2005－SLP－57（10）
2005／7／15

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

藤本 雅清　中村 哲

ATR 音声言語コミュニケーション研究所
〒 619-0288 京都府相楽郡精華町光台 2-2-2
Tel: 0774-95-1301　　Fax: 0774-95-1308
E-mail: {masakiyo.fujimoto, satoshi.nakamura}@atr.jp

あらまし 本研究では，パーティクルフィルタを用いた非定常雑音の逐次推定法及び抑圧法を提案し，従来法と比較して非定常雑音下での音声認識性能改善に効果的である事を示す．提案手法において，非定常雑音は再サンプリング法を含むパーティクルフィルタ及びマルコフ連鎖モンテカルロ法を通じて逐次推定され，クリーン音声は推定された非定常雑音を MMSE 推定に基づく雑音抑圧法に適用することにより得られる．また，パーティクルフィルタで用いる状態空間モデルに Polyak averaging と feedback を導入することにより音声認識性能を大幅に改善できることを示す．

キーワード: 雑音下音声認識，非定常雑音，逐次推定，パーティクルフィルタ，Polyak averaging and feedback

A Non-stationary Noise Suppression Method Based on Particle Filtering and Polyak Averaging

Masakiyo Fujimoto　Satoshi Nakamura

ATR Spoken Language Communication Research Laboratories
2-2-2, Hikaridai, Seika-cho, Souraku-gun, Kyoto, 619-0288 Japan
Tel: 0774-95-1301　　Fax: 0774-95-1308
E-mail: {masakiyo.fujimoto, satoshi.nakamura}@atr.jp

Abstract This paper addresses a speech recognition problem in non-stationary noise environments, especially, the estimation of noise sequences. To solve this problem, we present a particle filter-based sequential noise estimation method for front-end processing of speech recognition in noise. In the proposed method, a noise sequence is estimated by three steps, a sequential importance sampling step, a residual resampling step, and finally a Markov chain Monte Carlo step with Metropolis-Hastings sampling. The estimated noise sequence is used in the MMSE-based clean speech estimation. We also introduce a Polyak averaging and feedback into state transition process used for particle filtering. In the evaluation results, we observed that the proposed method improves speech recognition accuracy in non-stationary noise environments results by the noise compensation method with stationary noise assumptions.

Keywords: noisy speech recognition, non-stationary noise, sequential estimation, particle filter, Polyak averaging and feedback

1　はじめに

雑音下での音声認識性能の改善は，音声認識技術に課せられた重要な問題の一つである．この問題において，時間的に定常的な性質を持つ雑音に限定した環境下では，様々な研究成果が報告されており，高い技術水準に達して成功を納めたと言える [1]-[4]．しかし，実環境で観測される雑音の多くは，時間的に変動する非定常的な性質を持っており，実環境下における頑健な音声認識技術を確立するためには，このような非定常雑音への対処が必要不可欠となる．

一般に，非定常雑音下での音声認識性能を改善するためには，雑音の時間変動を正確に逐次推定する必要がある．しかし，音声認識を行う際に観測できる信号は通常，雑音が重畳した音声のみであり，クリーン音声のみならず雑音もが非定常的な性質を持つ場合，各信号がそれぞれどのような時間推移を行うかを推定することは困難な問題である．この問題に関して，逐次 EM アルゴリズムを用いた非定常雑音の逐次推定法が提案され，その有効性が報告されている [6]-[8]．しかし逐次 EM アルゴリズムでは，1 フレーム毎に EM アルゴリズムの繰り返し推定を行ってパラメータを収束させる必要があるため，計算量が膨大なものとなり実時間処理に向かないという問題がある．よって実用面の

観点から，高速かつ高精度な逐次推定法が望まれる．

この問題に関して，近年，ベイズ推定法の一種であるパーティクルフィルタ [9, 10] に基づく逐次推定法が注目されており，様々な研究分野で応用されている．パーティクルフィルタは逐次モンテカルロ法に基づく手法であり，逐次 EM アルゴリズムのような繰り返し推定法を必要としないので計算量が少なく，実時間向け処理に適した手法であるという利点がある．以上の点を踏まえて，本研究では，パーティクルフィルタによる非定常雑音の逐次推定について検討する．また，パーティクルフィルタ推定された雑音を MMSE (Minimum Mean Square Error) 推定に基づく雑音抑圧法 [3, 4] に適用し，非定常雑音下での音声認識精度の改善に効果があることを示す．

