

(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2013-160938
(P2013-160938A)

(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)
G 1 0 L 15/04	(2013.01)	G 1 0 L 15/04	3 0 0 C	5 D 0 1 5
G 1 0 L 25/78	(2013.01)	G 1 0 L 11/00	4 0 2 L	
		G 1 0 L 11/02		

審査請求 未請求 請求項の数 6 O L (全 8 頁)

<p>(21) 出願番号 特願2012-23136 (P2012-23136)</p> <p>(22) 出願日 平成24年2月6日 (2012.2.6)</p>	<p>(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</p> <p>(74) 代理人 100123434 弁理士 田澤 英昭</p> <p>(74) 代理人 100101133 弁理士 濱田 初音</p> <p>(74) 代理人 100173934 弁理士 久米 輝代</p> <p>(74) 代理人 100156351 弁理士 河村 秀央</p> <p>(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</p> <p>Fターム(参考) 5D015 DD03</p>
--	--

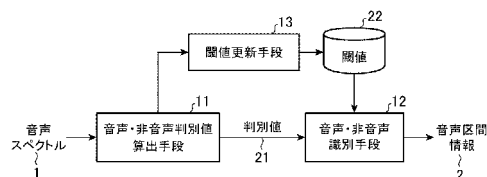
(54) 【発明の名称】 音声区間検出装置

(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、音声と騒音の混在した信号から音声区間と非音声区間に対応した判別値21を算出する。音声・非音声識別手段12は、判別値21と閾値22とを比較することで音声・非音声を識別する。閾値更新手段13は、判別値21に基づいて閾値22を動的に決定する。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

音声と騒音の混在した信号から音声区間と非音声区間に対応した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

、
前記閾値を前記判別値に基づいて動的に決定する閾値更新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】

閾値更新手段は、過去の履歴の音声区間における判別値と騒音区間における判別値の分離性が最も高くなる値に閾値に更新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。

10

【請求項 3】

閾値更新手段は、音声区間と騒音区間のクラスタリング分析を行い、その結果から更新する閾値の値を決定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。

【請求項 4】

騒音のレベルに対応した複数の閾値を有すると共に、当該複数の閾値を選択する閾値選択手段を備え、閾値更新手段は、前記閾値選択手段が選択した閾値の更新を行う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】

騒音の種類に対応した複数の閾値を有すると共に、前記騒音の種類に応じて前記複数の閾値を選択する閾値選択手段を備え、閾値更新手段は、前記閾値選択手段が選択した閾値の更新を行う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。

20

【請求項 6】

音声の種類に対応した複数の閾値を有すると共に、前記音声の種類に応じて前記複数の閾値を選択する閾値選択手段を備え、閾値更新手段は、前記閾値選択手段が選択した閾値の更新を行う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。

【発明の詳細な説明】

30

【技術分野】

【0001】

本発明は、音声と騒音とが混在した信号から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、複数特徴量を並列に利用したもの（例えば、特許文献 1 参照）や、別途学習した音声モデルを用いたもの（例えば、特許文献 2 参照）が提案されている。いずれにおいても閾値との比較によって音声・非音声の識別を行う。

40

【先行技術文献】

【特許文献】

【0003】

【特許文献 1】特開 2009 - 63700 号公報

【特許文献 2】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

50

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】

上記各文献に示されている手法の閾値は予め決めたものを用いるか、ノイズ区間における判別値を平均して安全率を掛けたものを用いるかといったものであった。この方法では音声の情報を用いていないため、例えば、騒音に対して音声のレベルが大きい場合（閾値を高め設定できる）と、小さい場合（閾値を低く設定する）において、最適な閾値とすることができず、音声区間の検出性能が低下するといった問題があった。

10

【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、環境が変化しても音声区間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、音声と騒音の混在した信号から音声区間と非音声区間に対応した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、閾値を判別値に基づいて動的に決定する閾値更新手段とを備えたものである。

20

【発明の効果】

【0008】

この発明の音声区間検出装置は、閾値を判別値に基づいて動的に決定する閾値更新手段を備えたので、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】音声区間検出装置において、判別分析で閾値を決定する方法を示す説明図である。

30

【図3】この発明の実施の形態1による音声区間検出装置の閾値を決定する方法を示す説明図である。

【図4】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5】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6】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。

