

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第6370751号
(P6370751)

(45) 発行日 平成30年8月8日 (2018.8.8)

(24) 登録日 平成30年7月20日 (2018.7.20)

(51) Int.Cl.	F I
G 1 O L 15/14 (2006.01)	G 1 O L 15/14
G 1 O L 21/0264 (2013.01)	G 1 O L 21/0264 Z
G O 6 F 17/16 (2006.01)	G O 6 F 17/16 Z
G O 6 F 17/10 (2006.01)	G O 6 F 17/10 D

請求項の数 8 (全 19 頁)

(21) 出願番号	特願2015-162474 (P2015-162474)	(73) 特許権者	000004226
(22) 出願日	平成27年8月20日 (2015.8.20)		日本電信電話株式会社
(65) 公開番号	特開2017-40781 (P2017-40781A)		東京都千代田区大手町一丁目5番1号
(43) 公開日	平成29年2月23日 (2017.2.23)	(74) 代理人	100121706
審査請求日	平成29年8月29日 (2017.8.29)		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
		(72) 発明者	小泉 悠馬
			東京都千代田区大手町一丁目5番1号 日
			本電信電話株式会社内
		(72) 発明者	丹羽 健太
			東京都千代田区大手町一丁目5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】 ガウス混合モデルパラメータ計算装置、情報推定装置、音強調装置、これらの方法及びプログラム

(57) 【特許請求の範囲】

【請求項 1】

・ T は行列又はベクトル・の転置を表すとして、入力変数 x 及び出力変数 y の学習データに基づいて、ガウス混合モデルで表現された結合分布 $p(A^T x, y; \cdot)$ のガウス混合モデルパラメータ 及び圧縮行列 A を求めるガウス混合モデルパラメータ計算装置であって、

入力されたガウス混合モデルパラメータ に基づいて、EMアルゴリズムにおける負担率を計算する負担率計算部と、

入力されたガウス混合モデルパラメータ 及び上記負担率に基づいてガウス混合モデルパラメータ を更新するガウス混合モデルパラメータ更新部と、

入力されたガウス混合モデルパラメータ 、上記負担率及び入力された圧縮行列 A に基づいて、上記入力された圧縮行列 A を勾配法を用いた最適化により更新する圧縮行列更新部と、

上記更新されたガウス混合モデルパラメータ 及び上記更新された圧縮行列 A を上記入力されたガウス混合モデルパラメータ 及び上記入力された圧縮行列 A として、上記負担率計算部、上記ガウス混合モデルパラメータ更新部及び上記圧縮行列更新部の処理を繰り返す制御を行う制御部と、

を含むガウス混合モデルパラメータ計算装置。

【請求項 2】

請求項 1 のガウス混合モデルパラメータ計算装置と、

上記繰り返し制御により最終的に生成された圧縮行列Aを用いて、情報推定の対象となる入力変数 x を圧縮した $A^T x$ を計算する次元圧縮部と、

上記繰り返し制御により最終的に生成されたガウス混合モデルパラメータ を用いて、上記計算された $A^T x$ が得られた下での y の期待値である \hat{y} を計算する推定部と、を含む情報推定装置。

【請求項3】

請求項1のガウス混合モデルパラメータ計算装置であって、

上記入力変数 x は目的音が含まれる音信号の音響特徴量であり、上記出力変数 y は事前信号雑音比である、

ガウス混合モデルパラメータ計算装置。

10

【請求項4】

強調の対象となる目的音が含まれる音信号を周波数領域信号に変換する周波数領域変換部と、

上記周波数領域信号に基づいて、強調の対象となる目的音が含まれる音信号の音響特徴量 x を抽出する音響特徴量抽出部と、

請求項3のガウス混合モデルパラメータ計算装置により最終的に生成された圧縮行列Aを用いて、上記抽出された音響特徴量 x を圧縮した $A^T x$ を計算する次元圧縮部と、

請求項3のガウス混合モデルパラメータ計算装置により最終的に生成されたガウス混合モデルパラメータ を用いて、上記計算された $A^T x$ が得られた下での事前信号雑音比 y の期待値である \hat{y} を計算する推定部と、

20

上記計算された \hat{y} を用いてウィナーフィルタを求めるウィナーフィルタ計算部と、

上記求めたウィナーフィルタを上記周波数領域信号に適用してフィルタリング後周波数領域信号を得るフィルタリング部と、

上記フィルタリング後周波数領域信号を時間領域信号に変換する時間領域変換部と、

を含む音強調装置。

【請求項5】

\cdot^T は行列又はベクトル \cdot の転置を表すとして、入力変数 x 及び出力変数 y の学習データに基づいて、ガウス混合モデルで表現された結合分布 $p(A^T x, y; \cdot)$ のガウス混合モデルパラメータ 及び圧縮行列Aを求めるガウス混合モデルパラメータ計算方法であって、

30

負担率計算部が、入力されたガウス混合モデルパラメータ に基づいて、EMアルゴリズムにおける負担率を計算する負担率計算ステップと、

ガウス混合モデルパラメータ更新部が、入力されたガウス混合モデルパラメータ 及び上記負担率に基づいてガウス混合モデルパラメータ を更新するガウス混合モデルパラメータ更新ステップと、

圧縮行列更新部が、入力されたガウス混合モデルパラメータ 、上記負担率及び入力された圧縮行列Aに基づいて、上記入力された圧縮行列Aを勾配法を用いた最適化により更新する圧縮行列更新ステップと、

制御部が、上記更新されたガウス混合モデルパラメータ 及び上記更新された圧縮行列Aを上記入力されたガウス混合モデルパラメータ 及び上記入力された圧縮行列Aとして、上記負担率計算部、上記ガウス混合モデルパラメータ更新部及び上記圧縮行列更新部の処理を繰り返す制御を行う制御ステップと、

40

を含むガウス混合モデルパラメータ計算方法。

【請求項6】

請求項5のガウス混合モデルパラメータ計算方法の各ステップと、

次元圧縮部が、上記繰り返し制御により最終的に生成された圧縮行列Aを用いて、情報推定の対象となる入力変数 x を圧縮した $A^T x$ を計算する次元圧縮ステップと、

推定部が、上記繰り返し制御により最終的に生成されたガウス混合モデルパラメータ を用いて、上記計算された $A^T x$ が得られた下での y の期待値である \hat{y} を計算する推定ステップと、

50

を含む情報推定方法。

【請求項 7】

周波数領域変換部が、強調の対象となる目的音が含まれる音信号を周波数領域信号に変換する周波数領域変換ステップと、

音響特徴量抽出部が、上記周波数領域信号に基づいて、強調の対象となる目的音が含まれる音信号の音響特徴量 x を抽出する音響特徴量抽出ステップと、

次元圧縮部が、請求項 5 のガウス混合モデルパラメータ計算方法により最終的に生成された圧縮行列 A を用いて、上記抽出された音響特徴量 x を圧縮した $A^T x$ を計算する次元圧縮ステップと、

推定部が、請求項 5 のガウス混合モデルパラメータ計算方法により最終的に生成されたガウス混合モデルパラメータを用いて、上記計算された $A^T x$ が得られた下での事前信号雑音比 y の期待値である \hat{y} を計算する推定ステップと、

