

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4975213号
(P4975213)

(45) 発行日 平成24年7月11日(2012.7.11)

(24) 登録日 平成24年4月20日(2012.4.20)

(51) Int.Cl.

G 1 O L 19/00 (2006.01)

F 1

G 1 O L 19/00 330 E

請求項の数 5 (全 22 頁)

(21) 出願番号 特願2000-612928 (P2000-612928)
 (86) (22) 出願日 平成12年4月19日 (2000.4.19)
 (65) 公表番号 特表2002-542520 (P2002-542520A)
 (43) 公表日 平成14年12月10日 (2002.12.10)
 (86) 國際出願番号 PCT/US2000/010577
 (87) 國際公開番号 WO2000/063884
 (87) 國際公開日 平成12年10月26日 (2000.10.26)
 審査請求日 平成19年4月16日 (2007.4.16)
 (31) 優先権主張番号 60/130,016
 (32) 優先日 平成11年4月19日 (1999.4.19)
 (33) 優先権主張国 米国(US)

(73) 特許権者 390035493
 エイ・ティ・アンド・ティ・コーポレーション
 A T & T C O R P.
 アメリカ合衆国 10013-2412
 ニューヨーク ニューヨーク アヴェニュー
 オブ ジ アメリカズ 32
 (74) 代理人 100075258
 弁理士 吉田 研二
 (74) 代理人 100096976
 弁理士 石田 純
 (72) 発明者 カピロウ ディビッド エイ
 アメリカ合衆国 ニュージャージー州 バークリー ハイツ ワシントン ストリート 122

最終頁に続く

(54) 【発明の名称】フレーム消去隠蔽処理装置

(57) 【特許請求の範囲】

【請求項 1】

フレーム消去隠蔽処理装置であつて、
 検出器と、
 前記検出器に接続された復号器と、
 前記復号器に結合されたバッファ、および前記バッファに結合された遅延モジュールを
 含んで構成され、前記フレーム消去隠蔽処理装置の出力ポートに音声サンプルを出力する
 メモリと、

前記バッファと前記検出器とに接続された FEC モジュールと、
 を備え、

前記検出器は、
 音声情報を含む符号化情報に応じて動作し、前記符号化情報が復号化可能なフレームに
 相当すると判断した場合は、復号化可能なフレームを前記復号器に与え、前記符号化情報
 が予期されていないフレームに相当すると判断した場合は、前記 FEC モジュールにその
 旨を知らせ、前記予期されていないフレームには、フレームが遺失、破損、消去、または
 遅延したものが含まれ、

前記復号器は、

前記復号化可能なフレームを音声サンプルを含む音声セグメントに復号化し、その音声
 セグメントを前記バッファに記憶し、

前記 FEC モジュールは更新手段を備え、

10

20

前記更新手段は、

前記予期されていないフレームに関する知らせが前記FECモジュールに与えられ、且つ、前記予期されていないフレームの直後のフレームが復号化可能である場合に、前記バッファに記憶された音声サンプルを含む合成音声セグメントを、前記予期されていないフレームの前記直後のフレームとのOLA処理を用いることで生成し、前記OLA処理は、前記合成音声セグメントと前記予期されていないフレームの前記直後のフレームとの間の境界部に適用し、前記合成音声セグメントを前記バッファに記憶し、前記OLA処理に用いるOLA窓の時間長は、他の先行する予期されていないフレームに対する前記予期されていないフレームの位置に基づくことを特徴とするフレーム消去隠蔽処理装置。

【請求項2】

10

請求項1に記載のフレーム消去隠蔽処理装置において、

前記予期されていないフレームが長い位置にある程、前記OLA処理における前記OLA窓の時間長を長くすることを特徴とするフレーム消去隠蔽処理装置。

【請求項3】

請求項1または請求項2に記載のフレーム消去隠蔽処理装置において、

前記FECモジュールは、

前記バッファに記憶された音声サンプルに基づいて、音声ピッチ周期を推定する手段を備え、

前記更新手段は、

第1の予期されていないフレームに関する知らせが前記FECモジュールに与えられた場合に、前記推定された音声ピッチ周期に基づいて定められる第1時間長の音声セグメントを前記バッファから取得し、前記第1時間長の音声セグメントを用いて第1合成音声セグメントを生成し、

第2の予期されていないフレームに関する知らせが前記FECモジュールに与えられ、前記第2の予期されていないフレームが前記予期されていないフレームである場合に、推定された音声ピッチ周期に基づいて定められ前記第1時間長より長い第2時間長の音声セグメントを前記バッファから取得し、前記第2時間長の音声セグメントを用いて第2合成音声セグメントを生成し、その際に、OLA処理を前記第1合成音声セグメントと前記第2合成音声セグメントとの間の境界部に適用し、前記第1合成音声セグメントを前記履歴バッファに記憶する、

ことを特徴とするフレーム消去隠蔽処理装置。

【請求項4】

30

請求項3に記載のフレーム消去隠蔽処理装置において、

前記第1時間長は、

推定された音声ピッチ周期の自然数N倍の時間に、その音声ピッチ周期についての4分の1波長に相当する時間を加えた時間長であり、

前記第2時間長は、

推定された音声ピッチ周期の自然数(N+1)倍の時間に、その音声ピッチ周期についての4分の1波長に相当する時間を加えた時間長であることを特徴とするフレーム消去隠蔽処理装置。

【請求項5】

40

請求項1または請求項2に記載のフレーム消去隠蔽処理装置において、

前記FECモジュールは、

前記履歴バッファに記憶された音声サンプルに基づいて、音声ピッチ周期を推定する手段を備え、

前記更新手段は、

推定された音声ピッチ周期および連続する予期されていないフレームの数に基づいた時間長の音声セグメントを前記バッファから取得し、取得した音声セグメントを用いて前記合成音声セグメントを生成する、

ことを特徴とするフレーム消去隠蔽処理装置。

50

【発明の詳細な説明】**【0001】****(技術分野)**

この仮特許出願ではない特許出願は、1999年4月19日に出願された、米国特許仮出願第60/130,016号の利益を主張し、その主題をここに引用し編入する。以下の文書もまた、ここに引用し編入する。すなわち、ITU-T Recommendation G.711- Appendix I, "A high quality low complexity algorithm for packet loss concealment with G.711" (9/99) 及び American National Standard for Telecommunications - Packet Loss Concealment for Use with ITU-T Recommendation G.711 (T1.521-1999) である。

【0002】

10

本発明はパケットロス又はフレーム消去の隠蔽 (Frame Erasure Concealment) (FEC) を実行する技法に関する。

【0003】**(背景技術)**

フレーム消去の隠蔽 (FEC) アルゴリズムは、音声通信システムにおける送信ロスを隠し、このシステムでは、入力スピーチ信号が、送信機において符号化及びパケット化され、ネットワーク上（どのような種類のものでもよい）に送信され、パケットを復号しスピーチ出力を再生する受信器において受信される。G.723.1やG.728やG.729などの標準的なCELPに基づくスピーチコーダの多くが、FECアルゴリズムを内蔵する又はその規格において提案されている。

20

【0004】

FECの目的は、合成スピーチ信号を生成し、受信したビットストリーム内の欠落データを補うことである。理想的には合成信号は、欠落した信号と同じ音色 (timbre) 及びスペクトル特性を有し、不自然なアーティファクト (unnatural artifact) を作り出さない。スピーチ信号はしばしば、局地的に静止しているので、信号の過去の履歴を使い、欠落部分の妥当な近似を生成することが可能である。消去が長すぎなければ、そして、信号が素早く変化している領域において消去が起きなければ、隠蔽によって、消去を聞こえないほどにすることもできる。

【0005】

30

従来のシステムは、フレーム消去を隠蔽するためにピッチ波形複製技術 (pitch waveform replication techniques) を利用する。この複製技法は例えば、D. J. Goodman et al., Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice Communications, Vol. 34, No. 6, IEEE Trans. on Acoustics, Speech, and Signal Processing 1440 - 48 (December 1996) や、O. J. Wasem et al., The Effect of Waveform Substitution on the Quality of PCM Packet Communications, Vol. 36, No. 3, IEEE Transactions on Acoustics, Speech, and Signal Processing 342 - 48 (March 1988)などを含む。

【0006】

40

ピッチ波形複製及びオーバーラップ加算技術を使い、音声データの失われたフレームを隠蔽する信号を合成していたが、これらの技術では時に、聞き手に不快感を与える「ピー」という音のアーティファクトを生じる。

【0007】**(発明の開示)**

本発明は、音声符号化システムの復号器によって生成された音声中の、欠落した音声情報の影響を隠蔽する処理に関する。特に本発明は、利用できないパケット（例えば、フレーム消去による欠落した音声情報など）を隠蔽するために合成された音声と、消去終了後、受信された音声情報から続いて復号される音声との、潤滑な組合せの方法に関する。この組合せは、オーバーラップ加算技術を使用して実行され、このオーバーラップ加算技術は、合成信号と、復号された信号とを混和 (blend) する。しかしながら従来技術と違い、オーバーラップ加算窓の正しい本質 (nature) は、消去の長さの関数である。したが

