(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第4416643号
(P4416643)

(45)発行日　平成22年2月17日(2010.2.17)　(24)登録日　平成21年12月4日(2009.12.4)

(51)Int.Cl.		F I		
G06F 3/16	(2006.01)	G06F	3/16	320A
G06F 3/01	(2006.01)	G06F	3/01	
G10L 15/24	(2006.01)	G10L	15/24	R

請求項の数 5　(全 14 頁)

(21)出願番号	特願2004-379948 (P2004-379948)
(22)出願日	平成16年12月28日 (2004.12.28)
(65)公開番号	特開2006-48628 (P2006-48628A)
(43)公開日	平成18年2月16日 (2006.2.16)
審査請求日	平成19年12月10日 (2007.12.10)
(31)優先権主張番号	特願2004-191632 (P2004-191632)
(32)優先日	平成16年6月29日 (2004.6.29)
(33)優先権主張国	日本国(JP)

(73)特許権者　000001007
キヤノン株式会社
東京都大田区下丸子３丁目３０番２号
(74)代理人　100090538
弁理士　西山　恵三
(74)代理人　100096965
弁理士　内尾　裕一
(72)発明者　池田　裕美
東京都大田区下丸子３丁目３０番２号キヤノン株式会社内

審査官　円子　英紀

最終頁に続く

(54)【発明の名称】マルチモーダル入力方法

(57)【特許請求の範囲】

【請求項１】

音声認識された情報とＧＵＩ入力された情報とを統合した認識結果を出力する情報処理装置の情報処理方法であって、

音声情報を受信する音声情報受信工程と、

前記音声情報を受信している時間内にユーザに操作された１または複数のボタンに対応するＧＵＩ入力情報を受信するＧＵＩ情報受信工程と、

前記音声情報を音声認識し、尤度が高い順に複数の解釈結果の候補を取得する音声認識工程と、

前記尤度が最も高い解釈結果の候補に含まれる未確定語の数と、前記ＧＵＩ入力情報の数とが一致しているか否か判断する判断工程と、

前記尤度が最も高い解釈結果の候補に含まれる未確定語の数と、前記ＧＵＩ入力情報の数とが一致していないと判断された場合、解釈結果の候補に含まれる未確定語の数と、前記ＧＵＩ入力情報の数とが一致している別の解釈結果の候補を検索する検索工程と、

検索された解釈結果の候補に含まれる未確定語を前記ＧＵＩ入力情報で置換えた情報を、認識結果として出力する出力工程とを有する情報処理方法。

【請求項２】

前記検索工程は、前記尤度が高い解釈結果の候補から順に、解釈結果の候補に含まれる未確定語の数と、前記ＧＵＩ入力情報の数とが一致している別の解釈結果の候補を検索することを特徴とする請求項１記載の情報処理方法。

【請求項３】
前記ＧＵＩ入力情報の数とは、前記ＧＵＩ入力手段から、前記音声情報を受信している時間内にユーザに操作されたボタンの個数または回数であることを特徴とする請求項１記載の情報処理方法。
【請求項４】
音声入力手段から音声情報を受信する音声情報受信手段と、
ＧＵＩ入力手段から、前記音声情報を受信している時間内にユーザに操作された１または複数のボタンに対応するＧＵＩ入力情報を受信するＧＵＩ情報受信手段と、
前記音声情報を音声認識し、尤度が高い順に複数の解釈結果の候補を取得する音声認識手段と、
前記尤度が最も高い解釈結果の候補に含まれる未確定語の数と、前記ＧＵＩ入力情報の数とが一致しているか否か判断する判断手段と、
前記尤度が最も高い解釈結果の候補に含まれる未確定語の数と、前記ＧＵＩ入力情報の数とが一致していないと判断された場合、解釈結果の候補に含まれる未確定語の数と、前記ＧＵＩ入力情報の数とが一致している別の解釈結果の候補を検索する検索手段と、
検索された解釈結果の候補に含まれる未確定語を前記ＧＵＩ入力情報で置換えた情報を、認識結果として出力する出力手段とを有する情報処理装置。
【請求項５】
請求項１乃至請求項３に記載の情報処理方法をコンピュータに実行させるためのプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、マルチモーダル・ユーザインタフェースに係る技術に関するものである。
【背景技術】
【０００２】
ＧＵＩ入力や音声入力といった複数種類の入力手段から、ユーザの所望の入力手段をもって情報の入力を可能にするマルチモーダル・ユーザインタフェースは、ユーザにとって利便性が高いものである。特に、複数種類の入力手段を同時に用いて入力を行った場合の利便性は高く、例えば音声で「これをここに移動」等の指示を発声しながら、ＧＵＩで「これ」に対応する対象と、「ここ」に対応する対象をクリックする等の操作を行うことにより、コマンド等の専門的な言語に不慣れなユーザであっても自由に対象を操作することができる。このような操作を可能にするためには複数種類の入力手段による入力を統合するための処理が必要になる。
【０００３】
複数種の入力手段による入力を統合する処理の例として、マウスイベントの種類や速度に関する情報を用いる方法（特許文献１、特許文献２）のほか、音声認識結果に対して言語解析を行う方法（特許文献３）や文脈情報を用いる方法（特許文献４）、入力時刻の近いものをまとめて意味解析単位として出力する方法（特許文献５）、入力データの認識結果遅着を考慮した方法（特許文献６）、利用者の意図を統計的な学習により検出する方法（特許文献７、特許文献８）、文法解析方法を用いた方法（特許文献９）、言語解析を行って意味構造を用いる方法（特許文献１０）等や、マウスなどのポインティングデバイスによるポインティング入力をリストに登録し、音声入力データ中の指示表現の数とリスト中の数とを比較し、ポインティング入力の数が少ない場合に、次のポインティング入力を得ることで数を合わせ、音声入力とポインティング入力を統合する方法（特許文献１１）が開示されている。
【特許文献１】特開平８－２８６８８７号公報
【特許文献２】特開平９－８１３６４号公報
【特許文献３】特許第２９９３８７２号公報
【特許文献４】特許第３３７５４４９号公報

