

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2018-513996
(P2018-513996A)

(43) 公表日 平成30年5月31日(2018.5.31)

(51) Int.Cl.

G 1 O L 19/008 (2013.01)
G 1 O L 19/00 (2013.01)

F 1

G 1 O L 19/008 1 O O
G 1 O L 19/00 3 3 O B

テーマコード(参考)

審査請求 未請求 予備審査請求 未請求 (全 31 頁)

(21) 出願番号	特願2017-552843 (P2017-552843)
(86) (22) 出願日	平成28年3月10日 (2016.3.10)
(85) 翻訳文提出日	平成29年12月4日 (2017.12.4)
(86) 國際出願番号	PCT/EP2016/055135
(87) 國際公開番号	W02016/162165
(87) 國際公開日	平成28年10月13日 (2016.10.13)
(31) 優先権主張番号	15305536.3
(32) 優先日	平成27年4月10日 (2015.4.10)
(33) 優先権主張国	歐州特許庁 (EP)
(31) 優先権主張番号	15306144.5
(32) 優先日	平成27年7月10日 (2015.7.10)
(33) 優先権主張国	歐州特許庁 (EP)
(31) 優先権主張番号	15306425.8
(32) 優先日	平成27年9月16日 (2015.9.16)
(33) 優先権主張国	歐州特許庁 (EP)

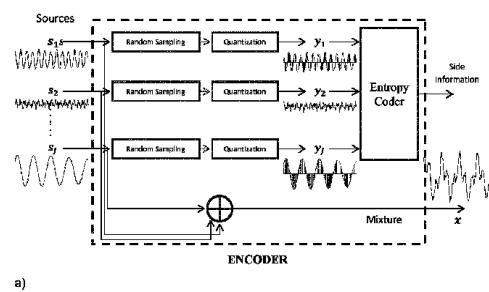
(71) 出願人	501263810 トムソン ライセンシング Thomson Licensing フランス国, 92130 イッシー レ ムーリノー, ル ジヤンヌ ダルク, 1-5 1-5, rue Jeanne d' A rc, 92130 ISSY LES MOULINEAUX, France
(74) 代理人	100107766 弁理士 伊東 忠重
(74) 代理人	100070150 弁理士 伊東 忠彦
(74) 代理人	100091214 弁理士 大貫 進介

最終頁に続く

(54) 【発明の名称】複数のオーディオ信号を符号化する方法およびデバイス、ならびに、分離を改善した、複数のオーディオ信号の混合物を復号する方法およびデバイス

(57) 【要約】

複数のオーディオ信号を符号化する方法は、複数のオーディオ信号の各々をランダム・サンプリングおよび量子化することと、サンプリングおよび量子化した複数のオーディオ信号を、複数のオーディオ信号の混合物から複数のオーディオ信号を復号および分離するために使用することができる副情報をとして符号化することとを含む。複数のオーディオ信号の混合物を復号する方法は、副情報を復号および逆多重化することであって、副情報は複数のオーディオ信号の各々の量子化されたサンプルを含むことと、任意のデータ・ソースから、複数のオーディオ信号の混合物を受信する、または取り出すことと、複数のオーディオ信号に近似した複数の推定オーディオ信号を生成することであって、複数のオーディオ信号の各々の量子化されたサンプルが使用されることとを含む。



a)

Fig.2

【特許請求の範囲】

【請求項 1】

複数の時間領域オーディオ信号を符号化する方法であって、
前記複数の時間領域オーディオ信号の各々をランダム・サンプリングおよび量子化する
ステップと、

前記サンプリングおよび量子化した複数の時間領域オーディオ信号を、前記複数の時間
領域オーディオ信号の混合物から前記複数の時間領域オーディオ信号を復号および分離す
るために使用することができる副情報として符号化するステップと、
を含む、前記方法。

【請求項 2】

前記ランダム・サンプリングは、既定の疑似ランダム・パターンを使用する、請求項 1
に記載の方法。

【請求項 3】

前記複数の時間領域オーディオ信号の混合物は、到着するにつれて順次符号化される、
請求項 1 または 2 に記載の方法。

【請求項 4】

どのソースがどの期間に無音であるかを割り出すステップと、前記割り出した情報を前
記副情報内に符号化するステップとをさらに含む、請求項 1 から 3 の何れか 1 項に記載の
方法。

【請求項 5】

複数のオーディオ信号の混合物を復号する方法であって、
副情報を復号および逆多重化するステップであって、前記副情報は前記複数のオーディ
オ信号の各々の量子化された時間領域サンプルを含む前記ステップと、
記憶装置または任意のデータ・ソースから、前記複数のオーディオ信号の混合物を受信
する、または取り出すステップと、
前記複数のオーディオ信号に近似した複数の推定オーディオ信号を生成するステップで
あって、前記複数のオーディオ信号の各々の前記量子化されたサンプルが使用される前記
ステップと、
を含む、前記方法。

【請求項 6】

前記複数の推定オーディオ信号を生成するステップは、
ランダム非負値から分散テンソル V を計算するステップと、
前記複数のオーディオ信号の前記量子化されたサンプルのソース・パワー・スペクトル
の条件付き期待値を計算するステップであって、推定ソース・パワー・スペクトル $P(f, n, j)$
が得られ、かつ、前記複数のオーディオ信号の前記分散テンソル V および複素
短時間フーリエ変換 (S T F T) 係数が使用される、前記ステップと、
前記推定ソース・パワー・スペクトル $P(f, n, j)$ から前記分散テンソル V を繰り
返し再計算するステップと、
結果として得られる分散テンソル V から、S T F T 係数

\hat{S}

の配列を計算するステップと、

S T F T 係数

\hat{S}

の前記配列を前記時間領域に変換するステップであって、前記複数の推定オーディオ信号
が得られる前記ステップと、
を含む、請求項 5 に記載の方法。

【請求項 7】

前記複数のオーディオ信号のうちの少なくとも 1 つについてオーディオ・インペインテ 50

イングを行うことをさらに含む、請求項 5 または 6 に記載の方法。

【請求項 8】

前記副情報を、どのオーディオ・ソースがどの時間に無音であるかを示す情報をさらに含み、前記分散テンソル V を定義する行列 H および Q を自動的に定めることをさらに含む、請求項 5 から 7 の何れか 1 項に記載の方法。

【請求項 9】

複数のオーディオ信号を符号化する装置であって、
プロセッサと、

実行されたときに、複数の時間領域オーディオ信号を符号化する方法を前記装置に実行させる命令を記憶したメモリと、

を有し、前記方法は、

前記複数の時間領域オーディオ信号の各々をランダム・サンプリングおよび量子化するステップと、

前記サンプリングおよび量子化した複数の時間領域オーディオ信号を、前記複数のオーディオ信号の混合物から前記複数の時間領域オーディオ信号を復号および分離するために使用することができる副情報をとして符号化するステップと、
を含む、前記装置。

【請求項 10】

前記ランダム・サンプリングは、既定の疑似ランダム・パターンを使用する、請求項 9 に記載の装置。

【請求項 11】

複数のオーディオ信号の混合物を復号する装置であって、
プロセッサと、

実行されたときに、複数のオーディオ信号の混合物を復号する方法を前記装置に実行させる命令を記憶したメモリと、

を有し、前記方法は、

副情報を復号および逆多重化するステップであって、前記副情報は前記複数のオーディオ信号の各々の量子化された時間領域サンプルを含む前記ステップと、

記憶装置または任意のデータ・ソースから、前記複数のオーディオ信号の混合物を受信する、または取り出すステップと、

前記複数のオーディオ信号に近似した複数の推定オーディオ信号を生成するステップであって、前記複数のオーディオ信号の各々の前記量子化されたサンプルが使用される前記ステップと、
を含む、前記装置。

【請求項 12】

前記複数の推定オーディオ信号を生成するステップは、
ランダム非負値から分散テンソル V を計算するステップと、

前記複数のオーディオ信号の前記量子化されたサンプルのソース・パワー・スペクトルの条件付き期待値を計算するステップであって、推定ソース・パワー・スペクトル P (f , n , j) が得られ、かつ、前記複数のオーディオ信号の前記分散テンソル V および複素短時間フーリエ変換 (S T F T) 係数が使用される、前記ステップと、

前記推定ソース・パワー・スペクトル P (f , n , j) から前記分散テンソル V を繰り返し再計算するステップと、

結果として得られる分散テンソル V から S T F T 係数

\hat{S}

配列を計算するステップと、

S T F T 係数

\hat{S}

10

20

30

40

50

の前記配列を前記時間領域に変換するステップであって、前記複数の推定オーディオ信号が得られる前記ステップと、
を含む、請求項 1 1 に記載の装置。

【請求項 1 3】

前記複数の時間領域オーディオ信号のうちの少なくとも 1 つについてオーディオ・インペインティングを行うことをさらに含む、請求項 1 1 または 1 2 に記載の装置。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、複数のオーディオ信号を符号化する方法およびデバイス、ならびに、複数のオーディオ信号の分離を改善した、複数のオーディオ信号の混合物を復号する方法およびデバイスに関する。 10

【背景技術】

【0 0 0 2】

オーディオ・ソース分離の問題は、ソース（例えば会話、楽器、雑音など）の混合物から個々のソースを推定することにある。オーディオとの関係においては、混合物とは、1つまたは複数のマイクロホンによって複数のソースを記録したものと意味する。オーディオ信号のインフォームド・ソース・セパレーション (informed source separation : ISS) は、ソースについてのある程度の情報が利用可能であるものとして、個々のオーディオ・ソースをソースの混合物から抽出する問題として見ることができる。ISS は、ソースの混合物が符号化段階および復号段階の両方で既知であるものとして、オーディオ・オブジェクト（ソース）の圧縮 [6]、すなわちマルチソース・オーディオの符号化にも関する。これらの両問題は、互いに結びついている。これらの問題は、幅広い適用分野で重要である。 20

【0 0 0 3】

既知の解決策（例えば、[3]、[4]、[5]）は、元のソースが符号化段階で利用可能であるという前提に立っている。副情報を計算して混合物とともに伝送し、復号段階ではその両方を処理して、ソースを回復する。いくつかの ISS 方法が既知であるが、それらの手法では全て、符号化段階は、復号段階より複雑で計算コストが高い。従って、これらの手法は、符号化を実行するプラットフォームが、エンコーダが要求する計算の複雑さを取り扱うことができない場合には、好ましくない。最後に、既知の複雑なエンコーダは、いくつかの適用分野で非常に重要なオンライン符号化、すなわち信号が到着するにつれて順次符号化していく符号化には使用できない。 30