50

って例えれば、より長い消去においては、より長いオーバーラップ加算窓が使われる。

【0008】

遺失フレーム検出器が、符号化されたフレームが消去されたと判断した場合は、FECモジュールがフレーム隠蔽処理を信号に施す。FEC処理は、消去されたフレームにおいて、自然に聞こえる合成音声を生成する。

【0009】

本発明は、音声符号化システム処理のパケットロス又はフレーム消去の隠蔽（FEC）を実行する方法又は装置に関する。符号化されたフレームが消去された場合、フレーム隠蔽処理が信号に施される。この処理は、欠落した音声を合成するためにピッチ波形の複製を使うが、従来の技術と違いこの処理は、消去の長さが増加するに従って、ピッチ波形の数を増やして複製（replicate）する。FEC処理は、消去されたフレームに対して、自然に聞こえる合成音声を生成することによりこの分野における進歩を作り出す。10

【0010】

（発明を実施するための最良の形態）

近年、サービス品質の保証無しにG.711をパケットネットワーク上で使い公衆電話交換ネットワーク（Plain-Old Telephony Service）（POTS）をサポートすることに関心が集まっている。フレーム消去（又はパケットロス）が、これらのネットワーク上で発生した場合、呼の品質を著しく劣化させないために、隠蔽技法が必要とされる。高品質で複雑さが低いフレーム消去隠蔽（FEC）技法が開発され、以下に詳述する。

【0011】

FECを備える音声システムを例示するブロック図を図1に示す。図1において、符号器110が入力音声フレームを受信し、符号化されたビットストリームを出力する。ビットストリームは、遺失フレーム検出器115によって受信され、この遺失フレーム検出器115は、いずれかのフレームが失われたかどうかを判断する。遺失フレーム検出器115が、フレームが失われたと判断すると、遺失フレーム検出器115は、FECモジュール130に信号を送り、FECアルゴリズム又は処理を施し、欠落したフレームを再構築する。20

【0012】

よって、FEC処理は、音声システムにおける送信の遺失を隠す。ここで音声システムは、入力信号が送信器において符号化及びパケット化され、ネットワーク上に送られ、フレームが失われたことを判断する遺失フレーム検出器115において受信されるようなものである。図1において、遺失フレーム検出器115は、予期されるフレームが到着しないこと又は使用不能なほど遅れて到着したことを判断する方法を備えているものと仮定される。IPネットワーク上ではこれは通常、送信されるフレーム内のデータに、シーケンスナンバ又はタイムスタンプを追加することにより実施される。遺失フレーム検出器115は、到着したフレームのシーケンスナンバと、フレームが失われていない場合に予期されるシーケンス数とを比較する。遺失フレーム検出器115が、予期された時にフレームが到着したと検出した場合、フレームは復号器120によって復号され、音声の出力フレームが出力システムに与えられる。フレームが失われていた場合、FECモジュール130が、当該フレーム分の合成音声を、その欠落したフレームの代わりに生成することにより、欠落した音声フレームを隠す処理を施す。30

【0013】

G.723.1、G.728、及びG.729などの標準ITU-T CELPに基づくスピーチコーダの多くは、その復号器において音声の再生をモデル化する。よって復号器は、FEC処理を直接復号器に一体化させるために十分な状態情報を有する。これらのスピーチコーダでは、FECアルゴリズム又は処理が、その規格の一部として指定されている。

【0014】

これに対してG.711は、サンプル毎の符号化機構であり、音声の再現をモデル化しない。FECを助けるための状態情報がコーダ内に存在しない。その結果、G.711にお4050

ける F E C 处理は、コーダとは独立したものである。

【 0 0 1 5 】

G . 7 1 1 コーダと共に使われるシステムを例示するブロック図を図 2 に示す。図 1 と同様に、G . 7 1 1 符号器 2 1 0 は、ビットストリームデータを符号化し、遺失フレーム検出器 2 1 5 に送信する。ここでも、遺失フレーム検出器 2 1 5 が、到着するフレームのシーケンスナンバを、フレームが一つも失われていない時に予期されるシーケンスナンバと比較する。予期された時にフレームが到着した場合、フレームは復号器 2 2 0 による復号のために転送され、信号を記憶する履歴バッファ 2 4 0 に出力される。フレームが失われていた場合、遺失フレーム検出器 2 1 5 は、F E C モジュール 2 3 0 に知らせ、F E C モジュール 2 3 0 は、そのフレーム分の合成音声を、当該欠落したフレームの代わりに生成することにより、欠落した音声フレームを隠す処理を施す。
10

【 0 0 1 6 】

しかしながら、欠落フレームを隠すためには、F E C モジュール 2 3 0 は、G . 7 1 1 F E C 処理を施す。この処理では、履歴バッファ 2 4 0 によって提供される、復号された出力信号の過去の履歴を使い、欠落フレーム内の信号がどのようなものであるべきかを推定する。これに加え、消去されたフレームと消去されなかったフレームとの間の潤滑な移行を保証するために、遅延モジュール 2 5 0 が、システムの出力を、例えば 3 . 7 5 ミリ秒 (msec) などの所定の期間、遅延させる。この遅延により、合成消去信号が、消去の開始時において本物の出力信号とゆっくりと混合されるようになる。
20

【 0 0 1 7 】

F E C モジュール 2 3 0 と、履歴バッファ 2 4 0 のブロック及び遅延モジュール 2 5 0 の各ブロックの間の矢印は、F E C 処理によって記憶された履歴が使用され、合成信号が生成されることを示す。これに加え、F E C モジュール 2 3 0 の出力は、消去部分の期間中、履歴バッファ 2 4 0 を更新するために使われる。F E C 処理は、G . 7 1 1 の復号された出力のみに依存するので、この処理はスピーチコーダが存在しない場合でも同様に正しく機能することに留意すべきである。
20

【 0 0 1 8 】

F E C モジュール 2 3 0 において、F E C 処理によって入力信号がどのように処理されるのかを、図 3 にグラフで例示する。
30

【 0 0 1 9 】

図中最上段の波形は、男性の話者による有声音声の、ある領域において 2 0 ミリ秒の消去が発生した場合のシステムへの入力を示す。その下の波形では、F E C 処理がこの間隙内で合成音声を生成することにより、欠落した部分を隠蔽している。比較の目的で、消去されていないオリジナルの入力信号も示す。理想的なシステムでは、隠蔽された音声は、元の音声と同じように聞こえる。図から明らかなように、合成波形は、欠落部分の元の音声に非常に似ている。「隠蔽された」波形が、「入力」波形からどのように生成されるかを、より詳しく下に述べる。
30

【 0 0 2 0 】

F E C モジュール 2 3 0 によって使われるF E C 処理は、履歴バッファ 2 4 0 に記憶されている音声と同様の特徴を持つ合成音声を生成することにより、欠落フレームを隠蔽する。基本的な考え方は次のようなものである。信号が声に出されていた場合、信号は、準周期的であり局地的に一様であると仮定する。ピッチを推定し、履歴バッファ 2 4 0 内で最後に現れたピッチ周期を数回繰り返す。しかしながら、消去が長い場合又はピッチが短い場合（周波数が高い場合）、同じピッチ周期を繰り返しすぎると、自然な音声と比較して、正弦成分 (harmonic) の多すぎる出力に繋がる。これらの、ビーという音やボーンという音として聞こえる人工的な正弦波的音声 (ハーモニック・アーティファクト) を避けるために、履歴バッファ 2 4 0 から使われるピッチ周期の数は、消去の長さが大きくなるにつれて増やされる。短い消去部分では、最後のピッチ周期又は最後のいくつかのピッチ周期のみを履歴バッファ 2 4 0 から使い、合成信号を生成する。長い消去部分に対しては、履歴バッファ 2 4 0 の更に以前のピッチ周期も使う。長い消去部分に対しては、履歴バッ
40
50

ファ 2 4 0 からのピッチ周期は、オリジナルの音声に現れた順序と異なる順序で再生される。しかしながら、実験により、長い消去の場合に生成される合成音声信号も、自然な音を作り出すことが確かめられている。

【 0 0 2 1 】

消去部分が長ければ長いほど、合成信号が本物の信号から逸脱する可能性が高くなる。特定の種類の音を長く持続させることによって引き起こされるアーティファクトを回避するために、合成信号は、消去部分が長くなるにしたがって減衰される。10ミリ秒以下の期間の消去部分においては、減衰は必要とされない。10ミリ秒よりも長い消去部分には、合成信号は追加の10ミリ秒毎に20%の割合で減衰される。60ミリ秒を過ぎると、合成信号はゼロ（無音）に設定される。これは、合成信号が元の信号と異なりすぎるため、平均的には、欠落した音声を60ミリ秒以降も隠蔽しようと努力し続けても、利点よりも害が多くなるためである。10

