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

特許第4738847号
(P4738847)

(45)発行日　平成23年8月3日(2011.8.3)　(24)登録日　平成23年5月13日(2011.5.13)

(51)Int.Cl.　F I
*G10L 15/00 (2006.01)*　G10L 15/00 200T
*G10L 15/10 (2006.01)*　G10L 15/10 300G

請求項の数 12　(全 18 頁)

(21)出願番号　特願2005-63149(P2005-63149)
(22)出願日　平成17年3月7日(2005.3.7)
(65)公開番号　特開2006-243673(P2006-243673A)
(43)公開日　平成18年9月14日(2006.9.14)
審査請求日　平成20年2月19日(2008.2.19)

(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　100076428
弁理士　大塚　康徳
(74)代理人　100112508
弁理士　高柳　司郎
(74)代理人　100115071
弁理士　大塚　康弘
(74)代理人　100116894
弁理士　木村　秀二
(72)発明者　山本　寛樹
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】データ検索装置および方法

(57)【特許請求の範囲】
【請求項１】
検索対象のデータがそれぞれ音声データと関連付けられて記憶されたデータベースから、ユーザにより入力されたキーワードを基にデータを検索するデータ検索装置であって、
前記データベース内の各データに関連付けられた音声データに対し音声認識を行い、サブワード表現形式で認識結果を出力する第１の音声認識手段と、
前記キーワードをサブワード表現形式に変換する変換手段と、
前記第１の音声認識手段により得られたサブワード表現形式の前記認識結果と、前記変換手段によりサブワード表現形式に変換された前記キーワードとに基づいて、前記キーワードと前記データベース内の各データに関連付けられた音声データとの類似度を計算する類似度計算手段と、
前記類似度計算手段により計算された前記類似度に基づき選択される１または２以上のデータの各々について、そのデータに関連付けられた音声データを入力とし、前記サブワード表現形式に変換された前記キーワードを認識対象語とする音声認識を行う第２の音声認識手段と、
前記第２の音声認識手段の認識スコアに基づいて検索スコアを計算する検索スコア計算手段と、
前記検索スコア計算手段により計算された前記検索スコアに基づいて選択される前記データベース内のデータを検索結果としてユーザに提示する検索結果提示手段と、
を有し、

前記検索スコア計算手段は、前記検索スコアとして、前記類似度計算手段により計算された類似度と前記第２の音声認識手段により得られた認識スコアとの重み付き和を計算する
ことを特徴とするデータ検索装置。
【請求項２】
前記第１の音声認識手段および前記変換手段は、前記キーワードが入力される前にあらかじめ実行されるものであり、前記類似度計算手段、前記第２の音声認識手段、前記検索スコア計算手段、および前記検索結果提示手段は、前記キーワードが入力されたことに応じて動作することを特徴とする請求項１に記載のデータ検索装置。
【請求項３】
前記類似度計算手段は、前記類似度として、前記変換手段によりサブワード表現形式に変換された前記キーワードを正解とする前記第１の音声認識手段により得られたサブワード表現形式の前記認識結果のサブワード正解率またはサブワード正解精度を計算することを特徴とする請求項１または２に記載のデータ検索装置。
【請求項４】
前記サブワード正解精度は、正解サブワード数から挿入誤りサブワード数、置換誤りサブワード数、および削除誤りサブワード数をそれぞれ引いて得たサブワード数と、前記正解サブワード数との比でもって表されるものであって、前記挿入誤りサブワード数に所定の重み係数が乗じられることを特徴とする請求項３に記載のデータ検索装置。
【請求項５】
前記サブワードは、音素または音節であることを特徴とする請求項１から４までのいずれか１項に記載のデータ検索装置。
【請求項６】
前記第２の音声認識手段により実行される音声認識は、前記キーワードを認識対象語とするキーワードスポッティングであることを特徴とする請求項１から５までのいずれか１項に記載のデータ検索装置。
【請求項７】
前記第２の音声認識手段は、前記類似度が大きい順に所定個数のデータを選択し、当該選択されたデータの各々について前記音声認識を行うことを特徴とする請求項１から６までのいずれか１項に記載のデータ検索装置。
【請求項８】
前記第２の音声認識手段は、前記類似度が所定の値よりも大きい１または２以上のデータを選択し、当該選択されたデータの各々について前記音声認識を行うことを特徴とする請求項１から６までのいずれか１項に記載のデータ検索装置。
【請求項９】
前記検索結果提示手段は、前記検索スコアが大きい順に所定個数のデータを検索結果として表示することを特徴とする請求項１から８までのいずれか１項に記載のデータ検索装置。
【請求項１０】
前記検索結果提示手段は、前記検索スコアが所定の値よりも大きいデータを検索結果として表示することを特徴とする請求項１から８までのいずれか１項に記載のデータ検索装置。
【請求項１１】
検索対象のデータがそれぞれ音声データと関連付けられて記憶されたデータベースから、ユーザにより入力されたキーワードを基にデータを検索するデータ検索装置によって実行されるデータ検索方法であって、
第１の音声認識手段が、前記データベース内の各データに関連付けられた音声データに対し音声認識を行い、サブワード表現形式で認識結果を出力する第１の音声認識ステップと、
変換手段が、前記キーワードをサブワード表現形式に変換する変換ステップと、

類似度計算手段が、前記第１の音声認識ステップにより得られたサブワード表現形式の前記認識結果と、前記変換ステップによりサブワード表現形式に変換された前記キーワードとに基づいて、前記キーワードと前記データベース内の各データに関連付けられた音声データとの類似度を計算する類似度計算ステップと、

第２の音声認識手段が、前記類似度計算ステップにより計算された前記類似度に基づき選択される１または２以上のデータの各々について、そのデータに関連付けられた音声データを入力とし、前記サブワード表現形式に変換された前記キーワードを認識対象語とする音声認識を行う第２の音声認識ステップと、

検索スコア計算手段が、前記第２の音声認識ステップでの認識スコアに基づいて検索スコアを計算する検索スコア計算ステップと、

検索結果提示手段が、前記検索スコア計算ステップにより計算された前記検索スコアに基づいて選択される前記データベース内のデータを検索結果としてユーザに提示する検索結果提示ステップと、

を有し、

前記検索スコア計算ステップでは、前記検索スコア計算手段が、前記検索スコアとして、前記類似度計算ステップで計算された類似度と前記第２の音声認識ステップで得られた認識スコアとの重み付き和を計算する

ことを特徴とするデータ検索方法。

【請求項１２】

請求項１１に記載のデータ検索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、データ検索技術に関し、特に、音声認識を用いて、音声情報が付与されたマルチメディアデータを検索するデータ検索技術に関する。

【背景技術】

【０００２】

音声はキー入力に不慣れな人やキーボードを設置するのが物理的に困難な小型機器での情報入力手段として有効である。現在市販されているデジタルカメラの多くは録音機能を搭載しており、撮影した画像に音声でメモをつけることができる。また、ＩＣレコーダの利用事例として、ＩＣレコーダを携帯し、日常のメモ代わりに音声メモ機能を利用する事例も増えている。これらの録音された音声は単に再生するにとどまらず、音声認識を利用して検索に利用することができる。

