

(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2006-146887

(P2006-146887A)

(43) 公開日 平成18年6月8日(2006.6.8)

(51) Int. Cl.	F I	テーマコード (参考)
G06F 17/24 (2006.01)	G06F 17/24 554G	5B009
G10L 15/22 (2006.01)	G10L 3/00 561E	5D015
G06F 3/16 (2006.01)	G10L 3/00 561C	
	G06F 3/16 320B	

審査請求 未請求 請求項の数 25 O L (全 19 頁)

(21) 出願番号	特願2005-299050 (P2005-299050)	(71) 出願人	500046438
(22) 出願日	平成17年10月13日 (2005.10.13)		マイクロソフト コーポレーション
(31) 優先権主張番号	10/997, 255		アメリカ合衆国 ワシントン州 9805
(32) 優先日	平成16年11月24日 (2004.11.24)		2-6399 レッドモンド ワン マイ
(33) 優先権主張国	米国 (US)		クロソフト ウェイ
		(74) 代理人	100077481
			弁理士 谷 義一
		(74) 代理人	100088915
			弁理士 阿部 和夫
		(72) 発明者	デビッド マウアット
			アメリカ合衆国 98052 ワシントン
			州 レッドモンド ワン マイクロソフト
			ウェイ マイクロソフト コーポレーシ
			ョン内

最終頁に続く

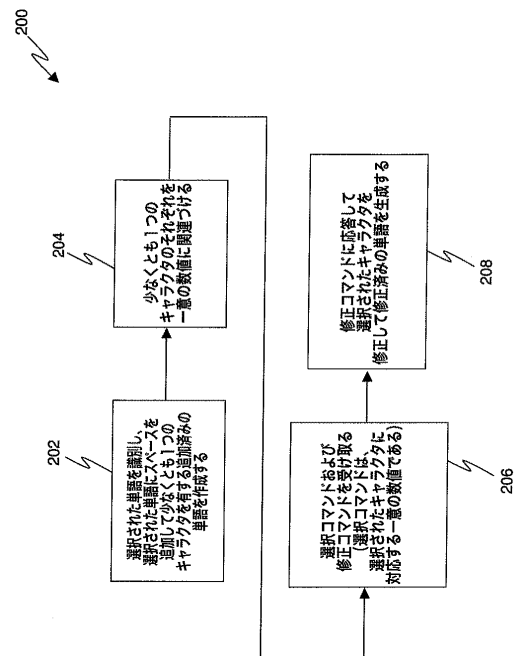
(54) 【発明の名称】 キャラクタの制御された操作

(57) 【要約】

【課題】ディスプレイ画面上に表示されるキャラクタを操作するための方法およびシステムを提供する。

【解決手段】この方法は、修正すべき少なくとも1つのキャラクタを含む選択された単語を識別することを含む。この方法は、少なくとも1つのキャラクタのそれぞれを一意的数値に関連づけること、ならびに選択されたキャラクタに対応する一意的数値である選択コマンドと、修正コマンドを受け取ることをさらに含んでいる。さらに、この方法は、この修正コマンドに応答して選択されたキャラクタを修正して、修正済みの単語を生成することを含む。

【選択図】図2



【特許請求の範囲】**【請求項 1】**

ディスプレイ画面上に表示されるキャラクタを操作するための方法であって、
修正すべき少なくとも 1 つのキャラクタを含む選択された単語を識別するステップと、
前記少なくとも 1 つのキャラクタのそれぞれを一意の数値に関連づけるステップと、
前記選択された単語中の選択されたキャラクタに対応する前記一意の数値である選択コマンド、および修正コマンドを受け取るステップと、
前記修正コマンドに応答して前記選択されたキャラクタを修正して、修正済みの単語を生成するステップと
を含むことを特徴とする方法。

10

【請求項 2】

前記識別するステップは、
ユーザが前記選択された単語を口頭で伝えることができるようにするスペリングユーザインターフェースをスペリングコマンドが呼び出すことができるようにするステップをさらに含むことを特徴とする請求項 1 に記載の方法。

【請求項 3】

前記関連づけるステップは、
前記選択された単語を前記ディスプレイ画面上に表示して、前記少なくとも 1 つのキャラクタのそれぞれと、前記一意の数値の各々との間の関連づけを視覚的に伝えるステップを
さらに含むことを特徴とする請求項 1 に記載の方法。

20

【請求項 4】

前記表示するステップは、
前記ディスプレイ画面上の前記選択された単語の周囲にボックスを描画するステップと、
前記少なくとも 1 つのキャラクタのそれぞれに隣接して前記一意の数値の各々を表示するステップと
を含むことを特徴とする請求項 3 に記載の方法。

【請求項 5】

前記識別するステップは、選択された単語を識別し、前記選択された単語に少なくとも 1 つのスペースを最後に追加して、追加済みの単語を作成し、前記追加済みの単語は、修正すべき少なくとも 1 つのキャラクタを含むことを特徴とする請求項 1 に記載の方法。

30

【請求項 6】

前記関連づけるステップは、
前記ディスプレイ画面上に前記追加済みの単語を表示して、前記少なくとも 1 つのキャラクタのそれぞれと前記一意の数値の各々との間の関連づけを視覚的に伝えるステップをさらに含むことを特徴とする請求項 5 に記載の方法。

【請求項 7】

前記表示するステップは、
前記ディスプレイ画面上の前記追加済みの単語の周囲にボックスを描画するステップと、
前記少なくとも 1 つのキャラクタのそれぞれに隣接して前記一意の数値の各々を表示するステップと
を含むことを特徴とする請求項 6 に記載の方法。

40

【請求項 8】

前記修正コマンドが、「削除」である場合、前記修正するステップは、前記選択されたキャラクタを前記追加済みの単語から削除するステップを含むことを特徴とする請求項 5 に記載の方法。

【請求項 9】

前記受け取るステップは、

50

オーディオ入力デバイスを介して前記選択コマンドおよび前記修正コマンドを受け取るステップを

含むことを特徴とする請求項 1 に記載の方法。

【請求項 10】

前記修正コマンドが、置換キャラクタである場合、前記修正するステップは、前記選択されたキャラクタを前記置換キャラクタで置き換えるステップを含むことを特徴とする請求項 1 に記載の方法。

【請求項 11】

前記修正コマンドが、「削除」である場合、前記修正するステップは、前記選択されたキャラクタを前記選択された単語から削除するステップを含むことを特徴とする請求項 1 に記載の方法。 10

【請求項 12】

前記修正コマンドが、「挿入」である場合、前記修正するステップは、前記選択されたキャラクタと直前のキャラクタの間にスペースを挿入するステップを含むことを特徴とする請求項 1 に記載の方法。

【請求項 13】

前記修正するステップは、前記スペースを選択し、前記ディスプレイ画面上に前記スペースを視覚的に示すステップをさらに含むことを特徴とする請求項 12 に記載の方法。

【請求項 14】

前記修正するステップは、前記選択されたキャラクタの直後に配置された後続のキャラクタを選択するステップをさらに含むことを特徴とする請求項 1 に記載の方法。 20

【請求項 15】

前記修正するステップは、前記ディスプレイ画面上に前記後続のキャラクタを視覚的に示すステップをさらに含むことを特徴とする請求項 14 に記載の方法。

【請求項 16】

ユーザが、終了コマンドを介して前記スピーリングユーザインターフェースを口頭で終了させることを可能にするステップをさらに含むことを特徴とする請求項 2 に記載の方法。

【請求項 17】

前記可能にするステップは、前記終了コマンドに応答して前記選択された単語を前記修正済みの単語で置き換えるステップをさらに含むことを特徴とする請求項 16 に記載の方法。 30

【請求項 18】

前記終了コマンドは、「OK」であることを特徴とする請求項 16 に記載の方法。

【請求項 19】

前記受け取るステップは、少なくとも 1 つの許容可能なキャラクタの修正のリストを表示するステップをさらに含み、前記リストは、前記少なくとも 1 つの許容可能なキャラクタの修正のそれぞれに対応する第 2 の一意の数値を含むことを特徴とする請求項 1 に記載の方法。

【請求項 20】

前記修正コマンドは、前記第 2 の一意の数値であることを特徴とする請求項 19 に記載の方法。 40

【請求項 21】

前記修正コマンドは、単語「変更」および「アンドゥ」のうちの少なくとも 1 つであることを特徴とする請求項 1 に記載の方法。

【請求項 22】

前記識別するステップは、オーディオ入力デバイスを介してオーディオ信号を受信する前記スピーリングユーザインターフェースをさらに含むことを特徴とする請求項 2 に記載の方法。

