(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2004-526173
(P2004-526173A)
(43)公表日　平成16年8月26日(2004.8.26)

(51)Int.Cl.[7]	F I		テーマコード（参考）
G10L 19/04	G10L 9/14	J	5J064
G10L 13/00	G11B 20/18	512C	5K014
G10L 19/00	G11B 20/18	560A	
G11B 20/18	G11B 20/18	574D	
H03M 7/36	H03M 7/36		

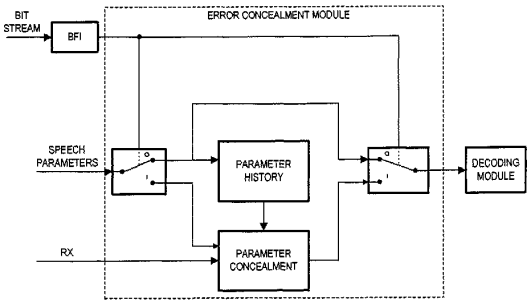
審査請求 有　予備審査請求 有　（全 67 頁）　最終頁に続く

(21)出願番号　特願2002-540142(P2002-540142)
(86)(22)出願日　平成13年10月29日(2001.10.29)
(85)翻訳文提出日　平成15年4月28日(2003.4.28)
(86)国際出願番号　PCT/IB2001/002021
(87)国際公開番号　WO2002/037475
(87)国際公開日　平成14年5月10日(2002.5.10)
(31)優先権主張番号　09/702,540
(32)優先日　平成12年10月31日(2000.10.31)
(33)優先権主張国　米国(US)

(71)出願人　399040520
ノキア　コーポレーション
フィンランド共和国、０２１５０　エスポー、ケイララハデンチエ　４
(74)代理人　100065226
弁理士　朝日奈　宗太
(74)代理人　100098257
弁理士　佐木　啓二
(72)発明者　メキネン、ヤリ
フィンランド共和国、フィン－３３１００　タムペレ、タッメラン　プイストカツ　３０－３２　セー５２
(72)発明者　ミッコラ、ハッヌ　イー
フィンランド共和国、フィン－３３３００　タムペレ、イッピセンカツ　１５
最終頁に続く

(54)【発明の名称】音声復号における音声フレームのエラー隠蔽のための方法およびシステム

(57)【要約】
デコーダにおいて受信される符号化されたビットストリームの部分としての音声シーケンスにおける１または２以上の不良フレームのエラーを隠蔽するための方法およびシステム。音声シーケンスが有声である場合、不良フレームのＬＴＰパラメータが最終のフレームの対応するパラメータに置き換えられる。音声フレームが無声である場合、不良フレームのＬＴＰパラメータが適応的に制限されるランダム項とともにＬＴＰヒストリーにもとづいて計算された値に置き換えられる。



【特許請求の範囲】
【請求項１】
音声デコーダに受信された音声信号を示す符号化されたビットストリームにおけるエラーを隠蔽するための方法であって、該符号化されたビットストリームが、音声シーケンスにより構成された複数の音声フレームを含み、該音声フレームが１または２以上の非劣化フレームによって先行される少なくとも１つの部分的に劣化したフレームを含み、該部分的に劣化したフレームが第１の長期予測ラグ値と第１の長期予測利得値とを含み、前記非劣化フレームが第２の長期予測ラグ値と第２の長期予測利得値とを含み、該第２の長期予測ラグ値が最終の長期予測ラグ値を含み、該第２の長期予測利得値が最終の長期予測利得値を含み、
前記方法が、
前記第２の長期予測ラグ値に基づいて上限と下限とを与える工程と、
前記第１の長期予測ラグ値が、前記上限および下限の範囲内または前記上限および下限の範囲の外側にあるかどうかを決定する工程と、
前記第１の長期予測ラグ値が前記上限および下限の範囲の外側にある場合、前記部分的に劣化したフレームにおける前記第１の長期予測ラグ値を第３のラグ値と交換する工程と、
前記第１の長期予測ラグ値が前記上限および下限の範囲内にある場合、前記部分的に劣化したフレームにおける前記第１の長期予測ラグ値を保持する工程
とを含む方法。
【請求項２】
前記第１の長期ラグ値が前記上限および下限の範囲の外側にある場合、前記部分的に劣化したフレームにおける前記第１の長期予測利得値を第３の利得値と交換する工程をさらに含む請求項１記載の方法。
【請求項３】
前記第３のラグ値が、前記第２の長期予測ラグ値および前記第２の長期予測ラグ値に基づいて決定されたさらなる限界に拘束される適応的に制限されたランダムラグジッタにもとづいて計算される請求項１記載の方法。
【請求項４】
前記第３のラグ値が、前記第２の長期予測利得値および前記第２の長期予測利得値にもとづいて決定された限界に拘束される適応的に制限されたランダム利得ジッタにもとづいて計算される請求項２記載の方法。
【請求項５】
音声デコーダに受信された音声信号を示す符号化されたビットストリームにおけるエラーを隠蔽するための方法であって、該符号化されたビットストリームが音声シーケンスにおいて構成された複数の音声フレームを含み、該音声フレームが１または２以上の非劣化フレームによって先行される少なくとも１つの劣化したフレームを含み、該劣化したフレームが第１の長期予測ラグ値と第１の長期予測利得値とを含み、前記非劣化フレームが第２の長期予測ラグ値と第２の長期予測利得値とを含み、該第２の長期予測ラグ値が最終の長期予測ラグ値を含み、該第２の長期予測利得値が最終の長期予測利得値を含み、前記音声シーケンスが定常的音声シーケンスと非定常的音声シーケンスとを含み、前記劣化したフレームが全体的に劣化したフレームか、または部分的に劣化したフレームであり得て、
前記方法が、
前記劣化したフレームが、部分的に劣化したのか、または全体的に劣化したのかを決定する工程と、
前記劣化したフレームが全体的に劣化している場合、当該劣化したフレームにおける第１の長期予測ラグ値を第３のラグ値と交換する工程と、
前記劣化したフレームが部分的に劣化している場合、当該劣化したフレームにおける第１の長期予測ラグ値を第４のラグ値と交換する工程
とを含んでなる方法。
【請求項６】

前記部分的に劣化したフレームが構成される音声シーケンスが定常的か非定常的かを判断する工程と、
前記音声シーケンスが定常的である場合、前記第４のラグ値を前記最終の長期予測ラグ値に等しく設定する工程と、
前記音声シーケンスが非定常的である場合、前記劣化したフレームに先立つ非劣化フレームに関係する適合コードブックから検索される復号された長期予測ラグ値にもとづいて前記第４のラグ値を設定する工程と
をさらに含む請求項５記載の方法。
【請求項７】
完全に劣化したフレームにおいて構成された音声シーケンスが定常的か、非定常的かを判断する工程と、
前記音声シーケンスが定常的である場合、前記第３のラグ値を前記最終の長期予測ラグ値に等しく設定する工程と、
前記音声シーケンスが非定常的である場合、前記第２の長期予測値および適応的に制限されるランダムラグジッタにもとづいて第３のラグ値を決定する工程と
をさらに含む請求項５記載の方法。
【請求項８】
前記第２の長期予測ラグ値が、最終から２番目の長期予測ラグ値と最終から３番目の長期予測予測ラグ値とを含み、前記第２の長期予測利得値が、最終から２番目の長期予測利得値と最終から３番目の長期予測利得値とをさらに含み、
前記方法が、
前記第２の長期予測ラグ値の中で最小の値であるｍｉｎＬａｇを決定する工程と、
前記第２の長期予測ラグ値の中で最大の値であるｍａｘＬａｇを決定する工程と、
前記第２の長期予測ラグ値の平均であるｍｅａｎＬａｇを決定する工程と、
ｍａｘＬａｇとｍｉｎＬａｇとの差であるｄｉｆＬａｇを決定する工程と、
前記第２の長期予測利得値の中で最小の値であるｍｉｎＧａｉｎを決定する工程と、
前記第２の長期予測利得値の中で最大の値であるｍａｘＧａｉｎを決定する工程と、
前記第２の長期予測利得値の平均であるｍｅａｎＧａｉｎを決定する工程
とをさらに含み、
ｄｉｆＬａｇ＜１０であり、かつ（ｍｉｎＬａｇ－５）＜第４のラグ値＜（ｍａｘＬａｇ＋５）である場合、または
前記最終の長期予測利得値が０．５より大きく、前記最終から２番目の長期予測利得値が０．５より大きく、前記第４のラグ値が前記最終の長期予測値と１０との和より小さく、当該第４のラグ値と１０との和が前記最終の長期予測値より大きい場合、または
ｍｉｎＧａｉｎ＜０．４であり、かつ前記長期予測利得値がｍｉｎＧａｉｎに等しく、前記第４のラグ値がｍｉｎＬａｇより大きくｍａｘＬａｇより小さい場合、または
ｄｉｆＬａｇ＜７０であり、かつ第４のラグ値がｍｉｎＬａｇより大きくｍａｘＬａｇより小さい場合、または
前記第４のラグ値がｍｅａｎＬａｇより大きくｍａｘＬａｇより小さい場合、
前記劣化したフレームが部分的に劣化していると決定される
請求項６記載の方法。
【請求項９】
前記音声シーケンスが非定常的であり、前記方法が、音声フレームのフレーム誤り率を決定する工程をさらに含み、
該フレーム誤り率が決められた値に達すると、前記第４のラグ値が前記復号された長期予測ラグ値に基づいてきめられ、かつ
該フレーム誤り率が決められた値より小さい場合、前記第４のラグ値が前記最終の長期予測ラグ値に等しく設定されてなる
請求項６記載の方法。
【請求項１０】

前記定常的音声シーケンスが有声シーケンスを含み、前記非定常的音声シーケンスが無声シーケンスを含む請求項５記載の方法。
【請求項１１】
音声信号を符号化されたビットストリームに符号化し、該符号化されたビットストリームを合成された音声に復号するための音声信号の送信および受信システムであって、前記符号化されたビットストリームが、音声シーケンスで構成された複数の音声フレームを含み、該音声フレームが１または２以上の非劣化フレームに先行される少なくとも１つの劣化したフレームを含み、該劣化したフレームが第１の長期予測ラグ値と第１の長期予測利得値とを含み、前記非劣化フレームが第２の長期予測ラグ値と第２の長期予測利得値とを含み、該第２の長期予測ラグ値が最終の長期予測ラグ値を含み、前記第２の長期予測利得値が最終の長期予測利得値を含み、前記音声シーケンスが、定常的音声シーケンスおよび非定常的音声シーケンスを含み、前記劣化したフレームを示すために、第１の信号が用いられ、
前記システムが、
該第１の信号に応答して、前記劣化したフレームが構成されている音声シーケンスが定常的または非定常的であるかの決定と、当該決定を表示する第２の信号の提供とを行なうための第１の手段と、
該第２の信号に応答して、前記音声シーケンスが定常的である場合、前記劣化したフレーム中の前記第１の長期予測ラグ値を前記最終の長期予測ラグ値と交換し、前記音声シーケンスが非定常的である場合、前記劣化したフレーム中の第１の長期予測ラグ値を第３のラグ値と交換するための第２の手段
とを備えたシステム。
【請求項１２】
前記第３のラグ値が、前記第２の長期予測ラグ値および適応的に制限されるランダムラグジッタにもとづいて決定される請求項１１記載のシステム。
【請求項１３】
前記音声シーケンスが非定常的である場合、前記第２の手段が、さらに劣化したフレームにおける第１の長期予測利得値を第３の利得値と交換する請求項１１記載のシステム。
【請求項１４】
前記第３の利得値が、前記第２の長期予測利得値および適応的に制限されるランダム利得ジッタにもとづいて決定される請求項１３記載のシステム。
【請求項１５】
前記定常的音声シーケンスが有声シーケンスを含み、前記非定常的音声シーケンスが無声シーケンスを含む請求項１１記載のシステム。
【請求項１６】
符号化されたビットストリームから音声を合成するためのデコーダであって、前記符号化されたビットストリームが、音声シーケンスで構成された複数の音声フレームを含み、該音声フレームが１または２以上の非劣化フレームに先行される少なくとも１つの劣化したフレームを含み、該劣化したフレームが第１の長期予測ラグ値と第１の長期予測利得値とを含み、前記非劣化したフレームが第２の長期予測ラグ値と第２の長期予測利得値とを含み、該第２の長期予測ラグ値が最終の長期予測ラグ値を含み、前記第２の長期予測利得値が最終の長期予測利得値を含み、前記音声シーケンスが、定常的音声シーケンスおよび非定常的音声シーケンスを含み、前記劣化したフレームを示すために、第１の信号が用いられ、
前記デコーダが、
該第１の信号に応答して、前記劣化したフレームが構成されている音声シーケンスが定常的かまたは非定常的であるかの決定と、当該決定を表示する第２の信号の提供とを行なうための第１の手段と、
該第２の信号に応答して、前記音声シーケンスが定常的である場合、前記劣化したフレーム中の前記第１の長期予測ラグ値を前記最終の長期予測ラグ値と交換し、前記音声シーケ

ンスが非定常的である場合、前記劣化したフレーム中の第１の長期予測ラグ値を第３のラグ値と交換するための第２の手段
とを備えたデコーダ。
【請求項１７】
前記ラグ値が前記第２長期予測ラグ値および適応的に制限されたランダムラグジッタにもとづいて決定される請求項１６記載のデコーダ。
【請求項１８】
前記第２の手段が、前記音声シーケンスが非定常的である場合、さらに劣化したフレームにおける前記第１の長期利得値を第３の利得値と交換する請求項１６記載のデコーダ。
【請求項１９】
前記第３の利得値が、前記第２の長期予測利得値および適応的に制限されるランダム利得ジッタにもとづいて決定される請求項１８記載のデコーダ。
【請求項２０】
前記定常的音声シーケンスが有声シーケンスを含み、前記非定常的音声シーケンスが無声シーケンスを含む請求項１６記載のデコーダ。
【請求項２１】
音声信号を示す音声データを含む符号化されたビットストリームを受信するように構成された移動局であって、前記符号化されたビットストリームが、音声シーケンスで構成された複数の音声フレームを含み、該音声フレームが１または２以上の非劣化フレームに先行される少なくとも１つの劣化したフレームを含み、該劣化したフレームが第１の長期予測ラグ値と第１の長期予測利得値とを含み、前記非劣化フレームが第２の長期予測ラグ値と第２の長期予測利得値とを含み、該第２の長期予測ラグ値が最終の長期予測ラグ値を含み、前記第２の長期予測利得値が最終の長期予測利得値を含み、前記音声シーケンスが、定常的音声シーケンスおよび非定常的音声シーケンスを含み、前記劣化したフレームを示すために、第１の信号が用いられ、
前記移動局が、
該第１の信号に応答して、前記劣化したフレームが構成されている音声シーケンスが定常的または非定常的であるかの決定と、当該決定を表示する第２の信号の提供とを行なうための第１の手段と、
該第２の信号に応答して、前記音声シーケンスが定常的である場合、前記劣化したフレーム中の前記第１の長期予測ラグ値を前記最終の長期予測ラグ値と交換し、前記音声シーケンスが非定常的である場合、前記劣化したフレーム中の第１の長期予測ラグ値を第３のラグ値と交換するための第２の手段
とを備えた移動局。
【請求項２２】
前記第３のラグ値が、前記第２の長期予測ラグ値および適応的に制限されたランダムラグジッタにもとづいて決定される請求項２１記載の移動局。
【請求項２３】
前記音声シーケンスが非定常的である場合、前記第２の手段が、劣化したフレームにおける第１の長期利得値を第３の利得値と交換する請求項２１記載の移動局。
【請求項２４】
前記第３の利得値が、前記第２の長期予測利得値および適応的に制限されたランダム利得ジッタにもとづいて決定される請求項２３記載の移動局。
【請求項２５】
前記定常的音声シーケンスが有声シーケンスを含み、非定常的音声シーケンスが無声シーケンスを含む請求項２１記載の移動局。
【請求項２６】
移動局から音声データを含む符号化されたビットストリームを受信するように構成された電気通信ネットワークにおける要素であって、前記音声データが、音声シーケンスで構成された複数の音声フレームを含み、該音声フレームが１または２以上の非劣化フレームに

先行される少なくとも１つの劣化したフレームを含み、該劣化したフレームが第１の長期予測ラグ値と第１の長期予測利得値とを含み、前記非劣化フレームが第２の長期予測ラグ値と第２の長期予測利得値とを含み、該第２の長期予測ラグ値が最終の長期予測ラグ値を含み、前記第２の長期予測利得値が最終の長期予測利得値を含み、前記音声シーケンスが、定常的音声シーケンスおよび非定常的音声シーケンスを含み、前記劣化したフレームを示すために、第１の信号が用いられ、
前記要素が、
該第１の信号に応答して、前記劣化したフレームが構成されている音声シーケンスが定常的または非定常的であるかの決定と、当該決定を表示する第２の信号の提供とを行なうための第１の手段と、
該第２の信号に応答して、前記音声シーケンスが定常的である場合、前記劣化したフレーム中の前記第１の長期予測ラグ値を前記最終の長期予測ラグ値と交換し、前記音声シーケンスが非定常的である場合、前記劣化したフレーム中の第１の長期予測ラグ値を第３のラグ値と交換するための第２の手段
とを備えた要素。
【請求項２７】
前記第３の長期予測ラグ値が、前記第２の長期予測ラグ値および適応的に制限されたランダムラグジッタにもとづいて決定される要素。
【請求項２８】
前記音声シーケンスが非定常的である場合、前記第３の手段がさらに前記第１の長期予測利得値を第３の利得値と交換する請求項２６記載の要素。
【請求項２９】
前記第３の利得値が、前記第２の長期予測利得値および適応的に制限されるランダム利得ジッタにもとづいて決定される請求項２８記載の要素。
【請求項３０】
前記定常的音声シーケンスが有声シーケンスを含み、非定常的音声シーケンスが無声シーケンスを含む請求項２６記載の要素。
【請求項３１】
前記第２の長期予測利得値が最終から２番目の長期予測利得値をさらに含み、かつ
ｄｉｆＬａｇ＜１０であり、かつ（ｍｉｎＬａｇ－５）＜ｄｅｃｏｄｅｄＬａｇ＜（ｍａｘＬａｇ＋５）である場合、または
ｌａｓｔＧａｉｎ＞０．５であり、かつｓｅｃｏｎｄｌｓａｓｔＧａｉｎ＞０．５であり、かつ（ｌａｓｔＬａｇ－１０）＜ｄｅｃｏｄｅｄＬａｇ＜（ｌａｓｔＬａｇ＋１０）である場合、または
ｍｉｎＧａｉｎ＜０．４であり、かつｌａｓｔＧａｉｎ＞０．５であり、ｍｉｎＬａｇ＜ｄｅｃｏｄｅｄＬａｇ＜ｍａｘＬａｇである場合、または
ｄｉｆＬａｇ＜７０であり、かつｍｉｎＬａｇ＜ｄｅｃｏｄｅｄＬａｇ＜ｍａｘＬａｇである場合、または
ｍｅａｎＬａｇ＜ｄｅｃｏｄｅｄＬａｇ＜ｍａｘＬａｇである場合、
第４の値がｄｅｃｏｄｅｄＬａｇに等しく設定され、
ｍｉｎＬａｇが前記第２の長期予測ラグ値の中でもっとも小さいラグ値であり、
ｍａｘＬａｇが前記第２の長期予測ラグ値の中でもっとも大きいラグ値であり、
ｍｅａｎＬａｇが前記第２の長期予測ラグ値の平均であり、
ｄｉｆＬａｇがｍａｘＬａｇとｍｉｎＬａｇとの差であり、
ｍｉｎＧａｉｎが前記第２の長期予測利得値の中でもっとも小さい利得値であり、
ｍｅａｎＧａｉｎが前記第２の長期予測利得値の平均であり、
ｌａｓｔＧａｉｎが前記最終の長期予測利得値であり、
ｌａｓｔＬａｇが前記最終の長期予測ラグ値であり、
ｓｅｃｏｎｄｌａｓｔＧａｉｎが前記最終から２番目の長期予測ラグ値であり、かつ
ｄｅｃｏｄｅｄＬａｇが復号された長期予測ラグであり、該復号された長期予測ラグが、

