

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5189979号
(P5189979)

(45) 発行日 平成25年4月24日(2013.4.24)

(24) 登録日 平成25年2月1日(2013.2.1)

(51) Int.Cl.

F I

G 1 O L 19/02 (2013.01)

G 1 O L 19/02 1 5 O

G 1 O L 19/00 (2013.01)

G 1 O L 19/00 4 O O Z

G 1 O L 19/20 (2013.01)

G 1 O L 19/00 3 3 O B

請求項の数 12 (全 31 頁)

(21) 出願番号 特願2008-525019 (P2008-525019)
 (86) (22) 出願日 平成18年7月24日(2006.7.24)
 (65) 公表番号 特表2009-503615 (P2009-503615A)
 (43) 公表日 平成21年1月29日(2009.1.29)
 (86) 国際出願番号 PCT/US2006/028874
 (87) 国際公開番号 W02007/016107
 (87) 国際公開日 平成19年2月8日(2007.2.8)
 審査請求日 平成21年7月23日(2009.7.23)
 (31) 優先権主張番号 60/705,079
 (32) 優先日 平成17年8月2日(2005.8.2)
 (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 カリフォルニア州 94
 103-4813 サンフランシスコ ポ
 トレロ アベニュー 100
 (74) 代理人 100068755
 弁理士 恩田 博宣
 (74) 代理人 100105957
 弁理士 恩田 誠
 (74) 代理人 100142907
 弁理士 本田 淳
 (74) 代理人 100071010
 弁理士 山崎 行造

最終頁に続く

(54) 【発明の名称】 聴覚事象の関数としての空間的オーディオコーディングパラメータの制御

(57) 【特許請求の範囲】

【請求項 1】

プロセッサが複数の入力チャンネルを受け取り、入力チャンネルの数よりも多い数のオーディオ出力チャンネルを生成する、オーディオの処理方法であって、

1 以上の前記複数のオーディオ入力チャンネルにおける時間に関するスペクトル形状の変化を検出するステップと、

前記 1 以上の前記複数のオーディオ入力チャンネルにおいて、オーディオ信号における聴覚事象の連続的に継続する境界を特定するステップであって、時間についての閾値を越えるスペクトル形状の変化により境界を定め、聴覚事象は、別々の分離したものとして識別される連続する境界間のオーディオのセグメントであり、連続的に継続する聴覚事象が得られるように、各境界は、先行する聴覚事象の終端であり、次の聴覚事象の先端である、特定するステップと、

前記オーディオ出力チャンネルを、少なくとも部分的には聴覚事象の境界及び/又は前記聴覚事象の境界に関連付けられた前記スペクトル形状の変化の程度に応答して生成するステップと、

を具備し、前記出力チャンネルを生成する前記ステップは、聴覚事象の境界でのみ更新される、オーディオの処理方法。

【請求項 2】

前記オーディオチャンネルの各々は、データブロック中のサンプルにより表現される、請求項 1 に記載の方法。

【請求項 3】

前記スペクトル形状は、ブロック中のオーディオのスペクトル形状である、請求項 2 に記載の方法。

【請求項 4】

前記時間についてのスペクトル形状の変化の検出は、ブロックとブロックとの間でのオーディオのスペクトル形状の変化の検出である、請求項 3 に記載の方法。

【請求項 5】

聴覚事象の時間的な開始と終了の境界はそれぞれ、データのブロックの境界に一致する、請求項 4 に記載の方法。

【請求項 6】

装置であって、

複数の入力チャンネルを受け取り、入力チャンネルの数よりも多い数のオーディオ出力チャンネルを生成するプロセッサを備え、該プロセッサは、

1 以上の前記複数のオーディオ入力チャンネルにおける時間に関するスペクトル形状の変化を検出する手段と、

前記 1 以上の前記複数のオーディオ入力チャンネルにおいて、オーディオ信号における聴覚事象の連続的に継続する境界を特定する手段であって、時間についての閾値を越えるスペクトル形状の変化により境界を定め、聴覚事象は、別々の分離したものとして識別される連続する境界間のオーディオのセグメントであり、連続的に継続する聴覚事象が得られるように、各境界は、先行する聴覚事象の終端であり、次の聴覚事象の先端である、前記特定する手段と、

前記オーディオ出力チャンネルを、少なくとも部分的には聴覚事象の境界及び / 又は前記聴覚事象の境界に関連付けられた前記スペクトル形状の変化の程度に応答して生成する手段とを含み、前記出力チャンネルを生成する前記手段は、聴覚事象の境界でのみ更新される、装置。

【請求項 7】

請求項 6 に記載の装置を制御して請求項 1 乃至請求項 5 のいずれか 1 項に記載の方法をコンピュータに実行させるためのプログラムを記録したコンピュータが読み取り可能な記録媒体。

【請求項 8】

請求項 1 乃至請求項 5 のいずれか 1 項に記載の方法をコンピュータに実行させるための、コンピュータが読み取り可能な記録媒体に記録されたコンピュータプログラム。

【請求項 9】

請求項 1 乃至請求項 5 のいずれか 1 項に記載の方法を用いてビットストリームを生成するステップを備える方法。

【請求項 10】

請求項 1 乃至請求項 5 のいずれか 1 項に記載の方法を用いてビットストリームを生成する手段を備える装置。

【請求項 11】

複数の入力チャンネルを受け取り、入力チャンネルの数よりも多い数のオーディオ出力チャンネルを生成するようになっているオーディオプロセッサであって、

1 以上の前記複数のオーディオ入力チャンネルにおける時間に関するスペクトル形状の変化を検出する手段（6"）と、

前記 1 以上の前記複数のオーディオ入力チャンネルにおいて、オーディオ信号における聴覚事象の連続的に継続する境界を特定する手段であって、時間についての閾値を越えるスペクトル形状の変化により境界を定め、聴覚事象は、別々の分離したものとして識別される連続する境界間のオーディオのセグメントであり、連続的に継続する聴覚事象が得られるように、各境界は、先行する聴覚事象の終端であり、次の聴覚事象の先端である、手段（4"）と、

前記オーディオ出力チャンネルを、少なくとも部分的には聴覚事象の境界及び / 又は前

10

20

30

40

50

記聴覚事象の境界に関連付けられた前記スペクトル形状の変化の程度に応答して生成する手段(26)と、

を具備し、前記出力チャンネルを生成する前記手段は、聴覚事象の境界でのみ更新される、オーディオプロセッサ。

【請求項12】

複数の入力チャンネルを受け取り、入力チャンネルの数よりも多い数のオーディオ出力チャンネルを生成するようになっているオーディオプロセッサであって、

1以上の前記複数のオーディオ入力チャンネルにおける時間に関するスペクトル形状の変化を検出するようになっていると共に、前記1以上の前記複数のオーディオ入力チャンネルにおいて、オーディオ信号における聴覚事象の連続的に継続する境界を特定するようになっている検出手段(4", 6")であって、時間についての閾値を越えるスペクトル形状の変化により境界を定め、聴覚事象は、別々の分離したものとして識別される連続する境界間のオーディオのセグメントであり、連続的に継続する聴覚事象が得られるように、各境界は、先行する聴覚事象の終端であり、次の聴覚事象の先端である、検出手段(4", 6")と、

前記オーディオ出力チャンネルを、少なくとも部分的には聴覚事象の境界及び/又は前記聴覚事象の境界に関連付けられた前記スペクトル形状の変化の程度に応答して生成するようになっているアップミキサーと、

を具備し、前記アップミキサーは聴覚事象の境界でのみ更新される、オーディオプロセッサ。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、エンコーダが複数のオーディオチャンネルをより少ない数のオーディオチャンネルと、このオーディオチャンネル同士の好ましい空間的關係を表す1以上のパラメータとにダウンミックスし、このパラメータの全て又は一部は、聴覚事象の関数として生成する、オーディオエンコーディング方法及び装置に関する。本発明はまた、複数のオーディオチャンネルを、聴覚事象の関数としての多数のオーディオチャンネルにアップミックスするオーディオについての方法及び装置に関する。本発明はまた、このような方法を実行するための、又は、このような装置を制御するためのコンピュータプログラムに関する。

【背景技術】

【0002】

[空間的コーディング]

限られたビットレートのデジタルオーディオコーディング技術によれば、入力マルチチャンネル信号を分析して、「ダウンミックス」コンポジット信号(入力信号より少ないチャンネルを有する信号)とサウンドフィールドのパラメトリックモデルを有するサイド情報が導き出される。このサイド情報とコンポジット信号はデコーダに送られ、デコーダで、適切な損失のあるデコーディング及び/又は無損失のデコーディングが適用され、その後、コンポジット信号を元のサウンドフィールドの近似を再現するより数の多いチャンネルに「アップミキシング」するために、パラメトリックモデルがこのデコードされたコンポジット信号に適用される。このような「空間的」コーディングシステム又は「パラメトリック」コーディングシステムの第1の目的は、非常に限られた量のデータでマルチチャンネルサウンドフィールドを再現することである。従って、このことは、元のサウンドフィールドをシミュレートするために用いられるパラメトリックモデルに制限を加える。このような空間的コーディングシステムの詳細は、以下に「参照としての編入」の表題で引用したものを含む、種々の文献に記載されている。

【0003】

このような空間的コーディングシステムでは、元のサウンドフィールドをモデル化するために、チャンネル間較差又はレベル差(ILD)、チャンネル間時間差又は位相差(I

10

20

30

40

50

P D)、及びチャンネル間相互相関(I C C)のようなパラメータを、元のサウンドフィールドをモデル化するために一般に採用する。一般に、このようなパラメータは、コード化された各チャンネルの複数のスペクトル帯域を推定するものであり、時間についてダイナミックに推定される。

【 0 0 0 4 】

一般的な先行技術である $M = 1$ の $N : M : N$ 空間的コーディングシステムにおいて、マルチチャンネル入力信号は重複させた D F T (離散的周波数変換) を用いて、周波数領域に変換される。D F T スペクトルは、次に、耳の臨界帯域を近似する帯域に分割される。チャンネル間較差、チャンネル間時間差又は位相差、及び、チャンネル間相互相関の推定値を各帯域について計算する。これらの推定値は、元の入力チャンネルをモノラル信号又は2チャンネル立体音響コンポジット信号にダウミックスするために用いられる。推定した空間的パラメータと共にこのコンポジット信号は、コンポジット信号が重複させた D F T を用い臨界帯域の間隔で周波数領域に変換される。この空間的パラメータは、次いで、元のマルチチャンネル信号を近似させるために、対応する帯域に適用される。

【 0 0 0 5 】

[聴覚事象と聴覚事象の検出]

サウンドを別々の分離したものとして識別される単位又はセグメントに分割することはしばしば、「聴覚事象分析」又は「聴覚シーン分析」又は「オーディオ事象」と呼ばれる。聴覚シーン分析の広範囲にわたる説明については、Albert S. Bregmanが自分の本、聴覚シーン分析、「Auditory Scene Analysis--The Perceptual Organization of Sound」マサチューセッツ工科大学、1991年、第4刷、2001年、MIT Press パパック第2版、で述べている。加えて、Bhadkamkar他による1999年12月14日の米国特許6,002,776に1976年までの「聴覚シーン分析によるサウンド分離に関する先行技術」として引用されている。しかし、Bhadkamkar他の特許では、「人の聴覚処理のモデルの科学的な観点からは興味があるが、聴覚シーン分析を用いる技術は、現状はコンピュータへの要求が大きく専門的過ぎるので、抜本的な進展がなされるまでサウンドの実用的な分離技術は考えられない」と結論づけているので、聴覚シーン分析の実用的な利用の意欲はそがれている。

