

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6012884号
(P6012884)

(45) 発行日 平成28年10月25日 (2016. 10. 25)

(24) 登録日 平成28年9月30日 (2016. 9. 30)

(51) Int. Cl. F I
G 1 0 L 19/008 (2013. 01) G 1 0 L 19/008 2 0 0
G 1 0 L 19/20 (2013. 01) G 1 0 L 19/20

請求項の数 22 (全 43 頁)

(21) 出願番号	特願2015-549414 (P2015-549414)	(73) 特許権者	507236292
(86) (22) 出願日	平成25年11月25日 (2013. 11. 25)		ドルビー ラボラトリーズ ライセンシング コーポレーション
(65) 公表番号	特表2016-509249 (P2016-509249A)		アメリカ合衆国 9 4 1 0 3 カリフォル ニア州 サンフランシスコ マーケット ストリート 1 2 7 5
(43) 公表日	平成28年3月24日 (2016. 3. 24)	(74) 代理人	100107766
(86) 国際出願番号	PCT/US2013/071679		弁理士 伊東 忠重
(87) 国際公開番号	W02014/099285	(74) 代理人	100070150
(87) 国際公開日	平成26年6月26日 (2014. 6. 26)		弁理士 伊東 忠彦
審査請求日	平成27年6月18日 (2015. 6. 18)	(74) 代理人	100091214
(31) 優先権主張番号	61/745, 401		弁理士 大貫 進介
(32) 優先日	平成24年12月21日 (2012. 12. 21)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	61/865, 072		
(32) 優先日	平成25年8月12日 (2013. 8. 12)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 知覚的基準に基づいてオブジェクト・ベースのオーディオ・コンテンツをレンダリングするためのオブジェクト・クラスタリング

(57) 【特許請求の範囲】

【請求項 1】

オブジェクト・ベースのオーディオ・データを圧縮する方法であって：

オーディオ・シーン内のオブジェクトの知覚的重要性を判別する段階であって、前記オブジェクトはオブジェクト・オーディオ・データおよび関連付けられたメタデータを含む、段階と；

前記オーディオ・オブジェクトの判別された知覚的重要性に基づいて、ある種のオーディオ・オブジェクトをオーディオ・オブジェクトのクラスターに組み合わせる段階であって、クラスターの数はオーディオ・シーン内のオーディオ・オブジェクトのものと数より少ない、段階とを含む、

ある種のオーディオ・オブジェクトをクラスターに組み合わせる段階が、最高の知覚的重要性をもつ諸オーディオ・オブジェクトに対応する諸クラスターについての諸重心を選択し、残りのオーディオ・オブジェクトのうち少なくとも一つを、パン技法によって、前記クラスターのうち二つ以上にわたって分配することを含む、方法。

【請求項 2】

前記知覚的重要性が、前記オーディオ・オブジェクトの前記オブジェクト・オーディオ・データから導出される、請求項 1 記載の方法。

【請求項 3】

前記知覚的重要性が、それぞれのオーディオ・オブジェクトのラウドネス値およびコン

テンツ型の少なくとも一方から導出される値であり、前記コンテンツ型が、ダイアログ、音楽、サウンド効果、周辺音およびノイズからなる群から選択される、請求項 1 または 2 記載の方法。

【請求項 4】

前記オーディオ・オブジェクトの判別される知覚的重要性が、オーディオ・シーンにおける前記オーディオ・オブジェクトの相対的な空間的位置に依存し、前記組み合わせる段階が：

いくつかの重心を決定する段階であって、前記重心は複数のオーディオ・オブジェクトをグループ化するクラスターの中心であり、重心位置は、一つまたは複数のオーディオ・オブジェクトの他のオーディオ・オブジェクトに対する知覚的重要性に依存する、段階と

10

；
オーディオ・オブジェクト信号を諸クラスターを横断して分配することによって前記オーディオ・オブジェクトを一つまたは複数のクラスターにグループ化する段階とを含む、請求項 1 または 2 記載の方法。

【請求項 5】

クラスター・メタデータが、高い知覚的重要性の一つまたは複数のオーディオ・オブジェクトによって決定される、請求項 1 または 2 記載の方法。

【請求項 6】

前記コンテンツ型がオーディオ分類プロセスによって決定され、前記ラウドネス値が知覚的モデルによって得られる、請求項 3 記載の方法。

20

【請求項 7】

前記知覚的モデルが、前記入力オーディオ信号の諸臨界周波数帯域における励起レベルの計算に基づき、当該方法がさらに：

前記オーディオ・オブジェクトのうちの第一のオーディオ・オブジェクトのまわりのクラスターについての重心を定義する段階と；

前記オーディオ・オブジェクトのすべての励起を総合する段階とを含む、請求項 6 記載の方法。

【請求項 8】

前記ラウドネス値が少なくとも部分的には、それぞれのオーディオ・オブジェクトの、他のオーディオ・オブジェクトへの空間的近接性に依存する、請求項 3 記載の方法。

30

【請求項 9】

前記空間的近接性が、少なくとも部分的には、それぞれのオーディオ・オブジェクトの関連付けられたメタデータの位置メタデータ値によって定義される、請求項 8 記載の方法。

【請求項 10】

前記組み合わせる段階が、各クラスタリングされたオーディオ・オブジェクトに関連するある空間的誤差を引き起こし、さらに、当該方法がさらに、相対的に高い知覚的重要性のオーディオ・オブジェクトについて空間的誤差が最小化されるようオーディオ・オブジェクトをクラスタリングすることを含む、請求項 1 または 2 記載の方法。

【請求項 11】

40

前記励起レベル、前記ラウドネスまたはそれから導出される属性を、グループ化されたオーディオ・オブジェクトの相対的な知覚的重要性によって導出される時定数に基づいて平滑化することをさらに含む、請求項 7 記載の方法。

【請求項 12】

オブジェクト・ベースのオーディオを処理する方法であって：

各オーディオ・オブジェクトの、前記複数のオーディオ・オブジェクトの他のオーディオ・オブジェクトに対する第一の空間的位置を決定する段階と；

前記複数のオーディオ・オブジェクトの各オーディオ・オブジェクトの部分ラウドネスを決定することによって前記複数のオーディオ・オブジェクトの各オーディオ・オブジェクトの相対的重要性を決定する段階であって、前記相対的重要性は、オーディオ・オブジ

50

エクトの前記相対的な空間位置に依存し、オーディオ・オブジェクトの前記部分ラウドネスは少なくとも部分的には、一つまたは複数の他のオーディオ・オブジェクトのマスキング効果に基づく、段階と；

いくつかの重心を決定する段階であって、各重心は複数のオーディオ・オブジェクトをグループ化するクラスターの中心であり、重心位置は一つまたは複数のオーディオ・オブジェクトの前記相対的重要性に依存する、段階と；

オーディオ・オブジェクト信号を複数のクラスターを横断して分配することによって、前記オーディオ・オブジェクトを一つまたは複数のクラスターにグループ化する段階とを含む、方法。

10

【請求項 13】

前記複数のオーディオ・オブジェクトの各オーディオ・オブジェクトのコンテンツ型および関連付けられたコンテンツ型重要性を判別する段階をさらに含む、請求項 12 記載の方法。

【請求項 14】

各オーディオ・オブジェクトの前記部分ラウドネスおよび前記コンテンツ型を組み合わせるそれぞれのオーディオ・オブジェクトの前記相対的重要性を決定する段階をさらに含む、請求項 13 記載の方法。

【請求項 15】

前記コンテンツ型が、ダイアログ、音楽、サウンド効果、周辺音およびノイズからなる群から選択される、請求項 13 または 14 記載の方法。

20

【請求項 16】

前記部分ラウドネスが、前記入力オーディオ信号の諸臨界周波数帯域における励起レベルの計算に基づく知覚的モデルによって得られ、当該方法がさらに：

前記オーディオ・オブジェクトのうちの第一のオーディオ・オブジェクトのまわりのクラスターについての重心を定義する段階と；

前記オーディオ・オブジェクトのすべての励起を総合する段階とを含む、請求項 12 ないし 14 のうちいずれか一項記載の方法。

【請求項 17】

前記オーディオ・オブジェクトをグループ化することが、各クラスタリングされたオーディオ・オブジェクトに関連するある空間的誤差を引き起こし、さらに、当該方法がさらに、相対的に高い知覚的重要性のオーディオ・オブジェクトについて空間的誤差が最小化されるよう前記オーディオ・オブジェクトをグループ化することを含む、請求項 12 ないし 14 のうちいずれか一項記載の方法。

30

【請求項 18】

最高の知覚的重要性をもつオーディオ・オブジェクトを、最高の知覚的重要性をもつオーディオ・オブジェクトを含むクラスターのクラスター重心として選択すること、あるいは最大ラウドネスをもつオーディオ・オブジェクトを、該最大ラウドネスをもつオーディオ・オブジェクトを含むクラスターについてのクラスター重心として選択することのうちの一方をさらに含む、請求項 17 記載の方法。

40

【請求項 19】

前記オーディオ・オブジェクトをグループ化することが：

同じクラスター内の構成要素オーディオ・オブジェクトについてのオーディオ・データを具現する波形を一緒に組み合わせるそれらの構成要素オーディオ・オブジェクトの組み合わせられた波形をもつ置換オーディオ・オブジェクトを形成する段階と；

同じクラスター内の構成要素オーディオ・オブジェクトについてのメタデータを組み合わせるそれらの構成要素オーディオ・オブジェクトについてのメタデータの置換セットを形成する段階とを含む、

請求項 12 ないし 14 のうちいずれか一項記載の方法。

【請求項 20】

50

請求項 1 ないし 1 1 のうちいずれか一項記載のオブジェクト・ベースのオーディオを圧縮する方法を実行するよう構成されている、オブジェクト・ベースのオーディオ・データを圧縮する装置。

【請求項 2 1】

請求項 1 2 ないし 1 9 のうちいずれか一項記載のオブジェクト・ベースのオーディオを処理する方法を実行するよう構成されている、オブジェクト・ベースのオーディオ・データを処理する装置。

【請求項 2 2】

プロセッサに請求項 1 ないし 1 9 のうちいずれか一項記載の方法を実行させる実行可能命令を有するプロセッサ読み取り可能なプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願への相互参照

本願は2012年12月21日に提出された米国仮特許出願第61/745,401号および2013年8月12日に提出された米国仮出願第61/885,072号の優先権の利益を主張するものである。両出願はここに参照によってその全体において組み込まれる。

【0002】

発明の技術分野

一つまたは複数の実施形態は概括的にはオーディオ信号処理に関し、より詳細には、効率的な符号化および/またはさまざまな再生システムを通じたレンダリングのためにオブジェクト・ベースのオーディオ・データを圧縮するために、知覚的基準に基づいてオーディオ・オブジェクトをクラスタリングすることに関する。

【背景技術】

【0003】

オブジェクト・ベースのオーディオの到来は、オーディオ・データの量およびハイエンド再生システム内でのこのデータのレンダリングの複雑さを著しく増大させた。たとえば、映画サウンドトラックは、スクリーン上の異なる位置から発し、背景音楽および周辺効果と組み合わせさせて全体的な聴覚経験を作り出す、スクリーン上の画像、ダイアログ、ノイズおよびサウンド効果に対応する多くの異なる音要素を含むことがある。正確な再生は、音が、音源位置、強度、動きおよび奥行きに関して、できるだけスクリーン上に示されているものに密接に対応するような仕方で再生されることを要求する。オブジェクト・ベースのオーディオは、オーディオ・コンテンツを聴取環境中の個々のスピーカーへのスピーカー・フィールドの形で送りよって個々のオーディオ・オブジェクトの空間的な再生に関して比較的制限されている伝統的なチャンネル・ベースのオーディオ・システムに対して、著しい改善を表わす。

【0004】

デジタル映画館の導入および三次元（「3D」）コンテンツの開発は、音についての新たなスタンダードを作り出した。たとえば、コンテンツ・クリエイターにとってのより大きな創造性を許容する複数チャンネル・オーディオの組み込みや、聴衆にとってより包み込むような、リアルな聴覚経験などである。空間的オーディオを配信する手段として伝統的なスピーカー・フィールドおよびチャンネル・ベースのオーディオを超えて拡張することは重要であり、聴取者が所望される再生構成を選択することを許容し、オーディオが選ばれた構成について個別にレンダリングされる、モデル・ベースのオーディオ記述にかなりの関心が寄せられてきた。音の空間的呈示はオーディオ・オブジェクトを利用する。オーディオ・オブジェクトは、見かけの源位置（たとえば3D座標）、見かけの源幅および他のパラメータの、関連付けられたパラメトリックな源記述をもつオーディオ信号である。さらなる進展として、次世代空間的オーディオ（「適応オーディオ（adaptive audio）」とも称される）フォーマットが開発されている。これは、オーディオ・オブジェクトについての位置メタデータとともに、オーディオ・オブジェクトおよび伝統的なチャンネル・ベース

10

20

30

40

50

のスピーカー・フィード(ベッド)の混合を含む。

【0005】

いくつかのサウンドトラックでは、オーディオを含むいくつかの(たとえば7、9または11個の)ベッド・チャンネルがあることがある。さらに、オーサリング・システムの機能に基づいて、レンダリングの間に組み合わせられて空間的に多様かつ没入的なオーディオ経験を生成する数十またはさらには数百の個々のオーディオ・オブジェクトがあることがある。いくつかの配信および伝送システムでは、ほとんどまたは全くオーディオ圧縮なしですべてのオーディオ・ベッドおよびオブジェクトを伝送するのに十分大きい利用可能な帯域幅があることがある。しかしながら、ブルーレイ・ディスク、放送(ケーブル、衛星および地上波)、モバイル(3Gおよび4G)およびオーバーザトップ(OTTまたはインターネット)配信のようないくつかの場合には、オーサリングの時点で生成されたベッドおよびオブジェクト情報のすべてをデジタル的に伝送するための利用可能な帯域幅に対する著しい制限があることがある。必要とされる帯域幅を減らすためにオーディオ符号化方法(不可逆または可逆)がオーディオに適用されてもよいが、オーディオ符号化は、特にモバイル3Gおよび4Gネットワークのような非常に限られたネットワーク上でオーディオを伝送するために必要とされる帯域幅を減らすのに十分でないことがある。

10

【発明の概要】

【発明が解決しようとする課題】

【0006】

クラスタリングによって入力オブジェクトおよびベッドの数を出力オブジェクトのより小さな集合に縮小するためにいくつかの従来の方法が開発されている。本質的には、同様の空間的またはレンダリング属性をもつオブジェクトが単一のまたはより少数の新しい、併合されたオブジェクトに組み合わせられる。併合プロセスは、オーディオ信号を組み合わせること(たとえば総和による)およびパラメトリックな源記述(たとえば平均することによる)を包含する。これら以前の方法におけるクラスターへのオブジェクトの割り当ては、空間的近接性に基づく。すなわち、個々には各オブジェクトについての小さな空間的誤差を保証しつつ、同様のパラメトリック位置データをもつ諸オブジェクトは一つのクラスターに組み合わせられる。このプロセスは一般に、コンテンツ中のすべての知覚的に有意なオブジェクトの空間的位置がそのようなクラスタリングを合理的な程度に小さな誤差をもって許容する限りは有効である。しかしながら、疎な空間的分布を有する同時にアクティブな多数のオブジェクトがある非常に複雑なコンテンツでは、ほどほどの空間的誤差しか許容されない場合には、そのようなコンテンツを正確にモデル化するための必要とされる出力クラスターの数著しくなることがある。あるいはまた、帯域幅または複雑さの制約条件に起因するなどして、出力クラスターの数制約される場合には、制約されたクラスタリング・プロセスおよび著しい空間的誤差のため、複雑なコンテンツは劣化した空間的品質をもって再生されることがある。よって、その場合、クラスターを定義するために近接性のみを使うことは、しばしば、最適でない結果を返す。この場合、オブジェクトの空間的位置だけではなく、オブジェクト自身の重要性が、クラスタリング・プロセスの知覚される品質を最適化するために考慮に入れられるべきである。

20

30

【0007】

クラスタリング・プロセスを改善するために他の解決策も開発されている。一つのそのような解決策は、マスキングのためまたはオブジェクトがサイレントであるためなどで知覚的に有意でないオブジェクトを除去する選別(culling)プロセスである。このプロセスはクラスタリング・プロセスを改善する助けになるものの、知覚的に有意なオブジェクトの数が利用可能な出力クラスターより多い場合には、改善されたクラスタリング結果を提供しない。

40

【0008】

背景セクションにおいて論じられた主題は、単に背景セクションにおいて言及されていることの結果として従来技術であると想定されるべきではない。同様に、背景セクションにおいて言及されているまたは背景セクションの主題に関連する問題は、従来技術におい

50

て以前に認識されていたと想定されるべきではない。背景セクションにおける主題は単に種々のアプローチを表わすものであり、それらのアプローチ自身も発明であることがある。

【課題を解決するための手段】

【0009】

いくつかの実施形態は、再生システムにおいてレンダリングされるべき第一の数のオーディオ・オブジェクトを同定する段階であって、各オーディオ・オブジェクトはオーディオ・データおよび関連付けられたメタデータを含む、段階と；各オーディオ・オブジェクトについての関連付けられたメタデータ内にエンコードされたある種のパラメータについての誤差閾値を定義する段階と；前記誤差閾値に基づいて、前記第一の数のオーディオ・オブジェクトのうちのオーディオ・オブジェクトを低下した数のオーディオ・オブジェクトにグループ化して、前記再生システムを通じて伝送されるオーディオ・オブジェクトについてのデータの量が減らされるようにする段階とを実行することによって、再生システムにおけるレンダリングのためにオブジェクト・ベースのオーディオ・データを圧縮することに向けられる。

10

【0010】

いくつかの実施形態はさらに、定義された時間間隔でいくつかのオブジェクトのうちの各オブジェクトの空間的位置を同定する段階と、オブジェクトの諸対の間の最大距離および/または該グループ化によって前記オブジェクトに関連付けられたある種の他の特性に対して引き起こされる歪み誤差に基づいて、前記オブジェクトのうちの少なくともいくつかを一つまたは複数の時間変化するクラスターにグループ化する段階とを実行することによってオブジェクト・ベースのオーディオをレンダリングすることに向けられる。

20

【0011】

いくつかの実施形態は、オーディオ・シーン内のオブジェクトの知覚的な重要性を決定する段階であって、前記オブジェクトはオブジェクト・オーディオ・データおよび関連付けられたメタデータを含む、段階と、前記オブジェクトの決定された知覚的な重要性に基づいてある種のオーディオ・オブジェクトをオーディオ・オブジェクトのクラスターに組み合わせる段階であって、クラスターの数、前記オーディオ・シーン内のもとのオブジェクトの数よりも少ない、段階とを実行することによって、再生システムにおけるレンダリングのためにオブジェクト・ベースのオーディオ・データを圧縮する方法に向けられる。この方法では、前記知覚的な重要性は、それぞれのオブジェクトのラウドネス値およびコンテンツ型の少なくとも一方から導出される値であってもよく、前記コンテンツ型はダイアログ、音楽、サウンド効果、周辺音およびノイズのうちの少なくとも一つである。

30

【0012】

本方法のある実施形態では、コンテンツ型は、前記オーディオ・オブジェクトについての入力オーディオ信号を受領するオーディオ分類プロセスによって決定され、前記ラウドネスは前記入力オーディオ信号の諸臨界周波数帯域における励起レベルの計算に基づいて知覚的モデルによって得られ、本方法はさらに、前記オーディオ・オブジェクトのうちの第一のオブジェクトのまわりのクラスターについての重心を定義し、前記オーディオ・オブジェクトのすべての励起を総合する (aggregating) ことを含む。前記ラウドネス値は、少なくとも部分的には、それぞれのオブジェクトの他のオブジェクトへの空間的な近接性に依存してもよく、該空間的な近接性は、少なくとも部分的には、それぞれのオブジェクトについての前記関連付けられたメタデータの位置メタデータ値によって定義されてもよい。組み合わせる工程は、各クラスターされるオブジェクトに関連付けられたある種の空間的誤差を引き起こしうる。ある実施形態では、本方法はさらに、相対的に高い知覚的な重要性のオブジェクトについて空間的誤差が最小化されるようオブジェクトをクラスターリングすることを含む。ある実施形態では、オブジェクトの決定された知覚的な重要性は、オーディオ・シーン内の諸オブジェクトの相対的な空間的位置に依存し、組み合わせる段階はさらに、いくつかの重心を決定する段階であって、各重心は複数のオーディオ・オブジェクトをグループ化するためのクラスターの中心を含み、重心位置は他のオーディオ・

