

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7314280号

(P7314280)

(45)発行日 令和5年7月25日(2023.7.25)

(24)登録日 令和5年7月14日(2023.7.14)

(51)国際特許分類

F I

G 1 0 L 21/0388(2013.01)

G 1 0 L 21/0388 1 0 0

G 1 0 L 19/02 (2013.01)

G 1 0 L 19/02

請求項の数 20 (全26頁)

(21)出願番号	特願2021-536022(P2021-536022)	(73)特許権者	591037214
(86)(22)出願日	令和1年12月12日(2019.12.12)		フラウンホッファー - ゲゼルシャフト
(65)公表番号	特表2022-516604(P2022-516604 A)		ツァ フェルダールング デア アンゲヴ アンテン フォアシュンク エー . ファオ
(43)公表日	令和4年3月1日(2022.3.1)		ドイツ連邦共和国 8 0 6 8 6 ミュンヘ ン ハンザシュトラッセ 2 7 ツェー
(86)国際出願番号	PCT/EP2019/084974	(74)代理人	100079577
(87)国際公開番号	WO2020/126857		弁理士 岡田 全啓
(87)国際公開日	令和2年6月25日(2020.6.25)	(74)代理人	100167966
審査請求日	令和3年8月20日(2021.8.20)		弁理士 扇谷 一
(31)優先権主張番号	18215691.9	(72)発明者	ディッシェ ザシャ
(32)優先日	平成30年12月21日(2018.12.21)		ドイツ連邦共和国 9 1 0 5 8 エアラン
(33)優先権主張国・地域又は機関	欧州特許庁(EP)		ゲン アム ヴォルフスマンテル 3 3 フ ラウンホッファー - インスティチュート
(31)優先権主張番号	19166643.7		フュア インテグリールテ シャルトウン
(32)優先日	平成31年4月1日(2019.4.1)		最終頁に続く
	最終頁に続く		

(54)【発明の名称】 パルス処理を使用して周波数拡張オーディオ信号を生成するためのオーディオプロセッサおよび方法

(57)【特許請求の範囲】

【請求項 1】

ソースオーディオ信号(30)から周波数拡張オーディオ信号(420)を生成するためのオーディオプロセッサであって、

前記ソースオーディオ信号の少なくとも一部分の時間エンベロープを決定するためのエンベロープ決定器(100)と、

前記時間エンベロープを解析して前記時間エンベロープの所定の特徴の時間値を決定するための解析器(200)と、

合成信号を生成するための信号合成器(300)であって、前記合成信号の生成には、決定された前記時間値に関連してのパルスの配置を含み、前記パルスは、前記パルスが配置されている、前記時間値に関連する前記時間エンベロープの振幅から導出される重みを使用して加重される、信号合成器(300)と、

前記ソースオーディオ信号に含まれていない前記合成信号の少なくとも1つの帯域と前記ソースオーディオ信号とを結合して前記周波数拡張オーディオ信号(420)を得るための結合器(400)と、

を備える、オーディオプロセッサ。

【請求項 2】

前記解析器(200)は極大または極小の前記時間値を前記所定の特徴として決定するように構成される、請求項1に記載のオーディオプロセッサ。

【請求項 3】

10

20

前記エンベロープ決定器 (1 0 0) は、
前記ソースオーディオ信号を複数のサブバンド信号に分解 (1 0 5) し、
前記複数のサブバンド信号のうちの選択されたサブバンド信号の選択された時間エンベロープを計算 (1 2 0) し、ここで前記選択された時間エンベロープは前記時間エンベロープである、または、

前記複数のサブバンド信号のうちの少なくとも 2 つのサブバンド信号から少なくとも 2 つの時間エンベロープを計算 (1 2 0) し、前記少なくとも 2 つのサブバンド信号を結合 (1 3 0) して、前記時間エンベロープとしての結合時間エンベロープを得る
ように構成される、請求項 1 または請求項 2 に記載のオーディオプロセッサ。

【請求項 4】

10

前記エンベロープ決定器 (1 0 0) は、結合 (1 3 0) する前に前記選択されたサブバンド信号または前記時間エンベロープを正規化 (1 1 5 、 1 2 5) またはフィルタ処理するように構成される、または、

前記結合は平均化プロセスを含む、または

前記エンベロープ決定器 (1 0 0) は、前記複数のサブバンド信号のうちの全てのサブバンド信号から時間エンベロープを計算するように構成される、または、

前記エンベロープ決定器 (1 0 0) は、前記ソースオーディオ信号の単一のブロードバンド時間エンベロープを前記時間エンベロープとして決定するように構成される、
請求項 3 に記載のオーディオプロセッサ。

【請求項 5】

20

前記エンベロープ決定器 (1 0 0) は、前記時間エンベロープを、
波形を修正 (1 4 5) し、修正された前記波形をローパスフィルタ処理 (1 5 0) するように構成されたエンベロープフォロワーを使用して、または、

デジタル波形の絶対値または絶対値のべき乗を計算し、そのうえで計算結果をローパスフィルタ処理して、または

規定のウィンドウ幅のスライディングウィンドウを通した前記波形の瞬間的な二乗平均平方根値の計算を使用して、または

波形の区分的線形近似を決定することであって、ここで前記時間エンベロープは、前記区分的線形近似の結果間を移動するスライディングウィンドウ内の前記波形のピークを検出、結合して決定される、波形の区分的線形近似を決定することによって、または、

30

前記波形のための解析信号を生成するためのヒルベルト変換を使用して、且つ前記ソースオーディオ信号および前記解析信号から平方演算、加算演算および平方根演算を用いて前記時間エンベロープを計算して、
決定するように構成される、請求項 1 ないし請求項 4 のいずれかに記載のオーディオプロセッサ。

【請求項 6】

前記解析器 (2 0 0) は、
前記所定の特徴の初期時間値を決定 (2 0 5) し、
前記初期時間値から、最適化関数を使用して、または前記ソースオーディオ信号に関連付けられた副情報を使用して、あるいは処理規則に従って付加値を選択または操作して、
前記時間値を導出 (2 1 0) する

40

ように構成される、請求項 1 ないし請求項 5 のいずれかに記載のオーディオプロセッサ。

【請求項 7】

前記処理規則または前記最適化関数は、前記時間値がラスト内にラスト間隔 (T) をもって配置されるように実装され、前記ラスト間隔 (T) および前記ラストの前記時間エンベロープ内での位置は、前記時間値と前記初期時間値との偏差値が既定の特性を持つようになっている、請求項 6 に記載のオーディオプロセッサ。

【請求項 8】

前記偏差値は差分の二乗の和であり、前記既定の特性は最小限の特性である、請求項 7 に記載のオーディオプロセッサ。

50

【請求項 9】

前記信号合成器 (3 0 0) は、
正のパルスのみまたは負のパルスのみを配置 (3 0 5) してパルス列を取得し、
次いで前記パルス列内の前記パルスを加重 (3 1 5)、または
パルス列内の前記パルスの前記時間値に関連付けられた対応する重みを使用して、負の
パルスのみまたは正のパルスのみを加重し、
加重された前記パルスをそれぞれの前記時間値に配置して前記パルス列を取得する
ように構成される、請求項 1 ないし請求項 8 のいずれかに記載のオーディオプロセッサ。

【請求項 1 0】

前記信号合成器 (3 0 0) は、圧縮関数を使用して前記振幅から前記重みを導出 (3 1 0) するように構成され、前記圧縮関数は、

1 乗未満のべき関数、対数関数、平方根関数、およびより高い値を低減し、より低い
値を増大させるように構成された非線形関数
を含む関数グループからの関数である、請求項 1 ないし請求項 9 に記載の装置。

【請求項 1 1】

前記信号合成器 (3 0 0) は後処理機能 (3 2 0) を実行するように構成され、前記後
処理機能は、ノイズ付加、欠損した高調波の追加、逆フィルタ処理、およびエンベロープ
調整 (3 3 0) から成る機能グループのうちの少なくとも 1 つを含む、請求項 1 ないし請
求項 1 0 に記載のオーディオプロセッサ。

【請求項 1 2】

前記エンベロープ決定器 (1 0 0) は前記時間エンベロープを低周波数部分と高周波数
部分に分解 (1 6 0) するように構成され、

前記解析器 (2 0 0) は解析のために前記時間エンベロープの前記低周波数部分を使用
するように構成される、
請求項 1 ないし請求項 1 1 に記載のオーディオプロセッサ。

【請求項 1 3】

前記信号合成器 (3 0 0) は、エネルギー調整済み (1 8 0) ノイズを生成 (1 7 0)
し、前記エネルギー調整済み (1 8 0) ノイズを加重されたまたは非加重のパルスを含む
信号に付加 (3 3 5) して、前記合成信号を得るように構成される、請求項 1 2 に記載の
オーディオプロセッサ。

【請求項 1 4】

前記信号合成器 (3 0 0) は配置され加重された前記パルスを含む信号をハイパスフ
ィルタ処理 (3 2 5) またはバンドパスフィルタ処理 (3 2 5) して、前記ソースオーディ
オ信号に含まれない前記合成信号の前記少なくとも 1 つの帯域を得て、前記合成信号の前
記帯域のスペクトルエンベロープ調整 (3 3 0) を行うように構成される、もしくは前記
スペクトルエンベロープ調整は、前記ソースオーディオ信号に関連付けられた副情報から
導出されたエンベロープ調整値を使用して、または前記ソースオーディオ信号から導出さ
れた、または既定のエンベロープ調整関数に従って導出されたエンベロープ調整値を使用
して行われる、

請求項 1 ないし請求項 1 3 のいずれかに記載のオーディオプロセッサ。

【請求項 1 5】

前記ソースオーディオ信号は時間領域オーディオ信号であり、
前記合成信号の前記少なくとも 1 つの帯域は時間領域オーディオ信号であり、
前記結合器 (4 0 0) は前記合成信号の前記少なくとも 1 つの帯域のサンプルと前記ソ
ースオーディオ信号 (3 0) の対応するサンプルとのサンプル毎の付加を使用して、時間
領域結合を実施するように構成される、
請求項 1 ないし請求項 1 4 のいずれかに記載のオーディオプロセッサ。

【請求項 1 6】

前記ソースオーディオ信号は L P C (L P C = 線形予測符号化) 領域の励起信号であり、
前記合成信号の前記少なくとも 1 つの帯域は L P C 領域の励起信号であり、

前記結合器(400)は前記ソースオーディオ信号と前記少なくとも1つの帯域とをLPC領域におけるサンプル毎の付加によって結合(405)するように構成され、

前記結合器(400)は前記サンプル毎の付加の結果を、LPC合成フィルタを使用してフィルタ処理(410)して前記周波数拡張オーディオ信号を得るように構成され、

前記LPC合成フィルタ(410)は、副情報として前記ソースオーディオ信号に関連付けられたLPCデータ(40a)によって制御され、前記LPC合成フィルタ(410)は前記合成信号の前記少なくとも1つの帯域のためのエンベロープ情報(40b)によってさらに制御される、

請求項1ないし請求項15のいずれかに記載のオーディオプロセッサ。

【請求項17】

10

前記解析器(200)、前記信号合成器(300)および前記結合器(400)は、時間領域またはLPC時間領域で動作する、請求項1ないし請求項16のいずれかに記載のオーディオプロセッサ。

