(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第6654195号
(P6654195)

(45)発行日　令和2年2月26日(2020.2.26)　(24)登録日　令和2年1月31日(2020.1.31)

(51)Int.Cl.　F I

G10L 21/028 (2013.01)	G10L	21/028	C
G10L 21/0308 (2013.01)	G10L	21/0308	A
G10L 19/008 (2013.01)	G10L	19/008	100
	G10L	19/008	200

請求項の数 11　(全 23 頁)

(21)出願番号　特願2017-533302(P2017-533302)
(86)(22)出願日　平成27年12月18日(2015.12.18)
(65)公表番号　特表2018-503864(P2018-503864A)
(43)公表日　平成30年2月8日(2018.2.8)
(86)国際出願番号　PCT/US2015/066798
(87)国際公開番号　WO2016/106145
(87)国際公開日　平成28年6月30日(2016.6.30)
審査請求日　平成30年11月30日(2018.11.30)
(31)優先権主張番号　201410814973.9
(32)優先日　平成26年12月22日(2014.12.22)
(33)優先権主張国・地域又は機関
中国(CN)
(31)優先権主張番号　62/108,254
(32)優先日　平成27年1月27日(2015.1.27)
(33)優先権主張国・地域又は機関
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
(74)代理人　100107766
弁理士　伊東　忠重
(74)代理人　100070150
弁理士　伊東　忠彦
(74)代理人　100091214
弁理士　大貫　進介

最終頁に続く

(54)【発明の名称】オーディオ・コンテンツからの投影ベースのオーディオ・オブジェクト抽出

(57)【特許請求の範囲】

【請求項１】

複数のチャネルのオーディオ信号によって表現されているオーディオ・コンテンツからのオーディオ・オブジェクト抽出のための方法であって：

前記複数のチャネルの相関に基づいて前記複数のチャネルをクラスタリングし、それによりチャネル・グループを得る段階と；

前記チャネル・グループから第一のチャネルおよび第二のチャネルを選択する段階と；

前記第一のチャネルについての第一の部分集合および前記第二のチャネルについての第二の部分集合を含む投影空間の第一の集合を同定する段階と；

前記第一および第二のチャネルの間の相関の第一の集合を決定する段階であって、前記第一の集合の相関のそれぞれは前記第一の部分集合の投影空間の一つおよび前記第二の部分集合の投影空間の一つに対応する、段階と；

前記第一の集合の相関のうちで最大の値をもつ前記第一の集合の相関のうちの第一の相関を同定する段階と；

少なくとも部分的には前記第一の相関と、前記第一の相関に対応する前記第一の部分集合からの投影空間とに基づいて、前記第一のチャネルのオーディオ信号からオーディオ・オブジェクトを抽出する段階とを含む、

方法。

【請求項２】

クラスタリングにおいて、前記複数のチャネルの一対のチャネルの間の相関が：

投影空間の第二の集合であって、その対のチャネルの一方についての第三の部分集合およびその対のチャネルの他方についての第四の部分集合を含むものを同定する段階と；
その対のチャネルの間の相関の第二の集合を決定する段階であって、前記第二の集合の相関のそれぞれは前記第三の部分集合の投影空間の一つおよび前記第四の部分集合の投影空間の一つに対応する、段階と；
前記第二の集合の相関のうちの一つを、その対のチャネルの間の相関として選択する段階であって、選択される相関は第二のあらかじめ定義された閾値より大きい、段階とを実行することによって決定される、
請求項１記載の方法。
【請求項３】
前記チャネル・グループから前記第一および第二のチャネルを選択することは：
前記第二のチャネルのオーディオ信号が前記チャネル・グループにおいて最大のエネルギーをもつように、前記チャネル・グループから前記第二のチャネルを選択することを含む、
請求項１記載の方法。
【請求項４】
前記チャネル・グループから前記第一および第二のチャネルを選択することはさらに：
前記第一および第二のチャネルの間の相関が第三のあらかじめ定義された閾値より大きいように前記チャネル・グループから前記第一のチャネルを選択し；
前記第二のチャネルのオーディオ信号から、少なくとも部分的には前記第一の相関と前記第一の相関に対応する前記第二の部分集合からの投影空間とに基づいてオーディオ・オブジェクトを抽出することを含む、
請求項３記載の方法。
【請求項５】
前記オブジェクト抽出の検証および調整を：
少なくとも部分的には前記第一および第二のチャネルのオーディオ信号からの抽出されたオーディオ・オブジェクトに基づいて、マルチチャネル・オブジェクトを生成し；
生成されたマルチチャネル・オブジェクトをモノ表現にダウンミックスし；
抽出されたオブジェクトのもとのマルチチャネル表現と前記モノ表現との間のダウンミックス類似性を決定し；
少なくとも部分的には前記第一の集合の相関のうちの前記第一の相関と、前記第一の相関に対応する前記第一の部分集合からの投影空間とに基づき、前記ダウンミックス類似性にさらに基づいて、前記第一のチャネルのオーディオ信号から前記オーディオ・オブジェクトを抽出することによって行なうことを含む、
請求項１記載の方法。
【請求項６】
前記オブジェクト抽出の検証および調整を：
少なくとも部分的には前記第一および第二のチャネルのオーディオ信号からの抽出されたオーディオ・オブジェクトに基づいて、マルチチャネル・オブジェクトを生成し；
生成されたマルチチャネル・オブジェクトをモノ表現にダウンミックスし；
少なくとも部分的には推定位置に基づいて前記モノ表現をプリ・レンダリングし；
抽出されたオブジェクトのもとのマルチチャネル表現と前記プリ・レンダリングされたモノ表現との間の、エネルギー分布に関するプリ・レンダリング類似性を決定し；
少なくとも部分的には前記第一の集合の相関のうちの前記第一の相関と、前記第一の相関に対応する前記第一の部分集合からの投影空間とに基づき、前記プリ・レンダリング類似性にさらに基づいて、前記第一のチャネルのオーディオ信号から前記オーディオ・オブジェクトを抽出することによって行なうことを含む、
請求項１記載の方法。
【請求項７】
前記オーディオ・コンテンツは、フル帯域オーディオ信号のフレームを、周波数領域お

よび時間領域の少なくとも一方において分割することによって得られる一つまたは複数のオーディオ・ブロックを含む、請求項１ないし６のうちいずれか一項記載の方法。

【請求項８】

複数のチャネルのオーディオ信号によって表現されているオーディオ・コンテンツからのオーディオ・オブジェクト抽出のためのシステムであって：

前記複数のチャネルの相関に基づいて前記複数のチャネルをクラスタリングすることによって得られるチャネル・グループから、第一のチャネルおよび第二のチャネルを選択するよう構成された選択ユニットと；

前記第一のチャネルについての第一の部分集合および前記第二のチャネルについての第二の部分集合を含む投影空間の第一の集合を同定するよう構成された同定ユニットと；

前記第一および第二のチャネルの間の相関の第一の集合を決定するよう構成された決定ユニットであって、前記第一の集合の相関のそれぞれは前記第一の部分集合の投影空間の一つおよび前記第二の部分集合の投影空間の一つに対応する、決定ユニットと；

前記第一の集合の相関のうちで最大の値をもつ前記第一の集合の相関のうちの第一の相関を同定し、少なくとも部分的には前記第一の相関と、前記第一の相関に対応する前記第一の部分集合からの投影空間とに基づいて、前記第一のチャネルのオーディオ信号からオーディオ・オブジェクトを抽出するよう構成された抽出ユニットとを有する、

システム。

【請求項９】

少なくとも部分的には前記第一および第二のチャネルのオーディオ信号からの抽出されたオーディオ・オブジェクトに基づいて、マルチチャネル・オブジェクトを生成するよう構成された生成ユニットと；

生成されたマルチチャネル・オブジェクトをモノ表現にダウンミックスするよう構成されたダウンミックス・ユニットと；

抽出されたオブジェクトのもとのマルチチャネル表現と前記モノ表現との間のダウンミックス類似性を決定するよう構成された類似性決定ユニットとをさらに有しており、

前記抽出ユニットは、少なくとも部分的には前記第一の集合の相関のうちの前記第一の相関と、前記第一の相関に対応する前記第一の部分集合からの投影空間とに基づき、前記ダウンミックス類似性にさらに基づいて前記第一のチャネルのオーディオ信号から前記オーディオ・オブジェクトを再抽出するよう構成されている、

請求項８記載のシステム。

【請求項１０】

少なくとも部分的には前記第一および第二のチャネルのオーディオ信号からの抽出されたオーディオ・オブジェクトに基づいて、マルチチャネル・オブジェクトを生成するよう構成された生成ユニットと；

