(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4987623号
(P4987623)
(45)発行日　平成24年7月25日(2012.7.25)　(24)登録日　平成24年5月11日(2012.5.11)

(51)Int.Cl.　　F I
G10L 15/22 (2006.01)　　G10L 15/22 470Z
G10L 15/00 (2006.01)　　G10L 15/00 200G
　　G10L 15/22 300Z

請求項の数 7　(全 20 頁)

(21)出願番号　特願2007-213828(P2007-213828)
(22)出願日　平成19年8月20日(2007.8.20)
(65)公開番号　特開2009-47920(P2009-47920A)
(43)公開日　平成21年3月5日(2009.3.5)
審査請求日　平成22年6月1日(2010.6.1)

(73)特許権者　000003078
株式会社東芝
東京都港区芝浦一丁目１番１号
(74)代理人　100089118
弁理士　酒井　宏明
(72)発明者　降幡　建太郎
東京都港区芝浦一丁目１番１号　株式会社東芝内
(72)発明者　知野　哲朗
東京都港区芝浦一丁目１番１号　株式会社東芝内
(72)発明者　釜谷　聡史
東京都港区芝浦一丁目１番１号　株式会社東芝内

最終頁に続く

(54)【発明の名称】ユーザと音声により対話する装置および方法

(57)【特許請求の範囲】
【請求項１】
入力した音声を認識し、認識結果の候補を複数生成する認識部と、
第１音声に対する複数の第１認識結果の候補を解析して、複数の第１認識結果の候補それぞれに対応する応答の候補と、第１認識結果の候補に対する応答の候補の確からしさを表す尤度とを生成する候補生成部と、
前記尤度が最大となる第１認識結果の第１候補に対する応答の候補を選択し、選択した前記第１認識結果の第１候補に対する応答の候補を表す語句を含む第１認識結果の第１候補に対する応答文を生成する応答文生成部と、
第１認識結果の第１候補に対する応答文を音声信号に変換した合成音声を出力する出力部と、
前記合成音声の出力中に第２音声が入力された場合、前記候補生成部で生成された第２音声に対する第２認識結果の候補を解析して、前記第１認識結果の第１候補に対する応答文に含まれる語句を修正した修正語句を生成する修正語句生成部と、
複数の第１認識結果の候補に対する応答の候補から、前記修正語句と同一の語句を含む第１認識結果の別の候補に対する応答の候補を取得し、第１認識結果の別の候補に対する応答の候補のうち前記尤度が最大の第１認識結果の別の候補に対する応答の候補を選択する選択部と、
選択された第１認識結果の別の候補に対する応答の候補の語句で前記応答文を更新する更新部と、を備え、

前記出力部は、前記応答文が更新された場合、更新前の前記応答文の合成音声に代えて、更新後の前記応答文の合成音声を出力し、
前記応答文生成部は、前記応答の候補を表す語句を、該語句の曖昧性が少ない順に文頭から含む前記応答文を生成すること、
を特徴とする音声対話装置。
【請求項２】
前記出力部は、前記応答文が更新された場合、更新前の前記応答文で出力されていない語句に対応する語句から更新後の前記応答文の合成音声を出力すること、
を特徴とする請求項１に記載の音声対話装置。
【請求項３】
入力した音声を認識し、認識結果の候補を複数生成する認識部と、
第１音声に対する複数の第１認識結果の候補を解析して、複数の第１認識結果の候補それぞれに対応する応答の候補と、第１認識結果の候補に対する応答の候補の確からしさを表す尤度とを生成する候補生成部と、
前記尤度が最大となる第１認識結果の第１候補に対する応答の候補を選択し、選択した前記第１認識結果の第１候補に対する応答の候補を表す語句を含む第１認識結果の第１候補に対する応答文を生成する応答文生成部と、
第１認識結果の第１候補に対する応答文を音声信号に変換した合成音声を出力する出力部と、
前記合成音声の出力中に第２音声が入力された場合、前記候補生成部で生成された第２音声に対する第２認識結果の候補を解析して、前記第１認識結果の第１候補に対する応答文に含まれる語句を修正した修正語句を生成する修正語句生成部と、
複数の第１認識結果の候補に対する応答の候補から、前記修正語句と同一の語句を含む第１認識結果の別の候補に対する応答の候補を取得し、第１認識結果の別の候補に対する応答の候補のうち前記尤度が最大の第１認識結果の別の候補に対する応答の候補を選択する選択部と、
選択された第１認識結果の別の候補に対する応答の候補の語句で前記応答文を更新する更新部と、を備え、
前記出力部は、前記応答文が更新された場合、更新前の前記応答文で出力されていない語句に対応する語句から、更新前の前記応答文の合成音声に代えて、更新後の前記応答文の合成音声を出力すること、
を特徴とする音声対話装置。
【請求項４】
前記出力部は、前記応答文に含まれる語句のうち、更新前の前記応答文で出力済みの語句が、更新された語句のうち最も文頭に近い語句より文末側に含まれる場合に、更新された語句のうち最も文頭に近い語句から更新後の前記応答文の合成音声を出力すること、
を特徴とする請求項１～３のいずれか１つに記載の音声対話装置。
【請求項５】
前記出力部は、前記応答文に含まれる語句のうち、更新前の前記応答文で出力済みの語句が、更新された語句のうち最も文頭に近い語句より文頭側に含まれる場合に、出力済みの語句の次に文末側に含まれる語句から更新後の前記応答文の合成音声を出力すること、
を特徴とする請求項１～３のいずれか１つに記載の音声対話装置。
【請求項６】
認識部が、入力した音声を認識し、認識結果の候補を複数生成する認識ステップと、
候補生成部が、第１音声に対する複数の第１認識結果の候補を解析して、複数の第１認識結果の候補それぞれに対応する応答の候補と、第１認識結果の候補に対する応答の候補の確からしさを表す尤度とを生成する候補生成ステップと、
応答文生成部が、前記尤度が最大となる第１認識結果の第１候補に対する応答の候補を選択し、選択した前記第１認識結果の第１候補に対する応答の候補を表す語句を含む第１認識結果の第１候補に対する応答文を生成する応答文生成ステップと、

