

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4843124号
(P4843124)

(45) 発行日 平成23年12月21日(2011.12.21)

(24) 登録日 平成23年10月14日(2011.10.14)

(51) Int.Cl.

F 1

G 1 O L 19/00 (2006.01)
G 1 O L 21/04 (2006.01)G 1 O L 19/00 212
G 1 O L 21/04 130A

請求項の数 27 (全 20 頁)

(21) 出願番号 特願平10-548950
 (86) (22) 出願日 平成10年5月15日(1998.5.15)
 (65) 公表番号 特表2001-525079(P2001-525079A)
 (43) 公表日 平成13年12月4日(2001.12.4)
 (86) 國際出願番号 PCT/GB1998/001414
 (87) 國際公開番号 WO1998/052187
 (87) 國際公開日 平成10年11月19日(1998.11.19)
 審査請求日 平成17年4月25日(2005.4.25)
 (31) 優先権主張番号 97303321.0
 (32) 優先日 平成9年5月15日(1997.5.15)
 (33) 優先権主張国 歐州特許庁(EP)

(73) 特許権者
 ヒューレット・パッカード・カンパニー
 アメリカ合衆国カリフォルニア州9430
 4 パロアルト ハノーバー・ストリート
 3000
 (74) 復代理人
 弁理士 西山 清春
 (74) 代理人
 弁理士 古谷 聰
 (74) 代理人
 弁理士 古谷 韶
 (74) 代理人
 弁理士 溝部 孝彦

最終頁に続く

(54) 【発明の名称】音声信号を符号化及び復号化するためのコーデック及び方法

(57) 【特許請求の範囲】

【請求項1】

エンコーダ及びデコーダを含む、音声信号を符号化及び復号化するためのコーデックであつて、

前記エンコーダが、

前記音声信号を高副帯域信号と低副帯域信号に分解するためのフィルタ手段と、

前記低副帯域信号を符号化するための低副帯域コーディング手段と、

前記高副帯域信号の雑音成分のみを符号化するための高副帯域コーディング手段と、

前記符号化された低副帯域信号と前記符号化された高副帯域信号の雑音成分を結合して結合された符号化信号を生成するための手段

を備え、

前記デコーダ手段が、前記符号化された低副帯域信号及び前記符号化された高副帯域信号の雑音成分を復号化し、復号化された信号から音声出力信号を再生するための手段を備え、

前記復号化する手段が、フィルタ手段を備え、該フィルタ手段に、前記結合された符号化信号から得られたスペクトルパラメータを加えると共に、該フィルタ手段を励起信号で励起することによって、合成非高調波高副帯域信号を生成するよう動作し、前記励起信号がホワイトノイズ信号のみからなり、前記合成非高調波高副帯域信号と復号化された低副帯域信号が再結合されて音声出力信号を形成することからなる、コーデック。

【請求項2】

10

20

前記デコーダ手段が、符号化された低副帯域信号及び高副帯域信号をそれぞれ受信して復号化するための、低副帯域復号化手段及び高副帯域復号化手段とを備える、請求項1に記載のコーデック。

【請求項3】

前記高副帯域コーディング手段が、高副帯域エネルギーまたは利得値と、1つ以上の高副帯域スペクトルパラメータとを得るために前記高副帯域信号を分析して符号化するための手段を備える、請求項1又は2に記載のコーデック。

【請求項4】

前記1つ以上の高副帯域スペクトルパラメータが2次のLPC係数を含む、請求項3に記載のコーデック。

10

【請求項5】

前記符号化手段が、前記高副帯域信号中のエネルギーを測定し、これにより前記高副帯域エネルギーまたは利得値を導き出すための手段を備える、請求項3又は4に記載のコーデック。

【請求項6】

前記符号化手段が、前記高副帯域信号中の雑音成分のエネルギーを測定し、これにより前記高副帯域エネルギーまたは利得値を導き出すための手段を備える、請求項3又は4に記載のコーデック。

【請求項7】

前記高副帯域信号中の前記エネルギーをモニタし、当該エネルギーを前記高副帯域エネルギーと低副帯域エネルギーのうちの少なくとも1つから得られたしきい値と比較し、さらに、前記モニタしたエネルギーが前記しきい値よりも低い場合は、前記高副帯域符号化手段に最小符号出力を提供させるための手段を備える、請求項5又は6に記載のコーデック。

20

【請求項8】

前記低副帯域コーディング手段が、スピーチコーダと、有声・無声判定を行なうための手段を備える、請求項1～7のいずれか1項記載のコーデック。

【請求項9】

前記デコーダ手段が、音声信号が有声であるか無声であるかに応じて、前記励起信号中の雑音エネルギーを調節するための符号化された高副帯域信号中のエネルギー及び前記有声・無声判定に応答する手段を備える、請求項8に記載のコーデック。

30

【請求項10】

前記低副帯域コーディング手段がMPEG音声コーダを含む、請求項1～7のいずれか1項記載のコーデック。

【請求項11】

前記高副帯域が、2.75kHzよりも高い周波数を含み、前記低副帯域が2.75kHzよりも低い周波数を含む、請求項1～10のいずれか1項記載のコーデック。

【請求項12】

前記高副帯域が4kHzよりも高い周波数を含み、前記低副帯域が4kHzよりも低い周波数を含む、請求項1～10のいずれか1項記載のコーデック。

40

【請求項13】

前記高副帯域が5.5kHzよりも高い周波数を含み、前記低副帯域が5.5kHzよりも低い周波数を含む、請求項1～10のいずれか1項記載のコーデック。

【請求項14】

前記高副帯域コーディング手段が、800bpsよりも低いビットレート、好ましくは約300bpsのビットレートで前記雑音成分を符号化する、請求項1～13のいずれか1項記載のコーデック。

【請求項15】

前記高副帯域信号が、前記スペクトルパラメータを決定するために長いフレーム周期で分析され、前記エネルギーまたは利得値を決定するために短いフレーム周期で分析される、

50

請求項 3 ~ 14 のいずれか 1 項記載の コードック。

【請求項 16】

音声信号を符号化及び復号化するための方法であって、
 前記音声信号を高副帯域信号と低副帯域信号に分解するステップと、
 前記低副帯域信号を符号化するステップと、
 前記高副帯域信号の雑音成分のみを符号化するステップと、
 前記符号化された低副帯域信号と前記符号化された高副帯域信号の雑音成分を結合して、
 結合された符号化信号を生成するステップと、
 前記符号化された低副帯域信号及び前記符号化された高副帯域信号の雑音成分を復号化して音声出力信号を再生するステップ
 を含み、
 前記再生するステップが、

前記結合された符号化信号からスペクトルパラメータを得るステップと、
 前記ホワイトノイズ信号のみからなる励起信号を提供するステップと、
 前記励起信号でフィルタを励起すると共に、前記スペクトルパラメータを該フィルタに加えて、合成非高調波副帯域信号を生成するステップと、
 前記合成非高調波副帯域信号と前記復号化された低副帯域信号を結合するステップ
 を含むことからなる、方法。

【請求項 17】

音声信号を符号化するためのエンコーダと音声信号を復号化するためのデコーダを備える
 コードックであって、
 前記エンコーダが、
 前記音声信号を高副帯域信号と低副帯域信号に分解するための手段と、
 前記低副帯域信号を符号化するための手段と、
 前記高副帯域信号の雑音成分のみを符号化するための手段と、
 前記符号化された低副帯域信号と前記符号化された高副帯域信号の雑音成分を結合して、
 結合された符号化信号を生成するための手段
 を備え、
 前記デコーダが、

前記符号化された低副帯域信号を復号化するための狭帯域デコーダと、
 前記結合された符号化信号から得られたスペクトルパラメータを受け、かつ、ホワイトノイズ信号によって励起されて、合成非高調波高副帯域信号を生成するためのフィルタと、
 前記雑音成分のエネルギー値に基づいて前記合成非高調波高副帯域信号の利得を調整するための手段と、
 前記復号化された低副帯域信号と前記利得を調整された合成非高調波高副帯域信号を結合して合成音声信号を生成するための手段
 を備えることからなる、コードック。

【請求項 18】

音声信号を符号化及び復号化する方法であって、
 前記音声信号を高副帯域信号と低副帯域信号に分解するステップと、
 前記低副帯域信号を符号化するためステップと、
 前記高副帯域信号の雑音成分のみを符号化するステップと、
 前記符号化された低副帯域信号と前記符号化された高副帯域信号の雑音成分を結合して、
 結合された符号化信号を生成するステップと、
 前記符号化された低副帯域信号を復号化するステップと、
 フィルタ手段に前記結合された符号化信号から得られたスペクトルパラメータを加え、かつ、前記フィルタ手段をホワイトノイズ信号で励起して、合成非高調波高副帯域信号を生成するステップと、
 前記雑音成分のエネルギー値に基づいて前記合成非高調波高副帯域信号の利得を調整するステップと、

