

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6190052号
(P6190052)

(45) 発行日 平成29年8月30日 (2017. 8. 30)

(24) 登録日 平成29年8月10日 (2017. 8. 10)

(51) Int. Cl.	F I
G 1 O L 19/09 (2013. 01)	G 1 O L 19/09
G 1 O L 19/005 (2013. 01)	G 1 O L 19/005

請求項の数 13 (全 77 頁)

(21) 出願番号	特願2016-520530 (P2016-520530)	(73) 特許権者	597159765
(86) (22) 出願日	平成26年6月23日 (2014. 6. 23)		フラウンホーファーゲゼルシャフト ツー
(65) 公表番号	特表2016-532143 (P2016-532143A)		ル フォルデルング デル アンゲヴァン
(43) 公表日	平成28年10月13日 (2016. 10. 13)		テン フォルシユング エー. フアー.
(86) 国際出願番号	PCT/EP2014/063176		ドイツ国, デー-80686 ミュンヘン
(87) 国際公開番号	W02014/202789		, ハンサシュトラーセ 27 ツェー
(87) 国際公開日	平成26年12月24日 (2014. 12. 24)	(74) 代理人	100112715
審査請求日	平成28年2月18日 (2016. 2. 18)		弁理士 松山 隆夫
(31) 優先権主張番号	13173154.9	(74) 代理人	100104444
(32) 優先日	平成25年6月21日 (2013. 6. 21)		弁理士 上羽 秀敏
(33) 優先権主張国	欧州特許庁 (EP)	(74) 代理人	100125704
(31) 優先権主張番号	14166998.6		弁理士 坂根 剛
(32) 優先日	平成26年5月5日 (2014. 5. 5)	(74) 代理人	100120662
(33) 優先権主張国	欧州特許庁 (EP)		弁理士 川上 桂子

最終頁に続く

(54) 【発明の名称】 T C X L T P についての向上した概念を実現する装置及び方法

(57) 【特許請求の範囲】

【請求項 1】

符号化音声信号を復号して再構成音声信号を得るための装置であって、
 当該装置は、複数のフレームを受信するように構成され、
 逆修正離散余弦変換を行うことによって前記複数のフレームを復号して復号された音声
 信号の音声信号サンプルを得るための逆修正離散余弦変換モジュール（683）と、
 長期予測を行うための長期予測部（684）とを備え、
 前記長期予測部（684）は、
 前記復号された音声信号の前記音声信号サンプルを記憶するための遅延バッファ（10
 20）と、
 前記遅延バッファ（1020）に記憶されている前記復号された音声信号の音声信号サ
 ンプルから複数の選択音声信号サンプルを選択するためのサンプル選択部（1030）と
 、
 前記複数の選択音声信号サンプルを処理して前記再構成音声信号の再構成音声信号サン
 プルを得るためのサンプル処理部（1040）とを含み、
 前記サンプル選択部（1030）は、現在のフレームが当該装置によって受信されてお
 り、且つ、当該装置によって受信されている前記現在のフレームが破損していない場合、
 前記現在のフレームに含まれるピッチ遅延情報に応じて、前記遅延バッファ（1020）
 に記憶されている前記音声信号サンプルから前記複数の選択音声信号サンプルを選択す
 るように構成され、

前記サンプル選択部（１０３０）は、前記現在のフレームが当該装置によって受信されていない場合、又は当該装置によって受信されている前記現在のフレームが破損している場合、当該装置によって以前に受信されている別のフレームに含まれるピッチ遅延情報に応じて、前記遅延バッファ（１０２０）に記憶されている前記音声信号サンプルから前記複数の選択音声信号サンプルを選択するように構成され、

前記サンプル選択部（１０３０）は、変更後ゲインに応じて前記複数の選択音声信号サンプルを再スケーリングすることによって前記再構成音声信号サンプルを得るように構成され、

前記変更後ゲインは、

【数４６】

$gain = gain_past * damping;$

の式によって定義され、ここで、 $gain$ は、前記変更後ゲインであり、

前記サンプル選択部（１０３０）は、 $gain$ の算出後に $gain_past$ を $gain$ に設定するように構成され、

$damping$ は、 $0 \leq damping \leq 1$ の範囲における実数値である、装置。

【請求項２】

請求項１に記載の装置であって、

前記サンプル処理部（１０４０）は、前記現在のフレームが当該装置によって受信されており、且つ、当該装置によって受信されている前記現在のフレームが破損していない場合、前記現在のフレームに含まれるゲイン情報に応じて前記選択音声信号サンプルを再スケーリングすることによって前記再構成音声信号サンプルを得るように構成され、

前記サンプル選択部（１０３０）は、前記現在のフレームが当該装置によって受信されていない場合、又は当該装置によって受信されている前記現在のフレームが破損している場合、当該装置によって以前に受信されている前記別のフレームに含まれる前記ゲイン情報に応じて前記選択音声信号サンプルを再スケーリングすることによって前記再構成音声信号サンプルを得るように構成される、装置。

【請求項３】

請求項２に記載の装置であって、

前記サンプル処理部（１０４０）は、前記現在のフレームが当該装置によって受信されており、且つ、当該装置によって受信されている前記現在のフレームが破損していない場合、前記選択音声信号サンプルと、前記現在のフレームに含まれる前記ゲイン情報に応じた値とを乗算することによって前記再構成音声信号サンプルを得るように構成され、

前記サンプル選択部（１０３０）は、前記現在のフレームが当該装置によって受信されていない場合、又は当該装置によって受信されている前記現在のフレームが破損している場合、前記選択音声信号サンプルと、当該装置によって以前に受信されている前記別のフレームに含まれる前記ゲイン情報に応じた値とを乗算することによって前記再構成音声信号サンプルを得るように構成される、装置。

【請求項４】

請求項１から請求項３のいずれか一項に記載の装置であって、前記サンプル処理部（１０４０）は、前記再構成音声信号サンプルを前記遅延バッファ（１０２０）に記憶させるように構成される、装置。

【請求項５】

請求項４に記載の装置であって、前記サンプル処理部（１０４０）は、更なるフレームが当該装置によって受信される前に前記再構成音声信号サンプルを前記遅延バッファ（１０２０）に記憶させるように構成される、装置。

【請求項６】

請求項４に記載の装置であって、前記サンプル処理部（１０４０）は、更なるフレームが当該装置によって受信された後に前記再構成音声信号サンプルを前記遅延バッファ（１０２０）に記憶させるように構成される、装置。

10

20

30

40

50

【請求項 7】

請求項 1 に記載の装置であって、前記サンプル処理部（1040）は、ゲイン情報に応じて前記選択音声信号サンプルを再スケーリングして再スケーリング音声信号サンプルを得て、前記再スケーリング音声信号サンプルを入力音声信号サンプルと組み合わせることにより処理済音声信号サンプルを得るように構成される、装置。

【請求項 8】

請求項 7 に記載の装置であって、

前記サンプル処理部（1040）は、前記現在のフレームが当該装置によって受信されており、且つ、当該装置によって受信されている前記現在のフレームが破損していない場合、前記再スケーリング音声信号サンプル及び前記入力音声信号サンプルの組み合わせを示す前記処理済音声信号サンプルを前記遅延バッファ（1020）に記憶させ、前記再スケーリング音声信号サンプルを前記遅延バッファ（1020）に記憶させないように構成され、

10

前記サンプル処理部（1040）は、前記現在のフレームが当該装置によって受信されていない場合、又は当該装置によって受信されている前記現在のフレームが破損している場合、前記再スケーリング音声信号サンプルを前記遅延バッファ（1020）に記憶させ、前記処理済音声信号サンプルを前記遅延バッファ（1020）に記憶させないように構成される、装置。

【請求項 9】

請求項 7 に記載の装置であって、前記サンプル処理部（1040）は、前記現在のフレームが当該装置によって受信されていない場合、又は当該装置によって受信されている前記現在のフレームが破損している場合、前記処理済音声信号サンプルを前記遅延バッファ（1020）に記憶させるように構成される、装置。

20

【請求項 10】

請求項 1 から請求項 9 のいずれか 1 項に記載の装置であって、前記サンプル選択部（1030）は、前記変更後ゲインを計算するように構成される、装置。

【請求項 11】

請求項 1 から請求項 10 のいずれか 1 項に記載の装置であって、当該装置によってフレームが最後に受信されてから少なくとも所定数のフレームが当該装置によって受信されていない場合、前記変更後ゲイン $gain$ は、ゼロに設定される、装置。

30

【請求項 12】

符号化音声信号を復号して再構成音声信号を得るための方法であって、

複数のフレームを受信するステップと、

逆修正離散余弦変換を行うことによって複数のフレームを復号して復号された音声信号の音声信号サンプルを得るステップと、

長期予測を行うステップとを備え、

前記長期予測を行うステップは、

前記復号された音声信号の前記音声信号サンプルを遅延バッファ（1020）に記憶するステップと、

前記遅延バッファ（1020）に記憶されている前記音声信号サンプルから前記復号された音声信号の複数の選択音声信号サンプルを選択するステップと、

40

前記複数の選択音声信号サンプルを処理して前記再構成音声信号の再構成音声信号サンプルを得るステップとによって行われ、

現在のフレームが受信され、かつ、前記受信されている現在のフレームが破損していない場合、前記遅延バッファ（1020）に記憶されている前記音声信号サンプルから前記複数の選択音声信号サンプルを選択するステップは、前記現在のフレームに含まれるピッチ遅延情報に応じて行われ、

前記現在のフレームが受信されていない場合、又は前記受信されている現在のフレームが破損している場合、前記遅延バッファ（1020）に記憶されている前記音声信号サンプルから前記複数の選択音声信号サンプルを選択するステップは、以前に受信されている

50

別のフレームに含まれるピッチ遅延情報に応じて行われ、

当該方法は、更に、

変更後ゲインに応じて前記複数の選択音声信号サンプルを再スケーリングするステップを備え、

前記変更後ゲインは、

【数 47】

$gain = gain_past * damping;$

の式によって定義され、ここで、 $gain$ は、前記変更後ゲインであり、

$gain_past$ は、 $gain$ の算出後に $gain$ に設定され、

$damping$ は、 $0 \leq damping \leq 1$ の範囲における実数値である、方法。

【請求項 13】

コンピュータ又は信号プロセッサにおいて実行された時に請求項 12 の方法を実現するためのコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声信号符号化、処理及び復号に関するものであり、特に、エラー封じ込め中の切替音声符号化システムについての向上した信号フェードアウトのための装置及び方法に関するものである。

【背景技術】

【0002】

以下、パケット損失封じ込め（PLC）中の言語音声及び音声コーデック・フェードアウトに関する先行技術について説明する。先行技術についての説明では、まずGシリーズ（G. 718、G. 719、G. 722、G. 722. 1、G. 729、G. 729. 1）のITU-Tコーデックについて述べ、次に3GPPコーデック（AMR、AMR-WB、AMR-WB+）及び1種類のIETFコーデック（OPUS）について述べ、最後にMP3コーデック（HE-AAC、HILN）（ITU=国際電気通信連合；3GPP=第3世代パートナーシッププロジェクト；AMR=適応マルチレート；WB=広帯域；IETF=インターネット工学タスクフォース）について述べる。続いて、背景ノイズレベルのトレースに関する先行技術についての分析を行い、その後概略となる概要を述べる。

【0003】

まず、G. 718について検討する。G. 718は、狭帯域及び広帯域の言語音声コーデックであり、DTX/CNG（DTX=デジタルシミュレーションシステム；CNG=快適ノイズ生成）をサポートする。実施例は、特に低遅延符号に関するため、ここでは低遅延バージョンモードについてより詳細に説明する。

【0004】

ACELP（第1層）（ACELP=代数符号励振線形予測）に関し、ITU-Tは、G. 718 [ITU 08a、7. 11節] についてフェード速度を制御するための線形予測領域における適応フェードアウトを推奨している。一般的に、封じ込めは以下の原理に従う。

【0005】

G. 718によると、フレーム消去の場合、封じ込め戦略は、信号エネルギー及びスペクトル包絡線の、背景ノイズの推定パラメータへの収束として概括することができる。信号の周期性は、ゼロへ収束される。収束の速度は、最後に正しく受信したフレームのパラメータと、連続する消去されたフレームの数とに依存し、減衰定数 α によって制御される。減衰定数 α は、更に、「無声」フレームについてのLPフィルタ（LP=線形予測）の安定度 θ に依存する。一般的に、収束は、最後の良好な受信フレームが安定なセグメントにある場合は遅く、フレームが遷移セグメントにある場合は速い。

【0006】

減衰定数 α は、[ITU08 a、6. 8. 1. 3. 1及び7. 11. 1. 1節]に記載の信号分類から導き出される言語音声信号クラスに依存する。安定度定数 θ は、隣接するISF（インミッタンススペクトル周波数）フィルタ[ITU08 a、7. 1. 2. 4. 2節]間の距離測定値に基づいて計算される。

【0007】

表1は、 α の計算方式を示す。

【0008】

【表1】

最後の良好な受信フレーム	連続する消去フレームの数	α
人工的なオンセット		0. 6
オンセット、有声	≤ 3	1. 0
	> 3	0. 4
有声遷移		0. 4
無声遷移		0. 8
無声	$= 1$	$0. 2 \cdot \theta + 0. 8$
	$= 2$	0. 6
	> 2	0. 4

【0009】

表1：減衰定数 α の値。値 θ は、隣接するLPフィルタ間の距離測定値から計算された安定度定数である。[ITU08 a、7. 1. 2. 4. 2節]

更に、G. 718は、スペクトル包絡線に変更を加えるためのフェード方法を提供している。一般的な考えは、適応ISF平均ベクトルへ向けて最後のISFパラメータを収束させることである。最初に、最後の3つの既知のISFベクトルから平均ISFベクトルを算出する。次に、この平均ISFベクトルは、オフラインで学習された長期ISFベクトル（これは一定のベクトルである）によって再び平均される[ITU08 a、7. 11. 1. 2節]。

【0010】

更に、G. 718は、長期挙動、及び従って背景ノイズとの相互作用、を制御するためのフェード方法を提供しており、ここでピッチ励振エネルギー（及び従って励振周期性）は、0へ収束する一方、ランダム励振エネルギーは、CNG励振エネルギーに収束する[ITU08 a、7. 11. 1. 6節]。新規ゲイン減衰は、

【0011】

【数1】

$$g_s^{[1]} = \alpha g_s^{[0]} + (1 - \alpha) g_n \quad (1)$$

【0012】

として計算され、ここで、 $g_s^{[1]}$ は、次のフレームの開始時の新規ゲインであり、 $g_s^{[0]}$ は、現在のフレームの開始時の新規ゲインであり、 g_n は、快適ノイズ生成中に用いられる励振のゲインであり、そして減衰定数 α である。

【0013】

周期的励振減衰と同様、ゲインは、 $g_s^{[0]}$ で始まってサンプル毎にフレーム全体に亘って線形に減衰させられ、次のフレームの開始時に $g_s^{[1]}$ に達する。

【0014】

図2は、G. 718のデコーダ構造の概略を示す。特に、図2は、ハイパスフィルタを特徴とするPLCの高レベルG. 718デコーダ構造を示す。

【0015】

上述のG. 718の方策によって、新規ゲイン g_s は、パケット損失の長いバーストについて快適ノイズ生成 g_n 中に用いられるゲインに収束する。[ITU08a、6.12.3節]に記載のように、快適ノイズゲイン g_n は、エネルギー E の平方根として与えられる。 E の更新の条件は、詳細に記載されていない。参照実現（浮動小数点Cコード、stat_noise_uv_mod.c）に従い、 E は、以下のように導出され、

【0016】

【数2】

10

```

if(unvoiced_vad==0) {
    if(unv_cnt>20) {
        ftmp=lp_gainc*lp_gainc;
        lp_ener=0.7f*lp_ener+0.3f*ftmp;
    }
    else {
        unv_cnt++;
    }
}
else {
    unv_cnt=0
}

```

20

【0017】

ここで、unvoiced_vadは、発話区間検出を保持し、unv_cntは、行における無声フレームの数を保持し、lp_gaincは、固定コードブックの低域通過ゲインを保持し、lp_enerは、低域通過CNGエネルギー推定値 E を保持する（これは0で初期化される）。

30

【0018】

更に、G. 718は、最後の良好なフレームの信号が「無声」とは異なって分類された場合、無声励振の信号経路へ導入されるハイパスフィルタを提供している。図2を参照。また、[ITU08a、7.11.1.6節]を参照。このフィルタは、低い放置特性を有し、直流での周波数応答はナイキスト周波数よりも約5dB低い。

【0019】

更に、G. 718は、結合解除されたLTPフィードバックループ（LTP=長期予測）を提案している。通常動作中は、適応コードブックについてのフィードバックループは、全励振に基づいてサブフレーム単位で更新される（[ITU08a、7.1.2.1.4節]）。封じ込め中は、このフィードバックループは、有声励振のみに基づいてフレーム単位で更新される（[ITU08a、7.11.1.4節、7.11.2.4節、7.11.1.6節、7.11.2.6節を参照。dec_GV_exc@dec_gen_voic.c及びsyn_bfi_post@syn_bfi_pre_post.c]）。この方策では、適応コードブックは、ランダムに選択された新規励振によって、由来を有するノイズで「汚染」されない。

40

【0020】

G. 718の変換符号化強化層（3-5）に関し、封じ込め中のデコーダの挙動は、高い層の復号に関して通常動作と同様であるが、但しMDC Tスペクトルはゼロに設定される。特別なフェードアウト挙動は封じ込め中に適用されない。

50

【0021】

CNGに関し、G. 718において、CNG合成は、以下の順序で行われる。最初に、快適ノイズフレームのパラメータが復号される。次に、快適ノイズフレームが合成される。その後、ピッチバッファがリセットされる。次に、FER（フレームエラー回復）分類についての合成が保存される。その後、スペクトル強調解除が行われる。そして、低周波数の後フィルタリングが行われる。そして、CNG変数が更新される。

【0022】

封じ込めの場合、丁度同じことが行われるが、但し、CNGパラメータは、ビットストリームから復号されない。これは、パラメータは、フレーム損失中、更新されないが、最後の良好なSID（無音挿入記述子）フレームからの復号パラメータが用いられることを意味する。

10

【0023】

次に、G. 719について検討する。G. 719は、サイレン（Siren）22に基づいており、変換ベースの全帯域音声コーデックである。ITU-Tは、G. 719について、スペクトル領域におけるフレーム繰り返しを伴うフェードアウトを推奨している [ITU08b、8.6節]。G. 719によると、フレーム消去封じ込めメカニズムがデコーダに組み込まれる。フレームが正しく受信された場合、復元された変換係数は、バッファに記憶される。デコーダに対して、フレームが失われた又はフレームが破損したことが通知されると、最も最近受信されたフレームにおいて復元された変換係数は、0.5の定数に比例して減少され、次に、現在のフレームについての復元された変換係数として使用される。次に、デコーダは、これらを時間領域に変換し、窓化・重複・加算の演算を実行する。

20

【0024】

以下、G. 722について説明する。G. 722は、50～7000Hz符号化システムであって、64kbit/秒までのビットレート内でサブバンド適応差分パルスコード変調（SB-ADPCM）を用いるものである。信号は、QMF分析（QMF＝直交ミラーフィルタ）を用いてより高いサブバンド及びより低いサブバンドに分割される。結果として得られる2つの帯域は、ADPCM符号化（ADPCM＝適応差分パルス符号変調）される。

【0025】

30

G. 722については、パケット損失封じ込めについての高複雑度のアルゴリズムが付録（Appendix）IIIに述べられており [ITU06a]、パケット損失封じ込めについての低複雑度のアルゴリズムが付録IVに述べられている [ITU07]。G. 722—付録III（ [ITU06a、III.5節] ）においては、徐々に実行されるミューティングであって、フレーム損失の20ミリ秒後に開始し、フレーム損失の60ミリ秒後に完了するものが提案されている。更に、G. 722—付録IVにおいては、フェードアウト技術であって、「サンプル毎に計算及び適合されるゲイン定数を各々のサンプルに」適用するものが提案されている [ITU07、IV.6.1.2.7節]。

【0026】

G. 722においては、ミューティングプロセスは、サブバンド領域において、QMF合成の直前に、PLCモジュールの最後のステップとして行われる。ミューティング定数の算出は、PLCモジュールの一部でもある信号分類子からクラス情報を用いて実行される。区別は、TRANSIENT、UV__TRANSITION等のクラスの間で行われる。更に、10ミリ秒フレームの単一の損失と、他の場合（10ミリ秒のフレームの多数の損失及び20ミリ秒のフレームの単一／多数の損失）との間での区別が行われる。

40

【0027】

これを図3に示す。特に、図3は、G. 722のフェードアウト定数がクラス情報に依存し、80個のサンプルが10ミリ秒に相当するシナリオを示す。

【0028】

G. 722によると、PLCモジュールは、損失フレームについての信号と、次の良好

50

なフレームとクロスフェードさせられることになっている或る追加の信号（10ミリ秒）とを生成する。この追加の信号についてのミューティングは、同じ規則に従う。G. 722の高帯域封じ込めにおいては、クロスフェードは行われない。

【0029】

以下、G. 722. 1について検討する。G. 722. 1は、サイレン7に基づいており、超広帯域拡張モードを有する変換ベースの広帯域音声コーデックであり、G. 722. 1Cと呼ばれる。G. 722. 1Cそれ自体は、サイレン14に基づいている。ITU-Tは、G. 722. 1について、後続のミューティングを伴うフレーム繰り返しを推奨している〔ITU05、4. 7節〕。デコーダに対して、この勧告には規定されていない外部シグナリングメカニズムによって、フレームが失われた又は破損したと通知されると、先行フレームの復号されたMLT（変調済ラップド変換）係数を繰り返す。次に、これらを時間領域に変換し、先行フレーム及び次のフレームの復号された情報によって重複及び加算の演算を実行する。先行フレームもまた失われ又は損失している場合、デコーダは、全ての現在のフレームMLT係数をゼロに設定する。

10

【0030】

次に、G. 729について検討する。G. 729は、音声のための音声データ圧縮アルゴリズムであって、10ミリ秒の期間の packets におけるデジタル音声を圧縮するものである。公式には、符号励振線形予測言語音声符号化（CS-ACELP）を用いた8kb/sでの言語音声の符号化として記載されている〔ITU12〕。

20

【0031】

〔CPK08〕で概略的に説明されるように、G. 729は、LP領域におけるフェードアウトを推奨している。G. 729規格において採用されるPLCアルゴリズムは、以前に受信した言語音声情報に基づいて現在のフレームについての言語音声信号を復元する。換言すると、PLCアルゴリズムは、紛失励振を、以前に受信したフレームの等価の特性と入れ替えるが、励振エネルギーは、徐々に減衰し、最終的には、適応コードブック及び固定コードブックのゲインは、一定の定数によって減衰させられる。

【0032】

減衰させられた固定コードブックゲインは、

【0033】

【数3】

30

$$g_C^{(m)} = 0.98 \cdot g_C^{(m-1)}$$

【0034】

によって与えられ、ここでmは、サブフレームインデックスである。

【0035】

適応コードブックゲインは、先行する適応コードブックゲインを減衰させたものに基づいている。

【0036】

【数4】

40

$$g_p^{(m)} = 0.9 \cdot g_p^{(m-1)} \text{ 但し、 } g_p^{(m)} < 0.9$$

【0037】

ナム・イン・パーク（Nam in Park）他は、G. 729について、線形回帰による予測を用いた信号振幅制御を提案している〔CPK08、PKJ+11〕。これは、バースト状の packets 損失に対応するものであり、線形回帰を中核技術として用いる。線形回帰は、

【0038】

【数 5】

$$g'_i = a + bi \quad (2)$$

【0 0 3 9】

として線形モデルに基づいている。ここで、 g'_i は、新たに予測された現在の振幅であり、 a 及び b は、一次線形関数についての係数であり、 i は、フレームのインデックスである。最適化された係数 a^* 及び b^* を求めるために、二乗予測エラーの合計が次式のように最小化される。

【0 0 4 0】

10

【数 6】

$$\varepsilon = \sum_{j=i-4}^{i-1} (g_j - g'_j)^2 \quad (3)$$

【0 0 4 1】

ここで、 ε は、二乗エラーであり、 g_j は、元の過去の j 番目の振幅である。このエラーを最小化するために、 a 及び b に関する導関数を単にゼロに設定する。最適化されたパラメータ a^* 及び b^* を用いて各々の g^*_i の推定値は、

【0 0 4 2】

20

【数 7】

$$g^*_i = a^* + b^*i \quad (4)$$

【0 0 4 3】

によって表される。

【0 0 4 4】

図4は、振幅予測を示し、特に、線形回帰を用いて振幅 g^*_i の予測を示す。

【0 0 4 5】

失われたパケット i の振幅 A'_i を得るために、比 σ_i 、

30

【0 0 4 6】

【数 8】

$$\sigma_i = \frac{g_i^*}{g_{i-1}} \quad (5)$$

【0 0 4 7】

を、スケール定数 S_i によって乗算し、即ち、

【0 0 4 8】

【数 9】

$$A'_i = S_i * \sigma_i \quad (6)$$

40

【0 0 4 9】

とし、ここで、スケール定数 S_i は、連続する封じ込められたフレーム $l(i)$ の数に依存し、即ち、

【0 0 5 0】

【数 1 0】

$$S_i = \begin{cases} 1.0, & \text{if } l(i)=1, 2 \\ 0.9, & \text{if } l(i)=3, 4 \\ 0.8, & \text{if } l(i)=5, 6 \\ 0, & \text{その他} \end{cases} \quad (7)$$

【0 0 5 1】

とする。

【0 0 5 2】

〔P K J + 1 1〕においては、僅かに異なるスケーリングが提案されている。

【0 0 5 3】

G. 7 2 9 によると、この後、 A'_i を平滑化して、フレーム境界における離散減衰を防ぐ。最後の平滑化された振幅 $A_i(n)$ は、先行するP L C成分から得た励振へ乗算される。

【0 0 5 4】

以下、G. 7 2 9. 1について検討する。G. 7 2 9. 1は、G. 7 2 9 ベースの埋め込み可変ビットレートコードである。即ち、8 ~ 3 2 k b i t / 秒のスケーリング可能な広帯域コーダビットストリームであって、G. 7 2 9 と相互動作可能なものである〔I T U 0 6 b〕。

【0 0 5 5】

G. 7 2 9. 1によると、G. 7 1 8（上記を参照）と同様、適合フェードアウトであって、信号特性の安定度に依存するものが提案されている（〔I T U 0 6 b、7. 6. 1 節〕）。封じ込め中、信号は、通常、減衰定数 α に基づいて減衰させられ、この減衰定数は、最後の良好な受信フレームクラスのパラメータと、連続する消去されたフレームの数とに依存する。減衰定数 α は、更に、「無声」フレームについてのL Pフィルタの安定度に依存する。一般的に、減衰は、最後の良好な受信フレームが安定なセグメントにある場合は遅く、フレームが遷移セグメントにある場合は速い。

【0 0 5 6】

【数 1 1】

更に、減衰定数 α は、1サブフレーム $\bar{g}_p$ 当りの平均ピッチゲインに依存し（〔I T U 0 6 b、式1 6 3, 1 6 4〕）、

$$\bar{g}_p = 0.1g_p^{(0)} + 0.2g_p^{(1)} + 0.3g_p^{(2)} + 0.4g_p^{(3)} \quad (8)$$

であり、ここで、 $g_p^{(i)}$ は、サブフレーム i におけるピッチゲインである。

【0 0 5 7】

表 2 は、 α の計算方式を示し、ここで、

【0 0 5 8】

【数 1 2】

$$\beta = \sqrt{\bar{g}_p} \quad \text{ここで、} 0.85 \leq \beta \leq 0.98 \quad (9)$$

【0 0 5 9】

である。封じ込めプロセス中、 α は、以下の封じ込めツールにおいて用いられる。

【0 0 6 0】

10

20

30

40

【表 2】

最後の良好な受信フレーム	連続する消去フレームの数	α
有声	1	β
	2, 3	—
	> 3	g_p 0.4
オンセット	1	0.8 β
	2, 3	—
	> 3	g_p 0.4
人工的なオンセット	1	0.6 β
	2, 3	—
	> 3	g_p 0.4
有声遷移	≤ 2	0.8
	> 2	0.2
無声遷移		0.88
無声	1	0.95
	2, 3	$0.6\theta + 0.4$
	> 3	0.4

【0061】

表 2：減衰定数 α の値。値 θ は、隣接する LP フィルタ間の距離測定値から計算された安定度定数である。[ITU06b、7.6.1 節]

G.729.1 によると、声門音パルス再合成に関し、先行フレームの励振の最後のパルスが周期的部分の構成に用いられるため、そのゲインは、封じ込められたフレームの開始時においてほぼ正しく、1 に設定され得る。次に、サンプル毎にフレーム全体においてゲインを線形的に減衰させ、フレームの終わりで α の値に達する。有声セグメントのエネルギー変化は、最後の良好なフレームの各サブフレームのピッチ励振ゲイン値を用いて外挿される。一般的に、これらのゲインが 1 よりも大きい場合、信号エネルギーは増加しており、1 よりも低い場合、エネルギーは減少している。従って、 α は、上述のように $\beta = (\gamma / g_p)^{1/2}$ に設定される。[ITU06b、式 163、164] を参照。強いエネルギー増加及び現象を避けるために、 β の値は、0.98 と 0.85 の間に削減する。[ITU06b、7.6.4 節] を参照。