出力部が、第１認識結果の第１候補に対する応答文を音声信号に変換した合成音声を出力する第１出力ステップと、
修正語句生成部が、前記合成音声の出力中に第２音声が入力された場合、前記候補生成ステップで生成された第２音声に対する第２認識結果の候補を解析して、前記第１認識結果の第１候補に対する応答文に含まれる語句を修正した修正語句を生成する修正語句生成ステップと、
選択部が、複数の第１認識結果の候補に対する応答の候補から、前記修正語句と同一の語句を含む第１認識結果の別の候補に対する応答の候補を取得し、第１認識結果の別の候補に対する応答の候補のうち前記尤度が最大の第１認識結果の別の候補に対する応答の候補を選択する選択ステップと、
更新部が、選択された第１認識結果の別の候補に対する応答の候補の語句で前記応答文を更新する更新ステップと、
出力部が、前記応答文が更新された場合、更新前の前記応答文の合成音声に代えて、更新後の前記応答文の合成音声を出力する第２出力ステップと、を備え、
前記応答文生成ステップは、前記応答の候補を表す語句を、該語句の曖昧性が少ない順に文頭から含む前記応答文を生成すること、
を特徴とする音声対話方法。
【請求項７】
認識部が、入力した音声を認識し、認識結果の候補を複数生成する認識ステップと、
候補生成部が、第１音声に対する複数の第１認識結果の候補を解析して、複数の第１認識結果の候補それぞれに対応する応答の候補と、第１認識結果の候補に対する応答の候補の確からしさを表す尤度とを生成する候補生成ステップと、
応答文生成部が、前記尤度が最大となる第１認識結果の第１候補に対する応答の候補を選択し、選択した前記第１認識結果の第１候補に対する応答の候補を表す語句を含む第１認識結果の第１候補に対する応答文を生成する応答文生成ステップと、
出力部が、第１認識結果の第１候補に対する応答文を音声信号に変換した合成音声を出力する第１出力ステップと、
修正語句生成部が、前記合成音声の出力中に第２音声が入力された場合、前記候補生成ステップで生成された第２音声に対する第２認識結果の候補を解析して、前記第１認識結果の第１候補に対する応答文に含まれる語句を修正した修正語句を生成する修正語句生成ステップと、
選択部が、複数の第１認識結果の候補に対する応答の候補から、前記修正語句と同一の語句を含む第１認識結果の別の候補に対する応答の候補を取得し、第１認識結果の別の候補に対する応答の候補のうち前記尤度が最大の第１認識結果の別の候補に対する応答の候補を選択する選択ステップと、
更新部が、選択された第１認識結果の別の候補に対する応答の候補の語句で前記応答文を更新する更新ステップと、
出力部が、前記応答文が更新された場合、更新前の前記応答文で出力されていない語句に対応する語句から、更新前の前記応答文の合成音声に代えて、更新後の前記応答文の合成音声を出力する第２出力ステップと、を含むこと、
を特徴とする音声対話方法。
【発明の詳細な説明】
【技術分野】
【０００１】
この発明は、入力した音声に応じた動作を実行することによりユーザと対話する装置および方法に関するものである。
【背景技術】
【０００２】
近年、音声認識、音声合成および対話理解といった要素技術の研究が進み、それらを組み合わせることによって、複雑なボタン操作やコマンド入力をせずとも、自然言語音声の

発話によって機械を操作できるような音声対話インターフェースが実用化されつつある。
【０００３】
また、デジタル家電やカーナビゲーションシステムの性能の向上に伴って、このような従来型のユーザ・インタフェースよりも高い処理性能が必要な音声対話インターフェースの実装も可能になりつつある。
【０００４】
しかし、上記のような各要素技術にはまだ多くの技術的課題が残されており、システムに対するユーザの入力音声を常に正しく解釈し、ユーザの要求を満たす動作の実行または応答の出力を可能とするほど精度の高いシステムの実現はきわめて困難である。
【０００５】
例えば、音声からユーザの要求意図を解釈するためには、最初に音声認識処理によって、音声波形から言語情報を抽出する必要がある。ところが、この音声認識処理でさえ、常に正しい結果が得られるわけではない。例えば、雑音環境下では、認識精度が著しく低下するという課題が存在する。
【０００６】
また、認識した言語情報（テキスト）から、形態素情報、構文情報を抽出し、さらに発話意図を解析する処理を行う必要があるが、いずれの過程でも誤りが生じる可能性が存在する。特に、発話意図を抽出するような対話理解には、文脈などを考慮した非常に高度な言語処理が必要である。このため、ユーザからの自由発話を入力できる音声対話処理システムが、ユーザの発話を常に正しく解釈し、曖昧性の発生を避けることは非常に困難である。
【０００７】
そこで、各処理段階における要素技術の改良とともに、ヒューマン・インターフェース（ＨＩ）を用いて、ユーザがシステムの解釈の曖昧性・誤りを訂正できるようにするという対策が採られている。
【０００８】
ところが、ユーザに対するシステムの解釈結果のフィードバックの仕方によっては、手順が複雑になる場合や、ユーザ入力－システムの解釈結果応答－ユーザの訂正入力－システムの解釈訂正－システム動作実行という一連の訂正処理の時間が増加する場合があり、ユーザにストレスを与える可能性がある。
【０００９】
例えば、ユーザの発話に対する複数の解釈候補が存在する場合に、各解釈候補をユーザに音声でフィードバックし、ユーザに正しい解釈候補を選択させる方法を考える。この方法では、解釈候補をテキストによって一覧表示することができないため、それぞれの解釈候補に対応する読み上げ音声を順番に出力する必要がある。このため、出力に時間がかかる上、ユーザがその音声を逐一聞いて確認するための処理負担も増大する。
【００１０】
これを避けるための方法としては、例えば、システムが第１位の解釈候補のみを出力し、ユーザからの訂正入力を受け付けるという方式が考えられる。しかし、単純に応答出力－訂正入力－確認応答出力という手順で訂正する方式では、訂正処理が煩雑になるという問題がある。
【００１１】
また、音声でフィードバックするのではなく、テキストで一覧表示してフィードバックするテキスト表示型インターフェースも考えられる。しかし、表示部が小さい場合は、スクロール等の操作が必要になるため、上記と同様に訂正処理が煩雑になるという問題が生じうる。
【００１２】
このように、音声対話型ＨＩでは、人（ユーザ）と機械間の対話を円滑に進められるような工夫が求められる。
【００１３】

例えば、特許文献１では、ユーザからの発話を音声認識する認識処理の過程で、認識誤りが生じたフレーズを自動的に検出し、検出部分のみを原言語話者にテキストまたは音声によって提示して訂正させることによって、円滑な訂正が可能な対話インターフェースを実現する技術が提案されている。この方法では、発話者に提示されるのは誤りフレーズのみであるため、文全体の確認や再入力が不要となり、訂正に要する時間を短くすることができる。

【００１４】

【特許文献１】特開２０００－２９４９２号公報

【発明の開示】

【発明が解決しようとする課題】

【００１５】

しかしながら、特許文献１の方法では、音声認識で誤認識が生じうるのと同様に、音声認識誤り箇所の特定にも誤りが生じうるため、誤認識箇所を正しく訂正できない場合があるという問題があった。また、特定された誤りフレーズ以外のフレーズを訂正することができないという問題があった。

【００１６】

このような問題を解消し、円滑な対話を実現するためには、誤り箇所のみでなく解釈結果全体を音声により確認し、音声により訂正可能とすることが望ましい。しかしこの場合も、解釈結果全体の音声をすべて出力してから訂正発話を受け付けるという一般的な確認・訂正方法では、対話の進行が妨げられるという問題が生じうる。

【００１７】

本発明は、上記に鑑みてなされたものであって、対話を阻害することなく誤り箇所を容易に訂正することができる装置および方法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】

【００１８】

上述した課題を解決し、目的を達成するために、本発明は、入力した音声を認識し、認識結果の候補を複数生成する認識部と、第１音声に対する複数の第１認識結果の候補を解析して、複数の第１認識結果の候補それぞれに対応する応答の候補と、第１認識結果の候補に対する応答の候補の確からしさを表す尤度とを生成する候補生成部と、前記尤度が最大となる第１認識結果の第１候補に対する応答の候補を選択し、選択した前記第１認識結果の第１候補に対する応答の候補を表す語句を含む第１認識結果の第１候補に対する応答文を生成する応答文生成部と、第１認識結果の第１候補に対する応答文を音声信号に変換した合成音声を出力する出力部と、前記合成音声の出力中に第２音声が入力された場合、前記候補生成部で生成された第２音声に対する第２認識結果の候補を解析して、前記第１認識結果の第１候補に対する応答文に含まれる語句を修正した修正語句を生成する修正語句生成部と、複数の第１認識結果の候補に対する応答の候補から、前記修正語句と同一の語句を含む第１認識結果の別の候補に対する応答の候補を取得し、第１認識結果の別の候補に対する応答の候補のうち前記尤度が最大の第１認識結果の別の候補に対する応答の候補を選択する選択部と、選択された第１認識結果の別の候補に対する応答の候補の語句で前記応答文を更新する更新部と、を備え、前記出力部は、前記応答文が更新された場合、更新前の前記応答文の合成音声に代えて、更新後の前記応答文の合成音声を出力し、前記応答文生成部は、前記応答の候補を表す語句を、該語句の曖昧性が少ない順に文頭から含む前記応答文を生成すること、を特徴とする。

