

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2010-197858
(P2010-197858A)

(43) 公開日 平成22年9月9日(2010.9.9)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 0 L 15/22 (2006.01)	G 1 0 L 15/22 3 0 0 Z	5 D 0 1 5
G 1 0 L 15/18 (2006.01)	G 1 0 L 15/22 2 0 0 V	
	G 1 0 L 15/18 4 0 0	

審査請求 未請求 請求項の数 4 O L (全 12 頁)

(21) 出願番号 特願2009-44460 (P2009-44460)
(22) 出願日 平成21年2月26日 (2009.2.26)

(出願人による申告)平成16~20年度、文部科学省、地域科学技術振興施策、委託研究(知的クラスター創成事業、岐阜・大垣地域ロボティック先端医療クラスター)、産業技術力強化法第19条の適用を受ける特許出願

(71) 出願人 304019399
国立大学法人岐阜大学
岐阜県岐阜市柳戸1番1
(71) 出願人 309004714
合同会社Urimina
岐阜県岐阜市玉宮町1-8-1
(74) 代理人 100098224
弁理士 前田 勳次
(74) 代理人 100140671
弁理士 大矢 正代
(72) 発明者 高橋 優三
岐阜県岐阜市柳戸1番1 国立大学法人岐阜大学内

最終頁に続く

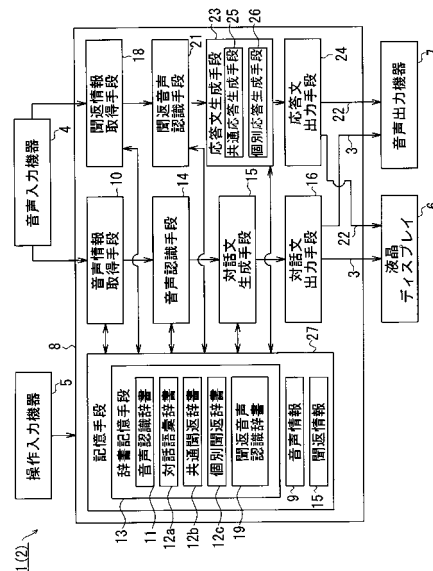
(54) 【発明の名称】 音声対話システム

(57) 【要約】

【課題】話者側からの聞返し及び確認に係る音声を認識し、これに基づいた回答や処理を実行することが可能な音声対話システムを提供することを課題とする。

【解決手段】音声対話システムの対話コンピュータ2は、音声情報9を音声情報取得手段10と、共通聞返辞書12b及び個別聞返辞書12cを有する対話語彙辞書12aを記憶する辞書記憶手段13と、対話語彙辞書12aに基づいて対話文3を生成する対話文生成手段15と、対話文3を音声出力する対話文出力手段16と、話者が発する聞返音声を検出し、聞返情報15として取得する聞返情報取得手段18と、共通聞返辞書12b及び個別聞返辞書12cに基づいて、応答する応答文22を生成する応答文生成手段23と、応答文22を音声出力する応答文出力手段24とを具備する。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

話者の音声を検出し、音声情報として取得する音声情報取得手段と、
前記音声情報に基づいて前記音声の内容を認識するための音声認識辞書、及び認識された前記音声に対応し、前記話者と対話するための語彙を含んで予め構築された対話辞書を記憶する辞書記憶手段と、
取得した前記音声情報及び前記音声認識辞書に基づいて前記音声の内容を認識する音声認識手段と、
認識された前記音声に対し、前記話者と対話するための対話文を前記対話辞書を利用して生成する対話文生成手段と、
生成された前記対話文を前記話者に対して出力する対話文出力手段と、
出力された前記対話文を認識した前記話者が発する音声による聞返し、前記対話文の内容に関する確認、及び前記対話文に関する要求を含む聞返り音声を検出し、聞返り情報として取得する聞返り情報取得手段と、
前記聞返り情報に基づいて前記聞返り音声の内容を認識するための聞返り音声認識辞書を記憶する聞返り辞書記憶手段と、
取得した前記聞返り情報及び前記聞返り音声認識辞書に基づいて前記聞返り音声の内容を認識する聞返り音声認識手段と、
認識された前記聞返り音声に対する応答文を前記対話辞書を利用して生成する応答文生成手段と、
生成された前記応答文を前記話者に対して出力する応答文出力手段と
を具備することを特徴とする聞返り対話システム。

10

20

【請求項 2】

前記対話辞書は、
前記話者との全ての対話において共通的に想定され、前記対話文の再出力要求、前記対話文の出力音量の調整、前記対話文の出力速度の調整を含む音声を前記聞返り音声として認識可能な複数の語彙が予め記憶された共通聞返り辞書を含んで構築され、
前記応答文作成手段は、
前記共通聞返り辞書を利用し、前記対話文の再出力、前記対話文の出力音量の増減、前記対話文の出力速度の加減を含んだ前記応答文を生成する共通応答生成手段をさらに具備することを特徴とする請求項 1 に記載の聞返り対話システム。

