

(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
2018年1月25日(25.01.2018)



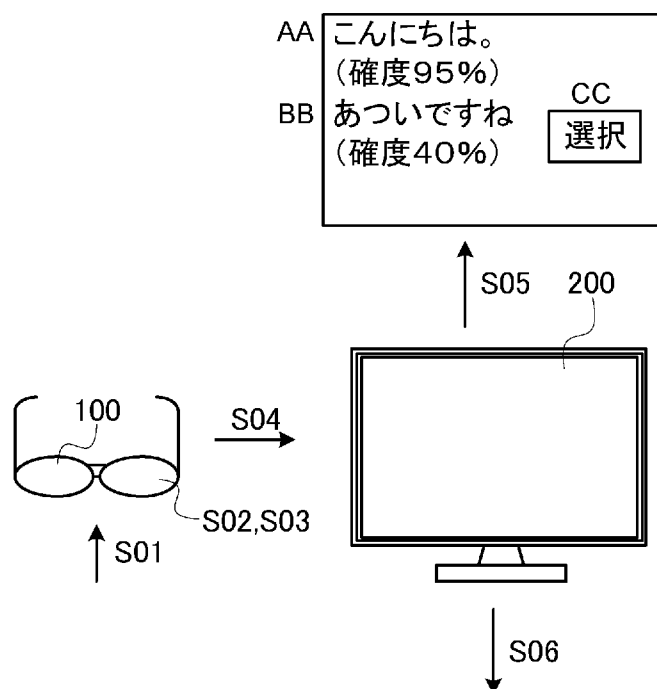
(10) 国際公開番号

WO 2018/015989 A1

- (51) 国際特許分類:
G10L 15/22 (2006.01)
- (21) 国際出願番号: PCT/JP2016/071110
- (22) 国際出願日: 2016年7月19日(19.07.2016)
- (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: 株式会社 オプティム (OPTIM CORPORATION) [JP/JP]; 〒8400047 佐賀県佐賀市与賀町4番18号 Saga (JP).
- (72) 発明者: 菅谷 俊二(SUGAYA Shunji); 〒1050022 東京都港区海岸1丁目2番20号 汐留ビルディング 21F 株式会社オプティム内 Tokyo (JP).
- (74) 代理人: 小木 智彦(KOGI Tomohiko); 〒8800804 宮城県宮崎市宮田町11-24 黒木ビル1F Miyazaki (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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,

(54) Title: SPEECH RECOGNITION SYSTEM, SPEECH RECOGNITION METHOD, AND PROGRAM

(54) 発明の名称: 音声認識システム、音声認識方法及びプログラム



AA Hello. (Accuracy is 95%.)
BB Hot, isn't it? (Accuracy is 40%.)
CC Selection

(57) Abstract: [Problem] The purpose of the present invention is to provide a speech recognition system, a speech recognition method, and a program that enable a user to listen to and check the results of the conversion of speech into text as to whether any portion of the desired text has been incorrectly converted. [Solution] Speech is recorded, the recorded speech is recognized, the recognized speech is converted into text, the converted text is displayed, a selection of a desired portion of the displayed text is received, and the speech corresponding to the text of the portion for which the selection

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))

was received is output.

(57) 要約: 【課題】 音声テキストに変換された結果に対して、所望するテキストの一部が間違っ変換がなされていないか否かを、ユーザが音声を聞いてチェックすることが可能な音声認識システム、音声認識方法及びプログラムを提供することを目的とする。 【解決手段】 音声を集音し、前記集音した音声を認識し、前記認識した音声をテキストに変換し、前記変換したテキストを表示し、前記表示したテキストから所望する一部分について、選択を受け付け、前記選択を受け付けた一部分のテキストに対応する音声を出力する。

明 細 書

発明の名称：音声認識システム、音声認識方法及びプログラム 技術分野

[0001] 本発明は、音声テキストに変換する音声認識システム、音声認識方法及びプログラムに関する。

背景技術

[0002] 従来、ユーザがマイク等の集音装置へ音声を入力し、この入力した音声をテキストに変換することが行われている。このような音声のテキスト変換は、例えば、ユーザによる文書入力、会議等の議事録作成等に用いられている。

[0003] このような議事録の作成として、会議の出席者の音声をマイクにより集音し、この音声の音声データをテキストに変換し、議事録として記録する構成が開示されている（特許文献1参照）。

先行技術文献

特許文献

[0004] 特許文献1：特開2013-222347号公報

発明の概要

発明が解決しようとする課題

[0005] しかしながら、特許文献1の構成では、音声をテキストに変換した後に、入力された音声と変換したテキストとが一致しているか否かを、チェックすることが困難であった。

[0006] 本発明の目的は、音声テキストに変換された結果に対して、所望するテキストの一部分が間違っ変換がなされていないか否かを、ユーザが音声を聞いてチェックすることが可能な音声認識システム、音声認識方法及び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

[0007] 本発明では、以下のような解決手段を提供する。

- [0008] 第1の特徴に係る発明は、音声を集音する集音手段と、
前記集音した音声进行認識する音声認識手段と、
前記認識した音声をテキストに変換するテキスト変換手段と、
前記変換したテキストを表示する表示手段と、
前記表示したテキストから所望する一部分について、選択を受け付ける部分選択受付手段と、
前記選択を受け付けた一部分のテキストに対応する音声を出力する音声出力手段と、
を備えることを特徴とする音声認識システムを提供する。
- [0009] 第1の特徴に係る発明によれば、音声を集音し、前記集音した音声を認識し、前記認識した音声をテキストに変換し、前記変換したテキストを表示し、前記表示したテキストから所望する一部分について、選択を受け付け、前記選択を受け付けた一部分のテキストに対応する音声を出力する。
- [0010] ここで、第1の特徴に係る発明は、音声認識システムのカテゴリであるが、方法又はプログラム等の他のカテゴリにおいても、そのカテゴリに応じた同様の作用・効果を發揮する。
- [0011] 第2の特徴に係る発明は、前記選択を受け付けた一部分のテキストを、前記出力した音声の結果に基づいて、変更を受け付けるテキスト変更受付手段と、
を備えることを特徴とする第1の特徴に係る発明である音声認識システムを提供する。
- [0012] 第2の特徴に係る発明によれば、第1の特徴に係る発明である音声認識システムは、前記選択を受け付けた一部分のテキストを、前記出力した音声の結果に基づいて、変更を受け付ける。
- [0013] 第3の特徴に係る発明は、前記変更を受け付けたテキストを、先に音声認識した音声と対応付けて記憶する記憶手段と、
を備えることを特徴とする第2の特徴に係る発明である音声認識システムを提供する。

[0014] 第3の特徴に係る発明によれば、第2の特徴に係る発明である音声認識システムは、前記変更を受け付けたテキストを、先に音声認識した音声と対応付けて記憶する。

