

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4875656号
(P4875656)

(45) 発行日 平成24年2月15日(2012.2.15)

(24) 登録日 平成23年12月2日(2011.12.2)

(51) Int.Cl. F I
G 1 O L 11/02 (2006.01) G 1 O L 11/02
G 1 O L 21/02 (2006.01) G 1 O L 21/02 2 O 2 Z
G O 1 S 3/808 (2006.01) G O 1 S 3/808

請求項の数 10 (全 16 頁)

(21) 出願番号	特願2008-119717 (P2008-119717)	(73) 特許権者	000004226
(22) 出願日	平成20年5月1日(2008.5.1)		日本電信電話株式会社
(65) 公開番号	特開2009-271183 (P2009-271183A)		東京都千代田区大手町二丁目3番1号
(43) 公開日	平成21年11月19日(2009.11.19)	(74) 代理人	100121706
審査請求日	平成22年7月26日(2010.7.26)		弁理士 中尾 直樹
		(74) 代理人	100128705
			弁理士 中村 幸雄
		(74) 代理人	100147773
			弁理士 義村 宗洋
		(74) 代理人	100066153
			弁理士 草野 卓
		(72) 発明者	荒木 章子
			東京都千代田区大手町二丁目3番1号 日 本電信電話株式会社内

最終頁に続く

(54) 【発明の名称】 信号区間推定装置とその方法と、プログラムとその記録媒体

(57) 【特許請求の範囲】

【請求項1】

複数のマイクロホンで収録された複数の音源からの音声信号から、各音源から信号が発せられている区間を推定する信号区間推定装置であって、

上記音声信号を、フレーム毎に周波数領域の信号に変換する周波数領域変換部と、

上記フレーム毎の周波数領域の信号からフレーム毎に音声の存在確率を推定する音声存在確率推定部と、

上記フレーム毎の周波数領域の信号からフレーム毎に各周波数成分についての音声到来方向を推定する到来方向推定部と、

上記各周波数成分についての音声到来方向からフレーム毎に上記各音源に関する音声到来方向確率を計算する到来方向確率計算部と、

上記音声存在確率と、上記音声到来方向確率との積を計算して上記各フレームにおける音源毎の存在確率を出力する乗算部と、

上記音源毎の存在確率が所定の閾値を越える場合に、当該フレームを当該音源の信号区間と判定する判定部と、

を具備する信号区間推定装置。

【請求項2】

請求項1に記載の信号区間推定装置において、

上記到来方向確率計算部は、クラスタリング部と、各クラスタの分布計算部と、確率計算部を備え、

上記クラスタリング部は、上記フレーム毎に計算された各周波数での音源方向をクラスタリングすることにより、上記音源方向をクラスタに分けるものであり、

上記各クラスタの分布計算部は、上記クラスタリング部で得られた各クラスタの上記音源方向の分布を計算するものであり、

上記確率計算部は、上記クラスタ毎の上記音源方向の分布を、上記音源方向の全体の分布で正規化して音声到来方向確率として出力するものである、

ことを特徴とする信号区間推定装置。

【請求項 3】

請求項 1 に記載の信号区間推定装置において、

上記到来方向確率計算部は、クラスタリング部と、確率計算部を備え、

上記クラスタリング部は、上記フレーム毎に計算された各周波数での音源方向と閾値との距離で上記音源方向をクラスタリングするものであり、

上記確率計算部は、上記クラスタの各メンバ数を上記音源方向の全体のメンバ数で除した値を音声到来方向確率として出力するものであることを特徴とする信号区間推定装置。

【請求項 4】

請求項 1 乃至 3 の何れかに記載した信号区間推定装置において、

上記到来方向確率計算部は、上記各フレーム、各周波数における正規化された振幅値を計算する振幅計算部を備え、

上記音声到来方向確率を計算する際に、上記正規化された振幅値を重み係数として用いることを特徴とする信号区間推定装置。

【請求項 5】

複数のマイクロホンで収録された複数の音源からの音声信号から、各音源から信号が発せられている区間を推定する信号区間推定方法であって、

周波数領域変換部が、上記音声信号を、フレーム毎に周波数領域の信号に変換する周波数領域過程と、

音声存在確率推定部が、上記フレーム毎の周波数領域の信号からフレーム毎に音声存在確率を推定する音声存在確率推定過程と、

到来方向推定部が、上記フレーム毎の周波数領域の信号からフレーム毎に各周波数成分についての音声到来方向を推定する到来方向推定過程と、

到来方向確率計算部が、上記各周波数成分についての音声到来方向からフレーム毎に上記各音源に関する音声到来方向確率を計算する到来方向確率計算過程と、

乗算部が、上記音声存在確率と、上記音声到来方向確率との積を計算して上記各フレームにおける音源毎の存在確率を出力する乗算過程と、

判定部が、上記音源毎の存在確率が所定の閾値を越える場合に、当該フレームを当該音源の信号区間と判定する判定過程と、

を含む信号区間推定方法。

【請求項 6】

請求項 5 に記載の信号区間推定方法において、

上記到来方向確率計算過程は、クラスタリング部が、上記フレーム毎に計算された各周波数での音源方向をクラスタリングすることにより、上記音源方向をクラスタに分ける分類ステップと、

各クラスタの分布計算部が、上記クラスタの上記音源方向の分布を計算する分布計算ステップと、

確率計算部が、上記クラスタ毎の上記音源方向の分布を、上記音源方向の全体の分布で正規化して音声到来方向確率として計算する確率計算ステップと、

を含むことを特徴とする信号区間推定方法。

【請求項 7】

請求項 5 に記載した信号区間推定方法において、

上記到来方向確率計算過程は、クラスタリング部が、上記フレーム毎に計算された各周波数での音源方向をクラスタリングする分類ステップと、

確率計算部が、上記クラスタの各メンバ数を上記音源方向の全体のメンバ数で除した値を音声到来方向確率として計算する確率計算ステップと、
を含むことを特徴とする信号区間推定方法。

