

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4708913号
(P4708913)

(45) 発行日 平成23年6月22日(2011.6.22)

(24) 登録日 平成23年3月25日(2011.3.25)

(51) Int. Cl.	F I	
G06F 3/16 (2006.01)	G06F 3/16	320H
G06T 7/00 (2006.01)	G06T 7/00	P
G06K 9/00 (2006.01)	G06K 9/00	S
G10L 15/00 (2006.01)	G10L 15/00	200P
G10L 15/24 (2006.01)	G10L 15/24	Q
請求項の数 6 (全 21 頁) 最終頁に続く		

(21) 出願番号 特願2005-234032 (P2005-234032)
 (22) 出願日 平成17年8月12日(2005.8.12)
 (65) 公開番号 特開2007-48177 (P2007-48177A)
 (43) 公開日 平成19年2月22日(2007.2.22)
 審査請求日 平成20年7月15日(2008.7.15)

(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 100126240
 弁理士 阿部 琢磨
 (74) 代理人 100124442
 弁理士 黒岩 創吾
 (72) 発明者 廣田 誠
 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
 (72) 発明者 深田 俊明
 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

最終頁に続く

(54) 【発明の名称】 情報処理方法及び情報処理装置

(57) 【特許請求の範囲】

【請求項1】

与えられた対象を図形認識することで図形の候補を取得する図形認識手段と、
 入力された音声情報を音声認識し図形の候補を取得する音声認識手段と、
 前記図形認識手段で取得した図形の候補及び前記音声認識手段で取得した図形の候補に
 基づいて、図形を選択する選択手段と、
前記与えられた対象のうち、前記選択手段で選択した図形に対応する特徴量を抽出し、
 抽出した該特徴量に基づく図形を生成する生成手段とを備えたことを特徴とする情報処理
 装置。

【請求項2】

ペン入力を受け付ける受付手段と、
 前記ペン入力の速度を検出する第1の検出手段と、
 前記音声情報に含まれる雑音の度合を検出する第2の検出手段とを更に備え、
 前記与えられた対象とは前記受付手段で受け付けたペン入力のストロークであり、
 前記図形認識手段は、図形の候補及び各候補がそれぞれ対象とする図形である確信の度
 合を示す確信度を取得し、
 前記音声認識手段は、図形の候補及び各候補がそれぞれ対象とする図形である確信の度
 合を示す確信度を取得し、
 前記選択手段は、前記図形認識手段で取得した図形の候補の確信度に前記ペン入力の速
 度に応じた重みを付けた第1の重み付き確信度と、前記音声認識手段で取得した図形の候

補の確信度に前記雑音の度合に応じた重みを付けた第2の重み付き確信度とに基づいて、前記図形を選択することを特徴とする請求項1記載の情報処理装置。

【請求項3】

与えられた対象を文字認識することで文字の候補を取得する文字認識手段と、
入力された音声情報を音声認識し文字の読みの候補を取得する音声認識手段と、
前記文字認識手段で取得した文字の候補及び前記音声認識手段で取得した文字の読みの候補に基づいて、文字を選択する選択手段と、

前記選択手段で選択された文字を前記与えられた対象に対応する位置に表示するよう制御する表示制御手段と、

前記表示制御手段で表示された文字を確定する確定手段と、

前記確定手段で一つまたは複数の文字が確定された後、前記与えられた対象と前記音声情報を受信する受信手段と、

単語辞書に含まれる単語のうち、前記確定した一つまたは複数の文字と表記が部分一致する単語を検索する検索手段と、

前記検索手段で検索された一つまたは複数の単語において、前記確定された一つまたは複数の文字に後続する後続文字を取得する後続文字取得手段と、

前記部分一致する単語に基づいて、前記後続文字取得手段によって取得された後続文字に対し、該後続文字が認識結果に含まれやすいよう前記受信手段で受信した連続する座標情報の文字認識を制御する文字認識制御手段と、

前記部分一致する単語の読みに基づいて、前記後続文字が認識結果に含まれやすいよう前記受信手段で受信した音声情報の音声認識を制御する音声認識制御手段とを備えたことを特徴とする情報処理装置。

【請求項4】

与えられた対象を処理する情報処理装置の情報処理方法であって、

前記与えられた対象を図形認識することで図形の候補を取得する図形認識工程と、

入力された音声情報を音声認識し図形の候補を取得する音声認識工程と、

前記図形認識工程で取得した図形の候補及び前記音声認識工程で取得した図形の候補に基づいて、図形を選択する選択工程と、

前記与えられた対象のうち、前記選択工程で選択した図形に対応する特徴量を抽出し、抽出した該特徴量に基づく図形を生成する生成工程とを備えたことを特徴とする情報処理方法。

【請求項5】

与えられた対象を処理する情報処理装置の情報処理方法であって、

前記与えられた対象を文字認識することで文字の候補を取得する文字認識工程と、

入力された音声情報を音声認識し文字の読みの候補を取得する音声認識工程と、

前記文字認識工程で取得した文字の候補及び前記音声認識工程で取得した文字の読みの候補に基づいて、文字を選択する選択工程と、

前記選択工程で選択された文字を前記与えられた対象に対応する位置に表示するよう制御する表示制御工程と、

前記表示制御工程で表示された文字を確定する確定工程と、

前記確定工程で一つまたは複数の文字が確定された後、前記与えられた対象と前記音声情報を受信する受信工程と、

単語辞書に含まれる単語のうち、前記確定した一つまたは複数の文字と表記が部分一致する単語を検索する検索工程と、

前記検索工程で検索された一つまたは複数の単語において、前記確定された一つまたは複数の文字に後続する後続文字を取得する後続文字取得工程と、

前記部分一致する単語に基づいて、前記後続文字取得工程によって取得された後続文字に対し、該後続文字が認識結果に含まれやすいよう前記受信工程で受信した連続する座標情報の文字認識を制御する文字認識制御工程と、

前記部分一致する単語の読みに基づいて、前記後続文字が認識結果に含まれやすいよう

10

20

30

40

50

に前記受信工程で受信した音声情報の音声認識を制御する音声認識制御工程とを備えたことを特徴とする情報処理装置。

【請求項 6】

請求項 4 または 5 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制御プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、入力された文字や図形等の情報を処理する情報処理方法に関し、特に入力された文字や図形等の認識率を向上させる技術に関する。

10

【背景技術】

【0002】

タッチスクリーン、電子ホワイトボード、タッチパネル付きの PDA、タブレット PC など、ペンやスタイラスによる画面への直接入力可能な機器では、手書きによる文字や図形の入力が可能である。これにより、紙やホワイトボードに手書きする自然な感覚で入力ができる。これに手書き文字認識や図形認識の技術を統合すると、手書きによるストローク情報をテキストなどの形でデジタルデータ化でき、その後の様々な処理につなげることができる。

