

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4950024号
(P4950024)

(45) 発行日 平成24年6月13日(2012.6.13)

(24) 登録日 平成24年3月16日(2012.3.16)

(51) Int.Cl. F I
G 1 0 L 15/22 (2006.01) G 1 0 L 15/22 3 0 0 U
 G 1 0 L 15/22 4 6 0 D

請求項の数 14 (全 15 頁)

(21) 出願番号	特願2007-505922 (P2007-505922)	(73) 特許権者	000005326
(86) (22) 出願日	平成18年2月27日 (2006. 2. 27)		本田技研工業株式会社
(86) 国際出願番号	PCT/JP2006/303613		東京都港区南青山二丁目1番1号
(87) 国際公開番号	W02006/093092	(74) 代理人	100077805
(87) 国際公開日	平成18年9月8日 (2006. 9. 8)		弁理士 佐藤 辰彦
審査請求日	平成21年1月13日 (2009. 1. 13)	(74) 代理人	100081477
(31) 優先権主張番号	60/657, 219		弁理士 堀 進
(32) 優先日	平成17年2月28日 (2005. 2. 28)	(74) 代理人	100099690
(33) 優先権主張国	米国 (US)		弁理士 鷺 健志
		(74) 代理人	100109232
			弁理士 本間 賢一
		(74) 代理人	100125210
			弁理士 加賀谷 剛

最終頁に続く

(54) 【発明の名称】 会話システムおよび会話ソフトウェア

(57) 【特許請求の範囲】

【請求項1】

ユーザの発話を認識する第1発話部と、発話を出力する第2発話部とを備えている会話システムであって、

前記第1発話部により認識された発話に含まれている1次入力言語単位に音響的に類似する言語単位を第1辞書DBから検索可能であることを要件として、当該1次入力言語単位に関連する複数の言語単位を第2辞書DBから検索して複数の1次出力言語単位として認識する第1処理部と、

前記第1処理部により認識された前記複数の1次出力言語単位のそれぞれの認識難易度を表す因子に基づき、当該複数の1次出力言語単位から1つを選定し、当該選定した1次出力言語単位に基づいてユーザの真意を問う1次質問を生成して前記第2発話部に出力させ、当該1次質問に対する前記ユーザの回答として前記第1発話部により認識された1次回答に基づき、当該ユーザの真意と前記1次入力言語単位との整合および不整合を判別する第2処理部とを備えていることを特徴とする会話システム。

【請求項2】

請求項1記載の会話システムにおいて、

前記第2処理部が、前記第1処理部により認識された前記複数の1次出力言語単位のそれぞれの観念的な認識難易度もしくは所定範囲における出現頻度を表す第1因子、および音響的な認識難易度もしくは所定数の他の言語単位との音響距離の最低平均値を表す第2因子のうち一方または両方に基づき、前記複数の1次出力言語単位から1つを選定するこ

10

20

とを特徴とする会話システム。

【請求項 3】

請求項 1 記載の会話システムにおいて、

前記第 2 処理部が前記 1 次入力言語単位と、前記第 1 処理部により認識された前記複数の 1 次出力言語単位のそれぞれとの音響距離に基づき、当該複数の 1 次出力言語単位から 1 つを選定することを特徴とする会話システム。

【請求項 4】

請求項 1 記載の会話システムにおいて、

前記第 1 処理部が、

前記 1 次入力言語単位とこれに音響的に類似する言語単位との相違部分を含む第 1 種言語単位と、

当該相違部分の本来の読み方と異なる読み方を表す第 2 種言語単位と、

他の言語体系において当該相違部分に対応する言語単位の読み方を表す第 3 種言語単位と、

当該相違部分に含まれる 1 つの音素を表す第 4 種言語単位と、

前記 1 次入力言語単位と概念的に類似する第 5 種言語単位とのうち、一部または全部を前記 1 次出力言語単位として認識することを特徴とする会話システム。

【請求項 5】

請求項 4 記載の会話システムにおいて、

前記第 1 処理部が、第 k 種言語単位群 ($k = 1 \sim 5$) から複数の言語単位を前記複数の 1 次出力言語単位として認識することを特徴とする会話システム。

【請求項 6】

請求項 1 記載の会話システムにおいて、

前記第 2 処理部により前記ユーザの真意と i 次入力言語単位 ($i = 1, 2, \dots$) とが整合していないと判定された場合、

前記第 1 処理部が、前記 i 次入力言語単位に音響的に類似する言語単位を前記第 1 辞書 DB から検索して $i + 1$ 次入力言語単位として認識し、前記 $i + 1$ 次入力言語単位に関連する言語単位を前記第 2 辞書 DB から検索して $i + 1$ 次出力言語単位として認識し、かつ

前記第 2 処理部が、前記第 1 処理部により認識された前記 $i + 1$ 次出力言語単位に基づき、前記ユーザの真意を問う $i + 1$ 次質問を生成して前記第 2 発話部へ出力させ、当該 $i + 1$ 次質問に対する前記ユーザの回答として前記第 1 発話部により認識された $i + 1$ 次回答に基づき、当該ユーザの真意と前記 $i + 1$ 次入力言語単位との整合および不整合を判別する特徴とする会話システム。

【請求項 7】

請求項 6 記載の会話システムにおいて、

前記第 1 処理部が複数の $i + 1$ 次出力言語単位を認識し、

前記第 2 処理部が、前記第 1 処理部により認識された前記複数の $i + 1$ 次出力言語単位のそれぞれの認識難易度を表す因子に基づき、当該複数の $i + 1$ 次出力言語単位から 1 つを選定し、当該選定した $i + 1$ 次出力言語単位に基づいて前記 $i + 1$ 次質問を生成することを特徴とする会話システム。

【請求項 8】

請求項 7 記載の会話システムにおいて、

前記第 2 処理部が、前記第 1 処理部により認識された前記複数の $i + 1$ 次出力言語単位のそれぞれの観念的な認識難易度もしくは所定範囲における出現頻度を表す第 1 因子、および音響的な認識難易度もしくは所定数の他の言語単位との音響距離の最低平均値を表す第 2 因子のうち一方または両方に基づき、当該複数の $i + 1$ 次出力言語単位から 1 つを選定することを特徴とする会話システム。

