(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2011-90311
(P2011-90311A)
(43)公開日　平成23年5月6日(2011.5.6)

(51)Int.Cl.		F I			テーマコード（参考）
G10L 19/14	(2006.01)	G10L	19/14	400B	5J064
G10L 19/12	(2006.01)	G10L	19/12	Z	
G10L 19/02	(2006.01)	G10L	19/02	150	
H03M 7/30	(2006.01)	H03M	7/30	Z	

審査請求 有　請求項の数 1　OL　外国語出願　（全 28 頁）

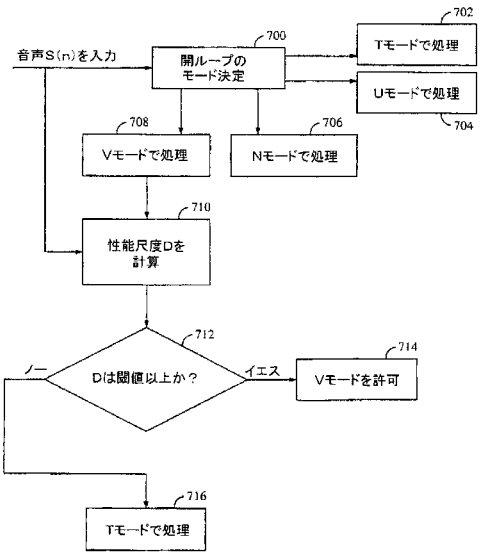
(21)出願番号　特願2010-249991 (P2010-249991)
(22)出願日　平成22年11月8日 (2010.11.8)
(62)分割の表示　特願2001-564148 (P2001-564148)の分割
原出願日　平成12年2月29日 (2000.2.29)

(71)出願人　595020643
クゥアルコム・インコーポレイテッド
ＱＵＡＬＣＯＭＭ　ＩＮＣＯＲＰＯＲＡＴＥＤ
アメリカ合衆国、カリフォルニア州　９２１２１－１７１４、サン・ディエゴ、モアハウス・ドライブ　５７７５
(74)代理人　100108855
弁理士　蔵田　昌俊
(74)代理人　100091351
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】閉ループのマルチモードの混合領域の線形予測音声コーダ

(57)【要約】　（修正有）
【課題】音声を閉ループのマルチモードの混合領域でコード化する方法の提供。
【解決手段】閉ループのマルチモードの混合領域の線形予測（MDLP）の音声コーダは、高レートの時間領域コード化モードと、低レートの周波数領域コード化モードと、入力されたフレームの音声内容に基づいてモードを選択するモード選択機構とを含む。遷移音声のフレームは、高レートのＣＥＬＰモードでコード化される。有声音音声のフレームは、低レートの高調波モードでコード化される。位相パラメータは、準位相モデルによりモデル化される。初期位相値は、周波数領域モードでコード化された直前の音声フレームの初期位相値になり、直前の時間領域でコード化された音声フレーム情報から計算される。周波数領域モードでコード化される各音声フレームを、対応する入力音声フレームと比較して、性能尺度が所定の閾値よりも低いときは、時間領域モードでコード化される。
【選択図】図９



【特許請求の範囲】
【請求項１】
少なくとも１つの時間領域コード化モードおよび少なくとも１つの周波数領域コード化モードをもつコーダと、
コーダに接続され、かつ音声プロセッサによって処理されるフレーム内容に基づいてコーダのコード化モードを選択するように構成されている閉ループのモード選択デバイスとを含むマルチモードの混合領域の音声プロセッサ。
【請求項２】
コーダが、音声フレームをコード化する請求項１記載の音声プロセッサ。
【請求項３】
コーダが、音声フレームの線形予測残余をコード化する請求項１記載の音声プロセッサ。
【請求項４】
少なくとも1つの時間領域コード化モードが、第１のコード化レートでフレームをコード化するコード化モードを含み、少なくとも1つの周波数領域コード化モードが、第２のコード化レートでフレームをコード化するコード化モードを含み、第２のコード化レートが第１のコード化レートよりも低い請求項１記載の音声プロセッサ。
【請求項５】
少なくとも1つの周波数領域コード化モードが、高調波のコード化モードを含む請求項１記載の音声プロセッサ。
【請求項６】
コーダに接続された比較回路であって、コード化されていないフレームを、少なくとも１つの周波数領域コード化モードでコード化されたフレームと比較して、比較に基づいて性能尺度を生成する比較回路をさらに含み、コーダが、性能尺度が所定の閾値よりも低いときだけ、少なくとも１つの時間領域コード化モードを適用し、さもなければコーダは、少なくとも１つの周波数領域コード化モードを適用する請求項１記載の音声プロセッサ。
【請求項７】
コーダが、少なくとも１つの時間領域コード化モードを、少なくとも１つの周波数領域コード化モードでコード化された所定数の連続的に処理されるフレームの直ぐ後の各フレームに適用する請求項１記載の音声プロセッサ。
【請求項８】
少なくとも1つの周波数領域コード化モードが、周波数、位相、および振幅を含む１組のパラメータをもつ複数のシヌソイドで各フレームの短期間のスペクトルを表わし、位相は多項式表現および初期位相値でモデル化されていて、初期位相値が、（１）前フレームが、少なくとも1つの周波数領域コード化モードでコード化されたときは、前フレームの推定された最終位相値であるか、または（２）前フレームが、少なくとも１つの時間領域コード化モードでコード化されたときは、前フレームの短期間のスペクトルから求められる位相値である請求項１記載の音声プロセッサ。
【請求項９】
各フレームにおけるシヌソイドの周波数が、フレームのピッチ周波数の整数倍である請求項８記載の音声プロセッサ。
【請求項１０】
各フレームにおけるシヌソイドの周波数が、０ないし２πの1組の実数から得られる請求項８記載の音声プロセッサ。
【請求項１１】
フレームを処理する方法であって、
開ループのコード化モード選択プロセスを各連続する入力フレームへ適用して、入力フレームの音声内容に基づいて、時間領域コード化モードか、または周波数領域コード化モードの何れかを選択するステップと、
入力フレームの音声内容が定常状態の有声音の音声を示すときは、入力フレームを周波数領域でコード化するステップと、

入力フレームの音声内容が定常状態の有声音の音声以外のものを示すときは、入力フレームを時間領域でコード化するステップと、
周波数領域でコード化されたフレームと入力フレームとを比較して、性能尺度を求めるステップと、
性能尺度が所定の閾値よりも低いときは、入力フレームを時間領域でコード化するステップとを含むフレームを処理する方法。
【請求項１２】
フレームが、線形予測残余フレームである請求項１１記載の方法。
【請求項１３】
フレームが音声フレームである請求項１１記載の方法。
【請求項１４】
時間領域でコード化するステップが、第1のコード化レートでフレームをコード化することを含み、周波数領域でコード化するステップが、第２のコード化レートでフレームをコード化することを含み、第２のコード化レートが第１のコード化レートよりも低い請求項１１記載の方法。
【請求項１５】
周波数領域でコード化するステップが、高調波でコード化することを含む請求項１１記載の方法。
【請求項１６】
周波数領域でコード化するステップが、周波数、位相、および振幅を含む１組のパラメータをもつ複数のシヌソイドで各フレームの短期間のスペクトルを表わし、位相は多項式表現および初期位相値でモデル化されていて、初期位相値が、（１）前フレームが周波数領域でコード化されたときは、前フレームの推定された最終位相値であるか、または（２）前フレームが時間領域でコード化されたときは、前フレームの短期間のスペクトルから求められる位相値である請求項１１記載の方法。
【請求項１７】
各フレームのシヌソイド周波数が、フレームのピッチ周波数の整数倍である請求項１６記載の方法。
【請求項１８】
各フレームのシヌソイド周波数が、０ないし２πの1組の実数から得られる請求項１６記載の方法。
【請求項１９】
マルチモードの混合領域の音声プロセッサであって、
開ループのコード化モード選択プロセスを入力フレームへ適用して、入力フレームの音声内容に基づいて、時間領域コード化モードか、または周波数領域コード化モードの何れかを選択する手段と、
入力フレームの音声内容が定常状態の有声音の音声を示すときは、入力フレームを周波数領域でコード化する手段と、
入力フレームの音声内容が定常状態の有声音の音声以外のものを示すときは、入力フレームを時間領域でコード化する手段と、
周波数領域でコード化されたフレームと入力フレームとを比較して、性能尺度を求める手段と、
性能尺度が所定の閾値よりも低いときは、入力フレームを時間領域でコード化する手段とを含むマルチモードの混合領域の音声プロセッサ。
【請求項２０】
フレームが線形予測残余フレームである請求項１９記載の音声プロセッサ。
【請求項２１】
入力フレームが音声フレームである請求項１９記載の音声プロセッサ。
【請求項２２】
時間領域でコード化する手段が、第1のコード化レートでフレームをコード化する手段を

含み、周波数領域でコード化する手段が、第２のコード化レートでフレームをコード化する手段を含み、第２のコード化レートが第１のコード化レートよりも低い請求項１９記載の音声プロセッサ。
【請求項２３】
周波数領域でコード化する手段が、高調波コーダを含む請求項１９記載の音声プロセッサ。
【請求項２４】
周波数領域でコード化する手段が、周波数、位相、および振幅を含む１組のパラメータをもつ複数のシヌソイドで各フレームの短期間のスペクトルを表わす手段を含み、位相は多項式表現および初期位相値でモデル化されていて、初期位相値が、（１）直前のフレームが周波数領域でコード化されたときは、直前のフレームの推定された最終位相値であるか、または（２）直前のフレームが時間領域でコード化されたときは、直前のフレームの短期間のスペクトルから求められる位相値である請求項１９記載の音声プロセッサ。
【請求項２５】
各フレームのシヌソイド周波数が、フレームのピッチ周波数の整数倍である請求項２４記載の音声プロセッサ。
【請求項２６】
各フレームのシヌソイド周波数が、０ないし２πの1組の実数から得られる請求項２４記載の音声プロセッサ。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、概ね音声処理の分野、とくに音声を閉ループのマルチモードの混合領域でコード化するための方法および装置に関する。
【背景技術】
【０００２】
ディジタル技術による音声（voice）の伝送は、とくに長距離のディジタル無線電話の応用において普及してきた。これにより、チャンネル上で送ることができる最少情報量を判断し、一方で再構成された音声の知覚品質を維持することに関心が生まれた。音声を単にサンプリングして、ディジタル形式にすることによって送るとき、従来のアナログ電話の音声品質を実現するには、毎秒６４キロビット秒（kbps）のオーダのデータレートが必要である。しかしながら、音声解析を使用し、その後で適切にコード化し、伝送し、受信機において再合成をすることによって、データレートを相当に低減することができる。
【０００３】
人間の音声の生成モデルに関係するパラメータを抽出することによって音声を圧縮する技術を採用したデバイスは、音声コーダと呼ばれている。音声コーダは、入力音声信号を時間のブロック、すなわち解析フレームに分割する。一般的に音声コーダはエンコーダとデコーダとを含む。エンコーダは、入力音声フレームを解析して、一定の関連するパラメータを抽出して、パラメータを二値表現、すなわち１組のビットまたは二値データパケットに量子化する。データパケットは通信チャンネル上で受信機およびデコーダへ送られる。デコーダはデータパケットを処理し、非量子化して（unquantize）パラメータを生成し、非量子化したパラメータを使用して音声フレームを再合成する。
【０００４】
音声コーダの機能は、音声が本質的にもっている固有の冗長の全てを取去ることによって、ディジタル化された音声信号を低ビットレートの信号へ圧縮することである。ディジタル圧縮は、入力音声フレームを１組のパラメータで表わし、量子化を採用して、このパラメータを１組のビットで表わすことによって実現される。入力音声フレームが多数のビットN_iをもち、音声コーダによって生成されるデータパケットが多数のビットN_0をもつとき、音声コーダによって得られる圧縮係数は、$C_r = N_i / N_0$である。デコードされた音声(speech)の高い音声品質（voice quality）を維持し、一方で目標の圧縮係数を

