

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4166673号
(P4166673)

(45) 発行日 平成20年10月15日(2008.10.15)

(24) 登録日 平成20年8月8日(2008.8.8)

| | | | |
|---------------|-----------|---------------|-------|
| (51) Int.Cl. | | F I | |
| G 1 O L 19/08 | (2006.01) | G 1 O L 19/08 | D |
| G 1 O L 19/02 | (2006.01) | G 1 O L 19/02 | 1 4 4 |
| G 1 O L 11/06 | (2006.01) | G 1 O L 11/06 | |

請求項の数 58 外国語出願 (全 28 頁)

| | | | |
|--------------|-------------------------------|-----------|--|
| (21) 出願番号 | 特願2003-383483 (P2003-383483) | (73) 特許権者 | 503417604 |
| (22) 出願日 | 平成15年11月13日(2003.11.13) | | デジタル・ボイス・システムズ・インコーポレーテッド |
| (65) 公開番号 | 特開2004-287397 (P2004-287397A) | | アメリカ合衆国マサチューセッツ州01886, ウェストフォード, リトルトン・ロード 234 |
| (43) 公開日 | 平成16年10月14日(2004.10.14) | (74) 代理人 | 100089705 |
| 審査請求日 | 平成15年11月13日(2003.11.13) | | 弁理士 社本 一夫 |
| (31) 優先権主張番号 | 10/292460 | (74) 代理人 | 100076691 |
| (32) 優先日 | 平成14年11月13日(2002.11.13) | | 弁理士 増井 忠武 |
| (33) 優先権主張国 | 米国 (US) | (74) 代理人 | 100075270 |
| | | | 弁理士 小林 泰 |
| | | (74) 代理人 | 100080137 |
| | | | 弁理士 千葉 昭男 |

最終頁に続く

(54) 【発明の名称】 相互使用可能なボコーダ

(57) 【特許請求の範囲】

【請求項 1】

デジタル音声サンプル・シーケンスをビット・ストリームにエンコードする方法であって、

前記デジタル音声サンプルを1つ以上のフレームに分割するステップと、
多数のフレームに対してモデル・パラメータを計算するステップであって、該モデル・パラメータがピッチ情報を搬送する第1パラメータを少なくとも含む、ステップと、
フレームの発声状態を判定するステップと、

判定した前記フレームの発声状態が1組の保存してある発声状態の1つに等しい場合、判定した前記フレームの発声状態を示すように前記ピッチ情報を搬送する第1パラメータを変更するステップと、

前記モデル・パラメータを量子化して量子化ビットを発生し、これらを用いて前記ビット・ストリームを生成するステップと、
を備えた方法。

【請求項 2】

請求項1記載の方法において、前記モデル・パラメータは、更に、スペクトル強度情報を決定する1つ以上のスペクトル・パラメータを含むことを特徴とする方法。

【請求項 3】

請求項1記載の方法において、

前記フレームの発声状態を多数の周波数帯域について判定し、前記モデル・パラメータ

は、更に、前記多数の周波数帯域において前記判定した発声状態を示す1つ以上の発声パラメータを含むことを特徴とする方法。

【請求項4】

請求項3記載の方法において、前記発声パラメータは、各周波数帯域における前記発声状態を、有声、無声またはパルス状のいずれかとして示すことを特徴とする方法。

【請求項5】

請求項4記載の方法において、前記1組の保存してある発声状態は、有声として示す周波数帯域がない発声状態に対応することを特徴とする方法。

【請求項6】

請求項3記載の方法において、判定した前記フレームの発声状態が1組の保存してある発声状態の1つに等しい場合、全ての周波数帯域を無声として示すように、前記発声パラメータを設定することを特徴とする方法。

10

【請求項7】

請求項4記載の方法において、判定した前記フレームの発声状態が1組の保存してある発声状態の1つに等しい場合、全ての周波数帯域を無声として示すように、前記発声パラメータを設定することを特徴とする方法。

【請求項8】

請求項5記載の方法において、判定した前記フレームの発声状態が1組の保存してある発声状態の1つに等しい場合、全ての周波数帯域を無声として示すように、前記発声パラメータを設定することを特徴とする方法。

20

【請求項9】

請求項6記載の方法において、前記ビット・ストリームを生成するステップは、誤り訂正コーディングを前記量子化ビットに適用するステップを含むことを特徴とする方法。

【請求項10】

請求項9記載の方法において、前記生成したビット・ストリームは、APCO Project 25に用いられる標準的なボコーダと相互使用可能であることを特徴とする方法。

【請求項11】

請求項3記載の方法において、前記フレームの発声状態を判定するステップは、前記フレームが発声活動ではなく背景ノイズに対応する場合、全ての周波数帯域において前記発声状態を無声に設定するステップを含むことを特徴とする方法。

30

【請求項12】

請求項4記載の方法において、前記フレームの発声状態を判定するステップは、前記フレームが発声活動ではなく背景ノイズに対応する場合、全ての周波数帯域において前記発声状態を無声に設定するステップを含むことを特徴とする方法。

【請求項13】

請求項5記載の方法において、前記フレームの発声状態を判定するステップは、前記フレームが発声活動ではなく背景ノイズに対応する場合、全ての周波数帯域において前記発声状態を無声に設定するステップを含むことを特徴とする方法。

【請求項14】

請求項2記載の方法であって、更に、
デジタル音声サンプルのフレームを分析してトーン信号を検出するステップと、
トーン信号を検出した場合、該検出したトーン信号を表すように前記フレームに対して前記1組のモデル・パラメータを選択するステップと、
を含むことを特徴とする方法。

40

【請求項15】

請求項14記載の方法において、前記検出したトーン信号はDTMFトーン信号を含むことを特徴とする方法。

【請求項16】

請求項14記載の方法において、前記検出したトーン信号を表すように前記1組のモデル・パラメータを選択するステップは、前記検出したトーン信号の振幅を表すように、前

50

記スペクトル・パラメータを選択するステップを含むことを特徴とする方法。

【請求項 17】

請求項 14 記載の方法において、前記検出したトーン信号を表すように前記 1 組のモデル・パラメータを選択するステップは、前記検出したトーン信号の周波数に少なくとも部分的に基づいて、ピッチ情報を搬送する前記第 1 パラメータを選択するステップを含むことを特徴とする方法。

【請求項 18】

請求項 16 記載の方法において、前記検出したトーン信号を表すように前記 1 組のモデル・パラメータを選択するステップは、前記検出したトーン信号の周波数に少なくとも部分的に基づいて、ピッチ情報を搬送する前記第 1 パラメータを選択するステップを含むことを特徴とする方法。

10

【請求項 19】

請求項 6 記載の方法において、前記フレームのスペクトル強度情報を決定する前記スペクトル・パラメータは、前記ピッチ情報を搬送する第 1 パラメータから決定した基本周波数の高調波から計算した 1 組のスペクトル強度パラメータを含むことを特徴とする方法。

【請求項 20】

デジタル音声サンプル・シーケンスをビット・ストリームにエンコードする方法であって、

前記デジタル音声サンプルを 1 つ以上のフレームに分割するステップと、

フレームの前記デジタル音声サンプルがトーン信号に対応するか否か判定するステップと、

20

多数のフレームについてモデル・パラメータを計算するステップであって、該モデル・パラメータが、前記ピッチを表す第 1 パラメータと、該ピッチの高調波倍数におけるスペクトル強度を表すスペクトル・パラメータとを少なくとも含む、ステップと、

フレームの前記デジタル音声サンプルがトーン信号に対応すると判定した場合、該検出したトーン信号を近似するように、前記ピッチ・パラメータおよび前記スペクトル・パラメータを選択するステップと、

前記モデル・パラメータを量子化して量子化ビットを発生し、これらを用いて前記ビット・ストリームを生成するステップと、

を備えた方法。

30

【請求項 21】

請求項 20 記載の方法において、前記 1 組のモデル・パラメータは、更に、多数の周波数帯域において発声状態を示す 1 つ以上の発声パラメータを含むことを特徴とする方法。

【請求項 22】

請求項 21 記載の方法において、前記ピッチを表す前記第 1 パラメータは、基本周波数であることを特徴とする方法。

【請求項 23】

請求項 21 記載の方法において、前記周波数帯域の各々において、前記発声状態を有声、無声またはパルス状として示すことを特徴とする方法。

【請求項 24】

請求項 22 記載の方法において、前記ビット・ストリームを生成するステップは、誤り訂正コーディングを前記量子化ビットに適用するステップを含むことを特徴とする方法。

40

【請求項 25】

請求項 21 記載の方法において、前記生成したビット・ストリームは、APCO Project 25 に用いる標準的なボコーダと相互使用可能であることを特徴とする方法。

【請求項 26】

請求項 24 記載の方法において、前記生成したビット・ストリームは、APCO Project 25 に用いる標準的なボコーダと相互使用可能であることを特徴とする方法。

【請求項 27】

請求項 21 記載の方法において、前記フレームの発声状態を判定するステップは、前記

50

フレームが発声活動ではなく背景ノイズに対応する場合、全ての周波数帯域において前記発声状態を無声に設定するステップを含むことを特徴とする方法。

【請求項 28】

ビット・シーケンスからデジタル音声サンプルをデコードする方法であって、前記ビット・シーケンスを個々のフレームに分割するステップであって、各フレームが多数のビットを含む、ステップと、

1 フレーム分のビットから量子化値を形成するステップであって、該形成した量子化値が、ピッチを表す第 1 量子化値と、発声状態を表す第 2 量子化値とを少なくとも含む、ステップと、

前記第 1 および第 2 量子化値が 1 組の保存してある量子化値に属するか否か判定を行うステップと、

前記量子化値から、フレームの音声モデル・パラメータを再現するステップであって、前記第 1 および第 2 量子化値が、前記 1 組の保存してある量子化値に属すると判定された場合、前記音声モデル・パラメータが、ピッチを表す前記第 1 量子化値から再現される前記フレームの発声状態を表す、ステップと、

前記再現した音声モデル・パラメータから 1 組のデジタル音声サンプルを計算するステップと、

を備えた方法。

【請求項 29】

請求項 28 記載の方法において、フレームに対する前記再現した音声モデル・パラメータは、前記フレームのピッチ・パラメータとスペクトル強度情報を表す 1 つ以上のスペクトル・パラメータも含むことを特徴とする方法。

【請求項 30】

請求項 29 記載の方法において、フレームを周波数帯域群に分割し、フレームの発声状態を表す前記再現した音声モデル・パラメータが、前記周波数帯域の各々における発声状態を示すことを特徴とする方法。

【請求項 31】

請求項 30 記載の方法において、各周波数帯域における前記発声状態を、有声、無声またはパルス状として示すことを特徴とする方法。

【請求項 32】

請求項 30 記載の方法において、前記周波数帯域のうちの 1 つ以上のものの帯域幅を前記ピッチ周波数に関係付けることを特徴とする方法。

【請求項 33】

請求項 31 記載の方法において、前記周波数帯域のうちの 1 つ以上のものの帯域幅を前記ピッチ周波数に関係付けることを特徴とする方法。

【請求項 34】

請求項 28 記載の方法において、前記第 2 量子化値が既知の値に等しい場合にのみ、前記第 1 および第 2 量子化値が前記 1 組の保存してある量子化値に属すると判定することを特徴とする方法。

