

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6453249号  
(P6453249)

(45) 発行日 平成31年1月16日(2019.1.16)

(24) 登録日 平成30年12月21日(2018.12.21)

(51) Int.Cl.		F I	
G 1 O L 19/26	(2013.01)	G 1 O L 19/26	B
G 1 O L 19/22	(2013.01)	G 1 O L 19/22	
G 1 O L 19/08	(2013.01)	G 1 O L 19/08	Z

請求項の数 32 (全 34 頁)

(21) 出願番号	特願2015-560497 (P2015-560497)	(73) 特許権者	504226412 ヴォイスエイジ・コーポレーション カナダ・ケベック・H3R・2H6・モン トリオール・ルサヌ・ロード・750・ スイート・250
(86) (22) 出願日	平成26年1月9日(2014.1.9)	(74) 代理人	100108453 弁理士 村山 靖彦
(65) 公表番号	特表2016-513812 (P2016-513812A)	(74) 代理人	100110364 弁理士 実広 信哉
(43) 公表日	平成28年5月16日(2016.5.16)	(74) 代理人	100133400 弁理士 阿部 達彦
(86) 国際出願番号	PCT/CA2014/000014	(72) 発明者	トミー・ヴァイヤンクール カナダ・ケベック・J1N・2K1・シャ ーブルック・リュ・ファラン・835
(87) 国際公開番号	W02014/134702		
(87) 国際公開日	平成26年9月12日(2014.9.12)		
審査請求日	平成29年1月6日(2017.1.6)		
(31) 優先権主張番号	61/772,037		
(32) 優先日	平成25年3月4日(2013.3.4)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 時間領域デコーダにおける量子化雑音を低減するためのデバイスおよび方法

(57) 【特許請求の範囲】

【請求項1】

合成フィルタを介して処理され、合成を生じるためのデコードされた時間領域励振に含まれる音響信号中の量子化雑音を低減するためのデバイスであって、前記デバイスは、  
合成前に、前記デコードされた時間領域励振の周波数領域励振への変換器と、  
前記周波数領域励振にตอบสนองして、前記量子化雑音中に失われたスペクトル情報を取り出すための重み付けマスクを生じるためのマスクビルダと、  
前記重み付けマスクの前記周波数領域励振への適用によりスペクトルダイナミクスを増大させるための前記周波数領域励振の変更器と、  
前記変更された周波数領域励振の、前記音響信号の量子化雑音が低減されたバージョンを含む変更された時間領域励振への変換器と  
を備えるデバイス。

10

【請求項2】

前記デコードされた時間領域励振の前記合成を生じるための前記合成フィルタと、  
前記デコードされた時間領域励振の前記合成の第1の組の励振カテゴリおよび第2の組の励振カテゴリのうち的一方への分類器を備え、  
前記第2の組の励振カテゴリが、INACTIVEまたはUNVOICEDカテゴリを含み、  
前記第1の組の励振カテゴリが、OTHERカテゴリを含む、請求項1に記載のデバイス。

【請求項3】

前記デコードされた時間領域励振の周波数領域励振への前記変換器は、前記デコードさ

20

れた時間領域励振の前記合成が前記第1の組の励振カテゴリに分類されたとき、前記デコードされた時間領域励振に適用される、請求項2に記載のデバイス。

【請求項4】

前記デコードされた時間領域励振の前記合成の第1の組の励振カテゴリおよび第2の組の励振カテゴリのうち的一方への前記分類器が、エンコーダから時間領域デコードに伝送され、前記時間領域デコードにおいて、デコードされたビットストリームから取り出された分類情報を使用する、請求項2または3に記載のデバイス。

【請求項5】

前記変更された時間領域励振の合成を生じるために第2の合成フィルタを備える、請求項2から4のいずれか一項に記載のデバイス。

10

【請求項6】

前記デコードされた時間領域励振の前記合成および前記変更された時間領域励振の前記合成のうち的一方から前記音響信号を生成するためにディエンファサイジングフィルタおよびリサンブラを備える、請求項5に記載のデバイス。

【請求項7】

前記デコードされた時間領域励振の前記合成が、前記第2の組の励振カテゴリに分類されたときには、前記デコードされた時間領域励振の前記合成として、

前記デコードされた時間領域励振の前記合成が、前記第1の組の励振カテゴリに分類されたときには、前記変更された時間領域励振の前記合成として出力合成を選択するための2段分類器を備える、請求項5または6に記載のデバイス。

20

【請求項8】

前記周波数領域励振が音楽を含むかどうかを決定するために前記周波数領域励振の分析器を備える、請求項1から7のいずれか一項に記載のデバイス。

【請求項9】

前記周波数領域励振の前記分析器が、前記周波数領域励振のスペクトルエネルギー差の統計偏差を閾値と比較することにより、前記周波数領域励振が音楽を含むことを決定する、請求項8に記載のデバイス。

【請求項10】

将来のフレームの励振を評価するために励振外挿器を備え、前記変更された周波数領域励振の変更された時間領域励振への変換が遅延なしである、請求項1から9のいずれか一項に記載のデバイス。

30

【請求項11】

前記デコードされた時間領域励振の前記周波数領域励振への前記変換器に供給される、過去のフレーム、現在のフレーム、および外挿された将来のフレーム時間領域励振の励振連結器を備える、請求項10に記載のデバイス。

【請求項12】

前記マスクビルダが、時間平均化または周波数平均化、または時間および周波数平均化の組合せを使用して前記重み付けマスクを生じる、請求項1から11のいずれか一項に記載のデバイス。

【請求項13】

前記マスクビルダが、  
スケールされたエネルギースペクトルを生じるための前記周波数領域励振のスペクトルエネルギーの正規化器と、  
周波数軸に沿った、前記スケールされたエネルギースペクトルの平均化器と、  
フレーム間の周波数スペクトル値を平滑化するための、時間領域軸に沿った、前記平均化されたエネルギースペクトルの平滑化器と  
 を備える、請求項1から12のいずれか一項に記載のデバイス。

40

【請求項14】

前記正規化器は、正規化されたエネルギースペクトルを生じ、累乗の値を前記正規化されたエネルギースペクトルに適用して前記スケールされたエネルギースペクトルを生

50

じ、前記スケーリングされたエネルギースペクトルの値を上限に制限する、請求項13に記載のデバイス。

【請求項15】

前記デコードされた時間領域励振の選択された帯域において信号対雑音比を推定するために、および前記信号対雑音比に基づいて周波数領域雑音低減を実施するために雑音低減装置を備える、請求項1から14のいずれか一項に記載のデバイス。

【請求項16】

前記デコードされた時間領域励振は、デコードされたCELP時間領域励振であり、前記合成フィルタは、LP合成フィルタである、請求項1から15のいずれか一項に記載のデバイス

。

【請求項17】

合成フィルタを介して処理され、合成を生じるためのデコードされた時間領域励振に含まれる音響信号中の量子化雑音を低減するための方法であって、前記方法は、

合成前に、時間領域から周波数領域への変換器を使用して、前記デコードされた時間領域励振を周波数領域励振に変換するステップと、

マスクビルダを使用して、かつ前記周波数領域励振にตอบสนองして、前記量子化雑音中に失われたスペクトル情報を取り出すための重み付けマスクを生じるステップと、

前記重み付けマスクの前記周波数領域励振への適用によりスペクトルダイナミクスを増大させるために前記周波数領域励振を変更するステップと、

周波数領域から時間領域への変換器を使用して、前記変更された周波数領域励振を、前記音響信号の量子化雑音が低減されたバージョンを含む変更された時間領域励振に変換するステップと

を含む方法。

【請求項18】

前記デコードされた時間領域励振の前記合成を生じるための前記合成フィルタを介して、前記デコードされた時間領域励振を処理するステップと、

前記デコードされた時間領域励振の前記合成を第1の組の励振カテゴリおよび第2の組の励振カテゴリのうち的一方に分類するステップを含み、

前記第2の組の励振カテゴリが、INACTIVEまたはUNVOICEDカテゴリを含み、

前記第1の組の励振カテゴリが、OTHERカテゴリを含む、請求項17に記載の方法。

【請求項19】

前記デコードされた時間領域励振の前記合成が前記第1の組の励振カテゴリに分類されたとき、前記デコードされた時間領域励振の周波数領域励振への変換を前記デコードされた時間領域励振に適用するステップを含む、請求項18に記載の方法。

【請求項20】

前記デコードされた時間領域励振の前記合成を第1の組の励振カテゴリおよび第2の組の励振カテゴリのうちの前記一方に分類するために、エンコーダから時間領域デコーダに伝送され、前記時間領域デコーダにおいて、デコードされたビットストリームから取り出された分類情報を使用するステップを含む、請求項18または19に記載の方法。

【請求項21】

前記変更された時間領域励振の合成を生じるステップを含む、請求項18から20のいずれか一項に記載の方法。

【請求項22】

前記デコードされた時間領域励振の前記合成および前記変更された時間領域励振の前記合成のうち的一方から前記音響信号を生成するステップを含む、請求項21に記載の方法。

【請求項23】

前記デコードされた時間領域励振の前記合成が前記第2の組の励振カテゴリに分類されたときには、前記デコードされた時間領域励振の前記合成として、

前記デコードされた時間領域励振の前記合成が前記第1の組の励振カテゴリに分類されたときには、前記変更された時間領域励振の前記合成として出力合成を選択するステップ

10

20

30

40

50

を含む、請求項21または22に記載の方法。

【請求項24】

前記周波数領域励振が音楽を含むかどうかを決定するために、前記周波数領域励振を分析するステップを含む、請求項17から23のいずれか一項に記載の方法。

【請求項25】

前記周波数領域励振のスペクトルエネルギー差の統計偏差を閾値と比較することにより、前記周波数領域励振が音楽を含むことを決定するステップを含む、請求項24に記載の方法。

【請求項26】

将来のフレームの外挿された励振を評価するステップを含み、前記変更された周波数領域励振の変更された時間領域励振への変換が遅延なしである、請求項17から25のいずれか一項に記載の方法。

10

【請求項27】

前記周波数領域励振へ変換するために、過去のフレーム、現在のフレーム、および外挿された将来のフレーム時間領域励振を連結するステップを含む、請求項26に記載の方法。

【請求項28】

前記重み付けマスクを、時間平均化または周波数平均化または時間および周波数平均化の組合せを使用して生じる、請求項17から27のいずれか一項に記載の方法。

【請求項29】

重み付けマスクを生じるステップは、  
スケールされたエネルギースペクトルを生じるために前記周波数領域励振のスペクトルエネルギーを正規化するステップと、  
周波数軸に沿って、前記スケールされたエネルギースペクトルを平均化するステップと、

20

フレーム間の周波数スペクトル値を平滑化するために、時間領域軸に沿って、前記平均化されたエネルギースペクトルを平滑化するステップと  
 を含む、請求項17から28のいずれか一項に記載の方法。

【請求項30】

前記周波数領域励振の前記スペクトルエネルギーを正規化するステップは、正規化されたエネルギースペクトルを生じるステップと、累乗の値を前記正規化されたエネルギースペクトルに適用して前記スケールされたエネルギースペクトルを生じるステップと、前記スケールされたエネルギースペクトルの値を上限に制限するステップとを含む、  
請求項29に記載の方法。

30

【請求項31】

前記デコードされた時間領域励振の選択された帯域において信号対雑音比を推定するステップと、

前記推定された信号対雑音比に基づいて周波数領域雑音低減を実施するステップとを含む、請求項17から30のいずれか一項に記載の方法。

【請求項32】

前記デコードされた時間領域励振は、デコードされたCELP時間領域励振であり、前記合成フィルタは、LP合成フィルタである、請求項17から31のいずれか一項に記載の方法。

40

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、音響処理の分野に関する。より具体的には、本開示は、音響信号中の量子化雑音を低減することに関する。

【背景技術】

【0002】

現在の会話型コーデックは、8kbps程度のビットレートにおいてきれいな音声信号を非常に良い品質で表現し、16kbpsのビットレートにおいて透明性に近づく。この高い音声品