【 0 0 0 6 】

聴覚事象を特定する有益な方法は、以下に「参照としての編入」の表題で記載したCrockett and Crockett他による、種々の特許出願及び論文に述べられている。これらの書類によれば、オーディオ信号(又はマルチチャンネル信号内のチャンネル)は、時間についてのスペクトル成分(周波数の関数としての振幅)の変化を検出することにより、それぞれが別々の分離したものとして識別される聴覚事象に分けられる。これは例えば、オーディオ信号の連続する時間ブロックのスペクトル内容を計算し、オーディオ信号の連続する時間ブロック間での差異を計算し、このような時間ブロック間でのスペクトル内容の差異が閾値を越えたとき、連続する時間ブロック同士の境界を聴覚事象の境界と特定する。あるいは、時間に対するスペクトル内容の変化の代わりに又はそれに加えて、時間に対する振幅の変化をを計算してもよい。

【 0 0 0 7 】

計算における要求を最小限にする実施の形態において、そのプロセスでは、全周波数帯域(全帯域幅のオーディオ)又は実質的に全周波数帯域(実際の実施の形態では、スペクトルの端部で帯域を制限するフィルタをしばしば採用する)を分析し、最大音量のオーディオ信号成分に最大の重み付けを行うことによりオーディオを時間セグメントに分割する。このアプローチは、小さな時間スケール(20ミリ秒(ms)以下)では、耳はその時間単一の聴覚事象に集中する傾向があるという、心理音響現象の利点を採用している。これは、同時に複数の事象が生じているとき、1つの成分が感覚的に突出している傾向があり、それを生じているただ1つの事象であるかのように処理することができることを意味する。この現象の利点を採用することで、聴覚事象の検出を処理中の複雑なオーディオに対応させることもできる。例えば、処理中の入力オーディオ信号が独奏楽器であるならば

、特定されるオーディオ事象は、演奏中の個々の響きとなるであろう。音声信号の入力についても同様で、スピーチの個々の成分、例えば母音と子音とが、個々のオーディオ事象として特定されるであろう。ドラムビート又は複数の楽器と音声を伴う音楽のように、オーディオの複雑さが増すにつれて、聴覚事象の検出において、その瞬間の「最も突出した」（すなわち、音量の最も大きい）オーディオ要素を特定することになる。

【 0 0 0 8 】

計算を複雑にするという代償を払って、このプロセスで、全帯域幅ではなく離散周波数サブ帯域（固定サブ帯域又は動的に定めたサブ帯域又は、固定サブ帯域と動的に定めたサブ帯域の両方）における時間についてのスペクトル構成における変化を考慮してもよい。この代替的なアプローチでは、特定の時間で単一のオーディオストリームだけを想定するのではなく異なった周波数サブ帯域における 1 以上のオーディオストリームを考慮に入れる。

10

【 0 0 0 9 】

聴覚事象の検出は、時間領域のオーディオ波形を時間区間又は時間ブロックに分割し、FFTのような、フィルタバンク又は時間周波数変換のどちらかを使い、各ブロックのデータを周波数領域に変換することにより実行する。各ブロックのスペクトル成分の振幅は、振幅の変化による影響を削減又は減少させるために正規化してもよい。結果生じた周波数領域での表現で、特定のブロックにおけるオーディオのスペクトル内容を表示する。連続するブロックのスペクトル内容が比較され、変化が閾値より大きい場合は、聴覚事象の開始時又は終了時を示すものと捉える。

20

【 0 0 1 0 】

以下に説明するように、周波数領域のデータは正規化されていることが望ましい。周波数領域のデータが必要な正規化される程度から振幅の表示が得られる。従って、この程度が所定の閾値を越えたならば、これも事象の境界を示すものとみなすことができる。スペクトルの変化と振幅の変化から得られる事象の開始時及び終了時は、どちらのタイプの変化で得られたものからでも事象の境界を特定できるように、OR 結合してもよい。

【 0 0 1 1 】

Crockett and Crockett他による出願及び論文は、本発明の特徴に関連して特に有益であるが、聴覚事象と事象の境界を特定する他の技術を本発明に採用することもできる。

30

【 発明の開示 】

【 0 0 1 2 】

本発明の 1 つの特徴によれば、オーディオエンコーダは、複数の入力オーディオチャンネルを受け取り、1 以上の出力オーディオチャンネルと、この 1 以上の出力オーディオチャンネルから導くことのできる複数のオーディオチャンネル間の望ましい空間的關係を記述する 1 以上のパラメータとを出力する。1 以上の前記複数の入力チャンネルにおける時間に関する信号特性の変化が検出され、そして 1 以上の前記複数の入力チャンネルにおける時間に関する信号特性の変化が聴覚事象の境界として特定され、連続する境界と境界との間のオーディオセグメントがチャンネルにおける聴覚事象を構成する。前記 1 以上のパラメータの一部又は全部は、少なくとも部分的には聴覚事象及び / 又は前記聴覚事象の境界に関連付けられた信号特性の変化の程度に応答して生成される。一般に、聴覚事象は、別々の分離したものとして識別されるオーディオのセグメントである。信号特性の 1 つの有用な尺度には、例えば、引用したCrockett and Crockett他の論文に記載されているような、オーディオの一定のスペクトル内容が含まれる。1 以上のパラメータの一部又は全部は、少なくとも部分的には 1 以上の聴覚事象の存在又は不存在に応答して生成される。時間についての閾値を越える信号特性の変化を聴覚事象の境界として特定することができる。あるいは、前記聴覚事象の境界に関連する信号特性における変化の程度の連続的な指標に少なくとも部分的に応答して、前記 1 以上のパラメータの一部又は全部が生成される。原則として、本発明の特徴はアナログ領域及び / 又はデジタル領域で実行することができるが、実際の実施の形態では、それぞれのオーディオ信号がデータブロック内のサ

40

50

ンプルとして表現されるデジタル領域で実施される場合が多い。この場合、信号特性はブロック内のオーディオのスペクトル内容とすることができ、時間についての信号特性の変化の検出が、ブロックからブロックへのオーディオのスペクトル内容の変化の検出とすることができ、時間的な聴覚事象の開始と終了の境界がデータブロックの境界と一致する。

【 0 0 1 3 】

本発明の他の特徴によれば、オーディオプロセッサは、複数の入力チャンネルを受け取り、1以上の複数のオーディオ入力チャンネルにおける時間についての信号特性の変化を検出し、前記1以上の複数のオーディオ入力チャンネルにおける時間についての信号特性の変化を聴覚事象の境界として特定し、ここで連続する境界と境界との間のオーディオセグメントはそのチャンネルにおける聴覚事象を構成し、少なくとも部分的に、聴覚事象及び/又は前記聴覚事象の境界に関連した信号特性の変化の程度に応答して、オーディオ出力チャンネルを生成することにより、入力チャンネル数よりも大きな数の前記オーディオ出力チャンネルを生成する。一般に、聴覚事象は、別々の分離したものとして識別されるオーディオのセグメントである。信号特性の1つの有用な尺度には、例えば、引用したCrockett and Crockett他の論文に記載されているような、オーディオの一定のスペクトル内容が含まれる。1以上のパラメータの一部又は全部は、少なくとも部分的には1以上の聴覚事象の存在又は不存在に応答して生成される。聴覚事象の境界は、時間についての閾値を越える信号特性の変化として特定することができる。あるいは、前記聴覚事象の境界に関連する信号特性における変化の程度連続的な指標に少なくとも部分的に応答して、前記1以上のパラメータの一部又は全部が生成される。原則として、本発明の特徴はアナログ領域及び/又はデジタル領域で実行することができるが、実際の実施の形態では、それぞれのオーディオ信号がデータブロック内でのサンプルとして表現されるデジタル領域で実施される場合が多い。この場合、信号特性はブロック内のオーディオのスペクトル内容とすることができ、時間についての信号特性の変化の検出が、ブロックからブロックへのオーディオのスペクトル内容の変化の検出とすることができ、時間的な聴覚事象の開始と終了の境界がデータブロックの境界と一致する。

【 0 0 1 4 】

本発明の特徴は、ここでは、他の発明の特徴を含む空間的コーディング環境において記載する。このような他の発明は、種々の出願中の本願発明の出願人であるドルビー・ラブラトリーズ・ライセンシング・コーポレーションの米国特許出願及び国際特許出に記載されており、これらの出願は本明細書中に明記されている。

【発明を実施するための最良の形態】

【 0 0 1 5 】

本発明の特徴を採用する空間的エンコーダの例を図1, 2, 及び3に示す。一般に、空間的コーダは、N個の元のオーディオ信号又はオーディオチャンネルを取り出し、M個の信号又はチャンネルを有するコンボジット信号にミキシングダウンする。ここで $M < N$ である。一般に $N = 6$ (5.1オーディオ)であり、 $M = 1$ 又は2である。同時に、種々のチャンネル間で知覚的に無音の空間的手がかりを示す低データレートのサイドチェーン信号を元のマルチチャンネル信号から抽出する。次いで、このコンボジット信号は、MPEG 2/4 AACエンコーダのような既存のオーディオコーダによりコード化することができる。前記デコーダにおいて、このコンボジット信号がデコードされ、アンパッケージ化されたサイドチェーン情報が、このコンボジットを元のマルチチャンネル信号に近似するようアップミックスするために用いられる。あるいは、前記デコーダは、このサイドチェーン情報を無視し、単にコンボジット信号を出力する。

【 0 0 1 6 】

最近の種々の技術論文(以下に引用するような)及びMPEG標準化団体に提案されている空間的コーディングシステムは一般に、チャンネル間レベル差(ILD)、チャンネル間位相差(IPD)、及びチャンネル間相関(ICC)のような元のサウンドフィールドにモデル化するためのパラメータを採用する。通常、このようなパラメータはコード化された各チャンネルの複数のスペクトル帯域を推定し、時間について動的に推定される。

本発明の特徴には、このようなパラメータの１つ以上を計算する新しい技術が含まれている。本発明の特徴についての有用な環境を説明するために、本明細書には、デコリレーションフィルタ及び元のマルチチャンネル信号の細かな時間構造を保存する技術を含む、このアップミックスされた信号をデコリレーションする方法の詳細な説明が含まれる。ここに記載した本発明の特徴についての他の有用な環境は、オーディオ素材を２チャンネルの内容から直接空間的デコーディングシステムと互換性のある素材に変換する「ブラインド」アップミキシング（補助的な制御信号なしにオーディオ信号のみに応答して動作するアップミキシング）を行うのに適したデコーダと連動して動作する空間的エンコーダの中である。このような有用な環境の特徴は、ドルビー・ラボラトリーズ・ライセンシング・コーポレーションの他の米国特許出願及び国際特許出の主題となっており、これらは知ら

10

【 0 0 1 7 】

〔コード概要〕

本発明の特徴を採用する空間的エンコーダの例を図１，２，及び３に示す。図１のエンコーダの例において、 N チャンネルの元の信号（例えばPCMフォーマットのデジタルオーディオ）は、装置又は機能（時間から周波数へ）２により、よく知られた短時間離散フーリエ変換（STDFT）のような適切な時間・周波数変換を用いて周波数領域に変換される。一般に、この変換は、１以上の周波数ビンが耳の臨界帯域を近似する帯域にグループ化されるようにして行われる。しばしば「空間的パラメータ」と称されるチャンネル間振幅差又はレベル差（ILD）、チャンネル間時間差又は位相差（IPD）、及びチャンネル間相関（ICC）の推定値は、（空間的サイド情報の導出）機能４の装置により帯域毎に計算される。以下に詳細を説明するように、聴覚シーン分析装置又は分析機能（聴覚シーン分析）６はまた、 N チャンネルの元の信号を受け取り、この明細書の別の箇所