【００１９】

また、本発明は、上記装置を実行することができる方法である。

【発明の効果】

【００２０】

本発明によれば、対話を阻害することなく誤り箇所を容易に訂正することができるという効果を奏する。

【発明を実施するための最良の形態】

【００２１】
以下に添付図面を参照して、この発明にかかる装置および方法の最良な実施の形態を詳細に説明する。
【００２２】
本実施の形態にかかる音声対話装置は、ユーザの入力音声を解釈し、解釈結果に対応する応答文を音声出力するとともに、応答文の出力中に入力された応答文を修正するための修正音声を利用して解釈結果と応答文を同時に更新し、更新後の応答文を出力するものである。
【００２３】
なお、以下では、ハードディクレコーダーやマルチメディアパソコンなどの、録画した放送番組等を録画再生可能なビデオ録画再生装置として音声対話装置を実現した例について説明する。なお、適用可能な装置はビデオ録画再生装置に限られず、ユーザの入力音声に対応する応答を出力するものであればあらゆる装置に適用できる。
【００２４】
図１は、本実施の形態にかかるビデオ録画再生装置１００の構成を示すブロック図である。図１に示すように、ビデオ録画再生装置１００は、主はハードウェア構成として、マイク１３１と、スピーカ１３２と、記憶部１２０と、を備えている。また、ビデオ録画再生装置１００は、主はソフトウェア構成として、受付部１０１と、対話処理部１１０と、出力部１０２と、録画再生部１０３とを備えている。
【００２５】
マイク１３１は、ユーザの発話した音声を入力するものである。また、スピーカ１３２は、応答を合成した合成音声などのデジタル形式の音声信号をアナログ形式の音声信号に変換（ＤＡ変換）して出力するものである。
【００２６】
記憶部１２０は、対話処理部１１０で生成されるアクション候補群、アクション断片、および応答フレーズリストなどの各種データ（詳細は後述）を記録するものである。記憶部１２０は、ＨＤＤ（Hard Disk Drive）、光ディスク、メモリカード、ＲＡＭ（Random Access Memory）などの一般的に利用されているあらゆる記憶媒体により構成することができる。
【００２７】
受付部１０１は、マイク１３１から入力された音声のアナログ信号に対してサンプリングを行い、ＰＣＭ（パルスデジタルコードモジュレーション）形式などのデジタル信号に変換して出力する処理を行うものである。受付部１０１の処理では、従来から用いられているＡ／Ｄ変換技術などを適用することができる。
【００２８】
対話処理部１１０は、ユーザから入力された音声に対応する応答および応答の内容を表す応答文を生成して出力することにより、ユーザとの対話処理を実行するものである。具体的には、対話処理部１１０は、まず、デジタル信号を音声認識してユーザの要求を解釈する。次に、対話処理部１１０は、その解釈結果に応じた応答の候補を生成する。さらに、対話処理部１１０は、最尤の候補に対応する応答文を生成する。
【００２９】
以下に、対話処理部１１０の詳細な機能と構成について説明する。図１に示すように、対話処理部１１０は、認識部１１１と、候補生成部１１２と、応答文生成部１１３と、修正語句生成部１１４と、選択部１１５と、更新部１１６と、を備えている。
【００３０】
認識部１１１は、受付部１０１が出力した音声のデジタル信号を音声認識してユーザの要求を表す認識結果の候補を生成するものである。具体的には、認識部１１１は、入力したデジタル信号を音声認識して、少なくとも１つの認識候補テキストからなる認識候補群を生成する。認識部１１１による音声認識処理では、ＬＰＣ分析、隠れマルコフモデル（ＨＭＭ：Hidden Markov Model）、ダイナミックプログラミング、ニューラルネットワー

ク、Ｎグラム言語モデルなどを用いた、一般的に利用されているあらゆる音声認識方法を適用することができる。

【００３１】

図２は、音声認識結果の一例を示す説明図である。図２は、「ＭＨＫで朝、英語講座を録ってね」を意味する日本語に対応する音声Ｉ０（「えむえっちけーであさえいごこうざをとってね」）に対する音声認識結果の例を示している。また、図２は、ラティス表現形式により音声認識結果を表した例を示している。

【００３２】

この例では、ノード２０１（「朝」）とノード２０２（「あさって」）との間、およびノード２０３（「英語講座を」）とノード２０４（「囲碁講座を」）との間に、それぞれ解釈の曖昧性が生じている。

【００３３】

なお、ラティスのノード間の線に付された数値は、ラティスの生成過程で統計的な共起頻度などから計算されたコストを表す。同図では、例えば、ノード２０５（「ＭＨＫで」）とノード２０１（「朝」）との間のコストが２であること、ノード２０２（「あさって」）とノード２０３（「英語講座を」）との間のコストが４であることが示されている。

【００３４】

認識部１１１は、このような認識結果のラティス表現およびコストを元に、確からしさを表す尤度が上位の所定数の候補を含む認識候補群を生成する。図３は、生成された認識候補文の一例を示す説明図である。図３は、図２のスタートノードからエンドノードまでのコストの総和に対応する尤度にしたがって、第１位候補から第４位候補まで順位付けを決定した結果を示している。

【００３５】

図３に示すように、認識部１１１は、認識候補を識別する候補番号と、認識候補の内容を表す候補テキストと、尤度とを対応づけた認識候補を生成する。なお、図３の例では、ユーザの要求に対応する正しい認識結果が第３位候補となっている。このように、音声認識処理では、第１位候補が誤りであっても、他の候補に正しい認識結果が含まれる場合が生じうる。

【００３６】

図１に戻り、候補生成部１１２は、このような状況を考慮し、最上位の候補に対する応答を生成するだけでなく、認識結果の候補それぞれについて、対応する応答の候補を生成するものである。なお、応答とは、ユーザの入力音声に対応して実行する処理または出力する内容を言う。本実施の形態は、ビデオ録画再生装置の例であるため、例えば、テレビ番組の再生・録画などの処理が応答となる。なお、以下では、応答をアクションといい、応答の候補をアクション候補という。

【００３７】

図４は、アクションの一例を示す説明図である。図４に示すように、アクションは、「操作」、「日時」、「チャンネル」、および「番組名」の４つの属性（以下、アクション属性という）を含む。なお、図４の表の２行目以降がアクションに相当する。

【００３８】

例えば、２行目は、「朝」（日時）に「ＭＨＫ」（チャンネル）の「英語講座」（番組名）を録画する（操作）というシステムの動作を表している。また、３行目は、「録画データ１」を再生するという動作を表す。ここで、「再生」は、ユーザ要求があった場合に、即時再生する動作を表すため、「日時」の値は空（「－」で表す）である。また、「チャンネル」の値も空である。

【００３９】

このように、アクションの表現形式は固定されるものではなく、少なくとも１つの語句によって、実行する処理や出力内容を表せればよい。図４の例では、少なくとも「操作」が設定されていればアクションの内容を特定することができる。

【００４０】

候補生成部１１２は、認識候補群に対して、形態素解析、構文解析、意味解析などの言語解析手法を適用することにより、ユーザの要求に対応するアクション候補群を生成する。このとき、候補生成部１１２は、音声認識処理で算出された認識候補それぞれの尤度および言語解析処理における確信度などから、各アクション候補についての尤度を算出し、各候補を順位付ける。