【請求項 8】

請求項 5 乃至 7 の何れかに記載した信号区間推定方法において、

上記到来方向確率計算過程は、振幅計算部が、上記各フレーム、各周波数における正規化された振幅値を計算する振幅計算ステップを含み、

上記正規化された振幅値を重み係数として用いて上記音声到来方向確率を計算することを特徴とする信号区間推定方法。

【請求項 9】

請求項 1 乃至 4 の何れかに記載した信号区間推定装置としてコンピュータを機能させるためのプログラム。

【請求項 10】

請求項 9 に記載したプログラムを記録したコンピュータで読み取り可能な記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、複数人による会話などを複数のマイクロホンで収録し、「いつ誰が話したか」を推定する技術に関連する。特に、複数の音源からの信号が混在しているデータについて、各音源から信号が発せられている区間を推定する複数信号区間推定装置とその方法と、プログラムと記録媒体に関する。

【背景技術】

【0002】

複数人の発話者の各話者の発言している音声区間を検出する技術は、例えば会議録自動作成において各発言に発話者を自動的に付与したり、会議収録データに話者情報を付与して録音データの検索や頭出しを容易にしたりする際に重要である。

【0003】

従来、音源方向推定装置として、非特許文献 1 に開示された方法が知られている。図 10 に非特許文献 1 の音源方向推定装置 200 の機能構成を示して簡単に説明する。音源方向推定装置 200 は、周波数領域変換部 201 と、音声区間推定部 202 と、到来方向推定部 203 と、到来方向分類部 204 を備える。周波数領域変換部 201 は、離散値化された複数のマイクロホンで収録された観測信号を例えば 32ms 毎に窓関数で切り出したあと（切り出した 1 区間を以降、「フレーム」と称する。）、観測信号をフーリエ変換などで周波数領域の信号に変換する。音声区間推定部 202 は、周波数領域に変換された観測信号から音声区間を推定する。到来方向推定部 203 は、音声区間とされた各フレームの観測信号から音声到来方向を推定する。到来方向分類部 204 は、音声到来方向が類似した音声区間をその方向の話者が話した区間として出力する。

【非特許文献 1】S.Araki,H.Sawada,and S.Makino," Blind speech separation in a meeting situation with maximum SNR beamformers," ICASSP2007,vol.1,pp.41-44, Apr.2007

【発明の開示】

【発明が解決しようとする課題】

【0004】

従来、方法では次の問題点があった。その 1 つは、音声区間推定部 202 が、音声区間か否かを決定論的に出力する点である。これは音声区間推定部 202 において、音声区間であるのにそうではないと判定する誤棄却や、音声が無い区間を音声区間と判定する誤受理の推定誤りが発生することを意味する。非音声区間と判定されたフレームは以後扱われないので、誤棄却は音声の取りこぼしの原因になる。一般に誤棄却と誤受理は、トレードオフの関係にある。誤棄却と誤受理との関係を複数話者の音声区間検出に適したものに設定することは大変難しく、従来、方法では音声区間の取りこぼしが発生していた。

10

20

30

40

50

【 0 0 0 5 】

また、問題点の2つ目としては、到来方向推定部203が、各フレームにおいて1つの到来方向しか出力しないため、フレーム内に複数人の発言が混在する場合でも1つの到来方向の情報しか得られない。このため、検出されなかった方向からの話者についての音声区間を取りこぼしてしまう。このように従来の方法では、音声区間推定部202と到来方向推定部203のそれぞれに、音声区間を欠損させてしまう問題点があった。

この発明は、このような点に鑑みてなされたものであり、音声区間を欠損させることのない複数信号区間推定装置と、その方法とプログラムと、その記録媒体を提供することを目的とする。

【課題を解決するための手段】

10

【 0 0 0 6 】

この発明の信号区間推定装置は、複数のマイクロホンで収録された複数の音源からの音声信号から、各音源から信号が発せられている区間を推定するものであって、周波数領域変換部と、音声存在確率推定部と、到来方向推定部と、到来方向確率計算部と、乗算部と、判定部とを具備する。周波数領域変換部は、音声信号をフレーム毎に周波数領域の信号に変換する。音声存在確率推定部は、フレーム毎の周波数領域の信号からフレーム毎に音声の存在確率を推定する。到来方向推定部は、フレーム毎の周波数領域の信号からフレーム毎に各周波数成分についての音声到来方向を推定する。到来方向確率計算部は、各周波数成分についての音声到来方向からフレーム毎に各音源に関する音声到来方向確率を計算する。乗算部は、音声存在確率と、音声到来方向確率との積を計算して各フレームにおける音源毎の存在確率を出力する。判定部は、音源毎の存在確率が所定の閾値を超える場合に、当該フレームを当該音源の信号区間と判定する。

20

【発明の効果】

【 0 0 0 7 】

この発明の複数信号区間推定装置は、各フレームにおける音声存在確率を計算すると共に、各フレームの全ての周波数における音声到来方向確率を推定する。そして、音声存在確率と音声到来方向確率を乗算した値を音源毎の発音確率として出力する。これにより、音声区間検出部の決定的な推定誤りによる性能低下を防げる。また、各フレームで複数の音源の到来方向を確率的に推定できる。よって、音声区間の取りこぼしの少ない複数信号区間推定装置を実現することができる。