20

説明するように、装置又は機能４により空間的パラメータの生成に作用を及ぼす。聴覚シーン分析６は、 N チャンネルの元の信号におけるどんなチャンネルの組み合わせも採用することができる。説明を簡単にするために別に示したが、装置又は機能４及び６は単一の装置又は機能とすることができる。 N チャンネルの元の信号に対応する M チャンネルコンボジット信号がまだ存在しない場合は（ $M < N$ ）、ダウンミキサー又はダウンミキシング機能（ダウンミックス）８にて、 N チャンネルの元の信号を M チャンネルコンボジット信号にダウンミックスするために空間的パラメータを用いることができる。次いで、 M チャンネルコンボジット信号を、（周波数から時間へ）装置又は機能１０により、装置又は機能２とは逆の変換を行う適切な周波数・時間変換を用いて、時間領域に戻すことができる。装置又は機能４からの空間的パラメータ及び時間領域における M チャンネルコンボジット信号は、次に、損失のあるビットリダクションエンコーディング及び／又は損失のないビットリダクションエンコーディングを含む装置又は機能（フォーマット）１２において、例えば、シリアルビットストリーム又はパラレルビットストリームの適切な形にフォーマットされる。フォーマット１２からの出力の形式は、本発明において重要ではない。

20

30

【 0 0 1 8 】

この明細書を通して、同じ参照番号は、同じ構成又は同じ機能を果たす装置又は機能に用いられる。装置又は機能が類似の機能構成であるが例えば、付加的な入力があるというように少し差異がある場合は、この少し違うが類似の装置又は機能は、プライムマーク（例えば４'）をつけて指定する。また、種々のブロック図は、単一の機能又は装置に種々の機能または全ての機能を統合した実際の実施の形態においても、機能を実行する機能又は装置を別々に示した、機能ブロック図であることは了解されよう。例えば、図１のようなエンコーダの実施の形態では、コンピュータプログラムの一部が種々の機能を実行するコンピュータプログラムを動作させる、デジタル信号プロセッサにより実施することができる。以下の「実施」の表題をつけた部分を参照のこと。

40

【 0 0 1 9 】

あるいは、図２に示すように、 N チャンネルの元の信号と、それに関連する M チャンネルコンボジット信号とが（それぞれが、例えば、複数チャンネルのPCMデジタルオーデ

50

ィオである場合)エンコーダへの入力として利用可能である場合、これらは、時間から周波数への変換2(明確にするために2つのブロックで示されている)で同時に処理することができ、図4の装置又は機能と類似するが2組の入力信号を受け取る装置又は機能(空間的サイド情報の導出)4'により、Mチャンネルコンポジット信号の空間的パラメータに関して、Nチャンネルの元の信号の空間的パラメータを計算することができる。Nチャンネルの元の信号のセットが利用できない場合は、利用可能なMチャンネルコンポジット信号を時間領域(不図示)でアップミックスし「Nチャンネルの元の信号」を生成することができる。マルチチャンネル信号はそれぞれ、図1の例において時間・周波数変換装置又は機能への入力のセットを提供する。図1のエンコーダ及び図2の代案の両方において、Mチャンネルコンポジット信号と空間的パラメータとは、図1の例に示すように、装置又は機能(フォーマット)12により適切な形式にエンコードされる。図1のエンコーダの例に示すように、フォーマット12からの出力の形式は、本発明において重要ではない。以下に詳細に説明するように、聴覚シーン分析装置又は機能

(聴覚シーン分析)6'は、Nチャンネルの元の信号とMチャンネルコンポジット信号を受け取り、本明細書の別の箇所で説明するように、装置又は機能4'における空間的パラメータの生成に作用を及ぼす。説明を簡単にするために別に示したが、装置又は機能4'及び6'は単一の装置又は機能とすることができる。聴覚シーン分析6'は、Nチャンネルの元の信号及びMチャンネルコンポジット信号におけるどんなチャンネルの組み合わせも採用することができる。

【0020】

本発明の特徴を採用するエンコーダのさらに他の例では、適切なデコーダと共に、「ブラインド」アップミキシングを実行する、空間的コーディングエンコーダを用いることを特徴とする。このようなエンコーダは、2006年5月26日に、Seefeldt他により出願された、同時係属の国際出願PCT/US2006/020882、表題「Channel Reconfiguration with Side Information」に記載されている。なおこの出願は参照として全文を本明細書に編入する。図1及び図2の空間的コーディングエンコーダは、空間的コーディングパラメータを生成する上で、既存のNチャンネルの空間的イメージをここでは採用する。多くの場合、しかしながら、空間的コーディングのアプリケーションに対するオーディオコンテンツプロバイダは、豊富なステレオコンテンツを有するが、元のマルチチャンネルコンテンツを欠いている。この問題に対するひとつの方法は、空間的コーディングの前にブラインドアップミキシングシステムを用いて、2チャンネルのステレオコンテンツをマルチチャンネル(例えば、5.1チャンネル)コンテンツに変換することである。上述のように、ブラインドアップミキシングシステムは、元の2チャンネルのステレオ信号自身をマルチチャンネル信号に合成するためにのみ有用な情報を用いる。このようなアップミキシングシステムの多くは、例えばドルビープロロジック(Dolby Pro Logic)II(「Dolby」、「Pro Logic」、及び「Pro Logic II」はドルビー・ラボラトリーズ・ライセンシング・コーポレーションの登録商標である)。空間的コーディングエンコーダと結合されたとき、このコンポジット信号は、ここで図1のエンコーダの例で示したように、ブラインドアップミックスされた信号をダウンミキシングすることによりエンコーダにて生成することができ、又は、ここで図2エンコーダの例で示したように、既存の2チャンネルのステレオ信号を用いることができる。

【0021】

代案として、図3の例に示したように、空間的エンコーダブラインドアップミキサーの一部として採用することができる。このようなエンコーダは、中間的なアップミックスされた信号を必要としないで2チャンネルのステレオ信号から直接、望ましいマルチチャンネルの空間的イメージのパラメトリックモデルを合成するために既存の空間的コーディングパラメータを利用する。結果生じたエンコードされた信号は、既存の空間的デコーダと互換性がある(このデコーダは、望ましいブラインドアップミックスを生成するためにサイド情報を用いてもよく、又は、リスナーに元の2チャンネルのステレオ信号を提供するためにこのサイド情報を無視してもよい)。

【 0 0 2 2 】

図 3 の例のエンコーダにおいて、M チャンネルの元の信号（例えば、P C M フォーマットのデジタルオーディオの複数のチャンネル）は N チャンネルの元の信号におけるどんなチャンネルの組み合わせも採用することができる（時間から周波数へ）装置又は機能 2 により、他のエンコーダの例で示したように、よく知られた短時間離散フーリエ変換（S T D F T）のような適切な時間・周波数変換を用いて、1 以上の周波数ビンが耳の臨界帯域を近似する帯域にグループ化されるようにして、周波数領域に変換される。一般に、この変換は、1 以上の周波数ビンが耳の臨界帯域を近似する帯域にグループ化されるようにして行われる。空間的パラメータは、（空間的サイド情報としてのアップミックス情報の導出）装置又は機能 4 ” により、帯域毎に計算される。以下に詳細を説明するように、聴覚

10

シーン分析装置又は機能（聴覚シーン分析）6 ” は、M チャンネルの元の信号を受け取り、本明細書の別の箇所で説明するように、装置又は機能 4 ” における空間的パラメータの生成に作用を及ぼす。説明を簡単にするために別に示したが、装置又は機能 4 ” 及び 6 ” は単一の装置又は機能とすることができる。装置又は機能 4 ” からの空間的パラメータ及び M チャンネルコンボジット信号（まだ時間領域にある）は、次に、損失のあるビットリダクションエンコーディング及び / 又は損失のないビットリダクションエンコーディングを含む装置又は機能（フォーマット）1 2 において、例えば、シリアルビットストリーム又はパラレルビットストリームの適切な形にフォーマットされる。図 1 と図 2 のエンコーダの例に示したように、フォーマット 1 2 から出力された形式は本発明において重要ではない。さらに、図 3 のエンコーダの詳細は、以下に「ブ

20

【 0 0 2 3 】

図 4 に示した空間的デコーダは、図 1、図 2、及び図 3 に示したエンコーダのようなエンコーダからコンボジット信号と空間的パラメータとを受け取る。ビットストリームは、空間的パラメータサイド情報と共に M チャンネルコンボジット信号を生成するために、装置又は機能（フォーマット）2 2 によりデコードされる。コンボジット信号は、装置又は機能（時間から周波数へ）2 4 により周波数領域に変換され、そこで、デコードされた空間的パラメータが装置又は機能（空間的サイド情報の適用）2 6 に適用され、周波数領域に N チャンネルの元の信号を生成する。少ない数のチャンネルからこのような数の多いチャンネルを生成することはアップミキシングである（装置又は機能 2 6 は「アップミキサー」と位置付けることができる）。最後に、周波数から時間変換（周波数から時間へ）2 8（図 1、2、及び 3 の時間から装置への変換装置又は機能 2 の逆変換）が N チャンネルの元の信号（もし、エンコーダが図 1 及び図 2 の例に示されたものである場合）の近似、又は、図 3 の M チャンネルの元の信号のアップミックスの近似を生成するために適用される。

30

【 0 0 2 4 】

本発明の他の特徴は、オーディオシーン分析の機能としてのアップミキシングを行う「スタンドアローン」又は「シングルエンド」プロセッサに関する。このような本発明の特徴は以下に図 5 に示した例のの詳細説明として記載されている。

【 0 0 2 5 】

本明細書全体を通じて、本発明とその環境についての特徴をさらに規定するために、以下の記号を用いる：

40

【 式 1 】

【 0 0 2 6 】

x は元の N チャンネル信号； y は M チャンネルコンポジット信号 ($M=1$ 又は 2) である； z は、 ILD 及び IPD パラメータのみを用いて y からアップミックスされた N チャンネル信号である； \hat{x} は z にデコリレーションを適用した後の元の信号 x の最終的な推定である； x_i 、 y_i 、 z_i 、 及び \hat{x}_i は信号 x 、 y 、 z 、 及び \hat{x} のチャンネル i における信号である； $X_i[k, t]$ 、 $Y_i[k, t]$ 、 $Z_i[k, t]$ 、 及び $\hat{X}_i[k, t]$ は、ビン k 及び時間ブロック t におけるチャンネル x_i 、 y_i 、 z_i 、 及び \hat{x}_i の $STDF$ T である。

コンポジット信号 y を生成するためのアクティブダウンミキシングは、以下の式により 10
帯域毎に周波数領域で行われる

【数 1】

$$Y_i[k, t] = \sum_{j=1}^N D_{ij}[b, t] X_j[k, t], \quad kb_b \leq k < ke_b \quad (1)$$

【0027】

ここで、 kb_b は帯域 b の低い方のビンであり、 ke_b は帯域 b の高い方のビンであり、 $D_{ij}[b, t]$ は、元のマルチチャンネル信号のチャンネル j に関するコンポジット信号のチャンネル i についての複素ダウンミックス係数である。

【0028】

アップミックスされた信号 z は、コンポジット y から周波数領域で同様に計算する、

【数 2】

$$Z_i[k, t] = \sum_{j=1}^M U_{ij}[b, t] Y_j[k, t], \quad kb_b \leq k < ke_b \quad (2)$$

【0029】

ここで、 $U_{ij}[b, t]$ は、コンポジット信号のチャンネル j に関するアップミックス信号のチャンネル i についてのアップミックス係数である。 ILD パラメータと IPD パラメータは、アップミックス係数の振幅及び位相として得られる。

