

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4617497号

(P4617497)

(45) 発行日 平成23年1月26日 (2011. 1. 26)

(24) 登録日 平成22年11月5日 (2010. 11. 5)

(51) Int. Cl.	F I
G 1 O L 15/20 (2006. 01)	G 1 O L 15/20 3 7 O F
G 1 O L 21/02 (2006. 01)	G 1 O L 21/02 1 O 2 B
	G 1 O L 21/02 1 O 1 Z
	G 1 O L 21/02 1 O 3 Z

請求項の数 7 (全 40 頁)

(21) 出願番号	特願2005-356665 (P2005-356665)	(73) 特許権者	393031586
(22) 出願日	平成17年12月9日 (2005. 12. 9)		株式会社国際電気通信基礎技術研究所
(65) 公開番号	特開2007-41499 (P2007-41499A)		京都府相楽郡精華町光台二丁目2番地2
(43) 公開日	平成19年2月15日 (2007. 2. 15)	(74) 代理人	100099933
審査請求日	平成19年5月30日 (2007. 5. 30)		弁理士 清水 敏
(31) 優先権主張番号	特願2005-193470 (P2005-193470)	(72) 発明者	藤本 雅清
(32) 優先日	平成17年7月1日 (2005. 7. 1)		京都府相楽郡精華町光台二丁目2番地2
(33) 優先権主張国	日本国 (JP)		株式会社国際電気通信基礎技術研究所内
		(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
		審査官	菊池 智紀

(出願人による申告) 平成17年度独立行政法人情報通信研究機構、研究テーマ「大規模コーパス音声対話翻訳技術の研究開発」に関する委託研究、産業活力再生特別措置法第30条の適用を受ける特許出願

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、前記雑音の時間推移に関し所定の拘束条件の設けられた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、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、

前記観測信号の特徴量と、前記推定パラメータと、前記目的音声に関する所定の音響モデルとを用いて、前記フレームごとに前記目的音声の推定特徴量を算出するための目的音声推定手段とを含み、

前記雑音推定手段は、

前記雑音の初期分布を推定し、前記複数のパーティクルの各々において、初期フレームの雑音を表す確率分布の推定パラメータを前記初期分布にしたがった確率でそれぞれ設定するための初期パラメータ設定手段と、

各パーティクルにおいて、第1のフレームでの前記推定パラメータと当該第1のフレームに先行する所定数のフレームでの前記推定パラメータとから、前記拘束条件に基づき当該推定パラメータに対する変更量を算出するための変更量算出手段と、

前記音響モデルと、前記観測信号の特徴量と、前記変更量算出手段により算出された変更量とを基に、拡張カルマンフィルタを用いて、各パーティクルにおける前記第1のフレ

10

20

ームの推定パラメータを、それぞれ当該第 1 のフレームに後続する第 2 のフレームに対応するものに更新するための更新手段と、

前記第 2 のフレームにおける前記複数のパーティクルの各々に対する重みを算出するための重み算出手段と、

前記更新手段により更新された雑音の推定パラメータと前記重み算出手段により算出された重みとを用いて、前記第 2 のフレームにおける前記雑音を表す確率分布の推定パラメータを生成するための手段とを含み、

前記拘束条件は、前記推定パラメータに対する過去のフレームからのフィードバックを含み、

前記変更量算出手段は、各パーティクルについて、前記第 1 のフレーム及び当該第 1 のフレームに先行する所定数のフレームにおける前記推定パラメータの平均を求め、当該推定パラメータの平均と前記第 1 フレームにおける前記推定パラメータとの差分を基に、前記変更量を算出するための手段を含む、雑音抑圧装置。

【請求項 2】

さらに、前記変更量算出手段が前記変更量を算出するのに使用するフレームの数を、予め定められた複数種類の数の中から、所定の確率モデルにしたがってフレームごとに選択するためのフレーム数選択手段を含む、請求項 1 に記載の雑音抑圧装置。

【請求項 3】

前記確率モデルは、前記変更量算出手段が前記変更量を算出するのに使用するフレームの数からの変化量の絶対値に対する単調減少関数として、次のフレームで前記変更量算出手段が前記変更量を算出するのに使用するフレームの数が選択される確率を割当て、請求項 2 に記載の雑音抑圧装置。

【請求項 4】

前記フレーム数選択手段は、

互いに異なる、予め定められた複数種類の自然数にそれぞれ対応付けられた複数の状態を準備するための手段と、

前記複数の状態の間での状態遷移確率を決定するための状態遷移確率決定手段とを含み、

前記変更量算出手段が前記変更量を算出するのに用いるフレームの数は、前記複数の状態の中から選択された状態に対応付けられた自然数であり、

前記フレーム数選択手段はさらに、

前記変更量算出手段が前記変更量を算出するのに用いたフレームの数に対応する状態から、前記状態遷移確率にしたがって次の状態を選択するための手段と、

当該選択された状態に対応する自然数を、次のフレームで前記変更量算出手段が使用する際に使用する前記フレームの数に設定するための手段とを含む、請求項 2 に記載の雑音抑圧装置。

【請求項 5】

前記準備するための手段は、互いに異なる、予め定められた複数種類の自然数にそれぞれ対応付けられた複数の状態を準備し、当該状態の各々に対し所定の順序にしたがい番号を付与するための手段を含む、

前記状態遷移確率決定手段は、前記複数の状態のうちの任意の二つの組合せに対し、当該二つの状態にそれぞれ付与された二つの番号の差の絶対値に対する単調減少関数として、当該二つの状態の間での状態遷移確率を決定する、請求項 4 に記載の雑音抑圧装置。

【請求項 6】

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 5 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 7】

請求項 1 ～請求項 5 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、前記

10

20

30

40

50

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。

【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】

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。

【0004】

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。

【0005】

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。

【0006】

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。

【0007】

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EM (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EM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。

【0008】

後掲の非特許文献4及び非特許文献5に開示された、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。

【0009】

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。

【0010】

【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.)

10

【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：「カルマンフィルター」、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」、丸善

20

【特許文献1】特開2002-251198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。

30

【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。

40

【0014】

特許文献1に記載の音声認識システムのように、雑音を考慮したモデルを用いて音声認識を行なう技術では、雑音が重畳された音声と確率モデルとのマッチングが行なわれる。そのため、例えば音響モデル適応など、雑音のない音声に基づいて行なうべき前処理を実行できない。

【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、

50

かつ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】

【0016】

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、雑音の時間推移に関し所定の拘束条件の設けられた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、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、観測信号の特徴量と、推定パラメータと、目的音声に関する所定の音響モデルとを用いて、フレームごとに目的音声の推定特徴量を算出するための目的音声推定手段とを含む。

10

【0017】

好ましくは、雑音推定手段は、雑音の初期分布を推定し、複数のパーティクルの各々において、初期フレームの雑音を表す確率分布の推定パラメータを初期分布にしたがった確率でそれぞれ設定するための初期パラメータ設定手段と、各パーティクルにおいて、第1のフレームでの推定パラメータと当該第1のフレームに先行する所定数のフレームでの推定パラメータとから、拘束条件に基づき当該推定パラメータに対する変更量を算出するための変更量算出手段と、音響モデルと、観測信号の特徴量と、変更量算出手段により算出された変更量とを基に、拡張カルマンフィルタを用いて、各パーティクルにおける第1のフレームの推定パラメータを、それぞれ第1のフレームに後続する第2のフレームに対応するものに更新するための更新手段と、第2のフレームにおける複数のパーティクルの各々に対する重みを算出するための重み算出手段と、更新手段により更新された雑音の推定パラメータと重み算出手段により算出された重みとを用いて、第2のフレームにおける雑音を表す確率分布の推定パラメータを生成するための手段とを含む。

20

【0018】

より好ましくは、拘束条件は、推定パラメータに対する過去のフレームのフィードバックを含む。第1のフレーム及び当該第1のフレームに先行する所定数のフレームにおける推定パラメータの平均を求め、推定パラメータの平均と第1フレームにおける推定パラメータとの差分を基に、変更量を算出するための手段を含む。

【0019】

雑音抑圧装置はさらに、第1のフレームについて、各パーティクルにおける推定パラメータと当該パーティクルに対する重みとを用いて、第1のフレームでの推定パラメータの加重平均を算出するための加重平均算出手段を含んでもよい。更新手段は、音響モデルと、観測信号の特徴量と、変更量及び加重平均とを基に、拡張カルマンフィルタを用いて、各パーティクルにおける第1のフレームの推定パラメータを、それぞれ第1のフレームに後続する第2のフレームに対応するものに更新するための手段を含む。

30

【0020】

さらに好ましくは、雑音抑圧装置はさらに、変更量算出手段が変更量を算出するのに使用するフレームの数を、予め定められた複数種類の数の中から、所定の確率モデルにしたがってフレームごとに選択するためのフレーム数選択手段を含む。

40

【0021】

確率モデルは、変更量算出手段が変更量を算出するのに使用するフレームの数からの変化量の絶対値に対する単調減少関数として、次のフレームで変更量算出手段が変更量を算出するのに使用するフレームの数が選択される確率を割当ててもよい。

【0022】

フレーム数選択手段は、互いに異なる、予め定められた複数種類の自然数にそれぞれ対応付けられた複数の状態を準備するための手段と、複数の状態の間での状態遷移確率を決定するための状態遷移確率決定手段とを含んでもよい。変更量算出手段が変更量を算出するのに用いるフレームの数は、複数の状態の中から選択された状態に対応付けられた自然数である。フレーム数選択手段はさらに、変更量算出手段が変更量を算出するのに用いた

50

フレームの数に対応する状態から、状態遷移確率にしたがって次の状態を選択するための手段と、選択された状態に対応する自然数を、次のフレームで変更量算出手段が使用する際に使用するフレームの数に設定するための手段とを含む。

【0023】

準備するための手段は、互いに異なる、予め定められた複数種類の自然数にそれぞれ対応付けられた複数の状態を準備し、状態の各々に対し所定の順序にしたがい番号を付与するための手段を含んでもよい。状態遷移確率決定手段は、複数の状態のうちの任意の二つの組合せに対し、それら二つの状態にそれぞれ付与された二つの番号の差の絶対値に対する単調減少関数として、それら二つの状態の間での状態遷移確率を決定する。

【0024】

付与するための手段は、状態の各々に対し、自然数の昇順又は降順に番号を付与してもよい。

【0025】

好ましくは、拘束条件は、所定の制御パラメータにより規定される。雑音抑圧装置はさらに、制御パラメータの値を、予め定められた複数種類の数の中から、所定の確率モデルにしたがってフレームごとに選択するための制御パラメータ値選択手段を含む。

【0026】

より好ましくは、確率モデルは、拘束条件を規定する制御パラメータの値からの変化量の絶対値に対する単調減少関数として、次のフレームでの拘束条件を規定する制御パラメータの値が選択される確率を割当てて。