【0003】

一方、音声認識技術が発展し、また機器のハードウェア性能が向上したことにより、PC やワークステーションに限らず、さまざまなコンピュータ制御機器において、音声入力が可能になりつつある。

20

【0004】

そんな中、ペン入力単独、あるいは音声入力単独の入力ではなく、ペンと音声を組み合わせたマルチモーダルな入力により、操作を効率化しようとする技術も提案されている。特許文献 1 では、ペンによる手書き入力で図形を描き、同時に音声で図形の種類を発声することで、手書き図形を整形して表示するシステムを提案している。

【特許文献 1】特開平 06 - 085983 号公報

【発明の開示】

【発明が解決しようとする課題】

30

【0005】

手書き入力、音声入力とも、直感的な入力を可能にするというメリットがある反面、「誤認識」というデメリットがある。例えば、ユーザが手書きで図形を描き、図形認識処理によってこれを自動整形するシステムを考えてみる。ユーザが「四角形」を意図して描いたとしても、誤認識により「円」と認識されて円に整形表示される可能性がある。これに対し、特許文献 1 では、ユーザが、図形を描くと同時に音声でその図形の種類を発声し、その発声を音声認識することで図形の種類を特定し、その図形の種類に応じて入力図形を整形するようにしている。しかし、音声認識にも誤認識の問題があり、ユーザが「シカケイ」と発声しているのに、「三角形」と認識されたら、図形は三角形として整形表示され、ユーザの意図しない整形がなされてしまう。つまり、誤認識により目的を達成できないという問題があった。ここでは特に図形認識と音声認識の例を挙げて説明したが、文字認識、顔認識やジェスチャ認識などのパターン認識において同様の問題があった。本発明はこのような問題に鑑み、入力された文字や図形等の情報の認識精度を向上することを目的とする。

40

【課題を解決するための手段】

【0006】

上記目的を達成するために本発明は、与えられた対象を図形認識することで図形の候補を取得する図形認識手段と、入力された音声情報を音声認識し図形の候補を取得する音声認識手段と、前記図形認識手段で取得した図形の候補及び前記音声認識手段で取得した図形の候補に基づいて、図形を選択する選択手段と、前記与えられた対象のうち、前記選択

50

手段で選択した図形に対応する特徴量を抽出し、抽出した該特徴量に基づく図形を生成する生成手段とを備えたことを特徴とする。

【0007】

また本発明は、与えられた対象を文字認識することで文字の候補を取得する文字認識手段と、入力された音声情報を音声認識し文字の読みの候補を取得する音声認識手段と、前記文字認識手段で取得した文字の候補及び前記音声認識手段で取得した文字の読みの候補に基づいて、文字を選択する選択手段と、前記選択手段で選択された文字を前記与えられた対象に対応する位置に表示するよう制御する表示制御手段と、前記表示制御手段で表示された文字を確定する確定手段と、前記確定手段で一つまたは複数の文字が確定された後、前記与えられた対象と前記音声情報を受信する受信手段と、単語辞書に含まれる単語のうち、前記確定した一つまたは複数の文字と表記が部分一致する単語を検索する検索手段と、前記検索手段で検索された一つまたは複数の単語において、前記確定された一つまたは複数の文字に後続する後続文字を取得する後続文字取得手段と、前記部分一致する単語に基づいて、前記後続文字取得手段によって取得された後続文字に対し、該後続文字が認識結果に含まれやすいよう前記受信手段で受信した連続する座標情報の文字認識を制御する文字認識制御手段と、前記部分一致する単語の読みに基づいて、前記後続文字が認識結果に含まれやすいよう前記受信手段で受信した音声情報の音声認識を制御する音声認識制御手段とを備えたことを特徴とする。

10

【発明の効果】

【0008】

本発明によれば、入力された文字や図形等の情報の認識精度を向上することができる。

20

【発明を実施するための最良の形態】

【0009】

以下、図面を参照しながら本発明の好適な実施例について説明していく。

【実施例1】

【0010】

本実施例では、会議等の場面において使用されるような大型のタッチスクリーンを備えた情報処理装置に対してペンで図形を描き、同時にその図形の種類を発声し、入力図形を整形表示する例をあげて説明する。図1は、本実施例に係る情報処理装置におけるマルチモーダル入力処理のブロック図である。同図において、101は音声情報を受信する音声情報受信部である。102は、受信した音声情報を音声認識する音声認識部である。103は、ペン入力情報を受信するペン入力情報受信部である。104は、受け取った座標情報から図形を認識する図形認識部である。105は、音声認識結果と図形認識結果を総合評価して図形を選択する総合評価部である。106は、ペン入力情報受信部103で受け取ったペン入力情報から特徴量を抽出する特徴量抽出部である。107は生成部であり、特徴量抽出部106で抽出した特徴量に基づいて総合評価部で選択された図形を生成する。108は表示制御部であり、生成部107で生成された図形等を表示するよう制御する。

30

【0011】

図2は、本実施例に係る情報処理装置のハードウェア構成を示す構成図である。同図において、201はCPUであり、後述する情報処理装置の動作手順を実行するプログラムに従って、上記マルチモーダル入力処理を制御する。202はRAMであり、上記プログラムの動作に必要な記憶領域を提供する。203はROMであり、上記プログラムなどを保持する。204はペンや指での入力可能なタッチスクリーンである。ペン入力情報受信部103は、タッチスクリーン204で入力されたペン入力情報を受信する。また表示制御部108は、タッチスクリーン204に前記図形等を表示するよう制御する。205はハードディスクであり、上記プログラムなどを保持する。本実施例で実行されるプログラムは、ROM203に保持されたプログラムでもHDD205に保持されたプログラムでもどちらでも構わない。またプログラムをROM203、HDD205で分割して保持し、両方を参照して実行するようにしても構わない。206はA/Dコンバータであり、

40

50

入力音声をデジタル信号に変換する。207はマイクであり、音声を受け付ける。208はバスである。

【0012】

本実施例では、図6のように、ユーザが、タッチスクリーン601に対して、ペン602を用いて、手書きストローク603を描きながら、「シカケイ」と発声したケースを考える。上述の構成よりなる情報処理装置の動作について、以下フローチャートを参照して説明する。

【0013】

図3はペン入力情報受信部103、図形認識部104の動作を示すフローチャートである。該フローチャートを実行するプログラムはROM203またはHDD205に保持され、CPU201の制御のもと実行される。まず、タッチスクリーン204に対して、ユーザからのペン入力があったかどうかをチェックする(S301)。ペン入力があった場合、ペン入力情報受信部103でペン入力情報を受信し、受信したペン入力情報から、図形認識部104で図形を認識する(S302)。その認識結果を不図示の図形認識結果スタックにプッシュする(S303)。図形認識結果スタックはRAM202やHDD205に保持されている。図7は、図形認識結果スタックにプッシュされた図形認識結果の例である。このように図形認識結果は、ペン入力開始、終了のタイムスタンプ(701、702)と、図形種別の確信度付きN - b e s t 703(ここでは、N = 3)を含む。確信度は公知の手法で求められる。相違度等、確信度と同様の概念のものを求められるのであればどのような手法を用いても構わない。

