

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2025年2月27日(27.02.2025)

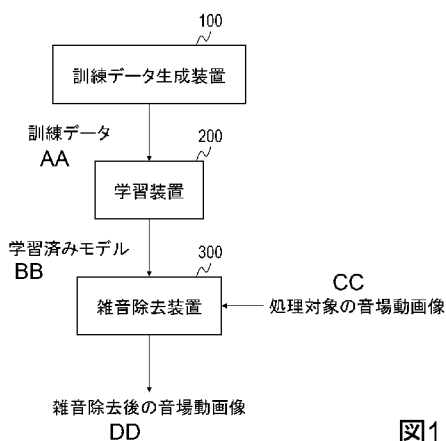


(10) 国際公開番号
WO 2025/041282 A1

- (51) 国際特許分類:
H04S 7/00 (2006.01)
- (21) 国際出願番号: PCT/JP2023/030226
- (22) 国際出願日: 2023年8月23日(23.08.2023)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).
- (72) 発明者: 石川 憲治 (ISHIKAWA, Kenji); 〒1808585 東京都武蔵野市緑町三丁目9番11号 NTT 知的財産センタ内 Tokyo (JP). 竹内 大起 (TAKEUCHI, Daiki); 〒1808585 東京都武蔵野市緑町三丁目9番11号 NTT 知的財産センタ内 Tokyo (JP). 原田 登 (HARADA, Noboru); 〒1808585 東京都武蔵野市緑町三丁目9番11号 NTT 知的財産センタ内 Tokyo (JP). 守谷 健弘 (MORIYA, Takehiro); 〒1808585 東京都武蔵野市緑町三丁目9番11号 NTT 知的財産センタ内 Tokyo (JP).
- (74) 代理人: 中尾 直樹, 外 (NAKAO, Naoki et al.); 〒1600022 東京都新宿区新宿三丁目1番22号 新宿NSビル6階 Tokyo (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(54) Title: LEARNING DEVICE, NOISE REMOVAL DEVICE, AND TRAINING DATA GENERATION DEVICE

(54) 発明の名称: 学習装置、雑音除去装置、および訓練データ生成装置



- 100 Training data generation device
- 200 Learning device
- 300 Noise removal device
- AA Training data
- BB Trained model
- CC Sound field video image to be processed
- DD Sound field video image after noise removal

図1

(57) Abstract: The present invention provides a noise removal device and the like that can perform noise removal on sound field video images with higher precision than the prior art and perform favorable noise removal even on data of an unknown nature or data containing nonlinear noise. The noise removal device is formed from a DNN, accepts a sound field image as input, and outputs a sound field image from which noise included in the sound field image has been removed.

(57) 要約: 従来技術よりも高精度に音場動画の雑音除去を行い、性質が未知のデータや、非線形雑音を含むデータに対しても良好な雑音除去を行うことができる雑音除去装置等を提供する。雑音除去装置は、DNNからなり、音場画像を入力とし、音場画像に含まれるノイズを除去した音場画像を出力する。

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一 国際調査報告 (条約第21条(3))

明 細 書

発明の名称：

学習装置、雑音除去装置、および訓練データ生成装置

技術分野

[0001] 本発明は、音場イメージング技術に関する。

背景技術

[0002] 音場イメージングは、空間中の音を測定し、動画像として可視化する技術であり、マイクロホンアレイや光計測装置等を用いて行われる。音場動画像は、動画像の各ピクセルの値が測定空間中の音の振幅に対応する値となった離散データであり、音の時空間的な挙動や音響現象を観測および測定するのに用いられる。

[0003] 音場動画像に関する信号処理タスクのひとつに雑音除去がある。非特許文献1では、音場動画像に対して時空間周波数フィルタリングを用いて雑音除去を行う。測定対象である音の物理的な性質に合わせて設計したフィルタを用いることで、雑音を除去し、音信号を強調することが可能である。

先行技術文献

非特許文献

[0004] 非特許文献1：N. Chitanont, K. Yatabe, K. Ishikawa, and Y. Oikawa, "Spatio-temporal filter bank for visualizing audible sound field by schlieren method", Appl. Acoust. 115, 109-120 (2017).

発明の概要

発明が解決しようとする課題

[0005] しかしながら、従来技術には、雑音除去性能に一定の限界があり、事前に音信号の性質を知っている必要があり、非線形雑音の除去には適さないなどの課題がある。

[0006] 本発明は、従来技術よりも高精度に音場動画像の雑音除去を行い、性質が未知のデータや、非線形雑音を含むデータに対しても良好な雑音除去を行う

ことができる雑音除去装置、雑音除去装置で用いるモデルの学習装置、学習装置で用いる訓練データを生成する訓練データ生成装置を提供することを目的とする。

課題を解決するための手段

- [0007] 上記の課題を解決するために、本発明の一態様によれば、雑音除去装置は、DNNからなり、音場画像を入力とし、その音場画像に含まれるノイズを除去した音場画像を出力する。
- [0008] 上記の課題を解決するために、本発明の他の態様によれば、学習装置は、DNNからなり、音場画像に対する複素振幅音場画像に基づく情報を入力とし、音場画像に含まれるノイズを除去した音場画像に対する複素振幅音場画像に基づく情報を出力する雑音除去モデルを用いて、訓練データに含まれる学習用入力データに対する音場画像からノイズを除去し、ノイズ除去後の学習用入力データを得る雑音除去部と、訓練データに含まれる正解データとノイズ除去後の学習用入力データとの誤差が小さくなるように、雑音除去モデルのパラメータを更新する更新部と、を含む。
- [0009] 上記の課題を解決するために、本発明の他の態様によれば、訓練データ生成装置は、正解データを生成する際に用いるパラメータを生成するパラメータ生成部と、パラメータを用いて音響シミュレーションにより、正解データを生成し、音場動画像を取得する計測システムに合わせて、任意のノイズを付加した正解データを学習用入力データとし、正解データと学習用入力データとの組からなる訓練データを得る訓練データ生成部とを含み、訓練データは、DNNからなり、音場画像に対する複素振幅音場画像に基づく情報を入力とし、音場画像に含まれるノイズを除去した音場画像に対する複素振幅音場画像に基づく情報を出力する雑音除去モデルを学習する際に用いられる。

発明の効果

- [0010] 本発明によれば、音場動画像の雑音除去を従来技術よりも高精度にすることができ、性質が未知のデータや、非線形雑音を含むデータに対しても良好な雑音除去を行うことができるという効果を奏する。

図面の簡単な説明

- [0011] [図1]第一実施形態に係る雑音除去システムの機能ブロック図。
[図2]第一実施形態に係る雑音除去システムの処理フローの例を示す図。
[図3]訓練データ生成装置の機能ブロック図。
[図4]訓練データ生成装置の処理フローの例を示す図。
[図5]訓練データを生成する際のシミュレーションの模式図。
[図6]学習装置の機能ブロック図。
[図7]学習装置の処理フローの例を示す図。
[図8]雑音除去装置の機能ブロック図。
[図9]雑音除去装置の処理フローの例を示す図。
[図10]雑音除去結果の例を示す図。
[図11]本手法を適用するコンピュータの構成例を示す図。

発明を実施するための形態

- [0012] 以下、本発明の実施形態について、説明する。なお、以下の説明に用いる図面では、同じ機能を持つ構成部や同じ処理を行うステップには同一の符号を記し、重複説明を省略する。以下の説明において、テキスト中で使用する記号「[^]」等は、本来直前の文字の真上に記載されるべきものであるが、テキスト記法の制限により、当該文字の直後に記載する。式中においてはこれらの記号は本来の位置に記述している。また、ベクトルや行列の各要素単位で行われる処理は、特に断りが無い限り、そのベクトルやその行列の全ての要素に対して適用されるものとする。

- [0013] <第一実施形態のポイント>

本実施形態では、DNN(Deep Neural Network)を用いて、高精度な雑音除去を実現する。DNNを用いることで、性質が未知のデータや、非線形雑音を含むデータに対しても良好な雑音除去を行うことができる。