30

【発明を実施するための最良の形態】

【 0 0 0 8 】

以下、この発明の実施の形態を図面を参照して説明する。複数の図面中同一のものには同じ参照符号を付し、説明は繰り返さない。

【実施例1】

【 0 0 0 9 】

図1にこの発明の複数信号区間推定装置100の機能構成例を示す。図2に動作フローを示す。複数信号区間推定装置100は、周波数領域変換部11と、音声存在確率推定部12と、到来方向推定部13と、到来方向確率計算部14と、乗算部15とを具備する。周波数領域変換部11に入力される観測信号 $x()$ は、複数のマイクロホンで収録された複数の音源からの音声信号であり、例えばサンプリング周波数16kHzで離散値化された信号である。図1では観測信号を離散値化するAD変換器については省略している。複数信号区間推定装置100は、例えばROM、RAM、CPU等で構成されるコンピュータに所定のプログラムが読み込まれて、CPUがそのプログラムを実行することで実現されるものである。

40

【 0 0 1 0 】

周波数領域変換部11は、離散値化された観測信号を、例えば512点毎に窓関数で切り出し、フーリエ変換などで周波数領域の信号に変換する(ステップS11、図2参照)。この場合、フレーム長は $512/16\text{kHz} = 32\text{ms}$ である。音声存在確率推定部12は、各フレーム $()$ における音声の存在確率 $p_v()$ を推定する(ステップS12

50

)。到来方向推定部 1 3 は、各フレームの各周波数成分についての音声到来方向 $q(f, \tau)$ を推定する(ステップ S 1 3)。到来方向確率計算部 1 4 は、音声到来方向を分類して各音源に関する音声到来方向の分布を求め、各音源に関する音声到来方向確率 $p_k(f, \tau)$ を計算する(ステップ S 1 4)。乗算部 1 5 は、音声存在確率 $p_v(f, \tau)$ と、音声到来方向確率 $p_k(f, \tau)$ との積を計算して音源毎の発音確率 $P_k(f, \tau)$ を出力する(ステップ S 1 5)。

【0011】

以上のように音声区間を音声存在確率として、また音声到来方向を各フレームで複数音源に関する音声到来方向確率として処理するのでフレームが欠損することが少ない。つまり音声信号の取りこぼしを少なくした複数信号区間推定装置が実現できる。以下、複数信号区間推定装置 100 の各部の動作を詳しく説明する。ただし、周波数領域変換部 1 1 と乗算部 1 5 については、従来技術で簡単に構成できるので、詳しい説明は省略する。

10

【0012】

〔音声存在確率推定部〕

図 3 に音声存在確率推定部 1 2 の機能ブロックを示す。音声存在確率推定部 1 2 は、GMM パラメータ記録部 1 2 0 と、カルマンフィルタ 1 2 1 と、GMM 尤度計算部 1 2 2 と、単一ガウス分布尤度計算部 1 2 3 と、推移確率記録部 1 2 4 と、前向き確率算出部 1 2 5 と、前向き確率保持部 1 2 6 を備える。音声存在確率推定部 1 2 は、入力の特徴ベクトルを混合ガウス分布で表現した GMM (Gaussian Mixture model) を用いて、式 (1) と (2) に示すように音声存在確率 $p_v(\tau)$ を前向き確率 $p_j(\tau)$ として算出するものである。

20

【0013】

〔数 1〕

$$p_v(\tau) = \alpha_1(\tau) \quad (1)$$

$$\alpha_j(\tau) = \sum_{i=0}^1 a_{ij} \alpha_i(\tau-1) b_j(\tau) \quad (2)$$

ここで、時刻 $\tau-1$ における音声状態(音声+雑音の状態)を $i=1$ とし、非音声状態(無音+雑音の状態)を $i=0$ とする。また時刻 τ における音声状態を $j=1$ とし、非音声状態を $j=0$ とする。 a_{ij} は時刻 $\tau-1$ での状態 i から j に推移する推移確率である。 $b_j(\tau)$ は音声 GMM 又は非音声 GMM の出力確率である。

30

【0014】

カルマンフィルタ 1 2 1 は、観測信号 $x(f, \tau)$ と GMM パラメータを入力として、時刻 $\tau-1$ における音声/非音声 GMM から、時刻 τ における各ガウス分布(k 番目)の平均値 μ_{jmk} と分散値 Σ_{jmk} を推定する。単一ガウス分布尤度計算部 1 2 3 は、ガウス分布の平均値 μ_{jmk} と分散値 Σ_{jmk} を入力として各ガウス分布の尤度 $b_{jk}(f, \tau)$ を式 (3) で計算する。

【0015】

〔数 2〕

$$b_{jk}(\tau) = \prod_m \frac{1}{\sqrt{2\pi\Sigma_{jmk\tau}}} \exp\left\{-\frac{(x(m,\tau) - \mu_{jmk\tau})^2}{2\Sigma_{jmk\tau}}\right\} \quad (3)$$

40

ここで $x(m, \tau)$ はフレーム τ の m 次のメルスペクトルである。

GMM 尤度計算部 1 2 2 は、各ガウス分布の尤度 $b_{jk}(f, \tau)$ と、重み係数 w_{jk} を入力として音声 GMM $b_1(f, \tau)$ 及び非音声 GMM $b_0(f, \tau)$ の尤度 $b_j(f, \tau)$ を式 (4) で計算する。

【0016】

【数3】

$$b_j(\tau) = \sum_k \omega_{jk} b_{jk}(\tau) \quad (4)$$

前向き確率算出部125は、音声GMM $b_1(\cdot)$ と、非音声GMM $b_0(\cdot)$ と、前向き確率保持部126に記録された1時刻前の前向き確率 $p_j(\tau-1)$ と、推移確率記録部124に記録された推移確率 a_{ij} を入力として、式(2)で得られた前向き確率 $p_j(\tau)$ を音声存在確率 $p_v(\tau)$ として出力する。

