

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6518254号
(P6518254)

(45) 発行日 令和1年5月22日(2019.5.22)

(24) 登録日 平成31年4月26日(2019.4.26)

(51) Int.Cl.

F 1

G 1 O L 19/008 (2013.01)
G 1 O L 19/00 (2013.01)G 1 O L 19/008 200
G 1 O L 19/00 330 B

請求項の数 15 (全 35 頁)

(21) 出願番号 特願2016-544661 (P2016-544661)
 (86) (22) 出願日 平成27年1月5日 (2015.1.5)
 (65) 公表番号 特表2017-508175 (P2017-508175A)
 (43) 公表日 平成29年3月23日 (2017.3.23)
 (86) 國際出願番号 PCT/US2015/010126
 (87) 國際公開番号 WO2015/105748
 (87) 國際公開日 平成27年7月16日 (2015.7.16)
 審査請求日 平成29年12月26日 (2017.12.26)
 (31) 優先権主張番号 P201430016
 (32) 優先日 平成26年1月9日 (2014.1.9)
 (33) 優先権主張国 スペイン (ES)
 (31) 優先権主張番号 61/951,048
 (32) 優先日 平成26年3月11日 (2014.3.11)
 (33) 優先権主張国 米国 (US)

(73) 特許権者 507236292
 ドルビー ラボラトリーズ ライセンシング コーポレイション
 アメリカ合衆国 94103 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 1275
 (73) 特許権者 510185767
 ドルビー・インターナショナル・アーベー
 オランダ王国, セーベン アムステルダム
 ズイドースト 1101, ヘリケルベル
 グウェグ 1-35, アポロ ビルディング
 グ 3エー
 (74) 代理人 100107766
 弁理士 伊東 忠重

最終頁に続く

(54) 【発明の名称】オーディオ・コンテンツの空間的誤差メトリック

(57) 【特許請求の範囲】

【請求項 1】

一つまたは複数のフレームにおける入力オーディオ・コンテンツに存在している複数のオーディオ・オブジェクトを判別する段階であって、前記複数のオーディオ・オブジェクトは $N_{objects}$ 個のオーディオ・オブジェクトを含み、 $N_{objects} > 2$ である、段階と；

前記一つまたは複数のフレームにおける出力オーディオ・コンテンツに存在している複数の出力クラスターを判別する段階であって、前記入力オーディオ・コンテンツにおける前記複数のオーディオ・オブジェクトが前記出力オーディオ・コンテンツにおける前記複数の出力クラスターに変換され、前記複数の出力クラスターは $N_{clusters}$ 個の出力クラスターを含み、 $N_{objects} > N_{clusters} > 1$ である、段階と；

少なくとも部分的には前記複数のオーディオ・オブジェクトの位置メタデータおよび前記複数の出力クラスターの位置メタデータに基づいて、一つまたは複数の空間的誤差メトリックを計算する段階であって、前記一つまたは複数の空間的誤差メトリックは少なくとも部分的にはオブジェクト重要度に依存する、段階とを含む、
 一つまたは複数のコンピューティング装置によって実行される方法。

【請求項 2】

前記オブジェクト重要度が、前記複数のオーディオ・オブジェクトにおけるオーディオ・データ、前記複数の出力クラスターにおけるオーディオ・データ、前記複数のオーディオ・オブジェクトにおけるメタデータまたは前記複数の出力クラスターにおけるメタデータの一つまたは複数を解析することから得られる、または、

10

20

前記オブジェクト重要度の少なくとも一部がユーザー入力に基づいて決定される、
請求項 1 記載の方法。

【請求項 3】

前記複数のオーディオ・オブジェクトにおける少なくとも一つのオーディオ・オブジェクトが、前記複数の出力クラスターにおける二つ以上の出力クラスターに配分される、または、前記複数の出力クラスターにおけるある出力クラスターに割り当てられる、請求項 1 記載の方法。

【請求項 4】

前記入力オーディオ・コンテンツにおける前記複数のオーディオ・オブジェクトを前記出力クラスターにおける前記複数の出力クラスターに変換することによって引き起こされる知覚的オーディオ品質劣化を、前記一つまたは複数の空間的誤差メトリックに基づいて決定する段階をさらに含む、
10
請求項 1 記載の方法。

【請求項 5】

前記知覚的オーディオ品質劣化は、知覚的オーディオ品質試験に関係する一つまたは複数の予測された試験スコアによって表わされる、請求項 4 記載の方法。

【請求項 6】

前記一つまたは複数の空間的誤差メトリックはフレーム内空間的誤差メトリックを含み、前記フレーム内空間的誤差メトリックは：重要度で重み付けされたフレーム内オブジェクト位置誤差メトリック、重要度で重み付けされたフレーム内オブジェクト・パン誤差メトリック、規格化された重要度で重み付けされたフレーム内オブジェクト位置誤差メトリックまたは規格化された重要度で重み付けされたフレーム内オブジェクト・パン誤差メトリックのうちの少なくとも一つを含む、請求項 1 記載の方法。
20

【請求項 7】

前記一つまたは複数の空間的誤差メトリックはフレーム間空間的誤差メトリックを含み、前記フレーム間空間的誤差メトリックは：利得係数フローに基づき、オブジェクト重要度で重み付けされたフレーム間空間的誤差メトリックを含む、請求項 1 記載の方法。

【請求項 8】

前記複数のオーディオ・オブジェクトは複数の利得係数を介して前記複数の出力クラスターに関係する、請求項 1 記載の方法。
30

【請求項 9】

前記フレームのそれぞれは、前記入力オーディオ・コンテンツにおける第一の時間セグメントおよび前記出力オーディオ・コンテンツにおける第二の時間セグメントに対応し、前記出力オーディオ・コンテンツにおける前記第二の時間セグメントに存在する出力クラスターは、前記入力オーディオ・コンテンツにおける前記第一の時間セグメントに存在するオーディオ・オブジェクトによってマッピングされる、請求項 1 記載の方法。

【請求項 10】

前記複数のオーディオ・オブジェクトのうちのオーディオ・オブジェクトまたは聴取空間内の前記複数の出力クラスターにおける出力クラスターの一つまたは複数を表わす一つまたは複数のユーザー・インターフェース構成要素を構築する段階と；
40

前記一つまたは複数のユーザー・インターフェース構成要素をユーザーに対して表示させる段階とをさらに含む、
請求項 1 記載の方法。

【請求項 11】

前記複数のオーディオ・オブジェクトにおけるオーディオ・オブジェクトのそれぞれのオブジェクト重要性、前記複数の出力クラスターにおける出力クラスターのそれぞれのオブジェクト重要性、前記複数のオーディオ・オブジェクトにおけるオーディオ・オブジェクトのそれぞれのラウドネス、前記複数の出力クラスターにおける出力クラスターのそれぞれのラウドネス、前記複数のオーディオ・オブジェクトにおけるオーディオ・オブジェクトのそれぞれの、発話もしくはダイアログ・コンテンツの確率、前記複数の出力クラス

ターにおける出力クラスターの発話もしくはダイアログ・コンテンツの確率のうちの一つまたは複数を表わす一つまたは複数のユーザー・インターフェース構成要素を構築する段階と；

前記一つまたは複数のユーザー・インターフェース構成要素をユーザーに対して表示させる段階とをさらに含む、

請求項1記載の方法。

【請求項12】

前記一つまたは複数の空間的誤差メトリックまたは少なくとも部分的には前記一つまたは複数の空間的誤差メトリックに基づいて決定された一つまたは複数の予測された試験スコアの一つまたは複数を表わす一つまたは複数のユーザー・インターフェース構成要素を構築する段階と；

10

前記一つまたは複数のユーザー・インターフェース構成要素をユーザーに対して表示させる段階とをさらに含む、

請求項1記載の方法。

【請求項13】

前記複数の出力クラスターにおけるある出力クラスターは、前記複数のオーディオ・オブジェクトにおける二つ以上のオーディオ・オブジェクトによってマッピングされる部分を含む、請求項1記載の方法。

【請求項14】

請求項1ないし13のうちいずれか一項記載の方法を実行するよう構成された、プロセッサを有する装置。

20

【請求項15】

一つまたは複数のプロセッサによって実行されたときに、請求項1ないし13のうちいずれか一項記載の方法の実行を引き起こすソフトウェア命令を記憶している非一時的なコンピュータ可読記憶媒体。

【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2014年1月9日に出願されたスペイン国特許出願第P201430016号および2014年3月1日に出願された米国仮特許出願第61/951,048号の優先権を主張するものである。各出願の内容はここに参照によってその全体において組み込まれる。

30

【0002】

技術分野

本発明は、概略的にはオーディオ信号処理に関し、より詳細にはオーディオ・オブジェクトのフォーマット変換、レンダリング、クラスタリング、リミックスまたは組み合わせに関連する空間的誤差メトリックおよびオーディオ品質劣化を決定することに関する。

【背景技術】

【0003】

オリジナルとしてオーサリング／制作されたオーディオ・コンテンツなどの入力オーディオ・コンテンツは、オーディオ・オブジェクト・フォーマットで個々に表現される多数のオーディオ・オブジェクトを含むことがある。入力オーディオ・コンテンツにおける多数のオーディオ・オブジェクトは、空間的に多様で、没入的で、正確なオーディオ経験を作り出すために使用できる。

40

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、多数のオーディオ・オブジェクトを含む入力オーディオ・コンテンツのエンコード、デコード、伝送、再生などは、高い帯域幅、大きなメモリ・バッファ、高い処理パワーなどを必要とすることがある。いくつかのアプローチのもとでは、入力オーデ

50

イオ・コンテンツは、より少数のオーディオ・オブジェクトを含む出力オーディオ・コンテンツに変換されることがある。同じ入力オーディオ・コンテンツが、多くの異なるオーディオ・コンテンツ配信、伝送および再生セッティング、たとえばブルーレイ・ディスク、放送（たとえばケーブル、衛星、地上波など）、モバイル（たとえば3G、4Gなど）、インターネットなどに関係したものに対応する出力オーディオ・コンテンツの多くの異なるバージョンを生成するために使われることがある。出力オーディオ・コンテンツの各バージョンは、対応するセッティングのために特に適応されていてもよい。該セッティングにおける共通に導出されたオーディオ・コンテンツの効率的な表現、処理、伝送およびレンダリングのための特定の課題に対処するためである。

【0005】

10

このセクションで記述されたアプローチは、追求されたことができたが必ずしも以前に着想または追求されたアプローチではない。したがって、特に断りのない限り、このセクションにおいて記述されるアプローチはいずれも、このセクションに含まれているというだけのために従来技術の資格をもつと想定されるべきではない。同様に、特に断りのない限り、一つまたは複数のアプローチに関して特定されている問題は、このセクションに基づいて何らかの従来技術において認識されていたと想定されるべきではない。

【図面の簡単な説明】

【0006】

本発明は、限定ではなく例として、付属の図面において示される。図面において、同様の参照符号は同様の要素を指す。

20

【図1】オーディオ・オブジェクト・クラスタリングに関する例示的なコンピュータ実装されるモジュールを示す図である。

【図2】例示的な空間的複雑さ解析器を示す図である。

【図3A】一つまたは複数のフレームにおける空間的複雑さの視覚化のための例示的なユーザー・インターフェースを示す図である。

20

【図3B】一つまたは複数のフレームにおける空間的複雑さの視覚化のための例示的なユーザー・インターフェースを示す図である。

【図3C】一つまたは複数のフレームにおける空間的複雑さの視覚化のための例示的なユーザー・インターフェースを示す図である。

【図3D】一つまたは複数のフレームにおける空間的複雑さの視覚化のための例示的なユーザー・インターフェースを示す図である。

30

【図4】二つの例示的な視覚的複雑さメーター事例を示す図である。

【図5】利得フローを計算するための例示的なシナリオを示す図である。

【図6】例示的なプロセス・フローを示す図である。

【図7】本稿に記載されるコンピュータまたはコンピューティング装置が実装されうる例示的なハードウェア・プラットフォームを示す図である。

【発明を実施するための形態】

【0007】

オーディオ・オブジェクト・クラスタリングに関する空間的誤差メトリックおよびオーディオ品質劣化を決定することに関する例示的実施形態が本稿で記載される。以下の記述では、説明の目的のため、本発明の十全な理解を提供するために、数多くの個別的詳細が記載される。しかしながら、本発明がそうした個別的詳細なしでも実施されうることは明白であろう。他方、本発明を無用に隠蔽し、かすませ、あるいは埋没させるのを避けるために、よく知られた構造および装置は網羅的な詳細さでは記述されない。

40

【0008】

例示的実施形態は、本稿では次のアウトラインに従って記載される：

1. 全般的概観
2. オーディオ・オブジェクト・クラスタリング
3. 空間的複雑さ解析器
4. 空間的誤差メトリック

50

- 4.1 フレーム内オブジェクト位置誤差
 - 4.2 フレーム内オブジェクト・パン誤差
 - 4.3 重要度で重み付けされた誤差メトリック
 - 4.4 規格化された誤差メトリック
 - 4.5 フレーム間空間的誤差
 - 5.主観的オーディオ品質の予測
 - 6.空間的誤差および空間的複雑さの視覚化
 - 7.例示的なプロセス・フロー
 - 8.実装機構 ハードウェアの概観
 - 9.等価物、拡張、代替その他
- 10

。

【0009】

1. 全般的概観

この概観は、本発明の実施形態のいくつかの側面の基本的な記述を提示する。この概観は該実施形態の諸側面の包括的ないし網羅的な要約ではないことは注意しておくべきである。さらに、この概観は、該実施形態の何らかの特に有意な側面もしくは要素を特定するものと理解されることも、特に該実施形態の、一般には本発明の、何らかの範囲を画定するものと理解されることも、意図されていないことを注意しておくべきである。この概観は単に、その例示的実施形態に関係するいくつかの概念を凝縮された単純化された形式で提示するものであり、単に後続の例示的な諸実施形態のより詳細な説明への概念的な導入部として理解されるべきである。

20

【0010】

あるフォーマットから別のフォーマットに変換、ダウンミックス、転換、トランスコードなどできる幅広い多様なオーディオ・オブジェクト・ベースのオーディオ・フォーマットが存在しうる。一例では、あるフォーマットはオーディオ・オブジェクトまたは出力クラスターの位置を記述するためにデカルト座標系を用いてもよく、他のフォーマットは、可能性としては距離で増強された角度アプローチを用いてもよい。別の例では、オブジェクト・ベースのオーディオ・コンテンツを効率的に記憶および伝送するために、一組の入力オーディオ・オブジェクトに対してオーディオ・オブジェクト・クラスタリングが実行されて、比較的多数の入力オーディオ・オブジェクトを比較的少数の出力オーディオ・オブジェクトまたは出力クラスターに減らしてもよい。