【請求項18】

前記エンベロープ決定器(100)はスペクトル変換(105)を適用して、フレームのシーケンスについて複数のバンドパス信号を抽出するように構成され、

前記信号合成器(300)はスペクトル変換を適用して前記合成信号の前記少なくとも1つの帯域を抽出(325)し、前記少なくとも1つの帯域にエンベロープ調整を施す(330)ように構成され、

前記結合器(400)は、スペクトル領域において結合し、時間領域への変換を適用して前記周波数拡張オーディオ信号(420)を得るように構成される、
請求項1ないし請求項16のいずれかに記載のオーディオプロセッサ。

20

【請求項19】

ソースオーディオ信号(30)から周波数拡張オーディオ信号(420)を生成する方法であって、

前記ソースオーディオ信号の少なくとも一部分の時間エンベロープを決定するステップ(100)と、

前記時間エンベロープの所定の特徴の時間値を決定するために、前記時間エンベロープを解析するステップ(200)と、

前記決定された時間値との関連で合成信号内にパルスを配置するステップ(305)であって、前記合成信号において、配置された前記パルスは、前記時間値に関連した前記時間エンベロープの振幅から導出された重みを使用して加重される、配置するステップと、前記周波数拡張オーディオ信号(420)を得るために、前記ソースオーディオ信号に含まれていない前記合成信号の少なくとも1つの帯域と、前記ソースオーディオ信号とを結合するステップ(400)と、
を含む、方法。

30

【請求項20】

コンピュータまたはプロセッサ上で動作する時に、請求項19に記載の生成する方法を実行するためのコンピュータプログラム。

【発明の詳細な説明】

40

【技術分野】

【0001】

本発明はオーディオ信号処理、とりわけソースオーディオ信号から周波数拡張信号を生成するための概念に関連する。

【背景技術】

【0002】

オーディオ信号の格納または伝送にはしばしば、厳しいビットレート制約が課せられる。過去の、非常に低いビットレートしか使用できなかった時には、コードは伝送されたオーディオ帯域を著しく削減しなければならなかった。今日、帯域幅拡張(BWE)方法を使用することで、広帯域信号を符号化することができる[1-2]。

50

【 0 0 0 3 】

これらのアルゴリズムは、復号済み信号の波形符号化済み低周波数（ＬＦ）部分から、高周波数（ＨＦ）スペクトル領域への転位（「パッチング」）によって生成される、高周波数コンテンツ（ＨＦ）のパラメトリック表現に依拠する。その際、まずは「生」パッチが生成され、次に「生」パッチに対してパラメータ駆動型後処理が適用される。

【 0 0 0 4 】

典型的に、前記後処理は、高周波数の生成から転位に至るまでの間顧慮されることなく、そのため結果として生じる「生」パッチに対して事後的に調整する必要がある重要な知覚特性を調整するために適用される。

【 0 0 0 5 】

しかし、例えば任意のターゲット領域にコピーされたパッチ内のスペクトル微細構造が元のコンテンツのスペクトル微細構造と著しく異なっていれば、望まないアーチファクトが生じて、復号済みオーディオ信号の知覚品質が損なわれる可能性がある。これらのケースではしばしば、適用された後処理では「生」パッチの誤った特性を完全に訂正することはできなかった。

【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、ＬＦ信号に知覚的に適合された、間隙充填または推定高周波数信号の「生」パッチコンテンツの新しい信号適応的生成を通して知覚品質を向上させることである。

【 0 0 0 7 】

知覚的に適応された「生」信号を既に得ておくことで、そうでない場合に必要となる事後訂正措置が最小化される。さらに、知覚的に適応された生信号によって、ＬＦとＨＦの間のクロスオーバー周波数の選択において、従来のアプローチよりもより低いものの選択が可能になる〔 3 〕。

【 0 0 0 8 】

BWE方式において、所定のいわゆるクロスオーバー周波数を上回るＨＦスペクトル領域の復元はしばしば、スペクトルパッチングに基づく。典型的に、ＨＦ領域は複数のスタックされたパッチから成り、これらのパッチはそれぞれ、所定のクロスオーバー周波数未満のＬＦスペクトルのバンドパス（ＢＰ）領域にソースをとる。

【 0 0 0 9 】

最新のシステムはフィルタバンクまたは時間 - 周波数変換表現において、ソーススペクトルからターゲットスペクトル領域に隣接するサブバンド係数のセットをコピーすることで、パッチングを効率的に実行する。

【 0 0 1 0 】

次の工程において、調性、ノイジネス、およびスペクトルエンベロープが、エンコーダで測定されBWE副情報としてビットストリームに伝送された元のＨＦ信号の知覚的特性およびエンベロープに緊密に類似するように、調整される。

【 0 0 1 1 】

スペクトル帯域複製（SBR）は、ハイ・エフィシエンシー・アドバンスト・オーディオ・コーディング（HE-AAC）等、現代のオーディオコーデックに使用される周知のBWEであって、上記に記載の技術を使用する〔 1 〕。

【 0 0 1 2 】

インテリジェントギャップ充填（IGF）は、MPEG-H 3Dオーディオまたは3gpp EVSコーデック等の近代のコーデックのセミパラメトリック符号化技術を意味する〔 2 〕。IGFはエンコーダにおける量子化プロセスにおいて、低ビットレートの制約によってもたらされたスペクトルホールを充填するために用いることができる。

【 0 0 1 3 】

典型的に、限られたビット予算のために透明符号化ができない場合、まずは信号の高周

10

20

30

40

50

波数（H F）領域にスペクトルホールが出現し、ビットレートが最も低い場合には、上部スペクトル領域に及ぼす影響が漸次的に増えていく。

【 0 0 1 4 】

デコーダ側では、このようなスペクトルホールはI G Fを介して、低周波数（L F）コンテンツからセミパラメトリックに生成された合成H Fコンテンツと、スペクトルエンベロープ調整およびスペクトル「ホワイトニングレベル」等の追加のパラメトリック副情報に制御される後処理を用いて置き換えられる。

【 0 0 1 5 】

しかしながら、前記後処理の後に、アーチファクトの知覚につながる可能性のあるミスマッチがまだ残っている可能性がある。このようなミスマッチは典型的には、

- 高調波のミスマッチ：高調波コンポーネントの配置場所の誤りによる拍動のミスマッチ
- 位相のミスマッチ：音声スピーチまたはプラス信号における鳴音の知覚される喪失につながる、パルス状励起信号の分散
- 調性のミスマッチ：過剰または過小な調性

【 0 0 1 6 】

したがって、追加の後処理を通してこれらのタイプのミスマッチを訂正するために、周波数および位相訂正メソッドが提案されてきた[3]。本発明においては、最新技術で確立されているように後処理工程で治すのではなく、これらのアーチファクトが「生」信号に組み込まれるのを最初から避けることを提案する。

【 0 0 1 7 】

他のB W Eの実施態様は、典型的には修正関数、二乗関数またはべき関数等の非線形関数を時間領域L F波形に適用してH F信号を推定するための時間領域技術[4]に基づいている。こうして、L Fを歪ませることによって、H Fコンテンツを復元するための「生」信号として使用することのできる、調和した高調波と不協和の高調波の濃いミクスチャが生成される。

【 0 0 1 8 】

ここで、とりわけ調和のミスマッチが問題になってくる。というのも多声コンテンツに対しては、これらの技術は、望まない不調和なコンポーネントと不可避免的にミックスされている所望の調和した高調波の濃厚なミクスチャを作り出すことになるからである。

【 0 0 1 9 】

このため、後処理はノイズネスを容易に増大させる可能性があり、ひとたび「生」と推定されたH Fに組み込まれてしまえば、望まない不調和なトーン成分を全く除去することができない。

【 0 0 2 0 】

本発明は、ソースオーディオ信号から周波数拡張オーディオ信号を生成するためのより良い概念を提供することを目的とする。

【課題を解決するための手段】

【 0 0 2 1 】

この目的は、請求項1に記載のオーディオプロセッサ、請求項17に記載のオーディオ信号を処理する方法、または請求項18に記載のコンピュータプログラムによって達成される。

【 0 0 2 2 】

本発明は、間隙充填または推定された高周波数（H F）信号の「生」パッチコンテンツの新しい信号適応的な生成によって、オーディオ帯域幅拡張または間隙充填または一般的に周波数拡張の知覚的品質の向上が得られるとの知見に基づいている。知覚的に適応された「生」信号を得ることで、そうでなければ必要となる事後の訂正措置が最小化あるいは除去すらすることができる。

【 0 0 2 3 】

波形エンベロープ合成パルス励起（W E S P E）と示される本発明の実施態様は、実際のパルス配置が時間領域エンベロープに合成される、時間領域におけるパルス列状の信号

10

20

30

40

50

の生成に基づいている。時間領域エンベロープは、例えばコアコードの出力において入手可能な、あるいはソースオーディオ信号の他のソースから入手可能な低周波数（L F）信号から導出される。こうして、知覚的に適応された「生」信号が得られる。

【 0 0 2 4 】

本発明の一側面によるオーディオプロセッサは、ソースオーディオ信号から周波数拡張オーディオを生成するように構成され、前記ソースオーディオ信号の少なくとも一部分の時間エンベロープを決定するためのエンベロープ決定器を備える。解析器は前記時間エンベロープを解析して時間エンベロープの所定の特徴の値を決定するように構成される。これらの値は時間値またはエネルギー、その他の特徴に関係する値でありうる。信号合成器は合成信号を生成するように配置され、合成信号の生成には、決定された時間値に関連したパルスの配置を含み、ここでパルスは時間エンベロープの振幅から導出された重みを使用して加重され、振幅はパルスが配置されている時間値に関連する。ソースオーディオ信号に含まれない合成信号の少なくとも1つの帯域とソースオーディオ信号とを結合して周波数拡張オーディオ信号を得るために、結合器が存在する。

10

【発明の効果】

【 0 0 2 5 】

例えば非線形処理等を使用することによる、ソースオーディオ信号からのある意味「無暗矢鱈な」高周波数の生成とは対照的に、本発明はソース信号の時間エンベロープを決定し、時間エンベロープの極大値または時間エンベロープの極小値等の時間エンベロープの所定の特性にパルスを配置すること、あるいはパルスを常に時間エンベロープの2つの極小値の間、あるいは時間エンベロープの所定の特徴に対して任意の関係性の場所に配置することによって、容易に制御することができる方法を提供している。パルスは、一般に、考慮されている周波数レンジ全体にわたって平坦な周波数コンテンツを有する。このように、理論上理想的ではないが、例えばそれに近いパルスが使用されたとしても、このような理想的ではないパルス、すなわち理想的なディラックの形状に即していないパルスの周波数コンテンツも、関心周波数レンジ内、例えばI G F（インテリジェントギャップ充填）の文脈では0 ~ 2 0 k H z、あるいは、例えばソース信号の帯域幅が制限されているオーディオ帯域幅拡張の文脈では8 ~ 1 6 k H zまたは2 0 k H zの周波数レンジ内では、比較的平坦である。