【請求項 9】

請求項 6 記載の会話システムにおいて、

10

20

30

40

50

前記第2処理部が前記*i*次入力言語単位と前記第1処理部により認識された前記複数の*i*+1次出力言語単位のそれぞれとの音響距離、および前記*i*+1次入力言語単位と前記複数の*i*+1次出力言語単位のそれぞれとの音響距離のうち一方または両方にに基づき、当該複数の*i*+1次出力言語単位から1つを選定することを特徴とする会話システム。

【請求項10】

請求項7記載の会話システムにおいて、

前記第1処理部が、

前記*i*+1次入力言語単位およびこれに音響的に類似する言語単位の相違部分を含む第1種言語単位と、

当該相違部分の本来の読み方と異なる読み方を表す第2種言語単位と、

他の言語体系において当該相違部分に対応する言語単位の読み方を表す第3種言語単位と、

当該相違部分に含まれる1つの音素を表す第4種言語単位と、

前記*i*+1次入力言語単位と概念的に類似する第5種言語単位とのうち、一部または全部を2次出力言語単位として認識することを特徴とする会話システム。

【請求項11】

請求項8記載の会話システムにおいて、

前記第1処理部が、第*k*種言語単位群($k = 1 \sim 5$)から複数の言語単位を前記複数の*i*+1次出力言語単位として認識することを特徴とする会話システム。

【請求項12】

請求項6記載の会話システムにおいて、

前記第2処理部により前記ユーザの真意と*j*次入力言語単位($j \geq 2$)とが整合していないと判定された場合、

前記第2処理部が、前記ユーザの再度の発話を促す質問を生成して、これを前記第2発話部へ出力させることを特徴とする会話システム。

【請求項13】

ユーザの発話を認識する第1発話機能と、発話を出力する第2発話機能とを有するコンピュータの記憶機能に格納される会話ソフトウェアであって、

前記第1発話機能により認識された発話に含まれている1次入力言語単位に音響的に類似する言語単位を第1辞書DBから検索可能であることを要件として、当該1次入力言語単位に関連する複数の言語単位を第2辞書DBから検索して複数の1次出力言語単位として認識する第1処理機能と、

前記第1処理機能により認識された前記複数の1次出力言語単位のそれぞれの認識難易度を表す因子に基づき、当該複数の1次出力言語単位から1つを選定し、当該選定した1次出力言語単位に基づいてユーザの真意を問う1次質問を生成して前記第2発話機能によって出力し、当該1次質問に対する前記ユーザの回答として前記第1発話機能により認識された1次回答に基づき、当該ユーザの真意と前記1次入力言語単位との整合および不整合を判別する第2処理機能とを前記コンピュータに付与することを特徴とする会話ソフトウェア。

【請求項14】

請求項13記載の会話ソフトウェアにおいて、

前記第2処理機能によりユーザの真意と*i*次入力言語単位($i = 1, 2, \dots$)とが整合していないと判定された場合、

前記第1処理機能として、前記*i*次入力言語単位に音響的に類似する言語単位を前記第1辞書DBから検索して*i*+1次入力言語単位として認識し、前記*i*+1次入力言語単位に関連する言語単位を前記第2辞書DBから検索して*i*+1次出力言語単位として認識する機能と、

前記第2処理機能として、前記第1処理機能により認識された*i*+1次出力言語単位に基づき、ユーザの真意を問う*i*+1次質問を生成して前記第2発話機能によって出力し、当該*i*+1次質問に対する前記ユーザの回答として前記第1発話機能により認識された*i*

10

20

30

40

50

+ 1 次回答に基づき、当該ユーザの真意と前記 i + 1 次入力言語単位との整合および不整合を判別する機能とを前記コンピュータに対して付与することを特徴とする会話ソフトウェア。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ユーザの発話を認識し、かつ、ユーザに対して発話を出力するシステム、およびユーザとの会話に必要な機能をコンピュータに付与するソフトウェアに関する。

【背景技術】

【0002】

ユーザおよびシステムの会話に際して、周辺雑音等の諸原因によって、システムによるユーザの発話認識に誤り（聞き誤り）が生じる可能性がある。このため、システムにおいてユーザの発話内容を確認するための発話を出力する技術が提案されている（たとえば、特開2002-351492号公報参照）。当該システムによれば、単語の「属性」「属性値」および「属性値間距離」が定義され、共通の属性を有しながら属性値が異なり、かつ、当該属性値の偏差（属性値間距離）が閾値以上となるような複数の単語が同一のユーザとの会話中に認識された場合、当該単語を確認するための発話が出力される。

【0003】

しかし、前記システムによれば、聞き誤りが生じた場合、属性値間距離が不適当に評価される場合がある。このため、ユーザが「A」と発話したにもかかわらず、システムがユーザの発話は「A」と音響的に近い「B」とであると認識しているという齟齬が解消されないまま、会話が進行してしまう可能性があった。

【0004】

そこで、本発明は、ユーザの発話と、認識された発話との齟齬をより適当に解消しながら、当該ユーザと会話しうるシステム、および当該会話機能をコンピュータに対して付与するソフトウェアを提供することを解決課題とする。

【発明の開示】

【0005】

前記課題を解決するための本発明の会話システムは、ユーザの発話を認識する第1発話部と、発話を出力する第2発話部とを備えている会話システムであって、前記第1発話部により認識された発話に含まれている1次入力言語単位に音響的に類似する言語単位を第1辞書DBから検索可能であることを要件として、当該1次入力言語単位に関連する複数の言語単位を第2辞書DBから検索して複数の1次出力言語単位として認識する第1処理部と、前記第1処理部により認識された前記複数の1次出力言語単位のそれぞれの認識難易度を表す因子に基づき、当該複数の1次出力言語単位から1つを選定し、当該選定した1次出力言語単位に基づいてユーザの真意を問う1次質問を生成して前記第2発話部に出力させ、当該1次質問に対する前記ユーザの回答として前記第1発話部により認識された1次回答に基づき、当該ユーザの真意と前記1次入力言語単位との整合および不整合を判別する第2処理部とを備えていることを特徴とする。

【0006】

第1発話部により認識された発話に含まれている「1次入力言語単位」に音響的に類似する言語単位が第1辞書DBから検索されうる場合、1次入力言語単位ではなく他の言語単位がユーザの発話に含まれていた可能性がある。すなわち、この場合、第1発話部が1次入力言語単位について聴き違えをした可能性が多少なりともある。この点に鑑みて1次入力言語単位に関連する複数の「1次出力言語単位」が第2辞書DBから検索される。