【0014】

図4は音声情報受信部101、音声認識部102の動作を示すフローチャートである。該フローチャートを実行するプログラムはROM203またはHDD205に保持され、CPU201の制御のもと実行される。まず、音声情報受信部は、ユーザからの音声入力があったかどうかをチェックする(S401)。音声入力の検出は、公知の技術であり、例えば入力音声のパワーの大きさなどから判断する。音声入力があった場合は、受信してこれを音声認識する(S402)。そして認識結果を不図示の音声認識結果スタックにプッシュする(S403)。音声認識結果スタックはRAM202やHDD205に保持されている。図8は、音声認識結果スタックにプッシュされた音声認識結果の例である。このように音声認識結果は、発声開始、終了のタイムスタンプ(801、802)と、図形種別の確信度付きN - b e s t 803(ここでは、N = 3)を含む。音声認識の確信度は既知の手法で求めることができる。

【0015】

図5は総合評価部105の動作を示すフローチャートである。該フローチャートを実行するプログラムはROM203またはHDD205に保持され、CPU201の制御のもと実行される。総合評価部105は、図形認識結果スタックを監視し、スタックに何らかの図形認識結果が保持されているかどうかをチェックする(S501)。図形認識結果が保持されていれば、音声認識結果スタックの中に、この図形認識結果に対応する音声認識結果が保持されているかどうかをチェックする(S502)。ここで、図形認識結果スタックは常時監視していても良いし、定期的にチェックするようにしても良い。また、図形認識結果スタックではなく音声認識結果スタックを監視するよう構成しても構わない。その場合、S501とS502の順序が逆になる。音声認識結果が図形認識結果に対応するか否かは、例えばまず、タイムスタンプで判断する。具体的には図10のように、音声の開始時刻よりT1秒前から、終了時刻のT2秒後までの期間Dに、図形認識結果のペン入力開始タイムスタンプ701、もしくはペン入力終了タイムスタンプ702が含まれていれば、音声認識結果と図形認識結果を対応させる。ここでは図8の音声認識結果が、図7の図形認識結果に対応するものとして説明する。

【0016】

図形認識結果に対応する音声認識結果が見つかった場合は、図形認識結果に含まれる一つまたは複数個の図形種別と音声認識結果に含まれる一つまたは複数個の図形種別の中で

10

20

30

40

50

、図形種別が一致するもの同士の確信度の値を足し合わせる。図7、図8では、図形認識結果の楕円の確信度60と音声認識結果の楕円の確信度10を足し合わせて70になり、三角形、四角形はそれぞれ90、110になる。そして、足し合わせた確信度の値が最も大きな図形種別を選択する(S503)。ここでは、足し合わせた確信度が最も大きい四角形が選択される。なお、S502で、対応する音声認識結果がなかった場合は、図形認識結果に含まれる一つまたは複数個の図形種別の中で最も確信度の値が大きい図形種別を選択する(S504)。以上のようにして、図形認識結果に含まれる一つまたは複数個の図形種別から一つを選択する処理が完了し、図形認識結果スタック、音声認識結果スタックをすべてクリアする(S505)。以上の処理の結果、図9のような総合評価結果から、図形種別は「四角形」と判断される。

10

【0017】

図11は特徴量抽出部106、生成部107、表示制御部108の動作を示すフローチャートである。該フローチャートを実行するプログラムはROM203またはHDD205に保持され、CPU201の制御のもと実行される。総合評価部105が判定した図形種別に応じて、入力ストロークの特徴量としてここでは特徴点を抽出する。四角形の場合、図12の1201から1204のように、4つの頂点を見つけ出す(S1101)。特徴点を見つけ出すアルゴリズムは公知技術である。この特徴点に従って、図13の1301のように図形を生成し(S1102)、これを表示する(S1103)。表示の際、元のストロークの表示は消す。整形された図形はRAM202やHDD205に記憶される。

20

【0018】

以上の例では、ペン入力ストロークからの図形認識では、認識結果の第一位は「楕円」であり、ユーザの意図とは違うものであった。一方、音声認識結果の第一位は「三角形」であり、これもユーザの意図とは異なる。しかし、両方の認識結果を総合評価することで、ユーザの意図通りの「四角形」として図形を整形表示することができた。このように、図形認識結果と音声認識結果を総合評価することで、ユーザの意図に対する認識精度が高まるという効果がある。

【0019】

なお、本実施例では、図形を生成する処理まで記載したが、図5の図形種別の選択処理まででも本発明の要件は満たされる。

30

【実施例2】

【0020】

実施例2では、情報処理装置が備えるタッチスクリーンに対してペンで文字を描き、同時にその文字の読みを発声し、入力文字を整形表示するケースで説明する。図14は、本実施例に係る情報処理装置におけるマルチモーダル入力処理のブロック図である。同図において、1401は音声情報受信部である。1402は音声認識部である。1403はペン入力情報受信部である。1404は文字認識部である。1405は総合評価部である。1406は総合評価部1405が出力する文字等を表示するよう制御する表示制御部である。総合評価部1407は単漢字辞書である。1408は音声認識文法である。

【0021】

実施例2に係る情報処理装置のハードウェア構成は図2と同様である。本実施例では、図18のように、ユーザが、タッチスクリーン1801に対して、ペン1802を用いて、手書きで文字ストローク1803を描きながら、「アイ」と発声したケースを考える。

40

【0022】

図15はペン入力情報受信部1403、文字認識部1404の動作を示すフローチャートである。該フローチャートを実行するプログラムはROM203またはHDD205に保持され、CPU201の制御のもと実行される。まず、タッチスクリーン204に対して、ユーザからのペン入力があったかどうかをチェックする(S1501)。ペン入力があった場合、ペン入力情報受信部でペン入力情報を受信し、文字認識部1404で該ペン入力情報を文字認識する(S1502)。その認識結果を不図示の文字認識結果スタック

50

にプッシュする (S 1 5 0 3)。文字認識結果スタックは R O M 2 0 3 または H D D 2 0 5 に保持されている。図 1 9 は、文字認識結果スタックにプッシュされた文字認識結果の例である。このように文字認識結果は、ペン入力開始、終了のタイムスタンプ (1 9 0 1、1 9 0 2) と、文字種別の確信度付き N - b e s t 1 9 0 3 (ここでは、N = 3) を含む。認識結果の各文字には、その文字の読み (複数ある場合はすべて) が付与されている。この読み情報は、文字がカタカナ、ひらがな、アルファベットの場合は、その表記から取得し、漢字の場合は、単漢字辞書 1 4 0 7 から取得する (単漢字辞書は、“?”、“#”などの記号も一つの漢字として扱うことで、記号にも対応できる)。