【特許文献５】特許第３３６３２８３号公報
【特許文献６】特開平１０－１９８５４４号公報
【特許文献７】特開平１１－２８８３４２号公報
【特許文献８】特開２００１－１００８７８号公報
【特許文献９】特開平６－２８２５６９号公報
【特許文献１０】特開２０００－２３１４２７号公報
【特許文献１１】特開平７－１１０７３４号公報
【発明の開示】
【発明が解決しようとする課題】
【０００４】
上記従来例では、各入力の入力時刻や入力順序を考慮しているが、１つの入力結果に対する複数の候補を解析するには複雑な処理を行わなければならない。また、音声入力を正確に認識できることを前提としているが、現在の音声認識技術では１００％正しく認識するのは困難である。そのため誤認識への対応が重要となるが、上記従来例には誤認識を起こした場合の対応や誤認識率を下げることについて言及していない。
【０００５】
特許文献１１では、音声入力データ中の指示入力の数に対してポインティング入力の数が足りない場合に次のポインティング入力を待って統合する技術が記載されているが、上述したように基本的に音声入力データ中の指示入力の数が正確に認識できることを前提としており、誤認識に関する記載はなく、また誤認識率を下げることついては記載されていない。特許文献１１は、ポインティング入力の数が音声入力データ中の指示入力の数よりも多い場合には、エラー処理を行い入力をやり直す構成となっているが、入力をやり直すことはユーザにとって負担となるため、このような事態を減らす技術が重要となる。
【０００６】
本発明は、このような事情を鑑みてなされたものであり、少なくとも２種類の入力手段からの入力が意図する指示内容の認識精度を向上することを目的とする。
【課題を解決するための手段】
【０００７】
上記課題を解決するために、本発明の情報処理方法は、音声認識された情報とＧＵＩ入力された情報とを統合した認識結果を出力する情報処理装置の情報処理方法であって、音声情報を受信する音声情報受信工程と、前記音声情報を受信している時間内にユーザに操作された１または複数のボタンに対応するＧＵＩ入力情報を受信するＧＵＩ情報受信工程と、前記音声情報を音声認識し、尤度が高い順に複数の解釈結果の候補を取得する音声認識工程と、前記尤度が最も高い解釈結果の候補に含まれる未確定語の数と、前記ＧＵＩ入力情報の数とが一致しているか否か判断する判断工程と、前記尤度が最も高い解釈結果の候補に含まれる未確定語の数と、前記ＧＵＩ入力情報の数とが一致していないと判断された場合、解釈結果の候補に含まれる未確定語の数と、前記ＧＵＩ入力情報の数とが一致している別の解釈結果の候補を検索する検索工程と、検索された解釈結果の候補に含まれる未確定語を前記ＧＵＩ入力情報で置換えた情報を、認識結果として出力する出力工程とを有することを特徴とする。
【発明の効果】
【０００９】
以上説明したように、本発明によれば、少なくとも２種類の入力手段からの入力が意図する指示内容の認識精度を向上することができる。
【発明を実施するための最良の形態】
【００１０】
以下、図面を参照して、本発明に係るマルチモーダル入力方法の好適な実施例ついて説明する。
【実施例１】
【００１１】

図１は、本発明の実施例１におけるシステムの基本構成を示す図である。ここでは、音声入力とＧＵＩ入力を受け付けるシステムを例にあげて説明する。ＧＵＩ入力部１０１、ＧＵＩ入力解釈部１０２、音声入力部１０３、音声認識・解釈部１０４、マルチモーダル入力統合部１０５、記憶部１０６、マークアップ解釈部１０７、制御部１０８、音声合成部１０９、表示部１１０、通信部１１１から構成される。

【００１２】

ＧＵＩ入力部１０１はＧＵＩ上で指示を与えるボタン群やキーボード、マウス、タッチパネル、ペン、タブレット等から構成され、各種の指示を本装置に入力するための入力インタフェースとして機能する。本システムはこれら入力部から入力情報を受信する。ＧＵＩ入力解釈部１０２は、ＧＵＩ入力部１０１から入力された情報に対して解釈を行う。上記解釈については、例えば手書き認識技術等、公知の技術を利用する。

【００１３】

音声入力部１０３はマイクロフォンやＡ／Ｄ変換器等により構成されており、ユーザの音声を入力する。音声認識・解釈部１０４は音声入力部１０３より入力された音声に対して音声認識を行う。上記音声認識技術については公知の技術を利用する。マルチモーダル入力統合部１０５は、ＧＵＩ入力解釈部１０２、音声認識・解釈部１０４で解釈された情報を統合する。

【００１４】

記憶部１０６は、各種の情報を保存するためのハードディスクドライブ装置や、システムに各種の情報を提供するためのＣＤ－ＲＯＭやＤＶＤ－ＲＯＭ等の記憶媒体等により構成されている。またこのハードディスクドライブ装置や記憶媒体には、各種のアプリケーションプログラム、ユーザインタフェース制御プログラム、そして各プログラムを実行する際に必要な各種のデータ等が記憶されており、これらは後段の制御部１０８の制御により、本システムに読み込まれる。マークアップ解釈部１０７はマークアップで記述された文書を解釈する。制御部１０８はワークメモリやＣＰＵ、ＭＰＵ等により構成されており、記憶部１０５に記憶されたプログラムやデータを読み出して各種の処理を実行する。またＧＵＩ入力解釈部、音声認識・解釈部、マルチモーダル入力統合部などの制御も実行する。

【００１５】

音声合成部１０９はスピーカやヘッドフォン、Ｄ／Ａ変換器等により構成されており、制御部１０８の制御により読み上げテキストから音声データを作成してＤ／Ａ変換し、音として外部に出力する処理を行う。上記音声合成技術については公知の技術を利用する。表示部１１０は液晶ディスプレイ等の表示装置から構成され、画像や文字等により構成される各種の情報を表示する。なお、表示部１１０としてタッチパネル式の表示装置を用いてもよく、その場合、表示部１１０はＧＵＩ入力部１０１としての機能（各種の指示を本システムに入力する機能）をも有することになる。通信部１１１は、インターネットやＬＡＮ等のネットワークを介して他の装置とのデータ通信を行うためのネットワークインタフェースである。

【００１６】

以下では、上記マルチモーダル入力統合部１０５における統合処理方法について図２のフローチャートを用いて説明する。ＧＵＩ入力解釈部１０２、音声認識・解釈部１０４で解釈された情報、つまり認識候補がマルチモーダル入力統合部１０５に渡されると、まず、ＧＵＩ入力のｉｎｔｅｇｒａｔｉｏｎ＝０の解釈結果を出力する（ステップＳ２０１）。上記ｉｎｔｅｇｒａｔｉｏｎは他の入力手段の入力結果と統合する必要があるか否かを示す情報であり、必要がある場合は“１”、必要がない場合は“０”が、ＧＵＩ入力解釈部１０２、音声認識・解釈部１０４において入力される。他の入力手段の入力結果と統合する必要があるか否かを判別する方法については、例えば、値を格納する場所が決まっているか否かを判別する等、公知の技術を利用する。

【００１７】

続いて、すべての音声認識・解釈結果においてｉｎｔｅｇｒａｔｉｏｎ＝０の場合（ス

テップＳ２０２にてＹＥＳ）、第１位の音声認識・解釈結果を出力して終了する（ステップＳ２０３）。音声認識・解釈結果にｉｎｔｅｇｒａｔｉｏｎ＝１の結果がある場合（ステップＳ２０２にてＮＯ）、音声入力時間Ｔ内のＧＵＩ入力の中でｉｎｔｅｇｒａｔｉｏｎ＝１の個数ＮＵＭをカウントする（ステップＳ２０４）。ここで、音声入力時間Ｔは、図３（縦軸：音声入力のパワー、横軸：時間）の３０１に示すように閾値ｔｈ以上のパワーが検出されている時間とする。あるいは、図３の３０２のように、閾値ｔｈ以上のパワーが検出されている時間の前後に任意の時間（例えば数秒）を足す等、音声入力部にて設定した値でもよい。図３の３０１の例では、ＧＵＩ入力がすべてｉｎｔｅｇｒａｔｉｏｎ＝１とするとＮＵＭ＝２となる。