パーティクルフィルタを適用する際には，状態空間モデル（動的モデル）と呼ばれる信号モデルを定義する必要がある．一般に，状態空間モデルは目的信号の時間（状態）遷移過程を表現した状態方程式と，観測信号の生成過程を表現した観測方程式から構成されている．我々は以前，状態方程式に Random walk 過程を用いた非定常雑音の逐次推定法 [11] を提案したが，Random walk 過程は信号の時間遷移をランダム雑音により規定するので，推定結果が不安定なものになるという問題があった．この問題に関しては，状態方程式に Polyak averaging と feedback [8, 12, 13] を導入することにより，安定した推定結果を得ることができ，音声認識性能の大幅な改善効果があることを示す．

本研究と同様の研究は Yao らによっても行われているが，Yao らの手法は音声認識時の音響モデルに対する雑音処理であり，推定された雑音と HMM 合成法を用いて音響モデルを逐次更新し，非定常雑音が重畳した音声を認識している [10]．一方，本研究は音声認識の前処理部（特徴抽出）に対する雑音処理であるため，雑音抑圧後のデータを用いた音響モデル適応など，複合処理が可能という利点がある．

2 パーティクルフィルタによる非定常雑音の抑圧

図 1 は，提案する非定常雑音の抑圧法の概要を示しており，パーティクルフィルタに基づく非定常雑音の推定部 [11] と MMSE 推定に基づく雑音抑圧部 [3, 4] の二つに大きく分かれている．

非定常雑音の推定部において，本研究で用いるパーティクルフィルタは，

(1) 拡張カルマンフィルタによるパラメータ更新
(2) サンプル重みの計算
(3) 再サンプリング
(4) マルコフ連鎖モンテカルロ（Metropolis-Hastings 法 [14]）によるサンプリング

の 4 つの要素技術により構成されている．以下，各部の詳細について述べる．

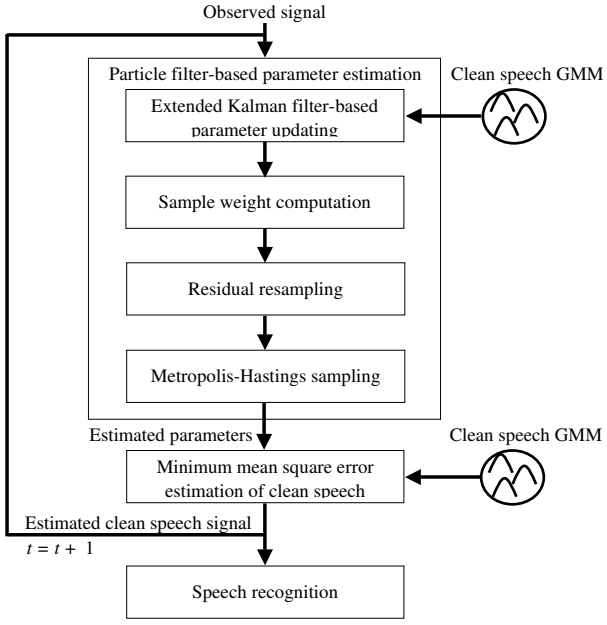


図 1: 提案手法の概要

2.1 状態空間モデルの定義

まずパーティクルフィルタを適用する際に必要となる，状態空間モデル（動的モデル）を定義する．一般に，状態空間モデルは目的信号の時間（状態）遷移過程を表現した状態方程式と，観測信号の生成過程を表現した観測方程式から構成されている．

t 番目の短時間フレームにおいて，雑音重畳音声，クリーン音声，雑音の対数メルスペクトルを要素に持つベクトルをそれぞれ $\mathbf{X}_t$，$\mathbf{S}_t$，$\mathbf{N}_t$ と定義する．次に，クリーン音声 $\mathbf{S}_t$ の背後に確率モデルとして，GMM (Gaussian Mixture Model) が存在し，時刻 t において，GMM 内のある要素分布 k_t (平均 $\mu_{\mathbf{S},k_t}$，分散 $\mathbf{\Sigma}_{\mathbf{S},k_t}$) から，パラメータ $\mathbf{S}_{k_t,t}$ が出力されると仮定する．また，$\mathbf{N}_t$ を用いて $\mathbf{X}_t$ が誤差 $\mathbf{V}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{S},k_t})$ を伴い，次式のように表現されるとする．

$$\begin{aligned}\mathbf{X}_t &= \mathbf{S}_{k_t,t} + \log\left(\mathbf{I} + \exp\left(\mathbf{N}_t - \mathbf{S}_{k_t,t}\right)\right) + \mathbf{V}_t \\ &= \mathbf{f}\left(\mathbf{S}_{k_t,t}, \mathbf{N}_t\right) + \mathbf{V}_t \end{aligned} \quad (1)$$

