

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5070591号
(P5070591)

(45) 発行日 平成24年11月14日(2012.11.14)

(24) 登録日 平成24年8月31日(2012.8.31)

(51) Int. Cl.	F I		
G 1 0 L 15/06 (2006.01)	G 1 0 L 15/06	3 1 0 T	
G 1 0 L 15/20 (2006.01)	G 1 0 L 15/20	3 7 0 D	
G 1 0 L 21/02 (2006.01)	G 1 0 L 21/02	1 0 3 Z	

請求項の数 10 (全 32 頁)

(21) 出願番号	特願2007-139634 (P2007-139634)	(73) 特許権者	393031586 株式会社国際電気通信基礎技術研究所 京都府相楽郡精華町光台二丁目2番地2
(22) 出願日	平成19年5月25日(2007.5.25)	(74) 代理人	100099933 弁理士 清水 敏
(65) 公開番号	特開2008-292858 (P2008-292858A)	(72) 発明者	松田 繁樹 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
(43) 公開日	平成20年12月4日(2008.12.4)	(72) 発明者	中村 哲 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
審査請求日	平成22年5月18日(2010.5.18)	(72) 発明者	黒岩 眞吾 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
前置審査			最終頁に続く

(54) 【発明の名称】 雑音抑圧装置、コンピュータプログラム、及び音声認識システム

(57) 【特許請求の範囲】

【請求項1】

雑音が発生する環境下での目的音声の観測により得られる観測信号における雑音の成分を抑圧するための雑音抑圧装置であって、

前記観測信号について所定周期ごとにフレーム化された所定時間長のフレームよりそれぞれ抽出される特徴量を受け、複数のパーティクルを有するパーティクルフィルタを用いて、予め準備された、複数個の要素分布からなる、クリーン音声のための音響モデルに基づき、前記雑音を表す確率分布の推定パラメータを前記フレームごとに逐次生成するための雑音推定手段と、

前記雑音推定手段によりパラメータが推定された雑音の確率分布にしたがって、前記音響モデルを雑音に適応化するための手段と、

前記雑音に適応化された前記音響モデルと、観測信号の特徴量とを用いて、前記フレームごとに前記目的音声の推定特徴量をMMSE推定法により算出するための目的音声推定手段とを含み、

前記複数個の要素分布は、予め複数個のクラスタに分類されており、

前記目的音声推定手段は、前記MMSE推定法において、クリーン音声の特徴量を推定するための、各パーティクルにおける前記雑音に適応化された前記音響モデルの各要素分布のための重みを、前記複数個のクラスタのうち、前記観測信号の特徴量に対応する点に音響空間において最も近い代表点を持つクラスタに含まれる要素分布のみを用いて算出することを特徴とする、雑音抑圧装置。

10

20

【請求項 2】

前記音響モデルは、前記複数個の要素分布からなるガウス混合分布である、請求項 1 に記載の雑音抑圧装置。

【請求項 3】

前記複数個の要素分布のうちいずれか二つの要素分布の分散は互いに異なっている、請求項 2 に記載の雑音抑圧装置。

【請求項 4】

前記複数個の要素分布は、互いに等しい分散を有する、請求項 2 に記載の雑音抑圧装置。

【請求項 5】

前記目的音声推定手段は、前記 M M S E 推定法において、クリーン音声の特徴量を推定するための、各パーティクルにおける前記雑音に適応化された前記音響モデルの各要素分布のための重みを、前記複数個のクラスタのうち、前記観測信号の特徴量に対応する点に音響空間において最も近い代表点を持つ単一のクラスタに含まれる要素分布のみを用いて算出することを特徴とする、請求項 1 ~ 請求項 4 のいずれかに記載の雑音抑圧装置。

10

【請求項 6】

前記複数個のクラスタに含まれる要素分布の数は互いに等しい、請求項 1 ~ 請求項 5 のいずれかに記載の雑音抑圧装置。

【請求項 7】

前記複数個の要素分布のいずれかは、前記複数個のクラスタのうち、2 個以上に属することを特徴とする、請求項 6 に記載の雑音抑圧装置。

20

【請求項 8】

前記複数個のクラスタの各々において、前記代表点は当該クラスタに属する要素分布の重心である、請求項 1 ~ 請求項 7 のいずれかに記載の雑音抑圧装置。

【請求項 9】

コンピュータにより実行されると、当該コンピュータを請求項 1 ~ 請求項 8 のいずれかに記載の雑音抑圧装置として動作させる、コンピュータプログラム。

【請求項 10】

請求項 1 ~ 請求項 8 のいずれかに記載の雑音抑圧装置と、

前記雑音抑圧装置により算出される前記目的音声の推定特徴量を受けて、前記音響モデルと、認識対象言語に関する所定の言語モデルとを用いて、前記目的音声に関する音声認識を行なうための音声認識手段とを含む、音声認識システム。

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、雑音が発生する実環境下での音声認識技術に関し、特に、非定常的な雑音が発生する環境下での音声認識率を改善するための雑音抑圧装置及びそれを使用した音声認識システムに関する。

【背景技術】

【0002】

40

人間にとって容易かつ自然なヒューマンマシンインタフェースを実現するための技術として、音声認識技術が研究されている。近年では、大規模な音声・テキストデータベースと統計確率的な音声認識手法とにより、高い認識率での音声認識が実現されるようになった。今日ではさらに、人間と機械とが接する実環境下において、高速にかつ高い認識率で音声認識を実現するための応用技術開発が進められている。

【0003】

実環境が実験室等の環境と大きく異なる点の一つに、雑音の存在がある。雑音は、無視できない音量で、絶え間なく、かつ不規則に発生し、時間の経過とともに変動する。雑音は、音声認識を行なう際の妨げとなる。雑音が発生する実環境下での音声認識率を改善することは、音声認識の応用技術開発を行なう上で、早急に解決されるべき問題である。

50

【 0 0 0 4 】

雑音が発生する環境下での音声認識率を改善するための技術の一つに、時間の経過に対し定常的な性質を持つ雑音について、音声認識の前処理の段階で雑音を推定し抑圧する技術がある。

【 0 0 0 5 】

後掲の非特許文献 1 には、定常的な雑音の一般的な抑圧方法であるスペクトルサブトラクション法が開示されている。この方法では、発話の前の区間において観測された雑音の振幅スペクトルと発話中の区間における雑音の振幅スペクトルとが同じであると仮定する。そしてこの仮定に基づき、発話時に観測された音声信号の振幅スペクトルから、発話直前に観測された雑音の振幅スペクトルを減算して、雑音を抑圧する。

10

【 0 0 0 6 】

後掲の非特許文献 2 には、分散型音声認識における雑音抑圧方法が開示されている。この方法では、発話直前に観測された雑音の振幅スペクトルを用いて、ウィナフィルタ理論に基づく雑音の抑圧を行なう。

【 0 0 0 7 】

音声認識の前処理の段階において雑音を逐次的に推定し抑圧する技術もある。後掲の非特許文献 3 には、逐次 E M (Expectation Maximization) アルゴリズムを適用して雑音の最尤推定値を逐次的に求める方法が開示されている。逐次 E M アルゴリズムを用いて逐次的に雑音を推定する方法では、雑音の時間変動に対処しつつ高精度に雑音の推定及び抑圧を行なうことができる。

20

【 0 0 0 8 】

後掲の非特許文献 4 及び非特許文献 5 が開示された、カルマンフィルタを用いて雑音の推定値を逐次的に求める方法も一般的に用いられている。この方法では、一期先予測とフィルタリングとを交互に行なうことによって、雑音を逐次的に推定し抑圧する。

【 0 0 0 9 】

また、雑音環境下での音声認識率を改善するための技術として、雑音を考慮した確率モデルを用いて適応的に音声認識を行なう技術がある。例えば後掲の特許文献 1 には、パーティクルフィルタと呼ばれる逐次推定法を用いて、雑音パラメータの推定と、HMM (Hidden Markov Model: 隠れマルコフモデル) を構成する隠れ状態の時間的成長とを行ない、当該 HMM に基づく音声認識を行なう音声認識システムが開示されている。

30

【 0 0 1 0 】

【非特許文献 1】S. F. ボル: 「スペクトルサブトラクションを用いた、音声内の音響ノイズの抑圧」、IEEE Trans. ASSP, Vol. 27, No. 2, 113 - 120 頁、1979 年 (S.F. Boll: "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," IEEE Trans. ASSP, Vol. 27, No. 2, pp. 113-120, 1979)

【非特許文献 2】欧州電気通信標準化機構 (ETSI: European Telecommunications Standards Institute) 勧告 ES 202 050 V1.1.3 “音声の処理、伝送、及び品質の局面 (STQ), 分配型音声認識: 上級フロントエンド 特徴抽出アルゴリズム; 圧縮アルゴリズム”、2003 年 11 月 (ETSI ES 202 050 V1.1.3, “Speech Processing, Transmission and Quality Aspects (STQ), Distributed Speech Recognition: Advanced Front-end Feature Extraction Algorithm; Compression Algorithms,” Nov. 2003.)

40

【非特許文献 3】M. アフィフィ、O. シオアン: 「ロバスト音声認識のための最適な忘却による逐次推定」、IEEE Trans. SAP, Vol. 12, No. 1, 19 - 26 頁、2004 年 (M. Afify, O. Siohan: “Sequential Estimation with Optimal Forgetting for Robust Speech Recognition,” IEEE Trans. SAP, Vol. 12, No.1, pp. 19-26, 2004)

【非特許文献 4】有本 卓: 「カルマンフィルタ」、産業図書

【非特許文献 5】中野 道雄 監修、西山 清 著: 「パソコンで解くカルマンフィルタ

50

」、丸善

【非特許文献6】A.M.ペイナド他、「分散音声認識のための、MMSEによるチャネル誤りの緩和」、ユーロスピーチ2001 スカンジナビア(第7回 音声コミュニケーション及びテクノロジー ヨーロッパ大会)予稿集、pp.2707-2710、2001年(Peinado A M, Sanchez V, Segura J C, Perez-Cordoba J L, "MMSE-Based Channel Error Mitigation for Distributed Speech Recognition," Eurospeech 2001 - Scandinavia (7th European Conference on Speech Communication and Technology), pp.2707-2710, 2001)

【特許文献1】特開2007-41499号公報

【発明の開示】

【発明が解決しようとする課題】

【0011】

非特許文献1及び非特許文献2に記載の技術はいずれも、雑音が定常的なものであるという前提のもとで雑音の推定及び抑圧を行なう技術である。しかし、実環境において雑音の多くは非定常である。すなわち、雑音の音響的特徴は時間の経過に伴い変動する。そのため、非特許文献1及び非特許文献2に記載の技術では、雑音の時間変動に対応できず、高精度に雑音を抑圧することができない。

【0012】

非特許文献3に記載の技術では、逐次EMアルゴリズムが用いられている。逐次EMアルゴリズムで雑音を推定する場合、観測された音声信号のフレーム毎に、そのフレームにおけるパラメータが尤度関数の局所最適値に収束するまで反復計算を行なう必要がある。そのため、雑音の変動する度に膨大な量の計算が必要となり、計算に時間を要する。よって、この手法により実時間で雑音を推定し抑圧するのは困難である。

【0013】

非特許文献4及び非特許文献5に記載の技術では、カルマンフィルタを用いて雑音を推定する。この推定方法は、一期先予測とフィルタリングとを交互に行なう方法であり、逐次EMアルゴリズムのような反復計算を必要とはしない。しかし、カルマンフィルタを用いた手法は、雑音の事後確率分布が単一正規分布であるものとして確率分布を推定する。真の事後確率分布が混合分布であった場合には、単一正規分布で近似される。そのため、精度が劣化する。

【0014】

特許文献1に記載の音声認識システムでは、雑音を考慮したモデルを用いて音声認識を行なうため、音声認識の精度が高くなるという効果が得られている。しかし、パーティクルフィルタを用いているために計算量が大きく、計算資源の限られた装置でこのシステムを高速に動作させるのは困難である。

【0015】

それゆえに、本発明の目的は、非定常雑音が発生する環境下での音声認識率を改善し、かつ限られた計算資源を用いて雑音を短時間で抑圧することができる雑音抑圧装置を提供することである。

【0016】

本発明の他の目的は、非定常雑音が発生する環境下での音声認識率を改善し、かつ計算量を削減しながら雑音を短時間で抑圧することができる雑音抑圧装置を提供することである。

【課題を解決するための手段】

【0017】

本発明の第1の局面に係る雑音抑圧装置は、雑音が発生する環境下での目的音声の観測により得られる観測信号における雑音の成分を抑圧するための雑音抑圧装置である。この雑音抑圧装置は、観測信号について所定周期ごとにフレーム化された所定時間長のフレームよりそれぞれ抽出される特徴量を受け、複数のパーティクルを有するパーティクルフィルタを用いて、予め準備された、複数個の要素分布からなる、クリーン音声のための音響

10

20

30

40

50

モデルに基づき、雑音を表す確率分布の推定パラメータをフレームごとに逐次生成するための雑音推定手段と、雑音推定手段によりパラメータが推定された雑音の確率分布にしたがって、上記音響モデルを雑音に適応化するための手段と、雑音に適応化された音響モデルと、観測信号の特徴量とを用いて、フレームごとに目的音声の推定特徴量をM M S E (最小平均2乗誤差: Minimum Mean Square Error) 推定法により算出するための目的音声推定手段とを含む。複数個の要素分布は、予め複数個のクラスタに分類されている。目的音声推定手段は、M M S E 推定法において、各パーティクルにおける雑音に適応化された音響モデルのための重みを、複数個のクラスタのうち、観測信号の特徴量に対応する点に音響空間において最も近い代表点を持つクラスタに含まれる要素分布のみを用いて算出することを特徴とする。

10

【0018】

M M S E 法の詳細については、非特許文献6に開示されている。

【0019】

なお、本明細書では、雑音抑圧装置に与えられる音声(観測音声)は、雑音のない目的音声と雑音とが重畳した音と考える。このように考えたときの目的音声を「クリーン音声」と呼ぶ。

