

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5154934号
(P5154934)

(45) 発行日 平成25年2月27日 (2013. 2. 27)

(24) 登録日 平成24年12月14日 (2012. 12. 14)

(51) Int. Cl.

F I

G 1 O L 19/00 (2013. 01)
G 1 O L 19/16 (2013. 01)
H O 3 M 7/30 (2006. 01)
G 1 O L 19/24 (2013. 01)
G 1 O L 19/20 (2013. 01)

G 1 O L 19/00 2 5 O
G 1 O L 19/14 4 O O Z
H O 3 M 7/30 Z
G 1 O L 19/14 4 O O D
G 1 O L 19/14 4 O O B

請求項の数 17 (全 23 頁)

(21) 出願番号 特願2007-531883 (P2007-531883)
(86) (22) 出願日 平成17年9月2日 (2005. 9. 2)
(65) 公表番号 特表2008-513823 (P2008-513823A)
(43) 公表日 平成20年5月1日 (2008. 5. 1)
(86) 国際出願番号 PCT/IB2005/052879
(87) 国際公開番号 W02006/030340
(87) 国際公開日 平成18年3月23日 (2006. 3. 23)
審査請求日 平成20年8月29日 (2008. 8. 29)
(31) 優先権主張番号 04104514. 7
(32) 優先日 平成16年9月17日 (2004. 9. 17)
(33) 優先権主張国 欧州特許庁 (EP)

(73) 特許権者 590000248
コーニンクレッカ フィリップス エレク
トロニクス エヌ ヴィ
オランダ国 5 6 2 1 ベーアー アイン
ドーフエン フルーネヴァウツウェッハ
1
(74) 代理人 100070150
弁理士 伊東 忠彦
(74) 代理人 100091214
弁理士 大貫 進介
(74) 代理人 100107766
弁理士 伊東 忠重
(74) 代理人 100135079
弁理士 宮崎 修

最終頁に続く

(54) 【発明の名称】 知覚的歪みを最小化する結合オーディオ符号化

(57) 【特許請求の範囲】

【請求項 1】

オーディオ信号を符号化するように構成されたオーディオエンコーダであって：

第 1 の符号化信号部分を発生するように構成された第 1 のエンコーダと、

少なくとも第 2 の符号化信号部分を発生するように構成された第 2 のエンコーダと、

制御部であって：

前記オーディオ信号を表し、歪みの大きさに関する前記第 1 と第 2 の符号化信号部分を有する符号化オーディオ信号を評価するように構成された、前記歪みの大きさは前記符号化オーディオ信号と前記オーディオ信号との間の差の大きさである、評価手段と、

所定基準に従って符号化パラメータを最適化するために、前記第 1 と第 2 のエンコーダの両方の符号化パラメータを調整し、その調整に応じて前記歪みの大きさを監視するように構成された最適化手段とを有する制御部と、

を有するオーディオエンコーダ。

【請求項 2】

歪みの大きさは知覚的歪みの大きさを含む、請求項 1 に記載のオーディオエンコーダ。

【請求項 3】

最適化手段は前記歪みの大きさを最小化するように前記符号化パラメータを調節するように構成された、請求項 1 に記載のオーディオエンコーダ。

【請求項 4】

最適化手段は、第 1 と第 2 のエンコーダの所定の最大トータルビットレートの制約下で

歪みの大きさを最小化するように構成された、請求項 3 に記載のオーディオエンコーダ。

【請求項 5】

最適化手段は、所定の最大トータルビットレート内で、第 1 と第 2 のビットレートをそれぞれ第 1 と第 2 のエンコーダに分配することにより歪みの大きさを最小化するように構成された、請求項 4 に記載のオーディオエンコーダ。

【請求項 6】

第 1 のエンコーダはオーディオ信号を第 1 の符号化信号部分に符号化するように構成され、第 2 のエンコーダはオーディオ信号と第 1 の符号化信号部分の差である第 1 の残差信号を第 2 の符号化信号部分に符号化するように構成された、請求項 1 に記載のオーディオエンコーダ。

10

【請求項 7】

歪みの大きさは、第 1 の残差信号と第 2 の符号化信号部分の差である第 2 の残差信号に基づく、請求項 6 に記載のオーディオエンコーダ。

【請求項 8】

オーディオ信号を第 1 と第 2 の部分に分離するように構成された信号スプリッタをさらに有し、

第 1 のエンコーダは第 1 のオーディオ信号部分を第 1 の符号化信号部分に符号化するように構成され、第 2 のエンコーダは第 2 のオーディオ信号部分を第 2 の符号化信号部分に符号化するように構成された、請求項 1 に記載のオーディオエンコーダ。

【請求項 9】

20

最適化手段は、所定の最大の歪みの大きさの制約下で第 1 と第 2 の信号部分のトータルビットレートを最小化するように構成された、請求項 1 に記載のオーディオエンコーダ。

【請求項 10】

第 1 のエンコーダは：パラメトリックエンコーダ、変換エンコーダ、サブバンドエンコーダ、定常パルス励起エンコーダ、符号ブック励起リニア予測エンコーダよりなる群から選択されたエンコーダである、請求項 1 に記載のオーディオエンコーダ。

【請求項 11】

第 2 のエンコーダは：パラメトリックエンコーダ、変換エンコーダ、サブバンドエンコーダ、定常パルス励起エンコーダ、符号ブック励起リニア予測エンコーダよりなる群から選択されたエンコーダである、請求項 1 に記載のオーディオエンコーダ。

30

【請求項 12】

オーディオエンコーダは重ならないセグメントに分割されたオーディオ信号を受信するように構成され、最適化手段はオーディオ信号の 1 つ以上の後続セグメントにわたって符号化パラメータを最適化するように構成された、請求項 1 に記載のオーディオエンコーダ。

【請求項 13】

オーディオエンコーダは重なっているセグメントに分割されたオーディオ信号を受信するように構成され、最適化手段はオーディオ信号の 1 つ以上の後続セグメントにわたって符号化パラメータを最適化するように構成された、請求項 1 に記載のオーディオエンコーダ。

40

【請求項 14】

第 3 の符号化信号部分を発生するように構成された第 3 のエンコーダをさらに有し、前記符号化オーディオ信号は第 1、第 2、及び第 3 の符号化信号部分を含む、請求項 1 に記載のオーディオエンコーダ。

【請求項 15】

オーディオ信号の符号化方法であって、

第 1 のエンコーダを用いて第 1 の符号化信号部分を発生する段階と、

第 2 のエンコーダを用いて少なくとも第 2 の符号化信号部分を発生する段階と、

前記オーディオ信号を表し、歪みの大きさに関する前記第 1 と第 2 の符号化信号部分を有する符号化オーディオ信号を評価する、前記歪みの大きさは前記符号化オーディオ信号

50

と前記オーディオ信号との間の差の大きさである、段階と、

所定の基準に従って前記歪みの大きさに応じて前記第 1 と第 2 のエンコーダの両方の符号化パラメータを最適化する段階とを有する方法。

【請求項 16】

請求項 1 に記載のオーディオエンコーダを有する装置。

【請求項 17】

請求項 15 の方法に記載のオーディオ信号を符号化するように構成されたコンピュータ読み取り可能なプログラムコード。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は高品質低ビットレートのオーディオ信号符号化に関する。本発明は、特に、目標ビットレートを考慮しつつ、知覚される音声品質に関して最適化された効果的な符号化に関する。より具体的には、本発明は複数のエンコーダを用いて結合符号化信号表現を生成するオーディオ信号符号化に関する。また、本発明は、エンコーダ、デコーダ、符号化及び復号方法、符号化オーディオ信号、かかる符号化信号を表すデータを有する記憶及び伝送媒体、エンコーダ及び/またはデコーダを有するオーディオ装置にも関する。

【背景技術】

【0002】

高品質のオーディオ符号化の分野では、多様なオーディオ信号に対して音声品質対ビットレートの点で最適な結果を提供するには、異なる符号化方法が必要だということは周知である。1つの符号化方法は、あるタイプのオーディオ信号に対してはよい結果を与えるが、他のタイプのオーディオ信号では性能がよくない。非常に低ビットレートでは正弦波エンコーダとノイズモデルが最も効率的であり、波形符号化包はビットレートが高いときによりよい結果を与える。

20

【0003】

現在の M P E G 2 と M P E G 4 標準では、ビットレートに応じて符号化ストラテジ (strategies) を変更すると効率的であることが分かっている。そのため、この標準には幅広いオーディオエンコーダが含まれており、限られたビットレートの範囲で最良の結果を得られるようになっている。