- [0014] 本実施形態では、音場動画像をフーリエ変換によって複素振幅として表現し、DNNで扱いやすくするために複素振幅で表現される音場画像（以下、複素振幅音場画像ともいう）を実部と虚部からなる2チャンネル画像とみなす。こ

れによって、任意の画像ノイズ除去用DNNを音場雑音除去の目的に用いることができる。

[0015] DNNを学習するためには、大量の訓練データを必要とするが、音場動画像に関して、実験で雑音を含まないデータと雑音を含むデータの対を大量に収集することは極めて困難である。そこで、本実施形態では、音響シミュレーションを用いて訓練データを生成する。

[0016] <第一実施形態>

図1は第一実施形態に係る雑音除去システムの構成例を示す。図2は雑音除去システムの処理フローを示す。

[0017] 雑音除去システムは、訓練データ生成装置100、学習装置200および雑音除去装置300を含む。

[0018] 訓練データ生成装置100は、正解データと学習用入力データとの組からなる訓練データを生成し(S100)、出力する。なお、正解データは音響シミュレーションを用いて生成したノイズを含まないデータであり、学習用入力データはノイズを含むデータである。

[0019] 学習装置200は、訓練データを入力とし、雑音除去モデルを学習し(S200)、学習済みの雑音除去モデルを出力する。雑音除去モデルは、DNNからなり、音場画像に対する複素振幅音場画像に基づく情報を入力とし、音場画像に含まれるノイズを除去した音場画像に対する複素振幅音場画像に基づく情報を入力するモデルである。

[0020] 雑音除去装置300は、雑音除去処理の前に、学習済みの雑音除去モデルを受け取る。雑音除去装置300は、処理対象の音場動画像を入力とし、音場動画像に含まれる雑音を除去し(S300)、雑音除去後の音場動画像を出力する。音場動画像は、時系列の音場画像からなり、雑音除去装置300は、処理対象の音場動画像を構成する音場画像を入力とし、音場動画像を構成する音場画像に含まれる雑音を除去し、雑音除去後の音場動画像を構成する音場画像を出力するとも言える。

[0021] 訓練データ生成装置100、学習装置200および雑音除去装置300は

、例えば、中央演算処理装置（CPU: Central Processing Unit）、主記憶装置（RAM: Random Access Memory）などを有する公知又は専用のコンピュータに特別なプログラムが読み込まれて構成された特別な装置である。訓練データ生成装置100、学習装置200および雑音除去装置300は、例えば、中央演算処理装置の制御のもとで各処理を実行する。訓練データ生成装置100、学習装置200および雑音除去装置300に入力されたデータや各処理で得られたデータは、例えば、主記憶装置に格納され、主記憶装置に格納されたデータは必要に応じて中央演算処理装置へ読み出されて他の処理に利用される。訓練データ生成装置100、学習装置200および雑音除去装置300の各処理部は、少なくとも一部が集積回路等のハードウェアによって構成されていてもよい。訓練データ生成装置100、学習装置200および雑音除去装置300が備える各記憶部は、例えば、RAM (Random Access Memory) などの主記憶装置、またはリレーショナルデータベースやキーバリューストアなどのミドルウェアにより構成することができる。ただし、各記憶部は、必ずしも訓練データ生成装置100、学習装置200および雑音除去装置300がその内部に備える必要はなく、ハードディスクや光ディスクもしくはフラッシュメモリ (Flash Memory) のような半導体メモリ素子により構成される補助記憶装置により構成し、訓練データ生成装置100、学習装置200および雑音除去装置300の外部に備える構成としてもよい。

[0022] 以下、各装置について説明する。

[0023] <訓練データ生成装置100>

図3は訓練データ生成装置100の機能ブロック図を、図4はその処理フローの例を示す。

[0024] 訓練データ生成装置100は、パラメータ生成部110、正解データ生成部120、ノイズ生成部130およびノイズ付加部140を含む。

[0025] <パラメータ生成部110および正解データ生成部120>

正解データの生成には、音場データを数値的に計算することのできる任意の理論式・数値シミュレーション方法を用いることができる。データ作成に

際して、生成手法の任意のパラメータを所望の確率分布から決定することで多様な学習データを生成する。一例として次のような方法がある。

[0026] パラメータをランダム化した2次元音場シミュレーションを使用する。図5はシミュレーションの模式図である。内側の長方形が測定領域で、その外側が音源領域となり、点音源がランダムに配置されている。単純なものから複雑なものまで多様な空間特性を持つ音場を生成するために、点音源の数は1～5個とし、各音源の位置と相対振幅をランダムに割り当てる。正解データは、これらの点音源から発生する音波の重ね合わせであり、次の式で計算される。

[数1]

$$p(r, k) = A \sum_{i=1}^N a_i \frac{j}{4} H_0^{(2)}(k|r_i - r|) \quad (1)$$

ここで、 $r=(x, y)$ は位置、 k は音響波数の大きさ、 A は音場全体の大きさを決める定数、 N は音源の数、 a_i と $r_i=(x_i, y_i)$ はそれぞれ i 番目の音源の相対振幅と位置、 $H_0^{(2)}$ は0次第2種ハンケル関数である。総和記号の内部は i 番目の音源の相対振幅と2次元ヘルムホルツ方程式のグリーン関数の積を表している。パラメータを所望の確率分布から生成し、上式に基づいて各点の音圧を計算することで、正解データを生成する。

[0027] パラメータ生成部110は、正解データを生成する際に用いるパラメータを生成し(S110)、出力する。パラメータとしては、各音源の相対振幅 a_i と位置 r_i 等が考えられる。例えば、音源領域の中から所望の確率分布に従ってランダムに各音源の位置を割り当て、想定される相対振幅の範囲の中から所望の確率分布に従ってランダムに各音源に相対振幅を割り当てる。なお、測定領域、音源領域および点音源の個数は予め与えられているものとする。

[0028] 正解データ生成部120は、パラメータ生成部110で生成されたパラメータを入力とし、パラメータを用いて音響シミュレーションにより、正解デ

ータを生成し（S 1 2 0）、出力する。例えば、音源の個数 N を1～5に変えながら、各音源の相対振幅 a_i と位置 r_i を入力とし、式(1)により、測定領域の各点における音圧を求め、正解データを生成する。

[0029] <ノイズ生成部130およびノイズ付加部140>

音場動画像を取得する計測システムに合わせて、任意のノイズが与えることができる。例えば、加法性ガウスホワイトノイズである。所望の振幅を有するホワイトノイズ生成し、正解データに加算することで、ノイズありデータを生成する。

[0030] ノイズ生成部130は、音場動画像を取得する計測システムに合わせて、任意のノイズを生成し（S 1 3 0）、出力する。例えば、加法性ガウスホワイトノイズが生成される。

[0031] ノイズ付加部140は、正解データと生成されたノイズを入力とし、正解データに生成されたノイズを付加し（S 1 4 0）、ノイズを付加した正解データを学習用入力データとし、正解データと学習用入力データとの組からなる訓練データを出力する。例えば、ホワイトノイズを加算することで、ノイズを付加する。

正解データ生成部120、ノイズ生成部130およびノイズ付加部140を合わせて訓練データ生成部150ともいい、訓練データ生成部150における処理を訓練データ生成処理S 1 5 0ともいう。

[0032] また、スペックル干渉計と呼ばれる光学システムを用いる音場イメージング結果の雑音除去を考える場合、スペックルノイズを付加することもできる。スペックルノイズデータの生成方法は、例えば参考文献1の方法などを用いることができる。ただしノイズを含むデータの生成方法は任意であり、これらに限定するものではない。

(参考文献1) Q. Fang, H. Xia, Q. Song, M. Zhang, R. Guo, S. Montresor, and P. Picart, "Speckle denoising based on deep learning via a conditional generative adversarial network in digital holographic interferometry", Opt. Express 30, 20666-20683 (2022).