得ることが課題とされている。音声コーダの性能は、（１）音声モデル、すなわち上述の解析および合成プロセスの組合せがどのくらい適切に行われるか、および（２）パラメータ量子化プロセスが１フレーム当りN_0ビットの目標ビットレートでどのくらい適切に実行されるかに依存する。したがって音声モデルは、各フレームごとの小さい組のパラメータを使用して、音声信号の本質（essence）、すなわち目標の音声品質を得ることを目的としている。

【０００５】

音声コーダは時間領域のコーダ、すなわち音声の小さいセグメント（一般的に５ミリ秒（millisecond, ms）のサブフレーム）を一度にコード化する高度な時間分解処理（time-resolution processing）を採用することによって時間領域の音声波形を得ることを試みる時間領域のコーダとして構成することができる。各サブフレームごとに、この技術において知られている種々のサーチアルゴリズムによって、コードブック空間から高精度の見本（representative）を見付ける。その代わりに、音声コーダは周波数領域のコーダとして構成されていてもよく、１組のパラメータを使用して入力音声フレームの短期間の音声スペクトルを捕らえて（解析）、対応する合成プロセスを採用して、スペクトルパラメータから音声波形を再現することを試みる。パラメータ量子化器は、文献（A Gersho & R.M. Gray, Vector Quantization and Signal Compression (1992)）に記載されている既知の量子化技術にしたがって、コードベクトルの記憶されている表現を使用してパラメータを表わすことによってそれらのパラメータを保存する。

【０００６】

よく知られている時間領域の音声コーダは、ＣＥＬＰ（Code Excited Linear Predictive）コーダであり、これはL.B. Rabiner & R.W. Schaferによる文献（Digital Processing of Speech Signals 396-453 (1978)）に記載されており、ここでは参考文献として全体的にこれを取り上げている。ＣＥＬＰコーダでは、線形予測（linear prediction, LP）解析によって、短期間のフォルマントフィルタの係数を見付け、音声信号における短期間の相関関係、すなわち冗長を取去る。短期間の予測フィルタを入力音声フレームに適用して、ＬＰの残余信号（residue signal）を生成し、このＬＰの残余信号をさらに長期間の予測フィルタパラメータおよび次の確率コードブックでモデル化して、量子化する。したがってＣＥＬＰのコード化は、時間領域の音声波形をコード化するタスクを、ＬＰの短期間のフィルタ係数をコード化するタスクおよびＬＰの残余をコード化するタスクの別々のタスクへ分ける。時間領域のコード化は、固定レート（すなわち、各フレームごとに、同数のビットN_0を使用するレート）で、または可変レート（すなわち、異なるビットレートが異なるタイプのフレームの内容に対して使用されるレート）で実行することができる。可変レートのコーダは、目標の品質を得るのに適したレベルまでコーデックパラメータをコード化するのに必要なビット量のみを使用することを試みる。例示的な可変レートのＣＥＬＰのコーダは米国特許第5,414,796号に記載されており、これは本発明の譲受人に譲渡され、ここでは参考文献として全体的に取り上げている。

【０００７】

ＣＥＬＰコーダのような時間領域のコーダは、通常は、フレームごとに多数のビットN_0に依存して、時間領域の音声波形の精度を保持する。このようなコーダは、通常はフレーム当りのビット数N_0が比較的に多いとき（例えば、８キロビット秒以上）、優れた音声品質を伝える。しかしながら低ビットレート（４キロビット秒以下）では、時間領域のコーダは、使用可能なビット数が制限されているために、高品質で丈夫な性能を維持しない。低ビットレートではコードブック空間が制限されているので、従来の時間領域のコーダには備えられている波形を整合する能力を取去って、より高レートの市販のアプリケーションにおいてこのようなコーダを実行するのに成功した。

【０００８】

現在、研究に対する関心および活発な商業上の要求が急激に高まり、中程度から低いビットレート（すなわち、２．４ないし４キロビット秒の範囲およびそれ以下）で動作する高品質の音声コーダを発展させた。応用分野には、無線電話通信、衛星通信、インターネ

ット電話通信、種々のマルチメディアおよび音声ストリーミングアプリケーション、音声メール、並びに他の音声保存システムを含む。駆動力については、大きい容量が必要とされ、かつパケットが失われた情況下での丈夫な性能が要求されている。種々の最近の音声のコード化を標準化する努力は、低レートの音声コード化アルゴリズムの研究および発展を推進する別の直接的な駆動力に当てられている。低レートの音声コーダは、許容可能な適用バンド幅ごとに、より多くのチャンネル、すなわちユーザを生成し、低レートの音声コーダを適切なチャンネルコーディングの追加の層と結合して、コーダの全体的なビット予定値（bit budget）の仕様に適合させ、チャンネルを誤った情況のもとでも丈夫な性能を発揮させることができる。

【０００９】

より低いビットレートでコード化するために、音声のスペクトル、すなわち周波数領域でコード化する種々の方法が開発され、この方法では音声信号は、時間にしたがって漸進的に変化するスペクトル（time-varying evolution of spectra）として解析される。例えば、R.J. McAulay & T.F. Quatieriによる文献（Sinusoidal Coding, in Speech Coding and Synthesis ch. 4 (W.B. Kleijin & K.K. Paliwal eds., 1995)参照。スペクトルコーダは、時間にしたがって変化する音声波形を精密にまねるのではなく、１組のスペクトルパラメータを使用して、音声の各入力フレームの短期間の音声スペクトルをモデル化、すなわち予測することを目的とする。スペクトルパラメータはコード化され、音声の出力フレームはデコードされたパラメータを使用して生成される。生成された合成された音声は、元の入力音声波形と整合しないが、同様の知覚品質を与える。この技術においてよく知られている周波数領域コーダの例には、マルチバンド励起コーダ（multiband excitation coder, MBE）、シヌソイド変換コーダ（sinusoidal transform coder, STC）、高調波コーダ（harmonic coder, HC）を含む。このような周波数領域のコーダは、低ビットレートで使用可能な少数のビットで正確に量子化できるコンパクトな組のパラメータをもつ高品質のパラメータモデルを与える。

【００１０】

それにも関わらず、低ビットレートのコード化は、制限されたコード化分解能、すなわち制限されたコードブック空間に重大な制約を加えて、単一のコード化機構の効果を制限し、コーダが、等しい精度の種々の背景条件のもとで、種々のタイプの音声セグメントを表わすことができないようにしている。例えば、従来の低ビットレートの周波数領域のコーダは、音声フレームの位相情報を送らない。その代わりに、位相情報は、ランダムな人工的に生成された初期位相値および線形補間技術（linear interpolation technique）を使用することによって再構成される。例えば、H.Yang、他による文献（Quadratic Phase Interpolation for Voiced Speech Synthesis in the MBE Model, in 29 Electronic Letters 856-57 (May 1993)）参照。位相情報は人工的に生成されるので、シヌソイドの振幅は量子化－非量子化プロセスによって完全に保持されるときでも、周波数領域のコーダによって生成される出力音声は元の入力音声と整合しない（例えば、大半のパルスは同期しない）。したがって、周波数領域のコーダでは、例えば信号対雑音比（signal-to-noise ratio, SNR）または知覚のＳＮＲのような、閉ループの性能尺度（performance measure）を採用することが難しいことが分かった。

【００１１】

開ループのモード決定プロセスに関連して低レートの音声のコード化を行なうために、マルチモードコード化技術が採用された。１つのこのようなマルチモードコード化技術は、Amitava Das、他による文献（Multimode and Variable-Rate Coding of Speech, in Speech Coding and Synthesis ch. 7 (W.B. Kleijin & K.K. Paliwal eds., 1995)）に記載されている。従来のマルチモードコーダは異なるモード、すなわちコード化－デコード化アルゴリズムを、異なるタイプの入力音声フレームへ適用する。各モード、すなわちコード化－デコード化プロセスは、最も効率的なやり方で、例えば、有声音音声、無声音音声、または背景ノイズ（非音声(nonspeech)）のような一定のタイプの音声セグメントを表わすために特化される。外部の開ループのモード決定機構は、入力音声フレームを検査し

て、何れのモードをフレームに適用するかに関して判断する。通常は、開ループのモード決定は、入力フレームから多数のパラメータを抽出して、一定の時間およびスペクトルの特性に関するパラメータを評価して、この評価に対するモード決定に基づくことによって行なわれる。したがってモード決定は、出力音声の抽出状態、すなわち出力音声が音声品質または他の性能尺度に関して入力音声にどのくらい近くなるかを前もって知ることなく行われる。

【００１２】

上述に基づいて、位相情報をより精密に推定する低ビットレートの周波数領域のコーダを用意することが望ましい。マルチモードの混合領域のコーダを用意して、フレームの音声内容に基づいて、一定の音声フレームを時間領域でコード化し、他の音声フレームを周波数領域でコード化することがさらに好都合である。閉ループのコード化モード決定機構にしたがって、一定の音声フレームを時間領域でコード化して、他の音声フレームを周波数領域でコード化することができる混合領域のコーダを用意することが、なおいっそう望ましい。したがって、コーダによって生成される出力音声と、コーダへ入力される元の音声との時間の同期性を保証する、閉ループのマルチモードの混合領域の音声コーダが必要とされている。

【発明の概要】

【００１３】

本発明は、コーダによって生成される出力音声と、コーダへ入力される元の音声との時間の同期性を保証する、閉ループのマルチモードの混合領域の音声コーダに関する。したがって、本発明の１つの態様では、マルチモードの混合領域の音声プロセッサが、少なくとも１つの時間領域コード化モードおよび少なくとも１つの周波数領域コード化モードをもつコーダと、コーダに接続され、かつ音声プロセッサによって処理されるフレーム内容に基づいてコーダのコード化モードを選択するように構成されている閉ループのモード選択デバイスとを含むことが好都合である。

【００１４】

本発明の別の態様では、フレームを処理する方法は、各連続する入力フレームへ開ループのコード化モード選択プロセスを適用して、入力フレームの音声内容に基づいて時間領域コード化モードか、または周波数領域コード化モードの何れかを選択するステップと、入力フレームの音声内容が定常状態の有声音の音声を示すときは、入力フレームを周波数領域でコード化するステップと、入力フレームの音声内容が定常状態の有声音の音声以外のものを示すときは、入力フレームを時間領域でコード化するステップと、周波数領域でコード化されたフレームと入力フレームとを比較して、性能尺度を求めるステップと、性能尺度が所定の閾値より低いときは入力フレームを時間領域でコード化するステップとを含むことが好都合である。

【００１５】

本発明の別の態様では、マルチモードの混合領域の音声プロセッサは、開ループのコード化モード選択プロセスを入力フレームへ適用して、入力フレームの音声内容に基づいて、時間領域コード化モードか、または周波数領域コード化モードの何れかを選択する手段と、入力フレームの音声内容が定常状態の有声音の音声を示すときは、入力フレームを周波数領域でコード化する手段と、入力フレームの音声内容が定常状態の有声音の音声以外のものを示すときは、入力フレームを時間領域でコード化する手段と、周波数領域でコード化されたフレームと入力フレームとを比較して、性能尺度を求める手段と、性能尺度が所定の閾値より低いときは、入力フレームを時間領域でコード化する手段とを含むことが好都合である。

