

(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6月18日(18.06.2020)



(10) 国際公開番号

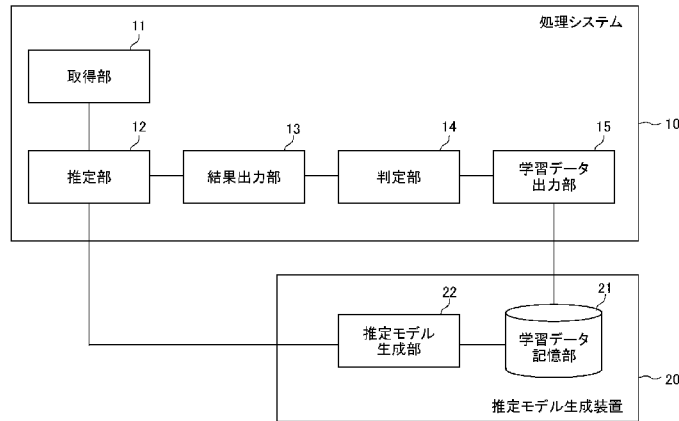
WO 2020/121616 A1

- (51) 国際特許分類:  
*G10L 15/06* (2013.01)    *G10L 15/10* (2006.01)  
*G10L 15/00* (2013.01)    *H04M 3/56* (2006.01)
- (21) 国際出願番号:                    PCT/JP2019/036062
- (22) 国際出願日:                    2019年9月13日(13.09.2019)
- (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 2018-231461    2018年12月11日(11.12.2018) 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: 松浦 廣樹 (MATSUURA Hiroki); 〒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). 古明地 秀治 (KOMEJI Shuji); 〒1368627 東京都江東区新木場一丁目18番7号 NECソリューションイノベータ株式会社内 Tokyo (JP). 城風 孝行 (SHIROKAZE Takayuki); 〒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). 吉田 玲司 (YOSHIDA Ryoji); 〒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: 速水 進治 (HAYAMI Shinji); 〒1410031 東京都品川区西五反田7丁目9番2号 KDX五反田ビル9階 Tokyo (JP).

(54) Title: PROCESSING SYSTEM, PROCESSING METHOD, AND PROGRAM

(54) 発明の名称: 処理システム、処理方法及びプログラム

[図2]



- 10 Processing system
- 11 Acquisition unit
- 12 Estimation unit
- 13 Result output unit
- 14 Determination unit
- 15 Learning data output unit
- 20 Estimation model generation device
- 21 Learning data storage unit
- 22 Estimation model generation unit

(57) Abstract: The present invention provides a processing system (10) comprising: an acquisition unit (11) for acquiring either target speech data in which a target speech which is a speech subject to language estimation is recorded or a target feature amount that represents the feature of the target speech; an estimation unit (12) for estimating the language of the target speech on the basis of an estimation model for estimating the language of a speech from speech data in which the speech is recorded or a speech feature amount that represents the feature of the speech, and the target speech data or the



WO 2020/121616 A1

(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, 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, 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))

---

target feature amount; a result output unit (13) for outputting estimation results obtained by the estimation unit (12); a determination unit (14) for determining whether an estimation result is correct; and a learning data output unit (15) for outputting the estimation result determined to be correct by the determination unit (14) and the target speech data or the target feature amount as learning data for generating the estimation model.

(57) 要約 : 本発明は、言語推定対象の音声である対象音声を記録した対象音声データ、又は、対象音声の特徴を示す対象特徴量を取得する取得部 (11) と、音声を記録した音声データ又は音声の特徴を示す音声特徴量から音声の言語を推定する推定モデルと、対象音声データ又は対象特徴量とに基づき、対象音声の言語を推定する推定部 (12) と、推定部 (12) による推定結果を出力する結果出力部 (13) と、推定結果の正誤判定を行う判定部 (14) と、判定部 (14) により正しいと判定された推定結果と、対象音声データ又は対象特徴量とを、推定モデルを生成するための学習データとして出力する学習データ出力部 (15) と、を有する処理システム (10) を提供する。

## 明 細 書

**発明の名称：処理システム、処理方法及びプログラム**

### 技術分野

[0001] 本発明は、処理システム、処理方法及びプログラムに関する。

### 背景技術

[0002] 特許文献1は、機械学習で生成した推定モデルを用いて、発話データから発話の言語を推定する装置を開示している。

[0003] 特許文献2は、通話相手である外国人と、当該外国人からの電話に対応する対応者と、当該外国人と当該対応者とのコミュニケーションをサポートするサポートチームとを含む三者の通話を実現する多言語オペレートシステムを開示している。具体的には、当該多言語オペレートシステムは、外国人又は対応者から言語の指定を受付けた後、指定された言語に対応したサポートチームを接続して三者通話を開始する。すなわち、当事者が指定した内容に基づき、外国人の言語が特定される。

### 先行技術文献

#### 特許文献

[0004] 特許文献1：特開2012-103554号公報

特許文献2：特開2003-32373号公報

### 発明の概要

#### 発明が解決しようとする課題

[0005] 言語の推定精度を向上させるためには機械学習する学習データを増やす必要があるが、特許文献1及び2はいずれも、当該課題及びその解決手段を開示していない。本発明は、機械学習で生成した推定モデルを用いて言語を推定する技術において、推定精度を向上させることを課題とする。

#### 課題を解決するための手段

[0006] 本発明によれば、

言語推定対象の音声である対象音声を記録した対象音声データ、又は、前

記対象音声の特徴を示す対象特徴量を取得する取得手段と、

音声を記録した音声データ又は前記音声の特徴を示す音声特徴量から前記音声の言語を推定する推定モデルと、前記対象音声データ又は前記対象特徴量とに基づき、前記対象音声の言語を推定する推定手段と、

前記推定手段による推定結果を出力する結果出力手段と、

前記推定結果の正誤判定を行う判定手段と、

前記判定手段により正しいと判定された前記推定結果と、前記対象音声データ又は前記対象特徴量とを、前記推定モデルを生成するための学習データとして出力する学習データ出力手段と、

を有する処理システムが提供される。

[0007] また、本発明によれば、

コンピュータが、

言語推定対象の音声である対象音声を記録した対象音声データ、又は、前記対象音声の特徴を示す対象特徴量を取得する取得工程と、

音声を記録した音声データ又は前記音声の特徴を示す音声特徴量から前記音声の言語を推定する推定モデルと、前記対象音声データ又は前記対象特徴量とに基づき、前記対象音声の言語を推定する推定工程と、

前記推定工程での推定結果を出力する結果出力工程と、

前記推定結果の正誤判定を行う判定工程と、

前記判定工程で正しいと判定された前記推定結果と、前記対象音声データ又は前記対象特徴量とを、前記推定モデルを生成するための学習データとして出力する学習データ出力工程と、

を実行する処理方法が提供される。

[0008] また、本発明によれば、

コンピュータを、

言語推定対象の音声である対象音声を記録した対象音声データ、又は、前記対象音声の特徴を示す対象特徴量を取得する取得手段、

音声を記録した音声データ又は前記音声の特徴を示す音声特徴量から前記

音声の言語を推定する推定モデルと、前記対象音声データ又は前記対象特徴量とに基づき、前記対象音声の言語を推定する推定手段、  
前記推定手段による推定結果を出力する結果出力手段、  
前記推定結果の正誤判定を行う判定手段、  
前記判定手段により正しいと判定された前記推定結果と、前記対象音声データ又は前記対象特徴量とを、前記推定モデルを生成するための学習データとして出力する学習データ出力手段、  
として機能させるプログラムが提供される。

### 発明の効果

[0009] 本発明によれば、機械学習で生成した推定モデルを用いて言語を推定する技術において、推定精度を向上させることができる。

### 図面の簡単な説明

[0010] 上述した目的、およびその他の目的、特徴および利点は、以下に述べる好適な実施の形態、およびそれに付随する以下の図面によってさらに明らかになる。

- [0011] [図1]本実施形態の処理システムのハードウェア構成の一例を示す図である。  
[図2]本実施形態の処理システムの機能ブロック図の一例である。  
[図3]本実施形態の処理システムの処理の流れの一例を示すフローチャートである。  
[図4]本実施形態の処理システムの機能ブロック図の一例である。  
[図5]本実施形態の処理システムの機能ブロック図の一例である。  
[図6]本実施形態の処理システムが処理する情報の一例を模式的に示す図である。  
[図7]本実施形態の処理システムの処理の流れの一例を示すシーケンス図である。  
[図8]本実施形態の処理システムが出力する画面の一例を模式的に示す図である。  
[図9]本実施形態の処理システムが出力する画面の一例を模式的に示す図であ

る。

[図10]本実施形態の処理システムが出力する画面の一例を模式的に示す図である。

[図11]本実施形態の処理システムが出力する画面の一例を模式的に示す図である。

[図12]本実施形態の処理システムの処理の流れの一例を示すフローチャートである。

[図13]本実施形態の処理システムの処理の流れの一例を示すシーケンス図である。

[図14]本実施形態の処理システムが出力する画面の一例を模式的に示す図である。

[図15]本実施形態の処理システムの処理の流れの一例を示すシーケンス図である。

[図16]本実施形態の処理システムが出力する画面の一例を模式的に示す図である。

[図17]本実施形態の処理システムの処理の流れの一例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

### [0012] <第1の実施形態>

まず、本実施形態の処理システムの概要を説明する。処理システムは、機械学習で生成した推定モデルと、言語推定対象の音声を記録した音声データ又はその音声の特徴量とに基づき、言語推定対象の音声の言語を推定する。次いで、処理システムは、推定結果を出力するとともに、推定結果の正誤判定を行う。そして、正しいと判定された推定結果と、言語推定対象の音声を記録した音声データ又はその音声の特徴量とを、推定モデルを生成するための学習データとして出力する。

[0013] 次に、処理システム10の構成を詳細に説明する。まず、処理システム10のハードウェア構成の一例について説明する。処理システム10が備える

各機能部は、任意のコンピュータのCPU (Central Processing Unit)、メモリ、メモリにロードされるプログラム、そのプログラムを格納するハードディスク等の記憶ユニット（あらかじめ装置を出荷する段階から格納されているプログラムのほか、CD (Compact Disc) 等の記憶媒体やインターネット上のサーバ等からダウンロードされたプログラムをも格納できる）、ネットワーク接続用インターフェイスを中心にハードウェアとソフトウェアの任意の組合せによって実現される。そして、その実現方法、装置にはいろいろな変形例があることは、当業者には理解されることである。

[0014] 図1は、処理システム10のハードウェア構成を例示するブロック図である。図1に示すように、処理システム10は、プロセッサ1A、メモリ2A、入出力インターフェイス3A、周辺回路4A、バス5Aを有する。周辺回路4Aには、様々なモジュールが含まれる。処理システム10は周辺回路4Aを有さなくてもよい。なお、処理システム10は物理的及び／又は論理的に分かれた複数の装置で構成されてもよい。この場合、複数の装置各々が上記ハードウェア構成を備えることができる。

[0015] バス5Aは、プロセッサ1A、メモリ2A、周辺回路4A及び入出力インターフェイス3Aが相互にデータを送受信するためのデータ伝送路である。プロセッサ1Aは、例えばCPU、GPU (Graphics Processing Unit) などの演算処理装置である。メモリ2Aは、例えばRAM (Random Access Memory) やROM (Read Only Memory) などのメモリである。入出力インターフェイス3Aは、入力装置、外部装置、外部サーバ、外部センサ、カメラ等から情報を取得するためのインターフェイスや、出力装置、外部装置、外部サーバ等に情報を出力するためのインターフェイスなどを含む。入力装置は、例えばキーボード、マウス、マイク等である。出力装置は、例えばディスプレイ、スピーカ、プリンター、メーラ等である。プロセッサ1Aは、各モジュールに指令を出し、それらの演算結果をもとに演算を行うことができる。

[0016] 次に、処理システム10の機能構成の一例について説明する。図2の機能ブロック図に示すように、処理システム10は、取得部11と、推定部12

と、結果出力部 13 と、判定部 14 と、学習データ出力部 15 とを有する。

[0017] なお、図では、処理システム 10 と推定モデル生成装置 20 とを分けて記載しているが、処理システム 10 は推定モデル生成装置 20 を含んでもよい。すなわち、処理システム 10 と推定モデル生成装置 20 とは物理的及び／又は論理的に分かれてもよいし、処理システム 10 と推定モデル生成装置 20 は物理的及び論理的に一体となってもよい。

[0018] 学習データ記憶部 21 は、音声を記録した音声データ又は音声の特徴を示す音声特徴量と、その音声の言語（例：日本語、英語、中国語等）とをペアにした学習データを記憶する。音声特徴量は、音声データから抽出される。

[0019] 推定モデル生成部 22 は、学習データ記憶部 21 に記憶されている学習データを用いた機械学習により、音声データ又は音声特徴量から音声の言語を推定する推定モデルを生成する。機械学習の技法は特段制限されず、あらゆる技法を採用できる。

[0020] 取得部 11 は、言語推定対象の音声である対象音声を記録した対象音声データ、又は、対象音声の特徴を示す対象特徴量を取得する。対象特徴量は、対象音声データから抽出される。