30

【0011】

本稿に記載される技法は、入力オーディオ・コンテンツをなす一組の（たとえば動的、静的などの）オーディオ・オブジェクトの、出力オーディオ・コンテンツをなす別の一組のオーディオ・オブジェクトへのフォーマット変換、レンダリング、クラスタリング、リミックスまたは組み合わせなどに関連する空間的誤差メトリックおよび／またはオーディオ品質劣化を決定するために使用できる。単に例解のために、入力オーディオ・コンテンツにおけるオーディオ・オブジェクトまたは入力オーディオ・オブジェクトは、時に、単に「オーディオ・オブジェクト」と称されることがある。出力オーディオ・コンテンツにおけるオーディオ・オブジェクトまたは出力オーディオ・オブジェクトは、一般に、「出力クラスター」と称されることがある。さまざまな実施形態において、用語「オーディオ・オブジェクト」および「出力クラスター」は、該オーディオ・オブジェクトを該出力クラスターに変換する特定の変換動作との関係で使われることを注意しておくべきである。たとえば、ある変換動作における出力クラスターは、その後の変換動作において入力オーディオ・オブジェクトとなることもある。同様に、現在の変換動作における入力オーディオ・オブジェクトは、前の変換動作における出力クラスターであることもある。

40

【0012】

入力オーディオ・オブジェクトが比較的少数または疎である場合、入力オーディオ・オブジェクトの少なくともいくつかについて、入力オーディオ・オブジェクトから出力クラスターへの一対一マッピングが可能である。

50

【0013】

いくつかの実施形態では、オーディオ・オブジェクトは、固定位置における一つまたは複数の音要素（たとえば、オーディオ・ベッドまたはオーディオ・ベッドの一部、物理的なチャネルなど）を表わしていてもよい。いくつかの実施形態では、出力クラスターも、固定位置における一つまたは複数の音要素（たとえば、オーディオ・ベッドまたはオーディオ・ベッドの一部、物理的なチャネルなど）を表わしていてもよい。いくつかの実施形態では、動的な位置（または非固定位置）をもつ入力オーディオ・オブジェクトが、固定位置をもつ出力クラスターにクラスタリングされてもよい。いくつかの実施形態では、固定位置をもつ入力オーディオ・オブジェクト（たとえば、オーディオ・ベッドまたはオーディオ・ベッドの一部など）が出力クラスター（たとえば、オーディオ・ベッドまたはオーディオ・ベッドの一部など）にマッピングされてもよい。いくつかの実施形態では、すべての出力クラスターが固定位置をもつ。いくつかの実施形態では、出力クラスターの少なくとも一つが動的位置をもつ。10

【0014】

入力オーディオ・コンテンツにおける入力オーディオ・オブジェクトが出力オーディオ・コンテンツにおける出力クラスターに変換される際、出力クラスターの数は、オーディオ・オブジェクトの数より少なくとも、そうでなくてもよい。入力オーディオ・コンテンツにおけるオーディオ・オブジェクトは、出力オーディオ・コンテンツにおける二つ以上の出力クラスターに配分されてもよい。オーディオ・オブジェクトは、該オーディオ・オブジェクトが位置しているのと同じ位置に位置していてもよい。オーディオ・オブジェクトが位置しているのと同じ位置に位置していなくてもよいある出力クラスターのみに割り当てられてもよい。オーディオ・オブジェクトの位置の出力クラスターの位置へのシフトが空間的誤差を誘起する。本稿に記載される技法は、入力オーディオ・コンテンツにおけるオーディオ・オブジェクトから出力オーディオ・コンテンツにおける出力クラスターへの変換に起因する、空間的誤差メトリックおよび／または空間的誤差に関するオーディオ品質劣化を決定するために使用されることができる。20

【0015】

本稿に記載される技法のもとで決定される空間的誤差メトリックおよび／またはオーディオ品質劣化は、不可逆コーデック、量子化誤差などによって引き起こされる符号化誤差を測る他の品質メトリック（たとえばPEAQなど）に加えて、またはその代わりに使われてもよい。一例では、空間的誤差メトリック、オーディオ品質劣化などは、オーディオ・オブジェクトまたは出力クラスターにおける位置メタデータおよび他のメタデータと一緒に、マルチチャネル・マルチオブジェクト・ベースのオーディオ・コンテンツにおけるオーディオ・コンテンツの空間的複雑さを視覚的に伝えるために使われることができる。30

【0016】

追加的、任意的または代替的に、いくつかの実施形態では、オーディオ品質劣化は、一つまたは複数の空間的誤差メトリックに基づいて生成される、予測された試験スコアの形で提供されてもよい。予測された試験スコアは、入力オーディオ・コンテンツおよび出力オーディオ・コンテンツの知覚的なオーディオ品質のいかなるユーザー調査も実際に実施することなく、出力オーディオ・コンテンツまたはその一部（たとえばフレーム内など）の、入力オーディオ・コンテンツに対する知覚的なオーディオ品質劣化の指標として使用されてもよい。予測された試験スコアは、MUSHRA (MUltiple Stimuli with Hidden Reference and Anchor [隠れた参照およびアンカーをもつ複数刺激]) 試験、MOS (Mean Opinion Score [平均意見スコア]) 試験などのような主観的なオーディオ品質試験に關していくてもよい。いくつかの実施形態では、一つまたは複数の空間的誤差メトリックが、トレーニング・オーディオ・コンテンツ・データの一つまたは複数の代表的な集合から決定／最適化された予測パラメータ（たとえば相関因子など）を使って、一つまたは複数の予測される試験スコアに変換される。40

【0017】

たとえば、トレーニング・オーディオ・コンテンツ・データの該集合内の各要素（または抜粋）が、該要素（または抜粋）内の入力オーディオ・オブジェクトが対応する出力ク50

ラスターに変換またはマッピングされる前にまたは後に、知覚的なオーディオ品質の主観的なユーザー調査にかけられてもよい。ユーザー調査から決定された試験スコアは、予測パラメータを決定または最適化する目的のために、該要素（または抜粋）中の入力オーディオ・オブジェクトおよび対応する出力クラスターに基づいて計算された空間的誤差メトリックと相關付けされてもよい。予測パラメータは、その後、必ずしもトレーニング・データの集合中にはないオーディオ・コンテンツについての試験スコアを予測するために使用できる。

【0018】

本稿に記載される技法のもとでのシステムは、入力オーディオ・コンテンツ（におけるオーディオ・オブジェクト）を出力オーディオ・コンテンツ（における出力クラスター）に変換するプロセス、動作、アルゴリズムなどを指揮するオーディオ・エンジニアに、客観的な仕方で、空間的誤差メトリックおよび / またはオーディオ品質劣化を提供するよう構成されてもよい。本システムは、オーディオ品質劣化を軽減または防止し、出力オーディオ・コンテンツのオーディオ品質に著しく影響する空間的誤差を最小限にするなどの目的のために、前記プロセス、動作、アルゴリズムなどを最適化するために、オーディオ・エンジニアからユーザー入力を受け入れるまたはフィードバックを受領するよう構成されてもよい。

【0019】

いくつかの実施形態では、オブジェクト重要性が、個々のオーディオ・オブジェクトまたは出力クラスターについて推定または決定され、空間的複雑さおよび空間的誤差を推定するために使われる。たとえば、無音であるまたは相対ラウドネスおよび位置近接性の点で他のオーディオ・オブジェクトによってマスクされるオーディオ・オブジェクトは、そのようなオーディオ・オブジェクトにより少ないオブジェクト重要性を割り当てることによって、より大きな空間的誤差を被ってもよい。それほど重要でないオーディオ・オブジェクトは、シーンにおいてより優勢である他のオーディオ・オブジェクトと違って比較的静かなので、かかるそれほど重要でないオーディオ・オブジェクトのより大きな空間的誤差は、ほとんど可聴アーチファクトを生じないことがある。

【0020】

フレーム内空間的誤差メトリックおよびフレーム間空間的誤差メトリックを計算するために使用できる技法が本稿において記載される。フレーム内空間的誤差メトリックの例は：オブジェクト重要性、オブジェクト重要性によって重み付けされた規格化された空間的誤差メトリックなどの任意のものを含むが、それに限定されない。いくつかの実施形態では、フレーム内空間的誤差メトリックは：(i) オーディオ・オブジェクトの、それぞれのコンテキストにおける個々のオブジェクト重要性を含むがそれに限られないオーディオ・オブジェクトにおけるオーディオ・サンプル・データ；および (ii) 変換前のオーディオ・オブジェクトのもとの位置と変換後のオーディオ・オブジェクトの再構成された位置との間の差に基づく客観的な品質メトリックとして計算ができる。

【0021】

フレーム間空間的誤差メトリックの例は、(時間的に)隣接するフレームどうしにおける出力クラスターの利得係数差および位置差の積に関係するもの、(時間的に)隣接するフレームどうしにおける利得係数フローに関係したものなどを含むがそれに限定されない。フレーム間空間的誤差メトリックは、(時間的に)隣接するフレームにおける非一貫性を示すために特に有用でありうる。たとえば、時間的に隣接するフレームを横断したオーディオ・オブジェクトから出力クラスターへの割り当て / 配分における変化は、あるフレームから次のフレームへの補間の際に生じるフレーム間空間的誤差のため、可聴なアーチファクトを生じることがある。

【0022】

いくつかの実施形態では、フレーム間空間的誤差メトリックは：(i) 時間を通じた(たとえば二つの隣接するフレーム間などの)出力クラスターに関する利得係数差；(ii) 時間を通じた出力クラスターの位置変化(たとえば、あるオーディオ・オブジェクト

10

20

30

40

50

がクラスターにパンされるとき、出力クラスターへのオーディオ・オブジェクトの対応するパン・ベクトルが変化する) ; (iii) オーディオ・オブジェクトの相対ラウドネス; などに基づいて計算されることができる。いくつかの実施形態では、フレーム間空間的誤差メトリックは、少なくとも部分的に出力クラスター間での利得係数フローに基づいて計算されることができる。

【0023】

本稿に記載される空間的誤差メトリックおよび/またはオーディオ品質劣化は、ユーザーと対話するよう一つまたは複数のユーザー・インターフェースを駆動するために使用されてもよい。いくつかの実施形態では、オーディオ・オブジェクトの集合の空間的複雑さ(たとえば高品質/低い空間的複雑さ、低品質/高い空間的複雑さなど)を、それらのオーディオ・オブジェクトが変換される出力クラスターの集合と比して示すために、ユーザー・インターフェースにおいて視覚的複雑さメーターが設けられる。いくつかの実施形態では、視覚的空間的複雑さメーターは、オーディオ品質劣化の指標(たとえば、知覚的MOS試験、MUSHRA試験などに関係する予測された試験スコア)を、入力オーディオ・オブジェクトを出力クラスターに変換する対応する変換プロセスへのフィードバックとして、表示する。変換プロセスに関連する空間的複雑さおよび/または空間的誤差メトリックを視覚的に伝えるために、空間的誤差メトリックおよび/またはオーディオ品質劣化の値は、VUメーター、棒グラフ、クリップ・ライト、数値インジケータ、他の視覚的コンポーネントなどを使ってディスプレイ上のユーザー・インターフェースにおいて視覚化されてもよい。

10

【0024】

いくつかの実施形態では、本稿に記載される機構は:ハンドヘルド装置、ゲーム機、テレビジョン、ホームシアター・システム、セットトップボックス、タブレット、モバイル装置、ラップトップ・コンピュータ、ネットブック・コンピュータ、セルラー無線電話、電子書籍リーダー、ポイントオブセール端末、デスクトップ・コンピュータ、コンピュータ・ワークステーション、コンピュータ・キオスク、さまざまな他の種類の端末およびメディア処理ユニットなどの任意のものを含むがそれに限定されないメディア処理システムの一部をなす。

20

【0025】

本稿に記載される好ましい実施形態および一般原理および特徴に対するさまざまな修正が当業者にはすぐに明白となるであろう。このように、本開示は、示される実施形態に限定されることは意図されておらず、本稿に記載される原理および特徴と整合する最も広い範囲を与えられるべきものである。

30

【0026】

本稿に記載される実施形態の任意のものは、単独で、あるいは任意の組み合わせにおいて互いと一緒に使用されうる。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。

40

【0027】

2. オーディオ・オブジェクト・クラスタリング

オーディオ・オブジェクトは、聴取空間(または環境)における特定の物理的位置(単数または複数)から発していると知覚されうる個々の音要素またはその集合と考えられることができる。オーディオ・オブジェクトの例は:オーディオ・プロダクション・セッションにおけるトラックの任意のものを含むが、それに限定されない。オーディオ・オブジェクトは静的(たとえば定常的)であるまたは動的である(たとえば動いている)ことができる。オーディオ・オブジェクトは、一つまたは複数の音要素を表わすオーディオ・サ

50

ンプル・データとは別個のメタデータを含む。メタデータは、所与の時点における（たとえば一つまたは複数のフレームにおける、フレームの一つまたは複数の部分における、など）音要素のうち一つまたは複数のものの一つまたは複数の位置（たとえば、動的なまたは固定された重心位置、聴取空間におけるスピーカーの固定された位置、周囲効果を表わす一つ、二つまたはそれ以上の動的なまたは固定された位置の集合など）を定義する位置メタデータを含む。いくつかの実施形態では、オーディオ・オブジェクトが再生されるとき、該オーディオ・オブジェクトは、実際の再生環境に存在しているスピーカーを使って、その位置メタデータに従ってレンダリングされ、必ずしも、オーディオ・オブジェクトを下流のオーディオ・デコーダのためのオーディオ信号にエンコードする上流のオーディオ・エンコーダが想定した参照オーディオ・チャネル構成のあらかじめ定義された物理的チャネルに出力されるのではない。

10

【0028】

図1は、オーディオ・オブジェクト・クラスタリングのための例示的なコンピュータ実装されるモジュールを示している。図1に示されるように、集団的に入力オーディオ・コンテンツを表わす入力オーディオ・オブジェクト102は、オーディオ・オブジェクト・クラスタリング・プロセス106を通じて出力クラスター104に変換される。いくつかの実施形態では、出力クラスター104は、集団的に、出力オーディオ・コンテンツを表現し、入力オーディオ・オブジェクトよりもコンパクトな入力オーディオ・コンテンツの表現（たとえばより少数のオーディオ・オブジェクトなど）をなす。これにより、低減した記憶および伝送要件ならびに入力オーディオ・コンテンツの再生のための低減した計算およびメモリ要件が許容される。特に、限られた処理能力、限られたバッテリー・パワー、限られた通信機能、限られた再生機能などをもつ消費者ドメイン装置においてそうである。しかしながら、特に多数の疎に分布した入力オーディオ・オブジェクトが存在する実施形態では、他のオーディオ・オブジェクトとクラスタリングされたときにすべての入力オーディオ・オブジェクトが空間的忠実さを維持できるわけではないので、オーディオ・オブジェクト・クラスタリングはある量の空間的誤差を生じる。

