, 그에 의해 영향을 받는 파라미터들은 더 많다.

[0163] 가중치는 또한 오디오 유형들의 정보적 및 방해 효과를 반영할 수 있다. 예를 들면, 방해 오디오 유형에 대해, 마이너스 가중치가 주어질 수 있다. 몇몇 특정 예들은 특정 오디오 개선 디바이스들에 대해 후속 파트들(2 내지 5)에서 주어질 것이다.

[0164] 본 출원의 콘텍스트에서, "가중치"는 다항 분포에서의 계수들보다 더 넓은 의미를 가진다는 것을 주의하자. 다항 분포에서의 계수들 외에, 그것은 또한 지수 또는 거듭제곱의 형태를 취할 수 있다. 다항 분포에서의 계수들일 때, 가중 계수들은 정규화되거나 또는 정규화되지 않을 수 있다. 간단히 말해서, 가중치는 단지 가중된 오브젝트가 조정될 파라미터에 얼마나 많은 영향을 주는지를 나타낸다.

[0165] 몇몇 다른 실시예들에서, 동일한 오디오 세그먼트에 포함된 다수의 오디오 유형들에 대해, 그것의 신뢰도 값들은 정규화되는 것을 통해 가중치들로 변환될 수 있으며, 그 후 최종 파라미터는 각각의 오디오 유형에 대해 미리 정의되며 신뢰도 값들에 기초하여 가중치들에 의해 가중된 파라미터 사전 설정 값들의 합을 산출하는 것을 통해 결정될 수 있다. 즉, 조정 유닛(300)은 신뢰도 값들에 기초하여 다수의 오디오 유형들의 효과들을 가중시키는 것을 통해 다수의 오디오 유형들을 고려하도록 구성될 수 있다.

[0166] 가중의 특정 예로서, 조정 유닛은 신뢰도 값들에 기초하여 적어도 하나의 우세 오디오 유형을 고려하도록 구성된다. 너무 낮은 신뢰도 값들(임계치보다 낮은)을 가진 이들 오디오 유형들에 대해, 그것들은 고려되지 않을 수 있다. 이것은 임계치보다 작은 신뢰도 값들이 0으로서 설정되는 다른 오디오 유형들의 가중치들과 같다. 몇몇 특정 예들은 특정 오디오 개선 디바이스들에 대해 후속 파트들(2 내지 5)에서 주어질 것이다.

[0167] 콘텐트 유형 및 콘텍스트 유형이 함께 고려될 수 있다. 일 실시예에서, 그것들은 동일한 레벨 상에 있는 것으로 간주될 수 있으며 그것들의 신뢰도 값들은 각각의 가중치들을 가질 수 있다. 또 다른 실시예에서, 단지 지명이 보여주는 바와 같이, "콘텍스트 유형"은 "콘텐트 유형"이 위치되는 콘텍스트 또는 환경이며, 따라서 조정 유닛(200)은 상이한 콘텍스트 유형의 오디오 신호에서의 콘텐트 유형이 오디오 신호의 콘텍스트 유형에 의존하여 상이한 가중치를 할당받도록 구성될 수 있다. 일반적으로 말하면, 임의의 오디오 유형은 또 다른 오디오 유형의 콘텍스트를 구성하며, 따라서 조정 유닛(200)은 또 다른 오디오 유형의 신뢰도 값을 갖고 하나의 오디오 유형의 가중치를 변경하도록 구성될 수 있다. 몇몇 특정 예들은 특정 오디오 개선 디바이스들에 대해 후속 파트들(2 내지 5)에 주어질 것이다.

[0168] 본 출원의 콘텍스트에서, "파라미터"는 그것의 문자 그대로의 의미보다 넓은 의미를 가진다. 하나의 단일 값을 가진 파라미터 외에, 그것은 또한, 상이한 파라미터들의 세트, 상이한 파라미터들로 구성된 벡터, 또는 프로파일을 포함하여, 이전에 언급된 바와 같은 프리셋을 의미한다. 구체적으로, 후속 파트들(2 내지 5)에서, 다음의 파라미터들이 논의될 것이지만 본 출원은 이에 제한되지 않는다: 다이얼로그 강화의 레벨, 다이얼로그-강화될

주파수 대역들을 결정하기 위한 임계치들, 배경 레벨, 서라운드 부스트 양, 서라운드 버추얼라이저에 대한 시작 주파수, 볼륨 레벨러의 동적 이득 또는 동적 이득의 범위, 상기 파라미터들은 오디오 신호의 정도가 새로운 지각 가능한 오디오 이벤트, 등화 레벨, 등화 프로파일들 및 스펙트럼 균형 프리셋들을 표시한다.

[0169] 1.5 파라미터 평활화

[0170] 섹션 1.3에서, 우리는 그것의 갑작스런 변화를 회피하기 위해, 및 그에 따라 오디오 개선 디바이스(들)의 파라미터들의 갑작스런 변화를 회피하기 위해 오디오 유형의 신뢰도 값을 평활화하는 것을 논의하였다. 다른 수단들이 또한 가능하다. 하나는 오디오 유형에 기초하여 조정된 파라미터를 평활화하는 것이며, 이 섹션에서 논의될 것이고; 다른 것은 오디오 분류기의 결과들의 변화를 지연시키도록 오디오 분류기 및/또는 조정 유닛을 구성하는 것이며, 이것은 섹션 1.6에서 논의될 것이다.

[0171] 일 실시예에서, 파라미터는 다음과 같이, 전이 포인트에서 가청 아티팩트들을 도입할 수 있는 빠른 변화를 회피하기 위해 추가로 평활화될 수 있다:

[0172] $$\tilde{L}(t) = \tau\tilde{L}(t-1) + (1-\tau)L(t) \quad (3')$$

[0173] 여기에서 $\tilde{L}(t)$ 은 평활화된 파라미터이고, L(t)는 비-평활화된 파라미터이고, τ는 시간 상수를 표현한 계수이고, t는 현재 시간이며 t-1은 지난 시간이다.

[0174] 즉, 도 8에 도시된 바와 같이, 오디오 프로세싱 장치는 조정 유닛(300)에 의해 조정된 오디오 개선 디바이스(다이얼로그 강화기(402), 서라운드 버추얼라이저(404), 볼륨 레벨러(406) 및 등화기(408) 중 적어도 하나와 같은)의 파라미터에 대해, 현재 시간에 조정 유닛에 의해 결정된 파라미터 값 및 지난 시간의 평활화된 파라미터 값의 가중 합을 산출함으로써 현재 시간에 조정 유닛(300)에 의해 결정된 파라미터 값을 평활화하기 위한 파라미터 평활화 유닛(814)을 포함할 수 있다.

[0175] 시간 상수(τ)는 오디오 개선 디바이스(400)의 애플리케이션 및/또는 구현의 특정 요건에 기초한 고정된 값일 수 있다. 그것은 또한 오디오 유형에 기초하여, 특히 음악에서 스피치로, 및 스피치에서 음악으로와 같이, 하나의 오디오 유형에서 또 다른 것으로의 상이한 전이 유형들에 기초하여 적응적으로 변경될 수 있다.

[0176] 예로서 등화기를 취하자(추가 상세들은 파트 5에 나타내어질 수 있다). 등화는 스피치 콘텐트 상에서가 아닌 음악 콘텐트 상에 적용하기에 좋다. 따라서, 등화의 레벨을 평활화하기 위해, 시간 상수는 오디오 신호가 음악에서 스피치로 전이할 때 비교적 작을 수 있으며, 따라서 보다 작은 등화 레벨이 스피치 콘텐트 상에서 보다 빠르게 적용될 수 있다. 다른 한편으로, 스피치에서 음악으로의 전이를 위한 시간 상수는 전이 포인트들에서 가청 아티팩트들을 회피하기 위해 비교적 클 수 있다.

[0177] 전이 유형(예로서, 스피치에서 음악으로 또는 음악에서 스피치로)을 추정하기 위해, 콘텐트 분류 결과들이 직접 사용될 수 있다. 즉, 음악 또는 스피치로 오디오 콘텐트를 분류하는 것은 전이 유형을 얻는 것을 간단하게 한다. 보다 연속적 방식으로 전이를 추정하기 위해, 우리는 또한 오디오 유형들의 경 판정들을 직접 비교하는 대신에, 추정된 평활화되지 않은 등화 레벨에 의존할 수 있다. 일반적인 사상은, 평활화되지 않은 등화 레벨이 증가한다면, 그것은 스피치에서 음악(또는 보다 음악 형)으로의 전이를 표시한다는 것이며; 그렇지 않다면 그것은 음악에서 스피치로(또는 보다 스피치 형으로)의 전이에 더 가깝다. 상이한 전이 유형들을 구별함으로써, 시간 상수는 그에 대응하여 설정될 수 있으며, 일 예는:

[0178] $$\tau(t) = \begin{cases} \tau_1 & L(t) \geq L(t-1) \\ \tau_2 & L(t) < L(t-1) \end{cases} \quad (4')$$이다.

[0179] 여기에서 τ(t)는 콘텐트에 의존한 시변 시간 상수이고, τ1 및 τ2는 보통 τ1 > τ2를 만족하는, 두 개의 사전 설정 시간 상수 값들이다. 직관적으로, 상기 함수는 등화 레벨이 증가할 때 비교적 느린 전이, 및 등화 레벨이 감소할 때 비교적 빠른 전이를 표시하지만, 본 출원은 이에 제한되지 않는다. 뿐만 아니라, 파라미터는 등화 레벨에 제한되지 않으며, 다른 파라미터들일 수 있다. 즉, 파라미터 평활화 유닛(814)은 가중 합을 산출하기 위한 가중치들이 조정 유닛(300)에 의해 결정된 파라미터 값의 증가 또는 감소 경향에 기초하여 적응적으로 변경되도록 구성될 수 있다.

[0180] 1.6 오디오 유형들의 전이

[0181] 도 9 및 도 10을 참조하여 오디오 유형의 갑작스런 변화를 회피하며 그에 따라 오디오 개선 디바이스(들)의 파라미터들의 갑작스런 변화를 회피하기 위한 또 다른 기법이 설명될 것이다.

[0182] 도 9에 도시된 바와 같이, 오디오 프로세싱 장치(100)는 오디오 분류기(200)가 동일한 새로운 오디오 유형을 연속적으로 출력하는 지속 시간을 측정하기 위한 타이머(916)를 더 포함할 수 있으며, 여기에서 조정 유닛(300)은 새로운 오디오 유형의 지속 시간의 길이가 임계치에 도달할 때까지 현재 오디오 유형을 계속해서 사용하도록 구성될 수 있다.

[0183] 다시 말해서, 관찰(또는 지속) 단계가, 도 10에 예시된 바와 같이 도입된다. 관찰 단계(지속 시간의 길이의 임계치에 대응하는)를 갖고, 오디오 유형의 변화는, 조정 유닛(300)이 실제로 새로운 오디오 유형을 사용하기 전에, 오디오 유형이 실제로 변화하였는지를 확인하기 위해 연속적인 시간량 동안 추가로 모니터링된다.

[0184] 도 10에 도시된 바와 같이, 화살표(1)는 현재 상태가 유형(A)이며 오디오 분류기(200)의 결과가 변하지 않는 상황을 예시한다.

[0185] 현재 상태가 유형 A이며 오디오 분류기(200)의 결과가 유형 B가 된다면, 타이머(916)는 타이밍을 시작하거나, 또는 도 10에 도시된 바와 같이, 프로세스가 관찰 단계(화살표(2))에 들어가며, 행오버 카운트(cnt)의 초기 값이 설정되어, 관찰 지속 기간의 양(임계치와 같은)을 표시한다.

[0186] 그 후, 오디오 분류기(200)가 연속적으로 유형 B를 출력한다면, cnt는 cnt가 0과 같을 때까지(즉, 새로운 유형 B의 지속 시간의 길이가 임계치에 도달하는) 계속해서 감소하며(화살표(3)), 그 후 조정 유닛(300)은 새로운 오디오 유형 B(화살표(4))를 사용할 수 있거나 또는 다시 말해서, 단지 지금까지만 오디오 유형이 유형 B로 실제로 변하였다고 간주될 수 있다.

[0187] 그렇지 않다면, cnt가 제로가 되기 전에(지속 시간의 길이가 임계치에 도달하기 전에), 오디오 분류기(200)의 출력이 다시 오래된 유형 A로 돌아가게 된다면, 관찰 단계는 종료되며 조정 유닛(300)은 여전히 오래된 유형 A(화살표(5))를 사용한다.

[0188] 유형 B에서 유형 A로의 변화는 상기 설명된 프로세스와 유사할 수 있다.

[0189] 상기 프로세스에서, 임계치(또는 행오버 카운트)는 애플리케이션 요건에 기초하여 설정될 수 있다. 그것은 미리 정의된 고정된 값일 수 있다. 그것은 또한 적응적으로 설정될 수 있다. 일 변형에서, 임계치는 일 오디오 유형에서 또 다른 오디오 유형으로의 상이한 전이 쌍들에 대해 상이하다. 예를 들면, 유형 A에서 유형 B로 변할 때, 임계치는 제 1 값일 수 있으며; 유형 B에서 유형 A로 변할 때, 임계치는 제 2 값일 수 있다.

[0190] 또 다른 변형에서, 행오버 카운트(임계치)는 새로운 오디오 유형의 신뢰도 값과 음으로 상관될 수 있다. 일반적인 사상은, 신뢰도가 두 개의 유형들 사이에서 혼란스러움을 도시한다면(예로서, 신뢰도 값이 단지 약 0.5일 때), 관찰 지속 기간은 더 길 필요가 있으며; 그렇지 않다면, 지속 기간은 비교적 짧을 수 있다는 것이다. 이러한 가이드라인에 이어서, 예시적인 행오버 카운트는 다음의 공식에 의해 설정될 수 있다,

[0191] $$HangCnt = C \cdot |0.5 - Conf| + D$$

[0192] 여기에서 HangCnt는 행오버 지속 기간 또는 임계치이고, C 및 D는 애플리케이션 요건에 기초하여 설정될 수 있는 두 개의 파라미터들이며, 보통 C는 음인 반면 D는 양의 값이다.

[0193] 부수적으로, 타이머(916)(및 따라서 상기 설명된 전이 프로세스)는 오디오 프로세싱 장치의 일부이지만 오디오 분류기(200)의 밖에 있는 것으로서 상기 설명되었다. 몇몇 다른 실시예들에서, 그것은 단지 섹션 7.3에 설명된 바와 같이, 오디오 분류기(200)의 일부로서 간주될 수 있다.

[0194] 1.7 실시예들 및 애플리케이션 시나리오들의 조합

[0195] 상기 논의된 모든 실시예들 및 그것들의 변형들은 그것의 임의의 조합으로 구현될 수 있으며, 상이한 파트들/실시예들에 언급되지만 동일하거나 또는 유사한 기능들을 가진 임의의 구성요소들이 동일하거나 또는 별개의 구성요소들로서 구현될 수 있다.

[0196] 구체적으로, 앞에서 실시예들 및 그것들의 변형들을 설명할 때, 이전 실시예들 또는 변형들에서 이미 설명된 것들과 유사한 참조 부호들을 가진 이들 구성요소들은 생략되며, 단지 상이한 구성요소들이 설명된다. 사실상, 이

들 상이한 구성요소들은 다른 실시예들 또는 변형들의 구성요소들과 조합될 수 있거나, 또는 별개의 해결책들을 단독으로 구성할 수 있다. 예를 들면, 도 1 내지 도 10을 참조하여 설명된 해결책들 중 임의의 둘 이상이 서로 조합될 수 있다. 가장 완전한 해결책으로서, 오디오 프로세싱 장치는 오디오 콘텐트 분류기(202) 및 오디오 콘텍스트 분류기(204) 양쪽 모두, 뿐만 아니라 유형 평활화 유닛(712), 파라미터 평활화 유닛(814) 및 타이머(916)를 포함할 수 있다.

[0197] 이전에 언급된 바와 같이, 오디오 개선 디바이스들(400)은 다이얼로그 강화기(402), 서라운드 버추얼라이저(404), 볼륨 레벨러(406) 및 등화기(408)를 포함할 수 있다. 오디오 프로세싱 장치(100)는, 그것들에 적응된 조정 유닛(300)을 갖고, 그것들 중 임의의 하나 이상을 포함할 수 있다. 다수의 오디오 개선 디바이스들(400)을 수반할 때, 조정 유닛(300)은 각각의 오디오 개선 디바이스들(400)에 특정적인 다수의 서브-유닛들(300A 내지 300D)(도 15, 도 18, 도 20 및 도 22)을 포함하는 것으로서 간주될 수 있거나, 또는 여전히 하나의 통합된 조정 유닛으로서 간주될 수 있다. 오디오 개선 디바이스에 특정적일 때, 오디오 분류기(200), 뿐만 아니라 다른 가능한 구성요소들과 함께 조정 유닛(300)은 특정 오디오 개선 디바이스의 제어기로서 간주될 수 있으며, 이것은 후속 파트들(2 내지 5)에 상세히 논의될 것이다.

[0198] 또한, 오디오 개선 디바이스들(400)은 언급된 바와 같이 예들에 제한되지 않으며 임의의 다른 오디오 개선 디바이스를 포함할 수 있다.

[0199] 뿐만 아니라, 이미 논의된 임의의 해결책들 또는 그것의 임의의 조합들은 본 개시의 다른 파트들에 설명되거나 또는 내포된 임의의 실시예와 추가로 조합될 수 있다. 특히, 파트들(6 및 7)에 논의될 바와 같이 오디오 분류기들의 실시예들은 오디오 프로세싱 장치에서 사용될 수 있다.

[0200] 1.8 오디오 프로세싱 방법

[0201] 이전 실시예들에서 오디오 프로세싱 장치를 설명하는 프로세스에서, 또한 몇몇 프로세스들 및 방법들이 분명히 개시된다. 이후, 이들 방법들의 요약은 상기 이미 논의된 상세들 중 일부를 반복하지 않고 주어지지만, 그것은 방법들이 오디오 프로세싱 장치를 설명하는 프로세스에서 개시될지라도, 방법들은 반드시 설명된 바와 같이 이들 구성요소들을 채택하는 것은 아니거나 또는 반드시 이들 구성요소들에 의해 실행되는 것은 아님이 주의될 것이다. 예를 들면, 오디오 프로세싱 장치의 실시예들은 하드웨어 및/또는 펌웨어를 갖고 부분적으로 또는 완전히 실현될 수 있는 반면, 이하에 논의되는 오디오 프로세싱 방법은, 방법들이 또한 오디오 프로세싱 장치의 하드웨어 및/또는 펌웨어를 채택할 수 있을지라도, 컴퓨터-실행 가능한 프로그램에 의해 완전히 실현될 수 있는 것이 가능하다.

[0202] 방법들은 도 11 내지 도 14를 참조하여 이하에서 설명될 것이다. 오디오 신호의 스트리밍 속성에 부응하여, 다양한 동작들이, 방법이 실시간으로 구현될 때 반복되며, 상이한 동작들이 반드시 동일한 오디오 세그먼트에 대한 것은 아님을 주의하자.

[0203] 도 11에 도시된 바와 같은 실시예에서, 오디오 프로세싱 방법이 제공된다. 첫 번째로, 프로세싱될 오디오 신호는 적어도 하나의 오디오 유형으로 실시간으로 분류된다(동작(1102)). 적어도 하나의 오디오 유형의 신뢰도 값에 기초하여, 오디오 개선을 위한 적어도 하나의 파라미터가 연속적으로 조정될 수 있다(동작(1104)). 오디오 개선은 다이얼로그 강화(동작(1106)), 서라운드 버추얼라이징(동작(1108)), 볼륨 레벨링(1110) 및/또는 등화(동작(1112))일 수 있다. 그에 따라, 적어도 하나의 파라미터가 다이얼로그 강화 프로세싱, 서라운드 버추얼라이징 프로세싱, 볼륨 레벨링 프로세싱 및 등화 프로세싱 중 적어도 하나에 대한 적어도 하나의 파라미터를 포함할 수 있다.

[0204] 여기에서, "실시간으로" 및 "연속적으로"는 오디오 유형을 의미하며, 따라서 파라미터는 오디오 신호의 특정 콘텐트에 따라 실시간으로 변할 것이며, "연속적으로"는 또한 조정이 갑작스런 또는 별개의 조정보다는, 신뢰도 값에 기초한 연속적 조정임을 의미한다.

[0205] 오디오 유형은 콘텐트 유형 및/또는 콘텍스트 유형을 포함할 수 있다. 그에 따라, 조정의 동작(1104)은 적어도 하나의 콘텐트 유형의 신뢰도 값 및 적어도 하나의 콘텍스트 유형의 신뢰도 값에 기초하여 적어도 하나의 파라미터를 조정하도록 구성될 수 있다. 콘텐트 유형은 단기 음악, 스피치, 배경 사운드 및 잡음의 콘텐트 유형들 중 적어도 하나를 더 포함할 수 있다. 콘텍스트 유형은 장기 음악, 영화-형 미디어, 게임 및 VoIP의 콘텍스트 유형들 중 적어도 하나를 더 포함할 수 있다.

[0206] VoIP 및 비-VoIP를 포함한 VoIP 관련 콘텍스트 유형들, 및 고-품질 오디오 또는 저-품질 오디오를 포함한 오디

오 품질 유형들처럼, 몇몇 다른 콘텍스트 유형 기법들이 또한 제안된다.

[0207] 단기 음악은 상이한 표준들에 따라 서브-유형들로 추가로 분류될 수 있다. 우세 소스의 존재에 의존하여, 그것은 우세 소스들이 없는 음악 및 우세 소스들을 가진 음악을 포함할 수 있다. 또한, 단기 음악은 음악의 리듬, 템포, 음색 및/또는 임의의 다른 음악적 속성들에 기초하여 분류된 적어도 하나의 장르-기반 클러스터 또는 적어도 하나의 악기-기반 클러스터 또는 적어도 하나의 음악 클러스터를 포함할 수 있다.

[0208] 양쪽 콘텐트 유형들 및 콘텍스트 유형들이 식별될 때, 콘텐트 유형의 중요성은 콘텐트 유형이 위치되는 콘텍스트 유형에 의해 결정될 수 있다. 즉, 상이한 콘텍스트 유형의 오디오 신호에서의 콘텐트 유형은 오디오 신호의 콘텍스트 유형에 의존하여 상이한 가중치를 할당받는다. 보다 일반적으로, 하나의 오디오 유형은 영향을 줄 수 있거나 또는 또 다른 오디오 유형의 전체일 수 있다. 그러므로, 조정의 동작(1104)은 또 다른 오디오 유형의 신뢰도 값을 갖고 일 오디오 유형의 가중치를 변경하도록 구성될 수 있다.

[0209] 오디오 신호가 다수의 오디오 유형들로 동시에 분류될 때(동일한 오디오 세그먼트에 대한 것인), 조정의 동작(1104)은 상기 오디오 세그먼트를 개선하기 위한 파라미터(들)를 조정하기 위해 식별된 오디오 유형들 중 일부 또는 모두를 고려한다. 예를 들면, 조정의 동작(1104)은 적어도 하나의 오디오 유형의 중요성에 기초하여 적어도 하나의 오디오 유형의 신뢰도 값들을 가중시키도록 구성될 수 있다. 또는, 조정의 동작(1104)은 그것들의 신뢰도 값들에 기초하여 그것들을 가중시키는 것을 통해 오디오 유형들의 적어도 일부를 고려하도록 구성될 수 있다. 특수한 경우에, 조정의 동작(1104)은 신뢰도 값들에 기초하여 적어도 하나의 우세 오디오 유형을 고려하도록 구성될 수 있다.

[0210] 결과들의 갑작스런 변화들을 회피하기 위해, 평활화 기법들이 도입될 수 있다.

[0211] 조정된 파라미터 값은 평활화될 수 있다(도 12의 동작(1214)). 예를 들면, 현재 시간에 조정의 동작(1104)에 의해 결정된 파라미터 값은 현재 시간에 조정의 동작에 의해 결정된 파라미터 값 및 지난 시간의 평활화된 파라미터 값의 가중 합으로 대체될 수 있다. 따라서, 반복된 평활화 동작을 통해, 파라미터 값이 타임 라인 상에서 평활화된다.

[0212] 가중 합을 산출하기 위한 가중치들이 오디오 신호의 오디오 유형에 기초하여, 또는 일 오디오 유형에서 또 다른 오디오 유형으로의 상이한 전이 쌍들에 기초하여 적응적으로 변경될 수 있다. 대안적으로, 가중 합을 산출하기 위한 가중치들은 조정의 동작에 의해 결정된 파라미터 값의 증가 또는 감소 경향에 기초하여 적응적으로 변경된다.

[0213] 또 다른 평활화 기법이 도 13에 도시된다. 즉, 방법은, 각각의 오디오 유형에 대해, 현재에 실제 신뢰도 값 및 지난 시간의 평활화된 신뢰도 값의 가중 합을 산출함으로써 현재 시간에 오디오 신호의 신뢰도 값을 평활화하는 단계를 더 포함할 수 있다(동작(1303)). 파라미터 평활화 동작(1214)과 유사하게, 가중 합을 산출하기 위한 가중치들은 오디오 신호의 오디오 유형의 신뢰도 값에 기초하여, 또는 일 오디오 유형에서 또 다른 오디오 유형으로의 상이한 전이 쌍들에 기초하여 적응적으로 변경될 수 있다.

[0214] 또 다른 평활화 기법은 오디오 분류 동작(1102)의 출력이 변할지라도 일 오디오 유형에서 또 다른 오디오 유형으로의 전이를 지연시키기 위한 버퍼 메커니즘이다. 즉, 조정의 동작(1104)은 새로운 오디오 유형을 즉시 사용하지 않지만 오디오 분류 동작(1102)의 출력의 안정화를 기다린다.

[0215] 구체적으로, 방법은 분류 동작이 동일한 새로운 오디오 유형을 연속적으로 출력하는 지속 시간을 측정하는 단계(도 14에서의 동작(1403))를 포함할 수 있으며, 상기 조정의 동작(1104)은 새로운 오디오 유형의 지속 시간의 길이가 임계치(동작(14035) 및 동작(11042)에서 "Y")에 도달할 때까지 현재 오디오 유형(동작(14035) 및 동작(11041)에서 "N")을 계속해서 사용하도록 구성된다. 구체적으로, 오디오 분류 동작(1102)으로부터 출력된 오디오 유형이 오디오 파라미터 조정 동작(1104)(동작(14031)에서 "Y")에서 사용된 현재 오디오 유형에 대하여 변할 때, 그 후 타이밍이 시작된다(동작(14032)). 오디오 분류 동작(1102)이 새로운 오디오 유형을 계속해서 출력하면, 즉 동작(14031)에서의 판단이 계속해서 "Y"이면, 타이밍이 계속된다(동작(14032)). 마지막으로, 새로운 오디오 유형의 지속 시간이 임계치(동작(14035)에서 "Y")에 도달할 때, 조정 동작(1104)은 새로운 오디오 유형(동작(11042))을 사용하며, 타이밍은 오디오 유형의 다음 스위치를 준비하기 위해 리셋된다(동작(14034)). 임계치(동작(14035)에서 "N")에 도달하기 전에, 조정 동작(1104)은 현재 오디오 유형을 계속해서 사용한다(동작(11041)).

[0216] 여기에서 타이밍은 타이머의 메커니즘(카운트 업 또는 카운트 다운)을 갖고 구현될 수 있다. 타이밍이 시작한 후 그러나 임계치에 도달하기 전에, 오디오 분류 동작(1102)이 조정 동작(1104)에서 사용된 현재 오디오 유형으

로 돌아가게 된다면, 조정 동작(1104)에 의해 사용된 현재 오디오 유형에 대하여 어떤 변화(동작(14031)에서 "N")도 없다고 간주되어야 한다. 그러나 현재 분류 결과(오디오 신호에서 분류될 현재 오디오 세그먼트에 대응하는)는 오디오 분류 동작(1102)(동작(14033)에서 "Y")의 이전 출력(오디오 신호에서 분류될 이전 오디오 세그먼트에 대응하는)에 대하여 변하며, 따라서 타이밍은, 다음 변화(동작(14031)에서 "Y")가 타이밍을 시작할 때까지, 리셋된다(동작(14034)). 물론, 오디오 분류 동작(1102)의 분류 결과가 오디오 파라미터 조정 동작(1104)(동작(14031)에서 "N")에 의해 사용된 현재 오디오 유형에 대하여 변하지 않으며, 이전 분류(동작(14033)에서 "N")에 대하여도 변하지 않는다면, 그것은 오디오 분류가 안정된 상태에 있으며 현재 오디오 유형이 계속해서 사용됨을 보여준다.

[0217] 여기에서 사용된 임계치는 또한, 상태가 그렇게 안정되지 않을 때, 일반적으로 우리는 오디오 개선 디바이스가 다른 것들보다 그것의 디폴트 상태들에 있는 것을 선호할 수 있기 때문에, 일 오디오 유형에서 또 다른 오디오 유형으로의 상이한 전이 쌍들에 대해 상이할 수 있다. 다른 한편으로, 새로운 오디오 유형의 신뢰도 값이 비교적 높다면, 새로운 오디오 유형으로 전이하는 것이 더 안전하다. 그러므로, 임계치는 새로운 오디오 유형의 신뢰도 값과 음으로 상관될 수 있다. 신뢰도 값이 높을수록, 임계치는 더 낮으며, 이는 오디오 유형이 새로운 것으로 보다 빨리 전이할 수 있음을 의미한다.

[0218] 오디오 프로세싱 장치의 실시예들과 유사하게, 오디오 프로세싱 방법 및 그것들의 변형들의 실시예들의 임의의 조합이 한편으로는 실현 가능하며, 다른 한편으로, 오디오 프로세싱 방법 및 그것들의 변형들의 실시예들의 모든 측면은 별개의 해결책들일 수 있다. 특히, 오디오 프로세싱 방법들 모두에서, 파트 6 및 파트 7에 논의된 바와 같은 오디오 분류 방법들이 사용될 수 있다.

[0219] 파트 2: 다이얼로그 강화기 제어기 및 제어 방법

[0220] 오디오 개선 디바이스의 일 예는 다이얼로그 강화기(DE)이며, 이것은 재생 중인 오디오를 계속해서 모니터링하는 것, 다이얼로그의 존재를 검출하는 것, 및 특히, 감소하는 청력을 가진 연장자들을 위해, 그것들의 명료성 및 양해도(다이얼로그를 듣고 이해하는데 더 용이하게 하는)를 증가시키기 위해 다이얼로그를 강화하는 것을 목표로 한다. 다이얼로그가 존재하는지를 검출하는 것 외에, 양해도에 가장 중요한 주파수들이 또한 다이얼로그가 존재한다면 검출되며, 그 후 그에 따라 강화된다(동적 스펙트럼 재균형을 갖고). 예시적인 다이얼로그 강화 방법은 WO 2008/106036 A2로서 공개된, H. Muesch의 "엔터테인먼트 오디오에서의 스피치 강화"에 제공되며, 그 전체는 여기에 참조로서 통합된다.

[0221] 다이얼로그 강화기에 대한 일반적인 수동 구성은, 다이얼로그 강화가 음악 신호들에 대해 너무 많이 거짓으로 트리거할 수 있기 때문에, 그것이 보통 영화-형 미디어 콘텐트에 대해 가능하게 되지만 음악 콘텐트에 대해 불능된다는 것이다.

[0222] 이용 가능한 오디오 유형 정보를 갖고, 다이얼로그 강화 및 다른 파라미터들의 레벨이 식별된 오디오 유형들의 신뢰도 값들에 기초하여 튜닝될 수 있다. 이전 논의된 오디오 프로세싱 장치 및 방법의 특정 예로서, 다이얼로그 강화기는 파트 1에 논의된 실시예들 모두 및 이들 실시예들의 임의의 조합들을 이용할 수 있다. 구체적으로, 다이얼로그 강화기를 제어하는 경우에, 도 1 내지 도 10에 도시된 바와 같이 오디오 프로세싱 장치(100)에서의 오디오 분류기(200) 및 조정 유닛(300)은 도 15에 도시된 바와 같이 다이얼로그 강화기 제어기(1500)를 구성할 수 있다. 이 실시예에서, 조정 유닛은 다이얼로그 강화기에 특정적이므로, 그것은 300A로서 불리울 수 있다. 이전 파트에 논의된 바와 같이, 오디오 분류기(200)는 오디오 콘텐트 분류기(202) 및 오디오 콘텍스트 분류기(204) 중 적어도 하나를 포함할 수 있으며, 다이얼로그 강화기 제어기(1500)는 유형 평활화 유닛(712), 파라미터 평활화 유닛(814) 및 타이머(916) 중 적어도 하나를 더 포함할 수 있다.

[0223] 그러므로, 이 파트에서, 우리는 이전 파트에서 이미 설명된 이들 콘텐트들을 반복하지 않을 것이며, 단지 그것의 몇몇 특정 예들을 제공할 것이다.

[0224] 다이얼로그 강화기에 대해, 조정 가능한 파라미터들이 이에 제한되지 않지만 다이얼로그 강화의 레벨, 배경 레벨, 및 강화될 주파수 대역들을 결정하기 위한 임계치들을 포함한다. 그 전체가 여기에 참조로서 통합되는, WO 2008/106036 A2로서 공개된 H. Muesch의, "엔터테인먼트 오디오에서의 스피치 강화"를 참조하자.

[0225] 2.1 다이얼로그 강화의 레벨

[0226] 다이얼로그 강화의 레벨을 수반할 때, 조정 유닛(300A)은 스피치의 신뢰도 값과 다이얼로그 강화기의 다이얼로그 강화의 레벨을 양으로 상관시키도록 구성될 수 있다. 부가적으로 또는 대안적으로, 레벨은 다른 콘텐트 유형들의 신뢰도 값에 음으로 상관될 수 있다. 따라서, 다이얼로그 강화의 레벨은 스피치 신뢰도에 비례(선형적으로

또는 비-선형적으로)하도록 설정될 수 있으며, 따라서 다이얼로그 강화는 음악 및 배경 사운드(사운드 효과들)와 같은, 비-스피치 신호들에서 덜 효과적이다.

[0227] 콘텍스트 유형에 대해, 조정 유닛(300A)은 영화-형 미디어 및/또는 VoIP의 신뢰도 값과 다이얼로그 강화기의 다이얼로그 강화의 레벨을 양으로 상관시키거나, 및 또는 장기 음악 및/또는 게임의 신뢰도 값과 다이얼로그 강화기의 다이얼로그 강화의 레벨을 음으로 상관시키도록 구성될 수 있다. 예를 들면, 다이얼로그 강화의 레벨은 영화-형 미디어의 신뢰도 값이 비례(선형적으로 또는 비-선형적으로)하도록 설정될 수 있다. 영화-형 미디어 신뢰도 값이 0(예로서, 음악 콘텐트에서)일 때, 다이얼로그 강화의 레벨이 또한 0이며, 이것은 다이얼로그 강화기를 불능시키는 것과 같다.

[0228] 이전 파트에서 설명된 바와 같이, 콘텐트 유형 및 콘텍스트 유형은 함께 고려될 수 있다.

[0229] 2.2 강화될 주파수 대역들을 결정하기 위한 임계치들

[0230] 다이얼로그 강화기의 작동 동안, 그것이 강화될 필요가 있는지를 결정하기 위해 각각의 주파수 대역에 대한 임계치(보통 에너지 또는 라우드니스 임계치)가 있으며, 즉 각각의 에너지/라우드니스 임계치들 이상의 이들 주파수 대역들이 강화될 것이다. 임계치들을 조정하는 것에 대해, 조정 유닛(300A)은 단기 음악 및/또는 잡음 및/또는 배경 사운드들의 신뢰도 값과 임계치들을 양으로 상관시키고 및/또는 스피치의 신뢰도 값과 임계치들을 음으로 상관시키도록 구성될 수 있다. 예를 들면, 임계치들은, 보다 많은 주파수 대역들이 강화되도록 허용하기 위해, 보다 신뢰성 있는 스피치 검출을 가정할 때, 스피치 신뢰도가 높다면 아래로 낮춰질 수 있으며; 다른 한편으로, 음악 신뢰도 값이 높을 때, 임계치들은 보다 적은 주파수 대역들이 강화되게 하기 위해(및 그에 따라 보다 적은 아티팩트들) 증가될 수 있다.

[0231] 2.3 배경 레벨에 대한 조정

[0232] 다이얼로그 강화기에서의 또 다른 구성요소는 도 15에 도시된 바와 같이, 최소 추적 유닛(4022)이며, 이것은 오디오 신호에서의 배경 레벨을 추정하기 위해 사용된다(SNR 추정, 및 섹션 2.2에 언급된 바와 같은 주파수 대역 임계치 추정). 그것은 또한 오디오 콘텐트 유형들의 신뢰도 값들에 기초하여 튜닝될 수 있다. 예를 들면, 스피치 신뢰도가 높다면, 최소 추적 유닛은 배경 레벨을 현재 최소치로 설정하기에 더 확신적일 수 있다. 음악 신뢰도가 높다면, 배경 레벨은 상기 현재 최소치보다 약간 높게 설정될 수 있거나, 또는 또 다른 방식으로, 현재 최소치에 대한 큰 가중치로, 현재 프레임의 에너지 및 현재 최소치의 가중 평균으로 설정될 수 있다. 잡음 및 배경 신뢰도가 높다면, 상기 배경 레벨은 상기 현재 최소치보다 훨씬 더 높게 설정될 수 있으며, 또는 다른 방법으로, 현재 최소치에 대한 작은 가중치로, 현재 프레임의 에너지 및 현재 최소치의 가중 평균으로 설정될 수 있다.

[0233] 따라서, 조정 유닛(300A)은 최소 추적 유닛에 의해 추정된 배경 레벨에 조정을 할당하도록 구성될 수 있으며, 여기에서 조정 유닛은 단기 음악 및/또는 잡음 및/또는 배경 사운드의 신뢰도 값과 조정을 양으로 상관시키고, 및/또는 스피치의 신뢰도 값과 조정을 음으로 상관시키도록 추가로 구성된다. 변형에서, 조정 유닛(300A)은 단기 음악보다 더 양으로 잡음 및/또는 배경의 신뢰도 값과 조정을 상관시키도록 구성될 수 있다.

[0234] 2.4 실시예들 및 애플리케이션 시나리오들의 조합

[0235] 파트 1과 유사하게, 상기 논의된 모든 실시예들 및 그것들의 변형들은 그것의 임의의 조합으로 구현될 수 있으며, 상이한 파트들/실시예들에 언급되지만 동일한 또는 유사한 기능들을 가진 임의의 구성요소들이 동일한 또는 별개의 구성요소들로서 구현될 수 있다.

[0236] 예를 들면, 섹션 2.1 내지 섹션 2.3에 설명된 해결책들 중 임의의 둘 이상은 서로 조합될 수 있다. 이들 조합들은 파트 1 및 나중에 설명될 다른 파트들에 설명되거나 또는 내포된 임의의 실시예와 추가로 조합될 수 있다. 특히, 많은 공식들이 실제로 오디오 개선 디바이스 또는 방법의 각각의 종류에 적용 가능하지만, 그것들은 반드시 본 개시의 각각의 파트에서 열거되거나 논의되는 것은 아니다. 이러한 상황에서, 상호-참조는 하나의 파트 내지 또 다른 파트에서 논의된 특정 공식을 적용하기 위해 본 개시의 파트들 중에서 이루어질 수 있으며, 단지 관련된 파라미터(들), 계수(들), 거듭제곱(들)(지수들) 및 가중(들)만이 특정 애플리케이션의 특정 요건들에 따라 적절히 조정된다.

[0237] 2.5 다이얼로그 강화기 제어 방법

[0238] 파트 1과 유사하게, 상기 실시예들에서 다이얼로그 강화기 제어기를 설명하는 프로세스에서, 몇몇 프로세스들 또는 방법들이 또한 분명히 개시된다. 이후 이들 방법들의 요약은 이전에 이미 논의된 상세들 중 일부를 반복하

지 않고 주어진다.

