

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4586577号  
(P4586577)

(45) 発行日 平成22年11月24日(2010.11.24)

(24) 登録日 平成22年9月17日(2010.9.17)

(51) Int.Cl.		F I			
<b>G 1 0 L</b>	<b>15/20</b>	<b>(2006.01)</b>	<b>G 1 0 L</b>	15/20	3 7 0 D
<b>G 1 0 L</b>	<b>21/02</b>	<b>(2006.01)</b>	<b>G 1 0 L</b>	21/02	1 0 1 B
			<b>G 1 0 L</b>	21/02	1 0 3 C

請求項の数 4 (全 25 頁)

<p>(21) 出願番号 特願2005-57993 (P2005-57993)</p> <p>(22) 出願日 平成17年3月2日(2005.3.2)</p> <p>(65) 公開番号 特開2006-243290 (P2006-243290A)</p> <p>(43) 公開日 平成18年9月14日(2006.9.14)</p> <p>審査請求日 平成18年12月20日(2006.12.20)</p> <p>(出願人による申告)平成16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</p> <p>前置審査</p>	<p>(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</p> <p>(74) 代理人 100099933 弁理士 清水 敏</p> <p>(72) 発明者 藤本 雅清 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</p> <p>(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</p> <p>審査官 間宮 嘉誉</p>
--	--

最終頁に続く

(54) 【発明の名称】 外乱成分抑圧装置、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】

外乱により加法性雑音及び乗法性歪みが生じる環境下で目的音声を観測することにより得られる観測信号の外乱成分を抑圧する外乱成分抑圧装置であって、前記外乱成分は、加法性雑音による成分、残響による成分、及び乗法性歪みによる成分を含み、

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て、前記残響による成分を加法性の外乱とみなし、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、前記外乱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外乱パラメータ推定手段と、

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、

前記外乱に含まれる残響成分を、前記観測信号と、前記外乱パラメータ推定手段により推定された前記加法性雑音による成分との差によって近似し、前記外乱パラメータ推定手段に入力するための手段とを含む、外乱成分抑圧装置。

【請求項2】

前記外乱パラメータ推定手段は、

前記外乱の初期分布を設定し、当該初期分布にしたがった確率で、前記複数のパーティクルの各々における外乱を表す確率分布の初期パラメータをそれぞれ設定するための初期パラメータ設定手段と、

前記音響モデルと前記観測信号の特徴量とを基に、前記残響による成分を加法性の外乱とみなし、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先行する第1のフレームの前記推定パラメータをそれぞれ前記第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の重みを算出するための重み算出手段とを含む、請求項1に記載の外乱成分抑圧装置。

【請求項3】

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又は請求項2に記載の外乱成分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4】

請求項1又は請求項2に記載の外乱成分抑圧装置と、

前記外乱成分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、音声に影響を与えるような外乱が発生する実環境下での音声認識技術に関し、特に、非定常的な加法性雑音と残響等の乗法性歪みとが発生する環境下での音声認識精度を改善するための、外乱成分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、人間が機械と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間なく不規則に発生する。加えて実環境ではさらに、その環境下での音声の空間伝達特性に依存して、又は残響等によって、音声に乗法性の歪みが生じる。このような外乱は、音声認識を行なう際の妨げとなる。これらの外乱が生じる環境下での音声認識性能の改善は、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】

雑音環境下での音声認識性能を改善するための技術のひとつに、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。後掲の非特許文献1には、雑音抑圧の一般的な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が発話中の区間における雑音の振幅スペクトルと同じであると仮定し、発話から得ら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することで雑音を抑圧する。

【0005】

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。非特許文献2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手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手法では、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。

【0006】

10

20

30

40

50

非特許文献3及び非特許文献4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手法も一般的に用いられている。この手法では、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。

【0007】

また、雑音環境下での音声認識性能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、隠れマルコフモデル(HMM: Hidden Markov Model)を構成する隠れ状態の時間的成長とを行ない、当該隠れマルコフモデル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。

10

【0008】

乗法性歪みが生じる環境下での音声認識性能を改善するための技術として、ケプストラム平均減算法(Cepstrum Mean Subtraction: CMS)を用いて乗法性歪みを除去する技術がある。この手法は、収録用のマイクロホンの特性による歪みなど、分析窓長より短いインパルス応答長の伝達特性を持つ乗法性歪みを除去することができる。

【0009】

非特許文献5には、反射音を加法性雑音とみなして残響下での音声認識を行なう技術が開示されている。この技術では、残響下で観測される音声(以下、「残響音声」と呼ぶ。)を、1次線形予測により表現している。ここに、時刻tにおける目的音声及び残響音声の線形メルスペクトルを要素に持つベクトルを $S_t^{Lin}$ 、及び $X_{S,t}^{Lin}$ とし、各メル周波数領域での音声の伝達特性、すなわち乗法性歪みの線形メルスペクトルを対角成分に持つ行列を $H^{Lin}$ とする。また、各メル周波数領域での残響の線形予測係数を対角成分に持つ行列を $A^{Lin}$ とする。この技術では、残響音声のベクトル $X_{S,t}^{Lin}$ を次の再帰式によって表現する。

20

【0010】

$$X_{S,t}^{Lin} = H^{Lin} S_t^{Lin} + A^{Lin} X_{S,t-1}^{Lin}$$

また、この技術では、行列 $H^{Lin}$ の要素すなわち乗法性歪みの線形メルスペクトルと、行列 $A^{Lin}$ の要素すなわち残響の線形予測係数を、それぞれ時間固定のパラメータとみなし、EMアルゴリズムによりこれらのパラメータを推定する。上記の再帰式により、分析窓長よりも長いインパルス応答長の歪みも表現されるため、反射音の影響等をモデル化することができる。

30

【0011】

【特許文献1】特開2002-251198号公報

【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】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)

40

【非特許文献3】有本 卓:「カルマンフィルタ」、産業図書

【非特許文献4】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

【非特許文献5】タキグチ テツヤ他:「1次線形予測を用いた残響音声の認識」、ICASSP'04, 869-872頁、2004年5月(T.Takiguchi et al: "Reverberant speech recognition using first-order linear prediction," ICASSP'04, pp. 896-872, May 2004)

50

## 【発明の開示】

## 【発明が解決しようとする課題】

## 【0012】

実環境において雑音の多くは非定常な雑音である。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。非特許文献1に記載のスペクトルサブトラクション法のように、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。

## 【0013】

非特許文献2に記載された逐次EMアルゴリズムを用いた手法は、尤度関数の局所最適値に値が収束するまで反復計算を行なう。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。

10

## 【0014】

非特許文献3及び非特許文献4に開示されたカルマンフィルタを用いる推定方法は、一期先予測とフィルタリングとを交互に行ない逐次推定を行なう。そのため、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。

## 【0015】

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。

20

## 【0016】

非特許文献5に記載の技術では、上記の再帰式により反射音の影響についてモデル化されている。しかし、一般に残響は、音源から離れた地点で音を観測又は収録する際に生じる現象である。音源と観測地点との距離が離れている環境下では、反射音のみならず、音源と観測地点とを取り巻く環境で発生する雑音の存在が無視できなくなる。非特許文献5に記載の技術では、この点について考慮されていない。また、非特許文献5に記載の技術では、行列 $H^{Lin}$ の要素すなわち乗法性歪みの線形メルスペクトルと、行列 $A^{Lin}$ の要素すなわち残響の線形予測係数を、それぞれ時間固定のパラメータとみなしている。しかし、実環境においては、例えば音源及びその周囲で音を反射する物体が移動することがある。このような環境下では、乗法性歪みのパラメータも、残響の線形予測係数も時間の経過とともに変動する。そのため、非特許文献5に記載の技術では、残響の時間変動に対応できず、外乱の影響に高精度に対処することができない。

30

## 【0017】

それゆえに、本発明の目的は、非定常雑音及び残響等の乗法性歪みが生じる環境下での音声認識性能を改善し、かつ外乱成分の抑圧を短時間で行なうことができる外乱成分抑圧装置を提供することである。

40

## 【課題を解決するための手段】

## 【0018】

本発明の第1の局面に係る外乱成分抑圧装置は、加法的雑音及び乗法性歪みが生じる環境下で目的音声を観測することにより得られる観測信号の、外乱の成分を抑圧する装置である。この装置は、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て、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、外乱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外乱パラメータ推定手段と、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。

50

## 【 0 0 1 9 】

好ましくは、外乱パラメータ推定手段は、外乱の初期分布を設定し、当該初期分布にしたがった確率で、複数のパーティクルの各々における外乱を表す確率分布の初期パラメータをそれぞれ設定するための初期パラメータ設定手段と、音響モデルと観測信号の特徴量とを基に、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先行する第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の重みを算出するための重み算出手段とを含む。

