

(11) 特許出願公開番号

特開2019-8274

(P2019-8274A)

(43) 公開日 平成31年1月17日(2019.1.17)

(51) Int.Cl.

F I

テーマコード (参考)

**G 1 0 L 15/22 (2006.01)**

G10L 15/22 300U

**G 1 O L 15/30 (2013.01)**

G10L 15/30

**G 1 O L 17/00 (2013.01)**

G10L 17/00 200C

**G 1 0 L 25/63 (2013.01)**

G10L 25/63

**G 1 O L 15/10 (2006.01)**

G10L 15/10 500Z

審査請求 有 請求項の数 18 O L (全 22 頁) 最終頁に続く

(21) 出願番号 特願2018-75244 (P2018-75244)  
 (22) 出願日 平成30年4月10日 (2018. 4. 10)  
 (31) 優先権主張番号 特願2017-124675 (P2017-124675)  
 (32) 優先日 平成29年6月26日 (2017. 6. 26)  
 (33) 優先権主張国 日本国 (JP)

(71) 出願人 513190830  
フェアリーデバイセズ株式会社  
東京都文京区湯島2丁目31-22 湯島  
アーバンビル7階

(74) 代理人 100110559  
弁理士 友野 英三

(72) 発明者 藤野 真人  
東京都文京区湯島2丁目31-22 湯島  
アーバンビル7階 フェアリーデバイセズ  
株式会社内

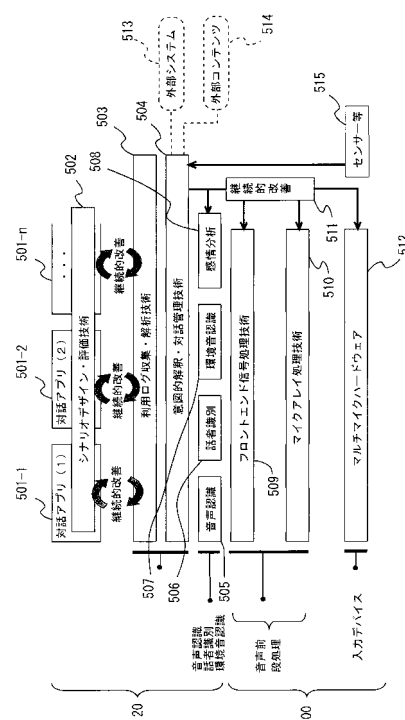
(54) 【発明の名称】 音声情報処理システム、音声情報処理システムの制御方法、音声情報処理システムのプログラム及び記録媒体

(57) 【要約】

【課題】 高精度で、しかも安価に音声内容の分析、解析、認識、評価、修正することにより最適化するとともに、会話の行われている状況をよりの確に把握する技術的思想の提示を目指すこと。

【解決手段】 音声に関する音声情報を入力する入力手段と、入力された音声情報に対し、識別処理が容易となるような前段処理を行う前段処理手段と、前段処理手段により処理された音声情報に所定の加工を施し、第1の情報に基づいてタスク処理を行い、タスク処理の評価が十分でない場合に第1の情報を修正し、評価が十分になるまで一連の処理を繰り返すことで最適化する最適化手段とを備える。

【選択図】図5



**【特許請求の範囲】****【請求項 1】**

音声に関する音声情報を入力する入力手段と、

前記入力された音声情報に対し、識別処理が容易となるような前段処理を行う前段処理手段と、

前記前段処理手段により処理された音声情報に所定の加工を施し、第 1 の情報に基づいてタスク処理を行い、前記タスク処理の評価が十分でない場合に前記第 1 の情報を修正し、前記評価が十分になるまで一連の処理を繰り返すことで最適化する最適化手段とを備えたことを特徴とする音声情報処理システム。

**【請求項 2】**

前記最適化手段は、

前記タスク処理の結果を評価する第 1 の評価手段と、

前記評価が十分でない場合に前記第 1 の情報を修正する修正手段と、

前記前段処理手段から前記修正手段までの一連の処理を繰り返す繰返手段と

を備えたことを特徴とする請求項 1 に記載の音声情報処理システム。

**【請求項 3】**

室内で前記音声の内容を分析して応答する際に、クライアント側の音声入出力装置の処理能力が対応可能な場合に前記音声入出力装置で情報処理を行い、前記音声入出力装置の処理能力が対応可能でない場合にクラウド側が情報処理を行う判断手段を備えたことを特徴とする請求項 1 に記載の音声情報処理システム。

**【請求項 4】**

室内の環境の設定、意図的解釈、及び対話を管理する外部システムを備えたことを特徴とする請求項 1 に記載の音声情報処理システム。

**【請求項 5】**

前記外部システムは、前記音声入出力装置の筐体外を撮像する撮像手段、前記筐体を振動させる振動手段、前記筐体を回転させる回転手段、及び前記筐体外の壁に画像を投影する投影手段を少なくとも一つ備えたことを特徴とする請求項 1 に記載の音声情報処理システム。

**【請求項 6】**

意図的解釈、前記対話の管理に外部コンテンツの利用が可能なことを特徴とする請求項 1 に記載の音声情報処理システム。

**【請求項 7】**

前記音声の内容の分析処理、解析処理、及び認識処理を含む情報加工処理を行い、話者の、年齢、性別を含む属性について推論する推論手段と、

第 2 の情報をデザインする際に利用したログを収集する収集手段と、

前記ログを解析する解析手段と、

前記応答及び前記第 2 の情報を評価する第 2 の評価手段と

を備え、

前記評価に基づいて継続的に改善することで最適化することを特徴とする請求項 1 に記載の音声情報処理システム。

**【請求項 8】**

前記推論手段は、

前記話者との対話を意図的に解釈する解釈手段と、

前記話者との対話を管理する管理手段と

を備えたことを特徴とする請求項 7 に記載の音声情報処理システム。

**【請求項 9】**

前記環境判断手段は、

前記室内のサイズを判断するサイズ判断手段と、

前記室内のノイズレベルを認識するノイズレベル認識手段と、

前記室内の残響レベルを認識する残響レベル認識手段と

10

20

30

40

50

を備えたことを特徴とする請求項 4 に記載の音声情報処理システム。

【請求項 10】

前記筐体に設けられ画像を表示する画像表示手段を備えたことを特徴とする請求項 5 に記載の音声情報処理システム。

【請求項 11】

前記筐体に設けられユーザを認識する指紋認証手段を備えたことを特徴とする請求項 5 に記載の音声情報処理システム。

【請求項 12】

クライアント側の音声入出力装置の処理能力は、プロセッサの演算速度、メモリーサイズ、センサ種類、マイクロフォンアレイ、スピーカの数、LED の数、内蔵カメラ、アプリケーションソフトウェアの数、他社の装置に対応可能なフロントエンド信号処理を含むことを特徴とする請求項 1 に記載の音声情報処理システム。

10

【請求項 13】

前記話者の音声から話し方を特徴として抽出する特徴抽出手段と、  
前記特徴を前記話者の情報に紐づけて記憶し、新たに入力した音声の特徴を前記記憶手段に記憶された話者の情報と照合して話者を識別する話者識別手段と  
を備えたことを特徴とする請求項 1 に記載の音声情報処理システム。

【請求項 14】

前記話者の感情を識別する感情識別手段を備えたことを特徴とする請求項 7 に記載の音声情報処理システム。

20

【請求項 15】

前記筐体の外周に設けられ、軌道上の一部の発光色が残りの部分の発光色と異なるように周回点灯し、前記話者を検知したときに前記一部の発光色が話者の方向で停止するように発光する発光手段を備えたことを特徴とする請求項 14 に記載の音声情報処理システム。

【請求項 16】

音声に関する音声情報を入力し、  
前記入力された音声情報に対し、識別処理が容易となるような前段処理を行い、  
前記前段処理された音声情報に所定の加工を施し、第 1 の情報に基づいてタスク処理を行い、前記タスク処理の評価が十分でない場合に前記第 1 の情報を修正し、前記評価が十分になるまで一連の処理を繰り返すことで最適化することを特徴とする音声情報処理システムの制御方法。

30

【請求項 17】

コンピュータが読み取り可能なプログラムであって、  
コンピュータを、  
音声に関する音声情報を入力する入力手段、  
前記入力された音声情報に対し、識別処理が容易となるような前段処理を行う前段処理手段、  
前記前段処理手段により処理された音声情報に所定の加工を施し、第 1 の情報に基づいてタスク処理を行い、前記タスク処理の評価が十分でない場合に前記第 1 の情報を修正し、前記評価が十分になるまで一連の処理を繰り返すことで最適化する最適化手段、  
として機能させるための音声情報処理システムのプログラム。