【0027】

より好ましくは、制御パラメータ値選択手段は、互いに異なる、予め定められた複数種類の値にそれぞれ対応付けられた複数の状態を準備するための手段と、複数の状態の間での状態遷移確率を決定するための状態遷移確率決定手段とを含む。制御パラメータの値は、複数の状態の中から選択された状態に対応付けられた値であり、制御パラメータ値選択手段はさらに、拘束条件を規定する制御パラメータの値に対応する状態から、状態遷移確率にしたがって次の状態を選択するための手段と、選択された状態に対応する値を、次のフレームでの拘束条件を規定する制御パラメータに設定するための手段とを含む。

【0028】

さらに好ましくは、準備するための手段は、互いに異なる、予め定められた複数種類の値にそれぞれ対応付けられた複数の状態を準備し、状態の各々に対し所定の順序にしたがい番号を付与するための手段を含む。状態遷移確率決定手段は、複数の状態のうちの任意の二つの組合せに対し、それら二つの状態にそれぞれ付与された番号の差の絶対値の単調減少関数として、それら二つの状態の間での状態遷移確率を決定する。

【0029】

付与するための手段は、状態の各々に対し、値の昇順又は降順に番号を付与してもよい。

【0030】

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本発明の第1の局面に係るいずれかの雑音抑圧装置として動作させる。

【0031】

本発明の第3の局面に係る音声認識システムは、本発明の第1の局面に係るいずれかの雑音抑圧装置と、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】

【0032】

〔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

10

20

30

40

50

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例えば「ベクトル X_t 」、「行列 Σ_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。

【0033】

〔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接続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。

【0034】

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。

【0035】

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が発生する雑音源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、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。

【0036】

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。

【0037】

計測部112は、観測信号を時間長数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。

【0038】

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

ル 1 2 6 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 1 2 6 を「推定クリーン音声の特徴量」と呼ぶ。

【0039】

探索部 1 1 0 は、推定クリーン音声の特徴量 1 2 6 を用いて、認識用音響モデル部 1 0 9 により準備された音響モデルと、言語モデル部 1 0 8 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 1 2 8 として出力する。

【0040】

〈前処理用の音響モデル〉

以下に、前処理用音響モデル部 1 0 6 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 1 0 6 は、クリーン音声 1 2 0 に対する音響モデルとして、ガウス混合モデル (GMM: Gaussian Mixture Model) 1 3 0 を準備し保持する。前処理用音響モデル部 1 0 6 は、予め用意されたクリーン音声 1 2 0 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 1 3 2 と、学習データ記憶部 1 3 2 内の学習データを用いて GMM 1 3 0 に対する学習を行なうためのモデル学習部 1 3 4 と、モデル学習部 1 3 4 による学習で得られる GMM 1 3 0 を記憶するための GMM 記憶部 1 3 6 とを含む。

【0041】

図 2 に、GMM 1 3 0 の概念を模式的に示す。図 2 を参照して、GMM 1 3 0 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 1 3 0 においては、クリーン音声 1 2 0 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 1 4 0 によって表現される。GMM 1 3 0 における混合正規分布 1 4 0 は、複数の要素分布 1 4 8 A, 1 4 8 B, ..., 1 4 8 K を含む。これらの要素分布 1 4 8 A, 1 4 8 B, ..., 1 4 8 K はいずれも単一正規分布である。例えば、混合正規分布 1 4 0 に含まれるある要素分布 1 5 0 を k_t とする。要素分布 k_t は、単一正規分布であり、分布の平均 μ_{S, k_t} と分散 Σ_{S, k_t} とにより表現される。この要素分布 k_t 1 5 0 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 1 3 0 から出力されるパラメータベクトル $S_{k_t, t}$ を、「(GMM 1 3 0 の) 出力パラメータ」と呼ぶ。

【0042】

〈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 3 に状態空間モデル 1 6 0 を模式的に示す。

【0043】

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 1 2 4 (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。

【0044】

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。

【0045】

ここで、クリーン音声120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、GMM130内のある要素分布 k_t 150（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 $S_{k_t,t}$ 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。

【0046】

10

【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GMM130内のある要素分布 k_t 150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。

【0047】

上記の仮定に基づき、観測信号の特徴量 X_t 124の生成過程は、雑音の特徴量ベクトル N_t 、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。

20

【0048】

【数2】

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

なお、式（1）で 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。

30

【0049】

次に、状態空間モデル160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160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。

40

【0050】

【数3】

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

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。

【0051】

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。

【0052】

【数 4】

$$\mathbf{N}_{t+1} = \mathbf{N}_t + \mathbf{W}_t \quad (2)$$

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 $\mathbf{N}_t$ の変化をランダムガウス雑音ベクトル $\mathbf{W}_t$ 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 $\mathbf{N}_t$ 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、図 1 に示す拘束条件パラメータ 1 3 8 を用いて、雑音の特徴量ベクトル $\mathbf{N}_t$ 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 1 3 8 の詳細については、後述する。

【0 0 5 3】

10

〈雑音抑圧部 1 1 4 の構成〉

図 4 に、雑音抑圧部 1 1 4 (図 1 参照) の構成をブロック図で示す。図 4 を参照して、雑音抑圧部 1 1 4 は、観測信号の特徴量 $\mathbf{X}_t$ 1 2 4、GMM 1 3 0、及び拘束条件パラメータ 1 3 8 を用いて、雑音の特徴量ベクトル $\mathbf{N}_t$ 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 2 0 0 を含む。雑音抑圧部 1 1 4 はさらに、雑音確率分布推定部 2 0 0 により出力された雑音確率分布の推定パラメータ 2 0 6 及び GMM 1 3 0 を基に、観測信号の特徴量 $\mathbf{X}_t$ 1 2 4 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 2 0 8 (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 2 0 2 と、雑音確率分布の推定パラメータ 2 0 6、観測信号分布のパラメータ 2 0 8、及び GMM 1 3 0 を基に、推定クリーン音声の特徴量 1 2 6 を生成するためのクリーン音声推定部 2 0 4 とを含む。

20

【0 0 5 4】

雑音確率分布推定部 2 0 0 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 2 0 6 を出力する機能を持つ。ここに、観測信号の特徴量 $\mathbf{X}_0, \dots, \mathbf{X}_t$ からなるベクトルの系列を系列 $\mathbf{X}_{0:t} = \{\mathbf{X}_0, \dots, \mathbf{X}_t\}$ とし、雑音の特徴量ベクトル $\mathbf{N}_0, \dots, \mathbf{N}_t$ からなるベクトルの系列を系列 $\mathbf{N}_{0:t} = \{\mathbf{N}_0, \dots, \mathbf{N}_t\}$ とする。観測信号ベクトルの系列 $\mathbf{X}_{0:t}$ が与えられた時の系列 $\mathbf{N}_{0:t}$ の事後確率分布 $p(\mathbf{N}_{0:t} | \mathbf{X}_{0:t})$ 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。

30

【0 0 5 5】

【数 5】

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

【0 0 5 6】

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 2 0 0 は、観測信号の特徴量 $\mathbf{X}_t$ 1 2 4 と、GMM 1 3 0 と、状態空間モデル 1 6 0 と、雑音の状態遷移に関する上記の拘束条件パラメータ 1 3 8 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 2 0 0 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。

40

【0 0 5 7】

観測信号分布推定部 2 0 2 は、観測信号分布のパラメータ 2 0 8 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

50

分布のパラメータ 208 の算出には、例えば、V T S (Vector Taylor Series) 法と呼ばれる HMM 合成法が用いられる。

【0058】

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、最小 2 乗誤差 (Minimum Mean Square Error: MMSE) 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。

【0059】

〈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。

【0060】

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 $N_t^{(j)}$ 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 $p(N_{0:t} | X_{0:t})$ 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。

【0061】

【数 6】

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

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

【0062】

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 $N_{0:t}^{(j)}$ を出力する確率分布を $q(N_{0:t}^{(j)} | X_{0:t})$ 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。

【0063】

10

20

30

40

【数 7】

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

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。

【0064】

【数 8】

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

10

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。

【0065】

【数 9】

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

20

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。

【0066】

【数 10】

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

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。

30

【0067】

【数 11】

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

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。

【0068】

【数 12】

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

40

【0069】

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 3 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

50

雑音の特徴量ベクトル $N_t^{(j)}$ の平均ベクトル $\hat{N}_t^{(j)}$ と共分散行列 $\Sigma_{N_t^{(j)}}$ とを含む。以下、確率密度関数 $p(N_{0:t} | X_{0:t})$ の平均ベクトル $\hat{N}_t^{(j)}$ と共分散行列 $\Sigma_{N_t^{(j)}}$ 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。

【0070】

〈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 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 $N_t^{(j)}$ ($1 \leq j \leq J$) の変化に対し、次の式(11)に示す状態方程式を導入する。

【0071】

【数13】

$$N_{t+1}^{(j)} = (1 - \alpha)N_t^{(j)} + \alpha\hat{N}_t + \alpha\beta\left(\mu_{N_t}^{(j)} - N_t^{(j)}\right) + W_t^{(j)} \quad (11)$$

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。

【0072】

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。

【0073】

第1の拘束条件において、ベクトル $\hat{N}_t$ 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 $N_t^{(1)}, \dots, N_t^{(J)}$ の加重平均であり、次の式(12)により与えられる。

【0074】

【数14】

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

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。

【0075】

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。

【0076】

【数15】

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

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 α 及び第2の拘束条件に対するスケール係数 β 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。

【0077】

10

20

30

40

50

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。

【0078】

〈雑音確率分布推定部 200 の構成〉

図 5 に、雑音確率分布推定部 200 の構成をブロック図で示す。図 5 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。

10

【0079】

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 ($t \geq 1$) 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。

【0080】

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 ($t \geq 1$) 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。

20

【0081】

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。

【0082】

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。

30

【0083】

【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 124 が雑音 121 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、 $0 \leq t \leq 9$ の 10 フレーム分の区間が雑音 121 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。

40

【0084】

【数 1 7】

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

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

そして雑音初期分布推定部 2 2 2 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 $\mathbf{N}_0^{(j)}$ と共分散行列 $\Sigma_{\mathbf{N}_0}^{(j)}$ とを、それぞれ、式（17）及び式（18）のように設定する。

【0085】

【数 1 8】

$$\mathbf{N}_0^{(j)} \sim \mathcal{N}(\mu_N, \Sigma_N) \quad (17)$$

$$\Sigma_{\mathbf{N}_0}^{(j)} = \Sigma_N \quad (18)$$

すなわち、雑音初期分布推定部 2 2 2 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 $\mathbf{N}_0^{(j)}$ を、初期分布 $p(\mathbf{N}_0)$ からのサンプリングによって生成し、共分散行列 $\Sigma_{\mathbf{N}_0}^{(j)}$ を、初期分布 $p(\mathbf{N}_0)$ の共分散行列 Σ_N に設定する。雑音初期分布推定部 2 2 2 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。

