



PARTICLE FILTERING FOR NON-STATIONARY SPEECH MODELLING AND ENHANCEMENT

J. Vermaak, C. Andrieu, A. Doucet

Signal Processing Group, Department of Engineering,
University of Cambridge,
Trumpington Street, Cambridge, CB2 1PZ, UK
{jv211, ca226, ad2}@eng.cam.ac.uk

ABSTRACT

This paper applies time-varying autoregressive (TVAR) models with stochastically evolving parameters to the problem of speech modelling and enhancement. The stochastic evolution models for the TVAR parameters are Markovian diffusion processes. The main aim of the paper is to perform on-line estimation of the clean speech and the model parameters, and to determine the adequacy of the chosen statistical models. An efficient simulation-based method is developed to solve the optimal filtering problem. The algorithm combines sequential importance sampling and a selection step, and employs several variance reduction strategies to make the best use of the statistical structure of the model. The modelling and enhancement performance of the model and algorithm are evaluated in simulation studies on real speech data sets.

1. INTRODUCTION

For enhancement purposes, speech is commonly modelled as the output of an autoregressive (AR) process observed in additive white Gaussian noise (AWGN). The main shortcoming of this model is obvious. Associated with the AR coefficients is an articulatory configuration that remains fixed throughout the analysis interval. In reality, however, the vocal tract is continually changing, sometimes slowly, sometimes rapidly. To partly reconcile the time-varying character of the vocal tract with the time invariance of the model, speech is normally processed in short (possibly overlapping) segments or frames, during each of which the signal is assumed to be stationary. However, since the framing is defined *a priori* with no relation to the phonetic information, non-stationary frames are still likely to occur, even for very short analysis intervals. In these circumstances non-stationary models may provide more true-to-life approximations of the behaviour of the vocal tract.

One such model is the time-varying AR (TVAR) process. It is a generalisation of the standard AR process where the model parameters are allowed to vary with time. Here a specific sub-class of TVAR models are considered, where the parameters are assumed to evolve according to Markovian diffusion processes. The parameters of these

models are normally estimated using iterative batch algorithms. However, these algorithms are not suited for real-time or near real-time implementations, and in speech processing applications, where the samples become available sequentially, on-line estimation methods make more sense. The development of such strategies is the main focus of this paper.

The TVAR speech and noise process model facilitates a state-space representation. Within a sequential framework general recursive expressions may be derived for the filtering distributions, from which estimates of the clean speech and model parameters may be obtained. However, if the model is non-linear and non-Gaussian, as it is here, no closed-form analytic solutions exist for the filtering distributions and the required estimates, and approximate methods need to be investigated.

One such strategy is that of sequential Monte Carlo integration, also commonly known as Monte Carlo particle methods (see [3] for an introduction and [2] for a summary of the state-of-the-art). Within this framework the distributions of interest are represented by a large number of samples, called particles. As will be evident later, these particles and their associated importance weights evolve randomly in time according to a simulation-based rule. Using these particles Monte Carlo estimates of the quantities of interest may be obtained, with the accuracy of these estimates being independent of the dimension of the state-space. This method is easier to implement than classical numerical methods, and allows complex non-linear and non-Gaussian estimation problems to be solved efficiently in an on-line manner.

The Monte Carlo particle filter developed here is not just a straightforward application of the basic methods, but is designed to make efficient use of the structure of the model, and incorporates various variance reduction strategies. Furthermore, it is straightforwardly combined with frequentist methods to perform model validation. The validation methods considered here give an objective and absolute measure of the goodness of fit of a particular model, and are only feasible using Monte Carlo simulation. At each iteration the algorithm has a computational complexity that is linear in the number of particles, and can easily be implemented on parallel computers, thus facilitating near real-time processing. Although not shown here, the Monte Carlo filter can also be combined with Markov chain Monte Carlo (MCMC) methods to obtain an efficient fixed-lag smoothing algorithm.

