

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5151103号
(P5151103)

(45) 発行日 平成25年2月27日(2013.2.27)

(24) 登録日 平成24年12月14日(2012.12.14)

(51) Int.Cl. F I
G 1 0 L 17/00 (2013.01) G 1 0 L 17/00 2 0 0 B
 G 1 0 L 17/00 2 0 0 C

請求項の数 7 (全 13 頁)

(21) 出願番号 特願2006-249161 (P2006-249161)
 (22) 出願日 平成18年9月14日(2006.9.14)
 (65) 公開番号 特開2008-70597 (P2008-70597A)
 (43) 公開日 平成20年3月27日(2008.3.27)
 審査請求日 平成21年7月17日(2009.7.17)

(73) 特許権者 000004075
 ヤマハ株式会社
 静岡県浜松市中区中沢町10番1号
 (74) 代理人 100125689
 弁理士 大林 章
 (74) 代理人 100121108
 弁理士 高橋 太郎
 (74) 代理人 100125335
 弁理士 矢代 仁
 (72) 発明者 吉岡 靖雄
 静岡県浜松市中沢町10番1号 ヤマハ株
 式会社内
 (72) 発明者 川▲原▼ 毅彦
 静岡県浜松市中沢町10番1号 ヤマハ株
 式会社内

最終頁に続く

(54) 【発明の名称】 音声認証装置、音声認証方法およびプログラム

(57) 【特許請求の範囲】

【請求項1】

認証時に被認証者の周囲に発生する認証時雑音の特性を分析する特性分析手段と、
予め登録された登録音声の登録時の登録時雑音と前記登録音声との音量比に対して、前記特性分析手段が分析した認証時雑音の特性に応じた関係を満たすように、 閾値を設定する閾値設定手段と、

前記登録音声と被認証者から採取された認証音声との特徴量の類否を示す指標値と前記閾値設定手段が設定した閾値との比較の結果に応じて当該被認証者を認証する認証手段とを具備する音声認証装置。

【請求項2】

前記閾値設定手段は、前記登録時雑音と前記登録音声との音量比に対し、前記認証時雑音に応じた直線または曲線に沿って閾値が変化するように、前記閾値を設定する

請求項1に記載の音声認証装置。

【請求項3】

前記閾値設定手段は、前記登録時雑音と前記登録音声との音量比に対し、前記認証時雑音および前記認証音声の音量比と前記登録時雑音および前記登録音声の音量比との相違に応じた直線または曲線に沿って閾値が変化するように、前記閾値を設定する

請求項2に記載の音声認証装置。

【請求項4】

前記閾値設定手段は、前記認証時雑音と前記登録時雑音との相違に応じて閾値を補正す

る補正手段を含む

請求項 1 から請求項 3 の何れかに記載の音声認証装置。

【請求項 5】

前記閾値設定手段は、前記認証音声または前記登録音声の時間長に応じて閾値を補正する補正手段を含む

請求項 1 から請求項 4 の何れかに記載の音声認証装置。

【請求項 6】

認証時に被認証者の周囲に発生する認証時雑音の特性を分析し、
予め登録された登録音声の登録時の登録時雑音と前記登録音声との音量比に対して、前記分析した認証時雑音の特性に応じた関係を満たすように、閾値を設定し、

前記登録音声と被認証者から採取された認証音声との特徴量の類否を示す指標値と前記設定した閾値との比較の結果に応じて当該被認証者を認証する
 音声認証方法。

【請求項 7】

コンピュータに、
 認証時に被認証者の周囲に発生する認証時雑音の特性を分析する特性分析処理と、
予め登録された登録音声の登録時の登録時雑音と前記登録音声との音量比に対して、前記特性分析処理で分析した認証時雑音の特性に応じた関係を満たすように、閾値を設定する閾値設定処理と、

前記登録音声と被認証者から採取された認証音声との特徴量の類否を示す指標値と前記閾値設定処理で設定した閾値との比較の結果に応じて当該被認証者を認証する認証処理と
 を実行させるプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声を利用した認証の技術に関する。

【背景技術】

【0002】

正当な利用者から事前に採取された音声（以下「登録音声」という）の特徴量と認証の対象者（以下「被認証者」という）から採取された音声（以下「認証音声」という）の特徴量との距離を閾値と比較することで被認証者の正当性を判別する音声認証の技術が従来から提案されている。また、特許文献 1 には、認証の目的や必要な精度に応じて閾値を変更する構成が開示されている。

【特許文献 1】特開 2003 - 248661 号公報

【発明の開示】

【発明が解決しようとする課題】

【0003】

図 8 は、音声認証の評価に使用されるグラフである。同図における FRR (False Rejection Rate) は、被認証者が正当な利用者であるにも拘わらず認証で正当性が否定される確率（本人拒否率）を意味し、FAR (False Acceptance Rate) は、被認証者が正当な利用者ではないにも拘わらず認証で正当性が肯定される確率（他人受入率）を意味する。同図から理解されるように、認証に使用される閾値を図 8 の数値 a に設定すれば、不当な被認証者が拒否される確度を十分に高水準に維持しながら、正当な利用者が拒否される可能性は十分に低減される。

【0004】

しかし、認証音声の特性は認証時に周囲で発生している音声（以下「認証時雑音」という）に影響されるから、認証音声と登録音声との距離は認証時雑音に応じて変動する。したがって、FRR や FAR の各々の曲線は、認証時雑音の特性に応じて横軸（距離）に沿って平行に移動する。そして、例えば図 8 に実線で図示された FAR が破線 L1 に変動した場合には、他人の正当性が誤認される確率が上昇（すなわち認証の精度が低下）し、図