【0086】

逐次計算部 2 2 4 は、GMM 1 3 0 から出力パラメータ 2 4 0 をサンプリングするための GMM サンプリング部 2 2 6 を含む。逐次計算部 2 2 4 はさらに、観測信号の特徴量 1 2 4 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 2 3 0 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 2 3 2 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 2 3 4 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 2 0 6 を生成するための推定パラメータ生成部 2 3 6 とを含む。

【0087】

GMM サンプリング部 2 2 6 は、パーティクル j （ $1 \leq j \leq J$ ）ごとに、GMM 1 3 0（図 2 参照）内の混合分布 1 4 0 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 2 2 6 はさらに、出力パラメータベクトル $\mathbf{S}_{k_t}^{(j)}|_{k_t}^{(j)}$ を、要素分布 $k_t^{(j)}$ からサンプリングして、更新部 2 3 0 に与える。ここに GMM 1 3 0 における要素分布 1 4 8 A, ..., 1 4 8 K 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、GMM 1 3 0 から次の式（19）に示すサンプリングによって得られる。

【0088】

【数 1 9】

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

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 GMM 1 3 0 の出力パラメータベクトル $\mathbf{S}_{k_t}^{(j)}|_{k_t}^{(j)}$ 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。

【0089】

【数20】

$$\mathbf{s}_{k_t^{(j)},t}^{(j)} \sim \mathcal{N}(\mu_{\mathbf{s},k_t^{(j)}}, \Sigma_{\mathbf{s},k_t^{(j)}}) \quad (20)$$

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。

【0090】

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3)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 $\mathbf{s}_{k_t^{(j)},t}^{(j)}$ 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。

10

【0091】

図6に、更新部230の構成をブロック図で示す。図6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 $\hat{\mathbf{N}}_{t-1}$ 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。

20

【0092】

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 $\hat{\mathbf{N}}_{t-1}^{(j)}$ との差分 $\mu_{N_{t-1}}^{(j)} - \hat{\mathbf{N}}_{t-1}^{(j)}$ を算出する。

30

【0093】

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 124と、j番目のパーティクルにおけるGMM130(図2参照)の出力パラメータベクトル $\mathbf{s}_{k_t^{(j)},t}^{(j)}$ 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 $\hat{\mathbf{N}}_{t-1}$ と、差分 $\mu_{N_{t-1}}^{(j)} - \hat{\mathbf{N}}_{t-1}^{(j)}$ とを用いる。

40

【0094】

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。

【0095】

【数 2 1】

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

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

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

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

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

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

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。

【0096】

再び図5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 $\mathbf{X}_{t124}$ と、第 t フレームの各パーティクルにおける GMM130 の出力パラメータベクトル $\mathbf{S}_{k_t^{(j)}, t}^{(j)}$ 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 $\hat{\mathbf{N}}_t^{(j)}$ 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。

【0097】

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る。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 $\mathbf{X}_{t124}$ の尤度に対応しているからである。

【0098】

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7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7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6に

10

20

30

40

50

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。

【0099】

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。

【0100】

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 ν 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 ν 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。

10

【0101】

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 ν を算出する機能を持つ。

【0102】

20

【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 ν 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。

【0103】

〈コンピュータによる実現〉

以下の説明からも明らかなように、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 8 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。

30

【0104】

図 8 を参照して、雑音抑圧処理が開始されると、ステップ 302 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 (15) 及び式 (16) 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 (17) 及び式 (18) 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。

40

【0105】

ステップ 304 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 306 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 $\mu_{N_t}^{(j)}$ 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 206 を生成する。

50

このステップでの処理については、図9を用いて後述する。

【0106】

ステップ308では、観測信号分布のパラメータ208を推定する。すなわち、ステップ306で定めた各パーティクルの雑音のパラメータ $\wedge N_t^{(j)}$ 、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 124の確率分布を推定する。さらに、GMM130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 124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。

【0107】

ステップ310では、MMSE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126を算出する。すなわちまず、ステップ306及びステップ308の処理で得られたパラメータを用いて、MMSE推定法によって、MMSE推定値ベクトル $\wedge S_t$ を算出し、推定クリーン音声の特徴量126（図1参照）として出力する。

【0108】

【数23】

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

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

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

この式において、 $P(k|\mathbf{X}_t, (j))$ は、 j 番目のパーティクルにおける、GMM130内の要素分布 k に対する混合重みを表す。混合重み $P(k|\mathbf{X}_t, (j))$ は、次の数式により算出される。

【0109】

【数24】

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

ステップ312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304に戻る。

【0110】

図9に、ステップ306（図8参照）において行なわれる雑音確率分布の推定パラメータ206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9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320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121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 $\wedge N_{t-1}$ を式(12)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 $\wedge N_{t-1}^{(j)}$ との差分 $\mu_{N_{t-1}}^{(j)} - \wedge N_{t-1}^{(j)}$ を算出する。

【0111】

ステップ322では、式(21)～式(26)に示す拡張カルマンフィルタを用いて、

10

20

30

40

50

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。

【0112】

ステップ324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(10)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326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328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。

【0113】

10

図10にステップ328(図9参照)における処理の詳細をフローチャートで示す。図10を参照して、ステップ328における処理が開始されると、ステップ340において、図9に示すステップ320と同様に、加重平均ベクトル $\hat{N}_{t-1}$ を、式(12)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。

【0114】

続くステップ342では、ステップ326(図9参照)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(21)～式(26)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322(図9参照)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344では、ステップ342で準備したパーティクルに対する重み $w_t^{*(j)}$ を、図9に示すステップ324の処理と同様の処理で算出し正規化する。

20

【0115】

ステップ346では、ステップ324の処理で算出された重み $w_t^{(j)}$ と、ステップ344で算出された重み $w_t^{*(j)}$ との比較により、ステップ342で準備されたパーティクルの許容確率 v を定める。ステップ348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350では、ステップ348で発生した乱数 u の値と、ステップ346で定めた許容確率 v 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352へ進む。さもなければステップ354に進む。ステップ352では、ステップ342で準備されたパーティクルを許容する。すなわち、ステップ326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354では、ステップ342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。

30

【0116】

[動作]

40

本実施の形態に係る音声認識システム100は以下のように動作する。

【0117】

まず、図5に示す雑音確率分布推定部200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206を生成する動作について説明する。図1に示す計測部112が、音源102から雑音重畳音声122を受け、観測信号の特徴量 X_t 124を抽出する。抽出された特徴量 X_t 124は、雑音抑圧部114の図5に示す雑音確率分布推定部200に与えられる。図5を参照して、雑音確率分布推定部200のフレーム選択部220は、特徴量 X_t 124のうち最初の10フレーム分を、雑音初期分布推定部222に与える。雑音初期分布推定部222は、上記の式(14)～式(16)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

50

7) 及び式 (18) 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 $\Sigma_{N_0}^{(j)}$ が決定される。雑音確率分布推定部 200 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 206 として出力する。

【0118】

次に、雑音確率分布推定部 200 の逐次推定部 224 が、第 t フレーム ($t \geq 1$) における雑音確率分布の推定パラメータ 206 を生成する動作について説明する。図 5 を参照して、次のフレームの処理の開始要求 210 に応答して、フレーム選択部 220 は、GMM サンプリング部 226 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 124 を更新部 230 に与える。

10

【0119】

GMM サンプリング部 226 は、GMM 130 から、出力パラメータベクトル $S_{k_t}^{(j)}$ (j, t 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 226 が、図 2 に示す GMM 130 内の混合正規分布 140 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 150 がサンプリングされたものとする。GMM サンプリング部 226 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 $S_{k_t}^{(j)}$ をサンプリングする。GMM サンプリング部 226 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 $S_{k_t}^{(j)}$ をそれぞれ、以上の手順でサンプリングし、図 5 に示す更新部 230 に与える。

20

【0120】

図 11 に、逐次計算部 224 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 11 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 11 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。

【0121】

図 11 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 420 が近似的に表現されているものとする。更新部 230 は、以下のようにして、状態空間 420 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 430 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。

30

【0122】

まず、図 6 に示す更新部 230 の拡張カルマンフィルタ部 258 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。

【0123】

図 6 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 $\hat{N}_{t-1}$ をに基づき、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。

40

【0124】

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

50

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 $\wedge N_{t-1}^{(j)}$ との差分 $\mu_{N_{t-1}}^{(j)} - \wedge N_{t-1}^{(j)}$ 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。

【0125】

図 12 に、Polyak Average及びフィードバックの概念を模式的に示す。図 12 (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 $N_{t-4}^{(j)}$, ..., $N_{t+1}^{(j)}$ との関係を表している。なお、図 12 (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 12 (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 12 (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。

【0126】

図 12 (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 $\mu_{N_t}^{(j)}$ との間には、差分 $\mu_{N_t}^{(j)} - N_t^{(j)}$ が生じる。図 12 (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 $\mu_{N_t}^{(j)}$ との間には、差分 $\mu_{N_t}^{(j)} - N_t^{(j)}$ が生じる。図 12 (A) における雑音の特徴量ベクトル $N_{t-4}^{(j)}$, ..., $N_t^{(j)}$ の変動に比べて、図 12 (B) における雑音の特徴量ベクトル $N_{t-4}^{(j)}$, ..., $N_t^{(j)}$ の変動は大きい。すなわち図 12 (A) における雑音の特徴量ベクトル $N_{t-4}^{(j)}$, ..., $N_t^{(j)}$ 同士の差異は、図 12 (B) における当該それらの差異より小さい。

【0127】

Polyak Averageベクトル $\mu_{N_t}^{(j)}$ は、 $N_{t-4}^{(j)}$, ..., $N_t^{(j)}$ 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、 $N_{t-4}^{(j)}$, ..., $N_t^{(j)}$ の最小から最大までの範囲である。したがって、図 12 (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 12 (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。

【0128】

拡張カルマンフィルタ部 258 (図 6 参照) は、加重平均ベクトル $\wedge N_{t-1}$ と、差分ベクトル $\mu_{N_{t-1}}^{(j)} - N_{t-1}^{(j)}$ と、拘束条件パラメータ 138 により定められる忘却係数 α 及びスケリング係数 β と、観測信号の特徴量 X_{t-1} と、出力パラメータ 240 とを基に、式 (21) ~ 式 (26) 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。

【0129】

この更新において、式 (21) 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、 $\wedge N_{t-1}^{(j)}$ 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。

【0130】

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11に示す状態空間420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430が表現される。

【0131】

これに応答して、重み算出部232が、状態空間430内の各パーティクルに対する重み $w_t^{(j)}$ を、式(22)及び式(23)によって算出する。再サンプリング部234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234はまず、状態空間430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430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440を表現するパーティクルがそれぞれ形成される。

【0132】

再サンプリング部234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236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7に示す再更新部262は、状態空間440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420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264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266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 v を算出する。パラメータ選択部270は、許容確率 v と、乱数発生部268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 v 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。

【0133】

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 $\Sigma_{N_t}^{(j)}$ 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 $\hat{N}_t^{(j)}$ 及び共分散行列 $\Sigma_{N_t}^{(j)}$ と、各パーティクルに対する重み $w_t^{(j)}$ 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4に示す観測信号分布推定部202に与える。

