

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6205493号
(P6205493)

(45) 発行日 平成29年9月27日(2017.9.27)

(24) 登録日 平成29年9月8日(2017.9.8)

(51) Int.Cl.

F I

G 1 O K 15/00 (2006.01)

G 1 O K 15/00 M

G 1 O L 19/008 (2013.01)

G 1 O L 19/008

H O 4 S 1/00 (2006.01)

H O 4 S 1/00 5 0 0

H O 4 S 7/00 (2006.01)

H O 4 S 7/00 3 2 0

H O 4 R 3/00 (2006.01)

H O 4 R 3/00 3 1 0

請求項の数 20 (全 38 頁)

(21) 出願番号 特願2016-528429 (P2016-528429)
 (86) (22) 出願日 平成26年7月16日(2014.7.16)
 (65) 公表番号 特表2016-532149 (P2016-532149A)
 (43) 公表日 平成28年10月13日(2016.10.13)
 (86) 国際出願番号 PCT/EP2014/065227
 (87) 国際公開番号 W02015/010983
 (87) 国際公開日 平成27年1月29日(2015.1.29)
 審査請求日 平成28年3月8日(2016.3.8)
 (31) 優先権主張番号 13177362.4
 (32) 優先日 平成25年7月22日(2013.7.22)
 (33) 優先権主張国 欧州特許庁(EP)
 (31) 優先権主張番号 13189230.9
 (32) 優先日 平成25年10月18日(2013.10.18)
 (33) 優先権主張国 欧州特許庁(EP)

(73) 特許権者 500341779
 フラウンホーファー・ゲゼルシャフト・ツ
 ール・フェルデルング・デル・アンゲヴァ
 ンテン・フォルシュング・アインゲトラ
 ゲネル・フェライン
 ドイツ連邦共和国, 80686 ミュンヘ
 ン, ハンザシュトラッセ 27 ツェー
 (74) 代理人 100205981
 弁理士 野口 大輔
 (74) 代理人 100085464
 弁理士 野口 繁雄
 (72) 発明者 ジモーネ・フュグ
 ドイツ連邦共和国 90652 カルヒロ
 イト ヴァイスガーセ 24

最終頁に続く

(54) 【発明の名称】 室内インパルス応答に従ってオーディオ信号を処理するための方法、信号処理装置、オーディオ符号化器、オーディオ復号器、及びバイノーラルレンダラ

(57) 【特許請求の範囲】

【請求項 1】

室内インパルス応答(300、804)に従ってオーディオ信号(400、802)を処理するための方法であって、該方法は、

前記室内インパルス応答(300、804)の早期部分(301、302)及び後期残響(304)を用いて前記オーディオ信号(400、802)を別個に処理すること(402、404、406、422、424、812、816a、816b)と、

前記室内インパルス応答(300、804)の前記早期部分(414、428、814)によって処理された前記オーディオ信号と残響付与済み信号(416、430、826a、826b)とを組み合わせること(418、432、828)と、を含み、

相関の大きさが一度閾値に達すると、前記室内インパルス応答(300、804)における前記早期部分(301、302)から前記後期残響(304)までの遷移が決定され、

前記相関の大きさは、前記室内インパルス応答(300、804)に関して、初期状態を含む音響エネルギーの減衰の類似度、及び所定の周波数範囲にわたる前記初期状態に続く時点において始まる音響エネルギーの減衰の類似度を記述し、

前記閾値は前記室内インパルス応答(300、804)の前記早期部分(301、302)における早期反射(302、602)のうちの選択された1つについての相関の大きさに応じて設定され、

前記早期反射のうちの前記選択された1つは最初の反射(602)である方法。

10

20

【請求項 2】

前記遷移を決定することは、

前記室内インパルス応答（300、804）に基づいて音響エネルギーの分布を決定すること（500）と、

前記決定された分布の複数の部分について、前記決定された分布のそれぞれの部分における前記音響エネルギーと初期状態における前記音響エネルギーとの間の相関を示す複数の相関の大きさを決定すること（502）と、を含む請求項 1 に記載の方法。

【請求項 3】

前記分布を決定することは前記音響エネルギーの時間周波数分布を決定することを含み、

前記分布の一部分は所定長の時間ブロックを含み、前記初期状態は前記時間周波数分布の前記複数の時間ブロックのうちの第 1 の時間ブロックによって規定される請求項 2 に記載の方法。

【請求項 4】

前記分布を決定することは、前記室内インパルス応答（300、804）からエネルギー減衰緩和（EDR）を計算することを含む請求項 2 又は 3 に記載の方法。

【請求項 5】

前記 EDR は以下のように計算される請求項 4 に記載の方法。

$$E(t, \omega) = \left| \int_t^{\infty} h(\tau) e^{-j\omega\tau} d\tau \right|^2$$

ただし、

$E(t, \omega)$ = エネルギー減衰緩和、

$h(\omega)$ = 室内インパルス応答（300、804）、

$t = 2/f$

である。

【請求項 6】

前記室内インパルス応答（300、804）は所定の有効長を有し、前記時間周波数分布を決定することは、前記室内インパルス応答の前記有効長に対応する長さを有するウィンドウを使用して前記室内インパルス応答（300、804）の FFT スペクトルを計算することを含む請求項 3 に記載の方法。

【請求項 7】

前記初期状態における前記音響エネルギーは、前記室内インパルス応答（300、804）の前記有効長全体を取り、前記 FFT スペクトルを計算して絶対値の二乗をすることによって求められ、

時間ブロックの前記音響エネルギーは、前記時間ブロックと関連付けられる時間だけ前記ウィンドウをシフトさせ、ウィンドウイングされたサンプルを前記有効長までゼロ詰めし、前記 FFT を計算して絶対値の二乗をすることによって求められる請求項 6 に記載の方法。

【請求項 8】

前記相関の大きさは以下のように計算される請求項 1 から 7 のいずれか一項に記載の方法。

$$\rho(t) = \frac{\sum_{\omega} (E(1, \omega) - \bar{E}(1, \omega)) \cdot \sum_{\omega} (E(t, \omega) - \bar{E}(t, \omega))}{\sqrt{\sum_{\omega} (E(1, \omega) - \bar{E}(1, \omega))^2} \cdot \sqrt{\sum_{\omega} (E(t, \omega) - \bar{E}(t, \omega))^2}}$$

ここで、

$\rho(t)$ は相関の大きさ、

10

20

30

40

50

$E(1, \quad)$ は周波数 f における全周波数範囲エネルギー減衰緩和、

$$\bar{E}(1, \omega)$$

は最初の全範囲エネルギー減衰緩和のすべての周波数にわたる平均値、

$E(t, \quad)$ は時刻 t から始まる周波数 f におけるエネルギー減衰緩和、

$$\bar{E}(t, \omega)$$

10

は時刻 t から始まる全範囲エネルギー減衰緩和のすべての周波数にわたる平均値、

$$= 2 \quad f$$

である。

【請求項 9】

前記閾値は、一定値及び前記早期反射 (302、602) のうちの選択された 1 つの前記相関の大きさに基づいて決定される請求項 1 から 8 のいずれか一項に記載の方法。

【請求項 10】

前記一定値は $1/e$ であり、前記閾値は以下のように定義される請求項 9 に記載の方法。

$$(t) = c \cdot (t_F)$$

20

ここで、

(t_F) は前記早期反射 (302、602) のうちの選択された 1 つの相関の大きさ、
 t_F は前記室内インパルス応答 (300、804) の直接的な音成分である直接音 (300) の後の前記早期反射 (302、602) のうちの選択された 1 つが衝突する時間インデックス、

c は $1/e$ に基づく前記一定値であり、 e はオイラー数である。

【請求項 11】

コンピュータ上で実行されると、請求項 1 から 10 のいずれか一項に記載の方法を実行する命令を記憶しているコンピュータ読取り可能な媒体を含む非有形コンピュータプログラム製品。

30

【請求項 12】

オーディオ信号 (400、802) を受信するための入力と、

請求項 1 から 10 のいずれか一項に記載の方法に従って、室内インパルス応答 (300、804) に従って前記オーディオ信号 (400、802) を処理するように構成されているプロセッサと、

前記オーディオ信号の処理済み早期部分 (414、428、814) と、前記残響付与済み信号 (416、430、826a、826b) とを組み合わせ (418、432、828)、出力オーディオ信号にするための出力と、を備えている信号処理装置。

【請求項 13】

前記室内インパルス応答 (300、804) の前記早期部分 (301、302、602) に従って前記受信オーディオ信号 (400、802) を処理するための早期部分プロセッサ (402、404、422、812) と、

40

前記室内インパルス応答 (300、804) の前記後期残響 (304) に従って前記オーディオ信号 (400、802) を処理するための後期残響プロセッサ (406、424、816a、816b) と、を備えている請求項 12 に記載の信号処理装置。

【請求項 14】

オーディオ信号を符号化するためのオーディオ符号化器であって、

該オーディオ符号化器は、請求項 1 から 10 のいずれか一項に記載の方法に従って、符号化されるべきオーディオ信号を室内インパルス応答 (300、804) に従って処理するように構成されているオーディオ符号化器。

50

【請求項 15】

請求項 12 又は 13 に記載の信号処理装置を有する、請求項 14 に記載のオーディオ符号化器。

【請求項 16】

符号化済みオーディオ信号を復号するためのオーディオ復号器であって、

該オーディオ復号器は、請求項 1 から 10 のいずれか一項に記載の方法に従って、復号されたオーディオ信号（400、802）を、室内インパルス応答（300、804）に従って処理するように構成されているオーディオ復号器。

【請求項 17】

請求項 12 又は 13 に記載の信号処理装置を有する、請求項 16 に記載のオーディオ復号器。

10

【請求項 18】

前記復号されたオーディオ信号（400、802）を受信し、前記室内インパルス応答（300、804）に基づいて出力信号をレンダリングするように構成されているレンダラを含む請求項 16 に記載のオーディオ復号器。

【請求項 19】

前記レンダラはバイノーラルレンダラを含む請求項 18 に記載のオーディオ復号器。

【請求項 20】

請求項 12 又は 13 に記載の信号処理装置を含むバイノーラルレンダラ。

【発明の詳細な説明】

20

【技術分野】

【0001】

本発明は、オーディオ符号化／復号の分野に関し、特に、空間オーディオコード化及び空間オーディオオブジェクトコード化、例えば、3Dオーディオコーデックシステムの分野に関する。本発明の実施形態は、室内インパルス応答に従ってオーディオ信号を処理し、そのような室内インパルス応答において早期反射から後期残響までの遷移を決定するための手法に関する。

【背景技術】

【0002】

空間オーディオコード化ツールが当該技術分野において周知であり、例えば、MPEGサラウンド規格において標準化されている。空間オーディオコード化は、複数の原入力、例えば5つ又は7つの入力チャンネルから開始する。それらの原入力は、それらの再生設定における配置によって、例えば、左チャンネル、中央チャンネル、右チャンネル、左サラウンドチャンネル、右サラウンドチャンネル及び低周波強調チャンネルとして識別される。空間オーディオ符号化器が、原チャンネルから1つ又は複数のダウンミックスチャンネルを導出することができ、加えて、チャンネルコヒーレンス値のチャンネル間レベル差、チャンネル間位相差、チャンネル間時間差などのような、空間手掛かりに関係するパラメータデータを導出することができる。1つ又は複数のダウンミックスチャンネルは、最終的に、原入力チャンネルの近似バージョンである出力チャンネルを得るために、ダウンミックスチャンネルを復号するための空間オーディオ復号器に対して空間手掛かりを示すパラメータサイド情報、及び、関連するパラメータデータとともに送信される。出力設定におけるチャンネルの配置は固定することができ、例えば、5.1フォーマット、7.1フォーマットなどとすることができる。

30

40

【0003】

また、空間オーディオオブジェクトコード化ツールが当該技術分野において周知であり、例えば、MPEG SAOCサラウンド規格（SAOC = spatial audio object coding：空間オーディオオブジェクトコード化）において標準化されている。原チャンネルから開始する空間オーディオコード化とは対照的に、空間オーディオオブジェクトコード化は、自動的に特定のレンダリング再生装備に専用になるのではないオーディオオブジェクトから開始する。むしろ、再生シーンにおけるオーディオオブジェクトの配置は柔軟であり、ユーザによって、例えば、空間オーディオオブジェクトコード化復号器に特定のレンダ

50

リング情報を入力することによって設定することができる。あるいは又はさらに、レンダリング情報は、付加的なサイド情報又はメタデータとして送信することができ、レンダリング情報は、（例えば、経時的に）再生装置におけるいずれの位置に特定のオーディオオブジェクトが配置されるべきかの情報を含むことができる。特定のデータ圧縮を得るためにS A O C符号化器を使用して複数のオーディオオブジェクトが符号化される。S A O C符号化器は、特定のダウンミックス情報に従ってオブジェクトをダウンミックスすることによって、入力オブジェクトから1つ又は複数のトランスポートチャネルを計算する。さらに、S A O C符号化器は、オブジェクトレベル差（O L D : object level difference）、オブジェクトコヒーレンス値のようなオブジェクト間手掛かりを表すパラメータサイド情報を計算する。S A C（S A C = Spatial Audio Coding : 空間オーディオコード化）におけるように、オブジェクト間パラメータデータは、個々の時間/周波数タイルについて計算される。オーディオ信号の特定のフレーム（例えば、1 0 2 4又は2 0 4 8サンプル）について、複数の帯域（例えば、2 4、3 2、又は6 4帯域）が考慮され、それによって、各フレーム及び各周波数帯域についてパラメータデータが提供される。例えば、オーディオ片が2 0個のフレームを有し、各フレームが3 2個の周波数帯域に区分化される時、時間/周波数タイルの数は6 4 0である。

10

【0 0 0 4】

3 Dオーディオシステムにおいて、オーディオ信号が特定の部屋の中で聴かれているかのような、オーディオ信号の空間的印象を与えることが所望されることがある。そのような状況において、特定の部屋の室内インパルス応答が、例えばその測定に基づいて生じ、オーディオ信号を聴き手に提示するときにオーディオ信号を処理するのに使用される。そのような提示において後期残響から分離して直接音と早期反射を処理することが望ましいことがある。これは、早期反射がどこで終わり、後期残響がどこで始まるかを決定することを必要とする。

20

【先行技術文献】

【非特許文献】

【0 0 0 5】

【非特許文献1】[1] T. Hidaka et al: "A new definition of boundary point between early reflections and late reverberation in room impulse responses". Forum Acusticum, 2005.

30

【非特許文献2】[2] Jot et al: "Analysis and synthesis of room reverberation based on a statistical time frequency model".

【非特許文献3】[3] J.S. Abel, P. Huang: "A Simple, Robust Measure of Reverberation Echo Density". AES Convention, San Francisco, 2006.

【非特許文献4】[4] R. Stewart, M. Sandler: "Statistical Measures of Early Reflections of Room Impulse Responses". DAFx, 2007.

【非特許文献5】[5] Reilly et al: "Using Auralisation for Creating Animated 3-D Sound Fields Across Multiple Speakers". AES Convention, New York, 1995.

【非特許文献6】[6] Usher, J.: "An improved method to determine the onset timing of reflections in an acoustic impulse response". Journal of the Acoustical Society of America, (2010, volume 127) band 4, p. 172-177.

40

【非特許文献7】[7] Masri, P.: "Computer Modelling of Sound for Transformation and Synthesis of Musical Signals". PhD thesis, University of Bristol, 1996.

