

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4049732号
(P4049732)

(45) 発行日 平成20年2月20日 (2008. 2. 20)

(24) 登録日 平成19年12月7日 (2007. 12. 7)

(51) Int. Cl.	F I
G 1 0 L 17/00 (2006. 01)	G 1 0 L 17/00 2 0 0 Z
G 1 0 L 15/06 (2006. 01)	G 1 0 L 15/06 3 1 0 S
G 1 0 L 15/02 (2006. 01)	G 1 0 L 15/02 3 0 0 C
	G 1 0 L 15/02 3 0 0 H

請求項の数 9 (全 23 頁)

(21) 出願番号	特願2003-389665 (P2003-389665)	(73) 特許権者	393031586
(22) 出願日	平成15年11月19日 (2003. 11. 19)		株式会社国際電気通信基礎技術研究所
(65) 公開番号	特開2005-148640 (P2005-148640A)		京都府相楽郡精華町光台二丁目2番地2
(43) 公開日	平成17年6月9日 (2005. 6. 9)	(74) 代理人	100064746
審査請求日	平成17年5月16日 (2005. 5. 16)		弁理士 深見 久郎
		(74) 代理人	100085132
			弁理士 森田 俊雄
		(74) 代理人	100083703
			弁理士 仲村 義平
		(74) 代理人	100096781
			弁理士 堀井 豊
		(74) 代理人	100098316
			弁理士 野田 久登
		(74) 代理人	100109162
			弁理士 酒井 将行

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音声認証プログラム

(57) 【特許請求の範囲】

【請求項 1】

認証対象者からの音声入力に基づいて、声道モデルの形状パラメータを決定するための特徴抽出手段を備え、

前記声道モデルは、

口腔に対応する第1の音響管部分と、

前記第1の音響管部分に連結し、咽頭腔に対応する第2の音響管部分と、

前記第2の音響管部分の底面に連結し、喉頭腔に対応する連結小音響管と、

前記第2の音響管部分の底面に連結し、梨状窩に対応する少なくとも1つの円錐管とを含み、

学習時において、前記特徴抽出手段により決定された前記形状パラメータを登録形状パラメータとして前記認証対象者と関連付けて記憶するための記憶手段をさらに備え、

前記特徴抽出手段は、認証時において、話者からの音声入力に基づいて、前記声道モデルの形状パラメータを認証形状パラメータとして決定し、

前記話者が登録された前記認証対象者であるか否かを特定するために、前記認証形状パラメータと前記登録形状パラメータとの比較を行なう類似度比較手段をさらに備える、音声認証装置。

【請求項 2】

前記特徴抽出手段は、

前記音声入力に基づいて、前記形状パラメータの初期値を決定する初期値決定手段と、

前記初期値に基づく声道モデルの伝達関数と前記音声入力の入力スペクトルとの差を最小化するように前記形状パラメータを修正する修正手段とを含む、請求項1記載の音声認証装置。

【請求項3】

前記第1の音響管部分は、互いに連結した複数の第1の音響管を含み、

前記第2の音響管部分は、互いに連結した複数の第2の音響管を含む、請求項2記載の音声認証装置。

【請求項4】

学習時において、認証対象者からの音声入力に基づいて、声道モデルの形状パラメータを決定するステップを備え、

前記声道モデルは、

口腔に対応する第1の音響管部分と、

前記第1の音響管部分に連結し、咽頭腔に対応する第2の音響管部分と、

前記第2の音響管部分の底面に連結し、喉頭腔に対応する連結小音響管と、

前記第2の音響管部分の底面に連結し、梨状窩に対応する少なくとも1つの円錐管とを含み、

学習時において決定された前記形状パラメータを登録形状パラメータとして前記認証対象者と関連付けて記憶装置に記憶するステップと、

認証時において、話者からの音声入力に基づいて、前記声道モデルの形状パラメータを認証形状パラメータとして決定するステップと、

前記認証形状パラメータと前記登録形状パラメータとの比較結果に基づいて、前記話者が登録された前記認証対象者であるか否かを特定するステップとをさらに備える、音声認証方法。

【請求項5】

前記声道モデルの形状パラメータを決定するステップは、

前記音声入力に基づいて、前記形状パラメータの初期値を決定するステップと、

前記初期値に基づく声道モデルの伝達関数と前記音声入力の入力スペクトルとの差を最小化するように前記形状パラメータを修正するステップとを含む、請求項4記載の音声認証方法。

【請求項6】

前記第1の音響管部分は、互いに連結した複数の第1の音響管を含み、

前記第2の音響管部分は、互いに連結した複数の第2の音響管を含む、請求項5記載の音声認証方法。

【請求項7】

コンピュータに音声認証処理を実行させるための音声認証プログラムであって、

前記音声認証処理は、

学習時において、認証対象者からの音声入力に基づいて、声道モデルの形状パラメータを決定するステップを備え、

前記声道モデルは、

口腔に対応する第1の音響管部分と、

前記第1の音響管部分に連結し、咽頭腔に対応する第2の音響管部分と、

前記第2の音響管部分の底面に連結し、喉頭腔に対応する連結小音響管と、

前記第2の音響管部分の底面に連結し、梨状窩に対応する少なくとも1つの円錐管とを含み、

学習時において決定された前記形状パラメータを登録形状パラメータとして前記認証対象者と関連付けて記憶装置に記憶するステップと、

認証時において、話者からの音声入力に基づいて、前記声道モデルの形状パラメータを認証形状パラメータとして決定するステップと、

前記認証形状パラメータと前記登録形状パラメータとの比較結果に基づいて、前記話者が登録された前記認証対象者であるか否かを特定するステップとをさらに備える、音声認

10

20

30

40

50

証プログラム。

【請求項 8】

前記声道モデルの形状パラメータを決定するステップは、
前記音声入力に基づいて、前記形状パラメータの初期値を決定するステップと、
前記初期値に基づく声道モデルの伝達関数と前記音声入力の入力スペクトルとの差を最小化するように前記形状パラメータを修正するステップとを含む、請求項 7 記載の音声認証プログラム。

【請求項 9】

前記第 1 の音響管部分は、互いに連結した複数の第 1 の音響管を含み、
前記第 2 の音響管部分は、互いに連結した複数の第 2 の音響管を含む、請求項 8 記載の音声認証プログラム。 10

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、音声の個人差に基づいて、話者を自動的に判定する話者認識を用いて、個人の認証を行なうための音声認証装置、音声認証方法および音声認証プログラムに関する。

【背景技術】

【0002】

重要な施設や部屋への部外者の入室を制限したりする場合や、システム外部からのシステム破壊等およびシステム内部からの不正アクセスを防止するためのアクセス管理や、さらには、電子商取引におけるいわゆる「なりすまし」などの不正行為の防止のために「個人認証技術」が必要とされる局面が増大している。 20

【0003】

このような個人認証には、従来から、「ユーザー ID」と「パスワード」の組み合わせや、公開鍵暗号系における「秘密鍵」等を利用した方式が採用されている。

【0004】

さらに、より個人認証の信頼度を向上させるために、指紋や虹彩などの本人の身体的特徴や行動的特徴を用いたいわゆる「バイオメトリクス」による認証技術も多く利用されている。 30

【0005】

一方、「バイオメトリクス」の一種である、音声を用いた個人認証技術に対する期待も高まっている。これは、最近の音声処理技術の発展に伴い、認証対象者の音声という、従来からの通信システムをそのまま利用可能な特徴を個人認証に用いることができれば、容易に通信関連のシステムが実現できると期待されるからである（たとえば、非特許文献 1 を参照）。

【0006】

ただし、音声による本人認証（以下、「音声認証」とよぶ）では、上記のような利点があるものの、指紋や虹彩などと比較すると、個人の身体的特性との関連性が従来方式では低く、本人認証の技術として利用するためには、さらなる精度の向上が必要である。 40

