

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6567571号
(P6567571)

(45) 発行日 令和1年8月28日(2019.8.28)

(24) 登録日 令和1年8月9日(2019.8.9)

(51) Int.Cl. F I
G 1 0 L 19/008 (2013.01) G 1 0 L 19/008 1 0 0
G 1 0 L 19/00 (2013.01) G 1 0 L 19/00 3 3 0 B

請求項の数 12 (全 32 頁)

(21) 出願番号	特願2016-575016 (P2016-575016)	(73) 特許権者	510185767
(86) (22) 出願日	平成27年6月22日 (2015. 6. 22)		ドルビー・インターナショナル・アーベー
(65) 公表番号	特表2017-523456 (P2017-523456A)		オランダ王国, セーエン アムステルダム
(43) 公表日	平成29年8月17日 (2017. 8. 17)		ズイドースト 1101, ヘリケルベル
(86) 国際出願番号	PCT/EP2015/063912		グウェグ 1-35, アポロ ビルディング
(87) 国際公開番号	W02015/197512		グ 3エー
(87) 国際公開日	平成27年12月30日 (2015. 12. 30)	(74) 代理人	100107766
審査請求日	平成30年6月15日 (2018. 6. 15)		弁理士 伊東 忠重
(31) 優先権主張番号	14306023.4	(74) 代理人	100070150
(32) 優先日	平成26年6月27日 (2014. 6. 27)		弁理士 伊東 忠彦
(33) 優先権主張国・地域又は機関	欧州特許庁 (EP)	(74) 代理人	100091214
			弁理士 大貫 進介

最終頁に続く

(54) 【発明の名称】 非差分的な利得値を表現するのに必要とされる最低整数ビット数をHOAデータ・フレーム表現の圧縮のために決定する方法および装置

(57) 【特許請求の範囲】

【請求項1】

HOAデータ・フレーム表現 (C(k)) の圧縮のために、前記HOAデータ・フレームのチャンネル信号について振幅変化に対応する非差分的な利得値の表現を2の指数 (2[°]) として記述するための最低の整数ビット数 e_c を決定する方法であって、各フレームにおける各チャンネル信号はサンプル値のグループを含み、前記HOAデータ・フレームの各フレームの各チャンネル信号 (y₁(k-2), ..., y₁(k-2)) に対して差分利得値が割り当てられ、該差分利得値は現在HOAデータ・フレーム ((k-2)) におけるチャンネル信号の第一のサンプル値の振幅の、直前のHOAデータ・フレーム ((k-3)) におけるチャンネル信号の第二のサンプル値に対する変化を引き起こすものであり、結果として得られる利得適応されたチャンネル信号はエンコーダ (16) においてエンコードされ、

前記HOAデータ・フレーム表現は空間領域において0個の仮想スピーカー信号 w_j(t) にレンダリングされており、それらの仮想スピーカーの位置は単位球上にあり、その単位球上で一様に分布させられるよう目標とされており、前記レンダリングは行列乗算 w(t) = ()⁻¹ · c(t) によって表現され、w(t) はすべての仮想スピーカー信号を含むベクトルであり、

は仮想スピーカー位置モード行列であり、c(t) は前記HOAデータ・フレーム表現の対応するHOA係数シーケンスのベクトルであり、前記HOAデータ・フレーム表現 (C(k)) は

【数 6 7】

$$\|\mathbf{w}(t)\|_{\infty} = \max_{1 \leq j \leq O} |w_j(t)| \leq 1 \quad \forall t$$

となるよう正規化されており、当該方法は：

・チャンネル信号を、

a) 前記チャンネル信号における優勢音信号 ($x(t)$) を表現するために、HOA係数シーケンスのベクトル $c(t)$ に混合行列 A を乗算するサブステップであって、混合行列 A は正規化されたHOAデータ・フレーム表現の係数シーケンスの線形結合を表わす、サブステップ；

b) 前記チャンネル信号における周囲HOA成分 $c_{AMB}(t)$ を表現するために、前記正規化されたHOAデータ・フレーム表現から前記優勢音信号を減算し、結果として得られる最小周囲HOA成分 $c_{AMB,MIN}(t)$ を、 $w_{MIN}(t) = \mathbf{MIN}^{-1} \cdot c_{AMB,MIN}(t)$ を計算することによって変換し、 $\|\mathbf{MIN}^{-1}\|_2 < 1$ であり、 \mathbf{MIN} は前記最小周囲HOA成分 $c_{AMB,MIN}(t)$ についてのモード行列である、サブステップ；

c) 前記HOA係数シーケンス $c(t)$ のうち、空間変換が適用される前記周囲HOA成分の係数シーケンスに関係する一部を選択する、サブステップ；

を実行することによって形成する段階と；

・前記整数ビット数 e を

【数 6 8】

$$\beta_e = \lceil \log_2(\lceil \log_2(\sqrt{K_{MAX}} \cdot O) \rceil + 1) \rceil$$

に基づいて決定する段階であって、

【数 6 9】

$$K_{MAX} = \max_{1 \leq N \leq N_{MAX}} K(N, \boldsymbol{\Omega}_1^{(N)}, \dots, \boldsymbol{\Omega}_O^{(N)})$$

であり、 N はHOA次数であり、 N_{MAX} は関心対象の最大次数であり、 $\boldsymbol{\Omega}_1^{(N)}, \dots, \boldsymbol{\Omega}_O^{(N)}$ は前記仮想スピーカ方向であり、 $O = (N + 1)^2$ はHOA係数シーケンスの数であり、 K は前記モード行列の二乗されたユークリッド・ノルム $\|\cdot\|_2^2$ と O との間の比である、段階とを含む

方法。

【請求項 2】

HOAデータ・フレーム表現 ($C(k)$) の圧縮のために、前記HOAデータ・フレームのチャンネル信号について振幅変化に対応する非差分的な利得値の表現を2の指数 (2^e) として記述するための最低の整数ビット数 e を決定する装置であって、各フレームにおける各チャンネル信号はサンプル値のグループを含み、前記HOAデータ・フレームの各フレームの各チャンネル信号 ($y_1(k-2), \dots, y_1(k-2)$) に対して差分利得値が割り当てられ、該差分利得値は現在HOAデータ・フレーム ($(k-2)$) におけるチャンネル信号の第一のサンプル値の振幅の、直前のHOAデータ・フレーム ($(k-3)$) におけるチャンネル信号の第二のサンプル値に対する変化を引き起こすものであり、結果として得られる利得適応されたチャンネル信号はエンコーダ (16) においてエンコードされ、

前記HOAデータ・フレーム表現 ($C(k)$) は空間領域において O 個の仮想スピーカ信号 $w_j(t)$ にレンダリングされており、それらの仮想スピーカの位置は単位球上にあり、その単位球上で一様に分布させられるよう目標とされており、前記レンダリングは行列乗算 $w(t) = (\cdot)^{-1} \cdot c(t)$ によって表現され、 $w(t)$ はすべての仮想スピーカ信号を含むベクトルであり、 $(\cdot)^{-1}$ は仮想スピーカ位置モード行列であり、 $c(t)$ は前記HOAデータ・フレーム表現の対応するHOA係数シーケンスのベクトルであり、前記HOAデータ・フレーム表現 ($C(k)$) は

【数 7 0】

$$\|\mathbf{w}(t)\|_{\infty} = \max_{1 \leq j \leq O} |w_j(t)| \leq 1 \quad \forall t$$

となるよう正規化されており、当該装置は：

・前記チャンネル信号 $(y_1(k-2), \dots, y_I(k-2))$ を、

a) 前記チャンネル信号における優勢音信号 $(x(t))$ を表現するために、HOA係数シーケンスの前記ベクトル $c(t)$ に混合行列 A を乗算する動作であって、混合行列 A は正規化されたHOAデータ・フレーム表現の係数シーケンスの線形結合を表わす、動作；

b) 前記チャンネル信号における周囲HOA成分 $c_{AMB}(t)$ を表現するために、前記正規化されたHOAデータ・フレーム表現から前記優勢音信号を減算し、結果として得られる最小周囲HOA成分 $c_{AMB,MIN}(t)$ を、 $w_{MIN}(t) = \mathbf{MIN}^{-1} \cdot c_{AMB,MIN}(t)$ を計算することによって変換し、 $\|\mathbf{MIN}^{-1}\|_2 < 1$ であり、 \mathbf{MIN} は前記最小周囲HOA成分 $c_{AMB,MIN}(t)$ についてのモード行列である、動作；

c) 前記HOA係数シーケンス $c(t)$ のうち、空間変換が適用される前記周囲HOA成分の係数シーケンスに関係する一部を選択する、動作；

を実行することによって形成する手段 (12、13、14) と；

・前記整数ビット数 e_e を

【数 7 1】

$$\beta_e = \lceil \log_2(\lceil \log_2(\sqrt{K_{MAX}} \cdot O) \rceil + 1) \rceil$$

に基づいて決定する手段 (15、...、151) であって、

【数 7 2】

$$K_{MAX} = \max_{1 \leq N \leq N_{MAX}} K(N, \boldsymbol{\Omega}_1^{(N)}, \dots, \boldsymbol{\Omega}_O^{(N)})$$

であり、 N はHOA次数であり、 N_{MAX} は関心対象の最大次数であり、 $\boldsymbol{\Omega}_1^{(N)}, \dots, \boldsymbol{\Omega}_O^{(N)}$ は前記仮想スピーカ方向であり、 $O = (N+1)^2$ はHOA係数シーケンスの数であり、 K は前記モード行列の二乗されたユークリッド・ノルム $\|\cdot\|_2^2$ と0との間の比である、手段とを含む

装置。

【請求項 3】

前記変換された最小周囲HOA成分に加えて、前記周囲HOA成分 $c_{AMB}(t)$ の変換されていない周囲係数シーケンスが前記チャンネル信号 $(y_1(k-2), \dots, y_I(k-2))$ に含まれる、請求項1記載の方法。

【請求項 4】

前記HOAデータ・フレームのうちの個々のものの前記チャンネル信号に関連付けられた前記非差分的な利得値の表現 (2°) がサイド情報として転送され、そのそれぞれが e_e ビットによって表現される、請求項1もしくは3記載の方法。

【請求項 5】

前記整数ビット数 e_e が

【数 7 3】

$$\beta_e = \lceil \log_2(\lceil \log_2(\sqrt{K_{MAX}} \cdot O) \rceil + e_{MAX} + 1) \rceil$$

に設定され、 $e_{MAX} > 0$ は利得制御 (15、151) 前のチャンネル信号のサンプル値の振幅が閾値より小さいとの判定に基づいて前記ビット数 e_e を増すはたらきをする、

請求項1、3および4のうちいずれか一項記載の方法。

【請求項 6】

$K_{MAX} = 1.5$ である、請求項1および3ないし5のうちいずれか一項記載の方法。

【請求項 7】

前記混合行列 A が、モノラル優勢音信号の方向分布を表わすすべてのベクトルから形成

10

20

30

40

50

されるモード行列のムーア・ペンローズの擬似逆行列を取ることによって、もとのHOA表現と優勢音信号のものとの間の残差のユークリッド・ノルムを最小にするよう決定される、請求項1および3ないし6のうちいずれか一項記載の方法。

【請求項8】

前記0個の仮想スピーカー信号の位置が θ_0 の計算のために想定された位置と一致しないとの判定に基づいて、

- ・該一致しない仮想スピーカー位置に基づいてモード行列 Ψ を計算し(51)；
- ・該モード行列のユークリッド・ノルム $\|\Psi\|_2$ を計算し(52)；
- ・前記正規化における最大の許容される振幅を置き換える最大許容される振幅値

【数74】

$$\gamma = \min\left(1, \frac{\sqrt{0} \cdot \sqrt{K_{MAX,DES}}}{\|\Psi\|_2}\right)$$

を計算する(53)ことを含み、

【数75】

$$K_{MAX,DES} = \max_{1 \leq N \leq N_{MAX,DES}} K(N, \Omega_{DES,1}^{(N)}, \dots, \Omega_{DES,0}^{(N)})$$

であり、Nは前記HOA次数であり、 $0 = (N+1)^2$ はHOA係数シーケンスの数であり、Kは、前記モード行列の二乗されたユークリッド・ノルムと0との比であり、 $N_{MAX,DES}$ は関心対象の次数であり $\Omega_{DES,1}^{(N)}, \dots, \Omega_{DES,1}^{(N)}$ は各次数について、前記HOAデータ・フレーム表現(C(k))の前記圧縮の実装のために想定された仮想スピーカーの方向であり、よって θ_0 は、前記非差分的な利得値の底2に対する指数(e)を符号化するために

【数76】

$$\beta_e = \lceil \log_2(\lceil \log_2(\sqrt{K_{MAX,DES}} \cdot 0) \rceil + 1) \rceil$$