10

20

30

40

50

前記復号化された低副帯域信号と前記利得を調整された合成非高調波高副帯域信号を結合して合成音声信号を生成するステップ
を含む、方法。

【請求項 19】

エンコーダ手段及びデコーダ手段を含む、スピーチ信号を符号化及び復号化するためのコードックであって、

前記エンコーダ手段が、

前記スピーチ信号を、少なくとも5.5kHzの帯域幅を合わせて画定する低副帯域及び高副帯域に分解するためのフィルタ手段と、

前記低副帯域について高次のボコーダ分析を実施して、前記低副帯域を表わすLPC係数を含むボコーダ係数を得るために低副帯域ボコーダ分析手段と、

10

前記高副帯域について低次のボコーダ分析を実施して、前記高副帯域を表わすLPC係数を含むボコーダ係数を得るために高副帯域ボコーダ分析手段と、

前記低副帯域係数及び高副帯域係数を含むボコーダパラメータを符号化して、記憶及び/又は伝送用に符号化信号を提供するためのコーディング手段を備え、

前記デコーダ手段が、

前記符号化信号を復号化して、前記低副帯域ボコーダ係数と高副帯域ボコーダ係数を結合する1組のボコーダパラメータを得るために復号化手段と、

前記1組のボコーダパラメータからLPCフィルタを構成し、前記フィルタ及び励起信号から前記スピーチ信号を合成するための合成手段

20

とを備え、前記励起信号は、有声スピーチについては周期的な励起信号であり、無声スピーチについてはホワイトノイズ信号であることからなる、コードック。

【請求項 20】

前記低副帯域ボコーダ分析手段及び前記高副帯域ボコーダ分析手段がLPCボコーダ分析手段である、請求項19に記載のコードック。

【請求項 21】

前記低副帯域LPC分析手段が10次以上の分析を実施する、請求項20に記載のコードック。

【請求項 22】

前記高帯域LPC分析手段が2次の分析を実施する、請求項20又は21に記載のコードック。

30

【請求項 23】

前記合成手段が、前記低副帯域及び前記高副帯域を再合成し、前記再合成された低副帯域と高副帯域を結合するための手段を備える、請求項19~22のいずれか1項記載のコードック。

【請求項 24】

前記合成手段が、低副帯域及び高副帯域のパワースペクトル密度をそれぞれ判定するための手段と、高次のLPCモデルを得るために前記パワースペクトル密度を結合するための手段とを備える、請求項23に記載のコードック。

【請求項 25】

前記結合するための手段が、前記結合されたパワースペクトル密度の自己相関を決定するための手段を備える、請求項24に記載のコードック。

40

【請求項 26】

前記結合するための手段が、前記低副帯域及び高副帯域のパワースペクトル密度関数の自己相関をそれぞれ決定し、前記自己相関を結合するための手段を備える、請求項25に記載のコードック。

【請求項 27】

スピーチ信号を符号化するための音声エンコーダと符号化されたスピーチ信号を合成するための音声デコーダからなるコードックであって、

前記音声エンコーダが、

50

前記スピーチ信号を低副帯域と高副帯域に分解するためのフィルタ手段と、
前記低副帯域信号について高次のボコーダ分析を実施して、前記低副帯域を表わすボコーダ係数を得るための低帯域ボコーダ分析手段と、
前記高副帯域信号について低次のボコーダ分析を実施して、前記高副帯域を表わすボコーダ係数を得るための高帯域ボコーダ分析手段と、
前記低副帯域ボコーダ係数及び高副帯域ボコーダ係数を符号化して、記憶及び／又は伝送用に符号化信号を提供するためのコーディング手段
を備え、
前記符号化されたスピーチ信号が、低副帯域及び高副帯域に関するLPC係数を含むパラメータを含み、

10

前記音声デコーダが、
前記符号化された信号を復号化して、前記低副帯域LPC係数及び高副帯域LPC係数を結合する1組のLPCパラメータを得るための復号化手段と、
前記高副帯域及び低副帯域に関する1組のLPCパラメータからLPCフィルタを構成し、前記スピーチ信号を前記フィルタ及び励起信号から合成するための合成手段
を備え、前記励起信号は、有声スピーチについては周期的な励起信号であり、無声スピーチについてはホワイトノイズ信号であることからなる、コーデック。

【発明の詳細な説明】

技術分野

本発明は、音声符号化装置及び方法に関し、より具体的には音声信号を低ビットレートで符号化するシステム及び方法に関するが、これに限定されない。

20

発明の背景

広範囲のアプリケーションにおいて、例えばコンピュータや携帯用口述記録機器、パーソナルコンピュータ機器等のメモリ容量を節約するために、音声信号を低ビットレートで効率的に記憶する設備を設けることが望ましい。同様に、例えばビデオ会議、オーディオストリーミング又はインターネットを介した電話通信等で音声信号を伝送する場合、低ビットレートであることが非常に望ましい。しかしながらいずれの場合においても明瞭度や品質が重要であり、したがって本発明は高いレベルの明瞭度及び品質を保ちつつ、非常に低いビットレートで符号化することの問題、また更にスピーチ及び音楽の両方を低ビットレートで充分満足に処理することができる符号化システムを提供するという問題を解決することに関するものである。

30

スピーチ信号で非常に低いビットレートを実現するためには、波形コーダではなくパラメトリックコーダ、即ち「ボコーダ」を利用すべきであることが一般的に知られている。ボコーダは、波形それ自体ではなく波形のパラメータのみを符号化し、スピーチのように聞こえはするものの、潜在的には非常に異なる波形を持つ信号を生成する。

代表的な例としては、T.E.Tremaineによる「The Government Standard Linear Predictive Coding Algorithm」:LPC10;Speech Technology、pp40 - 49、1982に記述のLPC 10ボコーダ(Federal Standard 1015)が挙げられる。これは同様のアルゴリズムであるLPC 10eに引き継がれているが、両者とも本願に参考資料として取り入れられる。LPC 10及びその他のボコーダは、従来電話周波数帯域(0 ~ 4 kHz)において動作されてきているが、それはスピーチを聞き取れるようにするために必要な情報を全てこの帯域幅に含むと考えられているためである。しかしながら我々は、この方法で2.4Kbit/sもの低いビットレートで符号化されたスピーチの音声品質と明瞭度が、現在の商業用アプリケーションの多くに適していないことを見いだした。

40

音声品質の向上にはスピーチモデルにおいてより多くのパラメータを必要とするが、これらの追加パラメータ手段を符号化しようとすると、既存のパラメータに使えるビットがより少なくなるという問題が生じる。LPC 10eモデルには、例えばA.V.McCree及びT.P.Barnwellによる「A Mixed Excitation LPC Vecoder Model for Low Bit Rate Speech Coding」;IEEE - Trans Speech and Audio Processing、Vol.3、No.4、1995年7月のように様々な強化策が提案されているが、これら全てを利用したとしても音声品質はわずかに適正

50

化されるにすぎない。

このモデルをさらに強化するために、我々は、より広い帯域幅（0～8 kHz）を符号化することに着目した。このことはボコーダについては考慮されたことがなかったが、これはより高い帯域幅の符号化に要する追加ビットが符号化による恩恵を大きく打ち消してしまうかのように見えるためである。広帯域幅の符号化は通常高品質コーダについてのみ考慮されており、これは明瞭度を増すというよりはむしろスピーチがより自然に聞こえるようにしたものであり、多くの追加ビットを必要とする。

広帯域のシステムを実現するための1つの一般的方法としては、信号を低副帯域及び高副帯域に分割し、高副帯域をより少ないビットで符号化できるようにする方法が挙げられる。ITU標準G722 (X.Maitreによる「7 kHz Audio Coding Within 64Kbit/s」(IEEE Journal on Selected Areas in Comm., Vol.6, No.2, pp283 - 298, 1988年2月)に記述されているように、2つの帯域は別々に復号化され、その後一つに合わせられる。この手法をボコーダに適用した場合、高帯域幅は低帯域幅よりも低次のLPCで分析されるべきであることが示唆された（我々は2次が適切であること見いだした）。それには別々のエネルギー値が必要であるが、低帯域幅からのものが利用できるために、別のピッチや有声・無声判定は必要ないことを、我々は見いだした。残念ながら、我々の推論するところ、2つの帯域間の位相の不整合が原因で2つの合成帯域を再結合することによりアーチファクトが発生してしまった。このデコーダにおける問題を、我々は各帯域のLPC及びエネルギーパラメータを組み合わせ、単一の高次広帯域フィルタを作り、これを広帯域励起信号で駆動することにより解決した。

