

(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
2020年3月5日(05.03.2020)



(10) 国際公開番号

WO 2020/045313 A1

(51) 国際特許分類:
G10L 25/51 (2013.01) *G10L 21/0208* (2013.01)
G10L 15/28 (2013.01) *G10L 25/30* (2013.01)

(21) 国際出願番号: PCT/JP2019/033184

(22) 国際出願日: 2019年8月23日(23.08.2019)

(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 2018-163856 2018年8月31日(31.08.2018) JP

(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).

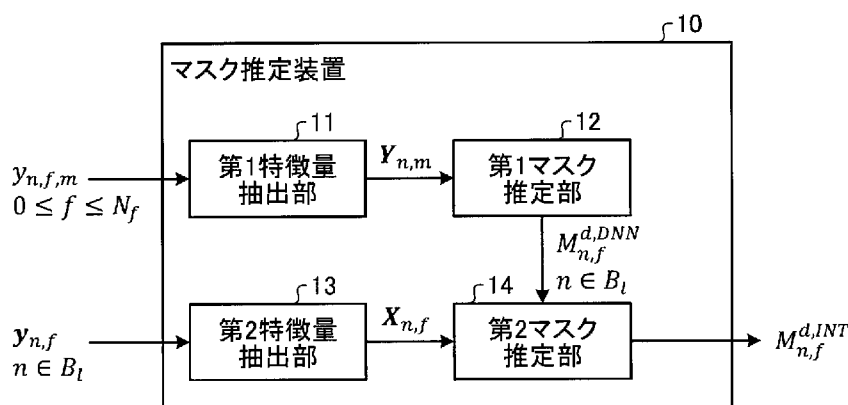
(72) 発明者: 中谷 智広 (NAKATANI, Tomohiro);
〒1808585 東京都武蔵野市緑町3丁目9-1 1

N T T 知的財産センタ内 Tokyo (JP). デルクロア マーク (DELCROIX, Marc); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 N T T 知的財産センタ内 Tokyo (JP). 木下 慶介 (KINOSHITA, Keisuke); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 N T T 知的財産センタ内 Tokyo (JP). 伊藤 信貴 (ITO, Nobutaka); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 N T T 知的財産センタ内 Tokyo (JP). 荒木 章子 (ARAKI, Shoko); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 N T T 知的財産センタ内 Tokyo (JP).

(74) 代理人: 特許業務法人酒井国際特許事務所 (SAKAI INTERNATIONAL PATENT OFFICE); 〒1000013 東京都千代田区霞が関3丁目8番1号 虎の門三井ビルディング Tokyo (JP).

(54) Title: MASK ESTIMATION DEVICE, MASK ESTIMATION METHOD, AND MASK ESTIMATION PROGRAM

(54) 発明の名称: マスク推定装置、マスク推定方法及びマスク推定プログラム



10 Mask estimation device
11 First feature amount extraction unit
12 First mask estimation unit
13 Second feature amount extraction unit
14 Second mask estimation unit

(57) Abstract: A first mask estimation unit (12) estimates, on the basis of a first feature amount obtained from an object segment's observed signal picked up at each of a plurality of positions, a first mask that is the degree of occupancy of a target signal with respect to the observed signal. A second mask estimation unit (14) estimates, on the basis of an estimation result of the first mask in the object segment, and a second feature amount obtained from the observed signal, a parameter that models the second feature amount, and a second mask that is the degree of occupancy of the target signal with respect to the observed signal.

(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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, 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))

(57) 要約: 第1マスク推定部(12)は、複数の位置で収録した対象セグメントの観測信号から得られる第1の特徴量を基に、観測信号に対する目的信号の占有度である第1のマスクを推定する。第2マスク推定部(14)は、対象セグメントにおける第1のマスクの推定結果と、観測信号から得られる第2の特徴量と、を基に、第2の特徴量をモデル化するパラメータと、観測信号に対する目的信号の占有度である第2のマスクとを推定する。

明 細 書

発明の名称：

マスク推定装置、マスク推定方法及びマスク推定プログラム

技術分野

[0001] 本発明は、マスク推定装置、マスク推定方法及びマスク推定プログラムに関する。

背景技術

[0002] 従来、音声を観測して得られた観測信号から、当該観測信号における目的の信号の占有度であるマスクを推定する技術が知られている。ここで、推定されたマスクは、自動音声認識（A S R : automatic speech recognition）における雑音除去のためのビームフォーミング等に用いられる。

[0003] 非特許文献 1 には、複数のマイクロホンで収録された観測信号から精度良くマスクを推定するために、ニューラルネットワークを用いたマスク推定の方法と、空間クラスタリングによりマスクを推定する方法とを組み合わせる技術が開示されている。

先行技術文献

非特許文献

[0004] 非特許文献1: Tomohiro Nakatani, Nobutaka Ito, Takuya Higuchi, Shoko Araki, and Keisuke Kinoshita, “INTEGRATING DNN-BASED AND SPATIAL CLUSTERING-BASED MASK ESTIMATION FOR ROBUST MVDR BEAMFORMING,” Proc. IEEE ICASSP2017, pp.286-290, 2017.

発明の概要

発明が解決しようとする課題

[0005] 非特許文献 1 に開示された技術は、全ての観測信号を読み込んでからバッチ処理によりマスクを推定するものである。ここで、マスクに基づく自動音声認識をスマートスピーカ等へ応用することを考えた場合、時々刻々と変化する環境に応じてマスクを逐次的に推定するオンライン型の技術が要求され

ることもある。しかしながら、非特許文献 1 に開示された技術では、オンラインでマスク推定を行うことができなかった。このように、従来の技術には、オンラインで精度良くマスク推定を行うことができない場合があるという問題がある。

課題を解決するための手段

[0006] 上述した課題を解決し、目的を達成するために、マスク推定装置は、連続する時間のセグメントのうち、処理対象とするセグメントを対象セグメントとして、複数の位置で収録した対象セグメントの観測信号から得られる第 1 の特徴量を基に、前記対象セグメントの観測信号に対する目的信号の占有度である第 1 のマスクを推定する第 1 マスク推定部と、前記対象セグメントにおける前記第 1 のマスクの推定結果と、前記対象セグメントの観測信号から得られる第 2 の特徴量と、を基に、前記第 2 の特徴量をモデル化するパラメータと前記観測信号に対する前記目的信号の占有度である第 2 のマスクとを推定する第 2 マスク推定部と、を有することを特徴とする。

発明の効果

[0007] 本発明によれば、オンラインで精度良くマスク推定を行うことができる。

図面の簡単な説明

[0008] [図1]図 1 は、第 1 の実施形態に係るマスク推定装置の構成の一例を示す図である。

[図2]図 2 は、第 1 の実施形態に係る第 2 マスク推定部の構成の一例を示す図である。

[図3]図 3 は、第 1 の実施形態に係るマスク推定装置の処理の流れを示すフローチャートである。

[図4]図 4 は、第 1 の実施形態に係るマスク推定装置の処理の流れを示すフローチャートである。

[図5]図 5 は、第 2 の実施形態に係るマスク推定装置の構成の一例を示す図である。

[図6]図 6 は、第 2 の実施形態に係るマスク推定装置の処理の流れを示すフロ

ーチャートである。

[図7]図7は、実験に用いた音声のデータを示す図である。

[図8]図8は、実験におけるハイパーパラメータを示す図である。

[図9]図9は、実験結果を示す図である。

[図10]図10は、実験結果を示す図である。

[図11]図11は、マスク推定プログラムを実行するコンピュータの一例を示す図である。

発明を実施するための形態

[0009] 以下に、本願に係るマスク推定装置、マスク推定方法及びマスク推定プログラムの実施形態を図面に基づいて詳細に説明する。なお、本発明は、以下に説明する実施形態により限定されるものではない。

[0010] [第1の実施形態]

第1の実施形態のマスク推定装置には、連続する時間のセグメントのうちの対象セグメントにおいて複数の位置で収録した観測信号、又は観測信号から抽出された特徴量が入力される。ここで、観測信号には、目的音源から発生する目的音声及び背景雑音の両方が含まれる。また、観測信号は、複数の異なる位置に設置されたマイクロホンによって収録される。

[0011] マスク推定装置10は、観測信号から目的信号を抽出するためのマスクを推定することができる。この場合、マスクは、各時間周波数点における、観測信号を目的音声の信号が占有している確率である。つまり、マスクは、各時間周波数点における観測信号に対する目的音声の信号の占有度である。同様に、マスク推定装置10は、観測信号から雑音を抽出するためのマスクを推定することができる。この場合、マスクは、各時間周波数点における、観測信号を雑音信号が占有している確率である。つまり、マスクは、各時間周波数点における観測信号に対する雑音信号の占有度である。以降、目的音声の信号を目的信号と呼び、目的音声以外の音の信号を雑音信号と呼ぶ。例えば、目的音声は、特定の話者が発する音声である。