によって選ばれたものである、

請求項1および3ないし7のうちいずれか一項記載の方法。

【請求項9】

コンピュータに請求項1および3ないし8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。

【請求項10】

音または音場の圧縮された高次アンビソニックス(HOA)音表現をデコードする方法であって：

前記圧縮されたHOA表現を含むビットストリームを受領する段階であって、前記ビットストリームは前記圧縮されたHOA表現に対応するいくつかのHOA係数を含む、段階と；

最低の整数数 β_e に基づいて前記圧縮されたHOA表現をデコードする段階であって、前記最低の整数数 β_e は

【数78】

$$\beta_e = \lceil \log_2(\lceil \log_2(\sqrt{K_{MAX}} \cdot 0) \rceil + 1) \rceil$$

に基づいて決定され、

【数79】

$$K_{MAX} = \max_{1 \leq N \leq N_{MAX}} K(N, \Omega_1^{(N)}, \dots, \Omega_0^{(N)})$$

であり、NはHOA次数であり、 N_{MAX} は関心対象の最大次数であり、 $\Omega_1^{(N)}, \dots, \Omega_0^{(N)}$ は前記仮想スピーカーの方向であり、 $0 = (N+1)^2$ はHOA係数シーケンスの数であり、Kは前記モード行列の二乗されたユークリッド・ノルム $\|\Psi\|_2^2$ と0との間の比である、段階とを含む、

方法。

10

20

30

40

50

【請求項 1 1】

音または音場の圧縮された高次アンビソニックス（HOA）音表現をデコードする装置であって：

前記圧縮されたHOA表現を含むビットストリームを受領する手段であって、前記ビットストリームは前記圧縮されたHOA表現に対応するいくつかのHOA係数を含む、手段と；

最低の整数数 β_e に基づいて前記圧縮されたHOA表現をデコードする手段であって、前記最低の整数数 β_e は

【数 8 0】

$$\beta_e = \lceil \log_2(\lceil \log_2(\sqrt{K_{\text{MAX}}} \cdot O) \rceil + 1) \rceil$$

10

に基づいて決定され、

【数 8 1】

$$K_{\text{MAX}} = \max_{1 \leq N \leq N_{\text{MAX}}} K(N, \boldsymbol{\Omega}_1^{(N)}, \dots, \boldsymbol{\Omega}_O^{(N)})$$

であり、NはHOA次数であり、 N_{MAX} は関心対象の最大次数であり、 $\boldsymbol{\Omega}_1^{(N)}, \dots, \boldsymbol{\Omega}_O^{(N)}$ は前記仮想スピーカーの方向であり、 $O = (N + 1)^2$ はHOA係数シーケンスの数であり、Kは前記モード行列の二乗されたユークリッド・ノルム $\| \cdot \|_2^2$ と0との間の比である、手段とを有する、装置。

【請求項 1 2】

20

$K_{\text{MAX}} = 1.5$ である、請求項 1 0 記載の方法。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、HOAデータ・フレーム表現の圧縮のために、該HOAデータ・フレームのうちの特定のもののチャンネル信号に関連する非差分的な利得値を表現するのに必要とされる最低整数ビット数を決定する方法および装置に関する。

【背景技術】

【0 0 0 2】

HOAと記される高次アンビソニックス（Higher Order Ambisonics）は、三次元的な音を表現する一つの可能性を提供する。他の技法は波面合成（WFS: wave field synthesis）または22.2のようなチャンネル・ベースのアプローチである。チャンネル・ベースの方法とは対照的に、HOA表現は特定のスピーカー・セットアップとは独立であるという利点をもたらす。しかしながら、この柔軟性は、特定のスピーカー・セットアップでのHOA表現の再生のために必要とされるデコード・プロセスの代償を伴う。必要とされるスピーカーの数が通例非常に多いWFSアプローチに比べ、HOAは少数のスピーカーのみからなるセットアップにレンダリングされてもよい。HOAのさらなる利点は、同じ表現を、いかなる修正もなしでヘッドフォンへのバイノーラル・レンダリングのために用いることもできるということである。

30

【0 0 0 3】

40

HOAは、複素調和平面波振幅の空間密度の、打ち切りされた球面調和関数（SH）展開による表現に基づく。各展開係数は角周波数の関数であり、これは時間領域関数によって等価に表現できる。よって、一般性を失うことなく、完全なHOA音場表現は、実際に、0個の時間領域関数からなると想定できる。ここで、0は展開係数の数を表わす。これらの時間領域関数は、以下では、等価だが、HOA係数シーケンスまたはHOAチャンネルと称される。

【0 0 0 4】

HOA表現の空間分解能は、展開の最大次数Nの増大とともに改善する。残念ながら、展開係数の数Oは次数Nとともに二次で、特に $O = (N + 1)^2$ の形で増大する。たとえば、次数N = 4を使う典型的なHOA表現はO = 25個のHOA（展開）係数を必要とする。HOA表現の伝送のための全ビットレートは、所望される単一チャンネル・サンプリング・レート f_s およびサンプル

50

当たりのビット数 N_b を与えられて、 $0 \cdot f_s \cdot N_b$ によって決定される。次数 $N=4$ のHOA表現を、 $f_s=48\text{kHz}$ のサンプリング・レートで、サンプル当たり $N_b=16$ ビットを用いて伝送することは、 19.2Mbits/s のビットレートにつながる。これは、たとえばストリーミングのような多くの実際的な用途にとって非常に高い。このように、HOA表現の圧縮がきわめて望ましい。

【0005】

以前に、HOA音場表現の圧縮が特許文献1、2、3において提案されている。非特許文献1参照。これらの手法は、音場解析を実行し、与えられたHOA表現を方向性成分 (directional component) と残差周囲成分 (residual ambient component) に分解することで共通している。一方では、最終的な圧縮された表現は、いくつかの量子化された信号からなると想定され、該量子化された信号は、方向性およびベクトル・ベースの信号と周囲HOA成分 (ambient HOA component) の関連する係数シーケンスとの知覚的符号化から帰結する。他方では、最終的な圧縮された表現は、量子化された信号に関する追加的なサイド情報を含む。このサイド情報は、HOA表現の、その圧縮されたバージョンからの再構成のために必要である。

10

【0006】

知覚的エンコーダに渡される前に、これらの中間時間領域信号は値範囲 $[-1, 1]$ 内の最大振幅をもつことが要求される。これは、現在利用可能な知覚的エンコーダの実装から生じる要件である。HOA表現を圧縮するときこの要件を満たすために、利得制御処理ユニット (特許文献4および上記の非特許文献1を参照) が知覚的エンコーダより先に使用される。これは入力信号をなめらかに減衰させるまたは増幅する。結果として得られる信号修正は可逆であり、フレームごとに適用されると想定される。特に、相続くフレーム間での信号振幅の変化は2の冪乗であると想定される。HOA圧縮解除器においてこの信号修正を反転させることを容易にするために、対応する正規化サイド情報が全サイド情報に含められる。この正規化サイド情報は2を底とする指数からなることができ、それらの指数が二つの相続くフレーム間での相対的な振幅変化を記述する。これらの指数は上述した非特許文献1に従ってランレングス符号を使って符号化される。相続くフレームの間では、より大きな変化よりも軽微な振幅変化のほうが可能性が高いからである。

20

【先行技術文献】

【特許文献】

30

【0007】

【特許文献1】欧州特許出願公開第2665208号

【特許文献2】欧州特許出願公開第2743922号

【特許文献3】欧州特許出願公開第2800401号

【特許文献4】欧州特許出願公開第2824661号

【非特許文献】

【0008】

【非特許文献1】ISO/IEC JTC1/SC29/WG11, N14264, WD1-HOA Text of MPEG-H 3D Audio, 2014年1月

【非特許文献2】J. Fliege, U. Maier, "A two-stage approach for computing cubature formulae for the sphere", Technical report, Fachbereich Mathematik, University of Dortmund, 1999

40

【非特許文献3】E. G. Williams, "Fourier Acoustics", vol.93 of Applied Mathematical Sciences. Academic Press, 1999

【非特許文献4】B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution", J. Acoust. Soc. Am., 4(116):2149-2157, October 2004

【非特許文献5】J. Daniel, "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia", PhD thesis, Université Paris

50

6, 2001

【発明の概要】

【発明が解決しようとする課題】

【0009】

HOA圧縮解除においてもとの信号振幅を再構成するために、差分符号化された振幅変化を使うことが、たとえば単一のファイルが最初から最後までいかなる時間的なジャンプもなしに圧縮解除される場合に、実用可能である。しかしながら、ランダム・アクセスを容易にするために、独立したアクセス単位が、符号化された表現（これは典型的にはビットストリームである）において存在している必要がある。所望される位置（または少なくともその近傍）から、先行するフレームからの情報とは独立に、圧縮解除を始めることを許容するためである。そのような独立したアクセス単位は、最初のフレームから現在フレームまで利得制御処理ユニットによって引き起こされた合計の絶対的な振幅変化（すなわち、非差分的な利得値）を含む必要がある。二つの相続くフレームの間の振幅変化が2の冪乗であるとする、合計の絶対的な振幅変化も底2の指数によって記述することが十分である。この指数の効率的な符号化のために、利得制御処理ユニットの適用前に信号の潜在的な最大利得を知っておくことが本質的である。しかしながら、この知識は、圧縮されるべきHOA表現の値範囲に対する制約条件の指定に強く依存する。残念ながら、非特許文献1のMPEG-H 3Dオーディオ文書は入力HOA表現のためのフォーマットの記述を提供するのみであり、値範囲に対するいかなる制約条件も設定していない。

10

【0010】

本発明によって解決されるべき課題は、非差分的な利得値を表現するために必要とされる最低整数ビット数を提供することである。この課題は、請求項1に開示される方法によって解決される。この方法を利用する装置は請求項2において開示される。本発明の有利な追加的实施形態はそれぞれの従属請求項において開示される。

20

【課題を解決するための手段】

【0011】

本発明は、入力HOA表現の値範囲と、HOA圧縮器内の利得制御処理ユニットの適用前の信号の潜在的な最大利得との間の相互関係を確立する。その相互関係に基づいて、要求されるビットの量が 入力HOA表現の値範囲についての所与の指定について 最初のフレームから現在フレームまでに利得制御処理ユニットによって引き起こされた修正された信号の合計の絶対的な振幅変化（すなわち、非差分的な利得値）をアクセス単位内で記述するための、2を底とする指数の効率的な符号化のために、決定される。

30

【0012】

さらに、ひとたび指数の符号化のための要求されるビットの量の計算のための規則が固定されたら、本発明は、所与のHOA表現が、正しく圧縮されることができるよう、要求される値範囲制約条件を満たすかどうかを検証するための処理を使う。

【0013】

原理的には、本発明の方法は、HOAデータ・フレーム表現の圧縮のために、前記HOAデータ・フレームのうちの特定のもののチャンネル信号についての非差分的な利得値を表現するために必要とされる最低整数ビット数 を決定するために好適である。ここで、各フレームにおける各チャンネル信号はサンプル値のグループを含み、前記HOAデータ・フレームの各フレームの各チャンネル信号に対して差分利得値が割り当てられ、そのような差分利得値は現在HOAデータ・フレームにおけるチャンネル信号のサンプル値の振幅の、直前のHOAデータ・フレームにおけるそのチャンネル信号のサンプル値に対する変化を引き起こすものであり、そのような利得適応されたチャンネル信号はエンコーダにおいてエンコードされ、

40

前記HOAデータ・フレーム表現は空間領域において0個の仮想スピーカー信号 $w_j(t)$ にレンダリングされており、それらの仮想スピーカーの位置は単位球上にあり、その単位球上で一様に分布させられるよう目標とされており、前記レンダリングは行列乗算 $w(t) = ()^{-1} \cdot c(t)$ によって表現され、 $w(t)$ はすべての仮想スピーカー信号を含むベクトルであり、 $()$ は仮想スピーカー位置モード行列であり、 $c(t)$ は前記HOAデータ・フレーム表現の対応

