

国際知財司法シンポジウム2021 パネルディスカッション事例

- 仮想オペレータ事件 -

1. 事件

本件特許発明の進歩性について、特許無効審判で争われているものとする。

2. 背景説明

従来ユーザは店舗に足を運ぶ、あるいは電話などによって事業者と直接口頭で通話することにより、事業者からサービスの提供を受けていた。現在では、通信ネットワークの発展により、ユーザは通信ネットワークを経由してサービスの提供を受けることができるようになってきた。典型的な例としてインターネットバンキングが挙げられ、ユーザは通信端末を操作することにより、時間と場所を選ばずに、各種商取引が可能となっている。

最近では人工知能を備えた音声認識技術が発展し、ユーザからの要望に対する対応や苦情の処理など、従来事業者のオペレータが対処してきた業務に対する支援システムも開発されている。例えば、ユーザが通信端末に対して音声情報を入力し、入力された音声情報の解析結果に基づいてユーザの要求や質問、苦情の意図を抽出することで、ユーザに対するより適切かつ迅速な対応を支援するシステムが開発されている。

3. 本件特許発明

出願日 2016年2月25日

請求項の記載

「サーバと通信端末とを含む情報提供システムにおいて、
前記サーバは、前記通信端末から送信されたユーザの音声情報に対する回答メッセージ、あるいは前記回答メッセージを特定できない場合には問合せメッセージを前記通信端末に送信し、
前記通信端末は、ユーザ対応を行う従業員を模造した仮想オペレータを表示するように構成され、前記回答メッセージ、前記問合せメッセージを再生する際、前記回答メッセージ、前記問合せメッセージを再生しない時と比較し、前記仮想オペレータの一部が大きな動作を行うように前記仮想オペレータを表示する、
情報提供システム。」

明細書及び図面の記載

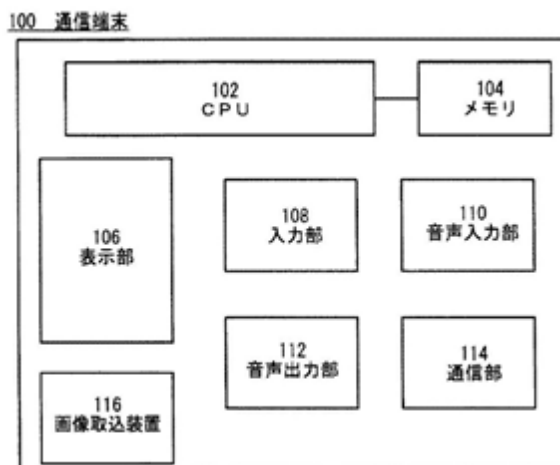
【課題】

本発明の課題は、携帯端末を介した自然な通話という手段を通じて、ユーザと事業者との距離感を低減させてより親密なユーザー事業者関係を構築することである。

【実施例】

図1に本実施例の情報提供方法で使用可能な通信端末100のシステム図を示す。通信端末100はユーザが操作可能な端末であり、携帯電話やタブレット、スマートフォンなどの携帯通信端末などが挙げられる。

