

(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2021-33466
(P2021-33466A)

(43) 公開日 令和3年3月1日(2021.3.1)

(51) Int.Cl.
G06N 3/02 (2006.01)

F I
G06N 3/02

テーマコード (参考)

審査請求 未請求 請求項の数 16 O L (全 26 頁)

(21) 出願番号 特願2019-150516 (P2019-150516)
(22) 出願日 令和1年8月20日 (2019.8.20)

(71) 出願人 504133110
国立大学法人電気通信大学
東京都調布市調布ヶ丘一丁目5番地1
(74) 代理人 100121131
弁理士 西川 孝
(74) 代理人 100082131
弁理士 稲本 義雄
(72) 発明者 中鹿 亘
東京都調布市調布ヶ丘一丁目5番地1 国立大学法人電気通信大学内

(54) 【発明の名称】 符号化装置、復号装置、パラメータ学習装置、およびプログラム

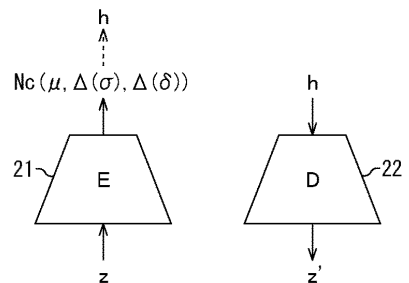
(57) 【要約】

【課題】 より高品位に符号化および復号する。

【解決手段】 符号化部は、複素数の観測データを、勾配法を利用した学習により求められた複素数のパラメータが設定されたニューラルネットワークである複素ニューラルネットワークを用いて複素数のまま符号化し、複素数の潜在変数の分布を符号化データとして出力する。復号部は、複素数の観測データが符号化された符号化データを、勾配法を利用した学習により求められた複素数のパラメータが設定されたニューラルネットワークである複素ニューラルネットワークを用いて複素数のまま復号して得られる復号データとして、再構築された前記複素数の観測データを出力する。本技術は、例えば、音声データや画像データなどの複素数データの符号化および復号に適用できる。

【選択図】 図3

FIG. 3



【特許請求の範囲】

【請求項 1】

複素数の観測データを、勾配法を利用した学習により求められた複素数のパラメータが設定されたニューラルネットワークである複素ニューラルネットワークを用いて複素数のまま符号化し、複素数の潜在変数の分布を符号化データとして出力する符号化部を備える符号化装置。

【請求項 2】

前記複素ニューラルネットワークは、確率分布を使って潜在変数を求める生成モデルであるVAE (variational autoencoder) に基づいて、複素数データをそのまま扱えるように表現された複素VAEである

10

請求項 1 に記載の符号化装置。

【請求項 3】

前記複素ニューラルネットワークの出力層となるノードが、前記複素数の潜在変数の分布の平均を出力する平均出力部、前記複素数の潜在変数の分布の分散を出力する分散出力部、および、前記複素数の潜在変数の分布の擬似分散を出力する擬似分散出力部により構成される

請求項 2 に記載の符号化装置。

【請求項 4】

前記符号化部は、前記複素数の観測データを z とし、前記複素数のパラメータを θ とし、前記平均を μ とし、前記分散を σ とし、前記擬似分散を δ とし、前記複素数の潜在変数を h としたときに、次の式で表される複素正規分布 $p(h|z)$ を前記符号化データとして求める

20

【数 1】

$$p_{\phi}(h|z) = N_c(h; \mu, \Delta(\sigma), \Delta(\delta))$$

請求項 3 に記載の符号化装置。

【請求項 5】

前記複素数の観測データを取得する複素数データ取得部と、前記複素数の観測データを単位時間ごとに切り出してスペクトル特徴量を算出し、そのスペクトル特徴量を正規化して前記符号化部に供給する前処理部と、前記複素数のパラメータを保持し、前記符号化部に設定するパラメータ保持部とをさらに備える請求項 1 に記載の符号化装置。

30

【請求項 6】

符号化装置のコンピュータに、複素数の観測データを、勾配法を利用した学習により求められた複素数のパラメータが設定されたニューラルネットワークである複素ニューラルネットワークを用いて複素数のまま符号化し、複素数の潜在変数の分布を符号化データとして出力することを含む符号化処理を実行させるためのプログラム。

【請求項 7】

複素数の観測データが符号化された符号化データである複素数の潜在変数を、勾配法を利用した学習により求められた複素数のパラメータが設定されたニューラルネットワークである複素ニューラルネットワークを用いて複素数のまま復号して得られる復号データとして、再構築された前記複素数の観測データを出力する復号部を備える復号装置。

40

【請求項 8】

前記複素ニューラルネットワークは、確率分布を使って潜在変数を求める生成モデルであるVAE (variational autoencoder) に基づいて、複素数データをそのまま扱えるように表現された複素VAEである

請求項 7 に記載の復号装置。

【請求項 9】

50

前記複素ニューラルネットワークの出力層となるノードが、複素平均を出力する複素平均出力部、分散行列を出力する分散行列出力部、および、擬似分散行列を出力する擬似分散行列出力部により構成される

請求項 8 に記載の復号装置。

【請求項 10】

前記復号部は、前記複素数の観測データを z とし、前記複素数のパラメータを θ とし、前記複素平均を a とし、分散行列を Γ とし、擬似分散行列を C とし、前記複素数の潜在変数を h としたときに、次の式で表される多変量複素正規分布 $p_{\theta}(z|h)$ を前記復号データとして求める

【数 2】

$$p_{\theta}(z|h) = N_c(z; a, \Gamma, C)$$

$$= \frac{1}{\pi^D \sqrt{\det(\Gamma) \det(\bar{\Gamma} - C^H \Gamma^{-1} C)}} \cdot \exp \left\{ -\frac{1}{2} \begin{bmatrix} z-a \\ \bar{z}-\bar{a} \end{bmatrix}^H \begin{bmatrix} \Gamma & C \\ C^H & \Gamma^H \end{bmatrix}^{-1} \begin{bmatrix} z-a \\ \bar{z}-\bar{a} \end{bmatrix} \right\}$$

請求項 9 に記載の復号装置。

【請求項 11】

前記復号部から出力される前記復号データに対して、単位時間ごとに正規化されたスペクトル特徴量を元に戻す処理を施す後処理部と、

前記後処理部において後処理が施された前記復号データを出力する出力部と、

前記複素数のパラメータを保持し、前記復号部に設定するパラメータ保持部とをさらに備える請求項 7 に記載の復号装置。

【請求項 12】

復号装置のコンピュータに、

複素数の観測データが符号化された符号化データを、勾配法を利用した学習により求められた複素数のパラメータが設定されたニューラルネットワークである複素ニューラルネットワークを用いて複素数のまま復号して得られる復号データとして、再構築された前記複素数の観測データを出力すること

を含む復号処理を実行させるためのプログラム。

【請求項 13】

複素数の観測データを、第 1 の複素数のパラメータが設定されたニューラルネットワークである第 1 の複素ニューラルネットワークを用いて複素数のまま符号化し、複素数の潜在変数の分布を符号化データとして出力する符号化部と、

前記符号化データを、第 2 の複素数のパラメータが設定されたニューラルネットワークである第 2 の複素ニューラルネットワークを用いて複素数のまま復号して得られる復号データとして、再構築された前記複素数の観測データを出力する復号部と

を備え、

前記第 1 の複素数のパラメータおよび前記第 2 の複素数のパラメータを、勾配法を利用した学習により求める

パラメータ学習装置。

【請求項 14】

前記符号化部により符号化された前記符号化データについて、誤差を逆伝播させるための補助演算を行う補助演算部と、

前記補助演算部から出力される複素数の潜在変数の分布に対するサンプリングを行い、複素数の潜在変数のサンプルを取得して、前記復号部に供給するサンプリング部と

をさらに備える請求項 13 に記載のパラメータ学習装置。

10

20

30

40

50

【請求項 15】

前記第1の複素数のパラメータおよび前記第2の複素数のパラメータが、所定の終了条件を満足したか否かを判定する判定部と、

前記判定部により所定の終了条件を満足していないと判定された場合に、前記第1の複素数のパラメータおよび前記第2の複素数のパラメータを更新するパラメータ更新部とをさらに備える請求項13に記載のパラメータ学習装置。

【請求項 16】

パラメータ学習装置のコンピュータに、

複素数の観測データを、第1の複素数のパラメータが設定されたニューラルネットワークである第1の複素ニューラルネットワークを用いて複素数のまま符号化し、複素数の潜在変数の分布を符号化データとして出力することと、

前記符号化データを、第2の複素数のパラメータが設定されたニューラルネットワークである第2の複素ニューラルネットワークを用いて複素数のまま復号して得られる復号データとして、再構築された前記複素数の観測データを出力することと

を含み、

前記第1の複素数のパラメータおよび前記第2の複素数のパラメータを、勾配法を利用した学習により求める

パラメータ学習処理を実行させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、符号化装置、復号装置、パラメータ学習装置、およびプログラムに関し、特に、より高品位に符号化および復号することができるようにした符号化装置、復号装置、パラメータ学習装置、およびプログラムに関する。

【背景技術】

【0002】

近年、ディープラーニングを用いた手法が飛躍的に高い精度を上げており、画像認識や音声認識などの幅広い分野において盛んに研究が行われている（例えば、非特許文献1参照）。

【0003】

また、これまでに数多くのディープラーニング手法が提案されており、その中の最も代表的な生成モデルの一つとして、変分オートエンコーダ（variational autoencoder; VAE）が挙げられる（例えば、非特許文献2参照）。

【先行技術文献】

【非特許文献】

【0004】

【非特許文献1】Y. LeCun et al., “Deep learning,” Nature, vol. 521, no. 7553, p. 436-444, 2015.

【非特許文献2】D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” in ICML, 2014.

