

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6480124号  
(P6480124)

(45) 発行日 平成31年3月6日(2019.3.6)

(24) 登録日 平成31年2月15日(2019.2.15)

(51) Int.Cl. F I  
G 1 O L 25/51 (2013.01) G 1 O L 25/51

請求項の数 11 (全 20 頁)

<p>(21) 出願番号 特願2014-166271 (P2014-166271)                  (22) 出願日 平成26年8月19日 (2014. 8. 19)                  (65) 公開番号 特開2016-42162 (P2016-42162A)                  (43) 公開日 平成28年3月31日 (2016. 3. 31)                  審査請求日 平成29年8月17日 (2017. 8. 17)</p>	<p>(73) 特許権者 504202472                  大学共同利用機関法人情報・システム研究                  機構                  東京都立川市緑町10番3号                  (74) 代理人 100103894                  弁理士 冢入 健                  (72) 発明者 山岸 順一                  東京都千代田区一ツ橋2-1-2 大学共                  同利用機関法人情報・システム研究機構                  国立情報学研究所内                  (72) 発明者 越前 功                  東京都千代田区一ツ橋2-1-2 大学共                  同利用機関法人情報・システム研究機構                  国立情報学研究所内</p>
--	---

最終頁に続く

(54) 【発明の名称】 生体検知装置、生体検知方法及びプログラム

(57) 【特許請求の範囲】

【請求項1】

話者の音声を取得する音声取得部と、  
 前記音声からポップノイズを検出するポップノイズ検出部と、  
 前記ポップノイズの検出結果に基づいて、前記話者が生体であるか否かを判断する判断部と、を有する  
 生体検知装置。

【請求項2】

前記音声取得部は、前記音声を第1の音声として取得する第1のマイクロフォンと、前記第1の音声に比較して前記ポップノイズが軽減された前記音声を第2の音声として取得する第2のマイクロフォンと、を含み、  
 前記ポップノイズ検出部は、前記第1の音声と前記第2の音声との差分を用いて、前記ポップノイズを検出する  
 請求項1記載の生体検知装置。

【請求項3】

前記第2のマイクロフォンは、前記ポップノイズを軽減するための覆いを備える  
 請求項2記載の生体検知装置。

【請求項4】

前記第1のマイクロフォンと前記第2のマイクロフォンとは、ある空間において、離れた異なる位置に配置される

10

20

請求項 2 記載の生体検知装置。

【請求項 5】

前記音声取得部は、前記音声を第 1 の音声として取得する第 1 のマイクロフォンと、前記第 1 の音声を出力するスピーカと、前記スピーカの出力する前記第 1 の音声を第 2 の音声として取得する第 2 のマイクロフォンと、を含み、

前記ポップノイズ検出部は、前記第 1 の音声と前記第 2 の音声との差分を用いて、前記ポップノイズを検出する

請求項 1 記載の生体検知装置。

【請求項 6】

前記音声取得部は、前記音声を取得するマイクロフォンを含み、

前記ポップノイズ検出部は、ローパスフィルタを用いて、前記音声から前記ポップノイズを検出する

請求項 1 記載の生体検知装置。

【請求項 7】

前記音声から複数の音素を識別する音素識別部をさらに有し、

前記ポップノイズ検出部は、前記音声を前記音素の時間長で区切ってなる複数のセグメント夫々について、前記ポップノイズの検出を行う

請求項 1 乃至 6 いずれか 1 項記載の生体検知装置。

【請求項 8】

前記判断部はさらに、前記セグメント夫々について、前記音素と、前記ポップノイズの検出結果と、の対応関係の妥当性を判断し、前記妥当性の判断結果に基づいて前記話者が生体であるか否かを判断する

請求項 7 記載の生体検知装置。

【請求項 9】

前記ポップノイズ検出部は、ポップノイズ及び非ポップノイズ夫々のモデルを予め学習した識別器を含む

請求項 1 乃至 8 いずれか 1 項記載の生体検知装置。

【請求項 10】

話者の音声を取得する音声取得ステップと、

前記音声からポップノイズを検出するポップノイズ検出ステップと、

前記ポップノイズの検出結果に基づいて、前記話者が生体であるか否かを判断する判断ステップと、を有する

生体検知方法。

【請求項 11】

コンピュータに、請求項 10 記載の方法を実行させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は生体検知装置、生体検知方法及びプログラムに関し、特に声を用いた生体検知技術に関する。

【背景技術】

【0002】

近年、個人認証技術のひとつとして、人の身体的特徴や行動に基づく生体認証技術が認知されるようになった。生体認証においては、個人の声、指紋、網膜、静脈等の身体的特徴が当該個人を認証するために利用される。なかでも、声の特徴に基づく個人認証技術である話者照合は、マイクロフォン等の汎用機器を用いて認証システムを構成可能であること、話者に練習を要しないこと、及び秘書機能アプリケーション等の普及に伴い機械に話しかけることに抵抗がなくなるつつあること等の要因により、今後さらなる普及が見込まれている。

【0003】

10

20

30

40

50

しかしながら、話者の音声に酷似した音声を機械的に生成することにより、話者を詐称し、話者照合システムを破ることが可能な種々の手法も発見されている。例えば、非特許文献1には、テキストデータに基づいて合成された音声により、話者照合システムを欺くことができる場合があることが示されている。また、非特許文献2には、ある声を特定の話者の声に似せて変換する声質変換技術により、話者照合システムを欺くことができる場合があることが示されている。

【0004】

一方、これらのなりすまし音声を見破るための技術も種々提案されている。非特許文献3及び4は、合成又は声質変換された音声と自然音声とでは音声パラメータの変化の様相が異なることに基づき、なりすまし音声を検出する手法を開示している。

10

【先行技術文献】

【非特許文献】

【0005】

【非特許文献1】Lindberg J.他, "Vulnerability in speaker verification - a study of technical impostor techniques", Proc. European Conference on Speech Communication and Technology (Eurospeech), 1999

【非特許文献2】Kinnunen, T., Wu, Z.Z., Lee, K.A., Sedlak, F., Chng, E.S., Li, H., 2012. Vulnerability of speaker verification systems against voice conversion spoofing attacks: The case of telephone speech, in: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).

20

【非特許文献3】Sato T.他, "A robust speaker verification system against imposture using an HMM-based speech synthesis system", Proc. European Conference on Speech Communication and Technology (Eurospeech), 2001

30

【非特許文献4】Wu, Z., Chng, E.S., Li, H., Detecting converted speech and natural speech for anti-spoofing attack in speaker recognition, in: Proc. Interspeech 2012.