【図1】



【図2】

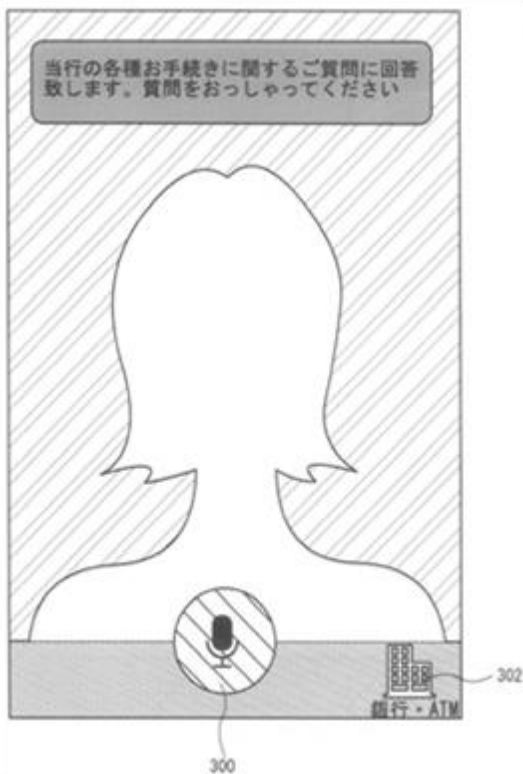


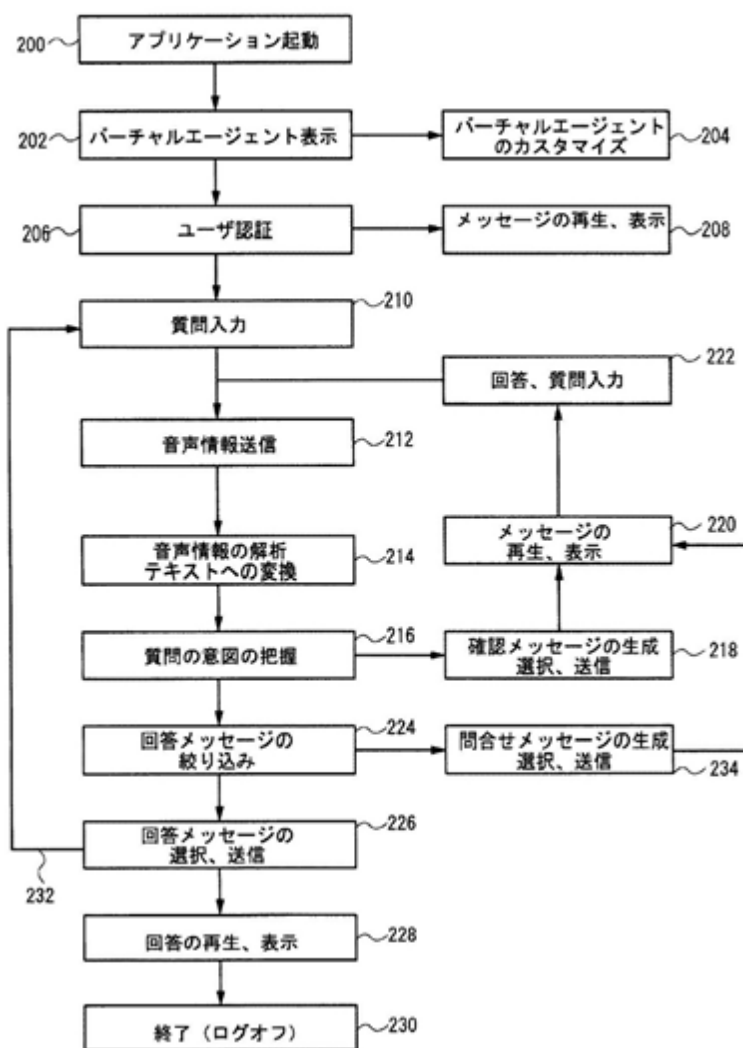
図2は仮想オペレータの一例であり、通信端末100の表示部106に表示した状態を例示したものである。仮想オペレータはユーザ対応を行う従業員などを模したものでよい。また、仮想オペレータが人間の自然な動作(例えば表情の変化、頭や手の動きなど)を行うように設定し、ユーザに親近感を抱かせるようにプログラムを構成してもよい。

図3に本実施例の情報提供方法のフローチャートを示す。ユーザがアプリケーションを起動すると(ステップ200), プログラムは通信端末の表示部106に仮想オペレータを表示する(ステップ202)。ユーザはまずアイコン300を操作して音声入力部110をアクティブにする。その後ユーザは, 事業者に対する質問を口頭で述べる(ステップ210)。

プログラムの命令に従い, 音声入力部110はユーザの質問を音声情報として通信部114へ伝達し, 通信部114は通信ネットワークを経由し, 音声認識サーバのインターフェースとなるAPIへ音声情報を送信する(ステップ212)。送信された音声情報は, 音声認識サーバにおいて解析され, テキストデータへ変換される(ステップ214)。テキストデータへ変換された音声情報は対話サーバで解析され, ユーザの質問の意図が理解, 把握される(ステップ216)。

ユーザの質問の意図を理解した後, 対話サーバは絞り込みエンジンを用いて回答メッセージの絞り込みを行う(ステップ224)。そして対話サーバ内のデータベースに蓄積された種々の回答メッセージから適合する回答メッセージを選択する(ステップ226)。選択された回答メッセージの文字情報は音声合成サーバで音声情報に変換され, 音声合成サーバのAPIを通してユーザの通信端末100へ送信される(ステップ226)。この時, 音声情報とともに文字情報として回答メッセージを通信端末100へ送信してもよい。プログラムの

【図3】



命令に従い、通信端末100が受信した音声情報を音声出力部112が再生する(ステップ228)。また同時に、回答メッセージを表示部106上にテキストとして表示してもよい。

データベースには事業者が作成した定型のメッセージを回答メッセージとして蓄積してもよく、対話サーバは蓄積されたメッセージから回答メッセージを選択することができる。ただし本実施例の情報提供方法はこれに限られず、回答メッセージを選択するだけでなく、人工知能としての機能を対話サーバに搭載し、ユーザとの対話を通じた学習によって新たな回答メッセージを作成し、これを音声合成サーバで音声情報に変換した後に通信端末100へ送信してもよい。すなわち、ユーザとの対話から新たに結論を獲得する機能、およびユーザとの対話履歴から学習する機能を対話サーバに付与し、これらの機能を利用して回答メッセージを作成するようにしてもよい。

ステップ224において、ユーザの質問に対する回答メッセージが絞り込めない場合、あるいは数多くの回答メッセージを選択せざるを得ない場合、ユーザに対する質問(問合せメッセージ)を生成あるいは選択して通信端末100に送信し、ユーザに対してさらなる情報提供を要請してもよい(ステップ234)。問合せメッセージは音声として、あるいはテキストとして、あるいはその両方として通信端末100で再生、表示される。ユーザは問合せメッセージに対して音声、あるいは文字で回答することができる(ステップ222)。問合せメッセージの送信とその回答というステップを繰り返すことにより、ユーザと仮想オペレータの対話を通じて、ユーザが要望する回答や情報を事業者が提供しえる数多くの情報の中から見つけ出すことが可能となる。