30

【0004】

しかし、通常、オーディオ信号には、短い時間であっても多様な信号特性が混ざっている。それゆえ、一般的には、数秒のオーディオ信号であっても純粋音、ノイズ、過渡音等が支配的な部分が含まれている。これらは特徴が異なるため、最適符号化のための符号化特性も異なる。すなわち、1種類のエンコーダを使用しても、信号のある部分に対しては、ビットレートまたは品質の点で非常に貧しい結果しか得られないことがある。

【0005】

Scott Levineによる博士論文 [1] (添付した参照文献リストを参照) は、正弦波 (すなわちパラメトリック) エンコーダと波形エンコーダが混在したエンコーダを記載している。オーディオ信号の大部分はパラメトリックエンコーダで符号化されるが、オーディオ信号の過渡的部分だけには波形エンコーダが使用される。この仕組みでは、パラメトリックエンコーダと波形エンコーダの間の分割は所定のものである。

40

【0006】

フィリップス社の米国特許第 5, 808, 569 号には、信号の異なる部分を 2 つの異なる符号化ストラテジを用いて符号化する符号化スキーム (scheme) が記載されている。しかし、ビットレートをエンコーダにわたって以下に配分するかを決定する仕様は与えられていない。

【0007】

このように、先行技術のオーディオエンコーダは、オーディオ信号のパラメータの変化に応じて 2 つ以上の異なる符号化スキームを制御する問題を解決していない。

50

【発明の開示】

【課題を解決するための手段】

【0008】

本発明の1つの目的は、多様なオーディオ信号特性と異なる目標ビットレートに対して効率が高い高品質なオーディオ符号化を提供できる、柔軟なオーディオエンコーダを提供することである。

【0009】

本発明の第1の態様によると、この目的は、オーディオ信号を符号化するオーディオエンコーダにより達成される。該オーディオエンコーダは：

- 第1の符号化信号部分を発生する第1のエンコーダと、
- 少なくとも、第2の符号化信号部分を発生する第2のエンコーダと、
- 制御部であって、
- 歪みの大きさに関して第1と第2の符号化信号部分を有するオーディオ信号のジョイント表現を評価する評価手段と、
- 第1と第2のエンコーダの少なくとも一方の符号化パラメータを調整し、その調整に応じて、所定基準に従って符号化パラメータを最適化するように、オーディオ信号のジョイント表現の歪みの大きさを監視する最適化手段とを有する制御部と、を有する。

10

【0010】

用語「歪みの大きさ(distortion measure)」は、オーディオ信号と符号化オーディオ信号(すなわち、オーディオ信号のジョイント表現(joint representation))間の差の任意の尺度と解釈すべきである。

20

【0011】

用語「符号化パラメータ」は、具体的なエンコーダに対して調整できる1つ以上の符号化変数の1つとして広く解釈すべきである。これらの符号化パラメータの性質はエンコーダの種類に応じて決まる。

【0012】

第1の態様によるオーディオエンコーダは、オーディオ信号の各信号(excerpt)の最適な符号化を適応させて、最大ビットレートを制限した場合に、2つのジョイントエンコーダを最適に利用して、知覚的歪みを最低にする、すなわち知覚される音声品質を最良にする。特に、まったく異なる符号化原理を使用する第1と第2のエンコーダを選択して、効率的に符号化する。例えば、ある信号特性を有する信号(excerpt)の場合、第1のエンコーダがほとんど全てのビットレートを使用した場合に符号化が最も効率的になり、次の信号(excerpt)の特性は異なり、最適な符号化をするには両方のエンコーダをミックスして使用する必要がある。第1の態様によるエンコーダは、異なるオーディオ信号の特性に適応することができ、異なる最大ビットレートの制限において性能が最適になる。あるエンコーダは特定のビットレートで最良の性能を発揮することが分かっている。このことは、2つのエンコーダを最適にミックスすることにより考慮され、目標ビットレートの広い範囲にわたって最適な符号化効率を得られる。第1と第2のエンコーダの両方の符号化パラメータが好ましくは最適化される。

30

【0013】

原理的には、本発明によるエンコーダにより、広い範囲の基準に従って、個々のエンコーダの符号化パラメータの最適化が可能である。一実施形態では、最適化手段は、歪みの大きさを最小化するように符号化パラメータを調整する、すなわち、この基準に従って、利用可能なビットレートを考慮せずに音声品質を最適化する。しかし、この実施形態は、第1と第2のエンコーダの所定の最大トータルビットレートの制約により修正されてもよい。

40

【0014】

他の実施形態では、最適化手段は、所定の最大トータルビットレート内で第1と第2のビットレートをそれぞれ第1と第2のエンコーダに分配することにより歪みの大きさを最小化する、このオーディオエンコーダの実施形態は、歪みを最小化するように、2つのエ

50

ンコーダ間で最も効率的にトータルビットレートを分配しようとするものである。2つのエンコーダの簡単な実施形態であって、固定のビットレートのセットに限定され、ビットレートの合計が一定である場合、最適化手段は2つのエンコーダ間のビットレートの分配を調整するのみでよい。

【0015】

他の実施形態では、最適化手段は、所定の最大の歪みの大きさの制約下で第1と第2の信号部分のトータルビットレートを最小化する。この実施形態によると、最適化基準は歪みの大きさが一定の場合にトータルビットレートを最小化する。好ましい実施形態では、歪みの大きさは知覚的歪みの大きさを含む、用語「知覚的歪みの大きさ(perceptual distortion measure)」は、例えば、音響心理的モデルに従って、知覚される音声品質に対して符号化信号が歪んでいる程度を表す量として、広く解釈すべきである。言い換えると、符号化信号の知覚的歪みの大きさは、リスナーが知覚できる、元の入力オーディオ信号の劣化の程度を表す量である。明らかに、符号化信号の音声品質を最適化するという目標を達成するためには、この大きさ(measure)を最小化することが好ましい。

【0016】

好ましい実施形態では、第1のエンコーダはオーディオ信号を第1の符号化信号部分に符号化し、第2のエンコーダはオーディオ信号と第1の符号化信号部分の差である第1の残差信号を第2の符号化信号部分に符号化する。この実施形態は、2つのエンコーダのカスケードであって、第1のエンコーダが符号化しなかった原信号の残余部分を第2のエンコーダが符号化するものである。歪みの大きさは、好ましくは、第1の残差信号と第2の符号化信号部分の差である第2の残差信号に基づく。すなわち、2つのエンコーダが符号化しなかった原信号の残余部分は、歪みの尺度(distortion measure)を発生するために、原オーディオ信号とともに使用される。一般的な言葉で言うと、3つ以上のエンコーダのカスケードにおいては、各エンコーダはそのカスケードの先行するエンコーダの残差信号を符号化し、カスケードの最後のエンコーダで符号化されなかった残りの信号を最適化プロセスの制御部への入力として使用する。

【0017】

他の好ましい実施形態では、オーディオエンコーダは、オーディオ信号を第1と第2の部分に分離する信号スプリッタをさらに有し、第1のエンコーダは第1のオーディオ信号部分を第1の符号化信号部分に符号化し、第2のエンコーダは第2のオーディオ信号部分を第2の符号化信号部分に符号化する。この実施形態では、第1と第2のエンコーダが並行して動作する。例えば、信号スプリッタはオーディオ信号を異なる周波数範囲に分離するフィルタバンクを有する。

【0018】

オーディオエンコーダは、第3の符号化信号部分を発生する第3のエンコーダをさらに有し、制御部は第1、第2、及び第3の符号化信号部分を含むオーディオ信号のジョイント表現を処理してもよい。3つのエンコーダは、上記のようにカスケードでも並行でも、またはこれらの組み合わせで動作してもよい。オーディオエンコーダのエンコーダは3つより多くてもよく、4つ、5つ、6つ、またはそれ以上であってもよい。これらは、カスケードされ、並列に結合され、またはカスケードと並列の組み合わせで結合されてもよい。複数のエンコーダは、異なる種類であってもよく、少なくとも2つの異なる種類のものである。

【0019】

最適化手段は、好ましくは、第1と第2のエンコーダの第1と第2の符号化テンプレートの所定のセットから、所定の基準に従って最良の性能になる第1と第2の符号化テンプレートのペアを選択する。ここで、「符号化テンプレート」は、エンコーダに対して、調節可能な符号化パラメータの所定のセットを意味すると解釈すべきである。「所定のテンプレートのセット」は、そのエンコーダに対して、異なる符号化パラメータのセットを意味すると解釈すべきである。

