

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6239521号
(P6239521)

(45) 発行日 平成29年11月29日(2017.11.29)

(24) 登録日 平成29年11月10日(2017.11.10)

(51) Int.Cl.	F I
G 1 O L 19/18 (2013.01)	G 1 O L 19/18
G 1 O L 19/16 (2013.01)	G 1 O L 19/16 2 O O A
H O 3 M 7/30 (2006.01)	H O 3 M 7/30 Z

請求項の数 32 (全 30 頁)

(21) 出願番号	特願2014-539195 (P2014-539195)	(73) 特許権者	504226412
(86) (22) 出願日	平成24年11月1日(2012.11.1)		ヴォイスエイジ・コーポレーション
(65) 公表番号	特表2015-501452 (P2015-501452A)		カナダ・ケベック・H3R・2H6・モン
(43) 公表日	平成27年1月15日(2015.1.15)		トリオール・ルサヌ・ロード・750・
(86) 国際出願番号	PCT/CA2012/001011		スイート・250
(87) 国際公開番号	W02013/063688	(74) 代理人	100108453
(87) 国際公開日	平成25年5月10日(2013.5.10)		弁理士 村山 靖彦
審査請求日	平成27年10月29日(2015.10.29)	(74) 代理人	100064908
(31) 優先権主張番号	61/555,246		弁理士 志賀 正武
(32) 優先日	平成23年11月3日(2011.11.3)	(74) 代理人	100089037
(33) 優先権主張国	米国 (US)		弁理士 渡邊 隆
		(74) 代理人	100110364
			弁理士 実広 信哉

最終頁に続く

(54) 【発明の名称】 低レートCELPデコーダに関する非音声コンテンツの向上

(57) 【特許請求の範囲】

【請求項 1】

時間領域デコーダによって復号化された時間領域励振の合成物を修正するための装置であって、

前記復号化された時間領域励振の前記合成物のいくつかのカテゴリのうちの1つへの分類器と、

前記復号化された時間領域励振の周波数領域励振への変換器と、

前記復号化された時間領域励振の前記合成物が前記分類器によって分類されたカテゴリに応じる前記周波数領域励振の修正器と、

修正された周波数領域励振の修正された時間領域励振への変換器と、

前記復号化された時間領域励振の修正された合成物を生成するために、前記修正された時間領域励振が供給される合成フィルタとを含む、時間領域励振の合成物を修正するための装置。

【請求項 2】

前記修正器は、時間領域の寄与度の使用が停止されるカットオフ周波数の計算機を含む、請求項1に記載の時間領域励振の合成物を修正するための装置。

【請求項 3】

前記修正器は、

前記カットオフ周波数より高い周波数領域励振のゼロアースと、

正規化された周波数領域励振を生成するための、前記カットオフ周波数より低い周波数

10

20

領域励振の正規化器とを含む、請求項2に記載の時間領域励振の合成物を修正するための装置。

【請求項4】

前記修正器は、ランダムノイズ発生器と、ランダムノイズの前記正規化された周波数領域励振への加算器とを含む、請求項3に記載の時間領域励振の合成物を修正するための装置。

【請求項5】

前記修正器は、修正の前および後の前記周波数領域励振のエネルギーを使用する、整合利得の計算機を含み、

前記計算機は、前記修正された周波数領域励振を生成するために、前記正規化された周波数領域励振に前記整合利得を適用する、請求項3または4のいずれか一項に記載の時間領域励振の合成物を修正するための装置。

10

【請求項6】

前記分類器は、前記復号化された時間領域励振の前記合成物を受動的として、または能動的な無声として分類する、請求項2から5のいずれか一項に記載の時間領域励振の合成物を修正するための装置。

【請求項7】

前記復号化された時間領域励振の前記合成物が、前記分類器によって前記カテゴリのうちの所与の1つとして分類されたとき、前記合成フィルタのスムーザを含む、請求項1から6のいずれか一項に記載の時間領域励振の合成物を修正するための装置。

20

【請求項8】

前記周波数領域励振は、それぞれが周波数ビンに分割される周波数バンドに分割され、

前記修正器は、正規化された周波数領域励振を生成するために、周波数バンドによって変化する正規化ファクターを使用する、前記周波数領域励振の正規化器を含む、請求項1から4のいずれか一項に記載の時間領域励振の合成物を修正するための装置。

【請求項9】

前記正規化器は、前記周波数ビンを含む周波数バンド中の前記正規化された周波数領域励振の最大値の分数より低い前記正規化された周波数領域励振における前記周波数ビンのゼロアを含む、請求項8に記載の時間領域励振の合成物を修正するための装置。

【請求項10】

30

前記修正器は、修正の前および後の前記周波数領域励振のエネルギーを使用する、周波数バンド当たりの整合利得の計算機を含む、請求項8または9に記載の時間領域励振の合成物を修正するための装置。

【請求項11】

前記修正器は、第1の周波数より低い周波数バンドに関して、前記修正された周波数領域励振を生成するために、前記正規化された周波数領域励振に前記整合利得を適用するための計算機を含む、請求項10に記載の時間領域励振の合成物を修正するための装置。

【請求項12】

第1のより低い周波数と第2のより高い周波数との間の周波数バンドに関して、

前記正規化された周波数領域励振の周波数バンド当たりの最大値の発見器と、

40

前記周波数ビン中の前記正規化された周波数領域励振が、前記周波数バンドの前記最大値に比例する値に等しいか、またはそれより高いとき、周波数ビン当たりの増幅ファクターだけ、前記整合利得を増幅するための増幅器と、

前記周波数ビン中で前記修正された周波数領域励振を生成するために、前記周波数ビン中の前記正規化された周波数領域励振に前記増幅された整合利得を適用するための計算機とを含む、請求項10に記載の時間領域励振の合成物を修正するための装置。

【請求項13】

第1のより低い周波数と第2のより高い周波数との間の周波数バンドに関して、

前記正規化された周波数領域励振の周波数バンド当たりの最大値の発見器と、

前記周波数ビン中の前記正規化された周波数領域励振が、前記周波数バンドの前記最大

50

値に比例する値より低いとき、前記周波数バンドの周波数ビン当たりの減衰ファクターだけ、前記整合利得を減衰させるための減衰器と、

前記周波数ビン中で前記修正された周波数領域励振を生成するために、前記周波数ビン中の前記正規化された周波数領域励振に前記減衰された整合利得を適用するための計算機とを含む、請求項10に記載の時間領域励振の合成物を修正するための装置。

【請求項14】

所与の周波数より高い周波数バンドに関して、

前記正規化された周波数領域励振の周波数バンド当たりの最大値の発見器と、

前記整合利得のための傾斜の計算機であって、前記周波数ビン中の前記正規化された周波数領域励振が、前記周波数バンドの前記最大値に比例する値より高いとき、前記計算機は、前記整合利得に計算された傾斜を適用する、計算機と、

前記周波数ビン中で前記修正された周波数領域励振を生成するために、前記計算された傾斜が適用されている前記整合利得を、前記周波数ビン中の前記正規化された周波数領域励振に適用するための計算機とを含む、請求項10に記載の時間領域励振の合成物を修正するための装置。

【請求項15】

所与の周波数より高い周波数バンドに関して、

前記正規化された周波数領域励振の周波数バンド当たりの最大値の発見器と、

前記周波数ビン中の前記正規化された周波数領域励振が、前記周波数バンドの前記最大値に比例する値より低いとき、前記周波数バンドの周波数ビン当たりの増幅ファクターだけ、前記整合利得を減衰させるための減衰器と、

前記周波数ビン中で前記修正された周波数領域励振を生成するために、前記周波数ビン中の前記正規化された周波数領域励振に前記減衰された整合利得を適用するための計算機とを含む、請求項10に記載の時間領域励振の合成物を修正するための装置。

【請求項16】

符号化パラメータによって符号化された音響信号を復号化するための装置であって、

前記音響信号の符号化パラメータに応答する時間領域励振のデコードと、

前記時間領域励振の合成物を生成するために、前記復号化された時間領域励振に応答する合成フィルタと、

前記時間領域励振の前記合成物を修正するための、請求項1から15のいずれか一項に記載の前記装置とを含む、装置。

【請求項17】

時間領域デコードによって復号化された時間領域励振の合成物を修正するための方法であって、

前記復号化された時間領域励振の前記合成物をいくつかのカテゴリのうちの1つに分類するステップと、

前記復号化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、

前記復号化された時間領域励振の前記合成物が分類されたカテゴリに応じて前記周波数領域励振を修正するステップと、

前記修正された周波数領域励振を修正された時間領域励振に変換するステップと、

前記復号化された時間領域励振の修正された合成物を生成するために、前記修正された時間領域励振を合成するステップとを含む、時間領域励振の合成物を修正するための方法。

【請求項18】

前記周波数領域励振を修正するステップは、時間領域の寄与度の使用が停止されるカットオフ周波数を計算するステップを含む、請求項17に記載の時間領域励振の合成物を修正するための方法。