【 0 0 2 3 】

音声情報受信部 1 4 0 1、音声認識部 1 4 0 2 の動作は図 4 と同様である。まず、音声情報受信部は、ユーザからの音声入力があったかどうかをチェックする (S 4 0 1)。音声入力の検出は、公知の技術であり、入力音声のパワーの大きさなどから判断する。音声入力があった場合は、受信した音声情報を音声認識する (S 4 0 2)。音声認識部 1 4 0 2 は、音声認識文法 1 4 0 8 に基づいて音声を認識する。音声認識文法 1 4 0 8 は、5 0 音、および、単漢字辞書 1 4 0 7 に含まれるすべての単漢字の読みを重複のないように集めたりストから構成される。従って、音声認識結果は、5 0 音、または、単漢字辞書 1 4 0 7 に含まれる単漢字の読みのいずれかになる。そして認識結果を不図示の音声認識結果スタックにプッシュする (S 4 0 3)。音声認識結果スタックは R O M 2 0 3 又は H D D 2 0 5 に保持される。図 2 0 は、音声認識結果スタックにプッシュされた音声認識結果の例である。このように音声認識結果は、発声開始、終了のタイムスタンプ (2 0 0 1、2 0 0 3) と、単漢字読みの確信度付き N - b e s t 2 0 0 3 (ここでは、N = 3) を含む。

【 0 0 2 4 】

図 1 7 は総合評価部 1 4 0 5 の動作を示すフローチャートである。該フローチャートを実行するプログラムは R O M 2 0 3 または H D D 2 0 5 に保持され、C P U 2 0 1 の制御のもと実行される。総合評価部 1 4 0 5 は、文字認識結果スタックを監視し、スタックに何らかの文字認識結果が保持されているかどうかをチェックする (S 1 7 0 1)。文字認識結果が保持されていれば、音声認識結果スタックの中に、この文字認識結果に対応する音声認識結果が保持されているかどうかをチェックする (S 1 7 0 2)。ここで、文字認識結果スタックは常時監視していても良いし、定期的にチェックするようにしても良い。また、文字認識結果スタックではなく音声認識結果スタックを監視するよう構成しても構わない。音声認識結果が文字認識結果に対応するか否かは、まず、タイムスタンプで判断する。その判断方法は、実施例 1 のケースと同じである。

【 0 0 2 5 】

文字認識結果に対応する音声認識結果が見つかった場合は、文字認識結果に含まれる一つまたは複数個の文字種別それぞれに付与された確信度を、音声認識結果も用いて修正する。確信度の修正は以下のようにして行う。文字認識結果第一候補の「監」は、“k a n”と“k e n”の2つの読みを持つ。それぞれについて、音声認識結果に含まれる3つの読み候補“e i”、“s a i”、“a i”との類似度を計算する。例えば、「監」の読み“k a n”と、音声認識結果の“e i”は、構成音素が全く一致しないので、類似度は0となる。従って、“k a n”の“e i”による修正確信度 M C (“k a n” | “e i”) は、

$$\begin{aligned} M C (“k a n” | “e i”) &= C 1 (“k a n”) + C 2 (“e i”) \times \text{類似度} \\ &= 7 0 + 6 0 \times 0 \\ &= 7 0 \end{aligned}$$

C 1 : 文字認識確信度、C 2 : 音声認識確信度
となる。“k a n”と“s a i”の場合、“k a n”の3つの音素のうち、“s a i”に含まれる音素と一致しているのは“a”の一つだけなので、類似度は、 $1 / 3 = 0 . 3 3$ となる。従って、

$$M C (“k a n” | “a i”) = 7 0 + 5 5 \times 0 . 3 3 = 8 8 . 3$$

となる。同様に、

$$MC("kan" | "sai") = 70 + 30 \times 0.33 = 80$$

となる。このうち最大の値を採用し、“kan”の修正確信度 $MC("kan")$ は、 88.3 となる。「監」のもう一つの読み“ken”についても同様に、修正確信度 $MC("ken")$ は、 $70 + 60 \times 0.33 = 90$ となる。文字「監」の修正確信度 $MC("監")$ は、これらの最大値を取り、

$$MC("監") = 90$$

となる。同様にして、

$$MC("蘭") = 50 + 55 \times 0.33 = 68.3$$

$$MC("藍") = 40 + 55 \times 1.0 = 95$$

となる。以上のようにして修正確信度が求まり、最大の修正確信度を持つ文字“藍”が最終的な認識結果として選択される(S1703)。なお、S1702で、対応する音声認識結果がなかった場合は、文字認識結果に含まれる一つまたは複数個の文字の中で最も確信度の値が大きい文字が選択される(S1704)。そして、選択された文字を、採用する文字として決定する。

【0026】

以上のようにして、文字認識結果に含まれる一つまたは複数個の文字から一つを選択する処理が完了し、文字認識結果スタック、音声認識結果スタックをすべてクリアする(S1705)。以上の処理の結果、図21のような総合評価結果から、文字は「藍」と判断され、図22のように、任意のフォントで「藍」の文字が表示される。選択された文字はRAMやHDDに記憶される。

【0027】

以上の例では、ペン入力ストロークからの文字認識では、認識結果の第一位は「監」であり、ユーザの意図とは違うものであった。一方、音声認識結果の第一位は「ei」であり、これもユーザの意図とは異なる。しかし、両方の認識結果を総合評価することで、ユーザの意図通りの「藍」という文字を表示することができた。このように、文字認識結果と音声認識結果を総合評価することで、ユーザの意図に対する認識精度が高まるという効果が得られる。

【実施例3】

【0028】

上記実施例では、ペン入力のタイムスタンプが入力音声の入力期間Dに含まれている場合に、音声認識結果と、図形認識結果または文字認識結果を対応付けていたが、本発明はこれに限られるものではない。書きながら発声するのが苦手なユーザにとっては、書き終わってから発声する方が使いやすい場合も考えられる。そのため本実施例では、ペン入力終了後に発声された音声将该ペン入力に対応付ける。具体的にはペン入力のタイムスタンプをとり、該タイムスタンプが表す時刻よりも後に入力された音声を該ペン入力に対応させる。ここでペン入力終了後所定時間内という制限をつけても構わない。また、これとは逆に音声入力終了後に入力されたペン入力を該音声に対応付けても良いことは言うまでもない。

【0029】

このような構成とすることで、ペン入力後に入力した図形や文字に対応する音声を発声する実施形態、また発声後対応する図形や文字をペン入力する実施形態を実行することが可能となる。