40

50

オブジェクトに対する一つまたは複数のオーディオ・オブジェクトの知覚的重要性に依存する、段階と、オブジェクト信号を前記諸クラスターを横断して分配することによって、前記オブジェクトを一つまたは複数のクラスターにグループ化する段階とを含む。クラスタリングはさらに、オブジェクトを最も近い近傍オブジェクトとグループ化することまたはパン方法を使ってオブジェクトを一つまたは複数のクラスターを通じて配送することを含む。

【0013】

オーディオ・オブジェクトを組み合わせる工程は、同じクラスター内の諸構成要素オブジェクトについてのオーディオ・データを具現する諸波形を一緒に組み合わせ、諸構成要素オブジェクトの組み合わせられた波形を有する置換オブジェクトを形成する段階と、同じクラスター内の諸構成要素オブジェクトについてのメタデータを一緒に組み合わせ、諸構成要素オブジェクトについてのメタデータの置換セットを形成する段階とを含んでいてもよい。

10

【0014】

いくつかの実施形態はさらに、いくつかの重心を定義する段階であって、各重心は複数のオーディオ・オブジェクトをグループ化するためのクラスターの中心を含む、段階と、各オブジェクトの、前記複数のオーディオ・オブジェクトのうちの他のオブジェクトに対する第一の空間的位置を決定する段階と、前記複数のオーディオ・オブジェクトの各オーディオ・オブジェクトの相対的重要性を決定する段階であって、前記相対的重要性はオブジェクトの相対的な空間的位置に依存する、段階と、いくつかの重心を定義する段階であって、各重心は複数のオーディオ・オブジェクトをグループ化するためのクラスターの中心を含み、重心位置は一つまたは複数のオーディオ・オブジェクトの相対的重要性に依存する、段階と、オブジェクト信号を前記諸クラスターを横断して分配することによって、前記オブジェクトを一つまたは複数のクラスターにグループ化する段階とを実行することによって、オブジェクト・ベースのオーディオをレンダリングする方法に向けられる。この方法はさらに、前記複数のオーディオ・オブジェクトの各オーディオ・オブジェクトの部分ラウドネスと、前記複数のオーディオ・オブジェクトの各オーディオ・オブジェクトのコンテンツ型および関連付けられたコンテンツ型重要性とを決定する段階を含んでいてもよい。ある実施形態では、各オーディオ・オブジェクトの部分ラウドネスおよびコンテンツ型は組み合わせられて、それぞれのオーディオ・オブジェクトの相対的重要性を決定する。オブジェクトは、相対的に高い知覚的重要性のオブジェクトについて空間的誤差が最小化されるよう、クラスタリングされる。ここで、空間的誤差は、オブジェクトを、他のオブジェクトとクラスタリングされるときに、第一の知覚される源位置から第二の知覚される源位置に動かすことによって引き起こされうるものである。

20

30

【0015】

上記の圧縮する方法またはレンダリングする方法についての実施形態を実装するシステムまたはデバイスおよびコンピュータ可読媒体について、いくつかのさらなる実施形態が記述される。

【0016】

本稿に記載される方法およびシステムは、進んだコンテンツ作成ツールの組によって可能にされる新しいスピーカーおよびチャネル構成ならびに新しい空間的記述フォーマットを含む適応オーディオ・システムに基づく、更新されたコンテンツ作成ツール、配送方法および向上されたユーザー経験を含む、オーディオ・フォーマットおよびシステムにおいて実装されうる。そのようなシステムでは、オーディオ・ストリーム（一般にチャネルおよびオブジェクトを含む）は、オーディオ・ストリームの所望される位置を含むコンテンツ・クリエーターまたはサウンド・ミキサーの意図を記述するメタデータとともに、伝送される。位置は、（あらかじめ定義されたチャネル構成内からの）名前を付けられたチャネルとしてまたは三次元（3D）空間位置情報として表現されることができる。

40

【0017】

参照による組み込み

50

本明細書において言及される各刊行物、特許および/または特許出願はここに参照によって、個々の各刊行物および/または特許出願が具体的かつ個別的に参照によって組み込まれることが示されている場合と同じ程度にその全体において組み込まれる。

【図面の簡単な説明】

【0018】

以下の図面では、同様の参照符号が同様の要素を指すために使われる。以下の図はさまざまな例を描いているが、前記一つまたは複数の実装は図面に描かれる例に限定されるものではない。

【図1】ある実施形態のもとでの、適応オーディオ混合を生成するためのチャンネルおよびオブジェクト・ベースのデータの組み合わせを示す図である。

10

【図2A】ある実施形態のもとでの、適応オーディオ・コンテンツをレンダリングするためのコーデック回路との関連での、クラスタリング・プロセスのブロック図である。

【図2B】ある実施形態のもとでの、適応オーディオ処理システムにおけるオブジェクトおよびベッドのクラスタリングを示す図である。

【図2C】ある実施形態のもとでの、全体的な適応オーディオ・レンダリング・システムにおける適応オーディオ・データのクラスタリングを示す図である。

【図3】Aは、ある実施形態のもとでの、組み合わせられたオブジェクトを生成するための二つのオブジェクトについてのオーディオ信号およびメタデータの組み合わせを示す図である。Bは、ある実施形態のもとでの、クラスタリング・プロセスのための例示的なメタデータ定義および組み合わせ方法を示す表である。

20

【図4】ある実施形態のもとでの、クラスタリング・プロセスによって用いられるクラスタリング方式のブロック図である。

【図5】AおよびBは、ある実施形態のもとでの、周期的な時間区間の間の、オブジェクトのクラスターへのグループ化を示す図である。

【図6A】ある実施形態のもとでの、定義されたオブジェクト境界および誤差閾値との関係での、オブジェクトのクラスターへのグループ化を示す図の一つである。

【図6B】ある実施形態のもとでの、定義されたオブジェクト境界および誤差閾値との関係での、オブジェクトのクラスターへのグループ化を示す図の一つである。

【図6C】ある実施形態のもとでの、定義されたオブジェクト境界および誤差閾値との関係での、オブジェクトのクラスターへのグループ化を示す図の一つである。

30

【図7】ある実施形態のもとでの、オブジェクトおよびベッドをクラスタリングする方法を示すフローチャートである。

【図8】ある実施形態のもとでの、空間的近接性に加えて知覚的重要性に基づいてオブジェクトおよびベッド・チャンネルをクラスターにクラスタリングするシステムを示す図である。

【図9】ある実施形態のもとでの、オーディオ・オブジェクトを出力クラスターにクラスタリングするためのプロセス・フローの構成要素を示す図である。

【図10】ある実施形態のもとでの、オーディオ分類コンポーネントの機能図である。

【図11】ある実施形態のもとでの、コンテンツ型およびラウドネスの知覚的因子に基づいてオーディオ・オブジェクトを処理する全体的な方法を示すフローチャートである。

40

【図12】ある実施形態のもとでの、クラスター重心を計算し、選択された重心にオブジェクトを割り当てるプロセスを示すフローチャートである。

【図13】AおよびBは、ある実施形態のもとでの、ある種の知覚的基準に基づく、オブジェクトのクラスターへのグループ化を示す図である。

【図14】ある実施形態のもとでの、オブジェクトおよびベッドをクラスタリングする方法を示すフローチャートである。

【図15】ある実施形態のもとでの、エンドポイント装置機能に基づくクラスタリングされたオブジェクト・データのレンダリングを示す図である。

【発明を実施するための形態】

【0019】

50

オブジェクト・ベースのオーディオ・データについてオブジェクト・クラスタリング・ベースの圧縮方式のためのシステムおよび方法が記述される。クラスタリング方式の諸実施形態は、オブジェクトをクラスターに割り当てるためにオブジェクトの知覚的重要性を利用し、位置および近接性に基づくクラスタリング方法に対して拡張する。知覚ベースのクラスタリング・システムは、近接性ベースのクラスタリングを、各オブジェクトのオーディオ信号から導出される知覚的相関量で増強し、知覚的に有意なオブジェクトの数が出力クラスターの数より多いときのような制約された条件においてオブジェクトのクラスターへの改善された割り当てを導出する。

【0020】

オーディオ処理システムのある実施形態では、オブジェクト組み合わせまたはクラスタリング・プロセスは、部分的には、オブジェクトの空間的近接性によって、かつある種の知覚的基準によっても制御される。一般に、オブジェクトのクラスタリングの結果として、ある量の誤差が生じる。すべての入力オブジェクトが、他のオブジェクトとクラスタリングされるときに空間的忠実性を維持することはできないからである。特に多数のオブジェクトが疎に分布している用途ではそうである。相対的に高い知覚される重要性をもつオブジェクトは、クラスタリング・プロセスでの空間的/知覚的誤差の最小化に関して、優遇される。オブジェクトの重要性は、シーン内の他のオブジェクトの間でのマスキング効果を考慮に入れたときの、オブジェクトの知覚されるラウドネスである部分ラウドネス (partial loudness) ならびにコンテンツ・セマンティクスまたは型 (たとえば、ダイアログ、音楽、効果など) のような因子に基づくことができる。

【0021】

本稿に記載される一つまたは複数の実施形態の諸側面は、ソフトウェア命令を実行する一つまたは複数のコンピュータまたは処理装置を含むミキシング、レンダリングおよび再生システムにおいて源オーディオ情報を処理するオーディオまたはオーディオビジュアル (AV) システムにおいて実装されてもよい。記載される実施形態のいずれも、単独でまたは互いと一緒に任意の組み合わせにおいて使用されてもよい。さまざまな実施形態が、本明細書の一つまたは複数の場所で論じられるまたは暗示されることがありうる従来技術でのさまざまな欠点によって動機付けられていることがありうるが、それらの実施形態は必ずしもこれらの欠点のいずれかに取り組むものではない。つまり、種々の実施形態は本明細書において論じられることがある種々の欠点に取り組むことがある。いくつかの実施形態は、本明細書において論じられることがあるいくつかの欠点または一つだけの欠点に部分的に取り組むだけであることがあり、いくつかの実施形態はこれらの欠点のどれにも取り組まないこともある。

【0022】

本記述の目的のためには、以下の用語は関連付けられた意味をもつ：用語「チャンネル」および「ベッド」は、オーディオ信号にメタデータを加えたものを意味する。メタデータにおいて、位置はチャンネル識別子、たとえば左前方または右上方サラウンドとして符号化される。「チャンネル・ベースのオーディオ」は、関連付けられた公称位置をもつスピーカー・ゾーンのあらかじめ定義されたセット、たとえば5.1、7.1などを通じた再生のためにフォーマットされたオーディオである。用語「オブジェクト」または「オブジェクト・ベースのオーディオ」は、見かけの源位置 (たとえば3D座標)、見かけの源幅などといったパラメトリックな源記述をもつ一つまたは複数のオーディオ・チャンネルを意味する。「適応オーディオ」は、チャンネル・ベースのおよび/またはオブジェクト・ベースのオーディオ信号に、オーディオ・ストリームに位置が空間内の3D位置として符号化されているメタデータを加えたものを使って、再生環境に基づいてオーディオ信号をレンダリングするメタデータを加えたものを意味する。「レンダリング」は、スピーカー・フィードとして使われる電気信号への変換を意味する。

【0023】

ある実施形態では、オブジェクト・クラスタリングを使ったシーン単純化プロセスは、「空間的オーディオ・システム」または「適応オーディオ・システム」と称されることが

ある音フォーマットおよび処理システムと協働するよう構成されているオーディオ・システムの一部として実装される。そのようなシステムは、向上した聴衆没入感、より大きな芸術的制御ならびにシステム柔軟性およびスケーラビリティを許容するためのオーディオ・フォーマットおよびレンダリング技術に基づく。全体的な適応オーディオ・システムは一般に、通常のチャンネル・ベースのオーディオ要素およびオーディオ・オブジェクト符号化要素の両方を含む一つまたは複数のビットストリームを生成するよう構成されたオーディオ・エンコード、配送およびデコード・システムを含む。そのような組み合わせられたアプローチは、別個に実施されるチャンネル・ベースまたはオブジェクト・ベースのアプローチのいずれと比べても、より大きな符号化効率およびレンダリング柔軟性を提供する。本願の実施形態との関連で使用されてもよい適応オーディオ・システムの例は、2012年6月27日に出願された、「適応オーディオ信号生成、符号化およびレンダリングのためのシステムおよび方法」と題する係属中の国際特許出願第PCT/US2012/044388号において記述されている。同出願はここに参照によって組み込まれる。適応オーディオ・システムおよび関連付けられたオーディオ・フォーマットの例示的な実装は、ドルビー（登録商標）Atmos（商標）プラットフォームである。そのようなシステムは、9.1サラウンド・システムまたは同様のサラウンドサウンド構成として実装されてもよい高さ（上下）次元を組み込む。

10

【0024】

オーディオ・オブジェクトは、聴取環境における一つまたは複数の特定の物理的位置から発するように知覚される個々の音要素または音要素の集合と考えることができる。そのようなオブジェクトは静的（すなわち、定常）または動的（すなわち、動いている）であることができる。オーディオ・オブジェクトは、他の機能とともに所与の時点における音の位置を定義するメタデータによって制御される。オブジェクトが再生されるとき、オブジェクトは、必ずしもあらかじめ定義された物理チャンネルに出力されるのではなく、位置メタデータに従って、存在している諸スピーカーを使ってレンダリングされる。セッションにおけるトラックがオーディオ・オブジェクトであることができ、標準的なパン・データが位置メタデータと類似する。このように、スクリーン上に配置されるコンテンツは、チャンネル・ベースのコンテンツと事実上同じようにしてパンされるが、サラウンドに配置されるコンテンツは、所望されるなら個別のスピーカーにレンダリングされることができる。オーディオ・オブジェクトの使用は離散的な効果に対する制御を提供するものの、サウンドトラックの他の諸側面は、チャンネル・ベースの環境において、より効果的に機能しうる。たとえば、多くの環境効果または残響は実際には、個々のドライバではなくスピーカーのアレイにフィードされることから裨益する。これらはアレイを満たすのに十分な幅をもつオブジェクトとして扱われることができるが、いくつかのチャンネル・ベースの機能を保持することが有益である。

20

30

【0025】

適応オーディオ・システムは、オーディオ・オブジェクトに加えて「ベッド」をサポートするよう構成される。ここで、ベッド（bed）は、事実上、チャンネル・ベースのサブミックスまたはステムである。これらは、コンテンツ・クリエイターの意図に依存して、個々にまたは単一のベッドに組み合わせられて、最終的な再生（レンダリング）のために送達されることができる。これらのベッドは、5.1、7.1および9.1ならびに頭上スピーカーを含むアレイのような種々のチャンネル・ベースの構成において生成されることができる。図1は、ある実施形態のもとでの、適応オーディオ混合を生成するためのチャンネルおよびオブジェクト・ベースのデータの組み合わせを示している。プロセス100において示されるように、たとえばパルス符号変調（PCM）されたデータの形で提供される5.1または7.1サラウンドサウンド・データであってもよいチャンネル・ベースのデータ102は、オーディオ・オブジェクト・データ104と組み合わせられて、適応オーディオ混合108を生成する。オーディオ・オブジェクト・データ104は、もとのチャンネル・ベースのデータの要素を、オーディオ・オブジェクトの位置に関するある種のパラメータを指定する関連するメタデータと組み合わせることによって生成される。図1において概念的に示されるよ

40

50

うに、オーサリング・ツールは、スピーカー・チャンネル・グループおよびオブジェクト・チャンネルの組み合わせを同時に含むオーディオ・プログラムを生成する能力を提供する。たとえば、オーディオ・プログラムは、任意的に諸グループ（または諸トラック、たとえばステレオまたは5.1トラック）に編成されている一つまたは複数のチャンネル、一つまたは複数のスピーカー・チャンネルについての記述メタデータ、一つまたは複数のオブジェクト・チャンネルおよび一つまたは複数のオブジェクト・チャンネルについての記述メタデータを含むことができる。

【0026】

適応オーディオ・システムは、空間的オーディオを配送する手段としてのスピーカー・フィールドを越えて拡張され、個別のニーズおよびシステム制約条件に適合する再生構成を調整するために進んだモデル・ベースのオーディオ記述を使い、オーディオが個々の構成のために特にレンダリングされることのできるようにする。オーディオ信号の空間的効果は、聴取者にとっての没入的経験を提供することにおいて枢要である。閲覧スクリーンまたは部屋の特定の領域から発することが意図される音は、その同じ相対位置に位置されるスピーカー（単数または複数）を通じて再生されるべきである。よって、モデル・ベースの記述における音イベントの主要なオーディオ・メタデータは位置である。ただし、サイズ、配向、速度および音響分散（acoustic dispersion）のような他のパラメータが記述されることもできる。

【0027】

上記のように、適応オーディオ・コンテンツは、いくつかのベッド・チャンネル102を、レンダリングの間に組み合わせられて空間的に多様かつ没入的なオーディオ経験を作り出す多くの個々のオーディオ・オブジェクト104とともに含んでもよい。大量の処理帯域幅のある映画館環境では、事実上、任意の数のベッドおよびオブジェクトがシアター内で生成され、正確にレンダリングされることのできる。しかしながら、映画館または他の複雑なオーディオ・コンテンツが家庭または個人的な聴取環境における配送および再生のために制作される際には、そのような装置およびメディアの比較的制限された処理帯域幅は、このコンテンツの最適なレンダリングまたは再生を妨げる。たとえば、消費者およびプロフェッショナルの用途のために使われる典型的な伝送媒体は、ブルーレイ・ディスク、放送（ケーブル、衛星および地上波）、モバイル（3Gおよび4G）およびオーバーザトップ（OTT）またはインターネット配送を含む。これらの媒体チャンネルは、適応オーディオ・コンテンツのベッドおよびオブジェクト情報のすべてをデジタル的に伝送するための利用可能な帯域幅に対して著しい制限を課すことがある。諸実施形態は、そのままオーディオ・ベッドおよびオブジェクト・データのすべてをレンダリングするためには十分大きな利用可能な帯域幅を有さないことがある伝送システムを通じて配送されよう、複雑な適応オーディオ・コンテンツを圧縮する機構に向けられる。

【0028】

現行のモノフォニック、ステレオおよびマルチチャンネル・オーディオ・コンテンツでは、上述した送達方法およびネットワークの帯域幅制約条件は、必要とされる帯域幅を配送方法の利用可能な帯域幅にマッチするよう低下させるために、一般にオーディオ符号化が必要とされるようなものである。現在の映画館システムは、典型的な7.1映画館フォーマットのための10Mbpsのオーダーの帯域幅で、圧縮されていないオーディオ・データを提供することができる。この容量と比較して、さまざまな他の送達方法および再生システムについての利用可能な帯域幅は実質的に少ない。たとえば、ディスク・ベースの帯域幅は、数百kbpsから数十Mbpsのオーダーである。放送帯域幅は数百kbpsから数十kbpsのオーダーである。OTTインターネット帯域幅は数百kbpsから数Mbpsのオーダーである。モバイル（3G/4G）はたった数百kbpsから数十kbpsのオーダーである。適応オーディオはフォーマットの一部である追加的なオーディオ・エッセンスを含むので、すなわちチャンネル・ベッド102に加えてオブジェクト104を含むので、伝送帯域幅に対するすでに著しい制約条件は、通常のチャンネル・ベースのオーディオ・フォーマットを超えて一層厳しくなり、低下した帯域幅の伝送および再生システムにおける正確な再生を容易にするために、オーディ

10

20

30

40

50

オ符号化ツールに加えて、追加的な帯域幅削減が必要とされる。

【0029】

オブジェクト・クラスタリングを通じたシーン単純化

ある実施形態では、適応オーディオ・システムは、オブジェクト・クラスタリングと、チャンネル・ベッドおよびオブジェクトの組み合わせによって作り出される空間的シーンの、知覚的に透明な単純化とを通じてオブジェクト・ベースのオーディオ・コンテンツの帯域幅を削減するコンポーネントを提供する。前記コンポーネントによって実行されるオブジェクト・クラスタリング・プロセスは、同様のオブジェクトをグループ化してもとのオブジェクトの代わりとなるオブジェクト・クラスターにすることによって空間的シーンの複雑さを低下させるために、空間的位置、コンテンツ型、時間的属性、オブジェクト幅およびラウドネスを含むオブジェクトについてのある種の情報を使う。

10

【0030】

もとの複雑なベッドおよびオーディオ・トラックに基づいて説得力のあるユーザー経験を配送およびレンダリングするための標準的なオーディオ符号化のための追加的なオーディオ処理は、一般に、シーン単純化および/またはオブジェクト・クラスタリングと称される。この処理の目的は、再生装置に送達される個別のオーディオ要素（ベッドおよびオブジェクト）の数を減らす、それでももともとオーサリングされたコンテンツとレンダリングされる出力との間の知覚される差が最小化されるよう十分な空間的情報を保持するクラスタリングまたはグループ化技法を通じて、空間的シーンを削減することである。