劣化したフレームに先行する非劣化フレームに関連する適応するコードブックから検索される請求項５記載の方法。
【請求項３２】
前記第１の長期予測利得値
がＵｐｄａｔｅｄ＿ｇａｉｎと交換され、
ｇａｉｎＤｉｆ＞０．５ ＡＮＤ ｌａｓｔＧａｉｎ＝ｍａｘＧａｉｎ＞０．９ ＡＮＤ ｓｕｂＢＦ＝１であれば、
Ｕｐｄａｔｅｄ＿ｇａｉｎ＝（ｓｅｃｏｎｄＬａｓｔＧａｉｎ＋ｔｈｉｒｄＬａｓｔＧａｉｎ）／２であり、
ｇａｉｎＤｉｆ＞０．５ ＡＮＤ ｌａｓｔＧａｉｎ＝ｍａｘＧａｉｎ＞０．９ ＡＮＤ ｓｕｂＢＦ＝２であれば、
Ｕｐｄａｔｅｄ＿ｇａｉｎ＝ｍｅａｎＧａｉｎ＋ｒａｎｄＶａｒ＊（ｍａｘＧａｉｎ－ｍｅａｎＧａｉｎ）であり、
ｇａｉｎＤｉｆ＞０．５ ＡＮＤ ｌａｓｔＧａｉｎ＝ｍａｘＧａｉｎ＞０．９ ＡＮＤ ｓｕｂＢＦ＝３であれば、
Ｕｐｄａｔｅｄ＿ｇａｉｎ＝ｍｅａｎＧａｉｎ－ ｒａｎｄＶａｒ＊（ｍｅａｎＧａｉｎ－ｍｉｎＧａｉｎ）であり、
ｇａｉｎＤｉｆ＞０．５ ＡＮＤ ｌａｓｔＧａｉｎ＝ｍａｘＧａｉｎ＞０．９ ＡＮＤ ｓｕｂＢＦ＝４であれば、
Ｕｐｄａｔｅｄ＿ｇａｉｎ＝ｍｅａｎＧａｉｎ＋ｒａｎｄＶａｒ＊（ｍａｘＧａｉｎ－ｍｅａｎＧａｉｎ）である。
Ｕｐｄａｔｅｄ＿ｇａｉｎが、ｌａｓｔＧａｉｎと等しいか、またはｌａｓｔＧａｉｎより小である場合、
または、
ｇａｉｎＤｉｆ＞０．５であれば、Ｕｐｄａｔｅｄ＿ｇａｉｎ＝ｌａｓｔＧａｉｎであり、
（８）ｇａｉｎＤｉｆ＜０．５ ＡＮＤ ｌａｓｔＧａｉｎ＝ｍａｘＧａｉｎであれば、Ｕｐｄａｔｅｄ＿ｇａｉｎ＝ｍｅａｎＧａｉｎであり、
（９）ｇａｉｎＤｉｆ＜０．５であれば、Ｕｐｄａｔｅｄ＿ｇａｉｎ＝ｌａｓｔＧａｉｎであり、
そのときＵｐｄａｔｅｄ＿ｇａｉｎはｌａｓｔＧａｉｎより大きく、
ｒａｎｄＶａｒは、０と１とのあいだの乱数であり、
ｇａｉｎＤｉｆは、もっとも大きい長期予測利得値ともっとも小さい長期予測利得値との差であり、
ｌａｓｔＧａｉｎは、最終の長期予測利得値であり、
ｓｅｃｏｎｄＬａｓｔＧａｉｎは、最終から２番目の長期予測利得値であり、
ｔｈｉｒｄＬａｓｔＧａｉｎは、最終から３番目の長期予測利得値であり、かつ
ｓｕｂＢＦは、サブフレームの次数である請求項８記載の方法。
【発明の詳細な説明】
【０００１】
［発明の分野］
本発明は、概して符号化されたビット・ストリームからの音声信号の復号に関し、より特定的には、音声の復号中に音声フレームにおいてエラーが検出された場合の劣化した音声パラメータの隠蔽に関する。
【０００２】
［発明の背景］
音声および音響の符号化アルゴリズム（ｃｏｄｉｎｇ ａｌｇｏｒｉｔｈｍ）は、通信、マルチメディアおよび記憶のシステムにおいて広範なアプリケーションを有している。符号化アルゴリズムの開発は、合成された信号の高い品質を維持しつつ送信および記憶容量を節約する必要に迫られている。コーダの複雑さは、たとえばアプリケーション・プラッ

トフォーム（ａｐｐｌｉｃａｔｉｏｎ ｐｌａｔｆｏｒｍ）の処理パワーによって制限される。たとえば音声記憶のようなあるアプリケーションでは、符号器はきわめて複雑でよいが、復号器（デコーダ）はできるだけ単純でなければならない。

【０００３】

近頃の音声コーデック（ｃｏｄｅｃ）は、音声信号をフレームと呼ばれる短いセグメントで処理して動作する。音声コーデックの典型的なフレーム長は２０ｍｓであり、これは、サンプリング周波数を８ｋＨｚと仮定した場合、１６０個の音声サンプルに相当する。広帯域コーデックでは、この２０ｍｓの典型的なフレーム長は、サンプリング周波数１６ｋＨｚを仮定すると３２０個の音声サンプルに相当する。フレームは、さらに多数のサブフレームに分割されてもよい。符号器（エンコーダ）は、全てのフレームについて入力信号のパラメータ表示を決定する。パラメータは量子化され、通信チャネルを介してデジタル形式で送信される（または、記憶媒体に記憶される）。デコーダは図１に示されるように、受信されたパラメータに基づいて合成された音声信号を生成する。

【０００４】

抽出される符号化パラメータの典型的なセットは、信号の短期予測に使用されるスペクトルパラメータ（線形予測符号化（ＬＰＣ）パラメータ等）、信号の長期予測（ＬＴＰ）に使用されるパラメータ、様々な利得パラメータおよび励振パラメータを含んでいる。ＬＴＰパラメータは、音声信号の基本周波数に密接に関連している。このパラメータは、しばしばいわゆるピッチラグ（ｐｉｔｃｈ－ｌａｇ）パラメータとして知られ、音声サンプルについての本的周期性を記述している。また、利得パラメータの１つはこの基本的周期性に高度に関連づけられていて、ＬＴＰ利得と呼ばれる。ＬＴＰ利得は、音声をできるだけ自然なものにする上できわめて重要なパラメータである。前記の符号化パラメータに関する記載は、おおまかには、かねてより最も成功している音声コーデックであるいわゆるコード励振線形予測（ＣＥＬＰ）コーデックを含む様々な音声コーデックに当てはまる。

【０００５】

音声パラメータは、通信チャネルを介してデジタル形式で送信される。通信チャネルの条件はときおり変化し、これがビット・ストリームのエラーの原因となる場合がある。これはフレーム・エラー（ｂａｄ ｆｒａｍｅ ：不良フレーム）を引き起こす。即ち、特定の音声セグメント（典型的には２０ｍｓ）を記述するパラメータの幾つかが劣化される。フレーム・エラーには、全体的に劣化したフレーム（ｔｏｔａｌｌｙ ｃｏｒｒｕｐｔｅｄ ｆｒａｍｅ）と部分的に劣化したフレーム（ｐａｒｔｉａｌｌｙ ｃｏｒｒｕｐｔｅｄ ｆｒａｍｅ）の２種類がある。これらのフレームは、デコーダで全く受信されない場合もある。パケットベースの送信システムでは、通常のインターネット接続のように、データパケットが全く受信機に到達しない、または該データパケットの到達が遅過ぎて、話し言葉の同時性のゆえに、データパケットが使用され得ないような状況が発生する可能性もある。部分的に劣化したフレームは、受信機に到達し、しかもエラーでないパラメータを幾つか含む可能性のあるフレームである。これは、通常、既存のＧＳＭ接続の場合のような回路切替接続（ｃｉｒｃｕｉｔ ｓｗｉｔｃｈｉｎｇ ｃｏｎｎｅｃｔｉｏｎ）における状況である。部分的に劣化したフレームにおけるビット・エラー率（ＢＥＲ）は、典型的には約０．５～５％である。

【０００６】

前記の説明から、不良フレームまたは劣化したフレームという２つのケースは、音声パラメータの損失に起因する再構成された音声の劣化（ｄｅｇｒａｄａｔｉｏｎ）に対応する際に異なるアプローチを必要とすることが分かる。

【０００７】

失われた、もしくはエラーのある音声フレームは、ビット・ストリームのエラーの原因となる通信チャネルの悪条件の結果である。受信された音声フレームにエラーが検出されると、エラー修正手順が開始される。エラー修正手順は通常、代替手順とミューティング手順とを含んでいる。従来技術では、不良フレームの音声パラメータが先行する優良な（ｇｏｏｄ）フレームからの減衰された、または変更された値に交換される。しかしながら、

劣化したフレームにおけるいくつかのパラメータ（ＣＥＬＰにおける励振パラメータ等）には、依然として復号化に使用することができるものがある。

【０００８】

図２は、従来技術による方法の原理を示している。図２に示されるように、「パラメータヒストリー」と標識されたバッファは、最終の優良フレーム（ｇｏｏｄ　ｆｒａｍｅ）の音声パラメータを格納するために使用される。不良フレームが検出されると、不良フレームインジケータ（ＢＦＩ）が１に設定され、エラー隠蔽手順が開始される。ＢＦＩが設定されなければ（ＢＦＩ＝０）、パラメータヒストリーは更新され、音声パラメータはエラー隠蔽なしで復号化に使用される。従来技術システムでは、エラー隠蔽手順は、劣化したフレームにおける失われた、もしくはエラーのあるパラメータを隠蔽するためにパラメータヒストリー（履歴）を使用する。受信されたフレームからの音声パラメータの中には、そのフレームが不良フレーム（ＢＦＩ＝１）として分類されていても、使用することができるものがある。たとえば、ＧＳＭ適応型マルチレート（ＡＭＲ）音声コーデック（ＥＴＳＩ仕様０６．９１）では、必ずそのチャネルからの励振ベクトルが使用される。（たとえば、幾つかのＩＰベースの送信システムにおいて）音声フレームが全体的に損失したフレームであるときは、受信された不良フレームからのパラメータは全く使用されない。場合によっては、フレームが全く受信されない、もしくはフレームの到着が遅すぎて失われたフレームとして分類されざるを得ないこともある。

【０００９】

ある先行技術システムでは、ＬＴＰラグ隠蔽は僅かに変更された分数部を有する最終の優良ＬＴＰラグ値を使用し、スペクトルパラメータは定数平均に向かい僅かにシフトされた最終の優良パラメータに交換される。利得（ＬＴＰおよび固定コードブック）は通常、減衰された最終の優良値に、または最終の幾つかの優良値の中央値（ｍｅｄｉａｎ）に交換される。全てのサブフレームに対して、同じ置換された音声パラメータが使用されるが、パラメータのいくつかには僅かな変更が加えられる。

【００１０】

従来技術によるＬＴＰ隠蔽は、定常的な音声信号、たとえば有声音声または定常的音声に関しては十分であると言える。しかしながら非定常的な音声信号に関しては、従来技術の方法では不快かつ可聴性のアーチファクト（ａｒｔｉｆａｃｔ）を引き起こすかも知れない。たとえば、音声信号が無声または非定常的である場合には、不良フレーム内のラグ値を単純に最終の優良ラグ値に置換すると、無声音声バーストの中央に短い有声音声セグメントが発生するという効果が出る（図１０参照）。「ビング（ｂｉｎｇ）」アーチファクトとして周知のこの効果は、煩わしいものになり得る。

【００１１】

音声の復号において、音声品質を向上させるためエラーを隠蔽する方法およびシステムを提供することが有益でありかつ望ましい。

【００１２】

［発明の要旨］

本発明は、音声信号における長期予測（ＬＴＰ）パラメータ間に認識できる関係性が存在するという事実を利用するものである。特にＬＴＰラグは、ＬＴＰ利得とのあいだに強い相関性を有している。ＬＴＰ利得が高くかつ十分に安定していれば、ＬＴＰラグは、典型的にはきわめて安定し、隣接するラグ値間の変動は小さい。その場合、音声パラメータは有声音声シーケンスを表わす。ＬＴＰ利得が低いか、または不安定であるとき、ＬＴＰラグは典型的には無声であり、音声パラメータは無声音声シーケンスを表す。いったん音声シーケンスが定常的（有声）または非定常的（無声）として分類されると、シーケンス内の劣化したフレームまたは不良フレームは異なる処理を施されることが可能である。

【００１３】

したがって、本発明の第１の態様は音声復号器（デコーダ）において受信された音声信号を示す符号化されたビット・ストリームにおけるエラーを隠蔽するための方法であって、該符号化されたビット・ストリームが音声シーケンスで構成された複数の音声フレームを

含み、該音声フレームが１または２以上の非劣化フレームによって先行される少なくとも１つの劣化したフレームを含み、該劣化したフレームが第１の長期予測ラグ値と第１の長期予測利得値とを含み、かつ該非劣化フレームが第２の長期予測ラグ値と第２の長期予測利得値とを含み、該第２の長期予測ラグ値は最終の長期予測ラグ値を含み、該第２の長期予測利得値は最終の長期予測利得値を含み、前記音声シーケンスは定常的および非定常的音声シーケンスを含み、前記劣化したフレームは部分的に劣化したか、または全体的に劣化したものであり得る。本方法は、

前記第１の長期予測ラグ値が、前記第２の長期予測ラグ値に基づいて決定された上限および下限の範囲内にあるか該範囲の外側にあるかを決定する工程と、

前記第１の長期予測ラグ値が該上限および下限の範囲の外側にある場合、前記部分的に劣化したフレームにおける前記第１の長期予測ラグ値を第３のラグ値に交換する工程と、

前記第１の長期予測ラグ値が該上限および下限の範囲内にある場合、前記部分的に劣化したフレームにおける前記第１の長期予測ラグ値を保持する工程

とを含んでいる。

【００１４】

あるいはこれに代えて、本方法は、

前記第２の長期予測利得値に基づいて、前記劣化したフレームが構成される前記音声シーケンスが定常的であるか非定常的であるかを判断する工程と、

前記音声シーケンスが定常的である場合、前記劣化したフレームにおける前記第１の長期予測ラグ値を前記最終の長期予測ラグ値に交換する工程と、

前記音声シーケンスが非定常的である場合、前記劣化したフレームにおける前記第１の長期予測ラグ値を、前記第２の長期予測ラグ値と適応的に制限された（ａｄａｐｔｉｖｅｌｙ－ｌｉｍｉｔｅｄ）ランダムラグジッタ（ｒａｎｄｏｍ　ｌａｇ　ｊｉｔｔｅｒ）とに基づいて決定された第３の長期予測ラグ値に交換し、前記劣化したフレームにおける前記第１の長期予測利得値を、前記第２の長期予測利得値と適応的に制限されたランダム利得ジッタ（ｒａｎｄｏｍ　ｇａｉｎ　ｊｉｔｔｅｒ）とに基づいて決定された第３の長期予測利得値に交換する工程とを含んでいる。

【００１５】

好適には、前記第３の長期予測ラグ値は、少なくとも部分的に前記第２の長期予測ラグ値の加重中央値に基づいて計算され、前記適応的に制限されたランダムラグジッタは、前記第２の長期予測ラグ値に基づいて決定された限定値に拘束された値である。

【００１６】

好適には、前記第３の長期予測利得値は、少なくとも部分的に前記第２の長期予測利得値の加重中央値に基づいて計算され、前記適応的に制限されたランダム利得ジッタは、前記第２の長期予測利得値に基づいて決定された限定値に拘束された値である。

【００１７】

あるいはこれに代えて、本方法は、

前記劣化したフレームが部分的に劣化しているか、全体的に劣化しているかを決定する工程と、

前記劣化フレームが全体的に劣化している場合、前記劣化したフレームにおける前記第１の長期予測ラグ値を第３のラグ値に交換する工程とを含み、前記全体的に劣化したフレームが構成されている音声シーケンスが定常的であるときは、前記第３のラグ値を前記最終の長期予測ラグ値に等しく設定し、前記音声シーケンスが非定常的である場合、前記第２の長期予測値と適応的に制限されたランダムラグジッタとに基づいて前記第３のラグ値を決定し、

前記劣化したフレームが部分的に劣化していれば、前記劣化したフレームにおける前記第１の長期予測ラグ値を第４のラグ値に交換する工程を含み、前記部分的に劣化したフレームが構成されている音声シーケンスが定常的である場合、前記第４のラグ値を前記最終の長期予測ラグ値に等しく設定し、前記音声シーケンスが非定常的である場合、前記劣化したフレームに先行する非劣化フレームに関連づけられた適応型コードブックから検索され

る復号された長期予測ラグ値に基づいて前記第４のラグ値を設定する。

【００１８】

本発明の第２の態様は、音声信号を符号化されたビット・ストリームに符号化し、かつ符号化されたビット・ストリームを合成音声に復号するための音声信号送受信機システムであって、当該システムにおいては、符号化されたビット・ストリームが音声シーケンスに配列された複数の音声フレームを含み、音声フレームが１または２以上の非劣化フレームに先行される少なくとも１つの劣化したフレームを含み、該劣化したフレームが第１の信号で表示されかつ第１の長期予測ラグ値と第１の長期予測利得値とを含み、該非劣化フレームが第２の長期予測ラグ値と第２の長期予測利得値とを含み、該第２の長期予測ラグ値が最終の長期予測ラグ値を含み、該第２の長期予測利得値が最終の長期予測利得値を含み、前記音声シーケンスが定常的および非定常的音声シーケンスを含んでいる。当該システムは、

前記第１の信号に応答して、前記第２の長期予測利得値に基づく、劣化したフレームが構成される音声シーケンスが定常的であるか、非定常的であるかの決定、および音声シーケンスが定常的であるか、非定常的であるかを表示する第２の信号の供給とを行なうための第１の機構と、

該第２の信号に応答して、前記音声シーケンスが定常的である場合、前記劣化したフレームにおける前記第１の長期予測ラグ値を前記最終の長期予測ラグ値に交換し、前記音声シーケンスが非定常的である場合、前記劣化したフレームにおける前記第１の長期予測ラグ値と第１の長期予測利得値とを各々第３の長期予測ラグ値と第３の長期予測利得値とに交換するための第２の機構とを備え、該第３の長期予測ラグ値が前記第２の長期予測ラグ値と適応的に制限されたランダムラグジッタとに基づいて決定され、該第３の長期予測利得値が前記第２の長期予測利得値と適応的に制限されたランダム利得ジッタとに基づいて決定される。