【0020】

10

20

30

40

50

第1のエンコーダは、好ましくはパラメトリックエンコーダ（例えば、正弦波エンコーダ）、変換エンコーダ、定常パルス励起エンコーダ（Regular Pulse Excitation encoders）、符号ブック励起リニア予測エンコーダ（Codebook Excited Linear Prediction encoders）よりなる群から選択されたエンコーダである。第2のエンコーダは好ましくは同じグループから選択されたエンコーダである。第1のエンコーダは結合エンコーダであってもよい。最も好ましくは、第1と第2のエンコーダは異なる種類であり、最良の助様態で互いに補完する。しかし、第1と第2のエンコーダは、同じ種類でもよいが、符号化テンプレートは異なる。

【0021】

オーディオエンコーダは、好ましくはセグメントに分割されたオーディオ信号を受信する。最適化手段は、好ましくは、オーディオ信号の1つ以上の後続セグメントにわたって符号化パラメータを最適化する。これらのセグメントは重なっていてもいなくてもよい。より好ましくは、3つ以上の後続のセグメントを最適化プロセスで使用する。

【0022】

本発明の第2の態様は、符号化オーディオ信号を復号するオーディオデコーダである。該オーディオエンコーダは、

- 第1の符号化信号部分から第1の復号信号部分を発生する第1のデコーダと、
- 第2の符号化信号部分から第2の復号信号部分を発生する第2のデコーダと、
- 第1と第2の復号信号部分の合計としてオーディオ信号の表現を発生する合計手段とを有する。

【0023】

第1と第2のデコーダは、符号化プロセスでを使用したものと同じ種類のものであることが必要である。さもないと、これらのデコーダは、正弦波パラメータ等のエンコーダ特有のデータを含む第1と第2の符号化信号を復号できない。デコーダは符号化信号の各部分で完全に平行して動作する。

【0024】

好ましい第1と第2のデコーダは、オーディオエンコーダに関して上で列挙したのエンコーダと対応する種類のものから選択できる。

【0025】

オーディオデコーダに関して、デコーダは、第3の符号化信号部分から第3の復号信号部分を発生する第3のデコーダをさらに有し、合計手段は第1、第2、及び第3の復号信号部分としてオーディオ信号の表現を発生する。オーディオデコーダは、各々が符号化オーディオ信号の個々の部分を復号する第4、第5、第6、またはそれ以上の別のデコーダをさらに有してもよい。全ての復号信号部分を加えて出力オーディオ信号を発生する。

【0026】

第3の態様では、本発明は、オーディオ信号の符号化方法を提供する。該方法は：

- 第1のエンコーダを用いて第1の符号化信号部分を発生する段階と、
- 第2のエンコーダを用いて少なくとも第2の符号化信号部分を発生する段階と、
- 歪みの大きさに関して第1と第2の符号化信号部分を有するオーディオ信号のジョイント表現を評価する段階と、
- 所定の基準に従って歪みの大きさに応じて第1と第2のエンコーダの符号化パラメータを最適化する段階とを有する。

【0027】

第1の態様に対するのと同じ説明が当てはまる。

【0028】

第4の態様では、本発明は、オーディオ信号の復号方法を提供する。該方法は：

- 第1のデコーダを用いて第1の符号化信号部分から第1の復号信号部分を発生する段階と、
- 第2のデコーダを用いて第2の符号化信号部分から第2の復号信号部分を発生する段階と、

- 第 1 と第 2 の復号信号部分を加える段階とを有する。

【 0 0 2 9 】

第 2 の態様に対するのと同じ説明が当てはまる。

第 5 の態様では、本発明は、異なるエンコーダが符号化した第 1 と第 2 の符号化信号部分を含む符号化オーディオ信号である。

【 0 0 3 0 】

符号化信号は、標準のデジタルオーディオフォーマットに従ったフォーマットを有するデジタル電気信号であってもよい。この信号は、電気接続ケーブルを用いて 2 つのオーディオ装置間を送信される。しかし、符号化信号は、ラジオ周波数の搬送波を用いた空中波等のワイヤレス信号でもよく、光ファイバを通して送信できる光信号であってもよい。

10

【 0 0 3 1 】

第 6 の態様では、本発明は、第 5 の態様による符号化オーディオ信号を表すデータを含む記憶媒体である。記憶媒体は、好ましくは、DVD、DVD-ROM、DVD-R、DVD+RW、CD、CD-R、CD-RW、コンパクトフラッシュ（登録商標）、メモリスティック等の標準的なオーディオデータ記憶媒体である。しかし、コンピュータのハードディスク、メモリ、フロッピー（登録商標）ディスク等のコンピュータデータ記憶媒体であってもよい。

【 0 0 3 2 】

第 7 の態様では、本発明は第 1 の態様によるオーディオエンコーダを有する装置である。

20

【 0 0 3 3 】

第 8 の態様では、本発明は、第 2 の態様によるオーディオデコーダを有するオーディオ装置である。

【 0 0 3 4 】

第 7 と第 8 の態様による好ましい装置は、テープ、ディスク、またはメモリベースのオーディオレコーダ、及び例えば、固体プレーヤ、DVDプレーヤ、コンピュータのオーディオプロセッサ等のプレーヤ等のオーディオ装置である。また、携帯電話にも有利である。

【 0 0 3 5 】

第 9 と第 10 の態様は、コンピュータ読み取り可能なプログラムコード、すなわちソフトウェアであって、該プログラムコードは、それぞれ第 3 と第 4 の態様による符号化方法と復号方法を実施するアルゴリズムを有している。

30

【 0 0 3 6 】

本発明の範囲内において様々な修正と代替形式が可能であるが、具体的な実施形態は図面中で例として示したものであり、以下に詳細に説明する。しかし、本発明は開示された具体的な形式に限定されるものではないことに留意すべきである。本発明は、添付した請求の範囲に記載した本発明の精神と範囲に含まれるすべての修正、等価物、代替物を含むものである。

【発明を実施するための最良の形態】

【 0 0 3 7 】

本発明を添付した図面を参照して詳細に説明する。

40

【 0 0 3 8 】

図 1 は、第 1 の単純なエンコーダの実施形態の原理を示すブロック図であり、一定のトータル目標ビットレート毎フレームで動作する 2 つの異なるエンコーダ AE 1 と AE 2 のカスケードを含む。フレームの定義は、単一のセグメントと時間的に等しいかそれより長い時間である。第 1 のエンコーダ AE 1 は好ましくは正弦波エンコーダを有し、第 2 のエンコーダ AE 2 は変換エンコーダを有する。正弦波符号化方法は低ビットレートで効率的であり、同じ低ビットレートにおける波形エンコーダと比較して音声品質がよい。変換エンコーダはより大きなビットレートを必要とするが、正弦波エンコーダよりも音声品質がよいことが知られている。よって、要するに組み合わせることにより柔軟なオーディオエ

50

ンコードとなる。

【 0 0 3 9 】

図 1 に示した符号化スキームでは、オーディオ信号 0 の一部分が、目標ビットレートのある比率 R_1 を用いて第 1 のエンコード A E 1 により符号化される。第 1 のエンコード A E 1 が使えるビットレートの比率 R_1 は制御部 C U が制御する。第 1 のエンコード A E 1 における正弦波符号化の後、第 1 の符号化信号部分 E 1 (すなわち、量子化されていない正弦波表示) が元の入力信号 0 から減算され、残差信号 1 (すなわち、正弦波エンコード A E 1 によりモデル化されていない信号の部分) が得られる。残差信号 1 は、第 2 のエンコード A E 2 (すなわち波形エンコード) により符号化されて第 2 の符号化信号部分 E 2 になる。これには、フレームの符号化に利用できるトータルビットレートの残りの部分 R_2 が使われる。

10

【 0 0 4 0 】

この実施形態では、制御部 C U は、結合符号化信号 E 1、E 2 の知覚的音声品質を最適化する。この最適化は、2つのエンコード A E 1、A E 2 の間でいくつかのビットレート R_1 、 R_2 の分配をテストして、知覚的歪みの大きさに関して符号化結果を評価することにより行う。知覚的歪みの大きさを提供するために、知覚的モデルを使用することが好ましい。知覚的歪みを予測する方法を明示的に提案する好ましいモデルは、参照文献 [4] のモデルである。一般的に、この最適化は、エンコードをローカルな信号特性に適應させるために、フレームごとに行われる必要がある。