【0020】

M M S E 推定法においては、非定常雑音に対応するために、各パーティクルにおいて、雑音に適応化された音響モデルのための重みを算出する必要がある。従来はこの算出において、音響モデルの全ての要素分布を用いていた。これに対し本雑音抑圧装置では、予め音響モデルを構成する要素分布をクラスタに分類しておき、音響モデルが雑音に適応化された後に、その代表点を定める。観測信号の特徴量が与えられると、音響空間においてその特徴量に対応する点に最も近い代表点を持つ1又は複数のクラスタが選択され、そのクラスタに含まれる混合分布のみについて、重みが算出される。他のクラスタに含まれる混合分布に対する重みを算出する必要はない。この結果、要素分布の重みを算出する、という計算量の大きな処理を一部の要素分布のみに対して行なえばよくなり、計算量が削減される。その結果、非定常雑音が発生する環境下での音声認識率を改善することが可能で、かつ限られた計算資源を用いて雑音を短時間で抑圧することができる雑音抑圧装置を提供することができる。

20

【0021】

音響モデルは、複数個の要素分布からなるガウス混合分布でもよい。

【0022】

ガウス混合分布を用いると、クリーン音声の特徴量が複雑な分布をしていても、それを統計的にモデル化することが容易になる。

【0023】

複数個の要素分布のうち、いずれか二つが、互いに異なる分散を持つようにしてもよい。

30

【0024】

通常、要素分布は音声サンプルに対する学習により統計的に得られるので、その分散は互いに異なることが多い。

40

【0025】

複数個の要素分布が、互いに異なる平均と、互いに等しい分散とを有するようにしてもよい。

【0026】

このようにすると、雑音推定手段における計算において分布を考慮する必要が事実上なくなり、計算量がさらに削減される。その結果、処理をより早くすることができる。その結果、非定常雑音が発生する環境下での音声認識率を改善することが可能で、かつ限られた計算資源を用いて雑音を短時間で抑圧することができる雑音抑圧装置を提供することができる。

【0027】

50

好ましくは、目的音声推定手段は、M M S E 推定法において、各パーティクルにおける雑音に適応化された音響モデルのための重みを、複数のクラスタのうち、観測信号の特徴量に対応する点に音響空間において最も近い代表点を持つ単一のクラスタに含まれる要素分布のみを用いて算出することを特徴とする。

【0028】

クラスタとして、複数のクラスタではなく単一のクラスタのみを用いる。複数のクラスタに含まれる混合分布の重みを算出する場合と比較して、重み算出のための計算量をさらに削減できる。その結果、定常雑音が発生する環境下での音声認識率を改善し、かつ計算量を削減しながら雑音を短時間で抑圧することができる雑音抑圧装置を提供できる。

【0029】

より好ましくは、複数のクラスタに含まれる要素分布の数は互いに等しい。

【0030】

各クラスタに含まれる要素分布の数を互いに等しくすることで、選択されたクラスタによって算出する重みの数を変化させる必要がない。その結果、重みの算出のための計算を単純化することができ、計算量をさらに削減できる。

【0031】

さらに好ましくは、雑音抑圧装置においては、複数の要素分布のいずれかは、複数のクラスタのうち、2個以上に属することを特徴とする。

【0032】

クラスタ間に重なりを許容することで、重み算出の対象となる要素分布の範囲を広く採ることができる。その結果、定常雑音が発生する環境下での音声認識率を改善し、かつ計算量を削減しながら信頼性高く雑音を短時間で抑圧できる雑音抑圧装置を提供できる。

【0033】

複数のクラスタの各々において、代表点は当該クラスタに属する要素分布の重心であってもよい。

【0034】

本発明の第2の局面に係るコンピュータプログラムは、コンピュータにより実行されると、当該コンピュータを上記したいずれかの雑音抑圧装置として動作させる。

【0035】

本発明の第3の局面に係る音声認識システムは、上記したいずれかの雑音抑圧装置と、この雑音抑圧装置により算出される目的音声の推定特徴量を受けて、目的音声に関する所定の音響モデルと、認識対象言語に関する所定の言語モデルとを用いて、目的音声に関する音声認識を行なうための音声認識手段とを含む。

【発明を実施するための最良の形態】

【0036】

[第1の実施の形態]

以下、図面を参照しつつ、本発明の一実施の形態について説明する。なお、以下の説明に用いる図面では、同一の部品には同一の符号を付してある。それらの名称及び機能も同一である。したがって、それらについての詳細な説明は繰返さない。以下の説明のテキスト中で使用する記号「 \wedge 」等は、本来はその直後の文字の直上に記載すべきものであるが、テキスト記法の制限により当該文字の直前に記載する。式中では、これらの記号等は本来の位置に記載してある。また以下の説明のテキスト中では、ベクトル又は行列については多くの場合、例えば「ベクトル X_t 」、「行列 W 」等のように直前に「ベクトル」、「行列」等を付した通常のテキストの形で記載するが、式中ではいずれも太字で記載する。

【0037】

[構成]

音声認識システム全体の構成

図1に、本実施の形態に係る音声認識システム100全体の構成を示す。図1を参照して、この音声認識システム100は、音源102が発生する音122から音声認識に用いる音声の特徴を表す特徴量ベクトル126を抽出するための前処理部104と、前処理部

10

20

30

40

50

104に接続され、音声の特徴と音素との関係を表す確率モデル（音響モデル）を準備するための前処理用音響モデル部106と、認識対象の言語における単語の連接確率等を表す確率モデル（言語モデル）を準備するための言語モデル部108と、言語モデル部108の言語モデル及び所定の音響モデルを用いて、前処理部104から出力された特徴量に対応する単語等を探索するための探索部110と、探索部110に接続され、探索部110による探索に用いられる音響モデルを準備するための認識用音響モデル部109とを含む。

【0038】

音声認識システム100はさらに、前処理部104による特徴量ベクトル126の抽出に用いられ、後述する状態空間モデルにおける拘束条件を定めるための係数からなる、拘束条件パラメータ138を含む。

10

【0039】

音源102は、認識されるべき音声（目的音声）120を発話する話者116と、話者116の周囲で雑音121を発生する雑音源118とを含む。音源102が発生し前処理部104により収録される音122は、話者116の発話により発生する雑音のない目的音声120と雑音121とが重畳した音となる。本明細書では、前述したように、雑音を含まない目的音声120を「クリーン音声」と呼ぶ。これに対して、前処理部104に到達し、前処理部104により収録される音122、すなわち、クリーン音声120と雑音121とが重畳した音122を、「雑音重畳音声」と呼ぶ。

【0040】

20

前処理部104は、雑音重畳音声122を収録し、その結果得られる観測信号に所定の信号処理を施すことにより、観測信号に関する所定の特徴量ベクトル（以下、この特徴量ベクトルを単に「観測信号の特徴量」と呼ぶことがある。）124を抽出するための計測部112と、計測部112により抽出された観測信号の特徴量124に含まれる雑音の成分を、前処理用音響モデル部106により準備される音響モデルと拘束条件パラメータ138とを用いて抑圧するための雑音抑圧部114とを含む。

【0041】

計測部112は、観測信号を、フレーム間隔10ミリ秒、時間長が数10ミリ秒のフレームごとに対数メルフィルタバンク分析し、得られる対数メルスペクトルを要素とするベクトルを観測信号の特徴量124として出力する。

30

【0042】

雑音抑圧部114は、前処理用音響モデル部106により準備される音響モデルと拘束条件パラメータ138とを用いて、観測信号の特徴量124を基に、クリーン音声120の特徴量ベクトルをフレームごとに逐次推定する機能を持つ。そしてこの逐次推定によって得られる特徴量ベクトルを、音声認識に用いる音声の特徴量ベクトル126として探索部110に出力する。この際、まず雑音121の特徴量ベクトルを推定し、その結果を基にクリーン音声120の特徴量ベクトルを推定する。なお、本明細書では、特徴量ベクトル126によって表される音声を「推定クリーン音声」と呼ぶ。また、特徴量ベクトル126を「推定クリーン音声の特徴量」と呼ぶ。

【0043】

40

探索部110は、推定クリーン音声の特徴量126を用いて、認識用音響モデル部109により準備された音響モデルと、言語モデル部108により準備された言語モデルとを基に、適合する目的言語の単語等を探索し、その結果を認識出力128として出力する。

【0044】

前処理用の音響モデル

以下に、前処理用音響モデル部106により準備される音響モデルについて説明する。図1に示す前処理用音響モデル部106は、クリーン音声120に対する音響モデルとして、ガウス混合モデル（GMM：Gaussian Mixture Model）130を準備し保持する。前処理用音響モデル部106は、予め用意されたクリーン音声120に関する学習データを記憶するための学習データ記憶部132と、学習データ記憶部132内の学習データを

50

用いてGMM 130に対する学習を行なうためのモデル学習部134と、モデル学習部134による学習で得られるGMM 130を記憶するためのGMM記憶部136とを含む。

【0045】

図2に、GMM 130の概念を模式的に示す。図2を参照して、GMM 130は、時系列信号を、一つの定常信号源(状態)によりモデル化した確率モデルである。このGMM 130においては、クリーン音声120の特徴量ベクトルとして出力される可能性のあるベクトルと、そのベクトルが出力される確率(以下、単に「出力確率」と呼ぶ。)とが定義される。出力確率は混合正規分布140によって表現される。GMM 130における混合正規分布140は、複数の要素分布148A, 148B, ..., 148Kを含む。これらの要素分布148A, 148B, ..., 148Kはいずれも単一正規分布である。例えば、混合正規分布140に含まれるある要素分布150を k_t とする。要素分布 k_t は、単一正規分布であり、多次元を考えると、分布の平均ベクトル μ_{s, k_t} (以下単に「平均」と呼ぶ。)と共分散行列 Σ_{s, k_t} (以下単に「分散」と呼ぶ。)とにより表現される。これらはいずれも予め様々な音声サンプルに基づいて統計的に学習(算出)される。この要素分布 k_{t150} にしたがった確率で出力されるパラメータのベクトルをベクトル $S_{k_t, t}$ とする。以下の説明では、GMM 130から出力されるパラメータベクトル $S_{k_t, t}$ を、「(GMM 130の)出力パラメータ」と呼ぶ。

10

【0046】

なお、本実施の形態では、後述するように、観測音声用に適応化された後のGMM 130を参照する際の計算量を削減するために、GMM 130の要素分布はクラスタ化される。図3にその状態を示す。

20

【0047】

図3を参照して、音響空間において、GMM 130は要素分布の集合161を形成している。図3では、図を簡略にするために、要素分布を構成する要素のうち、第1フィルタバンク係数と第2フィルタバンク係数とからなる音響空間での要素分布の分布を示している。実際の音響空間の次元数は特徴ベクトルの次元数と同じである。

【0048】

本実施の形態では、GMM 130は512個の要素分布を含む。図3において、要素分布は実線の楕円で示されている。これら要素分布は、128個のクラスタ162, 164, 166, ...にクラスタ化されている。各クラスタは64個の要素分布を含む。したがって、クラスタ166に属する要素分布170のように一つのクラスタのみに属する要素分布もあれば、クラスタ162及び164に属する要素分布172のように複数のクラスタに属する要素分布もある。このように、ある要素分布が二つ以上のクラスタに属することを許容することで、後述するMSE推定における重み演算のときに、演算対象となる要素分布の数を多くすることができ、その結果、精度が落ちることを防止できる。

30

【0049】

クラスタ化のための手法としては、例えばボトムアップクラスタリング法、K-平均法等、既知の手法を用いることができる。クラスタリングの方法としては、要素分布間の距離の関係を保存できるものであればどのようなものでもよい。すなわち、ある要素分布クラスタについて、そのクラスタに属する要素分布の重心としてそのクラスタの代表点を定めたときに、その中心点とそのクラスタに属する要素分布との間の距離は、その中心点とそのクラスタに属さない要素分布との間の距離のいずれよりも小さい。なお、本実施の形態では、距離としては、各要素分布の平均ベクトルの間のユークリッド距離を使用するものとする。

40

【0050】

状態空間モデル

以下に、状態空間モデルについて説明する。状態空間モデルは、観測信号の生成過程を表した観測方程式と、処理の対象の変化する過程(以下、この過程を「状態遷移過程」と呼ぶ。)を表した状態方程式とからなる動的モデルである。図4に状態空間モデル160を模式的に示す。

50

【 0 0 5 1 】

時刻 t のフレーム（以下、単に「第 t フレーム」と呼ぶ。）における観測信号の特徴量 X_t （図 1 参照）を X_t とする。観測信号の特徴量 X_t は、上記のとおり雑音重畳音声 1 2 2 から得られる対数メルスペクトルを要素に持つベクトルである。この観測信号の特徴量 X_t は、クリーン音声 1 2 0 と雑音 1 2 1 とが重畳した音の対数メルスペクトルを要素に持つ。ここに、第 t フレームにおけるクリーン音声 1 2 0 の対数メルスペクトルを要素に持つベクトルをクリーン音声の特徴量ベクトル S_t とする。また、雑音 1 2 1 の対数メルスペクトルを要素に持つベクトルを雑音の特徴量ベクトル N_t とする。ベクトル X_t 、 S_t 、及び N_t の次元数は同一である。なお、以下に説明する処理はこれらベクトル及び行列の要素についてそれぞれ行なわれるが、以下の説明では、簡単のために各要素を特に区別して言及することはしない。

10

【 0 0 5 2 】

まず、状態空間モデル 1 6 0 における観測信号の生成過程について説明する。観測信号の特徴量 X_t は、計測によって得られる既知のベクトルである。これに対し、クリーン音声の特徴量ベクトル S_t と雑音の特徴量ベクトル N_t とはいずれも、計測によっては得ることのできない未知のベクトルである。

【 0 0 5 3 】

ここで、クリーン音声 1 2 0 の出力過程が GMM でモデル化できるものと仮定する。すなわち、第 t フレームにおけるクリーン音声の特徴量ベクトル S_t が、GMM 1 3 0 内のある要素分布 k_t 1 5 0（図 2 参照）にしたがって出力される出力パラメータベクトル $S_{k_t, t}$ により表現されるものと仮定する。ただし、クリーン音声の特徴量ベクトル S_t と出力パラメータベクトル $S_{k_t, t}$ との間には誤差が存在する。この誤差もまたベクトルである。この誤差を誤差ベクトル V_t とする。誤差ベクトル V_t は、次の式に示すように、平均が 0 で分散が Σ_{S, k_t} の単一正規分布で表現される確率分布にしたがう値を要素に持つものとする。