30

【請求項 3】

前記対話辞書は、
前記話者とのそれぞれ個別の対話において想定され、前記対話文を全く同一に繰り返すオウム返し音声、前記対話文を同意の語彙に置換し、前記対話文の内容の確認を行う置換確認音声、及び前記対話文に関連する単語を含み、前記対話文の内容の確認を行う関連単語音声を前記聞返り音声として認識可能な複数の語彙が予め記憶された個別聞返り辞書を含んで構築され、
前記応答文生成手段は、
前記個別聞返り辞書を利用し、前記オウム返し音声に対する応答、前記置換確認音声に対する認否、及び前記関連単語音声に対する認否を含む前記応答文を生成する個別応答生成手段をさらに具備することを特徴とする請求項 1 に記載の聞返り対話システム。

40

【請求項 4】

前記対話文出力手段及び前記応答文出力手段は、
前記対話文及び前記応答文の少なくとも一方が音声によって出力されることを特徴とする請求項 1 乃至請求項 3 のいずれか一つに記載の聞返り対話システム。

【発明の詳細な説明】**【技術分野】**

50

【0001】

本発明は、音声対話システムに関するものであり、特に、通常の会話の中で頻繁に行われる会話内容の確認のための所謂「聞返し」に係る音声を認識し、これに基づいてスムーズな対話を成立させることが可能な音声対話システムに関するものである。

【背景技術】

【0002】

従来から、話者が発声する音声をマイク等の音声入力手段によって受付け、これを音声情報として取得し、さらに当該音声情報を解析することにより音声を認識する音声認識技術の開発が進められている。係る技術によって、キーボードやマウス等の操作入力機器を利用することなく、音声入力のみでコンピュータ等を操作することが可能となり、コンピュータの操作に不慣れな高齢者や手肢の動きが制限される要介護者であっても、コンピュータ等の操作が容易に行えるようになる。また、カーナビゲーションシステム等に音声入力及び音声認識技術を採用することにより、運転者がハンドルから手を離すことなく、目的地の設定や案内の開始を音声によって実行することが可能となり、安全性を高めることができる。このように、音声入力及び音声認識に係る技術は、幅広い分野で使用され、さらに新たな技術分野での使用が期待されている。

10

【0003】

ここで、音声認識に係る技術において、話者の発声した音声を正確に認識できない場合、システム等の誤動作やエラーに繋がるおそれがある。そのため、音声入力の際の認識率（判別率）が実用化の点で特に重要となる。そこで、音声認識の過程において、話者の音声を認識する際に、適切な聞返し処理を行うことにより、誤認識を抑制することができる音声認識装置が開発されている（特許文献1参照）。これによると、音声入力によって特定された文章に含まれる語彙の確信度に依りて、話者に対する聞返しの要否を判定し、聞返しが必要と判断された場合に話者に対して聞返しの問い掛けを行うものである。これにより、認識の確信度が低い語彙について再確認を装置側が問い掛けることにより、当該語彙についての確信度を向上させることができ、最終的に高い認識率を維持した状態で音声認識を行うことが可能な音声認識装置を構築することができる。

20

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、上記に示した音声認識装置の場合、下記に掲げるような問題点を生じる可能性があった。すなわち、従来技術において、聞返しの要否を判断し、問い掛けを行う主体は、あくまでも音声認識装置側に限定され、カーナビゲーション等の音声入力時の認識率を向上させることを主たる目的とするものであった。そのため、聞返しに応答する対象も人間である話者に限定されていた。

30

【0005】

一方、音声認識技術を利用し、医学生等の訓練者が問診技術に係る訓練を実施可能な問診シミュレータ等の教育訓練装置の開発も進められている。この場合、訓練者から問診シミュレータに対して質問、若しくは質問に対する回答に対し、さらに行う再質問（次質問）が音声によって行われることがある。このとき、問診シミュレータからは、音声に係る質問に対し、音声または画面上に表示された文字情報として質問に対する回答が出力されていた。このとき、訓練生が問診シミュレータからの回答に係る音声出力を聞き逃す場合、或いは周囲の雑音やノイズによって一部の回答内容しか聞き取れなかった場合がある。しかしながら、従来型の問診シミュレータの場合、訓練生が回答に係る音声出力に対し、再出力を求めたり、音声出力された回答の内容の認識が正しいのか否かについての確認を行うことができなかった。すなわち、話者（訓練者）側からの内容確認や再出力の要求等の所謂「聞返し」を行う処理ができない場合があった。その結果、音声認識技術に基づいて行われる訓練者と問診シミュレータ等との間で交わされる対話がスムーズに実施されることができず、問診シミュレータとしての十分な効果が得られないことがあった。さらに、聞返しによる確認ができないため、誤った認識の内容で話者がさらに発言をすすめるた