【0062】

励振のランダムな部分の構成に関し、G.729.1 によると、消去されたブロックの開始時に、新規ゲイン g_s は、最後の良好なフレームの各サブフレームの新規励振ゲインを用いて初期化され、

【0063】

【数 13】

$$g_s = 0.1g^{(0)} + 0.2g^{(1)} + 0.3g^{(2)} + 0.4g^{(3)}$$

【0064】

となる。ここで、 $g^{(0)}$ 、 $g^{(1)}$ 、 $g^{(2)}$ 及び $g^{(3)}$ は、最後の正しく受信されたフレームの

4つのサブフレームの固定コードブック又は新規ゲインである。新規ゲイン減衰は、

【0065】

【数14】

$$g_s^{(1)} = \alpha \cdot g_s^{(0)}$$

【0066】

として行われ、ここで、 $g_s^{(1)}$ は、次のフレームの開始時の新規ゲインであり、 $g_s^{(0)}$ は、現在のフレームの開始時の新規ゲインであり、 α は、上述の表2に定義した通りである。従って、周期的励振減衰と同様、ゲインは、 $g_s^{(0)}$ で始まってサンプル毎にフレーム全体を通して線形的に減衰させられて、次のフレームの開始時に達成される $g_s^{(1)}$ の値に達する。

10

【0067】

G. 729. 1によると、最後の良好なフレームが「無声」であれば、新規励振のみを用い、これを更に0.8の定数で減衰させる。この場合、励振の周期的な部分が利用可能ではないため、過去の励振バッファを新規励振で更新する。[ITU 06b、7.6.6節]を参照。

【0068】

以下、AMRについて検討する。3GPP AMR [3GP12b]は、ACELPアルゴリズムを利用した言語音声コーデックである。AMRは、8000サンプル/秒のサンプリングレート及び4.75~12.2kbit/秒のビットレートで言語音声を符号化することができ、シグナリング無音記述子フレーム(DTX/CNG)をサポートする。

20

【0069】

AMRにおいて、エラー封じ込め中([3GP12a]を参照)、エラー気味のフレーム(ビットエラー)と、完全に失われたフレーム(データが全くない)とを区別する。

【0070】

ACELP封じ込めの場合、AMRは、チャンネルの品質を推定する状態マシンを導入する。状態カウンタの値が大きいほど、チャンネル品質は劣悪である。システムは、状態0から始まる。劣悪なフレームが検出されるたびに、状態カウンタを1ずつ増加させ、6に達すると飽和する。良好な言語音声フレームが検出されるたびに、状態カウンタは、ゼロにリセットされるが、但し、状態が6である場合、状態カウンタを5に設定する。状態マシンの制御フローは、以下のCコードによって記述することができる(BFIは、劣悪フレームインジケータであり、Stateは状態変数である)。

30

【0071】

【数 1 5】

```

if(BFI!=0) {
    State=State+1;
}
else if(State==6) {
    State=5;
}
else {
    State=0;
}
if(State>6) {
    State=6;
}

```

10

【0 0 7 2】

この状態マシンに加えて、AMRにおいては、現在及び先行するフレームからの劣悪フレームフラグを調べる (prevBFI)。

20

【0 0 7 3】

3つの異なる組合せが可能である。

【0 0 7 4】

3つの組み合わせのうちの1番目のものが、BFI=0、prevBFI=0、状態=0である。即ち、受信言語音声フレーム又は先行する受信言語音声フレームにおいてエラーは検出されない。受信言語音声パラメータは、言語音声合成における通常の態様で使用する。言語音声パラメータの現在のフレームが保存される。

【0 0 7 5】

3つの組み合わせのうちの2番目のものが、BFI=0、prevBFI=1、状態=0又は5である。即ち、受信言語音声フレームにおいてエラーは検出されないが、先行する受信言語音声フレームは劣悪である。LTPゲイン及び固定コードブックゲインは、最後の受信された良好なサブフレームについて用いられる値未満に制限され、即ち

30

【0 0 7 6】

【数 1 6】

$$g_p = \begin{cases} g_p, & g_p \leq g_p(-1) \\ g_p(-1), & g_p > g_p(-1) \end{cases} \quad (10)$$

【0 0 7 7】

となり、ここで、 g_p = 現在の復号されたLTPゲイン、 $g_p(-1)$ = 最後の良好なサブフレーム (BFI=0) について用いられるLTPゲインであり、更に、

40

【0 0 7 8】

【数 1 7】

$$g_c = \begin{cases} g_c, & g_c \leq g_c(-1) \\ g_c(-1), & g_c > g_c(-1) \end{cases} \quad (11)$$

【0 0 7 9】

となり、ここで、 g_c = 現在の復号された固定コードブックゲイン、 $g_c(-1)$ = 最後の良好なサブフレーム (BFI=0) について用いられる固定コードブックゲインである。

50

【0080】

受信した言語音声パラメータの残りは、言語音声合成において通常の態様で使用される。言語音声パラメータの現在のフレームが保存される。

【0081】

3つの組み合わせのうちの3番目のものが、 $BFI = 1$ 、 $prevBFI = 0$ 又は1、状態 $= 1 \cdots 6$ である。即ち、受信言語音声フレームにおいてエラーが検出され、置き換え及びミュート手順が開始される。LTPゲイン及び固定コードブックゲインは、先行するサブフレームからの減衰した値と入れ替えられ、即ち

【0082】

【数18】

10

$$g_p = \begin{cases} P(state) \cdot g_p(-1), & g_p(-1) \leq \text{median5}(g_p(-1), \dots, g_p(-5)) \\ P(state) \cdot \text{median5}(g_p(-1), \dots, g_p(-5)) & g_p(-1) > \text{median5}(g_p(-1), \dots, g_p(-5)) \end{cases} \quad (12)$$

【0083】

となり、ここで、 g_p は、現在の復号されたLTPゲインを示し、 $g_p(-1), \dots, g_p(-n)$ は、最後の n 個のサブフレームについて用いられるLTPゲインを示し、 $\text{median5}()$ は、5点の中央値演算を示し、更に、

$P(state) = \text{減衰定数}$

であり、ここで、 $(P(1) = 0.98, P(2) = 0.98, P(3) = 0.8, P(4) = 0.3, P(5) = 0.2, P(6) = 0.2)$ であり、 $state = \text{状態数}$ であり、

20

【0084】

【数19】

$$g_c = \begin{cases} C(state) \cdot g_c(-1), & g_c(-1) \leq \text{median5}(g_c(-1), \dots, g_c(-5)) \\ C(state) \cdot \text{median5}(g_c(-1), \dots, g_c(-5)) & g_c(-1) > \text{median5}(g_c(-1), \dots, g_c(-5)) \end{cases} \quad (13)$$

【0085】

であり、ここで、 g_c は、現在の復号された固定コードブックゲインを示し、 $g_c(-1), \dots, g_c(-n)$ は、最後の n 個のサブフレームについて用いられる固定コードブックゲインを示し、 $\text{median5}()$ は、5点の中央値演算を示し、 $C(state) = \text{減衰定数}$ であり、ここで、 $(C(1) = 0.98, C(2) = 0.98, C(3) = 0.98, C(4) = 0.98, C(5) = 0.98, C(6) = 0.7)$ であり、 $state = \text{状態数}$ である。

30

【0086】

AMRにおいては、LTP遅れ値(LTP=長期予測)は、先行するフレームの4番目のサブフレームからの過去の値(12.2モード)又は最後の正しく受信された値に基づく僅かに変更した値(他の全てのモード)と入れ替えられる。

【0087】

AMRによると、エラーを含むフレームからの受信した固定コードブック新規パルスは、破損したデータが受信されたときに固定コードブック新規パルスが受信された状態において用いられる。データが受信されなかった場合には、ランダムな固定コードブックインデックスを用いることになる。

40

【0088】

AMRにおけるCNGに関し、[3GP12a、6.4節]によると、各々の最初の失われたSIDフレームは、以前に受信した有効なSIDフレームからのSID情報を用いて置き換えられ、有効なSIDフレームについての手順が適用される。後続の失われたSIDフレームについては、減衰技術を、徐々に出力レベルを減少させる快適ノイズに適用する。従って、最後のSID更新が50フレーム(=1秒)よりも前かどうかを調べ、もしそうであれば、出力は、ミュートとなる(1フレーム当たり $-6/8$ dBだけのレベル減衰[3GP12d、 $dtx_dec\{\} @ sp_dec.c$]、これは1秒当たり37.

50

5 dBをもたらす)。なお、CNGに適用されるフェードアウトはLP領域で実行される。

【0089】

以下、AMR-WBについて検討する。適応マルチレート-WB [ITU03、3GP09c] は、AMRに基づく言語音声コーデックACELPである（1.8節を参照）。これは、パラメータ的帯域幅拡張を用い、またDTX/CNGをサポートする。基準 [3GP12g] の説明において、封じ込め例の解決策が与えられており、これは僅かな逸脱があるもののAMR [3GP12a] についてのものと同じである。従って、ここでは、AMRとの相違点のみについて説明する。規格の説明については、上述の記載を参照されたい。

10

【0090】

ACELPに関し、AMR-WBにおいて、参照ソースコードに基づいてACELPフェードアウトを実行する [3GP12c] が、これは、ピッチゲイン g_p (LTPゲインと呼ばれる上述のAMRについて) に変更を加え、コードゲイン g_c に変更を加えることで行われる。

【0091】

失われたフレームについては、最初のサブフレームについてのピッチゲイン g_p は、0.95及び0.5の間に制限されることを除き、最後の良好なフレームにおけるものと同じである。2番目、3番目及び続くサブフレームについては、ピッチゲイン g_p は、0.95の定数だけ減少され、やはり制限される。

20

【0092】

AMR-WBにおいては、封じ込められたフレームにおいて、 g_c が最後の g_c に基づくことが提案されている。

【0093】

【数20】

$$g_{c, \text{current}} = g_{c, \text{past}} * (1.4 - g_{p, \text{past}}) \quad (14)$$

$$g_c = g_{c, \text{current}} * g_{c, \text{inov}} \quad (15)$$

$$g_{c, \text{inov}} = \frac{1.0}{\sqrt{\frac{\text{ener}_{\text{inov}}}{\text{サブフレーム_サイズ}}}} \quad (16)$$

30

$$\text{ener}_{\text{inov}} = \sum_{i=0}^{\text{サブフレーム_サイズ}-1} \text{code}[i] \quad (17)$$

【0094】

LTP遅れを封じ込めるために、AMR-WBにおいては、フレーム損失の場合に、5つの最後の良好なLTP遅れ及びLTPゲインの履歴を用いて最良の更新方法を見つける。フレームがビットエラーを伴って受信された場合、受信したLTP遅れが使用可能か否かについての予測を実行する [3GP12g]。

40

【0095】

CNGに関し、AMR-WBにおいては、最後の正しく受信されたフレームがSIDフレームであり、或るフレームが失われたものと分類された場合、最後の有効なSIDフレーム情報と置き替え、有効なSIDフレームについての手順を適用することになる。

【0096】

後続の失われたSIDフレームについては、AMR-WBにおいては、減衰技術を、出力レベルを徐々に減少させる快適ノイズに適用することが提案される。従って、最後のSID更新が50フレーム (= 1秒) よりも前かどうかを調べ、もしそうであれば、出力は

50

、ミュートとなる（1フレーム当り $-3/8$ dBだけのレベル減衰〔3GP12f、dtx_dec{}@dtx.c〕、これは1秒当たり18.75 dBをもたらす）。なお、CNGに適用されるフェードアウトは、LP領域において実行される。

【0097】

次に、AMR-WB+について検討する。適応マルチレート-WB+〔3GP09a〕は、ACELP及びTCX（TCX=変換符号化励振）をコアコーデックとして用いる切り替えコーデックである。これは、パラメータ的な帯域幅拡張を用い、また、DTX/CNGをサポートする。

【0098】

AMR-WB+において、モード外挿論理を適用して、歪められたスーパーフレーム内の失われたフレームのモードを外挿する。このモード外挿は、モードインジケータの定義において冗長性が存在するという事実に基づいたものである。AMR-WB+によって提案された決定論理（〔3GP09a、図18〕によって与えられる）は、以下の通りである。

【0099】

- ベクトルモード（ m_{-1} , m_0 , m_1 , m_2 , m_3 ）が規定され、ここで、 m_{-1} は、先行するスーパーフレームの最後のフレームのモードを示し、 m_0 , m_1 , m_2 , m_3 は、現在のスーパーフレーム（ビットストリームから復号されたもの）におけるフレームのモードを示し、ここで、 $m_k = -1, 0, 1, 2$ 又は 3 （ -1 : 失われた、 0 : ACELP、 1 : TCX20、 2 : TCX40、 3 : TCX80）であり、失われたフレームの数 n_{loss} は 0 と 4 との間であり得る。

【0100】

- $m_{-1} = 3$ 、且つフレーム $0 \sim 3$ のモードインジケータのうちの2つが3に等しい場合、全てのインジケータが3に設定される。それは、その場合、1つのTCX80フレームが当該スーパーフレーム内に示されたことが確実だからである。

【0101】

- フレーム $0 \sim 3$ のうちのただ1つのインジケータが3（且つ失われたフレームの数 n_{loss} が3）である場合、モードは、（1, 1, 1, 1）に設定される。それは、その場合、TCX80目標スペクトルの $3/4$ が失われ、グローバルTCXゲインが失われた可能性が極めて高いからである。

【0102】

- モードが（x, 2, -1, x, x）又は（x, -1, 2, x, x）を示している場合、これはTCX40フレームを示す（x, 2, 2, x, x）へと外挿される。モードが（x, x, x, 2, -1）又は（x, x, -1, 2）を示す場合、やはりTCX40フレームを示す（x, x, x, 2, 2）へと外挿される。なお、（x, [0, 1], 2, 2, [0, 1]）は無効な構成である。

【0103】

- その後、失われた各々のフレームについて（モード=-1）、先行するフレームがACELPであった場合、モードは、ACELPに設定され（モード=0）、その他全ての場合には、モードは、TCX20に設定される（モード=1）。

【0104】

ACELPに関し、AMR-WB+によると、モード外挿の後、失われたフレームモードが結果として $m_k = 0$ となった場合、このフレームについて〔3GP12g〕におけるのと同じ方策を適用する（上記を参照）。

【0105】

AMR-WB+において、失われたフレームの数及び外挿されたモードに応じて、以下のTCX関連の封じ込め方策を区別する（TCX=変換符号化励振）。

【0106】

- フレーム全体が失われた場合、ACELP様の封じ込めを適用する。最後の励振を繰り返し、封じ込められたISF係数（その適応平均へ僅かにずらされたもの）を用いて時

10

20

30

40

50

間領域信号を合成する。これに加えて、1フレーム（20ミリ秒）当り0.7のフェードアウト定数〔3GP09b、dec_tcx.c〕を、LPC（線形予測符号化）合成の直前に線形予測領域で乗算する。

【0107】

- 最後のモードがTCX80であり、（部分的に失われた）スーパーフレームの外挿されたモードがTCX80（nloss = [1, 2]、モード = (3, 3, 3, 3, 3)）である場合、位相及び振幅外挿を利用し、最後の正しく受信されたフレームを考慮しながら、FFT領域で封じ込めを実行する。位相情報の外挿方策は、ここでは関心の対象ではない（フェード戦略に無関係）ため、説明は行わない。更なる詳細については、〔3GP09a、6.5.1.2.4節〕を参照。AMR-WB+の振幅変更に関し、TCX封じ込めについて実行される方策は以下のステップからなる〔3GP09a、6.5.1.2.3節〕。

10

【0108】

- 先行するフレーム振幅スペクトルは、

【0109】

【数21】

$$\text{oldA}[k] = |\text{old}\hat{X}[k]|$$

【0110】

のように計算される。

20

【0111】

- 現在のフレーム振幅スペクトルは、

【0112】

【数22】

$$A[k] = |\hat{X}[k]|$$

【0113】

のように計算される。

30

【0114】

- 先行するフレームと現在のフレームとの間の失われていないスペクトル係数のエネルギーのゲイン差は、

【0115】

【数23】

$$\text{gain} = \sqrt{\frac{\sum A[k]^2}{\sum \text{oldA}[k]^2}}$$

40

【0116】

のように計算される。

【0117】

- 紛失スペクトル係数の振幅は、

【0118】

【数24】

$$\text{if}(\text{lost}[k]) \quad A[k] = \text{gain} \cdot \text{oldA}[k]$$

【0119】

50

を用いて外挿される。

【0120】

- $m_k = [2, 3]$ を伴う失われたフレームの他の各々全てのケースにおいて、TCX 目標（（ビットストリームから復号されたノイズレベルを用いて）復号されたスペクトル + ノイズ充填の逆FFT）は、全ての利用可能な情報（グローバルTCXゲインを含む）を用いて合成される。この場合、フェードアウトは適用されない。

【0121】

AMR-WB+におけるCNGに関し、AMR-WBにおけるのと同じ方策が用いられる（上記を参照）。

【0122】

以下、OPUSについて検討する。OPUS [IET12] は、2つのコーデックからの技術を組み込む。即ち、言語音声志向のSILK（スカイプコーデックとして知られる）及び低待ち時間CELT（CELT = 条件付きエネルギーラップド変換）である。OPUSは、高ビットレートと低ビットレートとの間で継ぎ目なく調節することが可能であり、内部的には、低ビットレートでの線形予測コーデック（SILK）と高ビットレートでの変換コーデック（CELT）並びに短い重複についてのハイブリッドとの間で切り替わる。

【0123】

SILK音声データの圧縮及び解凍に関し、OPUSにおいては、SILKデコーダルーチンにおける封じ込め中に減衰させられるいくつかのパラメータが存在する。LTPゲインパラメータは、連続する失われたフレームの数に応じて、全てのLPC係数を1フレーム当たり0.99、0.95又は0.90で乗算することによって減衰させられ、ここで、先行するフレームの励振からの最後のピッチサイクルを用いて励振が構築される。ピッチ遅れパラメータは、連続する損失の際に極めて遅く増加する。単一の損失については、最後のフレームと比較して一定に保たれる。更に、励振ゲインパラメータは、1フレーム当たり $0.99^{lostcnt}$ で指数的に減衰させられるため、1番目の励振ゲインパラメータについては、励振ゲインパラメータは、0.99であり、2番目の励振ゲインパラメータについては、励振ゲインパラメータは、0.992であり、以下同様となる。励振は、可変オーバーフローによってホワイトノイズを生成しているランダム数生成部を用いて生成される。更に、LPC係数は、最後の正しく受信された係数の組に基づいて外挿／平均される。減衰した励振ベクトルの生成後、封じ込められたLPC係数をOPUSにおいて用いて時間領域出力信号を合成する。

【0124】

次に、OPUSの文脈で、CELTについて検討する。CELTは、変換ベースのコーデックである。CELTの封じ込めは、ピッチベースのPLC方策を特徴とし、これは、最大5つの連続して失われたフレームに適用される。フレーム6から開始して、ノイズ様の封じ込め方策を適用し、これは、背景ノイズを生成する。その特徴は、先行する背景ノイズのように聞こえるものと仮定する。

【0125】

図5は、CELTのバースト状の損失挙動を示す。特に、図5は、CELTで封じ込められた言語音声セグメントのスペクトログラム（x軸：時間、y軸：周波数）を示す。薄いグレーのボックスは、最初の5つの連続して失われたフレームを示し、ここで、ピッチベースのPLC方策が適用される。それ以降は、ノイズ様の封じ込めを示す。なお、切り替えは、瞬間的に実行され、スムーズに遷移するものではない。

【0126】

ピッチベースの封じ込めに関し、OPUSにおいては、ピッチベースの封じ込めは、自己相関によって復号信号における周期性を求め、ピッチオフセット（ピッチ遅れ）を用いて（LPC分析及び合成を用いて励振領域で）窓化した波形を繰り返すことからなる。窓化した波形を重ね合わせて、先行するフレーム及び次のフレームによって時間領域エイリアシング消去を保持する [IET12]。追加的に、フェードアウト定数を導出し、以下

10

20

30

40

50

のコードによって適用する。

【0127】

【数25】

opus_val32 E1=1, E2=1;

int period;

if (pitch_index ≤ MAX_PERIOD/2) {

 period=pitch_index;

}

else {

 period=MAX_PERIOD/2;

}

for (i=0;i<period;i++)

{

 E1+=exc[MAX_PERIOD-period+i]*exc[MAX_PERIOD-period+i];

 E2+=exc[MAX_PERIOD-2*period+i]*exc[MAX_PERIOD-2*period+i];

}

if(E1>E2) {

 E1=E2;

}

decay=sqrt(E1/E2));

arrenuation=decay;

【0128】

このコードにおいては、excは、損失前のMAX__PERIOD個のサンプルまでの励振信号を含む。

【0129】

励振信号は、後に減衰で乗算され、それから合成されてLPC合成を介して出力される。

【0130】

時間領域方策についてのフェードアルゴリズムは、以下のように概括することができる。

【0131】

- 損失前の最後のピッチサイクルのピッチ同期エネルギーを求める。

【0132】

- 損失前の最後から2番目のピッチサイクルのピッチ同期エネルギーを求める。

【0133】

- エネルギーが増加している場合、一定となるように制限する。減衰=1。

【0134】

- エネルギーが減少している場合、封じ込め中に同じ減衰で継続する。

【0135】

ノイズ様の封じ込めに関し、OPUSによると、6番目及びその後連続する失われたフレームについて、MDC T領域におけるノイズ置き換え方策を実行して、快適背景ノイズをシミュレートする。

【0136】

背景ノイズレベル及び形状のトレースに関し、OPUSにおいては、背景ノイズ推定は、以下のように行われる。MDC T分析の後、MDC T帯域エネルギーの平方根を1周波

10

20

30

40

50

数帯域当りで算出し、MDC Tビンのまともりは、[I E T 1 2、表 5 5] によるバークスケール (b a r k s c a l e) に従う。次に、エネルギーの平方根を、

【0 1 3 7】

【数 2 6】

$$\text{bandLogE}[i] = \log_2(e) \cdot \log_e(\text{bandE}[i] - e\text{Means}[i]) \quad \text{for } i=0 \dots 21 \quad (18)$$

【0 1 3 8】

によって $\log_2$ 領域に変換し、ここで、 e は、オイラー数であり、 bandE は、MDC T帯域の平方根であり、 $e\text{Means}$ は、(結果として増強された符号化ゲインをもたらす結果、ゼロ平均を得るために必要な) 定数のベクトルである。

10

【0 1 3 9】

OPUSにおいては、背景ノイズは、以下のようにデコーダ側でログオンされる [I E T 1 2、amp 2 L o g 2 及び l o g 2 A m p @ q u a n t _ b a n d . c] 。

【0 1 4 0】

【数 2 7】

$$\text{backgroundLogE}[i] = \min(\text{backgroundLogE}[i] + 8 \cdot 0.001, \text{bandLogE}[i]) \\ \text{for } i=0 \dots 21 \quad (19)$$

【0 1 4 1】

トレースされた最小エネルギーは、基本的には、現在のフレームの帯域のエネルギーの平方根によって決定されるが、1 フレームから次のフレームへの増加は、0. 0 5 d B によって制限される。

20

【0 1 4 2】

背景ノイズレベル及び形状の適用に関し、OPUSによると、ノイズ様のPLCが適用される場合、最後の良好なフレームにおいて導出された backgroundLogE を用いて線形領域に変換し、即ち

【0 1 4 3】

【数 2 8】

$$\text{bandE}[i] = e^{(\log_e(2) \cdot (\text{backgroundLogE}[i] + e\text{Means}[i]))} \quad \text{for } i=0 \dots 21 \quad (20)$$

30

【0 1 4 4】

とし、ここで、 e は、オイラー数であり、 $e\text{Means}$ は、「線形からログへ」の変換についての定数の同じベクトルである。

【0 1 4 5】

現在の封じ込め手順は、ランダム数生成部によって生成されたホワイトノイズをMDC Tフレームに充填し、このホワイトノイズを、帯域単位で bandE のエネルギーに一致するようにスケーリングするものである。この後、逆MDC Tを適用し、その結果として時間領域信号が得られる。重ね合わせ加算及び強調解除(通常の復号と同様)の後、出力される。

【0 1 4 6】

40

以下、MPEG-4 HE-AACについて検討する(MPEG=動画エキスパートグループ; HE-AAC=高効率高度音声符号化)。高効率高度音声符号化は、変換ベースの音声コーデック(AAC)であって、パラメータ的な帯域幅拡張(SBR)を補ったものからなる。

【0 1 4 7】

AAC(AAC=高度音声符号化)に関し、DABコンソーシアムは、DAB+におけるAACについて、周波数領域におけるゼロへのフェードアウトについて述べている[EBU10、A1. 2 節](DAB=デジタル音声ブロードキャスト)。フェードアウト挙動、例えば、減衰ランプは、固定又はユーザによって調節可能であり得る。最後のAU(AU=アクセス単位)からのスペクトル係数は、フェードアウト特性に対応する定数によ

50

って減衰させられ、そして、周波数から時間へのマッピングへと渡される。減衰ランプに応じて、封じ込めは、或る数の連続する無効A Uの後でミュートに切り替わり、即ち、完全なスペクトルは、0に設定されることになる。

【0148】

DRM (DRM=デジタル権利管理) コンソーシアムは、DRMにおけるAACについて、周波数領域におけるフェードアウトについて述べている [EBU12、5.3.3節]。最後の周波数から時間への変換の直前に封じ込めがスペクトルデータに対して働く。多数のフレームが破損している場合、封じ込めは、まず、最後の有効なフレームからの僅かに変更したスペクトル値に基づいてフェードアウトを実行する。更に、DAB+と同様、フェードアウト挙動、例えば、減衰ランプは、固定又はユーザによって調節可能であり得る。最後のフレームからのスペクトル係数は、フェードアウト特性に対応する定数によって減衰させられ、そして、周波数から時間へのマッピングに渡される。減衰ランプに応じて、封じ込めは、或る数の連続する無効フレームの後でミュートに切り替わり、即ち、完全なスペクトルは、0に設定されることになる。

10

【0149】

3GPPは、強化aacPlusにおけるAACについて、DRMと同様の周波数領域におけるフェードアウトを導入している [3GP12e、5.1節]。最後の周波数から時間への変換の直前に封じ込めがスペクトルデータに対して働く。多数のフレームが破損している場合、封じ込めは、まず、最後の良好なフレームからの僅かに変更したスペクトル値に基づいてフェードアウトを実行する。完全なフェードアウトは、5フレームを要する。最後の良好なフレームからのスペクトル係数がコピーされ、

20

【0150】

【数29】

$$\text{fadeOutFac}=2^{-(n\text{FadeOutFrame}/2)}$$

【0151】

の定数によって減衰させられ、ここで、nFadeOutFrameは、最後の良好なフレームからのフレームカウンタである。フェードアウトの5フレームの後、封じ込めは、ミュートに切り替わり、即ち、完全なスペクトルは、0に設定されることになる。

30

【0152】

ラウバー及びシュペルシュナイダーは、AACについて、エネルギー外挿に基づいて、MDCTスペクトルのフレーム単位のフェードアウトを導入している [LS01、4.4節]。先行するスペクトルのエネルギー形状を用いて、推定スペクトルの形状を外挿することができる。エネルギー外挿は、或る種の後封じ込めとして、封じ込め技術とは独立して実行することができる。

【0153】

AACに関し、エネルギー算出をスケール定数帯域ベースで実行して、人間の聴覚系の臨界帯域に近くなるようにする。個々のエネルギー値をフレーム毎に減少させ、音量をスムーズに下げる、例えば、信号をフェードアウトさせる。これが必要なのは、推定値が現在の信号を表す確率が時間の経過に伴って急速に減少するからである。

40

【0154】

出力対象のスペクトルの生成については、ラウバー等はフレーム繰り返し又はノイズ置き換えを提案している [LS01、3.2節及び3.3節]。

【0155】

クヴァッケンブッシュ及びドリーゼンは、AACについて、ゼロへの指数的なフレーム単位のフェードアウトを提案している [QD03]。時間/周波数係数の隣接する組の繰り返し提案され、各々の繰り返しは、指数的に増加する減衰を有するため、停止期間が長引いた場合に徐々にミュートへとフェードする。

【0156】

50

MPEG-4 HE-AACにおけるSBR（SBR＝スペクトル帯域複写）に関し、3GPPは、強化aacPlusにおけるSBRについて、復号された包絡線データをバッファし、フレーム損失の場合に、送信された包絡線データのバッファされたエネルギーを再使用し、各々全ての封じ込められたフレームについて3dBの一定比によってこれらを減少させることを提案している。結果は、通常の復号プロセスに入力され、包絡線調節部は、これを用いて、HF生成部の生成したパッチ広帯域の調節に用いられるゲインを算出する。次に、SBR復号が通常と同様に行われる。更に、デルタ符号化ノイズフロア及び正弦レベル値を消去する。先行する情報との差が利用可能でない状態が続くため、復号ノイズフロア及び正弦レベルは、HF生成信号のエネルギーと比例し続ける〔3GPP12e、5.2節〕。