【0007】

また、複数の1次出力言語単位の中から、認識難易度を表す因子に基づき1次出力言語単位が選定され、当該選定された1次出力言語単位に応じた「1次質問」が生成されて出力される。このため、当該選定された1次出力言語単位のユーザにとっての認識を容易にすることができる。これにより、ユーザの真意と1次入力言語単位との整合および不整合

10

20

30

40

50

を判別する観点から適当な1次質問が生成される。そして、1次質問に対するユーザの発話として認識された「1次回答」に基づき、当該ユーザの真意と1次入力言語単位との整合および不整合が判別される。これにより、ユーザの発話（真意）と、システムにおいて認識された発話との齟齬をより確実に抑制しながら、当該ユーザとシステムとの会話が可能となる。

【0008】

なお「言語単位」とは、文字、単語、複数の単語より構成される文章、短い文章より構成される長い文章等を意味する。

【0011】

さらに、本発明の会話システムは、前記第2処理部が、前記第1処理部により認識された前記複数の1次出力言語単位のそれぞれの観念的な認識難易度もしくは所定範囲における出現頻度を表す第1因子、および音響的な認識難易度もしくは所定数の他の言語単位との音響距離の最低平均値を表す第2因子のうち一方または両方に基づき、前記複数の1次出力言語単位から1つを選定することを特徴とする。

10

【0012】

本発明の会話システムによれば、当該選定された1次出力言語単位のユーザにとっての観念的または音響的な認識を容易にすることができる。これにより、ユーザの真意と1次入力言語単位との整合および不整合の別を確認する観点から適当な1次質問が生成される。

【0013】

また、本発明の会話システムは、前記第2処理部が前記1次入力言語単位と、前記第1処理部により認識された前記複数の1次出力言語単位のそれぞれとの音響距離に基づき、当該複数の1次出力言語単位から1つを選定することを特徴とする。

20

【0014】

本発明の会話システムによれば、複数の1次出力言語単位の中から、1次入力言語単位との音響距離に基づき1次出力言語単位が選定されるので、当該選定された1次出力言語単位の1次入力言語単位とのユーザにとっての聴覚的な識別を容易にすることができる。

【0015】

さらに、本発明の会話システムは、前記第1処理部が、前記1次入力言語単位とこれに音響的に類似する言語単位との相違部分を含む第1種言語単位と、当該相違部分の本来の読み方と異なる読み方を表す第2種言語単位と、他の言語体系において当該相違部分に対応する言語単位の読み方を表す第3種言語単位と、当該相違部分に含まれる1つの音素を表す第4種言語単位と、前記1次入力言語単位と概念的に類似する第5種言語単位とのうち、一部または全部を前記1次出力言語単位として認識することを特徴とする。

30

【0016】

また、本発明の会話システムは、前記第1処理部が、第k種言語単位群（ $k = 1 \sim 5$ ）から複数の言語単位を前記複数の1次出力言語単位として認識することを特徴とする。

【0017】

本発明の会話システムによれば、1次質問の生成基礎である1次出力言語単位の選択肢の幅が広げられるので、ユーザの真意および1次入力言語単位の整合および不整合を判別する観点から最適な1次質問が生成されうる。

40

【0018】

さらに、本発明の会話システムは、前記第2処理部により前記ユーザの真意とi次入力言語単位（ $i = 1, 2, \dots$ ）とが整合していないと判定された場合、前記第1処理部が、前記i次入力言語単位に音響的に類似する言語単位を前記第1辞書DBから検索してi+1次入力言語単位として認識し、前記i+1次入力言語単位に関連する言語単位を前記第2辞書DBから検索してi+1次出力言語単位として認識し、かつ、前記第2処理部が、前記第1処理部により認識された前記i+1次出力言語単位に基づき、前記ユーザの真意を問うi+1次質問を生成して前記第2発話部に出力させ、当該i+1次質問に対する前記ユーザの回答として前記第1発話部により認識されたi+1次回答に基づき、当該ユーザ

50

の真意と前記 $i + 1$ 次入力言語単位との整合および不整合を判別する特徴とする。

【0019】

本発明の会話システムによれば、第1発話部により認識された発話に含まれている i 次入力言語単位に音響的に類似する言語単位としての「 $i + 1$ 次入力言語単位」がユーザの発話に含まれていた可能性があることに鑑みて、 $i + 1$ 次入力言語単位に関連する「 $i + 1$ 次出力言語単位」が第2辞書DBから検索される。また、 $i + 1$ 次出力言語単位に基づいて「 $i + 1$ 次質問」が生成されて出力される。そして、 $i + 1$ 次質問に対するユーザの発話として認識された「 $i + 1$ 次回答」に基づき、当該ユーザの真意と $i + 1$ 次入力言語単位との整合および不整合が判別される。このように、複数回にわたってユーザの真意を問うための質問が当該ユーザに向けて投げかけられる。これにより、ユーザの発話（真意）とシステムにおいて認識された発話との齟齬をさらに確実に抑制しながら、当該ユーザとシステムとの会話が可能となる。

10

【0020】

また、本発明の会話システムは、前記第1処理部が複数の $i + 1$ 次出力言語単位を認識し、前記第2処理部が、前記第1処理部により認識された前記複数の $i + 1$ 次出力言語単位のそれぞれの認識難易度を表す因子に基づき、当該複数の $i + 1$ 次出力言語単位から1つを選定し、当該選定した $i + 1$ 次出力言語単位に基づいて前記 $i + 1$ 次質問を生成することを特徴とする。

【0021】

本発明の会話システムによれば、複数の $i + 1$ 次出力言語単位の中から、認識難易度を表す因子に基づき $i + 1$ 次出力言語単位が選定されるので、当該選定された $i + 1$ 次出力言語単位のユーザにとっての認識を容易にすることができる。これにより、ユーザの真意と $i + 1$ 次入力言語単位との整合および不整合を判別する観点から適当な $i + 1$ 次質問が生成される。

20

【0022】

さらに、本発明の会話システムは、前記第2処理部が、前記第1処理部により認識された前記複数の $i + 1$ 次出力言語単位のそれぞれの観念的な認識難易度もしくは所定範囲における出現頻度を表す第1因子、および音響的な認識難易度もしくは所定数の他の言語単位との音響距離の最低平均値を表す第2因子のうち一方または両方に基づき、当該複数の $i + 1$ 次出力言語単位から1つを選定することを特徴とする。

