

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6908718号
(P6908718)

(45) 発行日 令和3年7月28日(2021.7.28)

(24) 登録日 令和3年7月5日(2021.7.5)

(51) Int.Cl. F I
G 1 0 L 25/69 (2013.01) G 1 0 L 25/69
G 1 0 L 19/008 (2013.01) G 1 0 L 19/008

請求項の数 22 (全 23 頁)

<p>(21) 出願番号 特願2019-548682 (P2019-548682) (86) (22) 出願日 平成30年3月6日(2018.3.6) (65) 公表番号 特表2020-509429 (P2020-509429A) (43) 公表日 令和2年3月26日(2020.3.26) (86) 国際出願番号 PCT/EP2018/055482 (87) 国際公開番号 W02018/162487 (87) 国際公開日 平成30年9月13日(2018.9.13) 審査請求日 令和1年11月6日(2019.11.6) (31) 優先権主張番号 17159903.8 (32) 優先日 平成29年3月8日(2017.3.8) (33) 優先権主張国・地域又は機関 欧州特許庁 (EP)</p>	<p>(73) 特許権者 591037214 フラウンホッフアーゲーゼルスシャフト ツ ァ フェルダールング デア アンゲヴァ ンテン フォアシュンク エー. ファオ ドイツ連邦共和国 80686 ミュンヘ ン ハンザシュトラッセ 27ツェー (74) 代理人 100079577 弁理士 岡田 全啓 (74) 代理人 100167966 弁理士 扇谷 一 (72) 発明者 スクーダ ウリ ドイツ連邦共和国 90542 エッケン タール ネルケンシュトラーセ 19A 審査官 渡部 幸和 最終頁に続く</p>
--	---

(54) 【発明の名称】 オーディオストリームに関連付けられた空間性の大きさを提供するための装置および方法

(57) 【特許請求の範囲】

【請求項 1】

オーディオストリームを評価するための装置(100、200、304)であって、
 前記オーディオストリーム(105)は、少なくとも2つの異なる空間層(420、410)で再生されるオーディオチャネル(106、206、305)を備え、前記2つの空間層は空間軸に沿って距離を開けて配置されており、

前記装置は、前記オーディオストリームの前記オーディオチャネルを評価して、

前記オーディオストリームのオーディオチャネルの第1のセットと前記オーディオストリームのオーディオチャネルの第2のセットとの間の類似性の大きさに基づいてアップミックス原点の推定(220d')を取得し、前記アップミックス原点の推定に基づいて空間性の大きさを決定すること

によって、前記オーディオストリームに関連付けられた空間性の大きさ(115、235)を提供するように構成される、装置。

【請求項 2】

前記装置は、前記オーディオストリームの前記オーディオチャネルがより少ないオーディオチャネルのオーディオストリームから導出されることを前記アップミックス原点の推定が示す場合、前記アップミックス原点の推定に基づいて前記空間性の大きさを低減するように構成される、請求項1に記載の装置。

【請求項 3】

前記装置は、前記空間性の大きさを前記アップミックス原点の推定を伴って出力するよ

うに構成される、請求項 1 または請求項 2 に記載の装置。

【請求項 4】

前記空間軸は水平方向に方向づけられる、または、前記空間軸は垂直方向に方向づけられる、請求項 1 ないし 3 のいずれかに記載の装置。

【請求項 5】

前記装置は、前記オーディオストリームのオーディオチャンネルの第 1 のセットに基づいて第 1 のレベル情報を取得し、また前記オーディオストリームのオーディオチャンネルの第 2 のセットに基づいて第 2 のレベル情報を取得するように構成され、

前記装置は、前記第 1 のレベル情報および前記第 2 のレベル情報に基づいて空間レベル情報 (220a') を決定し、また前記空間レベル情報に基づいて前記空間性の大きさを決定するように構成される、請求項 1 ないし 4 のいずれかに記載の装置。

10

【請求項 6】

前記オーディオストリームの前記オーディオチャンネルの第 1 のセットは、前記オーディオストリームの前記オーディオチャンネルの第 2 のセットから離れている、請求項 5 に記載の装置。

【請求項 7】

前記オーディオストリームの前記オーディオチャンネルの第 1 のセットは 1 つ以上の第 1 の空間層においてラウドスピーカーで再生され、前記オーディオストリームの前記オーディオチャンネルの第 2 のセットは 1 つ以上の第 2 の空間層においてラウドスピーカーで再生され、

20

前記 1 つ以上の第 1 の空間層および前記 1 つ以上の第 2 の空間層は空間的に離れている、請求項 5 または請求項 6 に記載の装置。

【請求項 8】

オーディオストリームを評価するための装置 (100、200、304) であって、前記オーディオストリーム (105) は、少なくとも 2 つの異なる空間層 (420、410) で再生されるオーディオチャンネル (106、206、305) を備え、前記 2 つの空間層は空間軸に沿って距離を開けて配置されており、

前記装置は、前記オーディオストリームの前記オーディオチャンネルを評価して、前記オーディオストリームのオーディオチャンネルの第 1 のセットに基づいて第 1 のレベル情報を取得し、また前記オーディオストリームのオーディオチャンネルの第 2 のセットに基づいて第 2 のレベル情報を取得すること、および、

30

前記第 1 のレベル情報および前記第 2 のレベル情報に基づいて空間性の大きさを決定すること、

によって、前記オーディオストリームに関連付けられた空間性の大きさ (115、235) を提供するように構成され、

前記オーディオストリームの前記オーディオチャンネルの第 1 のセットは 1 つ以上の第 1 の空間層においてラウドスピーカーで再生され、前記オーディオストリームの前記オーディオチャンネルの第 2 のセットは 1 つ以上の第 2 の空間層においてラウドスピーカーで再生され、

前記 1 つ以上の第 1 の空間層および前記 1 つ以上の第 2 の空間層は空間的に離れていて

40

、前記装置は、前記オーディオチャンネルの第 1 のセットのレベル情報に基づいてマスキング閾値を決定し、前記マスキング閾値を前記オーディオチャンネルの第 2 のセットのレベル情報と比較するように構成され、

前記比較によって前記オーディオチャンネルの第 2 のセットの前記レベル情報が前記マスキング閾値を超えていることが示された場合、前記装置は空間レベル情報を増強するように構成される、装置。

【請求項 9】

前記装置は、1 つ以上の第 1 の空間層で再生する前記オーディオストリームのオーディオチャンネルの第 1 のセットと、1 つ以上の第 2 の空間層で再生する前記オーディオストリ

50

ームのオーディオチャンネルの第2のセットとの間の類似性の大きさ(220b')を決定し、前記類似性の大きさに基づいて前記空間性の大きさを決定するように構成される、請求項1ないし請求項8のいずれかに記載の装置。

【請求項10】

前記装置は、前記類似性の大きさが小さいほど、前記空間性の大きさが大きくなるように、前記空間性の大きさを決定するように構成される、請求項9に記載の装置。

【請求項11】

オーディオストリームを評価するための装置(100、200、304)であって、前記オーディオストリーム(105)は、少なくとも2つの異なる空間層(420、410)で再生されるオーディオチャンネル(106、206、305)を備え、前記2つの空間層は空間軸に沿って距離を開けて配置されており、

前記装置は、前記オーディオストリームの前記オーディオチャンネルを評価して、1つ以上の第1の空間層で再生する前記オーディオストリームのオーディオチャンネルの第1のセットと、1つ以上の第2の空間層で再生する前記オーディオストリームのオーディオチャンネルの第2のセットとの間の類似性の大きさ(220b')を決定し、前記類似性の大きさに基づいて前記空間性の大きさを決定すること、

前記オーディオチャンネルの第1のセットのレベル情報に基づいてマスキング閾値を決定し、前記マスキング閾値を前記オーディオチャンネルの第2のセットのレベル情報と比較すること、および、

前記比較によって前記オーディオチャンネルの第2のセットの前記レベル情報が前記マスキング閾値を超えていることが示され、且つ、前記類似性の大きさが前記第1のセットと前記第2のセットとの間の類似性が低いことを示す場合、前記空間性の大きさを増大すること、

によって、前記オーディオストリームに関連付けられた空間性の大きさ(115、235)を提供するように構成される、装置。

【請求項12】

オーディオストリームを評価するための装置(100、200、304)であって、前記オーディオストリーム(105)は、少なくとも2つの異なる空間層(420、410)で再生されるオーディオチャンネル(106、206、305)を備え、前記2つの空間層は空間軸に沿って距離を開けて配置されており、

前記装置は、前記オーディオストリームの前記オーディオチャンネルを評価して、前記オーディオストリームに関連付けられた空間性の大きさ(115、235)を提供するように構成され、

前記装置は、前記オーディオチャンネルへの音源のパンニングの時間的変動に関して前記オーディオストリームの前記オーディオチャンネルを分析するように構成される、装置。

【請求項13】

オーディオストリームを評価するための装置(100、200、304)であって、前記オーディオストリーム(105)は、少なくとも2つの異なる空間層(420、410)で再生されるオーディオチャンネル(106、206、305)を備え、前記2つの空間層は空間軸に沿って距離を開けて配置されており、

前記装置は、前記オーディオストリームの前記オーディオチャンネルを評価して、前記オーディオストリームに関連付けられた空間性の大きさ(115、235)を提供するように構成され、

