(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2013-160937
(P2013-160937A)
(43)公開日　平成25年8月19日(2013.8.19)

(51)Int.Cl.		F I			テーマコード（参考）
G10L 15/04	(2013.01)	G10L	15/04	300C	5D015
G10L 25/78	(2013.01)	G10L	11/00	402L	

審査請求 未請求　請求項の数 7　OL　（全 11 頁）

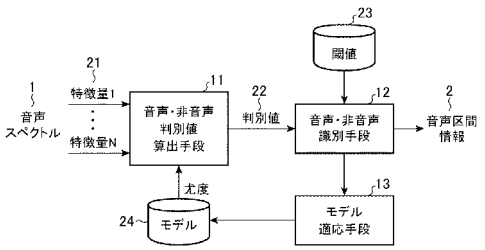
(21)出願番号　特願2012-23132(P2012-23132)
(22)出願日　平成24年2月6日(2012.2.6)

(71)出願人　000006013
三菱電機株式会社
東京都千代田区丸の内二丁目７番３号
(74)代理人　100123434
弁理士　田澤　英昭
(74)代理人　100101133
弁理士　濱田　初音
(74)代理人　100173934
弁理士　久米　輝代
(74)代理人　100156351
弁理士　河村　秀央
(72)発明者　太刀岡　勇気
東京都千代田区丸の内二丁目７番３号　三菱電機株式会社内

Ｆターム(参考)　5D015 DD03

(54)【発明の名称】音声区間検出装置

(57)【要約】
【課題】環境が変化した場合でも音声の検出を正しく行うことのできる音声区間検出装置を得る。
【解決手段】音声・非音声判別値算出手段１１は、入力された音声スペクトル１から特徴量２１を利用して、モデル２４により音声・非音声区間それぞれにおいて異なる判別値２２を算出する。音声・非音声識別手段１２は、判別値２２と予め設定した閾値２３とを比較することで音声・非音声を識別する。モデル適応手段１３は、音声・非音声識別手段１２の識別結果に基づいてモデル２４を適応する。
【選択図】図１



【特許請求の範囲】
【請求項１】
入力された音声のスペクトルからある特徴量もしくは複数の異なる特徴量の組み合わせを利用して、モデルにより音声・非音声区間それぞれにおいて異なる判別値を算出する音声・非音声判別値算出手段と、
前記判別値と予め設定した閾値とを比較することで音声・非音声を識別する音声・非音声識別手段と、
前記音声・非音声識別手段の識別結果に基づいて前記モデルを適応するモデル適応手段を備えたことを特徴とする音声区間検出装置。
【請求項２】
モデルは音声モデルとノイズモデルであり、音声・非音声判別値算出手段は、これらの尤度の比または差を算出し、判別値とすることを特徴とする請求項１記載の音声区間検出装置。
【請求項３】
モデルは、予め音声とノイズのデータから学習した密度比モデルであり、音声・非音声判別値算出手段は、前記密度比モデルに基づいて尤度比を算出し、判別値とすることを特徴とする請求項１記載の音声区間検出装置。
【請求項４】
学習環境が異なる複数のモデルを用意すると共に、実際の環境に対応して前記複数のモデルから特定のモデルを選択するモデル選択手段を備えたことを特徴とする請求項１から請求項３のうちのいずれか１項記載の音声区間検出装置。
【請求項５】
モデル適応手段は、音声・非音声識別手段で非音声または音声と判定されたフレームの特徴量を用いてモデルの適応を行うことを特徴とする請求項４記載の音声区間検出装置。
【請求項６】
音声・非音声判別値算出手段および音声・非音声識別手段とは別の特徴量に基づいて音声と非音声とを識別する音声・非音声判別値算出および識別手段を備え、
モデル適応手段は、前記音声・非音声判別値算出および識別手段で非音声または音声と判定されたフレームの特徴量に基づいてモデルの適応を行うことを特徴とする請求項１から請求項５のうちのいずれか１項記載の音声区間検出装置。
【請求項７】
モデル適応手段は、予め作成された音声モデルと適応したノイズモデルとを比較し、当該比較結果に基づいて適応の可否を判断することを特徴とする請求項２記載の音声区間検出装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、音声モデルやノイズモデルまたは尤度比モデルを用いて音声区間を検出する音声区間検出装置に関する。
【背景技術】
【０００２】
音声区間検出法としては音声のパワーがノイズのそれよりも大きなことを利用するパワーによるものがよく用いられている（例えば、非特許文献１参照）。また、音声区間、非音声区間の尤度比により音声区間検出を行う手法が提案されている（例えば、非特許文献２参照）。さらに、別途学習した音声モデルを用いたもの（例えば、特許文献１参照）が提案されている。
【先行技術文献】
【特許文献】
【０００３】
【特許文献１】特開２００９－２１０６４７号公報