## 【 0 0 2 0 】

より好ましくは、初期パラメータ設定手段は、観測信号の特徴量を基に加法性雑音の初期分布を推定し、当該初期分布にしたがった確率で、複数のパーティクルの各々における加法性雑音の確率分布の初期パラメータをそれぞれサンプリングするための手段と、複数のパーティクルの各々における乗法性歪みの確率分布の初期パラメータの値を所定値に設定するための手段とを含む。

10

## 【 0 0 2 1 】

より好ましくは、外乱パラメータ推定手段はさらに、再サンプリング手段により再サンプリングされたパラメータを基に、複数のパーティクルの各々において、第1のフレームに対応する推定パラメータをそれぞれ、第2のフレームに対応するものに再更新するための再更新手段と、複数のパーティクルの各々において、再更新手段により再更新された推定パラメータと、再サンプリング手段により再サンプリングされた推定パラメータとの一方を、所定の判定基準にしたがい第2のフレームの推定パラメータとして選択するための選択手段とを含む。

20

## 【 0 0 2 2 】

好ましくは、目的音声推定手段は、観測信号の特徴量と、推定パラメータと、音響モデルとを基に、フレームに対応する観測信号の確率モデルを合成するための観測信号モデル合成手段と、観測信号の特徴量、推定パラメータ、音響モデル、及び観測信号の確率モデルを基に、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推定特徴量算出手段とを含む。

## 【 0 0 2 3 】

より好ましくは、観測信号モデル合成手段は、複数のパーティクルの各々に対して、推定パラメータと、音響モデルとを基に、当該パーティクルにおける観測信号の確率モデルのパラメータを推定するためのパラメータ推定手段を含む。

30

## 【 0 0 2 4 】

推定特徴量算出手段は、フレームごとに、複数のパーティクルの各々の目的音声の推定パラメータを、観測信号の特徴量、音響モデル、推定パラメータ、及び観測信号の確率モデルを基に算出するための手段と、複数のパーティクルの各々における目的音声の推定パラメータを基に、当該フレームにおける目的音声の推定特徴量を算出するための手段とを含んでもよい。

## 【 0 0 2 5 】

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、コンピュータを本発明の第1の局面に係るいずれかの外乱成分抑圧装置として動作させる。

40

## 【 0 0 2 6 】

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外乱成分抑圧装置と、外乱成分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 】

## 【 0 0 2 7 】

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

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 $W$ 」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。

#### 【0028】

##### [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を収集し、収集した音から認識に用いる特徴量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、前処理部104から出力された特徴量に対応する単語等を、言語モデル部108の言語モデルを用いて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。

10

#### 【0029】

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の周囲において音の伝達に影響を及ぼす外乱要因118とを含む。前処理部104に到達する音122は、話者116の発話により発生した目的音声120ではなく、外乱要因118の影響を受けて変化した音となる。本明細書では、話者116の発話により発生する雑音のない目的音声120を、「クリーン音声」と呼ぶ。また、前処理部104により収録される音、すなわち外乱要因118の影響により変化した状態で前処理部104に到達する音122を「観測音」と呼ぶ。

20

#### 【0030】

前処理用音響モデル部106は、クリーン音声120に対するガウス混合モデル(GMM: Gaussian Mixture Model)からなる音響モデルを準備し保持する。前処理用音響モデル部106は、予め用意された大量の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に記憶された学習データを用いてGMMに対する学習処理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GMM130を記憶するためのGMM記憶部136とを含む。

30

#### 【0031】

図2に、GMM130の概念を模式的に示す。図2を参照して、GMM130は、時系列信号の値を、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GMM130においては、出力確率が定義される。具体的には、GMM130では、時刻 $t$ においてクリーン音声120として出力される可能性のある値と、その値が出力される確率とが定義される。GMM130においては、出力確率は混合正規分布によって表現される。例えばGMM130は単一正規分布148A, 148B, ..., 148Kからなる混合正規分布を持つ。

40

#### 【0032】

再び図1を参照して、前処理部104は、観測音122を収録し、得られる観測信号に所定の信号処理を施すことにより、当該観測信号に関する所定の特徴量ベクトル(以下単に「特徴量」と呼ぶ。)124を抽出するための計測部112と、計測部112が抽出する特徴量124に含まれる外乱の成分を、GMM130を用いて抑圧するための外乱成分抑圧部114とを含む。

#### 【0033】

具体的には、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特徴量1

50

24として出力する。

【0034】

外乱成分抑圧部114は、GMM130を用いて、観測信号の特徴量124からクリーン音声120の特徴量を推定する。そして推定により得られた特徴量126を探索部110に出力する。本明細書では、推定クリーン音声の特徴量126によって表される音声を「推定クリーン音声」と呼ぶ。

【0035】

図3に、外乱要因118の信号モデルを模式的に示す。図3を参照して、クリーン音声120は、図1に示す話者116から計測部112までの空間伝達特性等に依存して乗法性歪みを受けるため、話者116から計測部112に到達する音500は、クリーン音声120と異なる音となる。ここに、時刻tのフレーム（以下、単に「第tフレーム」と呼ぶ。）におけるクリーン音声120の線形メルスペクトルを要素に持つベクトルを $S_t^{Lin}$ とし、乗法性歪みの線形メルスペクトルを対角成分に持つ行列を $H_t^{Lin}$ とする。計測部112に到達する音500の線形メルスペクトルを要素に持つベクトルを $X_{S,t}^{Lin}(D)$ とすると、 $X_{S,t}^{Lin}(D)$ は一般に、次の式で表現される。すなわち、

$$X_{S,t}^{Lin}(D) = H_t^{Lin} S_t^{Lin}$$

【0036】

しかし、観測音122は、残響の影響を受ける。すなわち、直接的に到達する音500だけでなく、周囲の壁面等により反射して計測部112に到達する反射音502の影響を受ける。本実施の形態では、反射音を加法性雑音とみなす。反射音502の線形メルスペクトルを要素に持つベクトルを $X_{S,t}^{Lin}(R)$ とし、残響の影響を受けて計測部112に到達した音を $X_{S,t}^{Lin}$ とすると、 $X_{S,t}^{Lin}$ は、次の式で表現される。すなわち、

$$X_{S,t}^{Lin} = X_{S,t}^{Lin}(D) + X_{S,t}^{Lin}(R)$$

【0037】

直接音500と反射音502とはいずれも話者116が発した音であるが、伝播する経路の違いにより、反射音502は、直接音500より遅延して計測部112に到達する。非特許文献5によれば、各メル周波数帯域における残響の線形予測係数を対角成分に持つ行列を行列 $A_t^{Lin}$ とすると、 $X_{S,t}^{Lin}(R)$ は、次の式で表現される。すなわち、

$$X_{S,t}^{Lin}(R) = A_t^{Lin} X_{S,t-1}^{Lin}$$

【0038】

さらに実環境では、話者116及び計測部112の周囲において雑音504が発生し、計測部112に到達する。ここに雑音504の線形メルスペクトルを要素に持つベクトルを $N_t^{Lin}$ とし、観測音122の線形メルスペクトルを要素に持つベクトルを $X_{S+N,t}^{Lin}$ とする。 $X_{S+N,t}^{Lin}$ は、次の信号モデルによりモデル化できる。すなわち、

$$X_{S+N,t}^{Lin} = X_{S,t}^{Lin} + N_t^{Lin} = H_t^{Lin} S_t^{Lin} + N_t^{Lin} + A_t^{Lin} X_{S,t-1}^{Lin}$$

反射音は観測できないため、この式において反射音のベクトル $X_{S,t-1}^{Lin}$ を次のように近似する。すなわち、

$$X_{S,t-1}^{Lin} = X_{S+N,t-1}^{Lin} - N_{t-1}^{Lin}$$

【0039】

第tフレームにおける観測信号の特徴量124、すなわち観測音122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を特徴量ベクトル $X_t$ とする。なお、特徴量ベクトル $X_t$ は、ベクトル $X_{S,t}^{Lin}$ の各要素を対数メルスペクトル領域に変換することにより得られるベクトルである。特徴量ベクトル $X_t$ は、計測により得られる既知のパラメータである。特徴量ベクトル $X_t$ は、クリーン音声120の対数メルスペクトルを要素に持つベクトル $S_t$ が外乱の影響で変化したベクトルである。ベクトル $S_t$ は、未知のベクトルである。外乱には、乗法性歪み、残響、及び加法性雑音による影響分が含まれる。ここに、乗法性歪みの対数メルスペクトルを対角成分に持つ行列を $H_t$ とし、加法性雑音の対数メルスペクトルを要素に持つベクトルを $N_t$ とする。また、外乱には、残響による影響分も含まれる。さらに第tフレームにおける線形予測係数の対数を対角成分に持つ行列を行列 $A_t$ 、反射音のベクトル $X_{S,t-1}^{Lin}$ の各要素を対数化したベクトルを $X_{S,t-1}$ とする。