[0239] 첫 번째로, 파트 1에 논의된 바와 같이 오디오 프로세싱 방법의 실시예들은 다이얼로그 강화기를 위해 사용될 수 있으며, 그 파라미터(들)는 오디오 프로세싱 방법에 의해 조정될 타겟들 중 하나이다. 이러한 관점으로부터, 오디오 프로세싱 방법이 또한 다이얼로그 강화기 제어 방법이다.

[0240] 이러한 섹션에서, 다이얼로그 강화기의 제어에 특정적인 이들 측면들만이 논의될 것이다. 제어 방법의 일반적인 측면들에 대해, 참조가 파트 1에 대해 이루어질 수 있다.

[0241] 일 실시예에 따르면, 오디오 프로세싱 방법은 다이얼로그 강화 프로세싱을 더 포함할 수 있으며, 조정의 동작(1104)은 영화-형 미디어 및/또는 VoIP의 신뢰도 값과 다이얼로그 강화의 레벨을 양으로 상관시키는 것, 및 또는 장기 음악 및/또는 게임의 신뢰도 값과 다이얼로그 강화의 레벨을 음으로 상관시키는 것을 포함한다. 즉, 다이얼로그 강화는 주로 영화-형 미디어, 또는 VoIP의 콘텍스트에서 오디오 신호에 관한 것이다.

[0242] 보다 구체적으로, 조정의 동작(1104)은 스피치의 신뢰도 값과 다이얼로그 강화기의 다이얼로그 강화의 레벨을 양으로 상관시키는 것을 포함할 수 있다.

[0243] 본 출원은 또한 다이얼로그 강화 프로세싱에서 강화될 주파수 대역들을 조정할 수 있다. 도 16에 도시된 바와 같이, 각각의 주파수 대역들이 강화될지를 결정하기 위한 임계치들(보통 에너지 또는 라우드니스)은 본 출원에 따라 식별된 오디오 유형들의 신뢰도 값(들)에 기초하여 조정될 수 있다(동작(1602)). 그 후, 다이얼로그 강화기 내에서, 조정된 임계치들에 기초하여, 각각의 임계치들 이상의 주파수 대역들이 선택되고(동작(1604)) 강화된다(동작(1606)).

[0244] 구체적으로, 조정의 동작(1104)은 단기 음악 및/또는 잡음 및/또는 배경 사운드들의 신뢰도 값과 임계치들을 양으로 상관시키는 것, 및/또는 스피치의 신뢰도 값과 임계치들을 음으로 상관시키는 것을 포함할 수 있다.

[0245] 오디오 프로세싱 방법(특히 다이얼로그 강화 프로세싱)은 일반적으로 오디오 신호에서 배경 레벨을 추정하는 단계를 더 포함하며, 이것은 일반적으로 다이얼로그 강화기(402)에서 실현된 최소 추적 유닛(4022)에 의해 구현되며 SNR 추정 또는 주파수 대역 임계 추정에서 사용된다. 본 출원은 또한 배경 레벨을 조정하기 위해 사용될 수 있다. 이러한 상황에서, 배경 레벨이 추정된 후(동작(1702)), 그것은 먼저 오디오 유형(들)의 신뢰도 값(들)에 기초하여 조정되며(동작(1704)), 그 후 SNR 추정 및/또는 주파수 대역 임계 추정에 사용된다(동작(1706)). 구체적으로, 조정의 동작(1104)은 추정된 배경 레벨에 조정을 할당하도록 구성될 수 있으며, 여기에서 조정의 동작(1104)은 단기 음악 및/또는 잡음 및/또는 배경 사운드의 신뢰도 값과 조정을 양으로 상관시키고, 및/또는 스피치의 신뢰도 값과 조정을 음으로 상관시키도록 추가로 구성될 수 있다.

[0246] 보다 구체적으로, 조정의 동작(1104)은 단기 음악보다 더 양으로 잡음 및/또는 배경의 신뢰도 값과 조정을 상관시키도록 구성될 수 있다.

[0247] 오디오 프로세싱 장치의 실시예들과 유사하게, 오디오 프로세싱 방법의 실시예들 및 그것들의 변형들의 임의의 조합은 한편으로는 실현 가능하며; 다른 한편으로, 오디오 프로세싱 방법의 실시예들 및 그것들의 변형들의 모든 측면들은 별개의 해결책들일 수 있다. 또한, 이 섹션에 설명된 임의의 둘 이상의 해결책들은 서로 조합될 수 있으며, 이들 조합들은 파트 1 및 나중에 설명될 다른 파트들에 설명되거나 또는 내포된 임의의 실시예와 추가로 조합될 수 있다.

[0248] 파트 3: 서라운드 버추얼라이저 제어기 및 제어 방법

[0249] 서라운드 버추얼라이저는 서라운드 사운드 신호(다채널 5.1 및 7.1과 같은)가 PC의 내부 스피커들을 통해 또는 헤드폰들을 통해 렌더링될 수 있게 한다. 즉, 내부 랩탑 스피커들 또는 헤드폰들과 같은 스테레오 디바이스들을 갖고, 그것은 가상으로 사운드 효과를 생성하며 소비자들을 위한 영화적 경험을 제공한다. 헤드 관련 전달 함수들(Head Related Transfer Functions; HRTF들)은 보통 다-채널 오디오 신호와 연관된 다양한 스피커 위치들로부터 온 귀들에서의 사운드의 도착을 시뮬레이션하기 위해 서라운드 버추얼라이저에서 이용된다.

[0250] 현재 서라운드 버추얼라이저는 헤드폰들 상에서 잘 작동하지만, 그것은 내장 스피커들을 가진 상이한 콘텐트들 상에서 상이하게 작동한다. 일반적으로, 영화-형 미디어 콘텐트는 스피커들을 위한 서라운드 버추얼라이저를 가능하게 하지만, 음악은 그것이 너무 희박하게 들릴 수 있으므로 아니다.

[0251] 서라운드 버추얼라이저에서의 동일한 파라미터들이 영화-형 미디어 및 음악 콘텐트 양쪽 모두에 대한 양호한 사운드 이미지를 동시에 생성할 수 없으므로, 파라미터들은 콘텐트에 기초하여 보다 정확하게 튜닝되어야 한다.

이용 가능한 오디오 유형 정보, 특히 음악 신뢰도 값 및 스피치 신뢰도 값, 뿐만 아니라 몇몇 다른 콘텐트 유형 정보 및 콘텍스트 정보를 갖고, 작업이 본 출원에 따라 행해질 수 있다.

[0252] 파트 2와 유사하게, 파트 1에 논의된 오디오 프로세싱 장치 및 방법의 특정 예로서, 서라운드 버추얼라이저(404)는 파트 1에 논의된 실시예들의 모두 및 그 안에 개시된 이들 실시예들의 임의의 조합들을 이용할 수 있다. 구체적으로, 서라운드 버추얼라이저(404)를 제어하는 경우에, 도 1 내지 도 10에 도시된 바와 같이 오디오 프로세싱 장치(100)에서의 오디오 분류기(200) 및 조정 유닛(300)은 도 18에 도시된 바와 같이 서라운드 버추얼라이저 제어기(1800)를 구성할 수 있다. 이 실시예에서, 조정 유닛은 서라운드 버추얼라이저(404)에 특정적이기 때문에, 그것은 300B로서 불리울 수 있다. 파트 2와 유사하게, 오디오 분류기(200)는 오디오 콘텐트 분류기(202) 및 오디오 콘텍스트 분류기(204) 중 적어도 하나를 포함할 수 있으며, 서라운드 버추얼라이저 제어기(1800)는 유형 평활화 유닛(712), 파라미터 평활화 유닛(814) 및 타이머(916) 중 적어도 하나를 더 포함할 수 있다.

[0253] 그러므로, 이 파트에서, 우리는 파트 1에 이미 설명된 이들 콘텐트들을 반복하지 않을 것이며 단지 그것의 몇몇 특정 예들을 제공할 것이다.

[0254] 서라운드 버추얼라이저에 대해, 조정 가능한 파라미터들이 이에 제한되지 않지만 서라운드 버추얼라이저(404)에 대한 서라운드 부스트 양 및 시작 주파수를 포함한다.

[0255] 3.1 서라운드 부스트 양

[0256] 서라운드 부스트 양을 수반할 때, 조정 유닛(300B)은 잡음 및/또는 배경 및/또는 스피치의 신뢰도 값과 서라운드 버추얼라이저(404)의 서라운드 부스트 양을 양으로 상관시키고, 및/또는 단기 음악의 신뢰도 값과 서라운드 부스트 양을 음으로 상관시키도록 구성될 수 있다.

[0257] 구체적으로, 음악(콘텐트 유형)이 수용 가능하게 들리도록 서라운드 버추얼라이저(404)를 변경하기 위해, 조정 유닛(300B)의 예시적인 구현이 다음과 같은, 단기 음악 신뢰도 값에 기초하여 서라운드 부스트의 양을 튜닝할 수 있다:

[0258] $$SB \propto (1 - Conf_{music}) \quad (5)$$

[0259] 여기에서 SB는 서라운드 부스트 양을 표시하고, $Conf_{music}$은 단기 음악의 신뢰도 값이다.

[0260] 그것은 음악에 대한 서라운드 부스트를 감소시키며 그것이 없어지는 것 같은 것을 방지하도록 돕는다.

[0261] 유사하게, 스피치 신뢰도 값이 또한 이용될 수 있으며, 예를 들면:

[0262] $$SB \propto (1-Conf_{music})*Conf_{speech}^{\alpha} \quad (6)$$

[0263] 여기에서 $Conf_{speech}$는 스피치의 신뢰도 값이고, α는 지수의 형태에서의 가중 계수이며, 1-2의 범위에 있을 수 있다. 이러한 공식은 서라운드 부스트 양이 단지 완전한 스피치에 대해 높을 것임을 표시한다(높은 스피치 신뢰도 및 낮은 음악 신뢰도).

[0264] 또는 우리는 스피치의 신뢰도 값만을 고려할 수 있다:

[0265] $$SB \propto Conf_{speech} \quad (7)$$

[0266] 다양한 변형들이 유사한 방식으로 설계될 수 있다. 특히, 잡음 또는 배경 사운드에 대해, 공식들((5) 내지 (7))과 유사한 공식들이 구성될 수 있다. 또한, 4개의 콘텐트 유형들의 효과들은 임의의 조합으로 함께 고려될 수 있다. 이러한 상황에서, 잡음 및 배경은 앰비언스 사운드들이며 그것들은 큰 부스트 양을 갖기에 더 안전하고; 화자가 보통 스크린의 앞에 앉는다고 가정할 때, 스피치는 중간 부스트 양을 가질 수 있으며; 음악은 보다 적은 부스트 양을 사용한다. 그러므로, 조정 유닛(300B)은 콘텐트 유형 스피치보다 더 양으로 잡음 및/또는 배경의 신뢰도 값과 서라운드 부스트 양을 상관시키도록 구성될 수 있다.

[0267] 우리가 각각의 콘텐트 유형에 대한 예상된 부스트 양(그것은 가중치와 같다)을 미리 정의하였다고 가정하면, 또 다른 대안이 또한 적용될 수 있다:

$$\hat{a} = \frac{a_{speech} \cdot Conf_{speech} + a_{music} \cdot Conf_{music} + a_{noise} \cdot Conf_{noise} + a_{bkg} \cdot Conf_{bkg}}{Conf_{speech} + Conf_{music} + Conf_{noise} + Conf_{bkg}} \quad (8)$$

[0268]

[0269] 여기에서 $\hat{a}$는 추정된 부스트 양이고, 콘텐트 유형의 아래첨자를 가진 α는 콘텐트 유형의 예상된/미리 정의된 부스트 양(가중치)이고, 콘텐트 유형의 아래 첨자를 가진 Conf는 콘텐트 유형의 신뢰도 값이다(여기에서 bkg는 "배경 사운드"를 나타낸다). 상황들에 의존하여, a_{music}은, (반드시는 아니지만) 서라운드 버추얼라이저(404)가 순수 음악(콘텐트 유형)에 대해 불능될 것임을 표시하는, 0으로 설정될 수 있다.

[0270] 또 다른 관점으로부터, 공식(8)에서 콘텐트 유형의 아래첨자를 가진 α는 콘텐트 유형의 예상된/미리 정의된 부스트 양이며, 식별된 콘텐트 유형들 모두의 신뢰도 값들의 합으로 나누어진 대응하는 콘텐트 유형의 신뢰도 값의 몫은 대응하는 콘텐트 유형의 미리 정의된/예상된 부스트 양의 정규화된 가중치로서 간주될 수 있다. 즉, 조정 유닛(300B)은 신뢰도 값들에 기초하여 다수의 콘텐트 유형들의 미리 정의된 부스트 양들을 가중시키는 것을 통해 다수의 콘텐트 유형들 중 적어도 일부를 고려하도록 구성될 수 있다.

[0271] 콘텍스트 유형에 대해, 조정 유닛(300B)은 영화-형 미디어 및/또는 게임의 신뢰도 값과 서라운드 버추얼라이저(404)의 서라운드 부스트 양을 양으로 상관시키고, 및/또는 장기 음악 및/또는 VoIP의 신뢰도 값과 서라운드 부스트 양을 음으로 상관시키도록 구성될 수 있다. 그 후, (5) 내지 (8)과 유사한 공식들이 구성될 수 있다.

[0272] 특별한 예로서, 서라운드 버추얼라이저(404)는 완전한 영화-형 미디어 및/또는 게임에 대해 가능해질 수 있지만, 음악 및/또는 VoIP에 대해서는 불능될 수 있다. 한편, 서라운드 버추얼라이저(404)의 부스트 양은 영화-형 미디어 및 게임에 대해 상이하게 설정될 수 있다. 영화-형 미디어는 보다 높은 부스트 양을 사용하며, 게임은 보다 적은 부스트 양을 사용한다. 그러므로, 조정 유닛(300B)은 게임보다 더 양으로 영화-형 미디어의 신뢰도 값과 서라운드 부스트 양을 상관시키도록 구성될 수 있다.

[0273] 콘텐트 유형과 유사하게, 오디오 신호의 부스트 양은 또한 콘텍스트 유형들의 신뢰도 값들의 가중 평균으로 설정될 수 있다:

$$\hat{a} = \frac{a_{MOVIE} \cdot Conf_{MOVIE} + a_{MUSIC} \cdot Conf_{MUSIC} + a_{GAME} \cdot Conf_{GAME} + a_{VOIP} \cdot Conf_{VOIP}}{Conf_{MOVIE} + Conf_{MUSIC} + Conf_{GAME} + Conf_{VOIP}} \quad (9)$$

[0274]

[0275] 여기에서 $\hat{a}$는 추정된 부스트 양이고, 콘텍스트 유형의 아래첨자를 가진 α는 콘텍스트 유형의 예상된/미리 정의된 부스트 양(가중치)이고, 콘텍스트 유형의 아래 첨자를 가진 Conf는 콘텍스트 유형의 신뢰도 값이다. 상황들에 의존하여, a_{MUSIC} 및 a_{VOIP}는 (반드시는 아니지만), 서라운드 버추얼라이저(404)가 순수 음악(콘텍스트 유형) 및 또는 순수 VoIP에 대해 불능될 것임을 표시하는, 0으로 설정될 수 있다.

[0276] 다시, 콘텐트 유형과 유사하게, 공식 (9)에서 콘텍스트 유형의 아래 첨자를 가진 α는 콘텍스트 유형의 예상된/미리 정의된 부스트 양이며, 식별된 콘텍스트 유형들 모두의 신뢰도 값들의 합으로 나누어진 대응하는 콘텍스트 유형의 신뢰도 값의 몫은 대응하는 콘텍스트 유형의 미리 정의된/예상된 부스트 양의 정규화된 가중치로서 간주될 수 있다. 즉, 조정 유닛(300B)은 신뢰도 값들에 기초하여 다수의 콘텍스트 유형들의 미리 정의된 부스트 양들을 가중시키는 것을 통해 다수의 콘텍스트 유형들의 적어도 일부를 고려하도록 구성될 수 있다.

[0277] 3.2 시작 주파수

[0278] 다른 파라미터들이 또한 시작 주파수와 같은, 서라운드 버추얼라이저에서 변경될 수 있다. 일반적으로, 오디오 신호에서의 높은 주파수 구성요소들은 공간적으로 렌더링되기에 더 적합하다. 예를 들면, 음악에서, 그것은 베이스가 보다 많은 서라운드 효과들을 갖도록 공간적으로 렌더링된다면 이상하게 들릴 것이다. 그러므로, 특정한 오디오 신호에 대해, 서라운드 버추얼라이저는 주파수 임계치를 결정할 필요가 있으며, 그 이상의 구성요소들은 그 이하의 구성요소들이 보유되는 동안 공간적으로 렌더링된다. 주파수 임계치는 시작 주파수이다.

[0279] 본 출원의 실시예에 따르면, 서라운드 버추얼라이저에 대한 시작 주파수는 보다 많은 베이스가 음악 신호들을 위해 보유될 수 있도록 음악 콘텐트에 대해 증가될 수 있다. 그 후, 조정 유닛(300B)은 단기 음악의 신뢰도 값과 서라운드 버추얼라이저의 시작 주파수를 양으로 상관시키도록 구성될 수 있다.

[0280] 3.3 실시예들 및 애플리케이션 시나리오들의 조합

[0281] 파트 1과 유사하게, 상기 논의된 모든 실시예들 및 그것들의 변형들은 그것의 임의의 조합으로 구현될 수 있으

며, 상이한 파트들/실시예들에서 언급되지만 동일하거나 또는 유사한 기능들을 가진 임의의 구성요소들이 동일하거나 또는 별개의 구성요소들로서 구현될 수 있다.

[0282] 예를 들면, 섹션 3.1 및 섹션 3.2에 설명된 해결책들 중 임의의 둘 이상이 서로 조합될 수 있다. 조합들 중 임의의 것이 파트 1, 파트 2 및 나중에 설명될 다른 파트들에 설명되거나 또는 내포된 임의의 실시예와 추가로 조합될 수 있다.

[0283] 3.4 서라운드 버추얼라이저 제어 방법

[0284] 파트 1과 유사하게, 이전 실시예들에서 서라운드 버추얼라이저 제어기를 설명하는 프로세스에서, 몇몇 프로세스들 또는 방법들이 또한 분명히 개시된다. 이후, 이들 방법들의 요약은 이전 이미 논의된 상세들의 일부를 반복하지 않고 주어진다.

[0285] 첫 번째로, 파트 1에 논의된 바와 같은 오디오 프로세싱 방법의 실시예들이 서라운드 버추얼라이저를 위해 사용될 수 있으며, 그 파라미터(들)는 오디오 프로세싱 방법에 의해 조정될 타겟들 중 하나이다. 이러한 관점으로부터, 오디오 프로세싱 방법은 또한 서라운드 버추얼라이저 제어 방법이다.

[0286] 이 섹션에서, 서라운드 버추얼라이저의 제어에 특정한 이들 측면들만이 논의될 것이다. 제어 방법의 일반적인 측면들을 위해, 참조가 파트 1에 대해 이루어질 수 있다.

[0287] 일 실시예에 따르면, 오디오 프로세싱 방법은 서라운드 버추얼라이징 프로세싱을 더 포함할 수 있으며, 조정의 동작(1104)은 잡음 및/또는 배경 및/또는 스피치의 신뢰도 값과 서라운드 버추얼라이징 프로세싱의 서라운드 부스트 양을 양으로 상관시키고, 및/또는 단기 음악의 신뢰도 값과 서라운드 부스트 양을 음으로 상관시키도록 구성될 수 있다.

[0288] 구체적으로, 조정의 동작(1104)은 콘텐트 유형 스피치보다 더 양으로 잡음 및/또는 배경의 신뢰도 값과 서라운드 부스트 양을 상관시키도록 구성될 수 있다.

[0289] 대안적으로 또는 부가적으로, 서라운드 부스트 양은 또한 콘텍스트 유형(들)의 신뢰도 값(들)에 기초하여 조정될 수 있다. 구체적으로, 조정의 동작(1104)은 영화-형 미디어 및/또는 게임의 신뢰도 값과 서라운드 버추얼라이징 프로세싱의 서라운드 부스트 양을 양으로 상관시키고, 및/또는 장기 음악 및/또는 VoIP의 신뢰도 값과 서라운드 부스트 양을 음으로 상관시키도록 구성될 수 있다.

[0290] 보다 구체적으로, 조정의 동작(1104)은 게임보다 더 양으로 영화-형 미디어의 신뢰도 값과 서라운드 부스트 양을 상관시키도록 구성될 수 있다.

[0291] 조정될 또 다른 파라미터는 서라운드 가상화 프로세싱(surround virtualizing processing)을 위한 시작 주파수이다. 도 19에 도시된 바와 같이, 시작 주파수는 첫째로 오디오 유형(들)의 신뢰도 값(들)에 기초하여 조정되고(동작(1902)), 그 후 서라운드 버추얼라이저가 시작 주파수 위의 이들 오디오 구성요소들을 프로세싱한다(동작(1904)). 구체적으로, 조정의 동작(1104)은 단기 음악의 신뢰도 값과 서라운드 버추얼라이징 프로세싱의 시작 주파수를 양으로 상관시키도록 구성될 수 있다.

[0292] 오디오 프로세싱 장치의 실시예들과 유사하게, 오디오 프로세싱 방법의 실시예들 및 그것들의 변형들의 임의의 조합이 한편으로는 실현 가능하며; 다른 한편으로, 오디오 프로세싱 방법의 실시예들 및 그것들의 변형들의 모든 측면은 별개의 해결책들일 수 있다. 또한, 이 섹션에 설명된 임의의 둘 이상의 해결책들이 서로 조합될 수 있으며, 이들 조합들은 본 개시의 다른 파트들에 설명되거나 또는 내포된 임의의 실시예와 추가로 조합될 수 있다.

[0293] 파트 4: 볼륨 레벨러 제어기 및 제어 방법

[0294] 동일한 오디오 소스에서의 상이한 조각들 또는 상이한 오디오 소스들의 볼륨이 언젠가 많이 변한다. 그것은 사용자들이 빈번하게 볼륨을 조정해야 하기 때문에 성가시다. 볼륨 레벨러(VL)는 재생 중인 오디오 콘텐트의 볼륨을 튜닝하며 타겟 라우드니스 값에 기초하여 타임라인에 걸쳐 그것을 거의 일관되게 유지하는 것을 목표로 한다. 예시적인 볼륨 레벨러들이 US2009/0097676A1으로서 공개된, A. J. Seefeldt 외의 "오디오 신호의 지각된 라우드니스 및/또는 지각된 스펙트럼 균형을 산출하며 조정하는 방법"; WO2007/127023A1로서 공개된, B. G. Grockett 외의 "특정-라우드니스-기반 청각 이벤트 검출을 사용한 오디오 이득 제어"; 및 WO 2009/011827 A1으로서 공개된, A. Seefeldt 외의 "청각 장면 분석 및 스펙트럼 왜도를 사용한 오디오 프로세싱"에 제공된다. 3개의 문서들은 참조로서 여기에서 전체적으로 통합된다.

[0295] 볼륨 레벨러는 몇몇 방식으로 오디오 신호의 라우드니스를 연속적으로 측정하며 그 후 오디오 신호의 라우드니스를 변경하기 위한 스케일링 인자이며 보통 측정된 라우드니스, 원하는 타겟 라우드니스, 및 여러 개의 다른 인자들의 함수인 이득의 양만큼 신호를 변경한다. 양쪽 모두에 대한 기본 기준들을 갖고, 적절한 이득을 추정하기 위해 고려될 필요가 있는 다수의 인자들이 타겟 라우드니스에 도달하며 동적 범위를 유지한다. 그것은 보통 자동 이득 제어(AGC), 청각 이벤트 검출, 동적 범위 제어(DRC)와 같은 여러 개의 서브-요소들을 포함한다.

[0296] 제어 신호는 일반적으로 볼륨 레벨러에서 오디오 신호의 "이득"을 제어하기 위해 인가된다. 예를 들면, 제어 신호는 순수 신호 분석에 의해 도출된 오디오 신호의 규모에서의 변화의 표시자일 수 있다. 그것은 또한, 청각 장면 분석 또는 특정-라우드니스-기반 청각 이벤트 검출과 같은 음향 심리학적 분석을 통해, 새로운 오디오 이벤트가 나타나는지를 나타내기 위한 오디오 이벤트 표시자일 수 있다. 이러한 제어 신호는 볼륨 레벨러에서, 오디오 신호에서 이득의 빠른 변화로 인해 가능한 가청 아티팩트들을 감소시키기 위해, 예를 들면, 이득이 청각 이벤트 내에서 거의 일정함을 보장함으로써 및 이벤트 경계의 이웃에 대한 많은 이득 변화를 국한시킴으로써, 이득 제어를 위해 적용된다.

[0297] 그러나, 제어 신호들을 도출하는 종래의 방법들은 비-정보적(방해) 청각 이벤트들로부터 정보적 청각 이벤트들을 구별할 수 없다. 여기에서, 정보적 청각 이벤트는 의미 있는 정보를 포함하는 오디오 이벤트를 나타내며 다이얼로그 및 음악과 같은, 사용자들에 의해 보다 많은 관심을 갖게 될 수 있는 반면, 비-정보적 신호는 VoIP에서의 잡음과 같은, 사용자들에 대한 의미 있는 정보를 포함하지 않는다. 결과로서, 비-정보적 신호들은 또한 큰 이득에 의해 인가될 수 있으며 타겟 라우드니스에 가깝게 부스팅될 수 있다. 그것은 몇몇 애플리케이션들에서 만족을 주지 않을 것이다. 예를 들면, VoIP 호출들에서, 대화의 중지에 나타나는 잡음 신호는 종종 볼륨 레벨러에 의해 프로세싱된 후 큰 볼륨까지 부스팅된다. 이것은 사용자들에 의해 원치 않는다.

[0298] 적어도 부분적으로 이러한 문제를 처리하기 위해, 본 출원은 파트 1에 논의된 실시예들에 기초하여 볼륨 레벨러를 제어하도록 제안한다.

[0299] 파트 2 및 파트 3과 유사하게, 파트 1에 논의된 오디오 프로세싱 장치 및 방법의 특정 예로서, 볼륨 레벨러(406)는 파트 1에 논의된 실시예들 모두 및 그 안에 개시된 이들 실시예들의 임의의 조합들을 이용할 수 있다. 구체적으로, 볼륨 레벨러(406)를 제어하는 경우에, 도 1 내지 도 10에 도시된 바와 같이 오디오 프로세싱 장치(100)에서의 오디오 분류기(200) 및 조정 유닛(300)은 도 20에 도시된 바와 같이 볼륨 레벨러(406) 제어기(2000)를 구성할 수 있다. 이 실시예에서, 조정 유닛이 볼륨 레벨러(406)에 특정적이므로, 그것은 300C로서 불리울 수 있다.

[0300] 즉, 파트 1의 개시에 기초하여, 볼륨 레벨러 제어기(2000)는 오디오 신호의 오디오 유형(콘텐트 유형 및/또는 콘텍스트 유형과 같은)을 연속적으로 식별하기 위한 오디오 분류기(200); 및 식별된 대로 오디오 유형의 신뢰도 값에 기초하여 연속적인 방식으로 볼륨 레벨러를 조정하기 위한 조정 유닛(300C)을 포함할 수 있다. 유사하게, 오디오 분류기(200)는 오디오 콘텐트 분류기(202) 및 오디오 콘텍스트 분류기(204) 중 적어도 하나를 포함할 수 있으며, 볼륨 레벨러 제어기(2000)는 유형 평활화 유닛(712), 파라미터 평활화 유닛(814) 및 타이머(916) 중 적어도 하나를 더 포함할 수 있다.

[0301] 그러므로, 이 파트에서, 우리는 파트 1에 이미 설명된 이들 콘텐트들을 반복하지 않을 것이며 단지 그것의 몇몇 특정 예들을 제공할 것이다.

[0302] 볼륨 레벨러(406)에서의 상이한 파라미터들은 분류 결과들에 기초하여 적응적으로 튜닝될 수 있다. 우리는 예를 들면, 비-정보적 신호들에 대한 이득을 감소시킴으로써, 동적 이득 또는 동적 이득의 범위에 직접 관련된 파라미터들을 튜닝할 수 있다. 우리는 또한 새로운 지각 가능한 오디오 이벤트인 신호의 정도를 표시하는 파라미터들을 튜닝하며, 그 후 동적 이득을 간접적으로 제어할 수 있다(이득은 오디오 이벤트 내에서 느리게 변할 것이지만, 두 개의 오디오 이벤트들의 경계에서 빠르게 변할 수 있다). 본 출원에서는, 파라미터 튜닝 또는 볼륨 레벨러 제어 메커니즘의 여러 개의 실시예들이 제공된다.

[0303] 4.1 정보적 및 방해 콘텐트 유형들

[0304] 상기 언급된 바와 같이, 볼륨 레벨러의 제어와 관련되어, 오디오 콘텐트 유형들은 정보적 콘텐트 유형들 및 방해 콘텐트 유형들로서 분류될 수 있다. 조정 유닛(300C)은 오디오 신호의 정보적 콘텐트 유형들과 볼륨 레벨러의 동적 이득을 양으로 상관시키며, 오디오 신호의 방해 콘텐트 유형들과 볼륨 레벨러의 동적 이득을 음으로 상관시키도록 구성될 수 있다.

[0305] 예로서, 잡음은 방해적(비-정보적)이며 시끄러운 볼륨으로 부스팅되는 것이 성가실 것이라고 가정하면, 동적 이득을 직접 제어하는 파라미터, 또는 새로운 오디오 이벤트들을 표시하는 파라미터가 다음과 같은, 잡음 신뢰도 값($Conf_{noise}$)의 감소 함수에 비례하도록 설정될 수 있다.

[0306] $$\mathrm{GainControl} \propto 1 - Conf_{noise} \quad (10)$$

[0307] 여기에서, 간소화를 위해, 우리는, 볼륨 레벨러의 상이한 구현들이 상이한 기본 의미를 가진 파라미터들의 상이한 명칭들을 사용할 수 있으므로, 볼륨 레벨러에서 이득 제어에 관련된 파라미터들(또는 그것들의 효과들) 모두를 나타내기 위해 심볼 GainControl을 사용한다. 단일 용어 GainControl을 사용하는 것은 보편성을 잃지 않고 짧은 표현을 가질 수 있다. 본질적으로, 이들 파라미터들을 조정하는 것은 선형적 또는 비-선형적인, 원래 이득에 가중치를 적용하는 것과 같다. 일 예로서, GainControl은 GainControl이 작다면 이득이 작도록 이득을 스케일링하기 위해 직접 사용될 수 있다. 또 다른 특정 예로서, 이득은 여기에 전체적으로 참조로서 통합되는, WO2007/127023A1호로서 공개된, B.G.Grockett 외의 "특정-라우드니스-기반 청각 이벤트 검출을 사용한 오디오 이득 제어"에 설명된 이벤트 제어 신호를 GainControl을 갖고 스케일링함으로써 간접적으로 제어된다. 이 경우에, GainControl이 작을 때, 볼륨 레벨러의 이득의 제어들은 이득이 시간에 따라 상당히 변하는 것을 방지하기 위해 변경된다. GainControl이 높을 때, 제어들은 레벨러의 이득이 보다 자유롭게 변하도록 허용되도록 변경된다.

[0308] 공식(10)에 설명된 이득 제어를 갖고(원래 이득 또는 이벤트 제어 신호를 직접 스케일링하는), 오디오 신호의 동적 이득은 그것의 잡음 신뢰도 값에 상관된다(선형적 또는 비선형적으로). 신호가 높은 신뢰도 값을 가진 잡음이면, 최종 이득은 인자($1-Conf_{noise}$)로 인해 작을 것이다. 이러한 식으로, 그것은 잡음 신호를 불유쾌한 시끄러운 볼륨으로 부스팅하는 것을 회피한다.

[0309] 공식(10)으로부터의 예시적인 변형으로서, 배경 사운드가 또한 애플리케이션(VoIP와 같은)에 관심이 없다면, 그것은 유사하게 처리될 수 있으며 또한 작은 이득만큼 적용될 수 있다. 제어 함수는, 예를 들면, 잡음 신뢰도 값($Conf_{noise}$) 및 배경 신뢰도 값($Conf_{bkg}$) 양쪽 모두를 고려할 수 있다:

[0310] $$\mathrm{GainControl} \propto (1 - Conf_{noise}) \cdot (1 - Conf_{bkg}) \quad (11)$$

[0311] 상기 공식에서, 잡음 및 배경 사운드들 양쪽 모두는 원치 않기 때문에, GainControl은 잡음의 신뢰도 값 및 배경의 신뢰도 값에 의해 동일하게 영향을 받으며, 그것은 잡음 및 배경 사운드들이 동일한 가중치를 갖는다고 여겨질 수 있다. 상황들에 의존하여, 그것들은 상이한 가중치들을 가질 수 있다. 예를 들면, 우리는 잡음 및 배경 사운드들의 신뢰도 값들(또는 1을 가진 그것들의 차이)을 상이한 계수들 또는 상이한 지수(α 및 γ)를 제공할 수 있다. 즉, 공식(11)은 다음과 같이 재기록될 수 있다:

[0312] $$\mathrm{GainControl} \propto (1 - Conf_{noise})^{\alpha} \cdot (1 - Conf_{bkg})^{\gamma} \quad (12)$$

[0313] 또는

[0314] $$\mathrm{GainControl} \propto (1 - Conf_{noise}{}^{\alpha}) \cdot (1 - Conf_{bkg}{}^{\gamma}) \quad (13)$$

[0315] 대안적으로, 조정 유닛(300C)은 신뢰도 값들에 기초하여 적어도 하나의 우세 콘텐트 유형을 고려하도록 구성될 수 있다. 예를 들면:

[0316] $$\mathrm{GainControl} \propto 1 - \max(Conf_{noise}, Conf_{bkg}) \quad (14)$$

[0317] 공식(11)(및 그것의 변형들) 및 공식(14) 양쪽 모두는 잡음 신호들 및 배경 사운드 신호들에 대한 작은 이득을 표시하며, 볼륨 레벨러의 원래 행동은 GainControl이 1에 가깝도록 잡음 신뢰도 및 배경 신뢰도 양쪽 모두가 작을 때(스피치 및 음악 신호에서와 같이)만 유지된다.

[0318] 상기 예는 우세 방해 콘텐트 유형을 고려하는 것이다. 상황에 의존하여, 조정 유닛(300C)은 또한 신뢰도 값들에 기초하여 우세 정보 콘텐트 유형을 고려하도록 구성될 수 있다. 보다 일반적으로, 조정 유닛(300C)은 식별된 오디오 유형들이 정보적 및/또는 방해 오디오 유형들인지/을 포함하는지에 상관없이, 신뢰도 값들에 기초하여 적어도 하나의 우세 콘텐트 유형을 고려하도록 구성될 수 있다.

[0319] 공식(10)의 또 다른 예시적인 변형으로서, 스피치 신호가 가장 정보적 콘텐트이며 볼륨 레벨러의 디폴트 행동에 대한 보다 적은 변경을 요구한다고 가정하면, 제어 함수는 다음과 같이, 잡음 신뢰도 값($Conf_{noise}$) 및 스피치 신뢰도 값($Conf_{speech}$) 양쪽 모두를 고려할 수 있다:

[0320] $$\mathrm{GainControl} \propto 1-Conf_{noise} \cdot (1-Conf_{speech}) \quad (15)$$

[0321] 이 함수를 갖고, 작은 GainControl은 높은 잡은 신뢰도 및 낮은 스피치 신뢰도(예로서, 순수 잡음)를 갖고 이들 신호들에 대해서만 획득되며, GainControl은 스피치 신뢰도가 높다면(및 따라서 볼륨 레벨러의 원래 행동을 유지한다면) 1에 가까울 것이다. 보다 일반적으로, 그것은 ($Conf_{noise}$과 같은)하나의 콘텐트 유형의 가중치가 적어도 하나의 다른 콘텐트 유형($Conf_{speech}$와 같은)의 신뢰도 값을 갖고 변경될 수 있다고 여겨질 수 있다. 상기 공식(15)에서, 그것은 스피치의 신뢰도가 잡음의 신뢰도의 가중 계수를 변경한다고 여겨질 수 있다(공식(12 및 13)에서의 가중치들에 비교된다면 또 다른 종류의 가중치). 다시 말해서, 공식(10)에서, $Conf_{noise}$의 계수는 1로서 간주될 수 있는 반면; 공식(15)에서, 몇몇 다른 오디오 유형들(이에 제한되지 않지만, 스피치와 같은)은 잡음의 신뢰도 값의 중요성에 영향을 미칠 것이며, 따라서 우리는 $Conf_{noise}$의 가중치가 스피치의 신뢰도 값에 의해 변경된다고 말할 수 있다. 본 개시의 콘텍스트에서, 용어("가중치")는 이를 포함하는 것으로 해석될 것이다. 즉, 그것은 값의 중요성을 표시하지만, 반드시 정규화되는 것은 아니다. 참조가 섹션 1.4에 대해 이루어질 수 있다.

[0322] 또 다른 관점으로부터, 공식(12) 및 공식(13)과 유사하게, 지수들의 형태에서의 가중치들은 상이한 오디오 신호들의 우선 순위(또는 중요성)를 표시하기 위해 상기 함수에서의 신뢰도 값들 상에 적용될 수 있으며, 예를 들면, 공식(15)은 다음으로 변경될 수 있다:

[0323] $$\mathrm{GainControl} \propto 1- Conf_{noise}^{\alpha} \cdot (1- Conf_{speech})^{\gamma} \quad (16)$$

[0324] 여기에서 α 및 γ는 두 개의 가중치들이며, 이것은 그것이 레벨러 파라미터들을 변경하기에 보다 응답적인 것으로 예상된다면 더 작게 설정될 수 있다.

[0325] 공식들(10 내지 16)은 상이한 애플리케이션들에서 적절할 수 있는 다양한 제어 함수들을 형성하기 위해 자유롭게 조합될 수 있다. 음악 신뢰도 값과 같은, 다른 오디오 콘텐트 유형들의 신뢰도 값들은 또한 유사한 방식으로 제어 함수들에 쉽게 통합될 수 있다.

[0326] GainControl이 새로운 지각 가능한 오디오 이벤트인 신호의 정도를 표시하는 파라미터들을 튜닝하기 위해 사용되며, 그 후 동적 이득을 간접적으로 제어하는 경우에(이득은 오디오 이벤트 내에서 느리게 변할 것이지만, 두 개의 오디오 이벤트들의 경계에서 빠르게 변할 수 있다), 그것은 콘텐트 유형들의 신뢰도 값 및 최종 동적 이득 사이에 또 다른 전달 함수가 있다고 여겨질 수 있다.

[0327] 4.2 상이한 콘텍스트들에서의 콘텐트 유형들

[0328] 공식(10) 내지 공식(16)에서의 상기 제어 함수들은 잡음, 배경 사운드들, 단기 음악, 및 스피치와 같은, 오디오 콘텐트 유형들의 신뢰도 값들을 고려하지만, 영화-형 미디어 및 VoIP와 같은, 사운드들이 온 그것들의 오디오 콘텍스트들을 고려하지 않는다. 동일한 오디오 콘텐트 유형이 상이한 오디오 콘텍스트들, 예로서 배경 사운드들에서 상이하게 프로세싱되도록 요구할 수 있다는 것이 가능하다. 배경 사운드는 자동차 엔진, 폭발, 및 박수와 같은 다양한 사운드들을 포함한다. 그것은 VoIP 호출에서 의미가 없을 수 있지만, 그것은 영화-형 미디어에서 중요할 수 있다. 이것은 관심 있는 오디오 콘텍스트들이 식별될 필요가 있으며 상이한 제어 함수들이 상이한 오디오 콘텍스트들을 위해 설계될 필요가 있음을 표시한다.