50

するHOA係数シーケンスのベクトルであり、
前記HOAデータ・フレーム表現は
【数 1】

$$\|w(t)\|_{\infty} = \max_{1 \leq j \leq O} |w_j(t)| \leq 1 \quad \forall t$$

となるよう正規化されており、当該方法は：

・前記の正規化されたHOAデータ・フレーム表現から、サブステップ a)、 b)、 c)、
すなわち

a) 前記チャンネル信号における優勢音信号を表現するために、HOA係数シーケンスの前記ベクトル $c(t)$ に混合行列 A を乗算するサブステップであって、混合行列 A のユークリッド・ノルムは1より小さくなく、混合行列 A は前記正規化されたHOAデータ・フレーム表現の係数シーケンスの線形結合を表わす、サブステップ；

b) 前記チャンネル信号における周囲成分 $c_{AMB}(t)$ を表現するために、前記正規化されたHOAデータ・フレーム表現から前記優勢音信号を減算し、前記周囲成分 $c_{AMB}(t)$ の係数シーケンスの少なくとも一部を選択し、 $\|c_{AMB}(t)\|_2^2 / \|c(t)\|_2^2$ であり、結果として得られる最小周囲成分 $c_{AMB,MIN}(t)$ を、 $w_{MIN}(t) = w_{MIN}^{-1} \cdot c_{AMB,MIN}(t)$ を計算することによって変換し、 $\|w_{MIN}^{-1}\|_2 < 1$ であり、 w_{MIN} は前記最小周囲成分 $c_{AMB,MIN}(t)$ についてのモード行列である、サブステップ；

c) 前記HOA係数シーケンス $c(t)$ の一部を選択するサブステップであって、選択された係数シーケンスは、空間変換が適用される前記周囲HOA成分の係数シーケンスに関係し、前記選択された係数シーケンスの数を記述する最小次数 N_{MIN} は $N_{MIN} - 9$ である、サブステップ；

のうちの一つまたは複数によって前記チャンネル信号を形成する段階と；

・前記チャンネル信号についての前記非差分的な利得値を表現するために必要とされる前記最低整数ビット数 β_e を

【数 2】

$$\beta_e = \lceil \log_2(\lceil \log_2(\sqrt{K_{MAX}} \cdot O) \rceil + 1) \rceil$$

に設定する段階であって、

【数 3】

$$K_{MAX} = \max_{1 \leq N \leq N_{MAX}} K(N, \Omega_1^{(N)}, \dots, \Omega_O^{(N)})$$

であり、 N は前記次数であり、 N_{MAX} は関心対象の最大次数であり、 $\Omega_1^{(N)}, \dots, \Omega_O^{(N)}$ は前記仮想スピーカ方向であり、 $O = (N + 1)^2$ はHOA係数シーケンスの数であり、 K は前記モード行列の二乗されたユークリッド・ノルム $\| \cdot \|_2^2$ と0との間の比である、段階とを含む。

【0014】

原理的には、本発明の装置は、HOAデータ・フレーム表現の圧縮のために、前記HOAデータ・フレームのうちの特定のもののチャンネル信号についての非差分的な利得値を表現するために必要とされる最低整数ビット数 β_e を決定するために好適である。ここで、各フレームにおける各チャンネル信号はサンプル値のグループを含み、前記HOAデータ・フレームの各フレームの各チャンネル信号に対して差分利得値が割り当てられ、そのような差分利得値は現在HOAデータ・フレームにおけるチャンネル信号のサンプル値の振幅の、直前のHOAデータ・フレームにおけるそのチャンネル信号のサンプル値に対する変化を引き起こすもので

あり、そのような利得適応されたチャネル信号はエンコーダにおいてエンコードされ、前記HOAデータ・フレーム表現は空間領域において0個の仮想スピーカ信号 $w_j(t)$ にレンダリングされており、それらの仮想スピーカの位置は単位球上にあり、その単位球上で一様に分布させられるよう目標とされており、前記レンダリングは行列乗算 $w(t) = W^{-1} \cdot c(t)$ によって表現され、 $w(t)$ はすべての仮想スピーカ信号を含むベクトルであり、 W は仮想スピーカ位置モード行列であり、 $c(t)$ は前記HOAデータ・フレーム表現の対応するHOA係数シーケンスのベクトルであり、前記HOAデータ・フレーム表現は

【数4】

$$\|w(t)\|_{\infty} = \max_{1 \leq j \leq 0} |w_j(t)| \leq 1 \quad \forall t$$

10

となるよう正規化されており、当該装置は：

・前記の正規化されたHOAデータ・フレーム表現から動作 a)、 b)、 c)、すなわち a) 前記チャネル信号における優勢音信号を表現するために、HOA係数シーケンスの前記ベクトル $c(t)$ に混合行列 A を乗算する動作であって、混合行列 A のユークリッド・ノルムは1より大きくなり、混合行列 A は前記正規化されたHOAデータ・フレーム表現の係数シーケンスの線形結合を表わす、動作；

b) 前記チャネル信号における周囲成分 $c_{AMB}(t)$ を表現するために、前記正規化されたHOAデータ・フレーム表現から前記優勢音信号を減算し、前記周囲成分 $c_{AMB}(t)$ の係数シーケンスの少なくとも一部を選択し、 $\|c_{AMB}(t)\|_2^2 / \|c(t)\|_2^2$ であり、結果として得られる最小周囲成分 $c_{AMB,MIN}(t)$ を、 $W_{MIN}^{-1} \cdot c_{AMB,MIN}(t)$ を計算することによって変換し、 $\|W_{MIN}^{-1}\|_2 < 1$ であり、 W_{MIN} は前記最小周囲成分 $c_{AMB,MIN}(t)$ についてのモード行列である、動作；

20

c) 前記HOA係数シーケンス $c(t)$ の一部を選択する動作であって、選択された係数シーケンスは、空間変換が適用される前記周囲HOA成分の係数シーケンスに関係し、前記選択された係数シーケンスの数を記述する最小次数 N_{MIN} は $N_{MIN} - 9$ である、動作；

のうちの一つまたは複数によって前記チャネル信号を形成する手段と；

・前記チャネル信号についての前記非差分的な利得値を表現するために必要とされる前記最低整数ビット数 β_e を

30

【数5】

$$\beta_e = \lceil \log_2(\lceil \log_2(\sqrt{K_{MAX}} \cdot 0) \rceil + 1) \rceil$$

に設定する手段であって、

【数6】

$$K_{MAX} = \max_{1 \leq N \leq N_{MAX}} K(N, \Omega_1^{(N)}, \dots, \Omega_0^{(N)})$$

40

であり、 N は前記次数であり、 N_{MAX} は関心対象の最大次数であり、 $\Omega_1^{(N)}, \dots, \Omega_0^{(N)}$ は前記仮想スピーカの方向であり、 $0 = (N+1)^2$ はHOA係数シーケンスの数であり、 K は前記モード行列の二乗されたユークリッド・ノルム $\| \cdot \|_2^2$ と0との間の比である、手段とを含む。

【図面の簡単な説明】

【0015】

本発明の例示的な実施形態が付属の図面を参照して記述される。

【図1】HOA圧縮器を示す図である。

50

【図2】HOA圧縮解除器を示す図である。

【図3】HOA次数 $N=1, \dots, 29$ について、仮想方向 $j^{(N)}$ 、 $1 \leq j \leq 0$ についてのスケール値 K を示す図である。

【図4】HOA次数 $N_{MIN}=1, \dots, 29$ について、仮想方向 $_{MIN,d}^{(N)}$ 、 $d=1, \dots, 0_{MIN}$ についての逆モード行列 $^{-1}$ のユークリッド・ノルムを示す図である。

【図5】位置 $j^{(N)}$ 、 $1 \leq j \leq 0$ 、 $0 = (N+1)^2$ にある仮想スピーカ-の信号の最大許容大きさ $_{dB}$ の決定を示す図である。

【図6】球面座標系を示す図である。

【発明を実施するための形態】

【0016】

たとえ明示的に記述されなくても、以下の実施形態は任意の組み合わせまたはサブコンビネーションにおいて用いることができる。

【0017】

以下では、上述した課題が生起する、より詳細なコンテキストを提供するために、HOA圧縮および圧縮解除の原理が呈示される。この呈示の基礎は非特許文献1のMPEG-H 3Dオーディオ文書に記述された処理である。特許文献1、3、2も参照。非特許文献1においては、「方向性成分」は「優勢音成分 (predominant sound component)」に拡張される。方向性成分として、優勢音成分は、部分的には、方向性信号、つまり対応する方向 (その方向から聴取者に入射すると想定される方向) をもつモノラル信号に、該方向性信号からもとのHOA表現の諸部分を予測するためのいくつかの予測パラメータを合わせたものによって表現されると想定される。加えて、優勢音成分は、「ベクトル・ベースの信号」、つまり該ベクトル・ベースの信号の方向分布を定義する対応するベクトルをもつモノラル信号によって表現されると想定される。

【0018】

HOA圧縮

特許文献3に記載されるHOA圧縮器の全体的なアーキテクチャが図1に示されている。これは、図1のAに描かれる空間的HOAエンコード部と、図1のBに描かれる知覚的および源エンコード部とを有する。空間的HOAエンコード部は、1個の信号からなる第一の圧縮されたHOA表現を、そのHOA表現をどのように生成するかを記述するサイド情報とともに提供する。知覚的およびサイド情報源符号化器では、該1個の信号は知覚的にエンコードされ、該サイド情報は源エンコードにかけられる。その後、二つの符号化された表現が多重化される。

【0019】

空間的HOAエンコード

第一段階では、もとのHOA表現の現在の k 番目のフレーム $C(k)$ が方向およびベクトル推定処理段階またはステージ11に入力される。該段階はタプル集合 $M_{DIR}(k)$ および $M_{VEC}(k)$ を提供すると想定される。タプル集合 $M_{DIR}(k)$ は、第一の要素が方向性信号のインデックスを表わし、第二の要素がそれぞれの量子化された方向を表わす諸タプルからなる。タプル集合 $M_{VEC}(k)$ は、第一の要素がベクトル・ベースの信号のインデックスを表わし、第二の要素がそれらの信号の方向分布、すなわち該ベクトル・ベースの信号のHOA表現がどのようにして計算されるかを定義するベクトルを表わす諸タプルからなる。

【0020】

両方のタプル集合 $M_{DIR}(k)$ および $M_{VEC}(k)$ を使って、初期HOAフレーム $C(k)$ はHOA分解段階またはステージ12において、すべての優勢音 (すなわち、方向性およびベクトル・ベース) 信号のフレーム $X_{PS}(k-1)$ と、周囲HOA成分のフレーム $C_{AMB}(k-1)$ とに分解される。一フレームの遅延に注意されたい。これは、プロッキング・アーチファクトを回避するための重複加算処理のためである。さらに、HOA分解段階/ステージ12は、優勢音HOA成分を豊かにするために、これらの方向性信号からもとのHOA表現の諸部分をどのようにして予測するかを記述するいくつかの予測パラメータ $(k-1)$ を出力すると想定される。さらに、HOA分解処理段階またはステージ12において決定された優勢音信号の、1個の利用可能

10

20

30

40

50

なチャンネルへの割り当てについての情報を含む目標割り当てベクトル (target assignment vector) $v_{A,T}(k-1)$ が提供されると想定される。影響されるチャンネルは占有されると想定されることができる。つまり、それらはそれぞれの時間フレームにおいて周囲HOA成分のいかなる係数シーケンスを転送するためにも利用可能ではない。

【0021】

周囲成分修正処理段階またはステージ13では、周囲HOA成分のフレーム $C_{AMB}(k-1)$ は、目標割り当てベクトル $v_{A,T}(k-1)$ によって与えられる情報に従って修正される。特に、周囲HOA成分のどの係数シーケンスが所与の1個のチャンネルにおいて伝送されるべきかが、(他の側面もあるが中でも) どのチャンネルが利用可能であり、優勢音信号によってすでに占有されていないかについての(目標割り当てベクトル $v_{A,T}(k-1)$ に含まれる) 情報に依

10

【0022】

さらに、周囲HOA成分 $C_{AMB}(k-2)$ の最初の O_{MIN} 個の係数シーケンスは、常に、知覚的に符号化され伝送されるべく選ばれることが想定される。ここで、 $O_{MIN} = (N_{MIN} + 1)^2$ であり、 N_{MIN} は典型的にはもとのHOA表現のものより小さな次数である。これらのHOA係数シーケンスを脱相関するために、これらは、段階/ステージ13において、いくつかのあらかじめ定義された方向 $_{MIN,d}$ 、 $d = 1, \dots, O_{MIN}$ から入射する方向性信号(すなわち、一般平面波関数)に変換されることができる。

20

【0023】

修正された周囲HOA成分 $C_{M,A}(k-1)$ とともに、段階/ステージ13において、時間的に予測された修正された周囲HOA成分 $C_{P,M,A}(k-1)$ が計算され、合理的な先読みを許容するために、利得制御処理段階またはステージ15、151において使用される。ここで、周囲HOA成分の修正についての情報は、チャンネル割り当て段階またはステージ14における、すべての可能な型の信号の、利用可能なチャンネルへの割り当てに直接関係している。割り当てについての最終的な情報は、最終的な割り当てベクトル $v_A(k-2)$ に含まれると想定される。段階/ステージ13においてこのベクトルを計算するために、目標割り当てベクトル $v_{A,T}(k-1)$ に含まれる情報が活用される。