生成されたマルチチャネル・オブジェクトをモノ表現にダウンミックスするよう構成されたダウンミックス・ユニットと；

少なくとも部分的には推定位置に基づいて前記モノ表現をプリ・レンダリングするよう構成されたプリ・レンダリング・ユニットと；

抽出されたオブジェクトのもとのマルチチャネル表現と前記プリ・レンダリングされたモノ表現との間の、エネルギー分布に関するプリ・レンダリング類似性を決定するよう構成された類似性決定ユニットとをさらに有しており、

前記抽出ユニットは、少なくとも部分的には前記第一の集合の相関のうちの前記第一の相関と、前記第一の相関に対応する前記第一の部分集合からの投影空間とに基づき、前記プリ・レンダリング類似性にさらに基づいて前記第一のチャネルのオーディオ信号から前記オーディオ・オブジェクトを再抽出するよう構成されている、請求項８記載のシステム。

【請求項１１】

コンピュータに請求項１ないし７のうちいずれか一項記載の方法を実行させるためのコンピュータ・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2014年12月22日に出願された中国優先権出願第201410814973.9号および2015年1月27日に出願された米国仮特許出願第62/108,254号の優先権を主張するものである。これら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。
【０００２】
技術
本稿に開示される例示的実施形態は概括的にはオーディオ・コンテンツ処理に関し、より詳細にはオーディオ・コンテンツからのオーディオ・オブジェクト抽出のための方法およびシステムに関する。
【背景技術】
【０００３】
伝統的に、オーディオ・コンテンツはチャネル・ベースのフォーマットで作成され、記憶される。チャネル・ベースのフォーマットでは、オーディオ・コンテンツは通例、チャネルの媒体によって表現され、記憶され、伝達され、頒布される。本稿での用法では、用語「オーディオ・チャネル」または「チャネル」は、通例あらかじめ定義された物理的位置をもつオーディオ・コンテンツをいう。たとえば、ステレオ、サラウンド5.1、サラウンド7.1などはみな、オーディオ・コンテンツのためのチャネル・ベースのフォーマットである。各チャネルは固定位置の物理的スピーカーに対応する。マルチチャネル・コンテンツが再生されるとき、複数のスピーカーがライブで、没入的な音場を聴取者のまわりに作り出す。近年、いくつかの通常のマルチチャネル・システムは、チャネルおよびオーディオ・オブジェクトの両方を含む新しいフォーマットをサポートするよう拡張された。本稿での用法では、用語「オーディオ・オブジェクト」または「オブジェクト」は、定義された継続時間にわたって音場において存在する個別のオーディオ要素をいう。たとえば、オーディオ・オブジェクトはダイアログ、発砲音、雷鳴などを表わしてもよい。これらのオブジェクトは通例、その所望されるサウンド効果を作り出すためにミキサーによって使われる。たとえば、ダイアログは通例、中央前方に定位され、雷の音は通例頭上から発する。人間によるオブジェクトの位置の知覚は、同じオブジェクトのオーディオ信号を再生している複数のスピーカーの発射の結果である。たとえば、オブジェクトが前方左スピーカーおよび前方右スピーカーによって同様のエネルギー・レベルで再生されるときは、人は中央前方からのファントムを知覚する。
【０００４】
上述したように、コンテンツがチャネル・ベースのフォーマットで作成されるとき、それは通例、ミキサーによって特定の再生セッティングのために知覚経験が最適化されることを意味する。しかしながら、異なる再生セッティングによって再生されるときは、その聴取経験は、再生セッティング間の不一致のため劣化することがある。劣化の例は、オブジェクトの位置が変わることがあるということである。このように、チャネル・ベースのフォーマットは、多様なスピーカー再生構成に適応するには非効率的である。非効率性のもう一つの側面は、バイノーラル・レンダリングにある。ここでは、チャネル・ベースのフォーマットはスピーカー位置に固有の限られた数の頭部伝達関数（HRTF）を使うことができるだけであり、他の位置については、HRTFの補間が使われ、バイノーラル聴取経験を劣化させる。
【０００５】
この問題に対処する一つの潜在的な方法は、チャネル・ベースの表現から、位置とモノのクリーンな波形とを含むもとの源（またはオブジェクト）を復元し、それらの位置をメタデータとして、スピーカー再生装置のパン・アルゴリズムを制御するために使い、それによりオンザフライでオブジェクトを再レンダリングし、もとの音像と同様の音像を作り出すことである。バイノーラル・レンダリング・セッティングについては（限られた数の

HRTFを使う代わりに）、聴取経験をさらに向上させるために最も適切な諸HRTFを選ぶために位置を使うことができる。

【０００６】

しかしながら、メタデータを用いてレンダリングされるべきチャネル・ベースの表現におけるオブジェクトは常にクリーンであるとは限らない。かかるオブジェクトは、いくつかのチャネル内で他のオブジェクトと同時に混合されることがある。たとえば、芸術的意図を実施するために、ミキサーは二つのオブジェクトを聴取者の前方に同時に置いて、一方は中央と前方左の間に、他方は中央と前方右の間の何らかの位置に現われるようにすることがある。これは中央前方チャネルに二つのオブジェクトを含ませることができる。源分離技法が使われない場合には、これら二つのオブジェクトは一つのオブジェクトとみなされ、これはそれらの位置推定を誤らせることになる。

【０００７】

このように、クリーンなオブジェクトを得てその位置を推定するためには、オブジェクトをそのマルチチャネル混合から分離してクリーンなマルチチャネルまたはモノ表現を生成する源分離技法が必要とされる。上述した例において、単一のマルチチャネル入力が源分離コンポーネントによって、たとえばそれぞれ一つのクリーンなオブジェクトを含んでいるだけの二つのマルチチャネルまたはモノ出力に分割されることが所望される。

【発明の概要】

【発明が解決しようとする課題】

【０００８】

上記および他の潜在的な問題に対処するために、本稿に開示される例示的実施形態は、オーディオ・コンテンツからオーディオ・オブジェクトを抽出するための方法およびシステムを提案する。

【課題を解決するための手段】

【０００９】

ある側面では、例示的実施形態は、複数のチャネルのオーディオ信号によって表現されているオーディオ・コンテンツからのオーディオ・オブジェクト抽出のための方法を提供する。本方法は、前記複数のチャネルの第一のチャネルについての第一の部分集合および第二のチャネルについての第二の部分集合を含む投影空間の第一の集合を同定することを含む。本方法はさらに、前記第一および第二のチャネルの間の相関の第一の集合を決定することを含み、前記第一の集合の相関のそれぞれは前記第一の部分集合の投影空間の一つおよび前記第二の部分集合の投影空間の一つに対応する。本方法はまた、少なくとも部分的には相関の前記第一の集合のうちの第一の相関と、前記第一の相関に対応する前記第一の部分集合からの投影空間とに基づいて、前記第一のチャネルのオーディオ信号からオーディオ・オブジェクトを抽出することも含み、前記第一の相関は第一のあらかじめ定義された閾値より大きい。これに関する諸実施形態はさらに、対応するコンピュータ・プログラム・プロダクトを含む。

【００１０】

別の側面では、例示的実施形態は、複数のチャネルのオーディオ信号によって表現されているオーディオ・コンテンツからのオーディオ・オブジェクト抽出のためのシステムを提供する。本システムは、前記複数のチャネルの第一のチャネルについての第一の部分集合および第二のチャネルについての第二の部分集合を含む投影空間の第一の集合を同定するよう構成された同定ユニットを含む。本システムはさらに、前記第一および第二のチャネルの間の相関の第一の集合を決定するよう構成された決定ユニットを含み、前記第一の集合の相関のそれぞれは前記第一の部分集合の投影空間の一つおよび前記第二の部分集合の投影空間の一つに対応する。本システムはまた、少なくとも部分的には相関の前記第一の集合のうちの第一の相関と、前記第一の相関に対応する前記第一の部分集合からの投影空間とに基づいて、前記第一のチャネルのオーディオ信号からオーディオ・オブジェクトを抽出するよう構成された抽出ユニットも含み、前記第一の相関は第一のあらかじめ定義された閾値より大きい。