40

50

め、音声対話システムにおける話者の音声の認識率が著しく低下するおそれがあった。

【0006】

そこで、本発明は、上記実情に鑑み、話者側からの聞返しを可能とし、当該聞返しに基づいた音声出力処理をコンピュータが実行することが可能な音声対話システムの提供を課題とするものである。

【課題を解決するための手段】

【0007】

上記の課題を解決するため、本発明の音声対話システムは、「話者の音声を検出し、音声情報として取得する音声情報取得手段と、前記音声情報に基づいて前記音声の内容を認識するための音声認識辞書、及び認識された前記音声に対応し、前記話者と対話するための語彙を含んで予め構築された対話辞書を記憶する辞書記憶手段と、取得した前記音声情報及び前記音声認識辞書に基づいて前記音声の内容を認識する音声認識手段と、認識された前記音声に対し、前記話者と対話するための対話文を前記対話辞書を利用して生成する対話文生成手段と、生成された前記対話文を前記話者に対して出力する対話文出力手段と、出力された前記対話文を認識した前記話者が発する音声による聞返し、前記対話文の内容に関する確認、及び前記対話文に関する要求を含む聞返音声を検出し、聞返情報として取得する聞返情報取得手段と、前記聞返情報に基づいて前記聞返音声の内容を認識するための聞返音声認識辞書を記憶する聞返辞書記憶手段と、取得した前記聞返情報及び前記聞返音声認識辞書に基づいて前記聞返音声の内容を認識する聞返音声認識手段と、認識された前記聞返音声に対する応答文を前記対話辞書を利用して生成する応答文生成手段と、生成された前記応答文を前記話者に対して出力する応答文出力手段と」を具備して主に構成されている。

10

20

【0008】

ここで、音声情報取得手段、音声認識辞書等の辞書記憶手段、及び音声認識手段は、周知の音声認識技術を利用することができる。また、聞返情報取得手段とは、音声認識処理に基づいて話者に対して音声または文字情報として出力した対話文に対する確認や要求等を含む聞返音声を検出し、聞返情報として取得するものである。すなわち、通常の会話においても、聞き逃しや聞き漏らし等が発声し、発声者に対して再度発声してくれるように頼むことや、内容の確認を行うことがある。つまり、“・・・ですね?”、“・・・と同じ意味ですね?”、或いは、“もう一度仰ってください”等の確認及び要求を含む聞返しの音声相当する。この話者の聞返しの発言(音声)を聞返情報として取得し、これを認識することが行われる。なお、対話文に対して聞返音声以外の音声が発せられた場合には、通常の音声情報として取得され、新たな対話文の生成及び出力処理が行われる。

30

【0009】

聞返情報を取得し、これを認識した後、対話辞書に基づいて応答文が生成され、出力される。ここで、応答文は、聞返音声に対応する内容を含み、例えば、聞返音声が“もう一度仰ってください”であれば、先に音声出力をした対話文を再び出力できるような処理を行う。ここで、話者の音声に対する対話文及び聞返音声に対する応答文の出力は、話者がある内容を視覚又は聴覚を通じて認識することができるもので足り、例えば、スピーカ等の音声出力機器を介して音声出力するものであっても、或いは液晶ディスプレイ等を利用して文字情報として出力するもののいずれであっても構わない。

40

【0010】

したがって、本発明の音声対話システムによれば、音声若しくは文字等による対話文の出力に対し、話者が何らかの確認や要求等を含む聞返音声を発声した場合、これを認識し、その内容に応じた応答文を音声若しくは文字等によって出力することが可能となる。その結果、話者及び音声対話システムの間で交わされる対話において、通常の会話の中で頻繁に行われる聞返しを認識し、これに対する応答をすることができる。対話中の誤認識の可能性が低くなり、スムーズな対話が可能となる。特に、音声を発する話者(人間)側からの聞返しに対し、システム側が対応することが可能となり、シミュレータ等の音声認識機器を構築する上で有用な効果を得ることができる。

50

【 0 0 1 1 】

さらに、本発明の音声対話システムは、上記構成に加え、「前記対話辞書は、前記話者との全ての対話において共通的に想定され、前記対話文の再出力要求、前記対話文の出力音量の調整、前記対話文の出力速度の調整を含む音声の前記聞返音声として認識可能な複数の語彙が予め記憶された共通聞返辞書を含んで構築され、前記応答文作成手段は、前記共通聞返辞書を利用し、前記対話文の再出力、前記対話文の出力音量の増減、前記対話文の出力速度の加減を含んだ前記応答文を生成する共通応答生成手段」を具備するものであっても構わない。