30

【0023】

本発明の会話システムによれば、当該選定された $i + 1$ 次出力言語単位のユーザにとっての観念的または音響的な認識を容易にすることができる。これにより、ユーザの真意と $i + 1$ 次入力言語単位との整合および不整合を判別する観点から適当な $i + 1$ 次質問が生成される。

【0024】

また、本発明の会話システムは、前記第2処理部が前記 i 次入力言語単位と前記第1処理部により認識された前記複数の $i + 1$ 次出力言語単位のそれぞれとの音響距離、および前記 $i + 1$ 次入力言語単位と前記複数の $i + 1$ 次出力言語単位のそれぞれとの音響距離のうち一方または両方に基づき、当該複数の $i + 1$ 次出力言語単位から1つを選定することを特徴とする。

40

【0025】

本発明の会話システムによれば、複数の $i + 1$ 次出力言語単位の中から、 i 次入力言語単位との音響距離に基づき $i + 1$ 次出力言語単位が選定されるので、当該選定された $i + 1$ 次出力言語単位の i 次入力言語単位との音響的な識別を容易にすることができる。さらに、複数の $i + 1$ 次出力言語単位の中から、 $i + 1$ 次入力言語単位との音響距離に基づき $i + 1$ 次出力言語単位が選定されるので、当該選定された $i + 1$ 次出力言語単位の $i + 1$ 次入力言語単位との音響的な識別を容易にすることができる。

【0026】

さらに、本発明の会話システムは、前記第1処理部が、前記 $i + 1$ 次入力言語単位およ

50

びこれに音響的に類似する言語単位の相違部分を含む第1種言語単位と、当該相違部分の本来の読み方と異なる読み方を表す第2種言語単位と、他の言語体系において当該相違部分に対応する言語単位の読み方を表す第3種言語単位と、当該相違部分に含まれる1つの音素を表す第4種言語単位と、前記 $i + 1$ 次入力言語単位と概念的に類似する第5種言語単位とのうち、一部または全部を2次出力言語単位として認識することを特徴とする。

【0027】

また、本発明の会話システムは、前記第1処理部が、第 k 種言語単位群 ($k = 1 \sim 5$) から複数の言語単位を前記複数の $i + 1$ 次出力言語単位として認識することを特徴とする。

【0028】

本発明の会話システムによれば、 $i + 1$ 次質問の生成基礎としての $i + 1$ 次出力言語単位の選択肢の幅が広げられるので、ユーザの先の発話と $i + 1$ 次入力言語単位との整合および不整合を判別する観点から最適な $i + 1$ 次質問が生成されうる。

【0029】

さらに、本発明の会話システムは、前記第2処理部により前記ユーザの真意と j 次入力言語単位 ($j = 2$) とが整合していないと判定された場合、前記第2処理部が、前記ユーザの再度の発話を促す質問を生成して、これを前記第2発話部に出力させることを特徴とする。

【0030】

本発明の会話システムによれば、逐次出力される質問によってはユーザの真意が確認できないような場合、あらためてその真意を確認することができる。

【0031】

前記課題を解決するための本発明の会話ソフトウェアは、ユーザの発話を認識する第1発話機能と、発話を出力する第2発話機能とを有するコンピュータの記憶機能に格納される会話ソフトウェアであって、前記第1発話機能により認識された発話に含まれている1次入力言語単位に音響的に類似する言語単位を第1辞書DBから検索可能であることを要件として、当該1次入力言語単位に関連する複数の言語単位を第2辞書DBから検索して複数の1次出力言語単位として認識する第1処理機能と、前記第1処理機能により認識された前記複数の1次出力言語単位のそれぞれの認識難易度を表す因子に基づき、当該複数の1次出力言語単位から1つを選定し、当該選定した1次出力言語単位に基づいてユーザの真意を問う1次質問を生成して前記第2発話機能によって出力し、当該1次質問に対する前記ユーザの回答として前記第1発話機能により認識された1次回答に基づき、当該ユーザの真意と前記1次入力言語単位との整合および不整合を判別する第2処理機能とを前記コンピュータに付与することを特徴とする。

【0032】

本発明の会話ソフトウェアによれば、ユーザの発話(またはその真意)と、システムにおいて認識された発話との齟齬をより確実に抑制しながら、当該ユーザと会話する機能が当該コンピュータに付与される。

【0033】

また、本発明の会話ソフトウェアは、前記第2処理機能によりユーザの真意と i 次入力言語単位 ($i = 1, 2, \dots$) とが整合していないと判定された場合、前記第1処理機能として、前記 i 次入力言語単位に音響的に類似する言語単位を前記第1辞書DBから検索して $i + 1$ 次入力言語単位として認識し、前記 $i + 1$ 次入力言語単位に関連する言語単位を前記第2辞書DBから検索して $i + 1$ 次出力言語単位として認識する機能と、前記第2処理機能として、前記第1処理機能により認識された $i + 1$ 次出力言語単位に基づき、ユーザの真意を問う $i + 1$ 次質問を生成して前記第2発話機能によって出力し、当該 $i + 1$ 次質問に対する前記ユーザの回答として前記第1発話機能により認識された $i + 1$ 次回答に基づき、当該ユーザの真意と前記 $i + 1$ 次入力言語単位との整合および不整合を判別する機能とを前記コンピュータに対して付与することを特徴とする。

【0034】

10

20

30

40

50

本発明の会話ソフトウェアによれば、ユーザの真意を問う質問が複数回にわたって生成する機能が前記コンピュータに対して付与される。したがって、当該ユーザの真意をより正確に把握し、ユーザの発話とシステムにおいて認識された発話との齟齬をより確実に抑制しながら、当該ユーザと会話する機能が当該コンピュータに付与される。

【発明を実施するための最良の形態】

【0035】

本発明の会話システムおよび会話ソフトウェアの実施形態について図面を用いて説明する。

【0036】

図1は本発明の会話システムの構成例示図であり、図2は本発明の会話システムおよび会話ソフトウェアの機能例示図である。

【0037】

図1に示されている会話システム（以下「システム」という。）100は、自動車に搭載されたナビゲーションシステム（ナビシステム）10に組み込まれた、ハードウェアとしてのコンピュータと、当該コンピュータのメモリに格納された本発明の「会話ソフトウェア」とにより構成されている。