【発明の概要】

【発明が解決しようとする課題】

【0005】

ところで、VAEは、これまで様々な分野で用いられてきたにもかかわらず、潜在変数や入力特徴量はバイナリ値または実数値を扱うことが仮定されてきた。しかしながら、音声の複素スペクトル、MRI（Magnetic Resonance Imaging）画像、音響インテンシティ、衛星画像などのように、実データに基づいた画像認識や音声信号処理では複素数データを取り扱う場合が多い。また、音声認識や音声合成では、音響特徴量としてMFCC（Mel-Frequency Cepstrum Coefficients）やメルケプストラム特徴量、STRAIGHTスペクトルなど、振幅スペクトルに基づいた特徴量が利用されており、位相スペクトルと比較して振幅スペク

10

20

30

40

50

トルの方が音声認識や音声合成に効果的であることが知られている。さらに、VAEを用いた特徴抽出でも入力には実数値のメルケプストラム特徴量が利用される場合が多い。

【0006】

しかしながら、これらは位相情報が欠落しており、元の音声データに対して少なからず情報の損失が存在するため、位相情報を含めた複素数データをそのまま表現することが重要となる。例えば、ニューラルネットワークや、制限ボルツマンマシン (restricted Boltzmann machine)、非負値行列因子分解 (non-negative matrix factorization; NMF) などでは、それぞれ複素数データを表現する拡張表現が提案されている。

【0007】

そこで、このような位相情報の欠落や情報の欠損などを回避して、高品位に符号化および復号できるようにすることが求められている。

【0008】

本開示は、このような状況に鑑みてなされたものであり、より高品位に符号化および復号することができるようにするものである。

【課題を解決するための手段】

【0009】

本開示の第1の側面の符号化装置は、複素数の観測データを、勾配法を利用した学習により求められた複素数のパラメータが設定されたニューラルネットワークである複素ニューラルネットワークを用いて複素数のまま符号化し、複素数の潜在変数の分布を符号化データとして出力する符号化部を備える。

【0010】

本開示の第1の側面のプログラムは、複素数の観測データを、勾配法を利用した学習により求められた複素数のパラメータが設定されたニューラルネットワークである複素ニューラルネットワークを用いて複素数のまま符号化し、複素数の潜在変数の分布を符号化データとして出力することを含む。

【0011】

本開示の第1の側面においては、複素数の観測データが、勾配法を利用した学習により求められた複素数のパラメータが設定されたニューラルネットワークである複素ニューラルネットワークを用いて複素数のまま符号化され、複素数の潜在変数の分布が符号化データとして出力される。

【0012】

本開示の第2の側面の復号装置は、複素数の観測データが符号化された符号化データを、勾配法を利用した学習により求められた複素数のパラメータが設定されたニューラルネットワークである複素ニューラルネットワークを用いて複素数のまま復号して得られる復号データとして、再構築された前記複素数の観測データを出力する復号部を備える。

【0013】

本開示の第2の側面のプログラムは、複素数の観測データが符号化された符号化データを、勾配法を利用した学習により求められた複素数のパラメータが設定されたニューラルネットワークである複素ニューラルネットワークを用いて複素数のまま復号して得られる復号データとして、再構築された前記複素数の観測データを出力することを含む。

【0014】

本開示の第2の側面においては、複素数の観測データが符号化された符号化データが、勾配法を利用した学習により求められた複素数のパラメータが設定されたニューラルネットワークである複素ニューラルネットワークを用いて複素数のまま復号されて得られる復号データとして、再構築された前記複素数の観測データが出力される。

【0015】

本開示の第3の側面のパラメータ学習装置は、複素数の観測データを、第1の複素数のパラメータが設定されたニューラルネットワークである第1の複素ニューラルネットワークを用いて複素数のまま符号化し、複素数の潜在変数の分布を符号化データとして出力する符号化部と、前記符号化データを、第2の複素数のパラメータが設定されたニューラル

10

20

30

40

50

ネットワークである第2の複素ニューラルネットワークを用いて複素数のまま復号して得られる復号データとして、再構築された前記複素数の観測データを出力する復号部とを備え、前記第1の複素数のパラメータおよび前記第2の複素数のパラメータを、勾配法を利用した学習により求める。

【0016】

本開示の第3の側面のプログラムは、複素数の観測データを、第1の複素数のパラメータが設定されたニューラルネットワークである第1の複素ニューラルネットワークを用いて複素数のまま符号化し、複素数の潜在変数の分布を符号化データとして出力することと、前記符号化データを、第2の複素数のパラメータが設定されたニューラルネットワークである第2の複素ニューラルネットワークを用いて複素数のまま復号して得られる復号データとして、再構築された前記複素数の観測データを出力することとを含み、前記第1の複素数のパラメータおよび前記第2の複素数のパラメータを、勾配法を利用した学習により求める。

10

【0017】

本開示の第3の側面においては、複素数の観測データが、第1の複素数のパラメータが設定されたニューラルネットワークである第1の複素ニューラルネットワークを用いて複素数のまま符号化され、複素数の潜在変数の分布が符号化データとして出力され、その符号化データが、第2の複素数のパラメータが設定されたニューラルネットワークである第2の複素ニューラルネットワークを用いて複素数のまま復号して得られる復号データとして、再構築された複素数の観測データが出力される。そして、第1の複素数のパラメータおよび第2の複素数のパラメータが、勾配法を利用した学習により求められる。

20

【発明の効果】

【0018】

本開示の第1乃至第3の側面によれば、より高品位に符号化および復号することができる。

【0019】

なお、ここに記載された効果は必ずしも限定されるものではなく、本開示中に記載されたいずれかの効果であってもよい。

【図面の簡単な説明】

【0020】

30

【図1】VAEの構造を示す図である。

【図2】VAEにおけるReparameterization Trickについて説明する図である。

【図3】複素VAEの構造を示す図である。

【図4】複素VAEにおけるReparameterization Trickについて説明する図である。

【図5】本技術を適用した符号化装置の一実施の形態の構成例を示すブロック図である。

【図6】本技術を適用した復号装置の一実施の形態の構成例を示すブロック図である。

【図7】符号化処理を説明するフローチャートである。

【図8】復号処理を説明するフローチャートである。

【図9】パラメータ学習装置の構成例を示す図である。

【図10】パラメータ学習処理を説明するフローチャートである。

40

【図11】実験条件の一例を示す図である。

【図12】PESQ評価の一例を示す図である。

【図13】実験結果の一例を示す図である。

【図14】本技術を適用したコンピュータの一実施の形態の構成例を示すブロック図である。

【発明を実施するための形態】

【0021】

以下、本技術を適用した具体的な実施の形態について、図面を参照しながら詳細に説明する。

【0022】

50

< 従来のモデルであるVAE >

まず、本明細書において提案される新たなモデルである複素VAE (variational autoencoder) について説明する前に、従来のモデルであるVAEについて説明する。

【 0 0 2 3 】

VAEは、D次元の観測データ $x \in \mathbb{R}^D$ からH次元の潜在変数 $h \in \mathbb{R}^H$ の分布を表す確率的なエンコーダ $q_\phi(h|x)$ と、逆に潜在変数から観測データの分布を表す確率的なデコーダ $p_\theta(x|h)$ とを、それぞれ異なるニューラルネットワークで表現した生成モデルである。また、エンコーダ $q_\phi(h|x)$ のモデルパラメータ、および、デコーダ $p_\theta(x|h)$ のモデルパラメータ (正確には、 θ, ϕ はモデル全体のパラメータ) は、AEVB (auto-encoding variational bayes) アルゴリズムによって推定される。

10

【 0 0 2 4 】

図1には、VAEの構造が示されている。例えば、エンコーダ11は、観測データ x を入力し、潜在変数 h の確率分布を出力する。一方、デコーダ12は、潜在変数 h を入力し、観測データ x' を再構築する。

【 0 0 2 5 】

ここで、観測データ x の対数尤度は、次の式(1)に示すイェンセンの不等式より、下限 $L(\theta, \phi; x)$ を求めることができる。ただし、式(1)において、 $D_{KL}(q||p)$ は、分布 q および分布 p どうしの間のKLダイバージェンスを表す。

【 0 0 2 6 】

【 数 1 】

$$\log p_\theta(x) = D_{KL}(q_\phi(h|x) || p_\theta(h|x)) + L(\theta, \phi; x)$$

20

$$\geq L(\theta, \phi; x) \triangleq E_{q_\phi(h|x)} \left[\log \frac{p_\theta(x, h)}{q_\phi(h|x)} \right] \quad \dots (1)$$

【 0 0 2 7 】

さらに、この下限 $L(\theta, \phi; x)$ は、次の式(2)に示すように変形することができる。

【 0 0 2 8 】

【 数 2 】

$$L(\theta, \phi; x) = E_{q_\phi(h|x)} [\log p_\theta(x|h)] - D_{KL}(q_\phi(h|x) || p_\theta(h)) \quad \dots (2)$$

30

【 0 0 2 9 】

この式(2)の右辺第一項は、観測データ x から得られる潜在変数 h から生成されるデータの対数尤度の期待値を表している。また、この式(2)の右辺第二項は、観測データ x から得られる潜在変数 h の分布 $q_\phi(h|x)$ を、潜在変数 h の単純な事前分布 $p_\theta(h)$ に近付ける制約項を表している。

【 0 0 3 0 】

そして、補助関数法の考え方から、下限 $L(\theta, \phi; x)$ を最大化するエンコーダ11のモデルパラメータおよびデコーダ12のモデルパラメータは、対数尤度 $\log p_\theta(x)$ についても最適化する。したがって、VAEでは、下限 $L(\theta, \phi; x)$ を最大化するように勾配法を用いて、エンコーダ11のモデルパラメータおよびデコーダ12のモデルパラメータを最適化する。

40

【 0 0 3 1 】

しかしながら、上述した式(2)の右辺第一項について、エンコーダ11のモデルパラメータに関する勾配を求めることは困難である。そこで、以下で説明するReparameterization Trickが用いられる。

