

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4672003号
(P4672003)

(45) 発行日 平成23年4月20日 (2011. 4. 20)

(24) 登録日 平成23年1月28日 (2011. 1. 28)

(51) Int. Cl.

F I

G 1 O L 17/00 (2006. 01)

G 1 O L 17/00 2 O O D

G 1 O L 15/00 (2006. 01)

G 1 O L 15/00 2 O O S

請求項の数 10 (全 24 頁)

(21) 出願番号 特願2007-503538 (P2007-503538)
 (86) (22) 出願日 平成17年2月18日 (2005. 2. 18)
 (86) 国際出願番号 PCT/JP2005/002589
 (87) 国際公開番号 W02006/087799
 (87) 国際公開日 平成18年8月24日 (2006. 8. 24)
 審査請求日 平成19年2月23日 (2007. 2. 23)

(73) 特許権者 000005223
 富士通株式会社
 神奈川県川崎市中原区上小田中4丁目1番
 1号
 (74) 代理人 110000040
 特許業務法人池内・佐藤アンドパートナーズ
 (72) 発明者 早川 昭二
 神奈川県川崎市中原区上小田中4丁目1番
 1号 富士通株式会社内
 審査官 山下 剛史

最終頁に続く

(54) 【発明の名称】 音声認証システム

(57) 【特許請求の範囲】

【請求項 1】

キーワードを音声入力させて認証を行うテキスト依存型の音声認証システムであって、
 発声可能な単位を最小単位として複数の部分に分割されたキーワードの音声入力を、
 前記部分毎に時間間隔をおいて複数回にわたって受け付ける入力部と、

利用者の登録キーワードを、前記発声可能な単位で作成された話者モデルとして予め
 格納した話者モデル格納部と、

前記入力部において1回の音声入力を受け付けられたキーワードの部分から、当該部分
 に含まれる音声の特徴量を求める特徴量変換部と、

前記特徴量変換部で求められた特徴量と前記話者モデルとの類似度を求める類似度計算
 部と、

前記類似度計算部で求められた類似度に基づき、前記複数回の音声入力により発声内容
 を判定する発声内容判定部と、前記複数回の音声入力の発声内容が過不足なく登録キー
 ワードを構成しうるか否かを判定するキーワード確認部と、

前記キーワード確認部による判定結果と、前記類似度計算部で求められた類似度とに基
 づき、認証を受理するか棄却するかを判定する認証判定部とを備えた音声認証システム。

【請求項 2】

前記発声可能な単位が音節である、請求項 1 に記載の音声認証システム。

【請求項 3】

前記話者モデル格納部において、登録キーワードを構成する各音節の話者モデルに個別

10

20

のインデックスが付与されており、

前記特徴量変換部が、前記音声入力で受け付けられたキーワードの部分から、音節毎の特徴量を求め、

前記類似度計算部が、前記音節毎の特徴量と前記話者モデルとの類似度を求め、

前記類似度計算部により求められた類似度に基づき、前記音声入力で受け付けられたキーワードの部分が、登録キーワードのどの音節に最も類似するかを判定する音節判定部をさらに備え、

前記キーワード確認部が、前記音節判定部の判定結果に基づき、前記複数回の音声入力により判定された音節を用いて登録キーワードを構成しうるか否かを判定する、請求項 2 に記載の音声認証システム。

10

【請求項 4】

前記発声が可能な単位が、数字の読みまたはアルファベットの読みである、請求項 1 ~ 3 のいずれかに記載の音声認証システム。

【請求項 5】

前記特徴量変換部により求められた特徴量から、不特定話者音声認識により、前記キーワードの部分の音韻を認識する音声認識部をさらに備え、

前記キーワード確認部が、前記音声認識部の認識結果に基づいて、前記複数回の音声入力により音声認識された結果を用いて、登録キーワードを構成しうるか否かを判定する、請求項 1 に記載の音声認証システム。

【請求項 6】

20

前記キーワードのある部分の音声入力の完了から所定の時間が経過しても次の部分の音声入力がなく、かつ、それまでの音声入力の発声内容の情報を用いて登録キーワードを構成できない場合、前記認証判定部が認証を棄却する、請求項 1 に記載の音声認証システム。

【請求項 7】

前記キーワードの部分の音声入力がある度に利用者の所在位置情報を取得する位置情報取得部と、

前回の音声入力時に前記位置情報取得部で取得された所在位置情報と、今回の音声入力時に前記位置情報取得部で取得された所在位置情報とを比較し、前回の音声入力時から今回の音声入力時までの間に利用者が所定の距離以上移動したかを確認する位置確認部とをさらに備えた、請求項 1 に記載の音声認証システム。

30

【請求項 8】

前記複数回の音声入力で受け付けられたキーワードの部分の全てについて、前記類似度計算部で求められた類似度を統合して統合類似度を求める類似度統合部をさらに備え、

前記認証判定部が、前記類似度統合部により求められた統合類似度に基づいて認証を受理するか棄却するかを判定する、請求項 1 に記載の音声認証システム。

【請求項 9】

前記入力部が、利用者の携帯端末より音声入力を受け付ける、請求項 1 に記載の音声認証システム。

【請求項 10】

40

キーワードを音声入力させて認証を行うテキスト依存型の音声認証システムを具現化するコンピュータプログラムであって、

発声が可能な単位を最小単位として複数の部分に分割されたキーワードの音声入力を、前記部分毎に時間間隔をおいて複数回にわたって受け付ける入力処理と、

1 回の音声入力で受け付けられたキーワードの部分から、当該部分に含まれる音声の特徴量を求める特徴量変換処理と、

利用者のキーワードが、前記発声が可能な単位で作成された話者モデルとして予め登録された話者モデル格納部を参照し、前記特徴量変換処理で求められた特徴量と、前記話者モデルとの類似度を求める類似度計算処理と、

前記類似度計算処理で求められた類似度に基づき、前記複数回の音声入力により発声内

50

容を判定する発声内容判定処理と、

前記複数回の音声入力の発声内容が過不足なく登録キーワードを構成しうるか否かを判定するキーワード確認処理と、

前記キーワード確認処理による判定結果と、前記類似度計算処理で求められた類似度とに基づき、認証を受理するか棄却するかを判定する認証判定処理とをコンピュータに実行させることを特徴とするコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声を用いて本人確認を行う音声認証システムに関し、特に、テキスト依存型の音声認証システムに関する。

【背景技術】

【0002】

従来、システム利用者の本人確認の一手段として、音声認証が実用化されている。近年は、モバイル環境における電子商取引等の際の個人認証にバイオメトリクス認証の一つである指紋認証が用いられることが多くなっている。しかし指紋認証には別途特別なセンサが必要であるが、音声を用いて個人認証する場合には、例えば携帯電話ではセンサであるマイクが既に装備されているため、モバイル環境での音声認証に携帯端末を適用することが期待されている。

【0003】

音声認証は、テキスト依存型とテキスト自由型の二種類に大きく分類される。テキスト依存型は、あらかじめ、キーワード（パスワード）やフレーズを利用者に読み上げさせてその音声を登録しておき、認証時には、登録時と同じキーワード若しくはフレーズを利用者に発声させ、認証を行う方式である。テキスト自由型は、発話内容には関係なく声質のみで認証を行う方式である。従って、テキスト自由型の場合は、キーワード等を決めておく必要がなく、利用者は任意の発声内容で登録・認証を行える。本発明は前者のテキスト依存型の音声認証に関するものである。

【0004】

テキスト依存型音声認証では、声質と発声内容（キーワード等）の秘密性の情報の両方に基づいて認証処理を行うので、比較的高い認証精度が得られる。しかし、認証時に傍に他人が居る環境では、秘密のキーワードが聞かれてしまう可能性がある。このため、プライバシーが保たれない環境で認証を行うような用途（例えば、店舗のレジカウンタや自動販売機において決済機能内蔵の携帯電話を用いて支払いをする際の本人確認等）には、利用者の抵抗があるため、テキスト依存型音声認証を採用し難い。

【0005】

また、キーワードが暴露された場合、発声内容の秘密性を認証に利用できなくなるため、認証精度が落ちてしまう。さらに、利用者本人が発声した秘密のキーワードを、他人がテープレコーダやＩＣレコーダで不正に録音し、認証時にこれを再生することにより、詐称（録音再生詐称）が成功する恐れもある。

【0006】

これらの問題に対して、録音再生詐称を防止するために、音声の位相差情報に基づいてスピーカからの再生であることを検出する方法（特許文献１参照）や、伝達特性を比較したり音声にＤＴＭＦ信号をかぶせて音声透かしを入れたりすることにより録音再生を検出する方法（特許文献２参照）等が提案されている。また、利用者に、認証毎に異なる内容のテキストを発声させることにより、録音再生詐称を防ぐシステムもある（特許文献３参照）。

【0007】

また、複数種類のパスワードをインデックスと対応づけて登録しておき、認証時には、パスワードと共に、そのパスワードに対応するインデックスを入力させることにより、パスワードが暴露されても詐称を防ぐ方法が提案されている（特許文献４）。また、音声認

10

20

30

40

50

証に用いる暗証番号を回りに暴露されないようにする手段として、番号それぞれに色が指定された画面を表示し、色の名前を認証時に発声させることにより、暗証番号の漏洩を防ぐ方法も提案されている（特許文献５）。

【０００８】

また、数字複数桁の音声入力により操作者の認証を行うシステムにおいて、数字桁の入力順序を予め多種類準備し、認証時にそのうちの１種を指示して入力させる方法もある（特許文献６）。また、秘密の記号列を変形して発声するように指定することにより、秘密のパスワードの漏洩を防ぐシステムもある（特許文献７）。さらに、音声入力される暗証番号にダミー番号をランダムに挿入することを指示することによって暗証番号の漏洩を防ぐ音声応答認識装置も知られている（特許文献８）。