10

20

30

40

50

## 【 0 0 4 0 】

上記したベクトル  $X_t$ 、 $S_t$ 、 $N_t$ 、及び  $X_{S,t-1}$  の次元数は同一である。また、行列  $H_t$  及び  $A_t$  の行数及び列数は同一である。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、簡単のために各の要素を特に区別して言及することはしない。

## 【 0 0 4 1 】

図 4 に、観測信号の観測過程及び雑音の状態変化過程を表現する状態空間モデル 160 を示す。図 4 を参照して、状態空間モデル 160 において、クリーン音声 120 の出力過程は 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 120 の成分であるベクトル  $S_t$  は、GMM 130 内のある要素分布にしたがって出力されるものと仮定する。

10

## 【 0 0 4 2 】

GMM 130 において、第  $t$  フレームに対応する要素分布を  $k_t$  とする。なお、要素分布  $k_t$  は、平均を  $\mu_{S,k_t}$  とし分散を  $\Sigma_{S,k_t}$  とする単一正規分布とする。また、要素分布  $k_t$  から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、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。クリーン音声 120 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。また、 $X_{S+N,t}^{Lin}$  を対数メルスペクトル領域に変換する際にも誤差を伴う。これらの誤差もまたベクトルであり、これらの誤差のベクトルをまとめて、ベクトル  $V_t$  とする。また、外乱要因 118 による外乱を表す行列を  $\Lambda_t = (N_t, H_t, A_t)$  とする。観測信号の特徴量ベクトル  $X_t$  (124) の観測過程は、上記の  $X_{S+N,t}^{Lin}$  を対数メルスペクトル領域に変換することにより、GMM 130 の出力パラメータベクトル  $S_{k_t,t}$  及び誤差ベクトル  $V_t$ 、並びにベクトル  $X_t$ 、 $S_t$ 、 $N_t$ 、及び  $X_{S,t-1}$ 、並びに行列  $H_t$  及び  $\Lambda_t$  を用いて、次の式 (1) により表現される。

20

## 【 0 0 4 3 】

## 【数 1】

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

30

なお、式 (1) で  $\mathbf{I}$  は単位行列を表す。また行列の対数、行列の指数演算はそれぞれ、行列の各要素について対数を取り、又は指数計算をし、その結果を成分とする行列を表すものとする。誤差ベクトル  $V_t$  は、次の式 (2) のように平均が 0 で分散が  $\Sigma_{S,k_t}$  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。

## 【 0 0 4 4 】

## 【数 2】

$$V_t \sim \mathcal{N}(0, \Sigma_{S,k_t}), \quad (2)$$

40

ただしこの式において  $\Sigma_{S,k_t}$  は GMM 130 内のある要素分布  $k_t$  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、平均値ベクトル  $\mu$ 、分散  $\Sigma$  の単一正規分布を表す。

## 【 0 0 4 5 】

また状態空間モデル 160 において、外乱を表す行列  $\Lambda_t$  は、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  $t-1$  フレームにおける外乱を表す行列  $\Lambda_{t-1}$  と時刻  $t$  における外乱を表す行列  $\Lambda_t$  との間に誤差が生じるものと仮定する。ベクトル  $N_t$ 、行列  $H_t$ 、及び行列  $A_t$  に対するこの誤差をそれぞれ、ベクトル  $W_N$ 、行列  $W_H$ 、

50

及び行列 $W_A$ とし、これらをまとめて誤差を表す行列 $W_t = (W_{Nt}, W_{Ht}, W_{At})$ と定義する。外乱を表す行列 $W_t$ の時間変動は、次の式(3)により表現される。

【0046】

【数3】

$$\Lambda_t = \Lambda_{t-1} + W_t \quad (3)$$

この式において、誤差行列 $W_t = (W_{Nt}, W_{Ht}, W_{At})$ を構成する、ベクトル $W_{Nt}$ 、行列 $W_{Ht}$ 、及び行列 $W_{At}$ はそれぞれ予測誤差であり、それぞれ平均0、共分散行列が $\Sigma_{WN}$ 、 $\Sigma_{WH}$ 、及び $\Sigma_{WA}$ の単一正規分布で表現される確率分布にしたがう白色性ガウス雑音であるものとする。誤差を表す行列 $W_t$ は、次の式(4)のように単一正規分布にしたがう。

【0047】

【数4】

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

$$\Sigma_W = \text{diag}(\Sigma_{WN}, \Sigma_{WH}, \Sigma_{WA}) \quad (4)$$

ただし、式(4)において $\Sigma_W$ は、誤差を表す行列 $W_t$ の共分散行列を表す。

【0048】

図1に示す外乱成分抑圧部114は、上記の式(1)~式(4)により表現される状態空間モデル160を用いて、フレームごとに、クリーン音声の特徴量ベクトルを逐次推定する。

【0049】

図5に、外乱成分抑圧部114の構成をブロック図で示す。図5を参照して、外乱成分抑圧部114は、観測信号の特徴量 $X_t$ (124)を受けて、GMM130を用いて状態空間モデル160における外乱を表す行列 $W_t$ の確率分布(以下、「外乱確率分布」と呼ぶ。)を推定するための外乱確率分布推定部200と、外乱確率分布推定部200により推定された外乱確率分布とGMM130とから観測信号の確率モデルの平均ベクトルと共分散行列とを生成するためのパラメータ生成部202と、外乱確率分布、観測信号の平均ベクトル及び共分散行列、並びにGMM130を用いて、推定クリーン音声の特徴量126を算出するためのクリーン音声推定部204とを含む。

【0050】

外乱確率分布推定部200は、外乱確率分布をフレームごとに逐次推定し、外乱確率分布を表すパラメータ206(以下、単に「推定外乱分布206」と呼ぶ。)を出力する機能を持つ。ここに、外乱を表す行列 $W_0, \dots, W_t$ からなる行列の系列を系列 $W_{0:t} = \{W_0, \dots, W_t\}$ とする。系列 $W_{0:t}$ の事後確率分布 $p(W_{0:t} | X_{0:t})$ は、1次マルコフ連鎖を用いて、次の式(5)のように表される。

【0051】

【数5】

$$p(\Lambda_{0:t} | \mathbf{X}_{0:t}) = p(\Lambda_0 | \mathbf{X}_0) \prod_{t'=1}^t p(\Lambda_{t'} | \Lambda_{t'-1}) p(\mathbf{X}_{t'} | \Lambda_{t'}) \quad (5)$$

式(5)の $p(W_t | W_{t-1})$ は、単一正規分布を用いて次の式(6)のようにモデル化される。

【0052】

【数6】

$$p(\Lambda_t | \Lambda_{t-1}) = \mathcal{N}(\Lambda_t; \Lambda_{t-1}, \Sigma_W) \quad (6)$$

また、式(5)の $p(X_t | W_t)$ は、単一正規分布を用いて次の式(7)のようにモデル

10

20

30

40

50

化される。

【 0 0 5 3 】

【 数 7 】

$$p(\mathbf{X}_t | \Lambda_t) = \mathcal{N}(\mathbf{X}_t; f(\mathbf{S}_{k_t, t}, \Lambda_t), \Sigma_{\mathbf{S}, k_t}) \quad (7)$$

【 0 0 5 4 】

したがって、状態空間モデル 1 6 0 を基に外乱を表す行列  $\mathbf{X}_t$  の確率分布を逐次推定する問題は、観測信号ベクトル  $\mathbf{X}_t$  が与えられた時の事後確率を最大にするような系列  $\Lambda_{0:t}$  を推定する問題に帰着する。外乱確率分布推定部 2 0 0 は、観測信号ベクトル  $\mathbf{X}_t$  と状態空間モデル 1 6 0 とに基づき、この推定を行なう。