【発明を実施するための形態】

【0010】

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、音声スペクトル1を入力し、音声区間情報2を出力するもので、音声・非音声判別値算出手段11、音声・非音声識別手段12、閾値更新手段13を備えている。音声・非音声判別値算出手段11は、音声と騒音とが混在した音声スペクトル1を入力し、音声区間と騒音区間とに対応した判別値21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21と、予め求められた閾値22とを比較し、音声と非音声とを識別し、音声区間情報2を出力する手段である。閾値更新手段13は、判別値21に基づいて、閾値22を動的に更新する手段である。

40

【0011】

音声・非音声判別値算出手段11では、入力される音声スペクトル1の特徴量から何らかの判別値21（以下、Rと表記する）を算出する。例えば非特許文献2に示されている

50

ように、複素スペクトルを特徴量として入力し、判別値 2 1 として尤度比を出力する。それを音声・非音声識別手段 1 2 で閾値 2 2 と比較して音声区間を識別する。音声区間検出には閾値 2 2 の設定が不可欠であり、環境によって最適な閾値 2 2 が異なるため設定が難しい。そこで閾値 2 2 を学習して適応する手段として閾値更新手段 1 3 を設けている。

【 0 0 1 2 】

一般的にはノイズ区間の情報から閾値 2 2 を推定するが、上記課題にて述べたとおり、ノイズだけの情報から閾値を求めると精度が低下するため、音声と騒音両方の情報を用いて閾値 2 2 を決定する。音声の情報がないと閾値 2 2 を更新することはできないので、始めは何らかの初期値に従い音声区間検出を行う。音声を検出されたら、音声と騒音それぞれラベル付けされた閾値 2 2 を並べて、何らかの基準で新しい閾値 2 2 を決定する。

10

【 0 0 1 3 】

例えば、図 2 のように音声と騒音の別にそれぞれの平均 μ と標準偏差 σ を算出する。式 (1) のように判別分析の結果により θ を求め、 $\theta > \theta_0$ 以上であれば音声、 $\theta < \theta_0$ 以下であれば騒音とする方法が考えられる。 θ_0 は定数であり、音声の検出率を向上させるか (1 以下)、誤検出率を低減させるか (1 以上) によって値を変化させる。

$$\theta = \frac{\sigma_1 \mu_2 + \sigma_2 \mu_1}{\sigma_1 + \sigma_2} \quad (1)$$

20

【 0 0 1 4 】

この方法は、音声とノイズのラベル付けが正しければ性能がよいが、ラベル付けが間違っていた場合には性能が低下する。次にラベル付け不要の方法に関して述べる。構成は図 1 と同じである。音声、騒音のラベリングをせずに過去の R を累積する。ここでは、この R は騒音の時に小さな値を取り、音声の時に大きな値をとることにする。例えば、パワーなどが考えられる。ここでは非特許文献 2 の尤度比を想定するが、これ以外を用いることもできる。この際に騒音のダイナミックレンジに比べて音声のダイナミックレンジは大きいので、式 (2) のような非線形な変換を用いてレンジをある一定の範囲内に収める。

$$R' = \beta \frac{\tan^{-1}(R/\gamma)}{\sqrt{\beta/\gamma}} \quad (2)$$

30

【 0 0 1 5 】

ここで β はある定数である。適当な初期値のもとで音声区間が検出されたら、過去の R をクラスタリングする。例えば K - m e a n s アルゴリズムを用いて複数のクラスタに分ける。分けられたクラスタの平均値と分散を計算する。一般に騒音のダイナミックレンジは音声のそれよりも小さいことを考え合わせると、音声に割り当てられるクラスタ数が多くなるはずである。平均値の大きい上から n 番目と n + 1 番目のクラスタで図 3 のように式 (1) の内分点を閾値として用いることで閾値を決定できる。すなわち、過去の履歴の音声区間における判別値と騒音区間における判別値の分離性が最も高くなる閾値に決定する。この方法ではクラスタ数を増やすことによりクラスタ内の分散を小さくすることができ、音声と騒音の分離性が高くなるため、上記判別分析による方法よりも頑健に最適な閾値の推定が行える。このように閾値更新手段 1 3 においてクラスタリングにより適切な閾値を推定することで、頑健に最適な閾値の推定が行えるため、環境変化に対応できるという効果が得られる。

40

【 0 0 1 6 】

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、音声と騒音の混在した信号から音声区間と非音声区間に対応した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

50

、閾値を判別値に基づいて動的に決定する閾値更新手段とを備えたので、環境が変化しても音声区間の検出を正しく行うことができる。

【0017】

また、実施の形態1の音声区間検出装置によれば、閾値更新手段は、過去の履歴の音声区間における判別値と騒音区間における判別値の分離性が最も高くなる値に閾値に更新するようにしたので、頑健に最適な閾値の推定を行うことができる。

【0018】

また、実施の形態1の音声区間検出装置によれば、閾値更新手段は、音声区間と騒音区間のクラスタリング分析を行い、その結果から更新する閾値の値を決定するようにしたので、頑健に最適な閾値の推定を行うことができる。

10

【0019】

実施の形態2 .