一方，$\mathbf{N}_t$ の時間推移が以下のような，ランダムガウス雑音 $\mathbf{W}_t$ (平均 $\mathbf{0}$, 分散 $\mathbf{\Sigma}_{\mathbf{W}}$) を伴う Random walk 過程により表現できるものと仮定する．

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

本研究では，式 (1) を観測方程式，式 (2) を状態方程式として状態空間モデルを構成する．

2.2 パーティクルフィルタの定義 (Sequential Important Sampling)

式 (1), (2) で定義された状態空間モデルが与えられ, $\mathbf{N}_{0:t} = \{\mathbf{N}_0, \ldots, \mathbf{N}_t\}$ とすると，$\mathbf{X}_t$ が観測されたときの $\mathbf{N}_{0:t}$ の事後確率分布は，マルコフ連鎖を用いて次式のように表され,

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

式 (3) を最大にするような信号列 $\mathbf{N}_{0:t}$ を推定する問題に帰着する．パーティクルフィルタでは，時刻 t の事後確率分布を次式のようなモンテカルロサンプリングにより近似する．

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

上式において，j はサンプル番号，J はサンプルの総数, $\delta(\cdot)$ は Dirac-delta 関数，$w_t^{(j)}$ は各時刻におけるサンプル j の重みであり ($\sum_{j=1}^{N} w_t^{(j)} = 1$)，$w_t^{(j)}$ は次式により与えられる．

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

$q(\mathbf{N}_{0:t}^{(i)}|\mathbf{X}_{0:t})$ は，サンプル $\mathbf{N}_{0:t}^{(i)}$ を出力する確率分布であり，以下の連鎖モデルで表現されるものとする．

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

また，式 (3) の事後確率分布は，ベイズ則により次式のように表されるため,

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

式 (6)，(7) より，$w_t^{(j)}$ は次式により与えられる．

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

ここで，$p(\mathbf{N}_t^{(j)}|\mathbf{N}_{t-1}^{(j)}) = q(\mathbf{N}_t^{(j)}|\mathbf{N}_{0:t-1}^{(i)}, \mathbf{X}_{0:t})$ と仮定することにより，次式が得られる．

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

$p(\mathbf{X}_t|\mathbf{N}_t^{(j)})$ は，次式のような確率密度関数であり,

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

$\mathbf{N}_t^{(j)}$ は，式 (1)，(2) を状態空間モデルとする拡張カルマンフィルタにより過去の値から更新される．

一般に，以上に述べたパーティクルフィルタは，Sequential Importance Sampling (SIS) アルゴリズムと呼ばれている [9].

2.3 拡張カルマンフィルタによる更新

確率密度関数 $p(\mathbf{N}_{0:t}^{(j)}|\mathbf{X}_{0:t})$ のパラメータ（平均ベクトル $\hat{\mathbf{N}}_t^{(j)}$，共分散行列 $\mathbf{\Sigma}_{\mathbf{N}t}^{(j)}$) は，以下のような拡張カルマンフィルタを用いて，過去のパラメータより更新される．

$$\mathbf{N}_{t|t-1}^{(j)} = \hat{\mathbf{N}}_{t-1}^{(j)} \tag{11}$$

$$\mathbf{\Sigma}_{\mathbf{N}t|t-1}^{(j)} = \mathbf{\Sigma}_{\mathbf{N}t-1}^{(j)} + \mathbf{\Sigma}_{\mathbf{W}} \tag{12}$$

$$\mathbf{K}_t^{(j)} = \mathbf{\Sigma}_{\mathbf{N}t|t-1}^{(j)} \mathbf{F}_t^{(j)T} \left[\mathbf{F}_t^{(j)} \mathbf{\Sigma}_{\mathbf{N}t|t-1}^{(j)} \mathbf{F}_t^{(j)T} + \mathbf{\Sigma}_{\mathbf{S},k_t^{(j)}}\right]^{-1} \tag{13}$$

$$\mathbf{F}_t^{(j)} = \partial \mathbf{f} \left(\mathbf{S}_{k_t^{(j)},t}^{(j)}, \mathbf{N}_{t|t-1}^{(j)}\right) \Big/ \partial \mathbf{N}_{t|t-1}^{(j)} \tag{14}$$

