

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5753769号
(P5753769)

(45) 発行日 平成27年7月22日 (2015. 7. 22)

(24) 登録日 平成27年5月29日 (2015. 5. 29)

(51) Int. Cl.	F I
G 1 0 L 25/54 (2013. 01)	G 1 0 L 25/54
G 1 0 L 15/00 (2013. 01)	G 1 0 L 15/00 2 0 0 T
G 1 0 L 13/00 (2006. 01)	G 1 0 L 13/00 1 0 0 L

請求項の数 12 (全 13 頁)

(21) 出願番号	特願2011-252425 (P2011-252425)	(73) 特許権者	000005108
(22) 出願日	平成23年11月18日 (2011. 11. 18)		株式会社日立製作所
(65) 公開番号	特開2013-109061 (P2013-109061A)		東京都千代田区丸の内一丁目6番6号
(43) 公開日	平成25年6月6日 (2013. 6. 6)	(74) 代理人	110001689
審査請求日	平成26年7月7日 (2014. 7. 7)		青稜特許業務法人
		(74) 代理人	110000350
			ポレール特許業務法人
		(72) 発明者	神田 直之
			東京都国分寺市東恋ヶ窪一丁目280番地
			株式会社日立製作所 中央研究所内
		審査官	山下 剛史

最終頁に続く

(54) 【発明の名称】 音声データ検索システムおよびそのためのプログラム

(57) 【特許請求の範囲】

【請求項 1】

キーワードを入力する入力装置と、
 入力された前記キーワードを音素表記へ変換する音素変換部と、
 音素表記のキーワードを元に音声データ中で当該キーワードが発話された個所を検索する音声データ探索部と、
 音素表記のキーワードを元にユーザが聴取混同する可能性のある当該キーワードとは別の対照キーワードの集合を生成する対照キーワード生成部と、
ユーザが入力した前記キーワードと、前記対照キーワード生成部で生成した前記対照キーワードのいずれか一方もしくは両方を音声合成する音声合成部と、
 前記音声データ探索部からの検索結果および前記対照キーワード生成部からの前記対照キーワードをユーザへ提示する検索結果提示部を備え、
前記検索結果提示部は、前記音声合成部からの合成音声を用いてユーザへ提示する音声データ検索システム。

【請求項 2】

請求項 1 に記載の音声データ検索システムにおいて、
 ユーザごとの音素混同行列を備え、
 前記対照キーワード生成部は、前記音素混同行列に基づいて対照キーワード生成を行うことを特徴とする音声データ検索システム。

【請求項 3】

10

20

請求項 2 に記載の音声データ検索システムにおいて、
ユーザが理解可能な言語についての情報を入力する言語情報入力部と、
言語情報入力部から得られた情報に基づいて前記音素混同行列を生成する音素混同行列生成部を備えたことを特徴とする音声データ検索システム。

【請求項 4】

請求項 1 に記載の音声データ検索システムにおいて、
前記対照キーワード生成部は、前記音素表記されたキーワードと単語辞書に登録された単語の音素表記との間の編集距離を計算し、編集距離がしきい値以下の単語を対照キーワードとすることを特徴とする音声データ検索システム。

【請求項 5】

請求項 1 に記載の音声データ検索システムにおいて、
前記対照キーワード生成部で生成した対照キーワード候補と前記音声データ探索部の検索結果を比較して、不要な対照キーワード候補を除去する対照キーワードチェック部を備えたことを特徴とする音声データ検索システム。

【請求項 6】

請求項 5 に記載の音声データ検索システムにおいて、
前記対照キーワードチェック部は、前記対照キーワード候補と前記音声データ探索部の検索結果を比較して、不要な音声データ検索結果を除去することを特徴とする音声データ検索システム。

【請求項 7】

コンピュータを、
入力されたキーワードを音素表記へ変換する音素変換部と、
音素表記のキーワードを元に音声データ中で当該キーワードが発話された個所を検索する音声データ探索部と、
音素表記のキーワードを元にユーザが聴取混同する可能性のある当該キーワードとは別の対照キーワードの集合を生成する対照キーワード生成部と、
入力された前記キーワードと、前記対照キーワード生成部で生成した前記対照キーワードのいずれか一方もしくは両方を音声合成する音声合成部と、

前記音声データ探索部からの検索結果および前記対照キーワード生成部からの前記対照キーワードをユーザへ提示する検索結果提示部とを備え、

前記検索結果提示部は、前記音声合成部からの合成音声を用いてユーザへ提示する音声データ検索システムとして機能させるためのプログラム。

【請求項 8】

請求項 7 に記載のプログラムにおいて、
ユーザごとの音素混同行列を備え、前記対照キーワード生成部は、前記音素混同行列に基づいて対照キーワード生成を行うように機能させることを特徴とする、コンピュータを音声データ検索システムとして機能させるためのプログラム。

【請求項 9】

請求項 8 に記載のプログラムにおいて、更に、
ユーザが理解可能な言語についての情報を入力する言語情報入力部と、言語情報入力部から得られた情報に基づいて前記音素混同行列を生成する音素混同行列生成部として機能させることを特徴とする、コンピュータを音声データ検索システムとして機能させるためのプログラム。