実施の形態1の構成では、徐々に変化する騒音には対応できるものの、段階的に急激に変化する騒音には対応することが困難である。例えば、ホワイトノイズのような騒音が徐々にSN比を変えながら重畳している場合には実施の形態1の構成で対応可能であるが、パブルノイズのような騒音の場合には対応が困難である。そこで、実施の形態2では、いくつかの騒音を想定し、閾値を複数用意することでこのような騒音にも対応できるようにしたものである。

【0020】

図4は、実施の形態2の音声区間検出装置の構成図であり、図1に示した実施の形態1の構成に加えて閾値選択手段14が追加されている。閾値選択手段14は、複数の閾値22-1~22-nのうち、いずれかの閾値を選択する手段である。また、閾値更新手段13aは、閾値選択手段14が選択した閾値を更新するよう構成されている。その他の構成は実施の形態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。

20

【0021】

複数の閾値22-1~22-nとして、最も単純にはSN比による方法が考えられる。例えばノイズのレベルを5dBずつ分割し、それぞれに対応する閾値を複数用意し、これを閾値22-1~22-nとする。閾値更新手段13aでは、閾値選択手段14で選択された閾値を更新する。そして、音声・非音声識別手段12では、実際の環境に合わせた閾値22-1~22-nを用い、音声・非音声の識別を行う。このように閾値選択手段14を備えたことにより、閾値22-1~22-nを非連続的に環境にあったものを選ぶことができるため、急激に変化する環境にも対応できるという効果が得られる。

30

【0022】

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、騒音のレベルに対応した複数の閾値を有すると共に、複数の閾値を選択する閾値選択手段を備え、閾値更新手段は、閾値選択手段が選択した閾値の更新を行うようにしたので、急激に変化する環境にも対応することができる。

【0023】

実施の形態3 .

実施の形態2では騒音レベルに応じて閾値を選択したが、騒音の種類により選択するものも考えられる。単純には変動性の小さな騒音に関しては閾値を下げ、変動性の大きな騒音には閾値を引き上げることが考えられる。このような例を実施の形態3として図5に示す。

40

【0024】

図5に示すように、実施の形態3の音声区間検出装置では、実施の形態2の構成に対して騒音モデル23-1~23-nが追加されている。騒音モデル23-1~23-nは、騒音の種類別のモデルであり、それぞれのモデルが閾値22-1~22-nに紐付けられている。閾値選択手段14aは、騒音モデル23-1~23-nのうちの選択された騒音モデルに紐付けられている閾値を選択するよう構成されている。その他の構成は図4に示

50

した実施の形態 2 と同様である。

【0025】

例えば、騒音モデル 23 - 1 ~ 23 - n としては、下式の GMM (Gaussian Mixture Model) が考えられる。

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

ここで、N は平均 μ_k , 共分散 Σ_k , 混合率 π_k の正規分布である。

【0026】

このモデルを想定される複数の種類の騒音に対して事前に EM アルゴリズムなどで学習しておく。閾値選択手段 14 a は、音声スペクトル 1 から騒音 GMM の尤度 $p(x)$ を計算し、最も尤度の高いものを選択し、閾値更新手段 13 a によって閾値の更新・参照を行う。このように、騒音の種類に応じた適切な閾値を選択することができるため、騒音の特性に応じて音声の検出力を変化させることができる。

【0027】

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、騒音の種類に対応した複数の閾値を有すると共に、騒音の種類に応じて複数の閾値を選択する閾値選択手段を備え、閾値更新手段は、閾値選択手段が選択した閾値の更新を行うようにしたので、騒音の特性に応じて音声区間の検出を正しく行うことができる。

【0028】

実施の形態 4 .