10

【 0 0 5 5 】

外乱確率分布推定部 2 0 0 は、外乱を表す行列  $\mathbf{X}_t$  の確率分布を逐次的に推定する際に、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、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、状態空間内におけるパラメータの確率分布を、各パーティクルにおいて推定された確率分布を用いて近似的に表現する手法である。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、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し、その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。そして、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。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。状態空間におけるパラメータは、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。すなわち、パーティクルの数を  $J$ 、 $j$  番目のパーティクルにおいて外乱を表す行列  $\mathbf{X}_t$  に対応する各パラメータからなる行列を行列  $\mathbf{X}_t^{(j)}$  とし、当該パーティクルに対する重みを  $w_t^{(j)}$  とすると、式 ( 5 ) に示す系列  $\Lambda_{0:t}$  の事後確率分布  $p(\Lambda_{0:t} | \mathbf{X}_{0:t})$  は、次の式 ( 8 ) によって近似的に表現される。

20

30

【 0 0 5 6 】

【 数 8 】

$$p(\Lambda_{0:t} | \mathbf{X}_{0:t}) \simeq \sum_{j=1}^J w_t^{(j)} p(\Lambda_{0:t}^{(j)} | \mathbf{X}_{0:t}^{(j)}) \quad (8)$$

【 0 0 5 7 】

パラメータ生成部 2 0 2 は、具体的には V T S ( Vector Taylor Series ) 法と呼ばれる H M M 合成法によって、パーティクルフィルタにより推定された外乱確率分布を用い、複数のパーティクルにおける観測信号の特徴量ベクトル  $\mathbf{X}_t$  の平均ベクトル及び共分散行列 ( 2 0 8 ) をそれぞれ算出する機能を持つ。

40

【 0 0 5 8 】

クリーン音声推定部 2 0 4 は、最小 2 乗誤差 ( Minimum Mean Square Error : M M S E ) 推定法で、フレームごとに、複数のパーティクルにおけるクリーン音声のパラメータをそれぞれ推定し、それら推定されたパラメータの重み付き和によって推定クリーン音声の特徴量 1 2 6 を算出する機能を持つ。クリーン音声推定部 2 0 4 はさらに、外乱確率分布推定部 2 0 0 に、次のフレームへの移行に関する要求 2 1 0 を発行する機能を持つ。

【 0 0 5 9 】

図 6 に、外乱確率分布推定部 2 0 0 の構成をブロック図で示す。図 6 を参照して、外乱

50

確率分布推定部 200 は、観測信号の特徴量 124 とクリーン音声推定部 204 からの要求 210 とを受けて、処理対象となるフレームを選択し、当該フレームにおける観測信号の特徴量 124 をフレームに応じた出力先に出力するためのフレーム選択部 220 と、フレーム選択部 220 から最初の所定フレーム分の観測信号の特徴量 124 を受けて初期状態における外乱確率分布を推定し、各パーティクルにおける外乱の初期的なパラメータを決定するための外乱初期分布推定部 222 と、フレーム選択部 220 から  $t$  ( $t = 1$ ) 番目フレームにおける観測信号の特徴量 124 を受けて、逐次的に、パーティクルにおける雑音のパラメータと当該パーティクルに対する重みとを算出するための逐次計算部 224 とを含む。

【0060】

10

外乱初期分布推定部 222 は、時刻  $t = 0$  のフレームにおける外乱を表す行列  $\mathbf{N}_0 = (N_0, H_0, A_0)$  の確率分布（以下、「外乱初期分布」）を推定する。この際、加法性雑音の初期分布を以下のようにして推定する。

【0061】

外乱初期分布推定部 222 はまず、加法性雑音の初期分布、すなわち加法性雑音の初期値ベクトル  $\mathbf{N}_0$  の確率分布  $p(\mathbf{N}_0)$  が、単一正規分布であるものとみなし、加法性雑音の初期分布を推定する。加法性雑音の初期分布における平均ベクトルを  $\mu_N$  とし、共分散行列を行列  $\Sigma_N$  とすると、加法性雑音の初期分布  $p(\mathbf{N}_0)$  は次の式 (9) のように表される。

【0062】

20

【数 9】

$$p(\mathbf{N}_0) = \mathcal{N}(\mathbf{N}_0; \mu_N, \Sigma_N) \quad (9)$$

外乱初期分布推定部 222 は、最初の所定フレーム分の区間の観測信号の特徴量ベクトル  $\mathbf{X}_t$  が加法性雑音の成分のみからなるものとみなし、加法性雑音の初期分布の平均ベクトル  $\mu_N$ 、及び共分散行列  $\Sigma_N$  を推定する。例えば 0 ~ 9 の 10 フレーム分の区間がこの区間に該当する場合、外乱初期分布推定部 222 は、平均ベクトル  $\mu_N$  及び共分散行列  $\Sigma_N$  をそれぞれ、次の式 (10) 及び式 (11) によって算出する。ただし、ベクトルの右肩に付した「T」は転置を表す。

【0063】

30

【数 10】

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

$$\Sigma_N = \frac{1}{10} \sum_{t=0}^9 (\mathbf{X}_t - \mu_N) (\mathbf{X}_t - \mu_N)^T \quad (11)$$

次に外乱初期分布推定部 222 は、初期状態での各パーティクルにおける外乱を表す行列  $\mathbf{N}_0^{(i)}$  を構成するベクトル  $N_0^{(i)}$ 、行列  $\mathbf{H}_0^{(i)}$ 、及び行列  $\mathbf{A}_0^{(i)}$  を、式 (12) のように設定する。

40

【0064】

【数 11】

$$\mathbf{N}_0^{(j)} \sim p(\mathbf{N}_0), \mathbf{H}_0^{(j)} = \mathbf{A}_0^{(j)} = 0 \quad (12)$$

すなわち、各パーティクルにおける加法性雑音のベクトル  $N_0^{(i)}$  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、各パーティクルにおける乗法性歪みの行列  $\mathbf{H}_0^{(i)}$  及び残響の行列  $\mathbf{A}_0^{(i)}$  の各要素の値を 0 に設定する。

50

## 【 0 0 6 5 】

さらに外乱初期分布推定部 2 2 2 は、各パーティクルにおける外乱を表す行列  $\mathbf{N}_0^{(j)}$  を構成するベクトル  $\mathbf{N}_0^{(j)}$ 、行列  $\mathbf{H}_0^{(j)}$ 、及び行列  $\mathbf{A}_0^{(j)}$  の共分散行列  $\Sigma_{\mathbf{N}_0}^{(j)}$ 、 $\Sigma_{\mathbf{H}_0}^{(j)}$ 、及び  $\Sigma_{\mathbf{A}_0}^{(j)}$  を式 ( 1 3 ) のように設定する。

## 【 0 0 6 6 】

## 【 数 1 2 】

$$\Sigma_{\mathbf{N}_0}^{(j)} = \Sigma_{\mathbf{N}}, \Sigma_{\mathbf{H}_0}^{(j)} = \Sigma_{\mathbf{A}_0}^{(j)} = 0 \quad (13)$$

すなわち、各パーティクルにおける加法性雑音のベクトル  $\mathbf{N}_0^{(j)}$  の共分散行列を、初期分布  $p(\mathbf{N}_0)$  の共分散行列に設定し、各パーティクルにおける乗法性歪みの行列  $\mathbf{H}_0^{(j)}$  及び残響の行列  $\mathbf{A}_0^{(j)}$  の共分散行列の各要素を 0 に設定する。外乱初期分布推定部 2 2 2 は、式 ( 1 2 ) と式 ( 1 3 ) とに示す設定を、パーティクル  $j$  ( 1 j J ) ごとに行なう。

## 【 0 0 6 7 】

逐次計算部 2 2 4 は、GMM 1 3 0 の出力パラメータ 1 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 と、第  $t$  フレームにおける観測信号の特徴量 1 2 4 を受け、各パーティクルにおける外乱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それぞれ算出するための重み算出部 2 3 2 と、重み算出部 2 3 2 により算出された重みに基づき、パーティクルにおける外乱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パーティクルにおける外乱のパラメータと第  $t - 1$  フレームにおける各パーティクルにおける外乱のパラメータとに基づき、各パーティクルにおける外乱のパラメータを決定し、推定外乱分布 2 0 6 を生成するための推定外乱分布生成部 2 3 6 とを含む。

## 【 0 0 6 8 】

更新部 2 3 0 は、状態空間モデル 1 6 0 ( 図 4 ) を基に構成され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機能を持つ。拡張カルマンフィルタは、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。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 ( 1 4 ) ~ 式 ( 1 9 ) 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。