例えば、スペックルノイズデータの生成方法する場合、訓練データ生成部150は、パラメータ生成部110で生成されたパラメータを入力とし、パラメータを用いて音響シミュレーションにより、正解データを生成し、正解データを生成するアルゴリズムの途中でスペックルノイズを付加した正解データを生成する(S150)。

[0033] このような構成とすることで、膨大な手間や時間のかかる実験なしに、音響シミュレーションによって大量の訓練データを生成することができる。

[0034] <学習装置200>

図6は学習装置200の機能ブロック図を、図7はその処理フローの例を示す。

[0035] 学習装置200は、画像変換部210、雑音除去部220および更新部230を含む。

[0036] <画像変換部210>

画像変換部210は、訓練データを入力とする。訓練データに含まれる学習用入力データと正解データは、それぞれ複素振幅音場画像である。複素振幅音場画像の要素は、対応する空間位置とフーリエ周波数における複素振幅を表す。画像変換部210は、各複素振幅音場画像を実部と虚部からなる2チャンネル画像に変換する(S210)。例えば、複素振幅音場画像がK個の要素からなり、 $[a_1, \dots, a_k]$ と表現され、k番目の要素が $a_k = b_k + c_k j$ のとき、画像変換部210は、複素振幅音場画像 $[a_1, \dots, a_k]$ を、実部からなる画像(以下、実部画像ともいう) $[b_1, \dots, b_k]$ と、虚部からなる画像(以下、虚部画像ともいう) $[c_1, \dots, c_k]$ とに変換する。さらに、画像変換部210は、2チャンネル画像(実部画像 $[b_1, \dots, b_k]$ と虚部画像 $[c_1, \dots, c_k]$)を正規化し、出力する。なお、ここで行われる正規化は、データを雑音除去モデルで扱いやすいものに整える処理であり、データに正規化係数を乗じる処理である。

[0037] <雑音除去部220>

雑音除去部220は、訓練データに含まれる学習用入力データを入力とし、雑音除去モデルを用いて、学習用入力データからノイズを除去し(S22

0)、ノイズ除去後の学習用入力データを得、出力する。

[0038] 本実施形態では、雑音除去部220は、訓練データに含まれる学習用入力データ(複素振幅音場画像)を変換した2チャンネル画像(実部画像 $[b_{T,1}, \dots, b_{T,K}]$ と虚部画像 $[c_{T,1}, \dots, c_{T,K}]$)を入力とし、雑音除去モデルを用いて、2チャンネル画像(実部画像 $[b_{T,1}, \dots, b_{T,K}]$ と虚部画像 $[c_{T,1}, \dots, c_{T,K}]$)からノイズを除去し、ノイズ除去後の2チャンネル画像(実部画像 $[b'_{T,1}, \dots, b'_{T,K}]$ と虚部画像 $[c'_{T,1}, \dots, c'_{T,K}]$)を得る。なお、雑音除去モデルは、前述の通り、DNNからなり、音場画像に対する複素振幅音場画像に基づく情報を入力とし、音場画像に含まれるノイズを除去した音場画像に対する複素振幅音場画像に基づく情報を出力するモデルである。本実施形態では、雑音除去モデルは、実部画像と虚部画像を正規化したものを入力とし、実部画像と虚部画像に含まれるノイズを除去し、ノイズ除去後の実部画像と虚部画像を出力するモデルである。

[0039] 本実施形態では、2チャンネルの入力画像のノイズ除去にDNNを用いるため、画像のノイズ除去を行うネットワークであれば利用可能である。例えば、U-net構造のネットワークが使える。また、画像ノイズ除去に優れた性能を持ち、比較的小さなメモリと学習時間で動作可能なNAFNet (Nonlinear activation free network) (参考文献2参照)を使うこともできる。

(参考文献2) L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration", in Proceedings of the European Conference on Computer Vision (ECCV), (Springer Nature Switzerland, Cham, 2022), pp. 17-33

[0040] <更新部230>

更新部230は、訓練データに含まれる正解データとノイズ除去後の学習用入力データとを入力とし、訓練データに含まれる正解データと、ノイズ除去後の学習用入力データとの誤差が小さくなるように、雑音除去モデルのパラメータを更新する(S230)。

[0041] 本実施形態では、訓練データに含まれる正解データ(複素振幅音場画像)を2チャンネル画像に変換し、正規化したものと、ノイズ除去後の2チャンネル

ル画像（実部画像 $[b'_{T,1}, \dots, b'_{T,K}]$ と虚部画像 $[c'_{T,1}, \dots, c'_{T,K}]$ ）との誤差が小さくなるように、雑音除去モデルのパラメータを更新する。例えば、ロスにはRMSE(Root Mean Square Error)を用い、最適化アルゴリズムには最急降下法、Adam等を用いることができる。

[0042] 更新部230は、所定の条件を満たさない場合（S230-1のNO）、更新後のパラメータを雑音除去部220に出力し、S210~S230を繰り返す。

[0043] 所定の条件を満たす場合（S230-1のYES）、更新部230は、その時点の雑音除去モデルを学習済みの雑音除去モデルとして出力する。

[0044] 所定の条件としては、パラメータの更新が収束したか否かを判断するための条件を設定すればよく、例えば、学習を一定回数（例えば数回）繰り返したか?更新前後のパラメータの差分が所定の閾値以下か?損失が所定の閾値以下か?などの条件を利用できる。

[0045] このような学習方法を採用することで、任意の画像ノイズ除去用DNNを音場雑音除去の目的に用いることができる。

[0046] <雑音除去装置300>

図8は雑音除去装置300の機能ブロック図を、図9はその処理フローの例を示す。

[0047] 雑音除去装置300は、フーリエ変換部310、画像変換部320、雑音除去部330、画像逆変換部340および逆フーリエ変換部350を含む。

[0048] 雑音除去装置300は、学習済みの雑音除去モデルを雑音除去処理時に雑音除去部330で用いる。

[0049] <フーリエ変換部310>

フーリエ変換部310は、処理対象の音場動画像を入力とし、音場動画像の全画素に対して時間領域フーリエ変換を行い（S310）、周波数毎の複素振幅音場画像を得、出力する。フーリエ変換としては、高速フーリエ変換等を用いることができる。

[0050] <画像変換部320>

画像変換部320は、周波数毎の複素振幅音場画像を入力とし、複素振幅音場画像を実部と虚部からなる2チャンネル画像に変換し(S320)、2チャンネル画像を正規化し、出力する。例えば、画像変換部320は、画像変換部210と同様の処理を行う。

[0051] <雑音除去部330>

雑音除去部330は、処理対象の音場動画像に対する音場画像を入力とし、雑音除去モデルを用いて、処理対象の音場動画像からノイズを除去し(S330)、ノイズ除去後の処理対象の音場動画像に対する音場画像を得、出力する。

[0052] 本実施形態では、雑音除去部330は、処理対象の音場動画像に対する複素振幅音場画像を変換した2チャンネル画像（実部画像 $[b_{E,1}, \dots, b_{E,K}]$ と虚部画像 $[c_{E,1}, \dots, c_{E,K}]$ ）を入力とし、雑音除去モデルを用いて、2チャンネル画像からノイズを除去し、ノイズ除去後の2チャンネル画像（実部画像 $[b'_{E,1}, \dots, b'_{E,K}]$ と虚部画像 $[c'_{E,1}, \dots, c'_{E,K}]$ ）を得る。