【０００３】

音声認識を利用した音声データの検索のタイプは次の二通りが考えられる。

【０００４】

（１）あらかじめ検索に用いるキーワード（あるいはフレーズ）が決められているもの。

（２）任意のキーワード（あるいはフレーズ）を使用できるもの。

【０００５】

（１）の場合、検索対象となる音声データに対して、あらかじめ決められたキーワードを含んだ認識辞書あるいは認識文法を用いて音声データを音声認識した認識結果を記憶しておき、ユーザが指定した検索用キーワードと一致あるいは部分一致する認識結果となる音声データを検索結果として提示する。音声認識の方法としてキーワードスポッティングを用いて音声データからキーワードを抽出してもよい。この方法の利点は、検索に先立って音声入力時に音声認識を実行できる点である。計算量の多い音声認識処理を事前に行っておけば、認識結果と検索用キーワードとの文字列同士の比較という比較的計算量の少ない処理だけで検索できる。しかしながら、認識辞書あるいは認識文法に含まれるあらかじめ決められたキーワードしか音声認識できないため、付与する音声の内容が制限され、音声メモの付与・検索という活用目的で利用する場合にはその活用範囲が制限されることになる。この方法を実現した従来技術の一例が、特開２００３－２１９３２７公報（特許文

献１）に開示されている。
【０００６】
（２）の場合、検索時にユーザが入力したキーワード（あるいはフレーズ）を認識辞書あるいは認識文法を用いて音声データを音声認識し、ユーザが指定した検索用キーワードと一致あるいは部分一致する認識結果となる音声データを検索結果として提示する。音声認識の方法としてキーワードスポッティングを用いて、ユーザが指定した任意のキーワードが音声データ内に含まれるかどうかを判定しても良い。この方法を実現した従来技術の一例が、特開平１０－１７３７６９号公報（特許文献２）に開示されている。この特許文献２に開示された音声メッセージ検索装置では、ユーザが入力したキーワード文字列を不特定話者の標準パターンに変換して、この標準パターンを用いてキーワードスポッティング音声認識を行い、音声メッセージ中でキーワードが検出された箇所が再生される。この方法の利点は、任意の内容を音声入力でき、任意のキーワードで検索できる点である。しかしながら、検索用のキーワードを入力してから音声認識を実行するので、検索時に計算量の多い音声認識を実行する必要があり、検索対象とする音声データが大量にある場合は検索に時間がかかるという問題がある。
【０００７】
特許文献１および２の欠点を解消する検索方法の一例が、特開２００２－２７８５７９公報（特許文献３）に開示されている。特許文献３に開示されている音声データ検索装置では、検索対象となる音声データ（特許文献３では「音声波形データ」）を音声シンボル列に変換したものを記録しておき、キーワード（特許文献３中では「検索語」）を音声シンボル列に変換して、記憶されている音声データを変換した音声シンボル列と一致する部分を検索し、検索された部分に対してキーワードと一致しているか否かを判定する。この判定にキーワードスポッティング音声認識装置を用いることが特許文献３の請求項４に記載されている。この方法では、検索対象となる音声データの音声シンボル列への変換を検索に先立って事前に行うことにより、検索時にシンボルレベルでの高速な候補の絞込みが実現され、さらに絞り込まれた候補に対してより正確な音声認識を用いた検索を実施することで、任意のキーワードに対して、正確で従来よりも高速な検索が実現できる。
【０００８】
【特許文献１】特開２００３－２１９３２７公報
【特許文献２】特開平１０－１７３７６９号公報
【特許文献３】特開２００２－２７８５７９公報
【発明の開示】
【発明が解決しようとする課題】
【０００９】
特許文献３の方法では、シンボルレベルでのマッチングの際にキーワードと一致した部分のみを候補として絞り込むため、音声データを音声シンボル列に変換する際に誤った音声シンボル列に変換された候補は検索できないという問題がある。
【００１０】
特許文献３には、音声データから音声シンボル列への変換に音素認識結果を用いた例が開示されている。日本語の音素認識の場合、一般に７０～８０％の認識率で性能が良いとされている。つまり、いかに高性能な音素認識を用いても、音素認識結果には誤りが含まれることになる。特許文献３には、この対策として、音素認識結果から一般に誤りやすいとされる子音を無視して有音母音、長母音、発音、無音のみの音声シンボル列へ変換する方法や、コンフュージョンを起こしやすい子音をグループ化して１つの音節として扱う方法が開示されている。前者は例えば、「ｈａｋｏｎｅ（ハコネ）」という音素認識結果に対して、「ａａｅ（アアエ）」という音声シンボル列に変換する。後者は、例えばｋとｐとｔを同じグループとしてｔに置き換え、「ｈａｋｏｎｅ（ハコネ）」を「ｈａｔｏｎｅ（ハトネ）」にマッピングする。入力されたキーワードにも同様のマッピングを行うので、子音の音素認識を誤っても検索することができる。
【００１１】

しかしながら、音素認識誤りは子音だけではなく母音にもしばしば起こる問題であるし、音素認識誤りは置換誤りだけではなく、脱落誤り、挿入誤りなども考えられる。前述の「ｈａｋｏｎｅ（ハコネ）」の例では、例えば前述の「ハコネ」という音声データに対して、「ｈｏｋｏｎｅ（ホコネ）」（母音の認識誤り）、「ｈａｋｏｎｅｅ（ハコネー）」（挿入誤り）、「ｋｏｎｅ（コネ）」（脱落誤り）といった音素誤りが生じた場合、特許文献３に開示されている方法でキーワードを「ハコネ」として検索しても、検索できないという問題がある。これらの問題は、シンボルレベルで検索して候補を絞り込む際に、音声データの音声シンボル列からキーワードの音声シンボル列と一致する部分を検索するために生じている。

【００１２】

本発明は、このような問題点に鑑みてなされたものであり、音声認識を用いたデータ検索における検索精度を高めることを目的とする。

【課題を解決するための手段】

【００１３】

上記目的を達成するために、例えば本発明のデータ検索装置は以下の構成を備える。すなわち、本発明の一側面に係るデータ検索装置は、検索対象のデータがそれぞれ音声データと関連付けられて記憶されたデータベースから、ユーザにより入力されたキーワードを基にデータを検索するデータ検索装置であって、前記データベース内の各データに関連付けられた音声データに対し音声認識を行い、サブワード表現形式で認識結果を出力する第１の音声認識手段と、前記キーワードをサブワード表現形式に変換する変換手段と、前記第１の音声認識手段により得られたサブワード表現形式の前記認識結果と、前記変換手段によりサブワード表現形式に変換された前記キーワードとに基づいて、前記キーワードと前記データベース内の各データに関連付けられた音声データとの類似度を計算する類似度計算手段と、前記類似度計算手段により計算された前記類似度に基づき選択される１または２以上のデータの各々について、そのデータに関連付けられた音声データを入力とし、前記サブワード表現形式に変換された前記キーワードを認識対象語とする音声認識を行う第２の音声認識手段と、前記第２の音声認識手段の認識スコアに基づいて検索スコアを計算する検索スコア計算手段と、前記検索スコア計算手段により計算された前記検索スコアに基づいて選択される前記データベース内のデータを検索結果としてユーザに提示する検索結果提示手段とを有し、前記検索スコア計算手段は、前記検索スコアとして、前記類似度計算手段により計算された類似度と前記第２の音声認識手段により得られた認識スコアとの重み付き和を計算することを特徴とする。