20

【0029】

いくつかの実施形態では、オーディオ・オブジェクト・クラスタリング・プロセス106は、入力オーディオ・オブジェクトのサンプル・データ、オーディオ・オブジェクト・メタデータなどの一つまたは複数から生成されるオブジェクト重要性108に少なくとも部分的に基づいて、入力オーディオ・オブジェクト102をクラスタリングする。サンプル・データ、オーディオ・オブジェクト・メタデータなどは、オブジェクト重要性推定器110に入力される。これは、オーディオ・オブジェクト・クラスタリング・プロセス106が使うためのオブジェクト重要性108を生成する。

30

【0030】

本稿で記載されるように、オブジェクト重要性推定器110およびオーディオ・オブジェクト・クラスタリング・プロセス106は、時間の関数として実行されることができる。いくつかの実施形態では、入力オーディオ・オブジェクト102をもってエンコードされたオーディオ信号または入力オーディオ・オブジェクト102から生成された出力クラスター104をもってエンコードされた対応するオーディオ信号は、個々のフレーム（たとえば、20ミリ秒などの継続時間のユニット）にセグメント分割されることができる。そのようなセグメント分割は、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任意の他の変換領域を使ってもよい。オブジェクト重要性推定器（110）は、コンテンツ種別、部分ラウドネスなどを含むがそれに限られない入力オーディオ・オブジェクト（102）の一つまたは複数の特性に基づいて、入力オーディオ・オブジェクト（102）のそれぞれのオブジェクト重要性を生成するよう構成されることができる。

40

【0031】

本稿に記載される部分ラウドネス（partial loudness）は、音響心理学的原理に基づくオーディオ・オブジェクトのセット、集合、グループ、複数、クラスターなどのコンテキストにおけるオーディオ・オブジェクトの（相対的）ラウドネスを表わしていてもよい

50

。オーディオ・オブジェクトの部分ラウドネスは、オーディオ・オブジェクトのオブジェクト重要性を決定し、オーディオ・レンダリング・システムがすべてのオーディオ・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ーディオ・オブジェクトを選択的にレンダリングするために使用できる。

【0032】

オーディオ・オブジェクトは、所与の時点における（たとえばフレーム毎の、一つまたは複数のフレームにおける、あるフレームの一つまたは複数の部分における、など）ダイアログ、音楽、周囲音、特殊効果などといったいくつかの（たとえば定義されているなどの）コンテンツ種別のうちの一つに分類されてもよい。オーディオ・オブジェクトは、その継続時間を通じてコンテンツ種別を変えてよい。（たとえば一つまたは複数のフレームにおける、あるフレームの一つまたは複数の部分における、などの）オーディオ・オブジェクトは、そのオーディオ・オブジェクトがそのフレーム内で特定のコンテンツ種別である確率を割り当てられることができる。一例では、一定のダイアログ種別のオーディオ・オブジェクトは、100パーセントの確率として表現されてもよい。別の例では、ダイアログ種別から音楽種別に変容するオーディオ・オブジェクトは、50パーセント・ダイアログ / 50パーセント音楽、あるいはダイアログおよび音楽種別の異なる百分位組み合わせとして表現されてもよい。

【0033】

オーディオ・オブジェクト・クラスタリング・プロセス106またはオーディオ・オブジェクト・クラスタリング・プロセス106とともに動作するモジュールは、オーディオ・オブジェクトのコンテンツ種別（たとえば、ブーリアン値をもつ成分をもつベクトルなどとして表現される）と、オーディオ・オブジェクトのそれらのコンテンツ種別の確率（たとえば、百分位数値をもつ成分のベクトルとして表現される）とをフレーム毎に決定するよう構成されていてよい。オーディオ・オブジェクトのコンテンツ種別に基づいて、オーディオ・オブジェクト・クラスタリング・プロセス106は、そのオーディオ・オブジェクトを特定の出力クラスターにクラスタリングし、（フレーム毎の、一つまたは複数のフレームにおける、あるフレームの一つまたは複数の部分における）そのオーディオ・オブジェクトと出力クラスターなどとの間の相互の一対一マッピングを割り当てるよう構成されてもよい。

【0034】

例解の目的のために、複数のオーディオ・オブジェクト（たとえば入力オーディオ・オブジェクト102など）のうちの、m番目のフレームに存在するi番目のオーディオ・オブジェクトは、対応する関数 $x_i(n, m)$ によって表現されてもよい。ここで、nは、m番目のフレーム中の複数のオーディオ・データ・サンプルの間でのn番目のオーディオ・データ・サンプルを表わすインデックスである。m番目のフレームなどのフレーム中でのオーディオ・データ・サンプルの総数は、オーディオ・データ・サンプルを作り出すためにオーディオ信号がサンプリングされるサンプリング・レート（たとえば48kHzなど）に依存する。

【0035】

いくつかの実施形態では、前記複数のオーディオ・オブジェクトでm番目のフレームは、次式に示される（たとえばオーディオ・オブジェクト・クラスタリング・プロセスなどにおける）線形演算に基づいて、複数の出力クラスター $y_j(n, m)$ にクラスタリングされる：

$$y_j(n, m) = \sum_i g_{i,j} x_i(n, m) \quad (1)$$

ここで、 $g_{i,j}(m)$ はオブジェクトiのクラスターjへの利得係数を表わす。出力クラスター $y_j(n, m)$ における不連続を回避するため、クラスタリング動作は、フレームを横断した $g_{i,j}(m)$ の変化を補間するために、窓掛けされた部分的に重複するフレームに対して実行されることができる。本稿での用法では、利得係数は、特定の入力オーディオ・オブジェクトの一部の、特定の出力クラスターへの配分を表わす。いくつかの実施形態では、オーディオ・オブジェクト・クラスタリング・プロセス（106）は、式(1)に従って入力オーディ

10

20

30

40

50

オ・オブジェクトを出力クラスターにマッピングするための複数の利得係数を生成するよう構成っていてもよい。代替的、追加的または任意的に、利得係数 $g_{i,j}(m)$ は、補間された利得係数 $g_{i,j}(n,m)$ を生成するためにサンプル(n)を横断して補間されてもよい。代替的に、利得係数は周波数依存であることができる。そのような実施形態では、入力オーディオは、好適なフィルタバンクを使って周波数帯域に分割され、可能性としては、分割された各オーディオに利得係数の異なるセットが適用される。

【0036】

3. 空間的複雑さ解析器

図2は、フレーム内空間的誤差解析器204、フレーム間空間的誤差解析器206、オーディオ品質解析器208、ユーザー・インターフェース・モジュール210などといったいくつかのコンピュータ実装されるモジュールを有する例示的な空間的複雑さ解析器200を示している。図2に示されるように、空間的複雑さ解析器200は、オーディオ・オブジェクト・データ202を受領／収集するよう構成される。該オーディオ・オブジェクト・データが、一組の入力オーディオ・オブジェクト(たとえば図1の102など)と該入力オーディオ・オブジェクトが変換された一組の出力クラスター(たとえば図1の104など)とに関する空間的誤差およびオーディオ品質劣化について解析されるべきものである。オーディオ・オブジェクト・データ202は、入力オーディオ・オブジェクト(102)についてのメタデータ、出力クラスター(104)についてのメタデータ、入力オーディオ・オブジェクト(102)を式(1)に示されるように出力クラスター(104)にマッピングする利得係数、入力オーディオ・オブジェクト(102)の部分ラウドネス、入力オーディオ・オブジェクト(102)のオブジェクト重要性、入力オーディオ・オブジェクト(102)のコンテンツ種別、入力オーディオ・オブジェクト(102)のコンテンツ種別の確率のうちの一つまたは複数を含む。

【0037】

いくつかの実施形態では、フレーム内空間的誤差解析器(204)は、フレーム毎にオーディオ・オブジェクト・データ(202)に基づいて一つまたは複数の型のフレーム内空間的誤差メトリックを決定するよう構成される。いくつかの実施形態では、各フレームについて、フレーム内空間的誤差解析器(204)は：(i)利得係数、入力オーディオ・オブジェクト(102)の位置メタデータ、出力クラスター(102)の位置メタデータなどをオーディオ・オブジェクト・データ(202)から抽出し；(ii)フレーム内の各入力オーディオ・オブジェクトについて個々に、フレーム内のその入力オーディオ・オブジェクトにおけるオーディオ・オブジェクト・データ(202)からの抽出されたデータに基づいて、前記一つまたは複数の型のフレーム内空間的誤差メトリックのそれぞれを計算するよう構成される。

【0038】

フレーム内空間的誤差解析器(204)は、入力オーディオ・オブジェクト(102)について個々に計算された空間的誤差に基づいて、前記一つまたは複数の型のフレーム内空間的誤差メトリックにおける対応する型についての全体的なフレーム毎空間的誤差メトリックを計算するよう構成されることができる。全体的なフレーム毎空間的誤差メトリックは、個々のオーディオ・オブジェクトの空間的誤差を、フレーム内の入力オーディオ・オブジェクト(102)のそれぞれのオブジェクト重要性のような重み因子で重み付けすることなどによって計算されてもよい。追加的、任意的または代替的に、全体的なフレーム毎空間的誤差メトリックは、フレーム内の入力オーディオ・オブジェクト(102)のそれぞれのオブジェクト重要性を示す値の和のような重み因子の和に関係する規格化因子を用いて規格化されるなどしてもよい。

【0039】

いくつかの実施形態では、フレーム間誤差解析器(206)は、二つ以上の隣接するフレームについてのオーディオ・オブジェクト・データ(202)に基づいて一つまたは複数の型のフレーム間空間的誤差メトリックを決定するよう構成される。いくつかの実施形態では、二つの隣接するフレームについて、フレーム間空間的誤差解析器(206)は、

10

20

30

40

50

(i) オーディオ・オブジェクト・データ(202)から、利得係数、入力オーディオ・オブジェクト(102)の位置メタデータ、出力クラスター(102)の位置メタデータなどを抽出し；(ii) それらのフレーム内の各入力オーディオ・オブジェクトについて個々に、それらのフレーム内の入力オーディオ・オブジェクトにおけるオーディオ・オブジェクト・データ(202)からの抽出されたデータに基づいて、前記一つまたは複数の型のフレーム間空間的誤差メトリックのそれぞれを計算するなどする。

【0040】

前記フレーム間空間的誤差解析器(206)は、二つ以上の隣接するフレームについて、それらのフレーム内の入力オーディオ・オブジェクト(102)について個々に計算された空間的誤差に基づいて、前記一つまたは複数の型のフレーム間空間的誤差メトリックなどにおける対応する型についての全体的な空間的誤差メトリックを計算するよう構成されることができる。全体的な空間的誤差メトリックは、それらのフレーム内の入力オーディオ・オブジェクト(102)のそれぞれのオブジェクト重要性のような重み因子をもって個々のオーディオ・オブジェクトの空間的誤差を重み付けすることなどによって計算されてもよい。追加的、任意的または代替的に、全体的な空間的誤差メトリックは、規格化因子、たとえばそれらのフレームにおける入力オーディオ・オブジェクト(102)のそれぞれのオブジェクト重要性に関係するものを用いて規格化されてもよい。

10

【0041】

いくつかの実施形態では、オーディオ品質解析器(208)は、たとえばフレーム内空間的誤差解析器(204)またはフレーム間空間的誤差解析器(206)によって生成された、フレーム内空間的誤差メトリックまたはフレーム間空間的誤差メトリックの一つまたは複数に基づいて知覚的オーディオ品質を決定するよう構成される。いくつかの実施形態では、知覚的オーディオ品質は、空間的誤差メトリックの前記一つまたは複数に基づいて生成される一つまたは複数の予測された試験スコアによって示される。いくつかの実施形態では、前記予測された試験スコアのうち少なくとも一つは、MUSHRA試験、MOS試験などのようなオーディオ品質の主観的な評価試験に関する。オーディオ品質解析器(208)は、トレーニング・データの一つまたは複数のセットなどからあらかじめ決定された予測パラメータ(たとえば相関因子など)を用いて構成設定されてもよい。いくつかの実施形態では、オーディオ品質解析器(208)は、前記空間的誤差メトリックの前記一つまたは複数を、前記予測パラメータに基づいて一つまたは複数の予測された試験スコアに変換するよう構成される。

20

【0042】

いくつかの実施形態では、空間的複雑さ解析器(200)は、本稿に記載される技法のもとで決定される空間的誤差メトリック、オーディオ品質劣化、空間的複雑さなどのうちの一つまたは複数を、出力データ212として、ユーザーまたは他の装置に提供するよう構成される。追加的、任意的または代替的に、いくつかの実施形態では、空間的複雑さ解析器(200)は、入力オーディオ・コンテンツを出力オーディオ・コンテンツに変換することにおいて使用されるプロセス、アルゴリズム、動作パラメータなどに変更またはフィードバックを提供するユーザー入力214を受領するよう構成されることができる。そのようなフィードバックの例はオブジェクト重要性である。追加的、任意的または代替的に、いくつかの実施形態では、空間的複雑さ解析器(200)は、たとえばユーザー入力214において受領されるフィードバックまたは変更に基づいて、あるいは推定された空間的オーディオ品質に基づいて、入力オーディオ・コンテンツを出力オーディオ・コンテンツに変換することにおいて使用されるプロセス、アルゴリズム、動作パラメータなどに制御データ216を送るよう構成されることができる。

30

【0043】

いくつかの実施形態では、ユーザー・インターフェース・モジュール(210)は、一つまたは複数のユーザー・インターフェースを通じてユーザーと対話するよう構成される。ユーザー・インターフェース・モジュール(210)は、ユーザー・インターフェースを通じてユーザーに対して出力データ212の一部または全部を描くユーザー・インター

40

50

フェース構成要素を表示するまたはその表示を引き起こすよう構成されることがある。ユーザー・インターフェース・モジュール(210)はさらに、前記一つまたは複数のユーザー・インターフェースを通じてユーザー入力214の一部または全部を受領するよう構成されることがある。