【0031】

シーン単純化プロセスは、オブジェクトを削減された数に動的にクラスタリングするために空間的位置、時間的属性、コンテンツ型、幅および他の適切な特性を含む当該オブジェクトについての情報を使って、低下した帯域幅のチャンネルまたは符号化システムにおける、オブジェクトにベッドを加えたコンテンツのレンダリングを容易にする。このプロセスは、以下のクラスタリング動作を実行することによってオブジェクトの数を削減することができる：（1）オブジェクトをオブジェクトにクラスタリングする；（2）オブジェクトをベッドとクラスタリングする；（3）オブジェクトおよびベッドをオブジェクトにクラスタリングする。さらに、オブジェクトは、二つ以上のクラスターにわたって分配することができる。上記プロセスはさらに、オブジェクトのクラスタリングおよび脱クラスタリングを制御するために、オブジェクトについてのある種の時間的および/または知覚的情報を使う。オブジェクト・クラスターは、構成要素オブジェクトの個々の波形およびメタデータ要素を、単一の等価な波形およびメタデータ・セットで置き換え、それによりN個のオブジェクトについてのデータが単一のオブジェクトについてのデータで置き換えられ、本質的に、オブジェクト・データをNから1に圧縮する。上述したように、代替的または追加的に、オブジェクトまたはベッド・チャンネルは（たとえば振幅パン技法を使って）二つ以上のクラスターにわたって分配されてもよい。それによりオブジェクト・データはNからMに圧縮される。ここで、 $M < N$ である。クラスタリング・プロセスは、クラスタリング圧縮とクラスタリングされたオブジェクトの音劣化との間の最適なトレードオフを決定するために、クラスタリングされるオブジェクトの位置、ラウドネスまたは他の特性における変化に起因する歪みに基づく誤差メトリックを利用する。クラスタリング・プロセスは、同期的に実行されることができ、あるいはイベント駆動であって、たとえばクラスタリングを通じたオブジェクト単純化を制御するために聴覚的シーン解析（ASA: auditory scene analysis）およびイベント境界検出を使うことによることができる。いくつかの実施形態では、上記プロセスは、クラスタリングを制御するために、エンドポイント・レンダリング・アルゴリズムおよび装置の知識を利用してもよい。このようにして、再生装置のある種の特性または属性が、クラスタリング・プロセスに情報を与えるために使われてもよい。たとえば、ヘッドフォンまたは他のオーディオ・ドライバに比してスピーカーについては異なるクラスタリング方式が利用されてもよく、あるいは不可逆符号化に比して可逆符号化については異なるクラスタリング方式が利用されてもよい、など。

20

30

40

【0032】

50

以下の記述の目的のためには、用語「クラスタリング」および「グループ化」または「組み合わせること」は、適応オーディオ再生システムにおける伝送およびレンダリングのために適応オーディオ・コンテンツの単位内のデータの量を削減するためのオブジェクトおよび/またはベッド（チャンネル）の組み合わせを記述するために交換可能に使用される。用語「圧縮」または「削減」は、オブジェクトおよびベッドのそのようなクラスタリングを通じて適応オーディオのシーン単純化を実行する工程を指すために使われることがある。本記述を通じて用語「クラスタリング」、「グループ化」または「組み合わせること」は、オブジェクトまたはベッド・チャンネルの単一のクラスターのみへの厳密に一意的な割り当てに限定されるものではない。そうではなく、オブジェクトまたはベッド・チャンネルは、重みまたは利得ベクトルを使って二つ以上の出力ベッドまたはクラスターにわたって分配されてもよい。該重みまたは利得ベクトルは、オブジェクトまたはベッド信号の、出力クラスターまたは出力ベッド信号への相対的な寄与を決定する。

10

【 0 0 3 3 】

図 2 A は、ある実施形態のもとでの、適応オーディオ・コンテンツのレンダリングのためのコーデック回路との関連でクラスタリング・プロセスを実行するクラスタリング・コンポーネントのブロック図である。描画 2 0 0 に示されるように、回路 2 0 0 は、低下した帯域幅で出力オーディオ信号を生成するために入力オーディオ信号を処理するエンコーダ 2 0 4 およびデコーダ 2 0 6 段を含む。図 2 A に示される例については、入力信号の一部 2 0 9 が既知の圧縮技法を通じて処理されて、圧縮されたオーディオ・ビットストリーム 2 0 5 を生成してもよい。この圧縮されたオーディオ・ビットストリーム 2 0 5 がデコーダ段 2 0 6 によってデコードされて出力 2 0 7 の少なくとも一部を生成する。そのような既知の圧縮技法は、入力オーディオ・コンテンツ 2 0 9 を解析し、オーディオ・データを量子化し、次いでオーディオ・データ自身に対してマスキングなどの圧縮技法を実行することに関わる。圧縮技法は不可逆であっても可逆であってもよく、ユーザーが 192kbps、256kbps、512kbps などといった圧縮された帯域幅を選択することを許容しうるシステムにおいて実装される。

20

【 0 0 3 4 】

適応オーディオ・システムにおいて、入力オーディオの少なくとも一部は、オーディオおよびメタデータからなるオブジェクトを含む入力信号 2 0 1 を含む。メタデータは、オブジェクト空間位置、コンテンツ型、ラウドネスなどといった、関連するオーディオ・コンテンツのある種の特徴を定義する。いかなる実際的な数のオーディオ・オブジェクト（たとえば数百のオブジェクト）が再生のために上記システムを通じて処理されてもよい。幅広い多様な再生システムおよび伝送媒体におけるこれら多数のオブジェクトの正確な再生を容易にするために、システム 2 0 0 は、もとのオブジェクトをより少数のオブジェクト・グループに組み合わせることによってオブジェクトの数を、オブジェクトのより少ない、より扱いやすい数まで削減するクラスタリング・プロセスまたはコンポーネント 2 0 2 を含む。このように、クラスタリング・プロセスはオブジェクトのグループを構築して、個々の入力オブジェクト 2 0 1 のもとのセットから、より少数の出力グループ 2 0 3 を生成する。クラスタリング・プロセス 2 0 2 は本質的には、オーディオ・データ自身のほかにオブジェクトのメタデータを処理して、削減された数のオブジェクト・グループを生成する。任意の時点におけるどのオブジェクトが他のオブジェクトと最も適切に組み合わせられるかを決定するために、メタデータが解析され、次いで、組み合わせられる諸オブジェクトのための対応する諸オーディオ波形が合計されて、代替オブジェクトまたは組み合わせられたオブジェクトを生成する。組み合わせられたオブジェクト・グループは次いでエンコーダ 2 0 4 に入力され、該エンコーダ 2 0 4 が、デコーダ 2 0 6 への伝送のためのオーディオおよびメタデータを含むビットストリーム 2 0 5 を生成する。

30

40

【 0 0 3 5 】

一般に、オブジェクト・クラスタリング・プロセス 2 0 2 を組み込む適応オーディオ・システムは、もとの空間的オーディオ・フォーマットからメタデータを生成する諸コンポーネントを含む。コーデック回路 2 0 0 は、通常のチャンネル・ベースのオーディオ要素お

50

よびオーディオ・オブジェクト符号化要素の両方を含む一つまたは複数のビットストリームを処理するよう構成されたオーディオ・レンダリング・システムの一部を含む。諸オーディオ・オブジェクト符号化要素を含む拡張層が、チャンネル・ベースのオーディオ・コーデック・ビットストリームまたはオーディオ・オブジェクト・ビットストリームのいずれか一方に加えられる。このアプローチは、既存のスピーカーおよびドライバ設計または個々に指定可能なドライバおよびドライバ定義を利用する次世代スピーカーと一緒に使うためのレンダラーによって処理されるべき拡張層を含むビットストリーム205を可能にする。この空間的オーディオ・プロセッサからの空間的オーディオ・コンテンツは、オーディオ・オブジェクト、チャンネルおよび位置メタデータを含む。オブジェクトがレンダリングされるとき、該オブジェクトは、位置メタデータおよび再生スピーカーの位置に従って、一つまたは複数のスピーカーに割り当てられる。追加的なメタデータがオブジェクトに関連付けられていて、再生位置を変更したりまたは他の仕方でも再生のために使われるスピーカーを制限したりしてもよい。メタデータは、空間的パラメータ（たとえば位置、速度、強度、音色など）を制御するレンダリング手がかりを提供し、聴取環境におけるどのドライバ（単数または複数）またはスピーカー（単数または複数）が披露の間にそれぞれの音を再生するかを指定するエンジニアのミキシング入力にตอบสนองして、オーディオ・ワークステーションにおいて生成されてもよい。該メタデータは、空間的オーディオ・プロセッサによるパッケージングおよび転送のために、ワークステーションにおいてそれぞれのオーディオ・データと関連付けられる。

10

【0036】

20

図2Bは、ある実施形態のもとでの、適応オーディオ処理システムにおけるオブジェクトおよびベッドのクラスタリングを示している。描画250に示されるように、ある種のシーン単純化タスクを実行するオブジェクト処理コンポーネント256は、任意の数の入力オーディオ・ファイルおよびメタデータを読み込む。入力オーディオ・ファイルは入力オブジェクト252および関連付けられたオブジェクト・メタデータならびにベッド254および関連付けられたベッド・メタデータを含む。このように、この入力ファイル/メタデータは、「ベッド」または「オブジェクト」トラックに対応する。オブジェクト処理コンポーネント256は、より少数の出力オブジェクトおよびベッド・トラックを生成するために、メディア・インテリジェンス/コンテンツ分類、空間的歪み解析およびオブジェクト選択/クラスタリングを組み合わせる。具体的には、オブジェクトは一緒にクラスタリングされて、新たな等価な諸オブジェクトまたは諸オブジェクト・クラスター258を、関連付けられたオブジェクト/クラスター・メタデータとともに生成することができる。これらのオブジェクトは、ベッドへの「下方混合」のために選択されることもできる。これは、出力ベッド・オブジェクトおよび関連付けられたメタデータ270を形成するためにベッド262との組み合わせ268のためにレンダラー266に入力される下方混合されたオブジェクト260の出力として示されている。出力ベッド構成270（たとえば、家庭用の典型的な5.1）は必ずしも、たとえばAtmos（商標）映画館であることができる入力ベッド構成と一致する必要はない。入力トラックからのメタデータを組み合わせることによって、出力トラックについて新しいメタデータが生成される。入力トラックからのオーディオを組み合わせることによって、出力トラックについて新しいオーディオも生成される。

30

40

【0037】

オブジェクト処理コンポーネント256はある種の処理構成設定情報272を利用する。ある実施形態では、これは出力オブジェクトの数、フレーム・サイズおよびある種のメディア・インテリジェンス設定を含む。メディア・インテリジェンスとは、コンテンツ型（すなわち、ダイアログ/音楽/効果など）、領域（セグメント/分類）、前処理結果、聴覚的シーン解析結果および他の同様の情報といった、オブジェクトに関連付けられたいくつかのパラメータまたは特性を含むことができる。

【0038】

ある代替的な実施形態では、単純化メタデータ（たとえば、どのオブジェクトがどのク

50

ラスターに属するか、どのオブジェクトがベッドにレンダリングされるか、など)のほか、にすべてのもとのトラックへの参照を保持することによって、オーディオ生成は猶予されることできる。これは、スタジオとエンコード・ハウスとの間で、または他の同様のシナリオにおいて単純化プロセスを分散させるために有用であることがある。

【0039】

図2Cは、ある実施形態のもとでの、全体的な適応オーディオ・レンダリング・システムにおける適応オーディオ・データのクラスタリングを示している。全体的な処理システム220は、ポストプロダクション221、伝送(送達/ストリーミング)223および再生システム225(家庭/シアター/スタジオ)という三つの主要な段を有する。図2Cに示されるように、もとの数のオブジェクトを削減された数のオブジェクトまたはオブジェクト・クラスタに組み合わせることによってオーディオ・コンテンツを単純化するための動的クラスタリング・プロセスは、これらの段の一つまたは任意のもの間に行われてもよい。

10

【0040】

ポストプロダクション段221では、映画館および/または家庭ベースの適応オーディオ・コンテンツであることができる入力オーディオ・データ222がメタデータ生成プロセス224に入力される。このプロセスは、位置、幅、脱相関係およびレンダリング・モード情報を含むオブジェクトについての空間的メタデータと、コンテンツ型、オブジェクト境界および相対的重要性(エネルギー/ラウドネス)を含むコンテンツ・メタデータとを生成する。次いで、クラスタリング・プロセス226が入力データに適用されて、ある種のオブジェクトをその空間的近接性、時間的近接性または他の特性に基づいて一緒に組み合わせることによって、全体的な数の入力オブジェクトをより少数のオブジェクトに削減する。クラスタリング・プロセス226は、システムにおいて入力データが処理される際にコンスタントなまたは定期的なプロセスとしてクラスタリングを実行する動的クラスタリング・プロセスであってもよく、目標クラスタ数、オブジェクト/クラスタに対する重要性重み付け、フィルタリング効果などといったある種の制約条件を指定するユーザー入力228を利用してよい。ポストプロダクション段は、混合、脱相関係、リミッターなどといったクラスタのある種の処理を提供するクラスタ下方混合ステップをも含んでいてもよい。ポストプロダクション段は、オーディオ・エンジニアがクラスタリング・プロセスの結果をモニタリングまたは傾聴して、結果が十分でない場合に入力データ222またはユーザー入力228を修正することを許容するレンダリング/モニタリング・オプション232を含んでいてもよい。

20

30

【0041】

伝送段223は一般に、生データからコーデックへのインターフェーシング234および適切なコーデック(たとえばTrueHD、ドルビー・デジタル+など)を使った当該デジタル・データの送達またはストリーミングのための適切な出力フォーマットへのオーディオ・データのパッケージング236を含む。伝送段223では、さらなる動的クラスタリング・プロセス238がポストプロダクション段221の間に生成されたオブジェクトに適用されてもよい。

【0042】

再生システム225は伝送されたデジタル・オーディオ・データを受領し、適切な設備(たとえば増幅器およびスピーカー)を通じた再生のために最終的なレンダリング・ステップ242を実行する。この段の間に、ある種のユーザー入力244および再生システム(計算)機能245情報を使って、オブジェクトをクラスタにさらにグループ化するために、追加的な動的クラスタリング・プロセス240が適用されてもよい。

40

【0043】

ある実施形態では、伝送または再生段のいずれかにおいて実行されるクラスタリング・プロセス240および238は、形成されるクラスタの数および/またはクラスタリングを実行するために使われる情報の量および型の点で、オブジェクト・クラスタリングの量がポストプロダクションのクラスタリング・プロセス226に比べて制限されていても

50

よいという意味で、制限されたクラスタリング・プロセスであってもよい。

【 0 0 4 4 】

図3のAは、ある実施形態のもとでの、組み合わされたオブジェクトを生成するための二つのオブジェクトについてのオーディオ信号およびメタデータの組み合わせを示している。描画300に示されるように、第一のオブジェクトは、波形302として示されるオーディオ信号を、各定義された時間期間（たとえば20ミリ秒）についてのメタデータ312とともに含む。よって、たとえば、波形302が60ミリ秒のオーディオ・クリップである場合、第一のオブジェクトについては、MD1、MD2およびMD3と表わされる三つの異なるメタデータ・インスタンスがある。同じ時間区間について、第二のオブジェクトはオーディオ波形304およびMDa、MDbおよびMDcと表わされる三つの異なる対応するメタデータ・インスタンスを含む。クラスタリング・プロセス202は、これら二つのオブジェクトを組み合わせて、波形306および関連付けられたメタデータ316を含む組み合わされたオブジェクトを作り出す。ある実施形態では、もとの第一および第二の波形302および304がそれらの波形を合計することによって組み合わされて、組み合わされた波形306を生成する。あるいはまた、それらの波形は、システム実装に依存して他の波形組み合わせ方法によって組み合わされることができ、第一および第二のオブジェクトについての各期間におけるメタデータも組み合わされて、MD1a、MD2bおよびMD3cと表わされる組み合わされたメタデータ316を生成する。メタデータ要素の組み合わせは、定義されたアルゴリズムまたは組み合わせ関数に従って実行され、システム実装に依存して変わることができる。種々の型のメタデータはさまざまな異なる仕方で組み合わされることができ

10

20

【 0 0 4 5 】

図3のBは、ある実施形態のもとでの、クラスタリング・プロセスについての例示的なメタデータ定義および組み合わせ方法を示す表である。テーブル350の列352に示されるように、メタデータ定義は、他にもある可能なメタデータ型の中でも、オブジェクト位置、オブジェクト幅、オーディオ・コンテンツ型、ラウドネス、レンダリング・モード、制御信号のようなメタデータ型を含む。メタデータ定義は、各メタデータ型に関連付けられたある種の値を定義する要素を含む。各メタデータ型についての例示的なメタデータ要素はテーブル350の列354に挙げられている。二つ以上のオブジェクトがクラスタリング・プロセス202において一緒に組み合わされるとき、それぞれのメタデータ要素は、定義された組み合わせ方式を通じて組み合わされる。各メタデータ型についての例示的な組み合わせ方式がテーブル350の列356に挙げられている。図3のBに示されるように、二つ以上のオブジェクトの位置および幅はそれぞれ、組み合わされたオブジェクトの位置および幅を導出する重み付けされた平均を通じて組み合わされてもよい。位置に関しては、クラスタリングされる（構成要素）オブジェクトを包含する重心の幾何学的中心が、置換オブジェクトの位置を表わすために使用できる。メタデータの組み合わせは、構成要素オブジェクトのメタデータの（相対的な）寄与を決定するために重みを用いてもよい。そのような重みは、一つまたは複数のオブジェクトおよび/またはベッド・チャネルの（部分）ラウドネスから導出されてもよい。

30

【 0 0 4 6 】

組み合わされたオブジェクトのラウドネスは、構成要素オブジェクトのラウドネスを平均または合計することによって導出されてもよい。ある実施形態では、信号のラウドネス・メトリックは、信号の知覚的なエネルギーを表わし、これは周波数に基づいて重み付けされたエネルギーの指標である。よって、ラウドネスは、聴取者による音の知覚に対応する、スペクトル的に重み付けされたエネルギーである。代替的な実施形態では、ラウドネスの代わりにまたはラウドネスとともに、上記プロセスは信号の純粋なエネルギー（RMSエネルギー）または信号エネルギーの他の何らかの指標を、オブジェクトの重要性を決定する際の因子として使ってもよい。さらに代替的な実施形態では、組み合わされたオブジェクトのラウドネスは、クラスタリングされる諸オブジェクトの部分ラウドネス・データから導出される。ここで、部分ラウドネス（partial loudness）は、音響心理学的原理

40

50

に基づく、オブジェクトおよびベッドの完全なセットのコンテキストにおけるオブジェクトの（相対的な）ラウドネスを表わす。よって、テーブル350に示されるように、ラウドネス・メタデータ型は、絶対ラウドネス、部分ラウドネスまたは組み合わされたラウドネス・メタデータ定義として具現されうる。オブジェクトの部分ラウドネス（または相対的重要性）は、重要性メトリックとしてクラスタリングのために、またはレンダリング・システムがすべてのオブジェクトを個々にレンダリングするための十分な機能をもたない場合にオブジェクトを選択的にレンダリングするための手段として、使用されることができ

【0047】

他のメタデータ型は他の組み合わせ方法を必要とすることがある。たとえば、ある種のメタデータは、論理演算または算術演算を通じて組み合わせられることはできず、よって選択がなされる必要がある。たとえば、あるモードであるか別のモードであるかのいずれかであるレンダリング・モードの場合には、優勢なオブジェクトのレンダリング・モードが、組み合わされたオブジェクトのレンダリング・モードとなるよう割り当てられてもよい。制御信号などといった他の型のメタデータは、用途およびメタデータ特性に依存して選択または組み合わせられうる。

【0048】

コンテンツ型に関しては、オーディオは一般に、ダイアログ、音楽、周辺音、特殊効果などといったいくつかの定義されたコンテンツ型の一つに分類される。オブジェクトは、その期間を通じてコンテンツ型を変えてもよいが、どの特定の時点においても、それは一般にはコンテンツの一つの型のみである。コンテンツ型はこのように、オブジェクトが任意の時点においてコンテンツのある特定の型である確率として表現される。よって、たとえば、コンスタントなダイアログ・オブジェクトは百パーセントの確率のダイアログ・オブジェクトとして表現されることになる一方、ダイアログから音楽に変容するオブジェクトは五十パーセント・ダイアログ/五十パーセント音楽として表現されてもよい。異なるコンテンツ型をもつオブジェクトのクラスタリングは、各コンテンツ型についてのそれぞれの確率を平均し、最も優勢なオブジェクトについてのコンテンツ型確率またはコンテンツ型指標の他の何らかの論理的な組み合わせを選択することによって実行されることができ