50

質を低ビットレートで維持するために、マルチモーダル符号化スキームが一般に使用される。通常、入力信号はその特性を反映する異なるカテゴリの間で分割される。異なるカテゴリは、例えば、有声音声、無声音声、有声音オンセットなどを含む。次いで、コーデックは、これらのカテゴリに最適化された異なる符号化モードを使用する。

【0003】

音声モデルベースのコーデックは、通常、音楽などの汎用オーディオ信号をうまくレンダリングしない。したがって、一部の展開された音声コーデックは、特に低いビットレートにおいて良い品質で音楽を表現しない。コーデックが展開されたとき、ビットストリームが標準化されており、ビットストリームに何らかの変更を加えると、コーデックの相互運用性が破壊されることにより、エンコーダを変更することは困難である。

10

【0004】

したがって、音声モデルベースのコーデック、例えば、線形予測(LP)ベースのコーデックの音楽コンテンツレンダリングを改善することが必要とされている。

【先行技術文献】

【特許文献】

【0005】

【特許文献1】PCT特許公開WO 2009/109050 A1

【特許文献2】PCT特許公開WO 2003/102921 A1

【特許文献3】PCT特許公開WO 2007/073604 A1

【特許文献4】PCT国際出願PCT/CA2012/001011

20

【非特許文献】

【0006】

【非特許文献1】「Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding Functions」という名称のTechnical Specification (TS) 26.190 of the 3rd Generation Partnership Program (3GPP)

【非特許文献2】J. D. Johnston「Transform coding of audio signal using perceptual noise criteria」、IEEE J. Select. Areas Commun.、vol. 6、314～323ページ、1988年2月

【発明の概要】

【課題を解決するための手段】

30

【0007】

本開示によれば、時間領域デコーダによってデコードされた時間領域励振に含まれる信号中の量子化雑音を低減するためのデバイスが提供される。デバイスは、デコードされた時間領域励振の周波数領域励振への変換器を備える。また、量子化雑音中に失われたスペクトル情報を取り出すための重み付けマスクを生じるマスクビルダも含まれる。デバイスは、重み付けマスクの適用によりスペクトルダイナミクスを増大させるための周波数領域励振の変更器も備える。デバイスは、変更された周波数領域励振の変更された時間領域励振への変換器をさらに備える。

【0008】

本開示は、時間領域デコーダによってデコードされた時間領域励振に含まれる信号中の量子化雑音を低減するための方法にも関する。デコードされた時間領域励振は、時間領域デコーダによって周波数領域励振に変換される。重み付けマスクが、量子化雑音中に失われたスペクトル情報を取り出すために生じる。周波数領域励振は、重み付けマスクの適用によりスペクトルダイナミクスを増大させるために変更される。変更された周波数領域励振は、変更された時間領域励振に変換される。

40

【0009】

前述のおよび他の特徴は、添付の図面を参照して、例だけとして与えられる、以下の、それらの例示的实施形態の非限定的説明を読めばより明らかとなるであろう。

【図面の簡単な説明】

【0010】

50

本開示の実施形態は、添付の図面を参照して例だけとして説明される。

【0011】

【図1】一実施形態による、時間領域デコーダによってデコードされた時間領域励振に含まれる信号中の量子化雑音を低減するための方法の動作を示すフローチャートである。

【図2A】音楽信号および他の音響信号中の量子化雑音を低減するための周波数領域後処理機能を有するデコーダの簡略化された回路図であり、図2bと合わせて図2と呼ぶ。

【図2B】音楽信号および他の音響信号中の量子化雑音を低減するための周波数領域後処理機能を有するデコーダの簡略化された回路図であり、図2aと合わせて図2と呼ぶ。

【図3】図2のデコーダを形成するハードウェア構成要素の構成例の簡略化された構成図である。

【発明を実施するための形態】

【0012】

本開示の様々な態様は、一般に、音楽信号中の量子化雑音を低減することにより、音声モデルベースのコーデック、例えば、線形予測(LP)ベースのコーデックの音楽コンテンツレンダリングを改善する課題の1つまたは複数に対処する。本開示の教示は、他の音響信号、例えば、音楽以外の汎用オーディオ信号にも適用できることに留意すべきである。

【0013】

デコーダの変更は、受信側の知覚品質を改善することができる。本開示は、デコーダ側で、デコードされた合成のスペクトル中の量子化雑音を低減する音楽信号および他の音響信号の周波数領域後処理を実現するための取組みを開示する。後処理は任意の追加の符号化遅延なしで実現することができる。

【0014】

本明細書に使用されるスペクトルハーモニクスと周波数後処理との間の量子化雑音の周波数領域除去の原理は、その開示が参照により本明細書に組み込まれる、2009年9月11日付のVaillancourtらへのPCT特許公開WO 2009/109050 A1(以下「Vaillancourt'050」)に基づく。概して、そのような周波数後処理は、デコードされた合成に適用され、オーバーラップを含め、処理を追加して顕著な品質利得を得るために処理遅延の増大を必要とする。さらに、従来の周波数領域後処理の場合、限定された周波数分解能により、追加される遅延がより短ければ短いほど(すなわち、変換窓がより短ければ短いほど)、後処理がより効果的でなくなる。本開示によれば、周波数後処理は、合成に遅延を追加することなく、より高い周波数分解能を達成する(より長い周波数変換が使用される)。さらに、過去のフレームスペクトルエネルギー中に存在する情報を利用して、符号化雑音中に失われたスペクトル情報を取り出す、すなわち強化するために現在のフレームスペクトルに適用される重み付けマスクを生じる。合成に遅延を追加することなくこの後処理を達成するために、この例では、対称台形窓が使用される。窓が平坦である(定数値が1である)現在のフレームを中心にし、外挿を使用して将来の信号を作製する。後処理は、一般に、任意のコーデックの合成信号に直接適用され得るが、本開示は、後処理を、3GPPのウェブサイト上で入手可能な、参照によりその全内容が本明細書に組み込まれる、「Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding Functions」という名称のTechnical Specification (TS) 26.190 of the 3rd Generation Partnership Program (3GPP)に記載の符号励振線形予測(CELP)コーデックの枠組みにおける励振信号に適用する、例示的实施形態を導入する。合成信号ではなく励振信号に取り組み利点は、後処理によって導入された任意の潜在的断絶が、その後のCELP合成フィルタの適用によって平滑化されることである。

【0015】

本開示においては、12.8kHzの内部サンプリング周波数を有するAMR-WBが例示のために使用される。しかし、本開示は、合成が、合成フィルタ、例えば、LP合成フィルタを通してフィルタリングされた励振信号によって取得される、他の低いビットレートの音声デコーダに適用することができる。合成は、音楽が時間領域励振と周波数領域励振との組合せを用いて符号化される、マルチモーダルコーデックにも適用することができる。次の数行はポストフィルタの動作をまとめたものである。AMR-WBを使用する例示的实施形態の詳細

10

20

30

40

50

な説明がその後続く。

【0016】

まず、完全なビットストリームは、デコードされ、現在のフレーム合成が、その開示が参照により本明細書に組み込まれる、Jelinekらへの2003年12月11日付のPCT特許公開WO 2003/102921 A1、Vaillancourtらへの2007年7月5日付のPCT特許公開WO 2007/073604 A1、およびVaillancourtらの名前で2012年11月1日に出願したPCT国際出願PCT/CA2012/001011(以下「Vaillancourt'011」)に開示されたものと同様の第1の段の分類器を通して処理される。本開示のために、この第1の段の分類器は、フレームを分析し、INACTIVEフレームと、UNVOICEDフレーム、例えば、アクティブなUNVOICED音声に対応するフレームとを分離する。第1の段においてINACTIVEフレームとしてまたはUNVOICEDフレームとして分類されないすべてのフレームは、第2の段の分類器を用いて分析される。第2の段の分類器は、後処理を適用するかどうか、およびどの程度まで適用するかを判定する。後処理が適用されないとき、メモリに関連する後処理だけが更新される。

10

【0017】

第1の段の分類器によってINACTIVEフレームとしてまたはアクティブなUNVOICED音声フレームとして分類されないすべてのフレームには、過去のデコードされた励振、現在のフレームのデコードされた励振、および将来の励振の外挿を使用して、ベクトルが形成される。過去のデコードされた励振と外挿された励振との長さは同じであり、周波数変換の所望の分解能によって異なる。この例においては、使用される周波数変換の長さは、640サンプルである。過去のおよび外挿された励振を用いてベクトルを生じることにより、周波数の分解能を増大させることが可能となる。本例においては、過去のおよび外挿された励振の長さは同じであるが、窓の対称性は必ずしもポストフィルタが効率的に働くのには必要でない。

20

【0018】

連結励振(過去のデコードされた励振、現在のフレームのデコードされた励振および将来の励振の外挿を含む)の周波数表現のエネルギー安定性は、次に、音楽の存在下での確率を決定するために第2の段の分類器を用いて分析される。この例においては、音楽の存在下であることの決定は2段プロセスで実施される。しかし、音楽の検出は、例えば、周波数変換前に単一の動作で実施され得る、またはエンコーダで決定されビットストリームで伝送されさえするなど、異なるやり方で実施され得る。

30

【0019】

高調波間量子化雑音は、周波数ビンごとの信号対雑音比(SNR)を推定することにより、およびそのSNRにより各周波数ビンに利得を適用することにより、Vaillancourt'050の場合と同様に低減される。しかし、本開示においては、雑音エネルギー推定は、Vaillancourt'050において教示されるものとは異なって行われる。

【0020】

次いで、符号化雑音中に失われた情報を取り出し、スペクトルのダイナミクスをさらに増大させる、追加の処理が使用される。このプロセスは、エネルギースペクトルの0と1との間の正規化から始まる。次いで、定数オフセットが正規化エネルギースペクトルに追加される。最後に、8の累乗が変更されたエネルギースペクトルの各周波数ビンに適用される。その結果得られるスケーリングされたエネルギースペクトルが、低周波数から高周波数まで、周波数軸に沿って平均化関数により処理される。最後に、時間とともにスペクトルの長時間の平滑化がビンごとに実施される。

40

【0021】

この処理の第2の部分は、結果として、ピークが重要なスペクトル情報に対応し、谷が符号化雑音に対応するマスクとなる。次いで、このマスクを使用して、雑音をフィルタリング除去し、ピーク領域におけるスペクトルビンの大きさをわずかに増大させることにより、スペクトルダイナミクスを増大させ、谷におけるビンの大きさを減衰させ、したがって、ピーク対谷の比を増大させる。これらの2つの動作は、出力合成に遅延を追加せずに、高い周波数分解能を使用して行われる。

50

## 【 0 0 2 2 】

連結励振ベクトルの周波数表現が強調された(その雑音が低減され、そのスペクトルダイナミクスが増大された)後、逆周波数変換を実施して、連結励振の強化バージョンを製作する。本開示においては、現在のフレームに対応する変換窓の部分は、実質的に平坦であり、過去のおよび外挿された励振信号に適用される窓の部分だけを漸減する必要がある。これにより、逆変換の後、強化された励振の現在のフレームを根絶することが可能になる。この最後の操作は、時間領域の強化された励振を現在のフレームの位置における長方形窓で乗じることと同様である。この動作は合成領域において行うと、重要なブロックアーチファクトを追加するが、Vaillancourt'011に示すように、LP合成フィルタが1つのブロックから別のブロックへの移行を平滑化するのに役立つので、これは励振領域において代替的に行うことができる。

10

## 【 0 0 2 3 】

例示的なAMR-WB実施形態の説明

ここに説明する後処理は、音楽や残響音声などの信号のLP合成フィルタのデコードされた励振に適用される。信号の性質(音声、音楽、残響音声など)に関する判定および後処理を適用することに関する判定は、AMR-WBビットストリームの一部としてデコーダ分類情報に向けて送るエンコーダによって信号伝達することができる。そうでない場合には、信号分類は、デコーダ側で代替的に行うことができる。複雑性と分類信頼性とのトレードオフにより、合成フィルタは、一時的合成およびより良好な分類分析を得るために現在の励振に任意選択で適用することができる。この構成においては、合成は、分類が結果としてポストフィルタリングが適用されるカテゴリになる場合、上書きされる。複雑性が増すのを最小限にするために、分類は、過去のフレーム合成で行うこともでき、合成フィルタは、後処理の後、一度適用される。

