(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2006 年3 月16 日 (16.03.2006)

PCT

(10) 国際公開番号
WO 2006/027844 A1

(51) 国際特許分類[7]: G01L 17/00, E05B 49/00
(21) 国際出願番号: PCT/JP2004/013197
(22) 国際出願日: 2004 年9 月10 日 (10.09.2004)
(25) 国際出願の言語: 日本語
(26) 国際公開の言語: 日本語

(71) 出願人 (米国を除く全ての指定国について): 三菱電機株式会社 (MITSUBISHI DENKI KABUSHIKI KAISHA) [JP/JP]; 〒1008310 東京都千代田区丸の内二丁目２番３号 Tokyo (JP).
(72) 発明者; および
(75) 発明者/出願人 (米国についてのみ): 石井 純 (ISHII, Jun) [JP/JP]; 〒1008310 東京都千代田区丸の内二丁目２番３号 三菱電機株式会社内 Tokyo (JP).
(74) 代理人: 高橋 省吾, 外(TAKAHASHI, Shogo et al.); 〒1008310 東京都千代田区丸の内二丁目２番３号 三菱電機株式会社 知的財産センター内 Tokyo (JP).

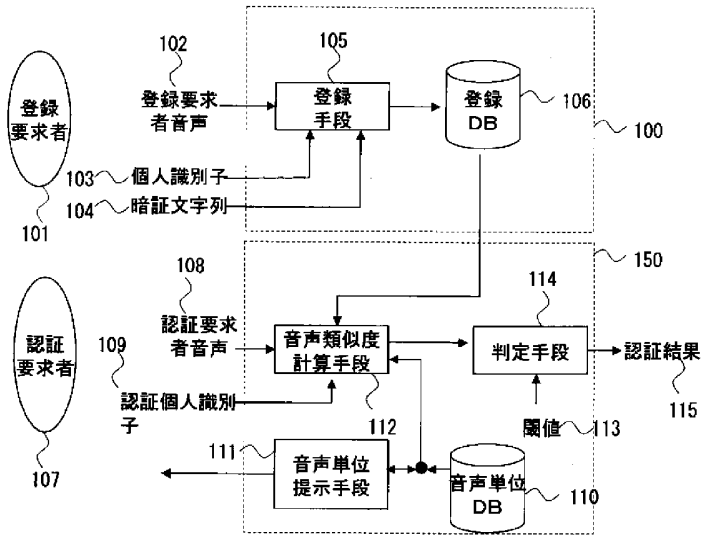
(81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[続葉有]

(54) Title: SPEAKER COLLATOR

(54) 発明の名称: 話者照合装置



101... REGISTRATION REQUESTER
107... AUTHENTICATION REQUESTER
102... REGISTRATION REQUESTER VOICE
103... PERSONAL IDENTIFIER
104... PASSWORD CHARACTER STRING
105... REGISTERING MEANS
106... REGISTRATION DB
108... AUTHENTICATION REQUESTER VOICE
109... AUTHENTICATION PERSONAL IDENTIFIER
112... VOICE SIMILARITY CALCULATING MEANS
111... VOICE UNIT PRESENTING MEANS
114... JUDGING MEANS
113... THRESHOLD
110... VOICE UNIT DB
115... AUTHENTICATION RESULT

(57) Abstract: Since conventional speaker collators are so constituted that the user utters voices corresponding to words and registers them, the voiced contents are limited within the word s registered by the user, and spoofing is easily performed when the voices of the words uttered by the user are recorded by another person. This invention has been made to solve the above problem to make spoofing difficult even if the uttered content is recorded by another person.

(57) 要約: 課題は、従来の話者照合装置は、予め、利用者が複数の単語に対応する音声をそれぞれ発生して登録するように構成されているので、発声する内容は、利用者が登録した単語に限定されるため、利用者が当該単語を発声した音声を他人に録音された場合には、簡単に詐称されてしまう問題があった。そこで、この発明は、上記のような課題を解決するためになされたものであり、発声した内容を他人に録音されても、詐称されにくくしたものである。

WO 2006/027844 A1

添付公開書類:
— 国際調査報告書

２文字コード及び他の略語については、定期発行される各*PCT*ガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

明 細 書

話者照合装置

技術分野

[0001] 本発明は、利用者の音声が、申告した本人の音声かどうかを判定することにより、利用者が正当な者か否かを判断する話者照合装置に関するものである。

背景技術

[0002] 従来の話者照合装置においては、予め、利用者が複数の単語に対応する音声をそれぞれ発声して登録するとともに、登録した複数の単語に対してインデックスをつける。照合時においては、利用者がインデックス及び個人識別子を指定し、このインデックスに対応した単語を発声する。この発声した音声とこれに対応し予め登録された音声とを照合することにより、利用者が正当な利用者であるか否かを判断する。

[0003] 特許文献1:特開2000−181490

特許文献2:特開2002−269047

特許文献3:特開2000−99090

特許文献4:特開2000−338987

特許文献5:特開平11082492

特許文献6:特開平10−214096

特許文献7:特開2001−331196

非特許文献1:「音声情報処理」古井 貞熙 森北出版株式会社

非特許文献2:中村哲、山本英里、永井論、鹿野清宏「HMMを用いた音声と唇画像の統合による音声認識と唇画像生成」音声言語情報処理研究会P15−P17

発明の開示

発明が解決しようとする課題

[0004] 従来の話者照合装置は、以上のように構成されているので、発声する内容は、利用者が登録した単語に限定されるため、利用者が当該単語を発声した音声を他人に録音された場合には、簡単に詐称されてしまう問題があった。

[0005] この発明は、上記のような課題を解決するためになされたものであり、発声した内容を

他人に録音されても、詐称されにくくしたものである。

課題を解決するための手段

[0006] この発明に係る話者照合装置は、認証要求者に発声させるべき音声単位を認証要求者に提示する際に前回提示した音声単位を他の音声単位に切り替えて提示する音声単位提示手段と、この提示する音声単位と個人識別子とに対応づけて準備される音声標準パターンと、この準備された音声標準パターンと上記認証要求者が発声した音声とを比較して音声の類似度を計算する音声類似度計算手段と、この音声類似度計算手段の計算結果に基づいて上記認証要求者が正当な利用者か否かを判定する判定手段と、を備えるものである。

発明の効果

[0007] この発明に係る話者照合装置は、認証要求者に発声させるべき音声単位を認証要求者に提示する際に前回提示した音声単位を他の音声単位に切り替えて提示するようにしたので、認証要求者が前回発声した内容を他人に録音されても、その録音内容を今回は利用できなくなるため、他人による詐称を防止することができる。

発明を実施するための最良の形態

[0008] 実施の形態1.

図1は、本発明の実施の形態1による話者照合装置の構成図である。図2、図5、図6は、本発明の実施の形態1による音声単位提示手段による提示内容を示す図である。図3は、本発明の実施の形態1による登録データベースにおける音声標準パターン118の格納態様の概念図である。図4は、本発明の実施の形態1による話者照合装置の処理内容を示すフローチャートである。

[0009] 次に、図1乃至3を用いて本実施の形態における話者照合装置の構成を説明する。

[0010] 図1において、100は、利用者登録手段である。利用者登録手段100は、登録時において登録要求者101が、予め登録要求者音声102、個人識別子103及び暗証文字列104を登録することにより、個人識別子103に対応させて音声標準パターン118（図3参照）及び暗証文字列104を登録データベース106に格納する手段である。

[0011] 101は、登録要求者である。登録要求者101は、例えば、本実施の形態にかかる話者照合装置がビルに入る際の認証に使用される場合には、このビルの居住人等、予

めビルに入ることが許可されている者が該当する。

[0012] 102は、登録要求者の音声である。この登録要求者音声102により後述する音声標準パターン118を生成する。音声の録音の仕方は、話者照合装置が指定した文章を登録要求者101が読み上げることで得る。なお、話者照合装置が指定する文章は多くの音節の種類が含まれるようにすることで、質の良い音声標準パターン118を生成することができる。

[0013] ここで「音声標準パターン118」とは、登録要求者音声102の特徴を効率よく表現するものである。音声標準パターンは、例えば、登録要求者音声102の音声波形をAD変換してデジタル信号を生成し、この信号に対して特徴量分析を行って得られた音声特徴量によって学習した音節単位のサブワードHMM(Hidden Markov Model)である。HMMに関しては非特許文献1に詳細が記されている。音声特徴量とは、音声信号を効率よく表現するものであり、例えばケプストラムを用いる。

[0014] 103は、個人識別子である。個人識別子103は、複数の登録要求者101を識別するために登録要求者101に割り当てられた符号であり、例えば英数字の組合せを用いる。個人識別子がアルファベットで構成されている場合、最初に登録した「鈴木太郎」には「AAA」、二番目に登録した「鈴木次郎」には「AAB」、三番目に登録した「鈴木三郎」には「AAC」という様に割り当てる。