【0007】

ここで、声道の下部構造と 3 次元 MRI 動画データとの対応関係から、声道下部構造のモデル化を試みた例は存在するが（たとえば、非特許文献 2 を参照）、音声認証をいかにして行なうかについては、従来、必ずしも明らかではなかった。

【非特許文献 1】古井著、“音声による本人認証 第 1 部 音声による本人認証のしくみと技術動向”，情報処理，40 巻 11 号，1999 年，11 月

【非特許文献 2】竹本，本多，正木，島田，藤本著，“3 次元 MRI 動画データに基づく声道下部構造のモデル化”，日本音響学会講演論文集 pp. 281-282，2003 年 9 月

【発明の開示】

【発明が解決しようとする課題】

【0008】

本発明は、上述したような問題点を解決するためになされたものであって、その目的は、音声スペクトル上の複数の特徴と、その生成要因となる声道内の各部位との対応関係を組み合わせて利用することにより、入力された音声の話者を特定することが可能な音声認証装置、音声認証方法および音声認証プログラムを提供することである。

【課題を解決するための手段】

【0009】

このような目的を達成するために、本発明の1つの局面に従うと、音声認証装置であって、認証対象者からの音声入力に基づいて、声道モデルの形状パラメータを決定するための特徴抽出手段を備え、声道モデルは、口腔に対応する第1の音響管部分と、第1の音響管部分に連結し、咽頭腔に対応する第2の音響管部分と、第2の音響管部分の底面に連結し、喉頭腔に対応する連結小音響管と、第2の音響管部分の底面に連結し、梨状窩に対応する少なくとも1つの円錐管とを含み、学習時において、特徴抽出手段により決定された形状パラメータを登録形状パラメータとして認証対象者と関連付けて記憶するための記憶手段をさらに備え、特徴抽出手段は、認証時において、話者からの音声入力に基づいて、声道モデルの形状パラメータを認証形状パラメータとして決定し、話者が登録された認証対象者であるか否かを特定するために、認証形状パラメータと登録形状パラメータとの比較を行なう類似度比較手段をさらに備える。

10

【0010】

好ましくは、特徴抽出手段は、音声入力に基づいて、形状パラメータの初期値を決定する初期値決定手段と、初期値に基づく声道モデルの伝達関数と音声入力の入力スペクトルとの差を最小化するように形状パラメータを修正する修正手段とを含む。

20

【0011】

好ましくは、第1の音響管部分は、互いに連結した複数の第1の音響管を含み、第2の音響管部分は、互いに連結した複数の第2の音響管を含む。

【0012】

この発明の他の局面に従うと、音声認証方法であって、学習時において、認証対象者からの音声入力に基づいて、声道モデルの形状パラメータを決定するステップを備え、声道モデルは、口腔に対応する第1の音響管部分と、第1の音響管部分に連結し、咽頭腔に対応する第2の音響管部分と、第2の音響管部分の底面に連結し、喉頭腔に対応する連結小音響管と、第2の音響管部分の底面に連結し、梨状窩に対応する少なくとも1つの円錐管とを含み、学習時において決定された形状パラメータを登録形状パラメータとして認証対象者と関連付けて記憶装置に記憶するステップと、認証時において、話者からの音声入力に基づいて、声道モデルの形状パラメータを認証形状パラメータとして決定するステップと、認証形状パラメータと登録形状パラメータとの比較結果に基づいて、話者が登録された認証対象者であるか否かを特定するステップとをさらに備える。

30

【0013】

好ましくは、声道モデルの形状パラメータを決定するステップは、音声入力に基づいて、形状パラメータの初期値を決定するステップと、初期値に基づく声道モデルの伝達関数と音声入力の入力スペクトルとの差を最小化するように形状パラメータを修正するステップとを含む。

40

【0014】

好ましくは、第1の音響管部分は、互いに連結した複数の第1の音響管を含み、第2の音響管部分は、互いに連結した複数の第2の音響管を含む。

【0015】

この発明のさらに他の局面にしたがうと、コンピュータに音声認証処理を実行させるための音声認証プログラムであって、音声認証処理は、学習時において、認証対象者からの音声入力に基づいて、声道モデルの形状パラメータを決定するステップを備え、声道モデルは、口腔に対応する第1の音響管部分と、第1の音響管部分に連結し、咽頭腔に対応する第2の音響管部分と、第2の音響管部分の底面に連結し、喉頭腔に対応する連結小音響

50

管と、第2の音響管部分の底面に連結し、梨状窩に対応する少なくとも1つの円錐管とを含み、学習時において決定された形状パラメータを登録形状パラメータとして認証対象者と関連付けて記憶するステップと、認証時において、話者からの音声入力に基づいて、声道モデルの形状パラメータを認証形状パラメータとして決定するステップと、認証形状パラメータと登録形状パラメータとの比較結果に基づいて、話者が登録された認証対象者であるか否かを特定するステップとをさらに備える。

【0016】

好ましくは、声道モデルの形状パラメータを決定するステップは、音声入力に基づいて、形状パラメータの初期値を決定するステップと、初期値に基づく声道モデルの伝達関数と音声入力の入力スペクトルとの差を最小化するように形状パラメータを修正するステップとを含む。

10

【0017】

好ましくは、第1の音響管部分は、互いに連結した複数の第1の音響管を含み、第2の音響管部分は、互いに連結した複数の第2の音響管を含む。

【発明の効果】

【0018】

本発明に係る音声認証装置、音声認証方法および音声認証プログラムは、音声認証において、個人の身体的特性との関連性を高めて本人認証を行なうことができ、音声認証の精度を向上させることが可能である。

【発明を実施するための最良の形態】

20

【0019】

以下、図面を参照して本発明の実施の形態について説明する。

〔ハードウェア構成〕

図1は、本発明の音声認証方法および音声認証プログラムが適用される音声認証装置を用いた音声認証システム1000の一例を示す概念図である。

【0020】

図1を参照して、音声認証システム1000は、認証対象者2の発話に基づいて認証対象者2に対してアクセスを許可するか否かを判断するコンピュータ100を備える。

【0021】

すなわち、以下では、本発明の音声認証方法をアクセス権の管理に適用する場合を例にとって説明することにする。

30

【0022】

図1を参照して、このコンピュータ100は、CD-ROM (Compact Disc Read-Only Memory) 上の情報を読込むためのCD-ROMドライブ108およびフレキシブルディスク (Flexible Disk、以下FD) 116に情報を読み書きするためのFDドライブ106を備えたコンピュータ本体102と、コンピュータ本体102に接続された表示装置としてのディスプレイ104と、同じくコンピュータ本体102に接続された入力装置としてのキーボード110およびマウス112と、音声入力装置としてのマイク132と、音声出力装置としてのスピーカ134とを含む。

【0023】

40

なお、本発明の音声認証方法を入室管理などに適用する場合には、コンピュータ100は入室管理システムの一部として動作し、本人認証された場合には、ゲートの開錠処理を行なうことになる。また、本発明の音声認証方法を電子商取引等に適用する場合には、マイク132から入力された音声は、通信に適したフォーマットに変換された後に、ネットワーク310を介して、相手先のコンピュータシステム300に伝送される。相手先のコンピュータシステム300において、以下に説明するような音声認証処理を行ない、認証対象者2の本人認証を行なうことになる。

【0024】

図2は、このコンピュータ100のハードウェア構成をブロック図形式で示す図である。

50

【0025】

図2に示されるように、このコンピュータ100を構成するコンピュータ本体102は、CD-ROMドライブ108およびFDドライブ106に加えて、それぞれバスBSに接続されたCPU (Central Processing Unit) 120と、ROM (Read Only Memory) およびRAM (Random Access Memory) を含むメモリ122と、直接アクセスメモリ装置、たとえば、ハードディスク124と、マイク132またはスピーカ134とデータの授受を行なうためのインタフェース128とを含んでいる。CD-ROMドライブ108にはCD-ROM118が装着される。FDドライブ106にはFD116が装着される。