[0021] なお、本明細書において、「取得」とは、ユーザ入力に基づき、又は、プログラムの指示に基づき、「自装置が他の装置や記憶媒体に格納されているデータを取りに行くこと（能動的な取得）」、たとえば、他の装置にリクエストまたは問い合わせして受信すること、他の装置や記憶媒体にアクセスして読み出すこと等、および、ユーザ入力に基づき、又は、プログラムの指示に基づき、「自装置に他の装置から出力されるデータを入力すること（受動的な取得）」、たとえば、配信（または、送信、プッシュ通知等）されるデータを受信すること、また、受信したデータまたは情報の中から選択して取得すること、及び、「データを編集（テキスト化、データの並び替え、一部データの抽出、ファイル形式の変更等）などして新たなデータを生成し、当該新たなデータを取得すること」の少なくともいずれか一方を含む。

[0022] 推定部 12 は、推定モデル生成部 22 が生成した推定モデルと、取得部 1

1 が取得した対象音声データ又は対象特徴量とに基づき、対象音声の言語を推定する。

[0023] 結果出力部 1 3 は、推定部 1 2 による推定結果を出力する。結果出力部 1 3 は、ディスプレイ、スピーカ、プリンター、投影装置、メーラ等の出力装置を介して推定結果を出力することができる。

[0024] 判定部 1 4 は、推定部 1 2 による推定結果の正誤判定を行う。判定部 1 4 は、所定の判定材料を取得し、当該判定材料に基づき推定結果の正誤判定を行うことができる。判定材料は、ユーザにより入力された推定結果の正誤を直接的に示す情報（「正しい」又は「間違っている」）であってもよいし、ユーザ入力以外の手段で取得した推定結果の正誤を間接的に示す情報であってもよい。なお、判定部 1 4 が、ユーザ入力以外の手段で取得した推定結果の正誤を間接的に示す情報に基づき推定結果の正誤判定を行う例は、以下の実施形態で説明する。

[0025] 学習データ出力部 1 5 は、判定部 1 4 により正しいと判定された推定結果と、推定部 1 2 によりその推定結果が導き出された対象音声データ又は対象特徴量とを、推定モデルを生成するための学習データとして出力する。学習データ出力部 1 5 により出力された学習データは、学習データ記憶部 2 1 に蓄積される。

[0026] 次に、図 3 のフローチャートを用いて、処理システム 1 0 の処理の流れの一例を説明する。

[0027] まず、取得部 1 1 が新たな対象音声データ又は対象特徴量を取得すると（S 1 0）、推定部 1 2 は、推定モデル生成部 2 2 により生成された推定モデルと、S 1 0 で取得された対象音声データ又は対象特徴量とに基づき、対象音声の言語を推定する（S 1 1）。次いで、結果出力部 1 3 は、S 1 1 で推定された推定結果を出力する（S 1 2）。

[0028] その後、判定部 1 4 は、S 1 2 で推定された推定結果の正誤判定を行う（S 1 3）。例えば、判定部 1 4 は、ユーザにより入力された推定結果の正誤を直接的に示す情報を取得し、当該情報に基づき推定結果の正誤判定を行っ

てもよい。

[0029] 判定部 14 の判定結果が「正しい」でない場合 (S 14 の No)、処理を終了する。

[0030] 一方、判定部 14 の判定結果が「正しい」である場合 (S 14 の Yes)、学習データ出力部 15 は、S 10 で取得された対象音声データ又は対象特徴量と、S 11 で推定された推定結果とを学習データとして出力する (S 15)。

[0031] S 15 で学習データ出力部 15 により出力された学習データは、学習データ記憶部 21 に蓄積される。その後、図示しないが、推定モデル生成部 22 は、学習データ記憶部 21 に記憶されている学習データを用いた機械学習により、音声データ又は音声特徴量から音声の言語を推定する推定モデルを生成する。

[0032] 推定モデル生成部 22 は、新たな 1 つの学習データが学習データ記憶部 21 に登録される毎に新たに機械学習を行い、新たな推定モデルを生成してもよい。その他、推定モデル生成部 22 は、新たな所定個数 (2 以上の任意の数) の学習データが学習データ記憶部 21 に登録される毎に新たに機械学習を行い、新たな推定モデルを生成してもよい。その他、推定モデル生成部 22 は、予め定められた時刻になる毎に新たに機械学習を行い、新たな推定モデルを生成してもよい。その他、推定モデル生成部 22 は、推定モデルを更新するユーザ入力があったタイミングで新たに機械学習を行い、新たな推定モデルを生成してもよい。なお、ここで例示した機械学習を行うタイミングは一例であり、これに限定されない。

[0033] 以上、言語推定と、推定結果を用いた学習データの生成とを連動して実行する処理システム 10 によれば、言語を推定する処理を実行しながら、学習データを増やすことができる。結果、学習データを増やすための面倒な作業を回避できる。

[0034] また、上記特徴を有する処理システム 10 によれば、実務において処理システム 10 を利用して言語を推定させながら、実務で得られた推定結果に基

づき学習データを増やすことができる。このような処理システム10によれば、実務に適した学習データ、具体的には実務において接する機会がある言語の学習データを効率的に増やすことができる。そして、実務に適した推定モデルが生成されることとなる。このように、処理システム10を利用する環境のニーズに特化した性能を処理システム10に与えることができる。

[0035] <第2の実施形態>

本実施形態の処理システム10は、通話手段を有し、通話相手の音声を対象音声とする点等が第1の実施形態と異なる。

[0036] 処理システム10のハードウェア構成の一例は、第1の実施形態と同様である。

[0037] 処理システム10の機能ブロック図の一例は、図4で示される。図示するように、処理システム10は、取得部11と、推定部12と、結果出力部13と、判定部14と、学習データ出力部15と、通話部16と、通話制御部17と、通訳担当者抽出部18と、通訳担当者出力部19とを有する。なお、図示しないが、処理システム10は、学習データ記憶部21及び推定モデル生成部22を備えてもよい。

[0038] 処理システム10は、図5に示す通話端末1、又は、通話端末1とサーバ2の両方により構成される。すなわち、処理システム10が有する機能部の全てを通話端末1が備えてもよい。その他、処理システム10が有する機能部の一部を通話端末1が備え、残りの機能部をサーバ2が備えてもよい。通話端末1は、パーソナルコンピュータ、タブレット端末、スマートフォン等の端末装置に所定のアプリケーションをインストールすることで所定の機能部を実現した装置であってもよいし、所定の機能部が予め実現された通話専用の端末であってもよい。

[0039] 通話部16は、他の通話端末と接続し、通話を実現する機能を有する。

[0040] 取得部11は、通話相手の音声を記録した対象音声データ、又は、通話相手の音声の特徴を示す対象特徴量を取得する。取得部11のその他の構成は、第1の実施形態と同様である。

- [0041] 推定部 12、結果出力部 13、判定部 14 及び学習データ出力部 15 の構成は、第 1 の実施形態と同様である。
- [0042] 通訳担当者抽出部 18 は、複数の言語各々の通訳担当者及び連絡先情報を登録した通訳者情報を参照し、推定結果で示される言語の通訳担当者を抽出する。図 6 に、通訳者情報の一例を模式的に示す。
- [0043] 通訳担当者出力部 19 は、通訳担当者抽出部 18 により抽出された通訳担当者を示す情報を出力する。通訳担当者出力部 19 は、ディスプレイ、スピーカ、プリンター、投影装置、メーラ等の出力装置を介して、通訳担当者抽出部 18 により抽出された通訳担当者を示す情報を出力することができる。例えば、通訳担当者出力部 19 は、抽出された通訳担当者の一覧を出力してもよい。
- [0044] 通話制御部 17 は、ユーザ入力に基づき通話相手との通話を保留する。そして、通話制御部 17 は、通訳担当者抽出部 18 により抽出された通訳担当者の中から 1 人を指定し、指定した通訳担当者との通話（例：内線通話）を開始させる。例えば、通話制御部 17 は、ユーザ入力に基づき 1 人の通訳担当者を指定してもよいし、予め定められたルールに従い 1 人の通訳担当者を指定してもよい。予め定められたルールは、例えば、「当日のその時まで三者通話した回数が最も少ない人」等が例示されるが、これに限定されない。
- [0045] そして、通話制御部 17 は、通話相手との通話を保留し、指定された通訳担当者との通話を開始させた後、所定の入力を受付けると、通話相手及び通訳担当者を含む三者通話を開始させることができる。
- [0046] 次に、図 7 のシーケンス図を用いて、処理システム 10 の処理の流れの一例を説明する。
- [0047] 第 1 の通話端末は、通話相手が利用する通話端末である。ここでは、通話相手は外国人であるものとする。なお、外国人は、本実施形態の処理システム 10 を利用する地域における公用語以外の言語を話す人物あるいは、当事者が指定した言語以外の言語を話す人物であることが想定される。

- [0048] 第2の通話端末は、電話で顧客対応を行うことを業務とする対応オペレータが利用する通話端末である。第2の通話端末は、処理システム10が有する機能部の全てを備えてもよいし、処理システム10が有する機能部の一部を備えてもよい。
- [0049] 第3の通話端末は、所定の言語に対応した通訳者が利用する通話端末である。
- [0050] まず、第1の通話端末から第2の通話端末に通話要求が送信され（S20）、当該通話要求に対する通話受諾が第2の通話端末から第1の通話端末に送信されると（S21）、第1の通話端末と第2の通話端末との間で通話が開始され（S22）、通話が可能な状態（以下、「通話状態」）となる。通話状態において、第2の通話端末に表示される画面の一例は図8に示される。図では、通話中であることを示す情報、通話相手の電話番号、及び、各種操作を行うためのUI（user interface）ボタンを含むウィンドウが表示されている。UIボタンは、通話を終了するための通話終了ボタン、通話を保留状態にするための保留ボタン、及び、通話相手の言語を推定する処理を実行するための言語推定ボタン等を含む。
- [0051] 対応オペレータは、通話中に、通話相手が外国人であることを認識したが、通話相手の言語を認識できない場合、保留ボタンを押して通話を保留状態にした後（S23、S24）、言語推定ボタンを押す。言語推定ボタンの押下に応じて、言語推定処理が実行される（S25）。
- [0052] S25では、取得部11は、通話相手の音声を記録した対象音声データ、又は、通話相手の音声の特徴を示す対象特徴量を取得する。例えば、第2の通話端末は、通話相手に関わらず、全ての通話を録音するように構成されてもよい。そして、取得部11は、上記録音された音声データの一部又は全部を、対象音声データとして取得してもよい。また、取得部11は、このように取得した対象音声データから対象特徴量を抽出してもよい。次いで、推定部12は、推定モデル生成部22により生成された推定モデルと、取得部11により取得された対象音声データ又は対象特徴量とに基づき、対象音声の

言語を推定する。

[0053] そして、結果出力部 13 は、推定された推定結果を出力する (S 26)。図 9 に、結果出力部 13 が推定結果を出力した時に第 2 の通話端末に表示される画面の一例を示す。図では、推定結果を示すウィンドウが表示されている。当該ウィンドウでは、推定結果で示される言語の通訳担当者を示す情報のリンクがはられている。

[0054] 例えば図 9 に示す画面において、通訳担当者を示す情報のリンクが押されると、通訳担当者出力部 19 は、通訳担当者抽出部 18 により抽出された通訳担当者を示す情報を取得し、出力する (S 27)。通訳担当者抽出部 18 は、複数の言語各々の通訳担当者及び連絡先情報を登録した通訳者情報 (図 6 参照) を参照し、推定結果で示される言語の通訳担当者を抽出する。図 10 に、通訳担当者出力部 19 が通訳担当者を示す情報を出力した時に第 2 の通話端末に表示される画面の一例を示す。図では、推定結果の言語の通訳担当者を示すウィンドウが表示されている。当該ウィンドウでは、通訳担当者の氏名の一覧を表示するとともに、各通訳担当者との通話を開始するための通話ボタンが表示されている。結果出力部 13 は、第 2 の通話端末が備える。通訳担当者出力部 19 は、第 2 の通話端末が備える。

[0055] 例えば図 10 に示す画面において、1 人の通訳担当者の通話ボタンが押されると (1 人の通訳担当者を指定する入力 (S 28))、第 2 の通話端末は指定された通訳担当者の連絡先情報を通訳者情報 (図 6 参照) に基づき特定し、その通訳担当者の第 3 の通話端末に対して通話要求を送信する (S 29)。そして、当該通話要求に対する通話受諾が第 3 の通話端末から第 2 の通話端末に送信されると (S 30)、第 2 の通話端末と第 3 の通話端末との間で通話が可能な状態となる (S 31)。当該状態において、第 2 の通話端末に表示される画面の一例は図 11 に示される。図では、通話中であることを示す情報、指定された通訳担当者の名称、及び、各種操作を行うための UI ボタンを含むウィンドウが表示されている。UI ボタンは、通話を終了するための通話終了ボタン、保留中の通話相手を含めた三者通話を開始するため