[0329] 그러므로, 조정 유닛(300C)은 오디오 신호의 콘텍스트 유형에 기초하여 오디오 신호의 콘텐트 유형을 정보적 또는 방해로서 간주하도록 구성될 수 있다. 예를 들면, 잡음 신뢰도 값 및 배경 신뢰도 값을 고려하며, VoIP 및 비-VoIP 콘텍스트들을 구별함으로써, 오디오 콘텍스트-의존적 제어 함수는

if 오디오 콘텍스트가 VoIP

$$\text{GainControl} \propto 1 - \max(Conf_{noise}, Conf_{bkg})$$

else

$$\text{GainControl} \propto 1 - Conf_{noise} \quad (17)$$

[0330] 일 수 있다.

[0331] 즉, VoIP 콘텍스트에서, 잡음 및 배경 사운드들은 방해 콘텐트 유형들로서 간주되는 반면; 비-VoIP 콘텍스트에서, 배경 사운드들은 정보적 콘텐트 유형으로서 간주된다.

[0332] 또 다른 예로서, 스피치, 잡음 및 배경의 신뢰도 값들을 고려하며, VoIP 및 비-VoIP 콘텍스트들을 구별하는 오디오 콘텍스트-의존적 제어 함수는

if 오디오 콘텍스트가 VoIP

$$\text{GainControl} \propto 1 - \max(Conf_{noise}, Conf_{bkg})$$

else

$$\text{GainControl} \propto 1 - Conf_{noise} \cdot (1 - Conf_{speech}) \quad (18)$$

[0333] 일 수 있다.

[0334] 여기에서, 스피치는 정보적 콘텐트 유형으로서 강조된다.

[0335] 음악이 또한 비-VoIP 콘텍스트에서 중요한 정보적 정보라고 가정하면, 우리는 공식(18)의 제 2 파트를 다음으로 확대할 수 있다:

[0336] $$\text{GainControl} \propto 1 - Conf_{noise} \cdot (1 - \max(Conf_{speech}, Conf_{music})) \quad (19)$$

[0337] 사실상, (10) 내지 (16) 또는 그것들의 변형들에서 제어 함수들의 각각은 상이한/대응하는 오디오 콘텍스트들에서 적용될 수 있다. 따라서, 그것은 오디오 콘텍스트-의존적 제어 함수들을 형성하기 위해 다수의 조합들을 생성할 수 있다.

[0338] 공식(17) 및 공식(18)에서 구별되며 이용된 바와 같은 VoIP 및 비-VoIP 콘텍스트들 외에, 영화-형 미디어, 장기 음악, 및 게임, 또는 저-품질 오디오 및 고-품질 오디오와 같은 다른 오디오 콘텍스트들이 유사한 방식으로 이용될 수 있다.

[0339] 4.3 콘텍스트 유형들

[0340] 콘텍스트 유형들은 또한 잡음과 같은, 이들 성가신 사운드들이 너무 많이 부스팅되는 것을 회피하도록 볼륨 레벨러를 제어하기 위해 직접 사용될 수 있다. 예를 들면, VoIP 신뢰도 값은 볼륨 레벨러를 조종하기 위해 사용될 수 있으며, 그것이 그것의 신뢰도 값이 높을 때 덜 민감하게 한다.

[0341] 구체적으로, VoIP 신뢰도 값($Conf_{VOIP}$)을 갖고, 볼륨 레벨러의 레벨은 ($1-Conf_{VOIP}$)에 비례하도록 설정될 수 있다. 즉, 볼륨 레벨러는 VoIP 콘텐트에서 거의 비활성화되며(VoIP 신뢰도 값이 높을 때), 이것은 VoIP 콘텍스트에 대한 볼륨 레벨러를 불능시키는 종래의 수동 셋업(프리셋)과 일치한다.

[0342] 대안적으로, 우리는 오디오 신호들의 상이한 콘텍스트들에 대한 상이한 동적 이득 범위들을 설정할 수 있다. 일반적으로, VL(볼륨 레벨러) 양은 오디오 신호 상에 적용된 이득의 양을 추가로 조정하며, 이득 상에서의 또 다른(비선형) 가중치로서 보여질 수 있다. 일 실시예에서, 셋업은 다음과 같이 될 수 있다:

[0343] 표 1

[0344]

	영화-형 미디어	장기 음악	VOIP	게임
VL 양	고	중간	오프(또는 최저)	저

[0345] 더욱이, 예상된 VL 양이 각각의 콘텍스트 유형에 대해 미리 정의된다고 가정하자. 예를 들면, VL 양은 영화-형 미디어에 대해 1, VoIP에 대해 0, 음악에 대해 0.6, 및 게임에 대해 0.3으로서 설정되지만, 본 출원은 이에 제한되지 않는다. 예에 따르면, 영화-형 미디어의 동적 이득의 범위가 100%이면, VoIP의 동적 이득의 범위는 60% 등이다. 오디오 분류기(200)의 분류가 경 판정에 기초한다면, 동적 이득의 범위는 상기 예로서 직접 설정될 수

있다. 오디오 분류기(200)의 분류가 연 판정에 기초한다면, 범위는 콘텍스트 유형의 신뢰도 값에 기초하여 조정될 수 있다.

[0346] 유사하게, 오디오 분류기(200)는 오디오 신호로부터 다수의 콘텍스트 유형들을 식별할 수 있으며, 조정 유닛(300C)은 다수의 콘텐트 유형들의 중요성에 기초하여 다수의 콘텐트 유형들의 신뢰도 값들을 가중시킴으로써 동적 이득의 범위를 조정하도록 구성될 수 있다.

[0347] 일반적으로, 콘텍스트 유형에 대해, (10) 내지 (16)과 유사한 기능들은 또한 콘텍스트 유형들로 대체된 콘텐트 유형들을 갖고, 적절한 VL 양을 적응적으로 설정하기 위해 여기에서 사용될 수 있으며, 실제로 표 1은 상이한 콘텍스트 유형들의 중요성을 반영한다.

[0348] 또 다른 관점으로부터, 신뢰도 값은 섹션 1.4에 논의된 바와 같이 정규화된 가중치를 도출하기 위해 사용될 수 있다. 특정 양이 표 1에서의 각각의 콘텍스트 유형에 대해 미리 정의된다고 가정하면, 그 후 공식(9)과 유사한 공식이 또한 적용될 수 있다. 부수적으로, 유사한 해결책들이 또한 다수의 콘텐트 유형들 및 임의의 다른 오디오 유형들에 적용될 수 있다.

[0349] 4.4 실시예들 및 애플리케이션 시나리오들의 조합

[0350] 파트 1과 유사하게, 상기 논의된 모든 실시예들 및 그것의 변형들은 그것의 임의의 조합으로 구현될 수 있으며, 상이한 파트들/실시예들에 언급되지만 동일하거나 또는 유사한 기능들을 가진 임의의 구성요소들이 동일하거나 또는 별개의 구성요소들로서 구현될 수 있다. 예를 들면, 섹션 4.1 내지 섹션 4.3에 설명된 해결책들 중 임의의 둘 이상이 서로 조합될 수 있다. 조합들 중 임의의 것이 파트 1 내지 파트 3 및 나중에 설명될 다른 파트들에서 설명되거나 또는 내포된 임의의 실시예와 추가로 조합될 수 있다.

[0351] 도 21은 원래 단기 세그먼트(도 21(A))를 비교함으로써 애플리케이션에서 제안된 볼륨 레벨러 제어기의 효과를 예시하며, 상기 단기 세그먼트는 파라미터 변경 없이 종래의 볼륨 레벨러에 의해 프로세싱되며(도 21(B)), 단기 세그먼트는 본 출원에 제공된 바와 같이 볼륨 레벨러에 의해 프로세싱된다(도 21(C)). 보여진 바와 같이, 도 21(B)에 도시된 바와 같은 종래의 볼륨 레벨러에서, 잡음의 볼륨(오디오 신호의 후반)은 또한 부스팅되며 성가시다. 반대로, 도 21(C)에 도시된 바와 같은 새로운 볼륨 레벨러에서, 오디오 신호의 유효 파트의 볼륨은 잡음의 볼륨을 명확하게 부스팅하지 않고 부스팅되어, 청중에게 양호한 경험을 제공한다.

[0352] 4.5 볼륨 레벨러 제어 방법

[0353] 파트 1과 유사하게, 이전 실시예들에서 볼륨 레벨러 제어기를 설명하는 프로세스에서, 몇몇 프로세스들 또는 방법들이 또한 명확하게 개시된다. 이후 이들 방법들의 요약이 이전에 이미 논의된 상세들의 일부를 반복하지 않고 제공된다.

[0354] 첫 번째로, 파트 1에 논의된 바와 같이 오디오 프로세싱 방법의 실시예들은 볼륨 레벨러를 위해 사용될 수 있으며, 그 파라미터(들)는 오디오 프로세싱 방법에 의해 조정될 타겟들 중 하나이다. 이러한 관점으로부터, 오디오 프로세싱 방법은 또한 볼륨 레벨러 제어 방법이다.

[0355] 이 섹션에서, 단지 볼륨 레벨러의 제어에 특정적인 이들 측면들만이 논의될 것이다. 제어 방법의 일반적인 측면들에 대해, 참조가 파트 1에 대해 이루어질 수 있다.

[0356] 본 출원에 따르면, 볼륨 레벨러 제어 방법이 제공되며, 상기 방법은 오디오 신호의 콘텐트 유형을 실시간으로 식별하는 단계, 및 오디오 신호의 정보적 콘텐트 유형들과 볼륨 레벨러의 동적 이득을 양으로 상관시키며, 오디오 신호의 방해 콘텐트 유형들과 볼륨 레벨러의 동적 이득을 음으로 상관시킴으로써, 식별된 대로 콘텐트 유형에 기초하여 연속적인 방식으로 볼륨 레벨러를 조정하는 단계를 포함한다.

[0357] 콘텐트 유형은 스피치, 단기 음악, 잡음 및 배경 사운드를 포함할 수 있다. 일반적으로, 잡음은 방해 콘텐트 유형으로서 여겨진다.

[0358] 볼륨 레벨러의 동적 이득을 조정할 때, 그것은 콘텐트 유형의 신뢰도 값에 기초하여 직접 조정될 수 있거나, 또는 콘텐트 유형의 신뢰도 값의 전달 함수를 통해 조정될 수 있다.

[0359] 이미 설명된 바와 같이, 오디오 신호는 동시에 다수의 오디오 유형들로 분류될 수 있다. 다수의 콘텐트 유형들을 수반할 때, 조정 동작(1104)은 다수의 콘텐트 유형들의 중요성에 기초하여 다수의 콘텐트 유형들의 신뢰도 값들을 가중시키는 것을 통해, 또는 신뢰도 값들에 기초하여 다수의 콘텐트 유형들의 효과들을 가중시키는 것을 통해 다수의 오디오 콘텐트 유형들 중 적어도 일부를 고려하도록 구성될 수 있다. 구체적으로, 조정 동작(110

4)은 신뢰도 값들에 기초하여 적어도 하나의 우세 콘텐트 유형을 고려하도록 구성될 수 있다. 오디오 신호가 방해 콘텐트 유형(들) 및 정보적 콘텐트 유형(들) 양쪽 모두를 포함할 때, 조정 동작은 신뢰도 값들에 기초하여 적어도 하나의 우세 방해 콘텐트 유형을 고려하고, 및/또는 신뢰도 값들에 기초하여 적어도 하나의 우세 정보적 콘텐트 유형을 고려하도록 구성될 수 있다.

[0360] 상이한 오디오 유형들은 서로 영향을 줄 수 있다. 그러므로, 조정 동작(1104)은 적어도 하나의 다른 콘텐트 유형의 신뢰도 값을 갖고 하나의 콘텐트 유형의 가중치를 변경하도록 구성될 수 있다.

[0361] 파트 1에 설명된 바와 같이, 오디오 신호의 오디오 유형의 신뢰도 값은 평활화될 수 있다. 평활화 동작의 상세에 대해, 파트 1을 참조하자.

[0362] 방법은 오디오 신호의 콘텍스트 유형을 식별하는 단계를 더 포함할 수 있으며, 여기에서 조정 동작(1104)은 콘텍스트 유형의 신뢰도 값에 기초하여 동적 이득의 범위를 조정하도록 구성될 수 있다.

[0363] 콘텐트 유형의 역할은 그것이 위치되는 콘텍스트 유형에 의해 제한된다. 그러므로, 콘텐트 유형 정보 및 콘텍스트 유형 정보 양쪽 모두가 동시에 오디오 신호를 위해 획득될 때(그것은 동일한 오디오 세그먼트를 위한 것이다), 오디오 신호의 콘텐트 유형은 오디오 신호의 콘텍스트 유형에 기초하여 정보적 또는 간섭으로서 결정될 수 있다. 뿐만 아니라, 상이한 콘텍스트 유형의 오디오 신호에서의 콘텐트 유형은 오디오 신호의 콘텍스트 유형에 의존하여 상이한 가중치를 할당받을 수 있다. 또 다른 관점으로부터, 우리는 콘텐트 유형의 정보적 특징 또는 방해 특징을 반영하기 위해 상이한 가중치(보다 크거나 또는 보다 작은, 플러스 값 또는 마이너스 값)를 사용할 수 있다

[0364] 오디오 신호의 콘텍스트 유형은 VoIP, 영화-형 미디어, 장기 음악 및 게임을 포함할 수 있다. 콘텍스트 유형(VoIP)의 오디오 신호에서, 배경 사운드는 방해 콘텐트 유형으로서 간주되지만; 콘텍스트 유형(비-VoIP)의 오디오 신호에서, 배경 및/또는 스피치 및/또는 음악은 정보적 콘텐트 유형으로서 간주된다. 다른 콘텍스트 유형은 고-품질 오디오 또는 저-품질 오디오를 포함할 수 있다.

[0365] 다수의 콘텐트 유형들과 유사하게, 오디오 신호가 동시에(동일한 오디오 세그먼트에 대하여) 대응하는 신뢰도 값들을 가진 다수의 콘텍스트 유형들로 분류될 때, 조정 동작(1104)은 다수의 콘텍스트 유형들의 중요성에 기초하여 다수의 콘텍스트 유형들의 신뢰도 값들을 가중시키는 것을 통해, 또는 신뢰도 값들에 기초하여 다수의 콘텍스트 유형들의 효과들을 가중시키는 것을 통해 다수의 콘텍스트 유형들 중 적어도 일부를 고려하도록 구성될 수 있다. 구체적으로, 조정 동작은 신뢰도 값들에 기초하여 적어도 하나의 우세 콘텍스트 유형을 고려하도록 구성될 수 있다.

[0366] 마지막으로, 이 섹션에 논의된 바와 같은 방법의 실시예들은 파트 6 및 파트 7에 논의될 바와 같이 오디오 분류 방법을 사용할 수 있으며, 상세한 설명은 여기에서 생략된다.

[0367] 오디오 프로세싱 장치의 실시예들과 유사하게, 오디오 프로세싱 방법 및 그것들의 변형들의 실시예들의 임의의 조합이 한편으로는 실현 가능하며; 다른 한편으로, 오디오 프로세싱 방법 및 그것들의 변형들의 실시예들의 모든 측면은 별개의 해결책들일 수 있다. 또한, 이 섹션에 설명된 임의의 둘 이상의 해결책들은 서로 조합될 수 있으며, 이들 조합들은 본 개시의 다른 파트들에 설명되거나 또는 내포된 임의의 실시예와 추가로 조합될 수 있다.

[0368] 파트 5: 등화기 제어기 및 제어 방법

[0369] 등화는 보통 "톤" 또는 "음색"으로서 알려진 바와 같이, 그것의 스펙트럼 균형을 조정하거나 또는 변경하기 위해 음악 신호 상에 적용된다. 종래의 등화기는 사용자들로 하여금, 특정한 기구들을 강조하거나 또는 바람직하지 않은 사운드들을 제거하기 위해, 각각의 개개의 주파수 대역 상에서 주파수 응답 (이득)의 전체 프로파일(곡선 또는 형태)을 구성하도록 허용한다. 윈도우즈 미디어 플레이어와 같은, 인기 있는 음악 플레이어들은 보통 각각의 주파수 대역에서 이득을 조정하기 위해 그래픽 등화기를 제공하며, 또한, 음악의 상이한 장르들을 청취할 때 최고의 경험을 얻기 위해, 록, 랩, 재즈 및 포크송과 같은, 상이한 음악 장르들에 대한 등화기 프리셋들의 세트를 제공한다. 일단 프리셋이 선택되거나, 또는 프로파일이 설정된다면, 동일한 등화 이득들이, 프로파일이 수동으로 변경될 때까지, 신호 상에 적용될 것이다.

[0370] 반대로, 동적 등화기는 원하는 음색 또는 톤에 관하여 스펙트럼 균형의 전체 일관성을 유지하기 위해 각각의 주파수 대역들에서 등화 이득들을 자동으로 조정하기 위한 방식을 제공한다. 이러한 일관성은 오디오의 스펙트럼 균형을 연속적으로 모니터링하고, 그것을 원하는 사전 설정된 스펙트럼 균형에 비교하며, 오디오의 원래 스펙트

럼 균형을 원하는 스펙트럼 균형으로 변환하기 위해 적용된 등화 이득들을 동적으로 조정함으로써 달성된다. 원하는 스펙트럼 균형은 프로세싱 전에 수동으로 선택되거나 또는 사전 설정된다.

[0371] 양쪽 모두의 종류들의 등화기들은 다음의 단점을 공유한다: 최상의 등화 프로파일, 원하는 스펙트럼 균형, 또는 관련된 파라미터들이 수동으로 선택되어야 하며 그것들은 재생 중인 오디오 콘텐트에 기초하여 자동으로 변경될 수 없다. 오디오 콘텐트 유형들을 식별하는 것은 상이한 종류들의 오디오 신호들에 대한 전체 양호한 품질을 제공하기에 매우 중요할 것이다. 예를 들면, 상이한 음악 조각들은 상이한 장르들의 것들과 같은, 상이한 등화 프로파일들을 요구한다.

[0372] 임의의 종류들의 오디오 신호들(단지 음악이 아닌)이 입력되기에 가능한 등화기 시스템에서, 등화기 파라미터들은 콘텐트 유형들에 기초하여 조정되도록 요구한다. 예를 들면, 등화기는, 그것이 너무 많이 스피치의 음색을 변경할 수 있으며 그에 따라 신호 사운드를 부자연스럽게 만들기 때문에, 보통 음악 신호들에 대해 가능해지지만, 스피치 신호들에 대해서는 불능된다.

[0373] 적어도 부분적으로 이러한 문제점을 처리하기 위해, 본 출원은 파트 1에 논의된 실시예들에 기초하여 등화기를 제어하도록 제안한다.

[0374] 파트 2 내지 파트 4와 유사하게, 파트 1에 논의된 오디오 프로세싱 장치 및 방법의 특정 예로서, 등화기(408)는 파트 1에 논의된 실시예들 모두 및 여기에 개시된 이들 실시예들의 임의의 조합들을 이용할 수 있다. 구체적으로, 등화기(408)를 제어하는 경우에, 도 1 내지 도 10에 도시된 바와 같이 오디오 프로세싱 장치(100)에서의 오디오 분류기(200) 및 조정 유닛(300)은 도 22에 도시된 바와 같이 등화기(408) 제어기(2200)를 구성할 수 있다. 이 실시예에서, 조정 유닛은 등화기(408)에 특정적이므로, 그것은 300D로서 불리울 수 있다.

[0375] 즉, 파트 1의 개시에 기초하여, 등화기 제어기(2200)는 오디오 신호의 오디오 유형을 연속적으로 식별하기 위한 오디오 분류기(200); 및 식별된 대로 오디오 유형의 신뢰도 값에 기초하여 연속적인 방식으로 등화기를 조정하기 위한 조정 유닛(300D)을 포함할 수 있다. 유사하게, 오디오 분류기(200)는 오디오 콘텐트 분류기(202) 및 오디오 콘텍스트 분류기(204) 중 적어도 하나를 포함할 수 있으며, 볼륨 등화기 제어기(2200)는 유형 평활화 유닛(712), 파라미터 평활화 유닛(814) 및 타이머(916) 중 적어도 하나를 더 포함할 수 있다.

[0376] 그러므로, 이 파트에서, 우리는 파트 1에 이미 설명된 이들 콘텐트들을 반복하지 않을 것이며 단지 그것의 몇몇 특정한 예들을 제공할 것이다.

[0377] 5.1 콘텐트 유형에 기초한 제어

[0378] 일반적으로 말하면, 음악, 스피치, 배경 사운드 및 잡음과 같은 일반적인 오디오 콘텐트 유형들에 대해, 등화기는 상이한 콘텐트 유형들에 대해 상이하게 설정되어야 한다. 종래의 셋업과 유사하게, 등화기는 음악 신호들에 대해 자동으로 가능하게 될 수 있지만, 스피치에 대해 불능될 수 있거나; 또는 보다 연속적인 방식으로, 음악 신호들에 대해 높은 등화 레벨을 및 스피치 신호들에 대해 낮은 등화 레벨을 설정할 수 있다. 이러한 식으로, 등화기의 등화 레벨은 상이한 오디오 콘텐트에 대해 자동으로 설정될 수 있다.

[0379] 구체적으로 음악에 대해, 그것은 등화기가, 우세 소스의 음색이 상당히 변할 수 있으며 적절하지 않은 등화가 적용된다면 부자연스럽게 들릴 수 있으므로, 우세 소스를 가진 음악 조각상에서 그렇게 잘 작동하지 않는다는 것이 관찰된다. 이를 고려하면, 우세 소스들을 가진 음악 조각들에 대해 낮은 등화 레벨을 설정하는 것이 더 양호할 것인 반면, 등화 레벨은 우세 소스들 없이 음악 조각들에 대해 높게 유지될 수 있다. 이러한 정보를 갖고, 등화기는 상이한 음악 콘텐트에 대한 등화 레벨을 자동으로 설정할 수 있다.

[0380] 음악은 또한, 장르, 악기 및 리듬, 템포, 및 음색을 포함한 일반적인 음악 특성들과 같은, 상이한 속성들에 기초하여 그룹핑될 수 있다. 상이한 등화기 프리셋들이 상이한 음악 장르들을 위해 사용되는 동일한 방식으로, 이들 음악 그룹들/클러스터들이 또한 그것들 자신의 최적의 등화 프로파일들 또는 등화기 곡선들(종래의 등화기에서) 또는 최적의 원하는 스펙트럼 균형(동적 등화기에서)을 가질 수 있다.

[0381] 상기 언급된 바와 같이, 등화기가 다이얼로그를 음색 변화로 인해 그렇게 양호하게 들리지 않게 할 수 있으므로 등화기는 일반적으로 음악 콘텐트에 대해 가능해지지만 스피치에 대해 불능된다. 그것이 자동으로 달성하기 위한 하나의 방법은 콘텐트, 특히 오디오 콘텐트 분류 모듈로부터 획득된 음악 신뢰도 값 및/또는 스피치 신뢰도 값에, 등화 레벨을 관련시키는 것이다. 여기에서, 등화 레벨은 적용된 등화기 이득들의 가중치로서 설명될 수 있다. 레벨이 높을수록, 적용된 등화는 더 강하다. 예에 대해, 등화 레벨이 1이면, 완전한 등화 프로파일이 적용되며; 등화 레벨이 0이면, 모든 이득들이 그에 따라 0dB이고 따라서 비-등화가 적용된다. 등화 레벨이 등화기

알고리즘들의 상이한 구현들에서의 상이한 파라미터들에 의해 표현될 수 있다. 이러한 파라미터의 예시적 실시예는, 여기에 전체적으로 참조로서 통합되는, US 2009/0097676 A1으로서 공개된, A. Seefeldt 외의 "오디오 신호의 지각된 라우드니스 및/또는 지각된 스펙트럼 균형을 산출 및 조정하는 방법"에 구현된 바와 같이 등화기 가중치이다.

[0382] 다양한 제어 기법들이 등화 레벨을 튜닝하기 위해 설계될 수 있다. 예를 들면, 오디오 콘텐트 유형 정보를 갖고, 스피치 신뢰도 값 또는 음악 신뢰도 값이 다음으로서, 등화 레벨을 설정하기 위해 사용될 수 있다:

[0383] $$L_{eq} \propto Conf_{music} \quad (20)$$

[0384] 또는

[0385] $$L_{eq} \propto 1 - Conf_{speech} \quad (21)$$

[0386] 여기에서 L_{eq}는 등화 레벨이며 $Conf_{music}$ 및 $Conf_{speech}$는 음악 및 스피치의 신뢰도 값을 나타낸다.

[0387] 즉, 조정 유닛(300D)은 단기 음악의 신뢰도 값과 등화 레벨을 양으로 상관시키거나 또는 스피치의 신뢰도 값과 등화 레벨을 음으로 상관시키도록 구성될 수 있다.

[0388] 스피치 신뢰도 값 및 음악 신뢰도 값은 등화 레벨을 설정하기 위해 추가로 공동으로 사용될 수 있다. 일반적인 사상은 등화 레벨이 단지 양쪽 모두 음악 신뢰도 값이 높으며 스피치 신뢰도 값이 낮을 때만 높아야 하며 그렇지 않다면 등화 레벨이 낮다는 것이다. 예를 들면,

[0389] $$L_{eq} = Conf_{music}(1 - Conf_{speech}^{\alpha}) \quad (22)$$

[0390] 여기에서 스피치 신뢰도 값은 빈번하게 일어날 수 있는 음악 신호들에서 비-제로 스피치 신뢰도를 처리하기 위해 α에 동력이 공급된다. 상기 공식을 갖고, 등화는 임의의 스피치 구성요소들 없이 순수 음악 신호들에 완전히 적용될 것이다(1과 같은 레벨을 갖고). 파트 1에 서술된 바와 같이, α는 콘텐트 유형의 중요성에 기초하여 가중 계수로서 간주될 수 있으며, 통상적으로 1 내지 2로 설정될 수 있다.

[0391] 스피치의 신뢰도 값에 보다 큰 가중치를 제기한다면, 조정 유닛(300D)은 콘텐트 유형 스피치에 대한 신뢰도 값이 임계치보다 클 때 등화기(408)를 불능시키도록 구성될 수 있다.

[0392] 상기 설명에서, 음악 및 스피치의 콘텐트 유형들은 예들로서 취해진다. 대안적으로 또는 부가적으로, 배경 사운드 및/또는 잡음의 신뢰도 값들이 또한 고려될 수 있다. 구체적으로, 조정 유닛(300D)은 배경의 신뢰도 값과 등화 레벨을 양으로 상관시키고, 및/또는 잡음의 신뢰도 값과 등화 레벨을 음으로 상관시키도록 구성될 수 있다.

[0393] 또 다른 실시예로서, 신뢰도 값은 섹션 1.4에 논의된 바와 같이 정규화된 가중치를 도출하기 위해 사용될 수 있다. 예상된 등화 레벨이 각각의 콘텐트 유형에 대해 미리 정의된다고 가정하면(예로서, 음악에 대해 1, 스피치에 대해 0, 잡음 및 배경에 대해 0.5), 공식(8)과 유사한 공식이 정확하게 적용될 수 있다.

[0394] 등화 레벨은 전이 포인트들에서 가청 아티팩트들을 도입할 수 있는 빠른 변화를 회피하기 위해 추가로 평활화될 수 있다. 이것은 섹션 1.5에 설명된 바와 같이 파라미터 평활화 유닛(814)을 갖고 행해질 수 있다.

[0395] 5.2 음악에서 우세 소스들의 가능성

[0396] 우세 소스들을 가진 음악이 높은 등화 레벨을 적용받는 것을 회피하기 위해, 등화 레벨은 음악 조각이 우세 소스를 포함하는지를 표시하는 신뢰도 값($Conf_{dom}$)에 추가로 상관될 수 있다, 예를 들면,

[0397] $$L_{eq} = 1 - Conf_{dom} \quad (23)$$

[0398] 이러한 방식으로, 등화 레벨은 우세 소스들을 가진 음악 조각들에 대해 낮으며, 우세 소스들이 없는 음악 조각들에 대해 높을 것이다.

[0399] 여기에서, 우세 소스를 가진 음악의 신뢰도 값이 설명되지만, 우리는 우세 소스가 없는 음악의 신뢰도 값을 또한 사용할 수 있다. 즉, 조정 유닛(300D)은 우세 소스들이 없는 단기 음악의 신뢰도 값과 등화 레벨을 양으로 상관시키고, 및/또는 우세 소스들을 가진 단기 음악의 신뢰도 값과 등화 레벨을 음으로 상관시키도록 구성될 수

있다.

[0400] 섹션 1.1에 서술된 바와 같이, 한편으로 음악 및 스피치, 및 다른 한편으로 우세 소스들을 갖거나 또는 그것이 없는 음악이 상이한 계층 레벨들 상에서의 콘텐트 유형들일지라도, 그것들은 동시에 고려될 수 있다. 우세 소스들의 신뢰도 값 및 상기 설명된 바와 같이 스피치 및 음악 신뢰도 값들을 함께 고려함으로써, 등화 레벨은 공식(20 및 21) 중 적어도 하나를 (23)과 조합함으로써 설정될 수 있다. 예는 3개의 공식 모두를 조합한다:

[0401] $$L_{eq} = Conf_{music}(1-Conf_{speech})(1-Conf_{dom}) \quad (24)$$

[0402] 콘텐트 유형의 중요성에 기초한 상이한 가중치들이 공식(22)의 방식에서와 같이, 일반성에 대한 상이한 신뢰도 값들에 추가로 적용될 수 있다.

[0403] 또 다른 예로서, $Conf_{dom}$이 단지 오디오 신호가 음악일 때만 계산된다고 가정하면, 단계적 함수가 다음과 같이 설계될 수 있다:

[0404] $$L_{eq} = \begin{cases} (1-Conf_{dom}) & Conf_{music} > \text{임계치} \\ Conf_{music}(1-conf_{speech}^{\alpha}) & \text{그렇지 않다면} \end{cases} \quad (25)$$

[0405] 이러한 함수는 분류 시스템이 오디오가 음악임을 공정하게 알아낸다면 우세 스코어들의 신뢰도 값에 기초하여 등화 레벨을 설정하며(음악 신뢰도 값은 임계치보다 크다); 그렇지 않다면, 그것은 음악 및 스피치 신뢰도 값들에 기초하여 설정된다. 즉, 조정 유닛(300D)은 단기 음악에 대한 신뢰도 값이 임계치보다 클 때 우세 소스들 없이/이를 갖고 단기 음악을 고려하도록 구성될 수 있다. 물론, 공식(25)에서의 전반 또는 후반은 공식(20) 내지 공식(24)의 방식으로 변경될 수 있다.

[0406] 섹션 1.5에 논의된 바와 동일한 평활화 기법이 또한 적용될 수 있으며, 시간 상수(α)는 우세 소스들을 가진 음악으로부터 우세 소스들이 없는 음악으로의 전이, 또는 우세 소스들이 없는 음악으로부터 우세 소스들을 가진 음악으로의 전이와 같은, 전이 유형에 기초하여 추가로 설정될 수 있다. 이러한 목적을 위해, 공식(4')과 유사한 공식이 또한 적용될 수 있다.

[0407] 5.3 등화기 프리셋들

[0408] 오디오 콘텐트 유형들의 신뢰도 값들에 기초하여 등화 레벨을 적응적으로 튜닝하는 것 외에, 적절한 등화 프로파일들 또는 원하는 스펙트럼 균형 프리셋들이 또한 그것들의 장르, 악기, 또는 다른 특성들에 의존하여, 상이한 오디오 콘텐트에 대해 자동으로 선택될 수 있다. 동일한 장르를 갖고, 동일한 악기를 포함하거나 또는 동일한 음악 특성들을 가진 음악은 동일한 등화 프로파일들 또는 원하는 스펙트럼 균형 프리셋들을 공유할 수 있다.

[0409] 일반성을 위해, 우리는 동일한 장르, 동일한 악기, 또는 유사한 음악 속성들을 가진 음악 그룹들을 나타내기 위해 용어 "음악 클러스터"를 사용하며, 그것들은 섹션 1.1에 서술된 바와 같이 오디오 콘텐트 유형들의 또 다른 계층 레벨로서 간주될 수 있다. 적절한 등화 프로파일, 등화 레벨, 및/또는 원하는 스펙트럼 균형 프리셋이 각각의 음악 클러스터와 연관될 수 있다. 등화 프로파일은 음악 신호에 적용된 이득 곡선이며 상이한 음악 장르들(클래식, 록, 재즈 및 포크송과 같은)을 위해 사용된 등화기 프리셋들 중 임의의 것일 수 있으며, 원하는 스펙트럼 균형 프리셋은 각각의 클러스터에 대한 원하는 음색을 나타낸다. 도 23은 돌비 홈 시어터 기술들에서 구현된 바와 같이 원하는 스펙트럼 균형 프리셋들의 여러 개의 예들을 예시한다. 각각의 것은 가청 주파수 범위에 걸쳐 원하는 스펙트럼 형태를 설명한다. 이러한 형태는 인입 오디오의 스펙트럼 형태에 계속해서 비교되며, 등화 이득들은 인입 오디오의 스펙트럼 형태를 프리셋의 것으로 변환하기 위해 이러한 비교로부터 계산된다.

[0410] 새로운 음악 조각에 대해, 가장 가까운 클러스터가 결정될 수 있거나(경 판정), 또는 각각의 음악 클러스터에 대한 신뢰도 값이 계산될 수 있다(연 판정). 이러한 정보에 기초하여, 적절한 등화 프로파일, 또는 원하는 스펙트럼 균형 프리셋이 주어진 음악 조각에 대해 결정될 수 있다. 가장 간단한 방식은 다음과 같이, 최상의 매칭된 클러스터의 대응하는 프로파일에 그것을 할당하는 것이다.

[0411] $$P_{eq} = P_{c*} \quad (26)$$

[0412] 여기에서 P_{eq}는 추정된 등화 프로파일 또는 원하는 스펙트럼 균형 프리셋이며, c*는 최고 매칭된 음악 클러스터

(우세 오디오 유형)의 인덱스이며, 이것은 최고 신뢰도 값을 가진 클러스터를 픽업함으로써 획득될 수 있다.

[0413] 게다가, 0보다 큰 신뢰도 값을 가진 하나 이상의 음악 클러스터가 있을 수 있으며, 이것은 음악 조각이 이들 클러스터들과 유사한 더 많거나 또는 더 적은 속성들을 가진다는 것을 의미한다. 예를 들면, 음악 조각은 다수의 악기들을 가질 수 있거나, 또는 그것은 다수의 장르들의 속성들을 가질 수 있다. 그것은, 단지 가장 가까운 클러스터만을 사용하는 것 대신에, 클러스터 모두를 고려함으로써 적절한 등화 프로파일을 추정하기 위한 또 다른 방식을 고취시킨다. 예를 들면, 가중 합이 사용될 수 있다:

[0414]
$$P_{eq} = \sum_{c=1}^{N} w_c P_c \quad (27)$$

[0415] 여기에서 N은 미리 정의된 클러스터들의 수이며, w_c는 각각의 미리 정의된 음악 클러스터(인덱스(c)를 가진)에 관한 설계된 프로파일(P_c)의 가중이며, 이것은 그것들의 대응하는 신뢰도 값에 기초하여 1로 정규화되어야 한다. 이러한 방식으로, 추정된 프로파일은 음악 클러스터들의 프로파일들의 혼합일 것이다. 예를 들면, 재즈 및 록의 양쪽 속성들 모두를 가진 음악 조각에 대해, 추정된 프로파일은 그 사이에서의 무엇일 것이다.

[0416] 몇몇 애플리케이션들에서, 우리는 공식(27)에 도시된 바와 같이 클러스터들 모두를 수반하길 원하지 않을 수 있다. 단지 클러스터들의 서브세트만 - 현재 음악 조각에 가장 관련된 클러스터들 -이 고려될 필요가 있으며, 공식(27)은 다음으로 약간 개정될 수 있다:

[0417]
$$P_{eq} = \sum_{c'=1}^{N'} w_{c'} P_{c'} \quad (28)$$

[0418] 여기에서 N'은 고려될 클러스터들의 수이며, c'는 그것들의 신뢰도 값들에 기초하여 클러스터들을 점감적으로 분류한 후 클러스터 인덱스이다. 서브세트를 사용함으로써, 우리는 가장 관련된 클러스터들에 대해 더 초점을 맞출 수 있으며 덜 관련된 것들을 배제할 수 있다. 다시 말해서, 조정 유닛(300D)은 신뢰도 값들에 기초하여 적어도 하나의 우세 오디오 유형을 고려하도록 구성될 수 있다.

[0419] 상기 설명에서, 음악 클러스터들이 예로서 취해진다. 사실상, 해결책들은 섹션 1.1에 논의된 바와 같이 임의의 계층 레벨 상에서 오디오 유형들에 적용 가능하다. 그러므로, 일반적으로, 조정 유닛(300D)은 각각의 오디오 유형에 등화 레벨 및/또는 등화 프로파일 및/또는 스펙트럼 균형 프리셋을 할당하도록 구성될 수 있다.

[0420] 5.4 콘텍스트 유형에 기초한 제어

[0421] 이전 섹션들에서, 논의는 다양한 콘텐트 유형들에 초점이 맞추어진다. 이 섹션에서 논의될 보다 많은 실시예들에서, 콘텍스트 유형이 대안적으로 또는 부가적으로 고려될 수 있다.

[0422] 일반적으로, 등화기는 영화-형 미디어에서의 다이얼로그들이 명백한 음색 변화로 인해 그렇게 양호하게 들리지 않게 할 수 있기 때문에 등화기는 음악에 대해 가능해지지만 영화-형 미디어 콘텐트에 대해서는 불능된다. 그것은 등화 레벨이 장기 음악의 신뢰도 값 및/또는 영화-형 미디어의 신뢰도 값에 관련될 수 있음을 표시한다:

[0423]
$$L_{eq} \propto Conf_{MUSIC} \quad (29)$$

[0424] 또는

[0425]
$$L_{eq} \propto 1 - Conf_{MOVIE} \quad (30)$$

[0426] 여기에서 L_{eq}는 등화 레벨이고, $Conf_{MUSIC}$ 및 $Conf_{MOVIE}$는 장기 음악 및 영화-형 미디어의 신뢰도 값을 나타낸다.

[0427] 즉, 조정 유닛(300D)은 장기 음악의 신뢰도 값과 등화 레벨을 양으로 상관시키거나, 또는 영화-형 미디어의 신뢰도 값과 등화 레벨을 음으로 상관시키도록 구성될 수 있다.

[0428] 즉, 영화-형 미디어 신호에 대해, 영화-형 미디어 신뢰도 값은 높으며(또는 음악 신뢰도는 낮다), 따라서 등화

레벨은 낮고; 다른 한편으로, 음악 신호에 대해, 영화-형 미디어 신뢰도 값은 낮을 것이며(또는 음악 신뢰도는 높다) 따라서 등화 레벨은 높다.

[0429] 공식(29) 및 공식(30)에 도시된 해결책들은 공식(22) 내지 공식(25)과 동일한 방식으로 변경될 수 있으며, 및/또는 공식(22) 내지 공식(25)에 도시된 해결책들 중 임의의 하나와 조합될 수 있다.

[0430] 부가적으로 또는 대안적으로, 조정 유닛(300D)은 게임의 신뢰도 값과 등화 레벨을 음으로 상관시키도록 구성될 수 있다.

[0431] 또 다른 실시예에서, 신뢰도 값은 섹션 1.4에 논의된 바와 같이 정규화된 가중치를 도출하기 위해 사용될 수 있다. 예상된 등화 레벨/프로파일이 각각의 콘텍스트 유형에 대해 미리 정의된다고 가정하면(등화 프로파일들은 다음의 표 2에 도시된다), 공식(9)과 유사한 공식이 또한 적용될 수 있다.