【0038】

会話システム100は、第1発話部101と、第2発話部102と、第1処理部111と、第2処理部112と、第1辞書DB121と、第2辞書DB122とを備えている。

【0039】

第1発話部101は、マイクロフォン（図示略）等により構成され、入力音声に基づいて隠れマルコフモデル法等、公知の手法にしたがってユーザの発話を認識する。

【0040】

第2発話部102は、スピーカ（図示略）等により構成され、音声（または発話）を出力する。

【0041】

第1処理部111は、第1発話部101により認識された発話に含まれている1次入力言語単位に音響的に類似する言語単位を第1辞書DB121から検索可能であることを要件として1次入力言語単位に関連する複数種類の言語単位を第2辞書DB122から検索して1次出力言語単位として認識する。さらに、第1処理部111は、後述するように必要に応じてより高次の出力言語単位を認識する。

【0042】

第2処理部112は、1次入力言語単位に基づき、第1処理部111により認識された複数種類の1次出力言語単位の中から1つを選定する。また、第2処理部112は、選定した1次出力言語単位に基づき、ユーザの真意を問う1次質問を生成して第2発話部102に出力させる。さらに、第2処理部112は、当該1次質問に対するユーザの回答として第1発話部101により認識された1次回答に基づき、該ユーザの真意と1次入力言語単位との整合および不整合を判別する。また、第2処理部112は、後述するように必要に応じてより高次の質問を生成し、かつ、高次の回答に基づいてユーザの真意を確認する。

【0043】

第1辞書DB121は、第1処理部111により $i + 1$ 次入力言語単位（ $i = 1, 2, \dots$ ）として認識されうる複数の言語単位を記憶保持している。

【0044】

第2辞書DB122は、第1処理部111により i 次出力言語単位として認識されうる複数の言語単位を記憶保持している。

【0045】

前記構成の会話システム100の機能について、図2を用いて説明する。

【0046】

まず、ユーザが目的地設定のためにナビシステム10を操作したことに応じて、第2発

10

20

30

40

50

話部 102 が「目的地はどこですか」という初期発話を出力する(図 2 / S 1)。初期発話に応じてユーザが目的地を表す単語を口にすると、第 1 発話部 101 がこの発話を認識する(図 2 / S 2)。このとき、入力言語単位、出力言語単位、質問および回答の次数を表す指数 i が「1」に設定される(図 2 / S 3)。

【0047】

また、第 1 処理部 111 が、第 1 発話部 101 により認識された発話を言語単位列に変換し、この言語単位列から第 1 辞書 DB 121 において「地域名称」や「建築物名称」等に分類されている言語単位を抽出して i 次入力言語単位 x_i として認識する(図 2 / S 4)。言語単位列から抽出される言語単位の分類は、ナビ装置 1 がユーザにその目的地までの案内ルートを提示するといったドメインに基づいている。

10

【0048】

さらに、第 1 処理部 111 が、 i 次入力言語単位 x_i と音響的に類似する言語単位を第 1 辞書 DB 121 から検索可能であるか否か、すなわち、当該音響類似単語が第 1 辞書 DB 121 に記憶されているか否かを判定する(図 2 / S 5)。ここで、言語単位 x_i および x_j が音響的に類似するとは、次式(1)によって定義される音響距離 $pd(x_i, x_j)$ が閾値未満であることを意味する。

【0049】

$$pd(x_i, x_j) = ed(x_i, x_j) / \ln[\min(|x_i|, |x_j|) + 1] \quad (1)$$

式(1)において $|x|$ は言語単位 x に含まれている音素(または音韻)の数である。音素とは、1つの言語で用いられる音を弁別機能の見地から規定された最小単位を意味する。

20

【0050】

また、 $ed(x_i, x_j)$ は言語単位 x_i および x_j の編集距離であり、言語単位 x_i の音素列を言語単位 x_j の音素列に変換するための音素の挿入、削除、置換に際して、モーラ(日本語の発音の最小単位を意味する。)または音素の数が増える場合のコストを「1」、モーラや音素の数が増えない場合のコストを「2」として、DP マッチングにより求められる。

【0051】

第 1 処理部 111 は i 次入力言語単位 x_i に音響的に類似する言語単位が第 1 辞書 DB 121 に登録されていると判定した場合(図 2 / S 5 YES)、 i 次入力言語単位 x_i に関連する複数種類の i 次出力言語単位 $y_{ki} = y_k(x_i)$ ($k = 1 \sim 5$) を第 2 辞書 DB 122 から検索する(図 2 / S 6)。

30

【0052】

具体的には、第 1 処理部 111 は i 次入力言語単位 x_i における当該音響類似言語単位 z_i との相違部分 $\delta_i = \delta(x_i, z_i)$ を含む言語単位を第 2 辞書 DB 122 から検索して、第 1 種の i 次出力言語単位 $y_{1i} = y_1(x_i)$ として認識する。たとえば i 次入力言語単位 x_i が「Boston」という地名を表す単語であり、音響類似言語単位 z_i が「Austin」という地名を表す単語である場合、相違部分 δ_i として i 次入力言語単位 x_i の頭文字である「b」が抽出される。また、この相違部分 δ_i を含む言語単位として「bravo」が検索される。

40

【0053】

また、第 1 処理部 111 は当該相違部分 δ_i の読み方(本来の読み方) $p_{1i} = p_1(\delta_i)$ と異なる読み方 $p_{2i} = p_2(\delta_i)$ を第 2 辞書 DB 122 から検索して、第 2 種の i 次出力言語単位 $y_{2i} = y_2(x_i)$ として認識する。たとえば、日本語では、大半の漢字に「音読み」および「訓読み」という異なる読み方が存在する。このため、相違部分 δ_i である漢字「銀」の本来の読み方が音読み「ギン」である場合、その訓読み「シロガネ」が第 2 種の i 次出力言語単位 y_{2i} として認識される。