【請求項19】

前記周波数領域励振を修正するステップは、

前記カットオフ周波数より高い周波数領域励振をゼロに設定するステップと、

正規化された周波数領域励振を生成するために、前記カットオフ周波数より低い周波数領域励振を正規化するステップとを含む、請求項18に記載の時間領域励振の合成物を修正するための方法。

【請求項20】

前記周波数領域励振を修正するステップは、
ランダムノイズを発生するステップと、

前記ランダムノイズを前記正規化された周波数領域励振に加えるステップとを含む、請求項19に記載の時間領域励振の合成物を修正するための方法。

【請求項21】

前記周波数領域励振を修正するステップは、

修正の前および後の前記周波数領域励振のエネルギーを使用して、整合利得を計算するステップと、

前記修正された周波数領域励振を生成するために、前記整合利得を前記正規化された周波数領域励振に適用するステップとを含む、請求項19または20に記載の時間領域励振の合成物を修正するための方法。

【請求項22】

前記復号化された時間領域励振の前記合成物は、受動的として、または能動的な無声として分類される、請求項18から21のいずれか一項に記載の時間領域励振の合成物を修正するための方法。

【請求項23】

前記復号化された時間領域励振の前記合成物が、前記カテゴリのうちの所与の1つとして分類されたとき、前記修正された時間領域励振の合成を平滑化するステップを含む、請求項17から22のいずれか一項に記載の時間領域励振の合成物を修正するための方法。

【請求項24】

前記周波数領域励振は、それぞれが周波数ビンに分割される周波数バンドに分割され、

前記周波数領域励振を修正するステップは、正規化された周波数領域励振を生成するために、周波数バンドによって変化する正規化ファクターを使用して前記周波数領域励振を正規化するステップを含む、請求項17から20のいずれか一項に記載の時間領域励振の合成物を修正するための方法。

【請求項25】

前記周波数領域励振を修正するステップは、前記周波数ビンを含む周波数バンド中の前記正規化された周波数領域励振の最大値の分数より低い前記正規化された周波数領域励振における前記周波数ビンをゼロに設定するステップを含む、請求項24に記載の時間領域励振の合成物を修正するための方法。

【請求項26】

前記周波数領域励振を修正するステップは、修正の前および後の前記周波数領域励振のエネルギーを使用して、周波数バンド当たりの整合利得を計算するステップを含む、請求項24または25に記載の時間領域励振の合成物を修正するための方法。

【請求項27】

前記周波数領域励振を修正するステップは、第1の周波数より低い周波数バンドに関して、前記修正された周波数領域励振を生成するために、前記整合利得を前記正規化された周波数領域励振に適用するステップを含む、請求項26に記載の時間領域励振の合成物を修正するための方法。

【請求項28】

第1のより低い周波数と第2のより高い周波数との間の周波数バンドに関して、

前記正規化された周波数領域励振の周波数バンド当たりの最大値を発見するステップと、

前記周波数ビン中の前記正規化された周波数領域励振が、前記周波数バンドの前記最大値に比例する値に等しいか、またはそれより高いとき、周波数ビン当たりの増幅ファクターだけ、前記整合利得を増幅するステップと、

10

20

30

40

50

前記周波数ビン中で前記修正された周波数領域励振を生成するために、前記周波数ビン中の前記正規化された周波数領域励振に前記増幅された整合利得を適用するステップとを含む、請求項26に記載の時間領域励振の合成物を修正するための方法。

【請求項 2 9】

第1のより低い周波数と第2のより高い周波数との間の周波数バンドに関して、

前記正規化された周波数領域励振の周波数バンド当たりの最大値を発見するステップと

、
前記周波数ビン中の前記正規化された周波数領域励振が、前記周波数バンドの前記最大値に比例する値より低いとき、前記周波数バンドの周波数ビン当たりの減衰ファクターだけ、前記整合利得を減衰させるステップと、

10

前記周波数ビン中で前記修正された周波数領域励振を生成するために、前記周波数ビン中の前記正規化された周波数領域励振に前記減衰された整合利得を適用するステップとを含む、請求項26に記載の時間領域励振の合成物を修正するための方法。

【請求項 3 0】

所与の周波数より高い周波数バンドに関して、

前記正規化された周波数領域励振の周波数バンド当たりの最大値を発見するステップと

、
前記周波数ビン中の前記正規化された周波数領域励振が、前記周波数バンドの前記最大値に比例する値より高いとき、前記整合利得のための傾斜を計算し、そして前記計算された傾斜を前記整合利得に適用するステップと、

20

前記周波数ビン中で前記修正された周波数領域励振を生成するために、前記計算された傾斜が適用されている前記整合利得を、前記周波数ビン中の前記正規化された周波数領域励振に適用するステップとを含む、請求項26に記載の時間領域励振の合成物を修正するための方法。

【請求項 3 1】

所与の周波数より高い周波数バンドに関して、

前記正規化された周波数領域励振の周波数バンド当たりの最大値を発見するステップと

、
前記周波数ビン中の前記正規化された周波数領域励振が、前記周波数バンドの前記最大値に比例する値より低いとき、前記周波数バンドの周波数ビン当たりの増幅ファクターだけ、前記整合利得を減衰させるステップと、

30

前記周波数ビン中で前記修正された周波数領域励振を生成するために、前記減衰された整合利得を、前記周波数ビン中の前記正規化された周波数領域励振に適用するステップとを含む、請求項26に記載の時間領域励振の合成物を修正するための方法。

【請求項 3 2】

符号化パラメータによって符号化された音響信号を復号化するための方法であって、

前記音響信号の符号化パラメータに応答して時間領域励振を復号化するステップと、

前記時間領域励振の合成物を生成するために、前記復号化された時間領域励振を合成するステップと、

請求項17から31のいずれか一項に記載の、前記時間領域励振の前記合成物を修正するための前記方法とを含む、方法。

40

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本開示は、時間領域デコーダによって復号化された時間領域励振の合成物を修正するための装置および方法に関する。

【背景技術】

【0 0 0 2】

最新式の会話コーデックは、ビットレートが約8kbpsの状態で、明瞭な音声信号を極めて高品質で表すことができ、16kbpsのビットレートにおける明瞭さに近づけることができ

50

る。低ビットレートにおいてさえも、この音声の高品質を維持するためには、マルチモーダル(multi modal)符号化スキームを使用することができる。通常、入力音響信号は、その特性を反映する異なるカテゴリの間に分割される。たとえば、異なるカテゴリは、有声、無声および出だし音(onset: 頭子音)のカテゴリを含むことができる。コーデックでは、これらのカテゴリのすべてに最適化される異なる符号化モードが使用される。

【先行技術文献】

【特許文献】

【0003】

【特許文献1】Milan JelinekおよびPhilippe Gournay、PCT特許出願国際公開第03102921号、「A method and device for efficient frame erasure concealment in linear predictive based speech codecs」

10

【特許文献2】T. Vaillancourt他、PCT特許出願国際公開第2007073604号、「Method and device for efficient frame erasure concealment in speech codecs」

【非特許文献】

【0004】

【非特許文献1】3GPP TS 26.190, 「Adaptive Multi-Rate-Wideband(AMR-WB)speech codec」;Transcoding functions

【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頁、1988年2月

20

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかし、いくつかの展開されている音声コーデックでは、このマルチモーダルのアプローチが使用されておらず、明瞭な音声と異なる音響信号では、特に低いビットレートにおいて準最適な品質になる。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対するあらゆる修正が、コーデックの相互運用性を壊すことになるということが原因で、エンコードを修正することは困難である。しかし、デコードの修正は、受信側で受け取られる品質を向上させるように実施することができる。

【課題を解決するための手段】

30

【0006】

第1の態様によれば、本開示は、時間領域デコードによって復号化された時間領域励振の合成物を修正するための装置に関し、該装置は、復号化された時間領域励振の合成物のいくつかのカテゴリのうちの1つへの分類器と;復号化された時間領域励振の周波数領域励振への変換器と;復号化された時間領域励振の合成物が分類器によって分類されたカテゴリに応じる、周波数領域励振の修正器と;修正された周波数領域励振の修正された時間領域励振への変換器と;復号化された時間領域励振の修正された合成物を生成するために、修正された時間領域励振が供給される合成フィルタとを含む。

【0007】

別の態様によれば、本開示は、符号化パラメータによって符号化された音響信号を復号化するための装置に関し、その装置は、音響信号の符号化パラメータにตอบสนองする時間領域励振のデコードと;前記時間領域励振の合成物を生成するように、復号化された時間領域励振にตอบสนองする合成フィルタと;時間領域励振の合成物を修正するための上記に述べた装置とを含む。

40

【0008】

第3の態様によれば、本開示は、時間領域デコードによって復号化された時間領域励振の合成物を修正するための方法に関し、その方法は、復号化された時間領域励振の合成物をいくつかのカテゴリの1つに分類するステップと;復号化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;復号化された時間領域励振の合成物が分類されたカテゴリに応じて、周波数領域励振を修正するステップと;修正された周波数領域励振を修正され