【発明の概要】

【発明が解決しようとする課題】

【0006】

非特許文献3及び4が開示するなりすまし検知技術は、いずれも Transmission point、すなわち音声波形から特徴量を抽出する段階におけるものである(図6)。これらの検知技術では、音声波形の特徴に基づいてなりすまし音声と自然音声を判別することにより、なりすましを検知する。しかしながら、音声合成技術や声質変換技術は時とともに精度が向上し、自然音声との差異は縮小している。そのため、Transmission pointにおけるなりすまし検知技術は、かかる技術向上に応じた改善が絶えず求められるという問題がある。

40

【0007】

さらに、Transmission pointにおけるなりすまし検知技術は、そこにいるはずの話者が本当に生きている人なのかを検知するもの(生体検知技術)ではない。そのため、なりすましに対する抜本的な解決策にはならないという問題がある。

【0008】

50

本発明は、このような問題点を解決するためになされたものであり、声による生体検知が可能な生体検知装置、生体検知方法及びプログラムを提供することを目的とする。

【0009】

その他の課題と新規な特徴は、本明細書の記述及び添付図面から明らかになるであろう。

【課題を解決するための手段】

【0010】

本発明に係る生体検知装置は、話者の音声を取得する音声取得部と、前記音声からポップノイズを検出するポップノイズ検出部と、前記ポップノイズの検出結果に基づいて、前記話者が生体であるか否かを判断する判断部と、を有するものである。

10

【0011】

また、本発明に係る生体検知方法は、話者の音声を取得する音声取得ステップと、前記音声からポップノイズを検出するポップノイズ検出ステップと、前記ポップノイズの検出結果に基づいて、前記話者が生体であるか否かを判断する判断ステップと、を有するものである。

【0012】

また、本発明に係るプログラムは、コンピュータに上記方法を実行させるためのプログラムである。

【発明の効果】

【0013】

本発明により、声による生体検知が可能な生体検知装置、生体検知方法及びプログラムを提供することができる。

20

【図面の簡単な説明】

【0014】

【図1】本発明の実施の形態にかかる生体検知装置100の構成を示す図である。

【図2】本発明の実施の形態にかかる生体検知装置100の動作を示す図である。

【図3】本発明の実施の形態にかかる生体検知装置100の動作を示す図である。

【図4】本発明の実施の形態3にかかる実験結果を示す図である。

【図5】本発明の実施の形態にかかる生体検知装置100の動作を示す図である。

【図6】一般的な話者認識システムの構成を示す図である。

30

【図7】本発明の実施の形態にかかる音声取得部110の例を示す図である。

【図8】本発明の実施の形態にかかる音声取得部110の例を示す図である。

【図9】本発明の実施の形態にかかる音声取得部110の例を示す図である。

【図10】本発明の実施の形態におけるポップノイズ検出部130の処理を示す図である。

【図11】本発明の実施の形態におけるポップノイズ検出部130の処理を示す図である。

【図12】本発明の実施の形態におけるポップノイズ検出部130の処理を示す図である。

【図13】本発明の実施の形態におけるポップノイズ検出部130の処理を示す図である。

40

【図14】本発明の実施の形態におけるポップノイズ検出部130の処理を示す図である。

【図15】本発明の実施の形態における判断部150の処理を示す図である。

【図16】本発明の実施の形態で用いるマイクロフォンの例を示す図である。

【図17】本発明の実施の形態5にかかる実験結果を示す図である。

【発明を実施するための形態】

【0015】

はじめに、本発明の理解を容易にするため、従来となりすまし検知手法と比較しつつ、本発明に係る生体検知手法の特徴について説明する。

50

## 【0016】

図6は、一般的な話者認識システムの概略を示す図である。従来のTransmission pointにおけるなりすまし検知とは異なり、本発明に係る手法は、Microphone pointすなわちマイクロフォンによって音声を取得する段階において生体検知を行う。

## 【0017】

声による生体検知を行うためには、生きている人間には可能であって、かつ装置等による再生音声では再現不可能な特徴に着目する必要がある。そこで、発明者は、人が発声する際のマイクロフォンへの息のかかり方に着目した。人は、発声する際、音と同時に息を吐出する。そして、マイクロフォンは、音声だけでなく息も拾うことができる。マイクロフォンが息を大量に拾うと、ポップノイズと呼ばれる独特のノイズが発生することが知られている。ポップノイズは、音声に対してはノイズであるが、観点を換えれば、その音声を発しているのが生きた人間であることを証明する情報でもある。ポップノイズは原理的にスピーカでは再現できないからである。

10

## 【0018】

本発明はこうした知見に鑑み、Microphone pointにおいて取得した音声からポップノイズを検出し、その検出結果を利用して話者の生体検知を行うことを大きな特徴とするものである。以下、本発明を適用した具体的な実施の形態について、図面を参照しながら詳細に説明する。

## 【0019】

つぎに、図1を用いて、本発明の実施の形態にかかる生体検知装置100の基本的な構成について説明する。

20

## 【0020】

生体検知装置100は、音声取得部110、ポップノイズ検出部130、判断部150を有する。生体検知装置100は、典型的には、制御プログラム等を格納する記憶装置、制御プログラムに基づいて各種処理を実行する制御装置、及び外部デバイスとの間で情報を入出力する入出力装置等を有する情報処理装置である。音声取得部110、ポップノイズ検出部130、判断部150は、上述のハードウェアと制御プログラムとを用いて、論理的な処理手段として実現される。

## 【0021】

音声取得部110は、話者の発する声を取得する機能を有する。典型的には、音声をアナログ信号に変換するマイクロフォンと、アナログ信号をデジタル信号に変換してポップノイズ検出部130に出力する変換部と、を含む。

30

## 【0022】

音声取得部110は、単数のマイクロフォンにより構成される場合と、複数のマイクロフォンを含む場合とがある。マイクロフォンが単数である場合は、マイクロフォンは話者の声及び息の双方を収録する。マイクロフォンが複数である場合は、一方のマイクロフォン(第1のマイクロフォンと称する)は話者の声及び息の双方を収録する。他方のマイクロフォン(第2のマイクロフォンと称する)は、可能な限り話者の声のみを収録し、息については減衰させるなどして収録が抑制されるよう構成される。

40

## 【0023】

複数のマイクロフォンを利用する場合の構成例としては、例えば、第2のマイクロフォンにのみマイクカバーやポップフィルタを設ける手法、及び第1のマイクロフォンと第2のマイクロフォンとを異なる空間に設置する手法(マイクロフォンアレイ)がある。

## 【0024】

ポップノイズ検出部130は、音声取得部110が出力する音声信号(デジタル信号)から、話者が発する息をマイクロフォンが拾う時に発生するポップノイズ成分を検出する処理を行う。ポップノイズ検出部130は、分離モジュール131、特徴量化モジュール132、音素アライメントモジュール133、及び識別モジュール134を含み得る。

## 【0025】

50

分離モジュール131は、音声取得部110が出力する音声信号から、ポップノイズ成分と疑われる音声信号を分離する処理を行う。すなわち、分離モジュール131は音声信号処理を行うモジュールである。音声取得部110が単数のマイクロフォンにより構成される場合、分離モジュール131は、音声成分とポップノイズ成分とが混じった1つの音声信号から、ポップノイズ成分と疑われる音声信号を分離する。具体的な分離手法としては、例えばローパスフィルタ（スムージング）、及び当該音声信号をスピーカ出力し空気伝搬したものを再収録する手法等がある。本発明の実施にあたっては、必要に応じこれらの手法のうち任意のもの、あるいは他の同等の手法を採用することができる。

【0026】

一方、音声取得部110が複数のマイクロフォンにより構成される場合、分離モジュール131は、音声成分とポップノイズ成分との双方を含む第1のマイクロフォンの出力信号（第1の音声という）と、音声成分を含むがポップノイズ成分は抑制されている第2のマイクロフォンの出力信号（第2の音声という）とを利用して、ポップノイズ成分と疑われる音声信号を分離する処理を行う。具体的な分離手法としては、例えば線形フィルタ、スペクトルサブトラクション等がある。本発明の実施にあたっては、必要に応じこれらの手法のうち任意のもの、あるいは他の同等の手法を採用することができる。

【0027】

特徴量化モジュール132は、分離モジュール131により分離された、ポップノイズ成分と疑われる音声信号から、特徴量を抽出する処理を行う。音声信号は時系列情報であるが、後段の識別モジュール134においては時系列情報を扱いにくい。そのため、ここで特徴量化モジュール132が、時系列情報を周波数に基づく情報に変換する。具体的な特徴量化手法としては、例えばMFCC（Mel-Frequency Cepstrum Coefficients）、MFCCの低周波数域への拡張、MFCCの低周波数域への拡張に加えて次元圧縮を行う手法等がある。本発明の実施にあたっては、必要に応じこれらの手法のうち任意のもの、あるいは他の同等の手法を採用することができる。なお、特徴量化モジュール132の採用は任意であり、存在しなくとも本発明の最低限の目的を達成することは可能である。