コンテンツ型は、 n 次元ベクトルとして表現されてもよい（ n は異なるコンテンツ型の総数であり、たとえばダイアログ/音楽/周辺音/効果の場合には4）。次いで、クラスタリングされる諸オブジェクトのコンテンツ型が適切なベクトル演算を実行することによって導出される。テーブル350に示されるように、コンテンツ型メタデータは、組み合わされたコンテンツ型メタデータ定義として具現されてもよい。ここで、コンテンツ型の組み合わせは組み合わされる諸確率分布（たとえば音楽、発話などの諸確率のベクトル）を反映する。

【0049】

オーディオの分類に関し、ある実施形態では、上記プロセスは、信号を解析し、信号の特徴を識別し、オブジェクトの特徴が特定のクラスの特徴にどのくらいよく一致するかを判別するために、識別された特徴を既知のクラスの特徴と比較するよう、時間フレーム毎に作用する。特徴が特定のクラスにどのくらいよく一致するかに基づいて、分類器は、オブジェクトが特定のクラスに属する確率を同定できる。たとえば、時刻 $t = T$ において、オブジェクトの特徴がダイアログ特徴に非常によく合う場合、オブジェクトは高い確率でダイアログとして分類されることになる。時刻 $t = T + N$ において、オブジェクトの特徴が音楽特徴に非常によく合う場合、オブジェクトは高い確率で音楽として分類されることになる。最後に、時刻 $t = T + 2N$ において、オブジェクトの特徴がダイアログとも音楽とも特によく合わない場合には、オブジェクトは50%音楽および50%ダイアログとして分類されてもよい。

【0050】

図3のBにおけるメタデータ定義の一覧は、ある種の例示的なメタデータ定義を例解することを意図されており、ドライバ定義（数、特性、位置、投射角）、部屋およびスピー

10

20

30

40

50

カー情報を含む較正情報および他の任意の適切なメタデータといった、他の多くのメタデータ要素も可能である。

【0051】

ある実施形態では、図2Aを参照するに、クラスタリング・プロセス202は、コーデックのエンコーダ204およびデコーダ206段とは別個のコンポーネントまたは回路において提供される。コーデック204は、既知の圧縮技法を使った圧縮のために生のオーディオ・データ209を処理するとともに、オーディオおよびメタデータ定義を含む適応オーディオ・データ201を処理するよう構成されていてもよい。一般に、クラスタリング・プロセスは、エンコーダ段204の前にオブジェクトをグループにクラスタリングし、クラスタリングされた諸オブジェクトをデコーダ段206後にレンダリングするエンコーダ前およびデコーダ後プロセスとして実装されてもよい。あるいはまた、クラスタリング・プロセス202は、統合されたコンポーネントとして、エンコーダ204段の一部として含められてもよい。

10

【0052】

図4は、ある実施形態のもとでの、図2のクラスタリング・プロセスによって用いられるクラスタリング方式のブロック図である。描画400に示されるように、第一のクラスタリング方式402は、個々のオブジェクトを他のオブジェクトとクラスタリングして、削減された情報で伝送されることのできるオブジェクトの一つまたは複数のクラスターを形成することに焦点を当てる。この削減は、複数のオブジェクトを記述する、より少ないオーディオまたはより少ないメタデータの形であることができる。オブジェクトのクラスタリングの一つの例は、空間的に関係しているオブジェクトをグループ化する、すなわち、同様の空間的位置に位置しているオブジェクトを組み合わせることである。ここで、空間的位置が「同様」であることは、構成要素オブジェクトを置換クラスターによって定義される位置にシフトさせることに起因する歪みに基づいて、最大誤差閾値によって定義される。

20

【0053】

第二のクラスタリング方式404は、空間的に多様でありうるオーディオ・オブジェクトを、固定した空間的位置を表わすチャンネル・ベッドと組み合わせることが適切であるときを判別する。この型のクラスタリングの例は、もともと三次元空間中を横切っていくものとして表現されていることがありうるオブジェクトを伝送するための十分な利用可能な帯域幅がなく、代わりに、そのオブジェクトをその水平面上への投影中に混合するというものである。これは、一つまたは複数のオブジェクトが静的なチャンネル中に動的に混合されることを許容し、それにより伝送される必要のあるオブジェクトの数を削減する。

30

【0054】

第三のクラスタリング方式406は、ある種の既知のシステム特性の事前の知識を使う。たとえば、エンドポイント・レンダリング・アルゴリズムおよび/または再生システム中の再生装置の知識が、クラスタリング・プロセスを制御するために使用されてもよい。たとえば、典型的な家庭シアター構成は、固定した位置に位置される物理的なスピーカーに依拠する。これらのシステムは、室内のいくつかのスピーカーの不在を埋め合わせて、室内に存在する聴取者仮想スピーカーを与えるためのアルゴリズムを使うスピーカー仮想化アルゴリズムにも依拠することがある。スピーカーの空間的多様性および仮想化アルゴリズムの正確さといった情報が既知であれば、スピーカー構成および仮想化アルゴリズムは限られた知覚的経験を聴取者に提供することができるだけなので、削減された数のオブジェクトを送ることが可能でありうる。この場合、ベッドにオブジェクトを加えたフルの表現を送ることは帯域幅の浪費になることがあり、よってある程度のクラスタリングが適切であろう。他の型の既知の情報がこのクラスタリング方式において使われることもできる。たとえば、クラスタリングを制御するためのオブジェクト(単数または複数)のコンテンツ型またはクラスタリングを制御するためのオブジェクト(単数または複数)の幅である。この実施形態のために、コーデック回路200は、再生装置に基づいて出力オーディオ信号207を適応させるよう構成されていてもよい。この機能は、ユーザーまたは他

40

50

のプロセスが、グループ化されたクラスター 203 の数および圧縮されたオーディオ 211 についての圧縮率を定義することを許容する。種々の伝送媒体および再生装置が著しく異なる帯域幅容量をもつことがあるので、標準的な圧縮アルゴリズムおよびオブジェクト・クラスタリング両方についての柔軟な圧縮方式が有利でありうる。たとえば、入力が第一の数、たとえば100のもとのオブジェクトを含んでいる場合、クラスタリング・プロセスは、ブルーレイ・システムのために20個の組み合わせられたグループ 203 を、あるいは携帯電話再生のために10個のオブジェクトなどを生成するよう構成されていてもよい。クラスタリング・プロセス 202 は、段階的により少数のクラスタリングされたグループ 203 を生成するよう再帰的に適用されてもよい。それにより、異なる再生用途のために出力信号 207 の異なるセットが提供されうる。

10

【0055】

第四のクラスタリング方式 408 は、オブジェクトの動的なクラスタリングおよび脱クラスタリングを制御するために時間的情報を使うことを含む。ある実施形態では、クラスタリング・プロセスは、規則的な間隔または周期で実行される（たとえば10ミリ秒毎に一回）。あるいはまた、個々のオブジェクトの継続時間に基づいて最適なクラスタリング構成を決定するためにオーディオ・コンテンツを解析して処理するために、聴覚的シーン解析（ASA）および聴覚的イベント境界検出のような技法を含む他の時間的事件が使われることができる。

【0056】

描画 400 において示される諸方式が、クラスタリング・プロセス 202 によって、スタンドアロンの工程として、あるいは一つまたは複数の他の方式との組み合わせにおいて実行されることができると注意しておくべきである。これらの方式はまた、他の方式に対していかなる順序で実行されてもよく、クラスタリング・プロセスの実行のためにいかなる特定の順序も必須とはされない。

20

【0057】

クラスタリングが空間的位置に基づく場合 402 については、もとのオブジェクトはクラスターにグループ化され、それらのクラスターについて空間的重心が動的に構築される。重心の位置がそのグループの新たな位置になる。そのグループについてのオーディオ信号は、そのグループに属する各オブジェクトについてのすべてのもとのオーディオ信号のミックダウンである。各クラスターは、そのもとのコンテンツを近似するが、もとの入力オブジェクトと同じコア属性/データ構造を共有する新たなオブジェクトと見ることができる。結果として、各オブジェクト・クラスターはオブジェクト・レンダラーによって直接処理されることができる。

30

【0058】

ある実施形態では、クラスタリング・プロセスは、もとの数のオーディオ・オブジェクトおよび/またはベッド・チャンネルを、目標数の新たな等価なオブジェクトおよびベッド・チャンネルに動的にグループ化する。たいていの実際の応用では、目標数はもとの数より実質的に少ない。たとえば、100個のもとの入力トラックが20個以下の組み合わせられたグループに組み合わせられる。これらの解決策は、ベッドおよびオブジェクト・チャンネルの両方がクラスタリング・プロセスに対して入力および/または出力として利用可能であるシナリオに当てはまる。オブジェクトおよびベッド・トラックの両方をサポートする第一の解決策は、入力ベッド・トラックを、空間内の固定したあらかじめ定義された位置をもつオブジェクトとして処理するというものである。これは、システムが、たとえばオブジェクトおよびベッドの両方を含むシーンを、目標数のオブジェクト・トラックのみに単純化することを許容する。しかしながら、クラスタリング・プロセスの一部として、出力ベッド・トラックの数を保存することが望ましいこともありうる。その場合、より重要でないオブジェクトは、前置プロセスとして、ベッド・トラックに直接レンダリングされることができ、一方、最も重要な諸オブジェクトは、より少ない目標数の等価なオブジェクト・トラックにさらにクラスタリングされることができる。結果として得られるクラスターのいくつかが高い歪みをもつ場合、それらのクラスターは、後置プロセスとしてベッドに

40

50

レンダリングされることもできる。このほうがもとのコンテンツのよりよい近似につながりうるからである。誤差 / 歪みは時間変化する関数なので、この決定は、時間変化する仕方になされることのできる。

【 0 0 5 9 】

ある実施形態では、クラスタリング・プロセスは、すべての個々の入力トラック（オブジェクトまたはベッド）201のオーディオ・コンテンツおよび付属のメタデータ（たとえばオブジェクトの空間的位置）を解析して、所与の誤差メトリックを最小にする等価な数の出力オブジェクト / ベッド・トラックを導出することに関わる。基本的な実装では、誤差メトリックは、クラスタリングされるオブジェクトをシフトさせることに起因する空間的歪みに基づき、時間を追った各オブジェクトの重要性の指標によってさらに重み付け 10
されることができ、オブジェクトの重要性は、ラウドネス、コンテンツ型および他の有意な因子といったオブジェクトの他の特性を表わすことができる。あるいはまた、これら他の因子は、空間的な誤差メトリックと組み合わせられることのできる別個の誤差メトリックを形成することのできる。

【 0 0 6 0 】

誤差計算

クラスタリング・プロセスは本質的には、システムを通じて伝送されるデータの量を削減するが、もとのオブジェクトをより少数のレンダリングされるオブジェクトに組み合わせることによって起因するある程度のコンテンツ劣化を本来的に導入する、不可逆圧縮方式の型 20
を表わす。上記のように、オブジェクトのクラスタリングに起因する劣化は、誤差メトリックによって定量化される。もとのオブジェクトの比較的少数の組み合わせられたグループへの削減が大きいほど、および / またはもとのオブジェクトを組み合わせられたグループにする空間的縮退の量が大きいほど、一般に、誤差が大きくなる。ある実施形態では、クラスタリング・プロセスにおいて使われる誤差メトリックは、式(1)に示されるように表現される。

【 0 0 6 1 】

$$E(s,c)[t] = \text{Importance}_s[t] * \text{dist}(s,c)[t] \quad (1)。$$

【 0 0 6 2 】

上記のように、オブジェクトは、他のオブジェクトと一緒に単一のクラスターにグループ化されるのではなく、二つ以上のクラスターにわたって分配されてもよい。インデックスsをもつオブジェクト信号x(s)[t]が二つ以上のクラスターcにわたって分配されるとき、代表クラスター・オーディオ信号y(c)[t]は振幅利得g(s,c)[t]を使って、式(2)に示されるように、 30

$$y(c)[t] = \sum_s g(s,c)[t] * x(s)[t] \quad (2)$$

である。各クラスターcについての誤差メトリックE(s,c)[t]は、式(1)で表わされる諸項の、振幅利得g(s,c)[t]の関数である重みによる重み付けされた組み合わせであることができ、式(3)に示されるようになる：

$$E(s,c)[t] = \sum_s (f(g(s,c)[t]) * \text{Importance}_s[t] * \text{dist}(s,c)[t]) \quad (3)。$$

【 0 0 6 3 】

ある実施形態では、クラスタリング・プロセスは、幅または広がり（spread）パラメータをもつオブジェクトをサポートする。幅は、ピンポイント源としてではなく、見かけの空間的広がりをもつ音としてレンダリングされるオブジェクトのために使われる。幅パラメータが増すにつれて、レンダリングされる音はより空間的に拡散したものとなり、結果として、その特定の位置はそれほど有意でなくなる。よって、幅が増すにつれてより多くの位置誤差を支持するよう、クラスタリング歪みメトリックに幅を含めることが有利である。誤差表現E(s,c)はよって、式(4)に示されるように、幅メトリックを取り入れるよう 40
修正されることができ：

$$E(s,c)[t] = \text{Importance}_s[t] * (* (1 - \text{Width}_s[t]) * \text{dist}(s,c)[t] + (1 -) * \text{Width}_s[t]) \quad (4)。$$

【 0 0 6 4 】

上記の式(1)および(3)において、重要性因子 s はオブジェクトの相対重要性であり、 c はクラスタの重心であり、 $\text{dist}(s,c)$ はオブジェクトとクラスタの重心との間の三次元的なユークリッド距離である。これらの量すべては、 $[t]$ の項によって表わされるように、時間的に変化する。オブジェクトの位置に対するサイズの相対的な重みを制御するために、重み付け項が導入されることもできる。

【0065】

重要性関数 $\text{Importance}_s[t]$ は、信号のラウドネスのような信号ベースのメトリックを、各オブジェクトが当該混合の残りに対してどのくらい顕著であるかの、より高レベルの指標と組み合わせたものであることができる。たとえば、同様の信号と一緒にグループ化される傾向となるよう、入力オブジェクトの各対について計算されるスペクトル類似性指標がさらにラウドネス・メトリックに重み付けすることができる。たとえば映画コンテンツについては、スクリーン上のオブジェクトに対してより大きな重要性を与えることが望ましいこともあり、その場合、上記重要性は、前方中央オブジェクトについて最大になりオブジェクトがスクリーン外に移るにつれて減衰していく、方向性のドット積項によってさらに重み付けされることことができる。

【0066】

クラスタを構築するとき、クラスタリングが時間的に一貫するよう、重要性関数は、比較的長い時間窓（たとえば0.5秒）にわたって時間的に平滑化される。このコンテキストでは、オブジェクト開始および停止時刻の先読みまたは事前の知識を含めることが、クラスタリングの精度を改善できる。対照的に、クラスタ重心の等価な空間的位置は、重要性関数のより高いレート（10ないし40ミリ秒）で適応されることことができる。重要性メトリックにおける突然の変化または増加（たとえば過渡検出器を使う）は、上記の比較的長い時間窓を一時的に短くしたり、あるいは該長い時間窓との関係で任意の解析状態をリセットしたりしてもよい。

【0067】

上記のように、コンテンツ型のような他の情報も、追加的な重要性重み付け項として誤差メトリックに含められることことができる。たとえば、映画サウンドトラックでは、ダイアログが音楽およびサウンド効果より重要であると考えられることがある。したがって、対応するオブジェクトの相対的な重要性を増加させることによって、一つまたは若干数のダイアログのみのクラスタ内にダイアログを分離することが好ましいであろう。各オブジェクトの相対的な重要性は、ユーザーによって提供されるまたは手動で調節されることできる。同様に、ユーザーが望むなら、もとのオブジェクトの特定の部分集合だけがクラスタリングまたは単純化されることができ、一方、他のオブジェクトは個々にレンダリングされるオブジェクトとして保持されることになる。コンテンツ型情報は、オーディオ・コンテンツを分類するためにメディア・インテリジェンス技法を使って自動的に生成されることできる。

【0068】

誤差メトリック $E(s,c)$ は、組み合わせられたメタデータ要素に基づくいくつかの誤差成分の関数であることことができる。このように、距離以外の他の情報がクラスタリング誤差において考慮されることができ。たとえば、ダイアログ、音楽、効果などといったオブジェクト型に基づいて、異なるオブジェクトではなく、同様のオブジェクトと一緒にクラスタリングされてもよい。両立しない異なる型のオブジェクトを組み合わせる結果として出力音の歪みまたは劣化が生じることがある。誤差は、クラスタリングされるオブジェクトの一つまたは複数についての不適切なまたは最適でないレンダリング・モードに起因して導入されることもある。同様に、特定の諸オブジェクトについてのある種の制御信号が、クラスタリングされるオブジェクトについて、度外視され、または妥協されることがある。このように、あるオブジェクトがクラスタリングされる時に組み合わせられる各メタデータ要素についての誤差の和を表わす全体的な誤差項が定義されてもよい。全体的な誤差の例示的な表式は式(5)に示される：

$$E_{\text{overall}}[t] = E_{\text{MDn}} \quad (5)。$$

10

20

30

40

50

【 0 0 6 9 】

式(5)において、MD_nは、あるクラスター内に併合される各オブジェクトについて組み合わせられるN個のメタデータ要素の特定のメタデータ要素を表わし、E_{MD_n}はそのメタデータをクラスター中の他のオブジェクトについての対応するメタデータ値と組み合わせることに付随する誤差を表わす。この誤差値は、平均されるメタデータ値（たとえば位置/ラウドネス）については百分率値として、あるいはある値または別の値として選択されるメタデータ値（たとえばレンダリング・モード）については二値の0パーセントもしくは100パーセント値として、表わされてもよく、あるいは他の任意の適切な誤差メトリックであってもよい。図3のBに示されるメタデータ要素については、全体的な誤差は式(6)に示されるように表わすことができる：

$$E_{\text{overall}[t]} = E_{\text{spatial}} + E_{\text{loudness}} + E_{\text{rendering}} + E_{\text{control}} \quad (6).$$

【 0 0 7 0 】

空間的誤差以外の種々の誤差成分が、オブジェクトのクラスタリングおよび脱クラスタリングのための基準として使用されることができる。たとえば、ラウドネスが、クラスタリング挙動を制御するために使われてもよい。個別ラウドネス（specific loudness）は、音響心理学的原理に基づくラウドネスの知覚的な指標である。種々のオブジェクトの個別ラウドネスを測定することによって、オブジェクトの知覚されるラウドネスが、該オブジェクトがクラスタリングされるか否かを案内しうる。たとえば、ラウドネスが大きいオブジェクトは、その空間的な軌跡が修正される場合に、聴取者にとって、より明白になる可能性が高い。一方、より静かなオブジェクトについては逆のことが一般に成り立つ。したがって、個別ラウドネスは、オブジェクトのクラスタリングを制御するための、空間的誤差に加えた重み付け因子として使われることができる。もう一つの例は、オブジェクト型である。ここで、オブジェクトのいくつかの型は、その空間的編成が修正される場合に、より知覚されやすくなりうる。たとえば、人間は発話信号に対して非常に敏感であり、これらの型のオブジェクトは、空間的知覚がそれほど鋭敏でないノイズ様または周辺効果のような他のオブジェクトとは異なる仕方で扱われる必要があることがある。したがって、オブジェクトのクラスタリングを制御するために、空間的誤差に加えて、オブジェクト型（発話、効果、周辺音など）が重み付け因子として使われることができる。

【 0 0 7 1 】

クラスタリング・プロセス202は、このように、オブジェクトのある種の特性と、超えられることのできない定義された誤差量とに基づいて、オブジェクトをクラスターに組み合わせる。図3のAに示されるように、時間的にオブジェクト・グループ化を最適にするために、種々のまたは周期的な時間間隔でオブジェクト・グループをコンスタントに構築するために、クラスタリング・プロセス202は動的にオブジェクト・グループ203を再計算する。代替または組み合わせられたオブジェクト・グループは、構成要素オブジェクトのメタデータの組み合わせを表わす新たなメタデータ・セットと、構成要素オブジェクト・オーディオ信号の総和を表わすオーディオ信号とを表わす。図3のAに示される例は、組み合わせられたオブジェクト306が、特定の時点についてのもとのオブジェクト302および304を組み合わせることによって導出される場合を例示している。のちの時点において、組み合わせられたオブジェクトは、クラスタリング・プロセスによって実行される動的な処理に依存して、一つまたは複数の他のまたは異なるもとのオブジェクトを組み合わせることによって導出されることができる。