【発明の概要】

【発明が解決しようとする課題】

【0 0 0 6】

本発明の目的は、室内インパルス応答に従ってオーディオ信号を処理するための改善された手法を提供することである。

【課題を解決するための手段】

【0 0 0 7】

50

この目的は、請求項 1 に記載の方法、請求項 16 に記載の信号処理装置、請求項 18 に記載のオーディオ符号化器、請求項 20 に記載のオーディオ復号器、及び請求項 24 に記載のバイノーラルレンダラによって達成される。

【0008】

本発明は、従来の手法においては、最初の反射が発生又は衝突さえしていないうちに、遷移の発生を判定するのに使用される相関がすでに閾値に達しているために、早期反射から後期残響までの遷移の決定が早過ぎる状況があるということにおいて問題があるという本発明者の所見に基づく。これらの所見に基づいて、また、最初の反射ははっきりと区別でき確実に後期拡散残響ではあり得ないため、遷移時間は最初の反射の到来時刻よりも大きくなかなければならないことが分かっているために、本発明者らは、固定閾値を使用することを回避することが必要であることを見出しており、むしろ、本発明の手法によれば、閾値は早期反射のうちの 1 つの衝突時刻における相関に依存するように規定される。これによって、最初の反射が常に遷移時間の前に位置することが保証される。

10

【0009】

(1) 本発明は室内インパルス応答に従ってオーディオ信号を処理するための方法を提供し、この方法は、室内インパルス応答の早期部分及び後期残響を用いてオーディオ信号を別個に処理することと、室内インパルス応答の早期部分によって処理されたオーディオ信号と残響付と済み信号とを組み合わせることを含む。室内インパルス応答における早期部分から後期残響までの遷移が、閾値に達する相関の大きさによって決定され、その閾値は室内インパルス応答の早期部分における早期反射のうちの選択された 1 つについての相関の大きさに応じて設定される。

20

【0010】

本発明の手法は、ロバストな遷移点に基づいてオーディオ信号の処理を改善することを可能にするため、有利である。本発明の手法は、部屋、バイノーラル手法が使用されているか否か、及び、入射角とは無関係である。従来技術の手法と比較して、本発明の手法は、バイノーラルインパルス応答の方位角、及び、直接音の振幅と最初の衝突反射との間の関係に強く依存しないため、さらに有利である。

【0011】

(2) 実施形態によれば、相関の大きさは、室内インパルス応答に関して、初期状態を含む音響エネルギーの減衰、及び、所定の周波数範囲にわたって、初期状態に続く任意の時点において始まる音響エネルギーの減衰の類似度を記述する。

30

【0012】

(3) 実施形態によれば、遷移を決定することは、室内インパルス応答に基づいて音響エネルギーの分布を決定することと、決定された分布の複数の部分について、決定された分布のそれぞれの部分における音響エネルギーと初期状態における音響エネルギーとの間の相関を示す複数の相関の大きさを決定することを含む。

【0013】

(4) 実施形態によれば、分布を決定することは、音響エネルギーの時間周波数分布を決定することを含み、分布の一部分は所定長の時間ブロックを含み、初期状態は時間周波数分布の複数の時間ブロックのうちの第 1 の時間ブロックによって規定される。

40

【0014】

これは、異なる複数の時点における周波数分布を分析することを可能にし、それによって周波数分布の特性の経時的な表現をもたらすため、有利である。

【0015】

(5) 実施形態によれば、分布を決定することは、室内インパルス応答からエネルギー減衰緩和 (EDR : energy decay relief) を計算することを含み、EDR は以下のように計算される。

$$E(t, \omega) = \left| \int_t^\infty h(\tau) e^{-j\omega\tau} d\tau \right|^2$$

ここで、

$E(t, \omega)$ はエネルギー減衰緩和、

$h(\omega)$ は室内インパルス応答、

$\omega = 2\pi f$ である。

【0016】

(6) 実施形態によれば、室内インパルス応答は所定の有効長を有し、時間周波数分布を決定することは、室内インパルス応答の有効長に対応する長さを有するウィンドウを使用して室内インパルス応答のFFTスペクトルを計算することを含む。

10

【0017】

FFT/DFITが明確に定義されており、スペクトル値を計算するための有効なアルゴリズムがあるため、これは有利である。ウィンドウ内の値が分かっている場合、複雑でない方法でFFT/DFITを計算することができる。

【0018】

(7) 実施形態によれば、初期状態における音響エネルギーは、室内インパルス応答の有効長全体を取り、FFTスペクトルを計算して絶対値の二乗を取ることによって求められ、時間ブロックの音響エネルギーは、時間ブロックと関連付けられる時間だけウィンドウをシフトさせ、ウィンドウイングされたサンプルを有効長までゼロ詰めし、FFTを計算して絶対値の二乗を取ることによって求められる。

20

【0019】

EDRの狭い帯域の計算に追加のフィルタバンクなどが必要とされず、ウィンドウをシフトすることだけが必要とされるため、これは有利である。

【0020】

(8) 実施形態によれば、相関の大きさは、初期状態を含む音響エネルギーの減衰、及び、初期状態に続く任意の時点において始まる音響エネルギーの減衰の類似度を記述する相関の大きさである。相関の大きさは以下のように計算することができる。

30

$$\rho(t) = \frac{\sum_{\omega} (E(1, \omega) - \bar{E}(1, \omega)) \cdot \sum_{\omega} (E(t, \omega) - \bar{E}(t, \omega))}{\sqrt{\sum_{\omega} (E(1, \omega) - \bar{E}(1, \omega))^2} \cdot \sqrt{\sum_{\omega} (E(t, \omega) - \bar{E}(t, \omega))^2}}$$

ここで、

$\rho(t)$ は相関の大きさ、

$E(1, \omega)$ は周波数 f における全周波数範囲エネルギー減衰緩和、

$$\bar{E}(1, \omega)$$

40

は最初全範囲エネルギー減衰緩和のすべての周波数にわたる平均値、

$E(t, \omega)$ は時刻 t から始まる周波数 f におけるエネルギー減衰緩和、

$$\bar{E}(t, \omega)$$

は時刻 t から始まる全範囲エネルギー減衰緩和のすべての周波数にわたる平均値、

$\omega = 2\pi f$

である。

【0021】

50

この式は、周知のピアソンの相関係数（ピアソンの積率相関：Pearson's Product-Moment correlation）を指すため、これは有利である。相関係数はE D Rから直接計算することができる。

【0022】

（9）実施形態によれば、閾値は、一定値と早期反射のうちの選択された1つの相関の大きさに基づいて決定される。閾値は以下のように定義することができる。

$$(t) = c \cdot (t_F)$$

ここで、

(t_F) は早期反射のうちの選択された1つの相関の大きさ、

t_F は直接音の後の早期反射のうちの選択された1つが衝突する時間インデックス、

c は $1/e$ に基づく一定値、 e はオイラー数である。

10

【0023】

閾値は一定ではなく、相関が閾値を下回って収まるのが早すぎないことを保証するために、選択された早期反射に依存するので有利である。

【0024】

（10）実施形態によれば、早期反射のうちの選択された1つの時間は、例えば、移動尖度演算子（running kurtosis operator）、閾値検出又はアタック検出によって求められる。

【0025】

これは、インパルス応答の時間ドメインサンプルから反射の衝突の時間を直接かつ自動的に計算することができるため、有利である。

20

【0026】

（11）実施形態によれば、早期反射のうちの選択された1つは、直接音に続く最初の反射である。

【0027】

（12）本発明は、オーディオ信号を受信するための入力と、本発明の方法に従って、室内インパルス応答に従って受信オーディオ信号を処理するように構成又はプログラムされているプロセッサと、受信オーディオ信号の処理済み早期部分及び残響付与済み信号を組み合わせる出力オーディオ信号にするための出力とを備える信号処理装置を提供する。この信号処理装置は、室内インパルス応答の早期部分に従って受信オーディオ信号を処理するための早期部分プロセッサと、室内インパルス応答の後期残響に従って受信オーディオ信号を処理するための後期残響プロセッサとを備えることができる。

30

【0028】

（13）本発明は、オーディオ信号を符号化するためのオーディオ符号化器を提供する。このオーディオ符号化器は、本発明の方法に従って、符号化されるべきオーディオ信号を室内インパルス応答に従って処理するように構成又はプログラムされている。このオーディオ符号化器は本発明の信号処理装置を含むことができる。

【0029】

（14）本発明は、符号化済みオーディオ信号を復号するためのオーディオ復号器を提供する。このオーディオ復号器は、本発明の方法に従って、復号済みオーディオ信号を室内インパルス応答に従って処理するように構成又はプログラムされている。このオーディオ復号器は本発明の信号処理装置を含むことができる。このオーディオ復号器は、復号済みオーディオ信号を受信し、室内インパルス応答に基づいて出力信号をレンダリングするように構成又はプログラムされている、バイノーラルレンダラのようなレンダラを含むことができる。

40

【0030】

（15）本発明は、本発明の信号処理装置を含むバイノーラルレンダラを提供する。

【図面の簡単な説明】

【0031】

【図1】3Dオーディオシステムの3Dオーディオ符号化器の概観図である。

50

【図 2】3D オーディオシステムの 3D オーディオ復号器の概観図である。

【図 3】図 2 の 3D オーディオ復号器内で実装することができるフォーマット変換器を実装するための一例を示す図である。

【図 4】図 2 の 3D オーディオ復号器内で実装することができるバイノーラルレンダラの一実施形態を示す図である。

【図 5】室内インパルス応答 $h(t)$ の一例を示す図である。

【図 6 A】本発明の一実施形態による、室内インパルスの早期部分及び後期残響を有するオーディオ信号を別個に処理するための（例えば、図 4 のバイノーラルレンダラ内の）信号処理装置のブロック図である。

【図 6 B】本発明のさらなる実施形態による、室内インパルスの早期部分及び後期残響を有するオーディオ信号を別個に処理するための（例えば、図 4 のバイノーラルレンダラ内の）別の信号処理装置のブロック図である。

【図 7】本発明の一実施形態による、室内インパルス応答の早期反射と後期残響との間の遷移時間を決定するための手法の流れ図である。

【図 8】FFT ベースの手法に従って決定されたインパルス応答について達成されるエネルギー減衰緩和（EDR）を示す図である。

【図 9】本発明の一実施形態による遷移時間決定を示す図である。

【図 10】従来の手法を使用して決定されている、測定されているバイノーラル室内インパルス応答の左チャンネル及び右チャンネルの遷移時間を示す図である。

【図 11】本発明の手法を使用して決定されている、測定されているバイノーラル室内インパルス応答の左チャンネル及び右チャンネルの遷移時間を示す図である。

【図 12】本発明の一実施形態による、バイノーラルレンダラにおけるオーディオ信号のバイノーラル処理を概略的に示す図である。

【図 13】本発明の一実施形態による、図 12 のバイノーラルレンダラの周波数ドメイン残響付与装置における処理を概略的に示す図である。

【発明を実施するための形態】

【0032】

本発明の実施形態を添付の図面に関連して説明する。

【0033】

室内インパルス応答に従ってオーディオ信号を処理し、室内インパルス応答において、早期反射から後期残響までの遷移を決定するための本発明の手法の実施形態を説明する。以下の説明は、本発明の手法を実施することができる 3D オーディオコーデックシステムのシステム概観によって開始する。

【0034】

図 1 及び図 2 は、実施形態による 3D オーディオシステムのアルゴリズムブロックを示す。より詳細には、図 1 は 3D オーディオ符号化器 100 の概観を示す。オーディオ符号化器 100 は、任意に設けることのできるプリレンダラ/混合器回路 102 において入力信号、より詳細には、オーディオ符号化器 100 に提供する複数の入力チャンネルを受信する。それらの入力信号は、複数のチャンネル信号 104、複数のオブジェクト信号 106 及び対応するオブジェクトメタデータ 108 である。プリレンダラ/混合器 102 によって処理されたオブジェクト信号 106（信号 110 参照）は、SAOC 符号化器 112（SAOC = Spatial Audio Object Coding: 空間オーディオオブジェクトコード化）に提供することができる。SAOC 符号化器 112 は、USAC 符号化器 116（USAC = Unified Speech and Audio Coding: 統合音声符号化）に提供される SAOC トランスポートチャンネル 114 を生成する。加えて、信号 SAOC-SI 118（SAOC-SI = SAOC side information: SAOC サイド情報）も、USAC 符号化器 116 に提供される。USAC 符号化器 116 はさらに、プリレンダラ/混合器からオブジェクト信号 120 を直接受信するとともに、チャンネル信号及びプリレンダリング済みオブジェクト信号 122 を受信する。オブジェクトメタデータ情報 108 は、圧縮オブジェクトメタデータ情報 126 を USAC 符号化器に提供する OAM 符号化器 124（OAM = object metadat

10

20

30

40

50

a: オブジェクトメタデータ) に与えられる。USAC符号化器116は、上述した入力信号に基づいて、128において示すような、圧縮出力信号mp4を生成する。

【0035】

図2は、3Dオーディオシステムの3Dオーディオ復号器200の概観を示す。図1のオーディオ符号化器100によって生成される符号化信号128(mp4)が、オーディオ復号器200、より詳細にはUSAC復号器202において受信される。USAC復号器202は、受信信号128を復号してチャンネル信号204、プリレンダリング済みオブジェクト信号206、オブジェクト信号208、及びSAOCトランスポートチャンネル信号210にする。さらに、圧縮オブジェクトメタデータ情報212及び信号SAOC-SI214が、USAC復号器202によって出力される。オブジェクト信号208は、レンダリング済みオブジェクト信号218を出力するオブジェクトレンダラ216に提供される。SAOCトランスポートチャンネル信号210は、レンダリング済みオブジェクト信号222を出力するSAOC復号器220に供給される。圧縮オブジェクトメタデータ情報212はOAM復号器224に供給され、OAM復号器224はレンダリング済みオブジェクト信号218及びレンダリング済みオブジェクト信号222を生成するために、オブジェクトレンダラ216及びSAOC復号器220にそれぞれの制御信号を出力する。この復号器は、図2に示すように、チャンネル信号228を出力するために入力信号204、206、218及び222を受信する混合器226をさらに備える。チャンネル信号は、230において示すように、スピーカ、例えば、32チャンネルスピーカに直接出力することができる。信号228はフォーマット変換回路232に提供することができ、フォーマット変換回路232はチャンネル信号228が変換されるべき方法を示す再生レイアウト信号を制御入力として受信する。図2に示す実施形態において、フォーマット変換回路232での変換は、234で示されるように、信号が5.1スピーカシステムに提供できるように行われるべきであると仮定される。また、チャンネル信号228は、238において示されるように、例えば、ヘッドホンのための2つの出力信号を生成するバイノーラルレンダラ236にも提供できる。

【0036】

本発明の実施形態において、図1及び図2に示す符号化/復号システムは、チャンネル及びオブジェクト信号(信号104及び106参照)のコード化のためのMPEG-DUSACコーデックに基づいている。大量のオブジェクトをコード化するための効率を増大させるために、MPEG SAOC技術を使用できる。3つのタイプのレンダラが、オブジェクトをチャンネルにレンダリングするタスク、チャンネルをヘッドホンにレンダリングするタスク、又は、チャンネルを異なるスピーカ施設にレンダリングするタスク(図2、参照符号230、234及び238参照)を実施することができる。オブジェクト信号が明示的に送信されるか、又はSAOCを使用してパラメトリックに符号化されると、対応するオブジェクトメタデータ情報108が圧縮され(信号126参照)、3Dオーディオビットストリーム128に多重化される。