【発明の効果】

【００１４】

本発明によれば、音声認識を用いたデータ検索における検索精度を高めることができる。

【発明を実施するための最良の形態】

【００１５】

以下、図面を参照して本発明の好適な実施形態について詳細に説明する。

【００１６】

（実施形態１）

本実施形態では、データ検索装置の一例として、画像データに関連付けられた音声データを用いて画像データを検索する画像データ検索装置について説明する。なお、本発明に係る検索の対象は画像データに限定されるものではなく、文書、図形などその他の種類のデータにも適用が可能である。

【００１７】

図１は、本実施形態における画像データ検索装置の構成を示すブロック図である。

【００１８】

図１において、１０１は制御メモリ（ＲＯＭ）、１０２は中央処理装置（ＣＰＵ）、１０３はメモリ（ＲＡＭ）、１０４はハードディスクなどの外部記憶装置、１０５はキー

やボタンなどの入力装置、１０６は液晶などの表示装置、１０７はバスである。画像データ検索処理を実現するための制御プログラム１０４ａやその制御プログラムで用いるデータ（後述する言語処理用データ２１０、マルチメディアデータ２１１、音声認識用データ２１３）やサブワード認識結果２１２は、例えば外部記憶装置１０４に記憶される。このような構成であるから、本画像データ検索装置は汎用のコンピュータによっても実現することが可能である。

【００１９】

これらの制御プログラムやデータは、中央処理装置１０２の制御のもと、バス１０７を通じて適宜メモリ１０３に取り込まれ、中央処理装置１０２によって実行される。言うまでもないことであるが、制御プログラムやデータは制御メモリ１０１に記憶してもよい。

【００２０】

図２は、本実施形態における画像データ検索装置の機能構成を示すブロック図である。

【００２１】

制御部２０１は、他の機能モジュール間の連携やシステム全体の処理を制御する。キーワード取得部２０３は、画像検索時にユーザが入力装置１０５から入力するキーワードを取得する。キーワード・サブワード変換部２０３は、後述の言語処理用データ２１０を参照してキーワード取得部２０２で取得したキーワードを後述するサブワード音声認識部２０５で用いるサブワード表現形式に変換する。「サブワード」とは、単語を構成する単語よりも小さい音声の単位の総称であり、音声認識では音節や音素を用いる場合が多い。

【００２２】

言語処理用データ２１０は、言語処理用辞書に含まれ、単語からサブワードに変換する際に必要なデータが記録されている。例えば、検索キーワードとして漢字やかなが混在した文字列が入力された場合に、キーワード・サブワード変換部２０３では、漢字の読みを推測し、推測された読みからかな文字列に変換し、かな文字列からサブワードに変換するといった処理を行うが、このときに必要な漢字に対する読みやかな文字からサブワードへの変換ルールなどが言語処理用データ２１０として記録されている。また、英語をはじめとする日本語以外の言語の場合は、単語に対応するサブワードを記述した変換テーブルを直接記述したテーブルを言語処理用データ２１０として用いるのが一般的である。

【００２３】

なお、以後、説明の簡略化のため、ユーザが入力したキーワードをサブワード表現形式に変換したものを「クエリサブワード」と記述する。

【００２４】

以下では、サブワードとして音節を用いた場合について説明する。キーワード取得部２０３で取得したキーワードが「箱根山」であった場合、キーワード・サブワード変換部により、「箱根山」は「は こ ね や ま」というクエリサブワードに変換される。

【００２５】

音声データ取り込み部２０４は、検索対象であるマルチメディアデータ２１１に記憶された画像データに関連付けられた音声データを取り込む。

【００２６】

本実施形態における画像データ検索装置が検索対象とするマルチメディアデータ２１１は、音声が関連付けられた画像データであり、データベースとして外部記憶装置１０４に記憶されている。画像データと関連する音声データは一つのデータとして統合されていても良いし、それぞれ独立したデータとして記憶し、画像データと音声データの関連を管理する別のデータを記憶しておいても良い。また、関連する音声データと画像データの拡張子部分を除いたファイル名を同じにして画像データと音声データの関連を管理しても良い。

【００２７】

サブワード認識部２０５は、検索対象となる全ての画像データについて、音声取り込み部２０４で取り込んだ音声を入力とし、サブワード表現形式 の認識結果を出力する音声認識を行い、得られた認識結果をサブワード認識結果２１２に記憶する。ここで用いる音

声認識は、認識結果としてサブワード表現形式で記述されたものが得られれば良いので、音素タイプライタや音節タイプライタなどサブワード表現形式の認識結果を出力するサブワードを単位とした音声認識を行っても良いし、一般にディクテーションとして知られる大語彙連続音声認識を行って得られた認識結果をサブワードに変換するようにしてもよい。

【００２８】

サブワード音声認識は、各画像データごとに実行し、認識結果を記憶する際に画像データに関連付けて記憶する。画像データと認識結果を関連付ける方法は、前述した画像データと音声データの関連付けと同様の方法で実現できる。また、サブワード音声認識では、単一の認識結果だけでなく複数の認識結果候補を求めても良い。また、図１および２では、サブワード認識結果２１２と、画像および音声データを記憶するマルチメディアデータ２１１とを別々に図示しているが、サブワード認識結果を画像データや音声データと同様にマルチメディアデータ２１１に記憶しても良い。サブワード音声認識の際に必要な音響モデルをはじめとする各種データは音声認識用データ２１３に記憶されている。本実施形態の画像データ検索装置では、サブワード音声認識として音節認識を行い、認識スコアの高い上位３候補をサブワード認識結果２１２に記憶するものとする。

【００２９】

図４に、「箱根山」という音声データに対して得られるサブワード認識結果の一例を示す。本実施形態では、サブワード認識結果として認識スコアの上位３候補を求め、４０１で示される順位、４０２で示される認識結果の音節列、４０３で示される認識スコアを、認識スコアの順に記憶する。

【００３０】

サブワード類似度計算部２０６は、サブワード音声認識部２０５で得られたサブワード認識結果２１２とクエリサブワードを比較し、サブワード類似度を計算する。サブワード類似度は、マルチメディアデータ記憶部に記憶された各画像データごとに計算する。本実施形態の画像データ検索装置では、サブワード認識結果２１２に記憶された複数の音声認識候補について、クエリサブワードを正解とした場合の音節正解精度を求め、その最大値をサブワード類似度して計算する。図４に示したサブワード認識結果に対して、クエリサブワードが「は こ ね や ま」であった場合は、第１位（４１１）、第２位（４１２）、第３位（４１３）の音節認識結果に対して、それぞれクエリサブワードを正解とするサブワード正解精度としての音節正解精度を求める。音節正解精度は、クエリサブワードを正解とした場合に、正解に含まれる音節数をＮ、サブワード音声認識結果に含まれる挿入誤り音節数をＩ、置換誤り音節数をＳ、削除誤りをＤとすると、（（Ｎ－Ｉ－Ｓ－Ｄ）／Ｎ）×１００で計算される。４１１に示した「ふぁ お ね や ま あ」を例にとると、正解音節数Ｎ＝５、挿入誤り音節数Ｉ＝１（「あ」）、置換誤り音節数Ｓ＝２（「ふぁ」、「お」）、削除誤り音節数Ｄ＝０であるので、（（５－１－２－０）／５）×１００＝４０（％）となる。