ウィナーフィルタ計算部が、上記計算された \hat{y} を用いてウィナーフィルタを求めるウィナーフィルタ計算ステップと、

フィルタリング部が、上記求めたウィナーフィルタを上記周波数領域信号に適用してフィルタリング後周波数領域信号を得るフィルタリングステップと、

時間領域変換部が、上記フィルタリング後周波数領域信号を時間領域信号に変換する時間領域変換ステップと、

を含む音強調方法。

【請求項 8】

請求項 1 又は 3 のガウス混合モデルパラメータ計算装置、請求項 2 の情報推定装置及び請求項 4 の音強調装置の何れかの装置の各部としてコンピュータを機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、ガウス混合モデルパラメータを計算するための技術、情報を推定するための技術又は特定の音を強調するための技術に関する。

【背景技術】

【0002】

ある入力 $x \in \mathbb{R}^D$ 、 $D = \{1, 2, \dots, T\}$ から、それに対応する出力 $y \in \mathbb{R}^M$ を推定する問題を考える。 D は、所定の正の整数とする。音源強調を例に挙げて説明すると、この問題は、スポーツフィールドである特定の競技音（e.g. キック音やホイッスル音）を収録する場合において、 M 本のマイクロホンで観測した信号から D 次元の音響特徴量 x を抽出し、目的音を抽出するフィルタ設計するパラメータ y を推定する問題である。

【0003】

音源強調の代表的な手法であるウィナーフィルタリングを行う場合、観測から推定すべきパラメータ $y = (y_1, y_2, \dots, y_T)^T$ は事前信号雑音比（以下、事前SNRとも表記する。）である。ここで、 \cdot^T は行列又はベクトル \cdot の転置を表す。以下ではウィナーフィルタリングと事前SNRの使い方について簡単に説明する。目的音（e.g. 収音したい特定の競技音）の複素スペクトルを S 、雑音（e.g. 歓声）の複素スペクトルを N 、時間と周波数のインデックスをそれぞれ $\{1, 2, \dots, T\}$ 、 $\{1, 2, \dots, F\}$ としたとき、観測信号は

$$F = S + N$$

と記述することができる。目的音を強調する時刻 t のウィナーフィルタ G_t は、

【0004】

【数 1】

$$G_{\omega,\tau} = \frac{y_{\omega,\tau}}{1+y_{\omega,\tau}}$$

$$y_{\omega,\tau} = \frac{|S_{\omega,\tau}|^2}{|N_{\omega,\tau}|^2}$$

【0005】

と求めることができる。また、 $y_{\omega,\tau}$ は近似計算された事前信号雑音比である。この $G_{\omega,\tau}$ と観測信号 $F_{\omega,\tau}$ とを掛け合わせるにより、目的音を強調した信号 $H_{\omega,\tau}$ を得ることができる。

10

【0006】

$$H_{\omega,\tau} = G_{\omega,\tau} F_{\omega,\tau}$$

つまり、適切なウィナーフィルタを設計するためには、観測信号 $F_{\omega,\tau}$ から事前SNRを精度よく推定する必要がある。

【0007】

推定の問題を解く一般的な手法として、最小二乗平均誤差推定（以下、MMSEとも表記する。）法がある。入力 x （i.e. 音響特徴量）を得たもとでの出力 y （i.e. 事前SNR）のMMSE推定量 \hat{y} は、以下のように計算できる。

【0008】

【数 2】

$$\hat{y}_\tau = \int y_\tau \times p(y_\tau; x_\tau, \Theta) dy_\tau$$

20

【0009】

ここで、 Θ は、後述する x と y の結合分布のパラメータである。つまり、MMSE推定量 \hat{y} は、入力 x を得たもとでの出力 y の期待値である。出力 y の推定精度を向上させるためには、学習データから条件付き分布を最適化する必要がある。ベイズの定理より、条件付き分布の最適化は、 x と y の結合分布 $p(x, y; \Theta)$ のパラメータ Θ を最適化すればよい。

【0010】

この結合分布には、例えばガウス混合モデル（Gaussian Mixture Model: GMM）を用いることができる。結合変数 $z = (x^T, y^T)^T$ の分布 $p(x, y; \Theta)$ をGMMで表現する場合、以下のような形になる。

30

【0011】

【数 3】

$$p(x_\tau, y_\tau; \Theta) = \sum_{k=1}^K w_k N(z_\tau; \mu_k^z, \Sigma_k^z)$$

【0012】

ここで、 T は転置、 $N(\mu, \Sigma)$ は平均 μ 、共分散行列 Σ の正規分布、 K はGMMの混合数、 μ_k^z は k 番目の正規分布の平均ベクトル、 Σ_k^z は k 番目の正規分布の共分散行列である。また、 $\Theta = \{w_k, \mu_k^z, \Sigma_k^z\}_{k=1}^K$ である。さらに、 μ_k^z, Σ_k^z は、以下のように書き下すことができる。

40

【0013】

【数 4】

$$\mu_k^z = \begin{bmatrix} \mu_k^y \\ \mu_k^x \end{bmatrix}, \quad \Sigma_k^z = \begin{bmatrix} \Sigma_k^{yy} & \Sigma_k^{yx} \\ \Sigma_k^{xy} & \Sigma_k^{xx} \end{bmatrix}$$

【0014】

μ_k^x, μ_k^y はそれぞれ k 番目の正規分布の x と y の平均ベクトル、 $\Sigma_k^{xx}, \Sigma_k^{yy}$ はそれぞれ k 番目の正規分布の x と y の共分散行列、 $\Sigma_k^{xy}, \Sigma_k^{yx}$ はそれぞれ k 番目の正規分布の x と y の各次元の相関を表す相関行列である。

50

【 0 0 1 5 】

上述のGMMに基づく、入力 x から y のMMSE推定量 \hat{y} は以下のように求めることができる。

【 0 0 1 6 】

【 数 5 】

$$\begin{aligned}\hat{y}_r &= \int y_r \times p(y_r; x_r, \Theta) dy \\ &= \sum_{k=1}^K \frac{w_k N(x_r; \mu_k^x, \Sigma_k^x)}{\sum_{j=1}^K w_j N(x_r; \mu_j^x, \Sigma_j^x)} \int y_r \times p(y_r; x_r, \mu_k^y, \Sigma_k^y) dy \\ &= \sum_{k=1}^K \frac{w_k N(x_r; \mu_k^x, \Sigma_k^x)}{\sum_{j=1}^K w_j N(x_r; \mu_j^x, \Sigma_j^x)} \{ \mu_k^y + \Sigma_k^y (\Sigma_k^x)^{-1} (x_r - \mu_k^x) \}\end{aligned}$$

10

【 0 0 1 7 】

結合分布にGMMを用いたMMSE推定法は、その計算の容易さから、声質変換（例えば、非特許文献1参照。）など、音響処理の分野で幅広く用いられている。

【 先行技術文献 】

【 非特許文献 】

【 0 0 1 8 】

【 非特許文献1 】 T. Toda, A.W. Black, K. Tokuda. "Voice conversion based on maximum likelihood estimation of spectral parameter trajectory," IEEE Transactions on Audio, Speech and Language Processing, Vol. 15, No. 8, pp. 2222-2235, Nov. 2007.

20

【 非特許文献2 】 小西貞則, "多変量解析入門, 9章主成分分析" pp.225-244, 岩波書店, 2010.

