

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号  
特許第6086999号  
(P6086999)

(45) 発行日 平成29年3月1日 (2017.3.1)

(24) 登録日 平成29年2月10日 (2017.2.10)

(51) Int. Cl.	F I
G 1 O L 19/22 (2013.01)	G 1 O L 19/22
G 1 O L 19/12 (2013.01)	G 1 O L 19/12
G 1 O L 19/02 (2013.01)	G 1 O L 19/02 1 5 O
H O 3 M 7/30 (2006.01)	H O 3 M 7/30 Z
G 1 O L 19/107 (2013.01)	G 1 O L 19/107

請求項の数 15 (全 30 頁)

(21) 出願番号	特願2015-563151 (P2015-563151)	(73) 特許権者	500341779
(86) (22) 出願日	平成27年7月21日 (2015.7.21)		フラウンホーファー・ゲゼルシャフト・ツ
(65) 公表番号	特表2016-535286 (P2016-535286A)		ール・フェルデルング・デル・アンゲヴァ
(43) 公表日	平成28年11月10日 (2016.11.10)		ンテン・フォルシュング・アインゲトラ
(86) 国際出願番号	PCT/EP2015/066677		ゲネル・フェライン
(87) 国際公開番号	W02016/016053		ドイツ連邦共和国, 8 0 6 8 6 ミュンヘ
(87) 国際公開日	平成28年2月4日 (2016.2.4)		ン, ハンザシュトラッセ 2 7 ツェー
審査請求日	平成27年11月12日 (2015.11.12)	(74) 代理人	100085497
(31) 優先権主張番号	14178809.1		弁理士 筒井 秀隆
(32) 優先日	平成26年7月28日 (2014.7.28)	(72) 発明者	ラベリ, エマニュエル
(33) 優先権主張国	欧州特許庁 (EP)		ドイツ連邦共和国 9 1 0 5 8 エルラン
早期審査対象出願		(72) 発明者	ゲン ブランデルヴェーク 7
			ムルトラス, マルクス
			ドイツ連邦共和国 9 0 4 6 9 ニュルン
			ベルク エツラウプヴェーク 7
			最終頁に続く

(54) 【発明の名称】 ハーモニクス低減を使用して第1符号化アルゴリズムと第2符号化アルゴリズムの一方を選択する装置及び方法

(57) 【特許請求の範囲】

【請求項 1】

オーディオ信号 ( 4 0 ) の一部分を符号化するために第 1 の特性を有する第 1 符号化アルゴリズムと第 2 の特性を有する第 2 符号化アルゴリズムの一方を選択して、前記オーディオ信号 ( 4 0 ) の前記一部分の符号化済みバージョンを得る装置 ( 1 0 ) であって、前記オーディオ信号を受信し、前記オーディオ信号内のハーモニクスの振幅を低減し、かつ前記オーディオ信号のフィルタ処理済みバージョンを出力する長期予測フィルタと、前記オーディオ信号のフィルタ処理済みバージョンを使用して、前記オーディオ信号の前記一部分の第 1 品質尺度として前記オーディオ信号の前記一部分の S N R ( 信号対ノイズ比 ) 又はセグメント S N R を推定する第 1 推定部 ( 1 2 ) であって、前記第 1 品質尺度は前記第 1 符号化アルゴリズムと関連しており、前記第 1 品質尺度を推定することは、前記第 1 符号化アルゴリズムの近似を実行して前記第 1 符号化アルゴリズムの歪み推定を取得し、かつ前記第 1 符号化アルゴリズムを使用した前記オーディオ信号の前記一部分の実際の符号化及び復号化を行わずに、前記オーディオ信号の前記一部分と前記第 1 符号化アルゴリズムの歪み推定とに基づいて第 1 品質尺度を推定することを含む、第 1 推定部 ( 1 2 ) と、前記オーディオ信号の前記一部分についての第 2 品質尺度として S N R 又はセグメント S N R を推定する第 2 推定部 ( 1 4 ) であって、前記第 2 品質尺度は前記第 2 符号化アルゴリズムと関連しており、前記第 2 品質尺度を推定することは、前記第 2 符号化アルゴリズムの近似を実行して、第 2 符号化アルゴリズムの歪み推定を取得し、かつ前記第 2 符号化

アルゴリズムを使用した前記オーディオ信号の前記一部分の実際の符号化及び復号化を行わずに、前記オーディオ信号の前記一部分と前記第2符号化アルゴリズムの歪み推定とを使用して前記第2品質尺度を推定することを含む、第2推定部(14)と、前記第1品質尺と前記第2品質尺度との間の比較に基づいて、前記第1符号化アルゴリズム又は前記第2符号化アルゴリズムを選択する制御部(16)と、を含み、前記第1符号化アルゴリズムは、変換符号化アルゴリズム、MDC T(修正離散コサイン変換)ベースの符号化アルゴリズム、又はTCX(変換符号化励振)符号化アルゴリズムのいずれかであり、前記第2符号化アルゴリズムは、CELP(符号励振線形予測)符号化アルゴリズム又はACELP(代数符号励振線形予測)符号化アルゴリズムである、装置。

10

#### 【請求項2】

請求項1に記載の装置(10)であって、前記長期予測フィルタの伝達関数は、ピッチラグの整数部分と、前記ピッチラグの小数部に依存したマルチタップフィルタの伝達関数とを含む、装置。

#### 【請求項3】

請求項1に記載の装置(10)であって、前記長期予測フィルタは、次式のような伝達関数を有し、

$$P(z) = 1 - \beta g B(z, T_{fr}) z^{-T_{int}}$$

20

ここで、 $T_{int}$ 及び $T_{fr}$ はピッチラグの整数部及び小数部であり、 $g$ はゲインであり、 $B(z, T_{fr})$ は前記ピッチラグの小数部に依存する係数を持つFIRローパスフィルタの伝達関数である、装置。

#### 【請求項4】

請求項1乃至3のいずれかに記載の装置であって、1つ又は複数のハーモニシティ尺度及び/又は1つ又は複数の時間的構造尺度を閾値と比較し、その比較結果に基づいて前記長期予測フィルタを無効にする無効化ユニットをさらに備える、装置。

#### 【請求項5】

請求項4に記載の装置であって、前記1つ又は複数のハーモニシティ尺度は正規化相関又は予測ゲインの少なくとも1つを含み、前記1つ又は複数の時間的構造尺度は時間的平坦度及びエネルギー変化の少なくとも1つを含む、装置。

30

#### 【請求項6】

請求項1乃至5のいずれかに記載の装置であって、前記長期予測フィルタは1フレーム毎に前記オーディオ信号に適用され、前記装置は、前記長期予測フィルタに起因して前記オーディオ信号内に生じる不連続部を除去するユニットをさらに含む、装置。

#### 【請求項7】

請求項1乃至6のいずれかに記載の装置(10)であって、前記第1推定部及び第2推定部は、前記オーディオ信号の重み付きバージョンの一部のSNR又はセグメントSNRを推定するよう構成された、装置。

40

#### 【請求項8】

請求項1乃至7のいずれかに記載の装置(10)であって、前記第1推定部(12)は、前記オーディオ信号の前記一部分を量子化する際に前記第1符号化アルゴリズムで使用される量子化器が導入するであろう、推定量子化歪みを決定するよう構成され、前記オーディオ信号の重み付きバージョンの一部のエネルギーと前記推定量子化歪みとに基づいて、前記第1品質尺度を推定するよう構成され、前記第1推定部(12)は、前記第1符号化アルゴリズムにおいて使用された量子化器とエントロピー符号器とを用いて符号化された場合に、前記オーディオ信号の前記一部分が所定の目標ピッ

50

トレートを生成するよう、前記オーディオ信号の前記一部分についてのグローバルゲインを推定するよう構成され、前記第 1 推定部 ( 1 2 ) はさらに前記推定されたグローバルゲインに基づいて前記推定量子化歪みを決定するよう構成された、装置。

【請求項 9】

請求項 1 乃至 8 のいずれかに記載の装置 ( 1 0 ) であって、

前記第 2 推定部 ( 1 4 ) は、前記オーディオ信号の前記一部分を符号化するために適応型コードブックを使用した際に前記第 2 符号化アルゴリズムで使用される前記適応型コードブックが導入するであろう、推定適応型コードブック歪みを決定するよう構成され、前記第 2 推定部 ( 1 4 ) はまた、前記オーディオ信号の重み付きバージョンの一部のエネルギーと前記推定適応型コードブック歪みとに基づいて、前記第 2 品質尺度を推定するよう構成され、前記オーディオ信号の前記一部分の複数のサブ部分のそれぞれについて、前記第 2 推定部 ( 1 4 ) は更に、前処理段階で決定されたピッチラグ分だけ過去へとシフトされた前記重み付きオーディオ信号のサブ部分のバージョンに基づいて、前記適応型コードブックを近似するよう構成され、前記重み付きオーディオ信号のサブ部分と前記近似された適応型コードブックとの間の誤差が最小化されるように、適応型コードブックゲインを推定するよう構成され、かつ前記重み付きオーディオ信号の前記一部分のサブ部分と前記適応型コードブックゲインによってスケールされた前記近似された適応型コードブックとの間の誤差のエネルギーに基づいて、前記推定適応型コードブック歪みを決定するよう構成されている、装置。

10

【請求項 10】

20

請求項 9 に記載の装置 ( 1 0 ) であって、

前記第 2 推定部 ( 1 4 ) は、さらに前記オーディオ信号の前記一部分の各サブ部分について決定された前記推定適応型コードブック歪みを一定ファクタで低減するよう構成された、装置。

【請求項 11】

請求項 1 乃至 8 のいずれかに記載の装置であって、

前記第 2 推定部 ( 1 4 ) は、前記オーディオ信号の前記一部分を符号化するために適応型コードブックを使用した際に前記第 2 符号化アルゴリズムで使用される前記適応型コードブックが導入するであろう、推定適応型コードブック歪みを決定するよう構成され、前記第 2 推定部 ( 1 4 ) はまた、前記オーディオ信号の重み付きバージョンの一部のエネルギーと前記推定適応型コードブック歪みとに基づいて、前記第 2 品質尺度を推定するよう構成され、前記第 2 推定部 ( 1 4 ) は、前処理段階で決定されたピッチラグ分だけ過去へとシフトされた前記重み付きオーディオ信号の前記一部分のバージョンに基づいて、前記適応型コードブックを近似するよう構成され、かつ前記重み付きオーディオ信号の前記一部分と前記近似された適応型コードブックとの間の誤差が最小化されるように、適応型コードブックゲインを推定するよう構成され、前記重み付きオーディオ信号の前記一部分と前記適応型コードブックゲインによってスケールされた前記近似された適応型コードブックとの間の誤差のエネルギーに基づいて、前記推定適応型コードブック歪みを決定するよう構成されている、装置。

30

【請求項 12】

40

オーディオ信号の一部分を符号化する装置 ( 2 0 ) であって、請求項 1 乃至 11 のいずれかに記載の装置 ( 1 0 ) と、前記第 1 符号化アルゴリズムを実行する第 1 符号化部 ( 2 6 ) と、前記第 2 符号化アルゴリズムを実行する第 2 符号化部 ( 2 8 ) とを備え、前記符号化する装置 ( 2 0 ) は、前記制御部 ( 1 6 ) による選択に依存して、前記第 1 符号化アルゴリズム又は前記第 2 符号化アルゴリズムを使用して前記オーディオ信号の前記一部分を符号化するよう構成された、装置。

【請求項 13】

請求項 12 に記載の符号化する装置 ( 2 0 ) と、前記オーディオ信号の前記一部分の符号化済みバージョンと前記オーディオ信号の前記一部分を符号化するために使用されたアルゴリズムの指示とを受信し、かつ前記オーディオ信号の前記一部分の符号化済みバージョン

50

ンを前記指示されたアルゴリズムを使用して復号化するよう構成された復号器と、を含む、符号化及び復号化のシステム。

【請求項 14】

オーディオ信号(40)の一部分を符号化するために第1の特性を有する第1符号化アルゴリズムと第2の特性を有する第2符号化アルゴリズムの一方を選択して、前記オーディオ信号(40)の前記一部分の符号化済みバージョンを得る方法であって、

前記オーディオ信号内のハーモニクスの振幅を低減し、かつ前記オーディオ信号のフィルタ処理済みバージョンを出力するために、長期予測フィルタを使用して前記オーディオ信号をフィルタ処理するステップと、