$$\hat{\mathbf{N}}_t^{(j)} = \mathbf{N}_{t|t-1}^{(j)} + \mathbf{K}_t^{(j)} \left(\mathbf{X}_t - \mathbf{f}\left(\mathbf{S}_{k_t^{(j)},t}^{(j)}, \mathbf{N}_{t|t-1}^{(j)}\right)\right) \tag{15}$$

$$\mathbf{\Sigma}_{\mathbf{N}t}^{(j)} = \mathbf{\Sigma}_{\mathbf{N}t|t-1}^{(j)} - \mathbf{K}_t^{(j)} \mathbf{F}_t^{(j)} \mathbf{\Sigma}_{\mathbf{N}t|t-1}^{(j)} \tag{16}$$

上式において，$t|t-1$ は，フレーム $t-1$ から予測されたパラメータを示し，$\mathbf{S}_{k_t^{(j)},t}^{(j)}$ は，式 (17)，(18) を用いてクリーン音声の GMM からサンプリングされたパラメータである．

$$\mathbf{S}_{k_t^{(j)},t}^{(j)} \sim \mathcal{N}\left(\mu_{\mathbf{S},k_t^{(j)}}, \mathbf{\Sigma}_{\mathbf{S},k_t^{(j)}}\right) \tag{17}$$

$$k_t^{(j)} \sim P_{\mathbf{S},k_t} \tag{18}$$

上式において，$P_{\mathbf{S},k_t}$ は，GMM の混合重みである．また，$p(\mathbf{N}_{0:t}^{(j)}|\mathbf{X}_{0:t})$ の初期パラメータは,

$$\mathbf{N}_0^{(j)} \sim \mathcal{N}\left(\mu_{\mathbf{N}}, \mathbf{\Sigma}_{\mathbf{N}}\right) \tag{19}$$

$$\mathbf{\Sigma}_{\mathbf{N}0}^{(j)} = \mathbf{\Sigma}_{\mathbf{N}} \tag{20}$$

としてサンプリングし，$\mu_{\mathbf{N}}$，$\mathbf{\Sigma}_{\mathbf{N}}$ は $\mathbf{X}_t$ の最初の 10 フレームを雑音のみが存在する区間とみなして推定する．

2.4 再サンプリング (Residual resampling)

重み $w_t^{(j)}$ は，各サンプルごとに割り当てられるが，$w_t^{(j)}$ の値が微小であるサンプルは，事後確率分布を近似するサンプルとして相応しくない．よって図 2 に示すように微小な $w_t^{(j)}$ を持つサンプルを破棄する．また，大きな $w_t^{(j)}$ を持つサンプルを幾つかの同じ値を持つ子サンプルに分割して，親サンプルに割り当てることにより，サンプルの総数を維持する (Residual re-sampling) [9]．これは，大きな $w_t^{(j)}$ を持つサンプルを重要サンプルと見なし，そのような重点的に利用することを意味している．なお，割り当てられる子サンプルの数は $w_t^{(j)}$ の値に依存する．

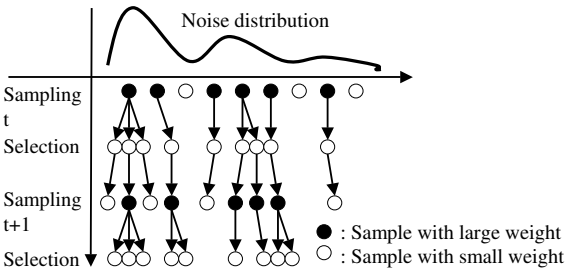


図 2: 再サンプリングの概念図

2.5 マルコフ連鎖モンテカルロ

再サンプリングによる重要サンプルの割り当てにおいて，時間が経過するに従い，1 つの親サンプルに多くの子サンプルが割り当てられる場合がある．最悪の場合，1 つの親サンプルに全ての同じ子サンプルが割り当てられ，分布の近似精度が低下する．この問題において，本研究では，マルコフ連鎖モンテカルロ法の Metropolis-Hastings サンプリング [14] を用いて新たな子サンプルを生成し，1 つの親サンプルに全ての同じ子サンプルが割り当てられるような状況を回避している．

