

(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2009-47920

(P2009-47920A)

(43) 公開日 平成21年3月5日(2009.3.5)

(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 4 7 0 Z	5 D 0 1 5
G 1 0 L 15/00 (2006.01)	G 1 0 L 15/00 2 0 0 G	
	G 1 0 L 15/22 3 0 0 Z	

審査請求 未請求 請求項の数 6 O L (全 19 頁)

(21) 出願番号	特願2007-213828 (P2007-213828)	(71) 出願人	000003078
(22) 出願日	平成19年8月20日 (2007.8.20)		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
		Fターム(参考)	5D015 LL04 LL06

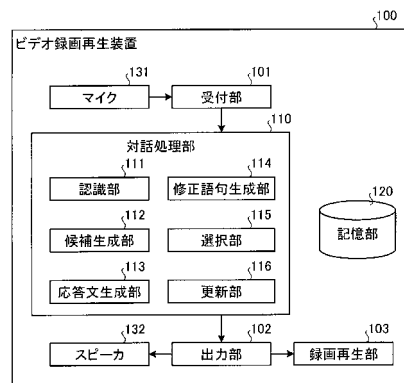
(54) 【発明の名称】 ユーザと音声により対話する装置および方法

(57) 【要約】

【課題】対話を阻害することなく誤り箇所を容易に訂正する音声対話装置を提供する。

【解決手段】 音声を認識し、応答の候補と、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 1 1 2 と、最尤の応答の候補を選択の内容を表す語句を含む応答文を生成する応答文生成部 1 1 3 と、応答文の合成音声を出力する出力部 1 0 2 と、合成音声の出力中にユーザが発話した音声に対する認識結果を解析して、応答文に含まれる語句に対応する少なくとも1つの修正語句を生成する修正語句生成部 1 1 4 と、生成された応答の候補から、生成された修正語句と同一の意味内容の語句を含む応答の候補を取得し、取得した応答の候補のうち最尤の応答の候補を選択する選択部 1 1 5 と、選択された応答の候補の内容を表す語句で応答文を更新する更新部 1 1 6 と、を備え、出力部 1 0 2 は、更新後の応答文の合成音声を出力する。

【選択図】 図 1



【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入力する出力部と、

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入力すること、

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入力すること、

を特徴とする請求項 1 に記載の音声対話装置。

【請求項 3】

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入力すること、

を特徴とする請求項 2 に記載の音声対話装置。

【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入力すること、

を特徴とする請求項 2 に記載の音声対話装置。

【請求項 5】

前記候補生成部は、さらに、前記応答の候補を表す語句ごとの確からしさを表す第 2 尤度を生成し、

前記応答文生成部は、前記応答の候補を表す語句を、前記第 2 尤度が小さい順に文頭から含む前記応答文を生成すること、

を特徴とする請求項 1 に記載の音声対話装置。

【請求項 6】

認識部によって、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、

候補生成部によって、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、

応答文生成部によって、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

10

20

30

40

50

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、

出力部によって、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、

修正語句生成部によって、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、

選択部によって、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、

更新部によって、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、を備え、

出力部によって、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、

を備えた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・インター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

10

20

30

40

50

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ます、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。

10

【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。

20

【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必要となり、訂正に要する時間を短くすることができる。

30

【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。

40

【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】

50

【 0 0 1 8 】

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入力する出力部と、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入力すること、を特徴とする。

10

【 0 0 1 9 】

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。

20

【 発明の効果 】

【 0 0 2 0 】

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 】

【 0 0 2 1 】

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。

【 0 0 2 2 】

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。

30

【 0 0 2 3 】

なお、以下では、ハードディス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。

【 0 0 2 4 】

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。

40

【 0 0 2 5 】

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。

【 0 0 2 6 】

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、

50

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、

10

20

30

40

50

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。

10

【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。

20

【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。

30

【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。

40

【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

50

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、 $P\{P_1 \sim P_N\}$ (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 $C A c t 1$ 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 $P_1 \sim P_4$ は、この順で出力部102から音声出力される。