10

【０００９】

【特許文献１】特開２００１－１０９４９号公報

【特許文献２】特表２００２－５１４３１８号公報

【特許文献３】特開２０００－１４８１８７号公報

【特許文献４】特開２０００－１８１４９０号公報

【特許文献５】特開２００２－３１１９９２号公報

【特許文献６】特開昭５９－１９１６４５号公報

【特許文献７】特開昭６３－２３１４９６号公報

【特許文献８】特開昭６３－２０７２６２号公報

【００１０】

20

しかしながら、上述の特許文献１～３に記載されているような録音再生詐欺対策を行ったとしても、録音された時点でパスワードは暴露されているため、認証精度は落ちてしまう。また、パスワードの漏洩を防いだり隠匿したりするには、上述の特許文献４～８に記載されている方法のように、発声内容の変更や特別な操作が必要であり、利用者にとって使いづらいという問題がある。また、認証の度に発声内容を指定する場合には、発話内容（何と発声したか）の秘密性の情報を認証に使えないため、高い精度は得られないという問題がある。

【発明の開示】

【００１１】

以上の問題に鑑み、本発明は、テキスト依存型の音声認証システムにおいて、キーワードの漏洩や録音詐欺を防ぐことにより、キーワードの秘密性を利用して高い認証精度を維持し得る音声認証システムを提供することを目的とする。

30

【００１２】

上記の目的を達成するために、本発明にかかる音声認証システムは、キーワードを音声入力させて認証を行うテキスト依存型の音声認証システムであって、発声が可能な単位を最小単位として複数の部分に分割されたキーワードの音声入力を、前記部分毎に時間間隔をおいて複数回にわたって受け付ける入力部と、利用者の登録キーワードを、前記発声が可能な単位で作成された話者モデルとして予め格納した話者モデル格納部と、前記入力部において１回の音声入力を受け付けられたキーワードの部分から、当該部分に含まれる音声の特徴量を求める特徴量変換部と、前記特徴量変換部で求められた特徴量と前記話者モデルとの類似度を求める類似度計算部と、前記類似度計算部で求められた類似度に基づき、前記複数回の音声入力により発声内容に関する情報を判定する発声内容判定部と、該発声内容判定部で判定された発声内容に関する情報に基づき、前記複数回の音声入力の発声内容を用いて登録キーワードを構成しうるか否かを判定するキーワード確認部と、前記キーワード確認部による判定結果と、前記類似度計算部で求められた類似度とに基づき、認証を受理するか棄却するかを判定する認証判定部とを備えたことを特徴とする。

40

【００１３】

この構成によれば、キーワードを発声が可能な単位を最小単位として複数の部分に分割して音声入力させることにより、周囲に他人が居る環境での認証時に、キーワード全体の漏洩を防止できる。また、キーワード全体を１回で続けて発声しないので、キーワード全

50

体が録音されるのを防げるので録音詐称も抑止できる。さらに、キーワード確認部において、複数回の音声入力により登録キーワードを構成するすべての音韻の音声入力完了したか否かを判定し、その判定結果と類似度計算部で求められた類似度とに基づいて、認証を受理するか棄却するかを判定することにより、精度の高いテキスト依存型音声認証システムを提供できる。

【 0 0 1 4 】

本発明にかかる音声認証システムにおいて、前記発声が可能な単位の一例は、音節である。この場合、前記話者モデル格納部において、登録キーワードを構成する各音節の話者モデルに個別のインデックスが付与されており、前記特徴量変換部が、前記音声入力を受け付けられたキーワードの部分から、音節毎の特徴量を求め、前記類似度計算部が、前記音節毎の特徴量と前記話者モデルとの類似度を求め、前記類似度計算部により求められた類似度に基づき、前記音声入力を受け付けられたキーワードの部分が、登録キーワードのどの音節に最も類似するかを判定する音節判定部をさらに備え、前記キーワード確認部が、前記音節判定部の判定結果に基づき、前記複数回の音声入力の発声内容を用いて登録キーワードを構成しうるか否かを判定することが好ましい。これにより、音声認識を行わなくても、認証時に音声入力されたキーワードの発話内容が登録キーワードの発話内容と一致しているかを判定できるため、音声認識用の文法辞書等を持つ必要がなく、システム構成を簡略化できる。

10

【 0 0 1 5 】

本発明にかかる音声認証システムにおいて、発声が可能な単位の他の例は、数字の読みまたはアルファベットの読みである。これにより、数字またはアルファベットあるいはこれらの組み合わせからなるキーワードに適応可能である。

20

【 0 0 1 6 】

本発明にかかる音声認証システムにおいて、前記特徴量変換部により求められた特徴量から、不特定話者音声認識により、前記キーワードの部分の音韻を認識する音声認識部をさらに備え、前記キーワード確認部が、前記音声認識部の認識結果に基づいて、登録キーワードを構成しうるか否かを判定することが好ましい。認証時のキーワードの発話内容が登録キーワードの発話内容と一致しているかを音声認識によって判定することにより、高精度な音声認証が可能となる。

【 0 0 1 7 】

本発明にかかる音声認証システムにおいて、前記キーワードのある部分の音声入力の完了から所定の時間が経過しても次の部分の音声入力がなく、かつ、それまでの音声入力の発声内容の情報を用いて登録キーワードを構成できない場合、前記認証判定部が認証を棄却することが好ましい。セキュリティを強化できるからである。

30

【 0 0 1 8 】

本発明にかかる音声認証システムにおいて、前記キーワードの部分の音声入力がある度に利用者の所在位置情報を取得する位置情報取得部と、前回の音声入力時に前記位置情報取得部で取得された所在位置情報と、今回の音声入力時に前記位置情報取得部で取得された所在位置情報とを比較し、前回の音声入力時から今回の音声入力時までの間に利用者が所定の距離以上移動したかを確認する位置確認部とをさらに備えたことが好ましい。キーワードを分割して音声入力する際に、音声入力毎に場所を移動した方が、周囲の他人にパスワード全体が知られてしまう可能性をさらに低くすることでセキュリティを強化できるからである。

40

【 0 0 1 9 】

また、本発明にかかる音声認証システムにおいて、前記複数回の音声入力を受け付けられたキーワードの部分の全てについて、前記類似度計算部で求められた類似度を統合して統合類似度を求める類似度統合部をさらに備え、前記認証判定部が、前記類似度統合部により求められた統合類似度に基づいて認証を受理するか棄却するかを判定することが好ましい。キーワード全体について類似度を統合して判定する方が、発声長が長くなるので、安定した類似度により判定できるからである。

50

【 0 0 2 0 】

また、本発明にかかる音声認証システムにおいて、前記入力部が、利用者の携帯端末より音声入力を受け付けることが好ましい。これにより、最初に音声入力した場所から離れた場所で2回目以降の音声入力を可能にするので、キーワード全体が他人に暴露される可能性を低くすることが可能だからである。

【 0 0 2 1 】

上記の目的を達成するために、本発明にかかるコンピュータプログラムは、キーワードを音声入力させて認証を行うテキスト依存型の音声認証システムを具現化するコンピュータプログラムであって、発声が可能な単位を最小単位として複数の部分に分割されたキーワードの音声入力を、前記部分毎に時間間隔をおいて複数回にわたって受け付ける入力処理と、1回の音声入力を受け付けられたキーワードの部分から、当該部分に含まれる音声の特徴量を求める特徴量変換処理と、利用者のキーワードが、前記発声が可能な単位で作成された話者モデルとして予め登録された話者モデル格納部を参照し、前記特徴量変換処理で求められた特徴量と、前記話者モデルとの類似度を求める類似度計算処理と、前記類似度計算処理で求められた類似度に基づき、前記複数回の音声入力により発声内容に関する情報を判定する発声内容判定処理と、前記発声内容判定処理で判定された発声内容に関する情報に基づき、前記複数回の音声入力の発声内容を用いて登録キーワードを構成するか否かを判定するキーワード確認処理と、前記キーワード確認処理による判定結果と、前記類似度計算処理で求められた類似度とに基づき、認証を受理するか棄却するかを判定する認証判定処理とをコンピュータに実行させることを特徴とする。

【 0 0 2 2 】

また、上記のコンピュータプログラムを記録した記録媒体も、本発明の一側面である。

【図面の簡単な説明】

【 0 0 2 3 】

【図1】図1は、本発明の第1の実施形態にかかる音声認証システムの概略構成を示すブロック図である。

【図2】図2は、前記音声認証システムにおいて、認証時入力キーワードの音節と登録キーワードの音節との類似度から、キーワードの正しさを判定する方法を説明する図である。

【図3】図3は、本発明の第1の実施形態にかかる音声認証システムの動作を示すフローチャートである。

【図4】図4は、本発明の第2の実施形態にかかる音声認証システムの概略構成を示すブロック図である。

【図5】図5は、本発明の第2の実施形態にかかる音声認証システムの動作を示すフローチャートである。

【図6】図6は、本発明の第3の実施形態にかかる音声認証システムの概略構成を示すブロック図である。

【図7】図7は、本発明の第3の実施形態にかかる音声認証システムの動作を示すフローチャートである。

【図8】図8は、本発明の第4の実施形態にかかる音声認証システムの概略構成を示すブロック図である。

【図9】図9は、本発明の第4の実施形態にかかる音声認証システムの動作を示すフローチャートである。

【発明を実施するための最良の形態】

【 0 0 2 4 】

(第1の実施形態)

本発明にかかる音声認証システムの一実施形態について、以下に説明する。

【 0 0 2 5 】

本実施形態にかかる音声認証システムでは、認証を受けようとする利用者は、携帯電話または携帯端末から、予め登録された秘密のキーワード(パスワード)を音声入力するこ

10

20

30

40

50

とにより、認証を受ける。この音声認証システムにおいて、利用者は、認証時に、キーワードを構成する全ての音節を、1音節または複数音節からなる部分に分割し、分割したキーワードを、時間間隔をあけて、さらに好ましくは場所を移動して、複数回に分けて音声入力する。分割したキーワードの入力順序は、キーワードの順番でも良いし、ランダムな順番でも良い。なお、認証時に、前記複数回の音声入力によって、キーワードを構成するすべての音節を発声することが必要である。

【0026】

例えば、キーワードが「カラオケ」である場合、利用者は、このキーワードを例えば「カラ」と「オケ」のように分割し、1回目の分割キーワード（例えば「カラ」）を音声入力してから数秒から数分経過した後に、2回目の分割キーワード（例えば「オケ」）を音声入力する。近くに居る他人にキーワードを聞かれることを避けるために、1回目の音声入力と2回目の音声入力との間で利用者が場所を移動することが好ましい。