Metropolis-Hastings サンプリングではまず，サンプル j のパラメータセットを $\mathbf{\Psi}_t^{(j)} = \left(w_t^{(j)}, \hat{\mathbf{N}}_t^{(j)}, \mathbf{\Sigma}_{\mathbf{N}t}^{(j)}\right)$ と定義し，同一の親サンプルから新たなパラメータセット $\mathbf{\Psi}_t^{*(j)}$ を発生させる．次に，次式により定義される許容確率 ν と一様乱数 $u \sim U_{[0,1]}$ を発生させる（$U_{[0,1]}$ は 0〜1 の範囲の一様分布）．

$$\nu = \min\left\{1, w_t^{*(j)}/w_t^{(j)}\right\} \tag{21}$$

その後，次式のように ν と u を比較して，$\mathbf{\Psi}_t^{*(j)}$ を受理するか否かを決定する．

$$\mathbf{\Psi}_t^{(j)} = \begin{cases} \mathbf{\Psi}_t^{*(j)} & \text{if } u \le \nu \ \text{(accept state transition)} \\ \mathbf{\Psi}_t^{(j)} & \text{otherwise (reject state transition)} \end{cases} \tag{22}$$

2.6 MMSE 推定による雑音抑圧

以上に述べた手法で推定された雑音の確率分布を用いて，MMSE 推定法 [3] に基づくクリーン音声の推定（雑音抑圧）を行う．パーティクルフィルタにより得られた 1 サンプルのパラメータセット $\mathbf{\Psi}_t^{(j)}$ を用いた MMSE 推定結果は次式により得られる．

$$\hat{\mathbf{S}}_t^{(j)} = \mathbf{X}_t - \sum_{k=1}^{K} P(k|\mathbf{X}_t,(j)) \left(\mu_{\mathbf{X}_{k,t}^{(j)}} - \mu_{\mathbf{S},k}\right) \tag{23}$$

$$\mu_{\mathbf{X}_{k,t}^{(j)}} = \mathbf{f}\left(\mu_{\mathbf{S},k}, \mathbf{N}_t^{(j)}\right) \tag{24}$$

上式において，$\mu_{\mathbf{X}_{k,t}^{(j)}}$ は $\mathbf{X}_t$ の平均ベクトルであり，K は，クリーン音声 GMM の混合分布数である．また，$P(k|\mathbf{X}_t,(j))$ は $\mathbf{X}_t$ の事後確率であり，

$$P(k|\mathbf{X}_t,(j)) = \frac{P_{\mathbf{S},k}\mathcal{N}\left(\mathbf{X}_t, \mu_{\mathbf{X}_{k,t}^{(j)}}, \mathbf{\Sigma}_{\mathbf{X}_{k,t}^{(j)}}\right)}{\sum_{k'=1}^{K} P_{\mathbf{S},k'}\mathcal{N}\left(\mathbf{X}_t, \mu_{\mathbf{X}_{k',t}^{(j)}}, \mathbf{\Sigma}_{\mathbf{X}_{k',t}^{(j)}}\right)} \tag{25}$$

により与えられる．$\mathbf{\Sigma}_{\mathbf{X}_{k,t}^{(j)}}$ は，$\mathbf{X}_t$ の共分散行列であり，VTS (Vector Taylor Series) 法 [4] とパラメータ $\mu_{\mathbf{S},k}$, $\mathbf{\Sigma}_{\mathbf{S},k}$, $\mathbf{N}_t^{(j)}$ and $\mathbf{\Sigma}_{\mathbf{N}t}^{(j)}$. を用いて近似的に推定する．

最終的に，推定クリーン音声 $\hat{\mathbf{S}}_t$ は，次式により得られる．

$$\hat{\mathbf{S}}_t = \sum_{j=1}^{J} w_t^{(j)} \hat{\mathbf{S}}_t^{(j)} \tag{26}$$

3 Polyak averaging and feedback の導入

2 章にて述べたパーティクルフィルタは，2.1 節で定義された状態空間モデルに基づいて推定を行う．ここで，状態空間モデルの状態方程式には，式 (2) に示した Random walk 過程を適用していたが，Random walk 過程はパラメータの時間推移をランダム雑音により規定しているので，パラメータの時間推移を正確に表現できないという問題がある．状態空間モデルに基づいて対象のパラメータを正確に逐次推定するためには，状態方程式の定義が極めて重要である．この問題において，本研究では次式のような状態方程式を導入する．

$$\mathbf{N}_{t+1}^{(j)} = (1-\alpha)\mathbf{N}_t^{(j)} + \alpha\hat{\mathbf{N}}_t + \alpha\beta\left(\mu_{\mathbf{N}t}^{(j)} - \mathbf{N}_t^{(j)}\right) + \mathbf{W}_t^{(j)} \tag{27}$$

