

(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
2022年7月7日(07.07.2022)



(10) 国際公開番号

WO 2022/145015 A1

(51) 国際特許分類:
G10L 21/028 (2013.01) G10L 21/0308 (2013.01)

(21) 国際出願番号: PCT/JP2020/049247

(22) 国際出願日: 2020年12月28日(28.12.2020)

(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: 木下 慶介 (KINOSHITA, Keisuke); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP).
中谷 智広 (NAKATANI, Tomohiro); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT

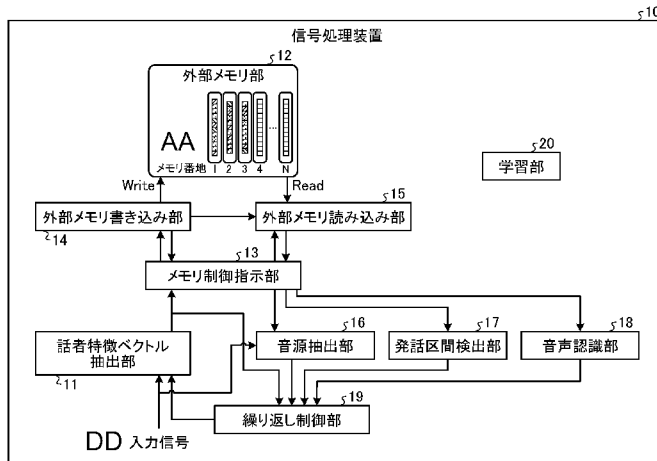
T 知的財産センタ内 Tokyo (JP). デルクロア マーク (DELCROIX, Marc); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP).

(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).

(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, IT, JO, JP, 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,

(54) Title: SIGNAL PROCESSING DEVICE, SIGNAL PROCESSING METHOD, AND SIGNAL PROCESSING PROGRAM

(54) 発明の名称: 信号処理装置、信号処理方法及び信号処理プログラム



10 Signal processing device
11 Speaker feature vector extraction unit
12 External memory unit
13 Memory control instruction unit
14 External memory writing unit
15 External memory reading unit
16 Sound source extraction unit
17 Utterance section detecting unit
18 Speech recognition unit
19 Repetition control unit
20 Learning unit
AA Memory number
DD Input signal

(57) Abstract: A signal processing device (10) has: a speaker feature vector extraction unit (11) for repeatedly extracting, for an inputted acoustic signal, a speaker feature vector for each time block a number of times commensurate with the number of speakers present in the time block; a memory control instruction unit (13) for instructing to write the speaker feature vector to an unused memory block of an external memory unit (12) when a speaker feature vector of a speaker who has not appeared in a time block thus far is extracted, and instructing to write the speaker feature vector to a memory slot of the external memory unit (12) that corresponds to a speaker who has already appeared when a speaker feature value vector of a speaker who has already appeared in a time block thus far is extracted; and a sound source extraction unit (16), an utterance section detecting unit (17), and a speech recognition unit (18) for executing signal processing on the basis of a speaker feature vector.

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, 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: 信号処理装置 (10) は、入力された音響信号に対して、時間ブロックごとに話者特徴ベクトルを、該時間ブロックに存在する話者の数の分だけ繰り返し抽出する話者特徴ベクトル抽出部 (11) と、今までの時間ブロックに出現したことの無い話者の話者特徴量ベクトルが抽出された場合には外部メモリ部 (12) の未使用のメモリスロットに話者特徴ベクトルを書き込むことを指示し、今までの時間ブロックに既に出現したことの無い話者の話者特徴量ベクトルが抽出された場合には外部メモリ部 (12) の既に出現したことの無い話者に対応するメモリスロットに話者特徴ベクトルを書き込むことを指示するメモリ制御指示部 (13) と、話者特徴ベクトルに基づき信号処理を実行する音源抽出部 (16)、発話区間検出部 (17) 及び音声認識部 (18) と、を有する。

明 細 書

発明の名称：

信号処理装置、信号処理方法及び信号処理プログラム

技術分野

[0001] 本発明は、信号処理装置、信号処理方法及び信号処理プログラムに関する。

背景技術

[0002] 複数の音源からの音が混合された音響信号を音源ごとの信号に分離する音源分離の技術では、複数のマイクで收音された音を対象とする第1の音源分離技術と、1つのマイクで收音された音を対象とする第2の音源分離技術がある。第2の音源分離技術は、マイクの位置に関する情報を利用することができないため、第1の音源分離技術よりも難しいとされている。

[0003] ここで、マイクの位置の情報を使わず、入力音響信号の情報に基づいて音源分離を行う第2の音源分離技術として、非特許文献1～3に記載の技術が知られている。

[0004] 非特許文献1に記載の技術は、入力音響信号を、事前に決められた数の音源に分離するものである。入力信号を双方向長短期記憶ニューラルネットワーク (Bi-directional Long Short-Term Memory neural network: BLSTM neural network, 以降BLSTMとする) に与えることで、各音源を抽出するためのマスクを推定することができる。BLSTMのパラメータの学習時には、ある入力信号が与えられた時に、予め与えられた正解分離信号と、推定されたマスクを観測信号に適用して取り出した分離信号との距離を最小化するように、パラメータを更新する。

[0005] 非特許文献2に記載の技術は、入力信号中に含まれる音源数に応じて分離信号数を変え、音源分離を行う。非特許文献1と同様にBLSTMを用いてマスクを推定するが、非特許文献2では、一度に推定する分離マスクは、入力信号中のある一つの音源に関するものだけである。そして、非特許文献2

に記載の技術では、分離マスクを用いて、観測信号を、分離信号と、観測信号から分離信号を取り除いた残留信号とに分離する。その上で、非特許文献2に記載の技術では、残留信号中にまだ別の音源信号が残っているか否かを自動的に判定し、残っている場合は、残留信号をBLSTMに入力し、別の音源を抽出する。一方、残留信号中に別の音源信号が残っていない場合は、その時点で、処理を終了する。

[0006] 非特許文献2に記載の技術では、最終的に残留信号中に音源が含まれなくなるまで、このマスク推定処理を繰り返し、各音源をひとつひとつ抽出することで、音源分離処理と音源数推定処理との両方を同時に達成する。非特許文献2に記載の技術において、判定には、残留信号の音量の閾値処理や、残留信号を別のニューラルネットワークに入力し、別の音源が残っているかどうかを検査させる方法などが提案されている。なお、非特許文献2に記載の技術では、マスク推定のためには、同じBLSTMを繰り返し使用する。

[0007] 非特許文献1～2に記載の技術は、バッチ処理であり、入力信号全体に処理を適用する形式をとっている。そのため、非特許文献1～2に記載の技術は、リアルタイム性に欠ける。例えば、ある会議音声に対してこれらの処理を適用しようとする場合には、少なくともその会議音声の録音が終了するまでは、処理を開始することができない。つまり、会議の開始時から音源分離を適用し、各分離音声を、自動音声認識を用いて逐次的に書き起こすような用途には、非特許文献1～2に記載の技術を適用できない。

[0008] 非特許文献3に記載の技術は、このような問題に鑑み、考案されたものである。非特許文献3に記載の技術では、入力信号を複数の時間ブロック（5～10秒程度の長さのブロック）に分け、ブロックごとに逐次的に処理を適用していく形式を取っている。最初のブロックに対しては、非特許文献2の方法を適用する。ただし、音源分離用マスクを推定すると同時に、抽出した話者に対応する話者特徴ベクトルを計算し、それらも出力する。

[0009] 非特許文献3に記載の技術において、2つ目のブロックでは、1つ目のブロックの処理の結果得られた全ての話者特徴ベクトルのそれぞれを用いて、

それらの話者に対応する音声を1つ目のブロックと同じ順番で繰り返し抽出する。そして、非特許文献3に記載の技術では、3つ目以降のブロックでも、過去のブロックで話者が検出された順番に従って話者を抽出する。

[0010] ここで、非特許文献3に記載の技術では、最初のブロックには出現しなかった新たな話者（新規話者）が2つ目のブロックに出現した場合、一つ目のブロックに出現していた話者すべてを2つ目のブロックの観測信号から抽出しても、その新規話者の音声成分が残留信号に残る。そのため、前記の判定処理を用いれば、新規話者がいることを検出することができ、結果として、新規話者のマスクと、新規話者の話者特徴ベクトルとを推定することも可能である。

[0011] このような処理をブロックごとに逐次的に繰り返していけば、ブロックオンライン処理の形で長時間データに関しても音源分離、音源数推定を行うことができる。また、話者抽出の順番を各ブロック間で共通とすることで、時間ブロック間で話者の追従（ある時間ブロックで得られた分離音と、それと異なる時間ブロックで得られた分離音のどれが同じ話者に属するものか見極める作業）を行うことができる。

[0012] なお、非特許文献3に記載の技術を用いれば、過去のブロックで抽出された話者が、現在のブロックでは発話をしていない場合も、他の話者の場合と同様に、その話者の特徴ベクトルを用いて、その話者に対応する信号の抽出が行われる。しかし、その話者の信号がそのブロックに含まれない場合は、その話者に対応するマスクは0となり、結果として、理想的には音圧0の信号が抽出される。

先行技術文献

非特許文献

[0013] 非特許文献1: Dong Yu, et al., "PERMUTATION INVARIANT TRAINING OF DEEP MODELS FOR SPEAKER-INDEPENDENT MULTI-TALKER SPEECH SEPARATION", in ICASSP, 2017, pp.241-245, [online], [令和2年12月9日検索]、インターネット<URL: <https://arxiv.org/pdf/1607>.