[0015] 104は、暗証文字列である。暗証文字列104は、正式な利用者であることを証明するために秘密にし、予め登録しておく符号であり、例えば英数字の組合せを用いる。暗証文字列104が4桁の数字列と指定されている場合、「9768」「4361」などが設定可能である。

[0016] 105は、登録手段である。登録手段105へ登録要求者音声102、個人識別子103、暗証文字列104を入力することにより、後述する登録データベース106へ個人識別子103に対応した音声標準パターン118及び暗証文字列104を登録する。登録手段105は、マイクやキーボード等を備えており、各登録要求者101はマイクを使用し登録要求者音声102を登録し、キーボードを利用し暗証文字列104を登録する。

[0017] 106は、登録データベースである。登録データベース106は、登録手段105により生成した音声標準パターン118、個人識別子103、暗証文字列104を格納する。音声

標準パターン118は、個人識別子103と標準パターン音声単位に対応するように格納する。標準パターン音声単位が音節から構成されている場合には、個人識別子103及び各音節117に対応するように格納する。例えば、図3に示すように個人識別子「AAB」で登録した話者の音節単位の音声標準パターン「あ(/a/)」である音声標準パターン118のように格納する。このように格納することにより認証個人識別子109及び音声単位提示手段111で提示された音声単位130(図2参照)に対応した音声標準パターン118を選択することができる。

[0018] 150は、利用者認識手段である。利用者認識手段150は、認証時において認証要求者107が正当な利用者か否かを判定する手段であり、認証要求者音声108と選択された音声標準パターン118との類似性を計算するものである。

[0019] 107は、認証要求者である。認証要求者107は、本実施の形態にかかる話者照合装置がビルに入る際の認証に使用される場合には、このビルに入ろうとする者が該当する。例えば、登録要求者101や詐称してビルに侵入しようとする者が該当する。

[0020] 108は、認証要求者音声である。認証要求者音声108は、後述する音声単位提示手段111で提示された音声単位130(図2参照)に対応して認証要求者107が発声する音声である。

[0021] 109は、認証個人識別子である。認証個人識別子109は、認証要求者107が認証時に申告する話者を識別するための符号であり、登録している個人識別子103のいずれかと一致していなければならない。

[0022] 110は、認証要求者107に提示する音声単位130を格納する音声単位データベースである。

[0023] 111は、音声単位提示手段である。音声単位提示手段111は、図2に示すように、暗証文字列構成文字120からなる暗証文字列構成文字群121、音声単位130、及び両者の対応関係を提示する。ここで「暗証文字列構成文字120」とは、暗証文字列104を構成する文字である。例えば暗証文字列104が10進数の数字列である場合に、暗証文字列構成文字120は「0」、「1」、「2」…「9」が該当する。また、「音声単位130」とは、認証要求者107が発声させるべき文字列である。文字列が単語に該当する場合には「はちのへ(/hatinohe/)」、「けせんぬま(/keseNnuma/)」、「ゆくはし(

/yukuhasi/)」、「さっぽろ(/saQporo/)」等がある。図2は、単語からなる音声単位130と暗証文字列構成文字群121とを対応させた組合せを複数提示したものである。「はちのへ(/hatinohe/)」という音声単位130に対して「0」、「2」、「6」が対応し、「けせんぬま(/keseNnuma/)」には「1」、「9」が対応し、「ゆくはし(/yukuhasi/)」には「3」、「4」、「7」が対応し、「さっぽろ(/saQporo/)」には「5」、「8」が対応するように割り当てる。この対応表に従って、認証要求者107は自己の暗証文字列104の数字列に対応した音声単位130を発声する。認証要求者107の暗証文字列104が「5218」であるならば、最初の数字5に対応した音声単位130「さっぽろ(/saQporo/)」を発声し、次に二番目の数字2に対応した「はちのへ(/hatinohe/)」を発声する。更に三番目の数字1に対応した「けせんぬま(/keseNnuma/)」、四番目の数字8に対応した「さっぽろ(/saQporo/)」を連続して発声させる。

[0024] 音声単位提示手段111は、認証要求者107に提示する際に前回提示した音声単位130を他の音声単位130に切り替えて提示する。即ち、音声単位提示手段111は、利用する毎に音声単位130が切り替わる場合、2回利用する毎に切り替わる場合、3回利用する毎に切り替わる場合、ランダムに切り替わる場合、一定の時間毎に切り替わる場合、同一の認証要求者107が利用する毎に切り替わる場合がある。

[0025] 112は、音声類似度計算手段である。音声類似度計算手段112は、認証個人識別子109の入力、及び、音声単位提示手段111の提示に基づいて登録データベース106から音声標準パターン118を選択する。例えば、認証要求者107が認証個人識別子109として「AAB」を入力し、音声単位提示手段111では暗証文字列104により「さっぽろ、はちのへ、けせんぬま、さっぽろ」を連続して発声することを提示された場合には、音声類似度計算手段112は、登録データベース106から、認証個人識別子109「AAB」に対応した、「さ(/sa/)」、「っ(/Q/)」、「ぽ(/po/)」、「ろ(/ro/)」、「は(/ha/)」…の音声標準パターン118が選択される。この選択された音声標準パターン118と認証要求者107が発声した認証要求者音声108の各音節とを比較し音声類似度を計算し類似度を出力する。音声類似度の計算は、音節ごとの音響的特徴を比較して、文章全体で判断する場合がある。類似度に関しては例えば「音声情報処理」古井　貞熙　1998年6月　森北出版株式会社(以下参考文献1とする)の5章に記さ

れている方法によって計算する。

[0026] 113は、閾値である。閾値113は、予め定められた基準値であって、認証要求者音声108が正当な利用者による音声か否かの判定の基準となる。音声類似度計算手段112での類似度がこの閾値113より大きければ、認証要求者107は正当な利用者と判定される。

[0027] 114は、認証要求者107が、正当な利用者か否かを判定する判定手段である。判定手段114は、音声類似度計算手段112での結果を基に、認証要求者音声108が正当な利用者による音声であるか否かを判定する。類似度が閾値113以上の場合に正当な利用者と判断し、逆に閾値113より小さい場合には、詐称を目的とする利用者と判定する。

[0028] 115は、認証結果である。認証結果115は、判定手段114からの出力であり、認証要求者107が正当な利用者と判断された場合は「受理」、詐称を目的とする利用者と判断された場合は「棄却」となる。例えば、本実施の形態にかかる話者照合装置がビルに入る際の認証に使用された場合、「受理」された場合にはドアのロックが解除され、「棄却」された場合にはロックがされたままとなる。

[0029] 次に、以上の構成からなる話者照合装置の動作について説明をする。図4はこの発明の実施の形態1による話者照合装置の処理内容を示すフローチャートである。以下、図4にしたがって動作を説明する。

[0030] 図4のステップ11は、登録要求者101の情報を登録するステップである。すなわち登録要求者101は登録手段105へ自己の情報、すなわち登録要求者音声102、個人識別子103、暗証文字列104を入力する。登録手段105は、登録要求者音声102に基づき音声標準パターン118を生成するとともに、この音声標準パターン及び暗証文字列を登録データベース106へ格納する。

[0031] ステップ12は、認証要求者107に発声させる音声単位と認証文字(群)との対応を認証要求者107に提示するステップである。すなわち、認証要求者107に認証個人識別子109を話者照合装置に入力させる。次に、音声単位提示手段111は、認証要求者107に発声させる音声単位130を提示する。音声単位提示手段111が、図2に示すように暗証文字列構成文字群121と音声単位130との両者の対応を提示する。

音声単位提示手段111は、認証要求者107に提示する際に前回提示した音声単位130を他の音声単位130に切り替えて提示する。即ち、音声単位提示手段111は、利用する毎に音声単位130が切り替わる場合、2回利用する毎に切り替わる場合、3回利用する毎に切り替わる場合、ランダムに切り替わる場合、一定の時間毎に切り替わる場合、同一の認証要求者107が利用する毎に切り替わる場合がある。なお、当該照合装置の管理者は、新たな音声単位130を登録したい場合には、音声単位データベース110を更新すればよい。音声単位データベース110に音声単位130を1つ増やせば、当該音声単位130を各登録要求者101に対して提示することができるという効果を得ることができる。また、登録要求者101は一度登録要求者音声102を登録すれば、音声標準パターン118が自動的に生成されるため、新たな音声単位130が追加されても新たに登録要求者音声102を録音しなくても良いという効果を得ることができる。