10

20

30

40

50

8のFRRが破線L2に変動した場合には正当な利用者の認証が拒絶される確率が上昇(すなわち利便性が低下)する。すなわち、従来の音声認証においては、認証時雑音の特性に応じて認証の精度と利便性との均衡が崩れるという問題がある。

【0005】

特許文献1のように認証の目的や必要な精度に応じて閾値を変更しても以上の問題は何ら解決しない。また、携帯電話機に代表される可搬型の電子機器で認証を実行する場合には電子機器の使用される環境に応じて認証時雑音の特性が多様に変化するから、以上の問題は特に深刻化する。このような事情に鑑みて、本発明は、認証時雑音に拘わらず認証の精度と利便性との均衡を維持するという課題の解決を目的としている。

【課題を解決するための手段】

【0006】

以上の課題を解決するために、本発明のひとつの形態に係る音声認証装置は、認証時に被認証者の周囲に発生する認証時雑音の特性を分析する特性分析手段と、予め登録された登録音声の登録時の登録時雑音と登録音声との音量比(例えば図4の音量比EN_SN)に対して、特性分析手段が分析した認証時雑音の特性に応じた関係を満たすように、閾値を設定する閾値設定手段と、登録音声と被認証者から採取された認証音声との特徴量の類否を示す指標値と閾値設定手段が設定した閾値との比較の結果に応じて当該被認証者を認証する認証手段とを具備する。以上の態様によれば、認証時雑音に応じて閾値が可変に設定されるから、認証時雑音に拘わらず認証の精度と利便性との均衡を維持することが可能である。

【0007】

本発明の好適な態様において、閾値設定手段は、登録時雑音と登録音声との音量比に対し、認証時雑音に応じた直線または曲線に沿って閾値が変化するように、閾値を設定する。さらに詳述すると、閾値設定手段は、登録時雑音と登録音声との音量比に対し、認証時雑音および認証音声の音量比(例えば音量比V_SN)と登録時雑音および登録音声の音量比との相違(例えば図4のDIF_SN1~DIF_SN3)に応じた直線または曲線に沿って閾値が変化するように、閾値を設定する。以上の態様によれば、登録時雑音や認証時雑音の特性に拘わらず簡易な処理によって認証の精度と利便性との均衡を維持することが可能となる。なお、閾値設定手段は、登録時雑音および登録音声の音量比と認証時雑音(より詳細には認証時雑音および認証音声の音量比と登録時雑音および登録音声の音量比との相違)と閾値との関係を定義するテーブルに基づいて閾値を設定してもよいし、これらの数値の関係を表現する数式を利用した演算によって閾値を算定してもよい。テーブルや数式の内容は、例えば利用者からの指示に応じて適宜に変更され得る。

【0008】

本発明のひとつの態様において、閾値設定手段は、認証時雑音と登録時雑音との相違(例えば図6の相関値NOISE_DIF)に応じて閾値を補正する補正手段(例えば図1の補正部54)を含む。以上の態様によれば、実際の認証時における認証時雑音と登録時雑音との相関が、登録時雑音および登録音声の音量比と閾値との関係を決定するときに想定した認証時雑音と登録時雑音との相関とは相違する場合であっても、補正手段が閾値を補正することで認証の精度と利便性との均衡を有効に維持することが可能となる。なお、補正手段は、認証時雑音および登録時雑音の相違と補正值(例えば図6の補正值A1)との関係を定義するテーブルに基づいて閾値に対する補正值を設定してもよいし、この関係を表現する数式を利用した演算によって閾値を算定してもよい。テーブルや数式の内容は、例えば利用者からの指示に応じて適宜に変更され得る。

【0009】

本発明のひとつの態様において、閾値設定手段は、認証音声または登録音声の時間長(例えば図7の発声長EN_SPEECH_LENや発声長V_SPEECH_LEN)に応じて閾値を補正する補正手段(例えば図1の補正部54)を含む。以上の態様によれば、実際の登録音声や認証音声の発声長が、登録時雑音および登録音声の音量比と閾値との関係を決定するときに想定した発声長とは相違する場合であっても、補正手段が閾値を補正することで認証の精度と

10

20

30

40

50

利便性との均衡を有効に維持することが可能となる。なお、補正手段は、認証音声または登録音声の時間長と補正值（例えば図7の補正值A2）との関係を定義するテーブルに基づいて閾値に対する補正值を設定してもよいし、この関係を表現する数式を利用した演算によって閾値を算定してもよい。テーブルや数式の内容は、例えば利用者からの指示に応じて適宜に変更され得る。

【0010】

なお、閾値に対する補正值を決定するための基準となる変数は、認証時雑音と登録時雑音との相違や認証音声または登録音声の時間長に限定されない。例えば、登録音声や認証音声のうち有声音と無声音との時間長の比率に応じて閾値を補正する補正手段、または、登録音声や認証音声の音節の個数に応じて閾値を補正する補正手段を、閾値設定手段に含ませてもよい。何れの態様においても、以上と同様に、テーブルや数式に応じて補正值を決定する構成や、テーブルや数式の内容が可変とされた構成が採用される。

10

【0011】

本発明は、以上の各態様に係る音声認証装置の動作方法（音声認証方法）としても特定される。本発明のひとつの態様に係る音声認証方法は、被認証者の周囲に認証時に発生する認証時雑音の特性を分析し、予め登録された登録音声の登録時の登録時雑音と登録音声との音量比に対して、分析した認証時雑音の特性に応じた関係を満たすように、閾値を設定し、登録音声と被認証者から採取された認証音声との特徴量の類否を示す指標値と設定した閾値との比較の結果に応じて当該被認証者を認証する。以上の方法によれば、本発明に係る音声認証装置と同様の作用および効果が奏される。