また、上述した種々のメッセージを音声出力部112で再生する際、仮想オペレータが実際に喋っているように見せるため、仮想オペレータの口や目を動かすようにしてもよい。あるいは手を動かすなど、説明を行うジェスチャーをするようにしてもよい。すなわち、メッセージが再生されていない時と比較し、仮想オペレータの一部がより大きな動作を行うようにプログラムを構成してもよい。

【効果】

上述したように本実施例では、仮想オペレータとユーザとの対話を通し、ユーザは様々な質問を事業者に伝えることができる。このため、ユーザはキーボードやタッチパネルを用いる文字入力操作から解放され、より簡便に種々の質問を事業者に伝達することが可能となる。また、音声合成技術を活用して仮想オペレータと対話するため、自然な対話を行うことができ、親密なユーザー事業者関係を構築することができる。

4. 文献1(公開特許公報)

公知日 2016年2月8日

文献1発明

「サーバ部とユーザ端末装置とを含む対話型処理システムにおいて、前記ユーザ端末装置からユーザ音声質問を前記サーバ部に送信し、前記サーバ部は、前記ユーザ音声質問に対応する想定回答又は聞き返し質問を前記ユーザ端末装置に送信し、前記ユーザ端末装置は前記想定回答又は前記聞き返し質問を、音声出力し、操作表示部に表示する、対話型処理システム。」

明細書及び図面の記載

【背景技術】

従来から、ユーザの音声情報に対してその回答内容をコンピュータで検索してユーザに返す音声対話装置が提供されている。しかし、従来の技術では、音声情報が曖昧な場合、音声対話装置が不適切な回答内容をユーザに返してしまう恐れがあった。

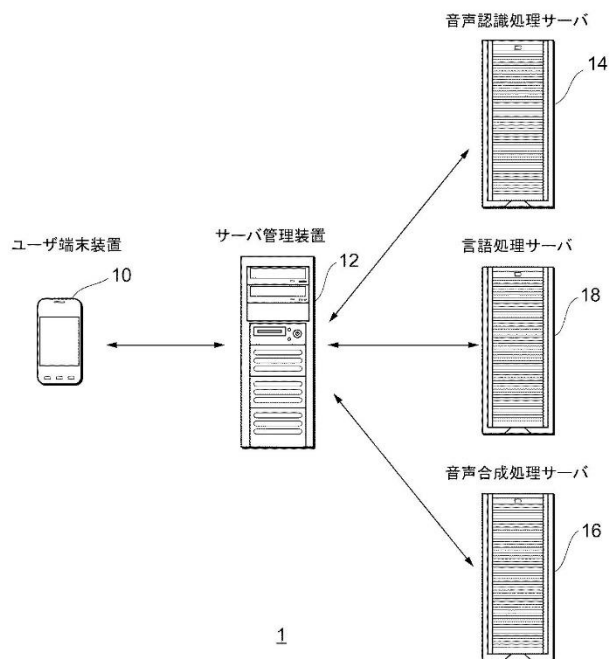
【課題】

本発明は、ユーザの質問に対して複数の回答内容が該当しうる場合に、ユーザに聞き返し質問をすることで、さらに具体的な質問内容に絞り込ませる対話型処理システムを提供する。

【実施例】

図1は、本実施例に係る対話型処理システムの構成図である。図1に示すように、本実施例に係る対話型処理システム1は、例示的に、ユーザ端末装置10、サーバ管理装置12、音声認識処理サーバ14、言語処理サーバ18、及び音声合成処理サーバ16を備えて構成されている。この対話型処理システム1は、上記構成にて対話型FAQシステムを実現するように機能する。すなわち、ユーザがユーザ端末装置10に対して音声で質問することで、音声認識処理サーバ14、言語処

【図1】

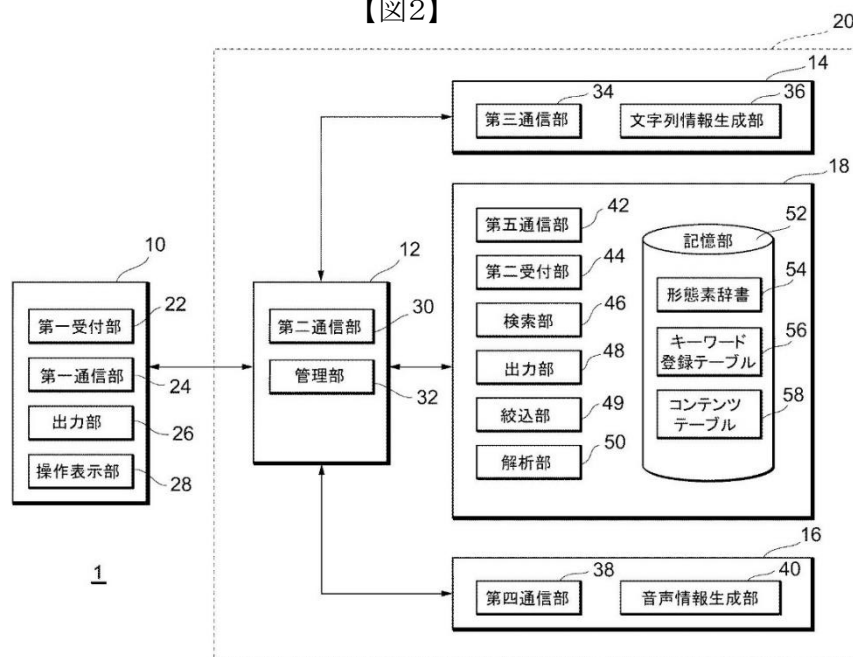


理サーバ18, 及び音声合成処理サーバ16の処理が行なわれて, その回答内容が返信されるシステムである。

ユーザ端末装置10は, スマートフォンなどの携帯電話, タブレット端末等であり, ユーザ情報, 音声情報(ユーザ質問), 及び文字列情報を送受信する通信機能を有する。