【００１８】

ＮＵＭ＝０であれば（ステップＳ２０５にてＹＥＳ）、ｉｎｔｅｇｒａｔｉｏｎ＝０の音声認識・解釈結果が存在するかどうかをチェックする（ステップＳ２０６）。存在する場合は（ステップＳ２０６にてＹＥＳ）ｉｎｔｅｇｒａｔｉｏｎ＝０の音声認識・解釈結果の中で最も確信度の高い結果を出力して終了する（ステップＳ２０７）。存在しない場合は（ステップＳ２０６にてＮＯ）結果を統合できなかった旨のエラーを出力して終了する（ステップＳ２０８）。

【００１９】

ＮＵＭ＝０でなければ（ステップＳ２０５にてＮＯ）、Ｎに１を代入して（ステップＳ２０９）ステップＳ２１０へと進む。Ｎ位（最初は１位）の音声認識・解釈結果が存在しない場合は（ステップＳ２１０にてＮＯ）、結果を統合できなかった旨のエラーを出力して終了する（ステップ２０８）。存在する場合は（ステップＳ２１０にてＹＥＳ）、ステップＳ２１１へと進む。ステップＳ２１１ではＮ位の音声認識・解釈結果のｉｎｔｅｇｒａｔｉｏｎが１か（統合する必要があるか否か）を判別する（ステップＳ２１１）。統合する必要がない場合は（ステップＳ２１１にてＮＯ）、Ｎに１を足し（ステップＳ２１２）、ステップＳ２１０へ進む。統合する必要がある場合は（ステップＳ２１１にてＹＥＳ）、“？”の数（図４の例では４０１～４０３のテーブルにおけるｕｎｋｎｏｗｎの値）が前述のＮＵＭと同じかどうかを判別する（ステップＳ２１３）。同じ値でなければ（ステップＳ２１３にてＮＯ）、Ｎに１を足し（ステップＳ２１２）、ステップＳ２１０へ進む。同じ値の場合は（ステップＳ２１３にてＹＥＳ）、Ｎ位の音声認識・解釈結果とＧＵＩの解釈結果を統合して出力する（ステップＳ２１４）。統合の具体例としては、音声入力「ここから」とＧＵＩ入力「恵比寿」を統合し、「恵比寿から」という結果になる。

【００２０】

以下では、上記マルチモーダル入力統合の例を図を用いて示す。図４～図７では、音声入力とボタン入力がなされた場合の、解釈処理結果の流れを示している。

【００２１】

１つ目の例を、図４を用いて説明する。図４の４０１、４０２は音声入力に対する音声認識・解釈結果を表すテーブルであり、確信度第１位の結果を４０１、確信度第２位の結果を４０２に示している。また、４０３はボタン入力に対する解釈結果を表すテーブルであり、この例では確信度が第１位の結果のみの場合を示している。

【００２２】

以下、４０１～４０３に示す各テーブルの項目について説明する。“ｒａｎｋ”は確信度の順位（高いものから１位，２位，・・・）、“ｕｎｋｎｏｗｎ”は確定していない値（後述の“ｖａｌｕｅ＝？”）の数、“ｖａｌｕｅ”は解釈値、“ｔｉｍｅ（ｓｔａｒｔ）”は入力開始時刻、“ｔｉｍｅ（ｅｎｄ）”は入力終了時刻、“ｓｃｏｒｅ”は確信度、“ｉｎｔｅｇｒａｔｉｏｎ”は統合が必要か否か（必要：１，不要：０）を表す。ＧＵＩ入力解釈部１０２、音声認識・解釈部１０４にて解釈された結果が上記テーブルに入力され、マルチモーダル入力統合部１０５へと渡される。上記テーブルがＸＭＬで記述されている場合はマークアップ解釈部１０７にて解釈される。

【００２３】

マルチモーダル入力統合部１０５では、前述のフローチャート図２に示す流れで処理を

行う。ＧＵＩ入力解釈処理結果である４０２は音声入力時間Ｔ（０２：１０：００～０２：１２：００）内に入力されたものであり、ｉｎｔｅｇｒａｔｉｏｎ＝１であるのでステップ２０４にてＮＵＭ＝１とし、続いてＮ＝１とする（ステップＳ２０９）。第１位の音声認識・解釈結果（４０１）が存在するので（ステップＳ２１０にてＹＥＳ）、ステップＳ２１１へと進む。続いてｉｎｔｅｇｒａｔｉｏｎ＝１（ステップＳ２１１にてＹＥＳ）、ｕｎｋｎｏｗｎ＝１＝ＮＵＭであるので（ステップＳ２１３にてＹＥＳ）、ステップＳ２１４へと進む。ステップＳ２１４では、音声入力の解釈結果「東京からここまで」とボタン入力の解釈結果「恵比寿」を統合し、「東京から恵比寿まで」を出力する。

【００２４】

同様に、図５の例では、５０３と５０４よりＮＵＭ＝２（ステップＳ２０４）である。第１位の音声認識・解釈結果（５０１）はｕｎｋｎｏｗｎ＝１であり、ＮＵＭ＝２とは異なるので（ステップＳ２１３にてＮＯ）、続いて第２位の音声認識・解釈結果（５０２）を調べる。５０２の結果はｕｎｋｎｏｗｎ＝２＝ＮＵＭであるので（ステップＳ２１３にてＹＥＳ）、ステップＳ２１４にて、音声入力の解釈結果「ここからここまで」とボタン入力の解釈結果「恵比寿」「横浜」を統合し、「恵比寿から横浜まで」を出力する。

【００２５】

図６の例では、６０４はｉｎｔｅｇｒａｔｉｏｎ＝０であるのでボタン入力の解釈結果「１」を出力する。また、６０３より、ＮＵＭ＝１（ステップＳ２０４）である。第１位の音声認識・解釈結果（６０１）はｕｎｋｎｏｗｎ＝１＝ＮＵＭであるので（ステップＳ２１３にてＹＥＳ）、ステップＳ２１４にて、音声入力の解釈結果「東京からここまで」とボタン入力の解釈結果「恵比寿」を統合して「東京から恵比寿まで」を出力する。

【００２６】

図７の例では、７０３はｉｎｔｅｇｒａｔｉｏｎ＝０であるのでボタン入力の解釈結果「１」を出力し、ＮＵＭ＝０とする（ステップＳ２０４）。音声認識・解釈結果７０１、７０２にｉｎｔｅｇｒａｔｉｏｎ＝０の結果が存在しないので（ステップＳ２０６にてＮＯ）、結果を統合できなかった旨のエラーを出力して終了する（ステップＳ２０８）。