## 【 0 0 6 9 】

## 【 数 1 3 】

$$\Lambda_{t|t-1}^{(j)} = \hat{\Lambda}_{t-1}^{(j)} \quad (14)$$

$$\Sigma_{\Lambda_{t|t-1}}^{(j)} = \Sigma_{\Lambda_{t-1}}^{(j)} + \Sigma_{\mathbf{w}} \quad (15)$$

$$\mathbf{K}_t^{(j)} = \Sigma_{\Lambda_{t|t-1}}^{(j)} \mathbf{F}_t^{(j)T} \left[ \mathbf{F}_t^{(j)} \Sigma_{\Lambda_{t|t-1}}^{(j)} \mathbf{F}_t^{(j)T} + \Sigma_{\mathbf{S}, k_t^{(j)}} \right]^{-1} \quad (16)$$

$$\mathbf{F}_t^{(j)} = \partial f \left( \mathbf{S}_{k_t^{(j)}, t}^{(j)}, \mathbf{X}_{\mathbf{S}, t-1}, \Lambda_{t|t-1}^{(j)} \right) / \partial \Lambda_{t|t-1}^{(j)} \quad (17)$$

$$\hat{\Lambda}_t^{(j)} = \Lambda_{t|t-1}^{(j)} + \mathbf{K}_t^{(j)} \left( \mathbf{X}_t - f \left( \mathbf{S}_{k_t^{(j)}, t}^{(j)}, \mathbf{X}_{\mathbf{S}, t-1}, \Lambda_{t|t-1}^{(j)} \right) \right) \quad (18)$$

$$\Sigma_{\Lambda_t}^{(j)} = \Sigma_{\Lambda_{t|t-1}}^{(j)} - \mathbf{K}_t^{(j)} \mathbf{F}_t^{(j)} \Sigma_{\Lambda_{t|t-1}}^{(j)} \quad (19)$$

10

20

30

40

50

ただし、式(17)及び式(18)のベクトル  $S_{k_t}^{(j)}, t$  は、j番目のパーティクルにおいてGMM130(図2参照)の出力パラメータベクトル  $S_{k_t, t}$  に相当するパラメータである。また前述した通り、行列  $w$  は、第  $t-1$  フレームから第  $t$  フレームへの状態変化の際に外乱を表す行列  $\epsilon_t$  に生じる誤差の行列  $W_t$  の共分散行列を表す。

【0070】

GMMサンプリング部226は、GMM130(図2参照)内の混合分布から、要素分布である単一正規分布  $k_t^{(j)}$  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、サンプリングされた要素分布  $k_t^{(j)}$  から出力パラメータベクトル  $S_{k_t}^{(j)}, t$  を確率分布にしたがってサンプリングして、更新部230に与える。GMM130における要素分布  $k_t$  の混合重みを  $P_{S, st}^{(j)}, k_t$  とすると、要素分布  $k_t^{(j)}$  は、混合重み  $P_{S, st}^{(j)}, k_t$  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(20)に示すサンプリングによって得られる。

10

【0071】

【数14】

$$k_t^{(j)} \sim P_{S, k} \quad (20)$$

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  $S_{k_t}^{(j)}, t$  は、要素分布  $k_t^{(j)}$  から、次の式(21)に示すサンプリングによって得られる。

20

【0072】

【数15】

$$S_{k_t}^{(j)}, t \sim \mathcal{N} \left( \mu_{S, k_t}^{(j)}, \Sigma_{S, k_t}^{(j)} \right) \quad (21)$$

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。

【0073】

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  $X_t$  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  $S_{k_t}^{(j)}, t$ 、及び外乱のパラメータ行列  $\epsilon_t^{(j)}$  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、次の式(22)及び式(23)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。

30

【0074】

【数16】

$$w_t^{(j)} \propto w_{t-1}^{(j)} p(\mathbf{X}_t | \Lambda_t^{(j)}), \quad (22)$$

$$p(\mathbf{X}_t | \Lambda_t^{(j)}) = \mathcal{N} \left( \mathbf{X}_t; f \left( S_{k_t}^{(j)}, t, \mathbf{X}_{S, t-1}, \Lambda_t^{(j)} \right), \Sigma_{S, k_t}^{(j)} \right) \quad (23)$$

40

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。

【0075】

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外乱のパラメータの確率分布にしたがい、時刻  $t$  に対応する各パーティクルにおける外乱のパラメータ行列  $\epsilon_t^{(j)}$  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確率分布からは、パラメータの再サンプリングを行わない。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、パラメータを重み  $w_t^{(j)}$  の大きさに応じた回数の再サンプリン

50

グを行ない、得られた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、式（22）から分かるように観測された特徴量ベクトル $X_t$ の尤度に対応しているからである。

【0076】

推定外乱分布生成部236は、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の再生成を行なう機能を持つ。図7に、推定外乱分布生成部236の構成をブロック図で示す。図7を参照して、推定外乱分布生成部236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外乱の確率分布を用いて状態空間モデル160における外乱の確率分布を表し、当該表した確率分布に基づき、第 $t-1$ フレームに対応するパーティクルにおける外乱のパラメータを第 $t$ フレームに対応するものへ、上記の式（14）～式（19）に示す拡張カルマンフィルタを用いて再更新するための再更新部262と、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（22）及び式（23）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 $w_t^{(j)}$ 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パラメータと、再更新されたパーティクルにおける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

20

【0077】

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（24）にしたがって、許容確率を算出する機能を持つ。

【0078】

【数17】

$$\nu = \min \left\{ 1, w_t^{*(j)} / w_t^{(j)} \right\} \quad (24)$$

30

【0079】

パラメータ選択部270は、 $u$ が許容確率以下であれば、当該パーティクルにおける外乱のパラメータを再更新で得られた新たなパラメータに変更する機能を持つ。

【0080】

〔プログラム構造〕

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により実現可能である。図8に、本実施の形態に係る前処理部104に含まれる外乱成分抑圧部114が行なう外乱成分の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

【0081】

図8を参照して、外乱成分の抑圧処理が開始されると、ステップ302において、初期状態における外乱 $o_0$ 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（10）及び式（11）に示す算出方法により、式（9）に示す加法性雑音の初期分布 $p(N_0)$ のパラメータ平均ベクトル $\mu_N$ 及び共分散行列 $\Sigma_N$ を算出する。さらに、式（12）及び式（13）にしたがい、初期分布 $p(N_0)$ からパラメータベクトル $N_0^{(j)}$ （ $j = 1, \dots, J$ ）をサンプリングし、各パーティクルにおける加法性雑音の初期的なパラメータに推定する。またこの際、各パーティクルにおける乗法性雑音の初期的なパラメータ行列 $H_0^{(j)}$ 及び残響の初期的なパラメータ行列 $A_0^{(j)}$ についても、それぞれ式（12）及び式（13）にしたがい設定を行なう。

50

## 【 0 0 8 2 】

ステップ 3 0 4 では、外乱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。ステップ 3 0 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おける外乱を表す行列に対応する確率分布のパラメータを推定する。すなわち、各パーティクルにおける外乱のパラメータ行列  $\mathbf{\Lambda}_t^{(i)}$ 、及び行列  $\mathbf{\Lambda}_t^{(i)}$  の共分散行列を推定し、さらに、各パーティクルに対する重み  $w^{(i)}$  を定める。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。

## 【 0 0 8 3 】

ステップ 3 0 8 では、ステップ 3 0 6 でパーティクルごとに定めた外乱のパラメータ行列  $\mathbf{\Lambda}_t^{(i)}$  と、その共分散行列とを用いて、各パーティクルにおける観測信号の特徴量ベクトル  $\mathbf{X}_t$  ( 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確率モデルの平均ベクトル  $\mu_{\mathbf{X}_{k,t}}^{(j)}$  と、共分散行列  $\Sigma_{\mathbf{X}_{k,t}}^{(j)}$  とを算出する。

## 【 0 0 8 4 】

ステップ 3 1 0 では、M M S E 推定法により、第  $t$  フレームにおけるクリーン音声の特徴量を推定する。すなわちまず、ステップ 3 0 6 及びステップ 3 0 8 の処理で得られたパラメータを用いて、M M S E 推定法によって、M M S E 推定値ベクトル  $\hat{\mathbf{S}}_t$  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 0 8 5 】

## 【 数 1 8 】

$$\hat{\mathbf{S}}_t = \mathbf{X}_t - \sum_{j=1}^J w_t^{(j)} \sum_{k=1}^K P(k|\mathbf{X}_t, (j)) \left( \mu_{\mathbf{X}_{k,t}}^{(j)} - \mu_{\mathbf{S},k} \right), \quad (25)$$

$$\mu_{\mathbf{X}_{k,t}}^{(j)} = \left( \mathbf{S}_{k,t}, \mathbf{X}_{\mathbf{S},t-1}, \mathbf{\Lambda}_t^{(j)} \right) \quad (26)$$

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、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  $P(k|\mathbf{X}_t, (j))$  は、次の数式により算出される。

