(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2007-232829
(P2007-232829A)
(43)公開日　平成19年9月13日(2007.9.13)

(51)Int.Cl.		F I		テーマコード（参考）
G10L 13/00	(2006.01)	G10L 13/00	100M	5D015
G10L 15/22	(2006.01)	G10L 15/22	200H	
G10L 13/02	(2006.01)	G10L 13/02	121A	
		G10L 13/02	130B	

審査請求 有　請求項の数 5　OL　（全 8 頁）

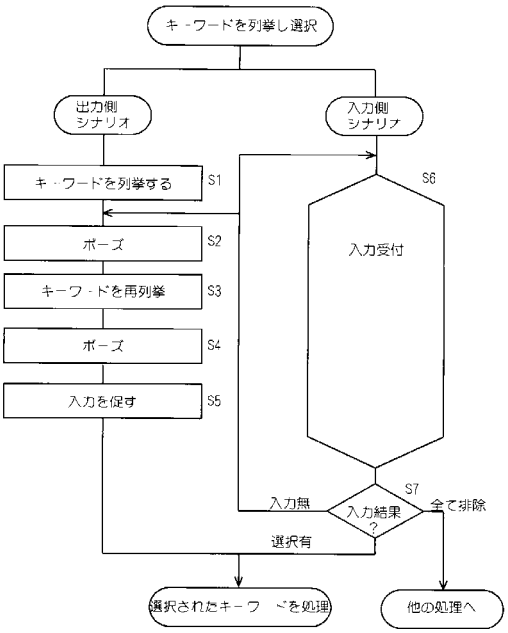
(21)出願番号　特願2006-51771(P2006-51771)
(22)出願日　平成18年2月28日(2006.2.28)

(71)出願人　000006297
村田機械株式会社
京都府京都市南区吉祥院南落合町３番地
(74)代理人　100086830
弁理士　塩入　明
(74)代理人　100096046
弁理士　塩入　みか
(72)発明者　新堂　安孝
京都市伏見区竹田向代町１３６番地　村田機械株式会社内

Fターム(参考)　5D015 AA01　AA04

(54)【発明の名称】音声対話装置とその方法及びプログラム

(57)【要約】
【課題】　音声対話装置が多数のキーワードを提示すると、人は個々のキーワードを聞き逃し易いし、また最適キーワードを選択するのが難しい。
【解決手段】　対話装置からキーワードを一旦列挙し、ポーズの後再度キーワードを提示して、選択を求める。有効な選択が有れば、選択肢に従ってシナリオを進め、選択がなければ再度キーワードを提示する。またどのキーワードも否定された場合、シナリオの他の場面へ分岐する。
【選択図】　図４