10

【0027】

キーワードの分割は、「カオ」と「ラケ」、「ケカ」と「オラ」、「ラカ」と「ケオ」のように、音節の順序をランダムにしても良い。また、「カ」と「ケオラ」のように、単音節からなる分割キーワードがあっても良い。さらに、分割数は2に限定されない。

【0028】

そして、本実施形態の音声認証システムは、このように分割されて発声されたキーワードに基づいて、音声認証を行う。これにより、近くに他人が居る環境でキーワードを発声したとしても、一回の発声ではキーワードの一部しか聞かれることがないので、キーワード全体が他人に知られることを防止できる。

20

【0029】

なお、本実施形態にかかる音声認証システムは、キーワードを、1音節または複数音節からなる分割キーワードに分割して音声認証を行うので、音節単位で単語を区切ることが容易な言語（例えば日本語）での音声認証に適している。

【0030】

本実施形態にかかる音声認証システムは、図1に示すように、入力部11、特徴量変換部12、類似度計算部13、音節判定部14、類似度保管部15、待機時間管理部16、キーワード確認部17、類似度統合部18、認証判定部19、登録話者別音節モデルデータベース20、および、キーワード登録部21を備えている。

30

【0031】

入力部11は、利用者の携帯電話や携帯端末から、認証を受けようとする利用者の利用者IDと音声を入力する。なお、携帯電話のように利用者が1名に限定される場合には、利用者IDの入力手順は不要である。特徴量変換部12は、入力した音声を、話者別音節モデルとの比較に用いられる特徴量に変換する。

【0032】

音節判定部14は、特徴量変換部12で得られた特徴量と、利用者が登録話者別音節モデルデータベース20に予め登録してある話者音節モデルとを比較して類似度を求め、求められた類似度に基づいて、入力音声の各音節が、話者音節モデルのどれに最も類似するかを判定する。類似度保管部15は、入力音声の各音節について、前記音節判定部14で求められた、話者音節モデルに対する類似度を保管する。

40

【0033】

待機時間管理部16は、キーワードが複数回に分けて発声されるとき発声の時間間隔を監視する。キーワード確認部17は、音節判定部14による音節の認識結果に基づき、複数回の発声によってキーワードの全音節が発声されたかを確認する。類似度統合部18は、類似度保管部15に保管されている音節毎の類似度に基づき、認証時に発声されたキーワード音声が発声されているキーワード音声に、全体としてどの程度類似しているかを表す統合類似度を求める。認証判定部19は、類似度統合部18で求められた統合類似度に基づき、利用者の認証を受理するか棄却するかを判定する。

【0034】

50

キーワード登録部 21 は、音声認証システムの利用者がキーワードを登録する際に、当該利用者が入力部 11 へ音声入力したキーワードから音節モデルを作成し、登録話者別音節モデルデータベース 20 へ登録する。

【0035】

なお、上記の音声認証システムは、コンピュータおよびその周辺機器によって実現することができる。この場合、類似度保管部 15 および登録話者別音節モデルデータベース 20 は、コンピュータ内部またはこのコンピュータからアクセス可能な記憶装置によって実現できる。入力部 11、特徴量変換部 12、類似度計算部 13、音節判定部 14、待機時間管理部 16、キーワード確認部 17、類似度統合部 18、認証判定部 19、キーワード登録部 21 は、上記コンピュータのプロセッサに、これら各部の機能を具現化させるプログラムを実行させることによって実現できる。従って、このようなプログラムまたはこのプログラムを記録したコンピュータ読み取り可能な記録媒体も、本発明の一実施形態である。

【0036】

ここで、以上の構成にかかる本実施形態の音声認証システムにおける音声認証の手順について説明する。

【0037】

最初に、利用者が、音声認証システムの登録話者別音節モデルデータベース 20 へキーワードを登録する際の手順について説明する。利用者は、所望のキーワードを決めて、入力部 11 に接続された携帯電話等のマイクロフォンに対して発声する。このとき、利用者は、キーワードを、1 音節ずつ、無音区間で区切って発声する。例えばキーワードが「カラオケ」の場合、このキーワードには「カ」「ラ」「オ」「ケ」の 4 音節が含まれるので、利用者は、これらの 4 音節を無音区間で区切って発声する。入力部 11 は、入力された音節発声を、特徴量変換部 12 に渡す。特徴量変換部 12 においては入力された音声を MFCC (Mel Frequency Cepstral Coefficients) や LPC ケプストラムなどの音声の特徴を抽出した特徴ベクトル系列に変換する。その後順次、キーワード登録部 21 へ渡す。

【0038】

キーワード登録部 21 は、入力された音節のそれぞれに対して、入力された順に、インデックスを割り当てる。すなわち、上記の例では、「カ(一番目の音節)」へインデックス I1、「ラ(二番目の音節)」へインデックス I2、「オ(三番目の音節)」へインデックス I3、「ケ(四番目の音節)」へインデックス I4 を、それぞれ割り当てる。なお、キーワード登録部 21 は、音節が入力された順番にインデックスを割り当てれば十分であって、各音節の発声(音韻)が何であることを認識する必要はない。ただし後述の通り、同じ内容の音節がキーワードに含まれているか否かのチェックをこの時点で行う場合もある。

【0039】

キーワード登録部 21 は、このようにインデックスを割り当てた各音節から話者音節モデルを作成し、上述のように割り当てたインデックスと共に、当該利用者の利用者 ID 等と関連づけて登録話者別音節モデルデータベース 20 へ登録する。話者別音節モデルは、例えば不特定話者 GMM (Gaussian Mixture Model) を MAP (Maximum A Posteriori) 推定により特定話者に適応し、その適応を音節別を実施することで作成することができる。ただし、話者音節モデルは GMM に限定されず、HMM (Hidden Markov Model) など、インデックス別に話者モデルを作成できるものであれば、任意の手法を適用できる。また適応をせずに、登録された特徴ベクトル系列をそのままテンプレートとして保管することによっても話者別音節モデルを作成することも可能である。

【0040】

なお、認証精度を向上させるために、利用者にキーワードをできるだけ多く(3 回以上)繰り返して発声させるのが望ましい。この場合、一音節につき、繰り返し回数分(例え

10

20

30

40

50

ば3回繰り返し発声させた場合は3個)の特徴ベクトル系列を用いて話者別音節モデルが適応され、登録話者別音節モデルデータベース20へ登録される。前述のテンプレートとして保管する形態の話者別音節モデルの場合には、繰り返し回数分のテンプレートをもった話者別音節モデルが登録話者別音節モデルデータベース20へ登録される。

【0041】

次に、上記の手順によって利用者が「カラオケ」というキーワードを登録した後に、当該利用者が認証を受ける場合の手順について、図3のフローチャートを参照しながら説明する。なお、ここでは、利用者が携帯電話から音声入力を行い、かつ、キーワードを2回に分割して発声する例を示すが、キーワードの分割数は2に限定されない。また、キーワードの入力手段は携帯電話に限らない。

10

【0042】

携帯電話に認証開始ボタンがある場合にはボタンを押すことにより、あるいは認証が必要なアプリを起動した際、あるいは他のアプリケーションの実行途中で認証要求があった際に、認証処理が開始される。

【0043】

最初に、入力部11は、利用者が携帯電話のキーパッドを用いて入力した利用者ID等を受け付ける(Op11)。なお、携帯電話のように利用者が1名に限定される場合には利用者IDの入力手順は不要となる。次に、利用者が、携帯電話のマイクロフォンに向けて1回目の分割キーワードを発声すると、入力部11はこの音声を受け取る(Op12)。なお、ここでは、前記利用者が、「カラ」と「オケ」にキーワードを分割し、音声入力するものとする。利用者は、1回目の分割キーワードとして、「カ」「ラ」を一音節ずつ無音区間で区切って音声入力する。以降、このように音声入力された「カ」「ラ」の各音節を、入力された順序に従ってS1、S2と表す。

20

【0044】

なお、1回目の分割キーワードの入力が完了したか否かは、無音区間が所定の時間以上続いたか否かによって判断することが可能である。あるいは、分割キーワードの入力が完了した時に、利用者に所定のキー操作をさせるようにしても良い。

【0045】

入力された音節S1、S2のそれぞれは、特徴量変換部12において、登録時と同一の分析方法によりMFCCやLPCケプストラムなどの、話者音節モデルとの比較に用いられる特徴ベクトル系列(特徴量)に変換される(Op13)。特徴量変換部12で得られた特徴量は類似度計算部13へ送られる。

30

【0046】

次に、類似度計算部13は、登録話者別音節モデルデータベース20から、当該利用者の利用者IDに関連づけて格納されている当該利用者の話者音節モデル(すなわち、上述のインデックスI1~I4の話者音節モデル)を抽出し、抽出した話者音節モデルに対する前記特徴量の類似度を計算する(Op14)。ここで、インデックスI1の話者音節モデルに対する音節S1の類似度を $C_{S1:I1}$ のように表記するものとする。すなわち、Op14において、例えば音節S1については、 $C_{S1:I1}$ 、 $C_{S1:I2}$ 、 $C_{S1:I3}$ 、 $C_{S1:I4}$ の4つの値が類似度として求めら、音節S2については、 $C_{S2:I1}$ 、 $C_{S2:I2}$ 、 $C_{S2:I3}$ 、 $C_{S2:I4}$ の4つの値が類似度として求められる。求められた類似度は、音節判定部14へ送られる。

40

【0047】

なお、音節モデルをテンプレートとして登録したときに、キーワードを繰り返して発声させた場合のように、同じ音節について複数回分の話者別音節テンプレートが登録されている場合は、それらの話者別音節テンプレートの全てについて、類似度の計算をそれぞれ行う。従って、キーワード登録時に、「カ」「ラ」「オ」「ケ」を3回繰り返し発声させてテンプレートを話者音節モデルとして登録した場合は、それらの話者音節モデルのインデックスを $I1(1)$ 、 $I1(2)$ 、 $I1(3)$ ・・・のように割り当てると、Op14では、例えば音節S1について、 $C_{S1:I1(1)}$ 、 $C_{S1:I1(2)}$ 、 $C_{S1:I1(3)}$ のように