20

【 0 0 5 4 】

【数 1】

$$V_t \sim \mathcal{N}(0, \Sigma_{S, k_t}),$$

ただし、この式において Σ_{S, k_t} は GMM 1 3 0 内のある要素分布 k_t 1 5 0 より得られるパラメータの共分散行列を表し、記号「 \sim 」は左辺の値が右辺に示される確率分布にしたがうことを示す。すなわち、左辺の値が右辺に示す確率分布にしたがったサンプリングにより推定できることを示す。また、この式において、「 $\mathcal{N}(\mu, \Sigma)$ 」は、平均が μ で分散が Σ の単一正規分布を表す。

30

【 0 0 5 5 】

上記の仮定に基づき、観測信号の特徴量 X_t 1 2 4 の生成過程は、雑音の特徴量ベクトル N_t 、GMM 1 3 0 からの出力パラメータベクトル $S_{k_t, t}$ 、及び誤差ベクトル V_t を用いて、次の式（1）に示す観測方程式により表現されるものとする。

【 0 0 5 6 】

【数 2】

$$\begin{aligned} X_t &= S_{k_t, t} + \log(\mathbf{I} + \exp(N_t - S_{k_t, t})) + V_t \\ &= \mathbf{f}(S_{k_t, t}, N_t) + V_t \end{aligned} \quad (1)$$

40

なお、式（1）で \mathbf{I} は単位ベクトルを表す。また、ベクトルの対数、ベクトルの指数演算はそれぞれ、ベクトルの各要素について対数を取り、又は指数計算し、その結果を成分とするベクトルを表すものとする。

【 0 0 5 7 】

次に、状態空間モデル 1 6 0 における処理対象の状態遷移過程について述べる。状態空間モデル 1 6 0 においては、雑音の特徴量ベクトル N_t が処理の対象になる。ここで、雑

50

音の特徴量ベクトル N_t がランダムウォーク過程にしたがって変化するものと仮定する。すなわち、第 $t-1$ フレームにおける雑音の特徴量ベクトル N_{t-1} が第 t フレームにおける雑音の特徴量ベクトル N_t との間に、ランダムな変化が生じるものと仮定する。このランダムな変化を表すベクトルをランダムガウス雑音ベクトル W_t とする。ランダムガウス雑音ベクトル W_t は、平均が0で分散が Σ_w の単一正規分布で表現される確率分布にしたがう値を要素に持つランダムガウス雑音であるものとする。

【0058】

【数3】

$$W_t \sim \mathcal{N}(0, \Sigma_w)$$

10

ただし、この式において Σ_w は、ランダムガウス雑音ベクトル W_t の共分散行列を表す。

【0059】

上記の仮定に基づき雑音の特徴量ベクトル N_t の状態遷移過程を表現する状態方程式を定義すると、状態方程式は、次の式(2)のようになる。

【0060】

【数4】

$$N_{t+1} = N_t + W_t \quad (2)$$

しかし、ランダムウォーク過程に基づく上記の仮定では、雑音の特徴量ベクトル N_t の変化をランダムガウス雑音ベクトル W_t で規定している。そのため、式(2)に示す状態方程式では、雑音の特徴量ベクトル N_t の時間変化を正確に表現することはできない。そこで、本実施の形態では、図1に示す拘束条件パラメータ138を用いて、雑音の特徴量ベクトル N_t の変化に対し、拘束条件を設ける。拘束条件及びそのための拘束条件パラメータ138の詳細については、後述する。

20

【0061】

雑音抑圧部114の構成

図5に、雑音抑圧部114(図1参照)の構成をブロック図で示す。図5を参照して、雑音抑圧部114は、観測信号の特徴量 X_t 124、GMM130、及び拘束条件パラメータ138を用いて、雑音の特徴量ベクトル N_t の出力される確率を表す確率分布(以下、「雑音確率分布」と呼ぶ。)をフレームごとに逐次的に推定し、雑音確率分布を表すパラメータ(以下、このパラメータを「雑音確率分布の推定パラメータ」と呼ぶ。)を生成するための雑音確率分布推定部200を含む。雑音抑圧部114はさらに、雑音確率分布推定部200により出力された雑音確率分布の推定パラメータ206及びGMM130を基に、観測信号の特徴量 X_t 124の出力される確率を表す確率分布(以下、「観測信号分布」と呼ぶ。)を推定し、観測信号確率分布を表すパラメータ208(以下、このパラメータを「観測信号分布のパラメータ」と呼ぶ。)を生成するための観測信号分布推定部202と、雑音確率分布の推定パラメータ206、観測信号分布のパラメータ208、及びGMM130を基に、推定クリーン音声の特徴量126を生成するためのクリーン音声推定部204とを含む。

30

【0062】

雑音確率分布推定部200は、雑音確率分布をフレームごとに逐次推定し、雑音確率分布の推定パラメータ206を出力する機能を持つ。ここに、観測信号の特徴量 X_0, \dots, X_t からなるベクトルの系列を系列 $X_{0:t} = \{X_0, \dots, X_t\}$ とし、雑音の特徴量ベクトル N_0, \dots, N_t からなるベクトルの系列を系列 $N_{0:t} = \{N_0, \dots, N_t\}$ とする。観測信号ベクトルの系列 $X_{0:t}$ が与えられた時の系列 $N_{0:t}$ の事後確率分布 $p(N_{0:t} | X_{0:t})$ は、1次マルコフ連鎖を用いて、次の式(3)のように表される。

40

【0063】

【数5】

$$p(\mathbf{N}_{0:t}|\mathbf{X}_{0:t}) = p(\mathbf{N}_0|\mathbf{X}_0) \prod_{t'=1}^t p(\mathbf{N}_{t'}|\mathbf{N}_{t'-1})p(\mathbf{X}_{t'}|\mathbf{N}_{t'}) \quad (3)$$

【0064】

したがって雑音の特徴量ベクトル \mathbf{N}_t の確率分布を逐次推定する問題は、観測信号ベクトルの系列 $\mathbf{X}_{0:t}$ が与えられた時の事後確率 $p(\mathbf{N}_{0:t}|\mathbf{X}_{0:t})$ を最大にするような系列 $\mathbf{N}_{0:t}$ を推定する問題に帰着する。雑音確率分布推定部200は、観測信号の特徴量 $X_{t,1}$ 、 $X_{t,2}$ 、 $X_{t,4}$ と、GMM130と、状態空間モデル160と、雑音の状態遷移に関する上記の拘束条件パラメータ138とに基づきこの推定を行なう。その際、雑音確率分布推定部200は、パーティクルフィルタと呼ばれる手法を用いる。この推定法は、ある状態空間モデルにより表現される状態空間内に、局限された状態空間(パーティクル)を多数生成して各パーティクルにおいてパラメータの確率分布を推定し、そして各パーティクルを用いて、状態空間内におけるパラメータの確率分布を近似的に表現する手法である。

10

【0065】

観測信号分布推定部202は、観測信号分布のパラメータ208として、各パーティクルにおける観測信号分布の平均ベクトル及び共分散行列を算出する機能を持つ。観測信号分布のパラメータ208の算出には、例えば、VTS(Vector Taylor Series)法と呼ばれるHMM合成法が用いられる。

20

【0066】

クリーン音声推定部204は、フレームごとに、各パーティクルにおけるクリーン音声のパラメータを推定し、推定クリーン音声の特徴量126を、それら推定されたパラメータの重み付き和によって算出する機能を持つ。推定クリーン音声の特徴量126の算出には、例えば、MSE推定法が用いられる。クリーン音声推定部204はさらに、雑音確率分布推定部200に、次のフレームへの移行に関する要求210を発行する機能を持つ。

【0067】

パーティクルフィルタ

以下に、パーティクルフィルタについて説明する。この手法では、多数のパーティクルにおける初期的なパラメータを、ランダムなサンプリングにより、又は当該パラメータの初期的な状態を表す確率分布からのサンプリングにより決定する。そして、以下の処理をフレームごとに行なう。すなわち、あるフレームに対応して各パーティクルにおいてパラメータが決定されると、まず、各パーティクルのパラメータを当該フレームに後続するフレームに対応するものに更新する。続いて、更新の尤度に応じて各パーティクルに対して重みを付与する。続いて、更新後のパーティクルにおけるパラメータの確率分布にしたがい、当該後続のフレームに対応する各パーティクルのパラメータを再サンプリングする。続いて、再サンプリングされたパラメータを基に、当該後続のフレームに対応する各パーティクルのパラメータを決定する。以上の処理をフレームごとに行なうことにより、逐次的に各パーティクルにおけるパラメータを決定する。

30

40

【0068】

パーティクルフィルタにおいて、状態空間モデル160におけるパラメータはそれぞれ、パーティクルにおけるパラメータの重み付き和によって近似的に表現される。ここに、パーティクルの数を J 個とし、第 t フレームでの、 j ($1 \leq j \leq J$)番目のパーティクルにおける雑音の特徴量ベクトルをベクトル $\mathbf{N}_t^{(j)}$ とする。さらに、第 t フレームにおける j 番目のパーティクルに対する重みを $w_t^{(j)}$ とする。式(3)に示す事後確率分布 $p(\mathbf{N}_{0:t}|\mathbf{X}_{0:t})$ は、次の式(4)に示すモンテカルロサンプリングにより近似的に表現される。

【0069】

【数6】

$$\begin{aligned}
 p(\mathbf{N}_{0:t}|\mathbf{X}_{0:t}) &\simeq \frac{1}{J} \sum_{j=1}^J \delta(\mathbf{N}_{0:t} - \mathbf{N}_{0:t}^{(j)}) \\
 &\simeq \sum_{j=1}^J w_t^{(j)} p(\mathbf{N}_{0:t}^{(j)}|\mathbf{X}_{0:t}) \quad (4)
 \end{aligned}$$

なお、この式において $\delta(\cdot)$ は、Dirac-delta関数を表す。

10

【0070】

j番目のパーティクルにおける雑音の特徴量ベクトルの系列 $\mathbf{N}_{0:t}^{(j)}$ を出力する確率分布を $q(\mathbf{N}_{0:t}^{(j)}|\mathbf{X}_{0:t})$ とすると、パーティクルに対する重み $w_t^{(j)}$ は、次の式(5)により与えられる。

【0071】

【数7】

$$w_t^{(j)} \propto \frac{p(\mathbf{N}_{0:t}^{(j)}|\mathbf{X}_{0:t})}{q(\mathbf{N}_{0:t}^{(j)}|\mathbf{X}_{0:t})} \quad (5)$$

20

確率分布 $q(\mathbf{N}_{0:t}^{(j)}|\mathbf{X}_{0:t})$ は、次の式(6)に示す連鎖モデルで表現されるものとする。

【0072】

【数8】

$$q(\mathbf{N}_{0:t}|\mathbf{X}_{0:t}) = q(\mathbf{N}_t|\mathbf{N}_{0:t-1}, \mathbf{X}_{0:t})q(\mathbf{N}_{0:t-1}|\mathbf{X}_{0:t-1}) \quad (6)$$

また、上記の式(3)の事後確率分布 $p(\mathbf{N}_{0:t}|\mathbf{X}_{0:t})$ は、ベイズ則により次の式(7)のように表現できる。

【0073】

30

【数9】

$$\begin{aligned}
 p(\mathbf{N}_{0:t}|\mathbf{X}_{0:t}) &= \frac{p(\mathbf{N}_t|\mathbf{N}_{t-1})p(\mathbf{X}_t|\mathbf{N}_t)}{p(\mathbf{X}_t|\mathbf{X}_{0:t-1})} p(\mathbf{N}_{0:t-1}|\mathbf{X}_{0:t-1}) \\
 &\propto p(\mathbf{N}_t|\mathbf{N}_{t-1})p(\mathbf{X}_t|\mathbf{N}_t)p(\mathbf{N}_{0:t-1}|\mathbf{X}_{0:t-1}) \quad (7)
 \end{aligned}$$

したがって、式(5)、式(6)、及び式(7)より、パーティクルに対する重み $w_t^{(j)}$ は、式(8)によって与えられることになる。

【0074】

【数10】

40

$$w_t^{(j)} \propto w_{t-1}^{(j)} \frac{p(\mathbf{N}_t^{(j)}|\mathbf{N}_{t-1}^{(j)})p(\mathbf{X}_t|\mathbf{N}_t^{(j)})}{q(\mathbf{N}_t^{(j)}|\mathbf{N}_{0:t-1}^{(j)}, \mathbf{X}_{0:t})} \quad (8)$$

ここで、 $p(\mathbf{N}_t^{(j)}|\mathbf{N}_{t-1}^{(j)}) = q(\mathbf{N}_t^{(j)}|\mathbf{N}_{0:t-1}^{(j)}, \mathbf{X}_{0:t})$ と仮定すると、式(8)より、式(9)が得られる。

【0075】

【数 1 1】

$$w_t^{(j)} \propto w_{t-1}^{(j)} p(\mathbf{X}_t | \mathbf{N}_t^{(j)}) \quad (9)$$

式(9)の $p(\mathbf{X}_t | \mathbf{N}_t^{(j)})$ は、次の式(10)に示す確率密度関数によりモデル化される。

【0076】

【数 1 2】

$$p(\mathbf{X}_t | \mathbf{N}_t^{(j)}) = \mathcal{N}\left(\mathbf{X}_t; \mathbf{f}\left(\mathbf{S}_{k_t^{(j)}, t}^{(j)}, \mathbf{N}_t^{(j)}\right), \Sigma_{\mathbf{S}, k_t^{(j)}}\right) \quad (10)$$