[0053] <画像逆変換部340>

画像逆変換部340は、ノイズ除去後の2チャンネル画像（実部画像 $[b'_{E,1}, \dots, b'_{E,K}]$ と虚部画像 $[c'_{E,1}, \dots, c'_{E,K}]$ ）を入力とし、音場の大きさを維持するために、ノイズ除去後の2チャンネル画像（実部画像 $[b'_{E,1}, \dots, b'_{E,K}]$ と虚部画像 $[c'_{E,1}, \dots, c'_{E,K}]$ ）に、前述の画像変換部210で乗じた正規化係数の逆数を乗じる。次に、画像逆変換部340は、ノイズ除去後の2チャンネル画像（実部画像 $[b'_{E,1}, \dots, b'_{E,K}]$ と虚部画像 $[c'_{E,1}, \dots, c'_{E,K}]$ ）を複素振幅音場画像 $[a'_{E,1}, \dots, a'_{E,K}]$ に変換し(S340)、出力する。例えば、複素振幅音場画像がK個の要素からなり、 $[a_1, \dots, a_K]$ と表現され、k番目の要素が $a_k = b_k + c_k j$ のとき、画像逆変換部340は、実部からなる画像(以下、実部画像ともいう) $[b_1, \dots, b_K]$ と、虚部からなる画像(以下、虚部画像ともいう) $[c_1, \dots, c_K]$ とを、複素振幅音場画像 $[a_1, \dots, a_K]$ に変換する。

[0054] <逆フーリエ変換部350>

逆フーリエ変換部350は、複素振幅音場画像 $[a'_{E,1}, \dots, a'_{E,K}]$ を入力とし、

時間方向に逆フーリエ変換し（S350）、雑音除去された音場動画像を得、出力する。逆フーリエ変換としては、前述のフーリエ変換部310で行ったフーリエ変換に対応する変換方法を用いればよい。

[0055] <効果>

以上の構成により、音場動画像の雑音除去を従来技術よりも高精度にすることができ、性質が未知のデータや、非線形雑音を含むデータに対しても良好な雑音除去を行うことができる。

[0056] <雑音除去結果>

次に本実施形態に基づく実装例および雑音除去結果例を示す。

[0057] 画像チャンネル数を除き、参考文献1と同じネットワークであるNAFNetを使用した。ネットワークは幅32のブロックからなり、画像チャンネル数は2（実部画像、虚部画像）、画像サイズは128 x 128とした。ロスにはRMSE（root mean square error）を、最適化アルゴリズムにはAdamを用い、学習率は0.001に設定した。学習データは、音源数ごとに400個ずつ、合計2,000個作成した。トレーニングバッチサイズは32、エポックは50であった。雑音は加法性白色ガウス雑音とした。

[0058] 図10は、雑音除去結果例を示す。本実施形態に係る雑音除去によって、雑音ありデータに含まれる雑音が著しく除去されていることが確認できる。

[0059] <変形例1>

第一実施形態と異なる部分を中心に説明する。

第一実施形態では、雑音除去装置300は、処理対象の音場動画像を入力とし、音場動画像に含まれる雑音を除去し（S300）、雑音除去後の音場動画像を出力する。しかしながら、処理対象は、音場動画像ではなく、一枚の音場画像であってもよい。この場合、雑音除去装置300は、処理対象の音場画像を入力とし、音場画像に含まれる雑音を除去し（S300）、雑音除去後の音場画像を出力する。処理対象の音場画像は、1枚の実数値の音場画像（例えば、グレースケール画像）であり、雑音除去装置300は、フーリエ変換部310、画像変換部320、画像逆変換部340および逆フーリエ変

換部350を含まなくともよく、少なくとも雑音除去部330を含めばよい。なお、必要に応じて正規化処理（データを雑音除去モデルで扱いやすいものに整える処理）を行う。

学習装置200の雑音除去部220は、訓練データに含まれる学習用入力データを入力とし、雑音除去モデルを用いて、学習用入力データからノイズを除去し（S220）、ノイズ除去後の学習用入力データを得、出力する。

本変形例では、雑音除去部220は、訓練データに含まれる学習用入力データ（複素振幅音場画像）を変換した2チャンネル画像（実部画像 $[b_{T,1}, \dots, b_{T,K}]$ と虚部画像 $[c_{T,1}, \dots, c_{T,K}]$ ）に含まれる実部画像 $[b_{T,1}, \dots, b_{T,K}]$ を少なくとも入力とし、雑音除去モデルを用いて、実部画像 $[b_{T,1}, \dots, b_{T,K}]$ からノイズを除去し、ノイズ除去後の実部画像 $[b'_{T,1}, \dots, b'_{T,K}]$ を得る。なお、雑音除去モデルは、前述の通り、DNNからなり、音場画像（実部画像）を入力とし、音場画像（実部画像）に含まれるノイズを除去した音場画像（実部画像）を出力するモデルである。ただし、本変形例では、雑音除去モデルは、実部画像を正規化したものを入力とし、実部画像に含まれるノイズを除去し、ノイズ除去後の実部画像を出力するモデルである。

本変形例では、1チャンネルの入力画像（実部画像）のノイズ除去にDNNを用いるため、第一実施形態と同様に画像のノイズ除去を行うネットワークであれば利用可能である。

[0060] <更新部230>

更新部230は、訓練データに含まれる正解データとノイズ除去後の学習用入力データとを入力とし、訓練データに含まれる正解データの実部と、ノイズ除去後の学習用入力データ（実部画像）との誤差が小さくなるように、雑音除去モデルのパラメータを更新する（S230）。

[0061] <変形例2>

第一実施形態と異なる部分を中心に説明する。

本変形例では、雑音除去モデルは、複素振幅音場画像に対応する振幅と位相を入力とし、ノイズ除去後の複素振幅音場画像に対応する振幅と位相を出

力するモデルである。

[0062] <画像変換部210>

学習装置200の画像変換部210は、訓練データを入力とする。訓練データに含まれる学習用入力データと正解データは、それぞれ複素振幅音場画像である。複素振幅音場画像の要素は、対応する空間位置とフーリエ周波数における複素振幅を表す。画像変換部210は、各複素振幅音場画像を振幅と位相に変換する(S210)。画像変換部210は、必要に応じて正規化処理(データを雑音除去モデルで扱いやすいものに整える処理)を行う。

[0063] <雑音除去部220>

学習装置200の雑音除去部220は、訓練データに含まれる学習用入力データを入力とし、雑音除去モデルを用いて、学習用入力データからノイズを除去し(S220)、ノイズ除去後の学習用入力データを得、出力する。

[0064] 本変形例では、雑音除去部220は、訓練データに含まれる学習用入力データ(複素振幅音場画像)を変換した振幅と位相を入力とし、雑音除去モデルを用いて、複素振幅音場画像からノイズを除去し、ノイズ除去後の複素振幅音場画像に対応する振幅と位相を得る。

[0065] 本変形例では、第一実施形態と同様のネットワークを利用することができる。例えば、第一実施形態と同様にUnet構造のネットワークが使える。また、NAFNetを使うこともできる。

[0066] <更新部230>

学習装置200の更新部230は、訓練データに含まれる正解データとノイズ除去後の学習用入力データとを入力とし、訓練データに含まれる正解データと、ノイズ除去後の学習用入力データとの誤差が小さくなるように、雑音除去モデルのパラメータを更新する(S230)。

[0067] 本変形例では、訓練データに含まれる正解データ(複素振幅音場画像)に対応する振幅と位相と、ノイズ除去後の複素振幅音場画像に対応する振幅と位相との誤差が小さくなるように、雑音除去モデルのパラメータを更新する。例えば、第一実施形態と同様の方法によりパラメータを更新することがで

きる。

[0068] <画像変換部 3 2 0>

雑音除去装置 3 0 0 の画像変換部 3 2 0 は、周波数毎の複素振幅音場画像を入力とし、複素振幅音場画像を振幅と位相に変換し (S 3 2 0)、出力する。例えば、画像変換部 3 2 0 は、画像変換部 2 1 0 と同様の処理を行う。

[0069] <雑音除去部 3 3 0>

雑音除去装置 3 0 0 の雑音除去部 3 3 0 は、処理対象の音場動画像に対する音場画像を入力とし、雑音除去モデルを用いて、処理対象の音場動画像からノイズを除去し (S 3 3 0)、ノイズ除去後の処理対象の音場動画像に対する音場画像を得、出力する。

[0070] 本変形例では、雑音除去部 3 3 0 は、処理対象の音場動画像に対する複素振幅音場画像を変換した振幅と位相を入力とし、雑音除去モデルを用いて、複素振幅音場画像からノイズを除去し、ノイズ除去後の複素振幅音場画像に対応する振幅と位相を得る。

