

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5715526号
(P5715526)

(45) 発行日 平成27年5月7日(2015.5.7)

(24) 登録日 平成27年3月20日(2015.3.20)

(51) Int.Cl.		F I			
G 1 0 L	15/22	(2006.01)	G 1 0 L	15/22	3 0 0 Z
G 1 0 L	13/00	(2006.01)	G 1 0 L	13/00	1 0 0 M
			G 1 0 L	15/22	2 0 0 V

請求項の数 7 (全 34 頁)

<p>(21) 出願番号 特願2011-175395 (P2011-175395)</p> <p>(22) 出願日 平成23年8月10日 (2011. 8. 10)</p> <p>(65) 公開番号 特開2012-42952 (P2012-42952A)</p> <p>(43) 公開日 平成24年3月1日 (2012. 3. 1)</p> <p>審査請求日 平成25年11月27日 (2013. 11. 27)</p> <p>(31) 優先権主張番号 61/372, 926</p> <p>(32) 優先日 平成22年8月12日 (2010. 8. 12)</p> <p>(33) 優先権主張国 米国 (US)</p> <p>特許法第30条第1項適用 一般社団法人 情報処理学会、研究報告音声言語情報処理、2011-SLP-86、12、2011年5月9日 社団法人 人工知能学会、2011年度人工知能学会全国大会(第25回)予稿集、3C2-OS19-1、2011年6月1日</p>	<p>(73) 特許権者 000005326 本田技研工業株式会社 東京都港区南青山二丁目1番1号</p> <p>(74) 代理人 100064908 弁理士 志賀 正武</p> <p>(74) 代理人 100108578 弁理士 高橋 詔男</p> <p>(74) 代理人 100146835 弁理士 佐伯 義文</p> <p>(74) 代理人 100094400 弁理士 鈴木 三義</p> <p>(74) 代理人 100107836 弁理士 西 和哉</p> <p>(74) 代理人 100108453 弁理士 村山 靖彦</p>
---	--

最終頁に続く

(54) 【発明の名称】 対話処理装置、対話処理方法、及び対話処理プログラム

(57) 【特許請求の範囲】

【請求項1】

音声を入力する音声入力部と、
 前記音声入力部が入力した音声を認識する音声認識部と、
 予め定めた処理を行うエキスパート部であって、前記処理を行う場合、前記音声認識部が認識した認識結果に基づいて前記処理を継続するか否かを判定し、前記処理を行わない場合、前記音声認識部が認識した認識結果に基づいて適合性を推定する複数のエキスパート部と、
 前記複数のエキスパート部のうち、前記処理を行う場合、前記処理を継続しないと判断したエキスパート部があるとき、前記複数のエキスパート部のうち前記処理を行わない場合に推定した適合性に基づいて前記処理を行うエキスパート部を選択する動作理解部と、
タスク計画部と、を備え、
前記複数のエキスパート部は、要求理解エキスパート部とインタラクティブプレゼンテーション部とを含み、
前記インタラクティブプレゼンテーション部は、
所定の題目に関する提示情報、及び前記提示情報に対する処理の形態を示す指示可能情報を記憶する動作情報記憶部と、
前記認識結果に含まれる単語と一致する指示可能情報が示す形態で前記提示情報を処理する動作処理部と、を備え、
前記要求理解エキスパート部は、

10

20

キーワードを含んだ質問内容を表す質問情報及び肯定発話情報を記憶し、キーワードを含んだ応答情報を前記質問情報と対応付けて記憶した処理情報記憶部と、

前記認識結果に含まれる単語と一致するキーワードを含んだ質問情報を前記処理情報記憶部から検索し、その後の認識結果に肯定発話情報が含まれる場合、検索した質問情報に含まれるキーワードを前記インタラクティブプレゼンテーション部の処理の開始を示す処理要求情報と定める動作処理部と、を備え、

前記タスク計画部は、

処理要求情報とインタラクティブプレゼンテーション部の識別情報とを対応付けて記憶する記憶部から、前記要求理解エキスパート部の動作処理部が定めた処理要求情報に対応する識別情報が示すインタラクティブプレゼンテーション部を起動させる

10

ことを特徴とする対話処理装置。

【請求項 2】

音声を入力する音声入力部と、

前記音声入力部が入力した音声を認識する音声認識部と、

予め定めた処理を行う機能を備える複数のエキスパート部のうち前記処理を行うエキスパート部について、前記音声認識部が認識した認識結果に基づいて前記処理を継続するか否かを判定し、

前記処理を継続しないと判断した場合、前記複数のエキスパート部のうち前記処理を行わないエキスパート部の各々について、前記音声認識部が認識した認識結果に基づいて適合性を推定し、前記推定した適合性に基づいて前記処理を行うエキスパート部を選択する動作理解部と、

20

タスク計画部と、を備え、

前記複数のエキスパート部は、要求理解エキスパート部とインタラクティブプレゼンテーション部とを含み、

前記インタラクティブプレゼンテーション部は、

所定の題目に関する提示情報、及び前記提示情報に対する処理の形態を示す指示可能情報を記憶する動作情報記憶部と、

前記認識結果に含まれる単語と一致する指示可能情報が示す形態で前記提示情報を処理する動作処理部と、を備え、

前記要求理解エキスパート部は、

30

キーワードを含んだ質問内容を表す質問情報及び肯定発話情報を記憶し、キーワードを含んだ応答情報を前記質問情報と対応付けて記憶した処理情報記憶部と、

前記認識結果に含まれる単語と一致するキーワードを含んだ質問情報を前記処理情報記憶部から検索し、その後の認識結果に肯定発話情報が含まれる場合、検索した質問情報に含まれるキーワードを前記インタラクティブプレゼンテーション部の処理の開始を示す処理要求情報と定める動作処理部と、を備え、

前記タスク計画部は、

処理要求情報とインタラクティブプレゼンテーション部の識別情報とを対応付けて記憶する記憶部から、前記要求理解エキスパート部の動作処理部が定めた処理要求情報に対応する識別情報が示すインタラクティブプレゼンテーション部を起動させる

40

ことを特徴とする対話処理装置。

【請求項 3】

前記要求理解エキスパート部の処理を継続するか否かの判定は、前記認識結果に肯定発話情報を含むか否かを示す特徴量に基づき、

前記要求理解エキスパート部の選択は、前記認識結果に前記所定の題目を含むか否かを示す特徴量を用いて推定された適合性に基づく

ことを特徴とする請求項 1 又は 2 に記載の対話処理装置。

【請求項 4】

前記動作理解部は、前記推定した適合性に基づいて前記処理を継続するか否かを判定すること

50

を特徴とする請求項 3 に記載の対話処理装置。

【請求項 5】

前記複数のエキスパート部において前記処理情報記憶部に記憶された
質問情報及び応答情報が各々異なること
 を特徴とする請求項 4 記載の対話処理装置。

【請求項 6】

予め定めた処理を行う機能を有する複数のエキスパート部を備える対話処理装置における対話処理方法において、

前記対話処理装置が、入力した音声を認識する第 1 の過程と、

前記複数のエキスパート部のうち前記処理を行うエキスパート部が、前記認識した認識結果に基づいて前記処理を継続するか否かを判定する第 2 の過程と、

前記処理を行うエキスパート部が前記処理を継続しないと判断した場合、前記複数のエキスパート部のうち前記処理を行わないエキスパート部の各々が、前記認識した認識結果に基づいて適合性を推定する第 3 の過程と、

前記対話処理装置が、前記推定した適合性に基づいて前記処理を行うエキスパート部を選択する第 4 の過程と

第 5 の過程と、を有し、

前記複数のエキスパート部は、要求理解エキスパート部とインタラクティブプレゼンテーション部とを含み、

前記インタラクティブプレゼンテーション部は、

所定の題目に関する提示情報、及び前記提示情報に対する処理の形態を示す指示可能情報を記憶する動作情報記憶部を備え、

前記認識結果に含まれる単語と一致する指示可能情報が示す形態で前記提示情報を処理し、

前記要求理解エキスパート部は、

キーワードを含んだ質問内容を表す質問情報及び肯定発話情報を記憶し、キーワードを含んだ応答情報を前記質問情報と対応付けて記憶した処理情報記憶部を備え、

前記認識結果に含まれる単語と一致するキーワードを含んだ質問情報を前記処理情報記憶部から検索し、その後の認識結果に肯定発話情報が含まれる場合、検索した質問情報に含まれるキーワードを前記インタラクティブプレゼンテーション部の処理の開始を示す処理要求情報と定め、

前記第 5 の過程は、

前記対話処理装置が、処理要求情報とインタラクティブプレゼンテーション部の識別情報とを対応付けて記憶する記憶部から、前記要求理解エキスパート部が定めた処理要求情報に対応する識別情報が示すインタラクティブプレゼンテーション部を起動させること
 を特徴とする対話処理方法。

【請求項 7】

予め定めた処理を行う機能を有する複数のエキスパート部を備える対話処理装置のコンピュータに、

入力した音声を認識する手順、

前記複数のエキスパート部のうち前記処理を行うエキスパート部について、前記認識した認識結果に基づいて処理を継続するか否かを判定する手順、

前記処理を継続しないと判断した場合、前記複数のエキスパート部のうち前記処理を行わないエキスパート部の各々について、前記認識した認識結果に基づいて適合性を推定する手順、

前記推定した適合性に基づいて前記処理を行うエキスパート部を選択する手順、

タスク計画手順、を実行させるための対話処理プログラムであって、

前記複数のエキスパート部は、要求理解エキスパート部とインタラクティブプレゼンテーション部とを含み、

前記インタラクティブプレゼンテーション部は、

10

20

30

40

50

所定の題目に関する提示情報、及び前記提示情報に対する処理の形態を示す指示可能情報を記憶する動作情報記憶部を備え、

前記認識結果に含まれる単語と一致する指示可能情報が示す形態で前記提示情報を処理し、

前記要求理解エキスパート部は、

キーワードを含んだ質問内容を表す質問情報及び肯定発話情報を記憶し、キーワードを含んだ応答情報を前記質問情報と対応付けて記憶した処理情報記憶部を備え、

前記認識結果に含まれる単語と一致するキーワードを含んだ質問情報を前記処理情報記憶部から検索し、その後の認識結果に肯定発話情報が含まれる場合、検索した質問情報に含まれるキーワードを前記インタラクティブプレゼンテーション部の処理の開始を示す処理要求情報と定め、

前記タスク計画手順は、

処理要求情報とインタラクティブプレゼンテーション部の識別情報とを対応付けて記憶する記憶部から、前記要求理解エキスパート部が定めた処理要求情報に対応する識別情報が示すインタラクティブプレゼンテーション部を起動させる手順である

対話処理プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、対話処理装置、対話処理方法、及び対話処理プログラムに関する。

【背景技術】

【0002】

利用者の発声内容に基づいて処理を行う対話処理技術が開発されている。例えば、ロボットに应用することで利用者が動作を指示する負担を軽減することが試みられている。この対話処理技術を用いることで、利用者が発話して指示した動作の種別を応答して動作を直ちに実行しない。これによりシステム側で誤認識された動作を直ちに実行することが回避され、利用者が意図する動作を導くことができる。

【0003】

例えば、特許文献1に記載の対話処理装置は、音声を入力する音声入力装置と、入力された音声を認識する音声認識部と、音声認識結果に応じて対話行動を制御する対話行動制御部と、を備え、対話行動制御部が、発話内容の記憶及びマッチングを行うトピック認識エキスパートと、モードの変更を管理するモード切替エキスパートが対話対象の発話に応じてモードの切り替えを行う。第1のモードでは発話された複数の単語をトピックとして記録するとともに、記録された複数のトピックに対して、第2のモードで発話された内容をマッチングして、最も尤度の高いトピックを選択する。

対話行動制御部は、音声認識結果を各エキスパートに配布し、選択したトピックと各エキスパートが選択される確からしさを表すスコアに基づいてエキスパートを選択する。エキスパートとは、利用者との対話において予め定めた種類の処理、又は利用者によって指示された種類の動作を行うための処理を行うモジュール（構成部）である。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】国際公開第2009/028647号

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、特許文献1に記載の対話処理装置は、不適切なエキスパートを選択しても、選択したトピックを変更しなければ選択したエキスパートを変更しないことがあった。

【0006】

10

20

30

40

50

本発明は上記の点に鑑みてなされたものであり、利用者の発話に適したエキスパートを選択する対話処理装置、対話処理方法、及び対話処理プログラムを提供する。

【課題を解決するための手段】

【0007】

(1) 本発明は上記の課題を解決するためになされたものであり、本発明の一態様は、音声を入力する音声入力部と、前記音声入力部が入力した音声を認識する音声認識部と、予め定めた処理を行うエキスパート部であって、前記処理を行う場合、前記音声認識部が認識した認識結果に基づいて前記処理を継続するか否かを判定し、前記処理を行わない場合、前記音声認識部が認識した認識結果に基づいて適合性を推定する複数のエキスパート部と、前記複数のエキスパート部のうち、前記処理を行う場合、前記処理を継続しないと判断したエキスパート部があるとき、前記複数のエキスパート部のうち前記処理を行わない場合に推定した適合性に基づいて前記処理を行うエキスパート部を選択する動作理解部と、タスク計画部と、を備え、前記複数のエキスパート部は、要求理解エキスパート部とインタラクティブプレゼンテーション部とを含み、前記インタラクティブプレゼンテーション部は、所定の題目に関する提示情報、及び前記提示情報に対する処理の形態を示す指示可能情報を記憶する動作情報記憶部と、前記認識結果に含まれる単語と一致する指示可能情報が示す形態で前記提示情報を処理する動作処理部と、を備え、前記要求理解エキスパート部は、キーワードを含んだ質問内容を表す質問情報及び肯定発話情報を記憶し、キーワードを含んだ応答情報を前記質問情報と対応付けて記憶した処理情報記憶部と、前記認識結果に含まれる単語と一致するキーワードを含んだ質問情報を前記処理情報記憶部から
検索し、その後の認識結果に肯定発話情報が含まれる場合、検索した質問情報に含まれるキーワードを前記インタラクティブプレゼンテーション部の処理の開始を示す処理要求情報と定める動作処理部と、を備え、前記タスク計画部は、処理要求情報とインタラクティブプレゼンテーション部の識別情報とを対応付けて記憶する記憶部から、前記要求理解エキスパート部の動作処理部が定めた処理要求情報に対応する識別情報が示すインタラクティブプレゼンテーション部を起動させることを特徴とする対話処理装置である。