40

【請求項 18】

請求項 17 に記載のプログラムを記録した記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、会話における音声情報処理システムに係る、音声情報処理システム、音声情報処理システムの制御方法、音声情報処理システムのプログラム及び記録媒体に関する。

【背景技術】

50

## 【 0 0 0 2 】

近年、音声情報処理技術の発展は目覚ましいものがある。例えば、利用者がシステムの状態を容易に把握できるようにし、利用者とシステムとが常に円滑な対話を実現できる音声対話システム（例えば、特許文献 1 参照）、苦情等を音声により受け付けて、後の処理に利用可能な形態で処理者に伝達することが可能な対話記録システム（例えば、特許文献 2 参照。）、及びユーザと円滑に対話できる、対話機能を有する電子機器（例えば、特許文献 3 参照。）が開示されている。

## 【 0 0 0 3 】

特許文献 1 に記載の発明は、マイク、音声入力手段、音声分析手段、音声認識手段、構文解析手段、意図抽出手段、対話管理手段、問題解決手段、応答文生成手段、音声合成手段、音声出力手段、スピーカ、複数の中途応答処理手段からなる音声対話システムにおいて、複数の中途応答処理手段は、入力系の手段である音声入力手段、音声分析手段、音声認識手段、構文解析手段、意図抽出手段のうち任意の 1 つあるいは複数の手段の処理結果を入力として、処理結果を出力系の手段である音声出力手段、音声合成手段、応答文生成手段のうち 1 つあるいは複数の手段へ出力するものである。

## 【 0 0 0 4 】

特許文献 2 に記載の発明は、対話の音声データを記録する記録装置と、記録される音声データについて、特定の箇所を識別するための識別子を生成して、記録装置に記録させる処理を行う情報処理装置とを備え、情報処理装置は、記記録装置に記録される音声データについて、識別子の生成の要求を受け付けて識別子を生成し、識別子を、記録すべき音声データと対応付けて記録装置に記録し、記録装置には、音声データと、識別子データとが記録され、また、音声データを音声認識部により、音声認識して得られたテキストデータが記録されるものである。

## 【 0 0 0 5 】

特許文献 3 に記載の発明は、冷蔵庫は、マイクおよびスピーカを備え、音声を取得し、取得した音声に応じて発話する対話機能を有するものであり、冷蔵庫は、冷蔵庫近傍の所定の範囲内におけるユーザの位置を特定する位置特定部と、位置特定部にて特定されたユーザの位置に応じた値となるように、マイクの感度を調整するマイク制御部、スピーカの音量を調整するスピーカ制御部を備えるものである。

## 【 先行技術文献 】

## 【 特許文献 】

## 【 0 0 0 6 】

【 特許文献 1 】 特許第 3 4 5 4 8 9 7 号公報

【 特許文献 2 】 特開 2 0 0 0 - 0 6 7 0 6 4 号公報

【 特許文献 3 】 特開 2 0 1 7 - 0 6 9 8 3 5 号公報

## 【 発明の概要 】

## 【 発明が解決しようとする課題 】

## 【 0 0 0 7 】

しかしながら、特許文献 1 に記載の発明は、オウム返し応答もしくは相槌応答によって、利用者は、自分の発話が音声として入力されていることを認識でき、安心して次の発話を行なえるが、定型文を利用しているため、ノイズやエコーの混在したイレギュラーな発話に対しては何ら評価をしたり、定型文に修正を施したりするようにはなっていない。

## 【 0 0 0 8 】

また、特許文献 2 に記載の発明は、予め定めた基準値以上かを判定し、基準値未満のときは、発言が途切れていると判定し、その後、基準値を超える状態となったとき、発言が始まったと判定して、頭出し信号を出力したりするものの、対話の相手の感情を把握したり、聞き間違いに対して改善するような処理はなされていない。

## 【 0 0 0 9 】

さらに、特許文献 3 に記載の発明は、取得した音声に応じて発話する受動的な応答をするようになっているものの相手に対して能動時に話しかけるようにはなっていない。

## 【 0 0 1 0 】

本願は、このような問題点を解決するために企図されたものであり、高精度で、しかも安価に音声内容の分析、解析、認識、評価、修正することにより最適化するとともに、会話の行われている状況をよりの確に把握する技術的思想の提示を目指すものである。

## 【課題を解決するための手段】

## 【 0 0 1 1 】

上記課題を解決するため、請求項 1 に記載の発明は、音声に関する音声情報を入力する入力手段と、前記入力された音声情報に対し、識別処理が容易となるような前段処理を行う前段処理手段と、前記前段処理手段により処理された音声情報に所定の加工を施し、第 1 の情報に基づいてタスク処理を行い、前記タスク処理の評価が十分でない場合に前記第 1 の情報を修正し、前記評価が十分になるまで一連の処理を繰り返すことで最適化する最適化手段と、を備えたことを特徴とする。

10

## 【 0 0 1 2 】

ここで、音声とは、物音（例えば、机やドアをたたく音等）と人の声と雑音（例えば、サイレンや動物の鳴き声、クシャミ等）とを含む音波である。

## 【 0 0 1 3 】

また、第 1 の情報とは、アプリケーションソフトウェア（以下、「アプリ」と記す。）に関連するシナリオデザイン、各種手段のうちどの手段を選択し、どの順番でどのように実行し、評価し、評価が不十分な場合に繰り返すフロー等を含む情報である。

## 【 0 0 1 4 】

請求項 2 に記載の発明は、請求項 1 に記載の構成に加え、前記最適化手段は、前記タスク処理の結果を評価する第 1 の評価手段と、前記評価が十分でない場合に前記第 1 の情報を修正する修正手段と、前記前段処理手段から前記修正手段までの一連の処理を繰り返す繰返手段と、を備えたことを特徴とする。

20

## 【 0 0 1 5 】

請求項 3 に記載の発明は、請求項 1 に記載の構成に加え、室内で前記音声の内容を分析して応答する際に、クライアント側の音声入出力装置の処理能力が対応可能な場合に前記音声入出力装置で情報処理を行い、前記音声入出力装置の処理能力が対応可能でない場合にクラウド側が情報処理を行う判断手段を備えたことを特徴とする。

## 【 0 0 1 6 】

請求項 4 に記載の発明は、請求項 1 に記載の構成に加え、室内の環境の設定、意図的解釈、及び対話を管理する外部システムを備えたことを特徴とする。

30

## 【 0 0 1 7 】

ここで、意図的解釈とは話者の意図を推定し、推定結果を反映した解釈を言う。

## 【 0 0 1 8 】

請求項 5 に記載の発明は、請求項 1 に記載の構成に加え、前記外部システムは、前記音声入出力装置の筐体外を撮像する撮像手段、前記筐体を振動させる振動手段、前記筐体を回転させる回転手段、及び前記筐体外の壁に画像を投影する投影手段を少なくとも一つ備えたことを特徴とする。

## 【 0 0 1 9 】

請求項 6 に記載の発明は、請求項 1 に記載の構成に加え、意図的解釈、前記対話の管理に外部コンテンツの利用が可能なることを特徴とする。

40

## 【 0 0 2 0 】

請求項 7 に記載の発明は、請求項 1 に記載の構成に加え、前記音声の内容の分析処理、解析処理、及び認識処理を含む情報加工処理を行い、話者の、年齢、性別を含む属性について推論する推論手段と、第 2 の情報をデザインする際に利用したログを収集する収集手段と、前記ログを解析する解析手段と、前記応答及び前記第 2 の情報を評価する第 2 の評価手段とをそなえ、前記評価に基づいて継続的に改善することで最適化することを特徴とする。

## 【 0 0 2 1 】

50

ここで、第２の情報とは、各種手段のうちどの手段を用い、どの順番で処理し、評価し、十分でない場合に繰り返すフローについての情報をいう。

【００２２】

認識処理は、収集した音声情報から、話し手の他に、笑い声、拍手、呼び声等の認識、さらに環境音を分析、解析、認識等の処理を行った結果から、話者識別、性別推定、年齢推定等を行うとともに、イントネーション判定から、出身地等に関する各種情報を提供するものである。

【００２３】

請求項８に記載の発明は、請求項７に記載の構成に加え、前記推論手段は、前記話者との対話を意図的に解釈する解釈手段と、前記話者との対話を管理する管理手段と、を備えたことを特徴とする。

10

【００２４】

ここで、話者との対話の管理とは、顧客満足度向上のため、話者がどのような発話に対しどのような感情を抱いたかを記録し、クライアント側の音声入出力装置をコールセンターに利用していた場合にオペレータに注意喚起したり、管理者に報告したりすることを含む。また、クライアント側の音声入出力装置を会議に利用していた場合に出席者が感情的になった場合に落ち着かせるように休憩を入れたり、冷静になるような旨の音声を発話したりすることを含む。