【0024】

段階/ステージ14におけるチャンネル割り当ては、割り当てベクトル $v_A(k-2)$ によって与えられる情報を用いて、フレーム $X_{PS}(k-2)$ に含まれる適切な信号およびフレーム $C_{M,A}(k-2)$ に含まれる適切な信号を、1個の利用可能なチャンネルに割り当て、信号フレーム $y_i(k-2)$ 、 $i = 1, \dots, l$ を与える。さらに、フレーム $X_{PS}(k-1)$ およびフレーム $C_{P,AMB}(k-1)$ に含まれる適切な信号も、1個の利用可能なチャンネルに割り当てられて、予測された信号フレーム $y_{P,i}(k-2)$ 、 $i = 1, \dots, l$ を与える。

30

【0025】

信号フレーム $y_i(k-2)$ 、 $i = 1, \dots, l$ のそれぞれは、最終的に利得制御15、151によって処理されて、指数 $e_i(k-2)$ および例外フラグ $_{i}(k-2)$ 、 $i = 1, \dots, l$ ならびに信号 $z_i(k-2)$ 、 $i = 1, \dots, l$ を与える。ここで、知覚的エンコーダ段階またはステージ16に好適な値範囲を達成するよう信号利得がなめらかに修正される。段階/ステージ16は、対応するエンコードされた信号フレーム

40

【数7】

$$\check{z}_i(k-2), \quad i = 1, \dots, l$$

を出力する。予測された信号フレーム $y_{P,i}(k-2)$ 、 $i = 1, \dots, l$ は、相続くブロックの間の激しい利得変化を避けるために一種の先読みを許容する。サイド情報データ $M_{DIR}(k-1)$ 、

50

$M_{VEC}(k-1)$ 、 $e_i(k-2)$ 、 $\tilde{z}_i(k-2)$ 、 $\tilde{r}(k-1)$ および $v_A(k-2)$ はサイド情報源符号化器段階またはステージ17において源符号化され、エンコードされたサイド情報フレーム【数8】

$$\tilde{r}(k-2)$$

を与える。マルチプレクサ18において、フレーム $(k-2)$ のエンコードされた信号【数9】

$$\tilde{z}_i(k-2)$$

10

およびこのフレームについてのエンコードされたサイド情報データ【数10】

$$\tilde{r}(k-2)$$

が組み合わされて、出力フレーム【数11】

20

$$\tilde{b}(k-2)$$

を与える。

【0026】

空間的HOAデコーダにおいては、段階/ステージ15、151における利得修正が、指数 $e_i(k-2)$ および例外フラグ $\tilde{z}_i(k-2)$ 、 $i=1, \dots, I$ を含む前記利得制御サイド情報を使って反転されると想定される。

30

【0027】

HOA圧縮解除

特許文献3に記載されるHOA圧縮解除器の全体的なアーキテクチャーが図2に示されている。これは、上記HOA圧縮器のコンポーネントの、逆順に配列された対応物からなり、図2のAに描かれる知覚的および源デコード部と、図2のBに描かれる空間的HOAデコード部とを含む。

【0028】

(知覚的およびサイド情報源デコーダを表わす)知覚的および源デコード部において、多重分離段階またはステージ21は、ビットストリームからの入力フレーム

【数12】

40

$$\tilde{b}(k)$$

を受領し、前記I個の信号の知覚的に符号化された表現

【数13】

$$\tilde{z}_i(k), i=1, \dots, I$$

50

と、そのHOA表現をどのようにして生成するかを記述する符号化されたサイド情報データ
【数 1 4】

$$\tilde{\mathbf{r}}(k)$$

とを与える。信号

【数 1 5】

$$\tilde{\mathbf{z}}_i(k)$$

10

は知覚的デコーダ段階またはステージ 2 2 において知覚的にデコードされて、デコードされた信号

【数 1 6】

$$\hat{\mathbf{z}}_i(k), \quad i = 1, \dots, I$$

を与える。符号化されたサイド情報データ

20

【数 1 7】

$$\tilde{\mathbf{r}}(k)$$

はサイド情報源デコーダ段階またはステージ 2 3 においてデコードされて、データ集合 $M_{IR}(k+1)$ 、 $M_{VEC}(k+1)$ 、指数 $e_i(k)$ 、例外フラグ $e_{i,EX}(k)$ 、予測パラメータ $\hat{e}_i(k+1)$ および割り当てベクトル $v_{AMB,ASSIGN}(k)$ を与える。 v_A と $v_{AMB,ASSIGN}$ の間の相違については、上述した MPEG の非特許文献 1 を参照。

【0 0 2 9】

空間的HOAデコード

30

空間的HOAデコード部では、知覚的にデコードされた信号

【数 1 8】

$$\hat{\mathbf{z}}_i(k), \quad i = 1, \dots, I$$

のそれぞれが、関連する利得補正指数 $e_i(k)$ および利得補正例外フラグ $e_{i,EX}(k)$ と一緒に逆利得制御処理段階またはステージ 2 4、2 4 1 に入力される。 i 番目の逆利得制御処理段階 / ステージは利得補正された信号フレーム

【数 1 9】

40

$$\hat{\mathbf{y}}_i(k)$$

$\{\hat{y}_i(k)\}$ を与える。

【0 0 3 0】

I 個の利得補正された信号フレーム

【数 2 0】

$$\hat{y}_i(k), \quad i = 1, \dots, I$$

のすべては割り当てベクトル $v_{AMB, ASSIGN}(k)$ およびタプル集合 $M_{DIR}(k+1)$ および $M_{VEC}(k+1)$ と一緒にチャンネル再割り当て段階またはステージ 2 5 に供給される。タプル集合 $M_{DIR}(k+1)$ および $M_{VEC}(k+1)$ の上記の定義を参照。割り当てベクトル $v_{AMB, ASSIGN}(k)$ は I 個の成分からなり、これらの成分は各伝送チャンネルについて、周囲HOA成分の係数シーケンスを含んでいるかどうかおよびどの係数シーケンスを含んでいるかを示す。チャンネル再割り当て段階 / ステージ 2 5 において、利得補正された信号フレーム $\hat{y}_i(k)$ は、すべての優勢音信号 (すなわちすべての方向性およびベクトル・ベースの信号) のフレーム

10

【数 2 1】

$$\hat{x}_{PS}(k)$$

$[\hat{x}_{PS}(k)]$ および周囲HOA成分の中間表現のフレーム $C_{I, AMB}(k)$ を再構成するために再分配される。さらに、 k 番目のフレームにおいてアクティブである、周囲HOA成分の係数シーケンスのインデックスの集合 $I_{AMB, ACT}(k)$ と、 $(k-1)$ 番目のフレームにおいて有効にされる、無効にされるまたはアクティブなままである必要がある周囲HOA成分の係数インデックスのデータ集合 $I_E(k-1)$ 、 $I_D(k-1)$ および $I_U(k-1)$ とが提供される。

20

【0 0 3 1】

優勢音合成段階またはステージ 2 6 では、優勢音成分

【数 2 2】

$$\hat{c}_{PS}(k-1)$$

$[\hat{c}_{PS}(k-1)]$ のHOA表現が、すべての優勢音信号のフレーム $\hat{x}_{PS}(k)$ から、タプル集合 $M_{DIR}(k+1)$ および予測パラメータの集合 $(k+1)$ 、タプル集合 $M_{VEC}(k+1)$ およびデータ集合 $I_E(k-1)$ 、 $I_D(k-1)$ および $I_U(k-1)$ を使って計算される。

30

【0 0 3 2】

周囲合成段階またはステージ 2 7 では、周囲HOA成分フレーム

【数 2 3】

$$\hat{c}_{AMB}(k-1)$$

$[\hat{c}_{AMB}(k-1)]$ が、周囲HOA成分の中間表現のフレーム $C_{I, AMB}(k)$ から、 k 番目のフレームにおいてアクティブである周囲HOA成分の係数シーケンスのインデックスの集合 $I_{AMB, ACT}(k)$ を使って生成される。一フレームぶんの遅延が、優勢音HOA成分との同期に起因して導入されている。最後に、HOA組成段階またはステージ 2 8 において、周囲HOA成分フレーム $\hat{c}_{AMB}(k-1)$ および優勢音HOA成分のフレーム $\hat{c}_{PS}(k-1)$ が重畳されて、デコードされたHOAフレーム $\hat{c}(k-1)$ を与える。

40

【0 0 3 3】

その後、空間的HOAデコーダは前記 I 個の信号および前記サイド情報から、前記再構成されたHOA表現を生成する。

【0 0 3 4】

エンコーダ側で周囲HOA成分が方向性信号に変換された場合、その変換はデコーダ側で段階 / ステージ 2 7 において反転される。

50

【0035】

HOA圧縮器内の利得制御処理段階／ステージ15、151より前の信号の潜在的な最大利得は、入力HOA表現の値範囲に強く依存する。よって、まず、入力HOA表現についての意味のある値範囲が定義され、その後、利得制御処理段階／ステージにはいる前の前記信号の前記潜在的な最大利得について結論する。

【0036】

入力HOA表現の正規化

本発明の処理を使うために、(全)入力HOA表現信号の正規化が、事前に実行される。HOA圧縮については、フレームごとの処理が実行される。ここで、もとの入力HOA表現のk番目のフレームC(k)は、高次アンビソニックスの基礎の節の式(54)において指定される時間連続的なHOA係数シーケンスのベクトルc(t)に関して

【数24】

$$C(k) := [c((kL+1)T_s) \quad c((kL+2)T_s) \quad \dots \quad c((k+1)LT_s)] \in \mathbb{R}^{O \times L} \quad (1)$$

のように定義される。ここで、kはフレーム・インデックス、Lはフレーム長(サンプル単位)を表わし、 $O = (N+1)^2$ はHOA係数シーケンスの数であり、 T_s はサンプリング周期を示す。

【0037】

特許文献4において述べられているように、実際の観点から見たHOA表現の意味のある正規化は、個々のHOA係数シーケンス $c_n^m(t)$ の値範囲に対して制約条件を課すことによって達成されない。これらの時間領域関数は、レンダリング後にスピーカーによって実際に再生される信号ではないからである。その代わりに、HOA表現を0個の仮想スピーカー信号 $w_j(t)$ 、 $1 \leq j \leq O$ にレンダリングすることによって得られる「等価な空間領域表現」を考へるほうが便利である。それぞれの仮想スピーカー位置は、球面座標系によって表わされると想定される。ここで、各位置は単位球上にあり、動径1をもつと想定される。よって、これらの位置は、次数に依存する諸方向 $\mathbf{j}^{(N)} = (\mathbf{j}_1^{(N)}, \mathbf{j}_2^{(N)})$ 、 $1 \leq j \leq O$ によって等価に表わすことができる。ここで、 $\mathbf{j}_1^{(N)}$ および $\mathbf{j}_2^{(N)}$ はそれぞれ傾斜角および方位角を表わす(球面座標系の定義については図6およびその説明を参照)。これらの方向は、

できるだけ一様に単位球上に分布させられるべきである。たとえば非特許文献2参照。特定の方向の計算のために、ノード数は<http://www.mathematik.uni-dortmund.de/lisx/research/projects/fliege/nodes/nodes.html>にある。これらの位置は一般に、「球状の一様分布」の定義の種類に依存するもので、よって曖昧さがなくもない。

【0038】

仮想スピーカー信号について値範囲を定義することが、HOA係数シーケンスについて値範囲を定義することに対して有利な点は、前者についての値範囲が、PCM表現を想定する通常のスピーカー信号についての場合のように、区間 $[-1, 1]$ に等しく直観的に設定されることができることである。これは、空間的に一様に分布した量子化誤差につながり、そのため有利なことに、量子化は、実際の聴取に関して有意な領域で適用される。このコンテキストにおける重要な側面は、通常ならサンプル当たりより多くのビット数(たとえば24あるいはさらには32)が必要とされるところ、サンプル当たりのビット数が通常のスピーカー信号について典型的にそうであるくらい低く、たとえば16に選ばれることができることである。これは、HOA係数シーケンスの直接量子化に比べて効率を高める。

【0039】

空間領域における正規化プロセスを詳細に記述するために、すべての仮想スピーカー信号は $w(t) := [w_1(t) \quad \dots \quad w_O(t)]^T$ (2) においてまとめられる。ここで、 $(\cdot)^T$ は転置を表わす。仮想方向 $\mathbf{j}^{(N)}$ 、 $1 \leq j \leq O$ に関するモード行列を

10

20

30

40

50

【数 2 5】

$$\Psi := [S_1 \quad \dots \quad S_0] \in \mathbb{R}^{O \times O} \quad (3)$$

ここで $S_j :=$ (4)

$$[S_0^0(\boldsymbol{\Omega}_j^{(N)}) \quad S_1^{-1}(\boldsymbol{\Omega}_j^{(N)}) \quad S_1^0(\boldsymbol{\Omega}_j^{(N)}) \quad S_1^1(\boldsymbol{\Omega}_j^{(N)}) \quad \dots \quad S_N^{N-1}(\boldsymbol{\Omega}_j^{(N)}) \quad S_N^N(\boldsymbol{\Omega}_j^{(N)})]^T$$