【0054】

さらに、第 1 処理部 111 は他の言語単位において当該相違部分 δ_i を意味する言語単

50

位 $f = f(x_i)$ の読み方 $p(f)$ を第 2 辞書 DB 1 2 2 から検索して、第 3 種の i 次出力言語単位 $y_{3i} = y_3(x_i)$ として認識する。たとえば日本語における漢字「銀」が相違部分 i である場合、当該漢字を意味する英単語「silver」の読み方「シルバー」が第 3 種の i 次出力言語単位 y_{3i} として認識される。

【0055】

また、第 1 処理部 1 1 1 は当該相違部分 i の読み方 $p(x_i)$ が複数のモーラ（または音素）により構成されている場合、その中から先頭モーラ等、1 つのモーラを表す音素文字、またはモーラを説明する文章を第 2 辞書 DB 1 2 2 から検索し、第 4 種の i 次出力言語単位 $y_{4i} = y_4(x_i)$ として認識する。たとえば、日本語における漢字「西」が相違部分 i である場合、その読み方 $p(x_i)$ 「ニシ」のうち最初のモーラ文字「ニ」が第 4 種の i 次出力言語単位 y_{4i} として認識される。また、日本語のモーラには、清音、半濁音（子音：p）および濁音（子音：g, z, d, b）という区分があるので、この区分を表す「清音」「半濁音」または「濁音」という単語が第 4 種の i 次出力言語単位 y_{4i} として認識される。

【0056】

さらに、第 1 処理部 1 1 1 は i 次入力言語単位 x_i と概念的に関連する言語単位を第 2 辞書 DB 1 2 2 から検索して、第 5 種の i 次出力言語単位 $y_{5i} = y_5(x_i)$ として認識する。たとえば、 i 次入力言語単位 x_i により表される目的地を包含する地域を表す言語単位（地名） $g = g(x_i)$ が第 5 種の i 次出力言語単位 y_{5i} として認識される。

【0057】

なお、第 k 種の i 次出力言語単位として、複数の言語単位が認識されてもよい。たとえば、当該相違部分 i が漢字「金」である場合、「故事成語」に分類される「沈黙は金」という文章、および「著名人の名称」に分類される「金 ×」という名称がともに第 1 種の i 次出力言語単位 y_{1i} として認識されてもよい。

【0058】

一方、第 1 処理部 1 1 1 は i 次入力言語単位 x_i に音響的に類似する言語単位が第 1 辞書 DB 1 2 1 に登録されていないと判定した場合（図 2 / S 5 NO）、 i 次入力言語単位 x_i がユーザの目的地名称を特定する言語単位であるという推定に応じた次の処理が実行される。これにより、たとえば第 2 発話部 1 0 2 が「それでは、目的地 x_i までのルートをご案内いたします」等の発話を出力する。また、ナビシステム 1 0 が、 i 次入力言語単位 x_i により特定される目的地までのルートの設定処理を実行する。

【0059】

続いて、第 2 処理部 1 1 2 が、第 1 処理部 1 1 1 によって認識された第 1 ~ 第 5 種の i 次出力言語単位 y_{ki} の中から 1 つを選定する（図 2 / S 7）。

【0060】

具体的には、第 2 処理部 1 1 2 は各種の i 次出力言語単位 y_{ki} について、次式（2）にしたがって 1 次指数 $score_1(y_{ki})$ を算出し、この i 次指数 $score_1(y_{ki})$ が最大の i 次出力言語単位 y_{ki} を選定する。

【0061】

$$\begin{aligned} score_1(y_{k1}) &= W_1 \cdot c_1(y_{k1}) + W_2 \cdot c_2(y_{k1}) + W_3 \cdot pd(x_1, y_{k1}), \\ score_{i+1}(y_{ki+1}) &= W_1 \cdot c_1(y_{ki+1}) + W_2 \cdot c_2(y_{ki+1}) + W_3 \cdot pd(x_i, y_{ki+1}) \\ &\quad + W_4 \cdot pd(y_{ki}, y_{ki+1}) \quad (2) \end{aligned}$$

式（2）において、 $W_1 \sim W_4$ は重み係数である。 $c_1(y_{ki})$ は第 k 種の i 次出力言語単位 y_{ki} の観念的な認識難易度（なじみやすさ）を表す第 1 因子である。第 1 因子としては、 i 次出力言語単位 y_{ki} をキーとしたときのインターネット検索エンジンのヒット数や、主要新聞や放送等のマスメディアにおける出現頻度等が採用される。また、 $c_2(y_{ki})$ は第 k 種の i 次出力言語単位 y_{ki} の音響的な認識難易度（発音一意性、ききわけやすさ）を表す第 2 因子である。第 2 因子としては、たとえば所定数（たとえば 10）の他の言

10

20

30

40

50

語単位（同音異義語など）との音響距離の最小平均値が採用される。 $pd(x, y)$ は、式(1)で定義される、言語単位 x および y の音響距離である。

【0062】

続いて、第2処理部112が、選定した1つの i 次出力言語単位 y_{ki} に基づき、ユーザの真意を問う i 次質問 $Q_i = Q(y_i)$ を生成して、第2発話部102に出力させる(図2/S8)。

【0063】

たとえば、第1種の i 次出力言語単位 y_{1i} が選定されたことに応じて「目的地名称には y_{1i} に含まれる α_i という文字が含まれますか」等の i 次質問 Q_i が生成される。この i 次質問 Q_i は前記相違部分 α_i を通じて、 i 次入力言語単位(たとえば、発話に含まれていた地名や建築物名称) x_i の認識の正誤を間接的にユーザに確認するための質問である。

10

【0064】

また、第2種の i 次出力言語単位 y_{2i} が選定されたことに応じて「目的地名称には p_{2i} と読まれる(または発音される)文字が含まれていますか」等の i 次質問 Q_i が生成される。この i 次質問 Q_i は相違部分 α_i の本来の読み方 p_{1i} とは異なる読み方 p_{2i} を通じて、 i 次入力言語単位 x_i の認識の正誤を間接的にユーザに確認するための質問である。

【0065】

さらに、第3種の i 次出力言語単位 y_{3i} が選定されたことに応じて「目的地名称には外国語(たとえば、日本語から見た英語)で p を意味する α_i という文字が含まれていますか」等の i 次質問 Q_i が生成される。この i 次質問 Q_i は他の言語単位において当該相違部分 α_i を意味する言語単位 $f = f(\alpha_i)$ の読み方 $p(f)$ を通じて、 i 次入力言語単位 x_i の認識の正誤を間接的にユーザに確認するための質問である。