【００２７】

以上のように、実施例１によれば、音声入力時間中のボタン入力の個数情報を利用して音声認識・解釈結果を選択することで、音声認識処理により生ずる候補に対して優先度をつけることができ、認識の精度を向上することができる。その結果、複数の候補から正しい認識結果が出力される可能性が高くなり、ユーザの再入力の手間を省くなどの効果が生まれる。

【実施例２】

【００２８】

続いて、本発明に係る情報処理システムの実施例２について説明する。前述した実施例１では、ＧＵＩ入力がボタン入力で認識率１００％である場合の例を示した。しかしながら、実際のマルチモーダル・ユーザインタフェースでは，統合する入力の解釈の確信度がいずれも１００％でない場合がある。このような場合は、第１位の解釈結果から順番に第１の実施形態と同様の処理を行えばよい。

【００２９】

本実施例では、音声入力とペン入力がなされた場合の例について図８～図１０を用いて説明する。図８に示すテーブル８０１～８０４の各項目は、前述の図４～図７に示した各テーブルの項目と同様であり、８０１は第１位の音声認識・解釈結果、８０２は第２位の音声認識・解釈結果、８０３は第１位のＧＵＩ入力解釈結果、８０４は第２位のＧＵＩ入力解釈結果である。

【００３０】

ＧＵＩ入力解釈結果の第１位の解釈結果から順番に第１の実施形態と同様の処理を行う。まず、第１位のＧＵＩ入力解釈結果８０３は音声入力時間Ｔ（０２：１０：００～０２：１２：００）内に入力されたものであり、ｉｎｔｅｇｒａｔｉｏｎ＝１である。また８０３よりｖａｌｕｅの数は１つであるので、ステップ２０４にてＮＵＭ＝１とし、続いて

Ｎ＝１とする（ステップＳ２０９）。第１位の音声認識・解釈結果（８０１）が存在するので（ステップＳ２１０にてＹＥＳ）、ステップＳ２１１へと進む。続いてｉｎｔｅｇｒａｔｉｏｎ＝１（ステップＳ２１１にてＹＥＳ）、ｕｎｋｎｏｗｎ＝１＝ＮＵＭであるので（ステップＳ２１３にてＹＥＳ）、ステップＳ２１４へと進む。ステップＳ２１４では、音声入力の解釈結果「ここ」とペン入力の解釈結果「恵比寿」を統合し、「恵比寿」を出力する。

【００３１】

図９の例では、まず第１位のＧＵＩ入力解釈結果９０３より、ＮＵＭ＝１（ステップＳ２０４）である。第１位の音声認識・解釈結果（９０１）はｕｎｋｎｏｗｎ＝１以上であり、ＮＵＭと一致するので（ステップＳ２１３にてＹＥＳ）、ステップＳ２１４にて、音声入力の解釈結果「このへん」とペン入力の解釈結果「恵比寿」を統合して「恵比寿」を出力する。

【００３２】

また、図１０では、ＧＵＩ入力としてペン入力とボタン入力の両方がなされた場合の例を示している。１００５はｉｎｔｅｇｒａｔｉｏｎ＝０であるのでボタン入力の解釈結果「１」を出力する。また、第１位のＧＵＩ入力解釈結果１００３より、ＮＵＭ＝１（ステップＳ２０４）である。第１位の音声認識・解釈結果（１００１）はｕｎｋｎｏｗｎ＝２以上であり、ＮＵＭとは異なるので（ステップＳ２１３にてＮＯ）、続いて第２位の音声認識・解釈結果（１００２）を調べる。第２位の音声認識・解釈結果（１００２）は、ｕｎｋｎｏｗｎ＝３であり、ＮＵＭとは異なるので（ステップＳ２１３にてＮＯ）統合できない。次に第２位のＧＵＩ入力解釈結果１００４より、ＮＵＭ＝２（ステップＳ２０４）とする。第１位の音声認識・解釈結果（１００１）はｕｎｋｎｏｗｎ＝２以上であり、ＮＵＭと一致するので（ステップＳ２１３にてＹＥＳ）、ステップＳ２１４にて、音声入力の解釈結果「これらを」とペン入力の解釈結果「Ａ，Ｂ」を統合して「Ａ，Ｂを」を出力する。

【００３３】

以上のように、実施例２によれば、統合する入力の解釈の確信度がいずれも１００％でない場合においても、音声入力時間中のＧＵＩ入力個数の情報を利用して音声認識・解釈結果を選択することで、音声認識結果の解釈の精度を向上することができる。

【実施例３】

【００３４】

上記実施例では、ＧＵＩ入力を受け付ける例をあげて説明したが、本発明はこれに限定されるものではなく、キーボードやテンキーなどの物理的なキー入力を受け付ける構成としてもかまわない。ここでは、テンキーと音声入力によって操作可能な複写機を例にあげて説明する。複写機における各指示コマンドが以下に示すようにテンキーの各キーに割り当てられていることを前提とする。キー１：用紙選択、キー２：枚数（部数）、キー３：倍率、キー４：濃さ、キー５：両面、キー６：ソータ、キー７：ステイプルソート、キー８：応用モード。

【００３５】

ユーザは、１０ページからなるＡ５サイズの資料を左上にステイプルしたものを５部コピーしたい場合、キー１を押して「Ａ５」、キー２を押して「５部」、キー７を押して「左上」と発声することで設定することができるが、この操作に慣れてきた場合は、１つ１つを入力するよりも「Ａ５、５部、左上」のように連続発声できたほうが効率的に作業を進めることができる。しかしながら現在の音声認識の精度は１００％ではないため、『５枚、左上』や『Ａ５、５部、左上、濃く』等の認識誤りが発生し、誤った認識候補が発生する可能性がある。

【００３６】

本実施例ではこのような場面において、音声入力とキー入力をキー入力の個数を用いて統合する。ユーザは、キー１、２、７を押しながら「Ａ５、５部、左上」と発声する。キーの押し方は３つ同時に押していても良いし、連続的に押しても構わない。ここでキー入

力の入力数は３である。音声入力の認識候補が１位「５部、左上」、２位「Ａ５、５部、左上」、３位「Ａ５、５部、左上、濃く」、４位「Ａ２、５０部、左上」であった場合、これとキー入力数３を統合することで、数が一致しない「Ａ５、５部」、「Ａ５、５部、左上、濃く」が除去されるかもしくは、数が一致する「Ａ５、５部、左上」、「Ａ２、５０部、左上」が選択されることにより、１位「Ａ５、５部、左上」、２位「Ａ２、５０部、左上」となって認識候補が絞られ、ここから認識尤度の最も高い認識候補が認識結果として選ばれることによって、ユーザが発声した「Ａ５、５部、左上」が正しく認識されることとなる。

【００３７】

この他、携帯電話に表示された番号付きのメニューに対して携帯電話のボタンで各メニューを指定できるような場面を想定し、例えばボタンで、３番、５番を押しながら、「これとこれのヘルプがみたい」と発声する場合などにも本発明が適用できることは言うまでもない。