【 0 0 1 2 】

したがって、本発明の音声対話システムによれば、対話辞書は、各々の対話で共通に想定される語彙を予め記憶した共通聞返辞書を含んでいる。すなわち、音声出力された対話文の再出力（例えば、“もう一度仰って下さい”、“えっ？”、“はぁ？”等）の音声、“もっと大きな声で”或いは“もっとゆっくり話して”等の話者からの確認や要求は、一般の会話の中であっても頻繁に使用され、いずれの対話においても発生する頻度が極めて高いことが予想される。そのため、これらの聞返音声を、特に共通聞返辞書に予め記憶しておくことにより、話者から対話文に対して発せられたこれらの言葉を容易に認識することができ、これに対する応答文の生成及び出力を容易に行うことができる。

【 0 0 1 3 】

さらに、本発明の音声対話システムは、上記構成に加え、「前記対話辞書は、前記話者とのそれぞれ個別の対話において想定され、前記対話文を全く同一に繰り返すオウム返し音声、前記対話文を同意の語彙に置換し、前記対話文の内容の確認を行う置換確認音声、及び前記対話文に関連する単語を含み、前記対話文の内容の確認を行う関連単語音声を前記聞返音声として認識可能な複数の語彙が予め記憶された個別聞返辞書を含んで構築され、前記応答文生成手段は、前記個別聞返辞書を利用し、前記オウム返し音声に対する応答、前記置換確認音声に対する認否、及び前記関連単語音声に対する認否を含む前記応答文を生成する個別応答生成手段を」具備するものであっても構わない。

【 0 0 1 4 】

したがって、本発明の音声対話システムによれば、対話辞書は、各対話毎に個別に想定される語彙を予め記憶した個別聞返辞書を含んで構成されている。すなわち、通常の会話の中で一方が発した音声を、他方がそのまま繰返して発する所謂「オウム返し」の音声や、対話文の内容を認識した上で確認するための同意の語彙に言替えたり、関連する単語による音声を含む聞返音声が出力されることが予想される。係る聞返音声は、各対話毎に特有の内容を含むものであり、先に示した共通聞返辞書と区別して構築することが行われる。そのため、それぞれの対話毎に個別聞返辞書のみを変更し、共通聞返辞書をそのまま残す構成の対話辞書を構築することも可能である。

【 0 0 1 5 】

さらに、本発明の音声対話システムは、上記構成に加え、「前記対話文出力手段及び前記応答文出力手段は、前記対話文及び前記応答文の少なくとも一方が音声によって出力される」ものであっても構わない。

【 0 0 1 6 】

したがって、本発明の音声対話システムによれば、話者による音声及び聞返音声の発声、及び音声対話システム側からの対話文または応答文のいずれもを音声によって行うことが可能となり、両者の間の対話をスムーズに行うことができる。

【 発明の効果 】

【 0 0 1 7 】

本発明の効果として、通常の会話の中で頻繁に交わされる所謂「聞返し」に対応した音声認識処理に係るシステムを構築することができる。これにより、聞返しによる音声の誤認識を解消し、安定した対話を音声対話システムとの間で実施することができる。

【 図面の簡単な説明 】

【 0 0 1 8 】

10

20

30

40

50

【図 1】本実施形態の音声対話システムの概略構成を示す説明図である。

【図 2】音声対話システムにおける対話コンピュータの機能的構成を示すブロック図である。

【図 3】対話コンピュータの処理の流れを示すフローチャートである。

【図 4】対話コンピュータの処理の流れ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】

【0019】

以下、本発明の一実施形態である対話システム 1（以下、単に「対話システム 1」と称す）について、図 1 乃至図 4 に基づいて説明する。ここで、図 1 は本実施形態の対話システム 1 の概略構成を示す説明図であり、図 2 は対話システム 1 における対話コンピュータ 2 の機能的構成を示すブロック図であり、図 3 及び図 4 は対話コンピュータ 2 の処理の流れを示すフローチャートである。

10

【0020】

本実施形態の対話システム 1 は、図 1 乃至図 4 に示すように、音声認識装置として機能する対話コンピュータ 2 によって主に構成されている。ここで、対話コンピュータ 2 は、図 1 及び図 2 に特に示すように、市販のパーソナルコンピュータを利用して構成され、話者 S が発する音声 V 及び対話コンピュータ 2 からの音声 V 及び対話コンピュータから出力される音声 V に対応する対話文 3 についての確認及び聞返し等を行う聞返し音声 V' を取得するマイク等の音声入力機器 4 と、各種データの入力及び操作を行うためのキーボード等の操作入力機器 5 と、各種データ等を表示する液晶ディスプレイ 6 と、対話文 3 を出力するためのスピーカ等の音声出力機器 7 とがそれぞれコンピュータ本体 8 に接続されて主に構成されている。