前記オーディオ信号のフィルタ処理済みバージョンを使用して、前記オーディオ信号の前記一部分の第1品質尺度として前記オーディオ信号の前記一部分のSNR(信号対ノイズ比)又はセグメントSNRを推定するステップであって、前記第1品質尺度は前記第1符号化アルゴリズムと関連しており、前記第1品質尺度を推定することは、前記第1符号化アルゴリズムの近似を実行して前記第1符号化アルゴリズムの歪み推定を取得し、かつ前記第1符号化アルゴリズムを使用した前記オーディオ信号の前記一部分の実際の符号化及び復号化を行わずに、前記オーディオ信号の前記一部分と前記第1符号化アルゴリズムの歪み推定とに基づいて第1品質尺度を推定することを含む、ステップと、

前記オーディオ信号の前記一部分についての第2品質尺度としてSNR又はセグメントSNRを推定するステップであって、前記第2品質尺度は前記第2符号化アルゴリズムと関連しており、前記第2品質尺度を推定することは、前記第2符号化アルゴリズムの近似を実行して、第2符号化アルゴリズムの歪み推定を取得し、かつ前記第2符号化アルゴリズムを使用した前記オーディオ信号の前記一部分の実際の符号化及び復号化を行わずに、前記オーディオ信号の前記一部分と前記第2符号化アルゴリズムの歪み推定とを使用して前記第2品質尺度を推定することを含む、ステップと、

前記第1品質尺度と前記第2品質尺度との間の比較に基づいて、前記第1符号化アルゴリズム又は前記第2符号化アルゴリズムを選択するステップと、を含み、

前記第1符号化アルゴリズムは、変換符号化アルゴリズム、MDCT(修正離散コサイン変換)ベースの符号化アルゴリズム、又はTCX(変換符号化励振)符号化アルゴリズムのいずれかであり、前記第2符号化アルゴリズムは、CELP(符号励振線形予測)符号化アルゴリズム又はACELP(代数符号励振線形予測)符号化アルゴリズムである、方法。

【請求項 15】

コンピュータ上で実行されたとき、請求項14に記載の方法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。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、オーディオコーディング(符/復号化)に関し、特に切り替え式オーディオコーディングであって、オーディオ信号の異なる部分について異なる符号化アルゴリズムを使用して符号化済み信号が生成されるコーディングに関する。

【背景技術】

【0002】

オーディオ信号の異なる部分について異なる符号化アルゴリズムを決定する切り替え式オーディオコードが公知である。一般に、切り替え式オーディオコードは、2つの異なるモード、即ちACELP(代数符号励振線形予測)及びTCX(変換符号励振)のようなアルゴリズム間での切り替えを提供するものである。

【0003】

MP3USAC(MP3統合スピーチオーディオ符号化)のLPDモードは、2つの異なるモードであるACELPとTCXとに基づいている。ACELPはスピーチ状及び過渡状の信号に対して良好な品質を提供する。TCXは音楽状及びノイズ状の信号に対

して良好な品質を提供する。符号器は、どちらのモードを使用するかをフレーム毎の基準で決定する。符号器によるこの決定は、コーデック品質にとって決定的である。誤った決定が一つだけ存在する場合でも、特に低ビットレートにおいて、強いアーチファクトを発生させるおそれがある。

#### 【 0 0 0 4 】

どちらのモードを使用するかを決定する最も単純な手法は、閉ループのモード選択である。即ち、両方のモードの完全な符号化 / 復号化を実行し、次にオーディオ信号と符号化済み / 復号化済みオーディオ信号とに基づいて両方のモードについての選択基準（例えばセグメント S N R）を計算し、最後にその選択基準に基づいて 1 つのモードを選択する。この手法は、一般的には安定でかつロバストな決定をもたらす。しかしながら、この手法はまた、各フレームにおいて両方のモードが作動しなければならないので、かなりの量の演算量を必要とする。

10

#### 【 0 0 0 5 】

演算量を低減する代替的な手法は、開ループのモード選択である。開ループ選択とは、両方のモードの完全な符号化 / 復号化を実行することなく、その代わり、低い演算量で計算された選択基準を使用して 1 つのモードを選択することである。そのため、最悪の場合でも、演算量は、演算量が最小のモード（通常は T C X）から、選択基準を計算するために必要な演算量を差し引いた分だけ低減される。演算量の節約は通常は有意であり、そのため、コーデックの最悪の場合の演算量が制限されるときに、この手法は魅力的となる。

#### 【 0 0 0 6 】

20

（非特許文献 1 に定義された）A M R - W B + 標準は、8 0 m s のフレーム内において A C E L P / T C X 2 0 / T C X 4 0 / T C X 8 0 の全ての組合せの間で決定を行うために使用される、開ループのモード選択を含む。この点については、非特許文献 1 の第 5 . 2 . 4 章に説明されている。更に、非特許文献 2 の会議録や、この会議録の著者による特許文献 1 及び特許文献 2 にも説明されている。

#### 【 0 0 0 7 】

特許文献 1 は、長期予測パラメータの分析に基づく開ループのモード選択を開示している。特許文献 2 は、オーディオ信号の各々のセクション内のオーディオコンテンツのタイプを示す信号特性に基づく開ループのモード選択を開示しており、そのような選択が実行不可能な場合には、その選択は、それぞれ隣接するセクションについて行われた統計的な評価に基づいて実行される。

30

#### 【 0 0 0 8 】

A M R - W B + の開ループの選択は、2 つの主要ステップで説明可能である。第 1 の主要ステップでは、エネルギーレベルの標準偏差、低周波 / 高周波のエネルギー関係、全体エネルギー、I S P ( i m m i t t a n c e s p e c t r a l p a i r ) 距離、ピッチラグ ( p i t c h - l a g ) 及びゲイン、スペクトル傾きなど、オーディオ信号についての複数の特徴が計算される。これらの特徴は、次に、簡易な閾値ベースの分類を用いながら、A C E L P と T C X との間の選択を行うために使用される。第 1 の主要ステップにおいて T C X が選択された場合には、第 2 の主要ステップは、閉ループ方式で T C X 2 0 / T C X 4 0 / T C X 8 0 の可能な組合せの間で決定を行う。

40

#### 【 0 0 0 9 】

特許文献 3 は、過渡検出の結果とオーディオ信号の品質の結果とに基づいて、異なる特性を有する 2 つの符号化アルゴリズムの間の決定を行う手法を開示している。加えて、ヒステリシスを適用することも開示されており、そのヒステリシスは過去に実行された選択、即ちオーディオ信号の先行部分についての選択に依存している。

#### 【 0 0 1 0 】

非特許文献 2 では、A M R - W B + の閉ループと開ループのモード選択が比較されている。主観的なリスニングテストによれば、開ループのモード選択は、閉ループのモード選択と比べて有意に悪い性能を示す。しかしまた、開ループのモード選択が、最悪の場合の演算量を 4 0 % 低減させることも示されている。

50

## 【先行技術文献】

## 【特許文献】

## 【0011】

【特許文献1】米国特許第7,747,430号

【特許文献2】米国特許第7,739,120号

【特許文献3】WO2012/110448A1

【特許文献4】PCT/EP2014/051557

【特許文献5】米国特許第5,012,517号

【特許文献6】EP0732687A2

【特許文献7】US5999899A

【特許文献8】米国特許第7,353,168号

## 【非特許文献】

## 【0012】

【非特許文献1】International Standard 3GPP TS 26.290 V6.1.0 2004-12

【非特許文献2】“Low Complex Audio Encoding for Mobile, Multimedia, VTC 2006, Mäkinen et al.”

【非特許文献3】Rec. ITU-T G.718

## 【発明の概要】

## 【発明が解決しようとする課題】

## 【0013】

本発明の目的は、良好な性能と低減された演算量とで、第1符号化アルゴリズムと第2符号化アルゴリズムとの間の選択を可能にする、改善された手法を提供することである。

## 【課題を解決するための手段】

## 【0014】

この目的は、請求項1に記載の装置と、請求項14に記載の方法と、請求項15に記載のコンピュータプログラムによって達成される。

## 【0015】

本発明の実施形態は、オーディオ信号の一部を符号化してオーディオ信号の符号化済みバージョンを取得するために、第1特性を有する第1符号化アルゴリズムと第2特性を有する第2符号化アルゴリズムとの一方を選択する装置を提供し、この装置は、オーディオ信号を受信し、オーディオ信号内のハーモニクス（高調波）の振幅を低減し、かつオーディオ信号のフィルタ処理済みバージョンを出力するよう構成された長期予測フィルタと、

オーディオ信号の前記部分の第1品質尺度としてオーディオ信号の前記部分のSNR（信号対ノイズ比）又はセグメントSNRを推定する際に、オーディオ信号のフィルタ処理済みバージョンを使用する第1推定部であって、第1品質尺度は第1符号化アルゴリズムと関連しており、実際に第1符号化アルゴリズムを使用してオーディオ信号の前記部分の符号化及び復号化を行わない、第1推定部と、

オーディオ信号の前記部分についての第2品質尺度としてSNR又はセグメントSNRを推定する第2推定部であって、第2品質尺度は第2符号化アルゴリズムと関連しており、実際に第2符号化アルゴリズムを使用してオーディオ信号の前記部分の符号化及び復号化を行わない、第2推定部と、

第1品質尺度と第2品質尺度との間の比較に基づいて、第1符号化アルゴリズム又は第2符号化アルゴリズムを選択する制御部と、を含む。

## 【0016】

本発明の実施形態は、オーディオ信号の一部を符号化してオーディオ信号の符号化済みバージョンを取得するために、第1特性を有する第1符号化アルゴリズムと第2特性を有する第2符号化アルゴリズムとの一方を選択する方法を提供し、この方法は、オーディオ信号内のハーモニクスの振幅を低減し、かつオーディオ信号のフィルタ処理済みバージョンを出力するために、オーディオ信号をフィルタ処理するステップと、

オーディオ信号の前記部分についての第1品質尺度としてオーディオ信号の前記部分のSNR又はセグメントSNRを推定する際に、オーディオ信号のフィルタ処理済みバージョンを使用するステップであって、第1品質尺度は第1符号化アルゴリズムと関連しており、実際に第1符号化アルゴリズムを使用してオーディオ信号の前記部分の符号化及び復号化を行わない、ステップと、

オーディオ信号の前記部分についての第2品質尺度としてSNR又はセグメントSNRを推定するステップであって、第2品質尺度は第2符号化アルゴリズムと関連しており、実際に第2符号化アルゴリズムを使用してオーディオ信号の前記部分の符号化及び復号化を行わない、ステップと、

第1品質尺度と第2品質尺度との間の比較に基づいて、第1符号化アルゴリズム又は第2符号化アルゴリズムを選択するステップと、を含む。

10

#### 【0017】

本発明の実施形態は、第1符号化アルゴリズム及び第2符号化アルゴリズムの各々について品質尺度を推定し、第1品質尺度と第2品質尺度との間の比較に基づいて、前記符号化アルゴリズムの一方を選択することにより、改善された性能を有する閉ループの選択が可能になる、という知見に基づいている。品質尺度は推定されるのであり、即ち、品質尺度を得るためにオーディオ信号が実際に符号化及び復号化されることはない。従って、低減された演算量を用いて品質尺度が取得可能となる。次いで、モード選択は、閉ループのモード選択に匹敵するような推定された品質尺度を使用して、実行できる。更に、本発明は、第1品質尺度の推定が、オーディオ信号のフィルタ処理されていないバージョンと比較してハーモニクスが低減されているオーディオ信号の一部分のフィルタ処理済みバージョンを使用する場合には、改善されたモード選択が達成できる、という知見にも基づいている。

20

#### 【0018】

本発明の実施形態においては、閉ループのモード選択が行われ、ここでは、ACELP及びTCXのセグメントSNRが低い演算量で最初に推定される。次に、これら推定されたセグメントSNR値を使用して、閉ループのモード選択と同様にモード選択が実行される。

#### 【0019】

本発明の実施形態は、AMR-WB+の開ループのモード選択において用いられるような、伝統的な特徴+分類部の手法を使用する訳ではない。その代わり、本発明の実施形態は、各モードの品質尺度を推定し、最良の品質を提供するモードを選択しようと試みる。

30

#### 【0020】

本発明の実施形態について、添付図面を参照しながら、以下において更に詳細に説明する。

#### 【図面の簡単な説明】

#### 【0021】

【図1】第1符号化アルゴリズムと第2符号化アルゴリズムとの一方を選択する装置の一実施形態の概略図を示す。

【図2】オーディオ信号を符号化する装置の一実施形態の概略図を示す。

40

【図3】第1符号化アルゴリズムと第2符号化アルゴリズムとの一方を選択する装置の一実施形態の概略図を示す。

【図4a】SNRの可能な計算式を示す。

【図4b】セグメントSNRの可能な計算式を示す。

#### 【発明を実施するための形態】

#### 【0022】

以下の説明では、異なる図面における類似の構成要素/ステップには同じ参照符号を付している。図面においては、信号接続などのような特徴であって、本発明を理解する上で必須でないものについては、説明を省略する点に留意されたい。