[0015] 第4の特徴に係る発明は、前記変換したテキストの正解率を表示する確度表示手段と、

を備えることを特徴とする第1の特徴に係る発明である音声認識システムを提供する。

[0016] 第4の特徴に係る発明によれば、第1の特徴に係る発明である音声認識システムは、前記変換したテキストの正解率を表示する。

[0017] 第5の特徴に係る発明は、音声を集音するステップと、

前記集音した音声を認識するステップと、

前記認識した音声をテキストに変換するステップと、

前記変換したテキストを表示するステップと、

前記表示したテキストから所望する一部分について、選択を受け付けるステップと、

前記選択を受け付けた一部分のテキストに対応する音声を出力するステップと、

を備えることを特徴とする音声認識方法を提供する。

[0018] 第6の特徴に係る発明は、音声認識システムに、

音声を集音するステップ、

前記集音した音声を認識するステップ、

前記認識した音声をテキストに変換するステップ、

前記変換したテキストを表示するステップ、

前記表示したテキストから所望する一部分について、選択を受け付けるステップ、

前記選択を受け付けた一部分のテキストに対応する音声を出力するステップ、

を実行させるためのプログラムを提供する。

発明の効果

[0019] 本発明によれば、音声テキストに変換された結果に対して、所望するテキストの一部分が間違っただけの変換がなされていないか否かを、ユーザが音声を聞いてチェックすることが可能な音声認識システム、音声認識方法及びプログラムを提供することを目的とする。

図面の簡単な説明

[0020] [図1]図1は、音声認識システム1の概要を示す図である。

[図2]図2は、音声認識システム1の全体構成図である。

[図3]図3は、ユーザ端末100、表示端末200の機能ブロック図である。

[図4]図4は、ユーザ端末100、表示端末200が実行する音声認識処理を示す図である。

[図5]図5は、ユーザ端末100、表示端末200が実行する音声認識処理を示す図である。

[図6]図6は、ユーザ端末100、表示端末200が実行する音声認識処理を示す図である。

[図7]図7は、表示端末200が表示するテキスト表示画面の一例を示す図である。

[図8]図8は、表示端末200が表示するテキスト表示画面の一例を示す図である。

[図9]図9は、表示端末200が表示するテキスト表示画面の一例を示す図である。

[図10]図10は、表示端末200が表示するテキスト表示画面の一例を示す図である。

発明を実施するための形態

[0021] 以下、本発明を実施するための最良の形態について、図を参照しながら説明する。なお、これはあくまでも一例であって、本発明の技術的範囲はこれに限られるものではない。

[0022] [音声認識システム1の概要]

本発明の好適な実施形態の概要について、図１に基づいて説明する。図１は、本発明の好適な実施形態である音声認識システム１の概要を説明するための図である。音声認識システム１は、ユーザ端末１００、表示端末２００から構成される。

[0023] なお、図１において、ユーザ端末１００又は表示端末２００は、１つに限らず複数であってもよい。また、ユーザ端末１００又は表示端末２００は、実在する装置に限らず仮想的な装置であってもよい。また、後述する各処理は、ユーザ端末１００又は表示端末２００のいずれか又は双方により実現されてもよい。

[0024] ユーザ端末１００は、表示端末２００とデータ通信可能な端末装置である。ユーザ端末１００は、例えば、携帯電話、携帯情報端末、タブレット端末、パーソナルコンピュータに加え、ネットブック端末、スレート端末、電子書籍端末、携帯型音楽プレーヤ等の電化製品や、スマートグラス、ヘッドマウントディスプレイ等のウェアラブル端末や、その他の物品である。

[0025] 表示端末２００は、ユーザ端末１００とデータ通信可能な端末装置である。表示端末２００は、ユーザ端末１００と同様の端末装置である。

[0026] ユーザ端末１００は、ユーザからの音声入力を、マイク等の集音装置により集音する（ステップＳ０１）。

[0027] ユーザ端末１００は、集音した音声を音声認識する（ステップＳ０２）。ユーザ端末１００は、音声認識することにより、ユーザが発した音声の内容を認識する。

[0028] ユーザ端末１００は、認識した音声を、テキストに変換する（ステップＳ０３）。

[0029] ユーザ端末１００は、認識した音声の音声データと、テキストのテキストデータとを、表示端末２００に送信する（ステップＳ０４）。

[0030] 表示端末２００は、音声データとテキストデータとを受信し、このテキストデータに基づいて、テキストを自身の表示部に表示する（ステップＳ０５）。なお、音声データを変換したテキストの正解率をテキストとともに表示

する構成であってもよい。

[0031] 表示端末200は、表示したテキストの一部分に対して、ユーザからの選択を受け付け、選択を受け付けた一部分のテキストに対応する音声を出力する（ステップS06）。なお、選択を受け付けた一部分のテキストに限らず、表示する全てのテキストに対応する音声を出力する構成であってもよい。

[0032] また、表示端末200は、出力した音声の結果に基づいて、選択を受け付けた一部分のテキストに対して、ユーザからの変更を受け付ける構成であってもよい。この場合、ユーザからの変更を受け付けることにより、ユーザが変更したテキストを、表示する構成であればよい。また、表示端末200は、この変更を受け付けたテキストを、先にユーザ端末100により音声認識した音声と対応付けて記憶する構成であってもよい。

[0033] 以上が、音声認識システム1の概要である。

[0034] [音声認識システム1のシステム構成]

図2に基づいて、本発明の好適な実施形態である音声認識システム1のシステム構成について説明する。図2は、本発明の好適な実施形態である音声認識システム1のシステム構成を示す図である。音声認識システム1は、ユーザ端末100、表示端末200、公衆回線網（インターネット網や、第3、第4世代通信網等）5から構成される。なお、ユーザ端末100又は表示端末200は、1つに限らず、複数であってもよい。また、ユーザ端末100又は表示端末200は、実在する装置に限らず、仮想的な装置であってもよい。また、後述する各処理は、ユーザ端末100又は表示端末200のいずれか又は双方により実現されてもよい。

[0035] ユーザ端末100は、後述の機能を備えた上述した端末装置である。

[0036] 表示端末200は、後述の機能を備えた上述した端末装置である。

[0037] [各機能の説明]

図3に基づいて、本発明の好適な実施形態である音声認識システム1の機能について説明する。図3は、ユーザ端末100、表示端末200の機能ブロック図を示す図である。

[0038] ユーザ端末100は、制御部110として、CPU (Central Processing Unit)、RAM (Random Access Memory)、ROM (Read Only Memory)等を備え、通信部120として、他の機器と通信可能にするためのデバイス、例えば、IEEE802.11に準拠したWiFi (Wireless Fidelity) 対応デバイスを備える。また、ユーザ端末100は、記憶部130として、ハードディスクや半導体メモリ、記録媒体、メモリカード等によるデータのストレージ部を備える。また、ユーザ端末100は、入出力部140として、制御部110で制御したデータや画像を出力表示する表示部や、ユーザからの入力を受け付けるタッチパネルやキーボード、マウス等の入力部や、音声を集音するマイク等の集音デバイスや、音声をテキスト変換する変換デバイスや、テキストの正解率を判断する確度判断デバイス等を備える。

[0039] ユーザ端末100において、制御部110が所定のプログラムを読み込むことにより、通信部120と協働して、データ送受信モジュール150を実現する。また、ユーザ端末100において、制御部110が所定のプログラムを読み込むことにより、記憶部130と協働して、記憶モジュール160を実現する。また、ユーザ端末100において、制御部110が所定のプログラムを読み込むことにより、入出力部140と協働して、集音モジュール170、音声認識モジュール171、変換モジュール172を実現する。

[0040] 表示端末200は、ユーザ端末100と同様に、制御部210として、CPU、RAM、ROM等を備え、通信部220として、他の機器と通信可能にするためのWiFi 対応デバイス等を備え、記憶部230として、データのストレージ部を備え、入出力部240として、表示部や入力部や確度判断デバイス等を備える。