前記装置は、

前記オーディオストリームの空間レベル情報、および/または、

前記オーディオストリームの類似性の大きさ、および/または、

前記オーディオストリームのパンニング情報、および/または、

前記オーディオストリームのアップミックス原点の推定

のうち少なくとも2つのパラメータの重み付け(230)に基づいて、前記空間性の大きさを提供するように構成される、装置。

【請求項 14】

前記装置は、前記空間性の大きさを視覚的に出力（320）するように構成される、請求項 1 ないし請求項 13 のいずれかに記載の装置。

【請求項 15】

オーディオストリームを評価するための装置（100、200、304）であって、
前記オーディオストリーム（105）は、少なくとも2つの異なる空間層（420、410）で再生されるオーディオチャンネル（106、206、305）を備え、前記2つの空間層は空間軸に沿って距離を開けて配置されており、

前記装置は、

前記オーディオストリームの前記オーディオチャンネルを評価して、前記オーディオストリームに関連付けられた空間性の大きさ（115、235）を提供し、

前記空間性の大きさを視覚的に出力（320）するとともに、前記空間性の大きさをグラフ（310）として提供するように構成され、ここで前記グラフは、経時的な前記空間性の大きさに関する情報を提供するように構成され、前記グラフの時間軸は前記オーディオストリームに整合される、装置。

【請求項 16】

前記装置は、前記空間性の大きさを数値（320）として提供するように構成され、前記数値は前記オーディオストリーム全体を表わす、請求項 1 ないし請求項 15 の1つに記載の装置。

【請求項 17】

前記装置は、前記空間性の大きさをログファイル（330）に書き込むように構成される、請求項 1 ないし請求項 16 の1つに記載の装置。

【請求項 18】

オーディオストリームを評価するための方法（500）であって、
前記オーディオストリームは、少なくとも2つの異なる空間層で再生されるオーディオチャンネルを備え、前記2つの空間層は空間軸に沿って距離を開けて配置されており、前記方法は、

前記オーディオストリームのオーディオチャンネルの第1のセットと前記オーディオストリームのオーディオチャンネルの第2のセットとの間の類似性の大きさに基づいてアップミックス原点の推定（220d'）を取得して、

前記アップミックス原点の推定に基づいて前記空間性の大きさを決定すること
によって、前記オーディオストリームに関連付けられた空間性の大きさを提供するために前記オーディオストリームのオーディオチャンネルを評価するステップ（510）を備える、方法。

【請求項 19】

オーディオストリームを評価するための方法（500）であって、
前記オーディオストリームは、少なくとも2つの異なる空間層で再生されるオーディオチャンネルを備え、前記2つの空間層は空間軸に沿って距離を開けて配置されており、前記方法は、

前記オーディオストリームのオーディオチャンネルの第1のセットに基づいて第1のレベル情報を取得し、また前記オーディオストリームのオーディオチャンネルの第2のセットに基づいて第2のレベル情報を取得すること、および

前記第1のレベル情報および前記第2のレベル情報に基づいて前記空間性の大きさを決定すること、

によって、前記オーディオストリームに関連付けられた空間性の大きさを提供するために前記オーディオストリームのオーディオチャンネルを評価するステップ（510）を備え、

前記オーディオストリームの前記オーディオチャンネルの第1のセットは1つ以上の第1の空間層においてラウドスピーカーで再生され、前記オーディオストリームの前記オーディオチャンネルの第2のセットは1つ以上の第2の空間層においてラウドスピーカーで再生

10

20

30

40

50

され、

前記 1 つ以上の第 1 の空間層および前記 1 つ以上の第 2 の空間層は空間的に離れていて

、

前記方法はさらに、

前記オーディオチャンネルの第 1 のセットのレベル情報に基づいてマスキング閾値を決定し、前記マスキング閾値を前記オーディオチャンネルの第 2 のセットのレベル情報と比較するステップと、

前記比較によって前記オーディオチャンネルの第 2 のセットの前記レベル情報が前記マスキング閾値を超えていることが示された場合、空間レベル情報を増強するステップと、を含む、方法。

10

【請求項 20】

オーディオストリームを評価するための方法 (500) であって、

前記オーディオストリームは、少なくとも 2 つの異なる空間層で再生されるオーディオチャンネルを備え、前記 2 つの空間層は空間軸に沿って距離を開けて配置されており、前記方法は、

1 つ以上の第 1 の空間層で再生する前記オーディオストリームのオーディオチャンネルの第 1 のセットと、1 つ以上の第 2 の空間層で再生する前記オーディオストリームのオーディオチャンネルの第 2 のセットとの間の類似性の大きさ (220b') を決定し、前記類似性の大きさに基づいて前記空間性の大きさを決定すること、

前記オーディオチャンネルの第 1 のセットのレベル情報に基づいてマスキング閾値を決定し、前記マスキング閾値を前記オーディオチャンネルの第 2 のセットのレベル情報と比較すること、および

20

前記比較によって前記オーディオチャンネルの第 2 のセットの前記レベル情報が前記マスキング閾値を超えていることが示され、且つ、前記類似性の大きさが前記第 1 のセットと前記第 2 のセットとの間の類似性が低いことを示す場合、前記空間性の大きさを増大すること

によって、前記オーディオストリームに関連付けられた空間性の大きさを提供するために前記オーディオストリームのオーディオチャンネルを評価するステップ (510) を備える、方法。

【請求項 21】

30

オーディオストリームを評価するための方法 (500) であって、

前記オーディオストリームは、少なくとも 2 つの異なる空間層で再生されるオーディオチャンネルを備え、前記 2 つの空間層は空間軸に沿って距離を開けて配置されており、前記方法は、

前記オーディオストリームに関連付けられた空間性の大きさを提供するために前記オーディオストリームのオーディオチャンネルを評価するステップ (510) を備え、

前記空間性の大きさは、

前記オーディオストリームの空間レベル情報、および/または、

前記オーディオストリームの類似性の大きさ、および/または、

前記オーディオストリームのパンニング情報、および/または、

前記オーディオストリームのアップミックス原点の推定

40

のうち少なくとも 2 つのパラメータの重み付け (230) に基づいて、提供される、方法。

【請求項 22】

コンピュータプログラムがコンピュータまたはマイクロコントローラ上で動作しているときに、請求項 18 ないし 21 のいずれかに記載する方法を実行するためのプログラムコードを有するコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

50

技術分野

本発明の実施形態は、オーディオストリームに関連付けられた空間特性、すなわち空間性の大きさの評価に関する。

【背景技術】

【0002】

背景技術

3D-ネスに焦点を当てた3D-オーディオコンテンツの評価は、特定のリスニングルームとすべてのコンテンツを聞く経験豊富なオーディオエンジニアを必要とする単調な作業である。

【0003】

プロフェッショナルレベルでオーディオを使用する場合、すべての制作段階は固有であり、その特定の分野の専門家が必要である。初期の制作段階からコンテンツを受け取り、編集する。最後に、次の製作段階または配布段階に渡される。コンテンツを受信すると、通常、品質チェックが実行され、素材が適切に機能し、指定された基準を満たしていることを確認する。例えば、放送局はすべての入ってくる素材に対してチェックを実行し、全体のレベルまたは動的範囲が目的の範囲[1、2、3]内にあるかどうかを確認する。したがって、必要なリソースを削減するために、説明したプロセスを可能な限り自動化することが望まれている。

【0004】

3D-オーディオを扱う場合、新しい態様が既存の状況に加わる。ラウドネス評価とダウンミックスの可能性とを監視する多くのチャンネルが存在するだけでなく、3D効果がいつ発生するのか、どの程度強力なのかという問題もある。後者は、次の理由で興味深いものである。これまで、5.1は国内市場で映画や長編映画の標準的なサウンド形式であった。制作および流通チェーンのすべてのワークフローおよびセグメント(例えば、ミキシング、マスタリング機能、ストリーミングプラットフォーム、放送局、A/Vレシーバーなど)は5.1サウンドを通過できるが、この再生方法は過去5年間に生まれたため、これは3D-オーディオの場合ではない。コンテンツ制作者は、現在、そのフォーマットの制作を始めている。

【0005】

3D-オーディオコンテンツが含まれている場合、より多くのリソースはレガシーコンテンツと比較して制作チェーンのすべてのポイントで提供されなければならない。多くの場合、サウンド編集スタジオ、ミキシングスタジオ、および、マスタリングスタジオは、3D-オーディオコンテンツで作業できるように、よい室内音響、より多くのスピーカーや拡張された信号フローを有するより大きな部屋を作り上げることによって、彼らの仕事環境をかなり改善する必要があるため、重要なコスト要因である。そのため、どの制作が3D-オーディオを使用してより高い予算と余分な作業を顧客にもたらすかについて、慎重に決定される。

【0006】

今まで、3D-オーディオコンテンツを評価すること、および、3D-オーディオ効果がどれほど印象的かに関して発表することは、それを聞くことによるのみ行われていた。これは、たいてい経験のあるサウンドエンジニアまたはトーンマイスターによって行われ、長くない場合、すべてのプログラムの時間が少なくともかかる。3D-オーディオリスニング設備には高い追加費用がかかるため、リスニングと評価は効果的である必要がある。