【 0 0 2 2 】

異なる源（source）からの信号間で移行（transition）が行われた場合には、この移行が非連続性を導入しないことが重要となる。この非連続性は、出力信号中のクリック又は不自然なアーティファクトとして聞こえる。これらの移行は、いくつかの場所で起きる。

【 0 0 2 3 】

1. 消去部分のはじめの部分、つまり合成信号の開始部と最後の良いフレームの尾部との境界。

【 0 0 2 4 】

2. 消去部分の終りの部分、つまり合成信号と消去部分の後の最初の良いフレームの開始部との境界。20

【 0 0 2 5 】

3. 履歴バッファ 2 4 0 から使われるピッチ周期の数が、信号のバリエーションを増加させるために変更された時。

【 0 0 2 6 】

4. 履歴バッファ 2 4 0 の繰り返し部分間の境界。

【 0 0 2 7 】

潤滑な移行を保証するため、オーバーラップ加算（Overlap Adds）（O L A）が全ての信号境界において実行される。O L A は、一つの端部において重複する二つの信号を潤滑に組合わせる方法である。信号が重複する領域では、信号が窓によって重みを与えられ、その後、足される（混合される）。窓は、どのサンプルにおいても、重みの和が1になるように設計される。すなわち、信号の全体的な和に対する利得又は減衰は適用されない。これに加え、窓は、左側の信号では重みが1から始まり徐々に0へとフェードアウトし、右側の信号では重みが0から始まり徐々に1へとフェードインするよう設計される。よって、重複する窓の左側の領域では、左信号のみが存在し、重複する窓の右側の領域では、右信号のみが存在する。重複した領域内では、信号が左側の信号から右側の信号へと徐々に移行する。F E C 処理では、三角形の窓が使われ、長さを変更可能な窓の計算の複雑さを低く保つが、例えばハニーフィング窓（Hanning window）などの他の窓をその代わりに使っても良い。30

【 0 0 2 8 】

図4は、20ミリ秒間の消去の終了時における、合成音声が、消去部分の後に始まる本物の音声とO L Aされるものを示す。この例では、O L A重み付け窓は、5.75ミリ秒の三角形の窓である。最上段の信号は、消去中に対応して生成された合成信号であり、その下のオーバーラップした信号は、消去部分の後の本物の音声である。O L A重み付け窓は、信号の下に示される。ここでは、消去期間中の本物の信号内でのピッチの変化によって、合成信号のピークと本物の信号のピークとが一致せず、O L A無しに信号を組合わせようとした場合に導入される非連続性を、「O L A無しの組合せ」とラベルを付けたグラフに示す。「O L A無しの組合せ」グラフは、O L A窓の開始までの合成信号をと、その期間中の本物の信号とを複写することにより作成されている。O L A作業の結果は、4050

境界での非連続性がどのように潤滑にされるかを示す。

【0029】

前述の論議は、ほぼ一様な、有聲音声に、処理の例がどのように働くかに関連するものであるが、音声信号が素早く変化している場合又は声に出されていない場合、音声信号は、周期的な構造を持たない場合もある。しかしながら、これらの信号も以下に示すように同じ方法で処理される。

【0030】

第一に、例示する実施形態で許容される最小のピッチ周期のピッチ推定値は5ミリ秒であり、これは200Hzの周波数に対応する。高周波の女性及び子供の話者の一部は、200Hz以上の基本振動数を有することが知られているが、窓を比較的大きく保つために、200Hzに制限する。この方法では、10ミリ秒間の消去フレーム内では、選択されたピッチ周期は、最大二回繰り返される。高周波の話者の場合、これは、ピッチ推定法が本物のピッチ周期の倍数を返すため、実際には出力が劣化しない。いずれの音声も多く繰り返しすぎないようにすることにより、処理は、非周期的な音声から周期的な合成音声を作り出さない。第二に、合成音声を生成するために使われるピッチ周期の数は、消去部分が長くなるにしたがって増やされるため、信号に十分な多様性が追加され、長い消去部分においても周期性が導入されない。

10

【0031】

音声の時間スケールのための波形類似オーバーラップ加算(Waveform Similarity Overlap Add)(WSOLA)処理もまた、大きな固定された大きさのOLA窓を使用し、周期的な音声信号及び非周期的な音声信号の両方の時間スケールを変更するために同じ処理が使えることに留意されたい。

20

【0032】

例示するFEC処理の全体像が上で説明されたが、個々のステップを下に、より詳しく説明する。

【0033】

説明の目的のために、フレームが10ミリ秒間の音声を含み、サンプリング速度が例えば8kHzであると仮定する。よって、消去は80サンプル($8000 * 0.010 = 80$)ごとに起きる可能性がある。FEC処理は、他のフレームの大きさ及びサンプリング速度に対しても容易に適合できることに留意すべきである。サンプリング速度を変更するには、単純に、ミリ秒単位で与えられる時間に0.001を掛け、その後サンプリング速度でかけて適切なバッファの大きさを取得する。例えば、履歴バッファ240は、最近 $48.75 \text{ミリ秒} * 8000 = 390$ サンプル分の長さであることを意味する。16kHzのサンプリングでは、バッファの大きさは二倍又は780サンプルとなる。

30

【0034】

バッファの大きさのいくつかは、処理が予想する最低の周波数に基づく。例えば、ここに例示する処理は、8kHzのサンプリングで見られる最低周波数が $66 + (2/3) \text{Hz}$ であると仮定する。これにより、最大ピッチ周期は、 $1 / (66 + (2/3)) = 0.015$ と導かれる。履歴バッファ240の長さは、最低周波数の周期の 3.25 倍である。よって、履歴バッファ240は $15 * 3.25 = 48.75$ ミリ秒となる。16kHzのサンプリングで、入力フィルタが50Hz(20ミリ秒の周期)までの低い周波数を許容する場合、履歴バッファ240は、 $20 * 3.25 = 65$ ミリ秒に伸ばされる。

40

【0035】

フレームの大きさもまた変更できる。10ミリ秒間がデフォルトとして選択されたのは、G.729などのいくつかの標準スピーチコーダによって使われるフレームの大きさであり、またいくつかの無線システムでも使われているからである。フレームの大きさを変更することは簡単である。望まれるフレームサイズが10ミリ秒の倍数である場合、処理は変更しない。単純に、消去処理のフレームサイズを10ミリ秒のままにして、それを一つ

50

レームに対して複数回呼び出す。望まれるパケットフレームサイズが、5ミリ秒など10ミリ秒の約数である場合、FEC処理は基本的には変更しない。しかしながら、ピッチバッファ内の周期を増大させたレートを、10ミリ秒間のフレーム数に基づいて変更する必要がある。12ミリ秒など、10ミリ秒の倍数でも約数でもないフレームサイズもまた順応させることができる。FEC処理は、ピッチバッファから使われるピッチ周期の数を増大させたレートを変更することに対し、柔軟性を有する。周期の数を、10ミリ秒毎から12ミリ秒毎に増やすことは、さほどの違いを生じない。

【0036】

図5は、図2に例示する実施形態において実行されるFEC処理のブロック図である。主要な機能のいくつかを実施するために必要とされる副ステップを、図7、図12、及び図16に更に詳しく示し、下で説明する。以下の論議では、値及びバッファを保持するためにはいくつかの変数が使われる。これらの変数の概要を次の表に示す。

【0037】

【表1】

変数	種類	説明	コメント
B	行列	ピッチバッファ	範囲 [-P*3.25 : -1]
H	行列	履歴バッファ	範囲 [-390 : -1]
L	行列	最終1/4バッファ	範囲 [-P*.25 : -1]
O	スカラ	ピッチバッファ内のオフセット	
P	スカラ	ピッチ推定値	40 <= P < 120
P4	スカラ	1/4ピッチ推定値	P4 = P >> 2
S	行列	合成スピーチ	範囲 [0 : 79]
U	スカラ	使われる波長	1 <= U <= 3

【0038】

図5のフローチャートに示すように、処理はステップ505から始まり、このステップでは、遺失フレーム検出器215によって次のフレームが受信される。ステップ510では、遺失フレーム検出器215が、フレームが消去されたかどうかを判断する。フレームが消去されていなければ、ステップ512においてフレームは復号器220によって復号される。次に、ステップ515において、復号されたフレームが、FECモジュール230による使用のために、履歴バッファ240に記憶される。