10

【0157】

DRMコンソーシアムは、AACとの関連でSBRについて、3GPPと同じ技術について述べている〔EBU12、5.6.3.1節〕。更に、DABコンソーシアムは、DAB+におけるSBRについて、3GPPと同じ技術について述べている〔EBU10、A2節〕。

【0158】

以下、MPEG-4 CELP及びMPEG-4 HVXC（HVXC＝高調波ベクトル励振符号化）について検討する。DRMコンソーシアムは、CELP及びHVXCとの関連でSBRについて〔EBU12、5.6.3.2節〕、言語音声コーデックのためのSBRについての最低要件の封じ込めは、破損したSBRフレームが検出された時には必ず所定の組のデータ値を適用することであると述べている。これらの値は、低い相対的再生レベルでの静的な広帯域スペクトル包絡線をもたらし、より高い周波数へのロールオフを呈する。その目的は、単に、（厳密なミュートではなく）「快適ノイズ」を挿入することによって、挙動の劣悪な、時に大きい音声バーストが聴取者の耳に届かないようにすることである。これは、実際には、現実のフェードアウトではなく、或るエネルギーレベルにジャンプして或る種の快適ノイズを挿入することである。

20

【0159】

その後、代替案として、最後の正しく復号されたデータを再使用して、AAC+SBRの場合と同様に0へとレベル（L）をゆっくりとフェードさせることについて述べられる〔EBU12、5.6.3.2節〕。

30

【0160】

次に、MPEG-4 HILNについて検討する（HILN＝高調波及び個々のライン＋ノイズ）。マイネ他は、パラメトリック領域においてパラメトリックなMPEG-4 HILNコーデック〔ISO09〕のためのフェードアウトを導入している〔MEP01〕。連続した高調波成分の場合、破損した、異なって符号化されたパラメータの代わりとなる良好なデフォルト挙動は、周波数を一定に保ち、振幅を減衰定数（例えば-6dB）だけ減少させ、スペクトル包絡線を平均低域通過特性のそれへと収束させることである。スペクトル包絡線についての代替案として、これを変化させないことがあり得る。振幅及びスペクトル包絡線に関し、ノイズ成分は、高調波成分と同様に扱うことができる。

【0161】

40

以下、先行技術における背景ノイズレベルのトレースについて検討する。ランガカーリ及びロイズ〔RL06〕は、いくつかの方法についての良い概観を提供しており、それらの限界のいくつかについて述べている。背景ノイズレベルをトレースする方法は、例えば、最小追跡手順〔RL06〕〔Coh03〕〔SFB00〕〔Dob95〕、VADベースのもの（VAD＝発話区間検出）、カルマンフィルタリング〔Gan05〕〔BJH06〕、サブスペース分解〔BP06〕〔HJH08〕、ソフト決定〔SS98〕〔MPC89〕〔HE95〕、及び最小統計である。

【0162】

最小統計の方策がUSAC-2についての範囲内で使用されるものとして選択されており（USAC＝統合言語音声及び音声符号化）、その後より詳細に述べられている。

50

【0163】

最適平滑化及び最小統計に基づくノイズパワースペクトル密度推定 [Mar 01] では、アクティブな言語音声又は背景ノイズである信号とは独立して働くことのできるノイズ推定部が導入される。他の方法とは対照的に、最小統計アルゴリズムは、言語音声区間と言語音声ポーズとの間を区別するために明示的な閾値を用いないため、従来の発話区間検出方法よりもソフト決定方法により近く関係している。ソフト決定方法と同様、言語音声区間中に推定ノイズ PSD (パワースペクトル密度) を更新することもできる。

【0164】

最小統計方法は、2つの所見に基づいている。即ち、言語音声とノイズとは、通常統計的に独立であること、及び、ノイズの多い言語音声信号のパワーは、頻繁にノイズのパワーレベルへと減衰することである。従って、ノイズの多い信号 PSD の最小値を追跡することによって正確なノイズ PSD (PSD = パワースペクトル密度) 推定値を導出することが可能である。最小値は、平均値よりも小さい (又は他の場合では等しい) ため、最小追跡方法は、バイアス補償を必要とする。

10

【0165】

バイアスは、平滑化信号 PSD の分散の関数であるため、PSD 推定部の平滑化パラメータに依存する。最小追跡についての以前の業績では、一定平滑化パラメータ及び一定最小バイアス訂正を利用していたが、これとは対照的に、時間及び周波数依存の PSD 平滑化が用いられ、これも時間及び周波数依存バイアス補償を必要とする。

20

【0166】

最小追跡を用いることで、ノイズパワーの凡その推定値が得られる。しかしながら、いくつかの欠点がある。固定の平滑化パラメータによる平滑化によって、平滑化 PSD 推定値の言語音声区間のピークが広がる。これにより、ノイズ推定値が不正確になるが、それは、最小探索のためのスライド窓が広いピークへと滑り込む恐れがあるからである。従って、1に近い平滑化パラメータを用いることはできず、その結果、ノイズ推定値は、比較的大きな分散を有することになる。更に、ノイズ推定は、より低い値へ偏る。更に、ノイズパワーを増加させる場合、最小追跡が遅れることになる。

【0167】

低複雑度の MMSE ベースのノイズ PSD 追跡 [HHJ 10] では、DFT (離散的フーリエ変換) スペクトルに対して用いられる MMSE 探索を利用した背景ノイズ PSD 方策が導入されている。このアルゴリズムは、以下の処理ステップからなる。

30

【0168】

- 先行するフレームのノイズ PSD に基づいて最大可能性推定部が計算される。

【0169】

- 最小平均平方推定部が計算される。

【0170】

- 決定によって指示される方策を用いて最大可能性推定部が推定される [EM 84]。

【0171】

- 言語音声及びノイズ DFT 係数がガウス分布したものであると仮定して逆バイアス定数が計算される。

40

【0172】

- 推定されたノイズパワースペクトル密度が平滑化される。

【0173】

- アルゴリズムの完全な竦みを回避するために適用される安全策もある。

【0174】

データを用いる再帰的ノイズパワー推定に基づく非定常ノイズの追跡 [EH 08] では、極めて非定常的なノイズ源によって汚染された言語音声信号からのノイズスペクトル分散の推定方法が導入されている。この方法は、また、時間/周波数方向における平滑化を用いている。

【0175】

50

ノイズパワー推定の平滑化及び推定バイアス訂正に基づく低複雑度のノイズ推定アルゴリズム [Y u 0 9] は、[E H 0 8] において導入された方策を強化するものである。主な相違点は、ノイズパワー推定のためのスペクトルゲイン関数が反復的なデータを用いた方法によって求められることである。

【0176】

ノイズの多い言語音声の強調のための統計的方法 [M a r 0 3] においては、[M a r 0 1] において与えられた最小統計方策を、ソフト決定ゲイン変更 [M C A 9 9]、アプライオリ SNR の推定 [M C A 9 9]、適応ゲイン制限 [M C 9 9] 及び M M S E ログ・スペクトル振幅推定部 [E M 8 5] によって組み合わせる。

【0177】

フェードアウトに対する関心は、複数の言語音声及び音声コーデックにおいて特に大きい。特に、AMR ([3 G P 1 2 B] を参照) (A C E L P 及び C N G を含む)、AMR-WB ([3 G P 0 9 c] を参照) (A C E L P 及び C N G を含む)、AMR-WB+ ([3 G P 0 9 a] を参照) (A C E L P、T C X 及び C N G を含む)、G.718 ([I T U 0 8 a] を参照)、G.719 ([I T U 0 8 b] を参照)、G.722 ([I T U 0 7] を参照)、G.722.1 ([I T U 0 5] を参照)、G.729 ([I T U 1 2、C P K 0 8、P K J + 1 1 を参照])、M P E G - 4 H E - A A C / 強化 a a c P l u s ([E B U 1 0、E B U 1 2、3 G P 1 2 e、L S 0 1、Q D 0 3] を参照) (A A C 及び S B R を含む)、M P E G - 4 H I L N ([I S O 0 9、M E P 0 1] を参照) 及び O P U S ([I E T 1 2] を参照) (S I L K 及び C E L T を含む) がある。

【0178】

コーデックに応じて、フェードアウトは、それぞれ異なる領域で実行される。

【0179】

L P C を利用するコーデックについては、フェードアウトは、線形予測領域（励振領域としても知られる）において実行される。このことは、以下のコーデックに当てはまる。即ち、A C E L P に基づくコーデック、例えば、AMR、AMR-WB、AMR-WB+ の A C E L P コア、G.718、G.729、G.729.1、O P U S における S I L K コア、時間周波数変換を用いた励振信号を更に処理するコーデック、例えば、AMR-WB+ の T C X コア、O P U S における C E L T コア、及び、線形予測領域において動作する快適ノイズ生成 (C N G) 方式、例えば、AMR における C N G、AMR-WB における C N G、AMR-WB+ における C N G、である。

【0180】

時間信号を周波数領域に直接変換するコーデックについては、フェードアウトは、スペクトル／サブバンド領域において実行される。このことは、M D C T 又は類似の変換に基づくコーデック、例えば、M P E G - 4 H E - A A C における A A C、G.719、G.722 (サブバンド領域) 及び G.722.1 に当てはまる。

【0181】

パラメトリックコーデックについては、フェードアウトは、パラメトリック領域で適用される。このことは、M P E G - 4 H I L N に当てはまる。

【0182】

フェードアウト速度及びフェードアウト曲線に関し、フェードアウトは、一般的に、適切な領域における信号表現に適用される減衰定数の適用によって実現される。減衰定数のサイズは、フェードアウト速度及びフェードアウト曲線を制御する。殆どの場合、減衰定数は、フレーム単位に適用されるが、サンプル単位の適用も利用される。例えば、G 7 1 8 及び G.722 を参照。

【0183】

特定の信号セグメントのための減衰定数は、2つの態様、即ち絶対的及び相対的に与えられ得る。

【0184】

減衰定数が絶対的に与えられる場合、参照レベルは、常に最後の受信フレームのもので

10

20

30

40

50

ある。絶対的減衰定数は、通常、最後の良好なフレームの直後の信号セグメントについて 1 に近い値で始まり、それから速く又は遅く 0 へと劣化する。フェードアウト曲線は、これらの要因に直接依存する。これは、例えば、G. 722 の付録 I V（特に [ITU07、図 I V. 7] を参照）に記載される封じ込めに該当し、ここでは、可能なフェードアウト曲線は、線形又は徐々に線形である。ゲイン定数 $g(n)$ を考慮する場合、 $g(0)$ は最後の良好なフレームのゲイン定数、絶対減衰定数 $\alpha_{abs}(n)$ を表すものとする、後続の失われたフレームのゲイン定数は

【0185】

【数30】

$$g(n) = \alpha_{abs}(n) \cdot g(0) \quad (21)$$

10

【0186】

として導出することができる。

【0187】

減衰定数が相対的に与えられる場合、参照レベルは、先行するフレームからのものである。これが有利となるのは、再帰的封じ込め手順の場合、例えば、既に減衰した信号が更に処理され、再び減衰させられる場合である。

【0188】

減衰定数が再帰的に適用される場合、これは、以下であり得る。即ち、連続する失われたフレームの数とは独立の固定値、例えば、G. 719 の場合は、0.5（上記を参照）、連続する失われたフレームの数に関連した固定値、例えば、[CPK08]における G. 729 について提案されたもの、最初の 2 つのフレームについて 1.0、次の 2 つのフレームについて 0.9、フレーム 5, 6 について 0.8、後続の全てのフレームについて 0（上記を参照）、又は、連続する失われたフレームの数に関連した値であって信号特性に依存するもの、例えば、不安定な信号については、より速いフェードアウト、安定な信号については、より遅いフェードアウト、例えば、G. 718（上記の節及び [ITU08a、表 44] を参照）である。

20

【0189】

相対的なフェードアウト定数を $0 \leq \alpha_{rel}(n) \leq 1$ 、 n が失われたフレームの数（ $n \geq 1$ ）と仮定すると、後続のフレームのゲイン定数は、

30

【0190】

【数31】

$$g(n) = \alpha_{rel}(n) \cdot g(n-1) \quad (22)$$

$$g(n) = \left(\prod_{m=1}^n \alpha(m) \right) \cdot g(0) \quad (23)$$

40

$$g(n) = \alpha_{rel}^n \cdot g(0) \quad (24)$$

【0191】

のように導出することができ、結果として、指数的フェードが得られる。

【0192】

フェードアウト手順に関し、通常、減衰定数が特定されるが、いくつかの応用規格（DRM、DAB+）では、減衰定数は、製造者に委ねられる。

【0193】

異なる信号部分が別個にフェードさせられる場合、異なる減衰定数を適用することによ

50

って、例えば、調性を有する成分を或る速度でフェードさせ、ノイズ様の成分を別の速度でフェードさせることができる（例えば、A M B、S I L K）。

【0194】

通常、或るゲインをフレーム全体に適用する。フェードがスペクトル領域で実行される場合、これが唯一可能な態様である。しかしながら、フェードが時間領域又は線形予測領域で行われる場合、よりグラニュール状のフェードが可能である。このようなよりグラニュール状のフェードがG. 718で適用されており、ここでは、個々のゲイン定数が、各々のサンプルにつき、最後のフレームのゲイン定数と現在のフレームのゲイン定数との間の線形補間によって導出される。

【0195】

可変フレーム期間を伴うコーデックの場合、フレーム期間に応じて、一定の相対的な減衰定数から、異なるフェードアウト速度が生じる。これは、例えばA A Cの場合に該当し、ここではフレーム期間がサンプリングレートに依存する。

【0196】

適用されたフェード曲線を、最後に受信された信号の一時的形状に対して適用するために、（静的）フェードアウト定数を更に調節することができる。このような更なる動的な調節は、例えばA M Bに適用され、ここでは先行する5つのゲイン定数の中央値が考慮される（[3 G P 1 2 b]及び1. 8. 1節を参照）。減衰を実行する前に、中央値が最後のゲインよりも小さければ、現在のゲインを中央値に設定し、その他の場合は、最後のゲインを用いる。更に、このような更なる動的な調節は、例えばG 7 2 9に適用され、ここでは、振幅は、先行するゲイン定数の線形回帰を用いて予測される（[C P K 0 8、P K J + 1 1]及び1. 6節を参照）。この場合、最初の封じ込められたフレームについての結果のゲイン定数は、最後の受信フレームのゲイン定数を超過する場合がある。

【0197】

フェードアウトの目標レベルに関し、G. 718及びC E L Tを除き、全ての分析されたコーデック（これらのコーデックの快適ノイズ生成（C N G）を含む）について目標レベルは、0である。

【0198】

G. 718において、ピッチ励振（調性成分を表す）のフェード及びランダム励振（ノイズ様の成分を表す）のフェードを別個に実行する。ピッチゲイン定数がゼロへフェードさせられる一方、新規ゲイン定数は、C N G励振エネルギーにフェードさせられる。

【0199】

相対的な減衰定数が与えられると仮定すると、ここから、式（23）に基づき、以下の絶対減衰定数が得られ、

【0200】

【数32】

$$g(n) = \alpha_{rel}(n) \cdot g(n-1) + (1 - \alpha_{rel}(n)) \cdot g_n \quad (25)$$

【0201】

ここで、 g_n は、快適ノイズ生成中に用いられる励振のゲインである。この式は、 $g_n = 0$ の場合、式（23）に対応する。

【0202】

G. 718は、D T X / C N Gの場合、フェードアウトを実行しない。

【0203】

C E L Tにおいては、目標レベルへのフェードがないが、調性を有する封じ込め（フェードアウトを含む）の5フレームの後、6番目の連続して失われたフレームで、レベルは、瞬間的に目標レベルに切り替えられる。このレベルは、式（19）を用いて帯域毎に導出される。

【0204】

フェードアウトの目標スペクトル形状に関し、全ての分析した純粋な変換ベースのコーデック（AAC、G. 719、G. 722、G. 722. 1）及びSBRは、フェードアウト中の最後の良好なフレームのスペクトル形状を単に長くするだけである。

【0205】

様々な言語音声コーデックは、LPC合成を用いてスペクトル形状を平均へフェードさせる。平均は、静的（AMR）又は適応的（AMR-WB、AMR-WB+、G. 718）であり得るが、後者は、静的な平均及び短期的平均（最後のn個のLP係数セットの平均を取ることで導出される）（LP＝線形予測）から導出される。

【0206】

上述のコーデックAMR、AMR-WB、AMR-WB+、G. 718における全てのCNGモジュールは、フェードアウト中の最後の良好なフレームのスペクトル形状を長くする。

10

【0207】

背景ノイズレベルトレースに関し、文献から5つの異なる方策が知られている。

【0208】

- 発話区間検出部ベース：SNR/VADに基づくが、調性が極めて困難であり、低SNR言語音声の場合に使いにくい。

【0209】

- ソフト決定方式：ソフト決定方策は、言語音声の存在の可能性を考慮に入れる [SS98] [MPC89] [HE95]。

20

【0210】

- 最小統計：時間に亘ってバッファ内に或る一定量の値を保持しながらPSDの最小値を追跡することによって、最小ノイズを過去のサンプルから求めることを可能にする [Mar01] [HHJ10] [EH08] [Yu09]。

【0211】

- カルマンフィルタリング：アルゴリズムは、時間に亘って観察された、ノイズ（ランダムな変動）を含む一連の測定値を用いて、単一の測定だけにに基づくものよりも正確である傾向があるノイズPSDの推定値を生成する。カルマンフィルタは、ノイズの多い入力データのストリームに対して再帰的に働き、システム状態の統計的に最適な推定値を生成する [Gan05] [BJH06]。

30

【0212】

- 部分空間分解：この方策は、例えば、KLT（カルフネン・ロエヴ（Karhunen-Loeve）変換、また主成分分析としても知られる）及び／又はDFT（離散時間フーリエ変換）を利用して、ノイズ様の信号をクリーンな言語音声信号及びノイズ部分に分解することを試みる。それから、任意の平滑化アルゴリズムを用いて固有ベクトル／固有値をトレースすることができる [BP06] [HJH08]。

【先行技術文献】

【非特許文献】

【0213】

【非特許文献1】 [3GP09a] 3GPP、技術仕様グループサービス及びシステム局面（Technical Specification Group Services and System Aspects）、「拡張適応マルチレート-広帯域（AMR-WB+）コーデック（Extended adaptive multi-rate - wideband (AMR-WB+) codec）」、3GPP TS 26. 290、第3世代パートナーシッププロジェクト（3rd Generation Partnership Project）、2009年

40

【非特許文献2】 [3GP09b] 「拡張適応マルチレート-広帯域（AMR-WB+）コーデック；浮動小数点ANSI-Cコード（Extended adaptive multi-rate - wideband (AMR-WB+) codec; floating-point ANSI-C code）」、3GPP TS 26. 304、第3世代パートナーシッププロジェクト（3rd Generation Partnership Project）、2009年

【非特許文献3】 [3GP09c] 「言語音声コーデック言語音声処理機能；適応マルチ

50

レートー広帯域 (AMRWB) 言語音声コーデック; トランスコード処理機能 (Speech codec speech processing functions; adaptive multi-rate - wideband (AMRWB) speech codec; transcoding functions)」、3GPP TS 26.190、第3世代パートナーシッププロジェクト (3rd Generation Partnership Project)、2009年

【非特許文献4】 [3GPP 12 a] 「適応マルチレート (AMR) 言語音声コーデック; 失われたフレームのエラー封じ込め (第11号) (Adaptive multi-rate (AMR) speech codec; error concealment of lost frames (release 11))」、3GPP TS 26.091、第3世代パートナーシッププロジェクト (3rd Generation Partnership Project)、2012年9月

【非特許文献5】 [3GPP 12 b] 「適応マルチレート (AMR) 言語音声コーデック; トランスコード処理機能 (第11号) (Adaptive multi-rate (AMR) speech codec; transcoding functions (release 11))」、3GPP TS 26.090、第3世代パートナーシッププロジェクト (3rd Generation Partnership Project)、2012年9月

【非特許文献6】 [3GPP 12 c] 「適応マルチレートー広帯域のためのANSI-Cコード (AMR-WB) 言語音声コーデック (ANSI-C code for the adaptive multi-rate - wideband (AMR-WB) speech codec)」、3GPP TS 26.173、第3世代パートナーシッププロジェクト (3rd Generation Partnership Project)、2012年9月

【非特許文献7】 [3GPP 12 d] 「浮動小数点適応マルチレート (AMR) 言語音声コーデックのためのANSI-Cコード (第11号) (ANSI-C code for the floating-point adaptive multi-rate (AMR) speech codec (release 11))」、3GPP TS 26.104、第3世代パートナーシッププロジェクト (3rd Generation Partnership Project)、2012年9月

【非特許文献8】 [3GPP 12 e] 一般的な音声コーデック音声処理機能: 高度aacプラス一般音声コーデック; 追加のデコーダツール (第11号) (General audio codec audio processing functions; Enhanced aacPlus general audio codec; additional decoder tools (release 11))」、3GPP TS 26.402、第3世代パートナーシッププロジェクト (3rd Generation Partnership Project)、2012年9月

【非特許文献9】 [3GPP 12 f] 「言語音声コーデック言語音声処理機能; 適応マルチレートー広帯域 (amr-wb) 言語音声コーデック; ansi-cコード (Speech codec speech processing functions; adaptive multi-rate - wideband (amr-wb) speech codec; ansi-c code)」、3GPP TS 26.204、第3世代パートナーシッププロジェクト (3rd Generation Partnership Project)、2012年

【非特許文献10】 [3GPP 12 g] 「言語音声コーデック言語音声処理機能; 適応マルチレートー広帯域 (AMR-WB) 言語音声コーデック; 誤りを含む又は失われたフレームのエラー封じ込め (Speech codec speech processing functions; adaptive multi-rate - wideband (AMR-WB) speech codec; error concealment of erroneous or lost frames)」、3GPP TS 26.191、第3世代パートナーシッププロジェクト (3rd Generation Partnership Project)、2012年9月

【非特許文献11】 [BJH06] I. バティーナ (Batina)、J. イェンゼン (Jensen)、R. ホイスデンス (Heusdens)、「言語音声パワースペクトルダイナミクスのための自動回帰モデルを用いた言語音声強調のためのノイズパワースペクトル推定 (Noise power spectrum estimation for speech enhancement using an autoregressive model for speech power spectrum dynamics)、第3回IEEE国際音響・言語音声・信号処理会議集録 (Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. 3)、(2006年)、1064~1067頁

【非特許文献12】 [BP06] A. ボロヴィッツ (Borowicz)、A. ペトロフスキー (Petrovsky)、「kltベースの言語音声強調のための最小制御ノイズ推定 (Minimal controlled noise estimation for klt-based speech enhancement)、CD-ROM、2006年、イタリア、フィレンツェ

【非特許文献13】 [Coh03] I. コーエン (Cohen)、「不利な環境でのノイズス

10

20

30

40

50

ペクトル推定：向上した最小制御再帰的平均化 (Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging)、I E E E 言語音声・音声処理会報 (IEEE Trans. Speech Audio Process) 第 1 1 巻 (2 0 0 3 年)、第 5 号、4 6 6 ~ 4 7 5 頁

【非特許文献 1 4】 [C P K 0 8] チョン・サン・チョ (Choong Sang Cho)、ナム・イン・パク (Nam In Park)、ホン・クック・キム (Hong Kook Kim)、「c e l p タイプの言語音声コードについてのバースト状のパケット損失に対してロバストなパケット損失封じ込めアルゴリズム (A packet loss concealment algorithm robust to burst packet loss for celp- type speech coders)」、韓国電子工学技術研究所技術レポート、グワン科学技術研究所、2 0 0 8 年、第 2 3 回国際回路・システム・コンピュータ・通信技術会議 (Tech. report, Korea Electronics Technology Institute, Gwang Institute of Science and Technology, 2008, The 23rd International Technical Conference on Circuits/Systems, Computers and Communications) (I T C - C S C C 2 0 0 8 年)

【非特許文献 1 5】 [D o b 9 5] G. ドー布林ガー (Doblinger)、「サブバンドにおけるスペクトル最小追跡による計算上効率的な言語音声強調 (Computationally efficient speech enhancement by spectral minima tracking in subbands)、ユーロスピーチ集録 (Proc. Eurospeech) (1 9 9 5 年)、1 5 1 3 ~ 1 5 1 6 頁

【非特許文献 1 6】 [E B U 1 0] E B U / E T S I J T C ブロードキャスト (EBU/ETSI JTC Broadcast)、「デジタル音声ブロードキャスト (D A B) ; 高度音声符号化 (A A C) 音声のトランスポート (Digital audio broadcasting (DAB); transport of advanced audio coding (AAC) audio)」、E T S I T S 1 0 2 5 6 3、欧州放送同盟 (European Broadcasting Union)、2 0 1 0 年 5 月 [E B U 1 2] 全世界デジタル無線 (D R M) ; システム仕様 (Digital radio mondiale (DRM); system specification)」、E T S I E S 2 0 1 9 8 0、E T S I、2 0 1 2 年 6 月

【非特許文献 1 7】 [E H 0 8] ヤーン・S. エルケレンス (Jan S. Erkelens)、リヒャルト・ホイスデンス (Richards Heusdens)、「データを用いた再帰的ノイズパワー推定に基づく非定常ノイズの追跡 (Tracking of Nonstationary Noise Based on Data-Driven Recursive Noise Power Estimation)」、音声・言語音声・言語処理、I E E E 会報 (Audio, Speech, and Language Processing, IEEE Transactions) 第 1 6 巻 (2 0 0 8 年)、第 6 号、1 1 1 2 ~ 1 1 2 3 頁

【非特許文献 1 8】 [E M 8 4] Y. エフライム (Ephraim)、D. マーラー (Malah)、「最小平均平方エラー短時間スペクトル振幅推定器を用いた言語音声強調 (Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator)、I E E E 音響・言語音声・信号処理会報 (IEEE Trans. Acoustics, Speech and Signal Processing) 第 3 2 巻 (1 9 8 4 年)、第 6 号、1 1 0 9 ~ 1 1 2 1 頁

【非特許文献 1 9】 [E M 8 5]「最小平均平方エラーログスペクトル振幅推定器を用いた言語音声強調 (Speech enhancement using a minimum mean-square error log-spectral amplitude estimator)、I E E E 音響・言語音声・信号処理会報 (IEEE Trans. Acoustics, Speech and Signal Processing) 第 3 3 巻 (1 9 8 5 年)、4 4 3 ~ 4 4 5 頁

【非特許文献 2 0】 [G a n 0 5] S. ガノート (Gannot)、「言語音声強調：推定最大 (e m フレームワーク) におけるカルマンフィルタの適用 (Speech enhancement: Application of the kalman filter in the estimate-maximize (em framework))、シュプリンガー (Springer)、2 0 0 5 年

【非特許文献 2 1】 [H E 9 5] H. G. ヒルシュ (Hirsch)、C. エーアリッヒャー (Ehrlicher)、「ロバストな言語音声認識のためのノイズ推定技術 (Noise estimation techniques for robust speech recognition)」、国際 I E E E 音響・言語音声・信号処理会議集録 (Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing)、1 5 3 ~ 1 5 6 頁、I E E E、1 9 9 5 年

【非特許文献 2 2】 [H H J 1 0] リヒャルト・C. ヘンドリクス (Richard C. Hendriks)、リヒャルト・ホイスデンス (Richard Heusdens)、イエスパー・イエンゼン (Jespe

10

20

30

40

50

r Jensen)、「低複雑度のMMSEベースのノイズPSD追跡(MMSE based noise PSD tracking with low complexity)」、音響言語音声及び信号処理(ICASSP)(Acoustics Speech and Signal Processing (ICASSP))、2010年IEEE国際会議(2010 IEEE International Conference)、2010年5月、4266～4269頁

【非特許文献23】[HJH08]リヒャルト・C.ヘンドリクス(Richard C. Hendriks)、イエスパー・イエンゼン(Jesper Jensen)、リヒャルト・ホイスデンス(Richard Heusdens)、「dft領域サブスペース分解を用いたノイズ追跡(Noise tracking using dft domain subspace decompositions)」、IEEE音声・言語音声・言語処理会報(IEEE Trans. Audio, Speech, Lang. Process.)第16巻(2008年)、第3号、541～553頁

10

【非特許文献24】[IET12]IETF、「Opus音声コーデックの定義(Definition of the Opus Audio Codec)」、技術レポートRFC6716、インターネット工學タスクフォース(Tech. Report RFC 6716, Internet Engineering Task Force)、2012年9月

【非特許文献25】[ISO09]ISO/IEC JTC1/SC29/WG11、「情報技術—音声視覚対象の符号化—第3部：音声(Information technology — coding of audio-visual objects — part 3: Audio)」、ISO/IEC IS 14496-3、国際標準化機構(International Organization for Standardization)、2009年

