

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4851578号

(P4851578)

(45) 発行日 平成24年1月11日(2012.1.11)

(24) 登録日 平成23年10月28日(2011.10.28)

(51) Int.Cl. F I
G 1 O L 19/14 (2006.01) G 1 O L 19/14 4 O O D
G 1 O L 19/00 (2006.01) G 1 O L 19/00 4 O O A

請求項の数 7 (全 20 頁)

(21) 出願番号	特願2009-262773 (P2009-262773)	(73) 特許権者	595020643
(22) 出願日	平成21年11月18日(2009.11.18)		クァアルコム・インコーポレイテッド
(62) 分割の表示	特願2004-219254 (P2004-219254)		Q U A L C O M M I N C O R P O R A T E D
原出願日	平成7年8月1日(1995.8.1)		アメリカ合衆国、カリフォルニア州 9 2
(65) 公開番号	特開2010-44421 (P2010-44421A)		1 2 1 - 1 7 1 4、サン・ディエゴ、モア
(43) 公開日	平成22年2月25日(2010.2.25)		ハウス・ドライブ 5 7 7 5
審査請求日	平成21年11月18日(2009.11.18)	(74) 代理人	100108855
(31) 優先権主張番号	286,842		弁理士 蔵田 昌俊
(32) 優先日	平成6年8月5日(1994.8.5)	(74) 代理人	100091351
(33) 優先権主張国	米国 (US)		弁理士 河野 哲
		(74) 代理人	100088683
			弁理士 中村 誠
		(74) 代理人	100109830
			弁理士 福原 淑弘

最終頁に続く

(54) 【発明の名称】 減少レート、可変レートの音声分析合成を実行する方法及び装置

(57) 【特許請求の範囲】

【請求項 1】

可変レート C E L P 符号器と通信するように結合された目標整合信号の雑音に対する比率 (T M S N R) 要素からの情報により決定される音声フレームを如何に適切に音声モデルがトラッキングするかに基づく C E L P 符号化アルゴリズムを用いて音声フレームを符号化する可変レート C E L P 符号器の平均データレートを、T M S N R 要素の出力に関する閾値を動的に変化させることによって、調整する方法であって、前記方法は

目標のレートを達成するために特定のレートで C E L P 符号化されるべき音声フレーム数を変更するように前記 T M S N R 要素の出力に関する前記閾値を増加させるまたは減少させることを含み、

前記 T M S N R 要素の前記出力が前記増加された閾値を超えない場合には、前記音声フレームの前記平均データレートは前記可変レート C E L P 符号器によって増加させられ、

前記 T M S N R 要素の前記出力が前記減少された閾値を超える場合には、前記音声フレームの前記平均データレートが前記可変レート C E L P 符号器によって減少させられる、方法。

【請求項 2】

前記音声フレームの前記平均データレートを増加させるため、2 分の 1 レートではなくフルレートで C E L P 符号化するのに必要な音声フレームの数を推定することをさらに含む請求項 1 記載の方法。

【請求項 3】

前記音声フレームの数の推定は、前記TMSNR要素の可能な出力値と記録された前記閾値の現在の値と間の複数の差分を含むヒストグラムを使用し、ここで複数の差分は2分の1レートにおいてCELP符号化するのにどれだけ多くの音声フレームが必要かを決定するために使用される請求項2記載の方法。

【請求項4】

前記音声フレームの前記平均データレートを減少させるため、フルレートではなく2分の1レートでCELP符号化するのに必要とする音声フレームの数を推定することをさらに含む請求項1記載の方法。

【請求項5】

前記音声フレームの数の推定は、前記TMSNR要素の可能な出力値と記録された前記閾値の現在の値と間の複数の差分を含むヒストグラムを使用し、なお複数の差分はフルレートでCELP符号化するのにいかに多くの音声フレームが必要かを決定するのに使用される請求項4記載の方法。

【請求項6】

可変レートCELP符号器と通信するように結合された目標整合信号の雑音に対する比率(TMSNR)要素からの情報により決定される音声フレームを如何に適切に音声モデルがトラッキングするかに基づいてCELP符号化アルゴリズムを用いて音声フレームを符号化する可変レートCELP符号器の平均データレートを、TMSNR要素の出力に関する閾値を動的に変化させることによって調整する装置であって、前記装置は

目標のレートを達成するために特定のレートでCELP符号化されるべき音声フレーム数を変更するように前記TMSNR要素の出力に関する前記閾値を増加させるまたは減少させるレート決定論理を含み、

前記TMSNR要素の前記出力が前記閾値を超えない場合には、前記音声フレームの前記平均データレートは前記可変レートCELP符号器によって増加させられ、

前記TMSNR要素の前記出力が前記閾値を超える場合には、前記音声フレームの前記平均データレートが前記可変レートCELP符号器によって減少させられる、装置。

【請求項7】

可変レートCELP符号器と通信するように結合された目標整合信号の雑音に対する比率(TMSNR)要素からの情報により決定される音声フレームを如何に適切に音声モデルがトラッキングするかに基づいてCELP符号化アルゴリズムを用いて音声フレームを符号化する可変レートCELP符号器の平均データレートを、TMSNR要素の出力に関する閾値を動的に変化させることによって調整する装置であって、前記装置は

目標のレートを達成するために特定のレートでCELP符号化されるべき音声フレーム数を変更するように前記TMSNR要素の出力に関する前記閾値を増加させるまたは減少させる手段を備え、

前記TMSNR要素の前記出力が前記増加された閾値を超えない場合には、前記音声フレームの前記平均データレートは前記可変レートCELP符号器によって増加させられ、

前記TMSNR要素の前記出力が前記減少された閾値を超える場合には、前記音声フレームの前記平均データレートが前記可変レートCELP符号器によって減少させられる、装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、通信に関する。具体的には、本発明は、斬新で且つ改良された線形予測(CELP)符号化によって駆動された可変レート符号を実行する方法及び装置に関する。

【背景技術】

【0002】

デジタル技術による音声の伝送は、一般に普及しつつあり、特に、遠距離及びデジタル無線電話分野に普及している。これは、言い替えれば、チャネルを通じて送られてくる再構築された音声の感知される品質が保たれる最小限の情報量を決定するのに関心が持たれ

10

20

30

40

50

ているということである。

【0003】

もし、音声は、ただ単にサンプリング及びデジタル化によって伝送される場合には、毎秒64キロビット(kbps)程度のデータレートが、通常のアナログ電話の音声品質を達成するために必要とされる。しかしながら、音声解析の使用を通し、次に適切な符号化を行ない、伝送し、そして受信器において再合成することにより、データレートにおいて重要な低減を達成することができる。

【0004】

音声を人間の音声生成のモデルに関連する抽出パラメータによって圧縮する技術を有する装置は、一般的にボコーダと呼ばれている。このような装置は、入ってくる音声を適切なパラメータを抽出するために解析する符号器と、伝送チャンネルを介して受信するパラメータを使用することにより音声を再合成する復号器とから構成されている。正確であるためには、このモデルは常に変化していなければならない。このようなことから音声は、パラメータが計算されている間、時間のブロック或いは解析フレームに分割される。このパラメータは、次に、それぞれの新しいフレームのために更新される。