## 【 0 0 8 6 】

## 【 数 1 9 】

$$P(k|\mathbf{X}_t, (j)) = \frac{P_{\mathbf{S},k} \mathcal{N} \left( \mathbf{X}_t, \mu_{\mathbf{X}_{k,t}}^{(j)}, \Sigma_{\mathbf{X}_{k,t}}^{(j)} \right)}{\sum_{k'=1}^K P_{\mathbf{S},k'} \mathcal{N} \left( \mathbf{X}_t, \mu_{\mathbf{X}_{k',t}}^{(j)}, \Sigma_{\mathbf{X}_{k',t}}^{(j)} \right)}, \quad (27)$$

## 【 0 0 8 7 】

ステップ 3 1 2 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外乱成分の抑圧処理を終了する。さもなければステップ 3 0 4 に戻る。

## 【 0 0 8 8 】

図 9 に、ステップ 3 0 6 ( 図 8 参照 ) において行なわれる外乱確率分布の推定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、外乱確率分布の推定処理が開始されると、ステップ 3 2 2 において、式 ( 1 4 ) ~ 式 ( 1 9 ) により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外乱確率分布から、第  $t$  フレームのパーティクルにおける外乱確率分布を推定する。

## 【 0 0 8 9 】

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 ( 2 2 ) 及び式 ( 2 3 ) によって算出し、正規化する。ステップ 3 2 6 では、各パーティクルに与えられた重み  $w_t^{(j)}$  に基づき、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外乱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 0 9 0 】

図 1 0 にステップ 3 2 8 ( 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 1 0 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 2 において、ステップ 3 2 6 ( 図 9 参照 ) での再サンプリングで得られたパーティクルにおけるパラメータを用いて、外乱確率分布の再更新を行なう。すなわち、時刻  $t$  の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 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 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 9 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 0 9 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 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 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破棄し、処理を終了する。

## 【 0 0 9 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。まず、図 6 に示す外乱確率分布推定部 2 0 0 による初期状態における外乱の確率分布の推定動作を説明する。図 1 に示す計測部 1 1 2 が、音源 1 0 2 から観測音 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、外乱成分抑圧部 1 1 4 の図 5 に示す外乱確率分布推定部 2 0 0 に与えられる。図 6 を参照して、外乱確率分布推定部 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、外乱初期分布推定部 2 2 2 に与える。外乱初期分布推定部 2 2 2 は、上記の式 ( 9 ) ~ 式 ( 1 1 ) に示す処理により加法性雑音の初期分布  $p ( N_0 )$  を推定する。さらに、雑音の初期分布  $p ( N_0 )$  から、上記の式 ( 1 2 ) 及び式 ( 1 3 ) 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ベクトル  $N_0^{(j)}$  及び共分散行列  $\Sigma_{N_0}^{(j)}$  が決定される。乗法性歪みの初期パラメータ行列  $H_0^{(j)}$  及びその共分散行列  $\Sigma_{H_0}^{(j)}$  をともに 0 に設定し、残響の初期パラメータ行列  $A_0^{(j)}$  及びその共分散行列  $\Sigma_{A_0}^{(j)}$  をともに 0 に設定する。外乱確率分布推定部 2 0 0 は、これらのパラメータを、時刻  $t = 0$  のフレームにおける推定外乱分布 2 0 6 のパラメータとして出力する。

## 【 0 0 9 3 】

次に、外乱確率分布推定部 2 0 0 による、第  $t$  フレーム (  $t - 1$  ) における推定外乱分布 2 0 6 の推定動作を説明する。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、観測信号の特徴量  $X_t ( 1 2 4 )$  を更新部 2 3 0

10

20

30

40

50

に与えるとともに、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。更新部230は、これに回答して、第t-1フレームの各パーティクルにおける推定外乱分布206のパラメータを取得する。

【0094】

GMMサンプリング部226は、GMM130から、出力パラメータベクトル $S_{k_t}^{(j)}(i)_t$ のサンプリングを行なう。図11に、出力パラメータベクトル $S_{k_t}^{(j)}(i)_t$ のサンプリングの概要を模式的に示す。例えば、j番目のパーティクルにおいて、GMM130内の混合正規分布400の中から、混合重みにしたがった確率で要素分布 $k_t^{(j)}(402)$ をサンプリングする。GMMサンプリング部226はさらに、要素分布 $k_t^{(j)}(402)$ 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 $S_{k_t}^{(j)}(i)_t(404)$ 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 $S_{k_t}^{(j)}(i)_t$ 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230に与える。

10

【0095】

図12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12においては、ある外乱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12においては、パーティクルを、白抜きの丸印、及び黒塗りの丸印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。

20

【0096】

図12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、式(14)~式(19)により示す拡張カルマンフィルタを用いて、状態空間420内の各パーティクルにおける外乱分布のパラメータ行列 $\hat{\Sigma}_{t-1}^{(j)}$ を、第tフレームに対応する推定外乱分布のパラメータ行列 $\hat{\Sigma}_t^{(j)}$ に更新する。これにより、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。

【0097】

続いて重み算出部232は、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外乱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、 $w_t^{(j)}$ 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外乱の確率分布に基づき、設定された回数ずつ、外乱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。

30

40

【0098】

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り返行なわれると、あるフレームに対応するパーティクルの多くにおける外乱のパラメータが、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外乱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、推定外乱分布生成部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における推定外乱分布にしたがい、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外乱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

50

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、許容確率を算出する。パラメータ選択部270は、許容確率と、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破棄する。

#### 【0099】

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推定外乱分布206のパラメータベクトル $N_t^{(j)} H_t^{(j)}$ 及び行列 $A_t^{(j)}$ 、並びに共分散行列 $N_t^{(j)} H_t^{(j)}$ 及び $A_t^{(j)}$ が推定される。外乱確率分布推定部200は、各パーティクルにおける推定外乱分布206のパラメータベクトル $N_t^{(j)} H_t^{(j)}$ 及び行列 $A_t^{(j)}$ 、並びに共分散行列 $N_t^{(j)} H_t^{(j)}$ 及び $A_t^{(j)}$ と、各パーティクルに対する重み $w_t^{(j)}$ と、観測信号の特徴量ベクトル $X_t$ とを、フレームごとに、図5に示すパラメータ生成部202に与える。

#### 【0100】

図5を参照して、パラメータ生成部202は、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の確率モデルの平均ベクトル及び共分散行列(208)を生成する。これにより、各パーティクルにおいて外乱の確率分布と、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、MMSE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、クリーン音声のMMSE推定値ベクトル $\hat{S}_t^{(j)}$ を算出する。さらに、MMSE推定値ベクトル $\hat{S}_t^{(j)}$ と重み $w_t^{(j)}$ とを用いて、時刻 $t$ におけるクリーン音声の推定特徴量ベクトル $\hat{S}_t$ を算出し、図1に示す探索部110に出力する。

#### 【0101】

図1に示す探索部110は、クリーン音声の推定特徴量ベクトル $\hat{S}_t$ 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。

#### 【0102】

##### 【実験】

本実施の形態に係る音声認識システム100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。

#### 【0103】

本実験においては、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、残響のインパルス応答を畳み込み、さらに加法性雑音を人工加算して、観測信号を生成した。残響のインパルス応答には、実環境音声・音響データベースに収録されている、残響時間0.3秒及び1.3秒のインパルス応答を用いた。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20dB、15dB、及び10dBのSNR(Signal-to-Noise Ratio)でそれぞれ雑音を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。

#### 【0104】

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外乱成分の抑圧処理の方法を含む次の5種の処理方法で、探索に用いる特徴量を生成した。すなわち、HTK Baselineすなわち外乱抑圧処理を行わない観測信号の特徴量(Baseline)、ETSI(European Telecommunications Standards Institute)により勧告されているETSI Advanced front-end(ES202)によ

10

20

30

40

50

る雑音抑圧処理を施した特徴量 ( E T S I )、従来の M M S E 推定により得られる推定特徴量 ( M M S E )、非特許文献 5 に記載の手法での処理により得られる推定特徴量 ( E M )、及びパーティクルフィルタを用いた本実施の形態に係る外乱成分の抑圧処理 ( P r o p o s e d ) により得られる推定特徴量である。

【 0 1 0 5 】

パーティクルフィルタを用いた外乱成分の抑圧処理を行なう際の、G M M 1 3 0 ( 図 2 参照 ) には、混合分布数 5 1 2 のモデルを用いた。この処理においては、誤差ベクトル  $W_t$  の共分散行列を、 $W_N = W_H = W_A = \text{diag}(0.01)$  に設定した。また、処理に用いるパーティクルの総数  $J$  を 2 0 に設定した。