【請求項 10】

請求項 7 に記載のプログラムにおいて、
前記対照キーワード生成部は、前記音素表記されたキーワードと単語辞書に登録された単語の音素表記との間の編集距離を計算し、編集距離がしきい値以下の単語を対照キーワードとするように機能させることを特徴とする、コンピュータを音声データ検索システムとして機能させるためのプログラム。

【請求項 11】

請求項7に記載のプログラムにおいて、更に、

前記対照キーワード生成部で生成した対照キーワード候補と前記音声データ探索部の検索結果を比較して、不要な対照キーワード候補を除去する対照キーワードチェック部として機能させることを特徴とする、コンピュータを音声データ検索システムとして機能させるためのプログラム。

【請求項12】

請求項11に記載のプログラムにおいて、

前記対照キーワードチェック部は、前記対照キーワード候補と前記音声データ探索部の検索結果を比較して、不要な音声データ検索結果を除去するように機能させることを特徴とする、コンピュータを音声データ検索システムとして機能させるためのプログラム。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声データを検索するシステムに関するものである。

【背景技術】

【0002】

近年のストレージデバイスの大容量化に伴い、大量の音声データが蓄積されるようになってきている。従来の多くの音声データベースでは、音声データを管理するために音声録音された時刻の情報を付与し、その情報を元に所望の音声データを検索することが行われてきた。しかしながら、時刻情報に基づく検索では、所望の音声が発話された時刻を予め知っている必要があり、特定のキーワードが発話中に含まれる音声を検索するといった用途には不向きであった。特定のキーワードが発話中に含まれる音声を検索する場合、音声を始めから終わりまで聴取する必要があった。

20

【0003】

そこで、音声データベース中の特定のキーワードが発話された時刻を自動的に検出する技術が開発されている。代表的な手法のひとつであるサブワード検索法では、まずサブワード認識処理によって、音声データをサブワード列へと変換しておく。ここでサブワードとは音素や音節など、単語よりも細かい単位を指す名称である。キーワードが入力されると、当該キーワードのサブワード表現と音声データのサブワード認識結果を比較し、サブワードの一致度が高い個所を検出することにより、音声データ中で当該キーワードが発話されている時刻を検出する（特許文献1、非特許文献1）。また、非特許文献2で示されているワードスポッティング法では、音素単位の音響モデルを組み合わせることで当該キーワードの音響モデルを生成し、当該キーワード音響モデルと音声データとの照合を行うことで、音声データ中で当該キーワードが発話された時刻の検出を行う。

30

【0004】

しかしながら、いずれの技術も発話の変動（なまりや話者性の違いなど）や雑音の影響を受け、検索結果には誤りが含まれ、実際には当該キーワードが発話されていない時刻が検索結果に現れることがある。そのため、ユーザは誤った検索結果を取り除くために、検索によって得られたキーワードの発話時刻から音声データを再生し、聴取により当該キーワードが本当に発話されているか否かを判断する必要がある。

40

【0005】

上記のような正解/不正解判定を補助するための技術も提案されている。特許文献2には、聴取により当該キーワードが本当に発話されているか否かを判断するために、当該キーワードの検出時刻を強調して再生する技術が開示されている。

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特開2002-221984号公報

【特許文献2】特開2005-38014号公報

【非特許文献】

50

【0007】

【非特許文献1】岩田耕平 他「語彙フリー音声文書検索手法における新しいサブワードモデルとサブワード音響距離の有効性の検証」情報処理学会論文誌, Vol. 48, No. 5, 2007

【非特許文献2】河原達也, 宗統敏彦, 堂下修司“ヒューリスティックな言語モデルを用いた会話音声の中の単語スポッティング” 信学論. D-I I, 情報・システム, I I-情報処理, vol. 78, no. 7, pp. 1013-1020, 1995.

【非特許文献3】鹿野清宏 他「IT Text 音声認識システム」オーム社(2001)

【発明の開示】

【発明が解決しようとする課題】

10

【0008】

特許文献2には、聴取により当該キーワードが本当に発話されているか否かを判断するために、当該キーワードの検出時刻を強調して再生する技術が開示されている。

【0009】

しかし、ユーザが検索対象としている音声データの言語を十分には理解できない状況では、しばしば上記のような正解/不正解の判定を聴取によって行うことに困難を伴うという問題がある。例えば、ユーザが「play」というキーワードで検索をした結果、実際には「pray」と発話された時刻が検出されることがある。この場合、英語を十分に解さない日本人のユーザはそれを「play」と言っていると判断してしまう可能性がある。特許文献2で提案されているような当該キーワードの検出位置を強調再生する技術では、上記の問題を解決することはできない。