【0008】

(2) 本発明のその他の態様は、音声を入力する音声入力部と、前記音声入力部が入力した音声を認識する音声認識部と、予め定めた処理を行う機能を備える複数のエキスパート部のうち前記処理を行うエキスパート部について、前記音声認識部が認識した認識結果に基づいて前記処理を継続するか否かを判定し、前記処理を継続しないと判断した場合、前記複数のエキスパート部のうち前記処理を行わないエキスパート部の各々について、前記音声認識部が認識した認識結果に基づいて適合性を推定し、前記推定した適合性に基づいて前記処理を行うエキスパート部を選択する動作理解部と、タスク計画部と、を備え、前記複数のエキスパート部は、要求理解エキスパート部とインタラクティブプレゼンテーション部とを含み、前記インタラクティブプレゼンテーション部は、所定の題目に関する提示情報、及び前記提示情報に対する処理の形態を示す指示可能情報を記憶する動作情報記憶部と、前記認識結果に含まれる単語と一致する指示可能情報が示す形態で前記提示情報を処理する動作処理部と、を備え、前記要求理解エキスパート部は、キーワードを含んだ質問内容を表す質問情報及び肯定発話情報を記憶し、キーワードを含んだ応答情報を前記
質問情報と対応付けて記憶した処理情報記憶部と、前記認識結果に含まれる単語と一致するキーワードを含んだ質問情報を前記処理情報記憶部から検索し、その後の認識結果に肯定発話情報が含まれる場合、検索した質問情報に含まれるキーワードを前記インタラクティブプレゼンテーション部の処理の開始を示す処理要求情報と定める動作処理部と、を備え、前記タスク計画部は、処理要求情報とインタラクティブプレゼンテーション部の識別情報とを対応付けて記憶する記憶部から、前記要求理解エキスパート部の動作処理部が定めた処理要求情報に対応する識別情報が示すインタラクティブプレゼンテーション部を起動させることを特徴とする対話処理装置。

(3) 本発明のその他の態様は、前記要求理解エキスパート部の処理を継続するか否かの判定は、前記認識結果に肯定発話情報を含むか否かを示す特徴量に基づき、前記要求理解

10

20

30

40

50

エキスパート部の選択は、前記認識結果に前記所定の題目を含むか否かを示す特徴量を用いて推定された適合性に基づくことを特徴とする(1)又は(2)の対話処理装置。

【0009】

(4)本発明のその他の態様は、前記動作理解部は、前記推定した適合性に基づいて前記処理を継続するか否かを判定することを特徴とする(3)の対話処理装置である。

【0011】

(5)本発明のその他の態様は、前記複数のエキスパート部において前記処理情報記憶部に記憶された前記処理情報記憶部に記憶された質問情報及び応答情報が各々異なることを特徴とする(4)の対話処理装置である。

【0012】

(6)本発明のその他の態様は、予め定めた処理を行う機能を有する複数のエキスパート部を備える対話処理装置における対話処理方法において、前記対話処理装置が、入力した音声を認識する第1の過程と、前記複数のエキスパート部のうち前記処理を行うエキスパート部が、前記認識した認識結果に基づいて前記処理を継続するか否かを判定する第2の過程と、前記処理を行うエキスパート部が前記処理を継続しないと判断した場合、前記複数のエキスパート部のうち前記処理を行わないエキスパート部の各々が、前記認識した認識結果に基づいて適合性を推定する第3の過程と、前記対話処理装置が、前記推定した適合性に基づいて前記処理を行うエキスパート部を選択する第4の過程と、第5の過程と、を有し、前記複数のエキスパート部は、要求理解エキスパート部とインタラクティブプレゼンテーション部とを含み、前記インタラクティブプレゼンテーション部は、所定の題目に関する提示情報、及び前記提示情報に対する処理の形態を示す指示可能情報を記憶する動作情報記憶部を備え、前記認識結果に含まれる単語と一致する指示可能情報が示す形態で前記提示情報を処理し、前記要求理解エキスパート部は、キーワードを含んだ質問内容を表す質問情報及び肯定発話情報を記憶し、キーワードを含んだ応答情報を前記質問情報と対応付けて記憶した処理情報記憶部を備え、前記認識結果に含まれる単語と一致するキーワードを含んだ質問情報を前記処理情報記憶部から検索し、その後の認識結果に肯定発話情報が含まれる場合、検索した質問情報に含まれるキーワードを前記インタラクティブプレゼンテーション部の処理の開始を示す処理要求情報と定め、前記第5の過程は、前記対話処理装置が、処理要求情報とインタラクティブプレゼンテーション部の識別情報とを対応付けて記憶する記憶部から、前記要求理解エキスパート部が定めた処理要求情報に対応する識別情報が示すインタラクティブプレゼンテーション部を起動させることを特徴とする対話処理方法である。

【0013】

(7)本発明のその他の態様は、予め定めた処理を行う機能を有する複数のエキスパート部を備える対話処理装置のコンピュータに、入力した音声を認識する手順、前記複数のエキスパート部のうち前記処理を行うエキスパート部について、前記認識した認識結果に基づいて処理を継続するか否かを判定する手順、前記処理を継続しないと判断した場合、前記複数のエキスパート部のうち前記処理を行わないエキスパート部の各々について、前記認識した認識結果に基づいて適合性を推定する手順、前記推定した適合性に基づいて前記処理を行うエキスパート部を選択する手順、タスク計画手順、を実行させるための対話処理プログラムであって、前記複数のエキスパート部は、要求理解エキスパート部とインタラクティブプレゼンテーション部とを含み、前記インタラクティブプレゼンテーション部は、所定の題目に関する提示情報、及び前記提示情報に対する処理の形態を示す指示可能情報を記憶する動作情報記憶部を備え、前記認識結果に含まれる単語と一致する指示可能情報が示す形態で前記提示情報を処理し、前記要求理解エキスパート部は、キーワードを含んだ質問内容を表す質問情報及び肯定発話情報を記憶し、キーワードを含んだ応答情報を前記質問情報と対応付けて記憶した処理情報記憶部を備え、前記認識結果に含まれる単語と一致するキーワードを含んだ質問情報を前記処理情報記憶部から検索し、その後の認識結果に肯定発話情報が含まれる場合、検索した質問情報に含まれるキーワードを前記インタラクティブプレゼンテーション部の処理の開始を示す処理要求情報と定め、前記タス

10

20

30

40

50

ク計画手順は、処理要求情報とインタラクティブプレゼンテーション部の識別情報とを対応付けて記憶する記憶部から、前記要求理解エキスパート部が定めた処理要求情報に対応する識別情報が示すインタラクティブプレゼンテーション部を起動させる手順である対話処理プログラムである。

【発明の効果】

【0014】

本発明の一態様(1)、(2)、(6)、(7)によれば、予め定めた種類の処理を行っているエキスパート部の処理を継続するか否かが考慮されるため、利用者の発話に適したエキスパート部を選択することができる。

本発明のその他の態様(3)によれば、さらに利用者の発話に適したエキスパート部を選択することができる。

10

本発明のその他の態様(4)によれば、さらに現在予め定めた種類の処理を行っていないエキスパート部が利用者の発話に適したエキスパート部である可能性を考慮して、利用者の発話に適したエキスパート部を選択することができる。

本発明のその他の態様(5)によれば、さらに利用者の発話と、発話に対する応答が異なるエキスパート部の中から、利用者の発話に適したエキスパート部を選択することができる。

【図面の簡単な説明】

【0015】

【図1】本発明の実施形態に係る対話処理装置の構成を示す概略図である。

20

【図2】本実施形態に係る対話処理を示すフローチャートである。

【図3】本実施形態に係る音声認識部の構成の一例を示す概略図である。

【図4】本実施形態に係る動作理解部が行うエキスパート選択処理を示すフローチャートである。

【図5】本実施形態に係るエキスパート部の構成を示す概略図である。

【図6】本実施形態に係るRUエキスパート部の処理情報記憶部が記憶する処理情報の一例を表す概念図である。

【図7】本実施形態に係るQAエキスパート部の処理情報記憶部が記憶する処理情報の一例を表す概念図である。

【図8】本実施形態に係るIPエキスパート部の処理情報記憶部が記憶する処理情報の一例を表す概念図である。

30

【図9】本実施形態に係るエキスパート共通の特徴量情報の一例を示す表である。

【図10】本実施形態に係るRUエキスパート部の特徴量情報の一例を示す表である。

【図11】本実施形態に係るIPエキスパート部の特徴量情報の一例を示す表である。

【図12】本実施形態に係るQAエキスパート部の特徴量情報の一例を示す表である。

【図13】本実施形態に係るRUエキスパート部が用いる特徴量情報の一例を示す表である。

【図14】本実施形態に係るIPエキスパート部が用いる特徴量情報の一例を示す表である。

【図15】本実施形態に係るQAエキスパート部が用いる特徴量情報の一例を示す表である。

40

【図16】本実施形態に係る対話処理装置による対話の一例を示す表である。

【図17】本実施形態に係る対話処理装置が選択したエキスパート部の回数の一例を示す表である。

【図18】従来技術によって選択したエキスパート部の回数の一例を示す表である。

【発明を実施するための形態】

【0016】

本発明の実施形態は、利用者が発話した音声を入力し、入力した音声信号に対して音声認識を行い、予め定めた種類の処理を行う複数のエキスパート部のうち現在処理を行っているエキスパート部について、音声認識結果に基づいて処理を継続するか否かを判定する

50

。本実施形態は、処理を継続しないと判断した場合、複数のエキスパート部のうち現在起動していないエキスパート部の各々について、音声認識結果に基づいて適合性を推定する。本実施形態は、推定した適合性に基づいて処理を行うエキスパート部を選択する。

また、本実施形態は、さらに推定した適合性に基づいて現在処理を行っているエキスパート部の処理を継続するか否かを判定する。

【0017】

以下、図面を参照しながら本実施形態について説明する。

図1は、本実施形態に係る対話処理装置1の構成を示す概略図である。

対話処理装置1は、音声入力部101、音声認識部102、動作理解部103、エキスパート部104-1、エキスパート部104-2、エキスパート部104-3、タスク計画部106、情報出力部107及び情報提示部108を含んで構成される。

なお、エキスパート部104-1~104-3を総称する場合や区別しない場合には、エキスパート部104-1~104-3又は単にエキスパート部104と呼ぶ。

【0018】

音声入力部101は、受信した音波を音響信号に変換し、変換した音響信号を音声認識部102に出力する。音声入力部101は、例えばマイクロホンである。ここで、音声入力部101は、利用者が発話した音声に伴う音波を受信する。

音声認識部102は、音声入力部101から入力された音響信号に基づいて認識した音声認識情報を生成し、生成した音声認識情報を動作理解部103に出力する。音声認識部102の構成については後述する。

【0019】

音声認識部102は、予め設定された時間（例えば、1秒）よりも長い時間、予め定めた値よりも音響信号の強度（例えば、パワー）が小さい場合、無音状態と検出する。音声認識部102は、前後が無音状態で挟まれる区間を発話区間と判断する。音声認識部102は、発話区間毎に上述の音声認識情報を動作理解部103に出力する。

【0020】

動作理解部103は、現在処理を行っていないエキスパート部104から適合度情報を入力される。動作理解部103は、現在処理を行っているエキスパート部104から処理を継続するか否かを示す処理継続信号を入力される。

処理継続信号が処理を継続しないことを示す場合、動作理解部103は、入力された適合性情報に基づいて処理を開始するエキスパート部104を選択する。

なお、どのエキスパート部104も処理を行っておらず処理継続信号が入力されない場合（例えば、初期状態）、動作理解部103は、予め設定されたエキスパート部104を当初処理を開始するエキスパート部104と選択する。

動作理解部103は、選択したエキスパート部104に、処理を開始することを示す起動信号を出力する。

【0021】

対話処理装置1は、上述のように処理の種類が各々異なるエキスパート部104-1~104-3を複数個備える。図1に示す例では、エキスパート部の個数は3個であるが、本実施形態ではエキスパート部の個数は、1個よりも多ければ何個でもよい。エキスパート部（ドメインエキスパートとも呼ぶ）は、利用者との対話における予め定めた種類の処理や、利用者によって指示された種類の動作を行うための処理を行うモジュール（構成部）である。

エキスパート部104は、次のように共通の処理を行う機能をさらに有する。エキスパート部104は、動作理解部103から起動信号を入力された場合、動作理解部103から入力された音声認識情報に基づき予め設定された処理を行う。この処理は、エキスパート部104が備える動作処理部1045（後述）が行い、出力パターン情報を生成する。処理選択部107から処理選択信号を入力されたとき、生成した出力パターン信号を処理選択部107に出力する。起動信号が入力されたエキスパート部（処理を行うエキスパート部）は、動作理解部103から入力された音声認識情報に基づき処理を継続するか否か

10

20

30

40

50

判定し、その処理を継続するか否かを示す処理継続信号を生成する。そのエキスパート部は、生成した処理継続信号を動作理解部 103 に出力する。

動作理解部 103 から起動信号が入力されないエキスパート部（処理を行わないエキスパート部）は、動作理解部 103 から入力された音声認識情報に基づいて適合性情報を推定し、推定した適合性情報を動作理解部 103 に出力する。

【0022】

情報出力部 107 は、各エキスパート部 104 - 1 ~ 104 - 3 の動作処理部 1045 に対して一定時間間隔（例えば、10ms）で出力パターン情報の出力を要求することを示す出力要求信号を出力する。情報出力部 107 は、予め設定された処理を行っているエキスパート部 104 から出力パターン情報を入力され、入力された出力パターン情報を情報提示部 108 に出力する。

情報提示部 108 は、情報出力部 107 から入力された出力パターン情報を利用者が知覚可能な形態に変換して利用者に提示する。出力パターン情報が、応答情報のように形態がテキスト列を備える場合には、その応答情報に対してテキスト音声合成を行って音声信号を生成する。情報提示部 108 は生成した音声信号に基づく音を再生する。但し、出力パターン情報が指示可能情報である場合には、情報提示部 108 は出力パターン上右方を表す文字の画像に変換し、変換した画像を表示する。指示可能情報とは、利用者が発話によって情報提示部 108 に実行させることができる情報である。指示可能情報の例については後述する。出力パターン情報が、画像情報である場合には、その画像信号に基づく画像を表示する。

【0023】

次に、本実施形態に係る対話処理装置 1 が行う対話処理について説明する。

図 2 は、本実施形態に係る対話処理を示すフローチャートである。

（ステップ S101）音声入力部 101 は、受信した音波を音響信号に変換し、変換した音響信号を音声認識部 102 に出力する。その後、ステップ S102 に進む。

（ステップ S102）音声認識部 102 は、音声入力部 101 から入力された音響情報に基づいて認識した音声認識情報を生成し、生成した音声認識情報を発話区間毎に動作理解部 103 に出力する。その後、ステップ S103 に進む。