【0028】

音素アライメントモジュール133は、特徴量化された音声信号を音素単位（セグメント）に分割する処理を行う。この処理を行うことにより、音素とポップノイズ発生箇所との対応関係の検証が可能となるため、風などによるポップノイズ類似の成分の影響を排除でき、より高精度なポップノイズ認識を行えるという利点がある。具体的な手法としては、例えば外部の自動音声認識器（ASR：Automatic Speech Recognition）等が提供する音素アライメント機能を利用する手法、予めセグメント境界が既知である参照音声と、入力音声との音響的比較（DTW：Dynamic Time Warping）により、入力音声のセグメントを特定する手法等がある。本発明の実施にあたっては、必要に応じこれらの手法のうち任意のもの、あるいは他の同等の手法を採用することができる。なお、後段の識別モジュール134において、HMM（Hidden Markov Model）等それ自身が音素境界を認識し得るモデルを用いる場合には、音素アライメントモジュール133に相当する構成を格別設ける必要はない。なお、音素アライメントモジュール133の採用は任意であり、存在しなくとも本発明の最低限の目的を達成することは可能である。

【0029】

ポップノイズ識別モジュール134は、特徴量化モジュール132が抽出した特徴量、又は、音素アライメントモジュール133がセグメント化した特徴量を入力し、それがポップノイズであるか否かを識別する処理を行う。具体的な識別手法としては、例えばGMM（Gaussian mixture model）等の統計処理モデル、SVM（Support vector machine）等の機械学習処理モデル、HMM、及び線形識別モデル等の利用が考えられる。音素アライメントを行う場合、識別モジュール134はセグメント単位でポップノイズの有無を識別する。一方、音素アライメントを行わ

10

20

30

40

50

ない場合は、識別モジュール 134 はセグメントに分割されていない、例えば文章単位でポップノイズ成分の有無を識別する。いずれの場合でも、識別モジュール 134 はポップノイズ成分を含まない音声成分の特徴量を予め学習しておくことにより、いわば異常値であるポップノイズ成分が入力されたときにそれを検出することができる。本発明の実施にあたっては、必要に応じこれらの手法のうち任意のもの、あるいは他の同等の手法を採用することができる。

【0030】

上述の分離モジュール 131、特徴量化モジュール 132、音素アライメントモジュール 133 及び識別モジュール 134 は、任意の組み合わせで利用することができる。

【0031】

判断部 150 は、ポップノイズ検出部 130 によるポップノイズの検出結果に基づいて、音声取得部 110 が取得した音声の話者が生きた人間であるか否かを判断する処理を行う。

【0032】

<実施の形態 1>

実施の形態 1 は、音声取得部 110 が複数のマイクロフォンを備え、一方のマイクロフォンにマイクカバー又はポップフィルタ等を備えた場合の本発明の構成例である。なお、本実施の形態においては、音素アライメントモジュール 133 は採用しない。

【0033】

音声取得部 110 は、複数のマイクロフォンを含む。一方のマイクロフォン（第 1 のマイクロフォン）は、話者の声や息を可能な限りそのまま拾うことを目的とする。典型的には、マイクカバーやポップフィルタを備えていないマイクロフォンである。あるいは、第 1 のマイクロフォンは、風など周囲の影響を軽減しつつ、発話に伴うポップノイズは取得できるよう、穴あきウィンドスクリーンを備えるものであっても良い（図 16）。

【0034】

他方のマイクロフォン（第 2 のマイクロフォン）は、話者の息を拾うことを抑制し、可能な限り話者の声だけを拾うことを目的とする。典型的には、マイクカバーやポップフィルタを備えたマイクロフォンである。一般に、マイクカバーやポップフィルタはスポンジやネット等で作られており、これらで振動膜を覆うことで、息が振動膜を直接振動させることにより発生するポップノイズを軽減することができる。

【0035】

図 7 乃至図 9 に、音声取得部 110 の構成例を示す。図 7 はステレオマイクを用いた構成例、図 8 は複数のコンデンサマイクを用いた構成例、図 9 は複数のヘッドセットマイクを用いた構成例である。これらはいずれも、マイクカバーを備えない第 1 のマイクロフォン、マイクカバーを備えた第 2 のマイクロフォンを、それぞれ 1 つ以上含んでいる。

【0036】

また、音声取得部 110 は、アレイマイクなど 3 つ以上のマイクロフォンを含むものであってもよい。この場合、アレイマイクのうち一部のマイクロフォンについてはマイクカバーを設けずに第 1 のマイクロフォンとして扱い、その余のマイクロフォンについてはマイクカバーを設けて第 2 のマイクロフォンとして扱う。

【0037】

つづいて、図 2 及び図 3 のフローチャートを用いて、実施の形態 1 にかかる生体検知装置 100 の動作について説明する。

【0038】

S101：音声取得

音声取得部 110 が、話者の音声を取得する。本実施の形態では、音声取得部 110 は第 1 のマイクロフォン及び第 2 のマイクロフォンを有しており、それぞれのマイクロフォンが、話者が発した同一の音声を同時に取得するものとする。

【0039】

音声取得部 110 の変換部は、第 1 のマイクロフォン及び第 2 のマイクロフォンが取得

10

20

30

40

50

した音声をそれぞれ別のデジタル信号に変換する。ここで、第1のマイクロフォン由来の音声データを第1の音声、第2のマイクロフォン由来の音声データを第2の音声と称する。変換部は、第1の音声及び第2の音声をポップノイズ検出部130に対して出力する。

【0040】

S104：ポップノイズ検出

まず、ポップノイズ検出部130の分離モジュール131は、第1の音声と第2の音声とを比較し、第1の音声のみに含まれる周波数を分離し、差分信号として出力する。これにより、ポップノイズである可能性のある音声信号を抽出できる。差分信号の抽出には、例えば線形フィルタ、スペクトルサブトラクションのほか、Independent component analysis、Independent Vector Analysis、Blind source separation等の公知の手法を利用できる。また、ローパスフィルタによる低域の変動を検出する手法、ブラインド信号源分離手法などを利用しても良い。

【0041】

図10は、第1の音声及び第2の音声を重ね合わせた状態を示している。また、図11は、第1の音声と第2の音声との差分信号を示している。ここで、横軸（時間軸）の1.0（単位：秒）近辺に突出している振幅成分（単位：dB）は、話者の声にあたる音声信号に対しては異常値と考えられ、ポップノイズである可能性がある。

【0042】

つぎに、ポップノイズ検出部130の特徴量化モジュール132は、差分信号から特徴量を抽出する。換言すれば、時系列情報である差分信号から周波数に基づく情報である特徴量を求める。特徴量の抽出は、例えばMFCC（Mel-Frequency Cepstrum Coefficients）等の公知の手法により実施可能である。又は、MFCCは70Hz以上の音声を識別可能に設計されているところ、識別可能領域を70Hz未満の周波数領域にも拡張する手法も採用し得る。一般に、ポップノイズは通常の音声に比べ低周波数であるためである。さらには、このように周波数領域を拡張すると特徴量の次元が増え、概して識別性能が低下することから、特徴量の次元圧縮を併用する手法を採用しても良い。これにより、MFCCにおいて低周波数領域に対応しつつ識別性能を維持することができる。