【 0 0 3 2 】

例えば、VAEを用いた多くの研究は、潜在変数 h の確率分布としてGaussian分布を仮定

50

している。この場合、エンコーダ 1 1 のニューラルネットワークは、Gaussian 分布の平均 $\mu \in \mathbb{R}^H$ および分散 $\Sigma \in \mathbb{R}^{H \times H}$ の結合ベクトル $[\mu^T \ \Sigma^T]^T$ を出力する。

【 0 0 3 3 】

そして、VAE の順伝播では、次の式 (3) より得られる潜在変数 h のサンプル \tilde{h} をデコーダ 1 2 のニューラルネットワークの入力とし、その出力である再構築データ x' を取得する。

【 0 0 3 4 】

【 数 3 】

$$\tilde{h} \sim N(h; \mu, \Delta(\sigma)) \quad \dots (3)$$

10

【 0 0 3 5 】

ここで、式 (3) において、 $N(\cdot; \mu, \Sigma)$ は、平均 μ 、分散共分散行列 Σ の多変量正規分布を表す。また、 $\Delta(\cdot)$ は、入力を対角成分とする行列を返す関数とする。

【 0 0 3 6 】

しかしながら、図 2 の A に示すように、潜在変数 h はサンプルされたものであるため微分可能ではない。したがって、学習時にデコーダ 1 2 のニューラルネットワークから逆伝播された誤差を、エンコーダ 1 1 のニューラルネットワークへ逆伝播することができない。

【 0 0 3 7 】

そこで、次の式 (4) に示すように、標準正規分布からサンプルされた標準正規乱数を用いて、潜在変数 h を計算する。

20

【 0 0 3 8 】

【 数 4 】

$$\begin{aligned} \tilde{h} &\sim \mu + \sqrt{\sigma} \circ \varepsilon \\ \varepsilon &\sim N(0, I) \end{aligned} \quad \dots (4)$$

【 0 0 3 9 】

これにより、図 2 の B に示すように、学習時にデコーダ 1 2 のニューラルネットワークから逆伝播された誤差を、エンコーダ 1 1 のニューラルネットワークへ逆伝播させることができる。

30

【 0 0 4 0 】

ここで、図 2 は、Gaussian 分布に従う潜在変数を仮定した VAE において、図 2 の A は、Reparameterization Trick を使用しない場合を示し、図 2 の B は、Reparameterization Trick を使用した場合を示している。また、図 2 において、実線は順伝播を表し、点線はサンプリングを表す。

【 0 0 4 1 】

次に、Gaussian 分布を仮定した VAE およびパラメータ最適化について説明する。

【 0 0 4 2 】

上述したように、VAE のパラメータは、上記の式 (2) を最大化するように勾配法を用いて最適化することで求められる。ここでは、デコーダ 1 2 のニューラルネットワークの出力確率として、連続値データを表現する単位分散の多変量正規分布を、次の式 (5) に示すように仮定する。

40

【 0 0 4 3 】

【 数 5 】

$$p_{\theta}(x|h) = N(x; a, I) \quad \dots (5)$$

【 0 0 4 4 】

ただし、式 (5) で用いられている複素平均 $a \in \mathbb{R}^D$ は、デコーダ 1 2 のニューラルネットワークの出力値である。また、期待値 $E_{q(h|x)}[f(h)]$ は、モンテカルロ法によって、

50

次の式(6)に示すように近似することができるが、この式(6)で用いられるサンプル数 $L = 1$ としても十分な性能となることが知られている。

【0045】

【数6】

$$E_{q_{\phi}(h|x)}[f(h)] \approx \frac{1}{L} \sum_{i=1}^L f(h^{(i)}) \quad \dots (6)$$

【0046】

以上より、上述した式(2)の右辺第一項は、次の式(7)により表される。なお、式(7)で用いられている K は、モデルのパラメータに依存しない項である。

10

【0047】

【数7】

$$E_{q_{\phi}(h|x)}[\log p_{\theta}(x|h)] \approx -\frac{1}{2} \|x-a\|_2^2 + K \quad \dots (7)$$

【0048】

また、潜在変数 h の分布として Gaussian 分布を仮定した場合、多くの既存研究は、潜在変数 h の事前分布 $p_{\theta}(h)$ を、次の式(8)に示すように、標準正規分布と同時に仮定している。

【0049】

【数8】

$$p_{\theta}(h) \triangleq N(0, I) \quad \dots (8)$$

【0050】

したがって、上述した式(2)の右辺第二項は、次の式(9)に示すように、解析的に求めることができる。

【0051】

【数9】

$$\begin{aligned} D_{KL}(q_{\phi}(h|x) || p_{\theta}(h)) &= D_{KL}(N(\mu, \Delta(\sigma)) || N(0, I)) \\ &= \frac{1}{2} (\mu^T \mu + \|\sigma - 1 - \log \sigma\|_1) \end{aligned} \quad \dots (9)$$

30

【0052】

以上のように、VAEでは、観測データの分布として Gaussian を仮定しているため、実数値を取り得る観測データを表現することができる。また、Bernoulli 分布を仮定して二値データを適切に表現することもできる。なお、この場合には、式(7)の loss 関数はクロスエントロピーとなる。

【0053】

しかしながら、いずれも分布の仮定上においても、従来、複素スペクトルなどの複素数データを直接表現することはできなかった。そこで、以下で説明するように、従来のVAEに基づいて、エンコーダにより得られる潜在変数から、デコーダを通して複素数データを表現する新たな生成モデルを提案する。

40

【0054】

< 提案のモデルである複素VAE >

図3に示すモデルを、複素変分自己符号化器(複素VAE; complex-valued variational autoencoder; CVAE)と称する。

【0055】

図3に示すように、エンコーダ21は、複素数の観測データ z を入力し、複素数の潜在変数 h の確率分布を出力する。一方、デコーダ22は、複素数の潜在変数 h を入力し、複

50

素数の観測データ z' を再構築する。

【 0 0 5 6 】

複素VAEでは、複素数の潜在変数 $h \in \mathbb{C}^H$ からデコーダ 2 2 を通して得られる観測データ $z \in \mathbb{C}^D$ は、次の式 (1 0) に示すように、複素平均 $a \in \mathbb{C}^D$ 、分散行列 $\Gamma \in \mathbb{C}^{D \times D}$ 、および、擬似分散行列 $C \in \mathbb{C}^{D \times D}$ の多変量複素正規分布に従うとする。

【 0 0 5 7 】

【 数 1 0 】

$$p_{\theta}(z|h) = N_{\mathbb{C}}(z; a, \Gamma, C)$$

$$= \frac{1}{\pi^D \sqrt{\det(\Gamma) \det(\bar{\Gamma} - C^H \Gamma^{-1} C)}} \quad \dots (10)$$

$$\cdot \exp \left\{ -\frac{1}{2} \begin{bmatrix} z-a \\ \bar{z}-\bar{a} \end{bmatrix}^H \begin{bmatrix} \Gamma & C \\ C^H & \Gamma^H \end{bmatrix}^{-1} \begin{bmatrix} z-a \\ \bar{z}-\bar{a} \end{bmatrix} \right\}$$

10

【 0 0 5 8 】

ここで、式 (1 0) で用いられている複素平均 a 、分散行列 Γ 、および擬似分散行列 C は、いずれもデコーダ 2 2 の出力であり、デコーダ 2 2 のパラメータ θ は、全て複素数である。なお、より簡易化するため、分散行列 Γ を単位分散とし、擬似分散行列 C を零擬似分散行列 (すなわち、 $p_{\theta}(z|h) = N_{\mathbb{C}}(x; \mu, I, 0)$) として、複素平均 a のみがデコーダ 2 2 の出力であることができる。

20

【 0 0 5 9 】

また、複素VAEでは潜在変数に対しても複素正規分布を仮定する。とりわけ、複素数を表現する単純な (次元独立な) 複素正規分布として対角な分散行列 $\Delta(\sigma)$ および擬似分散行列 $\Delta(\delta)$ の分布を、次の式 (1 1) に示すように仮定する。

【 0 0 6 0 】

【 数 1 1 】

$$p_{\phi}(h|z) = N_{\mathbb{C}}(h; \mu, \Delta(\sigma), \Delta(\delta)) \quad \dots (11)$$

30

【 0 0 6 1 】

ここで、 θ は、エンコーダ 2 1 の複素パラメータ集合であり、平均 $\mu \in \mathbb{C}^H$ 、分散 $\Delta(\sigma) \in \mathbb{R}^H$ 、擬似分散 $\Delta(\delta) \in \mathbb{C}^H$ はいずれも、入力される複素数の観測データ z に対するエンコーダ 2 1 の出力である。そして、複素VAEでは、潜在変数 h の事前分布として、次の式 (1 2) に示すように、最も単純な複素正規分布である標準複素正規分布と仮定する。

【 0 0 6 2 】

【 数 1 2 】

$$p_{\theta}(h) \stackrel{\Delta}{=} N_{\mathbb{C}}(0, I, 0) \quad \dots (12)$$

40

【 0 0 6 3 】

このとき、潜在変数に関する制約項は、次の式 (1 3) に示すように解析的に計算することができる。

【 0 0 6 4 】

【数 1 3】

$$\begin{aligned}
 D_{\text{KL}}(q_{\phi}(h|z) || p_{\theta}(h)) \\
 &= D_{\text{KL}}(N_c(\mu, \Delta(\sigma), \Delta(\delta)) || N_c(0, I, O)) \quad \dots (13) \\
 &= \mu^H \mu + ||\sigma - 1 - \frac{1}{2} \log(\sigma^2 - |\delta|^2)||_1
 \end{aligned}$$

【0065】

ただし、式(13)において、 \cdot^2 は、要素ごとの二乗を表し、 $|\cdot|$ は、絶対値を表す。また、式(13)で示す制約項は、エンコーダ21を通して得られる複素潜在変数の分布について、平均 μ や分散 Δ だけでなく擬似分散 $\Delta(\delta)$ も入力により変化する中で、なるべく単純な複素正規分布に近付ける役割を果たす。

