

(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
2024年9月19日(19.09.2024)



(10) 国際公開番号

WO 2024/190063 A1

(51) 国際特許分類:

G10L 21/0272 (2013.01)

(21) 国際出願番号:

PCT/JP2024/000990

(22) 国際出願日:

2024年1月16日(16.01.2024)

(25) 国際出願の言語:

日本語

(26) 国際公開の言語:

日本語

(30) 優先権データ:

特願 2023-039058 2023年3月13日(13.03.2023) JP

(71) 出願人: 日本電気株式会社 (NEC CORPORATION) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).

(72) 発明者: 遠藤 謙光(ENDO, Kenkoh); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 山本 仁(YAMAMOTO, Hitoshi); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). ヤン ホン(YANG, Hong); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP).

(74) 代理人: 江上 達夫, 外(EGAMI, Tatsuo et al.); 〒1040031 東京都中央区京橋一丁目16番10号 VPO 京橋3階 東京セントラル特許事務所内 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.

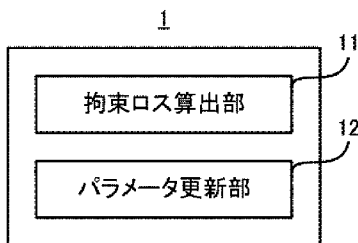
(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE,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

— 国際調査報告 (条約第21条(3))

(54) Title: INFORMATION PROCESSING DEVICE, INFORMATION PROCESSING METHOD, AND RECORDING MEDIUM

(54) 発明の名称: 情報処理装置、情報処理方法、及び、記録媒体



11 Constraint loss calculation unit
12 Parameter update unit

(57) Abstract: This information processing device comprises: a constraint loss calculation unit that calculates constraint loss by using an inference speech emphasis mask outputted by a speech emphasis mask inference model to which noise-mixed speech is inputted, and an inference noise emphasis mask outputted by a noise emphasis mask inference model to which the noise-mixed speech is inputted; and a parameter update unit that updates a parameter included in the speech emphasis mask inference model and a parameter included in the noise emphasis mask inference model by using the constraint loss, speech emphasis mask loss indicating the difference between the inference speech emphasis mask and a target speech emphasis mask, and noise emphasis mask loss indicating the difference between the inference noise emphasis mask and a target noise emphasis mask. The information processing device generates a speech emphasis mask inference model that can perform speech emphasis with high accuracy.

[続葉有]

(57) 要約：情報処理装置は、雑音混合音声が入力された音声強調マスク推定モデルが出力した推定音声強調マスク、及び雑音混合音声が入力された雑音強調マスク推定モデルが出力した推定雑音強調マスクを用いて拘束ロスを算出する拘束ロス算出部と、推定音声強調マスクと目標音声強調マスクとの違いを示す音声強調マスクロス、推定雑音強調マスクと目標雑音強調マスクとの違いを示す雑音強調マスクロス、及び拘束ロスを用いて、音声強調マスク推定モデルが含むパラメータ、及び雑音強調マスク推定モデルが含むパラメータを更新するパラメータ更新部とを備え、精度よく音声強調をすることができる音声強調マスク推定モデルを生成する。

明 細 書

発明の名称： 情報処理装置、情報処理方法、及び、記録媒体
技術分野

[0001] この開示は、情報処理装置、情報処理方法、及び、記録媒体の技術分野に関する。

背景技術

[0002] 非特許文献1には、モデルにより音声強調マスクと雑音強調マスクとを推定し、推定した各マスクを用いて指標を計算し、計算した指標と指標の理想値とのずれを用いてモデルを学習させる技術が記載されている。

先行技術文献

非特許文献

[0003] 非特許文献1：Investigations on Data Augmentation and Loss Functions for Deep Learning Based Speech-Background Separation; Interspeech 2018 ; P. 3499-3503

発明の概要

発明が解決しようとする課題

[0004] この開示は、先行技術文献に記載された技術の改良を目的とする情報処理装置、情報処理方法、及び、記録媒体を提供することを課題とする。

課題を解決するための手段

[0005] 情報処理装置の一の態様は、雑音混合音声が入力された音声強調マスク推定モデルが出力した推定音声強調マスク、及び前記雑音混合音声が入力された雑音強調マスク推定モデルが出力した推定雑音強調マスクを用いて拘束ロスを算出する拘束ロス算出手段と、前記推定音声強調マスクと雑音混合音声の音声と雑音が既知な場合に計算される目標音声強調マスクとの違いを示す音声強調マスキロス、前記推定雑音強調マスクと目標雑音強調マスクとの違いを示す雑音強調マスキロス、及び前記拘束ロスを用いて、前記音声強調マスク推定モデルが含むパラメータ、及び前記雑音強調マスク推定モデルが

含むパラメータを更新するパラメータ更新手段とを備える。

[0006] 情報処理方法の一の態様は、雑音混合音声が入力された音声強調マスク推定モデルが出力した推定音声強調マスク、及び前記雑音混合音声が入力された雑音強調マスク推定モデルが出力した推定雑音強調マスクを用いて拘束ロスを算出し、前記推定音声強調マスクと雑音混合音声の中の音声と雑音が既知な場合に計算される目標音声強調マスクとの違いを示す音声強調マスキロス、前記推定雑音強調マスクと目標雑音強調マスクとの違いを示す雑音強調マスキロス、及び前記拘束ロスをを用いて、前記音声強調マスク推定モデルが含むパラメータ、及び前記雑音強調マスク推定モデルが含むパラメータを更新する。

[0007] 記録媒体の一の態様は、コンピュータに、雑音混合音声が入力された音声強調マスク推定モデルが出力した推定音声強調マスク、及び前記雑音混合音声が入力された雑音強調マスク推定モデルが出力した推定雑音強調マスクを用いて拘束ロスを算出し、前記推定音声強調マスクと雑音混合音声の中の音声と雑音が既知な場合に計算される目標音声強調マスクとの違いを示す音声強調マスキロス、前記推定雑音強調マスクと目標雑音強調マスクとの違いを示す雑音強調マスキロス、及び前記拘束ロスをを用いて、前記音声強調マスク推定モデルが含むパラメータ、及び前記雑音強調マスク推定モデルが含むパラメータを更新する情報処理方法を実行させるためのコンピュータプログラムが記録されている。

図面の簡単な説明

[0008] [図1]図1は、第1実施形態における情報処理装置の構成を示すブロック図である。

[図2]図2は、第2実施形態における情報処理装置の構成を示すブロック図である。

[図3]図3は、第2実施形態における情報処理装置の情報処理動作の流れを示すフローチャートである。

[図4]図4は、第3実施形態における情報処理装置の構成を示すブロック図で

ある。

[図5]図5は、第4実施形態における情報処理装置の構成を示すブロック図である。

[図6]図6は、第5実施形態における情報処理装置の構成を示すブロック図である。

[図7]図7は、第5実施形態における情報処理装置の情報処理動作の流れを示すフローチャートである。

[図8]図8は、第6実施形態における情報処理装置の構成を示すブロック図である。

[図9]図9は、第6実施形態における情報処理装置の情報処理動作の流れを示すフローチャートである。

発明を実施するための形態

[0009] 以下、図面を参照しながら、情報処理装置、情報処理方法、及び、記録媒体の実施形態について説明する。

[1：第1実施形態]

[0010] 情報処理装置、情報処理方法、及び、記録媒体の第1実施形態について説明する。以下では、情報処理装置、情報処理方法、及び記録媒体の第1実施形態が適用された情報処理装置1を用いて、情報処理装置、情報処理方法、及び記録媒体の第1実施形態について説明する。

[1-1：情報処理装置1の構成]

[0011] 図1は、第1実施形態における情報処理装置1の構成を示すブロック図である。図1に示すように、情報処理装置1は、拘束ロス算出部11と、パラメータ更新部12とを備える。

[0012] 拘束ロス算出部11は、雑音混合音声が入力された音声強調マスク推定モデルが出力した推定音声強調マスク、及び雑音混合音声が入力された雑音強調マスク推定モデルが出力した推定雑音強調マスクを用いて拘束ロスを算出する。パラメータ更新部12は、推定音声強調マスクと目標音声強調マスクとの違いを示す音声強調マスクロス、推定雑音強調マスクと目標雑音強調マ

スクとの違いを示す雑音強調マスクロス、及び拘束ロスを用いて、音声強調マスク推定モデルが含むパラメータ、及び雑音強調マスク推定モデルが含むパラメータを更新する。なお、「雑音混合音声」、「音声強調マスク推定モデル」、「推定音声強調マスク」、「雑音強調マスク推定モデル」、「推定雑音強調マスク」、「拘束ロス」、「目標音声強調マスク」、「音声強調マスクロス」、「目標雑音強調マスク」、及び「雑音強調マスクロス」という語句については、後述する他の実施形態において詳しく説明する。

[1 - 2 : 情報処理装置 1 の技術的効果]

[0013] 第 1 実施形態における情報処理装置 1 は、入力音声に含まれる目的音声以外の雑音の時刻、周波数帯域の音量を相対的に減少する音声強調マスク推定モデルが出力した推定音声強調マスク、及び入力音声に含まれる目的音声の時刻、周波数帯域の音量を相対的に減少する雑音強調マスク推定モデルが出力した推定雑音強調マスクを用いて算出した拘束ロスを導入する。情報処理装置 1 は、音声強調マスクの推定値と目標値との違い、雑音強調マスクの推定値と目標値との違い、及び拘束ロスを用いて、音声強調マスク推定モデルが含むパラメータを更新するので、精度よく音声強調をすることができる音声強調マスク推定モデルを生成することができる。

[2 : 第 2 実施形態]

[0014] 続いて、情報処理装置、情報処理方法、及び記録媒体の第 2 実施形態について説明する。以下では、情報処理装置、情報処理方法、及び記録媒体の第 2 実施形態が適用された情報処理装置 2 を用いて、情報処理装置、情報処理方法、及び記録媒体の第 2 実施形態について説明する。

[2 - 1 - 1 : 音声と雑音]

[0015] 本実施形態において、「音声」とは、対象とする音信号を表す。本実施形態において、「音声」を「対象信号」と称する場合がある。また、本実施形態において、「雑音」とは、対象としない音信号を表す。本実施形態において、「雑音」を「対象外信号」と称する場合がある。

[0016] 音声は、注目したい音であってもよい。音声は、人物の声であってもよい

。音声は、特定の人物の声であってもよい。特定の人物は、1以上の人物であってもよい。特定の人物は、マイク等の音を取得する機構の近くに居る人物であってもよい。「音声」は、場面に応じて異なる音信号を表していてもよい。

[0017] 本実施形態において、「雑音混合音声」とは、「雑音」が混合されている「音声」である。「雑音混合音声」は、対象とする音信号である「音声」と、対象としない音信号である「雑音」とを含む。「雑音混合音声」に含まれる音信号は、「音声」、及び「雑音」の何れかである。言い換えると、「雑音混合音声」に含まれる「音声」ではない音信号は「雑音」である。また、「雑音混合音声」に含まれる「雑音」ではない音信号は「音声」である。また、本実施形態において、情報処理装置2に入力される入力信号は、「雑音混合音声」である。すなわち、情報処理装置2に入力される入力信号は、対象信号である「音声」と、対象外信号である「雑音」からなる。

[2-1-2：音声強調技術]