【0077】

雑音確率分布推定部200は、雑音確率分布の推定パラメータ206として、パーティクル j ($1 \leq j \leq J$) ごとに、パーティクルにおける雑音の特徴量ベクトル $\mathbf{N}_t^{(j)}$ に関する式(4)における確率密度関数 $p(\mathbf{N}_{0:t}^{(j)} | \mathbf{X}_{0:t})$ のパラメータと、そのパーティクルに対する重み $w_t^{(j)}$ とを、図4に示す状態空間モデル160に基づいて逐次的に算出する。確率密度関数 $p(\mathbf{N}_{0:t}^{(j)} | \mathbf{X}_{0:t})$ のパラメータは、そのパーティクルにおける雑音の特徴量ベクトル $\mathbf{N}_t^{(j)}$ の平均ベクトル $\hat{\mathbf{N}}_t^{(j)}$ と共分散行列 $\Sigma_{\mathbf{N}_t^{(j)}}$ とを含む。以下、確率密度関数 $p(\mathbf{N}_{0:t}^{(j)} | \mathbf{X}_{0:t})$ の平均ベクトル $\hat{\mathbf{N}}_t^{(j)}$ と共分散行列 $\Sigma_{\mathbf{N}_t^{(j)}}$ とを、「 j 番目のパーティクルにおける雑音のパラメータ」と呼ぶ。

【0078】

状態遷移過程に対する拘束条件

上記のとおり、式(2)に示す状態方程式では、雑音の特徴量ベクトル \mathbf{N}_t の時間変化を正確に表現することはできない。そこで、本実施の形態では、各パーティクルにおける雑音の特徴量ベクトル $\mathbf{N}_t^{(j)}$ ($1 \leq j \leq J$) の変化に対し、次の式(11)に示す状態方程式を導入する。

【0079】

【数 1 3】

$$\mathbf{N}_{t+1}^{(j)} = (1 - \alpha) \mathbf{N}_t^{(j)} + \alpha \hat{\mathbf{N}}_t + \alpha \beta \left(\mu_{\mathbf{N}_t}^{(j)} - \mathbf{N}_t^{(j)} \right) + \mathbf{W}_t^{(j)} \quad (11)$$

この状態方程式(11)において第1項と第2項とは、第 $t+1$ フレームにおけるパーティクルの散らばりを抑制するための拘束条件である。以下この拘束条件を第1の拘束条件と呼ぶ。また、状態方程式(11)において第3項は、 j 番目のパーティクルにおける雑音の特徴量ベクトルの時間推移に対する拘束条件である。以下、この拘束条件を第2の拘束条件と呼ぶ。

【0080】

状態方程式(11)において α は忘却係数であり、 β は第2の拘束条件に対するスケール係数である。

【0081】

第1の拘束条件において、ベクトル $\hat{\mathbf{N}}_t$ は、第 t フレームの各パーティクルにおける雑音の特徴量ベクトル $\mathbf{N}_t^{(1)}, \dots, \mathbf{N}_t^{(j)}$ の加重平均であり、次の式(12)により与えられる。

【0082】

【数 1 4】

$$\hat{\mathbf{N}}_t = \sum_{j=1}^J w_t^{(j)} \mathbf{N}_t^{(j)} \quad (12)$$

10

20

30

40

50

すなわち、第1の拘束条件により、各パーティクルにおける雑音の特徴量ベクトルは、加重平均ベクトル \hat{N}_t に近づくよう補正される。

【0083】

第2の拘束条件において、ベクトル $\mu_{N_t}^{(j)}$ は、j番目のパーティクルにおける過去Tフレーム分の雑音の特徴量ベクトル $N_{t-T+1}^{(j)}, \dots, N_t^{(j)}$ の平均 (Polyak Average) であり、次の式 (13) により与えられる。

【0084】

【数15】

$$\mu_{N_t}^{(j)} = \frac{1}{T} \sum_{s=t-T+1}^t N_s^{(j)} \quad (13) \quad 10$$

すなわち、第2の拘束条件により、パーティクルにおける雑音の特徴量ベクトルにそれぞれ、そのパーティクルにおけるPolyak Averageベクトル $\mu_{N_t}^{(j)}$ がフィードバックされる。本実施の形態では、式 (11) に示す状態方程式の忘却係数 及び第2の拘束条件に対するスケール係数 と、式 (13) におけるフレーム数 T とが、図1に示す拘束条件パラメータ 138 として与えられる。

【0085】

雑音確率分布推定部 200 は、観測方程式 (1) と上記の状態方程式 (11) とにより 20 表される状態空間モデルに基づくパーティクルフィルタを用いて、雑音確率分布の逐次推定を行なう。

【0086】

雑音確率分布推定部 200 の構成

図6に、雑音確率分布推定部 200 の構成をブロック図で示す。図6を参照して、雑音確率分布推定部 200 は、クリーン音声推定部 204 からの要求 210 を受けて、観測信号の特徴量 124 から処理対象となるフレームを選択し、当該フレームに対応する観測信号の特徴量 124 をフレームに応じた出力先に与えるためのフレーム選択部 220 を含む。

【0087】

雑音確率分布推定部 200 はさらに、フレーム選択部 220 から観測信号の特徴量 124 を受けて、初期的な状態における雑音を表す確率分布 (以下、「雑音初期分布」と呼ぶ。) を推定し、多数 (J個) のパーティクルについて、 $t=0$ のフレーム (以下、このフレームを「初期フレーム」と呼ぶ。) における雑音確率分布の推定パラメータ 206 を決定するための雑音初期分布推定部 222 と、フレーム選択部 220 から観測信号の特徴量 124 を受け、各パーティクルについて、 $t(t-1)$ 番目のフレームにおける雑音確率分布の推定パラメータ 206 を、逐次的に算出するための逐次計算部 224 とを含む。

【0088】

フレーム選択部 220 は、要求 210 が与えられる度に、処理対象のフレームを順次選択する。フレーム選択部 220 は、処理対象として初期フレームを選択すると、観測信号 40 の特徴量 X_t 124 のうち最初の所定フレーム分 (例えば10フレーム分) を、雑音初期分布推定部 222 に与える。またフレーム選択部 220 は、処理対象としてその他 ($t-1$) のフレームを選択すると、そのフレームにおける観測信号の特徴量 X_t 124 を逐次計算部 224 に与える。

【0089】

雑音初期分布推定部 222 は、雑音初期分布のパラメータを、以下のようにして推定する。

【0090】

すなわち、雑音初期分布推定部 222 は、雑音初期分布が、単一正規分布であるものとみなして、雑音初期分布を推定する。雑音の初期値ベクトルをベクトル N_0 とし、雑音初 50

期分布を $p(N_0)$ とする。雑音初期分布 $p(N_0)$ における平均ベクトルを μ_N とし、共分散行列を行列 Σ_N とすると、雑音初期分布 $p(N_0)$ は次の式 (14) のように表される。

【0091】
【数16】

$$p(N_0) = \mathcal{N}(N_0; \mu_N, \Sigma_N) \quad (14)$$

雑音初期分布推定部 222 は、最初の所定フレーム分の区間の観測信号の特徴量 $X_{t=0}$ が雑音 N_0 の成分のみからなるものとみなし、式 (14) に示す雑音初期分布 $p(N_0)$ の平均ベクトル μ_N と共分散行列 Σ_N とを推定する。例えば、0 から 9 の 10 フレーム分の区間が雑音 N_0 の成分のみからなる区間に該当する場合、雑音初期分布推定部 222 は、平均ベクトル μ_N と共分散行列 Σ_N とをそれぞれ、次の式 (15) と式 (16) とによって算出する。ただし、式 (16) においてベクトルの右肩に付した「T」は転置を表す。

【0092】
【数17】

$$\mu_N = \frac{1}{10} \sum_{t=0}^9 X_t \quad (15)$$

$$\Sigma_N = \frac{1}{10} \sum_{t=0}^9 (X_t - \mu_N) (X_t - \mu_N)^T \quad (16)$$

そして雑音初期分布推定部 222 は、初期フレーム ($t=0$) での j 番目のパーティクルにおける雑音のパラメータであるベクトル $N_0^{(j)}$ と共分散行列 $\Sigma_{N_0}^{(j)}$ とを、それぞれ、式 (17) 及び式 (18) のように設定する。

【0093】
【数18】

$$N_0^{(j)} \sim \mathcal{N}(\mu_N, \Sigma_N) \quad (17)$$

$$\Sigma_{N_0}^{(j)} = \Sigma_N \quad (18)$$

すなわち、雑音初期分布推定部 222 は、 j 番目のパーティクルにおける雑音の特徴量ベクトル $N_0^{(j)}$ を、初期分布 $p(N_0)$ からのサンプリングによって生成し、共分散行列 $\Sigma_{N_0}^{(j)}$ を、初期分布 $p(N_0)$ の共分散行列 Σ_N に設定する。雑音初期分布推定部 222 は、式 (17) と式 (18) とに示す設定をパーティクル j ($1 \leq j \leq J$) ごとに行なう。

【0094】

逐次計算部 224 は、GMM 130 から出力パラメータ 240 をサンプリングするための GMM サンプリング部 226 を含む。逐次計算部 224 はさらに、観測信号の特徴量 124 を受けて、各パーティクルにおける雑音のパラメータを更新するための更新部 230 と、更新後のパーティクルに対する重みを算出するための重み算出部 232 と、算出された重みに基づき、パーティクルにおける雑音のパラメータを再サンプリングするための再サンプリング部 234 と、再サンプリングされた各パーティクル及び第 $t-1$ フレームにおける各パーティクルに基づき、各パーティクルにおける雑音のパラメータを決定し、雑音確率分布の推定パラメータ 206 を生成するための推定パラメータ生成部 236 とを含む。

【0095】

GMMサンプリング部226は、パーティクル j ($1 \leq j \leq J$)ごとに、GMM130 (図2参照)内の混合分布140から、パーティクルに対応する要素分布 $k_t^{(j)}$ を、その混合重みに基づいてサンプリングする。GMMサンプリング部226はさらに、出力パラメータベクトル $S_{k_t^{(j)}, t}^{(j)}$ を、要素分布 $k_t^{(j)}$ からサンプリングして、更新部230に与える。ここにGMM130における要素分布148A, ..., 148Kの混合重みを P_{S, k_t} とすると、要素分布 $k_t^{(j)}$ は、混合重み P_{S, k_t} を出力確率とする確率分布にしたがう。すなわち、GMM130から次の式(19)に示すサンプリングによって得られる。

【0096】

【数19】

$$k_t^{(j)} \sim P_{S, k_t} \quad (19)$$

10

要素分布 $k_t^{(j)}$ の平均ベクトルをベクトル $\mu_{k_t^{(j)}}$ とし、要素分布 $k_t^{(j)}$ の共分散行列を行列 $\Sigma_{S, k_t^{(j)}}$ とすると、 j 番目のパーティクルにおけるGMM130の出力パラメータベクトル $S_{k_t^{(j)}, t}^{(j)}$ は、要素分布 $k_t^{(j)}$ から、次の式(20)に示すサンプリングによって得られる。

【0097】

【数20】

$$S_{k_t^{(j)}, t}^{(j)} \sim \mathcal{N}(\mu_{S, k_t^{(j)}}, \Sigma_{S, k_t^{(j)}}) \quad (20)$$

20

なお、フレーム選択部220はさらに、GMMサンプリング部226に対し、第 t フレームにおけるGMMの出力パラメータのサンプリングを要求する機能を持つ。

【0098】

更新部230は、上記の観測方程式(1)及び状態方程式(11)からなる動的モデルを状態空間モデルとする拡張カルマンフィルタによって、第 $t-1$ フレームに対応する各パーティクルにおける雑音のパラメータを、第 t フレームに対応するものに更新する機能を持つ。この際、拘束条件パラメータ138と、状態空間モデル160(図4)と、GMMサンプリング部226によりサンプリングされた出力パラメータ $S_{k_t^{(j)}, t}^{(j)}$ とを基にパラメータの更新を行なう。なお、拡張カルマンフィルタは、観測方程式(1)に示すように非線形項を含む状態空間モデルに対応したカルマンフィルタである。

【0099】

図7に、更新部230の構成をブロック図で示す。図7を参照して、更新部230は、第 $t-1$ フレームの雑音確率分布の推定パラメータ206を基に、第 $t-1$ フレームについて、状態方程式(11)の上記した第1の拘束条件に係る加重平均ベクトル \hat{N}_{t-1} を上記の式(12)を用いて算出するための加重平均算出部250を含む。

【0100】

更新部230はさらに、第 $t-1$ フレーム以前の各フレームについて、各パーティクルにおける雑音のパラメータを、パーティクルごとに蓄積するためのバッファメモリ部252と、バッファメモリ部252に蓄積された雑音のパラメータ及び拘束条件パラメータ138により定められるフレーム数 T を基に、各パーティクルについて、第 $t-1$ フレームにおける、上記の式(13)に示す T フレーム分のPolyak Averageベクトル $\mu_{N_{t-1}}^{(j)}$ を算出するためのPolyak Average算出部254と、Polyak Averageベクトル $\mu_{N_{t-1}}^{(j)}$ と第 $t-1$ フレームにおける雑音確率分布の推定パラメータ206とを基に、状態方程式(11)の第2の拘束条件におけるフィードバック分に対応するベクトルを算出するための、フィードバック部256とを含む。フィードバック部256は、Polyak Averageベクトル $\mu_{N_{t-1}}^{(j)}$ と第 $t-1$ フレームにおける平均ベクトル $\hat{N}_{t-1}^{(j)}$ との差分 $\mu_{N_{t-1}}^{(j)} - \hat{N}_{t-1}^{(j)}$ を算出する。

【0101】

更新部230はさらに、観測方程式(1)と状態方程式(11)とからなるモデルを状

30

40

50

態空間モデルとする拡張カルマンフィルタを用いて、第 $t - 1$ フレームに対応するパーティクルにおける雑音のパラメータを、第 t フレームに対応するものに更新するための拡張カルマンフィルタ部 258 とを含む。拡張カルマンフィルタ部 258 は、 j 番目のパーティクルにおける雑音のパラメータの更新に、第 t フレームにおける観測信号の特徴量 $X_{t-1,2,4}$ と、 j 番目のパーティクルにおける GMM 130 (図 2 参照) の出力パラメータベクトル $S_{k_t}^{(j)}(t)$ と、拘束条件パラメータ 138 として与えられる忘却係数 α 及びスケール係数 β と、加重平均ベクトル \hat{N}_{t-1} と、差分 $\mu_{N_{t-1}}^{(j)} - \hat{N}_{t-1}^{(j)}$ とを用いる。