【0044】

4. 空間的誤差メトリック

単一のフレームにおけるまたは複数の隣り合うフレームにおける全体的な空間的誤差に基づいて、複数の空間的誤差メトリックが計算されうる。全体的な空間的誤差メトリックおよび/または全体的なオーディオ品質劣化を決定/推定することにおいて、オブジェクト重要性は主要な役割を果たすことができる。無音である、比較的静かであるまたは他のオーディオ・オブジェクトによって(たとえばラウドネス、空間的隣接性などの点で)(部分的に)マスクされるオーディオ・オブジェクトは、現在シーンにおいて優勢なオーディオ・オブジェクトより、オーディオ・オブジェクト・クラスタリングのアーチファクトが可聴になる前より大きな空間的誤差を受けることがある。例解の目的で、いくつかの実施形態では、インデックス i をもつオーディオ・オブジェクトは対応するオブジェクト重要性(N_i と記される)をもつ。このオブジェクト重要性は、オブジェクト重要性推定器(図1の110)によって:知覚的ラウドネス・モデルに基づく、オーディオ・オブジェクトの、オーディオ・ベッドおよび他のオーディオ・オブジェクトに対する部分ラウドネス、ダイアログである確率のような意味的情報などの任意のものを含むがそれに限られないいくつかの属性に基づいて生成されてもよい。オーディオ・コンテンツの動的な性質を与えられて、 i 番目のオーディオ・オブジェクトのオブジェクト重要性 $N_i(m)$ は典型的には時間の関数として、たとえばフレーム・インデックス m (これは論理的に、メディア再生時間などのような時間を表わすまたはそのような時間にマッピングされる)の関数として変化する。加えて、オブジェクト重要性メトリックは、オブジェクトのメタデータに依存してもよい。そのような依存性の例は、オブジェクトの位置または移動速度に基づくオブジェクト重要性の修正である。

【0045】

オブジェクト重要性は、時間および周波数の関数として定義されてもよい。本稿で記載されるように、トランスコード、重要度推定、オーディオ・オブジェクト・クラスタリングなどは、離散フーリエ変換(DFT)、直交ミラー・フィルタ(QMF)バンク、(修正)離散コサイン変換(MDCT)、聴覚的フィルタバンク、同様の変換プロセスなどといった任意の好適な変換を使って、諸周波数帯域において実行されてもよい。一般性を失うことなく、 m 番目のフレーム(またはフレーム・インデックス m をもつフレーム)は、時間領域または好適な変換領域におけるオーディオ・サンプルの集合を含む。

【0046】

4.1 フレーム内オブジェクト位置誤差

フレーム内空間的誤差メトリックの一つは、オブジェクト位置誤差に関係し、フレーム内オブジェクト位置誤差メトリックと表わされてもよい。

【0047】

式(1)における各オーディオ・オブジェクト(たとえば i 番目のオーディオ・オブジェクトなど)は、各フレーム(たとえば m など)について関連付けられた位置ベクトルをもつ(たとえば $p_i(m)$ など)。同様に、式(1)における各出力クラスター(たとえば j 番目の出力クラスターなど)も、関連付けられた位置ベクトルをもつ(たとえば $p_j(m)$ など)。これらの位置ベクトルは、オーディオ・オブジェクト・データ(202)における位置メタデータに基づいて空間的複雑さ解析器(たとえば200など)によって決定されてもよい。オーディオ・オブジェクトの位置誤差は、そのオーディオ・オブジェクトの位置と、諸出力クラスターに配分されるそのオーディオ・オブジェクトの重心の位置との間の距離によって表現されてもよい。いくつかの実施形態では、 i 番目のオーディオ・オブジェクトの重心の位置は、そのオーディオ・オブジェクトが配分される諸出力クラスターの諸位置の重み付けされた和として決定され、利得係数 $g_{i,j}(m)$ が重み因子のはた

10

20

30

40

50

らきをする。そのオーディオ・オブジェクトの位置と、諸出力クラスターに配分されるそのオーディオ・オブジェクトの重心の位置との間の平方された距離は、次式を用いて計算されてもよい。

【0048】

【数1】

$$E_i^2(m) = \left| \vec{p}_i(m) - \frac{\sum_j g_{ij}(m) \vec{p}_j(m)}{\sum_j g_{ij}(m)} \right|^2 \quad (2)$$

10

式(2)の右辺 (RHS) の出力クラスターの位置の重み付けされた和は、 i 番目のオーディオ・オブジェクトの知覚される位置を表わす。 $E_i(m)$ は、フレーム m における i 番目のオーディオ・オブジェクトのフレーム内オブジェクト位置誤差と称されてもよい。

【0049】

例示的実装では、利得係数 (たとえば $g_{ij}(m)$ など) は、各オーディオ・オブジェクト (たとえば i 番目のオーディオ・オブジェクト) についてのコスト関数を最適化することによって決定される。式(1)における利得係数を得るために使われるコスト関数の例は、 $E_i(m)$ 、 $E_i(m)$ 以外のL2ノルムなどを含むがそれに限られない。本稿に記載される技法は、 $E_i(m)$ 以外の他の型のコスト関数を用いて最適化することを通じて得られた利得係数を使うよう構成されることができる注意しておく。

20

【0050】

いくつかの実施形態では、 $E_i(m)$ によって表わされるフレーム内オブジェクト位置誤差は、出力クラスターの凸包の外側の位置をもつオーディオ・オブジェクトについて大きいだけであり、凸包内では0である。

【0051】

4.2 フレーム内オブジェクト・パン誤差

式(2)で表わされるオーディオ・オブジェクトの位置誤差が0である場合 (たとえば出力クラスターの凸包内など) であっても、オーディオ・オブジェクトは、クラスタリングおよびレンダリング後には、クラスタリングなしで直接オーディオ・オブジェクトをレンダリングするのと比べて、かなり異なって聞こえることがある。これは、クラスター重心のどれもオーディオ・オブジェクトの位置の近傍に位置をもたず、よってオーディオ・オブジェクト (たとえばオーディオ・オブジェクトを表わすサンプル・データ部分、信号など) がさまざまな出力クラスターの間で分配される場合に起こりうる。フレーム m における i 番目のオーディオ・オブジェクトのフレーム内オブジェクト・パン誤差に関係する誤差メトリックは、次式によって表わされてもよい。

30

【0052】

【数2】

$$F_i^2(m) = \sum_j g_{ij}^2(m) |\vec{p}_i(m) - \vec{p}_j(m)|^2 \quad (3)$$

40

式(1)における利得係数 $g_{ij}(m)$ が重心最適化によって計算されるいくつかの実施形態では、式(3)における誤差メトリック $F_i^2(m)$ は、出力クラスターのうちの一つ (たとえば j 番目の出力クラスターなど) がオブジェクト位置 [付きの p_i] と一致する位置 [付きの p_j] をもつ場合に0になる。しかしながら、そのような一致がなければ、オブジェクトを出力クラスターの重心にパンすることは $F_i^2(m)$ の0でない値につながる。

【0053】

4.3 重要度で重み付けされた誤差メトリック

いくつかの実施形態では、空間的複雑さ解析器 (200) は、シーン内の各オーディオ・オブジェクトの個々のオブジェクト誤差メトリック (たとえば E_i 、 F_i など) を、(たと

50

えば部分ラウドネス N_i などに基づいて決定される)オブジェクト重要性に関して重み付けするよう構成される。オブジェクト重要性、部分ラウドネス N_i などは、受領されたオーディオ・オブジェクト・データ(202)から、空間的複雑さ解析器(200)によって推定または決定されてもよい。それぞれのオブジェクト重要性によって重み付けされたオブジェクト誤差メトリックは、合計されて、次式に示されるように、シーン内のすべてのオーディオ・オブジェクトについての全体的な誤差メトリックを生成することができる:

【数3】

$$A_{E_i}(m) = \sum_i E_i(m) N_i(m) \quad (4)$$

10

$$A_{F_i}(m) = \sum_i F_i(m) N_i(m)$$

代替的、追加的または任意的に、シーン内の各オーディオ・オブジェクトの個々の誤差メトリック(たとえば E_i 、 F_i など)は、合計されて、次式に示されるように、シーン内のすべてのオーディオ・オブジェクトについての平方領域における全体的な誤差メトリックを生成することができる:

【数4】

$$A_{E_i^2}(m) = \sqrt{\sum_i E_i^2(m) N_i^2(m)} \quad (5)$$

20

$$A_{F_i^2}(m) = \sqrt{\sum_i F_i^2(m) N_i^2(m)}$$

【0054】

4.4 規格化された誤差メトリック

式(4)および(5)における規格化されていない誤差メトリックは、次式に示されるように、全体的なラウドネスまたはオブジェクト重要性をもって規格化することができる:

【数5】

30

$$A'_{E_i}(m) = \frac{\sum_i E_i(m) N_i(m)}{\sum_i N_i(m) + N_0} \quad (6)$$

$$A'_{F_i}(m) = \frac{\sum_i F_i(m) N_i(m)}{\sum_i N_i(m) + N_0}$$

$$A'_{E_i^2}(m) = \sqrt{\frac{\sum_i E_i^2(m) N_i^2(m)}{\sum_i N_i^2(m) + N_0^2}} \quad (7)$$

$$A'_{F_i^2}(m) = \sqrt{\frac{\sum_i F_i^2(m) N_i^2(m)}{\sum_i N_i^2(m) + N_0^2}}$$

40

ここで、 N_0 は、部分ラウドネスまたは部分ラウドネスの二乗の和が0に近づく場合(たとえばオーディオ・コンテンツの一部が静かまたはほとんど静かであるなどのとき)に起こりうる数値的な不安定性を防止するための数値的な安定性因子である。空間的複雑さ解析器(200)は、部分ラウドネスまたは部分ラウドネスの二乗の和について特定の閾値(たとえば、最小の静かさなど)をもって構成されていてもよい。安定性因子は、この和が該特定の閾値以下である場合に式(7)に挿入されてもよい。本稿に記載される技法は、規格化されていないまたは規格化された誤差メトリックを計算することにおいて、数値的不安定性を防止する他の方法、例えばダンピングなどと一緒に機能するよう構成されること

50

もできることを注意しておくべきである。

【0055】

いくつかの実施形態では、空間的誤差メトリックは、各フレーム m について計算され、その後（たとえば500msなどの時定数をもつ一次の低域通過フィルタを用いて）低域通過フィルタリングされる。空間的誤差メトリックの最大値、平均、中央値などが、フレームのオーディオ品質の指標として使われてもよい。

【0056】

4.5 フレーム間空間的誤差

いくつかの実施形態では、時間的に隣り合うフレームにおける変化に関係した空間的誤差メトリックが計算されてもよく、本稿ではフレーム間空間的誤差メトリックと称されることがある。これらのフレーム間空間的誤差メトリックは、隣り合うフレームのそれぞれにおける空間的誤差（たとえばフレーム内空間的誤差）が非常に小さいまたはさらには0でありうる状況において使用されてもよいが、それに限定されない。フレーム内空間的誤差が小さくても、フレーム間でのオブジェクトからクラスターへの割り当ての変更は、たとえばあるフレームから次のフレームへの補間の際に生じる空間的誤差に起因して、可聴のアーチファクトを生じることがある。

【0057】

いくつかの実施形態では、本稿に記載されるオーディオ・オブジェクトのフレーム間空間的誤差は：オーディオ・オブジェクトがクラスタリングまたはパンされる出力クラスター重心の位置変化、オーディオ・オブジェクトがクラスタリングまたはパンされる出力クラスターに関する利得係数の変化、オーディオ・オブジェクトの位置変化、オーディオ・オブジェクトの相対または部分ラウドネスなどの任意のものを含むがそれに限られない一つまたは複数の空間的誤差関係因子に基づいて生成される。

【0058】

例として、フレーム間空間的誤差は、オーディオ・オブジェクトの利得係数の変化およびオーディオ・オブジェクトがクラスタリングまたはパンされる出力クラスターの位置変化に基づいて、次式に示されるように生成されることができる：

【数6】

$$\sum_j |g_{ij}(m) - g_{ij}(m+1)| |\vec{p}_j(m) - \vec{p}_j(m+1)| \quad (8)$$

上記のメトリックは、（1）オーディオ・オブジェクトの利得係数が著しく変化するおよび／または（2）オーディオ・オブジェクトがクラスタリングまたはパンされる出力クラスターの位置が著しく変化する場合に、大きな誤差を与える。さらに、上記のメトリックは、部分ラウドネスなどのようなオーディオ・オブジェクトの特定のオブジェクト重要性によって、次式に示されるように、重み付けができる：

【数7】

$$A_i^2(m \rightarrow m+1) = \sum_j N_i(m) N_i(m+1) |g_{ij}(m) - g_{ij}(m+1)| |\vec{p}_j(m) - \vec{p}_j(m+1)| \quad (9)$$

このメトリックはあるフレームから別のフレームへの遷移に関わるので、二つのフレームのラウドネス値の積が使用されることがある。よって、 m 番目のフレームまたは $(m+1)$ 番目のフレームの一方のオブジェクトのラウドネスが0であれば、上記の誤差メトリックの結果として得られる値も0である。これは、二つのフレームの後者においてオーディオ・オブジェクトが存在するようになるまたは存在しなくなる状況を扱うために使われても

10

20

30

40

50

よい。そのようなオーディオ・オブジェクトからの上記の誤差メトリックへの寄与は0である。

【0059】

フレーム間空間的誤差のもう一つの例は、オーディオ・オブジェクトについて、オーディオ・オブジェクトの利得係数の変化およびオーディオ・オブジェクトがクラスタリングまたはパンされる出力クラスターの位置変化のみならず、図5に示されるように、第一のフレーム（たとえばm番目のフレームなど）においてオーディオ・オブジェクトがレンダリングされる諸出力クラスターの第一の構成と第二のフレーム（たとえば(m+1)番目のフレームなど）においてオーディオ・オブジェクトがレンダリングされる諸出力クラスターの第二の構成との間の差または距離にも基づいて生成されることができる。図5に描かれる例では、出力クラスター2の重心は新たな位置にジャンプまたは移動し、結果として、（三角形として記されている）オーディオ・オブジェクトのレンダリング・ベクトルおよび利得係数（または利得係数分布）はしかるべき変化する。しかしながら、この例において、たとえ出力クラスター2の重心が長い距離をジャンプしたとしても、特定のオーディオ・オブジェクト（三角形）について、それはいまだ、出力クラスター3および4の両方の重心を使うことによってよく表現／レンダリングされることができる。出力クラスターの位置変化（または重心の変化）のジャンプまたは差を考えるだけでは、隣り合うフレーム（たとえばm番目と(m+1)番目のフレームなど）に関する変化の間で引き起こされるフレーム間の空間的誤差または潜在的アーチファクトを過大評価してしまうことがある。この過大評価は、隣り合うフレームに関するフレーム間空間的誤差を決定することにおける隣接フレームの利得係数分布の変化の根底にある利得フローを計算し、考慮に入れる事によって軽減されうる。

【0060】

いくつかの実施形態では、m番目のフレームにおけるオーディオ・オブジェクトの利得係数は利得ベクトル [$g_1(m), g_2(m), \dots, g_N(m)$] を用いて表現できる。ここで、利得ベクトルの各成分（たとえば1, 2, ..., Nなど）は、オーディオ・オブジェクトを、複数の出力クラスター（たとえばN個の出力クラスターなど）のうちの対応する出力クラスター（たとえば、第一出力クラスター、第二出力クラスター、...、第N出力クラスターなど）にレンダリングするために使われる利得係数に対応する。単に例解の目的のために、利得係数におけるオーディオ・オブジェクトのインデックスは、利得ベクトルの成分では無視される。 $(m+1)$ 番目のフレームにおけるオーディオ・オブジェクトの利得係数は、利得ベクトル [$g_1(m+1), g_2(m+1), \dots, g_N(m+1)$] を用いて表現できる。同様に、m番目のフレームにおける前記複数の出力クラスターの重心の位置は、ベクトル