【特許請求の範囲】
【請求項１】
人が音声入力するためのマイクロフォンと、マイクロフォンへの入力音声を音声認識するための音声認識装置と、スピーカを備えた音声出力装置と、シナリオの記憶部と、前記音声認識装置と音声出力装置とをシナリオに沿って制御する処理システム、とを備えた音声対話装置において、
複数のキーワードを列挙するようにスピーカから音声出力する際に、該複数のキーワードを最初に列挙した後に、音声出力をポーズし、次いで前記複数のキーワードを再度列挙して、人の入力音声を受け付けるように、前記記憶部に記憶するシナリオを構成することを特徴とする、音声対話装置。
【請求項２】
再度の列挙では、最初に列挙したキーワードを、同順で、かつ少なくとも一部のキーワードを同義語に変換して音声出力するように、前記シナリオを構成することを特徴とする、請求項１の音声対話装値。
【請求項３】
列挙されたキーワードに対する人からの入力音声を、遅くとも前記再度の列挙から処理するように、前記音声認識装置を構成することを特徴とする、請求項１の音声対話装置。
【請求項４】
マイクロフォンで人の入力音声を受け付けて音声認識装置で音声認識し、予め記憶したシナリオに沿って、処理システムで音声認識装置と音声出力装置とを制御する音声対話方法において、
スピーカから複数のキーワードを列挙し、次いで音声出力をポーズし、さらに前記複数のキーワードを再度列挙して、人の入力音声を音声認識装置で認識する、ことを特徴とする、音声対話方法。
【請求項５】
マイクロフォンで人の入力音声を受け付けて音声認識装置で音声認識し、予め記憶したシナリオに沿って、処理システムで音声認識装置と音声出力装置とを制御する音声対話用のプログラムにおいて、
スピーカから複数のキーワードを列挙する命令と、次いで音声出力をポーズする命令と、さらに前記複数のキーワードを再度列挙する命令と、少なくとも前記複数のキーワードを再度列挙する際に、人の入力音声を音声認識装置で認識する命令とを設けたことを特徴とする、音声対話プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
この発明は情報処理装置と人との音声による対話に関し、特にガイダンスなどのために装置側で予め記憶しているシナリオでの質問に人が答え易くすることに関する。
【背景技術】
【０００２】
音声対話装置が、人に多数のキーワードを列挙して選択を求める場面がある。このような場面で単純にキーワードを列挙されると、人は個々のキーワードを聞き逃し易い。そこでキーワード毎にポーズを入れると人に理解し易くなるが、各々のポーズ長の設定をどのようにするかなどで、シナリオの作成は複雑になる。
【特許文献１】特開平１１－２８８２９２号公報
【発明の開示】
【発明が解決しようとする課題】
【０００３】
この発明の課題は、多数のキーワードを音声で人に提示する際に、聞き取り易くかつ最適な選択をし易くすることにある。
請求項２の発明での追加の課題は、キーワードを繰り返すことにより対話が平板になり

、その結果冗長に感じられないように、しかも人が回答し易くすることにある。
請求項３，５の発明での追加の課題は、第２回目の列挙が終わるのを待たずに、人が回答できるようにすることにある。
【課題を解決するための手段】
【０００４】
この発明は、人が音声入力するためのマイクロフォンと、マイクロフォンへの入力音声を音声認識するための音声認識装置と、スピーカを備えた音声出力装置と、シナリオの記憶部と、前記音声認識装置と音声出力装置とをシナリオに沿って制御する処理システム、とを備えた音声対話装置において、複数のキーワードを列挙するようにスピーカから音声出力する際に、該複数のキーワードを最初に列挙した後に、音声出力をポーズし、次いで前記複数のキーワードを再度列挙して、人の入力音声を受け付けるように、前記記憶部に記憶するシナリオを構成することを特徴とする。
【０００５】
好ましくは、再度の列挙では、最初に列挙したキーワードを、同順で、かつ少なくとも一部のキーワードを同義語に変換して音声出力するように、前記シナリオを構成する。
また好ましくは、列挙されたキーワードに対する人からの入力音声を、遅くとも前記再度の列挙から処理するように、前記音声認識装置を構成する。
【０００６】
この発明は、マイクロフォンで人の入力音声を受け付けて音声認識装置で音声認識し、予め記憶したシナリオに沿って、処理システムで音声認識装置と音声出力装置とを制御する音声対話方法において、スピーカから複数のキーワードを列挙し、次いで音声出力をポーズし、さらに前記複数のキーワードを再度列挙して、人の入力音声を音声認識装置で認識する、ことを特徴とする。
人の入力音声の認識は例えば２回目の列挙から開始し、好ましくは１回目の列挙中、または列挙後のポーズから開始する。
【０００７】
この発明は、マイクロフォンで人の入力音声を受け付けて音声認識装置で音声認識し、予め記憶したシナリオに沿って、処理システムで音声認識装置と音声出力装置とを制御する音声対話プログラムにおいて、スピーカから複数のキーワードを列挙する命令と、次いで音声出力をポーズする命令と、さらに前記複数のキーワードを再度列挙する命令と、少なくとも前記複数のキーワードを再度列挙する際に、人の入力音声を音声認識装置で認識する命令とを設けたことを特徴とする。
【０００８】
この明細書において、音声対話装置に関する記載は音声対話方法や音声対話プログラムにもそのまま当てはまり、逆に音声対話方法に関する記載は音声対話装置や音声対話プログラムにもそのまま当てはまる。
キーワードの列挙に対する人の回答は例えばその選択である。
【発明の効果】
【０００９】
この発明では複数のキーワードを列挙して回答を求める際に、最初にキーワードを列挙し、一旦音声出力を停止して間を置き、次いでキーワードを再列挙する。このため最初の列挙でキーワードを聞き逃しても、次の列挙で正しく聞いて応答できる。最初の列挙と次の列挙の間にポーズが有るので、次の列挙が始まったときに、同じキーワードを繰り返していると直ちに分かる。さらに最初の列挙でキーワードの集合の概要を理解してから、キーワードが再度出力された時に回答すれば良く、回答が容易である。シナリオ作成の面では、ポーズ長を使い分ける必要が無く、ポーズの設定が簡単である。
【００１０】
２回目の列挙でキーワードを同義語に変換して出力すると、対話が平板になり難い。そしてキーワードの語順を１回目と同じにすると、１回目で人が聞いた順序でキーワードが繰り返されるので、回答し易い。