【0037】

図1及び図2に示す全体的な3Dオーディオシステムのアルゴリズムブロックを下記にさらに詳細に説明する。

【0038】

プリレンダラ/混合器102は、チャンネル+オブジェクト入力シーンを、符号化前にチャンネルシーンに変換するために任意に設けることができる。機能的には、これは後述するオブジェクトレンダラ/混合器と同一である。符号化器入力における決定論的(deterministic)信号エントロピーを保証するためにオブジェクトのプリレンダリングが望ましいことがある。符号化器入力は、基本的に、同時にアクティブなオブジェクト信号の数とは無関係である。オブジェクトをプリレンダリングすることによって、オブジェクトメタデータを送信する必要がなくなる。個別のオブジェクト信号が、符号化器が使用するように構成されているチャンネルレイアウトにレンダリングされる。各チャンネルについてのオブジェクトの重みが、関連するオブジェクトメタデータ(OAM)から得られる。

【 0 0 3 9 】

U S A C 符号化器 1 1 6 は、スピーカチャンネル信号、個別のオブジェクト信号、オブジェクトダウンミックス信号及びプリレンダリング済み信号のためのコアコーデックである。U S A C 符号化器 1 1 6 は、M P E G - D U S A C 技術に基づいている。U S A C 符号化器 1 1 6 は、入力チャンネル及びオブジェクト割り当ての幾何学的情報及び意味情報に基づいて、チャンネル及びオブジェクトのマッピング情報を作成することによって、上記信号のコード化を処理する。このマッピング情報は、入力チャンネル及びオブジェクトがチャンネル対要素 (C P E s : channel pair elements)、単一チャンネル要素 (S C E s : single channel elements)、低周波数効果 (L F E s : low frequency effects) 及び 4 チャンネル要素 (Q C E s : quad channel elements) のような U S A C チャンネル要素にどのようにマッピングされるかを記述し、C P E s、S C E s 及び L F E s、並びに対応する情報は復号器に送信される。S A O C データ 1 1 4、1 1 8 又はオブジェクトメタデータ 1 2 6 のようなすべての追加のペイロードが、符号化器のレート制御において考慮される。オブジェクトのコード化は、レンダラに対するレート / 歪み要件及び双方向性要件に応じて、様々な方法で可能である。実施形態によれば、以下のオブジェクトコード化変形形態が可能である。

10

【 0 0 4 0 】

・ プリレンダリング済みオブジェクト : オブジェクト信号は符号化前に 2 2 . 2 チャンネル信号にプリレンダリング及び混合される。後続のコード化チェーンは、2 2 . 2 チャンネル信号を考慮する。

20

【 0 0 4 1 】

・ 個別オブジェクト波形 : オブジェクトは単音波形として符号化器に供給される。符号化器は、チャンネル信号に加えてオブジェクトを送信するために、単一チャンネル要素 (S C E s) を使用する。復号済みオブジェクトは受信機側においてレンダリング及び混合される。圧縮オブジェクトメタデータ情報が、受信機 / レンダラに送信される。

【 0 0 4 2 】

・ パラメトリックオブジェクト波形 : オブジェクト特性及びそれらの互いに対する関係は、S A O C パラメータによって記述される。オブジェクト信号のダウンミックスは U S A C を用いてコード化される。パラメトリック情報が並行して送信される。ダウンミックスチャンネルの数は、オブジェクトの数及び全体的なデータレートに応じて選択される。圧縮オブジェクトメタデータ情報が S A O C レンダラに送信される。

30

【 0 0 4 3 】

オブジェクト信号のための S A O C 符号化器 1 1 2 及び S A O C 復号器 2 2 0 は、M P E G S A O C 技術に基づくことができる。このシステムは、より少数の送信チャンネル、及び、O L D s、I O C s (Inter Object Coherence : オブジェクト間コヒーレンス)、D M G s (DownMix Gains : ダウンミックス利得) のような追加のパラメータデータに基づいて、複数のオーディオオブジェクトを再生成、修正及びレンダリングすることが可能である。追加のパラメータデータは、すべてのオブジェクトを個々に送信するのに必要とされるよりも大幅に低いデータレートを示し、コード化を非常に効率的にする。S A O C 符号化器 1 1 2 は、入力として、単音波形としてのオブジェクト / チャンネル信号を取り込み、パラメトリック情報 (3 D オーディオビットストリーム 1 2 8 内にパケット化される) 及び S A O C トランスポートチャンネル (単一チャンネル要素を使用して符号化され、送信される) を出力する。S A O C 復号器 2 2 0 は、復号済み S A O C トランスポートチャンネル 2 1 0 及びパラメトリック情報 2 1 4 からオブジェクト / チャンネル信号を再構築し、再生レイアウト、解凍されたオブジェクトメタデータ情報、及び、任意ではあるがユーザ対話情報に基づいて出力オーディオシーンを生成する。

40

【 0 0 4 4 】

オブジェクトメタデータコーデック (O A M 符号化器 1 2 4 及び O A M 復号器 2 2 4 参照) は、各オブジェクトについて、3 D 空間におけるオブジェクトの幾何学的位置及び体積を指定する関連するメタデータが、時間及び空間におけるオブジェクト特性の量子化に

50

よって効率的にコード化されるように、設けられる。圧縮オブジェクトメタデータ c O A M 1 2 6 が、サイド情報として受信機 2 0 0 に送信される。

【 0 0 4 5 】

オブジェクトレンダラ 2 1 6 は、所定の再生フォーマットに従ってオブジェクト波形を生成するために、圧縮オブジェクトメタデータを利用する。各オブジェクトは、そのメタデータに従って特定の出力チャンネルにレンダリングされる。このブロックの出力は、部分的結果の合計からもたらされる。チャンネルベースの内容及び個別 / パラメータオブジェクトの両方が復号される場合、チャンネルベースの波形及びレンダリング済みオブジェクト波形が混合器 2 2 6 によって混合され、その後、結果としての波形 2 2 8 が出力され、又はそれらがバイノーラルレンダラ 2 3 6 もしくはスピーカレンダラモジュール 2 3 2 のような後処理モジュールに供給される。

10

【 0 0 4 6 】

バイノーラルレンダラモジュール 2 3 6 は、各入力チャンネルが仮想音源によって表わされるようにマルチチャンネルオーディオ材料のバイノーラルダウンミックスを生成する。この処理は、Q M F (Quadrature Mirror Filterbank : 直交ミラーフィルタバンク) ドメインにおいてフレームごとに行われ、バイノーラル化は測定バイノーラル室内インパルス応答に基づく。

【 0 0 4 7 】

スピーカレンダラ 2 3 2 は、送信されたチャンネル構成 2 2 8 と所望の再生フォーマットとの間で変換する。スピーカレンダラ 2 3 2 は「フォーマット変換器 (format converter) 」と呼ばれることもある。フォーマット変換器はより少数の出力チャンネルへの変換を実施する、すなわち、ダウンミックスを作成する。

20

【 0 0 4 8 】

図 3 はフォーマット変換器 2 3 2 を実装するための一例を示す。スピーカレンダラとも称されるフォーマット変換器 2 3 2 は、送信機チャンネル構成と所望の再生フォーマットとの間で変換する。フォーマット変換器 2 3 2 はより少数の出力チャンネルへの変換を実施する、すなわち、ダウンミックス (D M X) プロセス 2 4 0 を実施する。ダウンミキサ 2 4 0 は、好ましくは Q M F ドメインにおいて動作するものであり、混合器出力信号 2 2 8 を受信しスピーカ信号 2 3 4 を出力する。コントローラとも称されるコンフィギュレータ 2 4 2 を設けることができ、コンフィギュレータ 2 4 2 は、制御入力として、混合器出力レイアウト、すなわちそのために混合器出力信号 2 2 8 によって表されるデータが決定されるレイアウト、を示す信号 2 4 6、及び所望の再生レイアウトを示す信号 2 4 8 を受信する。この情報に基づいて、コントローラ 2 4 2 は、好ましくは自動的に、入力フォーマットと出力フォーマットとの所定の組み合わせのために最適化されたダウンミックス行列を生成し、これらの行列をダウンミキサ 2 4 0 に適用する。フォーマット変換器 2 3 2 は、標準的なスピーカ構成、及びスピーカ位置が非標準的なランダム構成を可能にする。

30

【 0 0 4 9 】

図 4 は、図 2 のバイノーラルレンダラ 2 3 6 の一実施形態を示す。バイノーラルレンダラモジュールは、マルチチャンネルオーディオ信号のバイノーラルダウンミックスを提供することができる。バイノーラル化は、測定されたバイノーラル室内インパルス応答に基づくことができる。室内インパルス応答は、実際の部屋の音響特性の「指紋」と考えることができる。室内インパルス応答は測定され、かつ記憶され、この「指紋」によって任意の音響信号を提供することができ、それによって聴き手において、室内インパルス応答と関連付けられた部屋の音響特性のシミュレーションが可能になる。バイノーラルレンダラ 2 3 6 は、頭部伝達関数 (head related transfer functions) 又はバイノーラル室内インパルス応答 (B R I R : binaural room impulse responses) を使用して出力チャンネルを 2 つのバイノーラルチャンネルにレンダリングするようにプログラム又は構成することができる。例えば、モバイルデバイスについて、そのようなモバイルデバイスに取り付けられているヘッドホン又はスピーカに対するバイノーラルレンダリングが所望される。そのようなモバイルデバイスにおいて、制約に起因して、復号器及びレンダリングの複雑度を制

40

50

限することが必要なことがある。そのような処理状況における非相関を省くことに加えて、最初に、ダウンミキサ250を使用して、中間ダウンミックス信号252、すなわちより少数の出力チャネル、へのダウンミックスを実施することが好ましい場合がある。これにより、実際のバイノーラル変換器254に対する入力チャネルがより少数になる。例えば、22.2チャンネル材料がダウンミキサ250によって5.1中間ダウンミックスにダウンミックスされてもよく、又は、あるいは中間ダウンミックスは、ある種の「ショートカット」モードにおいて図2のSAOC復号器220によって直接計算されてもよい。このとき、バイノーラルレンダリングは、22.2入力チャネルが直接レンダリングされるべきであるのであれば、44個のHRTF又はBRIR関数を適用するのとは対照的に、異なる位置にある5つの個々のチャンネルをレンダリングするための10個のHRTFs (10
Head Related Transfer Functions: 頭部伝達関数) 又はBRIR関数を適用しなければならないだけである。バイノーラルレンダリングに必要な畳み込みの動作は多大な処理能力を必要とし、それゆえ、許容可能なオーディオ品質を依然として得ながらこの処理能力を低減することが、モバイルデバイスにとって特に有用である。バイノーラルレンダラ236はマルチチャンネルオーディオ材料228のバイノーラルダウンミックス238を生成するので、各入力チャネル(LFEチャンネルは除外)が仮想音源によって表現される。この処理は、QMFドメインにおいてフレームごとに行うことができる。バイノーラル化は測定されたバイノーラル室内インパルス応答に基づき、直接音及び早期反射は、QMFドメインに加えて高速畳み込みを使用する疑似FFTドメインにおける畳み込み手法を介してオーディオ材料に印加することができ、一方、後期残響は別個に処理することができる 20
。

【0050】

図5は、室内インパルス応答 $h(t)$ 300の一例を示す。室内インパルス応答は、3つの成分、すなわち、直接音301、早期反射302及び後期残響304を含む。したがって、室内インパルス応答は、インパルスが再生されときの閉鎖された残響音響空間の反射挙動を描写する。早期反射302は密度が増大していく個別の反射であり、インパルス応答の、個々の反射をもはや区別することができない部分が、後期残響304と呼ばれる。直接音301は室内インパルス応答において容易に識別することができ、早期反射から分離することができるが、早期反射302から後期残響304への遷移はそれほど明白ではない。 30

【0051】

以下において、本発明の手法の実施形態をさらに詳細に説明する。本発明の実施形態によれば、オーディオ信号は、室内インパルス応答の早期部分及び後期残響を用いて別個に処理される。室内インパルス応答の早期部分によって処理されたオーディオ信号と残響付与済み信号とが組み合わされて、出力オーディオ信号として出力される。別個に処理するために、早期部分から後期残響への室内インパルス応答の遷移を知る必要がある。遷移は閾値に達する相関の大きさによって決定され、閾値は室内インパルス応答の早期部分における早期反射のうちの選択された1つについての相関の大きさに応じて設定される。相関の大きさは、室内インパルス応答に関して、初期状態を含む音響エネルギーの減衰、及び、所定の周波数範囲にわたる初期状態に続く任意の時点において始まる音響エネルギーの減衰の類似度を記述することができる。 40

【0052】

実施形態によれば、オーディオ信号を別個に処理することは、第1のプロセスの間に室内インパルス応答の早期反射部分301、302を用いてオーディオ信号を処理することと、第1のプロセスとは異なり分離している第2のプロセスの間に室内インパルス応答の拡散残響304を用いてオーディオ信号を処理することを含む。第1のプロセスから第2のプロセスへの変化は、遷移時間において発生する。さらなる実施形態によれば、第2のプロセスにおいて、拡散(後期)残響304は合成残響に置き換えることができる。その場合、与えられる室内インパルス応答は、早期反射部分301、302(図5参照)のみを含むことがあり、後期拡散残響304は含まれない。 50

【 0 0 5 3 】

図 6 (A) は、本発明の一実施形態による、室内インパルス応答の早期部分及び後期残響を有するオーディオ信号を別個に処理するための第 1 の例示的な信号処理装置のブロック図を示す。室内インパルス応答の異なる複数の部分によるオーディオ信号の処理は、上述したバイノーラルレンダラ 2 3 6 において実行することができる。オーディオ入力信号 4 0 0 は非残響オーディオ材料、例えば、マルチチャネルオーディオ入力信号とすることができ、それには室内インパルス応答、例えば、人工骨頭又はインナーイヤマイクロホンを使用して測定された室内インパルス応答が畳み込まれる。この畳み込みは、元の非残響オーディオ材料が室内インパルス応答と関連付けられる部屋で聴かれているような、そのオーディオ材料の空間印象を得ることを可能にする。例えば、上述したバイノーラルレンダラ 2 3 6 において、室内インパルス応答の直接音 3 0 1 及び早期反射 3 0 2 を用いてオーディオ信号を処理し、別個に、後期残響 3 0 4 を用いてオーディオ信号を処理することが所望されることがある。オーディオ入力信号 4 0 0 を処理するために、直接音処理のためのブロック 4 0 2、早期反射処理のためのブロック 4 0 4、及び後期残響処理のためのブロック 4 0 6 が設けられている。それぞれのブロック 4 0 2、4 0 6 の出力信号 4 0 8 と 4 1 0 は、早期処理済み信号 4 1 4 を生成するために第 1 の加算器 4 1 2 によって組み合わせられる。オーディオ信号が室内インパルス応答と関連付けられる部屋で聴かれているような印象を聴き手に与えるオーディオ出力信号 4 2 0 を生成するために、早期処理済み信号 4 1 4 とプロセッサ 4 0 6 によって与えられた残響付与済み信号 4 1 6 が第 2 の加算器 4 1 8 によって組み合わせられる。

【 0 0 5 4 】

直接音及び早期反射とは別個に後期残響処理 3 0 2 を処理することは、計算複雑度が低減することにより有利である。より具体的にいうと、インパルス応答全体の畳み込みを使用することは、非常に計算コストがかかる。それゆえ、より複雑度の低い残響アルゴリズムが、後期残響をシミュレートするためにオーディオ信号を処理するのに典型的に使用される。直接音とインパルス応答の早期反射部分は、例えば、畳み込みによってより正確に計算される。さらなる利点は、残響制御が可能であることである。これは、後期残響の修正を、例えば、ユーザ入力、測定される室内パラメータ又はオーディオ信号の内容に応じて行うことを可能にする。上記の利点を達成するために、早期反射 3 0 2 が終わり後期残響 3 0 4 が始まる遷移（例えば、時点）が分かる必要がある。後期残響処理の始まりが早すぎるとき、人間が聴いて失われた明確な早期反射を検知することができるため、オーディオ信号の品質がより低くなり得る。他方、遷移時間が遅すぎる場合、早期反射処理は一般的に後期残響処理よりもコストがかかるため、計算効率性が利用されなくなる。遷移、例えば時間ドメインサンプルにおける遷移は、入力パラメータとしてバイノーラルレンダラに供給することができ、入力パラメータはその後、受信された遷移に応じて、オーディオ信号を別個に処理するためのプロセッサ 4 0 2、4 0 4、4 0 6 を制御することになる。