【0039】

履歴バッファ更新ステップでは、このバッファ240の長さは、予想される最長ピッチ周期の長さの3.25倍である。8kHzのサンプリングでは、最長ピッチ周期は15ミリ秒又は120サンプルであり、よって履歴バッファ240の長さは48.75ミリ秒又は390サンプルである。したがって、各フレームが復号器220によって復号された後、履歴バッファ240は、最新の音声履歴を含むように更新される。履歴バッファ240の更新を図6に示す。この図に示すように、履歴バッファ240は、最新の音声サンプルを右側に含み、最古の音声サンプルを左側に含む。復号された音声の最新フレームが受信された時には、この最新フレームはバッファ240の右からシフトされ、最古の音声に対応するサンプルがバッファの左から外へとシフトされる(図6bを参照)。

【0040】

これに加え、ステップ520において、遅延モジュール250が、音声の出力を最長ピッチ周期の1/4だけ遅延させる。8kHzサンプリングではこれは、 $120 * 1 / 4 = 30$ サンプル又は3.75ミリ秒にあたる。この遅延により、FECモジュール230が1/4波長OLAを消去の開始時に実行することを可能にし、これにより、消去の前の本物の信号と、FECモジュール230によって作り出された合成信号との間の潤滑な移行を保証する。フレームを復号した直後は、次のフレームが消去されたかどうか分からぬため、出力を遅延させる必要がある。

【0041】

ステップ525において、音声が出力され、ステップ530において処理は、更にフレー

10

20

30

40

50

ムが存在するかどうかを判断する。更なるフレームが存在しなければ、処理は終了する。
更なるフレームが存在すれば、処理はステップ 505 に戻り次のフレームを取得する。

【0042】

しかしながら、ステップ 510において、喪失フレーム検出器 215 が、受信したフレームが消去されたものと判断した場合、処理はステップ 535 に進み、ここで FEC モジュール 230 が第一の消去されたフレームを隠蔽する。この処理は、下に図 7 に関して詳しく説明する。第一フレームが隠蔽された後、ステップ 540 において、喪失フレーム検出器 215 が、次のフレームを取得する。ステップ 545 において、喪失フレーム検出器 215 は、次のフレームが消去されているかどうかを判断する。もし次のフレームが消去されていなければ、ステップ 555 において FEC モジュール 230 が、消去後の第一フレームを処理する。この処理は、図 16 に関して下に詳しく説明する。第一フレームが処理された後、処理はステップ 530 に戻り、ここで喪失フレーム検出器 215 が更なるフレームが存在するかどうかを判断する。

【0043】

ステップ 545 で、喪失フレーム検出器 215 が、次又は後続のフレームが消去されていると判断した場合、FEC モジュール 230 が第二及び後続のフレームを隠蔽する。この処理は下に図 12 に関して詳しく説明する。

【0044】

図 7 に、消去の最初の 10 ミリ秒間を隠蔽するために採られるステップを詳しく示す。ステップは、以下に詳しく考察される。

20

【0045】

図 7 に示すように、ステップ 705 において、消去部分の開始時点における第一の作業は、ピッチの推定である。これを行うためには、タップ遅延が 40 から 120 サンプルの、20 ミリ秒 (160 サンプル) の窓を用いて、履歴バッファ 240 の信号に対し、正規化した自動相関が実行される。8 kHz のサンプリングにおいては、これらの遅延は、5 から 15 ミリ秒のピッチ周期又は 200 から 66 + (2/3) Hz の基本振動数に対応する。自動相関のピークにおけるタップが、ピッチ推定値 P である。H がこの履歴を含むものと仮定し、-1 (消去部分直前のサンプル) から -390 (消去部分が始まる 390 サンプル前のサンプル) まで索引を付けられると仮定すると、タップ j に対する自動相関は、数学的に次のように表すことができる。

30

【0046】

【数 1】

$$Autocor(j) = \frac{\sum_{i=1}^{160} H[-i]H[-i-j]}{\sqrt{\sum_{k=1}^{160} H^2[-k-j]}}$$

【0047】

自動相関のピーク又はピッチ推定値は、次のように表すことができる。

40

【0048】

【数 2】

$$P = \{ \max_j(Autocor(j)) \mid 40 \leq j \leq 120 \}$$

【0049】

上述のように、許容される最低ピッチ周期、5 ミリ秒又は 40 サンプルは、単一のピッチ周期が 10 ミリ秒である消去されたフレームの間に最大二回繰り返されるのに十分な大きさである。これにより、無音声中のアーティファクトを回避し、また、高ピッチの話者の不自然なハーモニックアーティファクトを回避する。

50

【0050】

図3の消去部分における、正規化した自動相関の計算例を、図8にグラフで示す。

【0051】

「履歴」とラベルを付けた波形は、消去部分直前の履歴バッファ240の内容である。水平な点線は、信号の基準部分、履歴バッファ240 H[-1] : H[-160]を示し、これは消去直前の20ミリ秒間の音声である。水平な直線は、40サンプル（最上段の線、5ミリ秒周期、200Hz周波数）から120サンプル（最下段の線、15ミリ秒周期、 $6.6 + (2/3) \text{ Hz}$ 周波数）のタップにおいて遅延された20ミリ秒間の窓である。相関の出力もまた、窓の場所と整列してグラフに表示される。相関中の垂直な点線は、曲線のピークであり、推定ピッチを表す。この線は、消去の開始時から一周期戻る。この場合、Pは、7ミリ秒のピッチ周期及び142.9Hzの基本振動数に対応する56サンプルに等しい。10

【0052】

自動相関の複雑さを低減するために、二つの特別な手順が使われる。これらの簡略化(shortcut)は、実質的には出力を変化させないが、処理の全体的な実行時間の複雑さに大きな影響を及ぼす。FEC処理の複雑さの大半は、自動相間に帰する。

【0053】

まず、全てのタップにおいて相関を計算する代わりに、ピークの大まかな推定を、省略された信号でまず判断し、その後、大まかなピークの近傍の領域において細かい検索が行われる。大まかな推定値のために、上述の自動相関(Autocor)関数を、2:1の省略された信号に作用する新しい関数に修正し、一つ置きのタップのみを検査する。20

【0054】

【数3】

$$Autocor_{rough}(j) = \frac{\sum_{i=1}^{80} H[-2i]H[-2i-j]}{\sqrt{\sum_{k=1}^{80} H^2[-2k-j]}}$$

$$P_{rough} = 2\{\max_j(Autocor_{rough}(2j)) | 20 \leq j \leq 60\}$$

【0055】

次に、この大まかな推定値を使い、元の検索処理を、 $P_{rough} - 1 \leq j \leq P_{rough} + 1$ の範囲のみで繰り返す。jが、元の40から120サンプルの範囲に納まるように注意が払われる。サンプリング速度が増やされた場合、省略率もまた増やすべきであり、これにより処理の全体的な複雑さがほぼ一定になるようにすべきであることに留意されたい。本発明者らは、8:1の省略率で、44.1kHzでサンプルされた音声において実験を行い、良好な結果を得た。図9は、簡略化自動相関(Autocor_{rough})のグラフと、通常の自動相関(Autocor)のグラフとを比較する図である。この図から明らかのように、Autocor_{rough}は、Autocorに近い近似であり、8kHzのサンプリングにおいては、複雑さがおよそ4倍に低減されている。この4倍は、一つおきのタップのみが検査されることにより2倍、与えられたタップにおいて、一つおきのサンプルのみが検査されることによる2倍、からである。40

【0056】

Autocor及びAutocor_{rough}のエネルギー計算の複雑さを低減させるために第二の手順が実行される。各ステップで、完全な和を計算する代わりに、エネルギーの継続和(running sum)が保持される。すなわち、

【0057】

【数4】

10

20

30

40

50

$$Energy(j) = \sum_{k=1}^{160} H^2[-k - j]$$

とすると、

【0058】

【数5】

$$Energy(j+1) = \sum_{k=1}^{160} H^2[-k - j - 1] = Energy(j) + H^2[-j - 161] - H^2[-j - 1]$$

となる。

10

【0059】

したがって、最初のエネルギーの項が計算された後は、2度の掛け算及び2度の足し算のみが、FEC処理の各ステップでエネルギーの項を更新するために必要となる。

【0060】

ピッチ推定値Pを得た後、消去部分の波形生成が開始される。図7のフローチャートに戻り、ステップ710において、最新の波長の3.25倍の部分(3.25 * Pサンプル)が、履歴バッファ240、すなわちHからピッチバッファBへと複写される。最新の1/4波長以外のピッチバッファの内容は、消去部分の間一定に保たれる。これに対し、履歴バッファ240は、消去部分の間、合成音声によって更新し続けられる。

【0061】

20

ステップ715において、履歴バッファ240からの最新の1/4波長分(0.25 * Pサンプル)が、最終クオーターバッファLに記憶される。この1/4波長分は、OLA作業のいくつかで必要とされる。簡便のため、履歴バッファ240にアクセスするために使ったものと同じ、負の数による索引付けの方法を使ってB及びLバッファにアクセスする。B[-1]は、消去部分が到着する前の最後のサンプルであり、B[-2]はその前のサンプル、などである。合成音声は、合成バッファSに入れられ、合成バッファSは、0から昇順に索引付けられる。よって、S[0]が最初の合成されたサンプルであり、S[1]が第二のサンプル、などとなる。