C. Andrieu is supported by AT&T Laboratories, Cambridge, UK.

A. Doucet is supported by EPSRC, UK.

The remainder of the paper is organised as follows. The model specification and estimation objectives are stated in Section 2. Section 3 develops a sequential simulation-based method to solve the filtering problem, and Section 4 shows how this method can be combined with frequentist strategies to determine the model adequacy. Finally, Section 5 presents and discusses some simulation results on real speech data sets.

2. MODEL SPECIFICATION AND ESTIMATION OBJECTIVES

The speech signal at time $t > 0$ is modelled as the output of a k -th order TVAR process, parameterised by a vector $\boldsymbol{\theta}_t \in \Theta \subset \mathbb{R}^{n_\theta}$, *i.e.*

$$x_t = \sum_{i=1}^k a_{i,t}(\boldsymbol{\theta}_t) x_{t-i} + \sigma_{e_t}(\boldsymbol{\theta}_t) e_t, \quad e_t \stackrel{iid}{\sim} \mathcal{N}(0, 1), \quad (1)$$

where $\mathbf{a}_t(\boldsymbol{\theta}_t) \triangleq (a_{1,t}(\boldsymbol{\theta}_t), \dots, a_{k,t}(\boldsymbol{\theta}_t))$ are the TVAR coefficients and $\sigma_{e_t}^2(\boldsymbol{\theta}_t)$ is the variance of the TVAR excitation sequence. The signal is assumed to be submerged in AWGN, so that the observed value at time $t > 0$ becomes

$$y_t = x_t + \sigma_{n_t}(\boldsymbol{\theta}_t) n_t, \quad n_t \stackrel{iid}{\sim} \mathcal{N}(0, 1), \quad (2)$$

where $\{n_t\}$ is a white noise process independent of $\{e_t\}$, and $\sigma_{n_t}^2(\boldsymbol{\theta}_t)$ is the variance of the observation noise.

Conditionally on $\{\boldsymbol{\theta}_t\}$ the signal model is linear, facilitating a conditionally Gaussian state-space (CGSS) representation. More precisely, defining the vectors $\boldsymbol{\alpha}_t \triangleq (x_t, \dots, x_{t-k+1})$, $\mathbf{y}_t \triangleq (y_t)$, $\mathbf{v}_t \triangleq (e_t)$ and $\mathbf{w}_t \triangleq (n_t)$, and the system matrices

$$\mathbf{A}_t(\boldsymbol{\theta}_t) \triangleq \begin{bmatrix} \mathbf{a}_t^T(\boldsymbol{\theta}_t) \\ \mathbf{I}_{k-1} & \mathbf{0}_{k-1 \times 1} \end{bmatrix} \quad \mathbf{B}_t(\boldsymbol{\theta}_t) \triangleq \begin{bmatrix} \sigma_{e_t}(\boldsymbol{\theta}_t) \\ \mathbf{0}_{k-1 \times 1} \end{bmatrix} \quad (3)$$

$$\mathbf{C}_t(\boldsymbol{\theta}_t) = \mathbf{C} \triangleq [1 \quad \mathbf{0}_{1 \times k-1}] \quad \mathbf{D}_t(\boldsymbol{\theta}_t) \triangleq [\sigma_{n_t}(\boldsymbol{\theta}_t)], \quad (4)$$

the signal model of (1) and (2) is readily expressed in the CGSS form given by

$$\boldsymbol{\alpha}_t = \mathbf{A}_t(\boldsymbol{\theta}_t) \boldsymbol{\alpha}_{t-1} + \mathbf{B}_t(\boldsymbol{\theta}_t) \mathbf{v}_t, \quad \mathbf{v}_t \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}_{n_v \times 1}, \mathbf{I}_{n_v}) \quad (5)$$

$$\mathbf{y}_t = \mathbf{C}_t(\boldsymbol{\theta}_t) \boldsymbol{\alpha}_t + \mathbf{D}_t(\boldsymbol{\theta}_t) \mathbf{w}_t, \quad \mathbf{w}_t \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}_{n_w \times 1}, \mathbf{I}_{n_w}), \quad (6)$$