10

【0005】

符号駆動線形予測符号化(CELP)、確率的符号化或いはベクトル駆動音声符号化は、種々の種類の音声符号器のうちの1つである。この特殊な種類の符号化アルゴリズムの例は、Thomas E. Tremain等による1988年の移動衛星会議の会報の“4.8 kbps 符号駆動線形予測符号器”の書類において述べられている。

20

【0006】

ボコーダの機能は、デジタル化された音声信号を圧縮して、音声における本来の自然な冗長性の全てを除去することにより、低ビットレート信号にすることにある。一般的に、音声は、主に音響管のフィルタリング作用による短時間の冗長性及び、音声符号による音響管の励振による長期間の冗長性を有している。

【0007】

CELP符号器においては、これらの作用は、短期間ホルマントフィルタ及び長期間ピッチフィルタの2つのフィルタによってモデル化されている。

【0008】

一度、これらの冗長性が取り除かれると、結果残余信号は白色ガウス雑音のようにモデル化され、また、符号化されなければならない。この技術の基礎は、人間の音響管モデルを使用した音声波形の短期間の予測を行なうLPCフィルタと呼ばれるフィルタのパラメータを計算することにある。

30

【0009】

加えて、音声のピッチに関連する長期間効果は、ピッチフィルタのパラメータの計算によってモデル化され、これは、本質的に人間の声帯を形に表わしている。

【0010】

最後に、これらのフィルタは駆動される。この駆動は、波形が前に述べた2つのフィルタを駆動した時に、本来の音声に最も近いコードブックの結果における雑音駆動波形のうちの1つを決定することにより行なわれる。

40

【0011】

このようなことから、転送パラメータは、(1) LPCフィルタ、(2) ピッチフィルタ及び(3) コードブック駆動の3つのパラメータに関連する。

【0012】

音声分析合成技術のさらなる目的は、再構築される音声の品質を保ちながらチャンネルを通じて送られてくる情報量を低減することを試みることにあるが、さらに低減を達成するためには他の技術が必要とされる。

【0013】

情報送信量の低減に使用される以前の1つの技術は、音声活性ゲート動作である。この技術においては、音声の休止中の間、情報は伝送されない。この技術では、目的のデータ

50

低減結果を達成することができるが、いくつかの欠陥に煩わされる。

【0014】

多くのケースでは、単語の最初の部分の振幅制限によって、音声品質が低減される。非活性の間にチャンネルをOFFにするゲート動作の他の問題は、システムのユーザが、通常、音声に付随する背景雑音及びチャンネルの品質レートが普通の電話呼び出しに比べて低くなってしまうことを感知してしまうことである。ゲート動作のさらなる問題は、背景において、時々発生する雑音が、音声が発生されていない時に送信器を作動してしまう場合があり、その結果、受信器において厄介な雑音のバーストとなる。

【0015】

音声活性ゲートシステムにおいて合成された音声の品質を改善するために、合成された心地よい雑音が解読処理の間に加えられる。快適な雑音を加えることにより、品質についていくつかの改良が達成されるが、このことは、快適な雑音が、符号器において実際の背景雑音をモデルとしていないことから全体の品質についての大幅な改良ではない。

【0016】

結果的に送信される必要のある情報を低減することに関するデータ圧縮を実現する好ましい技術は、可変レート音声分析合成を実施することである。音声は、本来、沈黙期間、すなわち、休止期間を含んでいるので、これらの期間を表わすために必要とされるデータ量は減らすことができる。

【0017】

可変レート音声分析合成は、この事実をこれらの沈黙期間のためのデータレートの低減によって、最も効果的に活用する。

【0018】

データ伝送における完全な停止とは対比的に、沈黙期間のデータレートにおける低減は、伝送された情報の低減を促進している間に音声活性ゲート動作に関連する問題を改善する。

【0019】

ここに、参照のために引用され、本発明の譲受人に譲渡され、1993年1月14日出願された米国特許出願第08/04,484号(1995年5月9日発行、米国特許第5,414,796号)明細書の“可変レートボコーダ”に、ここで述べた種類の音声符号器の音声分析合成アルゴリズム、符号駆動線形予測音声符号化(CELP)、確率的符号化或いはベクトル駆動音声符号化の詳細が述べられている。

【0020】

このCELP技術は、それ自身が、ある意味で音声を表わすために必要とされるデータ量の効果的な低減を提供し、結果的に高品質の音声となる再合成を行なう。前に述べたボコーダのパラメータは、それぞれのフレームにおいて更新される。係属中の特許出願に詳しく述べられているこのボコーダは、周波数変化による可変出力データレート及びモデルパラメータの精度を提供する。

【0021】

上述の特許出願の音声分析合成アルゴリズムは、音声の活性を基にした可変出力データレートの生成による従来のCELP技術と全く異なっている。この構成においては、音声の休止期間中に、パラメータが度々より少なく或いは低い精度で更新されるために定義される。この技術は、伝送されるべき情報量を大幅に低減することさえも可能にする。このデータレートを低減するために活用される現象は、音声活性要素であり、この音声活性要素は、会話の間中、話者が実際に話をしていることにより与えられる時間の平均レートである。典型的な、双方向の電話通話の平均データレートは、2倍以上低減される。音声における休止の間中、背景雑音のみがボコーダによって符号化されている。このような時においては、人間の音声管モデルに関連するいくつかのパラメータは、伝送される必要がない。

【0022】

前に述べた従来の沈黙の間の伝送された情報量を制限することの取り組みは、音声活性

10

20

30

40

50

ゲート動作と呼ばれており、この技術においては、沈黙の瞬間の間には、情報は伝送されない。

【0023】

受信器側においては、この期間は合成された“快適雑音”で満たされている。反対に、可変レートボコーダは、連続的にデータを送信しており、係属している出願の例示的な実施の形態における可変レートボコーダのレートの範囲は、ほぼ8kbp/sと1kbp/sとの間である。データの連続伝送を行なうボコーダは、合成された“快適な雑音”の必要性を背景雑音の符号化とともに除去し、より自然な品質を合成された音声に提供する。従って、前に述べた特許出願の発明は、合成された音声品質における効果的な改良を提供し、これは、音声と背景との間の円滑な遷移を可能にすることによる音声活性ゲート動作である。

10

【0024】

上述の特許出願の音声分析合成アルゴリズムは、音声における小休止を検出することが可能であり、その結果、有効な音声活性要素の減少を認識することができる。レート決定は、ハングオーバーのないフレーム毎になされ、データレートは、一般的な20msのフレーム継続時間の短さと同様に、音声における休止のために低くされる。従って、このような音節の間の休止が捕らえられる。句の間の長期間の休止だけではなく、短い休止も低いレートで符号化されることができると同様に、この技術は、伝統的に認識されているものにはできない音声活性要素の低減を行なう。

20

【0025】

レート決定は、フレームを基礎として行なわれるので、音声活性ゲート動作システムのように、単語の最初の部分の振幅制限はない。音声の検出とデータの再転送との間の遅れのために、音声活性ゲート動作システムにおいて、この種の振幅制限が起こる。それぞれのフレームを基礎にしたレート決定の使用は、結果的に、全ての遷移が自然な音を有する音声となる。