【図面の簡単な説明】

【００１６】

【図１】音声コーダによって各端部で終端している通信チャンネルのブロック図。

【図２】マルチモードの混合領域の線形予測（mixed-domain linear prediction, MDLP）の音声コーダにおいて使用できるエンコーダのブロック図。

【図３】マルチモードのＭＤＬＰの音声コーダにおいて使用できるデコーダのブロック図。
【図４】図２のエンコーダにおいて使用できるＭＤＬＰエンコーダによって実行されるＭＤＬＰのコード化ステップを示すフローチャート。
【図５】音声コード化決定プロセスを示すフローチャート。
【図６】閉ループのマルチモードのＭＤＬＰの音声コーダのブロック図。
【図７】図６のコーダまたは図２のエンコーダにおいて使用できるスペクトルコーダのブロック図。
【図８】高調波コーダのシヌソイドの振幅を示す振幅対周波数のグラフ。
【図９】マルチモードのＭＤＬＰの音声コーダにおけるモード決定プロセスを示すフローチャート。
【図１０】音声信号の振幅対時間のグラフ（図１０ａ）および線形予測（linear prediction, LP）の残余振幅対時間のグラフ（図１０ｂ）。
【図１１】閉ループのコード化決定のもとでのレート／モード対フレーム指標のグラフ（図１１ａ）、閉ループの決定のもとでの知覚の信号対雑音比（perceptual signal-to-noise ratio, PSNR）対フレーム指標のグラフ（図１１ｂ）、閉ループのコード化決定がないときのレート／モードおよびＰＳＮＲの両者対フレーム指標のグラフ（図１１ｃ）。
【発明を実施するための形態】
【００１７】
図１では、第１のエンコーダ10は、ディジタル形式の音声サンプルｓ（ｎ）を受信し、サンプルｓ（ｎ）をコード化して、伝送媒体12、すなわち通信チャンネル12上で第１のデコーダ14へ送る。デコーダ14はコード化された音声サンプルをデコードし、出力された音声信号S_{SYNTH}（ｎ）を合成する。反対方向で伝送するには、第２のエンコーダ16がディジタル形式の音声サンプルｓ（ｎ）をコード化し、それを通信チャンネル18上で送る。第２のデコーダ20はコード化された音声サンプルを受信し、デコードし、合成された出力音声信号S_{SYNTH}（ｎ）を生成する。
【００１８】
音声サンプルｓ（ｎ）は、この技術において知られている種々の方法、例えばパルスコード変調（pulse code modulation, PMC）、コンパンドされたμ法、すなわちＡ法（companded μ-law, or A-law）を含む方法にしたがって、ディジタル形式にされて量子化された音声信号を表わしている。この技術において知られているように、音声サンプルｓ（ｎ）は、各々が所定数のディジタル形式の音声サンプルｓ（ｎ）を含む入力データのフレームへ編成される。例示的な実施形態では、８キロヘルツのサンプリングレートが採用され、各２０ミリ秒のフレームは１６０サンプルを含んでいる。別途記載する実施形態では、データ伝送レートはフレームごとに８キロビット秒（フルレート）から４キロビット秒（２分の１レート）、２キロビット秒（４分の１レート）、１キロビット秒（８分の１レート）へ変化することが好都合である。その代わりに、他のデータレートを使用してもよい。ここで使用されているように、“フルレート（full rate）”または“高レート（high rate）”という用語は、通常は、８キロビット秒以上のデータレートを指し、“２分の１レート”または“低レート”という用語は、通常は、４キロビット秒以下のデータレートを指す。比較的に少ない音声情報を含むフレームに対して、より低いビットレートを選択的に採用できるので、データ伝送レートを変化させることが好都合である。当業者によって理解されるように、他のサンプリングレート、フレームサイズ、およびデータ伝送レートを使用してもよい。
【００１９】
第１のエンコーダ10および第２のデコーダ20は共に第１の音声コーダ、すなわち音声コーデックを含む。同様に、第２のエンコーダ16および第１のデコーダ14は共に第２の音声コーダを含む。音声コーダはディジタル信号プロセッサ（digital signal processor, DSP）、特定用途向け集積回路（application-specific integrated circuit, ASIC）、離散的ゲート論理（discrete gate logic）、ファームウエア、または従来のプログラマブル

ソフトウエアモジュールおよびマイクロプロセッサで構成されていてもよいことが分かるであろう。ソフトウエアモジュールは、ＲＡＭメモリ、フラッシュメモリ、レジスタ、またはこの技術において知られている他の形態の書き込み可能な記憶媒体内にある。その代わりに、従来のプロセッサ、制御装置、または状態機械をマイクロプロセッサと置換してもよい。音声のコード化のために特別に設計されたＡＳＩＣの例は、本発明の譲受人に譲渡され、かつここでは参考文献として全面的に取り上げている米国特許第5,727,123号、および1994年2月16日に出願され、本発明の譲受人に譲渡され、かつここでは参考文献として全面的に取り上げている米国特許出願第08/197,417号（発明の名称：VOCODER ASIC）に記載されている。

【００２０】

１つの実施形態にしたがって、図２に示されているように、音声コーダ内で使用できるマルチモードの混合領域の線形予測（mixed-domain linear prediction, MDLP）エンコーダ100は、モード決定モジュール102、ピッチ推定モジュール104、線形予測（linear prediction, LP）解析モジュール106、ＬＰ解析フィルタ108、ＬＰ量子化モジュール110、およびＭＤＬＰ残余エンコーダ112を含む。入力音声フレームｓ（ｎ）は、モード決定モジュール102、ピッチ推定モジュール104、ＬＰ解析モジュール106、およびＬＰ解析フィルタ108へ供給される。モード決定モジュール102は、各入力音声フレームｓ（ｎ）の周期性および他の抽出パラメータ、例えばエネルギー、スペクトルチルト、ゼロ交差レート、などに基づいて、モード指標I_MおよびモードＭを生成する。周期性にしたがって音声フレームを分類する種々の方法は、米国特許出願第08/815,354号（発明の名称：METHOD AND APPARATUS FOR PERFORMING REDUCED RATE VARIABLE RATE VOCODING）に記載されており、これは1997年3月11日に出願され、本発明の譲受人に譲渡され、ここでは参考文献として全面的に取り上げている。このような方法は、米国電気通信工業会の業界暫定標準（Telecommunication Industry Association Industry Interim Standards）のTIA/EIA IS-127およびTIA/EIA IS-733にも採用されている。

【数１】

ピッチ推定モジュール104は、各入力音声フレームｓ（ｎ）に基づいて、ピッチ指標Ｉｐおよびラグ値$Ｐ_0$を生成する。ＬＰ解析モジュール106は、各入力音声フレームｓ（ｎ）に対して線形予測解析を行って、ＬＰパラメータａを生成する。ＬＰパラメータａはＬＰ量子化モジュール110へ供給される。ＬＰ量子化モジュール110はさらにモードMを受信し、モードに依存するやり方で量子化プロセスを行なう。ＬＰ量子化モジュール110は、ＬＰ指標$Ｉ_{ＬＰ}$および量子化されたＬＰパラメータ$\hat{a}$を生成する。ＬＰ解析フィルタ108は、入力音声フレームｓ（ｎ）に加えて、量子化されたＬＰパラメータ$\hat{a}$を受信する。ＬＰ解析フィルタ108はＬＰ残余信号Ｒ[ｎ]を生成し、ＬＰ残余信号Ｒ[ｎ]は、入力音声フレームｓ（ｎ）と、量子化された線形予測パラメータ$\hat{a}$に基づいて再構成された音声との誤差を表わしている。ＬＰ残余信号Ｒ[ｎ]、モードM、および量子化されたＬＰパラメータ$\hat{a}$はＭＤＬＰ残余エンコーダ112へ供給される。これらの値に基づいて、ＭＤＬＰ残余エンコーダ112は、図４のフローチャートを参照して別途記載するステップにしたがって、残余指標$Ｉ_{Ｒ}$および量子化された残余信号$\hat{R}[n]$を生成する。

【数２】

図３では、音声コーダにおいて使用されるデコーダ200は、ＬＰパラメータデコーディングモジュール202、残余デコーディングモジュール204、モードデコーディングモジュール206、およびＬＰ解析フィルタ208を含む。モードデコーディングモジュール206は、モード指標I_Mを受信してデコードし、モードMを生成する。ＬＰパラメータデコーディングモジュール202はモードMおよびＬＰ指標I_{LP}を受信する。ＬＰパラメータデコーディングモジュール202は、受信した値をデコードして、量子化されたＬＰパラメータ$\hat{a}$を生成する。残余デコーディングモジュール204は、残余指標I_R、ピッチ指標I_P、およびモード指標I_Mを受信する。残余デコーディングモジュール204は受信した値をデコードして、量子化された残余信号$\hat{R}[n]$を生成する。量子化された残余信号$\hat{R}[n]$および量子化されたＬＰパラメータ$\hat{a}$はＬＰ合成フィルタ208へ供給され、ＬＰ合成フィルタ208はデコードされた出力音声信号$\hat{s}[n]$を合成する。

【００２１】

ＭＤＬＰ残余エンコーダ112を除いて、図２のエンコーダ100および図３のデコーダ200の種々のモジュールの動作および構成はこの技術において知られており、上述の米国特許第5,414,796号およびLB. Rabiner & R.W. Schaferによる文献（Digital Processing of Speech Signals 396-453 (1978)）に記載されている。

【００２２】

１つの実施形態にしたがって、ＭＤＬＰエンコーダ（図示されていない）は、図４のフローチャートに示したステップを実行する。ＭＤＬＰエンコーダは、図２のＭＤＬＰ残余エンコーダ112であってもよい。ステップ300では、ＭＤＬＰエンコーダは、モードＭがフルレート（full rate, FR）であるか、４分の１レート（quarter rate, QR）であるか、または８分の１レート（eighth rate, ER）であるかを検査する。モードＭがＦＲ、ＱＲ、またはＥＲであるときは、ＭＤＬＰエンコーダはステップ302へ進む。ステップ302では、ＭＤＬＰエンコーダは対応するレート（Ｍの値に依存して－ＦＲ，ＱＲ、またはＥＲ）を残余指標I_Rへ適用する。時間領域のコード化は、ＦＲモードでは高精度で高レートのコード化であり、かつＣＥＬＰのコード化であることが好都合であるが、この時間領域のコード化は、ＬＰの残余フレーム、またはその代わりに音声フレームへ適用される。次にフレームは（ディジタル対アナログ変換および変調を含む別の信号処理の後で）送られる。１つの実施形態では、フレームは、予測誤差を表わすＬＰ残余フレームである。代わりの実施形態では、フレームは、音声サンプルを表わす音声フレームである。

【００２３】

他方で、ステップ300では、モードＭがＦＲ、ＱＲ、またはＥＲでなかったとき（すなわち、モードＭが２分の１レート（half rate, HR）であるとき）、ＭＤＬＰエンコーダはステップ304へ進む。ステップ304では、スペクトルのコード化、好ましくは高調波のコード化を２分の１のレートでＬＰ残余、またはその代わりに音声信号へ適用する。次にＭＤＬＰエンコーダはステップ306へ進む。ステップ306では、コード化された音声をデコードして、それを元の入力フレームと比較することによって、ひずみ尺度Ｄを得る。次にＭＤＬＰエンコーダは、ステップ308へ進む。ステップ308では、ひずみ尺度Ｄは所定の閾値