【実施例４】

【００３８】

上記実施例では、ＧＵＩや物理的なキー入力数の情報を用いて、音声認識・解釈結果の第１位～第Ｎ位の候補の中から適切なものを選択する例を示したが、どの入力手段の入力個数情報をどの入力手段の入力情報に適応するかは上記例に限られない。例えば、音声認識・解釈結果より入力個数（前述の例でいえばｕｎｋｎｏｗｎの値）を判別して手書き文字入力の認識結果の第１位～第Ｎ位の候補の中から適切なものを選択してもよい。一般にボタン入力は音声入力（認識）に比べて確実性が高いことから、ボタン入力の個数情報を他方のモダリティに適用する等、任意に決めてもよいが、いずれの入力手段も曖昧性をもつような場合、どの入力手段の入力個数情報をどの入力手段の入力情報に適用すればいいのか、一意に決めることは適切でない。このような場合は、第１位の確信度と第２位以下の確信度の差が大きい方のモダリティを選択する等、確信度の結果から毎回決めてもよい。

【００３９】

図１１を用いてそのような例について説明する。図１１は音声入力で「ここ」と発声しながら、ペン入力で「恵比寿」に丸をつけた例を示している。ユーザは「恵比寿」にのみ丸をつけるつもりが「渋谷」にも少しかかってしまった状態である。音声入力の解釈処理により第１位が「ここ」、第２位が「ここととここ」であり、それぞれＳＣＯＲＥが９０、５５である。ペン入力の解釈処理により、第１位が「渋谷、恵比寿」、第２位が「恵比寿」であり、ＳＣＯＲＥが９５、９０である。第１位と第２位の確信度の差が音声入力の方が大きく、第１位が正解である確率が高いことから、入力数が正しい可能性も高いといえる。つまり入力数の確からしさの値が高いため、ここでは入力数は音声入力の方を信頼する。入力数の確からしさの値は、上述したように第１位と第２位の差から求めても良いし、例えば、確信度が上位の認識候補が共通して含む入力数がより多いものが確からしさの値が高くなるよう求めてもよい。例えば、１位と２位の差が大きくても１位と２位で入力数が異なるものは確からしさの値を低くし、１位から４位までの確信度の差が少なくても全て入力数が等しい場合は入力数の確からしさの値が高くなるように求めても良い。また、上述したように、「一般にボタン入力は音声入力（認識）に比べて確実性が高い」等の情報を加味して求めても良い。音声入力の第１位の入力数は１であり、ＧＵＩ入力のうち入力数が１である第２位が選ばれ、「ここ」と「恵比寿」が統合されて統合結果が「恵比寿」となる。

【実施例５】

【００４０】

上記実施例では、入力数を取得する入力手段が１種類の場合について説明してきたが、本発明はこれに限られる物ではない。例えばＧＵＩ入力とキー入力を併せ持つ複写機においては、これら両方の入力からの入力数を考慮することも可能である。音声入力で「Ｂ５、片面から両面をこの枚数で」と入力しながら、ＧＵＩ入力で用紙サイズと両面の指示を

選択し、キー入力で１０と入力した場合は、音声入力に含まれる入力数は３であり、ＧＵＩ入力とキー入力をあわせた入力数が３となり、これらの数が一致しないものを音声入力の認識結果から除外することで認識精度を向上することが可能となる。

【実施例６】

【００４１】

上記実施例では、音声入力と他の入力手段を統合する例を挙げて説明してきたが、本発明はこれに限定されるものではない。例えば、ジェスチャ入力と視線入力でのマルチモーダル入力に適用した場合にも適用可能である。ここでは、視線入力で操作対象を指定し、ジェスチャ入力でその操作対象に指示を与えるタスクを考えてみる。視線入力でオブジェクトを指定する場合は、そのオブジェクトを長く見ていた場合は選択されたとみなすなどの処理によって指定するが、それが正しく認識されず、操作対象の認識候補が複数得られる場合がある。それに対してジェスチャ入力により２つの指示が入力された場合は、操作対象が２つである可能性が高いため、視線入力の認識候補のうち、操作対象が２つの候補以外を除外することで認識精度を向上することができる。

【００４２】

なお、本発明の目的は、前述した実施例の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ（またはＣＰＵやＭＰＵ）が記憶媒体に格納されたプログラムコードを読出し実行することによっても、達成されることは言うまでもない。

【００４３】

この場合、記憶媒体から読み出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記憶した記憶媒体は本発明を構成することになる。

【００４４】

プログラムコードを供給するための記憶媒体としては、例えば、フレキシブルディスク，ハードディスク、光ディスク、光磁気ディスク、ＣＤ－ＲＯＭ、ＣＤ－Ｒ、磁気テープ、不揮発性のメモリカード、ＲＯＭなどを用いることができる。

【００４５】

また、コンピュータが読み出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているＯＳ（オペレーティングシステム）などが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【００４６】

さらに、記憶媒体から読出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるＣＰＵなどが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【図面の簡単な説明】

【００４７】

【図１】本発明の実施例１における情報処理システムの基本構成を示す図である。

【図２】本発明の実施例１に係る情報処理システムにおけるマルチモーダル入力統合部の処理の流れを説明するためのフローチャートである。

【図３】本発明の実施例１に係る入力の例を示す図である。

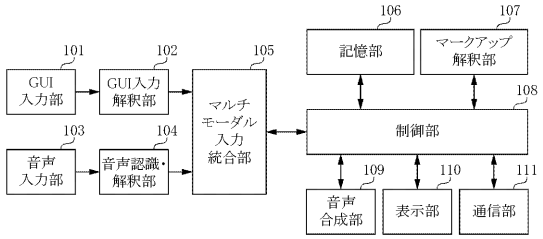
【図４】本発明の実施例１に係る情報処理システムにおけるマルチモーダル統合入力統合の例を示す図である。

【図５】本発明の実施例１に係る情報処理システムにおけるマルチモーダル統合入力統合の例を示す図である。

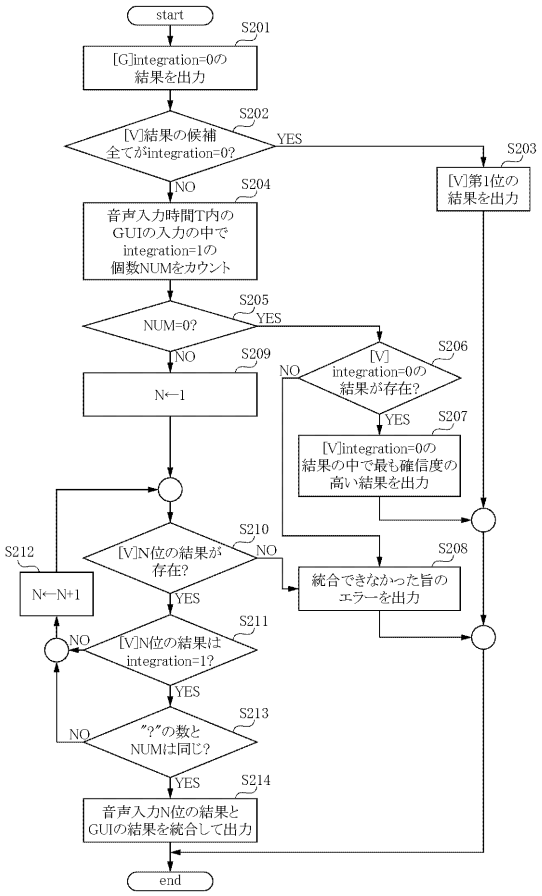
【図６】本発明の実施例１に係る情報処理システムにおけるマルチモーダル統合入力統合