20

【0010】

本発明は、このような課題を解決し、音声データ検索システムにおいて、検索結果の正解/不正解の判定を容易に行うことができるようにすることを目的とする。

【課題を解決するための手段】

【0011】

本発明は、上記課題を解決するために、例えば特許請求の範囲に記載の構成を採用する。

【0012】

本発明の音声データ検索システムの一例を挙げるならば、キーワードを入力する入力装置と、入力された前記キーワードを音素表記へ変換する音素変換部と、音素表記のキーワードを元に音声データ中で当該キーワードが発話された個所を検索する音声データ探索部と、音素表記のキーワードを元にユーザが聴取混同する可能性のある当該キーワードとは別の対照キーワードの集合を生成する対照キーワード生成部と、ユーザが入力した前記キーワードと、前記対照キーワード生成部で生成した前記対照キーワードのいずれか一方もしくは両方を音声合成する音声合成部と、前記音声データ探索部からの検索結果および前記対照キーワード生成部からの前記対照キーワードをユーザへ提示する検索結果提示部を備え、前記検索結果提示部は、前記音声合成部からの合成音声を用いてユーザへ提示する音声データ検索システムである。

30

【0013】

また、本発明のプログラムの一例を挙げるならば、コンピュータを、入力されたキーワードを音素表記へ変換する音素変換部と、音素表記のキーワードを元に音声データ中で当該キーワードが発話された個所を検索する音声データ探索部と、音素表記のキーワードを元にユーザが聴取混同する可能性のある当該キーワードとは別の対照キーワードの集合を生成する対照キーワード生成部と、入力された前記キーワードと、前記対照キーワード生成部で生成した前記対照キーワードのいずれか一方もしくは両方を音声合成する音声合成部と、前記音声データ探索部からの検索結果および前記対照キーワード生成部からの前記対照キーワードをユーザへ提示する検索結果提示部とを備え、前記検索結果提示部は、前記音声合成部からの合成音声を用いてユーザへ提示する音声データ検索システムとして機能させるためのプログラムである。

40

50

【発明の効果】

【0014】

本発明によれば、音声データ検索システムにおいて、ユーザが入力したキーワードを元に、ユーザが聴取混同する可能性のある対照キーワード集合を生成しユーザに提示することで、検索結果の正解/不正解の判定を容易に行うことができるようにすることができる。

【図面の簡単な説明】

【0015】

【図1】本発明を適用する計算機システムの構成を示すブロック図である。

【図2】本発明の構成要素を処理の流れに従って配置した図である。

10

【図3】本発明の処理の流れを示すフローチャートである。

【図4】対照キーワード候補の生成の処理の流れを示すフローチャートである。

【図5】単語辞書の一例を示す図である。

【図6】音素混同行列の一例を示す図である。

【図7】対照キーワード候補のチェックの処理の流れを示すフローチャートである。

【図8】ユーザへ情報を提示する画面の一例を示す図である。

【図9】音素混同行列の他の例を示す図である。

【図10】編集距離の算出過程の一例を示す図である。

【図11】編集距離の算出過程の他の例を示す図である。

【図12】ユーザが複数の言語を理解できる場合の、音素混同行列の一例を示す図である

20

【図13】編集距離計算の疑似コードを示す図である。

【発明を実施するための形態】

【0016】

以下、本発明の実施の形態を、添付図面に基づいて説明する。

【実施例1】

【0017】

図1は、第1の実施形態を示し、本発明を適用する計算機システムの構成を示すブロック図である。また、図2は、図1の構成要素を処理の流れに従って配置した図である。本実施形態の計算機システムは、計算機101と、表示装置111、入力装置112および音声出力装置113からなる。計算機101の内部には、音声データ蓄積装置102、音素混同行列103、単語辞書104を持ち、また、音声データ探索部105、音素変換部106、対照キーワード生成部107、対照キーワードチェック部108、音声合成部109、検索結果提示部110、言語情報入力部114および音素混同行列生成部115を持つ。

30

【0018】

音声データ検索システムは、計算機（コンピュータ）において、CPUが所定のプログラムをメモリ上にロードし、また、CPUがメモリ上にロードした所定のプログラムを実行することにより実現できる。この所定のプログラムは、図示していないが、読み取り装置を介して当該プログラムが記憶された記憶媒体から、または、通信装置を介してネットワークから入力して、直接メモリ上にロードするか、もしくは、一旦、外部記憶装置に格納してから、メモリ上にロードすれば良い。

40

【0019】

本発明におけるプログラムの発明は、このようにコンピュータに組み込まれ、コンピュータを音声データ検索システムとして動作させるプログラムである。本発明のプログラムをコンピュータに組み込むことにより、図1や図2のブロック図に示される音声データ検索システムが構成される。

【0020】

以下、各構成要素の処理の流れについて記述する。図3に、処理のフローチャートを示す。

【0021】

50

〔キーワード入力および音素表現への変換〕

ユーザが入力装置112からテキストでキーワードを入力すると（処理301）、まず音素変換部106が当該キーワードを音素表現へと変換する（処理302）。例えば、ユーザが入力として「play」というキーワードを入力した場合、これを「p l e I」と変換する。この変換は形態素解析処理として知られており、当業者に周知であるために説明を省略する。

【0022】

また入力装置としてマイクロフォンを用い、ユーザがマイクロフォンに対して音声でキーワードを発話することによって、キーワードの入力を行うことも可能である。この場合、音素変換部として音声認識技術を利用することで、当該音声波形を音素表現へと変換することが可能である。音声認識技術による音素表現への変換技術の詳細は非特許文献3などに記載されており、当業者に周知の技術であるため、詳細の説明は省略する。