20

【 0 0 2 6 】

このように、かかるパルスから成る合成信号は、濃密かつ容易に制御される高周波数コンテンツを提供する。例えばソースオーディオ信号のフレームから抽出された時間エンベロープごとにいくつかのパルスを配置することで、スペクトル領域における整形が得られる。というのも、所定の特徴に対して配置された異なるパルスの周波数コンテンツは、少なくとも最も優勢な特徴に合致するように、あるいは一般的にはソースオーディオ信号の時間エンベロープの所定の特徴に合致するように、スペクトル領域において互いに重畳するからである。パルスが表すスペクトル値の位相は互いにロックされているため、および、好ましくは、正のパルスまたは負のパルスのいずれかが信号合成器の隣に配置されるため、種々のパルスのうちの個別のパルスによって表されるスペクトル値の位相は互いにロックされる。したがって、非常に有用な周波数領域特性を持つ制御された合成信号が得られる。典型的に、合成信号は既存のオーディオ周波数レンジ全体にわたって広がる、すなわちL Fレンジにも広がるブロードバンド信号である。周波数拡張の目的のためにソースオーディオ信号と結合されることがある最終的な信号を実際に生成するためには、合成信号の高帯域等少なくとも1つの帯域、あるいはバンドパスによって決定される信号が抽出され、ソースオーディオ信号に付加される。

30

40

【 0 0 2 7 】

本発明の概念は、完全に時間領域において、すなわち任意の特殊な変換なしで実施することができるポテンシャルを秘めている。時間領域は典型的な時間領域あるいは線形予測符号化（L P C）フィルタ処理済み時間領域、つまりスペクトル的に白化され、オーディオ信号のレンダリングにおいて使用できるようにするために最終的にL P C合成フィルタ

50

を使用して処理して元のスペクトル形状を再度入れ込む必要がある、時間領域信号のいずれかである。このように、エンベロープの決定、解析、信号合成、合成信号帯域の抽出および最終的な結合は全て時間領域で実行することができ、典型的には遅延をもたらす時間-スペクトル変換またはスペクトル-時間変換を回避することができる。しかしながら、エンベロープ決定、信号合成および結合等いくつかの工程は部分的または完全にスペクトル領域で実行することもできる点で、本発明の概念は柔軟性を持つ。このように、本発明の実施、すなわち、発明が必要とする所定の工程が時間領域で実行されるかスペクトル領域で実行されるかは、所定のアプリケーションで必要とされる典型的なデコーダ設計の対応するフレームワークに常に全面的に適應させることができる。本発明の文脈は、例えば L P C 励起信号（例えば T C X 信号）の周波数拡張が実行される L P C 音声コーダを背景にした場合に、より一層柔軟性を持つ。合成信号とソースオーディオ信号の結合は L P C 時間領域で実行され、L P C 時間領域から通常的时间領域への最終的な変換は、L P C 合成フィルタをもって行われ、具体的には、L P C 合成フィルタステージ内で、合成信号の少なくとも 1 つの帯域が表す、対応するスペクトル部分のための合成信号の典型的に好ましいエンベロープ調整が実行される。このように、典型的には必要となる後処理作業が単一のフィルタステージ内でのエンベロープ調整と結合される。このような後処理作業は、L P C 合成フィルタ処理、音声デコーダからの既知のデエンファシスフィルタ処理、その他低音部ポストフィルタ処理作業や T C S X デコーダまたはその他のデコーダで見られる、L T P（長期予測）に基づく他の音声拡張ポストフィルタ処理方法などを含み得る。

10

【 0 0 2 8 】

20

本願の好ましい実施態様を以下に、添付の下記の図面に関連して論じる。

【図面の簡単な説明】

【 0 0 2 9 】

【図 1】図 1 は、本願にかかるオーディオプロセッサの実施態様の分解組立図である。

【図 2】図 2 は、図 1 のエンベロープ決定器の好ましい実施態様のより詳細な描写である。

【図 3 a】図 3 a は、サブバンドオーディオ信号またはフルバンドオーディオ信号の時間エンベロープを計算するための実施態様である。

【図 3 b】図 3 b は、時間エンベロープの生成のための別の実施態様である。

【図 3 c】図 3 c は、ヒルベルト変換を使用して図 3 a の解析信号を決定する実施態様のフローチャートを示す。

30

【図 4】図 4 は、図 1 の解析器の好ましい実施態様を図示する。

【図 5】図 5 は、図 1 の信号合成器の好ましい実施態様を図示する。

【図 6】図 6 は、コアデコーダの文脈において使用される装置または方法としてのオーディオプロセッサの好ましい実施態様を図示する。

【図 7】図 7 は、合成信号とソースオーディオ信号との結合が L P C 領域で実行される、好ましい実施態様を図示する。

【図 8】図 8 は、ハイパスフィルタまたはローパスフィルタ、エンベロープ調整およびソースオーディオ信号と合成信号との結合がスペクトル領域で実行される、本願の別の実施態様を図示する。

【図 9 a】図 9 a は、「男性によるドイツ語のスピーチ」という音声アイテムに対する周波数拡張プロセスにおけるいくつかの信号を図示している。

40

【図 9 b】図 9 b は、「男性によるドイツ語のスピーチ」という音声アイテムについてのスペクトログラムを図示している。

【図 1 0 a】図 1 0 a は、「ピッチパイプ」という音声アイテムに対する周波数拡張プロセスにおけるいくつかの信号を図示している。

【図 1 0 b】図 1 0 b は、「ピッチパイプ」という音声アイテムについてのスペクトログラムを図示している。

【図 1 1 a】図 1 1 a は、「マドンナのヴォーグ」という音声アイテムに対する周波数拡張プロセスにおけるいくつかの信号を図示している。

【図 1 1 b】図 1 1 b は「マドンナのヴォーグ」という音声アイテムに対する周波数拡張

50

プロセスにおけるいくつかの信号を図示している。

【 0 0 3 0 】

図 1 は、一方でエンベロープ決定器 1 0 0 に入力され、他方で結合器 4 0 0 に入力されたソースオーディオ信号から、前記結合器 4 0 0 の出力において周波数拡張オーディオ信号 4 2 0 を生成するためのオーディオプロセッサを図示している。

【 0 0 3 1 】

エンベロープ決定器 1 0 0 は、ソースオーディオ信号の少なくとも一部分の時間エンベロープを決定するように構成される。エンベロープ決定器は、フルバンドソースオーディオ信号、あるいは、例えば 1 0 0、2 0 0 または 5 0 0 H z 等の周波数等、境界周波数がより低いソースオーディオ信号の、例えば一帯域または一部分のみのいずれかを使用することができる。時間エンベロープはエンベロープ決定器 1 0 0 から解析器 2 0 0 に送られ、時間エンベロープが解析されて時間エンベロープの所定の特徴の値が決定される。これらの値は時間値またはエネルギーまたは特徴に関連するその他の値でありうる。所定の特徴は、例えば、時間エンベロープの極大、時間エンベロープの極小、時間エンベロープのゼロクロッシング、または、2 つの極大または 2 つの極小の間の、例えばこれらの特徴の間のポイントが隣接する特徴に対して同じ時間距離を持つ値であるポイントでありうる。このように、かかる所定の特徴は常に、2 つの極大または 2 つの極小の中間のポイントである可能性がある。しかしながら、好ましい実施態様においては、例えば弧長積分処理を使用した時間エンベロープの極大の決定が好ましいとされる。時間エンベロープの所定の特徴の時間値は、合成信号を生成するための信号合成器 3 0 0 に送られる。合成信号の生成には、決定された時間エンベロープに関連してのパルスの配置を含み、ここでパルスは配置前または配置後に時間エンベロープの振幅から導出された重みを使用して加重され、振幅は解析器から受信した時間値に関連、またはパルスが配置されている時間値に関連している。

【 0 0 3 2 】

合成信号の少なくとも 1 つの帯域または合成信号の高周波帯域全体または合成信号のいくつかの個別且つ別個の帯域あるいは合成信号全体が結合器 4 0 0 に送られて、ソースオーディオ信号に含まれていない合成信号の少なくとも 1 つの帯域とソースオーディオ信号が結合されて、周波数拡張オーディオ信号が得られる。

【 0 0 3 3 】

好ましい実施態様においては、エンベロープ決定器は図 2 に図示されるように構成される。この実施態様において、ソースオーディオ信号または少なくともソースオーディオ信号の一部分は、1 0 5 に図示されるように複数のサブバンド信号に分解される。1 1 0 に図示されるように 1 つまたは複数またはすべてのサブバンドが選択または使用されて、1 2 0 に図示されるように（選択された）それぞれのサブバンドについて個別の時間エンベロープが決定される。1 2 5 に図示されるように、時間エンベロープは正規化またはフィルタ処理され、個別の時間エンベロープは 1 3 0 に示されるように互いに結合され、エンベロープ決定器の出力において最終時間エンベロープが得られる。この最終時間エンベロープは、図 2 に示される手順で決定される結合エンベロープであることがある。実施によっては、個別の選択されたサブバンドを正規化またはフィルタ処理するために、追加のフィルタ処理ステージ 1 1 5 が設けられることがある。すべてのサブバンドが使用されている場合には、これらの全てのサブバンドはブロック 1 1 5 に示されるように、正規化またはフィルタ処理される。1 2 5 に記載される正規化工程は回避することができ、この決定された時間エンベロープの正規化またはフィルタ処理を回避する工程は、ブロック 1 2 0 において時間エンベロープが決定されることになるサブバンドが既に正規化または対応してフィルタ処理されている場合には有用である。当然、工程 1 1 5 と 1 2 5 のいずれも実行することができ、あるいは（選択された）それぞれのサブバンド 1 2 0 について時間エンベロープを決定する工程 1 2 0 とそれに続く時間エンベロープを結合する工程 1 3 0 のみが実行されて、ブロック 1 1 5 または 1 2 5 に図示される工程は一切実行されないこともありうる。

10

20

30

40

50

【 0 0 3 4 】

別の実施態様において、ブロック 1 0 5 における分解は一切実行することはできないが、周波数が 2 0、5 0、1 0 0 または例えば 5 0 0 H z 未満の低クロスオーバー周波数によるハイパスフィルタ処理に置き換えることができ、このハイパスフィルタ処理の結果からは 1 つの時間エンベロープしか決定することはできない。当然、ハイパスフィルタ処理も回避することができ、ソースオーディオ信号および典型的にはソースオーディオ信号のフレームから 1 つの時間エンベロープのみが導出され、ソースオーディオ信号は好ましくは典型的には重畳するフレームにおいて処理されるが、重畳しないフレームも同じく使用することができる。例えば、所定のサブバンド信号がサブバンド信号の特性に関する特定の基準を満たしていないとされた場合、または何らかの理由で最終の時間エンベロープを決定する工程から排除された場合、所定のシナリオにおいてブロック 1 1 0 に記載される選択が実行される。

10

【 0 0 3 5 】

図 5 は、信号合成器 3 0 0 の好ましい実施態様を図示している。信号合成器 3 0 0 は、解析器 2 0 0 からの入力として、特徴の時間値と、さらに、エンベロープに関する更なる情報を受信する。アイテム 3 1 0 において、図 5 に図示される信号合成器 3 0 0 は、時間値に関する時間エンベロープからスケーリング係数を導出する。このため、ブロック 3 1 0 はエンベロープ振幅等のエンベロープ情報と、時間値を受信する。スケーリング係数の導出は、例えば平方根関数、たとえば 1 . 0 乗未満のべき関数、対数関数等の圧縮関数を用いて実行される。