【0066】

また、複素VAEの目的関数は、従来のVAEと同様に、次の式(14)で示す制約付き対数尤度の下限 L を求める関数とする。

【0067】

【数 1 4】

$$\begin{aligned}
 L(\theta, \phi; z) = E_{q_{\phi}(h|z)} [\log p_{\theta}(z|h)] \\
 - D_{\text{KL}}(q_{\phi}(h|z) || p_{\theta}(h)) \quad \dots (14)
 \end{aligned}$$

【0068】

次に、複素VAEにおけるReparameterizationTrickおよびパラメータ最適化について説明する。

【0069】

複素VAEでは、上述した式(14)を最大化するように、例えば、複素勾配法を用いて複素VAEのパラメータ θ およびパラメータ ϕ を最適化する。例えば、最も単純な複素勾配法は複素最急降下法であり、複素数の学習率 α (α の実部は0より大きい: $\text{Re}(\alpha) > 0$)を用いて、パラメータ θ は、次の式(15)を繰り返し計算することで更新される。

【0070】

【数 1 5】

$$\theta^{(\text{new})} \leftarrow \theta^{(\text{old})} + \alpha \cdot 2 \frac{\partial L}{\partial \theta} \quad \dots (15)$$

【0071】

なお、このようなパラメータ θ の更新の拡張として、例えば、複素Adam(Adaptive moment estimation)を用いてもよい。また、式(15)ではパラメータ θ について示したが、パラメータ ϕ についても同様に、更新することができる。

【0072】

ただし、この式(15)における複素数の偏微分は、パラメータ θ の実部 $\text{Re}(\theta)$ および虚部 $\text{Im}(\theta)$ により、次の式(16)で示すWirtinger微分である。

【0073】

【数 1 6】

$$\frac{\partial L}{\partial \theta} = \frac{1}{2} \left(\frac{\partial L}{\partial \text{Re}(\theta)} - j \frac{\partial L}{\partial \text{Im}(\theta)} \right) \quad \dots (16)$$

【0074】

しかしながら、上述した式(11)に従って潜在変数を単純にサンプリングし、デコーダ22へ入力すると、このプロセスが微分可能ではないためエンコーダ21側へ誤差を逆

伝播することができないことになる。そこで、上述したような従来のVAEと同様に、標準正規分布に従う乱数を用いて潜在変数のサンプルを表現する、複素VAEにおけるReparameterization Trickを提案する。

【0075】

例えば、複素数の潜在変数 $h \in \mathbb{C}^H$ は、実部 $x \in \mathbb{R}^H$ および虚部 $y \in \mathbb{R}^H$ を用いて、 $h=x+iy$ と記述することができる。そこで、上述した式(11)の仮定より、複素数の潜在変数 h の各要素は互いに独立となり、複素数の潜在変数 h の実部 x および虚部 y は、それぞれ平均 μ の実部 $\text{Re}(\mu)$ および虚部 $\text{Im}(\mu)$ 、並びに、分散 σ の実部 σ_x および虚部 σ_y の正規分布に従う。ただし、分散 σ の実部 σ_x および虚部 σ_y は、次の式(17)で表される。

【0076】

【数17】

$$\begin{cases} \sigma_x \triangleq \frac{\sigma + \text{Re}(\delta)}{2} \\ \sigma_y \triangleq \frac{\sigma - \text{Re}(\delta)}{2} \end{cases} \quad \dots (17)$$

【0077】

また、複素数の潜在変数 h の実部 x および虚部 y の添字の対応する要素間には相関があり、その相関係数ベクトル $\rho \in \mathbb{R}^H$ は、次の式(18)で表される。

【0078】

【数18】

$$\rho = \frac{\text{Im}(\delta)}{\sigma + \text{Re}(\delta)} \quad \dots (18)$$

【0079】

そして、次の式(19)に示すように、標準正規乱数 $\tilde{x} \sim N(0, 1)$ を用いて複素数の潜在変数 h の実部 x をサンプリングしたとき、複素数の潜在変数 h の虚部 y は、次の式(20)に示すような条件付確率に従う。

【0080】

【数19】

$$\tilde{y} = \text{Im}(\mu) + \sigma_y \circ \varepsilon_x \quad \dots (19)$$

【数20】

$$p(y|\tilde{x}) = N\left(\text{Im}(\mu) + \rho \circ \sqrt{\frac{\sigma_y}{\sigma_x}} \circ (\tilde{x} - \text{Re}(\mu)), (1 - \rho^2) \circ \sigma_y\right) \quad \dots (20)$$

【0081】

したがって、標準正規乱数 $\tilde{y} \sim N(0, 1)$ を用いて、次の式(21)を計算することで、複素数の潜在変数 h の虚部 y のサンプル $y\sim$ を得ることができる。ただし、この式(1)において、 \cdot は要素ごとの平方根を表し、 \cdot^2 は要素ごとの二乗を表す。

【0082】

【数21】

$$\tilde{y} = \text{Im}(\mu) + \rho \circ \sqrt{\frac{\sigma_y}{\sigma_x}} \circ (\tilde{x} - \text{Re}(\mu)) + \sqrt{(1 - \rho^2) \circ \sigma_y} \circ \varepsilon_y \quad \dots (21)$$

【0083】

以上をまとめると、複素潜在変数のサンプル $h\sim$ は、図4に示すように、標準正規乱数

10

20

30

40

50

x および標準正規乱数 y を用いて、次の式 (22) に示すように表すことができる。

【0084】

【数22】

$$\begin{aligned} \tilde{h} &= \tilde{x} + i\tilde{y} \\ &= \mu + \kappa_x \circ \varepsilon_x + \kappa_y \circ \varepsilon_y \end{aligned} \quad \dots (22)$$

【0085】

なお、この式 (22) で用いられている補助演算係数 κ_x および κ_y は、次の式 (23) に示すように定義される。

【数23】

$$\begin{cases} \kappa_x \triangleq \frac{\sigma + \delta}{\sqrt{2\sigma + 2\operatorname{Re}(\delta)}} \\ \kappa_y \triangleq i \frac{\sqrt{\sigma^2 - |\delta|^2}}{\sqrt{2\sigma + 2\operatorname{Re}(\delta)}} \end{cases} \quad \dots (23)$$

【0086】

ここで、補助演算係数 κ_x および κ_y は、図4に示すようにデコーダ22からエンコーダ21へ誤差を逆伝播することが可能となるような複素数の潜在変数 h の分布を求めるための補助的な演算を行うのに用いられる。即ち、エンコーダ21から出力される分散 σ および擬似分散 δ から直接的に複素数の潜在変数 h の分布を求めるのではなく、分散 σ および擬似分散 δ から式 (23) に従って補助的に補助演算係数 κ_x および κ_y を求めた後に、補助演算係数 κ_x および κ_y から複素数の潜在変数 h の分布が求められる。

【0087】

なお、学習を安定化させるため、例えば、分散 σ および擬似分散 δ に代えて、対数分散 $\log \sigma$ および対数擬似分散 $\log \delta$ をエンコーダ21の出力としてもよい。このようなテクニックは、従来のVAEでもしばしば同様に用いられる。

【0088】

< 符号化装置および復号装置の構成例 >

図5は、本技術を適用した符号化装置の一実施の形態の構成例を示すブロック図である。

【0089】

図5に示すように、符号化装置31は、複素数データ取得部41、前処理部42、パラメータ保持部43、および符号化部44を備えて構成される。例えば、符号化装置31には、複素数の観測データ z が入力され、符号化装置31からは、複素数の潜在変数 h の分布が出力される。

【0090】

複素数データ取得部41は、符号化装置31において符号化の対象となる複素数の観測データ z が供給されると、その複素数の観測データ z を取得して、前処理部42に供給する。例えば、符号化装置31における符号化の対象としては、音声データや画像データなど様々なデータが適用可能であり、それらのデータは、複素数データとして供給される。

【0091】

前処理部42は、複素数データ取得部41から供給された複素数の観測データ z に対して、符号化部44において符号化が行われる前に必要となる前処理を施す。例えば、前処理部42は、符号化装置31において音声データを符号化の対象とする場合には、その音声データを単位時間（以下、フレームと称する）ごとに切り出す。そして、前処理部42は、フレームごとの音声データから、M F C C (Mel-Frequency Cepstrum Coefficients

10

20

30

40

50

：メル周波数ケプストラム係数)やメルケプストラム特徴量などのようなスペクトル特徴量を算出する。さらに、前処理部42は、フレームごとの音声データのスペクトル特徴量を正規化して、符号化部44に供給する。

【0092】

パラメータ保持部43は、符号化部44の複素ニューラルネットワーク45に設定されるパラメータを保持しており、後述する図9のパラメータ学習装置61において勾配法を利用した学習により求められたパラメータが供給される。

【0093】

符号化部44は、複素ニューラルネットワーク45により構成され、パラメータ保持部43で保持されているパラメータが設定された複素ニューラルネットワーク45により、前処理部42において前処理が施された複素数の観測データ z を符号化する。

10

【0094】

複素ニューラルネットワーク45は、上述したように、従来のVAEを、複素数データをそのまま扱うことができるように拡張表現されたモデルである複素VAEである。例えば、複素ニューラルネットワーク45の出力層となる各ノードは、複素正規分布の平均 μ を出力する平均出力部46、複素正規分布の分散を出力する分散出力部47、および、複素正規分布の擬似分散を出力する擬似分散出力部48により構成される。

【0095】

そして、符号化部44は、複素ニューラルネットワーク45の出力に基づき、複素数の観測データ z を符号化した符号化データとして、複素数の潜在変数 h の分布を出力する。

20

【0096】

例えば、符号化部44は、平均出力部46から出力される平均 μ 、分散出力部47から出力される分散、および、擬似分散出力部48から出力される擬似分散を用いて、上述した式(11)を演算して求められる複素正規分布 $p(h|z)$ を、複素数の潜在変数 h の分布として出力することができる。または、符号化部44は、平均 μ 、分散、および擬似分散をそのまま、複素数の潜在変数 h の分布として出力してもよい。または、符号化部44は、最も簡易化して、平均 μ のみを複素数の潜在変数 h の分布として出力してもよい。