20

【0021】

また、コンピュータ本体 8 の内部には、接続された各機器 4, 5, 6, 7 等との信号を送受するためのインターフェイスや制御機構、インターネット等のネットワーク環境への接続を可能とする通信機能、及び対話型対話システム 1 として機能させるための音声対話システム用ソフトウェア（図示しない）を内蔵するハードディスク等の記憶手段 27、音声対話システム用ソフトウェアに基づいて各種処理を行うための CPU を含む演算処理部等を含んで構成されている。これらのパーソナルコンピュータの構成及び機能については、既に周知のものであるため、ここでは説明を省略する。

30

【0022】

ここで、対話コンピュータ 2 は、その機能的構成として、図 2 に示すように、話者 S から対話コンピュータ 2 に対して発せられる音声 V を音声入力機器 4 を介して検出し、音声情報 9 として取得する音声情報取得手段 10 と、取得した音声情報 9 に基づいて音声 V の内容を認識するための音声認識辞書 11、認識された音声 V に対応し、話者 S と対話するための語彙を含んで予め構築された対話語彙辞書 12a、対話文 3 に対する話者 S による聞返し音声 V' を認識するための共通聞返し辞書 12b 及び個別聞返し辞書 12c を有する対話辞書 12 をそれぞれ記憶する辞書記憶手段 13 と、取得した音声情報 9 及び記憶された音声認識辞書 11 に基づいて音声 V の内容を認識する音声認識手段 14 と、認識された音声 V に対し、話者 S と対話するための対話文 3 を対話辞書 12 に含まれる対話語彙辞書 12a を利用して生成する対話文生成手段 15 と、生成された対話文 3 を話者 S に対し、スピーカ等の音声出力機器 7 を介して音声によって出力する対話文出力手段 16 と、音声出力された対話文 3 の内容について話者 S が応え、発声する内容確認等のための聞返し音声 V' を検出し、聞返し情報 17 として取得する聞返し情報取得手段 18 と、取得した聞返し情報 17 及び聞返し音声認識辞書 19 に基づいて聞返し音声 V' の内容を認識する聞返し音声認識手段 21 と、認識された聞返し音声 V' に対する応答文 22 を対話辞書 12 を利用して生成する応答文生成手段 23 と、生成された応答文 22 を話者に対して音声出力機器 7 を介して出力する応答文出力手段 24 とを主に具備している。また、辞書記憶手段 13 には、聞返し情報 17 に基づいて聞返し音声 V' の内容を認識するための聞返し音声認識辞書 19 が記憶されている。

40

50

【 0 0 2 3 】

さらに、本実施形態の対話システム 1 は、その他の機能的構成として、応答文生成手段 2 3 の中に、辞書記憶手段 1 3 に記憶された共通聞返辞書 1 2 b に基づいて、個々の対話文 3 に共通する聞返音声 V ' に係る応答文 2 2 を生成する共通応答生成手段 2 5 と、辞書記憶手段 1 3 に記憶された個別聞返辞書 1 2 c に基づいて、個々の対話文 3 に特有の聞返音声 V ' に係る応答文 2 2 を生成する個別応答生成手段 2 6 とを含んで構成されている。また、辞書記憶手段 1 3 を包含し、音声情報 9 等の各種情報を記憶するための記憶手段 2 7 を有している。

【 0 0 2 4 】

ここで、聞返音声 V ' に応答するための共通聞返辞書 1 2 b 及び個別聞返辞書 1 2 c について、さらに詳細に説明する。共通聞返辞書 1 2 b には、話者 S との対話において共通して想定される聞返しに係る音声を認識するための語彙が含まれている。例えば、対話文 3 に対し、“もう一度話して下さい”、“聞こえませんでした”等の対話文 3 の音声出力を再び対話コンピュータ 2 に対して要求するような音声（再出力要求音声）、或いは“もっと大きな声で言ってください”、“もっと小さな声で言ってください”等の音声出力の音量の増減の調整を要求する音声（出力音量の調整要求音声）、“もっとゆっくり言ってください”等の対話文 3 の再生速度の調整を要求する音声（再生速度の調整要求音声）を含む聞返音声を認識するための語彙が含まれている。これにより、対話コンピュータ 2 によって出力された対話文 3 を話者 S が聞き逃したり、一部を聞き漏らした場合であっても、対話コンピュータ 2 に対して聞返しを行い、話者 S が対話内容を誤認識することを防ぐことが可能となる。

【 0 0 2 5 】

一方、個別聞返辞書 1 2 c は、それぞれの対話において想定される特有の聞返に係る音声を認識するための語彙が含まれている。例えば、対話文 3 に対し、対話文 3 をそのまま繰り返す、所謂「オウム返し」の聞返音声 V ' を認識するもの、対話文 3 を同意の語彙で言換えまたは置換した言換音声、対話文に関連する単語を使用して確認した関連音声を含む聞返音声を認識するためのものである。すなわち、個別聞返辞書 1 2 c は、主に対話文 3 の内容についての確認を対話コンピュータ 2 に対して行うものを認識するためのものである。