【 0 0 5 5 】

図 6 (B) は、本発明の別の実施形態による、早期部分を有するオーディオ信号と室内インパルスの後期残響を別個に処理するための別の例示的な信号処理装置のブロック図を示す。入力信号 4 0 0、例えばマルチチャネルオーディオ入力信号、が受信され、早期部分を処理するための第 1 のプロセッサ 4 2 2 に印加される。すなわち、第 1 のプロセッサ 4 2 2 は、図 5 に示す直接音 3 0 1 及び室内インパルス応答 3 0 0 の早期反射 3 0 2 に従ってオーディオ信号を処理するためのものである。マルチチャネルオーディオ入力信号 4 0 0 は、室内インパルス応答の後期残響 3 0 4 に従ってオーディオ信号を処理するための第 2 のプロセッサ 4 2 4 にも印加される。バイノーラルレンダラにおいて、上述したように、主に、計算複雑度を低減するために、直接音及び早期反射を後期残響とは別個に処理することが所望されることがある。直接音及び早期反射の処理は、例えば、第 1 のプロセッサ 4 2 2 によって実行される畳み込み手法によってオーディオ信号に印加することができ、一方、後期残響は第 2 のプロセッサ 4 2 4 によって与えられる合成残響に置き換える

ことができる。したがって、全体的なバイノーラル出力信号 4 2 0 は、プロセッサ 4 2 2 によって与えられる畳み込み結果 4 2 8 と、プロセッサ 4 2 4 によって与えられる合成残響付与済み信号 4 3 0 との組み合わせである。実施形態によれば、信号 4 2 8 と 4 3 0 は、全体的なバイノーラル出力信号 4 2 0 を出力する加算器 4 3 2 によって組み合わせられる。

【 0 0 5 6 】

言及したように、第 1 のプロセッサ 4 2 2 は、オーディオ入力信号 4 0 0 と直接音及び室内インパルス応答の早期反射との畳み込みを引き起こすことができ、直接音及び室内インパルス応答の早期反射は複数の記録されているバイノーラル室内インパルス応答を保持する外部データベース 4 3 4 から第 1 のプロセッサ 4 2 2 に与えることができる。第 2 のプロセッサ又は残響付与装置 4 2 4 は残響 R T 6 0 及び残響エネルギーのような残響パラメータに基づいて動作することができ、残響パラメータは記憶されているバイノーラル室内インパルス応答から分析 4 3 6 によって得ることができる。留意することは、分析 4 3 6 はレンダラの必須の部分ではなく、むしろ、この分析は、データベース 4 3 4 に記憶されているそれぞれの応答からそれぞれの残響パラメータを導出することができることを示すためのものであり、この分析は外部で行われてもよいということである。残響付与装置パラメータは、例えば、オクターブ又は 3 分の 1 オクターブフィルタバンク分析においてエネルギー及び R T 6 0 残響時間を計算することによって求めてもよく、又は複数のインパルス応答分析の結果の平均値であってもよい。

【 0 0 5 7 】

加えて、両方のプロセッサ 4 2 2 及び 4 2 4 は、データベース 4 3 4 から直接又は分析 4 3 6 を介して早期部分から後期残響への室内インパルス応答の遷移に関する情報も入力パラメータとして受信する。遷移は、下記にさらに詳細に説明するように決定することができる。

【 0 0 5 8 】

実施形態によれば、遷移分析は、早期反射と後期残響とを分離するのに使用することができる。この分析は、入力パラメータとしてバイノーラルレンダラに供給することができる（例えば、残響付与装置を構成するのに使用される R T 6 0 値及びエネルギー値とともに専用ファイル/インターフェースから読み出すことができる）。この分析は、1 セットのバイノーラル室内インパルス応答（複数の方位角及び仰角の B R I R 対のセット）に基づくことができる。この分析は、すべてのインパルス応答について別個に実行される前処理段階であってもよく、このとき、すべての遷移値の中央値が 1 つの B R I R セットの全体的な遷移値として選ばれる。この全体的な遷移値はその後、バイノーラル出力信号の計算において後期残響から早期反射を分離するのに使用することができる。

【 0 0 5 9 】

遷移を決定するためのいくつかの手法が既知であるが、これらの手法は、これより説明するように不利である。従来技術文献 [1] には、早期反射から後期残響までの遷移時間を決定するためにエネルギー減衰緩和（ E D R ）及び相関の大きさを使用する方法が記載されている。しかしながら、従来技術文献 [1] に記載されている手法は不利である。

【 0 0 6 0 】

1 . この手法は、バイノーラルインパルス応答の方位角、及び直接音の振幅と最初の衝突反射の振幅との間の関係に強く依存する。

【 0 0 6 1 】

2 . 遷移時間は任意の周波数帯域において計算される。いずれの周波数帯域が、全体的なインパルス応答に使用されるための正しい遷移時間を与えるかに関しては、一般的に知られていない。

【 0 0 6 2 】

3 . この手法の基本的な相関ステップに関する情報はない。

【 0 0 6 3 】

別の既知の手法は、空間内のエコーの分散によって、例えば、秒あたりの反射の平均数

10

20

30

40

50

によって早期反射を記述し、この数が所定の閾値を超えるとときに後期残響の始まりを決定することである（従来技術文献〔2〕参照）。この手法は、すでに分かっている室内特性、すなわち室容積に依拠するが、室容積は未知であるが多い。室容積は、測定されたインパルス応答から容易に抽出することはできない。それゆえ、この方法は、測定されたインパルス応答からの遷移の計算には適用可能でない。また、後期残響と呼ばれるには、反射がどの程度密でなければならないかについては一般的に知られていない。

【0064】

従来技術文献〔3〕に記載されている別の可能性は、インパルス応答ウィンドウのある時点における実際の分布を、その時間ドメインにおけるガウス分布と比較することである。後期残響は正規分布を有すると仮定される。正規分布では、約3分の1（正確には $1/e$ ）のサンプルが平均の1標準偏差外にあり、3分の2のサンプルが平均の1標準偏差内にある。明確な早期反射は1標準偏差内により多くのサンプルを有し、外側により少ないサンプルを有する。1標準偏差外にあるサンプル対1標準偏差内にあるサンプルの比を使用して遷移時間を定義することができる。しかしながら、この手法の欠点は、比は閾値周辺で変動することがあるため、この大きさによって遷移を定義することが困難であることである。この大きさはまた、比が計算されるスライディングウィンドウのサイズ及びタイプに強く依存する。

【0065】

上述した手法に加えて、尖度（確率的信号のより高次のキユムラント（cumulant））も、遷移時間を決定するのに使用できる。尖度は、従来技術文献〔4〕に概説されているように、インパルス応答の後期部分に近づくと急速に低減する。しかしながら、遷移に対する閾値の定義（急速な低減又は最初にゼロに達する時間のいずれかの使用）は明瞭でない。

【0066】

〔2〕に記載されているような、測定されたインパルス応答の分析に依拠せず、室容積に依拠するまた別の手法がある。この手法は、遷移時間が容積のみに依存すると仮定するが、境界の拡散特性を考慮に入れない。それゆえ、結果は遷移時間の近似にすぎないものであり得、遷移時間が精密に決定されないときの上述した欠点を回避するのに必要とされるほど正確ではない。さらに、室容積はわからないことが多く、測定されたインパルス応答から容易に抽出することはできない。

【0067】

他の既知の手法は、環境を完全に無視し、遷移時間を単純に80msであると定義する。例えば、従来技術文献〔5〕参照。しかしながら、この数値は室内特性又は測定されるインパルス応答から完全にかき離れており、それゆえ、後期残響をインパルス応答の残りの部分から分離する目的のためには不正確に過ぎる。

【0068】

本発明は、実施形態によれば、オーディオ信号処理の改善に加えて、室内インパルス応答における早期反射と後期残響との間の遷移時間を決定するための手法の改善をももたらし、遷移時間のより正確な決定をもたらし。実施形態は、後述するように、FFT分析を使用して、測定されたインパルス応答から遷移時間を計算するための単純で効果的な可能性をもたらし。

【0069】

図7は、本発明の一実施形態による、室内インパルス応答の早期反射と後期残響との間の遷移時間を求めるための手法の流れ図を示す。早期反射から後期残響までの遷移時間を決定するために、第1のステップ500において、音響エネルギーの時間周波数分布が決定される。例えば、実施形態によれば、ステップ500においてエネルギー減衰緩和（ $E(t, f)$, EDR）を計算することができる。EDRは、測定された（例えば、パイノール）室内インパルス応答から直接計算することができ、一般的に使用されるエネルギー減衰曲線の周波数ドメイン展開（シュレーダー積分EDC（d））として解釈することができる。エネルギー減衰曲線は時刻t後のインパルス応答の残りのエネルギーを示す。

広帯域インパルス応答を使用する代わりに、時間周波数表現からEDRが導出され、多くの異なる時間周波数表現をこの目的のために使用することができる。ステップ500において音響エネルギーの時間周波数分布が決定されると、ステップ502において、時間周波数分布の時間ブロックにおける音響エネルギーと、初期状態における全体的な音響エネルギーとの間の相関の大きさが決定される。ステップ504において、相関の大きさが規定の閾値に達する（例えば、規定の閾値の下に下がる）か否かについて決定が行われる。相関の大きさが閾値に達しない場合、方法はステップ506に進み、現在の時間ブロックに続く次の時間ブロック及び分布が選択され、次の時間ブロックについてステップ502と504が繰り返される。したがって、ステップ502から506によれば、相関の大きさは、ステップ500において決定されるEDRの各時間ブロックと初期状態における全体的なエネルギーとの間の相関値を計算するのに使用される。相関の大きさが規定の閾値に達する（例えば、規定の閾値より下に下がる）とき、遷移時間に達する。言い換えれば、ステップ504において、現在の時間ブロックについて相関の大きさが閾値よりも低いと決定されると、方法はステップ508に進み、現在の時間ブロックの時間が遷移時間として出力される。

10

【0070】

以下において、本発明の手法の一実施形態をさらに詳細に説明する。最初に、測定されたバイノーラルインパルス応答を、遷移時間の計算のための入力として取り込むことができる。その後、Page又はLevin分布が、エネルギー減衰緩和（EDR）の計算に利用される。Page分布は過去のランニングスペクトルの導関数を指し、時間反転信号のPage分布がLevin分布と呼ばれる（従来技術文献[2]も参照されたい）。この分布は、瞬時パワースペクトルを記述し、インパルス応答 $h(t)$ のEDR（例えば、図5参照）は以下のように計算される。

20

$$E(t, \omega) = \left| \int_t^{\infty} h(\tau) e^{-j\omega\tau} d\tau \right|^2$$

ここで、

$E(t, \omega)$ はエネルギー減衰緩和、

$h(\omega)$ は室内インパルス応答、

$$= 2 \cdot f$$

である。

30

【0071】

上記の式による計算は直接音301（図5参照）において開始し、時間が増大するにつれて、エネルギー減衰緩和が含む明確な反射は少なくなり、確率的な残響が多くなる。説明されている実施形態によれば、エネルギー減衰緩和は、計算を容易にするために1msの長さを有する時間ブロックについて計算される。上述した機能によって、音響エネルギーの時間周波数分布は、図7のステップ500に関連して説明したように決定される。

【0072】

これに続いて、図7のステップ502から506に関連して説明したように、ピアソンの積率相関（.....'.....
.....）（相関係数としても知られている）に基づく相関の大きさ $\rho(t)$ が決定される。より詳細には、実施形態に従って、各時間ブロックの音響エネルギーの、初期状態における全体的なエネルギーとの相関が、以下のように決定される。

40

$$\rho(t) = \frac{\sum_{\omega} (E(1, \omega) - \bar{E}(1, \omega)) \cdot \sum_{\omega} (E(t, \omega) - \bar{E}(t, \omega))}{\sqrt{\sum_{\omega} (E(1, \omega) - \bar{E}(1, \omega))^2} \cdot \sqrt{\sum_{\omega} (E(t, \omega) - \bar{E}(t, \omega))^2}}$$

ここで、

50

$E(1, \quad)$ は周波数 f における全周波数範囲エネルギー減衰緩和、

$$\bar{E}(1, \omega)$$

は最初全範囲エネルギー減衰緩和のすべての周波数にわたる平均値、

$E(t, \quad)$ は時刻 t から始まる周波数 f におけるエネルギー減衰緩和、

$$\bar{E}(t, \omega)$$

10

は時刻 t から始まる全範囲エネルギー減衰緩和のすべての周波数にわたる平均値、

$$= 2 \quad f$$

である。

【0073】

上記相関は、初期状態を含む減衰と、任意の時刻 t において始まる減衰との類似度を記述する。この相関は、計算のために EDR の全周波数範囲を使用して広帯域 EDR から計算され、それによって、完全な初期エネルギー状況が時刻 t における状況と比較される。

【0074】

本発明はすべての周波数にわたる相関の計算には限定されない。むしろ、相関はまた、所定の周波数範囲にわたって計算されてもよい。周波数範囲は処理されるべきオーディオ信号から決定することができる。例えば、特定のオーディオ信号について、周波数範囲は、所定の範囲、例えば可聴周波数の範囲に限定することができる。実施形態によれば、周波数範囲は $20 \text{ Hz} \sim 20 \text{ kHz}$ とすることができる。留意すべきは、他の範囲も、例えば実証的研究によって選択できるということである。

20

【0075】

一実施形態によれば、EDR の有効 FFT ベースの実施形態を使用することができる。測定されるインパルス応答の有効長を有するウィンドウが適用され、測定されるインパルス応答は 2^{13} の有効長を有すると仮定され、これは 8192 周波数ビンに等しい。計算中、このウィンドウは、単一時間ブロックの個別の長さだけシフトされ、ウィンドウの終わりはゼロ詰めされる。実施形態によれば、 1 ms の時間ブロック長が使用され、EDR の単純かつ効率的な計算のために、以下の手法が適用される。

30

【0076】

(1) 測定されたインパルス応答の全有効長が FFT スペクトルを計算するために利用され、絶対値が二乗されて、 $E(1, \quad)$ がもたらされる。

【0077】

(2) インパルス応答の終わりに達するまで、インパルス応答の終わりに向かって 1 ms の個別の時間ブロック長だけウィンドウが動かされ、ウィンドウイングされたサンプルが有効長までゼロ詰めされ（すなわち、有効長を超えるサンプルがゼロにされ）、その後、 $E(t, \quad)$ をもたらす FFT スペクトルが計算される。

【0078】

40

上記の手法は有利である。というのは、EDR の狭い帯域の計算に追加のフィルタバンクなどが必要とされず、ウィンドウをシフトすることだけが必要とされるからである。図 8 は、上述した FFT ベースの手法に従ってインパルス応答について達成されるエネルギー減衰緩和の一例を示す。

【0079】

ステップ 504 及び 508 に関連して図 7 において説明したように、上述した方法で決定された相関が、その後、所定の閾値と比較される。閾値が小さくなるほど、遷移時間はインパルス応答の終わりに向かってより動く。例えば、バイノーラルインパルス応答について、閾値が $1/e = 0.3679$ になるように選択される場合（従来技術文献「2」も参照されたい）、相関は最初の反射が発生又は衝突するよりも前にすでに閾値の下に下が