50

た時間領域励振に変換するステップと;復号化された時間領域励振の修正された合成物を生成するために、修正された時間領域励振を合成するステップとを含む。

【0009】

さらなる態様によれば、本開示は、符号化パラメータによって符号化された音響信号を復号化するための方法に係わり、その方法は、音響信号の符号化パラメータに応答して時間領域励振を復号化するステップと;前記時間領域励振の合成物を生成するために、復号化された時間領域励振を合成するステップと;時間領域励振の合成物を修正するための上記に述べた方法とを含む。

【0010】

時間領域励振の合成物を修正するための装置および方法に関する前述および他の特徴は、限定しない実施例として設けた次の非限定の記述を、添付図面を参照して読むと、より明らかになるはずである。

【図面の簡単な説明】

【0011】

【図1】受動的なフレームおよび能動的な無声フレームを向上させるためのCELPデコーダの修正を示す簡単化したブロック図である。

【図2】受動的なフレームおよび能動的な無声フレームを向上させるためのCELPデコーダの修正を示す詳細なブロック図である。

【図3】一般のオーディオフィームを向上させるためのCELPデコーダの修正を示す簡単化したブロック図である。

【図4】一般のオーディオフィームを向上させるためのCELPデコーダの修正を示す詳細なブロック図である。

【発明を実施するための形態】

【0012】

本開示は、相互運用性を維持し、かつ知覚品質を高めるように、デコーダ側でマルチモーダルの復号化を実施するアプローチに関するものである。本開示では、参考文献[3GPP TS 26.190,「Adaptive Multi-Rate-Wideband(AMR-WB)speech codec」;Transcoding functions]に記載されているAMR-WBが、例示的な例として使用されているが、このアプローチは、他のタイプの低ビットレートの音声デコーダに同様に適用することができることを念頭に置くべきであり、この文献の全部の内容は、参照によって本明細書に組み込まれる。

【0013】

図1を参照すると、このマルチモーダルの復号化を達成するためには、時間領域励振デコーダ102が、まず、受け取られたビットストリーム101、たとえばAMR-WBビットストリームを全部復号化して、完全な時間領域の符号励振線形予測(CELP:Code-Excited Linear Prediction)による復号化された励振を得る。復号化された時間領域励振は、線形予測(LP:Linear Prediction)合成フィルタ103によって処理され、デコーダの内部サンプリング周波数における音声/音響信号の時間領域合成物が得られる。AMR-WBでは、この内部サンプリング周波数は12.8kHzであるが、別のコーデックでは、それは、異なることができるはずである。

【0014】

LP合成フィルタ103からの現在のフレームの時間領域合成物は、ビットストリーム101からの音声区間検出(VAD:voice activity detection)情報109が提供される分類器104-105-106-301(図1、図2および図3)によって処理される。分類器104-105-106-301は、時間領域合成物を分析して受動的な音声、能動的な有声音声、能動的な無声音声または一般のオーディオのいずれかとして分類する。受動的な音声(1051で検出される)は、音声バーストの間のすべてのバックグラウンドのノイズを含み、能動的な有声音声(1061で検出される)は、有声音声の特徴を有する能動的な音声バーストの間のフレームを表し、能動的な無声音声(1062で検出される)は、無声音声の特徴を有する音声バーストの間のフレームを表し、そして一般のオーディオ(3010で検出される)は、音楽または残響音声を表す。他のカテゴリが、追加されるか、または上記のカテゴリから導き出されることができる。開示するアプローチは

、特に、ただし排他的でなく、受動的な音声、能動的な無声音声および一般のオーディオの知覚品質を向上させることを目指している。

【 0 0 1 5 】

一度時間領域合成物のカテゴリが決定されると、変換器/修正器107は、非オーバーラップ周波数変換を使用して、時間領域励振デコーダ102からの復号化された励振を周波数領域に変換する。同様にオーバーラップ変換を使用することができるが、しかしそれは、端から端まで(end-to-end: エンドツーエンド)の遅延が増加するという意味を含み、それは、ほとんどの場合望まれない。次いで、励振の周波数表現は、変換器/修正器107中で異なる周波数バンドに分割される。周波数バンドは、固定のサイズを有することができるか、またはクリティカルバンドに依存することができる[全体の内容が参照によって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頁、1988年2月参照]か、または任意の他の組み合わせである。次いで、バンド当たりのエネルギーが計算され、変換器/修正器107中のメモリに保存されて、修正がグローバルなフレームのエネルギーレベルを変化させていないことを保証するために、再成形プロセスの後で使用される。

【 0 0 1 6 】

変換器/修正器107によって行われる周波数領域中の励振の修正は、合成物の分類によって異なることができる。受動的な音声および能動的な無声音声について、再成形は、ノイズを加えることによって低周波数を正規化することと、高周波数コンテンツをノイズだけと置き換えることとからなることができる。復号化された時間領域合成物のカットオフ周波数、すなわち低周波数と高周波数との間のリミットは、約1~1.2kHzの値に固定することができる。復号化された時間領域合成物の低周波数コンテンツのいくらかは、修正されないフレームと修正されたフレームとの間で切り替えられるとき、不自然な結果(artifact)を防止するために保存される。時間領域励振デコーダ102からの復号化されたピッチの関数として周波数ピンを選ぶことによって、カットオフ周波数をフレームごとに可変にすることも可能である。修正プロセスは、低ビットレートの音声コーデックと関連付けられた或る種類の電気ノイズを除去するという効果を有する。修正プロセスの後、6kHzより高い周波数におけるLPフィルタの利得の低下を補償するように、それらの周波数についてエネルギーをわずかに増加させることによって、周波数バンド当たりの初期エネルギーレベルを取り戻すために、周波数バンドごとの利得整合が適用される。

【 0 0 1 7 】

一般のオーディオとして分類されたフレームでは、変換器/修正器107中の処理が異なる。まず、正規化が、すべてのバンドについて周波数バンドごとに実施される。正規化動作では、バンド内の最大周波数値の分数より低い周波数バンド内部のすべてのピンは、ゼロに設定される。より高周波数バンドには、より多いピンが、バンドごとにゼロに設定される。これは、ビット配分が高いが、より多くのビットがより低周波数に割り当てられている周波数定量化スキームをシミュレートしている。正規化プロセスの後、ゼロに設定されたピンをランダムノイズに置き換えるために、ノイズ充填(noise fill)を適用することができるが、しかしビットレートに依存して、ノイズ充填は、常には使用されない。修正プロセスの後、周波数バンドごとの利得整合が、周波数バンド当たりの初期エネルギーレベルを取り戻すために適用され、そして一般のオーディオ入力の場合、LPフィルタのシステムの過小評価を補償するために、ビットレートに依存する傾斜是正が周波数バンドに沿って適用される。一般のオーディオ経路に対する別の差別化は、利得整合がすべての周波数ピンにわたって適用されないことに起因する。一般のオーディオのスペクトルが、通常、音声よりピーキー(peaky)なので、スペクトルのパルスを識別して、それをいくぶん強調することが可能であるとき、知覚品質は向上する。そうするためには、傾斜是正による十分な利得整合が、周波数バンド内部のもっとも高いエネルギーのピンだけに適用される。もっとも低いエネルギーのピンに関しては、わずかな利得整合だけがそれらのピンに適用される。これによって、スペクトルのダイナミックが高められることになる。

【 0 0 1 8 】

励振周波数の再成形および利得整合の後、修正された時間領域励振を得るために、変換器/修正器107が逆周波数変換を適用する。この修正された励振は、LP合成フィルタ108によって処理され、修正された時間領域合成物が得られる。ディエンファサイジング(de-emphasizing)フィルタおよび再サンブラ112における16kHz(AMR-WBの例について)までの最終的なディエンファシス(de-emphasis)および再サンプリングの前に、上書き器110は、LP合成フィルタ103からの時間領域の復号化された合成物を、時間領域の復号化された合成物の分類に依存して修正されたLP合成フィルタ108からの時間領域合成物によって単に上書きする。

【 0 0 1 9 】

受動的な音声の場合、能動的な無声音声の修正に比べて唯一の差は、より平坦なノイズ変動を与えるようにLP合成フィルタ108を平滑化するために、スムーザ111を使用することである。残された修正は、能動的な無声経路についてのものと同じである。次の文章では、開示するアプローチの実装形態のより詳細な実施例を、図2を参照して述べる。