【００２５】

請求項９に記載の発明は、請求項４に記載の構成に加え、前記環境判断手段は、前記室内のサイズを判断するサイズ判断手段と、前記室内のノイズレベルを認識するノイズレベル認識手段と、前記室内の残響レベルを認識する残響レベル認識手段と、を備えたことを特徴とする。

20

【００２６】

請求項１０に記載の発明は、請求項５に記載の構成に加え、前記筐体に設けられ画像を表示する画像表示手段を備えたことを特徴とする。

【００２７】

請求項１１に記載の発明は、請求項５に記載の構成に加え、前記筐体に設けられユーザを認識する指紋認証手段を備えたことを特徴とする。

【００２８】

30

請求項１２に記載の発明は、請求項１に記載の構成に加え、クライアント側の音声入出力装置の処理能力は、プロセッサの演算速度、メモリーサイズ、センサ種類、マイクアレイ、スピーカの数、ＬＥＤの数、内蔵カメラ、アプリケーションソフトウェアの数、他社の装置に対応可能なフロントエンド信号処理を含むことを特徴とする。

【００２９】

請求項１３に記載の発明は、請求項１に記載の構成に加え、前記話者の音声から話し方を特徴として抽出する特徴抽出手段と、前記特徴を前記話者の情報に紐づけて記憶し、新たに入力した音声の特徴を前記記憶手段に記憶された話者の情報と照合して話者を識別する話者識別手段と、を備えたことを特徴とする。

【００３０】

40

請求項１４に記載の発明は、請求項７に記載の構成に加え、前記話者の感情を識別する感情識別手段を備えたことを特徴とする。

【００３１】

請求項１５に記載の発明は、請求項１４に記載の構成に加え、前記筐体の外周に設けられ、軌道上の一部の発光色が残りの部分の発光色と異なるように周回点灯し、前記話者を検知したときに前記一部の発光色が話者の方向で停止するように発光する発光手段を備えたことを特徴とする。

【００３２】

請求項１６に記載の発明は、音声に関する音声情報を入力し、前記入力された音声情報に対し、識別処理が容易となるような前段処理を行い、前記前段処理された音声情報に所

50

定の加工を施し、第 1 の情報に基づいてタスク処理を行い、前記タスク処理の評価が十分でない場合に前記第 1 の情報を修正し、前記評価が十分になるまで一連の処理を繰り返すことで最適化することを特徴とする。

【 0 0 3 3 】

請求項 1 7 に記載の発明は、コンピュータが読み取り可能なプログラムであって、コンピュータを、音声に関する音声情報を入力する入力手段、前記入力された音声情報に対し、識別処理が容易となるような前段処理を行う前段処理手段、前記前段処理手段により処理された音声情報に所定の加工を施し、第 1 の情報に基づいてタスク処理を行い、前記タスク処理の評価が十分でない場合に前記第 1 の情報を修正し、前記評価が十分になるまで一連の処理を繰り返すことで最適化する最適化手段、として機能させるための音声情報処理システムのプログラムであることを特徴とする。

10

【 0 0 3 4 】

請求項 1 8 に記載の発明は、請求項 1 7 に記載のプログラムを記録した記録媒体であることを特徴とする。

【発明の効果】

【 0 0 3 5 】

本発明によれば、高精度で、しかも安価に音声内容の分析、解析、認識、評価することにより最適化するとともに、会話の行われている状況をよりの確に把握することを可能とする。

【図面の簡単な説明】

20

【 0 0 3 6 】

【図 1】本発明の一実施形態に係る音声情報処理システム全体の構成図の一例である。

【図 2】図 1 に示した音声情報処理システムに用いられるクラウド側サーバのハードウェアブロック図の一例である。

【図 3】図 1 に示した音声情報処理システムのハードウェアブロック図の一例である。

【図 4】図 2 に示した音声情報処理システムのソフトウェアブロック図の一例である。

【図 5】図 2 に示した音声情報処理システムの処理内容を示すソフトウェアスタック図の一例である。

【図 6】図 2 に示した音声情報処理システムの外觀図の一例である。

【図 7】図 2 に示した音声情報処理システムにおける全体動作を示すフローチャートの一例である。

30

【図 8】図 2 に示した音声情報処理システムにおける全体動作を示すフローチャートの他の一例である。

【図 9】図 2 に示した音声情報処理システムにおける全体動作を示すフローチャートの他の一例である。

【発明を実施するための形態】

【 0 0 3 7 】

本発明の実施の形態を、図面を参照して説明する。

【 0 0 3 8 】

< 構成 >

40

< システム全体 >

図 1 は、本発明の一実施形態に係る音声情報処理システム全体の構成図の一例である。本発明の一実施形態に係る音声情報処理システムは、ネットワーク 10 を介してクラウド側サーバ 20、及びクライアント側の音声入出力装置 100 が接続されて構成されている。

【 0 0 3 9 】

音声入出力装置 100 とスマートフォンとで連携して音声情報処理システムを構成してもよく、Wi-Fi ルータを用いてインターネット接続してもよい。なお、音声入出力装置 100 とスマートフォンとの間の通信手段としては、例えば、無線、赤外線、有線等どのような方法を用いてもよい。また本願を適用する、例えば音声情報処理システムが複数

50

存在する場合には、音声入出力装置 100 はその数だけ存在することとなる。

#### 【0040】

< クラウド側サーバ及び音声入出力装置のハードウェア構成 >

次に、図 2 に従ってクラウド側サーバ 20 の詳細を説明する。図 2 に示すように、クラウド側サーバ 20 は、データベース（以下、「DB」という。）21、プロセッサ 22、出力装置 23、入力装置 24、及びインターフェース 26 等を備えて構成されている。プロセッサ（「コンピュータ」とも称する）22 は、音声情報の管理に関するデータの処理を行い、DB 21 は音声情報の管理に関する情報等のデータ、及び制御プログラム等を記憶する。出力装置 23 は、ディスプレイ、プリンタ等を備えて構成され、必要に応じて各種情報を出力する。また、入力装置 24 は、キーボード、バーコードリーダ、及びスキャナ等を備えて構成され、必要に応じて情報の入力を行うが、情報の入力を可能とするすべての装置を含むものとする。なお、クラウド側サーバ 20 は最終的に音声情報処理システムとしての業務遂行が可能であれば、単独でも複数のシステムから構成されていてもよい。

10

#### 【0041】

< 音声入出力装置のハードウェア構成 >

次に、図 3 に従って音声入出力装置 100 の詳細を説明する。同図に示すように、音声入出力装置 100 は、主として拡張部 201、記憶部 202、マイクユニット 203、マイク制御部 204、信号処理部 205、通信部 206、音声発生部 207、非可聴音発生部 208、及び表示部 209 を備える。表示部 209 は LED (Light Emitting Diode: 発光ダイオード) 210 と LCD (Liquid Crystal Display: 液晶表示ディスプレイ) 211 とを有してもよい。LED 210 はリング状であってもよい。

20

#### 【0042】

音声入出力装置 100 は、さらに破線で示す撮像部 212、個人認証部 213、IR (Infrared: 赤外線) 部 214、投影部 215、振動部 216、及び回転部 217 を備えて構成してもよい。

#### 【0043】

拡張部 201 は音声入出力装置 100 に USB (Universal Serial Bus) メモリや USB 機器を接続するための部材である。

30

#### 【0044】

記憶部 202 は、音声入出力装置 100 の制御プログラム、音声データ、個人データ、画像データ等のデータを記憶する部材であり、例えば ROM (Read Only Memory: 読み出し専用メモリ)、RAM (Random Access Memory: 書き換え自在メモリ)、HDD (Hard Disk Drive)、SSD (Solid State Drive) が挙げられる。

#### 【0045】

マイクユニット 203 は、少なくとも 1 本のマイクからなり、マイク制御部 204 にて AGC (Automatic Gain Control: 自動利得制御) やフォーミング等の制御が行われる。

40

#### 【0046】

信号処理部 205 は、マイクからの音声信号に対し周囲雑音除去等の処理を施し、正確に認識処理した後、処理済みの情報を記憶部 202 に記憶し、音声発生部 207 から音声を発生させたり、マイクからの音声信号の話者識別処理や感情識別処理を行ったり、処理済みの音声情報を到来方位情報、話者識別情報、感情識別情報と共に記憶部 202 に記憶し、同時に表示部 209 に適合した表示を行ったりする。併せて通信部 206 や拡張部 201 より外部に送信し、クラウド処理等により詳細な情報分析を行うことができる。これらの処理により、特定方位に存在する雑音源からの音声情報をブロックアウトしたり、これとは逆に特定方位からの情報のみを記録したりすることができる。

