

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2020-170190

(P2020-170190A)

(43) 公開日 令和2年10月15日(2020.10.15)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 0 L 25/21 (2013.01)	G 1 0 L 25/21	
G 1 0 L 25/18 (2013.01)	G 1 0 L 25/18	
G 1 0 L 19/00 (2013.01)	G 1 0 L 19/00	4 0 0 A
G 1 0 L 21/0232 (2013.01)	G 1 0 L 21/0232	

審査請求 有 請求項の数 12 O L (全 18 頁)

(21) 出願番号	特願2020-113803 (P2020-113803)	(71) 出願人	500341779 フラウンホーファー・ゲゼルシャフト・ツール・フェルデルング・デル・アンゲヴァンテン・フォルシュング・アインゲトラーゲネル・フェライン ドイツ連邦共和国, 80686 ミュンヘン, ハンザシュトラッセ 27ツェー
(22) 出願日	令和2年7月1日(2020.7.1)	(74) 代理人	100205981 弁理士 野口 大輔
(62) 分割の表示	特願2018-174338 (P2018-174338) の分割	(72) 発明者	ベンジャミン・シューベルト ドイツ連邦共和国 90429 ニュルンベルグ ツィックシュトラッセ 6
原出願日	平成27年7月21日(2015.7.21)	(72) 発明者	マヌエル・ヤンダー ドイツ連邦共和国 91334 ヘムホルフェン ブルメンシュトラッセ 1
(31) 優先権主張番号	14178779.6		最終頁に続く
(32) 優先日	平成26年7月28日(2014.7.28)		
(33) 優先権主張国・地域又は機関	欧州特許庁 (EP)		

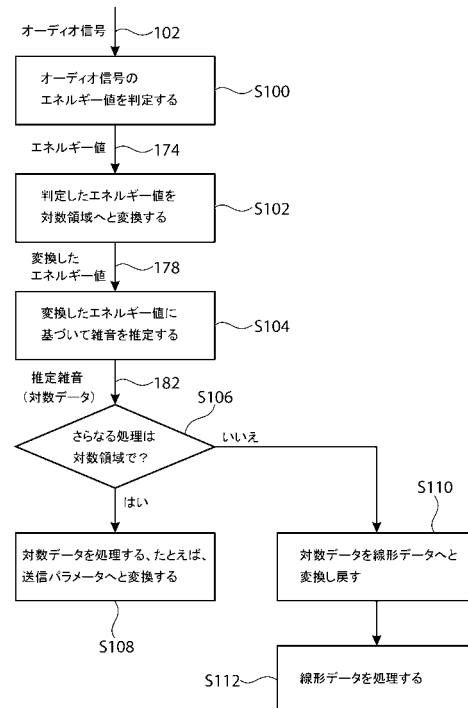
(54) 【発明の名称】 オーディオ信号内の雑音を推定するための方法、雑音推定器、オーディオ符号化器、オーディオ復号器、およびオーディオ信号を送信するためのシステム

(57) 【要約】 (修正有)

【課題】 オーディオ信号内の雑音を効率的に推定する方法を提供する。

【解決手段】 方法は、オーディオ信号102のエネルギー値174を推定しS100、対数領域へと変換しS102、変換したエネルギー値178に基づいて、オーディオ信号102の雑音レベルを推定するS104。

【選択図】 図3



【特許請求の範囲】

【請求項 1】

オーディオ信号 (102) 内の雑音を推定するための方法であって、
前記オーディオ信号 (102) のエネルギー値 (174) を判定すること (S100)
と、

前記エネルギー値 (174) を \log_2 領域へと変換すること (S102) と、
前記 \log_2 領域において直接的に、前記変換したエネルギー値 (178) に基づいて
前記オーディオ信号 (102) の雑音レベル (182) を推定すること (S104) と、
を含む方法。

【請求項 2】

前記雑音レベルを推定すること (S104) は、最小値統計アルゴリズムのような、所
定の雑音推定アルゴリズムを実施することを含む、請求項 1 に記載の方法。

【請求項 3】

前記エネルギー値 (174) を判定すること (S100) は、前記オーディオ信号 (1
02) を周波数領域へと変換することによって前記オーディオ信号 (102) のパワース
ペクトルを得ることと、前記パワースペクトルを心理音響的に動機付けられた帯域にグル
ープ化することと、各帯域のエネルギー値 (174) を形成するためにパワースペクトル
ピンを帯域内に累積することとを含み、各帯域の前記エネルギー値 (174) を \log_2
領域へと変換し、対応する前記変換したエネルギー値 (174) に基づいて、各帯域の雑
音レベルを推定する、請求項 1 または 2 に記載の方法。

【請求項 4】

前記オーディオ信号 (102) は複数のフレームを含み、各フレームについて、前記エ
ネルギー値 (174) を判定して前記 \log_2 領域へと変換し、前記変換したエネルギー
値 (174) に基づいてフレームの各帯域の前記雑音レベルを推定する、請求項 1 ~ 3 の
いずれか一項に記載の方法。

【請求項 5】

前記エネルギー値 (174) を、以下の式に従って前記 \log_2 領域へと変換し (S1
02)、

$$E_{n_log} = \frac{\lfloor (\log_2(1+E_{n_lin})) \cdot 2^N \rfloor}{2^N}$$

$$\lfloor x \rfloor$$

は $\text{floor}(x)$ であり、 E_{n_log} は前記 \log_2 領域における帯域 n のエネルギー
値であり、 E_{n_lin} は線形領域における帯域 n のエネルギー値であり、 N は量子化
分解能である、請求項 1 ~ 4 のいずれか一項に記載の方法。

【請求項 6】

前記変換したエネルギー値 (178) に基づいて前記雑音レベルを推定すること (S1
04) は、対数データをもたらし、前記方法は、

さらなる処理のために前記対数データを直接的に使用すること (S108)、または
さらなる処理のために前記対数データを線形領域へと変換し戻すこと (S110、S1
12)

をさらに含む、請求項 1 ~ 5 のいずれか一項に記載の方法。

【請求項 7】

前記 \log_2 領域において送信が行われる場合に、前記対数データを送信データへと直
接的に変換し (S108)、

前記対数データを送信データへと直接的に変換すること (S110) は、ルックアップ
テーブルまたは近似とともに、シフト関数、たとえば、

10

20

30

40

$$E_{n_lin} = 2^{(E_{n_log}-1)}$$

を使用する、請求項 6 に記載の方法。

【請求項 8】

コンピュータ上で実行されると、請求項 1 ~ 7 のいずれか一項に記載の方法を実行する命令を記憶しているコンピュータ可読媒体を備える、非一時的コンピュータプログラム製品。