によって定義される $w(t)$ で表わすと、レンダリング・プロセスは、行列乗算 10

$$w(t) = (\Psi)^{-1} \cdot c(t) \quad (5)$$

として定式化されることができる。

【0 0 4 0】

これらの定義を使うと、仮想スピーカー信号に対する合理的な要求は：

【数 2 6】

$$\|w(lT_S)\|_\infty = \max_{1 \leq j \leq O} |w_j(lT_S)| \leq 1 \quad \forall l \quad (6)$$

である。これは、各仮想スピーカー信号の大きさは範囲 $[-1, 1]$ 内にあることが要求されることを意味している。時間 t の時刻は、サンプル・インデックス l と前記 HOA データ・フレームのサンプル値のサンプル周期 T_S とによって表現される。 20

【0 0 4 1】

結果として、スピーカー信号の全パワーは、条件

【数 2 7】

$$\|w(lT_S)\|_2^2 = \sum_{j=1}^O |w_j(lT_S)|^2 \leq O \quad \forall l \quad (7)$$

を満たす。HOA データ・フレーム表現のレンダリングおよび正規化は、図 1 の A の入力 $C(k)$ の上流で実行される。 30

【0 0 4 2】

利得制御前の信号値範囲についての帰結

入力 HOA 表現の正規化が 入力 HOA 表現の正規化 の節の記述に従って実行されるとして、HOA 圧縮器における利得制御処理ユニット 1 5、1 5 1 に入力される信号 y_i 、 $i = 1, \dots, I$ の値範囲について以下で考察する。これらの信号は、HOA 係数シーケンスまたは優勢音信号 $x_{PS, d}$ 、 $d = 1, \dots, D$ および / または周囲 HOA 成分 $c_{AMB, n}$ 、 $n = 1, \dots, O$ の特定の諸係数シーケンス（その一部には空間変換が適用される）のうちの一つまたは複数の、利用可能な 1 個のチャンネルへの割り当てによって生成される。よって、式 (6) での正規化の想定のもとに、ここに挙げた異なる信号型の可能な値範囲を分析することが必要である。すべての種類の信号は、もとの HOA 係数シーケンスから中間的に計算されるので、それらの可能な値範囲を見ておく。 40

【0 0 4 3】

1 個のチャンネルにおいて、一つまたは複数の HOA 係数シーケンスのみが含まれる場合は図 1 の A および図 2 の B には描かれていない。すなわち、そのような場合は、HOA 分解、周囲成分修正および対応する合成ブロックは必要とされない。

【0 0 4 4】

HOA 表現の値範囲についての帰結

時間連続的な HOA 表現は仮想スピーカー信号から

$$c(t) = w(t) \quad (8)$$

によって得られる。これは、式(5)の逆演算である。よって、すべてのHOA係数シーケンスの全パワーは、式(8)および(7)を使って次のように制限される。

【 0 0 4 5 】

【 数 2 8 】

$$\|c(lT_S)\|_2^2 \leq \|\Psi\|_2^2 \cdot \|w(lT_S)\|_2^2 \leq \|\Psi\|_2^2 \cdot O \quad (9)$$

球面調和関数のN3D正規化の想定のもとでは、モード行列の二乗されたユークリッド・ノルムは

$$\| \cdot \|_2^2 = K \cdot O \quad (10a)$$

によって書くことができる。ここで、

$$K = \| \cdot \|_2^2 / O \quad (10b)$$

はモード行列の二乗されたユークリッド・ノルムとHOA係数シーケンスの数Oとの間の比を表わす。この比は特定のHOA次数Nおよび特定の諸仮想スピーカー方向 $j^{(N)}$ 、 $1 \leq j \leq O$ に依存する。このことは、

$$K = K(N, j_1^{(N)}, \dots, j_O^{(N)}) \quad (10c)$$

のように、この比の後に個々のパラメータ・リストを付けることによって表わせる。

【 0 0 4 6 】

図3は、上述した非特許文献2の論文に従って仮想方向 $j^{(N)}$ 、 $1 \leq j \leq O$ についてのKの値を、HOA次数N = 1, ..., 29について示している。

【 0 0 4 7 】

すべてのこれまでの議論および考察を組み合わせると、HOA係数シーケンスの絶対値についての上限が次のように与えられる。

【 0 0 4 8 】

【 数 2 9 】

$$\|c(lT_S)\|_\infty \leq \|c(lT_S)\|_2 \leq \sqrt{K} \cdot O \quad (11)$$

ここで、最初の不等号はノルムの定義から直接帰結する。

【 0 0 4 9 】

式(6)における条件は式(11)における条件を含意するが、逆は成り立たない、すなわち式(11)は式(6)を含意しないことに注意しておくことが重要である。

【 0 0 5 0 】

さらに重要な側面は、ほぼ一様に分布した仮想スピーカー位置の想定のもとで、仮想スピーカー位置に関するモード・ベクトルを表わすモード行列の列ベクトルは、ほぼ互いに直交であり、それぞれN+1のユークリッド・ノルムをもつ。この属性は、前記空間変換が、乗算定数を除いてユークリッド・ノルムをほぼ保存することを意味する。すなわち、

【 数 3 0 】

$$\|c(lT_S)\|_2 \approx (N+1) \|w(lT_S)\|_2 \quad (12)$$

モード・ベクトルに対する直交性の想定が破られるほど、真のノルム $\|c(lT_S)\|_2$ は式(12)の近似から異なってくる。

【 0 0 5 1 】

優勢音信号の値範囲についての帰結

優勢音信号の両方の型(方向性およびベクトル・ベース)は、HOA表現への寄与が、N+1のユークリッド・ノルムをもつ、すなわち

10

20

30

40

50

$$\|v_1\|_2 = N+1 \quad (13)$$

となる単一のベクトル v_1 、 R^0 によって記述されることで共通している。

【 0 0 5 2 】

方向性信号の場合、このベクトルは、ある信号源方向 $s_{,1}$ に関するモード・ベクトルに対応する、すなわち、

【 数 3 1 】

$$v_1 = S(\Omega_{s,1}) \quad (14)$$

$$:= [S_0^0(\Omega_{s,1}) \ S_1^{-1}(\Omega_{s,1}) \ S_1^0(\Omega_{s,1}) \ S_1^1(\Omega_{s,1}) \ \dots \ S_N^{N-1}(\Omega_{s,1}) \ S_N^N(\Omega_{s,1})]^T \quad (15)$$

10

このベクトルは、HOA表現によって、信号源方向 $s_{,1}$ への方向性ビームを記述する。ベクトル・ベースの信号の場合、ベクトル v_1 はいかなる方向に関するモード・ベクトルにも制約されず、よってモノラルのベクトル・ベースの信号の、より一般的な方向性分布を記述しうる。

【 0 0 5 3 】

以下では、D個の優勢音信号 $x_d(t)$ 、 $d=1, \dots, D$ の一般的な場合が考察される。これらの信号は、

$$x(t) = [x_1(t) \ x_2(t) \ \dots \ x_D(t)]^T \quad (16)$$

に従ってベクトル $x(t)$ に集められることができる。これらの信号は、モノラルの優勢音信号 $x_d(t)$ 、 $d=1, \dots, D$ の方向性分布を表わすすべてのベクトル v_d 、 $d=1, \dots, D$ から形成される行列

$$V := [v_1 \ v_2 \ \dots \ v_D] \quad (17)$$

に基づいて決定される必要がある。

【 0 0 5 4 】

優勢音信号 $x(t)$ の意味のある抽出のためには、以下の制約条件が定式化される：

a) 各優勢音信号はもとのHOA表現の係数シーケンスの線形結合として得られる、すなわち

$$x(t) = A \cdot c(t) \quad (18)$$

ここで、 $A \in R^{D \times O}$ は混合行列を表わす。

b) 混合行列 A は、そのユークリッド・ノルムが値1を超えない、すなわち

【 数 3 2 】

$$\|A\|_2 \stackrel{!}{\leq} 1 \quad (19)$$

ように、かつもとのHOA表現と優勢音信号のHOA表現との間の残差の二乗されたユークリッド・ノルム（または等価だがパワー）がもとのHOA表現の二乗されたユークリッド・ノルム（または等価だがパワー）より大きくない、すなわち

【 数 3 3 】

$$\|c(t) - V \cdot x(t)\|_2^2 \stackrel{!}{\leq} \|c(t)\|_2^2 \quad (20)$$

40

となるよう、選ばれるべきである。

【 0 0 5 5 】

式(18)を式(20)に代入すると、式(20)が制約条件

【数34】

$$\|I - V \cdot A\|_2 \leq 1 \quad (21)$$

と等価であることが見て取れる。ここで、 I は恒等行列を表わす。

【0056】

式(18)および(19)における制約条件ならびにユークリッド行列とベクトル・ノルムの整合性から、優勢音信号の絶対値についての上限は、式(18)、(19)および(11)を使って、

【数35】

10

$$\|x(lT_s)\|_\infty \leq \|x(lT_s)\|_2 \quad (22)$$

$$\leq \|A\|_2 \|c(lT_s)\|_2 \quad (23)$$

$$\leq \sqrt{K} \cdot O, \quad (24)$$

によって見出される。よって、優勢音信号がもとのHOA係数シーケンスと同じ範囲(式(11)参照)内に留まること、すなわち、

【数36】

20

$$\boxed{\|x(lT_s)\|_\infty \leq \sqrt{K} \cdot O} \quad (25)$$

となることが保証される。

【0057】

混合行列の選択のための例

制約条件(20)を満たす混合行列をどのようにして決定するかの例が、抽出後の残差のユークリッド・ノルムが最小化される、すなわち

【数37】

30

$$x(t) = \operatorname{argmin}_{x(t)} \|V \cdot x(t) - c(t)\|_2 \quad (26)$$

となるように優勢音信号を計算することによって得られる。式(26)の最小化問題に対する解は

$$x(t) = V^+ c(t) \quad (27)$$

によって与えられる。ここで、 $(\cdot)^+$ はムーア・ペンローズの擬似逆行列を示す。式(27)を式(18)と比較することによって、この場合、混合行列が行列 V のムーア・ペンローズ擬似逆行列に等しい、すなわち $A = V^+$ となることがわかる。

40

【0058】

にもかかわらず、行列 V は相変わらず制約条件(19)、すなわち

【数38】

$$\boxed{\|V^+\|_2 \leq 1} \quad (28)$$

を満たすよう選ばれる必要がある。

【0059】

方向性信号のみの場合、行列 V はいくつかの源信号方向 $s_{,d}$ 、 $d = 1, \dots, D$ に関するモー

50

ド行列、すなわち

【数 3 9】

$$\mathbf{V} = [\mathbf{s}(\Omega_{S,1}) \quad \mathbf{s}(\Omega_{S,2}) \quad \dots \quad \mathbf{s}(\Omega_{S,D})] \quad (29)$$

であり、この場合、制約条件(28)は、任意の二つの隣接する方向の距離が小さすぎないように源信号方向 $s_{,d}$ 、 $d=1, \dots, D$ を選ぶことによって満たされることができる。

【0060】

周囲HOA成分の係数シーケンスの値範囲についての帰結

周囲HOA成分は、もとのHOA表現から優勢音信号のHOA表現を引くことによって計算される。すなわち、

【数 4 0】

$$\mathbf{c}_{\text{AMB}}(t) = \mathbf{c}(t) - \mathbf{V} \cdot \mathbf{x}(t). \quad (30)$$

優勢音信号 $\mathbf{x}(t)$ のベクトルが基準(20)に従って決定される場合、

【数 4 1】

$$\|\mathbf{c}_{\text{AMB}}(lT_S)\|_{\infty} \leq \|\mathbf{c}_{\text{AMB}}(lT_S)\|_2 \quad (31)$$

$$\stackrel{(30)}{=} \|\mathbf{c}(lT_S) - \mathbf{V} \cdot \mathbf{x}(lT_S)\|_2 \quad (32)$$

$$\stackrel{(20)}{\leq} \|\mathbf{c}(lT_S)\|_2 \quad (33)$$

$$\stackrel{(11)}{=} \sqrt{K} \cdot 0. \quad (34)$$

と結論できる。

【0061】

周囲HOA成分の空間変換された係数シーケンスの値範囲

特許文献2および上述した非特許文献1のMPEG文書において提案されたHOA圧縮処理におけるさらなる側面は、周囲HOA成分の最初の 0_{MIN} 個の係数シーケンスが常に、トランスポート・チャンネルに割り当てられるよう選ばれるということである。ここで、 $0_{\text{MIN}} = (N_{\text{MIN}} + 1)^2$ であり、 N_{MIN} は典型的にはもとのHOA表現の次数よりも小さな次数である。これらのHOA係数シーケンスを脱相関させるために、これらは(入力HOA表現の正規化の節で述べた概念と同様に)いくつかのあらかじめ定義された方向 $_{\text{MIN},d}$ 、 $d=1, \dots, 0_{\text{MIN}}$ から入射する仮想スピーカ信号に変換されることができる。次数インデックス n N_{MIN} をもつ周囲HOA成分のすべての係数シーケンスのベクトルを $\mathbf{c}_{\text{AMB},\text{MIN}}(t)$ によって定義し、仮想方向 $_{\text{MIN},d}$ 、 $d=1, \dots, 0_{\text{MIN}}$ に関するモード行列を $_{\text{MIN}}$ によって定義すると、 $\mathbf{w}_{\text{MIN}}(t)$ という(によって定義される)すべての仮想スピーカ信号のベクトルは