Ｔと比較される。ひずみ尺度Ｄが閾値Ｔよりも大きいときは、２分の１レートのスペクトル的にコード化されたフレームについて、対応する量子化されたパラメータが変調されて、送られる。他方で、ひずみ尺度Ｄが閾値Ｔ以下であるときは、ＭＤＬＰエンコーダはステップ310へ進む。ステップ310では、デコードされたフレームは、この時間領域においてフルレートで再びコード化される。従来の高レートで高精度のコード化アルゴリズム、例えば好ましくはＣＥＬＰのコード化を使用してもよい。次に、フレームと関係するＦＲモードの量子化されたパラメータが変調されて、送られる。

【００２４】

図５のフローチャートに示したように、次に1つの実施形態にしたがって閉ループのマルチモードのＭＤＬＰの音声コーダは、音声サンプルを処理して送る1組のステップにしたがう。ステップ400では、音声コーダは、連続するフレーム内の音声信号のディジタルサンプルを受信する。所与のフレームを受信すると、音声コーダはステップ402へ進む。ステップ402では、音声コーダはフレームのエネルギーを検出する。エネルギーはフレームの音声活動（speech activity）の尺度である。音声検出は、ディジタル形式の音声サンプルの振幅の平方を加算して、生成されたエネルギーを閾値と比較することによって行なわれる。１つの実施形態では、背景ノイズの変化レベルに基づいて閾値を採用する。例示的な可変閾値の音声活動検出器は、上述の米国特許第5,414,796号に記載されている。若干の無声音の音声は非常に低いエネルギーのサンプルであり、誤って背景ノイズとしてコード化されてしまうことがある。このようなことが発生するのを防ぐために、上述の米国特許第5,414,796号に記載されているように、低エネルギーサンプルのスペクトルのチルトを使用して、無声音の音声を背景ノイズと区別する。

【００２５】

フレームのエネルギーを検出した後で、音声コーダはステップ404へ進む。ステップ404では、音声コーダは、音声情報を含んでいるかについてフレームを分類するのに、検出されたフレームエネルギーが十分であるかどうかを判断する。検出されたフレームエネルギーが所定の閾値レベルよりも低いときは、音声コーダはステップ406へ進む。ステップ406では、音声コーダは背景ノイズ（すなわち、非音声、または黙音）としてフレームをコード化する。１つの実施形態では、背景ノイズのフレームは、８分の１レート、すなわち１キロビット秒でコード化される時間領域である。ステップ404では、検出されたフレームのエネルギーが所定の閾値レベル以上であるとき、フレームは音声として分類され、音声コーダはステップ408へ進む。

【００２６】

ステップ408では、音声コーダは、フレームが周期的であるかどうかを判断する。周期性を判断する種々の既知の方法には、例えばゼロ交差の使用および正規化された自動相関関数（normalized autocorrelation function, NACF）の使用を含む。とくに、ゼロ交差およびＮＡＣＦを使用して、周期性を検出することは、米国出願第08/815,354号（発明の名称：METHOD AND APPARATUS FOR PERFORMING REDUCED RATE VARIABLE RATE VOCODING）に記載されており、これは1997年3月11日に出願され、本発明の譲受人に譲渡され、ここでは参考文献として全面的に取り上げている。さらに加えて、無声音の音声から有声音の音声を区別するのに使用される上述の方法は、米国電気通信工業会の業界暫定標準（Telecommunication Industry Association Industry Interim Standards）のTIA/EIA IS-127およびTIA/EIA IS-733に採用されている。ステップ408においてフレームが周期的でないと判断されるとき、音声コーダはステップ410へ進む。ステップ410では、音声コーダは、フレームを無声音の音声としてコード化する。1つの実施形態では、無声音の音声フレームは、４分の１レート、すなわち２キロビット秒でコード化される時間領域である。ステップ408では、フレームが周期的であると判断されるとき、音声コーダはステップ412へ進む。

【００２７】

ステップ412では、音声コーダは、例えば上述の米国特許出願第08/815,354号に記載されているように、この技術において知られている周期性検出方法を使用して、フレームが

十分に周期的であるかどうかを判断する。フレームが十分に周期性でないと判断されるときは、音声コーダはステップ414へ進む。ステップ414では、フレームは遷移音声（transition speech）（すなわち、無声音の音声から有声音の音声への遷移）として時間領域でコード化される。１つの実施形態では、遷移音声フレームはフルレート、すなわち８キロビット秒で時間領域でコード化される。

【００２８】

音声コーダは、ステップ412においてフレームが十分に周期的であると判断すると、ステップ416へ進む。ステップ416では、音声コーダは有声音の音声としてフレームをコード化する。１つの実施形態では、有声音の音声フレームは、とくに２分の１レート、すなわち４キロビット秒でスペクトル的にコード化される。図７を参照して別途記載するように、有声音の音声フレームは、高調波のコーダでスペクトル的にコード化されることが好都合である。その代わりに、他のスペクトルコーダは、この技術において知られているように、例えばシヌソイド変換コーダ（sinusoidal transmission coder）またはマルチバンド励起コーダ(multiband excitation coder)として使用できることが好都合である。次に音声コーダはステップ418へ進む。ステップ418では、音声コーダはコード化された有声音の音声フレームをデコードする。次に音声コーダはステップ420へ進む。ステップ420では、デコードされた有声音の音声フレームを、このフレームの対応する入力音声サンプルと比較して、合成された音声のひずみ尺度を得て、２分の１レートの有声音音声のスペクトルコード化モデルが許容限度内で動作しているかどうかを判断する。次に音声コーダはステップ422へ進む。

【００２９】

ステップ422では、音声コーダは、デコードされた有声音の音声フレームと、このフレームに対応する入力音声フレームとの誤差が所定の閾値より小さいかどうかを判断する。１つの実施形態では、この判断は、図６を参照して別途記載するやり方で行われる。コード化のひずみが所定の閾値よりも低いときは、音声コーダはステップ426へ進む。ステップ426では、音声コーダは、ステップ416のパラメータを使用して、フレームを有声音の音声として送る。ステップ422では、コード化のひずみが所定の閾値以上であるときは、音声コーダはステップ414へ進み、ステップ400において受信したディジタル形式の音声サンプルのフレームを遷移音声としてフルレートで時間領域でコード化する。

【００３０】

ステップ400ないし410は開ループのコード化決定モードを含むことに注目すべきである。他方で、ステップ412ないし426は閉ループのコード化決定モードを含む。

【００３１】

１つの実施形態では、図６に示したように、閉ループのマルチモードのＭＤＬＰの音声コーダはアナログ対ディジタルコンバータ（analog-to-digital converter, A/D）500を含み、Ａ／Ｄ500はフレームバッファ502に接続され、フレームバッファ502は制御プロセッサ504に接続される。エネルギー計算器506、有声音音声の検出器508、背景ノイズエンコーダ510、高レートの時間領域エンコーダ512、および低レートのスペクトルエンコーダ514は制御プロセッサ504へ接続される。スペクトルデコーダ516はスペクトルエンコーダ514に接続され、誤差計算器518はスペクトルデコーダ516および制御プロセッサ504へ接続される。閾値比較器520は、誤差計算器518および制御プロセッサ504へ接続される。バッファ522はスペクトルエンコーダ514、スペクトルデコーダ516、および閾値比較器520へ接続される。

【００３２】

図６の実施形態では、音声コーダの構成要素は、音声コーダ内にファームウエアまたは他のソフトウエア駆動モジュールとして構成されていることが好都合であり、音声コーダ自身はＤＳＰまたはＡＳＩＣ内にあることが好都合である。当業者には、音声コーダの構成要素は、多数の他の既知のやり方で同様に適切に構成できることが分かるであろう。制御プロセッサ504はマイクロプロセッサであることが好都合であるが、制御装置、状態機械、または離散的論理と共に構成されていてもよい。

【００３３】
図６のマルチモードのコーダでは、音声信号はＡ／Ｄ500へ供給される。Ａ／Ｄ500はアナログ信号をディジタル形式の音声サンプルＳ（ｎ）へ変換する。ディジタル形式の音声サンプルは、フレームバッファ502へ供給される。制御プロセッサ504は、フレームバッファ502からディジタル形式の音声サンプルを得て、それらをエネルギー計算器506へ供給する。エネルギー計算器506は、次の式にしたがって音声サンプルのエネルギーＥを計算する：
【数３】

$$E = \sum_{n=0}^{159} S^2(n)$$

【００３４】
なお、フレームは２０ミリ秒長であり、サンプリングレートは８キロヘルツである。計算されたエネルギーＥは制御プロセッサ504へ送られる。
【００３５】
制御プロセッサ504は、計算された音声エネルギーを音声活動（speech activity）の閾値と比較する。計算されたエネルギーが音声活動の閾値よりも小さいときは、制御プロセッサ504はディジタル形式の音声サンプルをフレームバッファ502から背景ノイズエンコーダ510へ送る。背景ノイズエンコーダ510は、背景ノイズの推定値を保持するために必要な最少数のビットを使用して、フレームをコード化する。
【００３６】
計算されたエネルギーが音声活動の閾値以上であるときは、制御プロセッサ504はディジタル形式の音声サンプルをフレームバッファ502から有声音音声の検出器508へ方向付ける。有声音音声の検出器508は、音声フレームの周期性が、低ビットレートのスペクトルのコード化を使用して効率的なコード化を可能にするかどうかを判断する。音声フレーム内の周期性のレベルを判断する方法は、この技術においてよく知られており、例えば正規化された自動相関関数（normalized autocorrelation function, NACF）およびゼロ交差の使用を含む。これらの方法および他の方法は、上述の米国特許出願第08/815,354号に記載されている。
【００３７】
有声音音声の検出器508は、スペクトルエンコーダ514が効率的にコード化するのに十分な周期性をもつ音声を音声フレームが含んでいるかどうかを示す信号を制御プロセッサ504へ供給する。有声音音声の検出器508が、音声フレームが十分な周期性を欠いていると判断するとき、制御プロセッサ504はディジタル形式の音声サンプルを高レートのエンコーダ512へ方向付け、エンコーダ512は所定の最大データレートで音声を時間領域でコード化する。１つの実施形態では、所定の最大データレートは８キロビット秒であり、高レートのエンコーダ512はＣＥＬＰのコーダである。
【００３８】
有声音音声の検出器508が最初に、音声信号が、スペクトルエンコーダ514が効率的にコード化するのに十分な周期性をもつと判断するとき、制御プロセッサ504は、フレームバッファ502からスペクトルエンコーダ514へディジタル形式の音声サンプルを方向付ける。例示的なスペクトルエンコーダは、図７を参照して別途詳しく記載する。

【数４】

スペクトルエンコーダ514は、推定されたピッチ周波数F_0、ピッチ周波数の高調波の振幅A_l、および音声情報（voicing information）V_cを抽出する。スペクトルエンコーダ514はこれらのパラメータをバッファ522およびスペクトルデコーダ516へ送る。スペクトルデコーダ516は、従来のＣＥＬＰエンコーダ内のエンコーダのデコーダに似ていることが好都合である。スペクトルデコーダ516は、（図７を参照して別途記載される）スペクトルデコーディングフォーマットのしたがって、合成された音声サンプル$\hat{s}[n]$を生成し、合成された音声サンプルを誤差計算器518へ供給する。制御プロセッサ504は音声サンプルＳ（ｎ）を誤差計算器518へ送る。