【請求項 9】

雑音推定器 (170) であって、

10

オーディオ信号 (102) のエネルギー値 (174) を判定するように構成されている検出器 (172) と、

前記エネルギー値 (174) を $\log 2$ 領域へと変換するように構成されている変換器 (176) と、

前記 $\log 2$ 領域において直接的に、前記変換したエネルギー値 (178) に基づいて前記オーディオ信号 (102) の雑音レベル (182) を推定するように構成されている推定器 (180) プロセッサと、を備える、雑音推定器 (170)。

【請求項 10】

請求項 9 に記載の雑音推定器を備える、オーディオ符号化器 (100)。

【請求項 11】

20

請求項 9 に記載の雑音推定器 (170) を備える、オーディオ復号器 (150)。

【請求項 12】

オーディオ信号 (102) を送信するためのシステムであって、

受信されたオーディオ信号 (102) に基づいてコード化されたオーディオ信号 (102) を生成するように構成されているオーディオ符号化器 (100) と、

前記コード化されたオーディオ信号 (102) を受信し、前記コード化されたオーディオ信号 (102) を復号し、復号されたオーディオ信号 (102) を出力するように構成されているオーディオ復号器 (150) と、を備え、

前記オーディオ符号化器および前記オーディオ復号器のうちの少なくとも一方は、請求項 9 に記載の雑音推定器 (170) を備える、システム。

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、オーディオ信号の処理の分野に関し、より詳細には、オーディオ信号、たとえば、符号化されるオーディオ信号、または、復号されたオーディオ信号内の雑音を推定する手法に関する。実施形態は、オーディオ信号内の雑音を推定する方法、雑音推定器、オーディオ符号化器、オーディオ復号器、およびオーディオ信号を送信するためのシステムを説明する。

40

【背景技術】

【0002】

オーディオ信号の処理の分野、たとえば、オーディオ信号の符号化または復号されたオーディオ信号の処理において、雑音を推定することが所望される状況がある。たとえば、参照により本明細書に組み込まれる国際出願 EP 2013/077525 号明細書および国際出願 EP 2013/077527 号明細書には、周波数領域において背景雑音のスペクトルを推定するために、雑音推定器、たとえば、最小値統計雑音推定器を使用することが記載されている。このアルゴリズムへと供給される信号は、たとえば、高速フーリエ変換 (FFT) または任意の他の適切なフィルタバンクによって、ブロックごとに周波数領域へと変換されている。この枠組みは通常、コーデックの枠組みと同一である。すなわち

50

、コーデック内にすでに存在する変換を再使用することができ、たとえば、EVS（拡張音声サービス）符号化器において、前処理のためにFFTが使用される。雑音推定を目的として、FFTのパワースペクトルが計算される。スペクトルは、心理音響的に動機付けられた帯域にグループ化され、帯域内のパワースペクトルピンは、帯域ごとのエネルギー値を形成するように蓄積される。最終的に、オーディオ信号の心理音響的処理に使用されることも多いこの手法によって、エネルギー値のセットが獲得される。各帯域は、それ自体の雑音推定アルゴリズムを有する。すなわち、各フレームにおいて、経時的な信号を分析し、任意の所与のフレームにおける各帯域の推定雑音レベルを与える雑音推定アルゴリズムを使用して、そのフレームのエネルギー値が処理される。

【0003】

高品質発話およびオーディオ信号に使用されるサンプル分解能は16ビットであり得、すなわち、信号は、96dBの信号対雑音比(SNR)を有する。パワースペクトルを計算するということは、信号を周波数領域へと変換し、各周波数ピンの2乗を計算することを意味する。2乗関数に起因して、これは32ビットのダイナミックレンジを必要とする。複数のパワースペクトルピンをまとめて帯域にするには、帯域内のエネルギー分布が実際には分からないため、ダイナミックレンジのためにさらなるヘッドルームが必要である。結果として、プロセッサ上で雑音推定器を作動させるためには、32ビットを超える、一般的には約40ビットのダイナミックレンジがサポートされる必要がある。

【0004】

バッテリーのようなエネルギー貯蔵ユニットから受け取られるエネルギーに基づいて動作する、オーディオ信号を処理するデバイス、たとえば、携帯電話のような携帯機器においては、エネルギーを維持するために、オーディオ信号の電力効率のよい処理が、バッテリー寿命のために必須である。既知の手法によれば、オーディオ信号の処理は、一般的に、16または32ビット固定小数点フォーマットのデータの処理をサポートする固定小数点プロセッサによって実施される。16ビットデータを処理することによって処理の最低の複雑度が達成され、一方、32ビットデータの処理は、すでにいくらかのオーバーヘッドを必要とする。40ビットのダイナミックレンジによるデータの処理は、データを2つ、すなわち、仮数および指数に分割することを必要とし、これらの両方が、データを修正するときに対処されなければならない、その結果として、計算がさらにより複雑になり、ストレージ要求がさらにより高くなる。

【先行技術文献】

【特許文献】

【0005】

【特許文献1】国際出願EP2013/077525号明細書

【特許文献2】国際出願EP2013/077527号明細書

【非特許文献】

【0006】

【非特許文献1】R. Martin「Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics」(2001)

【非特許文献2】T. GerkmannおよびR. C. Hendriks「Unbiased MMSE-based noise power estimation with low complexity and low tracking delay」(2012)

【非特許文献3】L. Lin、W. Holmes、およびE. Ambikairajah「Adaptive noise estimation algorithm for speech enhancement」(2003)

【発明の概要】

【発明が解決しようとする課題】

【0007】

10

20

30

40

50

上述した従来技術から開始して、本発明の目的は、不要な計算オーバーヘッドを回避するために固定小数点プロセッサを使用してオーディオ信号内の雑音を効率的に推定するための手法を提供することである。

【課題を解決するための手段】

【0008】

この目的は、独立請求項において定義されているものとしての主題によって達成される。

【0009】

本発明は、オーディオ信号内の雑音を推定するための方法であって、オーディオ信号のエネルギー値を判定することと、エネルギー値を対数領域へと変換することと、変換したエネルギー値に基づいてオーディオ信号の雑音レベルを推定することを含む、方法を提供する。

10

【0010】

本発明は、雑音推定器であって、オーディオ信号のエネルギー値を判定するように構成されている検出器と、エネルギー値を対数領域へと変換するように構成されている変換器と、変換したエネルギー値に基づいてオーディオ信号の雑音レベルを推定するように構成されている推定器とを備える、雑音推定器を提供する。