【請求項 35】

請求項 34 記載の方法において、前記既知の値は、全ての周波数帯域を無声として示す値であることを特徴とする方法。

【請求項 36】

請求項 34 記載の方法において、前記第 1 量子化値が数個の許容値の 1 つに等しい場合にのみ、前記第 1 および第 2 量子化値が前記 1 組の保存してある量子化値に属すると判定することを特徴とする方法。

【請求項 37】

請求項 30 記載の方法において、前記第 1 および第 2 量子化値が前記 1 組の保存してある量子化値に属すると判定した場合、各周波数帯域における発声状態を有声として示さないことを特徴とする方法。

10

20

30

40

50

【請求項 38】

請求項 28 記載の方法において、1 フレーム分のビットから前記量子化値を形成するステップは、前記 1 フレーム分のビットに対して誤りデコード処理を行うステップを含むことを特徴とする方法。

【請求項 39】

請求項 30 記載の方法において、前記ビット・シーケンスは、APCO Project 25 のボコーダ規格と相互使用可能な音声エンコーダによって生成されることを特徴とする方法。

【請求項 40】

請求項 38 記載の方法において、前記ビット・シーケンスは、APCO Project 25 のボコーダ規格と相互使用可能な音声エンコーダによって生成されることを特徴とする方法。

10

【請求項 41】

請求項 29 記載の方法であって、更に、フレームに対する前記再現した音声モデル・パラメータがトーン信号に対応すると判定した場合、前記再現したスペクトル・パラメータを変更するステップを含むことを特徴とする方法。

【請求項 42】

請求項 41 記載の方法において、前記再現したスペクトル・パラメータを変更するステップは、特定の望ましくない周波数成分を減衰させるステップを含むことを特徴とする方法。

【請求項 43】

請求項 41 記載の方法において、前記第 1 量子化値および前記第 2 量子化値が特定の既知のトーン量子化値に等しい場合にのみ、フレームに対する前記再現したモデル・パラメータがトーン信号に対応すると判定することを特徴とする方法。

20

【請求項 44】

請求項 41 記載の方法において、フレームの前記スペクトル強度情報が少数の優勢周波数成分を示す場合にのみ、フレームに対する前記再現したモデル・パラメータがトーン信号に対応すると判定することを特徴とする方法。

【請求項 45】

請求項 43 記載の方法において、フレームの前記スペクトル強度情報が少数の優勢周波数成分を示す場合にのみ、フレームに対する前記再現したモデル・パラメータがトーン信号に対応すると判定することを特徴とする方法。

30

【請求項 46】

請求項 44 記載の方法において、前記トーン信号は、DTFM トーン信号を含み、フレームの前記スペクトル強度情報が、既知の DTFM 周波数またはその付近にある 2 つの優勢な周波数成分を示す場合にのみ、前記 DTFM トーン信号を判定することを特徴とする方法。

【請求項 47】

請求項 32 記載の方法において、前記フレームのスペクトル強度情報を表す前記スペクトル・パラメータは、前記再現したピッチ・パラメータから決定した基本周波数の高調波を表す 1 組のスペクトル強度パラメータから成ることを特徴とする方法。

【請求項 48】

ビット・シーケンスからデジタル音声サンプルをデコードする方法であって、前記ビット・シーケンスを、各々多数のビットを含む個々のフレームに分割するステップと、

40

1 フレーム分のビットから音声モデル・パラメータを再現するステップであって、1 フレーム分の再現した音声モデル・パラメータが、当該フレームについてのスペクトル強度情報を表す 1 つ以上のスペクトル・パラメータを含む、ステップと、

前記再現した音声モデル・パラメータから、前記フレームがトーン信号を表すか否か判定を行うステップと、

前記フレームがトーン信号を表す場合、前記スペクトル・パラメータを変更し、該変更したスペクトル・パラメータが、前記判定したトーン信号のスペクトル強度情報を一層良

50

く表すようにするステップと、

前記再現した音声モデル・パラメータおよび前記変更したスペクトル・パラメータからデジタル音声サンプルを発生するステップと、
を備えた方法。

【請求項 49】

請求項 48 記載の方法において、フレームに対する前記再現した音声モデル・パラメータは、ピッチを表す基本周波数パラメータも含むことを特徴とする方法。

【請求項 50】

請求項 49 記載の方法において、フレームに対する前記再現した音声モデル・パラメータは、多数の周波数帯域における発声状態を示す発声パラメータも含むことを特徴とする方法。

10

【請求項 51】

請求項 50 記載の方法において、前記周波数帯域の各々における前記発声状態は、有声、無声またはパルス状として示すことを特徴とする方法。

【請求項 52】

請求項 49 記載の方法において、前記フレームのスペクトル・パラメータは、前記基本周波数パラメータの高調波における前記スペクトル強度情報を表す 1 組のスペクトル強度から成ることを特徴とする方法。

【請求項 53】

請求項 50 記載の方法において、前記フレームのスペクトル・パラメータは、前記基本周波数パラメータの高調波における前記スペクトル強度情報を表す 1 組のスペクトル強度から成ることを特徴とする方法。

20

【請求項 54】

請求項 52 記載の方法において、前記再現したスペクトル・パラメータを変更するステップは、前記判定したトーン信号に含まれない高調波に対応する前記スペクトル強度を減衰させるステップを含むことを特徴とする方法。

【請求項 55】

請求項 52 記載の方法において、前記 1 組のスペクトル強度における数個のスペクトル強度が前記 1 組における他の全スペクトル強度に対して優勢である場合にのみ、フレームに対する前記再現した音声モデル・パラメータはトーン信号に対応すると判定することを特徴とする方法。

30

【請求項 56】

請求項 55 記載の方法において、前記トーン信号は、D T F M トーン信号を含み、前記 1 組のスペクトル強度が、標準の D T F M 周波数またはその付近にある 2 つの優勢な周波数成分を含む場合にのみ、前記 D T F M トーン信号を判定することを特徴とする方法。

【請求項 57】

請求項 50 記載の方法において、前記基本周波数パラメータおよび前記発声パラメータが前記パラメータに対する一定の既知の値にほぼ等しい場合にのみ、フレームに対する前記再現した音声モデル・パラメータがトーン信号に対応すると判定することを特徴とする方法。

40

【請求項 58】

請求項 55 記載の方法において、前記ビット・シーケンスは、APCO Project 25 のボコーダ規格と相互使用可能な音声エンコーダによって生成されることを特徴とする方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般に、音声およびその他のオーディオ信号のエンコードおよび/またはデコード処理に関する。

【背景技術】

【0002】

50

音声のエンコードおよびデコード処理には多数の用途があり、広範囲にわたって研究されてきた。一般に、音声のコード化は、音声圧縮としても知られており、音声の品質即ち了解度を実質的に低下させることなく、音声信号を表すために必要なデータ・レートを低下させようとするものである。音声圧縮技法は、音声コードによって実現することができ、音声コードのことをボイス・コードまたはボコードと呼ぶこともある。

【0003】

音声コードは、一般に、エンコーダおよびデコーダを含むと見なされる。エンコーダは、マイクロフォンが生成するアナログ信号を入力として有するアナログ/デジタル変換器の出力に生成するような、音声のデジタル表現から圧縮ビット・ストリームを生成する。デコーダは、圧縮ビット・ストリームを、デジタル/アナログ変換器およびスピーカによる再生に適した、音声のデジタル表現に変換する。多くの用途では、エンコーダおよびデコーダは、物理的に分離されており、ビット・ストリームをこれらの間で送信するには、通信チャネルを用いる。

【0004】

音声コードの主要なパラメータの1つに、コードが達成する圧縮量があり、これは、エンコーダが生成するビット・ストリームのビット・レートによって測定する。エンコーダのビット・レートは、一般に、所望の忠実度（即ち、音声の品質）、および用いられる音声コードの形式の関数である。音声コードは、形式が異なれば、異なるビット・レートで動作するように設計されている。最近では、広範囲の移動通信用途のために、10 kbps未滿で動作する低から中程度のレートの音声コードに関心が集まっている（例えば、セルラ電話、衛星電話、陸線移動無線通信、および機内電話）。これらの用途では、高品質の音声、ならびに音響ノイズおよびチャネル・ノイズ（例えば、ビット・エラー）によって生ずるアーチファクトに対するロバスト性が要求されるのが通例である。

【発明の開示】

【発明が解決しようとする課題】

【0005】

音声は、一般に、時間の経過と共に変化する信号特性を有する非定常信号と見なされる。この信号特性の変化は、一般に、異なる音を生成する人の声道の特性において作られる変化と関連付けられる。音は、通例、ある短い期間、通例は10から100msの間維持され、次いで声道が再度変化して、次の音を生成する。音同士の間の変移は、遅く連続的であったり、あるいは変移は音声「開始」(onset)の場合のように素早いこともある。この信号特性の変化のために、ビット・レートが低くなるに連れて、音声をエンコードすることが増々難しくなる。何故なら、音によっては、他の音よりも本来的にエンコードが難しいものがあり、音声コードは、音声信号の特性変移に追従(adapt)する能力を保持しつつ、妥当な忠実度で全ての音をエンコードできなければならないからである。低から中程度のビット・レートの音声コードの性能を向上する1つの方法は、ビット・レートを可変とすることである。可変ビット・レート音声コードでは、音声の各セグメントに対するビット・レートは固定されておらず、逆に、ユーザの入力、システムの負荷、端末の設計または信号特性というような種々の要因に応じて、2つ以上の選択肢の間で変化させることができる。

【0006】

低から中程度のデータ・レートにおいて音声をコード化する主な手法には何種類がある。例えば、線形予測コーディング(LPC: linear predictive coding)に基づく手法では、短期および長期予測器を用いて、以前のサンプルから新しい音声の各フレームを予測しようとする。予測エラーは、いくつかの手法の1つを用いて量子化するのが通例であり、その中から、CELPおよび/またはマルチ・パルスの2例をあげておく。LPC法の利点は、時間分解能が高いことであり、無声音(unvoiced sound)のコーディングに役立つ。即ち、この方法には、破裂音および過渡音(transient)が結局は(in time)過度に不明瞭になることはないという効果がある。しかしながら、線形予測には、コード化した信号における周期性が不十分なことから、コード化した音声は粗雑にまたはしゃがれ声に聞こ

10

20

30

40

50

える場合が多く、有声音には難点がある。この問題は、データ・レートが低くなる程、一層深刻となる。これは、データ・レートが低い程長いフレーム・サイズが必要となるのが通例であることから、長期予測器は周期性再現の有効性が低下するためである。