【 0 0 7 2 】

ある実施形態では、クラスタリング・プロセスは、10ミリ秒毎に一度または他の任意の適切な時間期間など、規則的な周期的間隔で、オブジェクトを解析し、クラスタリングを実行する。図5のAおよびBは、ある実施形態のもとでの、周期的な時間間隔の間にオブジェクトをクラスターにグループ化することを示している。特定の諸時点におけるオブジェクトの位置または場所をプロットする描画500に示されるように、さまざまなオブジェクトが任意の一つの時点において種々の位置に存在することがあり、それらのオブジェクトは、図5のAに示されるように異なる幅のものであることができる。図5のAにおい

10

20

30

40

50

て、オブジェクト O_3 は他のオブジェクトより大きい幅をもつように示されている。クラスタリング・プロセスは、定義された最大誤差閾値に関して、互いに十分に空間的に近い諸オブジェクトの諸グループを形成するために、オブジェクトを解析する。互いから誤差閾値502によって定義される距離以内分離したオブジェクトは、一緒にクラスタリングされる資格がある。よって、オブジェクト O_1 および O_3 はオブジェクト・クラスターA内に一緒にクラスタリングされることができ、オブジェクト O_4 および O_5 は異なるオブジェクト・クラスターB内に一緒にクラスタリングされることができる。これらのクラスターは、ある時刻（たとえば $T=0$ ミリ秒）におけるそれらのオブジェクトの相対位置に基づいて形成される。次の時間期間においては、それらのオブジェクトは、移動しているまたはメタデータ特性の一つまたは複数の点で変化していることがありうる。その場合、オブジェクト・クラスターは定義し直されてもよい。各オブジェクト・クラスターは、構成要素オブジェクトを、異なる波形とメタデータのセットで置き換える。このように、オブジェクト・クラスターAは、オブジェクト O_1 ないし O_3 のそれぞれについての個々の波形およびメタデータの代わりにレンダリングされる、波形とメタデータのセットを含む。

【0073】

図5のBは、次の時間期間（たとえば $\text{Time}=10$ ミリ秒）におけるオブジェクトの異なるクラスタリングを示している。描画550の例では、オブジェクト O_5 はオブジェクト O_4 から離れ、別のオブジェクト、オブジェクト O_6 に近い近傍内に移っている。この場合、オブジェクト・クラスターBは今ではオブジェクト O_5 ないし O_6 を含み、オブジェクト O_4 は脱クラスタリングされ、スタンドアローン・オブジェクトとしてレンダリングされる。他の因子も、オブジェクトが脱クラスタリングされたり、あるいはオブジェクトがクラスターを変えたりするようにすることがある。たとえば、オブジェクトの幅またはラウドネス（または他のパラメータ）がその近隣オブジェクトと比べて十分大きいまたは異なるようになることがあり、そうすれば、該オブジェクトはもはやそれらの近隣オブジェクトと一緒にクラスタリングされるべきではない。このように、図5のBに示されるように、オブジェクト O_3 が十分幅広になってもよく、オブジェクト・クラスターAから脱クラスタリングされて単独でレンダリングされる。図5のA～Bにおける水平軸が時間を表わすのではなく、視覚的な編成および議論のために複数のオブジェクトを空間的に分布させる次元として使われていることを注意しておくべきである。これらの描画のトップ全体が、全オブジェクトの時刻 t における瞬間またはスナップショットおよびそれらのオブジェクトがどのようにクラスタリングされるかを表わしている。

【0074】

図5のAないしBに示されるような時間期間毎にクラスタリングを実行する代わりに、クラスタリング・プロセスは、オブジェクトに関連するトリガー条件またはイベントに基づいてオブジェクトをクラスタリングしてもよい。一つのそのようなトリガー条件は、各オブジェクトについての開始および停止時刻である。図6Aないし6Cは、ある実施形態のもとでの、定義されたオブジェクト境界および誤差閾値との関係で、オブジェクトをクラスターにグループ化することを示している。閾ステップとして、各オブジェクトは、特定の時間期間内に定義される必要がある。さまざまな異なる方法が時間においてオブジェクトを定義するために使用されうる。ある実施形態では、オブジェクト開始/停止の時間的情報が、クラスタリング・プロセスのためにオブジェクトを定義するために使われることができる。この方法は、オーディオ・オブジェクトの開始点および停止点を定義する明示的な時間ベースの境界情報を利用する。あるいはまた、時間においてオブジェクトを定義するイベント境界を識別するために、聴覚的シーン解析技法が使用されることができる。そのような技法は、特許文献1において記述されている。該文献はここに参照によって組み込まれ、物件Bとして本明細書に添付される。検出された聴覚的シーン・イベント境界は、オーディオにおける知覚的な変化がある、時間において知覚的に有意な瞬間であり、これが、聴取者に聞こえないオーディオに対して変化がなされることができる、オーディオ内での「知覚的マスキング」を提供するために使用されることができる。

【0075】

図6Aないし6Cは、ある実施形態のもとでの、クラスタリング・プロセスを使ったオーディオ・オブジェクトのクラスタリングを制御するための、聴覚的シーン解析およびオーディオ・イベント検出または他の同様の方法の使用を示している。これらの図の例は、クラスターを定義し、定義された誤差閾値に基づいてオブジェクト・クラスターからオーディオ・オブジェクトを除去するために、検出された聴覚的イベントを使うことを概観している。図6Aは、特定の時刻(t)における空間的誤差のプロットにおけるオブジェクト・クラスターの生成を示す描画600である。二つのオーディオ・オブジェクト・クラスターがクラスターAおよびクラスターBと表わされ、オブジェクト・クラスターAが四つのオーディオ・オブジェクト O_1 ないし O_4 から構成され、オブジェクト・クラスターBが三つのオーディオ・オブジェクト O_5 ないし O_7 から構成される。描画600の縦方向の次元は空間的誤差を示し、これはある空間的オブジェクトがクラスタリングされるオブジェクトの残りのものからどのくらい似ていないかの指標であり、そのオブジェクトをクラスターから除去するために使われることができる。描画600には、さまざまな個々のオブジェクト O_1 ないし O_7 についての検出された聴覚的イベント境界604も示されている。各オブジェクトがオーディオ波形を表わすので、任意の所与の時点において、オブジェクトが検出された聴覚的イベント境界604をもつことが可能である。描画600に示されるように、時刻=tにおいては、オブジェクト O_1 および O_6 が、それらのオーディオ信号のそれぞれにおいて、検出された聴覚的イベント境界をもつ。図6A~6Cにおける横軸は時間を表わすのではなく、視覚的な編成および議論のために複数のオブジェクトを空間的に分布させる次元として使われていることを注意しておくべきである。この描画のトップ全体が、全オブジェクトの時刻tにおける瞬間またはスナップショットおよびそれらのオブジェクトがどのようにクラスタリングされるかを表わしている。

【0076】

図6Aに示されるように、空間的誤差閾値602がある。この値は、クラスターからオブジェクトを除去するために超過される必要がある誤差の大きさを表わす。すなわち、あるオブジェクトが、この誤差閾値602を超える量だけ潜在的なクラスター中の他のオブジェクトから離れていれば、そのオブジェクトはそのクラスターに含められない。このように、図6Aの例については、個々のオブジェクトのいずれも、閾値602によって示される空間的誤差閾値を超える空間的誤差をもたず、したがって、脱クラスタリングは起こらない。

【0077】

図6Bは、図6Aのクラスタリング例を時刻=t+Nにおいて示している。この時刻は、tより何らかの有限の時間だけ後であり、オブジェクト O_1 ないし O_3 および O_5 ないし O_7 については、各オブジェクトの空間的誤差がわずかに変化している。この例において、オブジェクト O_4 は、上記のあらかじめ定義された空間的誤差閾値622を超える空間的誤差をもつ。時刻=t+Nでは、聴覚的イベント境界はオブジェクト O_2 および O_4 について検出されていることを注意しておくべきである。このことは、時刻=t+Nにおいては、 O_4 についての波形におけるイベント境界によって作り出される知覚的マスキングが、当該オブジェクトがクラスターから除去されることを許容することを示している。オブジェクト O_4 はt<時刻<t+Nの間に空間的誤差閾値を超えたことがありうるが、聴覚的イベントは検出されなかったもので、該オブジェクトはオブジェクト・クラスターA内に残っていたことを注意しておく。この場合、クラスタリング・プロセスは、オブジェクト O_4 がクラスターAから除去される(脱クラスタリングされる)ようにする。図6Cに示されるように、オブジェクト・クラスターAからオブジェクト O_4 を除去した結果として、時刻=t+N+1において新たなオブジェクト・クラスタリング編成が生じる。この時点において、オブジェクト O_4 は、レンダリングされる単一のオブジェクトとして存在してもよいし、あるいは好適なクラスターがあれば別のオブジェクト・クラスター中に統合されてもよい。

【0078】

適応オーディオ・システムでは、ある種のオブジェクトは、固定されたオブジェクト、たとえば特定のスピーカー・フィールドに関連付けられているチャンネル・ベッドとして定義

10

20

30

40

50

されてもよい。ある実施形態では、クラスタリング・プロセスは、ベッドと動的オブジェクトの相互作用を考慮に入れ、オブジェクトがクラスタリングされたオブジェクトとグループ化されると大きすぎる誤差を生じる（たとえば、そのオブジェクトが外れているオブジェクトである）ときは、そのオブジェクトは代わりにあるベッドに混合される。図7は、ある実施形態のもとでの、オブジェクトおよびベッドをクラスタリングする方法を示すフローチャートである。図7に示される方法700では、ベッドは固定位置のオブジェクトとして定義されることが想定される。次いで、外れているオブジェクトは、該オブジェクトが他のオブジェクトとクラスタリングするための誤差閾値より上であれば、一つまたは複数の適切なベッドとクラスタリングされる（混合される）（工程702）。次いで、該ベッド・チャンネル（単数または複数）は、クラスタリング後に上記オブジェクト情報でラベル付けされる（工程704）。次いで、プロセスは、オーディオをより多くのチャンネルにレンダリングし、追加的チャンネルをオブジェクトとしてクラスタリングし（工程706）、アーチファクト/脱相関、位相歪みなどを避けるために下方混合またはスマート・ダウンミックスに対してダイナミックレンジ管理を実行する（工程708）。工程710では、プロセスは2パスの選別/クラスタリング・プロセスを実行する。ある実施形態では、これは、N個の最も顕著なオブジェクトを別個に保持し、残りのオブジェクトをクラスタリングすることに関わる。こうして、工程712では、プロセスは、それほど顕著でないオブジェクトのみをグループまたは固定されたベッドにクラスタリングする。固定されたベッドは、動いているオブジェクトまたはクラスタリングされたオブジェクトに加えられることができ、これは、ヘッドフォン仮想化のような個別的なエンドポイント装置にとってより好適でありうる。何個のオブジェクトが、そしてどのオブジェクトが一緒にクラスタリングされるかおよびどこでそれらがクラスタリング後に空間的にレンダリングされるかの特性として、オブジェクト幅が使われてもよい。

【0079】

ある実施形態では、オブジェクト信号ベースの顕著性（saliency）は、混合の平均スペクトルと、各オブジェクトのスペクトルとの間の差であり、顕著性メタデータ要素がオブジェクト/クラスターに追加されてもよい。相対ラウドネスは、各オブジェクトが最終的な混合に寄与するエネルギー/ラウドネスの割合である。相対ラウドネス・メタデータ要素もオブジェクト/クラスターに加えられることができる。本プロセスは次いで、マスクされる源を選別するおよび/または最も重要な諸源を保存するために顕著性によってソートすることができる。クラスターは、重要性が低い/顕著性が低い源をさらに減衰させることによって単純化されることができる。

【0080】

クラスタリング・プロセスは、一般に、オーディオ符号化に先立つデータ・レート削減のための手段として使われる。ある実施形態では、オブジェクト・クラスタリング/グループ化は、デコード中に、エンドポイント装置のレンダリング機能に基づいて使われる。完全な映画館再生環境、家庭シアター・システム、ゲーム・システムおよびパーソナル・ポータブル装置およびヘッドフォン・システムからの任意のものといったさまざまな異なるエンドポイント装置が、本稿に記載されるようなクラスタリング・プロセスを用いるレンダリング・システムとの関連で使用されうる。このように、レンダリングに先立って、レンダラーの機能を超過しないために、ブルーレイ・プレーヤーのような装置においてオブジェクトおよびベッドをデコードする間に、同じクラスタリング技法が利用されうる。一般に、オブジェクトおよびベッド・オーディオ・フォーマットのレンダリングは、各オブジェクトが、各オブジェクトの空間的情報の関数としてレンダラーに関連付けられたチャンネルの何らかの集合にレンダリングされることを要求する。このレンダリングの計算コストは、オブジェクトの数とともにスケールし、したがって、いかなるレンダリング装置も該レンダリング装置がレンダリングすることができるオブジェクトの何らかの最大数を持ち、該最大数は該レンダリング装置の計算機能の関数である。AVRのようなハイエンド・レンダラーは、多数のオブジェクトを同時にレンダリングできる高度なプロセッサを含むことがある。ボックス内家庭シアター（HTIB: home theater in a box）また

10

20

30

40

50

はサウンドバーのようなそれほど高価でない装置は、より限られたプロセッサのため、より少数のオブジェクトをレンダリングできることがある。したがって、レンダラーがデコーダに対して、自分が受け容れることができるオブジェクトおよびベッドの最大数を通信することが有利である。この数がデコードされたオーディオに含まれているオブジェクトおよびベッドの数より少ない場合には、デコーダは、総数を通信された最大まで減らすよう、レンダラーへの送信に先立って、オブジェクトおよびベッドのクラスタリングを適用してもよい。機能のこの通信は、内蔵ブルーレイ・プレーヤーを含んでいるHTIBのような単一の装置内での別個のデコードおよびレンダリングのソフトウェア・コンポーネント間で、あるいはスタンドアローンのブルーレイ・プレーヤーとAVRのような二つの別個の装置の間でHDMIのような通信リンクを通じて、行なわれうる。オブジェクトおよびクラスタ 10
ーに関連付けられたメタデータは、レンダラーによってクラスタ数を最適に削減するよう情報を指示または提供してもよい。それはたとえば、重要性の順序を列挙すること、クラスタの（相対的）重要性を信号伝達することまたはレンダリングされるべきクラスタの全体的な数を削減するためにどのクラスタが逐次的に組み合わせられるべきかを指定することによる。これについては、図15を参照して後述する。

【0081】

いくつかの実施形態では、クラスタリング・プロセスは、各オブジェクトに内在的な情報以外に何ら追加の情報なしに、デコーダ段206において実行されてもよい。しかしながら、このクラスタリングの計算コストは、節約しようとしているレンダリング・コスト 20
以上であることがありうる。より計算効率のよい実施形態は、計算資源がずっと大きいことがありうるエンコード側204で階層的なクラスタリング方式を計算し、どのようにしてオブジェクトおよびベッドを漸進的により少数にクラスタリングするかをデコーダに指示するメタデータをエンコードされたビットストリームとともに送ることに関わる。たとえば、メタデータは、まずオブジェクト2をオブジェクト10と併合せよ、第二に、結果として得られるオブジェクトをオブジェクト5と併合せよ、などと述べるものであってもよい。

【0082】

ある実施形態では、オブジェクトは、オブジェクト・トラック内に含まれるオーディオのある種の属性を表わすために該オブジェクトに関連付けられた一つまたは複数の時間変 30
化するラベルを有していてもよい。上記のように、オブジェクトは、ダイアログ、音楽、効果、背景などといったいくつかのディスクリートなコンテンツ型の一つにカテゴリー分けされてもよく、これらの型がクラスタリングを案内するのを助けるために使われてもよい。同時に、これらのカテゴリーはレンダリング・プロセスの間に有用であってもよい。たとえば、ダイアログ向上アルゴリズムは、ダイアログとラベル付けされたオブジェクトに対してのみ適用されうる。しかしながら、オブジェクトがクラスタリングされるときは、クラスタは複数の異なるラベルをもつオブジェクトから構成されることがありうる。クラスタにラベル付けするために、いくつかの技法を用いることができる。たとえば、最大量のエネルギーをもつオブジェクトのラベルを選択することによって、クラスタにつ 40
いての単一のラベルが選ばれてもよい。この選択も時間変化してもよい。その場合、単一のラベルがクラスタの継続期間中に規則的な時間間隔で選ばれ、各特定の区間において、ラベルがその特定の区間内で最大エネルギーをもつオブジェクトから選ばれる。場合によっては、単一のラベルでは十分でないことがあり、新しい、組み合わせられたラベルが生成されてもよい。たとえば、規則的な間隔で、その区間の間、クラスタに寄与するすべてのオブジェクトのラベルがクラスタに関連付けられてもよい。あるいはまた、これら寄与するラベルのそれぞれに重みが関連付けられてもよい。たとえば、重みは、その特定の型に属する全体的なエネルギーの割合に等しく設定されてもよい：たとえば、50%ダイアログ、30%音楽および20%効果。そのようなラベル付けは、その後、レンダラーによって、より柔軟な仕方で使用されうる。たとえば、ダイアログ向上アルゴリズムは、少なくとも50%ダイアログを含むクラスタリングされたオブジェクト・トラックに対してのみ適用 50
されうる。

【0083】

ひとたび種々のオブジェクトを組み合わせるクラスターが定義されたら、各クラスターについて等価なオーディオ・データが生成される必要がある。ある実施形態では、図3のAに示されるように、組み合わせられたオーディオ・データは単にクラスター中の各もとのオブジェクトについてのもとのオーディオ・コンテンツの和である。しかしながら、この単純な技法はデジタル・クリッピングにつながりうる。この可能性を緩和するために、いくつかの異なる技法が用いられることができる。たとえば、レンダラーが浮動オーディオ・データをサポートする場合、高ダイナミックレンジ (high dynamic range) 情報が記憶され、のちの処理段において使われるべく、レンダラーに渡されることができ。限られたダイナミックレンジしか利用可能でない場合には、結果として得られる信号を制限するまたは結果として得られる信号を固定でも動的でもよい何らかの量だけ減衰させることが望ましい。この後者の場合、減衰係数は動的利得としてオブジェクト・データ中に運び込まれる。場合によっては、構成要素信号の直接的な和は楕形フィルタリング・アーチファクトにつながる可能性がある。この問題は、和を取る前に脱相関フィルタまたは同様のプロセスを適用することによって緩和できる。下方混合に起因する音色変化を緩和するもう一つの方法は、和をとる前にオブジェクト信号の位相整列を使うことである。楕形フィルタリングまたは音色変化を解決するさらにもう一つの方法は、合計された信号のスペクトルおよび個々のオブジェクト信号のスペクトルに応答して、合計されたオーディオ信号に対して周波数依存重みを適用することによって、振幅またはパワー無償総和 (complementary summation) を施行し直すことである。

10

20

【0084】

下方混合を生成するとき、プロセスはさらに、データの圧縮を増すために、クラスターのビット深さを削減することができる。これは、ノイズ整形 (noise-shaping) または同様のプロセスを通じて実行されることができ。ビット深さ削減は、構成要素オブジェクトより少数のビットをもつクラスターを生成する。たとえば、一つまたは複数の24ビット・オブジェクトが16または20ビットとして表現されるクラスターにグループ化されることができ。クラスターの重要性またはエネルギーまたは他の因子に依存して、異なるクラスターおよびオブジェクトについて異なるビット削減方式が使用されてもよい。さらに、下方混合を生成するとき、結果として得られる下方混合信号が、固定数のビットを用いたデジタル表現によって表現できる受け容れ可能な範囲外のサンプル値をもつことがある。そのような場合、範囲外のサンプル値を防止するために、下方混合信号は、ピーク制限器を使って制限されたり、あるいはある量だけ (一時的に) 減衰されてもよい。適用された減衰の量はクラスター・メタデータに含められてもよく、そうすればレンダリング、符号化または他のその後のプロセスの際に取り消す (または逆にする) ことができる。

30

【0085】

ある実施形態では、クラスタリング・プロセスはポインタ機構を用いてもよい。それによれば、メタデータはデータベースまたは他の記憶に記憶されている特定のオーディオ波形へのポインタを含む。オブジェクトのクラスタリングは、組み合わせられたメタデータ要素によって適切な波形をポイントすることによって実行される。そのようなシステムは、オーディオ・コンテンツの事前計算されたデータベースを生成し、符号化器およびデコーダ段からオーディオ波形を送信し、次いでクラスタリングされた諸オブジェクトについての特定のオーディオ波形へのポインタを使ってデコード段においてクラスターを構築するアーカイブ・システムにおいて実装されることができ。この型の機構は、異なるエンドポイント装置のためのオブジェクト・ベースのオーディオのパッケージングを容易にするシステムにおいて使われることができる。