【請求項 23】

前記方法は、スタンドアロンアプリケーションモジュールと、ターゲットソフトウェア 50

アプリケーション、スピーチ認識ソフトウェアアプリケーション、およびオペレーティングシステムのうちの少なくとも１つと統合された統合アプリケーションモジュールとのうちの少なくとも１つとして実施することができることを特徴とする請求項１に記載の方法。

【請求項２４】

前記方法は、ターゲットソフトウェアアプリケーションとして実施することができ、前記ターゲットソフトウェアアプリケーションは、ワードプロセッシングアプリケーション、スプレッドシートアプリケーション、および電子メールアプリケーションのうちの少なくとも１つを含むことを特徴とする請求項１に記載の方法。

【請求項２５】

修正すべき少なくとも１つのキャラクタを含む選択された単語を識別するステップと、前記少なくとも１つのキャラクタのそれぞれを一意的な数値に関連づけるステップと、選択されたキャラクタに対応する前記一意的な数値である選択コマンド、および修正コマンドを受け取るステップと、

前記修正コマンドに回答して前記選択されたキャラクタを修正して、修正済みの単語を生成するステップと

を含むディスプレイ画面上に表示されるキャラクタを操作するための方法を処理デバイスに実行させる命令を含む機械読取り可能コンピュータプログラムコードを記録したことを特徴とするコンピュータ読取り可能媒体。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、一般に音声認識ソフトウェアアプリケーションに関し、より詳細には、音声認識アプリケーションを介して単語のキャラクタを操作するための方法に関する。

【背景技術】

【０００２】

スピーチは、おそらく最も古い形式の人間のコミュニケーションであり、発声を介してコミュニケーションする能力が人間の脳の生態において先天的にもたらされるものであると、今では多数の科学者達は確信している。したがって、ユーザがスピーチなどのＮＵＩ（Natural User Interface：ナチュラルユーザインターフェース）を使用してコンピュータとコミュニケーションすることができるようにすることは、長い間追求されてきた目標であった。実際、最近になってこの目標を達成する上での大きな発展がなされてきている。例えば、一部のコンピュータは、今やユーザが、コンピュータを動作させるためのコマンドもテキストに変換すべき口述（dictation）も共に口頭で入力できるようにするスピーチ認識アプリケーションを含んでいる。これらのアプリケーションは、一般的にマイクロフォンを介して取得された音声サンプルを定期的に記録し、これらのサンプルを解析してこのユーザが発話している音素を認識し、これらの発話された音素により構成される単語を識別することによって動作する。

【０００３】

スピーチ認識がより一般的になりつつあるが、経験豊富なユーザをいらいらさせ、初心者ユーザを離反させる傾向がある従来のスピーチ認識アプリケーションを使用することには、短所が依然として存在する。かかる短所は、この話し手とこのコンピュータの間の対話に関与している。例えば、人間の対話では、人々は、人々が聞き手から感じ取る反応に基づいて自分のスピーチを制御する傾向がある。したがって、会話中に、聞き手は、「yes」または「uh-huh」など、うなずき、または口頭での応答を行って、この聞き手が自分に対して何が発言されているかを理解していることを示すことにより、フィードバックを行うことができる。さらに、この聞き手が、自分に対して何が発言されているかを理解できない場合には、この聞き手は、いぶかしげな表情を表し、身を乗り出し、または他の口頭または非口頭の合図を与えることができる。このフィードバックに応じて、この話し手は、一般的に自分が話している話し方を変えることになり、場合によっては、この話し手

10

20

30

40

50

は、この話し手が聞き手と対話している話し方を変えていることについて、通常この聞き手が気付くことさえないようにしながら、さらにゆっくり、さらに大きな声で話し、さらに頻繁に休止をとり、あるいは表現を繰り返すことさえできる。したがって、会話中のフィードバックは、話し手が理解されているかどうかについてこの話し手に伝える非常に重要な要素である。しかし、残念ながら、従来の音声認識アプリケーションでは、マンマシンインターフェースによって実行されるスピーチ入力/コマンドに対する、このタイプの「ナチュラルユーザインターフェース (NUI)」フィードバック応答を実現することはまだできない。

【0004】

現在では、音声認識アプリケーションは、90%から98%の認識率 (accuracy rate) を達成している。これは、ユーザが、典型的な音声認識アプリケーションを使用してドキュメントに口述入力するときに、これらユーザのスピーチは、この音声認識アプリケーションによって約90%から98%の割合で正確に認識されることになることを意味している。したがって、この音声認識アプリケーションが記録する100文字ごとのうち、約2文字から10文字は、訂正する必要があることになる。この問題に対処し、誤認識された文字または単語を訂正する2つの一般的な方法では、文字または単語を反復すること、すなわち再び発話することが必要になり、あるいは代替スピーチを要求することが必要になる。しかし、これら2つのアプローチは、ユーザが訂正を実施するたびごとにうまくいくとは限らず、したがって、訂正を実施するときにスピーチを使用しなければならないある種のクラスのユーザにとって、例えば物理的にキーボードを使用することができないこれらのユーザにとっては、特に不利になる。

【0005】

この問題に対処し、ディスプレイ画面上に表示される誤認識された文字または単語を訂正する他のアプローチでは、この単語全体を削除し、この単語を先頭から再スペリングする (respelling) ことを必要とする。例えば、単語「intent」を「indent」に変更するためには、このユーザは、「delete intent (intentを削除せよ)」と発声し、次いで「i」、「n」、「d」、「e」、「n」、「t」を発声することにより、この所望の単語を再スペリングする必要があるはずである。この問題に対処し、ディスプレイ画面上に表示される誤認識された単語を訂正するさらに他のアプローチでは、音声によってこのキーボードを制御して、間違っているこれらの文字を変更することを必要とする。この場合には、このユーザは、変更する必要がある文字まで単語中の文字のすべてを削除する必要がある。次いで、ユーザは、この残りを再スペリングする。例えば、単語「intent」を「indent」に変更するためには、このユーザは、「backspace (バックスペース) backspace backspace backspace」と発声し、次いで「d」、「e」、「n」、「t」と発声することにより、この所望の単語を再スペリングすることになる。

【発明の開示】

【発明が解決しようとする課題】

【0006】

しかし、残念ながら、これらのアプローチには、それらに関連するいくつかの短所がある。第1に、1文字を変更するために多数のコマンドが必要とされることである。第2に、これらのアプローチは、多数の文字を再スペリングすることに頼っていることである。現在のスピーチ認識の認識率では、10文字中の9文字しか正しくないの、このことは、再スペリングの手段に訴えることにより、たった2つまたは3つの単語を訂正しなければならなかった後に、このユーザが、統計的に1つのエラーを獲得してしまう可能性があることを意味する。このことは、このユーザが、各文字の後に休止して文字が正しいかを確認する必要がある (これには、さらに時間がかかる)、あるいはこのユーザが、ユーザが「backspace backspace...」と発声し、次いで多くの場合にその単語をもう一度再スペリングしなければならない可能性があることに我慢する必要があることを意味する。第3に、スピーチ認識の誤りは、多くの場合にこのユーザが意図した単語と2~3文字異なっているにすぎないので、非常に多くの場合にこのユーザがそのスペルを操作している単

語は、このユーザが意図した単語と非常に近いことになる。これらの短所は、ベテランユーザをいらだたせる傾向があるだけでなく、初心者ユーザをがっかりさせる傾向もあり、おそらく、このユーザがこの音声認識アプリケーションを使用し続けるのを拒否する結果がもたらされることになる。

【課題を解決するための手段】

【0007】

ディスプレイ画面上に表示されるキャラクタを操作する方法が提供されており、この方法は、修正すべき少なくとも1つのキャラクタを含む選択された単語を識別することを含んでいる。この方法は、少なくとも1つのキャラクタのそれぞれを一意の数値に関連づけること、ならびに選択されたキャラクタに対応する一意の数値である選択コマンド、および修正コマンドを受け取ることをさらに含んでいる。さらにこの方法は、この修正コマンドに応答してこの選択されたキャラクタを修正して、修正済みの単語を生成することを含んでいる。