【 発明の概要 】

【 発明が解決しようとする課題 】

【 0 0 1 9 】

ところが、入力 x の次元 D が、学習データに対して大きすぎる場合（例えば $T=3000$ 個程度の学習データで、 $D=1024$ 次元の入力を扱う場合）、結合分布 $p(x, y; \Theta)$ を正確に学習することができない。このデータの次元に対して学習データ数が足りず、学習が困難になる現象は「次元の呪い」と呼ばれている。こういった状況では結合分布を学習する前に、入力 x の次元を削減しなくてはならない。

30

【 0 0 2 0 】

入力 x の次元を小さくする方法として、2つのアプローチがある。1つは、開発者の経験や直感に基づき、入力 x に用いる音響特徴量を人手で設計/精査することである。例えばシュート音やバッティング音を収録したいとき、突発性の音源であるという特徴が事前に既知であるので、ある帯域における時間的なパワー差分を使えば、事前SNRを推定できそうだろう。つまり、「音響特徴量 x としてある周波数帯域のパワー時間差分だけを音響特徴量として用いる」と経験的に決定する。しかし、シュート音のように、「突発性」という、その音を特徴づける性質が明確であればよいが、人間の声や動物の鳴き声など、音を特徴づける性質が未知であることもある。こういった場合、強調/抑圧したい音源ごとに特徴量を人手で設計する必要があるが労力を要する。また目的音の時間周波数構造が複雑な場合は、ルールや特徴量自体を設計することが困難である。

40

【 0 0 2 1 】

そこで2つ目のアプローチとして、大量の音響特徴量候補 x を圧縮行列 $A \in \mathbb{R}^D \times Q$ ($Q < D$)を用いて次元圧縮する方法がある。つまり、人手で特徴量を精査するのをあきらめ、機械学習で有効な特徴量を探索する。そのため、 x としては考え付く限りの大量の次元 D の音響特徴量を用い、 $A^T x$ となる変換を行う。そして、結合分布 $p(A^T x, y; \Theta)$ を学習する。この圧縮行列 A の設計法の代表的なものとして、主成分分析がある（例えば、非特許

50

文献 2 参照)。主成分分析では、入力 x の分散が最大となるような圧縮行列 A を設計する。

【0022】

入力に対して圧縮処理を行うということは、入力 x の中から必要な情報を推定し、不要と判断された情報を捨てることに相当する。ところが、主成分分析の圧縮行列の設計には出力変数 y の情報を用いない。つまり、主成分分析では、出力 y を推定するために重要な情報を捨ててしまう可能性がある。そのため、主成分分析による次元圧縮は、GMMを用いたMMSE推定の枠組みから見たとき最適ではない。最適な圧縮行列 A を設計するためには、結合分布 $p(x, y; \cdot)$ の最適化と同時に、圧縮行列 A を最適化する必要がある。

【0023】

この発明の目的は、ガウス混合モデルパラメータを求めるときに圧縮行列 A の最適化を行うガウス混合モデルパラメータ計算装置、情報推定装置、音強調装置、これらの方法及び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】

【0024】

この発明の一態様によるガウス混合モデルパラメータ計算装置は、 \cdot^T は行列又はベクトル \cdot の転置を表すとして、入力変数 x 及び出力変数 y の学習データに基づいて、ガウス混合モデルで表現された結合分布 $p(A^T x, y; \cdot)$ のガウス混合モデルパラメータ及び圧縮行列 A を求めるガウス混合モデルパラメータ計算装置であって、入力されたガウス混合モデルパラメータに基づいて、EMアルゴリズムにおける負担率を計算する負担率計算部と、入力されたガウス混合モデルパラメータ及び上記負担率に基づいてガウス混合モデルパラメータを更新するガウス混合モデルパラメータ更新部と、入力されたガウス混合モデルパラメータ、上記負担率及び入力された圧縮行列 A に基づいて、上記入力された圧縮行列 A を勾配法を用いた最適化により更新する圧縮行列更新部と、上記更新されたガウス混合モデルパラメータ及び上記更新された圧縮行列 A を上記入力されたガウス混合モデルパラメータ及び上記入力された圧縮行列 A として、上記負担率計算部、上記ガウス混合モデルパラメータ更新部及び上記圧縮行列更新部の処理を繰り返す制御を行う制御部と、を備えている。

【0025】

この発明の一態様による情報推定装置は、上記ガウス混合モデルパラメータ計算装置と、上記繰り返し制御により最終的に生成された圧縮行列 A を用いて、情報推定の対象となる入力変数 x を圧縮した $A^T x$ を計算する次元圧縮部と、上記繰り返し制御により最終的に生成されたガウス混合モデルパラメータを用いて、上記計算された $A^T x$ が得られた下での y の期待値である \hat{y} を計算する推定部と、を備えている。

【0026】

この発明の一態様による音強調装置は、強調の対象となる目的音が含まれる音信号を周波数領域信号に変換する周波数領域変換部と、上記周波数領域信号に基づいて、強調の対象となる目的音が含まれる音信号の音響特徴量 x を抽出する音響特徴量抽出部と、請求項3のガウス混合モデルパラメータ計算装置により最終的に生成された圧縮行列 A を用いて、上記抽出された音響特徴量 x を圧縮した $A^T x$ を計算する次元圧縮部と、請求項3のガウス混合モデルパラメータ計算装置により最終的に生成されたガウス混合モデルパラメータを用いて、上記計算された $A^T x$ が得られた下での事前信号雑音比 y の期待値である \hat{y} を計算する推定部と、上記計算された \hat{y} を用いてウィナーフィルタを求めるウィナーフィルタ計算部と、上記求めたウィナーフィルタを上記周波数領域信号に適用してフィルタリング後周波数領域信号を得るフィルタリング部と、上記フィルタリング後周波数領域信号を時間領域信号に変換する時間領域変換部と、を備えている。

【発明の効果】

【0027】

ガウス混合モデルパラメータを求めるときに圧縮行列 A の最適化が行われる。

【図面の簡単な説明】

10

20

30

40

50

【 0 0 2 8 】

【図 1】 ガウス混合モデルパラメータ計算装置の例を説明するためのブロック図。

【図 2】 情報推定装置の例を説明するためのブロック図。

【図 3】 音強調装置の例を説明するためのブロック図。

【図 4】 画像の判別を行うためのガウス混合モデルパラメータを計算するガウス混合モデルパラメータ計算装置の例を説明するためのブロック図。

【図 5】 ガウス混合モデルパラメータ計算方法の例を説明するための流れ図。

【図 6】 情報推定方法の例を説明するための流れ図。

【図 7】 音強調方法の例を説明するための流れ図。

【発明を実施するための形態】

10

【 0 0 2 9 】

以下、図面を参照して、この発明の一実施形態であるガウス混合モデルパラメータ計算装置、情報推定装置、音強調装置及びこれらの方法について説明する。

【 0 0 3 0 】

まず、ガウス混合モデルパラメータ計算装置及び方法について説明し、その後、ガウス混合モデルパラメータ計算装置及び方法を用いた情報推定装置及び方法、音強調装置及び方法について説明する。

【 0 0 3 1 】

[ガウス混合モデルパラメータ計算装置及び方法]