#### 【0047】

50



また、記憶部 202 は多層構成とし、記録すべき音声情報の到来方位や話者識別、感情識別等の関連情報の整理が可能となる。

【0048】

信号処理部 205 は、Wi-Fi やブルートゥース（登録商標）等により外部機器と無線通信するための通信部 206 とハードワイヤにて外部機器と接続する拡張部 201 とを有し、外部マイクにより周囲雑音を集音して拡張部からかかる受信雑音を入力して周囲雑音の影響を低減したり、USB ポートにより外部機器と通信したりすることが可能である。

【0049】

非可聴音発生部 208 から超音波を発生し、その反射による話者や壁までの距離を測定することができる。

【0050】

表示部 209 のLED 210 は、リング状のLED を周回点灯させたり、点滅させたり発光間隔や発光色を変化させたりしてもよい。LCD 211 は、音声入出力装置 100 の筐体の天板や側面に設けてもよく、カラーでもモノクロでもよい。

【0051】

撮像部 212 は、音声入出力装置 100 の周囲の状況を撮像する部材であり、例えば CCD (Charge Coupled Device: 電荷結合素子) カメラが挙げられる。撮像部 212 による撮像画像は動画画像でも静止画像でもよい。

【0052】

個人認証部 213 は、ユーザの指紋や声紋を識別する部材であり、音声入出力装置 100 の天板に設けられた指紋識別装置であっても、話者の音声から声紋を識別する声紋識別装置（もしくはソフトウェア）であってもよい。

【0053】

IR 部 214 は、赤外線センサであり、人感センサとして人の侵入の監視や来客検知に用いることができる。

【0054】

投影部 215 は、音声入出力装置 100 の筐体に設けられ、例えば会議や旅行説明のため、室内のホワイトボードや壁やスクリーンに地図や議題を投影するプロジェクターである。

【0055】

振動部 216 は、クライアント側の音声入出力装置 100 の筐体を振動させることで、ユーザに注意を喚起させるものである。振動部 216 は、例えば、圧電素子や出力軸に偏心カムを有するモータ等が挙げられる。

【0056】

回転部 217 は、音声入出力装置 100 の底面に設けられたベースと、ベース上に設けられた回転軸と、回転軸上に設けられ筐体を回転させるモータとで構成される。この回転部 217 により、投影部 215 や LCD 211 の向きを変えることができる。

【0057】

<クラウド側サーバ及び音声入出力装置のソフトウェア構成>

図 4 に従ってクラウド側サーバ及び音声入出力装置のソフトウェア構成について説明する。

【0058】

<クラウド側サーバ>

クラウド側サーバ 20 は、入力手段 41、出力手段 42、記憶手段 43、判断手段 44、最適化手段 45、翻訳手段 49、第 1 制御手段 50、推論手段 51、及び通信手段 52 を備えて構成されている。

【0059】

最適化手段 45 は、評価手段 46、修正手段 47、及び繰返手段 48 を備える。推論手段 51 は、感情識別手段 51a、方位検出手段 51b、話者識別手段 51c、収集手段 5

10

20

30

40

50

1 d、解釈手段 5 1 e、管理手段 5 1 f、サイズ判断手段 5 1 g、ノイズレベル認識手段 5 1 h、及び残響レベル認識手段 5 1 i を備える。

【0060】

入力手段 4 1 は、必要に応じて情報の入力を行うが、情報の入力を可能とするすべての装置を含むものであり、図 2 に示した入力装置 2 4 によって実現される。

【0061】

出力手段 4 2 は、必要に応じて各種情報を出力する手段であり、図 2 に示した出力装置 2 3 によって実現される。

【0062】

記憶手段 4 3 は、クラウド側サーバの制御プログラム、及び音声情報の管理に関する情報等のデータ等を記憶する手段であり、図 2 に示したデータベース 2 1 によって実現される。制御プログラムには、起動時について能動的な動作、すなわち、例えば人を検知したときに先に挨拶を言うように設定されている。

【0063】

判断手段 4 4 は、室内で音声の内容を分析して応答する際に、クライアント側の音声入出力装置 1 0 0 の処理能力が対応可能な場合にクライアント側の音声入出力装置 1 0 0 で情報処理を行い、クライアント側の音声入出力装置 1 0 0 の処理能力が対応可能でない場合にクラウド側サーバ 2 0 が情報処理を行うよう判断する手段であり、図 2 に示したプロセッサ 2 2 によって実現される。

【0064】

ここで、クライアント側の音声入出力装置 1 0 0 の処理能力とは、プロセッサの演算速度、メモリーサイズ、センサ種類、マイクアレイ、スピーカの数、LED の数、内蔵カメラ、アプリケーションソフトウェアの数、他社の装置に対応可能なフロントエンド信号処理を含む。

【0065】

最適化手段 4 5 は、識別処理が容易となるような前段処理手段により処理された音声情報に所定の加工を施し、第 1 の情報に基づいてタスク処理を行い、タスク処理の評価が十分でない場合に第 1 の情報を修正し、評価が十分になるまで一連の処理を何回でも繰り返すことで最適化する手段であり、図 2 に示したデータベース 2 1 及びプロセッサ 2 2 によって実現される。

【0066】

第 1 の情報とは、アプリに関連するシナリオデザイン、各種手段のうちどの手段を選択し、どの順番でどのように実行し、評価し、結果が不十分な場合に何回でも繰り返すフロー等を含む情報である。

【0067】

最適化手段 4 5 の評価手段 4 6 は、タスク処理の結果を評価する手段である。最適化手段 4 5 の修正手段 4 7 は、結果が十分でない場合に第 1 の情報を修正する手段である。最適化手段 4 5 の繰返手段は、評価手段 4 6 から修正手段 4 7 までの一連の処理を何回でも繰り返す手段である。

【0068】

翻訳手段 4 9 は、自動的に言語を識別し、例えば日本語から日本語以外の多言語に翻訳し、日本語以外の多言語を日本語に翻訳する手段であり、図 2 に示したデータベース 2 1 及びプロセッサ 2 2 によって実現される。翻訳手段 4 9 は、日本語の音声日本語以外の多言語の音声に変換したり、日本語以外の多言語の音声日本語の音声に変換したりする、いわば、通訳機能（もしくは同時通訳機能）を有していてもよい。この場合、音声は音声合成手段により合成されるが、話者の性別、年齢を判断し、老若男女に対応した音声を発音するようにしてもよい。また、翻訳手段 4 9 は、翻訳する際はテキスト形式で翻訳内容を記録することもできるようになっている。テキストデータは話者と紐づけるようにするのが好ましい。テキストデータは話者に応じて色分けするようにしてもよい。

【0069】

第 1 制御手段 5 0 は、クラウド側サーバ 2 0 の各手段を統括制御する手段であり、図 2 に示したプロセッサ 2 2 によって実現される。

【 0 0 7 0 】

推論手段 5 1 は、音声の内容の分析処理、解析処理、及び認識処理を含む情報加工処理を行い、話者の、年齢、性別を含む属性について推論する手段であり、図 2 に示したデータベース 2 1 及びプロセッサ 2 2 によって実現される。

【 0 0 7 1 】

感情識別手段 5 1 a は、話者の感情を識別する手段であり、図 2 に示したデータベース 2 1 及びプロセッサ 2 2 によって実現される。感情識別手段 5 1 a によって識別された話者の感情について話者の発話と紐づけて記録するのが好ましい。

【 0 0 7 2 】

方位検出手段 5 1 b は、音声入出力装置 1 0 0 から見た話者の方位を検出する手段であり、図 2 に示した入力装置 2 4 及びプロセッサ 2 2 によって実現される。

【 0 0 7 3 】

話者識別手段 5 1 c は、話者の音声から話し方の平均的な音響モデルとの差を特徴として抽出する特徴抽出手段により得られた特徴を話者の情報に紐づけて記憶し、新たに入力した音声の特徴を記憶手段 4 3 に記憶された話者の情報と照合して話者を識別する手段であり、図 2 に示したデータベース 2 1 及びプロセッサ 2 2 によって実現される。話者識別に関しては、後述する音声入出力装置 1 0 0 に、撮像する撮像手段 7 3 や指紋認証手段 7 2 を設けることにより、顔画像認識処理や指紋認証処理によって実現するようにしてもよい。

【 0 0 7 4 】

収集手段 5 1 d は、第 2 の情報をデザインする際に利用したログを収集する手段であり、図 2 に示したデータベース 2 1 及びプロセッサ 2 2 によって実現される。第 2 の情報とは、前述したように各種手段のうちどの手段を用い、どの順番で処理し、評価し、結果が十分でない場合に繰り返すフローについての情報をいう。