【0026】

10

なお、インタフェース128は、たとえば、相手先のコンピュータシステム300との通信を行なうために使用することもできる。

【0027】

後に説明するように、本発明の音声認証プログラムが動作するにあたっては、その動作の基礎となる情報を格納するデータベースは、ハードディスク124に格納されるものとして説明を行なう。

【0028】

なお、CD-ROM118は、コンピュータ本体に対してインストールされるプログラム等の情報を記録可能な媒体であれば、他の媒体、たとえば、DVD-ROM (Digital Versatile Disc) やメモリカードなどでもよく、その場合は、コンピュータ本体102には、これらの媒体を読取ることが可能なドライブ装置が設けられる。

20

【0029】

本発明の音声認証装置の主要部は、コンピュータハードウェアと、CPU120により実行されるソフトウェアとにより構成される。一般的にこうしたソフトウェアはCD-ROM118、FD116等の記憶媒体に格納されて流通し、CD-ROMドライブ108またはFDドライブ106等により記憶媒体から読取られてハードディスク124に一旦格納される。または、当該装置がネットワーク310に接続されている場合には、ネットワーク上のサーバから一旦ハードディスク124にコピーされる。そうしてさらにハードディスク124からメモリ122中のRAMに読出されてCPU120により実行される。なお、ネットワーク接続されている場合には、ハードディスク124に格納することなくRAMに直接ロードして実行するようにしてもよい。

30

【0030】

図1および図2に示したコンピュータのハードウェア自体およびその動作原理は一般的なものである。したがって、本発明の最も本質的な部分は、FD116、CD-ROM118、ハードディスク124等の記憶媒体に記憶されたソフトウェアである。

【0031】

なお、一般的な傾向として、コンピュータのオペレーティングシステムの一部として様々なプログラムモジュールを用意しておき、アプリケーションプログラムはこれらモジュールを所定の配列で必要な時に呼び出して処理を進める方式が一般的である。そうした場合、当該音声認証装置を実現するためのソフトウェア自体にはそうしたモジュールは含まれず、当該コンピュータでオペレーティングシステムと協働してはじめて音声認証装置が実現することになる。しかし、一般的なプラットフォームを使用する限り、そうしたモジュールを含ませたソフトウェアを流通させる必要はなく、それらモジュールを含まないソフトウェア自体およびそれらソフトウェアを記録した記録媒体（およびそれらソフトウェアがネットワーク上を流通する場合のデータ信号）が実施の形態を構成すると考えることができる。

40

【0032】

[個人性生成要因にもとづく音声認証]

図3は、音声スペクトル上の特徴と声道内の部位の対応関係を示す図である。

【0033】

50

以下に説明するとおり、本発明の音声認証装置や音声認証方法は、音声スペクトル上の複数の特徴と、その生成要因となる声道内の各部位との対応関係を組み合わせて利用することにより、入力された音声の話者を特定することを可能にするものである。

【0034】

特に、本発明では、入力された音声（母音）から話者の個人性要因を抽出することにより音声認証を行なう。このとき、音声認証において、認証に用いる音声の発声内容（キーワード）を予め決めておく「テキスト依存型」の認証や、どんな言葉を発生してもよい「テキスト独立型」の認証や、装置を使うたびに新しいキーワードを装置側から認証対象者に対して指定する「テキスト指定型」の認証のいずれにの場合にも、本発明の音声認証を適用することができる。

10

【0035】

一方、図4は、音声生成系の正中断面図を示す概念図である。

【0036】

図3および図4を参照すると、音声スペクトルの極（極大点）の分布パターンは、「声道長」に対応している。「声道」とは、声門から咽頭腔、口腔を通り唇に至る空間を指す。

【0037】

低次フォルマントは、「咽頭腔と口腔の長さ、断面積および容積の関係」に対応している。低周波側から4番目の極大点である第4フォルマントは、「喉頭腔の形状」と対応している。

20

【0038】

さらに、高周波数帯域に存在する零点（極小点）の数、周波数、バンド幅、およびその周辺の極との相対的エネルギー差は、後に詳しく説明するように、「声道下部の梨状窩の形状」と対応している。

【0039】

声道形状は、個々の人間の声質、すなわち音声の個人性を決定づける主要因である。言い換えれば、音声の個人性の主たる生成要因は、声道形状の個人差であるといえる。

【0040】

以下、図3における声道内の部位について、さらに詳しく説明する。

【0041】

（声道長）

「声道長」とは声門から口唇までの長さを指す。声道長は年齢、性別、個人の体格との相関が高い。音響管の長さが長いほどその共鳴周波数が低くなるので、声道長と音声スペクトルの極の分布パターンには対応関係がある。従って、音声からその話者の声道長を求めることができる。

30

【0042】

（咽頭腔と口腔の長さ、断面積、容積の関係）

咽頭腔と口腔の長さ、断面積、容積の関係は低次フォルマントを決定する。

【0043】

図5は、声道の2区間モデルにおける咽頭腔と口腔の断面積変化、および低次フォルマントとの理論的關係を示す図である。

40

【0044】

図5（a）は、声道を2つの区間からなる音響管で代表させた2区間モデルにおいて、口腔に比べて咽頭腔の断面積が大きい場合と、口腔に比べて咽頭腔の断面積が小さい場合とを示している。一方、図5（b）は、図5（a）に示した咽頭腔と口腔との間の断面積の關係が異なる2つの場合にそれぞれ対応する、フォルマント周波数の変化を示す図である。

【0045】

まず、図5（a）の上側のように、咽頭腔の断面積が増大すると、図5（b）の上側に示すように、第1フォルマント（F1）の低下と第2フォルマント（F2）の上昇が起こる。

50

【0046】

これに対して、図5(a)の下側のように、口腔の断面積が増大すると、第1フォルマント(F1)の上昇と第2フォルマント(F2)の低下が起こる。

【0047】

この図5に示すような関係は、核磁気共鳴画像法(MRI:Magnetic Resonance Imaging)にもとづく実測値と録音音声との間にも観測され、第1フォルマント周波数は咽頭腔断面積と、第2フォルマント周波数は口腔断面積との相関が認められる。

【0048】

図6は、咽頭腔平均面積と第1フォルマント周波数との相関関係をMRIによる実測値で示す図である。図6(a)は、複数の被験者について、MRIにより実測された咽頭腔の平均面積と母音「あ」の第1フォルマントの周波数の実測値との関係を示す。また、図6(b)は、複数の被験者について、MRIにより実測された咽頭腔の平均面積と母音「え」の第1フォルマントの周波数の実測値との関係を示す。

10

【0049】

咽頭腔の平均面積の実測値と母音の第1フォルマントの周波数の実測値との間には負の相関がみられる。

【0050】

また、図7は、口腔平均面積と第1フォルマント周波数との相関関係をMRIによる実測値で示す図である。図7(a)は、複数の被験者について、MRIにより実測された口腔の平均面積と母音「あ」の第1フォルマントの周波数の実測値との関係を示す。また、図7(b)は、複数の被験者について、MRIにより実測された口腔の平均面積と母音「え」の第1フォルマントの周波数の実測値との関係を示す。

20

【0051】

口腔の平均面積の実測値と母音の第1フォルマントの周波数の実測値との間には正の相関がみられる。

【0052】

以上の関係を用いて、低次フォルマントから咽頭腔と口腔の概形を推定することができる。

【0053】

(喉頭腔の形状)

30

「喉頭腔」とは下咽頭腔の一部を構成する細い管である。

【0054】

図8は、喉頭腔の形状を説明するための図面である。図8(a)は、MRI画像において喉頭腔を白線で囲んで示し、図8(b)は、MRI画像から得られた下咽頭腔の3次元形状をワイヤフレームで示しており、喉頭腔部分は、ワイヤフレームを太線で示すとともにグレースケールを濃くして示してある。なお、図8(b)に示すとおり、下咽頭腔には、喉頭腔と、原則としては喉頭腔の両側後部に、後に説明する梨状窩が含まれている。