20

【0066】

また、第4種の i 次出力言語単位 y_{4i} が選定されたことに応じて「目的地名称には $p(\alpha_i)$ と発音される文字が含まれていますか」等の i 次質問 Q_i が生成される。この i 次質問 Q_i は相違部分 α_i の読み方 $p(\alpha_i)$ の中の1つのモーラを表すモーラを表す文字、またはモーラを説明する文章を通じて、 i 次入力言語単位 x_i の認識の正誤を間接的にユーザに確認するための質問である。

【0067】

さらに、第5種の i 次出力言語単位 y_{5i} が選定されたことに応じて「目的地は g に含まれていますか」等の i 次質問 Q_i が生成される。この i 次質問 Q_i は、 i 次入力言語単位 x_i と概念的に関連する言語単位を通じて、 i 次入力言語単位 x_i の認識の正誤を間接的にユーザに確認するための質問である。

30

【0068】

さらに、第1発話部101が、 i 次質問 Q_i に対するユーザの発話として i 次回答 A_i を認識する(図2/S9)。また、第2処理部112が、 i 次回答 A_i が「はい」のように肯定的なものであるか、または「いいえ」のように否定的なものであるかを判定する(図2/S10)。

【0069】

そして、第2処理部112により i 次回答 A_i が肯定的であると判定された場合(図2/S10 YES)、 i 次入力言語単位 x_i がユーザの目的地名称を特定する言語単位であるという推定に応じた次の処理が実行される。

40

【0070】

一方、第2処理部112により i 次回答 A_i が否定的であると判定された場合(図2/S10 NO)、指数 i が所定数 $j (> 2)$ 未満であるという条件が満たされているかが判定される(図2/S11)。そして、当該条件が満たされている場合(図2/S11 YES)、指数 i が1だけ増加され(図2/S12)、その上で前記のS4~S10の処理が繰り返される。この際、第1処理部111は、 $i-1$ 次入力言語単位 x_{i-1} ($i \geq 2$)に音響的に類似する言語単位を第1辞書DB121から検索して、 i 次入力言語単位 x_i として認識する。なお、 i 次入力言語単位 x_i として、 $i-1$ 次入力言語単位 x_{i-1}

50

の音響類似言語単位 z_{i-1} が認識されてもよい。また、当該条件が満たされていない場合（図2 / S 1 1 NO）、第2発話部 1 0 2 があらためて初期発話を出力する（図2 / S 1）等、ユーザとの会話が振り出しに戻ってやり直される。

【0071】

前記機能を発揮する会話システム 1 0 0（および会話ソフトウェア）によれば、それぞれの i 次出力言語単位 y_{ki} について、観念的な認識難易度を表す第1因子 c_1 、および音響的な認識難易度を表す第2因子 c_2 等に基づき、複数種類の i 次出力言語単位 y_{ki} の中から1つが選定される（図2 / S 6, S 7）。また、選定された1つの i 次出力言語単位 y_{ki} に基づき i 次質問 Q_i が生成される（図2 / S 8）。これにより、ユーザの真意と i 次入力言語単位 x_i の整合および不整合を判別する観点から最適な i 次質問 Q_i が生成される。また、ユーザの真意とシステムの認識とに齟齬があると判定された場合、さらなる質問が生成される（図2 / S 1 0 NO, S 4 ~ S 1 0）。したがって、ユーザの発話（真意）と、システム 1 0 0 において認識された発話との齟齬を確実に抑制しながら、当該ユーザとシステム 1 0 0 との会話が可能となる。

10

【0072】

さらに、ユーザの真意と j 次入力言語単位（ $j \geq 2$ ）とが整合していないと判定された場合、ユーザの再度の発話を促す初期質問が生成される（図2 / S 1 1 NO, S 1）。これにより、逐次出力される質問によってはユーザの真意が確認できないような場合、あらためてその真意を確認することができる。

【0073】

20

前記処理にしたがったユーザおよび会話システム 1 0 0 の第1の会話例を次に示す。U はユーザの発話を表し、S は会話システム 1 0 0 の発話を表している。

（第1の会話例）

S₀: 目的地はどこですか。

【0074】

U₀: 金閣寺です。

【0075】

S₁: 目的地名称に英語でシルバーを意味する「銀」という文字が含まれていますか。

【0076】

U₁: いいえ。

【0077】

S₂: では、目的地名称に「沈黙は金」における「金」という文字が含まれていますか。

30

【0078】

U₂: はい。

【0079】

S₃: それでは金閣寺までのルートを案内いたします。

【0080】

システム 1 0 0 の発話 S₀ は初期質問に該当する（図2 / S 1）。

40

【0081】

システム 1 0 0 の発話 S₁ は1次質問 Q₁ に該当する（図2 / S 8）。この1次質問 Q₁ は、1次入力言語単位 x_1 として「金閣寺」ではなく「銀閣寺」が認識（誤認）されたこと（図2 / S 4）、音響類似言語単位 z_1 として「金閣寺」が認識されたこと（図2 / S 5）、2つの言語単位 x_1 および z_1 の相違部分 Δ_1 である漢字「銀」に関連する5種類の1次出力言語単位 $y_{11} \sim y_{51}$ が認識されたこと（図2 / S 6）、および第3種の1次出力言語単位 y_{31} として当該相違部分 Δ_1 を表す英単語「silver」の日本語における読み方「シルバー」が選定されたこと（図2 / S 7）に応じて生成されたものである。

【0082】

システム 1 0 0 の発話 S₂ は2次質問 Q₂ に該当する（図2 / S 8）。この2次質問 Q₂

50

は、1次回答 A_1 として認識されたユーザの発話 U_1 が否定的なものであったこと(図2/S10 NO)、2次入力言語単位 x_2 として「金閣寺」が認識されたこと(図2/S4)、音響類似言語単位 z_2 として「銀閣寺」が認識されたこと(図2/S5)、2つの言語単位 x_2 および z_2 の相違部分 $_2$ である漢字「金」に関連する5種類の2次出力言語単位 $y_{12} \sim y_{52}$ が認識されたこと(図2/S6)、および第1種の2次出力言語単位 y_{12} として当該相違部分 $_2$ を含む故事成語「沈黙は金」が選定されたこと(図2/S7)に応じて生成されたものである。

【0083】

2次回答 A_2 として認識されたユーザの発話 U_2 が肯定的なものであったことに応じて(図2/S10 YES)、ユーザの目的地が金閣寺であるという判断に応じて、システム100から発話 U_4 が出力される。

10

【0084】

これにより、ユーザの目的地が「金閣寺」である一方、システム100により認識された目的地が「銀閣寺」であるといった齟齬が生じたまま、ユーザおよびシステム100の会話が進行する事態が回避される。すなわち、システム100は、ユーザの目的地が金閣寺であることを正確に認識することができる。そして、ナビシステム10は、システム100の当該認識に基づき、金閣寺までの案内ルートの設定等、ユーザの真意に鑑みて適切な処理を実行することができる。

【0085】

さらに、前記処理にしたがったユーザおよびシステム100の第2の会話例を次に示す。

20

(第2の会話例)

S_0 : Can you tell me the departure
city?