[0432] 표 2:

[0433]

	영화-형 미디어	장기 음악	VoIP	게임
등화 프로파일	프로파일 1	프로파일 2	프로파일 3	프로파일 4

[0434] 여기에서, 몇몇 프로파일들에서, 영화-형 미디어 및 게임과 같은, 상기 특정한 콘텍스트 유형에 대한 등화기를 불능시키기 위한 방식으로서, 모든 이득들은 0으로 설정될 수 있다.

[0435] 5.5 실시예들 및 애플리케이션 시나리오들의 조합

[0436] 파트 1과 유사하게, 상기 논의된 모든 실시예들 및 그 변형들이 그것의 임의의 조합으로 구현될 수 있으며, 상이한 파트들/실시예들에 언급되지만 동일하거나 또는 유사한 기능들을 가진 임의의 구성요소들이 동일하거나 또는 개별적인 구성요소들로서 구현될 수 있다.

[0437] 예를 들면, 섹션 5.1 내지 섹션 5.4에 설명된 해결책들 중 임의의 둘 이상이 서로 조합될 수 있다. 조합들 중 임의의 것은 파트 1 내지 파트 4 및 나중에 설명될 다른 파트들에 설명되거나 또는 내포된 임의의 실시예와 추가로 조합될 수 있다.

[0438] 5.6 등화기 제어 방법

[0439] 파트 1과 유사하게, 이전 실시예들에서 등화기 제어기를 설명하는 프로세스에서, 몇몇 프로세스들 또는 방법들이 또한 명확하게 개시된다. 이후 이들 방법들의 요약은 이전에 이미 논의된 상세들의 일부를 반복하지 않고 제공된다.

[0440] 첫 번째로, 파트 1에 논의된 바와 같이 오디오 프로세싱 방법의 실시예들은 등화기를 위해 사용될 수 있으며, 그 파라미터(들)는 오디오 프로세싱 방법에 의해 조정될 타겟들 중 하나이다. 이러한 관점으로부터, 오디오 프로세싱 방법은 또한 등화기 제어 방법이다.

[0441] 이 섹션에서, 단지 등화기의 제어에 특정한 이들 측면들만이 논의될 것이다. 제어 방법의 일반적인 측면들에 대해, 참조가 파트 1에 대해 이루어질 수 있다.

[0442] 실시예들에 따르면, 등화기 제어 방법은 오디오 신호의 오디오 유형을 실시간으로 식별하는 단계, 및 식별된 대로 오디오 유형의 신뢰도 값에 기초하여 연속적인 방식으로 등화기를 조정하는 단계를 포함할 수 있다.

[0443] 본 출원의 다른 파트들과 유사하게, 대응하는 신뢰도 값들을 가진 다수의 오디오 유형들을 수반할 때, 조정의 동작(1104)은 다수의 오디오 유형들의 중요성에 기초하여 다수의 오디오 유형들의 신뢰도 값을 가중시키는 것을 통해, 또는 신뢰도 값들에 기초하여 다수의 오디오 유형들의 효과들을 가중시키는 것을 통해 다수의 오디오 유형들 중 적어도 일부를 고려하도록 구성될 수 있다. 구체적으로, 조정 동작(1104)은 신뢰도 값들에 기초하여 적어도 하나의 우세 오디오 유형을 고려하도록 구성될 수 있다.

[0444] 파트 1에 설명된 바와 같이, 조정된 파라미터 값이 평활화될 수 있다. 참조가 섹션 1.5 및 섹션 1.8에 대해 이루어질 수 있으며 상세한 설명은 여기에서 생략된다.

[0445] 오디오 유형은 콘텐트 유형 또는 콘텍스트 유형, 또는 양쪽 모두일 수 있다. 콘텐트 유형을 수반할 때, 조정 동작(1104)은 단기 음악의 신뢰도 값과 등화 레벨을 양으로 상관시키고, 및/또는 스피치의 신뢰도 값과 등화 레벨을 음으로 상관시키도록 구성될 수 있다. 부가적으로 또는 대안적으로, 조정 동작은 배경의 신뢰도 값과 등화 레벨을 양으로 상관시키고, 및/또는 잡음의 신뢰도 값과 등화 레벨을 음으로 상관시키도록 구성될 수 있다.

[0446] 콘텍스트 유형을 수반할 때, 조정 동작(1104)은 장기 음악의 신뢰도 값과 등화 레벨을 양으로 상관시키고, 및/또는 영화-형 미디어 및/또는 게임의 신뢰도 값과 등화 레벨을 음으로 상관시키도록 구성될 수 있다.

[0447] 단기 음악의 콘텐트 유형에 대해, 조정 동작(1104)은 우세 소스들이 없는 단기 음악의 신뢰도 값과 등화 레벨을 양으로 상관시키고, 및/또는 우세 소스들을 가진 단기 음악의 신뢰도 값과 등화 레벨을 음으로 상관시키도록 구성될 수 있다. 이것은 단기 음악에 대한 신뢰도 값이 임계치보다 클 때만 행해질 수 있다.

[0448] 등화 레벨을 조정하는 것 외에, 등화기의 다른 측면들이 오디오 신호의 오디오 유형(들)의 신뢰도 값(들)에 기초하여 조정될 수 있다. 예를 들면, 조정 동작(1104)은 각각의 오디오 유형에 등화 레벨 및/또는 등화 프로파일 및/또는 스펙트럼 균형 프리셋을 할당하도록 구성될 수 있다.

[0449] 오디오 유형들의 특정 인스턴스들에 대해, 참조가 파트 1에 대해 이루어질 수 있다.

[0450] 오디오 프로세싱 장치의 실시예들과 유사하게, 오디오 프로세싱 방법의 실시예들 및 그것들의 변형들의 임의의 조합이 한편으로는 실현 가능하며; 다른 한편으로, 오디오 프로세싱 방법의 실시예들 및 그것들의 변형들의 모든 측면은 별개의 해결책들일 수 있다. 또한, 이 섹션에 설명된 임의의 둘 이상의 해결책들은 서로 조합될 수 있으며, 이들 조합들은 본 개시의 다른 파트들에 설명되거나 또는 내포된 임의의 실시예와 추가로 조합될 수 있다.

[0451] 파트 6: 오디오 분류기들 및 분류 방법들

[0452] 섹션 1.1 및 섹션 1.2에 서술된 바와 같이, 콘텐트 유형들 및 콘텍스트 유형들의 다양한 계층 레벨들을 포함하여, 본 출원에 논의된 오디오 유형들은, 기계-학습 기반 방법들을 포함하여, 임의의 기존의 분류 기법을 갖고 분류되거나 또는 식별될 수 있다. 이러한 파트 및 후속 파트에서, 본 출원은 이전 파트들에서 언급된 바와 같이 콘텍스트 유형들을 분류하기 위한 방법들 및 분류기들의 몇몇 신규 측면들을 제안한다.

[0453] 6.1 콘텐트 유형 분류에 기초한 콘텍스트 분류기

[0454] 이전 파트들에서 서술된 바와 같이, 오디오 분류기(200)는 오디오 신호의 콘텐트 유형을 식별하며 및/또는 오디오 신호의 콘텍스트 유형을 식별하기 위해 사용된다. 그러므로, 오디오 분류기(200)는 오디오 콘텐트 분류기(202) 및/또는 오디오 콘텍스트 분류기(204)를 포함할 수 있다. 오디오 콘텐트 분류기(202) 및 오디오 콘텍스트 분류기(204)를 구현하기 위해 기존의 기술들을 채택할 때, 두 개의 분류기들은, 그것들이 몇몇 특징들을 공유할 수 있으며 따라서 특징들을 추출하기 위한 몇몇 기법들을 공유할 수 있을지라도, 서로로부터 독립적일 수 있다.

[0455] 이러한 파트 및 후속 파트 7에서, 본 출원에 제안된 신규 측면에 따르면, 오디오 콘텍스트 분류기(204)는 오디오 콘텐트 분류기(202)의 결과들을 이용할 수 있으며, 즉 오디오 분류기(200)는: 오디오 신호의 콘텐트 유형을 식별하기 위한 오디오 콘텐트 분류기(202); 및 오디오 콘텐트 분류기(202)의 결과들에 기초하여 오디오 신호의 콘텍스트 유형을 식별하기 위한 오디오 콘텍스트 분류기(204)를 포함할 수 있다. 따라서, 오디오 콘텐트 분류기(202)의 분류 결과들은 이전 파트들에서 논의된 바와 같이 오디오 콘텍스트 분류기(204) 및 조정 유닛(300)(또는 조정 유닛들(300A 내지 300D)) 양쪽 모두에 의해 사용될 수 있다. 그러나, 도면들에 도시되지 않을지라도, 오디오 분류기(200)는 또한 조정 유닛(300) 및 오디오 콘텍스트 분류기(204)에 의해 각각 사용될 두 개의 오디오 콘텐트 분류기들(202)을 포함할 수 있다.

[0456] 뿐만 아니라, 섹션 1.2에 논의된 바와 같이, 특히 다수의 오디오 유형들을 분류할 때, 오디오 콘텐트 분류기(202) 또는 오디오 콘텍스트 분류기(204)는, 그것이 또한 하나의 단일 분류기로서 구현되는 것이 가능할지라도, 서로와 협력하는 분류기들의 그룹으로 포함될 수 있다.

[0457] 섹션 1.1에 논의된 바와 같이, 콘텐트 유형은 일반적으로 대략 수 내지 수십 개의 프레임들(1s와 같은)의 길이를 가진 단기 오디오 세그먼트들에 대한 오디오 유형의 종류이며, 콘텍스트 유형은 일반적으로 대략 수 내지 수십의 초들(10s와 같은)의 길이를 가진 장기 오디오 세그먼트들에 대한 오디오 유형의 종류이다. 그러므로, "콘텐트 유형" 및 "콘텍스트 유형"에 대응하여, 우리는 필요할 때 각각 "단기" 및 "장기"를 사용한다. 그러나, 후속 파트 7에 논의될 바와 같이, 콘텍스트 유형이 비교적 긴 시간스케일에서 오디오 신호의 속성을 표시하기 위한 것일지라도, 그것은 또한 단기 오디오 세그먼트들로부터 추출된 특징들에 기초하여 식별될 수 있다.

[0458] 이제 도 24를 참조하여 오디오 콘텐트 분류기(202) 및 오디오 콘텍스트 분류기(204)의 구조들로 가자.

[0459] 도 24에 도시된 바와 같이, 오디오 콘텐트 분류기(202)는 각각이 오디오 프레임들의 시퀀스를 포함한 단기 오디오 세그먼트들로부터 단기 특징들을 추출하기 위한 단기 특징 추출기(2022); 및 각각의 단기 특징들을 사용하여

장기 오디오 세그먼트에서의 단기 세그먼트들의 시퀀스를 단기 오디오 유형들로 분류하기 위한 단기 분류기(2024)를 포함할 수 있다. 단기 특징 추출기(2022) 및 단기 분류기(2024) 양쪽 모두는 기존의 기술들을 갖고 구현될 수 있지만, 또한 몇몇 변경들은 후속 섹션 6.3에서 단기 특징 추출기(2022)를 위해 제안된다.

[0460] 단기 분류기(2024)는 단기 세그먼트들의 시퀀스의 각각을 다음의 단기 오디오 유형들(콘텐트 유형들): 스피치, 단기 음악, 배경 사운드 및 잡음 중 적어도 하나로 분류하도록 구성될 수 있으며, 이것은 섹션 1.1에 설명되었다. 콘텐트 유형의 각각은 이에 제한되지 않지만 섹션 1.1에 논의된 바와 같이, 보다 낮은 계층 레벨 상에서의 콘텐트 유형들로 추가로 분류될 수 있다.

[0461] 이 기술분야에 알려진 바와 같이, 분류된 오디오 유형들의 신뢰도 값들은 또한 단기 분류기(2024)에 의해 획득될 수 있다. 본 출원에서, 임의의 분류기의 동작을 언급할 때, 신뢰도 값들은 그것이 명확하게 기록되는지 여부에 관계없이, 필요하다면 동시에 획득된다는 것이 이해될 것이다. 오디오 유형 분류의 예가 2003년 3월, ACM 멀티미디어 시스템들 저널 8 (6), 페이지 482-492, L. Lu, H.-J. Zhang, 및 S. Li의 "지지 벡터 기계들을 사용하는 것에 의한 콘텐트-기반 오디오 분류 및 분할"에서 발견될 수 있으며, 이것은 여기에 전체적으로 참조로서 통합된다.

[0462] 다른 한편으로, 도 24에 도시된 바와 같이, 오디오 콘텍스트 분류기(204)는 장기 특징들로서, 장기 오디오 세그먼트에서의 단기 세그먼트들의 시퀀스에 대하여 단기 분류기의 결과들의 통계들을 산출하기 위한 통계 추출기(2042); 및 장기 특징들을 사용하여, 장기 오디오 세그먼트를 장기 오디오 유형들로 분류하기 위한 장기 분류기(2044)를 포함할 수 있다. 유사하게, 통계 추출기(2042) 및 장기 분류기(2044) 양쪽 모두는 기존의 기술들을 갖고 구현될 수 있지만, 또한 몇몇 변경들이 후속 섹션 6.2에서 통계 추출기(2042)를 위해 제안된다.

[0463] 장기 분류기(2044)는 다음의 장기 오디오 유형들(콘텍스트 유형들): 영화-형 미디어, 장기 음악, 게임 및 VoIP 중 적어도 하나로 장기 오디오 세그먼트를 분류하도록 구성될 수 있으며, 이것은 섹션 1.1에 설명되었다. 대안적으로 또는 부가적으로, 장기 분류기(2044)는 장기 오디오 세그먼트를 VoIP 또는 비-VoIP로 분류하도록 구성될 수 있으며, 이것은 섹션 1.1에서 설명되었다. 대안적으로 또는 부가적으로, 장기 분류기(2044)는 장기 오디오 세그먼트를 고-품질 오디오 또는 저-품질 오디오로 분류하도록 구성될 수 있으며, 이것은 섹션 1.1에서 설명되었다. 실제로, 다양한 타겟 오디오 유형들은 애플리케이션/시스템의 요구들에 기초하여 선택되며 트레이닝될 수 있다.

[0464] 단기 세그먼트 및 장기 세그먼트(뿐만 아니라 섹션 6.3에 논의될 프레임)의 의미 및 선택에 대하여, 참조가 섹션 1.1에 대해 이루어질 수 있다.

[0465] 6.2 장기 특징들의 추출

[0466] 도 24에 도시된 바와 같이, 일 실시예에서, 단지 통계 추출기(2042)만이 단기 분류기(2024)의 결과들로부터 장기 특징들을 추출하기 위해 사용된다. 장기 특징들로서, 다음 중 적어도 하나가 통계 추출기(2042)에 의해 산출될 수 있다: 분류될 장기 세그먼트에서의 단기 세그먼트들의 단기 오디오 유형들의 신뢰도 값들의 평균 및 분산, 단기 세그먼트들의 중요도들에 의해 가중된 평균 및 분산, 각각의 단기 오디오 유형의 발생 주파수 및 분류될 장기 세그먼트에서 상이한 단기 오디오 유형들 사이에서의 전이들의 주파수.

[0467] 우리는 각각의 단기 세그먼트(1s의 길이의)에서 스피치 및 단기 음악 신뢰도 값들의 평균을 도 25에 예시한다. 비교를 위해, 세그먼트들은 3개의 상이한 오디오 콘텍스트들로부터 추출된다: 영화-형 미디어(도 25(A)), 장기 음악(도 25(B)), 및 VoIP(도 25(C)). 영화-형 미디어 콘텍스트에 대해, 높은 신뢰도 값들이 스피치 유형을 위해 또는 음악 유형을 위해 획득된다는 것이 관찰될 수 있으며 그것은 이들 두 개의 오디오 유형들 사이에서 빈번하게 교번한다. 반대로, 장기 음악의 세그먼트는 안정되고 높은 단기 음악 신뢰도 값 및 비교적 안정되고 낮은 스피치 신뢰도 값을 제공한다. 반면에, VoIP의 세그먼트는 안정되며 낮은 단기 음악 신뢰도 값을 제공하지만, VoIP 대화 동안 중지들 때문에 변동하는 스피치 신뢰도 값들을 제공한다.

[0468] 각각의 오디오 유형에 대한 신뢰도 값들의 분산은 또한 상이한 오디오 콘텍스트들을 분류하기 위한 중요한 특징이다. 도 26은 스피치, 단기 음악, 영화-형 미디어에서의 배경 및 잡음, 장기 음악 및 VoIP 오디오 콘텍스트들의 신뢰도 값들의 분산의 히스토그램들을 제공한다(가로 좌표는 데이터세트에서 신뢰도 값들의 분산이며, 세로 좌표는 데이터세트에서 분산 값들(s)의 각각의 빈의 발생들의 수이고, 이것은 분산 값들의 각각의 빈의 발생 확률을 표시하기 위해 정규화될 수 있다). 영화-형 미디어에 대해, 스피치, 단기 음악 및 배경의 신뢰도 값의 분산들 모두는 비교적 높으며 광범위하게 분포되고, 이들 오디오 유형들의 신뢰도 값들이 집중적으로 변화하고 있음을 표시하고; 장기 음악에 대해, 스피치, 단기 음악, 배경 및 잡음의 신뢰도 값의 분산들 모두는 비교적 낮으

며 좁게 분포되고, 이들 오디오 유형들의 신뢰도 값들이 안정되게 유지한다는 것을 표시하며: 스피치 신뢰도 값은 끊임없이 낮게 유지하며 음악 신뢰도 값은 끊임없이 높게 유지하고; VoIP에 대해, 단기 음악의 신뢰도 값의 분산들은 낮으며 좁게 분포되는 반면, 스피치의 것은 비교적 넓게 분포되고, 이것은 VoIP 대화들 동안 빈번한 중지들 때문이다.

[0469] 가중된 평균 및 분산을 산출할 때 사용된 가중치들에 대해, 그것들은 각각의 단기 세그먼트의 중요도에 기초하여 결정된다. 단기 세그먼트의 중요도는 그것의 에너지 또는 라우드니스에 의해 측정될 수 있으며, 이것은 많은 기존의 기술들을 갖고 추정될 수 있다.

[0470] 분류될 장기 세그먼트에서의 각각의 단기 오디오 유형의 발생 빈도는 장기 세그먼트에서의 단기 세그먼트들이 분류된 각각의 오디오 유형의 카운트이며, 장기 세그먼트의 길이를 갖고 정규화된다.

[0471] 분류될 장기 세그먼트에서의 상이한 단기 오디오 유형들 사이에서의 전이들의 빈도는 분류될 장기 세그먼트에서의 인접한 단기 세그먼트들 사이에서의 오디오 유형 변화들의 카운트이며, 장기 세그먼트의 길이를 갖고 정규화된다.

[0472] 도 25를 참조하여 신뢰도 값들의 평균 및 분산을 논의할 때, 각각의 단기 오디오 유형의 발생 빈도 및 이들 상이한 단기 오디오 유형들 중에서 전이 빈도가 또한 사실상 건드려진다. 이들 특징들은 또한 오디오 콘텍스트 분류에 매우 관련된다. 예를 들면, 장기 음악은 대부분 그것이 단기 음악의 높은 발생 빈도를 갖도록 단기 음악 오디오 유형을 포함하는 반면, VoIP는 대부분 스피치를 포함하며 그것이 스피치 또는 잡음의 높은 발생 빈도를 갖도록 중지한다. 또 다른 예에 대해, 영화-형 미디어는 장기 음악 또는 VoIP가 그런것 보다 더 빈번하게 상이한 단기 오디오 유형들 중에서 전이하며, 따라서 그것은 일반적으로 단기 음악, 스피치 및 배경 중에서 더 높은 전이 빈도를 가지며; VoIP는 보통 다른 것들이 하는 것보다 더 빈번하게 스피치 및 잡음 사이에서 전이하고, 따라서 그것은 일반적으로 스피치 및 잡음 사이에서 더 높은 전이 빈도를 가진다.

[0473] 일반적으로, 우리는 장기 세그먼트들이 동일한 애플리케이션/시스템에서 길이가 동일하다고 가정한다. 이것이 그 경우이면, 각각의 단기 오디오 유형의 발생 카운트, 및 장기 세그먼트에서의 상이한 단기 오디오 유형들 사이에서의 전이 카운트가 정규화 없이 직접 사용될 수 있다. 장기 세그먼트의 길이가 가변적이면, 상기 언급된 바와 같이 발생 빈도 및 전이들의 빈도가 사용될 것이다. 본 출원에서의 청구항들은 양쪽 상황들 모두를 커버하는 것으로서 해석될 것이다.

[0474] 부가적으로 또는 대안적으로, 오디오 분류기(200)(또는 오디오 콘텍스트 분류기(204))는 장기 오디오 세그먼트에서 단기 세그먼트들의 시퀀스의 단기 특징들에 기초하여 장기 오디오 세그먼트로부터 장기 특징들을 추가로 추출하기 위한 장기 특징 추출기(2046)(도 27)를 더 포함할 수 있다. 다시 말해서, 장기 특징 추출기(2046)는 단기 분류기(2024)의 분류 결과들을 사용하지 않으며, 장기 분류기(2044)에 의해 사용될 몇몇 장기 특징들을 도출하기 위해 단기 특징 추출기(2022)에 의해 추출된 단기 특징들을 직접 사용한다. 장기 특징 추출기(2046) 및 통계 추출기(2042)는 독립적으로 또는 공동으로 사용될 수 있다. 다시 말해서, 오디오 분류기(200)는 장기 특징 추출기(2046) 또는 통계 추출기(2042), 또는 양쪽 모두를 포함할 수 있다.

[0475] 임의의 특징들이 장기 특징 추출기(2046)에 의해 추출될 수 있다. 본 출원에서, 장기 특징들로서, 단기 특징 추출기(2022)로부터 단기 특징들의 다음의 통계들 중 적어도 하나를 산출하는 것이 제안된다: 평균, 분산, 가중 평균, 가중 분산, 높은 평균, 낮은 평균, 및 높은 평균 및 낮은 평균 사이에서의 비(대비).

[0476] 단기 특징들의 평균 및 분산은 분류될 장기 세그먼트에서의 단기 세그먼트들로부터 추출된다.

[0477] 단기 특징들의 가중 평균 및 분산은 분류될 장기 세그먼트에서 단기 세그먼트들로부터 추출된다. 단기 특징들은 방금 전에 언급된 바와 같이 그것의 에너지 또는 라우드니스를 갖고 측정되는 각각의 단기 세그먼트의 중요도에 기초하여 가중된다;

[0478] 높은 평균: 선택된 단기 특징들의 평균은 분류될 장기 세그먼트에서의 단기 세그먼트들로부터 추출된다. 단기 특징들은 다음의 조건들 중 적어도 하나를 만족할 때 선택된다: 임계치 이상; 또는 다른 단기 특징들 모두보다 낮지 않은 단기 특징들의 미리 결정된 비율 내에서, 예를 들면, 최고 10%의 단기 특징들;

[0479] 낮은 평균: 선택된 단기 특징들의 평균은 분류될 장기 세그먼트에서의 단기 세그먼트들로부터 추출된다. 단기 특징들은 다음의 조건들 중 적어도 하나를 만족시킬 때 선택된다: 임계치 미만; 또는 모든 다른 단기 특징들보다 높지 않은 단기 특징들의 미리 결정된 비율 내에서, 예를 들면, 최저 10%의 단기 특징들; 및

[0480] 대비: 장기 세그먼트에서 단기 특징들의 역학을 표현하기 위한 높은 평균 및 낮은 평균 사이에서의 비.

[0481] 단기 특징 추출기(2022)는 기존의 기술들을 갖고 구현될 수 있으며, 임의의 특징들이 그에 의해 추출될 수 있다. 그럼에도 불구하고, 몇몇 변경들은 후속 섹션 6.3에서 단기 특징 추출기(2022)를 위해 제안된다.

[0482] 6.3 단기 특징들의 추출

[0483] 도 24 및 도 27에 도시된 바와 같이, 단기 특징 추출기(2022)는 단기 특징들로서, 각각의 단기 오디오 세그먼트들로부터 직접 다음의 특징들 중 적어도 하나를 추출하도록 구성될 수 있다: 리듬 특성들, 인터럽션들/음소거들 특성들 및 단기오디오 품질 특징들.

[0484] 리듬 특성들은 리듬 강도, 리듬 규칙성, 리듬 명료성(여기에 전체적으로 참조로서 통합되는, 2006년 오디오, 스피치 및 언어 프로세싱에 대한 IEEE 트랜잭션들, 14(1):5 - 18, L. Lu, D. Liu, 및 H.-J. Zhang의 "음악 오디오 신호들의 자동 무드 검출 및 추적" 참조) 및 2D 서브-대역 변조(여기에 전체적으로 참조로서 통합되는, 2003년 Proc. ISMIR, M.F McKinney 및 J. Breebaart의 "오디오 및 음악 분류를 위한 특징들")를 포함할 수 있다.

[0485] 인터럽션들/음소거들 특성들은 스피치 인터럽션들, 급낙들, 음소거 길이, 부자연스러운 침묵, 부자연스러운 침묵의 평균, 부자연스러운 침묵의 총 에너지 등을 포함할 수 있다.

[0486] 단기 오디오 품질 특징들은 단기 세그먼트들에 대한 오디오 품질 특징들이며, 이것은 이하에 논의되는, 오디오 프레임들로부터 추출된 오디오 품질 특징들과 유사하다.

[0487] 대안적으로 또는 부가적으로, 도 28에 도시된 바와 같이, 오디오 분류기(200)는 단기 세그먼트에 포함된 오디오 프레임들의 시퀀스의 각각으로부터 프레임-레벨 특징들을 추출하기 위한 프레임-레벨 특징 추출기(2012)를 포함할 수 있으며, 단기 특징 추출기(2022)는 오디오 프레임들의 시퀀스로부터 추출된 프레임-레벨 특징들에 기초하여 단기 특징들을 산출하도록 구성될 수 있다.

[0488] 전-처리로서, 입력 오디오 신호는 모노 오디오 신호로 다운-믹싱될 수 있다. 전-처리는 오디오 신호가 이미 모노 신호이면 불필요하다. 그것은 그 후 미리 정의된 길이(통상적으로 10 내지 25 밀리초들)를 가진 프레임들로 분할된다. 그것에 부응하여, 프레임-레벨 특징들이 각각의 프레임으로부터 추출된다.

[0489] 프레임-레벨 특징 추출기(2012)는 다음의 특징들 중 적어도 하나를 추출하도록 구성될 수 있다: 다양한 단기 오디오 유형들의 특징들 특성화 속성들, 컷오프 주파수, 정적 신호-잡음-비(SNR) 특성들, 분절 신호-잡음-비(SNR) 특성들, 기본 스피치 디스크립터들, 및 성도 특성들.

[0490] 다양한 단기 오디오 유형들(특히 스피치, 단기 음악, 배경 사운드 및 잡음)의 특징들 특성화 속성들은 다음의 특징들 중 적어도 하나를 포함할 수 있다: 프레임 에너지, 서브-대역 스펙트럼 분포, 스펙트럼 플럭스, Mel-주파수 캡스트럼 계수(Mel-frequency Cepstral Coefficient; MFCC), 베이스, 잔여 정보, 채도 특징 및 제로-교차 레이트.

[0491] MFCC의 상세에 대해, 참조가 2003년 3월, ACM 멀티미디어 시스템들 저널 8(6), 페이지 482 내지 492, L. Lu, H.-J. Zhang, 및 S. Li의 "지지 벡터 기계들을 사용하는 것에 의한 콘텐트-기반 오디오 분류 및 분할"에 대해 이루어질 수 있으며, 이것은 여기에 전체적으로 참조로서 통합된다. 채도 특징의 상세에 대해, 참조가 1999년, SPIE에서 G. H. Wakefield의 "공동 시간 채도 분포들의 수학적 표현"에 대해 이루어질 수 있으며, 이것은 여기에서 전체적으로 참조로서 통합된다.

[0492] 컷오프 주파수는 콘텐트의 에너지가 0에 가까운 오디오 신호의 최고 주파수를 나타낸다. 그것은 대역 제한 콘텐트를 검출하도록 설계되며, 이것은 본 출원에서 오디오 콘텍스트 분류를 위해 유용하다. 컷오프 주파수는, 대부분의 코더들이 낮은 또는 중간 비트레이트들에서 높은 주파수들을 폐기하기 때문에, 보통 코딩에 야기된다. 예를 들면, MP3 코덱은 128 kbps에서 16kHz의 컷오프 주파수를 가지며; 또 다른 예에 대해, 많은 인기 있는 VoIP 코덱들은 8kHz 또는 16kHz의 컷오프 주파수를 가진다.

[0493] 컷오프 주파수 외에, 오디오 인코딩 프로세스 동안 신호 열화는 VoIP 대 비-VoIP 콘텍스트들, 고-품질 대 저-품질 오디오 콘텍스트들과 같은 다양한 오디오 콘텍스트들을 구별하기 위한 또 다른 특성으로서 고려된다. 객관적 스피치 품질 평가(여기에 전체적으로 참조로서 통합되는, 2006년 11월, 오디오, 스피치, 및 언어 프로세싱에 대한 IEEE 트랜잭션, VOL. 14, NO. 6, Ludovic Malfait, Jens Berger, 및 Martin Kastner의 "단일-출력 스피치 품질 평가를 위한 P.563- ITU-T 표준" 참조)에 대한 것들과 같은, 오디오 품질을 나타내는 특징들이 보다 풍부한 특성들을 캡처하기 위해 다수의 레벨들에서 추가로 추출될 수 있다. 오디오 품질 특징들의 예들은:

[0494] a) 추정된 배경잡음 레벨, 스펙트럼 명료성 등을 포함한 정적 SNR 특성들.

[0495] b) 스펙트럼 레벨 편차, 스펙트럼 레벨 범위, 상대적 잡음 플로어 등을 포함한 분절 SNR 특성들.

[0496] c) 피치 평균, 스피치 섹션 레벨 변화, 스피치 레벨 등을 포함한 기본 스피치 디스크립터들.

[0497] d) 로봇화, 피치 교차 전력 등을 포함한 성도 특성들.

[0498] 프레임-레벨 특징들로부터 단기 특징들을 도출하기 위해, 단기 특징 추출기(2022)가 단기 특징들로서 프레임-레벨 특징들의 통계들을 산출하도록 구성될 수 있다.

[0499] 프레임-레벨 특징들의 통계들의 예들은 단기 음악, 스피치, 배경 및 잡음과 같은, 다양한 오디오 유형들을 구별하기 위해 리듬 속성들을 캡처하는, 평균 및 표준 편차를 포함한다. 예를 들면, 스피치는 일반적으로 음절 레이트에서 음성 및 비음성 사운드들 사이에서 교번하는 반면 음악은 아니며, 스피치의 프레임-레벨 특징들의 변화는 보통 음악의 것보다 크다는 것을 표시한다.

[0500] 통계들의 또 다른 예는 프레임-레벨 특징들의 가중 평균이다. 예를 들면, 컷오프 주파수에 대해, 가중치로서 각각의 프레임의 에너지 또는 라우드니스를 갖고, 단기 세그먼트에서의 모든 오디오 프레임들로부터 도출된 컷오프 주파수들 중에서 가중 평균은 상기 단기 세그먼트에 대한 컷오프 주파수일 것이다.

[0501] 대안적으로 또는 부가적으로, 도 29에 도시된 바와 같이, 오디오 분류기(200)는 오디오 프레임들로부터 프레임-레벨 특징들을 추출하기 위한 프레임-레벨 특징 추출기(2012) 및 각각의 프레임-레벨 특징들을 사용하여 오디오 프레임들의 시퀀스의 각각을 프레임-레벨 오디오 유형들로 분류하기 위한 프레임-레벨 분류기(2014)를 포함할 수 있으며, 여기에서 단기 특징 추출기(2022)는 오디오 프레임들의 시퀀스에 대하여 프레임-레벨 분류기(2014)의 결과들에 기초하여 단기 특징들을 산출하도록 구성될 수 있다.

[0502] 다시 말해서, 오디오 콘텐트 분류기(202) 및 오디오 콘텍스트 분류기(204) 외에, 오디오 분류기(200)는 프레임 분류기(201)를 더 포함할 수 있다. 이러한 아키텍처에서, 오디오 콘텐트 분류기(202)는 프레임 분류기(201)의 프레임-레벨 분류 결과들에 기초하여 단기 세그먼트를 분류하며, 오디오 콘텍스트 분류기(204)는 오디오 콘텐트 분류기(202)의 단기 분류 결과들에 기초하여 장기 세그먼트를 분류한다.

[0503] 프레임-레벨 분류기(2014)는 오디오 프레임들의 시퀀스의 각각을 임의의 클래스들로 분류하도록 구성될 수 있으며, 이것은 "프레임-레벨 오디오 유형들"로서 불리울 수 있다. 일 실시예에서, 프레임-레벨 오디오 유형은 이전에 논의된 콘텐트 유형들의 아키텍처와 유사한 아키텍처를 가질 수 있고 또한 콘텐트 유형들과 유사한 의미를 가지며, 유일한 차이는 프레임-레벨 오디오 유형들이며 콘텐트 유형들은 프레임 레벨 및 단기 세그먼트 레벨인, 오디오 신호의 상이한 레벨들에서 분류된다. 예를 들면, 프레임-레벨 분류기(2014)는 다음의 프레임-레벨 오디오 유형들 중 적어도 하나로 오디오 프레임들의 시퀀스의 각각을 분류하도록 구성될 수 있다: 스피치, 음악, 배경 사운드 및 잡음. 다른 한편으로, 프레임-레벨 오디오 유형들은 또한 콘텐트 유형들의 아키텍처와 부분적으로 또는 완전히 상이하고, 프레임-레벨 분류에 더 적합하며, 단기 분류를 위한 단기 특징들로서 사용되기에 더 적절한 아키텍처를 가질 수 있다. 예를 들면, 프레임-레벨 분류기(2014)는 오디오 프레임들의 시퀀스의 각각을 다음의 프레임-레벨 오디오 유형들 중 적어도 하나로 분류하도록 구성될 수 있다: 음성, 비음성, 및 중지.

[0504] 프레임-레벨 분류의 결과들로부터 단기 특징들을 도출하는 방법에 대해, 유사한 기법이 섹션 6.2에서의 설명을 참조함으로써 채택될 수 있다.

[0505] 대안으로서, 프레임-레벨 분류기(2014)의 결과들에 기초한 단기 특징들 및 프레임-레벨 특징 추출기(2012)에 의해 획득된 프레임-레벨 특징들에 직접 기초한 단기 특징들 양쪽 모두가 단기 분류기(2024)에 의해 사용될 수 있다. 그러므로, 단기 특징 추출기(2022)는 오디오 프레임들의 시퀀스로부터 추출된 프레임-레벨 특징들 및 오디오 프레임들의 시퀀스에 대한 프레임-레벨 분류기의 결과들 양쪽 모두에 기초하여 단기 특징들을 산출하도록 구성될 수 있다.

[0506] 다시 말해서, 프레임-레벨 특징 추출기(2012)는, 다음의 특징들 중 적어도 하나를 포함하여, 섹션 6.2에 논의된 것들과 유사한 통계들 및 도 28과 관련되어 설명된 이들 단기 특징들 양쪽 모두를 산출하도록 구성될 수 있다: 다양한 단기 오디오 유형들의 특징들 특성화 속성들, 컷오프 주파수, 정적 신호-잡음-비 특성들, 분절 신호-잡음-비 특성들, 기본 스피치 디스크립터들, 및 성도 특성들.

[0507] 실시간으로 작동하기 위해, 실시예들 모두에서, 단기 특징 추출기(2022)는 미리 결정된 스텝 길이에서 장기 오디오 세그먼트의 시간적 치수에서 이동 윈도우 슬라이딩을 갖고 형성된 단기 오디오 세그먼트들 상에서 작동하

도록 구성될 수 있다. 단기 오디오 세그먼트에 대한 이동 윈도우, 뿐만 아니라 장기 오디오 세그먼트에 대한 오디오 프레임 및 이동 윈도우에 대해, 참조가 상세를 위해 섹션 1.1에 대해 이루어질 수 있다.

[0508] 6.4 실시예들 및 애플리케이션 시나리오들의 조합

[0509] 파트 1과 유사하게, 상기 논의된 모든 실시예들 및 그것의 변형들은 그것의 임의의 조합으로 구현될 수 있으며, 상이한 파트들/실시예들에서 언급되지만 동일하거나 또는 유사한 기능들을 가진 임의의 구성요소들이 동일한 또는 별개의 구성요소들로서 구현될 수 있다.

[0510] 예를 들면, 섹션 6.1 내지 섹션 6.3에 설명된 해결책들 중 임의의 둘 이상이 서로 조합될 수 있다. 조합들 중 임의의 것은 파트 1 내지 파트 5 및 나중에 설명될 다른 파트들에서 설명되거나 또는 내포된 임의의 실시예와 추가로 조합될 수 있다. 특히, 파트 1에 논의된 유형 평활화 유닛(712)은 프레임 분류기(2014), 또는 오디오 콘텐트 분류기(202), 또는 오디오 콘텍스트 분류기(204)의 결과들을 평활화하기 위해, 오디오 분류기(200)의 구성요소로서 이 파트에서 사용될 수 있다. 뿐만 아니라, 타이머(916)는 또한 오디오 분류기(200)의 출력의 갑작스런 변화를 회피하기 위해 오디오 분류기(200)의 구성요소로서 작용할 수 있다.

[0511] 6.5 오디오 분류 방법

[0512] 파트 1과 유사하게, 이전 실시예들에서 오디오 분류기를 설명하는 프로세스에서, 몇몇 프로세스들 또는 방법들이 또한 분명히 개시된다. 이후 이들 방법들의 요약은 이전에 이미 논의된 상세들의 일부를 반복하지 않고 제공된다.

[0513] 일 실시예에서, 도 30에 도시된 바와 같이, 오디오 분류 방법이 제공된다. 단기 오디오 세그먼트들(서로 중첩되거나 또는 중첩되지 않는)의 시퀀스로 구성된 장기 오디오 세그먼트의 장기 오디오 유형(콘텍스트 유형인)을 식별하기 위해, 단기 오디오 세그먼트들이 먼저 콘텐트 유형들인 단기 오디오 유형들로 분류되며(동작(3004)), 장기 특징들이 장기 오디오 세그먼트에서의 단기 세그먼트들의 시퀀스에 대하여 분류 동작의 결과들의 통계를 산출함으로써 획득된다(동작(3006)). 그 후 장기 분류(동작(3008))는 장기 특징들을 사용하여 수행될 수 있다. 단기 오디오 세그먼트는 오디오 프레임들의 시퀀스를 포함할 수 있다. 물론, 단기 세그먼트들의 단기 오디오 유형을 식별하기 위해, 단기 특징들이 그것들로부터 추출되도록 요구한다(동작(3002)).

[0514] 단기 오디오 유형들(콘텐트 유형들)은 이에 제한되지 않지만 스피치, 단기 음악, 배경 사운드 및 잡음을 포함할 수 있다.

[0515] 장기 특징들은 이에 제한되지 않지만: 단기 오디오 유형들의 신뢰도 값들의 평균 및 분산, 단기 세그먼트들의 중요도들에 의해 가중된 분산, 각각의 단기 오디오 유형의 발생 빈도 및 상이한 단기 오디오 유형들 사이에서의 전이들의 빈도를 포함할 수 있다.

[0516] 변형에서, 도 31에 도시된 바와 같이, 추가의 장기 특징들이 장기 오디오 세그먼트에서 단기 세그먼트들의 시퀀스의 단기 특징들에 기초하여 직접 획득될 수 있다(동작(3107)). 이러한 추가 장기 특징들은 이에 제한되지 않지만 단기 특징들의 다음의 통계들을 포함할 수 있다: 평균, 분산, 가중 평균, 가중 분산, 높은 평균, 낮은 평균, 및 높은 평균 및 낮은 평균 사이에서의 비.