【数 4 2】

$$\mathbf{w}_{\text{MIN}}(t) = \Psi_{\text{MIN}}^{-1} \cdot \mathbf{c}_{\text{AMB},\text{MIN}}(t) \quad (35)$$

によって得られる。

【0062】

よって、ユークリッド行列とベクトル・ノルムの整合性を使うと、
【数43】

$$\|\mathbf{w}_{\text{MIN}}(lT_S)\|_{\infty} \leq \|\mathbf{w}_{\text{MIN}}(lT_S)\|_2 \quad (36)$$

$$\stackrel{(35)}{\leq} \|\Psi_{\text{MIN}}^{-1}\|_2 \cdot \|\mathbf{c}_{\text{AMB,MIN}}(lT_S)\|_2 \quad (37)$$

$$\stackrel{(34)}{\leq} \|\Psi_{\text{MIN}}^{-1}\|_2 \cdot \sqrt{K} \cdot 0 \quad (38)$$

10

となる。

【0063】

上述した非特許文献1のMPEG文書においては、仮想方向 $\theta_{\text{MIN},d}$ 、 $d=1, \dots, 0_{\text{MIN}}$ は上述した非特許文献2の論文に従って選ばれている。モード行列 \mathbf{M}_{MIN} の逆行列のそれぞれのユークリッド・ノルムが次数 $N_{\text{MIN}}=1, \dots, 9$ について図4に示されている。

【0064】

【数44】

$$\|\Psi_{\text{MIN}}^{-1}\|_2 < 1 \quad N_{\text{MIN}} = 1, \dots, 9 \text{ について} \quad (39)$$

20

であることが見て取れる。

【0065】

しかしながら、 $N_{\text{MIN}} > 9$ についてはこのことは一般には成り立たない。この場合、 $\|\mathbf{M}_{\text{MIN}}^{-1}\|_2$ の値は典型的には1よりずっと大きくなる。それにもかかわらず、少なくとも $N_{\text{MIN}} = 9$ については、仮想スピーカ信号の振幅は次式によって制限される。

【0066】

【数45】

$$\|\mathbf{w}_{\text{MIN}}(lT_S)\|_{\infty} \stackrel{(38), \text{Fig.4}}{\leq} \sqrt{K} \cdot 0 \text{ for } 1 \leq N_{\text{MIN}} \leq 9 \quad (40)$$

30

HOA表現から生成される仮想スピーカ信号の振幅が値1を超えないことを要求する条件(6)を満たすよう入力HOA表現を制約することによって、利得制御前の信号の振幅が値 $(K) \cdot 0$ を超えないことが、次の条件のもとで、保証できる(式(25)、(34)、(40)参照)：

a) すべての優勢音信号 $x(t)$ のベクトルが式/制約条件(18)、(19)、(20)に従って計算される；

b) 仮想スピーカ位置として上述した非特許文献2の論文において定義されるものが使われる場合、空間変換が適用される周囲HOA成分の最初の諸係数シーケンスの数 0_{MIN} を決定する最小次数 N_{MIN} が9未満である必要がある。

40

【0067】

関心対象の最大次数 N_{MAX} までの任意の次数 N 、すなわち $1 \leq N \leq N_{\text{MAX}}$ について、利得制御前の信号の振幅が値 $(K_{\text{MAX}}) \cdot 0$ を超えないことが結論できる。ここで、

【数46】

$$K_{\text{MAX}} = \max_{1 \leq N \leq N_{\text{MAX}}} K(N, \Omega_1^{(N)}, \dots, \Omega_0^{(N)}) \quad (41a)$$

特に、図3から、初期空間変換について仮想スピーカ方向 $\theta_j^{(N)}$ 、 $1 \leq j \leq 0$ が非特許文

50

献2の論文における分布に従って選ばれていると想定される場合であり、加えて、関心対象の最大次数が $N_{MAX} = 29$ である（たとえば非特許文献1のMPEG文書のように）と想定される場合、この特別な場合には $K_{MAX} < 1.5$ なので、利得制御前の信号の振幅は1.50を超えない。すなわち、 $K_{MAX} = 1.5$ が選択されることができる。

【0068】

K_{MAX} は関心対象の最大次数 N_{MAX} および仮想スピーカー方向 $\theta_j^{(N)}$ 、 $1 \leq j \leq N$ に依存し、次のように表わせる。

【数47】

$$K_{MAX} = K_{MAX} \left(\left\{ \Omega_1^{(N)}, \dots, \Omega_0^{(N)} \mid 1 \leq N \leq N_{MAX} \right\} \right) . \quad (41b)$$

10

よって、知覚的符号化前の信号が区間 $[-1, 1]$ 内にあることを保証するために利得制御によって適用される最大利得は

【数48】

$2^{e_{MIN}}$, ここで

$$e_{MIN} = -\lceil \log_2(\sqrt{K_{MAX}} \cdot 0) \rceil < 0 . \quad (41c)$$

20

によって与えられる。

【0069】

利得制御前の信号の振幅があまりに小さい場合には、非特許文献1のMPEG文書において、それらの振幅を

【数49】

$$2^{e_{MAX}}$$

30

までの因子でなめらかに増幅することが可能であることが提案されている。ここで、 e_{MAX} は符号化されたHOA表現内でサイド情報として伝送される。

【0070】

このように、最初から現在フレームまでに利得制御処理ユニットによって引き起こされた、修正された信号の合計の絶対的な振幅変化をアクセス単位内で記述する底2に対するそれぞれの指数は、区間 $[e_{MIN}, e_{MAX}]$ 内の任意の整数値を取ることができる。結果として、それを符号化するために必要とされるビットの（最低の整数の）数 β_e は次式によって与えられる。

【0071】

【数50】

$$\beta_e = \lceil \log_2(|e_{MIN}| + e_{MAX} + 1) \rceil = \lceil \log_2(\lceil \log_2(\sqrt{K_{MAX}} \cdot 0) \rceil + e_{MAX} + 1) \rceil . \quad (42)$$

40

利得制御前の信号の振幅が小さすぎない場合には、式(42)は次のように単純化できる。

【0072】

【数5 1】

$$\beta_e = \lceil \log_2(|e_{\text{MIN}}| + 1) \rceil = \lceil \log_2(\lceil \log_2(\sqrt{K_{\text{MAX}}} \cdot O) \rceil + 1) \rceil . \quad (42a)$$

このビット数 β_e は、利得制御段階 / ステージ 1 5、...、1 5 1 の入力において計算されることができる。

【0 0 7 3】

指数のためのこのビット数 β_e を使うと、HOA圧縮器利得制御処理ユニット 1 5、...、1 5 1 によって引き起こされるすべての可能な絶対的な振幅変化が捕捉できることが保証され、圧縮された表現内のいくつかのあらかじめ定義された入場点において圧縮解除を開始することが許容される。

10

【0 0 7 4】

HOA圧縮解除器において、圧縮されたHOA表現の圧縮解除を開始するとき、いくつかのデータ・フレームについてサイド情報に割り当てられた合計の絶対的な振幅変化を表わし、受領されたデータ・ストリーム

【数5 2】

\tilde{B}

20

のうちからデマルチプレクサ 2 1 から受領される非差分的な利得値は、利得制御段階 / ステージ 1 5、...、1 5 1 において実行された処理の逆の仕方で、正しい利得制御を適用するために、逆利得制御段階またはステージ 2 4、...、2 4 1 において使われる。

【0 0 7 5】

さらなる実施形態

HOA圧縮、空間的HOAエンコード、HOA圧縮解除および空間的HOAデコードの節において述べたような具体的なHOA圧縮 / 圧縮解除システムを実装するとき、前記指数を符号化するためのビットの量 β_e が、スケーリング因子 $K_{\text{MAX,DES}}$ に依存して式(42)に従って設定される必要がある。この $K_{\text{MAX,DES}}$ 自身は圧縮されるべきHOA表現の所望される (desired) 最大次数 $N_{\text{MAX,DES}}$ およびある種の仮想スピーカー方向

30

【数5 3】

$$\Omega_{\text{DES},1}^{(N)}, \dots, \Omega_{\text{DES},0}^{(N)}, \quad 1 \leq N \leq N_{\text{MAX}}$$

に依存する。

【0 0 7 6】

たとえば、 $N_{\text{MAX,DES}} = 29$ を想定し、非特許文献 2 の論文に従って仮想スピーカー方向を選ぶとき、合理的な選択は $K_{\text{MAX,DES}} = 1.5$ であろう。その状況では、同じ仮想スピーカー方向 $\Omega_{\text{DES},1}^{(N)}, \dots, \Omega_{\text{DES},0}^{(N)}$ を使って 入力HOA表現の正規化 の節に従って正規化されている、 $1 \leq N \leq N_{\text{MAX}}$ となる次数 N のHOA表現については、正しい圧縮が保証される。しかしながら、この保証は、(効率性の理由のために) やはりPCMフォーマットで仮想スピーカー信号によって等価に表現されているが、仮想スピーカーの方向 $\Omega_j^{(N)}$ 、 $1 \leq j \leq 0$ がシステム設計段階で想定された上記の仮想スピーカー方向 $\Omega_{\text{DES},1}^{(N)}, \dots, \Omega_{\text{DES},0}^{(N)}$ とは異なるように選ばれているHOA表現の場合には、与えられないことができない。

40

【0 0 7 7】

仮想スピーカー位置のこの異なる選択のため、たとえこれらの仮想スピーカー信号が区間 $[1,1]$ 内にあったとしても、利得制御前の信号の振幅が値 $(K_{\text{MAX,DES}}) \cdot O$ を超えないことはもはや保証できない。よって、このHOA表現が、非特許文献 1 のMPEG文書において記述される処理に従った圧縮のために適正な正規化をもつことは保証できない。

50

【 0 0 7 8 】

この状況において、それぞれのHOA表現が非特許文献1のMPEG文書において記述される処理に従った圧縮のために好適であることを保証するために、仮想スピーカ位置の知識に基づいて、仮想スピーカ信号の最大限許容される振幅を与えるシステムをもつことが有利である。図5では、そのようなシステムが示されている。これは、 $0 = (N+1)^2$ 、 $N = N_0$ であるとして、入力として仮想スピーカ位置 $j^{(N)}$ 、 $1 \leq j \leq N_0$ を取り、出力として仮想スピーカ信号の(デシベルで測った)最大限許容される振幅 γ_{dB} を与える。段階またはステージ51では、諸仮想スピーカ位置に関するモード行列 Ψ が式(3)に従って計算される。続く段階またはステージ52では、該モード行列のユークリッド・ノルム $\|\Psi\|_2$ が計算される。第三の段階またはステージ53では、振幅 γ が、1、ならびに、仮想スピーカ位置の数の平方根と $K_{MAX,DES}$ の積とモード行列のユークリッド・ノルムとの間の商のうち、

すなわち、

$$\gamma = \min\left(1, \frac{\sqrt{O} \cdot \sqrt{K_{MAX,DES}}}{\|\Psi\|_2}\right) \quad (43)$$

デシベル単位での値は

$$\gamma_{dB} = 20 \log_{10}(\gamma) \quad (44)$$

によって得られる。

【 0 0 7 9 】

説明のために：上記の導出から、HOA係数シーケンスの大きさが値 $(K_{MAX,DES}) \cdot O$ を超えなければ、すなわち

【数55】

$$\|c(lT_S)\|_\infty \leq \sqrt{K_{MAX,DES}} \cdot O \quad (45)$$

であれば、利得制御処理ユニット15、151より前のすべての信号は相応してこの値を超えないことが見て取れる。これは、適正なHOA圧縮のための要件である。

【 0 0 8 0 】

式(9)から、HOA係数シーケンスの大きさが

【数56】

$$\|c(lT_S)\|_\infty \leq \|c(lT_S)\|_2 \leq \|\Psi\|_2 \cdot \|w(lT_S)\|_2 \quad (46)$$