驚くべきことに、純粋なスピーチに対する広帯域LPCボコーダの明瞭度は、同じビットレートの電話周波数帯域のものと比べると著しく高く、DRTスコア (W.D.Voiersによる「Diagnostic Evaluation of Speech Intelligibility」、in Speech Intelligibility and Speaker Recognition (M.E. Hawley, ed.) pp374 - 387, Dowden, Hutchinson&Ross, Inc., 1977) に記載）は、狭帯域コーダの84.4に比して86.8であった。

しかしながらバックグラウンド雑音が小さいスピーチにあってさえも、合成信号にはバズが目立ち、高帯域にアーチファクトが含まれていた。これは我々の解析結果から、符号化された高帯域エネルギーがバックグラウンド雑音により高められ、これが有声スピーチを合成する間に高帯域高調波を高めてバズ作用を生じることがわかった。

さらに詳細な調査の結果、我々は、明瞭度を向上させるには有声部分ではなく、主に無声の摩擦音や破裂音をより良好に復号化すればよいことを見いだした。このことにより我々の方向性は、雑音のみを合成し、有声スピーチの高調波を低帯域のみに限定するという、異なる高帯域復号化手法へと導かれた。これによりバズは除去されたが、復号化された高帯域エネルギーが高い場合、かわりに入力信号中の高帯域高調波が原因でヒスが加わってしまう場合があった。これは有声・無声判定を用いて解決可能であったが、我々は、最も信頼できる方法が、高帯域入力信号を雑音及び高調波（周期性）成分に分け、雑音成分のエネルギーのみを復号化することを見いだした。

この手法は、この技術の効力を大幅に強化する2つの思いがけない利益をもたらした。第一には、高帯域が雑音しか含んでいないため、高及び低帯域の位相を整合させる問題を解消したことであり、これはボコーダについてでさえ、それらを完全に分けて合成することができるることを意味する。実際、低帯域用のコーダは完全に別個のものでよく、市販の部品であっても良い。第二には、いかなる信号も雑音と高調波成分に分割することができるため、高帯域の符号化はスピーチに固有のものではなくなり、そうでなければその周波数帯域は再生される可能性が全く無かったところが、雑音成分再生の恩恵を受けることができる。これは強いパーカッション成分を含むロック音楽において特に言えることである。本システムは根本的に、McElroyらによる「Wideband Speech Coding in 7.2KB/s」(ICASSP 93, pp11-620 - 11-623) のような波形符号化に基づいた他の広帯域拡張技術とは異なる手法によるものである。波形符号化の問題は、G722 (Supra) のように多数のビットを必要とするか、さもなければ高帯域信号の不十分な再生 (McElroyら) によって大量の量子化雑音を高調波成分に加えることになるかのいずれかである点にある。

10

20

30

40

50

本願において「ボコーダ」という語は、選択されたモデルパラメータを符号化し、その中に残差波形の明示的な符号化を行わない、スピーチコーダを広義的に画定するのに使用され、またこの語には、スピーチスペクトルを複数の帯域に分割し、各帯域の基本パラメータセットを抽出することによって符号化を行う多帯域励起コーダ(MBE)も含まれる。

ボコーダ分析という語は、少なくとも線形予測符号化(LPC)係数及びエネルギー値を含むボコーダ係数を決定するプロセスを説明するために用いられる語である。また加えて低副帯域については、ボコーダ係数は有声・無声判定、さらに有声スピーチにはピッチ値を含む場合がある。

発明の開示

本発明の一態様によれば、エンコーダ及びデコーダを含む、音声信号を符号化及び復号化するための音声符号化システムが提供され、

前記エンコーダが：

前記音声信号を高副帯域信号及び低副帯域信号へと分解するための手段と；

前記低副帯域信号を符号化するための低副帯域符号化手段と；

ソースフィルタモデルに基づいて前記高副帯域信号の少なくとも非周期成分を符号化するための高副帯域符号化手段と；を含み、

前記デコーダ手段が、前記符号化された低副帯域信号及び前記符号化された高副帯域信号とを復号化するための、そしてそこから音声出力信号を再生するための復号化するための手段を含み、

前記復号化手段が、フィルタ手段と、そして前記フィルタ手段に通す励起信号を生成して合成音声信号を生成するための励起手段とを含み、該励起手段が、前記音声信号の高副帯域に対応する周波数帯域中の合成雑音の実質的成分を含む励起信号を生成するように作動可能であることを特徴とする。

復号化手段は、高及び低副帯域とともに変換するための単一の復号化手段から構成することができ、復号化手段として望ましいのは、符号化された低及び高副帯域信号をそれぞれに受信して復号化するための低副帯域復号化手段と高副帯域復号化手段とから構成されたものである。

特定の実施例においては、前記励起信号の前記高周波数帯域は実質的に全体が合成雑音信号により構成されているが、他の実施例においては、励起信号は合成雑音成分と、前記低副帯域音声信号の1つ以上の高調波に対応するさらなる成分とを混合したものから構成されている。

高副帯域エネルギー即ち利得値と、1つ以上の高副帯域スペクトルパラメータとを得るために好都合なように、高副帯域符号化手段は前記高副帯域信号を分析し及び符号化するための手段を備えている。1つ以上の高副帯域スペクトルパラメータはできれば2次LPC係数からなることが望ましい。

前記符号化手段が前記高副帯域における雑音エネルギーを測定する手段を含むことが望ましく、これにより前記高副帯域エネルギー即ち利得値を推論することが望ましい。代替的には前記符号化手段は、前記高副帯域信号中の全体のエネルギーを測定するための手段を含み、これにより前記高副帯域エネルギー即ち利得値を導き出す。

ビットレートの不必要的な使用を省くために、システムは、前記高副帯域信号中の前記エネルギーをモニタし、これを高及び低副帯域エネルギーの少なくとも1つから得たしきい値と比較し、そして前記モニタされたエネルギーが前記しきい値よりも低い場合に、前記高副帯域符号化手段に最低符号出力を供給する手段を含むことが望ましい。

主にスピーチの符号化を意図した構成においては、前記低副帯域符号化手段は有声・無声判定を行うための手段を含むスピーチコーダを含む。この場合、前記復号化手段は、前記高帯域符号化信号中のエネルギー及び前記有声・無声判定に応答して、音声信号が有声か無声かに依存する前記励起信号中の雑音エネルギーを調節する手段を含む。

システムが音楽用に意図されたものであれば、前記低副帯域符号化手段は、例えばMPEG音声コーダのような適当な波形コーダをいずれかの数量備える。

高及び低副帯域間の分割は特定の条件に基づいて選択され、したがって、約2.75kHz、4k

10

20

30

40

50

Hz、5.5kHz等が選択される。

前記高副帯域符号化手段は、前記雑音成分を800bpsよりも小さい、望ましくは300bps程度の非常に低いビットレートで符号化することが望ましい。

エネルギー利得値及び1つ以上のスペクトルパラメータを得るために高副帯域を分析する場合、前記高副帯域信号を前記スペクトルパラメータの決定には相対的に長いフレーム周期で、そして前記エネルギー即ち利得値の決定には相対的に短いフレーム周期で分析することが望ましい。

他の様において本発明は、入力信号が副帯域へと分割され、それぞれのボコーダ係数が得られ、その後再結合されてLPCフィルタに送られる、非常に低いビットレートで符号化するためのシステム及び方法を提供する。

したがってこの様においては、本発明は4.8kbit/s未満のビットレートで信号を圧縮し、またその信号を再合成するためのボコーダシステムが提供される。このシステムは符号化手段及び復号化手段を含み、該符号化手段が；

前記スピーチ信号を、ともに少なくとも5.5kHzの帯域幅を画定する低及び高副帯域へと分解するためのフィルタ手段と；

相対的に高次のボコーダ分析を前記低副帯域に実施して、前記低副帯域を表わすボコーダ係数を得るための低副帯域ボコーダ分析手段と；

相対的に低次のボコーダ分析を前記高副帯域に実施して、前記高副帯域を表わすボコーダ係数を得るための高副帯域ボコーダ分析手段と；

前記低及び高副帯域係数を含むボコーダパラメータを符号化して、記憶及び／又は伝送用に圧縮信号を供給するための符号化手段とを含み；さらに

前記復号化手段が；

前記圧縮信号を復号化して、前記低及び高副帯域ボコーダ係数を含むボコーダパラメータを得るための復号化手段と；

前記高及び低副帯域に関するボコーダパラメータからLPCフィルタを構成し、前記スピーチ信号を前記フィルタ及び励起信号から再合成するための合成手段とを含むことを特徴とする。