20

【0012】

以上の各態様に係る音声認証装置は、各処理に専用されるDSP（Digital Signal Processor）などのハードウェア（電子回路）によって実現されるほか、CPU（Central Processing Unit）などの汎用の演算処理装置とプログラムとの協働によっても実現される。本発明に係るプログラムは、被認証者の周囲に認証時に発生する認証時雑音の特性を分析する特性分析処理と、予め登録された登録音声の登録時の登録時雑音と登録音声との音量比に対して、特性分析処理で分析した認証時雑音の特性に応じた関係を満たすように、閾値を設定する閾値設定処理と、登録音声と被認証者から採取された認証音声との特徴量の類否を示す指標値と閾値設定処理で設定した閾値との比較の結果に応じて当該被認証者を認証する認証処理とを実行させる内容である。以上のプログラムによっても、以上の各態様に係る音声認証装置と同様の作用および効果が奏される。なお、本発明のプログラムは、CD-ROMなど可搬型の記録媒体に格納された形態で利用者に提供されてコンピュータにインストールされるほか、ネットワークを介した配信の形態でサーバ装置から提供されてコンピュータにインストールされる。

30

【発明を実施するための最良の形態】

【0013】

< A : 音声認証装置の構成 >

図1は、本発明のひとつの形態に係る音声認証装置の構成を示すブロック図である。音声認証装置100は、被認証者が特定の言葉を発声したときの音声に基づいて被認証者の正当性（予め登録された正規の利用者であるか否か）を判定する装置であり、携帯電話機や情報処理装置など各種の電子機器に搭載される。図1に図示された特性分析部20や認証部40や閾値設定部50は、例えばCPUなどの演算処理装置がプログラムを実行することで実現されてもよいし、DSPなどのハードウェア回路によって実現されてもよい。

40

【0014】

音声認証装置100の動作は初期登録と認証とに区分される。初期登録は、正当な利用者が発声した音声（登録音声）を認証に先立って登録する動作である。認証は、登録音声と被認証者が発声した音声（認証音声）との照合によって被認証者の正当性を認証する動作である。操作部10は、利用者によって操作される複数の操作子を含む。利用者は、操作部10を適宜に操作することで初期登録や認証の開始を音声認証装置100に指示することができる。

50

【 0 0 1 5 】

図 1 の入力部 1 5 および特性分析部 2 0 は、認証時には、認証音声や音声認証装置 1 0 0 の周囲の雑音（認証時雑音）の特性を検出するために使用され、初期登録時には、同図に破線 R で図示されるように、登録音声や音声認証装置 1 0 0 の周囲の雑音（登録時雑音）の特性を検出するために使用される。

【 0 0 1 6 】

入力部 1 5 は、周囲の音響（音声および雑音）に応じた音響信号 S を生成する收音機器である。図 2 に例示されるように、音響信号 S は、非発声区間 P 1 と発声区間 P 2 とに区分される。発声区間 P 2 は、初期登録に際して正当な利用者が登録音声を発声した区間や認証に際して被認証者が認証音声を発声した区間である。一方、非発声区間 P 1 は、登録音声や認証音声が発声されない区間である。音声認証装置 1 0 0 が設置された環境には各種の雑音が発生し得るから、非発声区間 P 1 においても完全な無音（音響信号 S の振幅がゼロ）ではなく、図 2 に示すように登録時雑音や認証時雑音が入力部 1 5 によって收音される。

10

【 0 0 1 7 】

入力部 1 5 が生成した音響信号 S は図 1 の特性分析部 2 0 に供給される。特性分析部 2 0 は、入力部 1 5 が採取した音響を分析する手段であり、区間検出部 2 2 と切換部 2 3 と雑音分析部 2 5 と音声分析部 2 6 と特徴分析部 2 8 とを含む。区間検出部 2 2 は、非発声区間 P 1 と発声区間 P 2 とを区分する。例えば、区間検出部 2 2 は、音響信号 S の振幅が不連続に増減した時点を非発声区間 P 1 と発声区間 P 2 との境界として検出する。なお、非発声区間 P 1 と発声区間 P 2 との区分には公知の様々な技術が採用される。

20

【 0 0 1 8 】

切換部 2 3 は、入力部 1 5 が生成した音響信号 S の供給先を選択的に切替える手段である。音響信号 S のうち区間検出部 2 2 が非発声区間 P 1 と認定した区間は雑音分析部 2 5 に供給され、区間検出部 2 2 が発声区間 P 2 と認定した区間は音声分析部 2 6 と特徴分析部 2 8 とに供給される。

【 0 0 1 9 】

雑音分析部 2 5 は、非発声区間 P 1 の音響信号 S に基づいて登録時雑音や認証時雑音の特性を分析する手段である。本形態の雑音分析部 2 5 は、非発声区間 P 1 内において周期的に音響信号 S の特性を分析する。そして、操作部 1 0 に対する操作に応じて初期登録または認証の開始が指示されると、雑音分析部 2 5 は、図 2 に示すように、当該指示の時点から所定の時間長だけ手前の時点までの区間（以下「検出区間」という）P における分析の結果を登録時雑音や認証時雑音の特性として確定する。なお、以下の説明において、登録時雑音や登録音声に関連する事項は「EN (enroll)」を含む符号で指示され、認証時雑音や認証音声に関連する事項は「V (verify)」を含む符号で指示される。