【００１９】

好適には、前記第３の長期予測ラグ値は、少なくとも部分的に前記第２の長期予測ラグ値の加重中央値に基づいて計算され、前記適応的に制限されたランダムラグジッタは、前記第２の長期予測ラグ値に基づいて決定された限定値に拘束された値である。

【００２０】

好適には、前記第３の長期予測利得値は、少なくとも部分的に前記第２の長期予測利得値の加重中央値に基づいて計算され、前記適応的に制限されたランダム利得ジッタは、前記第２の長期予測利得値に基づいて決定された限定値に拘束された値である。

【００２１】

本発明の第３の態様は、符号化されたビット・ストリームから音声を合成するためのデコーダであって、当該デコーダにおいては、符号化されたビット・ストリームは音声シーケンスに構成された複数の音声フレームを含み、音声フレームが１または２以上の非劣化フレームに先行される少なくとも１つの劣化したフレームを含み、該劣化したフレームが第１の信号で表示されかつ第１の長期予測ラグ値と第１の長期予測利得値とを含み、該非劣化フレームが第２の長期予測ラグ値と第２の長期予測利得値とを含み、該第２の長期予測ラグ値が最終の長期予測ラグ値を含み、該第２の長期予測利得値が最終の長期予測利得値を含み、前記音声シーケンスが定常的および非定常的音声シーケンスを含んでいる。当該デコーダは、

前記第１の信号に応答して、前記第２の長期予測利得値に基く、前記劣化したフレームが構成された音声シーケンスが定常的であるか、非定常的であるかの決定、および音声シーケンスが定常的であるか、非定常的であるかを表示する第２の信号を供給とを行なうための第１の機構と、

該第２の信号に応答して、前記音声シーケンスが定常的である場合、前記劣化したフレームにおける前記第１の長期予測ラグ値を前記最終の長期予測ラグ値に交換し、前記音声シーケンスが非定常的である場合、前記劣化したフレームにおける前記第１の長期予測ラグ値と前記第１の長期予測利得値とを各々第３の長期予測ラグ値と第３の長期予測利得値と

に交換するための第２の機構とを備え、該第３の長期予測ラグ値は前記第２の長期予測ラグ値と適応的に制限されたランダムラグジッタとに基づいて決定され、該第３の長期予測利得値は前記第２の長期予測利得値と適応的に制限されたランダム利得ジッタとに基づいて決定される。

【００２２】

本発明の第４の態様は、音声信号を表示する音声データを含む符号化されたビット・ストリームを受信するように構成された移動局であって、当該移動局においては、符号化されたビット・ストリームが音声シーケンスに構成された複数の音声フレームを含み、音声フレームが１または２以上の非劣化フレームに先行される少なくとも１つの劣化したフレームを含み、該劣化したフレームが第１の信号で表示されかつ第１の長期予測ラグ値と第１の長期予測利得値とを含み、該非劣化フレームが第２の長期予測ラグ値と第２の長期予測利得値とを含み、該第２の長期予測ラグ値が最終の長期予測ラグ値を含み、該第２の長期予測利得値が最終の長期予測利得値を含み、前記音声シーケンスが定常的および非定常的音声シーケンスを含んでいる。当該移動局は、

前記第１の信号に応答して、前記第２の長期予測利得値に基く、前記劣化したフレームが構成された音声シーケンスが定常的であるか、非定常的であるかの決定、および音声シーケンスが定常的であるか、非定常的であるかを表示する第２の信号を供給とを行なうための第１の機構と、

該第２の信号に応答して、前記音声シーケンスが定常的である場合、前記劣化したフレームにおける前記第１の長期予測ラグ値を前記最終の長期予測ラグ値に交換し、前記音声シーケンスが非定常的である場合、前記劣化したフレームにおける前記第１の長期予測ラグ値と前記第１の長期予測利得値とを各々第３の長期予測ラグ値と第３の長期予測利得値とに交換するための第２の機構とを備え、該第３の長期予測ラグ値は前記第２の長期予測ラグ値と適応的に制限されたランダムラグジッタとに基づいて決定され、該第３の長期予測利得値は前記第２の長期予測利得値と適応的に制限されたランダム利得ジッタとに基づいて決定される。

【００２３】

本発明の第５の態様は、音声データを含む符号化されたビット・ストリームを移動局から受信するように構成された電気通信網における要素であって、当該要素においては、音声データが音声シーケンスに構成された複数の音声フレームを含み、音声フレームが１または２以上の非劣化フレームに先行される少なくとも１つの劣化したフレームを含み、該劣化したフレームが第１の信号で表示されかつ第１の長期予測ラグ値と第１の長期予測利得値とを含み、該非劣化フレームが第２の長期予測ラグ値と第２の長期予測利得値とを含み、該第２の長期予測ラグ値は最終の長期予測ラグ値を含み、該第２の長期予測利得値は最終の長期予測利得値を含み、前記音声シーケンスは定常的および非定常的音声シーケンスを含んでいる。本要素は、

前記第１の信号に応答して、前記第２の長期予測利得値に基く、前記劣化したフレームが構成された音声シーケンスが定常的であるか、非定常的であるかの決定、および音声シーケンスが定常的であるか、非定常的であるかを表示する第２の信号を供給とを行なうための第１の機構と、

該第２の信号に応答して、前記音声シーケンスが定常的である場合、前記劣化したフレームにおける前記第１の長期予測ラグ値を前記最終の長期予測ラグ値に交換し、前記音声シーケンスが非定常的である場合、前記劣化したフレームにおける前記第１の長期予測ラグ値と前記第１の長期予測利得値とを各々第３の長期予測ラグ値と第３の長期予測利得値とに交換するための第２の機構とを備え、該第３の長期予測ラグ値は前記第２の長期予測ラグ値と適応的に制限されたランダムラグジッタとに基づいて決定され、該第３の長期予測利得値は前記第２の長期予測利得値と適応的に制限されたランダム利得ジッタとに基づいて決定される。

【００２４】

本発明は、図３ないし１１ｃに関連して行う説明を読めば明らかになるであろう。

【００２５】
［発明を実施するための最良の形態］
図３は、復号モジュール２０とエラー隠蔽モジュール３０とを含む復号器（デコーダ）１０を示している。復号モジュール２０は、通常は音声合成のための音声パラメータ１０２を示す信号１４０を受信する。この復号モジュール２０は、技術上周知である。エラー隠蔽モジュール３０は、符号化されたビット・ストリーム１００を受信するように構成されている。符号化されたビット・ストリーム１００は、音声シーケンス中で構成された複数の音声ストリームを含む。不良フレーム検出デバイス３２は、音声シーケンス中の劣化したフレームを検出するため、および劣化したフレームが検出された場合、不良フレームインジケータ（ＢＦＩ）フラグを表すＢＦＩ信号１１０を提供するために使用される。ＢＦＩもまた、技術上周知である。ＢＦＩ信号１１０は、２つのスイッチ４０および４２を制御するために使用される。通常、音声フレームは劣化されず、ＢＦＩフラグは０である。スイッチ４０および４２では、端子Ｓが端子０に動作可能なように接続されている。音声パラメータ１０２はバッファすなわち「パラメータヒストリー」記憶装置５０、および音声合成のための復号モジュール２０に伝達される。不良フレーム検出デバイス３２によって不良フレームが検出されると、ＢＦＩフラグは１に設定される。スイッチ４０および４２では、端子Ｓが端子１に接続される。したがって、音声パラメータ１０２はアナライザ７０に供給され、音声合成に必要な音声パラメータがパラメータ隠蔽モジュール６０により復号モジュール２０へ供給される。音声パラメータ１０２は、典型的には、短期予測のためのＬＰＣパラメータ、励振パラメータ、長期予測（ＬＴＰ）ラグ・パラメータ、ＬＴＰ利得パラメータおよび他の利得パラメータを含んでいる。パラメータヒストリー記憶装置５０は、多数の非劣化音声フレームのＬＴＰラグおよびＬＴＰ利得を格納するために使用される。パラメータヒストリー記憶装置５０の内容は絶えず更新され、記憶装置５０に格納された最終のＬＴＰ利得パラメータおよび最終のＬＴＰラグパラメータは、最終の非劣化音声フレームのＬＴＰ利得パラメータおよびＬＴＰラグパラメータである。音声シーケンスにおける劣化したフレームが復号器１０に受信されると、ＢＦＩフラグが１に設定され、劣化したフレームの音声パラメータ１０２はスイッチ４０を介してアナライザ７０へ伝達される。アナライザ７０は、劣化したフレームにおけるＬＴＰ利得パラメータと記憶装置５０に記憶されたＬＴＰ利得パラメータとを比較することにより、隣接フレームにおけるＬＴＰ利得パラメータの大きさおよびその変動に基づいて、音声シーケンスが定常的であるか、非定常的であるかを決定することができる。典型的には、定常的シーケンスでは、図７が示すように、ＬＴＰ利得パラメータは高い値でかなり安定しており、ＬＴＰラグ値は安定していて、隣接するＬＴＰラグ値の変動は小さい。これに対して非定常的シーケンスでは、図８が示すように、ＬＴＰ利得パラメータは低い値で不安定であり、ＬＴＰラグも不安定である。ＬＴＰラグ値は、多少はランダムに変化する。図７は、単語「ｖｉｉｎｉａ」の音声シーケンスを示している。図８は、単語「ｅｘｈｉｂｉｔｉｏｎ」の音声シーケンスを示している。
【００２６】
もし劣化したフレームを含む音声シーケンスが有声または定常的であれば、記憶装置５０から最終の優良ＬＴＰラグが検索され、パラメータ隠蔽モジュール６０に伝達される。検索された優良ＬＴＰラグは、劣化したフレームのＬＴＰラグと交換するために使用される。定常的音声シーケンスにおけるＬＴＰラグは安定していてその変動は小さいため、劣化したフレームにおける対応パラメータを隠蔽するため、先行するＬＴＰラグを僅かに変更して使用することが妥当である。続いて、ＲＸ信号１０４により、参照数字１３４が示す交換パラメータがスイッチ４２を介して復号モジュール２０に伝達される。
【００２７】
もし劣化したフレームを含む音声シーケンスが無声または非定常的であれば、アナライザ７０は、パラメータ隠蔽のための交換ＬＴＰラグ値および交換ＬＴＰ利得値を計算する。非定常的音声シーケンスにおけるＬＴＰラグは不安定であり、かつ隣接フレームにおけるその変動は典型的にはきわめて大きいため、パラメータの隠蔽は、エラーを隠蔽される非

定常的シーケンスにおけるＬＴＰラグがランダムに変動することを許容するものでなければならない。劣化したフレームにおけるパラメータが、損失フレームの場合のように全体的に劣化していれば、交換ＬＴＰラグが、先行する優良ＬＴＰラグ値の加重中央値および適応的に制限されたランダムジッタ（ａｄａｐｔｉｖｅｌｙ－ｌｉｍｉｔｅｄ ｒａｎｄｏｍ ｊｉｔｔｅｒ）を使用して計算される。適応的に制限されたランダムジッタは、ＬＴＰ値のヒストリから計算された限界内で変化することができるため、エラー隠蔽セグメントにおけるパラメータ変動は、同じ音声シーケンスの先行する優良部分に類似している。

【００２８】

ＬＴＰラグ隠蔽のための例示的規則は、下記のような条件セットによって規定される。

もし、

ｍｉｎＧａｉｎ＞０．５ かつ ＬａｇＤｉｆ＜１０；または

ｌａｓｔＧａｉｎ＞０．５ かつ ｓｅｃｏｎｄＬａｓｔＧａｉｎ＞０．５

であれば、全体的に劣化したフレームに関して最終に受信された優良ＬＴＰラグが使用される。

そうでなければ、全体的に劣化したフレームに関して、ランダム化によるＬＴＰラグバッファの加重平均であるＵｐｄａｔｅ＿ｌａｇが使用される。Ｕｐｄａｔｅ＿ｌａｇは、以下に述べる方法で計算される。

【００２９】

ＬＴＰラグバッファはソートされ、３つの最大バッファ値が検索される。これらの３つの最大値の平均は加重平均ラグ（ＷＡＬ）と呼ばれ、これらの最大値との差は加重ラグ差（ＷＬＤ）と呼ばれる。

ＲＡＮＤをスケール（－ＷＬＤ／２，ＷＬＤ／２）を有するランダム化（ｒａｎｄｏｍｉｚａｔｉｏｎ）であるとすると、

Ｕｐｄａｔｅ＿ｌａｇ＝ＷＡＬ＋ＲＡＮＤ（－ＷＬＤ／２，ＷＬＤ／２）

となる。ここで、

ｍｉｎＧａｉｎは、ＬＴＰ利得バッファの最小値であり、

ＬａｇＤｉｆは、最小および最大ＬＴＰラグ値の差であり、

ｌａｓｔＧａｉｎは、受信された最終の優良ＬＴＰ利得であり、

ｓｅｃｏｎｄＬａｓｔＧａｉｎは、受信された最終から２番目の優良ＬＴＰ利得である。

【００３０】

劣化したフレームにおけるパラメータが部分的に劣化していれば、該劣化したフレームにおけるＬＴＰラグ値が適宜交換される。フレームが部分的に劣化していることは、以下に与えられる典型的ＬＴＰ特徴基準のセットによって決定される。

もし、

（１）ＬａｇＤｉｆ＜１０ かつ （ｍｉｎＬａｇ－５）＜T_{bf}＜（ｍａｘＬａｇ＋５）；または

（２）ｌａｓｔＧａｉｎ＞０．５ かつ ｓｅｃｏｎｄＬａｓｔＧａｉｎ＞０．５ かつ （ｌａｓｔＬａｇ－１０）＜T_{bf}＜（ｌａｓｔＬａｇ＋１０）；または

（３）ｍｉｎＧａｉｎ＜０．４ かつ ｌａｓｔＧａｉｎ＝ｍｉｎＧａｉｎ かつ ｍｉｎＬａｇ＜T_{bf}＜ｍａｘＬａｇ；または

（４）ＬａｇＤｉｆ＜７０ かつ ｍｉｎＬａｇ＜T_{bf}＜ｍａｘＬａｇ；または

（５）ｍｅａｎＬａｇ＜T_{bf}＜ｍａｘＬａｇ

が真であれば、劣化したフレームにおけるＬＴＰラグの交換にT_{bf}が使用される。真でなければ、上述のように劣化したフレームは全体的に劣化したフレームとして処理される。上記条件において、

ｍａｘＬａｇは、ＬＴＰラグバッファの最大値であり、

ｍｅａｎＬａｇは、ＬＴＰラグバッファの平均値であり、

ｍｉｎＬａｇは、ＬＴＰラグバッファの最小値であり、

ｌａｓｔＬａｇは、受信された最終の優良ＬＴＰラグ値であり、

T_{bf}は、ＢＦＩが設定されているときに、ＢＦＩがあたかも設定されていないかのように適応型コードブックから検索される復号化されたＬＴＰラグである。

【００３１】

図９および１０は、パラメータ隠蔽の２つの例を示したものである。図が示すように、従来技術による不良フレームにおける交換ＬＴＰラグ値のプロファイルはどちらかといえば平坦であるが、本発明による交換のプロファイルは、エラーのないプロファイルと同様幾分かの変動を許容する。従来技術のアプローチと本発明との相違は、図１１ａに示されているようなエラーのないチャネルにおける音声信号に基づいて、各々図１１ｂおよび１１ｃにさらに詳しく示されている。

【００３２】

劣化したフレームにおけるパラメータが部分的に劣化している場合は、パラメータ隠蔽をさらに最適化することができる。部分的に劣化したフレームでは、劣化したフレームにおけるＬＴＰラグは、依然として許容される合成音声セグメントをもたらすことができる。ＧＳＭ仕様にしたがって、ＢＦＩフラグがサイクリック冗長検査（ＣＲＣ）機構または他のエラー検出機構により設定される。これらのエラー検出機構は、チャネル復号プロセスにおいて最上位（ｍｏｓｔ　ｓｉｇｎｉｆｉｃａｎｔ）のビットにおけるエラーを検出する。したがって、ほんの僅かのビットにエラーがあってもエラーが検出され得て、その結果ＢＦＩフラグが設定される。従来技術によるパラメータ隠蔽アプローチでは、フレーム全体が放棄される。その結果、正常なビットに含まれる情報が捨てられる。

【００３３】

典型的には、チャネル復号プロセスでは、フレーム当たりのＢＥＲがチャネル状態の良い指針となる。チャネル状態が良ければ、フレーム当たりのＢＥＲは小さく、エラーのあるフレームにおけるＬＴＰラグ値は高い率で適正である。たとえば、フレームエラー率（ＦＥＲ）が０．２％のとき、７０％を超えるＬＴＰラグ値は適正である。ＦＥＲが３％に届くような場合でも、ＬＴＰラグ値の約６０％は依然として適正であろう。ＣＲＣは、不良フレームを正確に検出して適宜ＢＦＩフラグを設定することができる。しかしながらＣＲＣは、フレームにおけるＢＥＲの推定値を供給しない。ＢＦＩフラグがパラメータ隠蔽に関する唯一の基準として使用されれば、適正なＬＴＰラグ値の多くの割合が廃棄される可能性がある。大量の適正なＬＴＰラグが放棄されることを防ぐためには、パラメータ隠蔽の決定基準をＬＴＰヒストリに基づいて適合化することが可能である。また、たとえばＦＥＲを決定基準として使用することも可能である。ＬＴＰラグが決定基準に適合すれば、パラメータ隠蔽の必要はない。この場合、アナライザ７０は、スイッチ４０を介して受信した通りの音声パラメータ１０２をパラメータ隠蔽モジュール６０に伝え、パラメータ隠蔽モジュール６０は次にこれをスイッチ４２を介して復号モジュール２０に伝える。もしＬＴＰラグが上記決定基準に適合していなければ、劣化したフレームはパラメータ隠蔽のため、上述のようにＬＴＰ特徴基準を使用してさらに調べられる。

【００３４】

定常的音声シーケンスでは、ＬＴＰラグはきわめて安定している。劣化したフレームにおけるＬＴＰラグ値の大部分が適正であるかエラーであるかは、高い確率で正確に予測することができる。したがって、きわめて厳密な基準をパラメータ隠蔽用に適応させることが可能である。非定常的音声シーケンスでは、ＬＴＰパラメータの非安定的性質により、劣化したフレームにおけるＬＴＰラグ値が適正であるかどうかの予測は困難であると言える。しかしながら、非定常的音声の場合、予測が正しいか誤りかということは定常的音声の場合ほど重要ではない。エラーのあるＬＴＰラグ値を定常的音声の復号に使用できるようにすることは、合成された音声を認識できないものにしてしまうかも知れない一方、エラーのあるＬＴＰラグ値を非定常的音声の復号に使用できるようすることは、通常可聴アーチファクトを増大させるだけである。したがって、非定常的音声におけるパラメータ隠蔽の決定基準は、比較的緩いものであり得る。