50

っているため、遷移はいくつかの方位角では早すぎる。しかしながら、最初の反射は明瞭に区別することができ、確実に、後期拡散残響ではあり得ないため、遷移時間は最初の反射の到来時刻よりも遅くなければならないことが分かることから、実施形態によれば、閾値は固定閾値としては定義されない。むしろ、本発明の手法によれば、閾値は、最初の反射の衝突時刻における相関に依存するように定義される。この定義によって、最初の反射が常に遷移時間の前に位置することが保証される。実施形態によれば、ステップ508に示すように、以下が当てはまるときに遷移時間に達すると考えられる。

$$(t) = c \cdot (t_F)$$

ここで、

(t_F) は早期反射のうちの選択された1つの相関の大きさ、

t_F は直接音の後の早期反射のうちの選択された1つが衝突する時間インデックス、

c は $1/e$ に基づく一定値であり、 e はオイラー数である。

【0080】

実施形態によれば、一定値は $1/e$ であってもよいが、本発明はこの値には限定されない。実施形態によれば、一定値は $1/e$ によって近似することができ、例えば $1/e$ を所定の少数位に対して四捨五入又は切り捨てすることによって近似することができる（下記参照）。

【0081】

説明されている実施形態において、 t_F は直接音の後の最初の反射が衝突する時間ブロックインデックスである。

【0082】

図9は、本発明の手法による遷移時間決定を示し、最初の反射の衝突点における相関とび固定値又は一定値 $1/e$ を乗算することによって、インパルス応答に依存する閾値が計算される。サンプル数に対する室内インパルス応答600の振幅が示されており、最初の反射602も示されている。波形604は、式(2)を適用することによって得られる相関値を示す。606において最初の反射における相関値が示されており、示されている例では0.58の値を有する。また、従来使用されている $1/e$ の固定閾値が608において示されている。最初の反射の相関値606と元の固定値 $1/e$ が乗算器610に適用され、乗算器610は、最初の反射における相関値に依存し、記載されている実施形態では612において示すように0.21の値を有する新たな閾値を生成する。したがって、従来の手法と比較すると、遷移点614に続くすべてのサンプルがここでは後期残響304と考えられ、それよりも前のすべてのサンプルは早期反射302と考えられるように、遷移点614はさらに右に向けて動かされている。その結果もたらされる決定時間614はよりロバストであることが見てとれる。例えば、バイノーラル室内インパルス応答において、これは、計算される遷移時間が、方位角にわたってはるかにより安定していることを意味する。これは、図10と図11の比較から見てとれる。図10は、上述したEDC実施態様を使用するが、 $1/e$ の固定位置を用いて測定されたバイノーラル室内インパルス応答の左チャンネル700と右チャンネル702について、従来技術文献[1]に記載されている手法を適用するときの遷移時間を示す。耳及び方位角に対する依存性、並びに、最初の反射が衝突する前に相関 (t) が閾値より下に下がるという事実によって10ms未満まで下がる遷移時間の深い下落が明瞭に見える。図11は、本発明の手法に従って計算されたときの、左チャンネル700及び右チャンネル702の遷移時間を示す。図10に関連して説明した従来の手法と比較すると、結果としてもたらされる遷移時間の、耳及び方位角への依存性ははるかに低いことが見てとれる。

【0083】

実施形態によれば、相関が初めて閾値以下になり、その後再び閾値を超えて増大しないときに、遷移時間に達したと考えられる。計算されている相関関数においてこのサンプルと関連付けられる時間値は、インパルス応答の後期残響が始まると考えられる時間である。本発明の手法によれば、第1の反射の衝突時間は、従来技術文献[6]において説明されているような移動尖度演算子によって決定することができる。あるいは、最初の反射は

10

20

30

40

50

、他の方法によって、例えば閾値検出によって、又は例えば従来技術文献[7]に記載されているようなアタック検出によって検出してもよい。

【0084】

実施形態によれば、 $e^{-1} = 0.3679$ は、例えば、従来技術文献[1]にも示されているような確率プロセスにおいて低い相関を示すための値として使用される。実施形態によれば、この値は、 e^{-1} は0.3679として近似されるように、有効数字4桁の十進数によって使用される。他の実施形態によれば、有効桁数がより多い又はより少ない十進数も使用することができ、検出される遷移時間は e^{-1} の正確な数からずれるに従って相応に変化することが観察されている。例えば、0.368の値を使用するとき、この結果としての遷移時間の変化は1msを下回る最小限のものである。

10

【0085】

さらなる実施形態によれば、インパルス応答は帯域制限することができ、この場合、EDRは制限された周波数範囲にわたって計算することができ、相関もEDRの制限された周波数範囲にわたって計算することができる。代替的な周波数変換又はフィルタバンク、例えばFFTドメインにおいて完全に動作する手法も使用することができ、それによって、例えばFFTベースのフィルタリング/畳み込みを使用するとき、さらなる変換が節約される。

【0086】

留意することは、実施形態の上記説明において、相関値の値が最初の反射について参照されているということである。しかしながら、他の実施形態は、別の早期反射について計算された相関値を使用してもよい。

20

【0087】

上述したように、本発明の手法は、実施形態によれば、オーディオ信号のバイノーラル処理のためのバイノーラルプロセッサにおいて使用することができる。以下において、オーディオ信号のバイノーラル処理の一実施形態を説明する。バイノーラル処理は、復号済み信号をバイノーラルダウンミックス信号に変換する復号器プロセスとして実行することができ、バイノーラルダウンミックス信号はヘッドホンを通じて聴かれるときにサラウンド音体験をもたらす。

【0088】

図12は、本発明の一実施形態によるオーディオ信号のバイノーラル処理のためのバイノーラルレンダラ800の概略図を示す。図12はまた、バイノーラルレンダラにおけるQMFドメイン処理の概説も提供する。入力802において、バイノーラルレンダラ800は、処理されるべきオーディオ信号、例えばN個のチャネルと64個のQMF帯域を含む入力信号を受信する。加えて、バイノーラルレンダラ800は、オーディオ信号の処理を制御するための複数の入力パラメータを受信する。入力パラメータは、 $2 \times N$ 個のチャネル及び64個のQMF帯域のためのバイノーラル室内インパルス応答(BRIR)804、オーディオ入力信号のBRIR804の早期反射部分との畳み込みに使用される最大帯域の指標 K_{max} 806、並びに、上述した残響付与装置パラメータ808及び810(RT60及び残響エネルギー)を含む。バイノーラルレンダラ800は、受信BRIR804の早期部分を用いて入力オーディオ信号802を処理するための高速畳み込みプロセッサ812を備えている。プロセッサ812は、出力において、2つのチャネル及び K_{max} QMF帯域を含む早期処理済み信号814を生成する。バイノーラルレンダラ800は、高速畳み込みプロセッサ812を有する早期処理分岐に加えて、各々がRT60情報808及び残響エネルギー情報810を入力パラメータとして受信する2つの残響付与装置816a及び816bを含む残響分岐をも含む。残響分岐はステレオダウンミックスプロセッサ818と相関分析プロセッサ820をさらに含み、両方とも同じく入力オーディオ信号802を受信する。加えて、ステレオダウンミックスプロセッサ818によって与えられるダウンミックス済み信号822の利得を制御するための2つの利得段821aと821bが、ステレオダウンミックスプロセッサ818とそれぞれの残響付与装置816a及び816bとの間に設けられている。ステレオダウンミックスプロセッサ818は、入

30

40

50

力信号 8 0 2 に基づいて、2 つの帯域と 6 4 個の Q M F 帯域を有するダウンミックス済み信号 8 2 2 を与える。利得段 8 2 1 a と 8 2 1 b の利得は、相関分析プロセッサ 8 2 0 によって与えられるそれぞれの制御信号 8 2 4 a と 8 2 4 b によって制御される。利得制御されたダウンミックス済み信号はそれぞれの残響付与装置 8 1 6 a と 8 1 6 b に入力され、残響付与装置 8 1 6 a と 8 1 6 b はそれぞれの残響付与済み信号 8 2 6 a、8 2 6 b を生成する。早期処理済み信号 8 1 4 と残響付与済み信号 8 2 6 a、8 2 6 b は混合器 8 2 8 によって受信され、混合器 8 2 8 は受信信号を組み合わせ、2 つのチャネルと 6 4 個の Q M F 帯域を有する出力オーディオ信号 8 3 0 にする。加えて、本発明によれば、高速畳み込みプロセッサ 8 1 2 並びに残響付与装置 8 1 6 a 及び 8 1 6 b は、室内インパルス応答 8 0 4 における上述したように決定される早期部分から後期残響への遷移を示す追加の入力パラメータ 8 3 2 を受信する。

10

【 0 0 8 9 】

バイノーラルレンダラモジュール 8 0 0 (例えば、図 2 又は図 4 のバイノーラルレンダラ 2 3 6) は、入力 8 0 2 として復号済みデータストリームを有する。その信号は、ISO/IEC 14496-3:2009, subclause 8.6.4.2 に述べられている修正を加えた ISO/IEC 14496-3:2009, subclause 4.B.18.2 に概説されている Q M F 分析フィルタバンクによって処理される。レンダラモジュール 8 0 0 は Q M F ドメイン入力データを処理することもでき、その場合は分析フィルタバンクは省略できる。バイノーラル室内インパルス応答 (B R I R s) 8 0 4 は、複素 Q M F ドメインフィルタとして表される。時間ドメインバイノーラル室内インパルス応答から複素 Q M F フィルタ表現への変換は、ISO/IEC FDIS 23003-1:2006, Annex B に概説されている。B R I R s 8 0 4 は複素 Q M F ドメインにおいては特定数の時間スロットに制限されるので、それらは早期反射部分 3 0 1、3 0 2 (図 5 参照) のみを含み、後期拡散残響 3 0 4 は含まれない。早期反射から後期残響への遷移点 8 3 2 は、上述したように、例えばバイノーラル処理の前処理ステップにおける B R I R s 8 0 4 の分析によって決定される。Q M F ドメインオーディオ信号 8 0 2 と Q M F ドメイン B R I R 8 0 4 はその後、バイノーラル処理を実施するために帯域ごとの高速畳み込み 8 1 2 によって処理される。Q M F ドメイン残響付与装置 8 1 6 a、8 1 6 b が使用されて、2 チャネル Q M F ドメイン後期残響 8 2 6 a、8 2 6 b が生成される。残響付与モジュール 8 1 6 a、8 1 6 b は、周波数依存残響時間 8 0 8 とエネルギー値 8 1 0 のセットを使用して残響の特性を適合させる。残響の波形はオーディオ入力信号 8 0 2 のステレオダウンミックス 8 1 8 に基づいており、マルチチャネルオーディオ信号 8 0 2 の相関分析 8 2 0 に応じて振幅が適応的にスケールリングされる 8 2 1 a、8 2 1 b。2 チャネル Q M F ドメイン畳み込み結果 8 1 4 と 2 チャネル Q M F ドメイン残響 8 1 6 a、8 1 6 b がその後組み合わせられて 8 2 8、最終的に、2 つの Q M F 合成フィルタバンクが、ISO/IEC 14496-3:2009, subclause 4.6.18.4.2 に概説されているように、バイノーラル時間ドメイン出力信号 8 3 0 を計算する。レンダラは Q M F ドメイン出力データを計算することもでき、このときは合成フィルタバンクは省かれる。

20

30

【 0 0 9 0 】

定義

バイノーラルレンダラモジュール 8 0 0 に供給されるオーディオ信号 8 0 2 を、以下において入力信号と称する。バイノーラル処理の結果であるオーディオ信号 8 3 0 を出力信号と称する。バイノーラルレンダラモジュール 8 0 0 の入力信号 8 0 2 は、コア復号器のオーディオ出力信号 (例えば、図 2 の信号 2 2 8 参照) である。以下の様々な定義が使用される。

40

【 0 0 9 1 】

N_{in}	入力チャネルの数	
N_{out}	出力チャネルの数、 $N_{out} = 2$	
M_{DMX}	実数値非負ダウンミックス係数（ダウンミックス利得）を含むダウンミックス行列。 M_{DMX} は $N_{out} \times N_{in}$ 次元である。	
L	時間ドメインオーディオサンプルにおいて測定されたフレーム長	
v	時間ドメインサンプルインデックス	
n	QMF時間スロットインデックス（部分帯域サンプルインデックス）	10
L_n	QMF時間スロットにおいて測定されたフレーム長	
F	フレームインデックス（フレーム数）	
K	QMF周波数帯域の数、 $K = 64$	
k	QMF帯域インデックス（1... 64）	
A, B, ch	チャネルインデックス（チャネル構成のチャネル数）	
L_{trans}	時間ドメインサンプルにおけるBRIRの早期反射部分の長さ	
$L_{trans,n}$	QMF時間スロットにおけるBRIRの早期反射部分の長さ	20
N_{BRIR}	BRIRデータセット内のBRIR対の数	
L_{FFT}	FFT変換の長さ	
$\Re(\cdot)$	複素値信号の実数部	
$\Im(\cdot)$	複素値信号の虚数部	
\mathbf{m}_{conv}	いずれの入力信号チャネルがBRIRデータセット内のいずれのBRIR対に属するかを信号伝達するベクトル	
f_{max}	バイノーラル処理に使用される最大周波数	30
$f_{max,decoder}$	復号器のオーディオ出力信号内に存在する最大信号周波数	
K_{max}	オーディオ入力信号のBRIRsの早期反射部分との畳み込みに使用される最大帯域	
a	ダウンミックス行列係数	
$c_{eq,k}$	帯域ごとのエネルギー均等化係数	
ε	数値定数、 $\varepsilon = 10^{-20}$	
d	QMFドメイン時間スロットにおける遅延	
$\tilde{\mathbf{y}}_{ch}^{n',k}$	周波数帯域 k における疑似FFTドメイン信号表現	40
n'	疑似FFT周波数インデックス	
$\tilde{\mathbf{h}}^{n',k}$	周波数帯域 k におけるBRIRの疑似FFTドメイン表現	
$\tilde{\mathbf{z}}_{ch,conv}^{n',k}$	周波数帯域 k における疑似FFTドメイン畳み込み結果	
$\hat{\mathbf{z}}_{ch,conv}^{n,k}$	中間信号：QMFドメインにおける2チャネル畳み込み結果	

$\hat{\mathbf{z}}_{\text{ch,rev}}^{n,k}$	中間信号：QMF ドメインにおける 2 チャンネル残響
K_{ana}	(残響付与装置に使用される) 分析周波数帯域の数
$f_{\text{c,ana}}$	分析周波数帯域の中心周波数
$N_{\text{DMX,act}}$	ステレオダウンミックスの 1 つのチャンネルにダウンミックスされ、実際の信号フレームにおいてアクティブであるチャンネルの数
c_{corr}	1 つの信号フレームに対する全体的な相関係数
$c_{\text{corr}}^{A,B}$	チャンネル A, B の組み合わせに対する相関係数
$\sigma_{\hat{y}_{\text{ch,A}}^n}$	信号 $\hat{y}_{\text{ch,A}}^n$ の時間スロット n の標準偏差
$\mathbf{c}_{\text{scale}}$	2 つのスケーリング係数のベクトル
$\tilde{\mathbf{c}}_{\text{scale}}$	経時的に平滑化された、2 つのスケーリング係数のベクトル

10

【 0 0 9 2 】

処理

ここで、入力信号の処理を説明する。バイノーラルレンダラモジュールは、入力オーディオ信号の長さ $L = 2048$ の時間ドメインサンプルの連続した重なり合わないフレームに対して動作し、長さ L の処理済み入力フレームあたり L 個のサンプルのうちの 1 つのフレームを出力する。