30

【 0 0 2 0 】

図 1 に示すように、本形態の雑音分析部 2 5 は、初期登録時には、登録時雑音について周波数特性 EN_NOISE_FC と雑音レベル EN_NOISE_LEVEL とを算定し、認証時には、認証時雑音について周波数特性 V_NOISE_FC と雑音レベル V_NOISE_LEVEL とを算定する。雑音レベル (EN_NOISE_LEVEL, V_NOISE_LEVEL) は、非発声区間 P 1 内の検出区間 P における音響信号 S のうち所定の周波数帯域に属する成分の強度（音圧）の平均値である。周波数特性 (EN_NOISE_FC, V_NOISE_FC) は、検出区間 P の音響信号 S を複数の周波数帯域に区分したときの各成分の強度を示す情報である。したがって、雑音分析部 2 5 は、例えば各々の通過帯域が相違する複数のバンドパスフィルタ（フィルタバンク）を含む。ただし、雑音分析部 2 5 は、FFT (Fast Fourier Transform) 処理などの周波数分析によって周波数スペクトルを周波数特性 (EN_NOISE_FC, V_NOISE_FC) として算定する手段であってもよい。

40

【 0 0 2 1 】

音声分析部 2 6 は、発声区間 P 2 の音響信号 S に基づいて登録音声や認証音声の特性を分析する。本形態の音声分析部 2 6 は、初期登録時には、登録音声について発声レベル EN_SPEECH_LEVEL と発声長 EN_SPEECH_LEN とを算定し、認証時には、認証音声について発声レ

50

ベル V_SPEECH_LEVEL と発声長 V_SPEECH_LEN とを算定する。発声レベル (EN_SPEECH_LEVEL , V_SPEECH_LEVEL) は、発声区間 P 2内の音響信号 S のうち所定の周波数帯域に属する成分の強度の平均値である。発声長 (EN_SPEECH_LEN , V_SPEECH_LEN) は発声区間 P 2の時間長 (すなわち発声が継続される時間長) を示す。音響信号 S の振幅が急峻に増大する時点 (発声区間 P 2の始点) から音響信号 S の振幅が急峻に減少する時点 (発声区間 P 2の終点) までの時間長が発声長 (EN_SPEECH_LEN , V_SPEECH_LEN) として検出される。

【 0 0 2 2 】

特徴分析部 2 8 は、登録音声や認証音声の特徴を分析する手段である。本形態の特徴分析部 2 8 は、初期登録時には登録音声の特徴量 EN_DATA を算定し、認証時には認証音声の特徴量 V_DATA を算定する。特徴量 (EN_DATA , V_DATA) は、発声区間 P 2内の音響信号 S から算定されるケプストラムの時系列的なベクトル列である。したがって、周波数分析 (例えば FFT 処理) を含む各種の演算を実行する手段が特徴分析部 2 8 として好適に採用される。

10

【 0 0 2 3 】

記憶装置 3 2 は、認証に使用される各種の情報を記憶する手段である。例えば図 1 に図示されるように、記憶装置 3 2 は、特性分析部 2 0 が登録音声および登録時雑音について特定した各種の情報を認証用の辞書として記憶する。すなわち、雑音分析部 2 5 が特定した周波数特性 EN_NOISE_FC および雑音レベル EN_NOISE_LEVEL と、音声分析部 2 6 が特定した発声レベル EN_SPEECH_LEVEL および発声長 EN_SPEECH_LEN と、特徴分析部 2 8 が特定した特徴量 EN_DATA とが、認証に先立って記憶装置 3 2 に格納される。記憶装置 3 2 は、音声認証装置 1 0 0 に固定的に設置された機器であっても、音声認証装置 1 0 0 に対して自在に着脱される可搬型の機器 (メモリ) であってもよい。

20

【 0 0 2 4 】

認証部 4 0 は、登録音声と認証音声との照合によって被認証者の正当性を認証する手段であり、距離算定部 4 2 と判定部 4 4 とを含む。距離算定部 4 2 は、特徴分析部 2 8 が認証音声について生成した特徴量 V_DATA と記憶装置 3 2 に記憶された特徴量 EN_DATA との距離 $DIST$ を算定する。距離 $DIST$ の算定には、特徴量 EN_DATA および V_DATA の各々のベクトル列について相互間の正規化距離を算定する DP マッチングなど各種のパターンマッチング技術が利用される。距離 $DIST$ が小さいほど認証音声は登録音声に類似する (すなわち被認証者が正当な利用者である可能性が高い)。

30

【 0 0 2 5 】

判定部 4 4 は、距離算定部 4 2 が算定した距離 $DIST$ を閾値 TH と比較することで被認証者の正当性を判定する。すなわち、判定部 4 4 は、距離 $DIST$ が閾値 TH を下回る場合 (すなわち登録音声と認証音声とが類似する場合) には被認証者の正当性を肯定し、距離 $DIST$ が閾値 TH を上回る場合 (すなわち登録音声と認証音声とが乖離する場合) には被認証者の正当性を否定する。判定部 4 4 による判定の結果は出力部 6 0 から出力される。例えば、認証の結果を画像として出力する表示機器や認証の結果を音声で出力する放音装置が出力部 6 0 として好適に採用される。