【0024】

（ステップ S103）エキスパート部 104 は、前回動作理解部 103 から起動信号を入力されたか否かを判断する。即ち、エキスパート部 104 は、予め設定された処理を行うか否かを判断する。エキスパート部 104 は、動作理解部 103 から起動信号を入力された場合（ステップ S103 Y）、ステップ S105 に進む。エキスパート部 104 は、動作理解部 103 から起動信号を入力されない場合（ステップ S103 N）、ステップ S104 に進む。

【0025】

（ステップ S104）エキスパート部 104 は、動作理解部 103 から入力された音声認識情報に基づいて適合性情報を推定し、推定した適合性情報を動作理解部 103 に出力する。その後、ステップ S107 に進む。

（ステップ S105）エキスパート部 104 は、動作理解部 103 から入力された音声認識情報に基づき予め設定された処理を行う。エキスパート部 104 は、自部が備える動作処理部 1045 において処理を行って出力パターン情報を生成する。処理選択部 107 から処理選択信号を入力されたとき、生成した出力パターン情報を処理選択部 107 に出力する。その後、ステップ S106 に進む。

（ステップ S106）エキスパート部 104 は、動作理解部 103 から入力された音声認識情報に基づき処理を継続するか否かを判定し、その処理を継続するか否かを示す処理継続信号を生成する。そのエキスパート部 104 は、生成した処理継続信号を動作理解部 103 に出力する。その後、ステップ S107 に進む。

【0026】

（ステップ S107）動作理解部 103 は、対話処理装置 1 が備える全てのエキスパート

10

20

30

40

50

部 104 - 1 ~ 104 - 3 についてステップ S 103 ~ S 106 の処理を行ったか否か判断する。動作理解部 103 が対話処理装置 1 が備える全てのエキスパート部 104 - 1 ~ 104 - 3 についてステップ S 103 ~ S 106 の処理を行ったと判断した場合（ステップ S 107 Y）、ステップ S 108 に進む。動作理解部 103 がステップ S 103 ~ S 106 の処理を行っていないエキスパート部 104 があると判断した場合（ステップ S 107 N）、ステップ S 103 に進む。

（ステップ S 108）動作理解部 103 は、現在予め設定された処理を行っていないエキスパート部 104 から適合性情報を入力される。動作理解部 103 は、現在予め設定された処理を行っているエキスパート部 104 から処理を継続するか否かを示す処理継続信号を入力される。処理継続信号が処理を継続しないこと（非継続）を示す場合、動作理解部 103 は、入力された適合性情報に基づいて次回起動するエキスパート部 104 を選択する。動作理解部 103 は選択したエキスパート部 104 に、処理を開始することを示す起動信号を出力する。その後、ステップ S 109 に進む。

【0027】

（ステップ S 109）情報出力部 107 は、各エキスパート部 104 - 1 ~ 104 - 3 の動作処理部 1045 に対して一定時間間隔で出力要求信号を出力する。情報出力部 107 は、予め設定された処理を行っているエキスパート部 104 から出力パターン情報を入力され、入力された出力パターン情報を情報提示部 108 に出力する。その後、ステップ S 110 に進む。

（ステップ S 110）情報提示部 108 は、情報出力部 107 から入力された出力パターン情報を利用者が知覚可能な形態に変換して利用者に提示する。その後、処理を終了する。

【0028】

次に、音声認識部 102 の構成の一例について説明する。

図 3 は、本実施形態に係る音声認識部 102 の構成の一例を示す概略図である。

音声認識部 102 は、特徴量算出部 1021、音響モデル記憶部 1022、音素決定部 1023、単語情報記憶部（単語辞書）1024、言語モデル記憶部 1025、単語決定部 1026 を含んで構成される。

特徴量算出部 1021 は、音声入力部 101 から入力された音響信号に基づいてフレーム（例えば 20ms）毎に音響特徴量（例えば、メル周波数ケプストラム（Mel frequency cepstral coefficient；MFCC））を算出する。特徴量算出部 1021 は、算出した音響特徴量を音素決定部 1023 に出力する。

【0029】

音響モデル記憶部 1022 には、複数個の音響特徴量に対応する対数確率値のセットを音素情報毎に備える音響モデル（例えば、隠れマルコフモデル（hidden Markov model；HMM））が予め記憶されている。

音素決定部 1023 は、特徴量算出部 1021 から入力された音響特徴量に対応する対数確率値のセットを含む音素情報を音響モデル記憶部 1022 から読み出す。音素決定部 1023 は、音響モデル記憶部 1022 から読み出した対数確率値を入力された音響特徴量の順番に加算して音響スコア（尤度）を算出する。音素決定部 1023 は、音響スコアが 1 番目から M（M は予め設定された 1 以上の整数）番目に大きい音素情報と各々の音響スコアを関連付けて音素候補情報として単語決定部 1026 及び動作理解部 103 に出力する。

【0030】

単語情報記憶部 1024 には、音素情報の組からなる単語情報（単語ごとの読み）が予め複数個記憶されている。

言語モデル記憶部 1025 には、複数（例えば 3 個）の単語情報の組と対数確率値（単語信頼度）のセット（単語間の接続制約）を複数個備える言語モデルが予め記憶されている。

単語決定部 1026 は、音素決定部 1023 から入力された音素候補情報に含まれる音

10

20

30

40

50

素情報の組に対応する単語情報を単語情報記憶部 1024 から読み出す。単語決定部 1026 は、読み出した単語情報の組に対応する単語信頼度を言語モデル記憶部 1025 から読み出し、読み出した単語信頼度を入力された音素候補情報が表す単語の順に加算して言語スコアを算出する。単語決定部 1026 は、算出した言語スコアが 1 番目から N (N は予め定めた 1 以上の整数) 番目に大きい単語情報と各々の言語スコアを単語候補情報と定める。単語決定部 1026 は、定めた単語候補情報に含まれる各単語情報の音素情報の組に対応する音響スコアを、入力された音素候補情報から抽出する。単語決定部 1026 は、定めた単語候補情報と抽出した音素候補情報を音声認識情報として動作理解部 103 に出力する。

【0031】

次に、動作理解部 103 が行うエキスパート選択処理について説明する。

図 4 は、本実施形態に係る動作理解部 103 が行うエキスパート選択処理を示すフローチャートである。

(ステップ S201) 動作理解部 103 は、音声認識部 102 から入力された音声認識情報を各エキスパート部 104 - 1 ~ 104 - 3 に出力する。その後、ステップ S202 に進む。

【0032】

(ステップ S202) 動作理解部 103 は、現在予め設定された処理を行っていないエキスパート部 104 から適合性情報を入力される。適合性情報とは、そのエキスパート部 104 が与えられた音声認識情報に適合する度合いを示す情報、例えばエキスパート部 104 が次回処理を行う確率値 P である。確率値 P を算出する処理については後述する。その後、ステップ S203 に進む。

【0033】

(ステップ S203) 動作理解部 103 は、現在予め設定された処理を行っているエキスパート部 104 から処理を継続するか否かを示す処理継続信号を入力される。

入力された処理継続信号が、そのエキスパート部 104 の処理を継続することを示す (例えば、信号値 1) 場合 (ステップ S203 Y)、動作理解部 103 は、処理を終了する。入力された処理継続信号が、そのエキスパート部 104 の処理を継続しないことを示す (例えば、信号値 0) 場合 (ステップ S203 N)、ステップ S204 に進む。

【0034】

(ステップ S204) 動作理解部 103 は、入力された適合性情報が示す適合性が最も高いエキスパート部 104 を、次回予め設定された処理を行うエキスパート部 104 と決定する。動作理解部 103 は、次回予め設定された処理を行うエキスパート部に、その処理を行うことを指示する起動信号を出力する。その後、処理を終了する。

【0035】

次に、エキスパート部 104 が共通に備える構成について説明する。

図 5 は、本実施形態に係るエキスパート部 104 の構成を示す概略図である。

エキスパート部 104 は、特徴量生成部 1041、処理情報記憶部 1042、適合性推定部 1043、処理継続判定部 1044、動作処理部 1045 及び対話履歴記憶部 1046 を含んで構成される。

【0036】

特徴量生成部 1041 は、動作理解部 103 から入力された音声認識情報に基づいて特徴量情報を生成する。特徴量情報とは、与えられた音声認識情報に対してエキスパート部 104 毎の動作の特徴を表す特徴量の組からなる情報である。特徴量の具体例については、後述する。特徴量生成部 1041 は、特徴量情報を生成するために入力された音声認識情報の一部から特徴量情報の一部を抽出する。特徴量生成部 1041 は、入力された音声認識情報の他の部分に対応する処理情報を処理情報記憶部 1042 から読み出し、読み出した処理情報に基づいて特徴量情報の他の部分を生成する。さらに他の特徴量情報については、特徴量生成部 1041 は、対話履歴記憶部 1046 から読み出した対話履歴情報に基づいて生成する。特徴量の具体例及び生成方法については後述する。

10

20

30

40

50

【 0 0 3 7 】

特徴量生成部 1 0 4 1 は、処理継続判定部 1 0 4 4 から処理を継続しないことを示す処理継続信号を入力されたときから動作理解部 1 0 3 から起動信号を入力されるまで、生成した特徴量情報を適合性推定部 1 0 4 3 に出力する。つまり、該当するエキスパート部 1 0 4 (具体的には動作処理部 1 0 4 5) が予め設定された処理を行わない場合に、適合性推定部 1 0 4 3 が適合性情報を推定する。

特徴量生成部 1 0 4 1 は、動作理解部 1 0 3 から起動信号を入力されたとき、又は処理継続判定部 1 0 4 4 から処理を継続することを示す処理継続信号を入力されたとき、生成した特徴量情報を処理継続判定部 1 0 4 4 に出力する。つまり、該当するエキスパート部 1 0 4 (具体的には動作処理部 1 0 4 5) が処理を行う場合に、処理継続判定部 1 0 4 4 が処理の継続・非継続を判定する。

なお、後述するように特徴量生成部 1 0 4 1 が出力する特徴量情報が含む特徴量の組は、適合性推定部 1 0 4 3 に出力するものと処理継続判定部 1 0 4 4 に出力するものとで異なってもよい。

【 0 0 3 8 】

処理情報記憶部 1 0 4 2 は、そのエキスパート部 1 0 4 が予め設定された処理を実行するための処理に用いる情報やその処理を特徴付ける情報を処理情報として記憶しておく。処理情報の詳細については後述する。

【 0 0 3 9 】

適合性推定部 1 0 4 3 は、特徴量生成部 1 0 4 1 から入力された特徴量情報 (特徴量ベクトル F) に基づいて、当該エキスパート部 1 0 4 に係る適合性情報を推定する。適合性情報とは、エキスパート部 1 0 4 が入力された音声認識情報に適合する度合いを示す情報である。具体的には、適合性情報は、そのエキスパート部 1 0 4 が次回起動する確率値 P である。適合性推定部 1 0 4 3 は、例えば式 (1) を用いて確率値 P を算出する。

【 0 0 4 0 】

【 数 1 】

$$P = \frac{1}{1 + \exp\left(-\left(a_0 + \sum_{l=1}^n a_l F_l\right)\right)} \quad \dots (1)$$

【 0 0 4 1 】

式 (1) において、 F_1, \dots, F_n は、特徴量ベクトル F の要素値 (特徴量) である。 n は、特徴量ベクトル F の要素数である。 a_1, \dots, a_n は、要素値 F_1, \dots, F_n の各々に対する乗算係数であって、実数値をとる。 a_0 は、バイアス値であって、実数値をとる。適合性推定部 1 0 4 3 は、エキスパート部 1 0 4 毎の特徴量情報と起動・非起動 (つまり予め設定した処理を行うか否か) の間の既知の関係を用いて、事前学習によって係数 a_0, \dots, a_n を予め算出し、設定しておく。

確率値 P は、特徴量の線形結合値 ($a_0 + \sum_{l=1}^n a_l F_l$) が増加するに従って単調に増加し、最小値 0 から最大値 1 までの間の実数値である。この線形結合値が 0 に近づくとき、確率値 P は 1 に漸近し、この線形結合値が $-\infty$ に近づくとき P は 0 に漸近する。

適合性推定部 1 0 4 3 は、推定した適合性情報を動作理解部 1 0 3 に出力する。

【 0 0 4 2 】

処理継続判定部 1 0 4 4 は、特徴量生成部 1 0 4 1 から入力された特徴量情報 (特徴量ベクトル F) に基づいて、当該エキスパート部 1 0 4 の動作を継続するか否か (即ち、動作処理部 1 0 4 5 が予め設定した処理を行うか否か) を判断する。処理継続判定部 1 0 4 4 は、動作を継続するか否かを示す処理継続信号を生成する。処理継続信号の値は、例え

ば1又は0である。処理継続信号の値が1の場合、予め設定した処理を継続することを示す。処理継続信号の値が0の場合、予め設定した処理を継続しないことを示す。処理継続判定部1044は、生成した処理継続信号を特徴量生成部1041、動作処理部1045及び動作理解部103に出力する。

【0043】

処理継続判定部1044は、現在予め設定された処理を行っているエキスパート部104が、その処理を継続するか否かを判断するために、例えばサポートベクターマシン(support vector machine; SVM)を用いる。サポートベクターマシンを用いる場合、処理継続判定部1044は、例えば、入力された特徴量情報を表す特徴量ベクトルFと重み係数を要素とする重み係数ベクトルwの内積を算出する。処理継続判定部1044は、算出した内積が予め定めた実数値b(例えば、ゼロ)と等しい又は大きい場合、処理を継続すると判断する。処理継続判定部1044は、算出した内積が予め定めた実数値bよりも小さい場合、処理を継続しない(非継続)と判断する。

10

【0044】

但し、処理継続判定部1044は、エキスパート部104毎に特徴量情報と動作継続・非継続(予め設定された処理を行うか否か)の既知の関係を用いて、事前学習によって重み係数ベクトルw及び実数値bを予め算出し、設定しておく。事前学習においては、例えば、上述の内積と実数値bの和である実数値yが1又は1よりも大きいという条件のもとで、重み係数ベクトルwのノルムが最小とする条件を課す。事前学習において算出する実数値yは、特徴量ベクトルFについて予め設定された処理を継続する場合に値1をとり、

20

【0045】

動作処理部1045は、処理継続判定部1044から処理を継続することを示す処理継続信号を入力された場合、又は、動作理解部103から起動信号を入力された場合、動作理解部103から入力された音声認識情報に基づいて予め定めた処理を行う。例えば、動作処理部1045は、音声認識情報に含まれる単語情報(キーワード)を備える発話パターン情報(例えば、質問情報)を処理情報記憶部1042から検索し、検索した発話パターン情報と対応する出力パターン情報(例えば、応答情報及び提示情報)を処理情報として処理情報記憶部1042から読み出す処理を行う。ここで、処理情報記憶部1042は、予め発話パターン情報と出力パターン情報を対応付けて処理情報として記憶している。