の三者通話開始ボタン等を含む。

- [0056] 当該画面において、対応オペレータが三者通話開始ボタンを押すと（S 3 2）、第 1 の通話端末と、第 2 の通話端末と、第 3 の通話端末との間で三者通話が可能な状態となる。
- [0057] ここで、本実施形態の変形例を説明する。図 7 のシーケンス図を用いて説明した例では、通話相手が外国人である場合、通話相手と、対応オペレータと、通話相手の言語に対応した通訳者とを含む三者通話を行った。変形例では、通話相手が外国人である場合、対応オペレータは、通話相手の言語に対応した特定言語用対応オペレータに通話を転送してもよい。特定言語用対応オペレータは、対応オペレータと同様のスキルを有し、特定の言語で顧客対応を行うことができるオペレータである。
- [0058] 変形例の場合、図 7 のシーケンス図を用いて行った処理の説明における「通訳担当者」を「特定言語用対応オペレータ」に読み替えればよい。そして、S 3 1 の後に第 2 の通話端末又は第 3 の通話端末において所定の入力が行われると、第 1 の通話端末と第 2 の通話端末との間の通話を、第 1 の通話端末と第 3 の通話端末との間の通話に切り替える転送が行われる。
- [0059] また、図 7 のシーケンス図を用いて説明した例では、保留ボタンに対する操作に応じて保留状態になった後に、言語推定ボタンに対する操作に応じて言語推定処理が開始された。変形例として、通話状態のままで言語推定ボタンに対する操作が可能であり、通話状態のままで言語推定処理が実行されてもよい。その他、保留ボタンに対する操作に応じて、保留状態になるとともに、言語推定処理が開始されてもよい。すなわち、保留ボタンに対する操作は、保留状態に指示、及び、言語推定処理を開始する指示の両方を兼ねてもよい。
- [0060] また、ユーザやオペレータの操作に応じて実行する処理の一部（例：S 3 2 等）を機械的な自動処理に置き換えてもよい。この場合も、本実施形態の作用効果が実現される。
- [0061] 以上説明した本実施形態の処理システム 10 によれば、第 1 の実施形態と

同様な作用効果を実現できる。

[0062] また、本実施形態の処理システム10は、通話で顧客対応を行うコールセンター等において利用することができる。コールセンターに外国人から電話がかかってきた場合に、その外国人の言語を認識できないと、その言語に対応した適切な通訳担当者を見つけるまでに時間がかかり、顧客を待たしてしまうという不都合が発生し得る。本実施形態の処理システム10によれば、このような不都合を軽減できる。結果、コールセンターにおける顧客満足度を高めることができる。

[0063] <第3の実施形態>

本実施形態の処理システム10は、第2の実施形態と同様の構成を有し、第1及び第2の実施形態と異なる手段で推定結果の正誤判定を行う点で、第1及び第2の実施形態と異なる。

[0064] 処理システム10のハードウェア構成の一例は、第1及び第2の実施形態と同様である。

[0065] 処理システム10の機能ブロック図の一例は、図4で示される。図示するように、処理システム10は、取得部11と、推定部12と、結果出力部13と、判定部14と、学習データ出力部15と、通話部16と、通話制御部17と、通訳担当者抽出部18と、通訳担当者出力部19とを有する。なお、図示しないが、処理システム10は、学習データ記憶部21及び推定モデル生成部22を備えてもよい。

[0066] 処理システム10は、図5に示す通話端末1、又は、通話端末1とサーバ2の両方により構成される。すなわち、処理システム10が有する機能部の全てを通話端末1が備えてもよい。その他、処理システム10が有する機能部の一部を通話端末1が備え、残りの機能部をサーバ2が備えてもよい。通話端末1は、パーソナルコンピュータ、タブレット端末、スマートフォン等の端末装置に所定のアプリケーションをインストールすることで所定の機能部を実現した装置であってもよいし、所定の機能部が予め実現された通話専用の端末であってもよい。

- [0067] 取得部 1 1、推定部 1 2、結果出力部 1 3、学習データ出力部 1 5、通話部 1 6、通話制御部 1 7、通訳担当者抽出部 1 8 及び通訳担当者出力部 1 9 の構成は、第 1 又は第 2 の実施形態と同様である。
- [0068] 判定部 1 4 は、通話相手及び通訳担当者を含む三者通話が開始されてからの経過時間が所定時間を超えた場合、推定結果が正しいと判定する。判定部 1 4 のその他の構成は、第 1 又は第 2 の実施形態と同様である。
- [0069] 次に、処理システム 1 0 の処理の流れの一例を説明する。通話相手及び通訳担当者を含む三者通話が開始されるまでの流れは、第 2 の実施形態で説明した通りである（図 7 参照）。
- [0070] 三者通話が開始されると、判定部 1 4 は三者通話開始からの経過時間を監視する。そして、図 1 2 のフローチャートに示すように、経過時間が所定時間を超えずに三者通話が終了した場合（S 4 0 の N o）、判定部 1 4 は S 2 5 の推定結果は正しくないと判定し、処理を終了する。
- [0071] 一方、経過時間が所定時間を超えた場合（S 4 0 の Y e s）、判定部 1 4 は S 2 5 の推定結果は正しいと判定する。そして、学習データ出力部 1 5 は、S 2 5 で取得された対象音声データ又は対象特徴量と、S 2 5 で推定された推定結果とを学習データとして出力する（S 4 1）。
- [0072] なお、判定部 1 4 が、特定言語用対応オペレータに通話が転送され、所定時間経過すると、推定結果が正しいと判定してもよい。
- [0073] 以上説明した本実施形態の処理システム 1 0 によれば、第 1 又は第 2 の実施形態と同様の作用効果を実現できる。また、本実施形態の処理システム 1 0 によれば、ユーザ入力以外の手段で取得した推定結果の正誤を間接的に示す情報に基づき推定結果の正誤判定を行うことができる。ユーザ入力なしで正誤判定できるため、ユーザ負担を軽減することができる。
- [0074] また、本実施形態の処理システム 1 0 は、通話相手及び通訳担当者を含む三者通話が継続した時間や、特定言語用対応オペレータに通話が転送されてから通話が継続した時間が所定時間を超えた場合、推定結果が正しいと判定する。推定結果が間違っている場合、他の通訳担当者に切り替えたり、他の

特定言語用対応オペレータに通話を再転送したりするため、上述した継続時間は比較的短くなる。上述した継続時間に基づき推定結果の正誤判定を行う本実施形態の処理システム10は、精度よく正誤判定を行うことができる。

[0075] <第4の実施形態>

本実施形態の処理システム10は、第2の実施形態と同様の構成を有し、第1乃至第3の実施形態と異なる手段で推定結果の正誤判定を行う点で、第1乃至第3の実施形態と異なる。

[0076] 処理システム10のハードウェア構成の一例は、第1乃至第3の実施形態と同様である。

[0077] 処理システム10の機能ブロック図の一例は、図4で示される。図示するように、処理システム10は、取得部11と、推定部12と、結果出力部13と、判定部14と、学習データ出力部15と、通話部16と、通話制御部17と、通訳担当者抽出部18と、通訳担当者出力部19とを有する。なお、図示しないが、処理システム10は、学習データ記憶部21及び推定モデル生成部22を備えてもよい。

[0078] 処理システム10は、図5に示す通話端末1、又は、通話端末1とサーバ2の両方により構成される。すなわち、処理システム10が有する機能部の全てを通話端末1が備えてもよい。その他、処理システム10が有する機能部の一部を通話端末1が備え、残りの機能部をサーバ2が備えてもよい。通話端末1は、パーソナルコンピュータ、タブレット端末、スマートフォン等の端末装置に所定のアプリケーションをインストールすることで所定の機能部を実現した装置であってもよいし、所定の機能部が予め実現された通話専用の端末であってもよい。

[0079] 取得部11、推定部12、結果出力部13、学習データ出力部15、通話部16、通訳担当者抽出部18及び通訳担当者出力部19の構成は、第1乃至第3の実施形態のいずれかと同様である。

[0080] 通話制御部17は、通話相手との通話を保留し、指定された通訳担当者との通話を開始させた後、対象音声データを再生し、再生音を通訳担当者の通

話端末に送信する。そして、通話制御部 17 は、その後所定の入力を受けると、通話相手及び通訳担当者を含む三者通話を開始させる。通話制御部 17 のその他の構成は、第 1 乃至第 3 の実施形態のいずれかと同様である。

[0081] 判定部 14 は、三者通話が開始されると、推定結果が正しいと判定する。判定部 14 のその他の構成は、第 1 乃至第 3 の実施形態のいずれかと同様である。

[0082] 次に、図 13 のシーケンス図を用いて、処理システム 10 の処理の流れの一例を説明する。

[0083] 第 1 の通話端末、第 2 の通話端末及び第 3 の通話端末は、第 2 の実施形態で説明した通りである。そして、S50 乃至 S61 の処理は、第 2 の実施形態で説明した図 7 の S20 乃至 S31 の処理と同様である。

[0084] S61 に示すように第 2 の通話端末と第 3 の通話端末との間で通話が可能な状態となった時に、第 2 の通話端末に表示される画面の一例は図 14 に示される。図では、通話中であることを示す情報、指定された通訳担当者の名称、及び、各種操作を行うための UI ボタンを含むウィンドウが表示されている。UI ボタンは、通話を終了するための通話終了ボタン、保留中の通話相手を含めた三者通話を開始するための三者通話開始ボタン、対象音声データを再生して対象音声を通訳担当者に聞かせるための対象音声再生ボタン等を含む。

[0085] 当該画面において、対応オペレータが対象音声再生ボタンを押すと、第 2 の通話端末は対象音声データを再生する (S62)。結果、通訳担当者は、対象音声を聞くことができる。そして、通訳担当者は、対象音声が自身の担当する言語か否かを判定し、判定結果を通話中の対応オペレータに伝える。判定結果は、口頭で伝えられてもよいし、システムを利用した通知で実現されてもよい。システムを利用した通知は、例えば第 2 の通話端末及び第 3 の通話端末で実行中のアプリケーション等を介したメッセージの送受信で実現されてもよい。

[0086] 対象音声が通訳担当者の担当する言語である場合、対応オペレータは、例

例えば図14に示す画面において、三者通話開始ボタンを押す(S63)。結果、第1の通話端末と、第2の通話端末と、第3の通話端末との間で三者通話が可能な状態となる(S64)。そして、判定部14はS55で推定された推定結果は正しいと判定し、学習データ出力部15はS55で取得された対象音声データ又は対象特徴量と、S55で推定された推定結果とを学習データとして出力する(S65)。

[0087] ここで、本実施形態の変形例を説明する。図13のシーケンス図を用いて説明した例では、通話相手が外国人である場合、通話相手と、対応オペレータと、通話相手の言語に対応した通訳者を含む三者通話を行った。変形例では、通話相手が外国人である場合、対応オペレータは、通話相手の言語に対応した特定言語用対応オペレータに通話を転送してもよい。特定言語用対応オペレータは、対応オペレータと同様のスキルを有し、特定の言語で顧客対応を行うことができるオペレータである。

[0088] 変形例の場合、図13のシーケンス図を用いて行った処理の説明における「通訳担当者」を「特定言語用対応オペレータ」に読み替えればよい。そして、S62の後に第2の通話端末又は第3の通話端末において所定の入力が行われると、第1の通話端末と第2の通話端末との間の通話を、第1の通話端末と第3の通話端末と間の通話に切り替える転送が行われる。

[0089] 判定部14は、第1の通話端末と第2の通話端末との間の通話を、第1の通話端末と第3の通話端末と間の通話に切り替える転送が行われると、S55で推定された推定結果は正しいと判定する。そして、学習データ出力部15は、S55で取得された対象音声データ又は対象特徴量と、S55で推定された推定結果とを学習データとして出力する。

[0090] また、ユーザやオペレータの操作に応じて実行する処理の一部(例：S62、S63等)を機械的な自動処理に置き換えてもよい。この場合も、本実施形態の作用効果が実現される。

[0091] 以上説明した本実施形態の処理システム10によれば、第1乃至第3の実施形態のいずれかと同様の作用効果を実現できる。また、本実施形態の処理

システム10によれば、ユーザ入力以外の手段で取得した推定結果の正誤を間接的に示す情報に基づき推定結果の正誤判定を行うことができる。ユーザ入力なしで正誤判定できるため、ユーザ負担を軽減することができる。

[0092] また、本実施形態の処理システム10は、三者通話や通話の転送を行う前に、対象音声データを再生し、対象音声を通訳担当者や特定言語用対応オペレータに聞かせることができる。そして、対象音声を聞いた通訳担当者や特定言語用対応オペレータの判定結果（対象音声の言語が自身の担当する言語か否かの判定結果）に基づき、三者通話や通話の転送を行うか否かを決定することができる。推定部12の推定結果が間違っている場合、三者通話や通話の転送が行われることはない。換言すれば、三者通話や通話の転送が行われた場合、推定部12の推定結果が正しいということができる。

[0093] 対象音声データを再生した後に三者通話や通話の転送が行われたか否かに基づき推定結果の正誤判定を行う本実施形態の処理システム10は、精度よく正誤判定を行うことができる。

[0094] <第5の実施形態>

本実施形態の処理システム10は、第2の実施形態と同様の構成を有し、第1乃至第4の実施形態と異なる手段で推定結果の正誤判定を行う点で、第1乃至第4の実施形態と異なる。

[0095] 処理システム10のハードウェア構成の一例は、第1乃至第4の実施形態と同様である。

[0096] 処理システム10の機能ブロック図の一例は、図4で示される。図示するように、処理システム10は、取得部11と、推定部12と、結果出力部13と、判定部14と、学習データ出力部15と、通話部16と、通話制御部17と、通訳担当者抽出部18と、通訳担当者出力部19とを有する。なお、図示しないが、処理システム10は、学習データ記憶部21及び推定モデル生成部22を備えてもよい。

[0097] 処理システム10は、図5に示す通話端末1、又は、通話端末1とサーバ2の両方により構成される。すなわち、処理システム10が有する機能部の

全てを通話端末1が備えてもよい。その他、処理システム10が有する機能部の一部を通話端末1が備え、残りの機能部をサーバ2が備えてもよい。通話端末1は、パーソナルコンピュータ、タブレット端末、スマートフォン等の端末装置に所定のアプリケーションをインストールすることで所定の機能部を実現した装置であってもよいし、所定の機能部が予め実現された通話専用の端末であってもよい。