【0026】

ボコーダは、いつも伝送を行なっているので、話者の周囲の背景雑音は、連続的に受信端で聞こえており、その結果、音声の休止の間、より自然な音をもたらされる。本発明は、このような円滑な遷移に背景雑音を与える。

【0027】

受話者に聞える話をしている間の背景は、音声活性ゲート動作システムにおける休止の間の合成された快適な雑音への突然の変化ではない。背景雑音は、伝送のために常に音声分析合成されているので、背景における興味ある出来事が全く明瞭に送信される。確かなケースにおいては、興味ある背景雑音までも高いレートで符号化される。

30

【0028】

たとえば、誰かが背景において大きな声で話しているとき時、或いは街角に立っているユーザの近くで救急車を運転している場合には、最大レートで符号化が行なわれる。

【0029】

しかしながら、一定の或いはゆっくりした変化の背景雑音は、遅いレートで符号化される。

40

【0030】

可変レート音声分析合成の使用には、符号分割多重接続(CDMA)を基礎としたデジタルセルラー電話システムの容量を2倍以上増加する見込みがある。CDMA及び可変レート音声分析合成は、一義的に合わせられ、CDMAにおいては、チャンネル間の干渉は、いくつかのチャンネルを減少させるデータ伝送レートのように、自動的に低下する。

【0031】

反対に、TDMA或いはFDMA等が考慮されたシステムにおいては、伝送スロットが割り当てられている。このようなシステムを採用することには、データ転送のレートをいくらか低下させることができるという利点があり、外部の発明が必要としない使用していないスロットの他のユーザへの再割り付けの調和のために必要とされる。

50

【0032】

このような方式における本質的な遅れは、長期の音声休止の間にのみチャンネルが再割り付けされることを黙示している。従って、音声活性要素の全ての利点を得ることができない。しかしながら、外部の調和により、可変レート音声分析合成が、他に述べた理由により、システムにおいてはCDMAよりも有用である。

【0033】

CDMAシステムにおける音声品質は、特別なシステムの能力が要求されたときに、時々わずかに低下する。要約していえば、ボコーダは、全てが異なるレートで動作し、異なる音声品質を有する複数のボコーダとして考えられている。

【0034】

その結果、音声品質は、データ転送の平均レートをさらに低減するために混ぜ合わされる。最初の実験は、フルレート及び2分の1のレートで音声分析合成された音声の混合を示しており、たとえば、最大可能データレートは、8 kbpsと4 kbpsとの間を基礎としたフレームによって変化させられ、この結果の音声の品質は、2分の1の可変レート、最大4 kbpsのものよりも良く、全可変レート、最大8 kbpsのものよりは良くない。

【0035】

殆どの電話の会話においては、1人のみが同時に話していることが知られている。レートと連動している全二重電話のために追加の機能が設けられる。もし、リンクの一方の方向が最高の伝送レートで伝送を行なっていると、リンクの他の方向は最低のレートで伝送を行なうことが強制される。リンクの2つの方向の間の連動は、リンクのそれぞれの方向の50%の平均利用より大きくならないように保証される。しかしながら、活性ゲート動作におけるレート連動のケースのように、チャンネルのゲートが閉じられたとき、会話における話者の役割を引き継ぐために、話者を遮る受話者のための方法がない。上述の特許出願の音声分析合成方法は、音声分析合成レートを設定する制御信号によって、容易に適応レートの能力を提供する。

【0036】

上述の特許出願において、ボコーダは、音声が存在するときのフルレート、あるいは音声が存在しないときの8分の1レートのいずれかで動作する。2分の1及び4分の1レートの音声分析合成アルゴリズムの手法は、能力に影響を与える特殊な条件あるいは他のデータが音声データと同時に転送された時のために確保される。

【0037】

ここに参照のために引用され、本発明の譲受人に譲渡され、1993年9月8日出願された係属中の米国特許出願第08/118,473号明細書の“マルチユーザ通信システムにおける伝送データレートを決定する方法及び装置”に、ここで述べた可変レートボコーダによる符号化されたフレームの平均データレートを制限するシステム能力測定に従った通信システムによる方法が述べられている。

【0038】

この装置は、低いレート、すなわち、2分の1のレートで符号化されるべきフルレートのフレームの一連の列において所定のフレームを強制することにより平均データレートを低減する。

【0039】

このような方法によって、活性音声フレームのための符号化レートを低減するときの問題は、制限が入力音声のどの特徴にも一致せず、そして音声圧縮の品質が最適化されないということである。

【0040】

ここに、参照のために引用され、本発明の譲受人に譲渡され、現在は、1994年8月23日に発行された米国特許番号第5,341,456であり、1992年12月2日出願された係属中の米国特許出願第07/984,602号明細書の“可変レートボコーダにおける音声符号化レートの決定方法”に、有声音から無声音を識別するための方法が

10

20

30

40

50

述べられている。

【0041】

この方法には、音声エネルギーの試験及び音声のスペクトルピッチ及び背景雑音から無声音を識別するためのスペクトルピッチの使用が開示されている。

【0042】

入力音声の音声活性に完全に基づいて符号化レートを変化する可変レートボコーダは、活性音声の間中、動的に変化する複雑性或いは情報内容に基づく符号化レートを変化する可変レート符号器の圧縮効率を認識することができない。

【0043】

入力波形の複雑性のために、符号化レートを整合させることにより、より効率的な音声符号器を設計することができる。さらに、可変レートボコーダの出力データレートを動的に調整することに努めるシステムが、望むべき平均データレートのための最適な音声品質を得るために、入力音声の特徴に従ってデータレートを変化する。

【発明の概要】

【0044】

本発明は、所定の最大レートと所定の最小レートとの間のレートで符号化された音声フレームにより低減されたデータレートによって、活性音声フレームを符号化する新規かつ改良された方法及び装置である。

【0045】

本発明は、活性音声動作モードの組を示す。本発明の例示的な実施の形態においては、4つの活性音声動作モード、フルレート音声、2分の1レート音声、無声音4分の1レート及び有声音4分の1レートがある。

【0046】

本発明の目的は、入力音声の符号化効率レートを提供する符号化モードを選択するための最適化された方法を提供することにある。

【0047】

本発明の第2の目的は、この動作モード選択に適した理想的なパラメータの組を認識し、このパラメータの組を生成する手段を提供することにある。本発明の第3の目的は、品質に関して最小限の犠牲の低レート符号化を可能にする2つの別々の状態の認識を提供することにある。この2つの状態は、無声音の存在及び時間的にマスクされた音声の存在である。本発明の第4の目的は、音声品質については、最小限の影響で音声符号器の平均出力データレートの動的調整を行うための方法を提供することにある。

【0048】

本発明は、モード測定に関連するレート決定基準の組を提供する。第1のモード測定は、前の符号化フレームにおける目標整合信号と雑音信号とのレート（TMSNR）であり、これは、どのようにしたら良く合成された音声が入力音声に整合するのかの情報、言い換えれば、どのようにしてうまく符号化モデルを実行するのかの情報を提供する。

【0049】

第2のモード測定は、正規化自己相関機能（NACF）であり、これは音声フレームの周期性を測定する。第3のモード測定は、零交差（ZC）パラメータであり、これは、入力音声フレームにおける高周波の内容を測定する計算的に安価な方法である。第4のモード測定は、LPCモデルがその予測効率を保っているか否かを決定する予測利得差分（PGD）である。第5の測定は、現在のフレームのエネルギーと平均のフレームエネルギーとを比較するエネルギー差分（ED）である。