実施の形態 2 では騒音レベルに応じて閾値を選択したが、音声の種類により選択するものも考えられる。例えばカーナビのように比較的使用者が固定的である装置において、使用者が女性であると見当がつけば、女性の閾値を下げ、男性の閾値を上げておくというようなものが考えられ、このような例を実施の形態 4 として図 6 に示す。

【0029】

実施の形態 4 では、図示のように、図 4 に示した実施の形態 2 の構成に対して音声モデル 24 - 1 ~ 24 - n が追加されている。音声モデル 24 - 1 ~ 24 - n は、それぞれ音声の種類に対応したモデルであり、各モデルがそれぞれ閾値 22 - 1 ~ 22 - n に紐付けられている。閾値選択手段 14 b は、音声モデル 24 - 1 ~ 24 - n のうちの選択された音声モデルに紐付けられている閾値を選択するよう構成されている。その他の構成は図 4 に示した実施の形態 2 と同様である。

【0030】

音声モデル 24 - 1 ~ 24 - n についても、実施の形態 3 と同様 GMM などによってモデルを用意しておき、閾値選択手段 14 b によって最尤のものを選択し、閾値 22 - 1 ~ 22 - n の更新および参照すればよい。

【0031】

このように、複数の音声モデル 24 - 1 ~ 24 - n と閾値選択手段 14 b とを備えたことにより、話者が変わらない場合には妨害音が人間の声であっても対象とする話者の音声だけを検出できるという効果が得られる。

【0032】

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、音声の種類に対応した複数の閾値を有すると共に、音声の種類に応じて複数の閾値を選択する閾値選択手段を備え、閾値更新手段は、閾値選択手段が選択した閾値の更新を行うようにしたので、音声の特性に応じて音声区間の検出を正しく行うことができる。

【0033】

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。

10

20

30

40

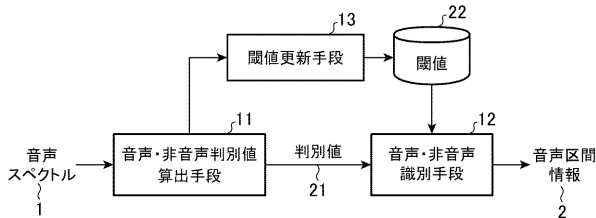
50

【符号の説明】

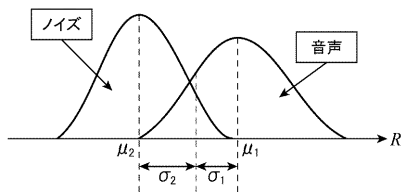
【0034】

1 音声スペクトル、2 音声区間情報、11 音声・非音声判別値算出手段、12 音声・非音声識別手段、13, 13a 閾値更新手段、14, 14a, 14b 閾値選択手段、21 判別値、22, 22-1~22-n 閾値、23-1~23-n 騒音モデル、24-1~24-n 音声モデル。

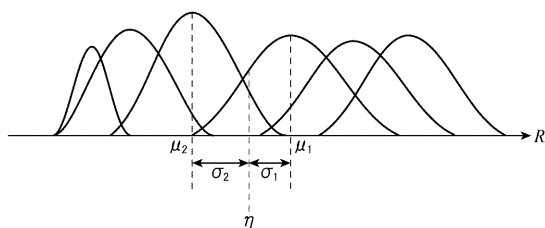
【図1】



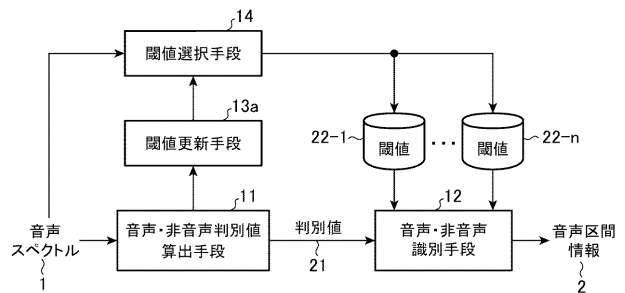
【図2】



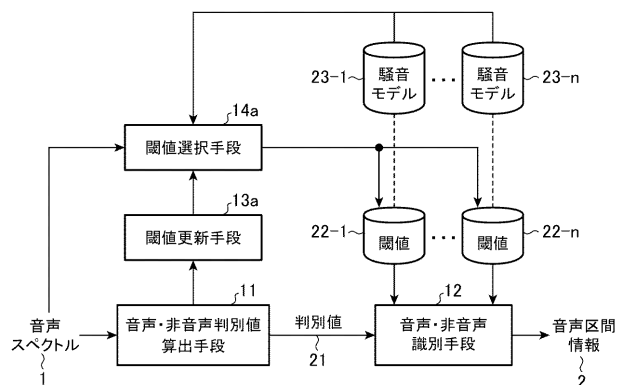
【図3】



【図4】



【図5】



【 図 6 】