【発明の概要】

【0 0 0 4】

上記に鑑みて、ISS の両方の問題に対する、完全に自動的且つ効率的な解決策があることが非常に望ましい。特に、エンコーダがデコーダよりも大幅に少ない処理しか必要としない解決策があれば望ましい。

【0 0 0 5】

本発明は、処理負担のほとんどをエンコーダ側からデコーダ側にシフトする簡単な符号化方式を提供する。副情報を生成するための提案する簡単な方法は、低複雑性符号化だけでなく、デコーダにおける効率的な回復も可能にする。最後に、符号化時に信号全体が分かれている必要があるいくつかの既存の効率的な方法（バッチ符号化と呼ばれる）と対照的に、提案する符号化方式では、オンライン符号化が可能である、すなわち、信号が、到着するにつれて順次符号化される。 40

【0 0 0 6】

このエンコーダは、オーディオ・ソースからランダム・サンプルをランダム・パターンでとる。一実施形態では、これは、既定の疑似ランダム・パターンである。サンプリングした値を、既定の量子化器が量子化し、その結果得られる量子化済みサンプルを、エントロピ・コーダが連結し、ロスレス圧縮して、副情報を生成する。混合物も符号化側で生成

10

20

30

40

50

することができ、あるいは、混合物は、復号側で他の方法で既に入手可能である。

【0007】

デコーダは、最初に、副情報から量子化済みサンプルを回復し、次いで、その量子化済みサンプルおよび混合物を前提として、混合物内の最尤ソース (most likely sources) を確率論的に推定する。

【0008】

一実施形態では、本発明の原理は、請求項1に開示する複数のオーディオ信号を符号化する方法に関する。一実施形態では、本発明の原理は、請求項3に開示する複数のオーディオ信号の混合物を復号する方法に関する。

【0009】

一実施形態では、本発明の原理は、以下で述べる符号化方法の各ステップにつき1つの複数の別々のハードウェア構成要素を含む符号化デバイスに関する。一実施形態では、本発明の原理は、以下で述べる復号方法の各ステップにつき1つの、複数の別々のハードウェア構成要素を含む復号デバイスに関する。

10

【0010】

一実施形態では、本発明の原理は、以下で述べるステップを含む符号化方法をコンピュータに実行させる実行可能命令を有するコンピュータ可読媒体に関する。一実施形態では、本発明の原理は、以下で述べるステップを含む復号方法をコンピュータに実行させる実行可能命令を有するコンピュータ可読媒体に関する。

20

【0011】

一実施形態では、本発明の原理は、オーディオ・ソースを分離する符号化デバイスであって、例えばハードウェア・プロセッサ、および少なくとも1つのソフトウェア構成要素を有形に実施する非一時的有形コンピュータ可読記憶媒体など、少なくとも1つのハードウェア構成要素を含む符号化デバイスに関し、少なくとも1つのハードウェア・プロセッサ上で実行されるときに、このソフトウェア構成要素は、以下で述べる符号化方法のステップを引き起こす。一実施形態では、本発明の原理は、オーディオ・ソースを分離する符号化デバイスであって、例えばハードウェア・プロセッサ、および少なくとも1つのソフトウェア構成要素を有形に実施する非一時的有形コンピュータ可読記憶媒体など、少なくとも1つのハードウェア構成要素を含む符号化デバイスに関し、少なくとも1つのハードウェア・プロセッサ上で実行されるときに、このソフトウェア構成要素は、以下で述べる復号方法のステップを引き起こす。

30

【0012】

本発明の原理のさらなる目的、特徴、および利点は、以下の説明および添付の特許請求の範囲を、添付の図面と関連付けて考慮すれば明らかになるであろう。

【0013】

添付の図面を参照して、例示的な実施形態について説明する。

【図面の簡単な説明】

【0014】

【図1】エンコーダおよびデコーダを含む、伝送および/または記憶システムの構造を示す図である。

40

【図2】例示的なエンコーダの簡略化した構造を示す図である。

【図3】例示的なデコーダの簡略化した構造を示す図である。

【図4】C S - I S S と従来の I S Sとの性能比較を示す図である。

【発明を実施するための形態】

【0015】

図1は、エンコーダおよびデコーダを含む、伝送および/または記憶システムの構造を示す図である。元の音声ソース s_1, s_2, \dots, s_j がエンコーダに入力され、エンコーダが、混合物 x および副情報を生成する。デコーダは、混合物 x および副情報を使用して音声を回復するが、ここで、ある程度の情報が喪失しており、従って、デコーダは音声ソースを推定する必要があり、推定音声ソース

50

$\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_J$

を提供するものと仮定している。

【0016】

元のソース s_1, s_2, \dots, s_j は、エンコーダで利用可能であり、エンコーダがこれらを処理して副情報を生成するものと仮定する。混合物は、エンコーダによって生成することもできるし、あるいは、デコーダにおいて他の手段によって入手可能であることもある。例えば、インターネット上で入手可能な既知のオーディオ・トラックでは、個々のソースから生成される副情報は、例えばそのオーディオ・トラックの作者または他者によって記憶することができる。本明細書で述べる1つの問題は、別個のチャネルのオーディオ・ソースが別個のマイクロホンを用いて記録され、それらを重ね合せて混合物を形成することである。例えば多チャネル・オーディオまたは複数のマイクロホンを用いた録音物などの他の構成は、記載する方法をそのまま延長することによって容易に取り扱うことができる。

10

【0017】

本明細書で上記の状況内で考慮する1つの技術的問題は、副情報を生成するエンコーダを有するときに、元のソース s_1, s_2, \dots, s_j に可能な限り近いソース

$\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_J$

を推定することができるデコーダを設計することにある。デコーダは、推定ソースの所との品質で必要な副情報のサイズを最小限に抑えるように、副情報および既知の混合物 x を効率的に使用するものとする。デコーダは、混合物のことも、それがソースを用いてどのように形成されたかも知っているものと仮定する。

20

【0018】

従って、本発明は、2つの部分、すなわちエンコーダおよびデコーダを含む。

【0019】

図2のa)は、例示的なエンコーダの簡略化した構造を示す図である。このエンコーダは、計算的に単純になるように設計されている。このエンコーダは、オーディオ・ソースからランダム・サンプルを探る。一実施形態では、エンコーダは、既定の疑似ランダム・パターンを使用する。別の実施形態では、エンコーダは、任意のランダム・パターンを使用する。サンプリングされた値は、(既定の)量子化器によって量子化され、その結果得られる量子化済みサンプル y_1, y_2, \dots, y_j を、エントロピ・コーダ(例えばハフマン・コーダまたは算術コーダ)によって連結し、ロスレス圧縮して、副情報を生成する。混合物も、復号側で既に利用可能になっていない場合には、生成する。

30

【0020】

図2のb)は、エンコーダ内の例示的な信号を拡大して示す図である。異なるソース信号 s_1, s_2, \dots, s_j を重畳または混合することによって、混合信号 x が得られる。ソース信号 s_1, s_2, \dots, s_j のそれぞれも、ランダム・サンプリング・ユニットでランダム・サンプリングされ、これらのサンプルを1つまたは複数の量子化器(この実施形態では各信号につき1つの量子化器)で量子化して、量子化済みサンプル y_1, y_2, \dots, y_j を得る。量子化済みサンプルを符号化して、副情報として使用する。なお、他の実施形態では、サンプリングと量子化の順序に入れ替わることもあることに留意されたい。

40

【0021】

図3は、例示的なデコーダの簡略化した構造を示す図である。このデコーダは、最初に、副情報から量子化済みサンプル y_1, y_2, \dots, y_j を回復する。次いで、デコーダは、観測したサンプル y_1, y_2, \dots, y_j および混合物 x を仮定し、ソース間の既知の構造および相関を活用して、最尤ソース

$\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_J$

を確率的に推定する。

50

【0022】

エンコーダの考えられる実施態様は、非常に単純である。デコーダの1つの可能な実施態様は、以下の2つの仮定に基づいて動作する。

(1) ソースが、ウィンドウ・サイズFおよびウィンドウ数Nの短時間フーリエ変換(STFT)領域に結合ガウス分布している。

(2) ガウス分布

$$V \in \mathbb{R}_+^{F \times N \times J}$$

の分散テンソルが、ランクKの低ランク非負値テンソル分解(NTF)を有し、

【数1】

$$V(f, n, j) = \sum_{k=1}^K H(n, k) W(f, k) Q(j, k), \quad H \in \mathbb{R}_+^{N \times K}, W \in \mathbb{R}_+^{F \times K}, Q \in \mathbb{R}_+^{J \times K}$$

となるようになっている。

【0023】

これら2つの仮定に従うと、デコーダの動作は、以下のステップで要約することができる。

1. ランダム非負値を有する行列

$$H \in \mathbb{R}_+^{N \times K}, W \in \mathbb{R}_+^{F \times K}, Q \in \mathbb{R}_+^{J \times K}$$

20

を初期化し、分散テンソル

$$V \in \mathbb{R}_+^{F \times N \times J}$$

を

【数2】

$$V(f, n, j) = \sum_{k=1}^K H(n, k) W(f, k) Q(j, k)$$

30

として計算する。

2. 収束または最大反復数に達するまで以下を繰り返す。

2.1

【数3】

$$P(f, n, j) = E\{|S(f, n, j)|^2 | x, y_1, y_2, \dots, y_j, V\}$$

となるように、ソース・パワー・スペクトルの条件付き予想を計算する。ここで、 $S \in \mathbb{C}^{F \times N \times J}$ は、ソースのSTFT複素係数の配列である。この条件付き予想計算についてのさらなる詳細は、以下において示す。

2.2

40

【数4】

$$\mathbf{Q}(j, k) \leftarrow \mathbf{Q}(j, k) \left(\frac{\sum_{f,n} \mathbf{W}(f, k) \mathbf{H}(n, k) \mathbf{P}(f, n, j) \mathbf{V}(f, n, j)^{-2}}{\sum_{f,n} \mathbf{W}(f, k) \mathbf{H}(n, k) \mathbf{V}(f, n, j)^{-1}} \right)$$

$$\mathbf{W}(f, k) \leftarrow \mathbf{W}(f, k) \left(\frac{\sum_{j,n} \mathbf{Q}(j, k) \mathbf{H}(n, k) \mathbf{P}(f, n, j) \mathbf{V}(f, n, j)^{-2}}{\sum_{j,n} \mathbf{Q}(j, k) \mathbf{H}(n, k) \mathbf{V}(f, n, j)^{-1}} \right)$$

$$\mathbf{H}(n, k) \leftarrow \mathbf{H}(n, k) \left(\frac{\sum_{f,j} \mathbf{W}(f, k) \mathbf{Q}(j, k) \mathbf{P}(f, n, j) \mathbf{V}(f, n, j)^{-2}}{\sum_{f,j} \mathbf{W}(f, k) \mathbf{Q}(j, k) \mathbf{V}(f, n, j)^{-1}} \right)$$