【非特許文献】
【０００４】
【非特許文献１】L.R. Rabiner and M.R. Sambur, “An algorithm for determining the endpoints of isolated utterances, ”Bell Syst. Tech. , vol.54(2), pp.297-315, 1975.
【非特許文献２】J. Sohn, N.S. Kim, and W. Sung, “Statistical Model-Based Voice Activity Detection, ”IEEE Signal Processing Letters, vol.6(1), pp. 1-3, 1999.
【発明の概要】
【発明が解決しようとする課題】
【０００５】
しかしながら、上記非特許文献１に記載されたような手法（以下、手法１とする）は音声のパワーがノイズに埋もれて検出できないため、低Ｓ／Ｎ下で有効でないという問題点があった。また、非特許文献２に記載されているようなノイズモデルを用いる手法（以下、手法２とする）は、ノイズらしさ音声らしさを判定できるため、手法１のような問題は起こりにくく、低Ｓ／Ｎ下でも音声を検出できるが、ノイズが非定常な場合には学習したモデルと実際の環境とのミスマッチにより、有効でないという問題があった。さらに、特許文献１に記載されたような手法（手法３とする）はオンラインでモデルを推定するため手法２のような問題は起こりにくいが、毎フレームモデルを合成するため、計算量が多いという問題があった。
【０００６】
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、環境が変化した場合でも音声の検出を正しく行うことのできる音声区間検出装置を得ることを目的とする。
【課題を解決するための手段】
【０００７】
この発明に係る音声区間検出装置は、入力された音声のスペクトルからある特徴量もしくは複数の異なる特徴量の組み合わせを利用して、モデルにより音声・非音声区間それぞれにおいて異なる判別値を算出する音声・非音声判別値算出手段と、判別値と予め設定した閾値とを比較することで音声・非音声を識別する音声・非音声識別手段と、音声・非音声識別手段の識別結果に基づいてモデルを適応するモデル適応手段とを備えたものである。
【発明の効果】
【０００８】
この発明の音声区間検出装置は、音声・非音声識別手段の識別結果に基づいてモデルを適応するモデル適応手段を備えたので、環境が変化した場合でも音声の検出を正しく行うことができる。
【図面の簡単な説明】
【０００９】
【図１】この発明の実施の形態１による音声区間検出装置を示す構成図である。
【図２】この発明の実施の形態１による音声区間検出装置の具体例を示す構成図である。
【図３】この発明の実施の形態２による音声区間検出装置の構成図である。
【図４】この発明の実施の形態３による音声区間検出装置の構成図である。
【図５】この発明の実施の形態４による音声区間検出装置の構成図である。
【図６】この発明の実施の形態５による音声区間検出装置に係わるモデル更新の検証を示す説明図である。
【図７】この発明の実施の形態５による音声区間検出装置の構成図である。
【発明を実施するための形態】
【００１０】
本発明では手法２の性能を改善することを目的とする。具体的には、学習したモデルと実際の環境とのミスマッチという問題に対応するために、モデルの適応を行う方法を提案する。また、モデルの頑健性を増し、頻繁にモデル適応をしなくてもよいように、尤度比

モデルの使用により頑健性を高める方法に関しても述べる。

【００１１】

実施の形態１．

図１は、この発明の実施の形態１による音声区間検出装置の構成図である。

図１に示す音声区間検出装置は、音声・非音声判別値算出手段１１、音声・非音声識別手段１２、モデル適応手段１３を備えている。音声・非音声判別値算出手段１１は、音声スペクトル１を入力し、その特徴量（１～Ｎ）２１とモデル２４との尤度の比または差に基づいて、判別値２２を出力する手段である。音声・非音声識別手段１２は、音声・非音声判別値算出手段１１から送出された判別値２２と、予め求められた閾値２３とを比較し、音声と非音声との識別結果をモデル適応手段１３に出力すると共に、音声と判定した場合は、その音声区間情報２を出力する手段である。モデル適応手段１３は、音声・非音声識別手段１２の音声・非音声の識別結果に基づいて、モデル２４を適応する手段である。