前記低副帯域分析手段は10次のLPC分析を適用し、前記高副帯域分析手段は2次のLPC分析を適用する。

また本発明は、上述のシステムと共に利用する音声エンコーダ及び音声デコーダ、並びにそれらに対応する方法にも及ぶ。

上記に本発明について説明したが、本発明は上記及び以下の説明で述べられた特長のあらゆる発明的組み合わせをも包含するものである。

【図面の簡単な説明】

本発明は様々な方法で実施することができるが、単に具体例を挙げる目的のために2つの実施例及びそれらの異なる変形態を、添付の図面を参照して詳細に説明する。図面は以下の通りである。

図1は、本発明に基づく広帯域コーデックの第一の実施例のエンコーダのブロック図である。

図2は、本発明に基づく広帯域コーデックの第一の実施例のデコーダのブロック図である。

図3は、第一の実施例において利用される符号化・復号化プロセスの結果得られたスペクトルを示すものである。

図4は、男性の声のスペクトル写真である。

図5は、代表的なボコーダによって仮定されるスピーチモデルのブロック図である。

図6は、本発明に基づくコーデックの第二の実施例のエンコーダのブロック図である。

図7は、16kHzでサンプリングされた無声スピーチフレームに関する2つの副帯域の短時間スペクトルを示す。

図8は、図7の無声スピーチフレームに関する2つの副帯域のLPCスペクトルを示す。

図9は、図7及び図8の無声スピーチフレームの、結合されたLPCスペクトルを示す。

10

20

30

40

50

図10は、本発明に基づくコーデックの第二の実施例のデコーダのブロック図である。

図11は、本発明の第二の実施例において利用されるLPCパラメータ符号体系のブロック図である。

図12は、本発明の第二の実施例において使用されるLSP予測器に対する好ましい重み付け方式を示すものである。

以下の説明において、本発明に基づく2つの異なる実施例を挙げるが、その両方が副帯域復号化を用いたものである。第一の実施例においては、高帯域の雑音成分のみが符号化され、デコーダにおいて再合成されるという符号体系が用いられる。

第二の実施例は、低及び高副帯域の両方に対してLPCボコーダ方式を使用し、結合して全極フィルタを制御するためのLPCパラメータの結合セットを生成するためのパラメータを得る。

第一の実施例を説明する前に、現在の音声及びスピーチコーダについて触れると、これらは拡張帯域幅を備える入力信号を与えられた場合、単に符号化前に入力信号の帯域を限定する。本願に説明する技術は、主コーダに比較して取るに足らないビットレートで拡張帯域幅を符号化できるようにしたものである。本技術は、高副帯域を完全に再生しようと試みるものではないが、それでも主要帯域限定信号の品質（スピーチに関しては明瞭度）を著しく向上させる符号化法を提供する。

高帯域は、全極フィルタが励起信号で駆動されると、通常の方法でモデリングされる。スペクトルを記述するには1つ又は2つのパラメータしか必要としない。励起信号はホワイトノイズ及び周期成分の組み合わせであると考えられ、周期成分はホワイトノイズに対して非常に複雑な関係を持っている可能性がある（多くの音楽においてはそうである）。以下に説明するコーデックの最も一般的な形式においては、周期成分が効果的に破棄される。伝送されるのは雑音成分の予測エネルギー及びスペクトルパラメータだけであり、デコーダにおいてはホワイトノイズのみが全極フィルタの駆動に使用される。

高帯域の符号化が完全にパラメータ形式で行われることが重要であり、独自の概念である。すなわち励起信号自体の符号化は行われないということである。唯一符号化されるパラメータはスペクトルパラメータ及びエネルギーパラメータである。

本発明のこの態様は、新しい形式のコーダとして、もしくは既存のコーダーに対する広帯域拡張として実現することができる。このような既存のコーダは第三者から供給を受けても良いし、あるいは既に同じシステム上にあるものでもおそらくは良い（例：Window95/NTのACMコーデック）。その意味においては、そのコーデックを使って主信号を符号化するが、その狭帯域コーデック自体が生成する信号よりも品質の高い信号を生成させる、コーデックに対するパラサイトとして機能する。高帯域を合成するためにホワイトノイズのみを利用することの重要な特長は、2つの帯域を結合することがさして難しくないという点にある。すなわちそれらの帯域を数ミリ秒以内に合わせなければならないだけで、解決しなければならない位相の連続性の問題が存在しないのである。事実、我々は異なるコーデックを利用して数多くの実証を行なったが、信号を合わせることに何等の困難はなかった。

本発明は2つの方法で利用することができる。1つは、既存の狭帯域（4kHz）コーダの品質を、入力帯域幅を非常にわずかのビットレート増で拡張することにより改善することである。もう1つは、低帯域コーダをより小さな入力帯域幅（代表的には2.75kHz）で動作させ、さらにそれを拡張して失われた帯域幅（代表的には5.5kHz）を補償することによって、より低いビットレートのコーダを作ることである。

図1及び図2は、コーデックの第一の実施例に対するエンコーダ10及びデコーダ12をそれぞれ図示する。まず最初に図1を参照すると、入力された音声信号はローパスフィルタ14を通過するが、ここでローパスフィルタによりろ波されることで低副帯域信号が形成され、大部分が捨てられる。また入力された音声信号はハイパスフィルタ16も通過するが、ここでハイパスフィルタによりろ波されることで高副帯域信号が形成され、大部分が捨てられる。

フィルタにはシャープカットオフ及び良好なストップバンド減衰が必要である。これを達

10

20

30

40

50

成するには、73タップFIRフィルタ又は8次楕円フィルタが利用されるが、これは使用されているプロセッサ上でどちらの方が高速動作できるかにより決定される。ストップバンド減衰は少なくとも40dB、好ましくは60dBであり、通過帯域リップルは最高でも -0.2dB と小さくなくてはならない。フィルタに関して 3 dB点が目標分割点（代表的には 4 kHz）である。

低副帯域信号は狭帯域エンコーダ18に供給される。狭帯域エンコーダはボコーダもしくは周波数帯域エンコーダである。高副帯域信号は、以下に説明するが、高副帯域のスペクトルを分析してパラメータ係数及びその雑音成分を判定する高副帯域分析器20へと供給される。

スペクトルパラメータ及び雑音エネルギー値の対数は量子化され、それらの以前の値から減算（例：差分符号化）され、そしてRiceコーダ22へと符号化のために供給され、その後狭帯域エンコーダ18からの符号化された出力と結合される。

デコーダ12において、スペクトルパラメータが符号化されたデータから得られ、スペクトル形成フィルタ23に加えられる。スペクトル形成フィルタ23は合成ホワイトノイズ信号により励起され、合成非高調波高副帯域信号を生成し、その利得値は24において雑音エネルギー値に基づいて調節される。その後合成信号は、信号を補間し、それを高副帯域に反映させるプロセッサ26を通過する。低副帯域信号を表わす符号化データは狭帯域デコーダ30を通過するが、この符号化データはさらに32で補間され、34で再結合されて低副帯域信号を復号化して合成出力信号を形成する。

上記の実施例において、記憶 / 伝送機構が可変ビットレートの符号化をサポートできる場合、又は充分に大きい遅延を許容してデータを固定サイズのパケット内にブロック化される場合には、Rice符号化法が唯一適切な符号化法である。それ以外では、従来の量子化法がビットレートにあまり影響を与えることなく利用可能である。

符号化 - 復号化プロセスの全てを実施した結果を図3のスペクトルに示す。上の図はエルトン・ジョンのNakitaから得た雑音及び強い高調波成分両方を含むフレームであり、下の図は同じフレームであるが、4 ~ 8 kHzの領域を上述した広帯域拡張を使用して符号化したものである。

高副帯域のスペクトル及び雑音成分分析についてより詳細を考察すると、スペクトル分析では安定したフィルタを確実に作成するとされる標準自己相関法を利用して2つのLPC係数を導出する。量子化のために、LPC係数は反射係数へと変換され、各々9レベルで量子化される。その後これらのLPC係数は、波形を逆ろ波して雑音成分分析用の白色化信号を生成するために使用される。