【0007】

マルチチャンネル信号を分析するための一般的な方法は、レベルやラウドネスを監視することである[4、5、6]。信号のレベルはピークメーターまたは過負荷インジケータを有するトゥルーピークメーターを用いて測定される。人間の知覚に近い大きさがラウドネス値である。インテグレートッドラウドネス(BS.1770-3)、ラウドネスレンジ(EBUR128LRA)、ATSC A/85(Calm Act)の後のラウ

10

20

30

40

50

ドネス、短期および瞬間的なラウドネス、ラウドネス値の分散またはラウドネスヒストリーは、よく用いられているラウドネスの測定である。これらのすべての測定は、ステレオおよび5.1信号によく使用される。3D-オーディオについてのラウドネスは、現在ITUで調査中である。

【0008】

2つ(ステレオ)または5つ(5.1)の信号の位相関係を比較するために、ゴニオメーター、ベクトルスコープ、相関メーターを利用することができる。エネルギーのスペクトル分布をリアルタイムアナライザー(RTA)またはスペクトルグラフを用いて分析することができる。5.1信号内のバランスを測定するためにサラウンドサウンドアナライザーも利用可能である。

10

【0009】

経時的な立体映像の3D効果を可視化する方法は、深度スクリプト、深度チャートまたは深度プロットである[7、8]。

【発明の概要】

【発明が解決しようとする課題】

【0010】

これらすべての方法は、共通の2つのことを有する。ステレオおよび5.1信号のために開発されているので、それらは3D-オーディオを分析することはできない。そして、3D-オーディオ信号の3D-ネスについての情報を得ることができない。

【0011】

それゆえに、オーディオストリームについての空間性の大きさを得るための改良された概念が望まれている。

20

【課題を解決するための手段】

【0012】

本発明の概要

本発明の実施形態は、オーディオストリームを評価するための装置であって、オーディオストリームは、少なくとも2つの異なる空間層で再生されるオーディオチャンネルを備える。2つの空間層は空間軸に沿って距離を開けて配置される。装置は、さらに、オーディオストリームに関連付けられた空間性の大きさを提供するようにオーディオストリームのオーディオチャンネルを評価するように構成される。

30

【0013】

説明される実施形態は、オーディオストリームに関連付けられた空間性を評価するための概念、すなわち、オーディオストリームに含まれるオーディオチャンネルによって説明されるオーディオシーンの空間性の大きさを提供するものである。このような概念により、評価はサウンドエンジニアによる評価よりも時間と費用効果が高くなる。特に、異なる空間層のラウドスピーカーに割り当てることができるオーディオチャンネルを含むオーディオストリームを評価することは、オーディオストリームを手動で評価するとき、高価なりスニングルーム施設が必要である。オーディオストリームのオーディオチャンネルは、空間層に配置されたラウドスピーカーに割り当てられてもよく、空間層は聴取者の正面および/または背面に配置されたラウドスピーカーによって形成されてもよい、すなわち、それらは正面および/または背面層であってもよく、および/または、空間層は、聴取者の頭が位置する層および/または聴取者の頭よりも上または下に配置される層などの水平層であってもよく、これらはすべて3D-オーディオの典型的な設定である。したがって、この概念は、再生設定を必要とせずに、前記オーディオストリームを評価するという利点を提供する。さらに、サウンドエンジニアがオーディオストリームを聞くことでオーディオストリームを評価するために投資しなければならない時間を節約できる。説明される実施形態は、例えば、サウンドエンジニアまたは他の当業者に、どの時間間隔がオーディオストリームの特別な関心があるかについての指示を提供し得る。それにより、サウンドエンジニアは、装置の評価結果を検証するために、オーディオストリームのこれらの示された時間間隔を聞くだけでよく、人件費の大幅な削減につながる可能性がある。

40

50

【 0 0 1 4 】

いくつかの実施形態において、空間軸は水平方向に方向づけられる、または、空間軸が垂直方向に方向づけられる。空間軸を水平方向に方向づけられる場合、第1層を聴取者の前に配置し、第2層を聴取者の後ろに配置することができる。垂直方向に方向付けられた空間軸の場合、第1層を聴取者の上に配置し、第2層を聴取者と同じ層または聴取者の下に配置することができる。

【 0 0 1 5 】

いくつかの実施形態において、装置は、オーディオストリームのオーディオチャンネルの第1のセットに基づいて第1のレベル情報を取得し、またオーディオストリームのオーディオチャンネルの第2のセットに基づいて第2のレベル情報を取得するように構成される。さらに、装置は、第1のレベル情報および第2のレベル情報に基づいて空間レベル情報を決定し、また空間レベル情報に基づいて空間性のレベルを決定するように構成される。グループ化のために、互いに近いラウドスピーカーで再生されるチャンネルを使用してグループを形成することができる。さらに、空間性を評価するため、または空間レベル情報を取得するために、好ましくはラウドスピーカーに割り当てられたグループが使用され、あるグループのラウドスピーカーは別のグループのラウドスピーカーから離れて配置される。それにより、音がおそらく聴取者の片側でのみ、例えば聴取者の上のラウドスピーカーのグループからのみ再生され、音が聞こえない、または音量の小さい音だけが別の側、例えば聴取者の下のラウドスピーカーのグループから再生される場合、強い空間効果が観察され、決定される場合がある。

【 0 0 1 6 】

いくつかの実施形態において、オーディオストリームのオーディオチャンネルの第1のセットは、オーディオストリームのオーディオチャンネルの第2のセットから離れている。例えば反対に配置されたラウドスピーカーのチャンネルを使用する場合に、離れたセットを使用することは、より意味のある空間レベル情報を決定できる。離れたセットは、聴取者とは異なる方向に向けられたラウドスピーカーで再生されることが好ましいため、そこから得られる空間レベル情報に基づいて、改善された空間性の大きさを得ることができる。

【 0 0 1 7 】

いくつかの実施形態において、オーディオストリームのオーディオチャンネルの第1のセットは1つ以上の第1の空間層においてラウドスピーカーで再生され、オーディオストリームのオーディオチャンネルの第2のセットは1つ以上の第2の空間層においてラウドスピーカーで再生される。1つ以上の第1層および1つ以上の第2層は、例えばそれらが離れたセットであるように、空間的に離れている。例えば、聴取者の上にある第1層と下にある第2層を使用すると、音源が上部のスピーカーからより顕著になり、下部または中間層のラウドスピーカーが周囲または低レベルのバックグラウンドサウンドを提供する場合、空間層の情報を導出することができる。

【 0 0 1 8 】

いくつかの実施形態において、装置は、オーディオチャンネルの第1のセットのレベル情報に基づいてマスキング閾値を決定し、マスキング閾値をオーディオチャンネルの第2のセットのレベル情報と比較するように構成される。さらに、比較によってオーディオチャンネルの第2のセットのレベル情報がマスキング閾値を超えていることが示された場合、装置は、空間レベル情報を増強するように構成される。レベル情報は、オーディオチャンネルのサウンドレベルの瞬間的または平均化された推定によって取得しうるサウンドレベルとすることができる。レベル情報は、例えば、オーディオチャンネルの信号の二乗値（例えば、平均化）によって推定できるエネルギーを説明することもできる。代わりに、レベル情報は、オーディオ信号の時間フレームの絶対値または最大値を使用して取得されてもよい。説明される実施形態は、例えば、心理音響知覚閾値を使用してマスキング閾値を定義することができる。マスキング閾値に基づいて、信号または音源がオーディオチャンネルのセット、例えばオーディオチャンネルの第2のセットのみから来ると認識されるかどうかを決定できる。

10

20

30

40

50

【 0 0 1 9 】

いくつかの実施形態において、装置は、1つ以上の第1の空間層で再生するオーディオストリームのオーディオチャンネルの第1のセットと、1つ以上の第2の空間層で再生するオーディオストリームのオーディオチャンネルの第2のセットとの間の類似性の大きさを決定するように構成される。さらに、装置は、類似性の大きさに基づいて空間性の大きさを決定するように構成される。オーディオチャンネルの第1のセットで再生される信号成分がオーディオチャンネルの第2のセットで再生される信号成分と無相関の場合、2つの異なるオーディオオブジェクトがオーディオチャンネルの各セットで再生されると想定でき、チャンネルは異なるラウドスピーカーに割り当てられる。つまり、無相関の信号は、異なるチャンネルで再生される非類似のオーディオコンテンツを示す。これにより、さまざまなチャンネルのセットから異なるオブジェクトが知覚される可能性があるため、聴取者に強い空間的印象を与えることができる。さらに、相互相関は、チャンネルのグループからの個々の信号を使用して、または和信号を相互相関することによって取得される。和信号は、チャンネルのグループまたはチャンネルのペアの個々の信号を合計することで取得できる。したがって、類似性の評価は、チャンネルのグループまたはチャンネルのペア間の平均相互相関に基づいてもよい。

10

【 0 0 2 0 】

いくつかの実施形態において、装置は、類似性の大きさが小さいほど、空間性の大きさが大きくなるように、空間性の大きさを決定するように構成される。類似性の大きさと空間性の大きさの間の説明された単純な関係（例えば、逆比例性）を使用することは、類似性の大きさに基づく空間性の大きさの単純な決定が可能になる。