30

【0046】

動作処理部1045は、情報出力部107から出力パターン情報の出力を要求することを示す出力要求信号を入力されたとき、読み出した出力パターン情報を情報出力部107に出力する。動作処理部1045は、読み出した発話パターン情報と出力パターン情報(又はこれらの識別情報)を時刻情報と対応付けて対話履歴記憶部1046に記憶する。従って、対話履歴記憶部1046には、時刻情報毎の発話パターン情報と出力パターン情報の情報列が対話履歴として記憶される。

【0047】

動作処理部1045は、エキスパート部104-1~104-3毎に各々異なる予め設定された処理を行う。エキスパート部104-1~104-3が、各々RU(request understanding; 要求理解)エキスパート部104-1、QA(question answering; 質問応答)エキスパート部104-2、IP(interactive presentation; インタラクティブプレゼンテーション)エキスパート部104-3である場合を例にとって説明する。

40

以下の説明では、各機能部(特徴量生成部1041等)を、各エキスパート部104毎に区別する場合には、特徴量生成部1041-1等と呼ぶ。ここで、末尾の-1等は、エキスパート部104-1等を示す。

【0048】

RUエキスパート部104-1は、入力された音声認識情報からIPエキスパート部104-3の処理を開始させることを示す処理要求情報を抽出し、抽出した処理要求情報を

50

確認するための対話処理（後述）を行う。RUエキスパート部104-1は、確認した処理要求情報をタスク計画部106（図示せず）に出力する。

【0049】

タスク計画部106は、対話処理装置1が備える構成部の一つである。タスク計画部106は、RUエキスパート部104-1の動作処理部1045から入力された処理要求情報に基づいて処理を開始させるIPエキスパート部104-3を選択する。タスク計画部106は、自部が備える記憶部に処理要求情報と処理を開始させるIPエキスパート部104-3の識別情報を対応付けて予め記憶させておく。タスク計画部106は、入力された処理要求情報に対応するIPエキスパート部104-3の識別情報を読み出し、読み出した識別情報が示すIPエキスパート部104-3の動作処理部1045-3に処理を開始することを指示する起動信号を出力する。動作処理部1045-3は、動作理解部103以外に、タスク計画部106からも起動信号を入力されることがある。動作処理部1045-3は、起動信号を入力されたとき、動作理解部103から入力された音声認識情報に基づいて予め設定された処理（後述）を行う。

10

【0050】

次に、RUエキスパート部104-1の処理情報記憶部1042-1が記憶する処理情報について動作処理部1045が行う処理とともに説明する。

図6は、本実施形態に係るRUエキスパート部104-1の処理情報記憶部1042-1が記憶する処理情報の一例を表す概念図である。

図6に示すように、処理情報記憶部1042-1は、発話パターン情報として質問情報、肯定発話情報、否定発話情報を記憶する。処理情報記憶部1042-1は、質問情報と対応付けて応答情報を出力パターン情報として記憶する。

20

質問情報とは、利用者が対話処理装置1に対して質問する内容を表す情報である。図6の最上行の「グリニッジ ニ ツイテ オシエテ クダサイ」とは、利用者が「グリニッジ」について質問することを表す質問情報である。そのうち、質問情報の一部として含まれ下線が付された部分「グリニッジ」、「オシエテ」は、キーワード情報を示す。キーワード情報とは、質問情報として必須であり、動作処理部1045-1が検索の手がかりとして音声認識情報に含まれる単語情報と照合する語句を表す情報である。

【0051】

図6の第2行の「カジットシ グリニッジ ノ セツメイ デスネ」とは、利用者に対してグリニッジに対する説明を要求することについて確認を促す応答情報（出力パターン情報）である。第3行の「ハイ」、「ウン」とは、直前に提示された応答情報又は提示情報に対して肯定的な発話を表す肯定発話情報である。第4行の「イエエ」、「イヤ」とは、直前に提示された応答情報又は提示情報に対して否定的な発話を表す否定発話情報である。

30

【0052】

従って、動作処理部1045-1は、入力された音声認識情報に含まれる単語情報として「グリニッジ」及び「オシエテ」を含む場合、処理情報記憶部1042-1から「グリニッジ ニ ツイテ オシエテ クダサイ」を示す質問情報を読み出す。動作処理部1045は、この質問情報に対応する応答情報として「カジットシ グリニッジ ノ セツメイ デスネ」を示す応答情報を処理情報記憶部1042-1から読み出す。動作処理部1045は、この読み出した応答情報を情報出力部107に出力し、読み出した質問情報と応答情報を時刻情報と対応付けて対話履歴記憶部1046-1に記憶する。動作処理部1045-1は、この読み出した質問情報のキーワードである「グリニッジ」、「セツメイ」を処理要求情報の候補として自部が備える記憶部に記憶する。

40

【0053】

動作処理部1045-1は、入力された音声認識情報に含まれる単語情報として「ハイ」を含む場合、処理情報記憶部1042-1から「ハイ」を示す肯定発話情報を読み出す。動作処理部1045-1は、読み出した肯定発話情報を対話履歴記憶部1046-1に時刻情報と対応付けて記憶する。

50

動作処理部 1045 - 1 は、肯定発話情報を読み出した場合、記憶した処理要求情報の候補「グリニッジ」、「セツメイ」を処理要求情報と定め、定めた処理要求情報をタスク計画部 106 に出力する。

【0054】

動作処理部 1045 - 1 は、入力された音声認識情報に含まれる単語情報として「イイエ」を含む場合、処理情報記憶部 1042 - 1 から「イイエ」を示す否定発話情報を読み出す。動作処理部 1045 - 1 は、読み出した否定発話情報を時刻情報と対応付けて対話履歴記憶部 1046 - 1 に記憶する。

動作処理部 1045 - 1 は、否定発話情報を読み出した場合、記憶した処理要求情報の候補「グリニッジ」、「セツメイ」を自部が備える記憶部から削除する。

10

【0055】

QAエキスパート部 104 - 2 は、予め定めた題目（例えば、世界遺産）に関する質問情報と応答情報の対を記憶する処理情報記憶部 1042 - 2 を備える。QAエキスパート部 104 - 2 は、音声認識情報に含まれる単語情報とキーワード情報が照合する質問情報に対応する応答情報を出力パターン情報として情報出力部 107 に出力する。

ここで、QAエキスパート部 104 - 2 の処理情報記憶部 1042 - 2 が記憶する処理情報について動作処理部 1045 が行う処理とともに説明する。

図7は、本実施形態に係るQAエキスパート部 104 - 2 の処理情報記憶部 1042 - 2 が記憶する処理情報の一例を表す概念図である。

図7に示すように、処理情報記憶部 1042 - 2 は、発話パターン情報として質問情報、肯定発話情報、否定発話情報、相槌情報を記憶する。処理情報記憶部 1042 - 2 は、質問情報と対応付けて応答情報と提示情報を出力パターン情報として記憶する。

20

図7の最上行の「イタリア ノ セカイイサン ノ リスト ヲ ミセテ クダサイ」とは、利用者が「イタリアの世界遺産のリストを見せ」ることを要求する質問情報である。そのうち、質問情報の一部として含まれ下線が付された部分「イタリア」、「セカイイサン」、「リスト」、「ミセテ」は、キーワード情報を示す。

【0056】

図7の第2行の「イタリア ノ セカイイサン ヲ チイキベツ ニ ショウカイ シマス」とは、利用者に対してイタリアの世界遺産を地域別に紹介することを示す応答情報（出力パターン情報）である。この応答情報は、同時に第4行に示す、イタリアの地域別の世界遺産のリストを示す画像情報（提示情報）を提示することについて利用者に対して注意を促す応答情報である。

30

第3行の「ドノ チイキ ノ リスト ガ ミタイ カ オッシャッテ クダサイ」とは、提示している画像情報を参照してどの地域のリストを提示すべきか利用者に対して指示を促す応答情報である。

図7の第5行の「ハイ」、「ウン」とは、直前に提示された応答情報又は提示情報に対しする肯定発話情報である。第6行の「イイエ」、「イヤ」とは、直前に提示された応答情報又は提示情報に対する否定発話情報である。

図7の第7行の「ナルホド」、「ソウデスネ」とは、直前に提示された応答情報に対して利用者が反応を示す相槌情報である。

40

【0057】

従って、動作処理部 1045 - 2 は、入力された音声認識情報に含まれる単語情報として「イタリア」、「セカイイサン」、「リスト」及び「ミセテ」を含む場合、処理情報記憶部 1042 - 2 から「イタリア ノ セカイイサン ノ リスト ヲ ミセテ クダサイ」を示す質問情報を読み出す。動作処理部 1045 - 2 は、この質問情報に対応する応答情報として「イタリア ノ セカイイサン ノ リスト ヲ ミセテ クダサイ。ドノ チイキ ノ リスト ガ ミタイ カ オッシャッテ クダサイ」を示す応答情報とイタリアの地域別の世界遺産のリストを示す画像情報を提示情報として読み出す。動作処理部 1045 - 2 は、この読み出した応答情報と提示情報を出力パターン情報として情報出力部 107 に出力し、読み出した質問情報を発話パターン情報として、応答情報及び提示

50

情報を出力パターン情報として時刻情報と対応付けて対話履歴記憶部 1046 - 2 に記憶する。

【0058】

動作処理部 1045 - 2 は、入力された音声認識情報に含まれる単語情報として「ハイ」を含む場合、処理情報記憶部 1042 - 2 から「ハイ」を示す肯定発話情報を読み出す。動作処理部 1045 - 2 は、読み出した肯定発話情報を対話履歴記憶部 1046 - 2 に時刻情報と対応付けて記憶する。

動作処理部 1045 - 2 は、入力された音声認識情報に含まれる単語情報として「いいえ」を含む場合、処理情報記憶部 1042 - 2 から「イエエ」を示す否定発話情報を読み出す。動作処理部 1045 - 2 は、読み出した否定発話情報を対話履歴記憶部 1046 - 2 に時刻情報と対応付けて記憶する。

10

動作処理部 1045 - 2 は、入力された音声認識情報に含まれる単語情報として「なるほど」を含む場合、処理情報記憶部 1042 - 2 から「ナルホド」を示す相槌情報を読み出す。動作処理部 1045 - 2 は、読み出した相槌情報を対話履歴記憶部 1046 - 2 に時刻情報と対応付けて記憶する。

【0059】

IPエキスパート部 104 - 3 は、予め定めた題目（トピック）のうち1つの事項（例えば、著名な世界遺産のうちの1つ）に関する提示情報（映像情報、音声情報）と応答情報を質問情報と対応付けて記憶する動作情報記憶部 1042 - 3 を備える。IPエキスパート部 104 - 3 は、音声認識情報に含まれる単語情報とキーワード情報が照合する質問情報に対応する応答情報及び提示情報を出力パターン情報として情報出力部 107 に出力する。

20

動作情報記憶部 1042 - 3 に記憶されている動作情報は、最初に出力すべき応答情報と提示情報の組を初期出力情報として含む。IPエキスパート部 104 - 3 は、タスク計画部 106 から処理を開始することを指示する起動信号が入力された場合、初期出力情報を情報出力部 107 に出力し、対話履歴記憶部 1046 - 3 に記憶する。

【0060】

ここで、IPエキスパート部 104 - 3 の処理情報記憶部 1042 - 3 が記憶する処理情報について動作処理部 1045 - 3 が行う処理とともに説明する。

図8は、本実施形態に係るIPエキスパート部 104 - 3 の処理情報記憶部 1042 - 3 が記憶する処理情報の一例を表す概念図である。

30

図8に示すように、処理情報記憶部 1042 - 3 は、発話パターン情報として質問情報、指示可能情報を記憶する。処理情報記憶部 1042 - 3 は、応答情報と提示情報を出力パターン情報として記憶する。処理情報記憶部 1042 - 3 は、質問情報と対応付けて記憶する応答情報と提示情報を出力パターン情報として記憶する。

【0061】

図8の最上行から第3行に太字で示される、応答情報と提示情報の組が初期出力情報である。図8の最上行及び第2行の「カイトシ グリニッジ ノ アンナイ ヲ シマス / グリニッジヒョウジュンジ ノ キジュン トナル トシトシテ ユイショ アル ミナトマチ デス」とは、利用者に海事都市グリニッジの案内を開始することを示し、第3行のグリニッジ市内を示す画像情報を提示することについて注意を喚起する応答情報である。

40

図8の第4行の「グリニッジヒョウジュンジ ッテ ナンデスカ?」とは、利用者が対話処理装置1に対してグリニッジ標準時に対して問い合わせる質問情報である。そのうち、質問情報の一部として含まれ下線が付された部分「グリニッジヒョウジュンジ」、「ナンデスカ」は、キーワード情報を示す。

【0062】

図8の第5行の「グリニッジテンモンダイ デノ ハイキンタイヨウジ ノ コトデス」とは、利用者に対してグリニッジ標準時に対する回答としてグリニッジ天文台での平均太陽時を示す応答情報（出力パターン情報）である。この応答情報は、同時に第6行に

50

示す、グリニッジ天文台の内容を示す画像情報（提示情報）を提示することについて利用者に対して注意を促す応答情報である。

図8の第7行の「トマル」、「ツギ」、「モデル」、「モウイチド」とは、利用者1が提示情報を提示する処理の形態を指示する指示可能情報である。「トマル」とは、画像情報の提示を停止することを指示する情報である。「ツギ」とは、現在提示されている画像情報の直後に提示される画像情報を提示することを指示する情報である。「モデル」とは、現在提示されている画像情報の直前に提示される画像情報を提示することを指示する情報である。「モウイチド」とは、現在提示している画像情報を再度初頭から提示することを指示する情報である。

【0063】

従って、動作処理部1045-3は、タスク計画部106から起動信号が入力された場合、初期出力情報（応答情報、提示情報）を処理情報記憶部1042-3から読み出す。動作処理部1045-3は、読み出した初期出力情報を出力パターン情報として情報出力部107に出力し、対話履歴記憶部1046に記憶する。

動作処理部1045-3は、入力された音声認識情報に含まれる単語情報として「グリニッジヒョウジュンジ」及び「ナンデスカ」を含む場合、処理情報記憶部1042-3から「グリニッジヒョウジュンジ ッテ ナンデスカ？」を示す質問情報を読み出す。動作処理部1045-3は、この質問情報に対応する応答情報として「グリニッジテンモンダイ デノ ヘイキンタイヨウジ ノ コトデス」を示す応答情報とグリニッジ天文台の内容を示す画像情報を提示情報として読み出す。動作処理部1045-3は、この読み出した応答情報と提示情報を出力パターン情報として情報出力部107に出力し、読み出した質問情報を発話パターン情報として、応答情報及び提示情報を出力パターン情報として、時刻情報と対応付けて対話履歴記憶部1046-3に記憶する。