【0043】

ポップノイズ検出部130の識別モジュール134は、抽出した差分信号の特徴量を認識器に入力する。ここで、識別モジュール134は、差分信号の特徴量を任意のセグメントに分割して識別子に入力することとしても良い。なお、ここでいうセグメントは、音素アライメントモジュール133が扱うような音素単位のセグメントである必要はない。

【0044】

例えば、ポップノイズ検出部130は、ポップノイズを含む文章と、ポップノイズを含まない文章と、をそれぞれ用いて予め学習した2つの識別器を有していても良い。識別モジュール134は、これらの識別器に差分信号の特徴量を入力することにより、それぞれ尤度を出力として得る。ここで、ポップノイズを含む文章で学習した識別器が出力する尤度と、ポップノイズを含まない文章で学習した識別器が出力する尤度との間に有意な差があれば、差分信号はポップノイズを含むと判定できる。2つの尤度の差がどの程度であるときにポップノイズが存在するものと判定するかについては、適切な閾値を予め決定しておくことができる。

【0045】

識別器としては、例えば2クラスのパターン識別器であるSVMのほか、GMM、HMM等、公知の構成を適宜採用できる。なお、識別器による判定を行う場合には、予め識別器にポップノイズを含む文章及びポップノイズを含まない文章夫々のモデルを学習させる工程が必要となる。この工程については後述する。

【0046】

また、ポップノイズ成分の検出は音声区間検出（VAD）と枠組みが近いので、話者照

10

20

30

40

50



合の分野だけでなくVADの分野で使われる任意の公知の手法を利用しても良い。

【0047】

ポップノイズ検出部130の識別モジュール134は、差分信号をポップノイズと識別した場合、ポップノイズを検出した旨判断部150に出力する。判断部150は、ポップノイズの検出結果に基づいて、話者が生きた人間であるか否かを判断する。典型的には、ポップノイズが検出された場合には話者は生体であると判断し、S105に遷移する。一方、ポップノイズが検出されなかった場合には話者は生体でないものと判断し、S106に遷移する。

【0048】

S105：話者照合

話者が生体であると判断された場合、生体検知装置100は任意の手法を用いた話者照合フェーズに移行することができる。話者照合については種々の手法が既知であるため、ここでは詳細な説明を省略する。なお、好ましくは、話者照合に際してはポップノイズ成分が比較的少ない第2の音声を用いることができる。

【0049】

S106：詐称音声として棄却

話者が生体でないと判断された場合、音声取得部110が取得した音声は人が発したのではなく、例えば合成や声質変換による音声である蓋然性が高い。よって、生体検知装置100はこれを詐称音声と判断し、話者照合を行うことなく棄却する。すなわち、エラー処理や終了処理等を行う。

【0050】

ここで、図3を用いて、識別器によるポップノイズ判定を行う場合に必要な、事前学習工程について説明する。ここでは、学習器及び識別器としてSVMを用いる場合を例として説明する。

【0051】

S201：音声取得

S101と同様に、音声取得部110が、話者の音声を取得する。

【0052】

S202：ポップノイズ/非ポップノイズモデルの学習

まず、S104と同様に、ポップノイズ検出部130の分離モジュール131が、第1の音声と第2の音声とを比較し、第1の音声のみに含まれる周波数を分離し、差分信号として出力する。次いで、特徴量化モジュール132が、差分信号を特徴量化する。そして、識別モジュール134は、差分信号の特徴量と、それがポップノイズである旨を示す教師信号と、を共に一方の学習器に入力する。また、特徴量化モジュール132は、ポップノイズ成分を含まない第2の音声を特徴量化する。そして、識別モジュール134は、第2の音声の特徴量と、それが非ポップノイズである旨を示す教師信号と、を共に他方の学習器に入力する。すなわち、本実施の形態では、ポップノイズ/非ポップノイズそれぞれの尤度を判定する学習器を1つずつ、合計2つ生成する。

【0053】

学習器としてSVM又はGMMを用いる場合は、差分信号の特徴量のうちポップノイズ成分にあたるセグメントを事前に切り出しておき、切り出されたセグメントを学習器に入力することが好ましい。一方、学習器としてHMMを用いる場合は、モデル自体が音素境界を自動的に認識する機能を有するため、上述のような切り出し処理は特段不要である。

【0054】

S201乃至S202に係る処理を複数回繰り返すことにより、学習器内に、ポップノイズ、非ポップノイズ音声それぞれのモデルが形成される。これにより、差分信号の特徴量の入力に応じ、該当するモデルを出力する識別器が形成される。

【0055】

本実施の形態によれば、音声取得部110が、Microphone pointにおいて、生体検知に不可欠な情報を取得する。これにより、Transmission p

10

20

30

40

50

ointにおけるなりすまし検知では不可能であった生体検知を実現することができる。

【0056】

また、本実施の形態によれば、判断部150は、ポップノイズ検出部130によるポップノイズ検出結果に基づいて、生体検知を行う。ポップノイズはスピーカでは原理的に再現不能な現象であるので、これにより、話者の詐称に頑健な生体認証を実現することができる。

【0057】

また、本実施の形態によれば、音声取得部110は、複数のマイクロフォンを用いることで、ポップノイズを含む音声信号及び含まない音声信号を出力する。これにより、ポップノイズ検出部130は、公知の分離技術を適用して効率的にポップノイズ成分を分離することができるようになった。

【0058】

<実施の形態2>

実施の形態2は、音声取得部110が複数のマイクロフォンを備え、それらのマイクロフォンをそれぞれ異なる空間に配置した構成例である。その余の構成については、実施の形態1と同様である。

【0059】

本実施の形態では、音声取得部110が有する複数のマイクロフォンのうち、一方のマイクロフォン(第1のマイクロフォン)は、話者の声とともにポップノイズを拾うことを目的とするため、話者の息が直接かかりやすい位置に配置される。例えば、第1のマイクロフォンは話者に正対する位置に配置される。他方のマイクロフォン(第2のマイクロフォン)は、ポップノイズを拾うことを抑制し、可能な限り話者の声だけを拾うことを目的とするため、話者の息が直接かかりにくい位置に配置される。例えば、第2のマイクロフォンは話者の側方や、第1のマイクロフォンよりも離れた位置に配置される。なお、本実施の形態においては、第2のマイクロフォンには必ずしもマイクカバーやポップフィルタを備えることを要しない。

【0060】

例えば、音声取得部110としてアレイマイクを用いる場合は、話者の近くに位置するマイクロフォンを第1のマイクロフォン、第1のマイクロフォンよりも話者から遠くに位置するマイクロフォンを第2のマイクロフォンとして扱うことができる。

【0061】

本実施の形態によれば、音声取得部110は、マイクカバーやポップフィルタを用いることなく、実施の形態1と同様のポップノイズ検出処理を実現することができる。

【0062】

<実施の形態3>

実施の形態3は、音声取得部110が単一のマイクロフォンを備え、分離モジュール131としてローパスフィルタを採用した構成例である。その余の構成については、実施の形態1と同様である。

【0063】

実施の形態3における音声取得部110は、1本のマイクロフォンにより構成される。このマイクロフォンは、話者の声やポップノイズを可能な限りそのまま拾うことを目的とする。したがって、話者の息がかかりやすい位置に配置された、マイクカバーやポップフィルタを備えていないマイクロフォンであることが好ましい。あるいは、第1のマイクロフォンは、風など周囲の影響を軽減しつつ、発話に伴うポップノイズのみ取得できるよう、穴あきウィンドスクリーンを備えるものであっても良い。