10

【0023】

〔音声データ探索〕

続いて、音声データ探索部105が、音声データ蓄積装置102に蓄積されている音声データ中で、当該キーワードが発話された時刻を検出する（処理303）。この処理には、例えば非特許文献2で提示されているワードスポッティング処理を用いることができる。もしくは、特許文献1や非特許文献1など、予め音声データ蓄積装置を前処理しておく方法を利用することも可能である。事業者はこれらのうちいずれかの手段を選択すればよい。

【0024】

〔対照キーワード候補の生成〕

20

続いて、対照キーワード生成部107が、ユーザが聴取混同する可能性のある対照キーワード集合を生成する（処理304）。以下の説明では、キーワードは英語で入力されており、一方でユーザは日本語を母語としているものとする。ただし、キーワードの言語及びユーザの母語は、英語と日本語に限定されず、いかなる言語の組み合わせでも実施可能である。

【0025】

図4に、処理の流れを示す。まず対照キーワード集合Cを空集合として初期化する（処理401）。続いて英語の単語辞書に登録された全ての単語 W_i について、その音素表記とユーザが入力したキーワードKの音素表記との間の編集距離 $Ed(K, W_i)$ を計算する（処理403）。当該単語 W_i に対する編集距離がしきい値以下であれば、当該単語を対照キーワード集合Cに追加する（処理404）。最後に、対照キーワード集合Cを出力する。

30

【0026】

図5に、単語辞書の例を示す。図5に示すように、単語辞書は、単語501とその音素表現502の組を多数記載したものである。

【0027】

図6に、日本語話者用の音素混同行列の例を示す。音素混同行列では、縦の列に示されている音素が横の行に示されている音素と混同されやすい場合には0に近い値、混同されにくい場合には1に近い値が、0から1の間の数値で記載されたものである。ただし、SPは「無音」を表す特殊記号である。例えば、音素bは音素aと混同されにくいために音素混同行列では1が割り当てられている。対して、音素lと音素rは日本語を母語とするユーザにとっては混同しやすい音素であるため、音素混同行列では0という値が割り当てられている。同一の音素の場合には常に0が割り当てられる。音素混同行列は、ユーザの母語言語ごとに1つ用意される。以下、音素混同行列において音素Xの行、音素Yの列に割り当てられた値を $Matrix(X, Y)$ と表す。

40

【0028】

編集距離とは、ある文字列Aと文字列Bの間の距離尺度を定義するものであり、文字列Aに置換、挿入、削除の各操作を施して文字列Bに変換するための最小操作コストとして定義される。例えば図10のように文字列Aがabcdeであり、文字列Bがacfeegであったときは、まず文字列Aの2文字目のbを削除し、文字列Aの4文字目のdをfに置換し、文字列Aの最後尾にgを追加することで文字列Bへ変換できる。ここで、置換、挿入、削除にかかるコ

50

ストがそれぞれ定義されており、操作コストの和が最小となるような操作を選んだ時の、操作コストの和を編集距離 $Ed(A, B)$ とする。

【0029】

本実施例では、ある音素 X の挿入にかかるコストは $Matrix(SP, X)$ 、ある音素 X の削除にかかるコストは $Matrix(X, SP)$ 、音素 X を音素 Y に置換するコストは $Matrix(X, Y)$ とする。これにより、音素混同行列を反映した編集距離を計算することができる。例えばキーワード「play」の音素表現「p l e I」と、単語「pray」の音素表現「p r e I」の編集距離を図6の音素混同行列に従って計算することを考える。「p l e I」の2文字目の l を r へ置換することで「p r e I」へと変換することができる。ここで図6の音素混同行列では l と r に対して0という値が割り当てられているために、 l を r へ置換するコスト $Matrix(l, r)$ は0であることから、「p l e I」はコスト0で「p r e I」へと変換でき、従って編集距離 $Ed(play, pray)=0$ と計算される。

10

【0030】

なお、編集距離の効率的な計算方法である動的計画法は当業者に周知であるため、ここでは疑似コードのみを示す。図13に、疑似コードを示した。ここで音素列 A の i 文字目の音素は $A(i)$ と表しており、音素列 A と音素列 B の長さはそれぞれ N と M としている。

【0031】

また、上記とは別の編集距離の定義として、文字列 A に置換、挿入、削除の各操作を施して、操作後の文字列が文字列 B に含まれるようにするための最小操作コストとして定義することもできる。例えば図11のように文字列 A が $abcde$ 、文字列 B が $xyzacfe$ であった場合には、まず文字列 A の2文字目の b を削除し、続いて3文字目の文字 d を f と置換することで、操作後の文字列 $acfe$ は文字列 B に含まれるようになる。このときの操作コストの和を編集距離 $Ed(A, B)$ とする。

20

【0032】

対照キーワード生成においては、編集距離の定義として上記の2種類のいずれを用いてもよい。また、上記で示した処理以外にも、文字列間の距離を計測する方法であれば、いずれの方法でも利用可能である。