なお、音声存在確率 $p_v(\tau)$ を式(5)に示す演算で求めても良い。

【0017】

【数4】

$$p_v(\tau) = \frac{\Lambda(\tau)}{1 + \Lambda(\tau)} \quad (5)$$

ここで $\Lambda(\tau)$ は、式(6)と(7)で表わせる。

【0018】

【数5】

$$\log \Lambda(\tau) = \frac{1}{L} \sum_{f=0}^{L-1} \{\gamma(f, \tau) - \log \gamma(f, \tau) - 1\} \quad (6)$$

$$\gamma(f, \tau) = \frac{|x_1(f, \tau)|^2}{\lambda_N(f)} \quad (7)$$

ここで $\lambda_N(f)$ は周波数 f におけるノイズの平均パワー（音声明らかに存在しない録音ファイルの冒頭区間などで求める。）、 L はフーリエ変換で用いる周波数の個数である。（例えば参考文献参照）

【参考文献】J.Sohn, N.S.Kim and W.Sung, "A Statistical Model-Based Voice Activity Detection, IEEE Signal Processing Letters", vol.6, no.1, pp.1-3, 1999.

【0019】

〔到来方向推定部〕

図4に到来方向推定部13と到来方向確率計算部14の機能構成例を示す。到来方向推定部13は、マイク間位相差計算部131と音源方向ベクトル計算部132を備える。マイク間位相差計算部131は、周波数領域に変換された観測信号 $x_j(f, \tau)$ の各フレーム τ 、各周波数 f におけるマイク間位相差 $q'_{jj}(f, \tau)$ を式(8)で計算する。

【0020】

【数6】

$$q'_{jj}(f, \tau) = \frac{1}{2\pi f} \arg[x_j(f, \tau)x_{j'}^*(f, \tau)] \quad (8)$$

ここで $x_j(f, \tau)$ はマイク j でのフレーム τ 、周波数 f における観測信号である。 $*$ は複素共役を表わす。全てのマイクペアにおけるマイク間位相差 $q'_{jj}(f, \tau)$ を並べたベクトルを $q'(f, \tau)$ と記載する。ベクトル $q'(f, \tau)$ は音源方向ベクトル計算部132に入力される。音源方向ベクトル計算部132は、ムーア・ペンローズ (Moore-Penrose) の擬似逆行列を用いて式(9)により音源方向ベクトル $q(f, \tau)$ を計算する。

【0021】

【数7】

$$q(f, \tau) = cD^+ q'(f, \tau) \quad (9)$$

ここで $+$ は Moore-Penrose の擬似逆行列を表わし、 $D = [d_1 - d_j, \dots, d_M - d_j]^T$ であり、 d_j はマイク j の座標を $[x, y, z]$ と並べたベクトルである。マイクから見た

10

20

30

40

50

音源の水平角を $\theta(f, \tau)$ 、仰角を $\phi(f, \tau)$ とすると音源方向ベクトル $q(f, \tau)$ は式 (10) で表わせる。

【0022】

【数8】

$$q(f, \tau) = [\cos\theta(f, \tau)\cos\phi(f, \tau), \sin\theta(f, \tau)\cos\phi(f, \tau), \sin\phi(f, \tau)]^T \quad (10)$$

音源方向ベクトル $q(f, \tau)$ は到来方向確率計算部 14 に入力される。以降、記載の簡単化のために水平角 $\theta(f, \tau)$ のみを用いて説明をする。

【0023】

〔到来方向確率計算部〕

到来方向確率計算部 14 は、クラスタリング部 140 と、各クラスタの分布計算部 160 と、確率計算部 170 を備える。この実施例では、クラスタリング部 140 が、音源方向ベクトル $q(f, \tau)$ の各フレームの各周波数 f における水平角 $\theta(f, \tau)$ をオンラインクラスタリングする。到来方向確率計算部 14 の動作フローを図 5 に示す。到来方向確率計算過程（ステップ S14、図 2 参照）は、クラスタリング部 140 が音源方向ベクトルとクラスタのセントロイドとの距離で音源方向ベクトルを分類する分類ステップ（ステップ S140、図 5 参照）と、各クラスタの分布計算部が分類毎の音源方向ベクトルの分布を計算する分布計算ステップ（ステップ S160）と、確率計算部 170 が分類毎の音源方向ベクトルの分布を、音源方向ベクトルの全体の分布で除して音声到来方向確率として計算する確率計算ステップ（ステップ S170）とを含む。図 6 に分類ステップ S140 の詳細な動作フローを示して説明する。

【0024】

<ステップ S141>

まず、分類するグループの中心値であるセントロイドを更新する大きさである更新ステップサイズ Δ と、グループ分けするための閾値 z を設定する。更新ステップサイズ Δ と閾値 z は、この発明を実施する環境に応じて適宜実験的に定められる値である。

<ステップ S142>

フレーム τ と周波数 f を初期化 ($\tau = 1$, $f = 1$) する。

<ステップ S143>

最初のフレーム $\tau = 1$ の最小周波数 $f = 1$ の音源方向ベクトル $q(f, \tau)$ の水平角 $\theta(f, \tau)$ を、第 1 のセントロイド c_1 とする。

<ステップ S144>

周波数 f を次の周波数にインクリメントする。

【0025】

<ステップ S145>

水平角 $\theta(f, \tau)$ に最も近い既存セントロイド c_k を見つけ、その番号を k とする。つまり式 (11) で、クラスタリングする周波数成分の水平角に最も近いクラスタ k を選択する。

【0026】

【数9】

$$k = \operatorname{argmin}_k \|\theta(f, \tau) - c_k\| \quad (11)$$

<ステップ S146>

ステップ S145 で求めた最も近いセントロイド c_k と $\theta(f, \tau)$ の距離と閾値 z を比較する。距離が閾値 z より小さければ（ステップ S146 の Yes）、 $\theta(f, \tau)$ も同方向（音源）からの周波数成分と判定してステップ S147 の処理を行う。距離が閾値 z よりも大きければ（ステップ S146 の No）、他の方向の音源からの周波数成分と判定してステップ S149 の処理を行う。