20

【 0 0 3 6 】

信号合成器 3 0 0 は、パルスを時間値に配置する工程 3 0 5 を含み、ここで好ましくは、パルスに関連付けられる関連スペクトル値の同期フェーズを得るため、負のパルスのみまたは正のパルスのみが配置される。しかしながら、別の実施態様では、ベースバンド信号の調性がさほど高くない時に、典型的には利用可能な間隙充填または帯域幅拡張サイド情報から導出される他の基準に依存して、パルスのランダムな配置が実行される。負のパルスまたは正のパルスの配置は、元の波形の極性によって制御することができる。パルスの極性は、波高率が最も高い元の波形の極性と等しくなるように選択することができる。言い換えると、これはつまり、正のピークが正のパルスによって形作られ、負のピークは負のパルスによって形作られることを意味する。

30

【 0 0 3 7 】

工程 3 1 5 において、ブロック 3 0 5 によって得られるパルスはブロック 3 1 0 の結果を用いてスケーリングされ、パルスに対して追加の後処理 3 2 0 が施される。パルス信号が得られ、パルス信号の周波数帯域を得るため、つまり結合器に送られる合成信号の少なくとも 1 つの帯域を得るために、パルス信号はブロック 3 2 5 に図示されるようにハイパスフィルタ処理またはバンドパスフィルタ処理される。しかしながら、フィルタ処理ステージ 3 2 5 によって出力される信号に対し、任意のスペクトルエンベロープ調整が施され、このスペクトルエンベロープ調整は、所定のエンベロープ関数または副情報から導出される、もしくは例えばブラインド帯域幅拡張アプリケーションを背景とした場合にソースオーディオ信号から導出される所定のエンベロープパラメータの選択によって実行される。

40

【 0 0 3 8 】

図 6 は、周波数拡張オーディオ信号を生成するためのオーディオプロセッサまたはオーディオ処理方法の好ましい実施態様を図示している。波形エンベロープ同期化パルス励起 (W E S P E) と名付けられる本発明のアプローチはパルス列状の信号の生成に基づいており、実際のパルス配置は専用の時間領域エンベロープに同期されている。この共通エンベロープと呼ばれるものはコアデコーダ 2 0 の出力において、個別のエンベロープが結合されて 1 つの共通エンベロープになる L F 信号から、バンドパス信号のセットを介して導出される。

【 0 0 3 9 】

図 6 は帯域幅拡張 (B W E) 機能の特徴とするオーディオデコーダへの W E S P E 処理

50

の典型的な統合を示しており、これは新技術の好ましい実施態様でもある。この実施態様は、任意でフレーム間で例えば50%時間的に重畳する、例えば持続時間が20msの時間フレームに対して作用する。

【0040】

新たに提案されているWESP E BWEの利点としては、以下が挙げられる。

- ・ 粗さと拍動のアーチファクトの軽減
- ・ 信号の調和的な継続
- ・ パルスの維持
- ・ 音声BWEとして適格
- ・ 音楽も扱うことができる。
- ・ 調性、ピッチ整列、位相に関して自動調整するBWEである。

10

【0041】

WESP E処理は、以下の工程を含む。

【0042】

1. 時間エンベロープの推定(100)

コアデコーダ20から得られるLF信号がバンドパス信号の集合に分割される(105)。次に、それぞれのバンドパス信号について時間エンベロープが決定される(120)。任意で、個々のエンベロープの正規化またはフィルタ処理が適用されることがある。そのうえで、すべての時間エンベロープが共通のエンベロープに結合される(130)。好ましくは、結合作業は平均化プロセスである。

20

【0043】

2. 同期化パルスの配置

工程1で導出された共通のエンベロープは、好ましくは極大の位置のための弧長積分の適用によって解析される(205)。得られた極大候補は任意で後選択されるまたは時間距離(210)に対して固定されることがある。ディラックパルスがそれぞれの極大位置におけるHF生成のための推定された「生」信号に配置される(305)。任意で、このプロセスは副情報のサポートを受けることがある。

【0044】

3. エンベロープからの個々のパルスの大きさのスケージングの導出

前の工程2で集められたパルス列が、共通のエンベロープから導出された時間的重みによって加重される(315)。

30

【0045】

4. 後処理、HF抽出または間隙充填の選択

工程3で生成された「生」信号が任意で、例えばノイズ付加によって後処理され(320)、BWEにおいてHFとして、もしくは間隙充填ターゲットタイル信号として使用するために、フィルタ処理される(325)。

【0046】

5. エネルギー調整

工程4からのフィルタ処理された信号のスペクトルエネルギー分布は、BWEにおいてHFとして、もしくは間隙充填ターゲットタイル信号として使用するために、調整される(330)。ここで、ビットストリームからの所望のエネルギー分布に関する副情報40が使用される。

40

【0047】

6. HFまたは間隙充填信号とLFとのミキシング

最後に、工程5からの調整された信号が、通常のBWEまたは間隙充填原則に従って、すなわちHPフィルタを通してLFを補足すること、あるいは間隙充填スペクトル領域におけるスペクトルホールを充填することによって、コアコーダ出力30とミックスされる(400)。

【0048】

下記に、WESP E処理に含まれる各工程の機能を、例となる信号と処理結果に対する

50

効果を示して、さらに詳細に説明する。

【 0 0 4 9 】

適正な共通の時間的エンベロープの推定は、W E S P E の鍵となる部分である。共通のエンベロープにより、個別の時間フレームそれぞれの平均化された、ひいては代表的な知覚特性の推定が可能になる。

【 0 0 5 0 】

L F 信号がピッチ f_0 且つ間隔 Δf_0 の強い高調波線スペクトルで非常に調性が高いのであれば、個々のバンドパス信号のそれぞれにいくつかの線が現れ、パスバンド幅が調整することができるのであれば、すべてのバンドパス帯域内での拍動を通じて強いコヒーレントなエンベロープ変調が作り出される。時間エンベロープを平均化すれば、バンドパスエンベロープ全体にわたって見いだされるこのようなコヒーレントなエンベロープ変調構造が維持され、 $\Delta T_0 = 1 / (\Delta f_0)$ の間隔のほぼ等距離の位置における強いピークがもたらされる。その後、弧長積分の適用を通じて、これらのピーク位置に強いパルスが配置され、位置 $n * \Delta f_0$, $n = 1 \dots N$ に離散した等距離の線から成るスペクトルを有するパルス列が形成される。

10

【 0 0 5 1 】

強いトーン信号に高調波が一切ない場合、もしくはバンドパスフィルタの帯域幅が個々の帯域内の 1 つ以上のこれらの高調波を調整できない場合、変調構造はすべてのバンドパス信号において現れず、ひいては平均化された共通のエンベロープを支配することはなくなる。その結果生じるパルス配置は、大部分が不規則に間隔のあいた極大に基づくこととなり、よって雑音が多くなる。

20

【 0 0 5 2 】

同じことが、共通のエンベロープ信号において極大の配置がランダムな、雑音の多い L F 信号にも言える。これらは偽ランダムなパルス配置へとつながる。

【 0 0 5 3 】

このケースでは、すべてのバンドパス信号は時間的に整列され、ひいては共通のエンベロープにも表れる共通の極大を共有するため、一時的現象が維持される。

【 0 0 5 4 】

バンドパスは、知覚的帯域にわたって広がり、且つ分解が必要とされる最高周波数について少なくとも 2 つの高調波を調整することができるような大きさに定められるものとする。優れた平均化を行うためには、バンドパスは移行帯域に多少の重畳があっても良い。こうすると、評価された信号の調性は本質的に L F 信号に適應する。バンドパスは、例えば 2 0 H z 未満の超低周波数を排除することがある。

30

【 0 0 5 5 】

同期化時間パルス配置およびスケールリングは W E S P E のもう一つの鍵となる寄与である。同期化パルス配置は、共通のエンベロープの時間変調において濃縮された代表的な知覚特性を継承しており、知覚的に適應された生フルバンド信号にインプリントする。

【 0 0 5 6 】

人間による高周波数コンテンツの知覚は、臨界帯域エンベロープにおける変調の評価を通じて作用することが知られている。上記で詳説したとおり、共通の L F エンベロープに同期された時間パルス配置は、L F と H F の間の知覚的に関連する時間およびスペクトル構造の類似性および整列を強いるものである。

40

【 0 0 5 7 】

例えばピッチパイプ等強く清い高調波を持つ非常に調性の高い信号の場合には、W E S P E は追加の任意の安定化によって、パルス配置が厳密に等距離となり、ひいては非常に調性の高い「生」信号の H F 高調波スペクトルに確実につながることとなる。

【 0 0 5 8 】

パルスを共通のエンベロープで加重することで、優勢な変調が強いパルスで維持され、一方でより重要性の低い変調は弱いパルスをもたらし、「生」信号の L F 信号への本質的な適應という W E S P E の特性にさらに寄与することが確実になる。

50

【 0 0 5 9 】

ノイズの多い信号の場合には、パルス配置と加重が次第にランダムになっていけば、極めて所望の特性である、漸次的にノイズが多くなっていく「生」信号につながる。

【 0 0 6 0 】

H F 抽出、エネルギー調整およびミキシングの残りの処理工程は、新しい W E S P E 処理をコーデックに組み込んで B W E または間隙充填の全機能性に適合させるために必要なさらなる工程である。

【 0 0 6 1 】

図 3 a は、時間エンベロープの決定のための好ましい実施態様を図示している。1 3 5 に示されるとおり、解析信号はヒルベルト変換を用いて決定される。ブロック 1 3 5 の出力、すなわちヒルベルト変換信号が、1 4 0 に図示されるエンベロープ $ENV(t)$ の算出のために用いられる。このため、エンベロープは所定の時刻における元のソースオーディオ信号の時間値と、所定の時刻における対応するヒルベルト変換値の平方とを平方し、平方値を加算し、個々の時刻のそれぞれについて加算結果から平方根を計算して、算出される。この方法により、時間エンベロープは元のソースオーディオ信号 $a(t)$ と同じサンプル解像度で決定される。当然、ブロック 1 3 5 および 1 4 0 への入力、図 2 のブロック 1 0 5 から得られたまたはブロック 1 1 0 が選択したまたはブロック 1 1 5 b によって正規化及びフィルタ処理された、サブバンド信号である場合に、同じ手順が実行される。

【 0 0 6 2 】

時間エンベロープを計算するための別の手順が図 3 b のブロック 1 4 5 および 1 5 0 に図示されている。この目的のために、ソースオーディオ信号またはソースオーディオ信号からのサブバンドの波形が修正されて (1 4 5)、修正済み信号がローパスフィルタ処理され (1 5 0)、このローパスフィルタ処理の結果がソースオーディオ信号のエンベロープ、または好ましくは図 2 の 1 3 0 に図示されるような平均化によって他のサブバンドの同様なエンベロープと結合された個々のサブバンド信号のエンベロープである。

【 0 0 6 3 】

2 0 1 7 年 3 月 2 0 日発行 C . ジャーン (J a r n e) 著「3 ステップで信号エンベロープを得るための単純な経験的アルゴリズム (Simple empirical algorithm to obtain signal envelope in the three step)」は、支持が有限なスライディングウィンドウを通じて波形の瞬間的な二乗平均平方根 (R M S) の計算等、時間エンベロープの計算のための別の方法を図示している。別の方法は、波形のピークを発見しデータ全体にわたって移動するウィンドウに結合することで振幅エンベロープが作り出される、波形の区分線形近似を算出することから成る。さらなる方法は、ソースオーディオ信号またはサブバンド信号における恒常的なピークの決定と、補間によるエンベロープの導出に依拠する。