10

となるように、推定ソース・パワー・スペクトル $P(f, n, j)$ の 3 値テンソルと NTF モデル近似 $V(f, n, j)$ の 3 値テンソルの間の IS 発散 [15] を最小限にする乗法的更新 (MU) 規則を用いて、NTF モデル・パラメータ

$$\mathbf{H} \in \mathbb{R}_+^{N \times K}, \mathbf{W} \in \mathbb{R}_+^{F \times K}, \mathbf{Q} \in \mathbb{R}_+^{J \times K}$$

を再推定する。これらの更新は、複数回反復することができる。

3. STFT 係数 $S \in \mathbb{C}^{F \times N \times J}$ の配列を、

【数5】

20

$$\hat{\mathbf{S}}(f, n, j) = E\{\mathbf{S}(f, n, j) | \mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_J, \mathbf{V}\}$$

として事後平均として計算し、時間領域に変換して戻して、推定ソース

$$\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \dots, \tilde{\mathbf{s}}_J$$

を回復する。この事後平均計算についてのさらなる詳細は、以下において示す。

【0024】

以下、上記の計算についてのいくつかの数学的基礎について説明する。

【0025】

30

テンソルは、より高次の行列と見ることができるデータ構造である。行列は 2 次元であるが、テンソルは N 次元とすることができます。ここでの場合では、V は、(立方体のような) 3 次元テンソルである。これは、ソースの結合ガウス分布の共分散行列を表す。

【0026】

40

行列は、低ランク・モデルでは、それぞれ 2 つのベクトルを乗算することによって形成されるいくつかのランク 1 行列の和として表すことができる。ここでの場合では、テンソルも、同様に K 個のランク 1 テンソルの和として表され、ここで、ランク 1 テンソルは、例えば $\mathbf{h}_i, \mathbf{q}_i$ および \mathbf{w}_i など 3 つのベクトルを乗算することによって形成される。これらのベクトルをまとめて、行列 H, Q および W を形成する。K 個のランク 1 テンソルには、K セットのベクトルがある。基本的に、テンソルは、K 個の成分で表され、行列 H, Q および W は、これらの成分が、それぞれ異なるフレームに沿って、STFT の異なる周波数に沿って、また異なるソースに沿って、どのように分布するかを表す。行列における低ランク・モデルと同様に、小さい K の方が、より良好に例えば音楽のようなオーディオ・データなどのデータの特性を定義するので、K は小さく保たれる。従って、V を低ランク・テンソルとすべきであるという情報を使用することによって、信号の未知の特徴を推測することができる。これにより、未知数の数が減少し、データの様々な部分の間の相互関係が定義される。

【0027】

上述の反復アルゴリズムのステップは、以下のように説明することができる。

【0028】

50

最初に、行列 H , Q および W と、従って V とを初期化する。

【0029】

V が与えられると、信号の確率分布が分かる。信号の観測部分を見ると（信号は部分的にしか観測されない）、例えばウィーナ・フィルタリングによって、S T F T 係数

\hat{S}

を推定することができる。これは、信号の事後平均である。さらに、信号の事後共分散も計算する。これは、以下で使用する。このステップは、信号のウインドウごとに独立して実行され、並列化可能である。これは、予想ステップまたは E ステップと呼ばれる。

10

【0030】

事後平均および事後共分散を計算したら、これらを使用して、事後パワー・スペクトル p を計算する。これは、所期のモデル・パラメータ、すなわち H , Q および W を更新するために必要である。より良好な推定を得るために、このステップを複数回（例えば 2 ~ 10 回）繰り返すと有利であることもある。これは、最大化ステップまたは M ステップと呼ばれる。

【0031】

モデル・パラメータ H , Q および W を更新したら、一実施形態では、ある程度の収束に到達するまで、(S T F T 係数

20

\hat{S}

の推定からの) 全てのステップを繰り返すことができる。収束に到達した後で、一実施形態では、S T F T 係数

\hat{S}

の事後平均を時間領域に変換して、最終結果としてオーディオ信号を得る。

【0032】

本発明の 1 つの利点は、複数のオーディオ・ソース信号のその混合物からの回復を改善することができることである。これにより、強力なデバイスを必要とすることなく、マルチソース・オーディオ記録の効率的な記憶および伝送が可能になる。重いバッテリ・ドレンまたはプロセッサを利用することなく、携帯電話または携帯タブレットを容易に使用して、オーディオ・トラックの複数のソースに関する情報を圧縮することができる。

30

【0033】

別の利点は、個々のソースについての圧縮情報が、それらが必要な場合にしか復号されないので、ソースを符号化および復号するための計算資源が、より効率的に利用されることである。音楽製作など、いくつかの適用分野では、個々のソースについての情報は必ず符号化されて記憶されるが、必ずしもその後に必要となって取り出されるとは限らない。従って、全ての符号化オーディオ・ストリームに対して複雑性の高い処理を実行する高価なエンコーダに対して、複雑性の低いエンコーダおよび複雑性の高いデコーダを備えるシステムは、個々のソースが後に実際に必要となるオーディオ・ストリームについてのみ処理能力を利用するという利点を有する。

40

【0034】

本発明が提供する第 3 の利点は、新たに良好な復号方法への適合性があることである。データ内の相関を活用する新しい改良された方法が発見されたときには、新たな復号方法 (x , y_1 , y_2 , … , y_j) を前提として

$\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_j$

を推定するより良好な方法) を考案することができ、古い符号化ビットストリームを、そ

50

のソースを再符号化しなくても、より良好な品質で復号することができる。従来の符号化／復号パラダイムでは、データ内の相関を活用する改良された方法から新たな符号化方法が得られたときに、その新たな手法の利点を活用するためには、そのソースを復号して再符号化する必要がある。さらに、既に符号化されているビットストリームを再符号化するプロセスは、元のソースに関するさらなる誤差を生じることが分かっている。

【0035】

本発明の第4の利点は、ソースをオンラインで符号化することができること、すなわちソースがエンコーダに到着するときにそれらのソースが符号化されること、ならびにストリーム全体が利用可能でなくても符号化できることである。

【0036】

本発明の第5の利点は、分離したオーディオ・ソース信号におけるギャップを修復することができることである。これは、オーディオ・インペインティングと呼ばれる。従って、本発明は、以下で述べるように、結合したオーディオ・インペインティングおよびソース分離を可能にする。

【0037】

本明細書に開示する手法は、やはり複雑性をエンコーダからデコーダにシフトさせることを目的とする、分散型ソース・コーディング [9]、および特に分散型ビデオ・コーディング [10] パラダイムに触発されて得られたものである。この手法は、ソースが、オーディオ・ソースが疎である場合の基底とインコヒーレント (incoherent) である [13] 基底のベクトルのランダムに選択された部分集合がまたがる線形部分空間に投影されるので、圧縮センシング／サンプリングの原理 [11]～[13] に依拠する。開示する手法は、圧縮サンプリング型 ISS (CS - ISS) と呼ぶことができる。

【0038】

さらに詳細には、ソースの時間的サンプルの部分集合を単純にランダム選択し、その後に一様な量子化およびエントロピ符号化を行うことによってソースを符号化することを提案する。一実施形態では、これが、デコーダに伝送される唯一の副情報となる。

【0039】

なお、時間領域のサンプリングの利点は2つあることに留意されたい。第1に、時間領域のサンプリングは、いかなる変換領域におけるサンプリングよりも高速である。第2に、時間的基底が、オーディオ信号が疎である短時間フーリエ変換 (STFT) フレームと十分にインコヒーレントであり、STFT係数の低ランクNTF表現とはさらにインコヒーレントである。測定領域および事前情報領域のインコヒーレンシがソースの回復に重要なことは、圧縮センシング理論で示されている [13]。

【0040】

デコーダにおいて、量子化済みのソース・サンプルおよび混合物からソースを回復するために、モデル型の圧縮センシングに従うモデル型の手法を使用することが提案されている [14]。特に、一実施形態では、[4]、[5] のように、ソース・スペクトル写真的 Itakura-Saito (IS) 非負値テンソル分解 (NTF) モデルを使用する。そのガウス確率論的定式化 [15] のおかげで、このモデルは、混合物とソース・サンプルの量子化され伝送された部分とから最尤 (ML) 的に推定することができる。このモデルを推定するために、乗法的更新 (MU) 規則 [15] に基づく新たな汎用期待値最大化 (GEM) アルゴリズム [16] を使用することができる。推定したモデルおよび他の全ての観測結果が与えられれば、ウィーナ・フィルタリング [17] によってソースを推定することができる。

【0041】

CS - ISS フレームワークの概要

提案するCS - ISS エンコーダ／デコーダの全体の構造を、上記で既に説明したように、図2に示す。エンコーダは、既定のランダム化パターンを使用して所望のレートでランダムにソースをサブサンプリングし、これらのサンプルを量子化する。次いで、量子化されたサンプルを、単一のストリームに整列させて、エントロピ・エンコーダで圧縮して

10

20

30

40

50

、最終的な符号化ビットストリームを形成する。ランダム・サンプリング・パターン（またはランダム・パターンを生成するシード）は、一実施形態では、エンコーダおよびデコーダの両方が知っているので、伝送する必要はない。別の実施形態では、ランダム・サンプリング・パターン、またはランダム・パターンを生成するシードを、デコーダに伝送する。オーディオ混合物も、デコーダが知っているものと仮定する。デコーダは、エントロピ復号を実行して、ソースの量子化済みサンプルを取り出し、その後に、以下で詳細に述べるように C S - I S S 復号を行う。

【0 0 4 2】

提案する C S - I S S フレームワークは、従来の I S S に優る利点をいくつか有する。それらは、以下のように要約することができる。

10

【0 0 4 3】

第 1 の利点は、図 2 の単純なエンコーダを、例えば低パワー・デバイスで必要とされる低複雑性符号化に使用することができることである。低複雑性符号化方式は、符号化を頻繁に使用するが、符号化したストリームのいくつかしか復号する必要がない適用分野でも有利である。このような適用分野の例はスタジオでの音楽製作であり、スタジオでの音楽製作においては、製作した各々の音楽のソースを将来の使用に備えて保持しておくが、それらが必要になることは滅多にない。従って、C S - I S S では、処理能力および処理時間の有意な節約が可能である。