【0134】

図4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。クリーン音声推定部204は、MMS E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MMS E推定値ベクトル $\hat{S}_t^{(j)}$ を算出する。さらに、MMS E推定値ベクトル $\hat{S}_t^{(j)}$ と重み $w_t^{(j)}$ 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。

【0135】

10

20

30

40

50

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 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。

【0136】

〔第 2 の実施の形態〕

式 (1) に示す観測方程式と式 (11) に示す状態方程式とに基づく状態空間モデル 160 (図 3 参照) は、過去の第 t フレームでの j 番目のパーティクルにおける雑音の特徴量ベクトル $N_t^{(j)}$ の時間変化状況に応じて未来のパラメータを予測するモデルである。第 1 の実施の形態においては、拘束条件パラメータ 138 (図 1 参照) を時間固定のものとして扱っており、状態方程式 (11) を時間変化に対して定常的なものとして、未来のパラメータを予測した。しかし、非定常雑音の時間変化の様相は時々刻々と変化するため、拘束条件パラメータもまた、時間の経過に伴い変化することが望ましい。そこで、以下に説明する第 2 の実施の形態では、それぞれ異なるパラメータの設定された複数の状態空間モデルの中から、未来予測に最適な状態空間モデルをフレームごとに選ぶ。そして各フレームにおいて選択した状態空間モデルを適用し、パーティクルの更新を行なう。

【0137】

〔構成〕

図 13 に、本実施の形態に係る音声認識システムにおいて、図 1 に示す前処理部 104 に代えて用いられる前処理部 500 の構成を示す。図 13 を参照して、前処理部 500 は、図 1 に示す前処理部 104 のものと同一の計測部 112 と、GMM 130 を用い、複数の状態空間モデルを選択的に適用して計測部 112 の出力する観測信号の特徴量 124 に含まれる雑音の成分を抑圧するための雑音抑圧部 510 とを含む。雑音抑圧部 510 は、図 1 に示す前処理部 104 と異なり、拘束条件パラメータとして所与のものを使用しない。その代り雑音抑圧部 510 は、後述のとおり、フレームの移行時に逐次的に拘束条件パラメータを決定し、当該拘束条件パラメータを使用して雑音の抑圧を行なう。

【0138】

図 14 に、本実施の形態に係る雑音抑圧部 510 の構成をブロック図で示す。図 14 を参照して、雑音抑圧部 510 は、第 1 の実施の形態に係る雑音抑圧部 114 のものとそれぞれ同一の雑音確率分布推定部 200、観測信号分布推定部 202、及びクリーン音声推定部 204 を含む。雑音抑圧部 510 はさらに、クリーン音声推定部 204 により発行される次のフレームへの移行に関する要求 210 を受けるように接続され、フレームが移行する度にそのフレームに対応する拘束条件パラメータ 522 を決定し、雑音確率分布推定部 200 に与えるための拘束条件パラメータ決定部 520 を含む。

【0139】

図 15 に、拘束条件パラメータ決定部 520 の構成をブロック図で示す。拘束条件パラメータ決定部 520 は、要求 210 を受ける度に忘却係数 α 、スケーリング係数 β 、及びフレーム数 T の値をそれぞれ決定するための α 決定部 530、 β 決定部 532、及び T 決定部 534 と、 α 決定部 530 により決定された忘却係数 α 、 β 決定部 532 により決定されたスケーリング係数 β 、及び T 決定部 534 により決定されたフレーム数 T を、拘束条件パラメータ 522 として出力し雑音確率分布推定部 200 (図 14 参照) に与えるための拘束条件パラメータ出力部 536 とを含む。

【0140】

α 決定部 530、 β 決定部 532、及び T 決定部 534 はそれぞれ、忘却係数 α 、スケーリング係数 β 、及びフレーム数 T のとり得る値を一覧にした α 状態-パラメータ表 540、 β 状態-パラメータ表 544、及び T 状態-パラメータ表 548 と、要求 210 を受ける度に α 状態-パラメータ表 540、 β 状態-パラメータ表 544、及び T 状態-パラメータ表 548 を用いてそれぞれ、忘却係数 α 、スケーリング係数 β 、及びフレーム数 T の値をエルゴディックモデルによる確率分布にしたがって選択し出力するためのエルゴディックモデル部 542、546、及び 550 とを含む。

【0141】

α 状態－パラメータ表 540、 β 状態－パラメータ表 544、及び T 状態－パラメータ表 548 の例を、それぞれ図 16 (A)、(B)、及び (C) に示す。図 16 (A) を参照して、 α 状態－パラメータ表 540 の各行は、忘却係数 α の状態に対応している。また、 α 状態－パラメータ表 540 には、状態番号の欄と、忘却係数 α の欄とが設けられている。すなわち、 α 状態－パラメータ表 540 は、忘却係数 α が 4 つの状態を有し、一つの状態が一つの状態番号と忘却係数 α のとり得る一つのパラメータ値との組により構成されていることを表す。なお状態番号は、対応のパラメータ値が小さな状態ほど若い番号となるように予め付与される。

【0142】

10

図 16 (B) に示す β 状態－パラメータ表 544、及び図 16 (C) に示す T 状態－パラメータ表 548 もまた、 α 状態－パラメータ表 540 と類似の構成である。ただし、列挙されているパラメータ値はそれぞれ、スケーリング係数 β に対応するもの、及びフレーム数 T に対応するものになっている。

【0143】

図 17 に、忘却係数 α の状態遷移過程を模式的に示す。図 17 を参照して、忘却係数 α の状態遷移過程は、エルゴディックモデルによってモデル化されている。忘却係数 α を出力しうる状態番号 1～4 の状態がある。これらの状態はそれぞれ、 α 状態－パラメータ表 540 により、忘却係数 α のとり得るパラメータ値に対応している。フレームが移行する度に、これらの状態のうち一つが忘却係数 α のパラメータ値を出力する。フレーム移行前の状態からある確率で他の状態に状態が遷移し、遷移先の状態がパラメータ α を出力する。この状態遷移を矢印で示す。本実施の形態では、自己ループを含めすべての状態の間の遷移が許容される。

20

【0144】

エルゴディックモデル部 542 は、忘却係数 α の状態を、フレーム移行の度に状態番号 1～4 のいずれかの状態へ遷移させることによって、各フレームにおける忘却係数 α の値を決定する。

【0145】

エルゴディックモデル部 542、546、及び 550 は、対応のパラメータが異なることを除けば、いずれも同一の機能を有し独立に動作する。エルゴディックモデル部 542、546、及び 550 を代表して、 α 決定部 530 のエルゴディックモデル部 542 の構成を図 18 にブロック図で示す。図 18 を参照して、エルゴディックモデル部 542 は、 α 状態－パラメータ表 540 における各状態の状態番号と所定のパラメータ γ ($1 \geq \gamma > 0$) 570 とをもとに、各状態間の遷移に対し状態遷移確率を割当てするための状態遷移確率割当部 572 と、状態遷移確率割当部 572 が割当てた状態遷移確率を一覧にした状態遷移確率表 574 と、フレーム移行の要求 210 を受ける度に、状態遷移確率表 574 を用いてフレーム移行前の状態から各状態への状態遷移確率を決定するための状態遷移確率決定部 562 と、状態遷移確率決定部 562 が決定したフレーム移行前の状態から各状態への状態遷移確率をもとにフレーム移行後の状態を選択し、その状態番号を出力するための状態遷移部 564 と、状態遷移部 564 により出力された状態番号を記憶するための状態番号記憶部 560 と、状態遷移部 564 により出力された状態番号に対応するパラメータ値を α 状態－パラメータ表 540 から読出して、拘束条件パラメータ出力部 536 に出力するためのパラメータ読出部 566 とを含む。

30

40

【0146】

状態遷移確率割当部 572 は、 α 状態－パラメータ表 540 が準備された時点で、 α 状態－パラメータ表 540 の状態のうちの任意の二つの組合せに対し、当該二つの状態の状態番号の差の絶対値の単調減少関数として、当該二つの状態間での状態遷移確率を決定する。図 20 に、状態番号とパラメータの値との関係を概略的に示す。図 20 を参照して、第 t フレーム及び第 $t+1$ フレームにおける状態の状態番号をそれぞれ、 $m_t^{(j)}$ 、 $n_{t+1}^{(j)}$ とする。状態遷移確率割当部 572 は、 α 状態－パラメータ表 540 (図 16 (A) 参

50

照)の行ごとに、状態番号 $m_t^{(j)}$ からその行の状態番号を $n_{t+1}^{(j)}$ とした場合の状態番号の変化量 $d_{mn,t}^{(j)} = |n_{t+1}^{(j)} - m_t^{(j)}|$ を算出する。そしてパラメータ $\gamma 570$ を用い、次の数式によって、自己ループへの状態遷移確率を1とした場合の状態 $m_t^{(j)}$ から状態 $n_{t+1}^{(j)}$ への状態遷移確率の相対量(以下、この量を「(状態遷移確率の)割当量」と呼ぶ。) $a_{mn,t}^{(j)}$ を決定する。

【0147】

【数25】

$$a_{mn,t}^{(j)} = \gamma^{d_{mn,t}^{(j)}}$$

10

そして、 $\sum_n a_{mn,t}^{(j)} = 1$ となるよう割当量 $a_{mn,t}^{(j)}$ を正規化し、正規化後の割当量 $a_{mn,t}^{(j)}$ を、 $m_t^{(j)}$ と $n_{t+1}^{(j)}$ との間の状態遷移確率とし、状態遷移確率表574に格納する。

【0148】

図19に状態遷移確率表574の一例を示す。図19を参照して、この状態遷移確率表574は、状態数4のエルゴディックモデルにおいて、パラメータ $\gamma 570$ の値を0.5に設定した場合の各状態間の状態遷移確率を示すものである。状態遷移確率表574の各行は遷移前の状態に対応しており、各欄は遷移先の状態に対応している。状態遷移確率表574のどの行においても、その行の状態番号と同じ状態番号に対応する欄に格納される状態遷移確率が、状態遷移確率の最大値をとっている。すなわち、他の状態に遷移するよりも自己ループする確率がより高く設定される。

20

【0149】

再び図18を参照して、状態遷移確率決定部562は、要求210が与えられる度に、状態番号記憶部560に記憶されている状態番号を参照し、さらに、状態遷移確率表574の当該状態番号の行を読出して、状態遷移部564に与える。

【0150】

状態遷移部564は、状態遷移確率表574の、状態遷移確率決定部562より与えられた行における各欄に表される状態遷移確率にしたがったモンテカルロサンプリングにより、遷移先の状態を選択し、選択した状態の状態番号を状態番号記憶部560及びパラメータ読出部566に与える機能を持つ。パラメータ読出部566は、状態遷移部564により出力された状態番号に対応する状態のパラメータ値を、 α 状態-パラメータ表540より読出して出力する機能を持つ。状態番号記憶部560は、状態遷移部564が出力した状態番号を記憶する。この番号は、次のフレーム移行時に、状態遷移確率決定部562により参照される。