[0012] [第1の実施形態の構成]

図 1 を用いて、第 1 の実施形態のマスク推定装置の構成について説明する。図 1 は、第 1 の実施形態に係るマスク推定装置の構成の一例を示す図である。図 1 に示すように、マスク推定装置 10 は、第 1 特徴量抽出部 11、第 1 マスク推定部 12、第 2 特徴量抽出部 13 及び第 2 マスク推定部 14 を有する。

[0013] まず、マスク推定装置 10 は、ミニバッチ単位で観測信号の入力を受け付ける。ここで、ミニバッチは所定の時間セグメントの単位である。例えば、観測信号の収録を開始してから 0 ms ~ 500 ms を 1 番目のミニバッチに設定し、500 ms ~ 750 ms を 2 番目のミニバッチに設定し、以降 250 ms ごとにミニバッチを設定することができる。また、各ミニバッチの長さは一定であってもよいし、異なってもよい。以降、 B_1 は、1 番目のミニバッチを表すものとする。つまり、観測信号全体を所定時間ごとに分割した部分区間をミニバッチという。

[0014] マスク推定装置 10 は、ミニバッチ単位で入力された観測信号に対し、短時間周波数分析に基づき短時間フレームごとの周波数領域信号に変換する。なお、マスク推定装置 10 には、この変換が行われた後の観測信号が入力されてもよい。以下では、この変換のために、一例として、短時間フーリエ変換 (STFT: short-time Fourier transform) を用いるものとして説明する。 $y_{n,f,m}$ は、観測信号の STFT を表すものとする。ここで、 n 及び f は、それぞれ時間及び周波数のインデックスである。また、 m は、観測信号を収録したマイクロホンを表すインデックスである。また、 $1 \leq n \leq N_t$ 、 $0 \leq f \leq N_f$ 、及び $1 \leq m \leq N_m$ が成り立つものとする。

[0015] 第 1 特徴量抽出部 11 は、観測信号の STFT $y_{n,f,m}$ からスペクトル特徴量 $Y_{n,m}$ を抽出する。具体的には、第 1 特徴量抽出部 11 は、(1) 式に示すように、 $y_{n,f,m}$ の対数を要素とするベクトル $Y_{n,m}$ をスペクトル特徴量として抽出する。

[0016] [数1]

$$Y_{n,f,m} = \log|y_{n,f,m}| \quad \cdots (1)$$

[0017] [数2]

$$Y_{n,m} = [Y_{n,0,m}, \dots, Y_{n,N_f,m}] \quad \dots (2)$$

[0018] 第1マスク推定部12は、1つ、もしくは、複数の位置で収録した対象セグメントの観測信号から得られるスペクトル特徴量を基に、第1のマスクを推定する。ここで、対象セグメントは、マスク推定装置10に入力された観測信号に対応するミニバッチである。また、スペクトル特徴量は、第1の特徴量の一例である。

[0019] 具体的には、第1マスク推定部12は、ニューラルネットワークを用いて第1のマスクを推定する。第1マスク推定部12は、第1特徴量抽出部11によって抽出されたスペクトル特徴量 $Y_{n,m}$ をニューラルネットワークに入力し、当該ニューラルネットワークの出力として、 m 番目のマイクで収録した観測信号のみに基づきマスク $M_{n,f}^{d,DNN}$ を得る。

[0020] また、第1マスク推定部12は、複数のマイクのそれぞれで収録した観測信号に基づきマスクを推定し、複数のマスクの推定値を得たのち、複数のマスクの推定値を統合して1つのマスクの推定値とすることもできる。マスクの統合法としては、推定値間で平均値をとる、中央値（メジアン）をとる等の方法がある。

[0021] 要するに、第1マスク推定部12は、複数の位置で収録した対象セグメントの観測信号から得られる第1の特徴量を基に、前記対象セグメントの観測信号に対する目的信号の占有度である第1のマスクを推定すればよく、第1のマスクの計算には、対象セグメントの観測信号の一部（例えば、 m 番目のマイクについての観測信号）を用いてもよいし、観測信号全体（ M 個のマイクについての観測信号）を用いてもよい。

[0022] 第1マスク推定部12では、逐次的に入力されるスペクトル特徴量をオンラインで処理可能なニューラルネットワークが用いられる。例えば、第1マスク推定部12では、LSTM（long short-term memory）ネットワークが用いられる。また、ニューラルネットワークのパラメータは、目的音声又は雑音の両方を含むシミュレーション音声等を用いて学習済みであるものと

する。

[0023] ここで、 d は、0又は1をとる。また、第1マスク推定部12は、 $M_{n,f}^{0,DNN}$ 、及び $M_{n,f}^{1,DNN}$ の2種類のマスクを得ることができる。 $M_{n,f}^{0,DNN}$ は、時間周波数点 (n, f) における観測信号から雑音信号を抽出するマスクである。一方、 $M_{n,f}^{1,DNN}$ は、時間周波数点 (n, f) における観測信号から目的信号を抽出するマスクである。 $M_{n,f}^{d,DNN}$ は、0から1の範囲の数値である。

[0024] また、 $M_{n,f}^{0,DNN} + M_{n,f}^{1,DNN} = 1$ のように決めておけば、第1マスク推定部12は、いずれか一方のマスクをニューラルネットワークから出力し、他方のマスクは、当該出力したマスクを1から引くことで計算することもできる。このため、第1マスク推定部12は、ニューラルネットワークから $M_{n,f}^{0,DNN}$ 及び $M_{n,f}^{1,DNN}$ の両方を出力するようにしてもよいし、どちらか一方を出力するようにしてもよい。

[0025] 第2特徴量抽出部13は、ベクトル $y_{n,f}$ から、(4)式に示すように、空間特徴量 $X_{n,f}$ を抽出する。つまり、第2特徴量抽出部13は、対象セグメントの観測信号から空間特徴量 $X_{n,f}$ を抽出する。また、(3)式に示すように、ベクトル $y_{n,f}$ の要素は、マイクロホンごとの観測信号のSTFTである。ここで、 $\|\cdot\|$ は、ユークリッドノルムを表す。また、 T は、非共役転置を表す。

[0026] [数3]

$$y_{n,f} = [y_{n,f,1}, \dots, y_{n,f,N_m}]^T \quad \dots (3)$$

[0027] [数4]

$$X_{n,f} = \frac{y_{n,f}}{\|y_{n,f}\|} \quad \dots (4)$$

[0028] 第2マスク推定部14は、対象セグメントにおける第1のマスクの推定結果と、対象セグメントの観測信号から得られる空間特徴量と、を基に、対象セグメントの空間特徴量をモデル化する空間パラメータと観測信号に対する目的信号の占有度である第2のマスクとを推定する。ここで、空間特徴量は

、第2の特徴量の一例である。

[0029] 具体的には、第2マスク推定部14は、対象セグメントごとに、空間パラメータを条件とした場合の空間特徴量の分布モデル及び第1のマスクを基に、第2のマスクを推定する。このとき、第2マスク推定部14は、空間特徴量の分布モデルとして、複素角度ガウス混合モデル（cACGMM：complex angular central Gaussian mixture model）を用いる。cACGMMは（5）式のように定義される。

[0030] [数5]

$$\mathcal{A}(\mathbf{X}|d_{n,f} = d; \theta^{SC}) = \frac{(N_m - 1)!}{2\pi^{(N_m)} \det \mathbf{R}_f^d} \frac{1}{(\mathbf{X}^H (\mathbf{R}_f^d)^{-1} \mathbf{X})^{N_m}} \quad \dots (5)$$

[0031] ここで、パラメータ集合 θ^{SC} は、 $\{\{w_d^f\}, \{R_d^f\}\}$ と表される。また、 w_d^f は、混合重みであり、 $d_{n,f}$ の事前確率である。つまり、 $w_d^f = p(d_{n,f} = d)$ と書ける。なお、後述（[0046]段落）のように、本実施形態では、 w_d^f は、第1のマスクと等価であり、その推定値で置き換えられるものとする。また、（5）式は、 d が与えられたときの、複素角度ガウス（cACG：complex angular central Gaussian）分布で定義される空間特徴量 $\mathbf{X}$ の条件付き分布を表す。このとき、空間パラメータ R_d^f は、複素角度ガウス分布の形状を定めるパラメータであり、 $N_m \times N_m$ 次元の正定値エルミート行列である。ここで、 $\det$ は行列式を表す。また、 H は共役転置を表す。