10

【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。

20

【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 I_1 (「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。

30

【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 I_1 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。

40

【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 $S E G 1$ という)が生成されたとする。この場合、選

50

択部 1 1 5 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 S E G 1 ((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 1 1 5 は、C A c t 3 および C A c t 4 を取得することができる。次に、選択部 1 1 5 は、C A c t 3 および C A c t 4 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、C A c t 3 の尤度 = 0 . 2 > C A c t 4 の尤度 = 0 . 1 であるため、C A c t 3 が選択される。

【 0 0 5 3 】

更新部 1 1 6 は、選択部 1 1 5 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 1 1 6 は、まず、選択部 1 1 5 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 1 1 6 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。

10

【 0 0 5 4 】

図 1 0 は、旧候補の一例を示す説明図である。また、図 1 1 は、新候補の一例を示す説明図である。図 1 0 および図 1 1 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 1 1 6 は、これらのアクション属性を抽出する。

【 0 0 5 5 】

次に、更新部 1 1 6 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 1 1 の例では、更新部 1 1 6 は、属性値 1 1 0 1 ((当日) 朝) および属性値 1 1 0 2 (英語講座) を新たな属性値として取得する。そして、更新部 1 1 6 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。

20

【 0 0 5 6 】

図 1 2 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 1 2 は、図 7 の応答フレーズリストを、図 1 1 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。

【 0 0 5 7 】

なお、上述のように、候補生成部 1 1 2 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 1 1 5 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。

30

【 0 0 5 8 】

出力部 1 0 2 は、応答文生成部 1 1 3 によって生成された応答文、または更新部 1 1 6 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じ、合成音声をスピーカ 1 3 2 に出力するものである。

【 0 0 5 9 】

具体的には、出力部 1 0 2 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 1 0 2 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 1 0 2 は、生成した音声信号を D A 変換してスピーカ 1 3 2 に出力する。

40

【 0 0 6 0 】

また、出力部 1 0 2 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 1 0 2 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。

【 0 0 6 1 】

録画再生部 1 0 3 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

50

を実行するものである。例えば、録画再生部 103 は、図 5 の C A c t 3 が最尤のアクション候補として選択された場合、C A c t 3 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。

10

【0064】

まず、受付部 101 は、マイク 131 から入力音声 I 0 が入力されたか否かを判断する (ステップ S 1301)。入力音声 I 0 が入力されていない場合は (ステップ S 1301: NO)、入力されるまで処理を繰り返す。

【0065】

入力音声 I 0 が入力された場合 (ステップ S 1301: YES)、認識部 111 は、入力音声 I 0 を音声認識し、認識候補群を生成する (ステップ S 1302)。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 C A c t { C A c t 1 ~ C A c t M } (M はアクション候補の個数) を生成する (ステップ S 1303)。

20

【0066】

次に、応答文生成部 113 が、尤度が最大のアクション候補 A C T を決定する (ステップ S 1304)。次に、応答文生成部 113 は、アクション候補 A C T に対応する応答フレーズリスト P { P 1 ~ P N } (N はフレーズ数) を生成する (ステップ S 1305)。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 A C T 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 P i (i = 1 ~ N) を取得し、音声合成した合成音声出力する (ステップ S 1306)。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。

30

【0068】

次に、受付部 101 は、マイク 131 から入力音声 I i が入力されたか否かを判断する (ステップ S 1307)。なお、入力音声 I i は、i 番目の応答フレーズ P i 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 P i 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 P i の前に出力された応答フレーズ P 1 ~ P i - 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 P i + 1 ~ P N をユーザが推測して発話した場合であれば、入力音声 I i が応答フレーズ P i + 1 ~ P N の修正内容を表す場合もある。

40

【0069】

入力音声 I i が入力された場合は (ステップ S 1307: YES)、入力音声 I i 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 S 1308)。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 S 1307 で入力音声 I i が入力されていない場合 (ステップ S 1307: 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 S 1309)。