【0064】

また、実施の形態3における分離モジュール131は、ローカットフィルタ及びローパスフィルタを備える。

【0065】

次いで、実施の形態3の特徴的な動作について説明する。

## 【 0 0 6 6 】

S 1 0 1 : 音声取得

音声取得部 1 1 0 が、話者の音声を取得する。本実施の形態では、音声取得部 1 1 0 は 1 本のマイクロフォンである。音声取得部 1 1 0 の変換部は、マイクロフォンが取得した音声をデジタル信号に変換してポップノイズ検出部 1 3 0 に出力する。

## 【 0 0 6 7 】

S 1 0 4 : ポップノイズ検出

ポップノイズ検出部 1 3 0 の分離モジュール 1 3 1 は、ローカットフィルタを利用して、音声取得部 1 1 0 が出力する音声信号から音声成分のみを抽出する。ポップノイズ検出部 1 3 0 は、抽出された音声成分を、実施の形態 1 における第 2 の音声と同等のものとして利用する。また、ポップノイズ検出部 1 3 0 は、ローパスフィルタを利用して、音声取得部 1 1 0 が出力する音声信号からノイズ成分のみを抽出する。ポップノイズ検出部 1 3 0 は、抽出されたノイズ成分を、実施の形態 1 における差分信号と同等のものとして利用できる。

## 【 0 0 6 8 】

ローパスフィルタは、音声波形を平滑化するが、音声信号に含まれる異常値は残す性質がある。そのため、ローパスフィルタにより、ポップノイズを含む音声信号から、ポップノイズ成分を顕出させることができる。

## 【 0 0 6 9 】

そして、ポップノイズ検出部 1 3 0 は、実施の形態 1 と同様に識別器を用いて、抽出したノイズ成分がポップノイズであるか否かを判定する。すなわち、特徴量化モジュール 1 3 2 が、実施の形態 1 における差分信号の代わりにローパスフィルタにより抽出されたノイズ成分を特徴量化し、好ましくは幾つかのセグメントに分割する。そして、識別モジュール 1 3 4 が、2 つの識別器にノイズ成分の特徴量を入力し、出力される尤度の差に基づいてポップノイズの存在の有無を判定する。

## 【 0 0 7 0 】

S 1 0 5 乃至 S 1 0 6 :

実施の形態 1 と同様に動作する。

## 【 0 0 7 1 】

本実施の形態の事前学習工程における動作は以下のとおりである。

S 2 0 1 : 音声取得

S 1 0 1 と同様に、音声取得部 1 1 0 が、話者の音声を取得する。

## 【 0 0 7 2 】

S 2 0 2 : ポップノイズ/非ポップノイズモデルの学習

まず、S 1 0 4 と同様に、分離モジュール 1 3 1 が、ローパスフィルタを使用して音声信号からノイズ成分を抽出する。次いで、特徴量化モジュール 1 3 2 がノイズ信号を特徴量化する。そして、識別モジュール 1 3 4 が、ノイズ信号の特徴量と、それがポップノイズである旨を示す教師信号とを共に一方の学習器に入力する。また、分離モジュール 1 3 1 が、ローカットフィルタを使用して音声信号から音声成分を抽出する。特徴量化モジュール 1 3 2 は、音声成分を特徴量化する。そして、識別モジュール 1 3 4 は、音声成分の特徴量と、それが非ポップノイズである旨を示す教師信号と、を共に他方の学習器に入力する。その余の動作については、実施の形態 1 と同様である。

## 【 0 0 7 3 】

( 実験結果 )

図 4 に、実施の形態 3 の構成を用いた実証実験結果を示す。発明者は、F 0 0 1 乃至 F 0 1 0 の 1 0 人の話者を対象として、ポップノイズの検出を試行した。その結果、すべての話者において、ポップノイズを含む文章を学習した識別器が出力する尤度(「ポップノイズあり音声」)が、ポップノイズを含まない文章を学習した識別器が出力する尤度(「ポップノイズなし音声」)を上回った。すなわち、話者の入力音声にポップノイズを含むものであることを、高精度で検出することが可能であることがわかった。

10

20

30

40

50

## 【0074】

本実施の形態によれば、音声取得部110は単一のマイクロフォンで構成され、ポップノイズ検出部130がローカットフィルタ及びローパスフィルタを用いて音声成分及びノイズ成分を抽出する。これにより、複数のマイクロフォンを使用する場合に比べ簡素な構成で生体検知を実現できる。

## 【0075】

<実施の形態4>

実施の形態4は、音声取得部110が話者の音声を直接取得するための単一のマイクロフォンを備え、分離モジュール131としてスピーカ及びマイクロフォンを採用した構成例である。その余の構成については、実施の形態1と同様である。

10

## 【0076】

実施の形態4における音声取得部110は、話者の音声を直接取得するための1本のマイクロフォン(第1のマイクロフォン)により構成される。第1のマイクロフォンは、話者の声やポップノイズを可能な限りそのまま拾うことを目的とする。したがって、話者の息がかかりやすい位置に配置された、マイクカバーやポップフィルタを備えていないマイクロフォンであることが好ましい。あるいは、第1のマイクロフォンは、風など周囲の影響を軽減しつつ、発話に伴うポップノイズのみ取得できるよう、穴あきウィンドスクリーンを備えるものであっても良い。

## 【0077】

また、ポップノイズ検出部130の分離モジュール131は、第1のマイクロフォンによる収録音声を出力するスピーカと、スピーカが出力し空気伝播した音声を収録する第2のマイクロフォンとを備える。なお、第2のマイクロフォンによる音声収録は生体検知装置100の内部で実施すれば良いため、スピーカ及び第2のマイクロフォンは話者に対して露出している必要はない。

20

## 【0078】

次いで、実施の形態4の特徴的な動作について説明する。

## 【0079】

S101:音声取得

音声取得部110の第1のマイクロフォンが、話者の音声を取得する。本実施の形態では、第1のマイクロフォンのみが話者の音声を直接収録する。第1のマイクロフォンが取得する音声は、話者が発生させるポップノイズを含むものである。

30

## 【0080】

次に、分離モジュール131のスピーカが、第1のマイクロフォンが収録した話者の音声を再生する。そして、第2のマイクロフォンが、スピーカから再生された音声を収録する。スピーカから出力される音声は、原理的にポップノイズを発生させないので、第2のマイクロフォンが取得する音声はポップノイズ成分が含まれないものとなる。

## 【0081】

音声取得部110及び分離モジュール131は、第1のマイクロフォン及び第2のマイクロフォンが取得した音声をそれぞれ別のデジタル信号に変換する。ここで、第1のマイクロフォン由来の音声データを第1の音声、第2のマイクロフォン由来の音声データを第2の音声と称する。変換部は、第1の音声及び第2の音声をポップノイズ検出部130に対して出力する。

40

## 【0082】

S104乃至S106、及びS201乃至S202にかかる動作は実施例1と同様であるため、詳細な説明を省略する。

## 【0083】

本実施の形態によれば、音声取得部110は、話者に対して露出する1本のマイクロフォンと、装置内部に設けられるスピーカ及び第2のマイクロフォンで構成される。これにより、話者に対して露出する複数のマイクロフォンを使用する場合に比べ、簡素な外観で生体検知装置100を構成できる。

50

## 【 0 0 8 4 】

## &lt;実施の形態 5 &gt;

実施の形態 1 乃至 4 では、生体検知装置 1 0 0 が、音声信号にポップノイズが含まれているか否かを判断することにより、生体検知を行う例について説明した。しかしながら、実施の形態 1 乃至 4 の手法では、例えば風など話者の息以外の要因によりポップノイズ類似の音が入力された場合、生体検知装置 1 0 0 はこれを生体が発声したポップノイズと誤認してしまうことがある。そこで実施の形態 5 では、実施の形態 1 乃至 4 と比較して、特に風などの影響に対して頑健な生体検知手法を提示する。