50

$C_{S1:I2(1)}$ 、 $C_{S1:I2(2)}$ 、 $C_{S1:I2(3)}$ ・・・の12個の値が、類似度として得られる。

【0048】

次に、音節判定部14は、類似度計算部13で求められた類似度に基づき、 $S1$ 、 $S2$ の各音節について、登録話者別音節モデルデータベース20に格納されている話者音節モデルの中で最も類似度の高いモデルのインデックスを選択する(Op15)。

【0049】

例えば、音節 $S1$ として利用者が発声したのは「カ」であるため、音節 $S1$ についての類似度 $C_{S1:I1}$ 、 $C_{S1:I2}$ 、 $C_{S1:I3}$ 、 $C_{S1:I4}$ のうち、インデックス $I1$ (登録時の音節「カ」)の話者音節モデルに対する類似度 $C_{S1:I1}$ が最大値となる。従って、図2に示すように、音節判定部14は、発声された音節 $S1$ がインデックス $I1$ の話者音節モデルに相当すると判定し、発声済み音節のインデックスデータとして「 $I1$ 」を類似度保管部15に保管する。また、これと共に、音節 $S1$ についての類似度の最大値である $C_{S1:I1}$ が、音節 $S1$ の類似度(音節類似度)として類似度保管部15に保管される。

10

【0050】

また、上記と同様にして、音節 $S2$ についても、話者音節モデルのうち最も類似度の高いモデルのインデックスが選択される。ここで、音節 $S2$ として利用者が発声したのは「ラ」であるため、音節 $S2$ についての類似度 $C_{S2:I1}$ 、 $C_{S2:I2}$ 、 $C_{S2:I3}$ 、 $C_{S2:I4}$ のうち、インデックス $I2$ (登録時の音節「ラ」)の話者音節モデルに対する類似度 $C_{S2:I2}$ が最大値となる。これにより、図2に示すように、音節判定部14は、音節 $S2$ がインデックス $I2$ に相当すると判定し、発声済み音節のインデックスデータとして「 $I2$ 」を類似度保管部15に保管する。また、これと共に、音節 $S2$ についての類似度の最大値である $C_{S2:I2}$ も、音節 $S2$ の音節類似度として類似度保管部15に保管される。

20

【0051】

ここで、2回目の分割キーワードの入力がまだ済んでいないので(Op16の結果がNO)、Op17へ進む。

【0052】

なお、待機時間管理部16は、入力部11に1回目の分割キーワード(「カラ」)が入力された後の経過時間を、タイマ(図示せず)により測定している。そして、2回目の分割キーワード入力が所定時間(例えば10分)内に開始されない場合は、タイムアウトとしてOp18へ進み、1回目の分割キーワードの処理結果を、類似度保管部15等から消去する(Op18)。この場合、利用者は、認証を受けるためには、1回目の分割キーワード入力からやり直さなければならない。

30

【0053】

なお、第三者へのキーワードの暴露を確実に防止するために、1回目の分割キーワードの音声入力から所定時間(例えば2分)が経過しなければ2回目の分割キーワードの音声入力を受け付けないようにしても良い。この場合、1回目の分割キーワードの入力が終了したときに、2回目の分割キーワードの音声入力までに一定時間をおくように指示するメッセージを利用者の携帯電話等のディスプレイへ表示したり、同様の内容の音声メッセージを当該携帯電話へ送ったりすることが好ましい。

40

【0054】

一方、Op17において、上記の所定時間内に2回目の分割キーワード入力開始された場合は、待機時間管理部16がタイマをストップ・クリアすると共に、Op19へ進み、2回目の分割キーワード入力を受け付け(Op19)、その後Op13へ進む。

【0055】

この例では、利用者は、2回目の分割キーワードとして、「オ」「ケ」を一音節ずつ区切って発声する。以降、このように音声入力された「オ」「ケ」の各音節を、入力された順序に従って $S3$ 、 $S4$ と表すものとする。

50

【 0 0 5 6 】

入力された音節 S_3 , S_4 のそれぞれは、特徴量変換部 1 2 において特徴量に変換される (Op 1 3) 。次に、類似度計算部 1 3 は、登録話者別音節モデルデータベース 2 0 から、当該利用者の利用者 ID に関連づけて格納されている話者音節モデル (すなわち、インデックス $I_1 \sim I_4$ の話者音節モデル) を抽出し、抽出した話者音節モデルに対する前記特徴量の類似度を計算する (Op 1 4) 。

【 0 0 5 7 】

これにより、インデックス $I_1 \sim I_4$ の話者音節モデルのそれぞれに対して、音節 S_3 , S_4 のそれぞれについて特徴量変換部 1 2 で得られた特徴量との類似度 $C_{S_3 : I_1}$ 、 $C_{S_3 : I_2}$ 、 $C_{S_3 : I_3}$ 、 $C_{S_3 : I_4}$ 、 $C_{S_4 : I_1}$ 、 $C_{S_4 : I_2}$ 、 $C_{S_4 : I_3}$ 、 $C_{S_4 : I_4}$ が計算される。

10

【 0 0 5 8 】

なお、キーワードの登録時にキーワードを繰り返し発声させることにより、同じ音節について複数回分の話者音節モデルが登録されている場合は、それらの話者音節モデルの全てについて、類似度の計算をそれぞれ行う。従って、キーワード登録時に、「カ」「ラ」「オ」「ケ」を 3 回繰り返し発声させて話者音節モデルが登録されている場合は、それらの話者音節モデルのインデックスを $I_1(1)$ 、 $I_1(2)$ 、 $I_1(3)$ … のように割り当てると、例えば音節 S_3 について、 $C_{S_3 : I_1(1)}$ 、 $C_{S_3 : I_1(2)}$ 、 $C_{S_3 : I_1(3)}$ 、 $C_{S_3 : I_2(1)}$ 、 $C_{S_3 : I_2(2)}$ 、 $C_{S_3 : I_2(3)}$ … の 1 2 個の値が、類似度として得られる。

20

【 0 0 5 9 】

次に、音節判定部 1 4 は、類似度計算部 1 3 で求められた類似度に基づき、 S_3 , S_4 の各音節について、登録話者別音節モデルデータベース 2 0 に格納されている話者音節モデルの中で最も類似度の高いモデルのインデックスを選択する (Op 1 5) 。

【 0 0 6 0 】

この例では、音節 S_3 として利用者が発声したのは「オ」であるため、音節 S_3 についての類似度 $C_{S_3 : I_1}$ 、 $C_{S_3 : I_2}$ 、 $C_{S_3 : I_3}$ 、 $C_{S_3 : I_4}$ のうち、 $C_{S_3 : I_3}$ が最大値となる。従って、音節判定部 1 4 は、音節 S_3 がインデックス I_3 に相当すると判定し、発声済み音節のインデックスデータとして「 I_3 」を類似度保管部 1 5 に保管する。また、これと共に、音節 S_3 についての類似度の最大値である $C_{S_3 : I_3}$ も、音節 S_3 の音節類似度として、類似度保管部 1 5 に保管される。

30

【 0 0 6 1 】

上記と同様にして、音節 S_4 についても、話者音節モデルのうち最も類似度の高いモデルのインデックスが選択される。ここで、音節 S_4 として利用者が発声したのは「ケ」であるため、音節 S_4 についての類似度 $C_{S_4 : I_1}$ 、 $C_{S_4 : I_2}$ 、 $C_{S_4 : I_3}$ 、 $C_{S_4 : I_4}$ のうち、 $C_{S_4 : I_4}$ が最大値となる。従って、音節判定部 1 4 は、音節 S_4 がインデックス I_4 に相当すると判定し、発声済み音節のインデックスデータとして「 I_4 」を類似度保管部 1 5 に保管する。これと共に、音節 S_4 についての類似度の最大値である $C_{S_4 : I_4}$ も、音節 S_3 の音節類似度として、類似度保管部 1 5 に保管される。

40

【 0 0 6 2 】

以上のとおり、2 回目の分割キーワードの処理が終了すると (Op 1 6 にて YES) 、キーワード確認部 1 7 が、類似度保管部 1 5 に保管されている発声済み音節のインデックスデータに基づき、1 回目と 2 回目の発声によってキーワードの全ての音節が発声されたかを確認する (Op 2 0) 。上記の例の場合、類似度保管部 1 5 には、発声済み音節のインデックスデータとして「 I_1 」、「 I_2 」、「 I_3 」、「 I_4 」が保管されている。すなわち、登録されているキーワードの全てのインデックス ($I_1 \sim I_4$) が揃っているため、キーワード確認部 1 7 は、キーワードの全ての音節が発声されたことを確認できる。Op 2 0 の判断結果が NO の場合は、その旨が認証判定部 1 9 へ通知され、認証判定部 1 9 は当該利用者の認証を棄却する (Op 2 1) 。

【 0 0 6 3 】

50

例えば、上記の例において、1回目の分割キーワードとして「カ」「ラ」と発声し、2回目の分割キーワードとして「カ」「ケ」と発声した場合は、「オ」が発声されていないために、発声済み音節のインデックスデータ中に「オ」のインデックスI3が存在しない。このため、認証は棄却される。

【0064】

一方、Op20の判断結果がYESの場合は、類似度統合部18が、類似度保管部15に保管されている、音節S1～S4のそれぞれの音節類似度($C_{S1} : I_1$, $C_{S2} : I_2$, $C_{S3} : I_3$, $C_{S4} : I_4$)に基づいて、発声されたキーワード全体が登録キーワードにどの程度似ているかを表す統合類似度を算出する(Op22)。統合類似度としては、各音節の音節類似度の総計(すなわち、この例では、 $C_{S1} : I_1 + C_{S2} : I_2 + C_{S3} : I_3 + C_{S4} : I_4$)を用いても良いし、各音節の音節類似度の平均値を用いても良い。あるいは、各音節の音節類似度の総和をフレーム数等の時間長で平均化したものを、統合類似度として用いても良い。

【0065】

このように類似度統合部18で求められた統合類似度は、認証判定部19へ送られ、認証判定部19が、所定の閾値と統合類似度とを比較する。認証判定部19は、統合類似度が所定の閾値を超えていれば当該利用者の認証を受理し、下回ってれば棄却する(Op23)。