【0071】

50

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 S 1 3 0 9 : N O）、出力部 1 0 2 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 S 1 3 0 6）。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 1 0 2 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。

【 0 0 7 2 】

すべての応答フレーズを処理した場合は（ステップ S 1 3 0 9 : Y E S）、録画再生部 1 0 3 が、最尤のアクション候補 A C T に対応するアクションを実行する（ステップ S 1 3 1 0）。

【 0 0 7 3 】

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。

【 0 0 7 4 】

次に、ステップ S 1 3 0 8 の候補更新処理の詳細について図 1 4 を用いて説明する。図 1 4 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。

【 0 0 7 5 】

まず、認識部 1 1 1 は、入力音声 I i を音声認識し、認識結果を出力する（ステップ S 1 4 0 1）。次に、修正語句生成部 1 1 4 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 S E G { S E G 1 ~ S E G K }（K はアクション断片の個数）を生成する（ステップ S 1 4 0 2）。

【 0 0 7 6 】

次に、選択部 1 1 5 は、アクション断片群 S E G が存在するか否かを判断し（ステップ S 1 4 0 3）、存在する場合は（ステップ S 1 4 0 3 : Y E S）、アクション断片群 S E 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 C A c t k を選択する（ステップ S 1 4 0 4）。

【 0 0 7 7 】

次に、選択部 1 1 5 は、アクション候補 C A c t k が存在するか否かを判断する（ステップ S 1 4 0 5）。アクション候補 C A c t k が存在する場合は（ステップ S 1 4 0 5 : Y E S）、更新部 1 1 6 が、アクション候補 C A c t k（新候補）と、現在の最尤のアクション候補 A C T（旧候補）とを比較する。そして、更新部 1 1 6 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 A t t { A t t 1 ~ A t t L }（L は不一致属性の個数）を生成する（ステップ S 1 4 0 6）。

【 0 0 7 8 】

次に、選択部 1 1 5 は、不一致属性群 A t t が存在するか否かを判断し（ステップ S 1 4 0 7）、存在する場合は（ステップ S 1 4 0 7 : Y E S）、アクション候補 C A c t k を最尤のアクション候補 A C T として設定する（ステップ S 1 4 0 8）。

【 0 0 7 9 】

次に、更新部 1 1 6 は、応答フレーズリスト P のうち、不一致属性群 A t t 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 A t t の属性値で置換する（ステップ S 1 4 0 9）。

【 0 0 8 0 】

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 1 0 2 が以下の処理を実行する（ステップ S 1 4 1 0 ~ ステップ S 1 4 1 2）。

【 0 0 8 1 】

まず、出力部 1 0 2 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 S 1 4 1 0）。次に、出力部 1 0 2 は、取得した属性値の位置

10

20

30

40

50

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 S 1 4 1 1）。

【 0 0 8 2 】

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。

【 0 0 8 3 】

位置 j が位置 i より前の場合は（ステップ S 1 4 1 1 : Y E S ）、出力部 1 0 2 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 S 1 4 1 2）。すなわち、出力部 1 0 2 は、j を i に代入する。

10

【 0 0 8 4 】

ステップ S 1 4 0 3 でアクション断片群 S E G が存在しないと判断された場合（ステップ S 1 4 0 3 : N O ）、ステップ S 1 4 0 5 でアクション候補 C A c t k が存在しないと判断された場合（ステップ S 1 4 0 5 : N O ）、ステップ S 1 4 0 7 で不一致属性群 A t t が存在しないと判断された場合（ステップ S 1 4 0 7 : N O ）、または、ステップ S 1 4 1 1 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 S 1 4 1 1 : N O ）、候補更新処理を終了する。