【0011】

本発明は、本発明の方法に従って動作するように構成されている雑音推定器を提供する。

20

【0012】

実施形態によれば、対数領域は、 \log_2 領域を含む。

【0013】

実施形態によれば、雑音レベルを推定することは、対数領域において直接的に、変換されたエネルギー値に基づいて所定の雑音推定アルゴリズムを実施することを含む。雑音推定は、R. Martin「Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics」(2001)によって記載されている最小値統計アルゴリズムに基づいて実行することができる。他の実施形態において、T. GerkmannおよびR. C. Hendriks「Unbiased MMSE-based noise power estimation with low complexity and low tracking delay」(2012)によって記載されているMMSEベースの雑音推定器、または、L. Lin、W. Holmes、およびE. Ambikairajah「Adaptive noise estimation algorithm for speech enhancement」(2003)によって記載されているアルゴリズムのような、代替的な雑音推定アルゴリズムが使用されてもよい。

30

【0014】

実施形態によれば、エネルギー値を判定することは、オーディオ信号を周波数領域へと変換することによってオーディオ信号のパワースペクトルを得ることと、パワースペクトルを心理音響的に動機付けられた帯域にグループ化することと、各帯域のエネルギー値を形成するためにパワースペクトルピンを帯域内に累積することとを含み、各帯域のエネルギー値は対数領域へと変換され、対応する変換されたエネルギー値に基づいて、各帯域の雑音レベルは推定される。

40

【0015】

実施形態によれば、オーディオ信号は複数のフレームを含み、各フレームについて、エネルギー値が判定されて対数領域へと変換され、変換されたエネルギー値に基づいて各帯域の雑音レベルは推定される。

【0016】

実施形態によれば、エネルギー値は以下のように対数領域へと変換される。

50

$$E_{n_log} = \frac{\lfloor (\log_2(1+E_{n_lin})) \cdot 2^N \rfloor}{2^N}$$

$\lfloor x \rfloor$

は $\text{floor}(x)$ であり、 E_{n_log} は \log_2 領域における帯域 n のエネルギー値であり、 E_{n_lin} は線形領域における帯域 n のエネルギー値であり、 N は分解能 / 精度である。

【0017】

実施形態によれば、変換されたエネルギー値に基づいて雑音レベルを推定することは、対数データをもたらし、方法は、さらなる処理のために対数データを直接使用すること、または、さらなる処理のために対数データを線形領域に変換し戻すことをさらに含む。

10

【0018】

実施形態によれば、対数データは、送信が対数領域で行われる場合には送信データに直接変換され、対数データを送信データへと直接的に変換するには、ルックアップテーブルまたは近似とともにシフト関数、たとえば、

$$E_{n_lin} = 2^{(E_{n_log}-1)}$$

を使用する。

20

【0019】

本発明は、コンピュータ上で実行されると、本発明の方法を実行する命令を記憶しているコンピュータ可読媒体を備える非一時的コンピュータプログラム製品を提供する。

【0020】

本発明は、本発明の雑音推定器を備えるオーディオ符号化器を提供する。

【0021】

本発明は、本発明の雑音推定器を備えるオーディオ復号器を提供する。

【0022】

本発明は、オーディオ信号を送信するためのシステムであって、受信オーディオ信号に基づいてコード化オーディオ信号を生成するように構成されているオーディオ符号化器と、コード化オーディオ信号を受信し、コード化オーディオ信号を復号し、復号オーディオ信号を出力するように構成されているオーディオ復号器とを備え、オーディオ符号化器およびオーディオ復号器のうちの少なくとも一方は、本発明の雑音推定器を備える、システムを提供する。

30

【0023】

本発明は、雑音推定アルゴリズムが線形エネルギーデータに対して作動する従来の手法とは対照的に、オーディオ / 発話材料内の雑音レベルを推定することを目的として、対数入力データに基づいてもアルゴリズムを作動させることが可能であるという本発明者らの知見に基づく。雑音推定に対して、データ精度に対する要求はそれほど高くなく、たとえば、両方とも参照により本明細書に組み込まれる国際出願 EP 2013 / 077525 号明細書または国際出願 EP 2013 / 077527 号明細書に記載されているような快適雑音生成のための推定値を使用するとき、帯域ごとのほぼ正確な雑音レベルを推定すれば十分であり、すなわち、雑音レベルが、たとえば、0.1 dB だけより高いと推定されるか否かは、最終的な信号において注目されるものではないことが分かっている。したがって、データのダイナミックレンジをカバーするためには 40 ビットが必要とされ得るが、従来の手法において、中 / 高レベル信号のためのデータ精度は、実際に必要であるよりもはるかに高い。これらの知見に基づいて、実施形態によれば、本発明の重要な要素は、帯域ごとのエネルギー値を対数領域、好ましくは \log_2 領域へと変換し、たとえば、最小値統計アルゴリズムまたは任意の他の適切なアルゴリズムに基づいて、対数領域において直接的に雑音推定を実行することであり、それによって、たとえば、16 ビットにおいて

40

50

エネルギー値を表現するがを可能になり、その結果として、たとえば、固定小数点プロセッサを使用して、より効率的な処理が可能になる。

【0024】

以下において、本発明の実施形態を、添付の図面を参照しながら説明する。

【図面の簡単な説明】

【0025】

【図1】符号化されるべきオーディオ信号または復号オーディオ信号内の雑音を推定するための本発明の手法を実施する、オーディオ信号を送信するためのシステムの単純化したブロック図である。

【図2】オーディオ信号符号化器および/またはオーディオ信号復号器において使用することができる一実施形態による雑音推定器の単純化したブロック図である。

【図3】一実施形態によるオーディオ信号内の雑音を推定するための本発明の手法を示す流れ図である。

【発明を実施するための形態】

【0026】

以下において、本発明の手法の実施形態をさらに詳細に説明する。添付の図面において、同一または類似の機能を有する要素は、同じ参照符号によって示されることに留意されたい。

【0027】

図1は、符号化器側および/または復号器側において本発明の手法を実施する、オーディオ信号を送信するためのシステムの単純化したブロック図を示す。図1のシステムは、入力102においてオーディオ信号104を受信する符号化器100を備える。符号化器は、オーディオ信号104を受信し、符号化器の出力108において提供される符号化オーディオ信号を生成する符号化プロセッサ106を含む。符号化プロセッサは、オーディオ信号の連続的なオーディオフレームを処理し、符号化されるべきオーディオ信号104内の雑音を推定するための本発明の手法を実施するようにプログラムまたは構築することができる。しかしながら、他の実施形態において、符号化器は、送信システムの一部である必要はなく、符号化器は、符号化オーディオ信号を生成する独立型デバイスであってもよく、または、オーディオ信号送信機の一部であってもよい。一実施形態によれば、符号化器100は、112において示されているように、オーディオ信号の無線送信を可能にするためのアンテナ110を備えることができる。他の実施形態において、符号化器100は、たとえば、参照符号114において示されているように、有線接続回線を使用して、出力108において提供される符号化オーディオ信号を出力してもよい。