where $\boldsymbol{\alpha}_t \in \mathbb{R}^{n_\alpha}$ is the system state, $\mathbf{y}_t \in \mathbb{R}^{n_y}$ is the observation, and $\mathbf{v}_t \in \mathbb{R}^{n_v}$ and $\mathbf{w}_t \in \mathbb{R}^{n_w}$ are the system disturbances at time t , respectively. It is further assumed that $\mathbf{D}_t(\boldsymbol{\theta}_t) \mathbf{D}_t^T(\boldsymbol{\theta}_t) > 0$, for all $t > 0$, $\boldsymbol{\alpha}_0 \sim \mathcal{N}(\mathbf{m}_0(\boldsymbol{\theta}_0), \mathbf{P}_0(\boldsymbol{\theta}_0))$, with $\mathbf{P}_0(\boldsymbol{\theta}_0) > 0$, and that $\boldsymbol{\alpha}_0$, \mathbf{v}_t and \mathbf{w}_t are mutually independent for all $t > 0$.

The model order k is assumed to be fixed and known throughout. The unknown parameters are then the TVAR coefficients and the excitation and observation noise variances. Here the TVAR coefficients are represented in their standard form, whereas the variances are parameterised by their corresponding logarithms, *i.e.* $\phi_{e_t} \triangleq \log \sigma_{e_t}^2$ and $\phi_{n_t} \triangleq \log \sigma_{n_t}^2$, so that the unknown parameter vector at time t may be expressed as $\boldsymbol{\theta}_t \triangleq (\mathbf{a}_t, \phi_{e_t}, \phi_{n_t})$, $n_\theta = k + 2$,

with corresponding support $\Theta \triangleq A_k \times \mathbb{R} \times \mathbb{R}$, where A_k is the region of stability for the coefficients of a k -th order stationary AR process.

The unknown parameters are assumed to evolve according to a first-order Markov process, which is fully specified by its initial state and state transition distributions, here taken to be $p(\boldsymbol{\theta}_0) = p(\mathbf{a}_0) p(\phi_{e_0}) p(\phi_{n_0}) = \mathcal{N}(\mathbf{a}_0; \mathbf{0}_{k \times 1}, \boldsymbol{\Delta}_{\mathbf{a}_0}) \mathbb{I}_{A_k}(\mathbf{a}_0) \mathcal{N}(\phi_{e_0}; 0, \delta_{e_0}^2) \mathcal{N}(\phi_{n_0}; 0, \delta_{n_0}^2)$ and $p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}) = p(\mathbf{a}_t | \mathbf{a}_{t-1}) p(\phi_{e_t} | \phi_{e_{t-1}}) p(\phi_{n_t} | \phi_{n_{t-1}}) = \mathcal{N}(\mathbf{a}_t; \mathbf{a}_{t-1}, \boldsymbol{\Delta}_{\mathbf{a}}) \mathbb{I}_{A_k}(\mathbf{a}_t) \mathcal{N}(\phi_{e_t}; \phi_{e_{t-1}}, \delta_e^2) \mathcal{N}(\phi_{n_t}; \phi_{n_{t-1}}, \delta_n^2)$, for $t > 0$, respectively, where the parameters of the Markov process $(\boldsymbol{\Delta}_{\mathbf{a}_0}, \boldsymbol{\Delta}_{\mathbf{a}}, \delta_{e_0}^2, \delta_e^2, \delta_{n_0}^2, \delta_n^2)$, with $\boldsymbol{\Delta}_{\mathbf{a}_0} \triangleq \text{diag}(\delta_{a_{1,0}}^2, \dots, \delta_{a_{k,0}}^2)$ and $\boldsymbol{\Delta}_{\mathbf{a}} \triangleq \text{diag}(\delta_{a_1}^2, \dots, \delta_{a_k}^2)$, are assumed to be fixed and known. The signal model defines a non-linear non-Gaussian state-space model for which no closed-form solutions exist for the filtering distributions, hence necessitating numerical estimation strategies.