[0032] ステップ13は、認証の際に、認証要求者107が発声した認証要求者音声108とこれに対応する音声標準パターン118とを比較するステップである。

[0033] ステップ14は、認証の際に、判定手段114が、音声類似度計算手段112の出力である音声類似度を、予め定められた閾値113と比較して、音声類似度が閾値113以上であれば正当な利用者であるとして「受理」、一方、音声類似度が閾値113より小さければ詐称を目的とする利用者として「棄却」を認証結果115として出力するステップである。

[0034] 以上のように、この実施の形態1によれば、前回提示した内容を他の内容に切り替えて提示するので、録音した音声によって他人に詐称されたり、暗証文字列104を他人に知られることで詐称されたりする可能性が低くなり安全性の高い話者照合を実現できる。

[0035] なお、登録手段105、音声単位提示手段111、音声類似度計算手段112、判定手段114をハードウェアで構成してもよいが、各手段の処理内容を記述した話者照合プログラムを作成し、コンピュータが当該話者照合プログラムを実行するようにしてもよい。

[0036] また、1つの閾値113によって受理か棄却を決定する場合を説明したが、複数の閾

値を設定してもよい。例えば閾値をA、Bの2種類用意して、音声類似度が閾値Aより大きい値ならば正当な利用者であると判定し、音声類似度がAとBの間なら明瞭な判定不可能であると判定し、Bより音声類似度が小さいならば詐称を目的とする者と判定するようにする。このように設定することにより、例えば、当該話者照合装置がビルに入る際の認証に使用する場合には、正当な利用者と判断すればドアのロックを解除し、判定不可能であると判断すれば再度認証要求者107に音声単位提示手段111を提示し、詐称者と判断した場合には、ドアのロックは解除しない。

[0037] また、図5に示すように、ひらがな一文字からなる音声単位130と1つの暗証文字列構成文字120とを1対1に対応させた組合せを複数提示する場合がある。例えば、認証要求者107の暗証文字列104が「5218」であるなら認証要求者107は、「せ」、「ゆ」、「け」、「の」と発声することになる。

[0038] また、図6に示すように、ひらがな一文字からなる音声単位130と暗証文字列構成文字群121とを対応させた組合せを複数提示する場合がある。例えば、「は」という音声単位130に対して、「0」、「2」、「6」が対応し、「け」には、「1」、「9」が対応し、「ゆ」には「3」、「4」、「7」が対応し、「さ」には「5」、「8」が対応するように割り当てられる。この対応表にしたがって認証要求者107は自己の暗証文字列104の数字列に対応した音声単位130を発声する。認証要求者107の暗証文字列104が「5218」であるならば、最初の数字に対応した音声単位130「さ」を発声し、次に2に対応した「は」を発声する。さらに3番目の数字である1に対応した「け」、4番目の数字に8対応した「さ」を連続して発声することになる。このように音声単位130に対して複数の数字が割り当てられているので、発声した内容を他人に知られたとしても暗証文字列104の数字列が一意に知られてしまうことはない。

実施の形態2.

[0039] 図7は、本発明の実施の形態2による話者照合装置の構成図である。図8は、本発明の実施の形態2による話者照合装置の処理内容を示すフローチャートである。図7を用いて本実施の形態における話者照合装置の構成を説明する。なお、実施の形態1と共通する部分については説明を省略する。

[0040] 図7において、201は、登録要求者属性である。登録要求者属性201は、登録要求

者101の属性に関する情報であり、性別、年齢、干支、血液型、出身地等の情報である。例えば「男性、22歳、ねずみ年、A型、東京都出身」等の情報がある。

[0041] 202は、登録手段である。登録手段202へ登録要求者音声102、個人識別子103、登録要求者属性201を入力することにより、登録データベース203へ音声標準パターン118を格納する。登録手段202はマイクやキーボード等を備えており、各登録要求者101は自己の個人識別子103に対応させるように、マイクを使用して登録要求者音声102を登録し、キーボードを使用して登録要求者属性201を登録する。

[0042] 203は、登録データベースである。登録データベース203は、登録手段202により生成した音声標準パターン118及び登録要求者属性201を格納する。

[0043] 204は、音声単位提示手段である。音声単位提示手段204は、認証要求者107に対して、属性に関する質問を提示する。例えば「あなたの年はおいくつですか？」、「あなたの出身はどちらですか？」、「あなたの血液型はなんですか？」等がある。この質問は、認証要求者107が答えるべき音声単位130を前回提示した音声単位130から他の音声単位130に切り替えて提示する。例えば、ある認証要求者107に「あなたの年はおいくつですか？」を提示し、その次にこの話者照合装置を利用する認証要求者107に「あなたの出身はどちらですか？」を提示する場合がある。

[0044] 205は、応答時間計算手段である。応答時間計算手段205は、認証要求者107が認証要求者音声108を発声するのにかかった時間を計る。例えば、音声単位提示手段204により質問がなされてから認証要求者107が発声開始までの応答するのにかかった時間を計る。

[0045] 206は、判定用スコア計算手段である。判定用スコア計算手段206は、応答時間計算手段205の出力である応答時間と音声類似度計算手段112の出力である音声類似度とから判定用スコアSを計算する。判定用スコアSは、音声類似度をL（大きければ類似度が高い）、応答時間をTrとすれば例えば数式1で求める。

[0046] $$S = L - \alpha \cdot Tr \cdots\cdots \text{数式1}$$

数式1において、αは重み係数である。数式1によれば、応答時間Trが長ければ判定用スコアが低くなる。

[0047] 次に、以上の構成からなる話者照合装置の動作を図8を用いて説明をする。図8に

おいて、ステップ21は、登録要求者101は登録手段202へ自己の登録要求者音声102、個人識別子103、登録要求者属性201を入力するステップである。登録要求者音声102及び登録要求者属性201は個人識別子103に対応するように入力する。次に登録手段202は、登録要求者音声102に基づき音声標準パターン118を生成し登録データベース203へ格納する。

[0048] ステップ22は、話者照合装置が認証要求者107に、属性に関する質問をするステップである。例えば「あなたのお年はおいくつですか？」と質問する。これらの質問は、音声単位提示手段204により前回提示した内容から切り替わって提示されるために、他人が認証要求者107の発声した音声を録音して、ビルに入ろうとしても同じ音声単位130が提示される可能性が少なくなるために詐称される可能性は低くなる。

[0049] ステップ23は、認証要求者107が認証要求者音声108を発声する時間を計るステップである。応答時間計算手段205は、認証要求者107が認証要求者音声108を発声する時間を計る。正当な利用者であれば、自己の属性を把握しているため、一般的に応答時間が短くなる。また他人であれば、この質問に対する正しい答えを準備するのに時間がかかり、応答時間が長くなる。

[0050] ステップ24は、判定用スコア計算手段206により、応答時間計算手段205の応答時間及び音声類似度計算手段112の音声の類似度に基づいて、認証要求者107が正当な利用者か否かの判定用スコアを算出するステップである。判定手段114は、判定用スコア計算手段の出力を、予め定められた閾値113と比較して、出力が閾値113以上であれば本人であるとして「受理」、一方、音声類似度が閾値より小さければ他人であるとして「棄却」を認証結果115として出力する。

[0051] 以上のように、この実施の形態2によれば、前回提示した音声単位130を他の音声単位130に切り替えて提示されるので、録音した音声によって他人に詐称されにくい。また、認証要求者107は質問に答えるという1つの行為だけで、音声の類似度及び応答時間の2つの観点から、認証要求者107が正当な利用者か否かが判断されるため、認証要求者107は面倒な手続きをすることなく、精度の高い音声照合ができる。

[0052] 実施の形態3.