[0032] 第2マスク推定部14は、上記の複素角度ガウス混合モデルを用いて、EM（expectation-maximization）アルゴリズムにより第2のマスクを推定する。図2は、第1の実施形態に係る第2マスク推定部の構成の一例を示す図である。図2に示すように、第2マスク推定部14は、設定部141、第1更新部142、第2更新部143、判定部144、及び記憶部145を有する。なお、図2では記憶部145が第2マスク推定部14内に設けられているが、記憶部145を第2マスク推定部14の外側、つまり、マスク推定装置10内の記憶部として設けてもよいことは言うまでもない。

[0033] 設定部 141 は、対象セグメントにおける第 2 のマスク及び空間パラメータの初期値として、対象セグメントに対して推定された第 1 のマスク及び 1 つ前のセグメントにおける空間パラメータをそれぞれ設定する。具体的には、設定部 141 は、(6) 式のように第 2 のマスク $M_{n,f}^{d,INT}$ の初期値を設定する。なお、第 2 マスク推定部 14 は、第 1 マスク推定部 12 から第 1 のマスク $M_{n,f}^{d,DNN}$ を取得する。また、対象セグメントに対応するミニバッチを B_l とすると、設定部 141 は、(7) 式のように空間パラメータ $R_{f,l}^d$ の初期値を設定する。また、設定部 141 は、第 1 のマスクの累積和 $\Lambda_{f,l-1}^d$ を (8) 式のように設定する。

[0034] [数6]

$$M_{n,f}^{d,INT} = M_{n,f}^{d,DNN} \quad \dots (6)$$

[0035] [数7]

$$R_{f,l}^d = R_{f,l-1}^d \quad \dots (7)$$

[0036] [数8]

$$\Lambda_{f,l}^d = \Lambda_{f,l-1}^d + \sum_{n \in B_l} M_{n,f}^{d,DNN} \quad \dots (8)$$

[0037] 第 1 更新部 142 は、対象セグメントまでの第 1 のマスクの累積和と、対象セグメントの空間特徴量及び第 2 のマスクと、を基に空間パラメータを更新する。具体的には、第 1 更新部 142 は、(9) 式のように空間パラメータ $R_{f,l}^d$ を更新する。このとき、第 1 更新部 142 は、(10) 式のように更新空間パラメータ $R_{f,new}^d$ を計算する。

[0038] [数9]

$$R_{f,l}^d = \frac{\Lambda_{f,l-1}^d}{\Lambda_{f,l}^d} R_{f,l-1}^d + \frac{1}{\Lambda_{f,l}^d} R_{f,new}^d \quad \dots (9)$$

[0039] [数10]

$$R_{f,new}^d = N_m \sum_{n \in B_l} \frac{M_{n,f}^{d,INT} \mathbf{X}_{n,f} \mathbf{X}_{n,f}^H}{\mathbf{X}_{n,f}^H (\mathbf{R}_{f,l}^d)^{-1} \mathbf{X}_{n,f}} \quad \dots (10)$$

[0040] 第2更新部143は、対象セグメントの空間特徴量、第1のマスク、及び空間パラメータを基に、第2のマスクを更新する。具体的には、第2更新部143は、(11)式のように第2のマスク $M_{n,f}^{d,INT}$ を更新する。

[0041] [数11]

$$M_{n,f}^{d,INT} = \frac{M_{n,f}^{d,DNN} \mathcal{A}(X_{n,f}|d; \theta^{SC})}{\sum_d M_{n,f}^{d,DNN} \mathcal{A}(X_{n,f}|d; \theta^{SC})} \quad \dots (11)$$

[0042] 判定部144は、第2更新部143によって第2のマスクが更新された場合、所定の収束条件が満たされているか否かを判定し、収束条件が満たされていないと判定した場合、第1更新部142及び第2更新部143に処理をさらに実行させる。つまり、第1更新部142及び第2更新部143は、所定の収束条件が満たされるまで処理を繰り返すことになる。その際、繰り返しのたびに第2のマスク及び空間パラメータが更新され、第2のマスクの目的音声の抽出精度が向上していく。

[0043] また、判定部144の収束条件は、繰り返し数が閾値を超えたことであってもよい。このとき、繰り返し数の閾値は1回とすることができる。すなわち、第1更新部142及び第2更新部143は、1つのミニバッチに対し、それぞれ1回のみ更新処理を行うようにしてもよい。また、判定部144が収束を判定する条件は、1回の更新における第2のマスクの更新量や空間パラメータの更新量が一定値以下になったことであってもよい。

[0044] また、判定部144は、(12)式で表される尤度関数 $\mathcal{L}(\theta^{SC})$ の値の更新量が一定値以下になった場合に収束したと判定してもよい。 X_l は、ミニバッチ B_l までに観測された空間特徴量 $X_{n,f}$ の集合である。また、 Y_l は、ミニバッチ B_l までに観測された空間パラメータ $Y_{n,m}$ の集合である。また、 θ^{DNN} は、第1マスク推定部12のニューラルネットワークのパラメータである。

[0045] [数12]

$$\mathcal{L}(\theta^{SC}) = p(X_l, Y_l; \theta^{DNN}, \theta^{SC}) \quad \dots (12)$$

[0046] また、(12)式は、(13)式のように書き換えられる。このとき、(13)式の右辺の $p(d_{n,f}=d|y_l;\theta^{DNN})$ は、第1マスク推定部12によって推定される第1のマスク $M_{n,f}^{d,DNN}$ と等価であるとみなすことができる。したがって、本実施形態では、 $p(d_{n,f}=d|y_l;\theta^{DNN})$ を $M_{n,f}^{d,DNN}$ に置き換えて、尤度関数を最大化する。このため、第2マスク推定部14は、非特許文献1に記載された方法と同様の方法で、各ミニバッチごとに尤度関数 $L(\theta^{SC})$ を最大化し、第2のマスク $M_{n,f}^{d,INT}$ 及びパラメータ θ^{SC} の推定を行うことができる。また、各ミニバッチで推定された空間パラメータを記憶部が記憶し、次のミニバッチで空間パラメータの初期値として用い、更新するようにすることで、ミニバッチごとにバラバラに尤度関数を最大化するよりも、高い精度でマスク推定をすることができる。

[0047] [数13]

$$L(\theta^{SC}) = \prod_{n,f} \sum_d p(X_{n,f}|d_{n,f}=d;\theta^{SC})p(d_{n,f}=d|y_l;\theta^{DNN}) \quad \dots (13)$$

[0048] 記憶部145は、前のセグメントでの計算値であって、対象セグメントの初期設定で用いられる値を記憶する。つまり、記憶部145は、ミニバッチ B_{l-1} において計算された空間パラメータ $R_{f,l-1}^d$ 及び第1のマスクの累積和 $\Lambda_{f,l-1}^d$ を記憶する。そして、設定部141は、ミニバッチ B_l において空間パラメータ $R_{f,l}^d$ 及び第1のマスクの累積和 $\Lambda_{f,l}^d$ を設定する際に、記憶部145から空間パラメータ $R_{f,l-1}^d$ 及び第1のマスクの累積和 $\Lambda_{f,l-1}^d$ を取得する。

[0049] なお、ミニバッチが先頭である場合、すなわち、 $l=1$ の場合、空間パラメータ $R_{f,l-1}^d$ は未計算である。この場合、設定部141は、非特許文献1に記載された方法と同様に、空間パラメータの初期値 $R_{f,0}^d$ に所定の学習データを使って学習した値を設定してもよい。例えば、目的信号の空間パラメータ $R_{f,0}^1$ の学習データは、特定の話者が雑音のない環境で発話した際に得られる観測信号である。また、設定部141は、空間パラメータの初期値 $R_{f,0}^1$ に単位行列を設定してもよい。さらに、雑音信号の空間パラメータ $R_{f,0}$

°は、雑音のみが含まれている観測信号から推定してもよい。

[0050] [第1の実施形態の処理]

図3を用いて、本実施形態のマスク推定装置10の処理の流れを説明する。図3は、第1の実施形態に係るマスク推定装置の処理の流れを示すフローチャートである。

[0051] 図3に示すように、まず、マスク推定装置10は、ミニバッチ単位の観測信号の入力を受け付ける（ステップS11）。ここで、マスク推定装置10は、観測信号のSTFTを計算してもよい。また、マスク推定装置10に入力される観測信号は、STFTが行われたものであってもよい。

[0052] 次に、マスク推定装置10は、マイクロホンごとの観測信号のSTFTからスペクトル特徴量を抽出する（ステップS12）。そして、マスク推定装置10は、スペクトル特徴量から第1のマスクを推定する（ステップS13）。このとき、マスク推定装置10は、ニューラルネットワークを用いて第1のマスクを推定することができる。