<ステップ S147>

セントロイド c_k を式 (12) で更新する。

10

20

30

40

50

【 0 0 2 7 】

【 数 1 0 】

$$c_k \leftarrow c_k + \beta(\theta(f, \tau) - c_k) \quad (12)$$

式(12)はセントロイド c_k を、水平角 (f, τ) に近づけることを意味する。これは、セントロイド c_k の初期値にクラスタリングの性能が左右されないようにするクラスタリングの一般的な手法である。

<ステップ S 1 4 8>

距離が閾値 z より小さいので同方向(音源)からの周波数成分と判断し、その時間周波数 (f, τ) にクラスタ K のクラスタ番号を付与する。ここでは、ある時間周波数 (f, τ) のクラスタ番号を $C(f, \tau)$ に保持する。

10

【 0 0 2 8 】

<ステップ S 1 4 9>

距離が閾値 z よりも大きいので、この音源方向ベクトル $q(f, \tau)$ は、他の方向の音源からの周波数成分と判定する(ステップ S 1 4 6 の No)。他の方向からの周波数成分として分類するために、 $\max(k) + 1$ 番目の新しいクラスタを生成し、そのセントロイドを $c_{\max(k) + 1} = (f, \tau)$ として与える。

<ステップ S 1 5 0>

その時間周波数 (f, τ) に新しいクラスタ番号を付与する。

<ステップ S 1 5 1>

20

周波数 f が、最後の周波数か否かを判定する。最後の周波数で無い場合(ステップ S 1 5 1 の No)、周波数をインクリメント(ステップ S 1 5 4)してステップ S 1 4 5 の動作に戻る。

【 0 0 2 9 】

<ステップ S 1 5 2>

周波数 f が、最後の周波数の場合(ステップ S 1 5 1 の Yes)、フレーム n が最後であるか否かを判定する。フレーム n が最後の場合、クラスタリング動作を終了する(ステップ S 1 5 2 の Yes)。フレーム n が最後で無い場合(ステップ S 1 5 2 の No)、フレーム n をインクリメントすると共に周波数を初期化(ステップ S 1 5 5)してステップ S 1 4 5 の動作に戻る。なお、メンバ数が少ないクラスタは除外しても良い(破線で示すステップ S 1 5 3)。

30

【 0 0 3 0 】

以上のように動作することで、音源方向ベクトル $q(f, \tau)$ の全てのフレーム、全ての周波数がクラスタリングされ、音源方向ベクトル $q(f, \tau)$ にクラスタ番号 k が付与される。

クラスタリング部 1 4 0 で分類された音源方向ベクトル $q(f, \tau)$ の水平角 (f, τ) の分布を、各クラスタの分布計算部 1 6 0 が計算する。各クラスタの分布計算部 1 6 0 は、式(13)を用いて各クラスタを平均値 c_k 、分散 σ_k^2 の正規分布でモデル化する(ステップ S 1 6 0、図 5 参照)。

【 0 0 3 1 】

40

【 数 1 1 】

$$N(\theta(f, \tau) | c_k, \sigma_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{-\frac{|\theta(f, \tau) - c_k|^2}{2\sigma_k^2}\right\} \quad (13)$$

ここで、平均値 c_k は、クラスタのセントロイド又は式(14)で計算した値を用いる。分散 σ_k^2 は式(15)で計算する。

【 0 0 3 2 】

【数 1 2】

$$c_k = \frac{1}{|C_k|} \sum_{C(f,\tau)=k} \theta(f,\tau) \quad (14)$$

$$\sigma_k^2 = \frac{1}{|C_k|} \sum_{C(f,\tau)=k} \{\theta(f,\tau) - c_k\}^2 \quad (15)$$

ここで、 $|C_k|$ は、クラスタ番号 $C(f, \tau) = k$ である成分の個数である。確率計算部 170 は、各フレーム τ で k の方向の音源が存在する確率を式 (16) と (17) を用いて計算する。

10

【0033】

【数 1 3】

$$p_k(\tau) = \frac{\sum_f p'_k(f,\tau)}{\sum_k \sum_f p'_k(f,\tau)} \quad (16)$$

$$p'_k(f,\tau) = \frac{N(\theta(f,\tau)|c_k,\sigma_k)}{\sum_k N(\theta(f,\tau)|c_k,\sigma_k)} \quad (17)$$

20

最後に乗算部 15 にて各フレーム τ における音声存在確率 $p_v(\tau)$ と到来方向確率 $p_k(\tau)$ との積を計算し、その確率値を音源 k の発話確率 $P_k(\tau)$ として出力する (ステップ S170)。

【0034】

【数 1 4】

$$P_k(\tau) = p_v(\tau) p_k(\tau) \quad (18)$$

これを全てのクラスタ k に対して計算することで、全ての音源の発話確率 $P_k(\tau)$ を得ることができる。

【実施例 2】

30

【0035】

次に到来方向確率 $p_k(\tau)$ を、クラスタリングされた音源方向ベクトル $q(f, \tau)$ の水平角 $\theta(f, \tau)$ の度数から求めるようにした実施例 2 を説明する。実施例 2 の到来方向確率計算部 14' は、クラスタリング部 140' と、確率計算部 170' を備える (図 4 参照)。他の構成は実施例 1 と同じである。動作フローを図 7 に示す。

クラスタリング部 140' は、ある時刻 τ までに存在するクラスタのセントロイド c_k について、式 (19) に示す計算をして音源方向ベクトル $q(f, \tau)$ をクラスタリングする (ステップ S140')。

【0036】

【数 1 5】

40

$$\phi_k(\theta(f,\tau)) = \begin{cases} 1 & \text{if } |\theta(f,\tau) - c_k(\tau)| \leq th \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