【0066】

本実施形態にかかる音声認証システムでは、以上の手順により、音声による利用者の認証が可能である。

【0067】

なお、上記の実施形態では、登録されたキーワードを構成する音節の全てが、複数回に分割されたキーワード発声によって過不足なく発声されたか否かを認識する。従って、上記の具体例では、「カラオケ」というキーワードを「カラ」と「オケ」との2つに分割したが、例えば、「カオ」と「ラケ」、「ケカ」と「オラ」、「ラカ」と「ケオ」のように、音節の順序をランダムにしたキーワードの分割方法をとっても、正しい音声認証が可能である。

【0068】

なお、本実施形態では、キーワードの登録時も認証時も音韻を認識しないので、キーワードが同じ音の音節を2つ以上含む場合は、それらの音節を区別することができない。そこで、認証時に音声入力された音節の複数が、登録されている話者音節モデルの複数に対して相互に高い類似度を示した場合は、これらは同じ音節であると見なして処理することが好ましい。

【0069】

例えば、キーワードが「スキヤキ」である場合、このキーワードの登録時には、キーワード登録部21が、「ス(一番目の音節)」ヘインデックスI1、「キ(二番目の音節)」ヘインデックスI2、「ヤ(三番目の音節)」ヘインデックスI3、「キ(四番目の音節)」ヘインデックスI4をそれぞれ割り当てて、登録話者別音節モデルデータベース20に話者音節モデルを格納する。

【0070】

そして、認証時に、例えば利用者が1回目の分割キーワードとして「ス(音節S1)」「キ(音節S2)」, 2回目の分割キーワードとして「ヤ(音節S3)」「キ(音節S4)」と発声したとする。この場合、音節S2と音節S4の両方が、インデックスI2, I4の話者音節モデルに対して共に高い類似度を示すが、音節S2とインデックスI2の話者音節モデルとの類似度、音節S4とインデックスI2の話者音節モデルとの類似度、音節S2とインデックスI4の話者音節モデルとの類似度、音節S4とインデックスI4の話者音節モデルとの類似度は、互いに大きな差はないため、音節S2との類似度が最も高いインデックスと、音節S4との類似度が最も高いインデックスとが一致してしまう可能性もある。例えば音節S2とS4との両方がインデックスI2と最も類似すると判断され

10

20

30

40

50

た場合であれば、インデックス I 4 の音節が発声されていないとして、認証が棄却されてしまう。

【 0 0 7 1 】

従って、このように、複数の音節が、話者音節モデルの複数に対して相互に高い類似度であって、かつ、他の話者音節モデルに対する類似度よりも顕著に高い類似度を示す場合は、Op 15において最も類似度の高いインデックスを選択するのではなく、それら複数の音節が同じ音であると見なし、それらの複数の音節との類似度が高い話者音節モデルのインデックスをすべて類似度保管部 15へ保持すれば良い。あるいは、登録時に、複数の音節発声が、話者別音節モデルの複数に対して相互に高い類似度か否かを調べる等により、それら複数の音節が同じ音であることをチェックし、それを示す情報を記憶しておき、それらの音節に類似する話者音節モデルのインデックスとして同一のインデックスが類似度保管部 15に保管されていれば、全てのインデックスが揃っていなくても認証を棄却しないようにしても良い。

10

【 0 0 7 2 】

これにより、上記の例において、「ス」「キ」「ヤ」「キ」の入力音節に対する類似話者音節モデルのインデックスとして、例えば、「I 1」、「I 2」、「I 3」、「I 2」が類似度保管部 15に保管されていても、キーワードの全ての音節が発声されたものと正しく判断することができる。

【 0 0 7 3 】

以上、本発明の一実施形態にかかる音声認証システムについて説明したが、上記の説明は、本発明を実施するための一具体例に過ぎず、発明の範囲内で種々の変更が可能である。

20

【 0 0 7 4 】

例えば、上記の実施形態では、認証時にキーワードを2回に分割して発声する例を説明したが、キーワードの分割数は、2に限定されず、任意である。キーワードの分割数が3以上の場合は、分割キーワードの音声入力と、入力された分割キーワードに対する図3のOp 13～Op 15の処理とを、分割数だけ繰り返し繰り返せば良い。従って、「カ」と「ラオケ」、「オケラ」と「カ」、「ラ」と「ケ」と「オカ」、あるいは、「オ」と「ラ」と「カ」と「ケ」といったような分割によっても、正しい音声認証が可能である。

【 0 0 7 5 】

30

さらに、認証を受けようとする利用者が分割数を任意に決定できるようにしても良い。この場合、分割キーワードを複数回発声することによってキーワード全体の発声が完了したときに、利用者に例えば携帯電話の特定のボタンを押させることとし、音声認証システムは、入力部 11において当該ボタンが押されたことを検知すると、図3のOp 20以降の処理へ移行するようにすれば良い。あるいは、分割キーワードの発声が終了する毎に、それまでに発声された音節の集合と登録キーワードを構成する音節とを比較し、キーワードを構成する音節が全て発声済みであると判断されたときに、図3のOp 20以降の処理へ移行するようにしても良い。

【 0 0 7 6 】

また、本実施形態では、2回目の分割キーワードが入力した後に、キーワード確認部 17でキーワード確認を行い、さらに、類似度統合部 18でキーワード全体としての統合類似度を算出し、認証判定部 19で認証を受理するか棄却するかを判定する構成とした。しかし、1回目の分割キーワードの音声入力終了した時点で、類似度統合部 18が、1回目の分割キーワードに含まれる音節だけから統合類似度を算出するものとしても良い。

40

【 0 0 7 7 】

この場合、1回目の分割キーワードから算出された統合類似度が所定の閾値を下回る場合は、1回目の分割キーワードの音声入力をやりなおさせても良い。このようにすれば、例えば2回目の分割キーワードの類似度が高いが1回目の分割キーワードの類似度が低い場合に、最初から音声入力をやり直す手間が軽減されるという利点がある。

【 0 0 7 8 】

50

なお、セキュリティを厳しくする観点からは、1回目の分割キーワードの類似度が低い場合は、2回目の分割キーワードを入力させることなく、認証判定部19が認証を棄却するようにしても良い。

【0079】

また、2回目の分割キーワードの入力が終了した後に、2回目の分割キーワードの類似度が低い場合は、キーワード確認部17により全音節が発声済みであることが確認されていることを条件として、2回目の分割キーワードの音声入力をやりなおさせても良い。上記の条件を付けるのは、全音節が発声済みであることが確認されていない場合に2回目の分割キーワードの音声入力のやりなおしを認めると、キーワードの試行を許すこととなり、詐称者による認証突破が可能となってしまうからである。

10

【0080】

なお、安定した認証結果を得るためには、分割キーワード毎に類似度を求める方法よりも、分割キーワードの音声入力が全て終了した後にキーワード全体としての統合類似度を算出する方法の方が有利である。分割キーワードは発声長が短いので安定した類似度が得られにくいことがあるのに対して、キーワード全体としての統合類似度は発声長が長いので比較的安定した類似度が得られるからである。そのため、1回目の分割キーワードの類似度で判定する場合のしきい値は全体よりもやや緩めに設定することになる。

【0081】

(第2の実施形態)

本発明にかかる音声認証システムの他の実施形態について、以下に説明する。なお、第1の実施形態で説明した構成と同様の機能を有する構成については、第1の実施形態で用いた参照記号と同じ記号を付記し、その詳細な説明は省略する。

20

【0082】

本実施形態にかかる音声認証システムは、英語等のように、音節単位でのキーワード分割が必ずしも容易ではない言語による音声認証にも適している。

【0083】

このため、本実施形態にかかる音声認証システムは、図4に示すように、入力部11、特徴量変換部12、類似度計算部13、類似度保管部15、待機時間管理部16、キーワード確認部17、類似度統合部18、認証判定部19、話者モデルデータベース22、キーワード登録部21、文法生成部23、認識文法格納部24、および、音声認識部25を備えている。

30

【0084】

ここで、本実施形態の音声認証システムにおける音声認証の手順について、図5を参照しながら説明する。

【0085】

最初に、利用者が、音声認証システムの話者モデルデータベース22へキーワードを登録する際の手順について説明する。利用者は、所望のキーワードを決めて、入力部11に接続された携帯電話等のマイクロフォンに対して発声する。本実施形態では、文法生成部23が、音声入力されたキーワードを音声認識(不特定話者音声認識)し、認識結果としての認識文法を認識文法格納部24へ登録する。なお登録前に、画面などに認識結果を表示して利用者が登録したキーワードの内容を確認させるのが望ましい。また認識結果を修正する手段(ボタン等)を与えるのが望ましい。

40

【0086】

なお、不特定話者音声認識で単音節認識を行うことは精度的に難しいので、利用者は、キーワードを登録する際に、キーワードの発声内容を、例えば携帯電話のキー等を用いて、入力部11に対してテキスト入力するようにするのが望ましい。入力テキスト(キーワードの発声内容)と発声されたキーワード音声は、入力部11から文法生成部23へ渡される。

【0087】

文法生成部23は、テキスト入力された発声内容から、利用者が入力し得るキーワード

50

分割のパターンを認識するための文法を生成する。この文法は、音節単位に限定されず、音素単位の分割による全てのパターンを網羅するように生成される。日本語のように音節単位で切りやすい言語の場合には音節単位で分割しうる全てのパターンを網羅するように生成されるが、英語のように音節単位で切ることが難しい言語の場合には、音素単位の分割による全てのパターンを網羅するように生成される。例えば `verification` というキーワードに対しては `ve`, `veri`, `verifi`, `tion`, `cation`, `fication`, `verifica` 等のパターンを受理する文法が生成される。生成された文法は、キーワード登録部 21 に渡されると共に、認識文法格納部 24 へ格納される。