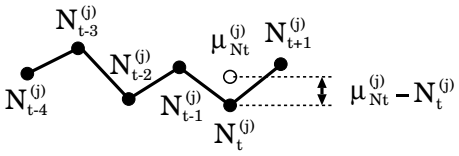
式 (27) の $\hat{\mathbf{N}}_t$ は，式 (28) により計算されるサンプル $\mathbf{N}_t^{(j)}$ の加重平均であり，α は忘却係数である．式 (27) の第 1, 第 2 項はサンプル $\mathbf{N}_t^{(j)}$ を平均値に近づけていることを意味しており，サンプルの散らばりを抑制する効果がある．これにより，真値とはかけ離れた値を持つ無意味なサンプルの出現を防ぐことができる．

$$\hat{\mathbf{N}}_t = \sum_{j=1}^{J} w_t^{(j)} \mathbf{N}_t^{(j)} \tag{28}$$

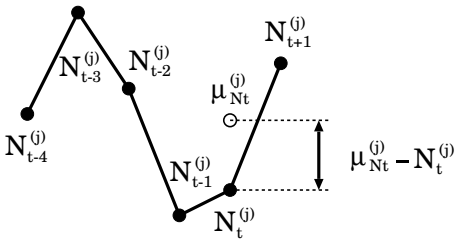
次に，式 (27) の $\mu_{\mathbf{N}t}^{(j)}$ は，式 (29) により計算される過去 T 点のサンプルの平均 (Polyak average [12]) である．式 (27) の第 3 項は，Polyak average のフィードバックを示しており，過去の平均値との差分を組み込むことにより，パラメータの時間変化量を表現している [8, 13]．なお，係数 β は，フィードバックのスケーリング係数である．

$$\mu_{\mathbf{N}t}^{(j)} = \frac{1}{T}\sum_{s=t-T+1}^{t} \mathbf{N}_s^{(j)} \tag{29}$$

図 3 は，Polyak averaging と feedback の概念図を示しており，図中 (a) のように $\mathbf{N}_t^{(j)}$ が過去に緩やかな動きをしている場合は，Polyak average $\mu_{\mathbf{N}t}^{(j)}$ と $\mathbf{N}_t^{(j)}$ の差分が小さくなる．この場合，式 (27) より $\mathbf{N}_t^{(j)}$ から $\mathbf{N}_{t+1}^{(j)}$ への変化量は小さいものとして予測される．一方，図中 (b) のように $\mathbf{N}_t^{(j)}$ が過去に激しい動きをしている場合は，$\mu_{\mathbf{N}t}^{(j)}$ と $\mathbf{N}_t^{(j)}$ の差分が大きくなる．よって，$\mathbf{N}_t^{(j)}$ から $\mathbf{N}_{t+1}^{(j)}$ への変化量は大きいものとして予測される．このように，過去の信号の変化度合いを考慮する Polyak averaging と feedback を導入することにより，Random walk 過程の場合に比べてパラメータの時間推移に対する拘束条件が強化され，より正確なパラメータの逐次推定を行うことができる．



(a) Slowly varying case



(b) Rapidly varying case

図 3: Polyak averaging と feedback の概念図

4 実験

4.1 実験条件

実験に用いた雑音重畳音声は，AURORA-2J [15] のクリーン音声 1001 文に，実環境で収録した雑音 [16] を人工的に加算して作成した．使用した雑音は，工場雑音と道路工事雑音であり，それぞれ非定常的な性質が強い雑音である．また，SNR は 20∼0dB とした．

パーティクルフィルタによる雑音推定及び，MMSE 推定に基づく雑音抑圧法で用いるクリーン音声 GMM は，AURORA-2J のクリーン学習データを用いて学習しており，混合分布数は 512 である（特徴量は 23 次対数メルスペクトル）．式 (2) のパラメータ $\mathbf{\Sigma_W}$ は，$\mathbf{\Sigma_W} = diag(0.01)$ に設定し，サンプルの総数は，$J = 50$ とした．また Polyak averaging と feedback のパラメータは，$\alpha = \{0.05, 0.1\}$, $\beta = \{1.0, 2.0\}$, $T = 5$ と設定した．

音声認識の際の特徴量は 0 次を含む 13 次 MFCC 及び，1 次，2 次の回帰係数を含む 39 次元の特徴量（CMS 処理有り）であり，音響モデルは，AURORA-2J 標準の HMM(16 状態，20 混合分布) を用いている．音響モデルの学習，認識は HTK ver. 3.2 [17] にて行った．