[0517] 단기 특징들을 추출하기 위한 상이한 방식들이 있다. 하나는 분류될 단기 오디오 세그먼트로부터 단기 특징들을 직접 추출하는 것이다. 이러한 특징들은 이에 제한되지 않지만 리듬 특성들, 인터럽션들/음소거들 특성들 및 단기 오디오 품질 특징들을 포함한다.

[0518] 제 2 방식은 각각의 단기 세그먼트에 포함된 오디오 프레임들로부터 프레임-레벨 특징들을 추출하며(도 32에서 동작(3201)), 그 후 단기 특징들로서 프레임-레벨 특징들의 통계들을 산출하는 것과 같이, 프레임-레벨 특징들에 기초하여 단기 특징들을 산출하는 것이다. 프레임-레벨 특징들은 이에 제한되지 않지만: 다양한 단기 오디오 유형들의 속성들을 특성화한 특징들, 컷오프 주파수, 정적 신호-잡음-비 특성들, 분절 신호-잡음-비 특성들, 기본 스피치 디스크립터들, 및 성도 특성들을 포함할 수 있다. 다양한 단기 오디오 유형들의 속성들을 특성화한 특징들은 프레임 에너지, 서브-대역 스펙트럼 분포, 스펙트럼 플럭스, Mel-주파수 켑스트럼 계수, 베이스, 잔여 정보, 채도 특징 및 제로-교차 레이트를 포함할 수 있다.

[0519] 제 3 방식은 장기 특징들의 추출과 유사한 방식으로 단기 특징들을 추출하는 것이다: 분류될 단기 세그먼트에서의 오디오 프레임들로부터 프레임-레벨 특징들을 추출한 후(동작(3201)), 각각의 프레임-레벨 특징들을 사용하여 각각의 오디오 프레임을 프레임-레벨 오디오 유형들로 분류하는 것(도 33에서의 동작(32011)); 및 단기 특징들이 프레임-레벨 오디오 유형들(선택적으로 신뢰도 값들을 포함한)에 기초하여 단기 특징들을 산출함으로써 추

출될 수 있다(동작(3002)). 프레임-레벨 오디오 유형들은 단기 오디오 유형(콘텐트 유형)과 유사한 속성들 및 아키텍처들을 가질 수 있으며, 또한 스피치, 음악, 배경 사운드 및 잡음을 포함할 수 있다.

[0520] 제 2 방식 및 제 3 방식은 도 33에서의 대시 기호로 된 화살표에 의해 도시된 바와 같이 함께 조합될 수 있다.

[0521] 파트 1에 논의된 바와 같이, 단기 오디오 세그먼트들 및 장기 오디오 세그먼트들 양쪽 모두가 이동 윈도우들을 갖고 샘플링될 수 있다. 즉, 단기 특징들을 추출하는 동작(동작(3002))은 미리 결정된 스텝 길이에서 장기 오디오 세그먼트의 시간적 치수에서의 이동 윈도우 슬라이딩을 갖고 형성된 단기 오디오 세그먼트들에 대해 수행될 수 있으며, 장기 특징들을 추출하는 동작(동작(3107)) 및 단기 오디오 유형들의 통계들을 산출하는 동작(동작(3006))은 또한 미리 결정된 스텝 길이에서 오디오 신호의 시간적 치수에서의 이동 윈도우 슬라이딩을 갖고 형성된 장기 오디오 세그먼트들에 대해 수행될 수 있다.

[0522] 오디오 프로세싱 장치의 실시예들과 유사하게, 오디오 프로세싱 방법의 실시예들 및 그것들의 변형들의 임의의 조합이 한편으로는 실현 가능하며; 다른 한편으로, 오디오 프로세싱 방법의 실시예들 및 그것들의 변형들의 모든 측면은 별개의 해결책들일 수 있다. 또한, 이 섹션에 설명된 임의의 둘 이상의 해결책들이 서로 조합될 수 있으며, 이들 조합들은 본 개시의 다른 파트들에 설명되거나 또는 내포된 임의의 실시예와 추가로 조합될 수 있다. 특히, 섹션 6.4에 이미 논의된 바와 같이, 오디오 유형들의 평활화 기법들 및 전이 기법은 여기에 논의된 오디오 분류 방법의 일 부분일 수 있다.

[0523] 파트 7: VoIP 분류기들 및 분류 방법들

[0524] 파트 6에서 신규의 오디오 분류기가 콘텐트 유형 분류의 결과들에 적어도 부분적으로 기초하여 오디오 콘텍스트 유형들로 오디오 신호를 분류하기 위해 제안된다. 파트 6에 논의된 실시예들에서, 장기 특징들은 수 내지 수십의 초들의 길이의 장기 세그먼트로부터 추출되며, 따라서 오디오 콘텍스트 분류는 긴 대기 시간을 야기할 수 있다. 오디오 콘텍스트는 또한 단기 세그먼트 레벨에서와 같이, 실시간으로 또는 거의 실시간으로 분류될 수 있는 것이 요구된다.

[0525] 7.1 단기 세그먼트에 기초한 콘텍스트 분류

[0526] 그러므로, 도 34에 도시된 바와 같이, 오디오 분류기(200A)가 제공되며, 상기 오디오 분류기(200A)는 오디오 신호의 단기 세그먼트의 콘텐트 유형을 식별하기 위한 오디오 콘텐트 분류기(202A), 및 오디오 콘텐트 분류기에 의해 식별된 콘텐트 유형에 적어도 부분적으로 기초하여 단기 세그먼트의 콘텍스트 유형을 식별하기 위한 오디오 콘텍스트 분류기(204A)를 포함한다.

[0527] 여기에서 오디오 콘텐트 분류기(202A)는 파트 6에 이미 언급된 기술들을 채택할 수 있지만, 섹션 7.2에서 이하에 논의될 바와 같이 상이한 기술들을 또한 채택할 수 있다. 또한, 오디오 콘텍스트 분류기(204A)는, 오디오 콘텍스트 분류기(204A) 및 오디오 콘텐트 분류기(202A) 양쪽 모두가 동일한 단기 세그먼트를 분류하기 때문에 콘텍스트 분류기(204A)가, 오디오 콘텐트 분류기(202A)로부터의 결과들의 통계들을 사용하기보다는, 오디오 콘텐트 분류기(202A)의 결과들을 직접 사용할 수 있다는 차이를 갖고, 파트 6에 이미 언급된 기술들을 채택할 수 있다. 뿐만 아니라, 파트 6과 유사하게, 오디오 콘텐트 분류기(202A)로부터의 결과들 외에, 오디오 콘텍스트 분류기(204A)는 단기 세그먼트로부터 직접 추출된 다른 특징들을 사용할 수 있다. 즉, 오디오 콘텍스트 분류기(204A)는 특징들로서, 단기 세그먼트의 콘텐트 유형들의 신뢰도 값들 및 단기 세그먼트로부터 추출된 다른 특징들을 사용함으로써 기계-학습 모델에 기초하여 단기 세그먼트를 분류하도록 구성될 수 있다. 단기 세그먼트로부터 추출된 특징들에 대해, 참조가 파트 6에 대해 이루어질 수 있다.

[0528] 오디오 콘텐트 분류기(200A)는 VoIP 스피치/잡음 및/또는 비-VoIP 스피치/잡음(VoIP 스피치/잡음 및 비-VoIP 스피치/잡음은 섹션 7.2에서 이하에 논의될 것이다)보다 더 많은 오디오 유형들로서 단기 세그먼트를 동시에 라벨링할 수 있으며, 다수의 오디오 유형들의 각각은 섹션 1.2에서 논의된 바와 같이 그 자신의 신뢰도 값을 가질 수 있다. 이것은 보다 풍부한 정보가 캡처될 수 있으므로 보다 양호한 분류 정확도를 달성할 수 있다. 예를 들면, 스피치 및 단기 음악의 신뢰도 값들의 공동 정보는 어떤 정도로 오디오 콘텐트가 그것이 순수 VoIP 콘텐트로부터 식별될 수 있도록 스피치 및 배경 음악의 혼합일 가능성이 있는지를 드러낸다.

[0529] 7.2 VoIP 스피치 및 VoIP 잡음을 사용한 분류

[0530] 본 출원의 이러한 측면은 VoIP/비-VoIP 분류 시스템에서 특히 유용하며, 이것은 짧은 결정 지연 시간을 위한 현재 단기 세그먼트를 분류하기 위해 요구될 것이다.

[0531] 이러한 목적을 위해, 도 34에 도시된 바와 같이, 오디오 분류기(200A)는 VoIP/비-VoIP 분류를 위해 특별하게 설

계된다. VoIP/비-VoIP를 분류하기 위해, VoIP 스피치 분류기(2026) 및/또는 VoIP 잡음 분류기가 오디오 콘텍스트 분류기(204A)에 의해 최종 강력한 VoIP/비-VoIP 분류를 위한 중간 결과들을 생성하기 위해 개발된다.

[0532] VoIP 단기 세그먼트는 VoIP 스피치 및 VoIP 잡음을 교대로 포함할 것이다. 높은 정확도가 스피치의 단기 세그먼트를 VoIP 스피치 또는 비-VoIP 스피치로 분류하기 위해 달성될 수 있지만, 지금까지 잡음의 단기 세그먼트를 VoIP 잡음 또는 비-VoIP 잡음으로 분류하기 위한 것은 아님이 관찰된다. 따라서, 그것은 스피치 및 잡음 사이에서의 차이를 고려하지 않고 및 따라서 함께 혼합된 이들 두 개의 콘텐트 유형들(스피치 및 잡음)의 특징들을 갖고 단기 세그먼트를 VoIP(VoIP 스피치 및 VoIP 잡음을 포함하지만 구체적으로 식별되지 않은 VoIP 스피치 및 VoIP 잡음을 갖는) 및 비-VoIP로 직접 분류함으로써 변별성을 흐릿하게 만들 것이라고 결론내려질 수 있다.

[0533] 분류기들은 스피치가 잡음보다 더 많은 정보를 포함하므로 VoIP 잡음/비-VoIP 잡음 분류 및 컷오프 주파수가 스피치를 분류하기에 더 효과적이므로 이러한 특징들에 대한 것보다 VoIP 스피치/비-VoIP 스피치 분류에 대해 더 높은 정확도들을 달성하는 것이 합당하다. adaBoost 트레이닝 프로세스로부터 획득된 가중치 랭킹에 따르면, VoIP/비-VoIP 스피치 분류에 대한 최상위 가중된 단기 특징들은: 대수 에너지의 표준 편차, 컷오프 주파수, 리듬 강도의 표준 편차, 및 스펙트럼 플럭스의 표준 편차이다. 대수 에너지의 표준 편차, 리듬 강도의 표준 편차, 및 스펙트럼 플럭스의 표준 편차는 비-VoIP 스피치에 대한 것보다 VoIP 스피치에 대해 일반적으로 더 높다. 하나의 가능성 있는 이유는 영화-형 미디어 또는 게임과 같은 비-VoIP 콘텍스트에서의 많은 단기 스피치 세그먼트들이 보통 배경 음악 또는 사운드 효과와 같은 다른 사운드들과 혼합된다는 것이며, 상기 특징들의 값들은 더 낮다. 한편, 컷오프 특징은 일반적으로 비-VoIP 스피치에 대한 것보다 VoIP 스피치에 대해 더 낮으며, 이것은 많은 인기 있는 VoIP 코덱들에 의해 도입된 낮은 컷오프 주파수를 표시한다.

[0534] 그러므로, 일 실시예에서, 오디오 콘텐트 분류기(202A)는 콘텐트 유형 VoIP 스피치 또는 콘텐트 유형 비-VoIP 스피치로 단기 세그먼트를 분류하기 위한 VoIP 스피치 분류기(2026)를 포함할 수 있으며; 오디오 콘텍스트 분류기(204A)는 VoIP 스피치 및 비-VoIP 스피치의 신뢰도 값들에 기초하여 단기 세그먼트를 콘텍스트 유형 VoIP 또는 콘텍스트 유형 비-VoIP로 분류하도록 구성될 수 있다.

[0535] 또 다른 실시예에서, 오디오 콘텐트 분류기(202A)는 단기 세그먼트를 콘텐트 유형 VoIP 잡음 또는 콘텐트 유형 비-VoIP 잡음으로 분류하기 위한 VoIP 잡음 분류기(2028)를 더 포함할 수 있으며; 오디오 콘텍스트 분류기(204A)는 VoIP 스피치, 비-VoIP 스피치, VoIP 잡음 및 비-VoIP 잡음의 신뢰도 값들에 기초하여 단기 세그먼트를 콘텍스트 유형 VoIP 또는 콘텍스트 유형 비-VoIP로 분류하도록 구성될 수 있다.

[0536] VoIP 스피치, 비-VoIP 스피치, VoIP 잡음 및 비-VoIP 잡음의 콘텐트 유형들은 파트 6, 섹션 1.2 및 섹션 7.1에 논의된 바와 같이 기존의 기술들을 갖고 식별될 수 있다.

[0537] 대안적으로, 오디오 콘텐트 분류기(202A)는 도 35에 도시된 바와 같이 계층 구조를 가질 수 있다. 즉, 우리는 먼저 단기 세그먼트를 스피치 또는 잡음/배경으로 분류하기 위해 스피치/잡음 분류기(2025)로부터의 결과들을 이용한다.

[0538] 단지 VoIP 스피치 분류기(2026)만을 사용하는 실시예에 기초하여, 단기 세그먼트가 스피치/잡음 분류기(2025)(이러한 상황에서 그것은 단지 스피치 분류기이다)에 의해 스피치로서 결정된다면, VoIP 스피치 분류기(2026)는 그것이 VoIP 스피치인지 또는 비-VoIP 스피치인지를 계속해서 분류하며, 이진 분류 결과를 산출하고; 그렇지 않다면, 그것은 VoIP 스피치의 신뢰도 값이 낮거나 또는 VoIP 스피치에 대한 결정이 불확실하다고 여겨질 수 있다.

[0539] 단지 VoIP 잡음 분류기(2028)만을 사용한 실시예에 기초하여, 단기 세그먼트가 스피치/잡음 분류기(2025)(이러한 상황에서 그것은 단지 잡음(배경) 분류기이다)에 의해, 잡음으로서 결정된다면, VoIP 잡음 분류기(2028)는 계속해서 그것을 VoIP 잡음 또는 비-VoIP 잡음으로서 분류하며, 이진 분류 결과를 산출한다. 그렇지 않다면, 그것은 VoIP 잡음의 신뢰도 값이 낮거나 또는 VoIP 잡음에 대한 결정이 불확실하다고 여겨질 수 있다.

[0540] 여기에서, 일반적으로 스피치는 정보적 콘텐트 유형이며 잡음/배경은 방해 콘텐트 유형이므로, 단기 세그먼트가 잡음이 아닐지라도, 이전 단락에서의 실시예에서, 우리는 단기 세그먼트가 콘텍스트 유형 VoIP가 아니라고 명확하게 결정할 수 없다. 단기 세그먼트가 스피치가 아니면, 단지 VoIP 스피치 분류기(2026)를 사용하는 실시예에서, 그것은 아마도 콘텍스트 유형 VoIP가 아니다. 그러므로, 일반적으로 단지 VoIP 스피치 분류기(2026)만을 사용한 실시예는 독립적으로 실현될 수 있는 반면, 단지 VoIP 잡음 분류기(2028)만을 사용한 다른 실시예는 예를 들면, VoIP 스피치 분류기(2026)를 사용한 실시예와 협력하는 보완적 실시예로서 사용될 수 있다.

[0541] 즉, VoIP 스피치 분류기(2026) 및 VoIP 잡음 분류기(2028) 양쪽 모두가 사용될 수 있다. 단기 세그먼트가 스피

치/잡음 분류기(2025)에 의해 스피치로서 결정된다면, VoIP 스피치 분류기(2026)는 계속해서 그것이 VoIP 스피치인지 또는 비-VoIP 스피치인지를 분류하며, 이진 분류 결과를 산출한다. 단기 세그먼트가 스피치/잡음 분류기(2025)에 의해 잡음으로서 결정된다면, VoIP 잡음 분류기(2028)는 계속해서 그것을 VoIP 잡음 또는 비-VoIP 잡음으로 분류하며, 이진 분류 결과를 산출한다. 그렇지 않다면, 단기 세그먼트가 비-VoIP로서 분류될 수 있다고 여겨질 수 있다.

[0542] 스피치/잡음 분류기(2025), VoIP 스피치 분류기(2026) 및 VoIP 잡음 분류기(2028)의 구현은 임의의 기존의 기술들을 채택할 수 있으며, 파트 1 내지 파트 6에 논의된 오디오 콘텐트 분류기(202)일 수 있다.

[0543] 상기 설명에 따라 구현된 오디오 콘텐트 분류기(202A)가 최종적으로, 모든 관련된 신뢰도 값들이 낮음을 의미하는, 단기 세그먼트를 스피치, 잡음 및 배경 중 어떤 것도 아님, 또는 VoIP 스피치, 비-VoIP 스피치, VoIP 잡음 및 비-VoIP 잡음 중 어떤 것도 아닌 것으로 분류한다면, 오디오 콘텐트 분류기(202A)(및 오디오 콘텍스트 분류기(204A))는 비-VoIP로서 단기 세그먼트를 분류할 수 있다.

[0544] VoIP 스피치 분류기(2026) 및 VoIP 잡음 분류기(2028)의 결과들에 기초하여 VoIP 또는 비-VoIP의 콘텍스트 유형들로 단기 세그먼트를 분류하기 위해, 오디오 콘텍스트 분류기(204A)는 섹션 7.1에서 논의된 바와 같이 기계-학습 기반 기술들을 채택할 수 있으며, 섹션 7.1에서 이미 논의된 바와 같이, 변경으로서, 단기 세그먼트로부터 직접 추출된 단기 특징들 및/또는 VoIP 관련 콘텐트 유형들과 다른 콘텐트 유형들로 향해진 다른 오디오 콘텐트 분류기(들)의 결과들을 포함하여, 보다 많은 특징들이 사용될 수 있다.

[0545] 상기 설명된 기계-학습 기반 기술들 외에, VoIP/비-VoIP 분류에 대한 대안적인 접근법은 도메인 지식을 이용하며 VoIP 스피치 및 VoIP 잡음과 관련되어 분류 결과들을 이용하는 경험적 규칙일 수 있다. 이러한 경험적 규칙들의 예가 이하에서 설명될 것이다.

[0546] 시간(t)의 현재 단기 세그먼트가 VoIP 스피치 또는 비-VoIP 스피치로서 결정된다면, 분류 결과는 VoIP/비-VoIP 스피치 분류가 이전에 논의된 바와 같이 강력하므로 VoIP/비-VoIP 분류 결과로서 직접 취해진다. 즉, 단기 세그먼트가 VoIP 스피치로서 결정된다면, 그것은 콘텍스트 유형 VoIP이고; 단기 세그먼트가 비-VoIP 스피치로서 결정된다면, 그것은 콘텍스트 유형 비-VoIP이다.

[0547] VoIP 스피치 분류기(2026)가 상기 언급된 바와 같이 스피치/잡음 분류기(2025)에 의해 결정된 스피치에 대하여 VoIP 스피치/비-VoIP 스피치에 관한 이진 결정을 할 때, VoIP 스피치 및 비-VoIP 스피치의 신뢰도 값들은 보완적일 수 있으며, 즉 그것의 합은 1이고(0이 100%가 아님을 나타내며 1이 100% 임을 나타낸다면), VoIP 스피치 및 비-VoIP 스피치를 구별하기 위한 신뢰도 값의 임계치들은 실제로 동일한 포인트를 표시할 수 있다. VoIP 스피치 분류기(2026)가 이진 분류기가 아니면, VoIP 스피치 및 비-VoIP 스피치의 신뢰도 값들은 보완적이지 않을 수 있으며, VoIP 스피치 및 비-VoIP 스피치를 구별하기 위한 신뢰도 값의 임계치들은 반드시 동일한 포인트를 나타내는 것은 아닐 수 있다.

[0548] 그러나, VoIP 스피치 또는 비-VoIP 스피치 신뢰도가 임계치에 가까우며 그 주변에서 변동하는 경우에, VoIP/비-VoIP 분류 결과들은 너무 빈번하게 스위칭하는 것이 가능하다. 이러한 변동을 회피하기 위해, 버퍼 기법이 제공될 수 있다: VoIP 스피치 및 비-VoIP 스피치에 대한 양쪽 임계치들 모두가 너무 크게 설정될 수 있으며, 따라서 현재 콘텐트 유형에서 다른 콘텐트 유형으로 스위칭하는 것은 너무 용이하지 않다. 설명의 용이함을 위해, 우리는 비-VoIP 스피치에 대한 신뢰도 값을 IP 스피치의 신뢰도 값으로 변환할 수 있다. 즉, 신뢰도 값이 높으면, 단기 세그먼트는 VoIP 스피치에 더 가까운 것으로 간주되며, 신뢰도 값이 낮으면, 단기 세그먼트는 비-VoIP 스피치에 더 가까운 것으로 간주된다. 상기 설명된 비-이진 분류기에 대해, 비-VoIP 스피치의 높은 신뢰도 값이 반드시 VoIP 스피치의 낮은 신뢰도 값을 의미하는 것은 아니지만, 이러한 간소화는 해결책의 본질을 잘 반영할 수 있으며 이진 분류기들의 언어를 갖고 설명된 관련 청구항들은 비-이진 분류기들에 대한 등가의 해결책들을 커버하는 것으로서 해석될 것이다.

[0549] 버퍼 기법은 도 36에 도시된다. 두 개의 임계치들(Th1 및 Th2)(Th1>=Th2) 사이에 버퍼 영역이 있다. VoIP 스피치의 신뢰도 값(v(t))이 영역에서 떨어질 때, 콘텍스트 분류는, 도 36에서의 좌측 및 우측 측면들 상에서의 화살표들에 의해 도시된 바와 같이, 변하지 않을 것이다. 단지 신뢰도 값(v(t))이 보다 큰 임계치(Th1)보다 클 때만, 단기 세그먼트가 VoIP(도 36에서의 최하부 상에서의 화살표에 의해 도시된 바와 같이)로 분류될 것이며; 단지 신뢰도 값이 보다 작은 임계치(Th2)보다 크지 않을 때만, 단기 세그먼트가 비-VoIP(도 36에서의 최상부 상에서의 화살표에 의해 도시된 바와 같이)로 분류될 것이다.

[0550] VoIP 잡음 분류기(2028)가 대신에 사용된다면, 상황은 유사하다. 해결책을 더 강력하게 만들기 위해, VoIP 스피

치 분류기(2026) 및 VoIP 잡음 분류기(2028)는 공동으로 사용될 수 있다. 그 후, 오디오 콘텍스트 분류기(204A)는: VoIP 스피치의 신뢰도 값이 제 1 임계치보다 크다면 또는 VoIP 잡음의 신뢰도 값이 제 3 임계치보다 크다면 단기 세그먼트를 콘텍스트 유형 VoIP로 분류하고; VoIP 스피치의 신뢰도 값이 제 2 임계치보다 크지 않다면(여기에서 제 2 임계치는 제 1 임계치보다 크지 않다) 또는 VoIP 잡음의 신뢰도 값이 제 4 임계치보다 크지 않다면(여기에서 제 4 임계치는 제 3 임계치보다 크지 않다) 단기 세그먼트를 콘텍스트 유형 비-VoIP로 분류하며; 그 외 지난 단기 세그먼트에 대해 단기 세그먼트를 콘텍스트 유형으로 분류하도록 구성될 수 있다.

[0551] 여기에서, 특히 이에 제한되지 않지만 이진 VoIP 스피치 분류기 및 이진 VoIP 잡음 분류기에 대해, 제 1 임계치는 제 2 임계치와 동일할 수 있으며, 제 3 임계치는 제 4 임계치와 동일할 수 있다. 그러나, 일반적으로 VoIP 잡음 분류 결과는 그렇게 강력하지 않기 때문에, 제 3 및 제 4 임계치들이 서로 동일하지 않다면 더 양호할 것이며, 양쪽 모두는 결코 0.5가 아니어야 한다(0은 비-VoIP 잡음이도록 높은 신뢰도를 표시하며 1은 VoIP 잡음이도록 높은 신뢰도를 표시한다).

[0552] 7.3 평활화 변동

[0553] 빠른 변동을 회피하기 위해, 또 다른 해결책은 오디오 콘텐트 분류기에 의해 결정된 바와 같이 신뢰도 값을 평활화하는 것이다. 그러므로, 도 37에 도시된 바와 같이, 유형 평활화 유닛(203A)은 오디오 분류기(200A)에 포함될 수 있다. 이전에 논의된 바와 같이, 4개의 VoIP 관련 콘텐트 유형들의 각각의 신뢰도 값에 대해, 섹션 1.3에 논의된 평활화 기법들이 채택될 수 있다.

[0554] 대안적으로, 섹션 7.2와 유사하게, VoIP 스피치 및 비-VoIP 스피치가 보완적 신뢰도 값들을 가진 쌍으로서 간주될 수 있으며; VoIP 잡음 및 비-VoIP 잡음은 보완적 신뢰도 값들을 가진 쌍으로서 또한 간주될 수 있다. 이러한 상황에서, 각각의 쌍 외의 단지 하나만이 평활화되도록 요구되며, 섹션 1.3에 논의된 평활화 기법들이 채택될 수 있다.

[0555] 예로서, VoIP 스피치의 신뢰도 값을 취하면, 공식(3)은 다음과 같이 다시 쓰여질 수 있다:

[0556] $$v(t) = \beta \cdot v(t-1) + (1-\beta) \cdot voipSpeechConf\,(t) \qquad (3")$$

[0557] 여기에서 v(t)는 시간(t)에서 평활화된 VoIP 스피치 신뢰도 값이고, v(t-1)은 지난 시간에서의 평활화된 VoIP 스피치 신뢰도 값이며, voipSpeechConf는 평활화하기 전에 현재 시간(t)에서 VoIP 스피치 신뢰도이며, α는 가중 계수이다.

[0558] 변형에서, 상기 설명된 바와 같이 스피치/잡음 분류기(2025)가 있다면, 단-세그먼트에 대한 스피치의 신뢰도 값이 낮다면, 단기 세그먼트는 강력하게 VoIP 스피치로서 분류될 수 없으며, 우리는 VoIP 스피치 분류기(2026)를 실제로 작동하게 하지 않고 voipSpeechConf(t)=v(t-1)를 직접 설정할 수 있다.

[0559] 대안적으로, 상기 설명된 상황에서, 우리는 불확실한 경우(여기에서, 신뢰도 = 1은 그것이 VoIP라는 높은 신뢰도를 표시하며 신뢰도 = 0 은 그것이 VoIP가 아니라는 높은 신뢰도를 표시한다)를 표시하는 voipSpeechConf(t)=0.5(또는 0.4 내지 0.5와 같이, 0.5보다 높지 않은 다른 값)를 설정할 수 있다.

[0560] 그러므로, 변형에 따르면, 도 37에 도시된 바와 같이, 오디오 콘텐트 분류기(200A)는 단기 세그먼트의 스피치의 콘텐트 유형을 식별하기 위한 스피치/잡음 분류기(2025)를 더 포함할 수 있으며, 유형 평활화 유닛(203A)은 미리 결정된 신뢰도 값(0.4 내지 0.5와 같은, 0.5 또는 다른 값과 같은)으로서 평활화하기 전에 현재 단기 세그먼트에 대한 VoIP 스피치의 신뢰도 값 또는 스피치/잡음 분류기에 의해 분류된 바와 같이 콘텐트 유형 스피치에 대한 신뢰도 값이 제 5 임계치보다 낮은 지난 단기 세그먼트의 평활화된 신뢰도 값을 설정하도록 구성될 수 있다. 이러한 상황에서, VoIP 스피치 분류기(2026)는 작동하거나 또는 작동하지 않을 수 있다. 대안적으로 신뢰도 값의 설정은 VoIP 스피치 분류기(2026)에 의해 행해질 수 있으며, 이것은 작업이 유형 평활화 유닛(203A)에 의해 행해지는 해결책과 같고, 청구항은 양쪽 상황들 모두를 커버하는 것으로 해석될 것이다. 또한, 여기에서 우리는 언어 "스피치/잡음 분류기에 의해 분류된 바와 같이 콘텐트 유형 스피치에 대한 신뢰도 값은 제 5 임계치보다 작다"를 사용하지만, 보호의 범위는 이에 제한되지 않으며, 그것은 단기 세그먼트가 스피치와 다른 콘텐트 유형들로 분류되는 상황과 같다.

[0561] VoIP 잡음의 신뢰도 값에 대해, 상황은 유사하며 상세한 설명은 여기에서 생략된다.

[0562] 빠른 변동을 회피하기 위해, 또 다른 해결책은 오디오 콘텍스트 분류기(204A)에 의해 결정된 바와 같이 신뢰도 값을 평활화하는 것이며, 섹션 1.3에 논의된 평활화 기법들이 채택될 수 있다.

[0563] 빠른 변동을 회피하기 위해, 또 다른 해결책이 VoIP 및 비-VoIP 사이에서의 콘텍스트 유형의 전이를 지연시키는 것이며, 섹션 1.6에 설명된 것과 동일한 기법이 사용될 수 있다. 섹션 1.6에 설명된 바와 같이, 타이머(916)는 오디오 분류기의 바깥쪽에 또는 그것의 일 부분으로서 오디오 분류기 내에 있을 수 있다. 그러므로, 도 38에 도시된 바와 같이, 오디오 분류기(200A)는 타이머(916)를 더 포함할 수 있다. 오디오 분류기는 새로운 콘텍스트 유형의 지속 시간의 길이가 제 6 임계치에 도달할 때까지(콘텍스트 유형은 오디오 유형의 인스턴스이다) 현재 콘텍스트 유형을 계속해서 출력하도록 구성된다. 섹션 1.6을 참조함으로써, 상세한 설명이 여기에서 생략될 수 있다.

[0564] 또한 또는 대안적으로, VoIP 및 비-VoIP 사이에서의 전이를 지연시키기 위한 또 다른 기법으로서, VoIP/비-VoIP 분류에 대해 이전에 설명된 바와 같이 제 1 및/또는 제 2 임계치는 지난 단기 세그먼트의 콘텍스트 유형에 의존하여 상이할 수 있다. 즉, 제 1 및/또는 제 2 임계치는 새로운 단기 세그먼트의 콘텍스트 유형이 지난 단기 세그먼트의 콘텍스트 유형과 상이할 때 더 커지는 반면, 새로운 단기 세그먼트의 콘텍스트 유형이 지난 단기 세그먼트의 콘텍스트 유형과 동일할 때 더 작아진다. 이러한 방식으로, 콘텍스트 유형은 현재 콘텍스트 유형에 유지되려는 경향이 있으며 따라서 콘텍스트 유형의 갑작스런 변동이 어느 정도 억제될 수 있다.

[0565] 7.4 실시예들 및 애플리케이션 시나리오들의 조합

[0566] 파트 1과 유사하게, 상기 논의된 모든 실시예들 및 그것의 변형들이 그것의 임의의 조합으로 구현될 수 있으며, 상이한 파트들/실시예들에 언급되지만 동일하거나 또는 유사한 기능들을 가진 임의의 구성요소들이 동일하거나 또는 별개의 구성요소들로서 구현될 수 있다.

[0567] 예를 들면, 섹션 7.1 내지 섹션 7.3에 설명된 해결책들 중 임의의 둘 이상은 서로 조합될 수 있다. 조합들 중 임의의 것이 파트 1 내지 파트 6에 설명되거나 또는 내포된 임의의 실시예와 추가로 조합될 수 있다. 특히, 이 파트에 논의된 실시예들 및 그것의 임의의 조합이 오디오 프로세싱 장치/방법 또는 파트 4에 논의된 볼륨 레벨러 제어기/제어 방법의 실시예들과 조합될 수 있다.

[0568] 7.5 VoIP 분류 방법

[0569] 파트 1과 유사하게, 이전 실시예들에서 오디오 분류기를 설명하는 프로세스에서, 몇몇 프로세스들 또는 방법들이 또한 분명히 개시된다. 이후 이들 방법들의 요약은 이전에 이미 논의된 상세들 중 일부를 반복하지 않고 주어진다.

[0570] 도 39에 도시된 바와 같은 일 실시예에서, 오디오 분류 방법은 오디오 신호의 단기 세그먼트의 콘텐트 유형을 식별하는 단계(동작(4004)), 그 후 식별된 대로 콘텐트 유형에 적어도 부분적으로 기초하여 단기 세그먼트의 콘텍스트 유형을 식별하는 단계(동작(4008))를 포함한다.

[0571] 동적으로 및 빠르게 오디오 신호의 콘텍스트 유형을 식별하기 위해, 이 파트에서 오디오 분류 방법은 콘텍스트 유형 VoIP 및 비-VoIP를 식별할 때 특히 유용하다. 이러한 상황에서, 단기 세그먼트는 먼저 콘텐트 유형 VoIP 스피치 또는 콘텐트 유형 비-VoIP 스피치로 분류될 수 있으며, 콘텍스트 유형을 식별하는 동작은 VoIP 스피치 및 비-VoIP 스피치의 신뢰도 값들에 기초하여 단기 세그먼트를 콘텍스트 유형 VoIP 또는 콘텍스트 유형 비-VoIP로 분류하도록 구성된다.

[0572] 대안적으로, 단기 세그먼트는 먼저 콘텐트 유형 VoIP 잡음 또는 콘텐트 유형 비-VoIP 잡음으로 분류될 수 있으며, 콘텍스트 유형을 식별하는 동작은 VoIP 잡음 및 비-VoIP 잡음의 신뢰도 값에 기초하여 단기 세그먼트를 콘텍스트 유형 VoIP 또는 콘텍스트 유형 비-VoIP로 분류하도록 구성될 수 있다.

[0573] 스피치 및 잡음은 함께 고려될 수 있다. 이러한 상황에서, 콘텍스트 유형을 식별하는 동작은 VoIP 스피치, 비-VoIP 스피치, VoIP 잡음 및 비-VoIP 잡음의 신뢰도 값들에 기초하여 단기 세그먼트를 콘텍스트 유형 VoIP 또는 콘텍스트 유형 비-VoIP로 분류하도록 구성될 수 있다.

[0574] 단기 세그먼트의 콘텍스트 유형을 식별하기 위해, 기계-학습 모델이 사용될 수 있으며, 특징들로서 단기 세그먼트의 콘텐트 유형들의 신뢰도 값들 및 단기 세그먼트로부터 추출된 다른 특징들 양쪽 모두를 취한다.

[0575] 콘텍스트 유형을 식별하는 동작은 또한 경험적 규칙들에 기초하여 실현될 수 있다. 단지 VoIP 스피치 및 비-VoIP 스피치만이 수반될 때, 경험적 규칙은 이와 같다: VoIP 스피치의 신뢰도 값이 제 1 임계치보다 크다면 단기 세그먼트를 콘텍스트 유형 VoIP로 분류하고; VoIP 스피치의 신뢰도 값이 제 2 임계치보다 크지 않다면(제 2 임계치는 제 1 임계치보다 크지 않다) 단기 세그먼트를 콘텍스트 유형 비-VoIP로 분류하며; 그 외 단기 세그먼

트를 지난 단기 세그먼트에 대한 콘텍스트 유형으로서 분류한다.

[0576] 단지 VoIP 잡음 및 비-VoIP 잡음만이 수반되는 상황에 대한 경험적 규칙은 유사하다.

[0577] 스피치 및 잡음 양쪽 모두가 수반될 때, 경험적 규칙은 이와 같다: VoIP 스피치의 신뢰도 값이 제 1 임계치보다 크다면 또는 VoIP 잡음의 신뢰도 값이 제 3 임계치보다 크다면 단기 세그먼트를 콘텍스트 유형 VoIP로 분류하고; VoIP 스피치의 신뢰도 값이 제 2 임계치보다 크지 않다면(제 2 임계치는 제 1 임계치보다 크지 않다) 또는 VoIP 잡음의 신뢰도 값이 제 4 임계치보다 크지 않다면(제 4 임계치는 제 3 임계치보다 크지 않다) 단기 세그먼트를 콘텍스트 유형 비-VoIP로 분류하며; 그렇지 않다면 단기 세그먼트를 지난 단기 세그먼트에 대한 콘텍스트 유형으로 분류한다.

[0578] 섹션 1.3 및 섹션 1.8에 논의된 평활화 기법이 여기에 채택될 수 있으며 상세한 설명은 생략된다. 섹션 1.3에 설명된 평활화 기법에 대한 변경으로서, 평활화 동작(4106) 전에, 방법은 단기 세그먼트로부터 콘텐트 유형 스피치를 식별하는 단계를 더 포함하며(도 40에서 동작(40040)), 여기에서 평활화하기 전에 현재 단기 세그먼트에 대한 VoIP 스피치의 신뢰도 값은 콘텐트 유형 스피치에 대한 신뢰도 값이 제 5 임계치(동작(40041)에서 "N")보다 낮은 지난 단기 세그먼트의 미리 결정된 신뢰도 값 또는 평활화된 신뢰도 값으로서 설정된다(도 40에서 동작(40044)).

[0579] 그 외 콘텐트 유형 스피치를 강력하게 식별하는 동작이 스피치(동작(40041)에서 "Y")로서 단기 세그먼트를 판단한다면, 단기 세그먼트는 평활화 동작(4106) 전에, VoIP 스피치 또는 비-VoIP 스피치로 추가로 분류된다(동작(40042)).

[0580] 사실상, 평활화 기법을 사용하지 않고도, 방법은 또한, 단기 세그먼트가 스피치 또는 잡음으로서 분류될 때, 콘텐트 유형 스피치 및/또는 잡음을 먼저 식별할 수 있으며, 추가 분류는 단기 세그먼트를 VoIP 스피치 및 비-VoIP 스피치, 또는 VoIP 잡음 및 비-VoIP 잡음 중 하나로 분류하도록 구현된다. 그 후 콘텍스트 유형을 식별하는 동작이 이루어진다.

[0581] 섹션 1.6 및 섹션 1.8에 언급된 바와 같이, 그 안에 논의된 전이 기법이 여기에서 설명된 오디오 분류 방법의 일 부분으로서 취해질 수 있으며, 상세는 생략된다. 간단하게, 방법은 콘텍스트 유형을 식별하는 동작이 계속해서 동일한 콘텍스트 유형을 출력하는 지속 시간을 측정하는 단계를 더 포함할 수 있으며, 여기에서 오디오 분류 방법은 새로운 콘텍스트 유형의 지속 시간의 길이가 제 6 임계치에 도달할 때까지 현재 콘텍스트 유형을 계속해서 출력하도록 구성된다.

[0582] 유사하게, 상이한 제 6 임계치들은 일 콘텍스트 유형에서 또 다른 콘텍스트 유형으로 상이한 전이 쌍들에 대해 설정될 수 있다. 또한, 제 6 임계치는 새로운 콘텍스트 유형의 신뢰도 값과 음으로 상관될 수 있다.

[0583] 특히 VoIP/비-VoIP 분류에 관한 오디오 분류 방법에서의 전이 기법에 대한 변경으로서, 현재 단기 세그먼트에 대한 제 1 내지 제 4 임계치 중 임의의 하나 이상은 지난 단기 세그먼트의 콘텍스트 유형에 의존하여 상이하게 설정될 수 있다.

[0584] 오디오 프로세싱 장치의 실시예들과 유사하게, 오디오 프로세싱 방법의 실시예들 및 그것들의 변형들의 임의의 조합은 한편으로 실현 가능하며; 다른 한편으로 오디오 프로세싱 방법의 실시예들 및 그것들의 변형들의 모든 측면은 별개의 해결책들일 수 있다. 또한, 이 섹션에 설명된 임의의 둘 이상의 해결책들은 서로 조합될 수 있으며; 이들 조합들은 본 개시의 다른 파트들에 설명되거나 또는 내포된 임의의 실시예와 추가로 조합될 수 있다. 구체적으로, 여기에 설명된 오디오 분류 방법은 이전에 설명된 오디오 프로세싱 방법, 특히 볼륨 레벨러 제어 방법에서 사용될 수 있다.