【0050】

本発明の例示的な実施の形態の音声分析合成アルゴリズムは、活性音声フレームの符号化モードを選択するための上に列挙された5つのモード測定を使用する。本発明のレート決定要素は、音声が無声音4分の1レートで符号化されるべきか否かを決定するために、第1の閾値に対するNACFと第2の閾値に対するZCとを比較する。

【0051】

もし、活性音声フレームが有声音フレームを含むと決定された場合には、ボコーダは、音声フレームが4分の1の有声音レートで符号化されるべきか否かを決定するために、パラメータEDを調べる。もし、音声フレームが4分の1レートで符号化されないと決定された場合には、次に、ボコーダは、音声フレームが2分の1のレートで符号化されるか否かをテストする。ボコーダは、音声フレームが2分の1のレートで符号化されるか否かを決定するために、TMSNR、PGD及びNA CFの値をテストする。もし、活性音声フレームが4分の1或いは2分の1レートで符号化されないと決定された場合には、フレームは、フルレートで符号化される。

【0052】

さらなる目的は、レート要求に適応させるために閾値を動的に変化させる方法を提供することにある。1つ又はそれ以上のモード選択閾値を変化させることにより、平均伝送データレートを増加或いは減少させることが可能になる。閾値を動的に調整することにより、出力レートが調整されることができる。

【0053】

本発明の特徴、目的及び利点は、図面と関連して理解される以下に述べる詳細な説明によって明らかになり、この詳細な説明において全体にわたって、それに対応する基準の特徴が認識される。

【図面の簡単な説明】

【0054】

【図1】本発明の符号化レート決定装置のブロックダイアグラムを示す図である。

【図2】レート決定論理の符号化レート選択プロセスを示すフローチャートである。

【発明を実施するための形態】

【0055】

例示的な実施の形態においては、160の音声サンプルの音声フレームが符号化される。本発明の例示的な実施の形態においては、符号化は4つのデータレート、フルレート、2分の1レート、4分の1レート及び8分の1レートで行なわれる。

【0056】

フルレートは、14.4 Kbps レートの出力データに対応する。2分の1レートは、7.2 Kbps レートの出力データに対応する。4分の1レートは、3.6 Kbps レートの出力データに対応する。8分の1レートは、1.8 Kbps レートの出力データに対応し、沈黙の期間の間の伝送のために確保されている。

【0057】

注目すべきことは、本発明は活性音声フレームの符号化にのみ関連することであり、このフレームは、活性音声フレームの中の現在の音声を得るために検出される。

【0058】

音声の現状を検出する方法については、前に述べた米国特許出願第08/004,484号（米国特許第5,414,796号）及び第07/984,602号（米国特許第5,341,456号）明細書に詳しく述べられている。

【0059】

図1を参照すると、モード測定要素12が活性音声フレームのための符号化レートを選択するレート決定論理14によって使用される5つのパラメータの値を決定する。

【0060】

例示的な実施の形態においては、モード測定要素12は5つのパラメータを決定し、この5つのパラメータをレート決定論理14に供給する。

【0061】

レート決定論理14は、モード測定要素12から供給されたパラメータに基づいて、フルレート、2分の1レート、或いは4分の1レートの符号化レートを選択する。

【0062】

レート決定論理14は、生成された5つのパラメータに従って、4つの符号化モードのうち、1つを選択する。この4つの符号化モードは、フルレートモード、2分の1レート

10

20

30

40

50

モード、4分の1の無声音レートモード及び4分の1の有声音レートモードを含んでいる。

【0063】

4分の1の有声音レートモード及び4分の1の無声音レートモードは、同じレートでデータを供給するが、これは異なる符号化方法によって行なわれる。

【0064】

2分の1レートモードは、定常的、周期的に十分にモデル化された音声を符号化するのに使用される。4分の1の無声音レート、4分の1の有声音レートの双方及び2分の1レートモードは、高い精度が要求されない音声の部分におけるフレームの符号化に利用される。

10

【0065】

4分の1の無声音レートモードは、声を発していない音声の符号化に使用される。4分の1の有声音レートモードは、時間的にマスクされた音声フレームの符号化に使用される。

【0066】

殆どのCELP音声符号器は、同時マスキングを利用しており、この同時マスキングにおいては、ある周波数の音声エネルギーが、同一の周波数及び聞き取れないような雑音の時間において外の雑音エネルギーをマスクしている。

【0067】

可変レート音声符号器は、時間的マスキングを利用することができ、この時間的マスキングにおいては、低エネルギーのアクティブ音声フレームが先行する類似する周波数内容の高エネルギーの音声フレームによってマスクされる。

20

【0068】

何故ならば、人間の耳は、種々の周波数帯域のエネルギーを時の経過とともに取り込み、低エネルギーのフレームは、低エネルギーのフレームの符号化の必要性を下げるために時間平均がとられるからである。

【0069】

この聴覚の複数の現象の時間的マスキングを利用することにより、可変レート音声符号器はこのモードにおける音声の間、符号化レートを低減することが可能になる。

【0070】

この精神聴覚学的現象は、E. Zwicker及びH. Fastlによる精神聴覚学のpp. 56-101. に詳しく述べられている。

30

【0071】

モード測定要素12は、4つの入力信号を受信し、5つのモードパラメータを生成する。モード測定要素12が受信する最初の信号は、 $S(n)$ であり、この $S(n)$ は、符号化されていない音声サンプルである。

【0072】

例示的な実施の形態においては、この音声サンプルは、160の音声サンプルを有するフレームから供給される。

【0073】

モード測定要素12に供給される音声フレームは、全てアクティブな音声を含んでいる。沈黙期間の間、本発明のアクティブ音声レート決定システムは、非活動状態にある。

40

【0074】

モード測定要素12が受信する2つめの信号は、合成音声信号 $S'(n)$ であって、この合成音声信号 $S'(n)$ は、可変レートCELP符号器の符号器の復号器からの解読された音声である。

【0075】

符号器の復号器は、CELP符号器を基にした合成による分析により、フィルタのパラメータとメモリとを更新する目的のために、符号化された音声のフレームを解読する。

【0076】

50

このような復号器の設計は、良く知られている技術であり、前に述べた米国特許出願第 08/004,484 号（米国特許第 5,414,796 号）明細書に詳しく述べられている。

【0077】

モード測定要素 12 が受信する 3 つめの信号は、ホルマント残余信号 $e(n)$ である。このホルマント残余信号は、CELP 符号器の線形予測符号化 (LPC) フィルタによってフィルタリングされた音声信号 $S(n)$ である。

【0078】

LPC フィルタの設計及びこのようなフィルタによる信号のフィルタリングは、良く知られた技術であり、前に述べた米国特許出願第 08/004,484 号（米国特許第 5,414,796 号）明細書に詳しく述べられている。

10

【0079】

モード測定要素 12 が受信する 4 つめの信号は、 $A(z)$ であり、この $A(z)$ は、CELP 符号器と関連した聴感重み付けフィルタのフィルタタップ値である。