【 0 1 0 6 】

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、3 9 次 M F C C ( Mel Frequency Cepstrum Coefficient ) ( 1 2 次 M F C C + C 0 + ) を用いた。また、図 1 に示す認識用音響モデル 1 0 9 には、1 6 状態 2 0 混合分布の H M M を用いた。

【 0 1 0 7 】

この認識実験における処理に、クロック周波数 3 . 2 ギガヘルツ、3 2 ビットの市販の C P U ( Central Processing Unit ) を用いた場合、処理に要した時間は、観測信号における実時間の 0 . 8 倍であった。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。

【 0 1 0 8 】

表 1 ~ 表 4 に、各試料に対する認識実験で得られた認識精度を、上記の処理方法別に示す。

【 0 1 0 9 】

10

20

【表 1】

表 1: 認識結果 (%) (工場雑音, 残響時間 0.3 秒)

SNR	Baseline	ETSI	MMSE	EM	Proposed
Clean	98.00	97.76	97.02	98.00	<b>98.43</b>
20 dB	86.92	92.32	92.29	93.95	<b>94.11</b>
15 dB	71.94	<b>87.53</b>	85.48	87.01	<b>87.07</b>
10 dB	48.27	<b>79.00</b>	73.32	77.62	<b>74.67</b>
Ave.	76.28	<b>89.15</b>	87.03	<b>89.15</b>	<b>88.57</b>

10

表 2: 認識結果 (%) (道路工事雑音, 残響時間 0.3 秒)

SNR	Baseline	ETSI	MMSE	EM	Proposed
Clean	98.00	97.76	97.02	98.00	<b>98.43</b>
20 dB	94.17	97.18	96.44	96.75	<b>98.62</b>
15 dB	88.76	95.70	96.25	96.32	<b>97.33</b>
10 dB	75.56	93.34	93.52	<b>94.47</b>	<b>93.55</b>
Ave.	89.12	96.00	95.81	96.39	<b>96.98</b>

20

表 3: 認識結果 (%) (工場雑音, 残響時間 1.3 秒)

SNR	Baseline	ETSI	MMSE	EM	Proposed
Clean	61.44	60.45	59.63	63.74	<b>70.77</b>
20 dB	39.27	52.44	54.68	58.27	<b>63.31</b>
15 dB	27.60	47.07	48.51	53.67	<b>56.40</b>
10 dB	20.20	37.27	41.08	44.89	<b>45.66</b>
Ave.	37.13	49.31	50.98	55.14	<b>59.04</b>

30

表 4: 認識結果 (%) (道路工事雑音, 残響時間 1.3 秒)

SNR	Baseline	ETSI	MMSE	EM	Proposed
Clean	61.44	60.45	59.63	63.74	<b>70.77</b>
20 dB	51.27	57.72	60.64	63.95	<b>68.10</b>
15 dB	42.34	55.79	58.21	62.36	<b>67.76</b>
10 dB	30.61	50.38	55.82	58.34	<b>64.14</b>
Ave.	46.42	56.09	58.58	62.10	<b>67.69</b>

40

表 1 ~ 表 4 を参照して、パーティクルフィルタによる抑圧処理 (Proposed) を行なうことで、良好な単語認識精度が得られることが分かる。特に、残響 1.3 秒の観測信号においては、パーティクルフィルタによる抑圧処理 (Proposed) により、高い単語認識精度が得られることが分かる。

【 0 1 1 0 】

以上の実験結果から、本実施の形態の外乱成分の抑圧処理により、非定常な加法性雑音

50

及び残響による歪みを受ける環境下での音声認識性能が改善され、かつ実時間処理が可能になることが明らかとなった。

【 0 1 1 1 】

[ 変形例等 ]

なお、本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外乱成分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、認識精度が向上することが期待される。

【 0 1 1 2 】

また、本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 ( 2 0 ) 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。

【 0 1 1 3 】

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。

【 図面の簡単な説明 】

【 0 1 1 4 】

【 図 1 】 本発明の一実施の形態に係る音声認識システム 1 0 0 の構成を示す概略図である。

【 図 2 】 GMM 1 3 0 の概念を示す概略図である。

【 図 3 】 外乱要因 1 1 8 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 1 6 0 の概念を示す概略図である。

【 図 5 】 外乱成分抑圧部 1 1 4 の構成を示すブロック図である。

【 図 6 】 外乱確率分布推定部 2 0 0 の構成を示すブロック図である。

【 図 7 】 推定外乱分布生成部 2 3 6 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 】 外乱確率分布推定処理の制御構造を示すフローチャートである。

【 図 1 0 】 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 1 1 】 GMM 1 3 0 からパラメータをサンプリングする動作の概要を示す図である。

【 図 1 2 】 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。

【 符号の説明 】

【 0 1 1 5 】

1 0 0 音声認識システム

1 0 2 音源

1 0 4 前処理部

1 0 6 前処理用音響モデル部

1 0 8 言語モデル部

1 0 9 認識用音響モデル部

1 1 0 探索部

1 1 2 計測部

1 1 4 外乱分布抑圧部

1 1 6 話者

1 1 8 外乱要因

1 2 0 クリーン音声

1 2 2 観測音

10

20

30

40

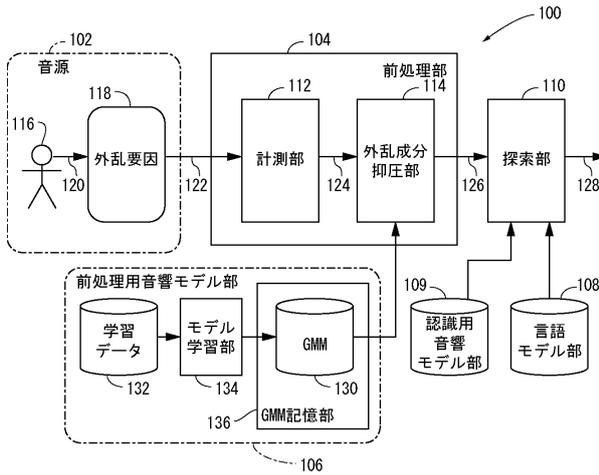
50

- 1 2 4 観測信号の特徴量
- 1 2 6 推定クリーン音声の特徴量
- 1 3 0 G M M
- 1 3 2 学習データ記憶部
- 1 3 4 モデル学習部
- 1 3 6 G M M記憶部
- 1 6 0 状態空間モデル
- 2 0 0 外乱確率分布推定部
- 2 0 2 パラメータ生成部
- 2 0 4 クリーン音声推定部
- 2 2 0 フレーム選択部
- 2 2 2 外乱初期分布推定部
- 2 2 4 逐次計算部
- 2 2 6 G M M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 推定外乱分布生成部
- 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