【００１２】

このように構成された音声区間検出装置は、入力の音声スペクトル１から音声区間情報２を出力する。まず、一つもしくは複数の特徴量２１を音声・非音声判別値算出手段１１に入力する。特徴量としては、帯域別のスペクトル、パワー、ＭＦＣＣ（メル周波数ケプストラム係数）、ゼロ交差点数などを用いることができる。音声・非音声判別値算出手段１１は、モデル２４により尤度を得て、これらを統合して判別値２２を出力する。音声・非音声識別手段１２では、受け取った判別値２２を、事前に設定しておいた閾値２３と比較し、音声区間であれば音声区間情報２として出力する。また、その結果をモデル適応手段１３にフィードバックし、モデル適応手段１３はモデル２４を適応する。

【００１３】

この処理を具体的に説明するために、図１のモデル２４として音声モデル２４ａとノイズモデル２４ｂを用いた構成を図２に示す。ここでは、特徴量の例として複素スペクトルを用いた手法２に即してモデル適応の方法を説明する。非音声区間H_0と音声区間H_1において、音声とノイズの離散フーリエ変換（ＤＦＴ）の係数のＬ次元のベクトル（複素スペクトル）をそれぞれＳ，Ｎとして観測音Ｘは下式のように表される。それぞれはＳ＝（S_0，S_1，…，S_k，…，S_{L-1}），Ｎ＝（N_0，N_1，…，N_k，…，N_{L-1}），Ｘ＝（X_0，X_1，…，X_k，…，X_{L-1}）である。

H_0：Ｘ＝Ｎ，H_1：Ｘ＝Ｎ＋Ｓ

【００１４】

手法２ではここでそれぞれの状態のＤＦＴ係数の確率密度関数が、下式のように独立なガウス分布で表せると仮定する。

$$p(X \mid H_0) = \prod_{k=0}^{L-1} \frac{1}{\pi \lambda_N(k)} \exp\left\{ -\frac{|X_k|^2}{\lambda_N(k)} \right\}$$

$$p(X \mid H_1) = \prod_{k=0}^{L-1} \frac{1}{\pi \{\lambda_N(k) + \lambda_S(k)\}} \exp\left\{ -\frac{|X_k|^2}{\lambda_N(k) + \lambda_S(k)} \right\}$$

ここで$\lambda_N(k)$，$\lambda_S(k)$はN_k，S_kの分散を表す。

するとk番目の周波数帯域の音声・非音声の尤度比は下式で表される。

$$\Lambda_k = \frac{p(X \mid H_1)}{p(X \mid H_0)} = \frac{1}{1+\xi_k} \exp\left(\frac{\gamma_k \xi_k}{1+\xi_k} \right)$$

【００１５】

モデルとしてノイズパワーの分散$\lambda_N(k)$と音声パワーの分散を$\lambda_S(k)$を予め学習しておく。このモデルを用いて、音声・非音声判別値算出手段１１で上式Λ（確率の比（音声の確率／ノイズの確率））を計算する。これが判別値（確率密度比）２２となる。音声・非音声識別手段１２において、判別値２２が予め設定した閾値２３より大きかった場合には音声であると判断し、小さかった場合にはノイズであると判断する。

【００１６】

この方法によると学習時の$\lambda_{\{S,N\}}(k)$と実際の環境でのそれが異なった場合には推定精度が大きく低下する。この場合にモデル適応手段１３では、判別値２２が閾値２３より非常に大きかった場合には$\lambda_S(k)$の学習データに加え、再学習する。一方、判別値２２が閾値２３より非常に小さかった場合にはノイズのそれを再学習する。これにより$\lambda_{\{S,N\}}(k)$が徐々に変化した場合には追従が可能となり、$\lambda_{\{S,N\}}(k)$が大きく変化した場合にも時間の経過と共に徐々に追従することができる。このようにモデル適応手段１３を備えたことにより、音声と騒音のモデルが常に最新のものに更新されるので、環境が変化した場合にも追従することができるという効果が得られる。

【００１７】

以上説明したように、実施の形態１の音声区間検出装置によれば、入力された音声のスペクトルからある特徴量もしくは複数の異なる特徴量の組み合わせを利用して、モデルにより音声・非音声区間それぞれにおいて異なる判別値を算出する音声・非音声判別値算出手段と、判別値と予め設定した閾値とを比較することで音声・非音声を識別する音声・非音声識別手段と、音声・非音声識別手段の識別結果に基づいてモデルを適応するモデル適応手段とを備えたので、環境が変化した場合でも音声の検出を正しく行うことができる。

【００１８】

また、実施の形態１の音声区間検出装置によれば、モデルは音声モデルとノイズモデルであり、音声・非音声判別値算出手段は、これらの尤度の比または差を算出し、それを判別値とするようにしたので、音声と騒音のモデルが常に最新のものに更新されるので、環境が変化した場合にも追従することができるという効果が得られる。