【数 3】

$$ILD_y[b, t] = |U_y[b, t]| \quad (3a)$$

$$ILD_y[b, t] = \angle U_y[b, t] \quad (3b)$$

最終的な推定信号 \hat{x} はアップミックスされた信号 z にデコリレーションを適用することにより導き出される。特定のデコリレーション技術は本発明において重要ではない。1つの技術は、Breebaart による、2003年10月30日に公開された国際出願公開WO 03/090206 A1、表題「Signal Synthesizing」に記載されている。あるいは、元の信号 x の特性に基づき他の1つ又は2つの技術を選択することもできる。デコリレーションの程度を変化させる第1番目の技術は、Seefeldt 等による2006年3月9日に公開された国際出願公開WO 2006/026452、表題「Multichannel Decorrelation in Spatial Audio Coding」に記載されている。Vinton 等による、2006年3月9日に公開された国際出願公開WO 2006/026161、表題「Temporal Envelope Shaping for Spatial Audio Coding Using Frequency Domain Wiener Filtering」に記載されている第2の技術では、推定した \hat{x} における x の各チャンネルの元の時間エンベロープを復元するためにスペクトルウィナーフィルタを $Z_i[k, t]$ に適用する。

【式 2】

【0030】

[コードパラメータ]

ILDパラメータ、IPDパラメータ、ICCパラメータ、及び「SWF」空間的パラメータの計算と適用について詳細をここに記載する。もし、先に引用したVinton等の特許出願に記載のデコリレーション技術を適用するならば、空間的エンコーダは、適切な「SWF」(空間的ウィナーフィルタ)パラメータも生成する。最初の3つのパラメータに共通するのは、元のマルチチャンネル信号 x の各帯域における共分散マトリックスの時間により変化する推定と独立していることである。 $N \times N$ の共分散マトリックス $R[b, t]$ は、 x の各チャンネルに亘る各帯域のスペクトル係数同士のドットプロダクトである(「ドットプロダクト」は、スカラー積としても知られている、2つのベクトルを使いスカラー量にする2項演算である)。時間の変化に対するこの推定を固定させるために、以下に示すように単純なリーキーインテグレータ(ローパスフィルタ)を用いて平滑化される。

【数 4】

$$R_y[b, t] = \lambda R_y[b, t-1] + \frac{1-\lambda}{ke_b - kb_b} \sum_{k=kb_b}^{ke_b-1} X_i[k, t] X_j^*[k, t], \quad (4)$$

10

20

30

40

50

【式 3】

【0031】

ここで $R_{ij}[b, t]$ は、 $R[b, t]$ の*i*番目の行、*j*番目の列の要素であり、時間ブロック *t* での帯域 *b* における *x* の *i* 番目のチャンネルと *j* 番目のチャンネルとの間の共分散を示し、 λ は、平滑時定数である。

[ILDとIPD]

元の信号 *x* のアクティブなダウンミックス *y* の生成し、ダウンミックス *y* をアップミックスして元の信号 *x* の推定値 *z* にする、ILD パラメータと IPD パラメータの計算について考える。以下の説明において、パラメータはサブ帯域 *b* と時間ブロック *t* について計算され、分かりやすくするために、この帯域と時間指標とは、はっきりとは示さない。加えて、ダウンミックス処理 / アップミックス処理を表現するベクトルを採用する。まず、コンポジット信号中のチャンネル数が $M = 1$ の場合を考え、次いで、 $M = 2$ の場合を考える。

10

【数 5】

[$M = 1$ のシステム]

$N \times 1$ の複素ランダムベクトル *x* としてサブ帯域 *b* 中の元の *N* チャンネル信号を表現するために、この元のベクトルの推定 *z* は、以下のようなダウンミキシング処理及びアップミキシング処理により計算される。

20

$$\mathbf{z} = \mathbf{u} \mathbf{d}^T \mathbf{x} \quad (5)$$

ここで *d* は $N \times 1$ の複素ダウンミキシングベクトル、*u* は $N \times 1$ の複素アップミキシングベクトルである。*z* と *x* 間での 2 乗平均誤差が最小になるベクトル *d* とベクトル *u* は以下のように示される。

$$\mathbf{U}^* = \mathbf{d} = \mathbf{v}_{\max} \quad (6)$$

ここで \mathbf{v}_{\max} は、*x* の共分散マトリックスである *R* の最大の固有値に対応する固有値である。

30

【0032】

最小 2 乗法において最適であるが、この手法では容認できない知覚可能なアーティファクトを持ち込むことがある。特に、元の信号の低いレベルのチャンネルの誤差を最小にするとき、この手法では低いレベルのチャンネルを「消去 (zero out)」してしまう傾向がある。知覚的に満足のできるように信号のダウンミックスとアップミックスの両方を行う目的で、良好な手法は、ダウンミックスした信号が固定した量の元の各信号チャンネルを具備し、各アップミックスしたチャンネルが元のチャンネルと等しくなるようなものである。しかしながらチャンネル間での相殺を最小限にするために、ダウンミキシングの前に各チャンネルを回転する上で、最小 2 乗法の段階を用いるのは有益である。同様に、アップミックスにおいて最小 2 乗法の段階を用いることは、チャンネル同士の元の位相関係を復元するために役に立つ。好ましい手法でのダウンミキシングベクトルは以下のように表すことができる。

40

【数 6】

$$\mathbf{d} = \alpha \bar{\mathbf{d}} \cdot e^{j\angle \mathbf{v}_{\max}} \quad (7)$$

【式 4】

【0033】

50

ここで $\bar{\mathbf{d}}$ は、例えば、標準のITUダウンミキシング係数を含むことのある固定のダウンミキシングベクトルである。ベクトル $\angle \mathbf{v}_{\max}$ は複素固有値の位相 \mathbf{v}_{\max} に等しく、演算 $\mathbf{a} \cdot \mathbf{b}$ は、2つのベクトルの要素毎の乗算を表す。スカラー α は、ダウンミックスした信号の指数が、固定のダウンミキシングベクトルにより重み付けられた元の信号チャンネルの指数の和と等しくなるような、正規化係数であり、以下のように計算される

【数 7】

$$\alpha = \sqrt{\frac{\sum_{i=1}^N \bar{d}_i^2 R_{ii}}{(\bar{\mathbf{d}} \cdot e^{j\angle \mathbf{v}_{\max}}) \mathbf{R}(\bar{\mathbf{d}} \cdot e^{j\angle \mathbf{v}_{\max}})^H}}, \quad (8)$$

【式 5】

【0 0 3 4】

ここで d_i は、ベクトル $\bar{\mathbf{d}}$ の i 番目の要素を表し、 R_{ij} は、共分散マトリックス \mathbf{R} の i 番目の行と、 j 番目の列の要素を表す。固有ベクトル \mathbf{v}_{\max} の使用により、複素スカラー乗算まで一意的であるという問題がでてくる。この固有ベクトルを一意的にするため、最も支配的なチャンネル g に対応する要素がゼロ位相であるような制限を加える。ここで、この支配的なチャンネルは最も大きなエネルギーを持つチャンネルとして以下のように定義される。

【数 8】

$$g = \arg \max_i (R_{ii}[b, t]). \quad (9)$$

【式 6】

【0 0 3 5】

アップミキシングベクトル \mathbf{u} は、 \mathbf{d} と同様に表現することができる。

【数 9】

$$\mathbf{u} = \beta \cdot \bar{\mathbf{u}} \cdot e^{-j\angle \mathbf{v}_{\max}}. \quad (10)$$

【式 7】

【0 0 3 6】

固定されたアップミキシングベクトル $\bar{\mathbf{u}}$ は、以下のように選択される。

【数 1 0】

$$\bar{u}_i \bar{d}_i = 1, \quad (11)$$

【式 8】

【0 0 3 7】

10

20

30

40

正規化ベクトル β の各要素は、アップミックスされた信号の各チャンネルにおける指数が、元の信号の対応するチャンネルの指数に等しくなるよう計算される。

【数 1 1】

$$\beta_i = \sqrt{\frac{\bar{d}_i^2 R_{ii}}{\sum_{j=1}^N \bar{d}_j^2 R_{jj}}}.$$

(12)

10

【式 9】

【0 0 3 8】

ILD パラメータ及び IPD パラメータは、アップミキシングベクトル u の振幅及び位相として得られる。

【数 1 2】

$$ILD_i[b, t] = |u_i|$$

$$IPD_i[b, t] = \angle u_i$$

20

$M=2$ System

【0 0 3 9】

[$M = 2$ のシステム]

(1) に類似のマトリックス式は、 $M = 2$ のとき以下のように記述することができる。

【数 1 3】

$$\mathbf{z} = \begin{bmatrix} \mathbf{u}_L & \mathbf{u}_R \end{bmatrix} \begin{bmatrix} \mathbf{d}_L^T \\ \mathbf{d}_R^T \end{bmatrix} \mathbf{x},$$

(14)

30

【0 0 4 0】

ここで、2 チャンネルのダウンミックスされた信号は左右のチャンネルを有するステレオ対に対応し、これらのチャンネルは両方とも対応するダウンミックスベクトルとアップミックスベクトルを有する。これらのベクトルは、 $M = 1$ のシステムの場合と同様に以下のように表すことができる。

【数 1 4】

$$\mathbf{d}_L = \alpha_L \bar{\mathbf{d}}_L \cdot e^{j\theta_{LR}} \quad (15a)$$

$$\mathbf{d}_R = \alpha_R \bar{\mathbf{d}}_R \cdot e^{j\theta_{LR}} \quad (15b)$$

40

$$\mathbf{u}_L = \beta_L \cdot \bar{\mathbf{u}}_L \cdot e^{-j\theta_{LR}} \quad (15c)$$

$$\mathbf{u}_R = \beta_R \cdot \bar{\mathbf{u}}_R \cdot e^{-j\theta_{LR}} \quad (15d)$$

【0 0 4 1】

5 . 1 チャンネルの元の信号に対して、固定のダウンミキシングベクトルは、標準の ITU ダウンミックス係数と等しくなるよう設定することができる (チャンネルは、L , C , R , L s , R s , L F E の順序であるとみなしている) 。

【数 1 5】

$$\bar{\mathbf{d}}_L = \begin{bmatrix} 1 \\ 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \bar{\mathbf{d}}_R = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 1 \\ 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}. \quad (16)$$

10

【0 0 4 2】

要素的な制限は、

【数 1 6】

$$\bar{d}_{Li}\bar{u}_{Li} + \bar{d}_{Ri}\bar{u}_{Ri} = 1, \quad (17)$$

【0 0 4 3】

対応する固定のアップミックスベクトルは以下のようにして得られる。

【数 1 7】

$$\bar{\mathbf{u}}_L = \begin{bmatrix} 1 \\ 1/\sqrt{2} \\ 0 \\ \sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \bar{\mathbf{u}}_R = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 1 \\ 0 \\ \sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}. \quad (18)$$

20

【0 0 4 4】

元の信号のイメージを2チャンネルのダウンミックスされたステレオ信号中に保持するためには、元の信号の左右のチャンネルの位相は回転させるべきでなく他のチャンネルの位相、特に中央の位相は、左右にダウンミックスした量と同じだけ回転すべきであることが分かった。これは、左チャンネルに関連する共分散マトリックスの要素と、右チャンネルに関連する共分散マトリックスの要素との間の、重み付けした合計の角として、以下のように、共通のダウンミックス位相回転を計算することにより実行する。