【0 0 4 4】

第 2 の利点は、（変換領域ではなく）時間領域でサンプリングを実行することにより、単純なサンプリング方式になるだけでなく、必要なときにはオンラインで符号化を実行することが可能になることであり、これは、他の方法では必ずしもそれほど直接的ではない [4, 5]。さらに、独立符号化方式により、復号の効率を損なうことなく、ソースを分散式に符号化することが可能になる。

20

【0 0 4 5】

第 3 の利点は、符号化ステップが、復号ステップについてのいかなる仮定もなく実行されることである。従って、本実施形態で提案するデコーダ以外のデコーダを使用することもできる。これにより、より良好な働きをするデコーダが設計されたときに、符号化したソースが、再符号化の必要なく、改善された復号の恩恵を直接受けるという点で、従来の I S S [2] ~ [5] に優る有意な利点が得られる。これは、エンコーダで使用されるランダム・サンプリングによって可能になる。圧縮センシング理論は、ランダム・サンプリング方式が多数の領域とのインコヒーレンシを提供するので、データについての様々な事前情報に依拠する効率的なデコーダを設計することが可能になることを示している。

30

【0 0 4 6】

C S - I S S デコーダ
ソース

$$j \in \llbracket 1, J \rrbracket$$

が時間インデックス

$$t \in \Omega_j'' \subseteq \llbracket 1, T \rrbracket$$

40

でサンプリングされるように、”によるランダム・サンプルの根拠を示す。エントロピ復号段階の後で、C S - I S S デコーダは、ソースの量子化済みサンプルの部分集合

$$y_{jt}''(\Omega_j''), j \in \llbracket 1, J \rrbracket$$

を有する。ここで、量子化済みサンプルは、

【数6】

$$y''_{jt} = s''_{jt} + b''_{jt} \quad (1)$$

として定義される。ここで、

$$s''_{jt}$$

は、真のソース信号を示し、

$$b''_{jt}$$

10

は、量子化された雑音である。

【0047】

なお、本明細書では、時間領域信号は、例えば x'' などツーダッシュ付きの文字で表すが、フレーム化またはウィンドウド・タイム領域 (windowed-time domain) 信号は、例えば x' などダッシュ付きの文字で表され、複素数値の短時間フーリエ変換 (STFT) 係数は、例えば x などダッシュなしの文字で表されることに留意されたい。

【0048】

混合物は、元のソースの和と仮定され、

【数7】

20

$$x''_t = \sum_{j=1}^J s''_{jt}, \quad t \in [1, T], j \in [1, J] \quad (2)$$

となる。

【0049】

混合物は、デコーダで既知であると仮定する。なお、本明細書では、混合物は、雑音がなく、量子化も行われていないものと仮定することに留意されたい。ただし、開示するアルゴリズムは、混合物に雑音を含むようにも容易に拡張することができる。

【0050】

STFT係数を計算するために、最初に、混合物およびソースを、ウィンドウ長 M を有し、合計 N 個のウィンドウを有する、ウィンドウド・タイム領域に変換する。 $y'_{j,m,n}$ 、 $s'_{j,m,n}$ および $x'_{m,n}$ で表されるその結果得られる係数は、 $j = 1, \dots, J$ 、 $n = 1, \dots, N$ 、および $m = 1, \dots, M$ (m のみについては、量子化済みソース・サンプルの場合には適当な部分集合 $'_{j,n}$ におけるもの) について、それぞれ、ウィンドウド・タイム領域における量子化済みソース、元のソース、および混合物を表す。ソースの STFT 係数 $s_{j,f,n}$ 、および混合物の STFT 係数 $x_{f,n}$ は、ウィンドウド・タイム領域の対応するものの各ウィンドウにユニタリ・フーリエ変換

$$U \in \mathbb{C}^{F \times M} \quad (F=M)$$

30

を適用することによって計算する。例えば、 $[x_{1,n}, \dots, x_{F,n}]^\top = U [x'_{1,n}, \dots, x'_{M,n}]^\top$ である。

40

【0051】

ソースは、正規分布 ($s_{j,f,n} \sim N_c(0, s_{j,f,n})$) によって STFT 領域でモデル化する。ここで、分散テンソル $V = [v_{j,f,n}]_{j,f,n}$ は、以下の低ランク NTF 構造 [18] を有する。

【数8】

$$v_{jfn} = \sum_{k=1}^K q_{jk} w_{fk} h_{nk}, K < \max(J, F, N) \quad (3)$$

【0052】

このモデルは、

$$Q = [q_{jk}] \in \mathbb{R}_+^{J \times K}, W = [w_{fk}] \in \mathbb{R}_+^{F \times K}$$

および

$$H = [h_{nk}] \in \mathbb{R}_+^{N \times K}$$

として、 $= \{ Q, W, H \}$ によってパラメータ化する。

【0053】

本原理の実施形態によれば、ソース信号は、アルゴリズム1に簡単に説明する汎用期待値最大化アルゴリズムで回復される。このアルゴリズムは、予想ステップでウィーナ・フィルタリングによって所与のモデル Ω を用いて観測値からソースおよびソース統計値を推定し、次いで、最大化ステップで事後ソース統計値を使用してモデルを更新する。このアルゴリズムの各ステップについての詳細を、以下に示す。

アルゴリズム1 NTFモデルを使用したCS-ISS復号のGEMアルゴリズム

- 1: 手続き CS-ISS DECODING(\mathbf{x}' , $\{\mathbf{y}'_j\}_1^J$, $\{\Omega'_j\}_1^J$, K)
- 2: 非負値 $\mathbf{Q}, \mathbf{W}, \mathbf{H}$ をランダムに初期化する
- 3: 反復
- 4: $\mathbf{Q}, \mathbf{W}, \mathbf{H}, \mathbf{x}', \{\mathbf{y}'_j\}_1^J, \{\Omega'_j\}_1^J$ を前提として、 \hat{s} (ソース)および
 \hat{P} (事後パワー・スペクトル)を推定する \triangleright Eステップ、セクション3.1参照
- 5: \hat{P} を前提として $\mathbf{Q}, \mathbf{W}, \mathbf{H}$ を更新する \triangleright Mステップ、セクション3.2参照
- 6: 収束基準が満たされるまで
- 7: 手手続き終了

【0054】

ソースの推定

基礎となる全ての分布がガウス分布であり、ソースと観測値の間の全ての関係が線形であるので、ソースは、モデル・パラメータ Q, W, H によって(3)で定義される共分散テンソル V を前提として、ウィーナ・フィルタ [17] によって最小平均二乗誤差 (MMSE) 的に推定することができる。

【0055】

n 次フレーム

$$\bar{o}'_n$$

についての観測データ・ベクトルを

$$\bar{o}'_n \triangleq [\bar{y}'_{1n}^T, \dots, \bar{y}'_{jn}^T, \bar{x}'_n^T]^T$$

と定義するものとする。ここで、

$$\bar{x}'_n \triangleq [x'_{1n}, \dots, x'_{Mn}]^T$$

および

$$\bar{y}'_{jn} \triangleq [y'_{jm_n}, m \in \Omega'_{jn}]^T$$

10

20

30

40

50

である。

【0 0 5 6】

対応する観測データ

$$\bar{o}'_n$$

およびN T F モデル を前提として、各ソース・フレーム s_{jn} の事後分散は、

$$s_{jn} | \bar{o}'_n; \Theta \sim N_c(\hat{s}_{jn}, \hat{\Sigma}_{s_{jn}s_{jn}})$$

と書くことができ、ここで

$$\hat{s}_{jn}$$

10

および

$$\hat{\Sigma}_{s_{jn}s_{jn}}$$

は、それぞれ、事後平均、および事後共分散行列である。これらはそれぞれ、ウィーナ・フィルタリングによって、

【数 9】

$$\hat{s}_{jn} = \Sigma_{\bar{o}'_n s_{jn}}^H \Sigma_{\bar{o}'_n \bar{o}'_n}^{-1} \bar{o}'_n, \quad (4) \quad 20$$

【数 10】

$$\hat{\Sigma}_{s_{jn}s_{jn}} = \Sigma_{s_{jn}s_{jn}} - \Sigma_{\bar{o}'_n s_{jn}}^H \Sigma_{\bar{o}'_n \bar{o}'_n}^{-1} \Sigma_{\bar{o}'_n s_{jn}}, \quad (5)$$

として計算することができ、これらは、以下の定義を前提としている。

【数 11】

$$\Sigma_{\bar{o}'_n \bar{o}'_n} = \begin{bmatrix} \Sigma_{\bar{y}'_{1n} \bar{y}'_{1n}} & \dots & \mathbf{0} & \Sigma_{\bar{x}'_n \bar{y}'_{1n}}^H \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \dots & \Sigma_{\bar{y}'_{Jn} \bar{y}'_{Jn}} & \Sigma_{\bar{x}'_n \bar{y}'_{Jn}}^H \\ \Sigma_{\bar{x}'_n \bar{y}'_{1n}} & \dots & \Sigma_{\bar{x}'_n \bar{y}'_{Jn}} & \Sigma_{\bar{x}'_n \bar{x}'_n} \end{bmatrix} \quad (6) \quad 30$$

【数 12】

$$\Sigma_{\bar{o}'_n s_{jn}} = \left[\underline{\mathbf{0}}_{S_1 \times F}^T, \Sigma_{\bar{y}'_{jn} s_{jn}}^T, \underline{\mathbf{0}}_{S_2 \times F}^T, \Sigma_{\bar{x}'_n s_{jn}}^T \right]^T, \quad 40$$

$$S_1 \triangleq \sum_{\hat{j}=1}^{j-1} |\Omega'_{\hat{j}n}|, \quad S_2 \triangleq \sum_{\hat{j}=j+1}^J |\Omega'_{\hat{j}n}| \quad (7)$$

、

【数13】

$$\Sigma_{\bar{s}_{jn} s_{jn}} = \text{diag} \left([v_{jfn}]_f \right) \quad (8)$$

【数14】

$$\Sigma_{\bar{y}'_{jn} \bar{y}'_{jn}} = \mathbf{U}(\Omega'_{jn})^H \text{diag} \left([v_{jfn}]_f \right) \mathbf{U}(\Omega'_{jn}) \quad (9)$$

10

【数15】

$$\Sigma_{\bar{y}'_{jn} s_{jn}} = \mathbf{U}(\Omega'_{jn})^H \text{diag} \left([v_{jfn}]_f \right) \quad (10)$$

【数16】

$$\Sigma_{\bar{x}'_n s_{jn}} = \mathbf{U}^H \text{diag} \left([v_{jfn}]_f \right) \quad (11)$$