10

【0008】

ディスプレイ画面上に表示されるキャラクタを操作する方法を実施するためのシステムが提供されており、このシステムは、入力されたコマンドを受け取るように構成されたオーディオモジュールを含むストレージデバイスを含んでいる。このシステムはまた、この入力されたコマンドを受け取る入力デバイスと、この入力されたコマンドを表示する表示画面を含むディスプレイデバイスとを含んでいる。さらに、このシステムは処理デバイスをさらに含み処理デバイスは、スペリングUIをディスプレイ画面上に表示させ、入力されたコマンドに回答して表示されたデータを操作する命令を受け取るように、ストレージデバイス、この入力デバイス、およびこのディスプレイデバイスと情報をやりとりする。

20

【0009】

機械読取り可能コンピュータプログラムコードが提供され、このプログラムコードは、ディスプレイ画面上に表示されるキャラクタを操作する方法を処理デバイスに実施させるための命令を含んでいる。この方法は、修正すべき少なくとも1つのキャラクタを含む選択された単語を識別すること、およびこの少なくとも1つのキャラクタのそれぞれを一意の数値に関連づけることを含んでいる。この方法は、選択されたキャラクタに対応する一意の数値である選択コマンド、および修正コマンドを受け取ること、ならびにこの修正コマンドに回答してこの選択されたキャラクタを修正して、修正済みの単語を生成すること

30

【0010】

機械読取り可能コンピュータプログラムコードを用いて符号化される媒体が提供され、このプログラムコードは、ディスプレイ画面上に表示されるキャラクタを操作する方法を処理デバイスに実施させるための命令を含んでいる。この方法は、修正すべき少なくとも1つのキャラクタを含む選択された単語を識別すること、およびこの少なくとも1つのキャラクタのそれぞれを一意の数値に関連づけることを含んでいる。この方法は、選択されたキャラクタに対応する一意の数値である選択コマンド、および修正コマンドを受け取ること、ならびにこの修正コマンドに回答してこの選択されたキャラクタを修正して、修正済みの単語を生成することをさらに含んでいる。

40

【0011】

本発明の前述および他の特徴、ならびに利点については、添付図面と併せ解釈することにより、例示の実施形態の以下の詳細な説明からさらに完全に理解されよう。いくつかの図面中で、同様なエレメントは同様に番号が付けられている。

【発明を実施するための最良の形態】

【0012】

本明細書中で説明している本発明は、スピーチ認識アプリケーションを使用して、ユーザが入力する音声コマンドを受け取り認識するシステムを実現する汎用コンピュータと共に使用されるスタンドアロンアプリケーションモジュールおよび/または統合されたアプリケーションモジュールのコンテキストで説明しているが、本明細書中に開示される本発

50

明は、所望の最終目的に適切な任意のコンテキストで 사용할 수 있는 것을理解されたい。例えば、本発明は、ディスプレイ画面上にソフトウェアUI (User Interface: ユーザインターフェース) ウィンドウを位置づける方法を実行するスピーチ認識モジュールを有するMicrosoft (登録商標) Wordなどのターゲットソフトウェアアプリケーション内の統合されたソフトウェアルーチンまたは機能とすることができ、かつ/または本発明は、汎用コンピュータのオペレーティングシステム内のルーチンまたは機能とすることもできる。オブジェクト指向アプリケーションとして、このアプリケーションモジュールは、クライアントプログラムがアクセスしてこのアプリケーションモジュールと情報をやりとりする標準インターフェースを公開することができる。このアプリケーションモジュールは、ワードプロセッシングプログラム、デスクトップパブリッシングプログラム、あるアプリケーションプログラムなどいくつかの異なるクライアントプログラムが、ローカルに、かつ/またはWAN、LAN、および/またはインターネットベースの手段などのネットワーク上でこのアプリケーションモジュールを使用できるようにすることもある。例えば、このアプリケーションモジュールは、ローカルにまたはインターネットアクセスポイントを介して、電子メールアプリケーションやMicrosoft (登録商標) Wordなど、テキストフィールドを有する任意のアプリケーションおよび/またはコントロールと共にアクセスし、使用することができる。しかし、本発明の態様を説明する前に、本発明を組み込み、本発明の恩恵を受けることができる適切なコンピューティング環境の一実施形態について以下で説明する。

10

20

30

40

50

【0013】

図1を参照すると、スプリングユーザインターフェース(UI)を有するスピーチ認識ソフトウェアアプリケーションを使用するターゲットソフトウェアアプリケーションによってディスプレイ画面上に表示されたキャラクタを操作する方法を実施するシステム100を示すブロック図が示されており、このシステムは、処理デバイス104、システムメモリ106、およびこのシステムメモリ106を処理デバイス104に結合するシステムバス108を含む汎用コンピュータシステム102を含んでいる。システムメモリ106は、ROM (read only memory: 読取り専用メモリ) 110およびRAM (random access memory: ランダムアクセスメモリ) 112を含むことができる。起動中などに汎用コンピュータシステム102内のエレメント間で情報を転送する助けをする基本ルーチンを含むBIOS (basic input/output system: 基本入出力システム) 114は、ROM 110に記憶される。汎用コンピュータシステム102は、ハードディスクドライブ118、例えば着脱可能な磁気ディスク122から情報を読み取り、またはそれに情報を書き込むための磁気ディスクドライブ120、例えばCD-ROMディスク126を読み取り、あるいは他の光媒体から情報を読み取り、またはそれに情報を書き込むための光ディスクドライブ124などのストレージデバイス116をさらに含んでいる。ストレージデバイス116は、ハードディスクドライブインターフェース130、磁気ディスクドライブインターフェース132、光ドライブインターフェース134などのストレージデバイスインターフェースによってシステムバス108に接続することができる。これらのドライブおよびこれらに関連するコンピュータ読取り可能媒体は、汎用コンピュータシステム102についての不揮発性ストレージを実現する。以上のコンピュータ読取り可能媒体の説明では、ハードディスク、着脱可能な磁気ディスク、およびCD-ROMディスクについて言及しているが、磁気カセット、フラッシュメモリカード、デジタルビデオディスク、ベルヌーイカートリッジ (Bernoulli cartridge) など、コンピュータシステムによって読取り可能であり、この所望の最終目的に適している他のタイプの媒体も使用することができることを理解されたい。

【0014】

ユーザは、キーボード136、マウス138などのポインティングデバイス、およびマイクロフォン140を含めて、従来の入力デバイス135を介して汎用コンピュータシステム102にコマンドおよび情報を入力することができ、ここでマイクロフォン140を使用してスピーチなどのオーディオ入力を汎用コンピュータシステム102に入力するこ

とができる。さらに、ユーザは、スタイラスを使用して書込みタブレット 142 上にグラフィック情報を描画することにより、描画や手書きなどのグラフィック情報を汎用コンピュータシステム 102 に入力することができる。汎用コンピュータシステム 102 は、ジョイスティック、ゲームパッド、サテライトディッシュ、スキャナなど、この所望の最終目的に適した追加の入力デバイスを含むこともできる。マイクロフォン 140 は、システムバス 108 に結合されるオーディオアダプタ 144 を介して処理デバイス 104 に接続することができる。さらに、他の入力デバイスが、多くの場合にシステムバス 108 に結合されるシリアルポートインターフェース 146 を介して処理デバイス 104 に接続されるが、これらは、ゲームポートや U S B (universal serial bus: ユニバーサルシリアルバス) など他のインターフェースによって接続することもできる。

10

【0015】

ディスプレイ画面 148 を有する、モニタや他のタイプのディスプレイデバイス 147 などのディスプレイデバイス 147 もまた、ビデオアダプタ 150 などのインターフェースを介してシステムバス 108 に接続される。ディスプレイ画面 148 に加えて、汎用コンピュータシステム 102 は、一般的にスピーカおよび/またはプリンタなど他のペリフェラル出力デバイスも含むことができる。汎用コンピュータシステム 102 は、1 つまたは複数のリモートコンピュータシステム 152 への論理接続を使用してネットワーク環境中で動作することもできる。リモートコンピュータシステム 152 は、サーバ、ルータ、ピアデバイス、または他の共通ネットワークノードとすることができ、汎用コンピュータシステム 102 に対して説明しているエレメントのうちのどれかまたはすべてを含むことができるが、1 つのリモートメモリストレージデバイス 154 しか、図 1 には示していない。図 1 に示すような論理接続は、L A N (local area network: ローカルエリアネットワーク) 156 および W A N (wide area network: ワイドエリアネットワーク) 158 を含んでいる。かかるネットワーキング環境は、オフィス、企業規模のコンピュータネットワーク、イントラネット、およびインターネットにおいては一般的なものである。