【００３５】

前述のとおり、ＬＴＰ利得は非定常的音声において大きく変動する。もし最終の優良フレ

ームからの同じＬＴＰ利得値が、音声シーケンスにおける１または２以上の劣化したフレームのＬＴＰ利得値に置換するため繰り返し使用されると、利得を隠蔽されたセグメントにおけるＬＴＰ利得プロファイルは（図７および８が示すように、従来技術によるＬＴＰラグの交換と同様に）平らになり、非劣化フレームの変動するプロファイルとは全く対照的である。ＬＴＰ利得プロファイルの突然の変化は、不快な可聴アーチファクトをもたらす可能性がある。これらの可聴アーチファクトを最小限に抑えるために、エラー隠蔽セグメントにおいて交換ＬＴＰ利得値を変動させることが可能である。この目的に沿ってアナライザ７０を限界値を決定するために使用することもできる。交換ＬＴＰ利得値は、ＬＴＰヒストリにおける利得値に基づき、該限界値のあいだで変動できる。

【００３６】

ＬＴＰ利得の隠蔽は、以下のようなやり方で実行することができる。ＢＦＩが設定されると、ＬＴＰ利得隠蔽規則のセットにしたがって交換ＬＴＰ利得値が計算される。交換ＬＴＰ利得は、Ｕｐｄａｔｅｄ＿ｇａｉｎで表される。

（１）ｇａｉｎＤｉｆ＞０．５　ＡＮＤ　ｌａｓｔＧａｉｎ＝ｍａｘＧａｉｎ＞０．９　ＡＮＤ　ｓｕｂＢＦ＝１であれば、

Ｕｐｄａｔｅｄ＿ｇａｉｎ＝（ｓｅｃｏｎｄＬａｓｔＧａｉｎ＋ｔｈｉｒｄＬａｓｔＧａｉｎ）／２であり、

（２）ｇａｉｎＤｉｆ＞０．５　ＡＮＤ　ｌａｓｔＧａｉｎ＝ｍａｘＧａｉｎ＞０．９　ＡＮＤ　ｓｕｂＢＦ＝２であれば、

Ｕｐｄａｔｅｄ＿ｇａｉｎ＝ｍｅａｎＧａｉｎ＋ｒａｎｄＶａｒ＊（ｍａｘＧａｉｎ－ｍｅａｎＧａｉｎ）であり、

（３）ｇａｉｎＤｉｆ＞０．５　ＡＮＤ　ｌａｓｔＧａｉｎ＝ｍａｘＧａｉｎ＞０．９　ＡＮＤ　ｓｕｂＢＦ＝３であれば、

Ｕｐｄａｔｅｄ＿ｇａｉｎ＝ｍｅａｎＧａｉｎ－ｒａｎｄＶａｒ＊（ｍｅａｎＧａｉｎ－ｍｉｎＧａｉｎ）であり、

（４）ｇａｉｎＤｉｆ＞０．５　ＡＮＤ　ｌａｓｔＧａｉｎ＝ｍａｘＧａｉｎ＞０．９　ＡＮＤ　ｓｕｂＢＦ＝４であれば、

Ｕｐｄａｔｅｄ＿ｇａｉｎ＝ｍｅａｎＧａｉｎ＋ｒａｎｄＶａｒ＊（ｍａｘＧａｉｎ－ｍｅａｎＧａｉｎ）である。

前の条件では、Ｕｐｄａｔｅｄ＿ｇａｉｎはｌａｓｔＧａｉｎより大きくなることはできない。前の条件が満たされ得ない場合は、以下の条件が使用される。

（５）ｇａｉｎＤｉｆ＞０．５であれば、

Ｕｐｄａｔｅｄ＿ｇａｉｎ＝ｌａｓｔＧａｉｎであり、

（６）ｇａｉｎＤｉｆ＜０．５　ＡＮＤ　ｌａｓｔＧａｉｎ＝ｍａｘＧａｉｎであれば、

Ｕｐｄａｔｅｄ＿ｇａｉｎ＝ｍｅａｎＧａｉｎであり、

（７）ｇａｉｎＤＩＦ＜０．５であれば、

Ｕｐｄａｔｅｄ＿ｇａｉｎ＝ｌａｓｔＧａｉｎである。

ここで、

ｍｅａｎＧａｉｎは、ＬＴＰ利得バッファの平均であり、

ｍａｘＧａｉｎは、ＬＴＰ利得バッファの最大値であり、

ｍｉｎＧａｉｎは、ＬＴＰ利得バッファの最小値であり、

ｒａｎｄＶａｒは、０と１のあいだのランダム値であり、

ｇａｉｎＤＩＦは、ＬＴＰ利得バッファにおける最小ＬＴＰ利得値と最大ＬＴＰ利得値との差であり、

ｌａｓｔＧａｉｎは、受信された最終の優良ＬＴＰ利得であり、

ｓｅｃｏｎｄＬａｓｔＧａｉｎは、受信された最終から２番目の優良ＬＴＰ利得であり、

ｔｈｉｒｄＬａｓｔＧａｉｎは、受信された最終から３番目の優良ＬＴＰ利得であり、

ｓｕｂＢＦは、サブフレームの次数である。

【００３７】

図４は、本発明によるエラー隠蔽の方法を示している。工程（ステップ）１６０で符号化

されたビット・ストリームが受信されると、工程１６２でフレームが劣化しているかどうかがチェックされる。フレームが劣化していなければ、工程１６４で音声シーケンスのパラメータヒストリーが更新され、工程１６６で現行フレームの音声パラメータが復号される。手順は、次に工程１６２に戻る。フレームが不良フレームであるか、または劣化していれば、工程１７０でパラメータがパラメータヒストリー記憶装置から検索される。工程１７２では、劣化したフレームが定常的音声シーケンスの一部であるか、または非定常的音声シーケンスの一部であるかが決定される。音声シーケンスが定常的であれば、工程１７４で最終の優良フレームのＬＴＰラグを使用して劣化したフレームにおけるＬＴＰラグが交換される。音声シーケンスが非定常的であれば、工程１８０でＬＴＰヒストリーに基づいて新たなラグ値と新たな利得値とが計算され、工程１８２でこれら新たなラグ値と新たな利得値を使用して劣化したフレームにおける対応するパラメータが交換される。

【００３８】

図５は、本発明の典型的な一実施形態による移動局２００のブロック図である。本移動局は、マイクロフォン２０１、キーパッド２０７、ディスプレイ２０６、イヤホン２１４、送信／受信スイッチ２０８、アンテナ２０９および制御ユニット２０５など、本デバイスの典型的部品を備えている。さらに本図は、移動局にとって典型的な送信機および受信機ブロック２０４、２１１を示している。送信機ブロック２０４は、音声信号を符号化するためのコーダ２２１を備えている。送信機ブロック２０４はまた、チャネル符号化、解読および変調並びにＲＦ機能に必要なオペレーションも備えているが、明瞭化のために図５には描かれていない。受信機ブロック２１１もまた、本発明による復号ブロック２２０を備えている。復号ブロック２２０は、図３が示すパラメータ隠蔽モジュール３０のようなエラー隠蔽モジュール２２２を備えている。マイクロフォン２０１から着信する信号は、増幅ステージ２０２で増幅され、Ａ／Ｄ変換器でデジタル化されて送信機ブロック２０４に送られ、典型的には送信ブロックに含まれる音声符号化デバイスに送られる。送信ブロックによって処理され、変調されかつ増幅された送信信号は、送信／受信スイッチ２０８を介してアンテナ２０９に送られる。受信される信号はアンテナから送信／受信スイッチ２０８を介して受信機ブロック２１１へ送られ、受信機ブロック２１１は受信された信号を復調し、解読およびチャネルコーディングを復号する。結果的に得られる音声信号は、Ｄ／Ａ変換器２１２を介して増幅器２１３に、さらにイヤホン２１４にと送られる。制御ユニット２０５は、移動局２００の動作を制御し、ユーザによってキーパッド２０７から与えられる制御コマンドを読取り、かつディスプレイ２０６によりユーザにメッセージを与える。

【００３９】

本発明によるパラメータ隠蔽モジュール３０はまた、一般的な電話網のような電気通信網３００において、またはＧＳＭ網のような移動局網においても使用することができる。図６は、こうした電気通信網のブロック図の一例である。たとえば、電気通信網３００は電話交換機（ｔｅｌｅｐｈｏｎｅ ｅｘｃｈａｎｇｅ）または対応する交換システム（ｓｗｔｉｃｈｉｎｇ ｓｙｓｔｅｍ）３６０を備えることが可能であり、これに電気通信網の通常の電話３７０、基地局３４０、基地局コントローラ３５０および他の中央デバイス３５５が結合されている。移動局３３０は、基地局３４０を介して電気通信網への接続を確立することができる。図３に示されるエラー隠蔽モジュール３０に類似するエラー隠蔽モジュール３２２を含む復号ブロック３２０は、たとえば基地局３４０に特に有利に配置されることが可能である。しかし復号ブロック３２０は、たとえば基地局コントローラ３５０または他の中央または交換デバイス３５５にも配置されることが可能である。移動局システムが、たとえば基地局と基地局コントローラとのあいだで別個のトランスコーダ（ｔｒａｎｓｃｏｄｅｒ）を使用して、無線チャネル上で取りこまれた符号化された信号を電気通信システム内で転送される典型的な毎秒６４キロビットの信号に変換する場合、かつ、この逆の変換を行う場合には、復号ブロック３２０をそのようなトランスコーダ内に配置することもできる。概して、パラメータ隠蔽モジュール３２２を含む復号ブロック３２０は、符号化されたデータストリームを符号化されていないデータストリームに変換する

電気通信網３００の任意の要素内に配置されることが可能である。復号ブロック３２０は、移動局３３０から着信する符号化された音声信号を復号して濾波し、音声信号はその後、電気通信網３００内の前方向へ圧縮されずに通常の方法で転送される。

【００４０】

本発明のエラー隠蔽方法は、定常的および非定常的の音声シーケンスに関連して説明されていること、および定常的音声シーケンスは一般に有声であり、非定常的音声シーケンスは一般に無声であることは留意されなければならない。したがって、開示された本方法は、有声および無声の音声シーケンスにおけるエラー隠蔽に適用可能である点は理解されるであろう。

【００４１】

本発明は、ＣＥＬＰ型の音声コーデックに適用可能であり、かつ他のタイプの音声コーデックにも適応させることができる。したがって、本発明はその好適な実施形態に関連して説明されているが、当業者には、その形式および詳細に関して、本発明の精神および範囲を逸脱することなく上述の、および他の様々な変更、省略および偏向を実行可能であることが理解されるであろう。

【図面の簡単な説明】

【図１】

音声データを含む符号化されたビット・ストリームが符号器から通信チャネルまたは記憶媒体を介して復号器（デコーダ）へ伝達される、総称的な分散音声コーデックを示すブロック図である。

【図２】

受信機における従来技術によるエラー隠蔽装置を示すブロック図である。

【図３】

受信機における本発明によるエラー隠蔽装置を示すブロック図である。

【図４】

本発明によるエラー隠蔽方法を示すフローチャートである。

【図５】

本発明によるエラー隠蔽モジュールを含む移動局のダイヤグラム表示である。

【図６】

本発明によるデコーダを使用する電気通信網のダイヤグラム表示である。

【図７】

有声音声シーケンスにおけるラグおよび利得プロファイルを示すＬＴＰパラメータのプロットである。

【図８】

無声音声シーケンスにおけるラグおよび利得プロファイルを示すＬＴＰパラメータのプロットである。

【図９】

従来技術によるエラー隠蔽アプローチと本発明によるアプローチとの相違を示す、一連のサブフレームにおけるＬＴＰラグ値のプロットである。

【図１０】

先行技術によるエラー隠蔽アプローチと本発明によるアプローチとの相違を示す、一連のサブフレームにおける他のＬＴＰラグ値のプロットである。

【図１１ａ】

図１１ｂおよび１１ｃに示されるような音声チャネルの不良フレームのロケーションを有するエラーのない音声シーケンスを示す音声信号のプロットである。

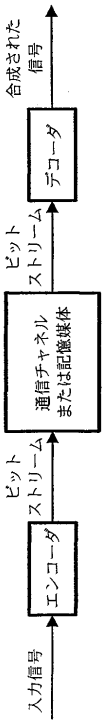
【図１１ｂ】

従来技術のアプローチによる不良フレームにおけるパラメータの隠蔽を示す音声信号のプロットである。

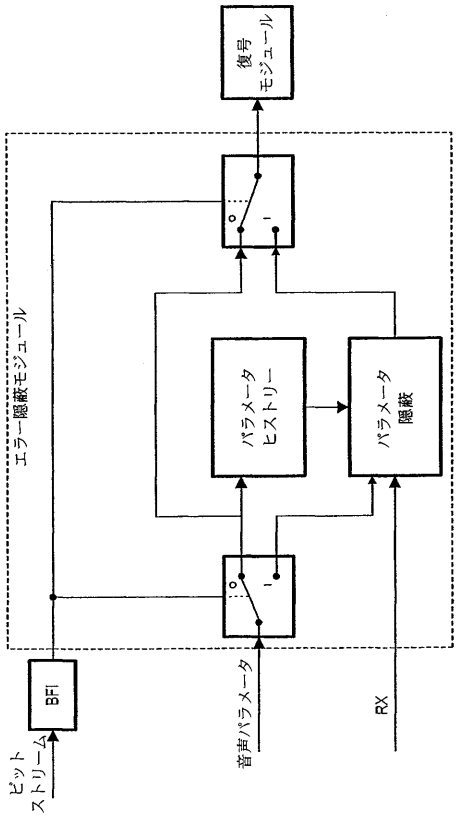
【図１１ｃ】

本発明による不良フレームにおけるパラメータの隠蔽を示す音声信号のプロットである。

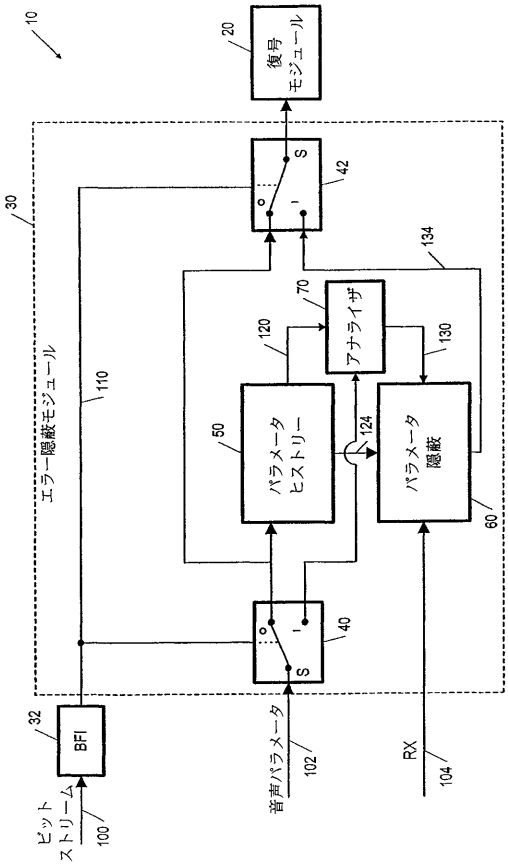
【図１】


入力信号
エンコーダ
ビットストリーム
通信チャネルまたは記憶媒体
ビットストリーム
デコーダ
合成された信号

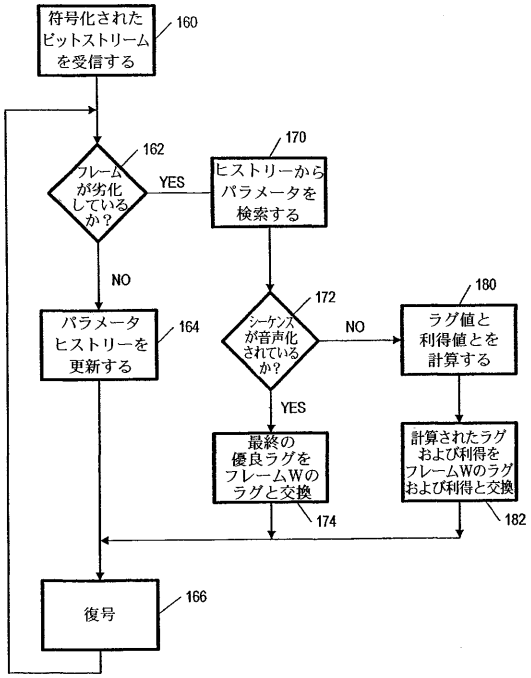
【図２】


復号モジュール
エラー隠蔽モジュール
パラメータヒストリー
パラメータ隠蔽
BFI
ビットストリーム
音声パラメータ
RX

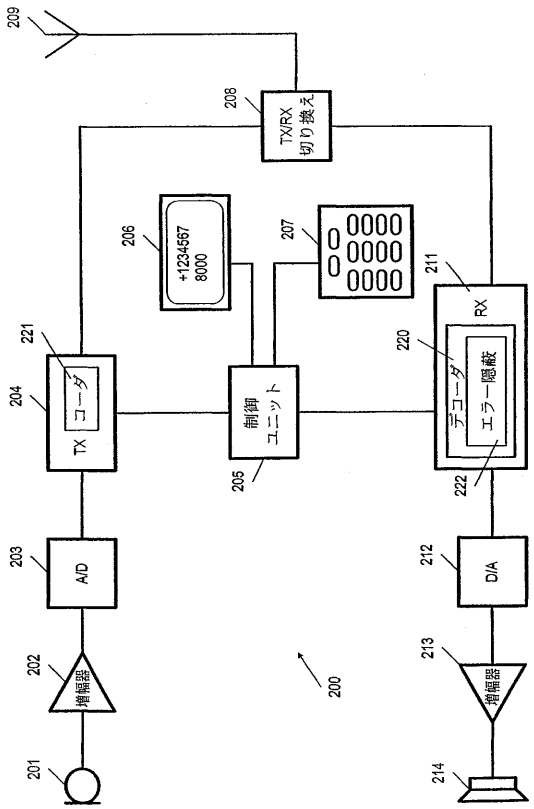
【図３】


10
20
復号モジュール
30
エラー隠蔽モジュール
42
134
70
アナライザ
120
130
110
50
パラメータヒストリー
124
パラメータ隠蔽
60
40
32
BFI
ビットストリーム
100
音声パラメータ
102
RX
104

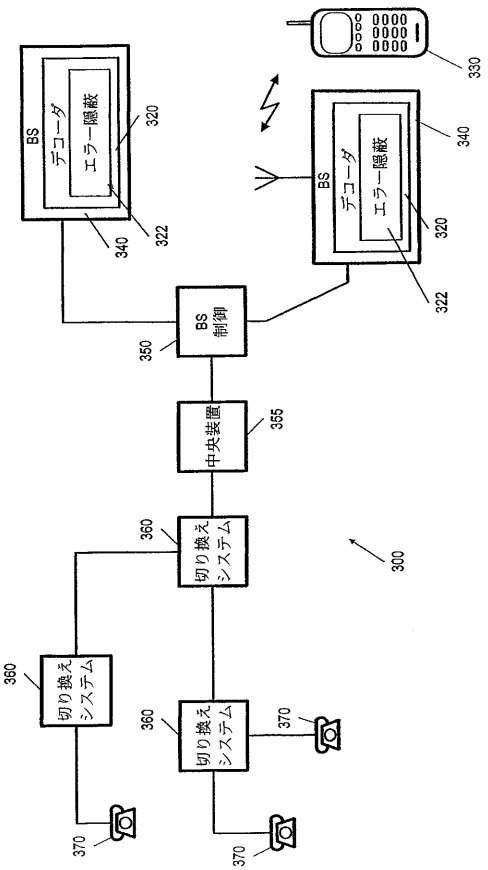
【図４】


符号化されたビットストリームを受信する
160
フレームが劣化しているか？
162
YES
NO
ヒストリーからパラメータを検索する
170
パラメータヒストリーを更新する
164
シーケンスが音声化されているか？
172
NO
YES
ラグ値と利得値とを計算する
180
最終の優良ラグをフレームWのラグと交換
174
計算されたラグおよび利得をフレームWのラグおよび利得と交換
182
復号
166