4.2 実験結果

図 4 は，工場雑音の逐次推定結果 (第 1 対数メルフィルタバンク出力値，SNR 0 dB) を示しており，“True noise” は真の雑音軌跡，“PF” はパーティクルフィルタの推定結果（2 章の手法，状態方程式に Random walk 過程を適用），“Polyak” は，Polyak averaging と feedback を用いた場合の結果を記述している．

図において，45 フレーム以降が雑音と音声が混在する区間であり，“PF” の推定誤差が大きくなっているのに対して，“Polyak” は真の雑音軌跡を追随することができていることがわかる．

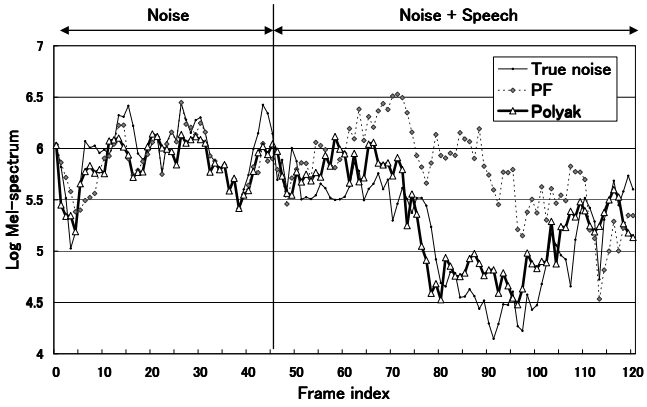


図 4: 工場雑音の推定結果

次に，表 1，2 に音声認識結果（単語正解精度）を示す．それぞれの表おいて，HTK Baseline，ETSI Advanced front-end [2]，雑音の逐次推定を行わない場合（MMSE），パーティクルフィルタ（2 章の手法，状態方程式に Random walk 過程を適用），Polyak averaging と feedback を用いた場合の結果を記述している．

表より，パーティクルフィルタを用いることにより，雑音の逐次推定を行わない場合に比べて音声認識性能の改善が得られることがわかる．特に，Polyak averaging と feedback を用いた場合の改善効果は大きく，3 章で述べたパラメータの時間推移に対する拘束条件が有効に作用したと言える．

また，提案手法の処理時間を Intel Pentium4 3.2GHz の CPU を用いて調査したところ，Polyak averaging と feedback の有無に関わらず，実時間のほぼ約 1.0 倍で動作した．このことから，提案法は実時間処理が可能かつ，大幅な音声認識性能の改善が得られる効果的な手法であると言える．

5 おわりに

本研究では，パーティクルフィルタを用いた非定常雑音の逐次推定及び，抑圧について検討を行い，評価の結果，提案手法が効果的であることを示した．また，状態空間モデルの状態方程式に，Polyak averaging と feedback を導入

表 1: 単語正解精度（工場雑音）(%)

SNR	HTK baseline	ETSI Advanced front-end	MMSE (Stationary noise compensation)	Particle filter	Polyak averaging and feedback			
					$\alpha = 0.05$ $\beta = 1.0$	$\alpha = 0.05$ $\beta = 2.0$	$\alpha = 0.1$ $\beta = 1.0$	$\alpha = 0.1$ $\beta = 2.0$
20 dB	93.61	92.88	96.41	96.13	**96.90**	96.84	96.84	96.78
15 dB	81.12	86.86	88.92	90.02	91.71	**91.93**	91.74	**91.93**
10 dB	54.81	76.73	74.27	75.87	81.39	81.98	**82.41**	82.04
5 dB	29.47	53.18	50.94	54.50	61.96	62.73	62.88	**63.28**
0 dB	18.73	23.15	24.72	28.92	35.92	36.75	**38.16**	37.95
Average	55.55	66.56	67.05	69.09	73.58	74.05	**74.41**	74.40

表 2: 単語正解精度（道路工事雑音）(%)

SNR	HTK baseline	ETSI Advanced front-end	MMSE (Stationary noise compensation)	Particle filter	Polyak averaging and feedback			
					$\alpha = 0.05$ $\beta = 1.0$	$\alpha = 0.05$ $\beta = 2.0$	$\alpha = 0.1$ $\beta = 1.0$	$\alpha = 0.1$ $\beta = 2.0$
20 dB	96.68	96.90	99.20	98.34	99.20	99.23	99.05	**99.39**
15 dB	89.93	94.81	97.61	95.61	97.79	97.79	98.10	**98.16**
10 dB	70.28	89.81	91.77	89.84	92.54	93.18	93.77	**93.86**
5 dB	38.81	76.02	71.57	75.28	78.14	78.48	**80.38**	80.35
0 dB	22.29	48.48	43.60	49.43	53.42	54.28	55.97	**57.08**
Average	63.60	81.20	80.75	81.70	84.22	84.59	85.45	**85.77**