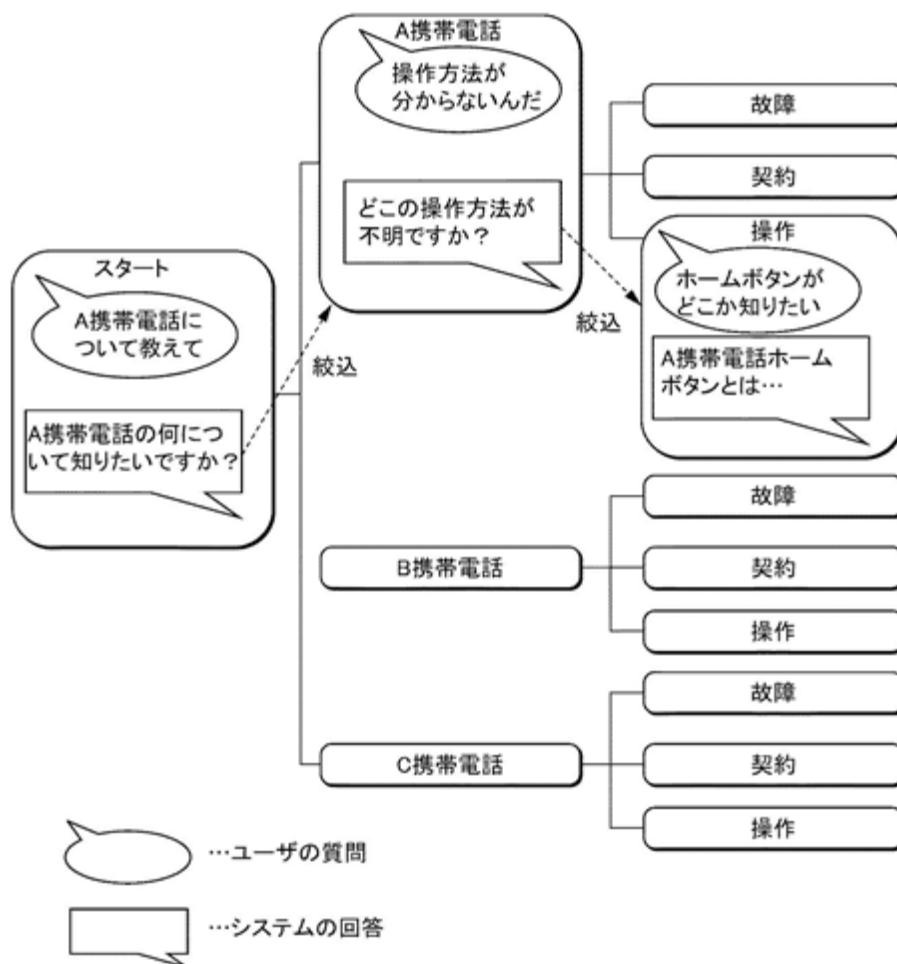
図2は, 対話型処理システム1の機能的構成の一例を示すブロック図である。図2に示すように, 対話型処理システム1は, ユーザ端末装置10と, サーバ部20を備える。ユーザ端末装置10は出力部26及び操作表示部28を備え, 出力部26は回答内容を音声に変換してスピーカから出力するとともに, 操作表示部28は回答内容を表示する。また, サーバ部20では, 検索部46は, 第二受付部44が受付けたユーザ文字列質問に基づき, コンテンツテーブル58内の複数のコンテンツ(複数の回答内容)の全部又は一部を検索して, ユーザ文字列質問に対応する, 複数の回答内容のうち想定回答及び聞き返し質問の何れか一方を取得する。具体的には, 検索部46は, コンテンツテーブル58を検索することによりユーザ質問文字列と各コンテンツとの相関関係を所定の計算式により算出して相互に比較して一つのコンテンツを特定する。一つのコンテンツを特定することができるのであれば, ユーザ文字列質問が応答するに足りる程度に具体的であることを意味しており, そのコンテンツに記録されている回答内容を取得する。また, 検索部46は, コンテンツテーブル58を検索しても複数のコンテンツが該当する可能性があるため一つのコンテンツを特定できない場合では, ユーザ文字列質問が曖昧であることを示唆しているため, さらに検索して曖昧なユーザ質問を絞るための聞き返し質問が格納されているコンテンツを特定し, 当該コンテンツに記録されている聞き返し質問を取得する。

【図2】



例えば、図3に示すように、ユーザ端末装置10からユーザが「A携帯電話について教えて」という音声による質問(ユーザ音声質問)をしたとする。このユーザ音声質問を入力した音声認識サーバ14では音声認識処理を実行し、ユーザ音声質問を対応する文字列に変換したユーザ文字列質問を言語処理サーバ18に送信する。言語処理サーバ18では、検索部46により、当該質問に対応するコンテンツテーブル58のうち特定されたコンテンツ59に記録されている聞き返し質問として回答内容、例えば「A携帯電話の何について知りたいですか?」という文字列が取得されて音声合成サーバ16に転送され、音声合成サーバ16が音声合成処理を行い、当該文字列に対応する音声としてユーザ端末装置10に出力される。