【0097】

このように符号化装置31は構成されており、複素数の観測データ z を、複素VAEによって複素数データのまま扱って符号化し、符号化データとして、複素数の潜在変数 h の分布を出力することができる。

30

【0098】

図6は、本技術を適用した復号装置の一実施の形態の構成例を示すブロック図である。

【0099】

図6に示すように、復号装置32は、復号部51、パラメータ保持部52、後処理部53、および出力部54を備えて構成される。例えば、復号装置32には、符号化装置31から出力される複素数の潜在変数 h の分布が入力され、復号装置32からは、再構築された複素数の観測データ z' が出力される。

【0100】

復号部51には、復号装置32において復号の対象となる符号化データとして複素数の潜在変数 h の分布が供給される。復号部51は、複素ニューラルネットワーク55により構成され、パラメータ保持部52で保持されているパラメータが設定された複素ニューラルネットワーク55により、複素数の潜在変数 h の分布を復号する。

40

【0101】

複素ニューラルネットワーク55は、上述したように、従来のVAEを、複素数データをそのまま扱うことができるように拡張表現されたモデルである複素VAEである。例えば、複素ニューラルネットワーク55の出力層となる各ノードは、複素平均 a を出力する複素平均出力部56、分散行列を出力する分散行列出力部57、および、擬似分散行列 C を出力する擬似分散行列出力部58により構成される。

50

【 0 1 0 2 】

そして、復号部 5 1 は、複素ニューラルネットワーク 5 5 の出力に基づき、複素数の潜在変数 h の分布を復号した復号データとして、符号化装置 3 1 に入力された複素数の観測データ z を再構築した複素数の観測データ z' を出力する。

【 0 1 0 3 】

例えば、復号部 5 1 は、複素平均出力部 5 6 から出力される複素平均 a 、分散行列出力部 5 7 から出力される分散行列 Σ 、および、擬似分散行列出力部 5 8 から出力される擬似分散行列 C を用いて、上述した式 (1 0) を演算することにより求められる多変量複素正規分布 $p(z|h)$ を、再構築した複素数の観測データ z' として出力する。または、復号部 5 1 は、複素平均 a 、分散行列 Σ 、および擬似分散行列 C をそのまま、再構築した複素数の観測データ z' として出力してもよい。または、復号部 5 1 は、最も簡易化して、複素平均 a のみを、再構築した複素数の観測データ z' として出力してもよい。

10

【 0 1 0 4 】

パラメータ保持部 5 2 は、復号部 5 1 の複素ニューラルネットワーク 5 5 に設定されるパラメータ θ を保持しており、後述する図 9 のパラメータ学習装置 6 1 において勾配法を利用した学習により求められたパラメータ θ が供給される。

【 0 1 0 5 】

後処理部 5 3 は、復号部 5 1 から出力される再構築した複素数の観測データ z' に対して、符号化装置 3 1 の前処理部 4 2 で行われた前処理を元に戻す後処理を施し、即ち、単位時間 (例えば、音声データであればフレーム) ごとに正規化されたスペクトル特徴量を元に戻し、出力部 5 4 に供給する。

20

【 0 1 0 6 】

出力部 5 4 は、後処理部 5 3 において後処理が施された再構築した複素数の観測データ z' を、復号装置 3 2 の後段に出力する。

【 0 1 0 7 】

このように復号装置 3 2 は構成されており、複素数の潜在変数 h を、複素VAEによって複素数データのまま扱って復号し、復号データとして、再構築された複素数の観測データ z' を出力することができる。

【 0 1 0 8 】

図 7 は、符号化装置 3 1 が実行する符号化処理を説明するフローチャートである。

30

【 0 1 0 9 】

ステップ S 1 1 において、複素数データ取得部 4 1 は、符号化装置 3 1 において符号化の対象となる複素数の観測データ z を取得して、前処理部 4 2 に供給する。

【 0 1 1 0 】

ステップ S 1 2 において、前処理部 4 2 は、ステップ S 1 1 で複素数データ取得部 4 1 から供給された複素数の観測データ z に対し、上述したような前処理を施して、符号化部 4 4 に供給する。

【 0 1 1 1 】

ステップ S 1 3 において、符号化部 4 4 は、ステップ S 1 2 で前処理部 4 2 により前処理が施された複素数の観測データ z を、パラメータ θ が設定された複素ニューラルネットワーク 4 5 により符号化して、複素数の潜在変数 h の分布を取得する。

40

【 0 1 1 2 】

ステップ S 1 4 において、符号化部 4 4 は、ステップ S 1 2 における符号化で取得された複素数の潜在変数 h の分布を、符号化データとして出力する。

【 0 1 1 3 】

以上のような符号化処理によって、符号化装置 3 1 は、複素数の観測データ z を符号化した符号化データとして、複素数の潜在変数 h の分布を出力することができる。

【 0 1 1 4 】

図 8 は、復号装置 3 2 が実行する復号処理を説明するフローチャートである。

【 0 1 1 5 】

50

ステップS 2 1において、復号部 5 1は、符号化装置 3 1から出力された符号化データ、即ち、複素数の潜在変数 h の分布を取得する。

【0 1 1 6】

ステップS 2 2において、復号部 5 1は、ステップS 2 1で取得した複素数の潜在変数 h の分布を、パラメータが設定された複素ニューラルネットワーク 5 5により復号して、再構築した複素数の観測データ z' を取得し、後処理部 5 3に供給する。

【0 1 1 7】

ステップS 2 3において、後処理部 5 3は、ステップS 2 2で復号部 5 1から供給される再構築した複素数の観測データ z' に対し、上述したような後処理を施して、出力部 5 4に供給する。

【0 1 1 8】

ステップS 2 4において、出力部 5 4から供給される再構築した複素数の観測データ z' を、復号データとして出力する。

【0 1 1 9】

以上のような復号処理によって、復号装置 3 2は、複素数の潜在変数 h の分布を復号した復号データとして、再構築した複素数の観測データ z' を出力することができる。

【0 1 2 0】

<パラメータ学習装置の構成例>

図 9 は、パラメータ学習装置の構成例を示すブロック図である。

【0 1 2 1】

図 9 に示すように、パラメータ学習装置 6 1は、複素数データ取得部 4 1、前処理部 4 2、符号化部 4 4、復号部 5 1、初期パラメータ設定部 6 2、補助演算部 6 3、サンプリング部 6 4、初期パラメータ設定部 6 5、判定部 6 6、およびパラメータ更新部 6 7を備えて構成される。

【0 1 2 2】

例えば、パラメータ学習装置 6 1には、符号化装置 3 1に供給される複素数の観測データ z と同様の学習用複素数データが供給される。そして、パラメータ学習装置 6 1は、符号化部 4 4が符号化に用いるパラメータ、および、復号部 5 1が復号に用いるパラメータを、勾配法を利用して学習するパラメータ学習処理を行い、その学習の結果として最適化されたパラメータ およびパラメータ を出力する。

【0 1 2 3】

また、パラメータ学習装置 6 1が備える複素数データ取得部 4 1、前処理部 4 2、および符号化部 4 4は、符号化装置 3 1が備える複素数データ取得部 4 1、前処理部 4 2、および符号化部 4 4と共通の構成となっている。同様に、パラメータ学習装置 6 1が備える復号部 5 1は、復号装置 3 2が備える復号部 5 1と共通の構成となっている。

【0 1 2 4】

即ち、パラメータ学習装置 6 1に供給される学習用複素数データは、複素数データ取得部 4 1により取得され、前処理部 4 2により前処理が施されて、符号化部 4 4に供給される。

【0 1 2 5】

初期パラメータ設定部 6 2は、パラメータ学習装置 6 1において行われるパラメータ学習処理の開始時に使用される初期のパラメータを保持しており、符号化部 4 4の複素ニューラルネットワーク 4 5に対して設定する。

【0 1 2 6】

符号化部 4 4は、複素ニューラルネットワーク 4 5の出力に基づき、学習用複素数データを符号化した符号化データとして、複素正規分布の平均 μ 、複素正規分布の分散、および、複素正規分布の擬似分散を出力する。そして、符号化部 4 4は、複素正規分布の平均 μ 、複素正規分布の分散、および、複素正規分布の擬似分散から求められる符号化データである複素正規分布 $p(h|z)$ を、判定部 6 6に供給する。また、符号化部 4 4は、パラメータ学習処理の開始時には初期パラメータ設定部 6 2により設定される初期のパ

10

20

30

40

50

ラメータ を用い、その後は、パラメータ更新部 6 7 により更新されたパラメータ を用いる。

【 0 1 2 7 】

補助演算部 6 3 は、図 4 を参照して上述したように、復号部 5 1 から符号化部 4 4 へ誤差を逆伝播することが可能となるような複素数の潜在変数 h の分布を求めるための補助演算を行う。即ち、補助演算部 6 3 は、符号化部 4 4 から出力される複素正規分布の分散および擬似分散 から補助演算係数 x および y を求めるために、上述した式 (2 3) を演算する。そして、補助演算部 6 3 は、補助演算係数 x および y と、符号化部 4 4 から出力される複素正規分布の平均 μ を用いて、複素数の潜在変数 h の分布を求めて、サンプリング部 6 4 に供給する。

10

【 0 1 2 8 】

サンプリング部 6 4 は、補助演算部 6 3 から出力される複素数の潜在変数 h の分布に対するサンプリングを行い、複素数の潜在変数 h のサンプルを取得して復号部 5 1 に供給する。

【 0 1 2 9 】

初期パラメータ設定部 6 5 は、パラメータ学習装置 6 1 において行われるパラメータ学習処理の開始時に使用される初期のパラメータ を保持しており、復号部 5 1 の複素ニューラルネットワーク 5 5 に対して設定する。