【 0 0 0 7 】

低から中程度のレートの音声コーディングの別の先端的手法に、モデルに基づく音声コード即ちボコードがある。ボコードは、音声を、短い時間期間における励起に対するシステムの応答としてモデル化する。ボコード・システムの例には、線形予測ボコード（例えば、MELP）、同形ボコード(homomorphic vocoder)、チャンネル・ボコード、正弦変換コード（「STC」）、高調波ボコード(harmonic vocoder)、およびマルチバンド励起（「MBE」）ボコードが含まれる。これらのボコードでは、音声は短いセグメント（通例、10から40ms）に分割され、各セグメントを1組のモデル・パラメータによって特徴化する。これらのパラメータは、通例、各音声セグメントの数個の基本的なエレメント、当該セグメントのピッチ、発声状態、およびスペクトル包絡線等を表す。これらのパラメータ毎に、多数の公知の表現の1つを用いるボコードも可能である。例えば、ピッチは、ピッチ周期、基本的周波数またはピッチ周波数（ピッチ周期の逆数）として、または長期予測遅延として表すことができる。同様に、発声状態は、1つ以上の発声計量、発声確率測定、または1組の発声判断(voicing decision)によって表すことができる。スペクトル包絡線は、全極フィルタ応答によって表されることが多いが、1組のスペクトル強度またはその他のスペクトル測定値によって表すこともできる。モデルに基づく音声コードは、少数のパラメータのみを用いて、音声セグメントを表現することができるので、ボコードのようなモデルに基づく音声コードは、通例では、中程度から低データ・レートで動作することができる。しかしながら、モデルに基づくシステムの品質は、基礎となるモデルの精度に左右される。したがって、これらの音声コードが高い音声品質を達成しなければならないとすると、忠実度が高いモデルを用いる必要がある。

【 0 0 0 8 】

MBEボコードは、MBE音声モデルに基づく高調波ボコードであり、多くの用途において優れた動作を行うことが示されている。MBEボコードは、有声音の高調波表現を、MBE音声モデルに基づく柔軟な周波数依存発声構造と組み合わせる。これによって、MBEボコードは、自然な発音の無声音(natural sounding unvoiced speed)を生成することができる。音響背景ノイズの存在に対するMBEボコードのロバスト性が高められる。これらの特性により、MBEボコードは、低から中程度のデータ・レートにおいて生成される音声の品質を高めることができ、多数の工業的移動通信用途においてMBEボコードが利用されるようになった。

【 0 0 0 9 】

MBE音声モデルは、ピッチに対応する基本周波数、1組の発声計量または判断、および声道の周波数応答に対応する1組のスペクトル強度を用いて、音声のセグメントを表す。MBEモデルは、従来のセグメント毎に1つのV/U/V判断を、1組の判断に一般化し、各判断は、特定の周波数帯域即ち領域における発声状態を表す。これによって、各フレームを、少なくとも有声および無声周波数領域に分割する。こうして、発声モデルにおいて柔軟性を高めることにより、MBEモデルは、一部の有声摩擦音のような、混合発声音に対する適応性を高め、音響背景ノイズによって潰された音声の表現精度を高めることができ、いずれの判断においてもエラーに対する感応性を低下させる。この一般化の結果、ボイス品質および了解度が向上したことが、広範な試験によって示されている。

【 0 0 1 0 】

MBEに基づくボコードには、IMBE（商標）音声コードや、AMBE（登録商標）音声コードが含まれる。IMBE（商標）音声コードは、APCO Project 25を含む多数のワイヤレス通信システムにおいて用いられている。AMBE（登録商標）音声コードは、これに含まれる励起パラメータ（基本周波数および発声判断）を推定する方法のロバスト性を高め、実際の音声において発見した変動やノイズをより良く追跡することができるように改良されたシステムである。通例では、AMBE（登録商標）音声コードは、フィ

10

20

30

40

50

ルタバンクを用いるが、多くの場合これには16のチャンネルおよび非線形性を含み、1組のチャンネル出力を生成し、これらから励起パラメータを容易に推定することができる。チャンネル出力を組み合わせて処理し、基本周波数を推定する。その後、数個（例えば、8つ）発声帯域の各々において、チャンネルを処理して、各発声帯域毎に発声判断（またはその他の発声計量）を推定する。AMBE+2（商標）ボコーダでは、三状態発声モデル（有声、無声、パルス状）を適用し、破裂音およびその他の過渡音声を一層良く表している。MBEモデル・パラメータを量子化する種々の方法が、多様なシステムにおいて適用されている。通例では、AMBE（登録商標）ボコーダおよびAMBE+2（商標）ボコーダが採用する量子化方法は、ベクトル量子化のように、一層進んでおり、ビット・レートが低い程高い品質の音声を生成する。

10

【0011】

MBEに基づく音声コーダのエンコーダは、各音声セグメント毎に1組のモデル・パラメータを推定する。MBEモデル・パラメータは、基本周波数（ピッチ周期の逆数）、発声状態を特徴化する1組のV/U/V計量または判断、およびスペクトル包絡線の特徴化する1組のスペクトル強度を含む。MBEモデル・パラメータをセグメント毎に推定した後、エンコーダは、パラメータを量子化して1フレーム分のビットを生成する。任意に、エンコーダは、誤り訂正/検出コードでこれらのビットを保護した後に、インターリーブし、その結果得られたビット・ストリームを対応するデコーダに送信する。

【0012】

MBEに基づくボコーダにおけるデコーダは、受信したビット・ストリームから、MBEモデル・パラメータ（基本周波数、発声情報、およびスペクトル強度）を音声のセグメント毎に再現する。この再現の一部として、デコーダは、デインターリーブ処理および誤り制御デコード処理を行い、ビット・エラーを訂正および/または検出する。加えて、位相再生(phase regeneration)もデコーダによって行われ、合成位相情報を計算するのが通例である。APCO Project 25 ボコーダの説明書に指定され、米国特許第5,081,681号および第5,664,051号に記載されている1つの方法では、発声判断に応じて、ランダム位相再生を、ランダム性の量と共に用いている。別の方法では、位相再生を行う際に、再現したスペクトル強度にスムージング・カーネルを適用する。これは、米国特許第5,701,390号に記載されている。

20

【0013】

デコーダは、再現したMBEモデル・パラメータを用いて、元の音声に知覚的に高度に類似した音声信号を合成する。有声、無声、そして任意にパルス状音声に対応する信号成分は、通常別個であり、各セグメント毎に合成され、次いで得られた成分を合計して、合成音声信号を形成する。このプロセスを音声のセグメント毎に繰り返し、完全な音声信号を再生し、D/A変換器およびラウドスピーカを介して出力する。無声信号成分を合成するには、ウィンドウ重複加算法(windowed overlap-add method)を用いて、白色ノイズ信号を濾過する。フィルタの時間変動スペクトル包絡線は、無声と指定された周波数領域において再生された一連のスペクトル強度から決定され、他の周波数領域は0に設定される。

30

【0014】

デコーダは、数種類の方法の内1つを用いて、有声信号成分を合成することができる。APCO Project 25 ボコーダの説明書において指定されている1つの方法では、1群の高周波発振器を用い、基本周波数の各高調波毎に1つずつ発振器を割り当て、発振器全てからの寄与を加算して、有声信号成分を形成する。別の方法では、有声信号成分を合成するには、有声インパルス応答とインパルス・シーケンスとの畳み込みを行い、隣接するセグメントからの寄与をウィンドウ重複加算によって組み合わせる。この2番目の方法の方が速く計算することができる。何故なら、これはセグメント間における成分の照合を全く必要とせず、任意のパルス信号成分にも適用できるからである。

40

【0015】

MBEに基づくボコーダの特定の一例に、APCO Project 25 移動無線通信システムの

50

標準として選択された 7200 bps の IMBE (商標) ボコーダがある。このボコーダは、APCO Project 25 ボコーダの説明書に記載されており、144 ビットを用いて各 20 ms フレームを表す。これらのビットは、56 ビットの冗長 FEC ビット (ゴレイおよびハミング・コーディングの組み合わせを適用する)、1 ビットの同期ビット、および 87 ビットの MBE パラメータ・ビットに分割される。87 ビットの MBE パラメータ・ビットは、基本周波数を量子化するための 8 ビットと、二進有声/無声判断を量子化する 3 - 12 ビットと、スペクトル強度を量子化する 67 - 76 ビットから成る。その結果生ずる 144 ビット・フレームは、エンコーダからデコーダに伝送される。デコーダは、誤り訂正を実行した後に、誤りデコード・ビットから MBE モデル・パラメータを再現する。次いで、デコーダは、再現したモデル・パラメータを用いて、有声および無声信号成分を合成し、これらを合計して、デコード音声信号を形成する。

10

【0016】

【特許文献1】米国特許第 5,081,681 号

【特許文献2】米国特許第 5,664,051 号

【特許文献3】米国特許第 5,701,390 号

【非特許文献1】APCO Project 25 ボコーダの説明書

【課題を解決するための手段】

【0017】

概括的な形態の 1 つでは、デジタル音声サンプル・シーケンスをビット・ストリームにエンコードする際、デジタル音声サンプルを 1 つ以上のフレームに分割し、多数のフレームについてモデル・パラメータを計算することを含む。モデル・パラメータは、ピッチ情報を搬送する第 1 パラメータを少なくとも含む。フレームの発声状態を判定し、判定したフレームの発声状態が、1 組の保存してある発声状態の 1 つに等しい場合、判定したフレームの発声状態を示すように、このフレームに対するピッチ情報を搬送するパラメータを変更する。次いで、モデル・パラメータを量子化して量子化ビットを発生し、これらを用いてビット・ストリームを生成する。

20

【0018】

実現例では、次の特徴を 1 つ以上含むことができる。例えば、モデル・パラメータは、更に、スペクトル強度情報を判定する 1 つ以上のスペクトル・パラメータを含むこともできる。

30

【0019】

フレームの発声状態を多数の周波数帯域について判定することもでき、モデル・パラメータは、周波数帯域において判定した発声状態を示す 1 つ以上のパラメータを含むこともできる。発声パラメータは、各周波数帯域における発声状態を、有声、無声またはパルス状として示すことができる。1 組の保存してある発声状態は、有声として示される周波数帯域がない発声状態に対応することができる。判定したフレームの発声状態が、1 組の保存してある発声状態の 1 つに等しい場合、全ての周波数帯域を無声として示すように、発声パラメータを設定することができる。また、発声状態は、フレームが発声活動ではなく、背景ノイズに対応する場合、全ての周波数帯域を無声として示すように設定することもできる。

40

【0020】

ビット・ストリームの生成では、量子化ビットに誤り訂正コーディングを適用することを含ませてもよい。生成したビット・ストリームは、APCO Project 25 に用いられる標準的なボコーダと相互使用可能とするとよい。

【0021】

1 フレーム分のデジタル音声サンプルを分析してトーン信号を検出することができ、トーン信号が検出された場合、フレームの 1 組のモデル・パラメータは、検出されたトーン信号を表すように選択することができる。この検出したトーン信号は、DTMF トーン信号を含むことがある。検出したトーン信号を表すように 1 組のモデル・パラメータを選択する場合、検出したトーン信号の振幅を表すようにスペクトル・パラメータを選択する

50

こと、および/または検出したトーン信号の周波数に少なくとも部分的に基づいて、ピッチ情報を搬送する第1パラメータを選択することを含むようにするとよい。

【0022】

フレームのスペクトル強度情報を決定するスペクトル・パラメータは、ピッチ情報を搬送する第1パラメータから決定した基本周波数の高調波から計算した1組のスペクトル強度パラメータを含む。

【0023】

別の概括的な態様では、デジタル音声サンプル・シーケンスをビット・ストリームにエンコードする際、デジタル音声サンプルを1つ以上のフレームに分割し、フレームのデジタル音声サンプルがトーン信号に対応するか否か判定することを含む。多数のフレームについて、モデル・パラメータを計算し、モデル・パラメータは、ピッチを表す第1パラメータと、ピッチの高調波倍数におけるスペクトル強度を表すスペクトル・パラメータとを少なくとも含む。フレームのデジタル音声サンプルがトーン信号に対応すると判定した場合、検出したトーン信号を近似するように、スペクトル・パラメータを選択する。モデル・パラメータを量子化して量子化ビットを発生し、これらを用いてビット・ストリームを生成する。