することにより，ランダム雑音でパラメータの時間推移を規定した Rodom walk 過程を用いた場合に比べて大幅な音声認識性能の改善が得られる事を示した．

今後，α 等の最適なパラメータの適応的決定法及び，空間伝達特性（特に移動音源の特性）の逐次推定への応用について検討を行う予定である．

謝辞 本研究は，情報通信研究機構の研究委託により実施したものである．

参考文献

[1] S. F. Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction,” *IEEE Trans. on ASSP*, Vol. 27, No. 2, pp. 113-120, Apr. 1979.

[2] 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,” Non. 2003.

[3] J. C. Segura, A. de la Torre, M. C. Benitez, and A. M. Peinado, “Model-Based Compensation of the Additive Noise for Continuous Speech Recognition. Experiments Using AURORA II Database and Tasks,” *Proc. EuroSpeech ’01*, Vol. I, pp. 221-224, Aalborg, Denmark, Sept. 2001.

[4] P. J. Moreno, B. Raj, and R. M. Stern, “A Vector Taylor Series Approach for Environment-Independent Speech Recognition,” *Proc. ICASSP ’96*, Vol. II, pp. 733-736, Atlanta, USA, May 1996.

[5] V. Krishnamurthy and J. B. Moore, “On-Line Estimation of Hidden Markov Model Parameters Based on the Kullback-Leibler Information Measure,” *IEEE Trans. on SP*, Vol. 41, No. 8, pp. 2557-2573, Aug. 1993.

[6] M. Afify and O. Siohan, “Sequential Estimation with Optimal Forgetting for Robust Speech Recognition,” *IEEE Trans. on SAP*, Vol. 12, No. 1, pp. 19-26, Jan. 2004.

[7] K. Yao, K. K. Paliwal, and S. Nakamura, “Noise Adaptive Speech Recognition Based on Sequential Noise Parameter Estimation,” *Speech Communication*, Vol. 42, Issue 1, pp. 5-23, Jan. 2004.

[8] T. A. Myrvoll and S.Nakamura, “Online Cepstral Filtering Using A Sequential EM Approach with Polyak Averaging and Feedback,” *Proc. ICASSP ’05*, Vol. I, pp. 261-264, Philadelphia, USA, March, 2005.

[9] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking,” *IEEE Trans. SP*, Vol. 50, No. 2, pp. 174.188, Feb. 2002.

[10] K. Yao and S. Nakamura, “Sequential noise compensation by sequential Monte Carlo method,” *Proc. NIPS ’01*, pp. 1205-1212, Vancouver, Canada Dec. 2001.

[11] M. Fujimoto and S. Nakamura, “Particle Filter-based Non-stationary Noise Tracking for Robust Speech Recognition,” *ICASSP ’05*, Vol. I, pp. 257-260, Philadelphia, Mar. 2005.

[12] B. T. Polyak and A. B. Juditsky, “Acceleration of Stochastic Approximation by Averaging,” *SIAM J. Contr. Optim.*, Vol. 30, No. 4, pp.838-855, July 1992.

[13] H. J. Kushner and J, Yang, “Stochastic Approximation with Averaging and Feedback: Rapidly Convergent “On-Line” Algorithm,” *IEEE Trans. on AC*, Vol. 40, No. 1, pp. 24-34, Jan. 1995.

[14] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, Vol. 57, No. 1, pp. 97-109, Jan. 1970.

[15] S. Nakamura, K. Yamamoto, K. Takeda, S. Kuroiwa, N. Kitaoka, T. Yamada, M. Mizumachi, T. Nishiura, M. Fujimoto, A. Sasou, and T. Endo, “Data Collection and Evaluation of AURORA2-J Japanese Corpus,” *Proc. ASRU ’03*, pp. 619-623, St. Thomas, US Virgin Islands, USA, Dec. 2003.

[16] 遠藤 俊樹，堀内 俊治，清水 徹，中村 哲，“ATR 実環境雑音 DB – ATRANS –を用いた雑音重畳音声認識実験，” 情報処理学会研究報告, SLP-57-8, July 2005. (to appear)

[17] HTK Web site, http://htk.eng.cam.ac.uk/