[0098] 取得部11、推定部12、結果出力部13、学習データ出力部15、通話部16、通訳担当者抽出部18及び通訳担当者出力部19の構成は、第1乃至第4の実施形態のいずれかと同様である。

[0099] 通話制御部17は、通話相手との通話を保留にしている間、推定結果で示される言語で所定のメッセージを出力する保留音を再生し、再生音を通話相手の通話端末に送信する。例えば、メッセージは、「このメッセージを読み上げている言語での通話を希望しますか？希望する場合は1を、他の言語での通話を希望する場合は2を押してください。」等であってもよい。

[0100] 判定部14は、保留音を再生中に通話相手から所定の入力を受付ける。そして、判定部14は、通話相手の入力が第1の入力である場合、推定部12による推定結果が正しいと判定し、通話相手の入力が第2の入力である場合、推定部12による推定結果が正しくないと判定する。また、判定部14は、所定の期限までに通話相手の入力がない場合、推定結果が正しくないと判定することができる。所定の期限は、例えば上記メッセージを最後まで出力し終えたタイミングからの経過時間で定義される。

[0101] 通話相手からの所定の入力は、上記メッセージに対して行われるものである。例えばメッセージが「このメッセージを読み上げている言語での通話を希望しますか？希望する場合は1を、他の言語での通話を希望する場合は2を押してください。」である場合、所定の入力はボタン1を押すことでなされる入力、又は、ボタン2を押すことでなされる入力である。この例の場合、ボタン1を押すことでなされる入力が第1の入力となり、ボタン2を押すことでなされる入力が第2の入力となる。

- [0102] なお、通話制御部 17 は、推定結果が複数の言語を含む場合、信頼度が最も高い言語のメッセージを最初に再生し、通話相手の入力が入力された場合、又は、所定の期限までに通話相手の入力がない場合には、次に信頼度が高い言語のメッセージを再生してもよい。
- [0103] 次に、図 15 のシーケンス図を用いて、処理システム 10 の処理の流れの一例を説明する。
- [0104] 第 1 の通話端末は、第 2 の通話端末及び第 3 の通話端末は、第 2 の実施形態で説明した通りである。
- [0105] まず、第 1 の通話端末から第 2 の通話端末に通話要求が送信され (S 70)、当該通話要求に対する通話受諾が第 2 の通話端末から第 1 の通話端末に送信されると (S 71)、第 1 の通話端末と第 2 の通話端末との間で通話が始まり (S 72)、通話が可能な状態 (通話状態) となる。通話状態において、第 2 の通話端末に表示される画面の一例は図 8 に示される。図 8 の画面の説明は第 2 の実施形態で行ったので、ここでは省略する。
- [0106] 対応オペレータは、通話中に、通話相手が外国人であることを認識したが、通話相手の言語を認識できない場合、言語推定ボタンを押して言語推定処理を実行させる。当該入力に応じて、言語推定処理が実行される (S 73)。
- [0107] S 73 において、取得部 11 は、通話相手の音声を記録した対象音声データ、又は、通話相手の音声の特徴を示す対象特徴量を取得する。例えば、第 2 の通話端末は、通話相手に関わらず、全ての通話が録音されるように構成されてもよい。そして、取得部 11 は、上記録音された音声データの一部又は全部を、対象音声データとして取得してもよい。また、取得部 11 は、このように取得した対象音声データから対象特徴量を抽出してもよい。
- [0108] 次に、推定部 12 は、推定モデル生成部 22 により生成された推定モデルと、取得部 11 により取得された対象音声データ又は対象特徴量とに基づき、対象音声の言語を推定する。
- [0109] そして、結果出力部 13 は、推定された推定結果を出力する (S 74)。

図 1 6 に、結果出力部 1 3 が推定結果を出力した時に第 2 の通話端末に表示される画面の一例を示す。図では、推定結果を示すウィンドウが表示されている。当該ウィンドウでは、推定結果で示される言語の通訳担当者を示す情報のリンクがはられている。

[0110] 例えば図 1 6 に示す画面において、通話者との通話を保留にする保留ボタンが押されると (S 7 5)、第 1 の通話端末と第 2 の通話端末との間の通話が保留状態になる (S 7 6)。そして、保留状態の間、第 2 の通話端末は図 1 7 に示す処理を実行する。

[0111] まず、通話制御部 1 7 は、S 7 3 で推定された推定結果で示される言語で所定のメッセージを出力する保留音を再生し、再生音を第 1 の通話端末に送信する (S 8 0)。推定結果で示される言語が複数ある場合、通話制御部 1 7 は、最も信頼度の高い言語で所定のメッセージを出力する保留音を再生し、再生音を第 1 の通話端末に送信する。

[0112] その後、第 2 の通話端末は、第 1 の通話端末を介して行われるユーザ入力待ちとなる。第 1 の入力が行われた場合 (S 8 1 の第 1 の入力)、判定部 1 4 は、推定結果は正しいと判定する。そして、学習データ出力部 1 5 は、出力したメッセージの言語と、S 7 3 で取得された対象音声データ又は対象特徴量とを学習データとして出力する (S 8 2)。以降、通話制御部 1 7 は、保留状態を終了する入力を受付けるまで、通常の保留音 (例：音楽等) を再生し、再生音を第 1 の通話端末に送信する (S 8 3)。

[0113] 一方、第 2 の入力が行われた場合、又は、所定の期限までに入力がない場合 (S 8 1 の第 2 の入力／入力なし)、通話制御部 1 7 は他の言語候補があるか判断する (S 8 5)。例えば、推定結果の中に他の言語が存在する場合、又は、推定結果の中に信頼度が所定レベル以上の他の言語が存在する場合、又は、推定結果に関係なく言語候補として予め登録されている他の言語が存在する場合などには、通話制御部 1 7 は他の言語候補があると判定することができる。

[0114] 他の言語候補がない場合 (S 8 5 の N o)、通話制御部 1 7 は、保留状態

を終了する入力を受付けるまで、通常の保留音（例：音楽等）を再生し、再生音を第1の通話端末に送信する（S83）。

[0115] 他の言語候補がある場合（S85のYes）、通話制御部17は、他の言語候補で所定のメッセージを出力する保留音を再生し、再生音を第1の通話端末に送信する（S86）。そして、S81に戻り、同様の処理を繰り返す。

[0116] なお、S76の保留状態の間、図7を用いて説明したS26乃至S32の処理や、図13を用いて説明したS56乃至S63の処理が実行されてもよい。

[0117] また、ユーザやオペレータの操作に応じて実行する処理の一部（例：S75、S76等）を機械的な自動処理に置き換えてもよい。この場合も、本実施形態の作用効果が実現される。

[0118] 以上説明した本実施形態の処理システム10によれば、第1乃至第4の実施形態のいずれかと同様の作用効果を実現できる。また、本実施形態の処理システム10は、保留状態の間に、推定結果の言語でメッセージを再生し、通話相手にそのメッセージを聞かせることができる。そして、本実施形態の処理システム10は、そのメッセージに対する通話相手の入力に基づき、推定結果の正誤判定を行うことができる。このような本実施形態の処理システム10によれば、高精度に推定結果の正誤判定を行うことができる。

[0119] 以下、参考形態の例を付記する。

1. 言語推定対象の音声である対象音声を記録した対象音声データ、又は、前記対象音声の特徴を示す対象特徴量を取得する取得手段と、

音声を記録した音声データ又は前記音声の特徴を示す音声特徴量から前記音声の言語を推定する推定モデルと、前記対象音声データ又は前記対象特徴量とに基づき、前記対象音声の言語を推定する推定手段と、

前記推定手段による推定結果を出力する結果出力手段と、

前記推定結果の正誤判定を行う判定手段と、

前記判定手段により正しいと判定された前記推定結果と、前記対象音声デ

ータ又は前記対象特徴量とを、前記推定モデルを生成するための学習データとして出力する学習データ出力手段と、  
を有する処理システム。

2. 1に記載の処理システムにおいて、

通話手段を有し、

前記取得手段は、通話相手の音声を記録した前記対象音声データ、又は、前記通話相手の音声の特徴を示す前記対象特徴量を取得し、

複数の言語各々の通訳担当者及び連絡先情報を登録した通訳者情報を参照し、前記推定結果で示される言語の前記通訳担当者を抽出する通訳担当者抽出手段と、

抽出された前記通訳担当者を示す情報を出力する通訳担当者出力手段と、

通話相手との通話を保留し、抽出された前記通訳担当者の中から指定された前記通訳担当者との通話を開始させる通話制御手段と、

を有する処理システム。

3. 2に記載の処理システムにおいて、

前記通話制御手段は、前記通話相手との通話を保留し、指定された前記通訳担当者との通話を開始させた後、所定の入力を受けると、前記通話相手及び前記通訳担当者を含む三者通話を開始し、

前記判定手段は、前記三者通話が開始されてからの経過時間が所定時間を超えた場合、前記推定結果が正しいと判定する処理システム。

4. 2に記載の処理システムにおいて、

前記通話制御手段は、前記通話相手との通話を保留し、指定された前記通訳担当者との通話を開始させた後、前記対象音声データを再生し、再生音を前記通訳担当者の通話端末に送信し、その後、所定の入力を受けると、前記通話相手及び前記通訳担当者を含む三者通話を開始させ、

前記判定手段は、前記三者通話が開始されると、前記推定結果が正しいと判定する処理システム。

5. 2から4のいずれかに記載の処理システムにおいて、

前記通話制御手段は、前記通話相手との通話を保留にしている間、前記推定結果で示される言語で所定のメッセージを出力する保留音を再生し、再生音を前記通話相手の通話端末に送信し、

前記判定手段は、前記保留音を再生中に前記通話相手から所定の入力を受け、前記通話相手の入力が第1の入力である場合、前記推定結果が正しいと判定し、前記通話相手の入力が第2の入力である場合、前記推定結果が正しくないと判定する処理システム。

6. 5に記載の処理システムにおいて、

前記判定手段は、所定の期限までに前記通話相手の入力がない場合、前記推定結果が正しくないと判定する処理システム。

7. 5又は6に記載の処理システムにおいて、

前記通話制御手段は、前記推定結果が複数の言語を含む場合、信頼度が最も高い言語の前記メッセージを再生し、前記通話相手の入力が第2の入力である場合、又は、所定の期限までに前記通話相手の入力がない場合には、次に信頼度が高い言語の前記メッセージを再生する処理システム。

8. コンピュータが、

言語推定対象の音声である対象音声を記録した対象音声データ、又は、前記対象音声の特徴を示す対象特徴量を取得する取得工程と、

音声を記録した音声データ又は前記音声の特徴を示す音声特徴量から前記音声の言語を推定する推定モデルと、前記対象音声データ又は前記対象特徴量とに基づき、前記対象音声の言語を推定する推定工程と、

前記推定工程での推定結果を出力する結果出力工程と、

前記推定結果の正誤判定を行う判定工程と、

前記判定工程で正しいと判定された前記推定結果と、前記対象音声データ又は前記対象特徴量とを、前記推定モデルを生成するための学習データとして出力する学習データ出力工程と、

を実行する処理方法。

9. コンピュータを、

言語推定対象の音声である対象音声を記録した対象音声データ、又は、前記対象音声の特徴を示す対象特徴量を取得する取得手段、

音声を記録した音声データ又は前記音声の特徴を示す音声特徴量から前記音声の言語を推定する推定モデルと、前記対象音声データ又は前記対象特徴量とに基づき、前記対象音声の言語を推定する推定手段、

前記推定手段による推定結果を出力する結果出力手段、

前記推定結果の正誤判定を行う判定手段、

前記判定手段により正しいと判定された前記推定結果と、前記対象音声データ又は前記対象特徴量とを、前記推定モデルを生成するための学習データとして出力する学習データ出力手段、

として機能させるプログラム。

[0120] この出願は、2018年12月11日に提出された日本出願特願2018-231461号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。

## 請求の範囲

- [請求項1] 言語推定対象の音声である対象音声を記録した対象音声データ、又は、前記対象音声の特徴を示す対象特徴量を取得する取得手段と、  
音声を記録した音声データ又は前記音声の特徴を示す音声特徴量から前記音声の言語を推定する推定モデルと、前記対象音声データ又は前記対象特徴量とに基づき、前記対象音声の言語を推定する推定手段と、  
前記推定手段による推定結果を出力する結果出力手段と、  
前記推定結果の正誤判定を行う判定手段と、  
前記判定手段により正しいと判定された前記推定結果と、前記対象音声データ又は前記対象特徴量とを、前記推定モデルを生成するための学習データとして出力する学習データ出力手段と、  
を有する処理システム。
- [請求項2] 請求項1に記載の処理システムにおいて、  
通話手段を有し、  
前記取得手段は、通話相手の音声を記録した前記対象音声データ、又は、前記通話相手の音声の特徴を示す前記対象特徴量を取得し、  
複数の言語各々の通訳担当者及び連絡先情報を登録した通訳者情報を参照し、前記推定結果で示される言語の前記通訳担当者を抽出する通訳担当者抽出手段と、  
抽出された前記通訳担当者を示す情報を出力する通訳担当者出力手段と、  
通話相手との通話を保留し、抽出された前記通訳担当者の中から指定された前記通訳担当者との通話を開始させる通話制御手段と、  
を有する処理システム。
- [請求項3] 請求項2に記載の処理システムにおいて、  
前記通話制御手段は、前記通話相手との通話を保留し、指定された前記通訳担当者との通話を開始させた後、所定の入力を受付けると、