20

【 0 0 2 1 】

いくつかの実施形態において、装置は、オーディオチャンネルの第1のセットのレベル情報に基づいてマスキング閾値を決定し、マスキング閾値をオーディオチャンネルの第2のセットのレベル情報と比較するように構成される。さらに、比較によってオーディオチャンネルの第2のセットのレベル情報がマスキング閾値を超えている（例えば、わずかに超えている）ことが示され、且つ、類似性の大きさがオーディオチャンネルの第1のセットとオーディオチャンネルの第2のセットとの間の類似性が低いことを示す場合、装置は、空間性の大きさを増大するように構成される。空間レベル情報と類似性の大きさとを組み合わせることは、空間性の大きさのより正確で信頼性の高い決定が可能になる。さらに、1つのインジケータ（例えば、空間レベル情報または類似性の大きさ）がニュートラルな空間性を示す場合、他のインジケータを使用して、オーディオストリームの高い空間性または低い空間性を決定する方向に進むことができる。

30

【 0 0 2 2 】

いくつかの実施形態において、装置は、オーディオチャンネルへの音源のパンニングの時間的変動に関してオーディオストリームのオーディオチャンネルを分析するように構成される。パンニングの変更に関してオーディオチャンネルを分析することは、オーディオチャンネル上のオーディオオブジェクトを簡単に追跡できる。時間の経過とともにオーディオチャンネル間のオーディオオブジェクトを移動することは、知覚される空間的印象を増大し、前記パンニングを分析することは、意味のある空間性の大きさに役立つ。

40

【 0 0 2 3 】

いくつかの実施形態において、装置は、オーディオストリームのオーディオチャンネルの第1のセットとオーディオストリームのオーディオチャンネルの第2のセットとの間の類似性の大きさに基づいてアップミックス原点の推定を取得するように構成される。さらに、アップミックス原点の推定に基づいて空間性の大きさを決定するように構成される。アップミックス原点の推定は、オーディオストリームが、より少ないオーディオチャンネルを有するオーディオストリームから取得されるかどうかを示す場合がある（例えば、ステレオを5.1または7.1にアップミックスするか、5.1オーディオストリームに基づく2.2のオーディオストリーム）。したがって、オーディオストリームがアップミックスに基づいている場合、オーディオチャンネルの信号成分は、一般により少ないソース信号の

50

数から導出されるため、類似性が高くなる。代わりに、例えば、第1層で主に音源の直接音が再生され（例えば、残響なしまたはほとんどない）、第2層で音源の拡散成分が再生される（例、遅い残響）ことが検出される場合、アップミックスが検出されてもよい。アップミックスに基づくオーディオストリームは、空間的な印象の品質に影響を与え、空間性の大きさを決定するのに役立つ。

【0024】

いくつかの実施形態において、装置は、オーディオストリームのオーディオチャンネルがより少ないオーディオチャンネルのオーディオストリームから導出されることをアップミックス原点の推定が示す場合、アップミックス原点の推定に基づいて空間性の大きさを低減するように構成される。一般に、オーディオチャンネルが少ないオーディオストリームから取得されたオーディオストリームは、空間的印象の点で品質が低いと認識される。したがって、オーディオストリームがより少ないチャンネルのオーディオストリームに基づいていることが検出された場合、空間性の大きさを低減することが適切である。

10

【0025】

いくつかの実施形態において、装置は、空間性の大きさをアップミックス原点の推定を伴って出力するように構成される。サウンドエンジニアが重要な副次情報として使用することができるため、アップミックス原点の推定を個別に出力することは便利である。サウンドエンジニアは、アップストリーム原点の推定を、例えばオーディオストリームの空間性の評価のための重要な情報として使用できる。

20

【0026】

いくつかの実施形態において、装置は、次のパラメータのうち少なくとも2つのパラメータの重み付けに基づいて空間性の大きさを提供するように構成され、パラメータは、オーディオストリームの空間レベル情報、および/または、オーディオストリームの類似性の大きさ、および/または、オーディオストリームのパンニング情報、および/または、オーディオストリームのアップミックス原点の推定である。説明された装置は、重要性に従って個々の因子に有利に重み付けして、空間性の大きさを得ることができる。この重み付けから得られた空間性の大きさは、説明されたインジケータの1つからのみ得られた空間性の大きさよりも改善される、すなわち、より意味があるかもしれない。

【0027】

いくつかの実施形態において、装置は、空間性の大きさを視覚的に出力するように構成される。視覚的な出力を使用して、サウンドエンジニアは視覚的な出力の視覚的な検査に基づくオーディオストリームの空間性を決定することができる。

30

【0028】

いくつかの実施形態において、装置は、空間性の大きさをグラフとして提供するように構成され、グラフは、経時的な空間性の大きさに関する情報を提供するように構成される。グラフの時間軸は、好ましくは、オーディオストリームの時間軸に整合される。サウンドエンジニアは、空間性の大きさのグラフで示されるオーディオストリームのセクションを検査（例えば、聞く）ことができるため、時間の経過に伴う空間性の大きさに関する情報を提供することは、空間的な印象的なコンテンツを含むので、サウンドエンジニアにとって役立つ。これにより、サウンドエンジニアは、空間的に印象的なオーディオシーンをオーディオストリームから高速に抽出したり、決定された空間性の大きさを検証したりできる。

40

【0029】

いくつかの実施形態において、装置は、空間性の大きさを数値として提供するように構成され、数値はオーディオストリーム全体を表わすように構成される。例えば、単純な数値は、異なるオーディオストリームの高速な分類とランク付けに使用することができる。

【0030】

いくつかの実施形態において、装置は、空間性の大きさをログファイルに書き込むように構成される。ログファイルを使用することは、特に自動評価に役立つ。

【0031】

50

本発明の実施形態は、オーディオストリームを評価するための方法を備える。方法は、オーディオストリームに関連付けられた空間性の大きさを提供するためにオーディオストリームのオーディオチャンネルを評価するステップを備える。さらに、オーディオストリームは、少なくとも2つの異なる空間層で再生されるオーディオチャンネルを備え、2つの空間層は空間軸に沿って距離を開けて配置される。

【0032】

図面の簡単な説明

以下において、本発明のより好ましい実施形態を、添付図面を参照して説明する。

【図面の簡単な説明】

【0033】

【図1】図1は、本発明の実施形態による装置のブロック図を示す。

【図2】図2は、本発明の実施形態による装置のブロック図を示す。

【図3】図3は、本発明の実施形態による装置のブロック図を示す。

【図4】図4は、3D-オーディオラウドスピーカーの配置を示す。

【図5】図5は、本発明の実施形態による方法のフローチャートを示す。

【発明を実施するための形態】

【0034】

実施形態の詳細な説明

図1は、本発明の実施形態による装置100のブロック図を示す。装置100は評価装置110を備える。

【0035】

装置100は、どのオーディオチャンネル106が評価装置110に提供されるかに基づいてオーディオストリーム105の入力を受け取る。評価装置110は、オーディオチャンネル106を評価し、評価に基づいて、装置100は空間性の大きさ115を提供する。

【0036】

空間性の大きさ115は、オーディオストリーム105の主観的な空間印象を表現する。慣例的に、人、より好ましくは、サウンドエンジニアは、オーディオストリームに関連付けられた空間性の大きさを提供するためにオーディオストリームを聞かなければならない。したがって、装置100は、評価のためにオーディオストリームを聞く当業者の必要性を回避する。さらに、信頼性のために、サウンドエンジニアは、装置100によって高い空間性の大きさを有することを示すことができるという検証に対してオーディオストリームの特定の部分だけを聞くことができる。したがって、オーディオエンジニアは示されたセクションまたは時間間隔を聞くことだけを必要とすることができるので、時間を節約することができる。例えば、サウンドエンジニアは、空間性の大きさ115を使用して、印象的な3D-オーディオ効果を有するような空間性の大きさ115によって、すなわち、主観的な空間印象であるオーディオストリームの時間間隔またはセクションだけ調べることができる。この指示に基づいて、サウンドエンジニアまたは熟練の聴取者はオーディオストリームの適切なセクションを見つけるまたは変更するために特定のセクションを聞く必要があるとされる。さらに、装置100は、高価な設備の取得を避けることができ、または、高価な設備の使用時間を低減することができる。例えば、オーディオチャンネル106を聞くための必要なプレイバック環境である(例えば、高価な)サウンドラボは、得られた空間性の大きさの確認のためだけに使用することができる。したがって、サウンドラボはより効果的に使用することができ、評価装置がすべて装置100に基づく場合、必須とされない。

【0037】

図2は、本発明の実施形態による装置200のブロック図を示す。言い換えると、図2は、異なる段階(例えば、分析段階)の信号フローとして解釈することができる。実線は、オーディオ信号を示し、(太い)破線は、3D-ネス(例えば空間性の大きさ)を評価するために使用される値を示し、小さい(または細い)破線は、異なる段階の間の情報交換を示す。装置200は、個々のまたは装置100との組み合わせの何れも含む特徴およ

10

20

30

40

50

び機能を備える。装置 200 は、追加の信号またはチャンネルアライナ/グルーパ 210、追加のレベル分析装置 220 a、追加の相関分析装置 220 b、追加の動的パンニング分析装置 220 c および追加のアップミックス推定装置 220 d を備える。さらに、装置 200 は追加の重み付け装置 230 を備える。個々の要素 210、220 a - d および 230 は、評価装置 110 に含まれる個々のまたは組み合わせである場合があり、オーディオチャンネル 206 はオーディオストリーム 105、同様にオーディオチャンネル 106 から得ることができる。