[0053] さらに、マスク推定装置10は、観測信号のSTFTから空間特徴量を抽出する（ステップS14）。そして、マスク推定装置10は、第1のマスク及び空間特徴量から第2のマスクを推定する（ステップS15）。

[0054] ここで、マスク推定装置10は、未処理のミニバッチがあるか否かを判定する（ステップS16）。未処理のミニバッチがある場合（ステップS16、Yes）、マスク推定装置10は、ステップS11に戻り、次のミニバッチの観測信号の入力を受け付ける。一方、未処理のミニバッチがない場合（ステップS16、No）、マスク推定装置10は処理を終了する。

[0055] 図4を用いて、マスク推定装置10が第2のマスクを推定する処理（図3のステップS15）を詳細に説明する。図4は、第1の実施形態に係るマスク推定装置の処理の流れを示すフローチャートである。

[0056] 図4に示すように、まず、マスク推定装置10は、第2のマスク、空間パラメータ、及び第1のマスクの累積和の初期値を設定する（ステップS151）。次に、マスク推定装置10は、第1のマスクの累積和、空間特徴量及

び第2のマスクを用いて空間パラメータを更新する（ステップS 1 5 2）。

そして、マスク推定装置10は、空間特徴量、第1のマスク、及び空間パラメータを基に第2のマスクを更新する（ステップS 1 5 3）。

[0057] ここで、マスク推定装置10は、第2のマスクが収束したか否かを判定する（ステップS 1 5 4）。マスク推定装置10は、第2のマスクが収束していないと判定した場合（ステップS 1 5 4、N o）、ステップS 1 5 2に戻り、さらに空間パラメータを更新する。一方、マスク推定装置10は、第2のマスクが収束したと判定した場合（ステップS 1 5 4、Y e s）、マスク推定装置10は処理を終了する。

[0058] [第1の実施形態の効果]

これまで説明してきたように、第1マスク推定部12は、連続する時間のセグメントのうち、処理対象とするセグメントを対象セグメントとして、複数の位置で収録した対象セグメントの観測信号から得られる第1の特徴量を基に、対象セグメントの観測信号に対する目的信号の占有度である第1のマスクを推定する。また、第2マスク推定部14は、対象セグメントにおける第1のマスクの推定結果と、対象セグメントの観測信号から得られる第2の特徴量と、を基に、第2の特徴量をモデル化するパラメータと観測信号に対する目的信号の占有度である第2のマスクを推定する。このように、マスク推定装置10は、2つのマスク推定方法を組み合わせることで、最終的なマスクを精度良く推定することができる。さらに、マスク推定装置10は、対象セグメントごとの観測信号に対して逐次的にマスクを推定することができる。このため、第1の実施形態によれば、オンラインで精度良くマスク推定を行うことができる。

[0059] また、マスク推定装置10は、スペクトル特徴量を入力するニューラルネットワークを用いた手法と、分布モデルを用いた手法を組み合わせている。このため、例えば、事前学習したニューラルネットワークのパラメータと観測信号との間にミスマッチがある場合でも、空間パラメータを用いることでマスクの精度を向上させることができる。また、信号対雑音比が特別に悪い

周波数帯がある場合でも、スペクトル特徴量に基づき目的信号の周波数パターンを考慮することで、高精度なマスク推定が可能になる。

[0060] [第2の実施形態]

第2の実施形態では、マスク推定装置10は、先頭のミニバッチから所定のミニバッチまでは、第1のマスクの推定値を第2のマスクの推定値として代用し、以降のミニバッチでは、空間パラメータの計算値を用いて第2のマスクの推定を行う。

[0061] ここで、目的信号を含む観測信号の量が多いほど、目的信号に対する空間パラメータの精度は向上する。逆に、目的信号を含む観測信号が少ないと、計算された目的信号に対する空間パラメータの精度が低く、実用的でない場合がある。つまり、目的信号を含む観測信号が少ないミニバッチで計算された目的信号に対する空間パラメータを第2マスク推定部14の推定に用いると、結果として推定される対象セグメントにおける第2のマスクの推定精度も低くなってしまうことがある。そこで、第2の実施形態では、マスク推定装置10は、十分な量の目的信号を含む観測信号を用いて空間パラメータが計算されるようになるまでの間、第1のマスクを第2のマスクの推定値として代用し、十分な量の目的信号を含む観測信号を用いて空間パラメータが計算されてからは、空間パラメータの計算値（推定値）を用いて第2のマスクの推定を行う。

[0062] [第2の実施形態の構成]

図5に示すように、第2の実施形態では、マスク推定装置10は、第1の実施形態と同様の処理部に加えて、制御部15をさらに有する。図5は、第2の実施形態に係るマスク推定装置の構成の一例を示す図である。

[0063] 制御部15は、マスクの推定対象のセグメントである対象セグメントまでの観測信号に含まれている目的信号の量が、所定の閾値を超えているか否かを判定する。ここで、制御部15は、目的信号の量が閾値を超えている場合は、第1の実施形態と同様に、第2マスク推定部14が、空間パラメータの計算値を用いて、第2のマスクを推定するように制御する。一方、制御部1

5は、目的信号の量が閾値を超えていない場合は、第2マスク推定部14が、第1のマスクの推定値を第2のマスクを推定値として代用するように制御する。これにより、第2の実施形態において、マスク推定装置10は、空間パラメータに適正な初期値が与えられない場合であっても、第2のマスクを精度良く推定することができる。

[0064] 制御部15は、所定の推定値を基に、対象セグメントを含む過去のセグメントの観測信号に含まれる目的信号の量が閾値を超えているか否かを判定する。例えば、制御部15は、目的信号についての第1のマスクの累積和 $\Lambda_{f,1}$ が閾値を超えているか否かを判定する。ここで、 $\Lambda_{f,1}$ は、目的信号が含まれるフレームの数が多いほど大きくなる。

[0065] 制御部15の判定の対象は、 $\Lambda_{f,1}$ に限られない。例えば、処理したミニバッチの数が増えるほど、目的信号を含む観測信号の量は増える（少なくとも減ることはない）ため、制御部15は、処理したミニバッチの数によって判定を行ってもよい。また、制御部15は、観測信号中に音声が含まれるか否かを判定する音声区間検出に基づいて検出した音声区間の長さを用いて判定を行ってもよい。

[0066] [第2の実施形態の処理]

図6を用いて、マスク推定装置10が第2のマスクを推定する処理を詳細に説明する。図6は、第2の実施形態に係るマスク推定装置の処理の流れを示すフローチャートである。図6の処理は、図4の処理と同様に、図3のステップS15に対応している。このため、図6の処理は、図3のステップS11で入力されたミニバッチ単位の観測信号を対象とする。

[0067] 図6に示すように、まず、マスク推定装置10は、各ミニバッチにおいて、第2のマスク、空間パラメータ、及び第1のマスクの累積和の初期値を設定する（ステップS251）。次に、マスク推定装置10は、第1のマスクの累積和、空間特徴量及び第2のマスクを用いて空間パラメータを更新する（ステップS252）。

[0068] ここで、マスク推定装置10は、第1のマスクの累積和が閾値以上である

か否かを判定する（ステップS 2 5 3）。第1のマスクの累積和が閾値以上である場合（ステップS 2 5 3、Y e s）、マスク推定装置10は、ステップS 2 5 2で更新した空間パラメータを用いて第2のマスクを更新する（ステップS 2 5 4）。一方、第1のマスクの累積和が閾値以上でない場合（ステップS 2 5 3、N o）、マスク推定装置10は、第1のマスクの推定値を第2のマスクの推定値の代用として用いる（ステップS 2 5 5）。

[0069] ここで、マスク推定装置10は、空間パラメータの更新が収束したか否かを判定する（ステップS 2 5 6）。マスク推定装置10は、空間パラメータの更新が収束していないと判定した場合（ステップS 2 5 6、N o）、ステップS 2 5 2に戻り、さらに空間パラメータを更新する。一方、マスク推定装置10は、空間パラメータの更新が収束したと判定した場合（ステップS 2 5 6、Y e s）、マスク推定装置10は処理を終了する。

[0070] なお、ここで、マスク推定装置10は、第1のマスクの累積和が閾値以上でない場合でも（ステップS 2 5 3、N o）、ステップS 2 5 2の空間パラメータの更新を行うようにしている。第1のマスクの累積値が閾値以下の場合でも、目的信号に対する空間パラメータの更新を行っていくことで、第1のマスクの累積和が閾値を超えた時点で、目的信号に対する空間パラメータの推定精度を高くすることができるからである。

[0071] [実験結果]