【００３１】

同様に、第２位（４１２）の、「ふぁ お ね や ま」（Ｎ＝５、Ｉ＝０、Ｓ＝２、Ｄ＝０）は６０（％）、第３位（４１３）、「ふぁ こ ね や ま」（Ｎ＝５、Ｉ＝０、Ｓ＝１、Ｄ＝０）は８０（％）となる。したがって、これらの最大値である８０（％）を、クエリサブワードと図４に示したサブワード認識結果との類似度とする。

【００３２】

以上のサブワード類似度計算部２０５におけるサブワード類似度の計算においては、サブワード音声認識で得られた複数のサブワード認識結果候補を用いたが、単に第１位のサブワード認識結果だけを用いても良い。また、複数のサブワード認識結果に対する音節正解精度の最大値を類似度として用いたが、各順位の音節正解精度に対して、順位（４０１）や認識スコア（４０３）に基づいて重み付けした値を求め、その最大値、あるいは和を類似度として用いても良い。また、音節正解精度ではなく、音節正解率や他の尺度を使って計算しても良い。また、言うまでもないことであるが、サブワード音声認識部２０５が

出力するサブワード認識結果の単位、音素その他のサブワードである場合には、サブワード類似度はその単位を基に計算される方法（音素正解精度など）を用いて計算することになる。

【００３３】

キーワードスポッティング音声認識部２０７は、音声データ取り込み部２０４が取得した音声データを入力し、キーワード・サブワード変換部２０３でキーワードをサブワード表現形式に変換したクエリサブワードを認識対象語とするキーワードスポッティングを行い、認識スコアを求める。具体的には、音声認識用データ２１３に記憶されている音響モデルを用いて、クエリサブワードを表す音響モデルを構成し、この音響モデルを用いて、音声データ取り込み部２０４が取得した音声データを入力した際に計算される認識スコアを求める。キーワードスポッティングに必要な音響モデルをはじめとする各種データは音声認識用データ２１３に記憶されている。キーワードスポッティング音声認識はサブワード類似度計算部２０６で計算されたサブワード類似度に基づいて選択された画像データについてのみ行う。例えば、サブワード類似度が大きい所定個数の画像データを選択しても良いし、所定の閾値を越える画像データを選択しても良い。

【００３４】

検索スコア計算部２０８は、キーワードスポッティング音声認識で求まる認識スコアを基に、検索スコアを求める。本実施形態における画像データ検索装置では、単にワードスポッティング音声認識で求まる認識スコアを検索スコアとして用いるが、これに加えて、サブワード類似度計算部２０６で計算されたサブワード類似度や、サブワード認識結果２１２に記憶されているサブワード音声認識で求まった認識スコアなどを組み合わせて検索スコアとして用いても良い。例えば、サブワード類似度をＡ、ワードスポッティング音声認識で求まる認識スコアをＢ、αを実験的に求める定数として、Ａ＋α×Ｂのような重み付き和を検索スコアとしてもよい。

【００３５】

検索結果表示部２０９は、検索スコア計算部２０８で求まった検索スコアに基づいて選択した画像データを表示装置１０６に表示する。この際、検索スコアの最も良い画像データのみを表示してもよいし、検索スコアが所定の値を超える画像データを表示しても良いし、検索スコアの良い所定個数の画像データを表示しても良い。また、表示の際に検索スコア順に並べ替えて表示しても良いし、検索スコアにかかわらず、ファイル名など画像データに付随したほかの情報を基準に並べえて表示しても良い。

【００３６】

以上説明した機能モジュール構成で実現される画像データ検索装置による処理の流れを、図３のフローチャートを用いて説明する。このフローチャートに対応するプログラムは制御プログラム１０４ａに含まれ、ＲＡＭ１０３にロードされた後、ＣＰＵ１０２によって実行される。

【００３７】

まず、音声データ取り込み部２０４により、マルチメディアデータ２１１に記憶された画像データに関連付けられた音声データを取り込む（ステップＳ３０１）。

【００３８】

次に、サブワード音声認識部２０６により、ステップＳ３０１で取り込んだ音声データを入力し、サブワード列を出力とする音声認識を行い（ステップＳ３０２）、その認識結果をサブワード認識結果２１２として記憶する（ステップＳ３０３）。前述した通り、ここで行う音声認識は、音素タイプライタや音節タイプライタなどのサブワードを単位とする音声認識を行っても良いし、ディクテーションに代表される大語彙連続音声認識を行って得られた認識結果をサブワードに変換して出力するようにしてもよい。ステップＳ３０３では、音節認識を行い、認識スコアの高い上位３候補の認識結果の音節列や認識スコアをサブワード認識結果２１２として記憶する。

【００３９】

ステップＳ３０１の音声の取り込み、ステップＳ３０２のサブワード音声認識、ステッ

プＳ３０３のサブワード認識結果の記憶の処理は、マルチメディアデータ２１１に記憶された全ての画像データについて、各画像データごとに実行する。

【００４０】

次に、キーワード取得部２０２は、ユーザにより入力装置１０５を介して入力された画像検索用のキーワードを取得する（ステップＳ３０４）。取得したキーワードは、キーワード・サブワード変換部２０３でサブワード表現形式に変換する（ステップＳ３０５）。繰り返しになるが、サブワード表現形式に変換したキーワードをクエリサブワードと記述する。 また、本実施形態の画像データ検索装置では、クエリサブワードの表現形式は音節とする。

【００４１】

次に、サブワード類似度計算部２０５で、サブワード認識結果２１２として記憶されたサブワード認識結果とクエリサブワードとを比較し、サブワード類似度を計算する（ステップＳ３０６）。類似度の計算方法は、このサブワード類似度計算部２０５の機能に関して説明した通りである。このステップＳ３０６では、サブワード類似度を各画像データごとに計算する。

【００４２】

次に、サブワード類似度が大きい所定個数の画像データを選択し（ステップＳ３０７）、キーワードスポッティング音声認識部２０７により、その選択された画像データに関連づけられた音声データについて、キーワード・サブワード変換部２０３で求めたクエリサブワードを認識対象語とするキーワードスポッティングを行い、認識スコアを求める（ステップＳ３０８）。

【００４３】

次に、検索スコア計算部２０８により、ステップＳ３０８で求まった認識スコアを基に検索スコアを求め（ステップＳ３０９）、検索結果表示部２０９により、検索スコアの良い画像データを表示装置１０６に表示する（ステップＳ３１０）。

【００４４】

本実施形態の画像データ検索装置ではキーワードスポッティング音声認識で求まった認識スコアを検索スコアとして用いる。キーワードスポッティング音声認識で求まる認識スコアは、入力されたキーワードをサブワード表現形式に変換したクエリサブワードを認識対象語として求めているため、ステップＳ３０２で行った言語制約が緩いサブワード音声認識で得られる認識結果に基づいてステップＳ３０６にてサブワード類似度計算部２０６で計算された類似度に比べ、音声データの“キーワードらしさ”をより正確に表していることが期待できる。すなわち、このキーワードスポッティング音声認識で求まる認識スコアに基づいて、画像データを選択することにより、検索精度が向上することが期待できる。