２回目にキーワードを列挙する際には、人はほぼ回答できる状態にあるので、ここでキーワードを出力しながら回答を音声認識すると、人がキーワードを聞いて直ちに回答することに対応できる。

【発明を実施するための最良の形態】

【００１１】

以下に本発明を実施するための最適実施例を示す。

【実施例】

【００１２】

図１～図６に、実施例を示す。各図において、２は音声対話装置で、４は音声入力用のマイクロフォンで、６はアンプで省略可能である。８は音声認識装置で、１０は音声認識に用いる辞書で、複数の辞書が対話装置２に記憶されている。１２は音声の認識結果を出力するためのレジスタで、１４は処理システムであり、１６は音声対話用のシナリオ記憶部である。なおシナリオの個々のシーンを場面と呼び、各場面の記憶位置を番地と呼ぶ。

【００１３】

シナリオ記憶部１６に記憶するシナリオの構成を図２に示す。一般的シナリオ１８はキーワードの列挙以外の部分のシナリオであり、キーワード列挙シナリオ２０がキーワードの列挙に関する部分のシナリオである。なおキーワードの列挙とは２個以上のキーワードを列挙することで、好ましくは３個以上のキーワードを列挙することである。第１回キーワード列挙場面２１で最初にキーワードを列挙し、ポーズ場面２２で一旦ポーズし、第２回キーワード列挙場面２３で再度キーワードを列挙する。ポーズ場面２４で２回目のキーワードの列挙後にポーズし、入力促し場面２５で２回目のポーズの後に入力を人に促す。シナリオには音声出力側のシナリオと音声入力側のシナリオとが同期して進行し、場面２１～２５は出力側のシナリオである。音声入力側のシナリオでは、入力受け付け場面２６で列挙されたキーワードに対する人の選択を受け付けて音声認識する。そして列挙されたキーワードに対する音声認識は、例えば第１回キーワード列挙場面２１、もしくはポーズ場面２２、あるいは第２回キーワード列挙場面２３から開始し、これに応じて図１の辞書１０を列挙したキーワードに対応したものに切り替え、認識開始前にレジスタ１２をゼロクリアする。

【００１４】

図３にレジスタ１２の構造を示す。ここではＡ～Ｇの７つのキーワードが列挙され、人に対してキーワードの選択を求めているものとする。有効な回答は、１以上のキーワードを選択することと、「全て要らない」、「要りません」などの全ての選択肢の否定である。レジスタ１２には質問ＩＤが記載され、これは入力側シナリオでの場面を示している。次の１ビットは回答が肯定か否定かを示し、肯定で０，否定でＦである。各キーワードには同義語があり、例えば大学の学部案内の場合、「工学部」と「工」並びに「工学」は同義語である。そこで同義語をまとめて抽象化したものをサブジェクトと呼ぶと、人の回答はサブジェクトの選択と言い換えることができる。図３のレジスタ１２では、Ａ～Ｇの７個のサブジェクトに対し各１ビットを割り当てる。サブジェクトの個数は質問毎に変化するので、肯定／否定のビットの後のビットから、各サブジェクト毎に１ビット割り当てることとし、レジスタ１２には充分大きな記憶容量を持たせる。