【0055】

図9は、3名分の下咽頭腔の3次元形状を示す図である。図9において、図9(a1)～(a3)は、それぞれ3名の被験者についてMRI画像から得られた下咽頭腔の3次元形状をワイヤフレームで表わしたものを正面から見た図であり、図9(b1)～(b3)は、これらのワイヤフレームをそれぞれ左側から見た図である。

40

【0056】

この図に示されるように、喉頭腔の形状と大きさには個人差がある。

【0057】

図10は、図9に示した3名について、各母音(/a/, /i/, /u/, /e/, /o/)を発声しているときの下咽頭腔の各部の横断面形状を声門からの距離をパラメータとして示す図である。図10(c1)～(c3)の各々が、図9(a1)～(a3)に示した各人に対応している。

【0058】

50

図10に示すように、発声する母音が変わっても、各人において、その形状変化が極めて小さい。

【0059】

図11は、母音「え」の音声スペクトルを示す図である。図11において、第4フォルマントには、矢印を付して示す。

【0060】

喉頭腔は声道内で音響的に独立しており、ヘルムホルツ共鳴器として働く。そして、喉頭腔の形状や大きさは音声スペクトルの第4フォルマントの周波数、バンド幅、エネルギーを決定する。すなわち、喉頭腔の形態上の個人差は、第4フォルマントに現れる。

【0061】

図12は、話者ア～コの第4フォルマントの周波数を示す図である。

【0062】

図12に示すように、第4フォルマント周波数は、個人間で異なっている。従って、喉頭腔は音声の個人性生成の一要因であるといえる。

【0063】

第4フォルマントに対応するヘルムホルツ共鳴器の形状を求めることによって、話者の喉頭腔の形状を求めることが可能である。

【0064】

なお、図12には、各話者について、梨状窩による零点の周波数も記載されているが、これについては後述する。

【0065】

(梨状窩の形状)

図13は、下咽頭腔における梨状窩の位置を示す図である。図13は、MRI画像から得られた下咽頭腔の3次元形状を正面から見てワイヤフレームで示しており、梨状窩部分は、ワイヤフレームを太線で示すとともにグレースケールを濃くして示してある。

【0066】

梨状窩は下咽頭腔に、原則として左右1つずつ存在する分岐管である。前面から見ると梨状窩は、図13のような形状をしているので、この形状は円錐で近似することができる。

【0067】

図10に示したとおり、喉頭腔と同様に、梨状窩の形状、長さ、大きさには個人差があり、なおかつ発声する母音が変わってもその形状変化が極めて小さい。

【0068】

梨状窩は声道内の分岐管であるため、音声スペクトル上で零点（極小点）を発生させる。梨状窩の形状、長さ、大きさは音声スペクトルの高周波数帯域に現れる零点の数、周波数、バンド幅、その零点の周辺の極との相対的エネルギー差を決定する。

【0069】

図14は、母音「え」の音声スペクトル上の梨状窩による零点の位置を示す図である。

【0070】

図14において、梨状窩による零点には、矢印を付加している。

【0071】

また、図12には、上述のとおり話者ア～コの10名の梨状窩による零点の周波数を示している。

【0072】

図12より、梨状窩による零点の周波数には個人差があることがわかる。この周波数は梨状窩の形態の個人差に対応する。従って、梨状窩も音声の個人性生成の一要因であるといえる。

【0073】

2つの梨状窩の形状、長さ、大きさが異なる場合には零点が2つ現れ、形状、長さ、大きさが等しいかもしくは近い場合には零点は1つのみ現れる。一般に、梨状窩は左右2つ

10

20

30

40

50

存在するが、図9（a3）に示した被験者のように梨状窩が片方にのみ存在する人もいる。この場合にも零点は1つのみ現れる。

【0074】

以上のことから、梨状窩による零点に関する情報を用いれば、話者の梨状窩の形状、長さ、大きさを求めることができる。

【0075】

なお、音声スペクトル上で梨状窩の影響が表われる周波数帯域は、固定電話の周波数帯域（4 kHz以下）よりも高い。そのため、本手法を電話に利用する場合には、より広い周波数帯域を持つ携帯電話やIP電話を対象にする必要がある。

【0076】

〔声道モデルの形状パラメータの最適化による話者の登録と認証〕

本発明では、上記の個人性生成要因を組み合わせることで個人の登録と認証を行なう。

【0077】

音声から声道断面積関数を逆推定することは難しい課題の一つであるが、その理由として、従来の音声生成モデルでは前述した梨状窩と喉頭腔の共鳴現象を考慮していないために、高域スペクトルの複雑性を逆推定に取り込むことができないことがあげられる。

【0078】

図15は、本発明の音声生成モデルの概念図説明するための図である。

【0079】

すなわち、本発明では、音声の生成を、音源からの音が、主声道の共鳴と下咽頭腔の共鳴との影響を受けた結果が、音声として発声されているものとしてモデル化している。

【0080】

図15に示すようなモデルを用いることで、高域スペクトルの複雑性を逆推定に取り込むことを可能とする。

【0081】

すなわち、従来の音声生成モデルでは音声を音源と声道の線形結合で表わすのに対し、本発明における音声生成モデルでは音源と主声道共鳴のほかに下咽頭腔共鳴を加えている。このモデルに基づいて、音声スペクトルに含まれている下咽頭共鳴の成分を取り除くことにより、主声道の断面積関数を正確に推定することができる。

【0082】

具体的には、個人性パラメータの決定には下記のいずれかの方法を用いることができる。

【0083】

（第1の個人性パラメータの決定方法）

まず、第1の個人性パラメータの決定方法としては、入力された音声のスペクトルから、咽頭腔・口腔の形状パラメータ、喉頭腔・梨状窩の形状パラメータを求め、これらをそのまま個人性パラメータとして採用するという方法を用いることができる。

【0084】

（第2の個人性パラメータの決定方法）

あるいは、上記のパラメータを声道モデルに適用して、入力された音声のスペクトルと声道モデルにより計算した伝達関数が一致するよう最適化し、そのときの声道モデルのパラメータを個人性パラメータとして採用する方法を用いることも可能である。

【0085】

以下、これら2つの個人性パラメータの決定方法について、さらに詳しく説明する。

【0086】

〔第1の個人性パラメータの決定方法の詳細〕

まず、主声道共鳴と下咽頭腔共鳴とは線形関係になく相互作用があるため、音声から個人性要因を抽出するには声道モデルより得られる伝達関数と入力された音声のスペクトルとの間で誤差最小化をはかることにより、個人性パラメータを求めなければならない。この最適化には一般的な誤差最小化の手法を用いることができる。

10

20

30

40

50

【0087】

以下、声道モデルの形状パラメータを最適化する手法を説明する。

【0088】

図16は、図15で説明した声道の各部分から構成される声道モデルの外形を示す図である。この声道モデルは、基本的に口腔と喉頭腔をそれぞれ2つの音響管で近似し、これら2つの音響管が連結されているものとしている。さらに、喉頭腔の音響管の底部には、2つの円錐で表わされる梨状窩と、2つの小音響管の連結により近似される喉頭腔とが連結されているものとする。音源からの音は、喉頭腔底部からこの声道モデルに入力されるものとする。

【0089】

図17は、図16に示した3次元声道モデルの形状を特定するための各パラメータを示す図である。

【0090】

図17に示すとおり、まず、口腔に対応する音響管は長さ L_{or} で、断面の半径 R_{or} の円筒形状であり、上面側（口腔側）は開口している。一方、咽頭腔に対応する音響管は長さ L_{ph} で、断面の半径 R_{ph} の円筒形状であり、その上面は口腔に対応する音響管の下側の開口部と連結している。一方、咽頭腔に対応する音響管の下面には、その中央部に、喉頭腔に対応する連結小音響管が連結するとともに、この連結音響管の両側に梨状窩に対応する2つの円錐管が連結する。喉頭腔に対応する連結小音響管は、咽頭腔に対応する音響管の下面と連結する断面半径 R_{la1} 、長さ L_{la1} の円筒形の第1の小音響管と、この第1の小音響管の下面と連結する断面半径 R_{la2} 、長さ L_{la2} の円筒形の第2の小音響管とを備え、第2の小音響管の下側は、開口している。この第2の小音響管の下側から音源の音が声道モデルに入力される。