#### 【0023】

50

図1は、オーディオ信号の一部分を符号化するための符号器として、TCXアルゴリズムなどの第1符号化アルゴリズムと、ACELPアルゴリズムなどの第2アルゴリズムとの一方を選択する装置10を示す。この装置10は、オーディオ信号の前記部分のSNR又はセグメントSNRを、その信号部分の第1品質尺度として推定する第1推定部12を含む。その第1品質尺度は、第1符号化アルゴリズムと関連している。装置10は、オーディオ信号を受信し、オーディオ信号内のハーモニクスの振幅を低減させ、オーディオ信号のフィルタ処理済みバージョンを出力するよう構成された、フィルタ2を含む。図1に示すように、フィルタ2は第1推定部12の内部に配置されてもよく、第1推定部12の外部に配置されてもよい。第1推定部12は、第1品質尺度を推定する際に、オーディオ信号のフィルタ処理済みバージョンを使用する。換言すれば、第1推定部12は第1品質尺度を推定するが、その第1品質尺度とは、第1符号化アルゴリズムを使用して符号化及び復号化された場合にオーディオ信号の前記部分が持つであろう品質尺度であり、実際に第1符号化アルゴリズムを使用してオーディオ信号の前記部分を符号化及び復号化することはない。装置10は、前記信号部分について第2品質尺度を推定するための第2推定部14を含む。第2品質尺度は第2符号化アルゴリズムと関連している。換言すれば、第2推定部14は第2品質尺度を推定するが、その第2品質尺度とは、第2符号化アルゴリズムを使用して符号化及び復号化された場合にオーディオ信号の前記部分が持つであろう品質尺度であり、実際に第2符号化アルゴリズムを使用してオーディオ信号の前記部分を符号化及び復号化することはない。更に、装置10は、第1品質尺度と第2品質尺度との間の比較に基づいて第1符号化アルゴリズム又は第2符号化アルゴリズムを選択する、制御部16を含む。その制御部は、選択された符号化アルゴリズムを示す出力18を含んでもよい。

#### 【0024】

以下の説明において、第1推定部は、オーディオ信号のフィルタ処理済みバージョンを使用する。即ち、たとえ明示的に示されなくても、ハーモニクスの振幅を低減するよう構成されたフィルタ2が提供され、かつそのフィルタが無効化されていない場合には、第1品質尺度を推定する際にオーディオ信号の前記部分のフィルタ処理済みバージョンが使用される。

#### 【0025】

一実施形態において、第1符号化アルゴリズムに関連する第1特性は、音楽状及びノイズ状の信号に対してより適しており、第2符号化アルゴリズムに関連する第2特性は、スピーチ状及び過渡状の信号に対してより適している。本発明の実施形態において、第1符号化アルゴリズムは、例えばMDC T（修正離散コサイン変換）符号化アルゴリズムなどの変換符号化アルゴリズムやTCX（変換符号励振）符号化アルゴリズムのような、オーディオ符/復号化アルゴリズムである。FFT変換、他のいずれかの変換、又はフィルタバンクに基づく他の変換符/復号化アルゴリズムもあり得る。本発明の実施形態において、第2符号化アルゴリズムは、CELP（符号励振線形予測）符/復号化アルゴリズムやACELP（代数符号励振線形予測）符/復号化アルゴリズムなどのような、スピーチ符号化アルゴリズムである。

#### 【0026】

実施形態の中で、品質尺度は知覚的な品質尺度を示す。第1符号化アルゴリズムの主観的な品質の推定値である1つの値と、第2符号化アルゴリズムの主観的な品質の推定値である1つの値とが計算されてもよい。これら2つの値の比較に基づいて、最良に推定された主観的な品質を与える符号化アルゴリズムが選択されてもよい。この点は、信号の種々の特徴を表す多数の特性が計算され、次に分類器を適用してどのアルゴリズムを選択するかが決定される、AMR-WB+標準とは相違している。

#### 【0027】

実施形態の中で、それぞれの品質尺度は、重み付きオーディオ信号、即ちオーディオ信号の重み付きバージョンの一部分に基づいて推定される。実施形態の中で、重み付きオーディオ信号は、重み付け関数によってフィルタ処理されたオーディオ信号として定義され得

10

20

30

40

50



る。その場合、重み付け関数は重み付き L P C フィルタ  $A(z/g)$  であり、ここで  $A(z)$  は L P C フィルタであり、 $g$  は例えば 0.68 等の 0 ~ 1 の間の重みである。このような方法で、知覚的品質の良好な尺度が得られることが分かってきた。L P C フィルタ  $A(z)$  及び重み付き L P C フィルタ  $A(z/g)$  は前処理段階で決定され、かつそれらは両方の符号化アルゴリズムにおいて使用される点に留意されたい。他の実施形態において、重み付け関数は線形フィルタ、F I R フィルタ又は線形予測フィルタであってもよい。

#### 【0028】

実施形態において、品質尺度は、重み付き信号ドメインにおけるセグメント S N R (信号対ノイズ比) である。重み付き信号ドメインにおけるセグメント S N R は、知覚的品質の良好な尺度を表し、従って、品質尺度として有益な方法で使用可能であることが分かってきた。これはまた、符号化パラメータを推定するために、A C E L P と T C X の両方の符号化アルゴリズムで使用される品質尺度である。

10

#### 【0029】

他の品質尺度は、重み付き信号ドメインにおける S N R でもよい。別の品質尺度は、セグメント S N R、つまり、非重み付き信号ドメインにおける、即ち (重み付き) L P C 係数によってフィルタ処理されていない、オーディオ信号の対応する部分の S N R であってもよい。

#### 【0030】

一般に、S N R はオリジナル及び処理済みのオーディオ信号 (スピーチ信号など) をサンプル毎に比較する。その目的は、入力波形を再生する波形コードの歪みを測定することである。S N R は図 4 a に示すように計算され得る。ここで、 $x(i)$  及び  $y(i)$  は、指標  $i$  を有するオリジナル及び処理済みのサンプルであり、 $N$  はサンプルの全体数である。セグメント S N R は、全体の信号について作用する代わりに、例えば 5 m s のように 1 ~ 10 m s などの短いセグメントの S N R 値の平均を計算する。S N R は図 4 b に示すように計算されてもよい。ここで、 $N$  と  $M$  とは、それぞれセグメント長とセグメントの数とを表す。

20

#### 【0031】

本発明の実施形態において、オーディオ信号の前記部分は、オーディオ信号を窓掛けして得られたオーディオ信号の 1 つのフレームを表し、オーディオ信号を窓掛けすることで得られた複数の連続的なフレームについて適切な符号化アルゴリズムの選択が行われる。以下の説明では、オーディオ信号に関連して、「部分」と「フレーム」という用語は交換可能に用いられている。実施形態において、各フレームは複数のサブフレームへと分割され、各サブフレームについて S N R を計算し、d B へと変換し、d B でサブフレーム S N R の平均値を計算することによって、各フレームについてセグメント S N R が推定される。

30

#### 【0032】

このように実施形態において、推定されるのは、入力オーディオ信号と復号化済みオーディオ信号との間の (セグメント) S N R ではなく、重み付き入力オーディオ信号と重み付き復号化済みオーディオ信号との間の (セグメント) S N R である。この (セグメント) S N R に関しては、A M R - W B + 標準 (非特許文献 1) の第 5.2.3 章を参照することができる。

40

#### 【0033】

本発明の実施形態においては、重み付きオーディオ信号の一部分のエネルギーと、それぞれのアルゴリズムによってその信号部分を符号化する際に導入される歪みの推定値とに基づいて、それぞれの品質尺度が推定されるが、第 1 推定部および第 2 推定部は、重み付きオーディオ信号のエネルギーに依存して、推定された歪みを決定するよう構成されている。

#### 【0034】

本発明の実施形態においては、オーディオ信号の前記部分を量子化する際に第 1 符号化アルゴリズムで使用される量子化器によって導入される、推定量子化歪みが決定され、第 1 品質尺度は、重み付きオーディオ信号の前記部分のエネルギーと推定量子化歪みとに基づ

50

いて決定される。そのような実施形態においては、オーディオ信号の前記部分が第1符号化アルゴリズムにおいて使用される量子化器とエントロピー符号器とによって符号化された場合に、所定の目標ビットレートを生成するように、オーディオ信号の前記部分についてグローバルゲインが推定されてもよく、その場合には、推定量子化歪みは、推定されたグローバルゲインに基づいて決定される。そのような実施形態において、推定量子化歪みは、推定されたゲインのパワーに基づいて決定されてもよい。第1符号化アルゴリズムで使用される量子化器が一様スカラー量子化器(uniform scalar quantizer)である場合、第1推定部は式 $D = G \cdot G / 12$ を用いて推定量子化歪みを決定してもよく、ここで、 $D$ は推定量子化歪みであり、 $G$ は推定されたグローバルゲインである。第1符号化アルゴリズムが別の量子化器を使用する場合には、量子化歪みは別の方法でグローバルゲインから決定されてもよい。

10

#### 【0035】

本発明の発明者らは、TCXなどの第1符号化アルゴリズムを用いてオーディオ信号の前記部分を符号化及び復号化する際に取得されるであろう、セグメントSNRのような品質尺度は、上述した特徴を任意の組合せにおいて使用することで、適切な方法で推定可能であることを認識している。

#### 【0036】

本発明の実施形態において、第1品質尺度はセグメントSNRであり、そのセグメントSNRは、オーディオ信号の前記部分の複数のサブ部分の各々に関連する推定されたSNRを、重み付きオーディオ信号の対応するサブ部分のエネルギーと推定された量子化歪みとに基づいて計算すること、及び、重み付きオーディオ信号の前記部分のサブ部分と関連する前記SNRの平均を計算することによって推定され、その結果、重み付きオーディオ信号の前記部分について推定されたセグメントSNRが得られる。

20

#### 【0037】

本発明の実施形態においては、オーディオ信号の前記部分を符号化するために適応型コードブックを使用した際に、第2符号化アルゴリズムで使用される適応型コードブックによって導入されるであろう、推定適応型コードブック歪みが決定され、第2品質尺度は、重み付きオーディオ信号の前記部分のエネルギーと推定適応型コードブック歪みとに基づいて推定される。

#### 【0038】

そのような実施形態においては、オーディオ信号の前記部分の複数のサブ部分のそれぞれについて、前処理段階で決定されたピッチラグ分だけ過去へとシフトされた重み付きオーディオ信号のサブ部分のバージョンに基づいて、適応型コードブックが近似されてもよく、また、重み付きオーディオ信号のサブ部分と近似された適応型コードブックとの誤差が最小化されるように、適応型コードブックゲインが推定されてもよく、更に、重み付きオーディオ信号の前記部分のサブ部分と適応型コードブックゲインによってスケールされた近似された適応型コードブックとの誤差のエネルギーに基づいて、推定適応型コードブック歪みが決定されてもよい。

30

#### 【0039】

本発明の実施形態において、オーディオ信号の前記部分の各サブ部分について決定された推定適応型コードブック歪みは、第2符号化アルゴリズムにおいて革新的コードブックによって達成される歪みの低減を考慮して、一定ファクタだけ低減されてもよい。

40

#### 【0040】

本発明の実施形態において、第2品質尺度はセグメントSNRであり、そのセグメントSNRは、重み付きオーディオ信号の対応するサブ部分のエネルギーと推定適応型コードブック歪みとに基づいて、各サブ部分に関連する推定されたSNRを計算し、かつ、前記サブ部分に関連するそれらSNRの平均を計算することで推定され、その結果、推定されたセグメントSNRが取得される。

#### 【0041】

本発明の実施形態においては、前処理段階で決定されたピッチラグ分だけ過去へとシフト

50

された重み付きオーディオ信号の前記部分のバージョンに基づいて、適応型コードブックが近似され、重み付きオーディオ信号の前記部分と近似された適応型コードブックとの誤差が最小化されるように、適応型コードブックゲインが推定され、更に、重み付きオーディオ信号の前記部分と適応型コードブックゲインによってスケールされた近似された適応型コードブックとの間のエネルギーに基づいて、推定適応型コードブック歪みが決定される。従って、推定適応型コードブック歪みは低い演算量で決定され得る。