[0585] 본 출원의 상세한 설명의 처음에 논의된 바와 같이, 출원의 실시예는 하드웨어에서 또는 소프트웨어에서, 또는 양쪽 모두에서 구체화될 수 있다. 도 41은 본 출원의 측면들을 구현하기 위한 대표적인 시스템을 예시한 블록도이다.

[0586] 도 41에서, 중앙 프로세싱 유닛(CPU)(4201)은 판독 전용 메모리(ROM)(4202)에 저장된 프로그램 또는 저장 섹션(4208)에서 랜덤 액세스 메모리(RAM)(4203)로 로딩된 프로그램에 따라 다양한 프로세스들을 수행한다. RAM(4203)에서, CPU(4201)가 다양한 프로세스들 등을 수행할 때 요구된 데이터가 또한 요구된 대로 저장된다.

[0587] CPU(4201), ROM(4202) 및 RAM(4203)은 버스(4204)를 통해 서로 연결된다. 입력/출력 인터페이스(4205)가 또한 버스(4204)에 연결된다.

[0588] 다음의 구성요소들이 입력/출력 인터페이스(4205)에 연결된다: 키보드, 마우스 등을 포함한 입력 섹션(4206); 음극선관(CRT), 액정 디스플레이(LCD) 등과 같은 디스플레이, 및 라우드스피커 등을 포함한 출력 섹션(4207); 하드 디스크 등을 포함한 저장 섹션(4208); 및 LAN 카드, 모뎀 등과 같은 네트워크 인터페이스 카드를 포함한 통신 섹션(4209). 통신 섹션(4209)은 인터넷과 같은 네트워크를 통해 통신 프로세스를 수행한다.

[0589] 드라이브(4210)는 요구된 대로 입력/출력 인터페이스(4205)에 또한 연결된다. 자기 디스크, 광 디스크, 자기-광학 디스크, 반도체 메모리 등과 같은, 착탈 가능한 매체(4211)가 요구된 대로 드라이브(4210) 상에 장착되며, 따라서 그로부터 판독된 컴퓨터 프로그램은 요구된 대로 저장 섹션(4208)으로 설치된다.

[0590] 상기 설명된 구성요소들이 소프트웨어에 의해 구현되는 경우에, 소프트웨어를 구성하는 프로그램은 인터넷과 같은 네트워크 또는 착탈 가능한 매체(4211)와 같은 저장 매체로부터 설치된다.

[0591] 여기에 사용된 용어는 단지 특정한 실시예들을 설명하기 위한 것이며 출원을 제한하도록 의도되지 않는다는 것을 주의하자. 여기에 사용된 바와 같이, 단수 형태("a", "an" 및 "the")는, 문맥이 달리 명확하게 표시하지 않는다면, 또한 복수 형태들을 포함하도록 의도된다. 용어들("포함하다" 및/또는 "포함하는")은, 본 명세서에 사용될 때, 서술된 특징들, 정수들, 동작들, 단계들, 요소들, 및/또는 구성요소들의 존재를 특정하며, 하나 이상의 다른 특징들, 정수들, 동작들, 단계들, 요소들, 구성요소들, 및/또는 그것의 그룹들의 존재 또는 부가를 배제하지 않는다는 것이 추가로 이해될 것이다.

[0592] 모든 수단 또는 동작의 대응하는 구조들, 재료들, 동작들, 및 등가물들 더하기 이하의 청구항들에서의 기능 요소들은 구체적으로 주장된 바와 같이 다른 주장된 요소들과 조합하여 기능을 수행하기 위해 임의의 구조, 재료, 또는 동작을 포함하도록 의도된다. 본 출원의 설명은 예시 및 설명의 목적들을 위해 제공되었지만, 철저하거나 또는 개시된 형태에서의 출원에 제한되도록 의도되지 않는다. 많은 변경들 및 변형들이 출원의 범위 및 사상으로부터 벗어나지 않고 이 기술분야의 숙련자들에게 명백할 것이다. 실시예는 출원 및 실현 가능한 출원의 원리들을 최고로 설명하기 위해, 및 이 기술분야의 다른 숙련자들이 고려된 특정한 용도에 적합한 것으로 다양한 변경들을 가진 다양한 실시예들에 대한 적용을 이해할 수 있게 하기 위해 선택되며 설명되었다.

부호의 설명

[0593] 100: 콘텐트-적응적 오디오 프로세싱 장치 200: 오디오 분류기

201: 프레임 분류기 202: 오디오 콘텐트 분류기

204: 오디오 콘텍스트 분류기 300: 조정 유닛

400: 오디오 개선 디바이스 402: 다이얼로그 강화기

404: 서라운드 버추얼라이저 406: 볼륨 레벨러

408: 등화기 712: 유형 평활화 유닛

814: 파라미터 평활화 유닛 916: 타이머

1500: 다이얼로그 강화기 제어기

1800: 서라운드 버추얼라이저 제어기 2000: 볼륨 레벨러 제어기

2012: 프레임-레벨 특징 추출기 2014: 프레임-레벨 분류기

2022: 단기 특징 추출기 2024: 단기 분류기

2025: 스피치/잡음 분류기 2026: VoIP 스피치 분류기

2028: VoIP 잡음 분류기 2042: 통계 추출기

2044: 장기 분류기 2046: 장기 특징 추출기

2200: 등화기 제어기 4022: 최소 추적 유닛

4201: 중앙 프로세싱 유닛 4202: 판독 전용 메모리

4203: 랜덤 액세스 메모리 4204: 버스

4205: 입력/출력 인터페이스

4206: 입력 섹션

4207: 출력 섹션

4208: 저장 섹션

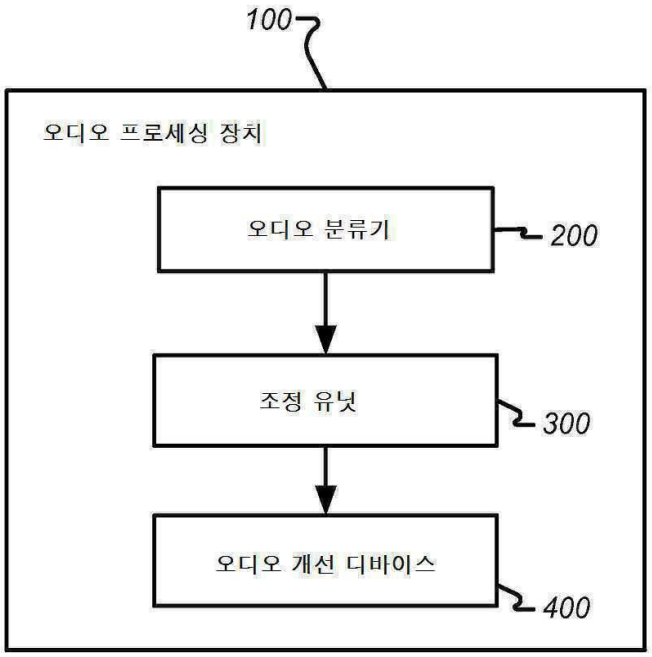
4209: 통신 섹션

4210: 드라이브

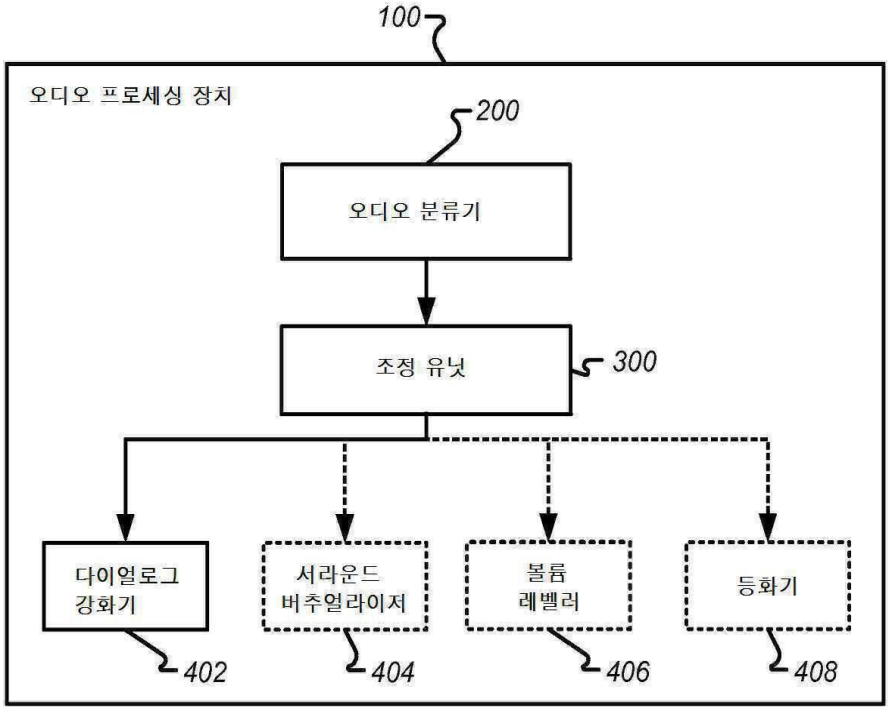
4211: 착탈 가능한 매체

도면

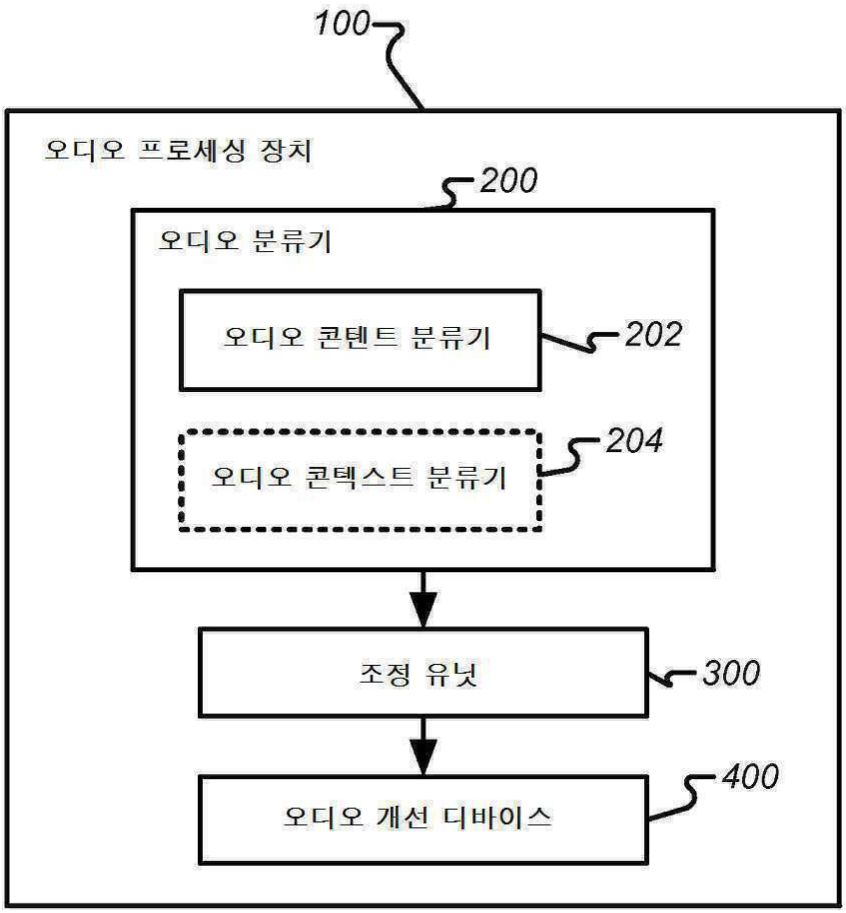
도면1



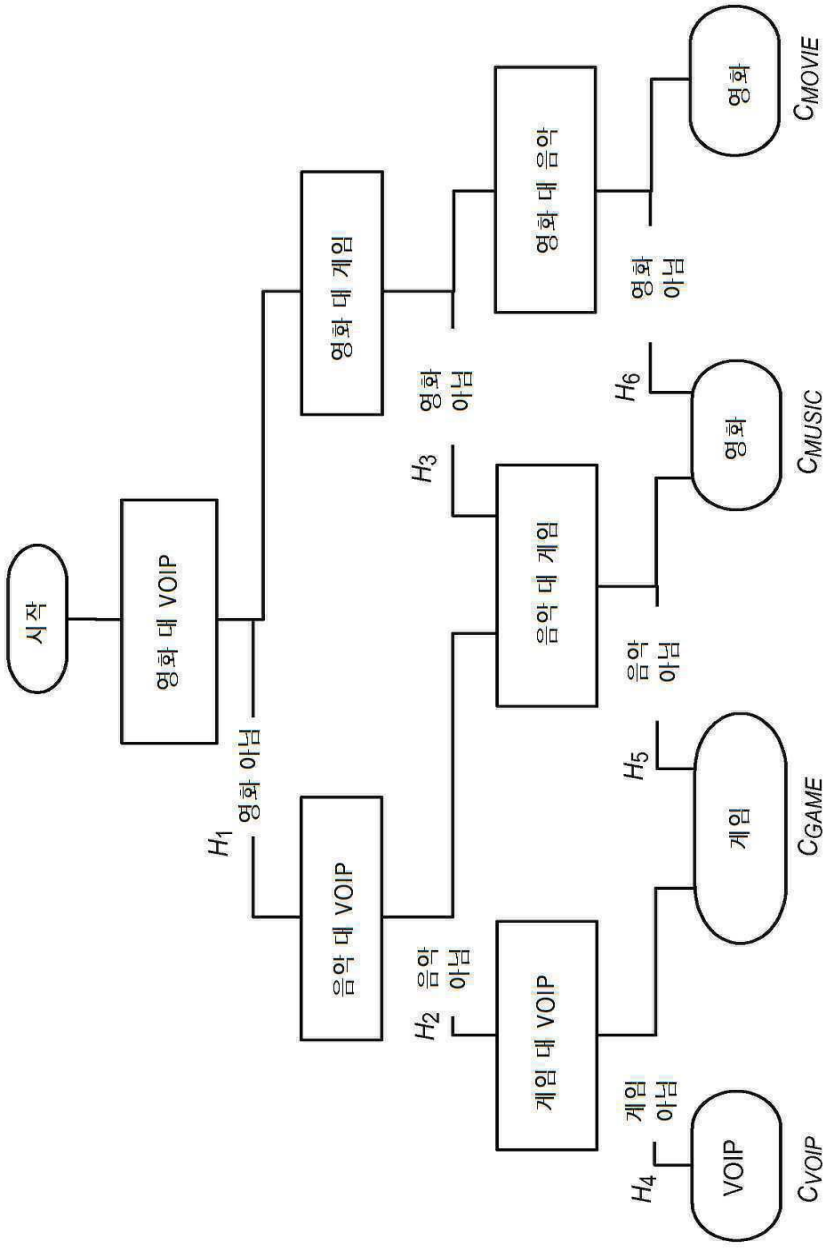
도면2



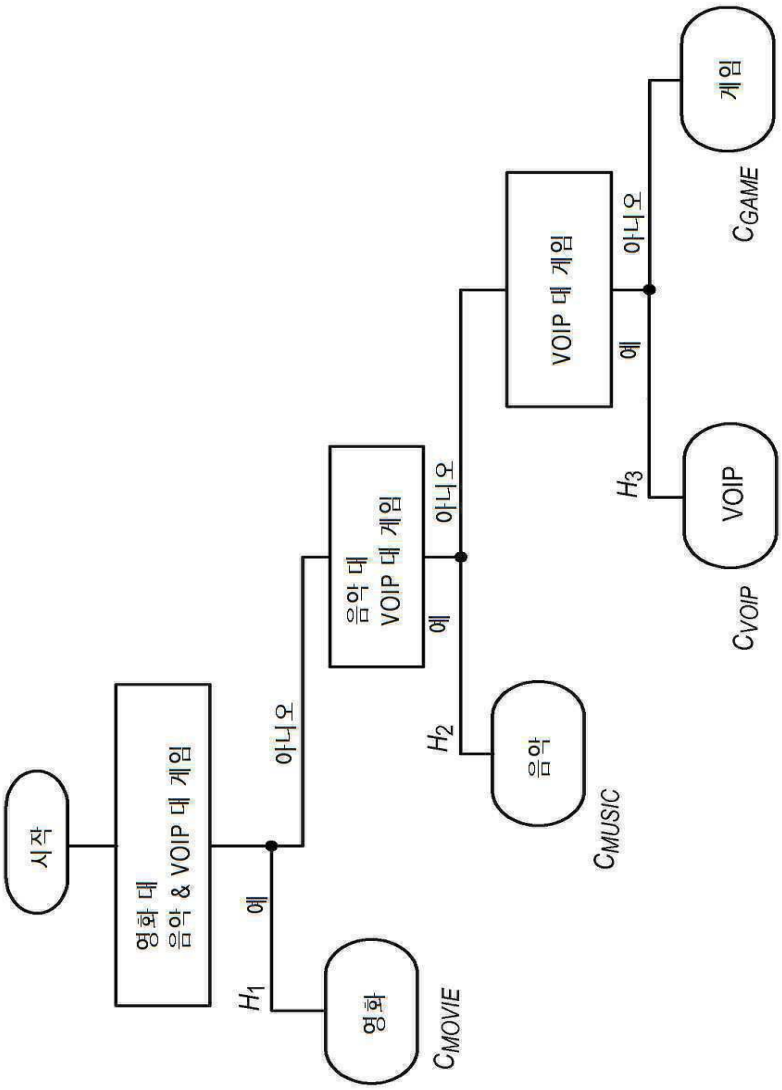
도면3


100
오디오 프로세싱 장치
200
오디오 분류기
오디오 콘텐트 분류기
202
오디오 콘텍스트 분류기
204
조정 유닛
300
오디오 개선 디바이스
400

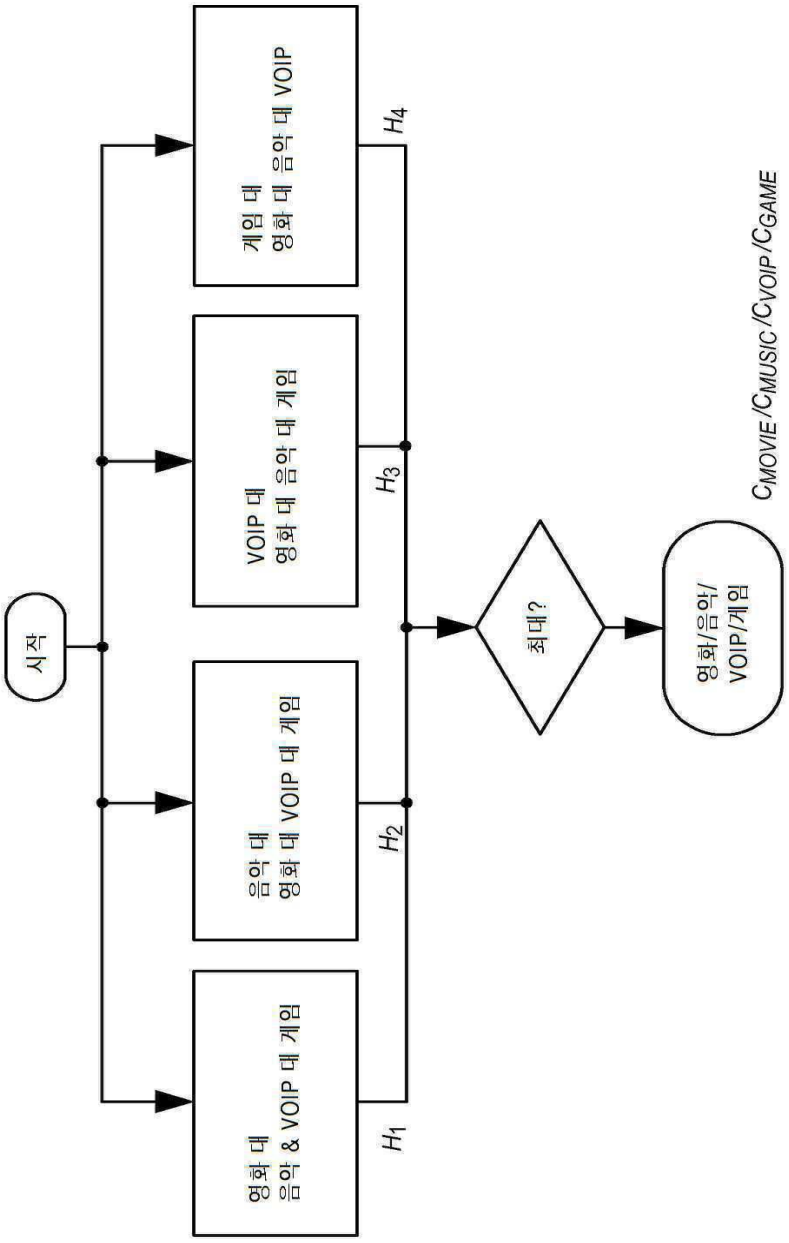
도면4


시작
영화 대 VOIP
영화 아님
H_1
영화 대 게임
음악 대 VOIP
영화
아님
H_3
영화 대 음악
음악 대 게임
영화
C_{MOVIE}
영화
아님
H_6
영화
C_{MUSIC}
음악
아님
H_5
음악
아님
H_2
게임 대 VOIP
게임
C_{GAME}
게임
아님
H_4
VOIP
C_{VOIP}

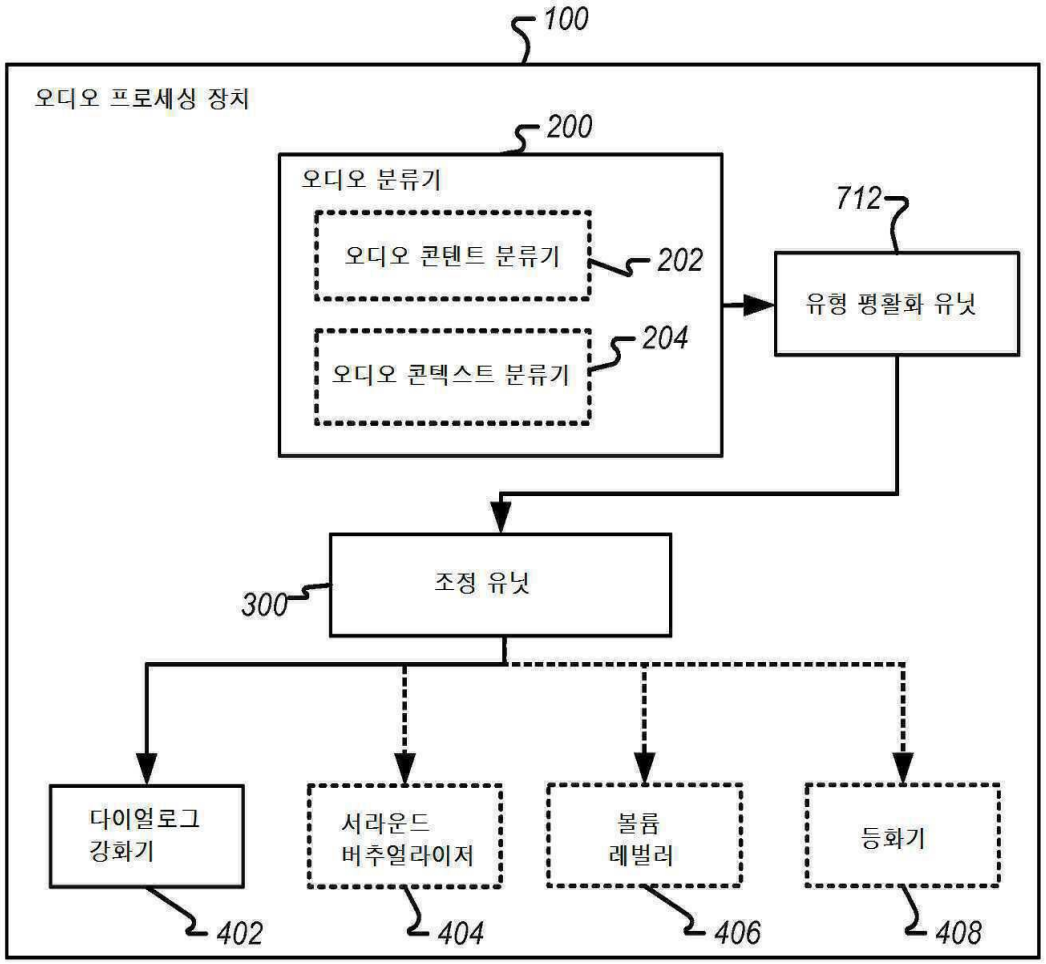
도면5


시작
영화 대
음악 & VOIP 대 게임
예
아니오
H_1
영화
C_{MOVIE}
음악 대
VOIP 대 게임
예
아니오
H_2
음악
C_{MUSIC}
VOIP 대 게임
예
아니오
H_3
VOIP
C_{VOIP}
게임
C_{GAME}

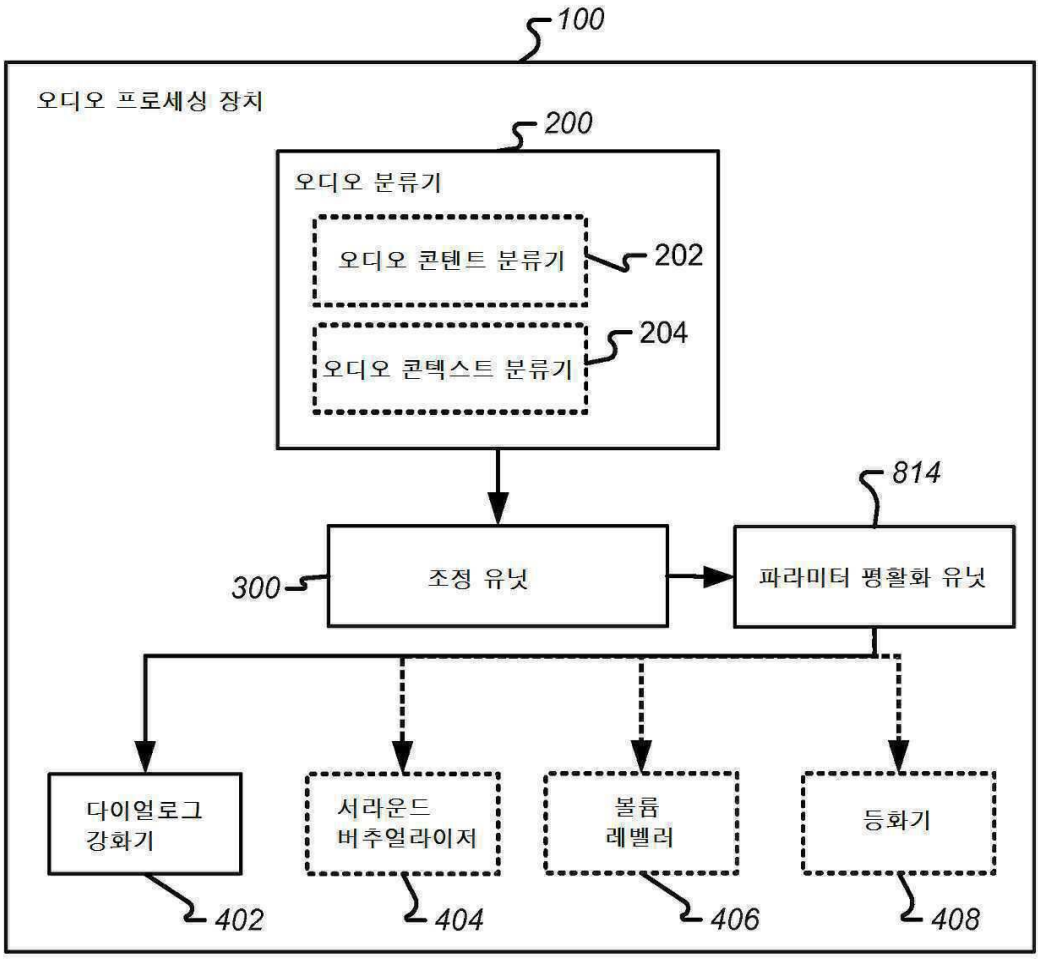
도면6


시작
영화 대
음악 & VOIP 대 게임
음악 대
영화 대 VOIP 대 게임
VOIP 대
영화 대 음악 대 게임
게임 대
영화 대 음악 대 VOIP
H1
H2
H3
H4
최대?
영화/음악/
VOIP/게임
CMOVIE/CMUSIC/CVOIP/CGAME

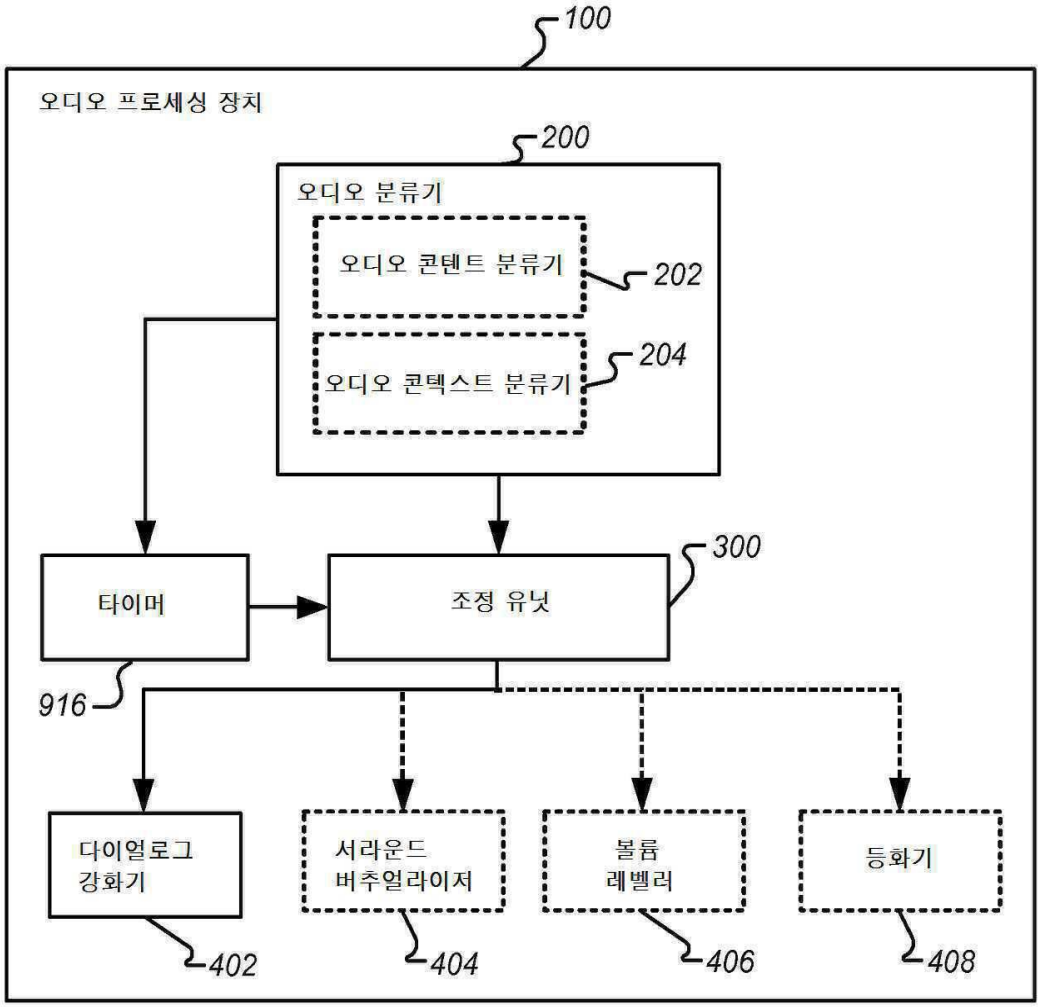
도면7


100
오디오 프로세싱 장치
200
오디오 분류기
오디오 콘텐트 분류기
202
오디오 콘텍스트 분류기
204
712
유형 평활화 유닛
조정 유닛
300
다이얼로그 강화기
서라운드 버추얼라이저
볼륨 레벌러
등화기
402
404
406
408

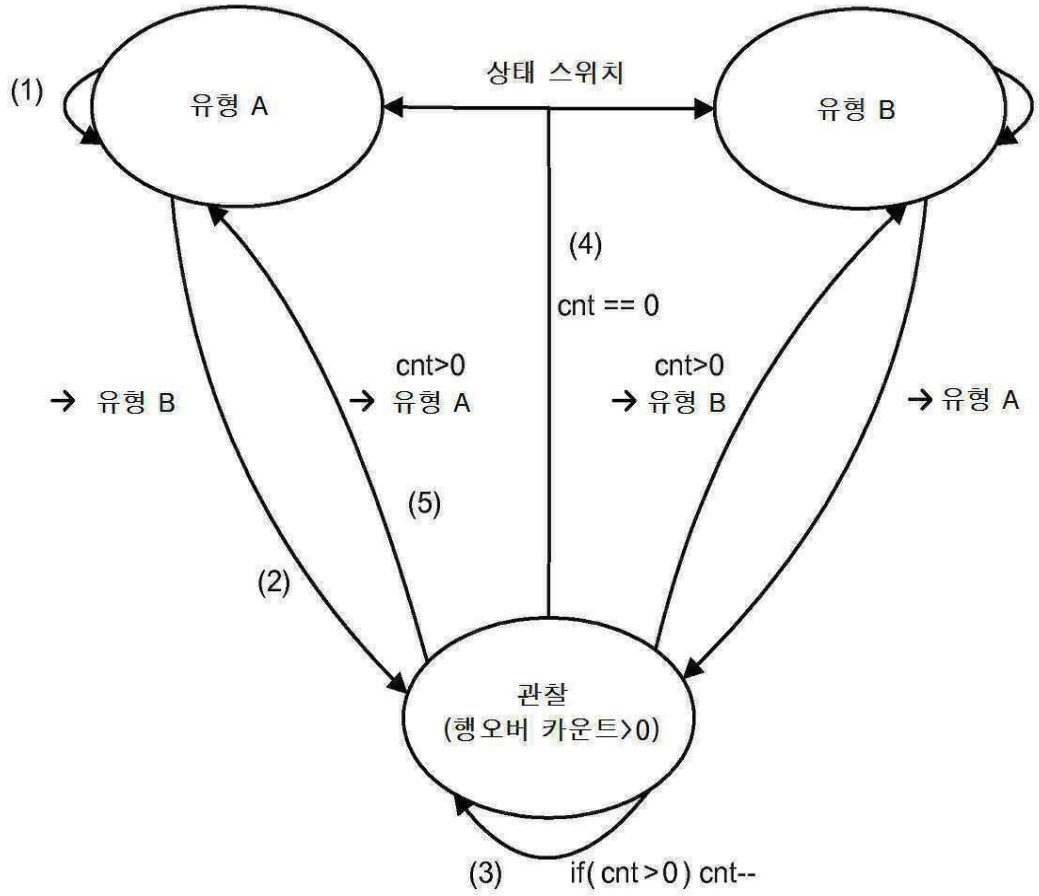
도면8


100
오디오 프로세싱 장치
200
오디오 분류기
오디오 콘텐트 분류기
202
오디오 콘텍스트 분류기
204
814
300
조정 유닛
파라미터 평활화 유닛
다이얼로그 강화기
서라운드 버추얼라이저
볼륨 레벨러
등화기
402
404
406
408

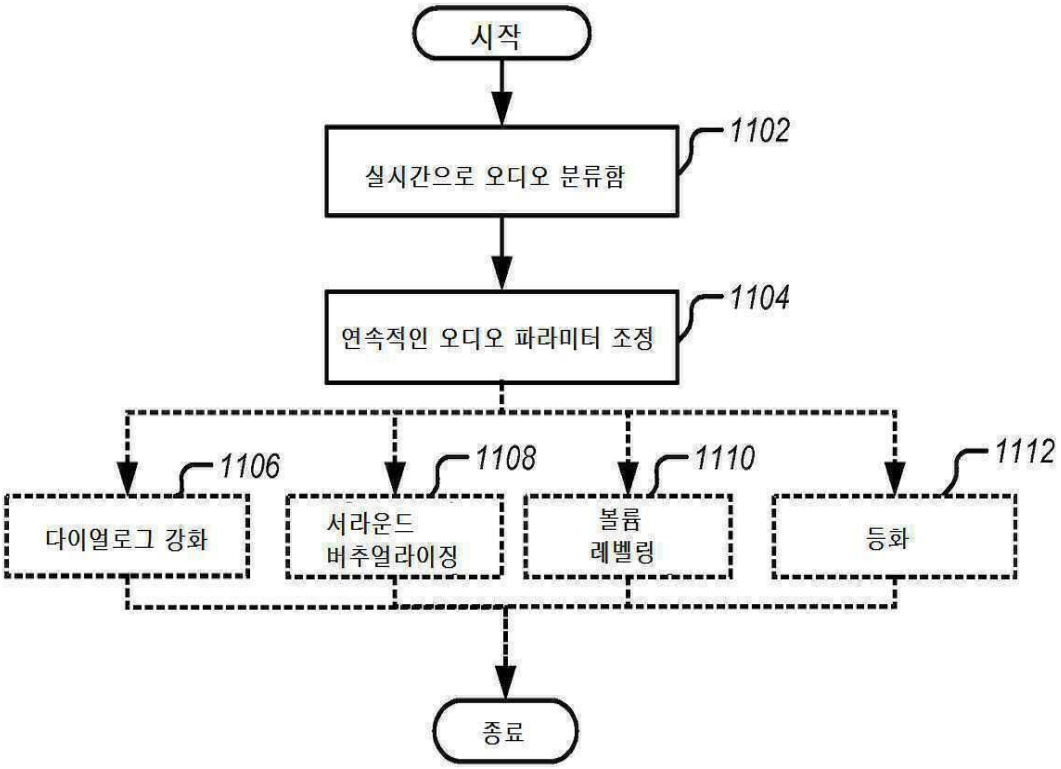
도면9


100
오디오 프로세싱 장치
200
오디오 분류기
오디오 콘텐트 분류기
202
오디오 콘텍스트 분류기
204
타이머
조정 유닛
300
916
다이얼로그 강화기
서라운드 버추얼라이저
볼륨 레벨러
등화기
402
404
406
408

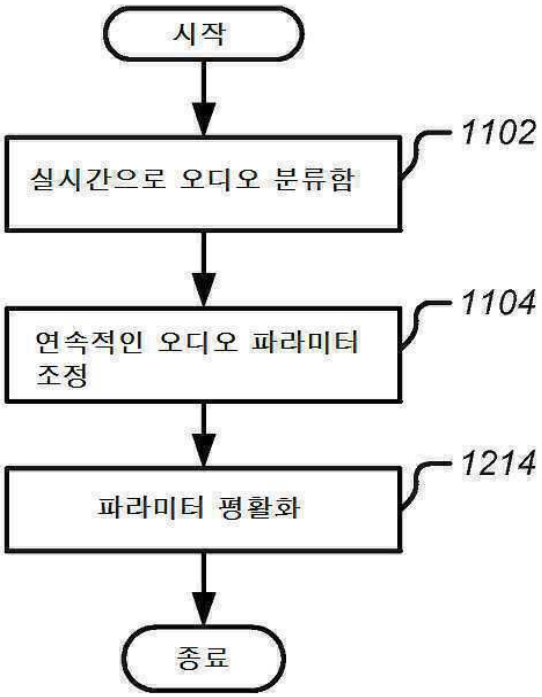
도면10


(1)
유형 A
상태 스위치
유형 B
(4)
cnt == 0
cnt>0
→ 유형 B
cnt>0
→ 유형 A
cnt>0
→ 유형 B
→ 유형 A
(5)
(2)
관찰
(행오버 카운트>0)
(3)
if(cnt > 0) cnt--