20

【数17】

$$\Sigma_{\bar{x}'_n \bar{x}'_n} = \mathbf{U}^H \text{diag} \left(\left[\sum_j v_{jfn} \right]_f \right) \mathbf{U} \quad (12)$$

30

【数18】

$$\Sigma_{\bar{x}'_n \bar{y}'_{jn}} = \mathbf{U}^H \text{diag} \left([v_{jfn}]_f \right) \mathbf{U}(\Omega'_{jn}) \quad (13)$$

ここで、 $\mathbf{U}(\cdot)_{jn}$ は、 \cdot_{jn} のインデックスを有する \mathbf{U} の列の $F \times |\cdot|_{jn}|$ 行列である。

【0057】

従って、以下で述べるようにNTFモデルを更新するために使用される事後パワー・スペクトル

$$\hat{P} = [\hat{p}_{jfn}]$$

40

は、

【数19】

$$\hat{p}_{jfn} = \mathbb{E} \left[|s_{jfn}|^2 \mid \bar{o}'_n; \theta \right] = |\hat{s}_{jfn}|^2 - \hat{\Sigma}_{s_{jn} s_{jn}}(f, f) \quad (14)$$

として計算することができる。

【0058】

50

モデルの更新

N T F モデル・パラメータは、

$$D_{IS}(\tilde{P} || \mathbf{V}) = \sum_{j,f,n} d_{IS}(\tilde{p}_{jfn} || v_{jfn})$$

として定義される、推定ソース・パワー・スペクトル

$$\tilde{P}$$

の3価テンソルとN T F モデル近似Vの3価テンソルの間のIS発散 [15] を最小限に 10
する乗法的更新 (M U) 規則を使用して再推定することができる。ここで、

$$d_{IS}(x||y) = \frac{x}{y} - \log\left(\frac{x}{y}\right) - 1$$

は、IS発散であり、

$$\tilde{p}_{jfn}$$

および j, f, n は、(14) および (3) に指定される。その結果として、Q, W, H は、[18] に示される M U 規則で更新することができる。これらの M U 規則を数回繰り返して、モデル推定を改善することができる。 20

【 0 0 5 9 】

さらに、N T F / N M F モデルを使用したソース分離適用分野では、個々のソースについてのいくつかの事前情報を有する必要があることが多い。この情報は、ソースのいくつかのサンプルであってもよいし、あるいはどのソースがどの瞬間に「非アクティブ」であるかについての知識であってもよい。ただし、この情報が実施されるときには、その前に必ず、各ソースがいくつの成分で構成されるかをアルゴリズムが予め規定する必要がある。これは、モデル・パラメータ

$$W \in \mathbb{R}_+^{M \times K}, \quad H \in \mathbb{R}_+^{N \times K}, \quad Q \in \mathbb{R}_+^{J \times K}$$

30

を初期化して、Q および H の特定の部分がゼロに設定され、各成分を特定のソースに割り当てるようにすることによって実施されることが多い。一実施形態では、モデルの計算は、成分の総数 K を前提として、各ソースが手動ではなく自動的に成分に割り当てられるように修正される。これは、S T F T 領域モデル・パラメータを通してではなく時間領域サンプル (0 である時間領域サンプルを有するという制約がある) を通してソースの「無音」を実施し、モデル・パラメータの初期条件を、それらが自動的に調整されるように緩和することによって実現される。ソース成分分布 (Q で定義される) のスパース構造を実施するための別の修正も、上記の乗法的更新の数式をわずかに修正することによって可能である。これにより、ソースから成分に自動的に割り当てられるようになる。

【 0 0 6 0 】

従って、一実施形態では、ソースの無音期間の形態の副情報 I_s が存在するときには、行列 H および Q は自動的に決定される。副情報 I_s は、どのソースがどの期間に無音であるかについての情報を含むことができる。このような特定の情報の存在下では、N M F を利用する従来の方法は、既定の k_i 個の成分が各ソースに割り当てられるように H および Q を初期化することとなる。改善された解決策は、このような初期化の必要がなく、H および Q を学習するので、k_i を事前に知る必要がない。これは、1) 時間領域サンプルを入力として使用して、S T F T 領域の操作が必須ではなくなること、および 2) スパース構造を有するよう行列 Q を制約することによって可能となる。これは、上述のように Q についての乗法的更新の数式を修正することによって実現される。

【 0 0 6 1 】

40

50

結果

本発明の手法の性能を評価するために、異なる量子化レベル（16ビット、11ビット、6ビット、および1ビット）および異なるソースあたりのサンプリング・ビットレート（0.64、1.28、2.56、5.12、および10.24 kbps / ソース）で提案するCS-ISSを使用して、16 kHzの音楽信号の3つのソースを符号化し、次いで復号する。この例では、ランダム・サンプリング・パターンは既定であり、符号時にも復号時にも分かっているものと仮定する。量子化済みサンプルは、ゼロ平均ガウス分布を仮定して算術エンコーダを使用して切り捨て、圧縮する。デコーダ側では、算術デコーダに統いて、ガウス・ウインドウ関数およびK = 18に固定された成分数（すなわちソースあたり6成分）を用いて1024サンプル（64ms）の半重なりサイン・ウインドウを使用して計算したSTFTによるGEMアルゴリズムを50回繰り返して、量子化済みサンプルからソースを復号する。再構築されたサンプルの品質は、[19]に記載されるように、信号対歪み比（SDR）で測定する。その結果得られる復号信号の符号化ビットレートおよびSDRを、表1に示す。括弧内は、符号化サンプルの割合である。なお、表1中の圧縮レートは、予想されるエントロピ・コーディング段階の性能のばらつきにより、対応する生ビットレートとは異なる。

10

【表1】

サンプルあたりのビット	圧縮レート/SDR(保持されるサンプルの%)				
	生レート(kbps/ソース)				
	0.64	1.28	2.56	5.12	10.24
16ビット	0.50 / -1.64 (0.25)	1.00 / 4.28 (0.50)	2.00 / 9.54 (1.00)	4.01 / 16.17 (2.00)	8.00 / 21.87 (4.00)
11ビット	0.43 / 1.30 (0.36)	0.87 / 6.54 (0.73)	1.75 / 13.30 (1.45)	3.50 / 19.47 (2.91)	7.00 / 24.66 (5.82)
6ビット	0.27 / 4.17 (0.67)	0.54 / 7.62 (1.33)	1.08 / 12.09 (2.67)	2.18 / 14.55 (5.33)	4.37 / 16.55 (10.67)
1ビット	0.64 / -5.06 (4.00)	1.28 / -2.57 (8.00)	2.56 / 1.08 (16.00)	5.12 / 1.59 (32.00)	10.24 / 1.56 (64.00)

20

表1：エントロピ・コーディング前の異なる（一様な）量子化レベルおよび異なる生ビットレートについての、対応するSDR(dB)でのCS-ISSのエントロピ・コーディング段階後の最終ビットレート（単位はソースあたりのkbps）。保持されるサンプルの割合も、それぞれの場合について括弧内に示す。最良のレート／歪みの兼合いに対応する結果は、太字で示す。

30

【0062】

CS-ISSの性能を、[4]に示されるより複雑なエンコーダおよび単純なデコーダを用いる従来のISS手法と比較する。ISSアルゴリズムは、NTFモデル量子化および[5]と同様の符号化とともに使用される。すなわち、NTF係数は対数領域で一様に量子化され、様々なNTF行列の量子化ステップのサイズは、[5]の式（31）～（33）を用いて計算され、インデックスは、2状態混合ガウス・モデル（GMM）に基づいて算術コーダを使用して符号化される（[5]の図5参照）。この手法を、異なる量子化ステップのサイズおよび異なるNTF成分の数、すなわち $= 2^{-2}, 2^{-1}, 2^{-1}, \dots, 2^4$ およびK = 4, 6, ..., 30について評価する。その結果は、250回のモデル更新の反復によって生成される。CS-ISSおよび従来のISSの両方の性能を図4に示す。この図において、本発明のデコーダが一定数の成分を使用する（エンコーダが非常に単純であり、この値を計算しないため）のに対して、ISS手法では最適な成分数および量子化を使用することができるにもかかわらず、CS-ISSの方がISS手法より明らかに性能が優れている。この性能差は、ランダム・サンプリングした時間領域と低ランクNTF領域のインコヒーレンシとによりCS-ISSデコーダが実現する高い効率によるものである。また、ISS手法は、[5]に説明されるようにエンコーダの構造に忠実性が欠けていることにより、10dBのSDRを超えると実行することができない。時間的制約のために本明細書では[5]に示されるISSアルゴリズムと比較することができないが、その結果は、レート歪み性能は同様の挙動を呈することを示している。提案する手法は、その低複雑性エンコーダによって優位であり、従ってより良好なレート歪み

40

50

性能を有する他の I S S 手法と比較しても依然として有利である可能性があることに留意されたい。

【 0 0 6 3 】

表 1 および図 4 に示す C S - I S S の性能は、レートによって量子化レベルを変えることが好ましいことがあることを示している。16 ビットの量子化も 1 ビットの量子化も良好な成績であるようには見えないが、この性能は、16 ビットの量子化は、これよりはるかに高いビットレートが利用可能であるときには、他の方式より優れている可能性があることを示している。同様に、1 ビットなどより粗な量子化も、かなり低いビットレートを考慮すれば、有利である可能性がある。量子化の選択は、エンコーダにおいて、単純な参照テーブルを参照として用いて実行することができる。また、C S - I S S のエンコーダは非常に単純であるが、提案するデコーダはかなり複雑性が高く、通常は従来の I S S 方法のエンコーダよりも高いことにも留意しなければならない。ただし、これも、提案するデコーダにおいて例えばグラフィック処理ユニット (G P U) などの並列処理を用いてフレーム間のウィーナ・フィルタリングの独立性を活用することによって克服することができる。

10

【 0 0 6 4 】

開示する解決策は、通常は、低ランク・テンソル構造が、再構築された信号のパワー・スペクトログラムに現れるという事実をもたらす。

【 0 0 6 5 】

「有する (c o m p r i s e) 」という動詞およびその活用形を使用している場合、それは請求項に記載される要素またはステップ以外の要素またはステップの存在を排除するものではないことに留意されたい。さらに、要素の前に「 a 」または「 a n 」という冠詞を使用していても、その要素が複数存在することを排除するものではない。いくつかの「手段」が、同じハードウェアで表されることもある。さらに、本発明は、全ての新規の特徴または特徴の組合せにある。本明細書で使用する「ディジタル・オーディオ信号」または「オーディオ信号」は、単なる数学的な抽象概念について述べているのではなく、機械または装置によって検出することができる物理的媒体に埋め込まれる情報、あるいはそのような物理的媒体によって搬送される情報を指している。この用語は、記録または伝送される信号を含み、パルス符号変調 (P C M) などの (ただし P C M に限定されるわけではない) 、任意の形態の符号化による搬送を含むものとして理解されたい。