【0080】

このタップ値の生成、及び聴感重み付けフィルタのフィルタリング動作は、良く知られた技術であり、前に述べた米国特許出願第 08/004,484 号（米国特許第 5,414,796 号）明細書に詳しく述べられている。

【0081】

雑音レートのためのターゲットマッチング整合信号 (SNR) 演算要素 2 は、合成された音声信号 $S'(n)$ 、音声サンプル $S(n)$ 、及び 1 組の聴感重み付けフィルタのタップ値 $A(z)$ を受信する。

20

【0082】

ターゲットマッチング SNR 演算要素 2 は、TMSNR で示されるパラメータを供給し、この TMSNR はどのようにしたらよく音声モデルが入力音声をトラッキングするかを示している。

【0083】

ターゲットマッチング SNR 演算要素 2 は、下記の (1) 式と一致する TMSNR を生成する。

【数 1】

30

$$\text{TMSNR} = 10 \cdot \log \left[\frac{\sum_{n=0}^{159} \hat{s}_w^2(n)}{\sum_{n=0}^{159} (s_w(n) - \hat{s}_w(n))^2} \right] \quad (1)$$

40

【0084】

ここで、添え字 W は、聴感重み付けフィルタによってフィルタリングされた信号を示している。

【0085】

ここで、注意すべきことは、この測定は、NACF, PGD, ED, ZC が現在の音声のフレームにおいて計算されている間に、前の音声のフレームのために計算されることである。

【0086】

TMSNR は、選択された符号化レートの機能により前の音声のフレームにおいて計算され、そして、複雑な計算であることから、符号化されたフレームの前のフレームにおい

50

て計算される。

【0087】

この聴感重み付けフィルタの設計及び実現は、良く知られた技術であり、前に述べた米国特許出願第08／004, 484号（米国特許第5, 414, 796号）明細書に詳しく述べられている。また、この聴感重み付けは、音声フレームの聴感的に重要な特徴の重み付けに適していることに注目すべきである。しかしながら、この測定は、信号の聴感的重み付けをすること無しに、測定が行なわれることをイメージしている。

【0088】

正規化自己相関演算要素4は、ホルマント残余信号、 $e(n)$ を受信する。この正規化自己相関演算要素4は、音声フレームにおけるサンプル周期の指示を供給するためのものである。

10

【0089】

正規化自己相関演算要素4は、下記の（2）式に従ってNACFで示されるパラメータを生成する。

【数2】

$$NACF = \max_{T \in [20, 120]} \frac{\sum_{n=0}^{159} e(n) \cdot e(n-T)}{\sum_{n=0}^{159} e^2(n)} \quad (2)$$

20

【0090】

ここで注意すべきことは、このパラメータの生成には、前のフレームの符号化からのホルマント残余信号のメモリが必要であることに留意すべきである。

【0091】

このことは、現在のフレームの周期だけではなく、前のフレームとともに現在のフレームの周期のテストを行なうことを可能にする。

【0092】

30

その理由は、最適な実施の形態においては、ホルマント残余信号、 $e(n)$ が音声サンプル、 $S(n)$ の代わりに使用されており、このNACFを生成するのに使用されるホルマント残余信号 $e(n)$ は、音声信号のホルマントの干渉を取り除くものである。

【0093】

ホルマントフィルタを通過する音声信号は、音声エンベロープを平滑化するのに役に立ち、故に、結果信号が白色化される。

【0094】

ここで、注意すべきことは、例示的实施例における遅れ T の値は、毎秒8000サンプルのサンプリング周波数のための66Hzと400Hzとの間の周波数のピッチに対応する。

40

【0095】

この遅れ値 T によって与えられるピッチ周波数は、下記の（3）式によって計算される。

【0096】

$$f_{pitch} = f_s / T \quad (3)$$

（但し、 f_s はサンプリング周波数）

ここで、注意すべきことは、周波数範囲は、1組の異なる遅れ値を単に選択することによって、拡大あるいは縮小される。

【0097】

さらに、ここで注意すべきことは、本発明は、どんなサンプリング周波数にも等しく適

50

用することができるということである。

【0098】

零交差カウンタ6は、音声サンプル $S(n)$ を受信し、音声サンプルの符号の変化の回数をカウントする。これは、音声信号における高周波部分を費用をかけずに計算する方法である。このカウンタは、以下の形のソフトウェアによるループで実現される。

【0099】

$cnt = 0$ (4)

for $n = 0, 158$ (5)

if $(S(n) \cdot S(n+1) < 0) cnt++$ (6)

式4-6のループは連続する音声サンプル同士を掛合わせ、その積が2つの連続したサンプル同士の符号が異なることを示す零以下であるかどうかをテストする。このことによって、音声信号にDC成分がないと推測する。信号からのDC成分をどのように除去するかは良く知られている技術である。

【0100】

予測利得差分要素8は、音声信号 $S(n)$ 及びホルマント残余信号 $e(n)$ を受信する。予測利得差分要素8は、PGDで示されるパラメータを生成し、このPGDはLPCモデルがその予測効率を保っているか否かを決定する。

【0101】

予測利得差分要素8は、下記の式(7)に従って、予測利得、 P_g 、を生成する。

【数3】

$$P_g = \frac{\sum_{n=0}^{159} S^2(n)}{\sum_{n=0}^{159} e^2(n)} \quad (7)$$

【0102】

現在のフレームの予測利得は、次に、下記の式(8)によって出カパラメータPGDが生成されている場合に、前のフレームの予測利得と比較される。

【0103】

$PGD = 10 \cdot \log((P_g(i)) / (P_g(i-1)))$, (8)

(但し、 i はフレーム番号を示す。)

最適な実施の形態においては、予測利得差分要素8は予測利得値 P_g 、を生成しない。ダービンの副産物であるLPC係数の生成は、予測利得 P_g であり、反復演算を必要としないものである。

【0104】

フレームエネルギー差動要素10は、現在のフレームの音声サンプル $s(n)$ を受信し、下記の(9)式に従った現在のフレームにおける音声信号のエネルギーを計算する。

【数 4】

$$E_i = \sum_{n=0}^{159} S^2(n) \quad (9)$$

10

【0105】

この現在のフレームのエネルギーは、前のフレームのエネルギーの平均 E_{ave} と比較される。例示的な実施の形態において、このエネルギーの平均、 E_{ave} は、漏れ積分器の形によって生成される。

【0106】

$$E_{ave} = \alpha \cdot E_{ave} + (1 - \alpha) \cdot E_i, \quad (10)$$

(但し、 $0 < \alpha < 1$)

係数 α は、フレームの範囲を決定し、この係数 α は、計算に関連するものである。例示的な実施の形態において、この α は、8 フレームの時間定数を提供する 0.8825 がセ

20

【0107】

$$ED = 10 \cdot \log(E_i / E_{ave}) \quad (11)$$

この5つのパラメータ、 $TMSNR$ 、 $NACF$ 、 ZC 、 PGD 及び ED は、レート決定論理14に供給される。レート決定論理14は、パラメータ及び予め設定されている選択規則に従って、次のフレームのサンプルのための符号化レートを選択する。今、図2を参照すると、レート決定論理要素14のレート選択手順を示す流れ図が示されている。