30

【0151】

〈コンピュータによる実現〉

以下の説明からも明らかなように、図13に示す前処理部500は、図1に示す前処理部104と同様に、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21に、前処理部500に含まれる雑音抑圧部510(図14参照)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。

40

【0152】

図21を参照して、本実施の形態に係る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は、図8に示す第1の実施の形態に係る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と類似する。ただし、本実施の形態に係る雑音抑圧処理においては、ステップ304において次のフレームへ処理対象を移行した後、ステップ600に進む。ステップ600では、移行後のフレームで雑音の確率分布の推定に使用する拘束条件パラメータ522を決定し、ステップ306に進む。なお、ステップ306では、ステップ600で決定された拘束条件パラメータ522を使用して処理を行なうことになる。

【0153】

50

ステップ600において、拘束条件パラメータ522を決定する処理は、忘却係数 α 、スケーリング係数 β 、及びフレーム数 T に対し、同様の処理がそれぞれ独立に実行される。図22に、これら独立に実行される処理を代表して、忘却係数 α を決定する処理の制御構造をフローチャートで示す。図22を参照して、忘却係数 α を決定する処理が開始されると、ステップ620において、予め記憶しておいた移行前のフレームにおいて忘却係数 α を出力した状態の状態番号 $m_{t-1}^{(j)}$ を取得する。ステップ622では、状態遷移確率表574（図19参照）の、ステップ620で取得した状態番号に対応する行を参照し、状態番号 $m_{t-1}^{(j)}$ の状態と各状態との間の状態遷移確率の割当を読出す。

【0154】

ステップ624では、ステップ622で読出した状態遷移確率にしたがい、状態番号 $n_t^{(j)}$ に関するモンテカルロサンプリングを実行し、サンプリングにより選択された状態番号 $n_t^{(j)}$ の状態を遷移先の状態として選択する。ステップ626では、ステップ624で選択された状態番号 $n_t^{(j)}$ の状態から忘却係数 α の値を出力し、処理を終了する。

【0155】

[動作]

本実施の形態に係る雑音抑圧部510は、フレーム移行の要求がある度に拘束条件パラメータ決定部520が拘束条件パラメータ522を決定し更新する。また、雑音確率分布推定部200は、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際各パーティクルにおける式(13)に示すPolyak Averageベクトル $\mu_{Nt}^{(j)}$ を算出する際、並びに式(21)～式(26)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際に、更新された拘束条件パラメータ522を使用して処理を実行する。これらの点を除き、雑音抑圧部510は、第1の実施の形態に係る雑音抑圧部114と同一の動作を実行する。拘束条件パラメータ決定部520の図15に示すエルゴディックモデル部542、546、及び550に要求210が与えられると、エルゴディックモデル部542、546、及び550はそれぞれ独立に動作し、忘却係数 α 、スケーリング係数 β 、及びフレーム数 T の値をそれぞれ決定する。

【0156】

以下、エルゴディックモデル部542、546、及び550を代表して、 α 決定部530のエルゴディックモデル部542の動作を説明する。まず、図18に示すエルゴディックモデル部542の状態遷移確率割当部572が状態遷移確率の割当を予め決定し、状態遷移確率表574（図19参照）を生成する動作について説明する。

【0157】

図18を参照して、エルゴディックモデル部542に α 状態パラメータ表540とパラメータ γ 570とが与えられると、状態遷移確率割当部572は、 α 状態パラメータ表540（図16(A)参照）の一つ目の行を処理対象に定める。そして、当該遷移前状態の状態番号と、 α 状態パラメータ表540の各行における状態番号とを基に、遷移前状態と各状態との間の状態番号の変化量 $d_{mn,t}^{(j)}$ を算出する。例えば、状態遷移時に自己ループする場合、状態番号の変化量 $d_{mn,t}^{(j)} = 0$ となる。

【0158】

各遷移先状態について変化量 $d_{mn,t}^{(j)}$ を算出すると、状態遷移確率割当部572は次に、パラメータ γ 570を状態番号の変化量 $d_{mn,t}^{(j)}$ でべき乗して割当量 $a_{mn,t}^{(j)}$ を算出する。パラメータ γ 570の値が $1 > \gamma > 0$ に設定されていれば、自己ループにおいて状態番号の変化量 $d_{mn,t}^{(j)}$ は0となり、割当量 $a_{mn,t}^{(j)}$ は、最大値 $\gamma^0 = 1$ となる。自己以外への状態遷移では、状態番号の変化量 $d_{mn,t}^{(j)}$ の値が大きくなるほど小さくなる。パラメータ γ 570の値が1に設定されていれば、割当量 $a_{mn,t}^{(j)}$ は、状態番号の変化量 $d_{mn,t}^{(j)}$ の値に関係なく $a_{mn,t}^{(j)} = 1$ となる。

【0159】

状態遷移確率割当部572はさらに、割当量 $a_{mn,t}^{(j)}$ の各々を、 $\sum_n a_{mn,t}^{(j)} = 1$ となるよう正規化する。このようにして算出された正規化後の割当量 $a_{mn,t}^{(j)}$ の各々を、状態遷移確率表574の遷移先の1行目の対応する欄に格納する。そして、 α 状態パラ

10

20

30

40

50

メータ表 540 (図 16 (A) 参照) の次の行を処理対象に定める。

【0160】

パラメータ γ_{570} が $1 > \gamma > 0$ に設定されていれば、各状態への状態遷移確率は、自己ループにおいて最大値となり、自己ループ以外の状態遷移においては、 $d_{mn,t}^{(j)}$ の値が大きくなるほど低い値となる。 α 状態-パラメータ表 540 においては、出力するパラメータ値が小さな状態ほど若い状態番号が付与されている。状態番号の変化量 $d_{mn,t}^{(j)}$ が小さければ、その分フレーム移行前後での忘却係数 α の値の変動は小さく、状態番号の変化量 $d_{mn,t}^{(j)}$ が大きければその分忘却係数 α の値の変動は大きくなる。したがって、パラメータ γ_{570} を $1 > \gamma > 0$ に設定すると、各状態への状態遷移確率は、忘却係数 α の値の変動が小さい状態に遷移する確率が高くなるように割当てられる。これに対し、パラメータ γ_{570} を 1 に設定すると、割当量 $a_{mn,t}^{(j)}$ は、状態番号の変化量 $d_{mn,t}^{(j)}$ の値に関係なく $a_{mn,t}^{(j)} = 1$ となる。そのため各状態への状態遷移確率は全て同じ確率になる。

10

【0161】

この動作を繰返すことにより α 状態-パラメータ表 540 の全ての行について処理が終了すると、状態遷移確率割当部 572 は動作を完了する。以上の一連の動作は、図 15 に示すエルゴディックモデル部 542、546、及び 550 においてそれぞれ独立に実行される。

【0162】

次に、エルゴディックモデル部 542 がフレーム移行の要求 210 を受けてフレーム移行後のフレームにおける忘却係数 α の値を決定する動作を説明する。図 14 に示すクリーン音声推定部 204 が次のフレームへの移行の要求 210 を発行すると、要求 210 は、雑音確率分布推定部 200 に与えられるばかりでなく、拘束条件パラメータ決定部 520 の図 18 に示す状態遷移確率決定部 562 にも与えられる。状態遷移確率決定部 562 に要求 210 が与えられる時点では既に、状態番号記憶部 560 は、フレーム移行前の状態について、その状態番号 $m_t^{(j)}$ を記憶している。状態遷移確率決定部 562 は、要求 210 に応答して、状態番号 $m_t^{(j)}$ を状態番号記憶部 560 から読出す。そして、状態遷移確率表 574 (図 19 参照) の状態番号 $m_t^{(j)}$ の行を読出し、状態遷移部 564 に与える。

20

【0163】

状態遷移部 564 は、与えられた行の各欄に格納されている、各状態への状態遷移確率にしたがったモンテカルロサンプリングにより遷移先の状態を選択する。そして選択した状態の状態番号を状態番号記憶部 560 及びパラメータ読出部 566 に与える。状態遷移部 564 が状態を選択すると、パラメータ読出部 566 は、状態遷移部 564 により選択された状態のパラメータ値を、 α 状態-パラメータ表 540 より読出して出力する。状態番号記憶部 560 は、状態遷移部 564 が出力した状態番号を記憶する。

30

【0164】

以上の処理と同様の処理を図 15 に示す β 決定部 532 及び T 決定部 534 もまた、要求 210 に応答して実行する。拘束条件パラメータ出力部 536 は、 α 決定部 530 から忘却係数 α を、 β 決定部 532 からスケール係数 β を、T 決定部からフレーム数 T を受けると、それらをまとめて拘束条件パラメータ 522 として、図 14 に示す雑音確率分布推定部 200 に与える。雑音確率分布推定部 200 は、第 1 の実施の形態と同様の動作で雑音確率分布の推定パラメータ 206 を逐次的に生成し、観測信号の特徴量ベクトル X_t 124 とともに、フレームごとに、図 14 に示す観測信号分布推定部 202 に与える。

40

【0165】

雑音確率分布の推定パラメータ 206 を生成する際の処理において、状態空間モデル 160 を形成する状態方程式 (11) における拘束条件パラメータは、処理対象のフレームごとに再設定される。これにより、フレームごとに状態空間モデルが変動するため、時間の推移とともに変動する非定常雑音をより適切に推定することが可能になる。

【0166】

[実験]

50

本発明の第1及び第2の実施の形態に係る音声認識システムによる効果を確認するために、観測信号からの雑音の推定実験と、観測信号の認識実験とを行なった。以下、実験方法及び結果について説明する。

【0167】

本実験では、日本語の雑音下音声認識評価用共通データベースに収録されたクリーン音声1001文のデータに、雑音を人工的に加算して、雑音重畳音声122（図1参照）の試料を生成した。また加算する雑音には、それぞれ実環境で収録された工場雑音と道路工事雑音とを使用した。本実験では、雑音を加算していない試料と、クリーン音声に雑音をそれぞれ20dB、15dB、10dB、5dB、及び0dBのSNR（Signal-to-Noise Ratio）で加算した試料とを用意した。用意した各試料を23次対数メルフィルタバンク処理し、得られた23次対数メルスペクトルの各成分を要素とするベクトルをそれぞれ生成し、認識対象の特徴量ベクトルとした。

10

【0168】

認識実験では、比較のために、上記の各試料から、本実施の形態に係る雑音抑圧処理の方法を含む次の各方法で、探索に用いる特徴量を生成した。すなわち、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い、得られた雑音確率分布のパラメータを用いてMMSE推定法により特徴量を生成した。また、雑音確率分布の逐次推定を行なうにあたり、パーティクルフィルタとして、次の3種のものを用いた。すなわち、状態遷移過程にランダムウォーク過程を仮定し上記の状態方程式（2）を導入したパーティクルフィルタ（PF1）と、状態遷移過程に時間固定の拘束条件パラメータ138に基づく拘束条件を設定し上記の状態方程式（11）を導入したパーティクルフィルタ（PF2）と、時間変動する拘束条件パラメータ522に基づく拘束条件を設定し上記の状態方程式（11）を導入したパーティクルフィルタ（PF3）とである。