[0053] 図9は、本発明の実施の形態3による話者照合装置の構成図である。図10は、本発

明の実施の形態3による話者照合装置の処理内容を示すフローチャートである。図9を用いて本実施の形態における話者照合装置の構成を説明する。なお、実施の形態1又は2と共通する部分については説明を省略する。

[0054] 図9において、301は、登録要求者唇画像である。登録要求者唇画像301は、予め、登録要求者101に、新聞や雑誌の記事を読ませ、この時の唇の形、動きを録画することにより作成する。この登録要求者唇画像301に基づいて後述する唇画像標準パターンを生成する。

[0055] ここで「唇画像パターン」は、唇の画像を音節単位のサブワードHMM(Hidden Markov Model)を利用して作成する。

[0056] また、「唇画像標準パターン」とは、上記パターンの画像特徴を効率よく表現したものをいう(以下、唇画像標準パターンは音節単位で構成されるとして説明する)。唇画像標準パターンは、例えば、中村哲、山本英里、永井論、鹿野清宏「HMMを用いた音声と唇画像の統合による音声認識と唇画像生成」音声言語情報処理研究会15−17、1997年2月(以下参考文献2とする)で報告されている方法で作成する。参考文献2によれば唇画像の標準パターン作成は、まず1フレーム(1フレームは33.3m秒)あたり横160×縦150＝19200画素の画像を256階調の濃淡画像に変換し、256×256のFFT(Fast Fourier Transform)によって周波数分析を行う。そして空間周波数領域おけるパワースペクトルを計算し、対数スケールのスムージングを行う。さらにフレーム間の差分をとることで動的な特徴を求める。唇画像標準パターンはパワースペクトル256分布、その差分に256分布の構造を持ったHMMであり、前記のように唇画像より求めたパワースペクトルと差分によって作成する。

[0057] 302は、登録手段である。登録手段302へ登録要求者音声102、個人識別子103、登録要求者属性201、登録要求者唇画像301を入力することにより、後述する登録データベース303へ個人識別子103に対応するように音声標準パターン118、唇画像標準パターン、及び登録要求者属性201を格納する。登録手段302はマイク及びカメラを備えており、各登録要求者101は自己の個人識別子103に対応させ、マイクを使用することにより登録要求者音声102を登録し、カメラを利用することにより登録要求者唇画像301を登録する。

[0058] 303は、登録データベースである。登録データベース303は、登録手段302により生成した音声標準パターン118、唇画像標準パターン、及び登録要求者属性201を格納する。

[0059] 304は、認証要求者唇画像である。認証要求者唇画像304は、音声単位提示手段204により提示された質問に対して答えた状態の唇画像を録画することにより得る。

[0060] 305は、音声及び唇画像類似度計算手段である。音声及び唇画像類似度計算手段305は、認証個人識別子109の入力、及び、音声単位提示手段204の音声単位130に基づいて登録データベース303から音声標準パターン118及び唇画像標準パターンを選択する。例えば、認証要求者107が認証個人識別子109として「AAB」を入力し、音声単位提示手段204「あなたの出身はどこですか？」という質問に対して、登録データベース303において認証個人識別子109の出身地が「札幌」に対応している場合には、音声及び唇画像類似度計算手段305は登録データベース303から「AAB」に対応した「さ(/sa/)」、「っ(/Q/)」、「ぽ(/po/)」、「ろ(/ro/)」の音声標準パターン118及び唇画像標準パターンを順次選択する。そしてこの選択された音声標準パターン118及び唇画像標準パターンに対応する認証要求者音声108及び認証要求者唇画像304とを比較する。

[0061] 次に、以上の構成からなる話者照合装置の動作について説明をする。図10はこの発明の実施の形態3による話者照合装置の処理内容を示すフローチャートである。以下、図10にしたがって動作を説明する。

[0062] 図10において、ステップ31は、登録要求者101の情報を登録するステップである。すなわち登録要求者101は登録手段302へ自己の登録要求者音声102、個人識別子103、登録要求者属性201、登録要求者唇画像301を入力する。登録手段302は、登録要求者音声102に基づき音声標準パターン118を生成し、登録要求者唇画像301に基づき唇画像標準パターンを生成し、両者を登録データベース303へ格納する。次に、格納の仕方は、音声標準パターン118及び唇画像標準パターンを個人識別子103及び音節117に対応するように格納する。このように格納することにより認証個人識別子109及び音声単位提示手段204に対応した音声標準パターン118及び唇画像標準パターンを選択することができる。

[0063] ステップ32は、認証要求者107が認証個人識別子109を話者照合装置に入力するステップである。次に、音声単位提示手段204は、属性に関する質問をする。この質問に対して認証要求者107は音声単位提示手段204で要求された音声単位130に対応する認証要求者音声108を発声する。

[0064] ステップ33は、認証要求者107が発声した認証要求者音声108とこれに対応する音声標準パターン118とを比較するステップである。また、認証要求者107が発声した時に撮影した認証要求者唇画像304とこれに対応する唇画像標準パターンとを比較する。認証要求者唇画像304と唇画像標準パターンとの比較は、例えば、認証要求者107が「AAB」の認証個人識別子109を入力し、音声単位提示手段204で「あなたの出身はどちらですか？」と質問された場合には、唇画像類似度計算手段は、登録データベース303から、認証個人識別子109「AAB」に対応した「さ(/sa/)」、「っ(/Q/)」、「ぽ(/po/)」、「ろ(/ro/)」にあたる唇画像標準パターンを選択する。この選択された各唇画像標準パターンとこれに対応した各認証要求者唇画像304とを比較し唇画像類似度を計算し類似度を出力する。

[0065] ここで認証要求者唇画像304は、認証要求者107が認証要求者音声108を発声した唇を撮影した画像であり、特徴量として、パワースペクトルおよびパワースペクトルの動的特徴を抽出する。この認証要求者唇画像304と登録データベース303から選択した唇画像標準パターンとの類似度計算は、文献1の5章で示される方法で尤度計算を行う。次に、音声による類似度Lと唇画像による類似度Mを用い、判定用の類似度を求める。判定用類似度は、例えば両者を重み付け加算した数式2に示すスコアとする。

[0066] [数2]

$$S = L + \beta M \quad \cdots\cdots \text{数式 } 2$$

数式2においてβは重み付け係数であり、高い照合精度が得られるように決定する。判定手段114は、音声と唇画像による類似度を音声及び唇画像類似度計算手段305の出力である類似度Sを入力し、予め定めた閾値113と比較して、類似度が閾値以上であれば本人であるとして「受理」、一方類似度が閾値より小さいのであるならば他人であるとして「棄却」を認証結果115として出力する。

[0067] ステップ34は、話者照合装置が、認証要求者107を正当な利用者か否かを判断するステップである。

[0068] 以上のように、この実施の形態3によれば、前回提示した音声単位130を他の音声単位130に切り替えて提示され、音声とともに唇画像も入力して類似度を計算するので、テープレコーダ等で録音した音声によって他人に詐称されにくい。また、認証要求者107は質問に答えるという1つの行為だけで、音声及び唇画像の2つの観点から、認証要求者107が正当な利用者か否かが判断することができる。

実施の形態4.

[0069] 図11は、本発明の実施の形態4による話者照合装置の構成図である。図12は、本発明の実施の形態4による話者照合装置の処理内容を示すフローチャートである。図11を用いて本実施の形態における話者照合装置の構成を説明する。なお、実施の形態1、2又は3と共通する部分については説明を省略する。

[0070] 図11において、401は、登録手段である。登録手段401へ個人識別子103、登録要求者属性201を入力することにより、後述する登録データベース402へ個人識別子103に対応させた登録要求者属性201を登録する。

[0071] 402は、登録データベースである。登録データベース402は、個人識別子に対応させて登録要求者属性201を格納する。

[0072] 403は、照合用音声標準パターン群である。照合用音声標準パターン群403とは、属性に依存した音声標準パターンの群である。例えば、属性が年齢と性別である場合には、10才代男性の音声標準パターンの群、20才代男性の音声標準パターンの群…というように分別して構成する。照合用音声標準パターン群403は、例えば、10才代男性の音声波形をAD変換したデジタル信号に対して特徴量分析を行って得られる音声特徴量を用いて作成する。404は、照合用音声標準パターン群集合である。照合用音声標準パターン群集合404は、照合用音声標準パターン群403の集合である。

[0073] 405は、照合用音声標準パターン選択手段である。照合用音声標準パターン選択手段405は、入力された認証個人識別子109により登録データベース402から登録要求者属性201のデータを取り出し、そのデータに基づいて照合用音声標準パター

ン群集合405から照合用音声標準パターン群を選択する。例えば、個人識別子103によって申告した認証要求者の属性が、20代男性であったならば、20代男性の照合用標準パターン群を選択する。

[0074] 次に、以上の構成からなる話者照合装置の動作について説明をする。図12はこの発明の実施の形態1による話者照合装置の処理内容を示すフローチャートである。以下、図12にしたがって動作を説明する。

[0075] 図12において、ステップ41は、登録要求者101の情報を登録するステップである。すなわち登録要求者101は登録手段401へ自己の個人識別子103、登録要求者属性201を入力する。登録要求者属性201は個人識別子103に対応するように登録する。

[0076] ステップ42は、認証要求者107は認証個人識別子109を話者照合装置に入力するステップである。次に、音声単位提示手段204は、属性に関するする質問を提示する。例えば、「あなたの出身はどこですか？」がある。この質問は、発声単位提示手段204により前回提示した内容を切り替えて提示する。

[0077] ステップ43は、認証要求者107が発声した認証要求者音声108と、これに対応する照合用音声標準パターン118とを比較するステップである。例えば、認証要求者107が「AAB」の認証個人識別子109を入力し、音声単位提示手段204では「出身地はどこですか？」と質問した場合には、照合用音声標準パターン選択手段405は、照合用音声標準パターン群集合405から、認証個人識別子109「AAB」の属性に対応した、照合用音声標準パターン群を選択し、更にその中から「さ(/sa/)」、「っ(/Q/)」、「ぽ(/po/)」、「ろ(/ro/)」の照合用音声標準パターン118を選択する。この選択された照合用音声標準パターン118と認証要求者音声108の各音節とを比較し音声類似度を計算し類似度を出力する。