[0041] 表示端末200において、制御部210が所定のプログラムを読み込むことにより、通信部220と協働して、データ送受信モジュール250を実現する。また、表示端末200において、制御部210が所定のプログラムを

読み込むことにより、記憶部 230 と協働して、記憶モジュール 260 を実現する。また、表示端末 200 において、制御部 210 が所定のプログラムを読み込むことにより、入出力部 240 と協働して、入力受付モジュール 270、表示モジュール 271、音声出力モジュール 272、変換モジュール 273 を実現する。

[0042] [音声認識処理]

図 4、図 5 及び図 6 に基づいて、音声認識システム 1 が実行する音声認識処理について説明する。図 4、図 5 及び図 6 は、ユーザ端末 100、表示端末 200 が実行する音声認識処理のフローチャートを示す図である。上述した各装置のモジュールが実行する処理について、本処理に併せて説明する。

[0043] 集音モジュール 170 は、ユーザが発した音声を集音し、ユーザからの音声の入力を受け付ける（ステップ S10）。集音モジュール 170 は、例えば、ユーザが発した文章や文節等の音声を集音する。

[0044] 音声認識モジュール 171 は、入力を受け付けた音声を、音声認識する（ステップ S11）。ステップ S11 において、音声認識モジュール 171 は、入力を受け付けた音声の内容を音声認識する。音声認識モジュール 171 は、例えば、既存の音声認識技術やパターン認識技術等により、音声を音声認識する。

[0045] なお、音声認識モジュール 171 は、上述した構成に限らず、他の構成により音声認識を行ってもよい。

[0046] 変換モジュール 172 は、音声認識した内容を、テキストに変換する（ステップ S12）。ステップ S12 において、変換モジュール 172 は、例えば、音声認識した内容が、「こんにちはあついです」である場合、テキストとして、「こんにちは。暑いですね。」と変換する。変換モジュール 172 は、テキストの変換として、例えば、前後の関係、文節の区切り位置、音声の抑揚等に基づいて、音声認識した内容をテキストに変換する。また、ステップ S12 において、変換モジュール 172 は、変換したテキストの候補が複数ある場合、複数のテキストを其々変換する。

- [0047] なお、変換モジュール１７２は、上述した構成に限らず、他の構成によりテキストに変換してもよい。
- [0048] 変換モジュール１７２は、変換したテキストの確度を算出する（ステップＳ１３）。ステップＳ１３において、変換モジュール１７２は、変換したテキストに複数の候補が存在する場合、前後の関係、文節の区切り位置、音声の抑揚、語彙の登場頻度等に基づいて、複数のテキストにおける其々の確度を算出する。テキストの確度とは、音声認識した内容と、変換したテキストとの正解率である。例えば、登場頻度の多い語彙は、確度が高く、登場頻度の少ない語彙は、確度が低くなるといった構成や、各語彙や文節毎に確度を数値化して設定しておき、この数値に基づいて確度を算出する構成である。変換モジュール１７２は、文章、文節又は単語の其々に対して確度を算出する。
- [0049] なお、変換モジュール１７２は、上述した構成に限らず、他の構成により確度を算出してもよい。また、変換モジュール１７２は、複数の候補が存在しない場合であっても、確度を算出してもよい。
- [0050] 記憶モジュール１６０は、入力を受け付けた音声と、この音声を変換したテキストと、このテキストの確度とを対応付けた変換情報を記憶する（ステップＳ１４）。記憶モジュール１６０は、変換したテキストが複数存在する場合、其々記憶する。すなわち、一の音声と、複数の変換テキストと、各テキストの確度とを対応付けた変換情報を記憶する。
- [0051] なお、記憶モジュール１６０は、変換したテキストが複数存在しない場合、一の音声と、一の変換したテキストと、このテキストの確度とを対応付けた変換情報を記憶する。
- [0052] また、ステップＳ１４の処理は、省略可能である。すなわち、ユーザ端末１００は、変換情報を記憶せずに、後述する処理を実行する構成であってもよい。
- [0053] データ送受信モジュール１５０は、変換情報を表示端末２００に送信する（ステップＳ１５）。

- [0054] データ送受信モジュール250は、変換情報を受信する。入力受付モジュール270は、テキストを変更する確度の入力を受け付けたか否かを判断する（ステップS16）。ステップS16において、入力受付モジュール270は、後述するテキストの変更入力を受け付けるか否かの基準となる確度の入力を受け付けたか否かを判断する。例えば、入力受付モジュール270は、確度の入力として、50%、60%等の数値の入力を受け付けたか否かを判断する。確度の数値としては、%表記に限らず、割合、数値等適宜変更可能である。
- [0055] ステップS16において、入力受付モジュール270は、入力を受け付けていないと判断した場合（ステップS16 NO）、表示モジュール271は、受信した変換情報に基づいて、テキストと確度とを表示する（ステップS17）。
- [0056] 図7は、ステップS17において、表示モジュール271が表示するテキスト表示画面の一例を示す図である。図7において、表示モジュール271は、テキスト表示領域300、選択アイコン310を表示する。テキスト表示領域300は、変換情報に基づいたテキストを表示する領域である。テキスト表示領域300は、テキストを文節毎又は文章毎に区切って表示する。図7において、表示モジュール271は、テキスト表示領域300に、「こんにちは」と「あ、ついですね」を其々区切って表示する。選択アイコン310は、ユーザからの入力を受け付け、該当するテキストへの選択入力を受け付ける。音声出力モジュール272は、選択アイコン310への入力を受け付けたテキストを、変換情報に基づいて音声出力する。図7において、表示モジュール271は、テキストの確度に関わらず、選択アイコン310を各テキストに表示させる。
- [0057] 一方、ステップS16において、入力受付モジュール270は、入力を受け付けたと判断した場合（ステップS16 YES）、表示モジュール271は、受信した変換情報に基づいて、テキストと確度とを表示する（ステップS18）。

[0058] 図8は、ステップS18において、表示モジュール271が表示するテキスト表示画面の一例を示す図である。図8において、所定の確度を50%の入力を受け付けた場合での一例を示す。図8において、表示モジュール271は、図7と同様に、テキスト表示領域300、選択アイコン310を表示する。図7と図8との相違点は、表示モジュール271は、図7では、全ての確度のテキストに対して選択アイコン310を表示するのに対して、図8では、入力を受け付けた確度以下の確度のテキストに対してのみ選択アイコン310を表示する点が相違している。すなわち、図8において、表示モジュール271は、確度が所定の値よりも高い「95%」である「こんにちは」に対しては、選択アイコン310を非表示とし、確度が所定の値よりも低い「40%」である「あついですね」に対しては、選択アイコン310を表示する。

[0059] 入力受付モジュール270は、テキストの選択入力を受け付けたか否かを判断する（ステップS19）。ステップS19において、入力受付モジュール270は、例えば、上述した選択アイコン310への入力操作を受け付けたか否か、表示するテキストへのタップ操作を受け付けたか否か等により判断する。

[0060] ステップS19において、入力受付モジュール270は、受け付けていないと判断した場合（ステップS19 NO）、記憶モジュール260は、変換情報を記憶する（ステップS20）。ステップS20において、記憶モジュール260は、音声と、表示したテキストと、確度とを対応付けて記憶する。表示端末200は、ステップS20の処理を実行後、本処理を終了する。