【００４５】

以上説明した画像データ検索装置の処理では、説明の簡単のため、図３に示した全ての処理を一度に行う場合について説明したが、ステップＳ３０１～Ｓ３０３の処理については、ユーザが入力するキーワードに依存しない処理であり、また、ステップＳ３０４以降の処理とは独立に実行できるため、ステップＳ３０４の処理に先立って事前に行っておくことが好ましい。つまり、ステップＳ３０１～Ｓ３０３の処理は、画像データが更新された場合にのみ実行し、ステップＳ３０４以降のキーワードに依存した処理だけを、キーワードが入力される度に実行するように構成するのが望ましい。

【００４６】

以上説明した実施形態によれば、シンボルレベルの検索で候補を絞り込む際に、サブワード表現形式に変換した音声データとキーワードの類似度を基準にすることにより、従来の技術では検索できなかったサブワード表現形式に変換したものに変換誤りが含まれる音声データを、検索することができるようになる。

【００４７】

（実施形態２）

サブワード類似度計算部２０６における類似度の計算では、クエリサブワードがサブワード認識結果２１２に対して部分一致する場合に類似度が大きくなるような計算方法を用いても良い。その一例を以下で説明する。

【００４８】

実施形態１で用いた音節正解精度の計算式（（Ｎ－Ｉ－Ｓ－Ｄ）／Ｎ）×１００のかわりに、挿入誤り音節数Ｉに対して重みβをかけた式（（Ｎ－β・Ｉ－Ｓ－Ｄ）／Ｎ）×１００を用い、βの値を調節することでクエリサブワードが部分一致した場合のサブワード類似度の値を大きくすることができる。

【００４９】

クエリサブワードが「は こ ね」、サブワード音声認識結果が「は こ ね や ま」であった場合、実施形態１で用いた音節正解精度をサブワード類似度とした場合は、正解音節数Ｎ＝５、挿入誤り音節数Ｉ＝２（「や」「ま」）、置換誤り音節数Ｓ＝０、削除誤り音節数Ｄ＝０となり、サブワード類似度は（（５－２－０－０）／５）＊１００＝６０となる。挿入誤り音節数に重みをかけた式（（５－β・２－０－０）／５）＊１００を用いた場合、β＝０．５にするとサブワード類似度は８０、β＝０にするとサブワード類似度は１００になり、音節正解精度を用いた場合に比べクエリサブワードが部分一致したときのサブワード類似度が大きくなる。このようにクエリサブワードが部分一致したときにサブワード類似度が大きくなるようなサブワード類似度の計算方法を用いることで、データに関連付けられた音声データに部分一致するような検索キーワードを使った検索が実現できる。なお、上式において、β＝０の時に求まる値は音節正解率、β＝１とおいた場合は実施形態１で用いた音節正解精度になる。

【００５０】

また、音節正解精度や音節正解率と別の方法として、例えば音節認識結果から音節同士の間違えやすさを記述したコンフュージョンマトリクスを作成し、これを基にして音節間の距離を定義して、この距離を基にクエリサブワードとサブワード認識結果２１２の類似度を求めるようにしてもよい。言うまでもないが、これらの方法は音節に限らず、音素など他の認識単位を用いた場合でも適用可能である。

【００５１】

（実施形態３）

実施形態１では、サブワード認識結果２１２に記憶するサブワード音声認識結果として、４０２（図４）に示すような認識結果の音節列を記憶する場合について説明したが、本発明はこれに限るものではなく、ラティス構造やグラフ構造でサブワード認識結果を表現したものを記憶しても良い。その一例を図５に示す。図５は、ノードとリンクを用いたグラフ構造で表現した音節音声認識結果である。ノード５０１、ノード５０５はそれぞれ認識結果の開始、終了を意味し、音節認識結果を構成する各音節は、５０１、５０５の間のノードで表現されている。５０２のノードを例に説明すると、音節名「お」が５０３に、さらに音節「お」の区間で計算された音声認識スコアが５０４に記述されている。開始ノード５０１から終了ノード５０５に至る経路を辿ることで、音節認識結果とその認識スコアを求めることができる。認識スコアは辿った経路上の各ノードに記録されている、その区間の音声認識スコアの和によって求まる。図５に示した例では、

「ふぁ お ね や ま あ」、
「ふぁ お ね や ま」、
「ふぁ こ ね や ま あ」、
「ふぁ こ ね や ま」、

の４個の音節認識結果が表現されており、それぞれの認識スコアは、

「ふぁ お ね や ま あ」：５０＋４１＋４０＋５０＋３０＋２２＝２３３、
「ふぁ お ね や ま」：５０＋４１＋４０＋５０＋３０＝２１１、
「ふぁ こ ね や ま あ」：５０＋３８＋４０＋５０＋３０＋２２＝２３０、
「ふぁ こ ね や ま」：５０＋３８＋４０＋５０＋３０＝２０８、

となる。

【００５２】
このようにサブワード認識結果２１２がグラフ構造で記憶されている場合、ステップＳ３０６でサブワード類似度の計算を行う際に、グラフが表現する全てのサブワード列をサブワード類似度の計算対象にしても良い。例えば、図５で示したグラフ構造では、先に説明した４つの音節認識結果からサブワード類似度を求める。このかわりに、グラフが表現するサブワード列のうち認識スコアの高い所定数の候補のみを計算対象にしても良い。例えば、図５で示したグラフ構造において、認識スコアが高い上位３候補のみをサブワード類似度の計算対象にする場合には、「ふぁ お ね や ま あ」、「ふぁ お ね や ま」、「ふぁ こ ね や ま あ」を用いてサブワード類似度を計算する。あるいは、認識スコアに閾値を設け、閾値以上のサブワード列のみ計算対象にしてもよい。例えば、図５で示したグラフ構造において、認識スコアの閾値を２３０とする場合には、「ふぁ お ね や ま あ」、「ふぁ こ ね や ま あ」からサブワード類似度を計算する。サブワード認識結果がラティス構造で表現されている場合についても、グラフ構造同様にラティスから全てのサブワード列を抽出することにより、本実施形態と同様の処理が適用できることは言うまでもない。
【００５３】
（実施形態４）
上述の実施形態１乃至実施形態３では、画像データ検索装置を例として、ユーザがキーボードなどで検索用キーワードを入力する態様を説明したが、同様の構成で、キーワードを音声によって入力する装置も実現が可能である。
【００５４】
実施形態１において、入力されたキーワードをサブワード表現形式のクエリサブワードに変換する処理を、ユーザが音声で入力したキーワードをサブワード表現形式のクエリサブワードに変換する処理に置き換えることで、キーワードの音声入力を実現できる。実施形態１で説明した画像データ検索装置は、音声データからサブワード認識結果を求めるサブワード音声認識部２０５を備えており、このサブワード音声認識部２０５を利用することにより、キーワードとして入力された音声をサブワード表現形式のクエリサブワードに変換することができる。
【００５５】
以下、この詳細を実施形態１と同様に画像データ検索装置を例に説明する。
【００５６】
図６は、本実施形態における画像データ検索装置のハードウェア構成を示すブロック図である。これは、図１に示した実施形態１の画像データ検索装置の構成に、マイクロフォンなどの音声を入力するための音声入力装置１０８が追加された構成である。また、本実施形態における画像データ検索装置の機能構成は、図２に示した実施形態１の画像データ検索装置と同様の機能構成を有するので、ここでは図２を援用する。ただし、実施形態１とは異なる処理を行う機能モジュールがあるので以下で説明する。
【００５７】
キーワード取得部２０２は、音声入力装置１０８を介して音声で入力されるキーワード（キーワード音声）を取得する。以後の説明では、画像データに関連付けられた音声データと、ユーザによりキーワードとして入力された音声データとを区別するため、前者をこれまでどおり単に「音声データ」と記述し、後者を「キーワード音声」と記述する。
【００５８】
キーワード・サブワード変換部２０３は、取得したキーワード音声をサブワード表現形式に変換したクエリサブワードに変換する。キーワード音声からクエリサブワードへの変換には、サブワード音声認識部２０５によるサブワード音声認識を用いる。キーワード音声データを入力としてサブワード音声認識を行い、得られたサブワード表現形式の認識結果をクエリサブワードとする。実施形態１で説明したように、ここで用いる音声認識は、認識結果としてサブワード表現形式で記述されたものが得られれば良いので、音素タイプライタや音節タイプライタなどサブワード表現形式の認識結果を出力するサブワードを単