【 0 0 2 0 】

1) 信号分類

図2を参照すると、分類器104-105-106-301は、デコーダで、上文に記載のように、修正が適用されるビットレートに関して音声/音響信号の時間領域合成物1021进行分类する。図面を簡単にするために、LP合成フィルタ103は、図2に示していない。デコーダでの分類は、全体の内容が参照によって本明細書に組み込まれる、参考文献[Milan JelinekおよびPhilippe Gournay、PCT特許出願国際公開第03102921号、「A method and device for efficient frame erasure concealment in linear predictive based speech codecs」]および[T. Vaillancourt他、PCT特許出願国際公開第2007073604号、「Method and device for efficient frame erasure concealment in speech codecs」]に記載されているものと同様であると共に、さらに、一般のオーディオ検出のためのいくつかの適用である。デコーダでのフレームの分類のために、次のパラメータが使用される;正規化相関 r_x 、スペクトル傾斜基準 e_t 、ピッチ安定性カウンタ p_c 、現在のフレームの最後における音響信号の相対的なフレームエネルギー E_s およびゼロ交差カウンタ z_c である。信号进行分类するために使用されるこれらのパラメータの計算は、下記で説明される。

【 0 0 2 1 】

正規化相関 r_x は、音声/音響信号の時間領域合成物 $S_{out}(n)$ に基づき、フレームの最後で計算される。時間領域励振デコーダ102からの最後のサブフレームのピッチ遅延が使用される。より具体的には、正規化相関 r_x は、次のようにピッチに同期して計算される。

【 0 0 2 2 】

【 数 1 】

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

【 0 0 2 3 】

ただし、 $x(n)=S_{out}(n)$ であり、 T は、最後のサブフレームのピッチ遅延であり、 $t=L-T$ であり、そして 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遅延が、 $3N/2$ より大きい場合(N はサブフレームサイズである)、 T は、最後の2つのサブフレームの平均のピッチ遅延に設定される。

【 0 0 2 4 】

したがって、正規化相関 r_x は、音声/音響信号の時間領域合成物 $S_{out}(n)$ を使用して計算される。サブフレームサイズ(64サンプル)より小さいピッチ遅延に関して、正規化相関は

10

20

30

40

50

、 $t=L-T$ の瞬間および $t=L-2T$ の瞬間で2回計算され、正規化相関 r_x は、これらの2つの計算値の平均値として与えられる。

【 0 0 2 5 】

スペクトルの傾斜パラメータ e_t は、エネルギーの周波数分布に関する情報を含む。非限定の実施例として、デコーダでのスペクトルの傾斜は、時間領域合成物の第1の正規化自己相関係数として推定される。これは、最後の3つのサブフレームに基づき、次のように計算される。

【 0 0 2 6 】

【 数 2 】

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

10

【 0 0 2 7 】

ただし、 $x(n)=S_{out}(n)$ は、時間領域合成信号であり、 N は、サブフレームサイズであり、そして L は、フレームサイズである(AMR-WBの例では $N=64$ および $L=256$)。

【 0 0 2 8 】

ピッチ安定性カウンタ pc は、ピッチ期間の変動を評価する。これは、次のようにデコーダで計算される。

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

【 0 0 2 9 】

値 p_0 、 p_1 、 p_2 および p_3 は、現在のフレームの4つのサブフレームからの閉ループのピッチ遅延に対応する(AMR-WBの例では)。

【 0 0 3 0 】

相対的なフレームエネルギー E_s は、dBでの現在のフレームエネルギー E_f とその長期平均値 E_{1t} との間の差として計算される。

$$E_s = E_f - E_{1t} \quad (4)$$

30

【 0 0 3 1 】

ただし、現在のフレームエネルギー E_f は、フレームの最後でピッチに同期して計算されるdBでの時間領域合成物 $S_{out}(n)$ のエネルギーであり、次のようである。

【 0 0 3 2 】

【 数 3 】

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

40

【 0 0 3 3 】

ただし、 $L=256$ (AMR-WBの例では)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のピッチ遅延である。 T がサブフレームサイズより小さい場合、 T は、 $2T$ に設定される(短いピッチ遅延については2つのピッチ期間を使用して計算されるエネルギー)。

【 0 0 3 4 】

長期平均化エネルギーは、次の関係を使用して能動的な音声フレームに対して更新される。

$$E_{1t} = 0.99 E_{1t} + 0.01 E_f \quad (6)$$

50

【 0 0 3 5 】

最後のパラメータは、時間領域合成物 $S_{out}(n)$ の1つのフレームに対して計算されるゼロ交差カウンタ zc である。非限定の実施例として、ゼロ交差カウンタ zc は、時間領域合成物の符号が、その期間の間、正から負に変化する回数をカウントする。

【 0 0 3 6 】

分類をよりロバストなものにするために、複数の分類パラメータを合わせて考慮し、1つのメリット関数 f_m を形成するものとする。その目的のために、分類パラメータは、まず、線形関数を使用してスケーリングされる。パラメータ p_x を考えると、そのスケーリングされたバージョンは、次式を使用して得られる。

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

10

【 0 0 3 7 】

スケーリングされたピッチ安定性カウンタ pc は、0と1の間にクリップされる。関数の係数 k_p および c_p は、そのパラメータのそれぞれについて実験的に見つけられている。実装形態のこの実施例で使用される値は、テーブル1(表1)に要約している。

【 0 0 3 8 】

【表 1】

テーブル 1. デコーダにおけるフレーム分類パラメータ

およびそれらのそれぞれのスケーリング関数の係数

20

パラメータ	意味	k_p	c_p
r_x	正規化相関	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	相対的な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

【 0 0 3 9 】

メリット関数は、次のように定義される。

30

【 0 0 4 0 】

【数 4】

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

【 0 0 4 1 】

ただし、上付き文字 s は、パラメータのスケーリングされたバージョンを示す。

40

【 0 0 4 2 】

次いで、フレームの分類が、メリット関数 f_m を使用し、テーブル2(表2)に要約したルールに従って行われる。

【 0 0 4 3 】

【表 2】

テーブル 2. デコーダにおける信号分類ルール

前のフレームクラス	ルール	現在のフレームクラス
出だし音 有声 有声遷移 人為的出だし音 一般のオーディオ音響	$f_m \geq 0.63$	有声
	$0.39 \leq f_m \leq 0.63$	有声遷移
	$f_m < 0.39$	無声
無声遷移 無声	$f_m > 0.56$	出だし音
	$0.56 \geq f_m > 0.45$	無声遷移
	$f_m \leq 0.45$	無声
現在のフレームの VAD 情報		
	VAD=0	無声

10

【 0 0 4 4 】

この分類に加えて、エンコーダによる音声区間検出(VAD:voice activity detection)に関する情報109は、AMR-WBの例の場合のように、ビットストリーム101中で伝達することができる(図1)。それゆえ、エンコーダが現在のフレームを能動的なコンテンツ(VAD=1)として見なしているのか、または受動的なコンテンツ(バックグラウンドノイズ、VAD=0)として見なしているのかを規定するために、1つのビットがビットストリーム101中に送られる。VAD情報が、コンテンツは受動的であると示すとき、分類器の一部分104、105、106および301は、分類を無声として上書きする。

20

【 0 0 4 5 】

分類スキームは、さらに、一般のオーディオ検出を含む(図3の分類器の一部分301参照)。一般のオーディオのカテゴリは、音楽、残響音声を含み、そしてまた、バックグラウンドの音楽を含むことができる。分類の第2のステップは、分類器104-105-106-301が、十分な確かさで、現在のフレームを一般のオーディオとして分類することができることを可能にする。2つのパラメータが、この第2の分類のステップを実現するために使用される。パラメータの1つは、方程式(5)で公式化される全フレームエネルギー E_f である。

30

【 0 0 4 6 】

まず、これまでの40個の全フレームエネルギー変動の平均値

【 0 0 4 7 】

【数 5】

$$\overline{E_{df}}$$

40

【 0 0 4 8 】

が、次の関係を使用して計算される。

【 0 0 4 9 】

【数 6】

$$\overline{E}_{df} = \frac{\sum_{i=40}^{i=1} \Delta'_E}{40}; \quad \text{ただし } \Delta'_E = E'_f - E_f^{(i-1)} \quad (9)$$

【0050】

次いで、最後の15個のフレームにわたるエネルギー変動履歴の統計的偏差 σ_E が、次の関係を使用して決定される。

【0051】

10

【数 7】

$$\sigma_E = 0.7745967 \cdot \sqrt{\sum_{i=15}^{i=1} \frac{(\Delta'_E - \overline{E}_{df})^2}{15}} \quad (10)$$

【0052】

その結果得られた偏差 σ_E は、復号化された合成物のエネルギー安定性についての表示を与える。通常、音楽は、エネルギー安定性が音声より高い(エネルギー変動履歴の統計的偏差がより低い)。

20

【0053】

さらに、方程式(5)で公式化されるフレームエネルギー E_f が -12dB より高いとき、無声として分類される2つのフレームの間の間隔 N_{uv} を評価するために、第1のステップの分類が使用される。フレームが無声として分類され、フレームエネルギー E_f が -9dB より大きいときは、信号が無声であるが無音でないことを意味し、方程式(6)で公式化される長期の能動的な音声エネルギー E_{l_i} が 40dB より低い場合、無声間隔カウンタが 16 に設定され、そうでなければ、無声間隔カウンタ N_{uv} が 8 だけ減少される。カウンタ N_{uv} は、また、能動的な音声信号については 0 と 300 の間に制限され、受動的な音声信号については 0 と 125 の間に制限される。この例示的な例では、能動的な音声信号と受動的な音声信号との間の差は、ビットストリーム 101 中に含まれる音声区間検出(VAD)情報から推測することができることに注意されたい。