【 0 0 6 4 】

時間エンベロープを計算するための他の方法には、エンベロープを表す副情報を解釈するステップ、または対応する予測係数がフレームの時間エンベロープを表す、T N S (時間的ノイズ整形) から知ることができる時間領域フレームから導出されるスペクトル値のセットに対してスペクトル領域における予測を実行するステップを含む。

【 0 0 6 5 】

10

20

30

40

図 3 c は、図 3 a の 1 3 5 に示されるヒルベルト変換を用いた解析信号の決定の好ましい実施態様を図示している。このようなヒルベルト変換を計算するための方法は、例えばヘ・レイ他著「ヒルベルト変換を使用して振幅エンベロープおよび時間的微細構造を抽出するための P r a a t をベースとするアルゴリズム (A Praat-Based Algorithm to Extract the Amplitude Envelope and Temporal Fine Structure Using the Hilbert Transform)」、I N T E R S P E E C H 2 0 1 6 - 1 4 4 7、5 3 0 - 5 3 4 に説明されている。工程 1 3 6 において、複素スペクトルが例えば信号 $a(t)$ 、ソースオーディオ信号またはサブバンド信号から算出される。工程 1 3 7 では、複素スペクトルの正の部分が選択されるまたは負の部分が選択解除される。工程 1 3 8 では、複素スペクトルの正の部分が「-j」で乗算され、工程 1 3 9 において、この乗算の積が時間領域に変換され、虚数分をとることで、解析信号 $\hat{a}(t)$ が得られる。

10

【 0 0 6 6 】

当然、時間エンベロープを決定するための他の多くの方法を利用することができ、時間エンベロープは必ずしも実際に時間領域信号を「包絡」する必要はなく、当然、時間領域信号のいくつかの極大または極小がこの時刻における対応するエンベロープ値より大きかったり小さかったりすることもありうるということが指摘される。

【 0 0 6 7 】

図 4 は、時間エンベロープの所定の特徴の時間値を決定する方法の好ましい実施態様を図示している。このため、特徴の初期時間値を決定するために平均時間エンベロープがブロック 2 0 5 に組み込まれる。これらの初期時間値は、例えば、時間エンベロープ内で実際に見つかった極大の時間値であっても良い。実際のパルスが配置される、特徴の最終時間値は、最適化関数、副情報またはブロック 2 1 0 に示されるように生の特徴を操作することによって、生時間値または「初期」時間値から導出される。好ましくは、ブロック 2 1 0 は初期値が処理ルールに従ってまたは最適化関数に従って操作されるようになっている。とりわけ、最適化関数または処理ルールは、時間値がラスト間隔 T のラストに配置されるように実施される。とりわけ、ラスト間隔 T および / または時間エンベロープ内のラストの位置は、時間値と初期時間値の間の偏差値が既定の特性を有し、実施態様において、偏差値の合計が差分の二乗を超え、および / または既定の特性が極小であるようになっている。このように、初期時間値の決定に引き続き、初期時間値の非均一的なラストに可能な限り密接に合致するが、明確かつ理想的なトーン挙動を今や示す、時間値が等距離のラストが配置される。ラストはアップサンプリングされていない領域に比べるとより精細な時間粒度を有するアップサンプリングされた領域で決定することができる、あるいはその代わりに、サンプル未満の単位の精度を持つパルス配置のための非整数遅延を使用することもできる。

20

30

【 0 0 6 8 】

図 7 は、L P C 処理の文脈での本発明の別の実施態様を図示している。例えば図 1 または図 6 に示されるように、図 7 のオーディオプロセッサはエンベロープ決定器 1 0 0 と解析器 2 0 0 (いずれも図 7 には図示されていない)、および信号合成器 3 0 0 を備えている。しかし、図 6 に反して、コアデコーダ出力データ、すなわち L F 出力 3 0 は時間領域オーディオ信号ではなく、L P C 時間領域のオーディオ信号である。このようなデータは典型的には T C X (変換符号化励起) コーダにおいて、内部信号表現としてみる事ができる。

40

【 0 0 6 9 】

図 7 のオーディオデコーダ 2 0 によって生成された T C X データは、図 7 においては L P C 領域加算器 4 0 5 として示されるミキサに転送される。信号合成器は T C X 周波数拡張データを生成する。このように、信号合成器によって生成される合成信号は、この実施態様においては T C X データ信号であるソースオーディオ信号から導出される。こうして、ブロック 4 0 5 の出力において、依然 L P C 時間領域内にある周波数拡張信号が得られる。次いで接続された L P C 合成フィルタ 4 1 0 によって、L P C 時間領域信号の時間領

50

域への変換が実行される。

【 0 0 7 0 】

L P C 合成フィルタは、必要に応じてある種のディエンファシスをさらに行うように構成され、またさらに、この時間領域フィルタは合成信号帯域についてスペクトルエンベロープ調整も行うように構成される。このように図 7 における L P C 合成フィルタ 4 1 0 は、オーディオデコーダ 2 0 が出力した T C X データ周波数レンジの合成フィルタ処理を行うだけでなく、オーディオデコーダ 2 0 が出力した T C X データに含まれていないスペクトル帯域内のデータについてもスペクトルエンベロープ調整を実行する。典型的に、このデータも、オーディオデコーダ 2 0 がコア周波数レンジのための L P C データ 4 0 a を抽出し、さらに高帯域のためのスペクトルエンベロープ調整を抽出する、あるいは I G F (インテリジェントギャップ充填) については、図 7 の 4 0 b に記される 1 つ以上の帯域を抽出することによって、符号化済みオーディオ信号 1 0 から得られる。このように、図 1 の結合器またはミキサは L P C 領域加算器 4 0 5 および引き続き接続される図 7 の L P C 合成フィルタ 4 0 によって実装され、4 2 0 に示される L P C 合成フィルタの出力が周波数拡張済み時間領域オーディオ信号となる。スペクトルエンベロープ調整 3 3 0 が結合器 4 0 0 によるミキシング作業を実行する前に行われる図 6 の方法とは対照的に、図 7 は高帯域のエンベロープ調整または帯域の充填を、ミキシングまたは両信号の結合後に実行している。

10

【 0 0 7 1 】

図 8 は、図 6 に図示される方法の別の実施態様を示している。基本的に図 6 の実施態様は時間領域で実行されていて、ブロック 3 2 0、3 2 5、3 3 0、4 0 0 は完全に時間領域で行われている。その代わり、図 8 の実施態様は、低帯域については、任意の措置であるスペクトル変換 1 0 5 に依拠しているが、低帯域のための図 8 におけるスペクトル変換作業 1 0 5 は、図 6 のバンドパスフィルタバンク 1 0 5 の実装のために有利に使用されている。さらに、図 8 の実施態様は、典型的には図 6 のパルス配置 3 0 5 とパルススケールリング 3 1 5 を含む、パルスプロセッサ 3 4 0 の出力を変換するためのスペクトル変換器 3 4 5 を備える。図 8 のパルスプロセッサ 3 4 0 はさらに、任意の特徴として安定化ブロック 2 1 0 と、任意の特徴として極値検索ブロック 2 0 5 を備えることがある。

20

【 0 0 7 2 】

しかしながら、ハイパスフィルタ処理 3 2 5、エンベロープ調整 3 3 0、および低帯域と高帯域との結合の手順は合成フィルタバンクによって行われる、つまりスペクトル領域で行われ、図 8 の合成フィルタバンク 4 0 0 の出力は時間領域周波数拡張オーディオ信号 4 2 0 である。一方で、要素 4 0 0 が異なる帯域を結合するための単なる結合器として実装されていれば、ブロック 4 0 0 の出力は、典型的にはいかなるようにも要求された通りにさらに処理される、スペクトル値の後続のブロックから成る、完全にスペクトル領域の信号であることもできる。

30

【 0 0 7 3 】

以下に、W E S P E B W E を用いて帯域幅拡張された特性信号の 3 つの例を示す。サンプルレートは 3 2 k H z で 5 1 3 本の線を持つ片側スペクトルによる D F T (図 8 の 1 0 5) が用いられ、8 つの重畳するバンドパス信号が抽出される。4 k H z のハイパス (図 8 の 3 2 5) を実装するためには、スペクトルエンベロープ調整 (図 8 の 3 3 0) および L F と H F とのミキシング (図 8 の 4 0 0)、5 0 % 重複した類似の D F T / I D F T (図 8 の 3 4 5) が 1 6 の同形のスケール係数帯域に整理されて使用される。結果として生じる、スペクトログラムに示される信号は、4 k H z から 1 6 k H z の W E S P E によって生成された、D C から 4 k H z までの符号化されていない P C M である。

40

【 0 0 7 4 】

図 9 a は波形の短い抜粋 (1 0 2 4 個のサンプルがあるブロック 1 つ)、共通のエンベロープおよび結果的に生じる W E S P E による合成・スケールリング済みパルス配置を示している。多少の分散がある大きなパルスが広い周期構造においてほぼ等距離に配置されている。

50

【 0 0 7 5 】

図 9 b はテストアイテム全体のスペクトログラムを描写している。音声スピーチの垂直パルス構造は L F と H F の間のコヒーレントに整列された状態を維持しており、一方で摩擦音は H F 構造のようなノイズを示している。

【 0 0 7 6 】

よって、図 9 a は、アイテムが「男性によるドイツ語のスピーチ」である場合に、W E S P E がどのように音声パルスを形作り、波形、共通のエンベロープおよびパルス生成を示すかを示している。

【 0 0 7 7 】

図 1 0 は、波形の短い抜粋（ 1 0 2 4 個のサンプルがあるブロック 1 つ）、共通のエンベロープおよび結果的に生じる W E S P E による合成・スケーリング済みパルス配置を示している。狭い周期構造の中に別個の鋭いパルスが等距離に配置されている。図 1 0 b は、テストアイテム全体のスペクトログラムを描写している。ピッチパイプの水平ライン構造は L F と H F との間で整列された状態を維持しているが、ただし H F も幾分ノイズの多いものであり、さらなる安定化で恩恵を被る。

10

【 0 0 7 8 】

図 1 0 a は W E S P E がどのように倍音を形作り、波形、共通のエンベロープおよびパルス生成を示すかを示している。アイテムは「ピッチパイプ」である。図 1 0 b は W E S P E がどのように倍音を形作り、スペクトログラムを示すかを示している。アイテムは「ピッチパイプ」である。

20

【 0 0 7 9 】

図 1 1 a は、テストアイテム「マドンナのヴォーグ」の波形の短い抜粋（ 1 0 2 4 個のサンプルがあるブロック 1 つ）、共通のエンベロープおよび結果的に生じる W E S P E による合成・スケーリング済みパルス配置を示している。パルスの配置とスケーリングは、ほぼランダムな構造をしている。図 1 1 b はテストアイテム全体のスペクトログラムを描写している。ポップミュージックの縦の一時的な構造は、L F と H F との間でコヒーレントに整列された状態を維持しているが、H F の調性はほとんどが低くなっている。