【0038】

装置 200 は、出力として空間性の大きさ 235 を備えることに基づいて、マルチチャンネルオーディオ信号 206 のオーディオ信号の入力を受け取る。装置 200 は、以下でより詳細に説明される評価装置 110 による評価装置 204 を備える。アライナ/グルーパ 210 において、信号またはチャンネルは、例えば、異なる空間層（例えば、空間的にグループ化される）で再生できるチャンネルに整合（例えば時間で）およびグループ化される。したがって、2 つまたはグループが取得され、分析および推定段階 220 a - d に提供される。グループ化は段階 220 a - d と異なる場合があり、この点に関する詳細は以下に記載される。例えば、グループは、図 4 に記載するように、2 つの層を持つラウドスピーカーの配置が示されている層に基づく。第 1 のグループは、層 410 に関連するオーディオチャンネルに基づき、第 2 のグループは、層 420 に関連するオーディオチャンネルに基づく場合がある。代わりに、第 1 のグループは、左側のラウドスピーカーに割り当てられたチャンネルに基づき、第 2 のグループは、右側のラウドスピーカーに割り当てられたチャンネルに基づく場合がある。さらに、可能なグループ化は以下でより詳細に説明する。

【0039】

レベル分析段階 220 a では、異なるグループのサウンドレベルが比較され、グループは 1 つ以上のチャンネルから構成されてもよい。音レベルは、例えば、自発的な信号値、平均化された信号値、最大信号値、または信号のエネルギー値に基づいて推定されてもよい。平均値、最大値、またはエネルギー値は、チャンネル 206 のオーディオ信号の時間フレームから取得されてもよく、または、再帰的推定を使用して取得されてもよい。第 1 のグループが第 2 のグループよりも高いレベル（例えば、平均レベルまたは最大レベル）を有すると決定され、第 1 のグループが第 2 のグループから空間的に離れている場合、空間レベル情報 220 a' が取得され、オーディオチャンネル 206 の高い空間性を示す。次いで、この空間レベル情報 220 a' は、重み付け段階 230 に提供される。空間レベル情報 220 a' は、以下の詳細に概説されるように、最終的な空間性の大きさの計算に寄与する。さらに、レベル分析段階 220 a は、オーディオチャンネルの第 1 グループに基づいてマスキング閾値を決定し、チャンネルの第 2 グループが決定されたマスキング閾値よりも高いレベルを有する場合に高い空間レベル情報 220 a' を取得してもよい。

【0040】

さらに、グルーパ/アライナ 210 による出力としてのチャンネルのグループまたはペアは、類似性を評価するために異なるグループまたはペアの個々の信号、すなわちチャンネルの信号間の相関（例えば、相互相関）を計算できる相関分析段階 220 b に提供される。代わりに、相関分析段階は、和信号間の相互相関を決定してもよい。各グループにおいて、個々の信号を合計することにより、異なるグループから和信号を取得することができ、それにより、グループ間の平均相互相関を取得し、グループ間の平均類似性を特徴付けることができる。相関分析段階 220 b がグループまたはペア間の高い類似性を決定する場合、類似性値 220 b' が、オーディオチャンネル 206 の低い空間性を示す重み付け段階 230 に提供される。相関は、サンプルごとに、または、チャンネル、チャンネルのグループ、またはチャンネルのペアの信号の時間フレームを相関させることによって、相関分析段階 220 b で推定することができる。さらに、相関分析段階 220 b は、レベル分析段階 220 a によって提供された情報に基づいて相関分析を実行するために、レベル情報 220 a' を使用してもよい。例えば、レベル分析段階 220 a から取得された異なるチャンネル、チャンネルのグループまたはチャンネルのペアの信号エンベロープは、レベル情報 22

10

20

30

40

50

0 a ' 'に含まれ得る。エンベロープに基づいて、相関を実行して、個々のチャンネル、チャンネルのグループ、またはチャンネルのペア間の類似性に関する情報を取得することができる。さらに、相関分析段階 2 2 0 b は、レベル分析段階 2 2 0 a に提供されたのと同じチャンネルグループ化を使用してもよく、または全く異なるグループ化を使用してもよい。

【 0 0 4 1 】

さらに、装置 2 0 0 は、ペアまたはグループに基づいて動的パンニング分析 / 検出 2 2 0 c を実行することができる。動的パンニング検出 2 2 0 c は、チャンネルの 1 つのペアまたはグループから別のチャンネルのペアまたはグループに移動するサウンドオブジェクトを検出することができ、例えば、チャンネルの第 1 のグループからチャンネルの第 2 のグループへのレベルの展開である。サウンドオブジェクトが異なるペアまたはグループ間を移動することにより、高い空間的印象が得られる。したがって、ソースの移動がパンニング分析段階 2 2 0 c によって検出される場合、動的パンニング情報 2 2 0 c ' が高い空間性を示す重み付け段階 2 3 0 に提供される。さらに、チャンネルのペアまたはグループ間で音源の動き（または、小さな動きのみ、例えばチャンネルのグループ内のみ）が検出されない場合、動的パンニング情報 2 2 0 c ' は、低い空間性を示し得る。パンニング検出段階 2 2 0 c は、サンプルごとに、またはフレームごとに、パンニング分析を実行することができる。さらに、動的パンニング検出段階 2 2 0 c は、レベル分析段階 2 2 0 a から取得されたレベル情報 2 2 0 a ' ' 'を使用して、パンニングを検出することができる。代わりに、パンニング検出段階 2 2 0 d は、パンニング検出を実行するためにそれ自体でレベル情報を推定してもよい。動的パンニング検出 2 2 0 c は、レベル分析段階 2 2 0 a または相関分析段階 2 2 0 b と同じグループ、またはグルーパー / アライナ 2 1 0 によって提供される異なるグループを使用してもよい。

【 0 0 4 2 】

さらに、アップミックス推定段階 2 2 0 d は、相関分析段階 2 2 0 b からの相関情報 2 2 0 b ' 'を使用するか、さらなる相関分析を実行して、チャンネル 2 0 6 がより少ないオーディオチャンネルを有するオーディオストリームを使用して形成されたかどうかを検出する。例えば、チャンネル 2 0 6 が相関情報 2 2 0 b ' 'から直接アップミックスに基づいているかどうかをアップミックス推定段階 2 2 0 d が評価し得る。代わりに、個々のチャンネル間の相互相関は、アップミックス推定段階 2 2 0 d で実行されてもよく、相関情報 2 2 0 b ' 'によって示される高い相関に基づいて、チャンネル 2 0 6 がアップミックスに由来するかどうかを評価する。相関分析段階 2 2 0 b またはアップミックス推定段階 2 2 0 c のいずれかによって実行される相関分析は、アップミックスを生成する一般的な方法が信号非相関機によるものであるため、アップミックス原点の検出に有用な情報である。アップミックス原点の推定値 2 2 0 d ' は、アップミックス推定段階 2 2 0 d によって重み付け段階 2 3 0 に提供される。アップミックス原点の推定値 2 2 0 d ' が、チャンネル 2 0 6 がより少ないチャンネルを有するオーディオストリームから導出されることを示す場合、アップミックス原点の推定値 2 2 0 d ' は、重み付け 2 3 5 にマイナスまたはわずかな寄与を与える場合がある。アップミックス推定段階 2 2 0 d は、レベル分析段階 2 2 0 a 、相関分析段階 2 2 0 b または動的パンニング検出段階 2 2 0 c と同じグループ、またはグルーパー / アライナ 2 1 0 によって提供される異なるグループを使用することができる。

【 0 0 4 3 】

例えば、重み付け段階 2 3 5 は、空間性の大きさへの寄与を平均化して、空間性の大きさを得ることができる。寄与は、因子 2 2 0 a ' 、 2 2 0 b ' 、 2 2 0 c ' および / または 2 2 0 d ' の組み合わせに基づいてもよい。平均化は均一であっても重み付けされていてもよく、重み付けは因子の有意性に基づいて実行されてもよい。

【 0 0 4 4 】

いくつかの実施形態では、空間性の大きさは、分析段階 2 2 0 a - c のうちの 1 つ以上のみに基づいて取得することができる。さらに、グルーパー / アライナは、分析段階 2 2 0 a - c のいずれか 1 つに統合されてもよく、例えば、各分析段階は独自にグループ化を実行する。

10

20

30

40

50

【 0 0 4 5 】

図3は、本発明の実施形態による装置300のブロック図を示す。言い換えれば、図3は、3D-ネスメーター304の一般的な信号の流れを示している。装置300は、装置100および200に匹敵し、入力としてマルチチャンネルオーディオ信号305を取り、それはそのまま出力されてもよい。3D-ネスメーター304は、評価装置110および評価装置204による評価装置である。マルチチャンネルオーディオ信号305に基づいて、図形出力またはディスプレイ310（例えば、グラフ）を使用して、数値出力またはディスプレイ320を使用して（例えば、オーディオストリーム全体に対して1つの数値スカラ値を使用して）、および/または、例えば、グラフまたはスカラ値が書き込まれ得るログファイル330を使用して、空間性の大きさをグラフィカルに出力することができる。さらに、装置300は、音声信号305または音声信号305を含む音声ストリームに含めることができる追加のメタデータ340を提供することができ、メタデータは空間性の大きさを含むことができる。さらに、追加のメタデータは、アップミックス原点の推定値または装置200における分析段階の出力のいずれかを含んでもよい。