[0061] 一方、ステップS19において、入力受付モジュール270は、受け付けたと判断した場合（ステップS19 YES）、音声出力モジュール272は、選択されたテキストを変換情報に基づいて、音声出力する（ステップS21）。

[0062] 入力受付モジュール270は、テキストの変更入力を受け付けたか否かを

判断する（ステップS 2 2）。ステップS 2 2において、入力受付モジュール2 7 0は、直接入力、音声入力、ジェスチャー入力等に基づいて、表示モジュール2 7 1が表示するテキストとは異なるテキストへ変更する入力を受け付けたか否かを判断する。

[0063] ステップS 2 2において、入力受付モジュール2 7 0は、テキストの変更入力を受け付けていないと判断した場合（ステップS 2 2 NO）、上述したステップS 2 0の処理を実行する。

[0064] 一方、ステップS 2 2において、入力受付モジュール2 7 0は、テキストの変更入力を受け付けたと判断した場合（ステップS 2 2 YES）、変換モジュール2 7 3は、変更後のテキストの確度を算出する（ステップS 2 3）。ステップS 2 3において、変換モジュール2 7 3は、上述したステップS 1 3の処理と同様の処理により、変更後のテキストの確度を算出する。

[0065] 変換モジュール2 7 3は、確度が所定の値以上であるか否かを判断する（ステップS 2 4）。ステップS 2 4において、変換モジュール2 7 3は、予め設定された値以上であるか否か、上述したステップS 1 6において受け付けた値以上であるか否か等を判断する。ステップS 2 4において、変換モジュール2 7 3は、所定の値以上であると判断した場合（ステップS 2 4 YES）、表示モジュール2 7 1は、変更後のテキストとその確度とを表示する（ステップS 2 5）。なお、ステップS 2 5において、表示モジュール2 7 1は、変更後のテキストのみを表示する構成であってもよい。

[0066] 図9は、表示モジュール2 7 1が表示する変更後のテキスト表示画面の一例を示す図である。図9において、表示モジュール2 7 1は、テキスト表示領域3 0 0を表示する。表示モジュール2 7 1は、テキスト表示領域3 0 0に、変更されなかったテキストと、変更されたテキストと、其々の確度とを表示する。すなわち、図9において、「こんにちは」のテキストは、変更されなかったテキストであり、「あ、ついですね（変更済）」は、変更されたテキストである。このように、表示モジュール2 7 1は、変更されたテキストには、変更済である旨のメッセージを付与して表示する。

- [0067] なお、表示モジュール２７１は、変更されたテキストのみをテキスト表示領域３００に表示する構成であってもよい。また、表示モジュール２７１は、変更済である旨のメッセージを付与せずに、変更されたテキストを表示する構成であってもよい。
- [0068] 記憶モジュール２６０は、変換情報に含まれる音声と、変更後のテキストと、確度とを対応付けた変更後変換情報を記憶する（ステップＳ２６）。ステップＳ２６において、記憶モジュール２６０は、変更後変換情報だけでなく、変更していないテキストの音声と、このテキストと、確度とを対応付けて記憶する。すなわち、ステップＳ２６において、記憶モジュール２６０は、表示モジュール２７１がステップＳ２５において表示した全てのテキストと、このテキストの音声と、確度とを対応付けて記憶する。表示端末２００は、ステップＳ２６の処理を実行した後、本処理を終了する。
- [0069] 一方、ステップＳ２４において、変換モジュール２７３は、所定の値以上ではないと判断した場合（ステップＳ２４　ＮＯ）、表示モジュール２７１は、変更後のテキストと、異なる変換候補の通知とを表示する（ステップＳ２７）。
- [0070] 図１０は、表示モジュール２７１が表示する変更後のテキスト表示画面の一例を示す図である。図１０において、確度が５０％以上でない場合における一例を示す。図１０において、表示モジュール２７１は、テキスト表示領域３００、選択アイコン３１０、変換候補通知表示領域４００を表示する。表示モジュール２７１は、テキスト表示領域３００に、図９と同様に、変更されなかったテキストと、変更されたテキストと、其々の確度とを表示する。ここで、表示モジュール２７１は、確度が所定の値以上でないことを示す旨の通知を、変換候補通知表示領域４００に表示する。表示モジュール２７１は、この変換候補通知表示領域４００に、最も確度が高い変換候補のテキストを表示するとともに、この変換候補のテキストの確度を表示する。すなわち、図１０において、「暑いですね（変更済）」の確度が「４０％」であることから、所定の確度を下回っているため、今回テキストの変更を受け付

けたものの、他により確度が高い変換候補が存在する旨の通知として「変更していますが、今回は、「あ、ついですね。」の可能性が高いです（確度70%）。」を、対応するテキストの一部に重畳させて表示する。

[0071] なお、この通知の表示態様は、適宜変更可能であり、変換候補のテキストのみが表示されてもよいし、所定の確度以上の複数の変換候補のテキストが表示されてもよい。また、確度は、表示されていなくてもよい。

[0072] 入力受付モジュール270は、テキストの変更入力を受け付けたか否かを判断する（ステップS28）。ステップS28の処理は、上述したステップS22の処理と同様である。

[0073] ステップS28において、入力受付モジュール270は、受け付けていないと判断した場合（ステップS28 NO）、上述したステップS26の処理を実行する。

[0074] 一方、ステップS28において、入力受付モジュール270は、受け付けたと判断した場合（ステップS28 YES）、上述したステップS23の処理を実行する。

[0075] 以上が、音声認識処理である。

[0076] なお、上述した音声認識処理において、ユーザ端末100が音声認識しているが、表示端末200が音声認識し、テキスト変換する構成であってもよい。例えば、ユーザ端末100が受け付けた音声を、音声データとして、表示端末200に送信し、表示端末200が音声認識等の処理を実行する構成であってもよい。また、サーバやその他の構成が、音声認識、テキスト変換、確度算出等を実行する構成であってもよい。

[0077] また、ユーザ端末100又は表示端末200が記憶した変更後のテキスト及び音声、変更前のテキスト及び音声等に基づいて、確度を算出する構成であってもよい。例えば、以前に変更したテキストや変更しなかったテキストのデータに基づいて、今回変更するテキストの変換候補を判断する構成であってもよい。

[0078] 上述した手段、機能は、コンピュータ（CPU、情報処理装置、各種端末

を含む)が、所定のプログラムを読み込んで、実行することによって実現される。プログラムは、例えば、フレキシブルディスク、CD (CD-ROM など)、DVD (DVD-ROM、DVD-RAM など)等のコンピュータ読取可能な記録媒体に記録された形態で提供される。この場合、コンピュータはその記録媒体からプログラムを読み取って内部記憶装置又は外部記憶装置に転送し記憶して実行する。また、そのプログラムを、例えば、磁気ディスク、光ディスク、光磁気ディスク等の記憶装置(記録媒体)に予め記録しておき、その記憶装置から通信回線を介してコンピュータに提供するようにしてもよい。

[0079] 以上、本発明の実施形態について説明したが、本発明は上述したこれらの実施形態に限るものではない。また、本発明の実施形態に記載された効果は、本発明から生じる最も好適な効果を列挙したに過ぎず、本発明による効果は、本発明の実施形態に記載されたものに限定されるものではない。