【0033】

さらに図4の処理403, 404において単語 W_i だけでなく、単語列 $W_1 \cdots W_N$ を用いてもよい。

30

【0034】

さらに、処理403において編集距離 $Ed(K, W_1 \cdots W_N)$ だけではなく、単語列 $W_1 \cdots W_N$ が生成される確率 $P(W_1 \cdots W_N)$ も合わせて求め、処理404において編集距離がしきい値以下で、かつ $P(W_1 \cdots W_N)$ がしきい値以上であれば $C \leftarrow C \cup \{W_1 \cdots W_N\}$ とする実装も可能である。この場合には、対照キーワード集合には単語列も含まれる。なお、 $P(W_1 \cdots W_N)$ の算出方法としては、例えば言語処理の分野でよく知られた N -gramモデルを利用することができる。 N -gramモデルの詳細については当業者に周知であるため、ここでは省略する。

【0035】

また上記の外に、 $Ed(K, W_1 \cdots W_N)$ と $P(W_1 \cdots W_N)$ を組み合わせた任意の尺度を利用することもできる。例えば $Ed(K, W_1 \cdots W_N) / P(W_1 \cdots W_N)$ や $P(W_1 \cdots W_N) * (\text{length}(K) - Ed(K, W_1 \cdots W_N)) / \text{length}(K)$ といった尺度を処理404において利用してもよい。ただし、 $\text{length}(K)$ はキーワード K の音素表現に含まれる音素数である。

40

【0036】

[音素混同行列の生成]

対照キーワード生成で使用する音素混同行列は、ユーザの母語もしくは使用可能言語によって切り替えることができる。この場合、ユーザは言語情報入力部114を通してシステムにユーザの母語もしくは使用可能言語に関する情報を入力する。ユーザからの入力を受け取ったシステムは音素混同行列生成部115がユーザの母語用の音素混同行列を出力する。例えば図6は、日本語話者用であるが、中国語を母語とするユーザに対しては、図9で示されたような音素混同行列を用いることが可能である。例えば図9では図6と違い、音

50

素lと音素rの交差する点はlであり、この2つの音素は中国語を母語とするユーザにとっては混同しにくいものであるという定義がなされている。

音素混同行列生成部は、ユーザの母語に限定せず、ユーザが理解できる言語の情報によって、音素混同行列を切り替えてもよい。

【0037】

さらにユーザが複数の言語を理解できる場合に、音素混同行列生成部115は、それらの言語情報を組み合わせた音素混同行列を生成することも可能である。実施例のひとつとして、α語とβ語の両方が理解できるユーザに対しては、α語ユーザ用の音素混同行列のi行j列要素とβ語ユーザ用の音素混同行列のi行j列要素の大きい方がi行j列要素となるような混同行列を生成することができる。3か国語以上の言語を理解できる場合にも、各言語の音素混同行列の中で各行列要素ごとに最も大きいものを選べばよい。

10

例えば日本語と中国語が理解できるユーザに対しては図12の音素混同行列が生成される。図12の音素混同行列の各要素は、日本語話者用音素混同行列(図6)と中国語話者用音素混同行列(図9)の各行列要素の大きい方を代入したものである。

【0038】

また、音素混同行列をユーザが直接操作し、行列の値を調整することも可能である。

なお、音素混同行列の生成は、対照キーワード生成部が動作する前の任意のタイミングで行うことができる。

【0039】

[対照キーワード候補のチェック]

20

対照キーワード生成部107によって生成された対照キーワード候補に対して、対照キーワードチェック部108が動作し、当該対照キーワードをユーザへ提示するか否かの選別を行う。これにより不要な対照キーワード候補を除去する。

【0040】

この処理の流れを図7に示す。

(1) まず対照キーワード生成部107によって生成された全ての対照キーワード候補 W_i ($i=1, \dots, N$)について $\text{flag}(W_i)=0$ とする(処理701)。

(2) 続いて、音声データ探索部から得られたキーワードの発話時刻候補全てに対して、以下の(i)~(iii)の処理を行う。

(i) キーワードの発話時刻の始端と終端を含む音声Xを切り出す(処理703)。

30

(ii) 全ての対照キーワード候補 W_i ($i=1, \dots, N$)について、当該音声に対するワードスポッティング処理を行う(処理705)。

(iii) ワードスポッティングの結果得られたスコア $P(*W_i*|X)$ がしきい値を超えた単語 W_i に対して $\text{flag}(W_i)=1$ とする(処理706)。

(3) $\text{flag}(W_i)$ が0であるキーワードを対照キーワード候補から取り除く(処理707)。

なお、ワードスポッティング処理では、音声X内でキーワード W_i が発話されている確率 $P(*key*|X)$ を数1式に従って算出する。

【0041】

【数1】

$$P(*key*|X) \simeq \max_{h_0} P(X|h_0)P(h_0)/(\max_{h_1} P(X|h_1)P(h_1))$$

40

【0042】

ここで、 h_0 は任意の音素集合のうちキーワードの音素表現を含む要素であり、 h_1 は任意の音素列集合の要素である。詳細は非特許文献2などに示されており、当業者には周知であるため、ここではこれ以上の説明は省略する。

【0043】

また、対照キーワードのチェックをする際に算出されるワードスポッティングの値 $P(*W_i*|X)$ がしきい値を超えた場合には、当該検索結果を検索結果から取り除くことも可能である。

50

【0044】

なお、対照キーワード候補のチェック処理は省略してもよい。

【0045】

[音声合成処理]