【００１１】
以下の記述を通じて、本稿に開示される例示的実施形態によれば、オーディオ・オブジェクトが複数チャネルに基づくオーディオ・コンテンツのオーディオ信号のそれぞれから分離されることができることが理解されるであろう。このようにして、オーディオ・コンテンツ入力が、その聴取経験を劣化させることなく、多様な再生構成に適応することが可能となる。例示的実施形態によって達成される他の利点は以下の記述を通じて明白となるであろう。
【図面の簡単な説明】
【００１２】
付属の図面を参照しての以下の詳細な説明を通じて、例示的実施形態の上記および他の目的、特徴および利点がよりわかりやすくなるであろう。図面においては、いくつかの例示的実施形態が限定ではなく例として示される。
【図１】複数のチャネルに基づくフォーマットのオーディオ信号のセグメントの例を示す図である。
【図２】例示的実施形態に基づくオーディオ・コンテンツからのオーディオ・オブジェクト抽出のための方法のフローチャートである。
【図３】ある例示的実施形態に基づくオーディオ・コンテンツからのオーディオ・オブジェクト抽出のためのシステム３００のブロック図である。
【図４】例示的実施形態を実装するのに好適な例示的なコンピュータ・システムのブロック図である。　図面を通じて、同じまたは対応する参照記号は同じまたは対応する部分を指す。
【発明を実施するための形態】
【００１３】
例示的実施形態の原理についてこれから図面に示されるさまざまな例示的実施形態を参照して述べる。これらの実施形態の描出は、単に当業者が例示的実施形態をよりよく理解し、さらに実装することができるようにするためだけであり、いかなる仕方であれ本稿に開示される例示的実施形態の範囲を限定することは意図されていないことは理解しておくべきである。また、「第一」、「第二」などの用語は異なる対象を指示するために使われているのであり、対象の序列についていかなる限定をも示唆するものではない。
【００１４】
上述したように、レガシーのチャネル・ベースのオーディオ・コンテンツは多様な再生セッティングに適応するには不十分である。特に、再生セッティングがミキサーの構成と一致しない場合、該再生セッティングによって表現される聴取経験は劣化する。さらに、芸術的意図を保存しつつ該再生セッティングでオーディオ・コンテンツを表現することも、オブジェクト分離技法にとっての課題となる。
【００１５】
したがって、チャネル・ベースのオーディオ・コンテンツからできるだけクリーンなオーディオ・オブジェクトを抽出することが所望される。図１は、複数のチャネルに基づくフォーマットのオーディオ信号のセグメントの例を示している。図１に示されるように、オーディオ信号のセグメント１００は時間および周波数領域において表現されている。横軸によって表現される時間領域では、オーディオ信号のセグメント１００は時間軸Tに沿っていくつかのフレームを含みうる。フレームはたとえば、t1からt2の時間長さであってもよい。オブジェクト抽出におけるその後の計算およびプロセスの便利のため、オーディオ信号のフレームは時間軸に沿って複数の部分にさらに分割されてもよい（図１の破線によって示される）。他方、垂直軸によって表現される周波数領域では、オーディオ信号のセグメント１００は、フル帯域信号を表現しており、これもオブジェクト抽出におけるその後の計算およびプロセスの便利のため、周波数領域に沿って複数のサブバンドに分割されることができる。多くの利用可能なスペクトル変換技法がサブバンド分割において適用されうる。たとえば、高速フーリエ変換（FFT）または複素直交ミラー・フィルター（CQMF）がある。人間の聴覚系の特性を考えると、周波数領域における分割は均一でなくても

よく、低周波数部分ではより細かく、高周波数部分ではより粗くてもよい。図１に示されるように、オーディオ信号１００は複数のチャネル、たとえばチャネルC1ないしC5に関係している。換言すれば、入力オーディオ信号１００は複数のオーディオ信号成分を含み、そのそれぞれはチャネルC1ないしC5の一つに対応する。したがって、ここでのオーディオ・コンテンツは、複数のチャネルに基づいて、フル帯域オーディオ信号またはサブバンド・オーディオ信号のチャネルでありうる。セグメントは、限定なしに、フレーム、フレームの一部、二つ以上のフレームでありうる。いくつかの例示的実施形態では、オーディオ・コンテンツは、フル帯域オーディオ信号のフレームを、周波数領域および時間領域の少なくとも一方において分割することによって得られる一つまたは複数のオーディオ・ブロックを含んでいてもよい。例示的実施形態によれば、オブジェクト抽出があるオーディオ・ブロック、たとえばブロックB1に対して実行されることが所望される場合、その上側の近傍のm個のブロック（単数または複数）および下側の近傍のm個のブロック（単数または複数）が典型的には考慮に入れられる。いくつかの実施形態では、mは1に設定されてもよい。この時点で、それぞれチャネルC1ないしC5に基づくブロックB0ないしB2が一緒に考慮に入れられ、その全体が、処理されるべきオーディオ・コンテンツをなす。

【００１６】

図２は、例示的実施形態に基づくオーディオ・コンテンツからのオーディオ・オブジェクト抽出のための方法２００のフローチャートを示している。上述したように、オーディオ・コンテンツは複数のチャネルのオーディオ信号によって表現されている。

【００１７】

図のように、段階S201において、前記複数のチャネルの第一のチャネルについての第一の部分集合および第二のチャネルについての第二の部分集合を含む投影空間の第一の集合が同定される。いくつかの実施形態では、第一および第二のチャネルは、前記複数のチャネルのうちの任意のチャネルであってもよいが、他の例示的実施形態では後述するいくつかの基準に基づいて選択されてもよい。

【００１８】

既知のように、チャネルのオーディオ信号は、個別の成分を得るためにさまざまな空間に投影されることがある。限定ではなく例解のために、オーディオ・コンテンツの第一および第二のチャネルのオーディオ信号表現についてそれぞれ行列$X \in R^{d \times n}$および$Y \in R^{k \times n}$が生成されるとする。ここで、dおよびkはそれぞれのオーディオ信号に含まれる周波数軸に沿ったサブバンド分割（単数または複数）の数を表わし（典型的にはd＝k）、nはオーディオ信号における時間軸に沿って分割された部分の数を表わす。すなわち、XおよびYはそれぞれ第一および第二のチャネルからのオーディオ・コンテンツのオーディオ信号を表わす。すると、XおよびYをそれぞれの投影空間に投影するために投影ベクトルω_xおよびω_yが使用できる。ここで、$\omega_x \in R^d$であり、$\omega_y \in R^k$である。換言すれば、$\omega_x^T X$および$\omega_y^T Y$はXおよびYについてのそれぞれの投影空間において投影された成分を表わしうる。ここで、ω_x^Tおよびω_y^Tはそれぞれω_xおよびω_yの転置である。さらに、複数のω_xについて、各ω_xを使ってXを投影することによって得られる対応する複数の空間がある。これら複数の空間の集合は、限定なしに、区別の簡単のために、段階S201の第一のチャネルについての第一の部分集合と称される。同様に、複数のω_yについて、各ω_yを使ってYを投影することによって得られる対応する複数の空間がある。これら複数の空間の集合は、限定なしに、区別の簡単のために、段階S201の第二のチャネルについての第二の部分集合と称される。いくつかの例では、第一の部分集合と第二の部分集合の和集合は投影空間の前記第一の集合をなす。

【００１９】

投影空間の第一の集合は通常、第一および第二のチャネルについて複数の投影空間を含むが、ただ一つの空間を含んでいてもよいことを注意しておくべきである。この場合、Xについての投影空間およびYについての投影空間は同じものである。例示的実施形態の範囲はこの点で限定されない。

【００２０】

次いで、方法は段階S202に進む。ここで、第一のチャネルと第二のチャネルの間の相関の第一の集合が決定される。第一の集合の相関のそれぞれは前記第一の部分集合の投影空間の一つおよび前記第二の部分集合の投影空間の一つに対応する。

【００２１】

投影空間の第一の部分集合および投影空間の第二の部分集合を含む投影空間の前記第一の集合が同定されたのち、オブジェクト抽出を容易にするためにいくつかの基準に基づいて、一対の投影空間が、投影空間の第一の部分集合および投影空間の第二の部分集合からそれぞれ選ばれることができる。例示的実施形態によれば、具体的に、XおよびYの両方に共通のオブジェクトが存在しているが、他の源またはノイズによって汚染されていて、該共通のオブジェクトがXまたはYからより容易に分離されるXおよびYについてのそれぞれの投影空間を見出すことが所望される。

【００２２】

例示的実施形態によれば、投影空間の各対について相関が計算される。投影空間の対の一方は第一の部分集合から選ばれ、投影空間の対の他方は第二の部分集合から選ばれる。それにより相関の集合（つまり、段階S202の相関の第一の集合）が形成される。たとえば、ω_xおよびω_yに関するXとYの間の相関ρは次のように計算されうる：

【数１】

$$\rho = \frac{\omega_x^T X Y^T \omega_y}{\sqrt{(\omega_x^T X X^T \omega_x)(\omega_y^T Y Y^T \omega_y)}} \tag{1}$$

ここで、ω_xおよびω_yの意味は上記と同じままであり、$\omega_x \in R^d$であり、$\omega_y \in R^k$である。

【００２３】

引き続き図２を参照するに、段階S203において、段階S203では、少なくとも部分的には相関の前記第一の集合のうちの第一の相関と、前記第一の相関に対応する前記第一の部分集合からの投影空間とに基づいて、前記第一のチャネルのオーディオ信号からオーディオ・オブジェクトが抽出される。ここで、前記第一の相関は第一のあらかじめ定義された閾値より大きい。