00325.pdf>.

非特許文献2: Keisuke Kinoshita, Lukas Drude, Marc Delcroix, Tomohiro Nakatani, “LISTENING TO EACH SPEAKER ONE BY ONE WITH RECURRENT SELECTIVE HEARING NETWORKS”, ICASSP, 2018, [online], [令和2年12月9日検索]、インターネット<URL: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8462646>>.

非特許文献3: Thilo von Neumann, Keisuke Kinoshita, Marc Delcroix, Shoko Araki, Tomohiro Nakatani, Reinhold Haeb-Umbach, “ALL-NEURAL ONLINE SOURCE SEPARATION, COUNTING, AND DIARIZATION FOR MEETING ANALYSIS”, ICASSP, 2019, [online], [令和2年12月9日検索]、インターネット<URL: <https://arxiv.org/pdf/1902.07881.pdf>>.

発明の概要

発明が解決しようとする課題

[0014] 非特許文献3に記載の技術を用いた場合、新たな時間ブロック（新規時間ブロック）において、ある特定の話者が発話をしているか否かを事前に知ることができないため、過去のブロックにおいて一度でも抽出された話者については、話者特徴ベクトルを用いて新規時間ブロックからの音源抽出を試みなければならない。そして、その結果として、話者が発話しているかどうかを、出力信号の音圧を調べることで判別することができる。

[0015] しかしながら、一般的な会話では、ある時間ブロックの中で発話を行う話者数は、2、3人程度であることが一般的である。つまり、過去に発話した全話者が新規ブロックで発話を行うような状況は極めて稀である。このため、非特許文献3に記載の技術は、発話をしていない話者（サイレント話者）については、本来必要のない抽出処理を行っていることとなる。

[0016] また、一般的にはサイレント話者についても、完全に音圧0の信号が抽出されることは稀であり、ある程度の音圧を持つ信号がその話者の信号として、誤って抽出されることが多い。そのため、サイレント話者が複数いる場合

に、繰り返しの音源抽出処理を行えば、真に発話を行っている話者に関する音源分離・抽出性能の劣化も引き起こすこととなる。

[0017] さらに、非特許文献3に記載の技術を用いた場合、過去に発話を行った順番で、新規時間ブロックの話者抽出を行うこととなる。しかしながら、一般的に、話者抽出の最適な順番は、各ブロックにおいて異なる可能性が高いことが分かっている（非特許文献1参照）。つまり、全てのブロックにおいて、同じ順番で話者抽出を試みることは、処理量が多くなり、処理の最適性を損なうことにつながる。

[0018] 本発明は、上記に鑑みてなされたものであって、処理量を削減しながら、音響信号に対する処理精度を向上することができる信号処理装置、信号処理方法及び信号処理プログラム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

[0019] 上述した課題を解決し、目的を達成するために、本発明に係る信号処理装置は、入力された音響信号に対して、時間ブロックごとに話者特徴ベクトルを、該時間ブロックに存在する話者の数の分だけ繰り返し抽出する抽出部と、抽出部によって抽出された話者特徴ベクトルを格納する外部メモリ部と、今までの時間ブロックに出現したことのない話者の話者特徴量ベクトルが抽出部によって抽出された場合には外部メモリ部の未使用のメモリスロットに話者特徴ベクトルを書き込むことを指示し、今までの時間ブロックに既に出現したことのある話者の話者特徴量ベクトルが抽出部によって抽出された場合には外部メモリ部の既に出現したことのある話者に対応するメモリスロットに話者特徴ベクトルを書き込むことを指示し、外部メモリ部からの話者特徴ベクトルの読み込みを指示する指示部と、指示部から指示を受け、外部メモリ部への話者特徴ベクトルの書き込みを行う書き込み部と、指示部から指示を受け、外部メモリ部から話者特徴ベクトルの読み込みを行う読み込み部と、読み込み部によって読み込まれた話者特徴ベクトルに基づき信号処理を実行する処理部と、を有することを特徴とする。

発明の効果

[0020] 本発明によれば、処理量を削減しながら、音響信号に対する処理精度を向上することができる。

図面の簡単な説明

[0021] [図1]図1は、実施形態に係る信号処理装置の構成の一例を模式的に示す図である。

[図2]図2は、実施形態に係る信号処理装置の構成の他の例を模式的に示す図である。

[図3]図3は、実施形態に係る信号処理装置の構成の他の例を模式的に示す図である。

[図4]図4は、実施形態に係る信号処理装置の構成の他の例を模式的に示す図である。

[図5]図5は、実施形態に係るデータ処理方法の処理手順を示すフローチャートである。

[図6]図6は、プログラムが実行されることにより、信号処理装置が実現され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

[0022] 以下、図面を参照して、本発明の一実施形態を詳細に説明する。なお、この実施形態により本発明が限定されるものではない。また、図面の記載において、同一部分には同一の符号を付して示している。なお、以下では、Aに対し、“ $\wedge$ A”と記載する場合は「“A”の直上に“ $\wedge$ ”が記された記号」と同等であるとする。

[0023] [実施形態]

[信号処理装置の構成]

まず、図1を用いて、実施形態に係る信号処理装置の構成について説明する。図1は、実施形態に係る信号処理装置の構成の一例を示す図である。実施形態に係る信号処理装置10は、例えば、ROM (Read Only Memory)、RAM (Random Access Memory)、CPU (Central Processing Unit) 等を含むコンピュータ等に所定のプログラムが読み込まれて、CPUが所

定のプログラムを実行することで実現される。

[0024] 図1に示すように、信号処理装置10は、話者特徴ベクトル抽出部11（抽出部）、外部メモリ部12、メモリ制御指示部13（指示部）、外部メモリ書き込み部14（書き込み部）、外部メモリ読み込み部15（読み込み部）、音源抽出部16、発話区間検出部17、音声認識部18、繰り返し制御部19及び学習部20を有する。信号処理装置10は、音響信号が入力されると、時間ブロックに分割し、分割した各時間ブロックの音響信号を話者特徴ベクトル抽出部11に入力する。

[0025] 話者特徴ベクトル抽出部11は、入力された音響信号（以降、観測信号とする）に対し、時間ブロックごとに話者特徴ベクトルを、そのブロックに存在する話者の数の分だけ繰り返し抽出する。話者特徴ベクトル抽出処理の繰り返し数は、繰り返し制御部19によって設定される。

[0026] 話者特徴ベクトル抽出部11は、話者特徴ベクトルの抽出方法として、各種方法を用いることができる。例えば、非特許文献3に記載の技術または参考文献1に記載の技術を用いる場合には、最初の時間ブロックにおいては、同ブロックに含まれる任意の数の話者のそれぞれに関して話者特徴ベクトルを抽出できる。

参考文献1：Shota Horiguchi, Yusuke Fujita, Shinji Watanabe, Yawen Xue, Kenji Nagamatsu, “End-to-End Speaker Diarization for an Unknown Number of Speakers with Encoder-Decoder Based Attractors”, arXiv:2005.09921, 2021.

[0027] ここで、一般的な形式として、話者特徴ベクトルの抽出は、式(1)のよう

[0028] [数1]

$$\mathbf{a}_{b,i}, \mathbf{h}_{b,i} = \mathcal{NN}_{\text{embed}} \left[[\mathbf{X}_b, \mathbf{h}_{b,i-1}] \right], i = 1, \dots, I_b \quad \dots (1)$$

[0029] I_b は時間ブロックbにおける話者の総数である。 $\mathbf{a}_{b,i}$ は、時間ブロックbにおいて抽出されたi番目の話者に関する話者特徴ベクトルである。 $\mathbf{X}_b$ は、時間ブロックbの観測信号、 $\mathcal{NN}_{\text{embed}}[\cdot]$ は、BLSTMなどのニューラ

ルネットワークである。 $h_{b,0}$ は、話者特徴ベクトル抽出における最初の繰り返しであることをネットワークに知らせるための初期ベクトル（場合によってはマトリックス）であり、 $NN_{\text{embed}}[\cdot]$ に合わせて適切に設定するものである。

[0030] 外部メモリ部12は、話者特徴ベクトル抽出部11によって抽出された話者特徴ベクトルを格納するためのメモリである。外部メモリ部12は、複数のメモリ番地を有し、1つのメモリ番地に1人の話者の話者特徴ベクトルを格納する。外部メモリ部12は、RAM (Random Access Memory)、フラッシュメモリ、NVS RAM (Non Volatile Static Random Access Memory) 等のデータを書き換え可能な半導体メモリであってもよい。外部メモリ部12は、HDD (Hard Disk Drive)、SSD (Solid State Drive)、光ディスク等の記憶装置であってもよい。

[0031] メモリ制御指示部13は、話者特徴ベクトル抽出部11によって話者特徴ベクトルが抽出されるごとに、その話者特徴ベクトルを受信する。

[0032] メモリ制御指示部13は、受信した話者特徴ベクトルが、新規話者のものである場合、外部メモリ部12の新しいメモリ番地にそのベクトルを書き込むように指示を行う。すなわち、メモリ制御指示部13は、今までの時間ブロックには出現したことのない話者の話者特徴量ベクトルが抽出された場合は、外部メモリ部12の未使用のメモリスロットに、この話者特徴ベクトルを書き込むことを指示する。

[0033] 一方、メモリ制御指示部13は、受信した話者特徴ベクトルが、過去に出現した話者のものである場合、外部メモリ部12のメモリ番地のうち、その過去に出現した話者に対応するメモリ番地に、その話者特徴ベクトルを適切な形で書き込むように指示を行う。すなわち、メモリ制御指示部13は、今までの時間ブロックにすでに出現したことのある話者の話者特徴量ベクトルが抽出された場合、外部メモリ部12のこの話者に対応するメモリスロットに同話者特徴ベクトルを書き込むことを指示する。

[0034] メモリ制御指示部13は、参考文献2～4に記載されたニューラルネット

ワークに基づくものを用いて、外部メモリ部12への話者特徴ベクトルの書き込み、読み込み指示を行う。

参考文献2 : A. Graves, G. Wayne, and I. Danihelka. “Neural Turing Machines”, arxiv:1410.5401., 2014.