【００４１】

図５は、アクション候補群の一例を示す説明図である。図５は、図３に示した各認識候補に対するアクション候補の例を示している。図５に示すように、アクション候補は、識別子である「候補」と、図４と同様の「操作」、「日時」、「チャンネル」、および「番組名」と、「尤度」とを含む。図５の表中、２行目以降の各行がアクションに相当し、第１位候補であるＡｃｔ１から昇順に並べてある。図５の例では、簡単のため、言語処理が正しく行われているものと仮定し、アクション候補の尤度の値として、図３に示した認識候補の尤度値をそのまま用いている。

【００４２】

図１に戻り、応答文生成部１１３は、尤度が最大のアクション候補が、ユーザの要求を満たすか否かをユーザに確認するための応答文を生成するものである。具体的には、応答文生成部１１３は、アクション属性によって記述したテンプレートを用いて応答文を生成する。

【００４３】

図６は、テンプレートの一例を示す説明図である。図６に示すように、テンプレートＴは、記号「｛｝」で指定した変数部と、その他の固定部とを含んでいる。変数部は、記号「｛｝」内にアクション属性を指定することにより、各アクション候補の対応するアクション属性の属性値を当てはめることを表している。また、テンプレートＴは、記号「/」によって、それぞれ１つのアクション属性が含まれるようにフレーズ単位で分割される。このように、予めフレーズ単位に分割するのは、後述の出力部１０２が、応答文をフレーズ単位で順次出力できるようにするためである。なお、以下では、フレーズ単位で区切られた応答文を応答フレーズリストといい、Ｐ{Ｐ１～ＰＮ}（Ｎはフレーズ数）と表す。

【００４４】

なお、応答文の生成方法はテンプレートを用いた方法に限られるものではなく、文法規則や生成規則を用いて文を生成する方法などの従来から用いられているあらゆる方法を適用できる。

【００４５】

図７は、テンプレートを用いて生成された応答フレーズリストの一例を示す説明図である。図７は、図５のアクション候補ＣＡｃｔ１を、図６のテンプレートに適用して生成した応答フレーズリストを表している。各応答フレーズＰ１～Ｐ４は、この順で出力部１０２から音声出力される。

【００４６】

図１に戻り、修正語句生成部１１４は、後述する出力部１０２によって出力された応答文に対してユーザが発話した応答文の修正内容を表す修正語句を生成するものである。具体的には、修正語句生成部１１４は、修正のために発話された音声に対する認識部１１１による認識結果の候補を元に、アクションを構成する複数のアクション属性のうち少なくとも１つに対応する属性値を含むアクション断片を修正語句として生成する。

【００４７】

ユーザが応答文を修正する場合、応答文のすべてを再度発話するのではなく、修正部分のみを発話する場合がある。すなわち、ユーザの発話に、アクションの全てのアクション属性（操作、日時、チャンネル、番組名）が含まれない場合がある。このような場合でも、修正語句生成部１１４は、認識結果の候補から、少なくともアクション属性の一部を抽出することができる。そして、このようにして抽出されたアクション属性の属性値は、ユーザが要求する修正内容を表すため、修正語句生成部１１４は、この属性値を修正語句として生成する。

【００４８】
図８は、認識部１１１により生成された認識候補文の別の例を示す説明図である。図８は、図７に示す応答フレーズを含む応答文に対して修正を要求するためユーザが発話した音声であり、アクション属性のうち「日時」を修正するために発話した、「朝だよ」を意味する日本語の入力音声Ｉ１（「あさだよ」）に対する音声認識結果の例を示している。また、図８は、認識結果の候補として唯一の候補（「朝だよ」）が生成されたことを示している。
【００４９】
このような認識結果に対し、修正語句生成部１１４は、アクション属性「日時」の値が「朝」であるという情報をアクション断片として抽出する。図９は、このようにして生成されたアクション断片の一例を示す説明図である。図９は、上述の入力音声Ｉ１から生成されたアクション断片の例である。
【００５０】
なお、修正語句生成部１１４と候補生成部１１２とは、アクション属性の一部のみを含むアクション断片を生成するか、すべてを含むアクション候補を生成するかが異なるのみである。すなわち、認識結果に対して、形態素解析、構文解析、意味解析などの言語解析手法を実行してユーザの要求を解釈する処理手順は共通する。したがって、両者のうちいずれか一方を他方に統合するように構成してもよい。
【００５１】
選択部１１５は、アクション候補群から、アクション断片の属性値を全て含むアクション候補群を選択し、選択したアクション候補群の中から最も尤度の大きい候補を新たな第１位候補として選択するものである。
【００５２】
例えば、図５に示すようなアクション候補群が生成され、さらに図９に示すようなアクション断片（以下、アクション断片ＳＥＧ１という）が生成されたとする。この場合、選択部１１５は、図５のアクション候補群の中で、属性「日時」がアクション断片ＳＥＧ１（（当日）朝）と一致するアクション候補を探す。図５の例では、選択部１１５は、ＣＡｃｔ３およびＣＡｃｔ４を取得することができる。次に、選択部１１５は、ＣＡｃｔ３およびＣＡｃｔ４のうち、尤度の大きい方を新たに第１位候補として選択する。この例では、ＣＡｃｔ３の尤度＝０．２＞ＣＡｃｔ４の尤度＝０．１であるため、ＣＡｃｔ３が選択される。
【００５３】
更新部１１６は、選択部１１５により選択されたアクション候補を元に応答フレーズリストを更新するものである。具体的には、更新部１１６は、まず、選択部１１５が新たに選択したアクション候補（以下、新候補という）と、選択前の第１位のアクション候補（以下、旧候補という）との間で、すべてのアクション属性値を比較する。そして、更新部１１６は、不一致部分に対応する新候補のアクション属性を抽出する。
【００５４】
図１０は、旧候補の一例を示す説明図である。また、図１１は、新候補の一例を示す説明図である。図１０および図１１の例では、アクション属性「日時」および「番組名」が相違しているため、更新部１１６は、これらのアクション属性を抽出する。
【００５５】
次に、更新部１１６は、旧候補から生成した応答フレーズリストのうち、抽出したアクション属性に対応する応答フレーズを、新たな属性値で更新する。図１１の例では、更新部１１６は、属性値１１０１（（当日）朝）および属性値１１０２（英語講座）を新たな属性値として取得する。そして、更新部１１６は、生成済みの応答フレーズリストの対応する応答フレーズの内容を新たな属性値で変更する。
【００５６】
図１２は、更新された後の応答フレーズリストの一例を示す説明図である。図１２は、図７の応答フレーズリストを、図１１に示すようなアクション候補の属性を用いて更新し

た後の応答フレーズリストを表している。

【００５７】

なお、上述のように、候補生成部１１２は、事前にすべての認識結果の候補に対応するアクション候補を生成している。このため、アクションを修正する場合は、選択部１１５が、ユーザの修正発話に応じて、生成済みのアクション候補から、より適切なアクション候補を選択するだけでよい。すなわち、応答文に対するユーザの修正発話に応じて、応答文（応答フレーズリスト）だけでなくアクション候補を同時に修正することが可能となる。

【００５８】

出力部１０２は、応答文生成部１１３によって生成された応答文、または更新部１１６によって更新された応答文を音声信号に変換した合成音声を生成し、合成音声をスピーカ１３２に出力するものである。

【００５９】

具体的には、出力部１０２は、まず、応答文を構成する各文字列を音声信号に変換する音声合成処理を行う。出力部１０２による音声合成処理は、音声素片編集音声合成、フォルマント音声合成、音声コーパスベースの音声合成などの一般的に利用されているあらゆる方法を適用することができる。そして、出力部１０２は、生成した音声信号をＤＡ変換してスピーカ１３２に出力する。