ここで th はある閾値であり、図 6 の閾値 z と例えば同じ値で構わない。あるクラスタ C_k に属する水平角 $\theta(f, \tau)$ については 1 を、そうでなければ 0 に分類する。この分類によって各クラスタの 度数 が求められる。その動作フローについては、前述した図 6 から明らかであるので省略する。

確率計算部 170' は、式 (20) で到来方向確率 $p_k(\tau)$ を計算する (ステップ S170')。

50

【 0 0 3 7 】

【 数 1 6 】

$$p_k(\tau) = \frac{\sum_f \phi_k(\theta(f, \tau))}{\sum_k \sum_f \phi_k(\theta(f, \tau))} \quad (20)$$

以上のように到来方向確率 $p_k(\tau)$ を求めることで、計算負荷を軽減することができる。計算負荷を軽減することで処理速度を向上させる効果が期待できる。

【 実施例 3 】

【 0 0 3 8 】

実施例 3 として雑音を抑圧するようにした到来方向確率計算部 60 の構成を図 4 に示して説明する。到来方向確率計算部 60 は、振幅計算部 61 を備える。他の構成は実施例 1, 2 と同じである。動作フローを図 8 に示す。

振幅計算部 61 は、音源方向ベクトル $q(f, \tau)$ の時間周波数 (f, τ) における正規化された振幅値 $a(f, \tau)$ を式 (21) で計算する (ステップ S61)。

【 0 0 3 9 】

【 数 1 7 】

$$a(f, \tau) = \frac{|x_1(f, \tau)|^b}{\sum_f |x_1(f, \tau)|^b} \quad (21)$$

$x_1(f, \tau)$ の 1 はマイクロホン番号である。定数 b は 1 ~ 4 の整数が望ましく、 $b = 1$ ならば振幅、 $b = 2$ ならばパワー、 $b = 4$ ならば尖度の正規化された振幅値 $a(f, \tau)$ となる。

確率計算部 62 は、振幅値 $a(f, \tau)$ を用いて到来方向確率 $p_k(\tau)$ を式 (22) で算出する (ステップ S170)。

【 0 0 4 0 】

【 数 1 8 】

$$p_k(\tau) = \frac{\sum_f a(f, \tau) \phi_k(\theta(f, \tau))}{\sum_k \sum_f a(f, \tau) \phi_k(\theta(f, \tau))} \quad (22)$$

ここで正規化された振幅値 $a(f, \tau)$ は、重み係数である。音声が存在する音源方向ベクトル $q(f, \tau)$ の時間周波数 (f, τ) の振幅は大きな値を持つ。それに対して音声が存在しない雑音だけの振幅は小さな値を持つ。したがって、音声区間における正規化された振幅値 $a(f, \tau)$ は大きくなり、非音声区間におけるそれは小さな値になる。

【 0 0 4 1 】

この正規化された振幅値 $a(f, \tau)$ を到来方向確率 $p_k(\tau)$ の算出の際に、式 (22) に示すように考慮することで、雑音を音声として誤検出してしまうことを抑制することができる。

なお、式 (22) は実施例 2 に振幅計算部 61 を設けた場合の式である。音源方向ベクトルの分布を正規分布として求めた実施例 1 に振幅計算部 61 を設けても、雑音を抑圧する効果が期待できる。

【 0 0 4 2 】

【 シミュレーション結果 】

実施例 2 の複数信号区間推定装置の性能を確認するシミュレーションを行った。シミュレーション条件を簡単に説明する。図 7 にシミュレーションに用いた部屋の平面図を示す。奥行きのある部屋の幅側の一边を、305cm の幅のパーティションで仕切り、幅が約 4m で奥行き約 9.3m の部屋を形成した。この部屋の残響時間は約 350ms である。パーティションの一方の隅にはパーソナルコンピュータ (PC) があり、そのファンノイ

10

20

30

40

50

ズが本システムに対する雑音となった。パーティション側に近い位置に長円形のテーブルを配置した。テーブルを挟んでパーティション側に話者 A と B の二人、反対側に話者 C と D の二人を座らせた。そして 4 人の話者のほぼ中央付近の位置に 3 個のマイクロホンを、4 cm の正三角形の頂点に位置するように配置した。

【0043】

話者 A ~ D の 4 名の会議を 5 分間、サンプリング周波数 16 kHz、フーリエ変換のフレーム長を 64 ms、フレームシフト長を 32 ms とし、上記した式 (16) の発話確率 P_k () が 0.4 以上となるクラス k の方向の話者が話したと判定した。評価指標としては、 $DER = (\text{誤受理} \cdot \text{誤棄却} \cdot \text{話者誤りの時間長}) / \text{全音声区間長} \times 100$ [% (Disarization Error Rate)] を利用した。

10

【0044】

ここで、誤受理 (FAT: false alarm speaker time) は、誰も話していないにもかかわらず誰かが話していると判定した時間長である。誤棄却 (MST: missed speaker time) は、誰かが話しているにもかかわらず話していないと判定した時間長である。話者誤り (SET: speaker error time) は、話者を誤って判定した時間長である。DER 値は、小さい方が話者区間推定の精度が高いことを意味する。全てを [%] で表わす。表 1 に結果を示す。

【0045】

【表 1】

表 1 実施例 2 の効果

20

Method	MST [%]	FAT [%]	SET [%]	DER [%]
従来法	35.3	8.1	9.7	53.2
実施例 2	25.1	6.5	3.6	35.2

従来法では、全てのエラーが多く、DER も大きかった。それに対して実施例 2 の方法では、特に誤棄却 (MST) が大きく改善され、その結果として DER の値が改善した。これは、フレーム毎に、音声区間と音声到来方向が、確率値として処理されること、及び各フレームで複数の方向を推定することにより音声区間が欠損することが少ないことによる。