【0108】

ブロック18において、レート決定手順が始まる。ブロック20においては、正規化自己相関演算要素4の出力 $NACF$ が予め設定された閾値、 $THR1$ に対して比較され、零交差カウンターの出力が予め設定された第2の閾値、 $THR2$ に対して比較される。

30

【0109】

もし、 $NACF$ が $THR1$ より小さく、且つ ZC が $THR2$ よりも大きい場合には、この流れは無声音4分の1レートとして音声を符号化するブロック22に進む。

【0110】

予め設定された閾値よりも小さい $NACF$ は、音声における周期性の欠如を示しており、予め設定された閾値よりも大きい ZC は、音声における高周波部分を示すものである。

【0111】

これら2つの状態の組み合わせは、フレームが無声音を含んでいることを示している。例示的な実施の形態において、 $THR1$ は 0.35 、 $THR2$ は 50 の零交差である。もし、 $NACF$ が $THR1$ よりも小さく或いは ZC が $THR2$ より大きくない場合には、流れはブロック24に進む。

40

【0112】

ブロック24においては、フレームエネルギー差動要素10の出力、 ED が第3の閾値 $THR3$ と比較される。もし、 ED が $THR3$ よりも小さい場合には、ブロック26において、現在の音声フレームは無声音4分の1レートとして符号化される。

【0113】

もし、現在のフレームの間のエネルギーの差が閾値量よりも大きく平均よりも小さい場合には、時間的にマスクされた音声の状態が示される。例示的な実施の形態においては、

50

THR 3は−14 dBである。もし、EDがTHR 3に到達しない場合には、流れはブロック28に進む。

【0114】

ブロック28においては、ターゲット整合SNR演算要素2の出力であるTMSNRは、第4の閾値THR 4と比較される。予測利得差分要素8の出力PGDは、第5の閾値THR 5と比較され、正規化自己相関演算要素4の出力NACFは、第6の閾値TH 6と比較される。

【0115】

もし、TMSNRがTHR 4を超え、PGDがTHR 5より小さく、NACFがTH 6よりも大きい場合には、流れはブロック30に進み、そして、音声は2分の1レートで符号化される。

10

【0116】

TMSNRがその閾値を上回るとは、モデル及びモデル化されたその音声は前のフレームにおいてマッチングしていたことを示している。パラメータPGDがその予め定められた閾値よりも小さいことは、LPCモデルがその予測効果を保ち続けていることを示している。パラメータNACFがその予め定められた閾値を超えることは、フレームが前の音声フレームに対して周期的である周期的音声を含むことを示している。

【0117】

例示的な実施の形態においては、THR 4は最初に10 dBにセットされ、THR 5は−5 dBにセットされ、THR 6は0.4にセットされる。ブロック28において、もしTMSNRがTHR 4を超えず、或いはPGDがTHR 5を超えず、或いはNACFがTHR 6を超えない場合、流れはブロック32に進み、そして現在の音声フレームがフルレートで符号化される。

20

【0118】

閾値の動的な調整を行なうことにより、任意の全体的なデータレートを達成することができる。この全体的な活性化された音声平均データレートRは、活性化音声フレームの解析窓Wで定義されることができる。

【数5】

$$R = \frac{R_f \cdot \# R_f \text{ frames} + R_h \cdot \# R_h \text{ frames} + R_q \cdot \# R_q \text{ frames}}{W} \quad (12)$$

30

【0119】

ここで、 R_f は、フルレートで符号化されたフレームのデータレート、
 R_h は、2分の1のレートで符号化されたフレームのデータレート、
 R_q は、4分の1のレートで符号化されたフレームのデータレート、
 $W = \# R_f \text{ フレーム} + \# R_h \text{ フレーム} + \# R_q \text{ フレーム}$ 。

【0120】

それぞれの符号化レートとそのようなレートで符号化された多くのフレームとを掛け合わせ、そして、サンプルにおける全ての数のフレームで除算することにより、活性化した音声のサンプルの平均データレートが計算される。”S“の音から引き出されるような無声音の長い持続時間によって平均レート統計値が歪められることを防止するのに十分なほど、フレームのサンプルサイズWを大きくとることが重要である。例示的な実施の形態において、平均レートを計算するためのフレームサンプルサイズWは、400フレームである。

40

【0121】

2分の1のレートで符号化されるべきであったがフルレートで符号化されたフレームの数を増大させることによってこの平均データレートは減少し、逆に、フルレートで符号化されるべきであったが2分の1のレートで符号化されたフレームの数が増大することによ

50

って、この平均データレートは増大する。この好適な実施の形態において、この変化をもたらすために調整される閾値は、T H R 4である。例示的な実施の形態においては、T M S N Rの値のヒストグラムが保存されている。例示的な実施の形態においては、この格納されたT M S N Rの値は、現在のT H R 4の値からデシベルの整数値に量子化される。この種のヒストグラムを保存することにより、前の解析ブロックにおいて、どのくらいの数のフレームがフルレートから2分の1のレートに変化しているかを推定し、このフルレートから2分の1のレートへの変化は、デシベルの整数値によって減少させられるT H R 4である。

【0122】

逆に言えば、どのくらいの数の2分の1のレートで符号化されたフレームがフルレートで符号化されたかの推定がデシベルの整数値によって増加させられる閾値となる。

【0123】

2分の1レートフレームからフルレートフレームへの変化するフレームの数を決定する方程式は、次の式によって決定される。

【数6】

$$\Delta = \frac{[\text{target rate} - \text{average rate}] \cdot W}{R_f - R_h} \quad (13)$$

【0124】

ここで、 Δ は、2分の1のレートで符号化され目標のレートを達成するためにフルレートで符号化されるべきフレームの数であり、

$$W = \# R_f \text{ フレーム} + \# R_h \text{ フレーム} + \# R_q \text{ フレーム}$$

$T M S N R_{N E W} = T M S N R_{O L D} + (\text{上述の (13) 式で定義される } T M S N R_{O L D} \text{ から } \Delta \text{ フレームに到達するまでの } d B \text{ 数の差})$

ここで、注意すべきことは、T M S N Rの初期値は、目標の関数であることが望ましい。R_f = 14.4 k b p s, R_f = 7.2 k b p s, R_f = 3.6 k b p sのシステムにおける目標レート8.7 K b p sの例示的な実施の形態においては、T M S N Rの初期値は10 d Bである。

【0125】

ここで、注意すべきことは、T M S N R値の閾値T H R 4からの距離のための数値への量子化は、2分の1或いは4分の1デシベルのように容易に細かく行なうことができ、或いは1.5或いは2デシベルのように荒く行うこともできる。

【0126】

目標レートのどちらか一方が、レート決定論理要素14のメモリ要素に格納されていることを想定しており、このようなケースにおいては、目標レートは、どちらかの動的に決定されるであろうT H R 4値に従って静的値となるであろう。加えて、この初期目標値では、通信システムがレート命令信号を、システムの現在の記憶容量に基づいて、符号化レート選択装置に送信することを想定している。

【0127】

このレート命令信号は、目標レート或いは平均レートにおける単なる増加或いは減少要求のどちらかを指定することができる。

【0128】

もし、システムが目標レートを指定するものである場合には、このレートは、(12)及び(13)式にしたがってT H R 4値を決定するために使用される。もし、このシステムが、ユーザが高い或いは低い転送レートの転送を行うべきことのみを指定している場合には、レート決定論理要素14は、予め定められた増分によって変化するT H R 4値によって変化され、或いはレートにおいて予め定められた増分増加或いは減少に従って増分変化を計算する。