40

【0086】

クラスタリング・プロセスは、エンドポイント・クライアント装置上でクラスタリングし直すことを許容するよう適応されることもできる。一般には代替クラスターがもとのオブジェクトを置き換えるが、この実施形態については、クラスタリング・プロセスは、各オブジェクトに関連付けられた誤差情報をも送る。クライアントが、オブジェクトが個々

50

にレンダリングされたオブジェクトであるかクラスタリングされたオブジェクトであるかを判定できるようにするためである。誤差値が0であれば、クラスタリングがなかったことが推定できる。しかしながら、誤差値が何らかの量に等しければ、そのオブジェクトは何らかのクラスタリングの結果であることが推定できる。次いで、クライアントにおけるレンダリング判断は、誤差の大きさに基づくことができる。一般に、クラスタリング・プロセスはオフライン・プロセスとして実行される。あるいはまた、コンテンツが生成される際のライブ・プロセスとして実行されてもよい。この実施形態については、クラスタリング・コンポーネントは、コンテンツ作成および/またはレンダリング・システムの一部として提供されてもよいツールまたはアプリケーションとして実装されてもよい。

【0087】

知覚ベースのクラスタリング

ある実施形態では、クラスタリング方法は、制約された条件においてオブジェクトおよび/またはベッド・チャンネルを組み合わせるよう構成される。たとえば、入力オブジェクトは、オブジェクトの多さおよび/またはその空間的に疎な分布のために、空間的な誤差基準を破ることなくしてはクラスタリングされることができない。そのような条件では、クラスタリング・プロセスは、(メタデータから導出される)空間的近接性によって制御されるばかりでなく、対応するオーディオ信号導出された知覚的基準によって補強される。より具体的には、コンテンツ中の高い(知覚される)重要性をもつオブジェクトは、空間的誤差を最小化することに関して、低い重要性をもつオブジェクトに対して優遇される。重要性を定量化することの例は、部分ラウドネスおよびセマンティクス(コンテンツ型)を含むがそれに限られない。

【0088】

図8は、ある実施形態のもとでの、空間的近接性に加えて知覚的重要性に基づいてオブジェクトおよびベッド・チャンネルをクラスターにクラスタリングするシステムを示している。図8に示されるように、システム360は前処理ユニット366と、知覚的重要性コンポーネント376と、クラスタリング・コンポーネント384とを有している。チャンネル・ベッドおよびまたはオブジェクト364は関連付けられたメタデータ362とともに、前処理ユニット366に入力されて、それらの相対的な知覚的重要性を決定するために処理され、次いで他のベッド/オブジェクトとクラスタリングされて、出力ベッドおよび/またはオブジェクトのクラスター(これは単独オブジェクトからなってもよく、あるいはオブジェクトの集合からなってもよい)を、これらのクラスターについての関連付けられたメタデータ390とともに、生成する。ある例示的な実施形態または実装では、入力は11.1ベッド・チャンネルおよび128以上のオーディオ・オブジェクトからなってもよく、出力は合計11~15個のオーダーの信号を各クラスターについての関連付けられたメタデータとともに含むクラスターおよびベッドの集合を含んでいてもよい。ただし、実施形態はこれに限定されるものではない。メタデータは、オブジェクト位置、サイズ、ゾーン・マスク、脱相関器フラグ、スナップ・フラグなどを指定する情報を含んでいてもよい。

【0089】

前処理ユニット366は、他にもあるコンポーネントの中でも、メタデータ処理器368、オブジェクト脱相関ユニット377、オフライン処理ユニット372および信号セグメンテーション・ユニット374のような個々の機能コンポーネントを含んでいてもよい。メタデータ出力更新レート396のような外部データが前処理器366に提供されてもよい。知覚的重要性コンポーネント376は、他にもあるコンポーネントの中でも、重心初期化コンポーネント378、部分ラウドネス・コンポーネント380およびメディア・インテリジェンス・ユニット382を有する。出力ベッドおよびオブジェクト構成データ398のような外部データが知覚的重要性コンポーネント376に提供されてもよい。クラスタリング・コンポーネント384は、信号併合386およびメタデータ併合388コンポーネントを有する。これらのコンポーネントは、クラスタリングされたベッド/オブジェクトを形成して、組み合わせられたベッド・チャンネルおよびオブジェクトについてのメ

10

20

30

40

50

タデータ390およびクラスター392を生成する。

【0090】

部分ラウドネス (partial loudness) に関し、オブジェクトの知覚されるラウドネスは、通例、他のオブジェクトのコンテキストにおいて低下する。たとえば、オブジェクトは、シーン内に存在する他のオブジェクトおよび/またはベッド・チャネルによって (部分的に) マスクされることがある。ある実施形態では、高い部分ラウドネスをもつオブジェクトが、空間的誤差最小化に関し、低い部分ラウドネスをもつオブジェクトより優遇される。このように、相対的にマスクされていない (すなわち、知覚的にラウドネスがより大きい) オブジェクトはクラスタリングされる可能性が低くなり、一方、相対的にマスクされているオブジェクトはクラスタリングされる可能性がより高くなる。このプロセスは、好ましくは、マスキングの空間的側面を含む。たとえば、マスクされるオブジェクトとマスクするオブジェクトが異なる空間的属性をもつ場合にマスキングからの解放を含む。換言すれば、ある関心オブジェクトのラウドネスに基づく重要性は、そのオブジェクトが他のオブジェクトから空間的に離れているときは、他のオブジェクトが関心オブジェクトの直近にあるときに比べ、より高くなる。

10

【0091】

ある実施形態では、オブジェクトの部分ラウドネス (partial loudness) は空間的マスキング解除現象をもって拡張された個別ラウドネス (specific loudness) を含む。下記の式で与えられるように、二つのオブジェクトの間の空間的距離に基づくマスキングの量を表現するために、マスキングからのバイノーラル解放が導入される：

20

$$N'_k(b) = (A + E_m(b)) + (A + E_m(b)(1 - f(k,m)))$$

。

【0092】

上式において、最初の和はすべてのmについて実行され、二番目の和はすべてのm kについて実行される。項 $E_m(b)$ はオブジェクトmの励起を表わし、項Aは絶対聴覚閾値 (absolute hearing threshold) を反映し、項 $(1 - f(k,m))$ はマスキングからの解放を表わす。この式に関するさらなる詳細は、下記で論じられる。

【0093】

コンテンツ・セマンティクスまたはオーディオ型に関し、ダイアログはしばしば背景音楽、周辺音、効果または他の型のコンテンツより重要である (またはより注意を引く) と考えられる。したがって、オブジェクトの重要性は、その (信号) コンテンツに依存し、相対的に重要でないオブジェクトは重要なオブジェクトよりも、クラスタリングされる可能性が高い。

30

【0094】

オブジェクトの知覚的重要性は、オブジェクトの知覚されるラウドネスおよびコンテンツ重要性を組み合わせることによって導出されることができる。たとえば、ある実施形態では、コンテンツ重要性は、ダイアログ信頼スコアに基づいて導出されることができ、この導出されたコンテンツ重要性に基づいて利得値 (dB単位) が推定されることができる。次いで、オブジェクトのラウドネスまたは励起は、推定されたラウドネスによって修正されることができ、修正されたラウドネスはオブジェクトの最終的な知覚的重要性を表わす

40

。

【0095】

図9は、ある実施形態のもとでの、知覚的重要性を使ったオブジェクト・クラスタリング・プロセスの機能コンポーネントを示している。描画900に示されるように、入力オーディオ・オブジェクト902はクラスタリング・プロセス904を通じて出力クラスター910に組み合わされる。クラスタリング・プロセス904は、少なくとも部分的には、オブジェクト信号および任意的にはそのパラメトリックなオブジェクト記述から生成される重要性メトリック908に基づいて、オブジェクト902をクラスタリングする。これらのオブジェクト信号およびパラメトリックなオブジェクト記述は、クラスタリング・プロセス904が使うための重要性メトリック908を生成する重要性推定906機能に

50

入力される。出力クラスター 910 は、もとの入力オブジェクト構成よりコンパクトな表現（たとえば、より少数のオーディオ・チャンネル）をなし、こうして、記憶および伝送要件の低減ならびに、特に限られた処理機能をもつおよび/またはバッテリーで動作する消費者ドメイン装置上でのコンテンツの再現のための計算およびメモリ要件の低減を許容する。

【0096】

ある実施形態では、重要性推定 906 およびクラスタリング 904 のプロセスは時間の関数として実行される。この実施形態については、入力オブジェクト 900 のオーディオ信号は、ある解析コンポーネントにかけられる個々のフレームにセグメント分解される。そのようなセグメント分解は、時間領域波形に対して適用されてもよいが、フィルタバンクまたは他の任意の変換領域を使って適用されてもよい。重要性推定機能 906 は、コンテンツ型および部分ラウドネスを含む入力オーディオ・オブジェクト 902 の一つまたは複数の特性に基づいて機能する。

【0097】

図 11 は、ある実施形態のもとでの、コンテンツ型およびラウドネスという知覚的因子に基づいてオーディオ・オブジェクトを処理する全体的な方法を示すフローチャートである。方法 1100 の全体的な諸工程は、入力オブジェクトのコンテンツ型を推定し（1102）、次いで、コンテンツ・ベースのオブジェクトの重要性を推定する（1104）ことを含む。ブロック 1106 に示されるように、オブジェクトの部分ラウドネスが計算される。部分ラウドネスは、システム構成に依存して、コンテンツ分類と並行して、あるいはコンテンツ分類の前または後に計算されることができる。ラウドネス指標およびコンテンツ解析は次いで組み合わされて（1108）、ラウドネスおよびコンテンツに基づく全体的な重要性を導出する。これは、オブジェクトの計算されたラウドネスを、そのオブジェクトがコンテンツに起因して知覚的に重要である確率によって修正することによってなされてもよい。ひとたび組み合わされたオブジェクト重要性が決定されたら、オブジェクトは、ある種のクラスタリング・プロセスに依存して、他のオブジェクトとクラスタリングされるまたはクラスタリングされないことができる。ラウドネスに基づくオブジェクトの過度のクラスタリングおよび非クラスタリングを防止するために、コンテンツ重要性に基づいてラウドネスをなめらかにする平滑化動作が使われてもよい（1110）。ラウドネス平滑化に関し、オブジェクトの相対的重要性に基づいて時定数が選択される。重要なオブジェクトについては、ゆっくり平滑化する大きな時定数が選択されることができ、それにより重要なオブジェクトは一貫して、クラスター重心として選択されることができ、コンテンツ重要性に基づいて適応的な時定数が使われてもよい。オブジェクトの平滑化されたラウドネスおよびコンテンツ重要性が次いで、適切な出力クラスターを形成するために使われる（1112）。方法 600 に示された主たるプロセス工程の各工程の諸側面は、下記でより詳細に述べる。システム制約条件および用途の要件に依存して、必要であれば、プロセス 1100 のある種の工程が省略されてもよいことを注意しておくべきである。たとえば、知覚的重要性をコンテンツ型または部分ラウドネスのうち的一方のみに基づくようにすることがありうる基本的なシステムや、ラウドネス平滑化を要求しないものである。

【0098】

オブジェクト・コンテンツ型の推定（1102）に関し、コンテンツ型（たとえばダイアログ、音楽およびサウンド効果）は、オーディオ・オブジェクトの重要性を示すための枢要な情報を提供する。たとえば、ダイアログは通例、ストーリーを伝えるので、映画における最も重要な構成要素であり、適正な再生は典型的には、ダイアログが他の動いているオーディオ・オブジェクトと一緒に動き回ることを許容しないことを要求する。図 9 における重要性推定機能 906 は、オーディオ・オブジェクトがダイアログであるか否かを、あるいは重要なまたは重要でない型のオブジェクトの他の何らかの型を判定するためにオーディオ・オブジェクトのコンテンツ型を自動的に推定するオーディオ分類コンポーネントを含んでいる。

【 0 0 9 9 】

図 10 は、ある実施形態のもとでの、オーディオ分類コンポーネントの機能図である。描画 1000 に示されるように、入力オーディオ信号 1002 は、入力オーディオ信号の時間的、スペクトルのおよび / または空間的属性を表わす特徴を抽出する特徴抽出モジュールにおいて処理される。各目標オーディオ型の統計的な属性を表わす事前トレーニングされたモデル 1006 の集合も提供される。図 10 の例については、モデルはダイアログ、音楽、サウンド効果およびノイズを含むが、他のモデルも可能であり、モデル・トレーニングのためにはさまざまな機械学習技法が適用されることができる。モデル情報 1006 および抽出された特徴 1004 はモデル比較モジュール 1008 に入力される。このモジュール 1008 は入力オーディオ信号の特徴を各目標オーディオ型のモデルと比較し、各目標オーディオ型の信頼スコアを計算し、最良一致した諸オーディオ型を推定する。各目標オーディオ型についての信頼スコアがさらに推定される。これは、識別されるべきオーディオ・オブジェクトと目標オーディオ型との間の確率または一致レベルを表わし、0 から 1 (または他の任意の適切な範囲) の値をもつ。信頼スコアは、種々の機械学習方法に依存して計算されることができる。たとえば、ガウシアン混合モデル (GMM: Gaussian Mixture Model) については事後確率が直接、信頼スコアとして使われることができ、サポートベクターマシン (SVM: Support Vector Machine) およびエイダブースト (AdaBoost) については信頼値を近似するためにシグモイド当てはめが使われることができる。他の同様の機械学習方法も使用できる。モデル比較モジュール 1008 の出力 1010 は、入力オーディオ信号 1002 についてオーディオ型 (単数または複数) およびその関連付けられた信頼スコア (単数または複数) を含む。

10

20

【 0 1 0 0 】

コンテンツ・ベースのオーディオ・オブジェクト重要性を推定することに関し、ダイアログ指向の用途のためには、上記のようにオーディオ中でダイアログが最も重要な成分であると想定して、コンテンツ・ベースのオーディオ・オブジェクト重要性は、ダイアログ信頼スコアのみに基づいて計算される。他の用途では、コンテンツの好まれる型に依存して、種々のコンテンツ型信頼スコアが使用されうる。ある実施形態では、下記の式で与えられるようなシグモイド関数が利用される：

【 数 1 】

$$I_k = \frac{1}{1 + e^{Ap_k + B}}$$

上式において、 I_k はオブジェクト k の推定されたコンテンツ・ベースの重要性であり、 p_k はオブジェクト k が発話 / ダイアログからなることの対応する推定される確率であり、 A および B は二つのパラメータである。

【 0 1 0 1 】

閾値 c より小さいダイアログ確率スコアをもつものについてはコンテンツ・ベースの重要性を一貫して 0 に近くさらに設定するために、上記の公式は次のように修正できる：

【 数 2 】

$$I_k = \frac{1}{1 + e^{A \cdot \max(p_k - c, 0) + B}}$$

ある実施形態では、定数 c は $c = 0.1$ の値を取ることができ、二つのパラメータ A および B は定数であるまたは確率スコア p_k に基づいて適応的に調整されることができる。

【 0 1 0 2 】

オブジェクト部分ラウドネスを計算することに関し、複雑な聴覚的シーンにおけるある

30

40

50

オブジェクトの部分ラウドネスを計算する一つの方法は、臨界帯域 (b) における励起レベル $E(b)$ の計算に基づく。ある関心オブジェクトについての励起レベル $E_{obj}(b)$ および残りすべての (マスキング) 信号の励起 $E_{noise}(b)$ は結果として、次式で与えられるような、帯域 b における個別ラウドネス (specific loudness) $N'(b)$ を与える：

$$N'(b) = C [(GE_{obj} + GE_{noise} + A) - A] - C [(GE_{noise} + A) - A]$$

ここで、G、C、A および E はモデル・パラメータである。その後、部分ラウドネス (partial loudness) N は、諸臨界帯域を通じて個別ラウドネス $N'(b)$ を合計することによって次のように得られる：

$$N = \sum_b N'(b)$$

10

【 0 1 0 3 】

聴覚的シーンが励起レベル $E_k(b)$ をもつ K 個のオブジェクトからなるとき ($k = 1, \dots, K$)、記法の簡単のため、モデル・パラメータ G および C が +1 に等しいとすると、オブジェクト k の個別ラウドネス $N'_k(b)$ は

$$N'_k(b) = (A + \sum_m E_m(b)) - (-E_k(b) + A + \sum_m E_m(b))$$

によって与えられる。

【 0 1 0 4 】

上式の第一項は聴覚的シーンの全体的な励起に、絶対聴覚閾値を反映する励起 A を加えたものを表わす。第二項は関心オブジェクト k を除いた全体的な励起を反映し、よって、第二項はオブジェクト k に適用される「マスキング項」として解釈されることができる。この定式化は、マスキングからのバイノーラル解放を考慮しない。マスキングからの解放は、次式によって与えられるように、関心オブジェクト k が別のオブジェクト m から遠方である場合に上記のマスキング項を低減させることによって組み込まれることができる：

20

$$N'_k(b) = (A + \sum_m E_m(b)) - (-E_k(b) + A + \sum_m E_m(b)(1 - f(k, m)))$$

【 0 1 0 5 】

上式において、 $f(k, m)$ は、オブジェクト k およびオブジェクト m が同じ位置をもつ場合には 0 に等しく、オブジェクト k と m の間の空間的距離の増大とともに +1 まで増大する値に等しい関数である。異なる言い方をすれば、関数 $f(k, m)$ はオブジェクト k および m のメトリック位置における距離の関数としてマスキング解除の量を表わす。あるいはまた、 $f(k, m)$ の最大値は、空間的に離れているオブジェクトについての空間的マスキング解除の量における上限を反映するために 0.995 のような +1 よりわずかに小さい値に制限されてもよい。

30

【 0 1 0 6 】

ラウドネスの計算は、定義されたクラスター重心によって考慮に入れられることができる。一般に、重心は、クラスターの中心を表わす属性空間における位置であり、属性は、測定 (たとえば、ラウドネス、コンテンツ型など) に対応する値のセットである。個々のオブジェクトの部分ラウドネスは、オブジェクトがクラスターリングされている場合および目標が、可能な最良のオーディオ品質を与える、クラスターおよび関連付けられたメトリック位置の制約されたセットを導出することである場合には、限られた有意性しかない。ある実施形態では、より典型的なメトリックは、特定のクラスター位置 (または重心) によって考慮に入れられる、その位置の近傍における全励起を総合する部分ラウドネスである。上記の場合と同様に、クラスター重心 c によって考慮に入れられる部分ラウドネスは次のように表現できる：

40

$$N'_c(b) = (A + \sum_m E_m(b)) - (A + \sum_m E_m(b)(1 - f(k, m)))$$

【 0 1 0 7 】

このコンテキストにおいて、出力ベッド・チャンネル (たとえば、再生システムにおける特定のラウドスピーカーによって再現されるべき出力チャンネル) は、目標ラウドスピーカーの位置に対応する固定した位置をもつ重心と見なすことができる。同様に、入力ベッド

50

信号は、対応する再生ラウドスピーカーの位置に対応する位置をもつオブジェクトと見なすことができる。よって、オブジェクトおよびベッド・チャンネルは、ベッド・チャンネル位置が固定されているという制約条件のもとで、厳密に同じ解析にかけられることができる。

【0108】

ある実施形態では、ラウドネスおよびコンテンツ解析データは、図11のブロック1108に示されるように、組み合わされたオブジェクト重要性値を導出するために組み合わされる。部分ラウドネスおよびコンテンツ解析に基づくこの組み合わされた値は、オブジェクトのラウドネスおよび/または励起を、そのオブジェクトが知覚的に重要である確率によって修正することによって得られる。たとえば、オブジェクトkの励起は次のように修正されることができる：

$$E'_k(b) = E_k(b)g(l_k)$$

【0109】

上式において、 l_k はオブジェクトkのコンテンツ・ベースのオブジェクト重要性であり、 $E'_k(b)$ は修正された励起レベルであり、 $g(\cdot)$ はコンテンツ重要性を励起レベル修正にマッピングする関数である。ある実施形態では、 $g(\cdot)$ は、コンテンツ重要性をdb単位での利得と解釈する指数関数である。