【００６０】

また、出力部１０２は、応答文が更新された場合、更新後の応答文をいずれの部分から出力するかを特定する。具体的には、出力部１０２は、更新前の応答文で出力されていない応答フレーズを特定し、特定した応答フレーズから更新後の応答文の合成音声を出力する。

【００６１】

録画再生部１０３は、決定されたアクション、すなわち、尤度が最大のアクション候補を実行するものである。例えば、録画再生部１０３は、図５のＣＡｃｔ３が最尤のアクション候補として選択された場合、ＣＡｃｔ３の各アクション属性に従い、指定された日時に、指定されたチャンネルの指定された番組名の番組を録画するアクションを実行する。

【００６２】

なお、録画再生部１０３などのような実際のアクションを実行する構成部を外部装置に備えるように構成してもよい。この場合は、決定したアクションに関する情報を音声対話装置から外部装置に出力し、外部装置はこの情報を参照してアクションを実行するように構成する。

【００６３】

次に、このように構成された本実施の形態にかかるビデオ録画再生装置１００による音声対話処理について図１３を用いて説明する。図１３は、本実施の形態における音声対話処理の全体の流れを示すフローチャートである。

【００６４】

まず、受付部１０１は、マイク１３１から入力音声Ｉ０が入力されたか否かを判断する（ステップＳ１３０１）。入力音声Ｉ０が入力されていない場合は（ステップＳ１３０１：ＮＯ）、入力されるまで処理を繰り返す。

【００６５】

入力音声Ｉ０が入力された場合（ステップＳ１３０１：ＹＥＳ）、認識部１１１は、入力音声Ｉ０を音声認識し、認識候補群を生成する（ステップＳ１３０２）。次に、候補生成部１１２が、認識候補群の各候補について、対応するアクション候補を求め、アクション候補群ＣＡｃｔ｛ＣＡｃｔ１～ＣＡｃｔＭ｝（Ｍはアクション候補の個数）を生成する（ステップＳ１３０３）。

【００６６】

次に、応答文生成部１１３が、尤度が最大のアクション候補ＡＣＴを決定する（ステップＳ１３０４）。次に、応答文生成部１１３は、アクション候補ＡＣＴに対応する応答フ

レーズリストＰ｛Ｐ１～ＰＮ｝（Ｎはフレーズ数）を生成する（ステップＳ１３０５）。具体的には、応答文生成部１１３は、図６に示すようなテンプレートを参照し、テンプレートの変数部に、アクション候補ＡＣＴの対応するアクション属性の属性値をそれぞれ当てはめることにより、応答フレーズリストＰを生成する。

【００６７】

次に、出力部１０２が、生成された応答フレーズリストＰから順次応答フレーズＰｉ（ｉ＝１～Ｎ）を取得し、音声合成した合成音声を出力する（ステップＳ１３０６）。なお、ｉは応答フレーズの出力順を表すカウンタ値である。

【００６８】

次に、受付部１０１は、マイク１３１から入力音声Ｉｉが入力されたか否かを判断する（ステップＳ１３０７）。なお、入力音声Ｉｉは、ｉ番目の応答フレーズＰｉの出力中に入力された音声であることを意味するが、応答フレーズＰｉの修正内容を表す音声であるとは限らない。すなわち、応答フレーズＰｉの前に出力された応答フレーズＰ１～Ｐｉ－１のいずれかの修正内容を表す場合もある。また、未出力の応答フレーズＰｉ＋１～ＰＮをユーザが推測して発話した場合であれば、入力音声Ｉｉが応答フレーズＰｉ＋１～ＰＮの修正内容を表す場合もある。

【００６９】

入力音声Ｉｉが入力された場合は（ステップＳ１３０７：ＹＥＳ）、入力音声Ｉｉの内容にしたがって最尤のアクション候補および対応する応答文を更新する候補更新処理が実行される（ステップＳ１３０８）。候補更新処理の詳細については後述する。

【００７０】

候補更新処理の後、またはステップＳ１３０７で入力音声Ｉｉが入力されていない場合（ステップＳ１３０７：ＮＯ）、出力部１０２は、すべての応答フレーズを処理したか否かを判断する（ステップＳ１３０９）。

【００７１】

すべての応答フレーズを処理していない場合は（ステップＳ１３０９：ＮＯ）、出力部１０２は、次の応答フレーズに対して出力処理を繰り返す（ステップＳ１３０６）。なお、後述するように、候補更新処理でアクション候補が変更された場合は、変更後のアクション候補に対応して応答文（応答フレーズリスト）が更新されるため、出力部１０２は、更新後の応答フレーズリストから、次の応答フレーズを取得して出力する。

【００７２】

すべての応答フレーズを処理した場合は（ステップＳ１３０９：ＹＥＳ）、録画再生部１０３が、最尤のアクション候補ＡＣＴに対応するアクションを実行する（ステップＳ１３１０）。

【００７３】

このようにして、ユーザの要求に対する応答であるアクションの内容を確認するための応答文を生成し、応答文の出力中に修正のための音声が入力された場合は、この音声にしたがってアクションおよび応答文を同時に変更することができる。これにより、音声によって容易に誤り箇所を修正可能としつつ、ユーザとの対話を円滑に進めることができる。

【００７４】

次に、ステップＳ１３０８の候補更新処理の詳細について図１４を用いて説明する。図１４は、本実施の形態における候補更新処理の全体の流れを示すフローチャートである。

【００７５】

まず、認識部１１１は、入力音声Ｉｉを音声認識し、認識結果を出力する（ステップＳ１４０１）。次に、修正語句生成部１１４は、認識結果を解析して少なくとも１つのアクション属性の属性値を含むアクション断片群ＳＥＧ｛ＳＥＧ１～ＳＥＧＫ｝（Ｋはアクション断片の個数）を生成する（ステップＳ１４０２）。

【００７６】

次に、選択部１１５は、アクション断片群ＳＥＧが存在するか否かを判断し（ステップＳ１４０３）、存在する場合は（ステップＳ１４０３：ＹＥＳ）、アクション断片群ＳＥ

Ｇの要素と同じアクション属性に対応する属性値が、すべての要素について一致するアクション候補を選択する。そして、選択したアクション候補のうち、尤度が最大のアクション候補ＣＡｃｔｋを選択する（ステップＳ１４０４）。

【００７７】

次に、選択部１１５は、アクション候補ＣＡｃｔｋが存在するか否かを判断する（ステップＳ１４０５）。アクション候補ＣＡｃｔｋが存在する場合は（ステップＳ１４０５：ＹＥＳ）、更新部１１６が、アクション候補ＣＡｃｔｋ（新候補）と、現在の最尤のアクション候補ＡＣＴ（旧候補）とを比較する。そして、更新部１１６は、不一致部分に対応する新候補のアクション属性（以下、不一致属性という）を含む不一致属性群Ａｔｔ{Ａｔｔ１～ＡｔｔＬ}（Ｌは不一致属性の個数）を生成する（ステップＳ１４０６）。

【００７８】

次に、選択部１１５は、不一致属性群Ａｔｔが存在するか否かを判断し（ステップＳ１４０７）、存在する場合は（ステップＳ１４０７：ＹＥＳ）、アクション候補ＣＡｃｔｋを最尤のアクション候補ＡＣＴとして設定する（ステップＳ１４０８）。

【００７９】

次に、更新部１１６は、応答フレーズリストＰのうち、不一致属性群Ａｔｔに含まれるアクション属性に対応する応答フレーズを、不一致属性群Ａｔｔの属性値で置換する（ステップＳ１４０９）。