【図3】



これに応答して、例えば「操作方法を教えて」という更なるユーザ音声質問が入力され、音声認識サーバ14からこれに対応するユーザ文字列質問が転送されてくると、検索部46は、当該ユーザ質問に対応するコンテンツを絞り込む。そして、今回のユーザ質問に含まれる「操作方法」が記録されているコンテンツが今回のユーザ質問に最も相関関係の高

いレコードとして特定され、当該コンテンツに記録された回答内容である「どこの操作方法が不明ですか？」が取得されて出力される。

これに応答して、例えば「ホームボタンが知りたい」という更なるユーザ音声質問が入力され、音声認識サーバ14からこれに対応するユーザ文字列質問が転送されてくると、検索部46は、コンテンツテーブル58において当該質問に対応する「操作」が分類されているコンテンツを特定し、記録された回答内容である、「A携帯電話ホームボタンとは・・・」が取得されて出力される。

【効果】

本発明によれば、ユーザ質問と聞き返し質問とを繰り返して回答内容を絞り込んでいくことにより、曖昧な内容を含むユーザの質問に対して適切な回答内容を返すことができる対話型処理システムを提供することができる。

5. 文献2(公開特許公報)

公知日 2015年2月12日

文献2技術

「エージェントを表示装置に表示するナビゲーション装置において、当該エージェントが話しているように表示するため、待機中と比較して、回答側センターの応答音声データをスピーカから出力させる際に、当該エージェントの口を開くように当該エージェントを表示するナビゲーション装置。」

明細書及び図面の記載

【背景技術】

ナビゲーション装置、及び当該ナビゲーション装置と無線通信を実施する情報センターを備えた応答システムにおいて、ナビゲーション装置は、ユーザの音声データを取得して情報センターに当該音声データを送信する。一方、情報センターは、当該音声データに対して音声認識処理を実施して、その認識処理結果に応じた応答音声データをナビゲーション装置に返送する。そして、応答音声データの返送を受けたナビゲーション装置では、当該応答音声データに従った音声出力を行う。

一方、携帯電話機においても、携帯電話会社の管理する情報センターと携帯電話機が無線通信することによって、ユーザの音声入力に対して種々の情報を提供する応答システムが普及してきている。

そして、近年では、ナビゲーション装置と携帯電話機とを連携させて動作させる技術も開発されている。このように、ナビゲーション装置と携帯電話機とを連携して動作させる場合には、ナビゲーション装置を介して複数の応答システムが利用可能な構成となる。

【課題】

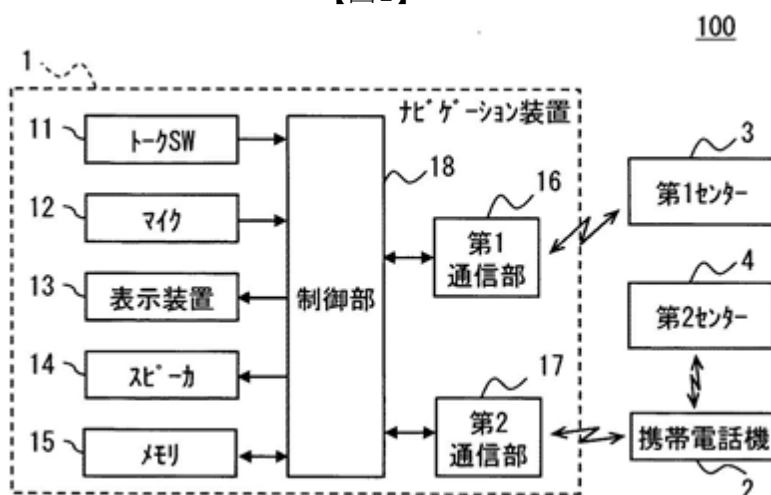
しかしながら、このような構成ではユーザは、ナビゲーション装置を操作するなどして、複数の応答システムのうち、目的に応じた応答システムを選択してから音声入力しなければならない。また、いったん応答システムを選択した後に、異なる応答システムを利用したい場合には、利用する応答システムを切り替えるための操作をしなければならない。

本発明は、この事情に基づいて成されたものであり、その目的とするところは、ユーザの音声入力に対する応答を行う応答システムを複数利用可能な場合に、応答を行わせる応答システムをユーザが選択する手間を省くことを可能にするナビゲーション装置を提供することにある。

【実施例】

図1に示すように応答制御システム100は、ナビゲーション装置1、携帯電話機2、第1センター3、および第2センター4を備えている。ナビゲーション装置1と第1センター3、ナビゲーション装置1と携帯電話機2、携帯電話機2と第2センター4とは、それぞれ公知の無線通信技術を用いてデータの送受信を実施する。

【図1】



第1センター3は、例えば自動車会社の情報センターであって、ナビゲーション装置の操作や、渋滞情報、自車両の操作に関連する質問に対応したり、ナビゲーション装置1の機能を利用するための音声入力による命令に対応する。第1センター3の動作の概要として