【００２４】

例示的実施形態によれば、第一のあらかじめ定義された閾値は所望に応じて任意の時点で設定され、調整されることができる。ある例示的実施形態では、第一のあらかじめ定義された閾値は、単に最大相関より小さいが、相関の前記第一の集合における他の相関よりは大きいものとして設定されることができる。この場合、段階S203における目的は、最大のρを見出し、それによりさらにオブジェクト抽出のためのω_xおよびω_yを同定することである。よって、段階S203では、次が意図される：

【数２】

$$\max_{\omega_x, \omega_y} \omega_x^T X Y^T \omega_y \tag{2}$$

$$\omega_x^T X X^T \omega_x = 1, \omega_y^T Y Y^T \omega_y = 1 \text{ の条件の下で}$$

ここで、X^T、Y^T、ω_x^T、ω_y^TはX、Y、ω_x、ω_yのそれぞれの転置である。YY^Tが非特異であれば、次の最適化問題を解くことによってω_xが得られることが示せる：

【数３】

$$\max_{\omega_x} \omega_x^T XY^T (YY^T)^{-1} YX^T \omega_x \qquad (3)$$

$$\omega_x^T XX^T \omega_x = 1 \text{ の条件の下で}$$

換言すれば、上記の式は、次の一般化された固有値問題の一番上の諸固有値に対応する諸固有ベクトルを見出そうとするものである：

【数４】

$$XY^T (YY^T)^{-1} YX^T \omega_x = \eta XX^T \omega_x \qquad (4)$$

ここで、ηは固有ベクトルω_xに対応する固有値を表わす。

【００２５】

上述したように、いくつかの例示的実施形態によれば、典型的には正規直交性の制約条件のもとで、複数の投影ベクトルω_xおよびω_yがあることがある。その場合、これらの複数の投影ベクトルが次の最適化問題を解くことによって同時に計算できる：

【数５】

$$\max_{W_x} trace(W_x^T XY^T (YY^T)^{-1} YX^T W_x) \qquad (5)$$

$$W_x^T XX^T W_x = I \text{ の条件の下で}$$

ここで、$W_x \in R^{d \times l}$は投影行列を表わし、$W_x = [\omega_x^1, \cdots, \omega_x^l]$であり、lは投影ベクトルの数を表わし、Iは恒等行列を表わす。

【００２６】

まとめると、第一および第二のチャネルのオーディオ入力について、オブジェクト抽出の準備のためにW_x、W_yおよびそれらの間の対応する諸相関Rが決定される。ここで、$W_x = [\omega_x^1, \cdots, \omega_x^l]$であり、$W_y = [\omega_y^1, \cdots, \omega_y^l]$であり、$\omega_x^i$または$\omega_y^i$のそれぞれは列ベクトルを表わし、それが投影空間の基底として使用できる。Rは対角線上に0でない要素（すなわちρ）をもつだけの正方相関行列を表わす。Rにおけるi番目の0でない対角要素ρ_iについて、それは$\omega_x^{iT}X$と$\omega_y^{iT}Y$との間の類似性スコアを測る。$\omega_x^{iT}X$または$\omega_y^{iT}Y$のそれぞれがn次元ベクトルを表わすことを注意しておくべきである。ここで、nはオーディオ信号のセグメント内の部分の数である。こうして、この測度は、オーディオ・ブロックに基づくオーディオ・コンテンツにおける類似性を反映する。上述したように、XおよびYをXおよびYの成分が両者の間の高い相関を示すそれぞれの投影空間に投影することによって、XとYの間の高い類似性が観察でき、XとYの間の共通のオブジェクトがしかるべく抽出されうる。

【００２７】

たとえば、i番目の投影空間について、オブジェクトX_i^*がXから次式により復元されうる：

【数６】

$$X_i^* = \omega_x^i \omega_x^{iT} X \qquad (6)$$

次いで、（前記第一の部分集合からのl個の投影空間に対応する）l個の投影ベクトルか

らなるW_xについて、X^*が次の代替的な式において計算されうる：
【数７】

$$F = W_x H W_x^T \qquad (7)$$

$$X^* = FX \qquad (8)$$

ここで、Hは重み付け対角行列を表わし、0でない要素はその対角線上にあり、その非対角要素はみな0である。Hの導入は、X^*の復元への諸投影ベクトルの寄与を区別することに有益である。具体的には、ある対の投影空間について、XとYがより類似しているほど、Hはより高くなる。結果として、諸Xおよび諸Yはそれぞれ前記ある対の投影空間において抽出されることができる。
【００２８】
本稿に開示される例示的実施形態によれば、Hの対角線上の値を決定するための一つの潜在的なアプローチは、それらを相関行列Rに依存して設定することである。上述したように、Rの対角要素は、W（たとえばWxまたはWy）の列ベクトルによって構築される投影空間にマッピングされる一対のチャネルの間の類似性を反映する。このように、より高い類似性スコアは、同じオブジェクトが存在し、これらの空間から復元できる、より大きな確率を示す。よって、高い類似性スコアをもつ空間から「より多くの」オブジェクトを抽出することが合理的である。すなわち、HはRの適切な関数によって制御できる。つまり、

$$H = f(R) \qquad (9)$$

ここで、関数fはいかなる関数でもよく、その値は入力値の増大とともに減少しない。たとえば、Hは、対角要素の和が1に等しい、規格化されたRであることができる。
【００２９】
上述したように、第一および第二のチャネルは前記複数のチャネルのうちの任意のチャネルでありうる。すなわち、段階S203における第一のチャネルのオーディオ信号からのオブジェクト抽出は第二のチャネルに関して実行されるように示されているが、実質的に前記複数のチャネルからのどのチャネルに関して実行されてもよい。さらに、段階S203では第一のチャネルのオーディオ信号についてのオーディオ・オブジェクト抽出について述べられているが、同様の動作は第二のチャネルのオーディオ信号についてのオーディオ・オブジェクト抽出を実行するために第二のチャネルに適用されてもよい。すなわち、第二のチャネルのオーディオ信号についてのオブジェクト抽出が、第一のチャネルまたは前記複数のチャネルからの他の任意のチャネルに関して実行されてもよい。これについては簡潔のためここでは詳述しない。例示的実施形態の範囲はこれに関して限定されない。
【００３０】
あるいはまた、いくつかの例示的実施形態では、第一および第二のチャネルはいくつかの基準に基づいて選択されてもよい。たとえば、両方のチャネルは、前記複数のチャネルをその相関に基づいてクラスタリングすることによって得られるあるチャネル・グループから選択されてもよい。いくつかの例示的実施形態では、前記複数のチャネルのうちの一対のチャネルの間の相関は、本稿では、該一対のチャネルの間の一般的な相関をいう。たとえば、前記複数のチャネルのうちの一対のチャネルの間のこの相関は、以下の諸段階によって得られてもよい。
【００３１】
第一に、その対のチャネルについての投影空間の第二の集合であって、その対のチャネルの一方についての第三の部分集合およびその対のチャネルの他方についての第四の部分集合を含むものが同定される。例として、この段階は、段階S201と同様の仕方で実装されることができ、ここでは詳述しない。投影空間の前記第二の集合は、投影空間の前記第一の集合とは異なっていてもよいことを注意しておく。ただし、場合によってはそれらは同じであってもよい。
【００３２】

次いで、その対のチャネルの間の相関の第二の集合が決定される。ここで、前記第二の集合の相関のそれぞれは前記第二の部分集合の投影空間の一つおよび前記第四の部分集合の投影空間の一つに対応する。また、この段階は、段階S202と同様の仕方で実装されてもよい。たとえば、その対のチャネルのそれぞれのオーディオ信号からそれぞれ生成される行列XおよびYについて、前記第二の集合の相関のそれぞれを計算するために式(1)を使う。ここでもまた、相関の前記第一の集合および相関の前記第二の集合は通常は、異なる対のチャネルについては異なる。

【００３３】

次に、前記第二の集合の相関のうちの一つがその対のチャネルの間の相関として選択される。ここで、選択された相関は第二のあらかじめ定義された閾値より大きい。この選択段階は、段階S203における第一の相関の選択と同様の仕方で実装されてもよく、ここでは詳述しない。たとえば、式(2)～(5)により実装されてもよい。第二のあらかじめ定義された閾値も所望に応じて任意の時点で設定および調整されうる。ある例示的実施形態では、第二のあらかじめ定義された閾値は、単に相関の第二の集合における最大相関より小さいが、他の相関よりは大きいものとして設定されてもよい。この場合、この段階は、相関の前記第二の集合における最大の相関を、その対のチャネルの間の相関として選択する。