20

【 0 0 6 6 】

本発明の基本的な新規の特徴を本発明の好ましい実施形態に適用して示し、説明し、指摘したが、当業者なら、本発明の趣旨を逸脱することなく、記載した装置および方法において、開示したデバイスの形態および詳細ならびにそれらの動作に対する様々な省略、置換、および変更を行うことができるることを理解されたい。実質的に同じ機能を実質的に同じ方法で実行して同じ結果を実現する要素の全ての組合せは、本発明の範囲に含まれるものと明示的に意図されている。記載した 1 つの実施形態から別の実施形態への要素の置換も完全に意図および企図されている。本明細書ならびに (適当である場合には) 特許請求の範囲および図面に開示するそれぞれの特徴は、独立して提供することもできるし、あるいは任意の適当な組合せで提供することもできる。特徴は、適宜、ハードウェア、ソフトウェア、またはその両者の組合せで実施することができる。接続は、適用可能である場合には、無線接続として実施することも有線接続として実施することもでき、必ずしも直接または専用の接続であるとは限らない。

30

【 0 0 6 7 】

引用文献

- [1] E . V i n c e n t 、 S . A r a k i 、 F . J . T h e i s 、 G . N o l t e 、 P . B o f i l l 、 H . S a w a d a 、 A . O z e r o v 、 B . V . G o w r e e s u n k e r 、 D . L u t t e r 、 および N . Q . K . D u o n g による「 T h e s i g n a l s e p a r a t i o n e v a l u a t i o n c a m p a i g n (2 0 0 7 ~ 2 0 1 0) : A c h i e v e m e n t s a n d r e m a i n i n g c h a l l e n g e s

40

50

」、*Signal Processing*、vol. 92、no. 8、pp. 1928~1936、2012年。

[2] M. Parvaix、L. Girin、およびJ. M. Brossierによる「A water marking based method for informed source separation of audio signals with a single sensor」、*IEEE Trans. Audio, Speech, Language Process.*、vol. 18、no. 6、pp. 1464~1475、2010年。

[3] M. ParvaixおよびL. Girinによる「Informed source separation of linear instantaneous under-determined audio mixtures by source index embedding」、*IEEE Trans. Audio, Speech, Language Process.*、vol. 19、no. 6、pp. 1721~1733、2011年。
10

[4] A. Liutkus、J. Pinel、R. Badeau、L. Girin、およびG. Richardによる「Informed source separation through spectrogram coding and data embedding」、*Signal Processing*、vol. 92、no. 8、pp. 1937~1949、2012年。

[5] A. Ozerov、A. Liutkus、R. Badeau、およびG. Richardによる「Coding-based informed source separation: Nonnegative tensor factorization approach」、*IEEE Transactions on Audio, Speech, and Language Processing*、vol. 21、no. 8、pp. 1699~1712、2013年8月。
20

[6] J. Engdegard、B. Resch、C. Falch、O. Hellmut h、J. Hilpert、A. Holzer、L. Terentiev、J. Breebaart、J. Koppens、E. Schuijers、およびW. Oomenによる「Spatial audio object coding (SAOC) - The upcoming MPEG standard on parametric object based audio coding」、124th Audio Engineering Society Convention (AES 2008)、オランダ、アムステルダム、2008年5月。
30

[7] A. Ozerov、A. Liutkus、R. Badeau、およびG. Richardによる「Informed source separation: source coding meets source separation」、*IEEE Workshop Applications of Signal Processing to Audio and Acoustics (WASPAA'11)*、米国ニューヨーク州ニューパルツ、2011年10月、pp. 257~260。

[8] S. Kirbizi、A. Ozerov、A. Liutkus、およびL. Girinによる「Perceptual coding-based informed source separation」、*Proc. 22nd European Signal Processing Conference (EUSIPCO)*、2014年、pp. 959~963。
40

[9] Z. Xiong、A. D. Liveris、およびS. Chengによる「Distributed source coding for sensor networks」、*IEEE Signal Processing Magazine*、vol. 21、no. 5、pp. 80~94、2004年9月。

[10] B. Girod、A. Aaron、S. Rane、およびD. Rebolledo-Monederoによる「Distributed video coding」*Pro*
50

ceedings of the IEEE、vol. 93、no. 1、pp. 71~83、2005年1月。

[11] D. Donohoによる「Compressed sensing」、IEEE Trans. Inform. Theory、vol. 52、no. 4、pp. 1289~1306、2006年4月。

[12] R. G. Baraniukによる「Compressive sensing」、IEEE Signal Processing Mag、vol. 24、no. 4、pp. 118~120、2007年7月。

[13] E. J. CandèsおよびM. B. Wakinによる「An introduction to compressive sampling」、IEEE Signal Processing Magazine、vol. 25、pp. 21~30、2008年。
10

[14] R. G. Baraniuk、V. Cevher、M. F. Duarte、およびC. Hegdeによる「Model-based compressive sensing」、IEEE Trans. Info. Theory、vol. 56、no. 4、pp. 1982~2001、2010年4月。

[15] C. Fevotte、N. Bertin、およびJ. L. Durrieuによる「Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis」、Neural Computation、vol. 21、no. 3、pp. 793~830、2009年3月。
20

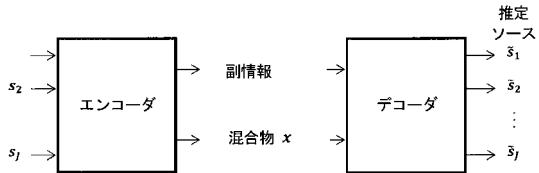
[16] A. P. Dempster、N. M. Laird、およびD. B. Rubinによる「Maximum likelihood from incomplete data via the EM algorithm」、Journal of the Royal Statistical Society. Series B (Methodological)、vol. 39、pp. 1~38、1977年。

[17] S. M. Kayによる「Fundamentals of Statistical Signal Processing: Estimation Theory」米国ニュージャージー州エングルウッド・クリフス、Prentice Hall、1993年。
30

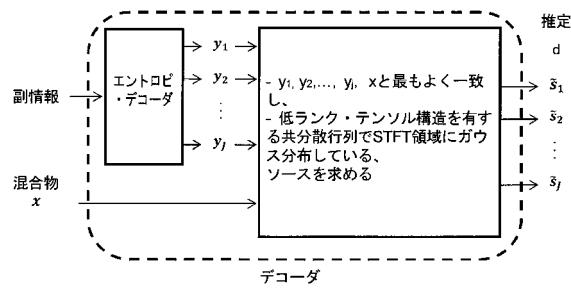
[18] A. Ozerov、C. Fevotte、R. Blouet、およびJ.-L. Durrieuによる「Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation」、IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'11)、プラハ、2011年5月、pp. 257~260。

[19] V. Emiya、E. Vincent、N. Harlander、およびV. Ohmannによる「Subjective and objective quality assessment of audio source separation」、IEEE Trans. Audio, Speech, Language Process、vol. 19、no. 7、pp. 2046~2057、2011年。
40

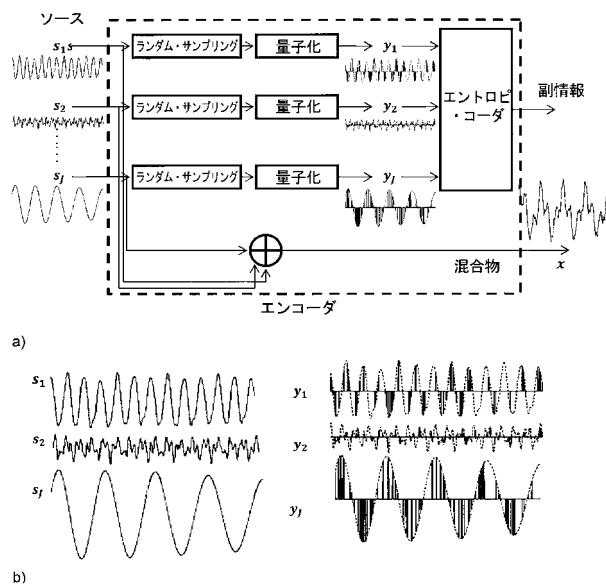
【図1】



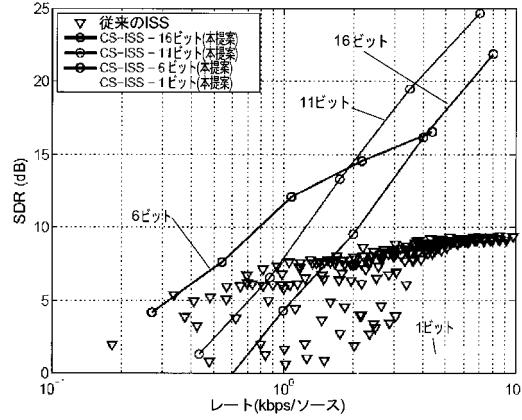
【図3】



【図2】



【図4】



【手続補正書】

【提出日】平成29年12月20日(2017.12.20)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】0 0 0 3

【補正方法】変更

【補正の内容】

【0 0 0 3】

既知の解決策（例えば、[3]、[4]、[5]、[20]、[21]）は、元のソースが符号化段階で利用可能であるという前提に立っている。副情報を計算して混合物とともに伝送し、復号段階ではその両方を処理して、ソースを回復する。

例えば、文書[21]では、マグニチュード・スペクトログラムを非負値行列分解としてその原子単位（原子と呼ぶ）に分解する合成モデリング（compositional modeling）が特に開示されている。従って、アクティベーション・ベクトルと原子で構成される行列との組合せから、スペクトル・ベクトルを得ることができる。特定の実施態様として、スペース表現および圧縮センシングの分野で利用される辞書学習技術が、スペース表現を有する辞書を発見するために活用される。これにより、復号側で依拠することができるモデルが得られる。

いくつかのISS方法が既知であるが、それらの手法では全て、符号化段階は、復号段階より複雑で計算コストが高い。従って、これらの手法は、符号化を実行するプラットフォームが、エンコーダが要求する計算の複雑さを取り扱うことができない場合には、好ましくない。最後に、既知の複雑なエンコーダは、いくつかの適用分野で非常に重要なオンライン符号化、すなわち信号が到着するにつれて順次符号化していく符号化には使用できない。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0011