20

【0016】

L A N ネットワーキング環境中で使用されるときには、汎用コンピュータシステム 102 は、ネットワークインターフェース 160 を介して L A N 156 に接続される。W A N ネットワーキング環境中で使用されるときには、汎用コンピュータシステム 102 は、一般的にインターネットなどの W A N 158 上で通信を確立するためのモデム 162 または他の手段を含んでいる。モデム 162 は、内蔵または外付けとすることができるが、シリアルポートインターフェース 146 を介してシステムバス 108 に接続することができる。ネットワーク環境においては、汎用コンピュータシステム 102 に関連して示されるプログラムモジュール、またはその一部分は、このリモートメモリストレージデバイス 154 に記憶することもできる。図に示すこれらのネットワーク接続は、例示的なものであり、コンピュータシステム間で通信リンクを確立する他の手段を使用することもできることを理解されたい。このアプリケーションモジュールは、汎用コンピュータシステム以外のホストコンピュータシステムまたはサーバコンピュータシステム上で同等に実装することができ、C D - R O M 以外の手段で、例えばネットワーク接続インターフェース 160 を介してこのホストコンピュータシステムに同等に伝送することも理解されたい。

30

40

【0017】

さらに、いくつかのプログラムモジュールは、汎用コンピュータシステム 102 のこれらのドライブおよび R A M 112 に記憶することができる。プログラムモジュールは、汎用コンピュータシステム 102 がどのように機能し、このユーザ、I / O デバイス、または他のコンピュータと対話するかを制御する。プログラムモジュールは、ルーチン、オペレーティングシステム 164、ターゲットアプリケーションプログラムモジュール 166、データ構造、ブラウザ、および他のソフトウェアコンポーネントまたはファームウェアコンポーネントを含んでいる。本発明の方法は、アプリケーションモジュール中に含めることができ、このアプリケーションモジュールは、本明細書中で説明する方法に基づいた

50

スピーチエンジン訂正モジュール170など1つまたは複数のプログラムモジュール中で便利に実施することができる。ターゲットアプリケーションプログラムモジュール166は、本発明に関連して使用される様々なアプリケーションを含むことができ、その一部は図2に示される。一部のこれらのプログラムモジュールの目的とその間の対話については、図2を説明するテキスト中でさらに十分に説明している。これらは、テキストフィールドを有する任意のアプリケーションおよび/またはコントロール、すなわち例えば電子メールアプリケーション、(ワシントン州、レッドモンド市のマイクロソフト社が生産するMicrosoft(登録商標)Wordなどの)ワードプロセッサプログラム、手書き認識プログラムモジュール、スピーチエンジン訂正モジュール170、およびIME(input method editor入力方式エディタ)を含んでいる。

10

【0018】

添付図面中に説明され示されるオペレーション、ステップ、およびプロシージャは、当業者なら本発明の例示の実施形態を実行できるようになるのに十分に開示されていると考えられるので、詳細な説明中で説明される様々なプロシージャを実行するための特定のプログラミング言語については、説明していないことを理解されたい。さらに、例示の実施形態を実行する際に使用することができる多数のコンピュータおよびオペレーティングシステムが存在しており、したがって、これらの多数の異なるシステムのすべてに適用可能となる詳細なコンピュータプログラムは、提供することができない。特定のコンピュータの各ユーザは、そのユーザのニーズおよび目的にとって最も有用な言語およびツールについて知っているはずである。

20

【0019】

図2を参照すると、スペリングユーザインターフェース(UI)を有するスピーチ認識ソフトウェアアプリケーションを使用するターゲットソフトウェアアプリケーションによってディスプレイ画面148上に表示されたキャラクタを操作する方法200を示すブロック図が示され、このスピーチ認識ソフトウェアアプリケーションを使用してこのターゲットソフトウェアアプリケーションに単語を入力するユーザの観点から説明している。

【0020】

このターゲットソフトウェアアプリケーションにテキストを入力するために、はじめにユーザは、このスピーチ認識ソフトウェアアプリケーションおよびターゲットソフトウェアアプリケーションを動作させて、図3に示すようにこのターゲットソフトウェアアプリケーションを介して少なくとも1つのキャラクタをディスプレイ画面148上に表示させる。処理デバイス104は、マイクロフォン入力デバイス135を介して入力されるユーザ命令に応答してこのスピーチ認識ソフトウェアアプリケーションを動作させることができ、かつ/または処理デバイス104は、起動直後にオペレーティングシステム164によって認識される「ブートアップ」命令などの命令に応答してこのスピーチ認識ソフトウェアアプリケーションを実行することができることを理解されたい。このスピーチ認識ソフトウェアアプリケーションを起動した後に、このユーザは、マイクロフォン入力デバイス140を介してこのスピーチ認識ソフトウェアアプリケーションにコマンドを口頭で伝えて、ターゲットソフトウェアアプリケーションを起動することができる。ここで、このターゲットソフトウェアアプリケーションは、電子メールアプリケーションやMicrosoft(登録商標)Wordなど、テキストフィールドを有する任意のアプリケーションおよび/またはコントロールとすることができる。このターゲットソフトウェアアプリケーションが起動された後に、ターゲットソフトウェアアプリケーションウィンドウ302が、ディスプレイ画面148上に表示される。次いで、ユーザは、マイクロフォン入力デバイス140を介してテキストを入力することにより、このスピーチ認識ソフトウェアアプリケーションを動作させて、ターゲットソフトウェアアプリケーションウィンドウ302を介してこのスピーチ認識ソフトウェアアプリケーションに入力済みのテキスト304を表示させることができる。この例では、一連の入力済みのテキスト304がシステム100に入力される。一連の入力済みのテキスト304は、「l」、「t」、「space(スペース)」、「i」、「s」、「space」、「t」、「h」、「e」、「space」、「i」、「n

30

40

50

「t」、「e」、「n」、「t」として入力される。しかし、このスピーチ認識ソフトウェアアプリケーションは、この一連の入力済みのテキスト304を「l」、「t」、「space」、「i」、「s」、「space」、「t」、「h」、「e」、「space」、「i」、「n」、「v」、「e」、「n」、「t」として「聞いて」おり、したがって、この最後の単語「invent」は、「intent」になるように訂正する必要がある。以上のテキストは、単語全体として、または文字（キャラクタ）として口述することができることを理解されたい。また、この口述筆記機能は、オペレーティングシステムレベルコンポーネントとすることができることを理解されたい。

【0021】

図4に示すように、これを訂正するために、スペリングコマンドが有効にされて、スペリングUIを呼び出し、ユーザは修正すべき単語、すなわち「spell intent (intentをスペリングする)」を汎用コンピュータシステム102に対して口頭で伝えることができるようになる。汎用コンピュータシステム102がこのスペリングコマンドを受け取るとすぐに、この選択された単語は、識別され、この選択された単語に少なくとも1つのスペース404を追加することによって追加済みの単語402が作成され、ここで、追加済みの単語402は、動作ブロック202（図2）に示すように、修正すべき少なくとも1つのキャラクタ406を含んでいる。次いで動作ブロック204（図2）に示すように、この追加済みの単語中の各キャラクタ408には、一意の数値410が割り当てられ、関連づけられる。次いで、追加済み単語402は、ディスプレイ画面148を介して表示されて、追加済みの単語402中の各キャラクタ408と、これらに割り当てられた一意の数値410の間の関連づけを視覚的に伝える。この関連づけは、ディスプレイ画面148上の追加済みの単語402の周囲にボックス412を描画し、追加済みの単語402中のキャラクタ408に隣接してそれらに割り当てられた一意の各数値410を表示することによって視覚的に伝えることができる。したがって、各キャラクタ408には、各キャラクタ408が関連づけられる一意の数値410が「割り当てられる」。例えば、単語「invent」414を「intent」に変更しようと望むユーザは、このスピーチ認識ソフトウェアアプリケーションに対して「spell invent (inventをスペリングする)」などのコマンドを口頭で入力する。これは、このスペリングUIを実施し、ボックス412がディスプレイ画面148上の単語「invent」414の周囲に表示されるようにするはずである。これはまた、単語「invent」414中の各キャラクタに一意の数値410が、割り当てられるようにし、この一意の数値は、その対応するキャラクタ408に隣接して表示され、これらの両方が図4に示されている。これにより、このユーザは、この単語「invent」中のどの文字を変更および/または訂正することもできるようになる。