20

## 【 0 0 2 4 】

次に図面を参照すると、図1は一実施形態による時間領域デコーダによってデコードされた時間領域励振に含まれる信号中の量子化雑音を低減するための方法の動作を示すフローチャートである。図1においては、シーケンス10が、可変順序で実行することができる複数の動作を含み、動作のうちの一部は場合により同時に実行され、動作のうちの一部は任意選択である。動作12では、時間領域デコーダはエンコーダによって生じたビットストリームを取り出し、デコードし、ビットストリームは、時間領域励振を再構成するのに使用可能なパラメータの形態での時間領域励振情報を含む。このため、時間領域デコーダは、入力インターフェースを介してビットストリームを受け取るまたはメモリからビットストリームを読み取ることができる。時間領域デコーダは、動作16において、デコードされた時間領域励振を周波数領域励振に変換する。動作16において励振信号を時間領域から周波数領域に変換する前に、動作14において、将来の時間領域励振を外挿することができ、したがって、時間領域励振から周波数領域励振への変換が遅延なしとなる。すなわち、余分な遅延を必要とすることなく、より良い周波数分析が実施される。このため、過去の、現在のおよび予測される将来の時間領域励振信号は、周波数領域に変換される前に連結することができる。時間領域デコーダは、次いで、動作18において、量子化雑音中に失われたスペクトル情報を取り出すための重み付けマスクを生じる。動作20では、時間領域デコーダは、重み付けマスクの適用によりスペクトルダイナミクスを増大させるために周波数領域励振を変更する。動作22では、時間領域デコーダは、変更された周波数領域励振を変更された時間領域励振に変換する。時間領域デコーダは、次いで、動作24において変更された時間領域励振の合成を生じ、動作26において、デコードされた時間領域励振の合成および変更された時間領域励振の合成のうち的一方から音響信号を生成することができる。

30

40

## 【 0 0 2 5 】

図1に示す方法は、いくつかの任意選択の特徴を使用して適合させることができる。例えば、デコードされた時間領域励振の合成は、第1の組の励振カテゴリおよび第2の組の励振カテゴリのうち的一方に分類することができ、その場合、第2の組の励振カテゴリは、INACTIVEまたはUNVOICEDカテゴリを含み、第1の組の励振カテゴリはOTHERカテゴリを含む

50

。デコードされた時間領域励振から周波数領域励振への変換は、第1の組の励振カテゴリに分類されたデコードされた時間領域励振に適用することができる。取り出されたビットストリームは、デコードされた時間領域励振の合成を第1の組の励振カテゴリまたは第2の組の励振カテゴリのいずれかに分類するのに使用可能な分類情報を含むことができる。音響信号を生成するために、出力合成を、時間領域励振が第2の組の励振カテゴリに分類されたときには、デコードされた時間領域励振の合成として選択することができ、時間領域励振が第1の組の励振カテゴリに分類されたときには、変更された時間領域励振の合成として選択することができる。周波数領域励振は、周波数領域励振が音楽を含むかどうかを決定するために分析することができる。具体的には、周波数領域励振が音楽を含むことを決定するには、周波数領域励振のスペクトルエネルギー差の統計偏差を閾値と比較することを利用できる。重み付けマスクは、時間平均化または周波数平均化または両方の組合せを使用して生じさせることができる。信号対雑音比が、デコードされた時間領域励振の選択された帯域に対して推定することができ、周波数領域の雑音低減を推定信号対雑音比に基づいて実施することができる。

#### 【 0 0 2 6 】

図2aおよび図2bは、音楽信号および他の音響信号中の量子化雑音を低減するための周波数領域後処理機能を有するデコーダの簡略化された回路図であり、両図を合わせて図2と呼ぶ。デコーダ100が図2aおよび図2bに示すいくつかの要素を備え、これらの要素は、図示するように矢印によって相互接続され、相互接続のうちの一部は、図2aの一部の要素がどのように図2bの他の要素に関連しているかを示す、コネクタA、B、C、DおよびEを使用して示される。デコーダ100は、例えば、無線通信インターフェースを介してエンコーダからAMR-WBビットストリームを受け取る受信機102を備える。あるいは、デコーダ100は、ビットストリームを記憶したメモリ(図示せず)に動作可能に接続することができる。デマルチプレクサ103が、時間領域励振、ピッチラグ情報および音声区間検出(VAD)情報を再構成するためにビットストリームから時間領域励振パラメータを抽出する。デコーダ100は、時間領域励振パラメータを受け取って、現在のフレームの時間領域励振をデコードする時間領域励振デコーダ104と、過去の励振バッファメモリ106と、2つのLP合成フィルタ108および110と、VAD信号を受け取る信号分類推定器114およびクラス選択テストポイント116を備えた第1の段の信号分類器112と、ピッチラグ情報を受け取る励振外挿器118と、励振連結器120と、窓掛けおよび周波数変換モジュール122と、第2の段の信号分類器124としてのエネルギー安定性分析器と、帯域ごとの雑音レベル推定器126と、雑音低減装置128と、スペクトルエネルギー正規化器131、エネルギー平均化器132およびエネルギー平滑化器134を備えたマスクビルダ130と、スペクトルダイナミクス変更器136と、周波数/時間領域変換器138と、フレーム励振抽出器140と、スイッチ146を制御する判定テストポイント144を備えた上書き器142と、ディエンファサイジングフィルタおよびリサンプラ148とを備える。判定テストポイント144によって行われた上書き判定が、第1の段の信号分類器112から取得されたINACTIVEまたはUNVOICEDの分類および第2の段の信号分類器124から取得された音響信号カテゴリ $e_{CAT}$ に基づいて、LP合成フィルタ108からのコア合成信号150またはLP合成フィルタ110からの変更された、すなわち、強化された合成信号152が、ディエンファサイジングフィルタおよびリサンプラ148に供給されるかどうかを決定する。ディエンファサイジングフィルタおよびリサンプラ148の出力は、アナログ信号を提供するデジタル/アナログ(D/A)変換器154に供給され、増幅器156によって増幅され、可聴音響信号を生成するスピーカ158にさらに提供される。あるいは、ディエンファサイジングフィルタおよびリサンプラ148の出力は、通信インターフェース(図示せず)を経てデジタル形式で伝送する、またはメモリ(図示せず)内に、コンパクトディスク上に、または任意の他のデジタル記憶媒体上にデジタル形式で記憶させることができる。別の代替として、D/A変換器154の出力は、直接かまたは増幅器を通してかのいずれかで、イヤホン(図示せず)に提供することができる。さらに別の代替として、D/A変換器154の出力は、アナログ媒体(図示せず)上に記録するまたはアナログ信号として通信インターフェース(図示せず)を介して伝送することができる。

10

20

30

40

50

## 【 0 0 2 7 】

以下の段落では、図2のデコーダ100の様々な構成要素によって実施される動作の詳細を提供する。

## 【 0 0 2 8 】

## 1) 第1の段の分類

例示的な実施形態においては、第1の段の分類は、デマルチプレクサ103からのVAD信号のパラメータに応答して、第1の段の分類器112におけるデコーダにおいて実施される。デコーダの第1の段の分類は、Vaillancourt'011の場合と同様である。以下のパラメータが、デコーダの信号分類推定器114において分類のために使用される。すなわち、正規化相関関係  $r_x$ 、スペクトル傾斜測定値  $e_t$ 、ピッチ安定性カウンタ  $pc$ 、現在のフレームの終端における信号の相対フレームエネルギー  $E_s$ 、およびゼロ交差カウンタ  $zc$  である。信号を分類するのに使用される、これらのパラメータの計算について、以下に説明する。

10

## 【 0 0 2 9 】

正規化相関関係  $r_x$  は、合成信号に基づいてフレームの終端において計算される。最後のサブフレームのピッチラグが使用される。

## 【 0 0 3 0 】

正規化相関関係  $r_x$  は、次式と同期して計算されたピッチである。

## 【 0 0 3 1 】

## 【数1】

$$r_x = \frac{\sum_{i=0}^{T-1} x(t+i)x(t+i-T)}{\sqrt{\sum_{i=0}^{T-1} x^2(t+i) \sum_{i=0}^{T-1} x^2(t+i-T)}} \quad (1)$$

20

## 【 0 0 3 2 】

ここで、 $T$  は最後のサブフレームのピッチラグ  $t=L-T$  であり、 $L$  はフレームサイズである。最後のサブフレームのピッチラグが  $3N/2$  ( $N$  はサブフレームサイズである) より大きい場合、 $T$  は最後の2つのサブフレームの平均ピッチラグに設定される。

## 【 0 0 3 3 】

相関関係  $r_x$  は、合成信号  $x(i)$  を使用して計算される。ピッチラグがサブフレームサイズ (64 サンプル) より低い場合、正規化相関関係は、 $t=L-T$  および  $t=L-2T$  の時点の2回計算され、 $r_x$  が2回の計算の平均として与えられる。

30

## 【 0 0 3 4 】

スペクトル傾斜パラメータ  $e_t$  は、エネルギーの周波数分布に関する情報を含む。本例示的な実施形態においては、デコーダにおけるスペクトル傾斜は、合成信号の第1の正規化自己相関係数として推定される。それは最後の3つのサブフレームに基づいて次式として計算される。

## 【 0 0 3 5 】

## 【数2】

$$e_t = \frac{\sum_{i=N}^{L-1} x(i)x(i-1)}{\sum_{i=N}^{L-1} x^2(i)} \quad (2)$$

40

## 【 0 0 3 6 】

ここで、 $x(i)$  は合成信号であり、 $N$  はサブフレームサイズであり、 $L$  はフレームサイズ (この例示的な実施形態においては  $N=64$  および  $L=256$ ) である。

## 【 0 0 3 7 】

50

ピッチ安定性カウンタ $p_c$ は、ピッチ周期の変動を評価する。それはデコーダにおいて次のように計算される。

$$p_c = |p_3 + p_2 - p_1 - p_0| \quad (3)$$

【 0 0 3 8 】

値 $p_0$ 、 $p_1$ 、 $p_2$ および $p_3$ は、4つのサブフレームからの閉ループピッチラグに対応する。

【 0 0 3 9 】

相対フレームエネルギー $E_s$ は、dB単位の現在のフレームエネルギーとその長時間平均との差として計算される。

$$E_s = E_f - E_{l_t} \quad (4)$$

【 0 0 4 0 】

ここで、フレームエネルギー $E_f$ は、フレームの終端において次式と同期してdBで計算されたピッチの合成信号 $s_{out}$ のエネルギーである。

【 0 0 4 1 】

【 数 3 】

$$E_f = 10 \log_{10} \left( \frac{1}{T} \sum_{i=0}^{T-1} s_{out}^2(i+L-T) \right) \quad (5)$$

【 0 0 4 2 】

ここで、 $L=256$ はフレーム長であり、 $T$ は最後の2つのサブフレームの平均ピッチラグである。 $T$ がサブフレームサイズより小さい場合、 $T$ は $2T$ (短時間ピッチラグの2つのピッチ周期を使用して計算されたエネルギー)に設定される。

【 0 0 4 3 】

長時間平均化エネルギーは、以下の関係を使用してアクティブなフレームにより更新される。

$$E_{l_t} = 0.99E_{l_t} + 0.01E_f \quad (6)$$

【 0 0 4 4 】

最後のパラメータは、1フレームの合成信号により計算されたゼロ交差パラメータ $z_c$ である。この例示的实施形態においては、ゼロ交差カウンタ $z_c$ は、正から負に信号の極性が変化する回数をその間隔の間カウントする。

【 0 0 4 5 】

第1の段の分類をより強固にするために、分類パラメータは、共に、メリットの関数 $f_m$ を形成するとみなされる。そのために、分類パラメータは、線形関数を使用してまずスケールされる。パラメータ $p_x$ を考えると、そのスケールされたバージョンは次式を使用して得られる。

$$p^s = k_p \cdot p_x + c_p \quad (7)$$

【 0 0 4 6 】

スケールされたピッチ安定性パラメータは、0と1との間でクリップされる。関数係数 $k_p$ および $c_p$ は、パラメータのそれぞれに対して実験的に求められている。この例示的实施形態に使用される値は、Table 1(表1)にまとめられている。