[0071] <画像逆変換部 3 4 0>

画像逆変換部 3 4 0 は、ノイズ除去後の複素振幅音場画像に対応する振幅と位相を入力とし、ノイズ除去後の複素振幅音場画像に対応する振幅と位相を複素振幅音場画像に変換し (S 3 4 0)、出力する。画像変換部 2 1 0 において正規化処理を行った場合には、元に戻す処理を行う。

<変形例 3>

第一実施形態と異なる部分を中心に説明する。

本変形例では、雑音除去モデルは、複素数を入力とする DNN (複素ニューラルネットワーク) であり、複素振幅音場画像を入力とし、ノイズ除去後の複素振幅音場画像を出力するモデルである。

[0072] 学習装置 2 0 0 は、雑音除去部 2 2 0 および更新部 2 3 0 を含む。

[0073] <雑音除去部 2 2 0>

雑音除去部 2 2 0 は、訓練データに含まれる学習用入力データを入力とし、雑音除去モデルを用いて、学習用入力データからノイズを除去し (S 2 2

0)、ノイズ除去後の学習用入力データを得、出力する。なお、必要に応じて学習用入力データに対して正規化処理(データを雑音除去モデルで扱いやすいものに整える処理)を行う。

[0074] 本変形例では、雑音除去部220は、訓練データに含まれる学習用入力データ(複素振幅音場画像)を入力とし、雑音除去モデルを用いて、複素振幅音場画像からノイズを除去し、ノイズ除去後の複素振幅音場画像を得る。

[0075] 本変形例では、複素振幅音場画像のノイズ除去にDNNを用いるため、画像のノイズ除去を行う複素ニューラルネットワークであれば利用可能である。

[0076] <更新部230>

更新部230は、訓練データに含まれる正解データとノイズ除去後の学習用入力データとを入力とし、訓練データに含まれる正解データと、ノイズ除去後の学習用入力データとの誤差が小さくなるように、雑音除去モデルのパラメータを更新する(S230)。

[0077] 本変形例では、訓練データに含まれる正解データ(複素振幅音場画像)と、ノイズ除去後の複素振幅音場画像との誤差が小さくなるように、雑音除去モデルのパラメータを更新する。

[0078] <雑音除去装置300>

雑音除去装置300は、フーリエ変換部310、雑音除去部330および逆フーリエ変換部350を含む。

[0079] <雑音除去部330>

雑音除去部330は、処理対象の音場動画像に対する音場画像を入力とし、雑音除去モデルを用いて、処理対象の音場動画像からノイズを除去し(S330)、ノイズ除去後の処理対象の音場動画像に対する音場画像を得、出力する。なお、雑音除去部220と同様に必要に応じて学習用入力データに対して正規化処理(データを雑音除去モデルで扱いやすいものに整える処理)を行い、雑音除去モデルの出力に対して元に戻す処理を行う。

[0080] 本変形例では、雑音除去部330は、処理対象の音場動画像に対する複素振幅音場画像を入力とし、雑音除去モデルを用いて、複素振幅音場画像から

ノイズを除去し、ノイズ除去後の複素振幅音場画像を得る。

[0081] 何れの実施形態および変形例においても、前記雑音除去モデルは、音場画像に対する複素振幅音場画像に基づく情報を入力とし、音場画像に含まれるノイズを除去した音場動画像に対する複素振幅音場画像に基づく情報を出力するモデルと言える。第一実施形態では実部画像と虚部画像が複素振幅音場画像に基づく情報に、変形例1では実部画像が複素振幅音場画像に基づく情報に、変形例2では位相と振幅が複素振幅音場画像に基づく情報に、変形例3では複素振幅音場画像自体が複素振幅音場画像に基づく情報に、相当する。

[0082] また、第一実施形態の画像変換部210を雑音除去部220の前処理を考え、画像変換部210を雑音除去部220の一部と考えてもよい。また、第一実施形態の画像変換部320と画像逆変換部340をそれぞれ雑音除去部330の前処理と後処理と考え、画像変換部320と画像逆変換部340を雑音除去部330の一部と考えてもよい。変形例2においても同様である。

[0083] <その他の変形例>

本発明は上記の実施形態及び変形例に限定されるものではない。例えば、上述の各種の処理は、記載に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的にあるいは個別に実行されてもよい。その他、本発明の趣旨を逸脱しない範囲で適宜変更が可能である。

[0084] <プログラム及び記録媒体>

上述の各種の処理は、図11に示すコンピュータ2000の記録部2020に、上記方法の各ステップを実行させるプログラムを読み込ませ、制御部2010、入力部2030、出力部2040、表示部2050などに動作させることで実施できる。

[0085] この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体

メモリ等のようなものでもよい。

[0086] また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させる構成としてもよい。

[0087] このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶装置に格納する。そして、処理の実行時、このコンピュータは、自己の記録媒体に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実行形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよく、さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆるASP (Application Service Provider) 型のサービスによって、上述の処理を実行する構成としてもよい。なお、本形態におけるプログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの（コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等）を含むものとする。

[0088] また、この形態では、コンピュータ上で所定のプログラムを実行させることにより、本装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

請求の範囲

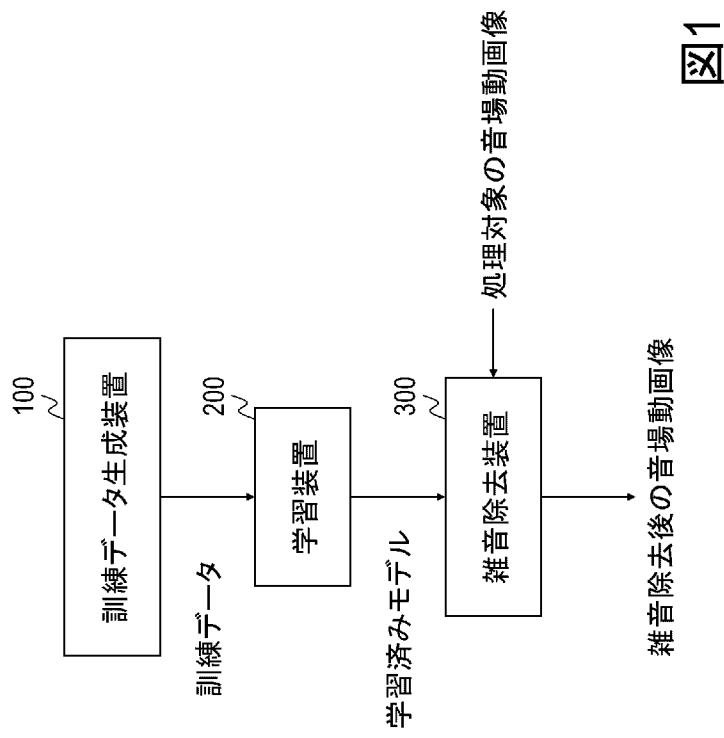
- [請求項1] DNNからなり、音場画像を入力とし、その音場画像に含まれるノイズを除去した音場画像を出力する、
雑音除去装置。
- [請求項2] 請求項1の雑音除去装置であって、
処理対象の音場動画画像の全画素に対して時間領域フーリエ変換を行い、周波数毎の複素振幅音場画像を得るフーリエ変換部と、
DNNからなり、音場画像に対する複素振幅音場画像に基づく情報を入力とし、その音場画像に含まれるノイズを除去した音場画像に対する複素振幅音場画像に基づく情報を出力する雑音除去モデルを用いて、ノイズ除去後の処理対象の音場動画画像に対する複素振幅音場画像を得る雑音除去部と、
前記ノイズ除去後の処理対象の音場動画画像に対する複素振幅音場画像を時間方向に逆フーリエ変換し、雑音除去された音場動画画像を得る逆フーリエ変換部と、を含む、
雑音除去装置。
- [請求項3] DNNからなり、音場画像に対する複素振幅音場画像に基づく情報を入力とし、その音場画像に含まれるノイズを除去した音場画像に対する複素振幅音場画像に基づく情報を出力する雑音除去モデルを用いて、訓練データに含まれる学習用入力データに対する音場画像からノイズを除去し、ノイズ除去後の学習用入力データを得る雑音除去部と、
前記訓練データに含まれる正解データと前記ノイズ除去後の学習用入力データとの誤差が小さくなるように、前記雑音除去モデルのパラメータを更新する更新部と、を含む、
学習装置。
- [請求項4] 正解データを生成する際に用いるパラメータを生成するパラメータ生成部と、
前記パラメータを用いて音響シミュレーションにより、正解データ