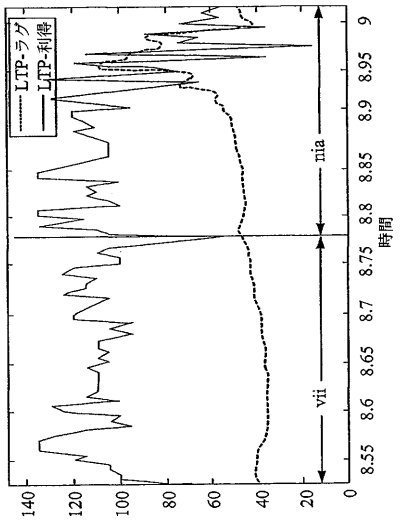
【図５】


209
208
TX/RX 切り換え
206
+1234567 8000
207
0000 0000 0000
211
RX
221
204
TX コーダ
制御 ユニット
220
デコーダ
エラー隠蔽
205
222
203
A/D
212
D/A
202
増幅器
213
増幅器
201
214
200

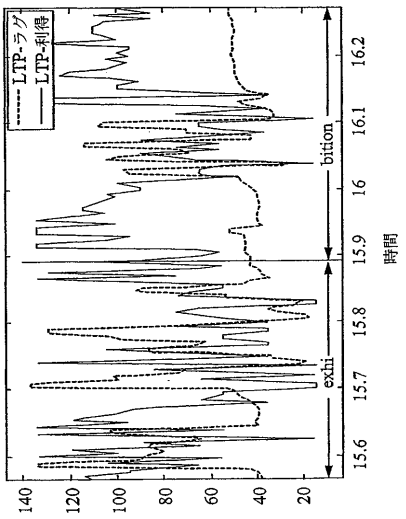
【図６】


330
BS
デコーダ
エラー隠蔽
320
340
322
BS
デコーダ
エラー隠蔽
340
320
322
BS 制御
350
中央装置
355
360
切り換え システム
300
360
切り換え システム
370
360
切り換え システム
370
370

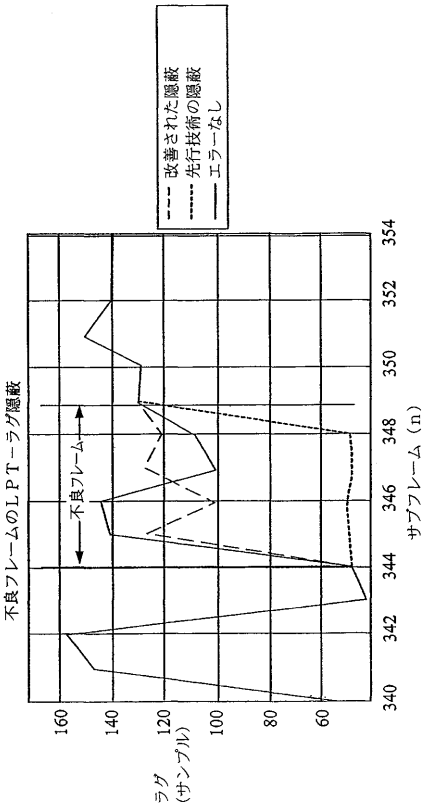
【図７】


LTP-ラグ
LTP-利得
mia
vii
時間
140
120
100
80
60
40
20
0
8.55
8.6
8.65
8.7
8.75
8.8
8.85
8.9
8.95
9

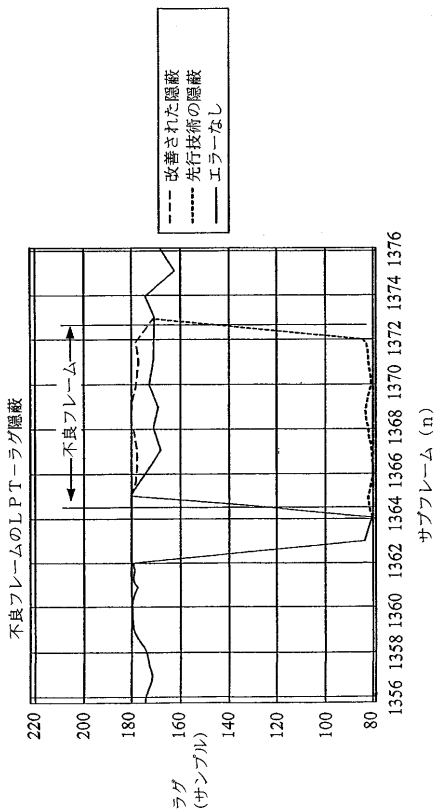
【図８】


LTP-ラグ
LTP-利得
bition
exhi
時間
140
120
100
80
60
40
20
15.6
15.7
15.8
15.9
16
16.1
16.2

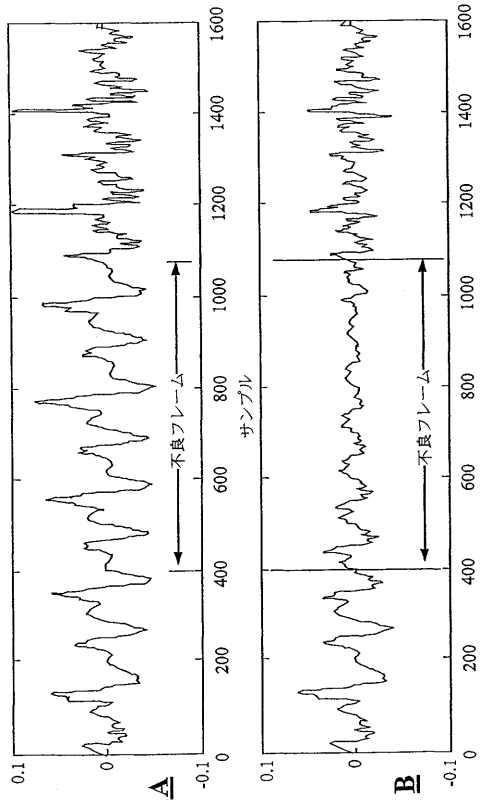
【図９】



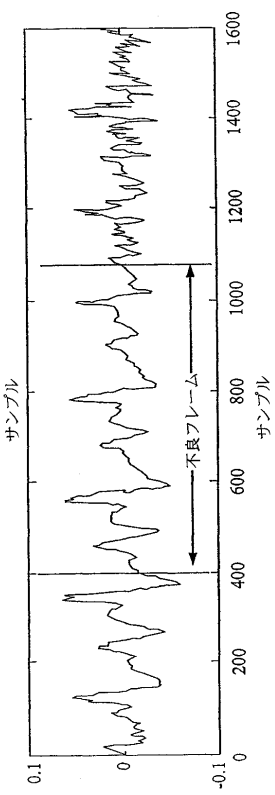
【図１０】



【図１１Ａ－Ｂ】



【図１１Ｃ】



【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 May 2002 (10.05.2002)

PCT

(10) International Publication Number
WO 02/37475 A1

(51) International Patent Classification7: G10L 19/00

(21) International Application Number: PCT/IB01/02021

(22) International Filing Date: 29 October 2001 (29.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/702,540 31 October 2000 (31.10.2000) US

(71) Applicant: NOKIA CORPORATION [FI/FI]; Keilalahdentie 4, FIN-02150 Espoo (FI).

(71) Applicant (for LC only): NOKIA INC. [US/US]; 6000 Connection Drive, Irving, TX 75039 (US).

(72) Inventors: MÄKINEN, Jari; Tammelan Puistokatu 30-32 C52, FIN-33100 Tampere (FI). MIKKOLA, Hannu, J.; Ippisenkatu 15, FIN-33300 Tampere (FI). VAINIO, Janne; Laurintie 16 C, FIN-33880 Lempäälä (FI). ROTOLA-PUKKILA, Jani; Lehvänkatu 24 E 44, FIN-33820 Tampere (FI).

(74) Agent: MAGUIRE, Francis, J.; Ware, Fressola, Van Der Sluys & Adolphson LLP, 755 Main Street, P.O. Box 224, Monroe, CT 06468 (US).

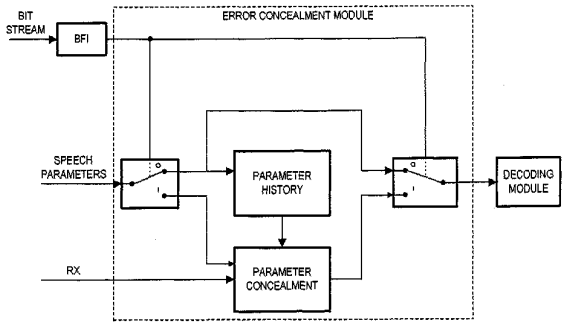
(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report

[Continued on next page]

(54) Title: METHOD AND SYSTEM FOR SPEECH FRAME ERROR CONCEALMENT IN SPEECH DECODING



(57) Abstract: A method and system for concealing errors in one or more bad frames in a speech sequence as part of an encoded bit stream received in a decoder. When the speech sequence is voiced, the LTP-parameters in the bad frames are replaced by the corresponding parameters in the last frame. When the speech sequence is unvoiced, the LTP-parameters in the bad frames are replaced by values calculated based on the LTP history along with an adaptively-limited random term.

WO 02/37475 A1

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD AND SYSTEM FOR SPEECH FRAME ERROR CONCEALMENT IN SPEECH DECODING

Field of the Invention

The present invention relates generally to the decoding of speech signals from an encoded bit stream and, more particularly, to the concealment of corrupted speech parameters when errors in speech frames are detected during speech decoding.

Background of the Invention

Speech and audio coding algorithms have a wide variety of applications in communication, multimedia and storage systems. The development of the coding algorithms is driven by the need to save transmission and storage capacity while maintaining the high quality of the synthesized signal. The complexity of the coder is limited by, for example, the processing power of the application platform. In some applications, for example, voice storage, the encoder may be highly complex, while the decoder should be as simple as possible.

Modern speech codecs operate by processing the speech signal in short segments called frames. A typical frame length of a speech codec is 20 ms, which corresponds to 160 speech samples, assuming an 8 kHz sampling frequency. In the wide band codecs, the typical frame length of 20 ms corresponds to 320 speech samples, assuming a 16 kHz sampling frequency. The frame may be further divided into a number of sub-frames. For every frame, the encoder determines a parametric representation of the input signal. The parameters are quantized and transmitted through a communication channel (or stored in a storage medium) in a digital form. The decoder produces a synthesized speech signal based on the received parameters, as shown in Figure 1.

A typical set of extracted coding parameters includes spectral parameters (such as Linear Predictive Coding (LPC) parameters) to be used in short term prediction of the signal, parameters to be used for long term prediction (LTP) of the signal, various gain parameters, and excitation parameters. The LTP parameter is closely related to the fundamental frequency of the speech signal. This parameter is often known as a so-called pitch-lag parameter, which describes the fundamental periodicity in terms of speech samples. Also, one of the gain parameters is very much related to the fundamental periodicity and so it is called LTP gain. The LTP gain is a very important parameter in

making the speech as natural as possible. The description of the coding parameters above fits in general terms with a variety of speech codecs, including the so-called Code-Excited Linear Prediction (CELP) codecs, which have for some time been the most successful speech codecs.

Speech parameters are transmitted through a communication channel in a digital form. Sometimes the condition of the communication channel changes, and that might cause errors to the bit stream. This will cause frame errors (bad frames), i.e., some of the parameters describing a particular speech segment (typically 20 ms) are corrupted. There are two kinds of frame errors: totally corrupted frames and partially corrupted frames. These frames are sometimes not received in the decoder at all. In the packet-based transmission systems, like in normal internet connections, the situation can arise when the data packet will never reach the receiver, or the data packet arrives so late that it cannot be used because of the real time nature of spoken speech. The partially corrupted frame is a frame that does arrive to the receiver and can still contain some parameters that are not in error. This is usually the situation in a circuit switched connection like in the existing GSM connection. The bit-error rate (BER) in the partially corrupted frames is typically around 0.5-5%.

From the description above, it can be seen that the two cases of bad or corrupted frames will require different approaches in dealing with the degradation in reconstructed speech due to the loss of speech parameters.

The lost or erroneous speech frames are consequences of the bad condition of the communication channel, which causes errors to the bit stream. When an error is detected in the received speech frame, an error correction procedure is started. This error correction procedure usually includes a substitution procedure and muting procedure. In the prior art, the speech parameters of the bad frame are replaced by attenuated or modified values from the previous good frame. However, some parameters (such as excitation in CELP parameters) in the corrupted frame may still be used for decoding.

Figure 2 shows the principle of the prior-art method. As shown in Figure 2, a buffer labeled "parameter history" is used to store the speech parameters of the last good frame. When a bad frame is detected, the Bad Frame Indicator (BFI) is set to 1 and the error concealment procedure is started. When the BFI is not set (BFI=0), the parameter history is updated and speech parameters are used for decoding without error

concealment. In the prior-art system, the error concealment procedure uses the parameter history for concealing the lost or erroneous parameters in the corrupted frames. Some speech parameters may be used from the received frame even though it is classified as a bad frame (BFI=1). For example, in a GSM Adaptive Multi-Rate (AMR) speech codec (ETSI specification 06.91), the excitation vector from the channel is always used. When the speech frames are totally lost frames (e.g., in some IP-based transmission systems), no parameters will be used from the received bad frame. In some cases, no frame will be received, or the frame will arrive so late that it has to be classified as a lost frame.

In a prior-art system, LTP-lag concealment uses the last good LTP-lag value with a slightly modified fractional part, and spectral parameters are replaced by the last good parameters slightly shifted towards constant mean. The gains (LTP and fixed codebook) may usually be replaced by the attenuated last good value or by the median of several last good values. The same substituted speech parameters are used for all sub-frames with slight modification to some of them.

The prior-art LTP concealment may be adequate for stationary speech signals, for example, voiced or stationary speech. However, for non-stationary speech signals, the prior-art method may cause unpleasant and audible artifacts. For example, when the speech signal is unvoiced or non-stationary, simply substituting the lag value in the bad frame with the last good lag value has the effect of generating a short voiced-speech segment in the middle of an unvoiced-speech burst (See Figure 10). The effect, as known as the "bing" artifact, can be annoying.

It is advantageous and desirable to provide a method and system for error concealment in speech decoding to improve the speech quality.

Summary of the Invention

The present invention takes advantage of the fact that there is a recognizable relationship among the long-term prediction (LTP) parameters in the speech signals. In particular, the LTP-lag has a strong correlation with the LTP-gain. When the LTP-gain is high and reasonably stable, the LTP-lag is typically very stable and the variation between adjacent lag values is small. In that case, the speech parameters are indicative of a voiced-speech sequence. When the LTP-gain is low or unstable, the LTP-lag is typically unvoiced, and the speech parameters are indicative of an unvoiced-speech sequence.

Once the speech sequence is classified as stationary (voiced) or non-stationary (unvoiced), the corrupted or bad frame in the sequence can be processed differently.

Accordingly, the first aspect of the present invention is a method for concealing errors in an encoded bit stream indicative of speech signals received in a speech decoder, wherein the encoded bit stream includes a plurality of speech frames arranged in speech sequences, and the speech frames include at least one corrupted frame preceded by one or more non-corrupted frames, wherein the corrupted frame includes a first long-term prediction lag value and a first long-term prediction gain value, and the non-corrupted frames include second long-term prediction lag values and second long-term prediction gain values, and wherein the second long-term prediction lag values include a last long-term prediction lag value, and the second long-term prediction gain values include a last long-term prediction gain value, and the speech sequences include stationary and non-stationary speech sequences, and wherein the corrupted frame can be partially corrupted or totally corrupted. The method comprises the steps of:

determining whether the first long-term prediction lag value is within or outside an upper limit and a lower limit determined based on the second long-term prediction lag values;

replacing the first long-term prediction lag value in the partially corrupted frame with a third lag value, when the first long-term prediction lag value is outside the upper and lower limits; and

retaining the first long-term prediction lag value in the partially corrupted frame when the first long-term prediction lag value is within the upper and lower limits.

Alternatively, the method comprises the steps of:

determining whether the speech sequence in which the corrupted frame is arranged is stationary or non-stationary, based on the second long-term prediction gain values;

when the speech sequence is stationary, replacing the first long-term prediction lag value in the corrupted frame with the last long-term prediction lag value; and

when the speech sequence is non-stationary, replacing the first long-term prediction lag value in the corrupted frame with a third long-term prediction lag value determined based on the second long-term prediction lag values and an adaptively-limited random lag jitter, and replacing the first long-term prediction gain value in the corrupted frame with a third long-term prediction gain value determined based on the second long-

term prediction gain values and an adaptively-limited random gain jitter.

Preferably, the third long-term prediction lag value is calculated based at least partially on a weighted median of the second long-term prediction lag values, and the adaptively-limited random lag jitter is a value bound by limits determined based on the second long-term prediction lag values.

Preferably, the third long-term prediction gain value is calculated based at least partially on a weighted median of the second long-term prediction gain values, and the adaptively-limited random gain jitter is a value bound by limits determined based on the second long-term prediction gain values.

Alternatively, the method comprises the steps of:

determining whether the corrupted frame is partially corrupted or totally corrupted;

replacing the first long-term prediction lag value in the corrupted frame with a third lag value if the corrupted frame is totally corrupted, wherein when the speech sequence in which the totally corrupted frame is arranged is stationary, set the third lag value equal to the last long-term prediction lag value, and when said speech sequence is non-stationary, determining the third lag value based on the second long-term prediction values and an adaptively-limited random lag jitter; and

replacing the first long-term prediction lag value in the corrupted frame with a fourth lag value if the corrupted frame is partially corrupted., wherein when the speech sequence in which the partially corrupted frame is arranged in stationary, set the fourth lag value equal to the last long-term prediction lag value, and when said speech sequence is non-stationary set the fourth lag value based on a decoded long-term prediction lag value searched from an adaptive codebook associated with the non-corrupted frame preceding the corrupted frame, when said speech sequence is non-stationary.

The second aspect of the present invention is a speech signal transmitter and receiver system for encoding speech signals in an encoded bit stream and decoding the encoded bit stream into synthesized speech, wherein the encoded bit stream includes a plurality of speech frames arranged in speech sequences, and the speech frames include at least one corrupted frame preceded by one or more non-corrupted frames, wherein the corrupted frame is indicated by a first signal and includes a first long-term prediction lag value and a first long-term prediction gain value, and the non-corrupted frames include

second long-term prediction lag values and second long-term prediction gain values, and wherein the second long-term prediction lag values include a last long-term prediction lag value, and the second long-term prediction gain values include a last long-term prediction gain value, and the speech sequences include stationary and non-stationary speech sequences. The system comprises:

a first mechanism, responsive to the first signal, for determining whether the speech sequence in which the corrupted frame is arranged is stationary or non-stationary, based on the second long-term prediction gain values, and for providing a second signal indicative of whether the speech sequence is stationary or non-stationary; and

a second mechanism, responsive to the second signal, for replacing the first long-term prediction lag value in the corrupted frame with the last long-term prediction lag value when the speech sequence is stationary, and replacing the first long-term prediction lag value and the first long-term gain value in the corrupted frame with a third long-term prediction lag value and a third long-term prediction gain value, respectively, when the speech sequence is non-stationary, wherein the third long-term prediction lag value is determined based on the second long-term prediction lag values and an adaptively-limited random lag jitter, and the third long-term prediction gain value is determined based on the second long-term prediction gain values and an adaptively-limited random gain jitter.