【 0 1 3 0 】

復号部 5 1 は、複素ニューラルネットワーク 5 5 の出力に基づき、サンプリング部 6 4 から供給される複素数の潜在変数 h のサンプルを復号した復号データとして、複素平均 a 、分散行列、および、擬似分散行列 C をから求められる復号データである多変量複素正規分布 $p(z|h)$ を出力する。

20

【 0 1 3 1 】

判定部 6 6 は、繰り返して求められるパラメータ およびパラメータ が、所定の終了条件を満足したか否かを判定する。例えば、判定部 6 6 は、現時点のパラメータ に基づいて得られる符号化データである複素正規分布 $p(h|z)$ と、現時点のパラメータ に基づいて得られる符号化データである多変量複素正規分布 $p(z|h)$ とを用いて、上述した式 (1 4) を演算し、制約付き対数尤度の下限 L が最大化された場合に、終了条件を満足したと判定することができる。その他、判定部 6 6 は、予め設定された回数に応じて、パラメータ およびパラメータ が繰り返して求められたことを終了条件として、判定を行ってもよい。

30

【 0 1 3 2 】

パラメータ更新部 6 7 は、上述した式 (1 5) を演算することで、複素勾配法を用いてパラメータ を更新する。また、パラメータ更新部 6 7 は、同様の演算を行ってパラメータ を更新することができる。

【 0 1 3 3 】

このようにパラメータ学習装置 6 1 は構成されており、勾配法を利用した学習により最適化されたパラメータ およびパラメータ を求めることができる。

【 0 1 3 4 】

図 1 0 は、パラメータ学習装置 6 1 が実行するパラメータ学習処理を説明するフローチャートである。

40

【 0 1 3 5 】

ステップ S 3 1 において、初期パラメータ設定部 6 2 は、符号化部 4 4 の複素ニューラルネットワーク 4 5 に対して初期のパラメータ を設定し、初期パラメータ設定部 6 5 は、復号部 5 1 の複素ニューラルネットワーク 5 5 に対して初期のパラメータ を設定する。

【 0 1 3 6 】

ステップ S 3 2 において、複素数データ取得部 4 1 は学習用複素数データを取得し、前処理部 4 2 は学習用複素数データに前処理を施して、符号化部 4 4 に供給する。

50

【0137】

ステップS33において、符号化部44は、ステップS32で供給された学習用複素数データを符号化し、複素正規分布の平均 μ 、複素正規分布の分散 σ^2 、および、複素正規分布の擬似分散 σ^2 を補助演算部63に供給する。さらに、符号化部44は、複素正規分布の平均 μ 、複素正規分布の分散 σ^2 、および、複素正規分布の擬似分散 σ^2 から求められる符号化データである複素正規分布 $p(h|z)$ を、判定部66に供給する。

【0138】

ステップS34において、補助演算部63は、ステップS33で供給された複素正規分布の分散 σ^2 および擬似分散 σ^2 から補助演算係数 α_x および α_y を求める補助演算を行う。さらに、補助演算部63は、ステップS33で供給された複素正規分布の平均 μ と、補助演算係数 α_x および α_y とを用いて、複素数の潜在変数 h の分布を求めて、サンプリング部64に供給する。

10

【0139】

ステップS35において、サンプリング部64は、ステップS34で供給された複素数の潜在変数 h の分布に対するサンプリングを行い、複素数の潜在変数 h のサンプルを復号部51に供給する。

【0140】

ステップS36において、復号部51は、ステップS35で供給された複素数の潜在変数 h のサンプルを復号し、複素平均 a 、分散行列 Σ 、および、擬似分散行列 C から求められる復号データである多変量複素正規分布 $p(z|h)$ を、判定部66に供給する。

20

【0141】

ステップS37において、判定部66は、繰り返して求められるパラメータ θ およびパラメータ ϕ が、所定の終了条件を満足したか否かを判定する。例えば、判定部66は、ステップS33で符号化部44から供給される複素正規分布 $p(h|z)$ 、および、ステップS36で復号部51から供給される多変量複素正規分布 $p(z|h)$ を用いて、上述した式(14)を演算することにより制約付き対数尤度の下限 L を求める。そして、判定部66は、制約付き対数尤度の下限 L が最大化された場合に、その時点で求められているパラメータ θ およびパラメータ ϕ が、所定の終了条件を満足したと判定する。

【0142】

ステップS37において、判定部66が終了条件を満足していないと判定した場合、処理はステップS38に進む。

30

【0143】

ステップS38において、パラメータ更新部67は、上述したようにパラメータ θ およびパラメータ ϕ を更新し、それぞれ符号化部44および復号部51に設定した後、処理はステップS33に戻り、以下、同様の処理が繰り返して行われる。

【0144】

一方、ステップS37において、判定部66が終了条件を満足したと判定した場合、処理はステップS39に進む。

【0145】

ステップS39において、判定部66は、最適化されたパラメータ θ およびパラメータ ϕ を出力し、パラメータ θ を符号化装置31に供給してパラメータ保持部43に保持させ、パラメータ ϕ を復号装置32に供給してパラメータ保持部52に保持させる。その後、パラメータ学習処理は終了される。

40

【0146】

以上のようなパラメータ学習処理によって、パラメータ学習装置61は、勾配法を利用した学習により最適化されたパラメータ θ およびパラメータ ϕ を求めることができる。

【0147】

<実験条件および実験結果>

図11乃至図13を参照して、複素VAEについて行われた実験について説明する。

【0148】

50

上述したように本実施の形態で提案した複素VAEの有効性を確認するため、一般的に提供されている音声データベース・セットの音声を用いて再構築音声の品質評価実験を行った。例えば、その音声データベース・セットから50文音声（約4.2分、サンプリングレート20kHzを16kHzにダウンサンプリング）を使用し、窓幅を512として64サンプルオーバーラップの短時間フーリエ変換を施した255次元の複素スペクトルを入力特徴量とした。そして、50次元の潜在変数を持つ複素VAEを学習させた後、別の53文音声に対してエンコーダ・デコーダを通して再合成された音声について客観評価基準値であるPESQ（Perceptual Evaluation of Speech Quality）を用いて品質評価を行った。

【0149】

さらに、比較手法として、同じ複素スペクトルデータの実部および虚部を連結したベクトルを入力特徴量としたVAE（以下、VAE(R+I)と称する）、および、255次元の振幅スペクトルを入力特徴量としたVAE（以下、VAE(GL)と称する）を用いた。

【0150】

図11には、複素VAE、VAE(R+I)、およびVAE(GL)について行われた実験条件の一例が示されている。

【0151】

例えば、いずれの手法においても、バッチサイズは100とし、ロスが下がらなくなった時点で停止するように学習を行った。また、図11に示すネットワークの構造を示す表記について、例えば、CVAEの“255-100-[50,50,50]”は、3層構造で入力側から順に255, 100, 50×3ユニットのfull-connected層が並んでいることを示している。

【0152】

なお、VAE(R+I)は、各ユニットが実部および虚部の2つの自由度を持つ複素VAEとの公平な比較のために、各層のユニット数を複素VAEの2倍にしている。また、VAE(GL)のみ得られた振幅スペクトルからGriffin-Lim法によって位相復元し、その他の手法は、そのまま逆短時間フーリエ変換を施して音声信号を復元する。

【0153】

図12には、図11に示した実験条件において行われた実験に対する評価の一例が示されている。例えば、図12には、VAE(GL)、VAE(R+I)、複素VAE、および複素VAE(w/o)の各手法によるPESQ評価が示されている。

【0154】

また、図13には、図11に示した実験条件において行われた実験で得られた実験結果の一例として、振幅スペクトルの一部と、対応するオリジナル音声の振幅スペクトルが示されている。例えば、図13に示すAは、オリジナルの振幅スペクトルであり、図13に示すBは、VAE(GL)により再構築された振幅スペクトルであり、図13に示すCは、VAE(R+I)により再構築された振幅スペクトルであり、図13に示すDは、複素VAEにより再構築された振幅スペクトルである。

【0155】

図12の評価では、複素VAEによって復元された音声が高品質であったことが示されている。

【0156】

また、図13の実験結果に示すように、従来のVAEにおいて、複素スペクトルの実部および虚部を連結したベクトルを特徴量に用いると、再構築スペクトルでは特に高周波域の情報が失われてしまっており、うまく学習されていないことが表れている。

【0157】

これに対し、複素VAEでは、再構築スペクトルのどの周波数帯域でも微細構造やフォルマントがよく表れている。これは、複素VAEでは、複素数データの複素数構造を保持したまま符号化および復号が行われ、複素勾配法によるパラメータ最適化を行なっていることにより、学習中に出現する複素スペクトルの頻出パターンをより捉えられていることによると考察される。

【0158】

10

20

30

40

50

また、参考として、複素VAEにおいて複素潜在変数の擬似分散 σ^2 を導入することによる効果を確認するため、擬似分散 σ^2 を含まない（すなわち、常に $\sigma^2 = 0$ ）モデルの性能を評価したところ、図12のCVAE(w/o σ^2)において、性能が低下することが示されている。これは、複素VAEの符号化部を通して実部と虚部との間に相関のある潜在変数が作成され、その相関を取り除くような制約を加えた方が質の高い潜在変数となることを示唆している。

【0159】

以上のように、本実施の形態では、近年、様々な分野で基盤として用いられるVAEを拡張して、複素数データを直接表現できる複素VAEを新たに提案し、音声の複素スペクトルモデリングに適用してその有効性を確認した。

【0160】

そして、複素VAEの特徴として、次の第1乃至第4の特徴などが挙げられる。第1の特徴は、入力、出力、潜在変数、およびパラメータ全てが複素数で構成され、複素数の構造を保ったまま情報伝播およびパラメータ最適化を行っていることである。また、第2の特徴は、デコーダの出力および潜在変数ともに複素正規分布を仮定していることである。また、第3の特徴は、潜在変数の分布について擬似分散を考慮して標準複素正規分布に近付ける制約を加え、さらにその制約が解析的に求まり微分可能であることである。そして、第4の特徴は、デコーダからエンコーダへ勾配を誤差逆伝播できるReparametrizationが存在することである。