ガウス混合モデルパラメータ計算装置は、図 1 に示すように、初期化部 1 1、負担率計算部 1 2、ガウス混合モデルパラメータ更新部 1 3、圧縮行列更新部 1 4 及び制御部 1 5 を例えば備えている。

20

【 0 0 3 2 】

ガウス混合モデルパラメータ計算装置が、図 5 に例示する各ステップの処理を行うことによりガウス混合モデルパラメータ計算方法が実現される。

【 0 0 3 3 】

ガウス混合モデルパラメータ計算装置は、通常の GMM の学習で用いられる期待値最大化 (expectation-maximization: EM) アルゴリズムに、勾配法を用いた圧縮行列の最適化を内包した一般化 EM (Generalized-EM: GEM) アルゴリズムを提供するものである。

【 0 0 3 4 】

30

ガウス混合モデルパラメータ計算装置 1 には、 T 個の入力変数 x の学習データ x_1, \dots, x_T と出力変数 y の学習データ y_1, \dots, y_T が入力される。入力の次元数は D 、出力の次元数は H とする。 D, H は正の整数である。また、ガウス混合モデルパラメータ計算装置 1 には、圧縮後の次元数 Q と GMM の混合数 K 、また後述する勾配法のステップサイズ α が入力される。 Q, K, α の値は任意であり、解きたい問題によって設定を変更すべきだが、例えば $Q=10, K=8, \alpha=10^{-2}$ などに設定できる。

【 0 0 3 5 】

< 初期化部 1 1 >

初期化部 1 1 は、GMM パラメータ $\theta = \{w_k, \mu_k, \sigma_k\}_{k=1}^K$ と圧縮行列 $A \in \mathbb{R}^{D \times Q}$ を初期化する (ステップ S 1 1)。以下、GMM パラメータのことを、ガウス混合モデルパラメータとも表記する。

40

【 0 0 3 6 】

初期化部 1 1 は、GMM パラメータ θ はランダム初期化してもよいし、K-means アルゴリズムなどで初期化することもできる。また、初期化部 1 1 は、圧縮行列 A はランダム初期化してもよいし、主成分分析などで初期化することもできる。

初期化された GMM パラメータ θ は、負担率計算部 1 2、ガウス混合モデルパラメータ更新部 1 3 及び必要に応じて制御部 1 5 に出力される。

【 0 0 3 7 】

初期化された圧縮行列 A は、負担率計算部 1 2、ガウス混合モデルパラメータ更新部 1 3、圧縮行列更新部 1 4 及び必要に応じて制御部 1 5 に出力される。

50

【 0 0 3 8 】

なお、GMMパラメータ と圧縮行列Aは入手で初期化してもよい。この場合は、初期化部 1 1 の処理に代えて、入手で初期化されたGMMパラメータ と圧縮行列Aがガウス混合モデルパラメータ計算装置 1 に入力される。

【 0 0 3 9 】

< 負担率計算部 1 2 >

負担率計算部 1 2 は、初期化されたGMMパラメータ 、及び、初期化された圧縮行列Aを入力として、通常のEMアルゴリズムと同様に負担率 $\gamma_{k, \tau}$ を計算する（ステップ S 1 2 ） 。計算された負担率 $\gamma_{k, \tau}$ は、ガウス混合モデルパラメータ更新部 1 3 及び圧縮行列更新部 1 4 に出力される。

10

【 0 0 4 0 】

【数 6】

$$\gamma_{k, \tau} = \frac{w_k N(\mathbf{v}_\tau; \boldsymbol{\mu}_k^\nu, \boldsymbol{\Sigma}_k^\nu, A)}{\sum_{j=1}^K w_j N(\mathbf{v}_j; \boldsymbol{\mu}_j^\nu, \boldsymbol{\Sigma}_j^\nu, A)}$$

ここで、

【 0 0 4 1 】

【数 7】

$$\mathbf{v}_\tau = (\mathbf{y}_\tau^T, (A^T \mathbf{x}_\tau)^T)^T$$

20

であり、

【 0 0 4 2 】

【数 8】

$$N(\mathbf{v}_\tau; \boldsymbol{\mu}_k^\nu, \boldsymbol{\Sigma}_k^\nu, A) = \frac{1}{(2\pi)^{(\Omega+Q)/2} |\boldsymbol{\Sigma}_k^\nu|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{v}_\tau - \boldsymbol{\mu}_k^\nu)^T (\boldsymbol{\Sigma}_k^\nu)^{-1} (\mathbf{v}_\tau - \boldsymbol{\mu}_k^\nu)\right\}$$

である。また、

【 0 0 4 3 】

【数 9】

$$\boldsymbol{\mu}_\tau^\nu = \begin{bmatrix} \boldsymbol{\mu}_k^y \\ \boldsymbol{\mu}_k^\xi \end{bmatrix}, \quad \boldsymbol{\Sigma}_k^\nu = \begin{bmatrix} \boldsymbol{\Sigma}_k^{yy} & \boldsymbol{\Sigma}_k^{y\xi} \\ \boldsymbol{\Sigma}_k^{\xi y} & \boldsymbol{\Sigma}_k^{\xi\xi} \end{bmatrix}$$

30

【 0 0 4 4 】

である。ここで、 \mathbf{v}_τ は $A^T \mathbf{x}_\tau$ を一文字で表すために表記上導入した変数であり、 $\boldsymbol{\mu}_k$ はそれぞれk番目の正規分布の $A^T \mathbf{x}$ と y の平均ベクトル、 $\boldsymbol{\Sigma}_k^{yy}$ はそれぞれk番目の正規分布の $A^T \mathbf{x}$ と y の共分散行列、 $\boldsymbol{\Sigma}_k^{y\xi}$ 、 $\boldsymbol{\Sigma}_k^{\xi y}$ のはそれぞれk 番目の正規分布の $A^T \mathbf{x}$ と y の各次元の相関を表す相関行列、 w_k はk番目の正規分布の混合比（重みパラメータ）である。

【 0 0 4 5 】

40

このようにして、負担率計算部 1 2 は、入力されたガウス混合モデルパラメータ に基づいて、EMアルゴリズムにおける負担率を計算する。

【 0 0 4 6 】

< ガウス混合モデルパラメータ更新部 1 3 >

ガウス混合モデルパラメータ更新部 1 3 は、初期化されたGMMパラメータ 、初期化された圧縮行列A、及び、負担率 $\gamma_{k, \tau}$ を入力として、以下の式でGMMパラメータ $\theta = \{w_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K$ を更新する（ステップ S 1 3 ）。更新されたGMMパラメータ は、圧縮行列更新部 1 4 及び制御部 1 5 に出力される。

【 0 0 4 7 】

【数 1 0】

$$\mu_k^v = \frac{1}{N_k} \sum_{\tau=1}^T \gamma_{k,\tau} v_\tau$$

$$\Sigma_k^v = \frac{1}{N_k} \sum_{\tau=1}^T \gamma_{k,\tau} (v_\tau - \mu_k^v)(v_\tau - \mu_k^v)^T$$

$$w_\tau = \frac{N_k}{T}$$

$$N_k = \sum_{\tau=1}^T \gamma_{k,\tau}$$

10

【0 0 4 8】

このようにして、ガウス混合モデルパラメータ更新部 1 3 は、入力されたガウス混合モデルパラメータ 及び上記負担率に基づいてガウス混合モデルパラメータ を更新する。