符号の説明

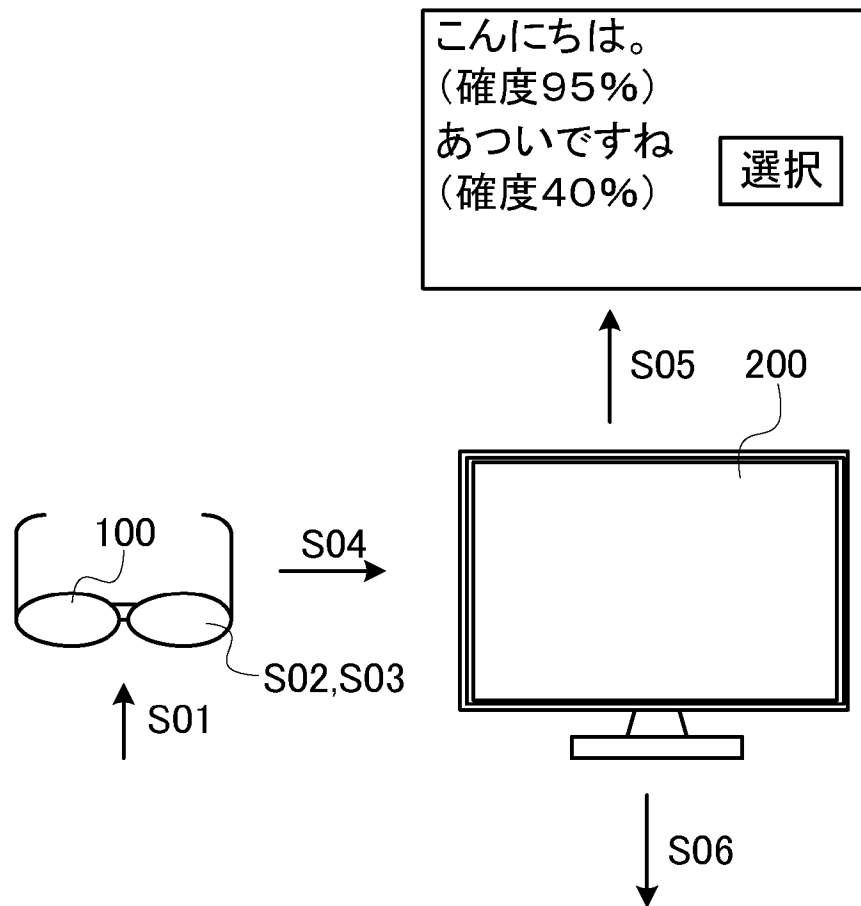
[0080] 1 音声認識システム、100 ユーザ端末、200 表示端末

請求の範囲

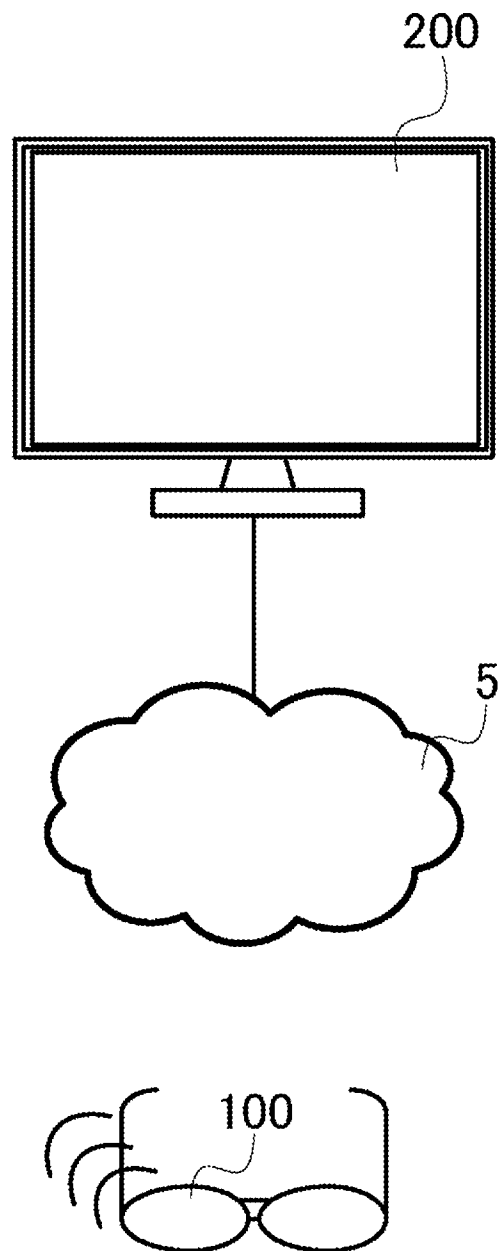
- [請求項1] 音声を集音する集音手段と、
前記集音した音声を認識する音声認識手段と、
前記認識した音声をテキストに変換するテキスト変換手段と、
前記変換したテキストを表示する表示手段と、
前記表示したテキストから所望する一部分について、選択を受け付ける部分選択受付手段と、
前記選択を受け付けた一部分のテキストに対応する音声を入力する音声出力手段と、
を備えることを特徴とする音声認識システム。
- [請求項2] 前記選択を受け付けた一部分のテキストを、前記出力した音声の結果に基づいて、変更を受け付けるテキスト変更受付手段と、
を備えることを特徴とする請求項1に記載の音声認識システム。
- [請求項3] 前記変更を受け付けたテキストを、先に音声認識した音声と対応付けて記憶する記憶手段と、
を備えることを特徴とする請求項2に記載の音声認識システム。
- [請求項4] 前記変換したテキストの正解率を表示する確度表示手段と、
を備えることを特徴とする請求項1に記載の音声認識システム。
- [請求項5] 音声を集音するステップと、
前記集音した音声を認識するステップと、
前記認識した音声をテキストに変換するステップと、
前記変換したテキストを表示するステップと、
前記表示したテキストから所望する一部分について、選択を受け付けるステップと、
前記選択を受け付けた一部分のテキストに対応する音声を入力するステップと、
を備えることを特徴とする音声認識方法。
- [請求項6] 音声認識システムに、

音声を集音するステップ、
前記集音した音声を認識するステップ、
前記認識した音声をテキストに変換するステップ、
前記変換したテキストを表示するステップ、
前記表示したテキストから所望する一部分について、選択を受け付けるステップ、
前記選択を受け付けた一部分のテキストに対応する音声を出力するステップ、
を実行させるためのプログラム。