[0018] 音声に混入している音声以外の雑音をマイクロフォンが拾う場合がある。この場合、音声に対して雑音が相対的に大きければ大きいほど、音声は聞き取りにくくなる。音声強調技術は、雑音混合音声から音声を雑音に対して相対的に大きくする技術である。本技術は、雑音が混合されている音声から雑音を抑制して音声を強調し、より聞き取り易い音声を提供する技術であってもよい。雑音が混合されている音声から、音声のみを強調し、より聞き取り易い音声を提供する技術であってもよい。音声強調技術は、高雑音の状況下において、通話音声の雑音を除去することができてもよい。また、音声強調技術は、通話相手の音声を強調して、より円滑なコミュニケーションを可能にしてもよい。また、音声強調技術は、音声認識器の認識率を向上させてもよい。

[0019] また、マスク型音声強調技術は、雑音の音量を下げる時刻、周波数帯、音量の減少量を指定する音声強調マスクを推定する。マスク型音声強調技術は、雑音が混合されている雑音混合音声に当該音声強調マスクを施すことで、

音声を聞き取り易くする技術であってもよい。本実施形態では、マスク型音声強調技術においては、音声強調マスクを推定する音声強調マスク推定モデルを機械学習してもよい。本実施形態において作成された音声強調マスク推定モデルは、音声を強調する場面に適用されてもよい。

[0020] 本実施形態において作成される音声強調マスク推定モデルは、クリーンな音声に雑音を重畳した雑音混合音声を入力信号として用いて学習をされる。クリーンな音声とは、雑音が極めて少ない音であってもよい。本実施形態における機械学習は、深層学習であってもよい。

[2-1-3：音声強調マスク、及び雑音強調マスク]

[0021] 音声強調マスク推定モデルは、雑音混合音声が入力されると音声強調マスクを出力するモデルである。音声強調マスク推定モデルが推定する音声強調マスクを推定音声強調マスクと呼んでもよい。推定音声強調マスクを推定値と呼ぶ場合がある。音声強調マスクは、下記式1のように表現してもよい。

[式1]

$$M_S(f, t) = \frac{S(f, t)^2}{S(f, t)^2 + N(f, t)^2}$$

[0022] $S(t, f)^2$ は、「音声」のパワーを示す。また、 $N(t, f)^2$ は、「雑音」のパワーを示す。

[0023] 音声強調マスクは、「雑音混合音声」のパワーのうちの「音声」のパワーの割合で表現する。

[0024] 上述の通り、本実施形態は、音声を強調することを目的として音声強調マスク推定モデルを学習をさせて作成するが、音声強調と共に雑音強調も学習させることで、音声強調マスク推定モデルの学習を促進できることが期待できる。

[0025] 雑音強調マスクは、入力音声に含まれる目的音声の時刻、周波数帯域の音量を相対的に減少させる。雑音強調マスク推定モデルは、雑音混合音声が入力されると雑音強調マスクを出力するモデルである。雑音強調マスク推定モデルが推定する雑音強調マスクを推定雑音強調マスクと呼んでもよい。推定

雑音強調マスクを推定値と呼ぶ場合がある。雑音強調マスクは、下記式 2 のように表現してもよい。

[式 2]

$$M_N(f, t) = \frac{N(f, t)^2}{S(f, t)^2 + N(f, t)^2}$$

[0026] 雑音強調マスクは、「雑音混合音声」のパワーのうちの「雑音」のパワーの割合を表す。

[2-1-4 : 学習]

[0027] 情報処理装置 2 は、音声強調マスク推定モデルと雑音強調マスク推定モデルとを学習させることにより、理想的な音声強調マスクを推定することのできる音声強調マスク推定モデルを作成してもよい。音声強調マスク推定モデル、及び雑音強調マスク推定モデルは、ニューラル・ネットワーク (Neural Network : NN) で実装してもよい。音声強調マスク推定モデル、及び雑音強調マスク推定モデルは、再帰的ニューラル・ネットワーク (Recurrent Neural Network : RNN) で実装してもよい。情報処理装置 2 は、音声強調マスク推定モデルによる推定値と雑音強調マスク推定モデルによる推定値とに応じて、音声強調マスク推定モデルが含むパラメータと雑音強調マスク推定モデルが含むパラメータとを更新し、理想的な音声強調マスクを推定することのできる音声強調マスク推定モデルを作成してもよい。

[0028] 教師入力信号、並びに教師入力信号に対応する理想的な音声強調マスク、及び雑音強調マスクを予め用意する。理想的な音声強調マスク（「目標音声強調マスク」と表すこともある）とは、雑音混合音声中の音声と雑音が既知な場合に計算される音声強調マスクである。また、理想的な雑音強調マスク（「目標雑音強調マスク」と表すこともある）とは、雑音混合音声中の音声と雑音が既知な場合に計算される雑音強調マスクである。

[0029] 以降、「理想的な音声強調マスク」および「理想的な雑音強調マスク」という言葉を用いて、情報処理装置 2 における処理を説明する。

[0030] 教師入力信号とは、理想的な音声強調マスク、及び雑音強調マスクが予め用意されている入力信号である。入力信号とは、上述の通り、対象信号である「音声」と、対象外信号である「雑音」からなっているもよい。理想的な音声強調マスクを理想音声強調マスク、又は理想値とよんでもよい。また、理想的な雑音強調マスクを理想雑音強調マスク、又は理想値とよんでもよい。また、教師入力信号、及び理想値を含む情報を、教師情報と呼んでもよい。教師情報は、後に説明する教師情報記憶部 222 に記憶されていてもよい。理想値は、パラメータの更新において、目標とする値（目標値）であってもよい。

[2-2：情報処理装置 2 の構成]

[0031] 図 2 は、第 2 実施形態における情報処理装置 2 の構成を示すブロック図である。図 2 に示すように、情報処理装置 2 は、演算装置 21 と、記憶装置 22 とを備えている。更に、情報処理装置 2 は、通信装置 23 と、入力装置 24 と、出力装置 25 とを備えていてもよい。但し、情報処理装置 2 は、通信装置 23、入力装置 24 及び出力装置 25 のうちの少なくとも一つを備えていなくてもよい。演算装置 21 と、記憶装置 22 と、通信装置 23 と、入力装置 24 と、出力装置 25 とは、データバス 26 を介して接続されていてもよい。

[0032] 演算装置 21 は、例えば、CPU (Central Processing Unit)、GPU (Graphics Processing Unit) 及び FPGA (Field Programmable Gate Array) のうちの少なくとも一つを含む。演算装置 21 は、コンピュータプログラムを読み込む。例えば、演算装置 21 は、記憶装置 22 が記憶しているコンピュータプログラムを読み込んでもよい。例えば、演算装置 21 は、コンピュータで読み取り可能であって且つ一時的でない記録媒体が記憶しているコンピュータプログラムを、情報処理装置 2 が備える図示しない記録媒体読み取り装置（例えば、後述する入力装置 24）を用いて読み込んでもよい。演算装置 21 は、通信装置 23（或いは、その他の通信装置）を介

して、情報処理装置 2 の外部に配置される不図示の装置からコンピュータプログラムを取得してもよい（つまり、ダウンロードしてもよい又は読み込んでもよい）。演算装置 21 は、読み込んだコンピュータプログラムを実行する。その結果、演算装置 21 内には、情報処理装置 2 が行うべき動作を実行するための論理的な機能ブロックが実現される。つまり、演算装置 21 は、情報処理装置 2 が行うべき動作（言い換えれば、処理）を実行するための論理的な機能ブロックを実現するためのコントローラとして機能可能である。

[0033] 図 2 には、情報処理動作を実行するために演算装置 21 内に実現される論理的な機能ブロックの一例が示されている。図 2 に示すように、演算装置 21 内には、後述する付記に記載された「拘束ロス算出手段」の一具体例である拘束ロス算出部 211 と、後述する付記に記載された「パラメータ更新手段」の一具体例であるパラメータ更新部 212 と、後述する付記に記載された「音声強調マスク推定手段」の一具体例である音声強調マスク推定部 213 と、後述する付記に記載された「雑音強調マスク推定手段」の一具体例である雑音強調マスク推定部 214 と、後述する付記に記載された「音声強調マスクロス算出手段」の一具体例である音声強調マスクロス算出部 215 と、後述する付記に記載された「雑音強調マスクロス算出手段」の一具体例である雑音強調マスクロス算出部 216 と、後述する付記に記載された「合算ロス算出手段」の一具体例である合算ロス算出部 217 と、雑音混合音声入力部 218 とが実現される。音声強調マスク推定部 213 は、後述する付記に記載された「音声強調マスク推定モデル」の一具体例である音声強調マスク推定モデルを用いて音声強調マスクを推定する。雑音強調マスク推定部 214 は、後述する付記に記載された「雑音強調マスク推定モデル」の一具体例である雑音強調マスク推定モデルを用いて雑音強調マスクを推定する。但し、演算装置 21 内には、音声強調マスク推定部 213、雑音強調マスク推定部 214、音声強調マスクロス算出部 215、雑音強調マスクロス算出部 216、合算ロス算出部 217、及び雑音混合音声入力部 218 の何れかが実現されなくてもよい。拘束ロス算出部 211、パラメータ更新部 212、

音声強調マスク推定部 213、雑音強調マスク推定部 214、音声強調マスクロス算出部 215、雑音強調マスクロス算出部 216、合算ロス算出部 217、及び雑音混合音声入力部 218 の各々の動作の詳細については、図 3 を参照しながら後に説明する。

[0034] 記憶装置 22 は、所望のデータを記憶可能である。例えば、記憶装置 22 は、演算装置 21 が実行するコンピュータプログラムを一時的に記憶していてもよい。記憶装置 22 は、演算装置 21 がコンピュータプログラムを実行している場合に演算装置 21 が一時的に使用するデータを一時的に記憶してもよい。記憶装置 22 は、情報処理装置 2 が長期的に保存するデータを記憶してもよい。尚、記憶装置 22 は、RAM (Random Access Memory)、ROM (Read Only Memory)、ハードディスク装置、光磁気ディスク装置、SSD (Solid State Drive) 及びディスクアレイ装置のうちの少なくとも一つを含んでいてもよい。つまり、記憶装置 22 は、一時的でない記録媒体を含んでいてもよい。記憶装置 22 は、パラメータ記憶部 221、及び教師情報記憶部 222 を実現してもよい。パラメータ記憶部 221 は、音声強調マスク推定モデルが含むパラメータ、及び雑音強調マスク推定モデルが含むパラメータを記憶する。但し、記憶装置 22 は、パラメータ記憶部 221、及び教師情報記憶部 222 の何れかを実現しなくてもよい。

[0035] 通信装置 23 は、不図示の通信ネットワークを介して、情報処理装置 2 の外部の装置と通信可能である。通信装置 23 は、イーサネット（登録商標）、Wi-Fi（登録商標）、Bluetooth（登録商標）、USB (Universal Serial Bus) 等の規格に基づく通信インターフェースであってもよい。

[0036] 入力装置 24 は、情報処理装置 2 の外部からの情報処理装置 2 に対する情報の入力を受け付ける装置である。例えば、入力装置 24 は、情報処理装置 2 のオペレータが操作可能な操作装置（例えば、キーボード、マウス及びタッチパネルのうちの少なくとも一つ）を含んでいてもよい。例えば、入力装

置 2 4 は情報処理装置 2 に対して外付け可能な記録媒体にデータとして記録されている情報を読み取り可能な読取装置を含んでいてもよい。

[0037] 出力装置 2 5 は、情報処理装置 2 の外部に対して情報を出力する装置である。例えば、出力装置 2 5 は、情報を画像として出力してもよい。つまり、出力装置 2 5 は、出力したい情報を示す画像を表示可能な表示装置（いわゆる、ディスプレイ）を含んでいてもよい。例えば、出力装置 2 5 は、情報を音声として出力してもよい。つまり、出力装置 2 5 は、音声を出力可能な音声装置（いわゆる、スピーカ）を含んでいてもよい。例えば、出力装置 2 5 は、紙面に情報を出力してもよい。つまり、出力装置 2 5 は、紙面に所望の情報を印刷可能な印刷装置（いわゆる、プリンタ）を含んでいてもよい。