【0024】

実現例では、以下の特徴の1つ以上、および先に記した特徴の1つ以上を含むことができる。例えば、1組のモデル・パラメータは、更に、多数の周波数帯域において発声状態を示す1つ以上の発声パラメータを含むこともできる。ピッチを表す第1パラメータは基本周波数とすることができる。

【0025】

別の概括的な態様では、ビット・シーケンスからデジタル音声サンプルをデコードする際、ビット・シーケンスを、各々多数のビットを含む、個々のフレームに分割することを含む。1フレーム分のビットから量子化値を形成する。形成した量子化値は、ピッチを表す第1量子化値と、発声状態を表す第2量子化値とを少なくとも含む。第1および第2量子化値が1組の保存してある量子化値に属するか否かについて判定を行う。その後、量子化値から、フレームの音声モデル・パラメータを再現する。第1および第2量子化値が1組の保存してある量子化値に属すると判定された場合、音声モデル・パラメータは、ピッチを表す第1量子化値から再現されたフレームの音声状態を表す。最後に、再現した音声モデル・パラメータから、デジタル音声サンプルを計算する。

【0026】

実現例では、以下の特徴の1つ以上、および先に記した特徴の1つ以上を含むことができる。例えば、フレームに対して再現した音声モデル・パラメータは、ピッチ・パラメータと、フレームについてのスペクトル強度情報を表す1つ以上のスペクトル・パラメータとを含むことができる。フレームを周波数帯域群に分割することができ、フレームの発声状態を表す、再現した音声モデル・パラメータは、周波数帯域の各々における発声状態を示すことができる。各周波数帯域における発声状態は、有声、無声またはパルス状のいずれかとして示すことができる。1つ以上の周波数帯域の帯域幅を、ピッチ周波数と関係付けることもできる。

【0027】

第2量子化値が既知の値と等しい場合にのみ、第1および第2量子化値が1組の保存してある量子化値に属すると判定することができる。既知の値は、周波数帯域の全てを無声として示す値とすることができる。第1量子化値が数個の許容値の1つに等しい場合にのみ、第1および第2量子化値が1組の保存してある量子化値に属すると判定することができる。第1および第2量子化値が1組の保存してある量子化値に属すると判定された場合、各周波数帯域における発声状態は、有声として示さないようにすることができる。

【0028】

1フレーム分のビットから量子化値を形成する際、この1フレーム分のビットに対して誤りデコード処理を実行することを含んでもよい。ビット・シーケンスは、APCO Project

10

20

30

40

50

25のボコーダ規格と相互使用可能な音声エンコーダによって生成することができる。

【0029】

フレームに対して再現した音声モデル・パラメータがトーン信号に対応すると判定された場合、再現したスペクトル・パラメータを変更することができる。再現したスペクトル・パラメータを変更する場合、特定の望ましくない周波数成分を減衰させることを含んでもよい。第1量子化値および第2量子化値が特定の既知のトーン量子化値に等しい場合、またはフレームのスペクトル強度情報が少数の有声周波数成分を示す場合にのみ、フレームに対して再現したモデル・パラメータがトーン信号に対応すると判定することができる。トーン信号は、DTMFトーン信号を含むことができ、フレームのスペクトル強度情報が、既知のDTMF周波数またはその付近にある2つの優勢な周波数成分を示す場合にのみ、このDTMFトーン信号を判定する。

10

【0030】

フレームのスペクトル強度情報を表すスペクトル・パラメータは、再現したピッチ・パラメータから決定した基本周波数の高調波を表す、1組のスペクトル強度パラメータで構成することができる。

【0031】

別の概括的な形態では、ビット・シーケンスからデジタル音声サンプルをデコードする際、ビット・シーケンスを、各々多数のビットを含む、個々のフレームに分割することを含む。1フレーム分のビットから音声モデル・パラメータを再現する。フレームに対して再現した音声モデル・パラメータは、当該フレームのスペクトル強度情報を表す1つ以上のスペクトル・パラメータを含む。再現した音声モデル・パラメータを用いて、フレームがトーン信号を表すか否かを判定し、フレームがトーン信号を表す場合、スペクトル・パラメータを変更し、変更したスペクトル・パラメータが、判定したトーン信号のスペクトル強度情報を一層良く表すようにする。再現した音声モデル・パラメータおよび変更したスペクトル・パラメータから、デジタル音声サンプルを発生する。

20

【0032】

実現例では、以下の特徴の1つ以上、および先に記した特徴の1つ以上を含むことができる。例えば、フレームに対して再現した音声モデル・パラメータは、ピッチを表す基本周波数パラメータと、多数の周波数帯域における発声状態を示す発声パラメータも含む。周波数帯域の各々における発声状態は、有声、無声またはパルス状のいずれかとして示すことができる。

30

【0033】

フレームのスペクトル・パラメータは、基本周波数パラメータの高調波におけるスペクトル強度情報を表す1組のスペクトル強度を含むことができる。再現したスペクトル・パラメータを変更するには、判定したトーン信号に含まれない高調波に対応するスペクトル強度を減衰させることを含めばよい。

【0034】

1組のスペクトル強度における数個のスペクトル強度が、1組における他の全スペクトル強度に対して優勢である場合、または基本周波数パラメータおよび発声パラメータが、当該パラメータに対する一定の既知の値にほぼ等しい場合にのみ、フレームに対して再現した音声モデル・パラメータがトーン信号に対応すると判定することができる。トーン信号は、DTMFトーン信号を含むことができ、1組のスペクトル強度が、標準的なDTMF周波数またはその付近にある2つの優勢な周波数成分を含む場合にのみ、このDTMFトーン信号を判定する。

40

【0035】

ビット・シーケンスは、APCO Project 25のボコーダ規格と相互使用可能な音声エンコーダによって生成することができる。

別の概括的な形態では、改良マルチバンド励起(MBE)ボコーダは、標準的なAPCO Project 25ボコーダと相互使用可能であるが、ボイス品質の向上、トーン信号に対する忠実度の向上、および背景ノイズに対するロバスト性の向上をもたらす。改良MBEエンコ

50

ーダ・ユニットは、M B Eパラメータ推定、M B Eパラメータ量子化、およびF E Cエンコード処理というようなエレメントを含むことができる。M B Eパラメータ推定エレメントは、発声活動検出、ノイズ抑制、トーン検出、および三状態発声モデルというような、先進の機構を含む。M B Eパラメータ量子化は、基本周波数データ・フィールドに、発声情報を挿入することができる。改良M B Eデコーダは、F E Cデコード処理、M B Eパラメータ再現、およびM B E音声合成というようなエレメントを含むことができる。M B Eパラメータ再現は、基本周波数データ・フィールドから発声情報を取り出せることを特徴とする。M B E音声合成は、有声、無声、およびパルス状信号成分の組み合わせとして、音声を合成することができる。

【 0 0 3 6 】

その他の特徴は、図面および特許請求の範囲を含む、以下の説明から明白であろう。

【発明を実施するための最良の形態】

【 0 0 3 7 】

図1は、マイクロフォン105からのアナログ音声または何らかのその他の信号をサンプリングする音声コーダ即ちボコーダ100を示す。A/D変換器110がマイクロフォンからのアナログ音声をデジタル化し、デジタル音声信号を生成する。デジタル音声信号は、改良M B E音声エンコーダ・ユニット115によって処理され、送信または格納に適したデジタル・ビット・ストリーム120を生成する。

【 0 0 3 8 】

通例では、音声エンコーダは、デジタル音声信号を短いフレーム単位で処理し、フレームを1つ以上のサブフレームに更に分割することもできる。デジタル音声サンプルの各フレームは、エンコーダのビット・ストリーム出力において、対応するビットのフレームを生成する。尚、フレームには1つのサブフレームしかない場合、フレームおよびサブフレームは通例では同等であり、同じ信号の区分を指すことを注記しておく。一実現例では、フレーム・サイズの期間は20msであり、8kHzのサンプリング・レートにおいて160個のサンプルから成る。用途によっては、各サンプルを2つの10msサブフレームに分割することによって、性能が向上する場合もある。

【 0 0 3 9 】

また、図1は、受信ビット・ストリーム125も示す。ビット・ストリーム125は、改良M B E音声デコーダ・ユニット130に入力され、M B E音声デコーダ・ユニット130は、各ビット・フレームを処理して、対応する合成音声サンプルのフレームを生成する。D/A変換ユニット135が、次に、デジタル音声サンプルをアナログ信号に変換し、これをスピーカ・ユニット140に受け渡し、人の聴取に適した音響信号に変換することができる。エンコーダ115およびデコーダ130は、異なる場所にあってもよく、送信ビット・ストリーム120および受信ビット・ストリームが同一であってもよい。

【 0 0 4 0 】

ボコーダ100は、改良型のM B Eに基づくボコーダであり、APCO Project 25通信システムにおいて用いられる標準的なボコーダと相互使用可能である。一実現例では、改良7200bpsボコーダが標準的なAPCO Project 25ボコーダ・ビット・ストリームを用いて相互使用可能となっている。この改良7200bpsボコーダでは、ボイス品質の向上、耐音響背景ノイズ性向上、最上位のトーン処理を含む、性能向上が得られる。ビット・ストリームの相互使用可能性(interoperability)を保存して、改良エンコーダが生成する7200bpsビット・ストリームを、標準的なAPCO Project 25ボイス・デコーダがデコードし、高品質の音声を生成できるようにする。同様に、改良エンコーダは、標準的なエンコーダが発生する7200bpsビット・ストリームを入力し、これから高品質の音声をデコードする。ビット・ストリームの相互使用可能性を備えることにより、改良エンコーダを組み込んだ無線機またはその他のデバイスを、既存のAPCO Project 25システムに継ぎ目無く組み込むことができ、システムのインフラストラクチャによる変換やトランスコード処理(transcoding)も不要である。標準的なボコーダとの下位互換性を備えることによって、改良ボコーダを用いると、相互使用可能性の問題を引き起こすことなく、

10

20

30

40

50

既存のシステムの性能を高度化する(upgrade)ことが可能となる。

【 0 0 4 1 】

図2を参照すると、改良MBEエンコーダ115は、音声エンコーダ・ユニット200を用いて実現することができる。音声エンコーダ・ユニット200は、まずパラメータ推定ユニット205によって入力デジタル音声信号を処理して、フレーム毎に、一般化したMBEモデル・パラメータを推定する。次いで、1つのフレームについて推定したモデル・パラメータを、MBEパラメータ量子化ユニット210によって量子化し、パラメータ・ビットを生成し、FECエンコード・パリティ付加ユニット215に供給し、量子化ビットと冗長順方向誤り訂正(FEC)データと組み合わせて、送信ビット・ストリームを形成する。冗長FECデータを付加することによって、デコーダは、伝送チャネルにおける劣化によって生ずるビット・エラーを訂正および/または検出することが可能となる。