【0091】

（音声認証システムの機能構成）

図18は、コンピュータ100上で動作するソフトウェアにより実現される音声認証システム1000の機能構成を説明するための機能ブロック図である。

【0092】

なお、このような図18に示す音声認証システムの基本的な構成は、上述した非特許文献1に記載されたものと同様であるが、以下の説明のとおり、本発明では、話者モデルが図16および図17で示される声道モデルのパラメータの組み合わせにより表現される構成となっている。

【0093】

以下、簡単に音声認証システム1000の機能構成について、簡単に説明する。

【0094】

図18を参照して、入力される音声波は、まず、音声分析部200において、20ミリ秒程度の細かい時間ごとにスペクトル変換される。このようなスペクトルの表現方法としては、特に限定されないが、たとえば、ケプストラム（cepstrum）パラメータを用いることができる。以下では、ケプストラムパラメータのように音声スペクトルを表現するためのパラメータを「音声パラメータ」と呼ぶ。

【0095】

話者モデルの登録処理（学習処理）では、切替部204は、特徴抽出部202から話者モデル作成部206に処理をつなぐように切り替えられている。

【0096】

そこで、特徴抽出部202は、音声パラメータの時系列に基づいて、話者の特徴を表現するパラメータ、すなわち、上述した声道モデルの形状を規定する各パラメータ（以下、「声道モデル形状パラメータ」と呼ぶ）の値を抽出する。

【0097】

話者モデル作成部206は、ハードディスク124のような記憶装置に、各話者に対応する声道モデル形状パラメータとを対応付けて登録する。

【0098】

続いて、しきい値設定部210は、予め各話者の音声の変動の幅を同一話者についての複数の入力音声から調べ、本人の音声と判定するための許容限界のしきい値を決定する。

【0099】

一方、認証処理においては、切替部204は、特徴抽出部202から類似度計算部220に処理をつなぐように切り替えられている。

【0100】

したがって、認証処理においても、学習処理時と同様にして、音声分析部200と特徴抽出部202との処理により、入力音声に対応した声道モデル形状パラメータを抽出する。

10

【0101】

類似度計算部220は、特徴抽出部202により抽出された声道モデル形状パラメータと、登録されている各話者モデルとの比較を行ない、類似の度合い、たとえば、両者の距離を計算し、しきい値比較部222は、類似の度合いが予め設定されているしきい値よりも大きければ、本人の音声と判定して受理する旨の認証結果を出力し、そうでない場合は、他人の音声として判定して、拒否するあるいは棄却する認証結果を出力する。

【0102】

すなわち、本発明では、上述のとおり、話者の音声からこの声道モデルの形状パラメータを決定し、これを用いて認証を行なう。ある話者の音声に対する声道モデルの形状パラメータの決定は以下のような方法で行なう。

20

【0103】

(ある話者の音声に対する声道モデルの形状パラメータの決定)

図19は、第1の個人性パラメータの決定方法により、話者の音声に対する声道モデルの形状パラメータを決定して登録するための話者モデルの登録処理の手続きを説明するためのフローチャートである。

【0104】

図19を参照して、話者モデルの登録処理が開始されると、まず、声道長の決定が行なわれる(ステップS100)。すなわち、音声分析部200が、音声をスペクトル分析する。そして、特徴抽出部202は、一定の周波数帯域に現れる極の数に基づいて、声道長を決定する。その際、MRI計測により予め得られている標準的な声道長を参考にする。

30

【0105】

続いて、特徴抽出部202は、声道長および基本周波数からしきい値処理によって男女を判定する(ステップS102)。このようなしきい値は、予め実験的に定めておくものとする。

【0106】

次に、特徴抽出部202は、喉頭管の形状パラメータを決定する(ステップS104)。すなわち、図17のパラメータLla1、Lla2、Rla2はMRI計測による標準的な値を参考に決定するものとする。喉頭管はヘルムホルツ共鳴器とみなせるので、これら3つのパラメータと第4フォルマント周波数からRla1を決定することができる。

【0107】

次に、特徴抽出部202は、梨状窩の形状パラメータを決定する(ステップS106)。

40

【0108】

このとき、上述のとおり、梨状窩は円錐形で近似されている。梨状窩の形状パラメータの決定のために、たとえば、あらかじめ円錐形の底面の半径と高さとその円錐形により作られる零点の周波数とバンド幅の関係をテーブルにしておく。次に、音声スペクトル上で4kHz以上の周波数帯域に現れる零点の数を特定し、零点が1つであれば1つの円錐形を、零点が2つあれば2つの円錐形を用いる。そして、音声スペクトル上の零点の周波数とバンド幅からテーブル逆引きにより円錐形の底面の半径(図17のパラメータRpr1、Rpr2)と高さ(図17のLpr1、Lpr2)を決定する。

50

【0109】

続いて、特徴抽出部202は、口腔および咽頭腔の形状パラメータを決定する（ステップS108）。

【0110】

ここでは、声道長を2等分し咽頭腔と口腔からなる2区間声道モデルをつくり、低次フォルマントの分析より咽頭腔と口腔の断面積を求める。

【0111】

さらに、特徴抽出部202は、2区間声道モデルへ下咽頭腔を追加して声道モデルを完成させる（ステップS110）。すなわち、ステップS108で得られた声道モデルに下咽頭腔を加える。

10

【0112】

次に、話者モデル作成部206は、ステップS110により得られた声道モデルの形状パラメータを当該話者に関する個人性パラメータとして、記憶装置に登録する（ステップS112）。

【0113】

以上で、第1の個人性パラメータの決定方法に基づく、話者モデルの登録処理が完了する。

【0114】

[第2の個人性パラメータの決定方法の詳細]

次に、上述した第2の個人性パラメータの決定方法、および第2の個人性パラメータの決定方法に基づく、話者モデルの登録処理について説明する。

20

【0115】

（第2の個人性パラメータの決定方法の第1の例）

図19のステップS108で得られた声道モデルにおいて、咽頭腔と口腔をそれぞれさらに2等分した4区間声道モデルを作る。初期値としては、2等分した各部分は、2等分前と同じ断面積を有するものとする。その上で、この4区間声道モデルの伝達関数と入力スペクトルとの差を最小化するように4区間の形状パラメータおよび下咽頭腔の形状パラメータを最適化する。必要に応じてさらに分割数を増やし8区間声道モデルを用いることもできる。上記最適化により、分割された各部分の断面積を個別に決定する。これにより得られた声道モデルの形状パラメータを当該話者に関する個人性パラメータとする。このような形状パラメータの決定方法は、登録時（学習時）においても、認証時においても実施される。

30

【0116】

なお、咽頭腔に相当する音響管と口腔に対応する音響管を分割する数については、上述した2分割や4分割に限られず、分割した結果に対応して得られる伝達関数と入力スペクトルの差を、計算により最小化することが可能な自由度であるかぎり、咽頭腔に相当する音響管と口腔に対応する音響管とのそれぞれで他の分割数とすることも可能である。

【0117】

（第2の個人性パラメータの決定方法の第2の例）

図20は、第2の個人性パラメータの決定方法の第2の例の手続きを示すフローチャートである。

40

【0118】

まず、特徴抽出部202は、音声から声道断面積関数を求める（ステップS200）。これは例えば、いわゆるPARCOR分析を用いることで可能である。

【0119】

次に、特徴抽出部202は、図19のステップS104と同じ方法で音声スペクトルから喉頭腔の形状パラメータを求める（ステップS202）。

【0120】

続いて、特徴抽出部202は、低次フォルマントと口腔平均面積、低次フォルマントと咽頭腔平均面積との相関関係から、ステップS200で求めた声道断面積関数の口腔およ

50

び咽頭腔に相当する部分を修正する（ステップS204）。

【0121】

ステップS200で求めた声道断面積関数には分岐管が含まれない。そこで、特徴抽出部202は、図19のステップS106と同じ方法で音声スペクトルから円錐形で近似した梨状窩の形状パラメータを求める（ステップS206）。