【非特許文献26】[ITU03]ITU-T、「適応マルチレート広帯域(amr-wb)を用いた約16kbit/sでの言語音声の広帯域符号化(Wideband coding of speech at around 16 kbit/s using adaptive multi-rate wideband (amr-wb))」、勧告(Recommendation)ITU-T G.722.2、ITU電気通信標準化セクター(Telecommunication Standardization Sector of ITU)、2003年7月

20

【非特許文献27】[ITU05]「低フレーム損失のシステムにおけるハンドフリー動作のための24・32kbit/sでの低複雑度符号化(Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss)」、勧告(Recommendation)ITU-T G.722.1、ITU電気通信標準化セクター(Telecommunication Standardization Sector of ITU)2005年5月

【非特許文献28】[ITU06a]「G.722付録III：G.722のためのパケット損失封じ込めのための高複雑度のアルゴリズム(G.722 Appendix III: A high-complexity algorithm for packet loss concealment for G.722)」、ITU-T勧告(Recommendation)、ITU-T、2006年11月

30

【非特許文献29】[ITU06b]「G.729.1：G.729ベースの埋め込み可変ビットレートコーダ：g.729と相互動作可能な8～32kbit/sスケール可能広帯域コーダビットストリーム(G.729.1: G.729-based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with g.729)」、勧告(Recommendation)ITU-T G.729.1、ITU電気通信標準化セクター(Telecommunication Standardization Sector of ITU)、2006年5月

【非特許文献30】[ITU07]「G.722付録IV：G.722によるパケット損失封じ込めのための低複雑度のアルゴリズム(G.722 Appendix IV: A low-complexity algorithm for packet loss concealment with G.722)」、ITU-T勧告(Recommendation)、ITU-T、2007年8月

40

【非特許文献31】[ITU08a]「G.718：8～32kbit/sからの言語音声及び音声のフレームエラーロバストな狭帯域及び広帯域埋め込み可変ビットレート符号化(G.718: Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s)」、勧告(Recommendation)ITU-T G.718、ITU電気通信標準化セクター(Telecommunication Standardization Sector of ITU)、2008年6月

【非特許文献32】[ITU08b]「G.719：高品質会話用途のための低複雑度全帯域音声符号化(G.719: Low-complexity, full-band audio coding for high-quality,

50

conversational applications)」、勧告 (Recommendation) ITU-T G.719、ITU 電気通信標準化セクター (Telecommunication Standardization Sector of ITU)、2008 年 6 月

【非特許文献 33】[ITU12]「G.729: 共役構造の代数的符号励振線形予測 (cs-acelp) を用いた 8 kbit/秒での言語音声の符号化 (G.729: Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (cs-acelp))」、勧告 (Recommendation) ITU-T G.729、ITU 電気通信標準化セクター (Telecommunication Standardization Sector of ITU)、2012 年 6 月

【非特許文献 34】[LS01] ピエール・ローベール (Pierre Lauber)、ラルフ・シュペルシュナイダー (Ralph Sperschneider)、「圧縮デジタル音声のためのエラー封じ込め (Error concealment for compressed digital audio)」、第 111 回音声工学協会会議 (Audio Engineering Society Convention 111)、第 5460 号、2001 年 9 月

【非特許文献 34】[Mar01] ライナー・マルティン (Rainer Martin)、「最適平滑化及び最小統計に基づくノイズパワースペクトル密度推定 (Noise power spectral density estimation based on optimal smoothing and minimum statistics)」、IEEE 言語音声・音声処理会報 (IEEE Transactions on Speech and Audio Processing) 第 9 巻 (2001 年)、第 5 号、504～512 頁

【非特許文献 35】[Mar03]「ノイズの多い言語音声の強調のための統計的方法 (Statistical methods for the enhancement of noisy speech)」、国際音響エコー・ノイズ制御ワークショップ (International Workshop on Acoustic Echo and Noise Control) (IWAENC2003)、ブラウンシュヴァイク工科大学 (Technical University of Braunschweig)、2003 年 9 月

【非特許文献 36】[MC99] R. マルティン (Martin)、R. コックス (Cox)、「低ビットレート言語音声符号化のための新規の言語音声強調技術 (New speech enhancement techniques for low bit rate speech coding)」、IEEE 言語音声符号化ワークショップ集録 (Proc. IEEE Workshop on Speech Coding) (1999 年)、165～167 頁

【非特許文献 37】[MCA99] D. マーラー (Malah)、R. V. コックス (Cox)、A. J. アッカルディ (Accardi)、「非定常ノイズ環境において言語音声強調を向上させるための言語音声の存在の不確実性の追跡 (Tracking speech-presence uncertainty to improve speech enhancement in nonstationary noise environments)」、国際 IEEE 音響・言語音声・信号処理会議集録 (Proc. IEEE Int. Conf. on Acoustics Speech and Signal Processing) (1999 年)、789～792 頁

【非特許文献 38】[MEP01] ニコラウス・マイネ (Nikolaus Meine)、ベルント・エードラー (Bernd Edler)、ハイコ・プルンハーゲン (Heiko Purnhagen)、「HILN MPEG-4 パラメータ的音声符号化のためのエラー保護及び封じ込め (Error protection and concealment for HILN MPEG-4 parametric audio coding)」、第 110 回音声工学協会会議 (Audio Engineering Society Convention 110)、第 5300 号、2001 年 5 月

【非特許文献 39】[MPC89] Y. マイユー (Mahieux)、J. -P. プチ (Petit)、A. シャルボニエ (Charbonnier)、「連続する変換ブロック間の相関を用いた音声信号の変換符号化 (Transform coding of audio signals using correlation between successive transform blocks)」、音響・言語音声・信号処理 (Acoustics, Speech, and Signal Processing)、1989 年、ICASSP-89、1989 年国際会議 (1989 International Conference)、1989 年、2021～2024 頁、第 3 巻

【非特許文献 40】[NMR+12] マックス・ノイエンドルフ (Max Neuendorf)、マルクス・ムルトルス (Markus Multrus)、ニコラウス・レッテルバッハ (Nikolaus Rettelbach)、ギヨーム・フックス (Guillaume Fuchs)、ジュリアン・ロビヤール (Julien Robilliard)、ジェレミー・ルコント (Jeremie Lecomte)、シュテファン・ヴィルデ (St

10

20

30

40

50

ephan Wilde)、シュテファン・バイヤー (Stefan Bayer)、ザシャ・ディッシュ (Sascha Disch)、クリスティアン・ヘルムリッヒ (Christian Helmrich)、ロッシュ・ルフェーブル (Roch Lefebvre)、フィリップ・グルネー (Philippe Gournay)、ブルーノ・ベセッテ (Bruno Bessette)、ジミー・ラピエール (Jimmy Lapierre)、クリストファー・キェルリング (Kristopfer Kjørling)、ハイコ・プルンハーゲン (Heiko Purnhagen)、ラルス・ヴィルモエス (Lars Villemoes)、ヴェルナー・オーメン (Werner Oomen)、エリック・シャイヤース (Erik Schuijers)、ケイ・キクイリ (Kei Kikuri)、トール・チネン (Toru Chinen)、タケシ・ノリマツ (Takeshi Norimatsu)、チョン・コク・セン (Chong Kok Seng)、ユーンミ・オー (Eunmi Oh)、ミヨン・キム (Miyoung Kim)、シュイラー・クヴァッケンブッシュ (Schuyler Quackenbush)、ベルンハルト・グリル (Bernhard Grill)、
「MPEG 統一言語音声及び音声符号化—全てのコンテンツタイプの高効率音声符号化のための ISO/MPEG 標準規格 (MPEG Unified Speech and Audio Coding - The ISO / MPEG Standard for High-Efficiency Audio Coding of all Content Types)、会議録 (Convention Paper)、8654 頁、AES、2012 年 4 月、第 132 回会議、ブダペスト、ハンガリー (Presented at the 132nd Convention Budapest, Hungary)

10

【非特許文献 41】 [PKJ+11] ナム・イン・パク (Nam In Park)、ホン・クック・キム (Hong Kook Kim)、ミン・ア・ジュン (Min A Jung)、ソン・ロ・リー (Seong Ro Lee)、ソン・ホ・チョイ (Seung Ho Choi)、
「無線センサネットワークにおける celp タイプの言語音声コーダのための多数のコードブック及び快適ノイズを用いたバースト状のパケット損失の封じ込め (Burst packet loss concealment using multiple codebooks and comfort noise for celp-type speech coders in wireless sensor networks)」、センサーズ (Sensors) 第 11 巻 (2011)、5323~5336 頁

20

【非特許文献 42】 [QD03] シュイラー・クヴァッケンブッシュ (Schuyler Quackenbush)、ペーター・F. ドリーセン (Peter F. Driessen)、
「MPEG-4 音声パケット通信システムにおけるエラー緩和 (Error mitigation in MPEG-4 audio packet communication systems)」、第 115 回音声工学協会会議 (Audio Engineering Society Convention 115)、第 5981 号、2003 年 10 月

【非特許文献 43】 [RL06] S. ランガカーリ (Rangachari)、P. C. ロイズ (Loizou)、
「高度に非定常な環境のためのノイズ推定アルゴリズム (A noise-estimation algorithm for highly non-stationary environments)」、言語音声通信 (Speech Commun.) 第 48 巻 (2006 年)、220~231 頁

30

【非特許文献 44】 [SFB00] V. シュタール (Stahl)、A. フィッシャー (Fischer)、R. ビップス (Bippus)、
「スペクトル減算及びウィーナフィルタリングのためのクオントイルベースのノイズ推定 (Quantile based noise estimation for spectral subtraction and wiener filtering)、国際 IEEE 音響・言語音声・信号処理会議集録 (Proc. IEEE Int. Conf. Acoust., Speech and Signal Process) (2000 年)、1875~1878 頁

【非特許文献 45】 [SS98] J. ザーン (Sohn)、W. ズング (Sung)、
「ソフト決定ベースのノイズスペクトル適応を用いた音声活動検出器 (A voice activity detector employing soft decision based noise spectrum adaptation)」、国際 IEEE 音響・言語音声・信号処理会議集録 (Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, no.)、365~368 頁、IEEE、1998 年

40

【非特許文献 46】 [Yu09] ロンシャン・ユー (Rongshan Yu)、
「ノイズパワー推定の平滑化及び推定バイアス訂正に基づく低複雑度のノイズ推定アルゴリズム (A low-complexity noise estimation algorithm based on smoothing of noise power estimation and estimation bias correction)」、音響・言語音声・信号処理 (Acoustics, Speech and Signal Processing)、2009 年、ICASSP 2009、IEEE 国際会議 (IEEE International Conference)、2009 年 4 月、4421~4424 頁

【発明の概要】

50

【発明が解決しようとする課題】

【0214】

本発明の目的は、音声符号化システムについての向上した概念を提供することである。本発明の目的は、請求項1に記載の装置、請求項12に記載の方法、請求項13に記載のコンピュータプログラムによって達成される。

【課題を解決するための手段】

【0215】

符号化音声信号を復号して再構成音声信号を得るための装置が提供される。当該装置は、複数のフレームを受信するための受信インターフェース、復号音声信号の音声信号サンプルを記憶するための遅延バッファ、遅延バッファに記憶されている音声信号サンプルから複数の選択音声信号サンプルを選択するためのサンプル選択部、及び、選択音声信号サンプルを処理して再構成音声信号の再構成音声信号サンプルを得るためのサンプル処理部を備える。サンプル選択部は、現在のフレームが受信インターフェースによって受信されており、且つ、受信インターフェースによって受信されている現在のフレームが破損していない場合、現在のフレームに含まれるピッチ遅延情報に応じて遅延バッファに記憶されている音声信号サンプルから複数の選択音声信号サンプルを選択するように構成される。更に、サンプル選択部は、現在のフレームが受信インターフェースによって受信されていない場合、又は受信インターフェースによって受信されている現在のフレームが破損している場合、受信インターフェースによって以前に受信されている別のフレームに含まれるピッチ遅延情報に応じて遅延バッファに記憶されている音声信号サンプルから複数の選択音声信号サンプルを選択するように構成される。

10

20

【0216】

実施例によると、サンプル処理部は、例えば、現在のフレームが受信インターフェースによって受信されており、且つ受信インターフェースによって受信されている現在のフレームが破損していない場合、現在のフレームに含まれるゲイン情報に応じて選択音声信号サンプルを再スケーリングすることによって再構成音声信号サンプルを得るように構成され得る。更に、サンプル選択部は、例えば、現在のフレームが受信インターフェースによって受信されていない場合、又は受信インターフェースによって受信されている現在のフレームが破損している場合、受信インターフェースによって以前に受信されている前記別のフレームに含まれるゲイン情報に応じて選択音声信号サンプルを再スケーリングすることによって再構成音声信号サンプルを得るように構成され得る。

30

【0217】

実施例においては、サンプル処理部は、例えば、現在のフレームが受信インターフェースによって受信されており、且つ、受信インターフェースによって受信されている現在のフレームが破損していない場合、選択音声信号サンプルと、現在のフレームに含まれるゲイン情報に応じた値とを乗算することによって再構成音声信号サンプルを得るように構成され得る。更に、サンプル選択部は、現在のフレームが受信インターフェースによって受信されていない場合、又は受信インターフェースによって受信されている現在のフレームが破損している場合、選択音声信号サンプルと、受信インターフェースによって以前に受信されている前記別のフレームに含まれるゲイン情報に応じた値とを乗算することによって再構成音声信号サンプルを得るように構成される。

40

【0218】

実施例によると、サンプル処理部は、例えば、再構成音声信号サンプルを遅延バッファに記憶させるように構成され得る。

【0219】

実施例においては、サンプル処理部は、例えば、更なるフレームが受信インターフェースによって受信される前に再構成音声信号サンプルを遅延バッファに記憶させるように構成され得る。

【0220】

実施例によると、サンプル処理部は、例えば、更なるフレームが受信インターフェース

50

によって受信された後に再構成音声信号サンプルを遅延バッファに記憶させるように構成され得る。

【0221】

実施例においては、サンプル処理部は、例えば、ゲイン情報に応じて選択音声信号サンプルを再スケーリングして再スケーリング音声信号サンプルを得て、再スケーリング音声信号サンプルを入力音声信号サンプルと組み合わせることにより処理済音声信号サンプルを得るように構成され得る。

【0222】

実施例によると、サンプル処理部は、例えば、現在のフレームが受信インターフェースによって受信されており、且つ、受信インターフェースによって受信されている現在のフレームが破損していない場合、再スケーリング音声信号サンプル及び入力音声信号サンプルの組み合わせを示す処理済音声信号サンプルを遅延バッファに記憶させ、再スケーリング音声信号サンプルを遅延バッファに記憶させないように構成され得る。更に、サンプル処理部は、現在のフレームが受信インターフェースによって受信されていない場合、又は受信インターフェースによって受信されている現在のフレームが破損している場合、再スケーリング音声信号サンプルを遅延バッファに記憶させ、処理済音声信号サンプルを遅延バッファに記憶させないように構成される。

10

【0223】

別の実施例によると、サンプル処理部は、例えば、現在のフレームが受信インターフェースによって受信されていない場合、又は受信インターフェースによって受信されている現在のフレームが破損している場合、処理済音声信号サンプルを遅延バッファに記憶させるように構成され得る。

20

【0224】

実施例においては、サンプル選択部は、例えば、変更後ゲインに応じて選択音声信号サンプルを再スケーリングすることによって再構成音声信号サンプルを得るように構成することができ、変更後ゲインは、

【0225】

【数33】

$gain = gain_past * damping;$

30

【0226】

の式によって定義され、ここで、 $gain$ は、変更後ゲインであり、サンプル選択部は、例えば、 $gain$ の算出後に $gain_past$ を $gain$ に設定するように構成することができ、 $damping$ は、実数値である。

【0227】

実施例によると、サンプル選択部は、例えば、変更後ゲインを計算するように構成され得る。

【0228】

実施例においては、 $damping$ は、例えば、 $0 \leq damping \leq 1$ によって定義され得る。

40

【0229】

実施例によると、変更後ゲイン $gain$ は、例えば、フレームが最後に受信インターフェースによって受信されてから少なくとも所定数のフレームが受信インターフェースによって受信されていない場合、ゼロに設定され得る。

【0230】

更に、符号化音声信号を復号して再構成音声信号を得るための方法が提供される。当該方法は、

- 複数のフレームを受信するステップと、
- 復号された音声信号の音声信号サンプルを記憶するステップと、

50

- 遅延バッファに記憶されている音声信号サンプルから複数の選択音声信号サンプルを選択するステップと、
- 選択音声信号サンプルを処理して再構成音声信号の再構成音声信号サンプルを得るステップと、を備える。

【0231】

現在のフレームが受信され、かつ、受信されている現在のフレームが破損していない場合、遅延バッファに記憶されている音声信号サンプルから複数の選択音声信号サンプルを選択するステップは、現在のフレームに含まれるピッチ遅延情報に応じて行われる。更に、現在のフレームが受信されていない場合、又は受信されている現在のフレームが破損している場合、遅延バッファに記憶されている音声信号サンプルから複数の選択音声信号サ

10

【0232】

更に、コンピュータ又は信号プロセッサにおいて実行された時に上述の方法を実現するためのコンピュータプログラムが提供される。

【0233】

実施例は、TCX LTP (TCX LTP=変換符号化励振長期予測)を採用する。通常動作中、ノイズと、再構成された調性成分とを含む合成信号でTCX LTPメモリが更新される。

【0234】

20

封じ込め中にTCX LTPを動作不能にする代わりに、その通常動作を、最後の良好なフレームにおいて受信されたパラメータによって封じ込め中に継続しても良い。これにより、信号のスペクトル形状、特に、LTPフィルタによってモデル化される調性成分が保管される。

【0235】

更に、実施例においては、TCX LTPフィードバックループが結合解除される。通常のTCX LTP演算を単純に継続させることによって追加のノイズが導入されるが、それは、各々の更新ステップによって、LTP励振からの更なるランダム生成ノイズが導入されるからである。従って、調性成分は、時間の経過に伴って、この追加のノイズによってますます歪められる。

30

【0236】

これを克服するために、更新されたTCX LTPバッファのみを(ノイズ追加なしに)フィードバックすることにより、望まれないランダムノイズによって調性情報が汚染されないようにすることができる。

【0237】

更に、実施例によると、TCX LTPゲインがゼロにフェードさせられる。

【0238】

これらの実施例は、TCX LTPを継続させることによって、信号特性を短期的に保管することが助けられるが、長期的には欠点があるという知見に基づいている。即ち、封じ込め中に再生される信号は、損失に先立って存在していた有声/調性情報を含む。特に、クリーンな言語音声又は背景ノイズ上の言語音声の場合、音調又は調波が長期間に亘って極めてゆっくりと減衰する可能性は非常に低い。特に、LTPメモリ更新が結合解除される(調性成分のみがフィードバックされ、符号がスクランブルされた部分はフィードバックされない)場合、封じ込め中にTCX LTP演算を継続することにより、有声/調性情報は、損失全体についての封じ込め信号において存在し続けることになり、全体的なフェードアウトによって快適ノイズレベルへ減衰されるのみである。更に、TCX LTPがバースト状の損失中に時間の経過に伴って減衰されることなく適用される場合、バースト状のパケット損失中に快適ノイズ包絡線に達することは不可能であるが、それは、その場合、信号がLTPの有声情報を常に組み込んでいるからである。

40

【0239】

50

従って、LTPによって表現される調性成分がゼロにフェードさせられ、同時に信号が背景信号レベル及び形状にフェードさせられ、フェードアウトが、望まれない調性成分を組み込むことなく、所望のスペクトル背景包絡線（快適ノイズ）に達するように、TCXLTPゲインが、ゼロへフェードさせられる。

【0240】

実施例においては、LTPゲインフェードについて、ホワイトノイズフェードと同じフェード速度が用いられる。

【0241】

これとは対照的に、先行技術では、封じ込め中にLTPを用いる変換コーデックは、知られていない。MPPEG-4 LTP [ISO09] の場合、先行技術において、封じ込め方策は、存在しない。LTPを利用する先行技術の別のMDCTベースのコーデックは、CELTであるが、このコーデックは、最初の5つのフレームについてACELP様の封じ込めを用い、後続のフレーム全てについては、LTPを利用しない背景ノイズを生成する。TCXLTPを用いない先行技術の欠点は、LTPでモデル化される全ての調性成分が突然消失することである。更に、先行技術のACELPベースのコーデックにおいては、LTP演算が封じ込め中に長くされ、適応コードブックのゲインは、ゼロへとフェードさせられる。フィードバックループ動作に関し、先行技術は、2つの方策を用いる。即ち、励振全体、例えば、新規励振と適応励振との合計をフィードバックする（AMR-WB）か、或いは、更新された適応励振、例えば調性信号部分のみをフィードバックする（G. 718）。上述の実施例は、先行技術の欠点を克服する。

【0242】

更に、音声信号を復号するための装置が提供される。

【0243】

当該装置は、受信インターフェースを備える。受信インターフェースは、複数のフレームを受信するように構成され、受信インターフェースは、複数のフレームのうちの第1のフレームを受信するように構成され、第1のフレームは、音声信号の第1の音声信号部分を含み、前記第1の音声信号部分は、第1の領域において表現され、受信インターフェースは、複数のフレームのうちの第2のフレームを受信するように構成され、第2のフレームは、音声信号の第2の音声信号部分を含む。

【0244】

更に、当該装置は、第2の音声信号部分、又は第2の音声信号部分から導出された値若しくは信号を、第2の領域からトレース領域に変換することによって第2の信号部分情報を得るための変換部を備え、第2の領域は、第1の領域とは異なり、トレース領域は、第2の領域とは異なり、トレース領域は、第1の領域と等しい又は異なる。

【0245】

更に、当該装置は、ノイズレベルトレース部を備え、ノイズレベルトレース部は、トレース領域において表現される第1の信号部分情報を受信するように構成され、第1の信号部分情報は、第1の音声信号部分に依存する。ノイズレベルトレース部は、トレース領域において表現される第2の信号部分を受信するように構成され、ノイズレベルトレース部は、トレース領域において表現される第1の信号部分情報と、トレース領域において表現される第2の信号部分情報とに応じてノイズレベル情報を決定するように構成される。

【0246】

更に、当該装置は、複数のフレームのうちの第3のフレームが受信インターフェースによって受信されていないが破損している場合、ノイズレベル情報に応じて音声信号の第3の音声信号部分を再構成するための再構成部を備える。

【0247】

音声信号は、例えば、言語音声信号、音楽信号、又は、言語音声及び音楽等を含む信号であり得る。

【0248】

第1の信号部分情報が第1の音声信号部分に依存するという記載は、第1の信号部分情

10

20

30

40

50

報が第1の音声信号部分であるか、又は、第1の信号部分情報が、第1の音声信号部分に応じて入手／生成されているか、若しくは或る他の態様で第1の音声信号部分に依存することを意味する。例えば、第1の音声信号部分は、第1の信号部分情報を得るために1つの領域から別の領域に変換されている場合がある。

【0249】

同様に、第2の信号部分情報が第2の音声信号部分に依存するという記載は、第2の信号部分情報が第2の音声信号部分であるか、又は、第2の信号部分情報が、第2の音声信号部分に応じて入手／生成されているか、若しくは或る他の態様で第2の音声信号部分に依存することを意味する。例えば、第2の音声信号部分は、第2の信号部分情報を得るために1つの領域から別の領域に変換されている場合がある。

10

【0250】

実施例においては、第1の音声信号部分は、例えば、第1の領域として時間領域において表現され得る。更に、変換部は、例えば、第2の音声信号部分、又は第2の音声信号部分から導出された値を、第2の領域である励振領域から、トレース領域である時間領域へ変換するように構成され得る。更に、ノイズレベルトレース部は、例えば、トレース領域としての時間領域において表現される第1の信号部分情報を受信するように構成され得る。更に、ノイズレベルトレース部は、例えば、トレース領域としての時間領域において表現される第2の信号部分を受信するように構成され得る。

【0251】

実施例によると、第1の音声信号部分は、例えば、第1の領域としての励振領域において表現され得る。更に、変換部は、例えば、第2の音声信号部分、又は第2の音声信号部分から導出された値を、第2の領域である時間領域から、トレース領域である励振領域に変換するように構成され得る。更に、ノイズレベルトレース部は、例えば、トレース領域としての励振領域において表現される第1の信号部分情報を受信するように構成され得る。更に、ノイズレベルトレース部は、例えば、トレース領域としての励振領域において表現される第2の信号部分を受信するように構成され得る。

20

【0252】

実施例においては、第1の音声信号部分は、例えば、第1の領域としての励振領域において表現することができ、ノイズレベルトレース部は、例えば、第1の信号部分情報を受信するように構成することができ、第1の信号部分情報は、トレース領域であるFFT領域において表現され、前記第1の信号部分情報は、励振領域において表現される第1の音声信号部分に依存し、変換部は、例えば、第2の音声信号部分、又は第2の音声信号部分から導出された値を、第2の領域である時間領域から、トレース領域であるFFT領域に変換するように構成することができ、ノイズレベルトレース部は、例えば、FFT領域において表現される第2の音声信号部分を受信するように構成され得る。

30

【0253】

実施例においては、当該装置は、例えば、更に、第1の音声信号部分に応じて第1の集約値を決定するための第1の集約部を備え得る。更に、当該装置は、例えば、更に、第2の音声信号部分に応じて、第2の音声信号部分から導出された値としての第2の集約値を決定するための第2の集約部を備え得る。更に、ノイズレベルトレース部は、例えば、トレース領域において表現される第1の信号部分情報として第1の集約値を受信するように構成することができ、ノイズレベルトレース部は、例えば、トレース領域において表現される第2の信号部分情報として第2の集約値を受信するように構成することができ、ノイズレベルトレース部は、例えば、トレース領域において表現される第1の集約値と、トレース領域において表現される第2の集約値とに応じて、ノイズレベル情報を決定するように構成され得る。

40

【0254】

実施例によると、第1の集約部は、例えば、第1の集約値が、第1の音声信号部分、又は第1の音声信号部分から導出された信号の、二乗平均平方根を示すように、第1の集約値を決定するように構成され得る。更に、第2の集約部は、例えば、第2の集約値が、第

50

2の音声信号部分、又は第2の音声信号部分から導出された信号の、二乗平均平方根を示すように、第2の集約値を決定するように構成され得る。

【0255】

実施例においては、変換部は、例えば、第2の音声信号部分から導出された値にゲイン値を適用することにより、第2の音声信号部分から導出された値を、第2の領域からトレース領域に変換するように構成され得る。

【0256】

実施例によると、ゲイン値は、例えば、線形予測符号化合成によって導入されたゲインを示すか、或いは、ゲイン値は、例えば、線形予測符号化合成及び強調解除によって導入されたゲインを示すことができる。

10

【0257】

実施例においては、ノイズレベルトレース部は、例えば、最小統計方策を適用することによってノイズレベル情報を決定するように構成され得る。

【0258】

実施例によると、ノイズレベルトレース部は、例えば、ノイズレベル情報として快適ノイズレベルを決定するように構成され得る。再構成部は、例えば、複数のフレームのうちの前記第3のフレームが受信インターフェースによって受信されていない場合、又は前記第3のフレームが受信インターフェースによって受信されているが破損している場合、ノイズレベル情報に応じて第3の音声信号部分を再構成するように構成され得る。

20

【0259】

実施例においては、ノイズレベルトレース部は、例えば、ノイズレベルスペクトルから導出されたノイズレベル情報として快適ノイズレベルを決定するように構成することができ、前記ノイズレベルスペクトルは、最小統計方策を適用することによって得られる。再構成部は、例えば、複数のフレームのうちの前記第3のフレームが受信インターフェースによって受信されていない場合、又は前記第3のフレームが受信インターフェースによって受信されているが破損している場合、複数の線形予測係数に応じて第3の音声信号部分を再構成するように構成することができる。

【0260】

別の実施例によると、ノイズレベルトレース部は、例えば、ノイズレベル情報として快適ノイズレベルを示す複数の線形予測係数を決定するように構成することができ、再構成部は、例えば、複数の線形予測係数に応じて第3の音声信号部分を再構成するように構成することができる。

30

【0261】

実施例においては、ノイズレベルトレース部は、ノイズレベル情報として快適ノイズレベルを示す複数のFFT係数を決定するように構成され、第1の再構成部は、複数のフレームのうちの前記第3のフレームが受信インターフェースによって受信されていない場合、又は前記第3のフレームが受信インターフェースによって受信されているが破損している場合、前記FFT係数から導出された快適ノイズレベルに応じて第3の音声信号部分を再構成するように構成される。

【0262】

40

実施例においては、再構成部は、例えば、複数のフレームのうちの前記第3のフレームが受信インターフェースによって受信されていない場合、又は前記第3のフレームが受信インターフェースによって受信されているが破損している場合、ノイズレベル情報と第1の音声信号部分とに応じて第3の音声信号部分を再構成するように構成され得る。

【0263】

実施例によると、再構成部は、例えば、第1の音声信号部分又は第2の音声信号部分から導出された信号を減衰させる又は増幅することによって、第3の音声信号部分を再構成するように構成され得る。