【００１９】

実施の形態２．

実施の形態１の図２に示した構成では、音声とノイズの二つのモデルを用いていた。ところで、尤度比Λを求めるためには、ある特徴量の音声区間の確率密度関数ｐ（Ｘ｜Ｈ$_1$）およびノイズ区間の確率密度関数ｐ（Ｘ｜Ｈ$_0$）がそれぞれ求まる必要はなく、それらの比の確率密度関数が直接推定できればよい。このように確率密度の比を直接推定する枠組みが密度比推定であり、これについては、例えば、杉山将，“密度比に基づく機械学習の新たなアプローチ，”統計数理ｖｏｌ．５８，ｎｏ．２，ｐｐ．１４１－１５５，２０１０．といった文献（以下、この文献を参考文献１とする）に示されている手法を用いることができる。この文献は、二つの異なる分布の出力値の尤度比を算出する一般的な枠組みであるが、ここではそれを音声区間検出に応用することで新たな効果が生まれることを説明する。

【００２０】

図３は、実施の形態２の音声区間検出装置を示す構成図である。

実施の形態２の音声区間検出装置では、音声・非音声判別値算出手段１１ａが用いるモデルとして密度比モデル２５となっている点が実施の形態１と異なる部分であり、音声・非音声判別値算出手段１１ａは密度比モデル２５に基づいて尤度比を算出し、これを判別値２２として出力するよう構成されている。他の構成は図１に示した実施の形態１と同様

であるため対応する部分に同一符号を付してその説明を省略する。

【００２１】

図示のように、実施の形態１では二つあったモデルが一つに減らせている。密度比モデル２５は事前の学習データにより学習しておく。この際に学習環境とのミスマッチを低減するために、密度比モデル２５を適応することで、推定精度を向上させることができる。

【００２２】

次に、密度比の推定法に関して説明する。密度比の推定にはいくつかの手法が提案されているが、ここでは最も基本的なカルバックライブラー重要度推定法（Ｋｕｌｌｂａｃｋ－Ｌｅｉｂｌｅｒ ｉｍｐｏｒｔａｎｃｅ ｅｓｔｉｍａｔｉｏｎ ｐｒｏｃｅｄｕｒｅ；ＫＬＩＥＰ）を例に述べる。

ＫＬＩＥＰでは、２組のデータ列 $\{x_i\}_{i=1}^{n}$ $\{x'_j\}_{j=1}^{n'}$ の確率密度関数ｑ，ｑ’の比 $r(x)=\frac{q'(x)}{q(x)}$ を下式（１）の線形モデルでモデル化する。

$$\hat{r}(x)=\sum_{l=1}^{b}\alpha_l\varphi_l(x) \qquad (1)$$

【００２３】

αは非負のパラメータであり、ψは非負の値をとる基底関数である。ｘ′が音声に対応する特徴量のデータ列であり、ｘがノイズに対応する特徴量のデータ列であると考え、参考文献１の方法に従って基底関数ψを事前に求める。参考文献１では下式のガウシアンカーネルKを用いその幅σの検定にはクロスバリデーションを使っている。

$$K_\sigma(x,x')=\exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$$

上式において、||・||はユークリッドノルムである。

【００２４】

ここで特徴量としてはパワー、ＦＦＴ後の帯域別パワー（ｌｏｇパワー）、ＭＦＣＣ、ゼロ交差数など音声とノイズを分けるのに有効な特徴量であればいずれを用いてもよい。またそれらを組み合わせて用いることもできる。ここではｌｏｇパワーを用いた場合を説明する。ノイズに比べて音声のパワーは狭い帯域に偏る傾向があるため、この特徴量を用いることでトータルのエネルギーとしてのＳ／Ｎを見るよりも判別性能を向上させることができる。

ｌｏｇパワーの特徴量のベクトルをＬＰ＝（$ＬＰ_0$，$ＬＰ_1$，…，$ＬＰ_{L-1}$）とする。

学習の際は音声・非音声でラベル付されたデータ $\{x_i\}_{i=1}^{n}$ $\{x'_j\}_{j=1}^{n'}$ よりα，φを学習する。

【００２５】

この方法で学習された密度比モデルは実際の環境で動作させる際にはミスマッチがあることが多いので、モデル適応手段１３でモデルを適応する必要がある。最も単純には学習時のノイズに対応する特徴量の平均$\frac{1}{N}\sum LP$と、始めの数フレームから算出された実際の環境の平均を取りその差を引き去ることで適応できる。またφとしてガウシアンカーネルを用いた場合は、分散σを正規化することでも適応できる。