【 0 0 2 6 】

閾値設定部 5 0 は、判定部 4 4 による判定に使用される閾値 TH を認証時雑音や登録時雑音に応じて可変に設定する手段であり、初期値設定部 5 2 と補正部 5 4 および 5 6 とを含む。初期値設定部 5 2 は、特性分析部 2 0 が初期登録時および認証時に生成した情報に基づいて閾値 TH の初期値を設定する。初期値設定部 5 2 が設定した閾値 TH は、補正部 5 4 および 5 6 による補正を経て判定部 4 4 に出力される。補正部 5 4 は、認証時雑音と登録時雑音との相違に応じて閾値 TH を補正する。補正部 5 6 は、初期登録時の発声長 EN_SPEECH_LEN に応じて閾値 TH を補正する。図 1 の記憶部 3 5 は、閾値 TH の設定や補正のために閾値設定部 5 0 が使用するテーブルを格納する。なお、記憶装置 3 2 内の特定の記憶領域を記憶部 3 5 としてもよい。

40

【 0 0 2 7 】

< B : 音声認証装置の動作 >

50

次に、認証時において閾値設定部 50 が閾値 TH を設定する処理を中心に音声認証装置 100 の動作を説明する。閾値設定部 50 は、認証の必要な時期が到来するたびに図 3 の処理を実行する。認証が必要な時機としては、例えば、音声認証装置 100 を搭載した電子機器が電源の投入を契機として動作を開始する時機や、電子機器が所定の動作（例えば特定の情報に対するアクセス）を開始する時機がある。認証が開始すると、被認証者は、操作部 10 を操作することで発声の開始を指示したうえで入力部 15 に対して所定の言葉を発声する。雑音分析部 25 は、操作部 10 への操作の時機を終点とする検出区間 P の音響信号 S から周波数特性 V_NOISE_FC および雑音レベル V_NOISE_LEVEL を特定するとともに、これに続く発声区間 P2 の音響信号 S から発声レベル V_SPEECH_LEVEL および発声長 V_SPEECH_LEN を特定する。

10

【0028】

図 3 に示すように、閾値設定部 50 は、登録時雑音と登録音声との音量比 EN_SN を算定する（ステップ S10）。音量比 EN_SN は、記憶装置 32 に格納された発声レベル EN_SPEECH_LEVEL と雑音レベル EN_NOISE_LEVEL との相対比であり、例えば以下の式(1)で算定される。

$$EN_SN = \log(EN_SPEECH_LEVEL / EN_NOISE_LEVEL) \quad \dots\dots (1)$$

【0029】

次に、閾値設定部 50 は、認証時雑音と認証音声との音量比 V_SN を算定する（ステップ S11）。音量比 V_SN は、雑音分析部 25 から供給される雑音レベル V_NOISE_LEVEL と音声分析部 26 から供給される発声レベル V_SPEECH_LEVEL との相対比であり、音量比 EN_SN と同様に以下の式(2)で算定される。

20

$$V_SN = \log(V_SPEECH_LEVEL / V_NOISE_LEVEL) \quad \dots\dots (2)$$

【0030】

次のステップ S12 において、閾値設定部 50 は、ステップ S10 で算定した初期登録時の音量比 EN_SN とステップ S11 で算定した今回の認証時の音量比 V_SN との差分値（以下「音量比差分値」という）DIF_SN を算定する（DIF_SN = V_SN - EN_SN）。さらに、閾値設定部 50 は、登録時雑音と認証時雑音との特性の相関（例えばスペクトル形状の類否）を示す相関値 NOISE_DIF を算定する（ステップ S13）。相関値 NOISE_DIF は、例えば以下の式(3)で算定される。

30

【数1】

$$NOISE_DIF = \frac{\sum_{i=M}^L (EN_MAG(i) - EN_MAG_AVE)(V_MAG(i) - V_MAG_AVE)}{\sqrt{\sum_{i=M}^L (EN_MAG(i) - EN_MAG_AVE)^2} \sqrt{\sum_{i=M}^L (V_MAG(i) - V_MAG_AVE)^2}} \quad \dots(3)$$

式(3)における数値 EN_MAG(i) は、複数の周波数帯域のうち変数 i で指定される周波数帯域における登録時雑音の強度であり、数値 EN_MAG(i) から減算される数値 EN_MAG_AVE は、変数 i で指定される周波数帯域における登録時雑音の強度の平均値である。同様に、数値 V_MAG(i) は、変数 i で指定される周波数帯域における認証時雑音の強度であり、数値 V_MAG_AVE は当該周波数帯域における認証時雑音の強度の平均値である。したがって、登録時雑音と認証時雑音とが完全に合致する場合には相関値 NOISE_DIF が「1」となり、両雑音の特性の相違が拡大するほど相関値 NOISE_DIF は減少していく（-1 < NOISE_DIF < 1）。

40

【0031】

次に、初期値設定部 52 は、ステップ S10 で算定した音量比 EN_SN とステップ S12 で算定した音量比差分値 DIF_SN とに基づいて閾値 TH の初期値を特定する（ステップ S14）。本願の発明者による試験によれば、認証の精度を高水準に維持するための閾値 TH は、音量比 EN_SN と音量比差分値 DIF_SN とに対して所定の関係を満たすという知見を得るに至った。すなわち、図 4 に示すように、音量比 EN_SN の数値（登録時雑音と登録音声との音量比）を変化させた各ケースにおいて認証の精度が所期の条件を満たす（例えば FRR や FAR が所期値を下回る）ように閾値 TH を設定し、横軸を音量比 EN_SN としたうえで各音