Given at time $t > 0$ the observations $\mathbf{y}_{1:t}$, all Bayesian inference for the signal model relies on the joint posterior distribution $p(\boldsymbol{\alpha}_{0:t}, \boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$. The optimal estimation problem of interest here is to compute the filtering distribution $p(\boldsymbol{\alpha}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t})$, as well as the MMSE estimate of $f_{t|t}(\boldsymbol{\alpha}_t, \boldsymbol{\theta}_t)$, with $f_{t|t} : \mathbb{R}^{n_\alpha} \times \Theta \rightarrow \mathbb{R}^{n_{f_{t|t}}}$, given by $I(f_{t|t}) \triangleq \mathbb{E}_{p(\boldsymbol{\alpha}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t})} [f_{t|t}(\boldsymbol{\alpha}_t, \boldsymbol{\theta}_t)]$. To obtain the filtered estimates of the clean speech signal and model parameters $f_{t|t}$ is set to $f_{t|t}(\boldsymbol{\alpha}_t, \boldsymbol{\theta}_t) = (\boldsymbol{\alpha}_t, \boldsymbol{\theta}_t)$.

3. A SIMULATION-BASED OPTIMAL FILTER

The problem of estimating $p(\boldsymbol{\alpha}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t})$ and $I(f_{t|t})$ can be reduced to one of sampling from $p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$. Indeed, $p(\boldsymbol{\alpha}_t, \boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t}) = p(\boldsymbol{\alpha}_t | \boldsymbol{\theta}_{0:t}, \mathbf{y}_{1:t}) p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$, where $p(\boldsymbol{\alpha}_t | \boldsymbol{\theta}_{0:t}, \mathbf{y}_{1:t})$ is a Gaussian distribution whose parameters may be computed using the Kalman filter. Thus, given an approximation of $p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$, an approximation of $p(\boldsymbol{\alpha}_t, \boldsymbol{\theta}_t | \mathbf{y}_{1:t})$ may straightforwardly be obtained. Since this scheme avoids the sampling of $\boldsymbol{\alpha}_t$, it greatly reduces the variance of the corresponding Monte Carlo estimates¹. Denoting by $\pi(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$ the so-called importance distribution, and assuming that a set of samples $\{\boldsymbol{\theta}_{0:t}^{(i)} : i = 1, \dots, N\}$ distributed according to this distribution is available, a Bayesian importance sampling [3] estimate of $I(f_{t|t})$ follows as

$$\widehat{I}_N(f_{t|t}) \triangleq \sum_{i=1}^N \widetilde{w}_{0:t}^{(i)} \mathbb{E}_{p(\boldsymbol{\alpha}_t | \boldsymbol{\theta}_{0:t}, \mathbf{y}_{1:t})} [f_{t|t}(\boldsymbol{\alpha}_t, \boldsymbol{\theta}_t^{(i)})], \quad (7)$$

where the normalised importance weights are given by

$$\widetilde{w}_{0:t}^{(i)} \triangleq \frac{w(\boldsymbol{\theta}_{0:t}^{(i)})}{\sum_{j=1}^N w(\boldsymbol{\theta}_{0:t}^{(j)})}, \quad w(\boldsymbol{\theta}_{0:t}) \propto \frac{p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})}{\pi(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})}. \quad (8)$$

For sequential estimation the aim is to obtain at any time t an estimate of the distribution $p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$ and to be able to propagate this estimate in time without modifying subsequently the past simulated trajectories $\{\boldsymbol{\theta}_{0:t}^{(i)} : i = 1, \dots, N\}$. This means that $\pi(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$

¹This result can be rigorously proved, but is omitted for the sake of conciseness.

should admit $\pi(\boldsymbol{\theta}_{0:t-1} | \mathbf{y}_{1:t-1})$ as marginal distribution. This is possible if the importance distribution is restricted to be of the general form

$$\pi(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t}) = \pi(\boldsymbol{\theta}_0) \prod_{k=1}^t \pi(\boldsymbol{\theta}_k | \boldsymbol{\theta}_{0:k-1}, \mathbf{y}_{1:k}). \quad (9)$$