【００２６】
実施の形態１ではどちらかのモデルの推定に誤りが生じた場合に推定精度が大きく低下する。それに対して実施の形態２では、密度比モデル２５の一つだけから尤度比を直接計算できるので、判別値算出が頑健に行えるようになる。そのため、モデルの学習環境と実環境がマッチしている場合には従来法よりも性能が良くなる効果が得られる。加えてモデル適応手段１３を備えたことにより、密度比モデル２５を適応することができるようになる。上述の通り、マッチしていれば従来法よりも性能が良くなる効果が得られるので、モデル適応によりモデルの学習環境と実環境を近づけることで、従来法よりも性能が良くなる効果が得られる。

【００２７】
以上説明したように、実施の形態２の音声区間検出装置によれば、モデルは、予め音声とノイズのデータから学習した密度比モデルであり、音声・非音声判別値算出手段は、密度比モデルに基づいて尤度比を算出し、判別値とするようにしたので、判別値算出が頑健に行えるようになり、頻繁にモデル適応をしなくてもよいといった効果を得ることができる。

【００２８】
実施の形態３．
音声とノイズの組み合わせによって、尤度比モデルや音声・ノイズモデルが異なるものになるのは明らかであるから、これらを複数用意し、学習環境が実際の環境に最も近いものを選び出すようにしてもよく、このような例を実施の形態３として以下に説明する。

【００２９】
図４は、実施の形態３の音声区間検出装置を示す構成図である。
図示の音声区間検出装置は、それぞれ学習環境が異なる複数の密度比モデル２５－１～２５－ｎを選択するモデル選択手段１４を備えている。それ以外の構成は図３に示した実施の形態２と同様であるため、対応する部分に同一符号を付してその説明を省略する。なお、実施の形態１の構成に追加するようにしてもよく、この場合は、図示のように、モデル選択手段１４は音声モデル２４ａ－１とノイズモデル２４ｂ－１～音声モデル２４ａ－ｎとノイズモデル２４ｂ－ｎのいずれかを選択する。

【００３０】
モデル選択手段１４は、モデルの学習環境が実際の環境に最も近いものを選び出す手段であるが、学習環境が実際の環境に近いかどうかを測る基準には、例えば背景ノイズの類似性、話者の類似性を用いることができる。まず、学習時のノイズと音声のモデルを作成しておく（いずれか一方でもよい）。例えば、下式のＧＭＭ（Ｇａｕｓｓｉａｎ　Ｍｉｘｔｕｒｅ　Ｍｏｄｅｌ）により、ＥＭアルゴリズムなどを用いてモデル化することができる。

$$p(x)=\sum_{k=1}^{K}\pi_k \mathrm{N}(\mu_k,\Sigma_k)$$

上式において、Ｎは平均μ_k，共分散Σ_k，混合率π_kの正規分布である。

【００３１】

モデル選択手段１４は、最も尤度ｐ（ｘ）の高いモデルに対応する密度比モデル２５－１～２５－ｎを選択することで、学習時と実際の環境のミスマッチの少ないモデルを選び出すことができる。また、ノイズモデル２４ｂ－１～２４ｂ－ｎはノイズが観測されれば一通りに選べるが、音声モデル２４ａ－１～２４ａ－ｎは音声が入力するまで選べないので、あるノイズに対する複数の音声モデル２４ａ－１～２４ａ－ｎを同時に用いて音声・非音声識別手段１２で音声区間検出を行い、検出された部分の特徴量から音声モデル２４ａ－１～２４ａ－ｎを選び出すこともできる。

【００３２】

以上説明したように、実施の形態３の音声区間検出装置によれば、学習環境が異なる複数のモデルを用意すると共に、実際の環境に対応して複数のモデルから特定のモデルを選択するモデル選択手段を備えたので、環境にマッチしたモデルを選択することができ、モデル適応の効果がより出やすいという効果が得られる。

【００３３】

また、実施の形態３の音声区間検出装置によれば、モデル適応手段は、音声・非音声識別手段で非音声または音声と判定されたフレームの特徴量を用いてモデルの適応を行うようにしたので、種々の環境でモデルの適応を行うことができる。

【００３４】

実施の形態４．

音声・非音声判別値算出手段１１ａと音声・非音声識別手段１２による音声・非音声の判別基準と、モデル適応手段１３における判別基準が同じであると、音声・非音声判別値算出手段１１ａや音声・非音声識別手段１２での誤りがモデル適応手段１３の適応によって訂正される効果を期待できなくなる。そこで、別のＭ個の特徴量により音声区間検出を行い、その結果に従ってモデルの適応を行うことで音声・非音声判別値算出手段１１ａと音声・非音声識別手段１２の誤りを正す効果が期待できる。別のＭ個の特徴量としては、例えば音声・非音声判別値算出手段１１ａで帯域別のパワーを用いた場合には、これとは異なるゼロ交差数を用いればよい。このような例を実施の形態４として図５に示す。