【0264】

実施例においては、当該装置は、例えば、更に、遅延バッファを含む長期予測部を備え

50

ることができる。更に、長期予測部は、例えば、第1の音声信号部分又は第2の音声信号部分と、遅延バッファに記憶されている遅延バッファ入力と、長期予測ゲインとに応じて、処理済信号を生成するように構成され得る。更に、長期予測部は、例えば、複数のフレームのうちの前記第3のフレームが受信インターフェースによって受信されていない場合、又は前記第3のフレームが受信インターフェースによって受信されているが破損している場合、長期予測ゲインをゼロヘフェードさせるように構成され得る。

【0265】

実施例によると、長期予測部は、例えば、長期予測ゲインをゼロヘフェードさせるように構成することができ、長期予測ゲインがゼロヘフェードさせられる速度は、フェードアウト定数に依存する。

10

【0266】

実施例においては、長期予測部は、例えば、複数のフレームのうちの前記第3のフレームが受信インターフェースによって受信されていない場合、又は前記第3のフレームが受信インターフェースによって受信されているが破損している場合、生成された処理済信号を遅延バッファに記憶させることによって遅延バッファ入力を更新するように構成され得る。

【0267】

実施例によると、変換部は、例えば、第1の変換部とすることができ、再構成部は、第1の再構成部である。当該装置は、更に、第2の変換部及び第2の再構成部を備える。第2の変換部は、例えば、複数のフレームのうちの第4のフレームが受信インターフェースによって受信されていない場合、又は前記第4のフレームが受信インターフェースによって受信されているが破損している場合、ノイズレベル情報をトレース領域から第2の領域に変換するように構成され得る。更に、前記第2の再構成部は、例えば、複数のフレームのうちの前記第4のフレームが受信インターフェースによって受信されていない場合、又は前記第4のフレームが受信インターフェースによって受信されているが破損している場合、第2の領域において表現されているノイズレベル情報に応じて音声信号の第4の音声信号部分を再構成するように構成され得る。

20

【0268】

実施例においては、第2の再構成部は、例えば、ノイズレベル情報と第2の音声信号部分とに応じて第4の音声信号部分を再構成するように構成され得る。

30

【0269】

実施例によると、第2の再構成部は、例えば、第1の音声信号部分又は第2の音声信号部分から導出された信号を減衰させる又は増幅することによって第4の音声信号部分を再構成するように構成され得る。

【0270】

更に、音声信号を復号するための方法が提供される。

【0271】

当該方法は、

- 複数のフレームのうちの第1のフレームを受信するステップを備え、前記第1のフレームは、前記音声信号の第1の音声信号部分を含み、前記第1の音声信号部分は、第1の領域において表現され、当該方法は、更に、

40

- 複数のフレームのうちの第2のフレームを受信するステップを備え、前記第2のフレームは、音声信号の第2の音声信号部分を含み、当該方法は、更に、

- 第2の音声信号部分、又は第2の音声信号部分から導出された値又は信号を、第2の領域からトレース領域に変換して第2の信号部分情報を得るステップを備え、第2の領域は、第1の領域とは異なり、トレース領域は、第2の領域とは異なり、トレース領域は、第1の領域と等しい又は異なり、当該方法は、更に、

- トレース領域において表現される第1の信号部分情報と、トレース領域において表現される第2の信号部分情報とに応じてノイズレベル情報を決定するステップを備え、第1の信号部分情報は、第1の音声信号部分に依存し、当該方法は、更に、

50

- 複数のフレームのうちの第3のフレームが受信されていない場合、又は前記第3のフレームが受信されているが破損している場合、トレース領域において表現されるノイズレベル情報に応じて音声信号の第3の音声信号部分を再構成するステップを備える。

【0272】

更に、コンピュータ又は信号プロセッサにおいて実行された時に上述の方法を実現するためのコンピュータプログラムが提供される。

【0273】

本発明のいくつかの実施例は、時間可変平滑化パラメータを提供し、これにより平滑化されたペリオドグラムの追跡能力と、その分散とのバランスがより良好に取られ、バイアス補償のためのアルゴリズムを開発し、ノイズ追跡全般の速度を向上させる。

10

【0274】

本発明の実施例は、フェードアウトに関し、以下のパラメータが関心の対象であるという知見に基づく。即ち、フェードアウト領域、フェードアウト速度又はより一般的にフェードアウト曲線、フェードアウトの目標レベル、フェードアウトの目標スペクトル形状、及び／又は背景ノイズレベルトレースである。この文脈で、実施例は、先行技術が顕著な欠陥を有しているという知見に基づく。

【0275】

エラー封じ込め中に切り替え音声符号化システムのための向上した信号フェードアウトのための装置及び方法が提供される。

【0276】

20

更に、コンピュータ又は信号プロセッサにおいて実行された時に上述の方法を実現するためのコンピュータプログラムが提供される。

【0277】

実施例は、快適ノイズレベルへのフェードアウトを実現する。実施例によると、励振領域における共通の快適ノイズレベルトレースが実現される。バースト状の packets 損失中に目標とされる快適ノイズレベルは、使用されているコアコーダ (ACELP/TCX) に関わらず同じであり、常に最新のものである。共通のノイズレベルトレースが必要な先行技術は知られていない。実施例は、バースト状の packets 損失中に、快適ノイズ様の信号への切り替えコーデックのフェードを提供する。

【0278】

30

更に、実施例は、機能 (PROM) 及びメモリが共有され得るため、2つの独立したノイズレベルトレースモジュールを有する場合と比較して全体の複雑度が低くなることを実現する。

【0279】

実施例においては、(時間領域におけるレベル導出と比較して) 励振領域におけるレベル導出は、言語音声情報の一部がLP係数によってカバーされるため、アクティブな言語音声により多くの最小値をもたらす。

【0280】

ACELPの場合、実施例によると、レベル導出は、励振領域で行われる。TCXの場合、実施例においては、レベルは、時間領域において導出され、LPC合成及び強調解除のゲインを訂正定数として適用することによって、励振領域におけるエネルギーレベルをモデル化する。励振領域におけるレベルをトレースする、例えば、FDNS前にこれを行うことは、理論的には可能であるが、TCX励振領域とACELP励振領域との間のレベル補償は、むしろ複雑と考えられる。

40

【0281】

先行技術で、このような異なる領域における共通の背景レベルトレースを組み込むものはない。先行技術は、切り替えコーデックシステムにおいて、例えば、励振領域における、そのような共通の快適ノイズレベルトレースを有さない。従って、先行技術では、バースト状の packets 損失中に目標とされる快適ノイズレベルは、レベルがトレースされた先行する符号化モード (ACELP/TCX) に応じて異なり得るものであり、先行技術で

50

は、各々の符号化モードについて別個のトレースは、不必要なオーバーヘッド及び追加の計算上の複雑度を引き起こすものであり、そして、先行技術では、いずれのコアにおいても、このコアへ最近切り替わったため最新の快適ノイズレベルは利用可能でない可能性があるため、実施例は、先行技術よりも有利である。

【0282】

いくつかの実施例によると、レベルトレースは、励振領域において行われるが、TCXフェードアウトは、時間領域において行われる。時間領域におけるフェードにより、エイリアシングを引き起こすTDACの失敗が回避される。これに対する関心は、調性を有する信号成分を封じ込める際に特に大きい。更に、ACELP励振領域とMDC Tスペクトル領域との間のレベル変換が回避されるため、例えば、計算リソースが節約される。励振領域と時間領域との間の切り替えのため、励振領域と時間領域との間でのレベル調節が必要となる。これを解決するために、LPC合成及び前強調によって導入されるゲインを導出し、このゲインを訂正定数として用いて2領域間でレベルを変換する。

10

【0283】

これとは対照的に、先行技術においては、励振領域におけるレベルトレース及び時間領域におけるTCXフェードアウトは、行われず。先行技術の変換ベースのコーデックに関し、減衰定数は、励振領域（時間領域／ACELP様の封じ込め方策については[3GP09a]を参照）又は周波数領域のいずれかにおいて適用される（フレーム繰り返し又はノイズ置き換えといった周波数領域方策については、[LS01]を参照）。減衰定数を周波数領域において適用する先行技術の方策の欠点は、時間領域における重複・追加区域においてエイリアシングが生じることである。これは、異なる減衰定数が適用される隣接するフレームの場合に該当するが、それは、フェード手順によってTDAC（時間領域エイリアス除去）が失敗するからである。これは、調性を有する信号成分の封じ込め時に特に重要である。従って、上述の実施例は、先行技術よりも有利である。

20

【0284】

実施例は、LPC合成ゲインに対するハイパスフィルタの影響を補償する。実施例によると、ハイパスフィルタを通過した無声の励振によって引き起こされるLPC分析及び強調における望まれないゲイン変化を補償するために、訂正定数が導出される。この訂正定数は、この望まれないゲイン変化を考慮に入れ、励振領域における目標快適ノイズレベルに変更を加えることで、時間領域において正しい目標レベルに到達するようにする。

30

【0285】

これとは対照的に、先行技術、例えばG. 718 [ITU08a]では、最後の良好なフレームの信号が「無声」と分類されなかった場合、図2に示すように、無声励振の信号経路にハイパスフィルタが導入される。これにより、先行技術では、望まれない副作用が生じるが、それは、後続のLPC合成のゲインが、このハイパスフィルタによって変更される信号特性に依存するからである。背景レベルがトレースされて励振領域において適用されるため、アルゴリズムは、LPC合成ゲインに依存し、これは、やはり励振信号の特性に依存する。換言すると、先行技術で行われるようなハイパスフィルタリングによる励振の信号特性の変更によって、LPC合成のゲインが変更（通常は減少）されることがある。これにより、励振レベルが正しくても出力レベルが誤ったものとなる。

40

【0286】

実施例は、先行技術のこれらの欠点を克服する。

【0287】

特に、実施例は、快適ノイズの適応スペクトル形状を実現する。G. 718とは対照的に、背景ノイズのスペクトル形状をトレースし、且つ、バースト状のパケット損失中に、この形状を適用する（これへフェードさせる）ことによって、先行する背景ノイズのノイズ特性は一致し、快適ノイズの心地よいノイズ特性が得られる。これによって、オフライン学習により導出されたスペクトル包絡線の使用により導入され得るスペクトル形状及び／又は最後の受信フレームのスペクトル形状の耳障りな不一致が回避される。

【0288】

50

更に、音声信号を復号するための装置が提供される。当該装置は、受信インターフェースを備え、受信インターフェースは、音声信号の第1の音声信号部分を含む第1のフレームを受信するように構成され、受信インターフェースは、音声信号の第2の音声信号部分を含む第2のフレームを受信するように構成される。

【0289】

更に、当該装置は、ノイズレベルトレース部を備え、ノイズレベルトレース部は、第1の音声信号部分及び第2の音声信号部分のうちの少なくとも1つに応じて（即ち、第1の音声信号部分及び／又は第2の音声信号部分に応じて）ノイズレベル情報を決定するように構成され、ノイズレベル情報は、トレース領域において表現される。

【0290】

更に、当該装置は、複数のフレームのうちの第3のフレームが受信インターフェースによって受信されていない場合、又は前記第3のフレームが受信インターフェースによって受信されているが破損している場合、ノイズレベル情報に応じて音声信号の第3の音声信号部分を第1の再構成領域において再構成するための第1の再構成部を備え、第1の再構成領域は、トレース領域と異なる又は等しい。

【0291】

更に、当該装置は、複数のフレームのうちの第4のフレームが受信インターフェースによって受信されていない場合、又は前記第4のフレームが受信インターフェースによって受信されているが破損している場合、ノイズレベル情報をトレース領域から第2の再構成領域に変換するための変換部を備え、第2の再構成領域は、トレース領域とは異なり、第2の再構成領域は、第1の再構成領域とは異なる。

【0292】

更に、当該装置は、複数のフレームのうちの第4のフレームが受信インターフェースによって受信されていない場合、又は前記第4のフレームが受信インターフェースによって受信されているが破損している場合、第2の再構成領域において表現されるノイズレベル情報に応じて音声信号の第4の音声信号部分を第2の再構成領域において再構成するための第2の再構成部を備える。

【0293】

いくつかの実施例によると、トレース領域は、例えば、時間領域、スペクトル領域、FFT領域、MDCT領域、又は励振領域であることができる。第1の再構成領域は、例えば、時間領域、スペクトル領域、FFT領域、MDCT領域、又は励振領域であり得る。第2の再構成領域は、例えば、時間領域、スペクトル領域、FFT領域、MDCT領域、又は励振領域であり得る。

【0294】

実施例においては、トレース領域は、例えば、FFT領域とすることができ、第1の再構成領域は、例えば、時間領域とすることができ、第2の再構成領域は、例えば、励振領域とすることができる。

【0295】

別の実施例においては、トレース領域は、例えば、時間領域とすることができ、第1の再構成領域は、例えば、時間領域とすることができ、第2の再構成領域は、例えば、励振領域とすることができる。

【0296】

実施例によると、第1の音声信号部分は、例えば、第1の入力領域において表現することができ、第2の音声信号部分は、例えば、第2の入力領域において表現することができる。変換部は、例えば、第2の変換部であり得る。当該装置は、例えば、更に、第2の音声信号部分、又は第2の音声信号部分から導出された値又は信号を、第2の入力領域からトレース領域に変換して第2の信号部分情報を得るための第1の変換部を備えることができる。ノイズレベルトレース部は、例えば、トレース領域において表現される第1の信号部分情報を受信するように構成することができ、第1の信号部分情報は、第1の音声信号部分に依存し、ノイズレベルトレース部は、トレース領域において表現される第2の信号

10

20

30

40

50

部分を受信するように構成され、ノイズレベルトレース部は、トレース領域において表現される第1の信号部分情報と、トレース領域において表現される第2の信号部分情報とに応じてノイズレベル情報を決定するように構成される。

【0297】

実施例によると、第1の入力領域は、例えば、励振領域とすることができ、第2の入力領域は、例えば、MDC T領域とすることができる。

【0298】

別の実施例においては、第1の入力領域は、例えば、MDC T領域とすることができ、第2の入力領域は、例えば、MDC T領域とすることができる。

【0299】

実施例によると、第1の再構成部は、例えば、ノイズ様のスペクトルへの第1のフェードを行うことにより第3の音声信号部分を再構成するように構成され得る。第2の再構成部は、例えば、ノイズ様のスペクトルへの第2のフェード及び／又はLTPゲインの第2のフェードを行うことによって第4の音声信号部分を再構成するように構成され得る。更に、第1の再構成部及び第2の再構成部は、例えば、ノイズ様のスペクトルへの第1のフェード及び第2のフェード、並びに／又はLTPゲインの第2のフェードを、同じフェード速度で行うように構成され得る。

【0300】

実施例において、当該装置は、例えば、更に、第1の音声信号部分に応じて第1の集約値を決定するための第1の集約部を備えることができる。更に、当該装置は、更に、例えば、第2の音声信号部分に応じて、第2の音声信号部分から導出された値として第2の集約値を決定するための第2の集約部を備えることができる。ノイズレベルトレース部は、例えば、トレース領域において表現される第1の信号部分情報として第1の集約値を受信するように構成することができ、ノイズレベルトレース部は、例えば、トレース領域において表現される第2の信号部分情報として第2の集約値を受信するように構成することができ、ノイズレベルトレース部は、トレース領域において表現される第1の集約値と、トレース領域において表現される第2の集約値とに応じてノイズレベル情報を決定するように構成される。

【0301】

実施例によると、第1の集約部は、例えば、第1の集約値が、第1の音声信号部分、又は第1の音声信号部分から導出された信号の二乗平均平方根を示すように、第1の集約値を決定するように構成され得る。第2の集約部は、第2の集約値が、第2の音声信号部分、又は第2の音声信号部分から導出された信号の二乗平均平方根を示すように、第2の集約値を決定するように構成される。

【0302】

実施例において、第1の変換部は、例えば、第2の音声信号部分から導出された値にゲイン値を適用することにより、第2の音声信号部分から導出された値を第2の入力領域からトレース領域に変換するように構成され得る。

【0303】

実施例によると、ゲイン値は、例えば、線形予測符号化合成によって導入されたゲインを示すことができ、又は、ゲイン値は、線形予測符号化合成及び強調解除によって導入されたゲインを示す。

【0304】

実施例においては、ノイズレベルトレース部は、例えば、最小統計方策を適用することによりノイズレベル情報を決定するように構成され得る。

【0305】

実施例によると、ノイズレベルトレース部は、例えば、ノイズレベル情報として快適ノイズレベルを決定するように構成され得る。再構成部は、例えば、複数のフレームのうちの前記第3のフレームが受信インターフェースによって受信されていない場合、又は前記第3のフレームが受信インターフェースによって受信されているが破損している場合、ノ

10

20

30

40

50

イズレベル情報に応じて第3の音声信号部分を再構成するように構成され得る。

【0306】

実施例においては、ノイズレベルトレース部は、例えば、ノイズレベルスペクトルから導出されたノイズレベル情報として快適ノイズレベルを決定するように構成することができ、前記ノイズレベルスペクトルは、最小統計方策を適用することによって得られる。再構成部は、例えば、複数のフレームのうちの前記第3のフレームが受信インターフェースによって受信されていない場合、又は前記第3のフレームが受信インターフェースによって受信されているが破損している場合、複数の線形予測係数に応じて第3の音声信号部分を再構成するように構成され得る。

【0307】

10

実施例によると、第1の再構成部は、例えば、複数のフレームのうちの前記第3のフレームが受信インターフェースによって受信されていない場合、又は前記第3のフレームが受信インターフェースによって受信されているが破損している場合、ノイズレベル情報と第1の音声信号部分とに応じて第3の音声信号部分を再構成するように構成され得る。

【0308】

実施例においては、第1の再構成部は、例えば、第1の音声信号部分を減衰させる又は増幅することにより第3の音声信号部分を再構成するように構成され得る。

【0309】

実施例によると、第2の再構成部は、例えば、ノイズレベル情報と第2の音声信号部分とに応じて第4の音声信号部分を再構成するように構成され得る。

20

【0310】

実施例においては、第2の再構成部は、例えば、第2の音声信号部分を減衰させる又は増幅することにより第4の音声信号部分を再構成するように構成され得る。

【0311】

実施例によると、当該装置は、例えば、更に、遅延バッファを含む長期予測部を備えることができ、長期予測部は、例えば、第1の音声信号部分又は第2の音声信号部分と、遅延バッファに記憶される遅延バッファ入力と、長期予測ゲインとに応じて、処理済信号を生成するように構成することができ、長期予測部は、複数のフレームのうちの前記第3のフレームが受信インターフェースによって受信されていない場合、又は前記第3のフレームが受信インターフェースによって受信されているが破損している場合、長期予測ゲインをゼロヘフェードさせるように構成される。

30

【0312】

実施例において、長期予測部は、例えば、長期予測ゲインをゼロヘフェードさせるように構成することができ、長期予測ゲインがゼロヘフェードさせられる速度は、フェードアウト定数に依存する。

【0313】

実施例においては、長期予測部は、例えば、複数のフレームのうちの前記第3のフレームが受信インターフェースによって受信されていない場合、又は前記第3のフレームが受信インターフェースによって受信されているが破損している場合、生成された処理済信号を遅延バッファに記憶させることによって遅延バッファ入力を更新するように構成され得る。

40

【0314】

更に、音声信号を復号するための方法が提供される。当該方法は、

- 音声信号の第1の音声信号部分を含む第1のフレームを受信し、音声信号の第2の音声信号部分を含む第2のフレームを受信するステップと、

- 第1の音声信号部分及び第2の音声信号部分のうちの少なくとも1つに応じてノイズレベル情報を決定するステップと、を備え、ノイズレベル情報は、トレース領域において表現され、当該方法は、更に、

- 複数のフレームのうちの第3のフレームが受信されていない場合、又は前記第3のフレームが受信されているが破損している場合、ノイズレベル情報に応じて音声信号の第3

50

の音声信号部分を第1の再構成領域において再構成するステップを備え、第1の再構成領域は、トレース領域と異なり又は等しく、当該方法は、更に、

- 複数のフレームのうちの第4のフレームが受信されていない場合、又は前記第4のフレームが受信されているが破損している場合、ノイズレベル情報をトレース領域から第2の再構成領域に変換するステップを備え、第2の再構成領域は、トレース領域とは異なり、第2の再構成領域は、第1の再構成領域とは異なり、当該方法は、更に、

- 複数のフレームのうちの前記第4のフレームが受信されていない場合、又は前記第4のフレームが受信されているが破損している場合、第2の再構成領域において表現されるノイズレベル情報に応じて音声信号の第4の音声信号部分を第2の再構成領域において再構成するステップを備える。

10

【0315】

更に、コンピュータ又は信号プロセッサにおいて実行された時に上述の方法を実現するためのコンピュータプログラムが提供される。

【0316】

更に、符号化音声信号を復号して再構成音声信号を得るための装置が提供される。当該装置は、1つ以上のフレームを受信するための受信インターフェース、係数生成部、及び信号再構成部を備える。係数生成部は、1つ以上のフレームのうちの現在のフレームが受信インターフェースによって受信されており、且つ受信インターフェースによって受信されている現在のフレームが破損していない場合、現在のフレームに含まれる1つ以上の第1の音声信号係数を決定するように構成され、1つ以上の第1の音声信号係数は、符号化音声信号の特性を示し、1つ以上のノイズ係数が符号化音声信号の背景ノイズを示す。更に、係数生成部は、現在のフレームが受信インターフェースによって受信されていない場合、又は受信インターフェースによって受信されている現在のフレームが破損している場合、1つ以上の第1の音声信号係数と1つ以上のノイズ係数とに応じて1つ以上の第2の音声信号係数を生成するように構成される。音声信号再構成部は、現在のフレームが受信インターフェースによって受信されており、且つ受信インターフェースによって受信されている現在のフレームが破損していない場合、1つ以上の第1の音声信号係数に応じて再構成音声信号の第1の部分を再構成するように構成される。更に、音声信号再構成部は、現在のフレームが受信インターフェースによって受信されていない場合、又は受信インターフェースによって受信されている現在のフレームが破損している場合、1つ以上の第2の音声信号係数に応じて再構成音声信号の第2の部分を再構成するように構成される。

20

30

【0317】

いくつかの実施例においては、1つ以上の第1の音声信号係数は、例えば、符号化音声信号の1つ以上の線形予測フィルタ係数であり得る。いくつかの実施例においては、1つ以上の第1の音声信号係数は、例えば、符号化音声信号の1つ以上の線形予測フィルタ係数であり得る。

【0318】

実施例によると、1つ以上のノイズ係数は、例えば、符号化音声信号の背景ノイズを示す1つ以上の線形予測フィルタ係数であり得る。実施例においては、1つ以上の線形予測フィルタ係数は、例えば、背景ノイズのスペクトル形状を表し得る。

40

【0319】

実施例においては、係数生成部は、例えば、1つ以上の第2の音声信号部分が、再構成音声信号の1つ以上の線形予測フィルタ係数となる、又は、1つ以上の第1の音声信号係数が、再構成音声信号の1つ以上のインミッタンススペクトル対となるように、1つ以上の第2の音声信号部分を決定するように構成され得る。

【0320】

実施例によると、係数生成部は、例えば、

【0321】

【数 3 4】

$$f_{\text{current}}[i] = \alpha \cdot f_{\text{last}}[i] + (1 - \alpha) \cdot p_{\text{t_mean}}[i]$$

【0 3 2 2】

の式を適用することにより 1 つ以上の第 2 の音声信号係数を生成するように構成することができ、ここで、 $f_{\text{current}}[i]$ は、1 つ以上の第 2 の音声信号係数のうちの 1 つを示し、 $f_{\text{last}}[i]$ は、1 つ以上の第 1 の音声信号係数のうちの 1 つを示し、 $p_{\text{t_mean}}[i]$ は、1 つ以上のノイズ係数のうちの 1 つであり、 α は、実数 ($0 \leq \alpha \leq 1$) であり、 i は、インデックスである。実施例では、 $0 < \alpha < 1$ である。

10

【0 3 2 3】

実施例によると、 $f_{\text{last}}[i]$ は、符号化音声信号の線形予測フィルタ係数を示し、 $f_{\text{current}}[i]$ は、再構成音声信号の線形予測フィルタ係数を示す。

【0 3 2 4】

実施例においては、 $p_{\text{t_mean}}[i]$ は、例えば、符号化音声信号の背景ノイズを示し得る。

【0 3 2 5】

実施例においては、係数生成部は、例えば、1 つ以上のフレームのうちの現在のフレームが受信インターフェースによって受信されており、且つ受信インターフェースによって受信されている現在のフレームが破損していない場合、符号化音声信号のノイズスペクトルを決定することにより 1 つ以上のノイズ係数を決定するように構成され得る。

20

【0 3 2 6】

実施例によると、係数生成部は、例えば、信号スペクトルに対して最小統計方策を用いて背景ノイズスペクトルを決定し、背景ノイズ形状を表す L P C 係数を背景ノイズスペクトルから計算することによって、背景ノイズを表す L P C 係数を決定するように構成され得る。

【0 3 2 7】

更に、符号化音声信号を復号して再構成音声信号を得るための方法が提供される。当該方法は、

- 1 つ以上のフレームを受信するステップと、
- 1 つ以上のフレームのうちの現在のフレームが受信されており、且つ受信されている現在のフレームが破損していない場合、現在のフレームに含まれる 1 つ以上の第 1 の音声信号係数を決定するステップと、を備え、前記 1 つ以上の第 1 の音声信号係数は、符号化音声信号の特性を示し、1 つ以上のノイズ係数が符号化音声信号の背景ノイズを示し、当該方法は、更に、
 - 現在のフレームが受信されていない場合、又は受信されている現在のフレームが破損している場合、1 つ以上の第 1 の音声信号係数と前記 1 つ以上のノイズ係数とに応じて 1 つ以上の第 2 の音声信号係数を生成するステップと、
 - 現在のフレームが受信されており、且つ受信されている現在のフレームが破損していない場合、1 つ以上の第 1 の音声信号係数に応じて再構成音声信号の第 1 の部分を再構成するステップと、
 - 現在のフレームが受信されていない場合、又は受信されている現在のフレームが破損している場合、1 つ以上の第 2 の音声信号係数に応じて再構成音声信号の第 2 の部分を再構成するステップと、を備える。

30

40

【0 3 2 8】

更に、コンピュータ又は信号プロセッサにおいて実行された時に上述の方法を実現するためのコンピュータプログラムが提供される。

【0 3 2 9】

フェードアウト中に快適ノイズのスペクトル形状をトレースし適用するための共通の手段を有することにはいくつかの利点がある。スペクトル形状をトレースして適用すること

50

が両方のコアコーデックについて同様に行われることで、簡単な共通の方策が可能となる。C E L Tにおいては、スペクトル領域におけるエネルギーの帯域単位のトレースと、スペクトル領域におけるスペクトル形状の帯域単位の形成とが教示されているのみであり、これは、C E L Pコアにおいては不可能である。

【0330】

これとは対照的に、先行技術では、バースト状の損失中に導入される快適ノイズのスペクトル形状は、完全に静的か、或いは、部分的に静的且つ部分的にスペクトル形状の短期平均に対して適応的であり（G. 718 [ITU08a] で実現される）、通常、パケット損失前の信号における背景ノイズに一致しない。この快適ノイズ特性の不一致は、不快なものとなる場合がある。先行技術によると、オフラインで学習させた（静的）背景ノイズ形状を用いることができ、これは、特定の信号については心地よい音であり得るが、他の信号については、それほど心地よいものではない場合があり、例えば、自動車のノイズは、オフィスのノイズとは全く異なって聞こえる。

10

【0331】

更に、先行技術においては、以前に受信したフレームのスペクトル形状の短期平均に対する適応が用いられることがあり、これにより以前に受信した信号に近い信号特性を得ることができるが、これは、必ずしも背景ノイズ特性に近いものではない。先行技術では、スペクトル領域において帯域単位でスペクトル形状をトレースすること（C E L T [ITU T 12] で実現される）は、M D C T領域ベースのコア（T C X）だけでなくA C E L Pベースのコアをも用いる切り替えコーデックには適用可能ではない。従って、上述の実施例は、先行技術よりも有利である。

20

【0332】

更に、符号化音声信号を復号して再構成音声信号を得るための装置が提供される。当該装置は、符号化音声信号の音声信号スペクトルの複数の音声信号サンプルについての情報を含む1つ以上のフレームを受信するための受信インターフェースと、再構成音声信号を生成するための処理部とを備える。処理部は、現在のフレームが受信インターフェースによって受信されていない場合、又は現在のフレームが受信インターフェースによって受信されているが破損している場合、変更後スペクトルを目標スペクトルにフェードさせることにより再構成音声信号を生成するように構成されており、変更後スペクトルは、複数の変更後信号サンプルを含み、変更後スペクトルの変更後信号サンプルの各々について、前記変更後信号サンプルの絶対値は、音声信号スペクトルの音声信号サンプルのうちの1つについての絶対値に等しい。更に、処理部は、1つ以上のフレームのうちの現在のフレームが受信インターフェースによって受信されており、且つ、受信インターフェースによって受信されている現在のフレームが破損していない場合、変更後スペクトルを目標スペクトルにフェードさせないように構成される。