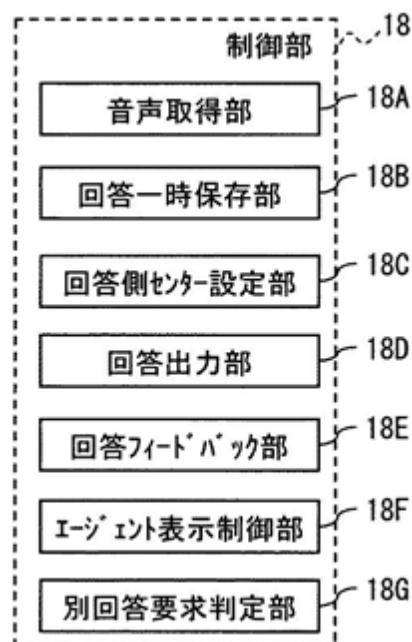
は、ナビゲーション装置1から送信されてくる音声データをもとに、音声認識処理を実施することでユーザの質問内容を解析する。そして、ユーザの質問に対して自センターが応答すべきかどうかを判定するとともに、ユーザの質問への応答となる応答音声データを生成してナビゲーション装置1に返送する。例えば、第1センター3は、ユーザの「コンビニはどこ？」という質問に対し、ユーザの現在地から最寄りのコンビニの位置を教えてくれるものである。

第2センター4は、例えば携帯電話会社の情報センターであって、当該携帯電話会社の管理する携帯電話網を利用する携帯電話機2のユーザに対して種々のサービスを提供する。第2センター4は、ユーザの質問に対して自センターが応答すべきかどうかの判定を実施しない点を除けば、第1センター3と同様の構成である。すなわち、ナビゲーション装置1から送信されてくる音声データをもとに、音声認識処理を実施することでユーザの質問内容を解析する。そして、ユーザの質問への応答となる応答音声データを生成してナビゲーション装置1に返送する。例えば、携帯電話会社が提供するサービスとしてスケジュール管理機能を想定した場合を例にとると、第2センター4は、ユーザの「今日の予定は？」という質問に対し、予め登録されているユーザのその日の予定を教えてくれるものである。

制御部18は、図2に示すように、種々の処理を実行するための機能ブロックを備える。回答側センター設定部18Cは、第1センター3から送られてくる判定結果信号に応じて、第1センター3および第2センター4のどちらから取得する応答音声データを、ユーザから入力された質問への回答としてスピーカ14から出力するかを決定する。

それぞれの応答システムが利用可能な状態となると、エージェント表示制御部18Fは、表示装置13に図3(A)に示すように、それぞれの応答システムに対応するエージェントA、Bを同時に表示させる。図3中のエージェントAは、第1センター3による応答システムに対応するエージェントの画像であり、エージェントBが第2センター4による応答システムに対応するエージェントの画像である。なお、エージェントとは、単純化された動物などのキャラクターである。

【図2】



エージェント表示制御部18Fは、公知の技術を用いて、それぞれの応答システムに対応するエージェント画像を生成し、それら複数のエージェントの画像を1つの画面に表示されるように合成または重畳して表示するものとする。

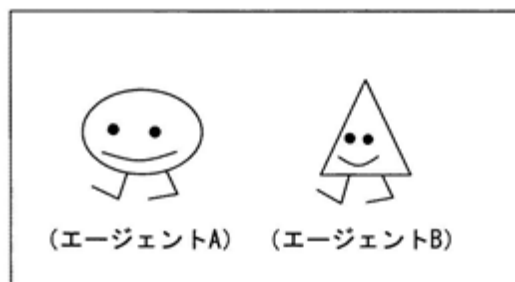
回答出力部18Dが、回答側センター設定部18Cによって回答側センターからの応答音声データをスピーカ14に音声出力させる際、図3の(B)や(C)に示すように回答側センターに対応するエージェントを相対的に大きく表示し、かつ、当該エージェントが話しているように表示するため口を開くように画像を表示させる。

【効果】

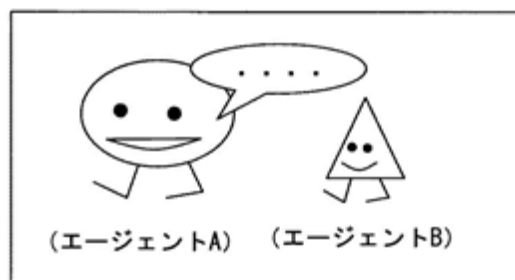
このような構成によれば、自動的に入力音声データの内容に応じて、ユーザの音声入力に対して応答すべきセンター(すなわち応答システム)を選択するため、ユーザは応答を行わせる応答システムを選択する手間を省くことができ、ユーザの利便性を向上させる事ができる。

さらに、エージェント表示制御部が、表示装置13に各センターに対応するエージェントA、Bを表示し、回答出力部18Dによってスピーカ14から音声出力されている応答音声データを生成したセンターに対応するエージェントが話しているように表示するため口を開くように表示させる。これによって、ユーザは、自身の音声入力に対してどちらのセンターが対応しているのかが一目で認識することができる。