【００１５】

「Ａは要らないが、Ｂは要る」などのように肯定と否定が組み合わさった回答に備え、レジスタ１２を複数段設け、「Ａは要らない」を初段のレジスタで処理し、「Ｂは要る」を次段のレジスタで処理しても良い。また肯定／否定やサブジェクトの選択を各１ビットのデータで記憶する必要はなく、よりビット長の長いデータで記憶しても良い。

【００１６】

辞書１０には列挙する各キーワードとその同義語、並びにキーワードの範囲や組み合わせを示す単語、及び肯定／否定を表す語が記憶されている。キーワードの範囲や組み合わせを示す単語として「全部」，「全て」などがあり、大学の学部案内での理学と工学の組み合わせを示す単語として「理工学」、文学部と経済学部及び商学部の組み合わせを示す

ものとして「文化系」などがある。これらのキーワードや同義語等は、入力側シナリオの場面毎に辞書を変更して切り替える。肯定を表す語として「はい」や「お願い」、接尾辞の「たい」「です」「ます」などがある。否定を表す語として「いいえ」や接尾辞の「ない」などがある。肯定を表す語も否定を表す語も入力されない場合、肯定／否定のビットは初期値の肯定のままである。

【００１７】

音声認識装置８は入力音声中に辞書１０に記述された語があると、レジスタ１２のこれに対応するビットへの書き込みを行う。単語が肯定あるいは否定を表す場合、肯定／否定のビットに対して０もしくはＦを出力し、サブジェクト毎のビットでは該当ビットをＦにする。またサブジェクトの集合に対応するキーワードを発見すると、その集合に含まれるサブジェクトのビットをＦにセットする。そして音声認識装置８はキーワードを発見する都度、レジスタ１２へＯＲ加算により書き込みする。例えば大学の学部案内で「文学をお願い」との回答が入力された場合、「文学」をキーワードとして検出し、これに対するサブジェクトのビットをＦにセットし、他のビットは０のままである。次に「お願い」が肯定に対応するので、先頭の肯定ビットを０に保ち、他のビットの値を変えない。この場合のレジスタ１２の出力は、肯定ビットが０で肯定であり、「文学」のビットがセットされ、他はセットされていないので、文学についてのみ案内を依頼したことになる。「文学と経済をお願い」の場合、「文学」のビットと「経済」のビットがセットされ、肯定／否定のビットは肯定の０ビットに保たれる。

【００１８】

列挙されたキーワードに対する選択を認識する際の特別のルールとして、「はい」「それ」などのキーワードを特定しない入力には、その直前に出力されたキーワードが肯定で選択されたものとする。これは２回目の列挙の途中で「はい」などが入力されることに備えたものであるが、このルールは設けなくても良い。また「文学は要らないが、経済学を知りたい」などの肯定／否定の述語が２語以上ある入力に対し、レジスタ１２を複数段設けると、１段目のレジスタで「文学は要らない」に対し肯定／否定のビットを否定のＦにセットし、「文学」のビットをＦにセットする。次の段のレジスタで「経済学を知りたい」を処理し、肯定／否定のビットを肯定の０にセットし、「経済学」のビットをＦにセットする。この場合の認識結果は、「経済学部について知りたい」というものと同じである。

【００１９】

図１に戻り、音声データ生成部３０はシナリオに基づいて音声を生成し、アンプ３２を介してスピーカ３４から音声出力する。なおアンプ３２は設けなくても良い。実施例では音声対話装置２をガイダンス用のロボットに組み込み、処理システム１４からのジェスチャー信号でロボット本体３６を動作させる。

【００２０】

図４に、実施例の音声対話方法を示す。シナリオ中のキーワードを列挙し選択する部分で、出力側のシナリオではステップ１でキーワードを列挙する。次いでステップ２でポーズし、ステップ３でキーワードを再列挙する。そしてステップ４でポーズした後、ステップ５で入力を促す。ただしステップ４のポーズは省略しても良い。実施例の場合、ステップ２あるいはステップ４で、ロボット本体３６に身振りをさせても良い。またステップ３でキーワードを再列挙する場合、ステップ１で列挙したキーワードの一部を同義語に変換し、特によりシンプルな単語に変更して同じ順で列挙すると、１回目と２回目とで表現を変えてしかも同じ順なので、人が答え易くでき、しかも冗長さを弱めることができる。