【 0 0 7 5 】

解釈手段 5 1 e は、話者との対話を意図的に解釈する手段であり、図 2 に示したデータベース 2 1 及びプロセッサ 2 2 によって実現される。

【 0 0 7 6 】

管理手段 5 1 f は、話者との対話を管理する手段であり、図 2 に示したデータベース 2 1 及びプロセッサ 2 2 によって実現される。

【 0 0 7 7 】

サイズ判断手段 5 1 g は、室内のサイズを判断する手段であり、図 2 に示したインターフェース 2 6、及びプロセッサ 2 2 によって実現される。サイズ判断手段 5 1 g は、インターフェース 2 6 を介してクライアント側の音声入出力装置 1 0 0 において、非可聴音を間欠発音し、周辺からの反射音をマイクで集音し、クライアント側の音声入出力装置 1 0 0 の環境を把握（2 次元方位と距離）するようにしてもよい。

【 0 0 7 8 】

ノイズレベル認識手段 5 1 h は、室内のノイズレベルを認識する手段であり、図 2 に示したインターフェース 2 6、及びプロセッサ 2 2 によって実現される。ノイズレベル認識手段 5 1 h は、クライアント側の音声入出力装置 1 0 0 のマイクで得られた室内の音声からノイズ除去処理前のノイズレベルの情報を、インターフェース 2 6 を介して得ることができるようにしている。室内のノイズレベルによってクライアント側の音声入出力装置 1 0 0 の環境が例えば受付か、会議室か、コールセンター室かそれ以外かが判断できる。

【 0 0 7 9 】

残響レベル認識手段 5 1 i は、室内の残響レベルを認識する手段であり、図 2 に示したインターフェース 2 6、及びプロセッサ 2 2 によって実現される。残響レベル認識手段 5 1 i は、クライアント側の音声入出力装置 1 0 0 のマイクで得られた室内の音声から残響レベルの情報を、インターフェース 2 6 を介して得ることができるようにしている。室

10

20

30

40

50

内の残響レベルによってクライアント側の音声入出力装置 100 の環境が判断できる。

【0080】

通信手段 52 は、クラウド側サーバ 20 とクライアント側の音声入出力装置 100 との間でネットワーク 10 を介して情報を授受するための手段であり、図 2 に示したインターフェース 26 によって実現できる。

【0081】

< 音声入出力装置 >

クライアント側の音声入出力装置 100 は、主として入力手段 61、出力手段 62、前段処理手段 63、発光手段 64、通信手段 65、第 2 制御手段 66、記憶手段 67、入出力手段 68、及び検知手段 69 を備えて構成されている。音声入出力装置 100 は、さら

10

【0082】

入力手段 61 は、音声に関する音声情報を入力する手段であり、図 3 に示すマイクユニット 203 及びマイク制御部 204 によって実現される。入力手段 61 は、ビームフォーミング処理、ブラインド音源分離処理、残響抑制処理、ノイズ抑圧処理、エコーキャンセル（バージン）処理、及び音声区間検出（Voice Activity Detection: VAD と略す。）処理を施す。

【0083】

出力手段 62 は、非可聴音（20 kHz ~ 40 kHz、好ましくは 30 kHz の音）を発生する手段であり、図 3 に示した非可聴音発生部 208 によって実現される。非可聴音発生部 208 としては、例えばムービングコイルを用いた超音波スピーカの他、圧電素子を用いた超音波スピーカが挙げられる。

20

【0084】

前段処理手段 63 は、マイクからの音声からノイズを除去したり、エコーをキャンセルしたり、ビームフォーミング処理、ブラインド音源分離処理、残響抑制処理を行う手段であり、図 3 に示したマイク制御部 204 によって実現される。

【0085】

発光手段 64 は、筐体の外周に設けられ、軌道上の一部の発光色が残りの部分の発光色と異なるように周回点灯したり、話者を検知したときに一部の発光色が話者の方向で停止するように発光したりする手段であり、図 3 に示した LED 210 によって実現される。

30

【0086】

通信手段 65 は、クライアント側の音声入出力装置 100 とクラウド側サーバ 20 との間でネットワーク 10 を介して情報を授受する手段であり、図 3 に示した通信部 206 によって実現される。

【0087】

第 2 制御手段 66 は、音声入出力装置 100 を統括制御する手段であり、図 3 に示した信号処理部 205 によって実現される。信号処理部 205 としては、例えばプロセッサが挙げられる。

【0088】

記憶手段 67 は、音声入出力装置 100 を統括制御するプログラムを記憶する手段であり、図 3 に示した記憶部 202 によって実現される。記憶手段 67 は、例えば ROM、RAM、HDD、SSD が挙げられ、音声情報、個人情報、画像情報、指紋情報を記憶するように構成してもよい。

40

【0089】

入出力手段 68 は、USB フラッシュメモリや USB 機器を接続するための手段であり、図 3 に示した拡張部 201 によって実現される。

【0090】

検知手段 69 は、音声入出力装置 100 に接近したり通過したりする人を検知する手段であり、図 3 に示した IR 部 214 が挙げられる。検知手段 69 としては、例えば人感セ

50

ンサが挙げられる。

【0091】

画像表示手段71は、文字情報を含む静止画像や動画像等の画像を表示する手段であり、図3に示したLCD211によって実現される。

【0092】

指紋認証手段72は、ユーザを認識する手段であり、図3に示した個人認証部213によって実現される。指紋認証手段72としては、例えば指紋センサが挙げられる。

【0093】

撮像手段73は、デジタルカメラであり、図3に示した撮像部212によって実現される。

10

【0094】

外部情報入力手段74は、外部からのコンテンツを入力する手段であり、図3に示した拡張部201によって実現される。

【0095】

振動手段75は、音声入出力装置100の筐体を振動させる手段であり、図3に示した振動部216によって実現される。

【0096】

回転手段76は、音声入出力装置100の筐体を鉛直な中心軸の周りに回転（旋回）させる手段であり、図3に示した回転部217によって実現される。

20

【0097】

投影手段77は、室内のスクリーン、ホワイトボード、壁面等に画像を投影する手段であり、図3に示した投影部215によって実現される。

【0098】

<ソフトウェアスタック>

上記に示す分析処理、解析処理、認識処理等の構成について、図5のソフトウェアスタック図の処理内容に従って説明する。処理内容は、利用ログ収集・解析部503、意図解釈・対話管理技術部504、音声認識部505、話者識別部506、環境音認識部507、感情分析部508、フロントエンド信号処理技術部509、マイクアレイ処理技術部510、マルチマイクアレイ処理技術部510、マルチマイクハードウェア部512、センサ515等から構成される。意図的解釈・対話管理技術部504は、外部システム513や外部コンテンツ514が接続されていてもよい。対話アプリ501-1～501-nに応じてシナリオデザインが評価される。その際利用ログ収集・解析して継続的に改善する。尚、図では継続的改善に3つの矢印が記載されているが、意図的解釈・対話管理技術部504、音声認識部505、話者識別部506、環境音認識部507、感情分析部508、フロントエンド信号処理技術部509、マイクアレイ処理技術部510、及びマルチマイクハードウェア部512についても、順番を入れ替えたり、一部を省略したりするとともに継続的改善が行われるので矢印が記載されていてもよい。

30

【0099】

マルチマイクハードウェア512やフロントエンド信号処理技術部509は他社製品にも柔軟に対応可能である。

40

【0100】

マイクアレイ処理技術部510は、ビームフォーミング処理、ブラインド音源分離処理、残響抑制処理等から構成され、フロントエンド信号処理技術部509はノイズ抑圧処理、エコーキャンセル（バージイン）処理、音声区間検出（「VAD」とも略す。）処理等から構成され、音声を文字等に変換する音声認識部505は多言語対応処理、自動言語識別処理、多言語混合処理等から構成され、話者識別部506は事前学習処理、クラスタリング処理、さらに話者照合（認証）処理等から構成され、感情分析部508は感情多クラス分類処理、感情マッピング処理、抑揚認識処理等から構成され、環境音認識部507は拍手音・笑い声認識処理、重なり検出処理、シーン推定処理、さらに異音検出処理は音響官能検出処理、正常/異常音識別処理等から構成される。

50

## 【0101】

以上、音声処理の構成を説明したが、音声処理により、笑い声、拍手、呼び鈴といった特徴のある音の認識、話者識別、性別推定、イントネーション判定等の処理を行う。ただし、上記各処理を可能とするためには、音の前処理として、音声切り出し、ノイズ抑圧、残響抑圧、音源定位すなわち指定角度の音を取り出すビームフォーミング等の処理を行うものとする。さらに、上記各処理による認識結果情報は、例えば関連システムと連携して、利用ログ収集・解析技術部503、意図解釈・対話管理技術部504により解析、管理処理を行うものとする。