の例を示す図である。
【図７】本発明の実施例１に係る情報処理システムにおけるマルチモーダル統合入力統合の例を示す図である。
【図８】本発明の実施例２に係る情報処理システムにおけるマルチモーダル統合入力統合の例を示す図である。
【図９】本発明の実施例２に係る情報処理システムにおけるマルチモーダル統合入力統合の例を示す図である。
【図１０】本発明の実施例２に係る情報処理システムにおけるマルチモーダル統合入力統合の例を示す図である。
【図１１】本発明の実施例４に係る情報処理システムにおけるマルチモーダル統合入力統合の例を示す図である。
【符号の説明】
【００４８】
１０１　ＧＵＩ入力部
１０２　ＧＵＩ入力解釈部
１０３　音声入力部
１０４　音声認識・解釈部
１０５　マルチモーダル入力統合部
１０６　記憶部
１０７　マークアップ解釈部
１０８　制御部
１０９　音声合成部
１１０　表示部
１１１　通信部

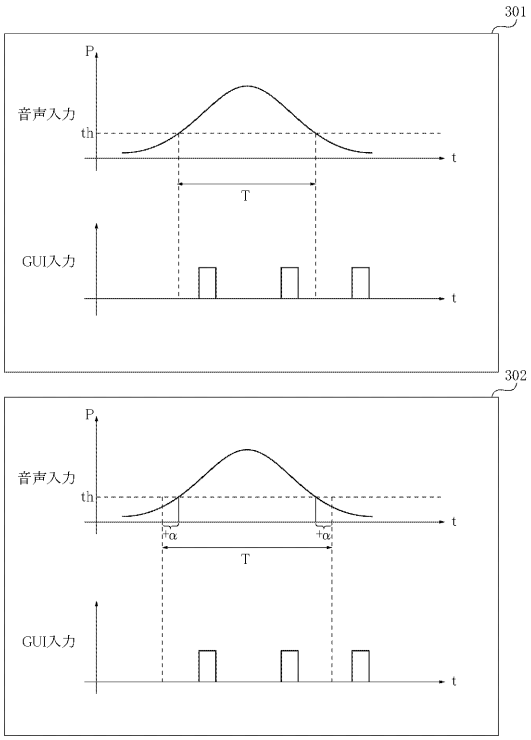
【図１】



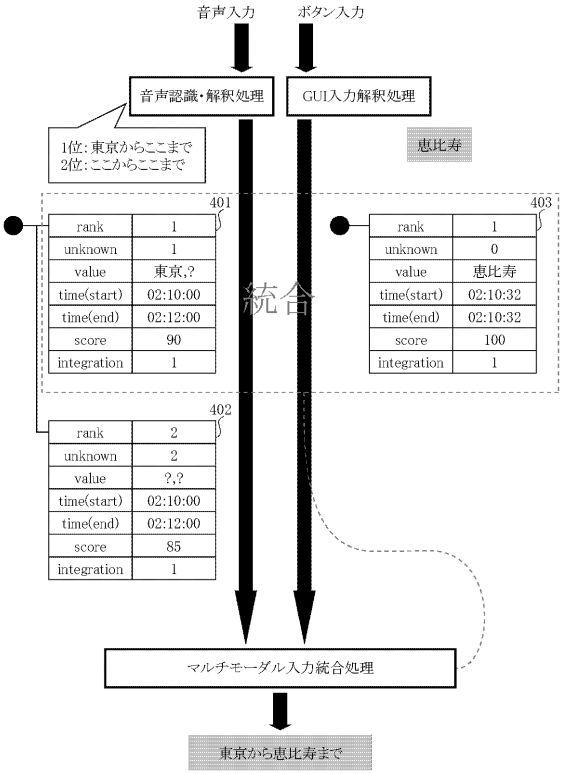
【図２】



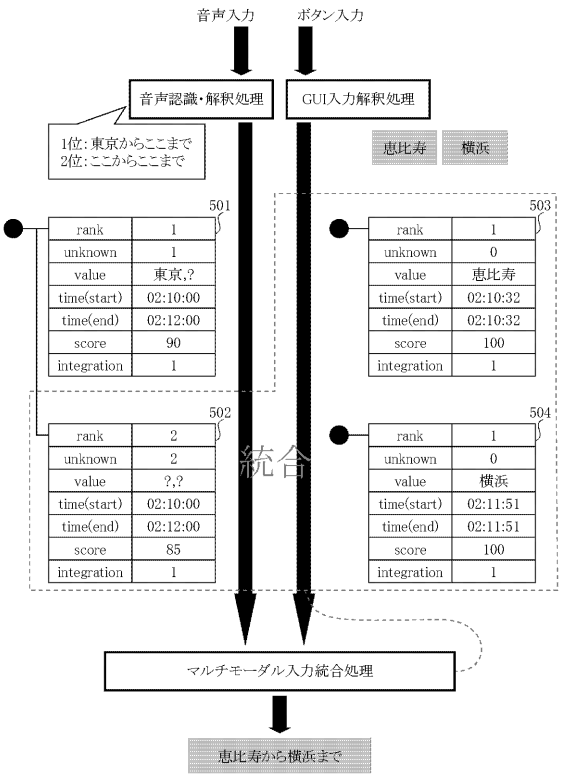
【図３】


301
P
音声入力
th
t
T
GUI入力
t
302
P
音声入力
th
t
+α
+α
T
GUI入力
t

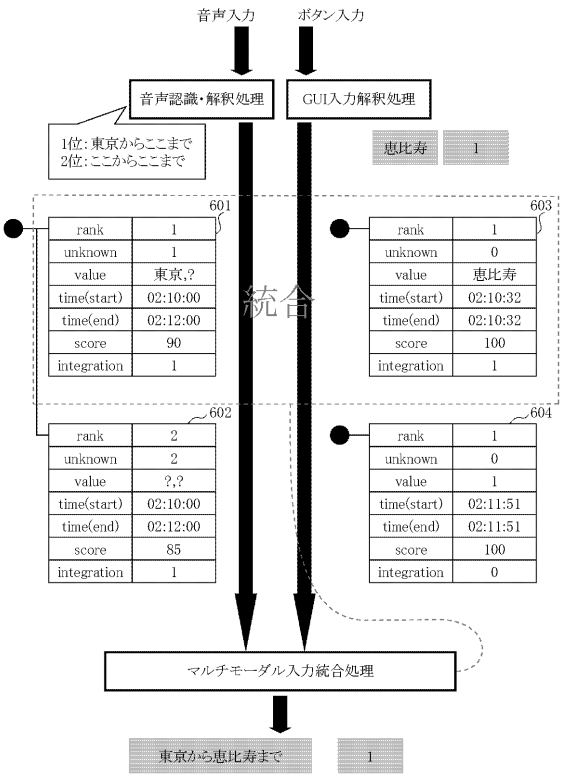
【図４】


音声入力
ボタン入力
音声認識・解釈処理
GUI入力解釈処理
1位:東京からここまで
2位:ここからここまで
恵比寿
401
rank 1
unknown 1
value 東京,?
time(start) 02:10:00
time(end) 02:12:00
score 90
integration 1
統合
403
rank 1
unknown 0
value 恵比寿
time(start) 02:10:32
time(end) 02:10:32
score 100
integration 1
402
rank 2
unknown 2
value ?,?
time(start) 02:10:00
time(end) 02:12:00
score 85
integration 1
マルチモーダル入力統合処理
東京から恵比寿まで

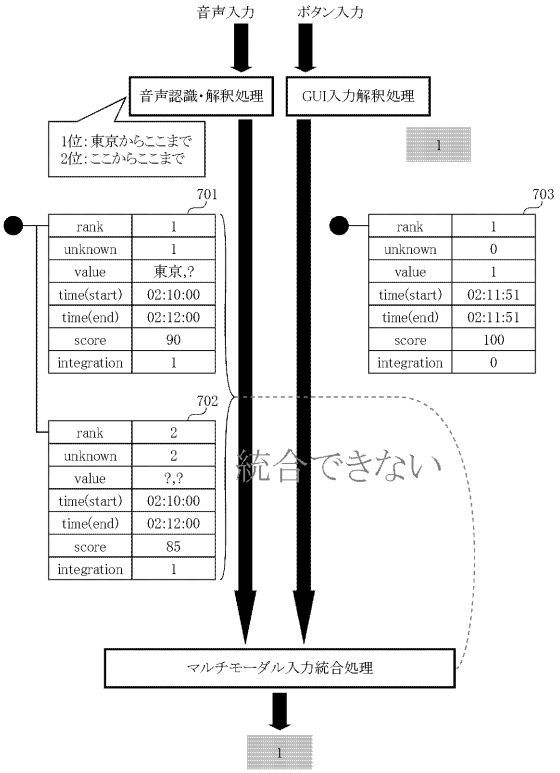
【図５】


音声入力
ボタン入力
音声認識・解釈処理
GUI入力解釈処理
1位:東京からここまで
2位:ここからここまで
恵比寿
横浜
501
rank 1
unknown 1
value 東京,?
time(start) 02:10:00
time(end) 02:12:00
score 90
integration 1
503