【 0 0 4 7 】

10

20

30

40

## 【表 1】

Table 1: デコーダにおける信号の第1段の分類パラメータ  
およびそれらのそれぞれのスケール関数の係数

パラメータ	意味	$k_p$	$c_p$
$r_x$	正規化相関関係	0.8547	0.2479
$e_t$	スペクトル傾斜	0.8333	0.2917
$pc$	ピッチ安定性カウンタ	-0.0357	1.6074
$E_s$	相対的フレームエネルギー	0.04	0.56
$zc$	ゼロ交差カウンタ	-0.04	2.52

10

## 【0048】

メリット関数は次式として定義されている。

## 【0049】

## 【数 4】

$$f_m = \frac{1}{6}(2 \cdot r_x^s + e_t^s + pc^s + E_s^s + zc^s) \quad (8)$$

20

## 【0050】

ここで、上付き文字sは、パラメータのスケールされたバージョンを示す。

## 【0051】

次いで、メリット関数 $f_m$ を使用し、以下のTable 2(表2)にまとめた規則に従って分類を行う(クラス選択テストポイント116)。

## 【0052】

## 【表 2】

Table 2: デコーダにおける信号分類規則

前のフレームクラス	規則	現在のフレームクラス
OTHER	$f_m \geq 0.39$	OTHER
	$f_m < 0.39$	UNVOICED
UNVOICED	$f_m > 0.45$	OTHER
	$f_m \leq 0.45$	UNVOICED
	VAD = 0	INACTIVE

30

## 【0053】

この第1の段の分類に加えて、エンコーダによる音声区間検出(VAD)の情報が、AMR-WBベースの例示的例の場合のようにビットストリームで伝送することができる。したがって、1ビットをビットストリームで送って、エンコーダが現在のフレームをアクティブコンテンツ(VAD=1)とみなすのか、またはINACTIVEコンテンツ(背景雑音VAD=0)とみなすのかを指定する。コンテンツがINACTIVEとみなされたとき、分類はUNVOICEDに上書きされる。第1の段の分類スキームは、GENERIC AUDIO検出も含む。GENERIC AUDIOカテゴリは、音楽、残響音声を含み、背景音楽も含むことができる。このカテゴリを識別するために2つのパラメータが使用される。パラメータの一方は、式(5)に公式化されているように全フレームエネルギー $E_f$ である。

40

## 【0054】

まず、モジュールが2つの隣接するフレームのエネルギー差

## 【0055】

50

【数5】

$$\Delta_E^t$$

【0056】

、具体的には現在のフレームのエネルギー

【0057】

【数6】

$$E_f^t$$

【0058】

と前のフレームのエネルギー

【0059】

【数7】

$$E_f^{(t-1)}$$

【0060】

との間の差を決定する。次いで、以下の関係を使用して過去の40フレームにわたる平均エネルギー差

【0061】

【数8】

$$\overline{E}_{df}$$

【0062】

を計算する。

【0063】

【数9】

$$\overline{E}_{df} = \frac{\sum_{t=-40}^{t=-1} \Delta_E^t}{40};$$

$$\text{ここで、} \Delta_E^t = E_f^t - E_f^{(t-1)}$$

(9)

30

【0064】

次いで、モジュールが、以下の関係を使用して最後の15フレームにわたるエネルギー変動の統計偏差  $\sigma_E$  を決定する。

【0065】

【数10】

$$\sigma_E = p \sqrt{\frac{\sum_{t=-15}^{t=-1} (\Delta_E^t - \overline{E}_{df})^2}{15}}$$

(10)

40

【0066】

例示の実施形態の実用化においては、倍率pは、実験的に求められ、約0.77に設定された。その結果得られた偏差  $\sigma_E$  によりデコードされた合成のエネルギー安定性についての指示が与えられる。典型的には、音楽は音声よりも高いエネルギー安定性を有する。

【0067】

第1の段の分類の結果は、UNVOICEDとして分類される2つのフレームの間のフレームの数  $N_{UV}$  をカウントするのにさらに使用される。実用化においては、-12dBよりも高いエネルギー  $E_f$  を有するフレームだけがカウントされる。一般に、フレームがUNVOICEDとして分類されたとき、カウンタ  $N_{UV}$  は0に初期設定される。しかし、フレームがUNVOICEDとして分類され、そのエネルギー  $E_f$  が-9dBよりも大きく、長時間平均エネルギー  $E_{1,t}$  が40dB未満である

50

とき、音楽の判定の方にわずかに偏向させるためにカウンタは16に初期設定される。それ以外の場合、フレームがUNVOICEDとして分類されたが、長時間平均エネルギー $E_{1t}$ が40dB超である場合、カウンタは音声の判定の方に収束させるために8だけ減少される。実用化においては、カウンタはアクティブな信号に対しては0と300との間に制限される。カウンタは、次のアクティブな信号が有効に音声であるとき、音声の判定への迅速な収束を得るためにINACTIVE信号に対しては0と125との間に制限もされる。これらの範囲は、限定するものではなく、他の範囲も特定の実現において企図することができる。この例示的例の場合、アクティブ信号とINACTIVE信号との判定は、ビットストリームに含まれる音声区間決定(VAD)から推測される。

【 0 0 6 8 】

長時間平均

【 0 0 6 9 】

【 数 1 1 】

$$\overline{N}_{uv}$$

【 0 0 7 0 】

は、アクティブな信号の場合、以下のように、このUNVOICEDフレームカウンタから導出され、

【 0 0 7 1 】

【 数 1 2 】

$$N_{UVkt} = 0.9 \cdot N_{UVkt} + 0.1 \cdot N_{UV}$$

【 0 0 7 2 】

【 数 1 3 】

$$\overline{N}_{uv}^t = 0.9 \cdot \overline{N}_{uv}^{(t-1)} + 0.1 \cdot N_{uv}, \quad (11)$$

【 0 0 7 3 】

INACTIVE信号の場合、以下のように、このUNVOICEDフレームカウンタから導出される。

【 0 0 7 4 】

【 数 1 4 】

$$\overline{N}_{uv}^t = 0.95 \cdot \overline{N}_{uv}^{(t-1)}, \quad (12)$$

【 0 0 7 5 】

ここで、tはフレームインデックスである。以下の擬似コードは、UNVOICEDカウンタの機能およびその長時間平均を示す。

【 0 0 7 6 】

10

20

30

## 【数 1 5】

if (UNVOICED &  $E_f > 9 \text{ dB}$ )

if ( $E_t \leq 40$ )

$$N_{uv} = 16$$

else

$$N_{uv} = N_{uv} - 8$$

elseif ( $E_f > 12$ )

$$N_{uv} = N_{uv} + 1$$

10

$$N_{uv} = \max(\min(300, N_{uv}), 0)$$

if (VAD=0)

$$\bar{N}_{uv} = 0.95 \cdot \bar{N}_{uv}$$

$$N_{uv} = \min(125, N_{uv})$$

else

$$\bar{N}_{uv} = 0.9 \cdot \bar{N}_{uv} + 0.1 \cdot N_{uv}$$

20

## 【0 0 7 7】

さらに、長時間平均

## 【0 0 7 8】

## 【数 1 6】

$$\bar{N}_{uv}$$

## 【0 0 7 9】

が非常に高く、偏差  $\sigma_E$  もある一定のフレーム(現在の例では

## 【0 0 8 0】

30

## 【数 1 7】

$$\bar{N}_{uv} > 140$$

## 【0 0 8 1】

および  $\sigma_E > 5$ )においてやはり高く、現在の信号が音楽である可能性がないことが意味されるとき、長時間平均はそのフレーム内で異なって更新される。100の値に収束し、判定を音声の方に偏向させるように長時間平均は更新される。これは以下に示すように行われる。

## 【0 0 8 2】

## 【数 1 8】

40

$$\bar{N}_{uv}^t = 0.2 \cdot \bar{N}_{uv}^{(t-1)} + 80$$

(13)

## 【0 0 8 3】

UNVOICEDに分類されたフレームの間のフレームの数の長時間平均によるこのパラメータは、フレームをGENERIC AUDIOとしてみなすべきかどうかを決定するのに使用される。UNVOICEDフレームがより多く時間的に近接していればいるほど、信号が音声特性を有する可能性がより多くある(GENERIC AUDIO信号である確率がより小さい)。例示的例においては、フレームがGENERIC AUDIO  $G_A$ とみなされるかどうかを判定する閾値は、以下のように定義される。

50

【 0 0 8 4 】

【 数 1 9 】

$$\overline{N}_{uv} > 100 \text{ および } \Delta_E^t < 12 \quad (14)$$

【 0 0 8 5 】

ならフレームは $G_A$ である。

【 0 0 8 6 】

大きなエネルギー変動をGENERIC AUDIOとして分類することを避けるために、式(9)に定義されたパラメータ

【 0 0 8 7 】

【 数 2 0 】

$$\Delta_E^t$$

【 0 0 8 8 】

が(14)で使用される。

【 0 0 8 9 】

励振により実施される後処理は、信号の分類に依存する。信号のある一定の種類の場合、後処理モジュールは、全く入力されない。次の表は、後処理が実施された場合をまとめたものである。

【 0 0 9 0 】

【 表 3 】

Table 3:励振変更のための信号カテゴリ

フレーム分類	後処理モジュールの入力 Y/N
VOICED	Y
GENERIC AUDIO	Y
UNVOICED	N
INACTIVE	N

【 0 0 9 1 】

後処理モジュールを入力したとき、以下に説明する、別のエネルギー安定性分析が連結励振スペクトルエネルギーに対して実施される。Vaillancourt'050の場合と同様に、この第2のエネルギー安定性分析により、スペクトルのどこで後処理が開始し、どの程度まで後処理が適用されるべきかとしての指示が得られる。

【 0 0 9 2 】

## 2) 励振ベクトルの作製

周波数分解能を増大させるために、フレーム長より長い周波数変換が使用される。そうするために、例示的实施形態においては、連結励振ベクトル $e_c(n)$ が、過去の励振バッファメモリ106に記憶された前のフレーム励振の最後の192サンプル、時間領域励振デコーダ104からの現在のフレーム $e(n)$ のデコードされた励振、および励振外挿器118から将来のフレーム $e_x(n)$ の192励振サンプルの外挿を連結することにより励振連結器120において作製される。これは以下に説明されるが、ただし、 $L_w$ が過去の励振の長さならびに外挿された励振の長さであり、 $L$ がフレーム長である。これは、それぞれ、192サンプルおよび256サンプルに対応し、例示的实施形態において全長 $L_c=640$ サンプルが得られる。

【 0 0 9 3 】

10

20

30

40

## 【数 2 1】

$$e_c(n) = \begin{cases} e(n) & n = -L_w, \dots, -1 \\ e(n) & n = 0, \dots, L-1 \\ e_x(n) & n = L, \dots, L+L_w-1 \end{cases} \quad (15)$$

## 【0094】

CELPデコーダにおいては、時間領域励振信号 $e(n)$ は、次式によって与えられる。

$$e(n) = bv(n) + gc(n)$$

10

## 【0095】

ここで、 $v(n)$ は適応コードブック寄与であり、 $b$ は適応コードブック利得であり、 $c(n)$ は固定コードブック寄与であり、 $g$ は固定コードブック利得である。将来の励振サンプル $e_x(n)$ の外挿は、現在のフレームの最後のサブフレームのデコードされた分数ピッチを使用して現在のフレーム励振信号 $e(n)$ を周期的に時間領域励振デコーダ104から延ばすことにより励振外挿器118において計算される。ピッチラグの分数分解能を仮定すると、35サンプルの長さのハミング窓掛けされた同期機能を使用して現在のフレーム励振のアップサンプリングが実施される。