Such an importance distribution allows recursive evaluation of the importance weights, *i.e.* $w(\boldsymbol{\theta}_{0:t}) = w(\boldsymbol{\theta}_{0:t-1}) w_t$, with

$$w_t \propto \frac{p(\mathbf{y}_t | \boldsymbol{\theta}_{0:t}, \mathbf{y}_{1:t-1}) p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})}{\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t})}. \quad (10)$$

There is an unlimited number of choices for the importance distribution, the only restrictions being that its support includes that of $p(\boldsymbol{\theta}_{0:t} | \mathbf{y}_{1:t})$, and that it is easily sampled from. For simplicity the importance distribution is here taken to be the prior distribution, *i.e.* $\pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{0:t-1}, \mathbf{y}_{1:t}) = p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$, so that w_t in (10) becomes $w_t \propto p(\mathbf{y}_t | \boldsymbol{\theta}_{0:t}, \mathbf{y}_{1:t-1})$. Evaluation of this requires only one step of the Kalman filter for each particle.

For importance distributions of the form specified by (9) the variance of the importance weights can only increase (stochastically) over time. It is thus impossible to avoid a degeneracy phenomenon. Practically, after a few iterations of the algorithm, all but one of the normalised importance weights are very close to zero, and a large computational effort is devoted to updating trajectories whose contribution to the final estimate is almost zero. For this reason it is of crucial importance to include a selection (or resampling) step in the algorithm. The purpose of this step is to discard particles with low normalised importance weights and multiply those with high normalised importance weights, so as to avoid the degeneracy of the algorithm. Numerous selection strategies exist (see *e.g.* [2, 4]), the discussion of which falls outside the scope of this paper.

To summarise, given at time $t-1$, $N \in \mathbb{N}^*$ particles $\{\boldsymbol{\theta}_{0:t-1}^{(i)} : i = 1, \dots, N\}$ distributed approximately according to $p(\boldsymbol{\theta}_{0:t-1} | \mathbf{y}_{1:t-1})$, the Monte Carlo filter proceeds as follows at time t .

Monte Carlo Filter

Sequential Importance Sampling Step

- For $i = 1, \dots, N$, sample $\tilde{\boldsymbol{\theta}}_t^{(i)} \sim \pi(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})$ and set $\tilde{\boldsymbol{\theta}}_{0:t}^{(i)} = (\boldsymbol{\theta}_{0:t-1}^{(i)}, \tilde{\boldsymbol{\theta}}_t^{(i)})$.
- For $i = 1, \dots, N$, evaluate the importance weights $w_t^{(i)}$ (see (10)) up to a normalising constant, and subsequently normalise.

Selection Step

- Multiply / discard particles $\{\tilde{\boldsymbol{\theta}}_{0:t}^{(i)} : i = 1, \dots, N\}$ *w.r.t.* high / low normalised importance weights to obtain N particles $\{\boldsymbol{\theta}_{0:t}^{(i)} : i = 1, \dots, N\}$.
-

The computational complexity of this algorithm at each iteration is clearly $O(N)$. At first glance, it could appear necessary to keep in memory the paths of all the

trajectories $\{\boldsymbol{\theta}_{0:t}^{(i)} : i = 1, \dots, N\}$, so that the storage requirements would increase linearly with time. In fact, the importance distribution adopted here and associated importance weights depend on $\boldsymbol{\theta}_{0:t-1}$ only via a set of low-dimensional sufficient statistics, which may be computed using the Kalman filter, and only these values need to be kept in memory. Thus, the storage requirements are also $O(N)$ and do not increase over time.

4. MODEL VALIDATION

Model validation is the process of determining how well a model fits the data. Within a Bayesian framework models can be compared using posterior model probabilities, but this strategy only provides relative performance indicators, and does not tell whether any particular model fits the data well. Here it is shown how the Monte Carlo filter and frequentist methods may be combined to determine the goodness of fit for any model of the data.