30

【0046】

以上述べたように、この発明の複数信号区間推定装置によれば、音声信号の取りこぼしを少なくした複数信号の区間推定を行うことができる。この発明の技術思想に基づく複数信号区間推定装置とその方法は、上述の実施形態に限定されるものではなく、この発明の趣旨を逸脱しない範囲で適宜変更が可能である。上記した装置及び方法において説明した処理は、記載の順に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的あるいは個別に実行されるとしてもよい。

例えば、音声存在確率 p_v () と到来方向確率 p_k () の一方を、計算を軽くする目的で、決定論的に算出するようにしても良い。一方を決定論的に算出しても、音声があると判定されたフレームにおいては、複数音源があれば複数方向の方向確率が計算されるので、そのフレーム内の複数の音源を取りこぼすことが従来法に比べて少なくなる。

40

【0047】

また、到来方向確率 p_k () は、式 (17) を満たす水平角 (θ, ϕ) が各周波数で $t h_2$ 個以上存在していれば p_k () = 1、そうでなければ p_k () = 0 としても良い。また、クラスタリング部におけるセントロイド c_k は、予めそれぞれの音源の方向 θ_k が分かっているならばその角度を $c_k = \theta_k$ として与えても良い。また、各フレーム、周波数 (f, ϕ) における水平角 (θ, ϕ) ではなく、従来の GCC - PHAT 法のように各フレーム 毎に 1 つだけ求めた水平角 (θ) をオンラインクラスタリングし、そのセントロイドを c_k として用いても良い。また、水平角 (θ, ϕ) を用いて音源の到来

50

方向を分類する例で説明を行ったが、音源方向ベクトル $q(f, \quad)$ そのものを用いて分類するようにしても良い。

【0048】

また、上記装置における処理手段をコンピュータによって実現する場合、各装置が有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、各装置における処理手段がコンピュータ上で実現される。

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよい。具体的には、例えば、磁気記録装置として、ハードディスク装置、フレキシブルディスク、磁気テープ等を、光ディスクとして、DVD (Digital Versatile Disc)、DVD-RAM (Random Access Memory)、CD-ROM (Compact Disc Read Only Memory)、CD-R (Recordable) / RW (ReWritable) 等を、光磁気記録媒体として、MO (Magneto Optical disc) 等を、半導体メモリとしてフラッシュメモリ等を用いることができる。

10

【0049】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記録装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させる構成としてもよい。

20

また、各手段は、コンピュータ上で所定のプログラムを実行させることにより構成することにしてもよいし、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