【0022】

この時点において、選択コマンドおよび修正コマンドを、汎用コンピュータシステム102が受け取ることができる。ここでこの選択コマンドは、動作ブロック206（図2）に示すように選択されたキャラクタに対応する一意の数値410である。このことは、汎用コンピュータシステム102に対して、追加済みの単語402中のどのキャラクタを変更すべきかを示す。汎用コンピュータシステム102が、修正コマンドを受け取った後に、動作ブロック208（図2）に示すように、修正コマンドに応答して選択されたキャラクタを修正することにより、修正済みの単語が生成される。このユーザは、「削除する (delete)」、「挿入する (insert)」、追加すべき文字/キャラクタなど、複数の修正コマンドを入力することができることを理解されたい。これらの各状況については以下で説明する。

【0023】

例として、ユーザがディスプレイ画面148上に表示された追加済みの単語「invent_」504中の文字「v」502を削除したいと思う上記の場合について、図5を参照して考察する。以上で説明したように、ユーザは、変更すべきキャラクタに対応する一意の数値を汎用コンピュータシステム102に伝える。これらの一意の数値508は、数字1から開始され、1ずつ増加するが、任意の一意の数値508および増分を割り当てることが

できることを理解されたい。図から分かるように、追加済みの単語「invent_」504中の文字「v」502には、「3」という一意の数値508が割り当てられている。したがって、このユーザは、数字「3」を汎用コンピュータシステム102に対して口頭で伝えることになる。これにより、カレット(caret)510によって示されるような、数字「3」に対応し、関連づけられた文字が「選択」され、この場合における文字は、この追加済みの単語「invent_」504中の文字「v」502となる。次いで、このユーザは、「delete」など所望の修正コマンドを入力することができ、この「delete」は、追加済みの単語「invent_」504から文字「v」502を削除させることになり、図6に示すようにその結果として「inent」512が残される。したがって、修正コマンド「delete」は、追加済みの単語から選択された文字、および対応するスペースを除去することになり、選択カレット510は、次の後続のキャラクタ、すなわち「e」を選択することになる。

【0024】

他方、ユーザが、ディスプレイ画面148上に表示される単語「invent」中の文字「n」704と文字「v」706の間の、文字やスペースなどのキャラクタを挿入したいと思う場合について、図7を参照して考察する。基本的には、ユーザは、一意の数値「3」に対応する位置(spot)にキャラクタを挿入したいと思っている。以上で説明したように、ユーザは、コマンド「spell invent」を口頭で伝えることにより、スペリングUIを実行することができる。これにより、スペースが単語「invent」に追加されて、追加済みの単語「invent_」708が作成され、ボックス710が追加済みの単語「invent_」708の周囲に表示され、一意の数値712が、追加済みの単語「invent_」708中の各キャラクタ408に隣接して割り当てられ表示されるようになる。図から分かるように、追加済みの単語「invent_」708中の文字「v」706には、「3」という一意の数値712が割り当てられる。したがって、ユーザは、汎用コンピュータシステム102に対して数字「3」を口頭で伝えることになる。このことは、汎用コンピュータシステム102に、カレット714によって示される数字「3」に対応し、相互に関連づけられた文字を「選択」させる。この例では選択される文字は、追加済みの単語「invent_」708中の文字「v」706である。次いで、ユーザは、修正コマンドを入力し、汎用コンピュータシステム102を適切な方法で応答させることができる。例えば、ユーザが、修正コマンド「insert」を伝え、次いで単語「space」を伝える場合には、スペースが文字「n」704と文字「v」706の間に挿入されることになり、図8に示すように、追加済みの単語「invent_」708を「in vent_」716に効果的に変更することになる。この場合に、カレット714は、その場にとどまって、一意の数値「3」に関連づけられたスペースが選択されていることを示すはずである。しかし、このユーザが、コマンド「insert」を伝え、次いで文字「p」を伝える場合には、文字「p」が文字「n」704と文字「v」706の間に挿入されることになり、図9に示すように、追加済みの単語「invent_」を「inpvent_」718に効果的に変更することになり、選択カレット714は、次のキャラクタに移動して、次のキャラクタ(すなわち、一意の数値「4」に対応するキャラクタ)が選択されていることを示すようになる。

【0025】

同様に、ユーザが、単にディスプレイ画面148上に表示された単語「invent」中の文字を変更しようと思う場合について、図10を参照して考察する。以上で説明したように、ユーザは、コマンド「spell invent」を口頭で伝えることにより、スペリングUIを実施することができる。これにより、スペースが、単語「invent」の最後に追加されて、追加済みの単語「invent_」902が作成され、ボックス904が、追加済みの単語「invent_」902の周囲に表示され、一意の数値906が、追加済みの単語「invent_」902中の各キャラクタ908に隣接して割り当てられ、表示される。図から分かるように、追加済みの単語「invent_」902中の文字「v」912には、「3」という一意の数値906が割り当てられる。したがって、ユーザは、コンピュータシステム102に対して数字「3」を口頭で伝えることになる。これにより、数字「3」に対応し、関連づけられた文字が、カレット910によって示されるように選択される。この例では選択される文字は、

追加済みの単語「invent_」902中の文字「v」912となる。次いでユーザは、コンピュータシステムを適切な方法で応答するようにする修正コマンド（この例ではコマンドは、単に文字である）を入力することができる。例えば、ユーザが、数字「3」の後に修正コマンド「t」を伝える場合には、文字「v」912は、文字「t」で置き換えられて、図11に示すように、追加済みの単語「invent_」902を単語「intent」914に効果的に変更することになる。この時点で、選択カレット910は、次のキャラクタに移動して、次のキャラクタ（すなわち、一意の数値「4」に対応するキャラクタ）が選択されていることを示すことになる。

【0026】

ユーザが変更すべき文字に対応する一意の数値を入力した後に、ドロップダウンメニューなど、推奨される修正コマンドのメニューを表示することができる。ここで、割り当てられるはずの推奨される各アクションは、それ自体が一意の数値であることを理解されたい。例えば、ユーザが、ディスプレイ画面148上に表示される単語「invent」中の1文字を変更したいと思う場合について、図12を参照して考察する。ユーザは、コマンド「spell invent」を口頭で伝えることにより、スペリングUIを実施することができる。これにより、スペースが、選択された単語「invent」に追加されて、追加済みの単語「invent_」1002が作成され、ボックス1004が、追加済みの単語「invent_」1002の周囲に表示され、一意の数値1006が、追加済みの単語「invent_」1002中の各文字に隣接して表示されるようになる。図から分かるように、追加済みの単語「invent_」1002中の文字「v」1008には、「3」という一意の数値1006が割り当てられる。したがって、ユーザは、汎用コンピュータシステム102に対して数字「3」を口頭で伝えて、カレット1010によって示される一意の数値「3」に対応して、関連づけられた文字を「選択する」ことになり、この例では選択される文字は、追加済みの単語「invent_」1002中の文字「v」1008である。図13に示すように、ユーザにいくつかの修正コマンド選択肢を提供するメニュー1012をディスプレイ画面148上に表示することができる。これらの各選択肢には、第2の一意の数値1014が割り当てられる。次いでユーザは、スピーチ認識ソフトウェアアプリケーションに適切な方法で応答させる所望の修正コマンドに関連づけられた第2の一意の数値1014である修正コマンドを入力することができる。例えば、ユーザが、数字「3」の後に数字「4」を伝える場合には、文字「v」1008は、文字「d」1016で置き換えられ、図14に示すように、効果的に追加済みの単語「invent_」1002を単語「indent」1018に変更することになる。上記と同様に、選択カレット1010は、次のキャラクタに移動して、次のキャラクタ（すなわち、一意の数値「4」に対応するキャラクタ）が選択されていることを示すことになる。