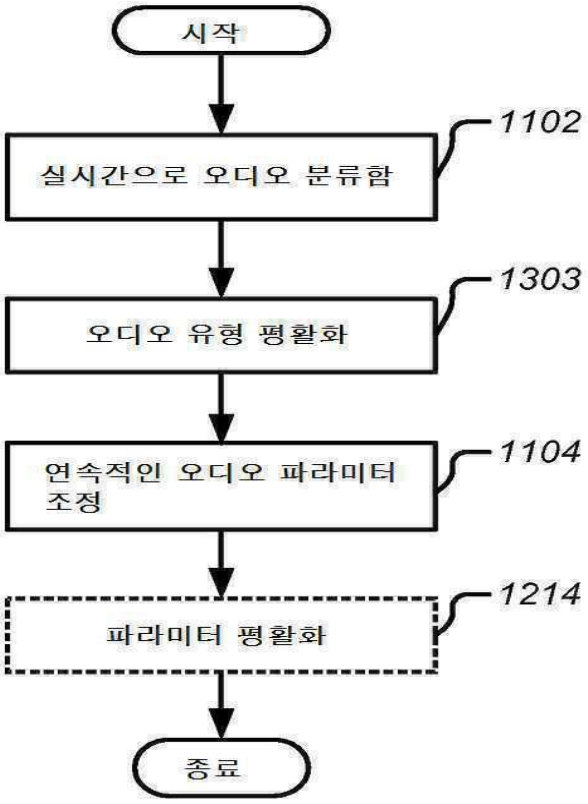
도면11


시작
실시간으로 오디오 분류함
1102
연속적인 오디오 파라미터 조정
1104
다이얼로그 강화
1106
서라운드 버추얼라이징
1108
볼륨 레벨링
1110
등화
1112
종료

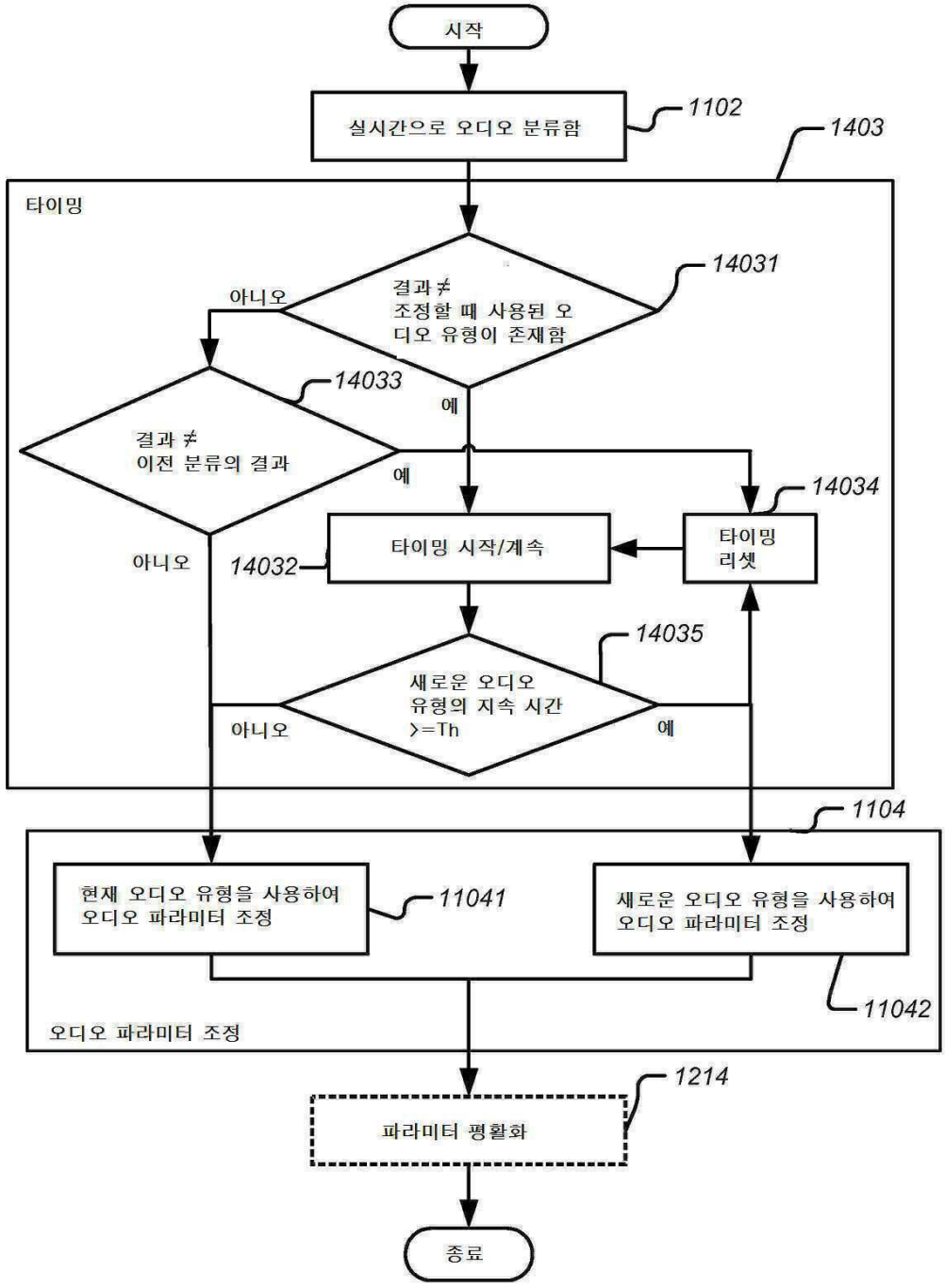
도면12


시작
실시간으로 오디오 분류함
1102
연속적인 오디오 파라미터 조정
1104
파라미터 평활화
1214
종료

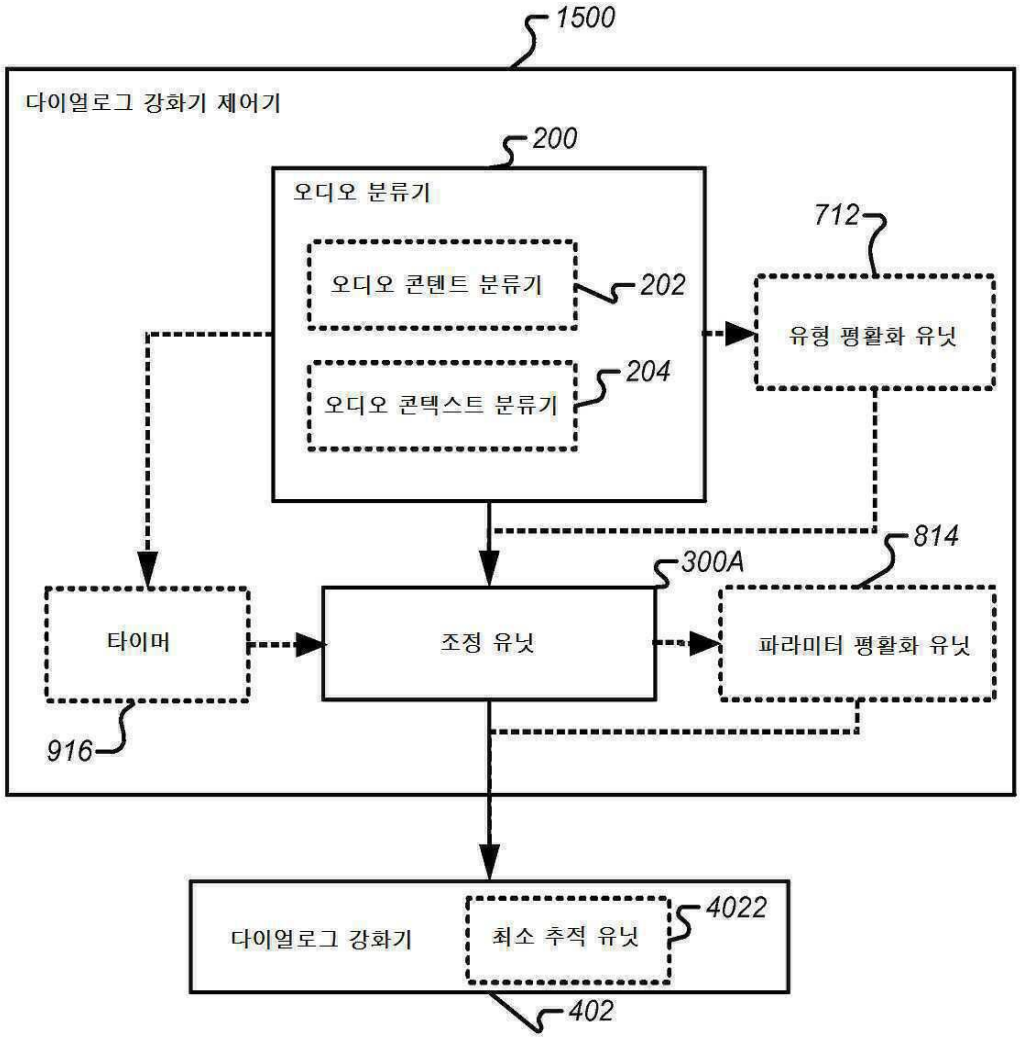
도면13


시작
1102
실시간으로 오디오 분류함
1303
오디오 유형 평활화
1104
연속적인 오디오 파라미터 조정
1214
파라미터 평활화
종료

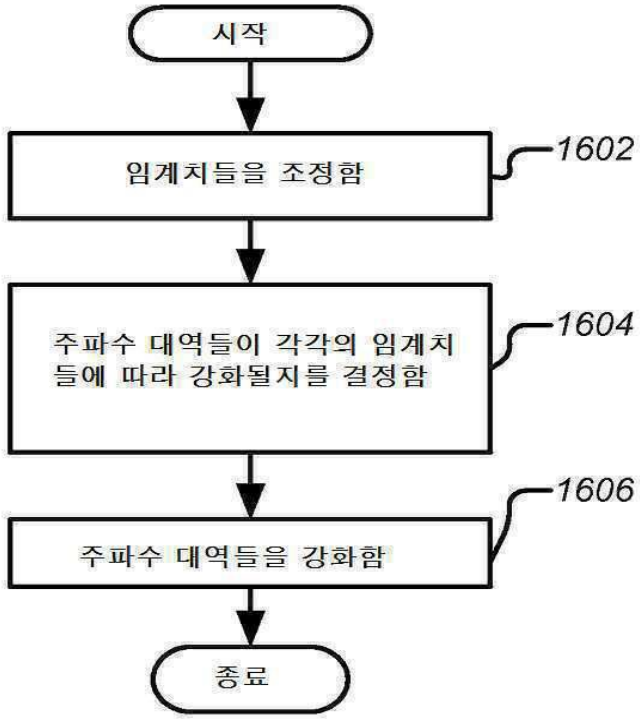
도면14


시작
실시간으로 오디오 분류함
1102
1403
타이밍
14031
결과≠
조정할 때 사용된 오
디오 유형이 존재함
아니오
14033
결과≠
이전 분류의 결과
예
예
14034
타이밍
리셋
아니오
14032
타이밍 시작/계속
14035
새로운 오디오
유형의 지속 시간
>=Th
아니오
예
1104
현재 오디오 유형을 사용하여
오디오 파라미터 조정
11041
새로운 오디오 유형을 사용하여
오디오 파라미터 조정
11042
오디오 파라미터 조정
1214
파라미터 평활화
종료

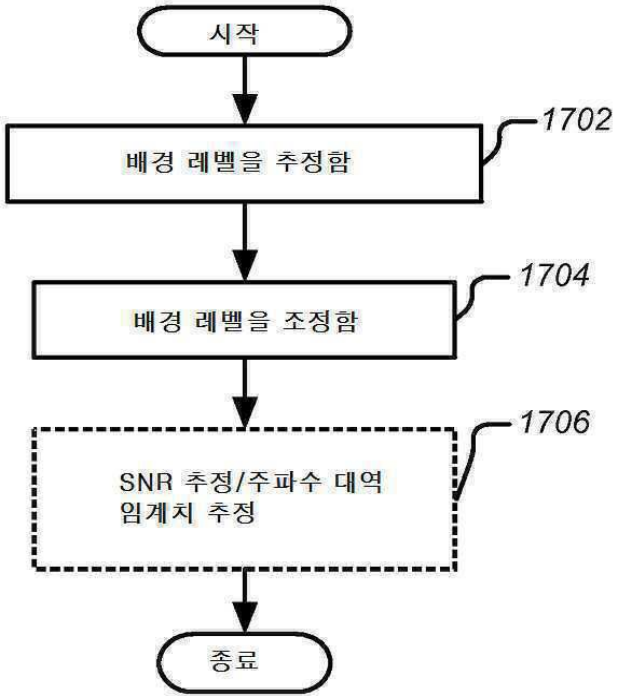
도면15


1500
다이얼로그 강화기 제어기
200
오디오 분류기
오디오 콘텐트 분류기
202
오디오 콘텍스트 분류기
204
712
유형 평활화 유닛
814
300A
타이머
조정 유닛
파라미터 평활화 유닛
916
다이얼로그 강화기
최소 추적 유닛
4022
402

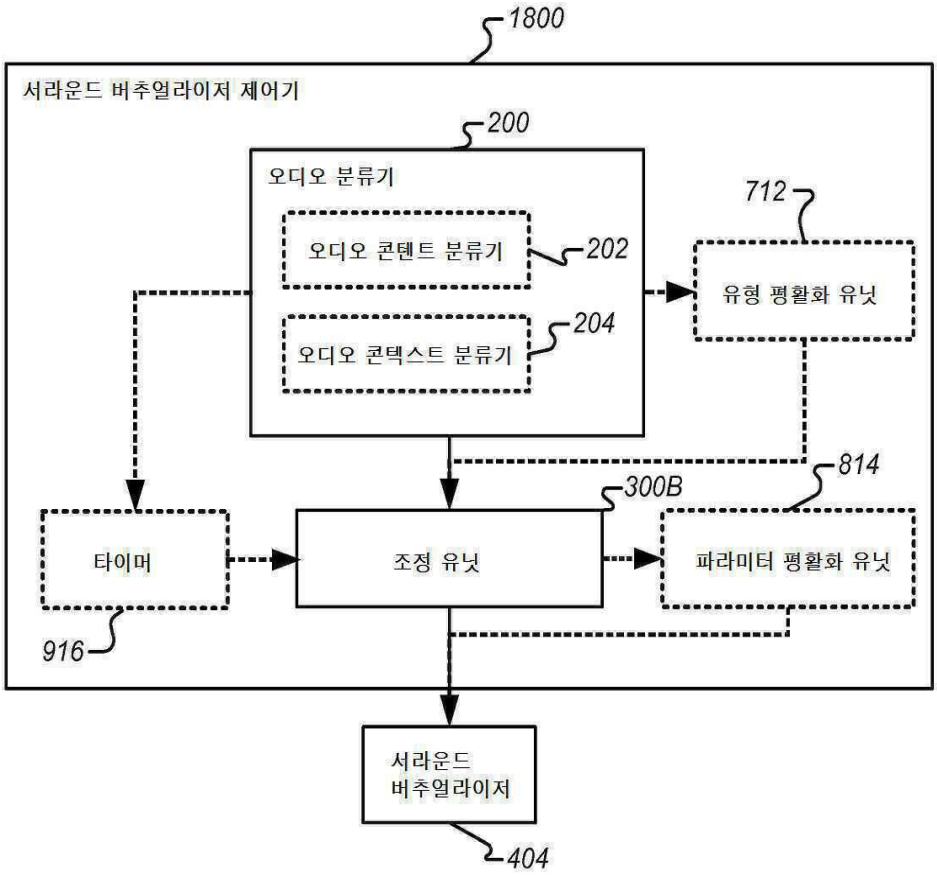
도면16


시작
임계치들을 조정함
1602
주파수 대역들이 각각의 임계치들에 따라 강화될지를 결정함
1604
주파수 대역들을 강화함
1606
종료

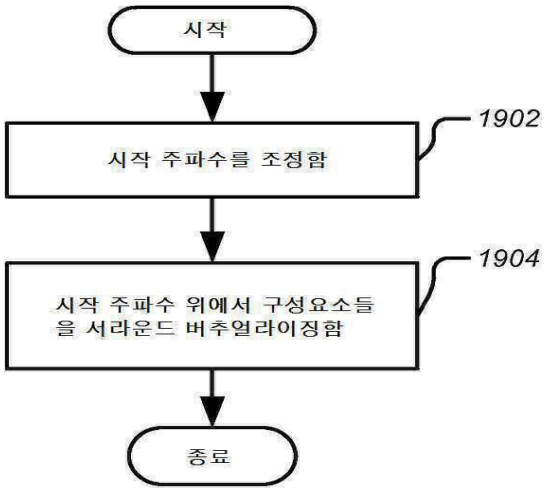
도면17


시작
배경 레벨을 추정함
1702
배경 레벨을 조정함
1704
SNR 추정/주파수 대역 임계치 추정
1706
종료

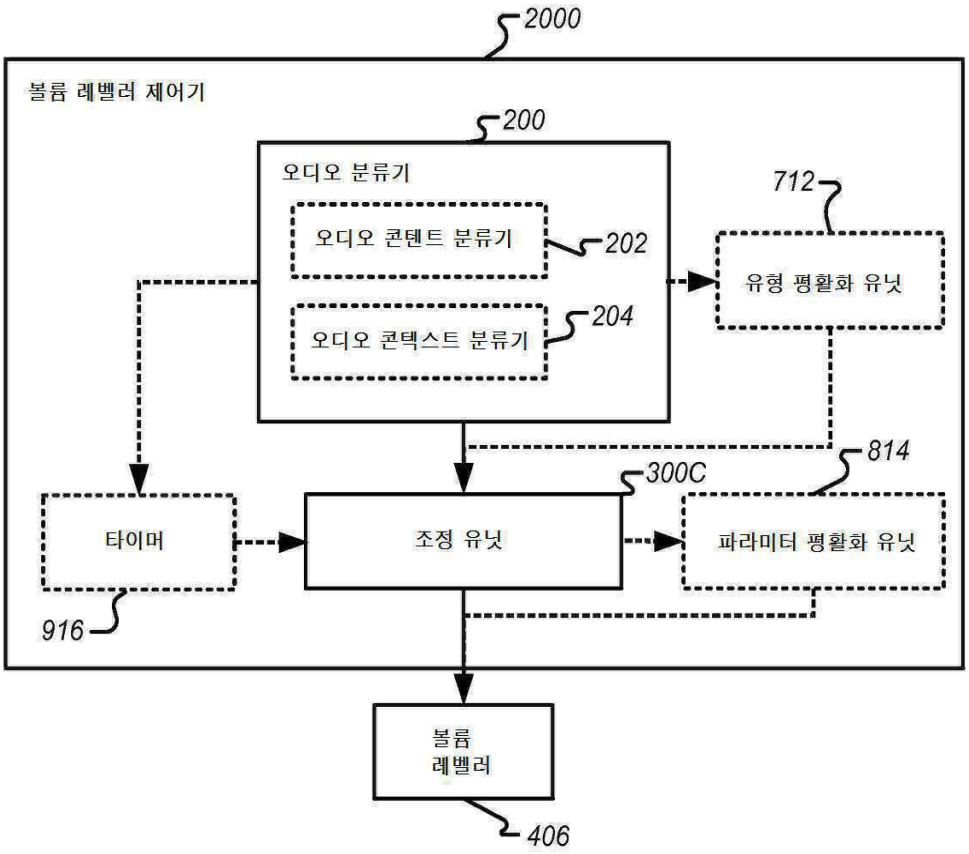
도면18


1800
서라운드 버추얼라이저 제어기
200
오디오 분류기
오디오 콘텐트 분류기
202
712
유형 평활화 유닛
오디오 콘텍스트 분류기
204
814
300B
타이머
조정 유닛
파라미터 평활화 유닛
916
서라운드
버추얼라이저
404

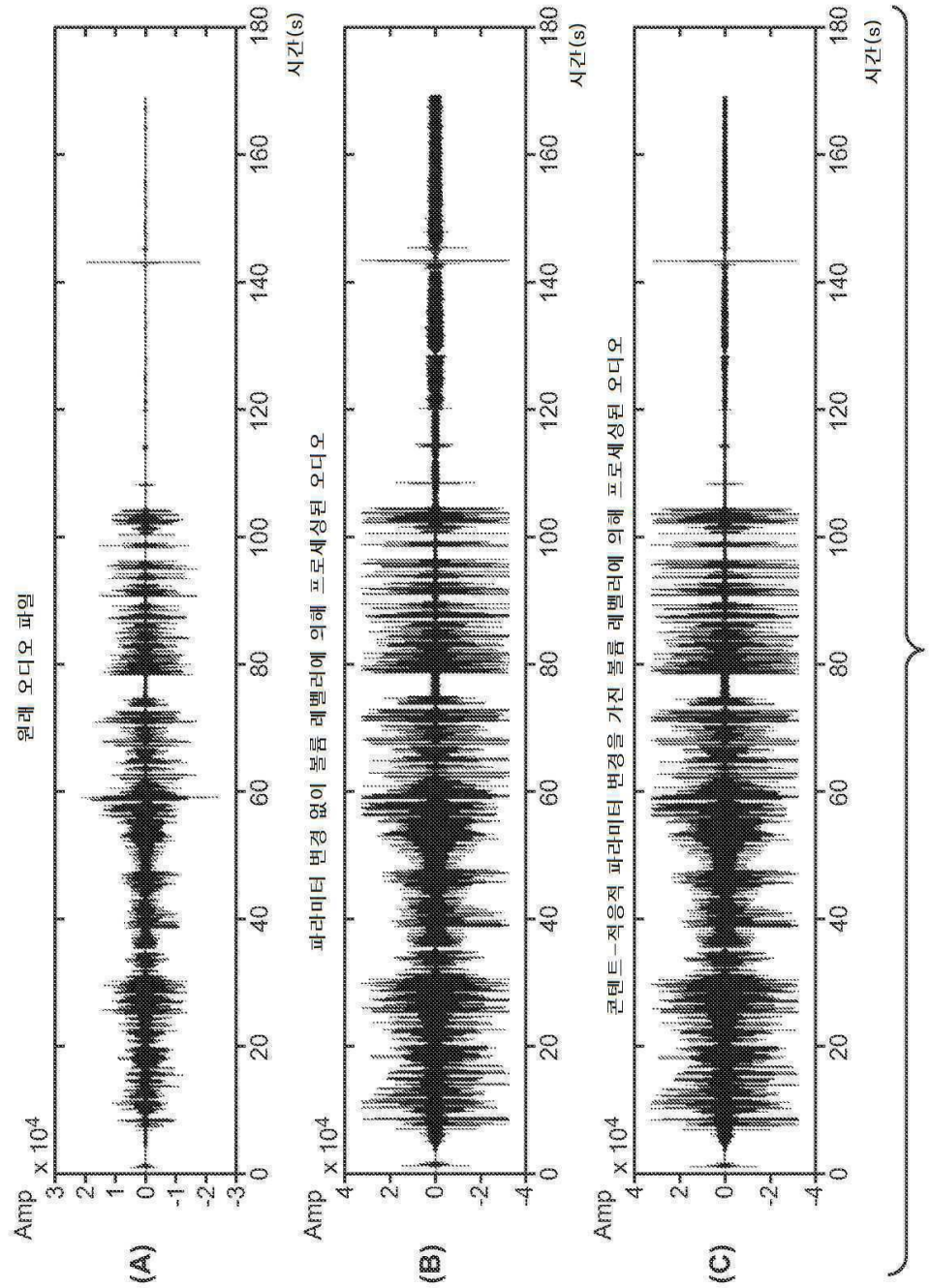
도면19


시작
1902
시작 주파수를 조정함
1904
시작 주파수 위에서 구성요소들
을 서라운드 버추얼라이징함
종료

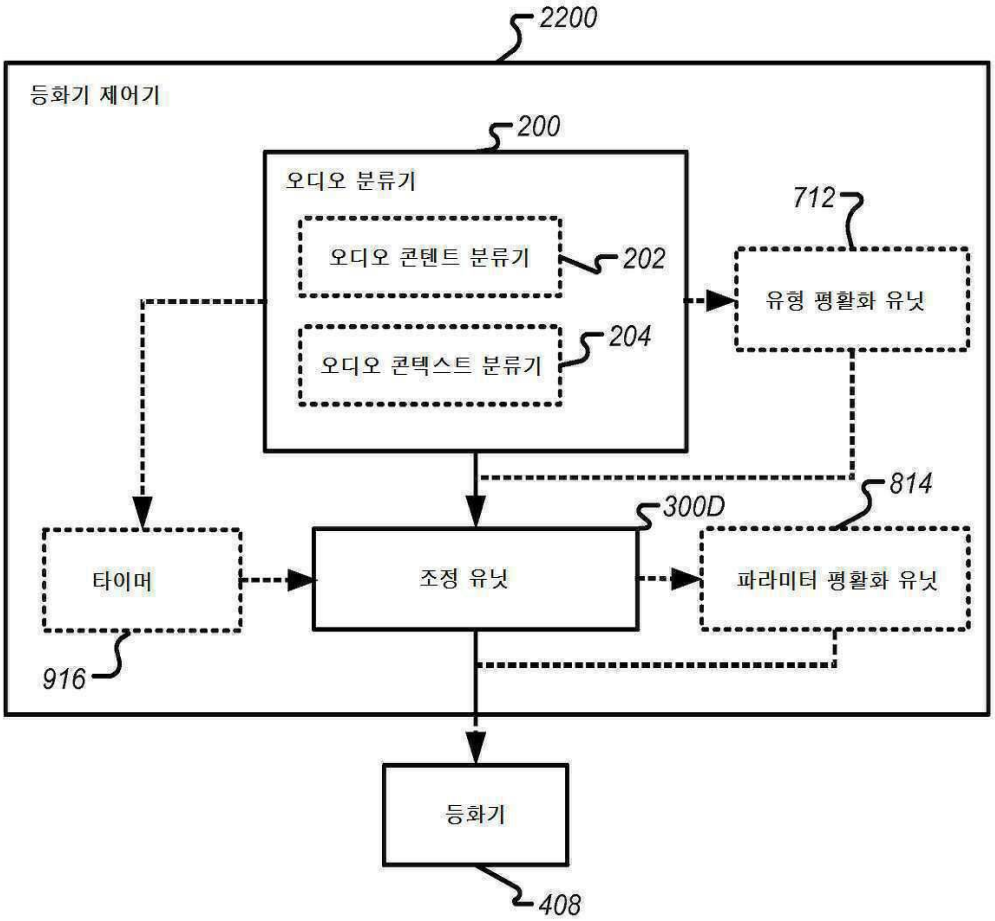
도면20


2000
볼륨 레벨러 제어기
200
오디오 분류기
오디오 콘텐트 분류기
202
오디오 콘텍스트 분류기
204
712
유형 평활화 유닛
814
300C
타이머
조정 유닛
파라미터 평활화 유닛
916
볼륨
레벨러
406

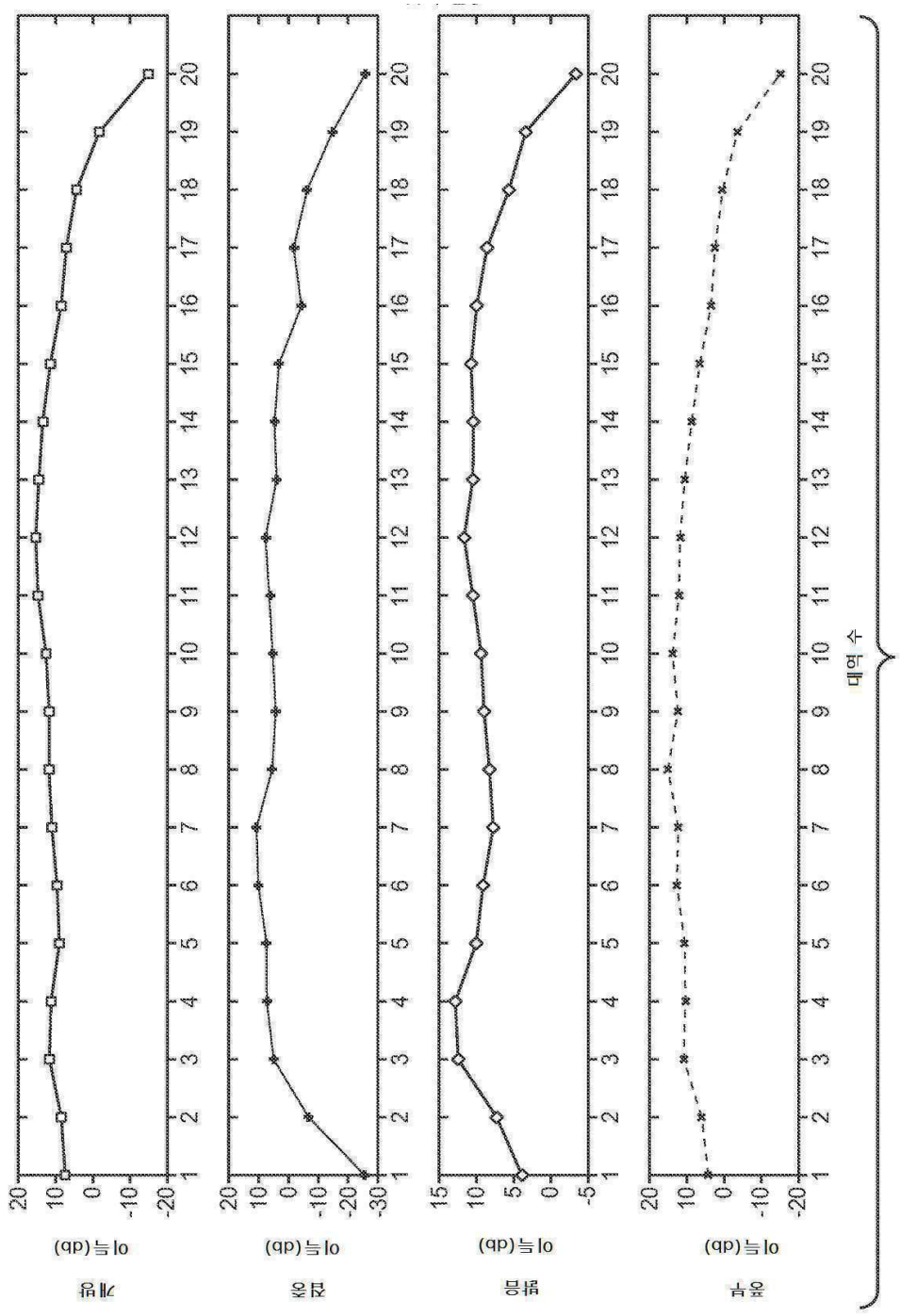
도면21


원래 오디오 파일
Amp
x 10^4
(A)
시간(s)
파라미터 변경 없이 볼륨 레벨러에 의해 프로세싱된 오디오
Amp
x 10^4
(B)
시간(s)
콘텐트-적응적 파라미터 변경을 가진 볼륨 레벨러에 의해 프로세싱된 오디오
Amp
x 10^4
(C)
시간(s)

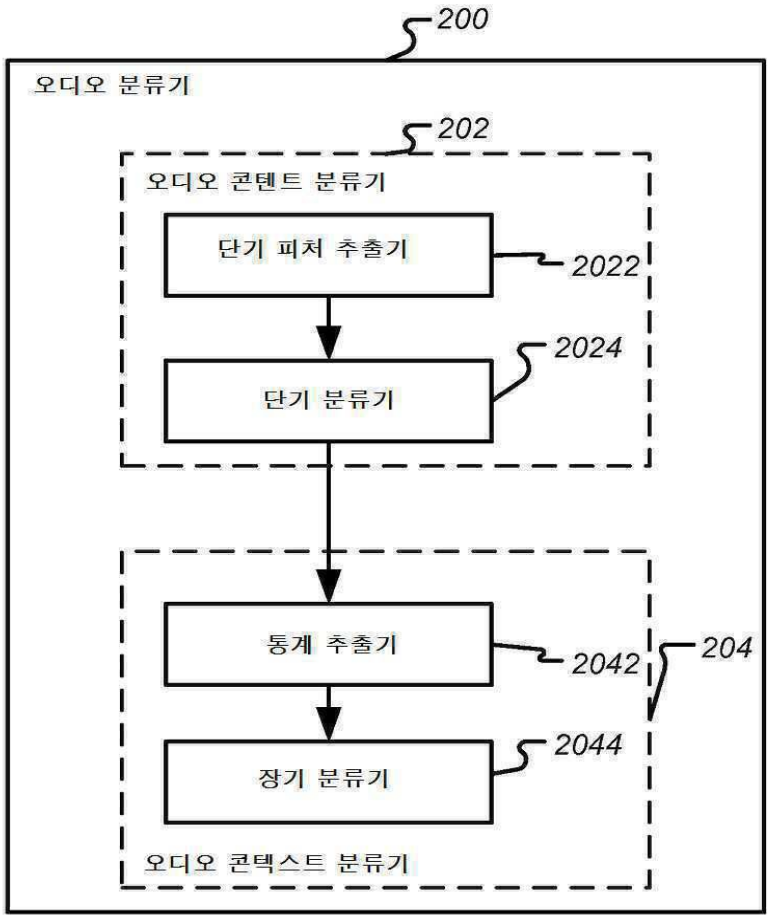
도면22


2200
등화기 제어기
200
오디오 분류기
오디오 콘텐트 분류기
202
오디오 콘텍스트 분류기
204
712
유형 평활화 유닛
814
300D
타이머
조정 유닛
파라미터 평활화 유닛
916
등화기
408

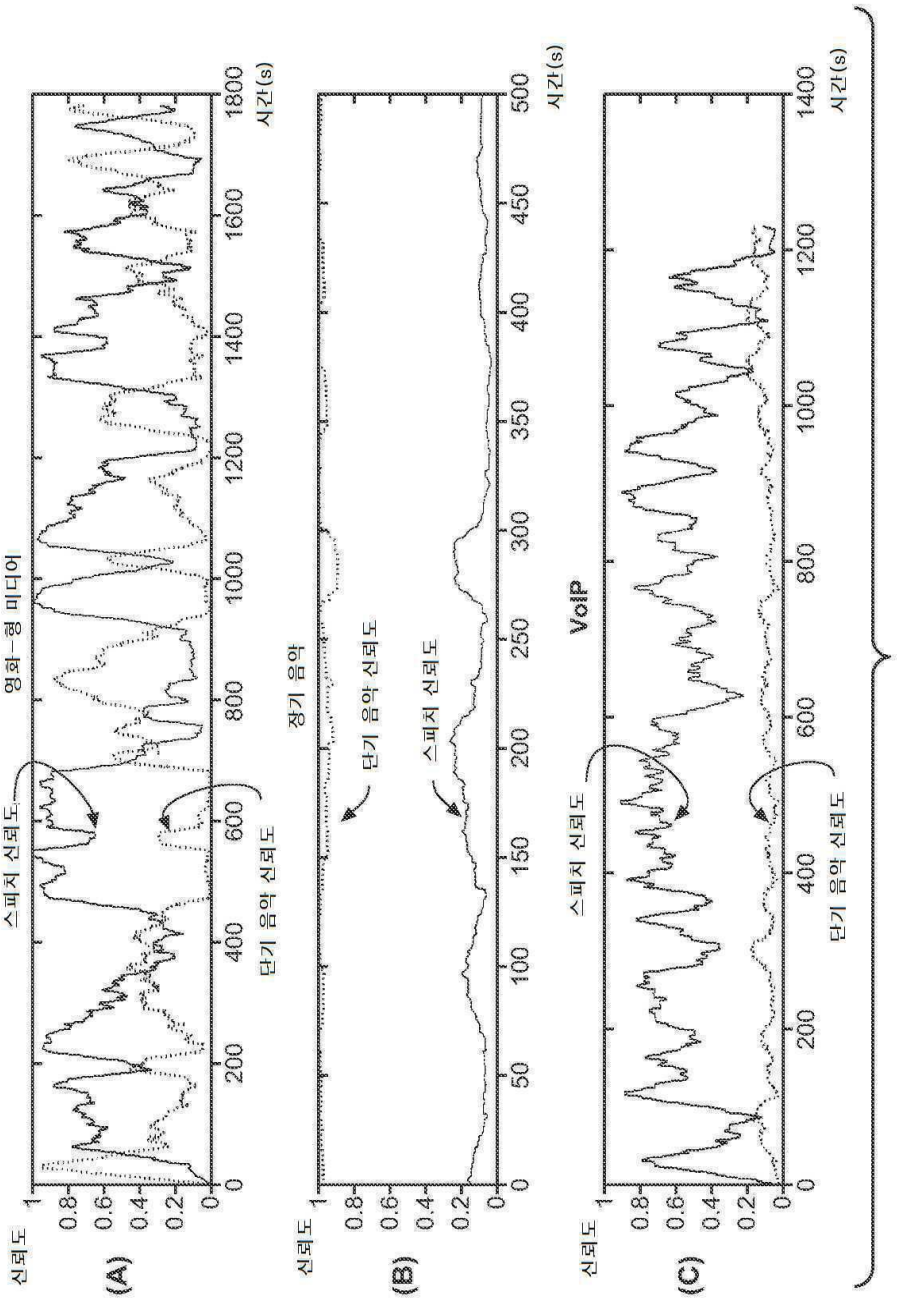
도면23



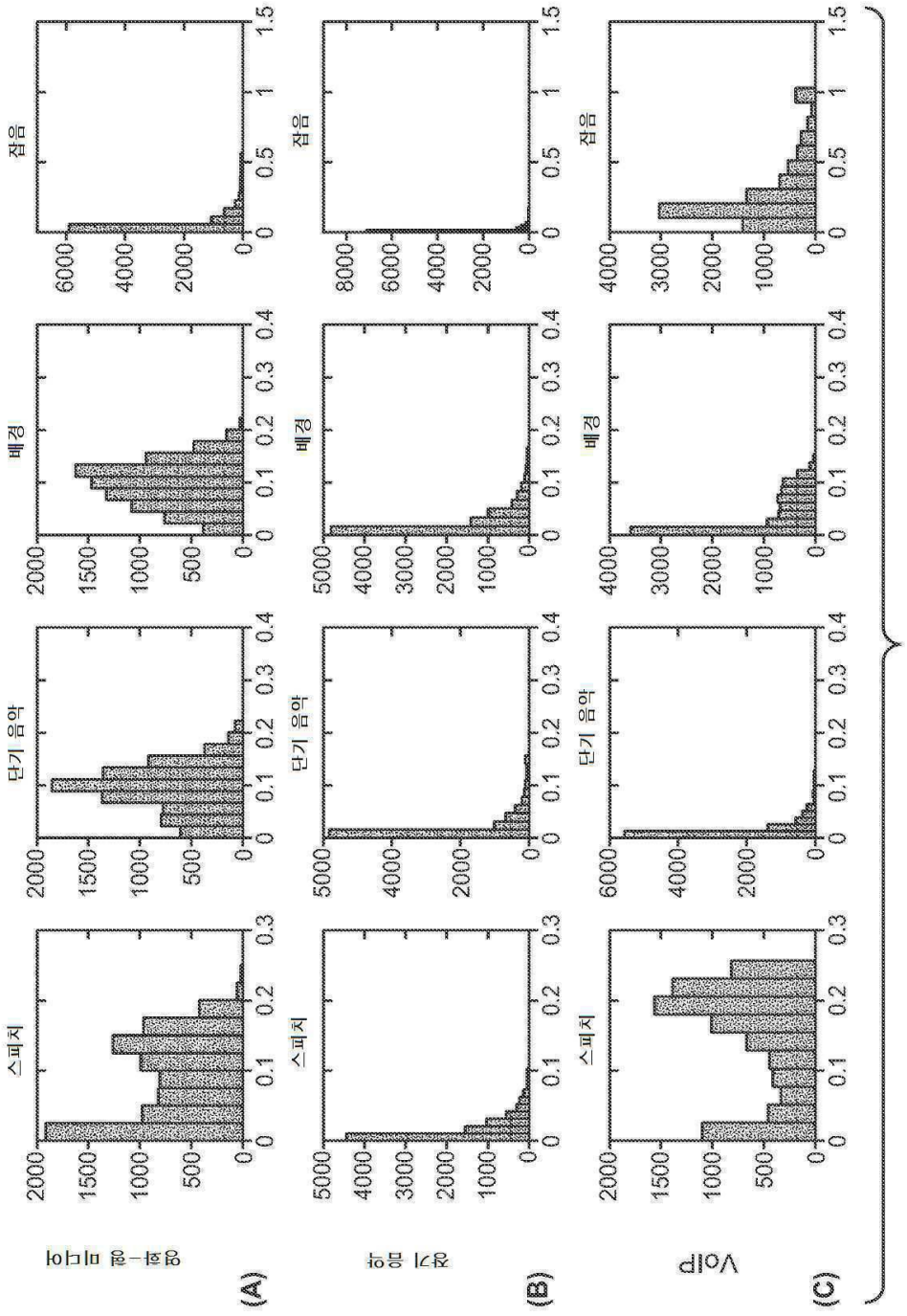
도면24


200
오디오 분류기
202
오디오 콘텐트 분류기
단기 피처 추출기
2022
2024
단기 분류기
통계 추출기
2042
204
2044
장기 분류기
오디오 콘텍스트 분류기