【図面の簡単な説明】

【0050】

【図1】この発明の複数信号区間推定装置100の機能構成例を示す図。

【図2】複数信号区間推定装置100の動作フローを示す図。

【図3】音声存在確率推定部12の機能ブロックを示す図。

【図4】到来方向推定部13と到来方向確率計算部14の機能ブロックを示す図。

【図5】到来方向確率計算部14の動作フローを示す図。

30

【図6】クラスタリング部140の動作フローを示す図。

【図7】実施例2の到来方向確率計算部14'の動作フローを示す図。

【図8】実施例3の到来方向確率計算部60の動作フローを示す図。

【図9】シミュレーションを行った部屋の平面を示す図。

【図10】非特許文献1に開示された従来の音声区間推定装置200の機能構成を示す図。

。

【図1】

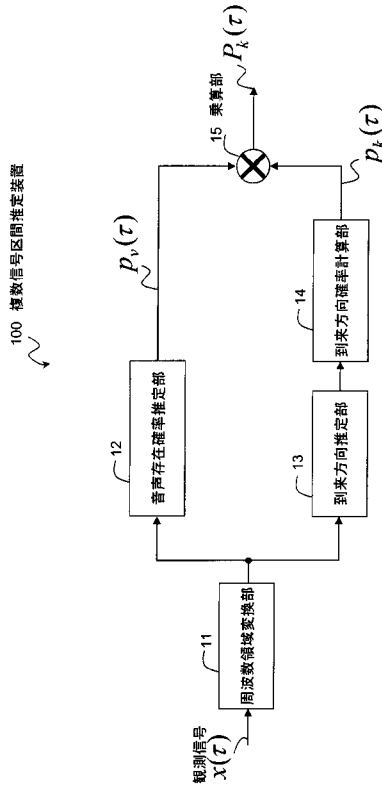


図1

【図2】

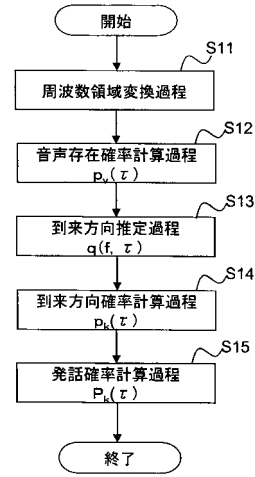


図2

【図3】

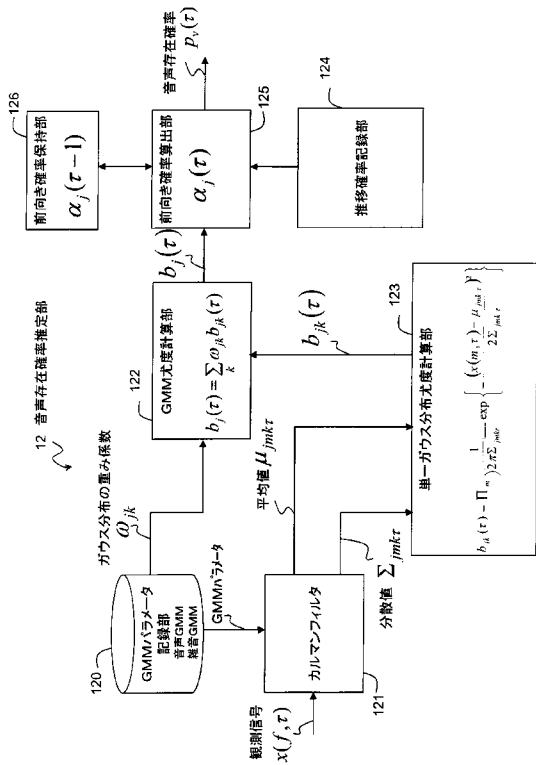


図3

【図4】

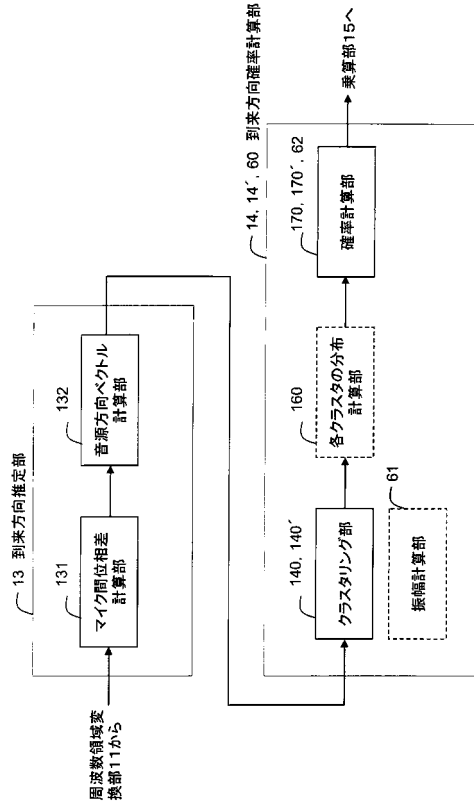


図4

【 図 5 】

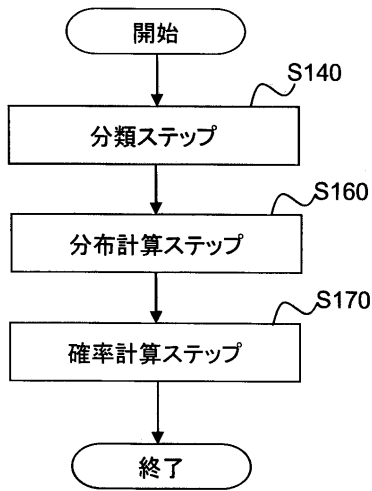


図5

【 図 6 】

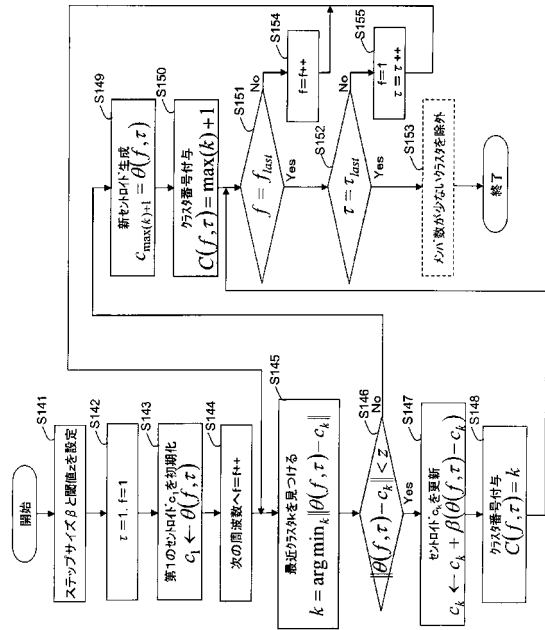


図6

【 図 7 】

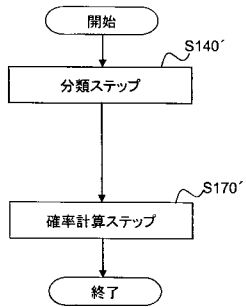


図7

【 図 9 】

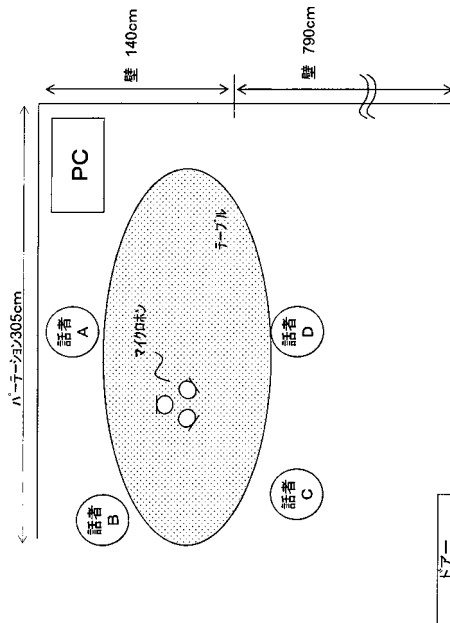


図9

【 図 8 】

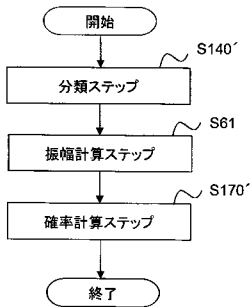


図8

【図10】

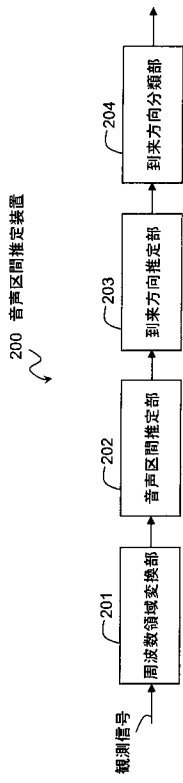


図10

フロントページの続き

- (72)発明者 石塚 健太郎
東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内
- (72)発明者 藤本 雅清
東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内
- (72)発明者 牧野 昭二
東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

審査官 安田 勇太

- (56)参考文献 特開2007-085734(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 11/02
G10L 21/02
G01S 3/808