【0110】

$$g(l_k) = 10^{G \cdot l_k}$$

ここで、Gはコンテンツ・ベースのオブジェクト重要性に対するもう一つの利得であり、これは最良のパフォーマンスを得るために調整されることができる。

【0111】

もう一つの実装では、 $g(\cdot)$ は：

$$g(l_k) = 1 + G \cdot l_k$$

のような線形関数である。

【0112】

上記の式は単に可能な実施形態の例である。代替的な方法は、励起の代わりにラウドネスに適用されることができ、単純な積に関わる以外の情報の組み合わせ法を含んでもよい。

【0113】

図11にも示されるように、諸実施形態は、コンテンツ重要性(1110)に基づいてラウドネスを平滑化する方法をも含む。ラウドネスは、通例、オブジェクト位置の急速な変化を避けるために、諸フレームわたって平滑化される。平滑化プロセスの時定数は、コンテンツ重要性に基づいて適応的に調節されることができる。このようにして、より重要なオブジェクトについては、時定数はより大きくなる(ゆっくり平滑化する)ことができ、それにより、より重要なオブジェクトは、諸フレームにわたって、一貫して、クラスター重心として選択されることができる。ダイアログは通例、話された言葉と合間を交互するので、これは、ダイアログについての重心選択の安定性をも改善する。ここで、合間にはラウドネスは低いことがあるので、他のオブジェクトが重心として選択されることになる。その結果として、最終的に選択された重心がダイアログと他のオブジェクトとの間で切り替わることになり、よって潜在的な不安定性を引き起こす。

【0114】

ある実施形態では、時定数はコンテンツ・ベースのオブジェクト重要性と

$$= \tau_0 + l_k \cdot \tau_1$$

のように正の相関をもつ。

【0115】

上式では、 τ_0 は推定された重要性依存の時定数であり、 τ_0 および τ_1 はパラメータである。さらに、コンテンツ重要性に基づく励起/ラウドネス・レベル修正と同様に、適応的な時定数方式は、ラウドネスまたは励起のいずれかに対して適用されることもできる。

【 0 1 1 6 】

上記のように、オーディオ・オブジェクトの部分ラウドネスは、定義されたクラスター重心に関して計算される。ある実施形態では、クラスター重心計算は、クラスターの総数が制約されているときに、諸重心の最大部分ラウドネスを考慮に入れる、クラスター重心の部分集合が選択されるよう、実行される。図 1 2 は、ある実施形態のもとでの、クラスター重心を計算し、オブジェクトを選択された重心に割り当てるプロセスを示すフローチャートである。プロセス 1 2 0 0 は、オブジェクト・ラウドネス値に基づいて重心の限られたセットを導出する実施形態を示している。本プロセスは、該限られたセットにおける重心の最大数を定義することによって始まる (1 2 0 1)。これは、空間的誤差のようなある種の基準が破られないよう、オーディオ・オブジェクトのクラスタリングを制約する。各オーディオ・オブジェクトについて、本プロセスは、そのオブジェクトの位置における重心が与えられたときに考慮に入れられるラウドネスを計算する (1 2 0 2)。本プロセスは次いで、最大ラウドネスを考慮に入れる重心であって、任意的にはコンテンツ型について修正されたものを選択し (1 2 0 4)、選択された重心によって考慮に入れられる全励起を除去する (1 2 0 6)。このプロセスは、判断ブロック 1 2 0 8 において判定されるところによりブロック 1 2 0 1 において定義された重心の最大数が得られるまで、繰り返される。

10

【 0 1 1 7 】

代替的な実施形態では、ラウドネス処理は、空間領域においてすべての可能な位置のサンプリングに対してラウドネス解析を実行し、続いてすべての位置にわたって極大を選択することに関わることができる。あるさらなる代替的な実施形態では、ホッホバウム (Hochbaum) 重心選択がラウドネスで増強される。ホッホバウム重心選択は、互いに対する最大距離をもつ位置のセットの選択に基づく。このプロセスは、重心を選択するための距離メトリックにラウドネスを乗算または加算することによって増強されることができる。

20

【 0 1 1 8 】

図 1 2 に示されるように、ひとたび上記最大数の重心が処理されたら、オーディオ・オブジェクトは適切な選択された重心に割り当てられる (1 2 1 0)。この方法のもとでは、クラスター重心の適正な部分集合が選択されたら、オブジェクトは、そのオブジェクトを最も近い近隣の重心に加えるまたはそのオブジェクトを重心のセットまたは部分集合中に混合することによって、重心に割り当てられることができる。それはたとえば、三角形分割、ベクトル分解の使用またはそのオブジェクトの空間的誤差を最小にするための他の任意の手段による。

30

【 0 1 1 9 】

図 1 3 の A および B は、ある実施形態のもとでの、ある種の知覚的基準に基づく、オブジェクトのクラスターへのグループ分けを示している。描画 1 3 0 0 は、X/Y 空間座標系として表わされる二次元オブジェクト空間における種々のオブジェクトの位置を示す。オブジェクトの相対サイズはそれらの相対的な知覚的重要性を表わし、より大きなオブジェクト (たとえば 1 3 0 6) はより小さなオブジェクト (たとえば 1 3 0 4) より高い重要性があるようになっている。ある実施形態では、知覚的重要性は、それぞれのオブジェクトの相対的な部分ラウドネス値およびコンテンツ型に基づく。クラスタリング・プロセスは、より大きな空間的誤差を許容する諸クラスター (オブジェクトの諸グループ) を形成するためにオブジェクトを解析する。ここで、空間的誤差は、最大誤差閾値 1 3 0 2 との関係で定義されうる。誤差閾値、クラスターの最大数および他の同様の基準のような適切な基準に基づいて、オブジェクトはいくつもある配置でクラスタリングされうる。

40

【 0 1 2 0 】

図 1 3 の B は、クラスタリング基準のある特定のセットについて、図 1 3 の A のオブジェクトのある可能なクラスタリングを示している。描画 1 3 5 0 は、描画 1 3 0 0 における七つのオブジェクトの、クラスター A~D と表わされる四つの別個のクラスターへのクラスタリングを示している。図 1 3 の B に示される例については、クラスター A は、より大きな空間的誤差を許容する低重要性オブジェクトの組み合わせを表わし；クラスター C お

50

よびDは別個にレンダリングされるべきであるほど高い重要性がある源に基づくクラスターであり；クラスターBは、低重要性オブジェクトが高重要性オブジェクトとグループ化されることが出来る場合を表わしている。図13のBの構成は、図13のAのオブジェクトについてのある可能なクラスタリング方式のほんの一例を表わすことが意図されており、多くの異なるクラスタリング配置が選択されることが出来る。

【0121】

ある実施形態では、クラスタリング・プロセスは、オブジェクトをクラスタリングするために、X/Y平面内でn個の重心を選択する。ここで、nはクラスター数である。本プロセスは、最高の重要性または考慮される最大ラウドネスに対応するn個の重心を選択する。次いで、残りのオブジェクトは(1)最も近い近隣重心または(2)パン技法によるクラスタ重心へのレンダリングに従ってクラスタリングされる。このように、オーディオ・オブジェクトは、クラスタリングされるオブジェクトのオブジェクト信号を最も近い重心に加えるまたは該オブジェクト信号をクラスターの(サブ)セットに混合することによって、クラスターに割り当てられることができる。選択されるクラスターの数は、動的であってもよく、クラスター中の空間的誤差を最小にする混合利得を通じて決定されてもよい。クラスター・メタデータは、クラスターに存在するオブジェクトの重み付けされた平均からなる。重みは、知覚されるラウドネスならびにオブジェクト位置、サイズ、ゾーン、排除マスク(exclusion mask)および他のオブジェクト特性に基づいていてもよい。一般に、オブジェクトのクラスタリングは、主として、オブジェクト重要性に依存してもよく、一つまたは複数のオブジェクトは複数の出力クラスターにわたって分散されてもよい。すなわち、オブジェクトは一つのクラスターに加えられてもよく(一意的にクラスタリングされる)、あるいは二つ以上のクラスターにわたって分配されてもよい(非一意的にクラスタリングされる)。

【0122】

図13のAおよびBに示されるように、クラスタリング・プロセスは、もとの数のオーディオ・オブジェクトおよび/またはベッド・チャンネルを目標数の新しい等価なオブジェクトおよびベッド・チャンネルに動的にグループ化する。たいていの実用的な用途では、目標数はもとの数より実質的に少ない。たとえば、100個のもとの入力トラックが20個以下の組み合わせられたグループに組み合わせられる。これらの解決策は、ベッドおよびオブジェクト・チャンネルの両方がクラスタリング・プロセスに対して入力および/または出力として利用可能であるシナリオに当てはまる。オブジェクトおよびベッド・トラックの両方をサポートする第一の解決策は、入力ベッド・トラックを、空間内の固定したあらかじめ定義された位置をもつオブジェクトとして処理するというものである。これは、システムが、たとえばオブジェクトおよびベッドの両方を含むシーンを、目標数のオブジェクト・トラックのみに単純化することを許容する。しかしながら、クラスタリング・プロセスの一部として、出力ベッド・トラックの数を保存することが望ましいこともありうる。その場合、より重要でないオブジェクトは、前置プロセスとして、ベッド・トラックに直接レンダリングされることができ、一方、最も重要な諸オブジェクトは、より少ない目標数の等価なオブジェクト・トラックにさらにクラスタリングされることが出来る。結果として得られるクラスターのいくつかが高い歪みをもつ場合、それらのクラスターは、後置プロセスとしてベッドにレンダリングされることもできる。このほうがもとのコンテンツのよりよい近似につながりうるからである。誤差/歪みは時間変化する関数なので、この決定は、時間変化する仕方になされることが出来る。

【0123】

ある実施形態では、クラスタリング・プロセスは、すべての個々の入力トラック(オブジェクトまたはベッド)のオーディオ・コンテンツおよび付属のメタデータ(たとえばオブジェクトの空間的位置)を解析して、所与の誤差メトリックを最小にする等価な数の出力オブジェクト/ベッド・トラックを導出することに関わる。基本的な実装では、誤差メトリック1302は、クラスタリングされるオブジェクトをシフトさせることに起因する空間的歪みに基づき、時間を追った各オブジェクトの重要性の指標によってさらに重み付

けされることができる。オブジェクトの重要性は、ラウドネス、コンテンツ型および他の有意な因子といったオブジェクトの他の特性を表わすことができる。あるいはまた、これら他の因子は、空間的な誤差メトリックと組み合わされることのできる別個の誤差メトリックを形成することができる。

【0124】

オブジェクトおよびチャンネル処理

適応オーディオ・システムでは、ある種のオブジェクトは固定されたオブジェクト、たとえば特定のスピーカー・フィールドに関連付けられているチャンネル・ベッドとして定義されてもよい。ある実施形態では、クラスタリング・プロセスは、ベッドと動的オブジェクトの相互作用を考慮に入れ、オブジェクトがクラスタリングされたオブジェクトとグループ化されるときに大きすぎる誤差を生じる（たとえば、そのオブジェクトが外れているオブジェクトである）ときは、そのオブジェクトは代わりにあるベッドに混合される。図14は、ある実施形態のもとでの、オーディオ・オブジェクトおよびベッドをクラスタリングするプロセス・フローの構成要素を示している。図14に示される方法1400では、ベッドは固定位置のオブジェクトとして定義されることが想定される。次いで、外れているオブジェクトは、該オブジェクトが他のオブジェクトとクラスタリングするための誤差閾値より上であれば、一つまたは複数の適切なベッドとクラスタリングされる（混合される）（1402）。次いで、該ベッド・チャンネル（単数または複数）は、クラスタリング後に上記オブジェクト情報でラベル付けされる（1404）。次いで、プロセスは、オーディオをより多くのチャンネルにレンダリングし、追加的チャンネルをオブジェクトとしてクラスタリングし（1406）、アーチファクト/脱相関、位相歪みなどを避けるために下方混合またはスマート・ダウンミックスに対してダイナミックレンジ管理を実行する（1408）。本プロセスは2パスの選別/クラスタリング・プロセスを実行する（1410）。ある実施形態では、これは、N個の最も顕著なオブジェクトを別個に保持し、残りのオブジェクトをクラスタリングすることに関わる。こうして、本プロセスは、それほど顕著でないオブジェクトのみをグループまたは固定されたベッドにクラスタリングする（1412）。固定されたベッドは、動いているオブジェクトまたはクラスタリングされたオブジェクトに加えられることができ、これは、ヘッドフォン仮想化のような個別的なエンドポイント装置にとってより好適でありうる。何個のオブジェクトが、そしてどのオブジェクトが一緒にクラスタリングされるかおよびどこでそれらがクラスタリング後に空間的にレンダリングされるかの特性として、オブジェクト幅が使われてもよい。

【0125】

再生システム

上記で論じたように、さまざまな異なるエンドポイント装置が、本稿に記載されるようなクラスタリング・プロセスを用いるレンダリング・システムとの関連で使用されてもよく、そのような装置はクラスタリング・プロセスに影響しうるある種の機能を有していてもよい。図15は、ある実施形態のもとでの、エンドポイント装置機能に基づく、クラスタリングされたデータ・のレンダリングを示している。描画1500に示されるように、ブルーレイ・ディスク・デコーダ1502は、サウンドバー、家庭シアター（home theater）・システム、個人用再生装置または他の何らかの制限された処理再生システム1504を通じたレンダリングのために、クラスタリングされたベッドおよびオブジェクトを含む単純化されたオーディオ・シーン・コンテンツを生成する。エンドポイント装置の特性および機能は、レンダラー機能情報1508として、デコーダ段1502に送信し返される。オブジェクトのクラスタリングが、使用される特定のエンドポイント装置に基づいて最適に実行されることができるようにするためである。

【0126】

本クラスタリング・プロセスの諸側面を用いる適応オーディオ・システムは、一つまたは複数の捕捉、前処理、オーサリングおよび符号化コンポーネントを通じて生成されるオーディオ・コンテンツをレンダリングおよび再生するよう構成されている再生システムを有していてもよい。適応オーディオ前処理器は、入力オーディオの解析を通じて適切なメ

タデータを自動的に生成する源分離およびコンテンツ型検出機能を含んでいてもよい。たとえば、位置メタデータは、チャンネル対間の関連した入力相対的なレベルの解析を通じて多チャンネル記録から導出されてもよい。発話または音楽のようなコンテンツ型の検出は、たとえば、特徴抽出および分類によって達成されてもよい。ある種のオーサリング・ツールは、サウンド・エンジニアの創造的な意図の入力および符号化を最適化することによって、オーディオ・プログラムをオーサリングすることを許容し、サウンド・エンジニアが、事実上任意の再生環境における再生のために最適化されている最終的なオーディオ・ミックスを一度で生成することを許容する。これは、オーディオ・オブジェクトおよびもとのオーディオ・コンテンツに関連付けられておりもとのオーディオ・コンテンツと一緒にエンコードされる位置データの使用を通じて達成されることのできる。聴衆席のまわり
10
に音を正確に配置するために、サウンド・エンジニアは、再生環境の実際の制約条件および特徴に基づいて、音が最終的にどのようにレンダリングされるかに対する制御を必要とする。適応オーディオ・システムは、サウンド・エンジニアが、オーディオ・コンテンツがどのようにデザインされ、オーディオ・オブジェクトおよび位置データの使用を通じて混合されるかを変えることを許容することによって、これを提供する。ひとたび適応オーディオ・コンテンツがオーサリングされ、適切なコーデック装置において符号化されたら、該オーディオ・コンテンツは、再生システムのさまざまなコンポーネントにおいてデコードされ、レンダリングされる。

【0127】

一般に、再生システムはいかなる業務用または消費者用オーディオ・システムであってもよく、これは家庭シアター（たとえばA/V受領器、サウンドバーおよびブルーレイ）、Eメディア（たとえばヘッドフォン再生を含むPC、タブレット、モバイル）、放送（たとえばTVおよびセットトップボックス）、音楽、ゲーミング、ライブ音、ユーザー生成コンテンツなどを含みうる。適応オーディオ・コンテンツは、すべてのエンドポイント装置のための消費者聴衆のための向上した没入感、オーディオ・コンテンツ・クリエイターにとっての拡張された芸術的制御、改善されたレンダリングのための改善されたコンテンツ依存（記述）メタデータ、消費者再生システムのための拡張された柔軟性およびスケラビリティ、音色保存およびマッチングならびにユーザー位置および対話に基づくコンテンツの動的レンダリングの機会を提供する。本システムは、コンテンツ・クリエイターのための新たなミキシング・ツール、頒布および再生のための更新された新しいパッケージング
20
30
および符号化ツール、家庭内動的混合およびレンダリング（種々の消費者構成について適切）、追加的なスピーカー位置および設計を含む、いくつかのコンポーネントを含む。

【0128】

本稿に記載されたオーディオ環境の諸側面は、適切なスピーカーおよび再生装置を通じたオーディオまたはオーディオ/ビジュアル・コンテンツの再生を表わし、聴取者が捕捉されたコンテンツの再生を経験している任意の環境を表わしうる。該環境はたとえば、映画館、コンサートホール、野外シアター、家または部屋、聴取ブース、自動車、ゲーム・コンソール、ヘッドフォンまたはヘッドセット・システム、公共案内（PA: public address）システムまたは他の任意の再生環境などである。オブジェクト・ベースのオーディオおよびチャンネル・ベースのオーディオを含む空間的オーディオ・コンテンツは、何らかの
40
関係したコンテンツ（たとえば関連付けられたオーディオ、ビデオ、グラフィックなど）との関連で使用されてもよく、あるいはスタンドアローンのオーディオ・コンテンツをなしていてもよい。再生環境は、ヘッドフォンまたは近距離場モニタ（near field monitors）から小さなまたは大きな部屋、自動車、野外アリーナ、コンサートホールなど、いかなる適切な聴取環境であってもよい。

【0129】

本稿に記載されるシステムの諸側面は、デジタルのまたはデジタイズされたオーディオ・ファイルを処理するための適切なコンピュータ・ベースの音処理ネットワーク環境において実装されてもよい。適応オーディオ・システムの諸部分は、コンピュータ間で伝送されるデータをバッファリングおよびルーティングするはたらきをする一つまたは複数のル
50

ーター（図示せず）を含め、いかなる所望される数の個別の機械を有する一つまたは複数のネットワークを含んでいてもよい。そのようなネットワークは、さまざまな異なるネットワーク・プロトコル上に構築されてもよく、インターネット、広域ネットワーク（WAN）、ローカル・エリア・ネットワーク（LAN）またはその任意の組み合わせであってもよい。ネットワークがインターネットを含むある実施形態では、一つまたは複数の機械がウェブ・ブラウザ・プログラムを通じてインターネットにアクセスするよう構成されていてもよい。

【0130】

コンポーネント、ブロック、プロセスまたは他の機能コンポーネントの一つまたは複数
が、システムのプロセッサ・ベースのコンピューティング装置の実行を制御するコンピ
ュータ・プログラムを通じて実装されてもよい。本稿に開示される様々な機能が、その挙動
、レジスタ転送、論理コンポーネントおよび/または他の特性に関し、ハードウェア、フ
ァームウェアの任意の数の組み合わせを使っておよび/またはさまざまな機械可読もしく
はコンピュータ可読媒体において具現されたデータおよび/または命令として記述されて
もよいことも注意しておくべきである。そのようなフォーマットされたデータおよび/ま
たは命令が具現されうるコンピュータ可読媒体は、光学式、磁気式または半導体記憶媒体
のようなさまざまな形の物理的な（非一時的な）不揮発性の記憶媒体を含むがそれに限ら
れるものではない。

10

【0131】

文脈が明瞭にそうでないことを要求するのではない限り、本記述および請求項を通じて、
単語「有する」「含む」などは、排他的もしくは網羅的な意味ではなく包含的な意味に解
釈されるものとする。すなわち、「.....を含むがそれに限定されない」の意味である。単
数または複数を使った単語は、それぞれ複数または単数をも含む。さらに、「本稿で」「
以下で」「上記で」「下記で」および類似の意味の単語は、全体としての本願を指すの
であって、本願のいかなる特定の部分を指すものでもない。単語「または」が二つ以上の項
目のリストを参照して使われるとき、その単語は該単語の以下の解釈のすべてをカバーす
る：リスト中の項目の任意のもの、リスト中の項目のすべておよびリスト中の項目の任意
の組み合わせ。