によって制限されることが見出される。結果として、 γ が式(43)に従って設定され、PCMフォーマットでの仮想スピーカ信号が

【数57】

$$\|w(lT_S)\|_\infty \leq \gamma \quad (47)$$

を満たす場合、式(7)から、

【数58】

$$\|w(lT_S)\|_2 \leq \gamma \cdot \sqrt{O} \quad (48)$$

10

20

30

40

50

となり、要件(45)が満たされていることになる。

【0081】

すなわち、式(6)における最大の大きさの値1が、式(47)では最大の大きさの値 によって置き換えられる。

【0082】

高次アンビソニックスの基礎

高次アンビソニックス(HOA)は、音源がないと想定されるコンパクトな関心領域内の音場の記述に基づく。その場合、関心領域内の位置 x および時刻 t における音圧の空間時間的挙動 $p(t, x)$ は、斉次の波の式(homogeneous wave equation)によって物理的に完全に決定される。以下では、図6に示される球面座標系を想定する。使用されるこの座標系では、 x 軸は前方位置を向き、 y 軸は左を向き、 z 軸は上を向く。空間内の位置 $x = (r, \theta, \phi)^T$ は動径 $r > 0$ (すなわち、座標原点までの距離)、極軸 z から測った傾斜角 $\theta \in [0, \pi]$ および xy 平面において x 軸から反時計回りに測った方位角 $\phi \in [0, 2\pi]$ によって表現される。さらに、 $(\cdot)^T$ は転置を表わす。

【0083】

すると、 ω が角周波数を表わし、 i は虚数単位を示すものとして、非特許文献3の教科書から、

$\mathcal{F}_t(\cdot)$ によって表わされる時間に関する音圧のフーリエ変換、すなわち

【数59】

$$P(\omega, x) = \mathcal{F}_t(p(t, x)) = \int_{-\infty}^{\infty} p(t, x) e^{-i\omega t} dt \quad (49)$$

は、

【数60】

$$P(\omega = kc_s, r, \theta, \phi) = \sum_{n=0}^N \sum_{m=-n}^n A_n^m(k) j_n(kr) S_n^m(\theta, \phi) \quad (50)$$

に従って球面調和関数級数に展開されることが示せる。ここで、 c_s は音速を表わし、 k は角波数を表わす。角波数は角周波数 ω に、 $k = \omega / c_s$ によって関係付けられる。さらに、 $j_n(\cdot)$ は第一種の球面ベッセル関数を表わし、 $S_n^m(\theta, \phi)$ は次数(order) n 、陪数(degree) m の実数値の球面調和関数を表わす。これは実数値球面調和関数の定義の節で定義される。展開係数 $A_n^m(k)$ は角波数 k のみに依存する。音圧が空間的に帯域制限されていることが暗黙的に想定されていることを注意しておく。よって、級数は次数インデックス n に関して上限 N で打ち切られる。この N はHOA符号化表現の次数と呼ばれる。

【0084】

音場が、角タプル (θ, ϕ) によって指定されるすべての可能な方向から到来する、異なる角周波数 ω の無限個の調和平面波の重ね合わせによって表現されたとすると、それぞれの平面波複素振幅関数 $C(\omega, \theta, \phi)$ は次の球面調和関数展開によって表わせることを示せる(非特許文献4)。

【0085】

【数61】

$$C(\omega = kc_s, \theta, \phi) = \sum_{n=0}^N \sum_{m=-n}^n C_n^m(k) S_n^m(\theta, \phi) \quad (51)$$

ここで、展開係数 $C_n^m(k)$ は展開係数 $A_n^m(k)$ に、

$$A_n^m(k) = i^n C_n^m(k) \quad (52)$$

によって関係付けられる。個々の係数 $C_n^m(k = \omega / c_s)$ が角周波数 ω の関数であるとする

10

20

30

40

50

、逆フーリエ変換 ($\mathcal{F}^{-1}(\cdot)$ によって表わされる) の適用は、各次数 n および陪数 m について、時間領域関数

【数 6 2】

$$c_n^m(t) = \mathcal{F}_t^{-1}(C_n^m(\omega/c_s)) = \frac{1}{2\pi} \int_{-\infty}^{\infty} C_n^m\left(\frac{\omega}{c_s}\right) e^{i\omega t} d\omega \quad (53)$$

を与える。これらの時間領域関数はここでは連続時間HOA係数シーケンスと称され、これは

【数 6 3】

10

$$\mathbf{c}(t) = \quad (54)$$

$$\left[c_0^0(t) \ c_1^{-1}(t) \ c_1^0(t) \ c_1^1(t) \ c_2^{-2}(t) \ c_2^{-1}(t) \ c_2^0(t) \ c_2^1(t) \ c_2^2(t) \ \dots \ c_N^{N-1}(t) \ c_N^N(t) \right]^T$$

によって単一のベクトル $\mathbf{c}(t)$ にまとめることができる。

【 0 0 8 6 】

ベクトル $\mathbf{c}(t)$ 内のHOA係数シーケンス $c_n^m(t)$ の位置インデックスは $n(n+1) + 1 + m$

20

によって与えられる。ベクトル $\mathbf{c}(t)$ 内の全体的な要素数は $O = (N+1)^2$ によって与えられる。

最終的なアンビソニックス・フォーマットは、サンプリング周波数 f_s を使って、 $\mathbf{c}(t)$ のサンプリングされたバージョンを、

【数 6 4】

$$\{\mathbf{c}(lT_s)\}_{l \in \mathbb{N}} = \{\mathbf{c}(T_s), \mathbf{c}(2T_s), \mathbf{c}(3T_s), \mathbf{c}(4T_s), \dots\} \quad (55)$$

30

として与える。ここで、 $T_s = 1/f_s$ はサンプリング周期を表わす。 $\mathbf{c}(lT_s)$ の要素は離散時間HOA係数シーケンスと称される。これは常に実数値であることが示せる。この属性は、連続時間バージョン $c_n^m(t)$ についても成り立つ。

【 0 0 8 7 】

実数値の球面調和関数の定義

実数値の球面調和関数 $S_n^m(\theta, \phi)$ (非特許文献 5、3.1章に基づくSN3D規格化を想定) は次式によって与えられる。

【 0 0 8 8 】

【数 6 5】

40

$$S_n^m(\theta, \phi) = \sqrt{(2n+1) \frac{(n-|m|)!}{(n+|m|)!}} P_{n,|m|}(\cos\theta) \text{trg}_m(\phi) \quad (56)$$

ここで

$$\text{trg}_m(\phi) = \begin{cases} \sqrt{2} \cos(m\phi) & m > 0 \\ 1 & m = 0 \\ -\sqrt{2} \sin(m\phi) & m < 0 \end{cases} \quad (57)$$

50

ルジャンドル陪関数 $P_{n,m}(x)$ は次式によって定義される。
【数 6 6】

$$P_{n,m}(x) = (1-x^2)^{m/2} \frac{d^m}{dx^m} P_n(x), m \geq 0 \quad (58)$$

ここで、ルジャンドル多項式 $P_n(x)$ を用いているが、非特許文献 3 とは異なり、コンドン・ショートリー (Condon-Shortley) 位相項 $(-1)^m$ がない。

【0089】

本発明は、単一のプロセッサまたは電子回路によって、あるいは並列に動作するおよび / または本発明の処理の異なる部分で動作するいくつかのプロセッサまたは電子回路によって実行されることができる。

【0090】

かかるプロセッサ (単数または複数) を動作させるための命令は一つまたは複数のメモリに記憶されることができる。

いくつかの態様を記載しておく。

〔態様 1〕

HOAデータ・フレーム表現 $(C(k))$ の圧縮のために、前記HOAデータ・フレームのうちの個々のもののチャンネル信号についての非差分的な利得値 (2°) を表現するために必要とされる最低の整数ビット数 ρ を決定する方法であって、各フレームにおける各チャンネル信号はサンプル値のグループを含み、前記HOAデータ・フレームの各フレームの各チャンネル信号 $(y_1(k-2), \dots, y_L(k-2))$ に対して差分利得値が割り当てられ、そのような差分利得値は現在HOAデータ・フレーム $(k-2)$ におけるチャンネル信号のサンプル値の振幅の、直前のHOAデータ・フレーム $(k-3)$ におけるそのチャンネル信号のサンプル値に対する変化を引き起こすものであり、そのような利得適応されたチャンネル信号はエンコード (16) においてエンコードされ、

前記HOAデータ・フレーム表現 $(C(k))$ は空間領域において O 個の仮想スピーカー信号 $w_j(t)$ にレンダリングされており、それらの仮想スピーカーの位置は単位球上にあり、その単位球上で一様に分布させられるよう目標とされており、前記レンダリングは行列乗算 $w(t) = ()^{-1} \cdot c(t)$ によって表現され、 $w(t)$ はすべての仮想スピーカー信号を含むベクトルであり、 $()$ は仮想スピーカー位置モード行列であり、 $c(t)$ は前記HOAデータ・フレーム表現 $(C(k))$ の対応するHOA係数シーケンスのベクトルであり、前記HOAデータ・フレーム表現 $(C(k))$ は

【数 6 7】

$$\|w(t)\|_\infty = \max_{1 \leq j \leq O} |w_j(t)| \leq 1 \quad \forall t$$

となるよう正規化されており、当該方法は：

・前記の正規化されたHOAデータ・フレーム表現 $(C(k))$ から、前記チャンネル信号 $(y_1(k-2), \dots, y_L(k-2))$ を、サブステップ a)、b)、c)、すなわち
a) 前記チャンネル信号における優勢音信号 $(x(t))$ を表現するために、HOA係数シーケンスの前記ベクトル $c(t)$ に混合行列 A を乗算するサブステップであって、混合行列 A のユークリッド・ノルムは1より大きくなく、混合行列 A は前記正規化されたHOAデータ・フレーム表現の係数シーケンスの線形結合を表わす、サブステップ；

b) 前記チャンネル信号における周囲成分 $c_{AMB}(t)$ を表現するために、前記正規化されたHOAデータ・フレーム表現 $(C(k))$ から前記優勢音信号を減算し、前記周囲成分 $c_{AMB}(t)$ の係数シーケンスの少なくとも一部を選択し、 $\|c_{AMB}(t)\|_2^2 - \|c(t)\|_2^2$ であり、結果として得られる最小周囲成分 $c_{AMB_MIN}(t)$ を、 $w_{MIN}(t) = MIN^{-1} \cdot c_{AMB_MIN}(t)$ を計算することによって変換し、 $\|MIN^{-1}\|_2 < 1$ であり、 MIN は前記最小周囲成分 $c_{AMB_MIN}(t)$ についてのモード行列である、サブステップ；

10

20

30

40

50

c) 前記HOA係数シーケンス $c(t)$ の一部を選択するサブステップであって、選択された係数シーケンスは、空間変換が適用される前記周囲HOA成分の係数シーケンスに関係し、前記選択された係数シーケンスの数を記述する最小次数 N_{MIN} は $N_{MIN} - 9$ である、サブステップ;

のうちの一つまたは複数によって形成する段階と;

・前記チャネル信号についての前記非差分的な利得値 (2°) を表現するために必要とされる前記最低の整数ビット数 β_e を

【数 6 8】

$$\beta_e = \lceil \log_2(\lceil \log_2(\sqrt{K_{MAX}} \cdot O) \rceil + 1) \rceil$$

10

に設定する段階であって、

【数 6 9】

$$K_{MAX} = \max_{1 \leq N \leq N_{MAX}} K(N, \Omega_1^{(N)}, \dots, \Omega_0^{(N)})$$

であり、 N は前記次数であり、 N_{MAX} は関心対象の最大次数であり、 $\Omega_1^{(N)}, \dots, \Omega_0^{(N)}$ は前記仮想スピーカ-方向であり、 $O = (N + 1)^2$ はHOA係数シーケンスの数であり、 K は前記モード行列の二乗されたユークリッド・ノルム $\| \cdot \|_2^2$ と O との間の比である、段階とを含む、

20

方法。

〔態様 2〕

HOAデータ・フレーム表現 ($C(k)$) の圧縮のために、前記HOAデータ・フレームのうちの個々のもののチャネル信号についての非差分的な利得値 (2°) を表現するために必要とされる最低の整数ビット数 β_e を決定する装置であって、各フレームにおける各チャネル信号はサンプル値のグループを含み、前記HOAデータ・フレームの各フレームの各チャネル信号 ($y_1(k-2), \dots, y_L(k-2)$) に対して差分利得値が割り当てられ、そのような差分利得値は現在HOAデータ・フレーム ($(k-2)$) におけるチャネル信号のサンプル値の振幅の、直前のHOAデータ・フレーム ($(k-3)$) におけるそのチャネル信号のサンプル値に対する変化を引き起こすものであり、そのような利得適応されたチャネル信号はエンコード (16) においてエンコードされ、