[0078] ステップ44は、話者照合装置が、認証要求者107が正当な利用者か否かを判断するステップである。

[0079] 以上のように、この実施の形態4によれば、前回提示した音声単位130を他の音声単位130に切り替えて提示されるので、テープレコーダ等で録音した音声によって他人に詐称されにくい。また、属性を利用して話者照合を行うので、事前の音声登録が

不要となり、登録要求者101が音声を登録できない場合であっても話者照合を実現できる。

図面の簡単な説明

[0080] [図1]この発明の実施の形態1を示す話者照合装置の構成図である。

[図2]この発明の実施の形態1における音声単位提示手段の概念を示す一例である。

[図3]この発明の実施の形態1における登録データベースにおける音声標準パターンの格納態様の概念図である。

[図4]この発明の実施の形態1における話者照合装置の処理内容を示すフローチャートである。

[図5]この発明の実施の形態1における音声単位提示手段の概念を示す一例である。

[図6]この発明の実施の形態1における音声単位提示手段の概念を示す一例である。

[図7]この発明の実施の形態2を示す話者照合装置の構成図である。

[図8]この発明の実施の形態2における話者照合装置の処理内容を示すフローチャートである。

[図9]この発明の実施の形態3を示す話者照合装置の構成図である。

[図10]この発明の実施の形態3における話者照合装置の処理内容を示すフローチャートである。

[図11]この発明の実施の形態4を示す話者照合装置の構成図である。

[図12]この発明の実施の形態4における話者照合装置の処理内容を示すフローチャートである。

符号の説明

[0081] 100 利用者登録手段

101 登録要求者

102 登録要求者音声

103 個人識別子

104 暗証文字列

105 登録手段

106 登録データベース

107 認証要求者

108 認証要求者音声

109 認証個人識別子

110 音声単位データベース

111 音声単位提示手段

112 音声類似度計算手段

113 閾値

114 判定手段

115 認証結果

117 音節

118 音声標準パターン

120 暗証文字列構成文字

121 暗証文字列構成文字群

130 音声単位

201 登録要求者属性

204 音声単位提示手段

205 応答時間計算手段

206 判定用スコア計算手段

301 登録要求者唇画像

304 認証要求者唇画像

305 音声及び唇画像類似度計算手段

403 照合用音声標準パターン群

404 照合用音声標準パターン群集合

405 照合用音声標準パターン選択手段

請求の範囲

[1] 認証要求者に発声させるべき音声単位を認証要求者に提示する際に前回提示した音声単位を他の音声単位に切り替えて提示する音声単位提示手段と、
この提示する音声単位と個人識別子とに対応づけて準備される音声標準パターンと、
この準備された音声標準パターンと上記認証要求者が発声した音声とを比較して音声の類似度を計算する音声類似度計算手段と、
この音声類似度計算手段の計算結果に基づいて上記認証要求者が正当な利用者か否かを判定する判定手段と、
を備えたことを特徴とする話者照合装置。

[2] 音声単位提示手段は、画面表示により認証要求者に発声させるべき音声単位を提示するものとし、
この画面表示する内容には、
認証要求者に発声させるべき複数の音声単位により構成される音声単位群と、
この音声単位群の各音声単位に対応づけて配置表示されかつ個人識別子に対応した暗証文字列を構成する文字からなる文字群と、
が含まれることを特徴とする請求項1に記載の話者照合装置。

[3] 判定手段は、
音声単位提示手段の提示内容に対して認証要求者が応答するのにかかった時間及び音声類似度計算手段の計算結果に基づいて認証要求者が正当な利用者か否かを判断する
ことを特徴とする請求項1に記載の話者照合装置。

[4] 音声標準パターンは、
登録利用者の属性情報に基づき作成され、かつ分別された照合用音声標準パターン群を構成し、
音声類似度計算手段は、
個人識別子により特定される属性に基づいて照合用音声標準パターン群集合から選択する音声標準パターンと音声とを比較して類似度を計算する、

ことを特徴とする請求項1に記載の話者照合装置。

[5] 認証要求者に発声させるべき音声単位を認証利用者に提示する際に前回提示した音声単位を他の音声単位に切り替えて提示する音声単位提示手段と、

この提示する音声単位と個人識別子とに対応づけて準備される音声標準パターンと、

この準備された音声標準パターンと上記認証要求者が発声した音声とを比較して音声の類似度を計算する音声類似度計算手段と、

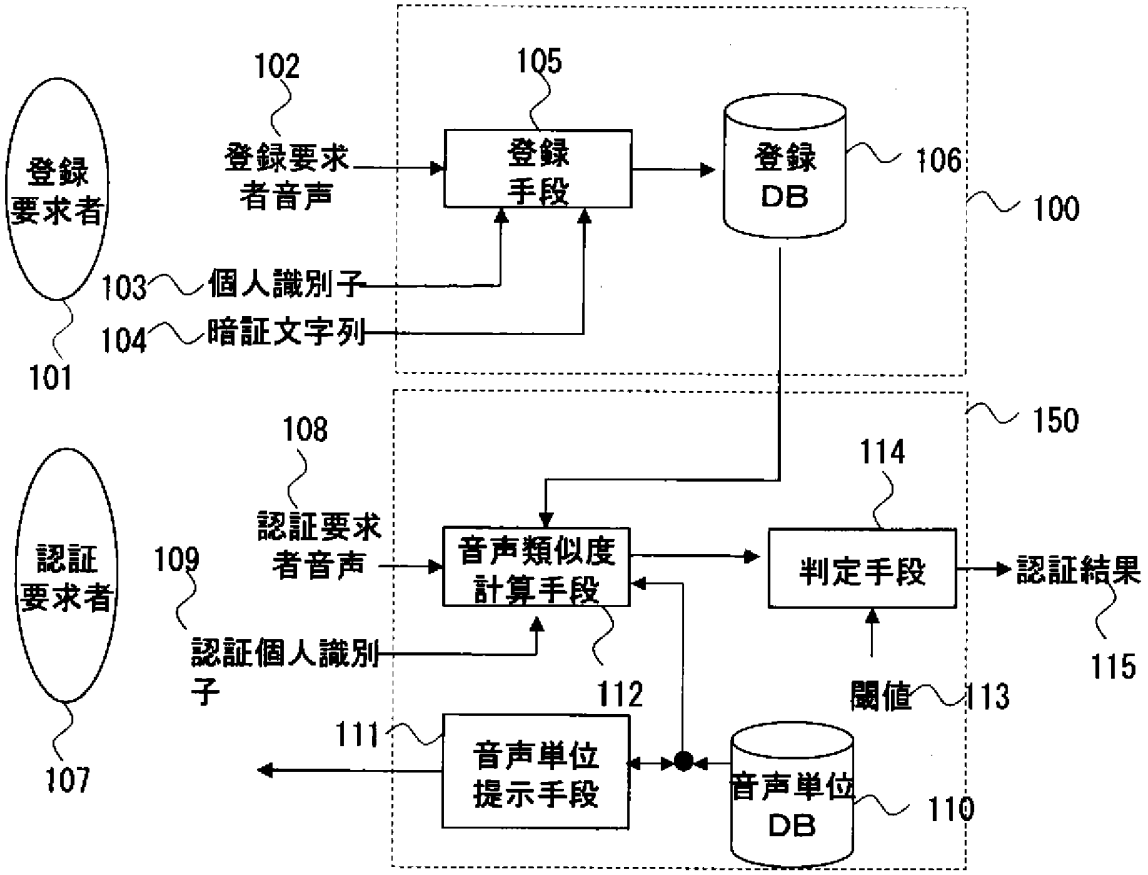
上記提示する音声単位と個人識別子とに対応づけて準備される唇画像標準パターンと、

この準備された唇画像標準パターンと上記認証要求者を撮影した唇画像とを比較して唇画像の類似度を計算する唇画像類似度計算手段と、

上記音声類似度計算手段の結果と上記唇画像類似度計算手段との計算結果に基づいて上記認証要求者が正当な利用者か否かを判定する判定手段と

を備えたことを特徴とする話者照合装置。

[図1]