【0102】

本実施の形態における拡張カルマンフィルタの分布更新式を、以下の式 (21) ~ 式 (26) に示す。なお、これらの数式において第 $t - 1$ フレームに対応するパラメータから予測される第 t フレームにおけるパラメータについては添え字として「 $t|t-1$ 」を付してある。

【0103】

【数 21】

$$\mathbf{N}_{t|t-1}^{(j)} = (1 - \alpha)\hat{\mathbf{N}}_{t-1}^{(j)} + \alpha\hat{\mathbf{N}}_{t-1} + \alpha\beta \left(\mu_{\mathbf{N}_{t-1}}^{(j)} - \hat{\mathbf{N}}_{t-1}^{(j)} \right) \quad (21)$$

$$\Sigma_{\mathbf{N}_{t|t-1}}^{(j)} = \Sigma_{\mathbf{N}_{t-1}}^{(j)} + \Sigma_{\mathbf{w}} \quad (22)$$

$$\mathbf{K}_t^{(j)} = \Sigma_{\mathbf{N}_{t|t-1}}^{(j)} \mathbf{F}_t^{(j)T} \left[\mathbf{F}_t^{(j)} \Sigma_{\mathbf{N}_{t|t-1}}^{(j)} \mathbf{F}_t^{(j)T} + \Sigma_{\mathbf{S}, k_t^{(j)}} \right]^{-1} \quad (23)$$

$$\mathbf{F}_t^{(j)} = \partial \mathbf{f} \left(\mathbf{S}_{k_t^{(j)}, t}^{(j)}, \mathbf{N}_{t|t-1}^{(j)} \right) / \partial \mathbf{N}_{t|t-1}^{(j)} \quad (24)$$

$$\hat{\mathbf{N}}_t^{(j)} = \mathbf{N}_{t|t-1}^{(j)} + \mathbf{K}_t^{(j)} \left(\mathbf{X}_t - \mathbf{f} \left(\mathbf{S}_{k_t^{(j)}, t}^{(j)}, \mathbf{N}_{t|t-1}^{(j)} \right) \right) \quad (25)$$

$$\Sigma_{\mathbf{N}_t}^{(j)} = \Sigma_{\mathbf{N}_{t|t-1}}^{(j)} - \mathbf{K}_t^{(j)} \mathbf{F}_t^{(j)} \Sigma_{\mathbf{N}_{t|t-1}}^{(j)} \quad (26)$$

ただし、行列 $\Sigma_{\mathbf{w}}$ は、前述したとおり、第 $t - 1$ フレームから第 t フレームへの状態変化の際に雑音の特徴量ベクトル \mathbf{N}_t に生じるランダムガウス雑音ベクトル \mathbf{W}_{t-1} の共分散行列を表す。

【0104】

再び図 6 を参照して、重み算出部 232 は、第 t フレームでの観測信号の特徴量ベクトル $X_{t-1,2,4}$ と、第 t フレームの各パーティクルにおける GMM 130 の出力パラメータベクトル $S_{k_t}^{(j)}(t)$ と、当該フレームの当該パーティクルにおける雑音のパラメータである平均ベクトル $\hat{N}_t^{(j)}$ 及び共分散行列 $\Sigma_{\mathbf{N}_t}^{(j)}$ と、第 $t - 1$ フレームのパーティクルに対する重み $w_{t-1}^{(j)}$ とを基に、上記の式 (9) 及び式 (10) に示す算出方法を用いて、第 t フレームのパーティクルに対する重み $w_t^{(j)}$ を算出する機能を持つ。なお、重み $w_t^{(j)}$ ($1 \leq j \leq J$) は、 $\sum_{j=1}^J w_t^{(j)} = 1$ となるように正規化される。

【0105】

再サンプリング部 234 は、パラメータが更新されたパーティクルにおける雑音の確率分布にしたがい、第 t フレームに対応する各パーティクルにおける雑音のパラメータを再サンプリングする機能を持つ。この際、再サンプリング部 234 は、微小な重み $w_t^{(j)}$ しか与えられていないパーティクルにおける雑音の確率分布からは、雑音のパラメータの再

10

20

30

40

50

サンプリングを行なわない。一方、大きな重み $w_t^{(j)}$ が与えられているパーティクルにおける確率分布からは、重み $w_t^{(j)}$ の大きさに応じた回数再サンプリングを行ない、得られた雑音のパラメータをそれぞれ、当該再サンプリングの回数と同数のパーティクルに割当てて。ただし再サンプリングの全回数及びパーティクルの全数は一定（ J ）である。このようにするのは、各パーティクルに割当てられる重みが、上記の式（9）から分かるように観測信号の特徴量 $X_{t-1, 2, 4}$ の尤度に対応しているからである。

【0106】

推定パラメータ生成部 236 は、マルコフ連鎖モンテカルロ法のMetropolis-Hastings アルゴリズムにより、第 t フレームに対応するパーティクルを再生成する機能を持つ。図 8 に、推定パラメータ生成部 236 の構成をブロック図で示す。図 8 を参照して、推定パラメータ生成部 236 は、第 $t-1$ フレームに対応する各パーティクルにおける雑音のパラメータを、第 t フレームに対応するものに再更新するための再更新部 262 を含む。再更新部 262 は、再サンプリング部 234 による再サンプリングで得られた各パーティクルにおける雑音のパラメータを用いて、状態空間モデル 160 における雑音確率分布を生成する。そして、当該生成した確率分布と拘束条件パラメータ 138 とに基づき、図 7 に示す更新部 230 と同様の手法を用いて、上記の式（21）～式（26）に示す分布更新式により表される拡張カルマンフィルタによって、各パーティクルにおける雑音のパラメータの再更新を行なう。

10

【0107】

推定パラメータ生成部 236 はさらに、再更新されたパーティクルに対する重み（これを以下「 $w_t^{*(j)}$ 」とする。）を上記の式（9）及び式（10）に示す算出方法を用いて算出するための重み再計算部 264 を含む。

20

【0108】

推定パラメータ生成部 236 はさらに、再サンプリングされたパーティクルに対する重み $w_t^{(j)}$ 及び再更新されたパーティクルに対する重み $w_t^{*(j)}$ から、再更新された雑音のパラメータを許容するか否かの判定に用いる許容確率 を算出するための許容確率算出部 266 と、所定の乱数発生方法により 0 から 1 までの閉区間内の乱数 u を発生させるための乱数発生部 268 と、許容確率 と乱数 u とに基づき、第 t フレームに対応するパーティクルにおけるパラメータとして、再サンプリングされたパーティクルにおける雑音のパラメータと、再更新されたパーティクルにおける雑音のパラメータとの一方を選択するためのパラメータ選択部 270 とを含む。

30

【0109】

許容確率算出部 266 は、重み $w_t^{(j)}$ 及び重み $w_t^{*(j)}$ から次の式（27）にしたがって、許容確率 を算出する機能を持つ。

【0110】

【数22】

$$\nu = \min \left\{ 1, w_t^{*(j)} / w_t^{(j)} \right\} \quad (27)$$

パラメータ選択部 270 は、乱数 u が許容確率 以下であれば、当該パーティクルにおける雑音のパラメータ及びその重みを再更新で得られた新たなパラメータ及びその重みに変更する機能を持つ。

40

【0111】

コンピュータによる実現

以下の説明からも明らかかなように、図 1 に示す音声認識システム 100 の前処理部 104、前処理用音響モデル部 106、及び探索部 110 は、いずれもコンピュータハードウェアと、その上で実行されるプログラムと、コンピュータハードウェアに格納されたデータとにより実現可能である。図 9 に、前処理部 104（図 1 参照）に含まれる雑音抑圧部 114 が行なう雑音抑圧処理を実現するコンピュータプログラムの制御構造をフローチャートで示す。

50

【 0 1 1 2 】

図 9 を参照して、雑音抑圧処理が開始されると、ステップ 2 8 2 において、初期状態における雑音の特徴量 N_0 の各要素の値に対応する初期分布を推定する。すなわち、上記の式 (1 5) 及び式 (1 6) に示す算出方法により、式 (4) に示す雑音初期分布 $p (N_0)$ のパラメータである平均ベクトル μ_N 及び共分散行列 Σ_N を算出する。さらに、式 (1 7) 及び式 (1 8) にしたがって雑音初期分布 $p (N_0)$ からベクトル $N_0^{(j)}$ ($j = 1, \dots, J$) をサンプリングし、初期フレームの各パーティクルにおける雑音のパラメータを推定する。

【 0 1 1 3 】

ステップ 2 8 4 では、雑音抑圧の対象となるフレームを次のフレームに移行させる。以下の説明では、移行後のフレームが第 t フレームであるものとする。ステップ 2 8 6 では、パーティクルフィルタを用いて、処理対象のフレームについて、各パーティクルにおける雑音のパラメータを推定する。すなわち、確率密度関数 $p (N_{0:t}^{(j)} | X_{0:t})$ のパラメータである平均ベクトル $\hat{N}_t^{(j)}$ 及び共分散行列 $\Sigma_{N_t}^{(j)}$ を推定し、さらに、各パーティクルに対する重み $w_t^{(j)}$ を定めて、雑音確率分布の推定パラメータ 2 0 6 を生成する。このステップでの処理については、図 1 2 を用いて後述する。

【 0 1 1 4 】

ステップ 2 8 8 では、観測信号分布のパラメータ 2 0 8 を推定する。すなわち、ステップ 2 8 6 で定めた各パーティクルの雑音のパラメータ $\hat{N}_t^{(j)}$ 、及び $\Sigma_{N_t}^{(j)}$ を用いて、各パーティクルにおける観測信号の特徴量 X_t 1 2 4 の確率分布を推定する。さらに、GMM 1 3 0 を構成する要素分布 $k (1 \leq k \leq K)$ ごとに、パーティクルにおける観測信号の特徴量 X_t 1 2 4 の確率分布の平均ベクトル $\mu_{X_{k,t}}^{(j)}$ と、共分散行列 $\Sigma_{X_{k,t}}^{(j)}$ とを算出する。

【 0 1 1 5 】

図 1 0 を参照して、この処理により、クリーン音声 GMM を構成する要素分布の集合 1 6 1 内の各要素分布が、音響空間内で背景雑音の予測量に応じて移動されることで、観測音声のために適応化された新たな要素分布の集合 3 0 0 が得られたことになる。

【 0 1 1 6 】

この後、MSE 推定法により、第 t フレームにおける推定クリーン音声の特徴量 1 2 6 を算出する。すなわちまず、ステップ 2 8 6 及びステップ 2 8 8 の処理で得られたパラメータを用いて、MSE 推定法によって、MSE 推定値ベクトル \hat{S}_t を算出し、推定クリーン音声の特徴量 1 2 6 (図 1 参照) として出力する。

【 0 1 1 7 】

【 数 2 3 】

$$\hat{S}_t^{(j)} = \mathbf{X}_t - \sum_{k=1}^K P(k | \mathbf{X}_t, (j)) \left(\mu_{\mathbf{X}_{k,t}}^{(j)} - \mu_{\mathbf{S},k} \right) \quad (28)$$

$$\mu_{\mathbf{X}_{k,t}}^{(j)} = \mathbf{f} \left(\mu_{\mathbf{S},k}, \mathbf{N}_t^{(j)} \right) \quad (29)$$

$$\hat{S}_t = \sum_{j=1}^J w_t^{(j)} \hat{S}_t^{(j)} \quad (30)$$

この式において、 $P (k | X_t, (j))$ は、 j 番目のパーティクルにおける、GMM 1 3 0 内の要素分布 k に対する混合重みを表す。混合重み $P (k | X_t, (j))$ は、特許文献 1 では、次の数式により算出されている。

【 0 1 1 8 】

10

20

30

40

【数 2 4】

$$P(k|\mathbf{X}_t, (j)) = \frac{P_{S,k} \mathcal{N} \left(\mathbf{X}_t, \mu_{\mathbf{X}_{k,t}^{(j)}}, \Sigma_{\mathbf{X}_{k,t}^{(j)}} \right)}{\sum_{k'=1}^K P_{S,k'} \mathcal{N} \left(\mathbf{X}_t, \mu_{\mathbf{X}_{k',t}^{(j)}}, \Sigma_{\mathbf{X}_{k',t}^{(j)}} \right)} \quad (31)$$

しかしこの処理には非常に大きな計算量が伴う。しかも、要素の分布 $K = 512$ 、フレーム間隔 10 ミリ秒のレートという設定で音声認識処理を行なう場合には、この計算を毎秒 100 回、合計 512×100 回行なう必要がある。これでは、計算コストが高く、計算機の性能が高い場合はともかく、計算資源の乏しい装置では実現がむずかしい。

10

【0119】

そこで、本実施の形態では、GMM130を予めクラスタ化しておき、このクラスタを用いて次のような一種の近似処理を行なうことにより、この処理を高速化する。

【0120】

まず、ステップ290において、図11に示すように、GMM130を構成する、雑音による適応化後の各クラスタ162A, 164A, 166A, ...内の要素分布の平均ベクトルの音響空間内における位置(ステップ288で算出される)に基づき、各クラスタの代表点302, 304, 306, ...を算出する。代表点としては、前述したとおり、クラスタに属する要素分布の重心を採用する。

20

【0121】

ステップ292において、上記した観測信号の特徴量 $X_{t,124}$ と、ステップ290で算出された128個の要素分布クラスタの代表点と、観測信号の特徴量 $X_{t,124}$ に対応する音響空間内の点との間の距離を算出する。特徴量 $X_{t,124}$ の音響空間における位置に最も近い代表点を持つクラスタをクラスタ C_L として選択する。クラスタ C_L に含まれる64個の要素分布のみを用い、以下の式(32)にしたがい、混合重み $P(k | X_{t, (j)})$ (ただし要素分布 $(k) \in C_L$) が算出される。他の要素分布については無視する。