【００２１】

入力側シナリオでは、ステップ１のキーワードの列挙から、入力を受け付け、入力音声を音声認識する。ただしステップ２のポーズから、もしくはステップ３のキーワードの再列挙から、入力音声を音声認識しても良い。そしてステップ７で入力結果を判別し、有効な入力がなかった場合、ステップ２のポーズまでもしくはステップ３のキーワードの再列挙まで戻って、あるいはキーワードの再列挙に準じた処理を行って、入力を再度受け付け

る。全ての選択肢が否定された場合他の処理へ移り、いずれかのキーワードが選択されると、選択されたキーワードやその組み合わせについてガイダンスする。

【００２２】

図５に音声対話プログラム４０の構造を示すと、一般シナリオ命令４１はキーワードの列挙以外のシナリオを処理し、キーワード列挙命令４２は１回目のキーワードの列挙を処理し、キーワード再列挙命令４３は２回目のキーワードの列挙を処理する。ポーズ命令４４は、１回目のキーワードの列挙と２回目のキーワードの列挙の間、並びに２回目のキーワードの列挙後の、ポーズやジェスチャーを処理する。音声認識命令４５は例えば１回目のキーワード列挙中から音声認識を開始し、図１の辞書１０を列挙されたキーワードに合わせて切り替える。音声認識命令４５は入力音声の認識結果に基づき、シナリオでのそれ以前の処理へ戻る、他の処理へ移る、選択されたキーワードについてガイダンスする、ようにシナリオを分岐させる。促し命令４６は、２回目のキーワードの列挙後と２回目のポーズの後に、入力を促す文を出力する。

【００２３】

図６に、大学の学部案内を例に音声ガイダンスの具体例を示す。なおこの具体例は実施例の音声対話装置、音声対話方法、音声対話プログラムのいずれにも該当する。ステップ１１で、大学の学部としてどのようなものがあるかを列挙する。ステップ１２で、ロボット本体のジェスチャーを交えてポーズし、ステップ１３でキーワードを再列挙する。ここで文学を「文」、経済学を「経済」としているように、キーワードをより短いキーワードで同義語のものに変換して同じ順序で列挙する。ステップ１４で再度ポーズし、ステップ１５で回答を促す文を出力する。

【００２４】

入力側シナリオでは、ステップ１１の段階で「経済学」などの回答がされることに備えて、ステップ１１のキーワードの列挙から入力音声を認識する。ただしステップ１２のポーズから、もしくはステップ１３の再列挙から、入力音声の認識を開始しても良い。そしてステップ１７で、入力結果に基づき分岐する。

【００２５】

実施例では以下の効果がある。

(1) キーワードが２回以上列挙されるので、人にとって聞き逃しにくい。

(2) １回目のキーワードの列挙でキーワードの全体をおおよそ理解し、２回目のキーワードの列挙で正確に聞き取って回答すればよいので、正確な回答が容易である。

(3) １回目と２回目とでキーワードの列挙の仕方を変えるので、対話が平板になりにくい。

(4) サブジェクト毎のビット和を用いるので、「文学」，「経済学」などの個別の回答から、「文化系」，「全部」などの範囲を示す回答まで対応できる。なお「Ａ，Ｂ，Ｃは要らない」などの入力に対し、これ以外のキーワードが選択されたものとして解釈すると、入力を認識できる範囲をさらに広げることができる。