20

【 0 0 9 3 】

(1) 初期化及び前処理

コア復号器 (例えば、図 2 の復号器 200 参照) によって発せられたオーディオサンプルの処理の前に、バイノーラル処理ブロックの初期化が行われる。初期化はいくつかの処理ステップから構成される。

【 0 0 9 4 】

(a) 分析値の読み出し

残響付与装置モジュール 816 a、816 b が、残響時間 808 及びエネルギー 810 の周波数依存セットを入力パラメータとして取り込む。これらの値は、バイノーラル処理モジュール 800 の初期化においてインターフェースから読み出される。加えて、時間ドメインサンプルにおける早期反射から後期残響付与への遷移時間 832 が読み出される。これらの値は、サンプルあたり 32 ビット、浮動小数点値、リトルエンディアン順序で書き込まれるバイナリファイルに記憶することができる。処理に必要とされる読み値を下記の表に記述する。

30

【 0 0 9 5 】

値の説明	数	データ型
遷移長 L_{trans}	1	整数
周波数帯域の数 K_{ana}	1	整数
周波数帯域の中心周波数 $f_{c,\text{ana}}$	K_{ana}	フロート
秒単位での残響時間 RT 60	K_{ana}	フロート
1つのBRIRの後期残響部分のエネルギー（振幅の2乗）を表すエネルギー値	K_{ana}	フロート

10

【 0 0 9 6 】

(b) BRIRs の読み出し及び前処理

20

バイノーラル室内インパルス応答 804 が、左耳 BRIRs 及び右耳 BRIRs を個々に記憶している 2 つの専用ファイルから読み出される。BRIRs の時間ドメインサンプルが、サンプルあたり 24 ビット及び 32 チャンネルの分解能で整数ウェーブファイルに記憶される。ファイル内での BRIRs の順序付けは、以下の表に記述する通りである。

【 0 0 9 7 】

チャンネル数	スピーカ ラベル
1	CH_M_L045
2	CH_M_R045
3	CH_M_000
4	CH_LFE1
5	CH_M_L135
6	CH_M_R135
7	CH_M_L030
8	CH_M_R030
9	CH_M_180
10	CH_LFE2
11	CH_M_L090
12	CH_M_R090
13	CH_U_L045
14	CH_U_R045
15	CH_U_000
16	CH_T_000
17	CH_U_L135
18	CH_U_R135
19	CH_U_L090
20	CH_U_R090
21	CH_U_180
22	CH_L_000
23	CH_L_L045
24	CH_L_R045
25	CH_M_L060
26	CH_M_R060
27	CH_M_L110
28	CH_M_R110
29	CH_U_L030
30	CH_U_R030
31	CH_U_L110
32	CH_U_R110

10

20

30

40

【0098】

スピーカ位置の1つにおいて測定されたBRI Rがない場合、ウェーブファイル内の対応するチャンネルがゼロ値を含む。LFEチャンネルはバイノーラル処理には使用されない。

【0099】

前処理ステップとして、バイノーラル室内インパルス応答(BRI R s)の所定のセットが時間ドメインフィルタから複素値QMFドメインフィルタへ変換される。複素値QMFドメインにおける所定の時間ドメインフィルタの実施は、ISO/IEC FDIS 23003-1:2006,

50

Annex Bにしたがって実行される。フィルタ変換のためのプロトタイプフィルタ係数は、ISO/IEC FDIS 23003-1:2006, Annex B, Table B.1に従って使用される。時間ドメイン表現

$$\tilde{\mathbf{h}}_{\text{ch}}^v = [\tilde{h}_1^v \cdots \tilde{h}_{N_{\text{BRIR}}}^v] \quad (\text{ただし、} 1 \leq v \leq L_{\text{trans}})$$

が処理されて、複素値QMFドメインフィルタ

$$\hat{\mathbf{h}}_{\text{ch}}^{n,k} = [\hat{h}_1^{n,k} \cdots \hat{h}_{N_{\text{BRIR}}}^{n,k}] \quad (\text{ただし、} 1 \leq n \leq L_{\text{trans}, n})$$

10

が得られる。

【0100】

(2) オーディオ信号処理

バイノーラルレンダラ800のオーディオ処理ブロックが N_{in} 個の入力チャンネルの時間ドメインオーディオサンプル802をコア復号器から得て、 $N_{\text{out}} = 2$ 個のチャンネルから構成されるバイノーラル出力信号830を生成する。

【0101】

処理は、入力として以下を取り込む。

- ・コア復号器からの復号済みオーディオデータ802、
- ・BRIRセット804の早期反射部分の複素QMFドメイン表現、及び
- ・後期残響826a、826bを生成するためにQMFドメイン残響付与装置816a、816bによって使用される周波数依存パラメータセット808、810、832。

20

【0102】

(a) オーディオ信号のQMF分析

第1の処理ステップとして、バイノーラルレンダラモジュールは、 N_{in} チャンネル時間ドメイン入力信号(コア復号器から入来する)

$$[\tilde{y}_{\text{ch},1}^v \cdots \tilde{y}_{\text{ch},N_{\text{in}}}^v] = \tilde{\mathbf{y}}_{\text{ch}}^v$$

の $L = 2048$ 個の時間ドメインサンプルを、 $L_n = 32$ 次元QMF時間スロット(スロットインデックス n)及び $K = 64$ 周波数帯域(帯域インデックス k)の N_{in} チャンネルQMFドメイン信号表現802に変換する。

30

【0103】

ISO/IEC 14496-3:2009, subclause 8.6.4.2に述べられている修正を加えたISO/IEC 14496-3:2009, subclause 4.B.18.2に概説されているQMF分析が、QMFドメイン信号

$$[\hat{y}_{\text{ch},1}^{n,k} \cdots \hat{y}_{\text{ch},N_{\text{in}}}^{n,k}] = \hat{\mathbf{y}}_{\text{ch}}^{n,k} \quad (\text{ただし、} 1 \leq v \leq L, \text{かつ } 1 \leq n \leq L_n)$$

のフレームを得るために、時間ドメイン信号

$$\tilde{\mathbf{y}}_{\text{ch}}^v$$

40

のフレームに対して実施される。

【0104】

(b) QMFドメインオーディオ信号及びQMFドメインBRIRsの高速畳み込み

次に、QMFドメインオーディオ信号802及びQMFドメインBRIRs804を処理するために、帯域ごとの高速畳み込み812が実行される。入力信号802の各チャンネル及び各BRIR804について、各QMF周波数帯域 k に対してFFT分析が実行され得る。

【0105】

QMFドメインにおける複素値に起因して、1つのFFT分析がQMFドメイン信号表

50

現の実数部に対して実行され、1つのFFT分析がQMFドメイン信号表現の虚数部に対して実行される。その結果がその後、組み合わされて、最終的な帯域ごとの複素値疑似FFTドメイン信号

$$\tilde{\mathbf{y}}_{\text{ch}}^{n',k} = \text{FFT}(\hat{\mathbf{y}}_{\text{ch}}^{n',k}) = \text{FFT}(\Re(\hat{\mathbf{y}}_{\text{ch}}^{n',k})) + j \cdot \text{FFT}(\Im(\hat{\mathbf{y}}_{\text{ch}}^{n',k}))$$

と帯域ごとの複素値BRIRs

左耳用の

$$\tilde{\mathbf{h}}_1^{n',k} = \text{FFT}(\hat{\mathbf{h}}_1^{n',k}) = \text{FFT}(\Re(\hat{\mathbf{h}}_1^{n',k})) + j \cdot \text{FFT}(\Im(\hat{\mathbf{h}}_1^{n',k}))$$

10

右耳用の

$$\tilde{\mathbf{h}}_2^{n',k} = \text{FFT}(\hat{\mathbf{h}}_2^{n',k}) = \text{FFT}(\Re(\hat{\mathbf{h}}_2^{n',k})) + j \cdot \text{FFT}(\Im(\hat{\mathbf{h}}_2^{n',k}))$$

が形成される。

【0106】

FFT変換の長さは、複素値QMFドメインBRIRフィルタ $L_{\text{trans},n}$ 及びQMFドメイン時間スロットにおけるフレーム長 L_n に従って、

$$L_{\text{FFT}} = L_n + L_{\text{trans},n} - 1$$

20

のように決定される。

【0107】

複素値疑似FFTドメイン信号はその後、複素値疑似FFTドメインBRIRフィルタと乗算されて高速畳み込み結果が形成される。入力信号のいずれのチャンネルがBRIRデータセット内のいずれのBRIR対に対応するかを信号伝達するために、ベクトル \mathbf{m}_{conv} が使用される。

【0108】

この乗算はすべてのQMF周波数帯域 k （ただし、 $1 \leq k < K_{\text{max}}$ ）について帯域ごとに行われる。最大帯域 K_{max} は、18kHz又はコア復号器からのオーディオ信号内に存在する最大信号周波数のいずれかの周波数を表すQMF帯域によって決定される。

30

$$f_{\text{max}} = \min(f_{\text{max,decoder}}, 18 \text{ kHz})$$

【0109】

各BRIR対との各オーディオ入力チャンネルの乗算結果が各QMF周波数帯域 k （ただし、 $1 \leq k < K_{\text{max}}$ ）において合計され、結果として、中間2チャンネル K_{max} 帯域疑似FFTドメイン信号がもたらされる。

$$\tilde{\mathbf{z}}_{\text{ch},1,\text{conv}}^{n',k} = \sum_{ch=1}^{ch=N_{\text{in}}} \tilde{\mathbf{y}}_{\text{ch},ch}^{n',k} \cdot \tilde{\mathbf{h}}_{1,\text{m}_{\text{conv}}[ch]}^{n',k} \text{ 及び } \tilde{\mathbf{z}}_{\text{ch},2,\text{conv}}^{n',k} = \sum_{ch=1}^{ch=N_{\text{in}}} \tilde{\mathbf{y}}_{\text{ch},ch}^{n',k} \cdot \tilde{\mathbf{h}}_{2,\text{m}_{\text{conv}}[ch]}^{n',k}$$

は、QMFドメイン周波数帯域 k における疑似FFT畳み込み結果

40

$$\tilde{\mathbf{z}}_{\text{ch,conv}}^{n',k} = [\tilde{\mathbf{z}}_{\text{ch},1,\text{conv}}^{n',k}, \tilde{\mathbf{z}}_{\text{ch},2,\text{conv}}^{n',k}]$$

である。

【0110】

次に、畳み込み結果をQMFドメインに変換し戻すために、帯域ごとのFFT合成が実行され、 L_{FFT} 個の時間スロット

$$\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k} = [\hat{\mathbf{z}}_{\text{ch},1,\text{conv}}^{n,k}, \hat{\mathbf{z}}_{\text{ch},2,\text{conv}}^{n,k}] \text{ (ただし、 } 1 \leq n < L_{\text{FFT}} \text{ かつ } 1 \leq k < K_{\text{max}} \text{)}$$

50

を有する中間 2 チャンネル K_{\max} 帯域 Q M F ドメイン信号をもたらす。

【 0 1 1 1 】

$L = 32$ 個の時間スロットを有する各 Q M F ドメイン入力信号フレームについて、 $L32$ 個の時間スロットを有する畳み込み結果信号フレームが返される。残りの $L_{FFT} - 32$ 個の時間スロットは記憶され、後続のフレーム（複数可）において重畳加算処理が実行される。

【 0 1 1 2 】

（ c ）後期残響の生成

第 2 の中間信号 8 2 6 a、8 2 6 b として、

$$\hat{\mathbf{z}}_{\text{ch,rev}}^{n,k} = \left[\hat{\mathbf{z}}_{\text{ch},1,\text{rev}}^{n,k}, \hat{\mathbf{z}}_{\text{ch},2,\text{rev}}^{n,k} \right]$$

10

と呼ばれる残響信号が、周波数ドメイン残響付与装置モジュール 8 1 6 a、8 1 6 b によって生成される。周波数ドメイン残響付与装置 8 1 6 a、8 1 6 b は、入力として以下を取り込む。

- ・ 入力信号の 1 つのフレームの Q M F ドメインステレオダウンミックス 8 2 2、
- ・ 周波数依存残響時間 8 0 8 及びエネルギー値 8 1 0 を含むパラメータセット。

【 0 1 1 3 】

周波数ドメイン残響付与装置 8 1 6 a、8 1 6 b は、2 チャンネル Q M F ドメイン後期残響尾部を戻す。

20

【 0 1 1 4 】

周波数依存パラメータセットの最大使用帯域数は、最大周波数に応じて計算される。

【 0 1 1 5 】

最初に、入力信号

$$\hat{\mathbf{y}}_{\text{ch}}^{n,k}$$

の 1 つのフレームの Q M F ドメインステレオダウンミックス 8 1 8 が実行されて、入力信号チャンネルの加重和によって残響付与装置の入力が形成される。重み付け利得は、ダウンミックス行列 \mathbf{M}_{DMX} に含まれている。それらは実数値かつ非負であり、ダウンミックス行列は $N_{\text{out}} \times N_{\text{in}}$ 次元である。それは、入力信号が 2 つの出力チャンネルのうちの一方にマッピングされる非ゼロ値を含む。

30

【 0 1 1 6 】

左半球上のスピーカ位置を表すチャンネルは左出力チャンネルにマッピングされ、右半球上に位置するスピーカを表すチャンネルは右出力チャンネルにマッピングされる。これらのチャンネルの信号は、係数 1 によって重み付けされる。正中面にあるスピーカを表すチャンネルは、バイノーラル信号の両方の出力チャンネルにマッピングされる。これらのチャンネルの入力信号は、以下の係数によって重み付けされる。

$$a = 0.7071 \approx \frac{1}{\sqrt{2}}$$

40

【 0 1 1 7 】

加えて、エネルギー均等化ステップがダウンミックスにおいて実施される。このステップは、1 つのダウンミックスチャンネルの帯域ごとのエネルギーを、このダウンミックスチャンネルに含まれている入力信号チャンネルの帯域ごとのエネルギーの合計に等しくなるように適合させる。このエネルギー均等化は、帯域ごとに以下の実数値係数を乗算することによって行われる。

$$c_{eq,k} = \sqrt{\frac{P_{in}^k}{P_{out}^k + \varepsilon}}$$

【 0 1 1 8 】

係数 $c_{eq,k}$ は $[0.5, 2]$ の間に制限される。ゼロによる除算を回避するために数値定数 ε が導入されている。ダウンミックスはまた周波数 f_{max} に帯域制限される。すなわち、それより高いすべての周波数帯域における値がゼロに設定される。

【 0 1 1 9 】

図 1 3 は、本発明の一実施形態による、バイノーラルレンダラ 8 0 0 の周波数ドメイン残響付与装置 8 1 6 a、8 1 6 b における処理を概略的に表す。

10

【 0 1 2 0 】

周波数ドメイン残響付与装置において、ステレオ入力のモノラルダウンミックスが入力混合器 9 0 0 を使用して計算される。これは、第 2 の入力チャンネルに 90° の位相シフトを非コヒーレントに適用して行われる。

【 0 1 2 1 】

このモノラル信号はその後、各周波数帯域 k 内のフィードバック減衰ループ 9 0 2 に供給され、これは、減衰インパルス系列を生成する。これに並列 FIR 相関分離器が続き、相関分離器はインパルスの間の間隔に減衰するように信号エネルギーを分散させ、出力チャンネル間に非コヒーレンス性をもたらす。エネルギー減衰をもたらすために減衰フィルタタップ密度が適用される。フィルタタップ位相動作は、疎で乗算器のない相関分離器を実装するために 4 つの選択肢に制限される。

20

【 0 1 2 2 】

残響の計算の後に、チャンネル間コヒーレンス (ICC) 補正 9 0 4 が、すべての QMF 周波数帯域について残響付与装置モジュールに含まれる。ICC 補正ステップにおいて、周波数依存直接利得 g_{direct} 及びクロスミックス利得 g_{cross} が ICC を適合させるのに使用される。