【補正方法】変更

【補正の内容】

【0011】

一実施形態では、本発明の原理は、オーディオ・ソースを分離する符号化デバイスであって、例えばハードウェア・プロセッサ、および少なくとも1つのソフトウェア構成要素を有形に実施する非一時的有形コンピュータ可読記憶媒体など、少なくとも1つのハードウェア構成要素を含む符号化デバイスに関し、少なくとも1つのハードウェア・プロセッサ上で実行されるときに、このソフトウェア構成要素は、以下で述べる符号化方法のステップを引き起こす。一実施形態では、本発明の原理は、オーディオ・ソースを分離する符号化デバイスであって、例えばハードウェア・プロセッサ、および少なくとも1つのソフトウェア構成要素を有形に実施する非一時的有形コンピュータ可読記憶媒体など、少なくとも1つのハードウェア構成要素を含む符号化デバイスに関し、少なくとも1つのハードウェア・プロセッサ上で実行されるときに、このソフトウェア構成要素は、以下で述べる復号方法のステップを引き起こす。

既存の解決策に対して驚くべきことに、符号化は、モデルおよび／またはフーリエ変換計算を用いるのではなく、時間領域オーディオ信号のランダム・サンプリングに依拠する。これは、潜在的には、場合によっては特定の実施態様では非常に限られた処理しか必要としない可能性がある非常に高速な符号化プロセスをもたらす可能性がある。この予期しない時間領域ランダム・サンプリングにもかかわらず、オーディオ信号を復号側でそれらの混合物から再構築することが可能になる。

【手続補正3】

【補正対象書類名】明細書

【補正対象項目名】0066

【補正方法】変更

【補正の内容】

【0066】

特徴は、適宜、ハードウェア、ソフトウェア、またはその両者の組合せで実施することができる。接続は、適用可能である場合には、無線接続として実施することも有線接続として実施することもでき、必ずしも直接または専用の接続であるとは限らない。

【手続補正4】

【補正対象書類名】明細書

【補正対象項目名】0067

【補正方法】変更

【補正の内容】

【0067】

引用文献

[1] E. Vincent、S. Araki、F. J. Theis、G. Nolte、P. Bofill、H. Sawada、A. Ozerov、B. V. Gowreesunker、D. Lutter、およびN. Q. K. Duongによる「The signal separation evaluation campaign (2007~2010) : Achievements and remaining challenges」、Signal Processing, vol. 92, no. 8, pp. 1928~1936, 2012年。

[2] M. Parvaix、L. Girin、およびJ. M. Brossierによる「A water marking based method for informed source separation of audio signals with a single sensor」、IEEE Trans. Audio, Speech, Language Process., vol. 18, no. 6, pp. 1464~

1475、2010年。

[3] M. Parvaix および L. Girin による「Informed source separation of linear instantaneous under-determined audio mixtures by source index embedding」、IEEE Trans. Audio, Speech, Language Process., vol. 19, no. 6, pp. 1721~1733, 2011年。

[4] A. Liutkus, J. Pinel, R. Badeau, L. Girin, および G. Richard による「Informed source separation through spectrogram coding and data embedding」、Signal Processing, vol. 92, no. 8, pp. 1937~1949, 2012年。

[5] A. Ozerov, A. Liutkus, R. Badeau, および G. Richard による「Coding-based informed source separation: Nonnegative tensor factorization approach」、IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 8, pp. 1699~1712, 2013年8月。

[6] J. Engdegard, B. Resch, C. Falch, O. Hellmuth, J. Hilpert, A. Holzer, L. Terentiev, J. Breebaart, J. Koppens, E. Schuijers, および W. Oomen による「Spatial audio object coding (SAOC) - The upcoming MPEG standard on parametric object based audio coding」、124th Audio Engineering Society Convention (AES 2008)、オランダ、アムステルダム、2008年5月。

[7] A. Ozerov, A. Liutkus, R. Badeau, および G. Richard による「Informed source separation: source coding meets source separation」、IEEE Workshop Applications of Signal Processing to Audio and Acoustics (WASPAA'11)、米国ニューヨーク州ニューパルツ、2011年10月、pp. 257~260。

[8] S. Kirbiz, A. Ozerov, A. Liutkus, および L. Girin による「Perceptual coding-based informed source separation」、Proc. 22nd European Signal Processing Conference (EUSIPCO)、2014年、pp. 959~963。

[9] Z. Xiong, A. D. Liveris, および S. Cheng による「Distributed source coding for sensor networks」、IEEE Signal Processing Magazine, vol. 21, no. 5, pp. 80~94, 2004年9月。

[10] B. Girod, A. Aaron, S. Rane, および D. Rebolledo-Monedero による「Distributed video coding」 Proceedings of the IEEE, vol. 93, no. 1, pp. 71~83, 2005年1月。

[11] D. Donoho による「Compressed sensing」、IEEE Trans. Inform. Theory, vol. 52, no. 4, pp. 1289~1306, 2006年4月。

[12] R. G. Baraniuk による「Compressive sensing」、IEEE Signal Processing Mag., vol. 24, no. 4,

pp. 118~120、2007年7月。

[13] E. J. Candès および M. B. Wakin による「An introduction to compressive sampling」、IEEE Signal Processing Magazine、vol. 25、pp. 21~30、2008年。

[14] R. G. Baraniuk、V. Cevher、M. F. Duarte、および C. Hegde による「Model-based compressive sensing」、IEEE Trans. Info. Theory、vol. 56、no. 4、pp. 1982~2001、2010年4月。

[15] C. Fevotte、N. Bertin、および J. L. Durrieu による「Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis」、Neural Computation、vol. 21、no. 3、pp. 793~830、2009年3月。

[16] A. P. Dempster、N. M. Laird、および D. B. Rubin による「Maximum likelihood from incomplete data via the EM algorithm」、Journal of the Royal Statistical Society. Series B (Methodological)、vol. 39、pp. 1~38、1977年。

[17] S. M. Kay による「Fundamentals of Statistical Signal Processing: Estimation Theory」、米国ニュージャージー州エングルウッド・クリフス、Prentice Hall、1993年。

[18] A. Ozerov、C. Fevotte、R. Blouet、および J.-L. Durrieu による「Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation」、IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'11)、プラハ、2011年5月、pp. 257~260。

[19] V. Emiya、E. Vincent、N. Harlander、および V. Hohmann による「Subjective and objective quality assessment of audio source separation」、IEEE Trans. Audio, Speech, Language Process、vol. 19、no. 7、pp. 2046~2057、2011年。

[20] J. Nikunen、T. Virtanen、および M. Välimäki による「Multichannel audio upmixing by time-frequency filtering using non-negative tensor factorization」、J. Audio Eng. Soc.、vol. 60、no. 10、pp. 794~806、2012年。

[21] T. Virtanen、J. F. Gemmeke、B. Raj、および P. Smaragdis による「Compositional models for audio processing」、IEEE Signal Processing Magazine、pp. 125~144、2015年。

【手続補正5】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項 1】

複数の時間領域オーディオ信号を、前記複数の時間領域オーディオ信号の混合物から前記複数の時間領域オーディオ信号を復号および分離するために使用することができる副情報として符号化する方法であって、

前記複数の時間領域オーディオ信号の各々をランダム・サンプリングおよび量子化することと、

前記サンプリングおよび量子化した複数の時間領域オーディオ信号を、前記副情報として符号化することと、
を含む、前記方法。

【請求項 2】

前記ランダム・サンプリングは、既定の疑似ランダム・パターンを使用する、請求項1に記載の方法。

【請求項 3】

前記複数の時間領域オーディオ信号の混合物は、到着するにつれて順次符号化される、請求項1に記載の方法。

【請求項 4】

どのソースがどの期間に無音であるかを割り出すステップと、前記割り出した情報を前記副情報内に符号化するステップとをさらに含む、請求項1に記載の方法。

【請求項 5】

複数のオーディオ信号の混合物を復号する方法であって、

記憶装置または任意のデータ・ソースから、前記複数のオーディオ信号の混合物を受信する、または取り出すことと、

前記複数のオーディオ信号に近似した複数の推定オーディオ信号を、前記複数のオーディオ信号の混合物に関連付けられた副情報から生成することと、

を含み、前記方法は、

前記複数のオーディオ信号の各々のランダム・サンプリングされ量子化された時間領域サンプルを含む前記副情報を復号および逆多重化することと、

前記複数のオーディオ信号の各々の前記量子化されたサンプルを使用して、前記複数の推定オーディオ信号を生成することとを含む、前記方法。

【請求項 6】

前記複数の推定オーディオ信号を生成することは、

ランダム非負値から分散テンソルVを計算することと、

前記複数のオーディオ信号の前記量子化されたサンプルのソース・パワー・スペクトルの条件付き期待値を計算することであって、推定ソース・パワー・スペクトルP(f, n, j)が得られ、かつ、前記複数のオーディオ信号の前記分散テンソルVおよび複素短時間フーリエ変換(STFT)係数が使用される、前記計算することと、

前記推定ソース・パワー・スペクトルP(f, n, j)から前記分散テンソルVを繰り返し再計算することと、

結果として得られる分散テンソルVから、STFT係数

\hat{S}

の配列を計算することと、

STFT係数

\hat{S}

の前記配列を前記時間領域に変換することであって、前記複数の推定オーディオ信号が得られる、前記変換することと、

を含む、請求項5に記載の方法。

【請求項 7】

前記複数のオーディオ信号のうちの少なくとも1つについてオーディオ・インペインテ

イングを行うことをさらに含む、請求項5に記載の方法。

【請求項 8】

前記副情報が、どのオーディオ・ソースがどの時間に無音であるかを示す情報をさらに含み、前記分散テンソルVを定義する行列HおよびQを自動的に定めることをさらに含む、請求項5に記載の方法。

【請求項 9】

複数のオーディオ信号を、複数の時間領域オーディオ信号の混合物から前記複数の時間領域オーディオ信号を復号および分離するために使用することができる副情報として符号化する装置であって、