【 0 0 4 1 】

20

制御部 C U は、2つのエンコード A E 1、A E 2 の間のビットレートの具体的分配の知覚的歪みの大きさを記憶し、最良の分配を見つけるまで、他の分配を試みる。この目的のため、制御部 C U は、第 2 のエンコード A E 2 の後のエラー信号 2 を元の入力信号 0 と比較する。エラー信号すなわち残差信号 2 の定義は、第 1 の残差信号 1 と第 2 の符号化信号部分 E 2 の差異であり、換言すると、2つのエンコード A E 1、A E 2 により符号化されていない最終的な残差信号 (rest signal) である。

【 0 0 4 2 】

ビットレートの分配 R_1 、 R_2 の所定のセットをテストしてから、制御部 C U は決定された知覚的歪みの大きさから、知覚的歪みが最小になる使用すべきビットレートの分配 R_1 、 R_2 を決定する。この分配 R_1 、 R_2 に従って、結果として得られた第 1 の信号部分 E 1 と第 2 の信号部分 E 2 (すなわち、エンコード A E 1、A E 2 からそれぞれ得られるパラメータとデータ) は、符号化出力ビットストリーム O U T を出力するようにビットストリームフォーマッタ B S F (Bit Stream Formatter) により処理される。

30

【 0 0 4 3 】

テストされるビットレートの分配 R_1 、 R_2 の所定のセットは、例えば、トータル目標ビットレート (すなわち、 $R_1 + R_2$) の 5 %、10 %、20 % または 25 % のステップである全ての組み合わせである。例えば、目標ビットレートが 64 k b p s の場合、(R_1 、 R_2) の組み合わせは (0, 64)、(16, 48)、(32, 32)、(48, 16)、(64, 0) とすることができる。

【 0 0 4 4 】

40

正弦波エンコード A E 1 が波形エンコード A E 2 よりも効率的になる交代点は、符号化される具体的なオーディオ素材に応じて異なる。例えば、ビットレートが 32 k b p s のオーディオ部分の符号化は正弦波エンコードによるのが最も効率的であろうが、同じビットレートでも、他のオーディオ部分の符号化は波形エンコードにより最も効率的であり得る。

【 0 0 4 5 】

上記の通り、制御部 C U はビットレートの分配 R_1 、 R_2 の所定のセットの全体をテストする。別の最適化プロセスでは、制御部 C U は、ビットレートの組み合わせ R_1 、 R_2 の結果の知覚的歪みの大きさが所定の基準値より小さくなったときに、さらに別のビットレート分配の組み合わせ R_1 、 R_2 をテストすることをやめる。

50

【 0 0 4 6 】

結果として、図 1 を参照して説明する実施形態は、関係する 2 つのオーディオエンコーダ A E 1 と A E 2 の能力を最大限利用することができる。各オーディオ部分に適応するからである。これにより、1) 符号化を要するオーディオのフレームに最良のオーディオエンコーダが自動的に選択され、2) 音声品質がよくなる場合、オーディオエンコーダを組み合わせて使用できる。

【 0 0 4 7 】

第 2 のエンコーダ A E 2 後に残る残差信号 2 は、ノイズエンコーダ (図示せず) の入力信号として使用できる。このように、2 つのエンコーダ A E 1、A E 2 によりモデル化できないスペクトル部分の少なくとも一部は、ノイズと置き換えることができ、通常は音声品質がよくなる。

10

【 0 0 4 8 】

第 1 の正弦波エンコーダ A E 1 の好ましい実施形態では、音響心理的マッチング追跡アルゴリズム (psycho-acoustical matching pursuit algorithm) (参考文献 5 参照) を使用して、正弦波を推定する。正弦波の分割と分配は、参考文献 6 に記載の方法に従って行うことが好ましい。

【 0 0 4 9 】

第 2 の変換エンコーダ A E 2 の好ましい実施形態は、参考文献 7 に記載されたフィルタバンクに基づく。第 2 のエンコーダ A E 2 の分割 (segmentation) は、第 1 のエンコーダ A E 1 の分割と同じか、一様分割 (uniform segmentation) であってもよい。

20

【 0 0 5 0 】

第 2 のエンコーダ A E 2 後の残差信号 2 は、好ましくは、トータルの知覚的歪みを測定するために、知覚モデル (参考文献 4 参照) により評価される。これは、元の信号 I N の各フレームのマスク関数 $v(f)$ を決定することにより為される事が好ましい。マスク関数 (masking function) は、問題のオーディオ信号が人間の聴覚系に入力されたときの人間のヒアリング閾値の周波数 f の関数としてスペクトル表示であると理解されている。時間領域の残差信号 2 を使用して、周波数 f の関数としてエラースペクトル $s(f)$ を求める。参考文献 4 の式 9 に示されているように、エラースペクトル信号とマスク関数の逆数との内積は、知覚された歪みのよい予測材料 (predictor) である。すなわち、知覚的歪み D は次式で計算できる：

30

【 0 0 5 1 】

【 数 1 】

$$D = \sum_f \frac{|s(f)|^2}{v(f)^2}$$

図 2 は、マスキング曲線 $v(f)$ の例を破線で示したグラフであり、上記の知覚モデルにより計算したものである。図 2 にはエラースペクトル $s(f)$ も実線で示している。これらは、上記のように、知覚的歪み尺度 D を求めるために使用する。グラフは、リニアな周波数スケール f に対してデシベル単位でレベル L_p を示している。図 2 は、低周波数 (例えば約 100 Hz) において、エラー信号 $s(f)$ がマスキング曲線 $v(f)$ と比較してレベルがかなり高く、この周波数範囲がトータルの知覚的歪み D に貢献していることを示している。10 - 12 kHz より上では、マスキング曲線は高くなるが、これは静かだと人間のヒアリング閾値が高くなることによるものである。

40

【 0 0 5 2 】

図 3 は、図 1 を参照して説明したような、正弦波エンコーダと波形エンコーダを有するオーディオエンコーダの場合に、正弦波エンコーダに割り当てられるビットレート部分へのトータルの知覚的歪み T P D の依存性を、異なるオーディオ信号に対して示す 2 つのグラフである。上のグラフのオーディオ信号はカスタネットのものであり、下のグラフのオ

50

オーディオ信号はハープシコードのものである。シンボルは異なるビットレートを示している：12 kbps（丸印）、24 kbps（プラス印）、48 kbps（星印）。太線は、様々なトータルビットレートについて選択されたビットレート分布を示す。

【0053】

上のグラフでカスタネットの場合について分かるように、少なくとも12 kbps（丸印）と24 kbps（プラス印）では、知覚的歪みはビットレート分布の関数としてほぼ一定である。しかし、48 kbps（星印）の場合、ほとんどのビットレートを波形エンコーダに送る方が、ほとんどのビットレートを正弦波エンコーダに送るよりも明らかに明らかに有利である。下の図のハープシコードの場合には、事情は異なる。ここで明らかにするのは、ビットレートが高くても、正弦波エンコーダがビットレートの約半分となることが明らかであり、ビットレートが低いと正弦波エンコーダにすべてのビットレートを使用する方がよいことが明らかである。

10

【0054】

図3に示した例は完全なオーディオ信号（complete audio excerpts）を評価し最適化することにより得られたが、この最適化方法はオーディオの短いセグメントに対して使用して、ビットレートR1、R2の分布を局所的な信号特性に対して適応できると考えられることに注意せよ。

【0055】

図4は、符号化されたオーディオ信号（例えば、図1を参照して説明したオーディオエンコーダにより符号化されたオーディオ信号）を復号するように構成されたオーディオデコーダのブロック図である。このオーディオデコーダは、第1と第2のエンコーダAE1、AE2のタイプにそれぞれ対応する第1と第2のデコーダAD1、AD2を有する。第1と第2のデコーダAD1、AD2は、エンコーダAE1、AE2からの第1と第2の符号化された信号部分E1、E2を受信するように構成されている。復号されたオーディオ信号は入力ビットストリームINとして受信され、第1と第2の復号された信号部分E1、E2がビットストリームデコーダBSDにより抽出される。そして、第1の復号された信号部分E1は第1のデコーダAD1に送られ、第2の復号された信号部分E2は第2のデコーダAD2に送られる。デコーダAD1、AD2は、それぞれのオーディオ部分を独立に復号でき、その結果の第1と第2の復号された信号部分D1、D2が単に加算されて、元のオーディオ信号の表示OUTを生成する。