雑音成分分析は複数の方法で実施可能である。例えば高副帯域は全波整流され、滑らかにされて、McCreeらの文献に記述されるような周期性についての分析を行なわれる。しかしながらその測定は、周波数領域における直接測定によってより簡単に実施される。したがって本実施例においては、256ポイントFFTを白色化された高副帯域信号に実施した。雑音成分エネルギーをFFTビンエネルギーの中央値として取った。このパラメータは重要な特性を持つ。すなわち信号が完全に雑音であった場合、中央値の期待値は単に信号のエネルギーである。しかし信号が周期成分を有している場合、平均間隔がFFTの周波数解像度の2倍よりも大きい限りは、中央値がスペクトル中のピークの間に来ることになる。しかし間隔が非常に狭い場合、かわりにホワイトノイズが使われていると、人の耳は小さな違いを認識する。

スピーチ（及び音声信号の一部）については、LPC分析よりもより短い間隔で雑音エネルギー計算を行なう必要がある。これは破裂音の急激な発生のため、そして無声スペクトルがあまり速く動かないためである。このような場合、FFTのエネルギーに対する中央値の比率（例えばわずかな雑音成分等）が測定される。これはその後、その分析周期に対する測定エネルギー値全てをスケーリングするために利用される。

雑音 / 周期判別は不完全であり、そして雑音成分分析それ自体も不完全である。これを許容するために、高副帯域分析器20は高帯域中のエネルギーを約50%の固定因数でスケーリングする。元の信号を復号化された拡張信号と比べると、高音域調整を若干下げたよう

10

20

30

40

50

聞こえる。しかし非拡張方式で復号化した信号における高音域の完全排除に比較すると、その差異はとるにたらない程度である。

通常雑音成分の再生は、雑音成分が高帯域中の高調波エネルギーと比べて小さい場合、又は低帯域中のエネルギーと比べて非常に小さい場合には行なう意味がない。前者の場合には、FFTピクセル間における信号リークにより、雑音成分の正確な測定はどんな方法を用いても難しい。これはまた、後者の場合においても低域フィルタのストップバンドにおける限られた減衰のためにある程度同じことが言える。したがって本実施例の修正形態において、高副帯域分析器20が測定された高副帯域雑音エネルギーを、高及び低副帯域エネルギーの少なくともいずれか1つから得たしきい値と比較し、それがしきい値よりも低い場合、雑音下限エネルギー値がかわりに伝送される。雑音下限エネルギー値とは、高帯域におけるバックグラウンド雑音レベルの推定値であり、通常これは出力信号の開始から測定された最低の高帯域エネルギー値に等しく設定される。10

次にこの実施例における性能を考察する。図4は男性の声のスペクトル写真である。周波数を示す縦軸は8000Hzに達しており、これは標準の電話コーダ（4kHz）範囲の2倍である。図中の暗い部分はその周波数における信号強度を表わしている。横軸は時間を表わしている。

4kHzより上においては信号は殆どが摩擦音もしくは破裂音からの雑音であるか、全く存在していないかであることが分かる。この場合における広帯域拡張は、高帯域のほぼ完全な再生を行なう。

女性の一部及び子供の声については、4kHzより高い周波数において有声スピーチがそのエネルギーの殆どを失う。この場合、理想的には若干高め（5.5kHz程度が良い）で帯域分割を行なうことが望ましい。しかし、そのようにしなくとも品質は無声スピーチにおいては非拡張コーデックよりも良好であり、有声スピーチでは全く同じである。さらに明瞭度の向上は摩擦音や破裂音の良好な再生から得られるものであり、母音のより良い再生からではないため、したがって分割点は音声品質に影響を与えるだけで明瞭度に影響することはない。20

音楽の再生については、広帯域拡張法の効果は音楽の種類に多少依存する。最も顕著な高帯域成分が打楽器や声（特に女性の声）の「柔らかさ」に由来するロック／ポップスについては、音をところどころで強調したとしても、雑音のみの合成が非常に効果的である。その他の音楽は、例えばピアノ演奏などのように高帯域には高調波成分しか持たない。この場合、高帯域では何も再生されない。しかしながら本質的に、低周波数の高調波が多く存在すれば、高周波数の欠如は音にとってあまり重要ではないようである。30

次に、図5～図12を参照して説明されるコーデックの第二の実施例を考察する。この実施例は周知のLPC10ボコーダ（T.E.Tremainの「The Government Standard Linear Predictive Coding Algorithm:LPC10」；Speech Technology、pp40-49、1982に記載）と同様の概念を基本としており、LPC10ボコーダが採用するスピーチモデルを図5に示す。全極フィルタ110としてモデリングされるボーカルトラクトは、有声スピーチについては周期的な励起信号112により、そして無声スピーチについてはホワイトノイズ114により駆動される。ボコーダはエンコーダ116及びデコーダ118の2つの部分から構成される。図6に示されるエンコーダ116は、入力スピーチを等しい時間間隔をおいたフレームへと分割する。その後各フレームは、スペクトルの0～4kHz及び4～8kHzの領域に対応する帯域へと分割される。これは計算的に効率的な方法で8次楕円フィルタを用いて行われる。ハイパスフィルタ120及びローパスフィルタ122がそれぞれに適用され、結果として得られた信号の大半を破棄して2つの副帯域を形成する。高副帯域には4～8kHzスペクトルを鏡映したものが含まれる。10個の線形予測符号化（LPC）係数が124において低副帯域から計算され、2つのLPC係数が126において高帯域から計算され、同様に各帯域の利得値も計算される。図7及び図8は、代表的な無声信号のサンプリング速度16kHzでの、2つの副帯域の短期スペクトル及び2つの副帯域LPCスペクトルをそれぞれに示し、図9は結合したLPCスペクトルを示す。有声フレームの有声・無声判定128及びピッチ値130もまた低副帯域から計算される。（有声・無声判定には任意で同時に高副帯域情報も利用することができる）。10個4050

の低帯域LPCパラメータは132において線スペクトル対(LSP)に変換され、その後全てのパラメータが予測量子化器134を用いて符号化され、低ビットレートデータストリームが作られる。

図10に示すデコーダ118は136においてパラメータを復号化し、有声スピーチの間は隣接するフレームのパラメータ間を各ピッチ周期の始まりで補間する。10個の低副帯域LSPは138においてLPC係数へと変換され、その後140で2つの高副帯域係数と結合されて18個のLPC係数のセットが作られる。これは以下に説明する自己相関領域結合(Autocorrelation Domain Combination)技術又はパワースペクトル領域結合(Power Spectral Domain Combination)技術を用いて実行される。LPCパラメータは全極フィルタ142を制御するが、このフィルタは励起信号発生器144からのホワイトノイズ又はピッチ周期の周期性を持つインパルス状の波形のいずれかにより励起され、図5に示すモデルをエミュレーションする。有声励起信号の詳細は後に説明する。

ボコーダの第二の実施例の特定の具体例を次に説明する。多岐にわたる様のより詳細な考察については、本願に参考資料として組み込まれるL.Rabiner及びR.W.Schaferによる「Digital Processing of Speech Signals」、Prentice Hall、1978を参照されたい。

LPC分析

標準自己相関法が使用されて低及び高副帯域両方のLPC係数及び利得を得る。これは安定した全極フィルタを確実に供する単純な手法であるが、しかしながらフォルマント帯域幅を過剰に見積もってしまう傾向がある。この問題は、A.V.McCree及びT.P.Barnwell IIIによる「A Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Encoding」、IEE Trans.Speech and Audio Processing、Vol.3、pp242 - 250、1995年7月に記述されるように、適応フォルマント強調によってデコーダ内で解決可能である。ここでは励起シーケンスを帯域幅拡張したLPC合成(全極)フィルタでろ波することによりフォルマントの回りのスペクトルが強調される。この結果生じるスペクトルの傾きを低減するために、より弱い全零フィルタもまた適用される。フィルタ全体は伝達関数：

$$H(z) = A(z/0.5) / A(z/0.8)$$

を有しており、ここで $A(z)$ は全極フィルタの伝達関数である。

再合成LPCモデル

2つの副帯域LPCモデルのパワースペクトル間の不連続性、及び位相応答の不連続性に起因する潜在的問題を回避するために、単一の高次再合成LPCモデルが副帯域モデルから発生される。このモデル(これには18次が適当であると判明)からは、標準LPCボコーダと同様にスピーチを合成することができる。本願では2つの手法を説明するが、第二の手法は計算的により単純な方法である。

以下において、「L」及び「H」の下付き文字を使用して、仮定されたローパスフィルタによりろ波された広帯域信号の特徴をそれぞれ表わし(4kHzのカットオフを持ち、通過帯域内で単位応答、外で零となるフィルタを想定)、そして「l」及び「h」の下付き文字を使用して、それぞれ低及び高副帯域信号の特徴を表わす。