【００３５】

図５に示す音声区間検出装置では、実施の形態２の構成に加えて、音声・非音声判別値算出および識別手段１５を備えたものである。音声・非音声判別値算出および識別手段１５は、音声・非音声判別値算出手段１１ａと音声・非音声識別手段１２の機能を備えたものであり、これら音声・非音声判別値算出手段１１ａおよび音声・非音声識別手段１２とは別の特徴量（Ｎ＋１～Ｎ＋Ｍ）２１ａに基づいて音声と非音声とを識別する手段である。また、モデル適応手段１３ａは、音声・非音声判別値算出および識別手段１５から出力されるモデル適応可否フラグが真である場合にのみモデル適応を行うよう構成されている。なお、図５では実施の形態２に対して音声・非音声判別値算出および識別手段１５を追加した例を示しているが、実施の形態１に対して追加するよう構成してもよい。

【００３６】

このように構成された音声区間検出装置では、音声・非音声判別値算出および識別手段１５において、モデルを更新することができると判断された場合（確実に非音声または音声と判定された場合）のみモデル適応可否フラグを真とする。モデル適応手段１３ａはモデル適応可否フラグが真であった場合にのみモデル適応を行う。

【００３７】

以上説明したように、実施の形態４の音声区間検出装置によれば、音声・非音声判別値算出手段および音声・非音声識別手段とは別の特徴量に基づいて音声と非音声とを識別する音声・非音声判別値算出および識別手段を備え、モデル適応手段は、音声・非音声判別値算出および識別手段で非音声または音声と判定されたフレームの特徴量に基づいてモデルの適応を行うようにしたので、判別誤りを訂正することができ、モデル適応がより効果的に行われるという効果が得られる。

【００３８】
実施の形態５．
実施の形態１においては音声モデルがあるのでその情報を積極的に使うことで、モデル適応の可否を判断することができる。図６のように信号が観測されたとする。図６の上図は信号波形、下図は適当な手段を用いて算出された尤度比Ｒである。閾値ηを超えたものを音声区間とみなすことにする。実際の音声は上図Ｖｏｉｃｅの部分であるので、音声区間の終端がノイズと誤って識別されている。このままモデル適応を行うとノイズモデルの適応データに音声データが混ざり、検出性能が著しく低下する。そこで、音声区間中の音声モデル（Ｓｐｅｅｃｈ ｍｏｄｅｌ）とノイズモデル（Ｎｏｉｓｅ ｍｏｄｅｌ）を比較することによって類似している場合には、モデル適応を行わないことにする。
【００３９】

例えば、実施の形態１では $\chi = \sum_k \frac{\lambda_N + \lambda_S}{\lambda_N}$ をモデル間の類似性を測る尺度として用いることができる。

図６の例ではＮｏｉｓｅ ｍｏｄｅｌ １においては音声の学習データが混入しているためλ$_N$が大きくなってχが小さくなり、Ｎｏｉｓｅ ｍｏｄｅｌ１とＳｐｅｅｃｈ ｍｏｄｅｌが類似していると判断される。このため適応が不可となる。Ｎｏｉｓｅ ｍｏｄｅｌ２は正しく適応されているため、類似性が小さく適応が可となる。
【００４０】
もしくは予め音声と分かっているテストデータを用意しておき、それに対する尤度ｐ（ｘ｜Ｈ$_0$）とｐ（ｘ｜Ｈ$_1$）を測り、それぞれの尤度差（比）による方法も考えられる。この方法を実施する音声区間検出装置を図７に示す。図７において、音声モデル２６ａ、ノイズモデル２６ｂは、予め音声と分かっているテストデータに基づいて適応したモデルである。モデル識別手段１６は、これら音声モデル２６ａとノイズモデル２６ｂとが類似しているか否かを識別し、類似していないと判断した場合にモデル適応可否フラグを真とする手段である。また、モデル更新手段１７ａ，１７ｂは、モデル適応可否フラグが真であった場合にそれぞれ音声モデル２４ａとノイズモデル２４ｂとを更新する手段である。なお、モデル適応手段１３、モデル識別手段１６およびモデル更新手段１７ａ，１７ｂは、予め作成された音声モデルと適応したノイズモデルとを比較し、その比較結果に基づいて適応の可否を判断するモデル適応手段を構成している。これ以外の構成は図２に示した実施の形態１の構成と同様であるため、対応する部分に同一符号を付してその説明を省略する。
【００４１】
このように、ノイズモデル２６ｂと音声モデル２６ａの双方を識別するモデル識別手段１６により、それらのモデルが類似していないと判断された場合には音声モデル２４ａとノイズモデル２４ｂを更新する。従って、ノイズモデルの適応データに音声データが混入し検出性能が著しく低下するのを避けることができる効果が得られる。
【００４２】
以上説明したように、実施の形態５の音声区間検出装置によれば、モデル適応手段は、予め作成された音声モデルと適応したノイズモデルとを比較し、比較結果に基づいて適応の可否を判断するようにしたので、ノイズモデルの適応データに音声データが混入することで検出性能が著しく低下するといったことを避けることができる。
【００４３】
なお、本願発明はその発明の範囲内において、各実施の形態の自由な組み合わせ、あるいは各実施の形態の任意の構成要素の変形、もしくは各実施の形態において任意の構成要