## 【0096】

## 3) 窓掛け

20

窓掛けおよび周波数変換モジュール122においては、時間/周波数変換の前に、窓掛けが連結励振に対して実施される。選択された窓 $w(n)$ は、現在のフレームに対応する平坦な頂部を有し、各端部においてハミング機能により0まで減少する。以下の式は使用される窓を表す。

## 【0097】

## 【数 2 2】

$$w(n) = \begin{cases} 0.5 \left( 1 - \cos \left( \frac{2\pi(n+L_w)}{2L_w-1} \right) \right) & n = -L_w, \dots, -1 \\ 1.0 & n = 0, \dots, L-1 \\ 0.5 \left( 1 - \cos \left( \frac{2\pi((n-L)+L_w)}{2L_w-1} \right) \right) & n = L, \dots, L+L_w-1 \end{cases} \quad (16)$$

30

## 【0098】

連結励振に適用されたとき、全長 $L_c=640$ サンプル( $L_c=2L_w+L$ )を有する周波数変換への入力の実用化において取得される。窓掛けされた連結励振 $e_{wc}(n)$ が現在のフレームで中心となり、以下の式により表現される。

## 【0099】

## 【数 2 3】

40

$$e_{wc}(n) = \begin{cases} e(n)w(n) & n = -L_w, \dots, -1 \\ e(n)w(n) & n = 0, \dots, L-1 \\ e_x(n)w(n) & n = L, \dots, L+L_w-1 \end{cases} \quad (17)$$

## 【0100】

## 4) 周波数変換

周波数領域後処理フェーズの間、連結励振は変換領域で表現される。この例示的实施形態においては、時間/周波数変換は、10Hzの分解能を与えるタイプII DCTを使用して窓掛

50

けおよび周波数変換モジュール122において達成されるが、任意の他の変換を使用することができる。別の変換(または異なる変換の長さ)を使用した場合、周波数分解能(上記に定義された)、帯域の数、帯域ごとのピンの数(さらに以下に定義された)は、それに応じて改訂する必要があり得る。連結され窓掛けされた時間領域のCELP励振 $f_e$ の周波数表現は、以下に与えられる。

【 0 1 0 1 】

【 数 2 4 】

$$f_e(k) = \begin{cases} \sqrt{\frac{1}{L_c}} \cdot \sum_{n=0}^{L_c-1} e_{wc}(n), & k=0 \\ \sqrt{\frac{2}{L_c}} \cdot \sum_{n=0}^{L_c-1} e_{wc}(n) \cdot \cos\left(\frac{\pi}{L_c}\left(n+\frac{1}{2}\right)k\right), & 1 \leq k \leq L_c-1 \end{cases} \quad (18)$$

【 0 1 0 2 】

ここで、 $e_{wc}(n)$ は、連結され、窓掛けされた時間領域励振であり、 $L_c$ は周波数変換の長さである。この例示的实施形態においては、フレーム長 $L$ は256サンプルであるが、周波数変換の長さ $L_c$ は、対応する内部サンプリング周波数が12.8kHzである場合640サンプルである。

【 0 1 0 3 】

5) 帯域ごとおよびピンごとのエネルギー分析

DCTの後、結果として得られたスペクトルは、臨界周波数帯域に分割される(実現化では、周波数範囲0~4000Hzにおいて17の臨界帯域および周波数範囲0~6400Hzにおいて20の臨界周波数帯域を使用する)。使用される臨界周波数帯域は、参照によりその内容が本明細書に組み込まれる、J. D. Johnston「Transform coding of audio signal using perceptual noise criteria」、IEEE J. Select. Areas Commun.、vol. 6、314~323ページ、1988年2月に指定されるものにできるだけ近くし、それらの上限は以下のように定義される。すなわち、 $C_B = \{100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400\}$  Hzである。

【 0 1 0 4 】

640ポイントのDCTは、結果として10Hzの周波数分解能となる(6400Hz/640ポイント)。臨界周波数帯域ごとの周波数ピンの数は、 $M_{CB} = \{10, 10, 10, 10, 11, 12, 14, 15, 16, 19, 21, 24, 28, 32, 38, 45, 55, 70, 90, 110\}$ である。

【 0 1 0 5 】

臨界周波数帯域ごとの平均スペクトルエネルギー $E_B(i)$ は、以下のように計算される。

【 0 1 0 6 】

【 数 2 5 】

$$E_B(i) = \frac{1}{L_c M_{CB}(i)} \sum_{h=0}^{M_B(i)-1} (f_e(h+j_i))^2, \quad i=0, \dots, 20 \quad (19)$$

【 0 1 0 7 】

ここで、 $f_e(h)$ は、臨界帯域の $h$ 番目の周波数ピンを表し、 $j_i$ は、 $j_i = \{0, 10, 20, 30, 40, 51, 63, 77, 92, 108, 127, 148, 172, 200, 232, 270, 315, 370, 440, 530\}$

によって与えられる $i$ 番目の臨界帯域における第1のピンのインデックスである。

【 0 1 0 8 】

スペクトル分析は、以下の関係を使用して周波数ピンごとのスペクトルのエネルギー $E_{iN}(k)$ も計算する。

【 0 1 0 9 】

10

20

30

40

50

【数 2 6】

$$E_{BIN}(k) = \frac{1}{L_c} f_e(k)^2, \quad k = 0, \dots, 639 \quad (20)$$

【0 1 1 0】

最後に、スペクトル分析は、以下の関係を使用して第1の17の臨界周波数帯域のスペクトルエネルギーの合計として連結励振の全スペクトルエネルギー $E_C$ を計算する。

【0 1 1 1】

【数 2 7】

10

$$E_C = 10 \log_{10} \left( \sum_{i=0}^{16} E_B(i) \right) - 3.0103, \text{ dB} \quad (21)$$

【0 1 1 2】

6) 励振信号の第2の段の分類

Vaillancourt'050に説明されるように、デコードされた汎用音響信号を強化するための方法は、どのフレームがトーン間雑音低減によく適しているかを識別することにより高調波間雑音低減の効率をさらに最大化するように設計された励振信号の追加の分析を含む。

【0 1 1 3】

第2の段の信号分類器124は、デコードされた連結励振を音響信号カテゴリにさらに分離するだけでなく、低減が開始できる最大レベルの減衰および最小周波数に関する命令を高調波間雑音低減装置128に与えもする。

20

【0 1 1 4】

提示される例示的例においては、第2の段の信号分類器124は、できるだけ簡略に保持され、Vaillancourt'050に説明される信号種類分類器に非常に類似している。第1の動作は、式(9)および(10)で行われるのと同様に、ただし、式(21)に公式化されているように連結励振 $E_C$ の全スペクトルエネルギーを入力として使用して、エネルギー安定性分析を実施することにある。

【0 1 1 5】

【数 2 8】

30

$$\bar{E}_a = \frac{\left( \sum_{t=-40}^{t=-1} \Delta_{E_C}^t \right)}{40}, \quad \text{ここで、} \Delta_{E_C}^t = E_C^t - E_C^{(t-1)} \quad (22)$$

【0 1 1 6】

ここで、

【0 1 1 7】

【数 2 9】

40

$$\bar{E}_a$$

【0 1 1 8】

は、2つの隣接するフレームの連結励振ベクトルのエネルギーの平均差を表し、

【0 1 1 9】

【数 3 0】

$$E_C^t$$

【0 1 2 0】

は、現在のフレーム $t$ の連結励振のエネルギーを表し、

50

【 0 1 2 1 】

【 数 3 1 】

$$E_C^{(t-1)}$$

【 0 1 2 2 】

は、前のフレームt-1の連結励振のエネルギーを表す。平均は最後の40フレームにわたって計算される。

【 0 1 2 3 】

次いで、最後の15フレームにわたってエネルギー変動の統計偏差  $\sigma_c$ が以下の関係を使用して計算される。

10

【 0 1 2 4 】

【 数 3 2 】

$$\sigma_c = p \cdot \sqrt{\sum_{t=-15}^{t=-1} \frac{(\Delta_{E_C}^t - \bar{E}_d)^2}{15}} \tag{23}$$

【 0 1 2 5 】

ここで、実用化においては、倍率pが実験的に求められ、約0.77に設定される。その結果得られた偏差  $\sigma_c$ は、高調波間の雑音をどの程度まで低減できるのかを決定するために4つの浮動閾値と比較される。この第2の段の信号分類器124の出力は、音響信号カテゴリ0から4までに命名された、5つの音響信号カテゴリ $e_{CAT}$ に分割される。各音響信号カテゴリは、それ自体のトーン間雑音低減調整を有する。

20

【 0 1 2 6 】

5つの音響信号カテゴリ0~4は、以下の表に示すように決定することができる。

【 0 1 2 7 】

【 表 4 】

Table 4:励振分類器の出力特性

カテゴリ	強化帯域 (広帯域)	許容低減
$e_{CAT}$	Hz	dB
0	NA	0
1	[920, 6400]	6
2	[920, 6400]	9
3	[770, 6400]	12
4	[630, 6400]	12

30

40

【 0 1 2 8 】

音響信号カテゴリ0は、トーン間雑音低減技法によって変更されない、非トーンの、非安定音響信号カテゴリである。デコードされた音響信号のこのカテゴリは、スペクトルエネルギー変動の最大の統計偏差を有し、概して、音声信号を含む。

【 0 1 2 9 】

50

音響信号カテゴリ1(カテゴリ0に続くスペクトルエネルギー変動の最大の統計偏差)は、スペクトルエネルギー変動の統計偏差  $\sigma_c$  が閾値1より小さく、最後に検出された音響信号カテゴリが 0であるとき、検出される。次いで、周波数帯域920～

【0130】

【数33】

$$\frac{F_S}{2}$$

【0131】

Hz(この例では6400Hz。ここでFsはサンプリング周波数)内のデコードされたトーン励振の量子化雑音の最大低減は、6dBの最大雑音低減 $R_{max}$ に制限される。

10

【0132】

音響信号カテゴリ2は、スペクトルエネルギー変動の統計偏差  $\sigma_c$  が閾値2より小さく、最後に検出された音響信号カテゴリが 1であるとき、検出される。次いで、周波数帯域920～

【0133】

【数34】

$$\frac{F_S}{2}$$

【0134】

Hz内のデコードされたトーン励振の量子化雑音の最大低減が最大9dBに制限される。

20

【0135】

音響信号カテゴリ3は、スペクトルエネルギー変動の統計偏差  $\sigma_c$  が閾値3より小さく、最後に検出された音響信号カテゴリが 2であるとき、検出される。次いで、周波数帯域770～

【0136】

【数35】

$$\frac{F_S}{2}$$

【0137】

Hz内のデコードされたトーン励振の量子化雑音の最大低減が最大12dBに制限される。

30

【0138】

音響信号カテゴリ4は、スペクトルエネルギー変動の統計偏差  $\sigma_c$  が閾値4より小さいとき、かつ最後に検出された信号種類カテゴリが 3であるとき、検出される。次いで、周波数帯域630～

【0139】

【数36】

$$\frac{F_S}{2}$$

【0140】

Hz内のデコードされたトーン励振の量子化雑音の最大低減が最大12dBに制限される。

40

【0141】

浮動閾値1～4は、間違っただ信号種類の分類を防止するのに役立つ。典型的には、音楽を表すデコードされたトーン音響信号は、そのスペクトルエネルギー変動の統計偏差が音声よりもずっと低くなる。しかし、音楽信号でさえ、より高い統計偏差セグメントを含むことができ、同様に音声信号はより小さい統計偏差を有するセグメントを含むことができる。それにもかかわらず、音声および音楽コンテンツは、フレームベースで一方から別のも

50

のに規則的に変化する可能性がない。浮動閾値は、判定ヒステリシスを追加し、高調波間雑音低減装置128の準最適な性能をもたらす得る任意の誤分類を実質的に防止するために前の状態の強化として働く。

【0142】

音響信号分類0の連続フレームのカウンタ、および音響信号カテゴリ3または4の連続フレームのカウンタは、それぞれ、閾値を低減または増大させるのに使用される。

【0143】

例えば、カウンタが音響信号カテゴリ3または4の一連の30フレーム超をカウントする場合、すべての浮動閾値(1から4までの)は、より多くのフレームが音響信号カテゴリ4とみなされることを可能にするために、所定の値だけ増加される。