20

30

【0056】

図5は、第1と第2の別々のエンコーダAE1、AE2のカスケードを有する他のオーディオエンコーダの実施形態のブロック図である。図1に示した第1の実施形態に関して説明した符号化方式（encoding scheme）が各所定の時間インターバルまたはセグメントで、一定のトータルビットレート（R1 + R2）の制約下で動作するが、この制約は図5の第2の実施形態では緩和される。この第2の実施形態は、原理的に、少なくとも第1のエンコーダAE1のすべての可能な符号化パラメータを考慮し、好ましくは第2のエンコーダAE2のすべての可能な符号化パラメータも考慮するものであり、その結果図1の第1のオーディオエンコーダと比較して知覚的歪みが低減される。しかし、第1のオーディオエンコーダの実施形態と比較して、第2のオーディオエンコーダの実施形態の実施はより複雑である。第1の実施形態と対照的に、第2の実施形態によりビットレートは各オーディオ信号（audio signal excerpt）の要求に適応可能となり、そのことにより2つのエンコーダAE1、AE2をより最適化できる。そのため、第2のオーディオエンコーダの実施形態は、多数のオーディオ信号の平均として考えられる同一のビットレートにおいて、知覚的歪みを低減することができ、すなわち音声品質を増大することができる。

40

【0057】

図5のオーディオエンコーダにおいて、第1と第2の異なるエンコーダAE1、AE2は、各々多数の異なる方法で受信入力信号0を符号化するように構成されている。これらの符号化オプションは符号化テンプレートと呼ばれる。例えば、正弦波エンコーダの場合、1つの符号化テンプレートは入力信号セグメントを表すために使用される制限はの1

50

つのセットを指定し、別のテンプレートは別の正弦波のセットを指定する。それゆえ、すべての可能なテンプレートのセットにより、エンコーダは可能なすべての符号化動作を実行でき、符号化を各オーディオ信号 (audio excerpt) に適応させられる。第 1 と第 2 のエンコーダ A E 1、A E 2 のテンプレートを、それぞれ第 1 と第 2 のテンプレート T 1、T 2 で示す。

【 0 0 5 8 】

2 つの符号化テンプレート T 1 と T 2 が選択されると、第 1 のエンコーダ A E 1 はオーディオ入力信号 0 を符号化して第 1 の符号化信号部分 E 1 を得る。符号化が不完全なので、符号化結果は残差信号 1 を含み、この残差信号 1 は第 2 のエンコーダ A E 2 により符号化され、第 2 の符号化信号部分 E 2 となる。第 2 の符号化処理の結果として残差信号 2 が得られる。この残差信号 2 は知覚的歪みの大きさを計算する知覚モデルを用いて制御部 C U により評価される。入力オーディオ信号 0 の最終的符号化について決定をするため、制御部 C U は、知覚的歪みの大きさが最小になる符号化テンプレート T 1 と T 2 を所定の符号化テンプレートから見つけるために、最適化手順を実行する。このため、知覚的歪みの大きさの他に、各エンコーダ A E 1、A E 2 のビットレート R 1、R 2 (またはその推測値) も考慮に入れる。

10

【 0 0 5 9 】

最終的な符号化テンプレート T 1 と T 2 が見つかり、これらのテンプレート T 1 と T 2 を用いて、第 1 と第 2 のエンコーダ A E 1、A E 2 からそれぞれ得られる第 1 と第 2 の符号化信号部分 E 1 と E 2 を生成する。これらの第 1 と第 2 の符号化信号部分 E 1、E 2 をビットストリームフォーマッタ (bit stream formatter) B S F に送る。このビットストリームフォーマッタ B S F はビットストリーム O U T を形成する。

20

【 0 0 6 0 】

第 1 のエンコーダ A E 1 は好ましくは正弦波エンコーダを有し、第 2 のエンコーダ A E 2 は変換エンコーダを有する。知覚的歪みの大きさ D は、第 1 のエンコーダの実施形態に関して説明したように、参考文献 4 に従って計算することが好ましい。

【 0 0 6 1 】

制御部 C U が解かねばならない最適化問題の正式な定義は次の通りである：

【 0 0 6 2 】

【 数 2 】

30

$$\arg \min_{T_1(n), T_2(n)} \sum_{n=1}^N D_2(T_1(n), T_2(n), n)$$

ここで、D 2 は 2 に基づき計算され、知覚モデル (例えば参考文献 4 を参照) により予測される知覚的歪みを表す。n はセグメント番号である。信号は、トータルの入力信号 0 から取ったいくつかの短い時間で符号化されると仮定している。この最小化問題は次の制約条件の下で最小化しなければならない。

【 0 0 6 3 】

【 数 3 】

40

$$c: \sum_{n=1}^N (R_1(T_1(n), n) + R_2(T_1(n), T_2(n), n)) \leq R_T$$

ここで、R T は目標ビットレートである。

【 0 0 6 4 】

ここで定式化された方法でこの問題を解くとき、原理的には、符号化テンプレート T 1、T 2 の全ての組み合わせをこの最小化問題を解くためにテストしなければならない。各セグメントについて第 1 と第 2 のエンコーダ A E 1、A E 2 にそれぞれ M 個の符号化テンプレートがあると仮定すると、テストしなければならない組み合わせの総数は次式の通り

50

である：

【 0 0 6 5 】

【 数 4 】

$$\# = M^{2N}$$

現実的な状況では、この問題は効果的に解くことができないので、ここでより効率的な解放を提示する。しかし、コアとなるアイデアはここに述べた問題かまたは少なくともその派生的問題を解くことである。制約条件付き最適化理論から、この種の問題は再定式化でき、セグメントごとに解く必要があるいくつかの独立な最適化問題に分割することができることが分かっている。これは、2つのエンコーダ A E 1、A E 2 のビットレート R 1、R 2 が独立であり、セグメントにわたって加えることができるのと制約条件下で行うことができる。同様に、セグメントをわたる知覚的歪みは加法的であり独立であることが必要である。

【 0 0 6 6 】

この問題に対する解は、全体的なビットレート制約条件に従って知覚的歪みの大きさにより予測される知覚的歪みを最小化する。暗にビットレートはセグメントごとに変わってもよい。また、知覚的歪みはセグメントにわたって一定ではない。しかし、セグメントにわたるこれらの変化を許すと、各セグメントについてビットレートまたは知覚的歪みが一定であるときより、全体的な知覚的歪みが小さくなる。

【 0 0 6 7 】

上記の制約条件の下で、N 個の独立なコスト関数を定義して、上記の制約条件の下で最小化することにより、問題を再定式化することができる。

【 0 0 6 8 】

【 数 5 】

$$J(T_1(n), T_2(n), n) = D_2(T_1(n), T_2(n), n) + \lambda [R_1(T_1(n), n) + R_2(T_1(n), T_2(n), n)] \quad (I)$$

解くべき問題は、以下の を見つけることである：

【 0 0 6 9 】

【 数 6 】

$$\lambda = \arg \sup_{\lambda} \left(\left[\sum_{n=1}^N J(T_{1,2 \min}(n), n) \right] - \lambda R_T \right) \quad (II)$$

$T_{1,2 \min}(n)$ は次式のように選択される：

【 0 0 7 0 】

【 数 7 】

$$T_{1,2 \min}(n) = \arg \min_{T_1(n), T_2(n)} J(T_1(n), T_2(n), n) \quad (III)$$

この問題の再定式化の利点は、N 個の独立の問題がラグランジュの未定定数 で結びつけられていることである。実際には、これは の初期値を選択することを意味する。この値を用いて式 (I I I) の最小化問題を各セグメント n について独立に解くことができる。この最適化の後、式 (I I) が満たされているかチェックできる。目標レート R T と使用するトータルビットレート間の差に基づいて を適応させる。このプロセスを、(式 I I に基づき) の最良値または満足のいく値が見つかるまで、繰り返す。

【 0 0 7 1 】

式 (I I I) の最適化問題を解くことは、考慮しているセグメント n について全ての符号化テンプレート T 1、T 2 の組み合わせをテストすることを意味する。具体的なエンコ