【００８０】

続いて、更新後の応答フレーズリストＰを、いずれの応答フレーズから出力するかを特定するため、出力部１０２が以下の処理を実行する（ステップＳ１４１０～ステップＳ１４１２）。

【００８１】

まず、出力部１０２は、置換した属性値のうち、最も文頭に近い属性値の文頭からの位置ｊを取得する（ステップＳ１４１０）。次に、出力部１０２は、取得した属性値の位置ｊが、更新前の応答フレーズリストＰで出力済みの応答フレーズの位置ｉより前か否かを判断する（ステップＳ１４１１）。

【００８２】

通常は、出力済みの応答フレーズに対する修正内容が発話され、対応する属性値が置換されるため、ｊはｉより小さくなる。しかし、上述のようにユーザが応答フレーズを推測して未出力の応答フレーズに対する修正内容が発話された場合などには、ｊがｉより小さくならない場合がある。

【００８３】

位置ｊが位置ｉより前の場合は（ステップＳ１４１１：ＹＥＳ）、出力部１０２は、置換した属性値の位置ｊを、次の出力位置に設定する（ステップＳ１４１２）。すなわち、出力部１０２は、ｊをｉに代入する。

【００８４】

ステップＳ１４０３でアクション断片群ＳＥＧが存在しないと判断された場合（ステップＳ１４０３：ＮＯ、ステップＳ１４０５でアクション候補ＣＡｃｔｋが存在しないと判断された場合（ステップＳ１４０５：ＮＯ）、ステップＳ１４０７で不一致属性群Ａｔｔが存在しないと判断された場合（ステップＳ１４０７：ＮＯ）、または、ステップＳ１４１１で位置ｊが位置ｉより前でないと判断された場合は（ステップＳ１４１１：ＮＯ）、候補更新処理を終了する。

【００８５】

次に、本実施の形態のかかるビデオ録画再生装置１００による音声対話処理の具体例について説明する。

【００８６】

まず、ユーザが、当日の朝、「ＭＨＫ」というチャンネルの、「英語講座」という名称の番組の録画予約をセットする目的で、「ＭＨＫで朝、英語講座を録ってね」を意味する日本語の入力音声Ｉ０（えむえっちけーであさえいごこうざをとってね）を入力する（ス

テップＳ１３０１）。続いて、認識部１１１が、入力音声Ｉ０を音声認識し、図３に示すような認識候補群を生成する（ステップＳ１３０２）。さらに、候補生成部１１２が、この認識候補群から図５に示すアクション候補群ＣＡｃｔを生成する（ステップＳ１３０３）。

【００８７】

なお、上述のように、図３の例では、ユーザの要求に適ったアクション候補は第３位候補であることに注意されたい。

【００８８】

アクション候補群ＣＡｃｔ中、最も尤度が大きい候補は、尤度０．４のＣＡｃｔ１であるため、ＣＡｃｔ１をＡＣＴに設定する（ステップＳ１３０４）。次に、応答文生成部１１３が、図６に示すようなテンプレートＴ（｛チャンネル｝で/｛日時｝放送される/｛番組名｝を/｛操作｝しますね？）の変数部に対応するアクション属性のそれぞれに、ＣＡｃｔ１の対応するアクション属性の属性値を挿入し、応答フレーズリストＰを生成する（ステップＳ１３０５）。図７は、このときに生成される応答フレーズリストＰを表している。

【００８９】

次に、出力部１０２が、カウンタｉ（＝１）に対応する応答フレーズＰ１（ＭＨＫで）を音声合成して出力する（ステップＳ１３０６）。ここでは、応答フレーズＰ１の出力処理中には、ユーザから入力音声Ｉ１が入力されなかったと仮定する（ステップＳ１３０７：ＮＯ）。続いて、出力部１０２が、次のカウンタｉ（＝２）に対応する応答フレーズＰ２（明後日放送される）を音声合成して出力する（ステップＳ１３０６）。

【００９０】

ここで、応答フレーズＰ２の音声出力中、ユーザが最初の入力音声Ｉ０の日時の指定（（今日の）朝）が、誤って解釈されていることに気づいたと仮定する。そして、ユーザが、録画する日時を朝に修正するために、「朝だよ」を意味する日本語の入力音声Ｉ２（あさだよ）を入力したと仮定する（ステップＳ１３０７：ＹＥＳ）。

【００９１】

この場合は、入力音声Ｉ２を元に最尤のアクション候補ＡＣＴおよび応答フレーズリストＰを更新する候補更新処理が実行される（ステップＳ１３０８）。

【００９２】

候補更新処理では、まず、認識部１１１が、入力音声Ｉ２を音声認識し、図８に示すような認識候補群を生成する（ステップＳ１４０１）。さらに、修正語句生成部１１４が、認識候補群に対応するアクション断片群ＳＥＧを生成する（ステップＳ１４０２）。ここでは、アクション候補の属性「日時」の情報のみが抽出されるため、アクション断片群ＳＥＧ｛ＳＥＧ１｝が得られる。

【００９３】

続いて、選択部１１５が、アクション断片群ＳＥＧの要素（ここではＳＥＧ１のみ）の属性「日時」の値が「（当日）朝」であるアクション候補群をアクション候補群ＣＡｃｔから選択する。この例では、選択部１１５は、図５のＣＡｃｔ３およびＣＡｃｔ４を選択する。そして、選択部１１５は、これら候補のうち、最も尤度の大きいＣＡｃｔ３（尤度０．３）を最尤候補ＣＡｃｔｋとする（ステップＳ１４０４）。

【００９４】

最尤候補ＣＡｃｔｋが見つかったため（ステップＳ１４０５：ＹＥＳ）、更新部１１６は、ＣＡｃｔ３とＡＣＴ（＝ＣＡｃｔ１）の各属性値を比較し、不一致属性群Ａｔｔを生成する（ステップＳ１４０６）。この例では、図１１に示すように、属性値１１０１に対応するアクション属性「日時」と、属性値１１０２に対応するアクション属性「番組名」とが不一致属性群Ａｔｔに含まれる。

【００９５】

そこで、更新部１１６は、応答フレーズリストＰ（｛ＭＨＫ｝で/｛明後日｝放送される/｛囲碁講座｝を/｛録画｝しますね？｝）の対応する属性値（｛明後日｝および｛囲碁

講座｝）を、ＣＡｃｔ３の属性値（「朝」および「英語講座」）で置き換える（ステップＳ１４０９）。図１２は、このようにして更新された応答フレーズリストＰを表している。

【００９６】

ここまでの処理によって、応答文に対応してユーザが発話した入力音声をフィードバックして、アクションおよびアクションに対応する応答フレーズも修正することができている。

【００９７】

しかし、応答フレーズを修正した場合に、途中まで出力した応答文（応答フレーズリスト）を再度、最初から出力するか、修正箇所だけ出力するか、といった出力の仕方によってユーザの利便性が大きく異なる。

【００９８】

そこで、本実施の形態では、上述のように、応答文のうち既に出力済みの部分は可能な限り再出力をさけつつ、変更箇所については必ず出力するように構成している。すなわち、更新した応答フレーズのうち、最も文頭に近い応答フレーズＰｊ（最も添え字ｊが小さい応答フレーズ）が既に出力済みであれば、出力部１０２は、応答フレーズＰｊから出力を再開する。また、応答フレーズＰｊが未出力であれば、出力部１０２は、現在の出力位置を表すカウンタｉが示す応答フレーズＰｉから続けて出力する。