30

【0333】

実施例によると、目標スペクトルは、例えば、ノイズ様のスペクトルであり得る。

【0334】

実施例においては、ノイズ様のスペクトルは、例えば、ホワイトノイズを表すことができる。

40

【0335】

実施例によると、ノイズ様のスペクトルは、例えば、形状付けられ得る。

【0336】

実施例においては、ノイズ様のスペクトルの形状は、例えば、以前に受信された信号の音声信号スペクトルに依存することができる。

【0337】

実施例によると、ノイズ様のスペクトルは、例えば、音声信号スペクトルの形状に応じて形状付けられ得る。

【0338】

実施例においては、処理部は、例えば、チルト定数を用いてノイズ様のスペクトルを形

50

状付けることができる。

【0339】

実施例によると、処理部は、例えば、

【0340】

【数35】

```
shaped_noise[i]=noise*power(tilt_factor, i/N)
```

【0341】

の式を用いることができ、ここで、 N は、サンプルの数を示し、 i は、インデックスであり、 $0 \leq i < N$ であり、 $tilt_factor > 0$ であり、 $power$ は、パワー関数である。

10

【0342】

$power(x, y)$ は、 x^y を示す。

【0343】

$power(tilt_factor, i/N)$ は、 $tilt_factor^{i/N}$ を示す。

【0344】

もし、 $tilt_factor$ が1よりも小さければ、これは、増加する i による減衰を意味する。 $tilt_factor$ が1よりも大きければ、増加する i による増幅を意味する。

20

【0345】

別の実施例によると、処理部は、例えば、

【0346】

【数36】

```
shaped_noise[i]=noise*(1+i/(N-1))*(tilt_factor-1)
```

【0347】

の式を用いることができ、ここで、 N は、サンプルの数を示し、 i は、インデックスであり、 $0 \leq i < N$ であり、 $tilt_factor > 0$ である。

30

【0348】

もし、 $tilt_factor$ が1よりも小さければ、これは、増加する i による減衰を意味する。 $tilt_factor$ が1よりも大きければ、増加する i による増幅を意味する。

【0349】

実施例によると、処理部は、例えば、現在のフレームが受信インターフェースによって受信されていない場合、又は受信インターフェースによって受信されている現在のフレームが破損している場合、音声信号スペクトルの音声信号サンプルのうちの1つ以上についての符号を変えることによって変更後スペクトルを生成するように構成され得る。

40

【0350】

実施例においては、音声信号スペクトルの音声信号サンプルの各々は、例えば、想像上の数ではなく実数によって表され得る。

【0351】

実施例によると、音声信号スペクトルの音声信号サンプルは、例えば、変更後離散余弦変換領域において表され得る。

【0352】

別の実施例においては、音声信号スペクトルの音声信号サンプルは、例えば、変更後離散正弦変換領域において表され得る。

【0353】

50

実施例によると、処理部は、例えば、第1の値又は第2の値をランダム又は擬似ランダムに出力するランダム符号関数を使用することによって変更後スペクトルを生成するように構成され得る。

【0354】

実施例においては、処理部は、例えば、続いて減衰定数を減少させることによって変更後スペクトルを目標スペクトルにフェードさせるように構成され得る。

【0355】

実施例によると、処理部は、例えば、続いて減衰定数を増加させることによって変更後スペクトルを目標スペクトルにフェードさせるように構成され得る。

【0356】

実施例においては、現在のフレームが受信インターフェースによって受信されていない場合、又は受信インターフェースによって受信されている現在のフレームが破損している場合、処理部は、例えば、

【0357】

【数37】

$$x[i] = (1 - \text{cum_damping}) * \text{noise}[i] + \text{cum_damping} * \text{random_sign}() * x_{\text{old}}[i]$$

【0358】

の式を用いて再構成音声信号を生成するように構成することができ、ここで、 i は、インデックスであり、 $x[i]$ は、再構成音声信号のサンプルを示し、 cum_damping は、減衰定数であり、 $x_{\text{old}}[i]$ は、符号化音声信号の音声信号スペクトルの音声信号サンプルのうちの1つを示し、 $\text{random_sign}()$ は、1又は-1を返し、 noise は、目標スペクトルを示すランダムなベクトルである。

【0359】

実施例においては、前記ランダムなベクトル noise は、例えば、その平方平均が、受信インターフェースによって最後に受信されたフレームのうちの1つに含まれる符号化音声信号のスペクトルの平方平均に類似するようにスケーリングされ得る。

【0360】

一般的な実施例によると、処理部は、例えば、ランダムなベクトルであって、その平方平均が、受信インターフェースによって最後に受信されたフレームのうちの1つに含まれる符号化音声信号のスペクトルの平方平均に類似するようにスケーリングされたものを用いて、再構成音声信号を生成するように構成され得る。

【0361】

更に、符号化音声信号を復号して再構成音声信号を得るための方法が提供される。当該方法は、

- 符号化音声信号の音声信号スペクトルの複数の音声信号サンプルについての情報を含む1つ以上のフレームを受信するステップ、及び、
 - 再構成音声信号を生成するステップ、
- を備える。

【0362】

再構成音声信号を生成するステップは、現在のフレームが受信されていない場合、又は現在のフレームが受信されているが破損している場合、変更後スペクトルを目標スペクトルにフェードさせることによって行われ、変更後スペクトルは、複数の変更後信号サンプルを含み、変更後スペクトルの変更後信号サンプルの各々について、前記変更後信号サンプルの絶対値は、音声信号スペクトルの音声信号サンプルのうちの1つについての絶対値に等しい。1つ以上のフレームのうちの現在のフレームが受信されており、且つ、受信されている現在のフレームが破損していない場合、変更後スペクトルは、ホワイトノイズスペクトルにフェードさせられない。

【0363】

10

20

30

40

50

更に、コンピュータ又は信号プロセッサにおいて実行された時に上述の方法を実現するためのコンピュータプログラムが提供される。

【0364】

実施例は、FDNS適用に先立つホワイトノイズへのMDC Tスペクトルのフェードを実現する（FDNS＝周波数領域ノイズ置き換え）。

【0365】

先行技術によると、ACELPベースのコーデックにおいて、新規コードブックをランダムなベクトル（例えば、ホワイトノイズ）と入れ換える。実施例では、新規コードブックをランダムなベクトル（例えば、ホワイトノイズ）と入れ換えることからなるACELP方策が、TCXデコーダ構造に対して採用される。ここでは、新規コードブックの等価物は、通常、ビットストリーム内で受信され、FDNSに入力されるMDC Tスペクトルである。

10

【0366】

古典的なMDC T封じ込め方策は、単に、このスペクトルをそのまま繰り返すか、又は或るランダム化プロセスを適用することであり、これは、基本的に最後の受信フレームのスペクトル形状を長くする[LS01]。これには、短期的スペクトル形状が長くされるという欠点があり、そのため、頻繁に繰り返される金属音が生じ、これは、背景音様ではないため、快適ノイズとしては用いられ得ない。

【0367】

本願において提案される方法を用いると、短期間のスペクトル形状付けは、FDNS及びTCX LTPによって実行され、長期間に亘るスペクトル形状付けは、FDNSのみによって実行される。FDNSによる形状付けは、短期スペクトル形状から、背景ノイズについてのトレースされた長期スペクトル形状へフェードさせられ、TCX LTPは、ゼロにフェードさせられる。

20

【0368】

FDNS係数を、トレースされた背景ノイズ係数にフェードさせることによって、最後の良好なスペクトル包絡線と、長期間に亘り目標とされるべきスペクトル背景包絡線との間で平滑な遷移が得られ、長いバースト状のフレーム損失の場合に心地よい背景ノイズを達成する。

【0369】

対照的に、先行技術によれば、変換ベースのコーデックについて、周波数領域におけるフレーム繰り返し又はノイズ置き換えによってノイズ様の封じ込めが行われる[LS01]。先行技術では、ノイズ置き換えは、通常、スペクトルビンの符号スクランブルによって実行される。先行技術において、封じ込め中に、TCX（周波数領域）符号スクランブルを用いる場合、最後に受信したMDC T係数を再使用し、各々の符号をランダム化してから、スペクトルを時間領域へ逆変換する。この先行技術の手順の欠点は、連続して失われたフレームについて、同じスペクトルが何度も使用され、符号のランダム化及びグローバル減衰が異なるにすぎないことである。粗い時間グリッドにおける時間の経過に亘るスペクトル包絡線を見ると、包絡線は、連続するフレーム損失中にほぼ一定であることが分かるが、それは、フレーム内で帯域エネルギーが互いに対して相対的に一定に保たれ、グローバルに減衰させられるのみであるからである。使用される符号化システムにおいて、先行技術によると、FDNSを用いてスペクトル値を処理することにより、元のスペクトルを回復する。これは、（FDNS係数、例えば、現在の背景ノイズを記述するものを用いて）MDC Tスペクトルを特定のスペクトル包絡線にフェードさせることを望む場合、結果は、FDNS係数にのみ依存するのではなく、符号においてスクランブルされた以前に復号されたスペクトルにも依存することを意味する。上述の実施例は、先行技術のこれらの欠点を克服する。

30

40

【0370】

実施例は、符号スクランブルに用いられるスペクトルをホワイトノイズにフェードさせてから、FDNS処理に入力することが必要であるという知見に基づく。これを行わない

50

限り、出力されたスペクトルは、F D N S 処理に用いられる目標包絡線に一致することはない。

【0371】

実施例においては、ホワイトノイズフェードについてと同様にLTPゲインフェードについても、同じフェード速度を用いる。

【0372】

以下、本発明の実施例について、図面を参照してより詳細に説明する。

【図面の簡単な説明】

【0373】

【図1A】図1Aは、実施例による音声信号を復号するための装置を示す図である。

10

【図1B】図1Bは、別の実施例による音声信号を復号するための装置を示す図である。

【図1C】別の実施例による音声信号を復号するための装置であって、更に第1の集約部及び第2の集約部を備えるものを示す図である。

【図1D】図1Dは、更なる実施例による音声信号を復号するための装置であって、更に遅延バッファを含む長期予測部を備えるものを示す図である。

【図2】G. 718のデコーダ構造を示す図である。

【図3】図3は、G. 722のフェードアウト定数がクラス情報に依存するシナリオを示す図である。

【図4】図4は、線形回帰を用いた振幅予測のための方策を示す図である。

【図5】図5は、条件付きエネルギーラップド変換(CELT)のバースト状の損失の挙動を示す図である。

20

【図6】図6は、エラーのない動作モード中のデコーダにおける実施例による背景ノイズレベルトレースを示す図である。

【図7】図7は、実施例によるLPC合成及び強調解除のゲイン導出を示す図である。

【図8】図8は、実施例によるパケット損失中の快適ノイズレベル適用を示す図である。

【図9】図9は、実施例によるACELP封じ込め中の高度広域通過ゲイン補償を示す図である。

【図10】図10は、実施例による封じ込め中のLTPフィードバックループの結合解除を示す図である。

【図11】図11は、実施例による、符号化音声信号を復号して再構成音声信号を得るための装置を示す図である。

30

【図12】図12は、別の実施例による、符号化音声信号を復号して再構成音声信号を得るための装置を示す図である。

【図13】図13は、更なる実施例による、符号化音声信号を復号して再構成音声信号を得るための装置を示す図である。

【図14】図14は、別の実施例による、符号化音声信号を復号して再構成音声信号を得るための装置を示す図である。

【発明を実施するための形態】

【0374】

図1aは、実施例による音声信号を復号するための装置を示す。

40

【0375】

当該装置は、受信インターフェース110を備える。受信インターフェースは、複数のフレームを受信するように構成され、受信インターフェース110は、複数のフレームのうちの第1のフレームを受信するように構成され、前記第1のフレームは、音声信号の第1の音声信号部分を含み、前記第1の音声信号部分は、第1の領域において表現される。更に、受信インターフェース110は、複数のフレームのうちの第2のフレームを受信するように構成され、前記第2のフレームは、音声信号の第2の音声信号部分を含む。

【0376】

更に、当該装置は、第2の音声信号部分、又は第2の音声信号部分から導出された値若しくは信号を、第2の領域からトレース領域に変換することによって第2の信号部分情報

50

を得るための変換部120を備え、第2の領域は、第1の領域とは異なり、トレース領域は、第2の領域とは異なり、トレース領域は、第1の領域と等しい又は異なる。

【0377】

更に、当該装置は、ノイズレベルトレース部130を備え、ノイズレベルトレース部は、トレース領域において表現される第1の信号部分情報を受信するように構成され、第1の信号部分情報は、第1の音声信号部分に依存し、ノイズレベルトレース部は、トレース領域において表現される第2の信号部分を受信するように構成され、ノイズレベルトレース部は、トレース領域において表現される第1の信号部分情報と、トレース領域において表現される第2の信号部分情報とに応じてノイズレベル情報を決定するように構成される。

10

【0378】

更に、当該装置は、複数のフレームのうちの第3のフレームが受信インターフェースによって受信されていないが破損している場合、ノイズレベル情報に応じて音声信号の第3の音声信号部分を再構成するための再構成部を備える。

【0379】

第1の音声信号部分及び／又は第2の音声信号部分に関し、例えば、第1の音声信号部分及び／又は第2の音声信号部分は、例えば、1つ以上のラウドスピーカーのための1つ以上のラウドスピーカー信号を生成するための1つ以上の処理部（図示せず）に入力されることができ、これにより、第1の音声信号部分及び／又は第2の音声信号部分に含まれる受信音情報を再生することができる。

20

【0380】

更に、第1の音声信号部分及び第2の音声信号部分は、また、例えば、後続のフレームが受信機に到達しない場合、或いは、後続のフレームにエラーが含まれる場合、封じ込めにも用いられる。

【0381】

特に、本発明は、ノイズレベルトレースが共通の領域、本願では「トレース領域」と呼ばれる、において行われるべきであるという知見に基づく。トレース領域は、例えば、励振領域であっても良く、これは、例えば、AMR-WB及びAMR-WB+（[3GP12a]、[3GP12b]、[3GP09a]、[3GP09b]、[3GP09c]を参照）に記載のLPC（LPC＝線形予測係数）又はISP（ISP＝インミッタンススペクトル対）によって信号が表現される領域である。単一の領域においてノイズレベルをトレースすることには、特に、第1の領域における第1の表現と第2の領域における第2の表現との間で信号が切り替わる（例えば信号の表現がACELPからTCXへ、又はその逆に切り替わる）時にエイリアシング効果が回避されるという利点がある。

30

【0382】

変換部120に関し、変換されるのは、第2の音声信号部分そのもの、又は第2の音声信号部分から導出された信号（例えば、第2の音声信号部分を処理して導出信号を得る）、又は第2の音声信号部分から導出された値（例えば、第2の音声信号部分を処理して導出値を得る）である。

【0383】

第1の音声信号部分に関し、いくつかの実施例では、第1の音声信号部分をトレース領域へ処理及び／又は変換することができる。

40

【0384】

しかし、他の実施例では、第1の音声信号部分は、既にトレース領域において表現されている場合もある。

【0385】

いくつかの実施例では、第1の信号部分情報は、第1の音声信号部分と同一である。他の実施例では、第1の信号部分情報は、例えば、第1の音声信号部分に応じた集約値である。

【0386】

50

以下において、まず、快適ノイズレベルへのフェードアウトをより詳細に検討する。

【0387】

本願に記載のフェードアウト方策は、例えば、 $xHE-AAC[NMB+12]$ ($xHE-AAC$ = 拡張高効率AAC) の低遅延バージョンにおいて実現することができ、これは、フレーム毎にACELP (言語音声) 符号化とMDCT (音楽/ノイズ) 符号化との間で継ぎ目なく切り替わることができる。

【0388】

トレース領域、例えば励振領域における共通レベルトレースに関し、パケット損失中に適切な快適ノイズレベルへの平滑なフェードアウトを適用するために、通常の変換プロセス中に、このような快適ノイズレベルを特定する必要がある。例えば、背景ノイズに類似のノイズレベルが最も快適であると想定することができる。従って、通常変換中に、背景ノイズを導出して絶えず更新することができる。

10

【0389】

本発明は、切り替えコアコーデック (例えば、ACELP及びTCX) を有する場合、選択されたコアコーデックとは独立の共通の背景ノイズレベルを考慮することが特に適切であるという知見に基づく。

【0390】

図6は、エラーのない動作モード、例えば通常変換中の、デコーダにおける好ましい実施例による背景ノイズレベルトレースを示す。

【0391】

トレースそのものは、例えば、最小統計方策 ([Mar 01] を参照) を用いて実行することができる。

20

【0392】

このトレースされた背景ノイズレベルは、例えば、上述のノイズレベル情報と見なすことができる。

【0393】

背景ノイズレベルトレースについては、例えば、ライナー・マルティン (Rainer Martin)、「最適平滑化及び最小統計に基づくノイズパワースpektral密度推定 (Noise power spectral density estimation based on optimal smoothing and minimum statistics)」、IEEE 言語音声・音声処理会報 (IEEE Transactions on Speech and Audio Processing) 第9巻 (2001年)、第5号、504~512頁 [Mar 01]、の文献に呈示された最小統計ノイズ推定を用いることができる。

30

【0394】

これに対応して、いくつかの実施例では、ノイズレベルトレース部130は、最小統計方策を適用する、例えば、[Mar 01] の最小統計ノイズ推定を適用することによってノイズレベル情報を決定するように構成される。

【0395】

続いて、このトレース方策についてのいくつかの考慮すべき点と詳細について説明する。

【0396】

レベルトレースに関し、背景は、ノイズ様であると想定される。従って、LPCによって取り出される前景調性成分をトレースすることを回避するために、励振領域においてレベルトレースを実行することが好ましい。例えば、ACELPノイズ充填も、また、励振領域における背景ノイズレベルを用いることができる。励振領域におけるトレースによって、背景ノイズレベルの単一のトレースだけで2つの目的を果たすことができ、これによって、計算上の複雑度が節約される。好ましい実施例においては、トレースは、ACELP励振領域において実行される。

40

【0397】

図7は、実施例によるLPC合成及び強調解除のゲイン導出を示す。

【0398】

50

レベル導出に関し、レベル導出は、例えば、時間領域又は励振領域のいずれかにおいて、或いは、その他の任意の適切な領域において行うことができる。レベル導出のための領域とレベルトレースのための領域とが異なる場合、例えば、ゲイン補償が必要となる場合がある。

【0399】

好ましい実施例においては、ACELPについてのレベル導出は、励振領域において実行される。従って、ゲイン補償は、必要ではない。

【0400】

TCXの場合、例えば、導出レベルをACELP励振領域に対して調節するためにゲイン補償が必要となる場合がある。

10

【0401】

好ましい実施例において、TCXについてのレベル導出は、時間領域において行われる。この方策について制御可能なゲイン補償が見出された。即ち、LPC合成及び強調解除によって導入されたゲインを、図7に示すように導出し、導出したレベルをこのゲインで除算する。

【0402】

これに代えて、TCXについてのレベル導出をTCX励振領域で実行することもできる。しかし、TCX励振領域とACELP励振領域との間のゲイン補償は、過度に複雑であると考えられる。

【0403】

20

従って、再び、図1aを参照して、いくつかの実施例においては、第1の音声信号部分は、第1の領域として時間領域において表現される。変換部120は、第2の音声信号部分、又は第2の音声信号部分から導出された値を、第2の領域である励振領域から、トレース領域である時間領域へ変換するように構成される。このような実施例では、ノイズレベルトレース部130は、トレース領域としての時間領域において表現される第1の信号部分情報を受信するように構成される。更に、ノイズレベルトレース部130は、トレース領域としての時間領域において表現される第2の信号部分を受信するように構成される。

【0404】

他の実施例においては、第1の音声信号部分は、第1の領域としての励振領域において表現される。変換部120は、第2の音声信号部分、又は第2の音声信号部分から導出された値を、第2の領域である時間領域から、トレース領域である励振領域に変換するように構成される。このような実施例では、ノイズレベルトレース部130は、トレース領域としての励振領域において表現される第1の信号部分情報を受信するように構成される。更に、ノイズレベルトレース部130は、トレース領域としての励振領域において表現される第2の信号部分を受信するように構成される。

30

【0405】

実施例においては、第1の音声信号部分は、例えば、第1の領域としての励振領域において表現することができ、ノイズレベルトレース部130は、例えば、第1の信号部分情報を受信するように構成することができ、前記第1の信号部分情報は、トレース領域であるFFT領域において表現され、前記第1の信号部分情報は、前記励振領域において表現される第1の音声信号部分に依存し、変換部120は、例えば、第2の音声信号部分、又は第2の音声信号部分から導出された値を、第2の領域である時間領域から、トレース領域であるFFT領域に変換するように構成することができ、ノイズレベルトレース部130は、例えば、FFT領域において表現される第2の音声信号部分を受信するように構成され得る。

40

【0406】

図1bは、別の実施例による装置を示す。図1bにおいては、図1aの変換部120は、第1の変換部120であり、図1aの再構成部140は、第1の再構成部140である。当該装置は、更に、第2の変換部121及び第2の再構成部141を備える。

50

【0407】

第2の変換部121は、複数のフレームのうちの第4のフレームが受信インターフェースによって受信されていない場合、又は前記第4のフレームが受信インターフェースによって受信されているが破損している場合、ノイズレベル情報をトレース領域から第2の領域に変換するように構成される。

【0408】

更に、第2の再構成部141は、複数のフレームのうちの前記第4のフレームが受信インターフェースによって受信されていない場合、又は前記第4のフレームが受信インターフェースによって受信されているが破損している場合、第2の領域において表現されているノイズレベル情報に応じて音声信号の第4の音声信号部分を再構成するように構成される。

10

【0409】

図1cは、別の実施例による音声信号を復号するための装置を示す。当該装置は、更に、第1の音声信号部分に応じて第1の集約値を決定するための第1の集約部150を備える。更に、図1cの装置は、更に、第2の音声信号部分に応じて、第2の音声信号部分から導出された値としての第2の集約値を決定するための第2の集約部160を備える。図1cの実施例では、ノイズレベルトレース部130は、トレース領域において表現される第1の信号部分情報として第1の集約値を受信するように構成され、ノイズレベルトレース部130は、トレース領域において表現される第2の信号部分情報として第2の集約値を受信するように構成される。ノイズレベルトレース部130は、トレース領域において表現される第1の集約値と、トレース領域において表現される第2の集約値とに応じて、ノイズレベル情報を決定するように構成される。

20

【0410】

実施例においては、第1の集約部150は、第1の集約値が、第1の音声信号部分、又は第1の音声信号部分から導出された信号の、二乗平均平方根を示すように、第1の集約値を決定するように構成される。更に、第2の集約部160は、第2の集約値が、第2の音声信号部分、又は第2の音声信号部分から導出された信号の、二乗平均平方根を示すように、第2の集約値を決定するように構成される。

【0411】

図6は、更なる実施例による音声信号を復号するための装置を示す。

30

【0412】

図6においては、背景レベルトレース部630は、図1aによるノイズレベルトレース部130を実現する。

【0413】

更に、図6においては、RMS部650（ $RMS = \text{二乗平均平方根}$ ）は、第1の集約部であり、RMS部660は、第2の集約部である。

【0414】

いくつかの実施例によると、図1a、図1b及び図1cの（第1の）変換部120は、第2の音声信号部分から導出された値にゲイン値（x）を適用すること、例えば、第2の音声信号部分から導出された値をゲイン値（x）によって除算することにより、第2の音声信号部分から導出された値を第2の領域からトレース領域に変換するように構成される。他の実施例では、ゲイン値は、例えば、乗算されても良い。

40

【0415】

いくつかの実施例においては、ゲイン値（x）は、例えば、線形予測符号化合成によって導入されたゲインを示すことができ、又は、ゲイン値（x）は、例えば、線形予測符号化合成及び強調解除によって導入されたゲインを示すことができる。

【0416】

図6において、ユニット622は、線形予測符号化合成及び強調解除によって導入されたゲインを示す値（x）を提供する。次に、ユニット622は、第2の集約部660によって提供された値であって第2の音声信号部分から導出された値を、提供されたゲイン値

50

(x)で除算する(例えば、 x によって除算する、又は $1/x$ の値を乗算する)。こうして、ユニット621、622を含む図6のユニット620は、図1a、図1b又は図1cの第1の変換部を実現する。

【0417】

図6の装置は、第1の音声信号部分を有する第1のフレームを受信し、これは、有声の励振及び／又は無声の励振であり、かつ、トレース領域、図6では(ACELP)LPC領域において表現されるものである。第1の音声信号部分をLPC合成及び強調解除部671に入力して処理し、時間領域の第1の音声信号部分出力を得る。更に、第1の音声信号部分をRMSモジュール650に入力して、第1の音声信号部分の二乗平均平方根を示す第1の値を得る。この第1の値(第1のRMS値)は、トレース領域において表現される。トレース領域において表現される第1のRMS値は、次に、ノイズレベルトレース部630に入力される。

10

【0418】

更に、図6の装置は、MDCTスペクトルを含みMDCT領域において表現される第2の音声信号部分を有する第2のフレームを受信する。ノイズ充填は、ノイズ充填モジュール681によって行われ、周波数領域ノイズ形状付けが周波数領域ノイズ形状付けモジュール682によって行われ、時間領域への変換が、iMDCT/OLAモジュール683(OLA=オーバーラップ・加算)によって行われ、長期予測が長期予測部684によって行われる。長期予測部は、例えば、遅延バッファ(図6では示さず)を含み得る。

20

【0419】

次に、第2の音声信号部分から導出された信号をRMSモジュール660に入力し、第2の音声信号部分から導出された信号の二乗平均平方根を示す第2の値を得る。この第2の値(第2のRMS値)は、時間領域において、尚、表現されている。次に、ユニット620は、第2のRMS値を、時間領域からトレース領域、ここでは(ACELP)LPC領域に変換する。トレース領域において表現される第2のRMS値は、次に、ノイズレベルトレース部630に入力される。

【0420】

実施例においては、レベルトレースは、励振領域において行われるが、TCXフェードアウトは、時間領域で行われる。

【0421】

30

通常の復号中は、背景ノイズレベルがトレースされるのに対し、パケット損失中は、これを、例えば、適切な快適ノイズレベルの標識として用いることができ、最後の受信された信号は、これへと平滑にレベル単位でフェードさせられる。

【0422】

トレースのためのレベルを導出することと、レベルのフェードアウトを適用することとは、一般的に互いに独立しており、それぞれ異なる領域で実行され得る。好ましい実施例では、レベル適用は、レベル導出と同じ領域で実行されるため、ACELPの場合は、ゲイン補償が不要であるということ、及びTCXの場合は、レベル導出(図6を参照)におけるような逆ゲイン補償が必要であるため、図7に示すように同じゲイン導出が用いられ得ること、という同じ利益が得られる。

40

【0423】

以下、実施例によるLPC合成ゲインに対するハイパスフィルタの影響の補償について説明する。

【0424】

図8は、この方策の概略を示す。特に、図8は、パケット損失中の快適ノイズレベル適用を示す。

【0425】

図8においては、ハイパスゲインフィルタ部643、乗算部644、フェード部645、ハイパスフィルタ部646、フェード部647及び組み合わせ部648と一緒に第1の再構成部を構成している。

50

【0426】

更に、図8において、背景レベル出力部631は、ノイズレベル情報を出力する。例えば、背景レベル供給部631は、図6の背景レベルトレース部630として等しく実現することができる。

【0427】

更に、図8において、LPC合成・強調解除ゲイン部649及び乗算部641は、一緒に第2の変換部640を構成する。

【0428】

更に、図8において、フェード部642は、第2の再構成部を表す。

【0429】

図8の実施例において、有声の励振と無声の励振とは別個にフェードさせられる。即ち、有声の励振は、ゼロヘフェードさせられるが、無声の励振は、快適ノイズレベルへとフェードさせられる。更に、図8は、ハイパスフィルタを示し、これを無声の励振の信号チェーンに導入して、信号が無声と分類された時を除き全ての場合において低周波数成分を抑制する。

【0430】

ハイパスフィルタの影響のモデル化に関し、LPC合成及び強調解除後のレベルは、ハイパスフィルタを伴って一度計算され、ハイパスフィルタを伴わずに一度計算される。この後、これら2つのレベルの比を導出し、これを用いて適用背景レベルを変更する。

【0431】

これを図9に示す。特に、図9は、実施例によるACELP封じ込め中の高度ハイパスゲイン補償を示す。

【0432】

この計算では、現在の励振信号の代わりに、簡単なインパルスを入力として用いる。これにより、複雑度が低減できるが、それは、インパルス応答は、急速に減衰するため、RMS導出を短い時間フレームで行うことができるからである。実際は、フレーム全体の代わりに、ただ1つのサブフレームを用いる。

【0433】

実施例によると、ノイズレベルトレース部130は、ノイズレベル情報として快適ノイズレベルを決定するように構成される。再構成部140は、複数のフレームのうちの前記第3のフレームが受信インターフェース110によって受信されていない場合、又は前記第3のフレームが受信インターフェース110によって受信されているが破損している場合、ノイズレベル情報に応じて第3の音声信号部分を再構成するように構成される。