30

【0054】

長期平均値は、能動的な音声信号に関して、次のように、この無声フレームカウンタから導き出される。

【0055】

【数 8】

$$N_{uvn} = 0.9 \cdot N_{uvn} + 0.1 \cdot N_{uv} \quad (11)$$

40

【0056】

そして、受動的な音声信号に関しては、次のようである。

【0057】

【数 9】

$$N_{uvn} = 0.95 \cdot N_{uvn} \quad (12)$$

50

【 0 0 5 8 】

さらにまた、長期平均値が大変大きく、偏差 σ_E が大きいとき、たとえば、実装形態の現在の実施例では、

【 0 0 5 9 】

【 数 1 0 】

$$N_{uv_{it}} > 140$$

10

【 0 0 6 0 】

および $\sigma_E > 5$ であるとき、長期平均値は、次のように修正される。

【 0 0 6 1 】

【 数 1 1 】

$$N_{uv_{it}} = 0.2 \cdot N_{uv_{it}} + 80 \quad (13)$$

【 0 0 6 2 】

20

無声として分類されたフレームの間のフレーム数の長期平均値についてのこのパラメータは、フレームを一般のオーディオとして見なすべきかどうかを決定するために、分類器 104-105-106-301によって使用される。無声フレームが時間内でより接近していればいるほど、フレームが音声特性をますます有しそうである(一般のオーディオである可能性がより少ない)。例示的な例では、フレームを一般のオーディオとして見なすのかどうかを決定するしきい値 G_A は、次のように定義される。

【 0 0 6 3 】

【 数 1 2 】

$$\text{フレームは、} G_A \text{ である : } N_{uv_{it}} > 140 \text{ かつ } \Delta'_E < 12 \text{ の場合} \quad (14)$$

30

【 0 0 6 4 】

方程式(9)で定義されるパラメータ

【 0 0 6 5 】

【 数 1 3 】

$$\Delta_E^t$$

40

【 0 0 6 6 】

は、大きいエネルギー変動を一般のオーディオとして分類しないように加えられ、その変動は、能動的な音声として維持される。

【 0 0 6 7 】

励振に対して実施される修正は、フレームの分類によって決定され、或るタイプのフレームには、全く修正が施されない。次のテーブル3(表3)に、修正を実施することができる、またはできない場合を要約する。

【 0 0 6 8 】

50

【表 3】

テーブル 3. 励振修正に関する信号カテゴリ

フレーム分類	音声区間が検出された? Y/N	カテゴリ	修正 Y/N
出だし音 有声 無声遷移 人為的出だし音	Y (VAD=1)	能動的な声	N
一般のオーディオ音響	Y	一般のオーディオ	Y*
有声遷移 無声	Y	能動的な無声	Y
出だし音 有声 無声遷移 人為的出だし音 一般のオーディオ音響 有声遷移 無声	N	受動的なオーディオ	Y

【 0 0 6 9 】

* 一般のオーディオカテゴリは、実装形態に依存して、修正してもよく、またはしなくてもよい。たとえば、一般のオーディオは、受動的であるだけ修正してよく、あるいは、一般のオーディオは、能動的であるときだけ修正してもよく、そして、いつでも修正してもよく、または、いつでも修正しなくてもよい。

【 0 0 7 0 】

2) 周波数変換

周波数領域の修正フェーズの間、励振は、変換領域中で表すことが必要である。たとえば、時間-周波数変換は、25Hzの周波数分解能を与えるタイプII DTC(離散コサイン変換; Discrete Cosine Transform)を使用して、変換器/修正器107の時間-周波数領域変換器201によって達成されるが、しかし、あらゆる他の適切な変換を使用することができる。別の変換が使用される場合、周波数分解能(上記で定義した)、周波数バンド数およびバンド当たりの周波数ビン数(さらに下記で定義する)は、それに従って改訂することが必要である。時間-周波数領域変換器201で計算される時間領域CELP励振の周波数表現 f_e は、下記で与えられる。

【 0 0 7 1 】

【数 1 4】

$$f_e(k) = \begin{cases} \sqrt{\frac{1}{L}} \cdot \sum_{n=0}^{L-1} e_{td}(n), & k = 0 \\ \sqrt{\frac{2}{L}} \cdot \sum_{n=0}^{L-1} e_{td}(n) \cdot \cos\left(\frac{\pi}{L} \left(n + \frac{1}{2}\right) k\right), & 1 \leq k \leq L-1 \end{cases} \quad (15)$$

【 0 0 7 2 】

ただし、 $e_{td}(n)$ は、時間領域CELP励振であり、 L はフレーム長である。AMR-WBの例では、フレーム長は、12.8kHzの対応する内部サンプリング周波数に対して256サンプルである

。

【 0 0 7 3 】

時間領域CELPデコーダ、たとえば102では、時間領域励振信号は、

$$e_{td}(n) = bv(n) + gc(n) \quad (15)$$

によって与えられる、ただし、 $v(n)$ は、適応コードブック寄与度(adaptive codebook contribution)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度(fixed codebook contribution)であり、 g は固定コードブック利得である。

【 0 0 7 4 】

3) バンド当たりのエネルギーの分析

時間領域励振に対するあらゆる修正の前に、変換器/修正器107は、それ自体が周波数領域励振のバンド当たりのエネルギー E_b を計算するサブ計算機209を含む利得計算機208-209-210を含み、そして励振スペクトル再成形後にエネルギーを調整するために、計算したバンド当たりのエネルギー E_b をメモリに保存する。12.8kHzのサンプリング周波数では、エネルギーは、次のように、サブ計算機209によって計算することができる。

【 0 0 7 5 】

【数 1 5 】

$$E_b(i) = \sqrt{\sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_e(j)^2} \quad (16)$$

【 0 0 7 6 】

ただし、 C_{Bb} は、バンド当たりの累積周波数ピンを表し、 B_b は、次のように定義される周波数バンド当たりのピン数を表す。

【 0 0 7 7 】

【数 1 6 】

$$B_b = \{4, 4, 4, 4, 4, 5, 6, 6, 6, 8, 8, 10, 11, 13, 15, 18, 22, 16, 16, 20, 20, 20, 16\}$$

$$C_{Bb} = \left\{ \begin{array}{l} 0, 8, 12, 16, 20, 25, 31, 37, 43, 51, 59, 69, 80, 93, \\ 108, 126, 148, 164, 180, 200, 220, 240 \end{array} \right\}$$

【 0 0 7 8 】

低周波数バンドは、参考文献[Milan JelinekおよびPhilippe Gournay、PCT特許出願国際公開第03102921号、「A method and device for efficient frame erasure concealment in linear predictive based speech codecs」]に記載のようにクリティカルなオーディオバンドに対応することができ、この文献の全部の内容は、参照によって本明細書に組み込まれるが、しかし、3700Hzより高い周波数バンドは、それらのバンド中の起こり得るスペクトルのエネルギー変動をより良く調和させるためには、少しだけより低いことがある。スペクトルのバンドのあらゆる他の構成も可能である。

【 0 0 7 9 】

4) 受動的なフレームおよび能動的な無声フレームのための励振修正

a) 時間領域寄与度対ノイズ充填のカットオフ周波数

受動的なフレームおよび能動的な無声フレームについて、非修正の励振と修正された励振との間の明瞭な切り替えを達成するためには、時間領域励振の寄与率の少なくともより低い周波数が保存される。変換器/修正器107は、時間領域寄与度の使用を停止すべき周波数を決定するカットオフ周波数計算機203を含み、カットオフ周波数 f_c は、1.2kHzの最小

10

20

30

40

50

値を有する。これは、復号化された励振の最初の1.2kHzが常に保存され、このカットオフ周波数は、時間領域励振デコーダ102からの復号化されたピッチ値に依存していて、より高くすることができることを意味する。第8次高調波が、すべてのサブフレームのもっとも低いピッチから計算され、時間領域寄与度は、この第8次高調波まで保存される。第8次高調波の推定は、次のように計算される。

【 0 0 8 0 】

【 数 1 7 】

$$h_{8th} = \frac{(8 \cdot F_s)}{\min_{0 \leq i < N_{sub}}(T(i))} \quad (17)$$

10

【 0 0 8 1 】

ただし、 $F_s=12800\text{Hz}$ であり、 N_{sub} は、サブフレーム数であり、 T は、復号化されたサブフレームのピッチである。 N_b が、周波数範囲 L_f に含まれる最大周波数バンドであって、すべての $i < N_b$ について、第8次高調波が配置されているバンドを見つけるために、次の不等式がなお検証されるもっとも高いバンドを探し求めることによって、検証が行われる。