位とした音声認識を行っても良いし、一般にディクテーションとして知られる大語彙連続音声認識を行って得られた認識結果をサブワードに変換するようにしてもよい。また、実施形態１における音声データのサブワード音声認識と同様に、キーワード音声のサブワード音声認識においても、単一の認識結果だけでなく複数の認識結果候補を求めてクエリサブワードとしてもよい。また、実施形態３で説明したようなラティス構造やグラフ構造で表現されたサブワード認識結果（以下「サブワードグラフ」と記述する。）を求めてクエリサブワードとしてもよい。すなわち、クエリサブワードの表現形式は、単一のサブワード列、複数のサブワード列、サブワードグラフのいずれの形式でも良い。

【００５９】

サブワード音声認識部２０５では、キーワード・サブワード変換部２０３におけるキーワード音声のサブワード音声認識に加え、実施形態１と同様に画像データに関連付けられた音声データのサブワード音声認識も行うので、本実施形態におけるサブワード音声認識部２０５では、検索用に入力されたキーワード音声と画像データに関連付けられた音声データの両方に対してサブワード音声認識を行うことになる。

【００６０】

サブワード類似度計算部２０６では、音声データをサブワード認識したサブワード認識結果２１２とキーワード・サブワード変換部２０３で求めたクエリサブワードとを比較し、サブワード類似度を計算する。前述したように、クエリサブワードの表現形式は、単一のサブワード列、複数のサブワード列、サブワードグラフのいずれの形式でも良い。同様に、実施形態１、実施形態３で説明したように、音声データのサブワード認識結果２１２の表現形式も、単一のサブワード列、複数のサブワード列、サブワードグラフのいずれの形式でも良い。

【００６１】

クエリサブワードがＱ個のサブワード列、サブワード認識結果がＮ個のサブワード列である場合のサブワード類似度の計算方法の一例を図７のフローチャートを用いて説明する。この方法では、クエリサブワードのｑ番目のサブワード列を正解とするときの、サブワード認識結果のｎ番目のサブワード列のサブワード正解精度ａｃｃ（ｑ、ｎ）を、１≦ｑ≦Ｑ、１≦ｎ≦Ｎ について全て求め、その最大値をサブワード類似度として計算する。

【００６２】

なお、図７および以下の説明において記述されている式は、Ｃ言語等の記法に従っていることに留意されたい。すなわち、単一の等号「＝」は「右辺の値を左辺に代入する」ことを意味し、二重等号「＝＝」は「左辺と右辺の値が等しい」ことを意味する。

【００６３】

まず、ステップＳ７０１で、サブワード正解精度の最大値を示す変数ｍａｘ、クエリサブワードのサブワード列のインデックスｑをそれぞれ０に初期化する。ステップＳ７０２で、クエリサブワードのサブワード列のインデックスｑの値を１増分するとともに、サブワード認識結果のサブワード列のインデックスｎを０に初期化する。続くステップＳ７０３では、サブワード認識結果のサブワード列のインデックスｎの値を１増分する。

【００６４】

次に、クエリサブワードのｑ番目のサブワードを正解として（ステップＳ７０４）、サブワード認識結果のｎ番目のサブワード列のサブワード正解精度ａｃｃ（ｑ、ｎ）を計算する（ステップＳ７０５）。ａｃｃ（ｑ、ｎ）がサブワード正解精度の最大値ｍａｘよりも大きい場合（ステップＳ７０６）はｍａｘの値を当該ａｃｃ（ｑ、ｎ）で置き換える（ステップＳ７０７）。全てのサブワード認識結果についてサブワード正解精度の計算を終えた場合（ステップＳ７０８においてｎがＮに等しくなった場合）は、ステップＳ７０３～Ｓ７０８と同様の処理をインデックスｑの値を１増分してクエリサブワードの次のサブワード列について行う。クエリサブワードの全てのサブワード列について、ステップＳ７０３～Ｓ７０８の処理を終了したとき（ステップＳ７０９においてｑがＱに等しくなったとき）に求まっているサブワード正解精度の最大値ｍａｘをサブワード類似度とする（ステップＳ７１０）。なお、実施形態１で示したサブワード類似度の計算方法は、図７のフ

ローチャートにＱを１、Ｎを３、サブワードを音節として適用した場合に相当する。

【００６５】

以上の説明では、サブワード正解精度ａｃｃ（ｑ、ｎ）（１≦ｑ≦Ｑ、１≦ｎ≦Ｎ）の最大値をサブワード類似度として用いたが、ａｃｃ（ｑ、ｎ）に認識スコアなどに基づいて重み付けした値を求めて、その最大値あるいは和をサブワード類似度として用いても良い。また、サブワード正解精度ではなく、実施形態２で説明したような音節正解率や他の尺度を使って計算しても良い。

【００６６】

クエリサブワードあるいはサブワード認識結果がサブワードグラフで得られている場合は、実施形態３で説明したようにサブワードグラフから複数のサブワード列を抽出して、抽出した複数（あるいは単数）のサブワード列に対して図７で示した処理を適用することでサブワード類似度を計算することができる。

【００６７】

キーワードスポッティング部２０７では、実施形態１と同様に音声データを入力、クエリサブワードを認識対象語とするキーワードスポッティングを行い、認識スコアを求める。クエリサブワードが複数のサブワード列として求まっている場合は、各サブワード列を認識対象語とし、クエリサブワードがサブワードグラフの形式で表される場合はサブワードグラフから抽出した複数のサブワード列を認識対象語とするキーワードスポッティングを行い、各認識対象語に対して求まった認識スコアの最大値を認識スコアとする。