【0 0 4 9】

< 圧縮行列更新部 1 4 >

圧縮行列更新部 1 4 は、初期化された圧縮行列A、負担率 w_k 、及び、更新されたガウス混合モデルパラメータ を入力として、圧縮行列A $R^D \times Q$ を勾配法で更新する（ステップS 1 4）。勾配法の具体的な手段には、最急降下法や確率的最急降下法などを用いることができる。例えば最急降下法の場合は、圧縮行列Aの各行ベクトル a_1, \dots, a_D を例えば 20

以下の式により更新する（ステップS 1 4 1）。

【0 0 5 0】

【数 1 1】

$$a_d^{new} = a_d^{old} - \varepsilon \nabla a_d$$

【0 0 5 1】

ここで、 ε はステップサイズである。勾配ベクトル a_d を算出する。まず、圧縮行列演算を以下のように書き下す。

【0 0 5 2】

【数 1 2】

$$A^T x_\tau = \sum_{d=1}^D a_d x_{d,\tau}$$

30

【0 0 5 3】

また、結合ガウス分布の性質を用いて、各ガウス分布の指数部分の二次形式をy の項、x の項、クロス項の3つに展開することで、勾配ベクトル a_d は以下のように計算できる。

【0 0 5 4】

【数 1 3】

$$\nabla a_d = \frac{1}{T} \sum_{\tau=1}^T x_{d,\tau} \sum_{k=1}^K \gamma_{k,\tau} Y_{k,\tau}$$

40

$$Y_{k,\tau} = d_{y,k,\tau}^T \Lambda_k^y + d_{x,k,\tau}^T \Lambda_k^{xx}$$

$$d_{y,k,\tau}^T = y_\tau - \mu_k^y$$

$$d_{x,k,\tau}^T = A^T x_\tau - \mu_k^x$$

【0 0 5 5】

ただし、 Λ_k^y と Λ_k^{xx} は以下の式で求まる分割行列である。なお、分割行列のサイズは Λ_k^y , Λ_k^{xx} に等しい。

50

【 0 0 5 6 】

【 数 1 4 】

$$(\Sigma_k^v)^{-1} = \begin{bmatrix} \Lambda_k^{yy} & \Lambda_k^{y\xi} \\ \Lambda_k^{\xi y} & \Lambda_k^{\xi\xi} \end{bmatrix}$$

【 0 0 5 7 】

また、圧縮後の入力変数 $A^T x$ の分散を大きくしたいときは、勾配ベクトル a_d を以下のように計算してもよい。

【 0 0 5 8 】

【 数 1 5 】

$$\nabla a_d = \frac{1}{T} \sum_{\tau=1}^T x_{d,\tau} \sum_{k=1}^K \gamma_{k,\tau} Y_{k,\tau} - \eta_{k,\tau} d_{x,k,\tau}^T (\Sigma_k^{\xi\xi})^{-1}$$

$$\eta_{k,\tau} = \frac{w_k N(A^T x_\tau; \mu_k^\xi, \Sigma_k^{\xi\xi})}{\sum_{j=1}^K w_j N(A^T x_\tau; \mu_j^\xi, \Sigma_j^{\xi\xi})}$$

10

【 0 0 5 9 】

圧縮行列更新部 1 4 は、以下に例示するステップ S 1 4 2 からステップ S 1 4 5 の処理において、圧縮行列 A の各基底が直交するように、最急降下法の各ステップで直行化する。

20

【 0 0 6 0 】

圧縮行列更新部 1 4 は、圧縮行列 A を、 $A^T A$ の最大固有値で割る（ステップ S 1 4 2）。

【 0 0 6 1 】

圧縮行列更新部 1 4 は、 $A - (3/2)A - (1/2)AA^T A$ とする（ステップ S 1 4 3）。すなわち、 $(3/2)A - (1/2)AA^T A$ を計算して、新たな A とする。

【 0 0 6 2 】

圧縮行列更新部 1 4 は、A の各列を、ノルムが 1 になるように正規化する（ステップ S 1 4 4）。

【 0 0 6 3 】

圧縮行列更新部 1 4 は、A が収束するまで、ステップ S 1 4 1 からステップ S 1 4 4 の処理を繰り返す（ステップ S 1 4 5）。

30

【 0 0 6 4 】

収束判定は、例えば、事前に指定した繰り返し回数に達するかどうかにより判定することができる。すなわち、圧縮行列更新部 1 4 は、ステップ S 1 4 1 からステップ S 1 4 4 の処理を行った回数をカウントして、カウントされた回数が事前に指定した繰り返し回数に達した場合には、A が収束したと判断し、ステップ S 1 4 5 の処理を終了する。事前に指定した繰り返し回数は、例えば 100 に設定することができる。

【 0 0 6 5 】

また、収束判定は、ステップ S 1 4 2 の処理開始前の圧縮行列 A とステップ S 1 4 5 の処理終了後の圧縮行列のフロベニウスノルムが閾値 γ_1 以下であるかを判定することにより行ってもよい。すなわち、圧縮行列更新部 1 4 は、ステップ S 1 4 2 の処理開始前の圧縮行列 A とステップ S 1 4 5 の処理終了後の圧縮行列のフロベニウスノルムが閾値 γ_1 以下である場合には、A が収束したと判断し、ステップ S 1 4 5 の処理を終了する。閾値 γ_1 は、例えば 10^{-4} などに設定できる。

40

【 0 0 6 6 】

このようにして、圧縮行列更新部 1 4 は、入力されたガウス混合モデルパラメータ、負担率及び入力された圧縮行列 A に基づいて、入力された圧縮行列 A を勾配法を用いた最適化により更新する。

【 0 0 6 7 】

< 制御部 1 5 >

50

制御部 15 は、GMMパラメータ と圧縮行列Aの更新が収束するまで、ステップ S 12 からステップ S 14 の処理を繰り返す（ステップ S 15）。

【0068】

収束判定は、例えば、事前に指定した繰り返し回数に達するかどうかにより判定することができる。すなわち、制御部 15 は、ステップ S 12 からステップ S 14 の処理を行った回数をカウントして、カウントされた回数が事前に指定した繰り返し回数に達した場合には、GMMパラメータ と圧縮行列Aの更新が収束したと判断し、ステップ S 145 の処理を終了する。事前に指定した繰り返し回数は、例えば1000に設定することができる。

【0069】

また、収束判定は、以下の式で表される対数尤度Lの上昇が閾値 ϵ_2 以下であるかを判定することにより行ってもよい。すなわち、制御部 15 は、以下の式で例えば表される対数尤度Lの上昇が閾値 ϵ_2 以下である場合には、GMMパラメータ と圧縮行列Aの更新が収束したと判断し、制御部 15 の処理を終了する。閾値 ϵ_2 は、例えば 10^{-5} などに設定できる。対数尤度Lの上昇とは、更新後の対数尤度Lと更新前の対数尤度Lとの差のことである。

【0070】

【数 16】

$$L = \sum_{\tau=1}^L \ln \left(\sum_{k=1}^K w_j N(v_{\tau}; \mu_k^v, \Sigma_k^v, A) \right)$$

【0071】

このようにして、制御部 15 は、更新されたガウス混合モデルパラメータ 及び更新された圧縮行列Aを、負担率計算部 12、ガウス混合モデルパラメータ更新部 13 及び圧縮行列更新部 14 へ新たな入力として、負担率計算部 12、ガウス混合モデルパラメータ更新部 13 及び圧縮行列更新部 14 の処理を繰り返す制御を行う。