【 0 0 2 6 】

次に、本実施形態の対話システム 1 における音声対話の処理の一例について、図 3 及び図 4 に基づいて説明する。初めに、話者 S は対話システム 1 の音声入力機器 4 に対して音声 V を発声する。これに対し、対話システム 1 の対話コンピュータ 2 は、話者 S の発した音声 V を音声入力機器 4 を介して検出する（ステップ S 1）。ここで、音声 V が検出される場合（ステップ S 1 において YES）、これを音声情報 9 として取得する（ステップ S 2）。一方、音声 V の検出がされない場合（ステップ S 1 において NO）、音声 V が検出されるまでステップ S 1 の処理を継続する。そして、対話システム 1 は、取得した音声情報 9 の内容を音声認識辞書 1 1 を利用して認識する（ステップ S 3）。ここで、係る認識は周知の音声認識技術を応用することによって達せられる。なお、音声認識辞書 1 1 には、予め話者 S が発すると想定される複数の語彙が登録され、取得された音声情報 9 に含まれる語彙と、当該音声認識辞書 1 1 に登録された語彙とを照合することにより、音声認識をすることが可能となっている。

【 0 0 2 7 】

これにより、対話システム 1 は、話者 S の発した音声 V の内容を認識することができる。そして、係る音声 V に対応し、話者 S と対話システム 1 との間で対話を成立させるための対話文 3 が生成される（ステップ S 4）。ここで、対話文 3 を生成するために、対話辞書 1 2 に含まれる対話語彙辞書 1 2 a が利用される。対話語彙辞書 1 2 a には、認識された音声 V に含まれる語彙と関連づけられた語彙が予め登録され、当該語彙を抽出し、文章化して作成することにより対話文 3 が生成される。係る対話文 3 の作成処理については、周知の音声認識技術及び文章作成技術を応用することができる。そして、電子データとし

10

20

30

40

50

て生成された対話文 3 が音声出力機器 7 (スピーカー等) を介して話者 S に対して音声出力される (ステップ S 5)。なお、音声認識による音声情報 9 及び作成した対話文 3 に係る電子データを記憶手段 2 7 に記憶することができる。

【 0 0 2 8 】

その後、対話システム 1 は、音声出力された対話文 3 に対し、話者 S から発せられる音声 V を検出する (ステップ S 6)。ここで、音声 V が検出される場合 (ステップ S 6 において Y E S)、当該音声 V が対話文 3 に対する確認や要求等を含む聞返音声 V' であるか否かを判定する (ステップ S 7)。すなわち、話者 S から発せられた音声 V が聞返音声 V' に相当する場合 (ステップ S 7 において Y E S)、これを聞返情報 1 7 として取得する (ステップ S 8)。一方、話者 S から発せられた音声 V が聞返音声 V' でない場合 (ステップ S 7 において N O)、換言すれば通常の音声 V として認識される場合、ステップ S 2 の処理に移行し、通常の音声 V による音声情報 9 として取得され、ステップ S 2 からステップ S 6 に係る処理が繰り返される。すなわち、本実施形態の場合、対話文 3 の出力に対して必ずしも聞返音声 V' が話者 S から発せるものではなく、通常の音声 V が発せられる可能性もある。そこで、聞返音声 V' の判定を行うことにより、通常の音声 V による対話が可能となる。なお、通常の音声 V が聞返音声 V' かの判定は、辞書記憶手段 1 3 に記憶された音声認識辞書 1 1 または対話辞書 1 2 に含まれる語彙によって行われる。このとき、それぞれの音声 V (聞返音声 V') の内容について認識処理は未だ実施されていない。

10

【 0 0 2 9 】

ここで、聞返情報 1 7 を取得した場合 (ステップ S 8)、取得した聞返情報 1 7 の内容を認識し、辞書記憶手段 1 3 の対話辞書 1 2 における共通聞返辞書 1 2 b または個別聞返辞書 1 2 c のいずれかに対応したものであるかの判定を行う (ステップ S 9)。ここで、聞返情報 1 7 が共通聞返辞書 1 2 b の内容に対応するものである場合 (ステップ 9 において Y E S)、共通聞返辞書 1 2 b を利用して聞返情報 1 7 の内容を認識する (ステップ S 1 0)。共通聞返辞書 1 2 b は、前述したように、話者 S とのそれぞれの対話において、共通化された確認のための問い掛けや、再出力等の要求を含む語彙が予め登録されている。そして、一方、聞返情報 1 7 が対話についての個別の内容に基づくものである場合 (ステップ S 9 において N O)、換言すれば、共通聞返辞書 1 2 b による認識がされない場合、個別聞返辞書 1 2 c を利用して聞返情報 1 7 の内容を認識する (ステップ S 1 1)。なお、聞返音声 V' に対する音声認識は、周知の音声認識技術を利用することができる。