ここで、従来の手法と実施形態とを比較するために行った実験について説明する。実験には、CH i ME - 3の音声認識用の音声データを用いた。音声データは、複数の雑音環境下で新聞を読み上げる音声を、複数のマイクロホンを備えたタブレット端末で収録したものである。また、図7に示すように、音声データは複数のサブセットを含む。また、各サブセットは、実際に収録したデータ（real data）及びシミュレーションにより生成したデータ（simu data）を含む。図7に各データの発話数を示す。図7は、実験に用いた音声のデータを示す図である。

[0072] また、実験では、従来の手法及び実施形態の手法を含む複数の手法でマス

クを推定し、推定したマスクを用いて目的音声を抽出した上で音声認識を行った。従来の手法は、DNNによるマスク推定手法(LSTM)及び空間クラスタリングによるマスク推定手法(cACGMM)である。

[0073] 空間パラメータの初期値は、事前学習を行ったものとした。また、先頭のミニバッチの長さを500msとし、2番目以降のミニバッチの長さを250msとした。その他の各ハイパーパラメータの設定は、図8の通りである。図8は、実験におけるハイパーパラメータを示す図である。

[0074] 図8に示すように、Number of EM iterationsの値は1である。これは、ミニバッチごとに、マスク推定装置10が第2のマスクを推定する際に、空間パラメータ及び第2のマスクの更新を1回のみ行うことを示している。

[0075] 図9に、各手法でマスクを推定し音声認識を行った際のWER(単語誤り率: Word error rate)を示す。図9は、実験結果を示す図である。図9に示すように、実施形態の手法(Proposed)でマスクを推定した場合の単語誤り率が、従来の手法と比べて概ね低かった。これより、実施形態のマスク推定手法は、従来の手法と比べて、音声認識の精度を向上させる効果があるといえる。

[0076] さらに、第2のマスクを推定する際の空間パラメータの設定方法を変化させて同様の実験を行った結果を図10に示す。図10は、実験結果を示す図である。なお、Bus、Caf、Ped、及びStrは、前述の複数の雑音環境を表している。

[0077] 図10のNoPriorは、第1の実施形態で空間パラメータの初期値に単位行列を設定する手法である。また、PostTrainedは、第2の実施形態の手法、すなわち制御部15によって第2のマスクの推定値に、第1のマスクの推定値を代用するかどうかを制御する手法である。なお、PostTrainedでは、制御部15は第1のマスクの累積和を用いて判定を行うものとし、閾値を1.5とした。図10に示すように、NoPrior及びPostTrainedの単語誤り率が、従来の手法と比べて低かった。また、NoPrior及びPostTrainedは、いずれも空間パラメータの事前学習を必要としない手法である。これより、実施形態のマ

ク推定手法は、事前学習を行うことなく、従来の手法よりも音声認識の精度を向上させることができるといえる。

[0078] [システム構成等]

また、図示した各装置の各構成要素は機能概念的なものであり、必ずしも物理的に図示のように構成されていることを要しない。すなわち、各装置の分散及び統合の具体的形態は図示のものに限られず、その全部又は一部を、各種の負荷や使用状況等に応じて、任意の単位で機能的又は物理的に分散又は統合して構成することができる。さらに、各装置にて行われる各処理機能は、その全部又は任意の一部が、CPU (Central Processing Unit) 及び当該CPUにて解析実行されるプログラムにて実現され、あるいは、ワイヤードロジックによるハードウェアとして実現され得る。

[0079] また、本実施形態において説明した各処理のうち、自動的に行われるものとして説明した処理の全部又は一部を手動的に行うこともでき、あるいは、手動的に行われるものとして説明した処理の全部又は一部を公知の方法で自動的に行うこともできる。この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。

[0080] [プログラム]

一実施形態として、マスク推定装置10は、パッケージソフトウェアやオンラインソフトウェアとして上記のマスク推定処理を実行するマスク推定プログラムを所望のコンピュータにインストールさせることによって実装できる。例えば、上記のマスク推定プログラムを情報処理装置に実行させることにより、情報処理装置をマスク推定装置10として機能させることができる。ここで言う情報処理装置には、デスクトップ型又はノート型のパーソナルコンピュータが含まれる。また、その他にも、情報処理装置にはスマートフォン、携帯電話機やPHS (Personal Handyphone System) 等の移動体通信端末、さらには、PDA (Personal Digital Assistant) 等のスレート端末等がその範疇に含まれる。

[0081] また、マスク推定装置 10 は、ユーザが使用する端末装置をクライアントとし、当該クライアントに上記のマスク推定処理に関するサービスを提供する学習サーバ装置として実装することもできる。例えば、マスク推定サーバ装置は、観測信号を入力とし、第 2 のマスクを出力とするマスク推定サービスを提供するサーバ装置として実装される。この場合、マスク推定サーバ装置は、Webサーバとして実装することとしてもよいし、アウトソーシングによって上記のマスク推定処理に関するサービスを提供するクラウドとして実装することとしてもかまわない。

[0082] 図 11 は、マスク推定プログラムを実行するコンピュータの一例を示す図である。コンピュータ 1000 は、例えば、メモリ 1010、CPU 1020 を有する。また、コンピュータ 1000 は、ハードディスクドライブレインタフェース 1030、ディスクドライブレインタフェース 1040、シリアルポートインタフェース 1050、ビデオアダプタ 1060、ネットワークインタフェース 1070 を有する。これらの各部は、バス 1080 によって接続される。

[0083] メモリ 1010 は、ROM (Read Only Memory) 1011 及び RAM (Random Access Memory) 1012 を含む。ROM 1011 は、例えば、BIOS (Basic Input Output System) 等のブートプログラムを記憶する。ハードディスクドライブレインタフェース 1030 は、ハードディスクドライブ 1090 に接続される。ディスクドライブレインタフェース 1040 は、ディスクドライブ 1100 に接続される。例えば磁気ディスクや光ディスク等の着脱可能な記憶媒体が、ディスクドライブ 1100 に挿入される。シリアルポートインタフェース 1050 は、例えばマウス 1110、キーボード 1120 に接続される。ビデオアダプタ 1060 は、例えばディスプレイ 1130 に接続される。

[0084] ハードディスクドライブ 1090 は、例えば、OS 1091、アプリケーションプログラム 1092、プログラムモジュール 1093、プログラムデータ 1094 を記憶する。すなわち、マスク推定装置 10 の各処理を規定す

るプログラムは、コンピュータにより実行可能なコードが記述されたプログラムモジュール１０９３として実装される。プログラムモジュール１０９３は、例えばハードディスクドライブ１０９０に記憶される。例えば、マスク推定装置１０における機能構成と同様の処理を実行するためのプログラムモジュール１０９３が、ハードディスクドライブ１０９０に記憶される。なお、ハードディスクドライブ１０９０は、ＳＳＤにより代替されてもよい。

[0085] また、上述した実施形態の処理で用いられる設定データは、プログラムデータ１０９４として、例えばメモリ１０１０やハードディスクドライブ１０９０に記憶される。そして、ＣＰＵ１０２０は、メモリ１０１０やハードディスクドライブ１０９０に記憶されたプログラムモジュール１０９３やプログラムデータ１０９４を必要に応じてＲＡＭ１０１２に読み出して、上述した実施形態の処理を実行する。

[0086] なお、プログラムモジュール１０９３やプログラムデータ１０９４は、ハードディスクドライブ１０９０に記憶される場合に限らず、例えば着脱可能な記憶媒体に記憶され、ディスクドライブ１１００等を介してＣＰＵ１０２０によって読み出されてもよい。あるいは、プログラムモジュール１０９３及びプログラムデータ１０９４は、ネットワーク（ＬＡＮ（Local Area Network）、ＷＡＮ（Wide Area Network）等）を介して接続された他のコンピュータに記憶されてもよい。そして、プログラムモジュール１０９３及びプログラムデータ１０９４は、他のコンピュータから、ネットワークインタフェース１０７０を介してＣＰＵ１０２０によって読み出されてもよい。

符号の説明

- [0087]
- １０ マスク推定装置
 - １１ 第１特徴量抽出部
 - １２ 第１マスク推定部
 - １３ 第２特徴量抽出部
 - １４ 第２マスク推定部
 - １５ 制御部