20

【0169】

第1の実施の形態に係るパーティクルフィルタは、PF2のパーティクルフィルタであり、第2の実施の形態に係るパーティクルフィルタはPF3である。本認識実験ではさらに、比較のために外乱抑圧処理を行なわない観測信号の特徴量（HTK）と、非特許文献2に記載の方式で雑音抑圧処理を施した観測信号の特徴量（ETSI）と、パーティクルフィルタを用いた雑音確率分布の逐次推定を行なわずにMMSE推定法により特徴量を生成する方式で雑音抑圧処理を施した観測信号の特徴量（MMSE）と、逐次EMアルゴリズムによる雑音の逐次推定の結果を用いて雑音抑圧処理を施した観測信号の特徴量（SEM: Sequential Expectation Maximization）とを用いた探索も行なった。

30

【0170】

上記のいずれのパーティクルフィルタを用いて処理を行なう場合にも、GMM130（図1参照）には、混合分布数512のモデルを用いた。この処理においては、ランダムガウス雑音ベクトル W_t の共分散行列を、 $\Sigma_w = \text{diag}(0.0001)$ に設定した。また、処理に用いるパーティクルの総数Jを50に設定した。

【0171】

なお、第1の実施の形態に係るパーティクルフィルタ（PF2）を用いて処理を行なう際、拘束条件パラメータ138には次の値を用いた。すなわち工場雑音環境下では、忘却係数 α 、フィードバック係数 β 、及びフレーム数Tのパラメータ値にそれぞれ、0.20、0.5、10を使用した。道路工事雑音環境下では、忘却係数 α 、フィードバック係数 β 、及びフレーム数Tのパラメータ値にそれぞれ、0.20、0.5、20を使用した。

40

【0172】

第2の実施の形態に係るパーティクルフィルタ（PF3）を用いて処理を行なう際、状態遷移確率の割当に使用するパラメータ γ には0.6を使用した。

【0173】

抑圧後の推定クリーン音声を用いた音声認識を行なう際の特徴量には、0次を含む13次MFCC（Mel Frequency Cepstrum Coefficient）と、1次及び2次の回帰係数とを含む39次元の特徴量を用いた。また、図1に示す認識用音響モデル109には、16

50

状態 20 混合分布の HMM を用いた。

【0174】

この認識実験における処理に、市販のクロック周波数 3.2 GHz、32 ビットの CPU (Central Processing Unit) を用いた場合、PF1、PF2、及び PF3 のいずれのパーティクルフィルタを用いた場合も、処理に要した時間は、観測信号における実時間と同程度であった。すなわち、認識処理を実時間で処理できることが明らかとなった。逐次 EM アルゴリズムによる雑音の推定の結果を用いて雑音抑圧処理を施した場合 (SEM)、その処理に要した時間は、結果が収束するまでの処理の反復回数がデータに依存して変化する。本実験においては処理時間は実時間の 2 倍～4 倍であった。

【0175】

図 23 に、雑音の推定実験の結果を示す。図 13 は、工場雑音の各フレームにおける、対数メルフィルタバンクの 1 番目のフィルタによる出力値 (以下、「真の雑音」と呼ぶ。) と、パーティクルフィルタを用いた雑音抑圧処理において当該出力値に対応する推定結果とを示すグラフである。このグラフにおいては、比較のため、状態遷移過程にランダムウォーク過程を仮定したパーティクルフィルタと、第 1 の実施の形態に係るパーティクルフィルタ (PF2) とによる推定結果を示す。このグラフの横軸は、フレーム番号を表す。すなわちグラフ上、左から右に進むにしたがい時間が経過する。

【0176】

図 23 を参照して、このグラフにおいて、44 フレームまでの区間は、雑音 121 のみが観測された区間であった。また、45 フレーム以降の区間は、雑音 121 と目的音声 120 との重畳した雑音重畳音声で観測された区間であった。雑音のみの区間においては、ランダムウォーク過程を仮定したパーティクルフィルタによる処理結果も、本実施の形態に係るパーティクルフィルタによる処理結果も、真の雑音に近い時間軌跡となっている。よってこの区間では、いずれの処理方法でも時間変動が推定できていることが分かる。しかし、雑音重畳音声で観測された区間において、本実施の形態に係るパーティクルフィルタによる雑音の推定誤差は、ランダムウォーク過程を導入したパーティクルフィルタによる雑音の推定誤差より小さくなっている。よって、本実施の形態に係る雑音抑圧処理により、特に雑音重畳音声で観測される区間において雑音が高精度に推定され、抑圧されることが分かる。

【0177】

表 1 及び表 2 に、各試料に対する認識実験で得られた単語認識率を、上記の処理方法別に示す。

【0178】

【表 1】

表 1 単語認識率 (工場雑音) (%)

SNR	HTK	ETSI	MMSE	SEM	PF 1	PF 2	PF 3
20 dB	93.61	92.88	96.41	96.50	96.13	96.71	97.54
15 dB	81.12	86.86	88.92	90.55	90.02	91.74	93.43
10 dB	54.81	76.73	74.27	75.07	75.87	82.13	85.08
5 dB	29.47	53.18	50.94	55.60	54.50	64.02	68.41
0 dB	18.73	23.15	24.72	26.48	28.92	38.66	43.20
Average	55.55	66.56	67.05	68.88	69.09	74.65	77.48

【0179】

【表 2】

表 2 単語認識率 (道路工事雑音) (%)

SNR	HTK	ETSI	MMSE	SEM	PF 1	PF 2	PF 3
20 dB	96.68	96.90	99.20	99.10	98.34	99.26	99.72
15 dB	89.93	94.81	97.61	97.90	95.61	98.28	98.83
10 dB	70.28	89.81	91.77	92.43	89.84	94.66	96.93
5 dB	38.81	76.02	71.57	77.20	75.28	81.79	86.15
0 dB	22.29	48.48	43.60	51.00	49.43	58.00	63.89
Average	63.60	81.20	80.75	83.53	81.70	86.40	89.10

10

【0180】

表 1 及び表 2 を参照して、パーティクルフィルタ P F 1、P F 2、又は P F 3 による雑音抑圧処理を行なうことで、高い単語認識率が得られることが分かる。中でも上記第 1 の実施の形態に係るパーティクルフィルタ P F 2 及び第 2 の実施の形態に係るパーティクルフィルタ P F 3 は、他のいずれの処理方法より高い単語認識率が得られることが分かる。よって、第 1 の実施の形態及び第 2 の実施の形態に係る拘束条件が有効な作用を奏したことは明らかである。特に第 2 の実施の形態に係るパーティクルフィルタ P F 3 による処理を施した試料に対する認識結果は、最良の結果を示しており、また、第 1 の実施の形態に係るパーティクルフィルタ (P F 2) による処理を施した試料に対する認識結果よりもさらに高い。よって、拘束条件パラメータを時間変動させることが、有効な作用を奏したこともまた明らかとなった。

20

【0181】

以上の実験結果から、本実施の形態の雑音抑圧処理により、非定常な雑音と目的音声とが重畳する環境下での音声認識率が改善され、かつ実時間処理が可能になることが明らかとなった。

30

【0182】

表 3 及び表 4 に、第 2 の実施の形態に係るパーティクルフィルタ (P F 3) において、パラメータ γ 570 を変化させた場合の単語認識率を示す。

【0183】

【表 3】

表 3 パラメータ γ を変化させた場合の単語認識率 (工場雑音) (%)

SNR	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.5$	$\gamma = 0.6$	$\gamma = 0.7$	$\gamma = 0.8$	$\gamma = 0.9$	$\gamma = 1.0$
20 dB	97.39	97.45	97.42	97.45	97.54	97.39	97.30	97.36	97.61	97.82
15 dB	92.91	93.09	92.88	92.78	93.43	93.15	93.55	93.49	93.55	93.52
10 dB	84.43	84.31	84.43	85.08	84.80	84.99	84.77	84.56	84.71	85.05
5 dB	67.95	67.76	68.07	67.70	68.41	68.31	68.10	67.67	68.25	67.82
0 dB	42.31	42.16	42.55	42.31	43.20	42.80	42.43	42.74	43.14	42.43
Average	77.00	76.95	77.07	77.06	77.48	77.33	77.23	77.16	77.45	77.33

40

【0184】

【表 4】

表 4 パラメータ γ を変化させた場合の単語認識率 (道路工事雑音) (%)

SNR	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.5$	$\gamma = 0.6$	$\gamma = 0.7$	$\gamma = 0.8$	$\gamma = 0.9$	$\gamma = 1.0$
20 dB	99.72	99.72	99.66	99.72	99.66	99.72	99.63	99.69	99.75	99.72
15 dB	98.99	98.89	98.99	98.93	99.08	98.83	99.14	98.86	98.96	99.02
10 dB	97.05	96.96	97.18	97.18	97.11	96.93	96.96	97.02	97.18	96.87
5 dB	86.18	86.03	86.12	85.75	85.97	86.15	86.37	85.54	85.94	85.88
0 dB	63.37	63.83	63.34	63.43	63.43	63.89	62.97	63.40	63.59	63.37
Average	89.06	89.09	89.06	89.00	89.05	89.10	89.01	88.90	89.08	88.97

表 3 及び表 4 より、パラメータ γ を変化させても、雑音種別、S N R に関わらず音声認識率の変動は 1 % 以下であった。単語認識率に対するパラメータ γ の依存性は低いと言える。したがって、拘束条件パラメータに含まれる忘却係数 α 、スケーリング係数 β 、及びフレーム数 T のフレームごとの変動を許容すること自体が、音声認識率の改善に寄与することと考えられる。

【0185】

[変形例等]

なお、上記各実施の形態においては、パーティクルフィルタによる処理を雑音の抑圧に用いている。そのため、雑音抑圧後の推定クリーン音声のパラメータを用いて探索を行なう前に、さらに音響モデル適応を行なうこともできる。音響モデル適応により、推定クリーン音声に適合した音響モデルを探索に用いることができるようになる。したがって、音声認識率が向上することが期待される。

【0186】

また、上記各実施の形態においては、前処理用の音響モデルに GMM を用いたが、前処理用の音響モデルに HMM を用いてもよい。この場合、上記の式 (20) に示す要素分布のサンプリングに先立ち、HMM の遷移確率にしたがって状態のサンプリングを行なえばよい。

【0187】

上記第 1 の実施の形態では、図 6 に示す Polyak Average 算出部 254 は、拘束条件パラメータ 138 により定められた T フレーム分の Polyak Average を算出した。しかし、第 1 の実施の形態においては、バッファメモリ部 252 が雑音確率分布の推定パラメータ 206 を T フレーム分保持し、Polyak Average 算出部 254 がバッファメモリ部 252 に保持されている当該 T フレーム分の雑音確率分布の推定パラメータ 206 を基に Polyak Average を算出するようにしてもよい。