【0088】

10

キーワード登録部 21 は、登録されたキーワードの音声を用いて適応した話者モデル（すなわち音素または音節単位の話者モデル）を作成し、当該利用者の利用者 ID と関連づけて、話者モデルデータベース 22 へ登録する。音素単位の話者モデルの作成方法としては、不特定話者音声認識によりキーワードの音声と音素の時間対応づけ（アライメントと呼ぶ）を行い、キーワードを構成する音素別に適応を行い、話者別音素モデルを話者モデルとして作成する方法がある。また、適応方法としては MLLR (Maximum Likelihood Linear Regression) 法、MAP 法等、モデル適応により話者適応を実現する方法であれば何でもよい。音節単位で話者モデルを作成する場合には音節単位でキーワードとの時間対応付けを行ったのち、話者適応を行い作成する。

【0089】

20

上記の手順によりキーワード登録が完了した後に、利用者が音声入力により認証を受ける際の手順について、以下に説明する。なお、以下の例においても、キーワードを 2 回に分割して音声入力する例を説明するが、キーワードの分割数は 2 に限定されない。

【0090】

最初に、入力部 11 は、利用者が携帯電話のキーパッドを用いて入力した利用者 ID 等を受け付ける (Op 31)。なお、携帯電話のように利用者が 1 名に限定される場合には利用者 ID の入力手順は不要となる。次に、利用者が、携帯電話のマイクロフォンに向けて 1 回目の分割キーワードを発声すると、入力部 11 はこの音声を受け取る (Op 32)。なお、本実施形態では、利用者は音節毎に区切って音声入力をしなくても良い。例えば、キーワードが「カラオケ」である場合、1 回目の分割キーワードを「カラ」とする場合、第 1 の実施形態のように「カ」と「ラ」の間に無音区間を入れずに、「カラ」と一気に発声しても良い。

30

【0091】

また、英語等は、話者が音節を厳密に意識しない傾向が強いため、キーワードを分割した際に、誤って音節が欠落したり重複したりする可能性がある。音節の重複はキーワード認証に際してあまり問題とならないため、本実施形態では、例えばキーワードが「`verification`」である場合に、「`verifi`」と「`fication`」のように、分割キーワードにおいて音節が重複することは許容される。

【0092】

入力された分割キーワードは、特徴量変換部 12 において、MFCC や LPC ケプストラムなどの、音声認識に用いられる特徴量に変換される (Op 33)。特徴量変換部 12 で得られた特徴量は音声認識部 25 へ送られる。

40

【0093】

次に、音声認識部 25 は、認識文法格納部 24 に格納されている認識文法を参照し、分割キーワードとして発声された音声の音韻を不特定話者音声認識で認識する (Op 34)。認識結果は、類似度計算部 13 へ送られる。キーワードと異なる発声があった場合等、認識スコアが非常に悪い場合には「認識できない」と判定した方が良く、この場合には「認識できない」という結果を類似度計算部 13 に送る。

この例では不特定話者音声認識で認識したが、登録時に適応した話者モデルを利用して認識させても良い。この場合、音声認識時に計算した類似度がそのまま話者モデルに対する

50

類似度となるので、類似度計算部 13 での類似度計算は不要となる。

【0094】

次に、類似度計算部 13 は、前記の認識結果に基づき、話者モデルデータベース 22 から、当該利用者の話者モデルのうち、前記 Op 34 で認識された音韻に対応する話者モデルを抽出し、抽出した話者モデルに対する前記特徴量の類似度を計算する (Op 35)。

【0095】

Op 35 の結果に基づき、発声済みの分割キーワードから認識された音韻が、登録キーワードのどの音素を網羅しているかを表す情報 (発声済み音素データ) が、類似度保管部 15 へ保管される (Op 36)。これと共に、認識された音韻の話者モデルに対する類似度も、類似度保管部 15 に保管される。

10

【0096】

ここで、2 回目の分割キーワードの入力がまだ済んでいないので (Op 37 の結果が NO)、Op 38 へ進む。

【0097】

なお、待機時間管理部 16 は、入力部 11 に 1 回目の分割キーワードが入力された後の経過時間を、タイマにより測定している。そして、2 回目の分割キーワード入力が所定時間内に開始されない場合は、タイムアウトとして Op 39 へ進み、1 回目の分割キーワードの処理結果を、類似度保管部 15 等から消去する (Op 39)。この場合、利用者は、認証を受けるためには、1 回目の分割キーワード入力からやり直さなければならない。

【0098】

20

一方、Op 38 において、上記の所定時間内に 2 回目の分割キーワード入力が開始された場合は、待機時間管理部 16 がタイマをストップ・クリアすると共に、Op 40 へ進み、2 回目の分割キーワード入力を受け付け (Op 40)、その後 Op 33 へ進む。

【0099】

そして、2 回目の分割キーワードに対して、Op 33 ~ Op 36 の処理を行った後、キーワード確認部 17 が、類似度保管部 15 に保管されている発声済み音素データに基づき、1 回目と 2 回目の発声によってキーワードの全ての音素が発声されたかを確認する (Op 41)。Op 41 の判断結果が NO の場合は、その旨が認証判定部 19 へ通知され、認証判定部 19 は当該利用者の認証を棄却する (Op 42)。

【0100】

30

一方、Op 41 の判断結果が YES の場合は、類似度統合部 18 が、類似度保管部 15 に保管されている分割キーワード毎の類似度に基づいて、発声されたキーワード全体が登録キーワードにどの程度似ているかを表す統合類似度を算出する (Op 43)。統合類似度としては、分割キーワードの類似度の総計を用いても良いし、分割キーワードの類似度の平均値を用いても良い。

【0101】

このように類似度統合部 18 で求められた統合類似度は、認証判定部 19 へ送られ、認証判定部 19 が、所定の閾値と統合類似度とを比較する。認証判定部 19 は、統合類似度が所定の閾値を超えていれば当該利用者の認証を受理し、下回ってれば棄却する (Op 44)。

40

【0102】

本実施形態にかかる音声認証システムでは、以上の手順により、音声による利用者の認証が可能である。本実施形態では、上述のとおり、複数音節をキーワード分割の最小単位としても良いし、形態素または単語など、発声の単位とできるものなら、音節以外をキーワード分割の最小単位としてもよい。

【0103】

また、キーワードが数字列 (例えば「1234」) である場合、日本語であれば「イチ」、「ニ」、「サン」、「シ」のように、あるいは英語であれば「one」、「two」、「three」、「four」のように、数字を単位としてキーワードの確認および音声認証を行ってもよい。ただし、この場合は、数字認識用の文法を用いる必要がある。

50

【 0 1 0 4 】

同様に、キーワードが数字とアルファベット等の混合（例えば「 1 a 2 b 」）である場合は、数字とアルファベットの認識用文法を用いる必要がある。

【 0 1 0 5 】

（第 3 の実施形態）

本発明にかかる音声認証システムのさらに他の実施形態について、以下に説明する。なお、第 1 または第 2 の実施形態で説明した構成と同様の機能を有する構成については、それらの実施形態で用いた参照記号と同じ記号を付記し、その詳細な説明は省略する。

【 0 1 0 6 】

本実施形態にかかる音声認証システムは、2 回目以降の分割キーワードが音声入力される毎に、前回の分割キーワードの音声入力時から利用者が場所を移動しているか否かを確認し、場所を移動していない場合は音声入力を受け付けない。

10

【 0 1 0 7 】

このため、本実施形態にかかる音声認証システムは、図 6 に示すように、入力部 1 1、特徴量変換部 1 2、類似度計算部 1 3、音節判定部 1 4、類似度保管部 1 5、待機時間管理部 1 6、キーワード確認部 1 7、類似度統合部 1 8、認証判定部 1 9、登録話者別音節モデルデータベース 2 0、キーワード登録部 2 1、位置情報取得部 3 1、および、位置確認部 3 2 を備えている。すなわち、本実施形態にかかる音声認証システムは、第 1 の実施形態にかかる音声認証システムに、位置情報取得部 3 1、および、位置確認部 3 2 をさらに備えた構成である。

20

【 0 1 0 8 】

位置情報取得部 3 1 は、利用者の携帯電話等に組み込まれた GPS (G l o b a l P o s i t i o n i n g S y s t e m) により、利用者が分割キーワードを音声入力する都度、利用者の現在位置を表す情報を取得する。位置確認部 3 2 は、位置情報取得部 3 1 により取得された情報に基づき、前回の分割キーワードの音声入力時から利用者が場所を移動しているか否かを確認する。

【 0 1 0 9 】

以上の構成にかかる、本実施形態の音声認証システムの動作について、図 7 を参照しながら説明する。

【 0 1 1 0 】

30

最初に、入力部 1 1 は、利用者が携帯電話のキーパッドを用いて入力した利用者 ID 等を受け付ける (O p 5 1)。なお携帯電話のように利用者が 1 名に限定される場合には利用者 ID の入力手順は不要となる。次に、利用者が、携帯電話のマイクロフォンに向けて 1 回目の分割キーワードを発声すると、入力部 1 1 はこの音声を受け取る (O p 5 2)。このとき同時に、位置情報取得部 3 1 が、利用者の携帯電話に組み込まれた GPS により、利用者の現在位置を表す情報を取得し、取得した情報を位置確認部 3 2 へ渡す (O p 5 3)。

【 0 1 1 1 】

入力された 1 回目の分割キーワードの各音節は、特徴量変換部 1 2 において特徴量に変換される (O p 5 4)。特徴量変換部 1 2 で得られた特徴量は類似度計算部 1 3 へ送られる。

40

【 0 1 1 2 】

次に、類似度計算部 1 3 は、登録話者別音節モデルデータベース 2 0 から、当該利用者の利用者 ID に関連づけて格納されている当該利用者の話者音節モデルを抽出し、抽出した話者音節モデルに対する前記特徴量の類似度を計算する (O p 5 5)。求められた類似度は、音節判定部 1 4 へ送られる。

【 0 1 1 3 】

次に、音節判定部 1 4 は、類似度計算部 1 3 で求められた類似度に基づき、1 回目の分割キーワードの各音節について、登録話者別音節モデルデータベース 2 0 の話者音節モデルの中で最も類似度の高いモデルのインデックスを選択し、その類似度と共に、類似度保