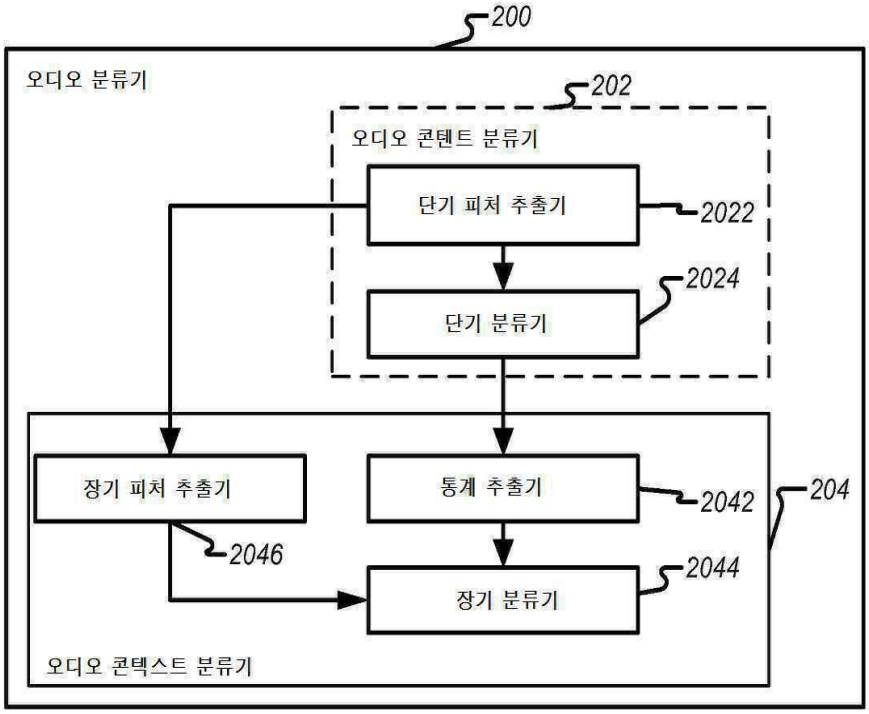
도면25


(A)
영화-형 미디어
스피치 신뢰도
단기 음악 신뢰도
신뢰도
시간(s)
(B)
장기 음악
단기 음악 신뢰도
스피치 신뢰도
신뢰도
시간(s)
(C)
VoIP
스피치 신뢰도
단기 음악 신뢰도
신뢰도
시간(s)

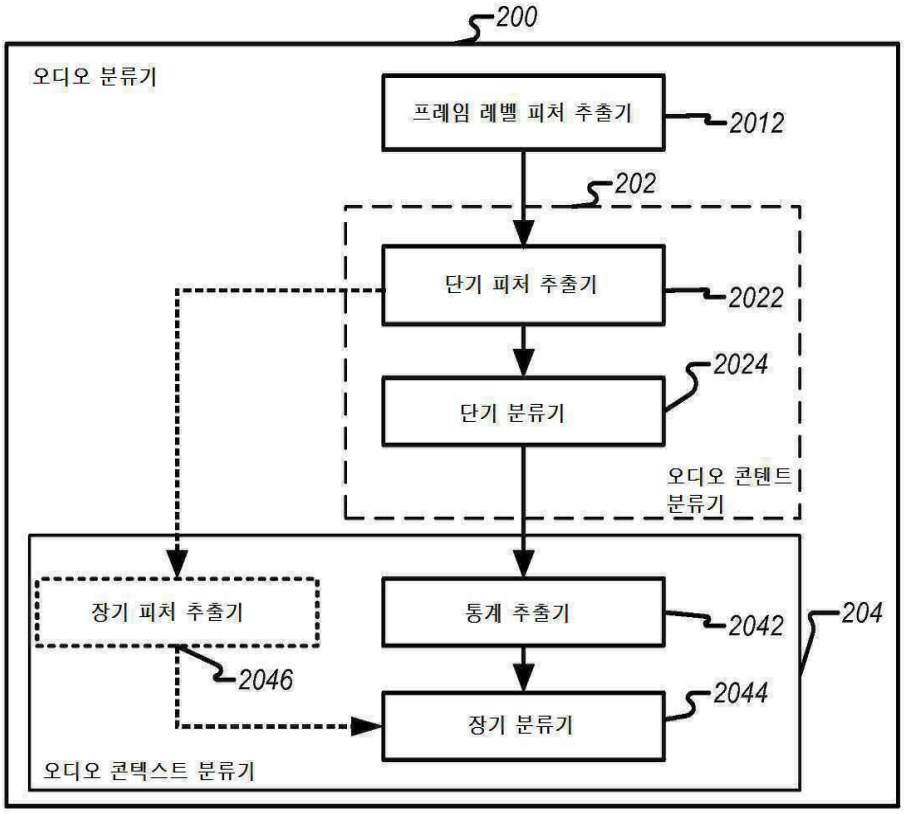
도면26


(A)
영화-영 미디어
스피치
단기 음악
배경
잡음
(B)
장기 음악
스피치
단기 음악
배경
잡음
(C)
VoIP
스피치
단기 음악
배경
잡음

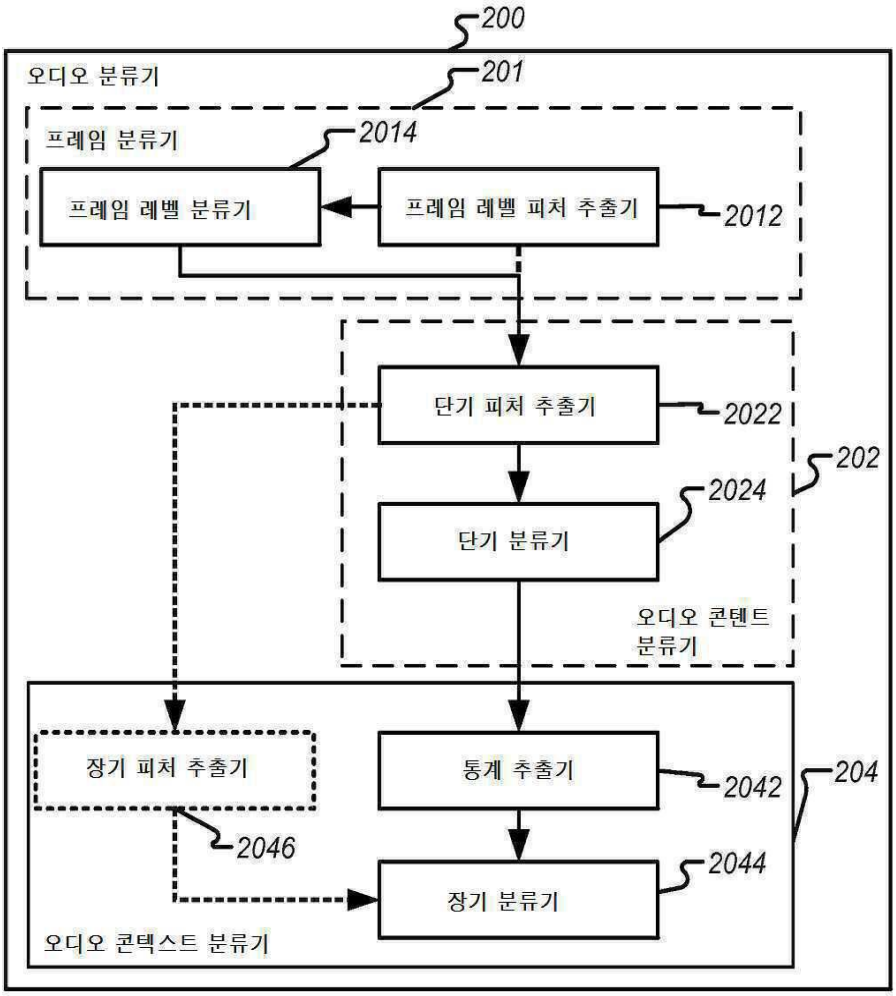
도면27


200
오디오 분류기
202
오디오 콘텐트 분류기
단기 피처 추출기
2022
2024
단기 분류기
장기 피처 추출기
통계 추출기
2042
204
2046
2044
장기 분류기
오디오 콘텍스트 분류기

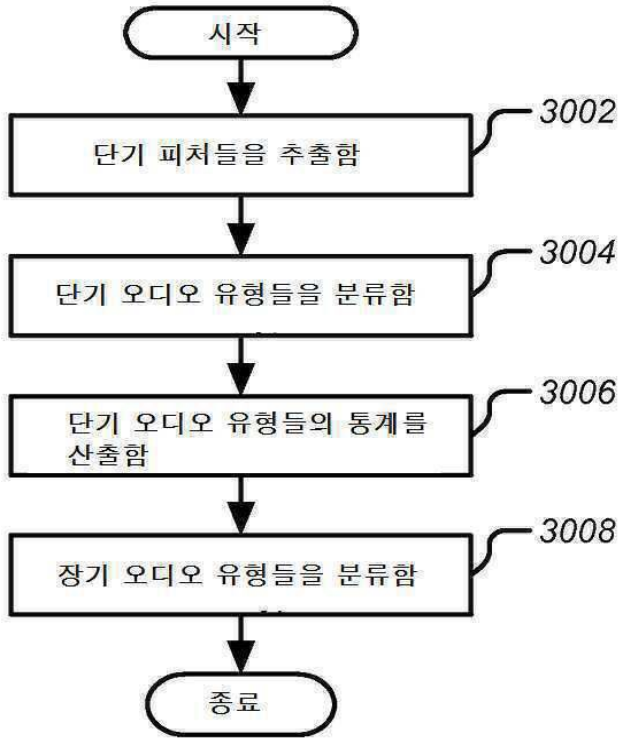
도면28


200
오디오 분류기
프레임 레벨 피처 추출기
2012
202
단기 피처 추출기
2022
2024
단기 분류기
오디오 콘텐트 분류기
장기 피처 추출기
통계 추출기
2042
204
2046
2044
장기 분류기
오디오 콘텍스트 분류기

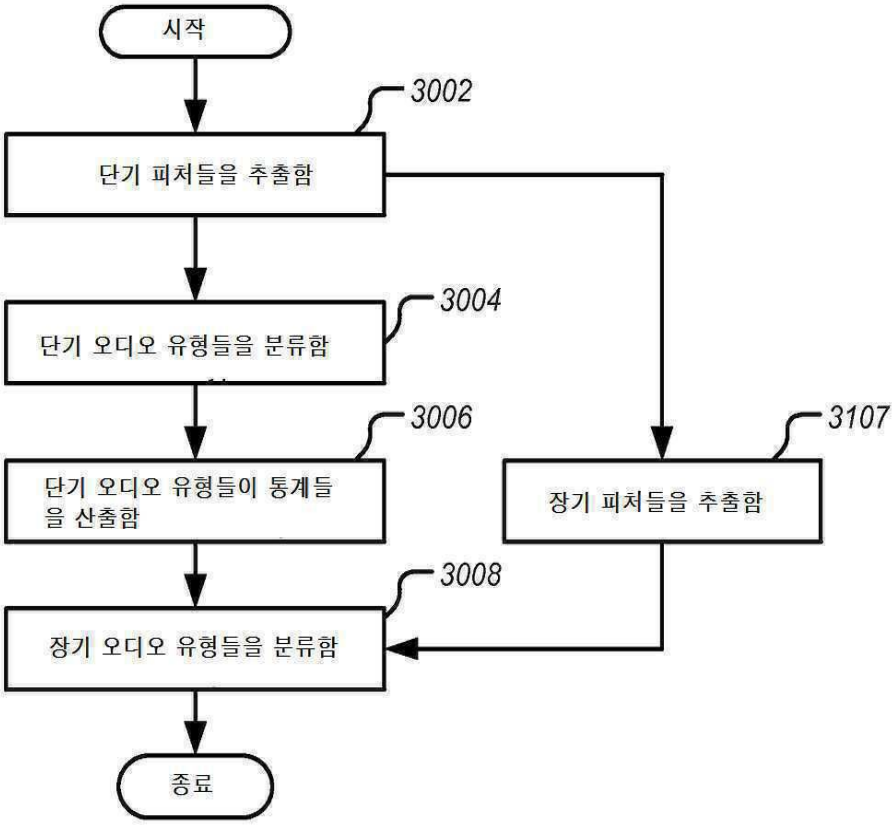
도면29


200
오디오 분류기
201
프레임 분류기
2014
프레임 레벨 분류기
프레임 레벨 피처 추출기
2012
단기 피처 추출기
2022
202
2024
단기 분류기
오디오 콘텐트 분류기
장기 피처 추출기
통계 추출기
2042
204
2046
2044
장기 분류기
오디오 콘텍스트 분류기

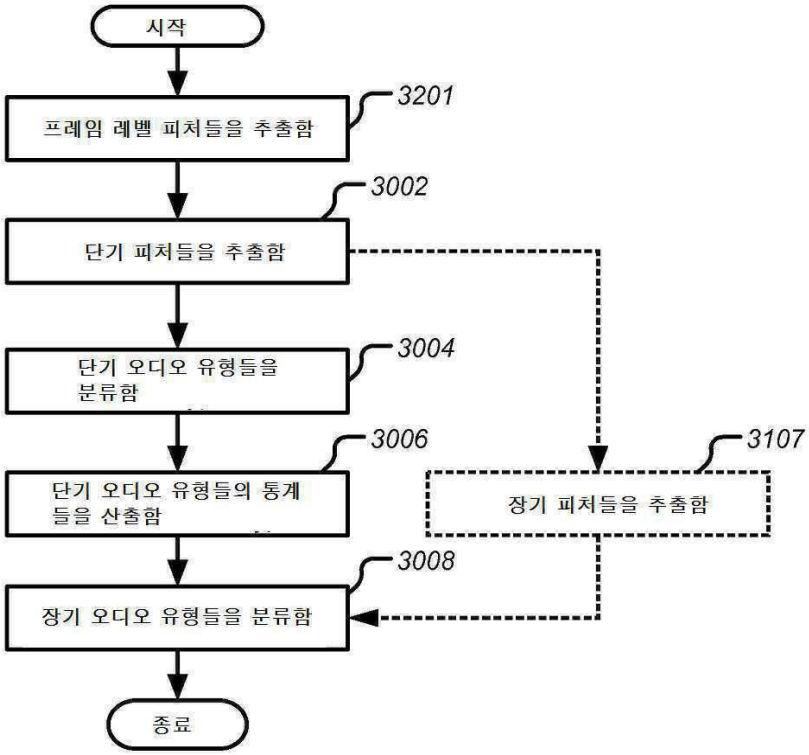
도면30


시작
단기 피처들을 추출함
3002
단기 오디오 유형들을 분류함
3004
단기 오디오 유형들의 통계를 산출함
3006
장기 오디오 유형들을 분류함
3008
종료

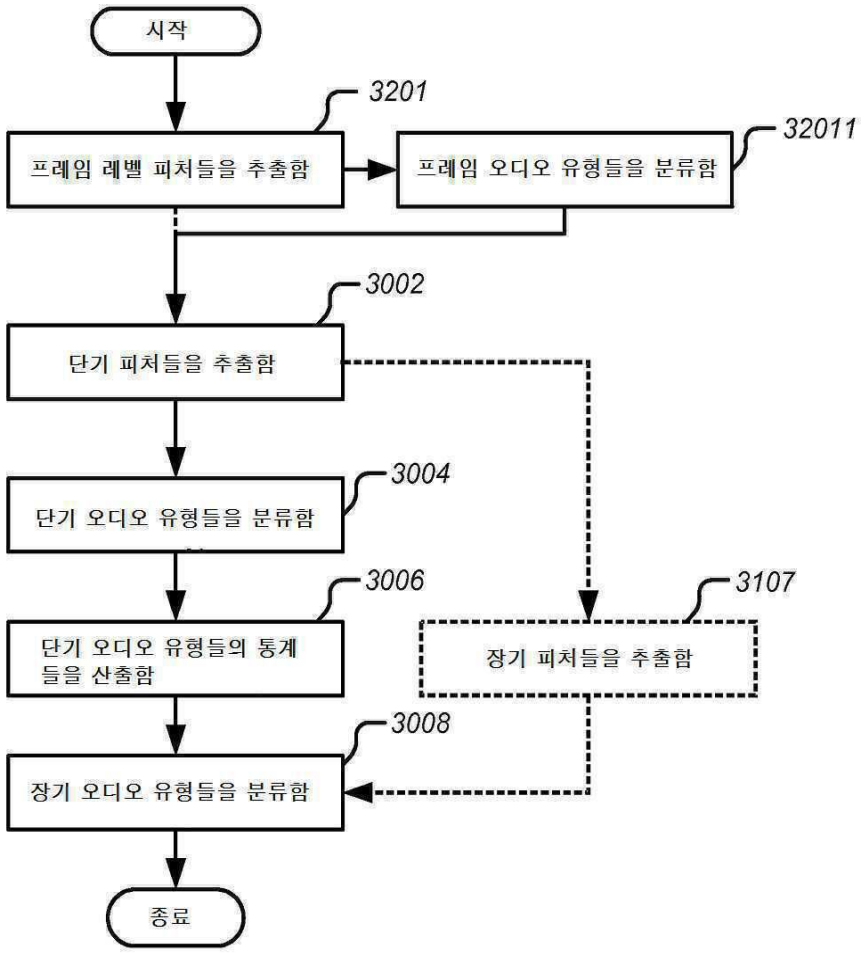
도면31


시작
단기 피처들을 추출함
3002
단기 오디오 유형들을 분류함
3004
단기 오디오 유형들이 통계들을 산출함
3006
장기 피처들을 추출함
3107
장기 오디오 유형들을 분류함
3008
종료

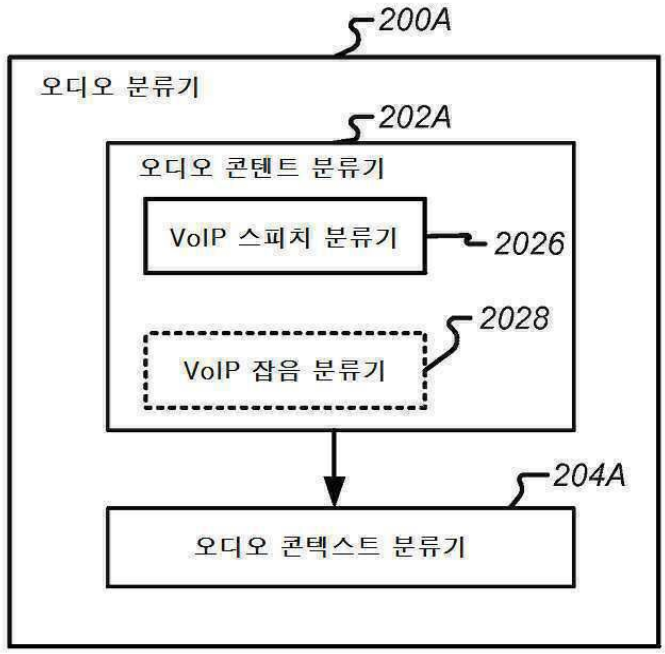
도면32


시작
프레임 레벨 피처들을 추출함
3201
단기 피처들을 추출함
3002
단기 오디오 유형들을 분류함
3004
단기 오디오 유형들의 통계들을 산출함
3006
장기 피처들을 추출함
3107
장기 오디오 유형들을 분류함
3008
종료

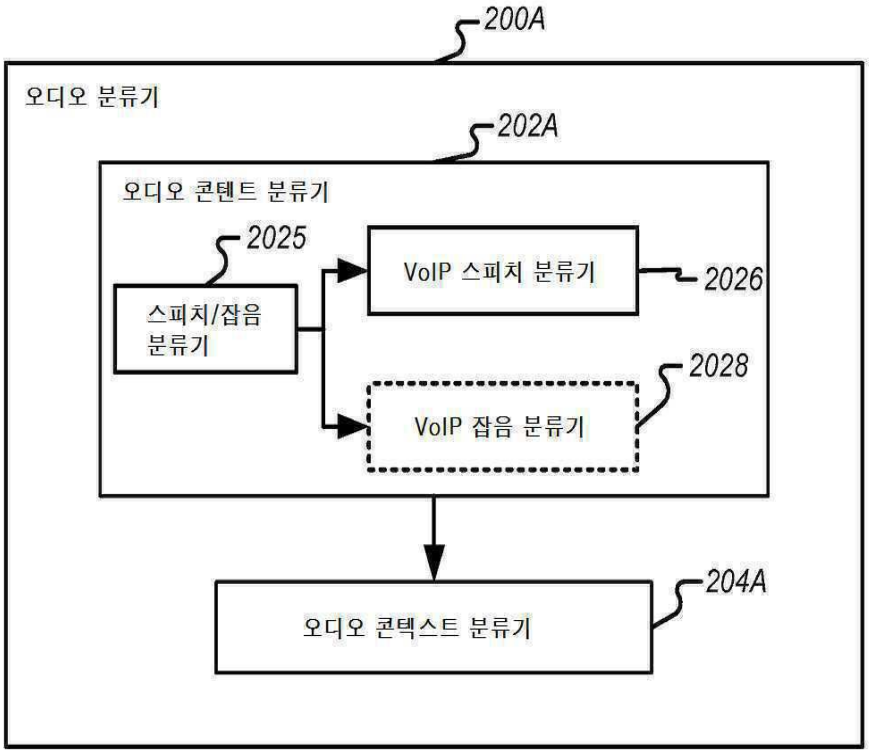
도면33


시작
3201
프레임 레벨 피처들을 추출함
32011
프레임 오디오 유형들을 분류함
3002
단기 피처들을 추출함
3004
단기 오디오 유형들을 분류함
3006
단기 오디오 유형들의 통계들을 산출함
3107
장기 피처들을 추출함
3008
장기 오디오 유형들을 분류함
종료

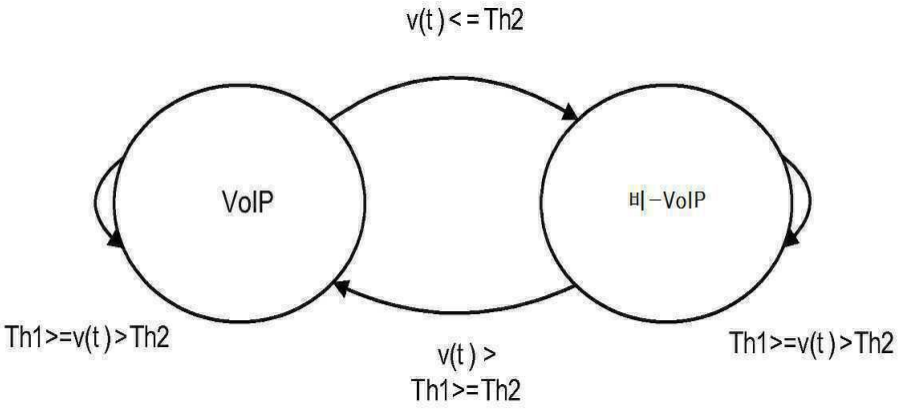
도면34


200A
오디오 분류기
202A
오디오 콘텐트 분류기
VoIP 스피치 분류기
2026
2028
VoIP 잡음 분류기
204A
오디오 콘텍스트 분류기

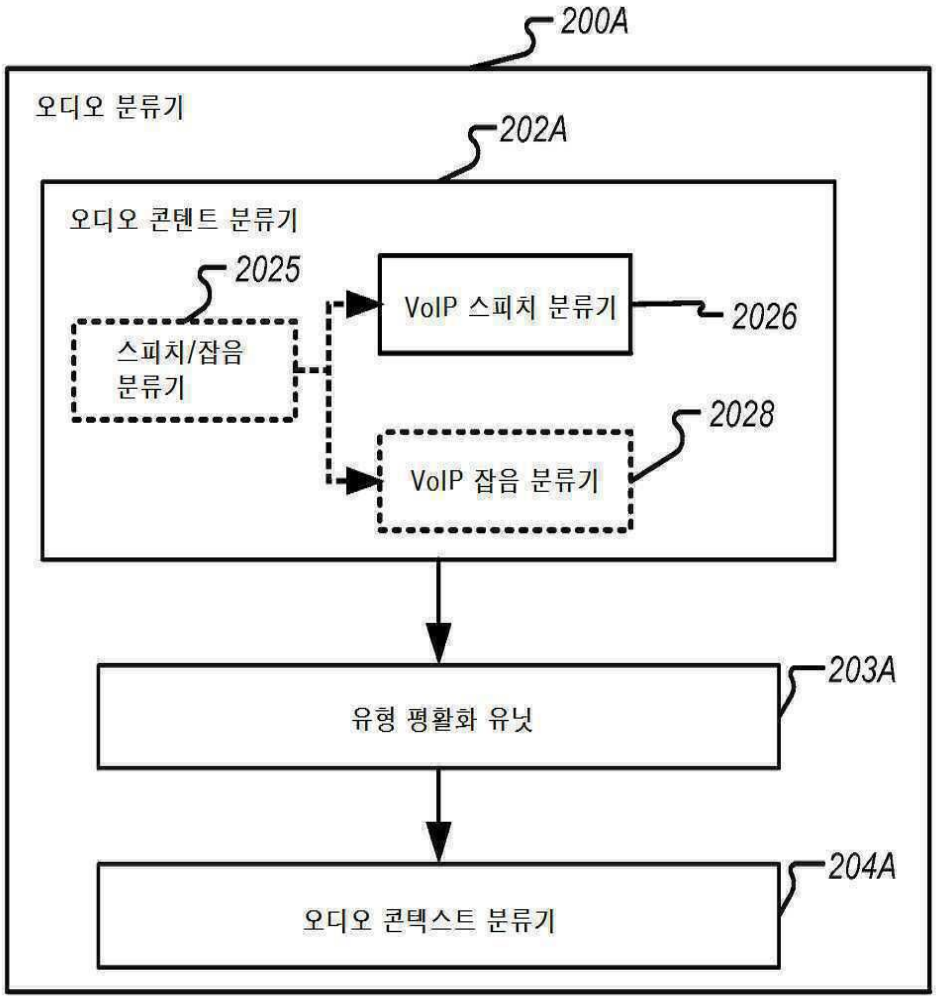
도면35


200A
오디오 분류기
202A
오디오 콘텐트 분류기
2025
스피치/잡음 분류기
VoIP 스피치 분류기
2026
2028
VoIP 잡음 분류기
204A
오디오 콘텍스트 분류기

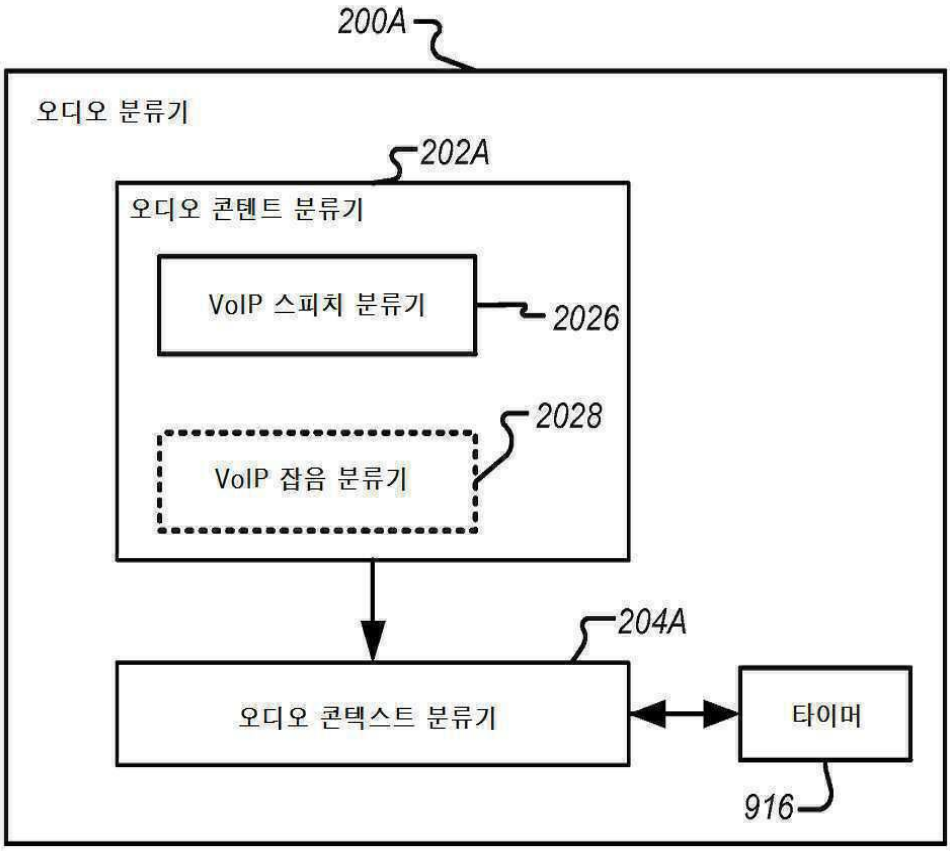
도면36


v(t)<=Th2
VoIP
비-VoIP
Th1>=v(t)>Th2
v(t)>
Th1>=Th2
Th1>=v(t)>Th2

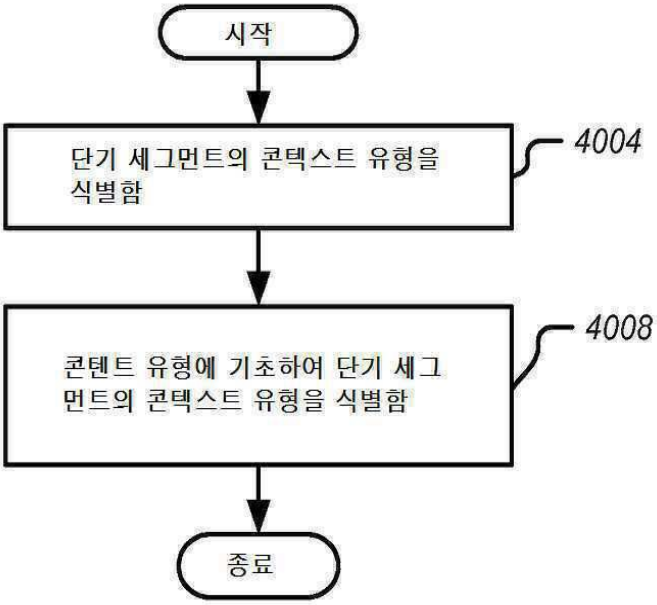
도면37


200A
오디오 분류기
202A
오디오 콘텐트 분류기
2025
스피치/잡음
분류기
VoIP 스피치 분류기
2026
2028
VoIP 잡음 분류기
203A
유형 평활화 유닛
204A
오디오 콘텍스트 분류기

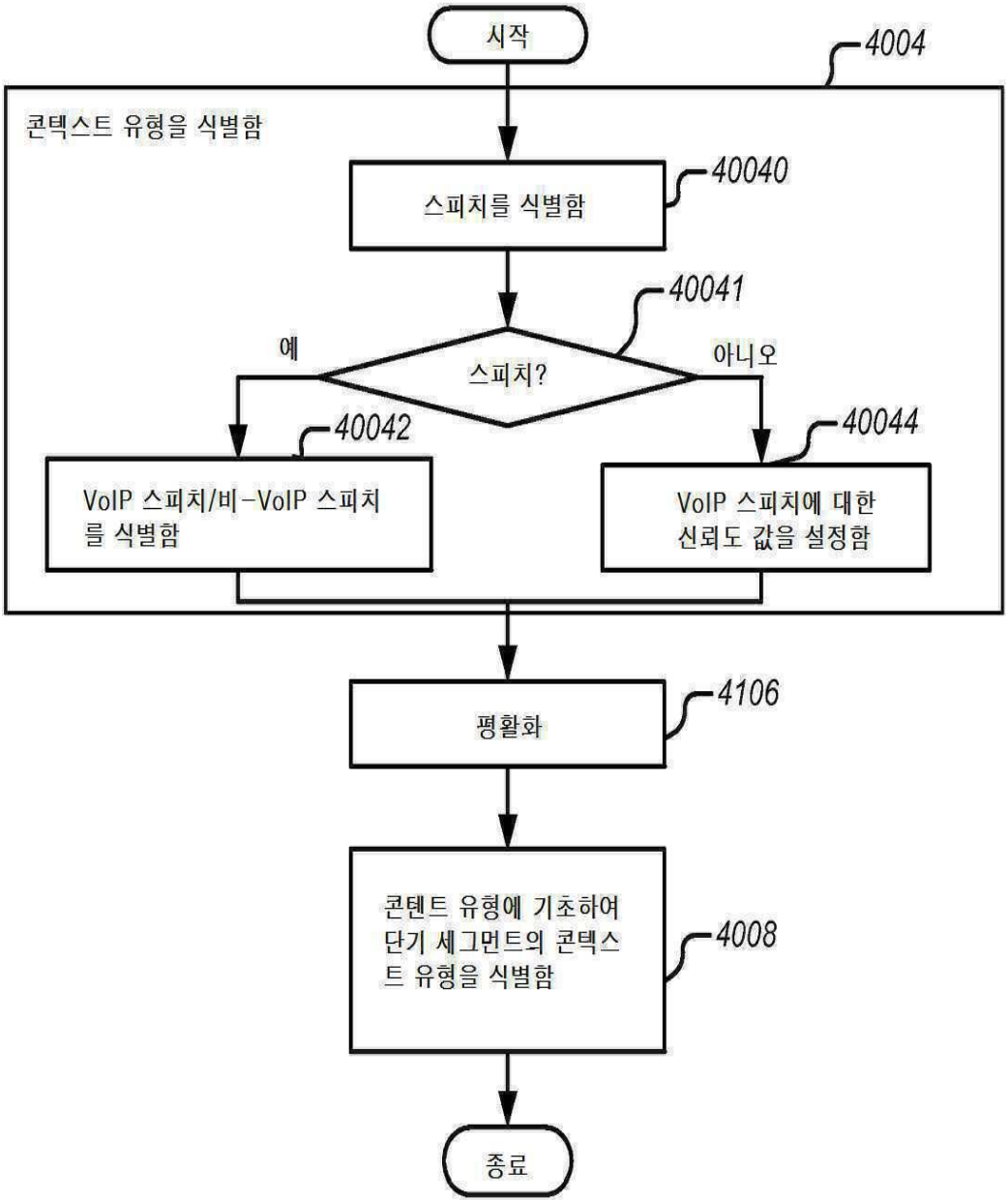
도면38


200A
오디오 분류기
202A
오디오 콘텐트 분류기
VoIP 스피치 분류기
2026
2028
VoIP 잡음 분류기
204A
오디오 콘텍스트 분류기
타이머
916

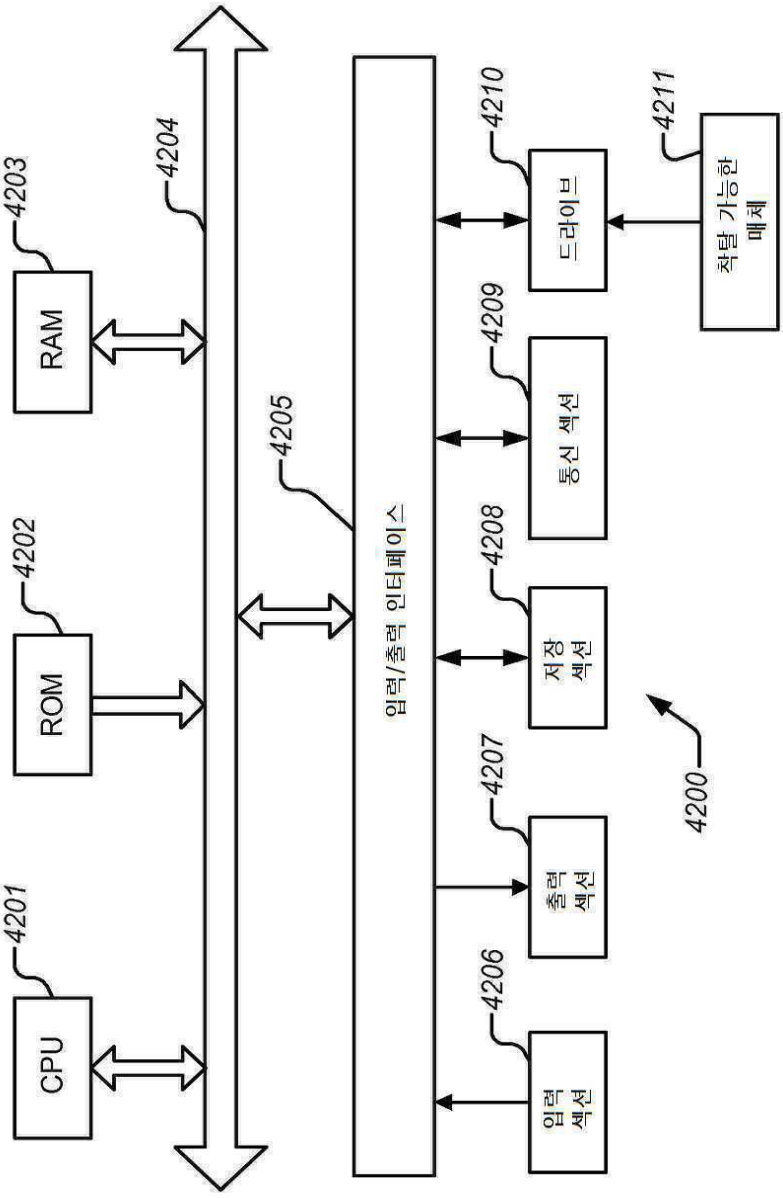
도면39


시작
단기 세그먼트의 콘텍스트 유형을 식별함
4004
콘텐트 유형에 기초하여 단기 세그먼트의 콘텍스트 유형을 식별함
4008
종료

도면40


시작
4004
콘텍스트 유형을 식별함
스피치를 식별함
40040
40041
스피치?
예
아니오
40042
40044
VoIP 스피치/비-VoIP 스피치를 식별함
VoIP 스피치에 대한 신뢰도 값을 설정함
평활화
4106
콘텐트 유형에 기초하여 단기 세그먼트의 콘텍스트 유형을 식별함
4008
종료

도면41


CPU
4201
ROM
4202
RAM
4203
4204
4205
입력/출력 인터페이스
입력 섹션
4206
출력 섹션
4207
저장 섹션
4208
통신 섹션
4209
드라이브
4210
착탈 가능한 매체
4211
4200