【0188】

上記第 2 の実施の形態では、状態番号がパラメータ値の昇順に付与されていた。しかしパラメータ値の降順に付与されていても、昇順に付与された場合と同様に動作する。

【0189】

また、上記第 2 の実施の形態では、拘束条件パラメータの状態遷移確率は、直前のフレームに対応する状態に依存して決定された。しかし、本発明はこのような実施の形態には限定されない。例えば、直前のフレームのみでなく、より過去の状態の影響を考慮して、拘束条件パラメータに関する状態遷移確率の決定を行なうようにしてもよい。

【0190】

また、上記第 2 の実施の形態では、各エルゴディックモデル部は、全ての状態間での状態遷移に関し状態遷移確率の割当を予め決定しておき、フレームを移行するたびに決定された状態遷移確率の割当にしたがって、フレーム移行後の拘束条件パラメータの値を決定した。しかし、本発明はこのような実施の形態には限定されない。例えば、フレームを移行する度に、状態遷移確率の割当を決定するようにしてもよい。

【0191】

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

10

20

30

40

50

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のすべての変更を含む。

【図面の簡単な説明】

【0192】

【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】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4】雑音抑圧部114の構成を示すブロック図である。

10

【図5】雑音確率分布推定部200の構成を示すブロック図である。

【図6】更新部230の構成を示すブロック図である。

【図7】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8】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9】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10】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。

【図11】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12】Polyak Average及びフィードバックの概念を示す模式図である。

20

【図13】本発明の第2の実施の形態に係る前処理部500の構成を示す図である。

【図14】第2の実施の形態に係る雑音抑圧部510の構成示すブロック図である。

【図15】拘束条件パラメータ決定部520の構成を示すブロック図である。

【図16】状態パラメータ表540、544、及び548の一例を示す図である。

【図17】エルゴディックモデルの概念を示す図である。

【図18】エルゴディックモデル部542の構成を示すブロック図である。

【図19】状態遷移確率表574の一例を示す図である。

【図20】拘束条件パラメータ α の状態遷移の概要を示す図である。

【図21】第2の実施の形態に係る雑音抑圧処理の制御構造を示すフローチャートである。

30

【図22】拘束条件パラメータの決定処理の制御構造を示すブロック図である。

【図23】第1の実施の形態に係る雑音抑圧処理による工場雑音の推定実験の結果を示す図である。

【符号の説明】

【0193】

100 音声認識システム

102 音源

104, 500 前処理部

106 前処理用音響モデル部

108 言語モデル部

109 認識用音響モデル部

110 探索部

112 計測部

114, 510 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

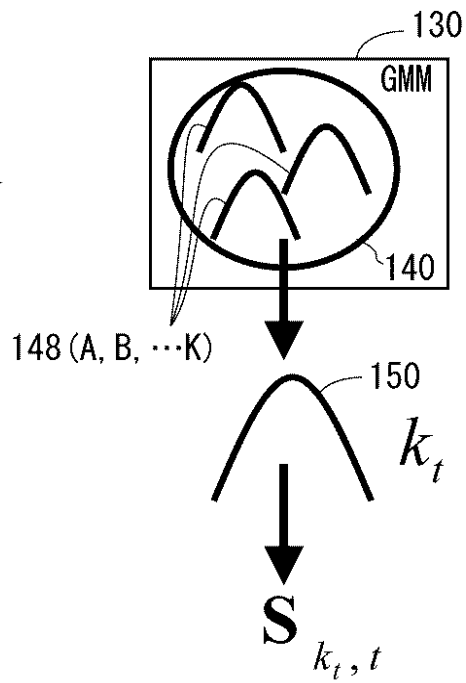
124 観測信号の特徴量

40

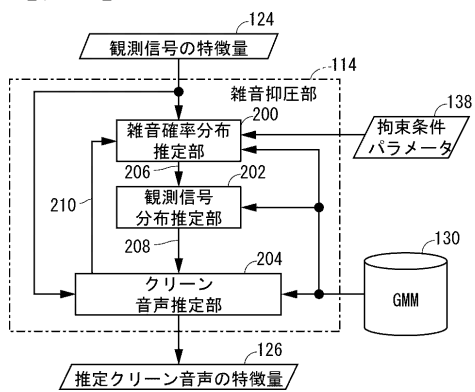
50

1 2 6	推定クリーン音声の特徴量	
1 3 0	GMM	
1 3 2	学習データ記憶部	
1 3 4	モデル学習部	
1 3 6	GMM記憶部	
1 3 8, 5 2 2	拘束条件パラメータ	
1 6 0	状態空間モデル	
2 0 0	雑音確率分布推定部	
2 0 2	観測信号分布推定部	
2 0 4	クリーン音声推定部	10
2 2 0	フレーム選択部	
2 2 2	雑音初期分布推定部	
2 2 4	逐次計算部	
2 2 6	GMMサンプリング部	
2 3 0	更新部	
2 3 2	重み算出部	
2 3 4	再サンプリング部	
2 3 6	推定パラメータ生成部	
2 4 0	出力パラメータ	
2 5 0	加重平均算出部	20
2 5 2	バッファメモリ部	
2 5 4	Polyak Average算出部	
2 5 6	フィードバック部	
2 5 8	拡張カルマンフィルタ部	
2 6 2	再更新部	
2 6 4	重み再計算部	
2 6 6	許容確率算出部	
2 6 8	乱数発生部	
2 7 0	パラメータ選択部	
5 2 0	拘束条件パラメータ決定部	30
5 3 0	α 決定部	
5 3 2	β 決定部	
5 3 4	T決定部	
5 3 6	拘束条件パラメータ出力部	
5 4 0	α 状態-パラメータ表	
5 4 2, 5 4 6, 5 5 0	エルゴディックモデル部	
5 4 4	β 状態-パラメータ表	
5 4 8	T状態-パラメータ表	
5 6 0	状態番号記憶部	
5 6 2	状態遷移確率決定部	40
5 6 4	状態遷移部	
5 6 6	パラメータ読出部	
5 7 0	パラメータ γ	
5 7 2	状態遷移確率割当部	
5 7 4	状態遷移確率表	