【 0 0 8 2 】

【 数 1 8 】

20

$$(h_{8th} \geq L_f(i)) \quad (18)$$

【 0 0 8 3 】

ただし、 L_f は、次のように定義される。

【 0 0 8 4 】

【 数 1 9 】

$$L_f = \left\{ \begin{array}{l} 175, 275, 375, 475, 600, 750, 900, 1050, 1250, 1450, 1700, 1975, \\ 2300, 2675, 3125, 3675, 4075, 4475, 4975, 5475, 5975, 6375 \end{array} \right\}$$

30

【 0 0 8 5 】

L_f 中のその周波数バンドの指標は、

【 0 0 8 6 】

【 数 2 0 】

$$i_{8th}$$

40

【 0 0 8 7 】

と呼ぶことにし、それは、第8次高調波が配置されていそうである周波数バンドを示す。計算機のカットオフ周波数計算機203は、1.2kHzと、第8次高調波が配置されていそうである周波数バンドの最後の周波数

【 0 0 8 8 】

【数 2 1】

$$\left(L_f(i_{g^{th}}) \right)$$

【0 0 8 9】

との間のより高いほうの周波数として、次の関係を使用して、最終的なカットオフ周波数 f_{tc} を計算する。

【0 0 9 0】

10

【数 2 2】

$$f_{tc} = \max \left(L_f(i_{g^{th}}), 1.2 \text{ kHz} \right) \quad (19)$$

【0 0 9 1】

b) 正規化およびノイズ充填

変換器/修正器107は、カットオフ周波数 f_c より高い周波数バンドの周波数ピンをゼロに設定するゼロアー (zeroer) 204をさらに含む。

20

【0 0 9 2】

受動的なフレームおよび能動的な無声フレームについて、変換器/修正器107の正規化器 (normalizer) 205は、[0, 4] の間において、時間領域CELP励振 f_e の周波数表現の周波数バンドの f_c より低い周波数ピンを、次の関係を使用して正規化する。

【0 0 9 3】

【数 2 3】

$$f_{eN}(j) = \begin{cases} \frac{4 \cdot f_e(j)}{\max_{0 \leq i < f_c} (|f_e(i)|)}, & 0 \leq j < f_c \quad \text{の場合} \\ 0, & f_c \leq j < 256 \quad \text{の場合} \end{cases} \quad (20)$$

30

【0 0 9 4】

その場合、変換器/修正器107は、ランダムノイズを発生するランダムノイズ発生器206を含み、そして周波数ピンのすべてにわたり一定レベルでノイズを加えるために、簡単なノイズ充填が加算器207によって実施される。このノイズ添加を記述する関数は、以下のよう定義される。

【0 0 9 5】

【数 2 4】

40

$j=0$ から $L-1$ に対して

$$f'_{eN}(j) = f_{eN}(j) + 0.75 \cdot \text{rand}(\quad) \quad (21)$$

【0 0 9 6】

ただし、 r_{and} は、-1と+1の間に制限される乱数の発生器である。

【0 0 9 7】

c) 修正された励振スペクトルのバンド当たりのエネルギー分析

利得計算機208-209-210のサブ計算機208は、上記の章3で述べたものと同じ方法を使用

50

して、スペクトル再成形 E_b' 後のバンド当たりのエネルギーを決定する。

【 0 0 9 8 】

d) エネルギー整合

受動的なフレームおよび能動的な無声フレームについて、エネルギー整合は、励振スペクトル修正後にバンド当たりのエネルギーをその初期値に調節することだけから成り立つ。各バンド i について、利得計算機208-209-210のサブ計算機210は、エネルギーを整合させるために、周波数バンド中のすべてのピンに適用される整合利得 G_b を次のように決定する。

【 0 0 9 9 】

【 数 2 5 】

10

$$G_b(i) = \frac{E_b(i)}{E_b'(i)} \quad (22)$$

【 0 1 0 0 】

ただし、 $E_b(i)$ は、上記の章3の方法を使用してサブ計算機209で決定される、励振スペクトル修正前のバンド当たりのエネルギーであり、 $E_b'(i)$ は、サブ計算機208で計算される、励振スペクトル修正後のバンド当たりのエネルギーである。特定のバンド i について、サブ計算機210で決定される、修正された(非正規化の)周波数領域励振

20

【 0 1 0 1 】

【 数 2 6 】

$$f_{edN}'$$

【 0 1 0 2 】

は、次のように書くことができる。

30

【 0 1 0 3 】

【 数 2 7 】

$$C_{Bb}(i) \leq j < C_{Bb}(i) + B_b(i) \text{ に対して} \\ f_{edN}'(j) = G_b(i) \cdot f_{eN}'(j) \quad (23)$$

【 0 1 0 4 】

ただし、 C_{Bb} および B_b は、上記の章3で定義される。

【 0 1 0 5 】

40

5) 一般のオーディオフレームのための励振修正

a) 正規化およびノイズ充填

ここで、図3を参照することにする。分類器の一部分301によって決定される一般のオーディオフレームに関して、正規化は、わずかに異なり、正規化器302によって行われる。まず、正規化ファクター N_f は、バンドごとに変化し、低周波数バンドにはより高い値を使用し、高周波数バンドにはより低い値を使用する。その考えは、パルスの位置がより正確である低い周波数バンド中に、より大きい振幅をもたらし、パルスの位置がそれほど正確でないより高い周波数バンド中に、より小さい振幅をもたらすということである。この例示的な例では、周波数バンドによって変化する正規化ファクター N_f は、次のように定義される。

50

$N_f = \{16, 16, 16, 16, 16, 16, 16, 12, 12, 12, 12, 8, 8, 8, 8, 8, 4, 4, 2, 2, 1, 1, 1\}$

【 0 1 0 6 】

特定の周波数バンド*i*に関して、一般のオーディオフレームの時間領域励振の周波数表現(周波数領域励振) f_e の正規化は、次のように記述することができる。

【 0 1 0 7 】

【数 2 8】

$$f_{eN}(j) = \frac{N_f(i) \cdot f_e(j)}{\max_{k=C_{Bb}(i)}^{|f_e(k)|}} , \quad C_{Bb}(i) \leq j < C_{Bb}(i) + B_b(i) \text{ の場合} \quad (24) \quad 10$$

【 0 1 0 8 】

ただし、 B_b は、周波数バンド当たりのピン数であり、バンド当たりの累積周波数ピンは、 C_{Bb} であり、 $f_{eN}(j)$ は、正規化された周波数領域励振である。 B_b および C_{Bb} は、上記の章3で記述している。

【 0 1 0 9 】

さらにまた、正規化器302は、

【 0 1 1 0 】

【数 2 9】

$$f'_{eN}(j):$$

【 0 1 1 1 】

【数 3 0】

$$f'_{eN}(j) = \begin{cases} 0 & (f_{eN}(j) < Z_f(i)) \text{ の場合} \\ f_{eN}(j) & \text{そうでなければ} \end{cases} \quad C_{Bb}(i) \leq j < C_{Bb}(i) + B_b(i) \text{ の場合} \quad (25) \quad 30$$

【 0 1 1 2 】

を得るために、各周波数バンド中の $f_{eN}(j)$ の最大値の分数 Z_f より低いすべての周波数ピンをゼロに設定するゼロアー(zeroer)(図示せず)を含む。

【 0 1 1 3 】

ただし、 Z_f は、次のように表すことができる。

$Z_f = \{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0.5, 0.5, 0.5\}$ 40

【 0 1 1 4 】

より挑戦的なゼロ設定は、スペクトルのピーキーな性質(peakiness: とがった性質)を増加させることが望まれる場合、ベクトル Z_f の値を増加することによって実施することができる。

【 0 1 1 5 】

b) 修正された励振スペクトルのバンド当たりのエネルギー分析

利得計算機303-304-306の計算機の一部303は、上記の章3で記述したものと同一方法を使用して、スペクトル再成形後のバンド当たりのエネルギー E_b' を決定する。

【 0 1 1 6 】

c) エネルギー整合

図3は、利得計算機303-304-306を示し、図4は、この利得計算機の計算機の一部306をより詳細に記述する。

【 0 1 1 7 】

一般のオーディオフレームに関して、エネルギー整合は、それがスペクトルのダイナミックも増加させることを目指すので、より難しい。各周波数バンド*i*について、利得計算機303-304-306の計算機の一部306のサブ計算機413は、方程式(22)中のものと同様に定義される推定される利得 G_e を計算する。

【 0 1 1 8 】

【数 3 1】

10

$$G_e(i) = \frac{E_b(i)}{E_b'(i)} \quad (26)$$

【 0 1 1 9 】

ただし、 $E_b(i)$ は、上記の章3で述べた方法を使用して、計算機の一部304で決定される励振スペクトル修正前のバンド当たりのエネルギーであり、 $E_b'(i)$ は、計算機の一部303で計算される、励振スペクトル修正後のバンド当たりのエネルギーである。