【0028】

図1のシステムは、復号器150をさらに備え、復号器150は、たとえば、有線回線114またはアンテナ154を介して、復号器150によって処理されるべき符号化オーディオ信号を受信する入力152を有する。復号器150は、符号化信号に対して動作し、出力160において復号オーディオ信号158を提供する復号プロセッサ156を備える。復号プロセッサは、復号オーディオ信号104内の雑音を推定するための本発明の手法を実施するための処理のためにプログラムまたは構築することができる。他の実施形態においては、復号器は、送信システムの一部である必要はなく、むしろ、復号器は、符号化オーディオ信号を復号するための独立型デバイスであってもよく、または、オーディオ信号受信機の一部であってもよい。

【0029】

図2は、一実施形態による雑音推定器170の単純化したブロック図を示す。雑音推定器170は、図1に示すオーディオ信号符号化器および/またはオーディオ信号復号器において使用することができる。雑音推定器170は、オーディオ信号102のエネルギー値174を判定するための検出器172と、エネルギー値174を対数領域(変換したエネルギー値178参照)へと変換するための変換器176と、変換したエネルギー値178に基づいてオーディオ信号102の雑音レベル182を推定するための推定器180と

10

20

30

40

50

を含む。推定器 170 は、共通のプロセッサによって実装されてもよく、または、検出器 172、変換器 176 および推定器 180 の機能を実施するようにプログラムまたは構築されている複数のプロセッサによって実装されてもよい。

【0030】

以下において、図 1 の符号化プロセッサ 106 および復号プロセッサ 156 のうちの少なくとも一方において、または、図 2 の推定器 170 によって実施することができる本発明の手法の実施形態をさらに詳細に説明する。

【0031】

図 3 は、オーディオ信号内の雑音を推定するための本発明の手法の流れ図を示す。オーディオ信号が受信され、第 1 のステップ S100 において、オーディオ信号のエネルギー値 174 が判定される。判定されたエネルギー値はその後、ステップ S102 において、対数領域へと変換される。変換されたエネルギー値 178 に基づいて、ステップ S104 において、雑音が推定される。実施形態によれば、ステップ S106 において、対数データ 182 によって表される推定雑音データのさらなる処理が行われるのは、対数領域であるべきか否かについて判定される。対数領域におけるさらなる処理が所望される（ステップ S106 において、はい）場合、推定雑音を表す対数データがステップ S108 において処理され、たとえば、送信が対数領域においても行われる場合に、対数データが送信パラメータへと変換される。そうでない場合（ステップ S106 において、いいえ）ステップ 110 において対数データ 182 が線形データへと変換し戻され、線形データは、ステップ S112 において処理される。

【0032】

実施形態によれば、ステップ S100 において、オーディオ信号のエネルギー値を判定することは、従来手法におけるように行われてもよい。オーディオ信号に適用されている FFT のパワースペクトルが計算され、心理音響的に動機付けられた帯域へとグループ化される。帯域内のパワースペクトルビンは、エネルギー値のセットが得られるように帯域ごとのエネルギー値を形成するように蓄積される。他の実施形態において、パワースペクトルを、MDCT（修正離散コサイン変換）、CLDFB（複素低遅延フィルタバンク）、または、スペクトルの種々の部分をカバーするいくつかの変換の組み合わせのような、任意の適切なスペクトル変換に基づいて計算してもよい。ステップ S100 において、各帯域のエネルギー値 174 が判定され、ステップ S102 において、各帯域のエネルギー値 174 はステップ S102 において対数領域へと変換され、実施形態によれば、log₂ 領域へと変換される。帯域エネルギーは、以下のように log₂ 領域へと変換することができる。

$$E_{n_log} = \frac{[(\log_2(1+E_{n_lin})) \cdot 2^N]}{2^N}$$

$\lfloor x \rfloor$

は $\text{floor}(x)$ であり、 E_{n_log} は log₂ 領域における帯域 n のエネルギー値であり、 R_{n_lin} は線形領域における帯域 n のエネルギー値であり、N は分解能 / 精度である。

【0033】

実施形態によれば、(int)log₂ 関数が通常、固定小数点数における先行ゼロの数を判定する「norm」関数を使用する固定小数点プロセッサ上で、非常に迅速に、たとえば、1 サイクルで計算することができるという点において有利である、log₂ 領域への変換が実施される。時折、上記の式において定数 N によって表現される、(int)log₂ 領域よりも高い精度が必要とされる。このわずかにより高い精度は、norm 命令または近似の後に最上位ビットを有する単純なルックアップテーブルによって達成することができる。これは、より低い精度が許容可能であるときに低複雑度対数計算を達成するための一般的な手法である。上記の式において、変換されたエネルギーが正のままであ

ることを保証するために、 \log_2 関数の内部に定数「1」が追加されている。実施形態によれば、これは、雑音推定器が雑音エネルギーの統計モデルに依拠する場合に重要であり得る。それは、負の値に対して雑音推定を実施することはそのようなモデルに違反することになり、結果として、推定器の予期せぬ挙動をもたらすことになるためである。

【0034】

一実施形態によれば、上記の式においてNは6に設定され、これは、 $2^6 = 64$ ビットのダイナミックレンジと等価である。これは、上述した40ビットのダイナミックレンジよりも大きく、それゆえ、十分である。このデータを処理するために、目標は16ビットデータを使用することであり、9ビットが仮数のために残され、1ビットが符号のために残される。そのようなフォーマットは、一般的に「6Q9」フォーマットとして示される。代替的に、正の値しか考慮されなくてもよいため、符号ビットを回避して仮数に使用することができ、合計10ビットが仮数のために残される。これは「6Q10」フォーマットとして参照される。