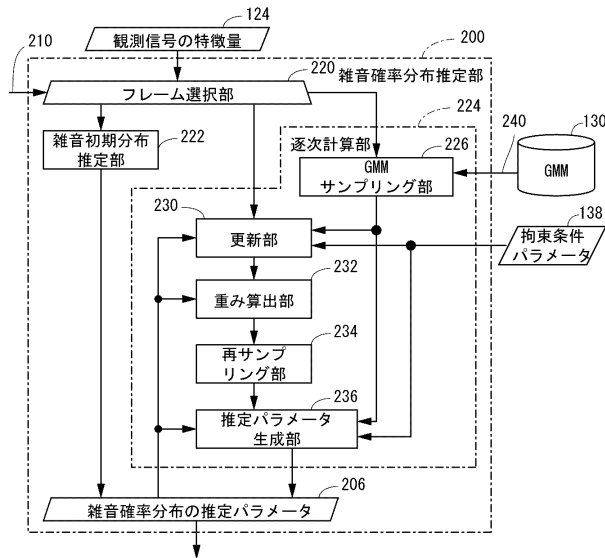
【図 2】



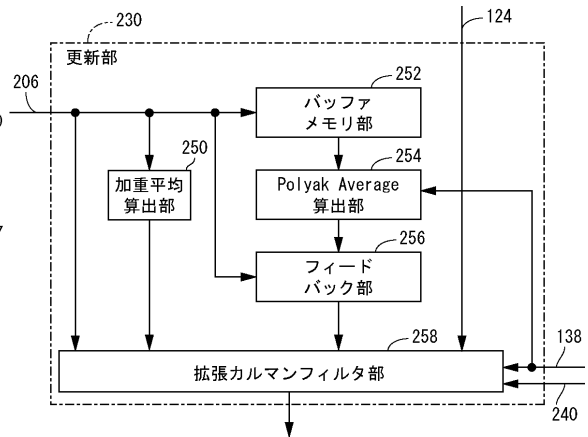
【图 4】



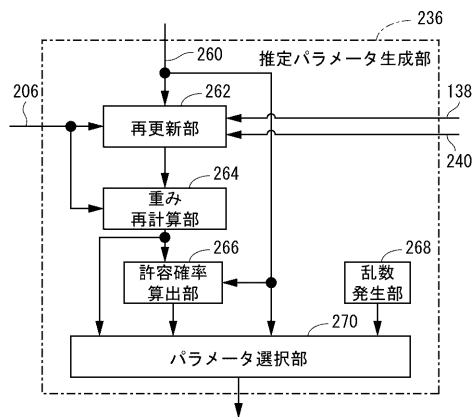
【図 5】



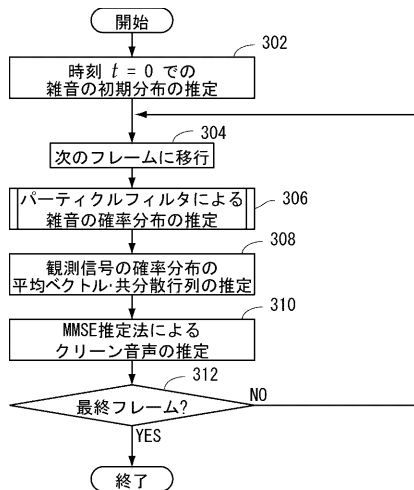
【図 6】



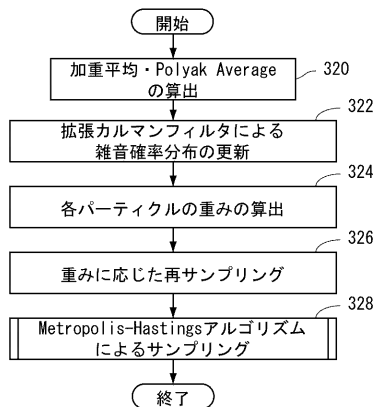
【図 7】



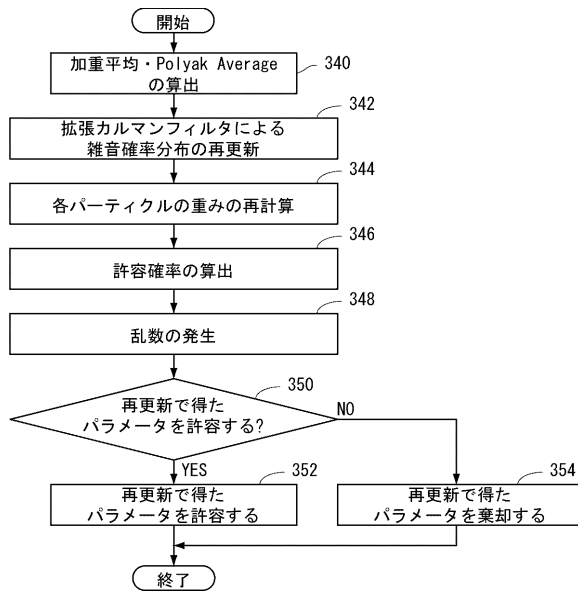
【図 8】



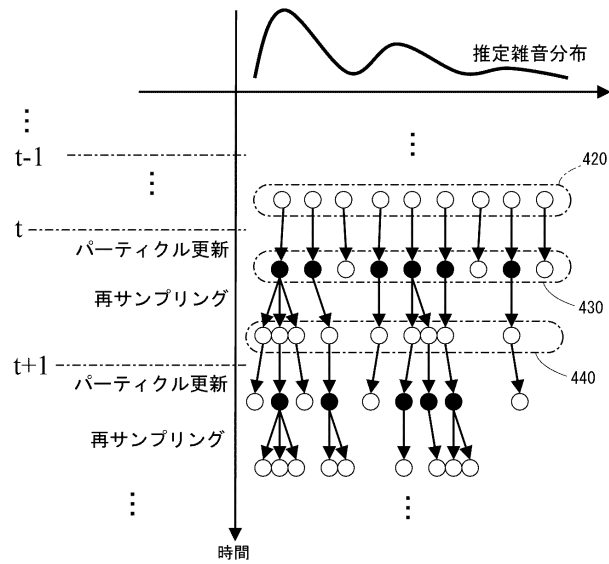
【図 9】



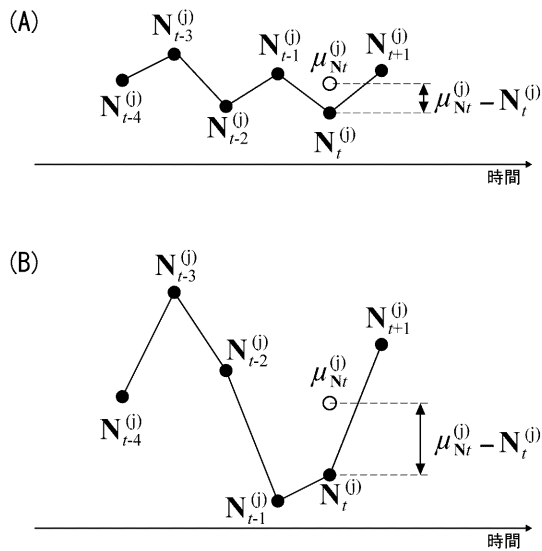
【図 10】



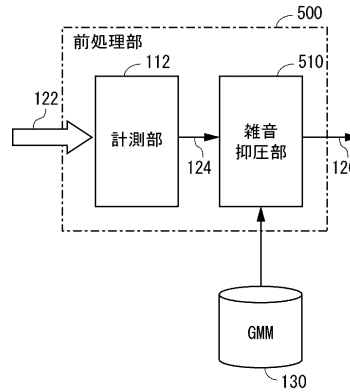
【図 11】



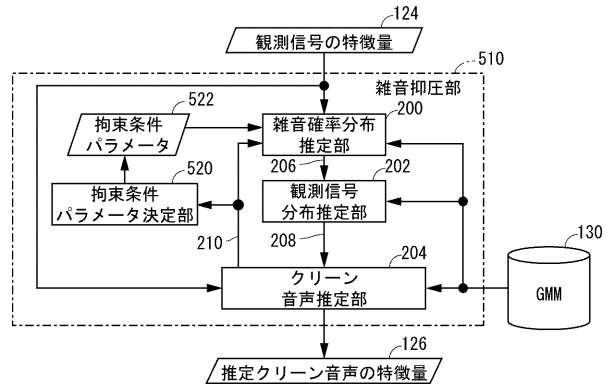
【図 12】



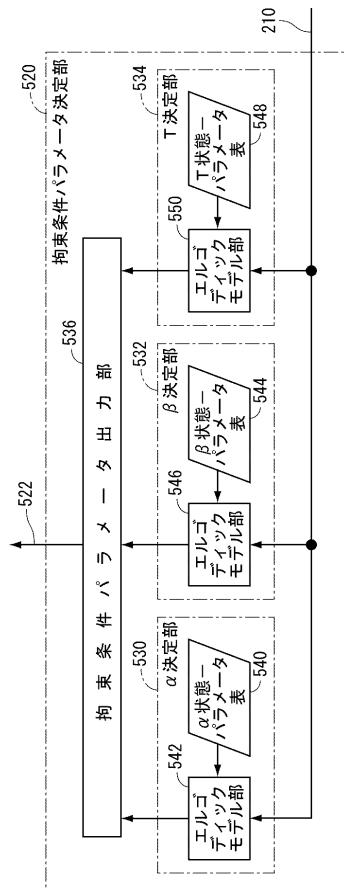
【図 13】



【図 14】



【図 15】

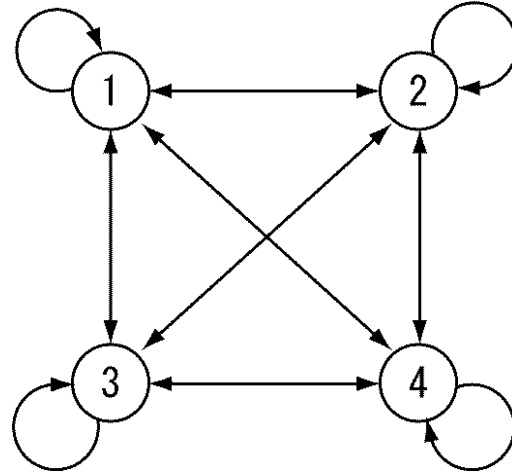


【図 16】

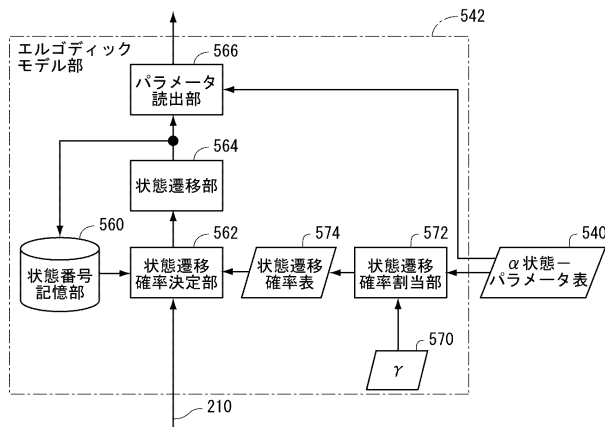
状態番号	α	状態番号	β	状態番号	γ
1	0.05	1	0.5	1	5
2	0.10	2	1.0	2	10
3	0.15	3	1.5	3	15
4	0.20	4	2.0	4	20

(A) (B) (C)

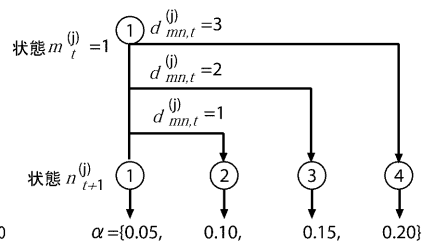
【図 17】



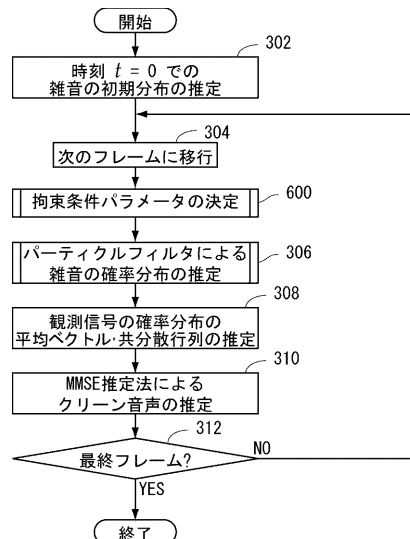
【図 18】



【図 20】



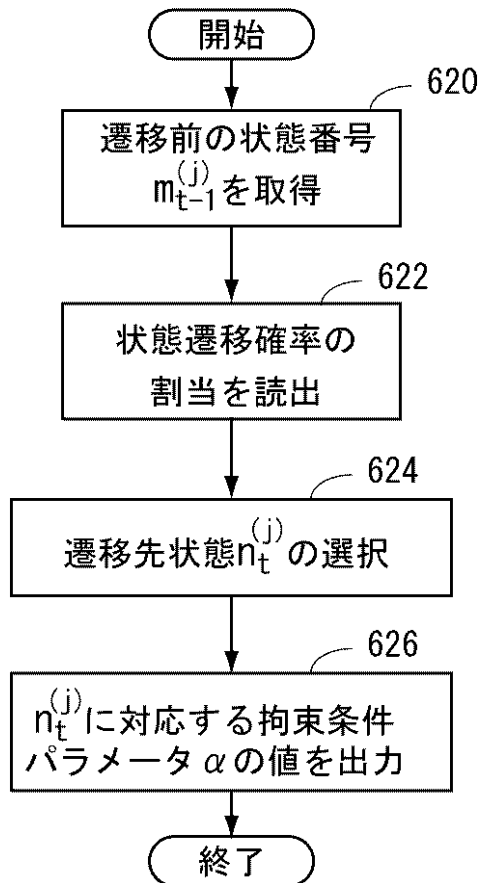
【図 21】



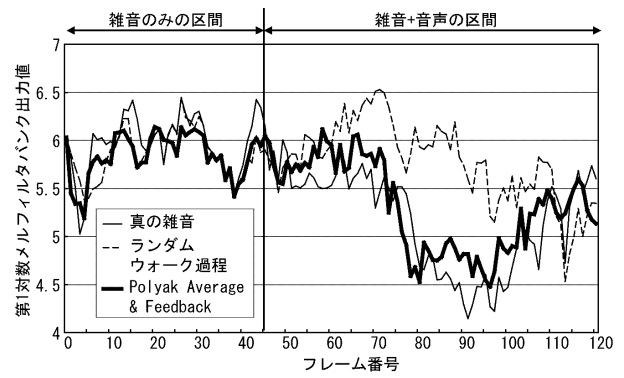
【図 19】

		遷移先状態			
		1	2	3	4
遷移前状態	1	0.533	0.267	0.133	0.067
	2	0.222	0.444	0.222	0.111
	3	0.111	0.222	0.444	0.222
	4	0.067	0.133	0.267	0.533

【図 2 2】



【図 2 3】



フロントページの続き

(56)参考文献 特開2002-251198 (JP, A)

特開2006-243290 (JP, A)

特開2006-201287 (JP, A)

藤本雅清 他, "パーティクルフィルタとPolyak Averagingを用いた非定常雑音の抑圧", 情報処理学会研究報告, 2005年 7月16日, Vol.2005, No.69, p.55-60

藤本雅清 他, "パーティクルフィルタに基づく雑音抑圧法を用いた非定常雑音及び残響環境下での音声認識", 日本音響学会2005年春季研究発表会講演論文集-I, 2005年 3月 8日, p.113-114

(58)調査した分野(Int.Cl., DB名)

G10L 11/00-21/06

JSTPlus (JDreamII)