Preferably, the third long-term prediction lag value is calculated based at least partially on a weighted median of the second long-term prediction lag values, and the adaptively-limited random lag jitter is a value bound by limits determined based on the second long-term prediction lag values.

Preferably, the third long-term prediction gain value is calculated based at least partially on a weighted median of the second long-term prediction gain values, and the adaptively-limited random gain jitter is a value bound by limits determined based on the second long-term prediction gain values.

The third aspect of the present invention is a decoder for synthesizing speech from an encoded bit stream, wherein the encoded bit stream includes a plurality of speech frames arranged in speech sequences, and the speech frames include at least one corrupted frame preceded by one or more non-corrupted frames, wherein the corrupted frame is indicated by a first signal and includes a first long-term prediction lag value and a first long-term prediction gain value, and the non-corrupted frames include second long-term

prediction lag values and second long-term prediction gain values, and wherein the second long-term prediction lag values include a last long-term prediction lag value and the second long-term prediction gain values include a last long-term prediction gain value and the speech sequences include stationary and non-stationary speech sequences. The decoder comprises:

a first mechanism, responsive to the first signal, for determining whether the speech sequence in which the corrupted frame is arranged is stationary or non-stationary, based on the second long-term prediction gain values, and for providing a second signal indicative of whether the speech sequence is stationary or non-stationary; and

a second mechanism, responsive to the second signal, for replacing the first long-term prediction lag value in the corrupted frame with the last long-term prediction lag value when the speech sequence is stationary, and replacing the first long-term prediction lag value and the first long-term gain value in the corrupted frame with a third long-term prediction lag value and a third long-term prediction gain value, respectively, when the speech sequence is non-stationary, wherein the third long-term prediction lag value is determined based on the second long-term prediction lag values and an adaptively-limited random lag jitter, and the third long-term prediction gain value is determined based on the second long-term prediction gain values and an adaptively-limited random gain jitter.

The fourth aspect of the present invention is a mobile station, which is arranged to receive an encoded bit stream containing speech data indicative of speech signals, wherein the encoded bit stream includes a plurality of speech frames arranged in speech sequences, and the speech frames include at least one corrupted frame preceded by one or more non-corrupted frames, wherein the corrupted frame is indicated by a first signal and includes a first long-term prediction lag value and a first long-term prediction gain value, and the non-corrupted frames include second long-term prediction lag values and second long-term prediction gain values, and wherein the second long-term prediction lag values include a last long-term prediction lag value and the second long-term prediction gain values include a last long-term prediction gain value and the speech sequences include stationary and non-stationary speech sequences. The mobile station comprises:

a first mechanism, responsive to the first signal, for determining whether the speech sequence in which the corrupted frame is arranged is stationary or non-stationary, based on the second long-term prediction gain values, and for providing a second signal

indicative of whether the speech sequence is stationary or non-stationary; and

a second mechanism, responsive to the second signal, for replacing the first long-term prediction lag value in the corrupted frame with the last long-term prediction lag value when the speech sequence is stationary, and replacing the first long-term prediction lag value and the first long-term gain value in the corrupted frame with a third long-term prediction lag value and a third long-term prediction gain value, respectively, when the speech sequence is non-stationary, wherein the third long-term prediction lag value is determined based on the second long-term prediction lag values and an adaptively-limited random lag jitter, and the third long-term prediction gain value is determined based on the second long-term prediction gain values and an adaptively-limited random gain jitter.

The fifth aspect of the present invention is an element in a telecommunication network, which is arranged to receive an encoded bit stream containing speech data from a mobile station, wherein the speech data includes a plurality of speech frames arranged in speech sequences, and the speech frames include at least one corrupted frame preceded by one or more non-corrupted frames, wherein the corrupted frame is indicated by a first signal and includes a first long-term prediction lag value and a first long-term prediction gain value, and the non-corrupted frames include second long-term prediction lag values and second long-term prediction gain values, and wherein the second long-term prediction lag values include a last long-term prediction lag value and the second long-term prediction gain values include a last long-term prediction gain value and the speech sequences include stationary and non-stationary speech sequences. The element comprises:

a first mechanism, responsive to the first signal, for determining whether the speech sequence in which the corrupted frame is arranged is stationary or non-stationary, based on the second long-term prediction gain values, and for providing a second signal indicative of whether the speech sequence is stationary or non-stationary; and

a second mechanism, responsive to the second signal, for replacing the first long-term prediction lag value in the corrupted frame with the last long-term prediction lag value when the speech sequence is stationary, and replacing the first long-term prediction lag value and the first long-term gain value in the corrupted frame with a third long-term prediction lag value and a third long-term prediction gain value, respectively, when the speech sequence is non-stationary, wherein the third long-term prediction lag value is

determined based on the second long-term prediction lag values and an adaptively-limited random lag jitter, and the third long-term prediction gain value is determined based on the second long-term prediction gain values and an adaptively-limited random gain jitter.

The present invention will become apparent upon reading the description taken in conjunction with Figures 3 to 11c.

Brief Description of the Drawings

Figure 1 is a block diagram illustrating a generic distributed speech codec, wherein the encoded bit stream containing speech data is conveyed from an encoder to a decoder via a communication channel or a storage medium.

Figure 2 is a block diagram illustrating a prior-art error concealment apparatus in a receiver.

Figure 3 is a block diagram illustrating the error concealment apparatus in a receiver, according to the present invention.

Figure 4 is a flow chart illustrating the method of error concealment according to the present invention.

Figure 5 is a diagrammatic representation of a mobile station, which includes an error concealment module, according to the present invention.

Figure 6 is a diagrammatic representation of a telecommunication network using a decoder, according to the present invention.

Figure 7 is a plot of LTP-parameters illustrating the lag and gain profiles in a voiced speech sequence.

Figure 8 is a plot of LTP-parameters illustrating the lag and gain profiles in an unvoiced speech sequence.

Figure 9 is a plot of LTP-lag values in a series of sub-frames illustrating the difference between the prior-art error concealment approach and the approach according to the present invention.

Figure 10 is another plot of LTP-lag values in a series of sub-frames illustrating the difference between the prior-art error concealment approach and the approach according to the present invention.

Figure 11a is a plot of speech signals illustrating an error-free speech sequence having the location of the bad frame of the speech channel, as shown in Figures 11b and

11c.

Figure 11b is a plot of speech signals illustrating the concealment of parameters in a bad frame according to the prior art approach.

Figure 11c is a plot of speech signals illustrating the concealment of parameters in a bad frame according to the present invention.

Best Mode for Carrying Out the Invention

Figure 3 illustrates a decoder **10**, which includes a decoding module **20** and an error concealment module **30**. The decoding module **20** receives a signal **140**, which is normally indicative of speech parameters **102** for speech synthesis. The decoding module **20** is known in the art. The error concealment module **30** is arranged to receive an encoded bit stream **100**, which includes a plurality of speech streams arranged in speech sequences. A bad-frame detection device **32** is used to detect corrupted frames in the speech sequences and provide a Bad-Frame-Indicator (BFI) signal **110** representing a BFI flag when a corrupted frame is detected. BFI is also known in the art. The BFI signal **110** is used to control two switches **40** and **42**. Normally, the speech frames are not corrupted and the BFI flag is 0. The terminal **S** is operatively connected to the terminal **0** in the switches **40** and **42**. The speech parameters **102** are conveyed to a buffer, or "parameter history" storage, **50** and the decoding module **20** for speech synthesis. When a bad frame is detected by the bad-frame detection device **32**, the BFI flag is set to 1. The terminal **S** is connected to the terminal **1** in the switches **40** and **42**. Accordingly, the speech parameters **102** are provided to an analyzer **70**, and the speech parameters needed for speech synthesis are provided by a parameter concealment module **60** to the decoding module **20**. The speech parameters **102** typically include LPC parameters for short term prediction, excitation parameters, a long-term prediction (LTP) lag parameter, an LTP gain parameter and other gain parameters. The parameter history storage **50** is used to store the LTP-lag and LTP-gain of a number of non-corrupted speech frames. The contents of the parameter history storage **50** are constantly updated so that the last LTP-gain parameter and the last LTP-lag parameter stored in the storage **50** are those of the last non-corrupted speech frame. When a corrupted frame in a speech sequence is received in the decoder **10**, the BFI flag is set to 1 and the speech parameters **102** of the corrupted frame are conveyed to the analyzer **70** through the switch **40**. By comparing the LTP-gain

parameter in the corrupted frame and the LTP-gain parameters stored in the storage **50**, it is possible for the analyzer **70** to determine whether the speech sequence is stationary or non-stationary, based on the magnitude and its variation in the LTP-gain parameters in neighboring frames. Typically, in a stationary sequence, the LTP-gain parameters are high and reasonably stable, the LTP-lag value is stable and the variation in adjacent LTP-lag values is small, as shown in Figure 7. In contrast, in an non-stationary sequence, the LTP-gain parameters are low and unstable, and the LTP-lag is also unstable, as shown in Figure 8. The LTP-lag values are changing more or less randomly. Figure 7 shows the speech sequence for the word "*viinii*". Figure 8 shows the speech sequence for the word "*exhibition*".

If the speech sequence that includes the corrupted frame is voiced or stationary, the last good LTP-lag is retrieved from the storage **50** and conveyed to the parameter concealment module **60**. The retrieved good LTP-lag is used to replace the LTP-lag of the corrupted frame. Because the LTP-lag in a stationary speech sequence is stable and its variation is small, it is reasonable to use a previous LTP-lag with small modification to conceal the corresponding parameter in corrupted frame. Subsequently, an RX signal **104** causes the replacement parameters, as denoted by reference numeral **134**, to be conveyed to the decoding module **20** through the switch **42**.

If the speech sequence that includes the corrupted frame is unvoiced or non-stationary, the analyzer **70** calculates a replacement LTP-lag value and a replacement LTP-gain value for parameter concealment. Because LTP-lag in an non-stationary speech sequence is unstable and its variation in adjacent frames is typically very large, parameter concealment should allow the LTP-lag in an error-concealed non-stationary sequence to fluctuate in a random fashion. If the parameters in the corrupted frame are totally corrupted, such as in a lost frame, the replacement LTP-lag is calculated by using a weighted median of the previous good LTP-lag values along with an adaptively-limited random jitter. The adaptively-limited random jitter is allowed to vary within limits calculated from the history of the LTP values, so that the parameter fluctuation in an error-concealed segment is similar to the previous good section of the same speech sequence.

An exemplary rule for LTP-lag concealment is governed by a set of conditions as follows:

If

minGain > 0.5 AND *LagDif* < 10; OR

lastGain > 0.5 AND *secondLastGain* > 0.5,

then the last received good LTP-lag is used for the totally corrupted frame. Otherwise, *Update_lag*, a weighted average of the LTP-lag buffer with randomization, is used for the totally corrupted frame. *Update_lag* is calculated in a manner as described below:

The LTP-lag buffer is sorted and the three biggest buffer values are retrieved. The average of these three biggest values is referred to as the weighted average lag (*WAL*), and the difference from these biggest values is referred to as the weighted lag difference (*WLD*).

Let *RAND* be the randomization with the scale of (-*WLD/2, WLD/2*), then

Update_lag = *WAL* + *RAND* (-*WLD/2,WLD/2*),

wherein

minGain is the smallest value of the LTP-gain buffer;

LagDif is the difference between the smallest and the largest LTP-lag values;

lastGain is the last received good LTP-gain; and

secondLastGain is the second last received good LTP-gain.

If the parameters in the corrupted frame are partially corrupted, then the LTP-lag value in the corrupted frame is replaced accordingly. That the frame is partially corrupted is determined by a set of exemplary LTP-feature criteria given below:

If

(1) *LagDif*<10 AND (*minLag* – 5)< T_{lf} < (*maxLag* + 5); OR

(2) *lastGain*>0.5 AND *secondLastGain*>0.5 AND (*lastLag*–10)< T_{lf} < (*lastLag*+10); OR

(3) *minGain* < 0.4 AND *lastGain* = *minGain* AND *minLag* < T_{lf} < *maxLag*; OR

(4) *LagDif* < 70 AND *minLag* < T_{lf} < *maxLag*; OR

(5) *meanLag* < T_{lf} < *maxLag*

is true, then T_{bf} is used to replace the LTP-lag in the corrupted frame. Otherwise, the corrupted frame is treated as a totally corrupted frame, as described above. In the above conditions:

maxLag is the largest value of the LTP-lag buffer;

meanLag is the average of the LTP-lag buffer;

minLag is the smallest value of the LTP-lag buffer;

lastLag is the last received good LTP-lag value; and

T_{bf} is a decoded LTP lag which is searched, when the BFI is set, from the adaptive codebook as if the BFI is not set.

Two examples of parameter concealment are shown in Figures 9 and 10. As shown, the profile of the replacement LTP-lag values in the bad frame, according to the prior art, is rather flat, but the profile of the replacement, according to the present invention, allows some fluctuation, similar to the error-free profile. The difference between the prior art approach and the present invention is further illustrated in Figures 11b and 11c, respectively, based on the speech signals in an error-free channel, as shown in Figure 11a.

When the parameters in the corrupted frame are partially corrupted, the parameter concealment can be further optimized. In partially corrupted frames, the LTP-lags in the corrupted frames may still yield an acceptable synthesized speech segment. Accordingly to the GSM specifications, the BFI flag is set by a Cyclic Redundancy Check (CRC) mechanism or other error detection mechanisms. These error detection mechanisms detect errors in the most significant bits in the channel decoding process. Accordingly, even when only a few bits are erroneous, the error can be detected and the BFI flag is set accordingly. In the prior-art parameter concealment approach, the entire frame is discarded. As a result, information contained in the correct bits is thrown away.

Typically, in the channel decoding process, the BER per frame is a good indicator for the channel condition. When the channel condition is good, the BER per frame is small and a high percentage of the LTP-lag values in the erroneous frames are correct. For example, when the frame error rate (FER) is 0.2%, over 70% of the LTP-lag values are correct. Even when the FER reaches 3%, about 60% of the LTP-lag values are still correct. The CRC can accurately detect a bad frame and set the BFI flag accordingly.

However, the CRC does not provide an estimation of the BER in the frame. If the BFI flag is used as the only criterion for parameter concealment, then a high percentage of the correct LTP-lag values could be wasted. In order to prevent a large amount of correct LTP-lags from being thrown away, it is possible to adapt a decision criterion for parameter concealment based on the LTP history. It is also possible to use the FER, for example, as the decision criterion. If the LTP-lag meets the decision criterion, no parameter concealment is necessary. In that case, the analyzer **70** conveys the speech parameters **102**, as received through the switch **40**, to the parameter concealment module **60** which then conveys the same to the decoding module **20** through the switch **42**. If the LTP-lag does not meet that decision criterion, then the corrupted frame is further examined using the LTP-feature criteria, as described hereinabove, for parameter concealment.

In stationary speech sequences, the LTP-lag is very stable. Whether most of the LTP-lag values in a corrupted frame are correct or erroneous can be correctly predicted with high probability. Thus, it is possible to adapt a very strict criterion for parameter concealment. In non-stationary speech sequences, it may be difficult to predict whether the LTP-lag value in a corrupted frame is correct, because of the unstable nature of the LTP parameters. However, that the prediction is correct or wrong is less important in non-stationary speech than in stationary speech. While allowing erroneous LTP-lag values to be used in decoding stationary speech may cause the synthesized speech to be unrecognizable, allowing erroneous LTP-lag values to be used in decoding non-stationary speech usually only increases the audible artifacts. Thus, the decision criterion for parameter concealment in non-stationary speech can be relatively lax.

As mentioned earlier, the LTP-gain fluctuates greatly in non-stationary speech. If the same LTP-gain value from the last good frame is used repeatedly to replace the LTP-gain value of one or more corrupted frames in a speech sequence, the LTP-gain profile in the gain concealed segment will be flat (similar to the prior-art LTP-lag replacement, as shown in Figures 7 and 8), in stark contrast to the fluctuating profile of the non-corrupted frames. The sudden change in the LTP-gain profile may cause unpleasant audible artifacts. In order to minimize these audible artifacts, it is possible to allow the replacement LTP-gain value to fluctuate in the error-concealed segment. For this purpose, the analyzer **70** can be also used to determine the limits between which the

replacement LTP-gain value is allowed to fluctuate based on the gain values in the LTP history.

LTP-gain concealment can be carried out in a manner as described below. When the BFI is set, a replacement LTP-gain value is calculated according to a set of LTP-gain concealment rules. The replacement LTP-gain is denoted as *Updated_gain*.

(1) If *gainDif*>0.5 AND *lastGain* = *maxGain* > 0.9 AND *subBF*=1, then
Updated_gain = (*secondLastGain* + *thirdLastGain*)/2;

(2) If *gainDif*>0.5 AND *lastGain* = *maxGain* > 0.9 AND *subBF*=2, then
Updated_gain = *meanGain* + *randVar** (*maxGain* - *meanGain*);

(3) If *gainDif*>0.5 AND *lastGain* = *maxGain* > 0.9 AND *subBF*=3, then
Updated_gain = *meanGain* – *randVar** (*meanGain* - *minGain*);

(4) If *gainDif*>0.5 AND *lastGain* = *maxGain* > 0.9 AND *subBF*=4, then
Updated_gain = *meanGain* + *randVar** (*maxGain* - *meanGain*);

In the previous conditions, *Updated_gain* cannot be larger than *lastGain*. If the previous conditions cannot be met, the following conditions are used:

(5) If *gainDif* > 0.5, then
Updated_gain = *lastGain*;

(6) If *gainDif* < 0.5 AND *lastGain* = *maxGain*, then
Updated_gain = *meanGain*;

(7) If *gainDIF* < 0.5, then
Updated_gain = *lastGain*,

Wherein

meanGain is the average of the LTP-gain buffer;
maxGain is the largest value of the LTP-gain buffer;
minGain is the smallest value of the LTP-gain buffer;
randVar is a random value between 0 and 1,
gainDIF is the difference between the smallest and the largest LTP-gain values in the LTP-gain buffer;
lastGain is the last received good LTP-gain;

seconLastGain is the second last received good LTP-gain;
thirdLastGain is the third last received good LTP-gain; and
subBF is the order of the subframe.

Figure 4 illustrates the method of error-concealment, according to the present invention. As the encoded bit stream is received at step **160**, the frame is checked to see if it is corrupted at step **162**. If the frame is not corrupted, then the parameter history of the speech sequence is updated at step **164**, and the speech parameters of the current frame are decoded at step **166**. The procedure then goes back to step **162**. If the frame is bad or corrupted, the parameters are retrieved from the parameter history storage at step **170**. Whether the corrupted frame is part of the stationary speech sequence or non-stationary speech sequence is determined at step **172**. If the speech sequence is stationary, the LTP-lag of the last good frame is used to replace the LTP-lag in the corrupted frame at step **174**. If the speech sequence is non-stationary, a new lag value and new gain value are calculated based on the LTP history at step **180**, and they are used to replace the corresponding parameters in the corrupted frame at step **182**.