【0062】

30

図3に示す消去時の、ピッチバッファB及び最終クオーターバッファLの内容を、図10に示す。上述の部分で、周期Pは56サンプルと計算されている。よってピッチバッファは、3.25 * 56 = 182サンプルの長さである。最終クオーターバッファは、0.25 * 56 = 14サンプルの長さである。図においては、消去部分の開始時からPサンプル戻る毎に、垂直な線が配置されている。

【0063】

最初の10ミリ秒間の消去の間は、ピッチバッファからの最終ピッチ周期のみが使われ、よってステップ720において、U = 1である。音声信号が、真に周期的であり、ピッチ推定値が推定値ではなく、正確に真正な値であった場合、ピッチバッファBから合成バッファSに直接波形を複写でき、この場合、合成信号は滑らかであり、かつ連続している。すなわち、S[0] = B[-P]、S[1] = B[-P + 1]、などとなる。ピッチが10ミリ秒間のフレームよりも短い場合、すなわちP < 80の場合、消去フレーム内で、单一のピッチ周期が一回よりも多く繰り返される。ここに示す例では、P = 56であり、よって複写は、S[56]で戻って(roll over)くる。すなわち、サンプル56近くのサンプル毎の複写順序は、S[54] = B[-2]、S[55] = B[-1]、S[56] = B[-56]、S[57] = B[-55]、などとなる。

40

【0064】

実際には、ピッチ推定値は正確ではなく、信号は真に周期的でない場合もある。(a)本物の信号と合成信号との間の境界及び(b)周期が繰り返される境界における非連続性を回避するために、OLAが必要とされる。これら両方の境界において、本物の音声の終了時B[-1]から、一周期戻った音声B[-P]までの潤滑な移行が望ましい。したがつ

50

て、ステップ 725において、B [-P] の前の1/4 波長と、履歴バッファ240の最終1/4 波長又はLの内容と、をオーバーラップ加算することにより、これが達成される。グラフではこれは、ピッチバッファ内の最終 $1 + (1/4)$ 波長を取り出し、右に一波長分ずらし、OLAを1/4 波長のオーバーラップした領域において行うことに等しい。ステップ 730において、OLAの結果が、履歴バッファ240の最終1/4 波長に複写される。合成波形の追加の周期を生成するために、ピッチバッファが追加の波長分だけずらされ、追加のOLAが実行される。

【0065】

図11は、最初の2回の反復におけるOLA作業を示す。この図では、全ての波形を交差する垂直な線は、消去部分の開始時を示す。短い垂直線は、ピッチマークであり、消去境界からPサンプルの位置に配置される。波形「ピッチバッファ」と波形「Pだけ、右にずらされたもの」の間のオーバーラップ領域は、波形「Pだけ右にずらされたもの」と波形「2Pだけ、右にずらされたもの」の間のオーバーラップ領域と正確に同じサンプルに対応することを理解すべきである。したがって、1/4 波長OLAは、一回だけ計算されればよい。

10

【0066】

ステップ 735において、OLAをまず計算し、その結果をピッチバッファの最終1/4 波長に入れることにより、真に周期的な信号に対する、合成波形の生成処理を使用できる。サンプルB [-P] からはじめ、ピッチバッファから合成バッファへと単純にサンプルを複写し、ピッチバッファの最後部に達した時には、ピッチバッファポインタをピッチ周期の最初に戻す。この技法を使い、どのような期間の合成波形でも生成できる。図11の「OLAと組合わせた」の波形における、消去開始時より左側のピッチ周期は、ピッチバッファの更新された内容に対応する。

20

【0067】

「OLAとの組合わせ」波形は、单一周期のピッチバッファが、周期Pの周期的な信号を非連続性なしに生成することを示している。この履歴バッファ240内の单一の波長から生成された合成音声は、消去部分の第一の10ミリ秒間を隠蔽するために使われる。OLAの効果は、「ピッチバッファ」内の消去部分の開始直前の1/4 波長と「OLAと組合せた」波形内のそれとを比較することにより見ることができる。ステップ 730において、この「OLAと組合せた」波形内の1/4 波長はまた、履歴バッファ240内の最終1/4 波長を置換える。

30

【0068】

三角形の窓を使ったOLA作業もまた、数学的に表現することができる。まず、変数P4を、サンプル内のピッチ周期の1/4と定義する。よって、 $P4 = P > > 2$ である。ここに示す例では、Pが56なので、P4は14となる。OLA作業は、 $1 \leq i \leq P4$ の範囲で次のように表現できる。

【0069】

【数6】

$$B[-i] = \frac{i}{P4} L[-i] + \left(\frac{P4-i}{P4} \right) B[-i-P]$$

40

【0070】

OLAの結果は、履歴バッファ240及びピッチバッファの両方の最後の1/4 波長部分を置換える。履歴バッファ240を置換えることにより、履歴バッファもまた出力を3.75ミリ秒間遅延させて、履歴バッファ240が更新される時に、1/4 OLA移行が出力される。消去の第一の10ミリ秒間における出力波形は、図3の「隠蔽された」波形の最初の2つの点線間の領域に見ることができる。

【0071】

ステップ 740において、フレームの合成音声の生成の終了時に、現在のオフセットが変数0としてピッチバッファに記憶される。このオフセットは、合成波形が次のフレームに

50

続くことを可能にし、これにより、次のフレームの本物の信号又は合成信号とのOLAを可能にする。Oはまた、消去が10ミリ秒を超える場合に、正しい合成信号の位相を保持することを可能にする。ここに示す80サンプルのフレーム及びP=56の例では、消去の開始時には、オフセットは-56である。56サンプル後、オフセットは-56に戻る。更に、 $80 - 56 = 24$ サンプル後、オフセットは $-56 + 24 = -32$ であり、よって、第一のフレーム終了時にはOは-32である。

【0072】

ステップ745において、合成バッファにS[0]からS[79]が満たされた後、このSが履歴バッファ240を更新するために用いられる。ステップ750において、履歴バッファ240はまた、3.75ミリ秒の遅延を追加する。履歴バッファ240の取り扱い(handling)は、消去されたフレームの間も消去されなかったフレームの間も同じである。この時点で、図5のステップ535に示す第一フレームの隠蔽作業が終了し、処理は、図5のステップ540に進む。

10

【0073】

FECモジュール230が、図5のステップ550に示すように、10ミリ秒よりも後の後続のフレームを隠蔽するためにどのように機能するかを、図12に詳しく示す。第二及び後続の消去されたフレームにおいて、合成信号を生成する技法は、第一の消去されたフレームに類似しており、これに更に、信号にいくらかの多様性を足すための追加の作業が必要とされる。

20

【0074】

ステップ1205において、消去コードは、第二又は第三フレームが消去されているかどうかを判断する。第二及び第三の消去されたフレームの期間中は、ピッチバッファから使われるピッチ周期の数が増やされる。これにより信号に、より大きな多様性を導入し、合成出力が、ハーモニックしすぎて聞こえることを防止する。他の全ての移行と同様に、ピッチ周期数が増加された時には、境界を滑らかにするためにOLAが必要とされる。第三フレーム(30ミリ秒間の消去)より後は、ピッチバッファは、三波長の長さに一定に保たれる。これらの三波長が、消去の期間中の全合成音声を生成する。よって、図12の左側の分岐は、第二及び第三の消去されたフレームに対してのみ行われる。

【0075】

次に、ステップ1210において、ピッチバッファで使われる波長の数を増加させる。すなわち、U=U+1とする。

30

【0076】

第二又は第三の消去されたフレームの開始時に、ステップ1215において、一つ前のフレームからの合成信号が、更に1/4波長分、現在のフレームの開始部に続く。例えばここに示す例では、第二のフレームの開始時には、図13に示すように合成信号が現れる。この1/4波長は、ピッチバッファからの古い波長を使う新しい合成信号にオーバーラップ加算される。

【0077】

第二の消去フレームの開始時に、波長の数は2に増やされ、すなわちU=2とされる。单一波長ピッチバッファ同様、二波長(2-wavelength)ピッチバッファがそれ自身を繰り返す境界では、OLAを実行する必要がある。ここでは、ステップ1220において、ピッチバッファBの尾部から戻って1/4波長分の終わりのU波長が、最終クオーターバッファLの内容とオーバーラップ加算される。OLA演算子は、1 i P4の範囲で次のように表すことができる。