30

【数 1 8】

$$\theta_{LRi} = \angle(d_{Li}d_{Li}R_{Li} + d_{Ri}d_{Ri}R_{Ri}), \quad (19)$$

【式 1 0】

40

【0 0 4 5】

ここで、lとrは、左右のチャンネルに対応する元のベクトル \mathbf{x} の指標である。式(10)で与えられたダウンミックスベクトルと共に、必要に応じて、定式は、 $\theta_{LRl} = \theta_{LRr} = 0$ となる。最後に、式(9a～9d)の正規化係数を、M=1のシステムにおける式(4)及び(7)のように計算する。ILDパラメータとIPDパラメータとは以下の式で与えられる。

【数 19】

$$ILD_{i1}[b, t] = |u_{Li}| \quad (20a)$$

$$ILD_{i2}[b, t] = |u_{Ri}| \quad (20b)$$

$$IPD_{i1}[b, t] = \angle u_{Li} \quad (20c)$$

$$IPD_{i2}[b, t] = \angle u_{Ri} \quad (20d)$$

【0046】

10

しかし、式(12)中の固定のアップミックスベクトルと共に、これらのパラメータのうちのいくつかは常にゼロであり、明示的にサイド情報として伝達する必要はない。

【0047】

[デコリレーション技術]

ILDパラメータとIPDパラメータとをコンボジット信号 y に適用することにより、中間チャンネルレベル及びアップミックスされた信号 z 中の元の信号 x の位相関係を復元する。これらの関係が元の空間的イメージの有意な手がかりとなり、このアップミックスされた信号 z は、そのそれぞれのチャンネルがコンボジット y 中の同じ少数のチャンネル(1又は2)から導き出されたものなので、高い相関関係を維持している。その結果、 z の空間的イメージは、元の信号 x と比べて、しばしば崩れたサウンドとなる。だからこそ、チャンネル間の相関関係が元の信号 x によく近似するよう信号 z を修正することが好ましいのである。この目的を達成するための2つの技術について記載する。第1の技術では、 z の各チャンネルに適用するデコリレーションの程度を制御するために、ICCの尺度を用いる。第2の技術、スペクトルウィナーフィルター(SWF)、では、周波数領域で信号 z をフィルターリングすることにより、 x の各チャンネルの元の時間エンベロープを復元する。

20

【0048】

[ICC]

【式11】

【0049】

30

元の信号の正規化されたチャンネル間相関マトリックス $C[b, t]$ は、共分散マトリックス $R[b, t]$ から以下のように計算することができる。

【数20】

$$C_{ij}[b, t] = \frac{|R_{ij}[b, t]|}{\sqrt{R_{ii}^2[b, t] R_{jj}^2[b, t]}} \quad (21)$$

【式12】

40

【0050】

$C[b, t]$ の i 番目の行 j 番目の列における要素は、信号 x のチャンネル i とチャンネル j との間の正規化された相関関係を計測する。修正後の相関マトリックスが $C[b, t]$ に等しくなるよう z を修正することが理想である。しかしながらサイドチェーンのデータレートの制約により、その近似として、各チャンネルと参照チャンネルとの間の相関関係が $C[b, t]$ の対応する要素とほぼ等しくなるよう z を修正する。参照値は式 (9) で定義された優勢なチャンネル g として選択される。サイド情報として送られてきた ICC パラメータは、次に、相関マトリックス $C[b, t]$ の行 g に等しく設定される。

10

【数 2 1】

$$ICC_i[b, t] = C_{gi}[b, t]. \quad (22)$$

【式 1 3】

【0 0 5 1】

デコーダでは、帯域毎に、 z とデコリレーションされた信号 \tilde{z} との線形結合を制御するために ICC パラメータが用いられる。

20

【数 2 2】

$$\hat{X}_i[k, t] = ICC_i[b, t]Z[k, t] + \sqrt{1 - ICC_i^2[b, t]}\tilde{Z}_i[k, t] \quad \text{for } kb_b \leq k \leq ke_b \quad (23)$$

【式 1 4】

【0 0 5 2】

デコリレーションされた信号 \tilde{z} は、信号 z の各チャンネルを一意的 LTI デコリレーションフィルタでフィルタすることにより生成する。

30

【数 2 3】

$$\tilde{z}_i = h_i * z_i. \quad (24)$$

【式 1 5】

【0 0 5 3】

フィルタ h_i は、 z 及び \tilde{z} の全てのチャンネルがお互いにほぼデコリレーションされるように設計する。

40

【数 2 4】

$$E\{z_i \tilde{z}_j\} \cong 0 \quad i = 1..N, j = 1..N$$

(25)

$$E\{\tilde{z}_i \tilde{z}_j\} \cong 0 \quad i = 1..N, j = 1..N, i \neq j$$

10

【式 1 6】

【0 0 5 4】

式(17)と式(19)における条件及び、 z が強く相関関係を持っているという前記条件から、最終的にアップミックスされた信号 \hat{x} の優勢なチャンネルと他のチャンネルとの間の相関関係は以下ようになる。

【数 2 5】

$$\hat{C}_{gi}[b, t] \cong ICC_i[b, t], \quad (26)$$

20

【0 0 5 5】

これが望ましい効果である。

【0 0 5 6】

本明細書で引用した国際公開WO 03 / 090206 A1において、単一の混合信号から、2チャンネルステレオが合成されるパラメトリックステレオコーディングシステムに対する、デコリレーション技術が提示されている。それ自体では、単一のデコリレーションフィルタだけが必要とされる。そこで提案されているフィルタは、周波数が増大するにつれて、時間遅れが最大値からゼロに直線的に減少する周波数により変化する時間遅れである。固定時間遅れと比べて、このようなフィルタは、式(17)に示されるように、フィルタされた信号をフィルタされていない信号に加えたとき、知覚可能なエコーが入ることなく顕著なデコリレーションを行うことができるという好ましい特性を有する。加えて、この周波数により変化する時間遅れは、周波数と共に増加するスペースを持った刻み目をスペクトル中に導入する。これは、固定時間遅れにより、櫛型フィルタによる直線的なスペースより自然な音として知覚される。

30

【0 0 5 7】

前記書類WO 03 / 090206 A1において、提案されたフィルタに関する唯一チューニング可能なパラメータはその長さである。引用したSeefeldt他による、国際公開WO 2006 / 026452において開示された発明の特徴は、N個の必要なデコリレーションフィルタのそれぞれに対して、より柔軟性のある周波数により変化する時遅れを導入することである。各フィルタのインパルス応答は、瞬間周波数が、シーケンスの継続中に からゼロに単調減少する、有限長の正弦曲線のシーケンスとして定められる。

40

【数 2 6】

$$h_i[n] = G_i \sqrt{|\omega'_i(n)|} \cos(\phi_i(n)), \quad n = 0..L_i$$

$$\phi_i(t) = \int \omega_i(t) dt, \quad (27)$$

【式 1 7】

50

【 0 0 5 8 】

ここで $\omega_i(t)$ は、単調減少の瞬時周波数関数であり、 $\omega_i'(t)$ は、この瞬時周波数関数の一次導関数であり、 $\phi_i(t)$ は、この瞬時周波数関数を積分することにより得られる瞬時位相であり、そして、 L_i は、フィルターの長さである。乗法項 $\sqrt{|\omega_i'(n)|}$ は、 $h_i[n]$ の周波数応答を全周波数でフラットにするために必要とされ、ゲイン G_i は、以下のように計算される。

10

【 数 2 7 】

$$\sum_{n=0}^{L_i} h_i^2[n] = 1. \quad (28)$$

【 0 0 5 9 】

指定されたインパルス応答は、チャープ（鳥のさえずりのような音）のようなシーケンスを持ち、結果として、このようなフィルターでオーディオ信号をフィルターすると、トランジエントの位置に可聴の「チャープング」アーティファクトを生じる。この効果は、

20

【 数 2 8 】

$$h_i[n] = G_i \sqrt{|\omega_i'(n)|} \cos(\phi_i(n) + N_i[n]). \quad (29)$$

【 0 0 6 0 】

の地位差は端数の分散を持つホワイトが薄ノイズと同等のこのノイズシーケンス $N_i[n]$ を作ることは、チャープというより十分ノイズに似たインパルス応答音を作る一方、周波数と $\omega_i(t)$ により特定される時間遅れとの関係は広く維持される。式 (23) のフィルターは、3つのフリーパラメータすなわち、 $\omega_i(t)$ 、 $L_i(t)$ 、及び $N_i[n]$ を持つ。N個のフィルターから相異なるこれらのパラメータを選択することにより、式 (19) における好ましいデコリレーション条件が満たされる。

30

【 式 1 8 】

【 0 0 6 1 】

デコリレーション信号 \tilde{z} は、時間領域で畳み込みを行うことにより生成することができるが、以下のように変換係数 z で乗算することによるフィルターを通すことによりもっと効率的に実行できる。

【 数 2 9 】

40

$$\tilde{Z}_i[k, t] = H_i[k] Z_i[k, t], \quad (30)$$

【 0 0 6 2 】

ここで $N_i[k]$ は、 $h_i[n]$ の DFT に等しい。厳密に言えば、この変換係数の乗算は時間領域における巡回畳み込みに対応するが、STDF T 分析と合成窓とデコリレーションフィルターの長さを適切に選択することにより、この演算は通常の畳み込みと等価となる。図 6 は、適切な分析窓と合成窓のペアを示す。この窓は 75% 重複するように設計され、デコリレーションフィルターを適用したとき、サーキュラーエイリアシング (circular aliasing) を避けるために、分析窓はメインローブに続いて顕著

50

なぜゼロパッド (zero padded) 領域を有する。各デコリレーションフィルターの長さが、図 6 の L_{\max} で示されるゼロパッド領域の長さ以下に選定される限り、式 (30) の乗算は時間領域における通常の畳み込みに対応する。ILD パラメータ、IPD パラメータ、及び ICC パラメータの帯域を横切つての変化に伴う因果関係のない畳み込みの漏れに対応するために、分析窓のメインローブに続くゼロパッドに加えて、少量の先行ゼロパッドを適用してもよい。

【式 19】

【0063】

[スペクトル・ウィナー・フィルタリング]

10

先の章では、帯域対帯域、ブロック対ブロックでデコリレーションの程度を制御するために ICC パラメータを用いることにより、元の信号 x のチャンネル間デコリレーションがどのように推定 \hat{x} に復元されるかを示した。

多くの信号に対してこれは非常にうまく働く。しかし、拍手のような信号に対しては、元のサウンドフィールドの発散性を再現させるためには、元の信号の個々のチャンネルの細かい時間構造を復元する必要がある。この細かい構造は、普通ダウンミキシングプロセスで壊されてしまい、採用された STDFTHop サイズと変換長さにより、時間での ILDP パラメータ、IPD パラメータ、及び ICC パラメータが十分にそれを復元できない。引用した Vinton 等による国際出願 WO 2006/026161 に記載された、SWF 技術は、このような問題の場合に、ICC に基づく技術に有利に取って代わるものである。スペクトルウィナーフィルタリング (SWF) と称される新しい方法は、周波数領域での畳み込みが時間領域での乗算と等価であるという、時間と周波数の双対性をうまく利用するものである。スペクトルウィナーフィルタリングにおいては、空間的デコーダの各出力チャンネルのスペクトルに FIR フィルターを適用し、出力チャンネルの時間的エンベロープを修正して、元の信号の時間エンベロープにうまく一致させる。この技術は、スペクトル領域での畳み込みにより時間エンベロープを修正する点で、MPEG-2/4 AAC で採用されるノイズ整形 (TNS) アルゴリズムに類似する。しかし、SWF アルゴリズムは、TNS とは違って、単一の末端であり、デコーダにのみ適用される。さらに、SWF アルゴリズムでは、フィルターをコーディングノイズではなく信号の時間的エンベロープを調整するために設計しているので、違ったフィルター設計上の制約を受ける。この空間的エンコーダは、デコーダに元の時間的エンベロープを適用するために必要な時間領域における乗算的な変化を表す、スペクトル領域において FIR フィルターを設計しなければならない。このフィルターの問題は、しばしばウィナーフィルターデザインと称される最小 2 乗問題として定式化することができる。しかし、時間領域において設計され適用されるウィナーフィルターの一般的な応用例とは異なり、ここで提案されたこのフィルター処理は、スペクトル領域において設計され適用される。