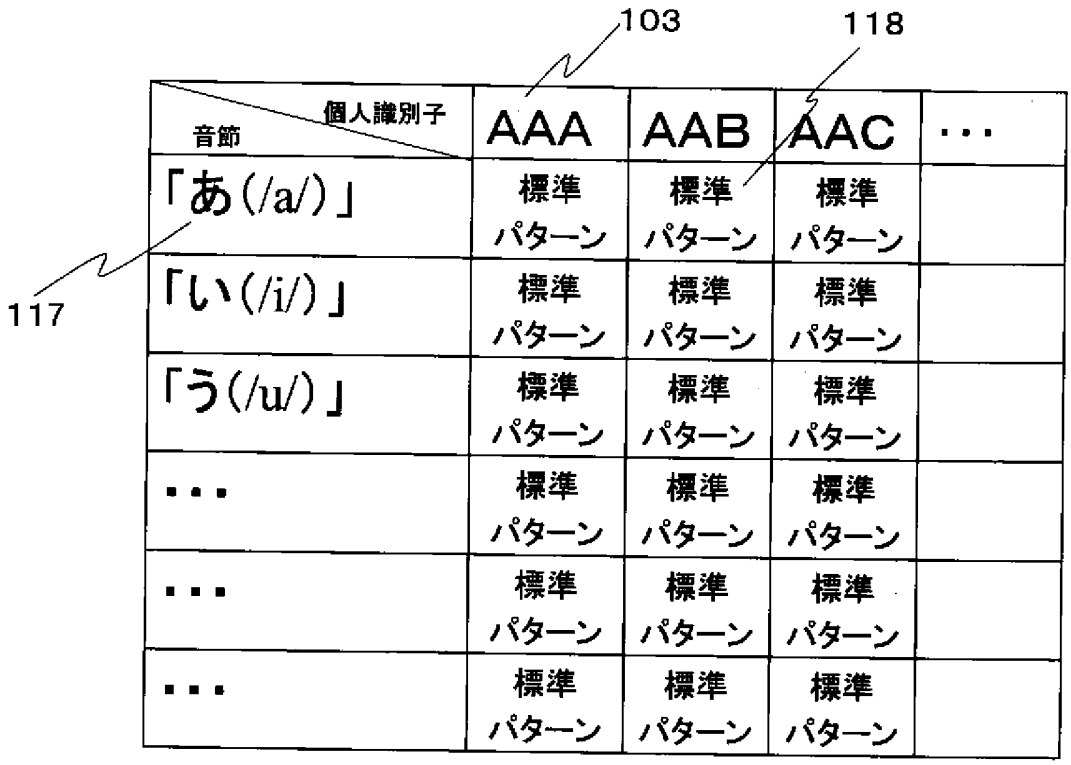


[図2]

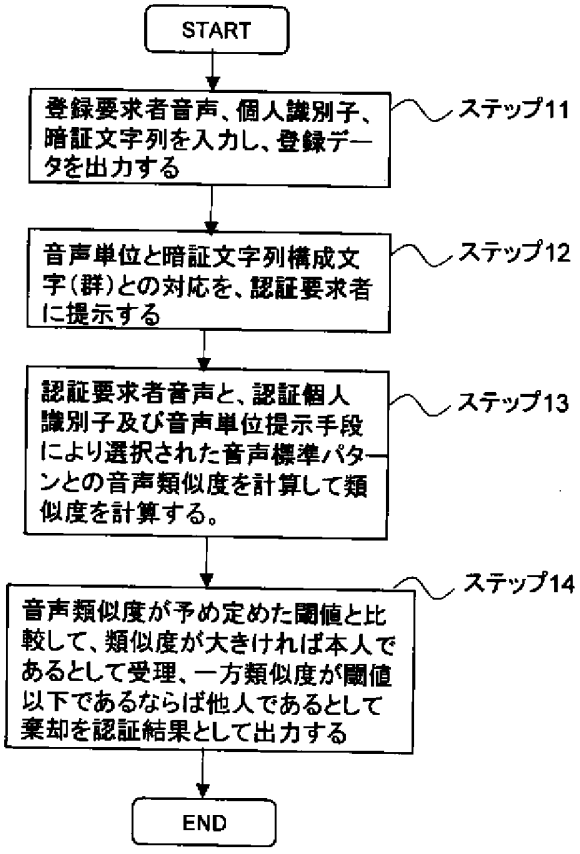
音声単位	暗証文字列構成文字群
はちのへ	0、2、6
けせんぬま	1、9
ゆくはし	3、4、7
さっぽろ	5、8

130
120
121

[図3]


103
118
117
個人識別子
音節
AAA
AAB
AAC
・・・
「あ(/a/)」
「い(/i/)」
「う(/u/)」
標準パターン
・・・

[図4]


START
登録要求者音声、個人識別子、暗証文字列を入力し、登録データを出力する
ステップ11
音声単位と暗証文字列構成文字(群)との対応を、認証要求者に提示する
ステップ12
認証要求者音声と、認証個人識別子及び音声単位提示手段により選択された音声標準パターンとの音声類似度を計算して類似度を計算する。
ステップ13
音声類似度が予め定めた閾値と比較して、類似度が大きければ本人であるとして受理、一方類似度が閾値以下であるならば他人であるとして棄却を認証結果として出力する
ステップ14
END

[図5]

音声単位	暗証文字列構成文字
は	0
け	1
ゆ	2
さ	3
ち	4
せ	5
く	6
つ	7
の	8
ん	9

130

120

[図6]

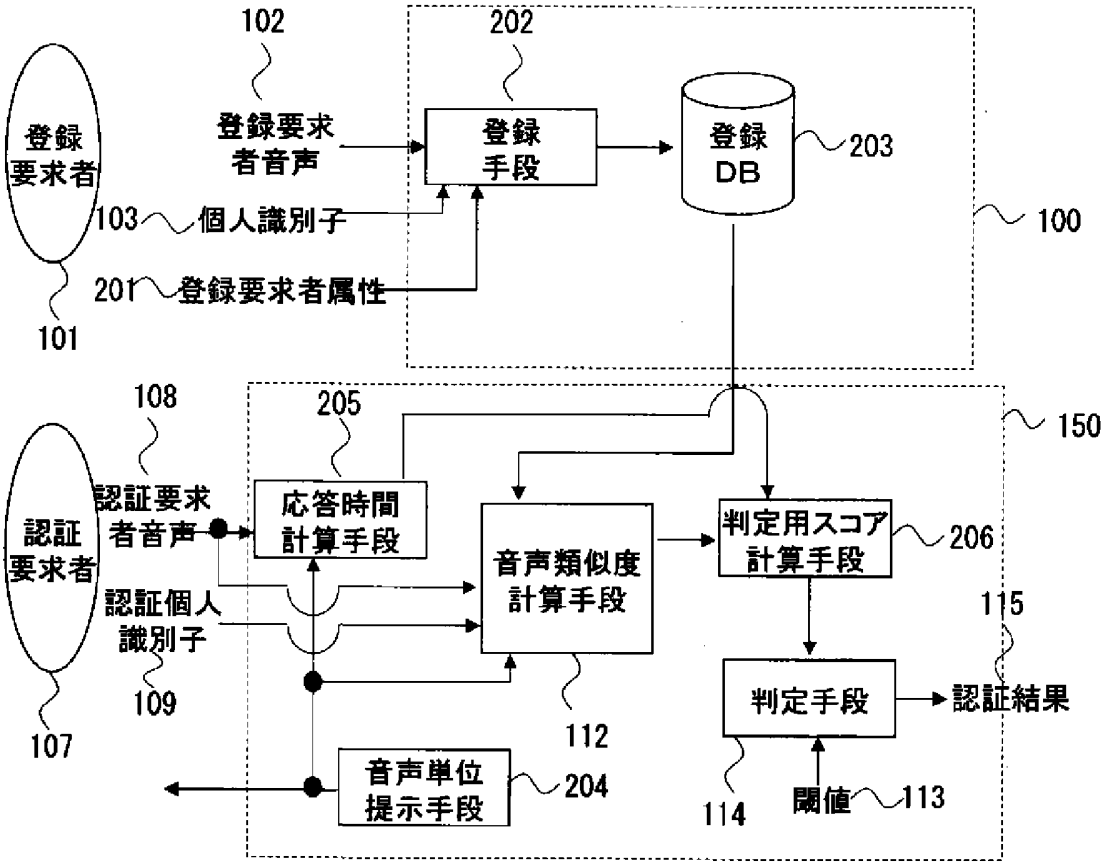
音声単位	暗証文字列構成文字群
は	0、2、6
け	1、9
ゆ	3、4、7
さ	5、8