【数8】

$$[\vec{p}_1(m), \vec{p}_2(m), \dots, \vec{p}_N(m)]$$

によって表現できる。 $(m+1)$ 番目のフレームにおける前記複数の出力クラスターの重心の位置は、ベクトル

【数9】

$$[\vec{p}_1(m+1), \vec{p}_2(m+1), \dots, \vec{p}_N(m+1)]$$

によって表現できる。 m 番目のフレームから $(m+1)$ 番目のフレームへのオーディオ・オブジェクトのフレーム間空間的誤差は、次式に示されるように計算できる（当面、オーディオ・オブジェクトのラウドネス、オブジェクト重要性などは無視しているが、のちに適用できる）：

【数10】

$$D(m \rightarrow m+1) = \sum_i \sum_j g_{i \rightarrow j} d_{i \rightarrow j} \quad (10)$$

ここで、 i は m 番目のフレームにおける出力クラスターの重心のインデックスであり、 j は($m+1$)番目のフレームにおける出力クラスターの重心のインデックスである。 $g_{i \rightarrow j}$ は m 番目のフレームにおける i 番目の出力クラスターの重心から $(m+1)$ 番目のフレームにおける j 番目の出力クラスターの重心への利得フローの値である。 $d_{i \rightarrow j}$ は m 番目のフレームにおける i 番目の出力クラスターの重心と $(m+1)$ 番目のフレームにおける j 番目の出力クラスターの重心との間の(たとえば利得フローなど)距離であり、次式に示されるように直接計算されうる：

【数11】

$$d_{i \rightarrow j} = |\vec{p}_i(t) - \vec{p}_j(t+1)| \quad (11)$$

いくつかの実施形態では、利得フロー値 $g_{i \rightarrow j}$ は、次のステップを含む方法によって推定される：

1. $g_{i \rightarrow j}$ を0に初期化する。 $g_i(m)$ および $g_j(m+1)$ が0より大きければ、 (i, j) の各対について $d_{i \rightarrow j}$ を計算する。

2. 最小距離をもつ重心対 (i^*, j^*) を選択する。ここで、重心対 (i^*, j^*) は以前に選択されたことがないものである。

3. 利得フロー値を

【数12】

$$g_{i^* \rightarrow j^*} = \min(g_{i^*}, g_{j^*})$$

として計算する。

4.

【数13】

$$g_{i^*} = g_{i^*} - g_{i^* \rightarrow j^*}, \quad g_{j^*} = g_{j^*} - g_{i^* \rightarrow j^*}$$

と更新する。

5. 更新された g_i 、 g_j のすべてが0であれば停止する。そうでなければ、上記のステップ2に進む。

【0061】

図5に描かれる例では、上記の方法を適用することによって得られる0でない利得フローは： $g_{1 \rightarrow 1} = 0.5$ 、 $g_{2 \rightarrow 3} = 0.2$ 、 $g_{2 \rightarrow 4} = 0.2$ 、 $g_{2 \rightarrow 1} = 0.1$ である。よって、(図5で三角形で記されている)オーディオ・オブジェクトについてのフレーム間空間的誤差は、次のように計算できる：

【数14】

$$\begin{aligned}
 D(m \rightarrow m+1) &= g_{1 \rightarrow 1} * d_{1 \rightarrow 1} + g_{2 \rightarrow 3} * d_{2 \rightarrow 3} + g_{2 \rightarrow 4} * d_{2 \rightarrow 4} + g_{2 \rightarrow 1} * d_{2 \rightarrow 1} \\
 &= 0.5 * d_{1 \rightarrow 1} + 0.2 * d_{2 \rightarrow 3} + 0.2 * d_{2 \rightarrow 4} + 0.1 * d_{2 \rightarrow 1} \quad (12)
 \end{aligned}$$

比較として、式(8)に基づいて計算されるフレーム間空間的誤差は次のようになる。

【0062】

10

【数15】

$$\begin{aligned}
 D(m \rightarrow m+1) &= |g_2(m) - g_2(m+1)| * |\vec{p}_2(m) - \vec{p}_2(m+1)| = 0.5 * |\vec{p}_2(t) - \\
 &\quad |\vec{p}_2(t+1)| \quad (13)
 \end{aligned}$$

20

式(12)および(13)において見て取れるように、式(13)で計算されるフレーム間空間的誤差は

【数16】

$$|\vec{p}_2(m) - \vec{p}_2(m+1)|$$

30

のみに依存し、実際の空間的誤差を過大評価することがありうる。出力クラスター2の重心が動いても、以前、m番目のフレームにおいて出力クラスター2にレンダリングされていた利得係数の部分（または利得フロー）を容易に（かつ空間的誤差の点で比較的正確に）受け継ぐことのできる近くの出力クラスター3および4の存在のため、オーディオ・オブジェクトの大きな空間的誤差を引き起こさないからである。

【0063】

オーディオ・オブジェクトkのフレーム間空間的誤差は D_k と記されてもよい。いくつかの実施形態では、全体的なフレーム間空間的誤差は、次のように計算できる：

【数17】

$$E_{\text{inter}}(m \rightarrow m+1) = \sum_k D_k(m \rightarrow m+1) \quad (14)$$

40

オーディオ・オブジェクトの部分ラウドネスなどのようなそれぞれのオブジェクト重要性を考えることにより、全体的なフレーム間空間的誤差がさらに次のように計算できる：

【数18】

$$E_{\text{inter}}(m \rightarrow m+1) = \sum_k N_k(m)N_k(m+1)D_k(m \rightarrow m+1) \quad (15)$$

ここで、 $N_k(m)$ および $N_k(m+1)$ は、それぞれm番目のフレームおよび(m+1)番目のフレームにおけるオーディオ・オブジェクトkの部分ラウドネスなどのようなオブジェクト重要性

50

である。

【0064】

いくつかの実施形態では、オーディオ・オブジェクトも動いているシナリオにおいて、オーディオ・オブジェクトの動きは、たとえば次式に示されるように、フレーム間空間的誤差の計算において補償される：

【数19】

$$E_{\text{inter}}(m \rightarrow m+1) = \sum_k N_k(m)N_k(m+1) \max\{D_k(m \rightarrow m+1) - O_k(m \rightarrow m+$$

1), 0\}

10

(16)

ここで、 $O_k(m \rightarrow m+1)$ は、 m 番目のフレームから $(m+1)$ 番目のフレームへのオーディオ・オブジェクトの実際の動きである。

【0065】

5. 主観的オーディオ品質の予測

いくつかの実施形態では、本稿に記載される空間的誤差メトリックの一つ、いくつかまたは全部が、空間的誤差メトリックが計算されるもとになった一つまたは複数のフレームの知覚されるオーディオ品質（たとえば、MUSHRA試験、MOS試験などのような知覚されるオーディオ品質の試験に関係するオーディオ品質）を予測するために使用されてもよい。トレーニング・データセット（たとえば、代表的なオーディオ・コンテンツ要素または抜粋の集合など）が、空間的誤差メトリックと複数のユーザーから集められた主観的なオーディオ品質の測定との間の相関（たとえば、負の値が、空間的誤差が大きいほどユーザーにより測定された主観的オーディオ品質が低くなることを反映するなど）を決定するために使われてもよい。トレーニング・データセットに基づいて決定された相関は、予測パラメータを決定するために使用されてもよい。これらの予測パラメータは、一つまたは複数のフレーム（たとえば非トレーニング・データなど）から計算された空間的誤差メトリックに基づいて、該一つまたは複数のフレームの知覚されるオーディオ品質の一つまたは複数の指標を生成するために使われてもよい。複数の空間的誤差メトリック（たとえば、フレーム内オブジェクト位置誤差、フレーム内オブジェクト・パン誤差など）が主観的オーディオ品質を予測するために使われるいくつかの実施形態では、主観的なオーディオ品質（たとえば、トレーニング・データセットに基づいて複数のユーザーに関してMUSHRA試験を通じて測定されたもの）と比較的高い相関（たとえば比較的大きな絶対値をもつ負の値など）をもつ空間的誤差メトリック（たとえばフレーム内オブジェクト・パン誤差メトリックなど）が、前記複数の空間的誤差メトリック（たとえば、フレーム内オブジェクト位置誤差、フレーム内オブジェクト・パン誤差など）の間で比較的高い重み値を与えられてもよい。本稿に記載される技法が、これらの技法によって決定される一つまたは複数の空間的誤差メトリックに基づいてオーディオ品質を予測する他の方法とともに機能するよう構成されることを注意しておくべきである。

20

30

40

【0066】

6. 空間的誤差および空間的複雑さの視覚化

いくつかの実施形態では、一つまたは複数のフレームについて本稿に記載される技法のもとで決定される一つまたは複数の空間的誤差メトリックは、前記一つまたは複数のフレームにおけるオーディオ・オブジェクトおよび/またはオーディオ・クラスターの属性（たとえばラウドネス、位置など）と一緒に、ディスプレイ（たとえばコンピュータ画面、ウェブ・ページなど）上に前記一つまたは複数のフレームにおけるオーディオ・コンテンツの空間的複雑さの視覚化を提供するために使われてもよい。視覚化は、VUメーター、オーディオ・オブジェクトおよび/または出力クラスターの（たとえば2D、3Dなどの）視覚

50

化、棒グラフ、他の好適な手段などといった幅広い多様なグラフィック・ユーザー・インターフェース構成要素〔コンポーネント〕を用いて提供されてもよい。いくつかの実施形態では、空間的複雑さの全体的指標が、たとえば空間的オーサリングもしくは変換プロセスが実行されている際、そのようなプロセスが実行された後などに、ディスプレイ上に提供される。

【0067】

図3Aないし図3Dは、一つまたは複数のフレームにおける空間的複雑さを視覚化するための例示的なユーザー・インターフェースを示す。ユーザー・インターフェースは、空間的複雑さ解析器（たとえば図2の200など）もしくはユーザー・インターフェース・モジュール（たとえば図2の210など）、ミキシング・ツール、フォーマット変換ツール、オーディオ・オブジェクト・クラスタリング・ツール、スタンドアローンの解析ツールなどによって提供されてもよい。ユーザー・インターフェースは、入力オーディオ・コンテンツ中のオーディオ・オブジェクトが出力オーディオ・コンテンツ中の（たとえばはるかに、など）より少ない数の出力クラスターに圧縮されるときの、可能なオーディオ品質劣化および他の関係した情報の視覚化を提供するために使用できる。可能なオーディオ品質劣化および他の関係した情報の視覚化は、同じ源オーディオ・コンテンツからのオブジェクト・ベースのオーディオ・コンテンツの一つまたは複数のバージョンの制作と並行して提供されてもよい。

【0068】

いくつかの実施形態では、ユーザー・インターフェースは、図3Aに示されるように、例示的な3D聴取空間におけるオーディオ・オブジェクトおよび出力クラスターの位置を視覚化する3D表示構成要素302を含む。ユーザー・インターフェースにおいて描かれているオーディオ・オブジェクトまたは出力クラスターの0個、一つまたは複数が、聴取空間における動的な位置または固定した位置を有していてもよい。

【0069】

いくつかの実施形態では、ユーザーまたは聴取者は、3D聴取空間のグラウンド平面の中央にいる。いくつかの実施形態では、ユーザー・インターフェースは、図3Bに示されるように3D聴取空間の種々の投影を表わす上面図、側面図、背面図などのような3D聴取空間の種々の2Dビューを含む。

【0070】

いくつかの実施形態では、ユーザー・インターフェースは、図3Cに示されるように、それぞれオブジェクト重要性（たとえばラウドネス、意味的ダイアログ確率などに基づいて決定／推定されるもの）およびオブジェクト・ラウドネスL（フォン（phon）単位）を視覚化する棒グラフ304および306をも含む。「入力インデックス」はオーディオ・オブジェクト（または出力クラスター）のインデックスを表わす。入力インデックスの各値における垂直の棒の高さが発話またはダイアログの確率を示す。縦軸Lは部分ラウドネスを表わす。これはオブジェクト重要性などを決定するための基礎として使われてもよい。縦軸Pは発話またはダイアログ・コンテンツの確率を表わす。棒グラフ304および306における垂直の棒（オーディオ・オブジェクトまたは出力クラスターの個々の部分ラウドネスおよび発話もしくはダイアログ・コンテンツの確率を表わす）は、フレームからフレームにかけて上がったり下がったりしうる。

【0071】

いくつかの実施形態では、ユーザー・インターフェースは、図3Dに示されるように、フレーム内空間的誤差に関係する第一の空間的複雑さメーター308と、フレーム間空間的誤差に関係する第二の空間的複雑さメーター310とを含む。いくつかの実施形態では、オーディオ・コンテンツの空間的複雑さは、フレーム内空間的誤差メトリック、フレーム間空間的誤差メトリックなどの一つまたは複数（たとえば種々の組み合わせなど）から生成される空間的誤差メトリックまたは予測されるオーディオ品質試験スコアによって定量化または表現されることができる。いくつかの実施形態では、トレーニング・データに基づいて決定される予測パラメータが、一つまたは複数の空間的誤差メトリックの値に基

10

20

30

40

50

づいて知覚的なオーディオ品質劣化を予測するために使われてもよい。予測される知覚的なオーディオ品質劣化は、MUSHRA試験、MOS試験などのような主観的な知覚的オーディオ品質試験を基準とした一つまたは複数の予測される知覚的試験スコアによって表現されてもよい。いくつかの実施形態では、少なくとも部分的にはそれぞれフレーム内空間的誤差およびフレーム間空間的誤差に基づいて、二組の知覚的試験スコアが予測されてもよい。少なくとも部分的にはフレーム内空間的誤差に基づいて生成される第一の組の知覚的試験スコアは、第一の空間的複雑さメーター 308 の表示を駆動するために使われてもよい。少なくとも部分的にはフレーム間空間的誤差に基づいて生成される第二の組の知覚的試験スコアは、第二の空間的複雑さメーター 310 の表示を駆動するために使われてもよい。