を生成し、音場動画像を取得する計測システムに合わせて、任意のノイズを付加した正解データを学習用入力データとし、正解データと学習用入力データとの組からなる訓練データを得る訓練データ生成部とを含み、

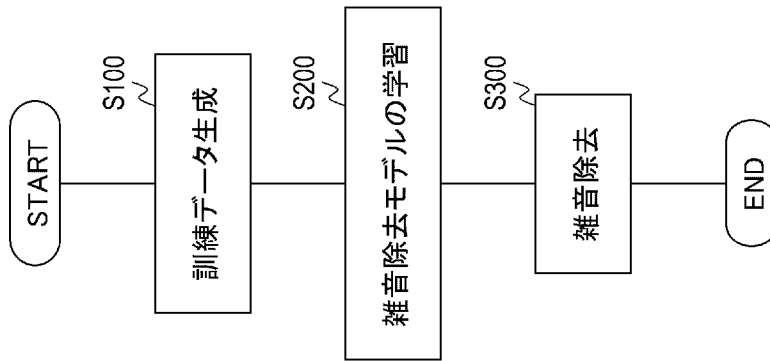
前記訓練データは、DNNからなり、音場画像に対する複素振幅音場画像に基づく情報を入力とし、その音場画像に含まれるノイズを除去した音場画像に対する複素振幅音場画像に基づく情報を出力する雑音除去モデルを学習する際に用いられる、

訓練データ生成装置。

[図1]

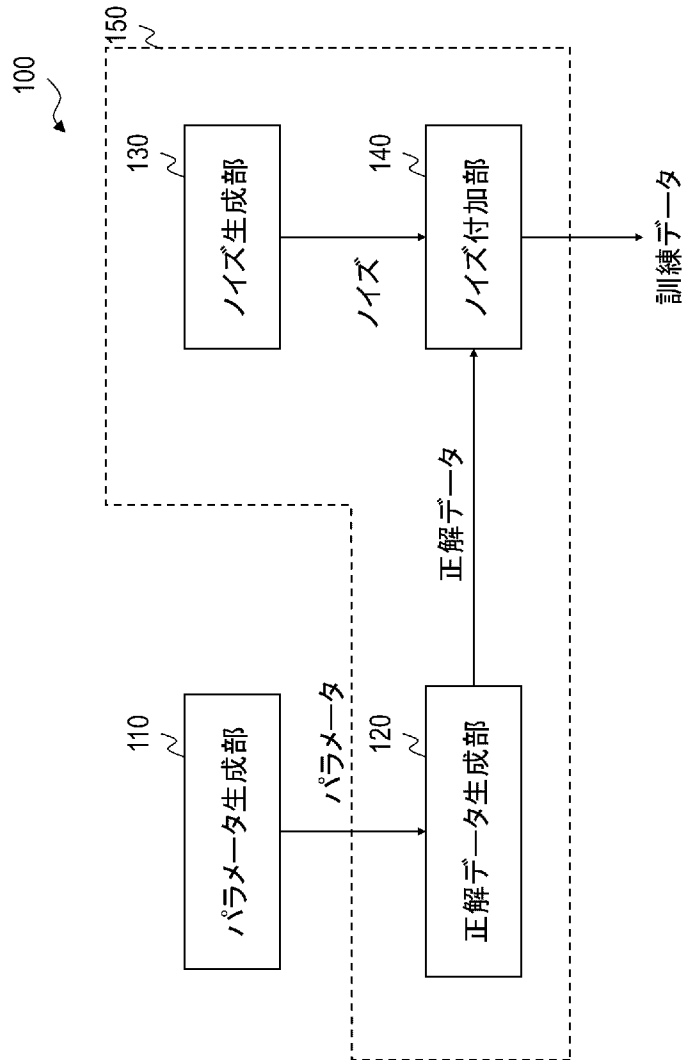


[図2]



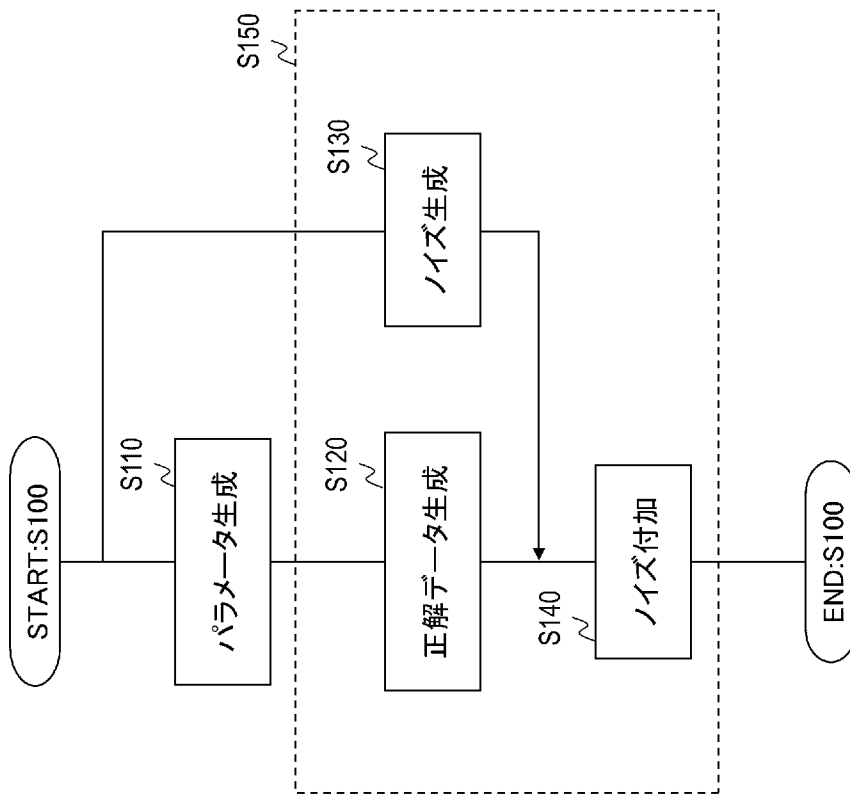
[図2]

[図3]



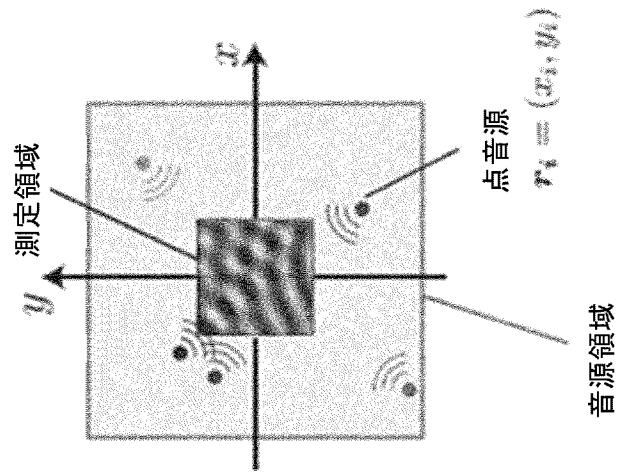
[図3]

[図4]