参考文献3 : Alex Graves, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwinska, Sergio Gomez; Edward Grefenstette, and Tiago Ramalho, “Hybrid computing using a neural network with dynamic external memory”. Nature. 538 (7626): 471-476. (2016-10-12).

参考文献4 : Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus, “End-To-End Memory Networks”, Advances in Neural Information Processing Systems 28, pp,2440-2448, 2015

[0035] 外部メモリ書き込み部14は、メモリ制御指示部13からの指示を受けて外部メモリ部12に話者特徴ベクトルの書き込みを行う。外部メモリ読み込み部15は、メモリ制御指示部13からの指示を受けて、外部メモリ部12から各話者に関する話者特徴ベクトルを読み込む。このように、信号処理装置10は、ニューラルネットワークを用いて外部メモリ部12との情報のやり取りを実装することで、全てのシステムを、誤差伝搬法により最適化することが可能となる。

[0036] メモリ制御指示部13は、外部メモリ部12への書き込みを行う外部メモリ書き込み部14、外部メモリ部12への読み込みを行う外部メモリ読み込み部15に対する指示のために、参考文献2～4等に記載の手順に倣って、指示ベクトルを生成する。

[0037] 例えば、メモリ制御指示部13は、まず、話者特徴ベクトル抽出部からの出力 $a_{b,i}$ と、メモリ制御指示部13の過去の出力（指示ベクトル） $w_{b,i-1}$ とを基に、ニューラルネットワークを用いて $1 \times M$ のサイズのキーベクトル $k_{b,i}$ を生成する。また、メモリ制御指示部13は、同様に、式（2）の計算に用いる $\beta_{b,i}$ も生成する。その上で、メモリ制御指示部13は、同キーベク

トルと現在の外部メモリ $M_{b,i}$ の各列 n との近さを測り、指示ベクトル $w_{b,i}$ を式 (2) のように計算する。

[0038] [数2]

$$w_{b,i}(n) \leftarrow \frac{\exp(\beta_{b,i} K[\mathbf{k}_{b,i}, \mathbf{M}_{b,i}(n)])}{\sum_j \exp(\beta_{b,i} K[\mathbf{k}_{b,i}, \mathbf{M}_{b,i}(j)])} \quad \dots (2)$$

[0039] 式 (2) において、 $M_{b,i}$ は、 b 番目の時間ブロックにおける i 番目の話者特徴ベクトルを書き込む前の外部メモリ行列 ($N \times M$ のサイズ) である。 N をメモリ番地の総数とし、 M を番地書き込めるベクトルの長さとする。

[0040] キーベクトルとメモリの各列との近さを測る尺度としては、式 (3) に示すコサイン類似度などが考えられる。

[0041] [数3]

$$K[\mathbf{u}, \mathbf{v}] = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|} \quad \dots (3)$$

[0042] $w_{b,i}$ を $1 \times N$ 次元の指示ベクトルとすると、 $w_{b,i}(n)$ は、その各要素であり、式 (4) に示す特徴を有する。

[0043] [数4]

$$\sum_{n=1}^N w_{b,i}(n) = 1, 0 \leq w_{b,i}(n) \leq 1, \forall n \quad \dots (4)$$

[0044] また、指示ベクトル $w_{b,i}$ 中の特定の要素が 1 に近くなるように (ベクトルがスパースになるように)、式 (5) に示す操作を加える。

[0045] [数5]

$$w_{b,i}(n) \leftarrow \frac{(w_{b,i}(n))^c}{\sum_j (w_{b,i}(j))^c} \quad \dots (5)$$

[0046] 式 (5) において、 c は、スパース性を高めるための定数である。

[0047] 外部メモリ書き込み部 14 は、メモリ制御指示部 13 から出力された指示ベクトル $w_{b,i}$ を基に、話者特徴ベクトルの外部メモリ部 12 への書き込み及び更新を行う。

[0048] 外部メモリ書き込み部 14 による書き込み処理は、以下に説明するように、消去処理と書き込み処理とを対として行われる。話者特徴ベクトル抽出部 11 において抽出された話者特徴ベクトル $\mathbf{a}_{b,i}$ は、外部メモリ書き込み部 14 に渡され、適切な形で外部メモリ部 12 によって書き出される。例えば、外部メモリ書き込み部 14 は、 $1 \times N$ のベクトルである消去ベクトル $\mathbf{e}_{b,i}$ と指示ベクトル $\mathbf{w}_{b,i}$ とを基に、式 (6) のようにメモリの消去を行う。なお、この消去ベクトル $\mathbf{e}_{b,i}$ もメモリ制御指示部 13 からの出力である。

[0049] [数6]

$$\tilde{\mathbf{M}}_{b,i}(n) \leftarrow \mathbf{M}_{b,i-1}[\mathbf{1} - \mathbf{w}_{b,i} \mathbf{e}_{b,i}] \quad \dots (6)$$

[0050] $\mathbf{1}$ は、 N 個の 1 からなる $1 \times N$ のベクトルである。外部メモリと $[\cdot]$ 内の成分との掛け算は、メモリスロットにおいて、point-wise の掛け算を行う。

[0051] なお、消去ベクトル $\mathbf{e}_{b,i}$ は、式 (7) のように設定することが多い。

[0052] [数7]

$$\sum_{n=1}^N e_{b,i}(n) = 1, 0 \leq e_{b,i}(n) \leq 1, \forall n \quad \dots (7)$$

[0053] つまり、式 (6)、式 (7) は、消去ベクトルの要素 $e_{b,i}(n)$ と、指示ベクトルの要素 $w_{b,i}(n)$ の両方が 1 の値である時に、そのメモリスロットに関する情報は 0 にリセットされることを示している。そして、上記の消去処理を行った上で、式 (8) に示すように新しい情報ベクトル $\mathbf{a}_{b,i}$ (ここでは、話者特徴ベクトル) が、外部メモリ書き込み部 14 によって外部メモリ部 12 に書き込まれる。

[0054] [数8]

$$\mathbf{M}_{b,i} \leftarrow \tilde{\mathbf{M}}_{b,i} + \mathbf{w}_{b,i} \mathbf{a}_{b,i} \quad \dots (8)$$

[0055] 外部メモリ読み込み部 15 は、メモリ制御指示部 13 からの指示を受けて、外部メモリ部 12 から各話者に関する話者特徴ベクトルを読み込み、メモリ制御指示部 13 に出力する。外部メモリ読み込み部 15 は、メモリ制御指示部 13 から出力された指示ベクトル $\mathbf{w}_{b,i}$ を基に、外部メモリ部 12 から、

更新済みの話者特徴ベクトルの読み込みを行う。

[0056] 外部メモリ読み込み部 15 は、指示ベクトル $w_{b,i}$ を用いることで、式 (9) に示すように、外部メモリ部 12 からデータを読み込むことができる。

[0057] [数9]

$$\mathbf{r}_{b,i} \leftarrow \mathbf{w}_{b,i} \mathbf{M}_{b,i} \quad \dots (9)$$

[0058] 式 (9) に示すように、行列 $\mathbf{M}$ と指示ベクトル $w_{b,i}$ との乗算で、読み込み対象の話者特徴ベクトル $r_{b,i}$ を出力することができる。指示ベクトルの n 番目の要素 $w_{b,i}(n)$ の値が 0 であることは、 n 番目の番地からは情報を読み込まないという指示に対応するため、指示ベクトルの要素 $w_{b,i}(n)$ の値が 1 に近いメモリ番地の情報が主に読み込まれ、0 に近いメモリ番地の情報は相対的に読み込まれない。

[0059] 信号処理装置 10 では、このように、外部メモリ書き込み部 14 による書き込み処理、外部メモリ読み込み部 15 による読み込み処理を、 $i = 1$ から $i = I_b$ まで繰り返すことで、外部メモリ部 12 への書き込み及び読み込みを全ての話者について行うことができる。そして、これら処理が時間ブロックごとに実行されるため、全ての時間ブロックの処理後の外部メモリ部 12 の書き込みが行われた番地の数は、全時間ブロック中に現れた話者数に相当する。したがって、全ての時間ブロックの処理後に外部メモリ部 12 の状態を確認することによって、全時間ブロック中に現れた話者数を確認することができる。

[0060] 音源抽出部 16 は、観測信号と、メモリ制御指示部 13 から出力された話者特徴ベクトルに基づき、この話者特徴ベクトルに対応する話者の音声を観測信号から抽出する（詳細は、非特許文献 3 参照）。

[0061] 音源抽出部 16 は、観測信号 $\mathbf{X}_b$ および外部メモリ部 12 から読みだされた話者特徴ベクトル $r_{b,i}$ を用いて、式 (10) に示すように分離音声 $\hat{S}_{b,i}$ を抽出する。

[0062] [数10]

$$\hat{\mathbf{s}}_{b,i} = \mathcal{NN}_{\text{extract}} \left[[\mathbf{X}_b, \mathbf{r}_{b,i}] \right], i = 1, \dots, I_b \quad \dots (10)$$