10

20

30

40

50

ーダ A E 1、A E 2 について、符号化テンプレート T 1、T 2 のあるサブセットの外にあるテンプレートは最適解にならないことがアприオリに分かっているとき、すべての可能な符号化テンプレート T 1、T 2 から符号化テンプレート T 1、T 2 のそのサブセットを選択することができる。式 (I I I) のジョイント最適化問題の場合、2 つのエンコーダ A E 1、A E 2 間の依存性により、最適化プロセスにおける考慮からアприオリに一定の符号化テンプレート T 1、T 2 を破棄することは、一層難しい。しかし、符号化テンプレート T 1 が知られていると仮定すると、最適化プロセスで考慮する必要のないテンプレート T 2 を選択することができる。テンプレート T 2 は最後のエンコーダ A E 2 に適用されるからである。より具体的には、第 2 のエンコーダ A E 2 に対して選択された符号化テンプレート T 2 は第 1 のエンコーダ A E 1 の符号化には影響しないからである。第 1 のエンコーダ A E 1 に対して、これは可能ではない。選択された T 1 は第 2 のエンコーダ A E 2 の振る舞いに影響するからである (式 I を参照、R 2 は T 1 と T 2 の両方に依存している)。それゆえ、エンコーダ A E 2 への影響を考えずにエンコーダ A E 1 の符号化テンプレート T 1 を破棄することはできない。エンコーダ A E 1 の符号化テンプレート T 1 のセットを制約することは、本来的により一層困難である。しかし、計算を複雑にしないため、例えば第 1 のエンコーダ A E 1 は孤立して動作すると仮定して、エンコーダ A E 1 の候補となるテンプレート T 1 の数を制限することもできる。

10

【 0 0 7 2 】

実際には、式 (I I I) の最適化問題を解くには、最初に符号化テンプレート T 1 を選択して、エンコーダ A E 2 に送られる残差信号 1 を計算する。T 1 は既知であるから、第 2 のエンコーダ A E 2 は式 (I I I) を単純化したものに従って最適化される：

20

【 0 0 7 3 】

【 数 8 】

$$\arg \min_{T_2(n)} J'(n) = D_2(T_1(n), T_2(n), n) + \lambda [R_2(T_1(n), T_2(n), n)] \quad (\text{IV})$$

上記の通り、可能な全ての符号化テンプレート T 2 を考慮しなくても、第 2 のエンコーダ A E 2 をほぼどのように選択しようが、この最適化問題をとくことができる。最小化問題を解くと、考えているセグメントに対して式 (I) の最適解が見つかるまで、第 1 のエンコーダ A E 1 の新しいテンプレート T 1 を選択することができる。

30

【 0 0 7 4 】

このセクションで説明した解法は以下のアルゴリズムでまとめることができる (A 1) : 各セグメントの最適符号化テンプレート T 1、T 2 とラグランジュの未定乗数 とを、目標ビットレートを満たすように決定する。

【 0 0 7 5 】

【数 9】

(A1):

Find λ :Loop n :Loop $T_1(n)$:Encode ε_0 with encoder AE1Loop $T_2(n)$:Encode ε_l with encoder AE2Derive $J'(n)$ (式IV参照)Remember best $T_2(n)$ and $J'(n)$ End Loop $T_2(n)$ Derive $J(n)$ (式I参照)Remember best $T_1(n)$, $T_2(n)$ and $J(n)$ End Loop $T_1(n)$ End Loop n Update λ End Find λ

10

20

(A 1) において、 T 1 のループにより式 (I I I) の最適解を見つける (例えばグローバルコスト関数を最小化する) 。この問題の一部として、式 (I V) で求めた第 2 のエンコーダ A E 2 についてコスト関数を最小化する T 2 のループがある。

【 0 0 7 6 】

ここで問題を定式化した方法において、同時にいくつかのセグメントの最適化が実行されることに留意せよ。このセグメントのセットでは、ビットレートはセグメントごとに変化することができる。実際的な状況では多くの場合、同時に評価できるセグメントのセットは限られている。この制約を処理するには 2 つのオプションがある：

30

1) セグメントの各セットに対して を決定し、そのセグメントのセット内のビットレートは常に要求される目標ビットレートを満たすようにする。

2) セグメントの各セットの後に を適応させて、過去の符号化動作におけるビットレートと目標ビットレート間のミスマッチを補償する。

【 0 0 7 7 】

ここでは、図 5 のエンコーダ A E 1 は正弦波エンコーダであり第 2 のエンコーダ A E 2 は変換エンコーダであると仮定している。第 1 のエンコーダ A E 1 について、すべての符号化テンプレート T 1 を考慮するわけではない。ある $1(n)$ のコスト関数を最小化する符号化テンプレート T 1 のみを考慮する：

40

【 0 0 7 8 】

【数 1 0】

$$J_1(n) = D_1(T_1(n), n) + \lambda_1(n) R_1(T_1(n), n) \quad (V)$$

ここで、D 1 は第 1 のエンコーダ A E 1 による符号化後に測定して知覚的歪みである。

【 0 0 7 9 】

2 つのエンコーダ A E 1、A E 2 は同じく分割 (segmentation) され、各エンコーダ A

50

E 1、A E 2 は符号化段階及び復号段階において重なり合うセグメントを使用する。このためにはアルゴリズム (A 1) を調整しなければならない。エンコーダ 2 が符号化セグメント n において必要な残差信号 $1(n)$ は符号化テンプレート $T 1(n - 1)$ 、 $T 1(n)$ 、 $T 1(n + 1)$ に依存するからである。

【0080】

この問題を明らかにするために、図 6 に符号化テンプレートを含む 2 つのエンコーダ A E 1、A E 2 のセグメント間の分割と重なりを、三角形のウィンドウで示した。図 6 から分かるように、第 1 のエンコーダの後の残差信号 $1(n)$ は、セグメント $n - 1$ 、 n 、 $n + 1$ において第 1 のエンコーダ A E 1 のために選択された符号化テンプレート $T 1$ に依存する。一般的に、符号化テンプレート $T 1(n + 1)$ はセグメント n を最適化するときには分からない。セグメントは 1 つずつ順番に最適化されるからである (アルゴリズム (A 1) 参照)。しかし、符号化テンプレート $T 1(n - 1)$ は、セグメント n で求められる解にも依存するから最適解ではないかも知れないが、セグメント n を最適化するときには分かっている。

10

【0081】

実際の解法は、先行するセグメント ($n - 1$) の最適化で求めた $T 1(n - 1)$ をとることである。次のセグメントでは、セグメント $n + 1$ に対してエンコーダ A E 1 が実行するであろう最終的な符号化について、詳細な情報を得た上での推測をする。このために、最近のセグメントの平均 1 を使用して、式 V に従って最適な符号化テンプレート $T 1(n + 1)$ を選択する。これに基づき、残差信号 $1(n)$ を計算でき、アルゴリズム (A 1) により 1 に従って最適な $T 2(n)$ を求めることができる。

20

【0082】

$1(n - 1)$ は、 $T 1(n)$ が終了したときにのみ分かり、その時に最終的な $T 2(n - 1)$ が求まる。

【0083】

説明を明確にするため、上で概要を説明した実際の解法を含む、アルゴリズム (A 1) をより詳細にしたもの (A 2) を以下に与える。(A 2) は、各セグメントの最適符号化テンプレート $T 1$ 、 $T 2$ とラグランジュの未定乗数 λ とを、目標ビットレートを満たすように決定する。セグメント間の重なりを考慮に入れる。

【0084】

30

【数 1 1】

(A2):

Find λ :Loop n :Loop $T_1(n)$:Encode $\varepsilon_0(n)$ with encoder AE1 and $T_1(n)$ Encode $\varepsilon_0(n+1)$ with encoder AE1 and informed
guess of $T_1(n+1)$ Retrieve $\varepsilon_1(n)$ based on $\varepsilon_0(n-1)$, $\varepsilon_0(n)$, $\varepsilon_0(n+1)$ with $T_1(n-1)$, $T_1(n)$, $T_1(n+1)$ Loop $T_2(n)$:Encode $\varepsilon_1(n)$ with encoder AE2Derive $J'(n)$ (式IV参照)Remember best $T_2(n)$ and $J'(n)$ End Loop $T_2(n)$ Derive $J(n)$ (式I参照)Remember best $T_1(n)$, $T_2(n)$ and $J(n)$ End Loop $T_1(n)$ # $T_1(n)$ の最終解が求まった# 最終の $\varepsilon_1(n-1)$ が求まり最終の $T_2(n-1)$ を求めることができるLoop $T_2(n-1)$:Encode $\varepsilon_1(n-1)$ with encoder AE2Derive $J'(n-1)$ (式IV参照)Remember best $T_2(n-1)$ and $J'(n-1)$ End Loop $T_2(n-1)$ End Loop n Update λ End Find λ