【0122】

【数 2 5】

$$P(k|\mathbf{X}_t, (j)) = \frac{P_{S,k} \mathcal{N} \left(\mathbf{X}_t, \mu_{\mathbf{X}_{k,t}^{(j)}}, \Sigma_{\mathbf{X}_{k,t}^{(j)}} \right)}{\sum_{m \in C_L} P_{S,m} \mathcal{N} \left(\mathbf{X}_t, \mu_{\mathbf{X}_{m,t}^{(j)}}, \Sigma_{\mathbf{X}_{m,t}^{(j)}} \right)} \quad (32)$$

30

すなわち、512個の要素分布の内、64個のみを $P(k | X_{t, (j)})$ の算出に使用する。この場合の式(32)による計算量は、こうした方式を使用しない場合の $64 / 512 = 1 / 8$ となる。各クラスタの代表点(128個)と観測信号の特徴量 $X_{t,124}$ との間の距離を算出する必要があるため、実際には必要な計算量は目安として $(64 + 128) = 192$ となるが、それでもこうした方式を使用しない場合と比較して計算量は $192 / 512 = 3 / 8$ となる。その結果、計算量が格段に削減され、音声認識処理が高速化できる。又は、計算機の性能が低くなっても同等の性能を維持することができる。

40

【0123】

なお、高速化という目的を達成するためには、GMM130を構成する要素分布の数を C_M 、クラスタの数を N_V 、クラスタ内の要素分布の数を N_D として、 $C_M > N_V + N_D$ が成立する必要がある。

【0124】

このようにして算出された混合重みを用い、ステップ294において、MMSE推定法によって、式(28)~式(30)にしたがってMMSE推定値ベクトル \hat{S}_t を算出し

50

、推定クリーン音声の特徴量 1 2 6 (図 1 参照) として出力する。

【 0 1 2 5 】

続いて、ステップ 2 9 6 では、終了判定を行なう。すなわち第 t フレームが最終のフレームであれば雑音抑圧処理を終了する。さもなければステップ 2 8 4 に戻る。

【 0 1 2 6 】

図 1 2 に、ステップ 2 8 6 (図 9 参照) において行なわれる雑音確率分布の推定パラメータ 2 0 6 の生成処理を実現するプログラムの制御構造をフローチャートで示す。図 1 2 を参照して、雑音確率分布の推定パラメータの生成処理が開始されると、ステップ 3 2 0 において、拡張カルマンフィルタによる更新を行なう際の雑音 1 2 1 の状態遷移過程に対する第 1 及び第 2 の拘束条件に係るパラメータベクトルを算出する。すなわち、第 $t - 1$ フレームのパーティクルでの雑音のパラメータの加重平均ベクトル \hat{N}_{t-1} を式 (1 2) を用いて算出する。そして、パーティクルの各々において、過去 T フレーム分の当該パーティクルにおける雑音のパラメータから Polyak Average ベクトル $\mu_{N_{t-1}}^{(j)}$ を算出し、平均ベクトル $\hat{N}_{t-1}^{(j)}$ との差分 $\mu_{N_{t-1}}^{(j)} - \hat{N}_{t-1}^{(j)}$ を算出する。

【 0 1 2 7 】

ステップ 3 2 2 では、式 (2 1) ~ 式 (2 6) に示す拡張カルマンフィルタを用いて、第 $t - 1$ フレームのパーティクルにおける雑音確率分布から、第 t フレームの各パーティクルにおける雑音のパラメータを推定する。

【 0 1 2 8 】

ステップ 3 2 4 では、第 t フレームの各パーティクルに対する重み $w_t^{(j)}$ を、式 (9) 及び式 (1 0) によって算出する。そして、重み $w_t^{(j)}$ を正規化する。ステップ 3 2 6 では、各パーティクルに対する重み $w_t^{(j)}$ を基に、各パーティクルからの再サンプリングの回数を決定し、当該パーティクルにおける雑音確率分布に基づいてパラメータを再サンプリングする。ステップ 3 2 8 では、Metropolis-Hastings アルゴリズムを用いて第 t フレームのパーティクルを再生成する。

【 0 1 2 9 】

図 1 3 にステップ 3 2 8 (図 1 2 参照) における処理の詳細をフローチャートで示す。図 1 3 を参照して、ステップ 3 2 8 における処理が開始されると、ステップ 3 4 0 において、図 1 2 に示すステップ 3 2 0 と同様に、加重平均ベクトル \hat{N}_{t-1} を、式 (1 2) に示す算出方法で算出する。そして、パーティクルの各々において、過去 T フレーム分の当該パーティクルにおける雑音のパラメータから Polyak Average ベクトル $\mu_{N_{t-1}}^{(j)}$ を算出し、平均ベクトル $\hat{N}_{t-1}^{(j)}$ との差分 $\mu_{N_{t-1}}^{(j)} - \hat{N}_{t-1}^{(j)}$ を算出する。

【 0 1 3 0 】

続くステップ 3 4 2 では、ステップ 3 2 6 (図 1 2 参照) での再サンプリングで得られた各パーティクルにおける雑音パラメータにより表現される雑音確率分布を用いて、式 (2 1) ~ 式 (2 6) に示す拡張カルマンフィルタにより、各パーティクルにおける雑音のパラメータの再更新を行なう。すなわち、第 t フレームのパーティクルを新たに準備し、ステップ 3 2 2 (図 1 2 参照) での処理と同様の処理により、第 $t - 1$ フレームのパーティクルに対応するパラメータから、第 t フレームのパーティクルに対応するパラメータへの再更新を行ない、準備したパーティクルのパラメータに設定する。ステップ 3 4 4 では、ステップ 3 4 2 で準備したパーティクルに対する重み $w_t^{*(j)}$ を、図 1 2 に示すステップ 3 2 4 の処理と同様の処理で算出し正規化する。

【 0 1 3 1 】

ステップ 3 4 6 では、ステップ 3 2 4 の処理で算出された重み $w_t^{(j)}$ と、ステップ 3 4 4 で算出された重み $w_t^{*(j)}$ との比較により、ステップ 3 4 2 で準備されたパーティクルの許容確率 α を定める。ステップ 3 4 8 では、区間 $[0, 1]$ の値からなる一様な集合 $U_{[0,1]}$ の中から任意の値を選択することにより乱数 u を発生する。ステップ 3 5 0 では、ステップ 3 4 8 で発生した乱数 u の値と、ステップ 3 4 6 で定めた許容確率 α の値とを比較する。 u が許容確率の値以下であれば、ステップ 3 5 2 へ進む。さもなければステップ 3 5 4 に進む。ステップ 3 5 2 では、ステップ 3 4 2 で準備されたパーティクルを許容す

10

20

30

40

50

る。すなわち、ステップ 3 2 6 での再サンプリングで得られたパラメータを、準備されたパーティクルのパラメータで置換して処理を終了する。ステップ 3 5 4 では、ステップ 3 4 2 で準備されたパーティクルを棄却する。すなわち、準備されたパーティクル及びそのパラメータを棄却し、処理を終了する。

【 0 1 3 2 】

〔動作〕

本実施の形態に係る音声認識システム 1 0 0 は以下のように動作する。

【 0 1 3 3 】

まず、図 6 に示す雑音確率分布推定部 2 0 0 が、初期フレーム ($t = 0$) における雑音確率分布の推定パラメータ 2 0 6 を生成する動作について説明する。図 1 に示す計測部 1 1 2 が、音源 1 0 2 から雑音重畳音声 1 2 2 を受け、観測信号の特徴量 X_t 1 2 4 を抽出する。抽出された特徴量 X_t 1 2 4 は、雑音抑圧部 1 1 4 の図 6 に示す雑音確率分布推定部 2 0 0 に与えられる。図 6 を参照して、雑音確率分布推定部 2 0 0 のフレーム選択部 2 2 0 は、特徴量 X_t 1 2 4 のうち最初の 1 0 フレーム分を、雑音初期分布推定部 2 2 2 に与える。雑音初期分布推定部 2 2 2 は、上記の式 (1 4) ~ 式 (1 6) に示す処理により雑音初期分布 $p(N_0)$ を推定する。さらに、雑音初期分布 $p(N_0)$ から、上記の式 (1 7) 及び式 (1 8) に示すサンプリングを J 回行なう。このサンプリングによって、各パーティクルにおける雑音の初期的なパラメータであるベクトル $N_0^{(j)}$ 及び共分散行列 $N_0^{(j)}$ が決定される。雑音確率分布推定部 2 0 0 は、これらのパラメータを、初期フレームにおける雑音確率分布の推定パラメータ 2 0 6 として出力する。

【 0 1 3 4 】

次に、雑音確率分布推定部 2 0 0 の逐次推定部 2 2 4 が、第 t フレーム ($t = 1$) における雑音確率分布の推定パラメータ 2 0 6 を生成する動作について説明する。図 6 を参照して、次のフレームの処理の開始要求 2 1 0 に応答して、フレーム選択部 2 2 0 は、GMM サンプリング部 2 2 6 に、第 t フレームにおける GMM の出力パラメータのサンプリングを要求するとともに、観測信号の特徴量 X_t 1 2 4 を更新部 2 3 0 に与える。

【 0 1 3 5 】

GMM サンプリング部 2 2 6 は、GMM 1 3 0 から、出力パラメータベクトル $S^{(j)}_{k_t}$ ($j = 1, \dots, J$) のサンプリングを行なう。例えば、 j 番目のパーティクルにおいて、GMM サンプリング部 2 2 6 が、図 2 に示す GMM 1 3 0 内の混合正規分布 1 4 0 の中から、混合重みにしたがった確率で要素分布 $k_t^{(j)}$ のサンプリングを行なう。その結果、要素分布 $k_t^{(j)}$ として、要素分布 1 5 0 がサンプリングされたものとする。GMM サンプリング部 2 2 6 はさらに、要素分布 $k_t^{(j)}$ により表される出力確率の分布にしたがい、出力パラメータベクトル $S^{(j)}_{k_t}$ をサンプリングする。GMM サンプリング部 2 2 6 は、総数 J の各パーティクルにおける出力パラメータベクトル $S^{(j)}_{k_t}$ をそれぞれ、以上の手順でサンプリングし、図 6 に示す更新部 2 3 0 に与える。

【 0 1 3 6 】

図 1 4 に、逐次計算部 2 2 4 によるパラメータの更新、及び再サンプリングの概要を模式的に示す。図 1 4 においては、ある雑音のパラメータが左右方向に分布し、時間が上から下に進行する。また、図 1 4 においては、パーティクルを白抜きの丸印と黒塗りの丸印とによって模式的に示す。例えば、白抜きの丸印で示すパーティクルが重み $w_t^{(j)}$ の値の微小なパーティクルであり、黒塗りの丸印で示すパーティクルが重み $w_t^{(j)}$ の値の大きなパーティクルであるものとする。

【 0 1 3 7 】

図 1 4 を参照して、第 $t - 1$ フレームに対応するパーティクルにより状態空間 4 2 0 が近似的に表現されているものとする。更新部 2 3 0 は、以下のようにして、状態空間 4 2 0 内の各パーティクルにおける雑音のパラメータを、第 t フレームに対応する状態空間 4 3 0 内の各パーティクルにおける雑音のパラメータに更新する。

【 0 1 3 8 】

まず、図 7 に示す更新部 2 3 0 の拡張カルマンフィルタ部 2 5 8 は、第 $t - 1$ フレーム

の各パーティクルにおける推定確率分布の推定パラメータ206を取得する。取得された推定確率分布の推定パラメータ206は、加重平均算出部250、バッファメモリ252、及びフィードバック部256に与えられる。なお、この時点で、バッファメモリ252には、少なくとも第 $t-1$ フレーム以前の T フレーム分について、推定確率分布の推定パラメータ206が格納されている。

【0139】

図7に示す加重平均算出部250は、推定確率分布の推定パラメータ206が与えられると、式(12)に示す加重平均ベクトル \hat{N}_{t-1} を算出する。この加重平均ベクトル \hat{N}_{t-1} に基づき、式(11)に示す状態方程式における第1の拘束条件を導入して、雑音の平均ベクトルを補正すると、補正後の雑音確率分布における雑音のパラメータは、補正前の平均ベクトル $\hat{N}_{t-1}^{(j)}$ より、加重平均ベクトル \hat{N}_{t-1} に近づく。したがって、パーティクルの散らばりが抑制される。

10

【0140】

新たな推定確率分布の推定パラメータ206がバッファメモリ部252に蓄積されると、Polyak Average算出部254は、バッファメモリ部252に蓄積されている T フレーム分の推定確率分布のパラメータ206を用いて、各パーティクルにおける式(13)に示すPolyak Averageベクトル $\mu_{N_t}^{(j)}$ を算出する。算出したPolyak Averageベクトル $\mu_{N_{t-1}}^{(j)}$ は、フィードバック部256に与えられる。フィードバック部256は、各パーティクルにおいて、Polyak Averageベクトル $\mu_{N_{t-1}}^{(j)}$ と、平均ベクトル $\hat{N}_{t-1}^{(j)}$ との差分 $\mu_{N_{t-1}}^{(j)} - \hat{N}_{t-1}^{(j)}$ を算出する。なお、バッファメモリ部252に推定確率分布の推定パラメータ206が T フレーム分蓄積されていない場合、Polyak Average算出部254は、バッファメモリ部252に蓄積されているだけのフレーム分の雑音確率分布の推定パラメータ206を用いて、Polyak Averageベクトル $\mu_{N_t}^{(j)}$ を算出する。

20

【0141】

図15に、Polyak Average及びフィードバックの概念を模式的に示す。図15(A)及び(B)はいずれも、 j 番目のパーティクルにおけるPolyak Averageベクトル $\mu_{N_t}^{(j)}$ とそのパーティクルに対応する雑音の特徴量ベクトル $N_{t-4}^{(j)}, \dots, N_{t+1}^{(j)}$ との関係を表している。なお、図15(A)は、雑音の特徴量ベクトルの時間遷移が緩やかである場合を示し、図15(B)は、時間遷移が激しい場合を示す。これらの図において、時間は左から右に進行し、雑音の特徴量は上下方向に変化する。図15(A)及び図(B)においては、第 t フレームにおけるPolyak Averageベクトル $\mu_{N_t}^{(j)}$ を、白抜きの丸印で示す。なお、この図に示すPolyak Averageベクトル $\mu_{N_t}^{(j)}$ においては、 $T=5$ フレーム分であるものとする。