【0072】

10

いくつかの実施形態では、空間的複雑さメーター（たとえば 308、310 など）の一つまたは複数によって表わされる予測される（たとえば 0 ないし 10 などの値域内の）オーディオ品質劣化が構成設定された「わずらわしくなる」閾値（たとえば 10 など）を超えたことを示すために、ユーザー・インターフェースにおいて「可聴な誤差」インジケータ・ランプが描かれてもよい。いくつかの実施形態では、「可聴な誤差」インジケータ・ランプは、空間的複雑さメーター（たとえば 308、310 など）のどれも構成設定された「わずらわしくなる」閾値（たとえば 数値 10 などをもつ閾値）を超えない場合には、描かれなくてもよく、空間的複雑さメーターの一つが該構成設定された「わずらわしくなる」閾値を超える際に、トリガーされることができる。いくつかの実施形態では、空間的複雑さメーター（たとえば 308、310 など）における予測されるオーディオ品質劣化の異なる部分範囲が異なる色の帯によって表現されてもよい（たとえば、0~3 の部分範囲は緑の帯にマッピングされて最小限のオーディオ品質劣化を示し、8~10 の部分範囲は赤の帯にマッピングされて深刻なオーディオ品質劣化を示すなど）。

【0073】

20

オーディオ・オブジェクトは図 3 A および図 3 B では円として描かれている。しかしながら、さまざまな実施形態において、オーディオ・オブジェクトまたは出力クラスターは、異なる形を使って描かれることができる。いくつかの実施形態では、オーディオ・オブジェクトまたは出力クラスターを表わす形のサイズが、該オーディオ・オブジェクトのオブジェクト重要性、該オーディオ・オブジェクトまたは出力クラスターの絶対的または相対的ラウドネスなどを示してもよい（たとえばそれに比例していてもよい、など）。ユーザー・インターフェースにおいてユーザー・インターフェース構成要素を色づけするために、種々のカラー・コード方式が使用される。たとえば、オーディオ・オブジェクトは緑の色を付けられてもよく、一方、出力クラスターは緑でない色を付けられてもよい。同じ色の種々の陰影がオーディオ・オブジェクトの属性の異なる値を区別するために使われてもよい。オーディオ・オブジェクトの色は、該オーディオ・オブジェクトの属性、該オーディオ・オブジェクトの空間的誤差、該オーディオ・オブジェクトの、該オーディオ・オブジェクトが配分されるまたは割り当てられる出力クラスターに対する距離などに基づいて変えられてもよい。

【0074】

30

図 4 は、VU メーターの形での視覚的複雑さメーターの二つの例示的なインスタンス 402 および 404 を示している。VU メーターは、図 3 A ないし図 3 D に描かれたユーザー・インターフェースの一部、あるいは図 3 A ないし図 3 D に描かれたユーザー・インターフェース以外の異なるユーザー・インターフェースでありうる。視覚的複雑さメーターの第一のインスタンス 402 は、低い空間的誤差に対応して、高いオーディオ品質および低い空間的複雑さを示す。視覚的複雑さメーターの第二のインスタンス 404 は、高い空間的誤差に対応して、低いオーディオ品質および高い空間的複雑さを示す。VU メーターにおいて示される複雑さメトリック値は、フレーム内空間的誤差、フレーム間空間的誤差、フレーム内空間的誤差に基づいて予測 / 決定された知覚的オーディオ品質試験スコア、フレーム間空間的誤差に基づいて予測 / 決定された予測オーディオ品質試験スコアなどでありうる。追加的、任意的または代替的に、VU メーターは、ある（たとえば過去などの）時間区

40

50

間において生起する最低の品質および最高の複雑さを表示するよう構成された「ピーク保持」機能を有して／実装していてもよい。時間区間は固定であってもよく（たとえば直近の10秒など）、あるいは可変であり、処理されているオーディオ・コンテンツの先頭に対するものであってもよい。また、複雑さメトリック値の数値表示がVUメーター表示と関連して、またはその代替として使われてもよい。

【0075】

図4に示されるように、複雑さクリップ・ライトが複雑さメーターを表わす垂直スケールの下に表示されることがある。このクリップ・ライトは、複雑さ値がある臨界閾値に達した／超えた場合にアクティブになってもよい。これは、明るくなること、色を変えることまたは視覚的に知覚できる他の任意の変化によって視覚化されうる。いくつかの実施形態では、複雑さラベル（たとえば高、良好、中間および低品質など）を示す代わりにまたはそれに加えて、垂直スケールも複雑さまたはオーディオ品質を示す（たとえば0から10などの）数値であってもよい。

【0076】

7. 例示的なプロセス・フロー

図6は、例示的なプロセス・フローを示している。いくつかの実施形態では、一つまたは複数のコンピューティング装置またはユニット（たとえば図2の空間的複雑さ解析器200など）がこのプロセス・フローを実行してもよい。

【0077】

ブロック602では、空間的複雑さ解析器200（図2に示されるものなど）は、一つまたは複数のフレームにおける入力オーディオ・コンテンツに存在している複数のオーディオ・オブジェクトを判別する。

【0078】

ブロック604では、空間的複雑さ解析器（200）は、前記一つまたは複数のフレームにおける出力オーディオ・コンテンツに存在している複数の出力クラスターを判別する。ここで、前記入力オーディオ・コンテンツにおける前記複数のオーディオ・オブジェクトは前記出力オーディオ・コンテンツにおける前記複数の出力クラスターに変換される。

【0079】

ブロック606では、空間的複雑さ解析器（200）は、少なくとも部分的には前記複数のオーディオ・オブジェクトの位置メタデータおよび前記複数の出力クラスターの位置メタデータに基づいて、一つまたは複数の空間的誤差メトリックを計算する。

【0080】

ある実施形態では、前記複数のオーディオ・オブジェクトにおける少なくとも一つのオーディオ・オブジェクトが前記複数の出力クラスターにおける二つ以上の出力クラスターに配分される。

【0081】

ある実施形態では、前記複数のオーディオ・オブジェクトのうちの少なくとも一つのオーディオ・オブジェクトが、前記複数の出力クラスターにおける出力クラスターに割り当てる。

【0082】

ある実施形態では、空間的複雑さ解析器（200）はさらに、前記入力オーディオ・コンテンツにおける前記複数のオーディオ・オブジェクトを前記出力クラスターにおける前記複数の出力クラスターに変換することによって引き起こされる知覚的オーディオ品質劣化を、前記一つまたは複数の空間的誤差メトリックに基づいて決定するよう構成されている。

【0083】

ある実施形態では、前記知覚的オーディオ品質劣化は、知覚的オーディオ品質試験に関係する一つまたは複数の予測された試験スコアによって表わされる。

【0084】

ある実施形態では、前記一つまたは複数の空間的誤差メトリックは：フレーム内空間的

10

20

30

40

50

誤差メトリックまたはフレーム間空間的誤差メトリックの少なくとも一方を含む。

【0085】

ある実施形態では、前記フレーム内空間的誤差メトリックは：フレーム内オブジェクト位置誤差メトリック、フレーム内オブジェクト・パン誤差メトリック、重要度で重み付けされたフレーム内オブジェクト位置誤差メトリック、重要度で重み付けされたフレーム内オブジェクト・パン誤差メトリック、規格化されたフレーム内オブジェクト位置誤差メトリック、規格化されたフレーム内オブジェクト・パン誤差メトリックなどのうちの少なくとも一つを含む。

【0086】

ある実施形態では、前記フレーム間空間的誤差メトリックは：利得係数フローに基づくフレーム間空間的誤差メトリック、利得係数フローに基づかないフレーム間空間的誤差メトリックなどのうちの少なくとも一つを含む。

10

【0087】

ある実施形態では、フレーム間空間的誤差メトリックは二つの異なるフレームに関して計算される。

【0088】

ある実施形態では、前記複数のオーディオ・オブジェクトは複数の利得係数を介して前記複数の出力クラスターに関係する。

【0089】

ある実施形態では、前記フレームのそれぞれは、前記入力オーディオ・コンテンツにおけるある時間セグメントおよび前記出力オーディオ・コンテンツにおける第二の時間セグメントに対応し、前記出力オーディオ・コンテンツにおける前記第二の時間セグメントに存在する出力クラスターは、前記入力オーディオ・コンテンツにおける前記第一の時間セグメントに存在するオーディオ・オブジェクトによってマッピングされる。

20

【0090】

ある実施形態では、前記一つまたは複数のフレームが二つの連続するフレームを含む。

【0091】

ある実施形態では、空間的複雑さ解析器（200）はさらに：前記複数のオーディオ・オブジェクトのうちのオーディオ・オブジェクト、聴取空間内の前記複数の出力クラスターにおける出力クラスターなどの一つまたは複数を表わす一つまたは複数のユーザー・インターフェース構成要素を構築する段階と；前記一つまたは複数のユーザー・インターフェース構成要素をユーザーに対して表示させる段階とを実行するよう構成される。

30

【0092】

ある実施形態では、前記一つまたは複数のユーザー・インターフェース構成要素におけるあるユーザー・インターフェース構成要素は、前記複数のオーディオ・オブジェクトのうちのあるオーディオ・オブジェクトを表わし；該オーディオ・オブジェクトは前記複数の出力クラスターのうちの一つまたは複数の出力クラスターにマッピングされ；前記ユーザー・インターフェース構成要素の少なくとも一つの視覚的特徴が前記オーディオ・オブジェクトの前記一つまたは複数の出力クラスターへのマッピングに関係した一つまたは複数の空間的誤差の総量を表わす。

40

【0093】

ある実施形態では、前記一つまたは複数のユーザー・インターフェース構成要素は、三次元（3D）形式での聴取空間の表現を有する。

【0094】

ある実施形態では、前記一つまたは複数のユーザー・インターフェース構成要素は、二次元（2D）形式での聴取空間の表現を有する。

【0095】

ある実施形態では、空間的複雑さ解析器（200）はさらに：前記複数のオーディオ・オブジェクトにおけるオーディオ・オブジェクトのそれぞれのオブジェクト重要性、前記複数の出力クラスターにおける出力クラスターのそれぞれのオブジェクト重要性、前記複

50

数のオーディオ・オブジェクトにおけるオーディオ・オブジェクトのそれぞれのラウドネス、前記複数の出力クラスターにおける出力クラスターのそれぞれのラウドネス、前記複数のオーディオ・オブジェクトにおけるオーディオ・オブジェクトのそれぞれの、発話もしくはダイアログ・コンテンツの確率、前記複数の出力クラスターにおける出力クラスターの発話もしくはダイアログ・コンテンツの確率などのうちの一つまたは複数を表わす一つまたは複数のユーザー・インターフェース構成要素を構築する段階と；前記一つまたは複数のユーザー・インターフェース構成要素をユーザーに対して表示させる段階とを実行するよう構成される。

【0096】

ある実施形態では、空間的複雑さ解析器（200）はさらに：前記一つまたは複数の空間的誤差メトリック、少なくとも部分的には前記一つまたは複数の空間的誤差メトリックに基づいて決定された一つまたは複数の予測された試験スコアなどの一つまたは複数を表わす一つまたは複数のユーザー・インターフェース構成要素を構築する段階と；前記一つまたは複数のユーザー・インターフェース構成要素をユーザーに対して表示させる段階とを実行するよう構成される。

【0097】

ある実施形態では、変換プロセスが前記入力オーディオ・コンテンツにおいて存在する時間依存のオーディオ・オブジェクトを、前記出力クラスターをなす時間依存の出力クラスターに変換し、前記一つまたは複数のユーザー・インターフェース構成要素は、前記一つまたは複数のフレームを含み前記一つまたは複数のフレームまでの過去の時間区間にについて前記変換プロセスにおいて生じる最悪のオーディオ品質劣化の視覚的指示を含む。

【0098】

ある実施形態では、前記一つまたは複数のユーザー・インターフェース構成要素は、前記一つまたは複数のフレームを含み前記一つまたは複数のフレームまでの過去の時間区間にについて変換プロセスにおいて生じるオーディオ品質劣化がオーディオ品質劣化閾値を超えたことの視覚的指示を含む。

【0099】

ある実施形態では、前記一つまたは複数のユーザー・インターフェース構成要素は、前記一つまたは複数のフレームにおけるオーディオ品質劣化を示す高さをもつ垂直の棒を含み、前記垂直の棒は前記一つまたは複数のフレームにおけるオーディオ品質劣化に基づいてカラーコーディングされる。

【0100】

ある実施形態では、前記複数の出力クラスターにおけるある出力クラスターは、前記複数のオーディオ・オブジェクトにおける二つ以上のオーディオ・オブジェクトによってマッピングされる部分を含む。

【0101】

ある実施形態では、前記複数のオーディオ・オブジェクトにおけるオーディオ・オブジェクトまたは前記複数の出力クラスターにおける出力クラスターの少なくとも一つが、時間とともに変化する動的位置をもつ。

【0102】

ある実施形態では、前記複数のオーディオ・オブジェクトにおけるオーディオ・オブジェクトまたは前記複数の出力クラスターにおける出力クラスターの少なくとも一つが、時間とともに変化しない固定した位置をもつ。

【0103】

ある実施形態では、前記入力オーディオ・コンテンツまたは前記出力オーディオ・コンテンツの少なくとも一つは、オーディオのみ信号またはオーディオビジュアル信号の一方の一部である。

【0104】

ある実施形態では、空間的複雑さ解析器（200）はさらに：前記入力オーディオ・コンテンツを前記出力オーディオ・コンテンツに変換する変換プロセスに対する変更を指定

10

20

30

40

50

するユーザー入力を受領する段階と；前記ユーザー入力を受領するのに応答して、前記入力オーディオ・コンテンツを前記出力オーディオ・コンテンツに変換する前記変換プロセスに対する前記変更を引き起こす段階とを実行するよう構成される。

【0105】

ある実施形態では、上記の方法のいずれかが、前記変換プロセスが前記入力オーディオ・コンテンツを前記出力オーディオ・コンテンツに変換している間に並行して実行される。

【0106】

実施形態は、本稿に記載される方法のいずれかを実行するよう構成されたメディア処理システムを含む。

10

【0107】

実施形態は、上記の方法のいずれかを実行するよう構成された、プロセッサを有する装置を含む。

【0108】

実施形態は、一つまたは複数のプロセッサによって実行されたときに、上記の方法の任意のものの実行を引き起こすソフトウェア命令を記憶している非一時的なコンピュータ可読記憶媒体を含む。別個の複数の実施形態が本稿において論じられているものの、本稿で論じられる実施形態および／または部分実施形態の任意の組み合わせが組み合わされてさらなる実施形態を形成してもよいことを注意しておく。

【0109】

20

8. 実装機構　　ハードウェアの概観

ある実施形態によれば、本稿に記載される技法は、一つまたは複数の特殊目的コンピューティング装置によって実装される。特殊目的コンピューティング装置は、本技法を実行するよう固定構成とされていてもよいし、あるいは本技法を実行するよう持続的にプログラムされた、一つまたは複数の特定用途向け集積回路（ASIC）またはフィールド・プログラマブル・ゲート・アレイ（FPGA）のようなデジタル電子デバイスを含んでいてもよいし、あるいはファームウェア、メモリ、他の記憶または組み合わせにおけるプログラム命令に従って本技法を実行するようプログラムされた一つまたは複数の汎用ハードウェア・プロセッサを含んでいてもよい。そのような特殊目的コンピューティング装置は、カスタムの固定構成論理、ASICまたはFPGAをカスタムのプログラミングと組み合わせて本技法を達成してもよい。特殊目的コンピューティング装置はデスクトップ・コンピュータ・システム、ポータブル・コンピュータ・システム、ハンドヘルド装置、ネットワーキング装置または本技法を実装するために固定構成および／またはプログラム論理を組み込んでいる他の任意の装置であってもよい。