30

前記HOAデータ・フレーム表現 ($C(k)$) は空間領域において O 個の仮想スピーカ-信号 $w_j(t)$ にレンダリングされており、それらの仮想スピーカ-の位置は単位球上にあり、その単位球上で一様に分布させられるよう目標とされており、前記レンダリングは行列乗算 $w(t) = (\cdot)^{-1} \cdot c(t)$ によって表現され、 $w(t)$ はすべての仮想スピーカ-信号を含むベクトルであり、 \cdot は仮想スピーカ-位置モード行列であり、 $c(t)$ は前記HOAデータ・フレーム表現 ($C(k)$) の対応するHOA係数シーケンスのベクトルであり、前記HOAデータ・フレーム表現 ($C(k)$) は

【数 7 0】

$$\|w(t)\|_\infty = \max_{1 \leq j \leq O} |w_j(t)| \leq 1 \quad \forall t$$

40

となるよう正規化されており、当該装置は:

・前記の正規化されたHOAデータ・フレーム表現 ($C(k)$) から、前記チャネル信号 ($y_1(k-2), \dots, y_L(k-2)$) を、動作 a)、b)、c)、すなわち

a) 前記チャネル信号における優勢音信号 ($x(t)$) を表現するために、HOA係数シーケンスの前記ベクトル $c(t)$ に混合行列 A を乗算する動作であって、混合行列 A のユークリッド・ノルムは1より大きくなく、混合行列 A は前記正規化されたHOAデータ・フレーム表現の係数シーケンスの線形結合を表わす、動作;

b) 前記チャネル信号における周囲成分 $c_{AMB}(t)$ を表現するために、前記正規化されたHOA

50

データ・フレーム表現 (C(k)) から前記優勢音信号を減算し、前記周囲成分 $c_{AMB}(t)$ の係数シーケンスの少なくとも一部を選択し、 $\|c_{AMB}(t)\|_2^2 - \|c(t)\|_2^2$ であり、結果として得られる最小周囲成分 $c_{AMB_MIN}(t)$ を、 $w_{MIN}(t) = \frac{1}{MIN} \cdot c_{AMB_MIN}(t)$ を計算することによって変換し、 $\|w_{MIN}\|_2 < 1$ であり、 $\frac{1}{MIN}$ は前記最小周囲成分 $c_{AMB_MIN}(t)$ についてのモード行列である、動作；

c) 前記HOA係数シーケンス $c(t)$ の一部を選択する動作であって、選択された係数シーケンスは、空間変換が適用される前記周囲HOA成分の係数シーケンスに関係し、前記選択された係数シーケンスの数を記述する最小次数 N_{MIN} は $N_{MIN} - 9$ である、動作；
 のうちの一つまたは複数によって形成する手段 (12、13、14) と；

・前記チャンネル信号についての前記非差分的な利得値 (2°) を表現するために必要とされる前記最低の整数ビット数 β_e を

【数71】

$$\beta_e = \lceil \log_2(\lceil \log_2(\sqrt{K_{MAX}} \cdot 0) \rceil + 1) \rceil$$

に設定する手段 (15、...、151) であって、

【数72】

$$K_{MAX} = \max_{1 \leq N \leq N_{MAX}} K(N, \Omega_1^{(N)}, \dots, \Omega_0^{(N)})$$

であり、 N は前記次数であり、 N_{MAX} は関心対象の最大次数であり、 $\Omega_1^{(N)}, \dots, \Omega_0^{(N)}$ は前記仮想スピーカ方向であり、 $0 = (N+1)^2$ はHOA係数シーケンスの数であり、 K は前記モード行列の二乗されたユークリッド・ノルム $\| \cdot \|_2^2$ と0との間の比である、手段とを含む、

装置。

〔態様3〕

前記変換された最小周囲成分に加えて、前記周囲成分 $c_{AMB}(t)$ の変換されていない周囲係数シーケンスが前記チャンネル信号 ($y_1(k-2), \dots, y_I(k-2)$) に含まれる、請求項1記載の方法または請求項2記載の装置。

〔態様4〕

前記HOAデータ・フレームのうちの個々のものの前記チャンネル信号に関連付けられた前記非差分的な利得値 (2°) がサイド情報として転送され、そのそれぞれが β_e ビットによって表現される、請求項1もしくは3記載の方法または請求項2もしくは3記載の装置。

〔態様5〕

前記最低の整数ビット数 β_e が

【数73】

$$\beta_e = \lceil \log_2(\lceil \log_2(\sqrt{K_{MAX}} \cdot 0) \rceil + e_{MAX} + 1) \rceil$$

に設定され、 $e_{MAX} > 0$ は利得制御 (15、151) 前のチャンネル信号のサンプル値の振幅が小さすぎる場合に前記ビット数 β_e を増すはたらきをする、

請求項1、3および4のうちいずれか一項記載の方法または請求項2ないし4のうちいずれか一項記載の装置。

〔態様6〕

$K_{MAX} = 1.5$ である、請求項1および3ないし5のうちいずれか一項記載の方法または請求項2ないし5のうちいずれか一項記載の装置。

〔態様7〕

前記混合行列 A が、モノラル優勢音信号の方向分布を表わすすべてのベクトルから形成されるモード行列のムーア・ペンローズの擬似逆行列を取ることによって、もとのHOA表現と優勢音信号のものとの間の残差のユークリッド・ノルムを最小にするよう決定される、請求項1および3ないし6のうちいずれか一項記載の方法または請求項2ないし6のうちいずれか一項記載の装置。

〔態様8〕

10

20

30

40

50

前記0個の仮想スピーカ信号の位置が θ_e の計算のために想定されたものと一致せず、
 ・これらの仮想スピーカ位置についてのモード行列 Ψ を計算し(51)；
 ・このモード行列のユークリッド・ノルム $\|\Psi\|_2$ を計算し(52)；
 ・前記正規化における最大の許容される振幅1を置き換える最大許容される振幅値
 【数74】

$$\gamma = \min\left(1, \frac{\sqrt{0} \cdot \sqrt{K_{MAX,DES}}}{\|\Psi\|_2}\right)$$

を計算する(53)ことを含み、
 【数75】

$$K_{MAX,DES} = \max_{1 \leq N \leq N_{MAX,DES}} K(N, \Omega_{DES,1}^{(N)}, \dots, \Omega_{DES,0}^{(N)})$$

であり、Nは前記次数であり、 $0 = (N+1)^2$ はHOA係数シーケンスの数であり、Kは、前記モード行列の二乗されたユークリッド・ノルムと0との比であり、 $N_{MAX,DES}$ は関心対象の次数であり $\Omega_{DES,1}^{(N)}, \dots, \Omega_{DES,1}^{(N)}$ は各次数について、前記HOAデータ・フレーム表現(C(k))の前記圧縮の実装のために想定された仮想スピーカ方向であり、よって θ_e は、前記非差分的な利得値の底2に対する指数(e)を符号化するために

【数76】

$$\beta_e = \lceil \log_2(\lceil \log_2(\sqrt{K_{MAX,DES}} \cdot 0) \rceil + 1) \rceil$$

によって選ばれたものである、
 請求項1および3ないし7のうちいずれか一項記載の方法または請求項2ないし7のうちいずれか一項記載の装置。

〔態様9〕

非差分的な利得値を含む符号化されたHOAデータ表現

【数77】

(\tilde{B})

であって、これらの非差分的な利得値を表現するために必要とされる最低の整数ビット数(β_e)が請求項1および3ないし8のうちいずれか一項記載の方法に従って決定されている、符号化されたHOAデータ表現。

〔態様10〕

請求項9記載の符号化されたHOAデータ・フレーム表現を含むまたは記憶しているまたは記録している記憶媒体。

〔態様11〕

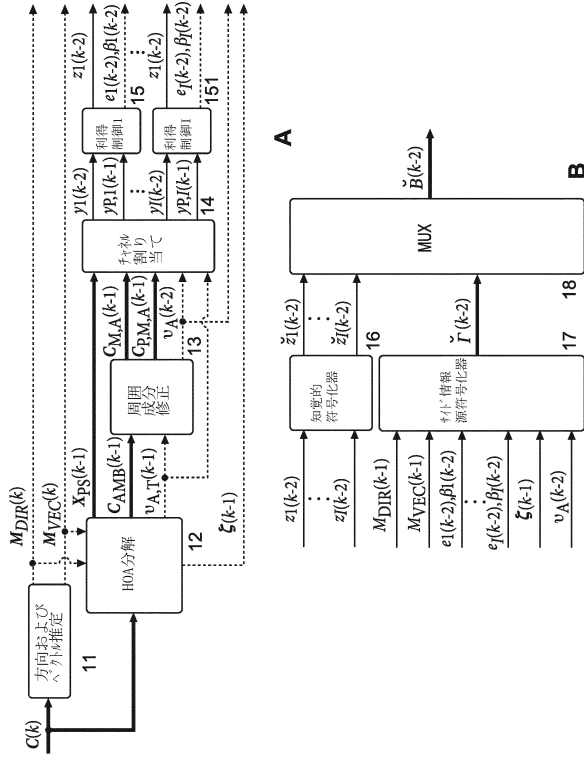
コンピュータ上で実行されたときに請求項1および3ないし8のうちいずれか一項記載の方法を実行する命令を含むコンピュータ・プログラム・プロダクト。

10

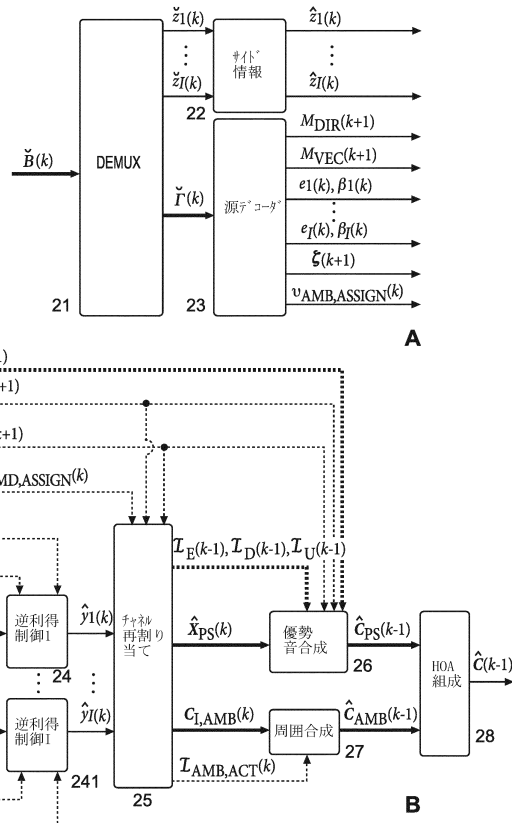
20

30

【 図 1 】



【 図 2 】



【 図 3 】

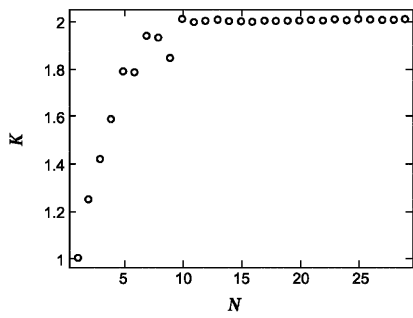


Fig. 3

【 図 4 】

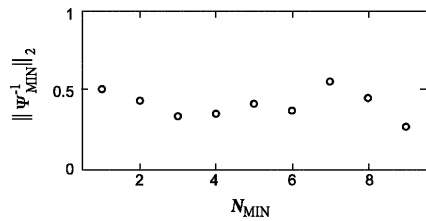
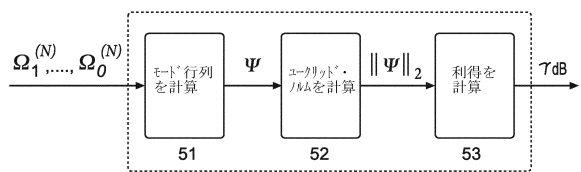
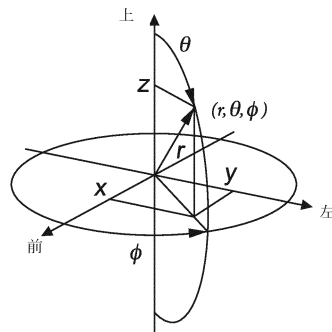


Fig. 4

【 図 5 】



【 図 6 】



フロントページの続き

- (72)発明者 コルドン, スヴェン
ドイツ連邦共和国, 3 1 5 1 5 ヴンストルフ, ミューレンカンブシュトラーセ 5 0 アー
(72)発明者 クルーガー, アレクサンダー
ドイツ連邦共和国, 3 0 6 5 5 ハノーヴァー, フォイエルバッハシュトラーセ 1 6

審査官 山下 剛史

- (56)参考文献 特開2012-133366(JP, A)
国際公開第2013/171083(WO, A1)
国際公開第2014/090660(WO, A1)
国際公開第2014/012944(WO, A1)

- (58)調査した分野(Int.Cl., DB名)
G10L 19/00 - 19/26
H04S 3/00 - 5/02