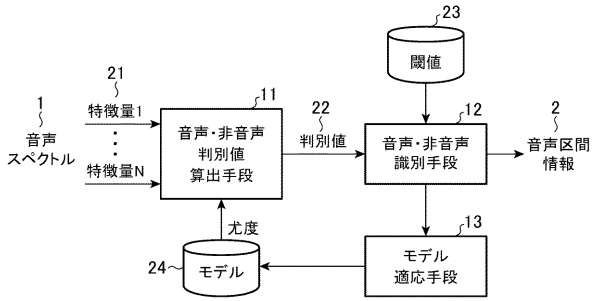
素の省略が可能である。

【符号の説明】

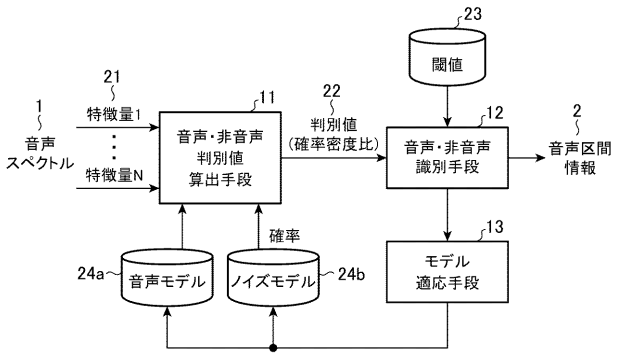
【００４４】

１　音声スペクトル、２　音声区間情報、１１，１１ａ　音声・非音声判別値算出手段、１２　音声・非音声識別手段、１３，１３ａ　モデル適応手段、１４　モデル選択手段、１５　音声・非音声判別値算出および識別手段、１６　モデル識別手段、１７ａ，１７ｂ　モデル更新手段、２１，２１ａ　特徴量、２２　判別値、２３　閾値、２４　モデル、２４ａ，２６ａ　音声モデル、２４ｂ，２６ｂ　ノイズモデル、２５，２５－１～２５－ｎ　密度比モデル。