30

【0110】

たとえば、図7は、本発明のある実施形態が実装されうるコンピュータ・システム700を示すブロック図である。コンピュータ・システム700は、情報を通信するためのバス702または他の通信機構と、情報を処理するための、バス702に結合されたハードウェア・プロセッサ704とを含む。ハードウェア・プロセッサ704はたとえば汎用マイクロプロセッサであってもよい。

40

【0111】

コンピュータ・システム700は、ランダム・アクセス・メモリ（RAM）または他の動的記憶装置のような、情報およびプロセッサ704によって実行されるべき命令を記憶するための、バス702に結合されたメイン・メモリ706をも含む。メイン・メモリ706はまた、一時変数または他の中間的な情報を、プロセッサ704によって実行されるべき命令の実行の間、記憶しておくために使われてもよい。そのような命令は、プロセッサ704にとってアクセス可能な非一時的な記憶媒体に記憶されたとき、コンピュータ・システム700を、前記命令において指定されている処理を実行するよう装置固有の特殊目的機械にする。

【0112】

50

コンピュータ・システム 700 はさらに、バス 702 に結合された、静的な情報およびプロセッサ 704 のための命令を記憶するための読み出し専用メモリ (ROM) 708 または他の静的記憶装置を含む。磁気ディスクまたは光ディスクのような記憶装置 710 が提供され、情報および命令を記憶するためにバス 702 に結合される。

【0113】

コンピュータ・システム 700 は、コンピュータ・ユーザーに対して情報を表示するための、液晶ディスプレイ (LCD) のようなディスプレイ 712 にバス 702 を介して結合されていてもよい。英数字その他のキーを含む入力装置 714 が、情報およびコマンド選択をプロセッサ 704 に伝えるためにバス 702 に結合される。もう一つの型のユーザー入力装置は、方向情報およびコマンド選択をプロセッサ 704 に伝えるとともにディスプレイ 712 上でのカーソル動きを制御するための、マウス、トラックボールまたはカーソル方向キーのようなカーソル・コントロール 716 である。この入力装置は典型的には、第一軸 (たとえば x) および第二軸 (たとえば y) の二つの軸方向において二つの自由度をもち、これにより該装置は平面内での位置を指定できる。

10

【0114】

コンピュータ・システム 700 は、本稿に記載される技法を実施するのに、装置固有の固定構成論理、一つまたは複数のASICもしくはFPGA、コンピュータ・システムと組み合わせてコンピュータ・システム 700 を特殊目的機械にするまたはプログラムするファームウェアおよび / またはプログラム論理を使ってもよい。ある実施形態によれば、本稿の技法は、プロセッサ 704 がメイン・メモリ 706 に含まれる一つまたは複数の命令の一つまたは複数のシーケンスを実行するのに応答して、コンピュータ・システム 700 によって実行される。そのような命令は、記憶装置 710 のような別の記憶媒体からメイン・メモリ 706 に読み込まれてもよい。メイン・メモリ 706 に含まれる命令のシーケンスの実行により、プロセッサ 704 は、本稿に記載されるプロセス段階を実行する。代替的な実施形態では、ソフトウェア命令の代わりにまたはソフトウェア命令と組み合わせて固定構成の回路が使用されてもよい。

20

【0115】

本稿で用いられる用語「記憶媒体」は、データおよび / または機械に特定の仕方で動作させる命令を記憶する任意の非一時的な媒体を指す。そのような記憶媒体は、不揮発性媒体および / または揮発性媒体を含んでいてもよい。不揮発性媒体は、たとえば、記憶装置 710 のような光学式または磁気ディスクを含む。揮発性媒体は、メイン・メモリ 706 のような動的メモリを含む。記憶媒体の一般的な形は、たとえば、フロッピーディスク、フレキシブルディスク、ハードディスク、半導体ドライブ、磁気テープまたは他の任意の磁気データ記憶媒体、CD-ROM、他の任意の光学式データ記憶媒体、孔のパターンをもつ任意の物理的媒体、RAM、PROM および EPROM、フラッシュ EPROM、NVRAM、他の任意のメモリ・チップまたはカートリッジを含む。

30

【0116】

記憶媒体は、伝送媒体とは異なるが、伝送媒体と関連して用いられてもよい。伝送媒体は、記憶媒体間で情報を転送するのに参加する。たとえば、伝送媒体は同軸ケーブル、銅線および光ファイバーを含み、バス 702 をなすワイヤを含む。伝送媒体は、電波および赤外線データ通信の際に生成されるような音響波または光波の形を取ることもできる。

40

【0117】

さまざまな形の媒体が、一つまたは複数の命令の一つまたは複数のシーケンスを実行のためにプロセッサ 704 に搬送するのに関与しうる。たとえば、命令は最初、リモート・コンピュータの磁気ディスクまたは半導体ドライブ上に担持されていてもよい。リモート・コンピュータは該命令をその動的メモリにロードし、該命令をモデムを使って電話線を通じて送ることができる。コンピュータ・システム 700 にローカルなモデムが、電話線上のデータを受信し、赤外線送信器を使ってそのデータを赤外線信号に変換することができる。赤外線検出器が赤外線信号において担持されるデータを受信することができ、適切な回路がそのデータをバス 702 上に載せることができる。バス 702 はそのデータをメ

50

イン・メモリ706に搬送し、メイン・メモリ706から、プロセッサ704が命令を取り出し、実行する。メイン・メモリ706によって受信される命令は、任意的に、プロセッサ704による実行の前または後に記憶装置710上に記憶されてもよい。

【0118】

コンピュータ・システム700は、バス702に結合された通信インターフェース718をも含む。通信インターフェース718は、ローカル・ネットワーク722に接続されているネットワーク・リンク720への双方向データ通信結合を提供する。たとえば、通信インターフェース718は、統合サービス・デジタル通信網（ISDN）カード、ケーブル・モデム、衛星モデムまたは対応する型の電話線へのデータ通信接続を提供するためのモデムであってもよい。もう一つの例として、通信インターフェース718は、互換LANへのデータ通信接続を提供するためのローカル・エリア・ネットワーク（LAN）カードであってもよい。無線リンクも実装されてもよい。そのようないかなる実装でも、通信インターフェース718は、さまざまな型の情報を表すデジタル・データ・ストリームを搬送する電気的、電磁的または光学的信号を送受信する。

10

【0119】

ネットワーク・リンク720は典型的には、一つまたは複数のネットワークを通じた他のデータ装置へのデータ通信を提供する。たとえば、ネットワーク・リンク720は、ローカル・ネットワーク722を通じてホスト・コンピュータ724またはインターネット・サービス・プロバイダー（ISP）726によって運営されているデータ設備への接続を提供してもよい。ISP726は、現在一般に「インターネット」728と称される世界規模のパケット・データ通信網を通じたデータ通信サービスを提供する。ローカル・ネットワーク722およびインターネット728はいずれも、デジタル・データ・ストリームを担持する電気的、電磁的または光学的信号を使う。コンピュータ・システム700に/からデジタル・データを搬送する、さまざまなネットワークを通じた信号およびネットワーク・リンク720上および通信インターフェース718を通じた信号は、伝送媒体の例示的な形である。

20

【0120】

コンピュータ・システム700は、ネットワーク（単数または複数）、ネットワーク・リンク720および通信インターフェース718を通じて、メッセージを送り、プログラム・コードを含めデータを受信することができる。インターネットの例では、サーバー730は、インターネット728、ISP726、ローカル・ネットワーク722および通信インターフェース718を通じてアプリケーション・プログラムのための要求されたコードを送信してもよい。

30

【0121】

受信されたコードは、受信される際にプロセッサ704によって実行されても、および/または、のちの実行のために記憶装置710または他の不揮発性記憶に記憶されてもよい。

【0122】

9. 等価物、拡張、代替その他

以上の明細書では、本発明の例示的実施形態について、実装によって変わりうる数多くの個別的詳細に言及しつつ述べてきた。このように、何が本発明であるか、何が出願人によって本発明であると意図されているかの唯一にして排他的な指標は、この出願に対して付与される特許の請求項の、その後の訂正があればそれも含めてかかる請求項が特許された特定の形のものである。かかる請求項に含まれる用語について本稿で明示的に記載される定義があったとすればそれは請求項において使用される当該用語の意味を支配する。よって、請求項に明示的に記載されていない限定、要素、属性、特徴、利点もしくは特性は、いかなる仕方であれかかる請求項の範囲を限定すべきではない。よって、明細書および図面は制約する意味ではなく例示的な意味で見なされるべきものである。

40

いくつかの態様を記載しておく。

〔態様1〕

50

一つまたは複数のフレームにおける入力オーディオ・コンテンツに存在している複数のオーディオ・オブジェクトを判別する段階と；

前記一つまたは複数のフレームにおける出力オーディオ・コンテンツに存在している複数の出力クラスターを判別する段階であって、前記入力オーディオ・コンテンツにおける前記複数のオーディオ・オブジェクトが前記出力オーディオ・コンテンツにおける前記複数の出力クラスターに変換される、段階と；

少なくとも部分的には前記複数のオーディオ・オブジェクトの位置メタデータおよび前記複数の出力クラスターの位置メタデータに基づいて、一つまたは複数の空間的誤差メトリックを計算する段階とを含む、

一つまたは複数のコンピューティング装置によって実行される方法。

10

〔態様 2〕

前記一つまたは複数の空間的誤差メトリックは少なくとも部分的にはオブジェクト重要度に依存する、態様 1 記載の方法。

〔態様 3〕

前記オブジェクト重要度が、前記複数のオーディオ・オブジェクトにおけるオーディオ・データ、前記複数の出力クラスターにおけるオーディオ・データ、前記複数のオーディオ・オブジェクトにおけるメタデータまたは前記複数の出力クラスターにおけるメタデータの一つまたは複数を解析することから得られる、態様 2 記載の方法。

〔態様 4〕

前記オブジェクト重要度の少なくとも一部がユーザー入力に基づいて決定される、態様 2 記載の方法。

20

〔態様 5〕

前記複数のオーディオ・オブジェクトにおける少なくとも一つのオーディオ・オブジェクトが前記複数の出力クラスターにおける二つ以上の出力クラスターに配分される、態様 1 記載の方法。

〔態様 6〕

前記複数のオーディオ・オブジェクトにおける少なくとも一つのオーディオ・オブジェクトが、前記複数の出力クラスターにおける出力クラスターに割り当てられる、態様 1 記載の方法。

〔態様 7〕

前記入力オーディオ・コンテンツにおける前記複数のオーディオ・オブジェクトを前記出力クラスターにおける前記複数の出力クラスターに変換することによって引き起こされる知覚的オーディオ品質劣化を、前記一つまたは複数の空間的誤差メトリックに基づいて決定する段階をさらに含む、

態様 1 記載の方法。

30

〔態様 8〕

前記知覚的オーディオ品質劣化は、知覚的オーディオ品質試験に関係する一つまたは複数の予測された試験スコアによって表わされる、態様 7 記載の方法。

〔態様 9〕

前記一つまたは複数の空間的誤差メトリックは：フレーム内空間的誤差メトリックまたはフレーム間空間的誤差メトリックの少なくとも一方を含む、態様 1 記載の方法。

40

〔態様 10〕

前記フレーム内空間的誤差メトリックは：フレーム内オブジェクト位置誤差メトリック、フレーム内オブジェクト・パン誤差メトリック、重要度で重み付けされたフレーム内オブジェクト位置誤差メトリック、重要度で重み付けされたフレーム内オブジェクト・パン誤差メトリック、規格化されたフレーム内オブジェクト位置誤差メトリックまたは規格化されたフレーム内オブジェクト・パン誤差メトリックのうちの少なくとも一つを含む、態様 9 記載の方法。

〔態様 11〕

前記フレーム間空間的誤差メトリックは：利得係数フローに基づくフレーム間空間的誤

50

差メトリックまたは利得係数フローに基づかないフレーム間空間的誤差メトリックのうちの少なくとも一つを含む、態様 9 記載の方法。

[態様 1 2]

前記フレーム間空間的誤差メトリックのそれぞれは二つ以上の異なるフレームに関して計算される、態様 9 記載の方法。

[態様 1 3]

前記複数のオーディオ・オブジェクトは複数の利得係数を介して前記複数の出力クラスターに関係する、態様 1 記載の方法。

[態様 1 4]

前記フレームのそれぞれは、前記入力オーディオ・コンテンツにおけるある時間セグメントおよび前記出力オーディオ・コンテンツにおける第二の時間セグメントに対応し、前記出力オーディオ・コンテンツにおける前記第二の時間セグメントに存在する出力クラスターは、前記入力オーディオ・コンテンツにおける前記第一の時間セグメントに存在するオーディオ・オブジェクトによってマッピングされる、態様 1 記載の方法。

10

[態様 1 5]

前記一つまたは複数のフレームが二つの連続するフレームを含む、態様 1 記載の方法。

[態様 1 6]

前記複数のオーディオ・オブジェクトのうちのオーディオ・オブジェクトまたは聴取空間内の前記複数の出力クラスターにおける出力クラスターの一つまたは複数を表わす一つまたは複数のユーザー・インターフェース構成要素を構築する段階と；

20

前記一つまたは複数のユーザー・インターフェース構成要素をユーザーに対して表示させる段階とをさらに含む、

態様 1 記載の方法。

[態様 1 7]

前記一つまたは複数のユーザー・インターフェース構成要素におけるあるユーザー・インターフェース構成要素は、前記複数のオーディオ・オブジェクトのうちのあるオーディオ・オブジェクトを表わし；該オーディオ・オブジェクトは前記複数の出力クラスターにおける一つまたは複数の出力クラスターにマッピングされ；前記ユーザー・インターフェース構成要素の少なくとも一つの視覚的特徴が前記オーディオ・オブジェクトの前記一つまたは複数の出力クラスターへのマッピングに関係した一つまたは複数の空間的誤差の総量を表わす、態様 1 6 記載の方法。

30

[態様 1 8]

前記一つまたは複数のユーザー・インターフェース構成要素は、三次元（3D）形式での聴取空間の表現を有する、態様 1 6 記載の方法。

[態様 1 9]

前記一つまたは複数のユーザー・インターフェース構成要素は、二次元（2D）形式での聴取空間の表現を有する、態様 1 6 記載の方法。

[態様 2 0]