30

【0142】

図15(A)を参照して、第 $t-1$ フレームにおける雑音の特徴量 $N_{t-1}^{(j)}$ と、Polyak Averageベクトル $\mu_{N_t}^{(j)}$ の間には、差分 $\mu_{N_t}^{(j)} - N_t^{(j)}$ が生じる。図15(B)に示すような時間遷移の激しい場合においても同様に、雑音の特徴量 $N_t^{(j)}$ と、Polyak Averageベクトル $\mu_{N_t}^{(j)}$ の間には、差分 $\mu_{N_t}^{(j)} - N_t^{(j)}$ が生じる。図15(A)における雑音の特徴量ベクトル $N_{t-4}^{(j)}, \dots, N_t^{(j)}$ の変動に比べて、図15(B)における雑音の特徴量ベクトル $N_{t-4}^{(j)}, \dots, N_t^{(j)}$ の変動は大きい。すなわち図15(A)における雑音の特徴量ベクトル $N_{t-4}^{(j)}, \dots, N_t^{(j)}$ 同士の差異は、図15(B)における当該それらの差異より小さい。

40

【0143】

Polyak Averageベクトル $\mu_{N_t}^{(j)}$ は、 $N_{t-4}^{(j)}, \dots, N_t^{(j)}$ の平均である。そのため、Polyak Averageベクトル $\mu_{N_t}^{(j)}$ のとり得る範囲は、 $N_{t-4}^{(j)}, \dots, N_t^{(j)}$ の最小から最大までの範囲である。したがって、図15(A)に示すように、これらの特徴量ベクトル同士の差異が小さければ、その分Polyak Averageベクトル $\mu_{N_{t-1}}^{(j)}$ のとり得る範囲は狭くなる。差分 $\mu_{N_{t-1}}^{(j)} - N_{t-1}^{(j)}$ の変動幅は自ずから小さくなる。これに対して、図15(B)に示すように雑音の特徴量ベクトル同士の差異が大きければ、その分Polyak Averageベクトル $\mu_{N_t}^{(j)}$ のとり得る範囲は広くなる。差分 $\mu_{N_t}^{(j)} - N_t^{(j)}$ の変動幅も自ず

50

から大きくなる。すなわち、差分 $\mu_{N_t}^{(j)} - N_t^{(j)}$ は、過去 T フレーム分の雑音の変化を反映する。この差分に基づき、次のフレームにおける雑音の特徴量ベクトル $N_{t+1}^{(j)}$ を予測すると、過去 T フレーム分の雑音の変化が反映された特徴量ベクトルが得られる。

【 0 1 4 4 】

拡張カルマンフィルタ部 2 5 8 (図 7 参照) は、加重平均ベクトル \hat{N}_{t-1} と、差分ベクトル $\mu_{N_t}^{(j)} - N_{t-1}^{(j)}$ と、拘束条件パラメータ 1 3 8 により定められる忘却係数及びスケリング係数 と、観測信号の特徴量 X_{t-1} と、出力パラメータ 2 4 0 とを基に、式 (2 1) ~ 式 (2 6) により示す拡張カルマンフィルタによって各パーティクルの更新を行なう。

【 0 1 4 5 】

この更新において、式 (2 1) に示す、雑音の一期先予測パラメータ $N_{t|t-1}^{(j)}$ においては、 $\hat{N}_{t-1}^{(j)}$ の散らばりが抑制される。また、過去 T フレーム分のパラメータの変動がフィードバックされる。すなわち、過去の変動が大きかった場合には、一期先予測パラメータ $N_{t|t-1}^{(j)}$ の変動も大きくなる。反対に過去の変動が小さかった場合には、一期先予測パラメータ $N_{t|t-1}^{(j)}$ の変動も小さくなる。したがって、パラメータの時間推移に対する拘束条件が、過去のパラメータの変動によって強化される。

【 0 1 4 6 】

以上のようにして、各パーティクルの更新を行なわれることにより、図 1 4 に示す状態空間 4 2 0 内の各パーティクルは更新され、パラメータが更新されたパーティクルにより第 t フレームに対応する状態空間 4 3 0 が表現される。

【 0 1 4 7 】

これに応答して、重み算出部 2 3 2 が、状態空間 4 3 0 内の各パーティクルに対する重み $w_t^{(j)}$ を、式 (2 2) 及び式 (2 3) によって算出する。再サンプリング部 2 3 4 は、重み $w_t^{(j)}$ に基づき、パーティクルにおける雑音のパラメータを再サンプリングする。この際、再サンプリング部 2 3 4 はまず、状態空間 4 3 0 内の各パーティクルからの再サンプリングの回数を、パーティクルに対する重み $w_t^{(j)}$ に応じてパーティクルごとに設定する。白抜きの丸印で表される重みの微小なパーティクルからのサンプリングの回数を 0 に設定する。また、黒塗りの丸印で表される重みの大きなパーティクルからのサンプリングの回数を、重みの大きさに応じて 1 ~ 3 に設定する。続いて、状態空間 4 3 0 内のパーティクルにおける雑音確率分布に基づき、設定された回数ずつ、雑音のパラメータの再サンプリングを行なう。このようにして、第 t フレームに対応する新たな状態空間 4 4 0 を表現するパーティクルがそれぞれ形成される。

【 0 1 4 8 】

再サンプリング部 2 3 4 によるこのような再サンプリングが繰返し行なわれると、あるフレームに対応するパーティクルの多くにおける雑音のパラメータが、それ以前の時点のフレームに対応する少数のパーティクルにおける雑音のパラメータの確率分布からサンプリングされたものとなるおそれがある。そこで、推定パラメータ生成部 2 3 6 は、Metropolis-Hastings アルゴリズムを用いて、新たに第 t フレームに対応するパーティクルにおけるパラメータを生成することにより、このような事態を防止する。図 8 に示す再更新部 2 6 2 は、状態空間 4 4 0 における雑音確率分布にしたがい、第 $t - 1$ フレームに対応する状態空間 4 2 0 内のパーティクルにおける雑音のパラメータを再更新する。重み再計算部 2 6 4 は、再更新されたパーティクルに対する重み $w_t^{*(j)}$ を算出する。許容確率算出部 2 6 6 は、再更新されたパーティクルに対する重み $w_t^{*(j)}$ と、再サンプリングされたパーティクルに対する重み $w_t^{(j)}$ とを基に、許容確率 を算出する。パラメータ選択部 2 7 0 は、許容確率 と、乱数発生部 2 6 8 が発生した $[0, 1]$ の区間の乱数 u とを比較し、乱数 u が許容確率 以下であれば、再サンプリングされたパーティクルにおけるパラメータを、再更新されたパーティクルにおけるパラメータで置換する。さもなければ、再更新されたパーティクルにおけるパラメータを棄却する。

【 0 1 4 9 】

以上のような動作をフレームごとに繰返すことにより、各フレームに対応して、各パー

10

20

30

40

50

ティクルにおける雑音のパラメータである、平均ベクトル $\hat{N}_t^{(i)}$ 及び共分散行列 $\Sigma_{N_t}^{(i)}$ が推定される。各パーティクルにおける雑音のパラメータである平均ベクトル $\hat{N}_t^{(i)}$ 及び共分散行列 $\Sigma_{N_t}^{(i)}$ と、各パーティクルに対する重み $w_t^{(i)}$ とが、雑音確率分布の推定パラメータ 206 となる。雑音確率分布推定部 200 は、雑音確率分布の推定パラメータ 206 と観測信号の特徴量ベクトル X_t 124 とを、フレームごとに、図 5 に示す観測信号分布推定部 202 に与える。

【0150】

図 5 を参照して、観測信号分布推定部 202 は、観測信号分布のパラメータ 208 として、VTS 法によって、第 t フレームに対応する各パーティクルにおける観測信号分布の平均ベクトル及び共分散行列を生成する。これにより、各パーティクルにおいて雑音の確率分布と観測信号の確率分布とが推定されたことになる。

10

【0151】

クリーン音声推定部 204 は、MSE 推定法により、第 t フレームに対応する各パーティクルにおいて、クリーン音声 120 の MSE 推定値ベクトル $\hat{S}_t^{(i)}$ を算出する。さらに、MSE 推定値ベクトル $\hat{S}_t^{(i)}$ と重み $w_t^{(i)}$ とを用いて、第 t フレームにおける推定クリーン音声の特徴量ベクトル \hat{S}_t 126 を算出し、図 1 に示す探索部 110 に出力する。

【0152】

MSE 推定値ベクトル $\hat{S}_t^{(i)}$ を算出するにあたり、クリーン音声推定部 204 は、得られた観測信号分布の GMM において、要素分布のクラスタごとに代表点を定める。そして、観測信号の特徴量 X_t 124 が与えられると、その特徴量により音響空間内に定められる点に最も近い代表点を持つクラスタを決定し、そのクラスタに含まれる要素分布のみを用い、式 (32) によって混合重みを算出する。

20

【0153】

したがって本実施の形態では、全ての要素分布 (512 個) に対して式 (31) の計算を行なう必要はない。このための計算は、観測信号の特徴量 X_t 124 により音響空間内に定まる観測点と各クラスタの中心点との間の距離の算出 (128 回)、及び観測点に最も近い中心点を持つクラスタに含まれる 64 個の要素分布を用いて混合重みを算出するだけである。その結果、この処理における計算量は (128 + 64) / 512 = 3 / 8 となり、5 / 8 に相当する計算量を削減することができる。

30

【0154】

図 1 に示す探索部 110 は、推定クリーン音声の特徴量ベクトル \hat{S}_t 126 を用いて、認識用音響モデル部 109 に保持された音響モデルと、言語モデル部 108 に保持された言語モデルとを基に、適合する目的言語の単語等を探索し、その結果を認識出力 128 として出力する。

【0155】

[変形例]

上記実施の形態では、クリーン音響モデルとして、予めサンプル音声に対する統計処理 (学習) により準備したガウス混合分布を用いている。ガウス混合分布は、各次元ごとに複数の要素分布を含む多次元の分布である。事前の学習により、要素分布ごとにその平均と分散とが算出される。したがって多くの場合、要素分布ごとにその平均と分散とは異なっている。そのため、複雑な分布でも統計的にモデル化することができる。この場合、二つの要素分布の平均が一致していたり、二つの要素分布の分散が一致したりしていることはあり得るが、両者が一致することは通常はないと考えられる。

40

【0156】

しかし、本発明による雑音抑圧を実現するためには、学習によって得られた要素分布をそのまま使用しなくてもよい。例えば、上記した実施の形態において、要素分布の平均のみを用い、分散は全ての要素分布において等しいものと仮定しても、上記実施の形態と全く同様の仕組みを用いて雑音抑圧を行なうことができる。この場合には、音響モデルとしては各要素分布の平均のみを記憶しておけばよい。クラスタの分類も、要素分布の平均の

50

みを使用して予め行なっておくことができる。

【0157】

さらに、要素分布の平均の算出にあたって、特徴量を連続的な値として算出してよいし、特徴量を予め離散的なものに定めておき、計算により得られた特徴量を、最も近い離散的特徴量により置換することで量子化してもよい。

【0158】

なお、上記実施の形態では、高速化という目的を達成するためには、GMM130を構成する要素分布の数を C_M 、クラスタの数を N_V 、クラスタ内の要素分布の数を N_D として、 $C_M > N_V + N_D$ が成立するようにしている。しかし、これは上記したように、一つのクラスタのみを用いたMMSEによる雑音除去をする場合についてのみにいえることである。例えば、与えられた観測信号の特徴量 $X_{t,124}$ の音響空間における位置に最も近い代表点を持つ二つのクラスタ C_{L1} と C_{L2} とを採用し、これらのいずれかに属する要素分布のみを用い、前述の式(32)にしたがい、混合重み $P(k | X_t, (j))$ (ただし要素分布 $(k) (C_{L1} C_{L2})$)を算出するようにしてもよい。この場合には、 $C_{L1} C_{L2}$ に属する要素分布の数を N_D として、 $C_M > N_V + N_D$ が成立するようにすればよい。採用するクラスタの数が3以上の場合も同様である。

10

【0159】

また、上記実施の形態では、各クラスタに属する要素分布の数は互いに等しい。しかし本発明はそのような実施の形態には限定されない。各クラスタに属する要素分布の数が互いに異なっても良い。ただしこの場合、精度を維持するためには各クラスタに属する要素分布の数は所定の数以上であることが望ましく、計算の高速化を行なうためには、各クラスタに属する要素分布の数は所定の上限以下であることが望ましい。

20

【0160】

上記実施の形態では、代表点としてクラスタに属する要素分布の重心を採用した。しかし本発明はそのような実施の形態には限定されない。クラスタに属する要素分布との距離の和が最小となるような点をクラスタの代表点としてもよい。

【0161】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内でのすべての変更を含む。

30

【図面の簡単な説明】

【0162】

【図1】本発明の一実施の形態に係る音声認識システム100の構成を示す概略図である。

【図2】GMM130の概念を示す概略図である。

【図3】GMM130のクラスタ化を模式的に示す図である。

【図4】観測信号の状態空間モデル160の概念を示す概略図である。

【図5】雑音抑圧部114の構成を示すブロック図である。

【図6】雑音確率分布推定部200の構成を示すブロック図である。

40

【図7】更新部230の構成を示すブロック図である。

【図8】推定パラメータ生成部236の構成を示すブロック図である。

【図9】雑音抑圧処理の制御構造を示すフローチャートである。

【図10】GMM130の雑音による適応化を示す図である。

【図11】雑音により適応化されたGMM130のクラスタとその代表点との例を示す図である。

【図12】雑音確率分布の推定パラメータ206の生成処理の制御構造を示すフローチャートである。

【図13】Metropolis-Hastingsアルゴリズムによるサンプリング処理の制御構造を示すフローチャートである。

50

【図14】パーティクルフィルタによる処理の概要を示す図である。

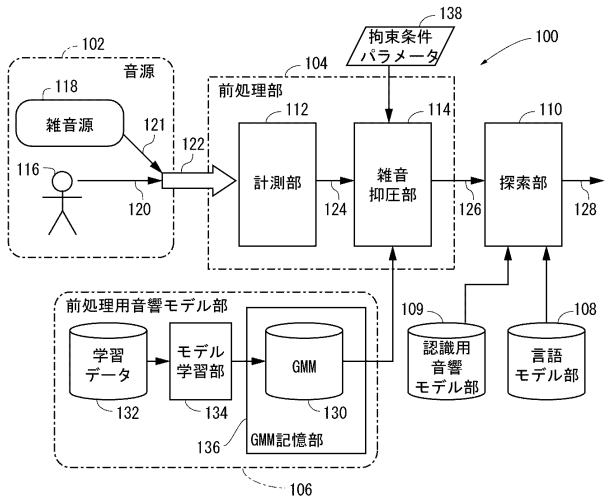
【図15】Polyak Average及びフィードバックの概念を示す模式図である。示す図である。