【0434】

実施例によると、ノイズレベルトレース部130は、ノイズレベル情報として快適ノイズレベルを決定するように構成される。再構成部140は、複数のフレームのうちの前記第3のフレームが受信インターフェース110によって受信されていない場合、又は前記第3のフレームが受信インターフェース110によって受信されているが破損している場合、ノイズレベル情報に応じて第3の音声信号部分を再構成するように構成される。

【0435】

実施例においては、ノイズレベルトレース部130は、ノイズレベルスペクトルから導出されたノイズレベル情報として快適ノイズレベルを決定するように構成され、前記ノイズレベルスペクトルは、最小統計方策を適用することによって得られる。再構成部140は、複数のフレームのうちの前記第3のフレームが受信インターフェースに110によって受信されていない場合、又は前記第3のフレームが受信インターフェース110によって受信されているが破損している場合、複数の線形予測係数に応じて第3の音声信号部分を再構成するように構成される。

【0436】

実施例においては、(第1及び/又は第2の)再構成部140、141は、例えば、複数のフレームのうちの前記第3(第4)のフレームが受信インターフェース110によ

10

20

30

40

50

て受信されていない場合、又は前記第3（第4）のフレームが受信インターフェース110によって受信されているが破損している場合、ノイズレベル情報と第1の音声信号部分とに応じて第3の音声信号部分を再構成するように構成され得る。

【0437】

実施例によると、（第1及び／又は第2の）再構成部140、141は、例えば、第1の音声信号部分を減衰させる又は増幅することによって第3（又は第4）の音声信号部分を再構成するように構成され得る。

【0438】

図14は、音声信号を復号するための装置を示す。当該装置は、受信インターフェース110を備え、受信インターフェース110は、音声信号の第1の音声信号部分を含む第1のフレームを受信するように構成され、受信インターフェース110は、音声信号の第2の音声信号部分を含む第2のフレームを受信するように構成される。

10

【0439】

更に、当該装置は、ノイズレベルトレース部130を備え、ノイズレベルトレース部130は、第1の音声信号部分及び第2の音声信号部分のうちの少なくとも1つに応じて（即ち、第1の音声信号部分及び／又は第2の音声信号部分に応じて）ノイズレベル情報を決定するように構成され、ノイズレベル情報は、トレース領域において表現される。

【0440】

更に、当該装置は、複数のフレームのうちの第3のフレームが受信インターフェース110によって受信されていない場合、又は前記第3のフレームが受信インターフェース110によって受信されているが破損している場合、ノイズレベル情報に応じて音声信号の第3の音声信号部分を第1の再構成領域において再構成するための第1の再構成部140を備え、第1の再構成領域は、トレース領域と異なる又は等しい。

20

【0441】

更に、当該装置は、複数のフレームのうちの第4のフレームが受信インターフェース110によって受信されていない場合、又は前記第4のフレームが受信インターフェース110によって受信されているが破損している場合、ノイズレベル情報をトレース領域から第2の再構成領域に変換するための変換部121を備え、第2の再構成領域は、トレース領域とは異なり、第2の再構成領域は、第1の再構成領域とは異なる。

【0442】

30

更に、当該装置は、複数のフレームのうちの前記第4のフレームが受信インターフェース110によって受信されていない場合、又は前記第4のフレームが受信インターフェース110によって受信されているが破損している場合、第2の再構成領域において表現されるノイズレベル情報に応じて音声信号の第4の音声信号部分を第2の再構成領域において再構成するための第2の再構成部141を備える。

【0443】

いくつかの実施例によると、トレース領域は、例えば、時間領域、スペクトル領域、FFT領域、MDCT領域、又は励振領域であることができる。第1の再構成領域は、例えば、時間領域、スペクトル領域、FFT領域、MDCT領域、又は励振領域であり得る。第2の再構成領域は、例えば、時間領域、スペクトル領域、FFT領域、MDCT領域、又は励振領域であり得る。

40

【0444】

実施例においては、トレース領域は、例えば、FFT領域とすることができ、第1の再構成領域は、例えば、時間領域とすることができ、第2の再構成領域は、例えば、励振領域とすることができ、

【0445】

別の実施例においては、トレース領域は、例えば、時間領域とすることができ、第1の再構成領域は、例えば、時間領域とすることができ、第2の再構成領域は、例えば、励振領域とすることができ、

【0446】

50

実施例によると、前記第1の音声信号部分は、例えば、第1の入力領域において表現することができ、前記第2の音声信号部分は、例えば、第2の入力領域において表現することができる。変換部は、例えば、第2の変換部であり得る。当該装置は、例えば、更に、第2の音声信号部分、又は第2の音声信号部分から導出された値又は信号を、第2の入力領域からトレース領域に変換して第2の信号部分情報を得るための第1の変換部を備えることができる。ノイズレベルトレース部は、例えば、トレース領域において表現される第1の信号部分情報を受信するように構成することができ、第1の信号部分情報は、第1の音声信号部分に依存し、ノイズレベルトレース部は、トレース領域において表現される第2の信号部分を受信するように構成され、ノイズレベルトレース部は、トレース領域において表現される第1の信号部分情報と、トレース領域において表現される第2の信号部分情報とに応じてノイズレベル情報を決定するように構成される。

10

【0447】

実施例によると、第1の入力領域は、例えば、励振領域とすることができ、第2の入力領域は、例えば、MDC T領域とすることができる。

【0448】

別の実施例においては、第1の入力領域は、例えば、MDC T領域とすることができ、第2の入力領域は、例えば、MDC T領域とすることができる。

【0449】

例えば、信号が時間領域において表現される場合、これは、例えば、信号の時間領域サンプルによって表現することができる。或いは、例えば、信号がスペクトル領域において表現される場合、これは、例えば、信号のスペクトルのスペクトルサンプルによって表現することができる。

20

【0450】

実施例においては、トレース領域は、例えば、FFT領域とすることができ、第1の再構成領域は、例えば、時間領域とすることができ、第2の再構成領域は、例えば、励振領域とすることができる。

【0451】

別の実施例では、トレース領域は、例えば、時間領域とすることができ、第1の再構成領域は、例えば、時間領域とすることができ、第2の再構成領域は、例えば、励振領域とすることができる。

30

【0452】

いくつかの実施例においては、図14に示す各ユニットは、例えば、図1a, 1b, 1c, 1dについて説明したように構成することができる。

【0453】

特定の実施例に関し、例えば、低レートモードにおいて、実施例による装置は、例えば、ACELPフレームを入力として受信することができ、これらフレームは、励振領域において表現され、それから、LPC合成を介して時間領域に変換される。更に、低レートモードにおいて、実施例による装置は、例えば、TCXフレームを入力として受信することができ、これらフレームは、MDC T領域において表現され、それから、逆MDC Tを介して時間領域に変換される。

40

【0454】

次に、トレースがFFT領域において行われ、FFT信号は、FFT（高速フーリエ変換）を行うことによって時間領域信号から導出される。トレースは、例えば、最小統計方策を全てのスペクトル線について別個に行って快適ノイズスペクトルを得ることによって行うことができる。

【0455】

次に、快適ノイズスペクトルに基づいてレベル導出を行うことによって封じ込めを行う。レベル導出は、快適ノイズスペクトルに基づいて行われる。時間領域へのレベル変換は、FD TCX PLCについて行われる。時間領域におけるフェードが行われる。励振領域へのレベル導出がACELP PLC及びTD TCX PLC（ACELP様）に

50

ついて行われる。次に、励振領域におけるフェードが行われる。

【0456】

これを以下のリストで要約する。

低レート

●入力

- a c e l p (励振領域→時間領域。L P C 合成を介する)
- t c x (m d c t 領域→時間領域。逆M D C Tを介する)

●トレース

- F F T を介して時間領域から導出される f f t 領域
- 全てのスペクトル線について別個の最小統計→快適ノイズスペクトル

10

●封じ込め

- 快適ノイズスペクトルに基づくレベル導出
- 以下についての時間領域へのレベル変換
 - ・ F D T C X P L C
 - 時間領域におけるフェード
- 以下についての励振領域へのレベル変換
 - ・ A C E L P P L C
 - ・ T D T C X P L C (A C E L P 様)
 - 励振領域におけるフェード

例えば、高レートモードにおいては、例えば、T C X フレームを入力として受信することができ、これらフレームは、M D C T 領域において表現され、それから、逆M D C T を介して時間領域に変換される。

20

【0457】

次に、トレースが時間領域において行われる。トレースは、例えば、エネルギーレベルに基づいて最小統計方策を行って快適ノイズレベルを得ることによって行うことができる。

【0458】

封じ込めについては、F D T C X P L C の場合、レベルをそのまま用いることができ、時間領域におけるフェードのみを行うことができる。T D T C X P L C (A C E L P 様) については、励振領域へのレベル変換及び励振領域におけるフェードが行われる。

30

【0459】

これを以下のリストで要約する。

高レート

●入力

- t c x (m d c t 領域→時間領域。逆M D C Tを介する)

●トレース

- 時間領域
- エネルギーレベルに対する最小統計→快適ノイズレベル

●封じ込め

- 「そのままの」レベル使用
 - ・ F D T C X P L C
 - 時間領域におけるフェード
- 以下についての励振領域へのレベル変換
 - ・ T D T C X P L C (A C E L P 様)
 - 励振領域におけるフェード

40

F F T 領域及びM D C T 領域は、両方ともスペクトル領域である一方、励振領域は、或る種の時間領域である。

【0460】

実施例によると、第1の再構成部140は、例えば、ノイズ様のスペクトルへの第1の

50

フェードを行うことにより第3の音声信号部分を再構成するように構成され得る。第2の再構成部141は、例えば、ノイズ様のスペクトルへの第2のフェード及び／又はLTPゲインの第2のフェードを行うことによって第4の音声信号部分を再構成するように構成され得る。更に、第1の再構成部140及び第2の再構成部141は、例えば、ノイズ様のスペクトルへの第1のフェード及び第2のフェード、並びに／又はLTPゲインの第2のフェードを、同じフェード速度で行うように構成され得る。

【0461】

次に、快適ノイズの適応スペクトル形状付けについて検討する。

【0462】

バースト状のパケット損失中に快適ノイズへの適応形状付けを達成するために、第1のステップとして、背景ノイズを表す適切なLPC係数を求めることができる。これらのLPC係数は、アクティブな言語音声の期間中に導出することができ、これを行うために、最小統計方策を用いて背景ノイズスペクトルを求めて、文献から公知のLPC導出のための任意のアルゴリズムを用いて、この背景ノイズスペクトルからLPC係数を算出する。いくつかの実施例では、例えば、背景ノイズスペクトルを或る表現に直接変換することができ、この表現をMDCT領域におけるFDNSについて直接用いることができる。

【0463】

快適ノイズへのフェードは、ISF領域において（LSF領域でも適用可能である。LSF＝線スペクトル周波数）、

【0464】

【数38】

$$f_{\text{current}}[i] = \alpha \cdot f_{\text{last}}[i] + (1 - \alpha) \cdot p_{\text{t_mean}}[i] \quad i=0 \dots 16 \quad (26)$$

【0465】

$p_{\text{t_mean}}$ を、快適ノイズを記述する適切なLP係数に設定することによって行うことができる。

【0466】

上述の快適ノイズの適応スペクトル形状付けに関し、より一般的な実施例を図11に示す。

【0467】

図11は、実施例による、符号化音声信号を復号して再構成音声信号を得るための装置を示す。

【0468】

当該装置は、1つ以上のフレームを受信するための受信インターフェース1110、係数生成部1120、及び信号再構成部1130を備える。

【0469】

係数生成部1120は、1つ以上のフレームのうちの現在のフレームが受信インターフェース1110によって受信されており、且つ受信インターフェース1110によって受信されている現在のフレームが破損していない／誤りを含まない場合、現在のフレームに含まれる1つ以上の第1の音声信号係数を決定するように構成され、前記1つ以上の第1の音声信号係数は、符号化音声信号の特性を示し、1つ以上のノイズ係数が符号化音声信号の背景ノイズを示す。更に、係数生成部1120は、現在のフレームが受信インターフェース1110によって受信されていない場合、又は受信インターフェース1110によって受信されている現在のフレームが破損している／誤りを含む場合、1つ以上の第1の音声信号係数と1つ以上のノイズ係数とに応じて1つ以上の第2の音声信号係数を生成するように構成される。

【0470】

音声信号再構成部1130は、現在のフレームが受信インターフェース1110によって受信されており、且つ受信インターフェース1110によって受信されている現在のフ

10

20

30

40

50

フレームが破損していない場合、1つ以上の第1の音声信号係数に応じて再構成音声信号の第1の部分再構成するように構成される。更に、音声信号再構成部1130は、現在のフレームが受信インターフェース1110によって受信されていない場合、又は受信インターフェース1110によって受信されている現在のフレームが破損している場合、1つ以上の第2の音声信号係数に応じて再構成音声信号の第2の部分再構成するように構成される。

【0471】

背景ノイズの決定は、先行技術で周知であり（例えば、[Mar01]：ライナー・マルティン（Rainer Martin）、「最適平滑化及び最小統計に基づくノイズパワースペクトル密度推定（Noise power spectral density estimation based on optimal smoothing and minimum statistics）」、IEEE言語音声・音声処理会報（IEEE Transactions on Speech and Audio Processing）第9巻（2001年）、第5号、504～512頁を参照）、実施例において、当該装置はこれに従って動作する。

10

【0472】

いくつかの実施例において、1つ以上の第1の音声信号係数は、例えば、符号化音声信号の1つ以上の線形予測フィルタ係数とすることができる。いくつかの実施例では、1つ以上の第1の音声信号係数は、例えば、符号化音声信号の1つ以上の線形予測フィルタ係数とすることができる。

【0473】

当該技術においては、音声信号、例えば、言語音声信号を、線形予測フィルタ係数又はインミッタンススペクトル対からどのように再構成するのかは周知であり（例えば、[3GP09c]：「言語音声コーデック言語音声処理機能。適応マルチレート広帯域（AMRWB）言語音声コーデック。トランスコード処理機能（Speech codec speech processing functions; adaptive multi-rate - wideband (AMRWB) speech codec; transcoding functions）、3GPP TS 26.190、第3世代パートナーシッププロジェクト（3rd Generation Partnership Project）、2009年、を参照）、実施例では、信号再構成部は、これに従って動作する。

20

【0474】

実施例によると、1つ以上のノイズ係数は、例えば、符号化音声信号の背景ノイズを示す1つ以上の線形予測フィルタ係数とすることができる。実施例においては、1つ以上の線形予測フィルタ係数は、例えば、背景ノイズのスペクトル形状を表すことができる。

30

【0475】

実施例においては、係数生成部1120は、例えば、1つ以上の第2の音声信号部分が、再構成音声信号についての1つ以上の線形予測フィルタ係数となる、又は、1つ以上の第1の音声信号係数が、再構成音声信号についての1つ以上のインミッタンススペクトル対となるように、1つ以上の第2の音声信号部分を決定するように構成され得る。

【0476】

実施例によると、係数生成部1120は、例えば、

【0477】

【数39】

40

$$f_{\text{current}}[i] = \alpha \cdot f_{\text{last}}[i] + (1 - \alpha) \cdot p_{\text{t_mean}}[i]$$

【0478】

の式を適用することにより1つ以上の第2の音声信号係数を生成するように構成することができ、ここで、 $f_{\text{current}}[i]$ は、1つ以上の第2の音声信号係数のうちの1つを示し、 $f_{\text{last}}[i]$ は、1つ以上の第1の音声信号係数のうちの1つを示し、 $p_{\text{t_mean}}[i]$ は、1つ以上のノイズ係数のうちの1つであり、 α は、実数（ $0 \leq \alpha \leq 1$ ）であり、 i は、インデックスである。

【0479】

50

実施例によると、 $f_{last}[i]$ は、符号化音声信号の線形予測フィルタ係数を示し、 $f_{current}[i]$ は、再構成音声信号の線形予測フィルタ係数を示す。

【0480】

実施例においては、 $pt_{mean}[i]$ は、例えば、符号化音声信号の背景ノイズを示す線形予測フィルタ係数であり得る。

【0481】

実施例によると、係数生成部1120は、例えば、少なくとも10個の第2の音声信号係数を、1つ以上の第2の音声信号係数として生成するように構成され得る。

【0482】

実施例においては、係数生成部1120は、例えば、1つ以上のフレームのうちの現在のフレームが受信インターフェース1110によって受信されており、且つ受信インターフェース1110によって受信されている現在のフレームが破損していない場合、符号化音声信号のノイズスペクトルを決定することにより1つ以上のノイズ係数を決定するように構成され得る。

【0483】

以下、FDNS適用に先立つホワイトノイズへのMDCTスペクトルのフェードについて検討する。

【0484】

MDCTビンの符号をランダムに変更する（符号スクランブル）代わりに、完全なスペクトルに、FDNSを用いて形状付けられたホワイトノイズを充填する。スペクトル特性における瞬間的な変化を回避するために、符号スクランブルとノイズ充填とのクロスフェードが適用される。クロスフェードは、以下のように実現することができる。

【0485】

【数40】

```
for(i=0; i<L_frame; i++) {
    if (old_x[i]!=0) {
        x[i]=(1-cum_damping)*noise[i]+cum_damping*random_sign()*x_old[i];
    }
}
```

【0486】

ここで、 $cum_damping$ は、（絶対）減衰定数である。これは、フレーム毎に減少し、1から始まって0へと減少する。

【0487】

x_old は、最後の受信フレームのスペクトルである。

【0488】

$random_sign$ は、1又は-1を返す。

【0489】

$noise$ は、ランダムなベクトル（ホワイトノイズ）であって、その平方平均（RMS）が最後の良好なスペクトルと類似となるようにスケールされるものを含む。

【0490】

$random_sign() * old_x[i]$ の項は、位相をランダム化して高調波の繰り返しを避けるための符号スクランブルプロセスを特徴づける。

【0491】

続いて、クロスフェードの後に、エネルギーレベルの別の正規化を実行しても良く、これにより合計エネルギーが2つのベクトルの相関によって逸脱しないことを確実にする。

【0492】

実施例によると、第1の再構成部140は、例えば、ノイズレベル情報と第1の音声信

10

20

30

40

50

号部分とに応じて第3の音声信号部分を再構成するように構成され得る。具体的な実施例では、第1の再構成部140は、例えば、第1の音声信号部分を減衰させる又は増幅することによって第3の音声信号部分を再構成するように構成され得る。

【0493】

いくつかの実施例では、第2の再構成部141は、例えば、ノイズレベル情報と第2の音声信号部分とに応じて第4の音声信号部分を再構成するように構成され得る。具体的な実施例では、第2の再構成部141は、例えば、第2の音声信号部分を減衰させる又は増幅することによって第4の音声信号部分を再構成するように構成され得る。

【0494】

上述のFDNS適用に先立つホワイトノイズへのMDCTスペクトルのフェードに関し、より一般的な実施例を図12に示す。

10

【0495】

図12は、実施例による、符号化音声信号を復号して再構成音声信号を得るための装置を示す。

【0496】

当該装置は、符号化音声信号の音声信号スペクトルの複数の音声信号サンプルについての情報を含む1つ以上のフレームを受信するための受信インターフェース1210と、再構成音声信号を生成するための処理部1220とを備える。

【0497】

処理部1220は、現在のフレームが受信インターフェース1210によって受信されていない場合、又は現在のフレームが受信インターフェース1210によって受信されているが破損している場合、変更後スペクトルを目標スペクトルにフェードさせることにより再構成音声信号を生成するように構成されており、変更後スペクトルは、複数の変更後信号サンプルを含み、変更後スペクトルの変更後信号サンプルの各々について、前記変更後信号サンプルの絶対値は、音声信号スペクトルの音声信号サンプルのうちの1つの絶対値に等しい。

20

【0498】

更に、処理部1220は、1つ以上のフレームのうちの現在のフレームが受信インターフェース1210によって受信されており、且つ受信インターフェース1210によって受信されている現在のフレームが破損していない場合、変更後スペクトルを目標スペクトルにフェードさせないように構成される。

30

【0499】

実施例によると、目標スペクトルは、ノイズ様のスペクトルである。

【0500】

実施例においては、ノイズ様のスペクトルは、ホワイトノイズを表す。

【0501】

実施例によると、ノイズ様のスペクトルは、形状付けられる。

【0502】

実施例においては、ノイズ様のスペクトルの形状は、以前に受信された信号の音声信号スペクトルに依存する。

40

【0503】

実施例によると、ノイズ様のスペクトルは、音声信号スペクトルの形状に応じて形状付けられる。

【0504】

実施例においては、処理部1220は、チルト定数を適用してノイズ様のスペクトルを形状付ける。

【0505】

実施例によると、処理部1220は、

【0506】

【数 4 1】

```
shaped_noise[i]=noise*power(tilt_factor,i/N)
```

【0 5 0 7】

の式を適用し、ここで、 N は、サンプルの数を示し、

i は、インデックスであり、

$0 \leq i < N$ であり、 $tilt_factor > 0$ であり、

$power$ は、パワー関数である。

【0 5 0 8】

もし、 $tilt_factor$ が1よりも小さければ、これは、増加する i による減衰を意味する。 $tilt_factor$ が1よりも大きければ、増加する i による増幅を意味する。

【0 5 0 9】

別の実施例によると、処理部 1 2 2 0 は、

【0 5 1 0】

【数 4 2】

```
shaped_noise[i]=noise*(1+i/(N-1)*(tilt_factor-1))
```

【0 5 1 1】

の式を適用することができ、ここで、 N は、サンプル数を示し、

i は、インデックスであり、 $0 \leq i < N$ であり、

$tilt_factor > 0$ である。

【0 5 1 2】

実施例によると、処理部 1 2 2 0 は、現在のフレームが受信インターフェース 1 2 1 0 によって受信されていない場合、又は受信インターフェース 1 2 1 0 によって受信されている現在のフレームが破損している場合、音声信号スペクトルの音声信号サンプルのうちの1つ以上の符号を変えることによって変更後スペクトルを生成するように構成される。

【0 5 1 3】

実施例においては、音声信号スペクトルの音声信号サンプルの各々は、想像上の数ではなく、実数によって表される。

【0 5 1 4】

実施例によると、音声信号スペクトルの音声信号サンプルは、変更後離散余弦変換領域において表される。

【0 5 1 5】

別の実施例においては、音声信号スペクトルの音声信号サンプルは、変更後離散正弦変換領域において表される。

【0 5 1 6】

実施例によると、処理部 1 2 2 0 は、第1の値又は第2の値をランダム又は疑似ランダムに出力するランダム符号関数を適用することによって変更後スペクトルを生成するように構成される。

【0 5 1 7】

実施例においては、処理部 1 2 2 0 は、続いて減衰定数を減少させることによって変更後スペクトルを目標スペクトルにフェードさせるように構成される。

【0 5 1 8】

実施例によると、処理部 1 2 2 0 は、続いて減衰定数を増加させることによって変更後スペクトルを目標スペクトルにフェードさせるように構成される。

【0 5 1 9】

実施例においては、現在のフレームが受信インターフェース 1 2 1 0 によって受信され

10

20

30

40

50

ていない場合、又は受信インターフェース1210によって受信されている現在のフレームが破損している場合、処理部1220は、

【0520】

【数43】

$$x[i] = (1 - \text{cum_damping}) * \text{noise}[i] + \text{cum_damping} * \text{random_sign}() * x_{\text{old}}[i]$$

【0521】

の式を適用して再構成音声信号を生成するように構成され、ここで、 i は、インデックスであり、 $x[i]$ は、再構成音声信号のサンプルを示し、 cum_damping は、減衰定数であり、 $x_{\text{old}}[i]$ は、符号化音声信号の音声信号スペクトルの音声信号サンプルのうちの1つを示し、 $\text{random_sign}()$ は、1又は-1を返し、 noise は、目標スペクトルを示すランダムなベクトルである。

10

【0522】

いくつかの実施例では、TCX LTP演算が継続される。これらの実施例では、TCX LTP演算は、最後の良好なフレームから導出されたLTPパラメータ（LTP遅れ及びLTPゲイン）によって封じ込め中に継続される。

【0523】

LTP演算は、以下のように要約することができる。

【0524】

- 以前に導出した出力に基づいてLTP遅延バッファを供給する。

20

【0525】

- LTP遅れに基づいて、LTP寄与分として用いられるLTP遅延バッファから適切な信号部分を選択して現在の信号を形状付ける。

【0526】

- LTPゲインを用いてこのLTP寄与分を再スケーリングする。

【0527】

- この再スケーリングされたLTP寄与分をLTP入力信号に加算してLTP出力信号を生成する。

【0528】

LTP遅延バッファ更新が実行される際、時間に関してそれぞれ異なる方策について検討することができる。

30

【0529】

フレーム n における最初のLTP演算として、最後のフレーム $n-1$ からの出力を用いる。これにより、フレーム n におけるLTP処理中に用いられるフレーム n におけるLTP遅延バッファが更新される。

【0530】

フレーム n における最後のLTP演算として、現在のフレーム n からの出力を用いる。これにより、フレーム $n+1$ におけるLTP処理中に用いられるフレーム n におけるLTP遅延バッファが更新される。

【0531】

以下、TCX LTPフィードバックループの結合解除について検討する。

40

【0532】

TCX LTPフィードバックループを結合解除することによって、封じ込めモード時にLTPデコーダの各フィードバックループ中における追加のノイズ（LTP入力信号に適用されるノイズ置き換えの結果得られるもの）の導入が回避される。

【0533】

図10は、この結合解除を示す。特に、図10は、封じ込め中のLTPフィードバックループの結合解除を示す（ $\text{bfi} = 1$ ）。

【0534】

図10は、遅延バッファ1020、サンプル選択部1030、及びサンプル処理部10

50

40 (サンプル処理部1040は、破線で示される)を示す。

【0535】

LTP遅延バッファ1020の更新が実行されるときに向けて、いくつかの実施例は、以下のように進行する。

【0536】

- 通常動作の場合、LTP遅延バッファ1020を最初のLTP演算として更新することが好ましいと考えられるが、それは、合計出力信号が、通常、永続的に記憶されるからである。この方策によって、専用のバッファを省略することができる。

【0537】

- 結合解除動作の場合、LTP遅延バッファ1020を最後のLTP演算として更新することが好ましいと考えられるが、それは、信号へのLTP寄与分が、通常、一時的に記憶されるだけであるからである。この方策によって、一時的なLTP寄与信号が保存される。実現例によっては、このLTP寄与バッファは、単に永続的なものとすることもできる。

10

【0538】

任意のケース(通常動作及び封じ込め)において後者の方策が用いられると仮定すると、実施例は、例えば、以下の事項を実現することができる。

【0539】

- 通常動作中、LTPデコーダの時間領域信号出力は、そのLTP入力信号への加算後に、LTP遅延バッファへの入力に用いられる。

20

【0540】

- 封じ込め中、LTPデコーダの時間領域信号出力は、そのLTP入力信号への加算前に、LTP遅延バッファへの入力に用いられる。

【0541】

いくつかの実施例では、TCX LTPゲインは、ゼロヘフェードさせられる。このような実施例では、TCX LTPゲインは、例えば、ある信号適応フェードアウト定数でゼロヘフェードさせることができる。これは、例えば、繰り返し行うことができ、例えば以下の疑似コードによって行うことができる。

【0542】

【数44】

30

```
gain=gain_past*damping;
```

```
[...]
```

```
gain_past=gain;
```

【0543】

ここで、

gainは、現在のフレームにおいて適用されるTCX LTPデコーダゲインであり、gain_pastは、先行するフレームにおいて適用されるTCX LTPデコーダゲインであり、

40

dampingは、(相対的な)フェードアウト定数である。

【0544】

図1dは、更なる実施例による装置を示し、当該装置は、更に、遅延バッファ180を含む長期予測部170を備える。長期予測部170は、第2の音声信号部分と、遅延バッファ180に記憶されている遅延バッファ入力と、長期予測ゲインとに応じて、処理済信号を生成するように構成される。更に、長期予測部は、複数のフレームのうちの前記第3のフレームが受信インターフェース110によって受信されていない場合、又は前記第3のフレームが受信インターフェース110によって受信されているが破損している場合、長期予測ゲインをゼロヘフェードさせるように構成される。

【0545】

50

他の実施例（図示せず）においては、長期予測部は、例えば、第1の音声信号部分と、遅延バッファに記憶されている遅延バッファ入力と、長期予測ゲインとに応じて、処理済信号を生成するように構成しても良い。

【0546】

図1dにおいては、第1の再構成部140は、例えば、更に、処理済信号に応じて第3の音声信号部分を生成することができる。

【0547】

実施例においては、長期予測部170は、例えば、長期予測ゲインをゼロヘフェードさせるように構成することができ、長期予測ゲインがゼロヘフェードさせられる速度は、フェードアウト定数に依存する。