パワースペクトル領域結合

ろ波された広帯域信号 $P_L(\omega)$ 及び $P_H(\omega)$ のパワースペクトル密度は以下のように計算することができる。

$$P_L(\omega/2) = \begin{cases} g_l^2 / \left| 1 + \sum_{n=1}^{p_l} a_l(n) e^{-j\omega n} \right|^2 & \omega \leq \pi \text{ のとき} \\ 0 & \pi < \omega \leq 2\pi \text{ のとき, (1)} \end{cases}$$

及び

$$P_H(\pi - \omega/2) = \begin{cases} g_h^2 / \left| 1 + \sum_{n=1}^{p_h} a_h(n) e^{-j\omega n} \right|^2 & \omega \leq \pi \text{ のとき} \\ 0 & \pi \geq \omega \leq 2\pi \text{ のとき, (2)} \end{cases}$$

ここで $a_l(n)$ 、 $a_h(n)$ 及び g_l 、 g_h はそれぞれスピーチのフレームからのLPCパラメータ及び利得値であり、 p_l 、 p_h はLPCモデル次数である。- /2の項が生じるのは、高副帯域スペクトルが鏡映されたためである。

10

20

30

40

50

広帯域信号のパワースペクトル密度 $P_w(\cdot)$ は以下により得られる。

$$P_w(\cdot) = P_L(\cdot) + P_H(\cdot) \quad (3)$$

広帯域信号の自己相関は $P_w(\cdot)$ の逆離散時間フーリエ変換により得られ、これからその広帯域信号のフレームに対応する(18次)LPCモデルが計算できる。ある実用的な例においては、逆離散フーリエ変換(DFT)を利用して反転変換が実施される。しかしながら、この場合には、適正な周波数解像度を得るために多数のスペクトル値(代表的には512個)が必要となり、過大な量の計算が必要となるという問題が生じる。

自己相関領域結合

この手法には、ローパス及びハイパス処理された広帯域信号のパワースペクトル密度を計算するかわりに、自己相関 $r_L(\cdot)$ 及び $r_H(\cdot)$ が生成される。ローパスフィルタでろ波された広帯域信号は因数2でアップサンプリングした低副帯域に等しい。時間領域においては、このアップサンプリングは交互の零の挿入(補間)、及びその後のローパスフィルタによるろ波で構成される。したがって自己相関領域においては、アップサンプリングは、補間、その後のローパスフィルタインパルス応答の自己相関が含まれる。

2つの副帯域信号の自己相関は、副帯域LPCモデルから効率的に計算することができる(例えば、R.A.Roberts及びC.T.Mullisによる「Digital Signal Processing」(第11章、p527、Addison-Wesley、1987)を参照)。 $r_L(m)$ が低副帯域の自己相関を表わす場合、補間自己相関 $r'_L(m)$ は以下により与えられる;

$$r'_L(m) = \begin{cases} r_L(m/2) & m = 0, \pm 2, \pm 4 \dots \text{のとき} \\ 0 & \text{それ以外} \end{cases}$$

(4)

低域フィルタでろ波された信号 $r_L(m)$ は以下から求められる。

$$r_L(m) = r'_{L1}(m) \times (h(m) \times h(-m)) \quad (5)$$

ここで $h(m)$ はローパスフィルタインパルス応答である。ハイパスフィルタでろ波された信号 $r_H(m)$ の自己相関も、ハイパスフィルタが適用されることを除いて同様に得られる。

広帯域信号 $r_w(m)$ の自己相関は以下の通りに表わすことができる;

$$r_w(m) = r_L(m) + r_H(m) \quad (6)$$

そしてこれにより広帯域LPCモデルが計算される。図5には、結果として得られた上記で考慮した無声スピーチのフレームのLPCスペクトルを示す。

パワースペクトル領域における結合と比較して、この手法の方が計算的に簡単であるという利点がある。30次のFIRフィルタがアップサンプリングを実行するに十分であることがわかった。この場合、低次フィルタが意味する低い周波数解像度でも適当である。なぜならそれは単に2つの副帯域間の交差点におけるスペクトルのリークを生じることにしかならない。これらの手法は共に広帯域スピーチに高次分析モデルを用いて得られたものと知覚的に酷似したスピーチを提供するものである。

図7、図8及び図9に示す無声スピーチのフレームをプロットしたものを参照すると、信号エネルギーの大部分がスペクトルのこの領域内に含まれることから、高帯域スペクトル情報を含んだことによる効果が明確にわかる。

ピッチ / 有声 - 無声分析

ピッチは標準ピッチトラッカーを用いて決定される。有声であると判定されたフレームの各々に、ピッチ周期に最低値を持つと予想されるピッチ関数が時間間隔の範囲について計算される。3つの異なる関数が、自己相関、平均振幅差異関数(AMDF)及び負ケプストラムに基づいて与えられる。これらは全て良好に機能する。計算的に最も効率的な利用すべき関数はコーダのプロセッサのアーキテクチャにより異なる。1つ以上の有声フレームのシーケンス毎に、ピッチ関数の最小値がピッチ候補として選択される。費用関数を最小化するピッチ候補のシーケンスは、予測ピッチの輪郭として選択される。費用関数は、ピッチ関数及び経路に沿ったピッチ変化の重み付きの和である。最良の経路はダイナミックプログラミングを利用して計算的に効率的な方法で得ることができる。

10

20

30

40

50

有声 - 無声選別器の目的は、スピーチの各フレームがインパルス励起モデル、もしくは雑音励起モデルのどちらの結果として生じたものかを判定することである。有声 - 無声判定を下すために広範な方法を利用することができる。本実施例で採用した方法は、線形判別関数を；低帯域エネルギー、低帯域（任意で高帯域）の第一の自己相関係数、ピッチ分析から得たコスト価格；に適用するという方法である。有声 - 無声判定を高レベルのバックグラウンド雑音中で満足に実行するために、雑音トラッカー（例えばA.Varga及びK.Pontingによる「Control Experiments on Noise Compensation in Hidden Markov Model Based Continuous Word Recognition (pp167 - 170、Eurospeech89) に記載のもの）を使用して雑音の確率を計算し、これを線形判別関数に含むことができる。

パラメータ符号化、有声 - 無声判定

10

有声 - 無声判定は単に1フレームにつき1ビットで符号化される。連続する有声 - 無声判定間の相関を考慮することにより、これを減らすことは可能であるが、低減出来るビットレートはわずかである。

ピッチ

無声フレームについては、ピッチ情報は符号化されない。有声フレームについては、ピッチはまず対数領域に変換され、知覚的に許容し得る解像度にするために定数（例えば20）によりスケーリングされる。現在と以前の有声フレームの変換ピッチの差異は最も近い整数に丸められ、その後符号化される。

利得

対数ピッチを符号化する方法が対数利得に対しても適用され、適正なスケーリング因子は低及び高帯域に対してそれぞれ1及び0.7である。

20

LPC係数

LPC係数は符号化データの大部分を生成する。LPC係数は、まず量子化に耐え得る表現（例えば安定性が保証されており、基本フォルマント周波数及び帯域幅の歪みが低いもの）に変換される。F.I takuraによる「Line Spectrum Representation of Linear Predictor Coefficients of Speech Signals」（J.Acoust.Soc.Ameri.、Vol.57、S35（A）、1975）に記述されるように、高副帯域LPC係数は反射係数として符号化され、低副帯域LPC係数は線形スペクトル対（LSP）へと変換される。高副帯域係数は対数ピッチや対数利得と全く同じ方法で符号化される（例えば連続する値の間の差異を符号化する方法 - 適正なスケーリング因子は5.0）。低帯域係数の符号化は以下に説明する。

30

Rice符号化

本実施例においては、パラメータは固定ステップサイズで量子化され、その後無損失符号化法を利用して符号化される。符号化の方法は、Rice符号化法（R.F.Rice及びJ.R.Plauntによる「Adaptive Variable-Length Coding for Efficient Compression of Spacecraft Television Data」（IEEE Transactions on Communication Technology、Vol.19、No.6、pp889 - 897、1971）に記載）であり、これは差異のラプラス密度を用いている。この符号化法では、差異の大きさと共に増加するビットの数が指定される。この方法は、フレーム当たりに生成されるビット数を固定する必要のないアプリケーションに適しているが、LPC10e方式に類似の固定ビットレート方式を利用することも可能である。