[2 - 3 : 情報処理装置 2 が行う情報処理動作]

[0038] 図 3 を参照しながら、情報処理装置 2 が行う情報処理動作について説明する。図 3 は、情報処理装置 2 が行う情報処理動作の流れを示すフローチャートである。

[0039] 図 3 に示す様に、雑音混合音声入力部 2 1 8 は、雑音混合音声を取得し、音声強調マスク推定部 2 1 3、及び雑音強調マスク推定部 2 1 4 に入力する（ステップ S 2 0）。

[0040] 音声強調マスク推定部 2 1 3 は、音声強調マスク推定モデルを用いて音声強調マスクを推定する（ステップ S 2 1）。音声強調マスク推定部 2 1 3 は、雑音混合音声が入力された音声強調マスク推定モデルが出力した推定音声強調マスクを推定値として出力する。

[0041] 雑音強調マスク推定部 2 1 4 は、雑音強調マスク推定モデルを用いて雑音強調マスクを推定する（ステップ S 2 2）。雑音強調マスク推定部 2 1 4 は、雑音混合音声が入力された雑音強調マスク推定モデルが出力した推定雑音強調マスクを推定値として出力する。

[2 - 3 - 1 : 音声強調マスクロス]

[0042] 音声強調マスクロス算出部 2 1 5 は、音声強調マスク推定モデルが推定した推定音声強調マスクと理想的な理想音声強調マスクとを用いて、音声強調

マスクロスを算出する（ステップS 2 3）。音声強調マスクロスは、推定音声強調マスクと理想音声強調マスクとの違いを示す。音声強調マスクロスは、音声強調マスク損失関数 L_S により算出されてもよい。音声強調マスク損失関数 L_S は、下記式3のように表現してもよい。すなわち、音声強調マスク損失関数 L_S は、各時刻における周波数瓶での推定値と理想値との差の二乗和で表現してもよい。 M_S は、音声強調マスクの推定値を表現している。また、 M_S^i は、音声強調マスクの理想値を表現している。

[式3]

$$L_S = MSE(M_S; M_S^i) = \sum_{f,t} (M_S - M_S^i)^2$$

なお、以下、各時刻における周波数瓶での事項を、「時間一周波数」という文言を用いて表現する場合がある。

[2-3-2：雑音強調マスクロス]

[0043] 雑音強調マスクロス算出部216は、雑音強調マスク推定モデルが推定した推定雑音強調マスクと理想的な理想雑音強調マスクとを用いて、雑音強調マスクロスを算出する（ステップS 2 4）。雑音強調マスクロスは、推定雑音強調マスクと理想雑音強調マスクとの違いを示す。雑音強調マスクロスは、雑音強調マスク損失関数 L_N により算出されてもよい。雑音強調マスク損失関数 L_N は、下記式4のように表現してもよい。すなわち、雑音強調マスク損失関数 L_N は、時間一周波数での推定値と理想値との差の二乗和で表現してもよい。 M_N は、雑音強調マスクの推定値を表現している。また、 M_N^i は、雑音強調マスクの理想値を表現している。

[式4]

$$L_N = MSE(M_N; M_N^i) = \sum_{f,t} (M_N - M_N^i)^2$$

[2-3-3：拘束ロス]

[0044] 上記式1で表現される音声強調マスク、及び上記式2で表現される雑音強調マスクから、下記式5を成り立たせることができる。すなわち、時間一周波数における音声強調マスク、及び雑音強調マスクの和は理想的には一定値

「1」となる。本実施形態では、これを拘束条件として用いる。

[式5]

$$M_S + M_N = \frac{S(f, t)^2}{S(f, t)^2 + N(f, t)^2} + \frac{N(f, t)^2}{S(f, t)^2 + N(f, t)^2} = 1$$

[0045] 拘束ロス算出部211は、音声強調マスク推定モデルが推定した推定音声強調マスク、及び雑音強調マスク推定モデルが推定した推定雑音強調マスクを用いて拘束ロスを算出する（ステップS25）。拘束ロスは、音声強調マスク、及び雑音強調マスクの和（拘束条件）の推定値と理想値との違いを示す。

[0046] 拘束条件から外れた場合、すなわち下記式6から求まる値が「0」から離れるに従いペナルティを課すように損失関数を設計する。本実施形態における拘束ロスは、下記式6から求めることができる。

[式6]

$$\frac{S(f, t)^2}{S(f, t)^2 + N(f, t)^2} + \frac{N(f, t)^2}{S(f, t)^2 + N(f, t)^2} - 1$$

[0047] 本実施形態では、音声と雑音との双方に関する情報として拘束条件を導入する。拘束条件は、後述するパラメータの更新の際に、音声と雑音との双方に関する情報として用いられてもよい。例えば、上記式6において、求まる値が「0」より大きい場合、すなわち、 $M_S + M_N$ が「1」よりも大きい場合は、音声強調マスクのパラメータは、音声強調がより強くなるように更新され、雑音強調マスクのパラメータも、雑音強調がより強くなるように更新されてもよい。逆に上記式6において求まる値が「0」より小さい場合は、すなわち、 $M_S + M_N$ が「1」よりも小さい場合は、音声強調マスクのパラメータは、音声強調がより弱くなるように更新され、雑音強調マスクのパラメータも、雑音強調がより弱くなるように更新されてもよい。

[0048] 推定された音声強調マスク及び雑音強調マスクが拘束条件から逸脱した量に基づいて、損失関数を導くことができる。本実施形態では、拘束条件として下記式7で示す拘束損失関数 L_{SN} を導入してもよい。下記式7は、時間一周波数の和を算出する。

[式 7]

$$L_{SN} = MSE(M_S + M_N; 1) = \sum_{f,t} (M_S + M_N - 1)^2$$

[0049] 拘束損失関数 L_{SN} は、「0」超過の場合に損失が存在する。この拘束条件の導入により、音声強調マスク及び雑音強調マスクの一方の推定値が正しくても、他方の推定値が正しくない場合は、損失関数の値は上昇する。音声強調マスク及び雑音強調マスクの各々を独立に学習させる比較例と比較して、正確な推定をすることができるようにモデルを学習させることができる。

[2-3-4：合算ロス]

[0050] 合算ロス算出部 217 は、音声強調マスクロス、雑音強調マスクロス、及び拘束ロスを合算した合算ロスを算出する（ステップ S26）。合算ロスは、合算損失関数 L_{ALL} により算出されてもよい。合算損失関数 L_{ALL} は、下記式 8 のように表現してもよい。合算損失関数 L_{ALL} から求まる合算ロスを、全ロスと呼んでもよい。

[式 8]

$$L_{ALL} = L_S + L_N + \lambda L_{SN}$$

[0051] λ は、「0」から「1」の間の値であってもよい。 λ は、1 未満の正数であってもよい。 λ は、「0.01」、「0.1」等の「1」と比較して小さい値であってもよい。 λ は、モデルの学習過程において、一定でもよいし、変化させてもよい。 λ は、モデルの学習過程において、開始の際には特に小さく、終了の際に大きくなってもよい。

[0052] なお、上記式 8 では、合算において、音声強調マスク損失関数 L_S と雑音強調マスク損失関数 L_N とに同じ重みを付しているが、音声強調マスク損失関数 L_S と雑音強調マスク損失関数 L_N とには異なる重みを付してもよい。例えば下記式 9 に示す合算損失関数 L_{ALL} により全ロスを算出してもよい。

[式 9]

$$L_{ALL} = aL_S + bL_N + \lambda L_{SN}$$

[0053] 上記「a」及び「b」は異なる値であってもよい。また、上記「a」及び

「b」は同じ値であってもよい。上記「a」及び「b」が共に「1」の場合、上記式9は上記式8と等しくなる。

[2-3-5: パラメータ更新]

[0054] パラメータ更新部212は、合算ロス算出部217の算出結果に応じて、音声強調マスク推定モデルが含むパラメータと雑音強調マスク推定モデルが含むパラメータとを更新する（ステップS27）。パラメータ更新部212は、推定音声強調マスクと理想的な理想音声強調マスクとの違いを示す音声強調マスクロス、推定雑音強調マスクと理想的な理想雑音強調マスクとの違いを示す雑音強調マスクロス、及び拘束ロスを用いて、音声強調マスク推定モデルが含むパラメータと雑音強調マスク推定モデルが含むパラメータとを更新する。パラメータ更新部212は、パラメータ記憶部221に記憶されている音声強調マスク推定モデルが含むパラメータと雑音強調マスク推定モデルが含むパラメータとを更新する。パラメータ更新部212は、例えば、逆誤差伝搬法を用いて、音声強調マスク推定モデルのパラメータと雑音強調マスク推定モデルのパラメータとを更新してもよい。

[0055] 本実施形態における情報処理装置2は、拘束損失関数 L_{SN} を通じて音声強調マスク推定モデル、及び雑音強調マスク推定モデルの双方の学習をすすめることができる。音声強調マスク推定モデルは、音声強調マスク推定モデルが含む情報、及び雑音強調マスク推定モデルが含む情報を結合させた情報を用いて学習されることができる。

[2-4: 拘束条件の許容範囲]

[0056] 上述した通り、音声強調マスクと雑音強調マスクとの和は理想的には「1」であり、本実施形態では、これを拘束条件として用いている。一方、この拘束条件は、厳密に「1」でなくてもよい場合がある。この場合、例えば、下記式10に示す様に、「1」を「 $1 + \varepsilon$ 」に置き換えてもよい。

[式10]

$$M_s + M_N = \frac{S(f, t)^2}{S^{i2}(f, t) + N^i(f, t)^2} + \frac{N(f, t)^2}{S^{i2}(f, t) + N^i(f, t)^2} = 1 + \varepsilon$$

「 ε 」は「1」と比較して非常に小さな正数である。「1」を「 $1 + \varepsilon$ 」に置き換えた場合、「 ε 」程度の誤差を許容することができる。「 ε 」は、設計により任意に変更することができてよい。上記式10の分母の S^i 、及び N^i は、理想値を表現しており、分子の S 、及び N は、推定値を表現している。

[0057] 「1」を「 $1 + \varepsilon$ 」に置き換得た場合、拘束損失関数 L_{SN} は、下記式11のように表現してもよい。

[式11]

$$L_{SN} = MSE(M_S + M_N; 1 + \varepsilon) = \sum (M_S + M_N - (1 + \varepsilon))^2$$

[0058] 拘束損失関数 L_{SN} から求まる拘束ロスからは、音声強調マスク、及び雑音強調マスクの何れが理想値と比較して大きいのか又は小さいのかが不明である。音声強調マスクと雑音強調マスクとの和が「1」を超える場合、すなわち、正数の「 ε 」を採用した場合とは、推定音声強調マスク、及び推定雑音強調マスクの少なくとも一方が理想値と比較して大きい場合である。すなわち、正数の「 ε 」を採用すると、音声から完全に雑音を抑制できない可能性がある。よって、多少の雑音を許容してもよい場合においては、正数の「 ε 」を許容してもよい。

[0059] 一方で、音声強調マスクと雑音強調マスクとの和が「1」未満の場合、すなわち、負数の「 ε 」を採用した場合とは、推定音声強調マスク、及び推定雑音強調マスクの少なくとも一方が理想値と比較して小さい場合である。すなわち、負数の「 ε 」を採用すると、残すべき音である音声を削りすぎてしまう場合がある。音声を削りすぎると音声聞き取りにくくなるので、負数の「 ε 」は許容しない。パラメータ更新部212は、少なくとも、音声強調マスクと雑音強調マスクとの和が「1」未満とはならないように、パラメータを更新する。