【 0 0 8 0 】

図 1 1 a は W E S P E がどのようにノイズの多いミクスチャを形作っているかを示しており、波形、共通のエンベロープおよびパルス生成を示している。アイテムは「ヴォーグ」である。図 1 1 b は W E S P E がどのようにノイズの多いミクスチャを形作っているかを示し、スペクトログラムを示している。アイテムは「ヴォーグ」である。

30

以上 C を挿入のこと

【 0 0 8 1 】

図 9 a、1 0 a、1 1 a の最初の画像は低帯域ソース信号の 1 0 2 4 個のサンプルのブロックの波形を図示している。さらに、サンプルのブロックを抽出するための解析フィルタの影響は、波形がブロックの始まりすなわちサンプル 0 ではゼロに等しく、ブロックの終わりすなわちサンプル 1 0 2 3 でもゼロに等しいことで示されている。このような波形は、例えば図 1 のブロック 1 0 0 への入力または図 6 の 3 0 で得られる。図 9 a、9 b および 9 c における縦軸は常に時間領域の振幅を示し、これらの図面における横軸は常に時間変数、とりわけ、典型的には 1 つのブロックについては 0 から 1 0 2 3 にわたるサンプルの数を示す。

40

【 0 0 8 2 】

図 9 a、1 0 a、1 0 b の 2 番目の画像は平均化された低帯域エンベロープ、とりわけ、低帯域エンベロープの正の部分のみを図示している。当然、低帯域エンベロープは典型的に対称であり、負のレンジにも伸長する。しかしながら、低帯域エンベロープの正の部分のみが必要とされる。図 9 a、1 0 a、1 1 a からは、この実施態様においては、エンベロープはブロックのサンプルの第 1 の組とブロックのサンプルの最後の組を除外して計算されただけであることが見て取れるが、これは、ブロックは好ましくは重畳して計算されるものであるため、一切問題とはならない。このように、図 9 a、1 0 a、1 1 a の 2

50

番目の画像は典型的に、例えば、図 1 のブロック 1 0 0 の出力または図 2 のブロック 1 3 0 の出力を図示している。

【 0 0 8 3 】

図 9 a、1 0 a、1 1 a の 3 番目の画像は、パルススケージングに続く、つまり、パルスがエンベロープの特徴の時間値に配置され、対応するエンベロープの振幅によって加重される処理に続く、合成信号を図示している。図 9 a、1 0 a、1 1 a は配置されたパルスはサンプル 2 5 6 からサンプル 7 6 8 にしか伸長していないことを図示している。よって、加重されたパルスから成る信号は 5 1 2 個のサンプルにしかわたっておらず、これらのサンプルの前およびこれらのサンプルの後の部分は一切ない、つまりフレームの中央部分のみをカバーしている。このことは、前のフレームに重畳があり、後続のフレームにも重畳があるという状況を反映している。後続のブロックを伴うパルス信号を生成する目的において、次のブロックからのパルス信号も最初の 4 分の 1 と最後の 4 分の 1 が失われるように処理され、したがって、次のブロックからのパルス信号は図 9 a、1 0 a、1 1 a の現在ブロックからの図示されたパルス信号の直後に配置されることになる。パルス信号の重畳 / 加算作業は必ずしも必須ではないため、この方法は極めて効率的である。しかしながら、重畳 / 加算作業またはあるフレームから次のフレームへのパルス信号に関するクロスフェーディングも、必要に応じて実行することができる。

10

【 0 0 8 4 】

図 9 b、1 0 b、1 1 b はスペクトログラムを図示している。横軸は時間を表しているが、図 9 a、1 0 a、1 1 a のようにサンプルに対する時間ではなく、D F T ブロックの数に対する時間を表している。縦軸は対応する図の下部の低周波数から対応する図の上部の高周波数までの周波数スペクトログラムを図示している。水平レンジは 0 ~ 1 6 k H z にわたり、低い方の 4 分の 1 が元の信号を示し、高い方の 4 分の 3 が合成信号を表している。このように、図 9 b、1 0 b、1 1 b は周波数拡張オーディオ信号を示すが、これらの図の 1 番低い 4 分の 1 のみがソースオーディオ信号を示している。

20

【 0 0 8 5 】

図は低帯域構造が高帯域においてきわめてよく反映されていることを示している。このことはピッチパイプを図示する図 1 0 に関してとりわけ見て取ることができ、ここでピッチパイプの 3 つの異なるトーンが図 1 0 b における左から右に次々に演奏される。とりわけ、図 1 0 b の左側の最初の部分はピッチパイプの最も低いトーンであり、中間部分はピッチパイプの中間トーンであり、図 1 0 b の右側部分はピッチパイプの最も高いトーンである。ピッチパイプは非常に調性の高いスペクトルを特に特徴としており、本発明はより高い 1 2 k H z において調和構造を複製するのに特に有用であるように思われる。

30

【 0 0 8 6 】

第 3 のテストアイテムについては、このようなポップミュージックアイテムのための低帯域構造は、本発明の方法によって高周波数レンジに非常にうまく変換することが可能であることが見て取れる。

【 0 0 8 7 】

図 1 2 は図 6 の実施態様にある意味類似する別の実施態様を図示している。したがって、図 6 と類似する参照符号は図 1 2 において同じ項目を示す。図 6 の特徴に加えて、図 1 2 の実施態様は L F / H F 周波数分解器 1 6 0、ノイズテーブル等のランダムノイズまたは偽ランダムノイズ生成器 1 7 0、およびエネルギー調整器 1 8 0 を含んでいる。

40

【 0 0 8 8 】

L F / H F 分解器 1 6 0 は、時間エンベロープの L F エンベロープおよび H F エンベロープへの分解を実行する。好ましくは、L F エンベロープはローパスフィルタ処理によって決定され、H F エンベロープは L F エンベロープから H F エンベロープを差し引いて決定される。

【 0 0 8 9 】

ランダムノイズまたは偽ランダムノイズ生成器 1 7 0 はノイズ信号を生成し、エネルギー調整器 1 8 0 はノイズのエネルギーを同じく 1 8 0 で推定される H F エンベロープのエ

50

エネルギーに合わせる。(LFエンベロープからの寄与は一切なく)HFエンベロープのエネルギーに調整されたエネルギーを有するノイズは、付加器335によって、ブロック315が出力する加重済みパルス列に付加される。しかし例えば処理ブロックまたは工程315、335の順番は、変更することもできる。

【0090】

一方で、205~315に関する手順は、ブロック160によって決定されるLFエンベロープにしか適用されない。

【0091】

少なくとも2つの部分へのフルバンドエンベロープの分解に依拠する好ましい実施態様は、下記の順番または他の技術的に実現可能な順番で、以下のブロックまたは工程を含む。

10

【0092】

時間エンベロープ推定100

修正;例えば関数 $x^{\wedge}0.75$ を使用した圧縮;そのうえでの、エンベロープのLFエンベロープおよびHFエンベロープへの分割160。LFエンベロープはローパスフィルタ処理を通じて得られ、その時クロスオーバー周波数は例えば2~6kHzである。ある実施態様においては、HFエンベロープは元のエンベロープと、好ましくは遅延調整されたLFエンベロープとの差分である。

【0093】

同期化パルス配置300

上記工程で導出されたLFエンベロープは、例えば弧長積分によって解析され、LFエンベロープの極値の位置にパルス配置が行われる。

20

【0094】

エンベロープから導出される個々のパルスの大きさのスケージング315

上記工程において集められたパルス列が、LFエンベロープから導出された時間的重みによって加重される。

【0095】

HFエンベロープのエネルギーが推定され、同エネルギーのランダムノイズが加重されたパルス列に付加(355)される。

【0096】

後処理、HF抽出または間隙充填の選択

30

上記の工程においてブロック335の出力に生成された「生」信号は、例えばノイズ付加によって任意で後処理(320)され、BWEにおいてHFとして使用するため、または間隙充填ターゲットタイル信号として使用するためにフィルタ処理(325)される。

【0097】

エネルギー調整330

上記の工程に説明されるエネルギー推定からのフィルタ済み信号のスペクトルエネルギー分布は、BWEにおいてHFとして使用するため、または間隙充填ターゲットタイル信号として使用するために調整される。ここで、所望のエネルギー分布に関するビットストリームからの情報が好ましくは使用される。

【0098】

40

HFまたは間隙充填信号とLFとのミキシング400

最後に、工程5の調整された信号は通常のBWWまたは間隙充填原則に従って、すなわちHPフィルタを通し、LFを補完すること、または間隙充填スペクトル領域におけるスペクトルホールを充填することによってコアコード出力とミックスされる。

【0099】

ここで、上記の全ての選択肢または側面、および下記請求の範囲の独立請求項に特定されるすべての側面は独立して、すなわち考慮されている選択肢、目的または独立請求項以外の他の選択肢または目的なしに使用することができることを指摘しておく。しかし、他の実施態様においては、2つ以上の選択肢または側面または独立請求項を互いに組み合わせることができ、また他の実施態様においては、すべての側面または選択肢と全ての独立

50

請求項を互いに組み合わせることができる。

【 0 1 0 0 】

本発明に従って符号化されたオーディオ信号は、デジタル記憶媒体または非一時的記憶媒体に格納することができる、あるいはインターネット等の無線伝送媒体または有線伝送媒体等の伝送媒体上で伝送することができる。

【 0 1 0 1 】

いくつかの側面は装置の文脈で記載したが、これらの側面は対応する方法の記載を表すことは明らかであり、ブロックまたは装置は方法工程または方法工程の特徴を示す。同様に、方法工程の文脈で記載された側面は対応するブロックまたはアイテムあるいは対応する装置の特徴も表す。

10

【 0 1 0 2 】

所定の実施要件によっては、本発明の実施態様はハードウェアでもソフトウェアでも実装することができる。実施は電子的に読み取り可能な制御信号を格納した、それぞれの方法が実行でき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（あるいは協働することができる）、例えばフロッピーディスク、DVD、CD、RO、PROM、EPROM、EEPROMまたはFLASHメモリ等のデジタル記憶媒体を使用して実行することができる。

【 0 1 0 3 】

本発明に記載の一部の実施態様は、本発明に記載の方法の1つが実行されるように、プログラム可能なコンピュータシステムと協働することができる、電子的に読み取り可能な制御信号を有するデータキャリアを備える。

20

【 0 1 0 4 】

概して、本発明の実施態様はプログラムコードを備えるコンピュータプログラム製品として実装ことができ、プログラムコードは、プログラムコードがコンピュータ上で実行されるときに、方法のうちの1つを実行するように動作する。プログラムコードは例えば機械可読のキャリア上に格納されていてもよい。

【 0 1 0 5 】

他の実施態様は、機械可読キャリアまたは非一時的記憶媒体に格納された、本明細書に記載の方法のうちの1つを実行するためのコンピュータプログラムを含む。

【 0 1 0 6 】

30

言い換えれば、本発明の方法の1つの実施態様はしたがって、コンピュータプログラムがコンピュータ上で実行されるときに、本明細書に記載の方法の1つを実行するためのプログラムコードを有するコンピュータプログラムである。

【 0 1 0 7 】

したがって、本発明の方法の別の実施態様は、本明細書に記載された方法の1つを実行するためのコンピュータプログラムをその上に記録したデータキャリア（またはデジタル記憶媒体、またはコンピュータ読み取り可能な媒体）である。