【実施例4】

【0030】

上記実施例では、明示はしていないが、毎回ペン入力と音声入力を総合評価する例について説明したが、本発明はこれに限定されるものではない。基本的にはペンによる入力を図形認識又は文字認識して認識した結果を表示するという処理を繰り返し、それをユーザが見て間違っていると判断した場合に音声入力を行うことで整形する実施形態も考えられる。

10

20

30

40

50

【0031】

図16にこの場合の処理のフローチャートを示す。該フローチャートを実行するプログラムはROM203またはHDD205に保持され、CPU201の制御のもと実行される。ここでは、図形を認識する場合の例で説明する。まず、ユーザからのペン入力があったかどうかをチェックする(S1601)。ペン入力があった場合、そのペンのストロークから図形を認識する(S1602)。その認識結果を図形認識結果スタックにプッシュする(S1603)。そして認識した結果最も確信度が高かった図形を表示する(S1604)。ここでユーザは表示された図形を確認し、図形が意図しているものと同じ場合は音声を入力しないため、S1605でNと判断され、図形認識結果をクリアし(S1606)、S1601に戻る。なお、図形認識結果スタックをクリアするタイミングは、次のペン入力が始まった段階や一定時間経過後にクリアするようにしても構わない。

10

【0032】

ユーザが図形を確認した結果、意図しているものと違うと判断した場合は、意図したものに修正するために発声を行う。本システムではこれを受け付け(S1605)、受け付けた発声を音声認識し(S1607)、音声認識結果を音声認識結果スタックにプッシュする(S1608)。その後の処理は図5と同様である。

【0033】

なお、一文字ずつではなく連続してペン入力を行う場合は、間違いに気づく前に次の図形を入力してしまう可能性がある。そのため、S1604で表示された図形の中で、修正対象の図形を指定し、該図形に対して意図する図形の名前を発声するよう構成しても構わない。

20

【0034】

このような構成とすることで、ユーザの観点からすると、基本的にはペン入力で入力した図形が整形されていき、間違えていると判断した場合に、意図している図形の名前を発声することで所望の図形に修正することが可能となる。

【0035】

なお、上記ではペン入力を入力していき、間違った場合に音声入力を行う例で説明したが、これとは逆に、音声で入力していき、間違った場合にペン入力での修正する構成でも良いことは言うまでもない。

【0036】

また、上記では図形を入力する例で説明したが、実施例2のように文字を入力する場合にも適用できることは言うまでもない。

30

【実施例5】

【0037】

上記実施例では、ペン入力をしている時に発声するケースに対応する例を挙げて説明したが、本発明はこれに限られるものではない。手書きで記載された紙文書等をスキャナで読み込んだ際に、記載内容を整形する場合にも適用することができる。

【0038】

本実施例では、手書きの図形が複数記載された紙をスキャンし、スキャンした図形を整形する機能を備えた複合機を例にあげて説明する。図24は本実施例の複合機のハードウェア構成図である。2401はCPU、2402はRAM、2403はROM、2404はハードディスクである。2405はA/Dコンバータであり、入力音声をデジタル信号に変換する。2406はマイクである。2407は紙の情報を読み取る画像読取部である。2408はタッチスクリーンである。2409はプリンタ部である。

40

【0039】

図25は本実施例の処理の流れを示すフローチャートである。該フローチャートを実行するプログラムはROM203またはHDD205に保持され、CPU201の制御のもと実行される。まず、画像読取部2407で手書きの図形が複数記載された紙を読み取り、画像データに変換する(S2501)。そして該画像データから図形を抽出する(S2502)。抽出した図形を図形認識し(S2503)、認識結果を図形認識結果スタック

50

にプッシュする（S2504）。認識結果から最も確信度の高い図形を選択し、選択された図形をもとに、S2503で抽出した図形から特徴量を抽出する（S2505）。抽出した特徴量に基づいて図形を生成し（S2506）、タッチスクリーン2408に該図形を表示する（S2507）。読み取った紙には複数の図形が記載されているため、ここでは複数の図形が表示されることになる。ユーザは表示された図形を確認し、意図しているものと異なる場合は、その図形を指定する。ここでは、表示されている図形をタッチスクリーン上で直接指示する。複合機はユーザが指示した図形を指定されたと判断する（S2508）。ユーザは意図する図形を発声し、複合機はこれを入力音声として受信する（S2509）。受信した音声を認識し（S2510）、音声認識結果と、S2504で図形認識結果スタックにプッシュされた図形認識結果とを総合評価する（S2511）。総合評価処理は実施例1と同様である。総合評価した結果選択された図形に基づいて、S2503で抽出した図形から特徴量を抽出し、該特徴量を用いて図形を生成し、生成した図形を表示する（S2512）。

10

【0040】

以上のような構成とすることで、リアルタイムのペン入力だけでなく、あらかじめ記載された図形についても、音声を用いて意図した図形に修正することが可能となる。なお、実施例2のように文字を対象としても良いことは言うまでもない。

【実施例6】

【0041】

実施例2では、一文字の認識を扱っていたが、本発明は2文字以上の単語を一度に認識するケースへも適用可能である。その場合、単漢字辞書1407の代わりに、単語辞書を用い、音声認識文法1408は、その単語の読みを重複なくリストアップしたのから構成されるようにする。単語辞書は、さまざまな単語の表記とその読みの対データを保持する。そして、文字認識部1404が複数の文字を認識し、認識した文字列で単語辞書を検索してその読みを取得するようにし、読み情報を含んだ結果を文字認識スタックにプッシュするように構成する。以降は、実施例2と同じアルゴリズムで対応可能である。

20

【0042】

以上のような構成とすることで、一文字の認識だけでなく、単語の認識についても精度を向上することが可能となる。

【実施例7】

【0043】

実施例6では、複数の文字を一度に認識するケースを述べたが、複数の文字を、上記実施例2の方法を用いて一文字ずつ認識してその結果を確定し、その後、次の文字を認識するようにしてもよい。この場合、各文字の認識結果の確定は、情報処理装置に備えられた所定のボタンを押す等して行う。さらに、ある一文字を認識する際、その直前までの文字の確定した認識結果を利用する。そのために、図14の構成に、さらに単語辞書を用いる。「憂鬱」という単語をタッチスクリーン上に書くケースを考える。ユーザは、まず、「憂」の文字をペンで書きながら「ユウ」と発声し、実施例2の方法で「憂」の文字を入力。確定ボタンで確定して、任意のフォントで「憂」の文字がタッチスクリーン上に表示される。そのすぐ隣に、続けて「鬱」の文字を意図して、図23の2303のようなストロークをペンで書きながら「ウツ」と発声する。この手書き文字を、文字認識と音声認識の総合評価で認識する前に、直前の「憂」という文字で始まる単語を単語辞書から検索する。そして、「憂鬱（ユウ・ウツ）」「憂国（ユウ・コク）」「憂愁（ユウ・シュウ）」「憂色（ユウ・シヨク）」「憂慮（ユウ・リョ）」が検索結果として得られる。これにより、2文字目は、「鬱（ウツ）」「国（コク）」「愁（シュウ）」「色（シヨク）」「慮（リョ）」である確率が高いことになる。音声認識文法1408に含まれる認識語彙の中で、「ウツ」「コク」「シュウ」「シヨク」「リョ」に対する重みを増やし、これらが認識結果に出やすくする。一方、文字認識部1404でも、ストローク2303に対して「鬱」「国」「愁」「色」「慮」といった文字が認識結果に出やすくなるよう適当に重み付けして認識する。このような構成とすることで、「単語」としてよりもっともらしい認識結