[2-5：変形例]

[0060] 音声は、複数の人物の声を含む場合、音声強調マスク推定部213は、複数の人物各々に対応する音声強調マスクを推定してもよい。各々の人物に対

応する音声強調マスク推定モデルが用意されており、各々の音声強調マスク推定モデルは、人物に対応する音声強調マスクを出力してもよい。

[0061] 例えば、音声は人物Aの声と人物Bの声を含む場合、人物Aの声に対応する音声強調マスク推定モデルA、及び人物Bの声に対応する音声強調マスク推定モデルBが学習されてもよい。音声強調マスク推定モデルAは、雑音混合音声が入力されると人物Aの声に対応する音声強調マスク M_{SA} を出力し、音声強調マスク推定モデルBは、雑音混合音声が入力されると人物Bの声に対応する音声強調マスク M_{SB} を出力してもよい。

[0062] 音声強調マスク M_{SA} 、及び音声強調マスク M_{SB} を出力する場合の拘束条件は、下記式12のように表現してもよい。

[式12]

$$\begin{aligned} & M_{SA} + M_{SB} + M_N \\ &= \frac{SA(f, t)^2}{SA(f, t)^2 + SB(f, t)^2 + N(f, t)^2} + \frac{SB(f, t)^2}{SA(f, t)^2 + SB(f, t)^2 + N(f, t)^2} \\ &+ \frac{N(f, t)^2}{SA(f, t)^2 + SB(f, t)^2 + N(f, t)^2} = 1 \end{aligned}$$

[0063] なお、本実施形態では、平均二乗誤差を用いた損失関数を例を挙げたが、平均二乗誤差以外の手法を用いた損失関数を採用してもよい。

[2-6：情報処理装置2の技術的効果]

[0064] 例えば上記非特許文献1には、モデルにより音声強調マスクと雑音強調マスクとを推定し、推定値と理想値との違いを損失として算出し、モデルを学習させる技術が記載されている。上記非特許文献1に記載されている技術を比較例とよぶ。比較例においては、モデルにより推定された音声強調マスクの評価に、モデルにより推定された雑音強調マスクの情報をを用いることがない。また、モデルにより推定された雑音強調マスクの評価に、モデルにより推定された音声強調マスクの情報をを用いることもない。すなわち、推定された音声強調マスクと推定された雑音強調マスクとは各々独立して評価されている。

[0065] 上述の通り、入力信号に含まれるのは、音声と雑音（上述の通り、本実施

形態では音声以外を「雑音」とよんでいる）である。したがって、音声強調マスクの理想値と雑音強調マスクの理想値との和は一定値「1」となる。上述の通り、音声強調マスクと雑音強調マスクとの和が「1」を超えた場合、音声強調マスクの推定値、及び雑音強調マスクの推定値の少なくとも一方が理想値と比較して大きく、音声から完全に雑音を抑制できない可能性がある。また、音声強調マスクと雑音強調マスクとの和が「1」未満の場合、音声強調マスクの推定値、及び雑音強調マスクの推定値の少なくとも一方が理想値と比較して小さく、残すべき音である音声を削りすぎてしまう場合がある。例えば、短時間に急変する雑音が入力信号に含まれている場合の推定において、音声を削りすぎてしまう場合が多い。音声を削りすぎると音声聞き取りにくくなることを避けるべく、特に音声を削りすぎは抑制することが望ましい。

[0066] 比較例では、推定された音声強調マスクと推定された雑音強調マスクとは各々独立して評価しており、音声強調マスクと雑音強調マスクとの和が一定値「1」であるか否かは不明である。したがって、音声から完全に雑音を抑制できないのみならず、残すべき音である音声を削りすぎてしまう場合がある。また、雑音の種類は多様であり、雑音強調マスクの精度は、音声強調マスクと比較して低精度となる場合が多いので、雑音強調マスクの独立した評価をモデルのパラメータの更新に用いると、音声強調の精度が悪くなる可能性がある。

[0067] 第2実施形態における情報処理装置2は、音声強調マスク推定部213による推定結果と、雑音強調マスク推定部214による推定結果とを用いて算出される拘束ロスを導入する。そして、音声強調マスクロス、雑音強調マスクロス、及び拘束ロスを合算した合算ロスを用いて、音声強調マスク推定モデルが含むパラメータを更新するので、音声から雑音を抑制し、残すべき音である音声を削りすぎることが防ぐことができる。情報処理装置2は、精度よく音声強調をすることができる音声強調マスク推定モデルを作成することができる。

[3 : 第 3 実施形態]

[0068] 続いて、情報処理装置、情報処理方法、及び記録媒体の第 3 実施形態について説明する。以下では、情報処理装置、情報処理方法、及び記録媒体の第 3 実施形態が適用された情報処理装置 3 を用いて、情報処理装置、情報処理方法、及び記録媒体の第 3 実施形態について説明する。

[0069] 図 4 は、第 3 実施形態における情報処理装置 3 の構成を示すブロック図である。第 3 実施形態における情報処理装置 3 は、拘束ロス算出部 3 1 1 の動作が第 2 実施形態における情報処理装置 2 と異なる。

[3 - 1 : 情報処理装置 3 が行う情報処理動作]

[0070] 第 3 実施形態において拘束ロス算出部 3 1 1 は、推定音声強調マスク、及び推定雑音強調マスクに雑音混合音声の大きさを乗じて、拘束ロスを算出する。拘束ロス算出部 3 1 1 は、時間一周波数での推定音声強調マスク、及び推定雑音強調マスクに雑音混合音声の大きさを乗じて、拘束ロスを算出してもよい。雑音混合音声の大きさは、雑音混合音声の振幅の絶対値であってもよい。雑音混合音声の大きさは、雑音混合音声の対数パワースペクトラム ($\text{Log power spectrum}$) であってもよい。雑音混合音声の大きさは、正規化した値であってもよい。当該値は、設計により任意に変更することができてもよい。

[0071] 雑音混合音声の大きさを LPS_{input} と表す場合、第 3 実施形態における拘束損失関数 L_{SN} は、下記式 1 3 のように表現してもよい。

[式 1 3]

$$L_{SN} = LPS_{input} \times MSE(M_S + M_N; 1) = \sum_{f,t} LPS_{input} \times (M_S + M_N - 1)^2$$

[0072] LPS_{input} を乗じると、 LPS_{input} が大きい時間一周波数における誤差をより重視することができる。これは、振幅スペクトラム近似 ($\text{magnitude spectrum approximation: MSA}$) の適用に相当する。

[0073] LPS_{input} を乗じると、例えば新幹線の通過音のような短時間に急変する

雑音の学習時に有効である。小さな入力信号よりも大きな入力信号の方に重みを与えることになり、大きな入力信号に関する誤差をより強調することができる。

[3 - 2 : 情報処理装置 3 の技術的効果]

[0074] 第 3 実施形態における情報処理装置 3 は、雑音混合音声の大きさを乗じるので、特に大きな入力信号に関する雑音を抑制することができる。

[4 : 第 4 実施形態]

[0075] 続いて、情報処理装置、情報処理方法、及び記録媒体の第 4 実施形態について説明する。以下では、情報処理装置、情報処理方法、及び記録媒体の第 4 実施形態が適用された情報処理装置 4 を用いて、情報処理装置、情報処理方法、及び記録媒体の第 4 実施形態について説明する。

[0076] 図 5 は、第 4 実施形態における情報処理装置 4 の構成を示すブロック図である。第 4 実施形態における情報処理装置 4 は、拘束ロス算出部 4 1 1 の動作が第 2 実施形態における情報処理装置 2、及び第 3 実施形態における情報処理装置 3 と異なる。

[4 - 1 : 情報処理装置 4 が行う情報処理動作]

[0077] 第 4 実施形態において拘束ロス算出部 4 1 1 は、推定音声強調マスク、及び推定雑音強調マスクを所定の指数で冪乗して、拘束ロスを算出する。拘束ロス算出部 4 1 1 は、時間一周波数での推定音声強調マスク、及び推定雑音強調マスクを所定の指数で冪乗して、拘束ロスを算出してもよい。所定の指数は、0 以上 1 以下の値を取り得てもよい。

[0078] 所定の指数を α と表す場合、第 4 実施形態における拘束損失関数 L_{SN} は、下記式 1 4 のように表現してもよい。

[式 1 4]

$$L_{SN} = MSE(M_S^\alpha + M_N^\alpha; 1)$$

[0079] α を冪乗すると、例えば、エアコンの音のような定常的な雑音をより重点的に学習することができる。これは、べき乗圧縮 (Power-law compression) の適用に相当する効果を期待することができる。

[0080] なお、第4実施形態においても、第3実施形態において適用した LPS_{input} を適用してもよい。この場合、第4実施形態における拘束損失関数 L_{SN} は、下記式15のように表現してもよい。

[式15]

$$L_{SN} = LPS_{input} \times MSE(M_S^\alpha + M_N^\alpha; 1)$$

[4-2: 情報処理装置4の技術的效果]

[0081] 第4実施形態における情報処理装置4は、 α を適用することにより、小さい信号をより重点的に学習することができる。

[5: 第5実施形態]

[0082] 続いて、情報処理装置、情報処理方法、及び記録媒体の第5実施形態について説明する。以下では、情報処理装置、情報処理方法、及び記録媒体の第5実施形態が適用された情報処理装置5を用いて、情報処理装置、情報処理方法、及び記録媒体の第5実施形態について説明する。

[5-1: 情報処理装置5の構成]

[0083] 図6に示すように、第5実施形態における情報処理装置5は、第2実施形態における情報処理装置2から第4実施形態における情報処理装置4と同様に、演算装置21と、記憶装置22とを備えている。更に、第5実施形態における情報処理装置5は、第2実施形態における情報処理装置2から第4実施形態における情報処理装置4と同様に、通信装置23と、入力装置24と、出力装置25とを備えていてもよい。但し、情報処理装置5は、通信装置23、入力装置24及び出力装置25のうちの少なくとも1つを備えていなくてもよい。第5実施形態における情報処理装置5は、演算装置21内に音響特徴抽出部519が更に実現される点で、第2実施形態における情報処理装置2から第4実施形態における情報処理装置4と異なる。情報処理装置5のその他の特徴は、第2実施形態における情報処理装置2から第4実施形態における情報処理装置4の少なくとも1つのその他の特徴と同一であってもよい。このため、以下では、すでに説明した各実施形態と異なる部分について詳細に説明し、その他の重複する部分については適宜説明を省略するもの

とする。

[5 - 2 : 情報処理装置 5 が行う情報処理動作]

[0084] 図 7 を参照しながら、情報処理装置 5 が行う情報処理動作について説明する。図 7 は、情報処理装置 5 が行う情報処理動作の流れを示すフローチャートである。

[0085] 図 7 に示す様に、雑音混合音声入力部 5 1 8 は、雑音混合音声を取得し、音響特徴抽出部 5 1 9 に入力する（ステップ S 5 0）。音響特徴抽出部 5 1 9 は、雑音混合音声から雑音混合音声の特徴である雑音混合音声特徴を抽出する（ステップ S 5 1）。音響特徴抽出部 5 1 9 は、雑音混合音声が入力され、雑音混合音声特徴を出力する音響特徴抽出モデルを有していてもよい。音響特徴抽出モデルは、RNN で実装してもよい。