【 0 1 2 0 】

20

計算機の一部306のサブ計算機414は、修正された(非正規化された)周波数領域励振

【 0 1 2 1 】

【数 3 2】

$$f_{edN}$$

【 0 1 2 2 】

を形成するために、正規化器302およびスペクトルスプリッタ401-420からの正規化された周波数領域励振

30

【 0 1 2 3 】

【数 3 3】

$$f_{eN}$$

【 0 1 2 4 】

の最初の400Hz(または最初の4つのバンド)に、次の関係を使用して、利得 G_e を適用する。

40

【 0 1 2 5 】

【数 3 4】

$$f'_{edN}(j) = G_e(i) \cdot f_{eN}(j),$$

$$C_{Bb}(i) \leq j < C_{Bb}(i) + B_b(i) |_{0 \leq i < 4} \text{ の場合} \quad (27)$$

【 0 1 2 6 】

発見器404は、400Hzより高いバンド*i*ごとに最大値 $\max_a \quad j < b (|f_{eN}(j)|)$ を決定し、ただ

50

し、 $a=C_{Bb}(i)$ および $b=C_{Bb}(i)+B_b(i)$ は、上記の章3で定義されている。

【 0 1 2 7 】

正規化された周波数領域励振の400Hzと2kHzの間(バンド4~12)に含まれる周波数バンドに関して(モジュール420および450参照)、周波数ビン中の正規化された周波数領域励振が、

【 0 1 2 8 】

【 数 3 5 】

$$f'_{eN}(j) \geq 0.86 \max_{a \leq j < b} (|f_{eN}(j)|)$$

10

【 0 1 2 9 】

である場合(モジュール451参照)、増幅器402は、方程式(28)の上の行に示すように、ファクター1.1だけ、サブ計算機413からの利得 G_0 を増幅する。サブ計算機403は、修正された(非正規化された)周波数領域励振

【 0 1 3 0 】

【 数 3 6 】

$$f'_{edN}$$

20

【 0 1 3 1 】

を得るために、方程式(28)の最初の行に従って、周波数ビン中の正規化されたスペクトルの励振

【 0 1 3 2 】

【 数 3 7 】

$$f'_{eN}$$

30

【 0 1 3 3 】

に、増幅器402からの増幅された利得を適用する。

【 0 1 3 4 】

さらに、正規化された周波数領域励振の400Hzと2kHzの間(バンド4~12)に含まれる周波数バンドに関して(モジュール420および450参照)、周波数ビン中の正規化された周波数領域励振が、

【 0 1 3 5 】

【 数 3 8 】

40

$$f'_{eN}(j) < 0.86 \max_{a \leq j < b} (|f_{eN}(j)|)$$

【 0 1 3 6 】

である場合(モジュール451参照)、減衰器405は、方程式(28)の下に行に示すように、ファクター0.86だけ、サブ計算機413からの利得 G_0 を減衰させる。サブ計算機406は、修正された(非正規化された)周波数領域励振

50

【 0 1 3 7 】

【 数 3 9 】

$$f'_{edN}$$

【 0 1 3 8 】

を得るために、方程式(28)の下に行に従って、周波数ビン中の正規化されたスペクトルの
励振

10

【 0 1 3 9 】

【 数 4 0 】

$$f'_{eN}$$

【 0 1 4 0 】

に、減衰器405からの減衰された利得を適用する。

【 0 1 4 1 】

20

要約すると、修正された(非正規化された)スペクトルの励振

【 0 1 4 2 】

【 数 4 1 】

$$f'_{edN}$$

【 0 1 4 3 】

は、次のように与えられる。

30

【 0 1 4 4 】

【 数 4 2 】

$$f'_{edN}(j) = \begin{cases} 1.1 \cdot G_e(i) \cdot f'_{eN}(j), & f'_{eN}(j) \geq 0.86 \cdot \max_{a \leq j < b} (|f_{eN}(j)|) \text{ の場合} \\ 0.86 \cdot G_e(i) \cdot f'_{eN}(j), & f'_{eN}(j) < 0.86 \cdot \max_{a \leq j < b} (|f_{eN}(j)|) \text{ の場合} \end{cases} \quad (28)$$

【 0 1 4 5 】

最後に、スペクトルのより高い部分、この実施例では正規化された周波数領域励振の2k
Hzより高い周波数バンド(バンド>12)に関して(モジュール420および450参照)、周波数ビ
ン中の正規化された周波数領域励振が、

40

【 0 1 4 6 】

【 数 4 3 】

$$f'_{eN}(j) \geq 0.86 \max_{a \leq j < b} (|f_{eN}(j)|)$$

【 0 1 4 7 】

50

である場合(モジュール452参照)、周波数バンド*i*の関数であり、またビットレートの関数とすることができる傾斜は、LPCフィルタの低すぎるエネルギー推定を補償するために、利得*G_e*に加えられる。周波数バンド当たりの傾斜の値 (i)は、次のように公式化される。

【 0 1 4 8 】
【 数 4 4 】

$$\delta(i) = 1.5 \cdot G_e(i) \cdot \frac{(j-12)}{32} \quad (29)$$

10

【 0 1 4 9 】

傾斜は、傾斜計算機407-408によって計算され、修正された(非正規化された)周波数領域励振

【 0 1 5 0 】
【 数 4 5 】

$$f'_{eN}$$

20

【 0 1 5 1 】

を得るために、サブ計算機409により、方程式(30)の上の行に従って、周波数ピンによる正規化された周波数領域励振

【 0 1 5 2 】
【 数 4 6 】

$$f'_{eN}$$

30

【 0 1 5 3 】

に適用される。

【 0 1 5 4 】

さらに、スペクトルのより高い部分、この例示的な例では正規化された周波数領域励振の2kHzより高い周波数バンド(バンド>12)に関して(モジュール420および450参照)、周波数ピン中の正規化された周波数領域励振が、

【 0 1 5 5 】
【 数 4 7 】

40

$$f'_{eN}(j) < 0.86 \max_{a \leq j < b} (|f_{eN}(j)|)$$

【 0 1 5 6 】

である場合(モジュール452参照)、減衰器410は、修正された(非正規化された)周波数領域励振

【 0 1 5 7 】

【数 4 8】

$$f'_{edN}$$

【0 1 5 8】

を得るために、サブ計算機411により、方程式(30)の下に行に従って、周波数ビンによって正規化されたスペクトルの励振

【0 1 5 9】

10

【数 4 9】

$$f'_{eN}$$

【0 1 6 0】

に適用される減衰利得

【0 1 6 1】

【数 5 0】

20

$$[f'_{eN}(j)/\max_{a \leq j < b}(|f_{eN}(j)|)]^2$$

【0 1 6 2】

を計算する。

【0 1 6 3】

要約すると、非正規化されたスペクトルの励振

【0 1 6 4】

30

【数 5 1】

$$f'_{edN}$$

【0 1 6 5】

が、次のように決定される。

【0 1 6 6】

【数 5 2】

40

$$f'_{edN}(j) = \begin{cases} \delta(i) \cdot f'_{eN}(j), & f'_{eN}(j) \geq 0.86 \cdot \max_{a \leq j < b}(|f_{eN}(j)|) \text{ の場合} \\ \left(\frac{f'_{eN}(j)}{\max_{a \leq j < b}(|f_{eN}(j)|)} \right)^2 \cdot f'_{eN}(j), & \text{そうでなければ} \end{cases} \quad (30)$$

【0 1 6 7】

ただし、aおよびbは、本明細書で上記に述べている。最新のバンドに適用される利得を

50

さらに増加することも可能であり、その場合、LPCのエネルギー整合は最も悪くなる。

【 0 1 6 8 】

6) 逆周波数変換

結合器453は、全部そろった修正された(非正規化された)周波数領域励振

【 0 1 6 9 】

【 数 5 3 】

$$f'_{edN}$$

10

【 0 1 7 0 】

を形成するために、サブ計算機414、403、406、409および411からの修正された(非正規化された)周波数領域励振

【 0 1 7 1 】

【 数 5 4 】

$$f'_{edN}$$

20

【 0 1 7 2 】

に寄与度を結合する。

【 0 1 7 3 】

周波数領域処理が完了した後、逆周波数-時間変換202が、時間領域の修正された励振を見出すために、結合器453からの修正された(非正規化された)周波数領域励振

【 0 1 7 4 】

【 数 5 5 】

$$f'_{edN}$$

30

【 0 1 7 5 】

に適用される。この例示的な実施形態では、周波数-時間変換が、25Hzの分解能を与える時間-周波数変換のために使用されるのと同じタイプII DCTの逆を用いて達成される。さらに、あらゆる他の変換を使用することができる。修正された時間領域励振

【 0 1 7 6 】

【 数 5 6 】

$$e'_{td}$$

40

【 0 1 7 7 】

が、以下のように得られる。

【 0 1 7 8 】

【数 5 7】

$$e'_{id}(k) = \begin{cases} \sqrt{\frac{1}{L}} \cdot \sum_{n=0}^{L-1} f'_{edN}(n), & k = 0 \\ \sqrt{\frac{2}{L}} \cdot \sum_{n=0}^{L-1} f'_{edN}(n) \cdot \cos\left(\frac{\pi}{L}\left(n + \frac{1}{2}\right)k\right), & 1 \leq k \leq L-1 \end{cases} \quad (31)$$