有声励起

40

有声励起は、雑音及び周期成分が一緒になったものから構成される混合励起信号である。周期成分は、周期重み付けフィルタを通過した、パルス分散フィルタ（McCreeらにより記述）のインパルス応答である。雑音成分は雑音重み付けフィルタを通過したランダムな雑音である。

周期重み付けフィルタは、ブレークポイント（kHz）及び振幅で表される20次の有限インパルス応答（FIR）フィルタである。

ブレークポイント	0	0.4	0.6	1.3	2.3	3.4	4.0	8.0
振幅	1	1.0	0.975	0.93	0.8	0.6	0.5	0.5

雑音重み付けフィルタは、逆の応答を備える20次のFIRフィルタであり、したがって両者

50

併せて周波数帯域全体にわたる一様な応答が生成されるのである。

LPCパラメータ符号化

本実施例においては、線形スペクトル対周波数（LSF）の符号化に予測が利用され、この予測は適応性のものである。ベクトル量子化を用いることもできるが、計算量と記憶容量の双方を節約するためにスカラー符号化法が用いられる。図11に符号体系の全体像を示す。LPCパラメータエンコーダ146において、入力 $l_i(t)$ が予測器150からの予測値の負の値 $\hat{l}_i(t)$

と共に加算器148へと供給されて、予測誤差が与えられ、これが量子化器152により量子化される。量子化された予測誤差は、Rice符号化法により154において符号化されて出力を得、また予測器150の出力と共に加算器156にも供給されて予測器150への入力が得られる。

10

LPCパラメータデコーダ158において、誤差信号がRice符号化法により160で符号化され、予測器164の出力と共に加算器162へと供給される。現在のLSF成分の予測値に対応する和が加算器162から出力され、そして予測器164の入力にも供給される。

LSF予測

予測段は、現在のLSF成分をデコーダが現在利用できるデータから予測する。予測誤差のばらつきは、元の値よりも小さいと考えられ、したがって与えられた平均誤差でこれをより低いピットレートで符号化することができる。

時間 t におけるLSF要素 i を $l_i(t)$ で表わし、デコーダにより回復されたLSF要素を $\bar{l}_i(t)$ で表わす。これらのLSFが、与えられた時間枠内の増加インデックス順で、時間に連続的に符号化された場合、 $l_i(t)$ を予測するために以下の値が利用される。

20

$$\{\bar{l}_j(t) | 1 \leq j < i\}$$

及び

$$\{\bar{l}_j(\tau) | \tau < t \text{ 及び } 1 \leq j \leq 10\}$$

従って一般線形LSF予測値は；

$$\hat{l}_i(t) = c_i + \sum_{\tau=t-t_0}^{t-1} \sum_{j=1}^{10} a_{ij}(\tau) \bar{l}_j(\tau) + \sum_{j=1}^{i-1} a_{ij}(0) \bar{l}_j(t), \quad (7)$$

30

となり、ここで $a_{ij}(\cdot)$ は

$$\bar{l}_j(t-\tau)$$

からの

$$\hat{l}_i(t)$$

の予測に関係した重み付けである。

一般的に、高次予測器は適用においても予測においても計算的に効率的ではないため、 $a_{ij}(\cdot)$ の値はわずかなセットしか利用するべきではない。非量子化LSFベクトルで実験を実施した（例えば様々な予測器の構成の性能を調べるために

$$\bar{l}_j(t)$$

40

からではなく、 $l_j(\cdot)$ から予測を行なった）。結果は以下の通りである。

システム	MAC	要素	誤差／dB
A	0	—	-23.47
B	1	$a_{ii}(1)$	-26.17
C	2	$a_{ii}(1), a_{ii-1}(0)$	-27.31
D	3	$a_{ii}(1), a_{ii-1}(0), a_{ii-1}(1)$	-27.74
E	2	$a_{ii}(1), a_{ii}(2)$	-26.23
F	19	$a_{ij}(1) 1 \leq j \leq 10,$ $a_{ij}(0) 1 \leq j \leq i-1$	-27.97

表1

10

装置D（図12に図示）が効率 - 誤差間のかねあいにおいて最良のものであった。

予測器が適応的に修正される体系が用いられた。適応的更新は以下に基づいて行われる；

$$C_{xx}^{(k+1)} = (1 - \rho) C_{xx}^{(k)} + \rho x_i x_i^T$$

(8)

$$C_{xy}^{(k+1)} = (1 - \rho) C_{xy}^{(k)} + \rho y_i x_i,$$

ここで ρ は適応率を決定する ($\rho = 0.005$ で4.5秒の時定数が得られ、この値が適していることが判明)。 C_{xx} 及び C_{xy} の項は以下のようなトレーニングデータから初期化される。

$$C_{xx} = \frac{1}{N} \sum_i x_i x_i^T$$

及び

$$C_{xy} = \frac{1}{N} \sum_i y_i x_i$$

ここで y_i は予測されるべき値 ($l_i(t)$) 及び x_i は予測器入力 (l 、 $l_i(t-1)$ 等を含む) のベクトルである。方程式(8)で画定される更新は、各フレームと周期的新最小平均自乗誤差 (MMSE) 予測器係数 ρ が $C_{xx}\rho = C_{xy}$ を解くことにより算出されてから適用される。

適応型予測器は、例えば話し手の違い、チャンネルもしくはバックグラウンド雑音の相違が原因でトレーニング条件と稼動条件との間に大きな違いがある場合にのみ必要となる。

量子化及び符号化

30

予測器の出力

$$\hat{l}_i(t)$$

が与えられ、予測誤差が

$$e_i(t) = l_i(t) - \hat{l}_i(t)$$

で計算される。これはスケーリングにより一様に量子化され、誤差

$$\bar{e}_i(t)$$

を得、この誤差はその後他の全てのパラメータと同様に無損失符号化法で符号化される。適したスケーリング因数は160.0である。無声に分類されたフレームについては、より粗い量子化法を用いることができる。

結果

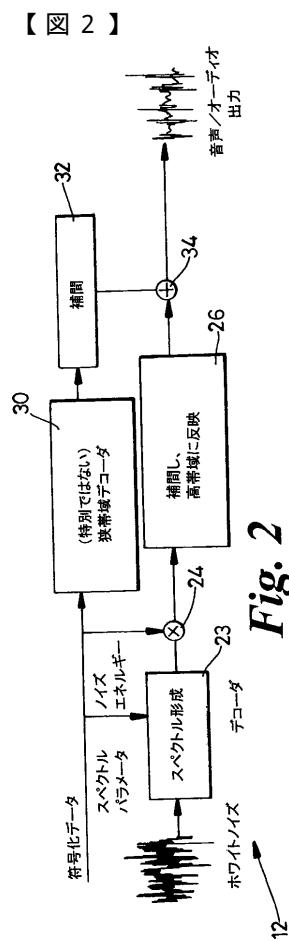
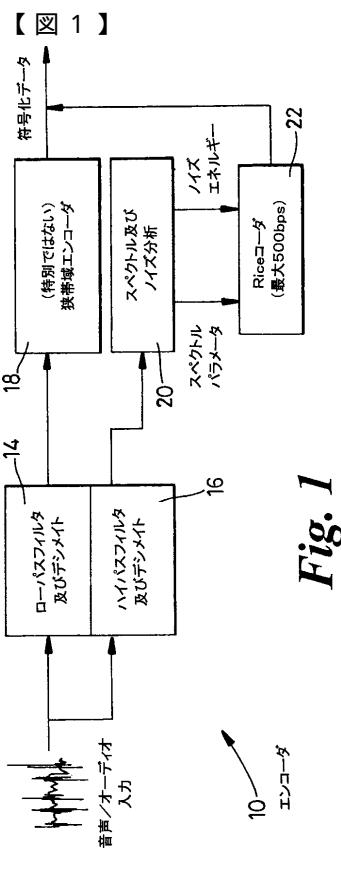
40

自己相関領域結合を利用した広帯域LPCボコーダの明瞭度を4800bpsのCELPコーダ (Federal Standard 1016) (狹帯域スピーチに利用される) の明瞭度と比較するために診断的押韻試験 (DRT) (W.D.Voiersによる「Diagnostic Evaluation of Speech Intelligibility」 (Speech Intelligibility and Speaker Recognition, M.E.Hawley, ed., pp374 - 387, Dowden, Hutchinson&Ross, Inc., 1977) に記載) を行なった。LPCボコーダについては、量子化レベル及びフレーム周期が、平均ビットレートが約2400bpsとなるように設定された。表2の結果からわかるように、広帯域LPCボコーダのDRTスコアはCELPコーダのスコアを上回っている。