前記通話相手及び前記通訳担当者を含む三者通話を開始し、

前記判定手段は、前記三者通話が開始されてからの経過時間が所定時間を超えた場合、前記推定結果が正しいと判定する処理システム。

[請求項4]

請求項2に記載の処理システムにおいて、

前記通話制御手段は、前記通話相手との通話を保留し、指定された前記通訳担当者との通話を開始させた後、前記対象音声データを再生し、再生音を前記通訳担当者の通話端末に送信し、その後所定の入力を受付けると、前記通話相手及び前記通訳担当者を含む三者通話を開始させ、

前記判定手段は、前記三者通話が開始されると、前記推定結果が正しいと判定する処理システム。

[請求項5]

請求項2から4のいずれか1項に記載の処理システムにおいて、

前記通話制御手段は、前記通話相手との通話を保留にしている間、前記推定結果で示される言語で所定のメッセージを出力する保留音を再生し、再生音を前記通話相手の通話端末に送信し、

前記判定手段は、前記保留音を再生中に前記通話相手から所定の入力を受付け、前記通話相手の入力が第1の入力である場合、前記推定結果が正しいと判定し、前記通話相手の入力が第2の入力である場合、前記推定結果が正しくないと判定する処理システム。

[請求項6]

請求項5に記載の処理システムにおいて、

前記判定手段は、所定の期限までに前記通話相手の入力がない場合、前記推定結果が正しくないと判定する処理システム。

[請求項7]

請求項5又は6に記載の処理システムにおいて、

前記通話制御手段は、前記推定結果が複数の言語を含む場合、信頼度が最も高い言語の前記メッセージを再生し、前記通話相手の入力が第2の入力である場合、又は、所定の期限までに前記通話相手の入力がない場合には、次に信頼度が高い言語の前記メッセージを再生する処理システム。

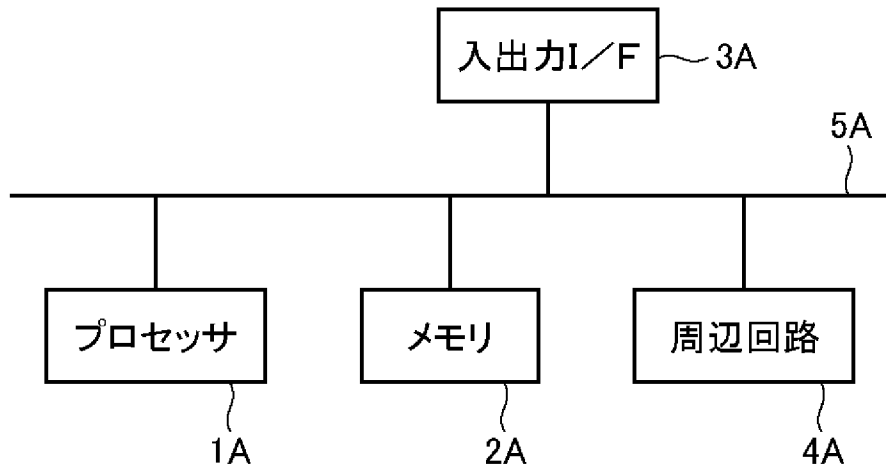
## [請求項8]

コンピュータが、  
言語推定対象の音声である対象音声を記録した対象音声データ、又は、前記対象音声の特徴を示す対象特徴量を取得する取得工程と、  
音声を記録した音声データ又は前記音声の特徴を示す音声特徴量から前記音声の言語を推定する推定モデルと、前記対象音声データ又は前記対象特徴量とに基づき、前記対象音声の言語を推定する推定工程と、  
前記推定工程での推定結果を出力する結果出力工程と、  
前記推定結果の正誤判定を行う判定工程と、  
前記判定工程で正しいと判定された前記推定結果と、前記対象音声データ又は前記対象特徴量とを、前記推定モデルを生成するための学習データとして出力する学習データ出力工程と、  
を実行する処理方法。

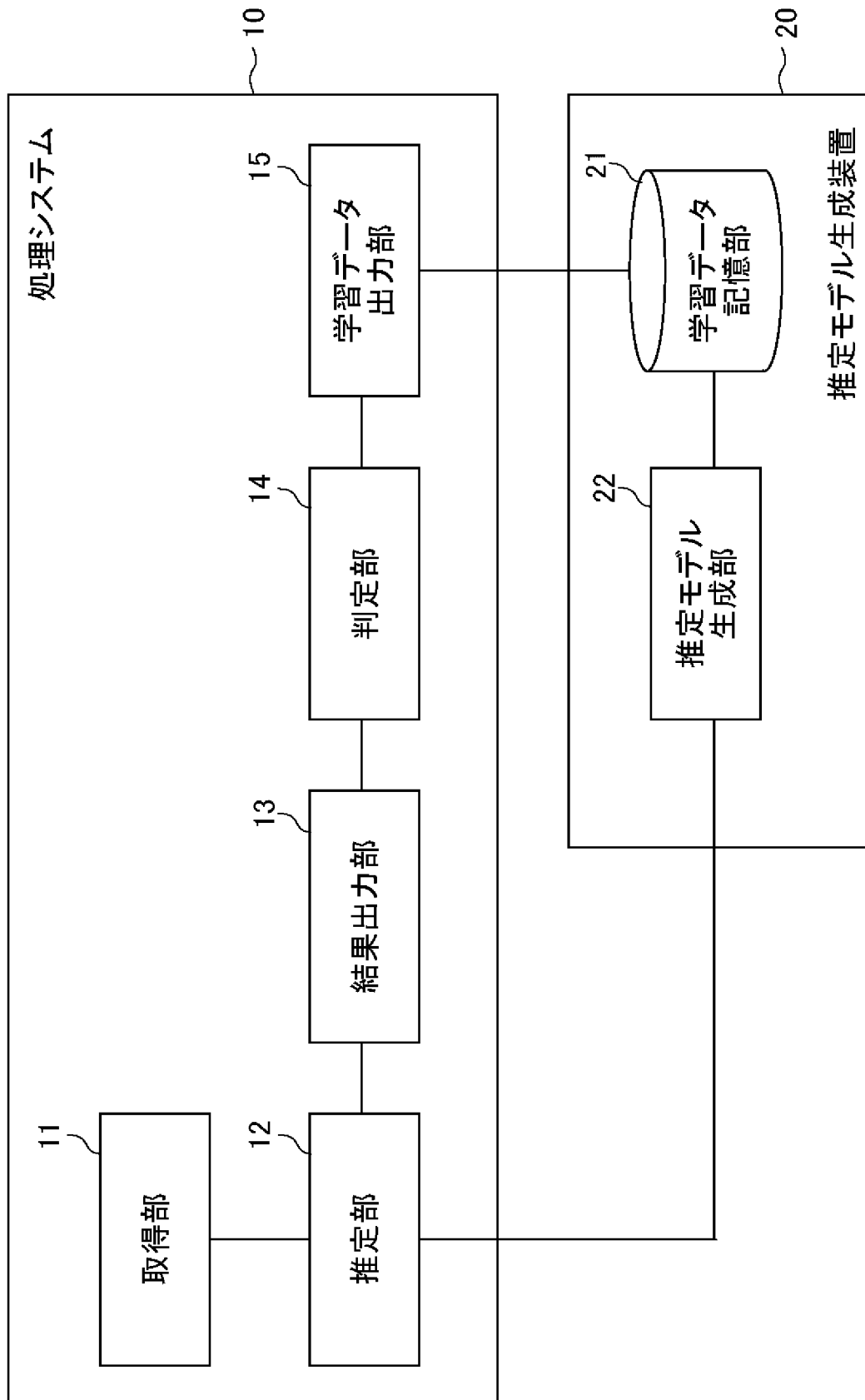
## [請求項9]

コンピュータを、  
言語推定対象の音声である対象音声を記録した対象音声データ、又は、前記対象音声の特徴を示す対象特徴量を取得する取得手段、  
音声を記録した音声データ又は前記音声の特徴を示す音声特徴量から前記音声の言語を推定する推定モデルと、前記対象音声データ又は前記対象特徴量とに基づき、前記対象音声の言語を推定する推定手段、  
、  
前記推定手段による推定結果を出力する結果出力手段、  
前記推定結果の正誤判定を行う判定手段、  
前記判定手段により正しいと判定された前記推定結果と、前記対象音声データ又は前記対象特徴量とを、前記推定モデルを生成するための学習データとして出力する学習データ出力手段、  
として機能させるプログラム。

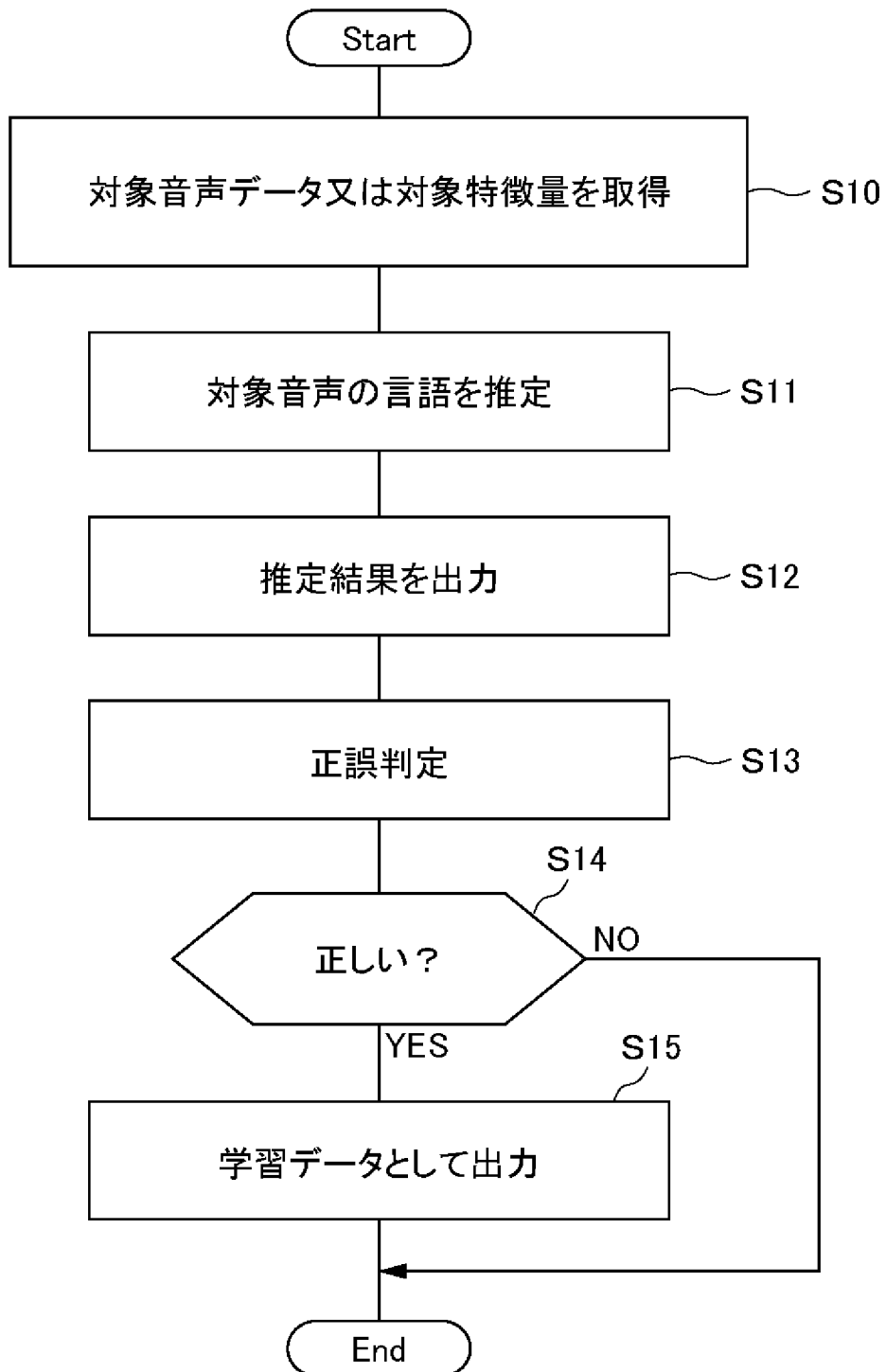
[図1]