U_0 : from Austin.

S_1 : Is the first letter of the city "b"
in "bravo"?

U_1 : No.

S_2 : Then is the first letter of the city
"a" in "alpha"?

30

U_2 : Yes.

システム100の発話 S_0 は初期質問に該当する(図2/S1)。

【0086】

システム100の発話 S_1 は1次質問 Q_1 に該当する(図2/S8)。この1次質問 Q_1 は、1次入力言語単位 x_1 として「Austin」ではなく「Boston」が認識(誤認)されたこと(図2/S4)、音響類似言語単位 z_1 として「Austin」が認識されたこと(図2/S5)、2つの言語単位 x_1 および z_1 の相違部分 $_1$ である英文字「b」に関連する5種類の1次出力言語単位 $y_{11} \sim y_{51}$ が認識されたこと(図2/S6)、および第1種の1次出力言語単位 y_{11} として当該相違部分 $_1$ を表す英単語「bravo」が選定されたこと(図2/S7)に応じて生成されたものである。

40

【0087】

システム100の発話 S_2 は2次質問 Q_2 に該当する(図2/S8)。この2次質問 Q_2 は、1次回答 A_1 として認識されたユーザの発話 U_1 が否定的なものであったこと(図2/S10 NO)、2次入力言語単位 x_2 として「Austin」が認識されたこと(図2/S4)、音響類似言語単位 z_2 として「Boston」が認識されたこと(図2/S5)、2つの言語単位 x_2 および z_2 の相違部分 $_2$ である英文字「a」に関連する5種類の2次出力言語単位 $y_{12} \sim y_{52}$ が認識されたこと(図2/S6)、および第1種の2次出力言語単位 y_{12} として当該相違部分 $_2$ を含む英単語「alpha」が選定されたこと(図2/S7)に応じて生成されたものである。

【0088】

50

2次回答 A_2 として認識されたユーザの発話 U_2 が肯定的なものであったことに応じて (図2 / S10 YES)、ユーザの目的地が Austin であるという判断に応じて、システム100から発話が出力される。

【0089】

これにより、ユーザの目的地が「Austin」である一方、システム100により認識された目的地が「Boston」であるといった齟齬が生じたまま、ユーザおよびシステム100の会話が進行する事態が回避される。すなわち、システム100は、ユーザの目的地が Austin であることを正確に認識することができる。そして、ナビシステム10は、システム100の当該認識に基づき、Austin までの案内ルートの設定等、ユーザの真意に鑑みて適切な処理を実行することができる。

10

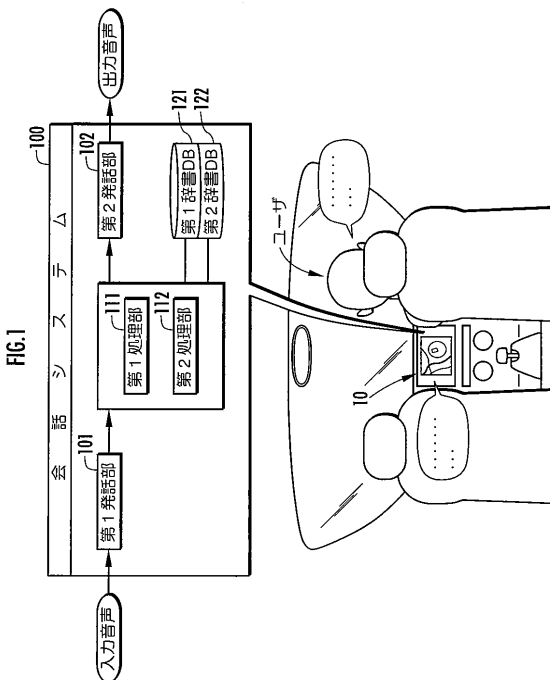
【図面の簡単な説明】

【0090】

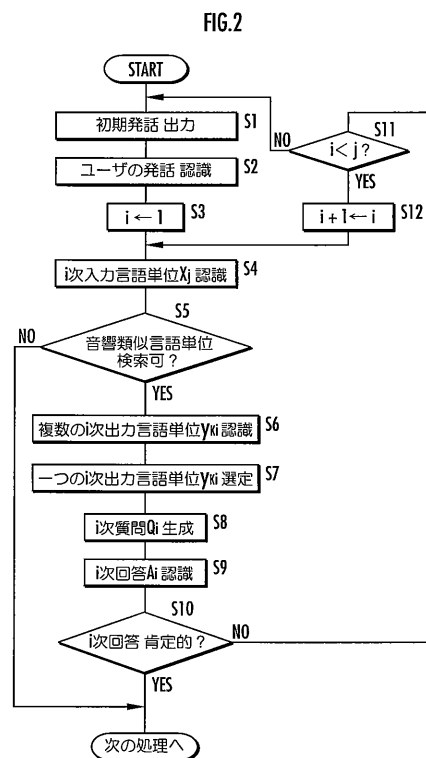
【図1】本発明の会話システムの構成例示図。

【図2】本発明の会話システムおよび会話ソフトウェアの機能例示図。

【図1】



【図2】



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丁目1番 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

京都府京都市中京区東洞院通三条下る三文字町205番地の3 フォルム東洞院三条1102号

(72)発明者 駒谷 和範

京都府京都市左京区吉田下大路町65-301

審査官 井上 健一

(56)参考文献 特開2003-228394(JP,A)

特開平11-153998(JP,A)

特開平10-269226(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 15/22

G10L 15/00