【 0 1 0 8 】

したがって、本発明の方法の別の実施態様は、本明細書に記載の方法の1つを実行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または信号のシーケンスである。データストリームまたは信号のシーケンスは例えば、例えばインターネットを介して等データ通信接続を介して伝送されるように構成することができる。

40

【 0 1 0 9 】

別の実施態様は、例えばコンピュータやプログラム可能な論理装置等、本明細書に記載の方法の1つを実行するように構成された、あるいは適応された、処理手段を含む。

【 0 1 1 0 】

別の実施態様は、本明細書に記載の方法の1つを実行するためのコンピュータプログラムをインストールしたコンピュータを含む。

【 0 1 1 1 】

一部の実施態様においては、本明細書に記載の方法の機能の一部または全部を実行する

50

ために、プログラム可能な論理装置（例えばフィールドプログラマブルゲートアレイ）を使用することができる。一部の実施態様においては、本明細書に記載の方法の1つを実行するために、フィールドプログラマブルゲートアレイがマイクロプロセッサと協働することがある。概して、方法は好ましくは何らかのハードウェア装置によって実行される。

【0112】

上記に記載の実施態様は本発明の原則に対して単に例証するものである。本明細書に記載の構成および詳細の改変および変形は、当業者に自明であるものと理解される。したがって、後述の特許請求の範囲にのみ限定され、本明細書に記載の実施態様の記載および説明によってあらわされる具体的な詳細に限定されることを意図するものではない。

【0113】

参考文献

- [1] Dietz, M., Liljeryd, L., Kjoerling, K., and Kunz, O., "Spectral Band Replication, a Novel Approach in Audio Coding," in Audio Engineering Society Convention 112, 2002.
- [2] Disch, S., Niedermeier, A., Helmrich, C. R., Neukam, C., Schmidt, K., Geiger, R., Lecomte, J., Ghido, F., Nagel, F., and Edler, B., "Intelligent Gap Filling in Perceptual Transform Coding of Audio," in Audio Engineering Society Convention 141, 2016.
- [3] Laitinen M-V., Disch S., Oates C., Pulkki V. "Phase derivative correction of bandwidth extended signals for perceptual audio codecs." In 140th Audio Engineering Society International Convention 2016, AES 2016. Audio Engineering Society. 2016.
- [4] Atti, Venkatraman, Venkatesh Krishnan, Duminda A. Dewasurendra, Venkata Chebiyyam, Shaminda Subasingha, Daniel J. Sinder, Vivek Rajendran, Imre Varga, Jon Gibbs, Lei Miao, Volodya Grancharov and Harald Pobloth. "Super-wideband bandwidth extension for speech in the 3GPP EVS codec." 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2015): 5927-5931.