10

【0144】

音響信号カテゴリ0については逆もまた真である。例えば、音響信号カテゴリ0の一連の30フレーム超がカウントされた場合、すべての浮動閾値(1から4までの)は、より多くのフレームが音響信号カテゴリ0とみなされることを可能にするために減少される。すべての浮動閾値1~4を絶対最大値および最小値に制限して、信号分類器が確実に固定カテゴリにロックされないようにする。

【0145】

フレーム消去の場合、すべての閾値1~4がそれらの最小値に再設定され、第2の段の分類器の出力が3連続フレーム(失われたフレームを含めて)に対して非トーン(音響信号カテゴリ0)とみなされる。

20

【0146】

音声区間検出器(VAD)からの情報が利用可能であり、その情報が音声活動を何も示していない(無音の存在)場合、第2の段の分類器の判定は、音響信号カテゴリ0( $e_{CAT}=0$ )に強制される。

【0147】

#### 7) 励振領域における高調波間雑音低減

トーン間または高調波間雑音低減は、強化の第1の動作として連結励振の周波数表現により実施される。トーン間量子化雑音の低減は、スケーリング利得 $g_s$ を最小利得 $g_{min}$ と最大利得 $g_{max}$ との間に制限して、各臨界帯域においてスペクトルをスケーリングすることにより雑音低減装置128において実施される。スケーリング利得は、その臨界帯域における推定信号対雑音比(SNR)から導出される。処理は、臨界帯域ベースではなく、周波数ピンベースで実施される。したがって、スケーリング利得は、すべての周波数ピンに適用され、そのピンを含む臨界帯域の雑音エネルギーの推定によって割られたピンエネルギーを使用して計算されたSNRから導出される。この特徴により、高調波またはトーンの近くの周波数におけるエネルギーを維持することが可能になり、したがって、実質的に歪みを防止し、高調波間の雑音を強力に低減することが可能になる。

30

【0148】

トーン間雑音低減がすべての640ピンにわたってピンごとのやり方で実施される。トーン間雑音低減をスペクトルに適用した後、スペクトル強化の別の動作が実施される。次いで、後述するように、強化された連結励振

40

【0149】

【数37】

$$e_{id}^{\prime}$$

【0150】

信号を再構成するのに逆DCTを使用する。

【0151】

最小スケーリング利得 $g_{min}$ は、dB単位の最大許容トーン間雑音低減 $R_{max}$ から導出される。上述したように、第2の段の分類により、最大許容低減が6dbから12dbまでの間で変動す

50

ることが可能になる。したがって、最小スケーリング利得は次式により与えられる。

【 0 1 5 2 】

【数 3 8】

$$g_{\min} = 10^{-R_{\max}/20} \quad (24)$$

【 0 1 5 3 】

スケーリング利得は、ビンごとのSNRに関連して計算される。次いで、ビンごとの雑音低減は、上述したように実施される。現在の例においては、ビンごとの処理が6400Hzの最大周波数までスペクトル全体に適用される。この例示的实施形態においては、雑音低減は6番目の臨界帯域から開始する(すなわち、630Hz未満では低減は何も実施されない)。技法の任意の悪影響を低減するために、第2の段の分類器は、開始する臨界帯域を8番目の帯域(920Hz)まで押し上げることができる。すなわち、雑音低減が実施される第1の臨界帯域が630Hzから920Hzまでの間にあり、フレームベースで変動することができる。より控えめな実現においては、雑音低減が開始する最小帯域は、より高く設定することができる。

10

【 0 1 5 4 】

ある一定の周波数ビン $k$ のスケーリングは、次式によって与えられる、SNRの関数として計算される。

【 0 1 5 5 】

【数 3 9】

$$g_s(k) = \sqrt{k_s \text{SNR}(k) + c_s}, \text{ によって境界される } g_{\min} \leq g_s \leq g_{\max} \quad (25)$$

20

【 0 1 5 6 】

通常、 $g_{\max}$ は1に等しく(すなわち、増幅は何も許容されず)、したがって、 $k_s$ および $c_s$ の値は、SNR=1dBの場合 $g_s=g_{\min}$ 、SNR=45dBの場合 $g_s=1$ などのように決定される。すなわち、1dB以下のSNRの場合、スケーリングは $g_{\min}$ に制限され、45dB以上のSNRの場合、雑音低減は何も実施されない( $g_s=1$ )。したがって、これらの2つの端点を考えると、式(25)における $k_s$ および $c_s$ の値は次式によって与えられる。

【 0 1 5 7 】

【数 4 0】

$$k_s = (1 - g_{\min}^2) / 44 \text{ および } c_s = (45g_{\min}^2 - 1) / 44 \quad (26)$$

30

【 0 1 5 8 】

$g_{\max}$ が1より高い値に設定された場合、処理が、最高のエネルギーを有するトーンをわずかに増幅することが可能となる。これは、実用化において使用される、CELPコーデックが周波数領域におけるエネルギーに完全には一致しないことを補償するのに使用することができる。これは一般に有声音声とは異なる信号の場合である。

【 0 1 5 9 】

ある一定の臨界帯域 $i$ におけるビンごとのSNRは、次式として計算される。

【 0 1 6 0 】

【数 4 1】

$$\text{NRF}_{\text{BIN}}(h) = \frac{0.3E_{\text{BIN}}^{(1)}(h) + 0.7E_{\text{BIN}}^{(2)}(h)}{N_B(i)}, \quad h = j_i, \dots, j_i + M_B(i) - 1 \quad (27)$$

40

【 0 1 6 1 】

ここで、

【 0 1 6 2 】

50

【数 4 2】

$$E_{BIN}^{(1)}(h)$$

【0 1 6 3】

および

【0 1 6 4】

【数 4 3】

$$E_{BIN}^{(2)}(h)$$

10

【0 1 6 5】

は、それぞれ、式(20)において計算される、過去のおよび現在のフレームのスペクトル分析に対する周波数ビンごとのエネルギーを表し、 $N_B(i)$ は、臨界帯域*i*の雑音エネルギー推定を表し、 $j_i$ は*i*番目の臨界帯域における第1のビンのインデックスであり、 $M_B(i)$ は上記に定義された、臨界帯域*i*におけるビンの数である。

【0 1 6 6】

平滑化係数は、適応でき、利得自体に逆相関される。この例示的实施形態においては、平滑化係数は  $g_s=1-g_s$  によって与えられる。すなわち、平滑化は利得 $g_s$ がより小さければより強力である。この取組みは、有声オンセットの場合のように、実質的に、低いSNR フレームによって先行される高いSNRセグメントにおける歪みを防止する。例示的实施形態においては、平滑化手順は、オンセットに対して迅速に適応し、より低いスケールリング利得を使用することができる。

20

【0 1 6 7】

インデックス*i*を有する臨界帯域におけるビンごとの処理の場合、式(25)におけるようにスケールリング利得を決定した後、および式(27)において定義されたSNRを使用して、実際のスケールリングを、以下のように周波数分析ごとに更新される平滑化されたスケールリング利得 $g_{BIN,LP}$ を使用して実施する。

$$g_{BIN,LP}(k) = g_s g_{BIN,LP}(k) + (1 - g_s) g_s \quad (28)$$

【0 1 6 8】

利得の時間平滑化は、実質的に可聴エネルギー発振を防止し、 $g_s$ を使用して平滑化を制御することにより、有声オンセットまたはアタックの場合のように、低いSNRフレームによって先行される高いSNRセグメントにおける歪みを実質的に防止する。

30

【0 1 6 9】

臨界帯域*i*におけるスケールリングは次式として実施される。

【0 1 7 0】

【数 4 4】

$$f_e'(h+j_i) = g_{BIN,LP}(h+j_i) f_e(h+j_i), \quad h=0, \dots, M_B(i)-1 \quad (29)$$

40

【0 1 7 1】

ここで、 $j_i$ は、臨界帯域*i*における第1のビンのインデックスであり、 $M_B(i)$ はその臨界帯域におけるビンの数である。

【0 1 7 2】

平滑化されたスケールリング利得 $g_{BIN,LP}(k)$ は、1に初期設定される。非トーン音響フレームが処理 $e_{CAT}=0$ されるたびに、平滑化された利得の値を1.0に再設定して、次のフレームにおいて任意の可能な低減があれば低減する。

【0 1 7 3】

あらゆるスペクトル分析において、平滑化されたスケールリング利得 $g_{BIN,LP}(k)$ は、スペクトル全体におけるすべての周波数ビンに対して更新されることに留意されたい。低エネ

50

ルギー信号の場合、トーン間雑音低減は-1.25dBに制限される。これは、すべての臨界帯域において最大雑音エネルギー $\max(N_B(i))$ ,  $i=0, \dots, 20$ が10以下であるとき起きる。

【0174】

#### 8) トーン間量子化雑音推定

この例示の実施形態においては、臨界周波数帯域ごとのトーン間量子化雑音エネルギーは、同じ帯域の最大ピンエネルギーを除外する、その臨界周波数帯域の平均エネルギーであるとして帯域ごとの雑音レベル推定器126において推定される。以下の公式は、具体的な帯域*i*の量子化雑音エネルギーの推定をまとめたものである。

【0175】

【数45】

$$N_B(i) = \frac{1}{q(i)} \left( \frac{\left( E_B(i) M_B(i) - \max_h (E_{BIN}(h+j_i)) \right)}{(M_B(i)-1)} \right), \quad h=0, \dots, M_B(i)-1 \quad (30)$$

10

【0176】

ここで、 $j_i$ は臨界帯域*i*における第1のピンのインデックスであり、 $M_B(i)$ は、その臨界帯域におけるピンの数であり、 $E_B(i)$ は帯域*i*の平均エネルギーであり、 $E_{BIN}(h+j_i)$ は、特定のピンのエネルギーであり、 $N_B(i)$ は、結果として得られた特定の帯域*i*の推定雑音エネルギーである。雑音推定式(30)において、 $q(i)$ は、実験的に求めた帯域ごとの雑音スケールリング倍率を表し、後処理が使用される実現により変更することができる。実用化においては、雑音倍率は、以下に示すように、低周波数においてより多くの雑音を除去することができ、高周波数においてより少ない雑音を除去することができるよう設定される。

20

$q=\{10, 10, 10, 10, 10, 10, 10, 11, 11, 11, 11, 11, 11, 11, 11, 11, 15, 15, 15, 15, 15\}$

【0177】

#### 9) 励振のスペクトルダイナミクスの増大

周波数後処理の第2の動作は、符号化雑音内に失われた周波数情報を取り出す能力を提供する。CELPコーデックは、特に低ビットレートで使用されたとき、3.5~4kHz超で周波数コンテンツを正しく符号化するのにそれほど効率的ではない。ここでの主な考え方は、音楽スペクトルがしばしば実質的にフレームごとに変化しないことを利用することである。したがって、長時間平均化を行うことができ、符号化雑音の一部を削除することができる。以下の動作は、周波数依存利得関数を定義するのに実施される。この関数は、次いで、時間領域にまた変換する前に励振をさらに強化するのに使用される。

30

【0178】

##### a. スペクトルエネルギーのピンごとの正規化

第1の動作は、連結励振のスペクトルの正規化エネルギーに基づいてマスクビルダ130において重み付けマスクを作製することにある。正規化は、トーン(または高調波)が1.0超の値を有し、谷が1.0未満の値を有するようにスペクトルエネルギー正規化器131において行われる。そうするために、ピンエネルギースペクトル $E_{BIN}(k)$ を0.925と1.925との間で正規化して、以下の式を使用して正規化エネルギースペクトル $E_n(k)$ を得る。

40

【0179】

【数46】

$$E_n(k) = \frac{E_{BIN}(k)}{\max(E_{BIN})} + 0.925, \quad k=0, \dots, 639 \quad (31)$$

【0180】

ここで、 $E_{BIN}(k)$ は、式(20)において計算されたピンエネルギーを表す。正規化がエネルギー領域において実施されるので、多くのピンは非常に低い値を有する。実用化においては、正規化エネルギーピンの小さな部分だけが1.0未満の値を有するように、オフセッ