- 1 4 1 設定部
- 1 4 2 第 1 更新部
- 1 4 3 第 2 更新部
- 1 4 4 判定部
- 1 4 5 記憶部

請求の範囲

- [請求項1] 連続する時間のセグメントのうち、処理対象とするセグメントを対象セグメントとして、
- 複数の位置で収録した対象セグメントの観測信号から得られる第1の特徴量を基に、前記対象セグメントの観測信号に対する目的信号の占有度である第1のマスクを推定する第1マスク推定部と、
- 前記対象セグメントにおける前記第1のマスクの推定結果と、前記対象セグメントの観測信号から得られる第2の特徴量と、を基に、前記第2の特徴量をモデル化するパラメータと前記観測信号に対する前記目的信号の占有度である第2のマスクとを推定する第2マスク推定部と、
- を有することを特徴とするマスク推定装置。
- [請求項2] 前記第2マスク推定部は、
- 前記対象セグメントまでの前記第1のマスクの累積和と、前記対象セグメントの前記第2の特徴量及び前記第2のマスクと、を基に前記パラメータを更新する第1更新部と、
- 前記対象セグメントの前記第2の特徴量、前記第1のマスク、及び前記パラメータを基に、前記第2のマスクを更新する第2更新部と、
- 所定の収束条件が満たされるまで、前記第1更新部及び前記第2更新部を繰り返し実行させる判定部と、
- を有することを特徴とする請求項1に記載のマスク推定装置。
- [請求項3] 前記第1マスク推定部は、ニューラルネットワークを用いて前記第1のマスクを推定し、
- 第2マスク推定部は、前記パラメータを条件とした場合の前記第2の特徴量の分布モデル及び前記第1のマスクを基に、前記第2のマスクを推定することを特徴とする請求項1又は2に記載のマスク推定装置。
- [請求項4] マスクの推定対象のセグメントである対象セグメントまでの観測信

号に含まれている目的信号の量が、所定の閾値を超えているか否かを判定し、

前記目的信号の量が前記閾値を超えている場合は、前記第2マスク推定部が、推定済みの前記パラメータを基に前記第2のマスクを推定するように制御し、

前記目的信号の量が前記閾値を超えていない場合は、前記第2マスク推定部が、前記第1のマスクの推定結果を前記第2のマスクの推定値として代用するように制御する制御部をさらに有することを特徴とする請求項1から3のいずれか1項に記載のマスク推定装置。

[請求項5]

コンピュータによって実行されるマスク推定方法であって、

連続する時間のセグメントのうち、処理対象とするセグメントを対象セグメントとして、

複数の位置で収録した対象セグメントの観測信号から得られる第1の特徴量を基に、前記対象セグメントの観測信号に対する目的信号の占有度である第1のマスクを推定する第1マスク推定工程と、

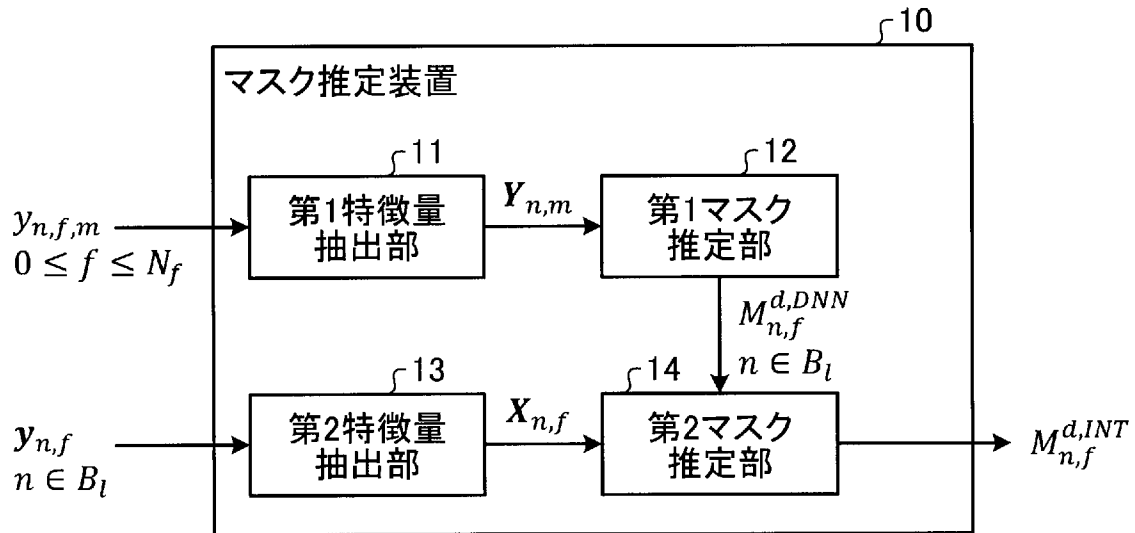
前記対象セグメントにおける前記第1のマスクの推定結果と、前記対象セグメントの観測信号から得られる第2の特徴量と、を基に、前記第2の特徴量をモデル化するパラメータと前記観測信号に対する前記目的信号の占有度である第2のマスクとを推定する第2マスク推定工程と、

を含むことを特徴とするマスク推定方法。

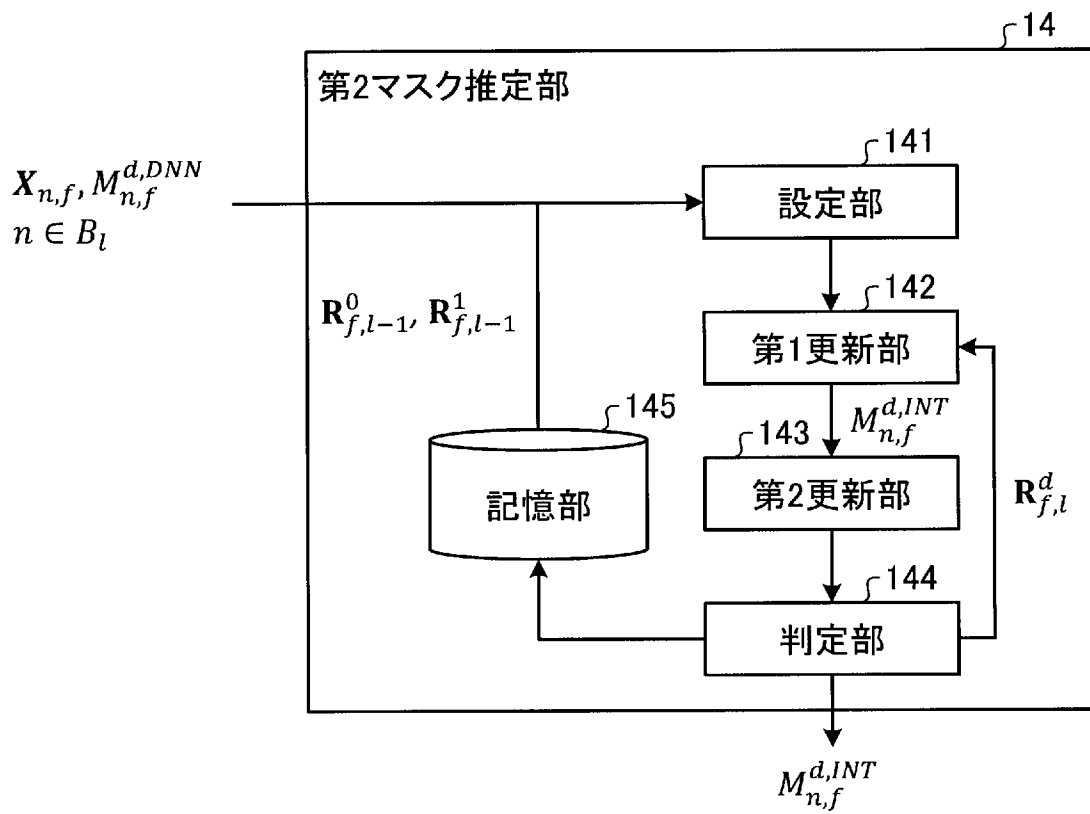
[請求項6]