rank 1
unknown 0
value 恵比寿
time(start) 02:10:32
time(end) 02:10:32
score 100
integration 1
502
rank 2
unknown 2
value ?,?
time(start) 02:10:00
time(end) 02:12:00
score 85
integration 1
統合
504
rank 1
unknown 0
value 横浜
time(start) 02:11:51
time(end) 02:11:51
score 100
integration 1
マルチモーダル入力統合処理
恵比寿から横浜まで

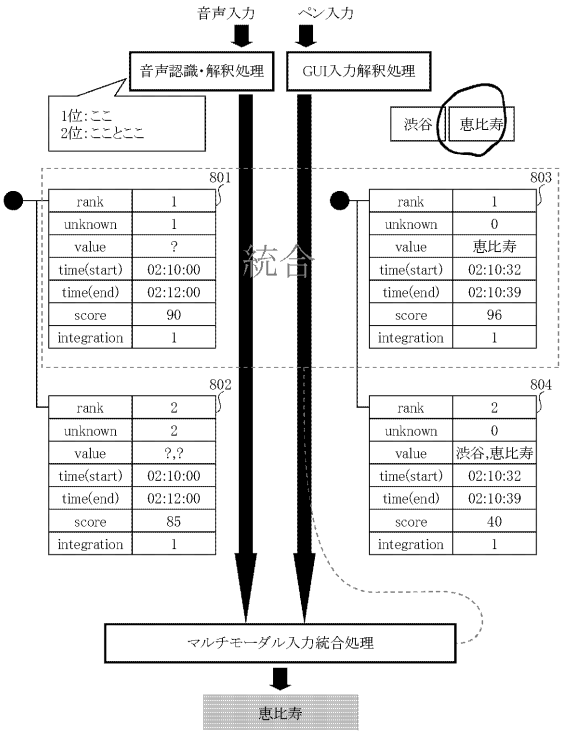
【図６】


音声入力
ボタン入力
音声認識・解釈処理
GUI入力解釈処理
1位:東京からここまで
2位:ここからここまで
恵比寿
1
601
rank 1
unknown 1
value 東京,?
time(start) 02:10:00
time(end) 02:12:00
score 90
integration 1
統合
603
rank 1
unknown 0
value 恵比寿
time(start) 02:10:32
time(end) 02:10:32
score 100
integration 1
602
rank 2
unknown 2
value ?,?
time(start) 02:10:00
time(end) 02:12:00
score 85
integration 1
604
rank 1
unknown 0
value 1
time(start) 02:11:51
time(end) 02:11:51
score 100
integration 0
マルチモーダル入力統合処理
東京から恵比寿まで
1

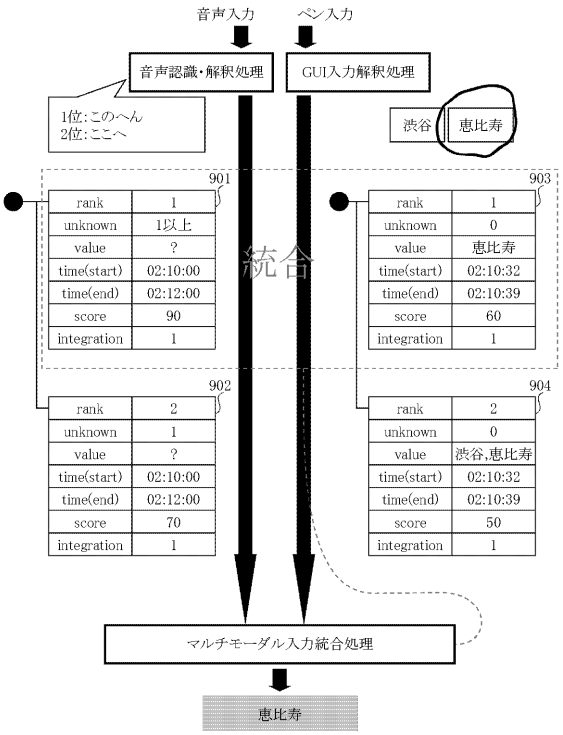
【図７】


音声入力
ボタン入力
音声認識・解釈処理
GUI入力解釈処理
1位:東京からここまで
2位:ここからここまで
1
701
rank 1
unknown 1
value 東京,?
time(start) 02:10:00
time(end) 02:12:00
score 90
integration 1
703
rank 1
unknown 0
value 1
time(start) 02:11:51
time(end) 02:11:51
score 100
integration 0
702
rank 2
unknown 2
value ?,?
time(start) 02:10:00
time(end) 02:12:00
score 85
integration 1
統合できない
マルチモーダル入力統合処理
1