## 【0102】

また、人の話し声を例えばマイクによって収集するだけでなく、例えばスピーカにより、人の聴覚では聞くことができない、例えば超可聴音もしくは非可聴音を発音し、この超可聴音もしくは非可聴音の反射音を例えばマイクによって收音し、この情報を音声処理により解析、認識処理を行い、音声入出力装置100の周辺状況に関し、反射する対象の材質、距離等を把握するものとする。これらを可能とするために、例えばマイクを複数搭載するとともに、例えば水平方向に円形状にマイクを搭載したマイク搭載部（図示せず）を縦方向に、例えば2層搭載することにより、垂直方向の検知を可能とする。また、発音する超可聴音もしくは非可聴音を例えばパルス状に成形し、あるパルス音から次のパルス音との間隔に、ある音声入出力装置100に特有の情報音を付加することにより、複数の音声入出力装置100を識別することを可能とする。

## 【0103】

ここで、本実施形態によれば、話者識別や感情識別等の音声処理や利用ログ収集、解析、意図解釈等のどの処理を使うか、どの順番で使うか、ノイズを除去してから残響処理を行うか、残響処理を行ってからノイズを除去するのかはユーザが自由に選べるようになっている。

## 【実施例】

## 【0104】

< 音声情報処理システム >

図6に示すシステムは、音声入出力装置100と、クラウド側サーバ20と、を有するシステムである。

## 【0105】

音声入出力装置100は、多数の貫通孔が形成された筐体101の天板外周部に配置されたLED(Light Emitting Diode: 発光ダイオード)リング102と、筐体101の同一平面の周面に円周状に配置された複数(例えば16個であるが限定されない。)のPDMマイク103-1~103-16と、筐体101内に下端向きに配置されたスピーカ群(スコカ104S、ツイータ104S)と、筐体101の底面に上向きに凸の円錐形状の反射板105と、を備える装置である。筐体101内には各種回路基板が設けられている。106は電源ランプとしてのLEDである。107は電源コードであるが、バッテリー搭載可能である。

## 【0106】

筐体101は、図では円筒状であるが限定されず角柱状であっても、円錐状であっても、角錐台状であっても、円錐台状であってもよい。

## 【0107】

LEDリング102は多数の3色LED多色発光型のデバイスであり、環状に形成されたものである。LEDリング102は、一部の数個の隣接配置されたLEDと他の残りの部分のLEDとが異なる発光色で発光したり、円周軌道にそって回転したり停止したりし、周回点灯したりすることが可能である。例えば、話者方向を数個のLEDで白色点灯し、残りのLEDを青色点灯することが挙げられるが、これに限定されるものではなく、点灯の代わりに点滅させたり、インジケータのように話者の音声の強度に応じて照度を変化させたりしてもよい。

## 【0108】

PDMマイク103-1~103-16は、筐体101の同一平面の周面に16個等間隔で配置されているため、反射音より音源の左右方向の識別が20度の範囲で可能であり、音声入出力装置100の近傍にいる音源としての人の音声を集音することが可能である。

#### 【0109】

スコカ104Sは通常の音声を発生するスピーカであり、ツイータ104Tは非可聴音（超音波）を発生するスピーカである。反射板105は、スコカ104S及びツイータ104Sからの音声や非可聴音を筐体101の外側に放射状に反射させる部材である。ツイータ104からの非可聴音は反射板105で反射され筐体101の外部に出射して話者に反射した場合には筐体101に向かい、PDMマイク103-1~103-16で集音することでレーダー（もしくはソナー）のように機能させることができる。この非可聴音はパルス変調されていてもよい。

クラウドシステム300は、複数のサーバを有するサーバ群301からなり、音声入出力装置100のディープラーニング処理等のソフトウェア処理を行ってもよい。

#### 【0110】

##### <動作1>

図7に示すフローチャートの動作の主体は、クラウド側サーバ20のプロセッサ22である。

#### 【0111】

本システムは、予めユーザが第1の情報としてのシナリオを設定可能である。シナリオとは、例えば、話者と音声入出力装置100との対話に対してどのように応答するかを示すストーリーである。

#### 【0112】

まずユーザがクラウド側サーバ20で複数情報（例えば、第1の情報、第2の情報、第3の情報）を作成する（ステップS10）。

#### 【0113】

音声入出力装置100の電源スイッチがオンされると（ステップS11）、プロセッサ22は音声入出力装置100の外部から、例えば図示しないスマートフォンでモード設定信号が有るか否か判断する（ステップS12）。

#### 【0114】

プロセッサ22は、外部からモード設定信号が有ると判断した場合（ステップS12 / YES）、コールセンターモード、受付モード、会議モード、...のうちのいずれかのモードが設定され（ステップS13）、外部からモード設定信号が無いと判断した場合（ステップS12 / NO）、ステップS14に進む。

#### 【0115】

ステップS14では、プロセッサ22が、クライアント側の音声入出力装置100の処理能力が十分か否かを判断し、音声入出力装置100の処理能力が十分であると判断した場合（ステップS14 / YES）、クライアント側の音声入出力装置100で処理し（ステップS15）、クライアント側の音声入出力装置100の処理能力が十分でないと判断した場合（ステップS14 / NO）クラウド側で処理し（ステップS16）、ステップS17に進む。

#### 【0116】

プロセッサ22は、最適化済か否かを判断し（ステップS17）、最適化済でない場合（ステップS17 / NO）、第1の情報を選択し（ステップS18）、識別処理が容易となるような前段処理を実行する。前段処理は、第1の情報として、例えば、ビームフォーミング、ブラインド音源分離、及び残響抑制のいずれかから少なくとも一つ選択して、順番を決定し、実行する。ノイズ抑圧、エコーキャンセル、及び音声区間検出から適宜選択して、順番を決定し、実行するものとする（ステップS19）。

#### 【0117】

次にタスクの評価（ステップS20）を行い、評価が十分か否かを判断し（ステップS

10

20

30

40

50

21)、評価が十分でないと判断した場合(ステップS21/NO)、ステップS10に戻り、クラウド側で第1の情報を修正し、ステップS10～ステップS23を何回でも繰り返す。

【0118】

プロセッサ22は、ステップS21でタスクの評価が十分であると判断した場合(ステップS21/YES)、第2の情報を選択し(ステップS22)、識別処理を実行する。識別処理は、第2の情報として、例えば、音声認識、話者識別、感情分析、及び環境音認識のいずれかから少なくとも一つ選択して、順番を決定し、実行するものとする(ステップS23)。

【0119】

次にタスクの評価(ステップS24)を行い、評価が十分か否かを判断し(ステップS25)、評価が十分でないと判断した場合(ステップS25/NO)、ステップS10に戻り、クラウド側で第2の情報を修正し、ステップS10～ステップS24を何回でも繰り返す。

【0120】

プロセッサ22は、ステップS24でタスクの評価が十分であると判断した場合(ステップS24/YES)、第3の情報を選択し(ステップS26)、対話アプリを実行する。対話アプリについては、第3の情報として、利用ログ収集、解析、意図的解釈、及び対話管理のいずれかから少なくとも一つ選択して、順番を決定し、実行するものとする(ステップS27)。

【0121】

次に、タスクの評価(ステップS28)を行い、評価が十分か否かを判断し(ステップS29)、評価が十分でないと判断した場合(ステップS29/NO)、ステップS10に戻り、クラウド側で第3の情報を修正し、ステップS10～ステップS28を何回でも繰り返す。

【0122】

プロセッサ22は、最適化が完了すると(ステップS30)、アプリを実行し(ステップS31)、終了しない場合(ステップS32/NO)、ステップS12に戻り、終了する場合(ステップS32/YES)、終了する。この場合、電源スイッチが自動的オフになるように構成してもよい。

【0123】

ここで、図7に示したフローチャートは一実施例に過ぎず、限定されるものではない。例えば、室内環境が同一であって、話者の数に変更があったり、室内環境に変更があっても話者や人数に変更がなかったり、使用するモードが変更になったり、使用するマイクの数やグレードに変更があったりしても、適宜ステップS18～ステップS21と、ステップS22～ステップS25と、ステップS26～ステップS29とを入れ替えたり、一部を省略したりすることで継続的改善が施されるとともに柔軟な対応が可能である。

【0124】

例えば、図7に示したフローチャートに基づいて処理する場合、マイクユニットのマイクが1本の場合にはビームフォーミングやブラインド音源分離は行わず、また、マイクの性能が高性能でない場合にはノイズ抑圧は1回だけ行い、マイクユニットのマイクが多数、例えば16個以上の場合にはビームフォーミングやブラインド音源処理を行い、ノイズ抑圧やエコーキャンセル等の処理を行うようになっている。

【0125】

<動作2>

図8に示すフローチャートの動作の主体は、クラウド側サーバ20のプロセッサ22である。図8に示したフローチャートの図7に示したフローチャートとの相違点は、ステップS18～ステップS21と、ステップS26～ステップS29とを入れ替えた点である。これは、クラウド側サーバで処理した場合に図7に示したフローチャートで処理しても結果が不十分な場合に行う処理の一例である。このような処理を行っても継続的改善が施