【 0 1 2 3 】

複数の異なる周波数帯域のエネルギーの量及び残響時間は入力パラメータセットに含まれている。それらの値は、 $K = 64$ 個の QMF 周波数帯域に中間マッピングされる複数の周波数点において与えられる。

30

【 0 1 2 4 】

最終的な中間信号

$$\hat{\mathbf{z}}_{ch,rev}^{n,k} = \left[\hat{z}_{ch,1,rev}^{n,k}, \hat{z}_{ch,2,rev}^{n,k} \right]$$

を計算するために、周波数ドメイン残響付与装置の 2 つのインスタンスが使用される。信号

$$\hat{z}_{ch,1,rev}^{n,k}$$

40

は残響付与装置の第 1 のインスタンスの第 1 の出力チャンネルであり、

$$\hat{z}_{ch,2,rev}^{n,k}$$

は残響付与装置の第 2 のインスタンスの第 2 の出力チャンネルである。それらは、2 つのチャンネル、64 個の帯域及び 32 個の時間スロットの次元を有する最終的な残響信号フレームに組み合わされる。

【 0 1 2 5 】

ステレオダウンミックス 8 2 2 は、残響付与装置出力の正しいスケーリングを保証する

50

ために、入力信号フレームの相関の大きさ 820 に従って両方で時間スケーリングされる 821 a、821 b。スケーリング係数は、0 と 1 との間の相関係数 c_{corr} に線形依存して

$$\left[\sqrt{N_{DMX,act}}, N_{DMX,act} \right]$$

の間隔内の値として定義される。ここで、

$$c_{corr} = \frac{1}{N_{in}^2} \cdot \sum_{A=1}^{A=N_{DMX,act}} \sum_{B=1}^{B=N_{DMX,act}} c_{corr}^{A,B} \quad 10$$

かつ

$$c_{corr}^{A,B} = \left| \frac{1}{K-1} \cdot \frac{\sum_k \sum_n \hat{y}_{ch,A}^{n,k} \cdot \hat{y}_{ch,B}^{n,k*}}{\sum_n \sigma_{\hat{y}_{ch,A}^n} \cdot \sigma_{\hat{y}_{ch,B}^n}} \right|$$

である。

式中、

20

$$\sigma_{\hat{y}_{ch,A}^n}$$

はチャンネル A の 1 つの時間スロット n にわたる標準偏差を意味し、演算子

$$\{*\}$$

は複素共役を示し、

$$\hat{y}$$

30

は実際の信号フレーム内の QMF ドメイン信号

$$\hat{y}$$

のゼロ平均バージョンである。

【0126】

c_{corr} は 2 回計算される。すなわち、実際の信号フレーム F においてアクティブであり、かつステレオダウンミックスの左チャンネルに含まれるすべてのチャンネル A、B について 1 回、及び、実際の信号フレーム F においてアクティブであり、かつステレオダウンミックスの右チャンネルに含まれるすべてのチャンネル A、B について 1 回である。

40

【0127】

$N_{DMX,act}$ は、1 つのダウンミックスチャンネル A にダウンミックスされた入力チャンネルの数 (ダウンミックス行列 M_{DMX} の第 A 行内の行列要素で、ゼロでないものの数) であり、かつその入力チャンネルは現在のフレームにおいてアクティブなものである。

【0128】

スケーリング係数はこのとき、以下のようになる。

$$\mathbf{c}_{\text{scale}} = \begin{bmatrix} c_{\text{scale},1}, c_{\text{scale},2} \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{N_{\text{DMX,act},1}} + c_{\text{corr}} \cdot (N_{\text{DMX,act},1} - \sqrt{N_{\text{DMX,act},1}}), \sqrt{N_{\text{DMX,act},2}} + c_{\text{corr}} \cdot (N_{\text{DMX,act},2} - \sqrt{N_{\text{DMX,act},2}}) \end{bmatrix}$$

【 0 1 2 9 】

スケーリング係数は、一次ローパスフィルタによってオーディオ信号フレームにわたって平滑化され、平滑化スケーリング係数

$$\tilde{\mathbf{c}}_{\text{scale}} = \begin{bmatrix} \tilde{c}_{\text{scale},1}, \tilde{c}_{\text{scale},2} \end{bmatrix}$$

10

がもたらされる。

【 0 1 3 0 】

スケーリング係数は、同じ手段を用いる時間ドメイン相関分析によって第 1 のオーディオ入力データフレームにおいて初期化される。

【 0 1 3 1 】

第 1 の残響付与装置インスタンスの入力はスケーリング係数

$$\tilde{c}_{\text{scale},1}$$

20

を用いてスケーリングされ、第 2 の残響付与装置インスタンスの入力はスケーリング係数

$$\tilde{c}_{\text{scale},2}$$

を用いてスケーリングされる。

【 0 1 3 2 】

(d) 畳み込み結果及び後期残響の組み合わせ

次に、1つのQMFドメインオーディオ入力フレームについて、畳み込み結果 8 1 4

$$\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k} = \begin{bmatrix} \hat{z}_{\text{ch},1,\text{conv}}^{n,k}, \hat{z}_{\text{ch},2,\text{conv}}^{n,k} \end{bmatrix}$$

30

と残響付与装置出力 8 2 6 a、8 2 6 b

$$\hat{\mathbf{z}}_{\text{ch,rev}}^{n,k} = \begin{bmatrix} \hat{z}_{\text{ch},1,\text{rev}}^{n,k}, \hat{z}_{\text{ch},2,\text{rev}}^{n,k} \end{bmatrix}$$

が、それらの2つの信号を帯域ごとに加算する混合プロセス 8 2 8 によって組み合わせられる。畳み込みは最大で K_{max} までの帯域で行われなため、 K_{max} よりも高い上方帯域は

$$\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k}$$

40

においてゼロであることに留意されたい。

【 0 1 3 3 】

後期残響出力は、混合プロセスにおいて、 $d = ((L_{\text{trans}} - 20 \cdot 64 + 1) / 64 + 0.5) + 1$ 個の時間スロットの量だけ遅延される。

【 0 1 3 4 】

遅延 d は、妥当な時間スロットにおける後期残響の挿入を保証するために、BRIRSにおける早期反射から後期残響への遷移時間、及び20個のQMF時間スロットの残響付与装置の初期遅延、並びにBRIRSのQMF分析のための0.5QMF時間スロットの

50

分析遅延を考慮に入れる。１つの時間スロット n における組み合わせ信号

$$\hat{\mathbf{z}}_{\text{ch}}^{n,k}$$

は

$$\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k} + \hat{\mathbf{z}}_{\text{ch,rev}}^{n-d,k}$$

によって計算される。

【 0 1 3 5 】

(e) バイノーマル Q M F ドメイン信号の Q M F 合成
Q M F ドメイン出力信号

$$\hat{\mathbf{z}}_{\text{ch}}^{n,k}$$

の 3 2 個の時間スロットの 1 つの 2 チャネルフレームが、ISO/IEC 14496-3:2009, subclause 4.6.18.4.2. による Q M F 合成によって、長さ L を有する 2 チャネル時間ドメイン信号フレームに変換され、最終的な時間ドメイン出力信号 8 3 0、

$$\tilde{\mathbf{z}}_{\text{ch}}^v = \left[\tilde{z}_{\text{ch},1}^v \cdots \tilde{z}_{\text{ch},2}^v \right]$$

がもたらされる。

【 0 1 3 6 】

いくつかの態様が装置に関して説明されているが、これらの態様はまた、対応する方法の説明をも表すことは明らかであり、ブロック又はデバイスが、方法ステップ又は方法ステップの特徴に対応する。同様に、方法ステップに関して説明されている態様も、対応する装置の対応するブロック又は項目又は特徴の説明を表す。方法ステップのいくつか又はすべては、例えば、マイクロプロセッサ、プログラム可能コンピュータ又は電子回路のようなハードウェア装置によって（又はそれを使用して）実行することができる。いくつかの実施形態において、最も重要な方法のうちの何らかの 1 つ又は複数は、そのような装置によって実行することができる。

【 0 1 3 7 】

特定の実施要件に応じて、本発明の実施形態は、ハードウェア又はソフトウェアにおいて実装することができる。実施態様は、それぞれの方法が実施されるようにプログラム可能コンピュータシステムと協働する（又は協働することが可能である）電子的に読取可能な制御信号が記憶されているデジタル記憶媒体、例えば、フロッピーディスク、DVD、Blu-Ray、CD、ROM、PROM、EPROM、EEPROM、又はフラッシュメモリのような非遷移型記憶媒体を使用して実施することができる。それゆえ、デジタル記憶媒体は、コンピュータ読取り可能であることができる。

【 0 1 3 8 】

本発明によるいくつかの実施形態は、本明細書に記載されている方法のうちの 1 つが実施されるように、プログラム可能なコンピュータシステムと協働することが可能である、電子的に読取可能な制御信号が記憶されているデータ担体を含む。

【 0 1 3 9 】

一般的に、本発明の実施形態は、プログラムコードを有するコンピュータプログラム製品として実装することができ、そのプログラムコードは、このコンピュータプログラム製品がコンピュータ上で作動すると、本発明の方法のうちの 1 つを実施するように動作可能である。そのプログラムコードは、例えば、機械的に取可り可能な担体上に記憶することができる。

【 0 1 4 0 】

10

20

30

40

50

他の実施形態は、機械的に読取り可能な担体上に記憶されている、本明細書に記載されている方法のうちの1つを実施するためのコンピュータプログラムを含む。

【0141】

言い換えれば、本発明の方法の一実施形態は、それゆえ、コンピュータプログラムであって、このコンピュータプログラムがコンピュータ上で作動すると、本明細書に記載されている方法のうちの1つを実施するためのプログラムコードを有するコンピュータプログラムである。

【0142】

本発明の方法のさらなる実施形態は、それゆえ、本明細書に記載されている方法のうちの1つを実施するためのコンピュータプログラムを記録しているデータ担体（又はデジタル記憶媒体、又はコンピュータ読取り可能な媒体）である。そのデータ担体、デジタル記憶媒体又は記録している媒体は、典型的には有形及び／又は非遷移型である。

10

【0143】

本発明の方法のさらなる実施形態は、それゆえ、本明細書に記載されている方法のうちの1つを実施するためのコンピュータプログラムを表すデータストリーム又は信号系列である。そのデータストリーム又は信号系列は、例えば、データ通信接続、例えば、インターネットを介して転送されるように構成することができる。

【0144】

さらなる実施形態は、本明細書に記載されている方法のうちの1つを実施するように構成又はプログラムされている処理手段、例えば、コンピュータ又はプログラム可能な論理デバイスを含む。

20

【0145】

さらなる実施形態は、本明細書に記載されている方法のうちの1つを実施するためのコンピュータプログラムがインストールされているコンピュータを含む。

【0146】

本発明によるさらなる実施形態は、本明細書に記載されている方法のうちの1つを実施するためのコンピュータプログラムを受信機に（例えば、電子的又は光学的に）転送するように構成されている装置又はシステムを含む。その受信機は、例えば、コンピュータ、モバイルデバイス、メモリデバイスなどとしてすることができる。その装置又はシステムは、例えば、コンピュータプログラムを受信機に転送するためのファイルサーバを含むことができる。

30

【0147】

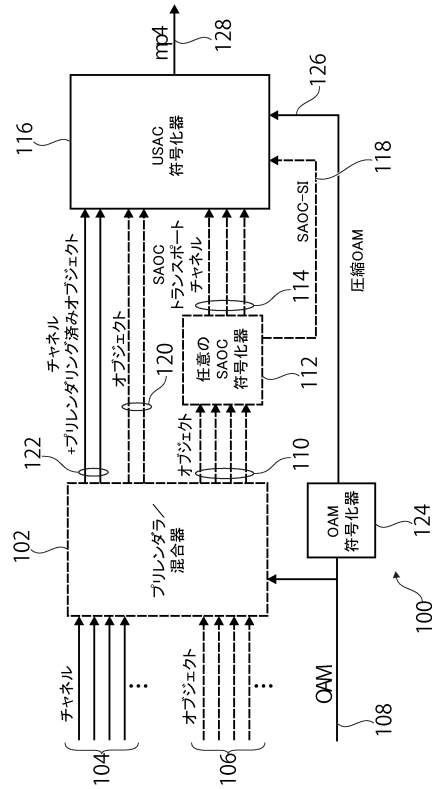
いくつかの実施形態において、プログラム可能な論理デバイス（例えば、フィールドプログラマブルゲートアレイ）を、本明細書に記載されている方法の機能のいくつか又はすべてを実施するために使用することができる。いくつかの実施形態において、フィールドプログラマブルゲートアレイは、本明細書に記載されている方法のうちの1つを実施するために、マイクロプロセッサと協働することができる。一般的に、本発明の方法は好ましくは、任意のハードウェア装置によって実施される。

【0148】

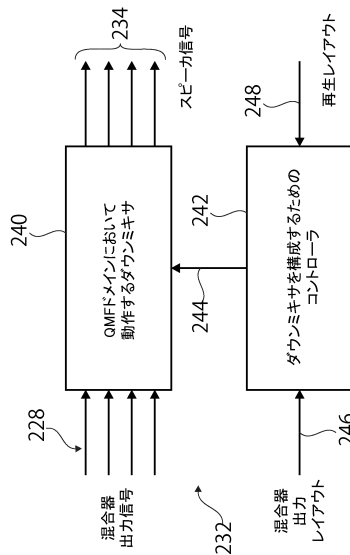
上述した実施形態は本発明の原理の例示にすぎない。当然ながら、本明細書に記載されている構成及び詳細の修正形態及び変形形態は、当業者には明らかとなる。それゆえ、意図するところは、本明細書における実施形態の記述及び説明によって提示されている特定の詳細ではなく、添付の特許請求項の範囲によってのみ限定されるということである。