【 0 0 8 5 】

次に、本実施の形態のかかるビデオ録画再生装置 1 0 0 による音声対話処理の具体例について説明する。

20

【 0 0 8 6 】

まず、ユーザが、当日の朝、「M H K」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、「M H Kで朝、英語講座を録ってね」を意味する日本語の入力音声 I 0（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 S 1 3 0 1）。続いて、認識部 1 1 1 が、入力音声 I 0 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 S 1 3 0 2）。さらに、候補生成部 1 1 2 が、この認識候補群から図 5 に示すアクション候補群 C A c t を生成する（ステップ S 1 3 0 3）。

【 0 0 8 7 】

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。

30

【 0 0 8 8 】

アクション候補群 C A c t 中、最も尤度が大きい候補は、尤度 0 . 4 の C A c t 1 であるため、C A c t 1 を A C T に設定する（ステップ S 1 3 0 4）。次に、応答文生成部 1 1 3 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、C A c t 1 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 S 1 3 0 5）。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。

40

【 0 0 8 9 】

次に、出力部 1 0 2 が、カウンタ i（= 1）に対応する応答フレーズ P 1（M H Kで）を音声合成して出力する（ステップ S 1 3 0 6）。ここでは、応答フレーズ P 1 の出力処理中には、ユーザから入力音声 I 1 が入力されなかったと仮定する（ステップ S 1 3 0 7 : N O）。続いて、出力部 1 0 2 が、次のカウンタ i（= 2）に対応する応答フレーズ P 2（明後日放送される）を音声合成して出力する（ステップ S 1 3 0 6）。

【 0 0 9 0 】

ここで、応答フレーズ P 2 の音声出力中、ユーザが最初の入力音声 I 0 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 I 2（あ

50

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。

10

【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。

20

【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る。

30

【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。

40

【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

50

を出力する（ステップ S 1 3 0 6）。ここで、ユーザが合成音声を聞くことにより入力音声 I 2 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。

【 0 1 0 1 】

以降、同様に、応答フレーズ P 3（{ 英語講座 } を）、および応答フレーズ P 4（{ 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 1 0 3 によって、確定されたアクションが実行される（ステップ S 1 3 1 0）。その後、ユーザからの入力受付状態にもどる（ステップ S 1 3 0 1）。

【 0 1 0 2 】

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。

10

【 0 1 0 3 】

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。

20

【 0 1 0 4 】

（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。

【 0 1 0 5 】

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。

【 0 1 0 6 】

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 P 1（{ M H K } で）のチャンネル名である「M H K」が「L H K」の誤りであったとする。しかし、応答フレーズ P 1 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。

30

【 0 1 0 7 】

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。

【 0 1 0 8 】

例えば、図 7 に対応する応答文を「明後日放送される / M H K 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「M H K」に係り、意味的に誤った応答文となる。

40

【 0 1 0 9 】

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 1 1 3 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。

【 0 1 1 0 】

図 1 5 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 1 5 では、

50

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 C A c t 1 の応答文を生成する場合、まず、応答文生成部 1 1 3 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 1 1 3 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 1 1 3 は、テンプレート T 2 ({ 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 1 1 3 は、応答フレーズリストとして、「 { 録画 } しますね? / { M H K } で / { 明後日 } 放送される / { 囲碁番組 } を / 」を生成する。

10

【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。

20

【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、C P U (Central Processing Unit) 5 1 などの制御装置と、R O M (Read Only Memory) 5 2 や R A M 5 3 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 / F 5 4 と、各部を接続するバス 6 1 を備えている。

30

【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、R O M 5 2 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 C D - R O M (Compact Disk Read Only Memory)、フレキシブルディスク(FD)、C D - R (Compact Disk Recordable)、D V D (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。

40

【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C P U 5 1 が上記 R O M 5 2 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