10

【 0 0 4 6 】

図4は、3D-オーディオラウドスピーカーの配置400を示す。言い換えると、図4は、5+4構成における3D-オーディオ再生のレイアウトを示す。中間層のラウドスピーカーは文字Mで示され、上部層のスピーカーはUとラベル付けされる。数字は、聴取者に対するスピーカーの方位角を指す（例えば、M30は30°の方位角で中間層にあるスピーカーである）。ラウドスピーカーの配置400は、オーディオストリーム（例えば、ストリーム105、オーディオチャンネル106、206または305）からオーディオチャンネルを割り当てることによって使用され、オーディオストリームを再生する。ラウドスピーカーの配置は、ラウドスピーカーの第1層410と、ラウドスピーカーの第1層410から垂直方向に離れて配置されたラウドスピーカーの第2層420を含む。ラウドスピーカーの第1層は5つのラウドスピーカー、すなわち、中央M0、正面右M-30、正面左M30、サラウンド右M-110およびサラウンド左M110を含む。さらに、ラウドスピーカーの第2層420は4つのラウドスピーカー、すなわち、左上U30、右上U-30、上後右U-110および後左上U110を含む。装置100、200、または300を使用する分析のために、層、すなわち層410および層420に基づいてグループ化を提供することができる。さらに、第2のグループを取得するために、例えば第1のグループから形成される聴取者から左側のラウドスピーカーと聴取者から右側のラウドスピーカーとを使用して層をまたいでグループを形成することができる。代わりに、第1のグループは、聴取者の前に位置するラウドスピーカーに基づき、第2のグループは、聴取者の後ろに位置するラウドスピーカーに基づき、第1のグループまたは第2のグループは、垂直に離れた、すなわちグループは垂直の層で形成されるラウドスピーカーを含む。さらに、別の任意のグループ化を定義でき、ラウドスピーカーの配置を検討できる。

20

30

【 0 0 4 7 】

図5は、本発明の実施形態による方法500のフローチャートを示す。方法は、オーディオストリームに関連付けられた空間性の大きさを提供するために、オーディオストリームのオーディオチャンネルを評価するステップ510を含む。さらに、オーディオストリームは、少なくとも2つの異なる空間層で再生されるオーディオチャンネルを含み、2つの空間層は空間軸に沿って距離を置いて配置される。

40

【 0 0 4 8 】

以下では、図2を参照して詳細を説明する。

【 0 0 4 9 】

実施形態は、与えられた3D-オーディオ信号の3D-オーディオ効果のパワー（または強度）を測定する方法を説明する。3D-オーディオコンテンツを見て、3D効果の特徴とする素材のセクションを見つけ、そのパワーを評価することは、手作業で行う必要がある主観的なタスクであることがわかっている。実施形態は、このプロセスをサポートするために使用することができ、3D効果が発生する位置を示し、3D効果の強さを評価す

50

ることによってそれを加速することができる3D-ネスメーターを説明する。

【0050】

「3D-ネス」という用語は、非常に広範な意味をカバーするため、これまで学術分野で3D-オーディオ効果の強さには使用されていなかった。したがって、より正確な用語と定義が詳しく説明されている[9, 10]。これらの用語は、印象全体ではなく、再生されたオーディオの特定の1つの態様にのみ適用される。一般的な印象として、全体的なリスニングエクスペリエンス(OLE)またはエクスペリエンスの品質(QoE)という用語が導入されている[11]。後者の用語は3D-オーディオに限定されない。3D-オーディオ効果の強さをOLEやQoEなどの用語と区別するために、このドキュメントでは3D-ネスという用語が使用されることがある。

10

【0051】

一般に、少なくとも2つの異なる垂直層で音源を生成できる場合(図4を参照)、再生システムは3D-オーディオまたは「没入型」と呼ばれる。一般的な3D-オーディオ再生レイアウトは、5.1+4、7.1+4または2.2.2である[12]。

【0052】

3D-オーディオに固有の効果は次のとおりである。

- ・高音の音源の知覚
- ・ローカライズ精度(方位角、仰角、距離)[9]
- ・動的なローカライズ精度(移動オブジェクトの場合)[9]
- ・巻き込み(音に覆われている感覚)[13, 14, 15]
- ・空間の明瞭さ(空間シーンをどれだけはっきりと認識できるか)[14, 15]

20

【0053】

これらの効果は、3D-オーディオの品質機能[9]または属性のカテゴリ[10, 16]と呼ばれる。3D-オーディオ効果のパワーは、OLEまたはQoEと直接関連しないことに留意すべきである。

【0054】

3D-ネスの実用的な例を示すために、いくつかのシナリオがリスト化されている。

- ・音源は異なる垂直層を移動し、例えば、ヒューという効果音は中間(または水平)層から上部層に移動する。
- ・音源は中間層と上部層で再生され、例えば、主音は中間層で知覚され、上から話しているときの音声セットまたは直接音は中間層で再生され、周囲音は上部層で再生される。

30

【0055】

さらに、製作者側では、サウンドトラックがファイナライズされるフィルムサウンドミキシング施設で3D-ネスを測定する要求がある。コンテンツがブルーレイ(登録商標)またはストリーミングサービスで配信されるように準備されている場合、3D-ネスの監視も重要である。放送局などのトップ(OTT)ストリーミングおよびダウンロードサービス[17]を介したコンテンツディストリビューターは、3D-ネスを測定して、3D-オーディオハイライトプログラムとして宣伝するコンテンツを決定する必要がある。研究、教育機関、映画批評は、異なる理由で3D-ネスを測定することに関心を持つ他の存在である。

40

【0056】

従来の方法は、3D-オーディオ信号の3D-ネスの測定には適していない。したがって、3D-ネスメーターがここで提案されている。一般的に、マルチチャネルオーディオ信号は、オーディオ分析が行われるメーターに送られる(図3を参照)。出力は、さまざまな表現の3D-ネス測定とともに、未処理かつ未変更のオーディオコンテンツであるかもしれない。3D-ネスメーターは、時間の関数として3D-ネスをグラフィカルに表示できる。代わりに、測定値を数値で表現し、統計を計算して異なる材料を比較可能にすることもできる。すべての結果はログファイルにエクスポートすることも、適切なメタデータ形式で元のオーディオ(ストリーム)に追加することもできる。オブジェクトベースま

50

たはシーンベースのオーディオの場合、例えば1次アンビソニックス（FOA）または高次アンビソニックス（HOA）、表現形式、オーディオチャンネルは、最初に基準スピーカレイアウトにレンダリングすることで評価できる。

【0057】

実施形態では、3D-ネスメータの動作モードは、並行作業の異なる分析段階にわたって共有される。各段階では、特定の3D-オーディオ効果に固有のオーディオ信号の特性を検出できる（図2を参照）。分析段階の結果は、重み付け、合計、および表示し得る。最後に、ディスプレイ上で、サウンドエンジニアに合計の3D-ネスインジケータ（例えば、空間性の大きさ）と最も重要なサブ結果（例えば、個々の分析段階の結果）を提供することができる。これにより、サウンドエンジニアは、関心のあるセクションを見つけたり、3D-ネスに関する決定を下したりするのに役立つさまざまなデータを有する。合計の3D-ネスインジケータは、0から2まで（0...2）の範囲の線形スケールであり、3D-ネス=0は、評価されたオーディオストリームに期待される3D-オーディオ効果がない、またはまったくないことを意味する。3D-ネス=2の最大値は、オーディオストリームで非常に強い3D-オーディオ効果が発生することを示す場合がある。範囲と合計の3D-ネスインジケータスケールの単位とは、事前に決定されている場合があり、他の値、単位または範囲（例えば、-1...1、0...10など）を使用できる。

10

【0058】

ステップでは、入力チャンネルを特定のチャンネルペアまたはチャンネルグループに割り当てることができる。可能なチャンネルペアは次のとおりである。

20

- ・ 中間層の左および上部層の左
- ・ 中間層の左サラウンドと上部層の左サラウンド
- ・ 中間層の中央と上部層の左
- ・ ...

可能なチャンネルグループは次のとおりである。

- ・ 中間層および上部層
- ・ 中間層の左右と上部層の左右
- ・ ...