【0042】

本発明者らは、ACELPなどの第2符号化アルゴリズムを用いてオーディオ信号の前記部分を符号化及び復号化する際に取得されるであろう、セグメントSNRのような品質尺度は、上述した特徴を任意の組合せで使用するにより、適切な方法で推定可能であることを認識している。

10

【0043】

本発明の実施形態においては、推定された品質尺度を比較する際に、ヒステリシス・メカニズムが使用される。これにより、どのアルゴリズムを使用すべきかを、より安定して決定することが可能になる。ヒステリシス・メカニズムは、推定された品質尺度（例えばそれらの間の差）及び他のパラメータ、例えば以前の決定に係る統計、時間的に定常なフレームの数、フレーム内の過渡などに依存し得る。そのようなヒステリシス・メカニズムに関して、例えば特許文献3を参照することができる。

【0044】

本発明の実施形態において、オーディオ信号を符号化する符号器は、装置(10)と、第1符号化アルゴリズムを実行するステージと、第2符号化アルゴリズムを実行するステージとを備え、符号器は、制御部16による選択に依存して、第1符号化アルゴリズム又は第2符号化アルゴリズムを使用してオーディオ信号の前記部分を符号化するように構成されている。本発明の実施形態において、符号化及び復号化のシステムは、前記符号器と、オーディオ信号の前記部分の符号化済みバージョンとオーディオ信号の前記部分を符号化するために使用されたアルゴリズムの指示とを受信し、かつオーディオ信号の前記部分の符号化済みバージョンを前記指示されたアルゴリズムを使用して復号化するように構成された復号器と、を含む。

20

【0045】

図1に示しかつ上述した(フィルタ2を除く)開ループのモード選択アルゴリズムは、先行出願(特許文献4)の中で開示されている。このアルゴリズムは、2つのモード、例えばACELPとTCXとの間の選択を、フレーム毎に行うために使用される。その選択は、ACELP及びTCXの両方のセグメントSNRの推定に基づいてもよい。最高の推定セグメントSNRを有するモードが選択される。任意ではあるが、更にロバストな選択を提供するために、ヒステリシス・メカニズムが使用されてもよい。ACELPのセグメントSNRは、適応型コードブック歪みの近似と、革新的コードブック歪みの近似とを使用して推定されてもよい。適応型コードブックは、ピッチ分析アルゴリズムによって推定されたピッチラグを使用して、重み付き信号ドメインにおいて近似されてもよい。その歪みは、最適なゲインを仮定して重み付き信号ドメインにおいて計算されてもよい。次に、その歪みは、革新的コードブック歪みを近似するように、一定ファクタだけ低減されてもよい。TCXのセグメントSNRは、現実のTCX符号器の簡易バージョンを使用して推定されてもよい。入力信号は、まずMDCTを用いて変換され、次に重み付きLPCフィルタを用いて整形されてもよい。最後に、その歪みが、グローバルゲイン及びグローバルゲイン推定部を使用して、重み付きMDCTドメインにおいて推定されてもよい。

30

40

【0046】

先行出願で開示されているこの開ループのモード選択アルゴリズムは、スピーチ状及び過渡状の信号に対してはACELPを選択し、音楽状及びノイズ状の信号に対してはTCXを選択して、期待される決定を殆どの場合に提供することが分かってきた。しかしながら、発明者らは、幾つかのハーモニックな楽音信号に対してACELPが時折選択される可能性があることを認識していた。そのような信号に対し、適応型コードブックは一般的に

50

高い予測ゲインを有する。その理由は、低い歪みを生成し、よってTCXよりも高いセグメントSNRを生成する、ハーモニックな信号の高い予測可能性に起因するものである。しかしながら、殆どのハーモニックな楽音信号に対してはTCXの方がより良好な音質を提供するので、このような場合にはTCXが推奨されるべきである。

【0047】

それ故、本発明は、入力信号のハーモニクスを低減するためにフィルタ処理された、入力信号のあるバージョンを使用して、第1品質尺度としてのSNR又はセグメントSNRの推定を実行することを提案する。これにより、ハーモニックな楽音信号に対する改善されたモード選択が達成できる。

【0048】

一般的に、ハーモニクスを低減するための如何なる適切なフィルタでも使用可能である。本発明の実施形態において、フィルタは長期予測フィルタである。長期予測フィルタの1つの簡単な例は、

$$F(z) = 1 - g \cdot z^{-T}$$

であり、ここで、フィルタパラメータはゲイン「g」とピッチラグ「T」であり、これらはオーディオ信号から決定される。

【0049】

本発明の実施形態は、TCXのセグメントSNR推定におけるMDC T分析の前にオーディオ信号に対して適用される、長期予測フィルタに基づいている。長期予測フィルタは、MDC T分析の前に、入力信号内のハーモニクスの振幅を低減させる。その結果、重み付きMDC Tドメイン内の歪みが低減され、TCXの推定されたセグメントSNRが増大し、最終的に、ハーモニックな楽音信号に対してTCXがより頻繁に選択されるようになる。

【0050】

本発明の実施形態において、長期予測フィルタの伝達関数は、ピッチラグの整数部と、ピッチラグの小数部に依存するマルチタップフィルタとを含む。これにより、整数部が通常のサンプリングレートの枠組み( $z^{-T_{int}}$ )の中だけで使用されるので、効率的な実装が可能になる。同時に、マルチタップフィルタにおける小数部の使用に起因して、高い精度も達成できる。マルチタップフィルタにおいて小数部を考慮することで、ハーモニクスのエネルギーの除去が達成できる一方で、ハーモニクスに近い部分のエネルギーの除去は防止できる。

【0051】

本発明の実施形態において、長期予測フィルタは以下の式で記述される。

$$P(z) = 1 - \beta g B(z, T_{fr}) z^{-T_{int}}$$

ここで、 $T_{int}$ と $T_{fr}$ とはピッチラグの整数部と小数部をそれぞれ示し、gはゲインであり、 $\beta$ は重みであり、 $B(z, T_{fr})$ はFIRローパスフィルタであって、その係数はピッチラグの小数部に依存している。そのような長期予測フィルタの実施例についての更なる詳細を、以下に説明する。

【0052】

ピッチラグとゲインはフレーム毎に推定されてもよい。

【0053】

予測フィルタは、1つ又は複数の(例えば正規化相関又は予測ゲインなどの)ハーモニシティ尺度、及び/又は、1つ又は複数の(例えば時間的平坦度又はエネルギー変化などの)時間的構造尺度に基づいて、無効化(ゲイン=0)されることができる。

【0054】

フィルタは入力オーディオ信号に対してフレーム毎に適用され得る。フィルタパラメータが1つのフレームから次のフレームへと変化する場合には、2つのフレーム間の境界に不連続部(discontinuity)が導入され得る。実施形態において、当該装置は、フィルタに

よって引き起こされたオーディオ信号内の不連続部を除去するためのユニットを更に含む。可能性のある不連続部を除去するために、如何なる技術、例えば特許文献 5、特許文献 6、特許文献 7 又は特許文献 8 に開示されている技術に匹敵する技術を用いてもよい。可能性のある不連続部を除去するための別の技術を、以下に説明する。

【 0 0 5 5 】

図 3 を参照しながら第 1 推定部 1 2 及び第 2 推定部 1 4 の実施例を詳細に説明する前に、図 2 を参照して符号器 2 0 の実施形態について説明する。

【 0 0 5 6 】

符号器 2 0 は、第 1 推定部 1 2、第 2 推定部 1 4、制御部 1 6、前処理ユニット 2 2、スイッチ 2 4、T C X アルゴリズムを実行するよう構成された第 1 符号器ステージ 2 6、A C E L P アルゴリズムを実行するよう構成された第 2 符号器ステージ 2 8 及び出力インターフェイス 3 0 を含む。前処理ユニット 2 2 は通常の U S A C 符号器の一部であってもよく、L P C 係数、重み付き L P C 係数、重み付きオーディオ信号及び 1 セットのピッチラグを出力するよう構成されてもよい。ここで、これら全てのパラメータは両方の符号化アルゴリズム、即ち T C X アルゴリズム及び A C E L P アルゴリズムにおいて使用されることに注意されたい。従って、そのようなパラメータは開ループのモード決定のために追加的に計算される必要がない。既に計算済みのパラメータを開ループのモード決定において使用する利点は、演算量を節約できることである。

【 0 0 5 7 】

図 2 に示すように、この装置はハーモニクス低減フィルタ 2 を含む。この装置は、任意ではあるが、一つ又は複数のハーモニシティ尺度（例えば正規化相関又は予測ゲインなど）及び / 又は一つ又は複数の時間的構造尺度（例えば時間的平坦度又はエネルギー変化）の組合せに基づいて、ハーモニクス低減フィルタ 2 を無効化させるための無効化ユニット 4 を更に含む。この装置は、任意ではあるが、オーディオ信号のフィルタ処理済みバージョンから不連続部を除去するための、不連続部除去ユニット 6 を更に含む。加えて、この装置は、任意ではあるが、ハーモニクス低減フィルタ 2 のフィルタパラメータを推定するためのユニット 8 を更に含む。図 2 において、これらの構成要素（2, 4, 6, 8）は第 1 推定部 1 2 の一部として示されている。指摘するまでもないが、これらの構成要素は、第 1 推定部の外部又は別の位置に構成され、かつ第 1 推定部に対してオーディオ信号のフィルタ処理済みバージョンを供給するよう構成されてもよい。

【 0 0 5 8 】

入力オーディオ信号 4 0 は入力ライン上に供給される。入力オーディオ信号 4 0 は第 1 推定部 1 2、前処理ユニット 2 2、及び両方の符号器ステージ 2 6, 2 8 に対して適用される。第 1 推定部 1 2 では、入力オーディオ信号 4 0 はフィルタ 2 に対して適用されて、入力オーディオ信号のフィルタ処理済みバージョンが第 1 品質尺度の推定に使用される。フィルタが無効化ユニット 4 によって無効化されている場合には、入力オーディオ信号のフィルタ処理済みバージョンに代えて、入力オーディオ信号 4 0 が第 1 品質尺度の推定に使用される。前処理ユニット 2 2 は、入力オーディオ信号を従来方式で処理して、L P C 係数及び重み付き L P C 係数 4 2 を導出し、オーディオ信号 4 0 をその重み付き L P C 係数 4 2 でフィルタ処理して、重み付きオーディオ信号 4 4 を得る。前処理ユニット 2 2 は、重み付き L P C 係数 4 2 と、重み付きオーディオ信号 4 4 と、1 セットのピッチラグ 4 8 とを出力する。当業者には明らかなように、重み付き L P C 係数 4 2 及び重み付きオーディオ信号 4 4 は、フレーム又はサブフレームへとセグメント化されてもよい。セグメンテーションは、オーディオ信号を適切な方法で窓掛けすることで取得することができる。

【 0 0 5 9 】

代替的实施形態において、重み付き L P C 係数及び重み付きオーディオ信号を、オーディオ信号のフィルタ処理済みバージョンに基づいて生成するよう構成された、前処理部が設けられてもよい。オーディオ信号のフィルタ処理済みバージョンに基づく重み付き L P C 係数及び重み付きオーディオ信号は、次に第 1 推定部に対して適用されて、重み付き L P C 係数 4 2 及び重み付きオーディオ信号 4 4 に代えて、第 1 品質尺度が推定される。

## 【 0 0 6 0 】

本発明の実施形態においては、量子化済み L P C 係数又は量子化済み重み付き L P C 係数が使用されてもよい。従って、用語「 L P C 係数」は「量子化済み L P C 係数」をも含むことを意図しており、用語「重み付き L P C 係数」は「重み付き量子化済み L P C 係数」をも含むことを意図している点を理解されたい。この点において、 U S A C の T C X アルゴリズムは、 M D C T スペクトルを整形するために、量子化済み重み付き L P C 係数を使用する点に注意する意義がある。

## 【 0 0 6 1 】

第 1 推定部 1 2 は、オーディオ信号 4 0、重み付き L P C 係数 4 2、及び重み付きオーディオ信号 4 4 を受信し、それらに基づいて第 1 品質尺度 4 6 を推定し、その第 1 品質尺度を制御部 1 6 へと出力する。第 2 推定部 1 4 は、重み付きオーディオ信号 4 4 とピッチラグ 4 8 のセットを受信し、それらに基づいて第 2 品質尺度 5 0 を推定し、その第 2 品質尺度 5 0 を制御部 1 6 へと出力する。当業者には公知のように、重み付き L P C 係数 4 2、重み付きオーディオ信号 4 4、及びピッチラグ 4 8 のセットは、先行するモジュール（即ち前処理ユニット 2 2）において既に計算されており、従って経費をかけずに使用可能である。