## 【 0 0 8 5 】

通常ポップノイズは、例えば破裂音など一部の特定の子音で主に発生することが知られている。そこで本実施の形態では、ポップノイズが適切な場所で発生しているか否かを検査することにより、頑健さを強化する。

10

## 【 0 0 8 6 】

本実施の形態における生体検知装置 1 0 0 は、実施の形態 1 乃至 4 に係る生体検知装置 1 0 0 の構成要素に加え、ポップノイズ検出部 1 3 0 内に音素アライメントモジュール 1 3 3 を有する点に特徴を有する。その余の構成は、特段の言及がない限り実施の形態 1 乃至 4 と同様である。

## 【 0 0 8 7 】

音声取得部 1 1 0 は、音声信号を音素アライメントモジュール 1 3 3 に対して出力する。ポップノイズ検出部 1 3 0 の分離モジュール 1 3 1 は、音声信号からポップノイズ成分を分離する。そして特徴量化モジュール 1 3 2 が、ポップノイズ成分を特徴量化する。

20

## 【 0 0 8 8 】

音素アライメントモジュール 1 3 3 は、音声取得部 1 1 0 から音声信号を入力し、音声信号中の音素を識別する処理を行う。そして、音声信号を各音素の時間長で分割したセグメントを定義し、ポップノイズ成分の特徴量をセグメントに分割する。

## 【 0 0 8 9 】

識別モジュール 1 3 4 は、セグメント単位でポップノイズの検出を実行する。すなわち、セグメント化された特徴量を識別器に投入し、セグメント毎にポップノイズの存在の有無を識別する。

## 【 0 0 9 0 】

判断部 1 5 0 は、ポップノイズ検出部 1 3 0 による各セグメントのポップノイズの検出結果と、音素との対応関係が正しいか否かを検証する。

30

## 【 0 0 9 1 】

次いで、図 5 のフローチャートを用いて、生体検知装置 1 0 0 の動作について説明する。

## 【 0 0 9 2 】

S 3 0 1 : 音声取得

S 1 0 1 同様、音声取得部 1 1 0 が、話者の音声を取得する。

## 【 0 0 9 3 】

S 3 0 2 : 音素識別

音素アライメントモジュール 1 3 3 は、音声取得部 1 1 0 から入力した音声信号から音素を抽出する。好ましくは、ポップノイズの比較的少ない第 2 の音声を利用することができる。

40

## 【 0 0 9 4 】

ここで音素とは、言語学上の価値を有する音声の最小単位をいう。例えば、個々の母音および子音が音素に相当する。音声信号からの音素の認識は、例えば H M M 等の公知の手法を利用して行うことができる。

## 【 0 0 9 5 】

音素アライメントモジュール 1 3 3 は、音声信号に対し、認識した各音素の時間長に対応する複数のセグメントを定義する ( 図 1 2 )。セグメントは、典型的には、各音素の始

50

点（時刻）および時間長によって定義できる。また、音素アライメントモジュール 133 は、各セグメントに対し、音素名をラベルとして付与する。例えば音素アライメントモジュール 133 は、音声信号に含まれる音素夫々について、音素名、音素の始点、及び音素の時間長を対応付けたレコードを作成し、図示しない記憶領域に保持させることでこれを実現できる（図 13）。

#### 【0096】

##### S303：ポップノイズ検出

ポップノイズ検出部 130 の分離モジュール 131 は、音声信号からポップノイズ成分を分離抽出する。つづいて、特徴量化モジュール 132 がポップノイズ成分を特徴量化する。そして、音素アライメントモジュール 133 が特徴量化されたポップノイズ成分をセグメントに分割する。識別モジュール 134 が、各セグメントごとにポップノイズの検出処理を行う。

10

#### 【0097】

例えば、音声信号に対して図 13 のようなセグメントが定義されている場合を考える。まず、始点が 00:00:00:00、時間長が 00:00:00:12 であるセグメント（ID = 1）が定義されているので、音素アライメントモジュール 133 は、ポップノイズ成分の特徴量のうち、このセグメントに相当する時間帯、すなわち 00:00:00:00 から 00:00:00:12 までの領域の特徴量を切り出す。そして、識別モジュール 134 は、切り出した特徴量を識別器に入力し、ポップノイズの検出結果を得る。

#### 【0098】

つづいて、音素アライメントモジュール 133 は、ポップノイズ成分の特徴量のうち、次のセグメント（ID = 2）に相当する時間幅、すなわち 00:00:00:12 から 00:00:00:20 までの領域の特徴量を切り出す。そして、同様に識別器を用いてポップノイズの検出を行う。同様に、ポップノイズ検出部 130 はすべてのセグメントについてポップノイズの検出を行う。ポップノイズ検出部 130 は、検出を試行した各セグメントについて、検出結果を記憶する（図 14）。

20

#### 【0099】

##### S304：ポップノイズと音素の関係の妥当性検証

判断部 150 は、S303 で得られたポップノイズの検出結果の妥当性を検証する。上述のように、ポップノイズは、破裂音（例えば“p”）など一部の特定の子音で主に発生することが知られている。本実施の形態では、生体検知装置 100 は、このような、音素と、ポップノイズ発生可能性の有無と、を対応付けたパターン表を予め保持しているものとする（図 15）。

30

#### 【0100】

判断部 150 は、S303 で得られたポップノイズの検出結果における音素とポップノイズ検出結果との対応関係（図 14）と、予め与えられた音素とポップノイズ発生可能性との対応関係（図 15）を比較し、両者が整合しているか否かを検証する。例えば、図 14 では、セグメント ID = 1 の音素“t”について、ポップノイズが“非検出”である。一方、図 15 では、音素“t”について、ポップノイズの発生可能性は“無”と定義されている。よって、判断部 150 は、セグメント ID = 1 の判定結果は妥当と判断する。

40

#### 【0101】

また、セグメント ID = 5 の音素“p”については、ポップノイズが“検出”されている。一方、図 15 でも、音素“p”について、ポップノイズの発生可能性は“有”と定義されている。よって、判断部 150 は、セグメント ID = 5 の判定結果も妥当と判断する。同様にして、判断部 150 は、S303 で得られたポップノイズの検出結果の各々について、妥当性を検証していく。

#### 【0102】

判断部 150 は、すべてのセグメントについて妥当性が確認できたならば、話者は生体であると判断し、S305 に遷移する。一方、すべてのセグメントで妥当性が確認できなかった場合には話者は生体でないものと判断し、S306 に遷移する。

50

## 【 0 1 0 3 】

なお、ここで判断部 1 5 0 は、生体検知の判断基準として、上記以外の任意の基準を適宜採用できる。例えば、妥当性が検証されたセグメントの割合が所定の閾値を超えた場合に、話者は生体であると判断するようにしても良い。

## 【 0 1 0 4 】

S 3 0 5 : 話者照合

話者が生体であると判断された場合、生体検知装置 1 0 0 は任意の手法を用いた話者照合フェーズに移行することができる。

## 【 0 1 0 5 】

S 3 0 6 : 詐称音声として棄却

話者が生体でないと判断された場合、音声取得部 1 1 0 が取得した音声は人が発したものであるのではなく、例えば合成や声質変換による音声である蓋然性が高い。よって、生体検知装置 1 0 0 はこれを詐称音声と判断し、話者照合を行うことなく棄却する。すなわち、エラー処理や終了処理等を行う。

## 【 0 1 0 6 】

本実施の形態によれば、音素アライメントモジュール 1 3 3 が音声信号を音素レベルに区分し、ポップノイズ検出部 1 3 0 が音素レベルでポップノイズの発生を検出する。そして、判断部 1 5 0 が、ポップノイズが適切な位置（音素）で発生しているか否かを検証する。これにより、人の発声によらない、例えば風などに由来するポップノイズによる影響を排除し得る、より頑健な生体検知を実現できる。