【数５】

誤差計算器518は、次の式にしたがって、各音声サンプルＳ（ｎ）と、各対応する合成された音声サンプル$\hat{s}[n]$との間の平均平方誤差（mean square error, MSE）を計算する：

$$MSE = \sum_{n=0}^{159} \left(S(n) - \hat{S}(n)\right)^2$$

計算されたＭＳＥは閾値比較器520へ供給され、閾値比較器520は、ひずみレベルが許容範囲内であるかどうか、すなわちひずみレベルが所定の閾値よりも小さいかどうかを判断する。

【００３９】

計算されたＭＳＥが許容範囲内であるときは、閾値比較器520は信号をバッファ522へ供給し、スペクトル的にコード化されたデータは音声コーダから出力される。他方で、ＭＳＥが許容限界内でないときは、閾値の比較器520は信号を制御プロセッサ504へ送り、制御プロセッサ504はディジタル形式のサンプルをフレームバッファ502から高レートの時間領域のエンコーダ512へ方向付ける。時間領域のエンコーダ512は、所定の最大レートでフレームをコード化し、バッファ522の内容は捨てられる。

【００４０】

図６の実施形態では、採用されたスペクトルのコード化のタイプは高調波のコード化であり、これについては図７を参照して別途記載するが、代わりの実施形態では、シヌソイド変換のコード化またはマルチバンド励起のコード化のような、スペクトルのコード化のタイプであってもよい。マルチバンド励起のコード化の使用は、米国特許第5,195,166号に記載されており、シヌソイド変換のコード化の使用は、例えば米国特許第4,865,068号に記載されている。

【００４１】

遷移フレーム、および位相ひずみ閾値が周期性パラメータ以下である有声音フレームで

は、図６のマルチモードコーダはフルレート、すなわち８キロビット秒で、高レートの時間領域のコーダ512によって、ＣＥＬＰのコード化を採用することが好都合である。その代わりに、このようなフレームに対して、他の既知の形態の高レートの時間領域のコード化を使用してもよい。したがって、遷移フレーム（および十分に周期的でない有声音フレーム）は高い精度でコード化され、入力および出力における波形は適切に整合し、位相情報は適切に保持される。１つの実施形態では、マルチモードコーダは、閾値比較器520の判断と無関係に、閾値が周期性の尺度を越えている所定数の連続する有声音フレームを処理した後で、各フレームごとに２分の１レートのスペクトルのコード化からフルレートのＣＥＬＰのコード化へスイッチする。

【００４２】

制御プロセッサ504に関連して、エネルギー計算器506および有声音音声の検出器508は開ループのコード化決定を含むことに注意すべきである。対照的に、制御プロセッサ504に関連して、スペクトルエンコーダ514、スペクトルデコーダ516、誤差計算器518、閾値比較器520、およびバッファ522は閉ループのコード化決定を含む。

【００４３】

図７を参照して記載した１つの実施形態では、スペクトルのコード化、好ましくは高調波のコード化を使用して、低ビットレートで十分に周期的な有声音フレームをコード化する。スペクトルコーダは、一般的に、周波数領域内の各音声フレームをモデル化してコード化することによって知覚的に重要なやり方で音声スペクトル特性の時間にしたがう漸進的変化（time-evolution）を保持することを試みるアルゴリズムとして規定される。このようなアルゴリズムの本質的な部分では、（１）スペクトルの解析またはパラメータの推定、（２）パラメータの量子化、（３）出力された音声波形とデコードされたパラメータとの合成を行う。したがって、１組のスペクトルパラメータをもつ短期間の音声スペクトルの重要な特性を保持し、デコードされたスペクトルパラメータを使用して、出力音声を合成することを目的とする。通常は、出力音声は、シヌソイドの重み付けされた和として合成される。シヌソイドの振幅、周波数、および位相は、解析中に推定されるスペクトルパラメータである。

【００４４】

“合成による解析”はＣＥＬＰのコード化においてよく知られた技術であるが、この技術はスペクトルのコード化には利用されていない。合成による解析がスペクトルコーダに適用されない主な理由は、初期位相の情報の損失によって、音声モデルが知覚の観点から適切に機能していても、合成された音声の平均二乗エネルギー（mean square energy, MSE）が高いからである。したがって、初期位相を正確に生成すると、音声サンプルと再構成された音声とを直接に比較して、音声モデルが音声フレームを正確にコード化しているかどうかを判断できるといった別の長所がある。

【００４５】

スペクトルのコード化では、出力された音声フレームは次に示すように合成することができる：　$S[n]=S_v[n]+S_{uv}[n]$, $n=1, 2, \ldots, N$, なお、Ｎは１フレーム当りのサンプル数であり、S_vおよびS_{uv}は、それぞれ有声音成分および無声音成分である。シヌソイド和合成プロセス（sum-of-sinusoid synthesis process）は次の式に示すように有声音成分を生成する：

【数６】

$$s[n]=\sum_{k=1}^{L}A(k,n)\bullet\cos(2\pi nf_k+\theta(k,n))$$

なお、Ｌはシヌソイドの合計数であり、

ｆ$_k$は短期間のスペクトルにおける目的の周波数であり、

Ａ（ｋ，ｎ）はシヌソイドの振幅であり、

θ（ｋ，ｎ）はシヌソイドの位相である。

【００４６】

振幅、周波数、および位相パラメータは、スペクトル解析プロセスによって入力フレームの短期間のスペクトルから推定される。無声音成分は、単一のシヌソイド和合成において有声音部分と一緒に生成されるか、または専用の無声音合成プロセスによって別々に計算され、Ｓ$_v$へ再び加えられる。

【００４７】

図７の実施形態では、高調波コーダと呼ばれる特定のタイプのスペクトルコーダを使用して、低ビットレートで十分に周期的な有声音フレームをスペクトル的にコード化する。高調波のコーダは、シヌソイド和としてフレームを特徴付け、フレームの小さいセグメントを解析する。シヌソイド和の中の各シヌソイドは、フレームのピッチＦ$_0$の整数倍の周波数をもつ。代わりの実施形態では、高調波のコーダ以外の特定のタイプのスペクトルコーダを使用し、各フレームに対するシヌソイド周波数は、０ないし２πの１組の実数から得られる。図７の実施形態では、和の中の各シヌソイドの振幅および位相が選択されることが好都合であり、その結果、図８のグラフによって示したように、和は１期間において信号と最良に整合する。高調波のコーダは一般的に外部の分類を採用し、各入力音声フレームは有声音または無声音として表示する。有声音フレームでは、シヌソイドの周波数は推定されたピッチ（Ｆ$_0$）の高調波に制限され、すなわちｆ$_k$＝ｋＦ$_0$である。無声音の音声では、短期間のスペクトルのピークを使用して、シヌソイドを判断する。次の式に示すように、振幅および位相が補間されて、フレームにおいて漸進的変化をまねる：

【数７】

$$A(k,n)=C_1(k)^*n+C_2(k)$$

$$\theta(k,n)=B_1(k)^*n^2+B_2(k)^*n+B_3(k)$$

係数[Ci(k),Bi(k)]は、ウインドウ処理された入力音声フレームの短期間のフーリエ変換（short-term Fourier Transform (STFT)）から、特定された周波数の位置ｆ$_k$（＝ｋｆ$_0$）における振幅、周波数、および位相の瞬間値によって推定される。

【００４８】

シヌソイドごとに送られるパラメータは振幅および周波数である。位相は送られないが、その代わりに、例えば準位相モデル（quadratic phase model）、または位相の従来の多項式表現を含むいくつかの既知の技術にしたがってモデル化される。

【００４９】

図７に示されているように、高調波コーダはピッチ抽出器600を含み、ピッチ抽出器600はウインドウ処理論理602へ接続され、ウインドウ処理論理602は離散フーリエ変換（Discrete Fourier Transform, DFT）、および高調波解析論理604へ接続される。入力として音

声サンプルＳ（ｎ）を受信するピッチ抽出器600はは、ＤＦＴおよび高調波解析論理604へも接続される。ＤＦＴおよび高調波解析論理604は、残余エンコーダ606へ接続される。ピッチ抽出器600、ＤＦＴおよび高調波解析論理604、並びに残余エンコーダ606は、パラメータ量子化器608へそれぞれ接続される。パラメータ量子化器608はチャンネルエンコーダ610へ接続され、チャンネルエンコーダ610は送信機612へ接続される。送信機612は、例えば、符号分割多重アクセス（code division multiple access, CDMA）のような標準の無線周波数（radio-frequency, RF）のインターフェイスによって空中インターフェイス（over-the-air interface）上で、受信機614へ接続される。受信機614はチャンネルデコーダ616へ接続され、チャンネルデコーダ616は非量子化器618へ接続される。非量子化器618はシヌソイド和音声合成器620へ接続される。シヌソイド和音声合成器620へさらに接続されるのは位相推定器622であり、位相推定器622は入力として前フレーム情報を受信する。シヌソイド和音声合成器620は合成された音声出力S_{SYNTH}（ｎ）を生成するように構成されている。

【００５０】

ピッチ抽出器600、ウインドウ処理論理602、ＤＴＦおよび高調波解析論理604、残余エンコーダ606、パラメータ量子化器608、チャンネルエンコーダ610、チャンネルデコーダ616、非量子化器618、シヌソイド和音声合成器620、並びに位相推定器622は、例えばファームウエアまたはソフトウエアモジュールを含む、当業者によく知られている種々の異なるやり方で構成することができる。送信機612および受信機614は、当業者には知られている対応する標準のＲＦの構成要素で実行されていてもよい。

【００５１】

図７の高調波コーダでは、入力サンプルＳ（ｎ）はピッチ抽出器600によって受信され、ピッチ抽出器600はピッチ周波数情報F_0を抽出する。次にサンプルは、ウインドウ処理論理602によって適切なウインドウ処理関数によって乗算され、音声フレームの小さいセグメントの解析を可能にしている。ピッチ抽出器600によって供給されるピッチ情報を使用して、ＤＦＴおよび高調波解析論理604はサンプルのＤＦＴを計算して、複合のスペクトル点を生成し、この複合のスペクトル点から、図８のグラフによって示されているように、高調波の振幅A_Iを抽出し、なお図８において、Ｌは高調波の合計数を示している。ＤＦＴは残余エンコーダ606へ供給され、残余エンコーダ606は音声情報（voicing information）V_cを抽出する。

【００５２】

V_cパラメータは、図８に示されているように、周波数軸上の点を示し、V_cがより高くなると、スペクトルは無声音の音声信号の特性を示し、最早高調波ではなくなることに注意すべきである。対照的に、点V_cより低くなると、スペクトルは高調波であり、有声音の音声の特性を示す。

【００５３】

A_I，F_0，およびV_cの成分は、パラメータ量子化器608へ供給され、パラメータ量子化器608では情報を量子化する。量子化された情報はパケットの形態でチャンネルエンコーダ610へ供給され、チャンネルエンコーダ610では、例えばハーフレート、すなわち４キロビット秒のような低ビットレートでパケットを量子化する。パケットは送信機612へ供給され、送信機612はパケットを変調して、生成された信号を受信機614へ空中で（over the air）送る。受信機614は信号を受信して、復調して、コード化されたパケットをチャンネルデコーダ616へ送る。チャンネルデコーダ616はパケットをデコードして、デコードされたパケットを非量子化器618へ供給する。非量子化器618は情報を非量子化する。情報はシヌソイド和音声合成器620へ供給される。