対照キーワード候補およびユーザが入力したキーワードの両方を音声合成部109によって音声波形へと変換する。ここでテキストを音声波形へ変換する音声合成技術については、当業者には周知であるため詳細は省略する。

【0046】

[検索結果提示]

最後に、検索結果提示部110が表示装置111および音声出力装置113を通して、ユーザへ検索結果および対照キーワードについての情報を提示する。この際に表示装置111へ表示する画面の例を、図8に示す。

10

【0047】

ユーザは検索窓801に検索キーワードを入力し、ボタン802を押すことで、音声データ蓄積装置102に蓄積された音声データ中でキーワードが発話されている個所を検索することができる。図8の例では、ユーザは「play」というキーワードが音声データ蓄積装置102に蓄積された音声データ中で発話されている個所を検索している。

【0048】

検索結果は、ユーザが入力したキーワードが発話されている音声ファイル名805と当該キーワードが当該音声ファイル内で発話されている時刻806であり、「キーワードから再生」807という個所をクリックすることで当該ファイルの当該時刻から音声出力装置113を通して音声再生される。また「ファイル冒頭から再生」808という個所をクリックすることで、当該ファイルの冒頭から音声出力装置113を通して音声再生される。

20

【0049】

また、「キーワード音声合成を聞く」803という個所をクリックすることにより、当該キーワードの音声合成が音声出力装置113を通して再生される。これによりユーザは当該キーワードの正しい発音を聞くことができ、当該検索結果が正しいかどうかの参考とすることができる。

【0050】

また、図8の804には対照キーワードの候補としてprayとclayが表示されており、「音声合成を聞く」809という個所をクリックすると、その音声合成が音声出力装置113を通して再生される。これらによってユーザは検索結果として「pray」や「clay」というキーワードが発話された個所が誤検出されている可能性に気づき、当該対照キーワードの合成音声聞くことにより、ユーザは当該検索結果が正しいかどうかを判定する際の参考とすることができる。

30

【符号の説明】

【0051】

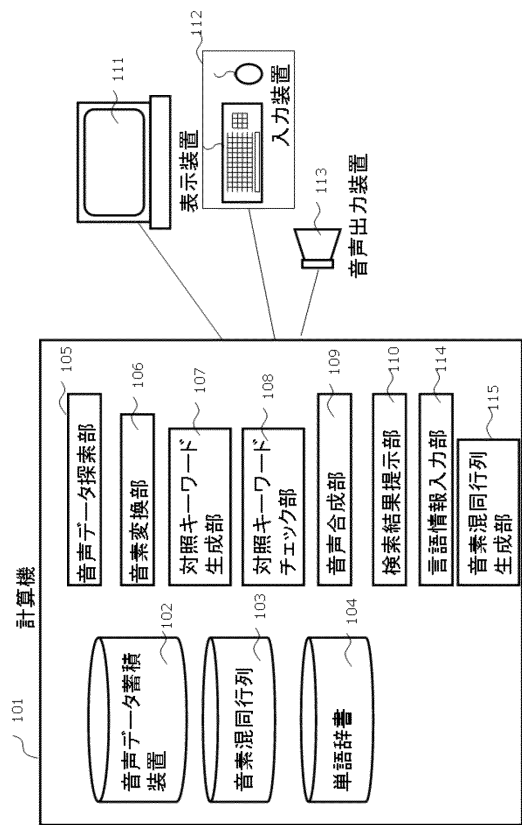
- 101 計算機
- 102 音声データ蓄積装置
- 103 音素混同行列
- 104 単語辞書
- 105 音声データ探索部
- 106 音素変換部
- 107 対照キーワード生成部
- 108 対照キーワードチェック部
- 109 音声合成部
- 110 検索結果提示部
- 111 表示装置
- 112 入力装置
- 113 音声出力装置