【0072】

このように、ガウス混合モデルパラメータを学習する際に圧縮行列Aの最適化を行うことにより、入力x が高次元な場合にも、最適な結合分布をGMMで学習できるようになる。また、結合分布の尤度を最大化する圧縮行列Aが求まるため、このようにして求めたガウス混合モデルパラメータ 及び圧縮行列Aを用いてMMSE 推定を行った場合には、その精度が向上する。

【0073】

また、ガウス混合モデルパラメータを学習する際に圧縮行列Aの最適化を行うことにより、これまでは人手で経験的に行ってきた入力変数の設計/選択を自動化することができる。これにより、このようにして求めたガウス混合モデルパラメータ 及び圧縮行列Aを用いたMMSE 推定を様々な分野で行うことができる。

【0074】

[情報推定装置及び方法]

情報推定装置は、図 2 に示すように、次元圧縮部 51 及び推定部 52 を例えば備えている。

【0075】

情報推定装置が、図 6 に例示する各ステップの処理を行うことにより情報推定方法が実現される。

【0076】

情報推定装置には、情報推定の対象となる入力変数x が入力される。

【0077】

< 次元圧縮部 51 >

次元圧縮部 51 は、収束後の圧縮行列Aを入力として、情報推定の対象となる入力変数x を、ガウス混合モデルパラメータ計算装置 1 及び方法で推定された圧縮行列Aで圧縮する（ステップ S 51）。圧縮された入力変数x である $A^T x$ は、推定部 52 に出力される。

10

20

30

40

50

【 0 0 7 8 】

すなわち、次元圧縮部 5 1 は、ガウス混合モデルパラメータ計算装置 1 及び方法における繰り返し制御により最終的に生成された圧縮行列Aを用いて、情報推定の対象となる入力変数 x を圧縮した $A^T x$ を計算する。

【 0 0 7 9 】

< 推定部 5 2 >

推定部 5 2 は、入力変数 x を圧縮した $A^T x$ を入力として、MMSE推定量を例えば以下のように求める（ステップS 5 2）。

【 0 0 8 0 】

【 数 1 7 】

$$\hat{y}_r = \frac{\sum_{k=1}^K w_k N(A^T x_r; \mu_k^\xi, \Sigma_k^{\xi\xi})}{\sum_{j=1}^K w_j N(A^T x_r; \mu_j^\xi, \Sigma_j^{\xi\xi})} \{ \mu_k^\eta + \Sigma_k^{\eta\xi} (\Sigma_k^{\xi\xi})^{-1} (A^T x_r - \mu_k^\xi) \}$$

10

【 0 0 8 1 】

すなわち、推定部 5 2 は、ガウス混合モデルパラメータ計算装置 1 及び方法における繰り返し制御により最終的に生成されたガウス混合モデルパラメータ を用いて、計算された $A^T x$ が得られた下での y の期待値である \hat{y} を計算する。

【 0 0 8 2 】

このように、ガウス混合モデルパラメータを学習する際に圧縮行列Aの最適化を行うガウス混合モデルパラメータ計算装置 1 及び方法で求めたガウス混合モデルパラメータ 及び圧縮行列Aを用いてMMSE 推定を行うことにより、入力 x が高次元な場合にもMMSE 推定を行うことができる。また、MMSE 推定精度を向上させることができる。また、MMSE 推定を様々な分野で行うことができる。

20

【 0 0 8 3 】

[音強調装置及び方法]

音強調装置は、図 3 に示すように、マイクロホン 2、周波数領域変換部 3、音響特徴量抽出部 4、情報推定装置 5、ウィナーフィルタ計算部 6、フィルタリング部 7、時間領域変換部 8、周波数領域変換部 9 1、重畳部 9 2、音響特徴量抽出部 9 3、事前信号雑音比計算部 9 4 及びガウス混合モデルパラメータ計算装置 1 を例えば備えている。

【 0 0 8 4 】

音強調装置が、図 7 に例示する各ステップの処理を行うことにより音強調方法が実現される。

30

【 0 0 8 5 】

音強調装置及び方法の処理は、学習処理（ステップS 9 1 からステップA 6 の処理）と強調処理（ステップS 2 からステップS 8 の処理）の 2 個の処理に分けることができる。まず、学習処理について説明する。学習処理では、GEMアルゴリズムを用いたガウス混合モデルパラメータの学習が行われる。学習処理は、強調処理に先立って行われてもよいし、強調処理と並行して行われてもよい。

【 0 0 8 6 】

学習処理は、強調処理に先立って行われる場合には、音強調装置は、周波数領域変換部 9 1、重畳部 9 2、音響特徴量抽出部 9 3、事前信号雑音比計算部 9 4 及びガウス混合モデルパラメータ計算装置 1 を備えていなくてもよい。

40

【 0 0 8 7 】

目的音の学習データと雑音の学習データの時間波形を用意する。ここで、時間波形のサンプリングレートや量子化ビット数は任意であるが、例えばサンプリングレートは48kHz、量子化ビット数は16bit などに設定できる。

【 0 0 8 8 】

< 周波数領域変換部 9 1 >

周波数領域変換部 9 1 は、目的音の学習データ及び雑音の学習データを短時間フーリエ変換（STFT）などを用いて周波数領域に変換し、それぞれ収音したい特定の音（目的音）

50

の複素スペクトル S と雑音の複素スペクトル N を得る（ステップS 9 1）。得られた S と N は、重畳部 9 2 及び事前信号雑音比計算部 9 4 に出力される。ここで、フーリエ変換長は1024 点（サンプリング周波数48kHz で約22ms）、シフト長は512 点（サンプリング周波数48kHz で約11ms）などに設定できる。

【 0 0 8 9 】

< 重畳部 9 2 >

重畳部 9 2 は、目的音が含まれる音信号である観測信号を模擬的に設計するために、以下の式に基づき S と N を重畳し、 F を得る（ステップS 9 2）。得られた F は、音響特徴量抽出部 9 3 に出力される。

$$F = S + N$$

10

【 0 0 9 0 】

< 音響特徴量抽出部 9 3 >

音響特徴量抽出部 9 3 は、入力された F から音響特徴量 x を抽出する（ステップS 9 3）。この抽出された音響特徴量 x は、ガウス混合モデルパラメータ計算装置 1 に出力され、GEMアルゴリズムを実行するガウス混合モデルパラメータ計算装置 1 における入力変数 x になる。

【 0 0 9 1 】

音響特徴量 x に用いる特徴量は任意であるが、例えば48次元のメル周波数ケプストラム係数（MFCC）およびその一階差分と二階差分や、48 次元のメルフィルタバンク出力（MFBO）およびその一階差分と二階差分などを用いることができる。また、観測に用いたマイクロホンの数 M が $M>1$ であれば、ビームフォーミングを行って、方向別にMFCC やMFBOを求めることもできる。またその他にも、スペクトルフラックスやスペクトルセントロイドなど、 $D = 512$ 程度の様々な音響特徴量を用いることができる。