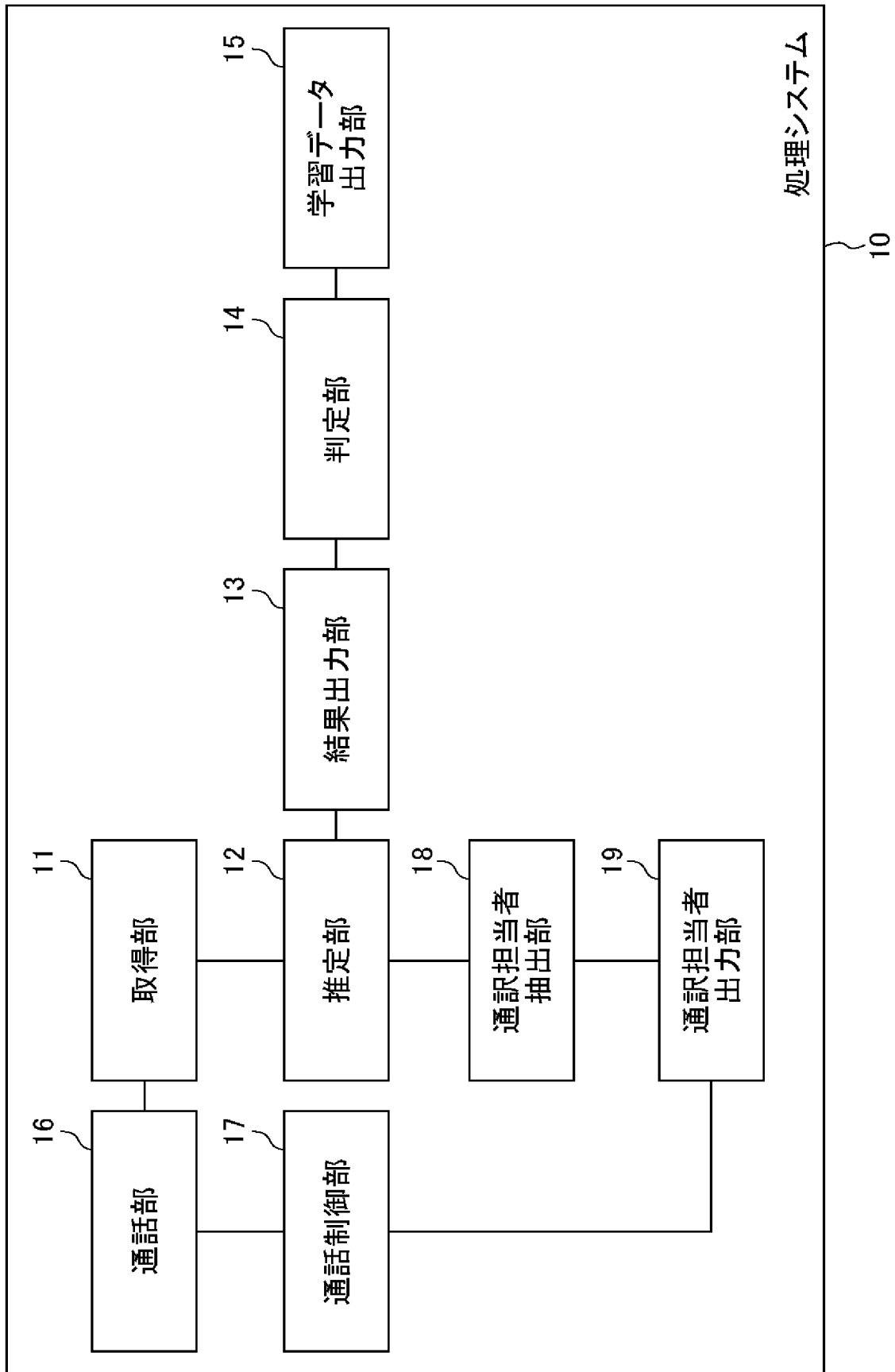
[図2]



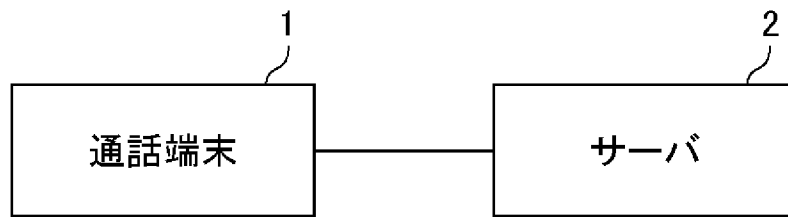
[図3]



[図4]



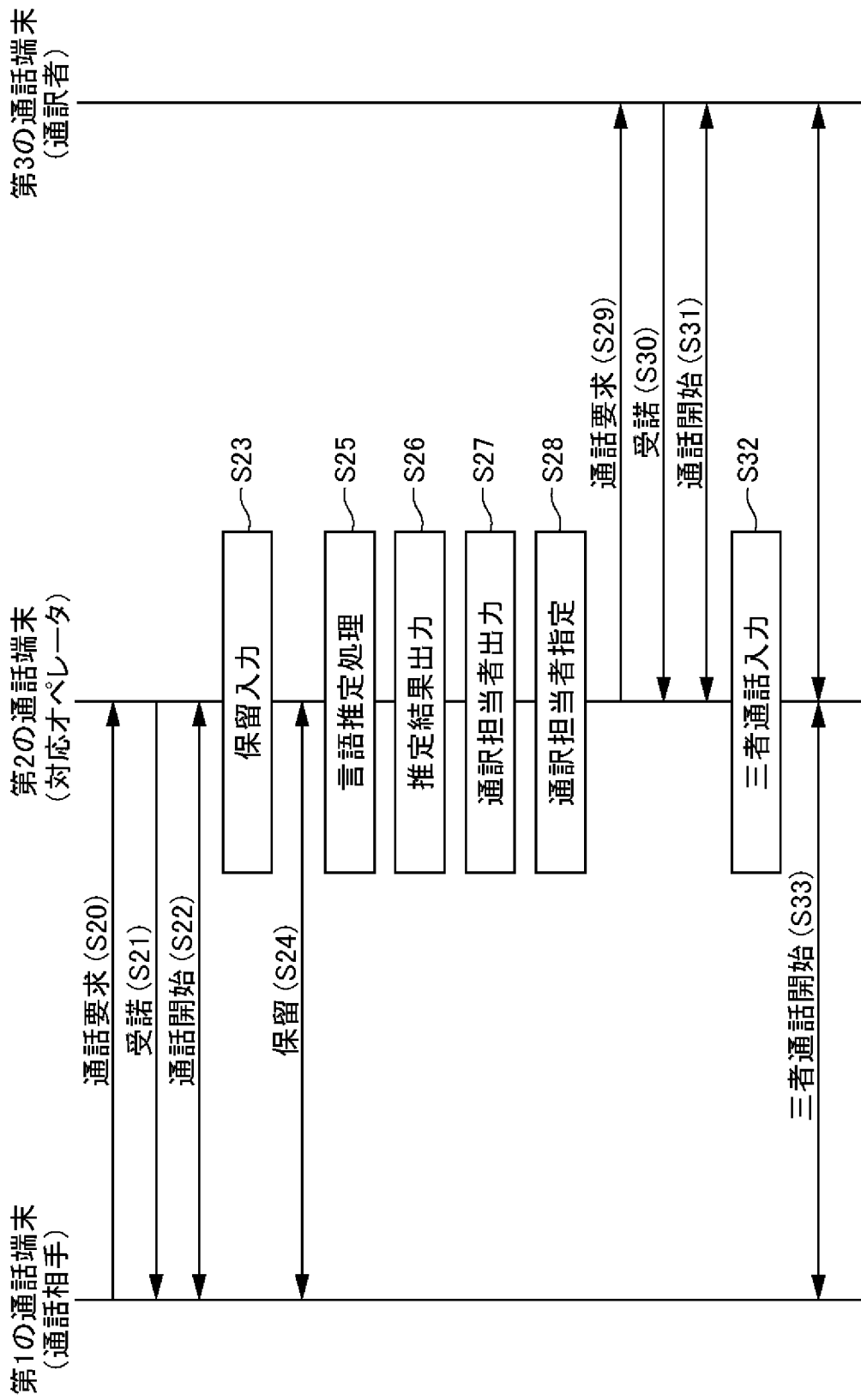
[図5]



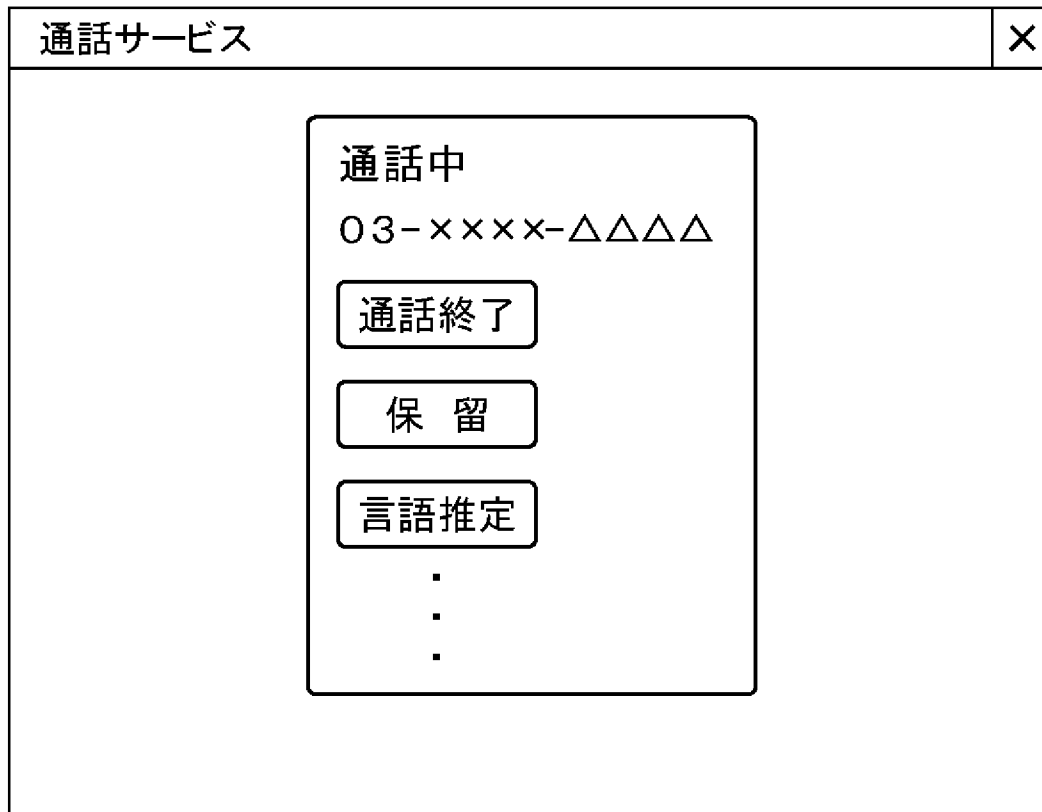
[図6]

| 通訳担当者 | 言語  | 連絡先情報 |
|-------|-----|-------|
| 日本太郎  | 中国語 | ***** |
| .     | .   | .     |
| .     | .   | .     |
| .     | .   | .     |
| .     | .   | .     |

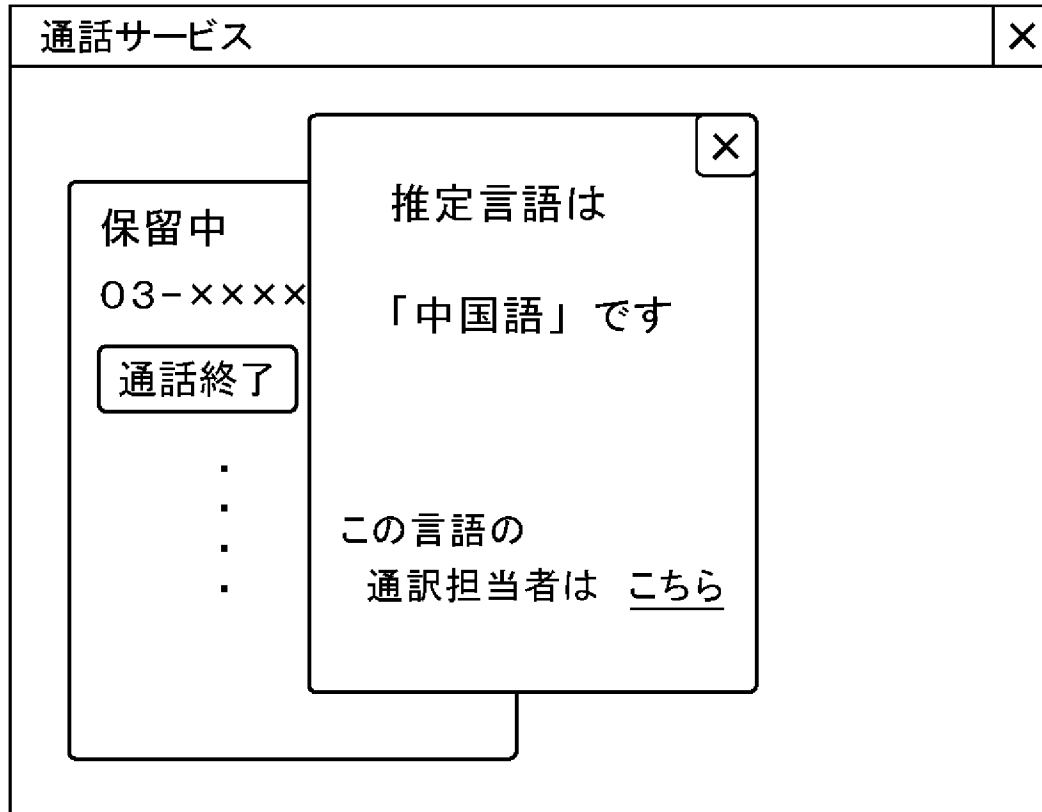
[図7]