10

20

30

40

50



されるとともに柔軟な対応が可能である。

【0126】

<動作3>

図9に示したフローチャートは、音声入出力装置100の使用が終了し、電源スイッチをオフにした状態で受付、会議室、もしくはコールセンターのオペレータの机の上に載置し、翌営業日に再度電源スイッチをオンにした場合について想定したものである。

【0127】

電源スイッチがオンされると(ステップS91)、プロセッサ22は、室内環境、話者の変更、人数及びモード変更が無いかなかを判断し(ステップS92)、変更が無い場合(ステップS92/YES)、アプリを実行する(ステップS93)。

10

【0128】

プロセッサ22は、アプリを実行した後終了かなかを判断し(ステップS94)、終了の場合には終了し(ステップS94/YES)、終了しない場合にはステップS92に戻る(ステップS94/NO)。

【0129】

プロセッサ22は、室内環境、話者の変更、人数及びモード変更があった場合(ステップS92/NO)、クライアント側の音声入出力装置100の処理能力が十分かなかを判断する(ステップS95)。

【0130】

プロセッサ22は、クライアント側の音声入出力装置100の処理能力が十分であると判断した場合(ステップS95/YES)、音声入出力装置100で処理し(ステップS96)、クライアント側の音声入出力装置100の処理能力が十分でないと判断した場合(ステップS96/NO)、クラウド側で処理し(ステップS97)、ステップS98に進む。

20

【0131】

プロセッサ22は、モード設定信号が有るかなかを判断し(ステップS98)、外部からモード設定信号が有ると判断した場合(ステップS98/YES)、コールセンターモード、受付モード、会議モード、...のうちのいずれかのモードが設定され(ステップS99)、外部からモード設定信号が無いと判断した場合(ステップS98/NO)、ステップS14(図7参照。)に進む。

30

【0132】

ステップS92の処理により、無駄な処理が省略され効率が向上する。

【0133】

以上で説明した本発明に係る音声入出力装置100は、コンピュータで処理を実行させる制御プログラムによって実現されている。一例として、プログラムにより本発明の機能を実現する場合の説明を以下で行う。

【0134】

コンピュータが読み取り可能なプログラムであって、

コンピュータを、

音声に関する音声情報を入力する入力手段、

入力された音声情報に対し、識別処理が容易となるような前段処理を行う前段処理手段

40

、  
前段処理手段により処理された音声情報に所定の加工を施し、第1の情報に基づいてタスク処理を行い、タスク処理の評価が十分でない場合に第1の情報を修正し、評価が十分になるまで一連の処理を繰り返すことで最適化する最適化手段、  
として機能させるための音声情報処理システムのプログラムが挙げられる。

【0135】

これにより、プログラムが実行可能なコンピュータ環境さえあれば、どこにおいても本発明にかかる音声入出力装置100を実現することができる。

【0136】

50

このようなプログラムは、コンピュータに読み取り可能な記録媒体に記憶されていてもよい。

#### 【0137】

##### <記録媒体>

ここで、記録媒体としては、例えばCD-ROM、フレキシブルディスク(FD)、CD-R等のコンピュータで読み取り可能な記録媒体、フラッシュメモリ、RAM、ROM、FeRAM等の半導体メモリやHDDが挙げられる。

#### 【0138】

CD-ROMは、Compact Disc Read Only Memoryの略である。フレキシブルディスクは、Flexible Diskを意味する。CD-Rは、CD Recordableの略である。FeRAMは、Ferroelectric RAMの略で、強誘電体メモリを意味する。HDDは、Hard Disc Driveの略である。

10

#### 【0139】

尚、上述した実施の形態は、本発明の好適な実施の形態の一例を示すものであり、本発明はそれに限定されることなく、その要旨を逸脱しない範囲内において、種々変形実施が可能である。

#### 【産業上の利用可能性】

#### 【0140】

本発明は、主に音声による情報をもとに対象の状況を、詳細に知ることができるだけでなく、将来にわたり、発生が予測される事項について認識可能とした上で、関係者全般に対し情報を提供することを可能とすることで、主に音声による状況把握を必要とする場面に適用可能である。

20

#### 【符号の説明】

#### 【0141】

- 10・・・ネットワーク
- 20・・・クラウド側サーバ
- 21・・・データベース(DB)
- 22・・・プロセッサ
- 23・・・出力装置
- 24・・・入力装置
- 26・・・インターフェース
- 100・・・音声入出力装置100
- 101・・・筐体
- 102・・・LEDリング
- 103-1～103-16・・・PDMマイク
- 104・・・スピーカ群
- 104S・・・スコーカ
- 104T・・・ツイータ
- 105・・・反射板
- 106・・・電源ランプ
- 107・・・電源コード
- 201・・・拡張部
- 202・・・記憶部
- 203・・・マイクユニット
- 204・・・マイク制御部
- 205・・・信号処理部
- 206・・・通信部
- 207・・・音声発生部
- 208・・・非可聴音発生部