40

【0078】

【数7】

$$B[-i] = \frac{i}{P4} L[-i] + \left(\frac{P4-i}{P4} \right) B[-i - PU]$$

【0079】

50

前述の方程式との違いは、右辺の、Bのインデックスに使われた定数Pが、PUに置き換えられていることのみである。二波長ピッチバッファの作成を図14にグラフで示す。

【0080】

図11と同様、消去開始時の左側の「OLAと組合せた」波形は、二周期ピッチバッファの更新された内容である。短い垂直線は、ピッチ周期を表す。「OLAと組合せた」波形の連続したピークの詳しい検査により、ピークが、消去部分の開始時点より1又は2波長分前に戻ったピークから交互になることが示される。

【0081】

第二のフレームの合成出力の開始時には、新しいピッチバッファからの信号を、図13において生成された $1/4$ 波長と合成させる必要がある。新しいピッチバッファからの合成信号は、使用されるバッファの最古の部分からのものであることが望ましい。しかしながら、新しい部分が、波形の類似した部分からのものであることに注意をしないと、耳障りなアーティファクトが作り出されるので、そのような注意が必要である。すなわち、正確な位相を保持することが望ましく、保持できなければ、波形が混合される際に、相殺的干渉を引き起こす可能性がある。10

【0082】

これは、ステップ1225(図12)において、一つ前のフレームの終了時に記憶されたオフセットOから、ピッチバッファの使用部分内の最古の波長を指すまで、周期Pを引くことにより達成される。

【0083】

例えば、第一の消去されたフレームでは、ピッチバッファBの有効なインデックスは、-1から-Pであった。よって、第一の消去されたフレームから記憶されたOは、必ずこの範囲内にある。第二の消去されたフレームでは、有効な範囲は、-1から-2Pである。よって、Oが、-2P < O < -Pの範囲内になるまで、OからPを引く。より一般的に言うと、Oが、-UP < O < -(U-1)Pの範囲内になるまで、OからPを引く。ここに示す例では、第一の消去されたフレームの終了時において、P=56であり、O=-32である。-32から56を引いて、-88を得る。よって、第二フレーム内の第一の合成サンプルは、B[-88]から来る。次の合成サンプルは、B[-87]、などとなる。20

【0084】

第二の消去されたフレームの開始時における、単一及び二周期ピッチバッファからの合成信号のOLA混合を、図15に示す。30

【0085】

OからPを引くことにより、正確な波形位相が保持され、「1Pピッチバッファ」波形及び「2Pピッチバッファ」波形の信号のピークが整列されることに留意されたい。「OLAと組合せた」波形は、第二の消去されたフレームの開始時において、異なるピッチバッファ間の潤滑な移行を示す。図15に示す「OLAと組合せた」波形の第二フレームを出力できるまでに、もう一つの作業が必要とされる。40

【0086】

ステップ1230(図12)において、新しいオフセットが使われ、ピッチバッファから一時バッファに $1/4$ 波長が複写される。ステップ1235において、 $1/4$ 波長が、オフセットに足される。次に、ステップ1240において、一時バッファが、出力バッファの開始部とOLAされ、この結果が、出力バッファの第一の $1/4$ 波長に入れられる。40

【0087】

ステップ1245において、オフセットが使われ、出力バッファ内の信号の残りの部分が生成される。ピッチバッファが、10ミリ秒間のフレームの間、出力バッファに複写される。ステップ1250において、現在のオフセットが、変数Oとしてピッチバッファに記憶される。

【0088】

第二及び後続の消去されたフレームの期間中は、ステップ1255において、線形の傾斜によって合成信号が減衰される。合成信号は、60ミリ秒後に設定される0又は沈黙へ徐50

々にフェードアウトする。消去が長くなるに従って、隠蔽された音声が真の信号から逸脱する可能性が高くなる。特定の種類の音を長時間保留すると、その音が短い期間中に隔離された状態では自然に聞こえても、隠蔽処理の出力中の不自然に聞こえるアーティファクトを生じる場合もある。これらの合成信号中のアーティファクトを回避するために、ゆっくりとしたフェードアウトが使われる。同様の作業が、G . 723、G . 728、及びG . 729などの標準スピーチコーダにおける隠蔽処理においても行われている。

【0089】

FEC処理は、第二のフレームの開始時から、10ミリ秒間のフレーム毎に20%の割合で信号を減衰する。合成バッファSが減衰前の合成信号を含み、Fが連続して消去されたフレームの数（第一の消去されたフレームがF=1、第二の消去されたフレームがF=2など）である場合、減衰は、次のように表すことができる。

10

【0090】

【数8】

$$S'[i] = [1 - .2(F - 2) - \frac{.2i}{80}]S[i]$$

【0091】

ここで、0 i 79であり、2 F 6である。例えば、第二の消去されたフレームの開始時のサンプルでは、F=2であり、F-2=0及び0.2/80=0.0025である。よって、S'[0]=1·S[0]、S'[1]=0.9975S[1]、S'[2]=0.995S[2]、及びS'[79]=0.8025S[79]である。第六の消去されたフレーム以降は、出力は単純にゼロに設定される。

20

【0092】

合成信号は、ステップ1255において減衰された後、ステップ1260において、履歴バッファ240に与えられ、ステップ1265において、出力が3.75ミリ秒間、遅延される。オフセットポインタOもまた、ピッチバッファ内の第二フレームの終了時の位置に更新され、これにより、合成信号が次のフレームに続くことができる。処理は、ステップ540に戻り、次のフレームを取得する。

【0093】

消去が二フレーム以上続いた場合、第三フレームの処理は、第二のフレームと同様に行われるが、ピッチバッファ内の周期の数は、1から2ではなく、2から3へと増やされる。ここに示す例では、消去は二フレームで終了するが、第三フレーム及びその後に使われる、三周期ピッチバッファを図17に示す。第三フレーム以降は、ピッチバッファ内の周期の数は、3に固定されるため、図12の右側の経路のみが使われる。この場合、単純にオフセットポインタOは単純にピッチバッファを合成出力に複写するために使われ、オーバーラップ加算作業は必要とされない。

30

【0094】

消去後の、第一の良好フレームに対するFECモジュール230の作業を図16に詳しく示す。消去部分の終了時点では、消去中に生成された合成音声と、本物の音声との間に、潤滑な移行が必要とされる。消去が一フレームのみの長さであった場合、ステップ1610において、1/4波長のための合成信号が続けられ、本物の音声とのオーバーラップ加算が実行される。

40

【0095】

ステップ1620において、FECモジュール230が、消去が10ミリ秒よりも長かったと判断した場合、合成信号と本物の信号とが一致しない可能性が高く、よって、ステップ1630において、合成音声生成が続けられ、OLA窓が更に、消去されたフレーム一つにつき4ミリ秒間分増やされる。このOLA窓の増加は、最大10ミリ秒間まで行われる。ピッチの推定値が、多少ずれていた場合又は本物の音声のピッチが消去期間中に変化していた場合、合成信号と本物の信号との位相の不一致の可能性は、消去が長くなればなるほど増加する。より長いOLA窓により、合成信号が、よりゆっくりとフェードアウト

50

し、本物の音声信号が、よりゆっくりとフェードインすることを可能にする。消去が 10 ミリ秒よりも長かった場合、OLA を実行する前に、ステップ 1640において合成音声を減衰する必要があり、これにより合成信号を、一つ前のフレームの信号レベルと一致させる。

【0096】

ステップ 1650において、出力バッファの内容（合成音声）と、新しい入力フレームの開始部とに OLA が実行される。入力バッファの開始部は、OLA の結果によって置き換えられる。上述の例の消去部分の終了時点における OLA は、図 4 に見ることができる。上述の例の隠蔽処理の完全な出力は、図 3 の「隠蔽」波形に見ることができる。

【0097】

ステップ 1660において、履歴バッファが入力バッファの内容によって更新される。ステップ 1670において、音声の出力が 3.75 ミリ秒間遅延され、処理は図 5 のステップ 530 に戻り、次のフレームを取得する。

【0098】

小さな調整により、FEC 処理を G.726 など、サンプル間又はフレーム間の状態情報を保持し隠蔽を提供しないような、他のスピーチコーダに応用することもできる。FEC 処理は上述と全く同じように使われ、消去期間中の合成波形を生成する。しかしながら、コーダの内部状態変数が、FEC 処理によって生成された合成音声を探知することを保証するよう注意することが必要である。そうでなければ、消去部分の終了後、復号器が間違った状態を使って再始動するに従って、アーティファクト及び非連続部分が出力に現れる。消去部分終了時点の OLA 窓はこの作業を助けるが、OLA 窓以上のことをする必要がある。

【0099】

図 18 に示すように、消去の期間中、復号器 1820 を符号器 1860 に変換し、FEC モジュール 1830 の合成出力を符号器 1860 の入力として使うことにより、より良い結果が得られる。

【0100】

これにより、復号器 1820 の変数状態は、隠蔽された音声を探知できる。典型的な符号器と違い、符号器 1860 は、状態情報を保持するためのみに作動され、その出力は使われないことに留意されたい。よって、簡略化の手法（short cut）を使って実行時間中の複雑さを相当に低減できる。