【図面の簡単な説明】

【００２６】

【図１】実施例の音声対話装置のブロック図

【図２】実施例で用いるシナリオのブロック図

【図３】実施例で用いる音声認識用のレジスタを示す図

【図４】実施例の音声対話方法を示すフローチャート

【図５】実施例の音声対話プログラムのブロック図

【図６】実施例を大学の学部案内に応用した例を示す図

【符号の説明】

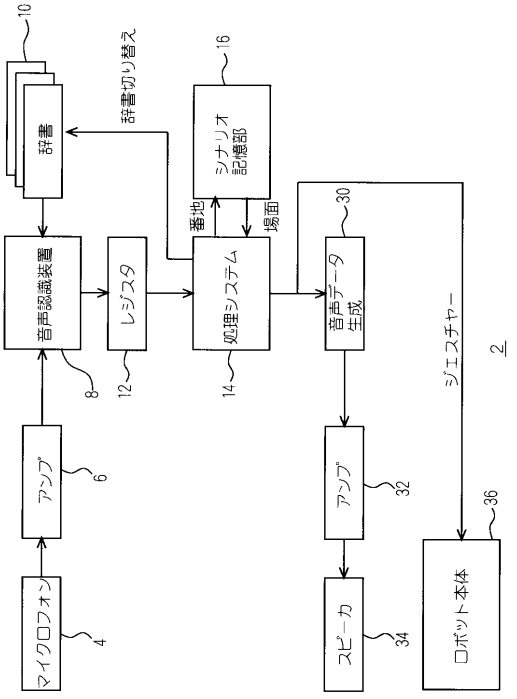
【００２７】

２ 音声対話装置

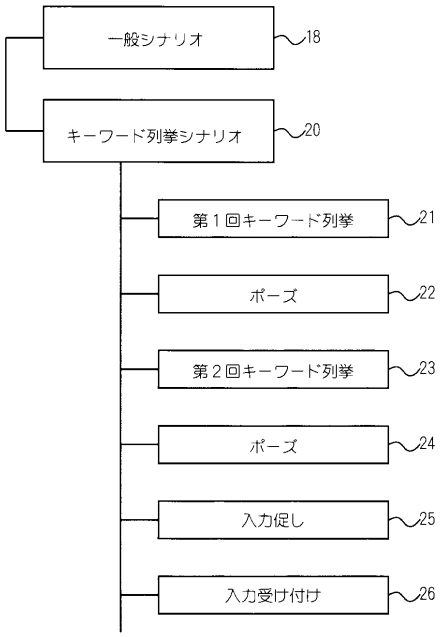
４ マイクロフォン

６，３２　アンプ
８　音声認識装置
１０　辞書
１２　レジスタ
１４　処理システム
１６　シナリオ記憶部
１８　一般的シナリオ
２０　キーワード列挙シナリオ
２１　第１回キーワード列挙場面
２２　ポーズ場面
２３　第２回キーワード列挙場面
２４　ポーズ場面
２５　入力促し場面
２６　入力受け付け場面
３０　音声データ生成部
３４　スピーカ
３６　ロボット本体
４０　音声対話プログラム
４１　一般シナリオ命令
４２　キーワード列挙命令
４３　キーワード再列挙命令
４４　ポーズ命令
４５　音声認識命令
４６　促し命令