## 【 0 0 6 2 】

制御部は、受信された品質尺度の比較に基づいて、T C X アルゴリズム又は A C E L P アルゴリズムの一方を選択する決定を行う。上述したように、制御部は、どちらのアルゴリズムを使用すべきか決定する際に、ヒステリシス・メカニズムを用いてもよい。第 1 符号器ステージ 2 6 又は第 2 符号器ステージ 2 8 の選択は、制御部 1 6 により出力される制御信号 5 2 によって制御されているスイッチ 2 4 を用いて実行され、この様子は図 2 において概略的に示されている。制御信号 5 2 は、第 1 符号器ステージ 2 6 又は第 2 符号器ステージ 2 8 のいずれが使用されるべきかを示している。その制御信号 5 2 に基づいて、図 2 において矢印 5 4 により示され、且つ L P C 係数、重み付き L P C 係数、オーディオ信号、重み付きオーディオ信号、及びピッチラグのセットを少なくとも含む所要信号が、第 1 符号器ステージ 2 6 又は第 2 符号器ステージ 2 8 の一方に対して適用される。選択された符号器ステージは、その関連する符号化アルゴリズムを適用し、符号化済み表現 5 6 又は 5 8 を出力インターフェイス 3 0 へ出力する。出力インターフェイス 3 0 は、符号化済みオーディオ信号 6 0 を出力するよう構成されてもよく、その信号 6 0 は、符号化済み表現 5 6 又は 5 8 の他に、L P C 係数又は重み付き L P C 係数、選択された符号化アルゴリズムのためのパラメータ、及び選択された符号化アルゴリズムについての情報を含んでもよい。

## 【 0 0 6 3 】

第 1 品質尺度及び第 2 品質尺度を推定するための特別な実施形態であって、第 1 及び第 2 の品質尺度が重み付き信号ドメインにおけるセグメント S N R であるものを、図 3 を参照しながら以下に説明する。図 3 は、第 1 推定部 1 2、第 2 推定部 1 4、及びそれらの機能を、個々の推定をステップ毎に示すフローチャートの形式で示している。

## 【 0 0 6 4 】

T C X セグメント S N R の推定

第 1 ( T C X ) 推定部は、オーディオ信号 4 0 ( 入力信号 )、重み付き L P C 係数 4 2、及び重み付きオーディオ信号 4 4 を入力として受信する。オーディオ信号 4 0 のフィルタ処理済みバージョンがステップ 9 8 において生成される。そのオーディオ信号 4 0 のフィルタ処理済みバージョンの中では、ハーモニクスが低減又は抑制されている。

## 【 0 0 6 5 】

オーディオ信号 4 0 は、一つ又は複数のハーモニシティ尺度（例えば正規化相関又は予測ゲインなど）及び / 又は一つ又は複数の時間的構造尺度（例えば時間的平坦度又はエネルギー変化など）を決定するために、分析されてもよい。これら尺度の 1 つ、又はこれら尺度の組合せに基づいて、フィルタ 2、つまりフィルタ処理 9 8 が無効化されてもよい。フィルタ処理 9 8 が無効化された場合には、第 1 品質尺度の推定は、オーディオ信号のフィ

10

20

30

40

50

ルタ処理済みバージョンではなく、オーディオ信号 40 を使用して実行される。

#### 【0066】

本発明の実施形態において、フィルタ処理 98 から生じたオーディオ信号内の不連続部を除去するために、不連続部を除去するステップ（図 3 には図示せず）をフィルタ処理 98 に続いて配置してもよい。

#### 【0067】

ステップ 100 において、オーディオ信号 40 のフィルタ処理済みバージョンが窓掛けされる。窓掛けは、10ms の低オーバーラップ・サイン窓を用いて実行されてもよい。過去のフレームが ACELP であった場合、ブロックサイズは 5ms 分増加されてもよく、窓の左側は矩形であってもよく、ACELP 合成フィルタの窓掛け済みゼロインパルス応答は、窓掛け済み入力信号から除去されてもよい。この処理は、TCX アルゴリズムで行われる処理と同様である。オーディオ信号の一部分を表現する、オーディオ信号 40 の窓掛け済みバージョンの 1 フレームがステップ 100 から出力される。

#### 【0068】

ステップ 102 において、窓掛け済みオーディオ信号、即ち結果として得られたフレームが、MDCT（修正離散コサイン変換）を用いて変換される。ステップ 104 において、重み付き LPC 係数を用いて MDCT スペクトルを整形することにより、スペクトル整形が実行される。

#### 【0069】

ステップ 106 において、エントロピーコード、例えば算術コードを用いて符号化された場合に、ゲイン G を用いて量子化された重み付きスペクトルが所与の目標 R をもたらすように、グローバルゲイン G が推定される。用語「グローバルゲイン」が使用される理由は、1 つのゲインが全体フレームのために決定されるからである。

#### 【0070】

グローバルゲイン推定の構成例を以下に説明する。このグローバルゲイン推定は、TCX 符号化アルゴリズムが算術符号器と共にスカラ量子化器を使用する実施形態に適している点に留意すべきである。算術符号器と共に使用されるそのようなスカラ量子化器は、MP EG USAC 標準において想定されている。

#### 【0071】

##### 初期化

最初に、ゲイン推定において使用される変数が以下により初期化される。

1. Set  $en[i] = 9.0 + 10.0 \cdot \log_{10}(c[4 \cdot i + 0] + c[4 \cdot i + 1] + c[4 \cdot i + 2] + c[4 \cdot i + 3])$ ,  
ここで、 $0 \leq i < L/4$ 、 $c[\ ]$  は量子化すべき係数のベクトルであり、 $L$  は  $c[\ ]$  の長さである。
2. Set  $fac = 128$ ,  $offset = fac$  and  $target = \text{any value (e.g. 1000)}$

#### 【0072】

##### 反復

次に、以下の操作ブロックが NITER 回実行される（例えば、ここでは  $NITER = 10$ ）。

1.  $fac = fac/2$
2.  $offset = offset - fac$
3.  $ener = 0$
4. for every  $i$  where  $0 \leq i < L/4$  do the following:  
if  $en[i] - offset > 3.0$ , then  $ener = ener + en[i] - offset$
5. if  $ener > target$ , then  $offset = offset + fac$

#### 【0073】

反復の結果はオフセット値である。この反復の後で、グローバルゲインが

$$G = 10^{(offset/20)}$$

10

20

30

40

50

として推定される。

#### 【 0 0 7 4 】

グローバルゲインが推定される特別な方法は、使用される量子化器及びエントロピーコードに依存して変化し得る。MPEG-1 USAC 標準においては、算術符号器と共にスカラ量子化器が想定されている。他のTCXの手法は、異なる量子化器を使用してもよく、そのような異なる量子化器のためのグローバルゲインを推定する方法は、当業者に理解できるであろう。例えば、AMR-WB+ 標準は、RE8 格子量子化器が使用されると想定する。そのような量子化器については、非特許文献1の5.3.5.7章(34頁)に説明されているように、グローバルゲインが推定されてもよく、ここでは固定の目標ビットレートが想定されている。

10

#### 【 0 0 7 5 】

ステップ106でグローバルゲインを推定した後で、ステップ108では歪み推定が行われる。更に具体的には、推定されたグローバルゲインに基づいて量子化歪みが近似される。この実施形態においては、一様スカラ量子化器が使用されると想定している。従って、量子化歪みは、 $D = G^2 / 12$  という簡略式を用いて決定され、ここで、Dは決定された量子化歪みを表し、Gは推定されたグローバルゲインを表す。これは、一様スカラ量子化歪みの高レート近似に対応する。

#### 【 0 0 7 6 】

決定された量子化歪みに基づいて、セグメントSNRの計算がステップ110において実行される。フレームの各サブフレームにおけるSNRが、重み付きオーディオ信号エネルギーと、サブフレーム内で一定だと仮定されている歪みDとの比として計算される。例えば、フレームが4つの連続的なサブフレームへと分割される。その場合、セグメントSNRはそれら4つのサブフレームのSNRの平均値であり、dBで示されてもよい。

20

#### 【 0 0 7 7 】

この手法は、TCXアルゴリズムを用いて対象となるフレームを実際に符号化及び復号化したときに得られるであろう第1のセグメントSNRの推定を可能にするが、実際にそのオーディオ信号を符号化及び復号化する必要はなく、従って、演算量の大幅な低減と計算時間の低減をもたらす。

#### 【 0 0 7 8 】

ACELPセグメントSNRの推定

30

#### 【 0 0 7 9 】

第2推定部14は、重み付きオーディオ信号44と、前処理ユニット22で既に計算されたピッチラグ48のセットとを受信する。

#### 【 0 0 8 0 】

ステップ112において示したように、各サブフレームにおいて、上記重み付きオーディオ信号と上記ピッチラグTとを単に使用して適応型コードブックが近似される。適応型コードブックは次式によって近似され、

$$xw(n-T), n = 0, \dots, N$$

ここで、xwは重み付きオーディオ信号であり、Tは対応するサブフレームのピッチラグであり、Nはサブフレーム長である。したがって、適応型コードブックは、Tの分だけ過去へとシフトされたサブフレームのバージョンを使用することによって近似される。それ故、本発明の実施形態では、適応型コードブックは非常に簡易な方法で近似される。

40

#### 【 0 0 8 1 】

ステップ114では、各サブフレームについて適応型コードブックゲインが決定される。さらに具体的には、各サブフレームにおいて、重み付きオーディオ信号と近似された適応型コードブックとの間の誤差を最小化するように、コードブックゲインGが推定される。この推定は、各サンプルについて両信号間の差を単純に比較し、これら差の合計が最小になるようにゲインを見つけることによって、実行される。

#### 【 0 0 8 2 】

ステップ116では、各サブフレームについて適応型コードブック歪みが決定される。各

50



サブフレームにおいて、適応型コードブックによって導入された歪み  $D$  は、単に、重み付きオーディオ信号とゲイン  $G$  によってスケールされた近似された適応型コードブックとの間の誤差のエネルギーである。

【 0 0 8 3 】

ステップ 1 1 6 で決定された歪みは、革新的コードブックを考慮に入れる、任意選択的なステップ 1 1 8 において調整されてもよい。A C E L P アルゴリズムにおいて使用される革新的コードブックの歪みは、一定値として単純に推定されてもよい。本発明の上述の実施例では、革新的コードブックは一定ファクタによって歪み  $D$  を低減するものと仮定されている。よって、各サブフレームについてステップ 1 1 6 で得られた歪みは、ステップ 1 1 8 で一定ファクタ、例えば 0 . 0 5 5 のような 0 ~ 1 のオーダーの一定ファクタによって乗算されてもよい。

10

【 0 0 8 4 】

ステップ 1 2 0 では、セグメント  $S N R$  の計算が行われる。各サブフレームにおいて、 $S N R$  が重み付きオーディオ信号エネルギーと歪み  $D$  との比として計算される。セグメント  $S N R$  は 4 個のサブフレームの  $S N R$  の平均であり、 $d B$  で示されても良い。

【 0 0 8 5 】

この手法は、A C E L P アルゴリズムを用いて対象となるフレームを実際に符号化 / 復号化した場合に得られる第 2 の  $S N R$  の推定を可能にするが、オーディオ信号を実際に符号化 / 復号化する必要はなく、それ故に、大幅に演算量を削減でき、かつ演算時間を削減できる。

20

【 0 0 8 6 】

第 1 と第 2 の推定部 1 2、1 4 は、推定されたセグメント  $S N R$  4 6、5 0 を制御部 1 6 へと出力し、制御部 1 6 は、推定されたセグメント  $S N R$  4 6、5 0 に基づいてオーディオ信号の関連する部分のためにどのアルゴリズムが使用されるべきかの決定を行う。制御部は、上記決定をより安定化させるためにヒステリシス・メカニズムを任意選択的に使用してもよい。例えば、閉ループ決定におけるのと同様のヒステリシス・メカニズムが僅かに異なる調整パラメータを用いて使用されてもよい。そのようなヒステリシス・メカニズムは、上記推定されたセグメント  $S N R$  (両者の差など) と他のパラメータとに依存し得る値「 $d s n r$ 」を計算してもよく、他のパラメータとしては、例えば以前の決定についての統計や、時間的に安定なフレームの数やフレーム内の過渡の数などである。

30

【 0 0 8 7 】

ヒステリシス・メカニズムを使用せずに、制御部はより高い推定された  $S N R$  を持つ符号化アルゴリズムを選択してもよい。すなわち、第 2 の推定された  $S N R$  が第 1 の推定された  $S N R$  よりも高い場合には A C E L P が選択され、第 1 の推定された  $S N R$  が第 2 の推定された  $S N R$  よりも高い場合には T C X が選択されてもよい。ヒステリシス・メカニズムを使用すれば、制御部は、以下のような決定規則に従って符号化アルゴリズムを選択してもよく、ここで、 $acelp\_snr$  が第 2 の推定された  $S N R$  であり、 $tcx\_snr$  が第 1 の推定された  $S N R$  である。

if  $acelp\_snr + dsnr > tcx\_snr$  then select ACELP, otherwise select TCX.