40

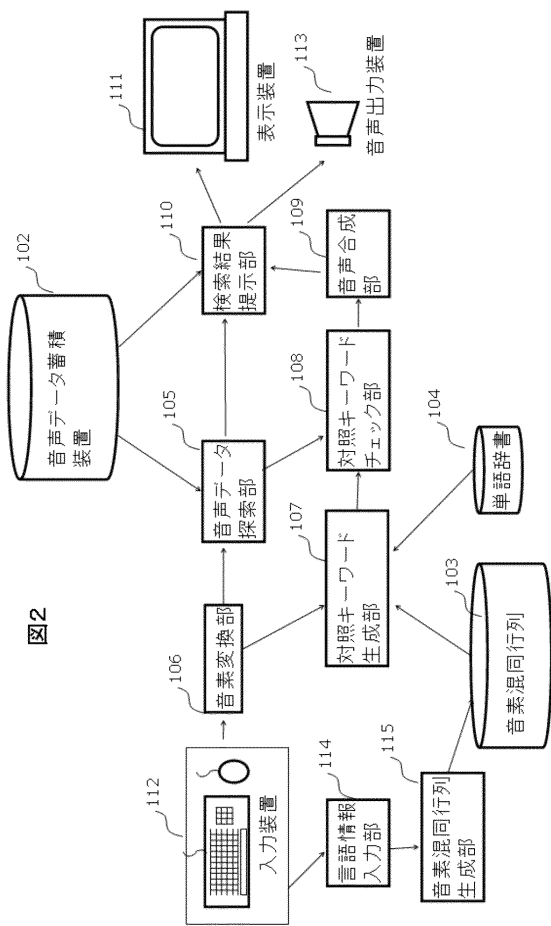
50

- 1 1 4 言語情報入力部
- 1 1 5 音素混同行列生成部

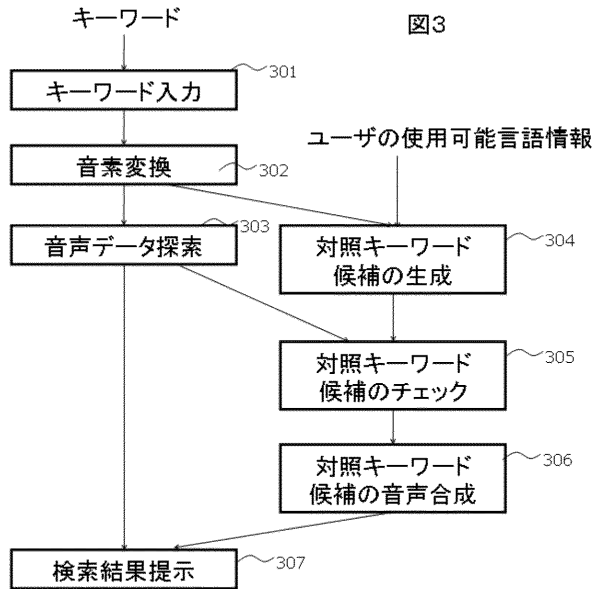
【図 1】



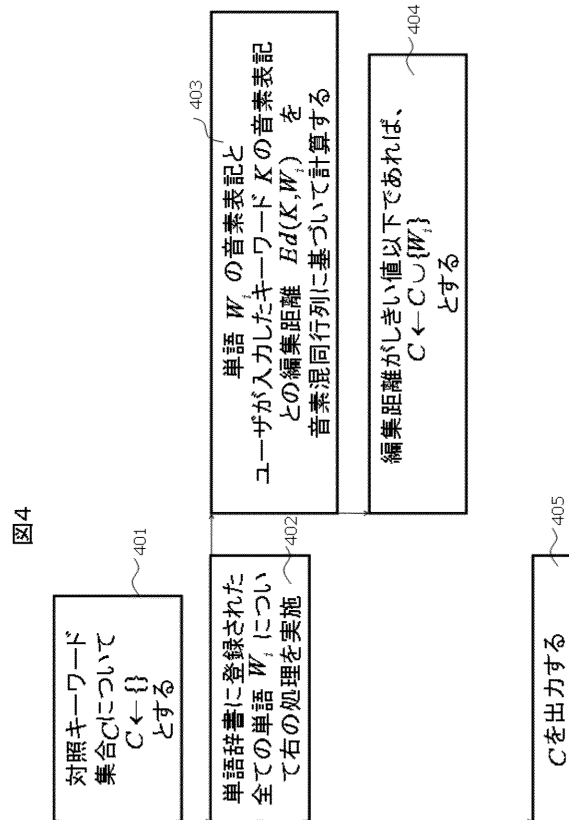
【図 2】



【図 3】



【図 4】



【図 5】

図5

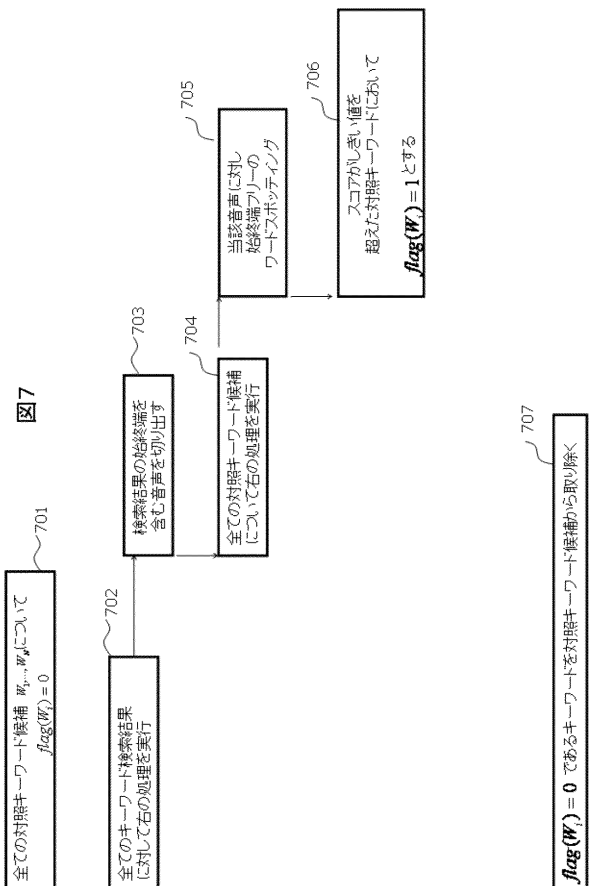
単語	音素表現
dog	d Q g
cat	k { t
...	
pray	pre l
...	