【００３４】

前記複数のチャネルの相関が計算された後、いくつかの例示的実施形態によれば、あらかじめ定義された閾値より大きな相関をもつチャネルどうしが一つのグループにクラスタリングされることができる。このあらかじめ定義された閾値は、クラスター間の最小の許容される相対類似性スコアとして解釈でき、時間に対して一定の値に設定されることができる。結果として、一つのグループにクラスタリングされるチャネルどうしは高いグループ内類似性を示す；一方、異なるグループにクラスタリングされるチャネルどうしは低いグループ間類似性を示す。したがって、一つのグループからのチャネルのオーディオ信号は通例共通のオブジェクトをもち、該共通のオブジェクトの関係した成分（つまり段階S203のオーディオ・オブジェクト）は段階S201～S203により各チャネルについて抽出でき、それによりマルチチャネル・オブジェクトを生成する。これについては後述する。いくつかの例示的実施形態では、チャネル・グループの数は、クラスタリング手順が終了するときに自動的に決定される。前記複数のチャネルのうちのチャネルが互いに似ているまたは前記複数のチャネルの各対の間の相関がみな前記あらかじめ定義された閾値より大きい場合には、前記複数のチャネルは単一のグループとみなされてもよいことを注意しておくべきである。

【００３５】

いくつかの例示的実装によれば、前記複数のチャネルの相関に基づいて前記複数のチャネルをクラスタリングすることは、次の手順によって実行されてもよい。

【００３６】

●初期化：あらかじめ定義された閾値を設定し、対ごとの類似性行列Sを計算する。ここで、項目S_{ij}はi番目とj番目のチャネルの間の類似性を表わす。各チャネルをクラスターとして初期化する。すなわち、$C_1,\cdots,C_T$であり、Tはチャネル数を表わす。

【００３７】

●ループ

○各クラスターについてのクラスター内類似性スコアを、クラスター内のチャネルの対ごとの類似性スコアを平均することによって計算する。すなわち、

【数８】

$$s_{intra}(m)=\frac{\sum_{i\in C_m}\sum_{j\in C_m}s_{ij}}{N_m}$$

ここで、N_mはm番目のクラスターの対の数である。

○各クラスター対について絶対的なクラスター間類似性スコアを、それぞれの自クラスター内にあるチャネルの対ごとの類似性スコアを平均することによって計算する。すなわち、

【数９】

$$s_{\text{int}\,er}(m,n)=\frac{\sum_{i\in C_m}\sum_{j\in C_n}s_{ij}}{N_{mn}}$$

ここで、N_{mn}はm番目とn番目のクラスターの間の対の数を表わす。

○各クラスター対についての相対的なクラスター間類似性スコアを、絶対的なクラスター間スコアを二つのクラスター内類似性スコアの平均で割ることによって計算する。すなわち、

【数１０】

$$s_{rela}(m,n)=\frac{s_{\text{int}\,er}(m,n)}{0.5\times(s_{\text{int}\,ra}(m)+s_{\text{int}\,ra}(n))}$$

○最大の相対クラスター間類似性スコアをもつクラスター対を見出す。最大スコアが前記あらかじめ定義された閾値未満である場合には、ループを終了し；それ以外の場合には、これら二つのクラスターを一つのクラスターに併合する。

【００３８】

●終了。

【００３９】

いくつかの例示的実施形態によれば、前記第一のチャネルが三つ以上のチャネルを含むグループに属する場合、前記第二のチャネルについては複数の候補がある。q個のチャネルからなるチャネル・グループ［$I_1,\cdots,I_{i-1},I_i,I_{i+1},\cdots,I_q$］が同定されるとする。$I_i$番目のチャネルについては、$I_i$番目のチャネルのオーディオ・オブジェクト抽出のためにq－1個の候補Wがある。すなわち

【数１１】

$$W_{I_i}^{(I_1,I_i)},\ldots,W_{I_i}^{(I_{i-1},I_i)},W_{I_i}^{(I_{i+1},I_i)},\ldots,W_{I_q}^{(I_q,I_i)}$$

このように、これらの候補からWを選択するための基準が必要である。

【００４０】

上述したように、いくつかの例示的実施形態では、第二のチャネルはチャネルのうちの任意のまたはランダムなチャネルであってもよい。さもなければ、他のいくつかの例示的実施形態では、チャネル・グループからの第二のチャネルの選択は、第二のチャネルのオーディオ信号がチャネル・グループにおいて最大のエネルギーをもつように実行されてもよい。換言すれば、最も優勢なチャネルが第二のチャネルとして選択されることができる。このように、第一のチャネルについておよびグループ内の他のチャネルについてのオブジェクト抽出はみな該第二のチャネル（つまり、最も優勢なチャネル）に関して実行されてもよい。

【００４１】

上記のように、第二のチャネルのオーディオ信号についてのオブジェクト抽出は、第一のチャネルまたは前記複数のチャネルからの他の任意のチャネルに関して実行されてもよい。代替として、いくつかの例示的実施形態によれば、限定なしに、第二のチャネルがチャネル・グループにおいて最大のエネルギーをもつ場合、第二のチャネルのオーディオ信号についてのオブジェクト抽出のためには、単に前記第一のチャネルを選ぶ代わりに、参

照チャネルを選択することが可能である。たとえば、第二のチャネルとの相関が第三のあらかじめ定義された閾値より大きいチャネルが参照チャネルとして選択されることができる。第三のあらかじめ定義された閾値は、所望に応じて任意の時点において設定され、調整されることができる。ある例示的実施形態では、第三のあらかじめ定義された閾値は、単にチャネル・グループ内の最大相関よりは小さいが他の相関よりは大きいものとして設定されてもよい。この場合、第二のチャネルに最も相関しているチャネルが参照チャネルとして選択される。方法２００の段階S201ないしS203は、第二のチャネルのオーディオ信号のオーディオ・オブジェクト抽出のために、第二のチャネルおよび参照チャネルに適用されることができる。

【００４２】

いくつかの例示的実施形態において前記第一のチャネルが、第一と第二のチャネルの間の相関が第三のあらかじめ定義された閾値より大きいように選択される場合、それがこの場合の参照チャネルであることができる。したがって、方法２００の段階S203において得られるところでは、少なくとも部分的には前記第一の相関と、前記第一の相関に対応する前記第二の部分集合からの投影空間とに基づいて、前記第二のチャネルのオーディオ信号からオーディオ・オブジェクトが抽出されてもよい。

【００４３】

図２に関して上記で示したように、オーディオ・オブジェクトは各チャネルについてさまざまな投影空間において抽出できる。したがって、いくつかの例では、一つのチャネル・グループからの諸チャネルのオーディオ信号から抽出されたオーディオ・オブジェクトに基づいて、マルチチャネル・オブジェクトが生成されることができる。いくつかのさらなる実施形態によれば、オブジェクト抽出を検証し、調整するために、「ソフト・ゲーティング」手順を導入することが有益でありうる。

【００４４】

特に、場合によってはある型のオブジェクトの再生がもとの表現への忠誠から逸脱しうるリスクを軽減するために「ソフト・ゲーティング」手順が導入される。「ソフト・ゲーティング」手順を導入するために、たとえば、利得ベクトルg_bが下記のように決定されることができる。

【００４５】

第一に、マルチチャネル・オブジェクトが少なくとも部分的には、第一および第二のチャネルのオーディオ信号からの抽出されたオーディオ・オブジェクトに基づいて生成される。例示的実施形態によれば、一般に、マルチチャネル・オブジェクトは、一つのチャネル・グループからのチャネルのオーディオ信号から抽出されるオーディオ・オブジェクトに基づいて生成されてもよい。

【００４６】

第二に、生成されたマルチチャネル・オブジェクトは、当技術分野で既知の任意の方法を使ってモノ表現にダウンミックスされることができる。抽出されたオブジェクトのモノ表現ともとのマルチチャネル表現との間のダウンミックス類似性が次に決定される。たとえば、ダウンミックス類似性は次式にように計算できる。

【００４７】

【数１２】

$$S_b^{(i)} = \left| \frac{\mathrm{Re}\left(\sum_t M_i(b,t) \times X_i(b,t)^*\right)}{\sqrt{\sum_t \left\|M_i(b,t)\right\|^2}\sqrt{\sum_t \left\|X_i(b,t)\right\|^2}} \right| \qquad (10)$$

ここで、$X_i(b,t)$はi番目のチャネルの表現であり、$M_i(b,t)$はダウンミックスされたモノ

表現であり、$X_i(b,t)^*$は$X_i(b,t)$の共役であり、|| ||は複素数の絶対値であり、Re()の作用は実部を意味する。bおよびtはそれぞれサブバンド・インデックスおよび時間部分インデックス、つまり周波数領域および時間領域でのそれぞれのインデックスを表わす。モノ表現ともとの表現との間の全体的なダウンミックス類似性は次式：