Under the null hypothesis that the model is correct, the sequence $\{u_k\}$, with $u_k \triangleq p(Y_k \leq y_k | \mathbf{y}_{1:k-1})$, is a realisation of *i.i.d.* random variables distributed according to $\mathcal{U}_{[0,1]}$ [2]. This result holds true for any time series model and may be used in statistical tests to determine the adequacy of the model. Using the one-step ahead prediction distribution, an expression for u_k follows straightforwardly as

$$u_k = \int p(Y_k \leq y_k | \boldsymbol{\theta}_{0:k}, \mathbf{y}_{1:k-1}) p(d\boldsymbol{\theta}_{0:k} | \mathbf{y}_{1:k-1}). \quad (11)$$

The integration in the above expression is analytically intractable, but knowing that $p(\boldsymbol{\theta}_{0:k} | \mathbf{y}_{1:k-1}) = p(\boldsymbol{\theta}_k | \boldsymbol{\theta}_{0:k-1}) p(\boldsymbol{\theta}_{0:k-1} | \mathbf{y}_{1:k-1})$, a Monte Carlo approximation of the one-step ahead prediction distribution may be obtained as $\widehat{p}_N(d\boldsymbol{\theta}_{0:k} | \mathbf{y}_{1:k-1}) \triangleq N^{-1} \sum_{i=1}^N \delta_{\boldsymbol{\theta}_{0:k}^{*(i)}}(d\boldsymbol{\theta}_{0:k})$, where $\boldsymbol{\theta}_{0:k}^{*(i)} \triangleq (\boldsymbol{\theta}_{0:k-1}^{(i)}, \boldsymbol{\theta}_k^{*(i)})$, with $\widehat{p}_N(d\boldsymbol{\theta}_{0:k-1} | \mathbf{y}_{1:k-1}) \triangleq N^{-1} \sum_{i=1}^N \delta_{\boldsymbol{\theta}_{0:k-1}^{(i)}}(d\boldsymbol{\theta}_{0:k-1})$ the Monte Carlo approximation of the filtering distribution at time $k-1$, and $\boldsymbol{\theta}_k^{*(i)} \sim p(\boldsymbol{\theta}_k | \boldsymbol{\theta}_{k-1}^{(i)})$ generated from the Markov process prior. With this approximation a Monte Carlo estimator for u_k follows straightforwardly as

$$\widehat{u}_k \triangleq \frac{1}{N} \sum_{i=1}^N p(Y_k \leq y_k | \boldsymbol{\theta}_{0:k}^{*(i)}, \mathbf{y}_{1:k-1}), \quad (12)$$

where $p(Y_k \leq y_k | \boldsymbol{\theta}_{0:k}, \mathbf{y}_{1:k-1}) = \int_{-\infty}^{y_k} p(dy_s | \boldsymbol{\theta}_{0:k}, \mathbf{y}_{1:k-1})$ can be computed using the one-step ahead prediction distribution for the observations, obtained from the Kalman filter. In the experiments reported here the Bowman-Shenton [1] and Ljung-Box [5] statistics are used to respectively test for the normality and whiteness of the sequence $\{v_k\}$, where $v_k \triangleq \Psi^{-1}(u_k)$, with Ψ the standard Gaussian cumulative distribution function.

5. SIMULATION STUDIES

Figure 1 shows two frames of speech and their corresponding noise-corrupted versions, with input SNRs of -0.61 dB and 6.10 dB, respectively. These sections of speech were

chosen to be representative of the kind of non-stationarities that are traditionally not well modelled by the standard AR model. The first (F1) shows the rather gradual transition between the fricative /sh/ and the vowel /uw/ in the word “should”, whereas the second (F2) depicts the much sharper transition between the fricative /s/ and the vowel /er/ in the word “service”.

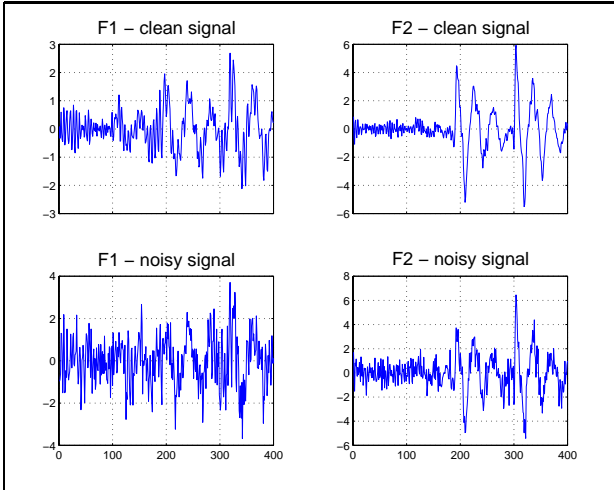


Figure 1: Speech data.