10

20

30

40

50

【図面】

【図 1】

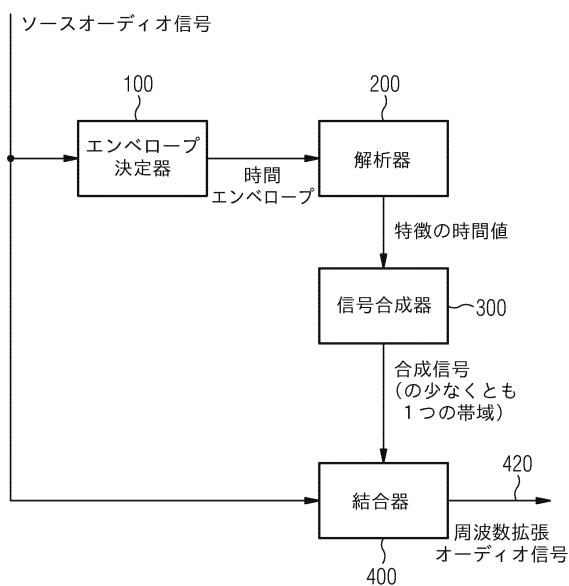


Fig. 1

【図 2】

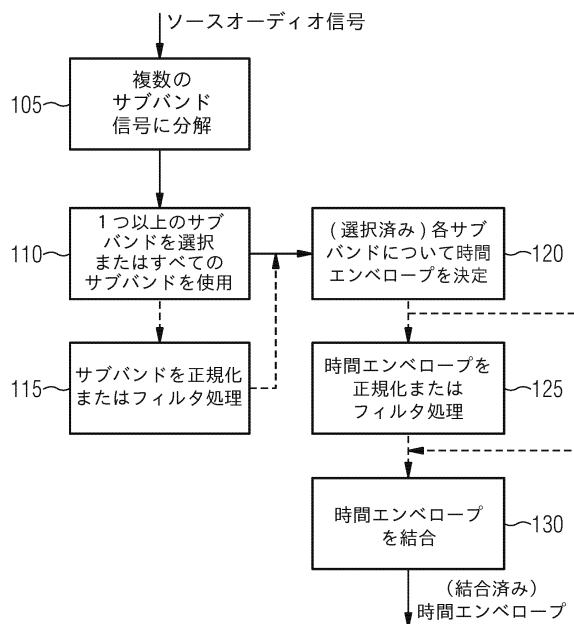


Fig. 2

【図 3 a】

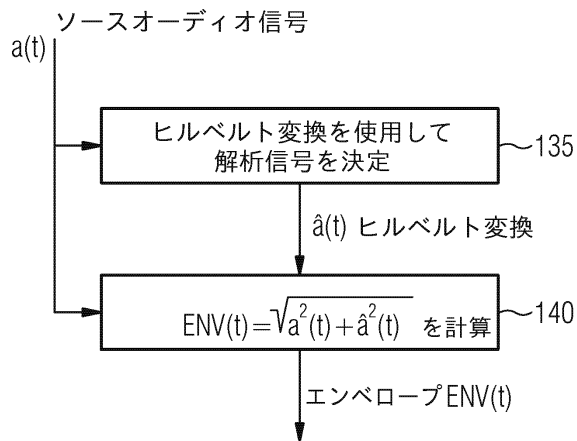


Fig. 3a

【図 3 b】

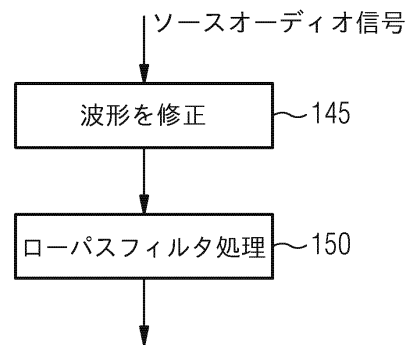


Fig. 3b

10

20

30

40

50

【 図 3 c 】

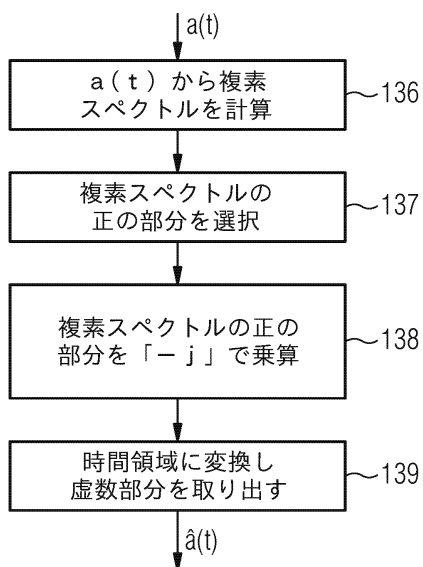


Fig. 3c

【圖 4】

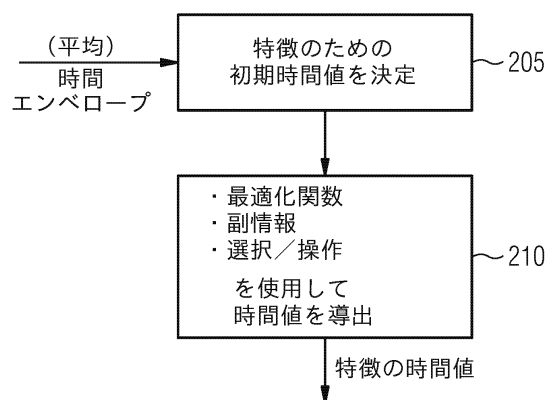


Fig. 4

【圖 5】

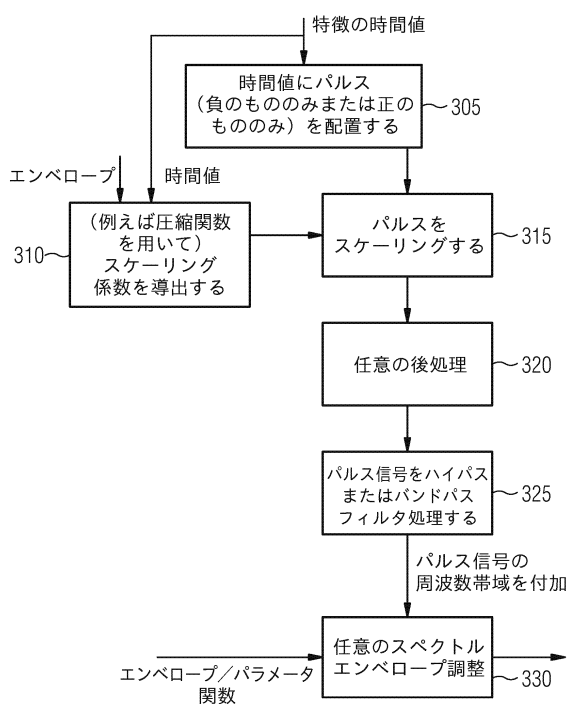


Fig. 5

【圖 6】

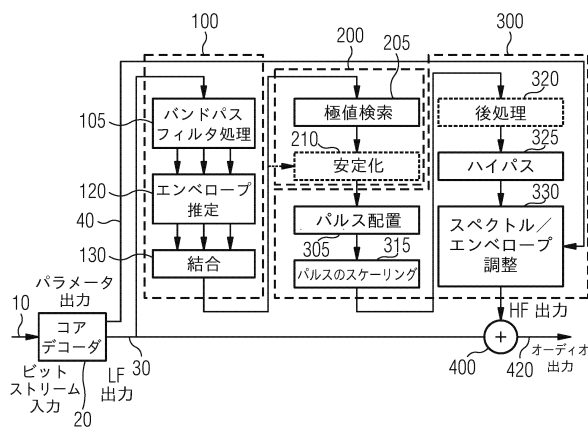


Fig. 6

【図 7】

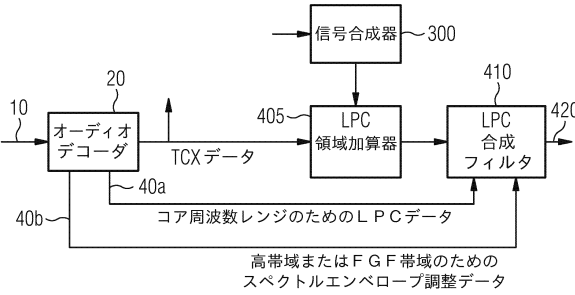


Fig. 7

【図 8】

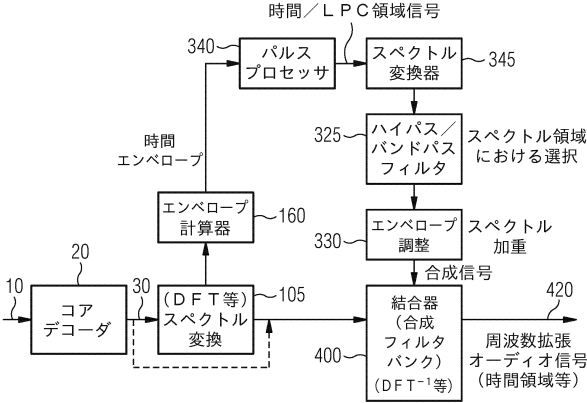


Fig. 8

【図 9 a】

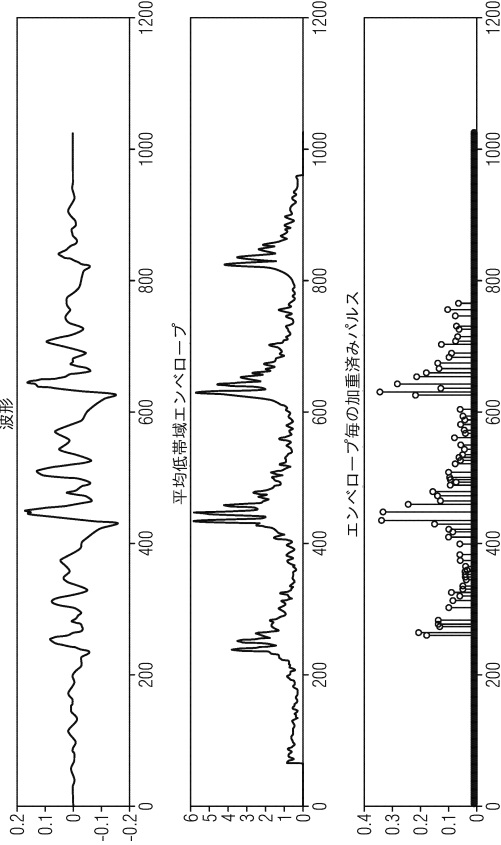


Fig. 9a

【図 9 b】

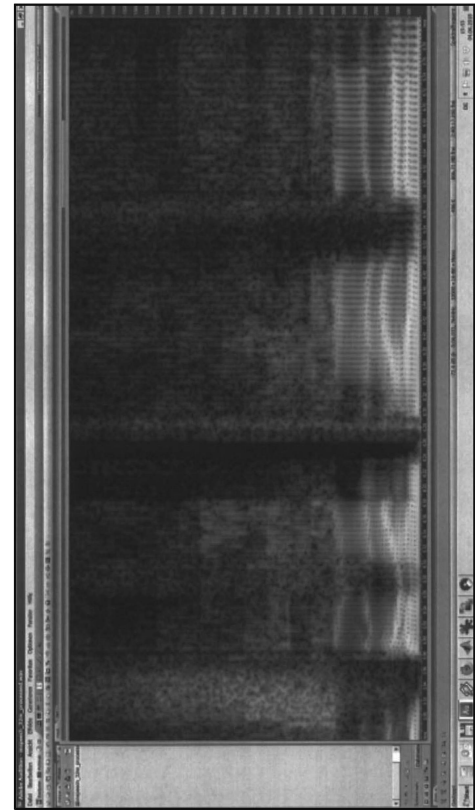


Fig. 9b

10

20

30

40

50

【図 10 a】

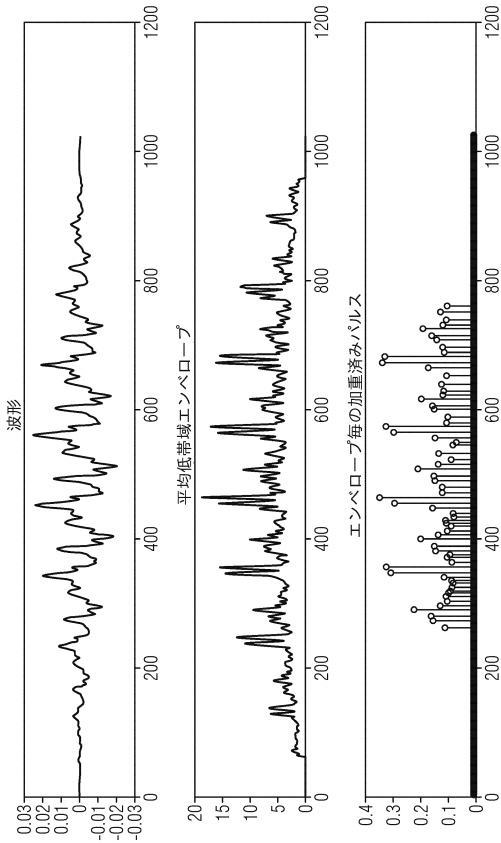


Fig. 10a

【図 10 b】

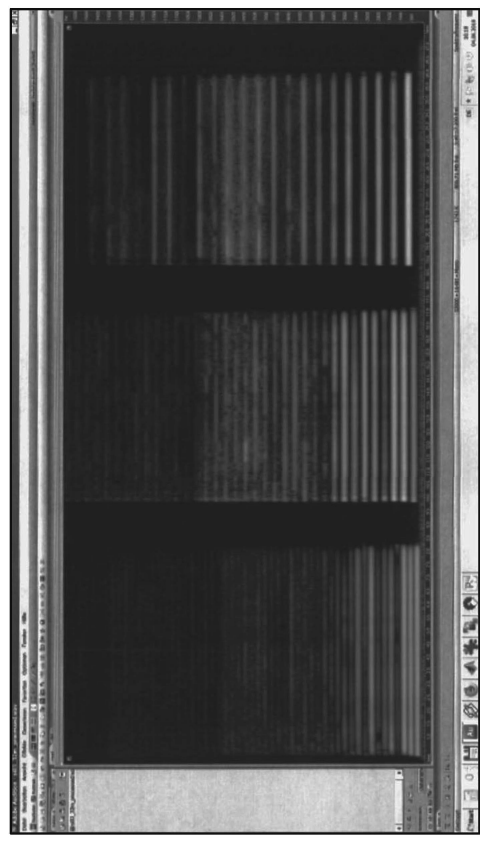


Fig. 10b

【図 11 a】

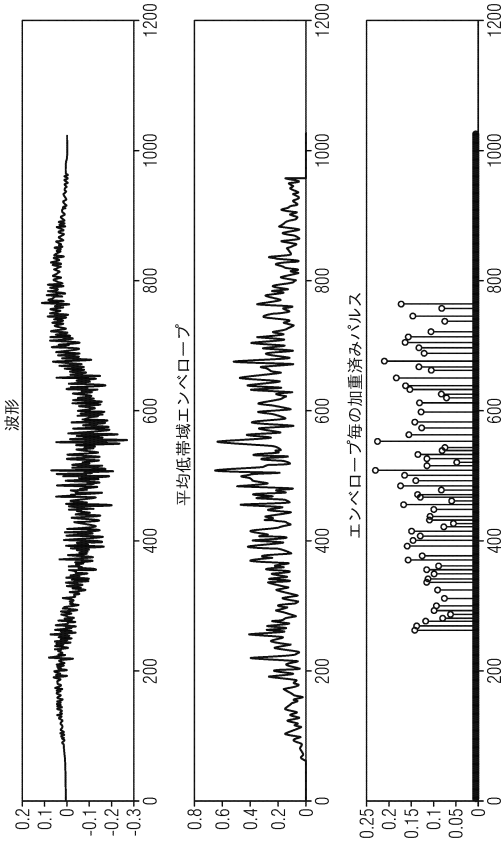


Fig. 11a

【図 11 b】

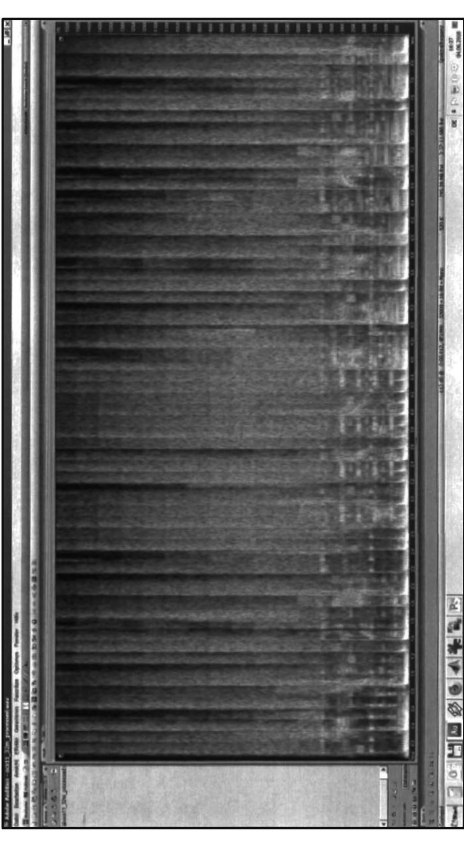


Fig. 11b

10

20

30

40

50

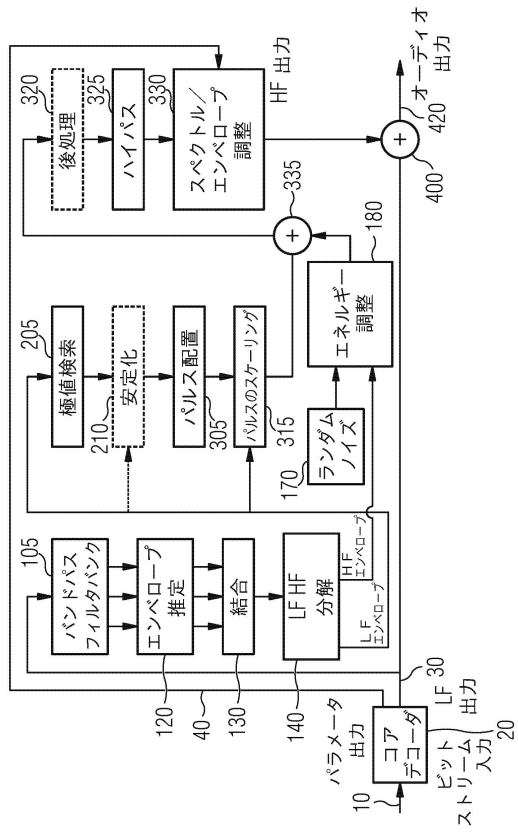


Fig. 12

フロントページの続き

(33)優先権主張国・地域又は機関

欧州特許庁(EP)

ゲン I I S 内

(72)発明者 シュトゥルム ミヒャエル

ドイツ連邦共和国 9 1 0 5 8 エアランゲン アム ヴォルフスマンテル 3 3 フラウンホッフ
ー - インスティテュート フュア インテグリールテ シャルトウンゲン I I S 内

審査官 山下 剛史

(56)参考文献 特開平 8 - 3 0 5 3 9 6 (J P , A)

特表 2 0 1 6 - 5 4 1 0 1 2 (J P , A)

特開 2 0 0 9 - 1 0 4 0 1 5 (J P , A)

特表 2 0 0 1 - 5 2 1 6 4 8 (J P , A)

(58)調査した分野 (Int.Cl. , D B 名)

G 1 0 L 1 9 / 0 0 - 2 5 / 9 3

I E E E X p l o r e