【0059】

以下において、実施形態において使用および/または決定され得るパラメータが説明される。さらに、以下では、層によるチャンネルのグループ化が主に考慮されるが、他の実施形態では他のグループ化が使用されてもよい。

30

【0060】

レベル分析段階

レベル分析段階220aは、上部層にレベルがあるかどうか、もしあればレベルが中間層に対してどれだけ高いかを監視することができる。重要な測定は、垂直音源のマスクング閾値である[18、19]。この分析段階では、中間層の信号のマスクング閾値が上部層によって大幅に超えた場合、またはその逆の場合にのみ、3D-ネスを検出できる。上部層で測定された信号（またはレベル）がない場合、またはその時点で対応する中間層の信号に対してレベルが低すぎる場合、3D-ネスメータは低い3D-ネス値（例えば、レベル分析段階から取得した情報に基づいて）を報告する場合がある。

40

実施形態では、3D-ネスメータを設定して、(i)上部層のレベルを中間層のマスクング閾値と比較する、(ii)中間層のレベルを上部層のマスクング閾値と比較するまたは(iii)指定されたすべての層を比較し、下位レベルの層のレベル（例えば、最低レベルの層）を対応する他の層と比較する。

【0061】

相関段階

実施形態では、相関段階220bを使用して、正規化された短期相互相関についてチャンネルペアまたはチャンネルグループを分析する。この測定は、2つの信号がどれほど似ているかを表し、時間の経過によるエネルギーの違いから導出される可能性がある。上部層信

50

号の非常に高い類似性は、中間層信号の最も可能性の高い要素、または中間層信号全体が上部層にも供給されることを示す。これは、特定の知覚された包絡線またはわずかに上に移動したサウンドシーンを提供する場合がある。

【0062】

低い相関関係は、中間層と上部層の信号が類似していないことを示しており、3D-オーディオ効果が強くなる。相関段階とレベル分析段階とは、情報を交換できる(図2の点線を参照)。例えば、上部層のレベルがマスキング閾値に近いが、わずかに上にある場合、相関段階が高い相関度を示すとき、示された3D-ネスは低くなる可能性がある。しかしながら、同じレベルの関係で相関が低い場合は、示された3D-ネスが高い可能性がある。

10

【0063】

動的なパンニング検出

実施形態では、パンニング段階220cは、異なる位置に異なる時間に現れるサウンド要素を探す。動的なパンニングは、中間層の左前の位置から上層の右後の位置に飛ぶヘリコプターのように、空間を移動する信号によって特徴付けられる。信号に関しては、パンニングの動きにより、1つのチャンネルまたはチャンネルのグループから別のチャンネルへのクロスフェードが発生する。そのようなクロスフェードが信号内で検出された場合、パンニング効果は3D-オーディオ効果(例えば、知覚された高い空間性)を生成する可能性がある。レベル分析段階からのレベル情報は、他の時定数でより詳細に処理される場合がある(例えば、平均化ウィンドウが長くなる)。

20

【0064】

アップミックス推定

アップミキシングアルゴリズムは、サウンド処理で確立される。通常、デコレーションと信号分離を使用して、より広く、より包み込み、より刺激的なサウンド再生を実現するために、使用するチャンネルの数を増やす。

アップミックス検出段階220dは、所定の非相関が以前に適用された自動アップミックスの結果であり得るかどうかを調べる。したがって、相関段階のデータ(例えば220a)が使用される。さらに、信号を分析して、最も一般的なアップミックス方法から生じる可能性のあるアーチファクトと結果を見つけることができる。

自動アップミックスのヒントを見つけることができるかどうかは、後続のダウンミックスの可能性がサウンドカラーレーションを引き起こす可能性があるため、重要な情報になる可能性がある。さらに、自動アップミックスは、芸術的に作成された3D-オーディオミックスに比べて価値が低いと見なされる可能性がある。したがって、オーディオストリームがアップミックスに基づいていると推定されている場合、取得された空間性の大きさから低い空間性が示されることがある。

30

【0065】

更なる応用

本発明の実施形態の有用性を説明するために、3D-ネスメーターのいくつかの実践的な使用事例が提示される。

【0066】

シナリオ1

サウンドエンジニアは、特定の映画ミックスに3D-オーディオが含まれているかどうかを求められる。3D-ネスメーターがない場合、エンジニアはサウンドトラック全体を聴いて、関連する3D-効果が発生するかどうかを確認する必要がある。3D-ネスメーターがある場合、オーディオはオフラインで分析される。これは、リアルタイムよりもはるかに高速であることを意味し、3D-効果が発生するセクションがマークされる。

40

【0067】

シナリオ2

エンジニアは、映画のサウンドトラックで最も印象的な3D-オーディオセクションを見つけるよう求められる。3D-ネスメーターの結果を見ると、3D効果のあるスポット

50

をすばやく見つけることができる。3D-ネスメーターで指摘されたセクションのみを聞く必要がある。

【0068】

シナリオ3

制作会社は、2つ可能性のあるタイトルのうち、どちらを追加の3D-オーディオトラックを有するブルーレイ（登録商標）用にリリースするかを決定する必要がある。3D-ネスメーターの結果は、どのタイトルが3D-オーディオ効果をより頻繁に使用しているかを示しており、経済的な判断の基礎となる。

【0069】

シナリオ4

3D-オーディオ制作がミキシングされる。所望の3D効果がとても強く、混乱する可能性がある場合に、3D-ネスメーターは、信号を監視し、ミキシングエンジニアに示すことができる。または、エンジニアが3D効果を作りたいと考えており、3D-ネスメーターが示すように、その効果は容易に知覚できるほど強くはない。

【0070】

シナリオ5

3Dオーディオミックスが配信され、クライアントは、ミックスが芸術的な意図を持つエンジニアによって作成されたものであるか、自動アップミックスのみであるかを調べたいと考えている。自動アップミキシングが適用されている場合、3D-ネスメーターが表示する場合がある。

【0071】

実施形態では、3D-ネスメーターの概念は、測定されたパラメータのグラフィックまたは数値の表現だけでなく、3Dオーディオ信号における聴覚3D-効果の存在および量を決定するプロセス全体を含む。

【0072】

さらに、3D-ネスメーターの方法は、非3D-オーディオコンテンツまたは2Dマルチチャンネルサラウンドコンテンツにも使用でき、どれぐらいのサラウンド効果が予想されるか、および、プログラムの何時にそれらが位置するかを示す。このため、垂直方向に間隔を空けた2つのチャンネルまたはチャンネルのグループを比較する代わりに、水平方向に間隔を空けたチャンネルまたはチャンネルのグループ、例えばフロントチャンネルおよびサラウンドチャンネルを比較できる。

【0073】

いくつかの態様が装置の文脈において記載されてきたが、これらの態様は対応する方法の記述をも表すことは明らかであり、ブロックまたはデバイスは方法ステップまたは方法ステップの機能に対応する。同様に、方法ステップの文脈において記載された態様は、対応する装置の対応するブロック、アイテムまたは機能の記述をも表す。いくつかのまたはすべての方法ステップは、例えば、マイクロプロセッサ、プログラム可能なコンピュータまたは電子回路のようなハードウェア装置によって（または用いて）実行することができる。いくつかの実施形態において、いくつかの1つ以上の最も重要な方法ステップは、このような装置によって実行することができる。

【0074】

特定の実現要求に依存して、本発明の実施形態は、ハードウェアにおいてまたはソフトウェアにおいて実施することができる。実施は、その上に記憶された電子的に読取可能な制御信号を有し、それぞれの方法が実行されるようにプログラム可能なコンピュータシステムと協働する（または協働することができる）、デジタル記憶媒体、例えばフロッピー（登録商標）ディスク、DVD、CD、ブルーレイ（登録商標）ディスク、ROM、PROM、EPROM、EEPROMまたはフラッシュメモリを用いて実行することができる。それ故に、デジタル記憶媒体は、コンピュータ読取可能とすることができる。

【0075】

本発明に係るいくつかの実施形態は、本願明細書に記載された方法の1つが実行される

10

20

30

40

50

ように、プログラム可能なコンピュータシステムと協働することができる、電子的に読取可能な制御信号を有するデータキャリアを備える。

【0076】

一般に、本発明の実施形態は、コンピュータプログラム製品がコンピュータ上で動作するとき、本発明の方法の1つを実行するように動作可能であるプログラムコードによるコンピュータプログラム製品として実施することができる。プログラムコードは、例えば機械読取可能なキャリアに記憶することができる。

【0077】

他の実施形態は、機械読取可能なキャリアに記憶された、本願明細書に記載された方法の1つを実行するコンピュータプログラムを備える。

10

【0078】

言い換えれば、本発明の方法の一実施形態は、それ故に、コンピュータプログラムがコンピュータ上で動作するとき、本願明細書に記載された方法の1つを実行するプログラムコードを有するコンピュータプログラムである。

【0079】

本発明の方法の更なる実施形態は、それ故に、その上に記録され、本願明細書に記載された方法の1つを実行するコンピュータプログラムを備えるデータキャリア（またはデジタル記憶媒体またはコンピュータ読取可能媒体）である。データキャリア、デジタル記憶媒体または記録媒体は、通常は有形および/または不揮発性である。

【0080】

20

本発明の方法の更なる実施形態は、それ故に、本願明細書に記載された方法の1つを実行するコンピュータプログラムを表すデータストリームまたは信号のシーケンスである。データストリームまたは信号のシーケンスは、例えば、データ通信接続、例えばインターネットによって転送されるように構成することができる。

【0081】

更なる実施形態は、本願明細書に記載された方法の1つを実行するように構成されたまたは適合された処理手段、例えばコンピュータまたはプログラマブルロジックデバイスを備える。

【0082】

更なる実施形態は、本願明細書に記載された方法の1つを実行するコンピュータプログラムがインストールされたコンピュータを備える。

30

【0083】

本発明に係る更なる実施形態は、本願明細書に記載された方法の1つを実行するコンピュータプログラムを、受信者に転送（例えば、電子的または光学的に）するように構成された装置またはシステムを備える。受信者は、例えば、コンピュータ、モバイルデバイス、メモリデバイスなどとすることができる。装置またはシステムは、例えば、コンピュータプログラムを受信者へ転送するファイルサーバを備えることができる。

【0084】

いくつかの実施形態において、本願明細書に記載された方法のいくつかまたは全ての機能を実行するために、プログラマブルロジックデバイス（例えばフィールドプログラマブルゲートアレイ）を用いることができる。いくつかの実施形態において、フィールドプログラマブルゲートアレイは、本願明細書に記載された方法の1つを実行するために、マイクロプロセッサと協働することができる。一般に、方法は、好ましくはいかなるハードウェア装置によっても実行される。