【0129】

ブロック22及び26は、有声音であることを示す音声サンプル或いは無声音であることを示す音声サンプルに基づいて、音声符号化を行なう方法の違いを示している。

【0130】

この無声音は、摩擦音の形をとる音声及び“f”、“s”、“sh”、“t”及び“z”のような一定の音である。

【0131】

4分の1レートの有声音は、時間的にマスクされた音声であり、周波数成分の近似した相対的に高音量の音声フレームに続く低音量音声フレームである。人間の耳は、高音量のフレームに続く低音量のフレームにおける音声の細かな点は聞くことができないので、4分の1レートによって音声を符号化することによって、ビットを節約することができる。

10

【0132】

無声音の4分の1レート符号化の例示的な実施の形態においては、音声フレームは4つのサブフレームに分割される。

【0133】

4つのサブフレームのそれぞれによって送信されるものは全て利得値G及びLPCフィルタ係数A(Z)である。例示的な実施の形態においては、それぞれのサブフレームの利得を表現するために5ビットが転送される。復号器において、それぞれのサブフレームのためのコードブックの索引はランダムに選択される。このランダムに選択されたコードブックのベクトルは、転送された利得値によって掛け合わされ、そして、合成された無声音を生成するために、LPCフィルタA(Z)を通過する。

20

【0134】

4分の1レートの有声音の符号化は、音声フレームが2つのサブフレームに分割され、そして、CELP符号器がコードブックの索引及び2つのサブフレームのそれぞれのための利得を決定する。この例示的な実施の形態においては、5つのビットがコードブックの索引を示すために割り当てられ、他の5つのビットが対応する利得値を指定するために割り当てられる。例示的な実施の形態において、4分の1レートの有声音の符号化のために使用されるコードブックは、2分の1及びフルレートの符号化のために使用されるコードブックのベクトルの部分組である。例示的な実施の形態においては、7つのビットは、フル及び2分の1のレート符号化モデルにおけるコードブックの索引を指定するために使用される。

30

【0135】

図1においては、ブロックは、設計された機能を実現するための構造ブロック或いはデジタル信号プロセッサ(DSP)或いは特定用途向け集積回路ASICの書き込みプログラムによって実現される機能を表わすブロックである。

【0136】

前に述べた最適な実施の形態の説明は、この分野における当業者に本発明を完成し、或いは使用することを可能にする。これらの実施の形態を種々に改良することは、この分野における当業者にとっては容易であり、この中に定義されている一般的な原理が発明的才能を使用することなく他の実施の形態に適用される。

40

【0137】

そのようなことから、本発明は、ここに示した実施の形態に限定されるものではなく、原理と一貫した最も広い範囲及びここに開示された新規な特徴と調和される。

【0138】

(1) 本願実施例に記載の発明は、所定の符号化レートの組から符号化レートを選択し、そして複数の音声サンプルを含む音声フレームを符号化する装置であって、

前記音声フレームの特徴を示す1組のパラメータを生成するために、前記音声サンプルおよび前記音声サンプルから得られた少なくとも1つの信号に応答する手段と、

前記1組のパラメータを受信し、前記1組のパラメータに対応する前記音声サンプルの音響心理学上の特徴を決定し、そして所定のレート選択規則を用いて前記所定の符号化レ

50

ートの組から符号化レートを選択する手段と、を含む装置を含む。

【0139】

(2) また本願実施例に記載の発明は、所定の符号化レートの組から符号化レートを選択し、そして複数の音声サンプルを含む音声フレームを符号化する装置であって、

前記音声サンプルおよび前記音声サンプルから得られた信号に対応する前記音声のフレームの特徴を示す1組のパラメータを生成するモード測定計算器と、

前記1組のパラメータを受信し、前記1組のパラメータに対応する前記音声サンプルの音響心理学上の特徴を決定し、そして前記所定の符号化レートの組から符号化レートを選択するレート決定論理と、を含む装置を含む。

【0140】

(3) また本願実施例に記載の発明は、遠隔局が中央通信局と通信を行う通信システムにおいて、前記遠隔局から伝送される音声フレームの伝送レートを動的に変化させるサブシステムであって、

前記音声フレームの特徴を示す1組のパラメータを生成するために、前記音声フレームおよび前記音声フレームから得られた信号に応答する手段と、

前記パラメータの組を受信し、前記パラメータの組に対応する音響心理学上の特徴を決定し、レート命令信号に対応する少なくとも1つの閾値を生成するためにレート命令信号を受信し、前記パラメータの組の少なくとも1つのパラメータを前記少なくとも1つの閾値と比較し、そして前記比較に応じて符号化レートを選択する手段と、を含むサブシステムを含む。

【0141】

(4) また本願実施例に記載の発明は、遠隔局が中央通信局と通信を行う通信システムにおいて、前記遠隔局から伝送される音声のフレームの伝送レートを動的に変化させるサブシステムであって、

前記音声サンプルおよび前記音声サンプルから得られた信号に対応する前記音声フレームの特徴を示す1組のパラメータを生成するモード測定計算器と、そして

前記1組のパラメータに対応する前記音声サンプルの音響心理学上の特徴を決定するために前記1組のパラメータを受信し、レート命令信号に対応する少なくとも1つの閾値を生成するためにレート命令信号を受信し、前記1組のパラメータの少なくとも1つのパラメータを前記少なくとも1つの閾値と比較し、そして前記比較に応じて符号化レートを選択するレート決定論理と、を含むサブシステムを含む。

【0142】

(5) また本願実施例に記載の発明は、複数の音声サンプルを含む音声フレームを符号化するために所定の符号化レートの組から符号化レートを選択する方法であって、

前記音声サンプルおよび前記音声サンプルから得られた信号に対応する前記音声フレームの特徴を示す1組のパラメータを生成し、そして

前記1組のパラメータに対応する前記所定の符号化レートの組から符号化レートを選択し、前記1組のパラメータは前記音声サンプルの音響心理学上の特徴を決定するものである方法を含む。

【0143】

(6) また本願実施例に記載の発明は、可変レート符号器と通信するように結合された目標整合信号の雑音に対する比率(TMSNR)要素からの情報により決定されるように、如何に適切に音声モデルが音声フレームをトラッキングするかに基づき、音声フレームを符号化する可変レート符号器の平均データレートを調整する方法であって、この方法は

TMSNR要素の出力が増加された閾値を超えず、そして音声フレームの平均データレートが可変レート符号器によって増加させられない場合には、TMSNR要素の出力に関する閾値を増加させ、

TMSNR要素の出力が減少された閾値を超え、そして音声フレームの平均データレートが可変レート符号器によって減少させられる場合には、TMSNR要素の出力に関する閾値を減少させる、方法を含む。

【0144】

(7) さらに、音声フレームの平均データレートを増加させるため、2分の1レートよりもむしろフルレートにおいて符号化するのに必要な音声フレームの数を推定することをさらに含む(6)の方法を含む。