前記複数のオーディオ・オブジェクトにおけるオーディオ・オブジェクトのそれぞれのオブジェクト重要性、前記複数の出力クラスターにおける出力クラスターのそれぞれのオブジェクト重要性、前記複数のオーディオ・オブジェクトにおけるオーディオ・オブジェクトのそれぞれのラウドネス、前記複数の出力クラスターにおける出力クラスターのそれぞれのラウドネス、前記複数のオーディオ・オブジェクトにおけるオーディオ・オブジェクトのそれぞれの、発話もしくはダイアログ・コンテンツの確率、前記複数の出力クラスターにおける出力クラスターの発話もしくはダイアログ・コンテンツの確率のうちの一つまたは複数を表わす一つまたは複数のユーザー・インターフェース構成要素を構築する段階と；

40

前記一つまたは複数のユーザー・インターフェース構成要素をユーザーに対して表示させる段階とをさらに含む、

態様 1 記載の方法。

50

〔 態 様 2 1 〕

前記一つまたは複数の空間的誤差メトリックまたは少なくとも部分的には前記一つまたは複数の空間的誤差メトリックに基づいて決定された一つまたは複数の予測された試験スコアの一つまたは複数を表わす一つまたは複数のユーザー・インターフェース構成要素を構築する段階と；

前記一つまたは複数のユーザー・インターフェース構成要素をユーザーに対して表示させる段階とをさらに含む、

〔 態 様 1 記 載 の 方 法。〕

〔 態 様 2 2 〕

変換プロセスが前記入力オーディオ・コンテンツにおいて存在する時間依存のオーディオ・オブジェクトを、前記出力クラスターをなす時間依存の出力クラスターに変換し、前記一つまたは複数のユーザー・インターフェース構成要素は、前記一つまたは複数のフレームを含み前記一つまたは複数のフレームまでの過去の時間区間について前記変換プロセスにおいて生じる最悪のオーディオ品質劣化の視覚的指示を含む、態様 2 1 記載の方法。

10

〔 態 様 2 3 〕

前記一つまたは複数のユーザー・インターフェース構成要素は、前記一つまたは複数のフレームを含み前記一つまたは複数のフレームまでの過去の時間区間について変換プロセスにおいて生じるオーディオ品質劣化がオーディオ品質劣化閾値を超えたことの視覚的指示を含む、態様 2 1 記載の方法。

〔 態 様 2 4 〕

20

前記一つまたは複数のユーザー・インターフェース構成要素は、前記一つまたは複数のフレームにおけるオーディオ品質劣化を示す高さをもつ垂直の棒を含み、前記垂直の棒は前記一つまたは複数のフレームにおけるオーディオ品質劣化に基づいてカラーコーディングされる、態様 2 1 記載の方法。

〔 態 様 2 5 〕

前記複数の出力クラスターにおけるある出力クラスターは、前記複数のオーディオ・オブジェクトにおける二つ以上のオーディオ・オブジェクトによってマッピングされる部分を含む、態様 1 記載の方法。

〔 態 様 2 6 〕

30

前記複数のオーディオ・オブジェクトにおけるオーディオ・オブジェクトまたは前記複数の出力クラスターにおける出力クラスターの少なくとも一つが、時間とともに変化する動的位置をもつ、態様 1 記載の方法。

〔 態 様 2 7 〕

前記複数のオーディオ・オブジェクトにおけるオーディオ・オブジェクトまたは前記複数の出力クラスターにおける出力クラスターの少なくとも一つが、時間とともに変化しない固定した位置をもつ、態様 1 記載の方法。

〔 態 様 2 8 〕

前記入力オーディオ・コンテンツまたは前記出力オーディオ・コンテンツの少なくとも一つは、オーディオのみ信号またはオーディオビジュアル信号の一方の一部である、態様 1 記載の方法。

40

〔 態 様 2 9 〕

前記入力オーディオ・コンテンツを前記出力オーディオ・コンテンツに変換する変換プロセスに対する変更を指定するユーザー入力を受領する段階と；

前記ユーザー入力を受領するのに応答して、前記入力オーディオ・コンテンツを前記出力オーディオ・コンテンツに変換する前記変換プロセスに対する前記変更を引き起こす段階とをさらに含む、

〔 態 様 1 記 載 の 方 法。〕

〔 態 様 3 0 〕

当該方法が、前記変換プロセスが前記入力オーディオ・コンテンツを前記出力オーディオ・コンテンツに変換している間に並行して実行される、態様 2 9 記載の方法。

50

【態様 3 1】

態様 1 ないし 3 0 のうちいずれか一項記載の方法を実行するよう構成されたメディア処理システム。

【態様 3 2】

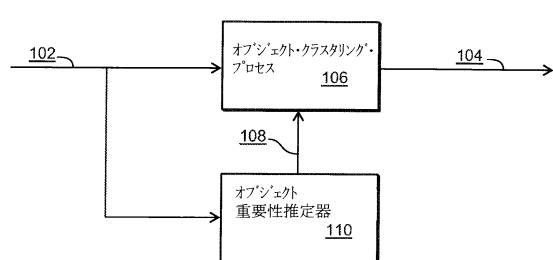
態様 1 ないし 3 0 のうちいずれか一項記載の方法を実行するよう構成された、プロセッサを有する装置。

【態様 3 3】

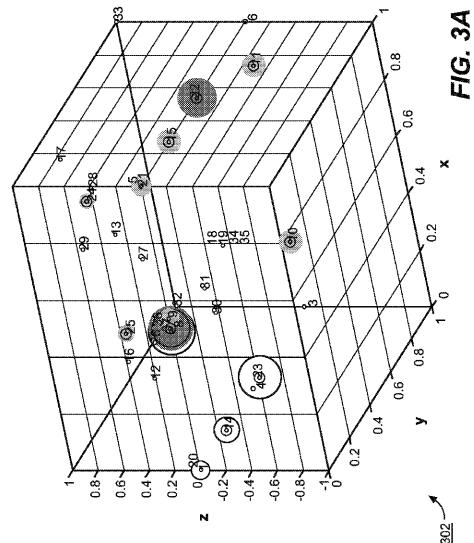
一つまたは複数のプロセッサによって実行されたときに、態様 1 ないし 3 0 のうちいずれか一項記載の方法の実行を引き起こすソフトウェア命令を記憶している非一時的なコンピュータ可読記憶媒体。

10

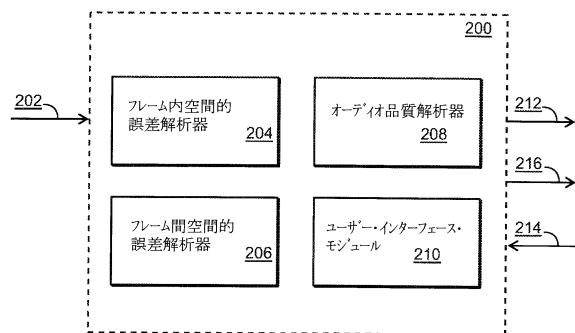
【図 1】



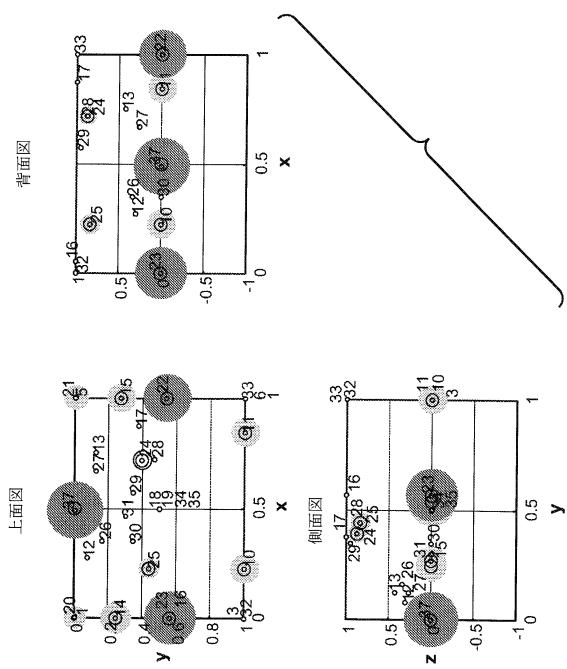
【図 3 A】



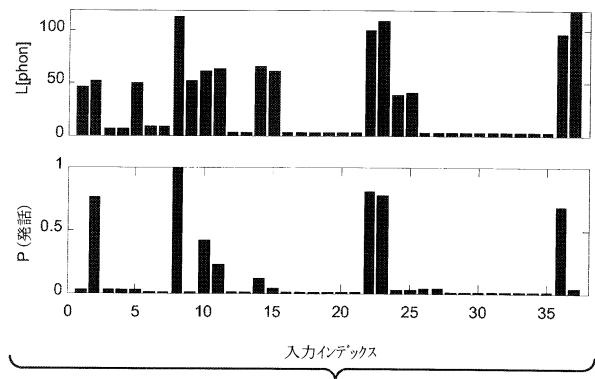
【図 2】



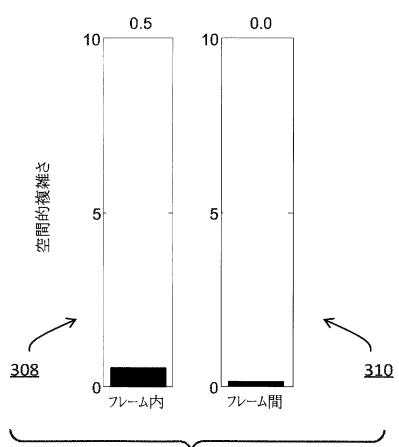
【図3B】



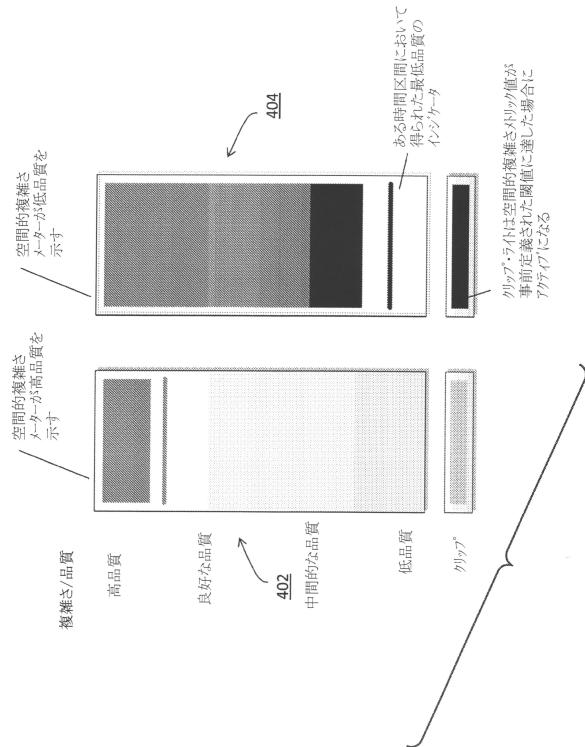
【図3C】



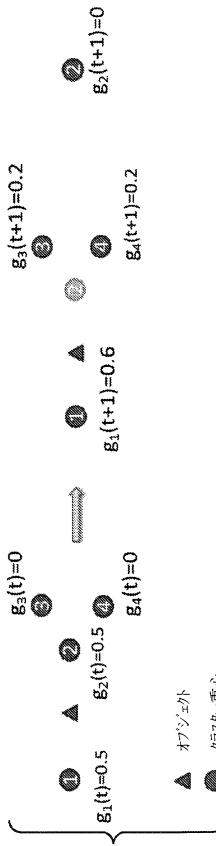
【図3D】



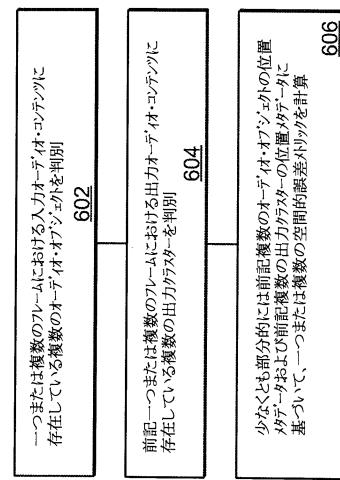
【図4】



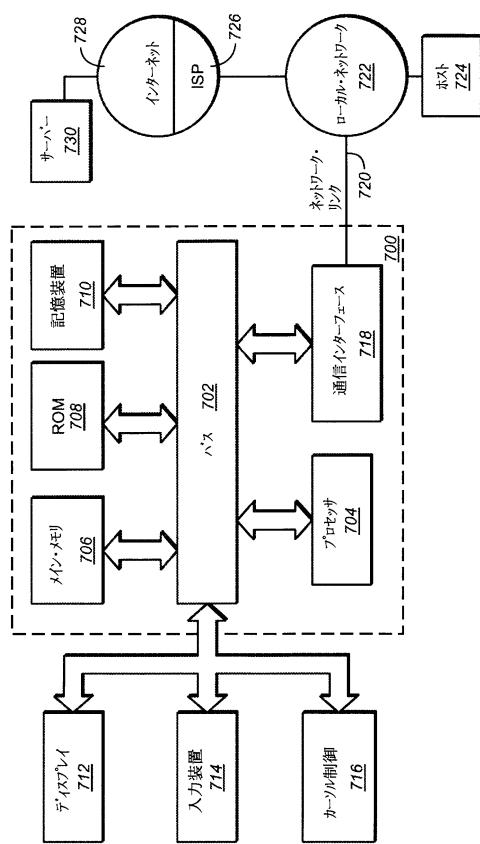
【図5】



【図6】



【図7】



フロントページの続き

(74)代理人 100070150
弁理士 伊東 忠彦

(74)代理人 100091214
弁理士 大貫 進介

(72)発明者 ブリーバルト, ディルク ジェロエン
オーストラリア国, ニュー・サウス・ウェールズ 2060, マクマホンズ・ポイント, ミッチャール・ストリート 35 レベル 3, 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 チエン, リアンワー
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 サービシーズ(ベイジン)コーポレイション 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 サービシーズ(ベイジン)コーポレイション リミテッド内

(72)発明者 ソレ, アントニオ マテオス
スペイン国, バルセロナ 08018, アヴェニダ ディアゴナル, 177, プランタ 10, エディフィシオ イマギナ, ドルビー イベリア エセ エレ内

(72)発明者 トワインゴ, ニコラ エール.
アメリカ合衆国, カリフォルニア州 94103, サンフランシスコ, マーケット ストリート 1275, ドルビー ラボラトリーズ インコーポレイテッド内

審査官 山下 剛史

(56)参考文献 米国特許出願公開第2005/0114121(US, A1)
特表2010-511190(JP, A)
特表2012-525600(JP, A)
特開2011-209588(JP, A)

(58)調査した分野(Int.Cl., DB名)

G10L 19/00 - 19/26