[0063] $NN_{extract}[\cdot]$ には、BLSTMや畳み込みニューラルネットワークなどのニューラルネットワークを用いるのが一般的である。

[0064] なお、時間ブロック b において抽出された分離音声 $\hat{S}_{b,i}$ ($i = 1, \dots, l_b$) と、時間ブロック b' ($b \neq b'$) で抽出された分離音声 $\hat{S}_{b',i}$ ($i = 1, \dots, l_{b'}$) とが、同じ話者のものであるか否かは、外部メモリ部 12 に対する指示ベクトル $w_{b,i}$ ($i = 1, \dots, l_b$) と、指示ベクトル $w_{b',i}$ ($i = 1, \dots, l_{b'}$) の各 i において最大値が検出されるインデックスが同じか否かを調べることで判定できる。言い換えれば、メモリ制御指示部 13 は、話者照合の役割を担っている。このため、話者特徴ベクトル抽出部 11 において、各時間ブロック間で異なる順序で話者特徴ベクトルが抽出されても、前述の外部メモリ部 12 への指示ベクトル内の最大値検出と比較処理とを用いることで、時間ブロック間で話者を追従することができる。

[0065] 発話区間検出部 17 は、観測信号と、メモリ制御指示部 13 から出力された話者特徴ベクトルに基づき、この話者特徴ベクトルに対応する話者の発話区間検出結果を出力する（詳細は、参考文献 1 参照）。

[0066] 音声認識部 18 は、観測信号と、メモリ制御指示部 13 から出力された話者特徴ベクトルに基づき、この話者特徴ベクトルに対応する話者の音声認識結果を出力する（詳細は、参考文献 5 参照）。

参考文献 5 : Marc Delcroix, Shinji Watanabe, Tsubasa Ochiai, Keisuke Kinoshita, Shigeki Karita, Atsunori Ogawa, Tomohiro Nakatani, “End-to-end SpeakerBeam for single channel target speech recognition”, Interspeech2019, pp.451-455, 2019.

[0067] なお、音源抽出部 16、発話区間検出部 17 及び音声認識部 18 は、音声処理する処理部の一例である。また、実施形態に係る信号処理装置として、音源抽出部 16、発話区間検出部 17 及び音声認識部 18 の複数の処理部を有する信号処理装置 10 について説明したが、これに限らない。例えば、実施形態に係る信号処理装置は、音源抽出部 16 を有する信号処理装置 10

A（図2参照）、発話区間検出部17を有する信号処理装置10B（図3参照）、または、音声認識部18を有する信号処理装置10C（図4参照）であってもよい。

[0068] 繰り返し制御部19は、話者特徴ベクトル抽出部11の抽出処理の状態、または、音源抽出部16と発話区間検出部17と音声認識部18とによる処理結果を基に、話者特徴ベクトル抽出部11による抽出処理の繰り返し数を決定する。言い換えると、繰り返し制御部19は、例えば、話者特徴ベクトル抽出部11の話者特徴ベクトルの抽出処理の状態を基に、話者特徴ベクトル抽出部11による抽出処理の繰り返し数を決定する。または、繰り返し制御部19は、発話区間検出部17及び音声認識部18からの出力結果を用いて、話者特徴ベクトル抽出部11の話者特徴ベクトルの抽出処理の繰り返し数を決定する。なお、理想的には、話者特徴ベクトル抽出部11は、各ブロック*b*において*l_b*個の話者特徴ベクトルを抽出することが望ましい。

[0069] 例えば、音源抽出部16のみを有する信号処理装置10A（図2参照）の場合、繰り返し制御部19は、話者特徴ベクトル抽出部11から出力される内部状態ベクトル $\mathbf{h}_{b,i}$ 、観測信号 $\mathbf{x}_b$ 、分離音声 $\hat{\mathbf{s}}_{b,i}$ を用いて、以下の式（11）のような繰り返しを停止すべきか否かを示すスカラー値 $\hat{f}_{b,i}$ （ $0 \leq \hat{f}_{b,i} \leq 1$ ）を算出する。スカラー値 $\hat{f}_{b,i}$ が所定の値よりも大きければ繰り返しを止め、所定の値よりも低ければ繰り返しを継続する。

[0070] [数11]

$$\hat{f}_{b,i} = \mathcal{NN}_{\text{stop-flag}} \left[\mathbf{h}_{b,i}, \mathbf{x}_b, \sum_{i'=1}^{i-1} \hat{\mathbf{s}}_{b,i'} \right], i = 1, \dots, l_b \quad \dots (11)$$

[0071] 繰り返し制御部19は、話者特徴ベクトル抽出部11が話者特徴ベクトルを抽出する度に、異なる補助情報を話者特徴ベクトル抽出部11のニューラルネットワークに入力することで、話者特徴ベクトル抽出部11に、異なる音源に対応する話者特徴ベクトルの抽出結果を出力させる。

[0072] 例えば、繰り返し制御部19は、入力される音響信号に含まれる全ての話者特徴ベクトルを仮想的に認識している場合を例に説明する。繰り返し制御

部 19 は、話者特徴ベクトル抽出部 11 が話者特徴ベクトルを抽出する度に、この話者特徴ベクトルがいずれの話者に対応するかを認識する。そして、繰り返し制御部 19 は、未抽出の話者がある場合には、次の話者の話者特徴ベクトルに関する補助情報を話者特徴ベクトル抽出部 11 に入力して、話者特徴ベクトル抽出部 11 に次の話者の話者特徴ベクトルを抽出される。繰り返し制御部 19 は、この時間ブロックの観測信号の全話者に対する話者特徴ベクトルの抽出処理が終われば、話者特徴ベクトル抽出部 11 による抽出処理の繰り返しを停止する。

[0073] 学習部 20 は、学習用データを用いて、信号処理装置 10 が使用するパラメータを最適化する。学習部 20 は、信号処理装置 10 を構成するニューラルネットワークのパラメータを所定の目的関数に基づき最適化する。

[0074] 学習データは、入力信号（観測信号）と、この入力信号に含まれる各音源に対応する正解クリーン信号、正解発話時間情報、正解発話内容と、入力信号に含まれる人数の総数情報（総話者数） I とから成る。学習部 20 は、学習データを基に、信号処理装置 10 の出力と正解情報との誤差が小さくなるようにパラメータを学習する。

[0075] 具体的には、図 2 の信号処理装置 10 A のパラメータを最適化するために学習処理を行う場合、入力信号 X_b ($b = 1, \dots, B$) を与えた際に、各時間ブロックにおいてシステムの出力結果（分離音声） $\hat{S}_{b,i}$ が正解の分離音声 $S_{b,i}$ に近くなるように、式（12）の事情誤差基準の損失関数を設ける。

[0076] [数12]

$$L_{reconst} = \sum_{b,i} |s_{b,i} - \hat{s}_{b,i}|^2 \quad \dots (12)$$

[0077] また、各時間ブロック b において、正しい音源数を推定するために繰り返し回数を決定するためのスカラー値について、式（13）のクロスエントロピー損失関数を設ける。

[0078]

[数13]

$$L_{local-count} = - \sum_{b,i} f_{b,i} \log \hat{f}_{b,i} \quad \dots (13)$$

[0079] なお、 $f_{b,i}$ は、 $i = I_b$ の場合のみ1を取り、それ以外は0を取る値である。

[0080] そして、学習部20は、外部メモリ内で使われるアドレスの総数が総話者数Iと同じ値となるように、式(14)も同時に小さくなるようにパラメータを更新する。

[0081] [数14]

$$L_{global-count} = \left| \sum \left(\min \left(\sum_{b=1}^B \sum_{i=1}^{I_b} \mathbf{w}_{b,i}, T_{count} \right) \right) - I \right|^2 \quad \dots (14)$$

[0082] 式(14)において、 T_{count} は事前設定した閾値であり、一般的には1を設定する。 $\min(\cdot)$ は、入力の値が T_{count} よりも大きい場合は T_{count} を出力し、 T_{count} よりも小さい場合、そのままの値を出力する関数である。式(14)は、外部メモリの中で使用されるメモリ番地の数が総話者数Iと合致するように促す損失関数である。

[0083] 学習部20は、最終的には、全ての損失関数の値を合算した式(15)の値が小さくなるようにニューラルネットワークのパラメータを学習する。

[0084] [数15]

$$L = \alpha L_{reconst} + \beta L_{local-count} + \gamma L_{global-count} \quad \dots (15)$$

[0085] なお、学習部20は、発話区間検出部17を有する信号処理装置10B、音声認識部18を有する信号処理装置10C、信号処理装置10に対しては、具備する処理部に応じた損失関数を用いて、ニューラルネットワークのパラメータを最適化すればよい。発話区間検出部17を有する信号処理装置10Bについては参考文献6を、音声認識部18を有する信号処理装置10Cについては参考文献7を参照されたい。

参考文献6：Yusuke Fujita, Naoyuki Kanda, Shota Horiguchi, Kenj

i Nagamatsu, Shinji Watanabe, “End-to-End Neural Speaker Diarization with Permutation-free Objectives”, Proc. Interspeech, pp. 4300-4304, 2019.

参考文献 7 : Shigeki Karita et al., “A comparative study on transformer vs RNN in speech applications”, IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2019.