【0064】

動作処理部1045-3は、入力された音声認識情報に含まれる単語情報として「トマル」を含む場合、処理情報記憶部1042-3から「トマル」を示す指示可能情報を読み出す。動作処理部1045-3は、現在出力している出力パターン情報の情報出力処理部107への出力を停止する。動作処理部1045-3は、読み出した指示可能情報を対話履歴記憶部1046-3に出力パターン情報として時刻情報と対応付けて記憶する。

動作処理部1045-3は、入力された音声認識情報に含まれる単語情報として「ツギ」を含む場合、処理情報記憶部1042-3から「ツギ」を示す指示可能情報を読み出す。動作処理部1045-3は、現在出力している提示情報よりも後に提示されるべき提示情報を出力パターン情報として情報出力部107へ出力する。動作処理部1045-3は、読み出した指示可能情報を対話履歴記憶部1046-3に出力パターン情報として時刻情報と対応付けて記憶する。

動作処理部1045-3は、入力された音声認識情報に含まれる単語情報として「モデル」を含む場合、処理情報記憶部1042-3から「モデル」を示す指示可能情報を読み出す。動作処理部1045-3は、現在出力している提示情報よりも前に提示されるべき提示情報を出力パターン情報として情報出力部107へ出力する。動作処理部1045-3は、読み出した指示可能情報を対話履歴記憶部1046-3に出力パターン情報として時刻情報と対応付けて記憶する。

動作処理部1045-3は、入力された音声認識情報に含まれる単語情報として「モウイチド」を含む場合、処理情報記憶部1042から「モウイチド」を示す指示可能情報を読み出す。動作処理部1045-3は、現在出力している提示情報を初頭から再度、出力パターン情報として情報出力部107へ出力する。動作処理部1045-3は、読み出した指示可能情報を対話履歴記憶部1046に出力パターン情報として時刻情報と対応付けて記憶する。

【0065】

次に、各エキスパート部の特徴量生成部1041が生成する特徴量情報の具体例について説明する。

10

20

30

40

50

以下の説明において、特徴量 F_{i, r_x} は、入力された音声認識情報に基づく特徴量情報である。特徴量 F_{i, h_x} は、対話履歴記憶部 1046 から読み出した対話履歴情報に基づく特徴量情報である。 $SRR_{i, j}$ は、音声認識情報においてエキスパートクラス i において j 番目に言語スコアが高い単語情報及び対応する音素情報である。エキスパートクラスとは、エキスパート部の種類を意味する。 i は、エキスパートクラスの何れかを示すインデックスである。 i が RU のとき、 RU エキスパート部 104-1 を示す。 i が QA のとき、 QA エキスパート部 104-2 を示す。 $i = IP$ のとき、 IP エキスパート部 104-3 を示す。 $SRR_{i, a_{11}}$ は、音声認識情報においてエキスパートクラス i において 1 番目から N 番目に言語スコアが高い単語情報及び対応する音素情報の全てである。 $SRR_{1v, j}$ は、大語彙言語モデルを用いた発話検証用の音声認識情報において j 番目に言語スコアが高い単語情報及び対応する音素情報の全てである。大語彙言語モデルとは、言語モデル記憶部 1025 に記憶される言語モデルであって、このモデルを構成する単語列に用いられる単語の種類の数（例えば、60、250語、）が多い言語モデルである（通常の言語モデルでは、例えば5、000語）。

10

【0066】

図9は、本実施形態に係るエキスパート部104共通の特徴量情報の一例を示す表である。

図9の最左列から右側へ順に、関連があるエキスパートクラス（種類）、特徴量、内容を示す。図9の最左列においてエキスパートクラスが共通とは、 RU エキスパート部 104-1、 QA エキスパート部 104-2、 IP エキスパート部 104-3 が共通に生成する特徴量情報を示す。

20

【0067】

図9の上から第2行の特徴量 F_{i, r_1} は、 $SRR_{i, 1}$ が得られたかどうかを示す特徴量である。 $SRR_{i, 1}$ が得られた場合、 F_{i, r_1} の値が1であり、 $SRR_{i, 1}$ が得られなかった場合、 F_{i, r_1} の値が0である。

図9の上から第3行の特徴量 F_{i, r_2} は、 $SRR_{i, j}$ に含まれる単語情報としてフィルターが含まれるかどうかを示す特徴量である。フィルターとは、意図をもった発話の間に挟みこむ語句、例えば「ええと」、「あのう」である。 $SRR_{i, 1}$ にフィルターが含まれる場合、 F_{i, r_2} の値が1であり、 $SRR_{i, 1}$ にフィルターが含まれない場合、 F_{i, r_2} の値が0である。

30

【0068】

図9の上から第4行の特徴量 F_{i, r_3} は、 $SRR_{i, 1}$ に含まれる単語情報に係る単語信頼度の最小値を示す。

図9の上から第5行の特徴量 F_{i, r_4} は、 $SRR_{i, 1}$ に含まれる単語情報に係る単語信頼度の平均値を示す。

図9の上から第6行の特徴量 F_{i, r_5} は、 $SRR_{i, 1}$ に含まれる音韻情報に係る音響スコアを発話区間の時間で除算した値を示す。

図9の上から第7行の特徴量 F_{i, r_6} は、 $SRR_{i, 1}$ の言語スコアを示す。

図9の上から第8行の特徴量 F_{i, r_7} は、 $SRR_{i, 1}$ に含まれる単語数を示す。

図9の上から第9行の特徴量 F_{i, r_8} は、 $SRR_{i, a_{11}}$ に含まれる単語数を示す

40

図9の上から第10行の特徴量 F_{i, r_9} は、特徴量 F_{i, r_5} と、 $SRR_{1v, 1}$ の音響スコアを発話区間の時間で除算した値との差分値を示す。

【0069】

図10は、本実施形態に係る RU エキスパート部 104-1 の特徴量情報の一例を示す表である。

図10の最左列から右側へ順に、関連があるエキスパートクラス（種類）、特徴量、内容を示す。図10の最左列においてエキスパートクラスが RU とは、 RU エキスパート部 104-1 の特徴量生成部 1041-1 のみが生成する特徴量情報を示す。

図10の上から第2行の特徴量 $F_{RU, r_{10}}$ は、 $SRR_{RU, 1}$ が肯定発話情報を含

50

むかどうかを示す特徴量である。SRR_{RU,1}が肯定発話情報を含む場合、F_{RU,r10}の値が1であり、SRR_{RU,1}が肯定発話情報を含まない場合、F_{RU,r10}の値が0である。特徴量生成部1041-1は、SRR_{RU,1}が肯定発話情報を含むかどうかを判断するために、処理情報記憶部1042-1に記憶されている肯定発話情報と、SRR_{RU,1}に含まれる単語情報を照合する。

【0070】

図10の上から第3行の特徴量F_{RU,r11}は、SRR_{RU,2}が否定発話情報を含むかどうかを示す特徴量である。SRR_{RU,2}が否定発話情報を含む場合、F_{RU,r11}は値1であり、SRR_{RU,2}が否定発話情報を含まない場合、F_{RU,r11}の値が0である。特徴量生成部1041-1は、SRR_{RU,2}が否定発話情報を含むかどうかを判断するために、処理情報記憶部1042-1に記憶されている否定発話情報と、SRR_{RU,2}に含まれる単語情報を照合する。

10

【0071】

図10の上から第4行の特徴量F_{RU,r12}は、音声認識情報に含まれる認識結果となる単語情報のうち、言語モデルによる認識結果の候補数を示す。

図10の上から第5行の特徴量F_{RU,r13}は、SRR_{RU,1}に世界遺産の名称が含まれているか否かを示す特徴量である。SRR_{RU,1}に世界遺産の名称が含まれている場合、F_{RU,r13}は値1であり、SRR_{RU,1}に世界遺産の名称が含まれていない場合、F_{RU,r13}は値0である。特徴量生成部1041-1は、SRR_{RU,1}が世界遺産の名称を含むかどうかを判断するために、処理情報記憶部1042-1に記憶されている質問情報に含まれるキーワードのうち世界遺産の名称を示す部分と照合する。

20

【0072】

図10の上から第6行の特徴量F_{RU,h1}は、対話履歴記憶部1046-1から読み出した音声認識情報であるSRR_{RU,1}が肯定発話情報を含むかどうかを示す特徴量である。SRR_{RU,1}が肯定発話情報を含む場合、F_{RU,h1}の値が1であり、SRR_{RU,1}が肯定発話情報を含まない場合、F_{RU,h1}の値が0である。特徴量生成部1041-1は、SRR_{RU,1}が肯定発話情報を含むかどうかを判断するために、処理情報記憶部1042-1に記憶されている肯定発話情報と、SRR_{RU,1}に含まれる単語情報を照合する。特徴量F_{RU,h1}は、処理継続判定部1044-1において処理を継続するか否かを判定するために用いられるが、適合性推定部1043-1において確率値Pを算出するためには用いられない。

30

【0073】

図10の上から第7行の特徴量F_{RU,h2}は、RUエキスパート部104-1に遷移した後のターン数を示す特徴量である。ターン数とは、利用者による発話、対話制御装置1からの情報提示（提示情報の提示、応答情報の提示）の回数である。特徴量生成部1041-1は、対話履歴記憶部1046-1から対話履歴を読み出し、動作理解部103-1から起動信号を入力された時刻から後の読み出した対話履歴に含まれる質問情報、応答情報、肯定発話情報及び否定発話情報の総数を算出する。

【0074】

図10の上から第8行の特徴量F_{RU,h3}は、RUエキスパート部104-1に遷移した後の否定発話の回数を示す特徴量である。特徴量生成部1041-1は、対話履歴記憶部1046-1から対話履歴を読み出し、動作理解部103-1から起動信号を入力された時刻から後の読み出した対話履歴に含まれる否定発話情報の総数を算出する。

40

図10の上から第9行の特徴量F_{RU,h4}は、特徴量F_{RU,h3}を特徴量F_{RU,h2}で除算した特徴量である。

【0075】

図11は、本実施形態に係るIPエキスパート部104-3の特徴量情報の一例を示す表である。

図11の最左列から右側へ順に、関連があるエキスパートクラス（種類）、特徴量、内容を示す。図11の最左列においてエキスパートクラスがIPとは、IPエキスパート部

50

104-3の特徴量生成部1041-3のみが生成する特徴量情報を示す。

【0076】

図11の上から第2行の特徴量 $F_{IP, r10}$ は、 $SRR_{IP, 1}$ に含まれる単語情報と照合するキーワード情報を含む質問情報が処理情報記憶部1042-3にあるか否かを示す特徴量である。このような質問情報がある場合、 $F_{IP, r10}$ の値は1であり、このような質問情報がない場合、 $F_{IP, r10}$ の値は0である。

図11の上から第3行の特徴量 $F_{IP, r11}$ は、質問情報に含まれるキーワードと照合する単語であって、 $SRR_{IP, j}$ に含まれる単語情報における単語数を、 $SRR_{IP, j}$ に含まれる単語数で除算した値を、全ての認識結果jにわたる総和をとり、さらに認識結果の個数で除算した値である。この質問情報とは、処理情報記憶部1042-3に記憶されている質問情報である。これは、特徴量 $F_{IP, r12} \sim F_{IP, r16}$ においても同様である。

10

【0077】

図11の上から第4行の特徴量 $F_{IP, r12}$ は、質問情報に含まれるキーワードkと照合する単語であって、 $SRR_{IP, all}$ に含まれる単語情報における単語数の全認識候補内での個数を、認識結果の個数で除算した値であって、キーワードk間の最小値を示す。

図11の上から第5行の特徴量 $F_{IP, r13}$ は、質問情報に含まれるキーワードkと照合する単語であって、 $SRR_{IP, all}$ に含まれる単語情報における単語数の全認識候補内での個数を、認識結果の個数で除算した値であって、キーワードk間の最大値を示す。

20

【0078】

図11の上から第6行の特徴量 $F_{IP, r14}$ は、質問情報に含まれるキーワードkに照合する単語であって第1候補 $SRR_{IP, 1}$ に含まれる単語情報における単語に対応する単語信頼度のキーワードk間の平均値である。

図11の上から第7行の特徴量 $F_{IP, r15}$ は、質問情報に含まれるキーワードkに照合する単語であって第1候補 $SRR_{IP, 1}$ に含まれる単語情報における単語に対応する単語信頼度のキーワードk間の最小値である。

図11の上から第8行の特徴量 $F_{IP, r16}$ は、質問情報に含まれるキーワードkに照合する単語であって第1候補 $SRR_{IP, 1}$ に含まれる単語情報における単語に対応する単語信頼度のキーワードk間の最大値である。

30

図11の上から第9行の特徴量 $F_{IP, h1}$ は、IPエキスパート部104-3が既出(即ち対話処理装置1が動作開始以降、予め設定された処理を行ったことがある)であるか否かを示す特徴量である。IPエキスパート部104-3が既出である場合、特徴量 $F_{IP, h1}$ は値1である。IPエキスパート部104-3が既出でない場合、特徴量 $F_{IP, h1}$ は値0である。IPエキスパート部104-3が既出であるか否かを判断するために、特徴量生成部1041は対話履歴記憶部1046-3から読み出した対話履歴情報に質問情報等の発話パターン情報及び応答情報等の出力パターン情報が含まれているか否かを判断する。

【0079】

図12は、本実施形態に係るQAエキスパート部104-2の特徴量情報の一例を示す表である。

40

図12の最左列から右側へ順に、関連があるエキスパートクラス(種類)、特徴量、内容を示す。図12の最左列においてエキスパートクラスがQAとは、QAエキスパート部104-2の特徴量生成部1041-2のみが生成する特徴量情報を示す。

【0080】

図12の上から第2行の特徴量 $F_{QA, r10}$ は、 $SRR_{QA, 1}$ に含まれる単語情報と照合するキーワード情報を含む質問情報が処理情報記憶部1042-2にあるか否かを示す特徴量である。このような質問情報がある場合、 $F_{QA, r10}$ の値が1であり、このような質問情報がない場合、 $F_{QA, r10}$ の値は0である。

50

図12の上から第3行の特徴量 $F_{QA, r11}$ は、質問情報に含まれるキーワードと照合する単語であって、 $SRR_{QA, j}$ に含まれる単語情報における単語数を、 $SRR_{QA, j}$ に含まれる単語数で除算した値を、全ての認識結果jにわたる総和をとり、さらに認識結果の個数で除算した値である。この質問情報とは、処理情報記憶部1042-2に記憶されている質問情報である。これは、特徴量 $F_{QA, r12} \sim F_{QA, r16}$ においても同様である。