10

【 0 0 4 2 】

また、図2に示すように、改良MBEデコーダ130は、MBE音声デコーダ・ユニット220を用いて実現することができる。MBE音声デコーダ・ユニット220は、まずFECデコーダ・ユニット225を用いて受信ビット・ストリームにおけるフレームを処理して、ビット・エラーを訂正および/または検出する。フレームのパラメータ・ビットは、次に、MBEパラメータ再現ユニット230によって処理され、フレーム毎に一般化されたMBEモデル・パラメータを再現する。次に、MBE音声合成ユニット235が、得られたモデル・パラメータを用いて、合成デジタル音声信号を生成する。これがデコーダの出力となる。

20

【 0 0 4 3 】

APCO Project 25 ボコーダ規格では、144ビットを用いて、各20msフレームを表す。これらのビットは、56ビットの冗長FECビット(ゴレイおよびハミング・コーディングの組み合わせを適用する)、1ビットの同期ビット、および87ビットのMBEパラメータ・ビットに分割される。標準的なAPCO Project 25ボコーダのビット・ストリームと相互使用可能とするには、改良ボコーダは、各フレーム内において、同じフレーム・サイズおよび同じ全体的なビット割り当てを用いる。しかしながら、改良ボコーダは、標準的なボコーダに対して、これらのビットにある種の修正を用いて、搬送する情報を増大し、ボコーダの性能を向上させつつ、標準的なボコーダとの下位互換性を維持している。

30

【 0 0 4 4 】

図3は、改良MBEボイス・エンコーダが実施する改良MBEパラメータ推定手順300を示す。手順300を実施するには、ボイス・エンコーダは、トーン判断(ステップ305)を実行して、フレーム毎に、入力信号が数個の既知のトーン形式(単一トーン、DTFMトーン、ノックス・トーン(Knox tone)、または呼進展トーン(call progress tone))の1つに対応するか否か判定を行う。

【 0 0 4 5 】

また、ボイス・エンコーダは、発声活動検出(VAD: voice activity detection)も実行して(ステップ310)、フレーム毎に、入力信号が人の声かまたは背景ノイズか判定を行う。VADの出力は、フレームがボイスかまたはボイスでないかを示す、フレーム毎の単一ビットの情報である。

40

【 0 0 4 6 】

次に、エンコーダは、MBE発声判断およびピッチ情報を搬送する基本周波数を推定し(ステップ315)、スペクトル強度を推定する(ステップ320)。発声判断は、VAD判断がフレームを背景ノイズ(ボイスでない)と判定した場合には、全て無声に設定すればよい。

【 0 0 4 7 】

スペクトル強度を推定した後、ノイズ抑制を適用し(ステップ325)、スペクトル強度から、知覚されるレベルの背景ノイズを除去する。実現例によっては、VAD判断を用いて、背景ノイズの推定値を改善する。

50

【 0 0 4 8 】

最後に、スペクトル強度が、無声またはパルス状と示された発声帯域にある場合、これらを補償する(ステップ330)。標準的なボコーダは、異なるスペクトル強度推定方法を用いるので、これを考慮して補償を行う。

【 0 0 4 9 】

改良MBEボイス・エンコーダは、トーン検出を実行し、入力信号においてある種別のトーン信号を特定する。図4は、エンコーダが実施するトーン検出手順400を示す。最初に、ハミング・ウィンドウまたはカイザー・ウィンドウ(Kaiser window)を用いて、入力信号をウィンドウ処理する(window)(ステップ405)。次いで、FFTを計算し(ステップ410)、FFT出力から全スペクトル・エネルギーを計算する(ステップ415)。通例、FFT出力を評価して、150 - 3800 Hzの範囲にある単一のトーン、DTMFトーン、ノックス・トーンおよびある種の呼進展トーンを含む数個のトーン信号の1つと対応するか否かを判定を行う。

10

【 0 0 5 0 】

次に、最良のトーン候補を判定する。この際、一般に、エネルギーが最大の1つ以上のFFTピンを発見する(ステップ420)。次いで、トーンが1つの場合には、FFTピンと選択したトーン周波数候補とを加算することによって、二重トーンの場合には、複数の周波数と加算することによって、トーン・エネルギーを計算する(ステップ425)。

【 0 0 5 1 】

次いで、SNR(トーン・エネルギーと全トーンとの間の比率)レベル、周波数、または捻れ(twist)のような、所要のトーン・パラメータをチェックすることによって、トーン候補の妥当性を判断する(ステップ430)。例えば、電気通信に用いられる標準化された二重周波数トーンであるDTMFトーンの場合、2つの周波数成分の各々の周波数は、有効なDTMFトーンに対する公称値の約3%以内でなければならない。このような検査によって有効なトーンが確認されたなら、表1に示すような1組のMBEモデル・パラメータを用いて、推定したトーン・パラメータを高調波群にマッピングする(ステップ435)。例えば、697 Hz、1336 HzのDTMFトーンを、基本周波数が70 Hz ($f_0 = 0.00875$)、2つの非ゼロ高調波(10、19)を有し、それ以外の全ての高調波が0に設定された高調波群にマッピングすることができる。次いで、非ゼロ高調波を含む発声帯域が有声となり、それ

20

30

【 0 0 5 2 】

【表 1 - 1】

表 1 : M B E トーン・パラメータ

| トーン形式 | 周波数成分 (Hz) | M B E モデル・パラメータ | | |
|-------------|---------------|-----------------|---------|--------|
| | | トーン・ インデックス | 基本 (Hz) | 非ゼロ高調波 |
| 単一トーン | 156.25 | 5 | 156.15 | 1 |
| 単一トーン | 187.5 | 6 | 187.5 | 1 |
| ... | ... | ... | ... | ... |
| 単一トーン | 375.0 | 12 | 375.0 | 1 |
| 単一トーン | 406.3 | 13 | 203.13 | 2 |
| ... | ... | ... | ... | ... |
| 単一トーン | 781.25 | 25 | 390.63 | 2 |
| 単一トーン | 812.50 | 26 | 270.83 | 3 |
| ... | ... | ... | ... | ... |
| 単一トーン | 1187.5 | 38 | 395.83 | 3 |
| 単一トーン | 1218.75 | 39 | 304.69 | 4 |
| ... | ... | ... | ... | ... |
| 単一トーン | 1593.75 | 51 | 398.44 | 4 |
| 単一トーン | 1625.0 | 52 | 325.0 | 5 |
| ... | ... | ... | ... | ... |
| 単一トーン | 2000.0 | 64 | 400.0 | 5 |
| 単一トーン | 2031.25 | 65 | 338.54 | 6 |
| ... | ... | ... | ... | ... |
| 単一トーン | 2375.0 | 76 | 395.83 | 6 |
| 単一トーン | 2406.25 | 77 | 343.75 | 7 |
| ... | ... | ... | ... | ... |
| 単一トーン | 2781.25 | 89 | 397.32 | 7 |
| 単一トーン | 2812.5 | 90 | 351.56 | 8 |
| ... | ... | ... | ... | ... |
| 単一トーン | 3187.5 | 102 | 398.44 | 8 |
| 単一トーン | 3218.75 | 103 | 357.64 | 9 |
| ... | ... | ... | ... | ... |
| 単一トーン | 3593.75 | 115 | 399.31 | 9 |
| 単一トーン | 3625.0 | 116 | 362.5 | 10 |
| ... | ... | ... | ... | ... |
| 単一トーン | 3812.5 | 122 | 381.25 | 10 |
| D T M F トーン | 941, 1336 | 128 | 78.50 | 12, 17 |
| D T M F トーン | 697, 1209 | 129 | 173.48 | 4, 7 |
| D T M F トーン | 697, 1336 | 130 | 70.0 | 10, 19 |
| D T M F トーン | 697, 1477 | 131 | 87.0 | 8, 17 |
| D T M F トーン | 770, 1209 | 132 | 109.95 | 7, 11 |
| D T M F トーン | 770, 1336 | 133 | 191.68 | 4, 7 |
| D T M F トーン | 770, 1447 | 134 | 70.17 | 11, 21 |
| D T M F トーン | 852, 1209 | 135 | 71.06 | 12, 17 |
| D T M F トーン | 852, 1336 | 136 | 121.58 | 7, 11 |
| D T M F トーン | 852, 1447 | 137 | 212.0 | 4, 7 |
| D T M F トーン | 697, 1633 | 138 | 116.41 | 6, 14 |
| D T M F トーン | 770, 1633 | 139 | 96.15 | 8, 17 |

10

20

30

40

【 0 0 5 3 】

【表 1 - 2】

| | | | | |
|-------------|-----------|-----|--------|--------|
| D T M F トーン | 852, 1633 | 140 | 71.0 | 12, 23 |
| D T M F トーン | 941, 1633 | 141 | 234.26 | 4, 7 |
| D T M F トーン | 941, 1209 | 142 | 134.38 | 7, 9 |
| D T M F トーン | 941, 1477 | 143 | 134.35 | 7, 11 |
| K n o x トーン | 820, 1162 | 144 | 68.33 | 12, 17 |
| K n o x トーン | 606, 1052 | 145 | 150.89 | 4, 7 |
| K n o x トーン | 606, 1162 | 146 | 67.82 | 9, 17 |
| K n o x トーン | 606, 1297 | 147 | 86.50 | 7, 15 |
| K n o x トーン | 672, 1052 | 148 | 95.79 | 7, 11 |
| K n o x トーン | 672, 1162 | 149 | 166.92 | 4, 7 |
| K n o x トーン | 672, 1297 | 150 | 67.70 | 10, 19 |
| K n o x トーン | 743, 1052 | 151 | 74.74 | 10, 14 |
| K n o x トーン | 743, 1162 | 152 | 105.90 | 7, 11 |
| K n o x トーン | 743, 1297 | 153 | 92.78 | 8, 14 |
| K n o x トーン | 606, 1430 | 154 | 101.55 | 6, 14 |
| K n o x トーン | 672, 1430 | 155 | 84.02 | 8, 17 |
| K n o x トーン | 743, 1430 | 156 | 67.83 | 11, 21 |
| K n o x トーン | 820, 1430 | 157 | 102.30 | 8, 14 |
| K n o x トーン | 820, 1052 | 158 | 117.0 | 7, 9 |
| K n o x トーン | 820, 1297 | 159 | 117.49 | 7, 11 |
| 呼進展 | 350, 440 | 160 | 87.78 | 4, 5 |
| 呼進展 | 440, 480 | 161 | 70.83 | 6, 7 |
| 呼進展 | 480, 630 | 162 | 122.0 | 4, 5 |
| 呼進展 | 350, 490 | 163 | 70.0 | 5, 7 |

10

20

【 0 0 5 4 】

通例、改良 M B E ボコーダは、発声活動検出 (V A D) を含み、各フレームをボイスまたは背景ノイズのいずれかに識別する。 V A D には種々の方法を適用することができる。しかしながら、図 5 に示す特定の V A D 方法 5 0 0 では、1 つ以上の周波数帯域 (1 6 帯域が通例) において 1 フレーム全体について入力信号のエネルギーを測定する (ステップ 5 0 5) ことを含む。

【 0 0 5 5 】

次に、各周波数帯域において、当該帯域における最小エネルギーを追跡することによって、背景ノイズ最低値 (floor) の推定値を推定し (ステップ 5 1 0) する。次いで、実際に測定したエネルギーとこの推定ノイズ最低値との間の誤差を、各周波数帯域に対して計算し (ステップ 5 1 5) 、次いで全周波数帯域においてエラーを蓄積する (ステップ 5 2 0) 。次いで、蓄積したエラーを閾値と比較し (ステップ 5 2 5) 、蓄積エラーが閾値を超過した場合、当該フレームに対してボイスが検出されたことになる。蓄積エラーが閾値を超過しない場合、背景ノイズ (ボイス以外) が検出されたことになる。