図1に示したエンコーダの実施形態に関して解くべき最適化問題をここで説明する。この実施形態では、図5の実施形態について説明したウィンドウが重なる問題を、後続のN個のセグメントに渡って1を一定にして解消し、対応する符号化テンプレート $T_1(1)$... $T_1(N)$ を各々が式(V)を最小化するエンコーダAE1に適用する。この場合、第1のエンコーダAE1のN個のセグメントのすべてを最初に求める。第2のエンコーダAE2について、に従って、式(IV)を最小化する符号化テンプレート $T_2(1)$... $T_2(N-1)$ を求める。このように、1のいくつかの値を式(1)を最小化するものが見つかるまでテストすることができる。目標ビットレートが可能な最小の知覚的歪みを満たすまで、の値をテストできる。セグメント1...N-1の解を求めた後、次のセグメントN...2N-1を最適化する。以下において、アルゴリズム(A3)

は、1を一定に保ってセグメント間の重なりを考慮にいて、目標ビットレートを満たすように、各セグメントの最適な符号化テンプレート T_1 と T_2 とラグランジュの未定乗数 λ を求める原理をまとめたものである。

【0085】

【数12】

(A3):

Find λ :

Loop λ_l :

Loop n_1 (1...N):

Encode $\varepsilon_0(n_1)$ with AE1 and use λ_l to min. $J_l(n_1)$ see

Eq. (V)

End Loop n_1

Loop n_2 (1...N-1):

Encode $\varepsilon_l(n_2)$ with AE2 and use λ to min. $J'(n_2)$ see Eq.

(IV)

End Loop n_2

Add all cost functions $J(n_2)$

Remember best λ_l and corresp. best templates for both encoders AE1,

AE2

End Loop λ_l

Remember best λ and corresponding best templates for both encoders AE1,

AE2

End Find λ

ネストになっているループの数は、アルゴリズム(A2)よりも(A3)の方が1つ少ないことに留意せよ。しかし、これは正しくない。1とに從う符号化には対応する符号化テンプレートを求める別のループが必要だからである。

【0086】

アルゴリズム(A3)の利点は、2つのエンコーダAE1、AE2の分割(segmentation)を一致させる必要がないことである。エンコーダAE1が符号化する時間的インターバル(例えばセグメント番号 $n = 1 \dots N$ を含む)はAE2が符号化する時間的インターバルと常に長さが少なくとも同じであることだけが要求される。

【0087】

アルゴリズム(A3)を実装してテストしたが、 n_2 のループが $N - 1$ ではなくて N までである点だけが異なる。これにより N 個のセグメントの終わりには符号化の正確性が若干落ちたが、品質には影響が無かったようである。実施において、第1のエンコーダAE1は異なる柔軟な分割を用いた(参考文献6参照)が、第2のエンコーダAE2は固定分割を用いた。

【0088】

ここまで説明したエンコーダの実施形態では、2つのカスケードされたエンコーダを用いた。しかし、本発明によれば、カスケードするエンコーダの数は容易に2より大きくできる。2つのシナリオがある：

全ての符号化テンプレートを考慮する（例えば、候補のテンプレートは制限しない）。この場合、第1のエンコーダを2つ（またはそれ以上）のエンコーダのカスケードで置き換えることができる。これらのエンコーダの各々の符号化テンプレートは、各セグメントについて、符号化テンプレートの可能なすべての組み合わせを含む大きな符号化テンプレートのセットにまとめられる。そうすると、この問題は2つのエンコーダのみがカスケードされているものとして解くことができる。

【0089】

すべての符号化テンプレートは考慮せず、式(V)で与えられるコスト関数を最小化するもののみを考慮する。この場合、第2のエンコーダは に従って最適化される2つのエンコーダのカスケードとして考えることができる。この「ネストされた」拡張は、カスケードされるエンコーダの数が増えても続けることができる。

10

【0090】

図7は、並行して動作する2つのエンコーダAE1、AE2を有する第3のオーディオエンコーダの実施形態を示す図である。これは図5に示した第2のエンコーダの実施形態と、オーディオ入力信号 0が分離部SPLITにより第1の信号部 1と第2の信号部 2に分離される点で異なる。第1と第2の信号部分 1、 2は加え合わせると入力信号 0となる。2つの信号 1と 2は、第1のエンコーダAE1と第2のエンコーダAE2にそれぞれ送られる。

【0091】

図7の第3のオーディオエンコーダの実施形態の制御部CUは、第1と第2のエンコーダにそれぞれ符号化テンプレートT1、T2を送り、符号化を実行させる。このように、選択されるすべての2つの符号化テンプレートT1とT2について、エンコーダAE1は第1の信号部分 01を処理し、これとは独立に、エンコーダAE2は第2の信号部分 02を処理する。エンコーダAE1、AE2はそれぞれ残差信号 3と 4を発生する。これらの残差信号は制御部に送られる。制御部は知覚モデルに従って、知覚的歪みの大きさを計算する。この知覚的歪みの大きさを用いて、可能な符号化テンプレートT1、T2のセットから最適な符号化テンプレートT1、T2を求め、信号の最終的な符号化を決定する。このため、知覚的歪みの大きさのだけでなく、各エンコーダAE1、AE2のビットレートR1、R2（またはその推測値）も考慮に入れる。第1と第2のオーディオエンコーダの実施形態について説明したように、参考文献4のモデルを使用して、知覚的歪みの大きさDを計算できる。

20

30

【0092】

第3のオーディオエンコーダの実施形態において、制御部が解かねばならない問題の正式な定義は次の通りである：

【0093】

【数13】

$$\arg \min_{T_1(n)T_2(n)} \sum_{n=1}^N D_1(T_1(n), n) + D_2(T_2(n), n)$$

40

ここで、D1とD2は、それぞれ 3と 4に基づき計算される。知覚的歪みは単に加えることができると仮定する。パラメータnはセグメント番号であり、信号はトータルの入力信号からいくつかの短い時間セグメントにより符号化されると仮定する。この最小化問題は次の制約条件の下で最小化しなければならない。

【0094】

【数 1 4】

$$c: \sum_{n=1}^N R_1(T_1(n), n) + R_2(T_2(n), n) \leq R_T$$

ここで、 R_T は目標ビットレートである。

【0 0 9 5】

前のセクションで説明した制約条件の下で、 $2N$ 個の独立なコスト関数を定義して、最小化することにより、問題を再定式化することができる。

10

【0 0 9 6】

【数 1 5】

$$J_1(T_1(n), n) = D_1(T_1(n), n) + \lambda R_1(T_1(n), n) \quad (\text{VI})$$

$$J_2(T_2(n), n) = D_2(T_2(n), n) + \lambda R_2(T_2(n), n) \quad (\text{VII})$$

解くべき問題は、以下の をを見つけることである：

20

【0 0 9 7】

【数 1 6】

$$\lambda = \arg \sup_{\lambda} \left(\sum_{n=1}^N J_1(T_{1\min}(n), n) + J_2(T_{2\min}(n), n) \right) - \lambda R_T \quad (\text{VIII})$$

$T_{1\min}(n)$ と $T_{2\min}(n)$ は次式をみたすように選択される：

【0 0 9 8】

【数 1 7】

$$T_{1\min}(n) = \arg \min_{T_{1\min}(n)} J_1(T_1(n), n) \quad (\text{IX})$$

$$T_{2\min}(n) = \arg \min_{T_{2\min}(n)} J_2(T_2(n), n) \quad (\text{X})$$

30

この問題の再定式化の利点は、 $2N$ 個の独立の問題がラグランジュの未定定数 で結びつけられていることである。実際には、これは の初期値を選択することを意味する。この値を用いて式 (IX) と (X) の最小化問題を各セグメント n と各エンコードについて独立に解くことができる。この最適化の後、式 (VIII) が満たされているかチェックする。目標レート R_T と使用するトータルビットレート ($R_1 + R_2$) 間の差に基づいて、を適応させる。このプロセスを、(式 VIII に基づき) の最良値または満足のいく値が見つかるまで繰り返す。

40

【0 0 9 9】

この並行する場合の最適化は個々のエンコード AE_1 、 AE_2 に対して別々であり独立にできるので、原理的には、可能な全ての符号化テンプレート T_1 、 T_2 から符号化テンプレート T_1 、 T_2 のサブセットを選択することが可能である。エンコード AE_1 、 AE_2 の特性により、そのサブセットに入らないテンプレート T_1 、 T_2 は最適解ではないことが分かっているからである。これは、カスケードされたエンコードと比較して並列エンコードの大きな利点である。