【００５４】

シヌソイド和音声合成器620は、Ｓ[ｎ]についての上述の式にしたがって短期間の音声スペクトルをモデル化する複数のシヌソイドのモデリングを合成するように構成されている。シヌソイドf_kの周波数は、基本周波数F_0の倍数または高調波であり、準周期的な（すなわち、遷移の）有声音の音声セグメントに対するピッチの周期性をもつ周波数であ

る。

【００５５】

さらに加えて、シヌソイド和の音声合成器620は位相推定器622から位相情報を受信する。位相推定器622は前フレームの情報、すなわち直前フレームについてのA_I，F_0，およびV_cのパラメータを受信する。位相推定器622は、前フレームの再構成されたＮのサンプルも受信し、なおＮはフレーム長（すなわち、Ｎは１フレーム当りのサンプル数）である。位相推定器622は、前フレームの情報に基づいて、フレームの初期位相を判断する。初期位相の判断は、シヌソイド和の音声合成器620へ供給される。現在のフレームに関する情報と、過去のフレーム情報に基いて位相推定器622によって行なわれた初期位相の計算とを基にして、シヌソイド和音声合成器620は上述のように音声フレームを生成する。

【００５６】

既に記載したように、高調波のコーダは、前フレームの情報を使用して、位相がフレームからフレームへ線形に変化することを予測することによって、音声フレームを合成、すなわち再構成する。上述の合成モデルは、一般的に準位相モデルと呼ばれており、このような合成モデルでは、係数B_3（ｋ）は、現在の有声音フレームの初期位相が合成されていることを表わしている。位相を判断するとき、従来の高調波のコーダは初期位相をゼロに設定するか、または初期位相値をランダムに、あるいは疑似ランダム生成方法を使用して生成する。位相をより正確に予測するために、位相推定器622は、直前のフレームが有声音の音声フレーム（すなわち、十分に周期的なフレーム）であるか、または遷移音声フレームであるかに依存して、初期位相を判断するための２つの可能な方法の一方を使用する。前フレームが有声音の音声フレームであったときは、このフレームの推定された最終位相値は、現在のフレームの初期位相値として使用される。他方で、前フレームが遷移フレームとして分類されたときは、現在のフレームの初期位相値は、前フレームのスペクトルから得られ、これは前フレームのデコーダ出力のＤＦＴを行なうことによって得られる。したがって位相推定器622は、（遷移フレームである前フレームがフルレートで処理されたので）既に使用可能である正確な位相情報を使用できる。

【００５７】

１つの実施形態では、閉ループのマルチモードのＭＤＬＰの音声コーダは、図９のフローチャート内に示されている音声処理ステップにしたがう。音声コーダは、最も適切なコード化モードを選択することによって、各入力音声フレームのＬＰの残余をコード化する。一定のモードは時間領域内でＬＰの残余、すなわち音声の残余をコード化し、一方で他のモードは周波数領域内でＬＰの残余、すなわち音声の残余を表わす。モードの組には、遷移フレームに対するフルレートの時間領域（Ｔモード）；有声音フレームに対する２分の１レートの周波数領域（Ｖモード）；無声音フレームに対する４分の１レートの時間領域（Ｕモード）；およびノイズフレームに対する８分の１レートの時間領域（Ｎモード）がある。

【００５８】

当業者には、図９に示したステップにしたがうことによって、音声信号または対応するＬＰの残余がコード化されることが分かるであろう。ノイズ、無声音、遷移、および有声音の音声の波形特性は、図１０ａのグラフにおいて時間関数として参照することができる。ノイズ、無声音、遷移、および有声音のＬＰの残余の波形特性は、図１０ｂのグラフにおいて時間関数として参照することができる。

【００５９】

ステップ700では、４つのモード（Ｔ、Ｖ、Ｕ，またはＮ）の何れか１つに関して、開ループのモード決定を行って、入力音声の残余Ｓ（ｎ）へ適用する。Ｔモードが適用されるときは、ステップ702では、時間領域においてＴモード、すなわちフルレートで音声の残余が処理される。Ｕモードが適用されるときは、ステップ704で、時間領域においてＵモード、すなわち４分の１レートで音声の残余が処理される。Ｎモードが適用されるときは、ステップ706では、時間領域においてＮモード、すなわち８分の１レートで音声の残

余が処理される。Ｖモードが適用されるときは、ステップ708では、周波数領域においてＶモードで、すなわち２分の１レートで音声の残余が処理される。

【００６０】

ステップ710では、ステップ708でコード化された音声がデコードされ、入力音声の残余Ｓ（ｎ）と比較され、性能尺度Ｄが計算される。ステップ712では、性能尺度Ｄが所定の閾値Ｔと比較される。性能尺度Ｄが閾値Ｔ以上であるときは、ステップ714では、ステップ708においてスペクトル的にコード化された音声の残余は送信を許可される。他方では、性能尺度Ｄが閾値Ｔよりも小さいときは、ステップ716では、入力音声の残余Ｓ（ｎ）はＴモードで処理される。別の実施形態では、性能尺度は計算されず、閾値は規定されない。その代わりに、所定数の音声残余フレームがＶモードで処理された後で、次のフレームはＴモードで処理される。

【００６１】

図９に示した決定のステップでは、高ビットレートのＴモードを必要なときだけ使用して、より低いビットレートのＶモードで有声音の音声セグメントの周期性を活用することができ、一方でＶモードが適切に実行されないときは、フルレートにスイッチすることによって品質の低下を防ぐことが好都合である。したがって、フルレートの音声品質に近づく非常に高い音声品質を、フルレートよりも相当に低い平均レートで生成することができる。さらに、選択された性能尺度および選ばれた閾値によって、目標の音声品質を制御することができる。

【００６２】

Ｔモードへの“更新”は、モデル位相追跡を入力音声の位相追跡の近くに維持することによって、後でＶモードを適用する動作を向上することができる。Ｖモードの性能が不適切であるときは、ステップ710および712の閉ループの性能検査はＴモードへスイッチし、初期位相値を“リフレッシュ”して、モデルの位相追跡を元の入力音声位相追跡に再び近付けることによって、次のＶモードの処理の性能を向上することができる。例えば、図１１ａないしｃのグラフに示したように、開始から５番目のフレームは、使用されているＰＳＮＲのひずみ尺度によって証明されているように、Ｖモードで適切に働かない。その結果、閉ループの決定および更新がないときは、モデル化された位相追跡は元の入力音声位相追跡から相当に外れ、図１１ｃに示したように、ＰＳＮＲを相当に劣化する。さらに、Ｖモードで処理される次のフレームの性能は劣化する。しかしながら、閉ループの決定のもとでは、５番目のフレームは、図１１ａに示したように、Ｔモードの処理へスイッチされる。５番目のフレームの性能は、図１１ｂに示したように、ＰＳＮＲにおける向上によって証明されているように、更新によって相当に向上する。さらに加えて、Ｖモードのもとで処理される次のフレームの性能も向上する。

【００６３】

図９に示した決定のステップでは、非常に正確な初期位相推定値を与えることによって、Ｖモードの表現品質を向上し、生成されたＶモードの合成された音声の残余信号は元の入力音声の残余Ｓ（ｎ）と正確に時間的に整合することを保証する。最初のＶモードで処理された音声の残余セグメントにおける初期位相は、次に示すやり方で直前のデコードされたフレームから求められる。各高調波では、前フレームがＶモードで処理されたときは、初期位相は前フレームの推定された最終位相に等しく設定される。各高調波では、前フレームがＴモードで処理されたときは、初期位相は前フレームの実際の高調波の位相に等しく設定される。前フレームの実際の高調波の位相は、全ての前フレームを使用して過去のデコードされた残余のＤＦＴをとることによって求められる。その代わりに、前フレームの実際の高調波の位相は、前フレームの種々のピッチ期間を処理することによって、ピッチが同期するやり方で、過去のデコードされたフレームのＤＦＴをとることによって求められる。

【００６４】

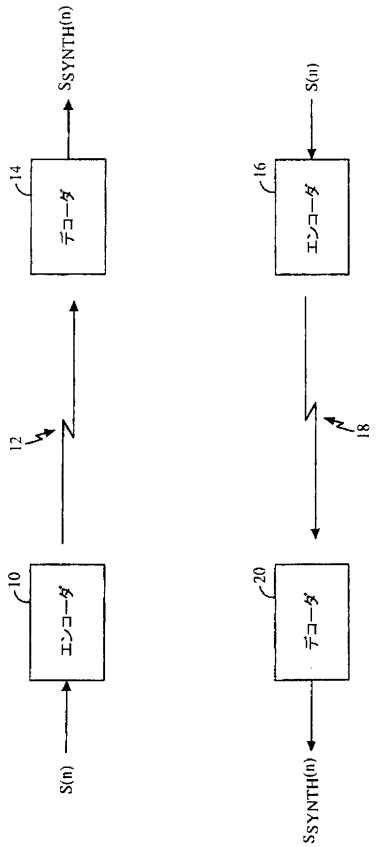
本明細書では、斬新な閉ループのマルチモードの混合領域の線形予測（mixed-domain linear prediction, MDLP）の音声コーダを記載した。当業者には、ここに開示した実施形

態に関係して記載した種々の例示的な論理ブロックおよびアルゴリズムのステップが、ディジタル信号プロセッサ（digital signal processor, DSP）、特定用途向け集積回路（application specific integrated circuit, ASIC）、離散的ゲートまたはトランジスタ論理、例えばレジスタおよびＦＩＦＯのような離散的ハードウエア構成要素、1組のファームウエア命令を実行するプロセッサ、または従来のプログラマブルソフトウエアモジュールおよびプロセッサで構成または実行できることが分かるであろう。プロセッサは、マイクロプロセッサであることが好都合であるが、その代わりに従来のプロセッサ、制御装置、マイクロプロセッサ、または状態機械であってもよい。ソフトウエアモジュールは、ＲＡＭメモリ、フラッシュメモリ、レジスタ、またはこの技術において知られている他の形態の書き込み可能な記憶媒体内にあってもよい。当業者にはさらに、上述の記述全体で参照したデータ、命令、コマンド、情報、信号、ビット、符号、およびチップが、電圧、電流、電磁波、磁界または磁粒、光の範囲または粒子（optical field or particles）、あるいはその組み合わせによって都合よく表わされることが分かるであろう。

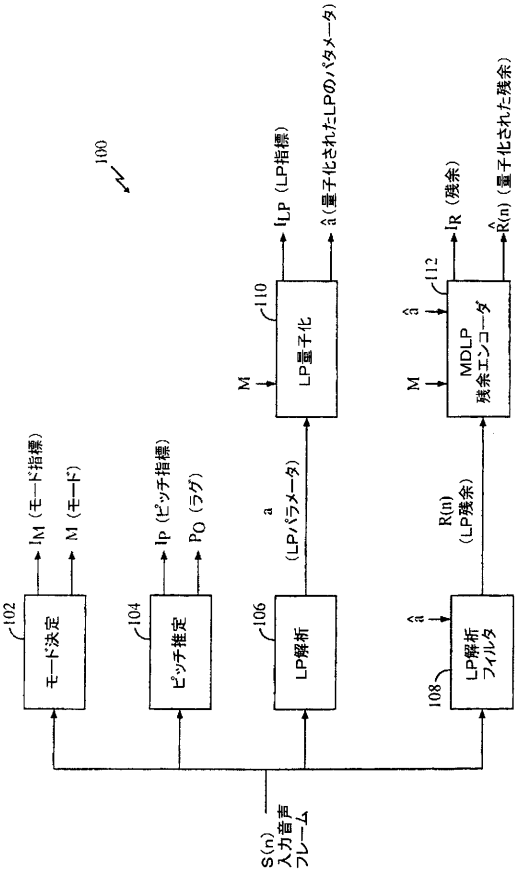
【００６５】

本明細書では、本発明の好ましい実施形態を示し、記載した。しかしながら、当業者の一人には、ここに記載した実施形態に対して、本発明の意図または技術的範囲から逸脱せずに多数の変更を加えられることが分かるであろう。したがって、本発明は、特許請求項にしたがうことを除いて制限されない。