50

管部 15 に保管する (Op 56)。ここで、2 回目の分割キーワードの入力がまだ済んでいないので (Op 57 の結果が NO)、Op 58 へ進む。

【0114】

第三者へのキーワードの暴露をより確実に防止するために、1 回目の分割キーワードの入力が終了したときに、2 回目の分割キーワードの音声入力までに場所を移動するように指示するメッセージを利用者の携帯電話等のディスプレイへ表示したり、同様の内容の音声メッセージを当該携帯電話へ送ったりすることが好ましい。

【0115】

このため、Op 58 において、1 回目の分割キーワードの音声入力後、適当な時間 (例えば 1 分) が経ったときに、位置情報取得部 31 が、利用者の携帯電話に組み込まれた GPS により、利用者の現在位置を表す情報を取得し、取得した情報を位置確認部 32 へ渡す。

10

【0116】

位置確認部 32 は、Op 53 で取得された 1 回目の分割キーワード入力時の位置情報と、Op 58 で取得された位置情報とを比較することにより、1 回目の分割キーワードを入力した後に、利用者が所定の距離 (例えば 50 m) 以上離れた場所へ移動しているかを確認する (Op 59)。なお、前記所定の距離は 50 m に限定されず、任意である。

【0117】

Op 59 の確認結果が NO であれば、音声認証システムは、利用者の携帯電話へ、場所を移動するように指示する音声メッセージを送る。あるいは、場所を移動するよう指示する警告メッセージを表示する (Op 60)。

20

【0118】

一方、Op 59 の確認結果が YES であれば、Op 61 以降の処理を続行する。Op 61 ~ Op 67 の処理は、第 1 の実施形態において説明した Op 17 ~ Op 23 と同様であるため、説明は省略する。

【0119】

以上のように、本実施形態にかかる音声認証システムは、GPS 等を利用して利用者の現在位置を確認し、分割キーワードの前回の入力時から利用者が所定の距離以上移動していない場合は、場所を移動するよう指示する。これにより、第三者へキーワードが暴露されることをより確実に防止できる。

30

【0120】

なお、本実施形態では、第 1 の実施形態にかかる音声認証システムに、位置情報取得部 31 および位置確認部 32 をさらに備えた構成を例示したが、第 2 の実施形態にかかる音声認証システムに位置情報取得部 31 および位置確認部 32 をさらに備えた構成も、本発明の一実施態様である。

【0121】

(第 4 の実施形態)

本発明にかかる音声認証システムのさらに他の実施形態について、以下に説明する。なお、第 1 ~ 第 3 の実施形態で説明した構成と同様の機能を有する構成については、それらの実施形態で用いた参照記号と同じ記号を付記し、その詳細な説明は省略する。

40

【0122】

本実施形態にかかる音声認証システムは、キーワードの分割方法等をシステムが利用者へ指示し、利用者が指示どおりに音声入力をしていない場合は、認証を棄却する構成である。

【0123】

このため、本実施形態にかかる音声認証システムは、図 8 に示すように、入力部 11、特徴量変換部 12、類似度計算部 13、音節判定部 14、類似度保管部 15、待機時間管理部 16、キーワード確認部 17、類似度統合部 18、認証判定部 19、登録話者別音節モデルデータベース 20、キーワード登録部 21、および、指示出力部 41 を備えている。すなわち、本実施形態にかかる音声認証システムは、第 1 の実施形態にかかる音声認証

50

システムに、指示出力部 4 1 をさらに備えた構成である。

【 0 1 2 4 】

指示出力部 4 1 は、予め定められた複数種類のパターンからランダムに選択したパターンに従って利用者にキーワードを分割入力させるよう指示を出す。

【 0 1 2 5 】

そのような指示としては、例えば、「 1 回目の分割キーワードとして最初の 3 音節、 2 回目の分割キーワードとして残りの全音節を入力して下さい。」、「奇数音節を 1 回目の分割キーワードとし、偶数音節を 2 回目の分割キーワードとして下さい。」、「キーワードを逆方向から 1 音節ずつ入力して下さい。」といった指示が考えられる。なお、これらの指示はあくまで一例であり、指示内容は任意である。

10

【 0 1 2 6 】

ここで、本実施形態にかかる音声認証システムの動作について、図 9 を参照しな説明する。なお、利用者によるキーワードの登録手順は第 1 の実施形態と同様であるため、説明を省略する。

【 0 1 2 7 】

最初に、入力部 1 1 は、利用者が携帯電話のキーパッドを用いて入力した利用者 ID 等を受け付ける (Op 7 1)。なお、携帯電話のように利用者が 1 名に限定される場合には利用者 ID の入力手順は不要となる。次に、指示出力部 4 1 が、予め定められた複数種類のパターンからランダムに選択したパターンに従って、利用者の携帯電話に対して、 1 回目の分割キーワードを入力させるための指示を、表示または音声メッセージとして出力する (Op 7 2)。指示出力部 4 1 による指示内容は、後の Op 7 6 における音節判定部 1 4 の処理のために、音節判定部 1 4 へも送られる。

20

【 0 1 2 8 】

例えばここで、「奇数音節を 1 回目の分割キーワードとして下さい。」という指示がされた場合、キーワードが「カラオケ」であったとすると、利用者は、 1 回目の分割キーワードとして「カ」「オ」、 2 回目の分割キーワードとして「ラ」「ケ」を入力しなければならない。

【 0 1 2 9 】

利用者が、この指示に従って、携帯電話のマイクロフォンに向けて 1 回目の分割キーワードを発声すると、入力部 1 1 はこの音声を受け取る (Op 7 3)。

30

【 0 1 3 0 】

入力された音節 S 1 , S 2 のそれぞれは、特徴量変換部 1 2 において、特徴量に変換される (Op 7 4)。特徴量変換部 1 2 で得られた特徴量は類似度計算部 1 3 へ送られる。

【 0 1 3 1 】

次に、類似度計算部 1 3 は、登録話者別音節モデルデータベース 2 0 から、当該利用者の利用者 ID に関連づけて格納されている当該利用者の話者音節モデルを抽出し、抽出した話者音節モデルに対する前記特徴量の類似度を計算する (Op 7 5)。求められた類似度は、音節判定部 1 4 へ送られる。

【 0 1 3 2 】

次に、音節判定部 1 4 は、類似度計算部 1 3 で求められた類似度に基づき、 1 回目の分割キーワードの各音節について、登録話者別音節モデルデータベース 2 0 に格納されている話者音節モデルの中で最も類似度の高いモデルのインデックスを選択し、その類似度と共に類似度保管部 1 5 に保管する (Op 7 6)。

40

【 0 1 3 3 】

また、Op 7 6 において、音節判定部 1 4 は、 1 回目の分割キーワードの各音節が、指示出力部 4 1 による指示どおりに発声されたかを判定する。すなわち、上記の例では、 1 回目の分割キーワードとして奇数音節、つまりキーワードの第 1 音節 (インデックス I 1) である「カ」と第 3 音節 (インデックス I 3) である「オ」を入力する旨の指示が出されている。従って、 1 回目の分割キーワードの 1 つ目の音節は、インデックス I 1 の話者音節モデルとの類似度が最も大きくなければならず、また、 2 つ目の音節は、インデック

50

ス I 3 の話者音節モデルとの類似度が最も大きくなければならない。そうでない場合は、音節判定部 1 4 は、その判定結果を認証判定部 1 9 へ通知し、認証判定部 1 9 が認証を棄却する。

【 0 1 3 4 】

なお、図 9 において、Op 7 7 以降は、第 1 の実施形態において説明した Op 1 6 ~ Op 2 3 と同様であるため、その重複した説明は省略する。

【 0 1 3 5 】

本実施形態にかかる音声認証システムでは、以上のように、指示出力部 4 1 が、キーワードの分割方法が認証毎にランダムになるように、分割方法を利用者へ指示することにより、キーワードの暴露をより確実に防止することができる。

10

【 0 1 3 6 】

なお、上記の説明では、指示出力部 4 1 が、キーワードの分割方法を一つ選択して利用者へ指示するものとしたが、指示出力部 4 1 が複数パターンの分割方法を利用者へ提示し、利用者がその中からいずれかのパターンを選択するようにしても良い。

【 0 1 3 7 】

また、本実施形態では、第 1 の実施形態にかかる音声認証システムに、指示出力部 4 1 をさらに備えた構成を例示したが、第 2 または第 3 の実施形態にかかる音声認証システムに指示出力部 4 1 をさらに備えた構成も、本発明の一実施態様である。

【 0 1 3 8 】

例えば、第 2 の実施形態にかかる音声認証システムに指示出力部 4 1 を加えた構成の場合であって、キーワードが数字列の場合は、例えば、「1 回目の分割キーワードとして 2 桁および 4 桁目の数字を入力し、2 回目の分割キーワードとして 1 桁および 3 桁目の数字を入力して下さい。」「暗証番号を逆方向から順番に入力して下さい。」といった指示が可能である。さらに、キーワードが数字とアルファベットの混合の場合は、「1 回目の分割キーワードとして奇数番目の数字またはアルファベット、2 回目に偶数番目の数字またはアルファベットを入力して下さい。」といった指示が可能である。

20

【 0 1 3 9 】

以上、各実施形態において説明したように、本発明によれば、キーワードを分割して入力することにより、キーワード全体を他人に聞かれることがない。従って、他人がキーワード内容を知ることが難しくなり、またキーワード全体を続けて録音することができないので、録音詐欺を防ぐことができる。

30

【 0 1 4 0 】

また、テキスト依存型の音声認証は、キーワード内容が暴露されることが多く、従来は、高い性能維持のためにキーワードを頻繁に変えなければならず、運用のコストが大きいという問題があった。しかし、本発明によれば、キーワード内容が暴露されにくくなり、キーワードを頻繁に変更する必要がなく、キーワード変更のためのコストを削減できるという効果がある。

【 0 1 4 1 】

さらに、例えば人気のない場所で 1 回目の分割キーワードを音声入力しておき、レジカウンタ前等の公衆の面前では 2 回目の分割キーワードのみを音声入力するようにすれば、公衆の面前での認証時間を短くすることができるので、スピーディーに決済を行うことができる、という効果もある。

40

【 0 1 4 2 】

なお、上記の各実施形態にかかる音声認証システムにおいて、例えば周囲に他人が居ない場合等にキーワードを分割せずに音声認証を行うことが可能となるように、利用者がキーワードを分割入力するか否かを選択できるようにしても良い。