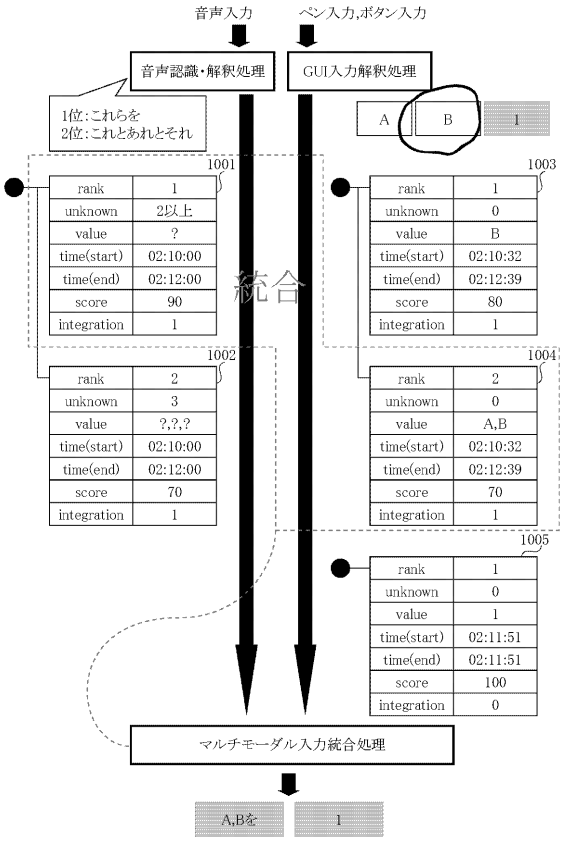
【図８】


音声入力
ペン入力
音声認識・解釈処理
GUI入力解釈処理
1位:ここ
2位:ここととここ
渋谷
恵比寿
801
rank 1
unknown 1
value ?
time(start) 02:10:00
time(end) 02:12:00
score 90
integration 1
803
rank 1
unknown 0
value 恵比寿
time(start) 02:10:32
time(end) 02:10:39
score 96
integration 1
統合
802
rank 2
unknown 2
value ?,?
time(start) 02:10:00
time(end) 02:12:00
score 85
integration 1
804
rank 2
unknown 0
value 渋谷,恵比寿
time(start) 02:10:32
time(end) 02:10:39
score 40
integration 1
マルチモーダル入力統合処理
恵比寿

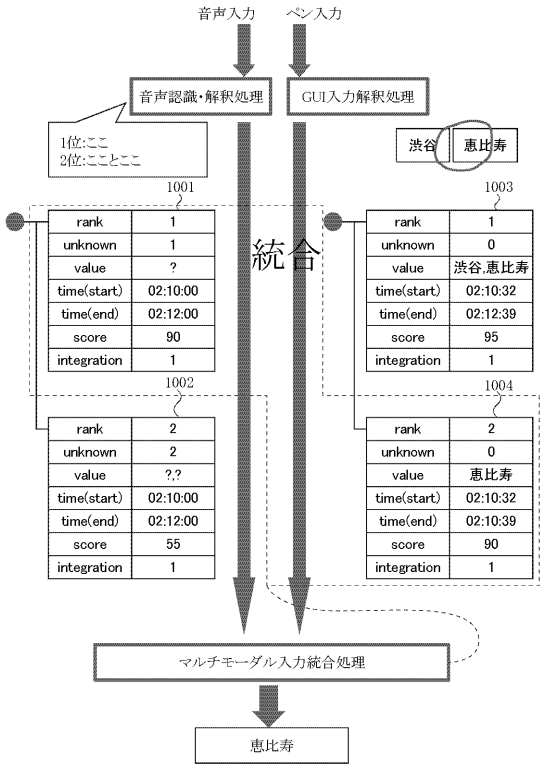
【図９】


音声入力
ペン入力
音声認識・解釈処理
GUI入力解釈処理
1位:このへん
2位:ここへ
渋谷
恵比寿
901
rank 1
unknown 1以上
value ?
time(start) 02:10:00
time(end) 02:12:00
score 90
integration 1
903
rank 1
unknown 0
value 恵比寿
time(start) 02:10:32
time(end) 02:10:39
score 60
integration 1
統合
902
rank 2
unknown 1
value ?
time(start) 02:10:00
time(end) 02:12:00
score 70
integration 1
904
rank 2
unknown 0
value 渋谷,恵比寿
time(start) 02:10:32
time(end) 02:10:39
score 50
integration 1
マルチモーダル入力統合処理
恵比寿

【図１０】


音声入力
ペン入力,ボタン入力
音声認識・解釈処理
GUI入力解釈処理
1位:これらを
2位:これとあれとそれ
A
B
1
1001
rank 1
unknown 2以上
value ?
time(start) 02:10:00
time(end) 02:12:00
score 90
integration 1
1003
rank 1
unknown 0
value B
time(start) 02:10:32
time(end) 02:12:39
score 80
integration 1
統合
1002
rank 2
unknown 3
value ?,?,?
time(start) 02:10:00
time(end) 02:12:00
score 70
integration 1
1004
rank 2
unknown 0
value A,B
time(start) 02:10:32
time(end) 02:12:39
score 70
integration 1
1005
rank 1
unknown 0
value 1
time(start) 02:11:51
time(end) 02:11:51
score 100
integration 0
マルチモーダル入力統合処理
A,Bを
1

【図１１】


音声入力
ペン入力
音声認識・解釈処理
GUI入力解釈処理
1位:ここ
2位:ここととここ
渋谷
恵比寿
1001
1003
統合
rank 1
unknown 1
value ?
time(start) 02:10:00
time(end) 02:12:00
score 90
integration 1
rank 1
unknown 0
value 渋谷,恵比寿
time(start) 02:10:32
time(end) 02:12:39
score 95
integration 1
1002
1004
rank 2
unknown 2
value ?,?
time(start) 02:10:00
time(end) 02:12:00
score 55
integration 1
rank 2
unknown 0
value 恵比寿
time(start) 02:10:32
time(end) 02:10:39
score 90
integration 1
マルチモーダル入力統合処理
恵比寿

フロントページの続き

(56)参考文献 特開平０８－０６３３１９（ＪＰ，Ａ）
特開２００３－０８４９００（ＪＰ，Ａ）
特開２０００－２０９３７８（ＪＰ，Ａ）
特開２００１－２８２４１３（ＪＰ，Ａ）
特開２００３－１４０６８７（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　３／０１
Ｇ０６Ｆ　３／１６
Ｇ１０Ｌ　１１／００－２１／０６