【数１３】

$$s_b = \frac{1}{T}\sum_{i=1}^{T} s_b^{(i)} \qquad (11)$$

のように、あるいは

【数１４】

$$s_b = \max_i s_b^{(i)} \qquad (12)$$

を介して計算できる。ダウンミックス類似性s_bによって制御される利得値g_b、つまり$g_b^{(1)}$は

【数１５】

$$g_b^{(1)} = f(s_b) \qquad (13)$$

によって表現できる。

【００４８】

関数f(x)がxに関する単調増加関数であることが理解される。fの定義の一例は次式で書ける。

【００４９】

【数１６】

$$f(x) = \frac{1}{1+\exp(a\times(x-b))} \qquad (14)$$

aの値を負に設定することにより、f(x)はxに関する単調増加関数になる。

【００５０】

いくつかの例示的実施形態によれば、計算された利得値は、Xのオブジェクト抽出に影響する重みとして、式(6)または(7)に適用されうる。すなわち、段階S203において第一のチャネルのオーディオ信号から、また第一のチャネルが属するチャネル・グループ内の他の任意のチャネルのオーディオから、オーディオ・オブジェクトを抽出する際、前記第一の相関または前記対応する投影空間に加えて、式(10)～(12)を介して計算されるダウンミックス類似性も、考慮されるべき因子である。換言すれば、段階S203において第一のチャネルのオーディオ信号からオーディオ・オブジェクトを抽出することはさらに、ダウンミックス類似性に基づいてオーディオ・オブジェクトを抽出することを含む。したがって、式(6)は次のように変形できる。

【００５１】

【数１７】

$$X_i^* = \omega_x^i g_b^{(1)} \omega_x^{iT} X \qquad (15)$$

式(7)は
【数１８】

$$F' = g_b^{(1)} F \tag{16}$$

に変形され、式(8)は
【００５２】
【数１９】

$$X^* = F' X \tag{17}$$

に変形される。
【００５３】
ダウンミックス類似性s_bによって制御される利得値$g_b^{(1)}$に加えてまたはその代わりに、諸例示的実施形態によれば、利得値g_bは、次の諸段階によって決定されることもできる：少なくとも部分的には第一および第二のチャネルのオーディオ信号からの抽出されたオーディオ・オブジェクトに基づいてマルチチャネル・オブジェクトが生成された後、生成されたマルチチャネル・オブジェクトがモノ表現にダウンミックスされる。次いで、モノ表現は少なくとも部分的には関係したメタデータ、たとえば推定位置に基づいてプリ・レンダリングされて、「新たな」（つまり、もとのマルチチャネル・オブジェクトとは異なる）マルチチャネル・オーディオ信号表現を生成することができる。その後、エネルギー分布に関する、抽出されたオブジェクトのプリ・レンダリングされた表現（つまり、新たなマルチチャネル・オーディオ信号表現）ともとのマルチチャネル表現との間のプリ・レンダリングされた類似性が決定される。
【００５４】
いくつかの例示的実装では、このプリ・レンダリング類似性は、もとのマルチチャネル・オブジェクトとモノ・オブジェクトのプリ・レンダリングから帰結したもののエネルギー分布の間の不一致によって反映されることが可能である。すなわち、不一致が大きいほど、プリ・レンダリング類似性は小さくなる。したがって、不一致を測る好適なメトリックが、
【数２０】

$$d_b = \sum_{i=1}^{T} \left| e_b^i - e_b^{*i} \right| \tag{18}$$

として、あるいは
【数２１】

$$d_b = \max_i \left| e_b^i - e_b^{*i} \right| \tag{19}$$

として適切に設計できる。ここで、$e_b{}^i$および$e_b{}^{*i}$はそれぞれ、レガシー・コンテンツと、推定されたメタデータと一緒にモノ・オブジェクトをレンダラーを使ってプリ・レンダリングすることから帰結したコンテンツとの、規格化されたエネルギー分布を表わす。bおよびiはそれぞれサブバンド・インデックスおよびチャネル・インデックス、つまり周波数領域およびチャネル領域でのそれぞれのインデックスを表わす。レンダラーを用いたプリ・レンダリングのチャネル構成はレガシー・コンテンツのチャネル構成と同じであることを注意しておくべきである。たとえば、サラウンド5.1レガシー・コンテンツについては、プリ・レンダリングのチャネル構成もサラウンド5.1であるはずである。規格化さ

れたエネルギー分布は

【数２２】

$$e_b^i = \frac{E_b^i}{\sum_{i=1}^{T} E_b^i} \qquad (20)$$

を介して計算できる。ここで、$E_b{}^i$はi番目のチャネルについてのb番目のサブバンド・エネルギーを表わす。

【００５５】

したがって、d_bによって制御される利得値（つまり$g_b{}^{(2)}$）は

【数２３】

$$g_b^{(2)} = f(d_b) \qquad (21)$$

と表現できる。ここで、$f(d_b)$はd_bに関する単調減少関数である。

【００５６】

いくつかの例示的実施形態では、この利得値$g_b{}^{(2)}$もXのオブジェクト抽出に影響する重みとして式(6)または(7)に適用されてもよい。すなわち、段階S203において第一のチャネルのオーディオ信号から（また第一のチャネルが属するチャネル・グループ内の他の任意のチャネルのオーディオから）オーディオ・オブジェクトを抽出する際、前記第一の相関または前記対応する投影空間に加えて、あるいは前記第一の相関、前記対応する投影空間および前記ダウンミックス類似性に加えて、プリ・レンダリング類似性を反映する、式(18)～(19)を介して計算される不一致も、考慮されるべき因子である。すなわち、段階S203において第一のチャネルのオーディオ信号からオーディオ・オブジェクトを抽出することはさらに、プリ・レンダリング類似性に基づいてオーディオ・オブジェクトを抽出することを含む。すると、式(6)は次のように変形できる。

【００５７】

【数２４】

$$X_i^* = \omega_x^i g_b^{(2)} \omega_x^{iT} X \qquad (22)$$

または

$$X_i^* = \omega_x^i g_b^{(2)} g_b^{(1)} \omega_x^{iT} X \qquad (23)$$

式(7)は

【数２５】

$$F'' = g_b^{(2)} F \qquad (24)$$

または

$$F'' = g_b^{(1)} \times g_b^{(2)} F \qquad (25)$$

に変形され、式(8)は

【数２６】

$$X^* = F'' X \qquad (26)$$

に変形される。

【００５８】

それぞれダウンミックス類似性およびプリ・レンダリング類似性に関連する利得ベクトル$g_b{}^{(1)}$および$g_b{}^{(2)}$の少なくとも一方の導入は、抽出されたオブジェクトの再生がもとの表現への忠誠から逸脱するかどうか、および抽出されたオブジェクトの再生が芸術的意図

を保存するかどうかを検証しうる。逸脱は、もしあれば、少なくとも、たとえばモノ表現ともとのマルチチャネル表現との間の音色の不一致があることを示すことができる。したがって、変形された式(15)、(17)、(22)、(23)、(26)は$g_b^{(1)}$または$g_b^{(2)}$の因子を導入することによる逸脱を減らすことができる。

【００５９】

図３は、ある例示的実施形態に基づく、オーディオ・コンテンツからのオーディオ・オブジェクト抽出のためのシステム３００のブロック図である。ここで、オーディオ・コンテンツは複数のチャネルのオーディオ信号によって表現されている。図のように、システム３００は、前記複数のチャネルの第一のチャネルについての第一の部分集合および前記複数のチャネルの第二のチャネルについての第二の部分集合を含む投影空間の第一の集合を同定するよう構成された同定ユニット３０１を含む。システム３００はさらに、前記第一および第二のチャネルの間の相関の第一の集合を決定するよう構成された決定ユニット３０２を含み、前記第一の集合の相関のそれぞれは前記第一の部分集合の投影空間の一つおよび前記第二の部分集合の投影空間の一つに対応する。システム３００はまた、少なくとも部分的には相関の前記第一の集合のうちの第一の相関と、前記第一の相関に対応する前記第一の部分集合からの投影空間とに基づいて、前記第一のチャネルのオーディオ信号からオーディオ・オブジェクトを抽出するよう構成された抽出ユニット３０３をも含み、前記第一の相関は第一のあらかじめ定義された閾値より大きい。

【００６０】

いくつかの実施形態では、システム３００はさらに、チャネル・グループから前記第一および第二のチャネルを選択するよう構成された選択ユニットを有していてもよい。前記チャネル・グループは、前記複数のチャネルの相関に基づいて前記複数のチャネルをクラスタリングすることによって得られる。

【００６１】

いくつかの実施形態では、前記複数のチャネルの一対のチャネルの間の相関が：投影空間の第二の集合であって、その対のチャネルの一方についての第三の部分集合およびその対のチャネルの他方についての第四の部分集合を含むものを同定し；その対のチャネルの間の相関の第二の集合を決定し、ここで、前記第二の集合の相関のそれぞれは前記第三の部分集合の投影空間の一つおよび前記第三の部分集合の投影空間の一つに対応し；前記第二の集合の相関のうちの一つを、その対のチャネルの間の相関として選択することによって決定される。ここで、選択された相関は第二のあらかじめ定義された閾値より大きい。