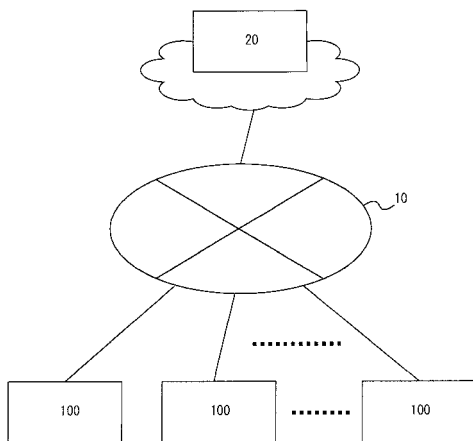
30

40

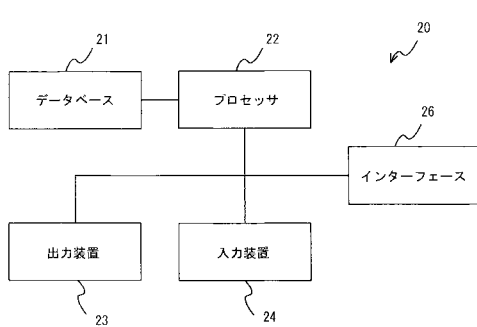
50

209 . . . 表示部

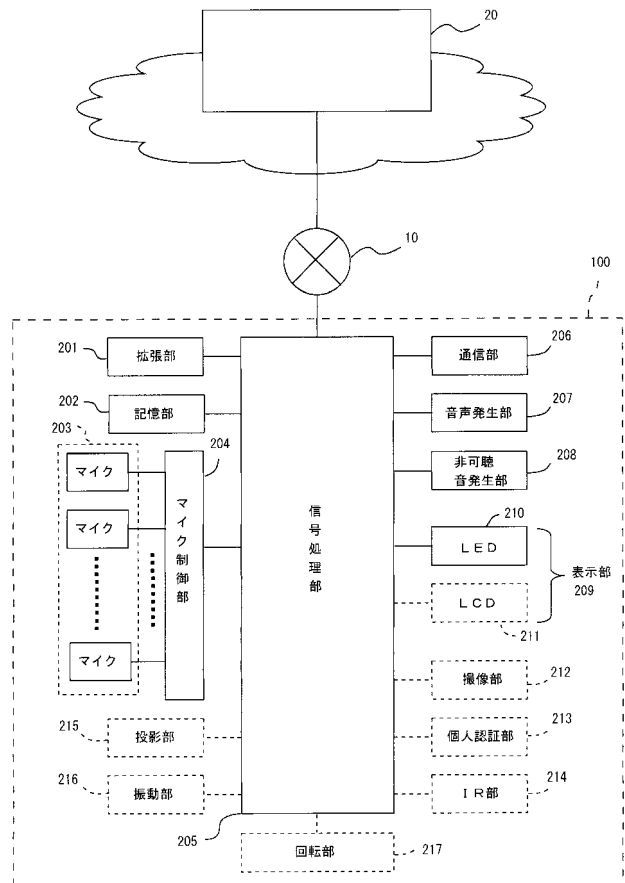
【図1】



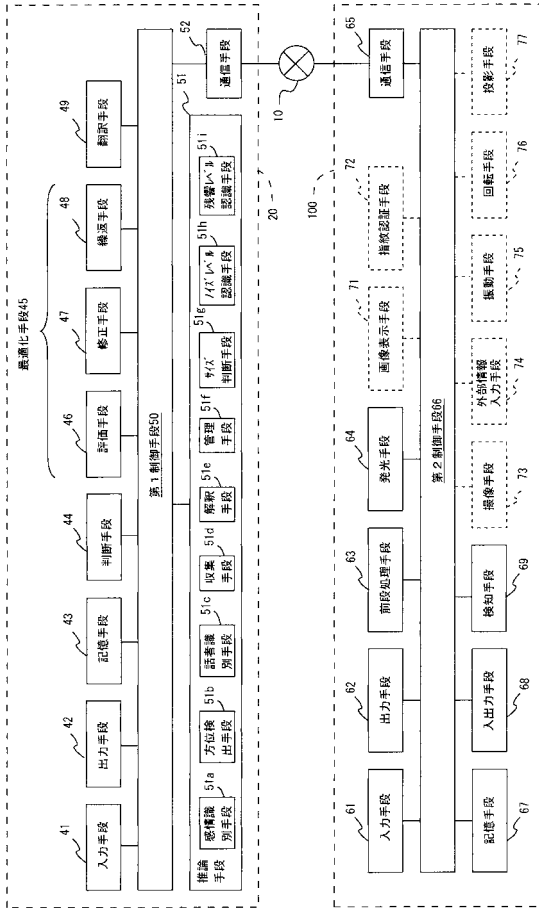
【図2】



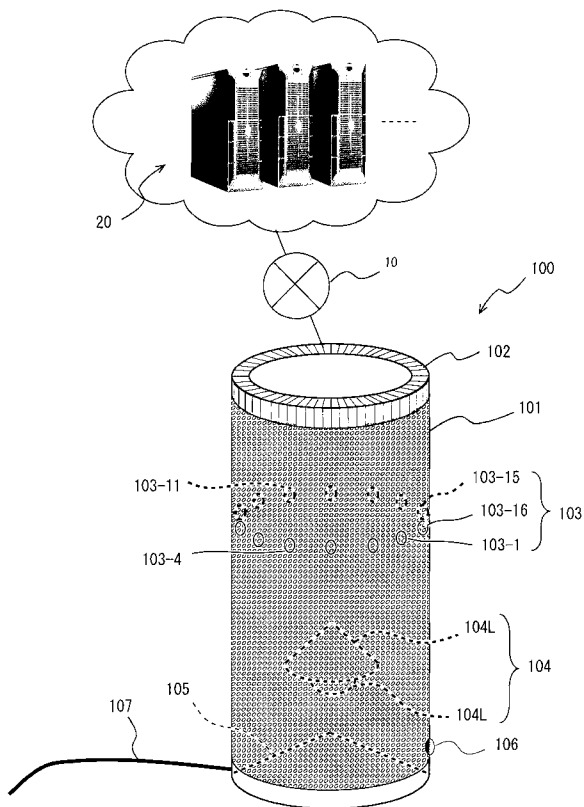
【図3】



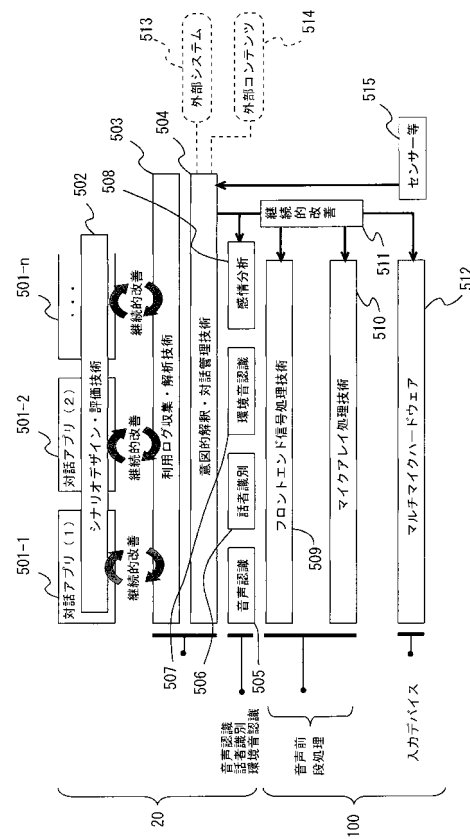
【図 4】



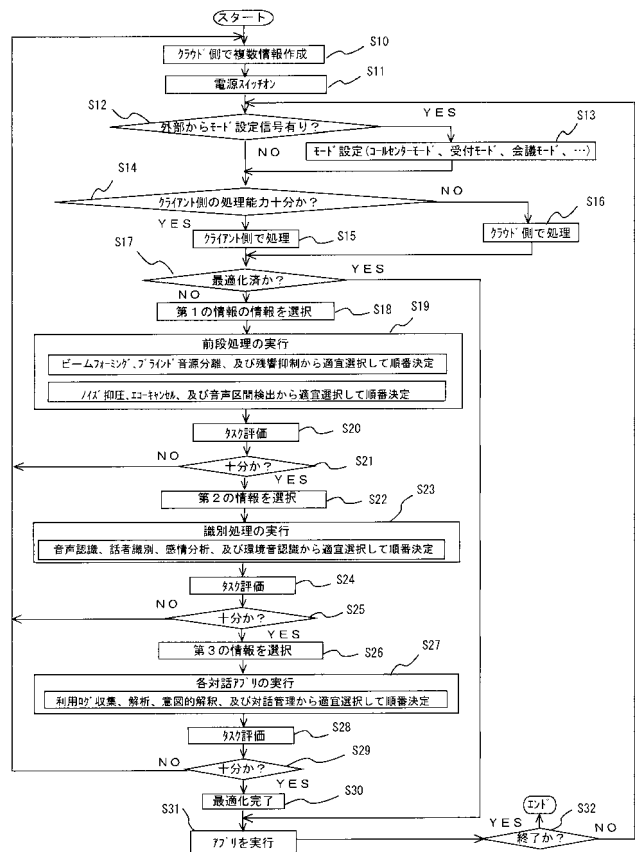
【図 6】



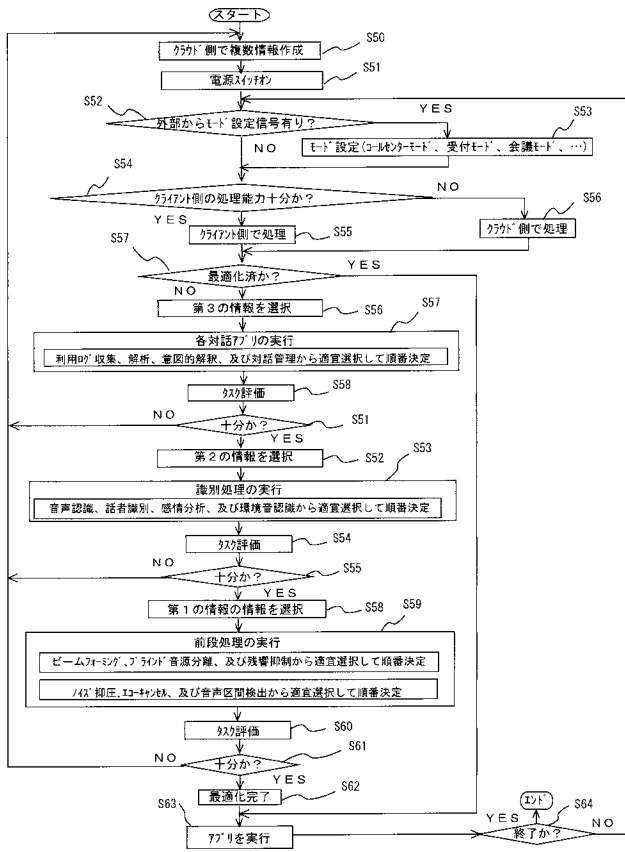
【図 5】



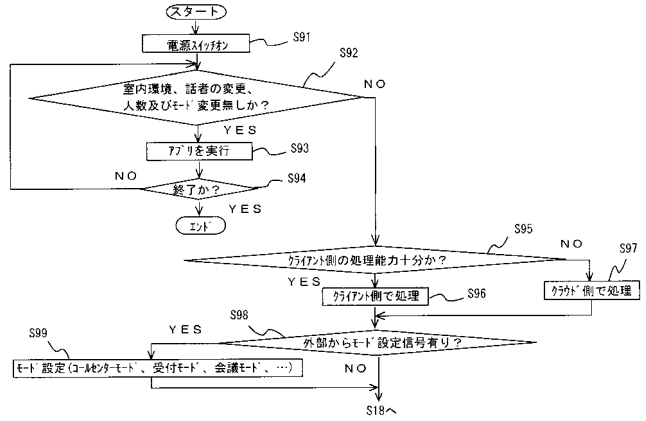
【図 7】



【図 8】



【図 9】



---

フロントページの続き

(51)Int.Cl.			F I			テーマコード(参考)
<b>G 0 6 F</b>	<b>3/16</b>	<b>(2006.01)</b>	<b>G 0 6 F</b>	<b>3/16</b>	<b>6 5 0</b>	