30

40

50

果を得られるという効果がある。

【実施例 8】

【0044】

実施例 2 では、音声認識部 1402 は、音声認識文法 1408 に基づいて音声認識を行い、音声認識文法 1408 は、50音及び単漢字辞書 1407 に含まれるすべての単漢字の読みを重複のないように集めたリストから構成していた。しかし、本発明はこれに限定されるものではない。音声認識文法 1408 として、任意の音素の組み合わせを受理する連続音声認識文法を用意し、音声認識部 1402 を連続音声認識処理可能なものとし、単漢字辞書 1407 に含まれる単漢字の読みに限定されない、任意の音素列を認識できるようにしてもよい。

10

【実施例 9】

【0045】

実施例 2 では、音声認識部 1402 は、文字認識部 1404 の認識処理とは独立に音声認識を行っていた。これに対し、文字認識部 1404 の認識が出るのを待ち、その N - best に含まれる文字の読みのリストを取り出し、その読みが音声認識結果に出やすくなるように、音声認識文法 1408 に適当な重み付けをして音声認識処理をするようにしてもよい。例えば、図 19 のような文字認識結果が得られた場合、“kan”、“ken”、“ran”、“ai” が音声認識結果に出やすくなるよう重み付けするように構成してもよい。

20

【実施例 10】

【0046】

実施例 9 とは逆に、音声認識部 1402 の認識結果が出るのを待ち、その N - best に含まれる読みのリストを取り出し、その読みを持つ文字が文字認識結果に出やすくなるように、文字認識部 1404 に適当な重み付けをして処理をするようにしてもよい。例えば、図 20 のような音声認識結果が得られた場合、“ei”、“ai”、“sai” という読みを持つ文字が文字認識結果に出やすくなるよう重み付けするようにしてもよい。

30

【実施例 11】

【0047】

上記実施例では、音声認識結果と、図形認識結果又は文字認識結果を総合評価する際に、確信度を加算する例をあげて説明したが、これに更に重み付けをするよう構成しても構

わない。

【0048】

例えば、周囲雑音が大きい環境で処理をする場合、音声認識の精度は雑音の影響を受けて落ちてしまうことが考えられる。そこで、周囲雑音を検出し、雑音の大きさにあわせて音声認識結果に対する重み付けを調整することが考えられる。雑音が大きい場合に音声認識結果の重みを減らせば、雑音の影響を緩和することが可能となる。実施形態としては、例えば図 1 において、雑音を検出する不図示の雑音検出部を更に備え、雑音検出部が検出した雑音を総合評価部 105 が受け取る。そして、総合評価をする際に、受け取った雑音の大きさに応じて音声認識結果の重みを調整するという構成をとればよい。

40

【0049】

また、ペン入力において、入力速度が速い場合には記載された図形又は文字が雑であるという仮説を立てることができる。そこで、ペン入力の入力速度を検出し、入力速度に応じてペン入力に対する図形認識結果又は文字認識結果の重み付けを調整することが考えられる。速度が速い場合に図形認識結果又は文字認識結果の重みを減らすことで、全体の認識精度を向上することが期待できる。実施形態としては、例えば図 1 において、ペン入力速度を検出する不図示のペン入力速度検出部を更に備え、ペン入力速度検出部が検出したペン入力速度を総合評価部 105 が受け取る。そして、総合評価をする際に、受け取ったペン入力速度に応じて図形認識結果又は文字認識結果の重みを調整するという構成をとればよい。

【実施例 12】

50

【 0 0 5 0 】

上記実施例では、図形認識又は文字認識と、音声認識を総合評価する例をあげて説明したが、顔認識やジェスチャ認識を対象としても構わない。具体的には、人物の特定をするために、名前の発声を音声認識し、顔認識と総合評価することで人物の認識率を向上することが可能となる。またジェスチャをしながら該ジェスチャに対応する発声を発声することでジェスチャの認識率を向上することが可能となる。

【実施例 1 3】

【 0 0 5 1 】

なお、本発明の目的は、前述した実施例の機能を実行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み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。

10

【 0 0 5 2 】

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行することになり、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。

【 0 0 5 3 】

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、不揮発性のメモリカード、ROMなどを用いることができる。

20

【 0 0 5 4 】

また、コンピュータが読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行されるだけでなく、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行される場合も含まれることは言うまでもない。

【 0 0 5 5 】

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行される場合も含まれることは言うまでもない。

30

【 0 0 5 6 】

以上の説明では、プログラムとハードウェアとの組み合わせで、発明を実現したが、プログラムのウエイトを少なくしてハードウェアを多用しても良い。更に、実施例においては、入力ステップから出力ステップまでの流れを説明したが、本発明は、実施例の一部にも適応される。

【 0 0 5 7 】

また、上記実施例では、情報処理方法の工程で音声認識をする工程、図形認識や文字認識をする工程を備えた例をあげて説明したが、音声認識、図形認識や文字認識などの認識を行う工程は備えず、認識結果を受け取る工程を備え、受け取った認識結果を処理する流れとしても構わない。OSが音声認識、図形認識や文字認識などを行った結果を受け取り、それらを総合評価する工程を備えれば本発明を実行することが可能である。

40

【 0 0 5 8 】

本発明において、与えられた対象とは、図形認識や文字認識などのパターン認識の対象であり、例えば実施例1のペンで描かれた図形、実施例2のペンで描かれた文字、実施例5のスキャンした図形等である。図形認識手段とは、例えば実施例1の図形認識部104である。音声認識手段とは例えば実施例1の音声認識部102である。選択手段とは、上記実施例では、総合評価した結果から図形や文字を選択している。抽出手段とは、例えば実施例1の特徴量抽出部106である。生成手段とは、例えば実施例1の生成部107で