50

ト0.925が選択されている。正規化が行われると、結果として得られた正規化エネルギースペクトルをべき関数により処理して、スケーリングされたエネルギースペクトルを得る。この例示的例においては、以下の公式に示すように、スケーリングされたエネルギースペクトルの最小値を0.5程度に制限するのに8の累乗が使用される。

$$E_p(k) = E_n(k)^8 \quad k=0, \dots, 639 \quad (32)$$

【0181】

ここで、 $E_n(k)$ は正規化エネルギースペクトルであり、 $E_p(k)$ はスケーリングされたエネルギースペクトルである。量子化雑音をさらに低減するために、より積極的なべき関数を使用することができ、例えば、10または16の累乗を選択することができ、場合によりオフセットを1により近づけることができる。しかし、あまりに多くの雑音を除去しようとすると、結果として重要な情報を失うことになることもある。

10

【0182】

べき関数をその出力を制限せずに使用すると、急速に1より高いエネルギースペクトル値に飽和をもたらす。スケーリングされたエネルギースペクトルの最大制限は、したがって、実用化において5に固定され、最大正規化エネルギー値と最小正規化エネルギー値との間でおよそ10の比を生じる。支配的なピンがフレームによってわずかに異なる位置を有することがあり、したがって、重み付けマスクが1つのフレームから次のフレームまで相対的に安定であることが好ましいならば、これは有用である。以下の式は、どのように関数を適用するのかを示す。

$$E_{p1}(k) = \min(5, E_p(k)) \quad k=0, \dots, 639 \quad (33)$$

20

【0183】

ここで、 $E_{p1}(k)$ は制限されスケーリングされたエネルギースペクトルを表し、 $E_p(k)$ は式(32)において定義されたスケーリングされたエネルギースペクトルを表す。

【0184】

b. 周波数軸および時間軸に沿ったスケーリングされたエネルギースペクトルの平滑化

最後の2つの動作により、最も活動的なパルスの位置が具体化し始める。正規化エネルギースペクトルのピンに8の累乗を適用することは、スペクトルダイナミクスを増大させるための効率的なマスクを作製する第1の動作である。次の2つの動作は、このスペクトルマスクをさらに強化する。まず、スケーリングされたエネルギースペクトルは、平均化フィルタを使用して低周波数から高周波数まで周波数軸に沿ってエネルギー平均化器132において平滑化される。次いで、結果として得られたスペクトルは、時間領域軸に沿ってエネルギー平滑化器134において処理して、フレームごとにピンの値を平滑化する。

30

【0185】

スケーリングされたエネルギースペクトルの周波数軸に沿った平滑化は、以下の関数を用いて説明することができる。

【0186】

【数47】

$$\bar{E}_{pl}(k) = \begin{cases} \frac{E_{pl}(k) + E_{pl}(k+1)}{2}, & k=0 \\ \frac{E_{pl}(k-1) + E_{pl}(k) + E_{pl}(k+1)}{3}, & k=1, \dots, 638 \\ \frac{E_{pl}(k-1) + E_{pl}(k)}{2}, & k=639 \end{cases} \quad (34)$$

40

【0187】

最後に、時間軸に沿った平滑化は、結果として、スペクトル

【0188】

【数 4 8】

 $f_e'$ 

【0 1 8 9】

に適用される時間平均増幅/減衰重み付けマスク $G_m$ となる。重み付けマスクは、利得マスクとも呼ばれ、以下の式を用いて説明される。

【0 1 9 0】

【数 4 9】

$$G_m^t(k) = \begin{cases} 0.95 \cdot G_m^{(t-1)}(k) + 0.05 \bar{E}_{pl}(k), & k = 0, \dots, 319 \\ 0.85 \cdot G_m^{(t-1)}(k) + 0.15 \bar{E}_{pl}(k), & k = 320, \dots, 639 \end{cases} \quad (35)$$

10

【0 1 9 1】

ここで、

【0 1 9 2】

【数 5 0】

 $\bar{E}_{pl}$ 

【0 1 9 3】

は、周波数軸に沿って平滑化されたスケーリングされたエネルギースペクトルであり、 $t$ はフレームインデックスであり、 $G_m$ は時間平均重み付けマスクである。

20

【0 1 9 4】

実質的に利得発振を防止するため、より遅い適応率がより低い周波数に選ばれている。トーンの位置がスペクトルのより高い部分において急速に変化する可能性がより多くあるので、より速い適応率が、より高い周波数に許容される。平均化が周波数軸で実施され、長時間平滑化が時間軸に沿って実施されると、(35)において得られた最終ベクトルが、式(29)の連結励振

【0 1 9 5】

【数 5 1】

 $f_e'$ 

30

【0 1 9 6】

の強化スペクトルに直接適用される重み付けマスクとして使用される。

【0 1 9 7】

10) 強化された連結励振スペクトルへの重み付けマスクの適用

上記に定義された重み付けマスクは、第2の段の励振分類器の出力(table 4(表4)に示す $e_{CAT}$ の値)によりスペクトルダイナミクス変更器136によって異なって適用される。重み付けマスクは、励振がカテゴリ0( $e_{CAT}=0$ 。すなわち、音声コンテンツの確率が高い)として分類された場合適用されない。コーデックのビットレートが高いとき、量子化雑音のレベルは、概して、より低く、周波数により変化する。すなわち、トーン増幅をスペクトル内のパルス位置およびエンコードされたビットレートにより制限することができる。CELP以外の別のエンコーディング方法を使用して、例えば、励振信号が時間および周波数領域で符号化された構成要素の組合せを含む場合、重み付けマスクの使用は、特定の事例ごとに調整され得る。例えば、パルス増幅は制限することができるが、方法は量子化雑音低減として依然として使用することができる。

40

【0 1 9 8】

最初の1kHz(実用化においては最初の100ピン)には、励振がカテゴリ0( $e_{CAT}=0$ )として分類されない場合、マスクは適用される。減衰は可能であるが、しかし、この周波数範囲

50

において増幅は何も実施されない(マスクの最大値が1.0に制限される)。

【0199】

25連続フレーム超がカテゴリ4( $e_{CAT}=4$ 。すなわち、音楽コンテンツの確率が高い)として分類されたが、多くて40フレームである場合、重み付けマスクは、残りのすべてのピン(ピン100から639まで)には増幅なしで適用される(最大利得 $G_{max0}$ が1.0に制限され、最小利得には何も制限がない)。

【0200】

1kHzから2kHzまでの間の周波数(実用化においてはピン100から199まで)に対して、40フレーム超がカテゴリ4として分類されたとき、最大利得 $G_{max1}$ は毎秒12650ビット(bps)未満のビットレートに対して1.5に設定される。それ以外の場合、最大利得 $G_{max1}$ は、1.0に設定される。この周波数帯域においては、ビットレートが15850bpsより高い場合のみ、最小利得 $G_{min1}$ は0.75に固定され、それ以外の場合、最小利得には何も制限がない。

10

【0201】

2kHzから4kHzまで(実用化においてはピン200から399まで)の帯域の場合、12650bps未満のビットレートには、最大利得 $G_{max2}$ は2.0に制限され、12650bps以上および15850bps未満のビットレートには、1.25に制限される。それ以外の場合、最大利得 $G_{max2}$ は1.0に制限される。この周波数帯域においては、ビットレートが15850bps超である場合のみ、最小利得 $G_{min2}$ は0.5に固定され、それ以外の場合、最小利得には何も制限がない。

【0202】

4kHzから6.4kHzまで(実用化においてはピン400から639まで)の帯域の場合、15850bps未満のビットレートには、最大利得 $G_{max3}$ は2.0に制限され、それ以外の場合には1.25に制限される。この周波数帯域においては、ビットレートが15850bps超である場合のみ、最小利得 $G_{min3}$ は、0.5に固定され、それ以外の場合、最小利得には何も制限がない。最大および最小利得の他の調整が、コーデックの特性により適当であり得ることに留意されたい。

20

【0203】

次の擬似コードは、重み付けマスク $G_m$ を強化スペクトル

【0204】

【数52】

$$f_e'$$

30

【0205】

に適用したとき、連結励振の最終スペクトル $f_e'$ がどのように影響されるかを示す。スペクトル強化の第1の動作(第7章で説明したように)は、ピンごとの利得変更のこの第2の強化動作を行うのに絶対に必要とはされないことに留意されたい。

【0206】

## 【数53】

$$\begin{aligned}
 & \text{if } (e_{CAT} \neq 0) \\
 & \text{if } (e_{CAT} == 4 \forall t = -1, \dots, -40) \\
 & f_e''(k) = \begin{cases} f_e'(k) \min(G_m(k), G_{max0}), & k = 0, \dots, 99 \\ f_e'(k) \max(\min(G_m(k), G_{max1}), G_{min1}), & k = 100, \dots, 199 \\ f_e'(k) \max(\min(G_m(k), G_{max2}), G_{min2}), & k = 200, \dots, 399 \\ f_e'(k) \max(\min(G_m(k), G_{max3}), G_{min3}), & k = 400, \dots, 639 \end{cases} \\
 & \text{else if } (e_{CAT} == 4 \forall t = -1, \dots, -25) \\
 & f_e''(k) = f_e'(k) \min(G_m(k), 1.0), \quad k = 0, \dots, 639 \\
 & \text{else} \\
 & f_e''(k) = f_e'(k), \quad k = 0, \dots, 639
 \end{aligned} \tag{36}$$

## 【0207】

ここで、 $f_e'$ は、前に式(28)のSNR関係関数 $g_{BIN,LP}(k)$ を用いて強化された連結励振のスペクトルを表し、 $G_m$ は、式(35)において計算された重み付けマスクであり、 $G_{max}$ および $G_{min}$ は、上記に定義したように周波数範囲ごとの最大および最小利得であり、 $t$ は、現在のフレームに対応する $t=0$ のフレームインデックスであり、最後に、 $f_e''$ は連結励振の最終強化スペクトルである。

## 【0208】

## 11) 逆周波数変換

周波数領域強化が完了した後、強化された時間領域励振を取り戻すために、逆周波数/時間変換が周波数/時間領域変換器138において実施される。この例示的な実施形態においては、周波数/時間変換は、時間/周波数変換に使用されるのと同じ種類のII DCTを用いて達成される。変更された時間領域励振

## 【0209】

## 【数54】

$$e_{td}'$$

## 【0210】

は、次式として得られる。

## 【0211】

## 【数55】

$$e_{td}'(n) = \begin{cases} \sqrt{\frac{1}{L_c}} \cdot \sum_{k=0}^{L_c-1} f_e''(k), & n = 0 \\ \sqrt{\frac{2}{L_c}} \cdot \sum_{k=0}^{L_c-1} f_e''(k) \cdot \cos\left(\frac{\pi}{L_c} \left(k + \frac{1}{2}\right) n\right), & 1 \leq n \leq L_c - 1 \end{cases} \tag{37}$$

## 【0212】

ここで、 $f_e''$ は、変更された励振の周波数表現であり、

## 【0213】

【数56】

$$e_{id}^{'}$$

【0214】

は、強化された連結励振であり、 $L_c$ は連結励振ベクトルの長さである。

【0215】

12) 現在のCELP合成をフィルタリングし、上書きする合成

合成に遅延を追加することは望ましくないので、実用化の構築においてオーバーラップおよび追加のアルゴリズムを避けることが決定されている。実用化は、以下の式に示されるようにオーバーラップなしで、合成を生成するのに使用される最終励振 $e_f$ の正確な長さを強化された連結励振から直接とる。

【0216】

【数57】

$$e_f(n) = e_{id}^{'(n+L_w)}, \quad n=0, \dots, 255 \quad (38)$$

【0217】

ここで、 $L_w$ は、式(15)で説明したように、周波数変換の前に過去の励振に適用される窓掛けの長さを表す。励振変更が行われ、周波数/時間領域変換器138からの強化され変更された時間領域励振の適正な長さが、フレーム励振抽出器140を使用して連結ベクトルから抽出されると、変更された時間領域励振が、現在のフレームの強化された合成信号を得るために合成フィルタ110を通して処理される。この強化された合成は、知覚品質を上げるために合成フィルタ108からの元々デコードされた合成に上書きするのに使用される。上書きする判定は、クラス選択テストポイント116からの、および第2の段の信号分類器124からの情報に回答して、上記に説明したようにスイッチ146を制御する判定テストポイント144を含む上書き器142によって下される。