130

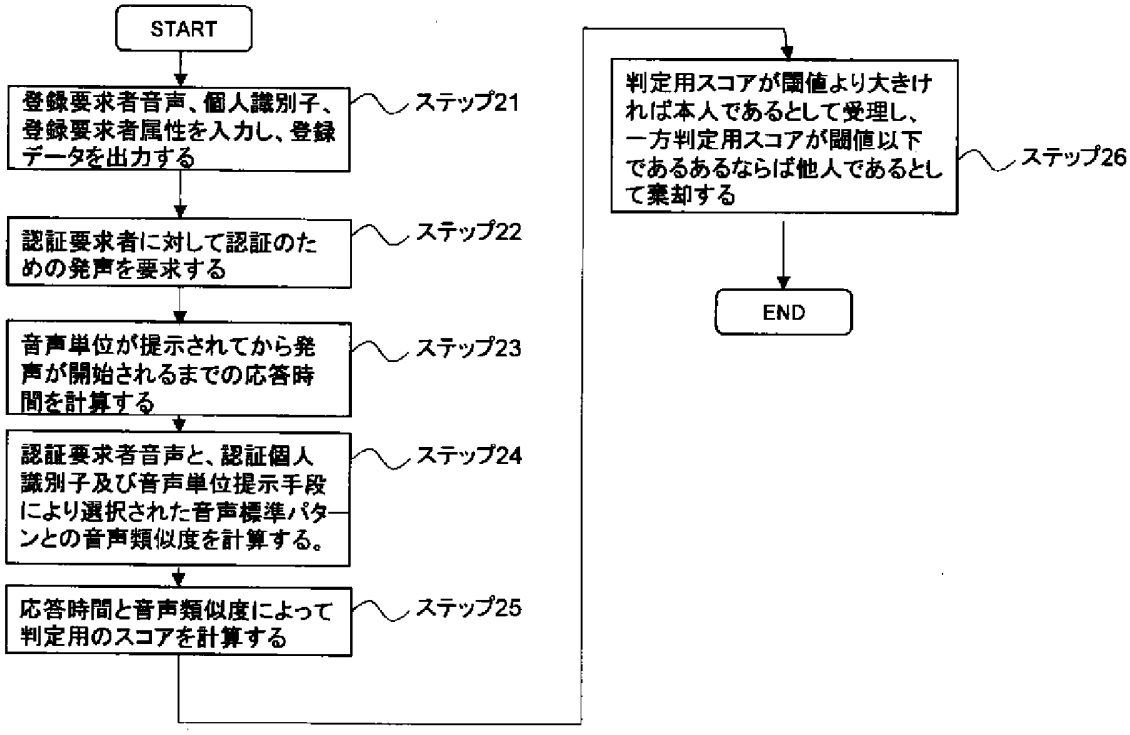
120

121

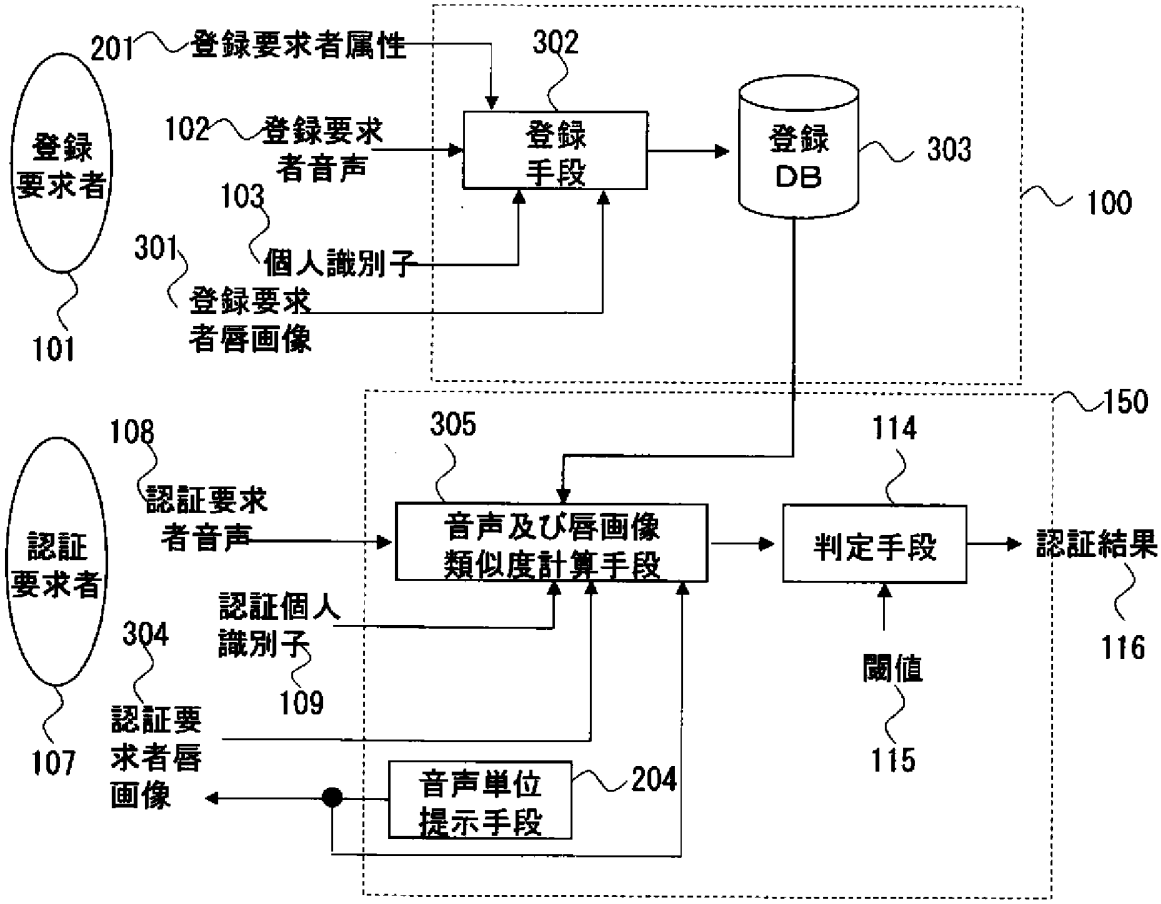
[図7]


102
202
登録要求者
登録要求者音声
登録手段
登録DB
203
100
103
個人識別子
201
登録要求者属性
101
108
205
150
認証要求者
認証要求者音声
応答時間計算手段
音声類似度計算手段
判定用スコア計算手段
206
認証個人識別子
115
109
判定手段
認証結果
107
112
音声単位提示手段
204
114
閾値
113

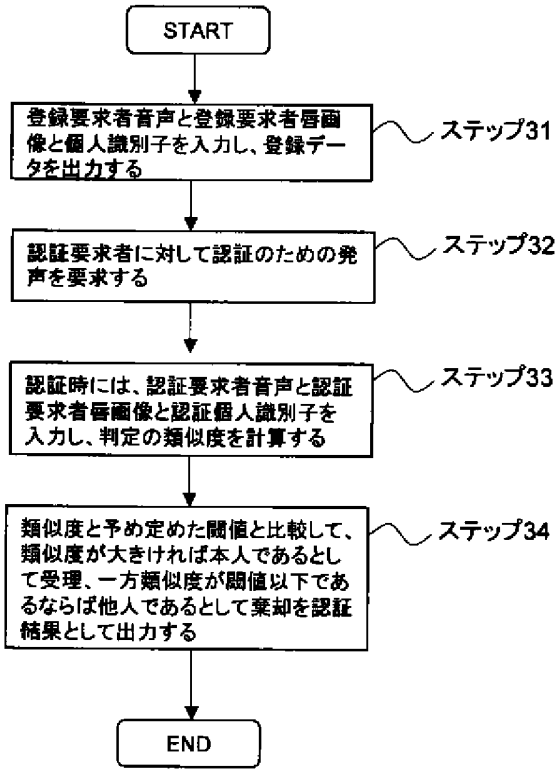
[図8]


START
登録要求者音声、個人識別子、登録要求者属性を入力し、登録データを出力する
ステップ21
認証要求者に対して認証のための発声を要求する
ステップ22
音声単位が提示されてから発声が開始されるまでの応答時間を計算する
ステップ23
認証要求者音声と、認証個人識別子及び音声単位提示手段により選択された音声標準パターンとの音声類似度を計算する。
ステップ24
応答時間と音声類似度によって判定用のスコアを計算する
ステップ25
判定用スコアが閾値より大きければ本人であるとして受理し、一方判定用スコアが閾値以下であるあるならば他人であるとして棄却する
ステップ26
END