複数の時間領域オーディオ信号を符号化する方法を前記装置に実行させるように構成された少なくとも1つのプロセッサを有し、

前記少なくとも1つのプロセッサが前記装置に、

前記複数の時間領域オーディオ信号の各々をランダム・サンプリングおよび量子化することと、

前記サンプリングおよび量子化した複数の時間領域オーディオ信号を、前記副情報として符号化することと、

を実行させるように構成されていることを特徴とする、前記装置。

【請求項 10】

前記ランダム・サンプリングは、既定の疑似ランダム・パターンを使用する、請求項9に記載の装置。

【請求項 11】

複数のオーディオ信号の混合物を復号する装置であって、

複数のオーディオ信号の混合物を復号する方法を前記装置に実行させるように構成された少なくとも1つのプロセッサを有し、前記方法は、

記憶装置または任意のデータ・ソースから、前記複数のオーディオ信号の混合物を受信する、または取り出すことと、

前記複数のオーディオ信号に近似した複数の推定オーディオ信号を、前記複数のオーディオ信号の混合物に関連付けられた副情報から生成することと、
を含み、

前記少なくとも1つのプロセッサが、

前記複数のオーディオ信号の各々のランダム・サンプリングされ量子化された時間領域サンプルを含む前記副情報を復号および逆多重化し、

前記複数のオーディオ信号の各々の前記量子化されたサンプルを使用して、前記複数の推定オーディオ信号を生成するように構成されていることを特徴とする、前記装置。

【請求項 12】

前記複数の推定オーディオ信号を生成することは、

ランダム非負値から分散テンソルVを計算することと、

前記複数のオーディオ信号の前記量子化されたサンプルのソース・パワー・スペクトルの条件付き期待値を計算することであって、推定ソース・パワー・スペクトルP(f, n, j)が得られ、かつ、前記複数のオーディオ信号の前記分散テンソルVおよび複素短時間フーリエ変換(STFT)係数が使用される、前記計算することと、

前記推定ソース・パワー・スペクトルP(f, n, j)から前記分散テンソルVを繰り返し再計算することと、

結果として得られる分散テンソルVからSTFT係数

\hat{S}

の配列を計算することと、

STFT係数

\hat{S}

の前記配列を前記時間領域に変換することであって、前記複数の推定オーディオ信号が得られる、前記変換することと、
を含む、請求項 1 1 に記載の装置。

【請求項 1 3】

前記少なくとも 1 つのプロセッサが、前記複数の時間領域オーディオ信号のうちの少なくとも 1 つについてオーディオ・インペインティングを行うようにさらに構成されている
、請求項 1 1 に記載の装置。

【請求項 1 4】

複数の時間領域オーディオ信号を、前記複数の時間領域オーディオ信号の混合物から前記複数の時間領域オーディオ信号を復号および分離するために使用することができる副情報として符号化する方法を実施するためにコンピュータによって実行可能な命令のプログラムを有形に含む、コンピュータ可読の非一時的プログラム記憶デバイスであって、前記方法は、

前記複数の時間領域オーディオ信号の各々をランダム・サンプリングおよび量子化することと、

前記サンプリングおよび量子化した複数の時間領域オーディオ信号を前記副情報として符号化することとを含む、前記非一時的プログラム記憶デバイス。

【請求項 1 5】

複数のオーディオ信号の混合物を復号する方法を実施するためにコンピュータによって実行可能な命令のプログラムを有形に含む、コンピュータ可読の非一時的プログラム記憶デバイスであって、前記方法は、

記憶装置または任意のデータ・ソースから、複数のオーディオ信号の混合物を受信する、または取り出すことと、

前記複数のオーディオ信号の混合物に関連付けられた副情報から前記複数のオーディオ信号に近似した複数の推定オーディオ信号を生成することとを含み、

前記方法は、

前記複数のオーディオ信号の各々のランダム・サンプリングされ量子化された時間領域サンプルを含む前記副情報を復号および逆多重化することと、

前記複数のオーディオ信号の各々の前記量子化されたサンプルを使用して、前記複数の推定オーディオ信号を生成することとを含む、前記非一時的プログラム記憶デバイス。

【請求項 1 6】

複数のオーディオ信号の混合物を復号するように構成された副情報を含む信号を有形に含む記憶媒体であって、

前記副情報は、前記複数のオーディオ信号の各々のランダム・サンプリングされ量子化された時間領域サンプルを含む、前記記憶媒体。

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International application No PCT/EP2016/055135									
A. CLASSIFICATION OF SUBJECT MATTER INV. G10L19/008 G10L21/0272 ADD. H03M1/12											
According to International Patent Classification (IPC) or to both national classification and IPC											
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G10L H03M											
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched											
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, WPI Data											
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left; padding: 2px;">Category*</th> <th style="text-align: left; padding: 2px;">Citation of document, with indication, where appropriate, of the relevant passages</th> <th style="text-align: left; padding: 2px;">Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td style="padding: 2px;">Y</td> <td style="padding: 2px;">OZEROV A ET AL: "Coding-Based Informed Source Separation: Nonnegative Tensor Factorization Approach", IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, USA, vol. 21, no. 8, August 2013 (2013-08), pages 1699-1712, XP011519779, ISSN: 1558-7916, DOI: 10.1109/TASL.2013.2260153 cited in the application page 1700, paragraph II. - page 1706, paragraph III.E; figure 3 page 1709, paragraph V. - page 1710</td> <td style="padding: 2px;">1-3,5-7, 9-13</td> </tr> <tr> <td style="padding: 2px;">A</td> <td style="padding: 2px;">----- -/-</td> <td style="padding: 2px;">4,8</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	Y	OZEROV A ET AL: "Coding-Based Informed Source Separation: Nonnegative Tensor Factorization Approach", IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, USA, vol. 21, no. 8, August 2013 (2013-08), pages 1699-1712, XP011519779, ISSN: 1558-7916, DOI: 10.1109/TASL.2013.2260153 cited in the application page 1700, paragraph II. - page 1706, paragraph III.E; figure 3 page 1709, paragraph V. - page 1710	1-3,5-7, 9-13	A	----- -/-	4,8
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.									
Y	OZEROV A ET AL: "Coding-Based Informed Source Separation: Nonnegative Tensor Factorization Approach", IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING, IEEE SERVICE CENTER, NEW YORK, NY, USA, vol. 21, no. 8, August 2013 (2013-08), pages 1699-1712, XP011519779, ISSN: 1558-7916, DOI: 10.1109/TASL.2013.2260153 cited in the application page 1700, paragraph II. - page 1706, paragraph III.E; figure 3 page 1709, paragraph V. - page 1710	1-3,5-7, 9-13									
A	----- -/-	4,8									
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C.		<input checked="" type="checkbox"/> See patent family annex.									
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed											
T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family											
Date of the actual completion of the international search 19 May 2016	Date of mailing of the international search report 03/06/2016										
Name and mailing address of the ISA/ European Patent Office, P.O. 5818 Patentlaan 2 NL-2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Virette, David										

INTERNATIONAL SEARCH REPORT

International application No PCT/EP2016/055135

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	NIKUNEN JOONAS ET AL: "Multichannel Audio Upmixing by Time-Frequency Filtering Using Non-Negative Tensor Factorization", JAES, AES, 60 EAST 42ND STREET, ROOM 2520 NEW YORK 10165-2520, USA, vol. 60, no. 10, 1 October 2012 (2012-10-01), pages 794-806, XP040574862, page 796, paragraph 1 - page 800 page 803, paragraph 4 - page 804	1-3,5,6, 9-12
A	VIRTANEN TUOMAS ET AL: "Compositional Models for Audio Processing: Uncovering the structure of sound mixtures", IEEE SIGNAL PROCESSING MAGAZINE, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 32, no. 2, March 2015 (2015-03), pages 125-144, XP011573080, ISSN: 1053-5888, DOI: 10.1109/MSP.2013.2288990 [retrieved on 2015-02-10]	4,7,8,13
A	page 125, line 9 - page 126, left-hand column, line 10 page 132, paragraph Source separation - page 134, paragraph Dictionary creation	4,8
Y	CANDES E J ET AL: "An Introduction To Compressive Sampling", IEEE SIGNAL PROCESSING MAGAZINE, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 25, no. 2, March 2008 (2008-03), pages 21-30, XP011225660, ISSN: 1053-5888, DOI: 10.1109/MSP.2007.914731 cited in the application page 29, paragraph 1) Nonuniform Sampler (NUS) - paragraph 2) Random Modulation Preintegration (RMP1)	2,10
A	US 2014/297294 A1 (KIM DONG SOO [KR] ET AL) 2 October 2014 (2014-10-02) page 1, paragraph [0003] - paragraph [0005] page 2, paragraph [0033] page 3, paragraph [0047] - paragraph [0049]	4,8
1		

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2016/055135

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014297294	A1 02-10-2014	AT 526659 T	15-10-2011
		AU 2008215230 A1	21-08-2008
		AU 2008215231 A1	21-08-2008
		AU 2008215232 A1	21-08-2008
		BR PI0802613 A2	30-08-2011
		BR PI0802614 A2	30-08-2011
		CA 2645912 A1	21-08-2008
		CA 2645913 A1	21-08-2008
		CA 2645915 A1	21-08-2008
		EP 2111616 A1	28-10-2009
		EP 2111617 A1	28-10-2009
		EP 2115739 A1	11-11-2009
		JP 5232795 B2	10-07-2013
		JP 5254983 B2	07-08-2013
		JP 5291227 B2	18-09-2013
		JP 2010506231 A	25-02-2010
		JP 2010506232 A	25-02-2010
		JP 2010508545 A	18-03-2010
		JP 2012198556 A	18-10-2012
		KR 20090030323 A	24-03-2009
		KR 20090082339 A	30-07-2009
		KR 20090082340 A	30-07-2009
		TW 200847136 A	01-12-2008
		TW 200907932 A	16-02-2009
		TW 200921642 A	16-05-2009
		US 2009210238 A1	20-08-2009
		US 2009326958 A1	31-12-2009
		US 2010076772 A1	25-03-2010
		US 2011200197 A1	18-08-2011
		US 2011202356 A1	18-08-2011
		US 2011202357 A1	18-08-2011
		US 2014297294 A1	02-10-2014
		WO 2008100098 A1	21-08-2008
		WO 2008100099 A1	21-08-2008
		WO 2008100100 A1	21-08-2008

フロントページの続き

(81)指定国 AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,MK,MT,NL,NO,PL,PT,R0,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DK,DM,D0,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,JP,KE,KG,KN,KP,KR,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,SM,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US

(74)代理人 100134094

弁理士 倉持 誠

(74)代理人 100123629

弁理士 吹田 礼子

(72)発明者 ビレン,カグダ

フランス国 エフ- 35576 セソン セビニエ セーエス 176 16 ゼドアーセー・デ
・シャン・プラン アベニユー・デ・シャン・プラン 975 テクニカラー・アール・アンド・
デイー フランス

(72)発明者 オゼロブ,アレクセイ

フランス国 エフ- 35576 セソン セビニエ セーエス 176 16 ゼドアーセー・デ
・シャン・プラン アベニユー・デ・シャン・プラン 975 テクニカラー・アール・アンド・
デイー フランス

(72)発明者 ペレス,パトリツク

フランス国 エフ- 35576 セソン セビニエ セーエス 176 16 ゼドアーセー・デ
・シャン・プラン アベニユー・デ・シャン・プラン 975 テクニカラー・アール・アンド・
デイー フランス