10

【0548】

これに代えて、又はこれに加えて、長期予測部170は、例えば、複数のフレームのうちの前記第3のフレームが受信インターフェース110によって受信されていない場合、又は前記第3のフレームが受信インターフェース110によって受信されているが破損している場合、生成された処理済信号を遅延バッファ180に記憶させることによって遅延バッファ180の入力を更新するように構成され得る。

【0549】

上述のTCX LTPの使用に関し、より一般的な実施例を図13に示す。

【0550】

図13は、符号化音声信号を復号して再構成音声信号を得るための装置を示す。

20

【0551】

当該装置は、複数のフレームを受信するための受信インターフェース1310、復号音声信号の音声信号サンプルを記憶するための遅延バッファ1320、遅延バッファ1320に記憶されている音声信号サンプルから複数の選択音声信号サンプルを選択するためのサンプル選択部1330、及び、選択音声信号サンプルを処理して再構成音声信号の再構成音声信号サンプルを得るためのサンプル処理部1340を備える。

【0552】

サンプル選択部1330は、現在のフレームが受信インターフェース1310によって受信されており、且つ受信インターフェース1310によって受信されている現在のフレームが破損していない場合、現在のフレームに含まれるピッチ遅れ情報に応じて遅延バッファ1320に記憶されている音声信号サンプルから複数の選択音声信号サンプルを選択するように構成される。更に、サンプル選択部1330は、現在のフレームが受信インターフェース1310によって受信されていない場合、又は受信インターフェース1310によって受信されている現在のフレームが破損している場合、受信インターフェース1310によって以前に受信されている別のフレームに含まれるピッチ遅れ情報に応じて遅延バッファ1320に記憶されている音声信号サンプルから複数の選択音声信号サンプルを選択するように構成される。

30

【0553】

実施例によると、サンプル処理部1340は、例えば、現在のフレームが受信インターフェース1310によって受信されており、且つ受信インターフェース1310によって受信されている現在のフレームが破損していない場合、現在のフレームに含まれるゲイン情報に応じて選択音声信号サンプルを再スケーリングすることによって再構成音声信号サンプルを得るように構成され得る。更に、サンプル選択部1330は、例えば、現在のフレームが受信インターフェース1310によって受信されていない場合、又は受信インターフェース1310によって受信されている現在のフレームが破損している場合、受信インターフェース1310によって以前に受信されている前記別のフレームに含まれるゲイン情報に応じて選択音声信号サンプルを再スケーリングすることによって再構成音声信号サンプルを得るように構成され得る。

40

【0554】

実施例においては、サンプル処理部1340は、例えば、現在のフレームが受信インタ

50

ーフェース1310によって受信されており、且つ受信インターフェース1310によって受信されている現在のフレームが破損していない場合、選択音声信号サンプルと、現在のフレームに含まれるゲイン情報に応じた値とを乗算することによって再構成音声信号サンプルを得るように構成され得る。更に、サンプル選択部1330は、現在のフレームが受信インターフェース1310によって受信されていない場合、又は受信インターフェース1310によって受信されている現在のフレームが破損している場合、選択音声信号サンプルと、受信インターフェース1310によって以前に受信されている前記別のフレームに含まれるゲイン情報に応じた値とを乗算することによって再構成音声信号サンプルを得るように構成される。

【0555】

10

実施例によると、サンプル処理部1340は、例えば、再構成音声信号サンプルを遅延バッファ1320に記憶させるように構成され得る。

【0556】

実施例においては、サンプル処理部1340は、例えば、更なるフレームが受信インターフェース1310によって受信される前に再構成音声信号サンプルを遅延バッファ1320に記憶させるように構成され得る。

【0557】

実施例によると、サンプル処理部1340は、例えば、更なるフレームが受信インターフェース1310によって受信された後に再構成音声信号サンプルを遅延バッファ1320に記憶させるように構成され得る。

20

【0558】

実施例においては、サンプル処理部1340は、例えば、ゲイン情報に応じて選択音声信号サンプルを再スケーリングして再スケーリング音声信号サンプルを得て、再スケーリング音声信号サンプルを入力音声信号サンプルと組み合わせて処理済音声信号サンプルを得るように構成され得る。

【0559】

実施例によると、サンプル処理部1340は、例えば、現在のフレームが受信インターフェース1310によって受信されており、且つ受信インターフェース1310によって受信されている現在のフレームが破損していない場合、再スケーリング音声信号サンプルと入力音声信号サンプルとの組み合わせを示す処理済音声信号サンプルを遅延バッファ1320に記憶させ、再スケーリング音声信号サンプルを遅延バッファ1320に記憶させないように構成され得る。更に、サンプル処理部1340は、現在のフレームが受信インターフェース1310によって受信されていない場合、又は受信インターフェース1310によって受信されている現在のフレームが破損している場合、再スケーリング音声信号サンプルを遅延バッファ1320に記憶させ、処理済音声信号サンプルを遅延バッファ1320に記憶させないように構成される。

30

【0560】

別の実施例によると、サンプル処理部1340は、例えば、現在のフレームが受信インターフェース1310によって受信されていない場合、又は受信インターフェース1310によって受信されている現在のフレームが破損している場合、処理済音声信号サンプルを遅延バッファ1320に記憶させるように構成され得る。

40

【0561】

実施例においては、サンプル選択部1330は、例えば、変更後ゲインに応じて選択音声信号サンプルを再スケーリングすることによって再構成音声信号サンプルを得るように構成することができ、変更後ゲインは、

【0562】**【数45】**
$$\text{gain} = \text{gain_past} * \text{damping};$$
【0563】

50

の式によって定義され、ここで、 $gain$ は、変更後ゲインであり、サンプル選択部1330は、例えば、 $gain$ の算出後に $gain_past$ を $gain$ に設定するように構成することができ、 $damping$ は、実数値である。

【0564】

実施例によると、サンプル選択部1330は、例えば、変更後ゲインを算出するように構成され得る。

【0565】

実施例においては、 $damping$ は、例えば、 $0 < damping < 1$ によって定義され得る。

【0566】

実施例によると、変更後ゲイン $gain$ は、例えば、フレームが最後に受信インターフェース1310によって受信されてから少なくとも所定数のフレームが受信インターフェース1310によって受信されていない場合、ゼロに設定され得る。

【0567】

以下、フェードアウト速度について検討する。或る種のフェードアウトを適用するいくつかの封じ込めモジュールが存在する。フェードアウトの速度は、これらのモジュール間で異なって選択され得るが、1つのコア（ACELP又はTCX）について全ての封じ込めモジュールに対して同じフェードアウト速度を用いることが有益である。例えば、以下の通りである。

【0568】

ACELPの場合、特に適応コードブック（ゲインの変更による）及び／又は新規コードブック信号（ゲインの変更による）について同じフェードアウト速度を用いることが求められる。

【0569】

また、TCXの場合、特に時間領域信号、及び／又はLTPゲイン（ゼロへのフェード）、及び／又はLPC重み付け（1へのフェード）、及び／又はLPC係数（背景スペクトル形状へのフェード）、及び／又はホワイトノイズへのクロスフェードについて、同じフェードアウト速度を用いることが求められる。

【0570】

更に、ACELP及びTCXについても同じフェードアウト速度を用いることが好ましいと考えられるが、コアの異なる特性のため、異なるフェードアウト速度を用いることを選択することも考えられる。

【0571】

このフェードアウト速度は、静的であっても良いが、信号特性に対して適応的であることが好ましい。例えば、フェードアウト速度は、例えば、LPC安定度定数（TCX）及び／又は分類及び／又は連続して失われたフレームの数に依存し得る。

【0572】

フェードアウト速度は、例えば、減衰定数に応じて決定されても良く、これは、絶対的又は相対的に与えられることができ、また特定のフェードアウト中に時間の経過に伴い変化し得る。

【0573】

実施例においては、ホワイトノイズフェードについてと同様にLTPゲインフェードについても同じフェード速度を用いる。

【0574】

上述のように快適ノイズ信号を生成するための装置、方法及びコンピュータプログラムが提供された。

【0575】

装置の文脈でいくつかの局面を記載したが、これらの局面は、対応の方法の記載をも表すものであり、ブロック又は装置は、方法ステップ又は方法ステップの特徴に対応することは明らかである。同様に、方法ステップの文脈で記載した局面は、対応の装置の対応の

10

20

30

40

50

ブロック若しくは項目又は特徴の記載をも表す。

【0576】

本発明による分解された信号は、デジタル記憶媒体で記憶することができ、或るいは、伝送媒体、例えば無線伝送媒体又はインターネットのような有線伝送媒体、で送信することができる。

【0577】

特定の実現要件に応じて、本発明の実施例は、ハードウェア又はソフトウェアによって実現され得る。その実現は、デジタル記憶媒体、例えば、フロッピー（登録商標）ディスク、DVD、CD、ROM、PROM、EPROM、EEPROM又はフラッシュメモリであって、電子的に読み出し可能な制御信号を格納しており、プログラム可能なコンピュータシステムと協働する（又は協働可能である）ことによりそれぞれの方法が実行されるようにするものを用いて実行され得る。

10

【0578】

本発明によるいくつかの実施例は、プログラム可能なコンピュータシステムと協働可能であることによって本願明細書に記載の方法の1つが実行されるようにする、電子的に読み出し可能な制御信号を有する非一時的データキャリアを含む。

【0579】

一般的には、本発明の実施例は、プログラムコードを有するコンピュータプログラム製品であって、このコンピュータプログラム製品がコンピュータにおいて実行されるときに上記プログラムコードが上記方法の1つを実行するように動作するものとして実現され得る。プログラムコードは、例えば、機械読み取り可能キャリアに格納され得る。

20

【0580】

他の実施例は、機械読み取り可能キャリアに格納された、本願明細書に記載の方法の1つを実行するためのコンピュータプログラムを含む。

【0581】

従って、換言すると、本発明の方法の一実施例は、コンピュータプログラムであって、このコンピュータプログラムがコンピュータにおいて実行されるときに、本願明細書に記載の方法の1つを実行するためのプログラムコードを有するものである。

【0582】

従って、本発明の方法の更なる実施例は、データキャリア（又はデジタル記憶媒体若しくはコンピュータ読み取り可能媒体）であって、そこに記録された、本願明細書に記載の方法の1つを実行するためのコンピュータプログラムを含むものである。

30

【0583】

従って、本発明の方法の更なる実施例は、本願明細書に記載の方法の1つを実行するためのコンピュータプログラムを表すデータストリーム又は信号シーケンスである。データストリーム又は信号シーケンスは、例えば、インターネットを介したデータ通信接続を介して転送されるように構成され得る。

【0584】

更なる実施例は、本願明細書に記載の方法の1つを実行するように構成又は適合された処理手段、例えば、コンピュータ又はプログラム可能論理装置を含む。

40

【0585】

更なる実施例は、本願明細書に記載の方法の1つを実行するためのコンピュータプログラムをインストールしたコンピュータを含む。

【0586】

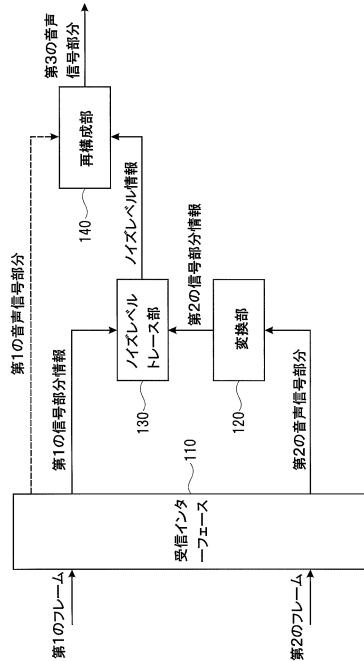
いくつかの実施例においては、プログラム可能論理装置（例えば、フィールドプログラマブルゲートアレイ）を用いて、本願明細書に記載の方法におけるいくつか又は全ての機能を実行しても良い。いくつかの実施例においては、フィールドプログラマブルゲートアレイは、マイクロプロセッサと協働して、本願明細書に記載の方法の1つを実行しても良い。一般的に、当該方法は、どのようなハードウェア装置によって実行されても良い。

【0587】

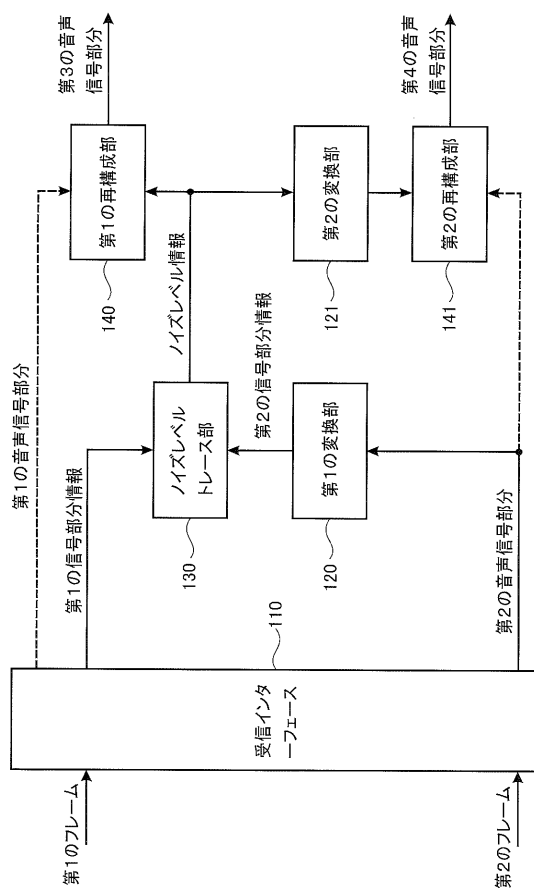
50

上述の各実施例は、単に本発明の原理を例示するものである。本願明細書に記載の構成及び詳細を変更及び変形したものが当業者には明らかであることが理解される。従って、本願明細書における各実施例の記載及び説明として呈示された特定の詳細によってではなく、添付の特許請求の範囲によってのみ限定されることが意図される。

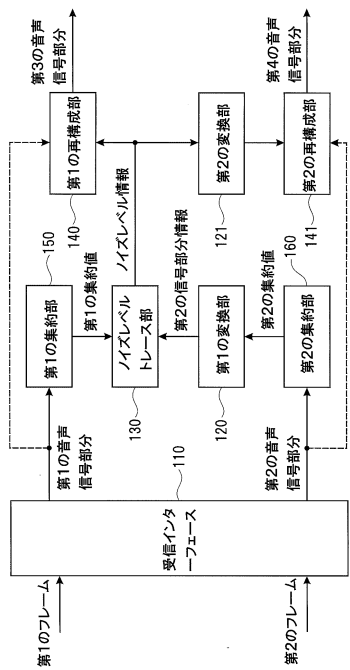
【図 1 A】



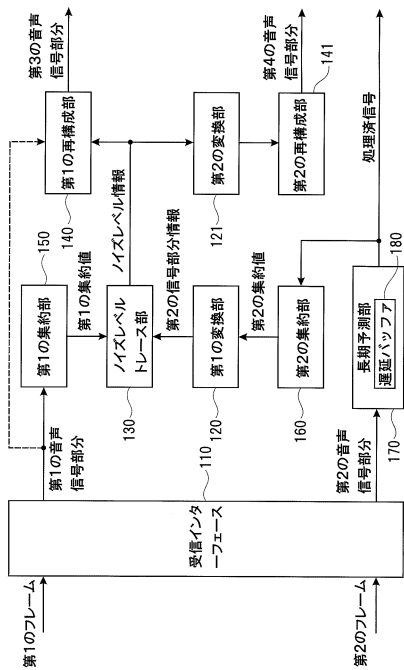
【図 1 B】



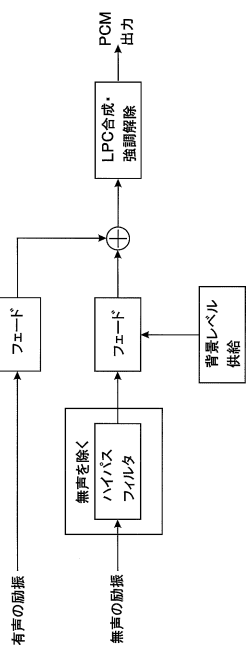
【図 1 C】



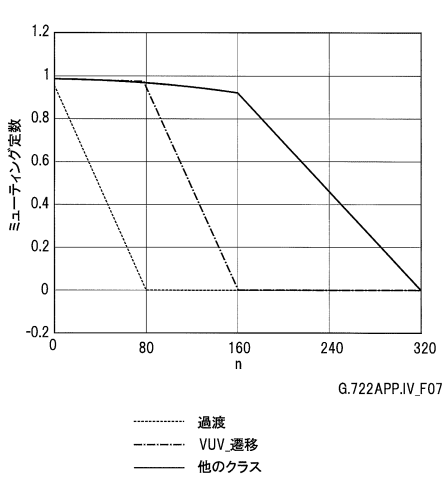
【図 1 D】



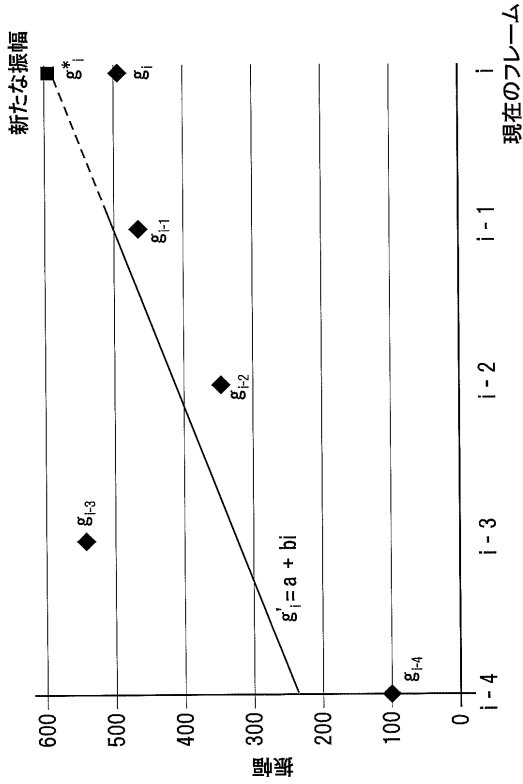
【図 2】



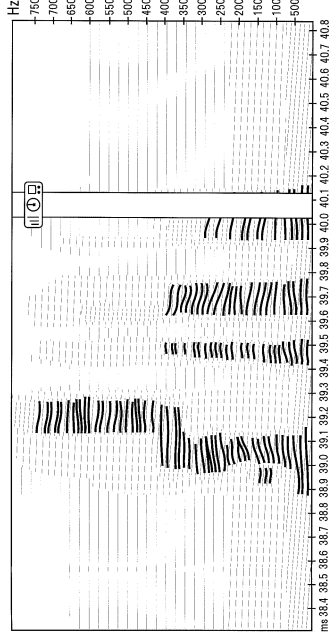
【図 3】



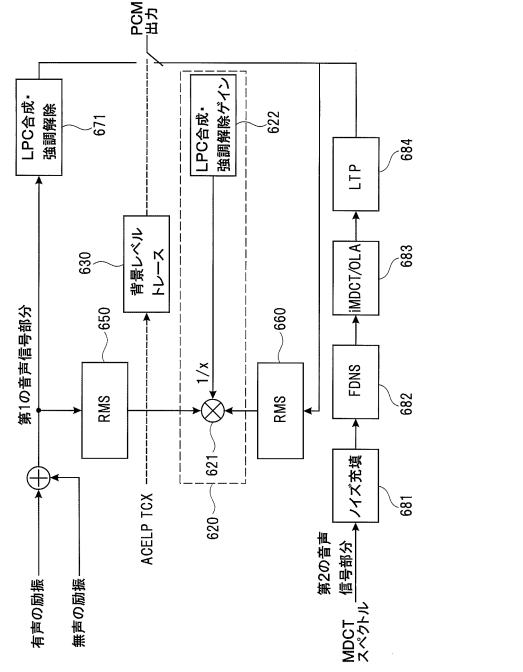
【図 4】



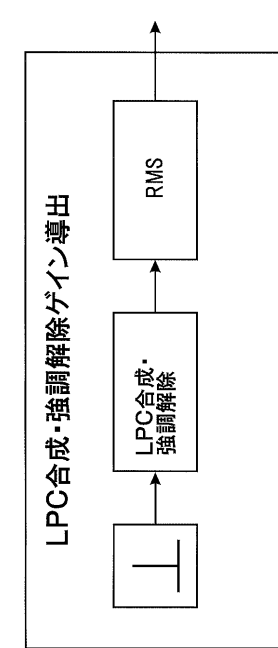
【図 5】



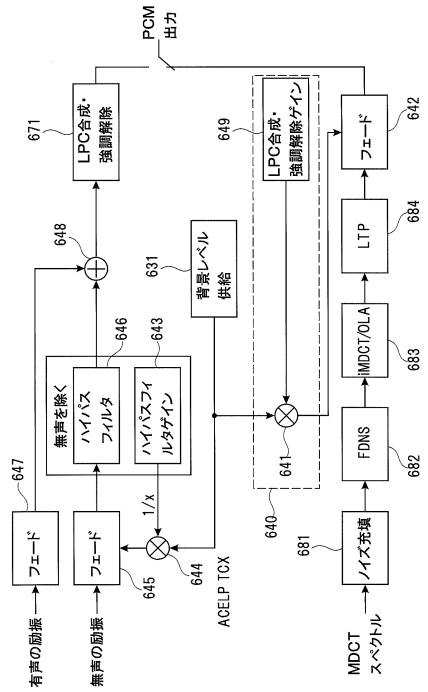
【図 6】



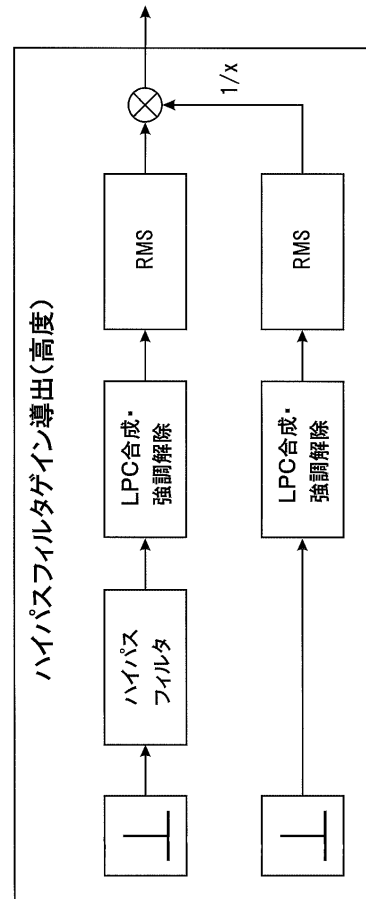
【図 7】



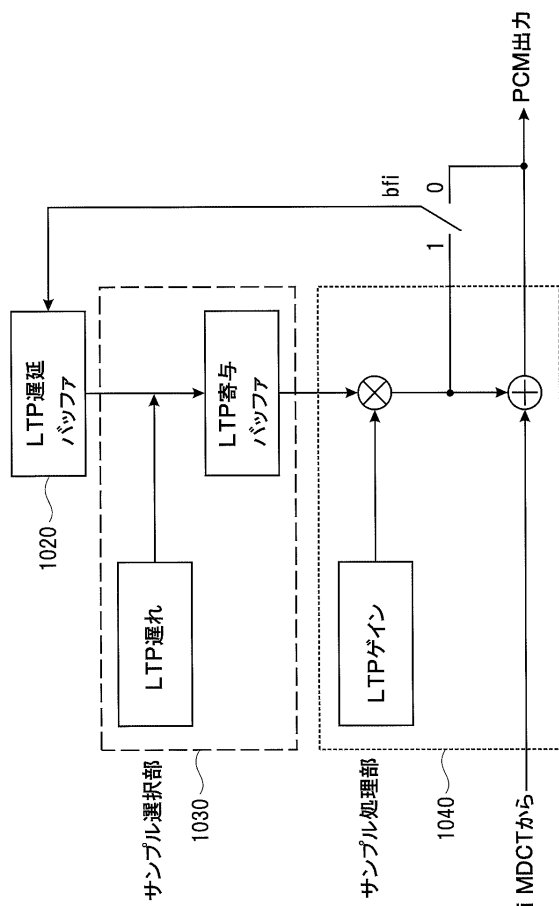
【図 8】



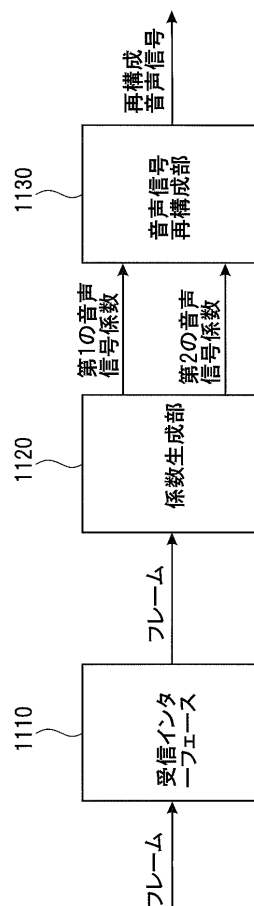
【図 9】



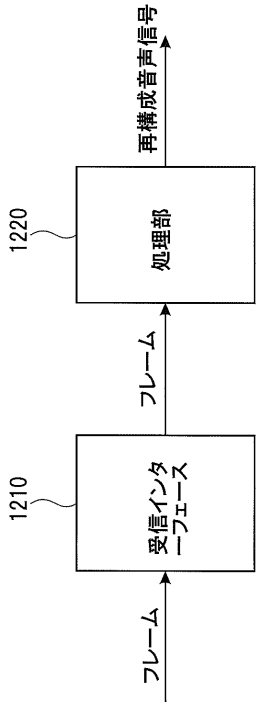
【図 10】



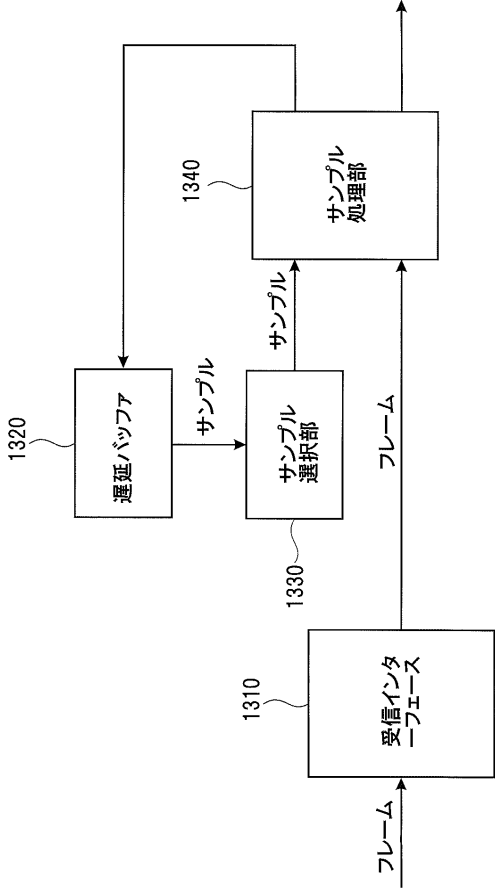
【图 1 1】



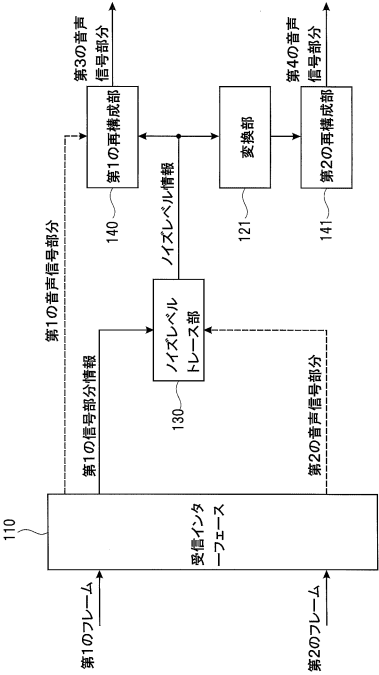
【図 1 2】



【図 1 3】



【図 1 4】



フロントページの続き

(74)代理人 100130638

弁理士 野末 貴弘

(72)発明者 シュナーベル、 ミヒャエル

ドイツ国 9 5 1 7 9 ゲーロルツグリューン、 シュナイトベルクヴェーク 4アー

(72)発明者 マールコヴィッチ、 ゴーラン

ドイツ国 9 0 4 2 5 ニュルンベルク、 アーヘナー・シュトラーセ 1 9

(72)発明者 シュペルシュナイダー、 ラルフ

ドイツ国 9 1 3 2 0 エーバーマンシュタット、 デーベルト 7 5

(72)発明者 ルコント、 ジェレミー

ドイツ国 9 0 7 6 3 フュルト、 ウルシュタインシュトラーセ 3 2

(72)発明者 ヘルムリヒ、 クリスティアン

ドイツ国 9 1 0 5 4 エルランゲン、 ハウプトシュトラーセ 6 8

審査官 五貫 昭一

(56)参考文献 特開2002-328700 (JP, A)

特表2004-501391 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G 1 0 L 1 9 / 0 9

G 1 0 L 1 9 / 0 0 5