【００９９】

上述の例では、最も文頭に近い更新された応答フレーズはＰ２（｛朝｝放送される）である。すなわち、更新された応答フレーズの添え字うち最も小さい添え字ｊは２であり、現在のカウンタｉ＝２と一致するため、カウンタｉは更新しない（ステップＳ１４１１：ＮＯ）。

【０１００】

この後、出力部１０２は、更新後の応答フレーズＰ２（｛朝｝放送される）の合成音声を出力する（ステップＳ１３０６）。ここで、ユーザが合成音声を聞くことにより入力音声Ｉ２が正しく解釈されたことを確認し、修正のための発話を行わなかったと仮定する。

【０１０１】

以降、同様に、応答フレーズＰ３（｛英語講座｝を）、および応答フレーズＰ４（｛録画｝しますね？）が順次出力される。その間、ユーザからの応答発話が検出されなかったとすると、応答文の出力後、録画再生部１０３によって、確定されたアクションが実行される（ステップＳ１３１０）。その後、ユーザからの入力受付状態にもどる（ステップＳ１３０１）。

【０１０２】

このように、本実施の形態にかかる音声対話装置では、ユーザの要求発話に応じた応答フレーズを順次出力し、ユーザからの修正のための応答があった場合は、アクション候補と応答フレーズリストを同時に修正することができる。また、修正箇所から応答フレーズの発話を続行するため、更新前で出力済みの部分は出力を省略することができる。これにより、余分な手順を踏んで対話を阻害することなく、容易に修正可能な音声対話装置を実現することができる。

【０１０３】

また、応答文の音声を聞いたユーザが、まだ出力されていない部分についての誤りを推測して言い直した場合であっても、修正箇所を特定し、適切な候補を選択しなおすことができる。これにより、ユーザの利便性を向上させ、対話をより円滑に進めることが可能となる。

【０１０４】

（変形例）

上記実施の形態では、図６に示したような固定のテンプレートにしたがって応答フレーズを生成し、生成した応答フレーズを順次出力していた。

【０１０５】

しかし、文の先頭に近い応答フレーズが誤っているような場合、誤った応答フレーズが出力された時点までに出力される情報が少ないため、その情報のみから、応答フレーズが誤っているか否かを適切に判断できない場合が生じうる。

【０１０６】

例えば、図７の応答フレーズリストの最初の応答フレーズＰ１（｛ＭＨＫ｝で）のチャンネル名である「ＭＨＫ」が「ＬＨＫ」の誤りであったとする。しかし、応答フレーズＰ１が出力された時点で、その断片的な情報のみから、その応答フレーズがチャンネル名に相当する箇所に対する応答フレーズであると、ユーザが瞬時に判別できるとは限らない。

【０１０７】

そこで、本変形例では、より解釈の曖昧性の少ない応答フレーズを先に出力することにより、このような問題を軽減する。ただし、単純に曖昧性の少ない順に応答フレーズを並べ替えただけでは、言語的な制約によって、不自然な意味の応答文や、文法的に不適格な応答文が生成されるおそれがある。

【０１０８】

例えば、図７に対応する応答文を「明後日放送される/ＭＨＫで/囲碁講座を/録画しますね？」のように並べ替えた場合、「放送される」が「ＭＨＫ」に係り、意味的に誤った応答文となる。

【０１０９】

そこで、並べ替えのための制約規則を構築し、その規則にしたがって応答フレーズリストを生成する。例えば、並べ替え可能なパターンを網羅した複数のテンプレートを予め用意し、最適なテンプレートを選択して応答文を生成するように構成する。具体的には、応答文生成部１１３が、このようなテンプレートから、曖昧性に応じて最適なテンプレートを選択して最尤のアクション候補の属性値を当てはめて応答文を生成する。

【０１１０】

図１５は、本変形例で利用するテンプレートの一例を示す説明図である。図１５では、応答フレーズの出力順が異なる４つのテンプレートの例が示されている。

【０１１１】

例えば、図５のアクション候補群が生成され、最尤のアクション候補Ｃａｃｔ１の応答文を生成する場合、まず、応答文生成部１１３は、アクション候補のアクション属性それぞれの曖昧性を判断する。図５の例では、アクション属性「操作」および「チャンネル」は、ただ１通りの属性値を有するため、曖昧性は低いと判断される。アクション属性「日時」および「番組名」は、それぞれ２通りの属性値を有するため曖昧性が高いと判断される。

【０１１２】

そこで、応答文生成部１１３は、アクション属性「操作」および「チャンネル」が先に出現するテンプレートを優先して選択する。図１５の例では、応答文生成部１１３は、テンプレートＴ２（｛操作｝しますね？/｛チャンネル｝で/｛日時｝放送される/｛番組名｝を/）を選択する。そして、この場合、応答文生成部１１３は、応答フレーズリストとして、「｛録画｝しますね？/｛ＭＨＫ｝で/｛明後日｝放送される/｛囲碁番組｝を/」を生成する。

【０１１３】

このように、事前に定められたテンプレートにしたがい応答文を生成しているため、文法的に誤った応答文が生成されることはない。また、曖昧性の少ない応答フレーズから順に出力するため、誤って認識された応答フレーズが出力されるまでに、多くの情報（応答フレーズ）が出力される可能性が高くなる。これにより、情報量が少ないことにより応答フレーズの適否を適切に判断できなくなるという上述の問題を解消することが可能となる。

【０１１４】

次に、本実施の形態にかかる音声対話装置のハードウェア構成について図１６を用いて説明する。図１６は、本実施の形態にかかる音声対話装置のハードウェア構成を示す説明

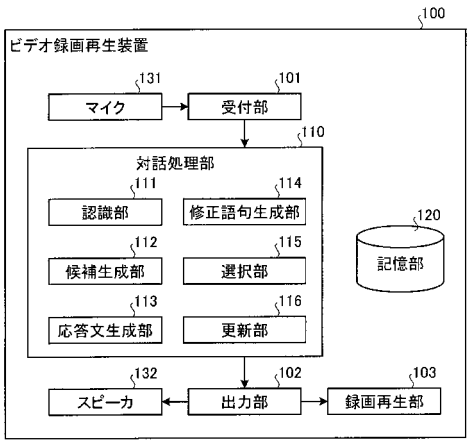
図である。
【０１１５】
本実施の形態にかかる音声対話装置は、ＣＰＵ（Central Processing Unit）５１などの制御装置と、ＲＯＭ（Read Only Memory）５２やＲＡＭ５３などの記憶装置と、ネットワークに接続して通信を行う通信Ｉ／Ｆ５４と、各部を接続するバス６１を備えている。
【０１１６】
本実施の形態にかかる音声対話装置で実行される音声対話プログラムは、ＲＯＭ５２等に予め組み込まれて提供される。
【０１１７】
本実施の形態にかかる音声対話装置で実行される音声対話プログラムは、インストール可能な形式又は実行可能な形式のファイルでＣＤ－ＲＯＭ（Compact Disk Read Only Memory）、フレキシブルディスク（ＦＤ）、ＣＤ－Ｒ（Compact Disk Recordable）、ＤＶＤ（Digital Versatile Disk）等のコンピュータで読み取り可能な記録媒体に記録して提供するように構成してもよい。
【０１１８】
さらに、本実施の形態にかかる音声対話装置で実行される音声対話プログラムを、インターネット等のネットワークに接続されたコンピュータ上に格納し、ネットワーク経由でダウンロードさせることにより提供するように構成してもよい。また、本実施の形態にかかる音声対話装置で実行される音声対話プログラムをインターネット等のネットワーク経由で提供または配布するように構成してもよい。
【０１１９】
本実施の形態にかかる音声対話装置で実行される音声対話プログラムは、上述した各部（受付部、対話処理部、出力部、録画再生部）を含むモジュール構成となっており、実際のハードウェアとしてはＣＰＵ５１が上記ＲＯＭ５２から音声対話プログラムを読み出して実行することにより上記各部が主記憶装置上にロードされ、各部が主記憶装置上に生成されるようになっている。
【産業上の利用可能性】
【０１２０】
以上のように、本発明にかかる装置および方法は、音声で入力された要求に応じて動作するビデオ録画再生装置、カーナビゲーションシステム、ゲーム機器などに適している。
【図面の簡単な説明】
【０１２１】
【図１】本実施の形態にかかるビデオ録画再生装置の構成を示すブロック図である。
【図２】音声認識結果の一例を示す説明図である。
【図３】認識候補文の一例を示す説明図である。
【図４】アクションの一例を示す説明図である。
【図５】アクション候補群の一例を示す説明図である。
【図６】テンプレートの一例を示す説明図である。
【図７】応答フレーズリストの一例を示す説明図である。
【図８】認識候補文の別の例を示す説明図である。
【図９】アクション断片の一例を示す説明図である。
【図１０】旧候補の一例を示す説明図である。
【図１１】新候補の一例を示す説明図である。
【図１２】更新された後の応答フレーズリストの一例を示す説明図である。
【図１３】本実施の形態における音声対話処理の全体の流れを示すフローチャートである。
【図１４】本実施の形態における候補更新処理の全体の流れを示すフローチャートである。
【図１５】変形例で利用するテンプレートの一例を示す説明図である。
【図１６】本実施の形態にかかる音声対話装置のハードウェア構成を示す説明図である。