[0086] [信号処理の処理手順]

次に、本実施形態に係る信号処理方法の処理手順について説明する。図 5 は、実施形態に係る信号処理方法の処理手順を示すフローチャートである。

[0087] 図 5 に示すように、信号処理装置 10 は、音響信号の入力を受け付けると（ステップ S 1）、音響信号を時間ブロックに分割する（ステップ S 2）。そして、信号処理装置 10 は、時間ブロックの数 b を $b = 1$ に初期化する（ステップ S 3）。

[0088] 時間ブロック b の音響信号（観測信号）が話者特徴ベクトル抽出部 11 に入力されると、話者特徴ベクトル抽出部 11 は、時間ブロック b の観測信号から、この時間ブロック b に存在するある一人の話者の話者特徴ベクトルを推定し、抽出する（ステップ S 4）。

[0089] そして、メモリ制御指示部 13 は、話者特徴ベクトル抽出部 11 が抽出した話者特徴ベクトルが、今までの時間ブロックには出現したことの無い話者の話者特徴量ベクトルであるか否かを判定する（ステップ S 5）。

[0090] 今までの時間ブロックには出現したことの無い話者の話者特徴量ベクトルである場合（ステップ S 5 : Yes）について説明する。この場合、メモリ制御指示部 13 は、外部メモリ部 12 の未使用のメモリスロットに、この話者特徴ベクトルを書き込むことを指示し、外部メモリ書き込み部 14 は、外部メモリ部 12 の未使用のメモリスロットに話者特徴ベクトルを書き込む（ステップ S 6）。

[0091] 今までの時間ブロックに出現したことの無い話者の話者特徴量ベクトルである場合（ステップ S 5 : No）について説明する。この場合、メモリ制御

指示部 13 は、外部メモリ部 12 のこの話者に対応するメモリスロットに、話者特徴ベクトルを書き込むことを指示し、外部メモリ書き込み部 14 は、外部メモリ部 12 のこの話者に対応するメモリスロットに話者特徴ベクトルを書き込む（ステップ S 7）。

[0092] そして、外部メモリ書き込み部 14 は、メモリ制御指示部 13 による指示にしたがって、外部メモリ部 12 から、話者特徴ベクトル抽出部 11 が抽出した一人の話者に対応する話者特徴ベクトルを読み込み（ステップ S 8）、メモリ制御指示部 13 に出力する。

[0093] そして、音源抽出部 16 は、観測信号と、メモリ制御指示部 13 から出力された一人の話者に対応する話者特徴ベクトルに基づき、この話者特徴ベクトルに対応する話者の音声を観測信号から抽出する（ステップ S 9）。

[0094] また、発話区間検出部 17 は、観測信号と、メモリ制御指示部 13 から出力された一人の話者に対応する話者特徴ベクトルに基づき、この話者特徴ベクトルに対応する話者の発話区間検出結果を出力する（ステップ S 10）。

[0095] そして、音声認識部 18 は、観測信号と、メモリ制御指示部 13 から出力された一人の話者に対応する話者特徴ベクトルに基づき、この話者特徴ベクトルに対応する話者の音声認識結果を出力する（ステップ S 11）。

[0096] ステップ S 9～ステップ S 11 は、図 5 に示すように、並行に処理されるほか、直列に処理されてもよい。直列に処理する場合、処理の順は、特に限定しない。また、信号処理装置 10A～10C の場合には、具備する音声処理機能部に応じた処理を実行すればよい。例えば、信号処理装置 10A の場合には、音源抽出部 16 による音源抽出処理（ステップ S 9）を実行し、ステップ S 12 に進む。

[0097] そして、繰り返し制御部 19 は、音源抽出部 16、発話区間検出部 17 及び音声認識部 18 の処理結果を基に、繰り返しを停止するか否かを判定する（ステップ S 12）。なお、繰り返し制御部 19 は、話者特徴ベクトル抽出部 11 の抽出処理の状態を基に、繰り返しを停止するか否かを判定してもよい。

[0098] 繰り返し制御部 19 は、繰り返しを停止しないと判定した場合（ステップ S 12 : N o）、この時間ブロック b における次の話者に関する処理を進めるため、ステップ S 4 に戻り、次の話者に対する話者特徴ベクトルの抽出を行う。

[0099] また、繰り返し制御部 19 は、繰り返しを停止すると判定した場合（ステップ S 12 : Y e s）、この時間ブロック b に対する処理結果を出力する（ステップ S 13）。なお、信号処理装置 10 の例であれば、出力結果は、音源抽出結果、発話区間検出結果、音声認識結果である。また、信号処理装置 10 は、全ての時間ブロックの処理結果をまとめて出力してもよい。

[0100] そして、信号処理装置 10 は、全ての時間ブロックの処理を終了したか否かを判定する（ステップ S 14）。信号処理装置 10 は、全ての時間ブロックの処理を終了した場合（ステップ S 14 : Y e s）、入力された音響信号に対する処理を終了する。また、信号処理装置 10 は、全ての時間ブロックの処理を終了していない場合（ステップ S 14 : N o）、次の時間ブロックに対する処理を行うため、時間ブロック b に 1 を加算し（ステップ S 15）、ステップ S 4 に戻り、処理を続ける。

[0101] [実施形態の効果]

実施形態に係る信号処理装置 10 は、時間ブロックごとに話者特徴ベクトルを繰り返し抽出し、外部メモリ部 12 に書き込みを行う。この際、信号処理装置 10 は、今までの時間ブロックに出現したことの無い話者の話者特徴量ベクトルが抽出された場合には外部メモリ部 12 の未使用のメモリスロットに話者特徴ベクトルを書き込む。信号処理装置 10 は、今までの時間ブロックに既に出現したことの無い話者の話者特徴量ベクトルが抽出された場合には外部メモリ部 12 のこの話者に対応するメモリスロットに話者特徴ベクトルを書き込む。

[0102] したがって、信号処理装置 10 は、発話をしていない話者（サイレント話者）については、抽出処理自体を実行しない。このため、信号処理装置 10 は、従来のように、時間ブロックごとにサイレント話者について音源抽出を

試みる必要がないため、従来と比して処理量を削減できるとともに、真に発話を行っている話者のみに対する処理を適切に行えるため、処理精度を向上することができる。

[0103] また、信号処理装置 10 では、時間ブロックごとに話者特徴ベクトルの抽出を行うため、従来のように全ての時間ブロックにおいて同じ順で話者特徴ベクトルを抽出する必要がないため、処理の最適性を損なうこともない。

[0104] このように、信号処理装置 10 は、処理量を削減しながら、音響信号に対する処理精度を向上させることができる。

[0105] [プログラム]

図 6 は、プログラムが実行されることにより、信号処理装置 10 が実現され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インタフェース 1030、ディスクドライブインタフェース 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。

[0106] メモリ 1010 は、ROM (Read Only Memory) 1011 及び RAM 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。

[0107] ハードディスクドライブ 1090 は、例えば、OS (Operating System) 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、信号処理装置 1

0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、信号処理装置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 (Solid State Drive) により代替されてもよい。

[0108] 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が、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に読み出して実行する。

[0109] 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。

[0110] 以上、本発明者によってなされた発明を適用した実施形態について説明したが、本実施形態による本発明の開示の一部をなす記述及び図面により本発明は限定されることはない。すなわち、本実施形態に基づいて当業者等によりなされる他の実施形態、実施例及び運用技術等は全て本発明の範疇に含まれる。

符号の説明

- [0111] 10, 10A, 10B, 10C 信号処理装置
- 11 話者特徴ベクトル抽出部
 - 12 外部メモリ部
 - 13 メモリ制御指示部
 - 14 外部メモリ書き込み部
 - 15 外部メモリ読み込み部
 - 16 音源抽出部
 - 17 発話区間検出部
 - 18 音声認識部
 - 19 繰り返し制御部
 - 20 学習部

請求の範囲

- [請求項1] 入力された音響信号に対して、時間ブロックごとに話者特徴ベクトルを、該時間ブロックに存在する話者の数の分だけ繰り返し抽出する抽出部と、
- 前記抽出部によって抽出された話者特徴ベクトルを格納する外部メモリ部と、
- 今までの時間ブロックに出現したことの無い話者の話者特徴量ベクトルが前記抽出部によって抽出された場合には前記外部メモリ部の未使用のメモリスロットに前記話者特徴ベクトルを書き込むことを指示し、今までの時間ブロックに既に出てきたことのある話者の話者特徴量ベクトルが前記抽出部によって抽出された場合には前記外部メモリ部の前記既に出てきたことのある話者に対応するメモリスロットに前記話者特徴ベクトルを書き込むことを指示し、前記外部メモリ部からの話者特徴ベクトルの読み込みを指示する指示部と、
- 前記指示部から指示を受け、前記外部メモリ部への話者特徴ベクトルの書き込みを行う書き込み部と、
- 前記指示部から指示を受け、前記外部メモリ部から話者特徴ベクトルの読み込みを行う読み込み部と、
- 前記読み込み部によって読み込まれた話者特徴ベクトルに基づき信号処理を実行する処理部と、
- を有することを特徴とする信号処理装置。
- [請求項2] 前記抽出部の抽出処理の状態または前記処理部による処理結果を基に、前記抽出部による抽出処理の繰り返し数を決定することを特徴とする請求項1に記載の信号処理装置。
- [請求項3] 前記抽出部、前記指示部及び前記処理部が用いるパラメータを所定の目的関数に基づき最適化する学習部をさらに有することを特徴とする請求項1または2に記載の信号処理装置。
- [請求項4] 前記処理部は、前記話者特徴ベクトルに基づき、前記話者特徴ベク

トルに対応する話者の音声を前記音響信号から抽出する音源抽出部であることを特徴とする請求項1～3のいずれか一つに記載の信号処理装置。

[請求項5] 前記処理部は、前記話者特徴ベクトルに基づき、前記話者特徴ベクトルに対応する話者の音声認識結果を出力する音源抽出部であることを特徴とする請求項1～4のいずれか一つに記載の信号処理装置。