【００６８】

以上の機能モジュール構成で実現した画像データ検索装置の処理手順を図８に示すフローチャートに示す。図３に示した実施形態１の画像データ検索装置の処理手順との違いは、ステップＳ３０４およびＳ３０５のかわりにステップＳ８０４およびＳ８０５が実行される点であるので、この部分のみ説明する。

【００６９】

実施形態１では、ステップＳ３０４においてキーワード取得部２０２が入力装置１０７から入力されたキーワードを取得したのに対し、本実施形態のステップＳ８０４では、キーワード取得部２０２は音声入力装置１０８から入力されたキーワード音声を取得する。

【００７０】

続くステップＳ８０５では、ステップＳ８０４で取得したキーワード音声をキーワード・サブワード変換部２０３でサブワード表現形式に変換するという点でステップＳ３０５と同一の処理であるが、本実施形態では先に説明したように、キーワード・サブワード変換部２０３からサブワード音声認識部２０５を駆動し、キーワード音声を入力とするサブワード音声認識を行い、得られた認識結果からサブワード表現形式のクエリサブワードを得る。

【００７１】

ステップＳ３０６以降の処理については、得られるクエリサブワードのサブワード列が複数になる場合の処理が追加されるが、実施形態１と同様の処理によって検索処理が実行される。

【００７２】

以上説明したように、本実施形態によれば、実施形態１と同様の構成で、検索用キーワードを音声入力可能なデータ検索装置が実現される。

【００７３】

（他の実施形態）

以上、本発明の実施形態を詳述したが、本発明は、複数の機器から構成されるシステムに適用してもよいし、また、一つの機器からなる装置に適用してもよい。

【００７４】

なお、本発明は、前述した実施形態の機能を実現するソフトウェアのプログラムを、システムあるいは装置に直接あるいは遠隔から供給し、そのシステムあるいは装置のコンピュータがその供給されたプログラムコードを読み出して実行することによっても達成され

る。その場合、プログラムの機能を有していれば、その形態はプログラムである必要はない。

【００７５】

従って、本発明の機能処理をコンピュータで実現するために、そのコンピュータにインストールされるプログラムコード自体およびそのプログラムを格納した記憶媒体も本発明を構成することになる。つまり、本発明の特許請求の範囲には、本発明の機能処理を実現するためのコンピュータプログラム自体、およびそのプログラムを格納した記憶媒体も含まれる。

【００７６】

その場合、プログラムの機能を有していれば、オブジェクトコード、インタプリタにより実行されるプログラム、ＯＳに供給するスクリプトデータ等、プログラムの形態を問わない。

【００７７】

プログラム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、ＭＯ、ＣＤ－ＲＯＭ、ＣＤ－Ｒ、ＣＤ－ＲＷ、磁気テープ、不揮発性のメモリカード、ＲＯＭ、ＤＶＤ（ＤＶＤ－ＲＯＭ、ＤＶＤ－Ｒ）などがある。

【００７８】

その他、プログラムの供給方法としては、クライアントコンピュータのブラウザを用いてインターネットのホームページに接続し、そのホームページから本発明のコンピュータプログラムそのもの、もしくは圧縮され自動インストール機能を含むファイルをハードディスク等の記憶媒体にダウンロードすることによっても供給できる。また、本発明のプログラムを構成するプログラムコードを複数のファイルに分割し、それぞれのファイルを異なるホームページからダウンロードすることによっても実現可能である。つまり、本発明の機能処理をコンピュータで実現するためのプログラムファイルを複数のユーザに対してダウンロードさせるＷＷＷサーバも、本発明のクレームに含まれるものである。

【００７９】

また、本発明のプログラムを暗号化してＣＤ－ＲＯＭ等の記憶媒体に格納してユーザに配布し、所定の条件をクリアしたユーザに対し、インターネットを介してホームページから暗号化を解く鍵情報をダウンロードさせ、その鍵情報を使用することにより暗号化されたプログラムを実行してコンピュータにインストールさせて実現することも可能である。

【００８０】

また、コンピュータが、読み出したプログラムを実行することによって、前述した実施形態の機能が実現される他、そのプログラムの指示に基づき、コンピュータ上で稼動しているＯＳなどが、実際の処理の一部または全部を行い、その処理によっても前述した実施形態の機能が実現され得る。

【００８１】

さらに、記憶媒体から読み出されたプログラムが、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き込まれた後、そのプログラムの指示に基づき、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、その処理によっても前述した実施形態の機能が実現される。

【図面の簡単な説明】

【００８２】

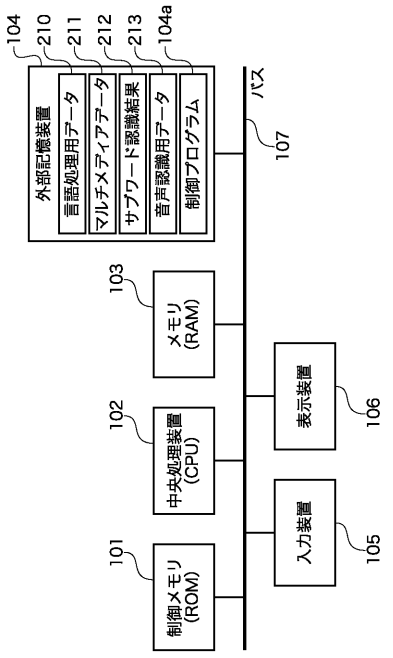
【図１】本発明の実施形態における画像データ検索装置のハードウェア構成を示すブロック図である。

【図２】本発明の実施形態における画像データ検索装置の機能構成を示すブロック図である。

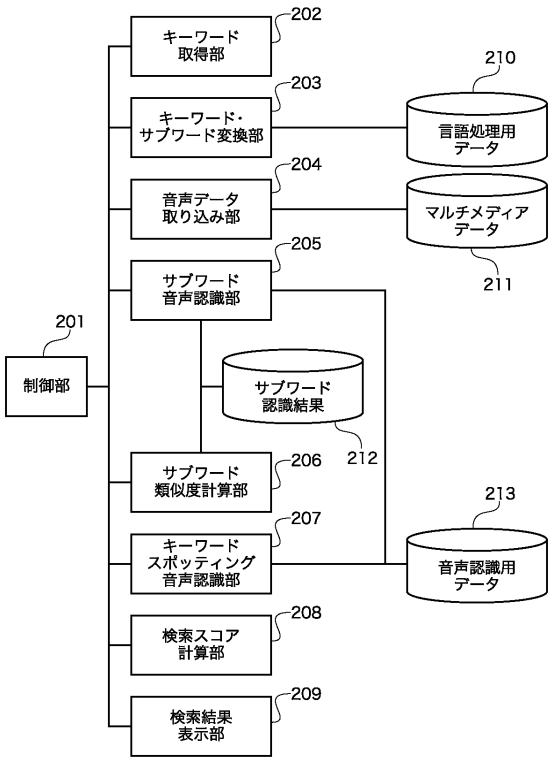
【図３】本発明の実施形態における画像データ検索装置によるデータ検索処理を示すフローチャートである。