【図 6】

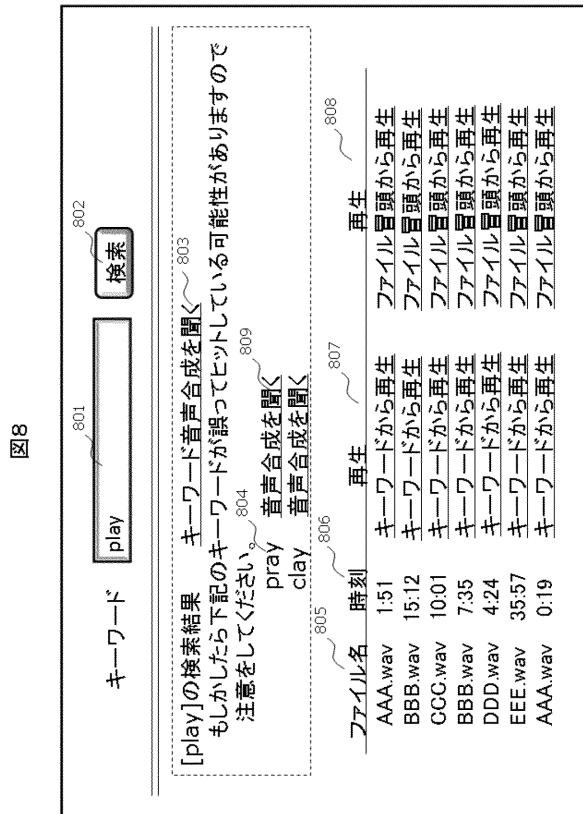
図6

	a	b	...	{	P	l	r	...	z	<sp>
a	0	1		0	1	1	1		1	1
b	1	0		1	1	1	1		1	1
...									...	...
{	0	1		0	1	1	1		1	1
p	1	1		1	0	1	1		1	1
l	1	1		1	1	0	0		1	1
r	1	1		1	1	0	0		1	1
...									...	...
z	1	1	...	1	1	1	1	...	0	1
<sp>	1	1	...	1	1	1	1	...	1	0

【図 7】



【图 8】



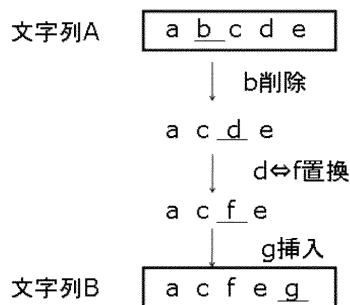
【図 9】

図9

	a	b	...	{	P	l	r	...	z	<sp>
a	0	1		0	1	1	1		1	1
b	1	0		1	0	1	1		1	1
⋮									⋮	⋮
{	0	1		0	1	1	1		1	1
p	1	0		1	0	1	1		1	1
l	1	1		1	1	0	1		1	1
r	1	1		1	1	1	0		1	1
⋮									⋮	⋮
z	1	1	...	1	1	1	1	...	0	1
<sp>	1	1	...	1	1	1	1	...	1	0

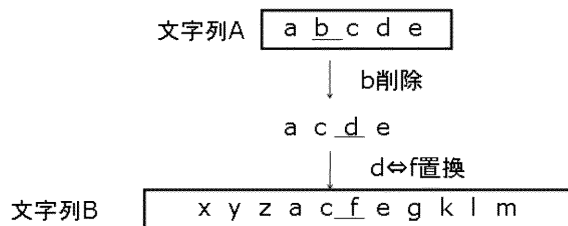
【図 10】

図10



【図 1 1】

図11



【図 12】

图12

	a	b	...	{	P		r	...	z	<sp>
a	0	1		0	1	1	1		1	1
b	1	0		1	1	1	1		1	1
:									:	:
{	0	1		0	1	1	1		1	1
p	1	1		1	0	1	1		1	1
	1	1		1	1	0	1		1	1
r	1	1		1	1	1	0		1	1
:									:	:
z	1	1	...	1	1	1	1	...	0	1
<sp>	1	1	...	1	1	1	1	...	1	0

【図 13】

図13

```

EditDistance(A,B){
    /* 初期化 */
    for i in 0,...,N
    {
        old[i] ← N
        new[i] ← N
    }
    Ed ← N
    old[0] ← 0

    /* 編集距離計算 */
    for j in 1,...,M
    {
        for i in 1,...,N
        {
            score0 ← old[i-1]+Matrix(A(i),B(j)) /*置換*/
            score1 ← old[i]+Matrix(SP,B(j)) /*挿入*/
            score2 ← new[i-1]+Matrix(A(i),SP) /*削除*/

            new[i] ← min(score0,score1,score2)
        }

        for i in 0,...,N
            old[i] ← new[i]

        if Ed > old[N]
            then Ed ← old[N]
    }

    /* 結果返却 */
    return Ed;
}

```

フロントページの続き

(56)参考文献 特開2010-267012 (JP, A)
特開2006-39954 (JP, A)
特開平7-36481 (JP, A)
特開2000-35799 (JP, A)
特表2004-508594 (JP, A)
米国特許出願公開第2010/0324900 (US, A1)

(58)調査した分野(Int.Cl., DB名)
G10L 13/00-15/34, 25/54
G06F 17/30