【0027】

推奨される修正コマンドのメニュー1012は、選択された文字または単語と音響的に同様な響きのキャラクタ/単語のメニューなど、所望の最終目的に適した任意の修正コマンドを含むことができること、例えば、「v」が選択される場合には、メニュー1012は、「d」、「t」、「e」、「g」、「3」を含むことになることを理解されたい。さらに、メニュー1012は、これらのキャラクタの大文字、例えば、「V」、ならびにスペルチェッカからのオートコンプリートのリストを含むこともできる。したがって、この実施例では、メニュー1012は、単語「indent」、「intent」、「amend」を含むことができる。さらに、本発明は、複数のファンクションを同時に実施する音声コマンド、例えば「Change 3 to "e" as in eagle」を含むこともでき、あるいは選択された単語中に文字「t」が1つしかない場合には、「Change t to g」は、文字「t」を文字「g」に変更することができることを理解されたい。選択された単語中に2つの文字「t」が存在した場合には、さらに正確にするためにこのユーザに対してフィードバックを行うことができる。さらに、変更されたキャラクタを元の状態に戻すことができる「アンドウ（undo）」など他のコマンドも提供することができ、例えば、ユーザが、（選択された文字を大文にするために）「cap that」と発声するが、この入力「caret（カレット）」と認識された場

合には、ユーザは、「アンドゥ」と発声して、この文字をその元の状態に戻すこともできる。

【0028】

例示の実施形態によれば、機械読取り可能コンピュータプログラムに応答して動作するコントローラにより、図2の処理の全体または一部を実施することができる。したがって、所定のファンクションおよび所望の処理、ならびに計算（例えば、1つ（または複数）の実行制御アルゴリズム、本明細書中で説明している制御プロセスなど）を実施するために、コントローラは、それだけには限定されないが、1つ（または複数）のプロセッサ、1つ（または複数）のコンピュータ、メモリ、ストレージ、1つ（または複数）のレジスタ、タイミング、1つ（または複数）の割込み、1つ（または複数）の通信インターフェース、および1つ（または複数）の入出力信号インターフェース、ならびに前述のうちの少なくとも1つを含む組合せを含むことができる。

10

【0029】

さらに、本発明は、コンピュータまたはコントローラに実装されたプロセスの形態で、実施することができる。本発明はまた、フロッピー（登録商標）ディスク、CD-ROM、ハードドライブ、および/または他の任意のコンピュータ読取り可能媒体などの有形媒体中に実施される命令を含むコンピュータプログラムコードの形態で実施することもできる。ここで、このコンピュータプログラムコードが、コンピュータまたはコントローラにロードされ実行されると、このコンピュータまたはコントローラは、本発明を実行するための装置になる。本発明はまた、例えば、ストレージ媒体に記憶され、コンピュータまたはコントローラによってロードおよび/または実行され、あるいは電気配線またはケーブル配線上、光ファイバを介して、電磁放射線を介してなど、何らかの伝送媒体上で伝送されるかにかかわらず、コンピュータプログラムコードの形態で実施することもできる。ここで、このコンピュータプログラムコードが、コンピュータまたはコントローラにロードされ実行されると、このコンピュータまたはコントローラは、本発明を実行するための装置になる。汎用マイクロプロセッサ上で実行されると、このコンピュータプログラムコードセグメントは、マイクロプロセッサを設定して、特定の論理回路を作り出すことができる。

20

【0030】

例示の実施形態に関して本発明を説明してきたが、本発明の趣旨および範囲を逸脱することなく、様々な変更、省略、および/または追加を行うことができ、また本発明の要素について同等物で代用することができる。したがって、本発明は、本発明を実行するように企図された最良の態様として開示された特定の実施形態だけには限定されず、本発明は、添付の特許請求の範囲の範囲内に含まれるすべての実施形態を包含することになることを意図している。さらに、特に明記されない限り、第1の、第2のなど、これらの用語のどのような使用も、どのような順序または重要度を示すものでもなく、そうでなくて、第1の、第2のなど、これらの用語は、1つの要素を別の要素から区別するために使用されている。

30

40

【図面の簡単な説明】

【0031】

【図1】例示の実施形態にしたがって、スペリングユーザインターフェース（UI）を有するスピーチ認識ソフトウェアアプリケーションを使用して、ディスプレイ画面上に表示されるキャラクタを操作する方法を実施するシステムのブロック図である。

【図2】例示の実施形態にしたがって、スペリングユーザインターフェース（UI）を有するスピーチ認識ソフトウェアアプリケーションを使用して、ディスプレイ画面上に表示されるキャラクタを操作する方法を示すブロック図である。