20

30

【 0 0 3 0 】

その後、それぞれの辞書 1 2 b, 1 2 c によって認識された聞返音声 V' の内容に基づいて応答文 2 2 を生成し (ステップ S 1 2)、これを話者 S に対して音声出力機器 7 を介して音声によって出力する (ステップ S 1 3)。その後、係る応答文 2 2 に対する話者 S から発せられる音声 V を検出する (ステップ S 1 4)。ここで、音声 V を検出する場合 (ステップ S 1 4 において Y E S)、ステップ S 7 の処理に移行し、当該音声 V が再び聞返音声 V' であるか、或いは応答文 1 8 との対話に係る対話文 3 かを判定処理し、上記の対話処理を繰り返す。一方、応答文 1 8 に対する音声 V が検出されない場合 (ステップ S 1 4 において N O)、システム終了の指示の有無を検出する (ステップ S 1 5)。これにより、音声 V または聞返音声 V' による対話システム 1 と話者 S との間の対話が成立することとなる。

40

【 0 0 3 1 】

一方、ステップ S 6 において応答文 2 2 に対する話者 S からの音声 V の検出がない場合 (ステップ S 6 において N O)、ステップ S 7 からステップ S 1 4 の処理をキャンセルし、上記と同様にシステム終了の指示の有無を検出する (ステップ S 1 5)。そして、システム終了の指示が有る場合 (ステップ S 1 5 において Y E S)、システムを終了する (ステップ S 1 6)。一方、システム終了の指示がない場合 (ステップ S 1 5 において N O)、ステップ S 1 の処理に戻り、新たに話者 S によって音声 V が発せられるのを待機する。

【 0 0 3 2 】

これにより、話者 S と対話システム 1 との間で行われる対話において、話者 S が聞き逃

50

したり、聞き漏らした対話文 3 について、聞返音声 V ' を発することにより、当該聞返音声 V ' に相当する応答文 2 2 を出力することが可能となる。これにより、話者 S と対話システム 1 との間の対話を確実に行うことができる。すなわち、従来の対話形式の音声認識システムでは、聞返しや確認に係る音声を単純に認識するのみであり、これに対する対応を音声認識システムはとることが困難であった。これに対し、例えば、本実施形態の対話システム 1 を音声認識システムの一部機能として採用した場合、上述の聞返音声 V ' の認識及びこれに対応する応答文 2 2 の出力に係る処理は、通常の音声認識に係る処理と区別することができる。その結果、音声認識によって最終的に出力される認識結果の認識率が向上し、誤認識の少ない安定した文章を出力することが可能となる。

【 0 0 3 3 】

以上、本発明について好適な実施形態を挙げて説明したが、本発明はこれらの実施形態に限定されるものではなく、以下に示すように、本発明の要旨を逸脱しない範囲において、種々の改良及び設計の変更が可能である。

【 0 0 3 4 】

すなわち、本実施形態の対話システム 1 において、共通聞返辞書 1 2 b 及び個別聞返辞書 1 2 c の双方を有する対話辞書 1 1 を具備するものを示したがこれに限定されるものではなく、想定される語彙が限定されるものである場合には、これらの辞書を統合した聞返辞書を構築するものであっても構わない。さらに、必要に応じて対話語彙辞書 1 2 a とも統合し、一つの辞書として構築するものであっても構わない。

【 0 0 3 5 】

さらに、本実施形態の対話システム 1 において、対話文 3 の出力及び応答文 2 2 の出力を音声出力機器 7 を介して音声によって出力するものを示したが、これに限定されるものではない。すなわち、対話システム 1 の対話コンピュータ 2 からのこれらの出力は、話者 S が視覚または聴覚等を通じて認識可能であるものでよく、音声による以外の方法、例えば、液晶ディスプレイ 6 に表示して出力するものであっても構わない。さらに、音声及び表示の双方で行うものであっても構わない。

【 符号の説明 】

【 0 0 3 6 】

- 1 対話システム (音声対話システム)
- 2 対話コンピュータ
- 3 対話文
- 4 音声入力機器
- 7 音声出力機器
- 9 音声情報
- 1 0 音声情報取得手段
- 1 1 音声認識辞書
- 1 2 対話辞書
- 1 2 a 対話語彙辞書
- 1 2 b 共通聞返辞書
- 1 2 c 個別聞返辞書
- 1 3 辞書記憶手段
- 1 4 音声認識手段
- 1 5 対話文生成手段
- 1 6 対話文出力手段
- 1 7 聞返情報
- 1 8 聞返情報取得手段
- 2 1 聞返音声認識手段
- 2 2 応答文
- 2 3 応答文生成手段
- 2 4 応答文出力手段