[請求項6] 前記処理部は、前記話者特徴ベクトルに基づき、前記話者特徴ベクトルに対応する話者の発話区間検出結果を出力する発話区間検出部であることを特徴とする請求項1～5のいずれか一つに記載の信号処理装置。

[請求項7] 信号処理装置が実行する信号処理方法であって、
前記信号処理装置は、データを格納する外部メモリを有し、
入力された音響信号に対して、時間ブロックごとに話者特徴ベクトルを、該時間ブロックに存在する話者の数の分だけ繰り返し抽出する抽出工程と、

前記抽出工程において抽出された話者特徴ベクトルを格納する外部メモリと、

今までの時間ブロックに出現したことの無い話者の話者特徴量ベクトルが前記抽出工程において抽出された場合には、前記外部メモリの未使用のメモリスロットに前記話者特徴ベクトルを書き込むことを指示し、今までの時間ブロックに既に出現したことの無い話者の話者特徴量ベクトルが前記抽出工程において抽出された場合には、前記外部メモリの前記既に出現したことの無い話者に対応するメモリスロットに前記話者特徴ベクトルを書き込むことを指示する指示工程と、

前記指示工程における指示を受け、前記外部メモリへの話者特徴ベクトルの書き込みを行う書き込み工程と、

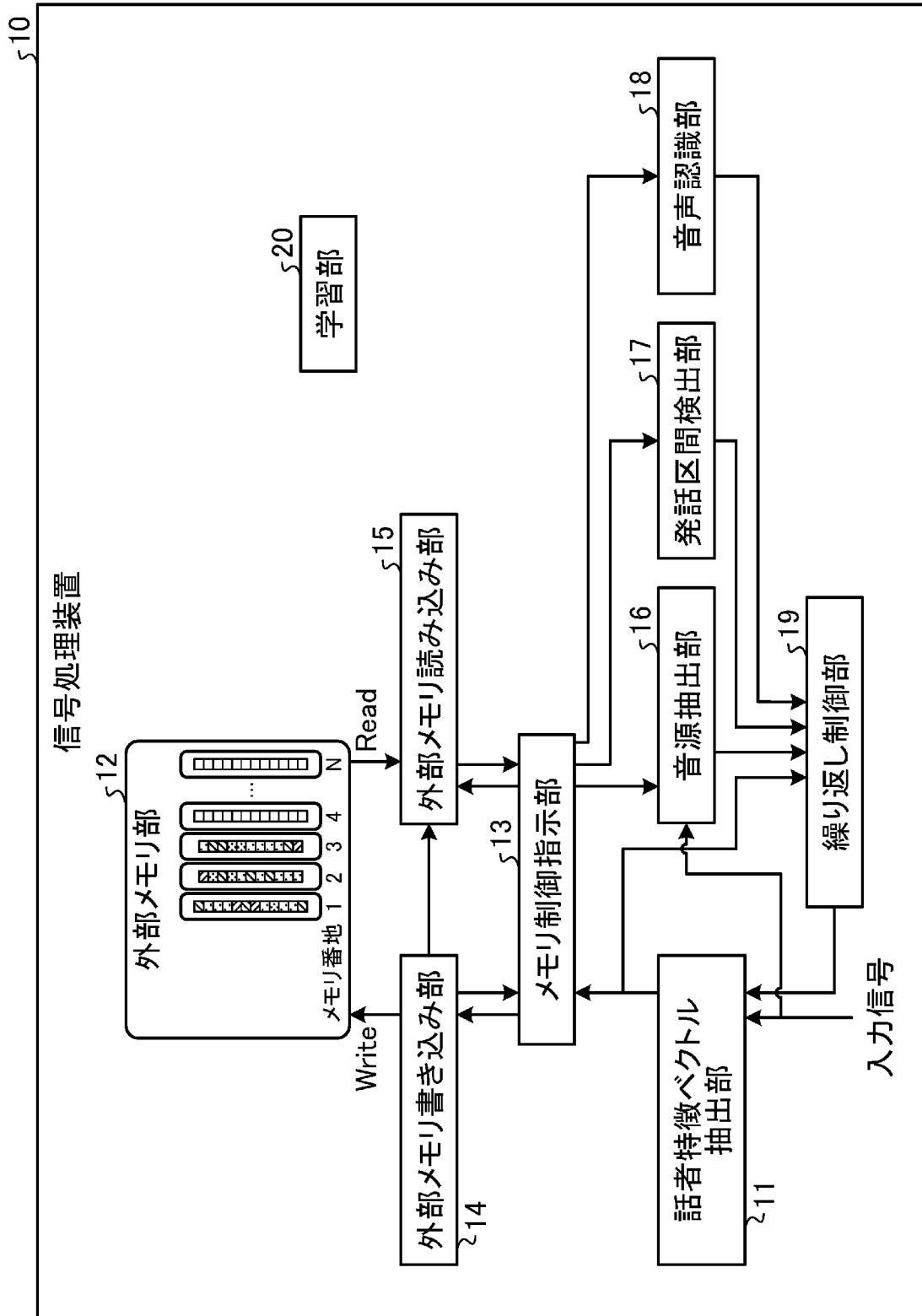
前記指示工程における指示を受け、前記外部メモリから話者特徴ベクトルの読み込みを行う読み込み工程と、