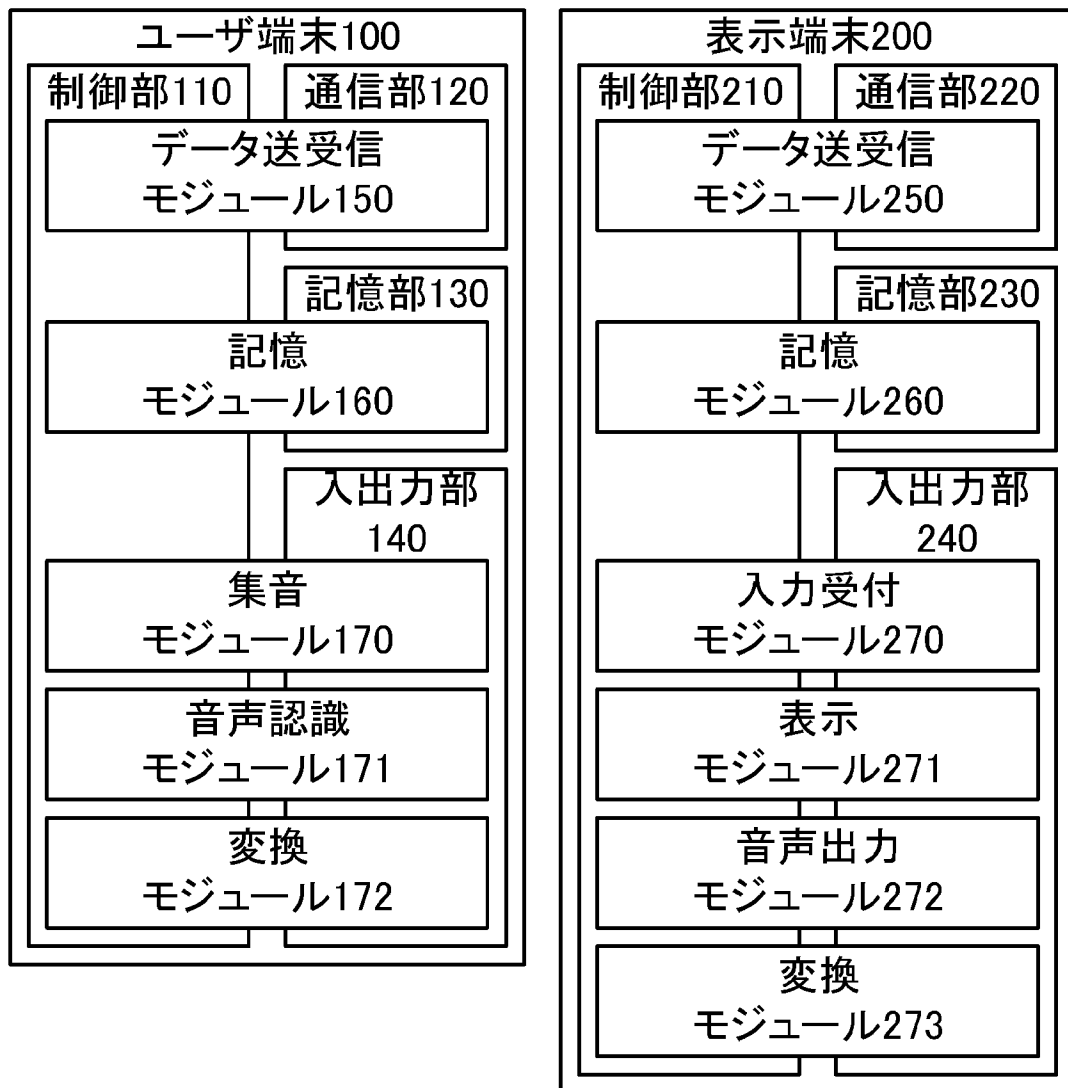
[図1]



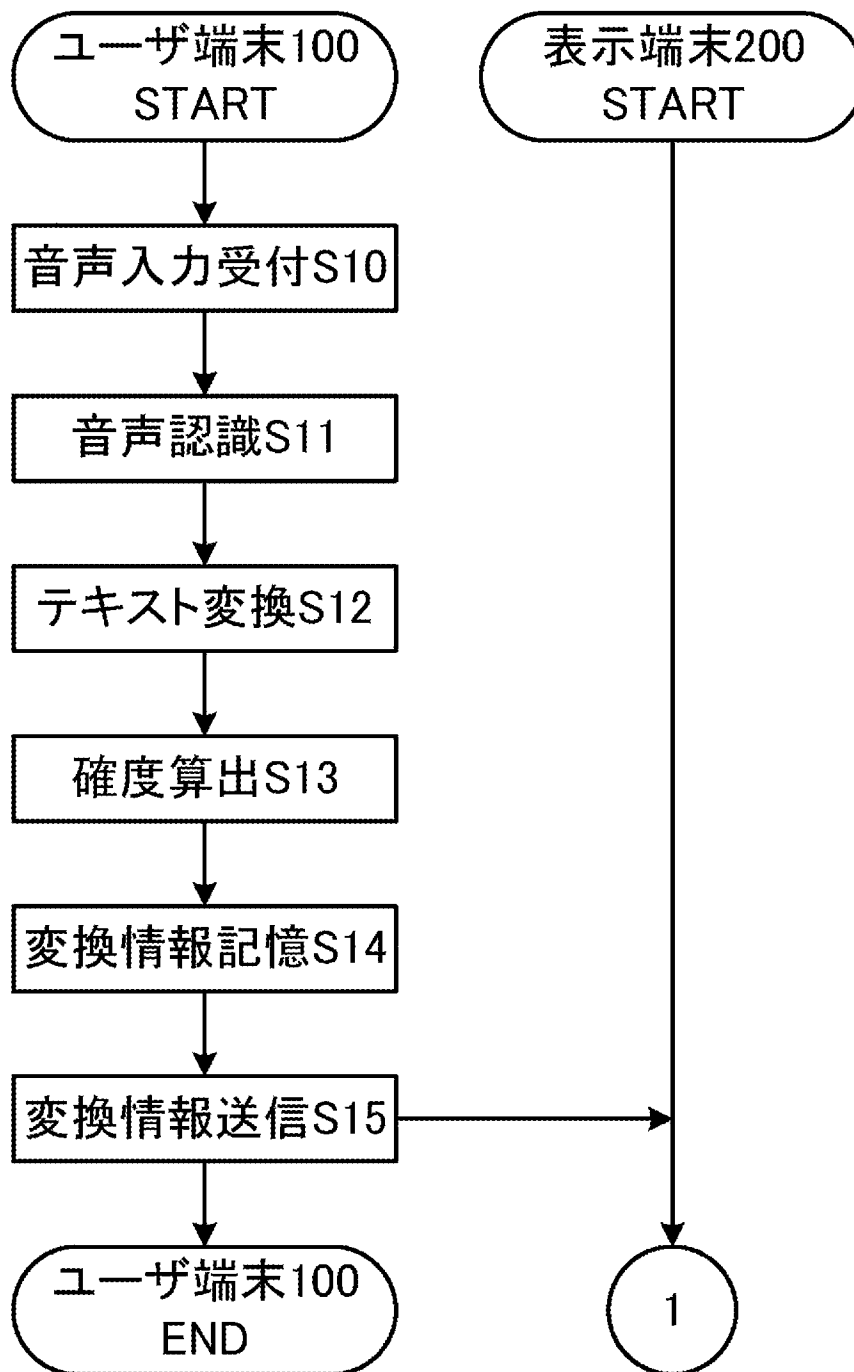
[図2]