50

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。

10

【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

20

。【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

30

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

40

114 修正語句生成部

115 選択部

116 更新部

120 記憶部

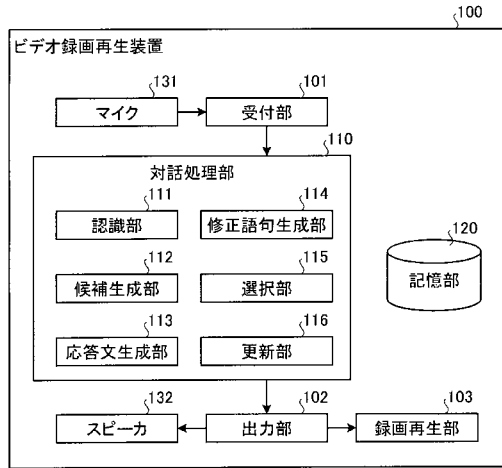
131 マイク

132 スピーカ

201~205 ノード

1101、1102 属性値

【 図 1 】



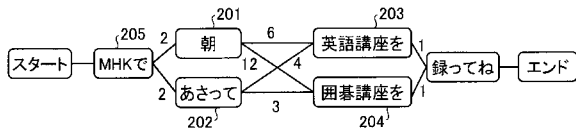
【 図 3 】

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 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 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 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 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 1 1 】

CAct3	録画	1101 (当日)朝	MHK	1102 英語講座	0.2
-------	----	---------------	-----	--------------	-----

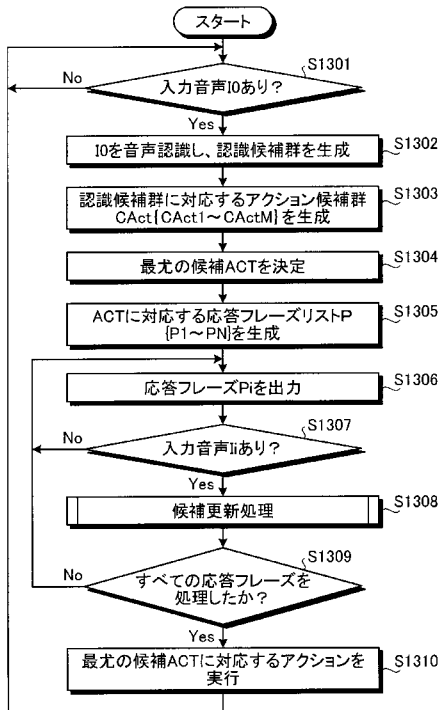
【 図 1 0 】

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

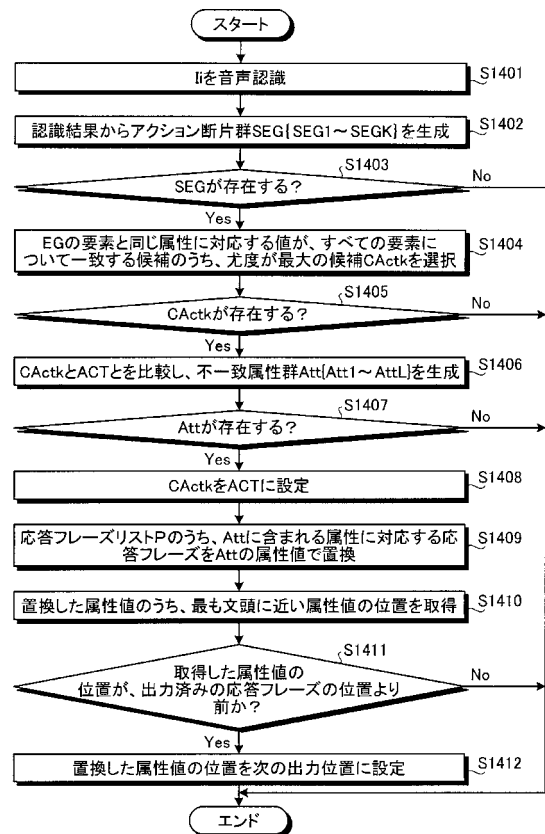
【 図 1 2 】

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

【 図 1 3 】



【 図 1 4 】



【 図 1 5 】

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 1 6 】