30

【 0 0 5 6 】

図 3 に示す改良 M B E エンコーダは、入力音声信号のフレーム毎に、1 組の M B E モデル・パラメータを推定する。通例では、発声判断および基本周波数 (ステップ 3 1 5) を最初に推定する。改良 M B E エンコーダは、先進の三状態発声モデルを用いて、所要の周波通領域を、有声、無声、またはパルス状のいずれかに定義する。この三状態発声モデルは、ボコーダの破裂音およびその他の過渡的音を表す機能を高め、知覚されるボイスの品質を大幅に高める。エンコーダは、1 組の発声判断を推定するが、各発声判断はフレーム内の個々の周波数領域の発声状態を示す。また、エンコーダは有声信号成分のピッチを示す基本周波数も推定する。

40

【 0 0 5 7 】

改良 M B E エンコーダが用いる特徴の 1 つとして、フレームが全体的に無声またはパルス状 (即ち、有声成分を有さない) であるときに、基本周波数はある程度任意であることがあげられる。したがって、フレームに有声の部分がない場合、基本周波数は、他の情報

50

を搬送するために用いることができる。これを図6に示し、以下で説明する。

【0058】

図6は、基本周波数および発声判断を推定する方法600を示す。入力音声は、最初に、非線形動作を含むフィルタバンクを用いて分割される(ステップ605)。例えば、一実現例では、入力音声を8つのチャンネルに分割し、各チャンネルは500Hzの範囲を有する。フィルタバンクの出力を処理して、このフレームの基本周波数を推定し(ステップ610)、各フィルタバンク・チャンネル毎に発声計量(voicing metric)を計算する(ステップ615)。これらのステップの詳細は、米国特許第5,715,365号および第5,826,222号において論じられており、その内容は、ここで引用したことによって、本願にも含まれることとする。加えて、三状態発声モデルでは、エンコーダがフィルタバンク・チャンネル毎にパルス計量を推定することが必要となる(ステップ620)。これは、2001年11月20日に出願した同時係属中の米国特許出願第09/988,809号において論じられている。その内容は、ここで引用したことによって、本願にも含まれることとする。次いで、チャンネル発声計量およびパルス計量を処理して、1組の発声判断を計算する(ステップ625)これらは、各チャンネルの発声状態を有声、無声、またはパルス状のいずれかとして表す。一般に、チャンネルが有声として示されるのは、発声計量が第1有声閾値よりも小さい場合であり、パルス状として示されるのは、発声計量が、第1有声スレシホールドよりも小さい第2有声閾値よりも小さいときであり、それ以外の場合は無声として示される。

10

【0059】

一旦チャンネル発声判断を決定したなら、いずれかのチャンネルが有声でないか判定するためにチェックを行う(ステップ630)。有声のチャンネルがない場合、当該フレームの発声状態は、全てのチャンネルが無声またはパルス状である1組の保存してある発声状態に属する。この場合、推定基本周波数を、表2からの値と置き換える(ステップ635)。この値は、ステップ625において決定したチャンネル発声判断に基づいて選択する。加えて、有声のチャンネルがない場合、標準的なAPCO Project 25ボコーダにおいて用いられる発声帯域の全てを無声に設定する(即ち、 $b_1 = 0$)。

20

【0060】

【表 2 - 1】

表 2 : 無声 M B E の基本周波数

| 基本周波数 (H z) | | チャンネル発声判断 | |
|--|-------|--|--|
| APCO Project 25 ボ コードの説明書か らの量子化器の値 (b_0) | (H z) | サブチャンネル 1 8つのフィルタバン ク・チャンネル 低周波 - 高周波 | サブチャンネル 0 8つのフィルタバン ク・チャンネル 低周波 - 高周波 |
| 25 | 248.0 | UUUUUUU | UUUUUUU |
| 128 | 95.52 | UUUUUUUP | UUUUUUU |
| 129 | 94.96 | UUUUUUUPU | UUUUUUUU |
| 130 | 94.40 | UUUUUUUPP | UUUUUUUU |
| 131 | 93.84 | UUUUUUUPUU | UUUUUUUU |
| 132 | 93.29 | UUUUUUUPPP | UUUUUUUU |
| 133 | 92.75 | UUUUUUUPPPP | UUUUUUUU |
| 134 | 92.22 | UUUUUUUPUUU | UUUUUUUU |
| 135 | 91.69 | UUUUUUUPUUU | UUUUUUUU |
| 136 | 91.17 | UUUUUUUPUUU | UUUUUUUU |
| 137 | 90.65 | UUUUUUUPUUU | UUUUUUUU |
| 138 | 90.14 | UUUUUUUPUUU | UUUUUUUU |
| 139 | 89.64 | UUUUUUUPUUU | UUUUUUUU |
| 140 | 89.14 | UUUUUUUPUUU | UUUUUUUU |
| 141 | 88.64 | UUUUUUUPUUU | UUUUUUUU |
| 142 | 88.15 | UUUUUUUPUUU | UUUUUUUU |
| 143 | 87.67 | UUUUUUUPUUU | UUUUUUUU |
| 144 | 87.19 | UUUUUUUPUUU | UUUUUUUU |
| 145 | 86.72 | UUUUUUUPUUU | UUUUUUUU |
| 146 | 86.25 | UUUUUUUPUUU | UUUUUUUU |
| 147 | 85.79 | UUUUUUUPUUU | UUUUUUUU |
| 148 | 85.33 | UUUUUUUPUUU | UUUUUUUU |
| 149 | 84.88 | UUUUUUUPUUU | UUUUUUUU |
| 150 | 84.43 | UUUUUUUPUUU | UUUUUUUU |
| 151 | 83.98 | UUUUUUUPUUU | UUUUUUUU |
| 152 | 83.55 | UUUUUUUPUUU | UUUUUUUU |
| 153 | 83.11 | UUUUUUUPUUU | UUUUUUUU |
| 154 | 82.69 | UUUUUUUPUUU | UUUUUUUU |
| 155 | 82.26 | UUUUUUUPUUU | UUUUUUUU |
| 156 | 81.84 | UUUUUUUPUUU | UUUUUUUU |
| 157 | 81.42 | UUUUUUUPUUU | UUUUUUUU |
| 158 | 81.01 | UUUUUUUPUUU | UUUUUUUU |
| 159 | 80.60 | UUUUUUUPUUU | UUUUUUUU |
| 160 | 80.20 | UUUUUUUPUUU | UUUUUUUU |
| 161 | 79.80 | UUUUUUUPUUU | UUUUUUUU |
| 162 | 79.40 | UUUUUUUPUUU | UUUUUUUU |
| 163 | 79.01 | UUUUUUUPUUU | UUUUUUUU |
| 164 | 78.62 | UUUUUUUPUUU | UUUUUUUU |
| 165 | 78.23 | UUUUUUUPUUU | UUUUUUUU |
| 166 | 77.86 | UUUUUUUPUUU | UUUUUUUU |
| 167 | 77.48 | UUUUUUUPUUU | UUUUUUUU |
| 168 | 77.11 | UUUUUUUPUUU | UUUUUUUU |
| 169 | 76.74 | UUUUUUUPUUU | UUUUUUUU |

10

20

30

40

【 0 0 6 1 】

【表 2 - 2】

| | | | |
|-----|-------|----------|----------|
| 170 | 76.37 | UUUUUUUU | PPUUUUUU |
| 171 | 76.01 | UUUUUUUU | UPPPPPPP |
| 172 | 75.65 | UUUUUUUU | PPUUUUUU |
| 173 | 75.29 | UUUUUUUU | UUUPPPPP |
| 174 | 74.94 | UUUUUUUU | UUUPPPPP |
| 175 | 74.59 | UUUUUUUU | PPPPPPUU |
| 176 | 74.25 | PPPPPPPP | PPPPPPPP |
| 177 | 73.90 | PPPPPPPP | UUUPPPPP |
| 178 | 73.56 | PPUUUUUU | PPUUUUUU |
| 179 | 73.23 | UUUPPPPP | UUUPPPPP |
| 180 | 72.89 | UUUPPPPP | UUUPPPPP |
| 181 | 72.56 | PPPPPPPP | PPUUUUUU |
| 182 | 72.23 | PPUUUUUU | PPUUUUUU |
| 183 | 71.91 | UUUUUUUU | UUUUUUUU |

10

【 0 0 6 2 】

1 フレーム中の発声帯域の数を計算する（ステップ 6 4 0）。発声帯域の数は、基本周波数に応じて 3 から 1 2 の間で変動する。所与の基本周波数に対する具体的な発声帯域の数は、APCO Project 25 ボコーダの説明書に記載されており、高調波の数を 3 で除算することによって近似的に得られ、最大 1 2 である。

【 0 0 6 3 】

1 つ以上のチャンネルが有声である場合、発声状態は、保存してある組には属せず、推定した基本周波数を保持して、標準的に量子化し、チャンネル発声判断を、標準的な APCO Project 25 発声帯域にマッピングする（ステップ 6 4 5）。

20

【 0 0 6 4 】

通例では、固定のフィルタバンク・チャンネル周波数から基本周波数に応じた発声帯域周波数までの周波数スケールリング(frequency scaling)を用いて、ステップ 6 4 5 に示すマッピングを行う。

【 0 0 6 5 】

図 6 は、チャンネル発声判断のいずれもが有声ではないときにはいつでも、発声判断に関する情報を搬送するための基本周波数の使用を示す（即ち、発声状態が、チャンネル発声判断が無声またはパルス状のいずれかに属する 1 組の保存してある発声状態に属する場合）。尚、標準的なエンコーダでは、発声帯域が全て無声である場合、基本周波数は任意に選択され、発声判断に関する情報を全く搬送しない。逆に、図 6 のシステムは、有声帯域がない場合にはいつでも、チャンネル発声判断に関する情報を搬送する新たな基本周波数を、好ましくは表 2 から選択する。

30

【 0 0 6 6 】

選択方法の 1 つは、ステップ 6 2 5 からのチャンネル発声判断を、表 2 における各基本周波数候補に対応するチャンネル発声判断と比較することである。チャンネル発声判断が最も近いテーブルのエントリを、新たな基本周波数として選択し、基本周波数量子化値 b_0 としてエンコードする。ステップ 6 2 5 の最終部分は、発声量子化値 b_1 を 0 に設定することであり、通常、標準的なデコーダでは全ての発声帯域を無声として示す。尚、改良エンコーダは、発声状態が無声および/またはパルス状帯域の組み合わせであるときはいつでも発声量子化値 b_1 を 0 に設定し、改良エンコーダが生成したビット・ストリームを標準的なデコーダが受信するときに、確実に全ての発声帯域を無声としてデコードするようにしていることを注記しておく。次に、どの帯域がパルス状であり、どの帯域が無声であるかについての具体的な情報を、前述のように、基本周波数量子化値 b_0 にエンコードする。APCO Project 25 ボコーダの説明書を参照すれば、量子化値 b_0 および b_1 のエンコードおよびデコード処理を含む、標準的なボコーダ処理に関する情報を更に得ることができる。