50

音量比EN_SNに対応した閾値THを縦軸にプロットして統計的に処理すると、各点は音量比差分値DIF_SNに応じた直線上に分布する傾向がある。図4には3種類の音量比差分値DIF_SN(DIF_SN1~DIF_SN3)に対応した3本の直線が図示されている。いま、音量比EN_SNが数値S Naで音量比差分値DIF_SNが数値DIF_SN1であるとすれば、閾値THを数値THaに選定することで所期の精度による認証が実現される。同図に示すように、音量比EN_SNや音量比差分値DIF_SNが増加するほど、認証を所期の精度に維持するための閾値THは増加する。

【0032】

以上の知見に基づいて、初期値設定部52は、ステップS10で算定された音量比EN_SNに対し、音量比差分値DIF_SNに応じた直線の関係を満たすように、閾値THの初期値を設定する。例えば図4に図示されるように、音量比EN_SNが数値S Naであるとすると、ステップS12で数値DIF_SN1が算定された場合には数値THaが閾値THの初期値として設定され、ステップS12で数値DIF_SN2が算定された場合には数値THbが閾値THの初期値として設定される。また、音量比差分値DIF_SNが予め設定された数値に該当しない場合には補間によって閾値THの初期値が算定される。例えば、数値DIF_SN1と数値DIF_SN2との中間の数値DIF_SNがステップS12で算定された場合には、数値DIF_SN1に対応した数値THaと数値DIF_SN2に対応した数値THbとの中間値THcが閾値THの初期値として算定される。

10

【0033】

本形態の初期値設定部52は、以上の条件を満たすように作成されたテーブルに基づいて音量比EN_SNおよび音量比差分値DIF_SNから閾値THの初期値を算定する。図5は、ステップS14にて使用されるテーブルの内容を示す概念図である。同図に示すように、別個の音量比差分値DIF_SN(DIF_SN1, DIF_SN2, DIF_SN3, ……)に対応した複数のテーブルが記憶部35に格納される。ひとつの音量比差分値DIF_SNに対応したテーブルには、音量比EN_SNと閾値THとが当該音量比差分値DIF_SNに対応した直線的な関係を満たすように、音量比EN_SNの各数値と閾値THの初期値とが対応づけられている。ステップS14において、初期値設定部52は、ステップS12で算定された音量比差分値DIF_SNに対応したひとつのテーブルを探索し、ここで探索したテーブルのうちステップS10で算定された音量比EN_SNに対応づけられた閾値THを初期値として設定する。ステップS12で算定された音量比差分値DIF_SNに対応したテーブルが存在しない場合、初期値設定部52は、当該音量比差分値DIF_SNの前後の音量比差分値DIF_SNに対応する各テーブルから音量比EN_SNに応じた閾値THを算定し、各閾値THの補間によって初期値を算定する。

20

30

【0034】

ところで、図4に例示した音量比EN_SN, 音量比差分値DIF_SNと閾値THとの関係は、例えば登録時雑音と認証時雑音との特性が同等であることを前提として決定される。しかし、実際には登録時雑音と認証時雑音との特性は相違する場合が多い。そこで、補正部54は、ステップS14で算定された閾値THの初期値を登録時雑音と認証時雑音との相関値NOISE_DIFに応じて補正する(ステップS15)。さらに詳述すると、補正部54は、相関値NOISE_DIFに応じた補正值A1を閾値THの初期値に加算することで補正後の閾値THを算定する。

40

【0035】

図6は、相関値NOISE_DIFと補正值A1との関係を示すグラフである。式(3)で算定される相関値NOISE_DIFは、登録時雑音と認証時雑音との相関に応じて「-1」から「1」までの範囲内で変動し、双方の雑音の特性が完全に合致する場合には「1」となる。登録時雑音と認証時雑音との関係が図4の関係の決定時と同等である(本形態では双方の雑音が合致する)ならば、相関値NOISE_DIFに応じて閾値THを補正する必要はない。一方、登録時雑音と認証時雑音との関係が図4の関係の決定時から乖離するほど閾値THを大きく補正すべきである。したがって、補正部54は、図6に示すように、相関値NOISE_DIFが「1」である場合に補正值A1をゼロに設定する(補正なし)とともに、相関値NOISE_DIFが「1」よりも小さいほど大きい数値となるように補正值A1を設定する。さらに詳述す

50

ると、相関値NOISE_DIFと補正值A1とが対応づけられたテーブルが記憶部35に格納され、補正部54はこのテーブルに基づいて補正值A1を決定する。

【0036】

また、図4に示した音量比EN_SNおよび音量比差分値DIF_SNと閾値THとの関係は、例えば登録音声が発声長L0にわたって継続する場合を想定して決定される。しかし、実際の初期登録における登録音声の発声長EN_SPEECH_LENは区々である。そこで、補正部56は、ステップS15における補正後の閾値THを発声長EN_SPEECH_LENに応じて補正する(ステップS16)。さらに詳述すると、補正部56は、発声長EN_SPEECH_LENに応じた補正值A2を閾値THに加算することで補正後の閾値THを算定する。補正部56による補正後の閾値THは、判定部44における距離DISTとの比較に使用される。

10

【0037】

図7は、発声長EN_SPEECH_LENと補正值A2との関係を示すグラフである。発声長EN_SPEECH_LENが図4の関係の決定時における時間長L0と同等であるならば、発声長EN_SPEECH_LENに応じて閾値THを補正する必要はない。したがって、記憶装置32に格納された発声長EN_SPEECH_LENが時間長L0と同等である場合、補正部56は補正值A2をゼロに設定する。また、発声長EN_SPEECH_LENが長いほど登録音声の発声が安定するから、特徴量EN_DATAは正当な利用者本来の基本的な声質を忠実に反映したものとなって距離DISTの正確性が向上する。したがって、他人の正当性が認証される可能性を低減するという観点から閾値THを低下させても、正当な利用者を否定する可能性が不当に上昇することはない。そこで、補正部56は、発声長EN_SPEECH_LENが時間長L0よりも長い場合には当該発声長EN_SPEECH_LENに応じた負数を補正值A2として選定し、発声長EN_SPEECH_LENが時間長L0よりも短い場合には当該発声長EN_SPEECH_LENに応じた正数を補正值A2として選定する。さらに詳述すると、発声長EN_SPEECH_LENと補正值A2とが対応づけられたテーブルが記憶部35に格納され、補正部56はこのテーブルに基づいて補正值A2を決定する。