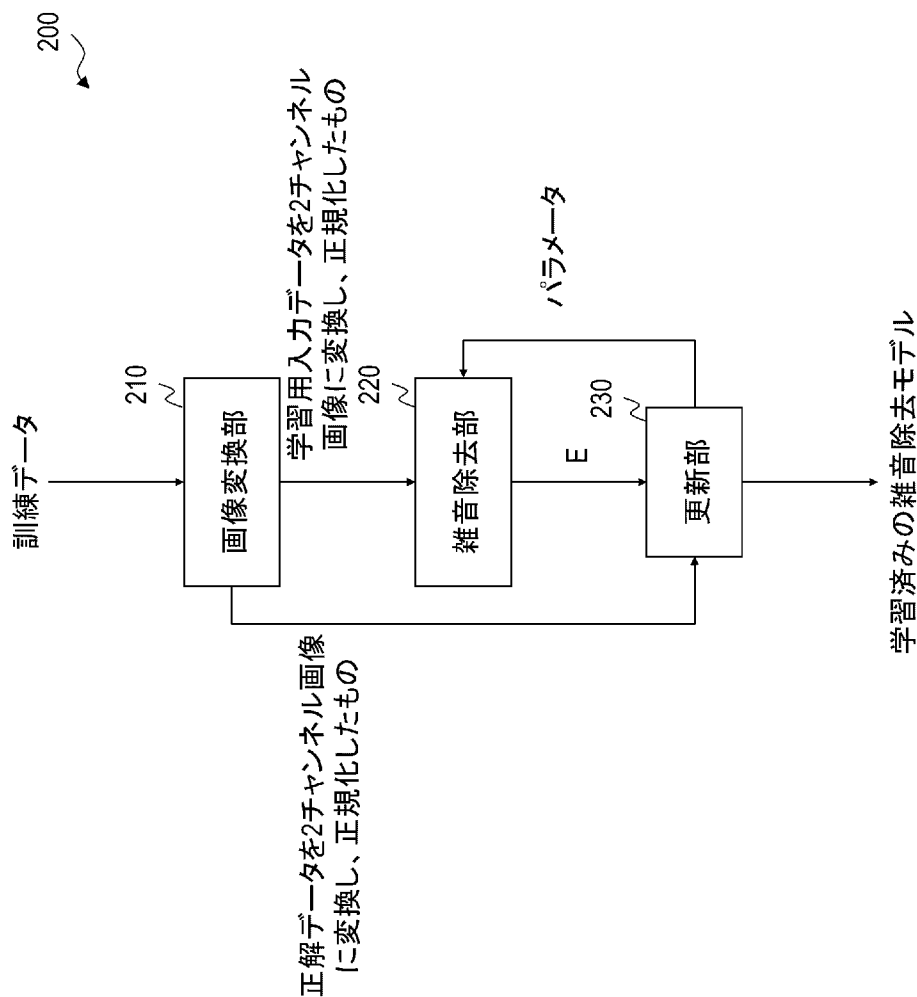
[図4]

[図5]



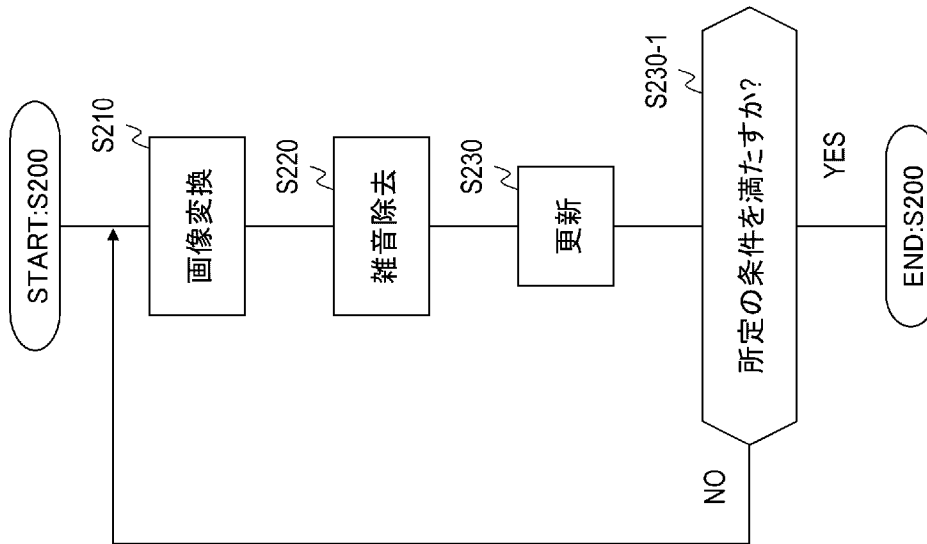
[図5]

[図6]



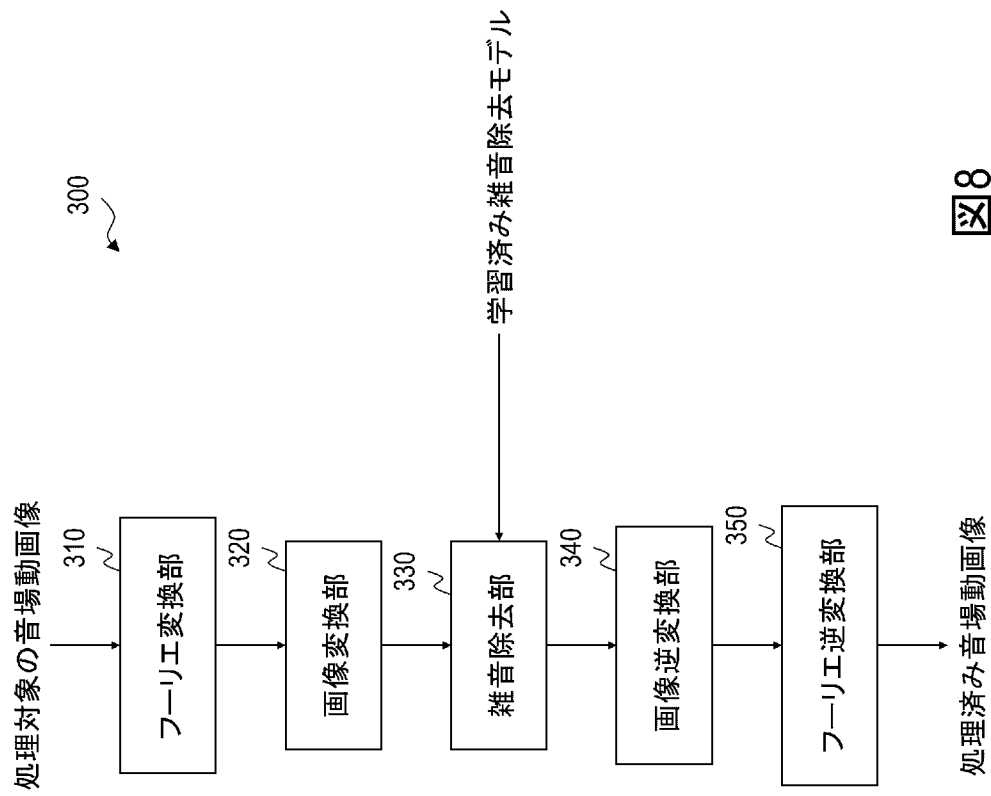
[図6]

[図7]



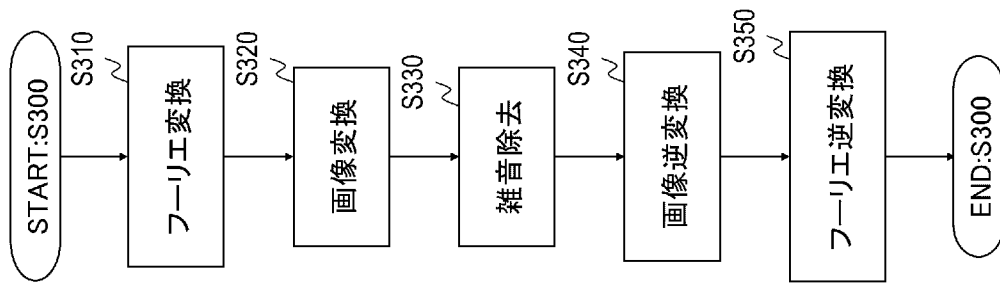
[図7]