10

【0035】

最小値統計アルゴリズムの詳細な記載は、R. Martin「Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics」(2001)に見出すことができる。このアルゴリズムは基本的に、一般的に数秒にわたる、各スペクトル帯域の所与の長さのスライドする時間窓にわたって、平滑化パワースペクトルの最小値を追跡することに存する。アルゴリズムはまた、雑音推定の精度を改善するためのバイアス補償をも含む。その上、時間変動雑音の追跡を改善するために、もたらされる推定雑音エネルギーの増大が穏やかであることを条件として、元の最小値の代わりに、はるかにより短い時間窓にわたって計算する局所的な最小値の追跡を使用することができる。増大の許容量はR. Martin「Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics」(2001)において、パラメータnoise_slope_maxによって決定される。一実施形態によれば、従来どおり、線形エネルギーデータに対して作動する最小値統計雑音推定アルゴリズムが使用される。しかしながら、本発明者らの知見によれば、オーディオ材料または発話材料内の雑音レベルを推定する目的で、アルゴリズムには代わりに対数入力データを供給することができる。信号処理自体は修正されないままであるが、最小限の再調整のみが必要とされる。この再調整は、線形データと比較した対数データのダイナミックレンジの低減に対処するためにパラメータnoise_slope_maxを低減することに存する。これまでのところ、最小値統計アルゴリズム、または、他の適切な雑音推定技法は、線形データに対して作動される必要があると仮定されていた。すなわち、実際には対数表現であるデータは適切でないと仮定されていた。この従来の仮定とは対照的に、本発明者らは、ほとんどの演算は16ビットで行うことができ、依然として32ビットを必要とするのはアルゴリズムのいくらかの部分のみであるため、雑音推定は実際には、16ビットでしか表されない入力データを使用することを可能にし、結果として、固定小数点実施態様において複雑度ははるかにより低くすることを可能にする対数データに基づいて作動され得ることを見出した。最小値統計アルゴリズムにおいて、たとえば、バイアス補償は、入力パワーの分散、したがって、一般的に依然として32ビット表現を必要とする4次統計に基づく。

20

30

40

【0036】

図3に関連して上述したように、雑音推定プロセスの結果は、種々の様式でさらに処理され得る。実施形態によれば、第1の様式は、たとえば、送信パラメータが、しばしばそうであるように対数領域においても送信される場合に、対数データ182を送信パラメータへと直接的に変換することによって、ステップS108に示すように、対数データ182を直接的に使用することである。第2の様式は、たとえば、表引きとともに、または、近似を使用することによって、通常、非常に高速で、一般的にプロセッサ上で1サイクル

50

しか必要としない、たとえば、以下のようなシフト関数を使用して、対数データがさらなる処理のために線形領域へと変換し戻されるように、対数データ 182 を処理することである。

$$E_{n,lin} = 2^{(E_{n,log}-1)}$$

【0037】

以下において、対数データに基づいて雑音を推定するための本発明の手法を実施するための詳細な例は、符号化器を参照しながら説明するが、上記で概説したように、本発明の手法は、たとえば、両方とも参照により本明細書に組み込まれる、国際出願 E P 2 0 1 2 / 0 7 7 5 2 5 号明細書または国際出願 E P 2 0 1 2 / 0 7 7 5 2 7 号明細書に記載されているように、復号器において復号されている信号に適用することもできる。以下の実施形態は、図 1 の符号化器 100 のような、オーディオ符号化器においてオーディオ信号内の雑音を推定するための本発明の手法の実施態様を説明する。より詳細には、拡張音声サービスコード (EVS コード) において受信するオーディオ信号内の雑音を推定するための本発明の手法を実施するための、EVS 符号化器の信号処理アルゴリズムの説明がなされる。

10

【0038】

16 ビット等速 PCM (パルスコード変調) フォーマットにおける、20 ms 長のオーディオサンプルの入力ブロックを仮定する。4 つのサンプリングレート、たとえば、8000、16000、32000 および 48000 サンプル / s、および、可能性として 5.9、7.2、8.0、9.6、13.2、16.4、24.4、32.0、48.0、64.0 または 128.0 kbit / s の、符号化ビットストリームのビットレートを仮定する。6.6、8.85、12.65、14.85、15.85、18.25、19.85、23.05 または 23.85 kbit / s の、符号化ビットストリームのビットレートにおいて動作する AMR - WB (適応的マルチレート広帯域 (コーデック)) 相互運用モードも提供され得る。

20

【0039】

以下の説明の目的で、以下の慣習を、数式に適用する。

$$\lfloor x \rfloor$$

30

は、 x 以下の最大の整数を示す。すなわち、

$$\lfloor 1.1 \rfloor = 1 \quad \lfloor 1.0 \rfloor = 1 \text{ および } \lfloor -1.1 \rfloor = -2$$

である。 \sum は、総和を示す。

【0040】

別途指定しない限り、 $\log(x)$ は、以下の説明全体を通じて、10 を底とする対数を示す。

【0041】

符号化器は、48、32、16 または 8 kHz においてサンプリングされる全帯域 (FB)、超広帯域 (SWB)、広帯域 (WB) または狭帯域 (NB) 信号を許容する。同様に、復号器出力は、48、32、16 または 8 kHz の FB、SWB、WB または NB であり得る。パラメータ R (8、16、32 または 48) を、符号化器における入力サンプリングレートまたは復号器における出力サンプリングレートを示すために使用する。

40

【0042】

入力信号は、20 ms フレームを使用して処理される。コーデック遅延は、入力および出力のサンプリングレートに依存する。WB 入力および WB 出力について、全体的なアルゴリズム遅延は 42.875 ms である。これは、1 つの 20 ms フレーム、入力および出力サンプリングフィルタの 1.875 ms の遅延、符号化器先読みの 10 ms、1 ms のポストフィルタリング遅延、および、復号器における、上位層変換コーディングの重

50

畳加算演算を可能にするための10msから構成される。NB入力およびNB出力について、上位層は使用されず、10msの復号器遅延は、フレーム消去の存在下でのコーデック性能の改善および音楽信号に使用される。NB入力およびNB出力の全体的なアルゴリズム遅延は、1つの20msフレーム、入力リサンプリングフィルタの2ms、符号化器先読みの10ms、出力リサンプリングフィルタの1.875ms、および符号化器における10msの遅延の、43.875msである。出力がレイヤ2に限定される場合、コーデック遅延は10msだけ低減することができる。

【0043】

符号化器の全体的な機能は、以下の処理セクション、すなわち、一般的な処理、CELLP（符号励振線形予測）コード化モード、MDCT（修正離散コサイン変換）コード化モード、切り替えコード化モード、フレーム消去隠蔽サイド情報、DTX/CNG（不連続送信/快適雑音生成器）動作、AMR-WB相互運用オプション、およびチャネルアウェア符号化を含む。

10

【0044】

本発明の実施形態によれば、本発明の手法は、DTX/CNG動作セクションにおいて実施される。コーデックは各入力フレームをアクティブまたは非アクティブとして分類するための信号アクティビティ検出(SAD)アルゴリズムを備える。これは、可変ビットレートにおいて背景雑音の統計を近似および更新するために周波数領域快適雑音生成(FD-CNG)モジュールが使用される、不連続送信(DTX)動作をサポートする。したがって、非アクティブ信号期間の間の伝送速度は可変であり、背景雑音の推定レベルに依存する。しかしながら、CNG更新速度はまた、コマンドラインパラメータによって固定することもできる。