20

【0038】

以上に説明したように、本形態においては、認証時雑音と認証音声との関係(V_SN)や登録時雑音と登録音声との関係(EN_SN)に応じて閾値THが可変に設定されるから、認証時雑音や登録時雑音の特性に拘わらず認証を所望の精度に維持することが可能である。すなわち、認証時雑音や登録時雑音に影響されることなく、FRRを低下させて利便性の向上を図りながら、FARの低下によって認証の精度を高水準に維持することが可能となる。

30

【0039】

本形態においては特に、音量比EN_SNと好適な閾値THとが音量比差分値DIF_SNに応じた直線の関係を満たすという知見に基づいて閾値THが設定される。したがって、閾値THの設定に必要な変数の個数を十分に削減しながら最適な閾値THを高精度に特定できるという利点がある。さらに、相関値NOISE_DIFや発声長EN_SPEECH_LENに応じて閾値THが補正されるから、初期登録時や認証時における音声認証装置100の環境を忠実に反映した閾値THを認証に利用することが可能となる。

【0040】

< C : 変形例 >

40

以上の形態には様々な変形を加えることができる。具体的な変形の態様を例示すれば以下の通りである。なお、以下の各態様を適宜に組み合わせてもよい。

【0041】

(1) 変形例1

以上の形態においては記憶部35に格納されたテーブルが利用される構成を例示したが、所定の演算式を利用した演算処理によって閾値THの初期値や補正值A1およびA2が選定される構成としてもよい。例えば、音量比差分値DIF_SNに応じた直線(音量比EN_SNと閾値THとの関係を定義する図4の各直線)を表わす複数の数式が記憶部35に格納され、初期値設定部52は、ステップS12で算定された音量比差分値DIF_SNに応じた数式にステップS10で算定された音量比EN_SNを代入することで閾値THを演算する。また、相関

50

値NOISE_DIFと補正值A1との関係を表わす数式に基づいて補正部54が補正值A1を算定する構成や、発声長EN_SPEECH_LENと補正值A2との関係を表わす数式に基づいて補正部56が補正值A2を算定する構成も採用される。

【0042】

(2) 変形例2

閾値THの補正の程度を決定する変数は相関値NOISE_DIFや発声長EN_SPEECH_LENに限定されない。例えば、発声長EN_SPEECH_LENに応じて閾値THを補正する構成に代えて、またはこの構成とともに、認証音声の発声長V_SPEECH_LENに応じて閾値THを補正する構成や、発声長EN_SPEECH_LENと発声長V_SPEECH_LENとの平均値に応じて閾値THを補正する構成を採用してもよい。例えば、発声長EN_SPEECH_LENと補正值A2との関係と同様に、発声長V_SPEECH_LENが長いほど閾値THが低下するように閾値THが補正される。

10

【0043】

また、以上に列挙した以外の変数に基づいて閾値THを補正してもよい。例えば、登録音声や認証音声(発声区間P2)のうち有声音と無声音との時間長の比率に応じて閾値THを補正してもよい。有声音の比率が高いほど特徴量(EN_DATA, V_DATA)は発声者の声質を忠実に反映した数値となるから距離DISTの正確性は向上する。したがって、閾値THを低下させてもFRRが不当に上昇することはない。そこで、登録音声や認証音声のうち有声音の比率が高いほど閾値THが低下するように閾値THを補正する構成が採用される。また、登録音声や認証音声において音節(モーラ)の個数が多いほど距離DISTの正確性は向上するから、例えば登録音声や認証音声の音節数が多いほど閾値THが低下するように閾値THを補正する構成としてもよい。

20

【0044】

(3) 変形例3

閾値THと各変数との関係が可変である構成も採用される。例えば、操作部10の操作に応じてテーブルを更新することで音量比EN_SNや音量比差分値DIF_SNと閾値THの初期値との関係が変更される構成としてもよい。同様に、各変数と閾値THに対する補正の程度との関係が可変である構成も採用される。例えば、相関値NOISE_DIFに対応した補正值A1や発声長EN_SPEECH_LENに応じた補正值A2は操作部10に対する操作に応じて変更され得る。これらの構成によれば、利用者の要求に応じた認証を実現することが可能となる。また、変形例1に例示したように数式の演算によって閾値THや補正值(A1, A2)が算定される構成においては、数式の内容(例えば各変数の係数)が操作部10に対する操作に応じて変更されるようにしてもよい。

30

【0045】

(4) 変形例4

以上の形態においては登録音声と認証音声との距離DISTが認証に利用される構成を例示したが、双方の音声の類似度の指標となる数値は距離DISTに限定されない。例えば、登録音声と認証音声とで特性が接近するほど数値が増加する性質の変数(指標値)に基づいて認証が実行される構成も採用される。この構成においては、各変数に対する閾値THの大小の関係が以上の形態とは逆転する。すなわち、例えば図4とは逆に、音量比EN_SNが増加するほど閾値THが減少するように閾値THが設定される。