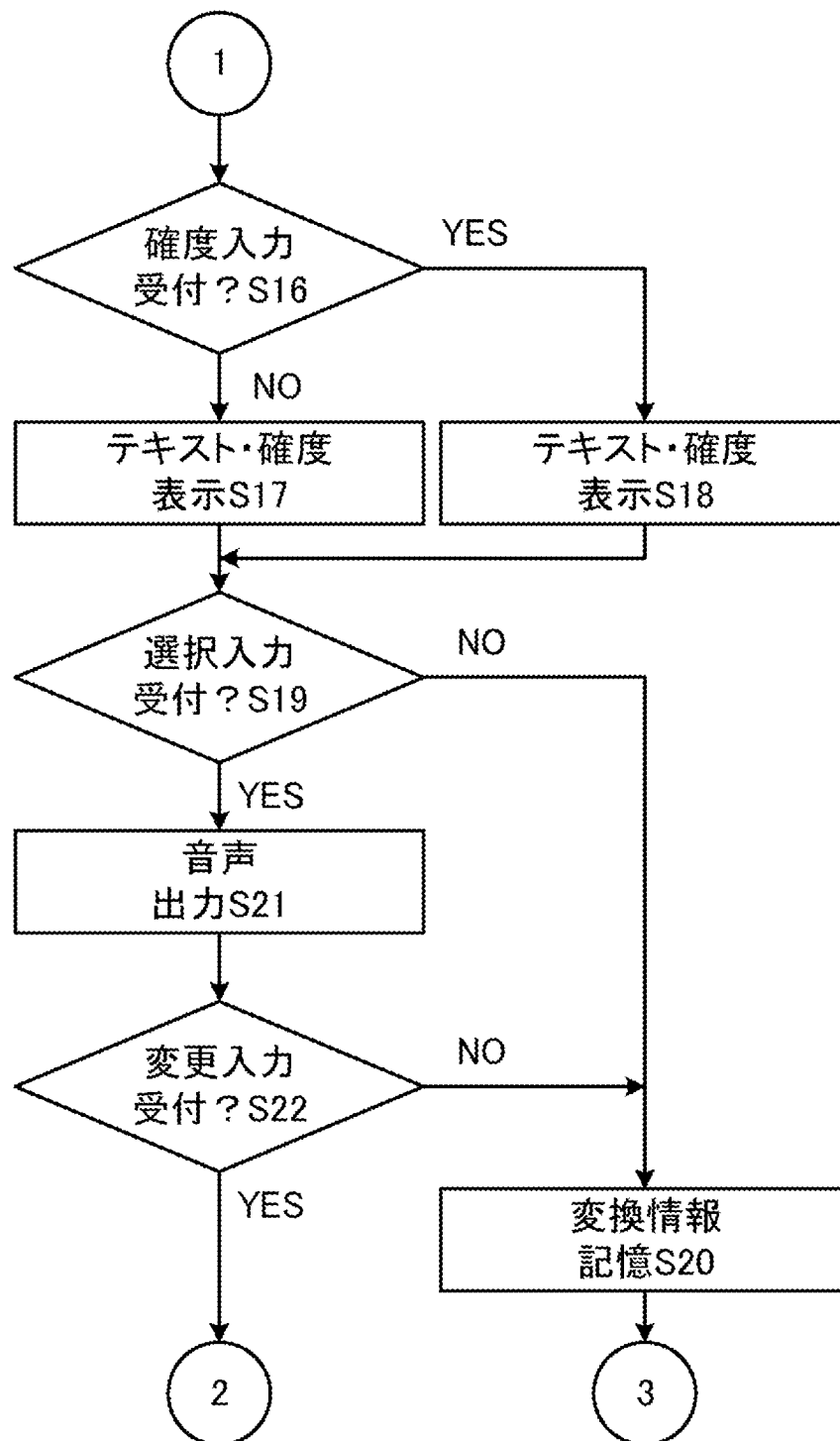
[図3]



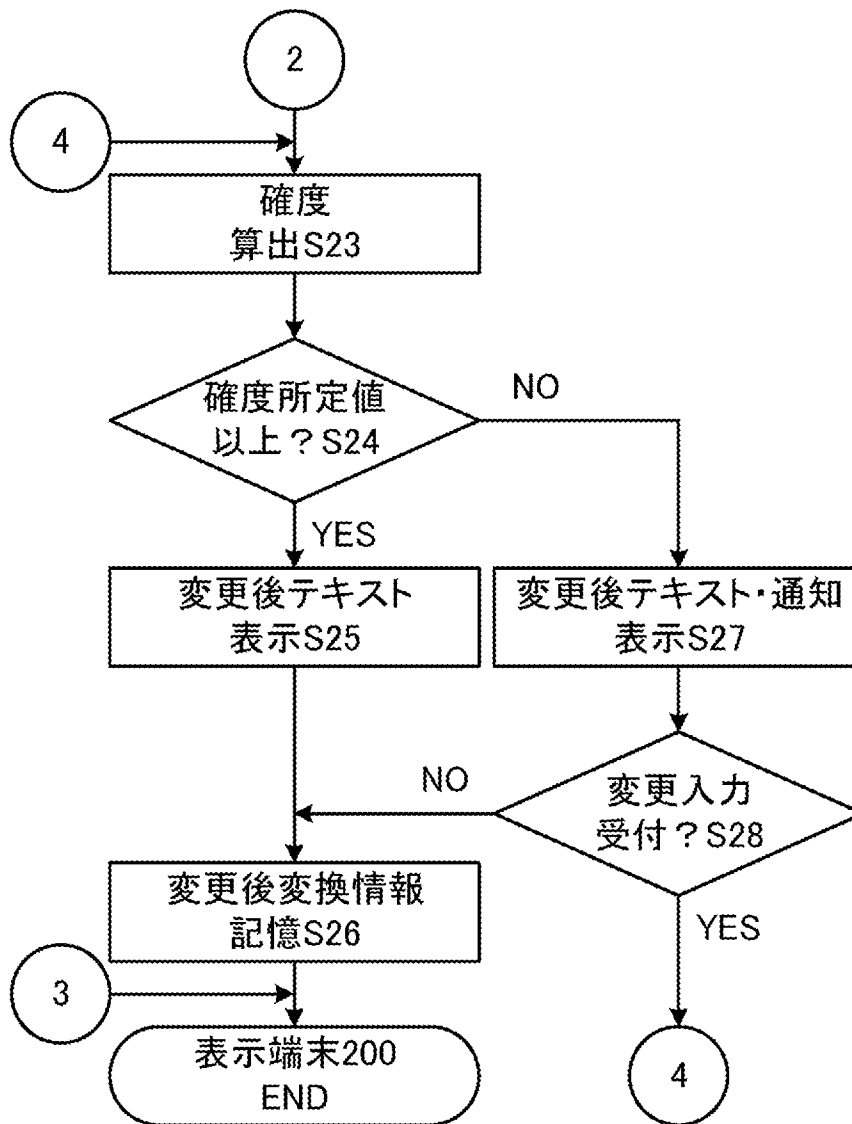
[図4]



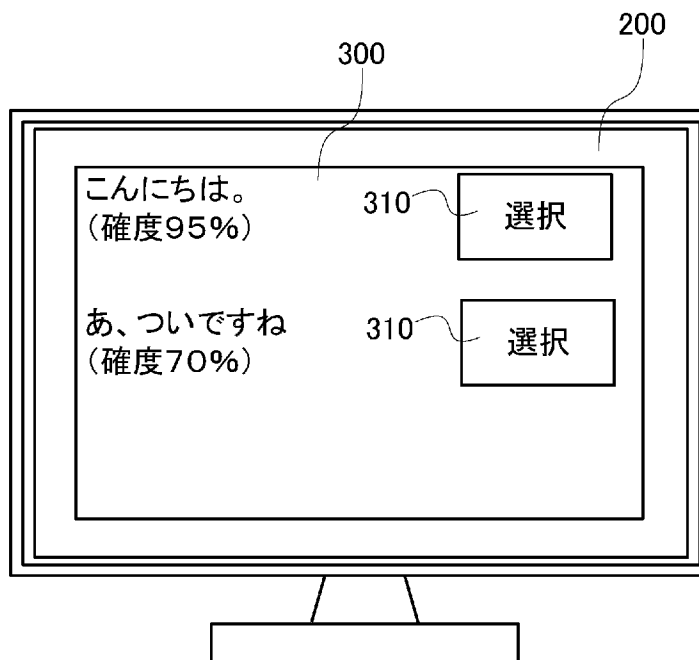
[図5]