前記読み込み工程において読み込まれた話者特徴ベクトルに基づき
信号処理を実行する処理工程と、

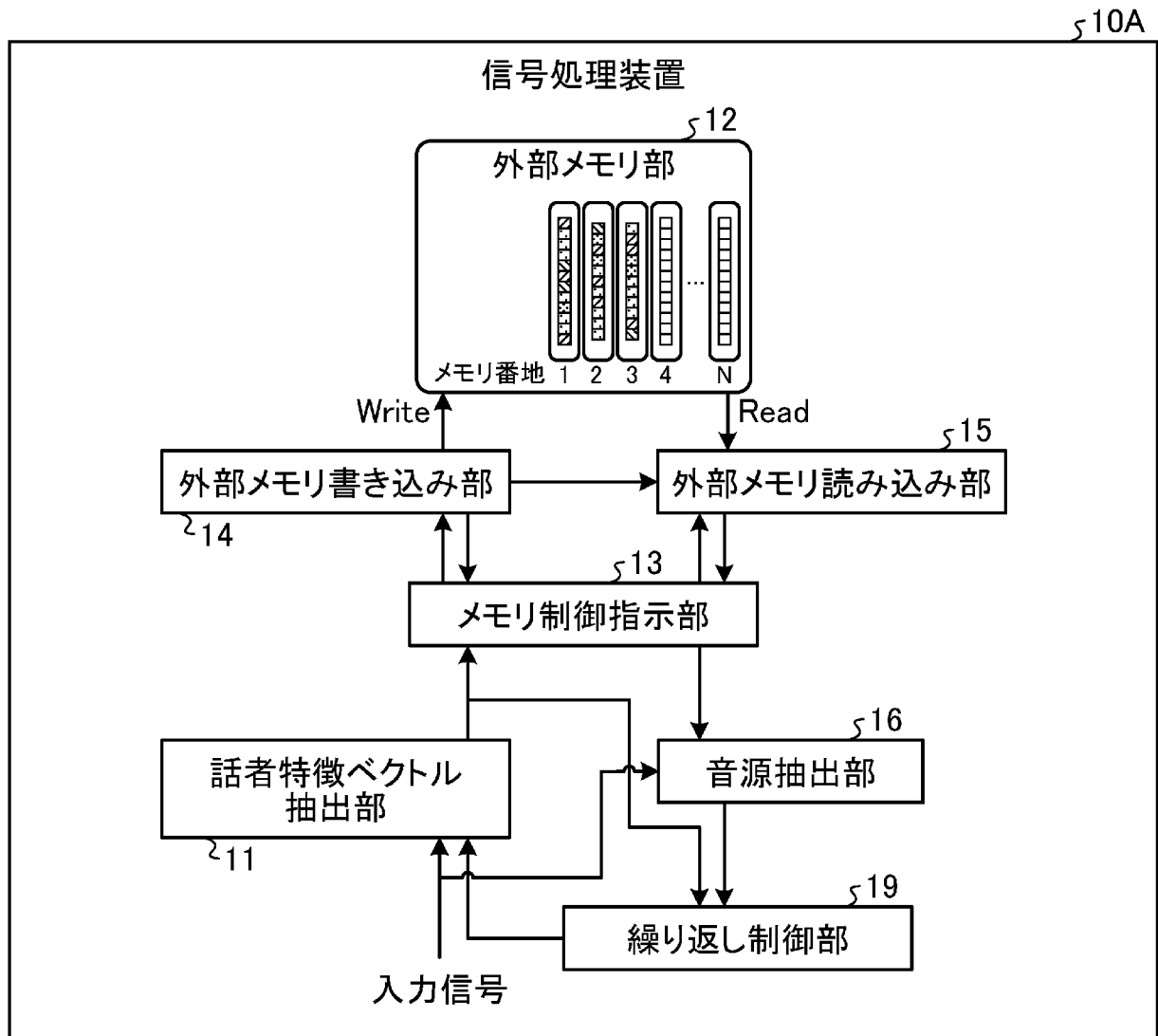
を含んだことを特徴とする信号処理方法。

[請求項8] コンピュータを、請求項1～6のいずれか一つに記載の信号処理装置として機能させるための信号処理プログラム。

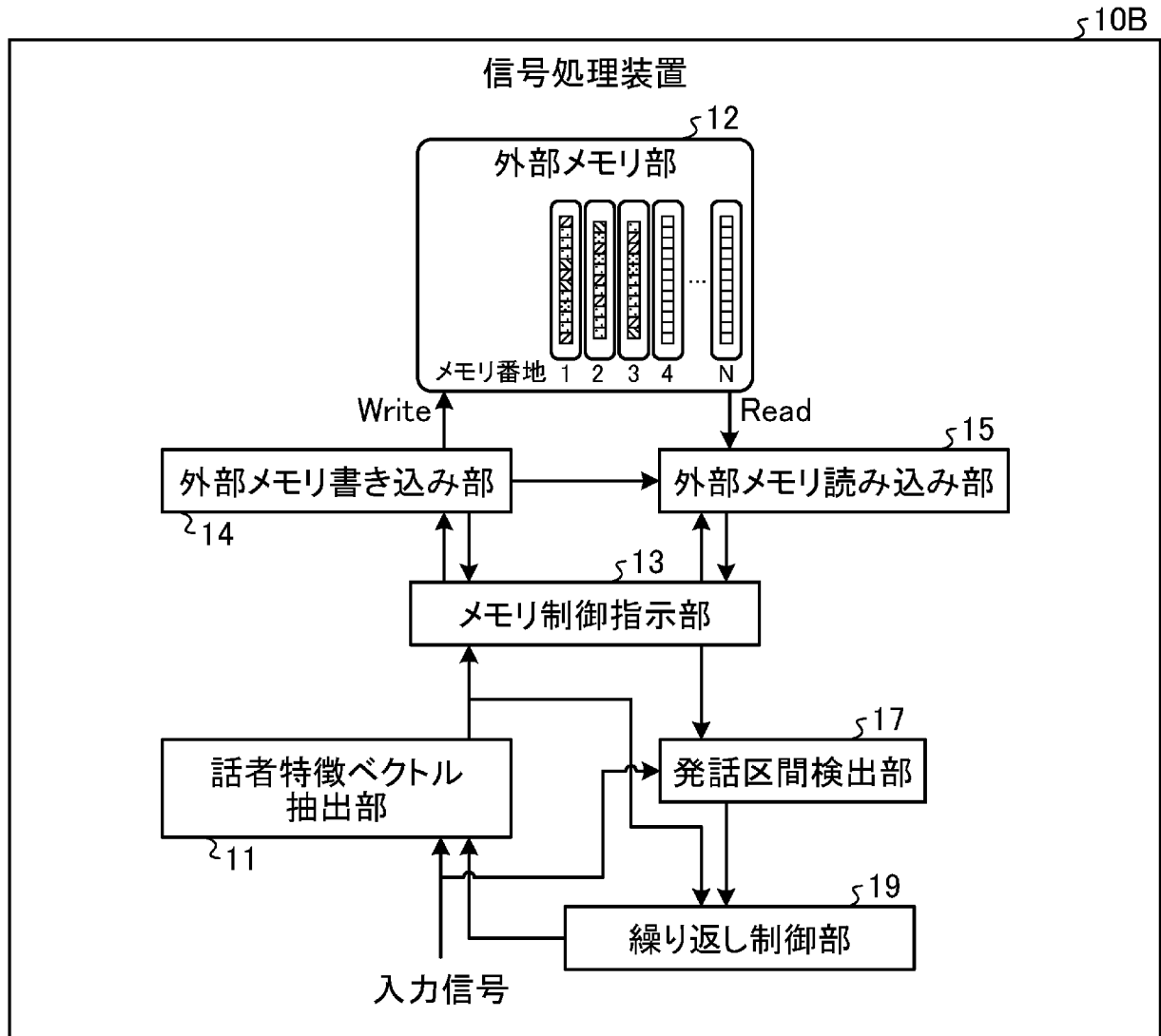
[図1]



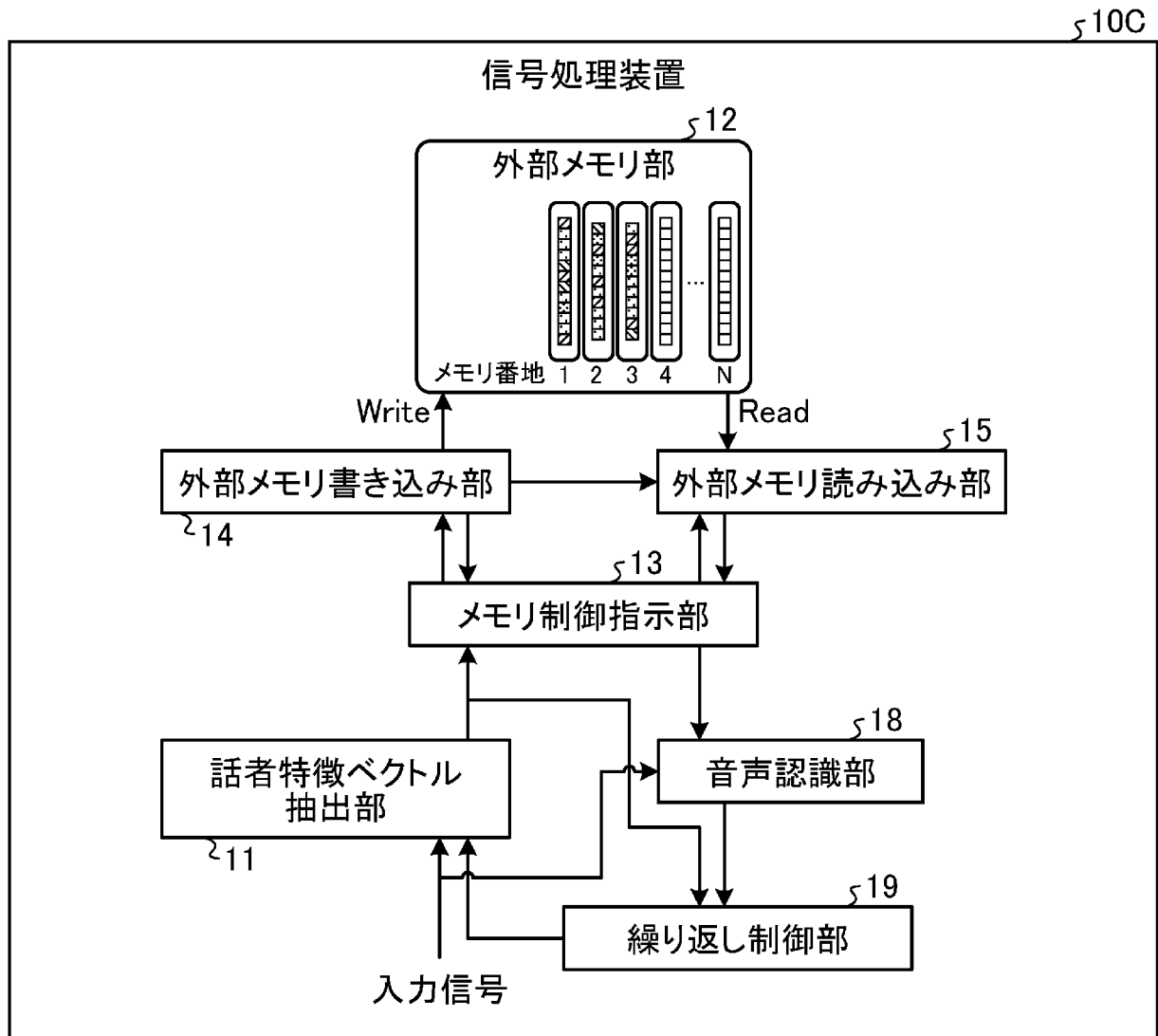
[図2]



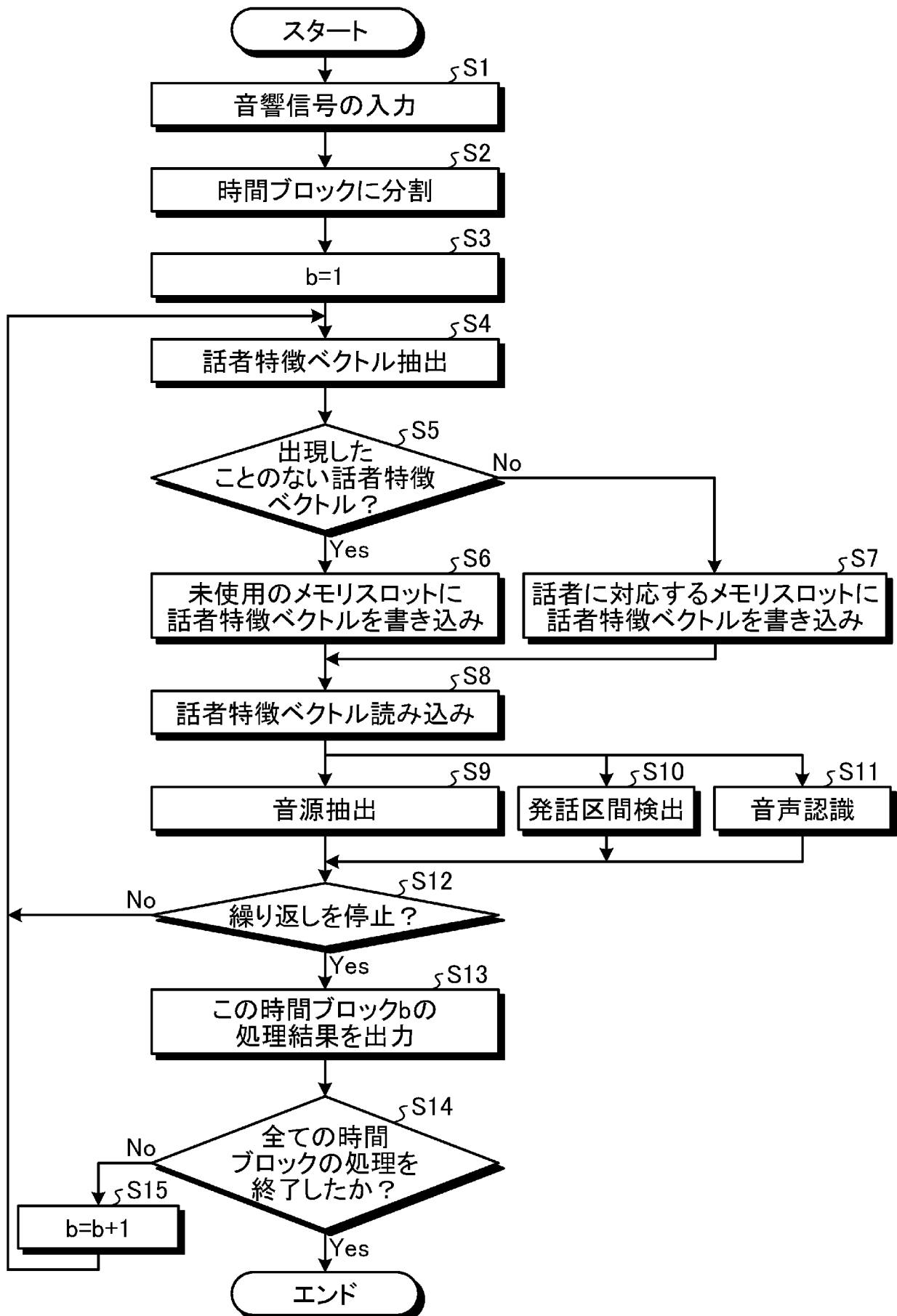
[図3]