40

【0046】

(5) 変形例5

以上の形態においては音量比EN_SNに応じて閾値THが直線的に変化する場合を例示したが、音量比EN_SNと閾値THとの関係は適宜に変更される。例えば、音量比EN_SNと閾値THとが音量比差分値DIF_SNに応じた曲線の間隔を満たすように(すなわち、音量比EN_SNに対し、音量比差分値DIF_SNに応じた曲線に沿って閾値THが変化するように)、音量比EN_SNと音量比差分値DIF_SNとに応じて閾値THが特定される構成も採用される。以上のように、本発明の好適な態様においては、音量比EN_SNと閾値THとが認証時雑音に応じて異なる関係(典型的には音量比差分値DIF_SNに応じた直線的または曲線的な関係)となるように閾値THが設定されれば足り、ひとつの音量比差分値DIF_SNに対応した音量比

50

EN_SNと閾値THとの具体的な関係は、例えば各変数の関係を導出するための試験の結果やこの結果に対する統計的な処理の如何に応じて適宜に変更される。

【0047】

(6) 変形例6

以上の形態においては正当な利用者と他人との区別に音声認証装置100が利用される場合を例示したが、正当な利用者の発声に基づいて事前に登録されたパスワードと認証音声から特定されるパスワードとの合致を判定する音声パスワード認証にも以上の形態に係る音声認証装置100を利用することが可能である。

【図面の簡単な説明】

【0048】

【図1】本発明のひとつの形態に係る音声認証装置の構成を示すブロック図である。

【図2】入力部が生成する音響信号の波形図である。

【図3】閾値設定部による処理の流れを示すフローチャートである。

【図4】音量比EN_SNおよび音量比差分値DIF_SNと閾値THとの関係を示すグラフである。

。

【図5】閾値の決定に使用されるテーブルの内容を示す概念図である。

【図6】相関値NOISE_DIFと補正值A1との関係を示すグラフである。

【図7】発声長EN_SPEECH_LENと補正值A2との関係を示すグラフである。

【図8】閾値とFARおよびFRRとの関係を示すグラフである。

【符号の説明】

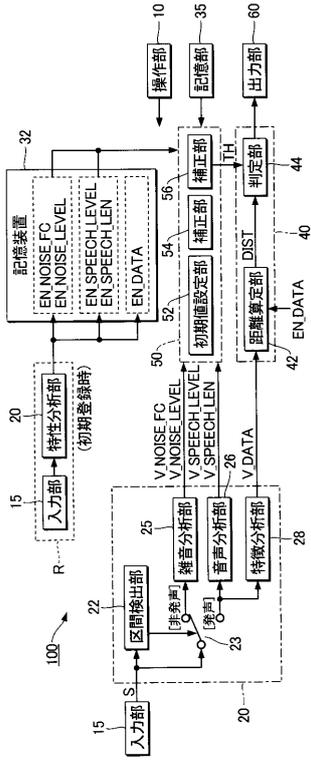
【0049】

100 …… 音声認証装置、10 …… 操作部、15 …… 入力部、20 …… 特性分析部、22 …… 区間検出部、23 …… 切換部、25 …… 雑音分析部、26 …… 音声分析部、28 …… 特徴分析部、32 …… 記憶装置、35 …… 記憶部、40 …… 認証部、42 …… 距離算定部、44 …… 判定部、50 …… 閾値設定部、52 …… 初期値設定部、54 …… 補正部、56 …… 補正部、60 …… 出力部。

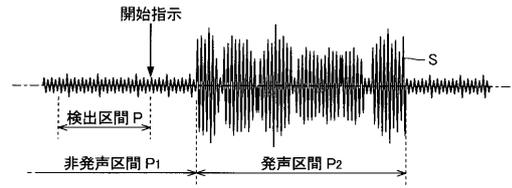
10

20

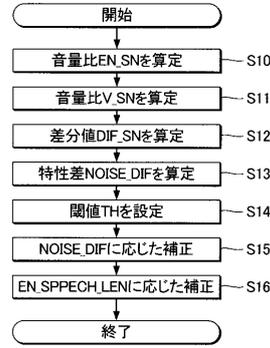
【図1】



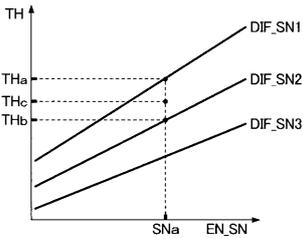
【図2】



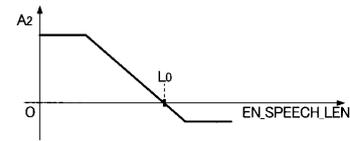
【図3】



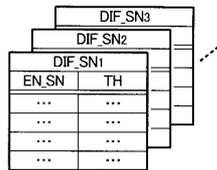
【図4】



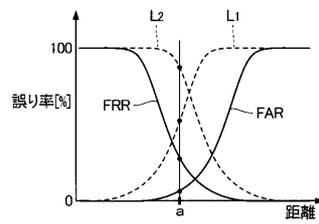
【図7】



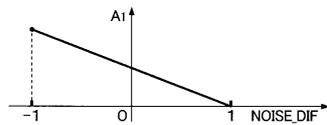
【図5】



【図8】



【図6】



フロントページの続き

審査官 菊地 陽一

- (56)参考文献 特開2003-308091(JP,A)
特開2004-109563(JP,A)
特開2002-236665(JP,A)
国際公開第2004/102527(WO,A1)
特開平10-055194(JP,A)
特開2003-316389(JP,A)