【0145】

(8) さらに、音声フレームの数の推定は、TMSNR要素の可能な出力値と記録された閾値の現在の値と間の複数の差分を含むヒストグラムを使用し、ここで複数の差分は2分の1レートにおいて符号化するのにどれだけ多くの音声フレームが必要かを決定するために使用される(7)の方法を含む。

【0146】

(9) さらに、音声フレームの平均データレートを減少させるため、フルレートよりもむしろ2分の1レートで符号化するのに必要とする音声フレームの数を推定することをさらに含む(6)の方法を含む。

【0147】

(10) さらに、音声フレームの数の推定は、TMSNR要素の可能な出力値と記録された閾値の現在の値と間の複数の差分を含むヒストグラムを使用し、なお複数の差分はフルレートで符号化するのにいかに多くの音声フレームが必要かを決定するのに使用される(9)の方法を含む。

【0148】

(11) また本願実施例に記載の発明は、フルレートフレーム、2分の1レートフレーム、4分の1レートフレームを具備する所定の組の符号化フレームを有するボコーダについて音声フレームを符号化する方法であって、

前記音声フレームにおける周期性を示す正規化自己相関測定と前記音声フレームの高周波部分の存在を示す零交差数とを決定するために前記音声フレームを評価するステップと、

前記正規化自己相関測定が第1の閾値未満であり、かつ前記零交差数が第2の閾値を超えている場合は、無声音のための4分の1レートのフレームを用いて前記音声フレームを符号化するステップと、

を具備する音声フレームを符号化する方法を含む。

【0149】

(12) さらに、前記音声フレームが4分の1レートの無声音として符号化されていない場合、前記音声フレームのエネルギーとフレームの平均エネルギーとの間のエネルギーの変化を示すフレームエネルギー差分測定を決定するために前記音声フレームを評価するステップと、

前記フレームエネルギー差分測定が第3の閾値未満である場合、4分の1レートの有声音のための所定の形式を用いて前記音声フレームを符号化するステップと、をさらに具備する(11)の方法を含む。

【0150】

(13) さらに、前記音声フレームが4分の1レートの有声音として符号化されていない場合、先の音声フレームとその音声フレームから得られた合成音声の整合度を示す目標整合信号対雑音比測定とホルマントのフレームからフレームへの安定性を示す予測利得差分測定とを決定するために、前記音声フレームを評価するステップと、

前記目標整合信号対雑音比測定が第4のしきい値を超えており、かつ前記予測利得差分測定が所定の第5の閾値未満であり、かつ前記自己相関測定が所定の第6の閾値である場合、2分の1レートのための所定の形式を用いて前記音声フレームを符号化するステップと、をさらに具備する(12)の方法を含む。

【0151】

(14) さらに、前記音声フレームが2分の1レート音声として符号化されていない場合、フルレートのための形式を用いて前記音声フレームを符号化するステップをさらに具備する(13)の方法を含む。

10

20

30

40

50

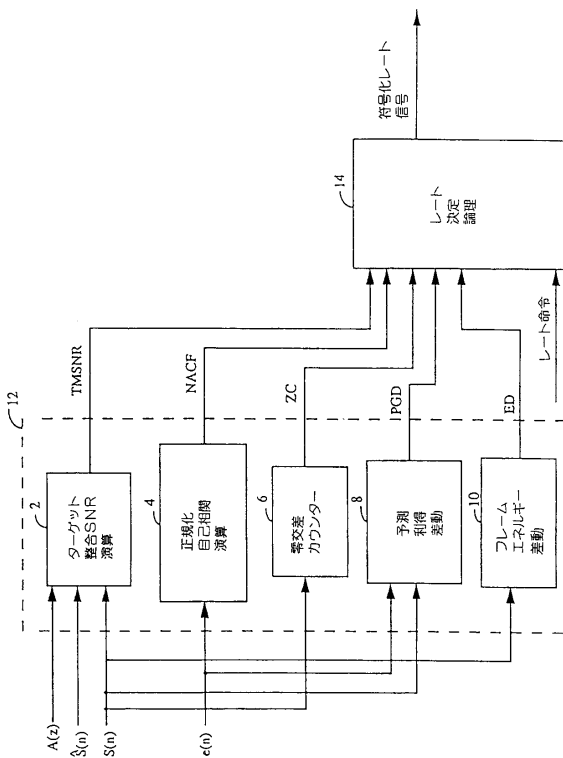
【符号の説明】

【0152】

2…ターゲットマッチング整合信号演算要素、 4…正規化自己相関演算要素、 6…零交差カウンタ、 8…予測利得差分要素、 10…フレームエネルギー差動要素、 12…モード測定要素、 14…レート決定論理要素

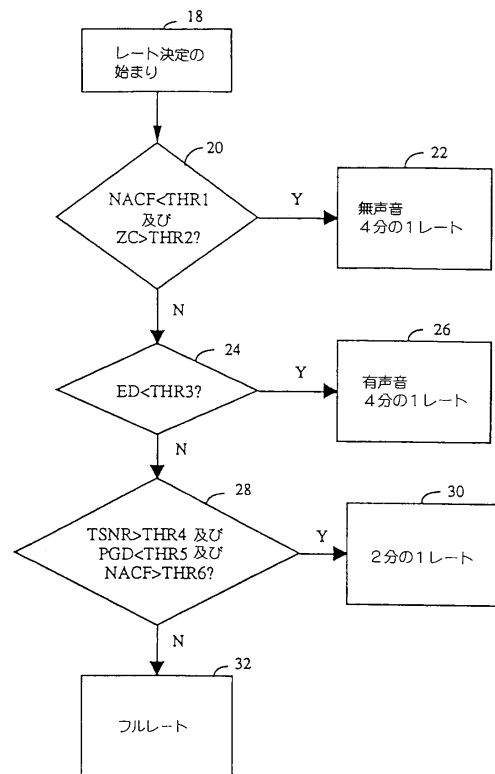
【図1】

図1



【図2】

図2



フロントページの続き

- (74)代理人 100075672
弁理士 峰 隆司
- (74)代理人 100095441
弁理士 白根 俊郎
- (74)代理人 100084618
弁理士 村松 貞男
- (74)代理人 100103034
弁理士 野河 信久
- (74)代理人 100119976
弁理士 幸長 保次郎
- (74)代理人 100153051
弁理士 河野 直樹
- (74)代理人 100140176
弁理士 砂川 克
- (74)代理人 100101812
弁理士 勝村 紘
- (74)代理人 100070437
弁理士 河井 将次
- (74)代理人 100124394
弁理士 佐藤 立志
- (74)代理人 100112807
弁理士 岡田 貴志
- (74)代理人 100111073
弁理士 堀内 美保子
- (74)代理人 100134290
弁理士 竹内 将訓
- (74)代理人 100127144
弁理士 市原 卓三
- (74)代理人 100141933
弁理士 山下 元
- (72)発明者 アンドリュー・ピー・デジャコ
アメリカ合衆国、カリフォルニア州 92126、サン・ディエゴ、フランダース・コーブ 10
424

審査官 山下 剛史

- (56)参考文献 特開平3-181232 (JP, A)
国際公開第92/22891 (WO, A1)
特開平1-233500 (JP, A)
特表平3-505929 (JP, A)

- (58)調査した分野(Int.Cl., DB名)
G10L 19/00-19/14
H04B 14/04
H03M 7/30