[0086] 音声強調マスク推定部 5 1 3 は、音声強調マスク推定モデルを用いて音声強調マスクを推定する（ステップ S 5 2）。第 5 実施形態において、音声強調マスク推定モデルは、雑音混合音声特徴が入力され、推定音声強調マスクを出力してもよい。音声強調マスク推定部 5 1 3 は、雑音混合音声特徴が入力された音声強調マスク推定モデルが出力した推定音声強調マスクを推定値として出力してもよい。

[0087] 雑音強調マスク推定部 5 1 4 は、雑音強調マスク推定モデルを用いて雑音強調マスクを推定する（ステップ S 5 3）。第 5 実施形態において、雑音強調マスク推定モデルは、雑音混合音声特徴が入力され、推定雑音強調マスクを出力してもよい。雑音強調マスク推定部 5 1 4 は、雑音混合音声特徴が入力された雑音強調マスク推定モデルが出力した推定雑音強調マスクを推定値として出力してもよい。

[0088] 音声強調マスククロス算出部 5 1 5 は、音声強調マスク推定モデルが推定した推定音声強調マスクと理想音声強調マスクとを用いて、音声強調マスククロスを算出する（ステップ S 5 4）。理想音声強調マスクは、教師入力信号から抽出された雑音混合音声特徴に対して理想的な音声強調マスクであってもよい。

[0089] 雑音強調マスクロス算出部516は、雑音強調マスク推定モデルが推定した推定雑音強調マスクと理想雑音強調マスクとを用いて、雑音強調マスクロスを算出する（ステップS55）。理想雑音強調マスクは、教師入力信号から抽出された雑音混合音声特徴に対し、理想的な雑音強調マスクであってもよい。

[0090] 拘束ロス算出部211は、音声強調マスク推定モデルが推定した推定音声強調マスク、及び雑音強調マスク推定モデルが推定した推定雑音強調マスクを用いて拘束ロスを算出する（ステップS25）。合算ロス算出部217は、音声強調マスクロス、雑音強調マスクロス、及び拘束ロスを合算した合算ロスを算出する（ステップS26）。

[0091] パラメータ更新部512は、合算ロス算出部217の算出結果に応じて、音声強調マスク推定モデルが含むパラメータ、及び雑音強調マスク推定モデルが含むパラメータと共に、音響特徴抽出モデルが含むパラメータを更新する（ステップS56）。音響特徴抽出モデルは、音声に関する情報と雑音に関する情報との両方を用いて学習されてもよい。

[0092] 第5実施形態における情報処理装置5は、音声強調マスク推定モデル、及び音響特徴抽出モデルを機械学習により作成した。第5実施形態において、音声を強調する場面では、作成した音声強調マスク推定モデルに加えて、音響特徴抽出モデルが適用される。

[5-3：情報処理装置5の技術的効果]

[0093] 第5実施形態における情報処理装置5は、音響特徴抽出部519を備えることにより、音声強調マスク推定モデル、及び雑音強調マスク推定モデルの双方への入力信号を大きくすることができるので、より好ましく音声強調をすることができる。一方で、第2実施形態のように、音響特徴抽出部519を備えない場合は、動作を軽くすることができる。

[0094] 例えば、上述した比較例の場合、音響特徴抽出部519を有することにより、逆伝搬の際に音声、及び雑音の一方の信号を削りすぎる方向へモデルのパラメータが調整される可能性がある。これに対し、本実施形態では、音声

強調マスク推定モデルによる推定結果と、雑音強調マスク推定モデルによる推定結果との情報を用いて算出される拘束ロスを導入するので、音響特徴抽出部519を有していても、逆伝搬の際に音声、及び雑音の一方の信号を削りすぎる方向へモデルのパラメータが調整されることがない。

[6：第6実施形態]

[0095] 続いて、情報処理装置、情報処理方法、及び記録媒体の第6実施形態について説明する。以下では、情報処理装置、情報処理方法、及び記録媒体の第6実施形態が適用された情報処理装置6を用いて、情報処理装置、情報処理方法、及び記録媒体の第6実施形態について説明する。

[6-1：情報処理装置6の構成]

[0096] 図8に示すように、第6実施形態における情報処理装置6は、第2実施形態における情報処理装置2から第5実施形態における情報処理装置5と同様に、演算装置21と、記憶装置22とを備えている。更に、第6実施形態における情報処理装置6は、第2実施形態における情報処理装置2から第5実施形態における情報処理装置5と同様に、通信装置23と、入力装置24と、出力装置25とを備えていてもよい。但し、情報処理装置6は、通信装置23、入力装置24及び出力装置25のうちの少なくとも1つを備えていなくてもよい。第6実施形態における情報処理装置6は、拘束ロス算出部611が雑音拘束ロス算出部6111、及び音声拘束ロス算出部6112を有する点で、第2実施形態における情報処理装置2から第5実施形態における情報処理装置5と異なる。情報処理装置6のその他の特徴は、第2実施形態における情報処理装置2から第5実施形態における情報処理装置5の少なくとも1つのその他の特徴と同一であってもよい。このため、以下では、すでに説明した各実施形態と異なる部分について詳細に説明し、その他の重複する部分については適宜説明を省略するものとする。

[6-2：情報処理装置6が行う情報処理動作]

[0097] 図9を参照しながら、情報処理装置6が行う情報処理動作について説明する。図9は、情報処理装置6が行う情報処理動作の流れを示すフローチャー

トである。

[0098] 図9に示す様に、雑音混合音声入力部218は、雑音混合音声を取得し、音声強調マスク推定部213、及び雑音強調マスク推定部214に入力する（ステップS20）。音声強調マスク推定部213は、音声強調マスク推定モデルを用いて音声強調マスクを推定する（ステップS21）。雑音強調マスク推定部214は、雑音強調マスク推定モデルを用いて雑音強調マスクを推定する（ステップS22）。音声強調マスクロス算出部215は、音声強調マスク推定モデルが推定した推定音声強調マスクと理想音声強調マスクとを用いて、音声強調マスクロスを算出する（ステップS23）。雑音強調マスクロス算出部216は、雑音強調マスク推定モデルが推定した推定雑音強調マスクと理想雑音強調マスクとを用いて、雑音強調マスクロスを算出する（ステップS24）。

[6-2-1：雑音拘束ロス]

[0099] 雑音拘束ロス算出部6111は、推定音声強調マスク、及び理想雑音強調マスクを用いて雑音拘束ロスを算出する（ステップS60）。雑音拘束ロス算出部6111は、推定音声強調マスク、及び上記式5で表現される拘束条件に基づいて、拘束雑音強調マスクを算出する。拘束雑音強調マスクを M_N^* と表す場合、拘束雑音強調マスクは、下記式16のように表現してもよい。

[式16]

$$M_N^* = 1 - M_S$$

[0100] 雑音拘束ロス算出部6111は、拘束雑音強調マスクと理想雑音強調マスクとの違いを示す雑音拘束ロスを算出する。雑音拘束ロスが雑音拘束損失関数 L_N^* により算出される場合、雑音拘束損失関数 L_N^* は、下記式17のように表現してもよい。

[式17]

$$L_N^* = M_N^* - M_N^i = (1 - M_S) - M_N^i$$

[6-2-2：音声拘束ロス]

[0101] 音声拘束ロス算出部6112は、推定雑音強調マスク、及び理想雑音強調

マスクを用いて音声拘束ロスを算出する（ステップS 6 1）。音声拘束ロス算出部 6 1 1 2 は、推定雑音強調マスク、及び上記式 5 で表現される拘束条件に基づいて、拘束音声強調マスクを算出する。拘束音声強調マスクを M_S^* と表す場合、拘束音声強調マスクは、下記式 1 8 のように表現してもよい。

[式 1 8]

$$M_S^* = 1 - M_N$$

[0102] 音声拘束ロス算出部 6 1 1 2 は、拘束音声強調マスクと理想音声強調マスクとの違いを示す音声拘束ロスを算出する。音声拘束ロスが音声拘束損失関数 L_S^* により算出される場合、音声拘束損失関数 L_S^* は、下記式 1 9 のように表現してもよい。

[式 1 9]

$$L_S^* = M_S^* - M_S^i = (1 - M_N) - M_S^i$$

[6 - 2 - 3 : 合算ロス]

[0103] 合算ロス算出部 6 1 7 は、音声強調マスクロス、雑音強調マスクロス、雑音拘束ロス及び音声拘束ロスを合算した合算ロスを算出する（ステップS 6 2）。合算ロスが合算損失関数 L_{ALL} により算出される場合、合算損失関数 L_{ALL} は、下記式 2 0 のように表現してもよい。

[式 2 0]

$$\begin{aligned} L_{All} &= (M_N - M_N^i) + (M_S - M_S^i) + ((1 - M_S) - M_N^i) + ((1 - M_N) - M_S^i) \\ &= (M_N + (1 - M_S)) + (M_S + (1 - M_N)) - 2M_N^i - 2M_S^i = 2M_S^{\sim} + 2M_N^{\sim} - 2M_S^i - 2M_N^i \end{aligned}$$

[0104] $M^{\sim}$ は、拘束条件も含んだマスクの推定値と考えることができる。上記式 2 0 の第 1 項と第 2 項の和は $M_S^{\sim} + M_N^{\sim} = 1$ を自動的に満たすので、第 2 実施形態で用いた式 5 で表現される拘束条件と同等の効果を得ることができる。すなわち、第 6 実施形態における情報処理装置 6 は第 2 実施形態における情報処理装置 2 と同等の効果を得ることができる。

[0105] 上記式 2 0 は、下記式 2 1 のように書き直してもよい。第 6 実施形態における合算ロスは、拘束条件も含んだマスクの推定値と理想値との違いを示し

ていると考えることができる。

[式 2 1]

$$L_{All} = 2(M_S^{\sim} - M_S^i) + 2(M_N^{\sim} - M_N^i)$$

[0106] パラメータ更新部 2 1 2 は、合算ロス算出部 2 1 7 の算出結果に応じて、音声強調マスク推定モデルが含むパラメータと雑音強調マスク推定モデルが含むパラメータとを更新する（ステップ S 2 7）。

[6 - 3 : 情報処理装置 6 の技術的效果]

[0107] 第 6 実施形態における情報処理装置 6 は、第 2 実施形態において採用している「λ」を採用しない。したがって、第 6 実施形態における情報処理装置 6 は、「λ」の決定にかかる処理負荷が小さい点において、第 2 実施形態における情報処理装置 2 よりも処理負荷を小さくすることができる。

[7 : 付記]

[0108] 以上説明した実施形態に関して、更に以下の付記を開示する。

[付記 1]

雑音混合音声が入力された音声強調マスク推定モデルが出力した推定音声強調マスク、及び前記雑音混合音声が入力された雑音強調マスク推定モデルが出力した推定雑音強調マスクを用いて拘束ロスを算出する拘束ロス算出手段と、

前記推定音声強調マスクと目標音声強調マスクとの違いを示す音声強調マスキロス、前記推定雑音強調マスクと目標雑音強調マスクとの違いを示す雑音強調マスキロス、及び前記拘束ロスを用いて、前記音声強調マスク推定モデルが含むパラメータ、及び前記雑音強調マスク推定モデルが含むパラメータを更新するパラメータ更新手段と

を備える情報処理装置。

[付記 2]