40

【0085】

本明細書で記載される装置は、ハードウェア装置を用いて、または、コンピュータを用いて、または、ハードウェア装置とコンピュータとの組み合わせを用いて、実装することができる。

【0086】

本明細書で記載される装置、または、本明細書で記載される装置のいずれかのコンポー

50

ネットは、ハードウェアでおよび／またはソフトウェアで少なくとも部分的に実装することができる。

【 0 0 8 7 】

本明細書に記載される方法は、ハードウェア装置を用いて、または、コンピュータを用いて、または、ハードウェア装置とコンピュータとの組み合わせを用いて、実装することができる。

【 0 0 8 8 】

本明細書に記載される方法、または、本明細書で記載される装置のいずれかのコンポーネントは、ハードウェアでおよび／またはソフトウェアで少なくとも部分的に実装することができる。

10

【 0 0 8 9 】

上記記載された実施形態は、単に本発明の原理に対して説明したものである。本願明細書に記載された構成および詳細の修正および変更は、当業者にとって明らかであると理解される。それ故に、本発明は、間近に迫った特許請求の範囲のスコープのみによって制限され、本願明細書の実施形態の記載および説明の方法によって表された特定の詳細によって制限されないことが意図される。

【 0 0 9 0 】

参考文献

[1] EBU. EBU TECH 3344: Practical guidelines for distribution systems in accordance with EBU R 128. Geneva, 2011.

20

[2] IRT. Technische Richtlinien - HDTV. Zur Herstellung von Fernsehproduktionen für ARD, ZDF und ORF. Frankfurt a.M., 2011.

[3] ARTE. Allgemeine technische Richtlinien. ARTE, Kehl, 2013.

[4] Gerhard Spikofski and Siegfried Klar. Levelling and Loudness in Radio and Television Broadcasting. European Broadcast Union, Geneva, 2004.

[5] ITU. ITU-R BS.2054-2: Audio Levels and Loudness, volume 2. International Telecommunication Union, Geneva, 2011.

[6] Robin Gareus and Chris Goddard. Audio Signal Visualisation and Measurement. In International Computer Music and Sound & Music Computing Conference, Athens, 2014.

30

[7] B Mendiburu. 3D Movie Making - Stereoscopic Digital Cinema from Script to Screen. Focal Press, 2009.

[8] B. Mendiburu. 3D TV and 3D Cinema. Tools and Processes for Creative Stereoscopy. Focal Press, 2011.

[9] Andreas Silzle. 3D Audio Quality Evaluation: Theory and Practice. In International Conference on Spatial Audio, Erlangen, 2014. VDT.

[10] Nick Zacharov and Torben Holm Pedersen. Spatial sound attributes - development of a common lexicon. In AES 139th Convention, New York, 2015. Audio Engineering Society.

[11] Michael Schoeffler, Sarah Conrad, and Jurgen Herre. The Influence of the Single / Multi-Channel-System on the Overall Listening Experience. In AES 55th Conference, Helsinki, 2014.

40

[12] Ulli Scuda. Comparison of Multichannel Surround Speaker Setups in 2D and 3D. In Malte Kob, editor, International Conference on Spatial Audio, Erlangen, 2014. VDT.

[13] R Sazdov, G Paine, and K Stevens. Perceptual Investigation into Envelopment, Spatial Clarity and Engulfment in Reproduced Multi-Channel Audio. In AES 31st Conference, London, 2007. Audio Engineering Society.

[14] R Sazdov. The effect of elevated loudspeakers on the perception of engulfment, and the effect of horizontal loudspeakers on the perception of envelopment

50

. In ICSA 2011. VDT.

[15] Robert Sazdov. Envelopment vs. Engulfment: Multidimensional scaling on the effect of spectral content and spatial dimension within a three-dimensional loudspeaker setup. In International Conference on Spatial Audio, Graz, 2015. VdT.

[16] Torben Holm Pedersen and Nick Zacharov. The development of a Sound Wheel for Reproduced Sound. In AES 138th Convention, Warsaw, 2015. AES.

[17] AES. Technical Document AESTD1005.1.16-09: Audio Guidelines for Over the Top Television and Video Streaming. AES, New York, 2016.

[18] Hyunkook Lee. The Relationship between Interchannel Time and Level Differences in Vertical Sound Localisation and Masking. In AES 131st Convention, number 10
r lclld, pages 1-13, 2011.

[19] Hanne Stenzel, Ulli Scuda, and Hyunkook Lee. Localization and Masking Thresholds of Diagonally Positioned Sound Sources and Their Relationship to Interchannel Time and Level Differences. In International Conference on Spatial Audio, Erlangen, 2014. VDT.

【図1】

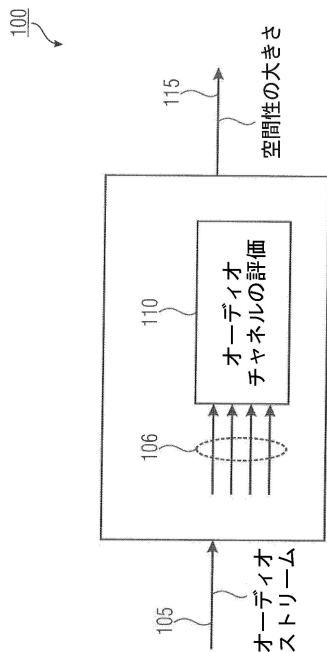


Fig. 1

【図2】

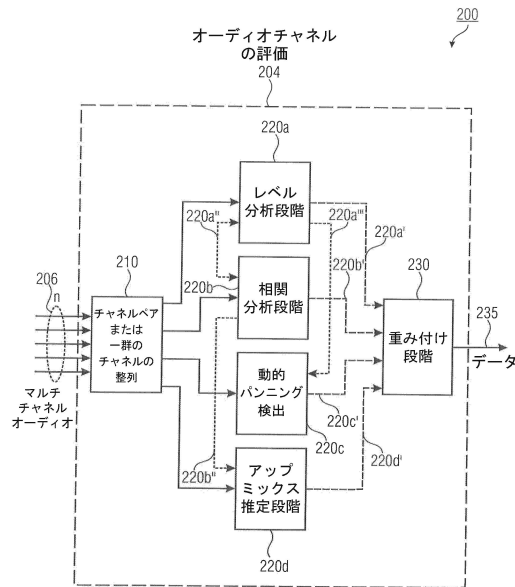


Fig. 2

【 図 3 】

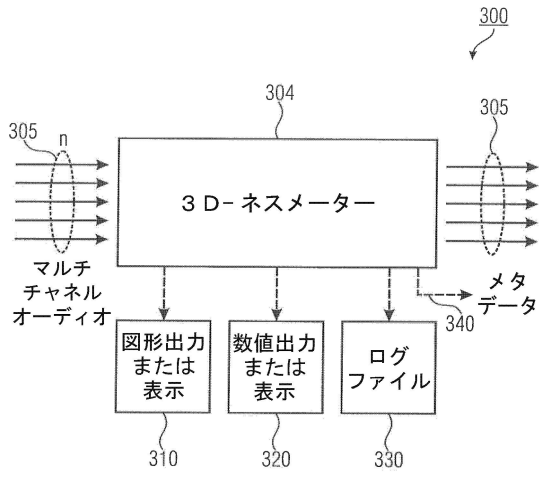


Fig. 3

【 図 4 】

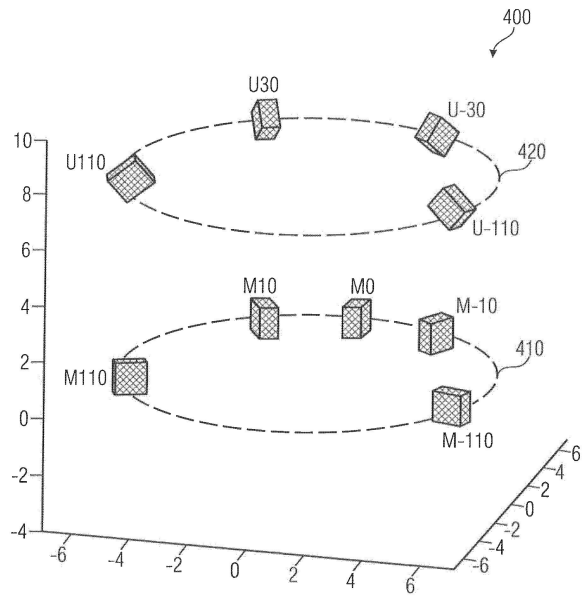


Fig. 4

【 図 5 】

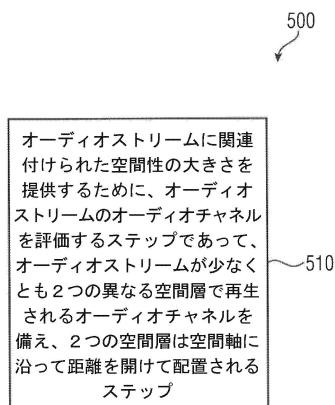


Fig. 5

フロントページの続き

(56)参考文献 米国特許出願公開第2007/0041592(US, A1)
特開2011-250049(JP, A)

(58)調査した分野(Int.Cl., DB名)
G10L 19/00-25/93