50

【 0 1 0 0 】

上記の並行最適化は、3個以上のエンコーダに容易に拡張できるが、これは式(V I)乃至(X)の性質から理解できるであろう。

【 0 1 0 1 】

図4の並行エンコーダの好ましい実施形態において、入力信号スプリッタ S P L I T は、オーディオ入力信号 0 の入力セグメントを変換係数に分離(split)するように構成された修正離散余弦変換(M D C T)を有する。変換係数は、別々に符号化される各々がスケールファクタ帯域を表すグループに分離される。各セグメントの各スケールファクタ帯域に対して、 の同じ値に従って式(V I)と(V I I)に与えられたコスト関数を最小化するように、1つのスケールファクタと1つの符号化ブック(coding book)を選択しなければならぬ。スケールファクタ帯域に応じて異なる符号ブック(code book)デザインを用いて、そのスケールファクタ帯域中の変換係数の異なる統計を任意的に利用してもよい。すべての個々のスケールファクタ帯域をセグメントにわたって最適化してから、トータルのビットレートを計算し、目標ビットレートになるように を適応させる。

10

【 0 1 0 2 】

本発明によるエンコーダとデコーダは、デジタルシグナルプロセッサを有する単一チップ上で実施してもよい。このチップは、オーディオ装置のシグナルプロセッサの能力にかかわらず、かかるオーディオ装置に組み込むことができる。あるいは、エンコーダとデコーダは、応用装置のメインのシグナルプロセッサ上で実行される純粋なアルゴリズムにより実施することもできる。

20

【 0 1 0 3 】

参照符号が請求項に含まれているが、単に請求項を明確にするためである。これらの図面の実施形態の例への参照は請求項の範囲を限定するものであると解してはならない。

参考文献リスト

[1] Scott N. Levine著、「Audio Representations for Data Compression and Compressed Domain Processing」、博士論文、1998年12月2日

[2] Wuppermann等著「Transmission system implementing different coding principles」米国特許第5,808,569号公報

[4] S. van de Par, A. Kohlrausch, G. Charestan, R. Heusdens (2002) 著「A new psychoacoustical masking model for audio coding applications」IEEE Int. Conf. Acoust, Speech and Signal Process., Orlando, USA、2002年ページ11-1805-1808

30

[5] R. Heusdens, R. Vafm, W.B. Kleijn (2002) 等著「Sinusoidal modeling using psychoacoustical matching pursuits」IEEE Signal Processing Lett, 9(8), pp. 262-265

[6] R. Heusdens and S. van de Par (2002) 等著「Rate-distortion optimal sinusoidal modeling of audio and Speech using psychoacoustical matching pursuits」IEEE Int. Conf. Acoust., Speech and Signal Process., Orlando, USA, 2002年ページ11-1809-1812

[7] J. Princen and A. Bradley (1986) 著「Analysis/synthesis filter bank design based on time domain aliasing cancellation」IEEE Trans. Acoust, Speech, Signal Processing, 34 pp. 1153-1161

40

【図面の簡単な説明】

【 0 1 0 4 】

【図1】オーディオの各部分についてトータル目標ビットレートの制約の下で動作する2つのエンコーダを有する第1のオーディオエンコーダの実施形態を示すブロック図である。

【図2】知覚的歪みの大きさを求めるために使用されるマスキング曲線とエラースペクトルの例を示すグラフである。

【図3】2つの異なる音声例の場合に、第1と第2のエンコーダの間のビットレートの分

50

配の、結果として得られるトータルな知覚的歪みに対する影響を示すグラフである。

【図4】2つのデコーダを有するオーディオデコーダのブロック図である。

【図5】各オーディオ部分に対して、いくつかの符号化テンプレートで動作する2つのエンコーダのカスケードを有する第2のエンコーダの実施形態を示す図である。

【図6】第2のエンコーダの実施形態による2つのエンコーダ間の分割と重なりを示す図である。

【図7】並行して動作する2つのエンコーダを有する第3のエンコーダの実施形態を示す図である。

【図1】

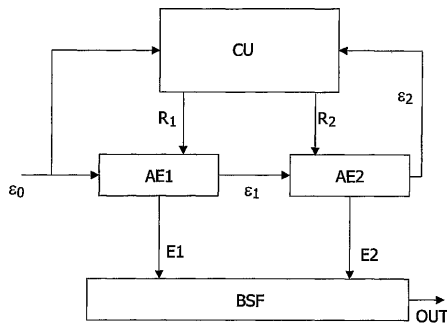


FIG.1

【図2】

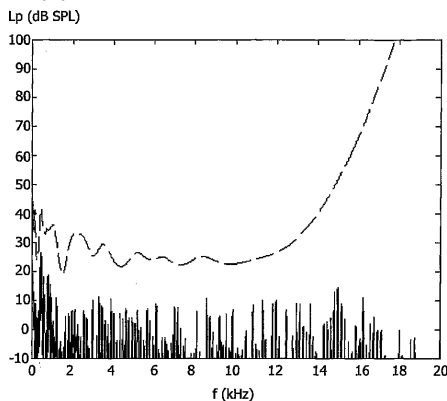


FIG.2

【図3】

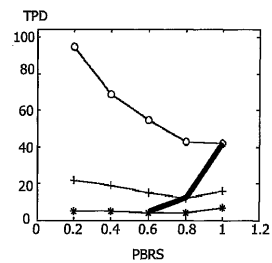
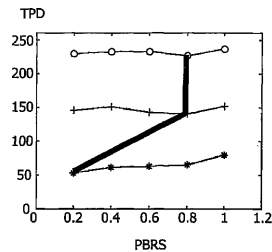


FIG.3

【図4】

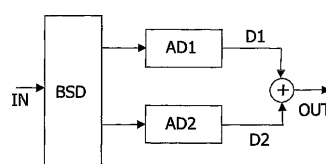


FIG.4

【図 5】

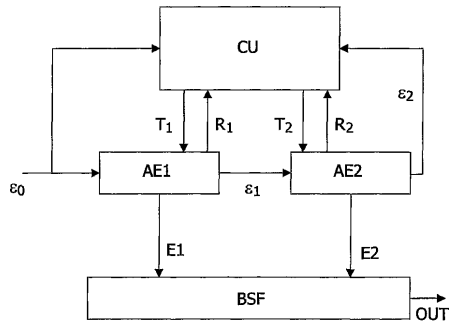


FIG.5

【図 6】

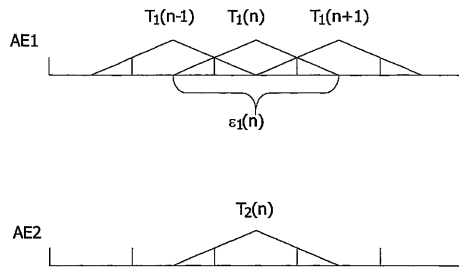


FIG.6

【図 7】

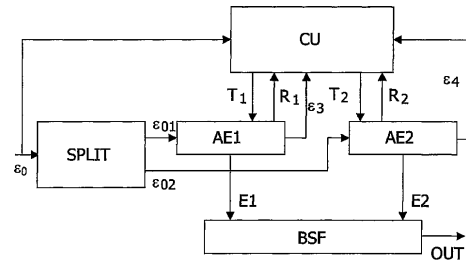


FIG.7

フロントページの続き

- (72)発明者 ファン デ パール, ステーフエン エル イェー デー エー
オランダ国, 5 6 5 6 アーアー アインドーフエン, プロフ・ホルストラーン 6
- (72)発明者 ファン スヘインデル, ニコル, ハー
オランダ国, 5 6 5 6 アーアー アインドーフエン, プロフ・ホルストラーン 6
- (72)発明者 コット, ヴァレリー, エス
オランダ国, 5 6 5 6 アーアー アインドーフエン, プロフ・ホルストラーン 6
- (72)発明者 ヒュースデンス, リハルト
オランダ国, 5 6 5 6 アーアー アインドーフエン, プロフ・ホルストラーン 6

審査官 田部井 和彦

- (56)参考文献 特開平04 - 170113 (JP, A)
特開平03 - 035300 (JP, A)
特開平03 - 108824 (JP, A)
特表2007 - 512572 (JP, A)
特開2004 - 004530 (JP, A)
欧州特許出願公開第00396121 (EP, A1)
米国特許第05808569 (US, A)
米国特許第06263312 (US, B1)

(58)調査した分野(Int.Cl., DB名)

G10L 11/00-21/06

H03M 7/30