【図3】図2の方法を説明するための、図1におけるシステムのディスプレイ画面である。

50

【図 4】図 2 の方法を説明するための、図 1 におけるシステムのディスプレイ画面である。

【図 5】図 2 の方法を説明するための、図 1 におけるシステムのディスプレイ画面である。

【図 6】図 2 の方法を説明するための、図 1 におけるシステムのディスプレイ画面である。

【図 7】図 2 の方法を説明するための、図 1 におけるシステムのディスプレイ画面である。

【図 8】図 2 の方法を説明するための、図 1 におけるシステムのディスプレイ画面である。

【図 9】図 2 の方法を説明するための、図 1 におけるシステムのディスプレイ画面である。

【図 10】図 2 の方法を説明するための、図 1 におけるシステムのディスプレイ画面である。

【図 11】図 2 の方法を説明するための、図 1 におけるシステムのディスプレイ画面である。

【図 12】図 2 の方法を説明するための、図 1 におけるシステムのディスプレイ画面である。

【図 13】図 2 の方法を説明するための、図 1 におけるシステムのディスプレイ画面である。

【図 14】図 2 の方法を説明するための、図 1 におけるシステムのディスプレイ画面である。

【符号の説明】

【0032】

147 ディスプレイデバイス

148 ディスプレイ画面

302 ターゲットソフトウェアアプリケーションウィンドウ

304 入力済みのテキスト

402 追加済みの単語

404 スペース

406 修正すべきキャラクタ

408 追加済みの単語中の各キャラクタ

410 追加済みの単語の各キャラクタに割り当てられた一意の数値

412 ボックス

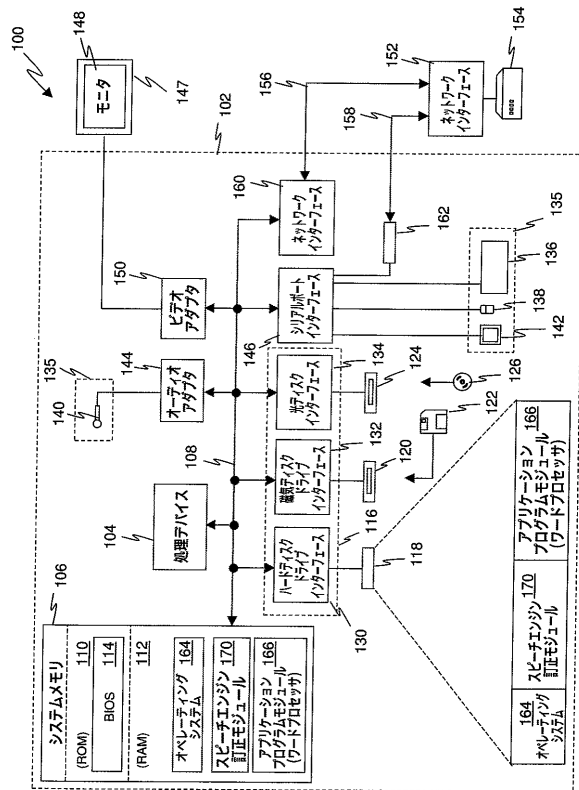
414 単語

10

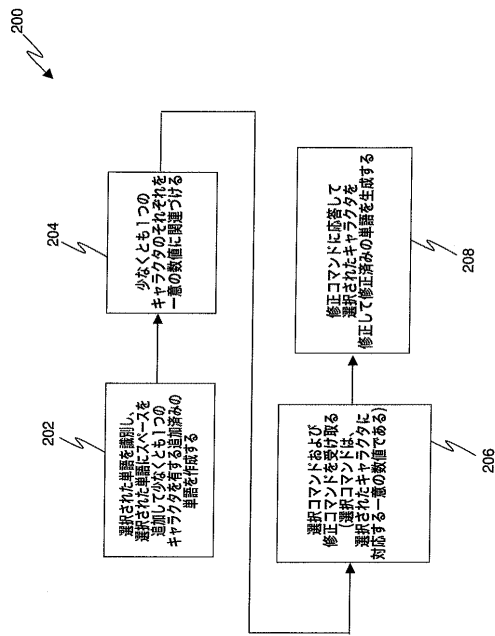
20

30