【符号の説明】

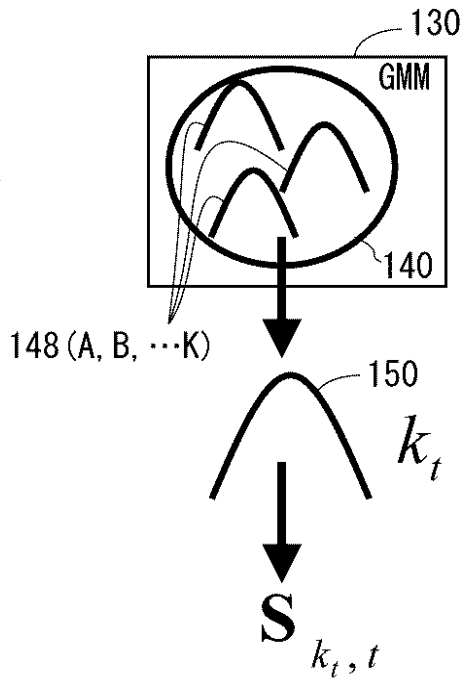
【0163】

100	音声認識システム	
102	音源	
104	前処理部	
106	前処理用音響モデル部	
108	言語モデル部	10
109	認識用音響モデル部	
110	探索部	
112	計測部	
114	雑音抑圧部	
116	話者	
118	雑音源	
120	クリーン音声	
121	雑音	
122	雑音重畳音声	
124	観測信号の特徴量	20
126	推定クリーン音声の特徴量	
130	GMM	
132	学習データ記憶部	
134	モデル学習部	
136	GMM記憶部	
138	拘束条件パラメータ	
160	状態空間モデル	
200	雑音確率分布推定部	
202	観測信号分布推定部	
204	クリーン音声推定部	30
220	フレーム選択部	
222	雑音初期分布推定部	
224	逐次計算部	
226	GMMサンプリング部	
230	更新部	
232	重み算出部	
234	再サンプリング部	
236	推定パラメータ生成部	
240	出力パラメータ	
250	加重平均算出部	40
252	バッファメモリ部	
254	Polyak Average算出部	
256	フィードバック部	
258	拡張カルマンフィルタ部	
262	再更新部	
264	重み再計算部	
266	許容確率算出部	
268	乱数発生部	
270	パラメータ選択部	

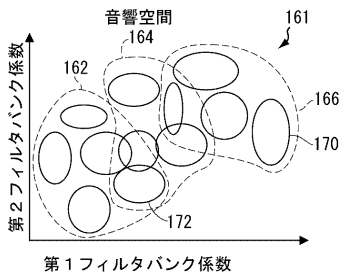
【図1】



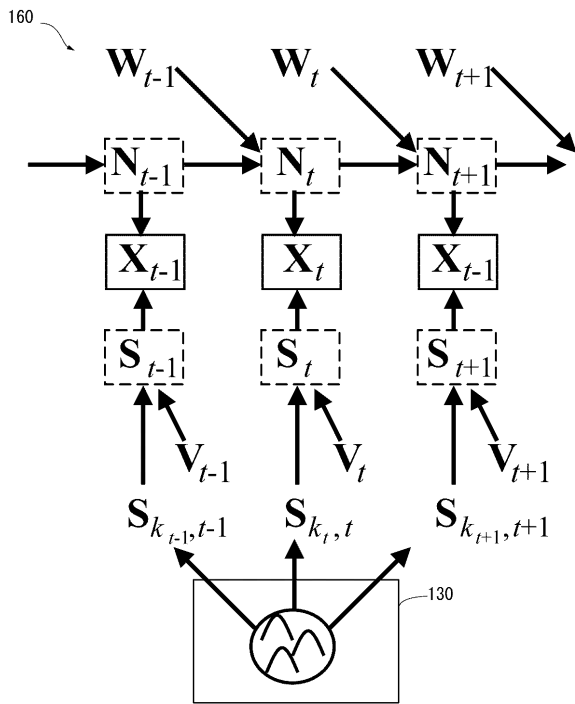
【図2】



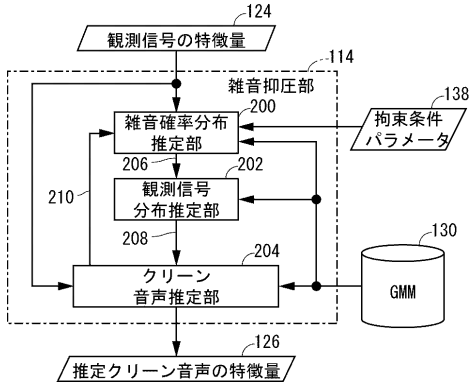
【図3】



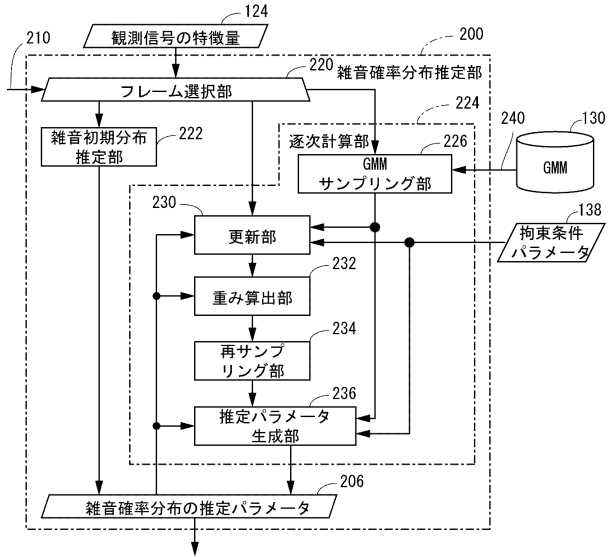
【図4】



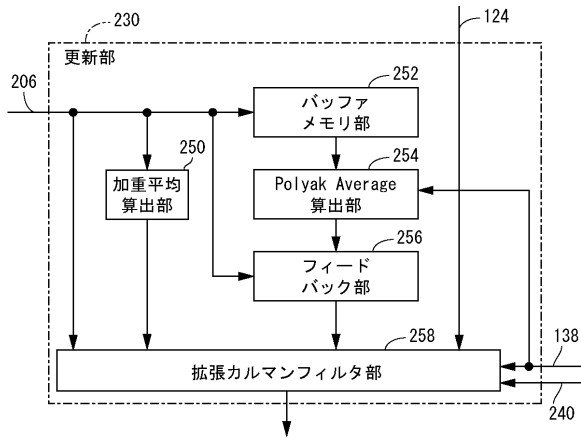
【図5】



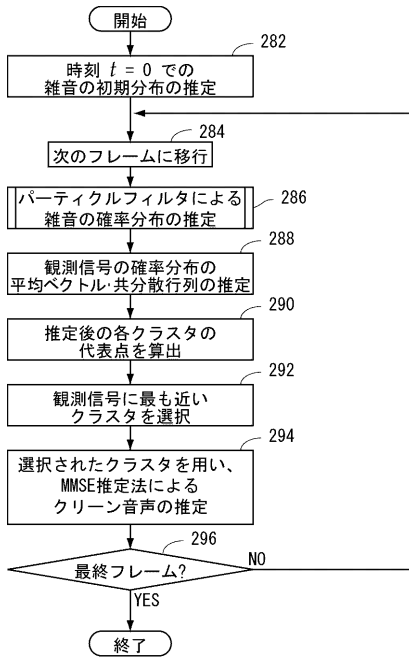
【図6】



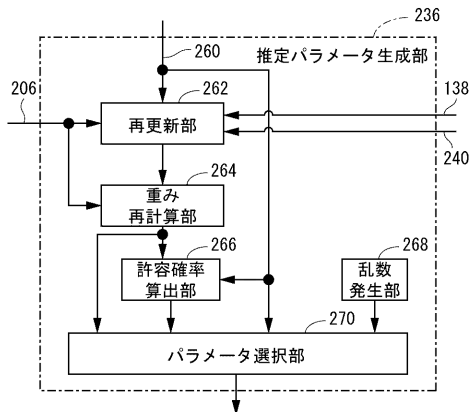
【図7】



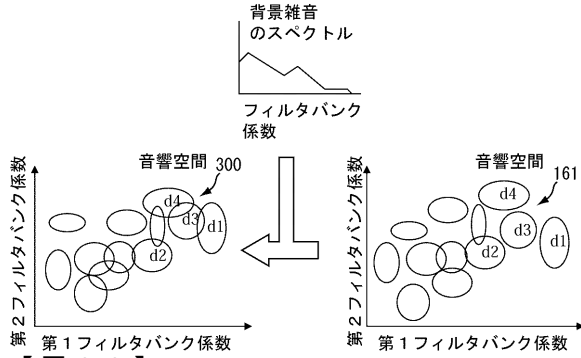
【図9】



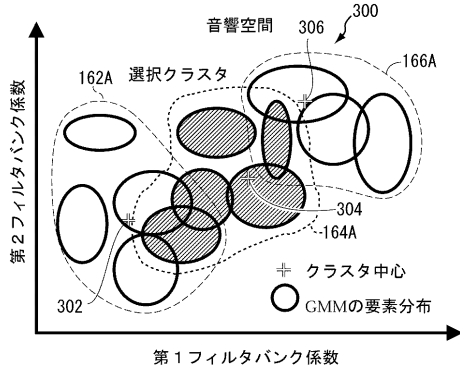
【図8】



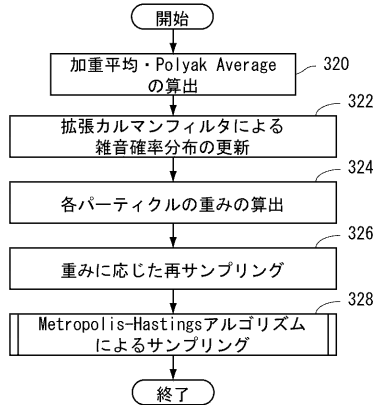
【図10】



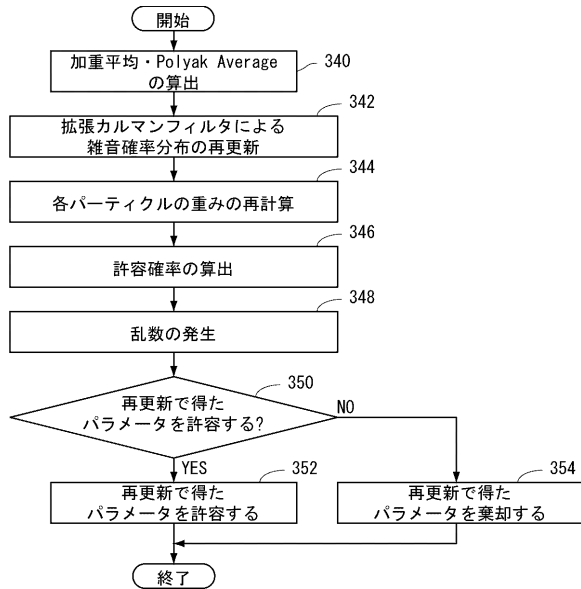
【図11】



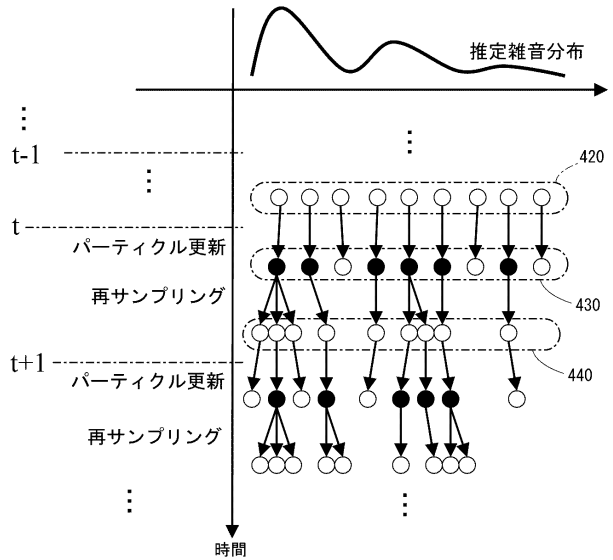
【図12】



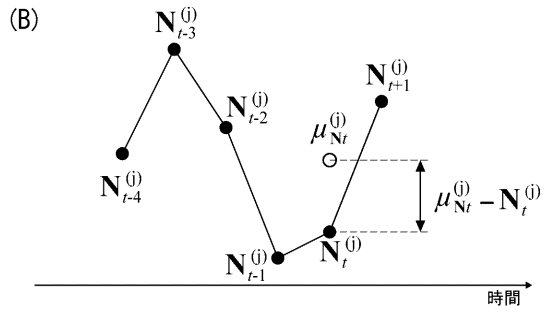
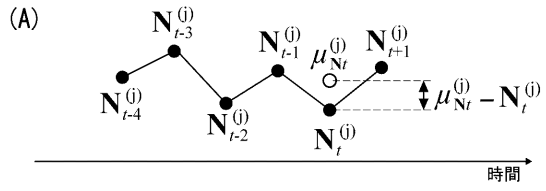
【図13】



【図14】



【 図 15 】



フロントページの続き

審査官 山下 剛史

- (56)参考文献 特開2006-243290(JP,A)
特開2006-201287(JP,A)
特開昭61-73199(JP,A)
特公平2-52278(JP,B2)

- (58)調査した分野(Int.Cl., DB名)
G10L 15/00 - 15/28, 21/02