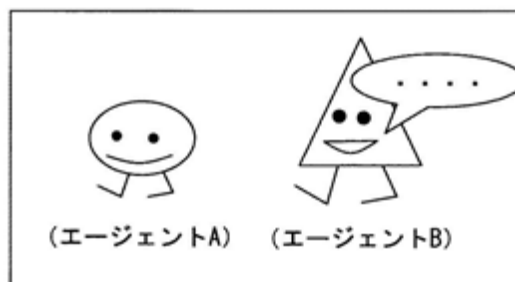
【図3】



(A) 待機中の表示例 (S103)



(B) 第1センター回答時の表示例



(C) 第2センター回答時の表示例

6. 文献3(公開特許公報)

公知日 2010年7月8日

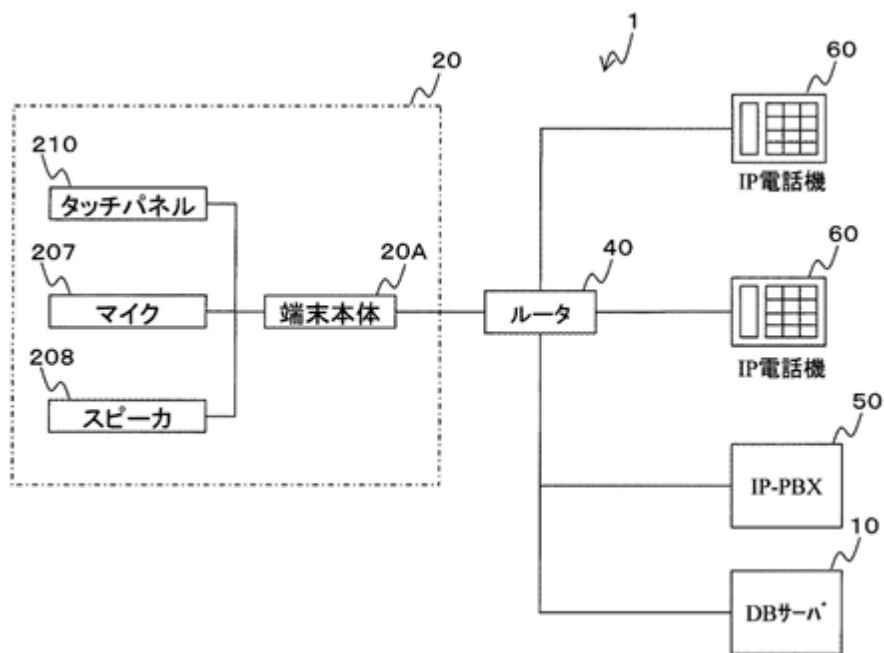
記載された周知の事項¹

「音声対話システムにおいて、ユーザ対応を行う従業員を模造した人物を静止画として表示すること。」

明細書及び図面の記載

図1は、本実施例の来訪者受付システム1の全体構成の概略構成を表すシステム構成図である。図1において、来訪者受付システム1は、例えば、ビルや会社その他の建造物への来訪者に対する受付業務を行うシステムである。この例では会社へ設置されている場合を例にとって説明する。

【図1】



来訪者受付システム1は、例えば会社の入口付近に設置された受付端末20(来訪者受付装置)と、周知のパーソナルコンピュータにより構成されるDBサーバ10と、会社の従業員それぞれに対応して設けられた複数のIP電話機60と、それら複数のIP電話機60の回線交換を行う周知の交換装置であるIP-PBX50とを有し、これらはすべてルータ40を介して接続されている。

¹ 当該周知の事項は、文献3以外にも多数の文献に記載されており、周知の事項とすることに対して両当事者の間に争いはない。

受付端末20は、端末本体20Aと、この端末本体20Aに接続された、タッチパネル210、マイク207、及びスピーカ208とを有している。

スピーカ208は、端末本体20Aから入力された音声信号を音声に変換して出力し、来訪者に対し報知を行う報知手段として機能する。マイク207は音声入力手段として機能し、来訪者により入力された音声を音声情報に変換し、端末本体20Aへ出力する。

図2は、上記受付端末20に備えられた、タッチパネル210及びマイク207の詳細外観構造の一例を表す斜視図である。

図2において、タッチパネル210は、この例では水平に設置されるベース212に対してアーム211を介し支持され、操作者(すなわちこの例では来訪者)の視線に対して直角となるように面方向が斜め上方を向いている。マイク207は、ベース212に対し先端を来訪者側へ向けるようにして略円弧状に配置されている。

【図2】

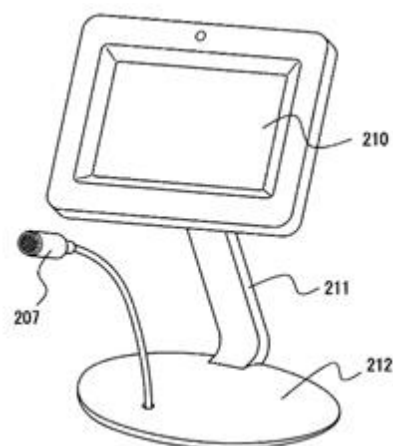
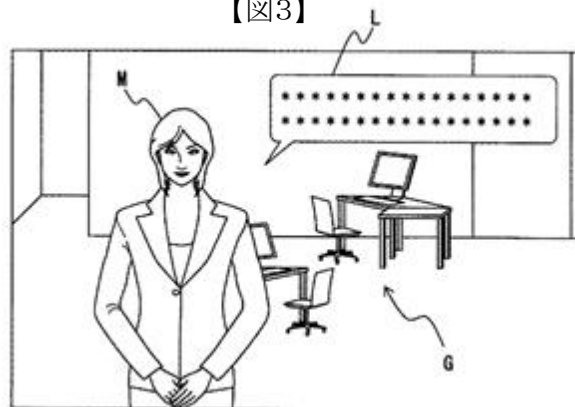