40

【 0 0 6 7 】

50

尚、チャンネル発声判断は、通常、フレーム毎に1回推定され、この場合、表2が基本周波数を選択する際には、推定したチャンネル発声判断を、表2の「サブフレーム1」と称する列における発声判断と比較し、最も近いテーブルのエントリを用いて、選択する基本周波数を決定する。この場合、「サブフレーム0」と称する表2の列は用いられない。しかしながら、前述の同じフィルタバンクに基づく方法を用いて、フレーム毎に2回（即ち、フレームにおける2つのサブフレームについて）チャンネル発声判断を推定することにより、性能を一層向上させることができる。この場合、フレーム当たり2組のチャンネル発声判断があり、表2から基本周波数を選択する際には、双方のサブフレームについて推定したチャンネル発声判断を、表2の両方の列に記入されている音声判断と比較する。この場合、両方のサブフレームに対して試験したときに最も近いテーブルのエントリを用いて、選択する基本周波数を決定する。

10

【0068】

再度図3を参照する。一旦励起パラメータ（基本周波数および発声情報）を推定したなら（ステップ315）、改良MBEエンコーダは、1組のスペクトル強度をフレーム毎に推定する（ステップ320）。トーン判断（ステップ305）によって、現フレームに対してトーン信号が検出されている場合、表1から指定した非ゼロの高調波を除いて、スペクトル強度を0に設定する。非ゼロの高調波には、検出したトーン信号の振幅を設定する。逆に、トーンが検出されない場合、フレームのスペクトル強度を推定するには、155点修正カイザー・ウィンドウのような短い重複ウィンドウ関数を用いて音声信号をウィンドウ処理し、次いでウィンドウ処理した信号についてFFTを計算する（通例では、 $K = 256$ ）。次に、推定した基本周波数の各高調波にエネルギーを加算し、和の二乗根が第1高調波のスペクトル強度 M_1 となる。スペクトル強度を推定する手法の1つが、米国特許第5,754,974号において論じられている。その内容は、ここで引用したことによって、本願にも含まれることとする。

20

【0069】

通例、改良MBEエンコーダは、ノイズ抑制方法（ステップ325）を含み、推定したスペクトル強度から、知覚される背景ノイズ量を低減するために用いる。1つの方法では、1組の周波数帯域において、局部ノイズ最低値(noise floor)の推定値を計算する。通例では、発声活動検出（ステップ310）からのVAD判断出力を用いて、ボイスが検出されないフレームの間に推定された局部ノイズを更新する。これによって、ノイズ最低値の推定値が、音声レベルではなく、背景ノイズ・レベルの測定値であることの確証が得られる。一旦ノイズの推定値を得たなら、このノイズ推定値のスムージングを行い、典型的なスペクトル減算技法を用いて、推定スペクトル強度から減算する。ここで、減衰の最大量は約15dBに制限されるのが通例である。ノイズ推定値が0に近い場合（即ち、背景ノイズが殆どまたは全くない場合）、ノイズ抑制を行っても、スペクトル強度には殆どまたは全く変化がない。しかしながら、かなりのノイズがある場合（例えば、窓を開けた車両の中で話すとき）、ノイズ抑制方法によって、推定スペクトル強度にはかなりの改善が得られる。

30

【0070】

APCO Project 25 ボコーダの説明書に指定されている標準的なMBEでは、スペクトル振幅は、有声および無声高調波毎に別々に推定する。逆に、改良MBEエンコーダは、米国特許第5,754,974号に記載されているように、同じ推定方法を用いて、全ての高調波を推定するのが通例である。この差を補正するために、改良MBEエンコーダは、無声およびパルス状高調波を補償し（即ち、無声またはパルス状であると明言された発声帯域内の高調波）、以下のように最終スペクトル強度 M_1 を求める。

40

【0071】

【数 1】

$$M_{l, m} = M_{l, m} / [K \cdot f_0]^{(1/2)}$$

第 1 高調波がパルス状または無声の場合

$$M_{l, n} = M_{l, n}$$

第 1 高調波が有声の場合 (1)

【0072】

ここで、 $M_{l, n}$ は、ノイズ抑制後の改善されたスペクトル強度であり、 K は FFT サイズ (通例では $K = 256$)、そして f_0 はサンプリング・レート (8000 Hz) に正規化した基本周波数である。最終的なスペクトル強度 M_l を量子化して、量子化値 b_2 、 b_1 、 \dots 、 b_{L+1} を形成する。 L は、フレームにおける高調波の数に等しい。最後に、FEC コーディングを量子化値に適用し、コーディングの結果、改良 MBE エンコーダから出力ビット・ストリームを形成する。

10

【0073】

改良 MBE エンコーダから出力したビット・ストリームは、標準的な APCO Project 25 ボコーダと相互使用可能である。標準的なデコーダは、改良 MBE エンコーダが生成したビット・ストリームをデコードし、高品質の音声を生成することができる。一般に、標準 20 的デコーダが生成する音声の品質は、標準的なビット・ストリームをデコードする場合よりは、改善したビット・ストリームをデコードしたときの方が高い。このボイス品質の向上は、発声活動検出、トーン検出、MBE パラメータ推定の改良、およびノイズ抑制というような、改良 MBE エンコーダの様々な形態によるものである。

【0074】

更に、改善したビット・ストリームを改良 MBE デコーダによってデコードすることによって、ボイス品質を向上させることができる。図 2 に示すように、改良 MBE デコーダは、通例、標準的なデコード処理 (ステップ 225) を含み、受信したビット・ストリームを量子化値に変換する。標準的な APCO Project 25 ボコーダでは、各フレームは、4 つの [23、12] ゴレイ・コードと、3 つの [15、11] ハミング・コードとを含み、 30 これらをデコードして、伝送中に発生し得るビット・エラーを訂正および/または検出する。FEC デコード処理に続いて、MBE パラメータ再現 (ステップ 230) を行い、量子化値を MBE パラメータに変換し、続いて MBE 音声合成によって合成を行う (ステップ 235)。

30

【0075】

図 7 は、特定の MBE パラメータ再現方法 700 を示す。方法 700 は、基本周波数および発声再現 (ステップ 705) を含み、続いてスペクトル強度再現 (710) を含む。次に、適用したスケールリングを全ての無声およびパルス状高調波から解除することによって、スペクトル強度を逆補償する (715)。

【0076】

次に、得られた MBE パラメータを表 1 と突き合わせてチェックし、有効なトーン・フレームに対応するか否か調べる (ステップ 720)。一般に、トーン・フレームが特定されるのは、基本周波数が表 1 におけるあるエントリにほぼ等しく、そのトーンの高調波の発声帯域が有声であり、他の発声帯域全てが無声であり、当該トーンについて表 1 に指定されている、その非ゼロ高調波のスペクトル強度が、他のスペクトル強度よりも優勢である場合である。トーン・フレームがデコーダによって識別される場合、指定された非ゼロ高調波以外の全ての高調波を減衰させる (20 dB の減衰が通例)。このプロセスによって、ボコーダに用いられるスペクトル強度量子化器が混入する望ましくない高調波サイドローブを減衰させる。サイドローブを減衰させることによって、歪み量が減少し、量子化器に全く変更を加える必要なく、合成したトーンの忠実度を高めることによって、 40 50

標準的なボコーダとの相互使用可能性を維持する。トーン・フレームが識別されない場合、サイドロープの抑制は、スペクトル強度には適用されない。

【 0 0 7 7 】

手順 7 0 0 における最終ステップとして、スペクトル強度の改善および適応スムージングを実行する（ステップ 7 2 5）。図 8 を参照すると、改良 M B E デコーダは、受信した量子化値 b_0 および b_1 から、手順 8 0 0 を用いて、基本周波数および発声情報を再現する。最初に、デコーダは b_0 から基本周波数を再現する（ステップ 8 0 5）。次いで、デコーダは、基本周波数から発声帯域の数を計算する（ステップ 8 1 0）。

【 0 0 7 8 】

次に、検査を適用して、受信した発声量子化値 b_1 の値が 0 で、全無声状態を示すか否か判定を行う（ステップ 8 1 5）。 b_1 の値が 0 の場合、第 2 の検査を適用して、受信した b_0 の値が、表 2 に収容されている b_0 の保存値の 1 つに等しいか否か判定を行う（ステップ 8 2 0）。これは、基本周波数が、発声状態に関する追加情報を含むことを示す。等しい場合、ある検査を用いて、状態変数 ValidCount が 0 以上か否かチェックする（ステップ 8 3 0）。0 以上である場合、デコーダは表 2 において、受信した量子化値 b_0 に対応するチャンネル発声判断を参照する（ステップ 8 4 0）。これに続いて、変数 ValidCount を最大 3 の値まで増分し（ステップ 8 3 5）、続いて表の参照から得たチャンネル判定を発声帯域にマッピングする（ステップ 8 4 5）。

【 0 0 7 9 】

b_0 が、保存されている値の 1 つとも等しくない場合、最小値 - 1 0 以上の値に ValidCount を減分する。（ステップ 8 2 5）。

変数 ValidCount が 0 未満の場合、変数 ValidCount を最大 3 の値まで増分する（ステップ 8 3 5）。

【 0 0 8 0 】

3 つの検査（ステップ 8 1 5、8 2 0、8 3 0）のいずれかが偽であった場合、APCO Project 25 ボコーダの説明書において標準的なボコーダについて記載されているように、受信した b_1 の値から発声帯域を再現する（ステップ 8 5 0）。

【 0 0 8 1 】

再度図 2 を参照する。一旦 M B E パラメータを再現したなら、改良 M B E デコーダは、出力音声信号を合成する（ステップ 2 3 5）。特定の音声合成方法 9 0 0 を図 9 に示す。この方法は、別個の有声、パルス状、および無声信号成分を合成し、3 つの成分を組み合わせ、出力合成音声を生成する。有声音声合成（ステップ 9 0 5）は、標準的なボコーダについて記載した方法を用いてもよい。しかしながら、他の手法では、インパルス・シーケンスおよび有声インパルス応答関数を畳み込み、次いで隣接するフレームからの結果を、ウィンドウ重複加算 (windowed overlap-add) を用いて組み合わせる。パルス状音声合成（9 1 0）は、通例、同じ方法を適用して、パルス状信号成分を計算する。この方法の詳細は、同時係属中の米国特許出願第 1 0 / 0 4 6 , 6 6 6 号に記載されている。これは、2 0 0 2 年 1 月 1 6 日に出版され、その内容は、ここで引用したことにより、本願にも含まれることとする。

【 0 0 8 2 】

無声信号成分の合成（9 1 5）では、白色ノイズ信号に重み付けを行い、標準的なボコーダについて説明したように、ウィンドウ重複加算を用いて、フレーム群を組み合わせる。最後に、3 つの信号成分を合計して（ステップ 9 2 0）、和を形成し、改良 M B E デコーダの出力とする。

【 0 0 8 3 】

尚、ここに記載した技法は、APCO Project 25 通信システムおよび当該システムが用いる標準的な 7 2 0 0 b p s M B E ボコーダに関するものであったが、ここに記載した技法は、他のシステムおよび / またはボコーダにも容易に適用可能である。例えば、他の既存の通信システム（例えば、FAA NEXCOM, Inmarsat、および ETSI GMR）が M B E 型ボコーダを用いると、前述の技法の効果が得られる。加えて、前述の技法は、異なるビット・レ