【0081】

図12の上から第4行の特徴量 $F_{QA, r12}$ は、質問情報に含まれるキーワードkと照合する単語であって、 $SRR_{QA, a11}$ に含まれる単語情報における単語数の全認識候補内での個数を、認識結果の個数で除算した値であって、キーワードk間の最小値を示す。

10

図12の上から第5行の特徴量 $F_{QA, r13}$ は、質問情報に含まれるキーワードkと照合する単語であって、 $SRR_{QA, a11}$ に含まれる単語情報における単語数の全認識候補内での個数を、認識結果の個数で除算した値であって、キーワードk間の最大値を示す。

【0082】

図12の上から第6行の特徴量 $F_{QA, r14}$ は、質問情報に含まれるキーワードkに照合する単語であって第1候補 $SRR_{QA, 1}$ に含まれる単語情報における単語に対応する単語信頼度のキーワードk間の平均値である。

図12の上から第7行の特徴量 $F_{QA, r15}$ は、質問情報に含まれるキーワードkに照合する単語であって第1候補 $SRR_{QA, 1}$ に含まれる単語情報における単語に対応する単語信頼度のキーワードk間の最小値である。

20

図12の上から第8行の特徴量 $F_{QA, r16}$ は、 $SRR_{IP, 1}$ に含まれる単語情報における単語が相槌情報であるか否かを示す特徴量である。 $SRR_{IP, 1}$ に含まれる単語情報における単語が相槌情報である場合、特徴量 $F_{QA, r16}$ の値は1である。 $SRR_{IP, 1}$ に含まれる単語情報における単語が相槌情報ではない場合、特徴量 $F_{QA, r16}$ の値は0である。特徴量生成部1041-2は、 $SRR_{IP, 1}$ に含まれる単語情報における単語が相槌情報であるかどうかを判断するために、処理情報記憶部1042-2に記憶されている相槌情報と、 $SRR_{IP, 1}$ に含まれる単語情報を照合する。

【0083】

図12の上から第9行の特徴量 $F_{QA, h1}$ は、対話履歴記憶部1046-2から読み出した音声認識情報である $SRR_{QA, 1}$ が肯定発話情報を含むかどうかを示す特徴量である。 $SRR_{QA, 1}$ が肯定発話情報を含む場合、 $F_{QA, h1}$ の値は1であり、 $SRR_{QA, 1}$ が肯定発話情報を含まない場合、 $F_{QA, h1}$ の値は0である。特徴量生成部1041-2は、 $SRR_{QA, 1}$ が肯定発話情報を含むかどうかを判断するために、処理情報記憶部1042-2に記憶されている肯定発話情報と、 $SRR_{QA, 1}$ に含まれる単語情報を照合する。特徴量 $F_{QA, h1}$ は、処理継続判定部1044-2において処理を継続するか否かを判定するために用いられるが、適合性推定部1043-2において確率値Pを算出するためには用いられない。

30

【0084】

図12の上から第10行の特徴量 $F_{QA, h2}$ は、QAエキスパート部104-2に遷移した後のターン数を示す特徴量である。ターン数とは、利用者が行う発話及び対話処理装置1からの応答の総数である。そこで、特徴量生成部1041-2は、対話履歴記憶部1046-2から対話履歴を読み出し、動作理解部103-2から起動信号を入力された時刻から後の読み出した対話履歴に含まれる質問情報、応答情報、肯定発話情報及び否定発話情報の総数を算出する。

40

【0085】

図12の上から第11行の特徴量 $F_{QA, h3}$ は、QAエキスパート部104-2に遷移(予め設定された処理を開始)した後の否定発話の回数を示す特徴量である。特徴量生成部1041-2は、対話履歴記憶部1046-2から対話履歴を読み出し、動作理解部

50

103-2から起動信号を入力された時刻から後の読み出した対話履歴に含まれる否定発話情報の総数を算出する。

図12の上から第12行の特徴量 $F_{QA, h4}$ は、特徴量 $F_{QA, h3}$ を特徴量 $F_{QA, h2}$ で除算した特徴量である。

【0086】

上述の特徴量を、各エキスパート部104-1~104-3の適合性推定部1043-1~1043-3は、確率値の算出に、処理継続判定部1044-1~1044-3は、エキスパート部の処理を継続するか否かの判断に用いる。

具体的には、RUエキスパート部104-1の適合性推定部1043-1は、特徴量 $F_{RU, r1} - F_{RU, r13}$ 、 $F_{RU, h2} - F_{RU, h4}$ を、式(1)の特徴量 $F_1 - F_n$ に代入して確率値Pを算出する。処理継続判定部1044-1は、特徴量 $F_{RU, r1} - F_{RU, r13}$ 、 $F_{RU, h1} - F_{RU, h4}$ を、特徴量ベクトルFの要素値として処理を継続するか否か判断する。

10

【0087】

具体的には、IPエキスパート部104-3の適合性推定部1043-3は、特徴量 $F_{IP, r1} - F_{IP, r16}$ 、 $F_{IP, h1}$ を、式(1)の特徴量 $F_1 - F_n$ に代入して確率値P(適合)を算出する。処理継続判定部1044-3は、特徴量 $F_{IP, r1} - F_{IP, r16}$ 、 $F_{IP, h1}$ を、特徴量ベクトルFの要素値として処理を継続するか否か判断する。

具体的には、QAエキスパート部104-2の適合性推定部1043-2は、特徴量 $F_{QA, r1} - F_{QA, r16}$ 、 $F_{QA, h2} - F_{QA, h4}$ を、式(1)の特徴量 $F_1 - F_n$ に代入して確率値Pを算出する。処理継続判定部1044-2は、特徴量 $F_{QA, r1} - F_{QA, r16}$ 、 $F_{QA, h1} - F_{QA, h4}$ を、特徴量ベクトルFの要素値として処理を継続するか否か判断する。

20

【0088】

なお、上述の特徴量を全て確率値の算出又は、エキスパート部の処理を継続するか否かの判断に用いると、演算量が過大になるうえ、各特徴量に乗算する係数を事前学習する際に過学習(overfitting)が生じ、却って推定精度や判断精度が低下するおそれがある。そこで、上述の特徴量のうち確率値、又は処理継続・非継続の判断に寄与する一部の特徴量のみを用い、確率値、又は処理継続・非継続の判断に寄与しないものを用いなくともよい。次に、その一部の特征量情報の例について説明する。

30

【0089】

図13は、本実施形態に係るRUエキスパート部104-1が用いる特徴量情報の一例を示す表である。

図13の最上行は、最左列から右側へ順に、特徴量、適合性推定、継続・非継続判定を示す。図13の最左列の第2行以降の行は、RUエキスパート部104-1が用いる特徴量の候補として $F_{RU, r1}, \dots, F_{RU, r13}, F_{RU, h1}, \dots, F_{RU, h4}, F_a$ を示す。ここで、特徴量 F_a とは、(具体的には動作処理部1045が)予め設定した処理を行わないエキスパート部104が推定した適合性情報(確率値P)の最大値を示す。

40

図13において、-印は該当する情報がないことを示す。図13の左から第2列において、は適合性推定に用いる特徴量であることを示す。図13の左から第3列において、は継続・非継続判断に用いる特徴量であることを示す。

即ち、図13は、適合性推定部1043-1は、適合性情報(確率値P)を算出する際に特徴量 $F_{RU, r2}, \dots, F_{RU, r6}, F_{RU, r8}, F_{RU, r13}$ を用いることを示す。従って、特徴量生成部1041は、これらの特徴量情報を生成し、生成した特徴量情報を適合性推定部1043に出力する。

また、図13は、処理継続判定部1044-1は、処理の継続・非継続を判定する際に特徴量 $F_{RU, r1}, F_{RU, r2}, F_{RU, r4}, F_{RU, r5}, F_{RU, h1}, F_{RU, h2}, F_a$ を用いることを示す。従って、特徴量生成部1041-1は、特徴量 F_R

50

$U, r_1, F_{RU}, r_2, F_{RU}, r_4, F_{RU}, r_5, F_{RU}, h_1, F_{RU}, h_2$ を表す特徴量情報を生成し、生成した特徴量情報を処理継続判定部 1044-1 に出力する。

【0090】

図 14 は、本実施形態に係る IP エキスパート部 104-3 が用いる特徴量情報の一例を示す表である。

図 14 の最上行は、最左列から右側へ順に、特徴量、適合性推定、継続・非継続判定を示す。図 14 の最左列の第 2 行以降の行は、IP エキスパート部 104-3 が用いる特徴量の候補として $F_{IP}, r_1, \dots, F_{IP}, r_{16}, F_{IP}, h_1, F_a$ を示す。

図 14 において、- 印は該当する情報がないことを示す。図 14 の左から第 2 列において、 F_{IP}, r_1 は適合性推定に用いる特徴量であることを示す。図 14 の左から第 3 列において、 F_{IP}, r_1 は継続非継続判定に用いる特徴量であることを示す。

即ち、図 14 は、適合性推定部 1043-3 は、適合性情報（確率値 P）を算出する際に特徴量 $F_{IP}, r_2, F_{IP}, r_4, F_{IP}, r_6, \dots, F_{IP}, r_{10}, F_{IP}, r_{12}, \dots, F_{IP}, r_{14}$ を用いることを示す。従って、特徴量生成部 1041-1 は、これらの特徴量情報を生成し、生成した特徴量情報を適合性推定部 1043-3 に出力する。

また、図 14 は、処理継続判定部 1044-3 は、処理の継続・非継続を判定する際に特徴量 $F_{IP}, r_1, \dots, F_{IP}, r_8, F_a$ を用いることを示す。従って、特徴量生成部 1041-3 は、特徴量 $F_{IP}, r_1, \dots, F_{IP}, r_8$ を表す特徴量情報を生成し、生成した特徴量情報を処理継続判定部 1044-3 に出力する。

【0091】

図 15 は、本実施形態に係る QA エキスパート部 104-2 が用いる特徴量情報の一例を示す表である。

図 15 の最上行は、最左列から右側へ順に、特徴量、適合性推定、継続・非継続判定を示す。図 15 の最左列の第 2 行以降の行は、QA エキスパート部 104-2 が用いる特徴量の候補として $F_{QA}, r_1, \dots, F_{QA}, r_{16}, F_{QA}, h_1, \dots, F_{QA}, h_4, F_a$ を示す。

図 15 において、- 印は該当する情報がないことを示す。図 15 の左から第 2 列において、 F_{QA}, r_1 は適合性推定に用いる特徴量であることを示す。図 15 の左から第 3 列において、 F_{QA}, r_1 は継続・非継続判定に用いる特徴量であることを示す。

即ち、図 15 は、適合性推定部 1043-2 は、適合性情報（確率値 P）を算出する際に特徴量 $F_{QA}, r_2, \dots, F_{QA}, r_4, F_{QA}, r_6, \dots, F_{QA}, r_8, F_{QA}, r_{10}, F_{QA}, r_{13}, \dots, F_{QA}, r_{15}$ を用いることを示す。従って、特徴量生成部 1041-2 は、これらの特徴量情報を生成し、生成した特徴量情報を適合性推定部 1043-2 に出力する。

また、図 15 は、処理継続判定部 1044-2 は、処理の継続・非継続を判定する際に特徴量 $F_{QA}, r_1, \dots, F_{QA}, r_6, F_{QA}, h_1, F_a$ を用いることを示す。従って、特徴量生成部 1041-2 は、特徴量 $F_{QA}, r_1, \dots, F_{QA}, r_6, F_{QA}, h_1$ を表す特徴量情報を生成し、生成した特徴量情報を処理継続判定部 1044-2 に出力する。

【0092】

なお、本実施形態では、動作理解部 103 は、予め設定された処理を行わないエキスパート部 104 から入力され確率値 P のうち最大値を特徴量 F_a として選択し、選択した特徴量 F_a を予め設定された処理を行うエキスパート部 104 の処理継続判定部 1044 に出力する。処理継続判定部 1044 は、動作理解部 103 から特徴量 F_a から入力されるようにする。これにより、処理継続判定部 1044 は音声認識結果に対するエキスパート部の適合性を考慮して、当該エキスパート部 104 が実行している処理の継続・非継続の判断を行うことができる。

【0093】

10

20

30

40

50

なお、適合性推定や継続・非継続判定に用いる特徴量を選択するために、本実施形態について事前検証を行っておく。この検証では、特徴量の候補の中から、特徴量を1個除外して各エキスパート部が正しく選択されるかを適合性推定、処理の継続・非継続の判断の各々について調べる。除外対象の特徴量を変更して検証を繰り返す。そして、除外対象の特徴量毎について性能の指標値であるF値を求めた。次に、F値が最も高くなる除外対象の特徴量を特徴量の候補から削除して検証を繰り返す。但し、特徴量の候補の中から、どの特徴量を除外してもF値がこれ以上高くならなかったときに検証を終了する。検証が終了して残った特徴量の候補の一例が、図13～15に示す特徴量である。この残った特徴量の候補が、適合性推定又は継続・非継続判定に用いる特徴量として選択される。

【0094】

F値とは、適合率(precision)と再現率(recall)の調和平均、つまり適合率と再現率の積を、適合率と再現率の相加平均で除算した値である。適合率とは選択された情報の数の選択対象の情報の数に対する割合であり、本実施形態では、エキスパート部104毎の選択された回数の試行回数に対する割合である。再現率とは選択された情報の数の正解情報の数に対する割合であり、本実施形態では、エキスパート部104毎の選択された回数の正解として選択されるべき試行回数に対する割合である。一般に適合率が増加すると再現率が低下する関係にあるが、F値が高いほどエキスパート部104を選択する精度が高いことを表す。

【0095】

なお、図1に示す対話処理装置1は、IPエキスパート部104-3を1個備えるが、本実施形態では、これには限られない。本実施形態では対話処理装置1は、各々異なる質問情報、応答情報、提示情報のセットを記憶しているIPエキスパート部104-3を複数個備えてもよい。

【0096】

なお、図1に示す対話処理装置1では、動作理解部103が、現在処理を行っているエキスパート部から処理を継続するか否かを示す処理継続信号を入力され、現在処理を行っていないエキスパート部から適合度情報を入力される。各エキスパート部は音声認識情報に基づいて特徴量情報を生成する。現在予め設定された処理を行っているエキスパート部は、生成した特徴量情報に基づき処理継続信号を生成する。現在予め設定された処理を行っていないエキスパート部は、生成した特徴量情報に基づき適合度情報を生成する。

しかし、本実施形態ではこれには限られず、動作理解部103は、各エキスパート部104-1～104-3から音声認識情報に基づいて生成した特徴量情報を入力されるようにしてもよい。このとき、動作理解部103は、処理継続判定部1044と同様に現在予め設定された処理を行っているエキスパート部から入力された特徴量情報に基づき処理継続信号を生成し、適合性推定部1043と同様に現在予め設定された処理を行っていないエキスパート部から入力された特徴量情報に基づき適合性情報を生成する。この場合、各エキスパート部104-1～104-3は、適合性推定部1043と処理継続判定部1044を省略してもよい。