図3は、タッチパネル210における表示画面の一例を表す図である。この画面においては、描画プログラムによって生成された、受付業務を行う仮想人物Mがオフィス風の背景Gとともに静止画として表示される。また、スピーカ208から発話される音声に対応する文章L(図中では「***」で略記している)が併せて表示される。

【図3】



7. 各公知文献のポイント

文献1～3のポイントは、以下のとおり。

文献1

- ユーザ音声質問に対し、文字及び音声のみで応答する対話型処理システムの発明である。
- 通信端末は、スマートフォンなどの携帯電話、タブレット端末等である。
- 本件特許発明の課題(ユーザと事業者との距離感を低減させてより親密なユーザー事業者関係を構築すること)は記載されていない。
- 仮想オペレータを表示することは記載されていない。

文献2

- ユーザの音声入力に対する応答を行う複数の応答システムが通信端末を介して利用可能な構成が前提となっている。
- 通信端末は、ナビゲーション装置である。
- 本件特許発明の課題は記載されていない。
- 複数の応答システムが利用可能な状態で、各応答システムに対応するエージェント画像を1つの画面に表示することが記載されている。エージェントは、単純化された動物などのキャラクターであって、音声データをスピーカから出力させる際に、応答中の応答システムに対応するエージェントが話しているように表示するため口を開くように表示させる。

文献3

- 音声対話システムにおいて、ユーザ対応を行う従業員を模造した人物を静止画として表示することが周知の事項であることを示す周知例である。
- 通信端末は、来訪者受付システムの受付端末(専用の固定端末)である。
- 本件特許発明の課題は記載されていない。
- ユーザ対応を行う従業員を模造した人物が、静止画として表示される。

8. 一致点と相違点

本件特許発明と文献1発明との一致点、相違点は、以下のとおり。

一致点

「サーバと通信端末とを含む情報提供システムにおいて、
前記サーバは、前記通信端末から送信されたユーザの音声情報に対する回答メッセージ、あるいは前記回答メッセージを特定できない場合には問合せメッセージを前記通信端末に送信し、
前記通信端末は、前記回答メッセージ、前記問合せメッセージを再生する、
情報提供システム。」の点

相違点

本件特許発明は、通信端末において「ユーザ対応を行う従業員を模造した仮想オペレータ」を表示するのに対し、文献1発明は仮想オペレータを表示しない点。それに伴い、本件特許発明は「前記通信端末において前記回答メッセージ、前記問合せメッセージを再生する際、前記回答メッセージ、前記問合せメッセージを再生しない時と比較し、前記仮想オペレータの一部が大きな動作を行うように前記仮想オペレータを表示する」のに対し、文献1発明はそのような特定がない点

一致点、相違点について、両当事者の間に争いはない。

9. 論点

(1) 課題

論点 「文献1及び文献2に明示的に記載のない課題, 例えば『メディアコミュニケーションの円滑化を図ること』を文献をあげるまでもない周知の課題として認定し², 文献2技術を文献1発明に適用する動機とすることは妥当か？」

妥当:Y 妥当でない:N

(2) 阻害要因

論点 「複数の応答システムが前提の文献2技術を, 単一の対話型処理システムの文献1発明に適用することは妥当か³？」

妥当:Y 妥当でない:N

(3) 主引例への副引例及び周知の事項の適用

論点 「文献1発明に, 文献2技術及び文献3に例示される周知の事項を適用して, 本件特許発明の進歩性を否定することは妥当か⁴？」

妥当:Y 妥当でない:N

(4) 結論

論点 「(1)～(3)を勘案し総合的に本件特許は, 無効か？」

無効である:Y 無効でない(有効):N

「[審判実務者研究会報告書2020](#)」事例10を元に特許庁作成



以上

² 文献をあげるまでもない周知の課題とすることに対して両当事者の間に争いはない。

³ 特許権者からは, 「文献2においてエージェントが口を開くように表示させるのは, ユーザの音声入力に対してどちらのセンターが対応しているのかが一目で認識可能とするためであり, しかも文献2の課題(ユーザの音声入力に対する応答を行う応答システムを複数利用可能な場合に, 応答を行わせる応答システムをユーザが選択する手間を省くことを可能にする)からみても, 文献2技術を文献1発明のような単一の対話型処理システムに適用することには阻害要因がある」旨の主張がされている。

⁴ 特許権者からは, 「文献1発明に文献2技術を組み合わせた上で更に文献3に例示される周知の事項を組み合わせることはいわゆる容易の容易であり認められず, しかも, 文献2技術は(単純化された動物などのキャラクターである)エージェントの口が発言に対応して動作するところ, 文献3に例示される周知の事項は静止画として表示することを前提とするものであるから, 文献2技術に文献3に例示される周知の事項を組み合わせることは後知恵である」旨の主張がされている。