コーダ	DRTスコア
CELP	83.8
広帯域LPC	86.8

表2

上述の第二の実施例にはLPCボコーダに対する最近の強化策が2つ施されている。具体的には、パルス分散フィルタ及び適応型スペクトル強化法であるが、しかし本発明の実施例に、最近発表された数多くの強化策の中から他のいずれの特長を取り込んでも良いことは言うまでもない。



【図3】

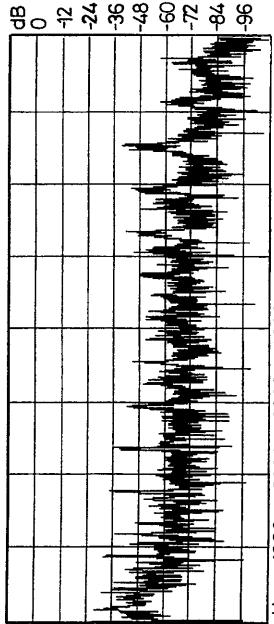


Fig. 3



【図4】

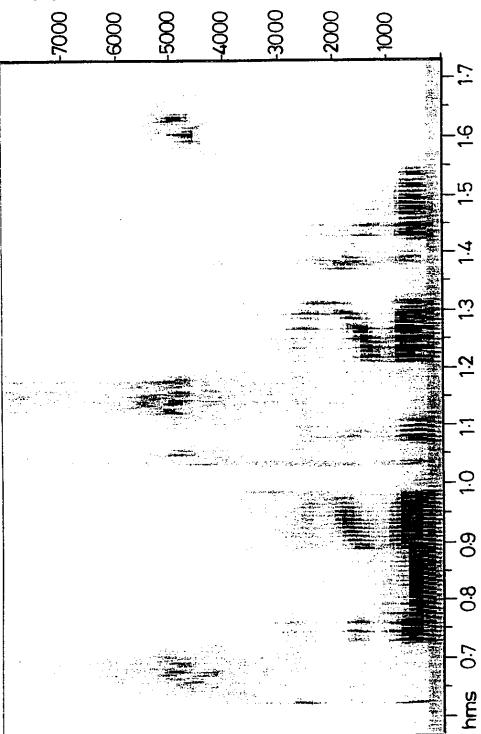


Fig. 4

【図5】

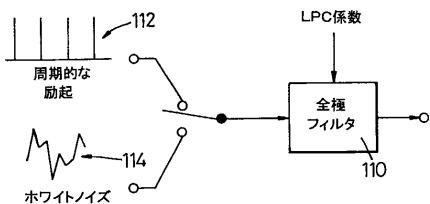


Fig. 5

【図6】

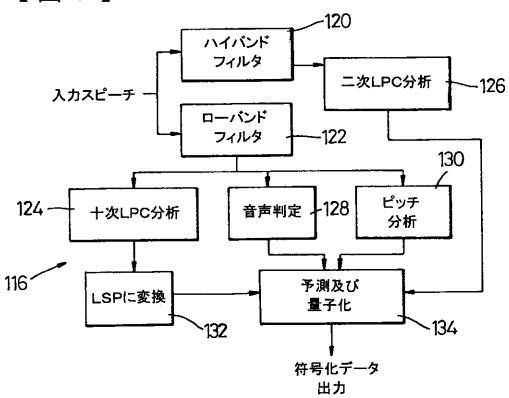


Fig. 6

【図7】

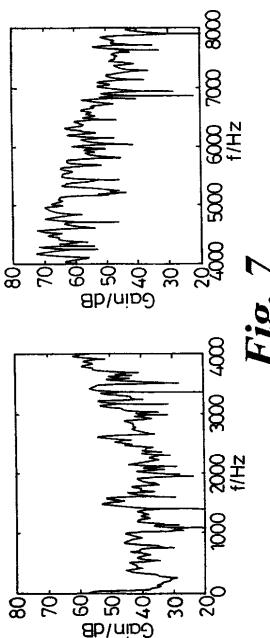


Fig. 7

【図 8】

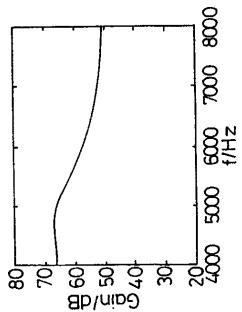
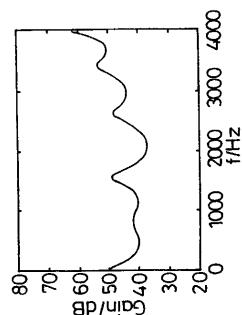


Fig. 8



【図 9】

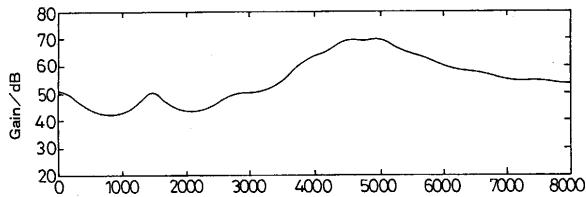


Fig. 9

【図 10】

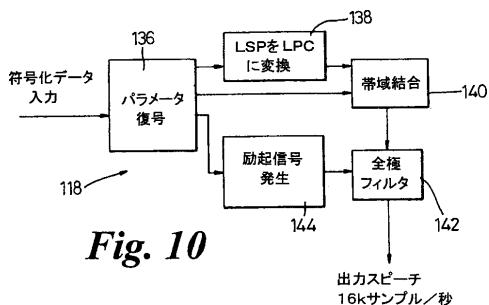


Fig. 10

【図 11】

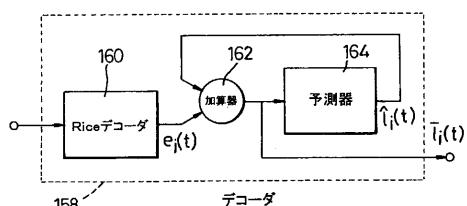
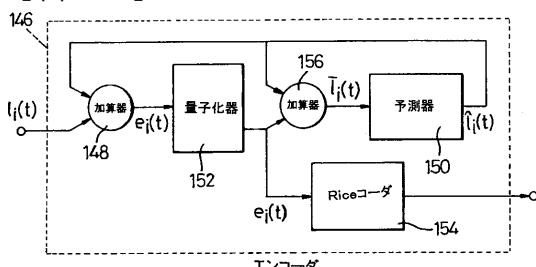


Fig. 11

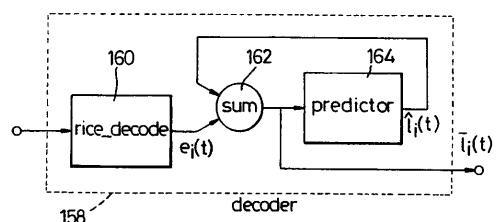
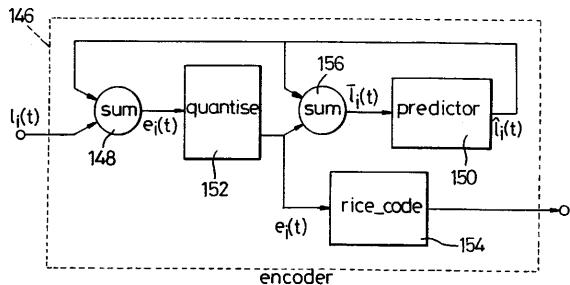


Fig. 11

【図 12】

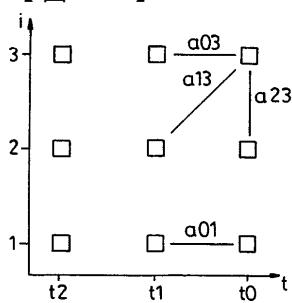


Fig. 12

フロントページの続き

(72)発明者 タッカー, ロジャー, セシル, フェリー
イギリス国モンマウスシャー・エヌピー5・7 ディーエイ, シエブストウ, タットシル, グロウ
セスター・ロード, ベラコー・ハウス(番地表示なし)
(72)発明者 セイムール, カール, ウィリアム
イギリス国ケンブリッジ・シービー5・8ディーエヌ, パルソネイジ・ストリート・26
(72)発明者 ロビンソン, アンソニー, ジョン
イギリス国ケンブリッジ・シービー4・3イーエックス, ハーベイ・グッドウィン・アベニュー・
39

審査官 山下 剛史

(56)参考文献 特開昭62-261238(JP, A)
特開平5-265492(JP, A)
特開昭54-88011(JP, A)
特表昭60-501918(JP, A)
特開平8-50500(JP, A)

(58)調査した分野(Int.Cl., DB名)

G10L 19/00 - 19/14, 21/04

H04B 14/04

H03M 7/30