20

30

【0064】

スペクトル領域におけるフィルター設計の最小 2 乗問題は、以下の通りである。すなわち $X_i[k, t]$ とフィルターされた $Z_i[k, t]$ との間の誤差を最小限にする 1 組のフィルター係数 $a_i[k, t]$ を以下の通り計算する。

40

【数 30】

$$\min_{a_i[k, t]} \left[E \left\{ X_i[k, t] - \sum_{m=0}^{L-1} a_i[m, t] Z_i[k-m, t] \right\}^2 \right], \quad (31)$$

【0065】

ここで E は、スペクトルビン k についての期待値演算子であり、 L は、設計したフィルターの長さである。ここで、 $X_i[k, t]$ と $Z_i[k, t]$ は、複素数なので、一般に、

50

$a_i[k, t]$ も複素数となる。式 (31) は、マトリックス表現を使って次のように書き換えることができる。

【数 3 1】

$$\min_{\mathbf{A}} \left\{ E \left\{ \mathbf{X}_k - \mathbf{A}^T \mathbf{Z}_k \right\} \right\}, \quad (32)$$

【0066】

ここで

【数 3 2】

$$\mathbf{X}_k = [X_i[k, t]],$$

10

$$\mathbf{Z}_k^T = [Z_i[k, t] \quad Z_i[k-1, t] \quad \cdots \quad Z_i[k-L+1, t]],$$

【0067】

そして

【数 3 3】

$$\mathbf{A}^T = [a_i[0, t] \quad a_i[1, t] \quad \cdots \quad a_i[L-1, t]].$$

【0068】

各フィルター係数について、式 (32) の偏導関数をゼロに設定することにより、この最小化問題の解は以下となる。

20

【数 3 4】

$$\mathbf{A} = \mathbf{R}_{ZZ}^{-1} \mathbf{R}_{ZX}, \quad (33)$$

【0069】

ここで

【数 3 5】

$$\mathbf{R}_{ZZ} = E \{ \mathbf{Z}_k \mathbf{Z}_k^H \},$$

30

$$\mathbf{R}_{ZX} = E \{ \mathbf{Z}_k \mathbf{X}_k^H \},$$

【式 20】

【0070】

エンコーダにおいて、最適な SWF 係数は、元の信号の各チャンネルに対して式 (33) により計算されて空間的サイド情報として送られる。デコーダにおいて、この係数は、以下のようにアップミックスされたスペクトル $Z_i[k, t]$ に適用され最終的な推定 $\hat{x}_i[k, t]$ が生成される。

40

【数 3 6】

$$\hat{X}_i[k, t] = \sum_{m=0}^{L-1} a_i[m, t] Z_i[k-m, t],$$

(34)

【0071】

図 7 は、SWF 処理の性能を示す。最初の 2 つのプロットは DFT 処理ブロック中の仮想の 2 つのチャンネル信号を示す。この 2 つのチャンネルを結合して 1 つのチャンネルに

50

コンポジットしたものを3番目のプロットに示した。ここでダウンミックスプロセスにおいて、2番目のプロット中の信号の詳細な時間的構成がなくなっていることが明らかである。予想したとおり、元の2番目のチャンネルの詳細な時間的構成の推定値が置き換えられている。2番目のチャンネルがSWF処理を用いないでアップミックスされていた場合は、時間的エンベロープは、3番目のプロットに示したコンポジット信号のようにフラットとなっていたであろう。

【0072】

〔ブラインドアップミキシング〕

図1の例と図2の例の空間的エンコーダは、既存のNチャンネル（通常5.1）信号の空間的イメージのパラメトリックモデルの近似を考えており、これにより、このイメージの近似がNより少ないチャンネルを有する関連のコンポジット信号から合成することができる。しかし、上述のように、多くの場合、コンテンツプロバイダは、元の5.1には足りないコンテンツを持つ。この問題に対応するひとつの方法は、最初に、空間的コーディングの前にブラインドアップミキシングシステムを用いることにより、既存の2チャンネルステレオコンテンツを5.1コンテンツに変換することである。このようなブラインドアップミキシングシステムでは、5.1信号を合成するために、元の2チャンネルステレオ信号自身でのみ有用な情報を用いる。多くのこのようなアップミキシングシステムは、例えばドルビープロロジック（Dolby

Pro logic）IIのように、商業的に利用可能となっている。空間的コーディングシステムと結合するとき、コンポジット信号は、図1に示すようにブラインドアップミックスされた信号をダウンミキシングすることにより、又は図2のように、既存の2チャンネルステレオ信号を用いることにより、エンコーダにて、生成される。

【0073】

代替的に、引用したSeefeldt等による出願中の国際特許出願PCT/US2006/020882の記載された空間的エンコーダがブラインドアップミキサーの一部として用いられる。この修正したエンコーダでは、中間的なブラインドアップミックスされた信号を生成することなく2チャンネルステレオ信号から望ましい直接5.1の空間的イメージのパラメトリックモデルを合成するために、既存の空間的コーディングパラメータを利用する。図3は、上記に概説した修正したエンコーダを示している。

【0074】

結果として生じたエンコードされた信号は、その結果、既存の空間的デコーダと互換性を有することになる。このデコーダは、好ましいブラインドアップミックスを生成するためにサイド情報を用いることができ、あるいは、元の2チャンネルステレオ信号のリサナーであることを条件として、このサイド情報を無視してもよい。

【0075】

先に説明した空間的コーディングパラメータ（ICC，IPD，及びICC）は、以下の例に従い、2チャンネルステレオ信号の5.1ブラインドアップミックスを作り出すために用いることができる。この例では、左右のステレオペアから3つのサラウンドチャンネルのみを合成することを考えているが、この技術を、中央チャンネルとLFE（低周波数効果）チャンネルを合成するために拡張することもできる。この技術は、ステレオ信号の左右のチャンネルのスペクトル部分がレコーディングにおける環境に対応するようにデコリレートされ、サラウンドチャンネルに方向付けられるというアイデアに基づくものである。左右のチャンネルが相互に相関があるスペクトル部分は、直接的なサウンドに対応し、前方右チャンネルと前方左チャンネルのまま維持される。

【式21】

【0076】

10

20

30

40

1 番目のステップとして、元の 2 チャンネルステレオの各帯域についての 2×2 の共分散マトリックス $\mathbf{Q}[b, t]$ が計算される。このマトリックスの各要素は、先に説明した $\mathbf{R}[b, t]$ と同じような再帰的な方法で更新される。

【数 3 7】

$$Q_{ij}[b, t] = \lambda Q_{ij}[b, t-1] + \frac{1-\lambda}{k e_b - k b_b} \sum_{k=k b_b}^{k=k e_b-1} Y_i[k, t] Y_j^*[k, t] \quad (35) \quad 10$$

【式 2 2】

【0 0 7 7】

次に、 $\mathbf{Q}[b, t]$ から左右のチャンネル間の正規化された相関関係 ρ が計算される。

【数 3 8】

$$\rho[b, t] = \frac{|Q_{12}[b, t]|}{\sqrt{Q_{11}^2[b, t] Q_{22}^2[b, t]}} \quad (36) \quad 20$$

【0 0 7 8】

I L D パラメータを用いて左右のチャンネルは、 ρ に比例する量だけ左右のサラウンドチャンネルに向けられる。もし $\rho = 0$ ならば、左右のチャンネルは完全にサラウンドチャンネルに向けられる。もし $\rho = 1$ ならば、左右のチャンネルは全て前方に残る。加えて、サラウンドチャンネルの I C C パラメータは、より拡がった空間的イメージを作るためにこれらのチャンネルが完全なデコリレーションを受けるように、ゼロに設定される。この 5 . 1 ブラインドアップミックスを達成するために用いられる空間的パラメータの全セットは、以下のテーブルにリスト化される。

【0 0 7 9】

チャンネル 1 (左)

$$\begin{aligned} I L D_{11}[b, t] &= \rho[b, t] \\ I L D_{12}[b, t] &= 0 \\ I P D_{11}[b, t] &= I P D_{12}[b, t] = 0 \\ I C C_1[b, t] &= 1 \end{aligned}$$

チャンネル 2 (中央)

$$\begin{aligned} I L D_{21}[b, t] &= I L D_{22} = I P D_{21}[b, t] = I P D_{22}[b, t] = 0 \\ I C C_2[b, t] &= 1 \end{aligned} \quad 40$$

チャンネル 3 (右)

$$\begin{aligned} I L D_{31}[b, t] &= 0 \\ I L D_{32}[b, t] &= -\rho[b, t] \\ I P D_{31}[b, t] &= I P D_{32}[b, t] = 0 \\ I C C_3[b, t] &= 1 \end{aligned}$$

チャンネル 4 (左サラウンド)

$$I L D_{41}[b, t] = (1 - \rho^2[b, t]) \quad 50$$

$$\begin{aligned} I L D_{42} [b, t] &= 0 \\ I P D_{41} [b, t] &= I P D_{42} [b, t] = 0 \\ I C C_4 [b, t] &= 1 \end{aligned}$$

チャンネル 5 (右サラウンド)

$$\begin{aligned} I L D_{51} [b, t] &= 0 \\ I L D_{52} [b, t] &= (1 - \quad^2 [b, t]) \\ I P D_{51} [b, t] &= I P D_{52} [b, t] = 0 \\ I C C_5 [b, t] &= 0 \end{aligned}$$

10

チャンネル 6 (L F E)

$$\begin{aligned} I L D_{61} [b, t] &= I L D_{62} = I P D_{61} [b, t] = I P D_{62} [b, t] = 0 \\ I C C_6 [b, t] &= 1 \end{aligned}$$

上記の単純なシステムで非常に説得力のあるサラウンド効果を合成するが、同じ空間的パラメータを用いたより洗練されたブラインドアップミキシング技術が可能である。特定のアップミキシング技術を用いることは本発明にとって重要ではない。

【0080】

空間的エンコーダ及びデコーダと共に動作させるのではなく、代わりに、記載したブラインドアップミキシングシステムをシングルエンド方式で動作させることができる。つまり、2チャンネルステレオ信号のようなマルチチャンネルステレオ信号から直接アップミックスした信号を合成すると同時に、空間的パラメータを導き出すことができる。このような構成は、旧来の膨大な量の2チャンネルステレオコンテンツを例えばコンパクトディスクから演奏する、オーディオ/ビデオ受信器のような、消費者の持つ装置で役に立つ。消費者は、このようなコンテンツを、再生時に直接マルチチャンネル信号に変換することを望むであろう。図5は、このようなシングルエンドモードにおけるブラインドアップミキサーの例を示す。