【符号の説明】

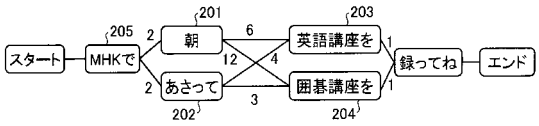
【０１２２】

５１　ＣＰＵ
５２　ＲＯＭ
５３　ＲＡＭ
５４　通信Ｉ／Ｆ
６１　バス
１００　ビデオ録画再生装置
１０１　受付部
１０２　出力部
１０３　録画再生部
１１０　対話処理部
１１１　認識部
１１２　候補生成部
１１３　応答文生成部
１１４　修正語句生成部
１１５　選択部
１１６　更新部
１２０　記憶部
１３１　マイク
１３２　スピーカ
２０１～２０５　ノード
１１０１、１１０２　属性値

【図１】



【図２】



【図３】

候補番号	候補テキスト	尤度
第1位候補	MHKであさって囲碁講座を録ってね	0.4
第2位候補	MHKであさって英語講座を録ってね	0.3
第3位候補	MHKで朝英語講座を録ってね	0.2
第4位候補	MHKで朝囲碁講座を録ってね	0.1

【図４】

操作	日時	チャンネル	番組名
録画	（当日）朝	MHK	英語講座
再生	―	―	録画データ1
消去	―	―	録画データ2

【図５】

候補	操作	日時	チャンネル	番組名	尤度
CAct1	録画	明後日	MHK	囲碁講座	0.4
CAct2	録画	明後日	MHK	英語講座	0.3
CAct3	録画	(当日)朝	MHK	英語講座	0.2
CAct4	録画	(当日)朝	MHK	囲碁講座	0.1

【図６】

テンプレートT	{チャンネル}で/{日時}放送される/{番組名}を/{操作}しますね？

【図７】

P1	P2	P3	P4
{MHK}で	{明後日}放送される	{囲碁講座}を	{録画}しますね？

【図８】

候補番号	候補テキスト	尤度
第1位候補	朝だよ	1.0

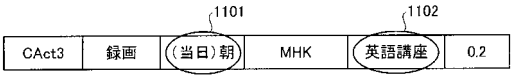
【図９】

日時
(当日)朝

【図１０】

候補	操作	日時	チャンネル	番組名	尤度
CAct1	録画	明後日	MHK	囲碁講座	0.4

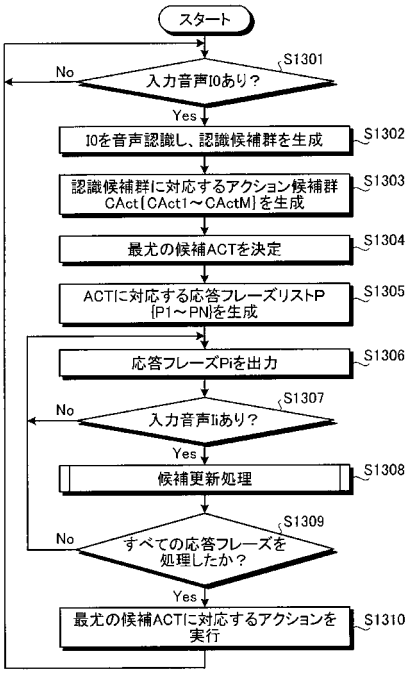
【図１１】



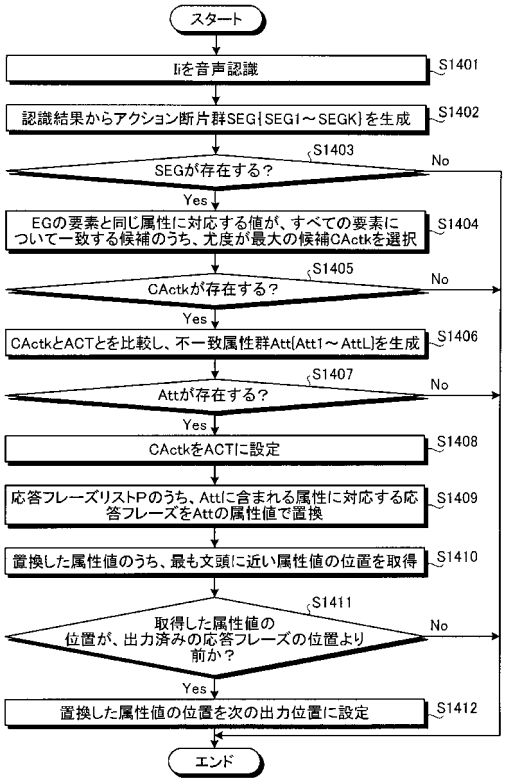
【図１２】

P1	P2	P3	P4
{MHK}で	{朝}放送される	{英語講座}を	{録画}しますね？

【図１３】



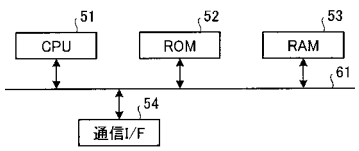
【図１４】



【図１５】

テンプレートT	{チャンネル}で/{日時}放送される/{番組名}を/{操作}しますね？
テンプレートT1	{日時}/{チャンネル}で放送される/{番組名}を/{操作}しますね？
テンプレートT2	{操作}しますね？/{チャンネル}で/{日時}放送される/{番組名}を/
テンプレートT3	{操作}しますね？/{日時}放送される{チャンネル}の/{番組名}を/

【図１６】



フロントページの続き

審査官　田部井　和彦

(56)参考文献　特開２００７－０９３７８９（ＪＰ，Ａ）
特開２００３－３３０４８８（ＪＰ，Ａ）
特開２００３－２０８１９６（ＪＰ，Ａ）
特開２００６－０３９１２０（ＪＰ，Ａ）
特開２０００－０２９４９２（ＪＰ，Ａ）
特開平０２－１２６３００（ＪＰ，Ａ）
特開平０１－２３７５９７（ＪＰ，Ａ）
特開昭６３－０９５５３２（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１５／００－１７／００