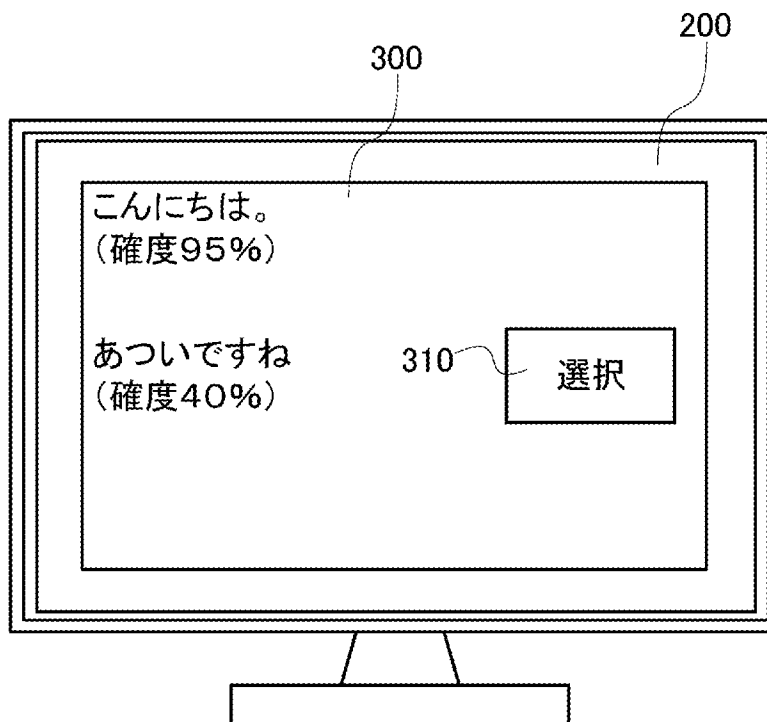
[図6]



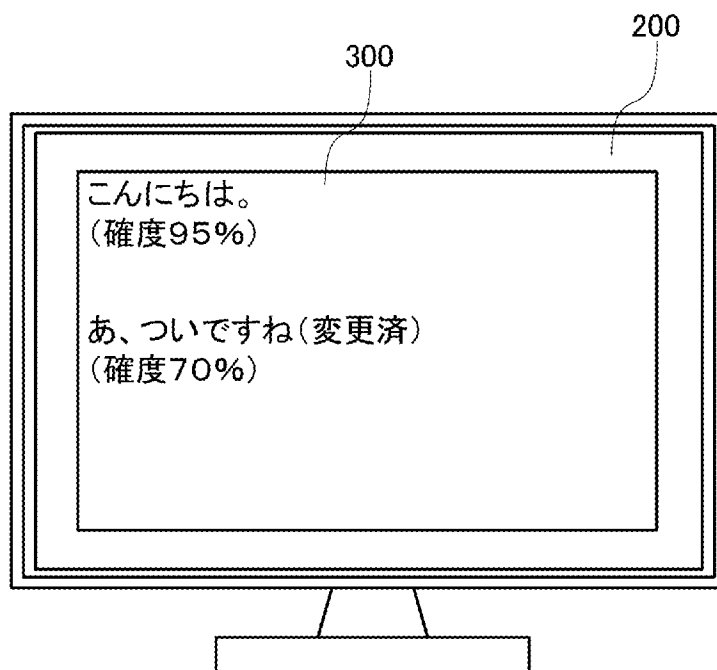
[図7]



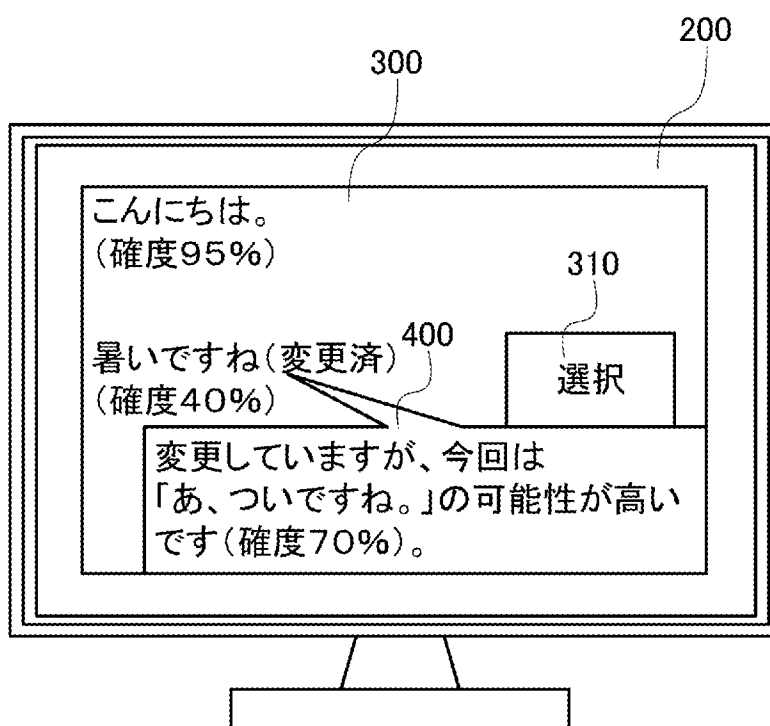
[図8]



[図9]



[図10]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2016/071110

A. CLASSIFICATION OF SUBJECT MATTER

G10L15/22(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)

G10L15/22

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2016
Kokai Jitsuyo Shinan Koho	1971-2016	Toroku Jitsuyo Shinan Koho	1994-2016

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.
X Y	JP 2005-228178 A (NEC Corp.), 25 August 2005 (25.08.2005), paragraphs [0037] to [0046]; fig. 2 to 12 (Family: none)	1-3, 5, 6 4
Y	JP 2011-232521 A (ON Semiconductor Trading, Ltd.), 17 November 2011 (17.11.2011), paragraphs [0036] to [0037]; fig. 6 (Family: none)	4

☐ 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
21 September 2016 (21.09.16)

Date of mailing of the international search report
04 October 2016 (04.10.16)

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. G10L15/22(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. G10L15/22

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	1971-2016年
日本国実用新案登録公報	1996-2016年
日本国登録実用新案公報	1994-2016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
X Y	JP 2005-228178 A (日本電気株式会社) 2005.08.25, 段落[0037]—[0046], 図2-12 (ファミリーなし)	1-3, 5, 6 4
Y	JP 2011-232521 A (オンセミコンダクター・トレーディング・リミテッド) 2011.11.17, 段落[0036]—[0037], 図6 (ファミリーなし)	4

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

21.09.2016

国際調査報告の発送日

04.10.2016

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)

安田 勇太

5Z

4066

電話番号 03-3581-1101 内線 3591