コンピュータを、請求項1から4のいずれか1項に記載のマスク推定装置として機能させるためのマスク推定プログラム。

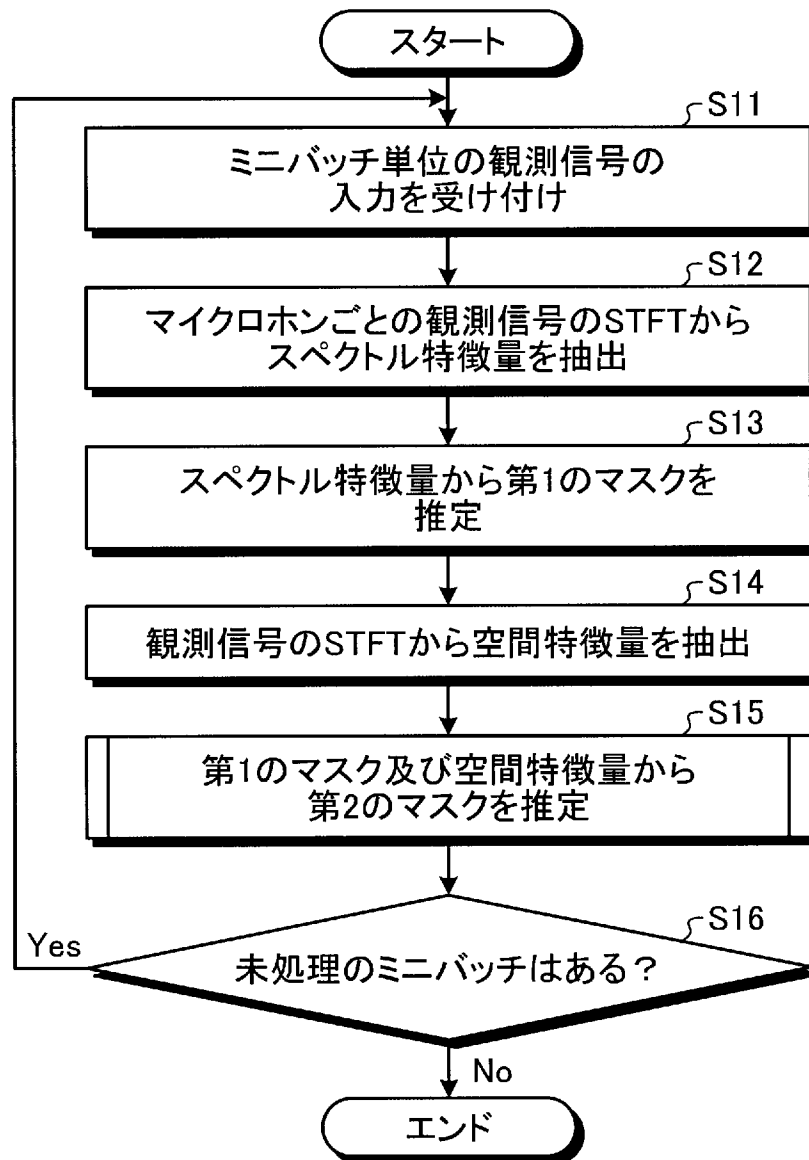
[図1]



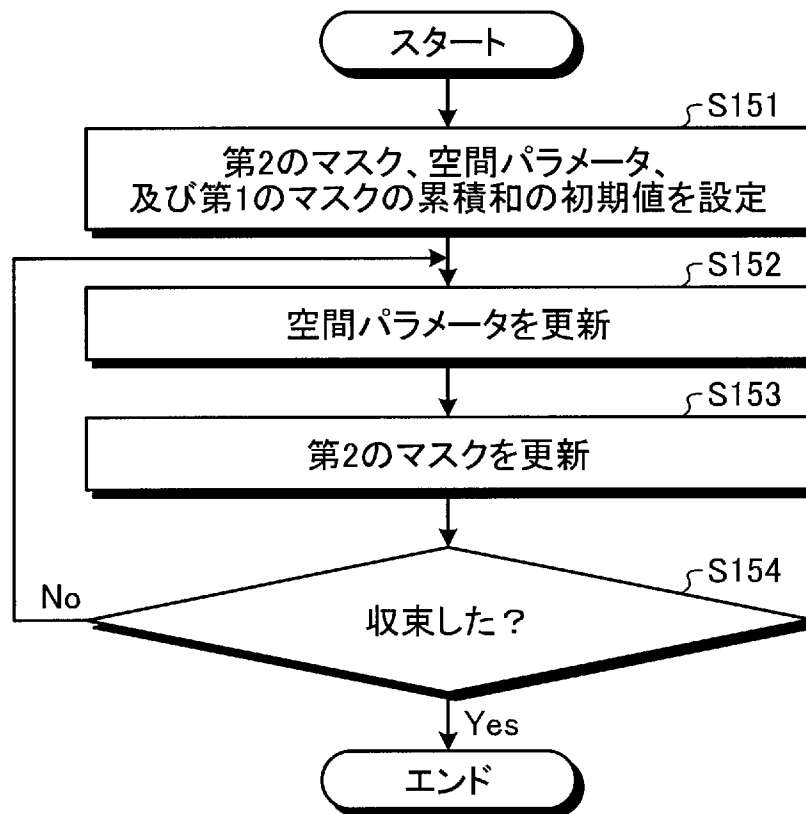
[図2]



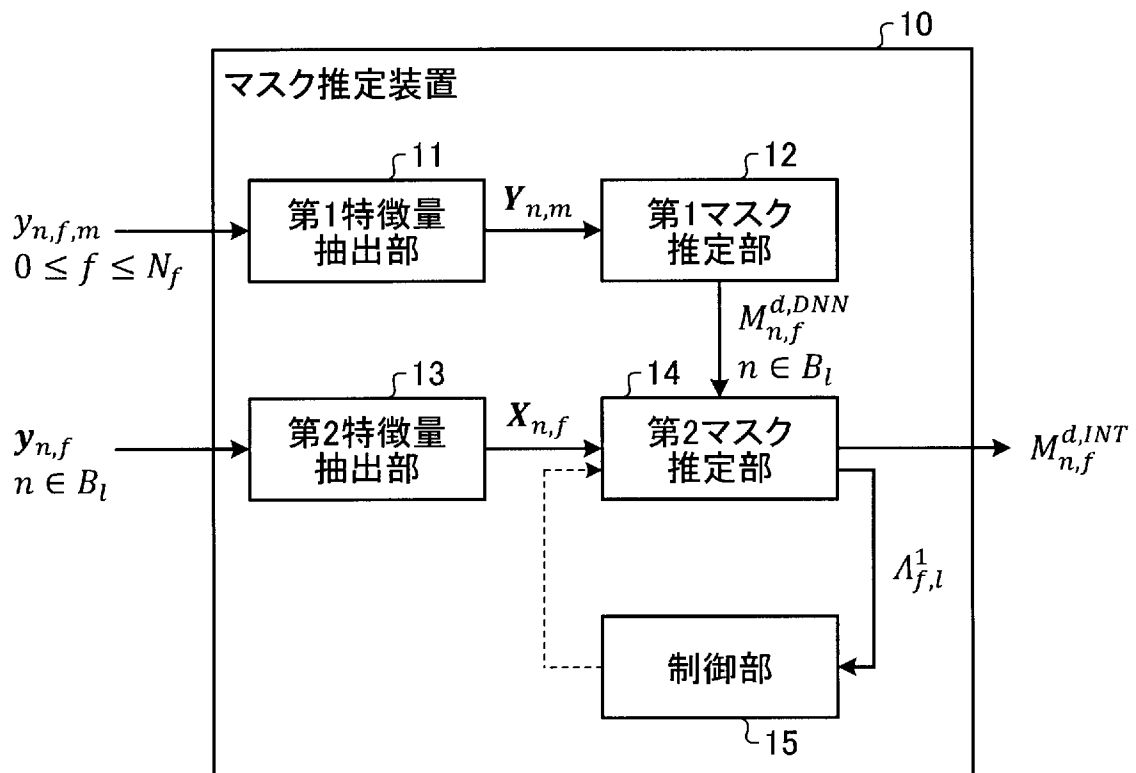
[図3]