[図8]



[図8]

[図9]



[図9]

[図10]

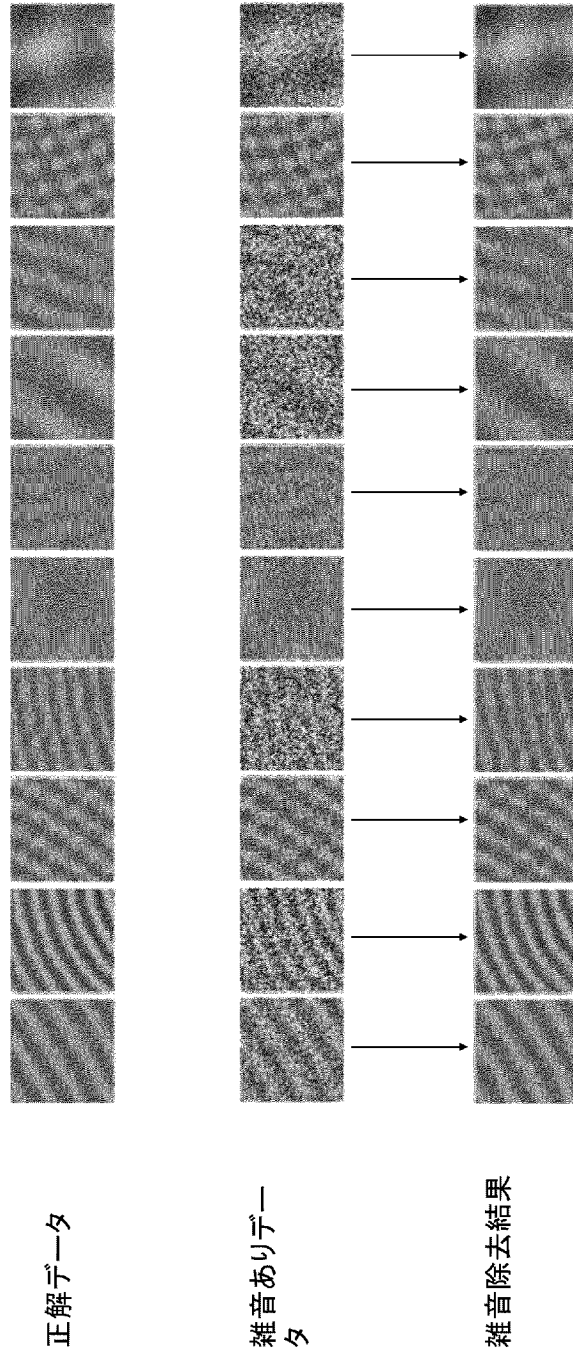


図10

[図11]

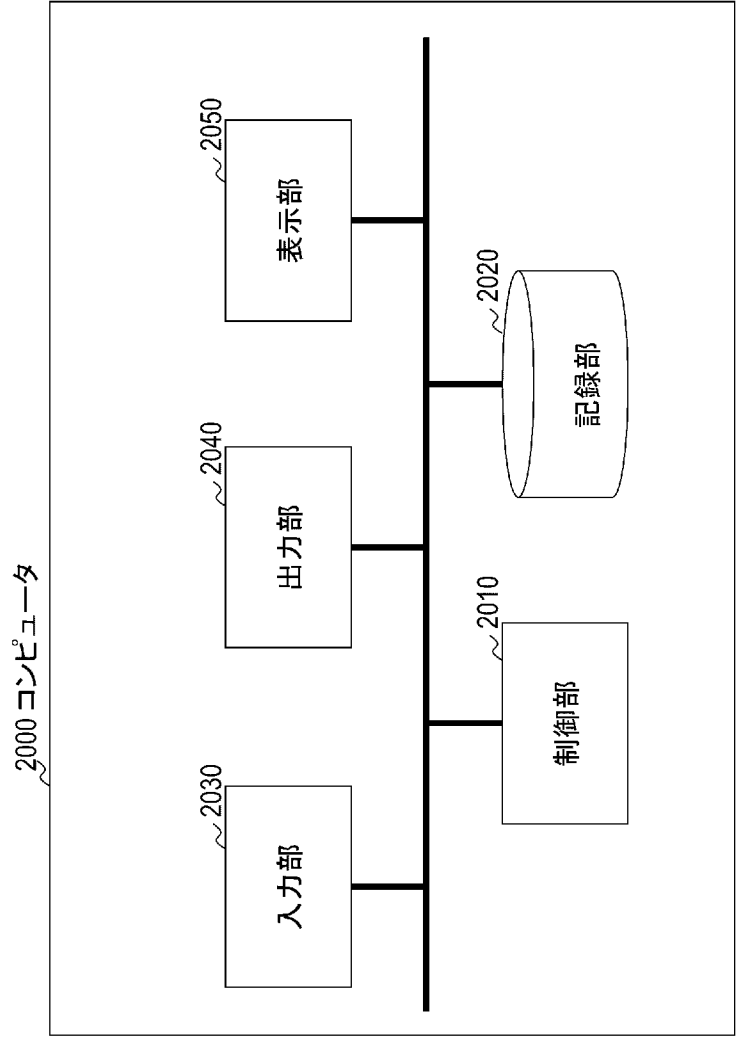


図11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2023/030226

A. CLASSIFICATION OF SUBJECT MATTER		
<i>H04S 7/00</i> (2006.01) FI: H04S7/00 300		
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) H04S7/00		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Published examined utility model applications of Japan 1922-1996 Published unexamined utility model applications of Japan 1971-2023 Registered utility model specifications of Japan 1996-2023 Published registered utility model applications of Japan 1994-2023		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2023-446 A (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 04 January 2023 (2023-01-04) paragraph [0040], fig. 4	1-4
Y	JP 2021-149937 A (INTEL CORPORATION) 27 September 2021 (2021-09-27) paragraphs [0211]-[0213], [0266]-[0269], [0310]	1-4
<input 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 26 October 2023		Date of mailing of the international search report 07 November 2023
Name and mailing address of the ISA/JP Japan Patent Office (ISA/JP) 3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915 Japan		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No. PCT/JP2023/030226

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP 2023-446 A	04 January 2023	(Family: none)	
JP 2021-149937 A	27 September 2021	US 2021/0287417 A1 paragraphs [0212]-[0214], [0267]-[0270], [0311] DE 102020131901 A CN 113409435 A KR 10-2021-0117137 A TW 202139129 A	

A. 発明の属する分野の分類（国際特許分類（IPC）） H04S 7/00(2006.01)i FI: H04S7/00 300		
B. 調査を行った分野 調査を行った最小限資料（国際特許分類（IPC）） H04S7/00 最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922 - 1996年 日本国公開実用新案公報 1971 - 2023年 日本国実用新案登録公報 1996 - 2023年 日本国登録実用新案公報 1994 - 2023年		
国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y	JP 2023-446 A（日本電信電話株式会社）04.01.2023（2023 - 01 - 04） 段落[0040]、図4	1-4
Y	JP 2021-149937 A（インテル・コーポレーション）27.09.2021（2021 - 09 - 27） 段落[0211]-[0213]、[0266]-[0269]、[0310]	1-4
<input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input checked="" type="checkbox"/> パテントファミリーに関する別紙を参照。		
* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献 “T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献		
国際調査を完了した日	国際調査報告の発送日	
26.10.2023	07.11.2023	
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 大石 剛 5Z 4882 電話番号 03-3581-1101 内線 3591	

国際調査報告
 パテントファミリーに関する情報

国際出願番号

PCT/JP2023/030226

引用文献	公表日	パテントファミリー文献	公表日
JP 2023-446 A	04.01.2023	(ファミリーなし)	
JP 2021-149937 A	27.09.2021	US 2021/0287417 A1 段落[0212]-[0214]、 [0267]-[0270]、[0311] DE 102020131901 A CN 113409435 A KR 10-2021-0117137 A TW 202139129 A	