20

【0045】

スペクトル-時間特性に関して実際の入力背景雑音を模倣する人工雑音を作り出すことを可能にするために、FD-CNGは、雑音推定アルゴリズムを利用して、符号化器入力に存在する背景雑音のエネルギーを追跡する。雑音推定値はその後、非アクティブ段階の間に復号器側で各周波数帯域において生成されるランダム系列の大きさを更新するために、SID（無音挿入記述子）フレームの形態のパラメータとして送信される。

【0046】

FD-CNG雑音推定器は、ハイブリッドスペクトル分析手法に依拠する。コア帯域幅に対応する低周波数は、高分解能FFT分析によってカバーされ、一方で、残りのより高い周波数は、400Hzの大幅により低いスペクトル分解能を呈するCLDFBによって捕捉される。CLDFBは、入力信号をコアサンプリングレートにダウンサンプリングするためのリサンプリングツールとしても使用されることに留意されたい。

30

【0047】

しかしながら、SIDフレームのサイズは、実際には限定される。背景雑音を記述するパラメータの数を低減するために、入力エネルギーは結局、パーティションと呼ばれるスペクトル帯域のグループの間で平均される。

【0048】

1. スペクトルパーティションエネルギー

40

パーティションエネルギーは、FFTおよびCLDFB帯域について別個に計算される。その後、FFTパーティションに対応する $L^{[FET]}_{SID}$ エネルギー、および、CLDFBパーティションに対応する $L^{[CLDFB]}_{SID}$ エネルギーが、サイズ $L_{SID} = L^{[FET]}_{SID} + L^{[CLDFB]}_{SID}$ の単一アレイ E_{FD-CNG} へと連結される。これは、後述する雑音推定器に対する入力としての役割を果たすことになる（「2. FD-CNG雑音推定」参照）。

【0049】

1.1 FFTパーティションエネルギーの計算

コア帯域幅をカバーする周波数のパーティションエネルギーは、以下のように得られる。

50

$$E_{\text{FD-CNG}}(i) = \frac{E_{\text{CB}}^{[0]}(i) + E_{\text{CB}}^{[1]}(i)}{2} H_{\text{de-emph}}(i) \quad i = 0, \dots, L_{\text{SID}}^{[\text{FFT}]} - 1$$

式中、 $E_{\text{CB}}^{[0]}(i)$ 及び $E_{\text{CB}}^{[1]}(i)$ はそれぞれ、第1の分析窓および第2の分析窓の臨界帯域 i における平均エネルギーである。コア帯域幅を捕捉するFFTパーティションの数 $L_{\text{SID}}^{[\text{FFT}]}$ は、使用される構成に従って、17から21の間に及ぶ（「1.3 FD-CNG符号化器構成」参照）。ディエンファシススペクトル重み $H_{\text{de-emph}}(i)$ は、ハイパスフィルタを補償するために使用され、以下のように定義される。

$$\{H_{\text{de-emph}}(0), \dots, H_{\text{de-emph}}(L_{\text{SID}}^{[\text{FFT}]} - 1)\} = \{9.7461, 9.5182, 9.0262, 8.3493, 7.5764, 6.7838, 5.8377, \\ 4.8502, 4.0346, 3.2788, 2.6283, 2.0920, 1.6304, 1.2850, \\ 1.0108, 0.7916, 0.6268, 0.5011, 0.4119, 0.3637\}.$$

10

【0050】

1.2 CLDFBパーティションエネルギーの計算

コア帯域幅を上回る周波数のパーティションエネルギーは、以下のように計算される。

$$E_{\text{FD-CNG}}(i) = \frac{1}{168(A_{\text{CLDFB}})^2} \frac{1}{j_{\text{max}}(i) - j_{\text{min}}(i) + 1} \sum_{j=j_{\text{min}}(i)}^{j_{\text{max}}(i)} E_{\text{CLDFB}}(j) \quad i = L_{\text{SID}}^{[\text{FFT}]} - 1, \dots, L_{\text{SID}}^{[\text{FFT}]} + L_{\text{CLDFB}} - 1$$

20

式中、 $j_{\text{min}}(i)$ 及び $j_{\text{max}}(i)$ はそれぞれ、 i 番目のパーティション内の第1のCLDFB帯域および最後のCLDFB帯域のインデックスであり、 $E_{\text{CLDFB}}(j)$ は j 番目のCLDFB帯域の総エネルギーであり、 A_{CLDFB} はスケール係数である。定数168は、CLDFB内の時間スロットの数を指す。CLDFBパーティションの数 L_{CLDFB} は、後述するように、使用される構成に依存する。

【0051】

1.3 FD-CNG符号化器構成

以下の表は、符号化器における種々のFD-CNG構成についてのパーティションの数およびそれらの上方境界をリストしている。

30

	ビット レート [k b p s]	$L_{\text{SID}}^{\text{[FFT]}}$	$L_{\text{SID}}^{\text{[CLDFB]}}$	$f_{\text{max}}(i),$ $i = 0, \dots, L_{\text{SID}}^{\text{[FFT]}} - 1$ [H z]	$f_{\text{max}}(i),$ $i = L_{\text{SID}}^{\text{[FFT]}} , \dots, L_{\text{SID}} - 1$ [H z]
N B	•	17	0	100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3975	×
W B	≤ 8	20	0	100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375	×
	$8 < \bullet \leq 13.2$	20	1	100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375	8000
	> 13.2	21	0	100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375, 7975	×
S W B / F B	≤ 13.2	20	4	100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375	8000, 10000, 12000, 14000
	> 13.2	21	3	100, 200, 300, 400, 500, 600, 750, 900, 1050, 1250, 1450, 1700, 2000, 2300, 2700, 3150, 3700, 4400, 5300, 6375, 7975	10000, 12000, 16000

10

20

30

表 1 : 符号化器における F D - C N G 雑音推定の構成

【 0 0 5 2 】

各パーティション $i = 0, \dots, L_{\text{SID}} - 1$ について、

$$f_{\text{max}}(i)$$

40

は、 i 番目のパーティション内の最後の帯域の周波数に対応する。各スペクトルパーティション内の第 1 の帯域および最後の帯域のインデックス $j_{\text{min}}(i)$ 及び $j_{\text{max}}(i)$ は、以下のように、コアの構成の関数として導出され得る。

$$j_{\max}(i) = \begin{cases} f_{\max}(i) \frac{\text{core_FFT_length}}{\text{core_sampling_rate}} & i = 0, \dots, L_{\text{SID}}^{\text{[FFT]}} - 1 \\ j_{\max}(L_{\text{SID}}^{\text{[FFT]}} - 1) + \frac{2f_{\max}(i) - \text{core_sampling_rate}}{800} & i = L_{\text{SID}}^{\text{[FFT]}} - 1, \dots, L_{\text{SID}} - 1 \end{cases}$$

$$j_{\min}(i) = \begin{cases} f_{\min}(0) \frac{\text{core_sampling_rate}}{\text{core_FFT_length}} & i = 0 \\ j_{\max}(i-1) + 1 & i > 0 \end{cases}$$