【 0 1 7 9 】

ただし、

【 0 1 8 0 】

【数 5 8】

10

$$f'_{edN}(n)$$

【 0 1 8 1 】

は、修正された励振の周波数表現であり、Lは、フレーム長である。この例示的な例では、フレーム長は、12.8kHzの対応する内部サンプリング周波数に対して256サンプルである(AMR-WB)。

20

【 0 1 8 2 】

7) 合成フィルタリングおよび現在のCELP合成物の上書き

一度励振修正が完了されると、修正された励振は、現在のフレームについて修正された合成物を得るために、合成フィルタ108によって処理される。上書き器110は、この修正された合成物を使用して、復号化された合成物の上に上書きし、それによって知覚品質が高められる。

【 0 1 8 3 】

次いで、16kHzまでの最後のディエンファシスおよび再サンプリングをディエンファシスフィルタおよび再サンプラ112で実施することができる。

30

【符号の説明】

【 0 1 8 4 】

- 101 AMR-WBビットストリーム
- 102 時間領域励振デコーダ
- 103 LP合成フィルタ
- 104、105、106、301 分類器の一部
- 107 変換器/修正器
- 108 LP合成フィルタ
- 109 音声区間検出(VAD)情報
- 111 LPフィルタスムーザ
- 112 ディエンファサイジングフィルタおよび再サンプラ
- 201 時間-周波数領域変換器
- 203 カットオフ周波数計算機
- 204 ゼロアー
- 205 正規化器
- 206 ランダムノイズ発生器
- 207 加算器
- 208、209、210 利得計算機のサブ計算機
- 302 正規化器

40

50

303、304、306 利得計算機の一部の計算機

401 スペクトルスプリッタ

402 増幅器

403 サブ計算機

404 発見器

405 減衰器

406 サブ計算機

407 傾斜計算機

408 傾斜計算機

409 サブ計算機

410 減衰器

411 サブ計算機

413 サブ計算機

414 サブ計算機

420 モジュール、スペクトルスプリッタ

450 モジュール

451 モジュール

452 モジュール

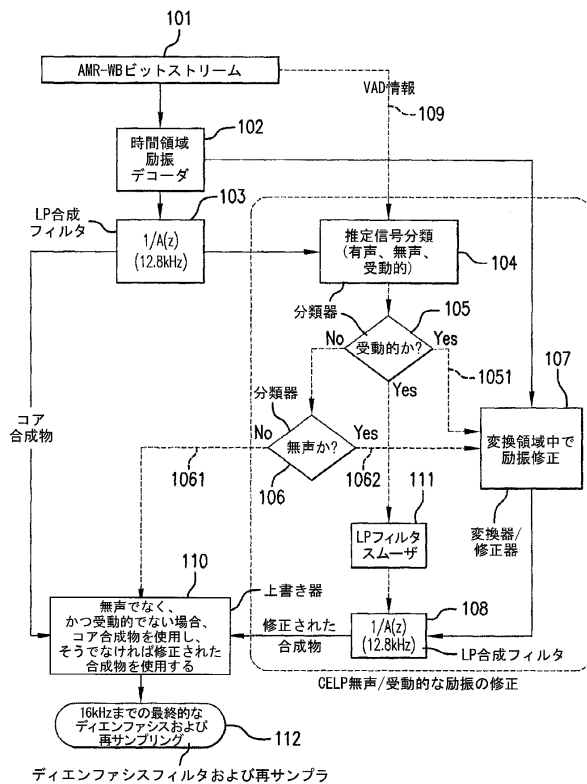
453 結合器

1021 音声/音響信号の時間領域合成物

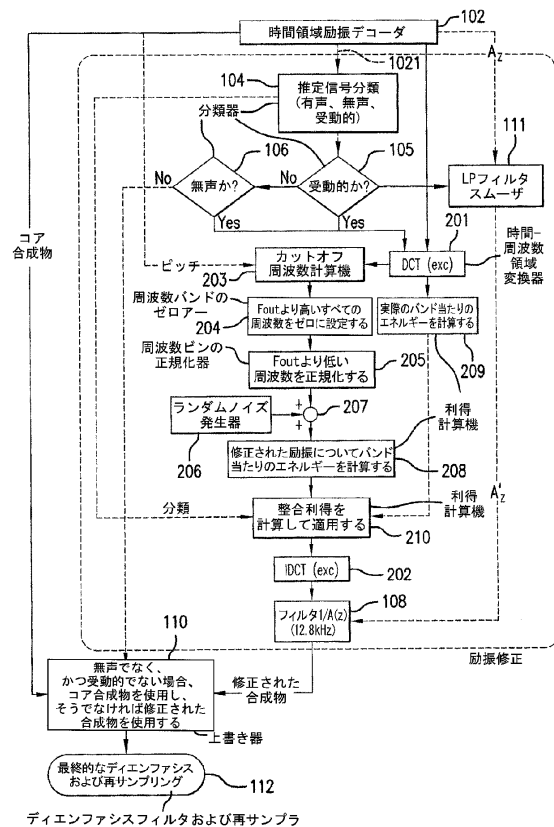
10

20

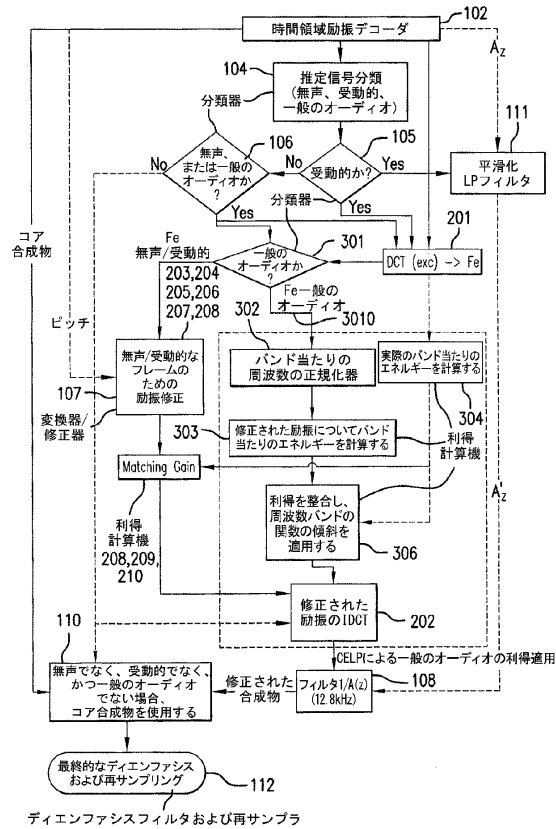
【図 1】



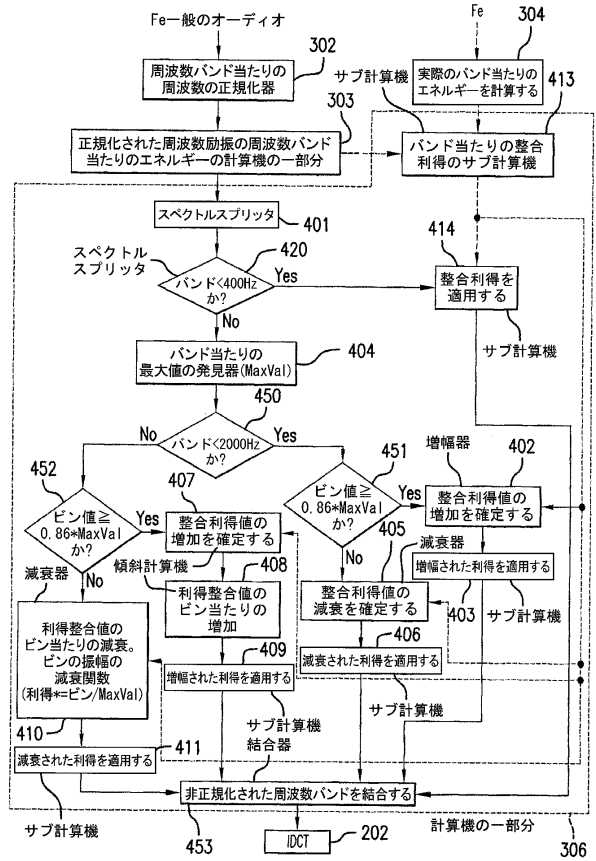
【図 2】



【 図 3 】



【 図 4 】



フロントページの続き

(72)発明者 トミー・ヴェヤンクール

カナダ・ケベック・J 1 N・2 K 1・シャープブルック・リュ・ファラン・8 3 5

(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・エミール・ネリガン・1 3 5 5

審査官 五貫 昭一

(56)参考文献 特表 2 0 0 9 - 5 0 8 1 4 6 (J P , A)

特表 2 0 0 9 - 5 2 4 1 0 1 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)

G 1 0 L 1 9 / 1 8

G 1 0 L 1 9 / 1 6

H 0 3 M 7 / 3 0