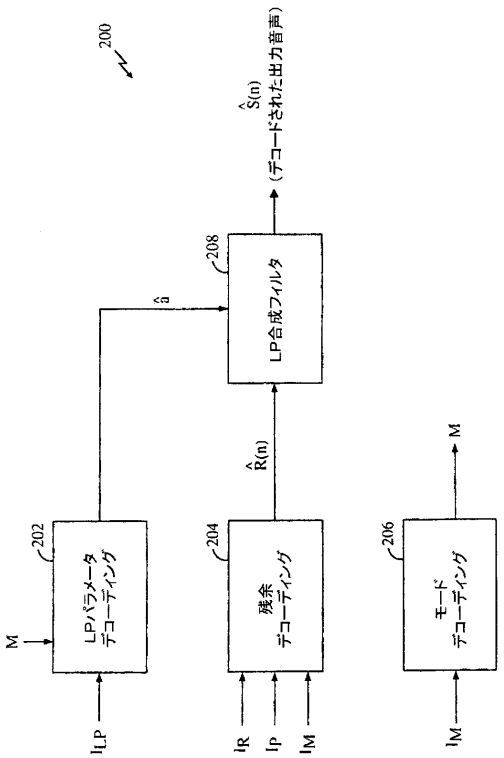
【図１】



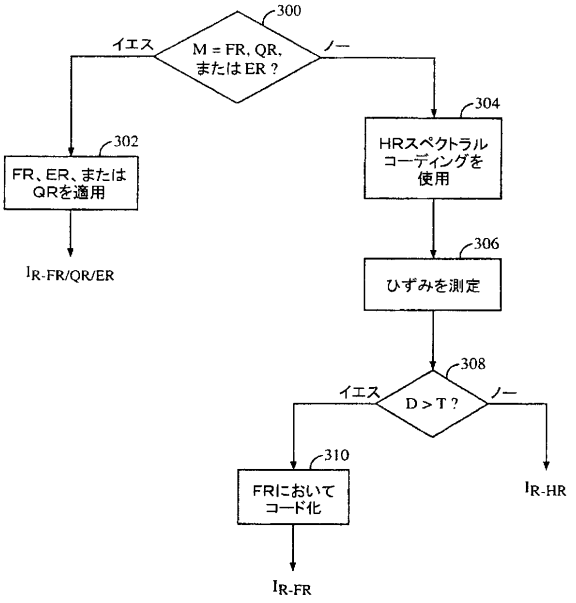
【図２】



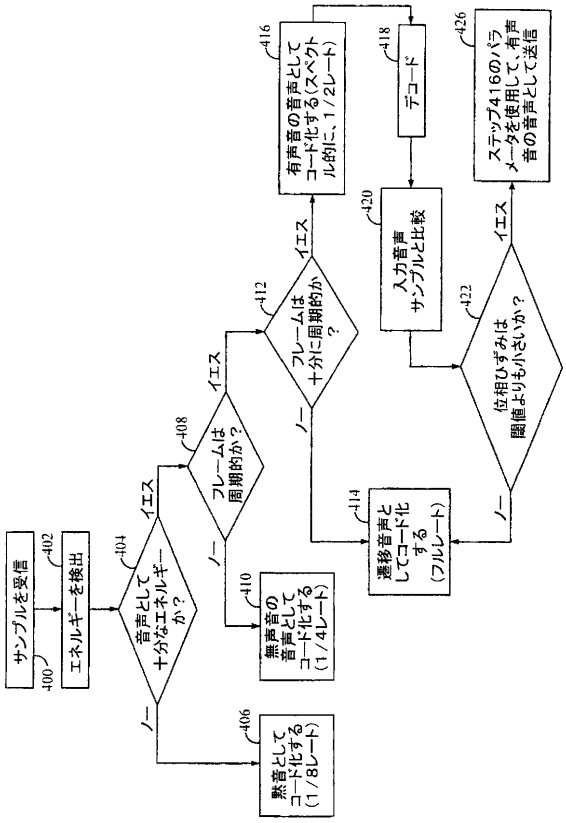
【図３】


200
$\hat{S}(n)$
(デコードされた出力音声)
208
LP合成フィルタ
$\hat{a}$
$\hat{R}(n)$
M
202
LPパラメータデコーディング
204
残余デコーディング
206
モードデコーディング
M
I_{LP}
I_R
I_P
I_M
I_M

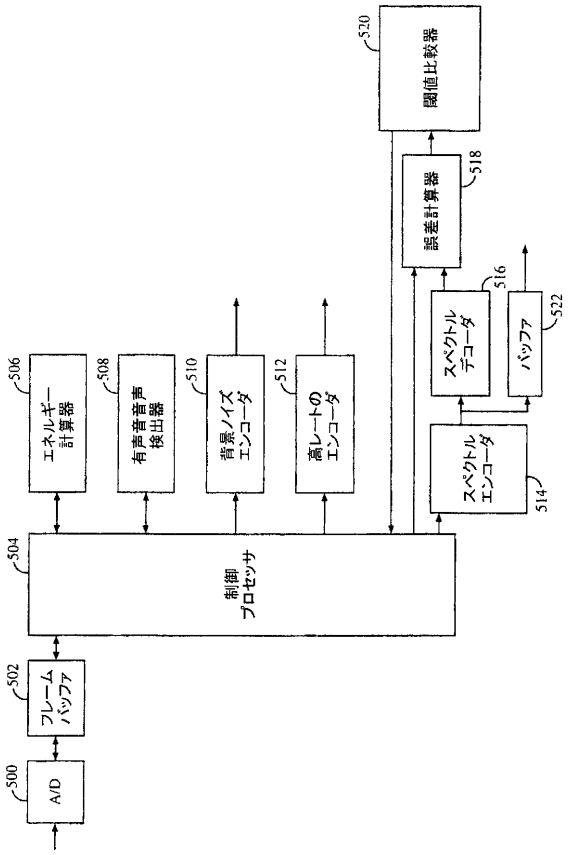
【図４】


300
M = FR, QR, またはER？
イエス
ノー
302
FR、ER、またはQRを適用
$I_{R-FR/QR/ER}$
304
HRスペクトラルコーディングを使用
306
ひずみを測定
308
D > T？
イエス
ノー
310
FRにおいてコード化
I_{R-FR}
I_{R-HR}

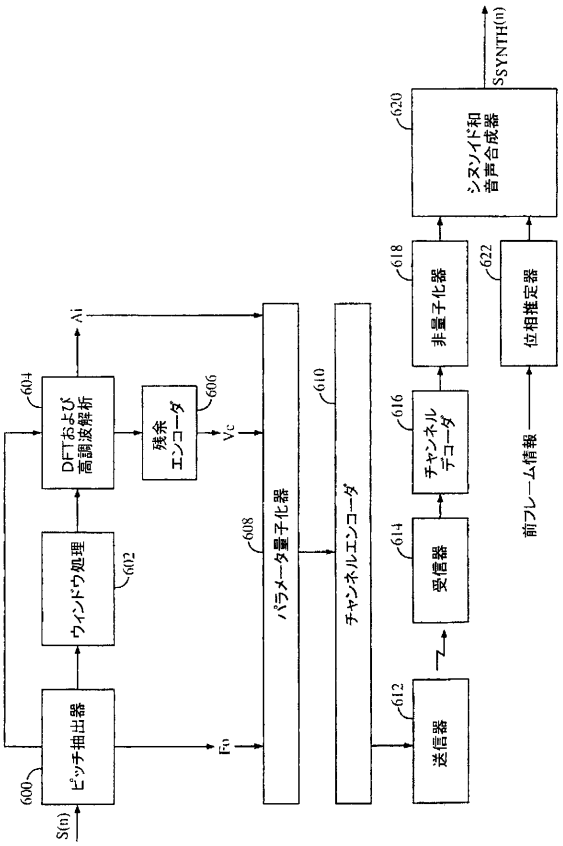
【図５】


サンプルを受信
400
エネルギーを検出
402
音声として十分なエネルギーか？
404
イエス
ノー
フレームは周期的か？
408
イエス
ノー
フレームは十分に周期的か？
412
イエス
ノー
位相ひずみは閾値よりも小さいか？
422
イエス
ノー
無音声の音声としてコード化する(1/4レート)
410
黙音としてコード化する(1/8レート)
406
遷移音声としてコード化する(フルレート)
414
有声音の音声としてコード化する(スペクトル的に、1/2レート)
416
デコード
418
入力音声サンプルと比較
420
ステップ416のパラメータを使用して、有声音の音声として送信
426

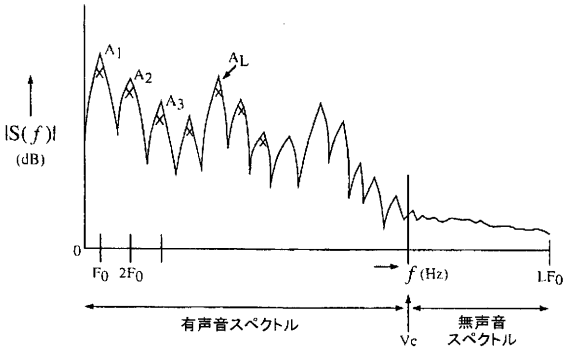
【図６】


A/D
500
フレームバッファ
502
制御プロセッサ
504
エネルギー計算器
506
有声音音声検出器
508
背景ノイズエンコーダ
510
高レートのエンコーダ
512
スペクトルエンコーダ
514
スペクトルデコーダ
516
バッファ
522
誤差計算器
518
閾値比較器
520

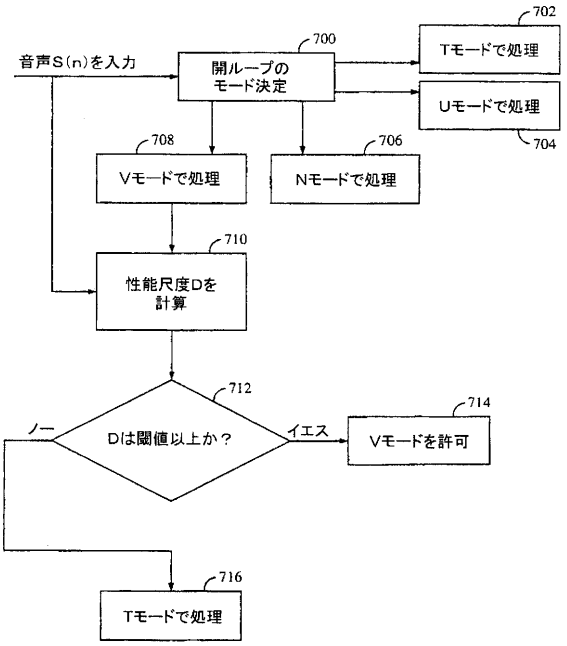
【図７】


S(n)
600
ピッチ抽出器
Fi
602
ウィンドウ処理
604
DFTおよび高調波解析
Ai
606
残余エンコーダ
Vc
608
パラメータ量子化器
610
チャンネルエンコーダ
612
送信器
614
受信器
616
チャンネルデコーダ
618
非量子化器
620
シヌソイド和音声合成器
SSYNTH(n)
622
位相推定器
前フレーム情報