【００６２】

いくつかの実施形態では、前記チャネル・グループからの前記第一および第二のチャネルの選択は、前記第二のチャネルのオーディオ信号が前記チャネル・グループにおいて最大のエネルギーをもつように、前記チャネル・グループから前記第二のチャネルを選択することを含む。

【００６３】

いくつかの実施形態では、前記チャネル・グループからの前記第一および第二のチャネルの選択はさらに、前記第一および第二のチャネルの間の相関が第三のあらかじめ定義された閾値より大きいように前記チャネル・グループから前記第一のチャネルを選択し；前記第二のチャネルのオーディオ信号から、少なくとも部分的には前記第一の相関と前記第一の相関に対応する前記第二の部分集合からの投影空間とに基づいてオーディオ・オブジェクトを抽出することを含んでいてもよい。

【００６４】

いくつかの実施形態では、システム３００はさらに、少なくとも部分的には前記第一および第二のチャネルのオーディオ信号からの抽出されたオーディオ・オブジェクトに基づいて、マルチチャネル・オブジェクトを生成するよう構成された生成ユニットと；生成されたマルチチャネル・オブジェクトをモノ表現にダウンミックスするよう構成されたダウンミックス・ユニットと；抽出されたオブジェクトのもとのマルチチャネル表現と前記モノ表現との間のダウンミックス類似性を決定するよう構成された類似性決定ユニットとを

有する。ここで、前記第一のチャネルのオーディオ信号からオーディオ・オブジェクトを抽出することは、前記ダウンミックス類似性にさらに基づいてオーディオ・オブジェクトを抽出することを含む。

【００６５】

いくつかの代替的実施形態では、システム３００はさらに、少なくとも部分的には前記第一および第二のチャネルのオーディオ信号からの抽出されたオーディオ・オブジェクトに基づいて、マルチチャネル・オブジェクトを生成するよう構成された生成ユニットと；生成されたマルチチャネル・オブジェクトをモノ表現にダウンミックスするよう構成されたダウンミックス・ユニットと；少なくとも部分的には推定位置に基づいて前記モノ表現をプリ・レンダリングするよう構成されたプリ・レンダリング・ユニットと；抽出されたオブジェクトのもとのマルチチャネル表現と前記プリ・レンダリングされたモノ表現との間の、エネルギー分布に関するプリ・レンダリング類似性を決定するよう構成された類似性決定ユニットとを有する。ここで、前記第一のチャネルのオーディオ信号からオーディオ・オブジェクトを抽出することは、前記プリ・レンダリング類似性にさらに基づいてオーディオ・オブジェクトを抽出することを含む。

【００６６】

いくつかの実施形態では、前記オーディオ・コンテンツは、フル帯域オーディオ信号のフレームを、周波数領域および時間領域の少なくとも一方において分割することによって得られる一つまたは複数のオーディオ・ブロックを含んでいてもよい。

【００６７】

明確のため、システム３００のいくつかの任意的なコンポーネントは図３には示していない。しかしながら、図１～図２を参照して上記した事項はみなシステム３００に適用可能であることは理解されるはずである。さらに、システム３００のコンポーネントは、ハードウェア・モジュールまたはソフトウェア・ユニット・モジュールでありうる。たとえば、いくつかの実施形態では、システム３００は、部分的にまたは完全に、たとえばコンピュータ可読媒体において具現されたコンピュータ・プログラム・プロダクトとして実装されるソフトウェアおよび／またはファームウェアとして実装されてもよい。代替的または追加的に、システム３００は部分的または完全に、たとえば集積回路（IC）、特定用途向け集積回路（ASIC）、システムオンチップ（SOC）、フィールド・プログラマブル・ゲート・アレイ（FPGA）などのようなハードウェアに基づいて実装されてもよい。例示的実施形態の範囲はこれに関して限定されない。

【００６８】

図４は、例示的実施形態を実装するために好適な例示的なコンピュータ・システム４００のブロック図を示している。図のように、コンピュータ・システム４００は、読み出し専用メモリ（ROM）４０２に記憶されたプログラムまたは記憶ユニット４０８からランダム・アクセス・メモリ（RAM）４０３にロードされたプログラムに従ってさまざまなプロセスを実行することのできる中央処理ユニット（CPU）４０１を有する。RAM　４０３では、CPU　４０１がさまざまなプロセスを実行するときに必要とされるデータなども必要に応じて記憶される。CPU　４０１、ROM　４０２およびRAM　４０３はバス４０４を介して互いに接続されている。入出力（I/O）インターフェース４０５もバス４０４に接続されている。

【００６９】

以下のコンポーネントがI/Oインターフェース４０５に接続される：キーボード、マウスなどを含む入力部４０６；陰極線管（CRT）、液晶ディスプレイ（LCD）などのようなディスプレイまたはスピーカーなどを含む出力部４０７；ハードディスクなどを含む記憶部４０８；およびLANカード、モデムなどのようなネットワーク・インターフェース・カードを含む通信部４０９である。通信部４０９は、インターネットのようなネットワークを介して通信プロセスを実行する。ドライブ４１０も必要に応じてI/Oインターフェース４０５に接続される。磁気ディスク、光ディスク、光磁気ディスク、半導体メモリなどのような着脱可能な媒体４１１が必要に応じてドライブ４１０にマウントされ、それにより必

要に応じて、そこから読まれたコンピュータ・プログラムが記憶部４０８にインストールされる。

【００７０】

特に、例示的実施形態によれば、図２を参照して上記したプロセスがコンピュータ・ソフトウェア・プログラムとして実装されてもよい。たとえば、例示的実施形態の実施形態は、方法２００を実行するためのプログラム・コードを含む、機械可読媒体上に有体に具現されたコンピュータ・プログラムを含むコンピュータ・プログラム・プロダクトを含む。そのような実施形態では、コンピュータ・プログラムは、通信ユニット４０９を介してネットワークからダウンロードおよびマウントされ、および／または着脱可能な媒体４１１からインストールされてもよい。

【００７１】

一般に、さまざまな例示的実施形態はハードウェアまたは特殊目的回路、ソフトウェア、論理またはそれらの任意の組み合わせにおいて実装されうる。いくつかの側面はハードウェアにおいて実装され、一方で他の側面がコントローラ、マイクロプロセッサまたは他のコンピューティング装置によって実行されうるファームウェアまたはソフトウェアにおいて実装されてもよい。例示的実施形態のさまざまな側面がブロック図、フローチャートとしてまたは他のいくつかの絵的表現を使って図示され、記述されているが、本稿に記載されるブロック、装置、システム、技法または方法は、限定しない例として、ハードウェア、ソフトウェア、ファームウェア、特殊目的回路または論理、汎用ハードウェアまたはコントローラまたは他のコンピューティング装置またはそれらの何らかの組み合わせにおいて実装されてもよいことは理解されるであろう。

【００７２】

さらに、フローチャートに示されるさまざまなブロックを方法ステップとしておよび／またはコンピュータ・プログラム・コードの動作から帰結する動作としておよび／または関連する機能（単数または複数）を実行するよう構築された複数の結合された論理回路要素として見ることができる。たとえば、実施形態は、機械可読媒体上に有体に具現されたコンピュータ・プログラムを有するコンピュータ・プログラム・プロダクトを含み、該コンピュータ・プログラムは、上記で述べた諸方法を実行するために構成されたプログラム・コードを含む。

【００７３】

本開示のコンテキストにおいて、機械可読媒体は、命令実行システム、装置またはデバイスによってまたはそれとの関連で使うためのプログラムを含むまたは記憶することができるいかなる有体の媒体であってもよい。機械可読媒体は機械可読信号媒体または機械可読記憶媒体でありうる。機械可読媒体は、電子式、磁気式、光学式、電磁式、赤外線または半導体のシステム、装置またはデバイスまたは上記の任意の好適な組み合わせを含みうるが、それに限られなくてもよい。機械可読記憶媒体のより具体的な例は、一つまたは複数のワイヤを有する電気接続、ポータブルなコンピュータ・ディスケット、ハードディスク、ランダム・アクセス・メモリ（RAM）、読み出し専用メモリ（ROM）、消去可能なプログラム可能型読み出し専用メモリ（EPROMまたはフラッシュ・メモリ）、光ファイバー、ポータブルなコンパクト・ディスク読み出し専用メモリ（CD-ROM）、光記憶デバイス、磁気記憶デバイスまたは上記の任意の好適な組み合わせを含む。