式中、

$$f_{\min}(0) = 50\text{Hz}$$

10

は、第1のスペクトルパーティション内の第1の帯域の周波数である。したがって、FD-CNGは、50Hzよりも上でのみ、何らかの快適雑音を生成する。

【0053】

2. FD-CNG雑音推定

FD-CNGは、入力スペクトル内に存在する背景雑音のエネルギーを追跡するために、雑音推定器に依拠する。これは主に、R. Martin「Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics」(2001)によって記載されている最小値統計アルゴリズムに基づく。しかしながら、入力エネルギーのダイナミックレンジ

20

$$\{E_{\text{FD-CNG}}(0), \dots, E_{\text{FD-CNG}}(L_{\text{SID}} - 1)\}$$

を低減し、したがって、雑音推定アルゴリズムの固定小数点実施態様を促進するために、雑音推定の前に非線形変換が適用される(「2.1 入力エネルギーに対するダイナミックレンジ圧縮」参照)。その後、結果もたらされる雑音推定値に対して逆変換を使用して、元のダイナミックレンジを復元する(「2.3 推定雑音エネルギーのダイナミックレンジ拡張」参照)。

【0054】

2.1 入力エネルギーに対するダイナミックレンジ圧縮

入力エネルギーを非線形関数によって処理し、以下のように、9ビット分解能で量子化する。

$$E_{\text{MS}}(i) = \frac{\lfloor \log_2((1 + E_{\text{FD-CNG}}(i))2^9) \rfloor}{2^9} \quad i = 0, \dots, L_{\text{SID}} - 1$$

30

【0055】

2.2 雑音追跡

最小値統計アルゴリズムの詳細な記載は、R. Martin「Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics」(2001)に見出すことができる。このアルゴリズムは基本的に、一般的に数秒にわたる、各スペクトル帯域の所与の長さのスライドする時間窓にわたって、平滑化パワースペクトルの最小値を追跡することに存する。アルゴリズムはまた、雑音推定の精度を改善するためのバイアス補償をも含む。その上、時間変動雑音の追跡を改善するために、もたらされる推定雑音エネルギーの増大が穏やかであることを条件として、元の最小値の代わりに、はるかに短い時間窓にわたって計算される局所的な最小値の追跡を使用することができる。増大の許容量はR. Martin「Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics」(2001)において、

40

50

パラメータ `noise_slope_max` によって決定する。

【0056】

雑音追跡器の主な出力は、雑音推定値

$$N_{MS}(i), i = 0, \dots, L_{SID} - 1$$

である。快適雑音においてより平滑な推移を得るために、1次再帰フィルタ、すなわち、

$$\bar{N}_{MS}(i) = 0.95\bar{N}_{MS}(i) + 0.05N_{MS}(i)$$

を適用することができる。

【0057】

さらに、入力エネルギー $E_{MS}(i)$ が最後の5フレームにわたって平均化される。これは、各スペクトルパーティション内の

$$\bar{N}_{MS}(i)$$

に対して上限を適用するために使用される。

【0058】

2.3 推定雑音エネルギーのダイナミックレンジ拡張

推定雑音エネルギーは、上述したダイナミックレンジ圧縮を補償するために非線形関数によって処理される。

$$N_{FD-CNG}(i) = 2^{\bar{N}_{MS}(i)-1} \quad i = 0, \dots, L_{SID} - 1.$$

【0059】

本発明によれば、特に、固定小数点計算を使用するプロセッサ上で処理されるオーディオ/発話信号について、雑音推定器の複雑度を低減することを可能にする、オーディオ信号内の雑音を推定するための改善された手法を説明する。本発明の手法は、たとえば、高スペクトル-時間分解能での快適雑音の生成について参照する国際出願EP2012/077527号明細書、または、低ビットレートにおける背景雑音のモデル化のための快適雑音付加について参照する国際出願EP2012/077527号明細書において記載されている環境における、オーディオ/発話信号処理のための雑音推定器に使用されるダイナミックレンジを低減することを可能にする。説明されているシナリオにおいて、雑音の多い発話信号、たとえば、電話通話において非常に一般的な状況である、背景雑音の存在下での発話、および、EVSコーデックの試験されるカテゴリのうちの1つについて、背景雑音の品質を増強するために、または、快適雑音生成のために、最小値統計アルゴリズムに基づいて動作する雑音推定器を使用する。EVSコーデックは、標準化によれば、固定演算を用いるプロセッサを使用することになり、本発明の手法は、もはや線形領域ではなく、対数領域においてオーディオ信号のエネルギー値を処理することによって、最小値統計雑音推定器に使用される信号のダイナミックレンジを低減することによって、処理複雑度を低減することを可能にする。

【0060】

説明されている概念のいくつかの態様は、装置の文脈において説明されているが、これらの態様が、対応する方法の説明をも表すことは明らかであり、ブロックまたはデバイスは方法ステップまたは方法ステップの特徴に対応する。同様に、方法ステップの文脈において説明されている態様は、対応するブロックもしくは項目または対応する装置の特徴の説明をも表す。

【0061】

特定の実施要件に応じて、本発明の実施形態は、ハードウェアまたはソフトウェアにおいて実装することができる。実施態様は、それぞれの方法が実施されるようにプログラム可能コンピュータシステムと協働する（または協働することが可能である）、電子可読制御信号を記憶しているデジタル記憶媒体、たとえば、フロッピーディスク、DVD、Blue-Ray、CD、ROM、PROM、EPROM、EEPROMまたはフラッシュメモ

10

20

30

40

50

モリを使用して実施することができる。それゆえ、デジタル記憶媒体は、コンピュータ可読であり得る。

【0062】

本発明によるいくつかの実施形態は、本明細書において説明されている方法のうちの1つが実施されるように、プログラム可能コンピュータシステムと協働することが可能である、電子可読制御信号を有するデータキャリアを含む。

【0063】

一般的に、本発明の実施形態は、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として実装することができ、プログラムコードは、コンピュータプログラム製品がコンピュータ上で作動するときに、本方法の1つを実施するように動作可能である。プログラ
ムコードを、たとえば、機械可読キャリア上に記憶してもよい。