50

ある。表示制御手段とは、例えば実施例 1 の表示制御部 108 である。確信度とは、認識した結果得られた候補がそれぞれ対象とするものである確信の度合を示す値であり、例えば図 7 に示されるような値である。文字認識手段とは、例えば実施例 2 の文字認識部 1404 である。パターンを特定するための情報とは、図形の場合は図形の名称・種類、文字の場合は文字の読みである。図形種別とは図形の種類を示す情報であり、図形の名称等である。

【図面の簡単な説明】

【0059】

【図 1】実施例 1 に係る情報処理装置におけるマルチモーダル入力処理のブロック図である。 10

【図 2】実施例 1 に係る情報処理装置のハードウェア構成図である。

【図 3】実施例 1 に係る図形認識部の動作を説明するフローチャートである。

【図 4】実施例 1 に係る音声認識部の動作を説明するフローチャートである。

【図 5】実施例 1 に係る総合評価部の動作を説明するフローチャートである。

【図 6】実施例 1 において、ユーザがタッチスクリーン上に手書きで図形を描き、同時に音声入力を行った場面をあらわす図である。

【図 7】実施例 1 に係る図形認識部の認識結果の例である。

【図 8】実施例 1 に係る音声認識部の認識結果の例である。

【図 9】実施例 1 に係る総合評価部の総合評価結果の例である。

【図 10】実施例 1 に係る入力音声に対応するペン入力が含まれるべき時間幅を説明する図である。 20

【図 11】実施例 1 に係る特徴量抽出部、生成部の動作を説明するフローチャートである。

【図 12】実施例 1 に係る特徴量抽出部が割り出した特徴点の例である。

【図 13】実施例 1 に係るマルチモーダル入力によって、生成された図形が表示された図である。

【図 14】実施例 2 に係る情報処理装置におけるマルチモーダル入力処理のブロック図である。

【図 15】実施例 2 に係る文字認識部の動作を説明するフローチャートである。

【図 16】実施例 4 に係る図形認識処理のフローチャートである。 30

【図 17】実施例 2 に係る総合評価部の動作を説明するフローチャートである。

【図 18】実施例 2 において、ユーザがタッチスクリーン上に手書きで文字を描き、同時に音声入力を行った場面をあらわす図である。

【図 19】実施例 2 に係る文字認識部の認識結果の例である。

【図 20】実施例 2 に係る音声認識部の認識結果の例である。

【図 21】実施例 2 に係る総合評価部の総合評価結果の例である。

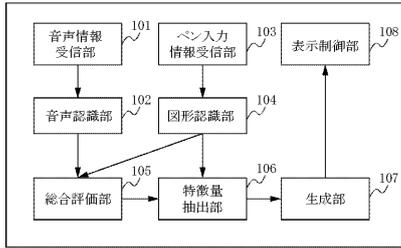
【図 22】実施例 2 に係るマルチモーダル入力によって、認識された文字が任意のフォントで表示された図である。

【図 23】実施例 4 において、ユーザがタッチスクリーン上に手書きで文字を描き、同時に音声入力を行った場面をあらわす図である。 40

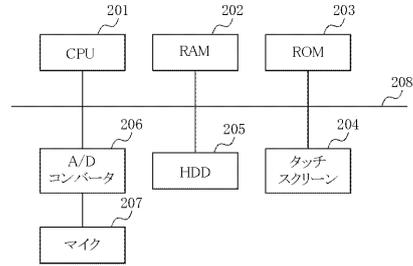
【図 24】実施例 5 に係る複合機のハードウェア構成図である。

【図 25】実施例 5 に係る図形生成処理のフローチャートである。

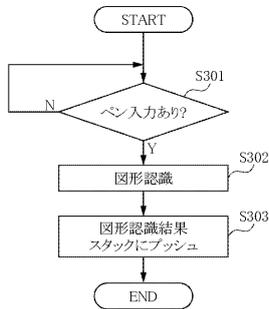
【図1】



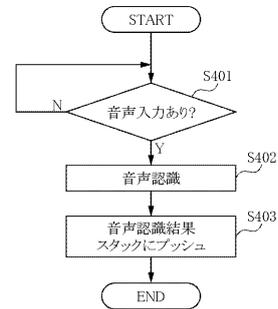
【図2】



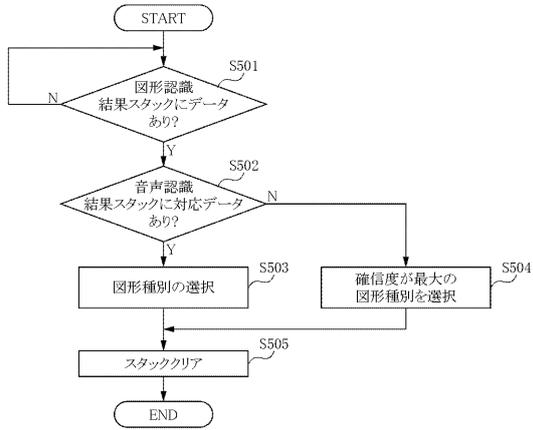
【図3】



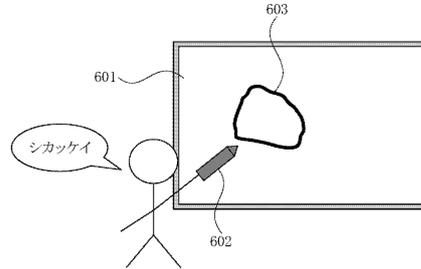
【図4】



【図5】



【図6】



【図7】

START	00:00:04:08		701
END	00:00:05:20		702
候補	第1位	図形種別	楕円
		確信度	60
	第2位	図形種別	四角形
		確信度	50
	第3位	図形種別	三角形
		確信度	20

} 703

【図8】

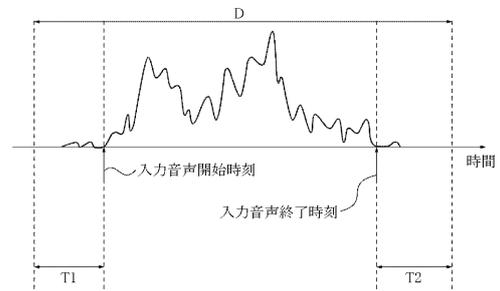
START	00:00:05:10		801
END	00:00:06:20		802
候補	第1位	図形種別	三角形
		確信度	70
	第2位	図形種別	四角形
		確信度	60
	第3位	図形種別	楕円
		確信度	10