【００７４】

本稿に開示される例示的実施形態の方法を実行するためのコンピュータ・プログラム・コードは、一つまたは複数のプログラミング言語の任意の組み合わせにおいて書かれうる。これらのコンピュータ・プログラム・コードは、汎用コンピュータ、特殊目的コンピュータまたは他のプログラム可能なデータ処理装置のプロセッサに提供されてもよく、それにより該プログラム・コードは、該コンピュータまたは他のプログラム可能なデータ処理装置のプロセッサによって実行されたとき、フローチャートおよび／またはブロック図において規定された機能／動作を実装させる。プログラム・コードは完全にコンピュータ上で、部分的にコンピュータ上で、スタンドアローンのソフトウェア・パッケージとして、

部分的にはコンピュータ上で部分的にはリモート・コンピュータ上で、あるいは完全にリモート・コンピュータまたはサーバー上で実行されてもよい。

【００７５】

さらに、動作は特定の順序で描かれているが、これは、そのような動作が示される特定の順序で、あるいは逐次順に実行されること、あるいは所望される結果を達成するために示されているすべての動作が実行されることを要求するものと理解されるべきではない。ある種の状況では、マルチタスクおよび並列処理が有利であることがある。同様に、いくつかの個別的な実装詳細が上記の議論に含まれるものの、これらはいずれかの実施形態のまたは特許請求されうるものの範囲に対する限定として解釈されるべきではなく、むしろ特定の実施形態に固有でありうる事項の記述と解釈されるべきである。別個の実施形態のコンテキストにおいて本明細書に記載されるある種の特徴は、単一の実施形態において組み合わせて実装されてもよい。逆に、単一の実施形態のコンテキストにおいて記述されているさまざまな特徴が、複数の実施形態において別個にまたは任意の好適なサブコンビネーションにおいて実装されることもできる。

【００７６】

付属の図面との関連で読まれるときの上記の記述に鑑み、本稿に開示される上記の例示的実施形態へのさまざまな修正および適応が当業者に明白となることがありうる。任意の、あらゆる修正がそれでも、本稿に開示される、限定しない、例示的な実施形態の範囲内にはいる。さらに、本稿に記載される他の実施形態が、上記の記述および図面に呈示される教示の恩恵をもつ当業者には思いつくであろう。

【００７７】

例示的実施形態は、本稿に記載される形の任意のもので具現されうる。たとえば、以下の付番実施例（EEE: enumerated example embodiment）は、例示的実施形態のいくつかの側面のいくつかの構造、特徴および機能を記述するものである。

〔ＥＥＥ１〕

複数のチャネルに基づくフォーマットのオーディオ・コンテンツからのオーディオ・オブジェクト抽出のための方法であって：

諸投影空間から導出されたフィルタ行列を通じたオブジェクト抽出と；

任意的に、

芸術的な意図を保存するよう、前記抽出されたオブジェクトまたは前記フィルタ行列に対して追加的な利得を提供するソフト・ゲーティングとを含む、

方法。

〔ＥＥＥ２〕

オブジェクト抽出を実行するために各オーディオ・ブロックについて、

各チャネル入力についての投影ベクトル集合を生成し、各対のチャネルの間の最大相関（類似性スコア）がそれらを諸投影空間に投影することによって計算され；

それらのチャネルを対応する相関（類似性スコア）に基づいてグループ化し；

グループ内の各チャネルについて、各オーディオ・ブロックについてのフィルタ行列を導出し；

各チャネルの入力オーディオ信号にそれ自身のフィルタ行列を乗算することによってオブジェクトを復元し、

前記オーディオ・ブロックは、フル帯域オーディオ信号のフレームを周波数領域および時間領域の少なくとも一方において分割することによって得られる、

ＥＥＥ１記載の方法。

〔ＥＥＥ３〕

前記投影ベクトル集合が、現在のオーディオ・ブロックおよび近隣のオーディオ・ブロックを使うことによってブロックごとに形成される、ＥＥＥ２記載の方法。

〔ＥＥＥ４〕

フィルタ行列Fの生成がWおよびHの選択に関わり、

Hの選択は式(9)を介してなされることができ、

Wの選択はグループ内での第二のチャネルの特定に関わる、
ＥＥＥ３記載の方法。
〔ＥＥＥ５〕
第二のチャネルの特定はチャネル・エネルギーに基づき、たとえばグループのうちで最大のエネルギーをもつチャネルが選択される、ＥＥＥ４記載の方法。
〔ＥＥＥ６〕
第一のチャネルについてのWの選択は、前記第二のチャネルに関して投影ベクトル集合を選択する、ＥＥＥ４記載の方法。
〔ＥＥＥ７〕
前記第二のチャネルについてのWの選択は、前記第二のチャネルへの最大の類似性を示す、前記グループのうちでの前記チャネルに関する投影ベクトル集合を選択する、ＥＥＥ４記載の方法。
〔ＥＥＥ８〕
前記ソフト・ゲーティング段階は、各オーディオ・ブロックについての利得ベクトルの生成に関わり、出力を生成するために、前記利得ベクトルはブロックごとに前記オーディオ信号入力を乗算される、ＥＥＥ１記載の方法。
〔ＥＥＥ９〕
前記利得ベクトルは、それぞれプリ・ダウンミックス処理およびプリ・レンダリング処理から生成される二つのサブ利得ベクトルの積として計算される、つまり式(22)である、ＥＥＥ８記載の方法。
〔ＥＥＥ１０〕
プリ・ダウンミックス処理からの前記サブ利得ベクトルは式(10)～(13)によって計算できる、ＥＥＥ９記載の方法。
〔ＥＥＥ１１〕
プリ・レンダリング処理からの前記サブ利得ベクトルは式(17)～(20)によって計算できる、ＥＥＥ９記載の方法。
〔ＥＥＥ１２〕
複数のチャネルに基づくフォーマットになっているオーディオ・コンテンツからのオーディオ・オブジェクト抽出のためのシステムであって、ＥＥＥ１ないし１１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諸ユニットを有する、システム。
〔ＥＥＥ１３〕
オーディオ・コンテンツからのオーディオ・オブジェクト抽出のためのコンピュータ・プログラム・プロダクトであって、当該コンピュータ・プログラム・プロダクトは、非一時的なコンピュータ可読媒体上に有体に記憶されており、実行されたときにＥＥＥ１ないし１１のうちいずれか一項記載の方法の段階を機械に実行させる機械実行可能命令を有する、コンピュータ・プログラム・プロダクト。
【００７８】
本稿に開示される例示的実施形態は開示される特定の実施形態に限定されず、他の実施形態も付属の請求項の範囲内に含まれることが意図されていることは理解されるであろう。本稿では個別的な用語が使われているが、それらは一般的で記述的な意味において使われているだけであり、限定のためではない。

【図１】

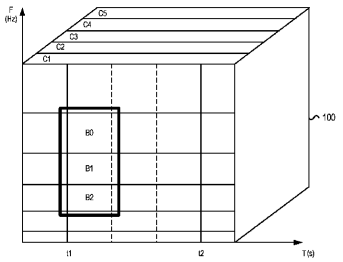
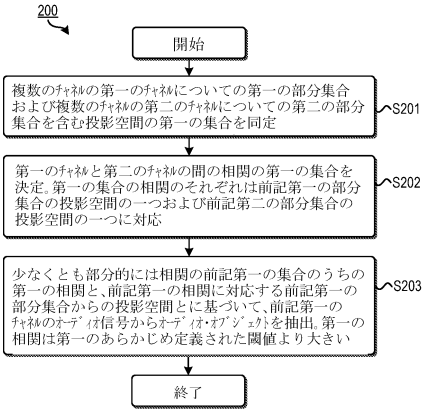
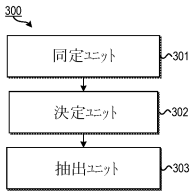


Figure 1

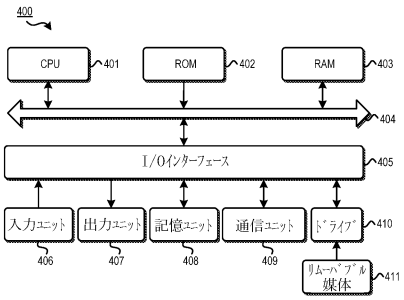
【図２】



【図３】



【図４】



フロントページの続き

(72)発明者　フー，ミンチン
中華人民共和国，ベイジン　１００２０，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　サービシーズ内
(72)発明者　ルー，リエ
中華人民共和国，ベイジン　１００２０，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　サービシーズ内
(72)発明者　チェン，リアンウー
中華人民共和国，ベイジン　１００２０，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　サービシーズ内

審査官　大野　弘

(56)参考文献　再公表特許第２０１０／０９２９１３（ＪＰ，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　２１／０２８
Ｇ１０Ｌ　１９／００８
Ｇ１０Ｌ　２１／０３０８