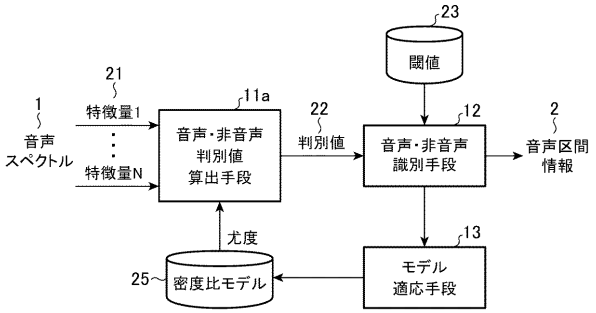
【図１】



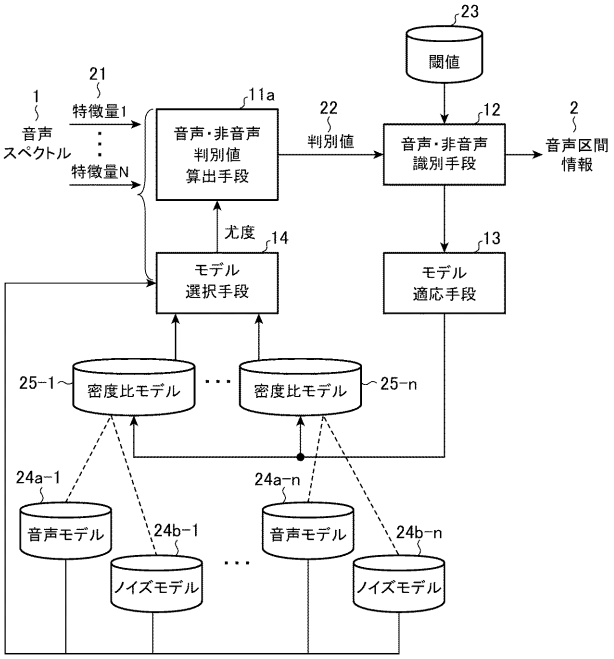
【図２】



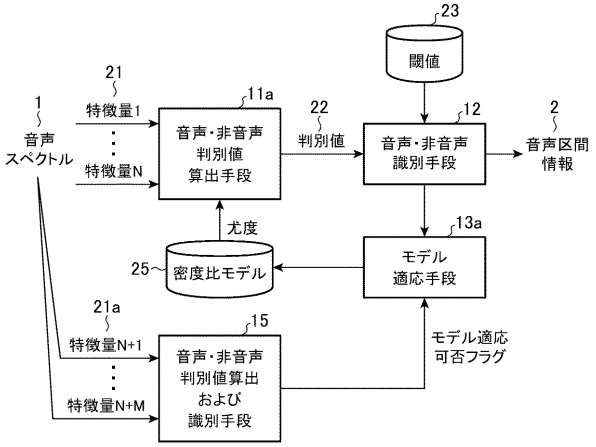
【図３】



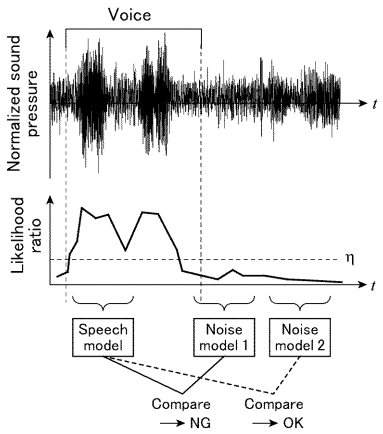
【図４】


1
音声
スペクトル
21
特徴量1
特徴量N
11a
音声・非音声
判別値
算出手段
22
判別値
12
音声・非音声
識別手段
2
音声区間
情報
23
閾値
尤度
14
モデル
選択手段
13
モデル
適応手段
25-1
密度比モデル
25-n
密度比モデル
24a-1
音声モデル
24b-1
ノイズモデル
24a-n
音声モデル
24b-n
ノイズモデル

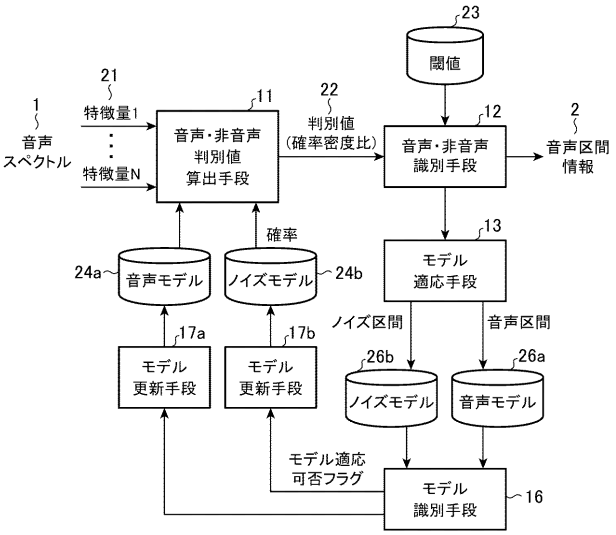
【図５】


1
音声
スペクトル
21
特徴量1
特徴量N
11a
音声・非音声
判別値
算出手段
22
判別値
12
音声・非音声
識別手段
2
音声区間
情報
23
閾値
尤度
25
密度比モデル
13a
モデル
適応手段
21a
特徴量N+1
特徴量N+M
15
音声・非音声
判別値算出
および
識別手段
モデル適応
可否フラグ

【図６】


Voice
Normalized sound pressure
t
Likelihood ratio
η
t
Speech model
Noise model 1
Noise model 2
Compare
NG
Compare
OK

【図７】


1
音声
スペクトル
21
特徴量1
特徴量N
11
音声・非音声
判別値
算出手段
22
判別値
(確率密度比)
12
音声・非音声
識別手段
2
音声区間
情報
23
閾値
確率
24a
音声モデル
24b
ノイズモデル
13
モデル
適応手段
ノイズ区間
音声区間
17a
モデル
更新手段
17b
モデル
更新手段
26b
ノイズモデル
26a
音声モデル
モデル適応
可否フラグ
モデル
識別手段
16