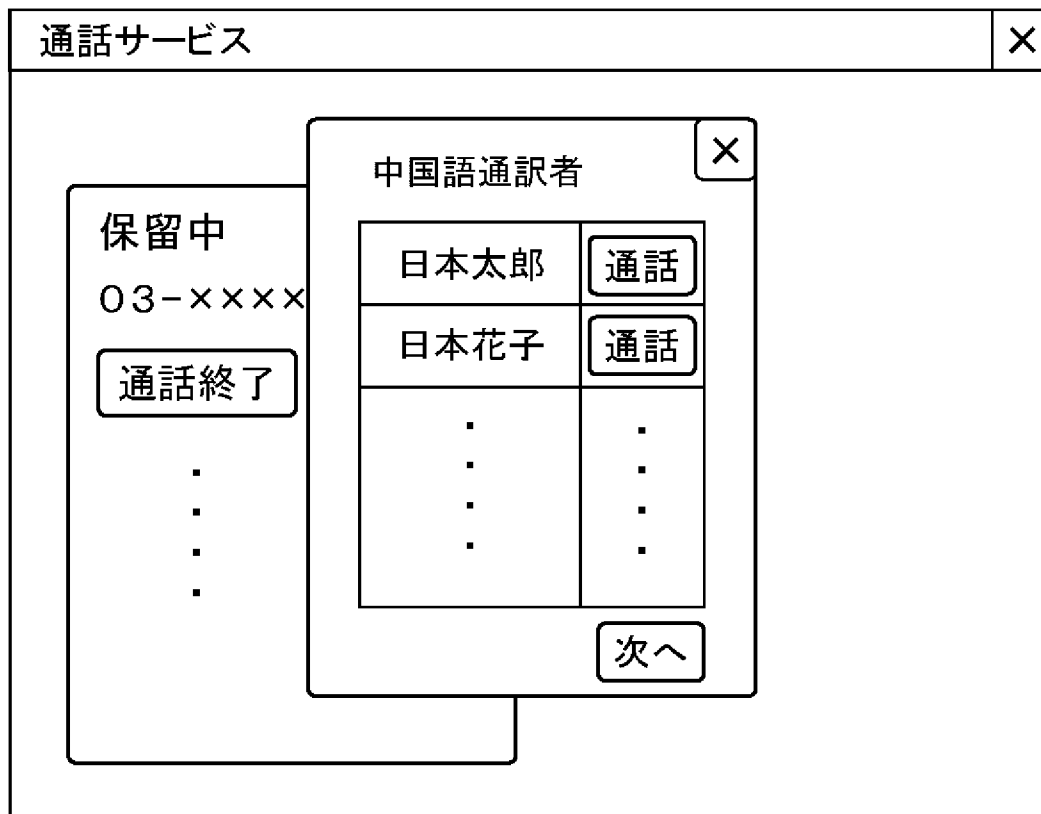
[図8]



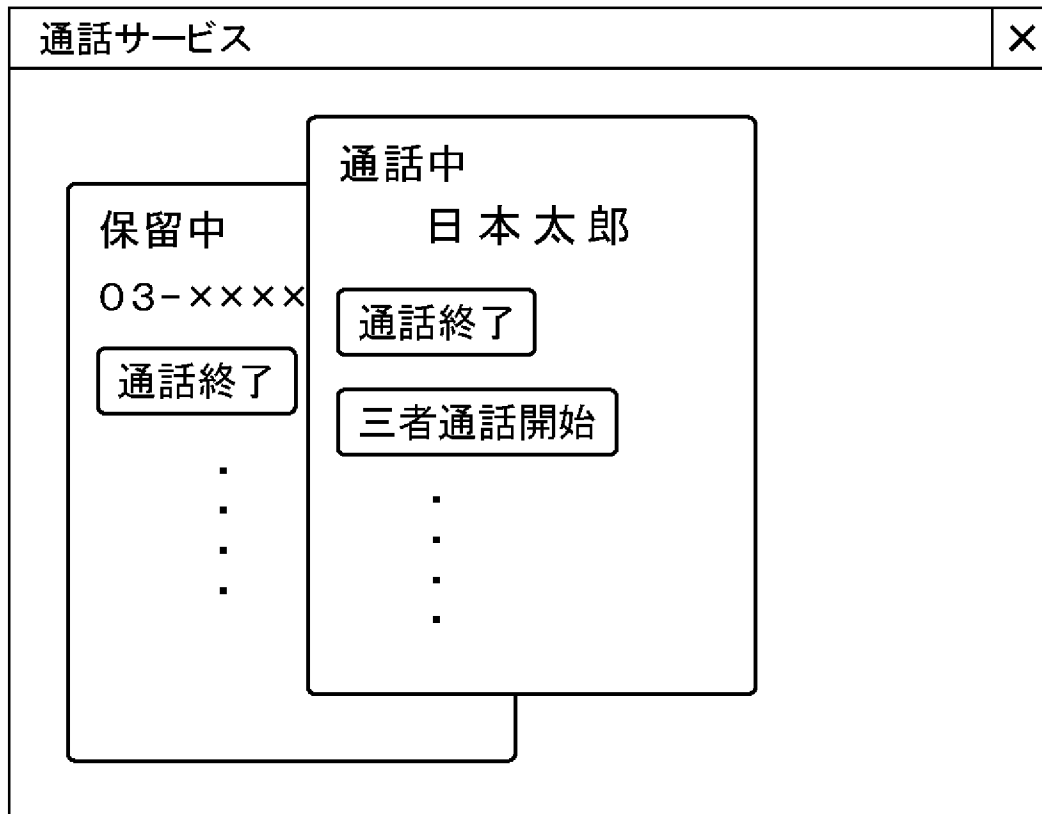
[図9]



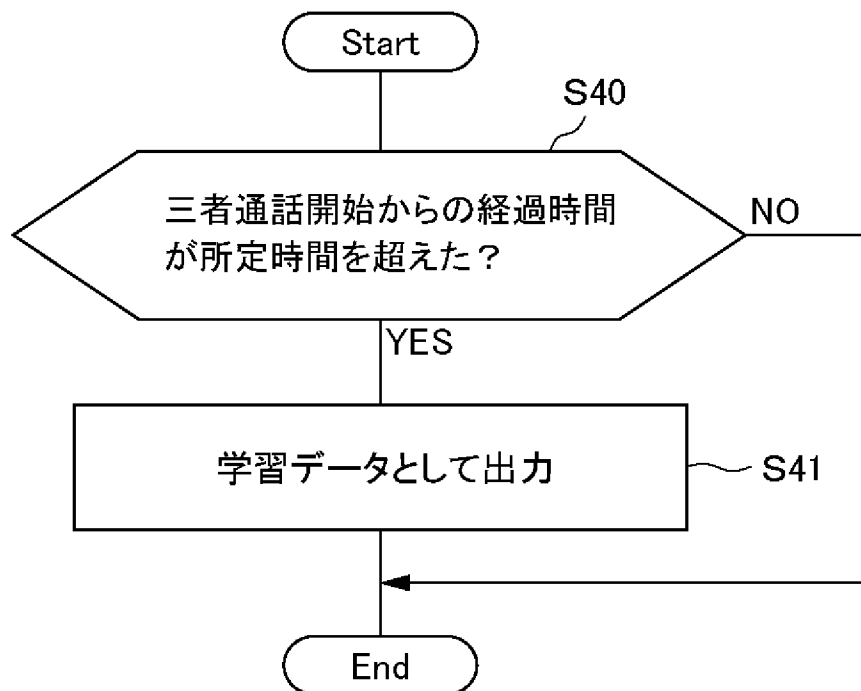
[図10]



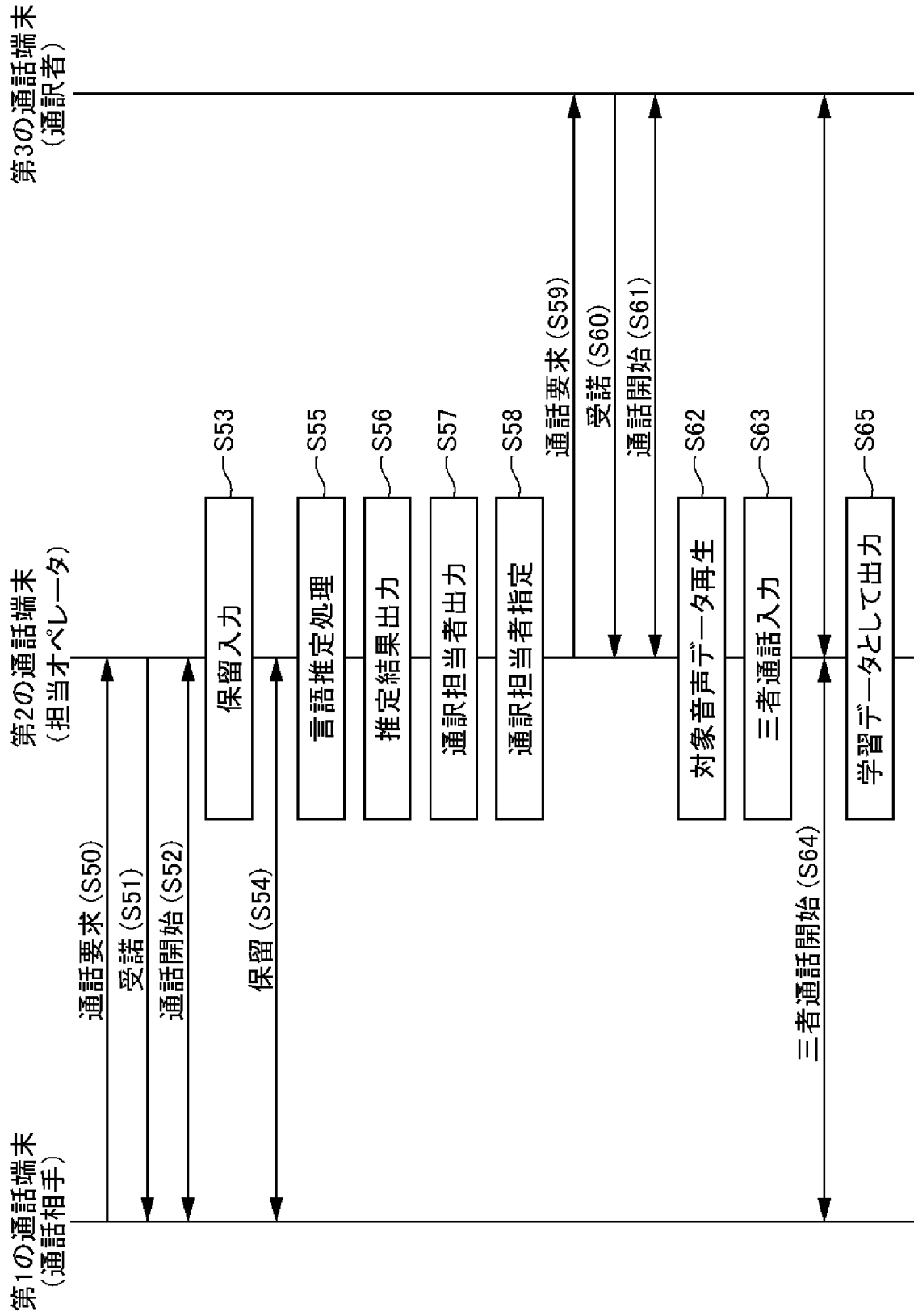
[図11]



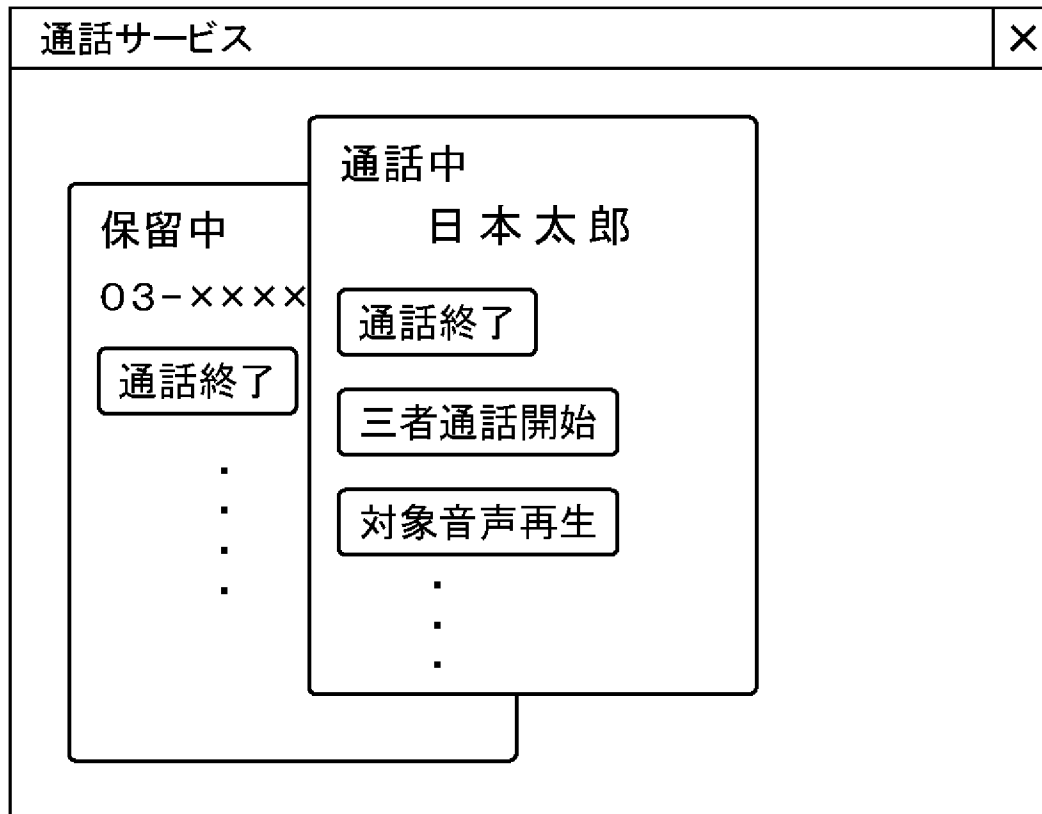
[図12]