20

【 0 0 9 2 】

< 事前信号雑音比計算部 9 4 >

事前信号雑音比計算部 9 4 は、以下の式に基づいて、学習データから得られた S と N を用いて事前SNRを計算する（ステップS 9 4）。計算された事前SNRは、ガウス混合モデルパラメータ計算装置 1 に出力される。この計算結果がGEMアルゴリズムを実行するガウス混合モデルパラメータ計算装置 1 における出力変数 y となる。

【 0 0 9 3 】

30

【数 1 8】

$$y_{\omega, f} = \frac{|S_{\omega, f}|^2}{|N_{\omega, f}|^2}$$

【 0 0 9 4 】

ここで、フーリエ変換長が大きいと、事前SNRの次元 も大きくなるため、上の式の計算結果をメルフィルタバンクで圧縮してもよい。この場合、圧縮された計算結果がガウス混合モデルパラメータ計算装置 1 における出力変数 y となる。メルフィルタバンクの個数はたとえば32 程度に設定できる。

【 0 0 9 5 】

40

< ガウス混合モデルパラメータ計算装置 1 >

ガウス混合モデルパラメータ計算装置 1 は、音響特徴量抽出部 9 3 が計算した音響特徴量を入力変数 x とし、事前信号雑音比計算部 9 4 が計算した事前信号雑音比を出力変数 y として、[ガウス混合モデルパラメータ計算装置及び方法] の欄で説明した手法により、ガウス混合モデルパラメータ 及び圧縮行列 A を計算する（ステップS 1）。計算されたガウス混合モデルパラメータ 及び圧縮行列 A は、情報推定装置 5 に出力される。

【 0 0 9 6 】

なお、ガウス混合モデルパラメータ計算装置 1 に、圧縮後の次元数 Q とGMMの混合数 K 、勾配法のステップサイズ が入力されるが、最終的に音強調をする場合には、これらのパラメータは、目的音のスペクトルの複雑さに応じてチューニングすべきである。スポーツ

50

フィールドでキック音を強調する場合、キック音のスペクトル形状は単純なため、例えば、 $Q=12$ 、 $K=12$ 、 $\alpha=10^{-2}$ などに設定できる。

【 0 0 9 7 】

以上が、音強調装置及び方法の学習処理である。以下、音強調装置及び方法の強調処理について説明する。

【 0 0 9 8 】

< マイクロホン 2 >

M個のマイクロホン 1 を用いて、強調の対象となる目的音が含まれる、所望の音が收音される（ステップ S 2）。ここで、Mは1以上の任意の自然数である。收音された音信号は、周波数領域変換部 3 に出力される。

10

【 0 0 9 9 】

< 周波数領域変換部 3 >

周波数領域変換部 3 は、音信号を周波数領域へ変換し、周波数領域信号 F を得る（ステップ S 3）。すなわち、周波数領域変換部 3 は、強調の対象となる目的音が含まれる音信号を周波数領域信号に変換する。得られた周波数領域信号 F は、音響特徴量抽出部 4 に出力される。

【 0 1 0 0 】

< 音響特徴量抽出部 4 >

音響特徴量抽出部 4 は、入力された周波数領域信号 F から音響特徴量 x を計算する（ステップ S 4）。すなわち、音響特徴量抽出部 4 は、周波数領域信号 F に基づいて、強調の対象となる目的音が含まれる音信号の音響特徴量 x を抽出する。計算された音響特徴量 x は、情報推定装置 5 に出力される。

20

【 0 1 0 1 】

< 情報推定装置 5 >

情報推定装置 5 は、ガウス混合モデルパラメータ計算装置 1 及び方法により計算されたガウス混合モデルパラメータ 及び圧縮行列 A に基づいて、[情報推定装置及び方法] の欄で説明した手法により、事前信号雑音比 y の期待値である \hat{y} を計算する（ステップ S 5）。

【 0 1 0 2 】

すなわち、情報推定装置 5 の次元圧縮部 5 1 は、ガウス混合モデルパラメータ計算装置 1 及び方法により最終的に生成された圧縮行列 A を用いて、抽出された音響特徴量 x を圧縮した $A^T x$ を計算する（ステップ S 5 1）。

30

【 0 1 0 3 】

また、情報推定装置 5 の推定部 5 2 は、ガウス混合モデルパラメータ計算装置 1 及び方法により最終的に生成されたガウス混合モデルパラメータ を用いて、計算された $A^T x$ が得られた下での事前信号雑音比 y の期待値である \hat{y} を計算する（ステップ S 5 2）。

【 0 1 0 4 】

< ウィナーフィルタ計算部 6 >

ウィナーフィルタ計算部 6 は、計算された \hat{y} を用いてウィナーフィルタを求める（ステップ S 6）。求めたウィナーフィルタ G は、フィルタリング部 7 に出力される。

40

【 0 1 0 5 】

具体的には、ウィナーフィルタ計算部 6 は、以下の式に基づき、推定した事前 $\text{SNR} \hat{y}$ からウィナーフィルタ G を計算する。

【 0 1 0 6 】

【 数 1 9 】

$$G_{\omega_f} = \frac{\hat{y}_f}{1 + \hat{y}_f}$$

【 0 1 0 7 】

ただし、学習処理において、事前信号雑音比計算部 9 4 が事前 SNR をメルフィルタバン

50

ク圧縮していた場合、上記式の計算の実行前に、事前 $SNR^{\wedge}y$ を線形周波数領域にスプライン補間する。

【0108】

<フィルタリング部7>

フィルタリング部7は、求めたウィナーフィルタを周波数領域信号 F に適用してフィルタリング後周波数領域信号 H を得る(ステップS7)。具体的には、フィルタリング部7は、以下の式に基づき、フィルタリング後周波数領域信号 H を得る。

【0109】

$$H = G \cdot F$$

得られたフィルタリング後周波数領域信号 H は、時間領域変換部8に出力される。

10

【0110】

<時間領域変換部8>

時間領域変換部8は、入力されたフィルタリング後周波数領域信号 H を時間領域信号に変換する(ステップS8)。

【0111】

具体的には、時間領域変換部8は、フィルタリング後周波数領域信号 H を逆フーリエ変換などの手法を用いて時間領域に変換する。

【0112】

時間領域信号は、目的音が強調された信号として音強調装置から出力される。

【0113】

20

[変形例]

ガウス混合モデルパラメータ装置、情報推定装置及びこれらの方法の用途は、音源強調に限らない。ガウス混合モデルパラメータ装置、情報推定装置及びこれらの方法は、音響強調以外の用途に用いることができる。

【0114】

つまり、入力変数 x は音響特徴量に限らず、画像等のなんらかのデータから特徴量を抽出した結果であってもよいし、位置座標などの生データであってもよい。同様に出力変数 y も事前 SNR に限らず、画像の属するクラス等を表す変数でもよいし、元信号の振幅スペクトル $|S|$ であってもよい。

【0115】

30

例えば、画像データに対する判別を行うMMSE推定を行うためのガウス混合モデルパラメータ及び圧縮行列 A を求める場合には、例えば以下のようにすればよい。判別対象の画像とそれ以外の画像の各データが学習データとして、図4の特徴量抽出部95に入力される。特徴量抽出部95は、コーナー検出やSHIFT(Scale-Invariant Feature Transform)特徴などを用いて、判別対象の画像とそれ以外の画像の各データから特徴量を抽出して入力変数 x とする。また、出力変数 y は、入力変数 x に対応する画像が判別対象の画像ならば $y=1$ それ以外は $y=0$ となるような2値変数とする。ガウス混合モデルパラメータ計算装置1は、これらの入力変数 x 及び出力変数 y に基づいて、[ガウス混合モデルパラメータ計算装置及び方法]の欄で説明した処理を行う。