【図１】

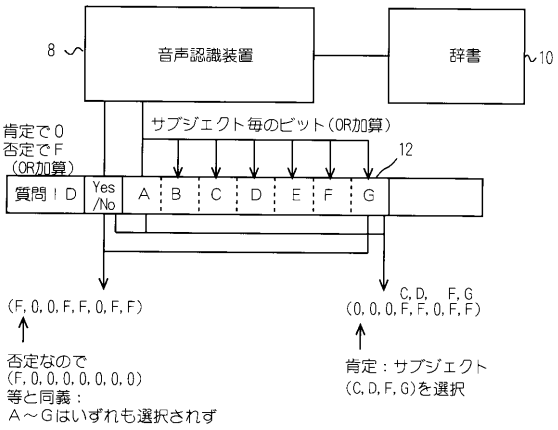


【図２】

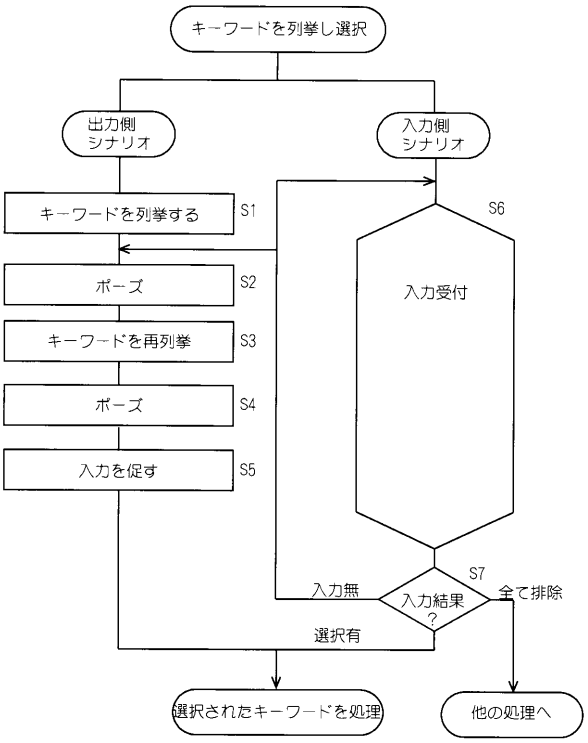


16

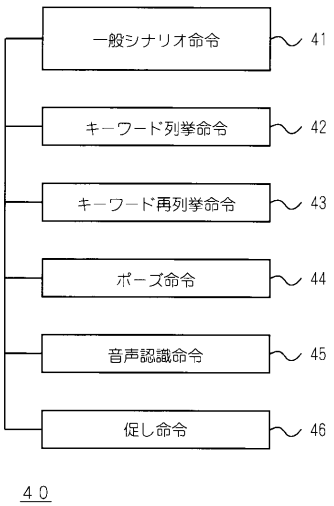
【図３】


8
音声認識装置
辞書
10
肯定で0
否定でF
(OR加算)
サブジェクト毎のビット(OR加算)
12
質問！D
Yes
/No
A
B
C
D
E
F
G
(F, 0, 0, F, F, 0, F, F)
否定なので
(F, 0, 0, 0, 0, 0, 0, 0)
等と同義：
A～Gはいずれも選択されず
C, D, F, G
(0, 0, 0, F, F, 0, F, F)
肯定：サブジェクト
(C, D, F, G)を選択

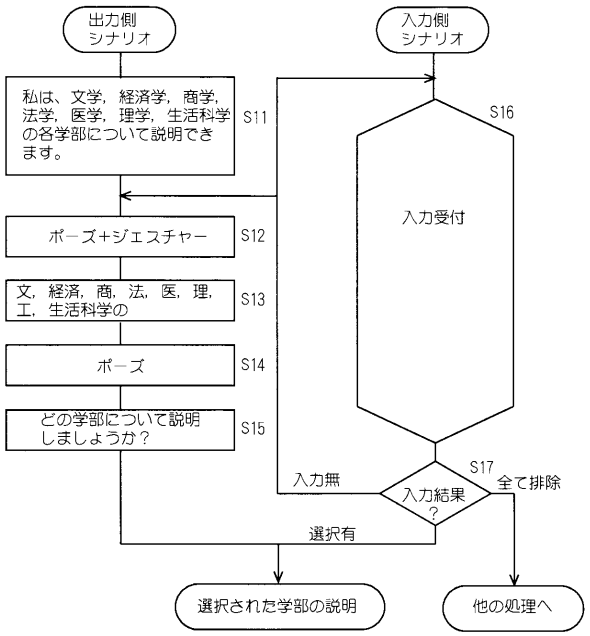
【図４】


キーワードを列挙し選択
出力側
シナリオ
入力側
シナリオ
キーワードを列挙する
S1
ポーズ
S2
キーワードを再列挙
S3
ポーズ
S4
入力を促す
S5
S6
入力受付
入力無
S7
入力結果
？
全て排除
選択有
選択されたキーワードを処理
他の処理へ

【図５】


一般シナリオ命令
41
キーワード列挙命令
42
キーワード再列挙命令
43
ポーズ命令
44
音声認識命令
45
促し命令
46
40

【図６】


出力側
シナリオ
入力側
シナリオ
私は、文学，経済学，商学，法学，医学，理学，生活科学の各学部について説明できます。
S11
S16
入力受付
ポーズ+ジェスチャー
S12
文，経済，商，法，医，理，工，生活科学の
S13
ポーズ
S14
どの学部について説明しましょうか？
S15
入力無
S17
入力結果
？
全て排除
選択有
選択された学部の説明
他の処理へ