【0161】

なお、本技術は、符号化装置31および復号装置32が一体で構成された符号化および復号システムに適用してもよい。また、パラメータ学習装置61が、符号化装置31または復号装置32に組み込まれるような構成としてもよく、その構成では、それぞれ同一の処理を行うブロックが共通で用いられることになる。

【0162】

また、図5および図9に示した複素ニューラルネットワーク45、並びに、図6および図9に示した複素ニューラルネットワーク55について、図示したモデルは模式的な例示であって、ノードの個数や中間層の数などは、用途に応じて適切に設計することができる。なお、複素ニューラルネットワーク45および55は、従来のVAEと異なって複素数のまま演算を行うことにより、適切な制約が自動的に加わることになる。

【0163】

<コンピュータの構成例>

次に、上述した一連の処理は、ハードウェアにより行うこともできるし、ソフトウェアにより行うこともできる。一連の処理をソフトウェアによって行う場合には、そのソフトウェアを構成するプログラムが、汎用のコンピュータ等にインストールされる。

【0164】

図14は、上述した一連の処理を実行するプログラムがインストールされるコンピュータの一実施の形態の構成例を示すブロック図である。

【0165】

プログラムは、コンピュータに内蔵されている記録媒体としてのハードディスク105やROM103に予め記録しておくことができる。

【0166】

あるいはまた、プログラムは、ドライブ109によって駆動されるリムーバブル記録媒体111に格納（記録）しておくことができる。このようなリムーバブル記録媒体111は、いわゆるパッケージソフトウェアとして提供することができる。ここで、リムーバブル記録媒体111としては、例えば、フレキシブルディスク、CD-ROM(Compact Disc Read Only Memory)、MO(Magneto Optical)ディスク、DVD(Digital Versatile Disc)、磁気ディスク、半導体メモリ等がある。

【0167】

なお、プログラムは、上述したようなリムーバブル記録媒体111からコンピュータにインストールする他、通信網や放送網を介して、コンピュータにダウンロードし、内蔵す

10

20

30

40

50

るハードディスク105にインストールすることができる。すなわち、プログラムは、例えば、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、コンピュータに無線で転送したり、LAN(Local Area Network)、インターネットといったネットワークを介して、コンピュータに有線で転送することができる。

【0168】

コンピュータは、CPU(Central Processing Unit)102を内蔵しており、CPU102には、バス101を介して、入出力インタフェース110が接続されている。

【0169】

CPU102は、入出力インタフェース110を介して、ユーザによって、入力部107が操作等されることにより指令が入力されると、それに従って、ROM(Read Only Memory)103に格納されているプログラムを実行する。あるいは、CPU102は、ハードディスク105に格納されたプログラムを、RAM(Random Access Memory)104にロードして実行する。

【0170】

これにより、CPU102は、上述したフローチャートにしたがった処理、あるいは上述したブロック図の構成により行われる処理を行う。そして、CPU102は、その処理結果を、必要に応じて、例えば、入出力インタフェース110を介して、出力部106から出力、あるいは、通信部108から送信、さらには、ハードディスク105に記録等させる。

【0171】

なお、入力部107は、キーボードや、マウス、マイク等で構成される。また、出力部106は、LCD(Liquid Crystal Display)やスピーカ等で構成される。

【0172】

ここで、本明細書において、コンピュータがプログラムに従って行う処理は、必ずしもフローチャートとして記載された順序に沿って時系列に行われる必要はない。すなわち、コンピュータがプログラムに従って行う処理は、並列的あるいは個別に実行される処理(例えば、並列処理あるいはオブジェクトによる処理)も含む。

【0173】

また、プログラムは、1のコンピュータ(プロセッサ)により処理されるものであっても良いし、複数のコンピュータによって分散処理されるものであっても良い。さらに、プログラムは、遠方のコンピュータに転送されて実行されるものであっても良い。

【0174】

さらに、本明細書において、システムとは、複数の構成要素(装置、モジュール(部品)等)の集合を意味し、すべての構成要素が同一筐体中にあるか否かは問わない。したがって、別個の筐体に収納され、ネットワークを介して接続されている複数の装置、及び、1つの筐体の中に複数のモジュールが収納されている1つの装置は、いずれも、システムである。

【0175】

また、例えば、1つの装置(または処理部)として説明した構成を分割し、複数の装置(または処理部)として構成するようにしてもよい。逆に、以上において複数の装置(または処理部)として説明した構成をまとめて1つの装置(または処理部)として構成されるようにしてもよい。また、各装置(または各処理部)の構成に上述した以外の構成を付加するようにしてももちろんよい。さらに、システム全体としての構成や動作が実質的に同じであれば、ある装置(または処理部)の構成の一部を他の装置(または他の処理部)の構成に含めるようにしてもよい。

【0176】

また、例えば、本技術は、1つの機能を、ネットワークを介して複数の装置で分担、共同して処理するクラウドコンピューティングの構成をとることができる。

【0177】

また、例えば、上述したプログラムは、任意の装置において実行することができる。そ

10

20

30

40

50

の場合、その装置が、必要な機能（機能ブロック等）を有し、必要な情報を得ることができるようにすればよい。

【0178】

また、例えば、上述のフローチャートで説明した各ステップは、1つの装置で実行する他、複数の装置で分担して実行することができる。さらに、1つのステップに複数の処理が含まれる場合には、その1つのステップに含まれる複数の処理は、1つの装置で実行する他、複数の装置で分担して実行することができる。換言するに、1つのステップに含まれる複数の処理を、複数のステップの処理として実行することもできる。逆に、複数のステップとして説明した処理を1つのステップとしてまとめて実行することもできる。

【0179】

なお、コンピュータが実行するプログラムは、プログラムを記述するステップの処理が、本明細書で説明する順序に沿って時系列に実行されるようにしても良いし、並列に、あるいは呼び出しが行われたとき等の必要なタイミングで個別に実行されるようにしても良い。つまり、矛盾が生じない限り、各ステップの処理が上述した順序と異なる順序で実行されるようにしてもよい。さらに、このプログラムを記述するステップの処理が、他のプログラムの処理と並列に実行されるようにしても良いし、他のプログラムの処理と組み合わせて実行されるようにしても良い。

【0180】

なお、本明細書において複数説明した本技術は、矛盾が生じない限り、それぞれ独立に単体で実施することができる。もちろん、任意の複数の本技術を併用して実施することもできる。例えば、いずれかの実施の形態において説明した本技術の一部または全部を、他の実施の形態において説明した本技術の一部または全部と組み合わせて実施することもできる。また、上述した任意の本技術の一部または全部を、上述していない他の技術と併用して実施することもできる。

【0181】

なお、本実施の形態は、上述した実施の形態に限定されるものではなく、本開示の要旨を逸脱しない範囲において種々の変更が可能である。また、本明細書に記載された効果はあくまで例示であって限定されるものではなく、他の効果があってもよい。

【符号の説明】

【0182】

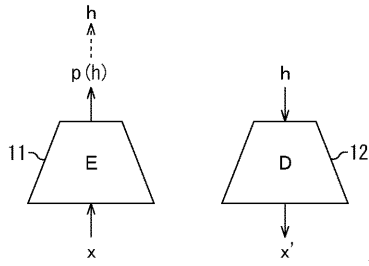
11 エンコーダ, 12 デコーダ, 21 エンコーダ, 22 デコーダ, 31 符号化装置, 32 復号装置, 41 複素数データ取得部, 42 前処理部, 43 パラメータ保持部, 44 符号化部, 51 復号部, 52 パラメータ保持部, 53 後処理部, 54 出力部, 61 パラメータ学習装置, 62 初期パラメータ設定部, 63 補助演算部, 64 サンプリング部, 65 初期パラメータ設定部, 66 判定部, 67 パラメータ更新部