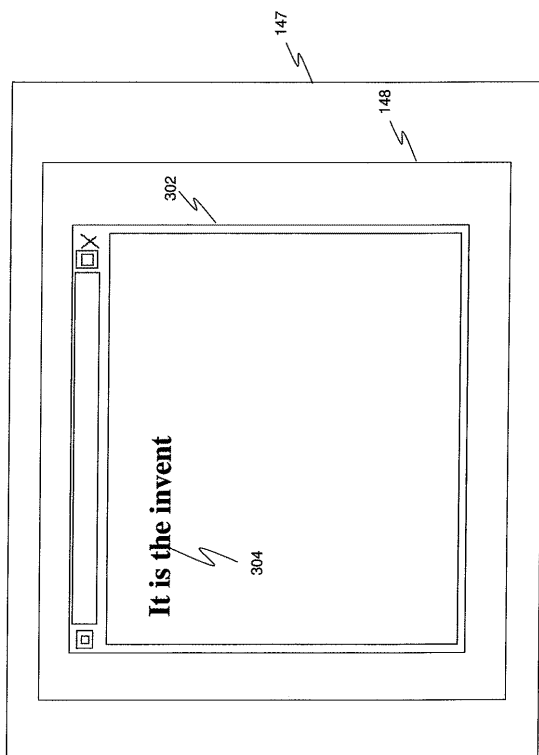
【図 1】



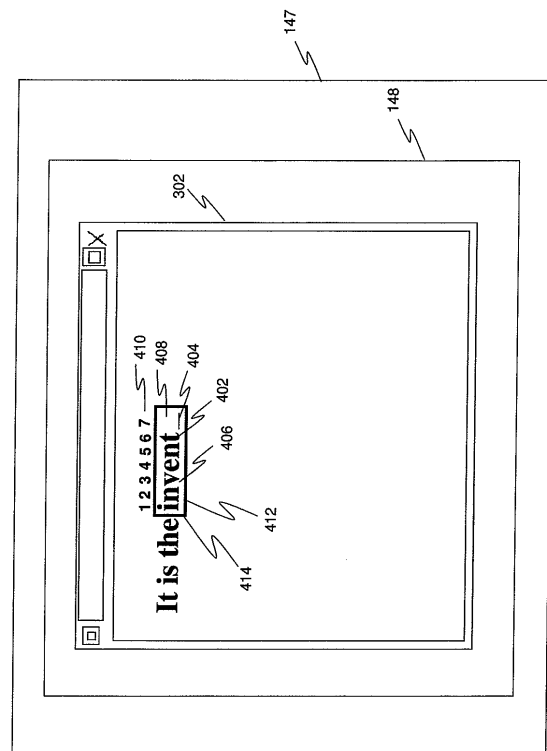
【図 2】



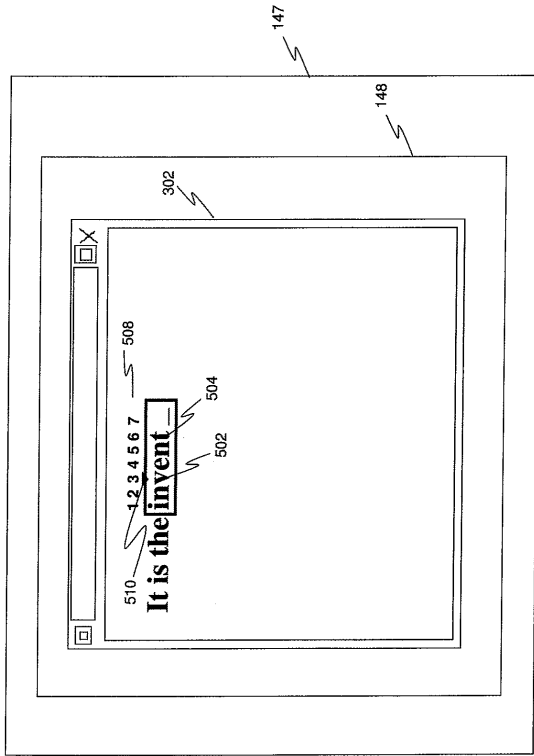
【図 3】



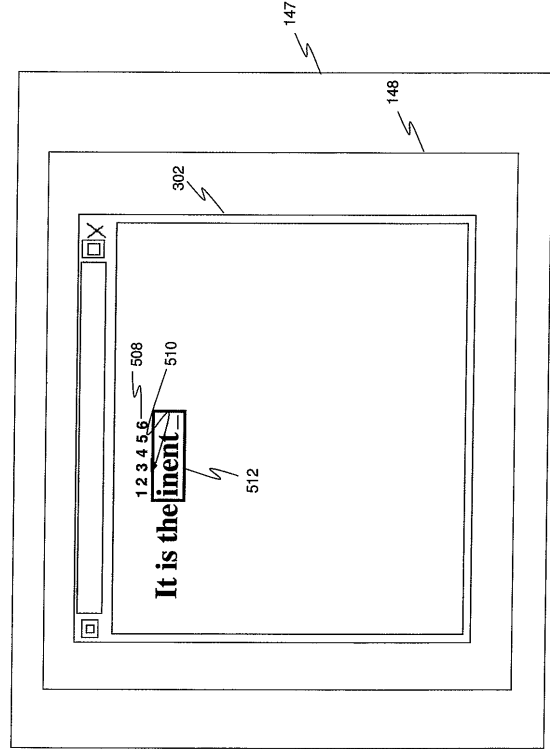
【図 4】



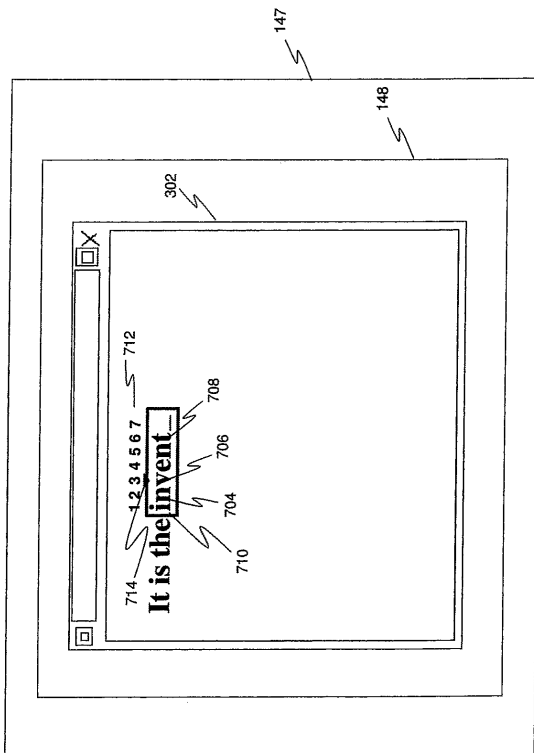
【 図 5 】



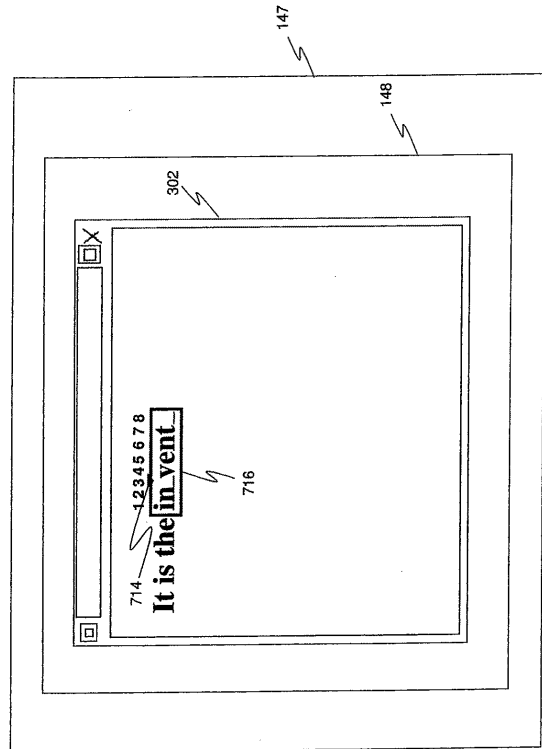
【 図 6 】



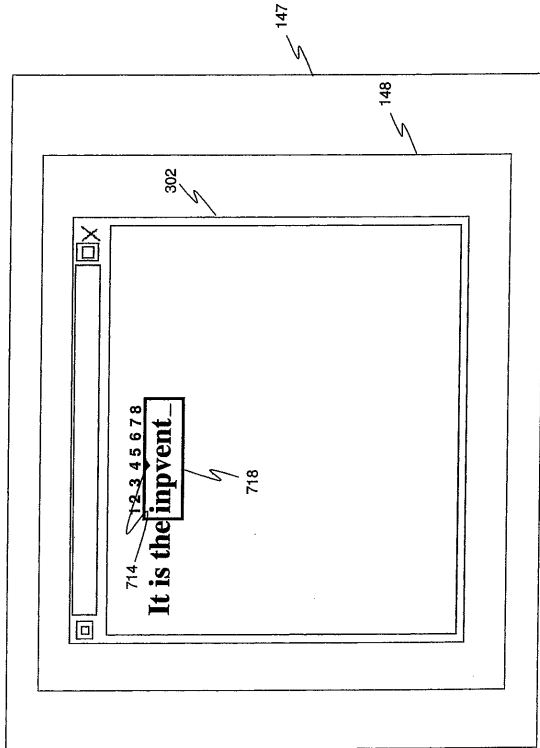
【 図 7 】



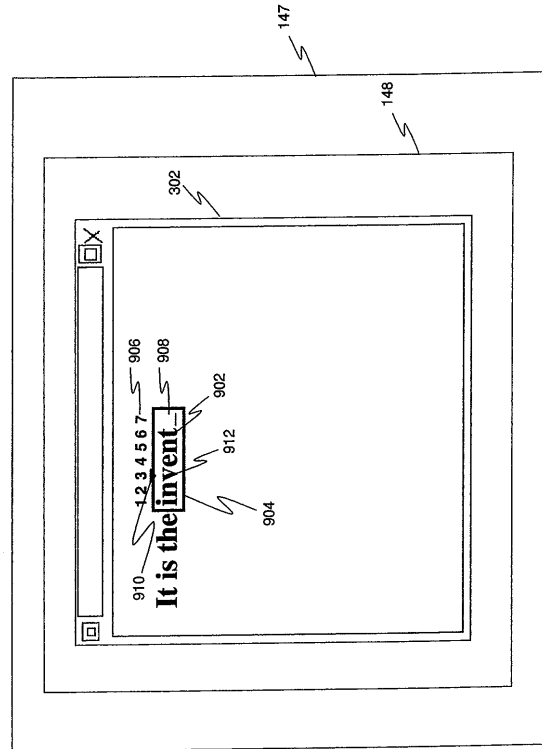
【 図 8 】



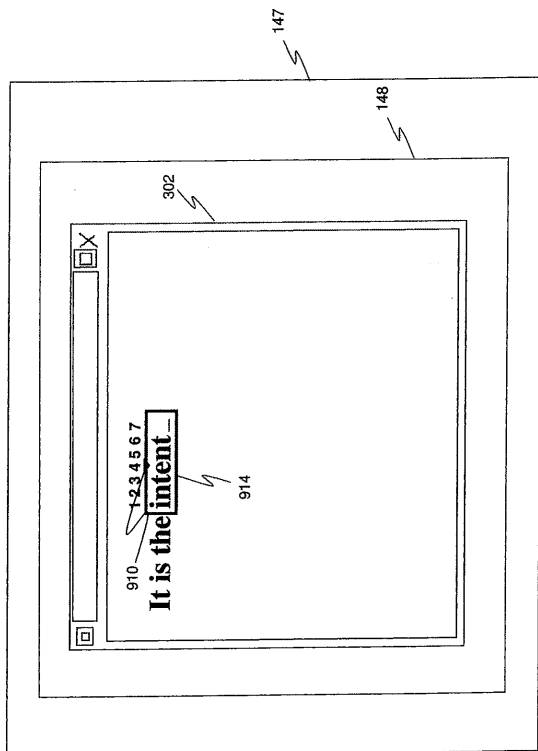
【図 9】



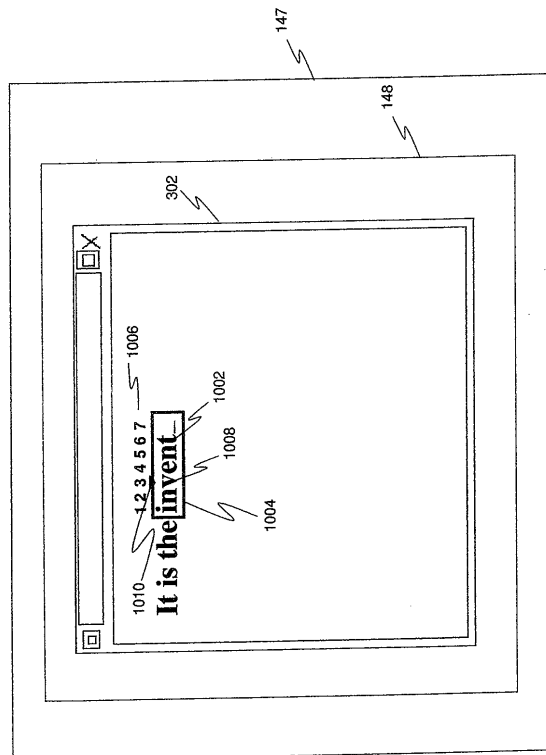
【図 10】



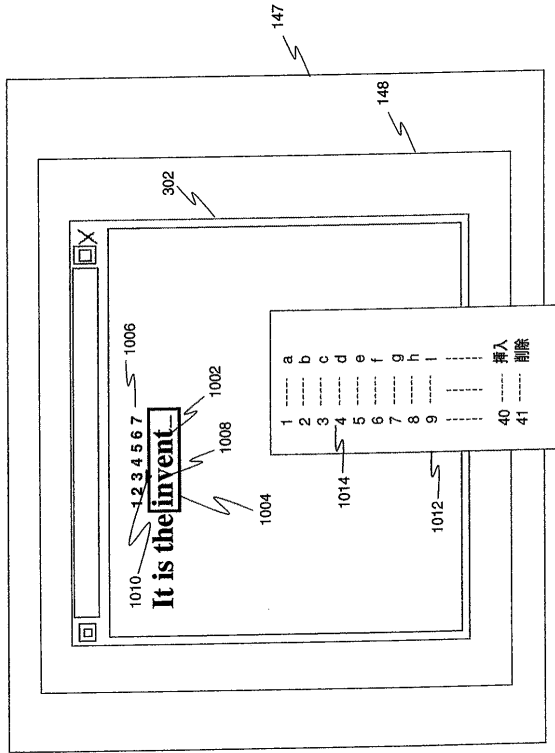
【図 11】



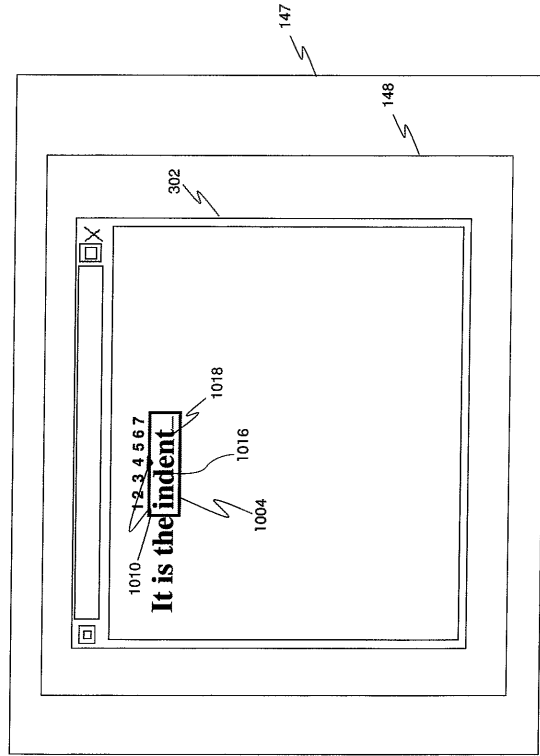
【図 12】



【図 13】



【図 14】



フロントページの続き

(72)発明者 フェリックス ジー ティー アイ アンドリュー
アメリカ合衆国 9 8 0 5 2 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション内

(72)発明者 ロバート エル . チャンバース
アメリカ合衆国 9 8 0 5 2 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション内

F ターム(参考) 5B009 KB01 QB03 QB05 QB11 VB02
5D015 LL04