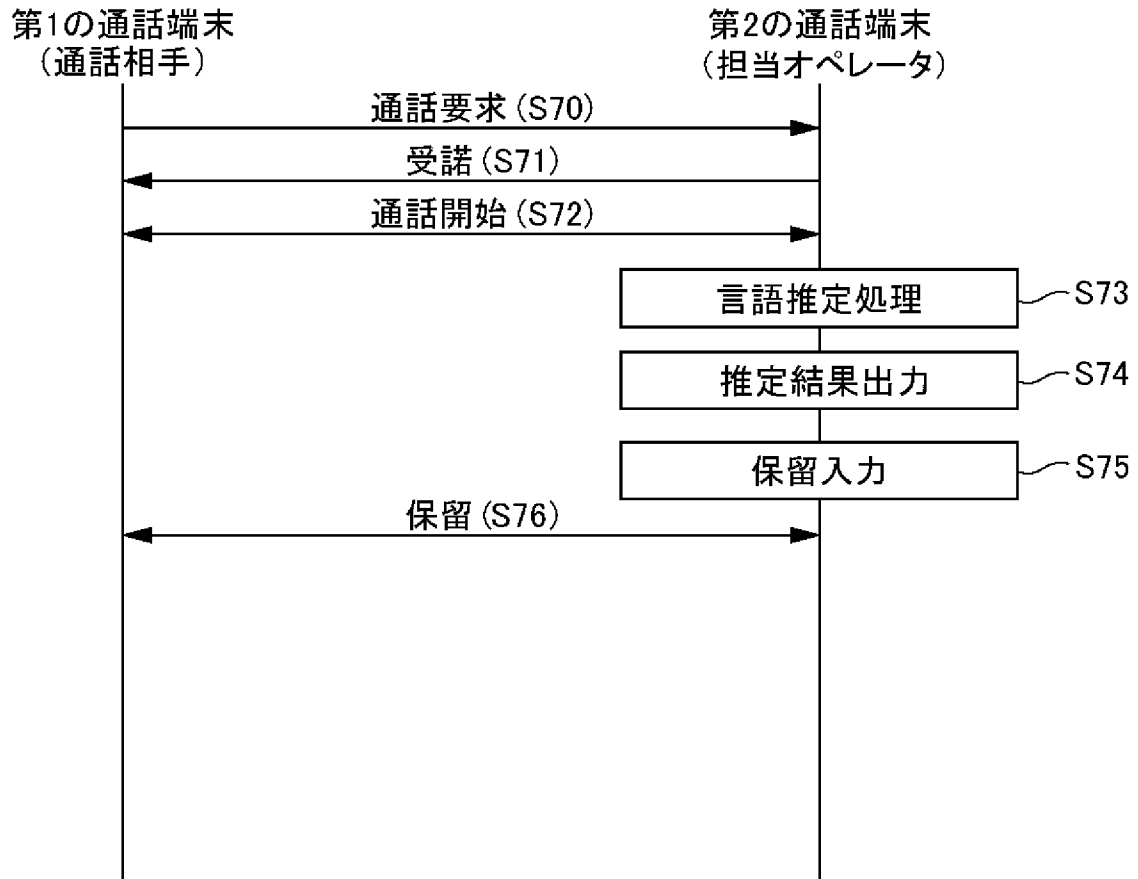
[図 13]



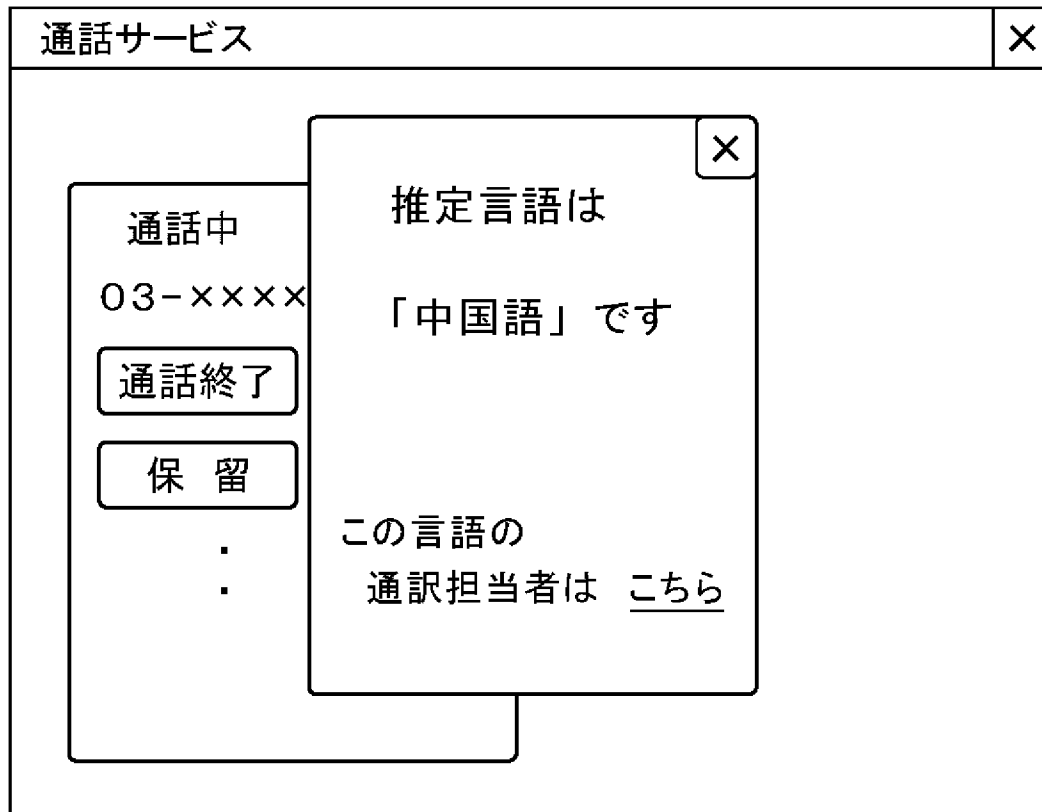
[図14]



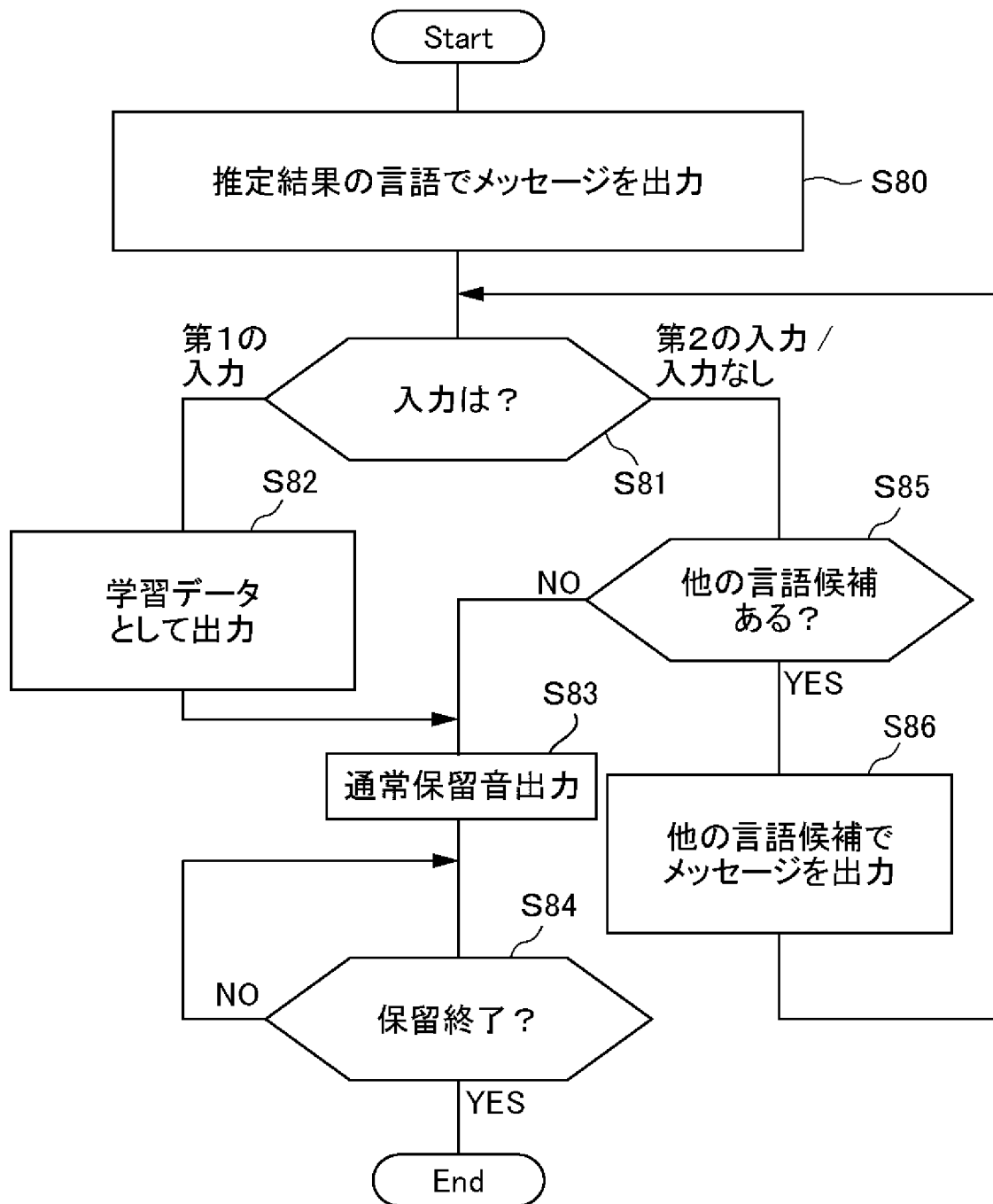
[図15]



[図16]



[図17]



**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/JP2019/036062

**A. CLASSIFICATION OF SUBJECT MATTER**

Int.Cl. G10L15/06 (2013.01) i, G10L15/00 (2013.01) i, G10L15/10 (2006.01) i, H04M3/56 (2006.01) 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. G10L15/06, G10L15/00, G10L15/10, H04M3/56

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         | JP 2012-103554 A (ADVANCED TELECOMMUNICATION RESEARCH INSTITUTE INTERNATIONAL) 31 May 2012, paragraphs [0016], [0021]<br>(Family: none) | 1-9                   |
| A         | JP 2018-128575 A (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 16 August 2018, paragraphs [0047]-[0060]<br>(Family: none)                | 1-9                   |
| A         | JP 2018-128576 A (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 16 August 2018, paragraphs [0035], [0048]-[0064]<br>(Family: none)        | 1-9                   |

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  | “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  |
| “E” earlier application or patent but published on or after the international filing date   | “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   |
| “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) | “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 |
| “O” document referring to an oral disclosure, use, exhibition or other means  | “&” document member of the same patent family  |
| “P” document published prior to the international filing date but later than the priority date claimed  |  |

|   |  |
|---|--|
| Date of the actual completion of the international search<br>21.11.2019 | Date of mailing of the international search report<br>03.12.2019 |
|---|--|

|  |   |
|--|---|
| Name and mailing address of the ISA/<br>Japan Patent Office<br>3-4-3, Kasumigaseki, Chiyoda-ku,<br>Tokyo 100-8915, Japan | Authorized officer<br><br>Telephone No. |
|--|---|

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/JP2019/036062

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages               | Relevant to claim No. |
|-----------|--|-----------------------|
| A         | JP 2018-174416 A (NEC CORPORATION) 08 November 2018, paragraphs [0008]-[0010]<br>(Family: none)  | 1-9                   |
| A         | JP 2010-021692 A (METEM CORPORATION) 28 January 2010, paragraphs [0008]-[0011]<br>(Family: none) | 1-9                   |

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. G10L15/06(2013.01)i, G10L15/00(2013.01)i, G10L15/10(2006.01)i, H04M3/56(2006.01)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. G10L15/06, G10L15/00, G10L15/10, H04M3/56

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<br>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<br>請求項の番号 |
|-----------------|---|----------------|
| A               | JP 2012-103554 A (株式会社国際電気通信基礎技術研究所)<br>2012.05.31, [0016],[0021] (ファミリーなし) | 1-9            |
| A               | JP 2018-128575 A (日本電信電話株式会社) 2018.08.16,<br>[0047]-[0060] (ファミリーなし)        | 1-9            |
| A               | JP 2018-128576 A (日本電信電話株式会社) 2018.08.16,<br>[0035],[0048]-[0064] (ファミリーなし) | 1-9            |

☑ 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  | の日の後に公表された文献   |
| 「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                                 | 「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     |
| 「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                         | 「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                     |
| 「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) | 「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 |
| 「O」口頭による開示、使用、展示等に言及する文献                                      | 「&」同一パテントファミリー文献   |
| 「P」国際出願日前で、かつ優先権の主張の基礎となる出願                                   |  |

国際調査を完了した日

21.11.2019

国際調査報告の発送日

03.12.2019

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)

岩田 淳

電話番号 03-3581-1101 内線 3591

5Z

4052

| C (続き) . 関連すると認められる文献 |  |                |
|-----------------------|--|----------------|
| 引用文献の<br>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<br>請求項の番号 |
| A                     | JP 2018-174416 A (日本電気株式会社) 2018. 11. 08, [0008]-[0010]<br>(ファミリーなし) | 1-9            |
| A                     | JP 2010-021692 A (株式会社メテム) 2010. 01. 28, [0008]-[0011] (フ<br>ァミリーなし) | 1-9            |