【0097】

次に、対話処理装置1が出力する発話と利用者の発話からなる対話の一例について説明する。

図16は、本実施形態に係る対話処理装置1による対話の一例を示す表である。

図16は、最左列から右側へ順に、発話主体、発話(動作)、選択エキスパートを示す。選択エキスパートとは、予め設定された処理を行うエキスパート部104として動作理解部103が選択したエキスパート部104である。

図16の第2行は、対話処理装置1が、利用者に対して「こんにちは、イタリア、ギリシャと他の地域の有名な世界遺産についての質問にお答えします。また、いくつかの世界遺産について詳しく説明できます。何でも聞いてみてください。」という音声を提示することを示す。これにより、対話処理装置1は利用者に対して問合せを促す。図16の第2行は、この選択エキスパートがQAエキスパート部104-1であることを示す。

10

20

30

40

50

図16の第3行は、対話処理装置1が上述の音声を受けて利用者が発話した「イタリアの世界遺産のリストを見せて下さい。」を示す音声を認識し、対話処理装置1はイタリアの世界遺産のリストを画像で表示することを示す。図16の第3行は、この選択エキスパートがQAエキスパート部104-1であることを示す。

【0098】

図16の第4行は、対話処理装置1が、利用者に対して「イタリアの世界遺産を地域別に紹介します。どの地域のリストを見たいかおっしゃって下さい。」という音声を提示することを示す。これにより、対話処理装置1は利用者に対して紹介する地域について音声で指示することを促す。図16の第4行は、この選択エキスパートがQAエキスパート部104-1であることを示す。

10

図16の第5行は、対話処理装置1が上述の音声を受けて利用者が発話した「他に何が聞けますか?」、を示す音声を認識し、対話処理装置1はイタリアの世界遺産のリストを画像で表示することを示す。図16の第5行は、この選択エキスパートがQAエキスパート部104-1であることを示す。

【0099】

図16の第6行は、対話処理装置1が利用者に対して「イタリア、ギリシャの世界遺産と、他の著名な世界遺産に関して質問にお答えします。」という音声を提示することを示し、世界遺産のリストを画像情報として提示する。

これにより、対話処理装置1は世界遺産に関する質問を利用者に対して促す。図16の第6行は、この選択エキスパートがQAエキスパート部104-1であることを示す。

20

【0100】

図16の第7行は、対話処理装置1が、利用者が発話した「グリニッジについて教えてください。」という音声を認識し、RUエキスパート部104-1が処理を開始したことを示す。

図16の第8行は、対話処理装置1が「海事都市グリニッジの説明ですね?」という音声を提示し、利用者に認識した情報の確認を促す。図16の第8行は、この処理に係るエキスパート部がRUエキスパート部104-1であることを示す。

図16の第9行は、対話処理装置1が、利用者が発話した「はい」という音声を認識し、認識した情報が肯定応答情報であることを認識する。図16の第9行は、この選択エキスパートがRUエキスパート部104-1であることを示す。

30

【0101】

図16の第10行は、対話処理装置1が「はい、海事都市グリニッジの案内をします。グリニッジ標準時の基準となる都市として由緒ある港町です。」という音声を提示し、提示する情報の概要を示す。図16の第10行は、この選択エキスパートがIPエキスパート部104-3であることを示す。

図16の第11行は、対話処理装置1が、利用者が発話した「グリニッジ標準時ってなんですか?」という音声を認識することを示す。図16の第11行は、この選択エキスパートが104-3であることを示す。

図16の第12行は、対話処理装置1が「グリニッジ天文台での平均太陽時です。」という音声を、利用者の発話に対する応答情報として提示することを示す。図16の第12行は、この選択エキスパートがIPエキスパート部104-3であることを示す。

40

【0102】

次に、本実施形態に係る対話処理装置1の検証結果として選択エキスパートの選択精度について説明する。なお、検証において、IPエキスパート部104-3を8個備える対話処理装置1を用いた。

図17は、本実施形態に係る対話処理装置1が選択したエキスパート部104の回数の一列を示す表である。

図17の各列は、対話処理装置1が選択したエキスパート部104(選択エキスパート)を示す。図17の各行は、正解として既知のエキスパート部104(正解エキスパート)を示す。選択エキスパート、正解エキスパートの種類は、それぞれRUエキスパート部

50

104-1を1種類(RU)、QAエキスパート部104-2を1種類(QA)、IPエキスパート部104-3を8種類(IP1~IP8)、計10種類である。IP1~IP8は、各々異なる処理情報(質問情報、応答情報、提示情報)を記憶し出力する。図17の第3~12行、第3~12列の各欄は、その行の正解エキスパートに対してその列の選択エキスパートと選択した回数を示す。図17の第13列は、全選択エキスパートにわたる正解エキスパート毎の回数の合計値(試行回数)を示す。図17の第13行は、全正解エキスパートにわたる選択エキスパート毎の選択された回数の合計値を示す。従って、図17の第3~12行の各対角要素は、エキスパート部毎の正しく選択された回数を示す。

【0103】

図17が示すように、RUについては、234回中162回が正解である。QAについては、1025回中941回が正解である。IP1については、12回中8回が正解である。IP2については、22回中18回が正解である。IP3については、72回中52回が正解である。IP4については、80回中61回が正解である。IP5については、29回中26回が正解である。IP6については、35回中24回が正解である。IP7については、57回中43回が正解である。IP8については、17回中12回が正解である。従って、全エキスパート部104にわたる正答率は85.09%である。

【0104】

比較として、従来技術を用いた場合のエキスパート部104の選択精度について説明する。選択対象となるエキスパート部104の個数及び種類は、上述と同様である。ここでは、エキスパート部104毎に異なる言語モデルを用いて音声認識してスコア値を算出する。また、直前のエキスパート部と同一のエキスパート部について予め定めた値を算出したスコア値に加算して、直前のエキスパート部を優先して選択する(参考文献: B. Lin, H. Wang, and L. Lee. A distributed architecture for cooperative spoken dialogue agents with coherent dialogue state and history. n Proc. ASRU-99, 1999)。

【0105】

図18は、従来技術によって選択したエキスパート部104の回数の一例を示す表である。

図18において、用いたエキスパート部の個数及び種類、試行回数、各行各列の配列は、図17が示す結果と同様である。

図18によれば、RUについては、234回中128回が正解である。QAについては、1025回中951回が正解である。IP1については、12回中4回が正解である。IP2については、22回中7回が正解である。IP3については、72回中21回が正解である。IP4については、80回中19回が正解である。IP5については、29回中12回が正解である。IP6については、35回中9回が正解である。IP7については、57回中15回が正解である。IP8については、17回中1回が正解である。従って、全エキスパート部104にわたる正答率は73.72%である。これより本実施形態では全エキスパート部104にわたる正答率が11.37%向上している。

【0106】

このように、本実施形態では、予め定めた処理を行う機能を有する複数のエキスパート部を備え、入力した音声を認識し、複数のエキスパート部のうち前記処理を行うエキスパート部が、前記認識した認識結果に基づいて前記処理を継続するか否かを判定し、前記処理を行うエキスパート部が前記処理を継続しないと判断した場合、前記複数のエキスパート部のうち前記処理を行わないエキスパート部の各々が、前記認識した認識結果に基づいて適合性を推定し、前記推定した適合性に基づいて前記処理を行うエキスパート部を選択する。これにより、処理を行っているエキスパート部の処理を継続するか否か考慮されるため、利用者の発話に適したエキスパート部を選択することができる。

【0107】

また、本実施形態では、さらに推定した適合性に基づいて前記処理を行うエキスパート部の処理を継続するか否かを判定する。これにより、本実施形態に係る対話処理装置は現在処理を行っていないエキスパート部が利用者の発話に適したエキスパート部である可能性を考慮して、利用者の発話に適したエキスパート部を選択することができる。

【0108】

また、本実施形態では、前記複数エキスパート部の各々は、利用者が行う質問内容を表す発話情報と、前記質問内容に対する出力情報を対応付けて記憶した処理情報記憶部を備え、前記処理は、前記音声認識部が認識した認識結果に対応する発話情報の一部が照合する発話情報に対応する出力情報を選択する。これにより、さらに利用者の発話に適した出力情報を出力するエキスパート部を選択することができる。

10

【0109】

また、本実施形態では、前記複数のエキスパート部において前記処理情報記憶部に記憶された前記発話情報及び出力情報が各々異なることを特徴とする。これにより、さらに利用者の発話と、発話に対する応答が異なるエキスパート部の中から、利用者の発話に適したエキスパート部を選択することができる。

【0110】

なお、上述した実施形態における対話処理装置1の一部、例えば、特徴量算出部1021、音素決定部1023、単語決定部1026、動作理解部103、特徴量生成部1041、適合性推定部1043、処理継続判定部1044、動作処理部1045、タスク計画部106、及び情報出力部107をコンピュータで実現するようにしても良い。その場合、この制御機能を実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することによって実現しても良い。

20

【0111】

また、対話処理装置1が備えるエキスパート部104のうち少なくとも1つは、対話処理装置1が備える他の機能部、例えば、特徴量算出部1021、音素決定部1023、単語決定部1026、動作理解部103、タスク計画部106、及び情報出力部107を実現する、別個のコンピュータシステムを備え、そのコンピュータシステムを用いて実現されるものであっても良い。その場合、対話処理装置1を実現するプログラムとは別個のプログラムを用いて実現する。

30

対話処理装置1が備える他の機能部、例えば、特徴量算出部1021、音素決定部1023、単語決定部1026、動作理解部103、タスク計画部106、及び情報出力部107を実現する、別個のコンピュータシステムを備え、そのコンピュータシステムを用いて実現されるものであってもよい。その場合、対話処理装置1を実現するプログラムとは別個のプログラムを用いて実現する。

【0112】

また、対話処理装置1が備えるエキスパート部104のうち少なくとも1つは、音声認識部を備え、その音声認識部が備える音響モデル部はエキスパート部104毎に異なる音響モデルを記憶しても良い。また、その音声認識部が備える言語モデル部はエキスパート部104毎に異なる言語モデルを記憶しても良い。その場合、対話処理装置1は、音声入力部101は動作理解部103に音響信号を出力する。動作理解部103は、音声入力部101から入力された音響信号を音声認識情報の代わりに、音声認識部を備えるエキスパート部104に出力する。そのエキスパート部104は、動作理解部103から入力された音響信号に基づいて音声認識情報を生成するようにする。対話処理装置1が備えるエキスパート部の全てが音声認識部を備える場合、音声認識部102を省略しても良い。

40

【0113】

なお、ここでいう「コンピュータシステム」とは、対話処理装置1に内蔵されたコンピュータシステムであって、OSや周辺機器等のハードウェアを含むものとする。また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、CD-ROM等の可搬媒体、コンピュータシステムに内蔵されるハードディスク

50

等の記憶装置のことをいう。さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムを送信する場合の通信線のように、短時間、動的にプログラムを保持するもの、その場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリのように、一定時間プログラムを保持しているものも含んでも良い。また上記プログラムは、前述した機能の一部を実現するためのものであっても良く、さらに前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるものであっても良い。

また、上述した実施形態における対話処理装置1の一部、または全部を、LSI (Large Scale Integration) 等の集積回路として実現しても良い。対話処理装置1の各機能ブロックは個別にプロセッサ化しても良いし、一部、または全部を集積してプロセッサ化しても良い。また、集積回路化の手法はLSIに限らず専用回路、または汎用プロセッサで実現しても良い。また、半導体技術の進歩によりLSIに代替する集積回路化の技術が出現した場合、当該技術による集積回路を用いても良い。

【0114】

以上、図面を参照してこの発明の一実施形態について詳しく説明してきたが、具体的な構成は上述のものに限られることはなく、この発明の要旨を逸脱しない範囲内において様々な設計変更等を行うことが可能である。

【符号の説明】

【0115】

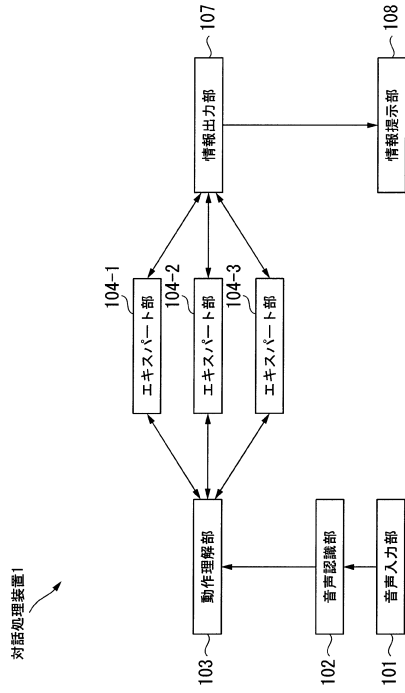
1 ... 対話処理装置、101 ... 音声入力部、102 ... 音声認識部、
 1021 ... 特徴量算出部、1022 ... 音響モデル記憶部、1023 ... 音素決定部、
 1024 ... 単語情報記憶部、1025 ... 言語モデル記憶部、1026 ... 単語決定部、
 103 ... 動作理解部、104 (104-1 ~ 104-3) エキスパート部、
 104-1 ... RUエキスパート部、104-2 ... QAエキスパート部、
 104-3 ... IPエキスパート部、
 1041 (1041-1 ~ 1041-3) ... 特徴量生成部、
 1042 (1042-1 ~ 1042-3) ... 処理情報記憶部、
 1043 (1043-1 ~ 1043-3) ... 適合性推定部、
 1044 (1044-1 ~ 1044-3) ... 処理継続判定部、
 1045 (1045-1 ~ 1045-3) ... 動作処理部、
 1046 (1046-1 ~ 1046-3) ... 対話履歴記憶部、
 106 ... タスク計画部、107 ... 情報出力部、108 ... 情報提示部