【0218】

図3は、図2のデコーダを形成するハードウェア構成要素の構成例の簡略化された構成図である。デコーダ200は、モバイル端末の一部として、ポータブルメディアプレーヤの一部として、または任意の同様のデバイスにおいて実現することができる。デコーダ200は、入力202と、出力204と、プロセッサ206と、メモリ208とを備える。

【0219】

入力202は、AMR-WBビットストリーム102を受け取るように構成される。入力202は、図2の受信機102を一般化したものである。入力202の非限定実現例は、モバイル端末の無線インターフェース、例えば、ポータブルメディアプレーヤのユニバーサルシリアルバス(USB)ポートなどの物理的インターフェースを備える。出力204は、図2のD/A変換器154、増幅器156およびスピーカ158を一般化したものであり、オーディオプレーヤ、スピーカ、記録デバイスなどを備えることができる。あるいは、出力204は、オーディオプレーヤ、スピーカ、記録デバイスなどに接続可能なインターフェースを備えることができる。入力202および出力204は、共通モジュール、例えば、シリアル入出力デバイスにおいて実現することができる。

【0220】

プロセッサ206は、入力202に、出力204に、およびメモリ208に動作可能に接続される。プロセッサ206は、時間領域励振デコーダ104の、LP合成フィルタ108および110の、第1の段の信号分類器112およびその構成要素の、励振外挿器118の、励振連結器120の、窓掛けおよび周波数変換モジュール122の、第2の段の信号分類器124の、帯域ごとの雑音レベル推定器126の、雑音低減装置128の、マスクビルダ130およびその構成要素の、スペクトルダイナミクス変更器136の、スペクトル/時間領域変換器138の、フレーム励振抽出器140の、上書き器142およびその構成要素の、ならびにディエンファージングフィルタおよび

10

20

30

40

50

リサンブラ148の機能を支持してコード命令を実行するための1つまたは複数のプロセッサとして実現される。

【0221】

メモリ208は、様々な後処理動作の結果を記憶する。より詳しくは、メモリ208は、過去の励振バッファメモリ106を備える。いくつかの変形において、プロセッサ206の様々な機能から生じる中間処理結果は、メモリ208に記憶させることができる。メモリ208は、プロセッサ206によって実行可能なコード命令を記憶するための非一時的メモリをさらに備えることができる。メモリ208は、ディエンファサイジングフィルタおよびリサンブラ148からのオーディオ信号も記憶することができ、プロセッサ206からの要求があり次第、記憶されたオーディオ信号を出力204に提供する。

10

【0222】

時間領域デコーダによってデコードされた時間領域励振に含まれる音楽信号または他の信号中の量子化雑音を低減するためのデバイスおよび方法の説明は、例示だけであり、決して限定することが意図されていないことを当業者は理解されよう。他の実施形態は、本開示の利益を有する当業者には容易に思いつくことであろう。さらに、開示されたデバイスおよび方法は、線形予測(LP)ベースのコーデックの音楽コンテンツレンダリングを改善する既存の要求および課題への価値ある解決策を提供するようにカスタマイズすることができる。

【0223】

明確さのために、デバイスおよび方法の実現の日常的な特徴のすべてが示され、説明されるわけではない。もちろん、時間領域デコーダによってデコードされた時間領域励振に含まれる音楽信号中の量子化雑音を低減するためのデバイスおよび方法のそのような任意の実現の実現の開発において、数多くの実現固有の判定を、アプリケーション、システム、ネットワーク、およびビジネス関連の制約への適合など、開発者固有の目標を達成するために行われることが必要であり得ること、およびこれらの固有の目標は、実現によって、および開発者によって異なることが理解されよう。さらに、開発努力は、複雑で時間がかかり得るが、それにもかかわらず、本開示の利益を有する音響処理の分野の当業者には日常的なエンジニアリングの仕事であることが理解されよう。

20

【0224】

本開示によれば、本明細書に説明する構成要素、プロセス動作、および/またはデータ構造は、様々な種類のオペレーティングシステム、コンピューティングプラットフォーム、ネットワークデバイス、コンピュータプログラム、および/または汎用機を使用して実現することができる。さらに、配線で接続されたデバイス、フィールドプログラマブルゲートアレイ(FPGA)、特定用途向け集積回路(ASIC)など、より汎用でない性質のデバイスも使用できることを当業者は認識されよう。一連のプロセス動作を含む方法がコンピュータまたはマシンによって実現され、それらのプロセス動作をマシンによって可読な一連の命令として記憶させることができる場合、それらは有形的媒体上に記憶させることができる。

30

【0225】

本開示は、それらの非限定的、例示的实施形態として本明細書の上記に説明してきたが、これらの実施形態は、本開示の趣旨および性質から逸脱することなく添付の特許請求の範囲内で随意に変更することができる。

40

【符号の説明】

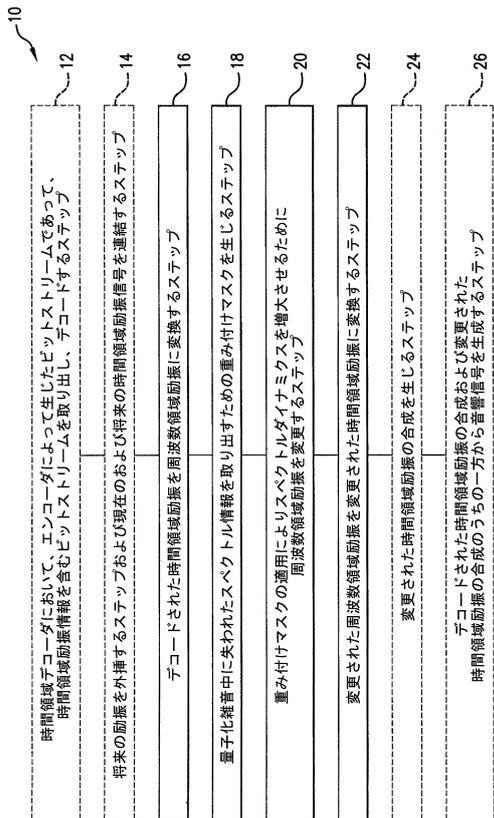
【0226】

- 100 デコーダ
- 102 受信機
- 103 デマルチプレクサ
- 104 時間領域励振デコーダ
- 106 過去の励振バッファメモリ
- 108 LP合成フィルタ

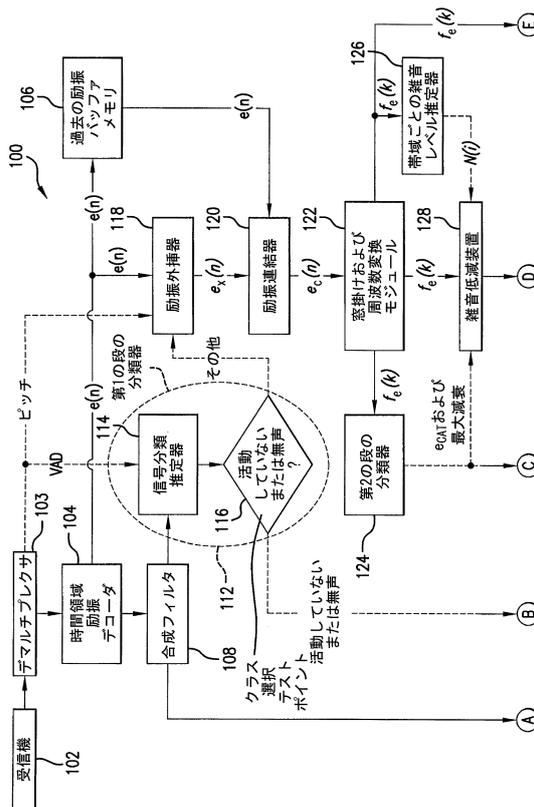
50

110	LP合成フィルタ	
112	第1の段の信号分類器	
114	信号分類推定器	
116	クラス選択テストポイント	
118	励振外挿器	
120	励振連結器	
122	窓掛けおよび周波数変換モジュール	
124	第2の段の信号分類器	
126	帯域ごとの雑音レベル推定器	
128	雑音低減装置	10
130	マスクビルダ	
131	スペクトルエネルギー正規化器	
132	エネルギー平均化器	
134	エネルギー平滑化器	
136	スペクトルダイナミクス変更器	
138	周波数/時間領域変換器	
140	フレーム励振抽出器	
142	上書き器	
144	判定テストポイント	
146	スイッチ	20
148	ディエンファサイジングフィルタおよびリサンブラ	
150	コア合成信号	
152	合成信号	
154	デジタル/アナログ変換器	
156	増幅器	
158	スピーカ	
200	デコーダ	
202	入力	
204	出力	
206	プロセッサ	30
208	メモリ	
A、B、C、D、E	コネクタ	

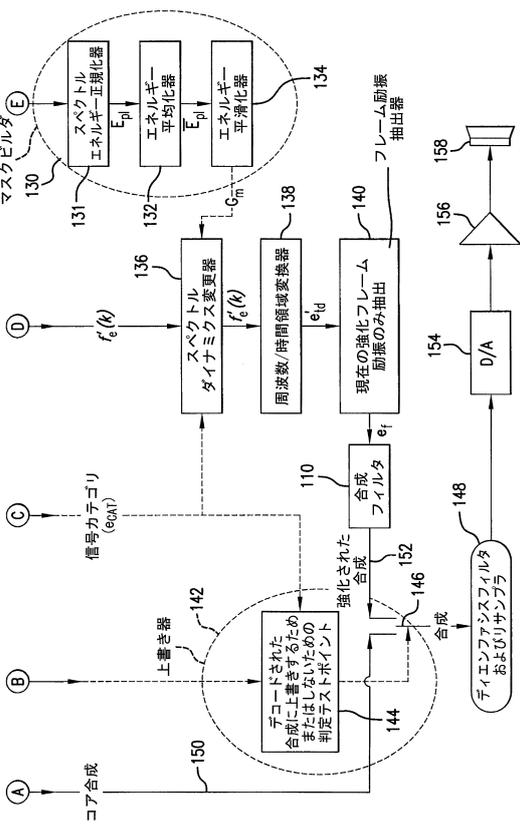
【図1】



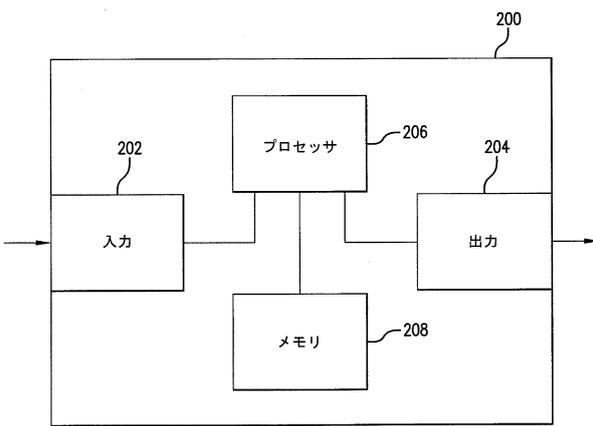
【図2A】



【図2B】



【図3】



---

フロントページの続き

(72)発明者 ミラン・ジェリネク

カナダ・ケベック・J 1 L・2 W 8・シャープブルック・エミリー - ネリニャン・1 3 5 5

審査官 安田 勇太

(56)参考文献 特表2011-514557(JP, A)

国際公開第2013/063688(WO, A1)

国際公開第1999/030315(WO, A1)

米国特許第07490036(US, B1)

(58)調査した分野(Int.Cl., DB名)

G 1 0 L 1 9 / 0 0 - 1 9 / 2 6

G 1 0 L 2 1 / 0 2 - 2 1 / 0 3 8 8