【0122】

そして、特徴抽出部202は、以上により求めた喉頭腔、口腔、咽頭腔、梨状窩の形状パラメータを初期値として、図17に示したような声道モデルを作成する（ステップS208）。なお、PARCOR分析の分析回数に応じて、口腔および咽頭腔の分割数、すなわち精度は変化する。

10

【0123】

続いて、特徴抽出部202は、この声道モデルの伝達関数を計算し、それと音声スペクトルとの誤差が最小となるまで声道モデルの形状パラメータを修正する（ステップS210）。

【0124】

ステップS210により得られた声道モデルの形状パラメータを当該話者に関する個人性パラメータとし、記憶装置に登録する（ステップS212）。

【0125】

このような形状パラメータの決定方法も、登録時（学習時）におけるだけでなく、認証時においても実施される。

20

【0126】

以上のようにして、図18に示したような音声認証システムに話者を登録する際には、第1または第2の個人性パラメータの決定方法を用いて、音声から登録話者の個人性パラメータを決定して登録する。話者を照合する場合には、入力音声からその話者の個人性パラメータを決定し、登録話者の個人性パラメータと照合して入力音声の話者と決定する。

【0127】

このような構成により、音声認証において、個人の身体的特性との関連性を高めて本人認証を行なうことができ、音声認証の精度を向上させることが可能である。

【0128】

今回開示された実施の形態はすべての点で例示であって制限的なものではないと考えられるべきである。本発明の範囲は上記した説明ではなくて特許請求の範囲によって示され、特許請求の範囲と均等の意味および範囲内でのすべての変更が含まれることが意図される。

30

【図面の簡単な説明】

【0129】

【図1】本発明の音声認証方法を実施するための音声認証装置1000の一例を示す概念図である。

【図2】コンピュータ100のハードウェア構成をブロック図形式で示す図である。

【図3】音声スペクトル上の特徴と声道内の部位の対応関係を示す図である。

【図4】音声生成系の正中断面図を示す概念図である。

40

【図5】声道の2区間モデルにおける咽頭腔と口腔の断面積変化、および低次フォルマントとの理論的關係を示す図である。

【図6】咽頭腔平均面積と第1フォルマント周波数との相関関係をMRIによる実測値で示す図である。

【図7】口腔平均面積と第1フォルマント周波数との相関関係をMRIによる実測値で示す図である。

【図8】喉頭腔の形状を説明するための図面である。

【図9】3名分の下咽頭腔の3次元形状を示す図である。

【図10】図9に示した3名について、各母音を発声しているときの下咽頭腔の各部の横断面形状を声門からの距離をパラメータとして示す図である。

50

【図 1 1】 母音「え」の音声スペクトルを示す図である。

【図 1 2】 話者ア～コの第 4 フォルマントの周波数を示す図である。

【図 1 3】 下咽頭腔における梨状窩の位置を示す図である。

【図 1 4】 母音「え」の音声スペクトル上の梨状窩による零点の位置を示す図である。

【図 1 5】 本発明の音声生成モデルの概念図説明するための図である。

【図 1 6】 図 1 5 で説明した声道の各部分から構成される声道モデルの外形を示す図である。

【図 1 7】 図 1 6 に示した 3 次元声道モデルの形状を特定するための各パラメータを示す図である。

【図 1 8】 コンピュータ 1 0 0 上で動作するソフトウェアにより実現される音声認証システム 1 0 0 0 の機能構成を説明するための機能ブロック図である。

【図 1 9】 話者の音声に対する声道モデルの形状パラメータを決定して登録するための話者モデルの登録処理の手続きを説明するためのフローチャートである。

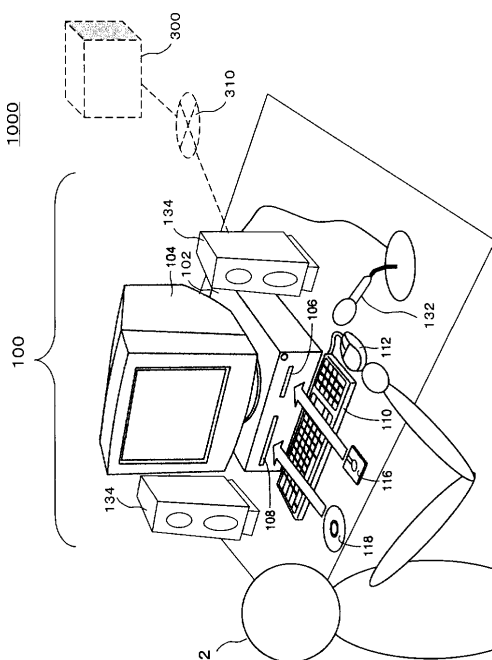
【図 2 0】 第 2 の個人性パラメータの決定方法の第 2 の例の手続きを示すフローチャートである。

【符号の説明】

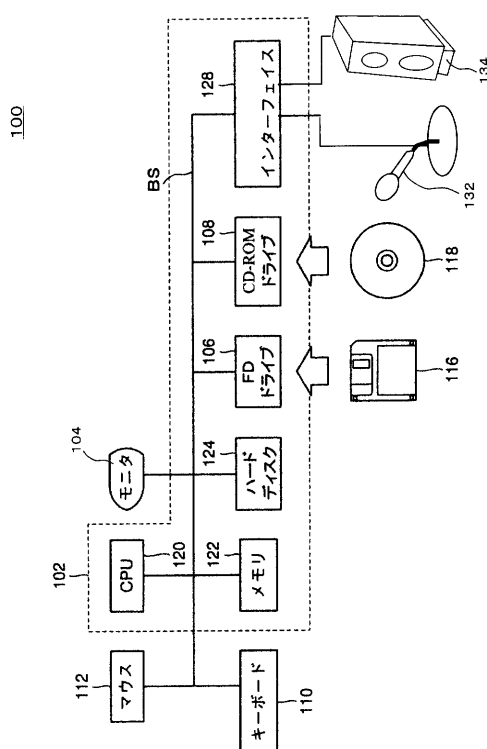
【0 1 3 0】

1 0 0 コンピュータ、1 0 2 コンピュータ本体、1 0 4 ディスプレイ、1 0 6 F D ドライブ、1 0 8 C D - R O M ドライブ、1 1 0 キーボード、1 1 2 マウス、1 1 6 フレキシブルディスク、1 1 8 C D - R O M、1 2 0 C P U、1 2 2 メモリ、1 2 4 ハードディスク、1 2 8 通信インターフェイス、1 3 2 マイク、1 3 4 スピーカ、3 0 0 相手先コンピュータ、3 1 0 ネットワーク、1 0 0 0 音声認証システム。

【図 1】



【図 2】

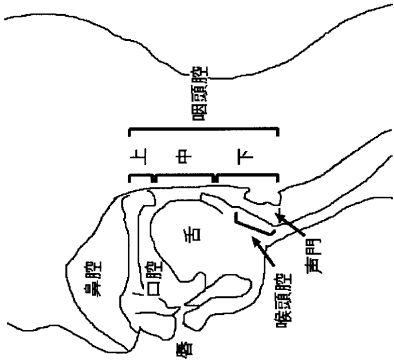


【図 3】

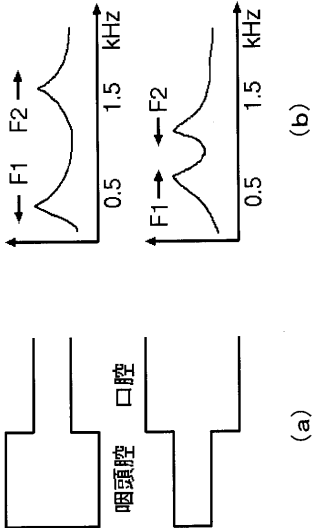
音声スペクトル上の特徴	声道内の部位
極の分布パターン	声道長
低次フォルマント	咽頭腔と口腔の長さ、断面積、容積の関係
第4フォルマント	喉頭腔の形状
高周波帯域に存在する帯域の数、周波数、バンド幅、およびその周辺の極との相対的エネルギー差	声道下部の梨状窩（リジョウカ）の形状