20

【0081】

図5のブラインドアップミキサーにおいて、Mチャンネルの元の信号(例えば、PCMフォーマットのマルチチャンネルデジタルオーディオ)は、先に例示したエンコーダの例のようによく知られた短時間離散フーリエ変換(STDFT)のような、適切な時間・周波数変換を用いて周波数領域に、(時間から周波数へ)の変換装置又は機能2により変換され、1以上の周波数ビンが耳の臨界帯域を近似する帯域にグループ化される。空間的パラメータ形式のアップミックス情報は、各帯域において(アップミックス情報の導出)機能の装置4”(この装置又は機能は、図3の空間的サイド情報の導出に対応する)により、計算される。上述の通り、聴覚シーン分析装置又は分析機能(聴覚シーン分析)6”もまた、Mチャンネルの元の信号を受信し、本明細書の別の所に記載したように、装置又は機能4”によりアップミックス情報の生成に関与する。説明を容易にするために、装置又は機能4”と6”とを別に表示したが、単一の装置又は機能としてもよい。装置又は機能4”からのアップミックス情報は、周波数領域におけるNチャンネルアップミックス信号を生成するために、(アップミックス情報の適用)装置又は機能26により、周波数領域におけるMチャンネルの元の信号の対応する帯域に適用される。少ない数のチャンネルからより多くのチャンネルを生成することがアップミキシングである(装置又は機能26は、アップミキサーとして特徴づけることができる)。最後に、周波数・時間変換(周波数から時間へ)28(時間・周波数変換装置又は機能2の逆機能)が、ブラインドアップミックスを構成するNチャンネルアップミックス信号を生成するために適用される。図5の例におけるアップミックス情報は、空間的パラメータの形式を取るが、聴覚事象及び/又は聴覚事象の境界と結びつく信号の特性の程度に少なくとも部分的に応答してオーディオ出力チャンネルを生成するスタンドアロンのアップミキサー装置又は機能におけるようなアップミックス情報は、空間的パラメータの形式を取る必要がない。

30

40

50

【式 23】

【0082】

[聴覚事象のパラメータ制御]

上記のように、N:M:N空間的コーディング及びブラインドアップミキシングの両方におけるILDパラメータ、IPDパラメータ、及びICCパラメータは、プレ帯域共分散マトリックスの時間的に変化する推定値、すなわち、N:M:N空間的コーディングの場合は、 $R[b, t]$ 、2チャンネルステレオブラインドアップミキシングの場合は、 $Q[b, t]$ 、に依存する。

10

コーダパラメータが好ましい空間的イメージの時間により変化する特徴を捉えるためには十分速く変化し、しかし、合成された空間的イメージに可聴な不安定性をもたらすほどには速く変化しないよう、対応する式(4)と(36)から得られる関連の平滑パラメータの選択には十分注意を払わなくてはならない。特に厄介なのは、 $M=1$ でありICCパラメータが $M=1$ と $M=2$ の両方となるシステムにおける、N:M:NシステムのIPDに関係付けられた優勢な参照チャンネル g の選定である。共分散推定値が時間ブロックを横切って顕著に平滑化されていたとしても、いくつかのチャンネルが同じ量のエネルギーを有する場合は、ブロックとブロックとの間で急速に変動することがある。IPDパラメータとICCパラメータとが急速に変化することにより、合成された信号中に可聴のアーティファクトが生じるという結果となる。

20

【0083】

この問題を解決する1つの方法は、優勢なチャンネル g を聴覚事象の境界でのみ更新することである。そうすることにより、このコーディングパラメータが各事象の継続期間中比較的安定し、各事象における知覚的完全性が保たれる。このオーディオのスペクトル形状の変化は、聴覚事象の境界を検出するために用いられる。このエンコーダの各時間ブロック t において、各チャンネル i の聴覚事象の境界の強さは、正規化した現在のブロックとその前のブロックとの対数スペクトルの大きさの差の絶対値の総和として計算される。

【数39】

30

$$S_i[t] = \sum_k |P_i[k, t] - P_i[k, t-1]|,$$

(37a)

【0084】

ここで

【数40】

$$P_i[k, t] = \log \left(\frac{|X_i[k, t]|}{\max_k \{|X_i[k, t]|\}} \right),$$

40

(37b)

【0085】

任意のチャンネル i において、事象の強さ $S_i[t]$ が固定の閾値 T_s より大きい場合は、優勢なチャンネル g は式(9)により更新される。それ以外の場合は、優勢なチャンネルは、その前の時間ブロックにおける値を保持する。

【0086】

ここに記載の技術は、聴覚事象に基づく「厳しい決断」の一例である。事象は検出されるかされないかのどちらか一方であり、優勢なチャンネルの更新は、この2値的な検出に

50

基づき行われる。聴覚事象はまた、「柔軟な決断」方法にも用いられる。

【式 2 4】

【 0 0 8 7 】

例えば、共分散パラメータ $R[b, t]$ 又は $Q[b, t]$ のどちらか一方を平滑化するために用いられるパラメータ λ を連続的に変化させるために、事象の強さ $S_i[t]$ を用いてもよい。

$S_i[t]$ が大きい場合は、強い事象が生じ、この強い事象に関連するオーディオの新しい統計値を速く捉えるために、マトリックスは少し平滑化して更新すべきである。 $S_i[t]$ が小さい場合は、オーディオは事象の範囲内にあり、比較的安定しているので、共分散マトリックスは、より強く平滑化されるべきである。最小（最小の平滑化）と最大（最大の平滑化）との間で λ を計算するための 1 つの方法は、この原理に基づく。

【数 4 1】

$$\lambda = \begin{cases} \lambda_{\min}, & S_i[t] > T_{\max} \\ \frac{S_i[t] - T_{\min}}{T_{\max} - T_{\min}} (\lambda_{\min} - \lambda_{\max}) + \lambda_{\max}, & T_{\max} \geq S_i[t] \geq T_{\min} \\ \lambda_{\max}, & S_i[t] < T_{\min} \end{cases}$$

(38)

【 0 0 8 8 】

〔実施の形態〕

本発明は、ハードウェア又はソフトウェア又は両方を組み合わせたもの（例えば、プログラマブルロジックアレー）で実施することができる。他に記載がない限り、本発明の 1 部に含まれるアルゴリズム又はプロセスは、特定のコンピュータ又は特定の装置に本質的に関連するようなものではない。とりわけ、種々の汎用機をここの記載に従って書かれたプログラムと共に用いてもよい、あるいは、要求の方法を実行するために、より特化した装置（例えば、集積回路）を構成することが便利かもしれない。このように、本発明は、それぞれ少なくとも 1 つのプロセッサ、少なくとも 1 つの記憶システム（揮発性及び非揮発性メモリ及び/又は記憶素子を含む）、少なくとも 1 つの入力装置又は入力ポート、及び少なくとも 1 つの出力装置又は出力ポートを具備する、1 つ以上のプログラマブルコンピュータシステム上で実行される 1 つ以上のコンピュータプログラムにより実現することができる。ここに記載した機能を遂行し、出力情報を出力させるために入力データにプログラムコードを適用する。この出力情報は、公知の方法で、1 以上の出力装置に適用される。

【 0 0 8 9 】

このようなプログラムの各々は、コンピュータシステムとの通信のために、必要とされるどんなコンピュータ言語（機械語、アセンブリ、又は、高級な、手続言語、論理型言語、又は、オブジェクト指向言語を含む）でも実現することができる。いずれにせよ、言語はコンパイル言語であってもインタープリタ言語であってもよい。

【 0 0 9 0 】

このようなコンピュータプログラムの各々は、ここに記載の手順を実行するために、コンピュータにより記憶媒体又は記憶装置を読み込んだとき、コンピュータを設定し動作させるための、汎用プログラマブルコンピュータ又は専用プログラマブルコンピュータにより、読み込み可能な記憶媒体又は記憶装置（例えば、半導体メモリ又は半導体媒体、又は磁気又は光学媒体）に保存又はダウンロードすることが好ましい。本発明のシステムはまた、コンピュータプログラムにより構成されるコンピュータにより読み込み可能な記憶媒体として実行することを考えることもできる。ここで、この記憶媒体は、コンピュータ

10

20

30

40

50

システムを、ここに記載した機能を実行するために、具体的にあらかじめ定めた方法で動作させる。

【 0 0 9 1 】

本発明の多くの実施の形態について記載した。しかしながら、本発明の精神と技術範囲を逸脱することなく多くの修正を加えることができることは明らかであろう。例えば、ここに記載したステップのいくつかの順序は独立であり、従って、記載とは異なる順序で実行することができる。

【 0 0 9 2 】

[参照としての編入]

以下の特許、特許出願、及び、刊行物は参照としてそのすべてを本明細書に編入する。

10

【 0 0 9 3 】

[空間的コーディング及びパラメトリックコーディング]

国際公開公報 WO 2 0 0 5 / 0 8 6 1 3 9 A 1、2 0 0 5 年 9 月 1 5 日公開、
 国際公開公報 WO 2 0 0 5 / 0 2 6 4 5 2、2 0 0 6 年 3 月 9 日公開、
 Seefeldt 等による国際特許出願 P C T / U S 2 0 0 6 / 0 2 0 8 8 2、2 0 0 6 年 5 月
 2 6 日出願、表題「Channel Reconfiguration with Side Information」、
 米国特許出願公開公報 U S 2 0 0 3 / 0 0 2 6 4 4 1、2 0 0 3 年 2 月 6 日公開、
 米国特許出願公開公報 U S 2 0 0 3 / 0 0 3 5 5 3、2 0 0 3 年 2 月 2 0 日公開、
 米国特許出願公開公報 U S 2 0 0 3 / 0 2 1 9 1 3 0 (Baumgarte & Faller)、2 0
 0 3 年 1 1 月 2 7 日公開、

20

Audio Engineering Society 論文 5 8 5 2、2 0 0 3 年 3 月、
 国際公開公報 WO 0 3 / 0 9 0 2 0 7、2 0 0 3 年 1 0 月 3 0 日公開、
 国際公開公報 WO 0 3 / 0 9 0 2 0 8、2 0 0 3 年 1 0 月 3 0 日公開、
 国際公開公報 WO 0 3 / 0 0 7 6 5 6、2 0 0 3 年 1 月 2 2 日公開、
 国際公開公報 WO 0 3 / 0 9 0 2 0 6、2 0 0 3 年 1 0 月 3 0 日公開、
 Baumgarte 他による、2 0 0 3 年 1 2 月 2 5 日に公開された、米国特許出願公開公報 U
 S 2 0 0 3 / 0 2 3 6 5 8 3 A 1、

Faller 他による、Audio Engineering Society Convention Paper 5574, 112th
 Convention, Munich, May 2002、「Binaural Cue Coding Applied to Stereo a
 nd Multi-Channel Audio Compression」、

30

Baumgarte 他による、Audio Engineering Society Convention Paper 5575, 112th
 Convention, Munich, May 2002「Why Binaural Cue Coding is Better than
 Intensity Stereo Coding」、

Baumgarte 他による、Audio Engineering Society Convention Paper 5706, 113th
 Convention, Los Angeles, October 2002、「Design and Evaluatin of Bina
 ural Cue Coding Schemes」、

Faller 他による、IEEE Workshop on Applications of Signal Processing to
 Audio and Acoustics 2001, New Paltz, New
 York, October 2001, pp.199-202、「Efficient Representation of Spatial Aud
 io Using Perceptual Parametrization」、

40

Baumgarte 他による、Proc. ICASSP 2002, Orlando, Florida, May 2002, pp.11
 -1801-1804、「Estimation of Auditory Spatial Cues for Binaural Cue Codin
 g」、

Faller 他による、Proc. ICASSP 2002, Orlando, Florida, May 2002, pp.11-18
 4111-1844、「Binaural Cue Coding: A Novel and Efficient Representation o
 f Spatial Audio」、

Breebaart 他による、Audio Engineering Society Convention Paper 6072, 116th
 Convention, Berlin, May 2004、「High-quality parametric spatial audio
 coding at low bitrates」、

Baumgarte 他による、Audio Engineering Society Convention Paper 6060, 116th

50

h Convention, Berlin, May

2004、「Audio Coder Enhancement using Scalable Binaural Cue Coding with Equalized Mixing」、