10

20

30

40

50

ートまたはフレーム・レートで動作する音声コーディング・システム、または代替りのパラメータ（例えば、S T C、M E L P、M B - H T C、C E L P、H V X Cまたはその他）を有する異なる音声モデルを用いる音声コーディング・システム、あるいは分析、量子化および/または合成に異なる方法を用いる音声コーディング・システム等、多くの他の音声コーディング・システムにも適用可能である。

【0084】

その他の実現例も、本発明の範囲内に該当するものとする。

【図面の簡単な説明】

【0085】

【図1】図1は、改良したM B Eエンコーダ・ユニットおよび改良したM B Eデコーダ・ユニットを有する、改良したM B Eボコーダを含むシステムのブロック図である。 10

【図2】図2は、図1のシステムの改良M B Eエンコーダ・ユニットおよび改良M B Eデコーダのブロック図である。

【図3】図3は、図2のエンコーダ・ユニットのM B Eパラメータ推定エレメントが用いる手順のフローチャートである。

【図4】図4は、図3のM B Eパラメータ推定エレメントのトーン検出エレメントが用いる手順のフローチャートである。

【図5】図5は、図3のM B Eパラメータ推定エレメントの発声活動検出エレメントが用いる手順のフローチャートである。

【図6】図6は、改良M B Eエンコーダにおいて基本周波数および発声パラメータを推定する際に用いる手順のフローチャートである。 20

【図7】図7は、図2のデコーダ・ユニットのM B Eパラメータ再現エレメントが用いる手順のフローチャートである。

【図8】図8は、改良M B Eデコーダにおける基本周波数および発声パラメータを再現するのに使用する手順のフローチャートである。

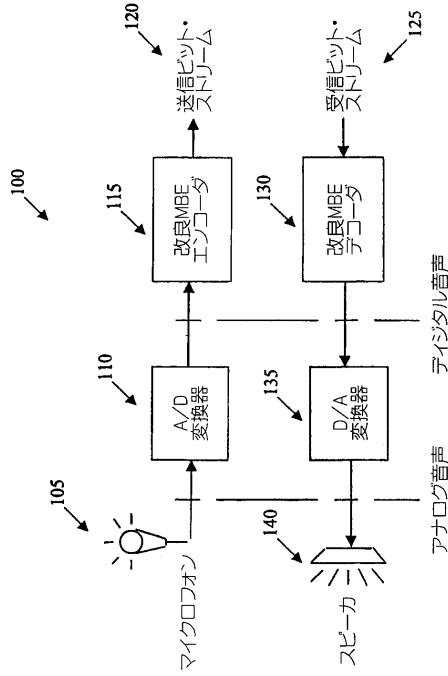
【図9】図9は、図2のデコーダのM B E音声合成エレメントのブロック図である。

【符号の説明】

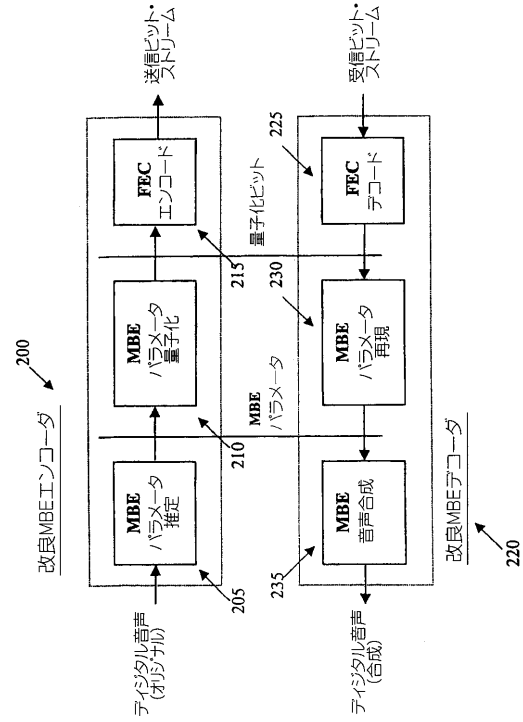
【0086】

| | | |
|-----|-----------------------|----|
| 100 | ボコーダ | |
| 105 | マイクロフォン | 30 |
| 110 | A / D変換器 | |
| 115 | 改良M B E音声エンコーダ・ユニット | |
| 120 | デジタル・ビット・ストリーム | |
| 125 | 受信ビット・ストリーム | |
| 130 | 改良M B E音声デコーダ・ユニット | |
| 135 | D / A変換ユニット | |
| 200 | 音声エンコーダ・ユニット | |
| 205 | パラメータ推定ユニット | |
| 210 | M B Eパラメータ量子化ユニット | |
| 215 | F E Cエンコード・パリティ付加ユニット | 40 |
| 220 | M B E音声デコーダ・ユニット | |
| 225 | F E Cデコーダ・ユニット | |
| 230 | M B Eパラメータ再現ユニット | |
| 235 | M B E音声合成ユニット | |

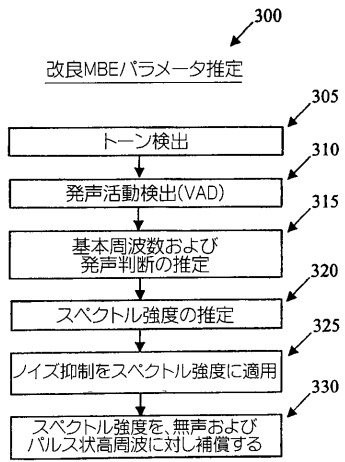
【図1】



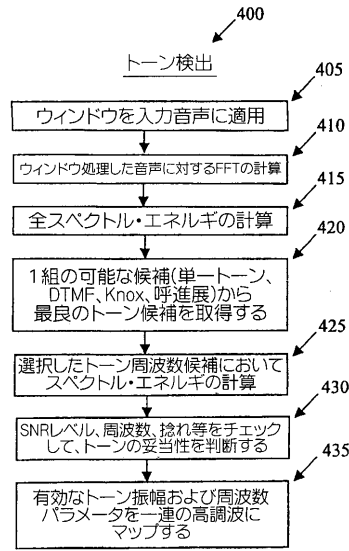
【図2】



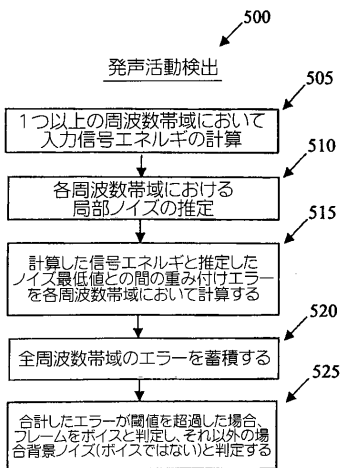
【図3】



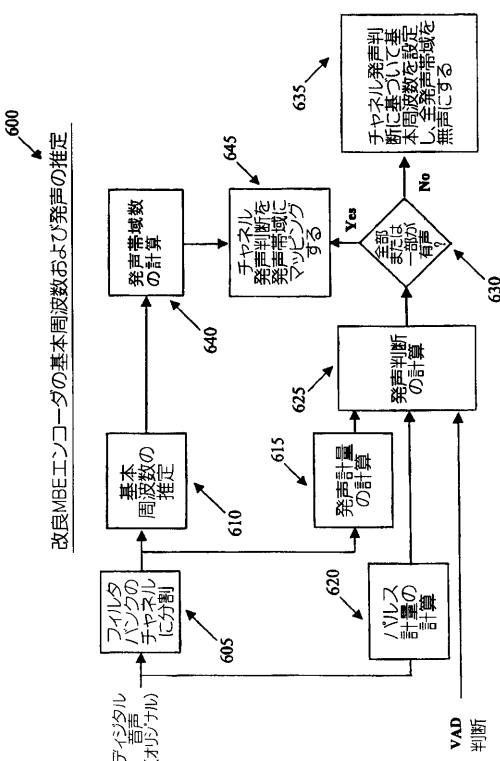
【図4】



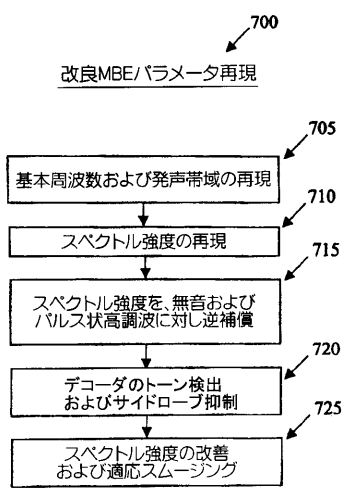
【図5】



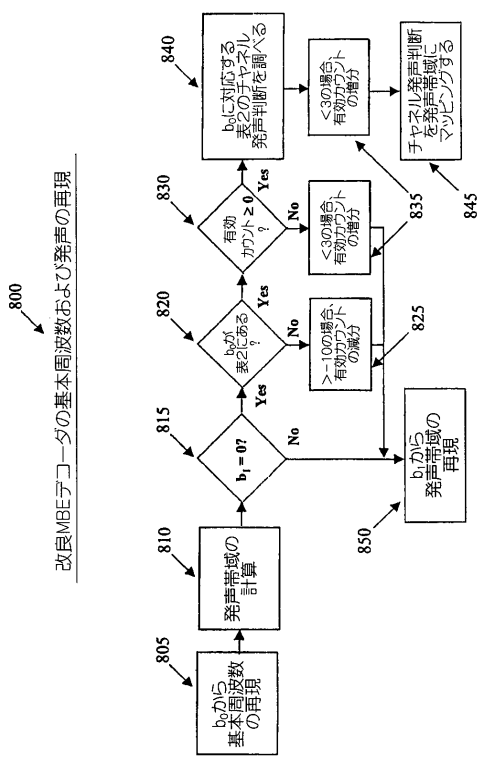
【図6】



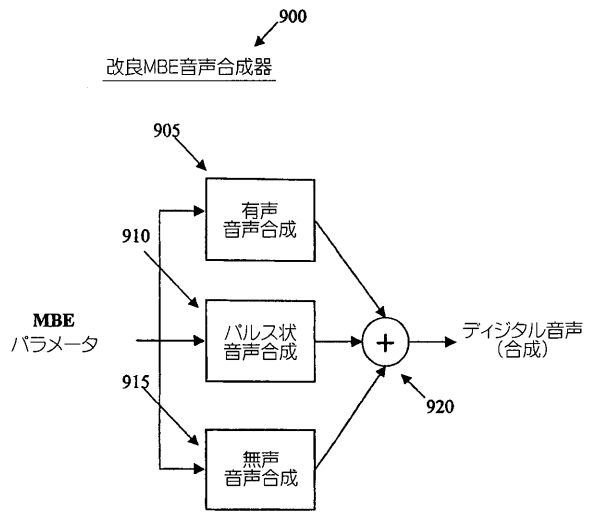
【図7】



【図8】



【図9】



フロントページの続き

(74)代理人 100096013

弁理士 富田 博行

(74)代理人 100120112

弁理士 中西 基晴

(72)発明者 ジョン・シー・ハードウィック

アメリカ合衆国マサチューセッツ州01776, サッドバリー, キャンパーダウン・レーン 75

審査官 山下 剛史

(56)参考文献 特開平05-346797(JP, A)

特開平10-293600(JP, A)

特開2001-177416(JP, A)

特開平11-122120(JP, A)

特開平07-044193(JP, A)

(58)調査した分野(Int.Cl., DB名)

G10L 11/00 - 11/06, 19/00 - 21/06