音声スペクトル上の特徴と声道内の部位の対応関係。

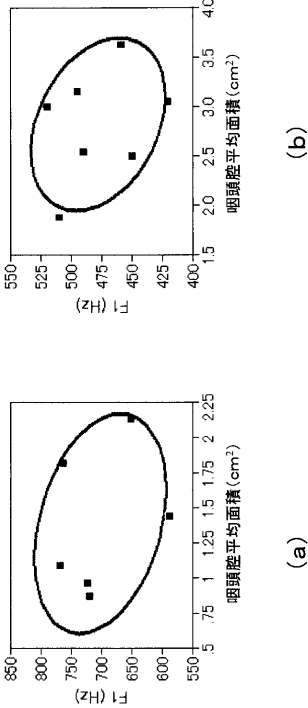
【図 4】



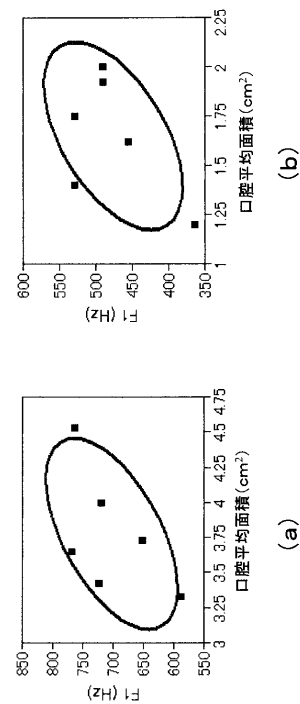
【図 5】



【図 6】



【図 7】

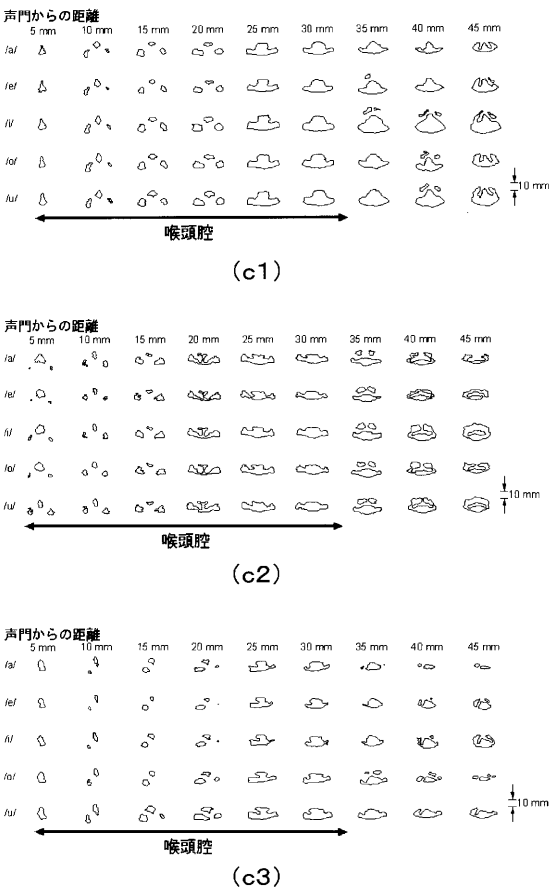


【図 12】

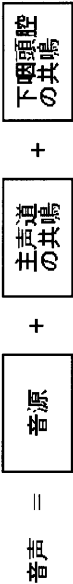
話者	F4	梨状窩による第点
ア	3862	5484
イ	3088	4212
ウ	3972	5621
エ	3475	3991
オ	3512	4489
カ	3825	4562
キ	3509	4083
ク	3770	4194
ケ	3807	4249
コ	3530	4452

成人男性 10 名の母音「え」の第 4 フォルマント周波数および梨状窩による第点の周波数 (Hz)。

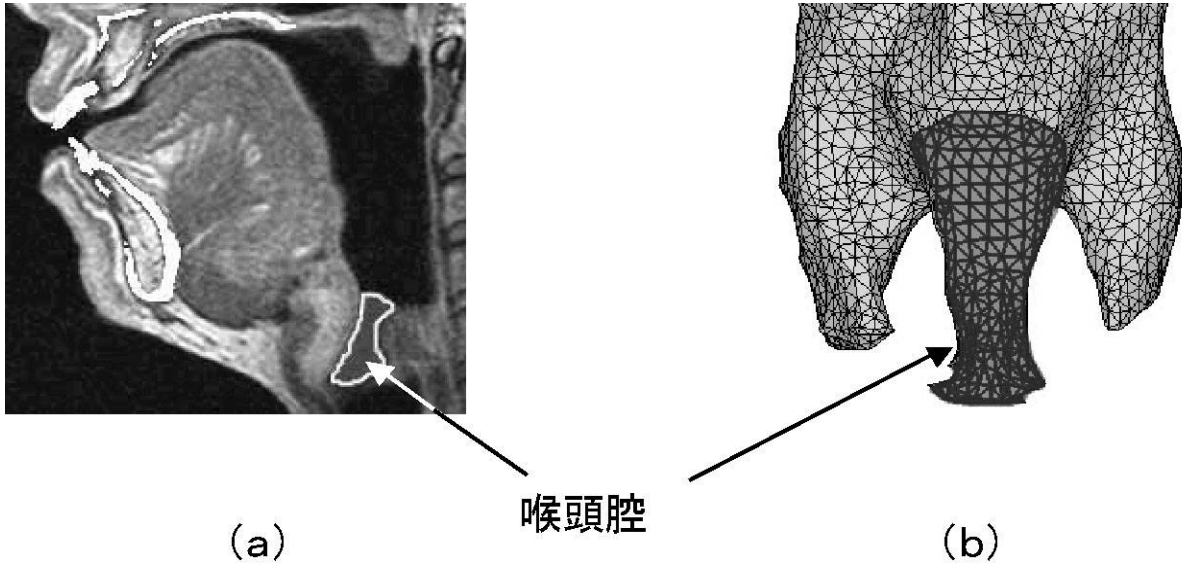
【図 10】



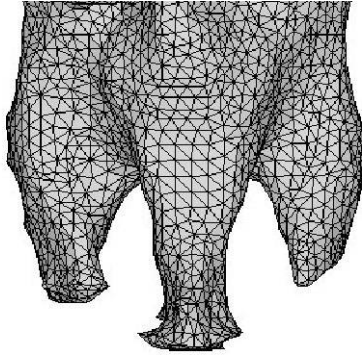
【図 15】



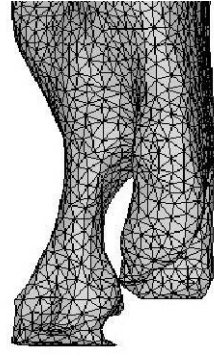
【図 8】



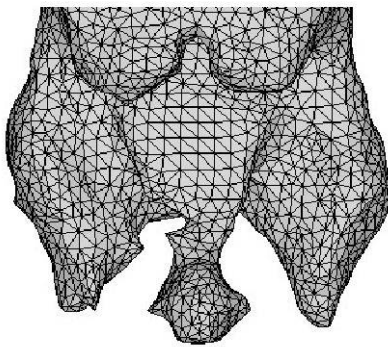
【図 9】



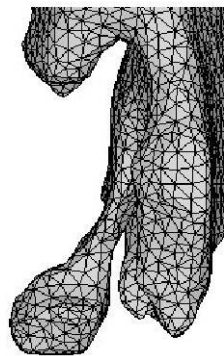
(a1)