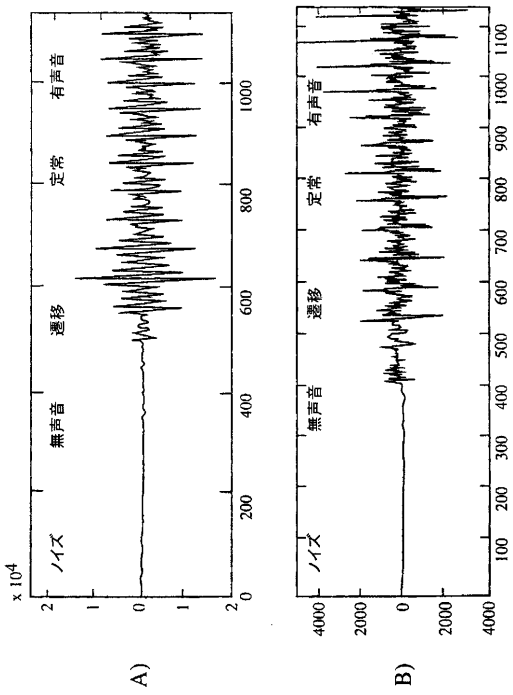
【図８】


|S(f)|
(dB)
A1
A2
A3
AL
0
F0
2F0
f (Hz)
LF0
有声音スペクトル
Vc
無声音
スペクトル

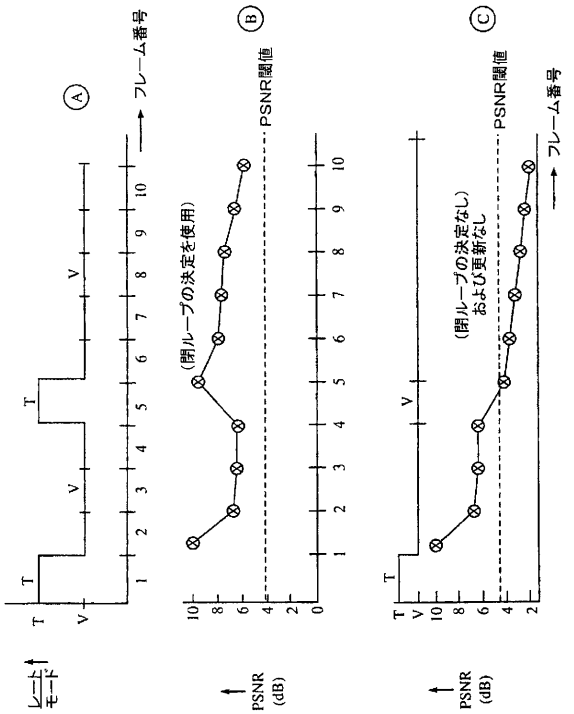
【図９】


音声S(n)を入力
700
開ループのモード決定
702
Tモードで処理
Uモードで処理
704
708
Vモードで処理
706
Nモードで処理
710
性能尺度Dを計算
712
Dは閾値以上か？
ノー
イエス
714
Vモードを許可
716
Tモードで処理

【図１０】


x 10^4
ノイズ
無声音
遷移
定常
有声音
A)
B)

【図１１】



【手続補正書】
【提出日】平成22年12月8日(2010.12.8)
【手続補正１】
【補正対象書類名】特許請求の範囲
【補正対象項目名】全文
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
少なくとも１つの時間領域コード化モードおよび少なくとも１つの周波数領域コード化モードをもつコーダと、
　コーダに接続され、かつ音声プロセッサによって処理されるフレーム内容に基づいてコーダのコード化モードを選択するように構成されている閉ループのモード選択デバイスとを含むマルチモードの混合領域の音声プロセッサ。
【手続補正２】
【補正対象書類名】明細書
【補正対象項目名】００６５
【補正方法】変更
【補正の内容】
【００６５】
　本明細書では、本発明の好ましい実施形態を示し、記載した。しかしながら、当業者の一人には、ここに記載した実施形態に対して、本発明の意図または技術的範囲から逸脱せずに多数の変更を加えられることが分かるであろう。したがって、本発明は、特許請求項にしたがうことを除いて制限されない。
　付記

（１）　少なくとも１つの時間領域コード化モードおよび少なくとも１つの周波数領域コード化モードをもつコーダと、
　コーダに接続され、かつ音声プロセッサによって処理されるフレーム内容に基づいてコーダのコード化モードを選択するように構成されている閉ループのモード選択デバイスとを含むマルチモードの混合領域の音声プロセッサ。
（２）　コーダが、音声フレームをコード化する（２）記載の音声プロセッサ。
（３）　コーダが、音声フレームの線形予測残余をコード化する（３）記載の音声プロセッサ。
（４）　少なくとも1つの時間領域コード化モードが、第１のコード化レートでフレームをコード化するコード化モードを含み、少なくとも1つの周波数領域コード化モードが、第２のコード化レートでフレームをコード化するコード化モードを含み、第２のコード化レートが第１のコード化レートよりも低い（１）記載の音声プロセッサ。
（５）　少なくとも1つの周波数領域コード化モードが、高調波のコード化モードを含む（１）記載の音声プロセッサ。
（６）　コーダに接続された比較回路であって、コード化されていないフレームを、少なくとも１つの周波数領域コード化モードでコード化されたフレームと比較して、比較に基づいて性能尺度を生成する比較回路をさらに含み、コーダが、性能尺度が所定の閾値よりも低いときだけ、少なくとも１つの時間領域コード化モードを適用し、さもなければコーダは、少なくとも１つの周波数領域コード化モードを適用する請求項１記載の音声プロセッサ。
（７）　コーダが、少なくとも１つの時間領域コード化モードを、少なくとも１つの周波数領域コード化モードでコード化された所定数の連続的に処理されるフレームの直ぐ後の各フレームに適用する（１）記載の音声プロセッサ。
（８）　少なくとも1つの周波数領域コード化モードが、周波数、位相、および振幅を含む１組のパラメータをもつ複数のシヌソイドで各フレームの短期間のスペクトルを表わし、位相は多項式表現および初期位相値でモデル化されていて、初期位相値が、（ａ）前フレームが、少なくとも1つの周波数領域コード化モードでコード化されたときは、前フレームの推定された最終位相値であるか、または（ｂ）前フレームが、少なくとも１つの時間領域コード化モードでコード化されたときは、前フレームの短期間のスペクトルから求められる位相値である（１）記載の音声プロセッサ。
（９）　各フレームにおけるシヌソイドの周波数が、フレームのピッチ周波数の整数倍である（８）記載の音声プロセッサ。
（１０）　各フレームにおけるシヌソイドの周波数が、０ないし２πの1組の実数から得られる（８）記載の音声プロセッサ。
（１１）　フレームを処理する方法であって、
　開ループのコード化モード選択プロセスを各連続する入力フレームへ適用して、入力フレームの音声内容に基づいて、時間領域コード化モードか、または周波数領域コード化モードの何れかを選択するステップと、
　入力フレームの音声内容が定常状態の有声音の音声を示すときは、入力フレームを周波数領域でコード化するステップと、
　入力フレームの音声内容が定常状態の有声音の音声以外のものを示すときは、入力フレームを時間領域でコード化するステップと、
　周波数領域でコード化されたフレームと入力フレームとを比較して、性能尺度を求めるステップと、
　性能尺度が所定の閾値よりも低いときは、入力フレームを時間領域でコード化するステップとを含むフレームを処理する方法。
（１２）　フレームが、線形予測残余フレームである（１１）記載の方法。
（１３）　フレームが音声フレームである（１１）記載の方法。
（１４）　時間領域でコード化するステップが、第1のコード化レートでフレームをコード化することを含み、周波数領域でコード化するステップが、第２のコード化レートで

フレームをコード化することを含み、第２のコード化レートが第１のコード化レートよりも低い（１１）記載の方法。

（１５） 周波数領域でコード化するステップが、高調波でコード化することを含む（１１）記載の方法。

（１６） 周波数領域でコード化するステップが、周波数、位相、および振幅を含む１組のパラメータをもつ複数のシヌソイドで各フレームの短期間のスペクトルを表わし、位相は多項式表現および初期位相値でモデル化されていて、初期位相値が、（ａ）前フレームが周波数領域でコード化されたときは、前フレームの推定された最終位相値であるか、または（ｂ）前フレームが時間領域でコード化されたときは、前フレームの短期間のスペクトルから求められる位相値である（１１）記載の方法。

（１７） 各フレームのシヌソイド周波数が、フレームのピッチ周波数の整数倍である（１６）記載の方法。

（１８） 各フレームのシヌソイド周波数が、０ないし２πの1組の実数から得られる（１６）記載の方法。

（１９） マルチモードの混合領域の音声プロセッサであって、

開ループのコード化モード選択プロセスを入力フレームへ適用して、入力フレームの音声内容に基づいて、時間領域コード化モードか、または周波数領域コード化モードの何れかを選択する手段と、

入力フレームの音声内容が定常状態の有声音の音声を示すときは、入力フレームを周波数領域でコード化する手段と、

入力フレームの音声内容が定常状態の有声音の音声以外のものを示すときは、入力フレームを時間領域でコード化する手段と、

周波数領域でコード化されたフレームと入力フレームとを比較して、性能尺度を求める手段と、

性能尺度が所定の閾値よりも低いときは、入力フレームを時間領域でコード化する手段とを含むマルチモードの混合領域の音声プロセッサ。

（２０） フレームが線形予測残余フレームである（１９）記載の音声プロセッサ。

（２１） 入力フレームが音声フレームである（１９）記載の音声プロセッサ。

（２２） 時間領域でコード化する手段が、第1のコード化レートでフレームをコード化する手段を含み、周波数領域でコード化する手段が、第２のコード化レートでフレームをコード化する手段を含み、第２のコード化レートが第１のコード化レートよりも低い（１９）記載の音声プロセッサ。

（２３） 周波数領域でコード化する手段が、高調波コーダを含む（１９）記載の音声プロセッサ。

（２４） 周波数領域でコード化する手段が、周波数、位相、および振幅を含む１組のパラメータをもつ複数のシヌソイドで各フレームの短期間のスペクトルを表わす手段を含み、位相は多項式表現および初期位相値でモデル化されていて、初期位相値が、（ａ）直前のフレームが周波数領域でコード化されたときは、直前のフレームの推定された最終位相値であるか、または（ｂ）直前のフレームが時間領域でコード化されたときは、直前のフレームの短期間のスペクトルから求められる位相値である（１９）記載の音声プロセッサ。

（２５） 各フレームのシヌソイド周波数が、フレームのピッチ周波数の整数倍である（２４）記載の音声プロセッサ。

（２６） 各フレームのシヌソイド周波数が、０ないし２πの1組の実数から得られる（２４）記載の音声プロセッサ。

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 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 アミタバ・ダス
アメリカ合衆国 カリフォルニア州 ９２１３１ サン・ディエゴ、ラムスデル・コート １１７３８

Ｆターム(参考) 5J064 AA02 BC11 BC14 BC16 BC22 BD02 BD03

【外国語明細書】
2011090311000001.pdf