Figure 5 shows a block diagram of a mobile station **200** according to one exemplary embodiment of the invention. The mobile station comprises parts typical of the device, such as a microphone **201**, keypad **207**, display **206**, earphone **214**, transmit/receive switch **208**, antenna **209** and control unit **205**. In addition, the figure shows transmitter and receiver blocks **204, 211** typical of a mobile station. The transmitter block **204** comprises a coder **221** for coding the speech signal. The transmitter block **204** also comprises operations required for channel coding, deciphering and modulation as well as RF functions, which have not been drawn in Figure 5 for clarity. The receiver block **211** also comprises a decoding block **220** according to the invention. Decoding block **220** comprises an error concealment module **222** like the parameter concealment module **30** shown in Figure 3. The signal coming from the microphone **201**, amplified at the amplification stage **202** and digitized in the A/D converter, is taken to the transmitter block **204**, typically to the speech coding device comprised by the transmit block. The transmission signal, which is processed, modulated and amplified by the transmit block, is taken via the transmit/receive switch **208** to the antenna **209**. The signal to be received is taken from the antenna via the transmit/receive switch **208** to the receiver

block **211**, which demodulates the received signal and decodes the deciphering and the channel coding. The resulting speech signal is taken via the D/A converter **212** to an amplifier **213** and further to an earphone **214**. The control unit **205** controls the operation of the mobile station **200**, reads the control commands given by the user from the keypad **207** and gives messages to the user by means of the display **206**.

The parameter concealment module **30**, according to the invention, can also be used in a telecommunication network **300**, such as an ordinary telephone network, or a mobile station network, such as the GSM network. Figure 6 shows an example of a block diagram of such a telecommunication network. For example, the telecommunication network **300** can comprise telephone exchanges or corresponding switching systems **360**, to which ordinary telephones **370**, base stations **340**, base station controllers **350** and other central devices **355** of telecommunication networks are coupled. Mobile stations **330** can establish connection to the telecommunication network via the base stations **340**. A decoding block **320**, which includes an error concealment module **322** similar to the error concealment module **30** shown in Figure 3, can be particularly advantageously placed in the base station **340**, for example. However, the decoding block **320** can also be placed in the base station controller **350** or other central or switching device **355**, for example. If the mobile station system uses separate transcoders, for example, between the base stations and the base station controllers, for transforming the coded signal taken over the radio channel into a typical 64 kbit/s signal transferred in a telecommunication system and vice versa, the decoding block **320** can also be placed in such a transcoder. In general, the decoding block **320**, including the parameter concealment module **322**, can be placed in any element of the telecommunication network **300**, which transforms the coded data stream into an uncoded data stream. The decoding block **320** decodes and filters the coded speech signal coming from the mobile station **330**, whereafter the speech signal can be transferred in the usual manner as uncompressed forward in the telecommunication network **300**.

It should be noted that the error concealment method of the present invention has been described with respect to stationary and non-stationary speech sequences, and that stationary speech sequences are usually voiced and non-stationary speech sequences are usually unvoiced. Thus, it will be understood that the disclosed method is applicable to error concealment in voiced and unvoiced speech sequences.

The present invention is applicable to CELP type speech codecs and can be adapted to other types of speech codecs as well. Thus, although the invention has been described with respect to a preferred embodiment thereof, it will be understood by those skilled in the art that the foregoing and various other changes, omissions and deviations in the form and detail thereof may be made without departing from the spirit and scope of this invention.

What is claimed is:

1. A method for concealing errors in an encoded bit stream indicative of speech signals received in a speech decoder, wherein the encoded bit stream includes a plurality of speech frames arranged in speech sequences, and the speech frames include at least one partially corrupted frame preceded by one or more non-corrupted frames, wherein the partially corrupted frame includes a first long-term prediction lag value and a first long-term prediction gain value, and the non-corrupted frames include second long-term prediction lag values and second long-term prediction gain values, and wherein the second long-term prediction lag values include a last long-term prediction lag value, and the second long-term prediction gain values include a last long-term prediction gain value, said method comprising the steps of:

providing an upper limit and a lower limit based on the second long-term prediction lag values;

determining whether the first long-term prediction lag value is within or outside the upper and lower limits;

replacing the first long-term prediction lag value in the partially corrupted frame with a third lag value, when the first long-term prediction lag value is outside the upper and lower limits; and

retaining the first long-term prediction lag value in the partially corrupted frame when the first long-term prediction lag value is within the upper and lower limits.

2. The method of claim 1, further comprising the step of replacing the first long-term prediction gain value in the partially corrupted frame with a third gain value, when the first long-term lag value is outside the upper and lower limits.

3. The method of claim 1, wherein the third lag value is calculated based the second long-term prediction lag values and an adaptively-limited random lag jitter bound by further limits determined based on the second long-term prediction lag values.

4. The method of claim 2, wherein the third gain value is calculated based on of the second long-term prediction gain values and an adaptively-limited random gain jitter bound by limits determined based on the second long-term prediction gain values.

5. A method for concealing errors in an encoded bit stream indicative of speech signals received in a speech decoder, wherein the encoded bit stream includes a plurality of speech frames arranged in speech sequences, and the speech frames include at least one corrupted frame preceded by one or more non-corrupted frames, wherein the corrupted frame includes a first long-term prediction lag value and a first long-term prediction gain value, and the non-corrupted frames include second long-term prediction lag values and second long-term prediction gain values, and wherein the second long-term prediction lag values include a last long-term prediction lag value, and the second long-term prediction gain values include a last long-term prediction gain value and the speech sequences include stationary and non-stationary speech sequences, and wherein the corrupted frame can be a totally corrupted frame or a partially corrupted frame, said method comprising the steps of:

determining whether the corrupted frame is partially corrupted or totally corrupted;

replacing the first long-term prediction lag value in the corrupted frame with a third lag value if the corrupted frame is totally corrupted; and

replacing the first long-term prediction lag value in the corrupted frame with a fourth lag value if the corrupted frame is partially corrupted.

6. The method of claim 5, further comprising the steps of:

determining whether the speech sequence in which the partially corrupted frame is arranged is stationary or non-stationary;

setting the fourth lag value equal to the last long-term prediction lag value, when said speech sequence is stationary ; and

determining the fourth lag value based on a decoded long-term prediction lag value searched from an adaptive codebook associated with the non-corrupted frame preceding the corrupted frame, when said speech sequence is non-stationary.

7. The method of claim 5, further comprising the steps of:

determining whether the speech sequence in which the totally corrupted frame is arranged is stationary or non-stationary;

setting the third lag value equal to the last long-term prediction lag value, when said speech sequence is stationary; and

determining the third lag value based on the second long-term prediction values and an adaptively-limited random lag jitter, when said speech sequence is non-stationary.

8. The method of claim 6, wherein the second long-term prediction lag values further include a second last long-term prediction lag value and a third last long-term prediction lag value, and the second long-term prediction gain values further include a second last long-term prediction gain value and a third last long-term prediction gain value, said method further comprising the steps of:

determining minLag, which is the smallest lag value among the second long-term prediction lag values;

determining maxLag, which is the largest lag value among the second long-term prediction lag values;

determining meanLag, which is an average of the second long-term prediction lag values;

determining difLag, which is the difference of maxLag and minLag;

determining minGain, which is the smallest gain value among the second long-term prediction gain values;

determining maxGain, which is the largest gain value among the second long-term prediction gain values; and

determining meanGain, which is an average of the second long term gain values; wherein

if difLag < 10, and (minLag - 5) < the fourth lag value < (maxLag + 5); or

if the last long-term prediction gain value is larger than 0.5, and the second last long-term prediction gain value is larger than 0.5, and the fourth lag value is smaller than a sum of the last long-term prediction value and 10, and a sum of the fourth lag value and 10 is larger than the last long-term prediction value; or

if minGain < 0.4, and the last long-term prediction gain value is equal to minGain, and the fourth lag value is larger than minLag but smaller than maxLag; or

if difLag < 70, and the fourth lag value is larger than minLag but smaller than maxLag; or

if the fourth lag value is larger than meanLag but smaller than maxLag; then the corrupted frame is determined as partially corrupted.

9. The method of claim 6, wherein when said speech sequence is non-stationary, said method further comprising the step of determining a frame-error rate of the speech frames such that

if the frame-error rate reaches a determined value, the fourth lag value is determined based on said decoded long-term prediction lag value, and

if the frame-error rate is smaller than the determined value, the fourth lag value is set equal to the last long-term prediction lag value.

10. The method of claim 5, wherein the stationary speech sequences include voiced sequences, and the non-stationary speech sequences include unvoiced sequences.

11. A speech signal transmitter and receiver system for encoding speech signals in an encoded bit stream and decoding the encoded bit stream into synthesized speech, wherein the encoded bit stream includes a plurality of speech frames arranged in speech sequences, and the speech frames include at least one corrupted frame preceded by one or more non-corrupted frames, wherein the corrupted frame includes frame a first long-term prediction lag value and a first long-term prediction gain value, and the non-corrupted frames include second long-term prediction lag values and second long-term prediction gain values, and wherein the second long-term prediction lag values include a last long-term prediction lag value and the second long-term prediction gain values include a last long-term prediction gain value, and the speech sequences include stationary and non-stationary speech sequences, and a first signal is used to indicate the corrupted frame, said system comprising:

a first means, responsive to the first signal, for determining whether the speech sequence in which the corrupted frame is arranged is stationary or non-stationary, and for providing a second signal indicative of said determining;

a second means, responsive to the second signal, for replacing the first long-term prediction lag value in the corrupted frame with the last long-term prediction lag value when said speech sequence is stationary, and replacing the first long-term prediction lag

value in the corrupted frame with a third lag value when said speech sequence is non-stationary.

12. The system of claim 11, wherein the third lag value is determined based on the second long-term prediction lag values and an adaptively-limited random lag jitter.

13. The system of claim 11, wherein the second means further replaces the first long-term prediction gain value in the corrupted frame with a third gain value when said speech sequence is non-stationary.

14. The system of claim 13, wherein the third gain value is determined based on the second long-term prediction gain values and an adaptively-limited random gain jitter.

15. The system of claim 11, wherein the stationary speech sequences include voiced sequences, and the non-stationary speech sequences include unvoiced sequences.

16. A decoder for synthesizing speech from an encoded bit stream, wherein the encoded bit stream includes a plurality of speech frames arranged in speech sequences, and the speech frames include at least one corrupted frame preceded by one or more non-corrupted frames, wherein the corrupted frame includes a first long-term prediction lag value and a first long-term prediction gain value, and the non-corrupted frames include second long-term prediction lag values and second long-term prediction gain values, and wherein the second long-term prediction lag values include a last long-term prediction lag value and the second long-term prediction gain values include a last long-term prediction gain value and the speech sequences include stationary and non-stationary speech sequences, and a first signal is used to indicate the corrupted frame, said decoder comprising:

a first means, responsive to the first signal, for determining whether the speech sequence in which the corrupted frame is arranged is stationary or non-stationary, and for providing a second signal indicative of said determining;

a second means, responsive to the second signal, for replacing the first long-term prediction lag value in the corrupted frame with the last long-term prediction lag value when said speech sequence is stationary, and replacing the first long-term prediction lag

value in the corrupted frame with a third lag value when said speech sequence is non-stationary.

17. The decoder of claim 16, wherein the lag value is determined based on the second long-term prediction lag values and an adaptively-limited random lag jitter.

18. The decoder of claim 16, wherein the second means further replaces the first long-term gain value in the corrupted frame with a third gain value when said speech sequence is non-stationary.

19. The decoder of claim 18, wherein the third gain value is determined based on the second long-term prediction gain values and an adaptively-limited random gain jitter.

20. The decoder of claim 16, wherein the stationary speech sequences include voiced sequences, and the non-stationary speech sequences include unvoiced sequences.

21. A mobile station, which is arranged to receive an encoded bit stream containing speech data indicative of speech signals, wherein the encoded bit stream includes a plurality of speech frames arranged in speech sequences, and the speech frames include at least one corrupted frame preceded by one or more non-corrupted frames, wherein the corrupted frame includes a first long-term prediction lag value and a first long-term prediction gain value, and the non-corrupted frames include second long-term prediction lag values and second long-term prediction gain values, and wherein the second long-term prediction lag values include a last long-term prediction lag value and the second long-term prediction gain values include a last long-term prediction gain value and the speech sequences include stationary and non-stationary speech sequences, and wherein a first signal is used to indicate the corrupted frame, said mobile station comprising:

a first means, responsive to the first signal, for determining whether the speech sequence in which the corrupted frame is arranged is stationary or non-stationary, and for providing a second signal indicative of said determining; and

a second means, responsive to the second signal, for replacing the first long-term prediction lag value in the corrupted frame with the last long-term prediction lag value

when said speech sequence is stationary, and replacing the first long-term prediction lag value in the corrupted frame with a third lag value when said speech sequence is non-stationary.

22. The mobile station of claim 21, wherein the third lag value is determined based on the second long-term prediction lag values and an adaptively-limited random lag jitter.

23. The mobile station of claim 21, wherein the second means further replaces the first long-term gain value in the corrupted frame with a third gain value when said speech sequence is non-stationary.

24. The mobile station of claim 23, wherein the third gain value is determined based on the second long-term prediction gain values and an adaptively-limited random gain jitter.

25. The mobile station of claim 21, wherein the stationary speech sequences include voiced sequences, and the non-stationary speech sequences include unvoiced sequences.

26. An element in a telecommunication network, which is arranged to receive an encoded bit stream containing speech data from a mobile station, wherein the speech data includes a plurality of speech frames arranged in speech sequences, and the speech frames include at least one corrupted frame preceded by one or more non-corrupted frames, wherein the corrupted frame includes a first long-term prediction lag value and a first long-term prediction gain value, and the non-corrupted frames include second long-term prediction lag values and second long-term prediction gain values, and wherein the second long-term prediction lag values include a last long-term prediction lag value and the second long-term prediction gain values include a last long-term prediction gain value and the speech sequences include stationary and non-stationary speech sequences, and wherein a first signal is used to indicate the corrupted frame, said element comprising:

a first means, responsive to the first signal, for determining whether the speech sequence in which the corrupted frame is arranged is stationary or non-stationary, and for providing a second signal indicative of said determining; and

a second means, responsive to the second signal, for replacing the first long-term prediction lag value in the corrupted frame with the last long-term prediction lag value when said speech sequence is stationary, and replacing the first long-term prediction lag value in the corrupted frame with a third lag value when said speech sequence is non-stationary.

27. The element of claim 26, wherein the third long-term prediction lag value is determined based on the second long-term prediction lag values and an adaptively-limited random lag jitter.

28. The element of claim 26, wherein the third means further replaces the first long-term prediction gain value with a third gain value when said speech sequence is non-stationary.

29. The element of claim 28, wherein the third gain value is determined based on the second long-term prediction gain values and an adaptively-limited random gain jitter.

30. The element of claim 26, wherein the stationary speech sequences include voiced sequences, and the non-stationary speech sequences include unvoiced sequences.

31. (New) The method of claim 5, wherein the second long-term prediction gain values further include a second last long-term prediction gain value, and

if $difLag < 10$, and $(minLag - 5) < decodedLag < (maxLag + 5)$; or
if $lastGain > 0.5$, and $secondlastGain > 0.5$, and
$(lastLag - 10) < decodedLag < (lastLag + 10)$; or
if $minGain < 0.4$, and $lastGain > 0.5$, and $minLag < decodedLag < maxLag$; or
if $difLag < 70$, and $minLag < decodedLag < maxLag$; or
if $meanLag < decodedLag < maxLag$,

then the fourth value is set equal to the *decodedLag*, wherein

minLag is a smallest lag value among the second long-term prediction lag values,

maxLag is a largest lag value among the second long-term prediction lag values,

meanLag is an average of the second long-term prediction lag values,

difLag is a difference of *maxLag* and *minLag*,

minGain is a smallest gain value among the second long-term prediction gain values,

meanGain an average of the second long-term prediction gain values,

lastGain is the last long-term prediction gain value,

lastLag is the last long-term prediction lag value,

secondlastGain is the second last long-term prediction lag value, and

decodedLag is a decoded long-term prediction lag which is searched from an adaptive codebook associated with the non-corrupted frame preceding the corrupted frame.

32. (New) The method of claim 8, wherein the first long-term prediction gain value is replaced by *Updated_gain*, and wherein

If *gainDif* >0.5 AND *lastGain* = *maxGain* > 0.9 AND *subBF*=1, then

Updated_gain = (*secondLastGain* + *thirdLastGain*)/2;

If *gainDif* >0.5 AND *lastGain* = *maxGain* > 0.9 AND *subBF*=2, then

Updated_gain = *meanGain* + *randVar** (*maxGain* - *meanGain*);

If *gainDif* >0.5 AND *lastGain* = *maxGain* > 0.9 AND *subBF*=3, then

Updated_gain = *meanGain* – *randVar** (*meanGain* - *minGain*);

If *gainDif* >0.5 AND *lastGain* = *maxGain* > 0.9 AND *subBF*=4, then

Updated_gain = *meanGain* + *randVar** (*maxGain* - *meanGain*);

and when *Updated_gain* is equal to or smaller than *lastGain* ;

or

If *gainDif* > 0.5, then

Updated_gain = *lastGain*;

(8) If *gainDif* < 0.5 AND *lastGain* = *maxGain*, then

Updated_gain = *meanGain*;

(9) If *gainDIF* < 0.5, then

Updated_gain = *lastGain*,

and when *Updated_gain* is larger than *lastGain*,

wherein

randVar is a random value between 0 and 1,

gainDIF is the difference between a smallest and a largest long-term prediction gain value;

lastGain is the last long-term prediction gain value;

secondLastGain is the second last long-term prediction gain value;

thirdLastGain is the third last long-term prediction gain value; and

subBF is an order of the subframe.