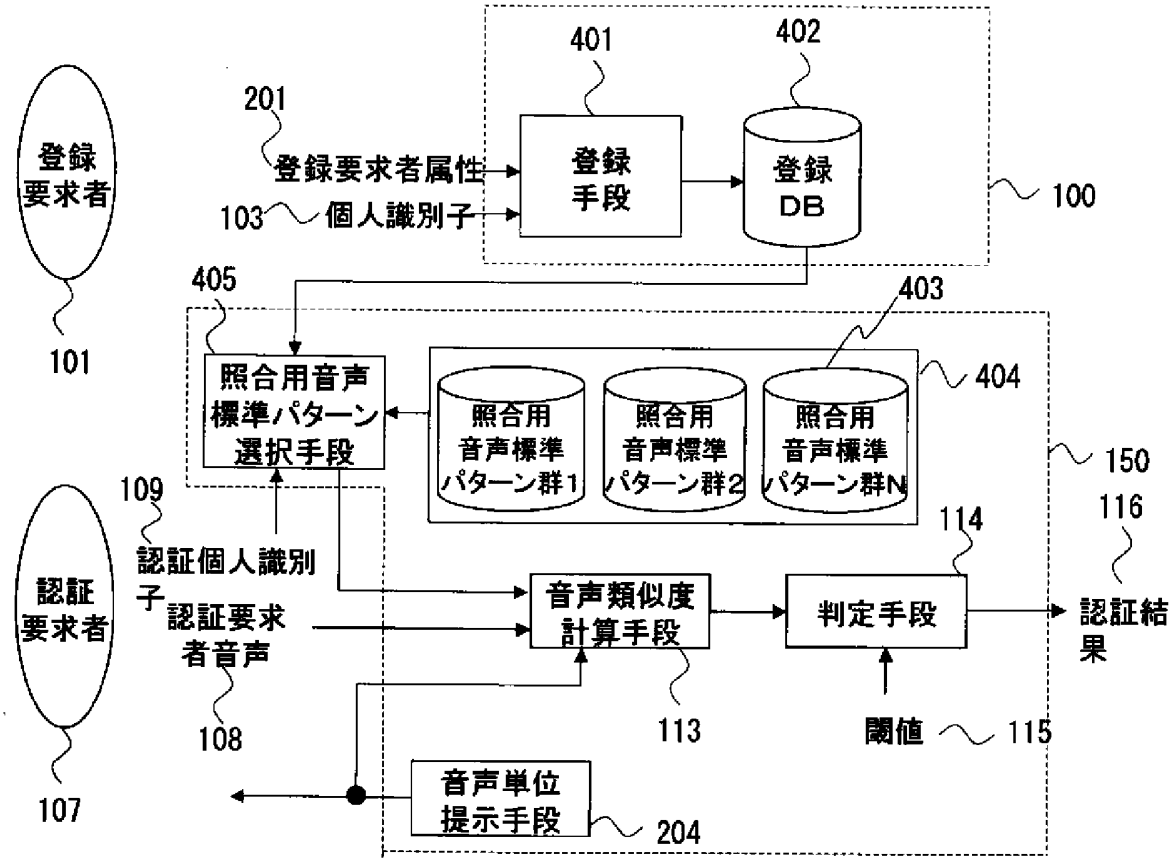
[図9]


201
登録要求者属性
302
登録要求者
102
登録要求者音声
登録手段
登録DB
303
100
103
個人識別子
301
登録要求者唇画像
101
150
108
認証要求者音声
305
114
認証要求者
音声及び唇画像類似度計算手段
判定手段
認証結果
認証個人識別子
304
109
116
閾値
認証要求者唇画像
107
音声単位提示手段
204
115

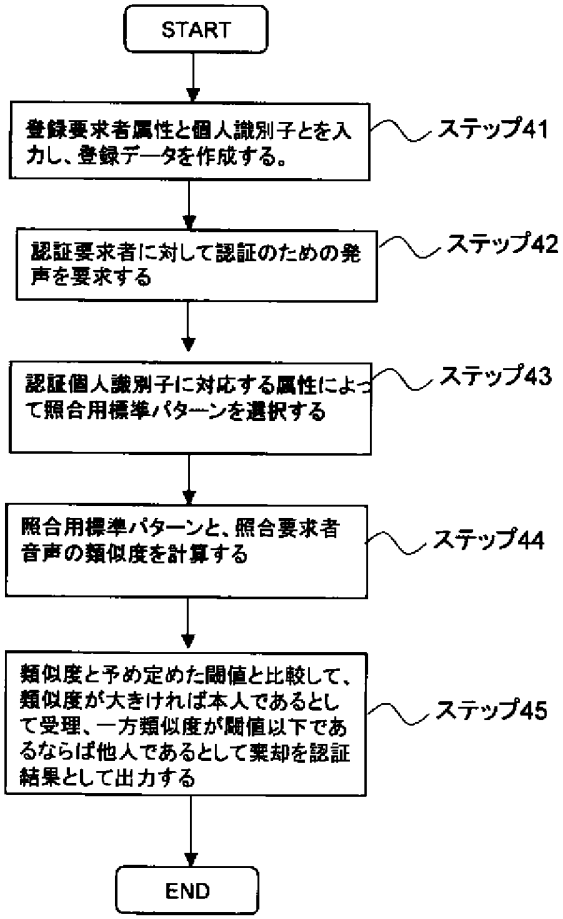
[図10]


START
登録要求者音声と登録要求者唇画像と個人識別子を入力し、登録データを出力する
ステップ31
認証要求者に対して認証のための発声を要求する
ステップ32
認証時には、認証要求者音声と認証要求者唇画像と認証個人識別子を入力し、判定の類似度を計算する
ステップ33
類似度と予め定めた閾値と比較して、類似度が大きければ本人であるとして受理、一方類似度が閾値以下であるならば他人であるとして棄却を認証結果として出力する
ステップ34
END

[図11]


登録要求者
101
201
登録要求者属性
103
個人識別子
401
登録手段
402
登録DB
100
405
照合用音声標準パターン選択手段
403
404
照合用音声標準パターン群1
照合用音声標準パターン群2
照合用音声標準パターン群N
150
109
認証個人識別子
認証要求者
107
認証要求者音声
108
音声類似度計算手段
113
114
判定手段
116
認証結果
閾値
115
音声単位提示手段
204

[図12]


START
登録要求者属性と個人識別子とを入力し、登録データを作成する。
ステップ41
認証要求者に対して認証のための発声を要求する
ステップ42
認証個人識別子に対応する属性によって照合用標準パターンを選択する
ステップ43
照合用標準パターンと、照合要求者音声の類似度を計算する
ステップ44
類似度と予め定めた閾値と比較して、類似度が大きければ本人であるとして受理、一方類似度が閾値以下であるならば他人であるとして棄却を認証結果として出力する
ステップ45
END

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2004/013197

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl[7] G10L17/00, E05B49/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl[7] G10L17/00, E05B49/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1922-1996 Toroku Jitsuyo Shinan Koho 1994-2004
Kokai Jitsuyo Shinan Koho 1971-2004 Jitsuyo Shinan Toroku Koho 1996-2004

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	JP 04-326409 A (Mitsubishi Electric Corp.), 16 November, 1992 (16.11.92), (Family: none)	1,3 2,4,5
Y	JP 2002-311992 A (Fujitsu Ltd.), 25 October, 2002 (25.10.02), (Family: none)	2,4
Y	JP 2000-338987 A (Mitsubishi Electric Corp.), 08 December, 2000 (08.12.00), (Family: none)	5

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 October, 2004 (18.10.04)	02 November, 2004 (02.11.04)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

Form PCT/ISA/210 (second sheet) (January 2004)

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２００４／０１３１９７

A.　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Ｉｎｔ．Ｃｌ⁷　Ｇ１０Ｌ１７／００、Ｅ０５Ｂ４９／００

B.　調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

Ｉｎｔ．Ｃｌ⁷　Ｇ１０Ｌ１７／００、Ｅ０５Ｂ４９／００

最小限資料以外の資料で調査を行った分野に含まれるもの
日本国実用新案公報　　１９２２－１９９６年
日本国公開実用新案公報　１９７１－２００４年
日本国登録実用新案公報　１９９４－２００４年
日本国実用新案登録公報　１９９６－２００４年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C.　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
X Y	ＪＰ　04-326409 A（三菱電機株式会社） 1992.11.16（ファミリーなし）	1，3 2，4，5
Y	ＪＰ　2002-311992 A（富士通株式会社） 2002.10.25（ファミリーなし）	2，4
Y	ＪＰ　2000-338987 A（三菱電機株式会社） 2000.12.08（ファミリーなし）	5

☐ Ｃ欄の続きにも文献が列挙されている。　　☐ パテントファミリーに関する別紙を参照。

＊　引用文献のカテゴリー
「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
１８．１０．２００４	02.11.2004

国際調査機関の名称及びあて先	特許庁審査官（権限のある職員）	
日本国特許庁（ＩＳＡ／ＪＰ） 郵便番号１００－８９１５ 東京都千代田区霞が関三丁目４番３号	山下　剛史 電話番号　０３－３５８１－１１０１　内線　３５４１	５Ｃ　３１４８

様式ＰＣＴ／ＩＳＡ／２１０（第２ページ）（２００４年１月）