前記音声強調マスク推定モデルを用いて前記推定音声強調マスクを推定する音声強調マスク推定手段と、

前記雑音強調マスク推定モデルを用いて前記推定雑音強調マスクを推定する雑音強調マスク推定手段と、

前記推定音声強調マスクと前記目標音声強調マスクとを用いて、前記音声強調マスクロスを算出する音声強調マスクロス算出手段と、

前記推定雑音強調マスクと前記目標雑音強調マスクとを用いて、前記雑音強調マスクロスを算出する雑音強調マスクロス算出手段と、

前記音声強調マスクロス、前記雑音強調マスクロス、及び前記拘束ロスを合算した合算ロスを算出する合算ロス算出手段と

を更に備え、

前記パラメータ更新手段は、前記合算ロス算出手段の算出結果に応じて、前記音声強調マスク推定モデルが含むパラメータと前記雑音強調マスク推定モデルが含むパラメータとを更新する

請求項 1 に記載の情報処理装置。

[付記 3]

前記拘束ロス算出手段は、前記推定音声強調マスク、及び前記推定雑音強調マスクに前記雑音混合音声の大きさを乗じて、前記拘束ロスを算出する

請求項 1 又は 2 に記載の情報処理装置。

[付記 4]

前記拘束ロス算出手段は、前記推定音声強調マスク、及び前記推定雑音強調マスクを所定の指数で冪乗して、前記拘束ロスを算出する

請求項 1 又は 2 に記載の情報処理装置。

[付記 5]

前記雑音混合音声から雑音混合音声特徴を抽出する音響特徴抽出手段を更に備え、

前記音声強調マスク推定モデルは、前記雑音混合音声特徴が入力され、前記推定音声強調マスクを出力し、

前記雑音強調マスク推定モデルは、前記雑音混合音声特徴が入力され、前記推定雑音強調マスクを出力する

請求項 1 又は 2 に記載の情報処理装置。

[付記 6]

前記拘束ロス算出手段は、

前記推定雑音強調マスク、及び前記目標音声強調マスクを用いて音声拘束ロスを算出する音声拘束ロス算出手段と、

前記推定音声強調マスク、及び前記目標雑音強調マスクを用いて雑音拘束ロスを算出する雑音拘束ロス算出手段とを有し、

前記拘束ロスは、前記音声拘束ロス、及び前記雑音拘束ロスを含む

請求項 1 又は 2 に記載の情報処理装置。

[付記 7]

雑音混合音声が入力された音声強調マスク推定モデルが出力した推定音声強調マスク、及び前記雑音混合音声が入力された雑音強調マスク推定モデルが出力した推定雑音強調マスクを用いて拘束ロスを算出し、

前記推定音声強調マスクと目標音声強調マスクとの違いを示す音声強調マスクロス、前記推定雑音強調マスクと目標雑音強調マスクとの違いを示す雑音強調マスクロス、及び前記拘束ロスを用いて、前記音声強調マスク推定モデルが含むパラメータ、及び前記雑音強調マスク推定モデルが含むパラメータを更新する

情報処理方法。

[付記 8]

コンピュータに、

雑音混合音声が入力された音声強調マスク推定モデルが出力した推定音声強調マスク、及び前記雑音混合音声が入力された雑音強調マスク推定モデルが出力した推定雑音強調マスクを用いて拘束ロスを算出し、

前記推定音声強調マスクと目標音声強調マスクとの違いを示す音声強調マスクロス、前記推定雑音強調マスクと目標雑音強調マスクとの違いを示す雑音強調マスクロス、及び前記拘束ロスを用いて、前記音声強調マスク推定モデルが含むパラメータ、及び前記雑音強調マスク推定モデルが含むパラメータ

タを更新する

情報処理方法を実行させるためのコンピュータプログラムが記録された記録媒体。

[0109] 上述の各実施形態の構成要件の少なくとも一部は、上述の各実施形態の構成要件の少なくとも他の一部と適宜組み合わせることができる。上述の各実施形態の構成要件のうちの一部が用いられなくてもよい。

[0110] この開示は上記実施形態に限定されるものではない。この開示は、請求の範囲及び明細書全体から読み取るこのできる技術的思想に反しない範囲で適宜変更可能である。そのような変更を伴う情報処理装置、情報処理方法、及び、記録媒体もまた、この開示の技術的思想に含まれる。また、法令で許容される限りにおいて、本願明細書に記載された全ての公開公報及び論文をここに取り込む。

[0111] 法令で許容される限りにおいて、この出願は、2023年3月13日に出願された日本出願特願2023-039058を基礎とする優先権を主張し、その開示の全てをここに取り込む。

符号の説明

[0112] 1, 2, 3, 4, 5, 6 情報処理装置
11, 211, 311, 411, 611 拘束ロス算出部
12, 212, 512 パラメータ更新部
221 パラメータ記憶部
222, 522 教師情報記憶部
213, 513 音声強調マスク推定部
214, 514 雑音強調マスク推定部
215, 515 音声強調マスキロス算出部
216, 516 雑音強調マスキロス算出部
217, 617 合算ロス算出部
218, 518 雑音混合音声入力部
519 音響特徴抽出部

6 1 1 1 雑音拘束ロス算出部

6 1 1 2 音声拘束ロス算出部

請求の範囲

[請求項1] 雑音混合音声が入力された音声強調マスク推定モデルが出力した推定音声強調マスク、及び前記雑音混合音声が入力された雑音強調マスク推定モデルが出力した推定雑音強調マスクを用いて拘束ロスを算出する拘束ロス算出手段と、

前記推定音声強調マスクと雑音混合音声の中の音声と雑音が既知な場合に計算される目標音声強調マスクとの違いを示す音声強調マスクロス、前記推定雑音強調マスクと目標雑音強調マスクとの違いを示す雑音強調マスクロス、及び前記拘束ロスを用いて、前記音声強調マスク推定モデルが含むパラメータ、及び前記雑音強調マスク推定モデルが含むパラメータを更新するパラメータ更新手段と

を備える情報処理装置。

[請求項2] 前記音声強調マスク推定モデルを用いて前記推定音声強調マスクを推定する音声強調マスク推定手段と、

前記雑音強調マスク推定モデルを用いて前記推定雑音強調マスクを推定する雑音強調マスク推定手段と、

前記推定音声強調マスクと前記目標音声強調マスクとを用いて、前記音声強調マスクロスを算出する音声強調マスクロス算出手段と、

前記推定雑音強調マスクと前記目標雑音強調マスクとを用いて、前記雑音強調マスクロスを算出する雑音強調マスクロス算出手段と、

前記音声強調マスクロス、前記雑音強調マスクロス、及び前記拘束ロスを合算した合算ロスを算出する合算ロス算出手段と

を更に備え、

前記パラメータ更新手段は、前記合算ロス算出手段の算出結果に応じて、前記音声強調マスク推定モデルが含むパラメータと前記雑音強調マスク推定モデルが含むパラメータとを更新する

請求項1に記載の情報処理装置。

[請求項3] 前記拘束ロス算出手段は、前記推定音声強調マスク、及び前記推定

雑音強調マスクに前記雑音混合音声の大きさを乗じて、前記拘束ロスを算出する

請求項 1 又は 2 に記載の情報処理装置。

[請求項4] 前記拘束ロス算出手段は、前記推定音声強調マスク、及び前記推定雑音強調マスクを所定の指数で乗算して、前記拘束ロスを算出する

請求項 1 又は 2 に記載の情報処理装置。

[請求項5] 前記雑音混合音声から雑音混合音声特徴を抽出する音響特徴抽出手段を更に備え、

前記音声強調マスク推定モデルは、前記雑音混合音声特徴が入力され、前記推定音声強調マスクを出力し、

前記雑音強調マスク推定モデルは、前記雑音混合音声特徴が入力され、前記推定雑音強調マスクを出力する

請求項 1 又は 2 に記載の情報処理装置。

[請求項6] 前記拘束ロス算出手段は、

前記推定雑音強調マスク、及び前記目標音声強調マスクを用いて音声拘束ロスを算出する音声拘束ロス算出手段と、

前記推定音声強調マスク、及び前記目標雑音強調マスクを用いて雑音拘束ロスを算出する雑音拘束ロス算出手段とを有し、

前記拘束ロスは、前記音声拘束ロス、及び前記雑音拘束ロスを含む請求項 1 又は 2 に記載の情報処理装置。

[請求項7] 雑音混合音声が入力された音声強調マスク推定モデルが出力した推定音声強調マスク、及び前記雑音混合音声が入力された雑音強調マスク推定モデルが出力した推定雑音強調マスクを用いて拘束ロスを算出し、

前記推定音声強調マスクと雑音混合音声中の音声と雑音が既知な場合に計算される目標音声強調マスクとの違いを示す音声強調マスクロス、前記推定雑音強調マスクと目標雑音強調マスクとの違いを示す雑音強調マスクロス、及び前記拘束ロスを用いて、前記音声強調マスク

推定モデルが含むパラメータ、及び前記雑音強調マスク推定モデルが含むパラメータを更新する

情報処理方法。

[請求項8]

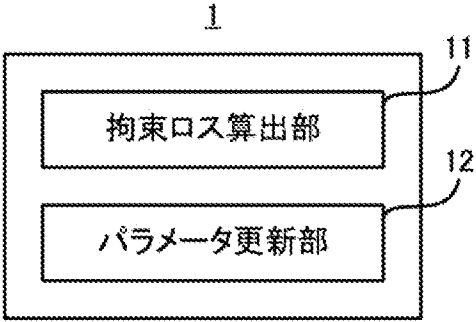
コンピュータに、

雑音混合音声が入力された音声強調マスク推定モデルが出力した推定音声強調マスク、及び前記雑音混合音声が入力された雑音強調マスク推定モデルが出力した推定雑音強調マスクを用いて拘束ロスを算出し、

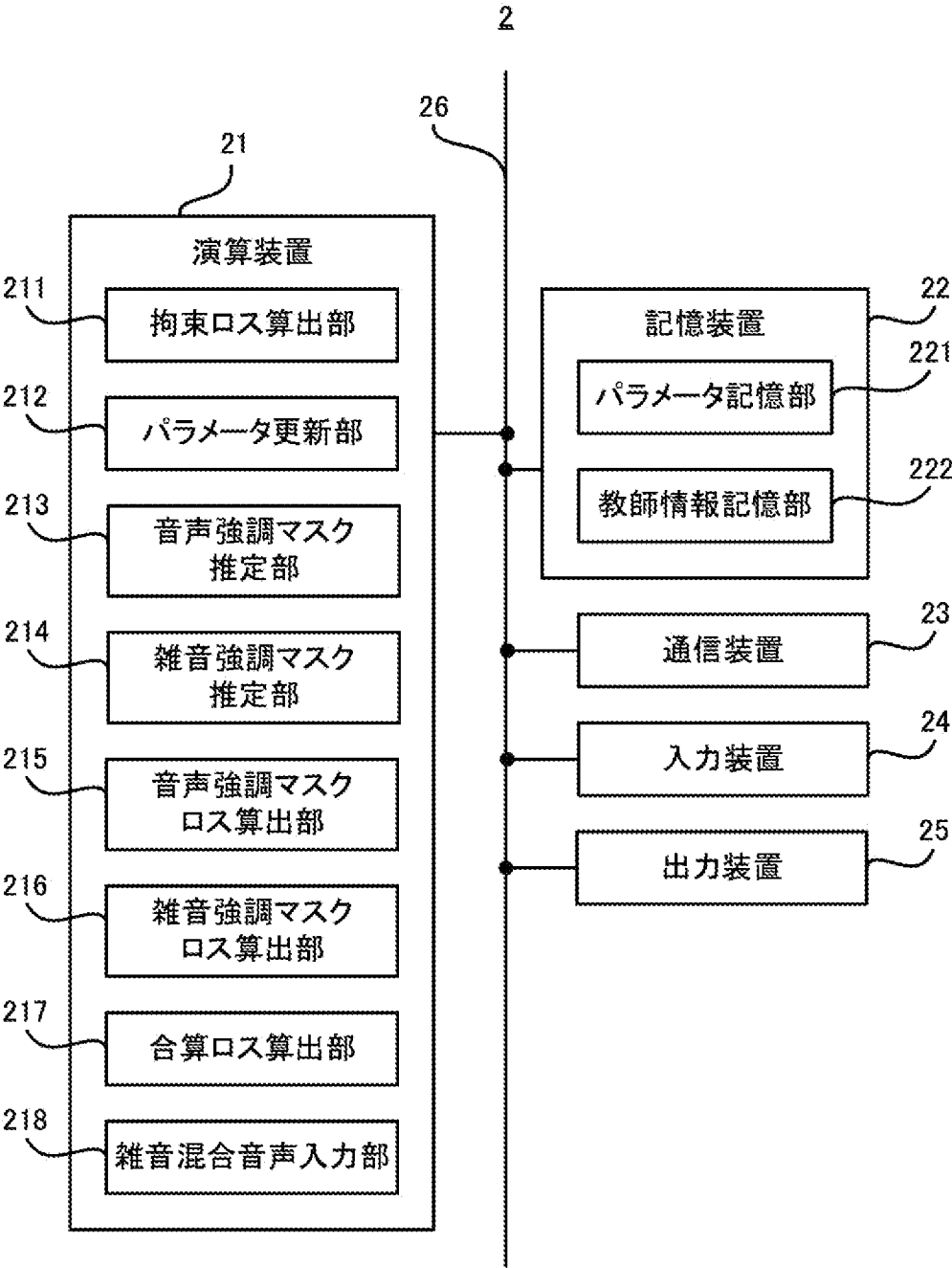
前記推定音声強調マスクと雑音混合音声の中の音声と雑音が既知な場合に計算される目標音声強調マスクとの違いを示す音声強調マスクロス、前記推定雑音強調マスクと目標雑音強調マスクとの違いを示す雑音強調マスクロス、及び前記拘束ロスを用いて、前記音声強調マスク推定モデルが含むパラメータ、及び前記雑音強調マスク推定モデルが含むパラメータを更新する