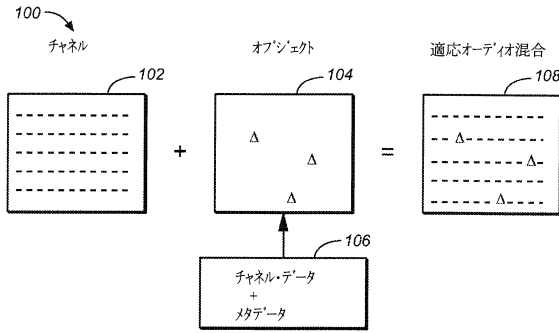
20

【0132】

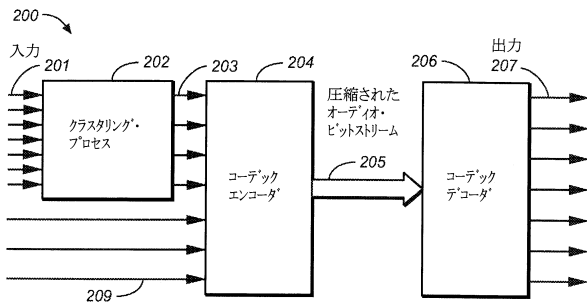
一つまたは複数の実装が、例として、個別的な実施形態を用いて記載されているが、一
つまたは複数の実装は開示される実施形態に限定されないことは理解されるものとする。
逆に、当業者に明白であろうさまざまな修正および類似の構成をカバーすることが意図さ
れている。したがって、付属の請求項の範囲は、そのようなすべての修正および類似の構
成を包含するような最も広い解釈を与えられるべきである。

30

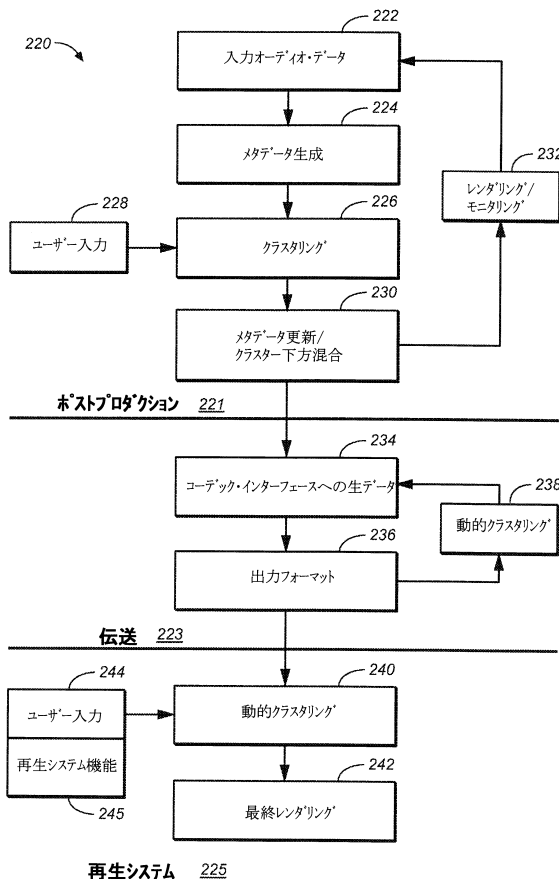
【図1】



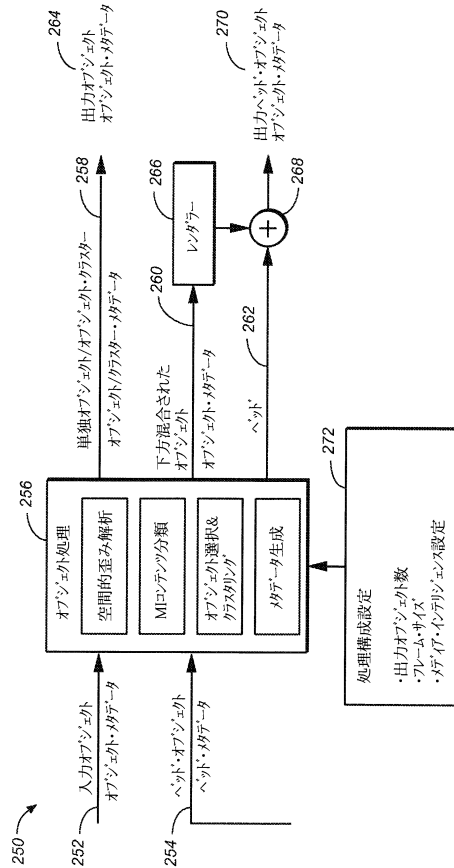
【図2A】



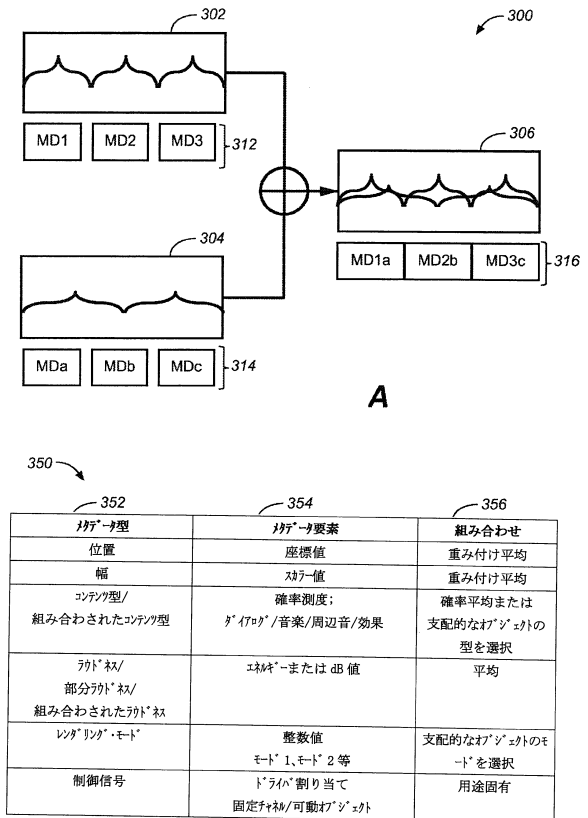
【図2C】



【図2B】

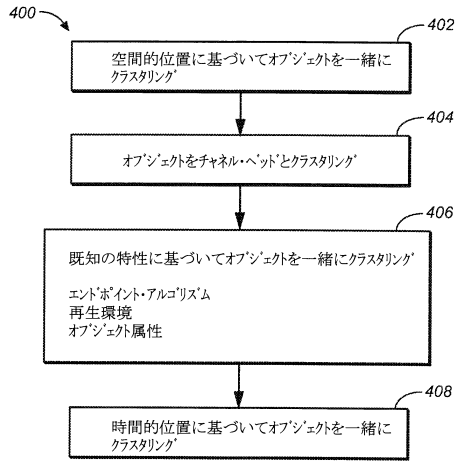


【図3】

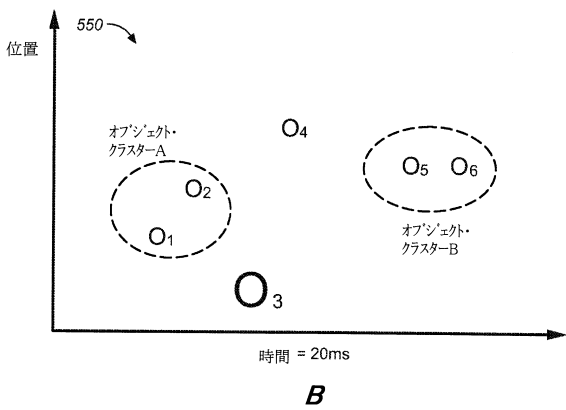
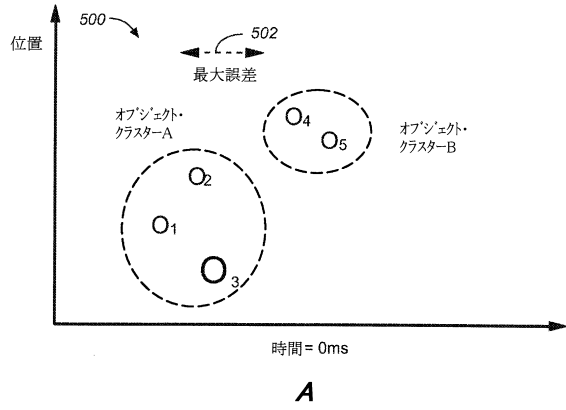


B

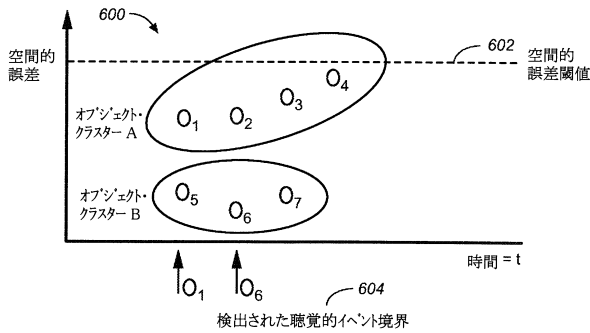
【図4】



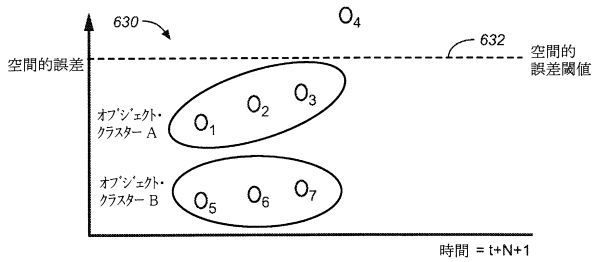
【図5】



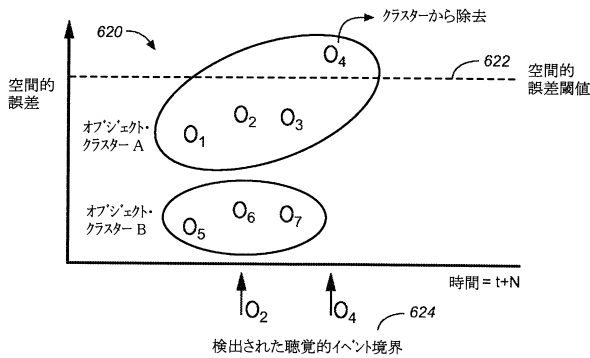
【図6A】



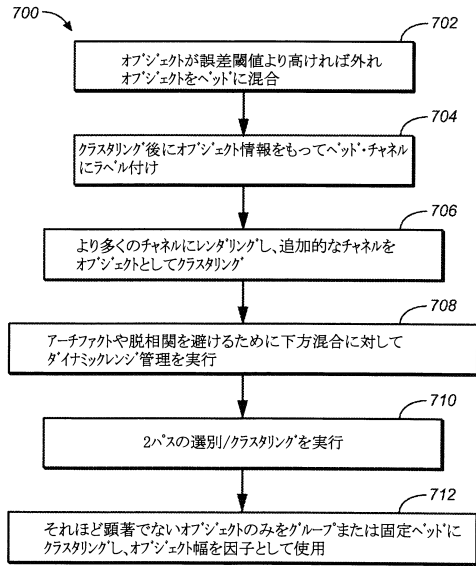
【図6C】



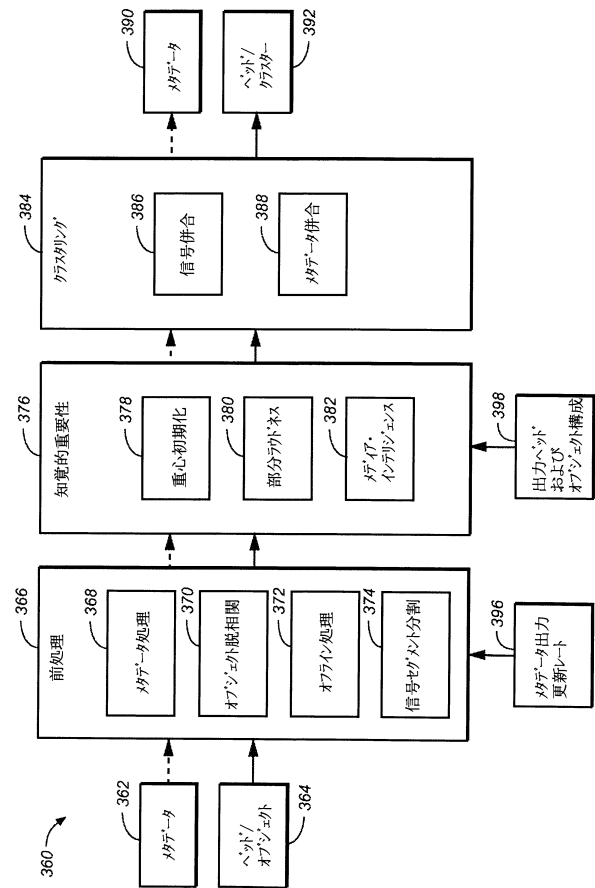
【図6B】



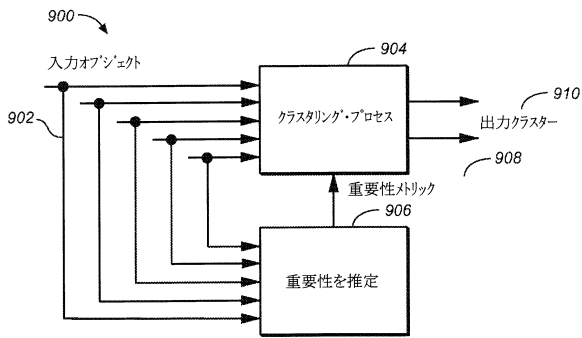
【図7】



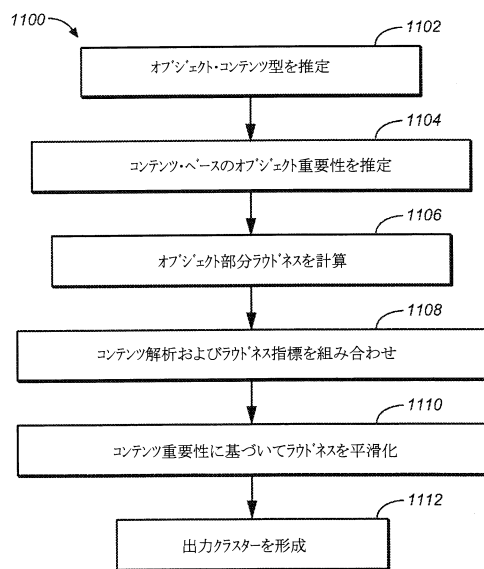
【図8】



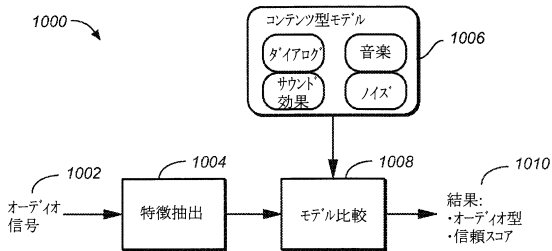
【図9】



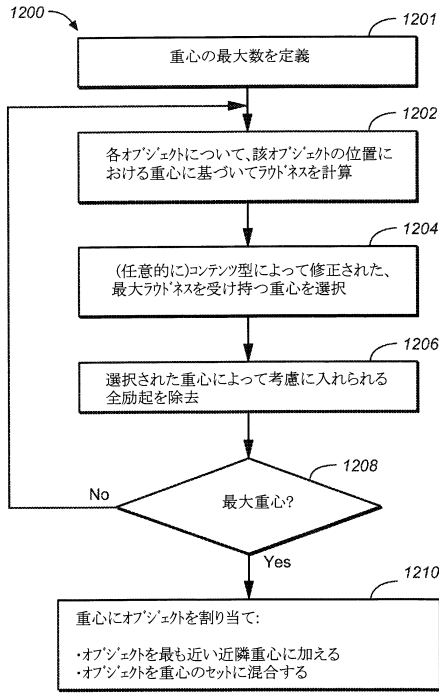
【図11】



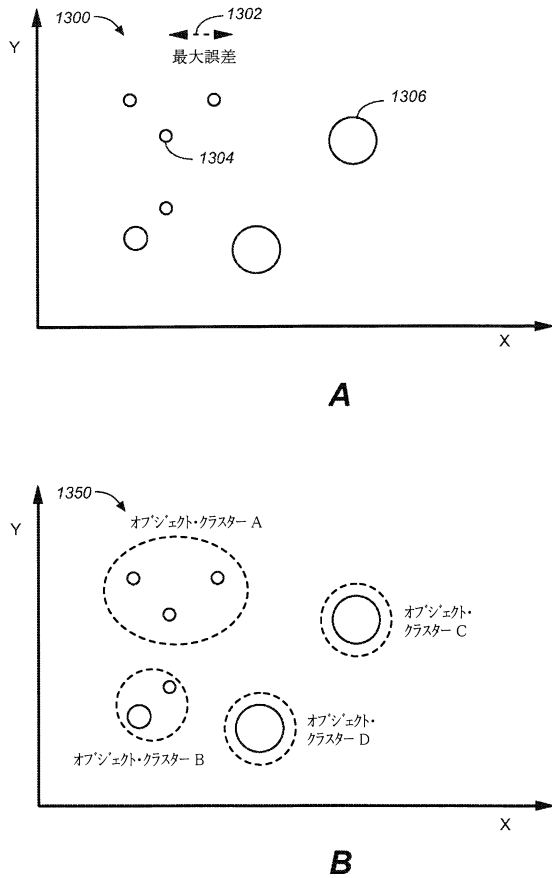
【図10】



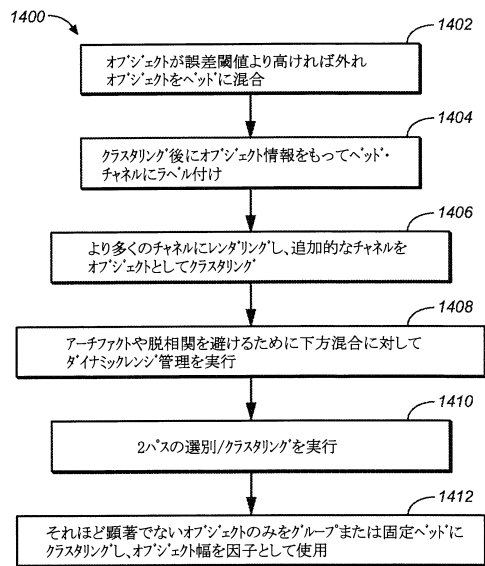
【図12】



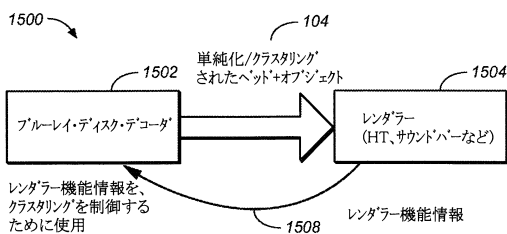
【図13】



【図14】



【図15】



フロントページの続き

- (72)発明者 クロケット, ブレット ジー
アメリカ合衆国, カリフォルニア州 94103, サンフランシスコ, ポットレロ アヴェニュー
100, ドルビー ラボラトリーズ, インコーポレイテッド内
- (72)発明者 シーフェルドット, アラン ジェイ
アメリカ合衆国, カリフォルニア州 94103, サンフランシスコ, ポットレロ アヴェニュー
100, ドルビー ラボラトリーズ, インコーポレイテッド内
- (72)発明者 ツインゴス, ニコラス アール
アメリカ合衆国, カリフォルニア州 94103, サンフランシスコ, ポットレロ アヴェニュー
100, ドルビー ラボラトリーズ, インコーポレイテッド内
- (72)発明者 ウィルソン, ロンダ
アメリカ合衆国, カリフォルニア州 94103, サンフランシスコ, ポットレロ アヴェニュー
100, ドルビー ラボラトリーズ, インコーポレイテッド内
- (72)発明者 ブリーバート, デイルク イエルーン
オーストラリア国, ニューサウスウェールズ 2060, シドニー, マクマホンズ・ポイント, レ
ベル 3, ミッチェル・ストリート 35-51 ドルビー・オーストラリア・プロプライエタリ
ー・リミテッド内
- (72)発明者 ルー, リエ
中華人民共和国, ベイジン 100020, シャオヤン・ディストリクト, イースト・サード・リ
ング・ミドル・ロード, ワールド・フィナンシャル・センター ナンバー 1, ドルビー ラボラ
トリーズ インターナショナル サービスズ カンパニー リミテッド内
- (72)発明者 チェン, リアンウ
中華人民共和国, ベイジン 100020, シャオヤン・ディストリクト, イースト・サード・リ
ング・ミドル・ロード, ワールド・フィナンシャル・センター ナンバー 1, ドルビー ラボラ
トリーズ インターナショナル サービスズ カンパニー リミテッド内

審査官 富澤 直樹

- (56)参考文献 特表2009-532372(JP, A)
特表2011-501823(JP, A)
米国特許出願公開第2009/017676(US, A1)
Kyungryeol Koo, Variable Subband Analysis for High Quality Spatial Audio Object Coding
, Advanced Communication Technology 2008, 米国, IEEE, 2008年 2月20日, p.1205-1
208, ISBN:978-89-5519-136-3

(58)調査した分野(Int.Cl., DB名)

G10L 19/00 - 19/26
H04S 1/00 - 7/00
IEEE Xplore