INPUT SIGNAL
ENCODER
BIT STREAM
COMM. CHANNEL OR STORAGE MEDIUM
BIT STREAM
DECODER
SYNTHESIZED SIGNAL

FIG. 1

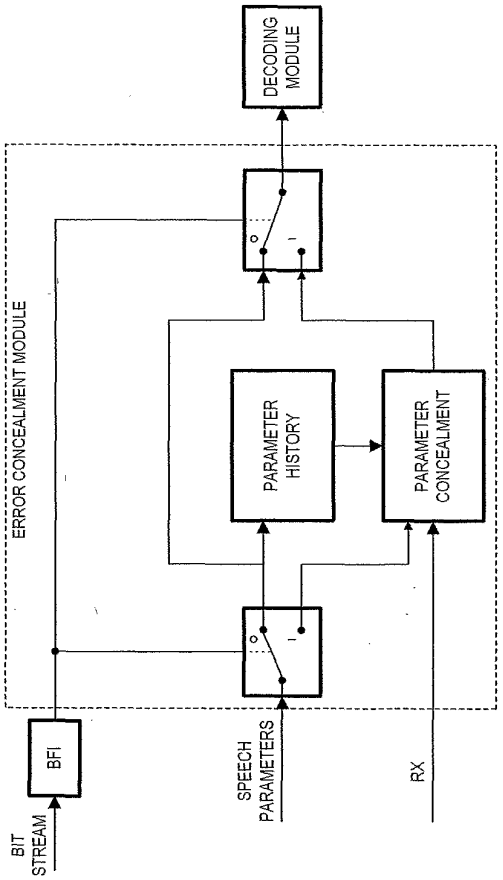

DECODING MODULE
ERROR CONCEALMENT MODULE
PARAMETER HISTORY
PARAMETER CONCEALMENT
BFI
BIT STREAM
SPEECH PARAMETERS
RX

FIG. 2

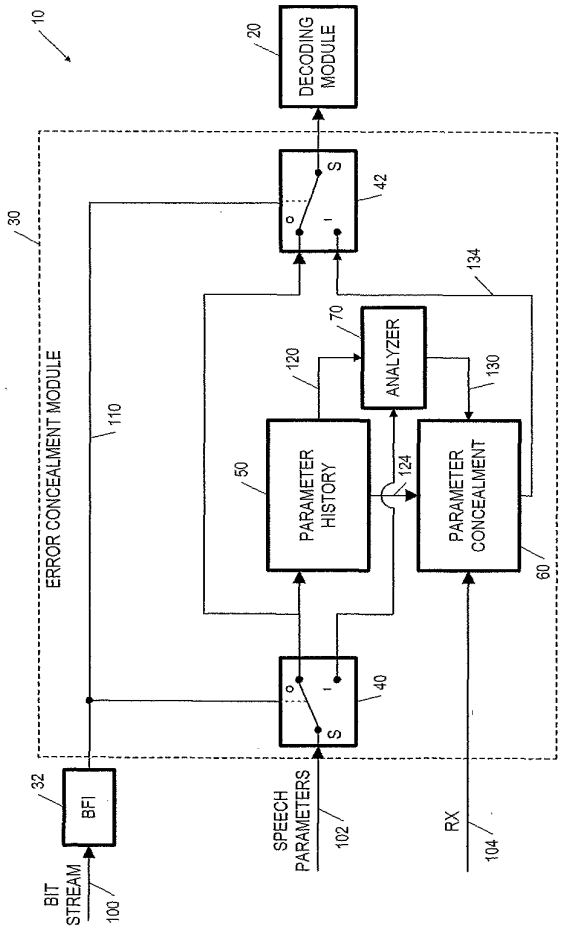

10
20
DECODING MODULE
30
ERROR CONCEALMENT MODULE
42
S
134
70
ANALYZER
120
130
110
50
PARAMETER HISTORY
124
PARAMETER CONCEALMENT
60
40
32
BFI
BIT STREAM
100
SPEECH PARAMETERS
102
RX
104

FIG. 3

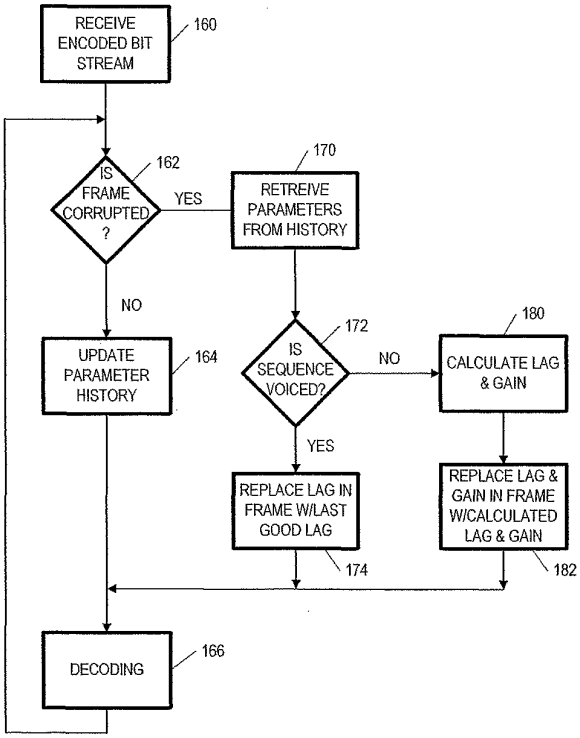

RECEIVE ENCODED BIT STREAM
160
IS FRAME CORRUPTED ?
162
YES
NO
RETREIVE PARAMETERS FROM HISTORY
170
UPDATE PARAMETER HISTORY
164
IS SEQUENCE VOICED?
172
NO
YES
CALCULATE LAG & GAIN
180
REPLACE LAG IN FRAME W/LAST GOOD LAG
174
REPLACE LAG & GAIN IN FRAME W/CALCULATED LAG & GAIN
182
DECODING
166

FIG. 4

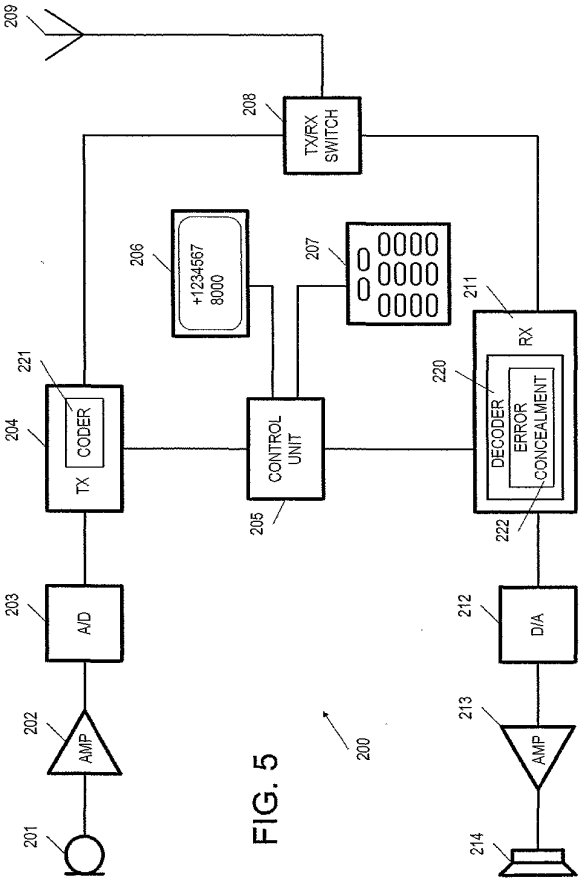

209
208
TX/RX SWITCH
206
+1234567 8000
207
0000 0000 0000
211
RX
221
CODER
204
TX
CONTROL UNIT
205
220
DECODER
ERROR CONCEALMENT
222
203
A/D
212
D/A
202
AMP
213
AMP
201
214
200

FIG. 5

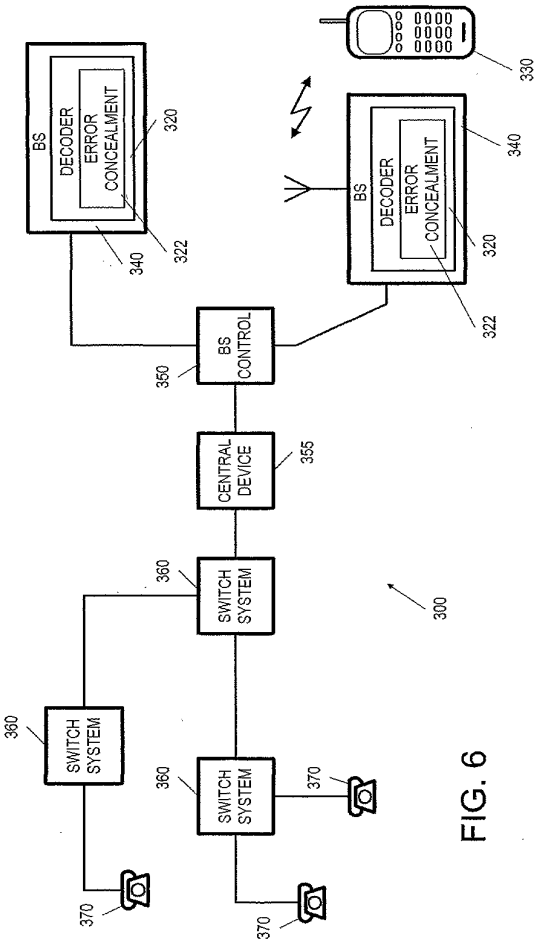

BS
DECODER
ERROR
CONCEALMENT
320
322
340
330
BS
DECODER
ERROR
CONCEALMENT
340
320
322
BS
CONTROL
350
CENTRAL
DEVICE
355
SWITCH
SYSTEM
360
300
SWITCH
SYSTEM
360
SWITCH
SYSTEM
360
370
370
370

FIG. 6

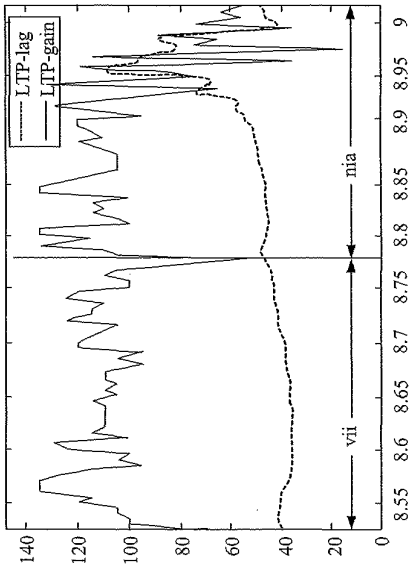

LTP-lag
LTP-gain
vii
nia
Time
0
20
40
60
80
100
120
140
8.55
8.6
8.65
8.7
8.75
8.8
8.85
8.9
8.95
9

FIG. 7

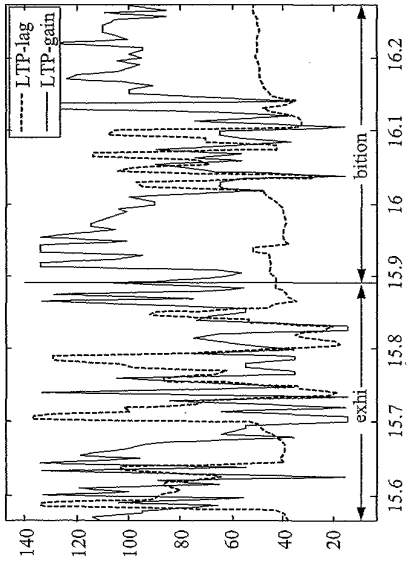

LTP-lag
LTP-gain
exhi
bition
Time
15.6
15.7
15.8
15.9
16
16.1
16.2
140
120
100
80
60
40
20

FIG. 8

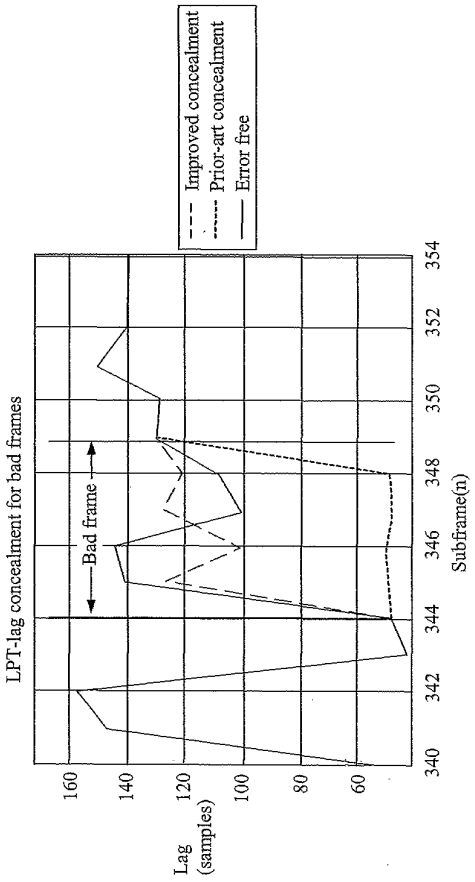

Improved concealment
Prior-art concealment
Error free
LPT-lag concealment for bad frames
Bad frame
Lag (samples)
160
140
120
100
80
60
340
342
344
346
348
350
352
354
Subframe(n)

FIG. 9

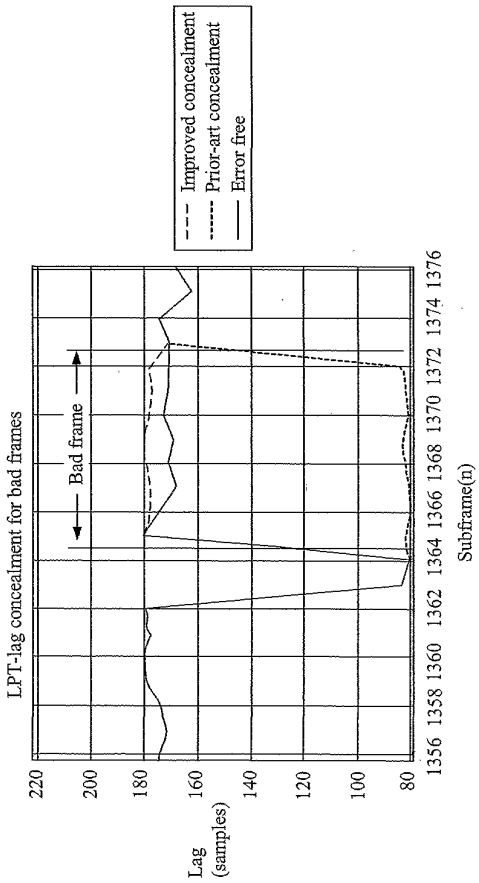

Improved concealment
Prior-art concealment
Error free
LPT-lag concealment for bad frames
Bad frame
Lag (samples)
220
200
180
160
140
120
100
80
1356 1358 1360 1362 1364 1366 1368 1370 1372 1374 1376
Subframe(n)

FIG. 10

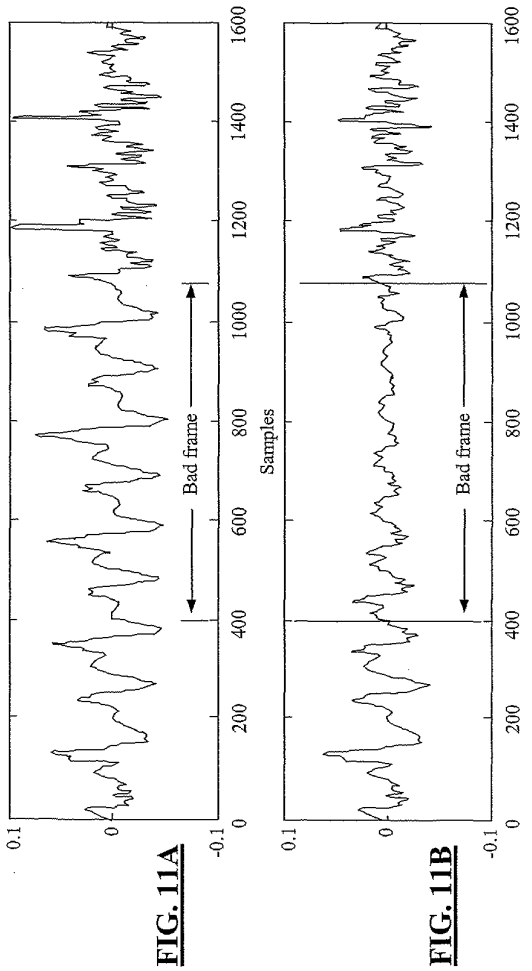


FIG. 11A

FIG. 11B

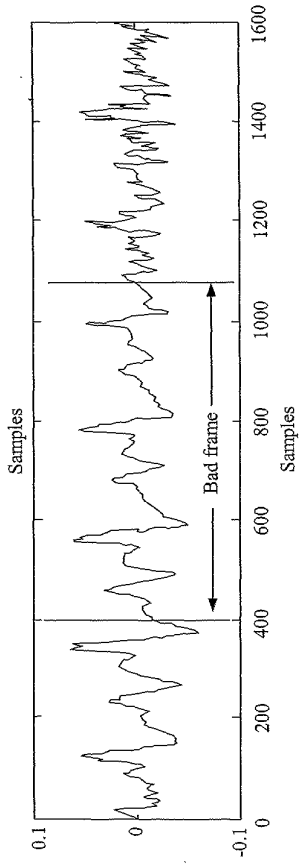

Samples
Bad frame
Samples
0.1
0
-0.1
0
200
400
600
800
1000
1200
1400
1600

FIG. 11C

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 01/02021

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L19/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	J. MAKINEN, J. VAINIO, H. MIKKOLA AND J. ROTTOLA-PUKKILA: "Improved substitution for erroneous LTP-parameters in a speech decoder" NORSIG SYMPOSIUM 2001, 18 - 20 October 2001, XP002195905 Trondheim the whole document -/--	1-30

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 April 2002	25/04/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Quélavoine, R

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 01/02021

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	TSG-SA CODEC WORKING GROUP: "3G TS 26.091" TECHNICAL SPECIFICATION GROUP SERVICES AND SYSTEM ASPECTS, 26 - 28 April 1999, XP002195906 Yokohama * section 6.2.2, the two first lines * * section 6.2.3.2 LTP-lag update * * section 7.2.2, the two first lines * * section 7.2.3.1 LTP-lag update *	1,5,11, 16,21,26
A	US 6 188 980 B1 (THYSSEN JES) 13 February 2001 (2001-02-13) abstract column 2, line 47-63 column 10, line 17-20 column 13, line 5-7 column 14, line 21-29 column 41, line 19-22 & WO 00 11651 A 2 March 2000 (2000-03-02)	1,5,11, 16,21,26

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/IB 01/02021

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 6188980	B1	13-02-2001	WO	0011651 A1	02-03-2000
			EP	1105871 A1	13-06-2001
			EP	1105872 A1	13-06-2001
			EP	1105870 A1	13-06-2001
			EP	1110209 A1	27-06-2001
			EP	1194924 A1	10-04-2002
			TW	448417 B	01-08-2001
			TW	440813 B	16-06-2001
			TW	440814 B	16-06-2001
			TW	448418 B	01-08-2001
			WO	0011658 A1	02-03-2000
			WO	0011652 A1	02-03-2000
			WO	0011655 A1	02-03-2000
			WO	0011659 A1	02-03-2000
			WO	0011648 A1	02-03-2000
			WO	0011653 A1	02-03-2000
			WO	0011649 A1	02-03-2000
			WO	0011656 A1	02-03-2000
			WO	0011660 A1	02-03-2000
			WO	0011650 A1	02-03-2000
			WO	0011661 A1	02-03-2000
			WO	0011657 A1	02-03-2000
			WO	0011654 A1	02-03-2000
			US	2001023395 A1	20-09-2001
			US	6104992 A	15-08-2000
			US	6330531 B1	11-12-2001
			US	6260010 B1	10-07-2001
			US	6173257 B1	09-01-2001
			US	6240386 B1	29-05-2001

Form PCT/ISA/210 (patent family annex) (July 1992)

フロントページの続き

(51)Int.Cl.[7]	FI		テーマコード（参考）
Ｈ０４Ｌ　1/00	Ｈ０４Ｌ	1/00	Ｂ
	Ｇ１０Ｌ	3/00	Ｆ
	Ｇ１０Ｌ	9/14	Ｎ

(81)指定国　AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,PH,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,UZ,VN,YU,ZA,ZW

(72)発明者　バイノ、ヤッネ
フィンランド共和国、フィン－３３８８０　レムペーレ、ラウリンチエ　１６　セー
(72)発明者　ロトラープッキラ、ヤニ
フィンランド共和国、フィン－３３８２０　タムペレ、レーベンカツ　２４　エー　４４

Ｆターム(参考)　5J064 AA01 BB01 BB03 BB08 BB12 BC25 BD02
5K014 AA01 BA06 EA08 GA02