} 803

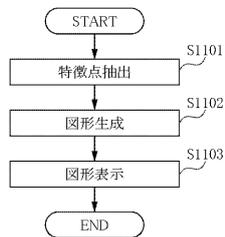
【 図 9 】

総合評価	第1位	図形	四角形
		確信度	110
	第2位	図形	三角形
		確信度	90
	第3位	図形	楕円
		確信度	70

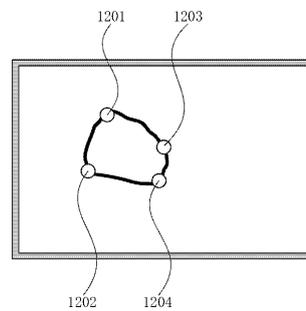
【 図 1 0 】



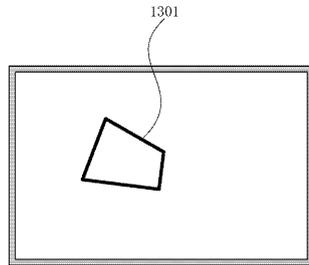
【 図 1 1 】



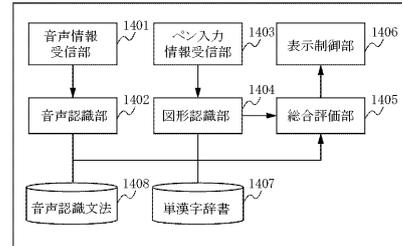
【 図 1 2 】



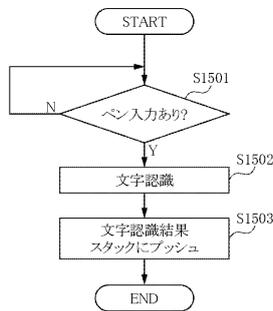
【図 13】



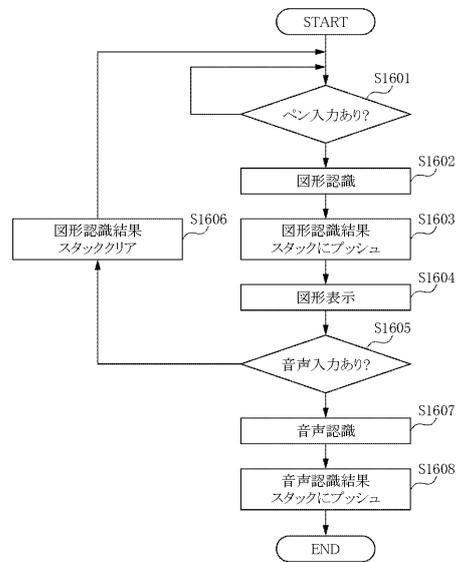
【図 14】



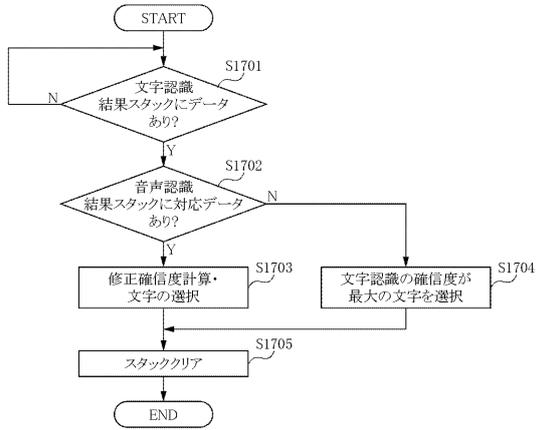
【図 15】



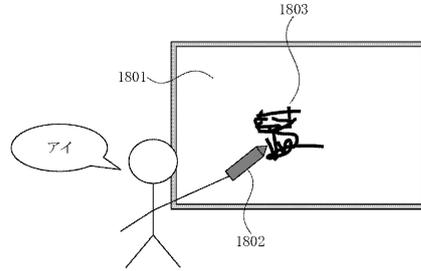
【図 16】



【図17】



【図18】



【図19】

START	00:00:04:80			1901	1902
END	00:00:05:20				
候補	第1位	文字	監(kan,ken)	1903	
		確信度	70		
	第2位	文字	蘭(ran)		
		確信度	50		
	第3位	文字	藍(ai,ran)		
		確信度	40		

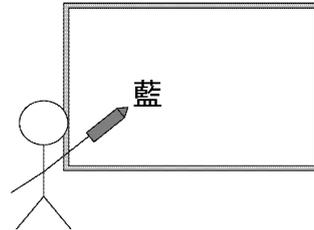
【図20】

START	00:00:05:10			2001	2002
END	00:00:06:20				
候補	第1位	読み	ei	2003	
		確信度	60		
	第2位	読み	ai		
		確信度	55		
	第3位	読み	sai		
		確信度	30		

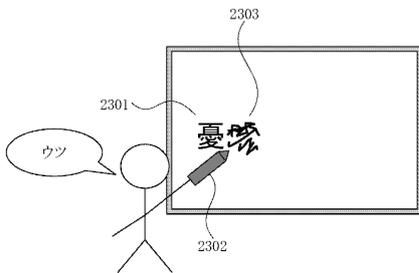
【図 2 1】

総合評価	第1位	文字	藍
		確信度	95
	第2位	文字	監
		確信度	90
	第3位	文字	蘭
		確信度	68.3

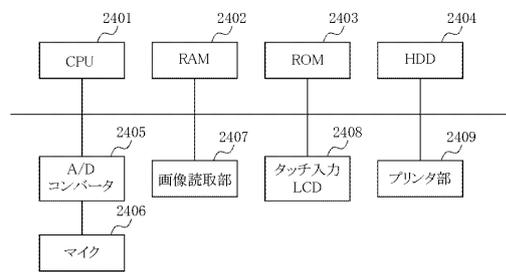
【図 2 2】



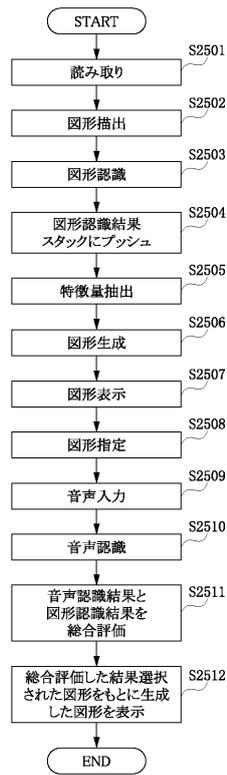
【図 2 3】



【図 2 4】



【図 25】



フロントページの続き

(51)Int.Cl. F I
G 0 6 F 3/01 (2006.01) G 0 6 F 3/01

(72)発明者 小森 康弘
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 吉 田 耕一

(56)参考文献 特開平08-083092(JP,A)
特開平11-311992(JP,A)
特開昭60-091435(JP,A)
特開平06-085983(JP,A)
特開2000-206986(JP,A)
特開平06-076116(JP,A)
特開平08-221408(JP,A)
特開平03-087800(JP,A)
特開昭63-159900(JP,A)

(58)調査した分野(Int.Cl., DB名)
G 0 6 F 3 / 1 6
G 0 6 F 3 / 0 1
G 0 6 K 9 / 0 0
G 0 6 T 7 / 0 0
G 1 0 L 1 5 / 0 0
G 1 0 L 1 5 / 2 4