40

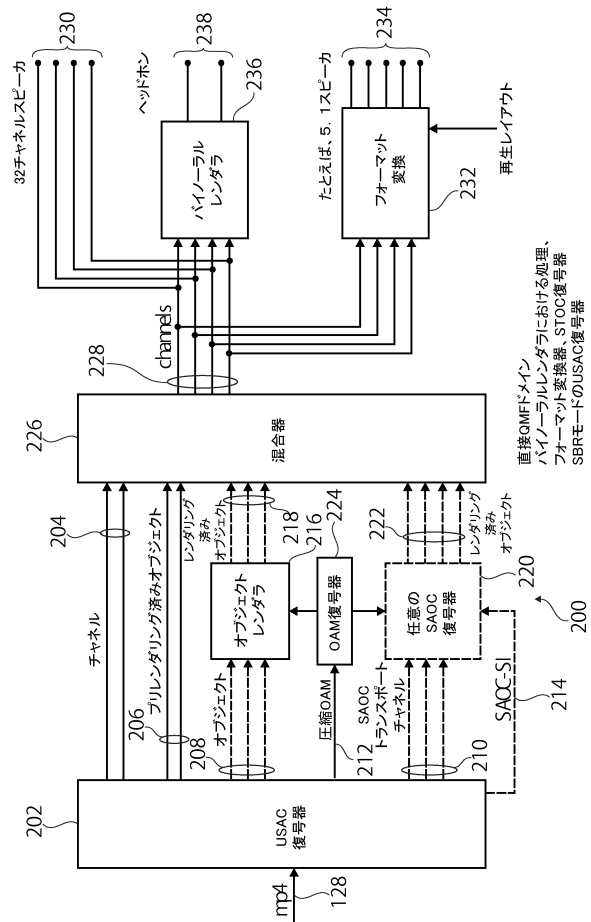
【図 1】



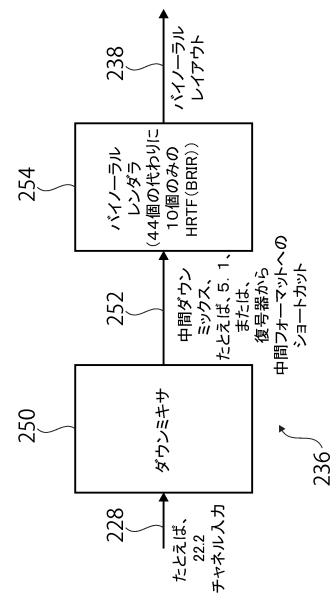
【図 3】



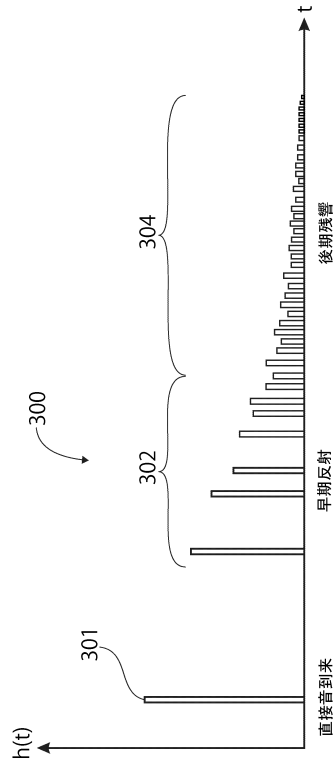
【図 2】



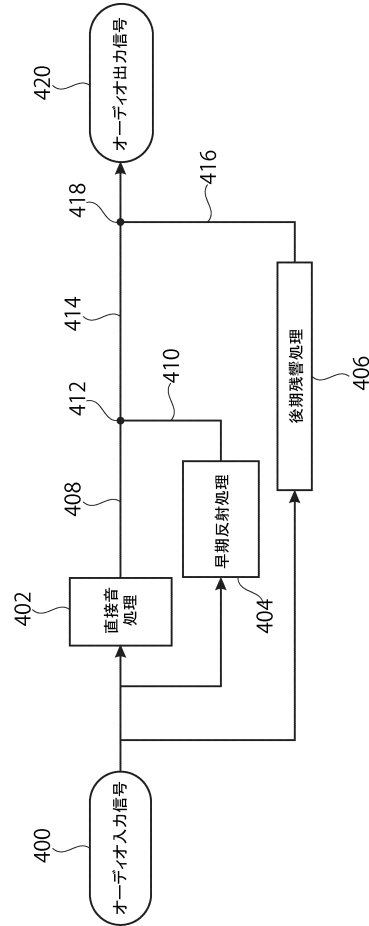
【図 4】



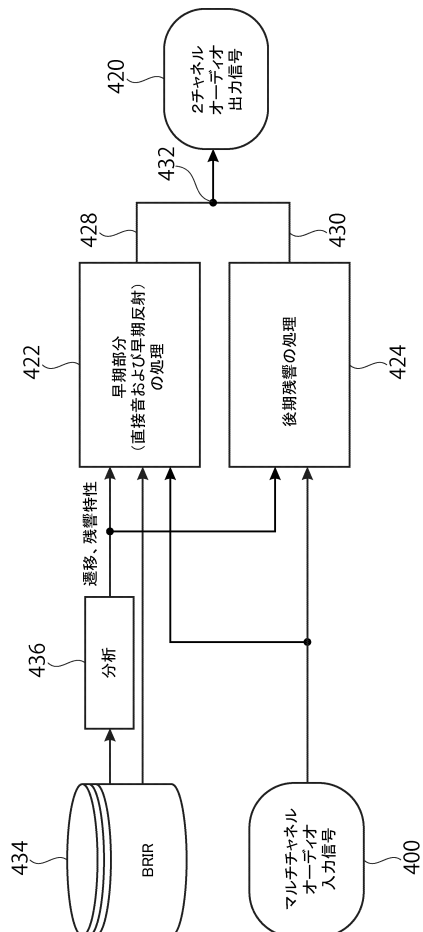
【図 5】



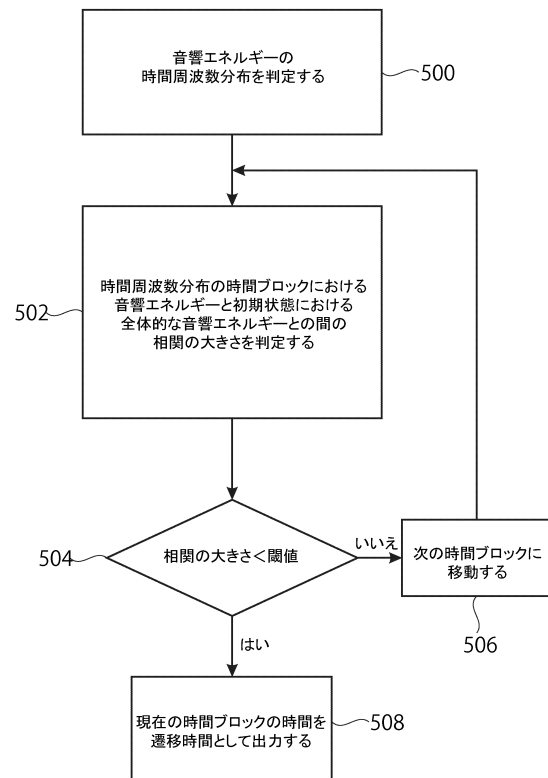
【図 6 A】



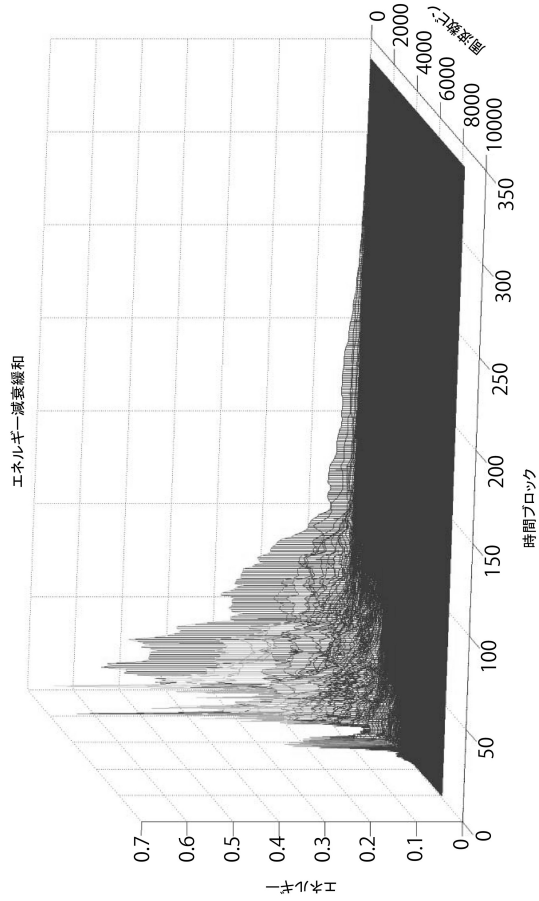
【図 6 B】



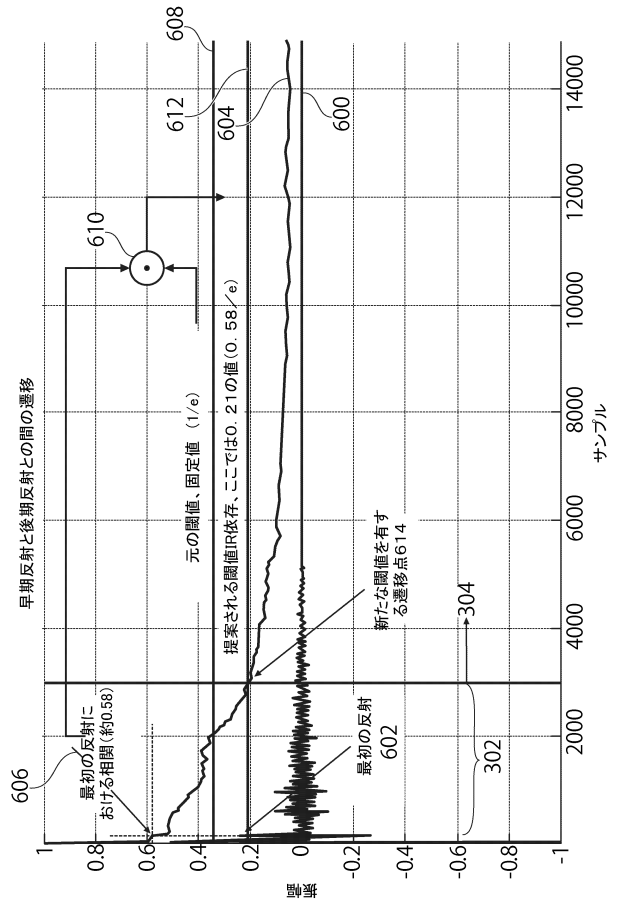
【図 7】



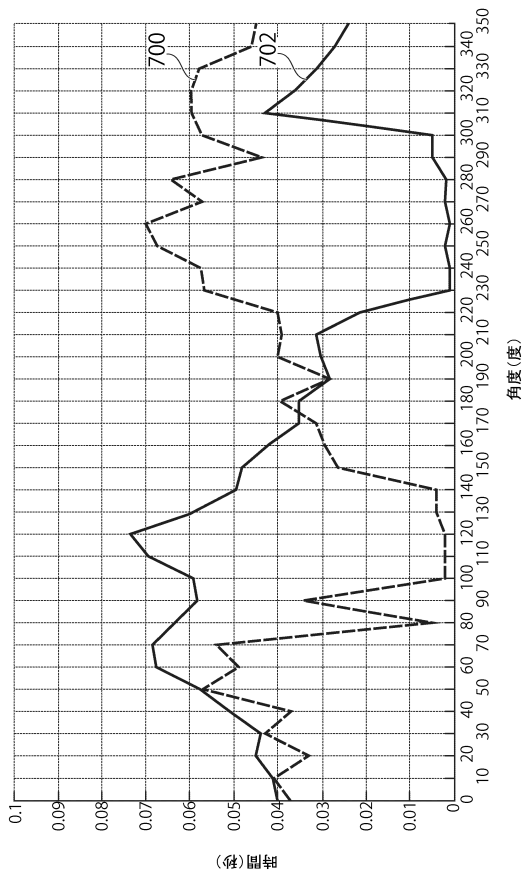
【図 8】



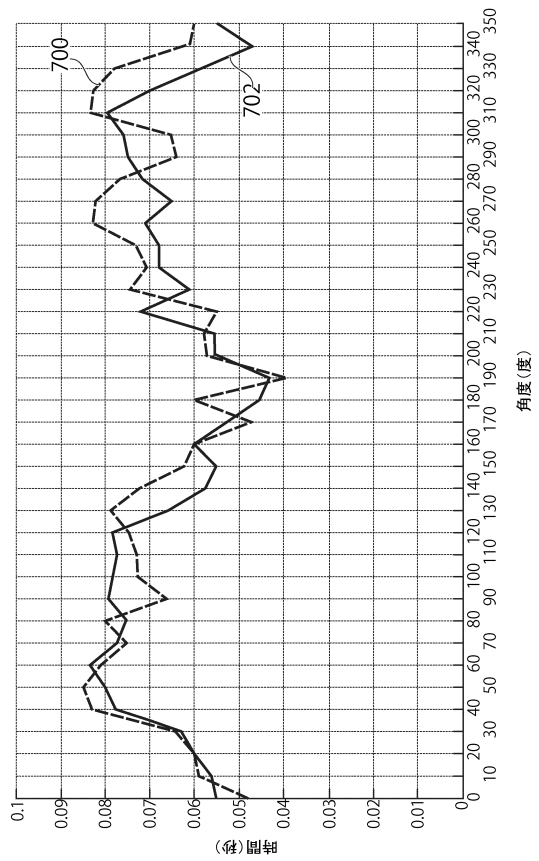
【図 9】



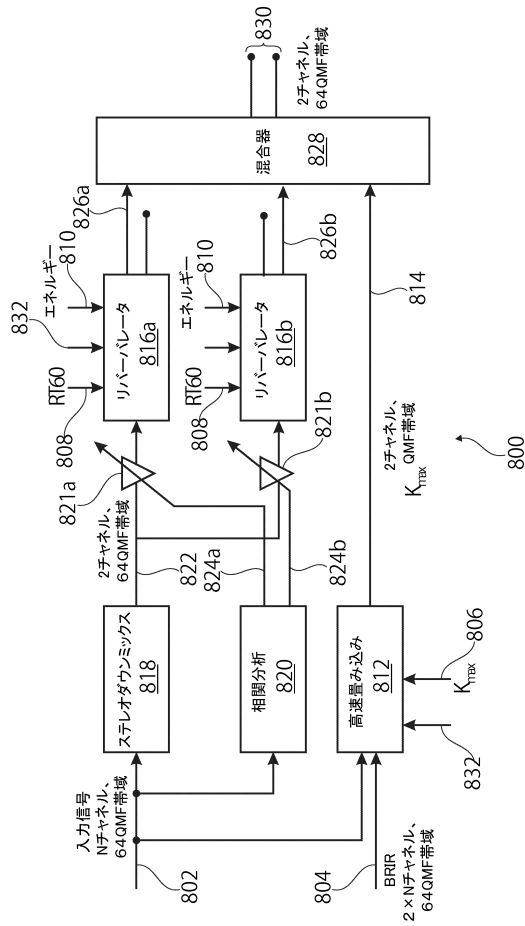
【図 10】



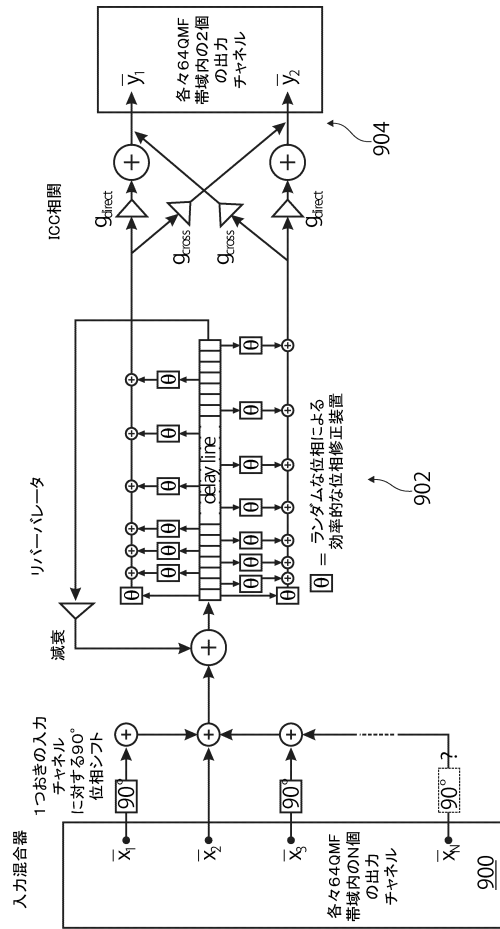
【図 11】



【圖 12】



【 図 1 3 】



フロントページの続き

(72)発明者 ヤン・プログスティエス

ドイツ連邦共和国 9 0 7 6 8 フュルト ザンドライテ 3 9 a

審査官 大石 剛

(56)参考文献 米国特許出願公開第 2 0 0 8 / 0 1 7 5 3 9 6 (U S , A 1)

特開 2 0 0 9 - 0 5 3 3 4 9 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)

G 1 0 K 1 5 / 0 0

G 1 0 L 1 9 / 0 0 8

H 0 4 R 3 / 0 0

H 0 4 S 1 / 0 0

H 0 4 S 7 / 0 0