【0101】

上述のように、本発明によって多くの利点及び特徴が提供される。特に、フレーム消去が進むにしたがって、合成信号を生成するために信号履歴から使われるピッチ周期が、時間の関数として増やされる。これにより、長い消去の際のハーモニック・アーティファクトをかなり低減できる。ピッチ周期は、元の順序で再生されないが、それでも出力は自然に聞こえる。

【0102】

サンプル間又はフレーム間の状態情報を保持する G.726 及び他のコーダでは、復号器を、隠蔽処理の合成出力の出力に対する符号器として作動させてもよい。これにより、復号器の内部状態変数が出力を探知し、復号器内の間違った状態情報に起因する、消去部分終了後の非連続性を回避し、又は少なくともそのような非連続性を低減できる。符号器からの出力は使われない（この符号器の唯一の目的は、状態情報を保持することである）ため、機能を省略した、複雑さが低い型の符号器を使うことができる。

【0103】

例示する実施形態（40 サンプル又は 200 Hz）で許容される最低ピッチ周期は、一部の女性及び子供の話者の基本振動数として予測されるものよりも大きい。よって、高周波数の話者の場合、消去部分の開始時点においても、合成音声を生成するために、1 よりも多いピッチ周期が使われる。基本振動数が高い話者の場合、より頻繁に波形が繰り返される。合成信号の複数のピッチ周期により、ハーモニック・アーティファクトの可能性が低

10

20

30

40

50

くなる。この技法はまた、音声内の無声部分や停止などの急激な移行部分において、信号が自然に聞こえるよう保つことを助ける。

【0104】

消去部分の後の最初の良好なフレームの終了時におけるOLA窓は、消去の長さと共に大きくなる。消去が長い場合、新しい良好フレームが到着した時に、位相の不一致が起きる可能性が高い。OLA窓を消去の長さの関数として伸ばすことにより、長い消去部分の際の位相不一致に起因する、突然の不調(glitch)を低減させ、同時に、消去が短い場合には、信号を迅速に回復させることを可能にする。

【0105】

本発明のFEC処理はまた、長さを変更可能なOLA窓を使う。このOLA窓は、1/4波長である推定ピッチの小さな断片であり、ピッチピークと整列していない。

【0106】

本発明のFEC処理は、有声(voiced)音声と、無声(unvoiced)音声とを区別しない。その代わりに、FEC処理は、処理の以下の二つの特性のため、無聲音声をうまく再現することができる。(A)最小の窓の大きさが、ほどよく大きいため、音声の無声領域もほどよい多様性を有する。(B)処理が進むにつれてピッチバッファの長さが増やされ、これによりハーモニック・アーティファクトが導入されないことを保証する。大きな窓を使い有聲音声と無聲音声に対する異なる取り扱いを避けることはまた、広く知られた時間スケール技法であるWSOLAにも存在することに留意すべきである。

【0107】

遅延を追加し、消去部分の開始時点におけるOLAを可能にすることは、本発明の処理においては、望ましくない態様と考えられるが、消去開始時において、本物の信号と合成信号との間の潤滑な移行を保証するために必要である。

【0108】

本発明は、上に概要を説明した特定の実施形態に関連して説明されたが、多くの代替案、変更案、及び変形案が、当業者には明らかであろう。したがって、上述の本発明の好適な実施形態は、例として示すものであり、制限するものではない。後述の請求項に定義される本発明の精神及び範囲から逸脱することなく、様々な変更を行ってもよい。

【0109】

本発明は、以下の図面を参照して詳述され、これらの図面において同様の符号は同様の要素を参照する。

【図面の簡単な説明】

【図1】 音声送信システムを例示する図である。

【図2】 G.711コーダ及びFECモジュールを備える音声送信システムを例示する図である。

【図3】 FEC技法を使った、出力音声信号を示す図である。

【図4】 消去の終了時における、オーバーラップ加算(overlap-add)(OLA)作業を示す図である。

【図5】 G.711コーダを使いFECを実行する処理を例示するフローチャートである。

【図6】 履歴バッファの更新処理を示すグラフである。

【図7】 信号の第一フレームの隠蔽処理を例示するフローチャートである。

【図8】 自動相関からのピッチの推定を示す図である。

【図9】 細かいピッチ推定と粗いピッチ推定との対比を示す図である。

【図10】 ピッチ及び最終クオーターバッファ(lastquarter buffer)内の信号を示す図である。

【図11】 単一周期ピッチバッファを使った、合成信号の生成を示す図である。

【図12】 信号の第二の又は後続の消去されたフレームを隠蔽する処理を例示するフローチャートである。

【図13】 第二の消去されたフレームへと続く合成信号を示す図である。

10

20

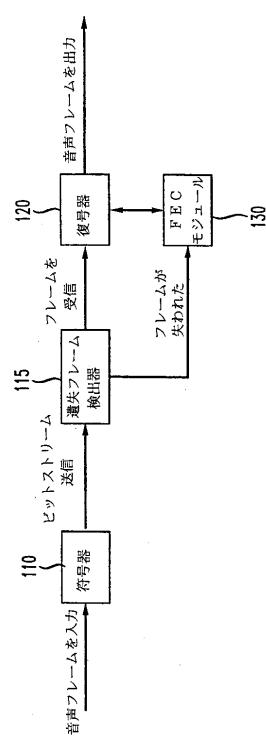
30

40

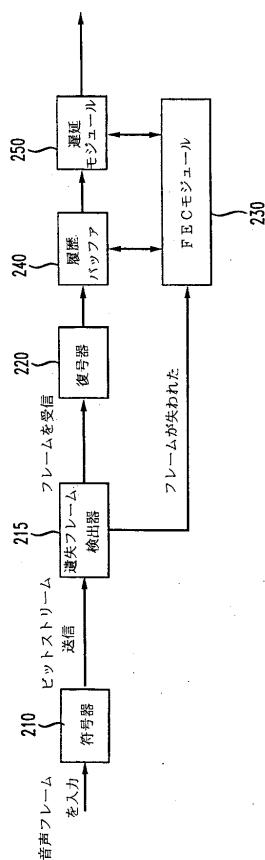
50

- 【図14】二周期ピッチバッファを使った合成信号の生成を示す図である。
 - 【図15】第二の消去されたフレームの開始時におけるOLAを示す図である。
 - 【図16】消去後の第一フレームを処理する方法を例示するフローチャートである。
 - 【図17】三周期ピッチバッファを使った、合成信号の生成を示す図である。
 - 【図18】FEC技法を他のスピーチコーダで使用する例を示すブロック図である。