【 0 0 8 8 】

40

ハーモニクスの振幅を低減するためのフィルタのパラメータの決定

【 0 0 8 9 】

ハーモニクスの振幅を低減するためのフィルタのパラメータの決定についての実施例について、次に説明する。フィルタパラメータは、例えばユニット 8 などの符号器側で推定されてもよい。

【 0 0 9 0 】

ピッチ推定

【 0 0 9 1 】

1 フレーム当たり 1 つのピッチラグ ( 整数部 + 小数部 ) が推定される ( フレームサイズは例えば 2 0 m s )。この推定は、演算量を低減しかつ推定精度を向上させるため、3 ステッ

50

ブで実施される。

【 0 0 9 2 】

a) ピッチラグの整数部の第 1 推定

滑らかなピッチ展開の輪郭を形成するピッチ分析アルゴリズムが使用される（例えば、非特許文献 3 の第 6 . 6 章に開示された開ループピッチ分析など）。この分析は、通常、サブフレーム・ベース（サブフレーム・サイズは例えば 1 0 m s ）で実行され、1 サブフレーム当たり 1 つのピッチラグ推定値が生成される。これらピッチラグ推定値は如何なる小数部も持たず、通常はダウンサンプリング済み信号（サンプリングレートは例えば 6 4 0 0 H z ）に対して推定される点に留意すべきである。使用される信号は如何なるオーディオ信号でもよく、例えば非特許文献 3 の第 6 . 5 章に開示された L P C 重み付きオーディオ信号であってもよい。

10

【 0 0 9 3 】

b) ピッチラグの整数部  $T_{int}$  の精製

コア符号器サンプリングレートで作動しているオーディオ信号  $x[n]$  に対してピッチラグの最終の整数部が推定され、そのサンプリングレートは通常、a) で使用されたダウンサンプリング済み信号のサンプリングレート（例えば 1 2 . 8 k H z 、1 6 k H z 、3 2 k H z など）よりも高い。信号  $x[n]$  は、例えば L P C 重み付きオーディオ信号のような、如何なるオーディオ信号であってもよい。

【 0 0 9 4 】

次に、ピッチラグの整数部  $T_{int}$  は、次の自己相関関数を最大化するラグであり、

20

$$C(d) = \sum_{n=0}^N x[n]x[n-d]$$

ここで、 $d$  は a) で推定されたピッチラグ  $T$  の周辺値である。

$$T - \delta_1 \leq d \leq T + \delta_2$$

【 0 0 9 5 】

c) ピッチラグの小数部  $T_{fr}$  の推定

30

ステップ b) で計算された自己相関関数  $C(d)$  を補間し、補間済み自己相関関数を最大化する小数ピッチラグを選択することによって、小数部  $T_{fr}$  が発見される。この補間は、例えば非特許文献 3 の第 6 . 6 . 7 章に開示されたローパス F I R フィルタを使用することで実行され得る。

【 0 0 9 6 】

ゲイン推定と量子化

【 0 0 9 7 】

ゲインは、一般的にコア符号器のサンプリングレートで入力オーディオ信号に対して推定されるが、その入力オーディオ信号は L P C 重み付きオーディオ信号のような如何なる信号でもよい。この信号は  $y[n]$  と表記され、 $x[n]$  と同じか又は相違していてもよい。

40

【 0 0 9 8 】

$y[n]$  の予測値  $y_p[n]$  は、まず次のフィルタを用いて  $y[n]$  をフィルタリングすることにより発見される。

$$P(z) = B(z, T_{fr})z^{-T_{int}}$$

ここで、 $T_{int}$  は、( b )において推定された)ピッチラグの整数部であり、 $B(z, T_{fr})$  は、その係数が ( c )において推定された)ピッチラグの小数部  $T_{fr}$  に依存するローパス F I R フィルタである。

【 0 0 9 9 】

50

ピッチラグ分解能が 1 / 4 である場合の B ( z ) の一例は次の通りである。

$$\begin{aligned} T_{fr} = \frac{0}{4} & \quad B(z) = 0.0000z^{-2} + 0.2325z^{-1} + 0.5349z^0 + 0.2325z^1 \\ T_{fr} = \frac{1}{4} & \quad B(z) = 0.0152z^{-2} + 0.3400z^{-1} + 0.5094z^0 + 0.1353z^1 \\ T_{fr} = \frac{2}{4} & \quad B(z) = 0.0609z^{-2} + 0.4391z^{-1} + 0.4391z^0 + 0.0609z^1 \\ T_{fr} = \frac{3}{4} & \quad B(z) = 0.1353z^{-2} + 0.5094z^{-1} + 0.3400z^0 + 0.0152z^1 \end{aligned}$$

10

【 0 1 0 0 】

次に、ゲイン g は以下のように計算され、

$$g = \frac{\sum_{n=0}^{N-1} y[n]y_P[n]}{\sum_{n=0}^{N-1} y_P[n]y_P[n]}$$

0 と 1 との間に制限される。

【 0 1 0 1 】

最後に、ゲイン g は、例えば一様量子化を使用して、例えば 2 ビットで量子化される。

【 0 1 0 2 】

20

はフィルタの強度を制御するために使用される。1 に等しい は全効果を生成する。0 に等しい はフィルタを無効化する。それ故、本発明の実施形態では、フィルタは を 0 の値に設定することによって無効化されてもよい。本発明の実施形態では、フィルタが有効化される場合、 は 0 . 5 ~ 0 . 7 5 の間の値に設定されてもよい。本発明の実施形態では、フィルタが有効化される場合、 は 0 . 6 2 5 の値に設定されてもよい。B ( z , T<sub>fr</sub> ) の一例が上述のように与えられる。また、B ( z , T<sub>fr</sub> ) の次数と係数は、ビットレートと出力サンプリングレートとに依存し得る。ビットレートと出力サンプリングレートとの各組合せについて、異なる周波数応答を設計でき、かつ調整できる。

【 0 1 0 3 】

フィルタの無効化

30

1 つ又は複数のハーモニシティ尺度及び / 又は 1 つ又は複数の時間的構造尺度の組合せに基づいて、フィルタは無効化されてもよい。このような尺度の実例を以下に説明する。

【 0 1 0 4 】

i ) ステップ b ) で推定された整数ピッチラグにおける正規化相関のようなハーモニシティ尺度

$$norm. corr. = \frac{\sum_{n=0}^N x[n]x[n - T_{int}]}{\sqrt{\sum_{n=0}^N x[n]x[n]} \sqrt{\sum_{n=0}^N x[n - T_{int}]x[n - T_{int}]}}$$

【 0 1 0 5 】

40

入力信号が整数ピッチラグによって完全に予測可能であれば、正規化相関は 1 であり、入力信号が全く予測不能であれば、正規化相関は 0 である。よって、高い値 ( 1 に近い ) は、ハーモニック信号を示すであろう。さらにロバストな決定のために、過去フレームの正規化相関もまたこの決定に使用されてもよく、例えば、

( norm.corr.( curr. ) \* norm.corr.( prev. ) ) > 0.25

であれば、フィルタは無効化されない。

【 0 1 0 6 】

i i ) 例えば、過渡検出のための過渡検出部によっても使用される、エネルギーサンプルに基づいて計算された時間的構造尺度 ( 例えば、時間的平坦度やエネルギー変化など ) 。

例えば、

50

(temporal flatness measure > 3.5

又は

energy change > 3.5)

であれば、フィルタは無効化される。

【 0 1 0 7 】

1つ又は複数のハーモニシティ尺度の決定に関する更なる詳細を以下に説明する。

【 0 1 0 8 】

ハーモニシティの尺度は、例えばオーディオ信号又はその前処理済みバージョンの、ピッチラグにおける若しくはその周辺の正規化相関によって計算される。ピッチラグは第1ステージと第2ステージとを含むステージ内で決定されてもよく、第1ステージ内では、第1サンプルレートのダウンサンプリング済みドメインでピッチラグの予備推定が決定され、第2ステージでは、ピッチラグの予備推定が第1サンプルレートよりも高い第2サンプルレートで精製される。ピッチラグは、例えば自己相関を用いて決定される。少なくとも1つの時間的構造尺度は、例えばピッチ情報に依存して時間的に配置された時間領域内で決定される。時間領域の時間的に過去に向かう側の端部は、例えばピッチ情報に依存して配置されている。時間領域の時間的に過去に向かう側の端部は、その時間領域の時間的に過去に向かう側の端部がピッチ情報の増大に伴って単調増加する時間量だけ過去に向かってずれるように、配置されてもよい。時間領域の時間的に将来に向かう側の端部は、時間的候補領域内でオーディオ信号の時間的構造に依存して配置されてもよく、その時間的候補領域は、時間領域の又は時間的構造尺度の決定に対して高い影響度を持つ領域の時間的に過去に向かう側の端部から、現在フレームの時間的に将来に向かう側の端部まで延びている。時間的候補領域内の最大エネルギーサンプルと最小エネルギーサンプルとの間の振幅又は比がこの目的で使用されてもよい。例えば、少なくとも1つの時間的構造尺度が、上記時間領域内でのオーディオ信号の平均又は最大エネルギー変化を示してもよく、また、その少なくとも1つの時間的構造尺度が所定の第1閾値より小さく、かつ現在のフレーム及び/又は先行するフレームについてハーモニシティ尺度が第2閾値より大きいとき、無効化の条件が満たされてもよい。また、現在フレームについてのハーモニシティ尺度が第3閾値より大きく、かつ現在フレーム及び/又は先行するフレームについてのハーモニシティ尺度が、ピッチラグの増大につれて減少する第4閾値より大きいとき、無効化の条件が満たされてもよい。

【 0 1 0 9 】

前記尺度を決定するための具体的実施例のステップ毎の説明を以下に行う。

【 0 1 1 0 】

ステップ1．過渡検出と時間的尺度

【 0 1 1 1 】

入力信号  $s_{HP}(n)$  は時間ドメインの過渡検出部に入力される。入力信号  $s_{HP}(n)$  はハイパスフィルタ処理される。過渡検出部のハイパスフィルタの伝達関数は、次の通りである。

$$H_{TD}(z) = 0.375 - 0.5z^{-1} + 0.125z^{-2} \quad (1)$$

【 0 1 1 2 】

過渡検出部のハイパスフィルタによってフィルタ処理された信号は、 $s_{TD}(n)$  で示される。このハイパスフィルタ処理された信号  $s_{TD}(n)$  は同じ長さの8個の連続したセグメントにセグメント化される。各セグメントについて、ハイパスフィルタ処理された信号  $s_{TD}(n)$  のエネルギーは次式のように計算される。

$$E_{TD}(i) = \sum_{n=0}^{L_{segment}-1} (s_{TD}(iL_{segment} + n))^2, \quad i = 0, \dots, 7 \quad (2)$$

10

20

30

40

50

ここで、

$$L_{segment} = \frac{L}{8}$$

は入力サンプリング周波数での 2 . 5 ミリ秒セグメント内におけるサンプルの個数である。

【 0 1 1 3 】

累積エネルギーは次式を使用して計算される。

$$E_{Acc} = \max(E_{TD}(i-1), 0.8125E_{Acc}) \quad (3) \quad 10$$

【 0 1 1 4 】

セグメント  $E_{TD}(i)$  のエネルギーが累積エネルギーを一定ファクタ

$$attackRatio = 8.5$$

分だけ超えた場合には、アタック (attack) が検出され、そのアタックインデックスは  $i$  に設定される。

$$E_{TD}(i) > attackRatio \cdot E_{Acc} \quad (4) \quad 20$$

【 0 1 1 5 】

もし、上記基準に基づいたアタックは検出されないが、セグメント  $i$  内で強いエネルギー上昇が検出された場合には、アタックインデックスはアタックの存在を示さずに  $i$  に設定される。そのアタックインデックスは基本的に幾つかの追加的制限の下で 1 フレーム内での最終アタックの位置に設定される。