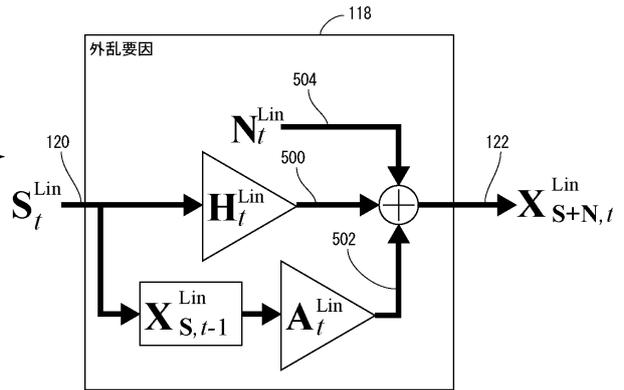
10

20

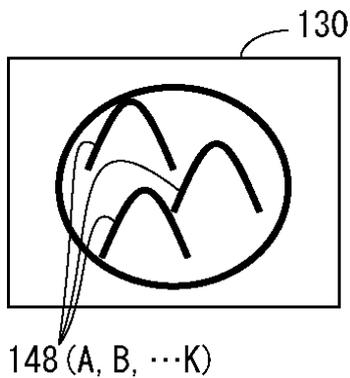
【図1】



【図3】

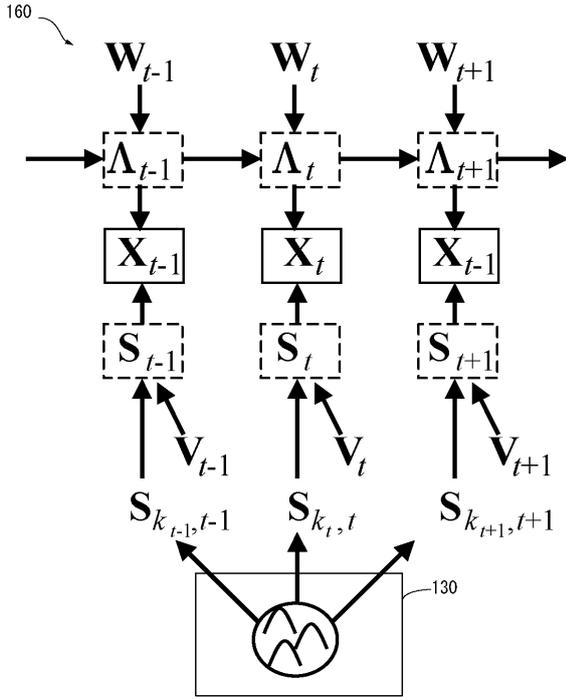


【図2】

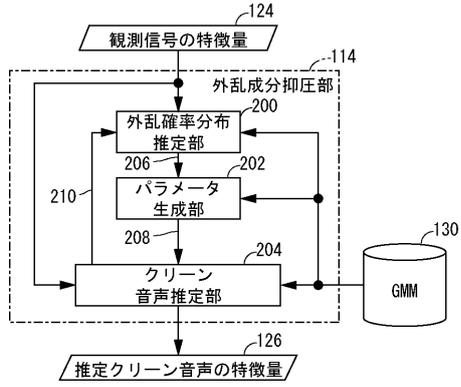


148 (A, B, ...K)

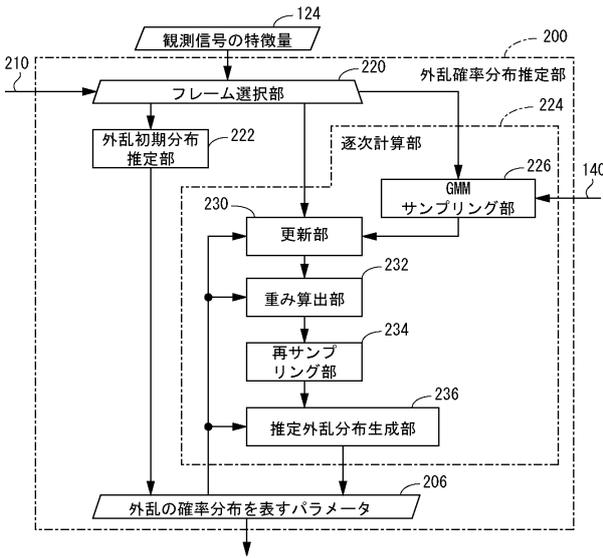
【図4】



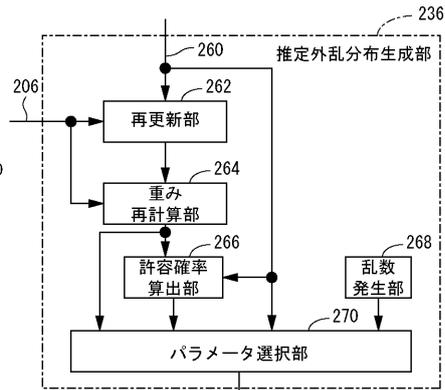
【図5】



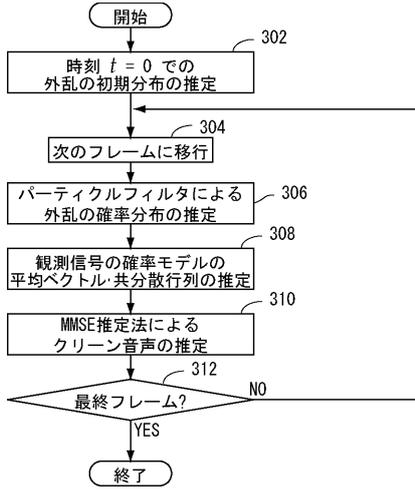
【図6】



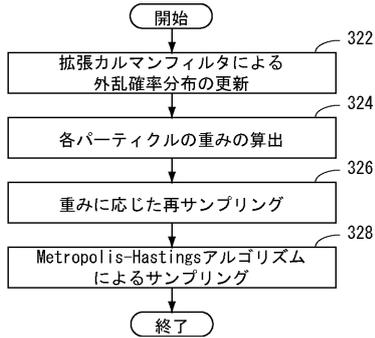
【図7】



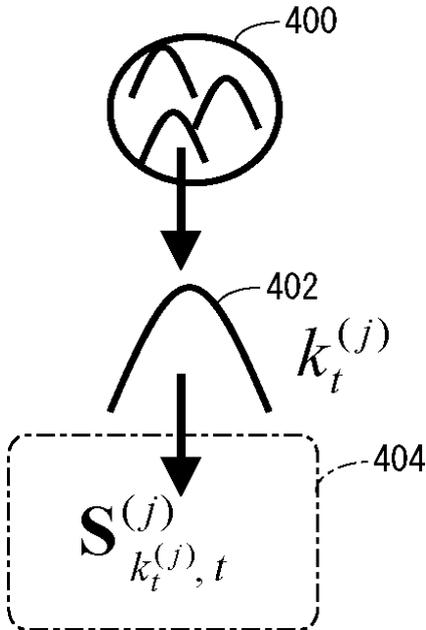
【図 8】



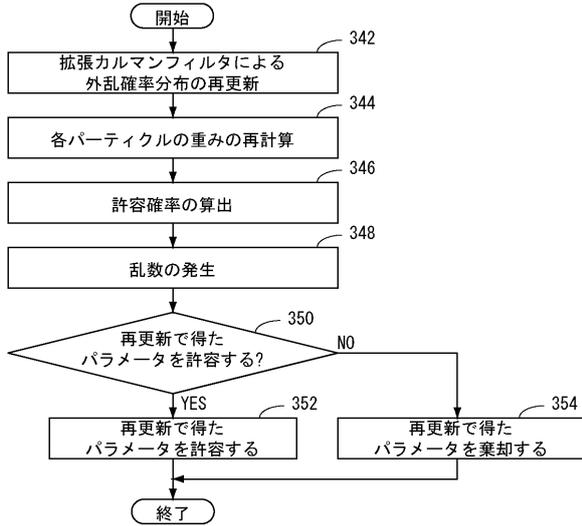
【図 9】



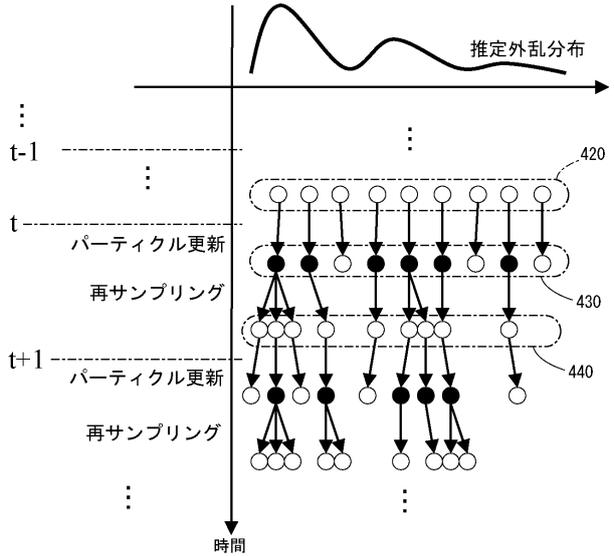
【図 11】



【図 10】



【図 12】



## フロントページの続き

- (56)参考文献 特開2004-347761(JP,A)  
特開2002-251198(JP,A)  
パーティクルフィルタを用いた非定常雑音下での音声認識, 日本音響学会講演論文集, 2004年 9月21日  
パーティクルフィルタに基づく雑音抑圧法を用いた定常雑音および残響雑音下での音声認識, 日本音響学会講演論文集, 2005年 3月 8日  
GMMに基づく音声信号推定法を用いた雑音下音声認識, 情報処理学会研究報告, 2002年12月20日  
山本 仁 Hitoshi YAMAMOTO, モデル合成法を用いた複数フレームにまたがる残響下の音声認識 Recognition of speech with reverberation spanning over multiple frames, 日本音響学会2003年秋季研究発表会講演論文集 - I - THE 2003 AUTUMN MEETING OF THE ACOUSTICAL SOCIETY OF JAPAN, 日本, 社団法人日本音響学会  
Tetsuya Takiguchi, Masafumi Nishimura, ACOUSTIC MODEL ADAPTATION USING FIRST ORDER PREDICTION FOR REVERBERANT SPEECH, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004, 2004年, vol.1, p.869-872

## (58)調査した分野(Int.Cl., DB名)

G10L 11/00 - 21/06  
IEEE Xplore  
JSTPlus(JDreamII)  
JST7580(JDreamII)