10

【0064】

他の実施形態は、機械可読キャリア上に記憶している、本明細書において説明されている方法の1つを実施するためのコンピュータプログラムを含む。

【0065】

すなわち、それゆえ、本発明の方法の一実施形態は、コンピュータプログラムがコンピュータ上で作動すると、本明細書において説明されている方法の1つを実施するためのプログラムコードを有するコンピュータプログラムである。

【0066】

それゆえ、本発明の方法のさらなる実施形態は、本明細書において説明されている方法の1つを実施するためのコンピュータプログラムを記録して含む、データキャリア（またはデジタル記憶媒体もしくはコンピュータ可読媒体）である。

20

【0067】

それゆえ、本発明の方法のさらなる実施形態は、本明細書において記載されている方法のうちの1つを実施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または信号系列である。データストリームまたは信号系列は、たとえば、データ通信接続、たとえばインターネットを介して転送されるように構成することができる。

【0068】

さらなる実施形態は、本明細書において記載されている方法のうちの1つを実施するように構成または適合されている処理手段、たとえば、コンピュータまたはプログラム可能な論理装置を含む。

30

【0069】

さらなる実施形態は、本明細書において説明されている方法の1つを実施するためのコンピュータプログラムをインストールされているコンピュータを含む。

【0070】

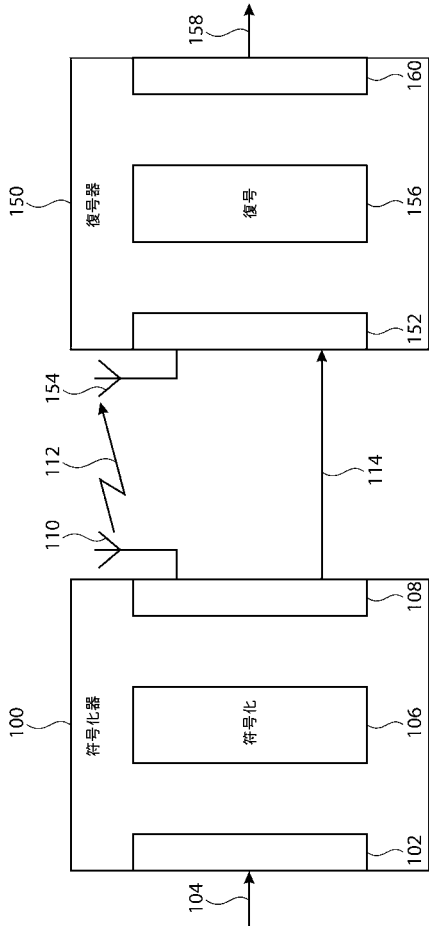
いくつかの実施形態において、プログラム可能な論理装置（たとえば、フィールドプログラマブルゲートアレイ F P G A ）が、本明細書において説明されている方法の機能の一部またはすべてを実施するために使用されてもよい。いくつかの実施形態において、フィールドプログラマブルゲートアレイは、本明細書において説明されている方法のうちの1つを実施するために、マイクロプロセッサと協働することができる。一般的に、方法は、
任意のハードウェア装置によって実施されることが好ましい。

40

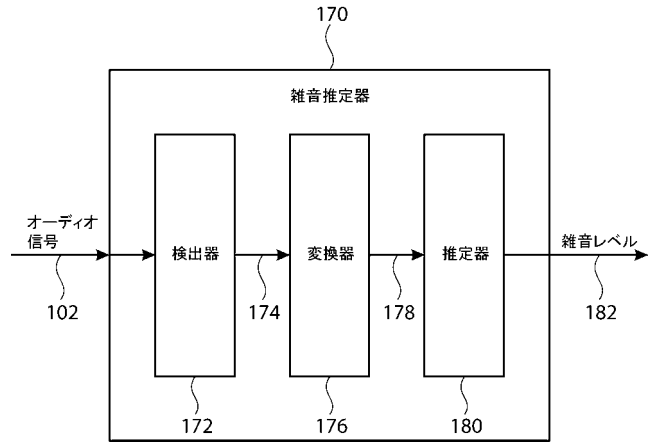
【0071】

上述した実施形態は、本発明の原理の例示に過ぎない。本明細書において記載されている構成および詳細の修正および変形は、当該技術分野においては明らかであると理解されたい。それゆえ、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、本明細書において実施形態の記述および説明によって示される特定の詳細によっては限定されない。

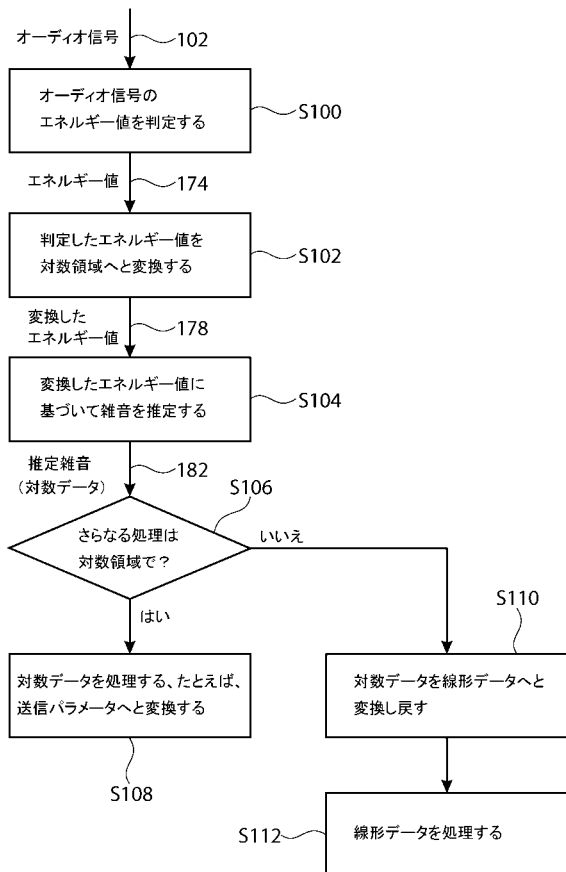
【 図 1 】



【 図 2 】



【 図 3 】



フロントページの続き

- (72)発明者 アンソニー・ロムバート
ドイツ連邦共和国 9 1 0 5 8 エアランゲン フォーゲルヘルト 9 2
- (72)発明者 マーティン・ディエツ
ドイツ連邦共和国 9 0 4 2 9 ニュルンベルグ ドイツェヘルンシュトラッセ 3 7
- (72)発明者 マルクス・ムルトウルス
ドイツ連邦共和国 9 0 4 6 9 ニュルンベルグ エッツラウヴェグ 7