【 0 1 1 6 】

各セグメントについてのエネルギー変化は、次のように計算される。

$$E_{chg}(i) = \begin{cases} \frac{E_{TD}(i)}{E_{TD}(i-1)}, & E_{TD}(i) > E_{TD}(i-1) \\ \frac{E_{TD}(i-1)}{E_{TD}(i)}, & E_{TD}(i-1) > E_{TD}(i) \end{cases} \quad (5) \quad 30$$

【 0 1 1 7 】

時間的平坦度は次のように計算される。

$$TFM(N_{past}) = \frac{1}{8 + N_{past}} \sum_{i=-N_{past}}^7 E_{chg}(i) \quad (6)$$

【 0 1 1 8 】

最大エネルギー変化は次のように計算される。

$$MEC(N_{past}, N_{new}) = \max(E_{chg}(-N_{past}), E_{chg}(-N_{past} + 1), \dots, E_{chg}(N_{new} - 1)) \quad (7) \quad 40$$

【 0 1 1 9 】

$E_{chg}(i)$  又は  $E_{TD}(i)$  のインデックスが負である場合には、そのインデックスは、先行するセグメントからの値を示し、そのセグメントインデックスは現在フレームに關係している。

【 0 1 2 0 】

$N_{past}$  は、過去のフレームからのセグメントの個数である。時間的平坦度が ACELP / TCX 決定において使用するために計算される場合には、その数は 0 に等しい。もし、時

間的平坦度が  $TCX$   $LTP$  決定のために計算される場合には、その数は次のようになる。

$$N_{past} = 1 + \min \left( 8, \left\lceil 8 \frac{pitch}{L} + 0.5 \right\rceil \right) \quad (8)$$

#### 【 0 1 2 1 】

$N_{new}$  は、現在のフレームからのセグメントの個数である。その個数は、非過渡フレームについては 8 個である。過渡フレームについては、まず最大及び最小エネルギーを持つセグメントの位置が発見される。

10

$$i_{\max} = \arg \max_{i \in \{N_{past}, \dots, 7\}} E_{TD}(i) \quad (9)$$

$$i_{\min} = \arg \min_{i \in \{N_{past}, \dots, 7\}} E_{TD}(i) \quad (10)$$

#### 【 0 1 2 2 】

もし、

$$E_{TD}(i_{\min}) > 0.375 E_{TD}(i_{\max}) \quad 20$$

であれば、 $N_{new}$  は  $i_{\max} - 3$  に設定され、それ以外であれば、 $N_{new}$  は 8 に設定される。

#### 【 0 1 2 3 】

#### ステップ 2 . 変換ブロック長の切り替え

#### 【 0 1 2 4 】

$TCX$  のオーバーラップ長及び変換ブロック長は、過渡の存在及び過渡の位置に依存している。

#### 【 0 1 2 5 】

表 1 : 過渡位置に基づいたオーバーラップ長及び変換長のコーディング

30

【表 1】

Attack-Index	Overlap with the first window of the following frame	Short/Long Transform decision (binary coded) 0 – Long, 1 - Short	Binary code for the overlap width	Overlap code
none	ALDO	0	0	00
-2	FULL	1	0	10
-1	FULL	1	0	10
0	FULL	1	0	10
1	FULL	1	0	10
2	MINIMAL	1	10	110
3	HALF	1	11	111
4	HALF	1	11	111
5	MINIMAL	1	10	110
6	MINIMAL	0	10	010
7	HALF	0	11	011

## 【 0 1 2 6 】

上述の過渡検出部は、基本的に、以下の制約、つまり、多数の過渡がある場合には、半オーバーラップの方が全オーバーラップよりも好ましく、最小オーバーラップの方が半オーバーラップよりも好ましいという制約の下で、最終アタックのインデックスをリターンする。もし、位置 2 又は 6 でのアタックが十分に強くない場合には、半オーバーラップが最小オーバーラップに代えて選択される。

## 【 0 1 2 7 】

## ステップ 3 . ピッチ推定

## 【 0 1 2 8 】

1 つのピッチラグ ( 整数部 + 小数部 ) は、演算量を削減しかつ推定精度を向上させるために、上述の 3 ステップ ( a ) ~ ( c ) の中で説明したように、フレーム毎 ( フレームサイズが例えば 2 0 m s ) に推定される。

## 【 0 1 2 9 】

## ステップ 4 . 決定ビット

## 【 0 1 3 0 】

入力オーディオ信号が如何なるハーモニックコンテンツを含まない場合、又は予測ベースの技術が時間的構造における歪み ( 例えば短い過渡の繰り返し ) を導入しそうな場合には、次にフィルタを無効化する決定がなされる。

## 【 0 1 3 1 】

その決定は、例えば整数ピッチラグにおける正規化相関や時間的構造尺度などの複数のパラメータに基づいて行われる。

## 【 0 1 3 2 】

整数ピッチラグにおける正規化相関 $\text{norm\_corr}$ は、上述したように推定される。入力信号が整数ピッチラグによって完全に予測可能であれば、正規化相関は1であり、入力信号が予測不能であれば、0である。その場合、高い値（1に近い）はハーモニック信号を示しているであろう。大半の口バスト決定にとって、現在のフレームについての正規化相関( $\text{norm\_corr}(\text{curr})$ )に加えて、過去のフレームの正規化相関( $\text{norm\_corr}(\text{prev})$ )が決定において使用され得る。例えば、

$$(\text{norm\_corr}(\text{curr}) * \text{norm\_corr}(\text{prev})) > 0.25$$

又は

$$\max(\text{norm\_corr}(\text{curr}), \text{norm\_corr}(\text{prev})) > 0.5$$

であれば、現在のフレームは幾分かのハーモニックコンテンツを含む。

## 【 0 1 3 3 】

強い過渡や大きな時間的变化を含む信号に対してフィルタが作動するのを避けるため、時間的構造尺度は過渡検出部（例えば時間的平坦度（式6）及び最大エネルギー変化（式7））によって計算されてもよい。時間的特徴は現在フレーム（ $N_{\text{new}}$ 個のセグメント）とピッチラグまでの過去のフレーム（ $N_{\text{past}}$ 個のセグメント）とを含む信号に対して計算される。ゆっくりと減衰する階段状の過渡については、過渡の位置（ $i_{\text{max}} - 3$ ）までだけ全て又は幾つの特徴が計算される。なぜなら、LTPフィルタリングによって導入されるスペクトルの非ハーモニック部分における歪みは、強く長く続く過渡（例えばクラッシュシンバル）のマスキングによって抑圧されるだろうからである。

## 【 0 1 3 4 】

低ピッチの信号におけるパルス列は、過渡検出部によって過渡として検出され得る。従って、低ピッチの信号については、過渡検出部からの特徴は無視され、それに代えてピッチラグに依存した正規化相関についての追加的閾値が与えられる。例えば、

$$\text{norm\_corr} > 1 - 2^{-T_{\text{int}}/L}$$

であれば、フィルタを無効化する。

## 【 0 1 3 5 】

以下に一例の決定を示す。そこでは、 $b1$ が何らかのビットレート、例えば48 kbpsであり、 $\text{TCX\_20}$ はフレームが単一のロングブロックを使用して符号化されることを示し、 $\text{TCX\_10}$ はフレームが2、3、4個又はそれ以上のショートブロックを使用して符号化されることを示し、 $\text{TCX\_20} / \text{TCX\_10}$ 決定は上述の過渡検出部の出力に基づいている。 $\text{tempFlatness}$ は（6）式で定義された時間的平坦度であり、 $\text{maxEnergyChange}$ は（7）式で定義された最大エネルギー変化である。

$$\text{norm\_corr}(\text{curr}) > 1 - 2^{-T_{\text{int}}/L}$$

という条件は、

$$(1 - 2^{-\text{norm\_corr}(\text{curr})}) * L < T_{\text{int}}$$

と書き換えることもできる。

## 【 0 1 3 6 】

## 【表2】

```
enableLTP =
(bitrate < b1 && tcxMode == TCX_20 && (norm_corr(curr) * norm_corr(prev)) > 0.25 && tempFlatness < 3.5) ||
(bitrate >= b1 && tcxMode == TCX_10 && max(norm_corr(curr), norm_corr(prev)) > 0.5 && maxEnergyChange < 3.5) ||
(bitrate >= b1 && norm_corr(curr) > 0.44 && norm_corr(curr) > 1.2 * T_int / L) ||
(bitrate >= b1 && tcxMode == TCX_20 && norm_corr(curr) > 0.44 &&
(tempFlatness < 6.0 || (tempFlatness < 7.0 && maxEnergyChange < 22.0)));

( bitrate >= b1 && tcxMode == TCX_20 && norm_corr > 0.44 &&
```

10

20

30

40

50



## 【 0 1 3 7 】

過渡の決定は、長期予測についてどの決定メカニズムが使用され、その決定に用いられる尺度のために信号のどの部分が使用されるかに影響を及ぼすものであって、過渡の決定が長期予測フィルタの無効化を直接的にトリガーするものではないということが、上述の実例から明らかである。

## 【 0 1 3 8 】

変換長の決定に使用される時間的尺度は、LTPフィルタ決定に使用される時間的尺度と完全に相違していても良く、又は、前者の尺度は後者の尺度と重複してもよく、若しくは全く同じであって異なる領域で計算されたものであってもよい。低ピッチの信号の場合、ピッチラグに依存する正規化相関についての閾値に到達したときには、過渡の検出は完全に無視されてもよい。

10

## 【 0 1 3 9 】

可能性のある不連続部を取り除く技術

## 【 0 1 4 0 】

線形フィルタ $H(z)$ をフレーム毎に適用することに起因する不連続部を取り除くための可能性のある技術をここで説明する。線形フィルタは上述のLTPフィルタであってもよい。線形フィルタは、FIR（有限インパルス応答）フィルタ又はIIR（無限インパルス応答）フィルタであってもよい。提案手法は、過去のフレームのフィルタパラメータを用いて現在のフレームの部分をフィルタリングせず、それによって公知手法の起こり得る問題を避けるものである。提案手法は、不連続部を取り除くためにLPCフィルタを使用する。このLPCフィルタは、（線形の時間 - 不変フィルタ $H(z)$ によってフィルタ処理済み、又は未処理の）オーディオ信号に対して推定され、そのため、（ $H(z)$ によってフィルタ処理済み、又は未処理の）オーディオ信号のスペクトル形状の良好なモデルとなる。それ故、LPCフィルタは、オーディオ信号のスペクトル形状が不連続部をマスクするように、使用される。

20

## 【 0 1 4 1 】

LPCフィルタは様々な方法で推定され得る。LPCフィルタは、例えば、オーディオ信号（現在及び/又は過去のフレーム）とLevinson-Durbin アルゴリズムとを使用して推定され得る。また、LPCフィルタは、Levinson-Durbin アルゴリズムを使用して、過去にフィルタ処理されたフレーム信号に対して計算され得る。

30

## 【 0 1 4 2 】

$H(z)$ がオーディオコーデックにおいて使用され、このオーディオコーデックが、例えば変換ベースのオーディオコーデックの中で量子化ノイズを整形するためにLPCフィルタ（量子化済み、又は未量子化）を既に使用している場合には、新たなLPCフィルタを推定するために必要な追加的演算量を使用せずに、上記LPCフィルタが不連続部を円滑化するために直接的に使用され得る。

## 【 0 1 4 3 】

以下に、FIRフィルタの場合とIIRフィルタの場合とにおける現在フレームの処理について説明する。過去フレームは既に処理済みであると仮定する。

## 【 0 1 4 4 】

40

FIRフィルタの場合：

- 1．現在フレームのフィルタパラメータを用いて現在フレームをフィルタ処理し、フィルタ処理済みの現在フレームを生成する。
- 2．オーディオ信号（フィルタ処理済み又は未処理）に対して推定された次数 $M$ のLPCフィルタ（量子化済み又は未量子化）を考慮する。
- 3．過去フレームの最後の $M$ 個のサンプルがフィルタ $H(z)$ と現在フレームの係数とを用いてフィルタ処理され、フィルタ処理済み信号の第1部分を生成する。
- 4．フィルタ処理済みの過去フレームの $M$ 個の最終サンプルは、次にフィルタ処理済み信号の第1部分から差し引かれ、フィルタ処理済み信号の第2部分を生成する。
- 5．次に、LPCフィルタとフィルタ処理済み信号の第2部分に等しいその初期状態とを