The Monte Carlo filter was run on F1 and F2 for various values of N , with the model order fixed to $k = 4$. No significant further improvements in the results were observed with an increase in k above 4. This useful result is due to the fact that the non-stationary character of the TVAR model allows for much more modelling flexibility than, say, a standard AR model of the same order. The fixed parameters of the Markov process on the TVAR parameters were set to $(\Delta_{\mathbf{a}_0}, \Delta_{\mathbf{a}}, \delta_{e_0}^2, \delta_e^2, \delta_{n_0}^2, \delta_n^2) = (0.5\mathbf{I}_4, 5 \times 10^{-3}\mathbf{I}_4, 0.5, 0.5 \times 10^{-3}, 0.5, 0.5 \times 10^{-3})$. The algorithm proved to be fairly robust to the specific values chosen for these quantities.

The SNR improvement results are summarised in Table 1, and were obtained by averaging over 50 independent runs of the algorithm for each value of N . The filtering performance for F2 steadily improves with an increase in the number of particles up to $N = 1000$, whereas good filtering performance is achieved for F1 with as few as $N = 10$ particles. This discrepancy is due to the relatively low input SNR of F1 compared to that of F2. In general it was found that the number of particles needed to accurately represent the filtering distribution increases with an increase in the input SNR. The reason for this becomes more obvious when considering the limiting case where there is no observation noise. In this case the true filtering distribution is a Dirac delta measure, and the probability of generating particles within its support is zero, so that an infinite number of particles would be required.

N	10	50	100	250	500	1000
F1	2.79	2.95	2.81	2.81	2.83	2.85
F2	-0.17	1.36	1.69	1.86	1.90	1.94

Table 1: SNR improvement results in dB.

To determine the adequacy of the model the Bowman-Shenton (BS) and Ljung-Box (LB) (using a lag of 5) tests were applied to F1 and F2. The results were obtained by averaging over 50 independent runs of the algorithm, and are presented in Table 2, together with the 5% critical values for the statistics. The results of the BS test show that the residuals are indeed standard normal distributed. The results of the LB test, however, indicate that there are still significant autocorrelations present in the residuals. Thus, even though the TVAR model is superior to the standard AR model, it is still less than ideal, and does not capture all the salient features present in the speech signals. A possible explanation for this inadequacy may be the presence of longer-term dependencies due to the glottal excitation in voiced speech signals. These dependencies cannot adequately be accounted for by models conditioning only on the recent past. Future work will focus on extending the basic TVAR model to overcome this problem.

Frame	N	BS		LB	
		val.	5% crit.	val.	5% crit.
F1	100	2.1930	5.9915	20.4634	11.0705
F2	250	4.2357	5.9915	25.7460	11.0705

Table 2: Model validation results.

Finally, the filter was run with $N = 100$ particles on an utterance of the sentence “*Good service should be rewarded by big tips.*” by a male American speaker. The clean signal was acoustically combined with a slowly time-varying AWGN process so that the input SNR over the whole utterance was 0.16 dB. The filter achieved a SNR improvement of 5.44 dB. Informal listening tests confirmed the reduction in the noise and revealed no musical or other undesired artifacts common to block-based enhancement algorithms.

6. REFERENCES

- [1] K. O. Bowman and L. R. Shenton. Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2 . *Biometrika*, 62(2):243–250, 1975.
- [2] A. Doucet, J. F. G. de Freitas, and N. J. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York, 2000. To Appear.
- [3] A. Doucet, S. J. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 2000. To Appear.
- [4] J. S. Liu and R. Chen. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93:1032–1044, 1998.
- [5] G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.