10

20

30

40

50

- 2 5 共通応答生成手段
- 2 6 個別応答生成手段
- S 話者
- V 音声
- V' 聞返音声

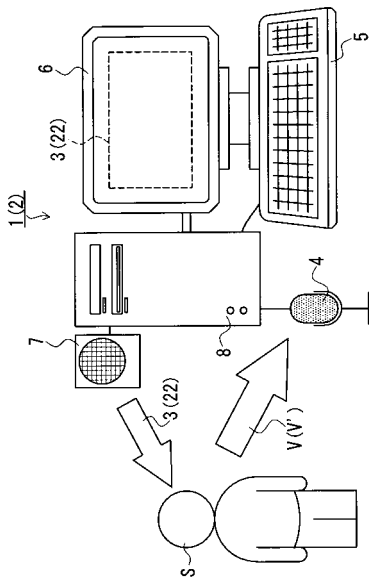
【先行技術文献】

【特許文献】

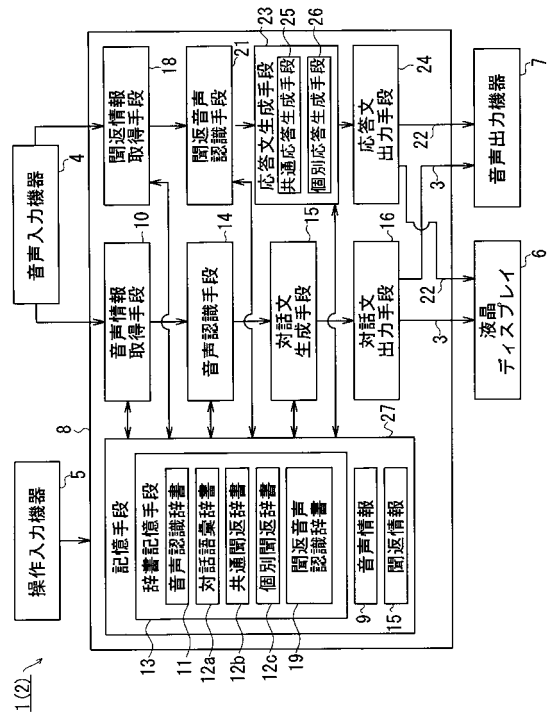
【0037】

【特許文献1】特開2008-52178号公報

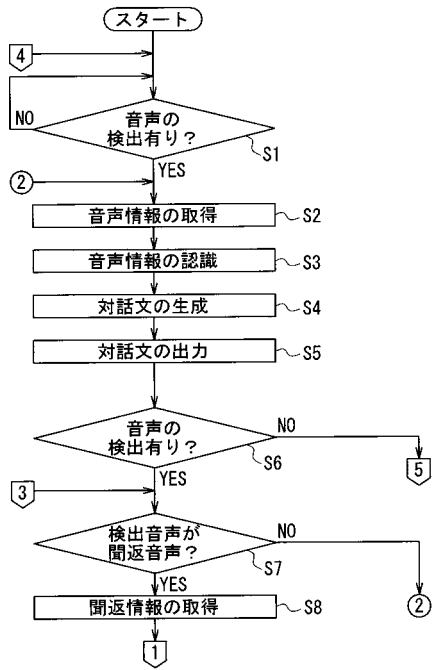
【図1】



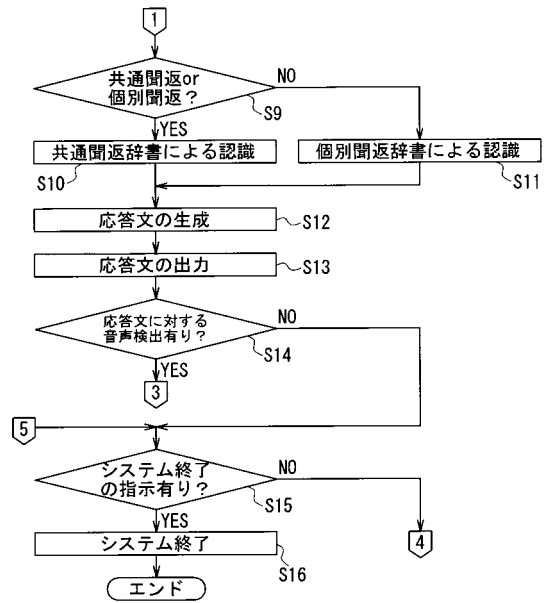
【図2】



【 図 3 】



【 図 4 】



フロントページの続き

(72)発明者 加藤 隆

岐阜県岐阜市玉宮町1 - 8 - 1 合同会社U r i m i n a内

Fターム(参考) 5D015 LL11