50

用いて、ゼロサンプルのフレームをフィルタ処理することにより、LPCフィルタのゼロインパルス応答(ZIR)が生成される。

6. 上記ZIRは、その振幅が速やかに0になるように、任意選択的に窓掛けされ得る。

7. 上記ZIRの開始部分はフィルタ処理済みの現在フレームの対応する開始部分から差し引かれる。

#### 【0145】

IIRフィルタの場合：

1. オーディオ信号(フィルタ処理済み又は未処理)に対して推定された、次数MのLPCフィルタ(量子化済み又は未量子化)を考慮する。

2. 過去フレームの最後のM個のサンプルがフィルタH(z)と現在フレームの係数とを用いてフィルタ処理され、フィルタ処理済み信号の第1部分を生成する。

3. フィルタ処理済みの過去フレームのM個の最終サンプルは、次にフィルタ処理済み信号の第1部分から差し引かれ、フィルタ処理済み信号の第2部分を生成する。

4. 次に、LPCフィルタとフィルタ処理済み信号の第2部分に等しいその初期状態とを用いて、ゼロサンプルのフレームをフィルタ処理することにより、LPCフィルタのゼロインパルス応答(ZIR)が生成される。

5. 上記ZIRは、その振幅が速やかに0になるように、任意選択的に窓掛けされ得る。

6. 現在フレームの開始部分は次に、現在フレームの第1サンプルから開始するように、サンプル毎に処理される。

7. そのサンプルはフィルタH(z)と現在フィルタのパラメータとを用いてフィルタ処理され、第1のフィルタ処理済みサンプルを生成する。

8. 上記ZIRの対応するサンプルは、次に第1のフィルタ処理済みサンプルから差し引かれ、フィルタ処理済みの現在フレームの対応するサンプルを生成する。

9. 次のサンプルへ移行する。

10. 現在フレームの開始部分の最終サンプルが処理されるまでステップ9~12(を繰り返す)。

11. 現在フレームの残りのサンプルを現在フレームのフィルタパラメータを用いてフィルタ処理する。

#### 【0146】

従って、本発明の幾つかの実施形態は、セグメントSNRを推定すること、及び簡易で正確な方法で適切な符号化アルゴリズムの選択を行うこと、を可能にする。特に、本発明の実施形態では、適切な符号化アルゴリズムの開ループ選択が可能になり、ハーモニクスを含むオーディオ信号の場合に、符号化アルゴリズムの不適切な選択が避けられる。

#### 【0147】

上述の実施形態では、個々のサブフレームについて推定されたSNRの平均値を計算することによって、セグメントSNRが推定される。代替的な実施形態においては、フレームをサブフレームに分割することなく、全体フレームのSNRを推定することも可能であろう。

#### 【0148】

本発明の実施形態では、閉ループ選択において必要となるステップ数を削減できるので、閉ループ選択に比べて演算時間を大幅に低減することができる。

#### 【0149】

従って、本発明の手法によって、良好な性能を持つ適切な符号化アルゴリズムの選択を可能にしながら、多数のステップとそれに関連する演算時間が節約可能になる。

#### 【0150】

これまで装置の文脈で幾つかの態様を示してきたが、これらの態様は対応する方法の説明をも表しており、そのブロック又は装置が方法ステップ又は方法ステップの特徴に対応することは明らかである。同様に、方法ステップを説明する文脈で示した態様もまた、対応する装置の対応するブロックもしくは項目又は特徴を表している。

#### 【0151】

ここで説明した装置の実施形態及びその特徴は、コンピュータ、1つ又は複数のプロセッサ、1つ又は複数のマイクロプロセッサ、フィールド・プログラマブル・ゲートアレイ（FPGAs）、アプリケーション特定型集積回路（ASICs）、及びそれに近似したもの又はそれらの組合せであって、上述の機能を提供するために構成され又はプログラムされたものによって実装されてもよい。

【0152】

方法ステップの幾つか又は全てが、例えばマイクロプロセッサ、プログラム可能なコンピュータ、又は電子回路のようなハードウェア装置によって（又は使用して）実行されてもよい。幾つかの実施形態では、最も重要な方法ステップの幾つか又はそれ以上がそれら装置によって実行されてもよい。

10

【0153】

所定の構成要件にもよるが、本発明の実施形態は、ハードウェア又はソフトウェアにおいて構成可能である。この構成は、その中に格納される電子的に読み取り可能な制御信号を有し、本発明の各方法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（又は協働可能な）、デジタル記憶媒体のような非一時的記憶媒体、例えばフレキシブルディスク、DVD、ブルーレイ、CD、ROM、PROM、EPROM、EEPROM、フラッシュメモリなどのデジタル記憶媒体を使用して実行することができる。したがって、デジタル記憶媒体はコンピュータ読み取り可能であってもよい。

【0154】

本発明に従う幾つかの実施形態は、上述した方法の1つを実行するようプログラム可能なコンピュータシステムと協働可能で、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含む。

20

【0155】

一般的に、本発明の実施例は、プログラムコードを有するコンピュータプログラム製品として構成することができ、そのプログラムコードは当該コンピュータプログラム製品がコンピュータ上で作動するときに、本発明の方法の一つを実行するよう作動可能である。そのプログラムコードは例えば機械読み取り可能なキャリアに記憶されていても良い。

【0156】

本発明の他の実施形態は、上述した方法の1つを実行するための、機械読み取り可能なキャリアに格納されたコンピュータプログラムを含む。

30

【0157】

換言すれば、本発明の方法のある実施形態は、そのコンピュータプログラムがコンピュータ上で作動するときに、上述した方法の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。

【0158】

本発明の他の実施形態は、上述した方法の1つを実行するために記録されたコンピュータプログラムを含む、データキャリア（又はデジタル記憶媒体又はコンピュータ読み取り可能な媒体など）である。そのデータキャリア、デジタル記憶媒体、又は記録された媒体は、典型的に有形及び/又は非一時的である。

【0159】

本発明の他の実施形態は、上述した方法の1つを実行するためのコンピュータプログラムを表現するデータストリーム又は信号列である。そのデータストリーム又は信号列は、例えばインターネットを介するデータ通信接続を介して伝送されるよう構成されてもよい。

40

【0160】

他の実施形態は、上述した方法の1つを実行するように構成又は適応された、例えばコンピュータ又はプログラム可能な論理デバイスのような処理手段を含む。

【0161】

他の実施形態は、上述した方法の1つを実行するためのコンピュータプログラムがインストールされたコンピュータを含む。

【0162】

50

本発明に従う他の実施形態は、ここで説明した方法の1つを実行するためのコンピュータプログラムを、受信器へ（例えば電子的に又は光学的に）伝送するよう構成された装置又はシステムを含む。受信器は、例えばコンピュータ、携帯機器、メモリーデバイス又はそれらの類似物であってもよい。装置又はシステムは、例えばコンピュータプログラムを受信器へと転送するファイルサーバを含んでもよい。

【0163】

幾つかの実施形態においては、（例えば書換え可能ゲートアレイのような）プログラム可能な論理デバイスが、上述した方法の幾つか又は全ての機能を実行するために使用されても良い。幾つかの実施形態では、書換え可能ゲートアレイは、上述した方法の1つを実行するためにマイクロプロセッサと協働しても良い。一般的に、そのような方法は、好適には任意のハードウェア装置によって実行される。

【0164】

上述した実施形態は、本発明の原理を単に例示的に示したに過ぎない。本明細書に記載した構成及び詳細について修正及び変更が可能であることは、当業者にとって明らかである。従って、本発明は、本明細書に実施形態の説明及び解説の目的で提示した具体的詳細によって限定されるものではなく、添付した特許請求の範囲によってのみ限定されるべきである。

10

【図1】

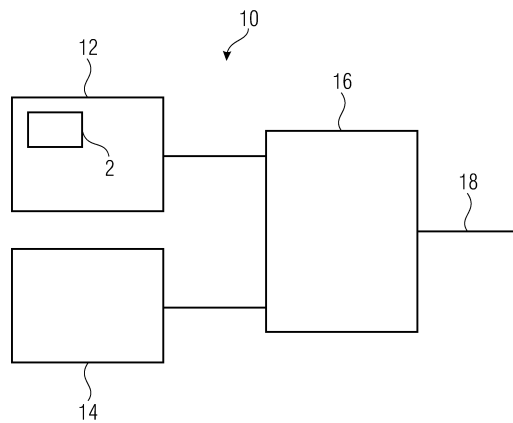


FIG 1

【図2】

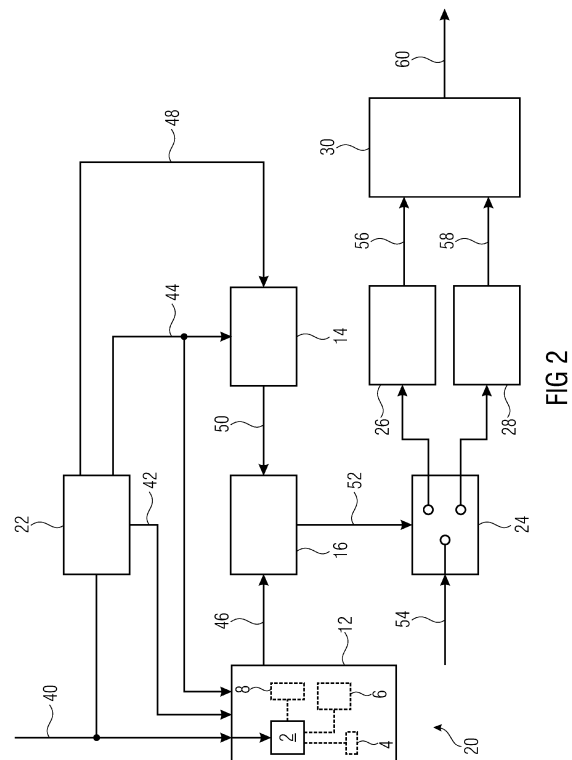
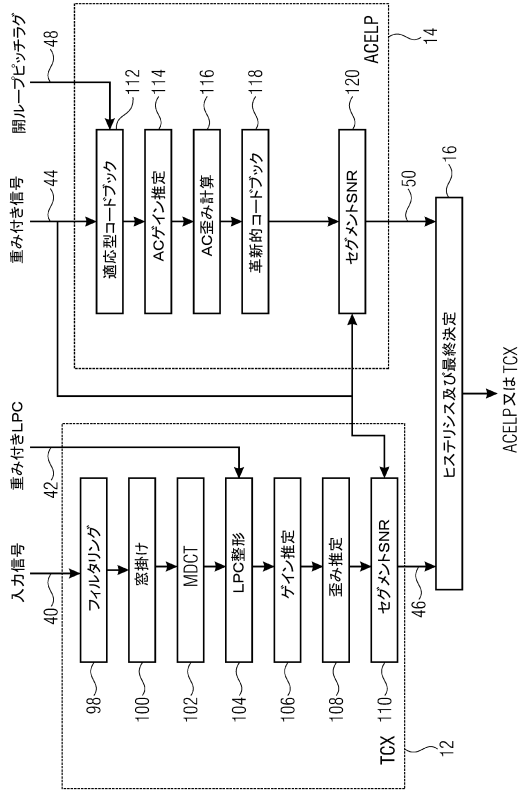


FIG 2

【図 3】



【図 4 a】

$$SNR = 10 \log_{10} \frac{\sum_{i=1}^N x^2(i)}{\sum_{i=1}^N (x(i) - y(i))^2}$$

FIG 4A

【図 4 b】

$$SNR_{seg} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \sum_{i=N_m}^{N_m+N-1} \left( \frac{\sum_{i=1}^N x^2(i)}{\sum_{i=1}^N (x(i) - y(i))^2} \right)$$

FIG 4B

---

フロントページの続き

(72)発明者 デーラ, ステファン

ドイツ連邦共和国 9 1 0 5 8 エルランゲン ザイデルシュタイク 6 1

(72)発明者 グリル, ベルンハルト

ドイツ連邦共和国 9 1 2 0 7 ラウフ ペーテル - ヘンライン - ストラーセ 7

(72)発明者 ヤンデル, マニユエル

ドイツ連邦共和国 9 1 0 5 2 エルランゲン リービックストラーセ 2

審査官 上田 雄

(56)参考文献 特表 2 0 1 3 - 5 3 1 8 2 0 ( J P , A )

特表 2 0 1 0 - 5 3 0 0 8 4 ( J P , A )

国際公開第 2 0 1 4 / 1 1 8 1 3 6 ( W O , A 1 )

特表 2 0 1 4 - 5 1 0 3 0 3 ( J P , A )

(58)調査した分野(Int.Cl., D B 名)

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

H 0 3 M 7 / 3 0