情報処理方法を実行させるためのコンピュータプログラムが記録された記録媒体。

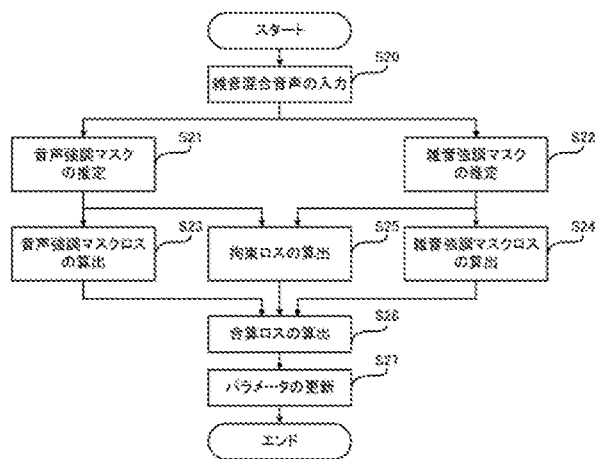
[図1]



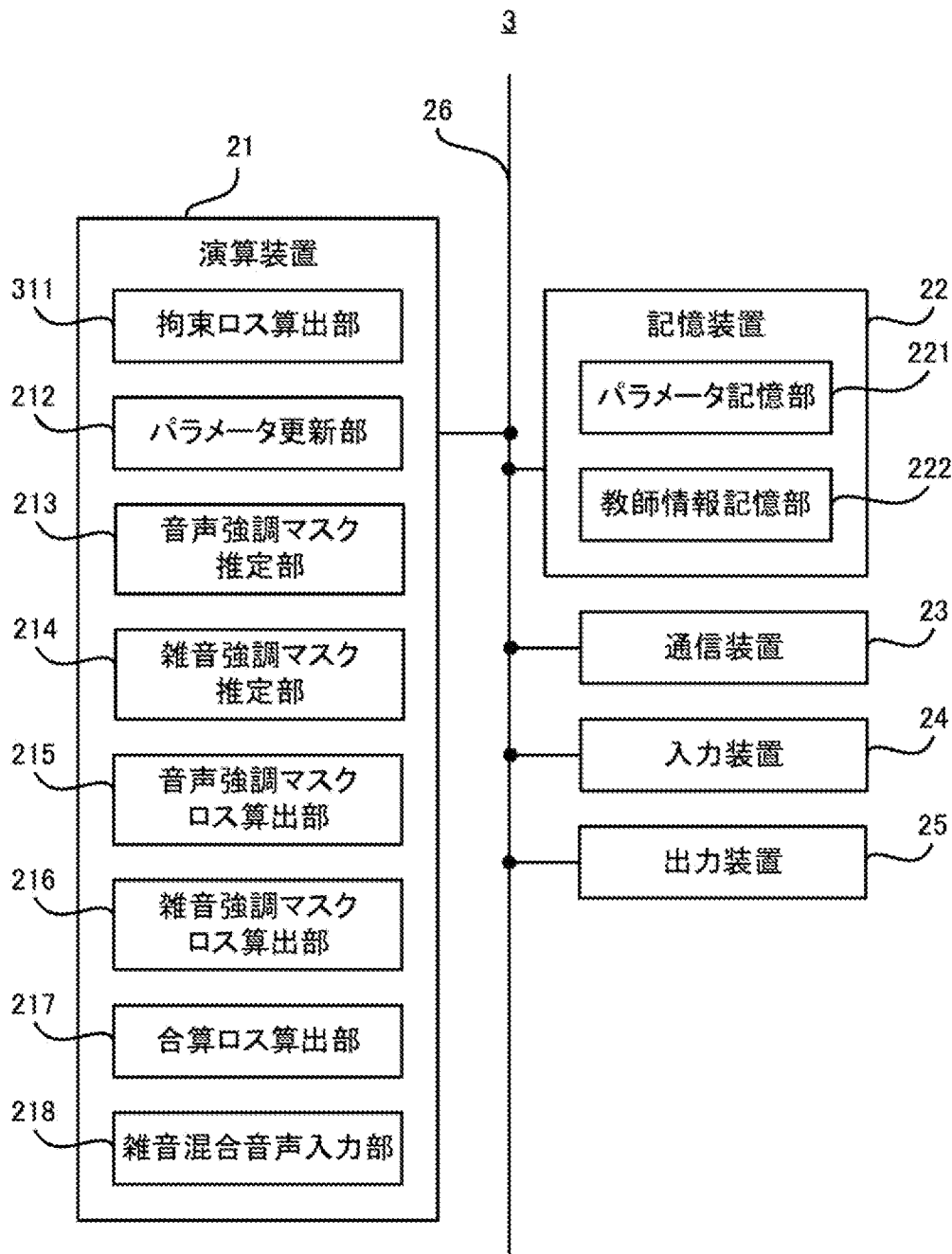
[図2]



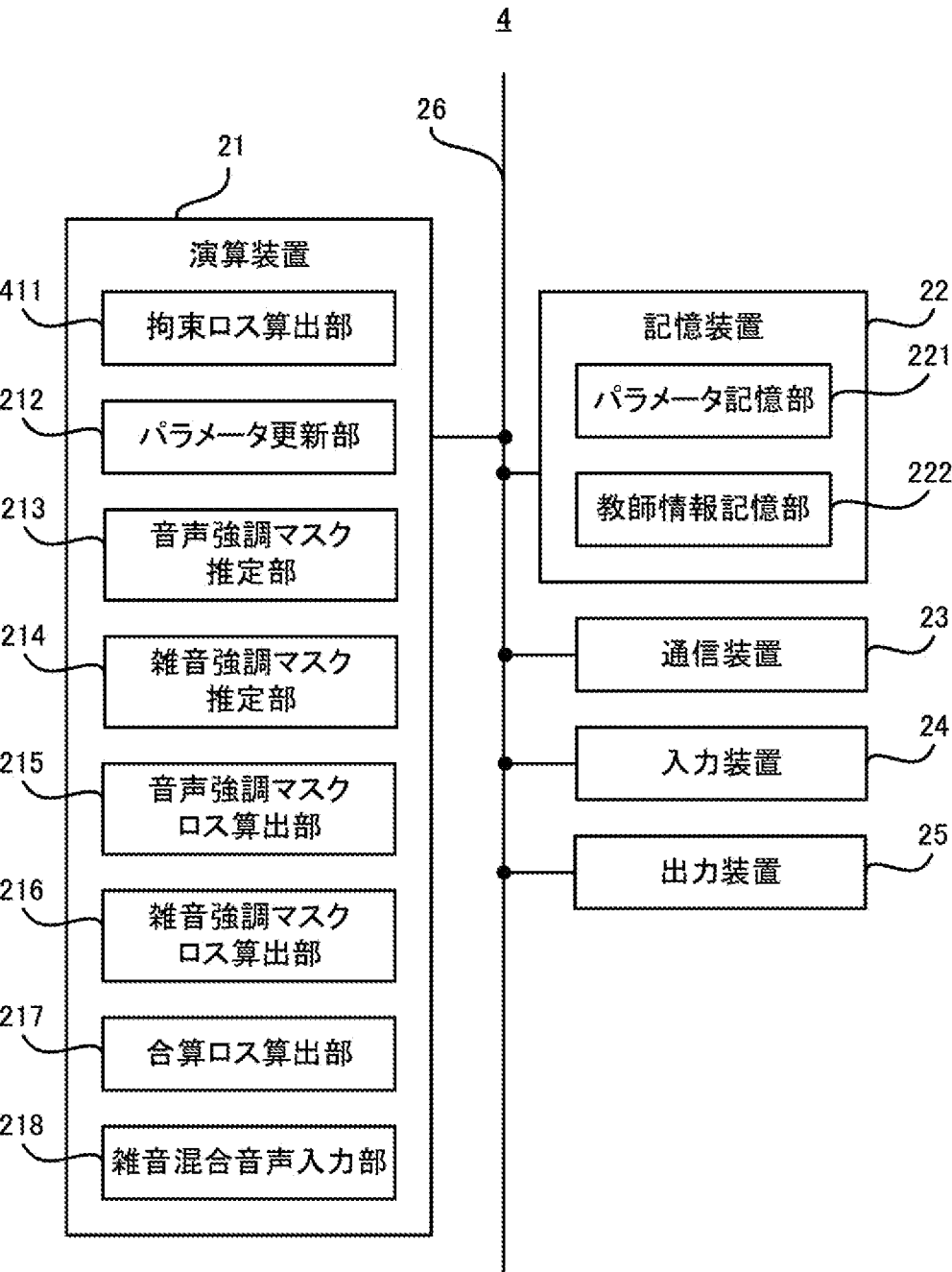
[図3]