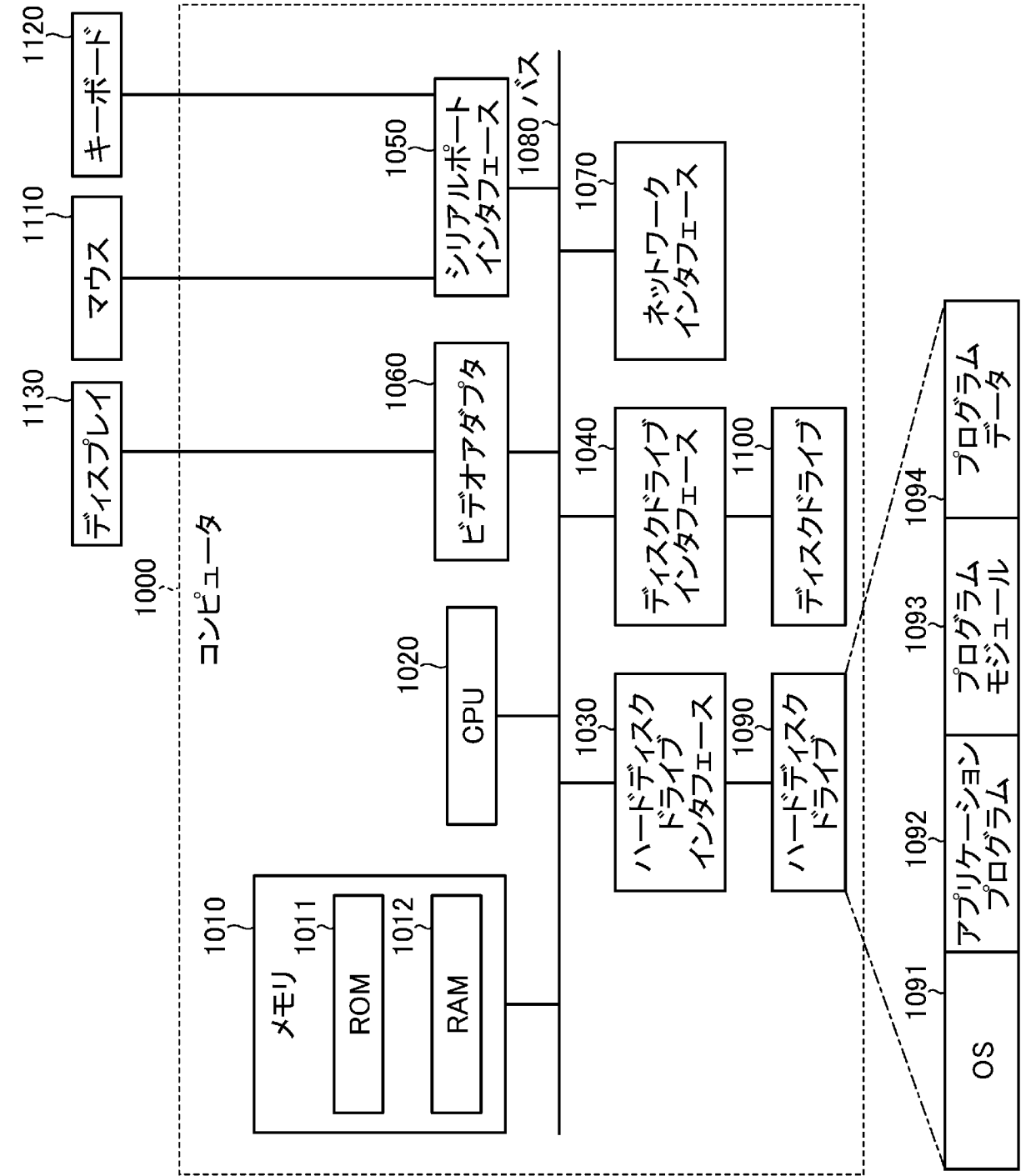
[図4]



[図5]



[図6]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/049247

A. CLASSIFICATION OF SUBJECT MATTER

G10L 21/028(2013.01)i; G10L 21/0308(2013.01)i

FI: G10L21/028 B; G10L21/0308 Z

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)

G10L17/00-17/26, 21/00-25/93

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 1971-2021

Registered utility model specifications of Japan 1996-2021

Published registered utility model applications of Japan 1994-2021

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 2020/039571 A1 (MITSUBISHI ELECTRIC CORP.) 27 February 2020 (2020-02-27) entire text	1-8
A	JP 2020-13034 A (HITACHI, LTD.) 23 January 2020 (2020-01-23) entire text	1-8
A	JP 2007-233239 A (NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY) 13 September 2007 (2007-09-13) entire text	1-8
A	JP 2020-134657 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 31 August 2020 (2020-08-31) entire text	1-8
A	NEUMANN, Thilo von et al., "ALL-NEURAL ONLINE SOURCE SEPARATION, COUNTING, AND DIARIZATION FOR MEETING ANALYSIS", ICASSP 2019[online], May 2019, [retrieved on 08 April 2021], Retrieved from the Internet: <URL: https://arxiv.org/pdf/1902.07881.pdf > pp. 1-5	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

"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
09 April 2021 (09.04.2021)Date of mailing of the international search report
20 April 2021 (20.04.2021)Name and mailing address of the ISA/
Japan Patent Office
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/JP2020/049247

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
WO 2020/039571 A1	27 Feb. 2020	(Family: none)	
JP 2020-13034 A	23 Jan. 2020	(Family: none)	
JP 2007-233239 A	13 Sep. 2007	(Family: none)	
JP 2020-134657 A	31 Aug. 2020	WO 2020/170907 A1 entire text	

A. 発明の属する分野の分類（国際特許分類（IPC）） G10L 21/028(2013.01)i; G10L 21/0308(2013.01)i FI: G10L21/028 B; G10L21/0308 Z										
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）） G10L17/00-17/26, 21/00-25/93										
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 <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922-1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971-2021年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996-2021年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994-2021年</td> </tr> </table>			日本国実用新案公報	1922-1996年	日本国公開実用新案公報	1971-2021年	日本国実用新案登録公報	1996-2021年	日本国登録実用新案公報	1994-2021年
日本国実用新案公報	1922-1996年									
日本国公開実用新案公報	1971-2021年									
日本国実用新案登録公報	1996-2021年									
日本国登録実用新案公報	1994-2021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 2020/039571 A1（三菱電機株式会社）27.02.2020（2020-02-27） 全文	1-8								
A	JP 2020-13034 A（株式会社日立製作所）23.01.2020（2020-01-23） 全文	1-8								
A	JP 2007-233239 A（独立行政法人産業技術総合研究所）13.09.2007（2007-09-13） 全文	1-8								
A	JP 2020-134657 A（日本電信電話株式会社）31.08.2020（2020-08-31） 全文	1-8								
A	Thilo von NEUMANN, et al., ALL-NEURAL ONLINE SOURCE SEPARATION, COUNTING, AND DIARIZATION FOR MEETING ANALYSIS, ICASSP 2019[online], 2019.05, [retrieved on 2021.04.08], Retrieved from the Internet: <URL: https://arxiv.org/pdf/1902.07881.pdf> pp.1-5	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 09.04.2021		国際調査報告の発送日 20.04.2021								
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） 山下 剛史 5Z 8946 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/JP2020/049247

引用文献			公表日	パテントファミリー文献	公表日
WO	2020/039571	A1	27.02.2020	(ファミリーなし)	
JP	2020-13034	A	23.01.2020	(ファミリーなし)	
JP	2007-233239	A	13.09.2007	(ファミリーなし)	
JP	2020-134657	A	31.08.2020	WO 2020/170907 A1	全文