10

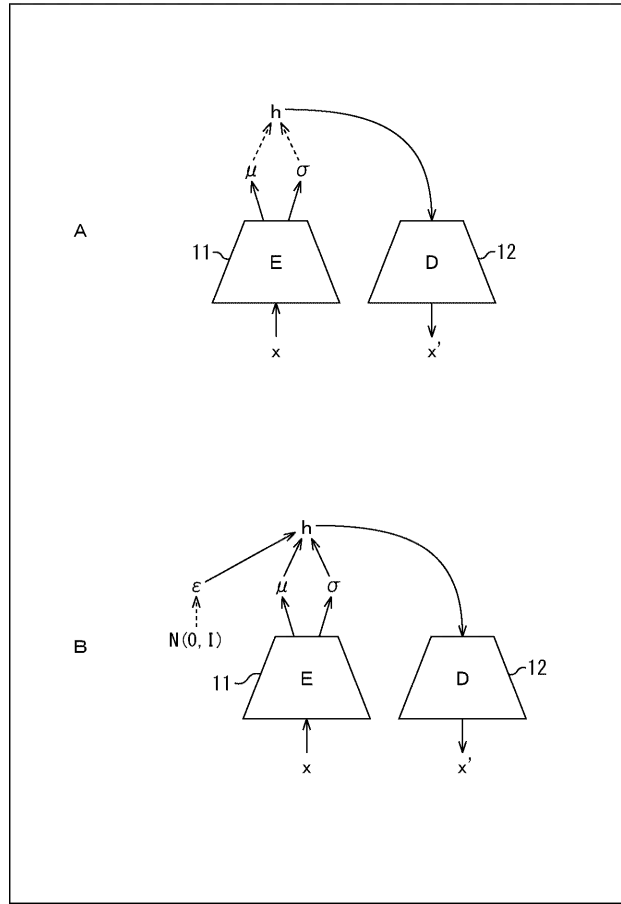
20

30

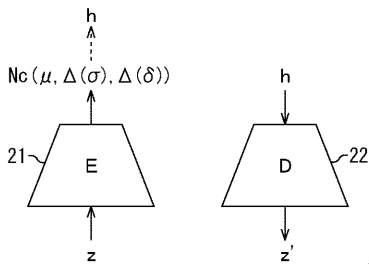
【 図 1 】
FIG. 1



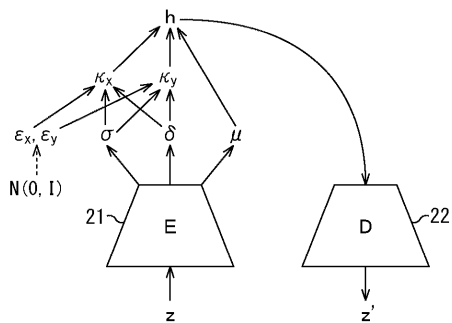
【 図 2 】
FIG. 2



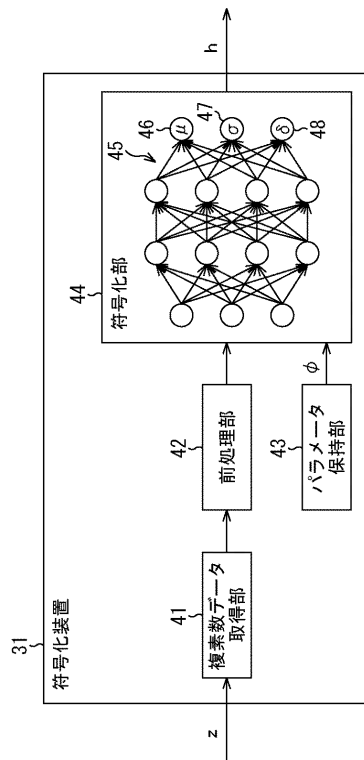
【 図 3 】
FIG. 3



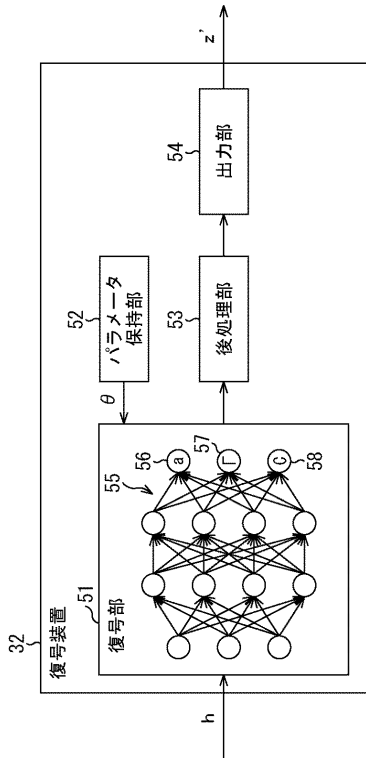
【 図 4 】
FIG. 4



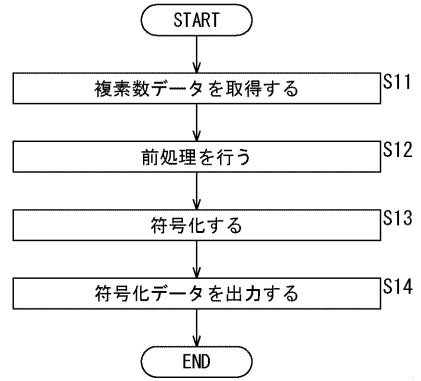
【 図 5 】
FIG. 5



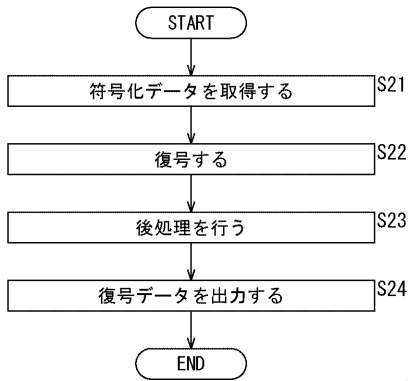
【図6】
FIG. 6



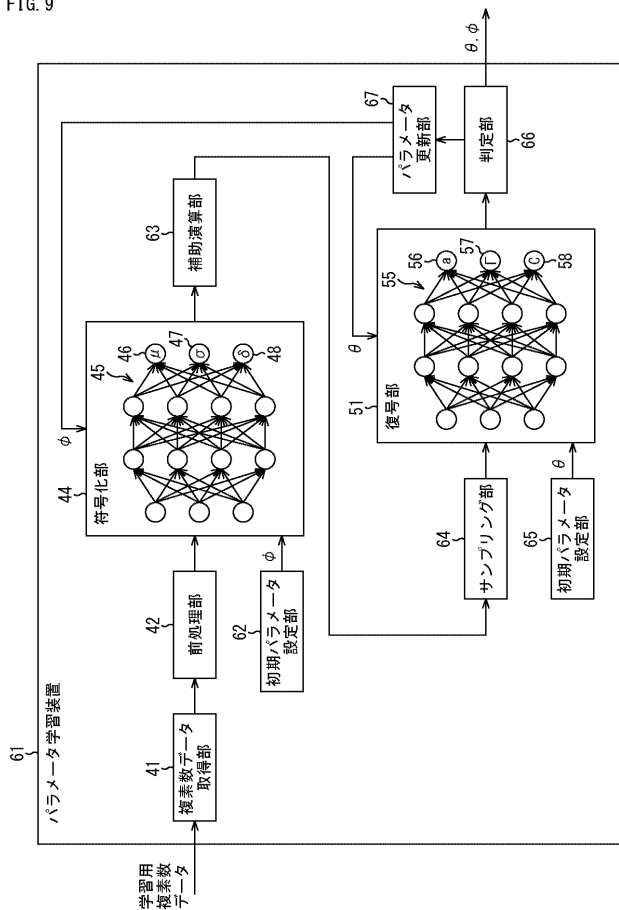
【図7】
FIG. 7



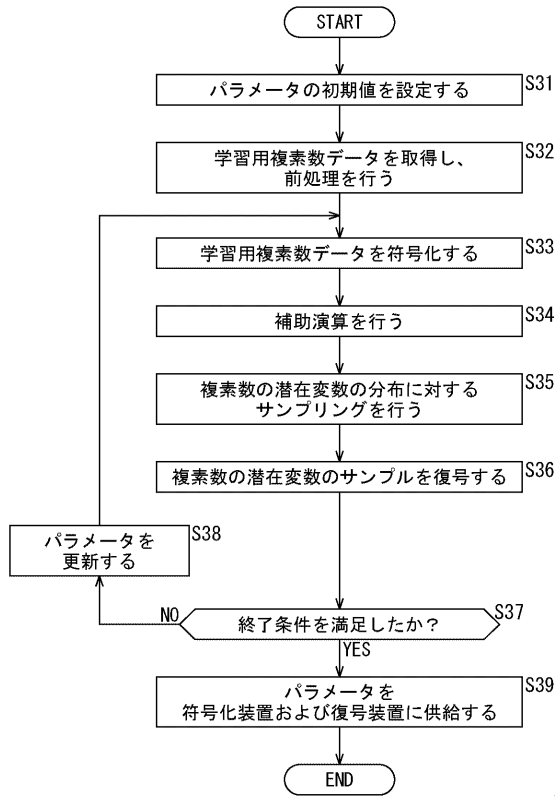
【図8】
FIG. 8



【図9】
FIG. 9



【 図 1 0 】
FIG. 10



【 図 1 1 】
FIG. 11

各手法の実験条件

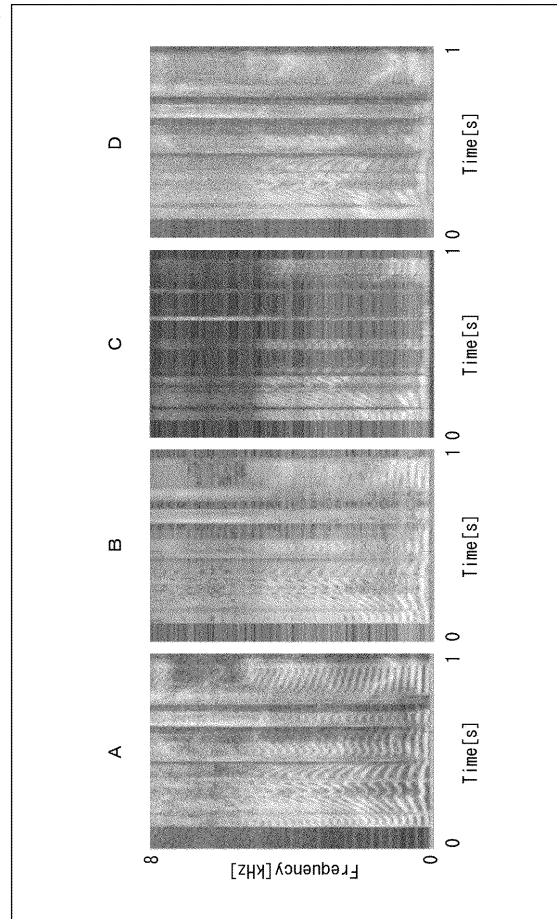
	VAE (GL)	VAE (R+I)	CVAE
input features	amplitude spectra	real & imag of complex spectra	complex spectra
speech reconstruction	Griffin-Lim	inverse SIFT	inverse SIFT
#epoch	290	540	80
optimizer	Adam	Adam	CAdam
learning rate	0.001	0.001	0.0001
encoder architecture	255-100-[50, 100]	510-200-[100, 100]	255-100-[50, 50]
decoder architecture	50-100-255	100-200-510	50-100-255

【 図 1 2 】
FIG. 12

各手法によるPESQ評価

Method	PESQ
VAE (GL)	1.90
VAE (R+I)	1.80
CVAE	2.44
CVAE (w/o δ)	2.39

【 図 1 3 】
FIG. 13



【図 14】
FIG. 14