Schuijers他による、Audio Engineering Society Convention Paper 6073, 116th Convention, Berlin, May

2004、「Low complexity parametric stereo coding」、

Engdegard他による、Audio Engineering Society Convention Paper 6074, 116th Convention, Berlin, May

2004、「Synthetic Ambience in Parametric Stereo Coding」。

【 0 0 9 4 】

10

[聴覚事象の検出と使用]

米国特許出願公開公報 U S 2 0 0 4 / 0 1 2 2 6 6 2 A 1、2 0 0 4 年 6 月 2 4 日 公 開

、

米国特許出願公開公報 U S 2 0 0 4 / 0 1 4 8 1 5 9 A 1、2 0 0 4 年 7 月 2 9 日 公 開

、

米国特許出願公開公報 U S 2 0 0 4 / 0 1 6 5 7 3 0 A 1、2 0 0 4 年 8 月 2 6 日 公 開

、

米国特許出願公開公報 U S 2 0 0 4 / 0 1 7 2 2 4 0 A 1、2 0 0 4 年 9 月 2 日 公 開、

米国特許出願公開公報 U S 2 0 0 6 / 0 1 9 7 1 9、2 0 0 6 年 2 月 2 3 日 公 開、

Brett Crockett 及び Michael Smithers による、Audio Engineering Society Convention Paper 6416, 118th Convention, Barcelona, May

20

28-31, 2005、「A Method for Characterizing and Identifying Audio Based on Auditory Scene Analysis」、

Brett Crockett による、Audio Engineering Society Convention Paper 5948, New York, October 2003、「High Quality Multichannel Time Scaling and Pitch-Shifting using Auditory Scene Analysis」。

【 0 0 9 5 】

[デコリレーション]

Breebaart による、2 0 0 3 年 1 0 月 3 0 日 に 公 開 さ れ た、国 際 公 開 公 報 W O 0 3 / 0 9 0 2 0 6 A 1、表 題 「Signal Synthesizing」、

30

2 0 0 6 年 3 月 9 日 公 開 さ れ た、国 際 公 開 公 報 W O 2 0 0 6 / 0 2 6 1 6 1、

2 0 0 6 年 3 月 9 日 公 開 さ れ た、国 際 公 開 公 報 W O 2 0 0 6 / 0 2 6 4 5 2。

【 0 0 9 6 】

[M P E G 2 / 4、A A C]

I S O / I E C J T C 1 / S C 2 9、「Information technology - very low bitrate audio-visual coding」、I S O / I E C I S 1 4 4 9 6 (P a r t 3、オーディオ)、1 9 9 6、

1) I S O / I E C 1 3 8 1 8 7 「MPEG-2 advanced audio coding, AAC」International Standard, 1997、

M.Bosi, K.Brandenburg, S.Quackenbush, L.Fielder, K.Akagiri, H.Fuchs, M.Dietz, J.Herre, G.Davidson, 及び Y.Oikawa による、Proc. of the 101st AES-Convention, 1996、「ISO/IEC MPEG-2 Advanced Audio Coding」、

40

M.Bosi, K.Brandenburg, S.Quackenbush, L.Fielder, K.Akagiri, H.Fuchs, M.Dietz, J.Herre, G.Davidson, 及び Y.Oikawa による、Journal of the AES, Vol.45, No.10, October 1997, pp.789-814、「ISO/IEC MPEG-2 Advanced Audio Coding」、

Karlheinz Brandenburg による、Proc. of the AES 17th International Conference on High Quality Audio Coding, Florence, Italy, 1999、「MP3 and AAC explained」、

G.A. Soulodre 等 による、J. Audio Eng. Soc., Vol.46, No.3, pp164-177, Mar

50

ch 1998、「Subjective Evaluation of State-of-the-Art Two-Channel Audio Co
decs」

【図面の簡単な説明】

【0097】

【図1】空間的コーディングシステムにおいてデコーダで再生したいNチャンネル信号をエンコーダが受け取る、空間的コーディングシステムにおけるエンコーダの一例を示す機能ブロック図である。

【図2】空間的コーディングシステムにおいてデコーダで再生したいNチャンネル信号をエンコーダが受け取り、また、このエンコーダからデコーダに送られるMチャンネルコンボジット信号もまた該エンコーダが受け取る、空間的コーディングシステムにおけるエン
10 コーダの一例を示す機能ブロック図である。

【図3】空間的エンコーダがブラインドミキシング構成の一部をなす、空間的コーディングシステムにおけるエンコーダの一例を示す機能ブロック図である。

【図4】図1～3のいずれか1つのエンコーダを用いるのに適した、空間的コーディングシステムにおけるデコーダの一例を示す機能ブロック図である。

【図5】シングルエンドのブラインドミキシング構成の機能ブロック図である。

【図6】本発明の特徴を実行する空間的エンコーディングシステムの有用なSTDFT分析及び合成窓の一例を示す。

【図7】時間に対する信号の時間領域での振幅のプロットである。最初の2つのプロットは、DFT処理ブロック内で想定した2チャンネル信号を示す。3番目のプロットは、2
20 つのチャンネル信号を1つのチャンネルに合成するダウンミキシングの効果を示し、4番目のプロットは、2番目のチャンネルにSWF処理を用いてアップミックスした信号を示す。

【図1】

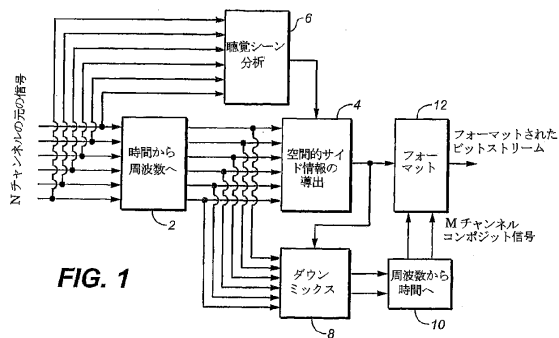


FIG. 1

【図2】

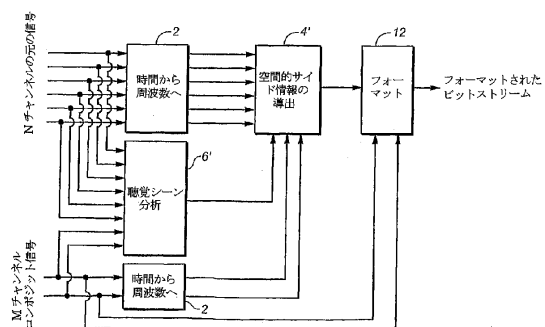


FIG. 2

【図3】

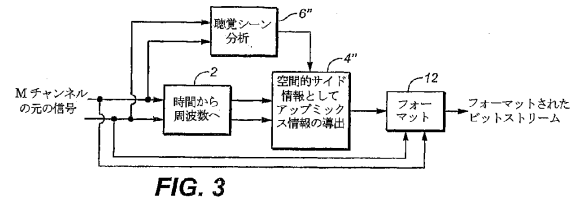


FIG. 3

【図4】

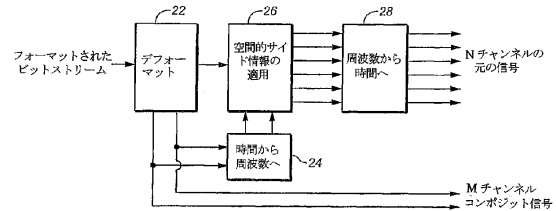


FIG. 4

【図5】

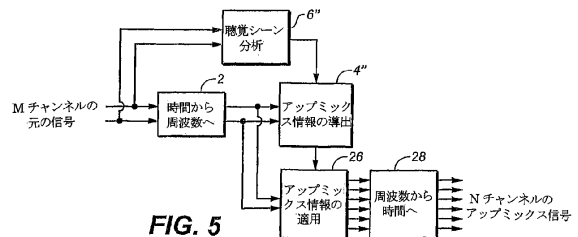


FIG. 5

【図 6】

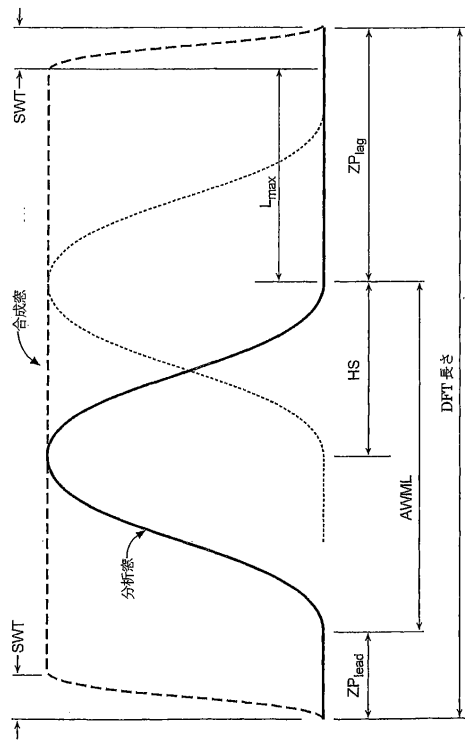


FIG. 6

【図 7】

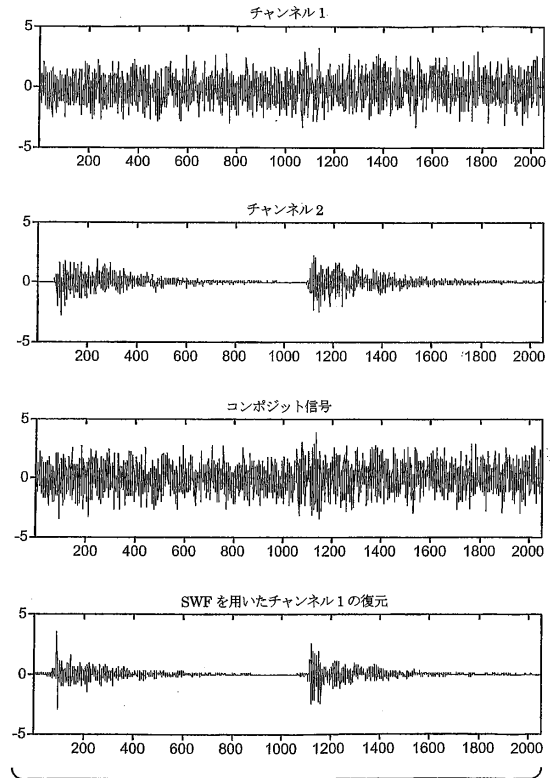


FIG. 7

フロントページの続き

- (74)代理人 100121762
弁理士 杉山 直人
- (74)代理人 100126767
弁理士 白銀 博
- (74)代理人 100118647
弁理士 赤松 利昭
- (74)代理人 100138519
弁理士 奥谷 雅子
- (74)代理人 100120145
弁理士 田坂 一郎
- (74)代理人 100138438
弁理士 尾首 亘聰
- (74)代理人 100147740
弁理士 保坂 俊
- (74)代理人 100123892
弁理士 内藤 忠雄
- (72)発明者 シーフェルト、アラン・ジェフリー
アメリカ合衆国、カリフォルニア州 94103、サンフランシスコ、ポトレロ・アベニュー 1
00
- (72)発明者 ピントン、マーク・ステュアート
アメリカ合衆国、カリフォルニア州 94103、サンフランシスコ、ポトレロ・アベニュー 1
00

審査官 山下 剛史

- (56)参考文献 特表2001-510953(JP,A)
特表2004-527000(JP,A)
特表2004-528601(JP,A)
特表2001-521259(JP,A)

- (58)調査した分野(Int.Cl., DB名)
G10L 19/00-19/26