10

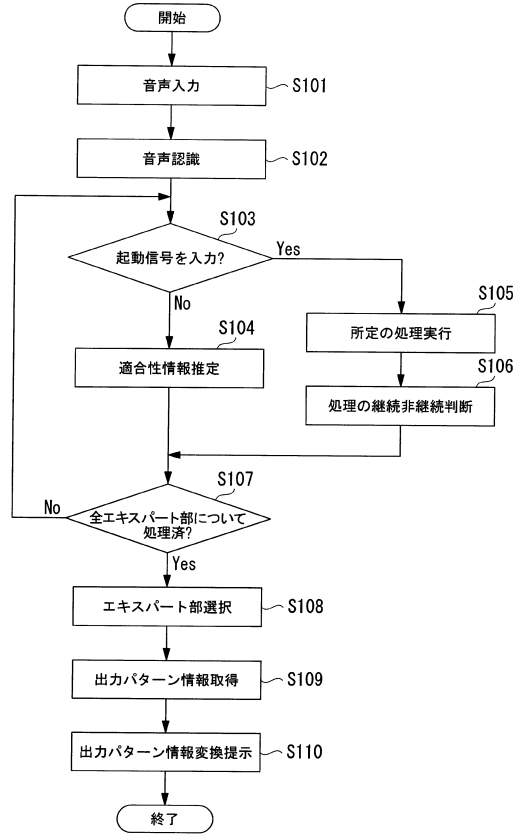
20

30

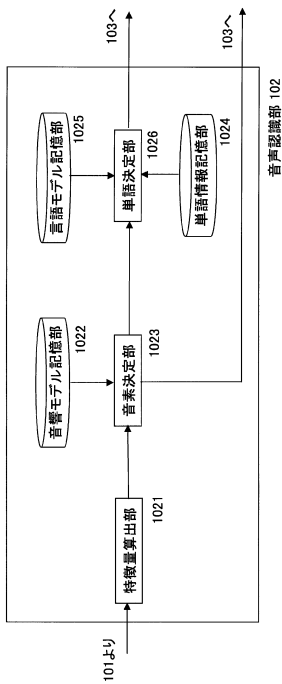
【図1】



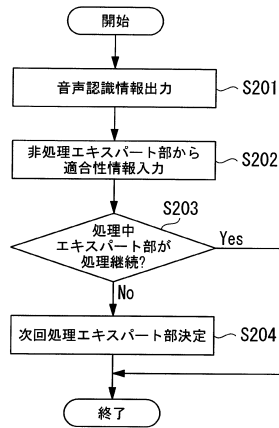
【図2】



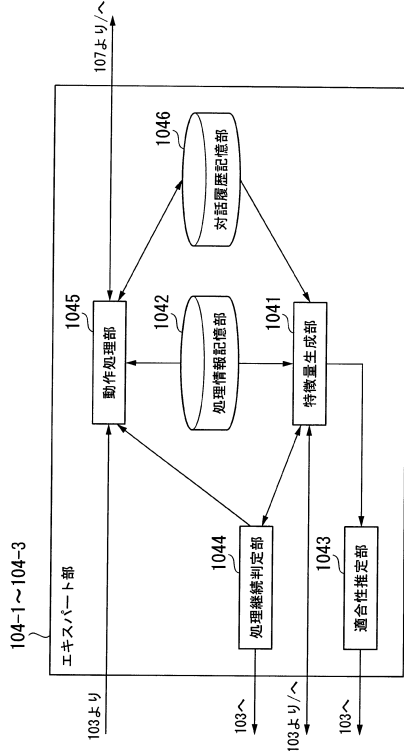
【図3】



【図4】



【 図 5 】



【 図 6 】

質問情報: グリニッジ ニ ツイテ オシエテ クダサイ
 応答情報: カイジツシ グリニッジ ノ セツメイ デスネ
 肯定発話情報: ハイ、ウン
 否定発話情報: イイエ、イヤ

【 図 7 】

質問情報: イタリア ノ セカイイサン ノ リスト ラ ミセテ クダサイ
 応答情報: イタリア ノ セカイイサン ラ タイキベツ ニ ショウカイ シマス
 ドノ タイキ ノ リスト ガ ミタイ カ オツシャツテ クダサイ
 提示情報: [イタリアの地域別の世界遺産のリストを示す画像情報]
 肯定発話情報: ハイ、ウン
 否定発話情報: イイエ、イヤ
 相続情報: ナルホド、ソウデスネ

【 図 8 】

応答情報: カイジツシ グリニッジ ノ アンナイ ラ シマス
 グリニッジヒョウジュンジ ノ キジュン トナル トントシテ ユイショ アル ミナトマチ デス
 提示情報: [グリニッジ所を示す画像情報]
 質問情報: グリニッジヒョウジュンジ ッテ ナンデスカ?
 応答情報: グリニッジテンモンダイ テノ ハイキンタイヨウジ ノ コトデス
 提示情報: [グリニッジ天文台の内容を示す画像情報]
 指示可能情報: トマル、ツギ、モドル、モウイチド

【 図 9 】

エキスパートクラス	特徴量	内容
共通 (I=RU, IP, OA)	$F_{i,r1}$	$SRR_{i,r1}$ が得られたかどうか(1/0)
	$F_{i,r2}$	$SRR_{i,r1}$ にフィラーが含まれるかどうか(1/0)
	$F_{i,r3}$	$SRR_{i,r1}$ の単語信頼度の最小値
	$F_{i,r4}$	$SRR_{i,r1}$ の単語信頼度の平均値
	$F_{i,r5}$	$SRR_{i,r1}$ の音響スコア / 時間
	$F_{i,r6}$	$SRR_{i,r1}$ の言語スコア
	$F_{i,r7}$	$SRR_{i,r1}$ の単語数
	$F_{i,r8}$	$SRR_{i,r1}$ の単語数
	$F_{i,r9}$	$(F_{i,r5} - SRR_{i,r5})$ の音響スコア / 時間

【 図 10 】

エキスパートクラス	特徴量	内容
RU	$F_{RU,r10}$	$SRR_{RU,r1}$ が肯定発話かどうか(1/0)
	$F_{RU,r11}$	$SRR_{RU,r1}$ が否定発話かどうか(1/0)
	$F_{RU,r12}$	言語モデルによる認識結果の候補数
	$F_{RU,r13}$	$SRR_{RU,r1}$ に世界遺産の名称が含まれているか(1/0)
	$F_{RU,h1}$	$SRR_{RU,r1}$ が肯定発話かどうか(1/0)
	$F_{RU,h2}$	RUエキスパート遷移後のターン数
	$F_{RU,h3}$	RUエキスパート遷移後の否定発話の回数
	$F_{RU,h4}$	$F_{RU,h3} / F_{RU,h2}$

【 図 1 1 】

エキスパートクラス	特徴量	内容
IP	$F_{IP,r10}$	$SRR_{IP,r1}$ と照合する質問情報があるかどうか(1/0)
	$F_{IP,r11}$	$\sum_j (SRR_{IP,rj})$ のキーワードの個数 / $(SRR_{IP,rj})$ の単語数 / (認識結果の個数)
	$F_{IP,r12}$	$(SRR_{IP,r1})$ のキーワードkの全認識候補内での個数 / (認識結果の個数)の最小値
	$F_{IP,r13}$	$(SRR_{IP,r1})$ のキーワードkの全認識候補内での個数 / (認識結果の個数)の最大値
	$F_{IP,r14}$	$SRR_{IP,r1}$ のキーワードkの第1候補における単語信頼度の平均値
	$F_{IP,r15}$	$SRR_{IP,r1}$ のキーワードkの第1候補における単語信頼度の最小値
	$F_{IP,r16}$	$SRR_{IP,r1}$ のキーワードkの第1候補における単語信頼度の最大値
	$F_{IP,h1}$	当該エキスパートが既出か否か(1/0)

【 図 1 2 】

エキスパートクラス	特徴量	内容
OA	$F_{OA,r10}$	$SRR_{OA,r1}$ とマッチする質問例がデータベースにあるかどうか(1/0)
	$F_{OA,r11}$	$\sum_j (SRR_{OA,rj})$ のキーワードが出現する個数 / $(SRR_{OA,rj})$ の単語数 / (認識結果の個数)
	$F_{OA,r12}$	$(SRR_{OA,r1})$ のキーワードkの全認識候補内での出現個数の最小値 / (認識結果の個数)
	$F_{OA,r13}$	$(SRR_{OA,r1})$ のキーワードkの全認識候補内での出現個数の最大値 / (認識結果の個数)
	$F_{OA,r14}$	$SRR_{OA,r1}$ のキーワードkの第1候補における単語信頼度の平均値
	$F_{OA,r15}$	$SRR_{OA,r1}$ のキーワードkの第1候補における単語信頼度の最小値
	$F_{OA,r16}$	$SRR_{OA,r1}$ が相違か否か(1/0)
	$F_{OA,h1}$	$SRR_{OA,r1}$ が肯定発話か否か(1/0)
	$F_{OA,h2}$	OAエキスパートに遷移後のターン数
	$F_{OA,h3}$	OAエキスパートに遷移後の否定発話の回数
	$F_{OA,h4}$	$F_{OA,h3} / F_{OA,h2}$

【 図 1 3 】

特徴量	適合性推定	継続・非継続判定
F _{RU, r1}	—	○
F _{RU, r2}	○	○
F _{RU, r3}	○	—
F _{RU, r4}	○	○
F _{RU, r5}	○	○
F _{RU, r6}	○	—
F _{RU, r7}	—	—
F _{RU, r8}	○	—
F _{RU, r9}	—	—
F _{RU, r10}	—	—
F _{RU, r11}	—	—
F _{RU, r12}	—	—
F _{RU, r13}	○	—
F _{RU, h1}	—	○
F _{RU, h2}	—	○
F _{RU, h3}	—	—
F _{RU, h4}	—	—
F _a	—	○

【 図 1 4 】

特徴量	適合性推定	継続・非継続判定
F _{IP, r1}	—	○
F _{IP, r2}	○	○
F _{IP, r3}	—	○
F _{IP, r4}	○	○
F _{IP, r5}	—	○
F _{IP, r6}	○	○
F _{IP, r7}	○	○
F _{IP, r8}	○	○
F _{IP, r9}	○	—
F _{IP, r10}	○	—
F _{IP, r11}	—	—
F _{IP, r12}	○	—
F _{IP, r13}	○	—
F _{IP, r14}	○	—
F _{IP, r15}	—	—
F _{IP, r16}	—	—
F _{IP, h1}	—	—
F _a	—	○

【 図 1 5 】

特徴量	適合性算出	継続・非継続判定
F _{OA, r1}	—	○
F _{OA, r2}	○	○
F _{OA, r3}	○	○
F _{OA, r4}	○	○
F _{OA, r5}	—	○
F _{OA, r6}	○	○
F _{OA, r7}	○	—
F _{OA, r8}	○	—
F _{OA, r9}	—	—
F _{OA, r10}	○	—
F _{OA, r11}	—	—
F _{OA, r12}	—	—
F _{OA, r13}	○	—
F _{OA, r14}	○	—
F _{OA, r15}	○	—
F _{OA, r16}	—	—
F _{OA, h1}	—	○
F _{OA, h2}	—	—
F _{OA, h3}	—	—
F _{OA, h4}	—	—
F _a	—	○

【 図 1 6 】

発話主体	発話(動作)	選択エキスパート
対話処理装置	こんにちは、イタリア、ギリシャと他の地域の有名な世界遺産についての質問にお答えします。また、いくつかの世界遺産については詳しく説明をします。何でも聞いてみてください。	OA
利用者	イタリアの世界遺産のリストを見せて下さい。	OA
対話処理装置	イタリアの世界遺産を地域別に紹介します。どの地域のリストをみたいかがおっしゃって下さい。(イタリアの世界遺産のリストを表示)	OA
利用者	他に何が聞けますか?	OA
対話処理装置	イタリア、ギリシャの世界遺産と、他の著名な世界遺産に関して質問にお答えします。(世界遺産のリストを表示)	OA
利用者	ギリニッジについて教えてください。	RU
対話処理装置	海軍都市ギリニッジの説明ですね?	RU
利用者	はい。	RU
対話処理装置	はい、海軍都市ギリニッジの案内をします。ギリニッジ標準時の基準となる都市として由緒ある港町です。	IP
利用者	ギリニッジ標準時ってなんですか?	IP
対話処理装置	ギリニッジ英文台での平均太陽時のことです。	IP

【 17 】

		選択エキスパート											合計
		RU	OA	IP1	IP2	IP3	IP4	IP5	IP6	IP7	IP8	合計	
RU	162	53	13	2	0	0	0	3	0	1	0	234	
OA	51	941	4	1	7	2	1	2	15	1	1025		
IP1	2	2	8	0	0	0	0	0	0	0	12		
IP2	0	4	0	18	0	0	0	0	0	0	22		
IP3	4	16	0	0	52	0	0	0	0	0	72		
IP4	5	14	0	0	0	61	0	0	0	0	80		
IP5	0	3	0	0	0	0	26	0	0	0	29		
IP6	1	10	0	0	0	0	0	24	0	0	35		
IP7	1	13	0	0	0	0	0	0	43	0	57		
IP8	1	4	0	0	0	0	0	0	0	12	17		
合計	227	1060	25	21	59	63	30	26	59	13	1583		
正解エキスパート													

【 18 】

		選択エキスパート											合計
		RU	OA	IP1	IP2	IP3	IP4	IP5	IP6	IP7	IP8	合計	
RU	128	80	0	0	0	7	6	5	0	4	4	234	
OA	64	951	0	0	5	3	0	0	0	2	0	1025	
IP1	3	5	4	0	0	0	0	0	0	0	12		
IP2	3	12	0	7	0	0	0	0	0	0	22		
IP3	16	35	0	0	21	0	0	0	0	0	72		
IP4	21	40	0	0	0	19	0	0	0	0	80		
IP5	5	12	0	0	0	0	12	0	0	0	29		
IP6	3	23	0	0	0	0	0	9	0	0	35		
IP7	7	34	0	0	0	0	0	0	15	1	57		
IP8	5	11	0	0	0	0	0	0	0	1	17		
合計	255	1203	4	7	33	28	17	9	21	6	1583		
正解エキスパート													

フロントページの続き

- (72)発明者 中野 幹生
埼玉県和光市本町 8 - 1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内
- (72)発明者 船越 孝太郎
埼玉県和光市本町 8 - 1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内
- (72)発明者 駒谷 和範
埼玉県和光市本町 8 - 1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内
- (72)発明者 奥乃 博
埼玉県和光市本町 8 - 1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内
- (72)発明者 佐藤 隼
埼玉県和光市本町 8 - 1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内

審査官 上田 雄

- (56)参考文献 国際公開第 2 0 0 7 / 0 1 3 5 2 1 (W O , A 1)
国際公開第 2 0 0 9 / 0 2 8 6 4 7 (W O , A 1)
特開 2 0 0 3 - 1 6 7 8 7 5 (J P , A)
特開 2 0 0 0 - 3 0 5 8 7 2 (J P , A)
神田直之、外 6 名、複数ドメイン音声対話システムにおける対話履歴を利用したドメイン選択の
高精度化、情報処理学会研究報告、日本、社団法人情報処理学会、2 0 0 6 年 2 月 4 日、V
ol. SLP-60, No. 11, pp. 55-60

(58)調査した分野(Int.Cl., DB名)

G 1 0 L 1 5 / 0 0 - 1 5 / 3 4
G 1 0 L 1 3 / 0 0