【0116】

40

その他、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ことはいうまでもない。

【0117】

[プログラム及び記録媒体]

ガウス混合モデルパラメータ計算装置、情報推定装置、音強調装置及びこれらの方法において説明した処理は、記載の順にしたが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にあるいは個別に実行されてもよい。

【0118】

また、ガウス混合モデルパラメータ計算装置、情報推定装置及び音強調装置における各処理をコンピュータによって実現する場合、ガウス混合モデルパラメータ計算装置、情報

50

推定装置及び音強調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、その各処理がコンピュータ上で実現される。

【 0 1 1 9 】

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。

【 0 1 2 0 】

また、各処理手段は、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。

10

【産業上の利用可能性】

【 0 1 2 1 】

この技術は、例えばサッカー等のスポーツ中継などで、ボールのキック音や選手の叫び声、ホイッスルなど、ある特定の競技音だけをクリアに抽出するために例えば用いることができる。もちろん、スポーツ中継以外の分野において、特定の音を強調するためにも用いることができる。

【符号の説明】

【 0 1 2 2 】

- 1 ガウス混合モデルパラメータ計算装置
- 1 1 初期化部
- 1 2 負担率計算部
- 1 3 ガウス混合モデルパラメータ更新部
- 1 4 圧縮行列更新部
- 1 5 制御部
- 2 マイクロホン
- 3 周波数領域変換部
- 4 音響特徴量抽出部
- 5 情報推定装置
- 5 1 次元圧縮部
- 5 2 推定部
- 6 ウィナーフィルタ計算部
- 7 フィルタリング部
- 8 時間領域変換部
- 9 1 周波数領域変換部
- 9 2 重畳部
- 9 3 音響特徴量抽出部
- 9 4 事前信号雑音比計算部
- 9 5 特徴量抽出部

20

30

【図 1】

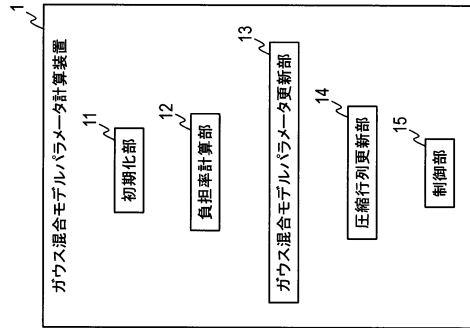


図 1

【図 2】

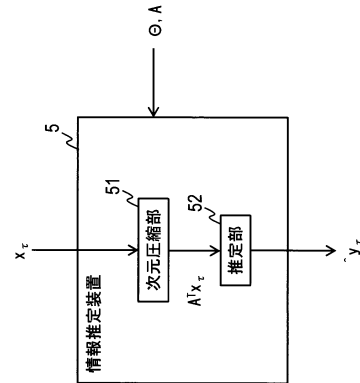


図 2

【図 3】

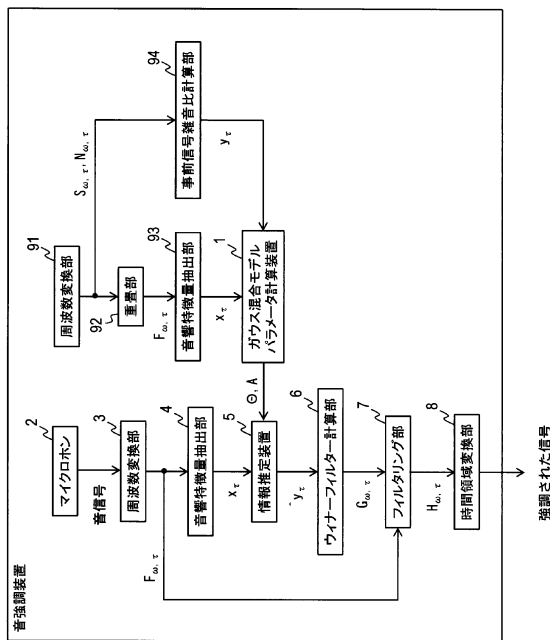


図 3

【図 4】

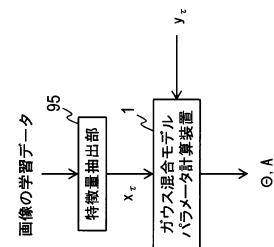


図 4

【図 5】

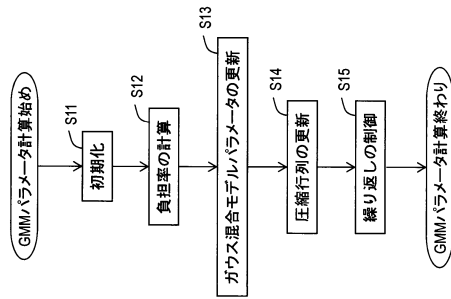


図 5

【図 6】

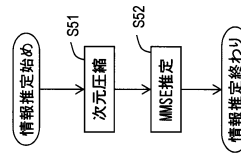


図 6

【図 7】

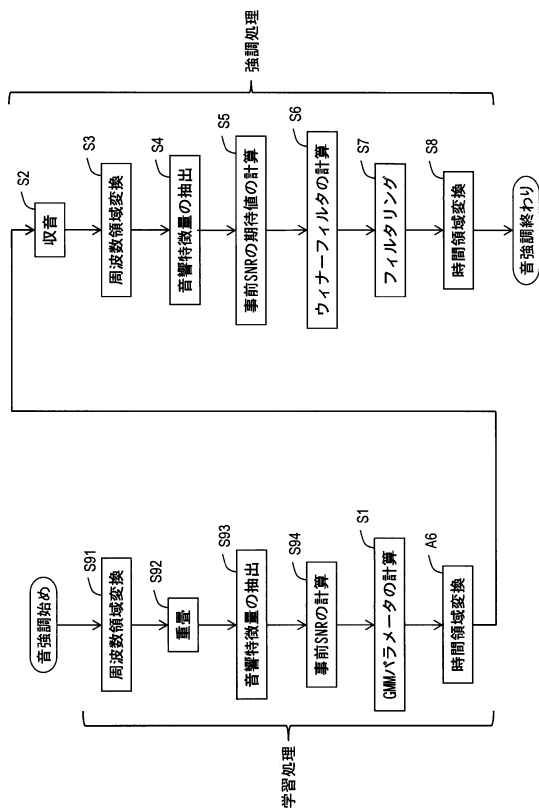


図 7

フロントページの続き

(72)発明者 小林 和則

東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内

(72)発明者 大室 伸

東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内

審査官 上田 雄

(56)参考文献 特開2013-120441(JP, A)

国際公開第2015/093025(WO, A1)

佐藤 美沙紀、外2名、距離評価に基づく認識のための次元圧縮、情報処理学会第67回(平成17年)全国大会講演論文集、社団法人情報処理学会、2005年 3月 2日、pp. 2-269 - 2-270

(58)調査した分野(Int.Cl., DB名)

G10L 15/00-15/34

G10L 21/00-21/18

G06F 17/00-17/18