## 【 0 1 0 7 】

( 実験結果 )

発明者は、実施の形態 5 の構成を用いて生体検知の実証実験を行った。実験においては、音声取得部 1 1 0 に対し、話者の生音声及びスピーカ出力される音声をそれぞれ入力した。音声取得部 1 1 0 としては、ポップフィルタのない第 1 のマイクロフォン及びポップフィルタを備える第 2 のマイクロフォンを採用した。第 1 のマイクロフォン及び第 2 のマイクロフォンとしては、コンデンサマイク、ステレオマイク及びヘッドセットマイクを使用した。また、特徴量化モジュール 1 3 2 は採用せず、第 1 の音声と第 2 の音声との差分信号をそのまま識別器に入力した。なお、識別器はポップノイズ発生時の差分信号を学習済みである。そして、識別モジュール 1 3 4 は、識別子の出力する尤度が所定の閾値以上

## 【 0 1 0 8 】

図 1 7 に、本実証実験の結果を示す。コンデンサマイク、ステレオマイクのいずれを用いた場合においても、話者の生音声を入力した場合は高い確率（ 1 0 0 % ）で生体と検知した。また、スピーカ出力した音声を入力した場合も、高い確率（ 1 0 0 % ）で生体でないものと判定した。なお、ヘッドセットマイクを用いた場合においても、 8 6 . 9 % と概ね良好な生体検知率が観察された。

## 【 0 1 0 9 】

< その他の実施の形態 >

なお、本発明は上記実施の形態に限られたものではなく、趣旨を逸脱しない範囲で適宜変更することが可能である。

## 【 0 1 1 0 】

例えば、実施の形態 5 では、 1 つの音素に対してポップノイズ発生可能性の有無が予め対応付けられている例（図 1 5 ）を示したが、複数の連続する音素に対して、ポップノイズ発生可能性の有無が対応付けられていても良い。

## 【 0 1 1 1 】

また、上述の実施の形態では、本発明を主にハードウェアの構成として説明したが、これに限定されるものではなく、任意の処理を、CPU（Central Processing Unit）にコンピュータプログラムを実行させることにより実現することも可能である。この場合、コンピュータプログラムは、様々なタイプの非一時的なコンピュー

10

20

30

40

50

タ可読媒体 (non-transitory computer readable medium) を用いて格納され、コンピュータに供給することができる。非一時的なコンピュータ可読媒体は、様々なタイプの実体のある記録媒体 (tangible storage medium) を含む。非一時的なコンピュータ可読媒体の例は、磁気記録媒体 (例えばフレキシブルディスク、磁気テープ、ハードディスクドライブ)、光磁気記録媒体 (例えば光磁気ディスク)、CD-ROM (Read Only Memory)、CD-R、CD-R/W、半導体メモリ (例えば、マスクROM、PROM (Programmable ROM)、EPROM (Erasable PROM)、フラッシュROM、RAM (random access memory)) を含む。また、プログラムは、様々なタイプの一時的なコンピュータ可読媒体 (transitory computer readable medium) によってコンピュータに供給されてもよい。一時的なコンピュータ可読媒体の例は、電気信号、光信号、及び電磁波を含む。一時的なコンピュータ可読媒体は、電線及び光ファイバ等の有線通信路、又は無線通信路を介して、プログラムをコンピュータに供給できる。

10

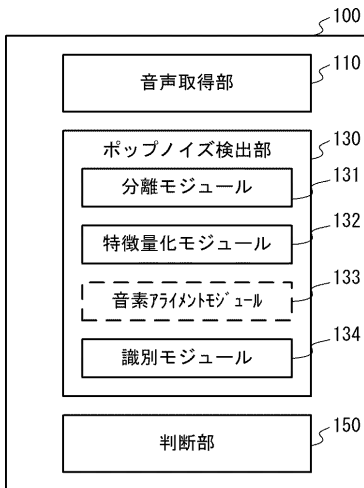
【符号の説明】

【0112】

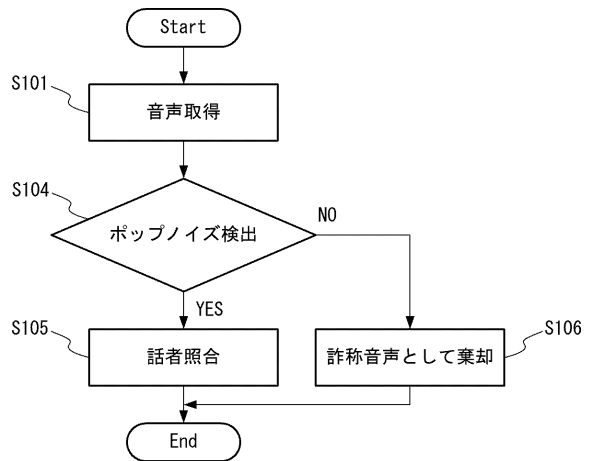
- 100 生体検知装置
- 110 音声取得部
- 130 ポップノイズ検出部
- 131 分離モジュール
- 132 特徴量化モジュール
- 133 音素アライメントモジュール
- 134 識別モジュール
- 150 判断部