【図４】本発明の実施形態におけるサブワード認識結果の一例を示す図である。
【図５】本発明の実施形態におけるグラフ構造で表現されたサブワード認識結果の一例を示す図である。
【図６】本発明の実施形態における、キーワードを音声によって入力するタイプの画像データ検索装置のハードウェア構成を示すブロック図である。
【図７】本発明の実施形態におけるキーワードを音声によって入力するタイプの画像データ検索装置によるサブワード類似度の計算処理を示すフローチャートである。
【図８】本発明の実施形態におけるキーワードを音声によって入力するタイプの画像データ検索装置による、データ検索処理を示すフローチャートである。

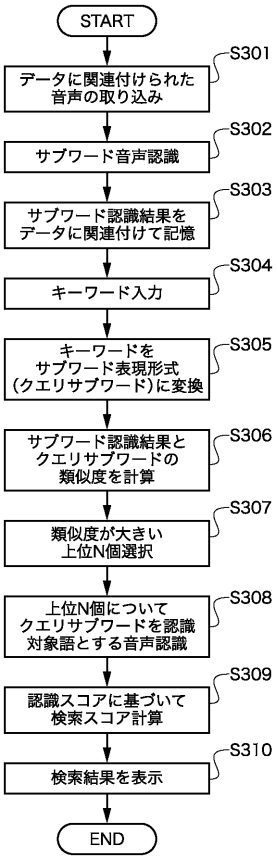
【図１】




【図２】



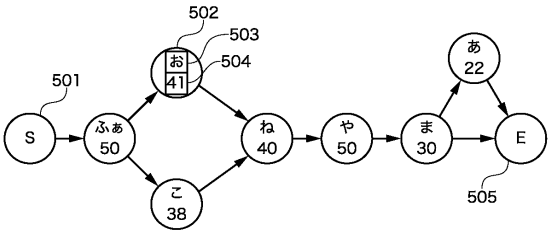
【図３】


START
データに関連付けられた音声の取り込み
S301
サブワード音声認識
S302
サブワード認識結果をデータに関連付けて記憶
S303
キーワード入力
S304
キーワードをサブワード表現形式(クエリサブワード)に変換
S305
サブワード認識結果とクエリサブワードの類似度を計算
S306
類似度が大きい上位N個選択
S307
上位N個についてクエリサブワードを認識対象語とする音声認識
S308
認識スコアに基づいて検索スコア計算
S309
検索結果を表示
S310
END

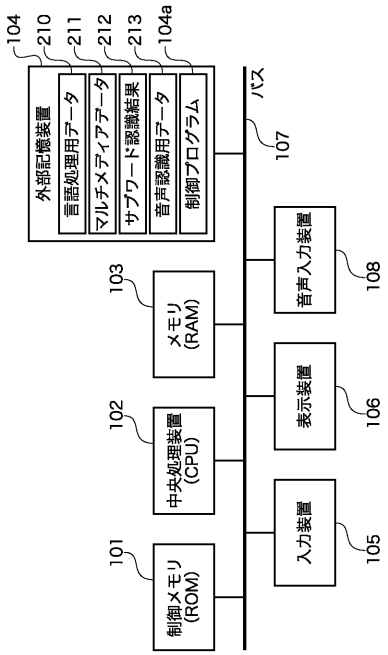
【図４】


411
412
413
順位
認識結果の音節列
認識スコア
1
ふぁ お ね や ま あ
233
2
ふぁ お ね や ま
211
3
ふぁ こ ね や ま
208
401
402
403

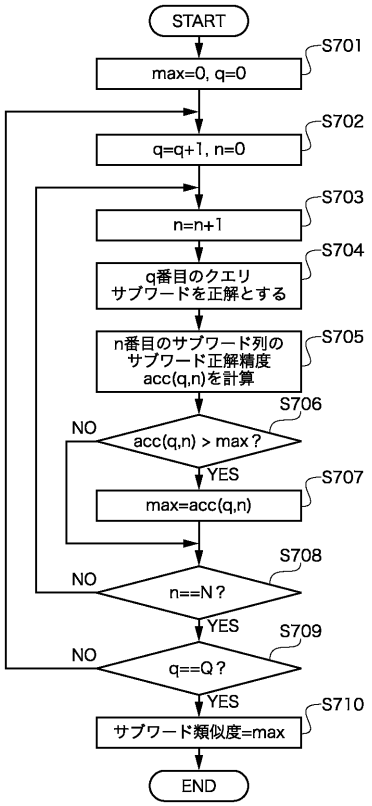
【図５】


502
503
504
501
S
ふぁ 50
お 41
こ 38
ね 40
や 50
ま 30
あ 22
E
505

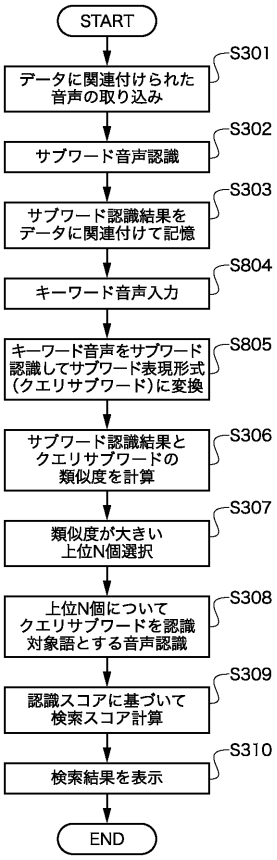
【図６】


104
210
211
212
213
104a
外部記憶装置
言語処理用データ
マルチメディアデータ
サブワード認識結果
音声認識用データ
制御プログラム
バス
107
103
メモリ(RAM)
102
中央処理装置(CPU)
101
制御メモリ(ROM)
音声入力装置
108
表示装置
106
入力装置
105

【図７】


START
S701
max=0, q=0
S702
q=q+1, n=0
S703
n=n+1
S704
q番目のクエリサブワードを正解とする
S705
n番目のサブワード列のサブワード正解精度acc(q,n)を計算
S706
NO
acc(q,n) > max?
YES
S707
max=acc(q,n)
S708
NO
n==N?
YES
S709
NO
q==Q?
YES
S710
サブワード類似度=max
END

【図８】


START
S301
データに関連付けられた音声の取り込み
S302
サブワード音声認識
S303
サブワード認識結果をデータに関連付けて記憶
S804
キーワード音声入力
S805
キーワード音声をサブワード認識してサブワード表現形式（クエリサブワード）に変換
S306
サブワード認識結果とクエリサブワードの類似度を計算
S307
類似度が大きい上位N個選択
S308
上位N個についてクエリサブワードを認識対象語とする音声認識
S309
認識スコアに基づいて検索スコア計算
S310
検索結果を表示
END

フロントページの続き

(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
(72)発明者　山田　耕平
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　井上　健一

(56)参考文献　特開２００２－２７８５７９（ＪＰ，Ａ）
特開２００４－３０２１７５（ＪＰ，Ａ）
特開２０００－２５９６４５（ＪＰ，Ａ）
特開平１１－０８５１８７（ＪＰ，Ａ）
特開昭６３－２３９４９９（ＪＰ，Ａ）
特開平１０－１７３７６９（ＪＰ，Ａ）
特開２００３－２１９３２７（ＪＰ，Ａ）
特開２００６－０４０１５０（ＪＰ，Ａ）
特開平０８－２１１８９３（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１５／００－１５／２８