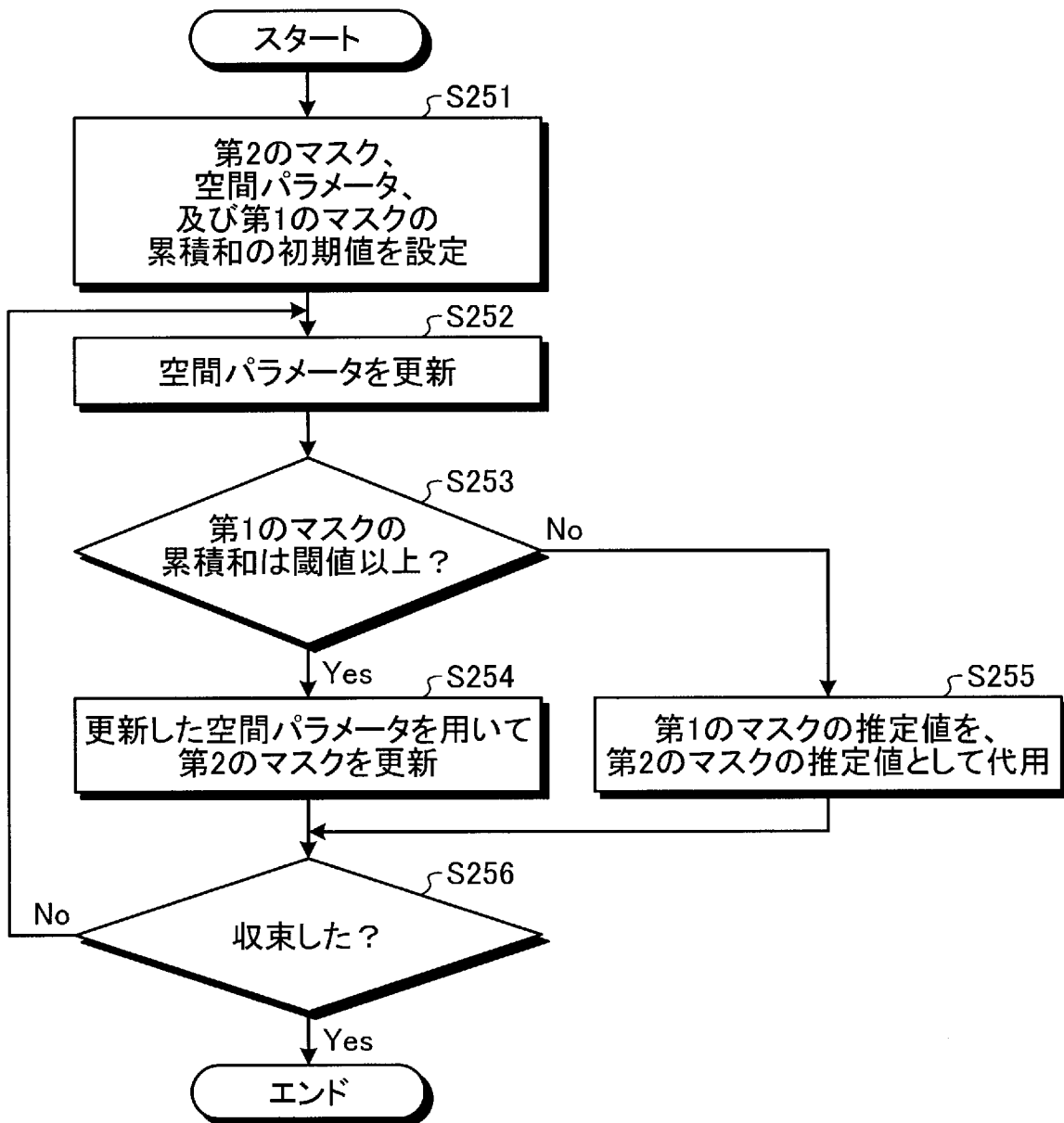
[図4]



[図5]



[図6]



[図7]

	real data	simulated (simu) data
Training set	1600	7138
Development (dev) set	1640	1640
Evaluation (eval) set	1320	1320

[図8]

Sampling frequency	16kHz
Frame length	64ms
Frame overlap	75%
Window function	Hanning
Number of EM iterations	1
Number of microphones	6

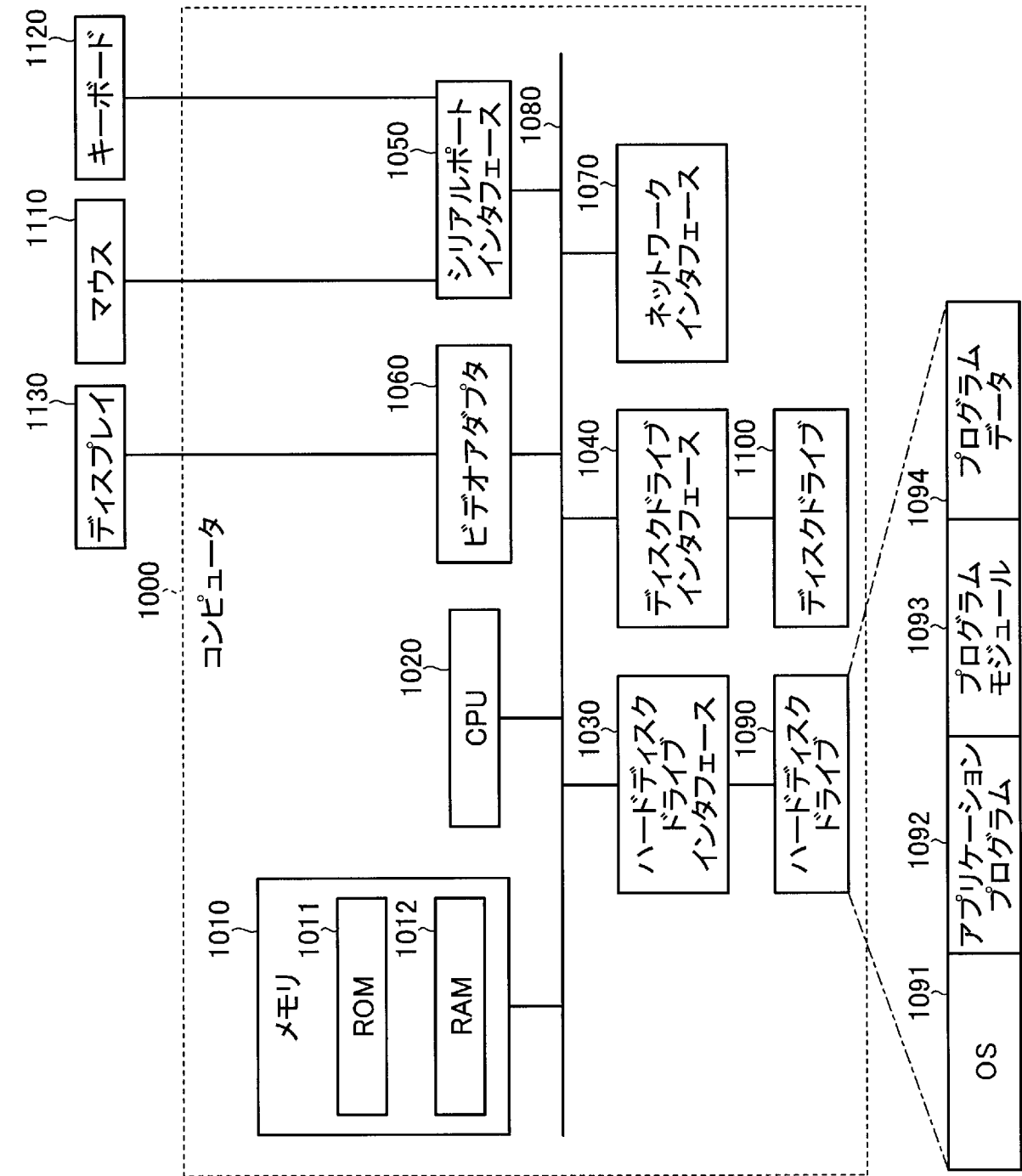
[図9]

	dev			eval		
	Ave	simu	real	Ave	simu	real
No frontend	8.62	8.24	9.01	12.89	10.17	15.60
LSTM	5.33	5.33	5.33	8.14	6.59	9.69
cACGMM	5.60	5.98	5.21	9.20	8.86	9.53
Proposed	4.98	5.33	4.63	7.35	6.81	7.89

[図10]

	Set	Bus	Caf	Ped	Str	Ave
LSTM	dev	6.34	5.32	4.69	4.96	5.33
NoPrior		6.06	4.76	4.59	4.59	5.00
PostTrained		5.96	4.68	4.48	4.62	4.94
LSTM	eval	13.48	9.81	8.00	7.45	9.69
NoPrior		12.15	8.42	6.86	6.65	8.52
PostTrained		11.09	8.48	7.29	6.65	8.38

[図11]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2019/033184

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G10L25/51 (2013.01) i, G10L15/28 (2013.01) i, G10L21/0208 (2013.01) i, G10L25/30 (2013.01) i

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)

Int.Cl. G10L25/51, G10L15/28, G10L21/0208, G10L25/30

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 1971-2019

Registered utility model specifications of Japan 1996-2019

Published registered utility model applications of Japan 1994-2019

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	WO 2017/094862 A1 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 08 June 2017, entire text, all drawings & US 2018/0366135 A1, entire text, all drawings & CN 108292508 A	1-6

☐ 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

“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
04.10.2019

Date of mailing of the international search report
15.10.2019

Name and mailing address of the ISA/
Japan Patent Office
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））

Int.Cl. G10L25/51(2013.01)i, G10L15/28(2013.01)i, G10L21/0208(2013.01)i, G10L25/30(2013.01)i

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））

Int.Cl. G10L25/51, G10L15/28, G10L21/0208, G10L25/30

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	1971-2019年
日本国実用新案登録公報	1996-2019年
日本国登録実用新案公報	1994-2019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	WO 2017/094862 A1（日本電信電話株式会社）2017.06.08, 全文、全図 & US 2018/0366135 A1, 全文、全図 & CN 108292508 A	1-6

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矛盾するものではなく、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

04.10.2019

国際調査報告の発送日

15.10.2019

国際調査機関の名称及びあて先

日本国特許庁（ISA/J P）

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

大野 弘

5Z

9175

電話番号 03-3581-1101 内線 3591