20

【図1】

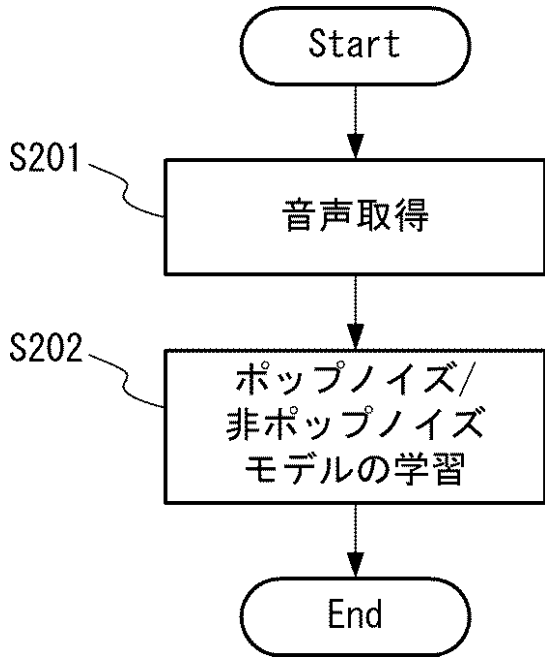


【図2】

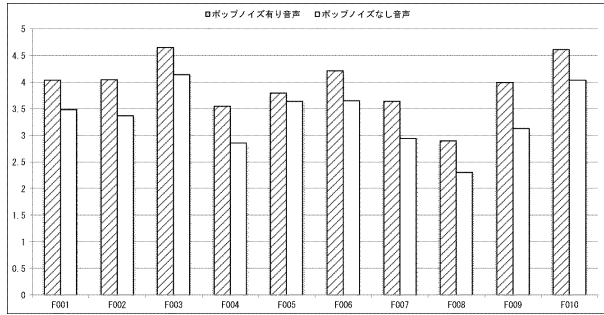




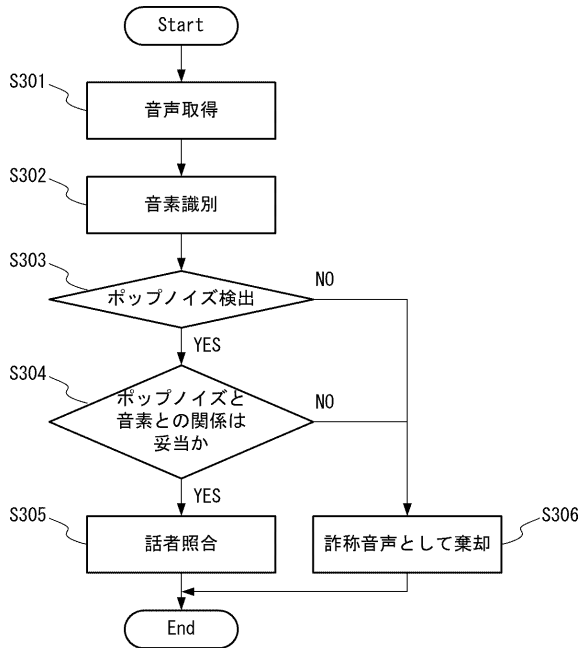
【図3】



【図4】



【図5】



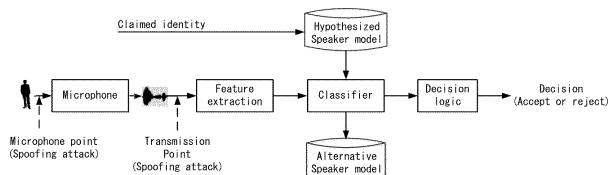
【図7】



【図8】



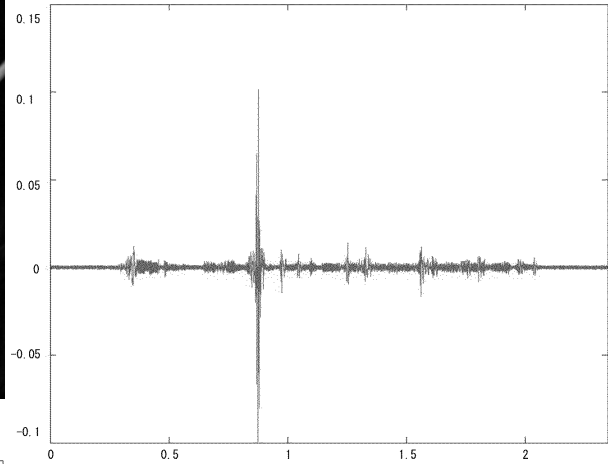
【図6】



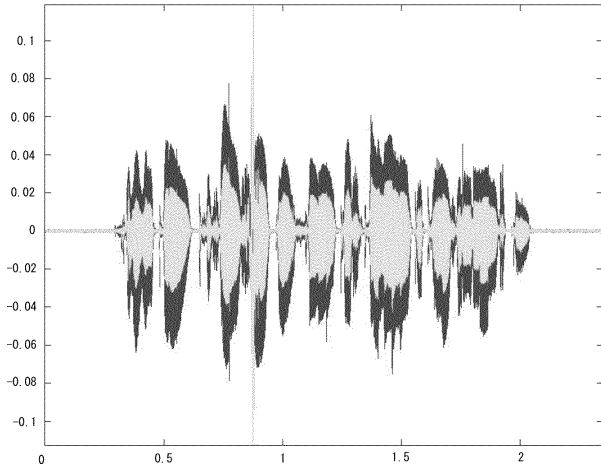
【図 9】



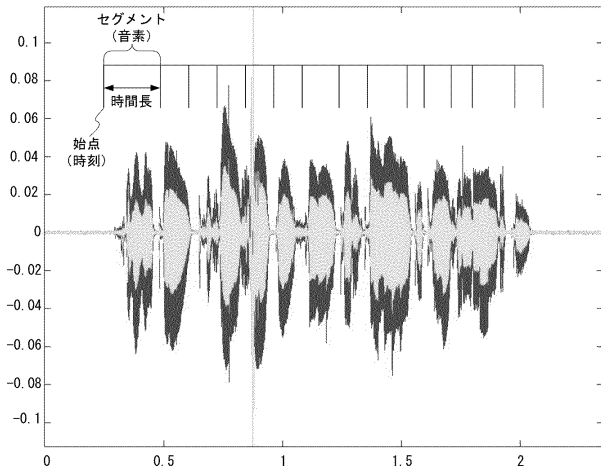
【図 11】



【図 10】



【図 12】



【図 14】

セグメントID	始点	時間長	音素名	検出結果
1	00:00:00:00	00:00:00:12	t	非検出
2	00:00:00:12	00:00:00:08	r	非検出
3	00:00:00:20	00:00:00:08	u	非検出
4	00:00:00:28	00:00:00:08	m	非検出
5	00:00:00:36	00:00:00:08	p	検出
6	00:00:00:44	00:00:00:08	e	非検出
7	00:00:00:52	00:00:00:12	t	非検出

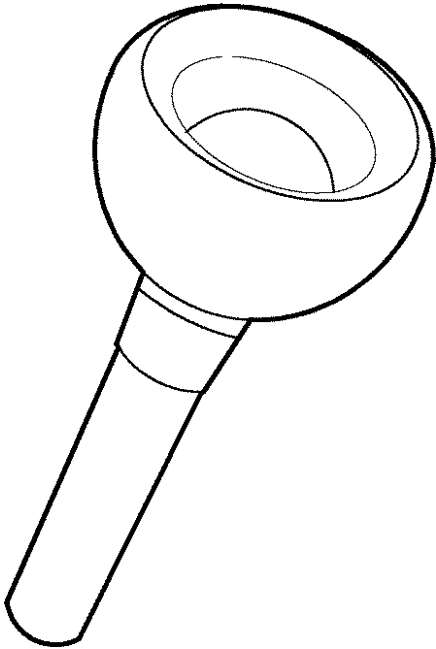
【図 13】

セグメントID	始点	時間長	音素名
1	00:00:00:00	00:00:00:12	t
2	00:00:00:12	00:00:00:08	r
3	00:00:00:20	00:00:00:08	u
4	00:00:00:28	00:00:00:08	m
5	00:00:00:36	00:00:00:08	p
6	00:00:00:44	00:00:00:08	e
7	00:00:00:52	00:00:00:12	t

【図 15】

音素名	ポップノイズ有無
o	無
p	有
q	無
r	無
s	無
t	無
u	無

【図16】



【図17】

入力	生体検知率 (生体と判定した回数/試行回数)
生音声 (ステレオマイク)	100% (510/510)
スピーカ出力された音声 (ステレオマイク)	100% (0/85)
生音声 (コンデンサマイク)	100% (510/510)
スピーカ出力された音声 (コンデンサマイク)	100% (0/85)
生音声 (ヘッドセットマイク)	86.9% (443/510)

## フロントページの続き

- (72)発明者 小野 順貴  
東京都千代田区一ツ橋2 - 1 - 2 大学共同利用機関法人情報・システム研究機構 国立情報学研究所内
- (72)発明者 松井 知子  
東京都立川市緑町10 - 3 大学共同利用機関法人情報・システム研究機構 統計数理研究所内
- (72)発明者 塩田 さやか  
東京都立川市緑町10 - 3 大学共同利用機関法人情報・システム研究機構 統計数理研究所内

審査官 菊池 智紀

- (56)参考文献 特開平2 - 124600 (JP, A)  
特開平8 - 211888 (JP, A)  
特開平4 - 61500 (JP, A)  
特開昭59 - 62295 (JP, A)  
特開昭59 - 49742 (JP, A)  
特開2007 - 171637 (JP, A)  
特開2000 - 276191 (JP, A)  
実開昭58 - 180600 (JP, U)  
特表2008 - 538882 (JP, A)  
国際公開第2006 / 117032 (WO, A1)

(58)調査した分野(Int.Cl., DB名)

G10L 15/00 - 17/26,  
25/00 - 25/93