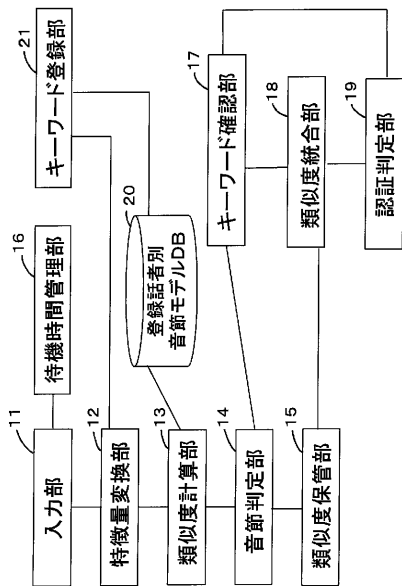
【 産業上の利用可能性 】

【 0 1 4 3 】

本発明は、キーワードの漏洩や録音詐欺を防ぐことにより、キーワードの秘密性を利用して高い認証精度を維持し得るテキスト依存型の音声認証システムとして有用である。

50

【図 1】



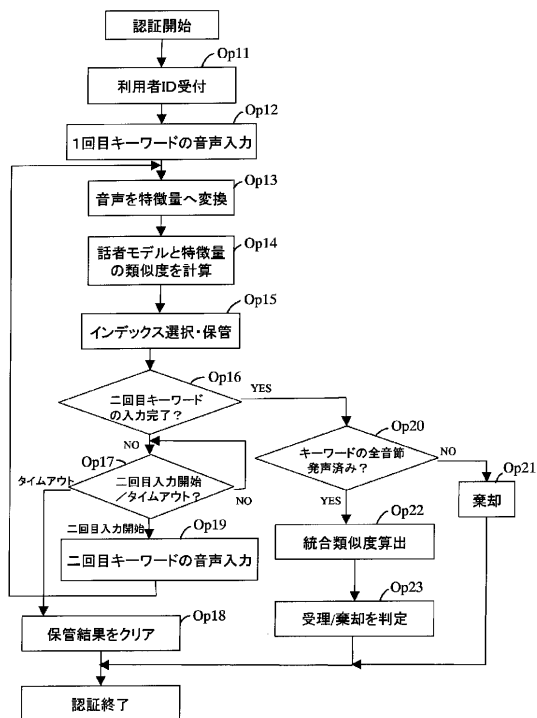
【図 2】

	I1 (「カ」)	I2 (「ラ」)	I3 (「オ」)	I4 (「ケ」)
S1 (「カ」)	C _{S1,I1}	C _{S1,I2}	C _{S1,I3}	C _{S1,I4}
S2 (「ラ」)	C _{S2,I1}	C _{S2,I2}	C _{S2,I3}	C _{S2,I4}
S3 (「オ」)	C _{S3,I1}	C _{S3,I2}	C _{S3,I3}	C _{S3,I4}
S4 (「ケ」)	C _{S4,I1}	C _{S4,I2}	C _{S4,I3}	C _{S4,I4}

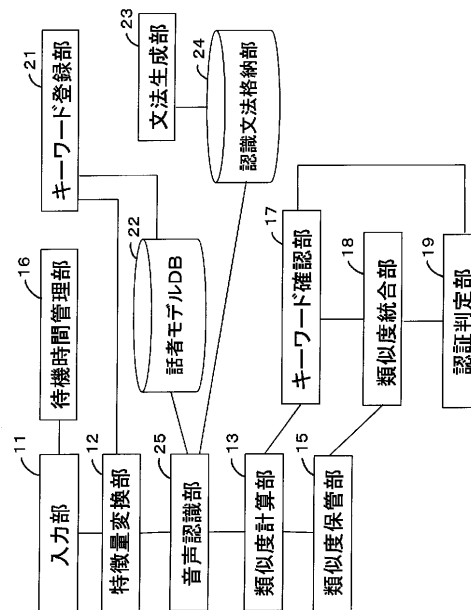
類似度保管部
へ格納

→ I1 → I2 → I3 → I4

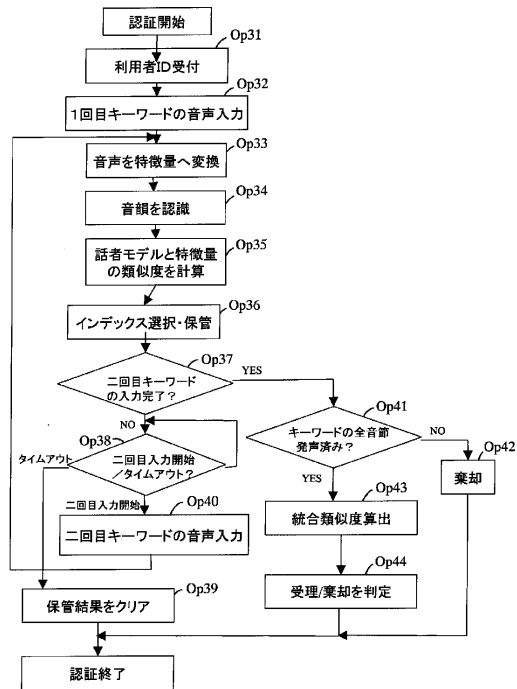
【図 3】



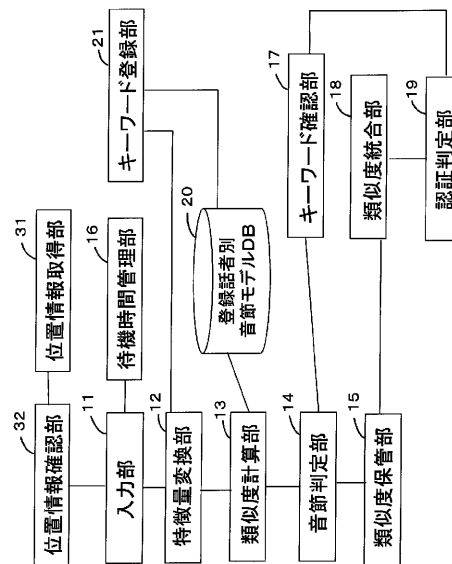
【図 4】



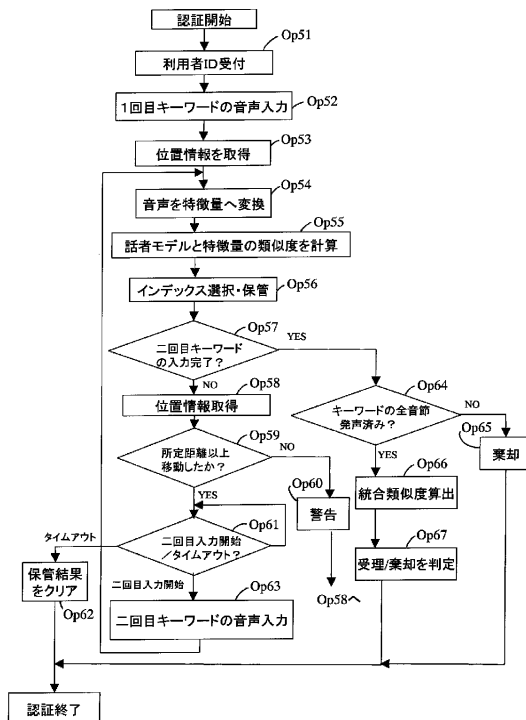
【図5】



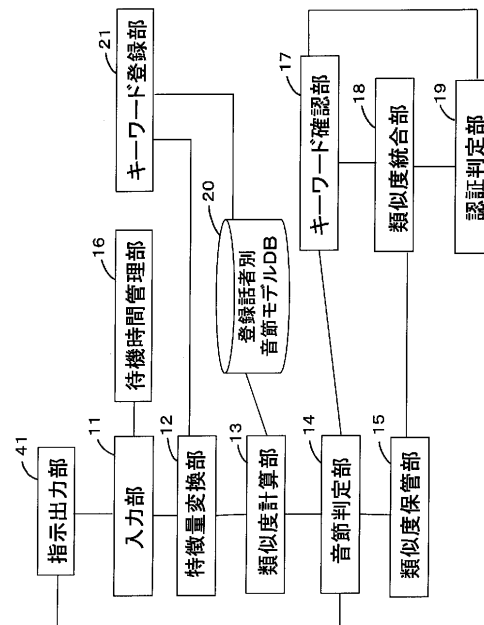
【図6】



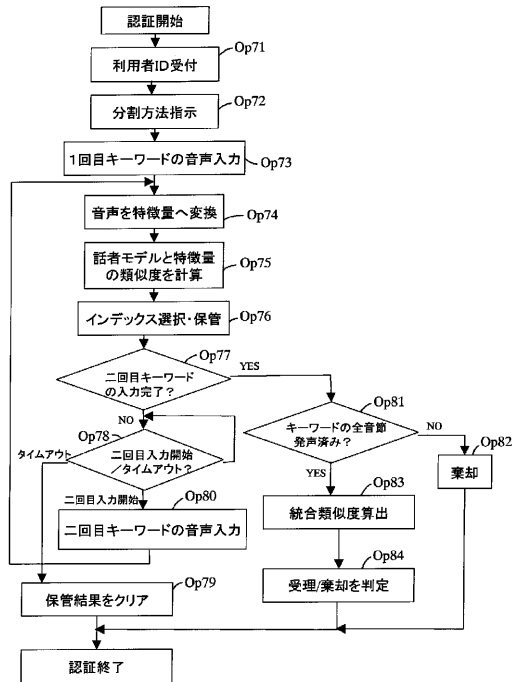
【図7】



【図8】



【図 9】



フロントページの続き

- (56)参考文献 特開 2 0 0 3 - 4 4 4 4 5 (J P , A)
特開 2 0 0 4 - 2 7 9 7 7 0 (J P , A)
特開平 1 0 - 5 6 4 4 9 (J P , A)
特開平 1 0 - 1 7 3 6 4 4 (J P , A)
特開 2 0 0 0 - 9 9 0 9 0 (J P , A)
特開 2 0 0 4 - 1 1 8 4 5 6 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)

G10L 15/00-17/00

H04L 9/32