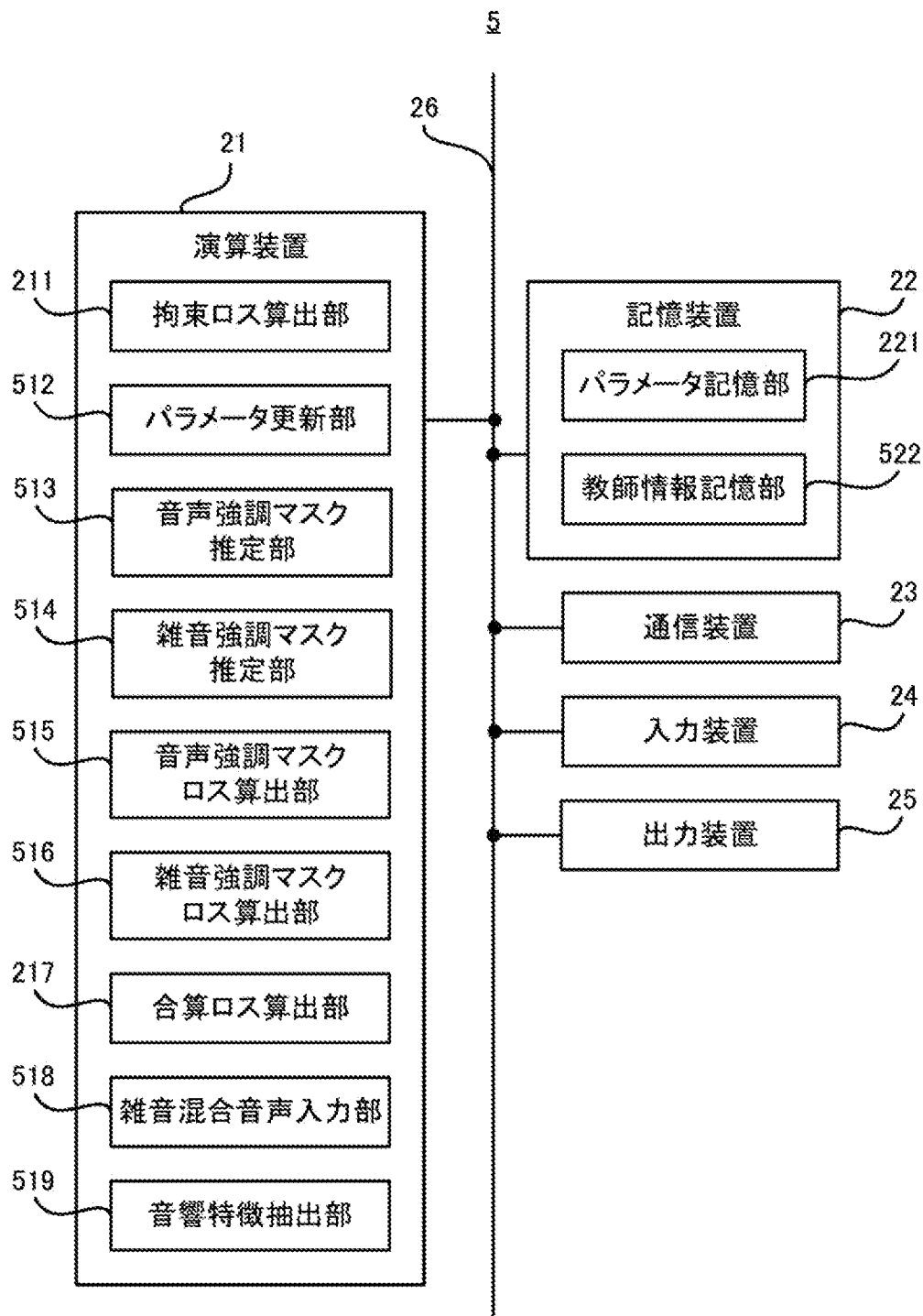
[図4]



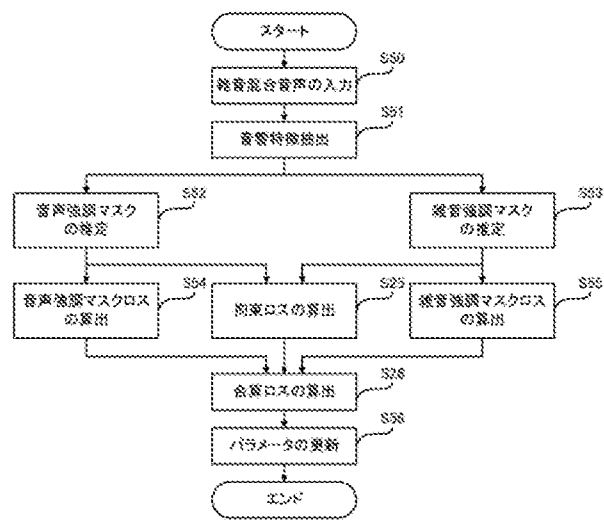
[図5]



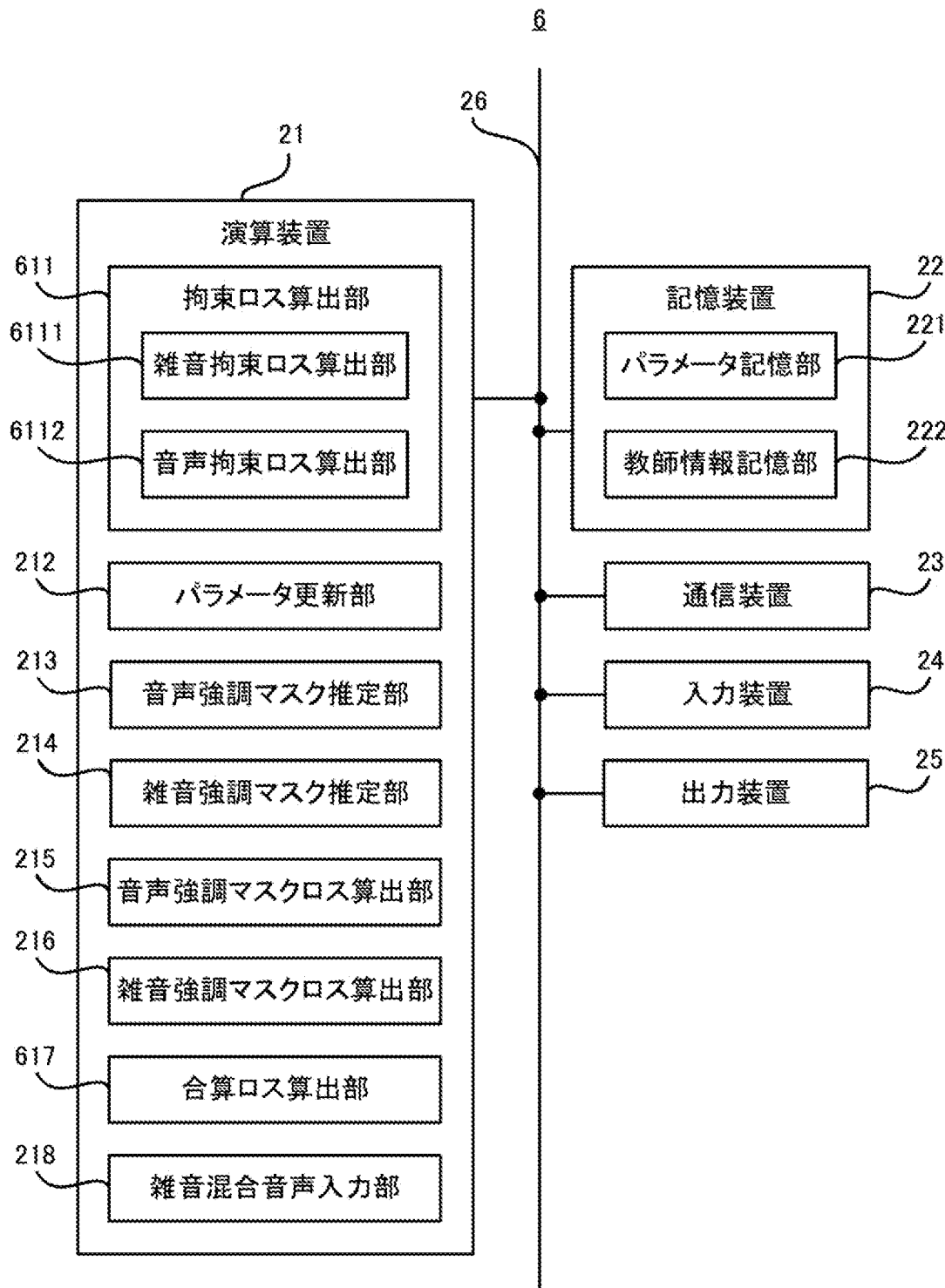
[図6]



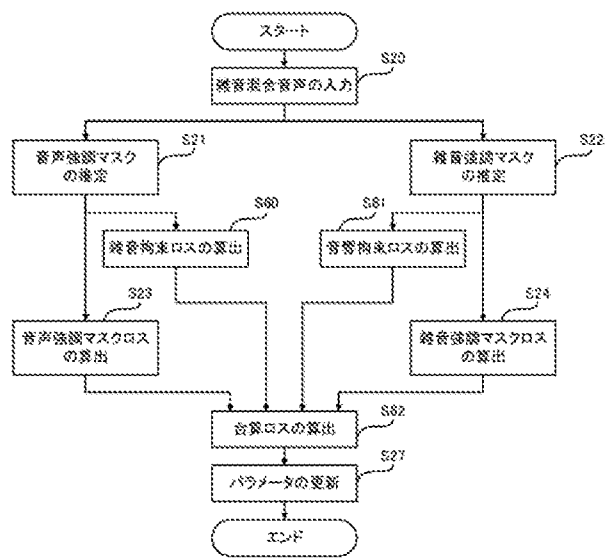
[図7]



[図8]



[図9]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2024/000990

A. CLASSIFICATION OF SUBJECT MATTER G10L 21/0272 (2013.01)i FI: G10L21/0272 100B 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) G10L21/02-21/0364 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 1971-2024 Registered utility model specifications of Japan 1996-2024 Published registered utility model applications of Japan 1994-2024 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.
A	ERDOGAN, Hakan et al., Investigations on Data Augmentation and Loss Functions for Deep Learning Based Speech-Background Separation, Proc. of the Interspeech2018, 02 September 2018, pages 3499-3503 particularly, 3. Mask Prediction - 4. Loss Functions for Mask Prediction	1-8
A	角田遼太 他, 妨害発話環境下におけるマルチモーダル音声認識モデルの事前学習方式の検討, 日本音響学会2022年春季研究発表会講演論文集CD-ROM, 23 February 2022, pages 1057-1060, (TSUNODA, Ryota et al., Pre-training method for audio-visual speech recognition in interference speaker environment, Reports of the 2022 spring meeting the Acoustical Society of Japan) particularly, 2 音声認識モデル - 3 提案手法, non-official translation (2 Speech recognition model - 3 Proposed method)	1-8
☐ Further documents are listed in the continuation of Box C. ☐ 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 “D” document cited by the applicant in the international application “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 27 March 2024		Date of mailing of the international search report 09 April 2024
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.

A. 発明の属する分野の分類（国際特許分類（IPC））

G10L 21/0272(2013.01)i
FI: G10L21/0272 100B

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））
G10L21/02-21/0364

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 1971-2024年
日本国実用新案登録公報 1996-2024年
日本国登録実用新案公報 1994-2024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
A	ERDOGAN, Hakan et al., "Investigations on Data Augmentation and Loss Functions for Deep Learning Based Speech-Background Separation", Proc. of the Interspeech2018, 2018.09.02, pp.3499-3503 特に、「3. Mask Prediction」-「4. Loss Functions for Mask Prediction」	1-8
A	角田遼太 他, "妨害発話環境下におけるマルチモーダル音声認識モデルの事前学習方式の検討", 日本音響学会2022年春季研究発表会講演論文集CD-ROM, 2022.02.23, pp.1057-1060 特に、「2 音声認識モデル」-「3 提案手法」	1-8

☐ C欄の続きにも文献が列挙されている。

☐ 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、一般的技术水準を示すもの

"D" 国際出願で出願人が先行技術文献として記載した文献

"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

"I" 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

27.03.2024

国際調査報告の発送日

09.04.2024

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）

菊池 智紀 5Z 3352

電話番号 03-3581-1101 内線 3591