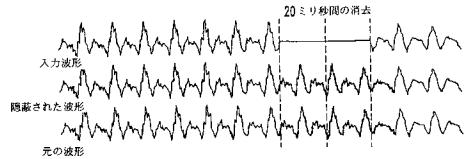
【図1】



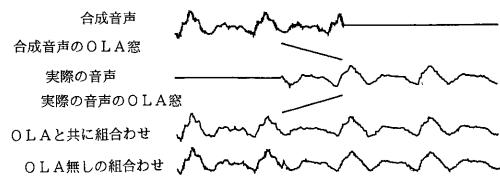
【 図 2 】



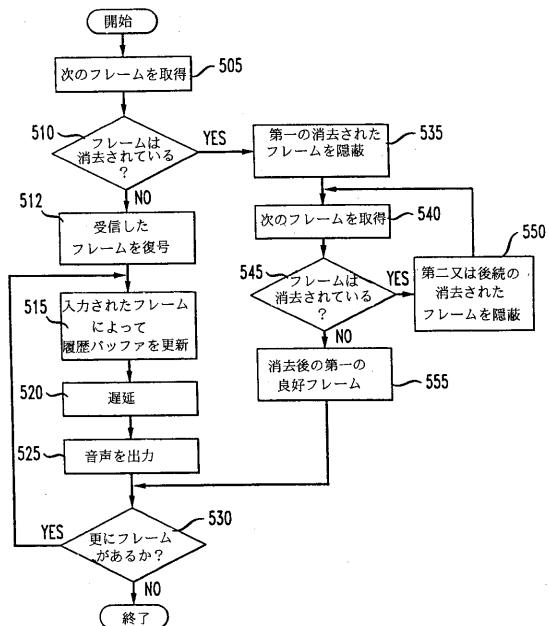
【図3】



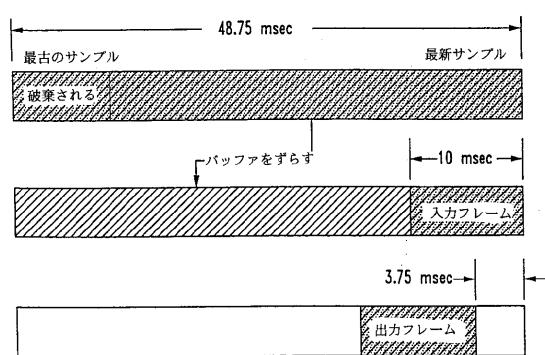
【図4】



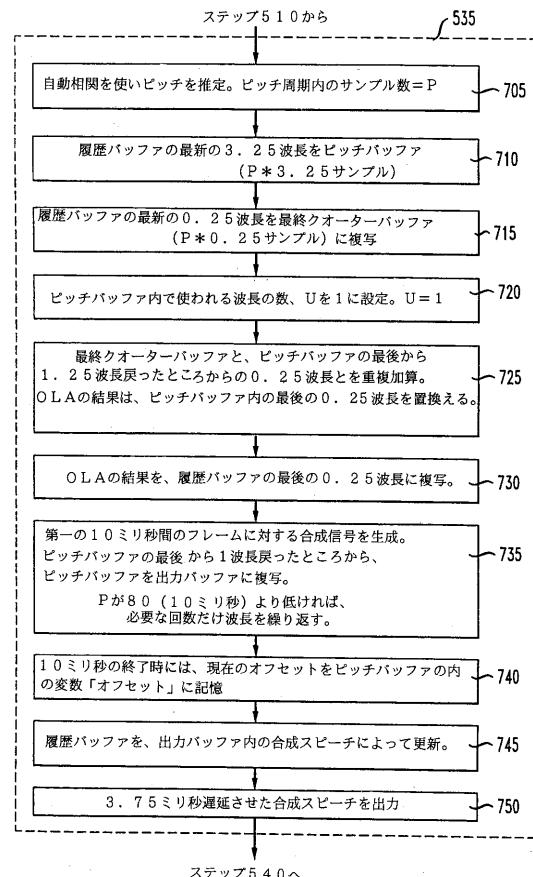
【図5】



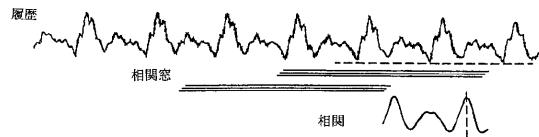
【図6】



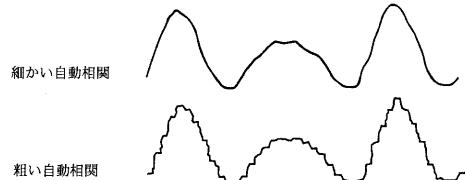
【図7】



【図 8】



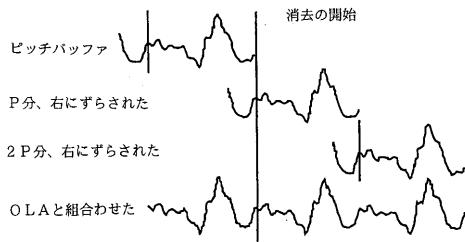
【図 9】



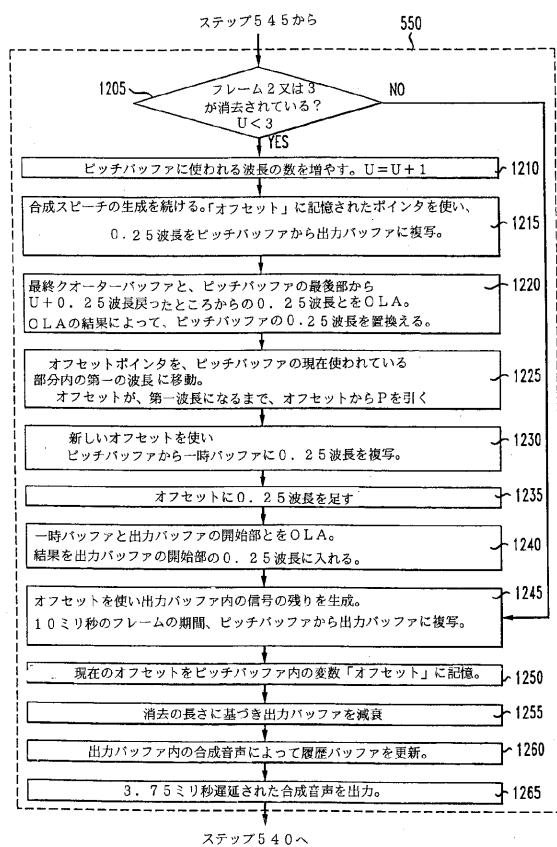
【図 10】



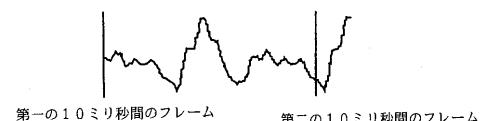
【図 11】



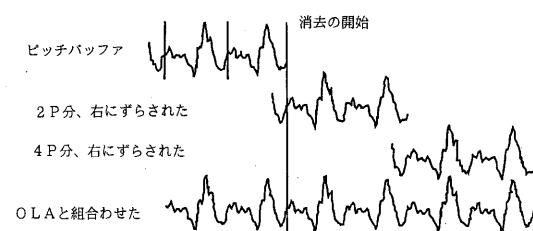
【図 12】



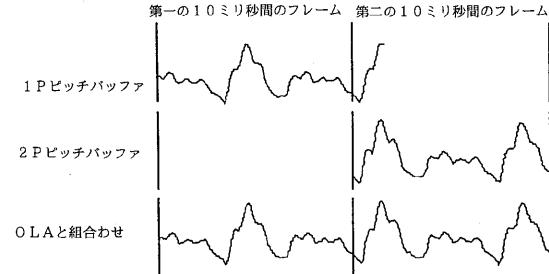
【図 13】



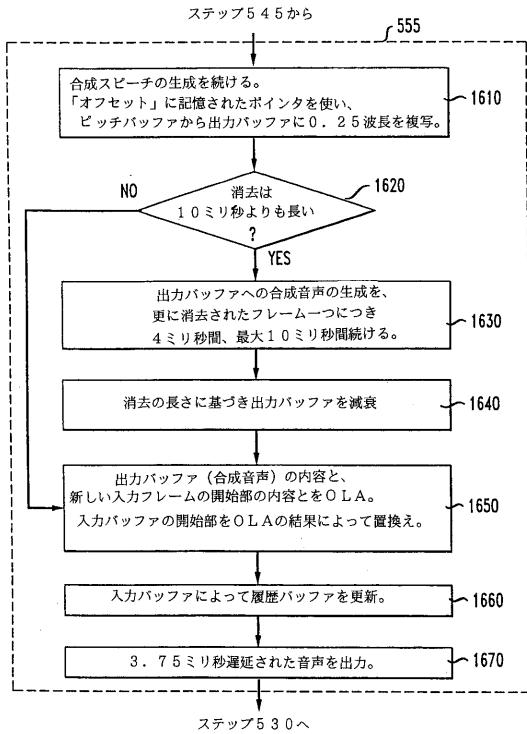
【図 14】



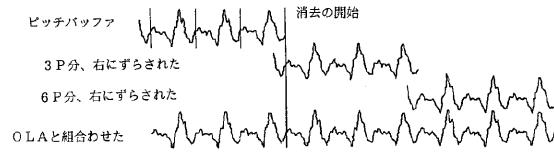
【図 15】



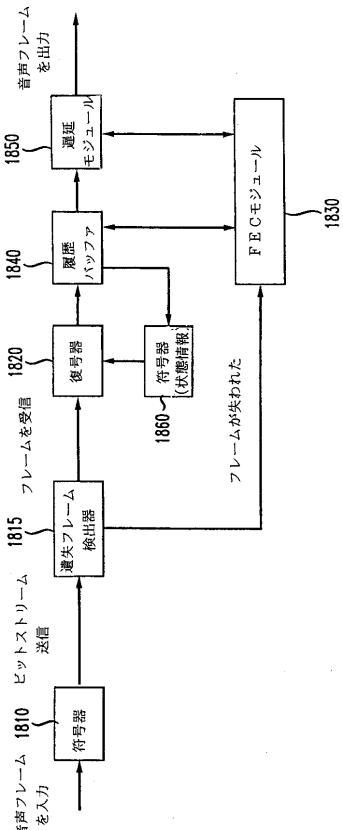
【図16】



【図17】



【図18】



フロントページの続き

審査官 井上 健一

(56)参考文献 特開平07-334191 (JP, A)
特開平10-282995 (JP, A)
特開平07-221714 (JP, A)
特開平08-305398 (JP, A)
特開昭55-049042 (JP, A)
特開平10-022936 (JP, A)
特表平08-500235 (JP, A)
特開平07-325594 (JP, A)
特表2002-530950 (JP, A)
特開平07-271391 (JP, A)
特開平10-069298 (JP, A)
国際公開第99/066494 (WO, A1)
国際公開第98/013941 (WO, A1)

(58)調査した分野(Int.Cl., DB名)

G10L 19/00