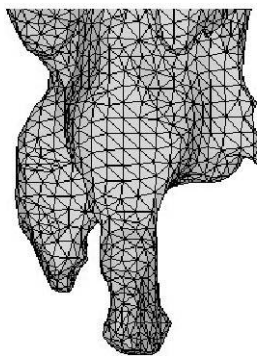
(b1)



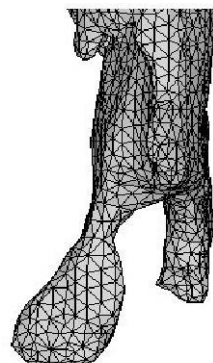
(a2)



(b2)

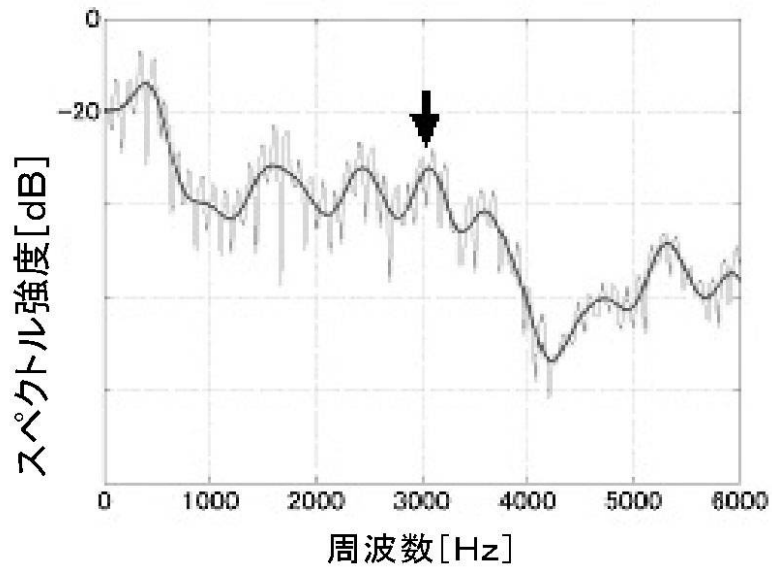


(a3)

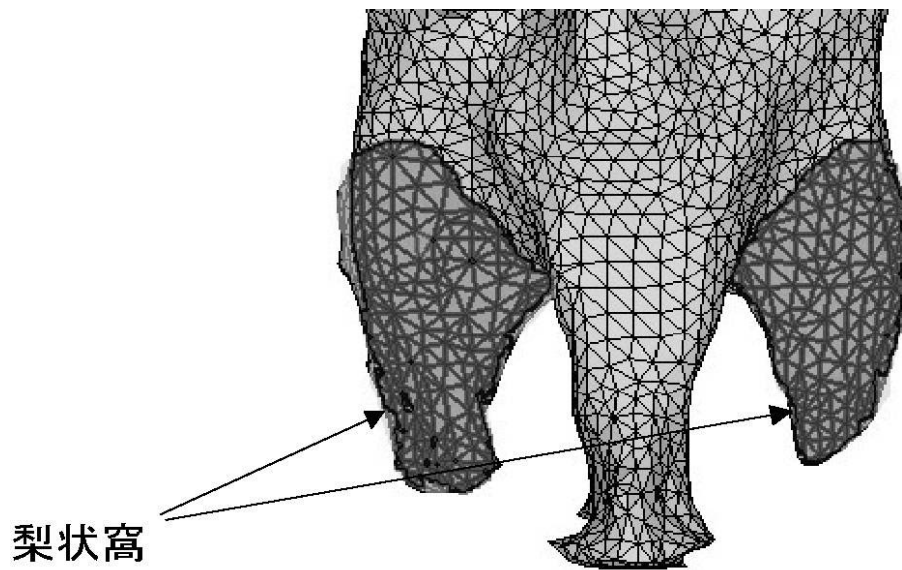


(b3)

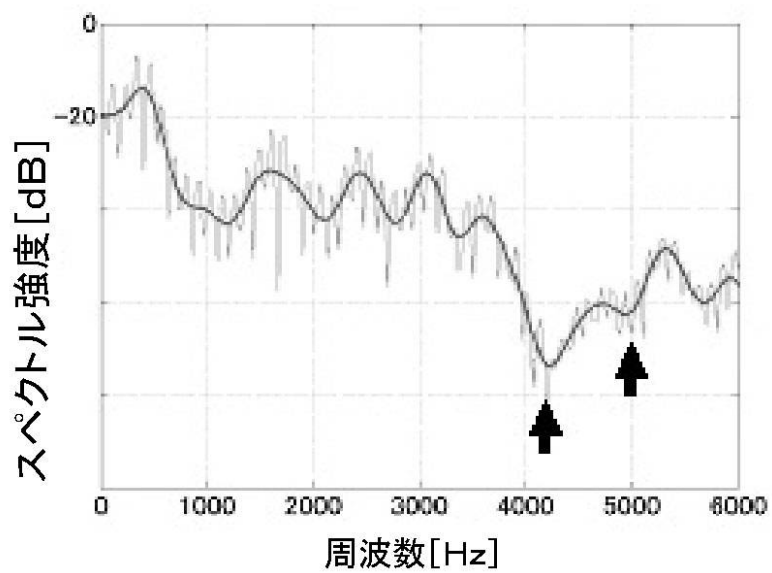
【図 1 1】



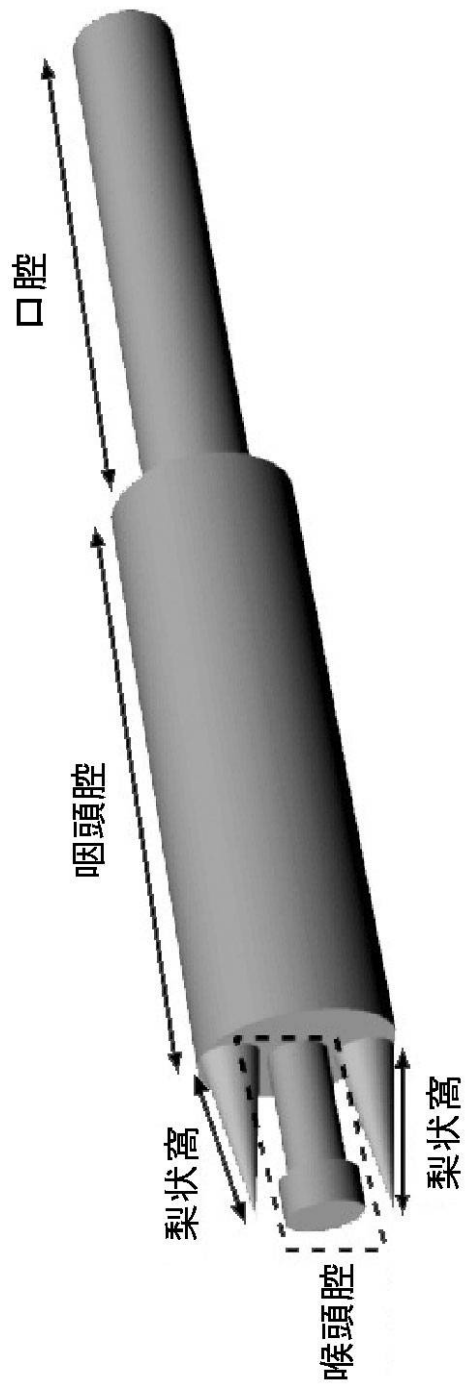
【図 1 3】



【図 1 4】



【図 16】



フロントページの続き

- (72)発明者 本多 清志
京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
- (72)発明者 北村 達也
京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
- (72)発明者 藤田 覚
京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
- (72)発明者 竹本 浩典
京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内

審査官 山下 剛史

- (56)参考文献 特表平05-508242 (JP, A)
特開平11-327592 (JP, A)
特開2002-182683 (JP, A)
党建武他, "母音発声時の音声スペクトルに対する梨状窩の影響", 電子情報通信学会技術研究報告, 1995年 5月19日, Vol.95, No.42, p.1-6

- (58)調査した分野(Int.Cl., DB名)
G10L 15/00-17/00