SIGNAL DENOISING WITH HIDDEN MARKOV MODELS USING HIDDEN MARKOV TREES AS OBSERVATION DENSITIES

Diego H. Milone, Leandro E. Di Persia and Diego R. Tomassi

Signals and Computational Intelligence Laboratory
FICH-UNL, Ciudad Universitaria, Santa Fe, Argentina

ABSTRACT

Wavelet-domain hidden Markov models have been found successful in exploiting statistical dependencies between wavelet coefficients for signal denoising. However, these models typically deal with fixed-length sequences and are not suitable neither for very long nor for real-time signals. In this paper, we propose a novel denoising method based on a Markovian model for signals analyzed on a short-term basis. The architecture is composed of a hidden Markov model in which the observation probabilities are provided by hidden Markov trees. Long-term dependencies are captured in the external model which, in each internal state, deals with the local dynamics in the time-scale plane. Model-based denoising is carried out by an empirical Wiener filter applied to the sequence of frames in the wavelet domain. Experimental results with standard test signals show important reductions of the mean-squared errors.

1. INTRODUCTION

The wavelet transform has shown to be a very interesting representation for signal and image analysis and it has led to simple but powerful approaches to statistical signal processing. Many of these methods assume that the wavelet coefficients are jointly Gaussian or statistically independent. However, actual signals show sparse wavelet representations and some residual dependency structure between the coefficients which do not agree with those models.

In order to account for these features, hidden Markov trees (HMT) were introduced in [1]. In this model, a Gaussian mixture density models the distribution of each wavelet coefficient. Each component in the mixture is related to the state taken by a hidden variable associated with the coefficient. By setting Markovian dependencies between the hidden state variables based on the natural tree structure of the wavelet decomposition, the model also captures the statistical dependencies at different scales. The HMT model has been improved in several ways in the last years, for example, using more states at each HMT node and developing more efficient algorithms for initialization and training [2, 3, 4]. However, the HMT still cannot deal neither with long-term dependencies nor with variable-length sequences or multiple image sizes. These shortcomings arise from the use of the discrete wavelet transform (DWT), which makes the structure of the representations depend on the length of the signal. Although this is not very restrictive when just a single observation is used to train a tied model, in many applications we have multiple observations available and we would want to use the whole information in order to train a full model. In these cases, if an HMT were to be considered, the model should be trained and used only with signals of the same length. Otherwise, a warping preprocessing would be required.

On the other hand, hidden Markov models (HMM) have been widely used for the statistical modeling of time series [5]. Despite of their relative simplicity, they are effective in handling correlations across time and they are very successful in dealing with sequences of different lengths. Although they have been traditionally used with Gaussian mixtures as observation densities, other models have been proposed for the observation distributions [6].

In this paper we propose a novel method to signal denoising based on a wavelet-domain Markovian model for sequences analyzed on a short-term basis, but not assuming stationarity within each frame. In order to combine the advantages of traditional HMM and those of the HMT to model the statistical dependencies between wavelet coefficients, we derive an EM algorithm to train a composite model in which each state of an external HMM uses an observation model provided by an HMT. In this HMM-HMT architecture, the external HMM handles the long term dynamics, while the local dynamics are appropriately captured in the wavelet domain by each HMT. We then apply the model in an empirical Wiener filter defined in the wavelet domain, thus allowing for a model-based approach to signal denoising in a short-term framework.

This work is supported by the National Research Council for Science and Technology (CONICET), the National Agency for the Promotion of Science and Technology (ANPCyT-UNL PICT 11-25984 and ANPCyT-UNER PICT 11-12700), and the National University of Litoral (UNL, project CAID 012-72).

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
D. H. Milone, L. Di Persia & D. R. Tomassi; "Signal denoising with hidden Markov models using hidden Markov trees as observation densities"
18th IEEE International Workshop on Machine Learning for Signal Processing, oct, 2008.

Dasgupta et al. [7] proposed a dual-Markov architecture trained by means of an iterative process where the most probable sequence of states is identified, and then each internal model is adapted with the selected observations. A similar approach applied to image segmentation was proposed in [8]. However, in both cases the model consists of two separated and independent entities, that are just forced to work in a coupled way. By contrast, in [9] an EM algorithm was derived for a full model composed of an external HMM in which, for each state, an internal HMM provides the observation probability distribution. The training algorithm proposed here follows this later approach rather than that in [7].

In the next section, we take a full Baum-Welch approach to parameter estimation and we state the reestimation formulas for the composite HMM-HMT model. Then we detail the model-based denoising approach for HMM-HMT and we provide results of benchmark experiments with the standard Doppler and Heavisine signals from Donoho and Johnstone [10].

2. THE HMM-HMT MODEL

The proposed model is a composition of two Markov models: the long term dependencies are modeled with an external HMM and each segment in the local context is modeled in the wavelet domain by an HMT. Figure 1 shows a diagram of this architecture, along with the analysis stage and the synthesis stage needed for signal denoising. In this section we define the HMM-HMT model and state the joint likelihood of the observations and the hidden states given the model. Then, the EM approach for the estimation of the model parameters is presented for single and multiple observations.

2.1. Model Definition and Notation

In order to model a sequence $\mathbf{W} = \mathbf{w}^1, \mathbf{w}^2, \ldots, \mathbf{w}^T$, with $\mathbf{w}^t \in \mathbb{R}^N$, we define a continuous HMM with the usual structure $\vartheta = \langle \mathcal{Q}, \mathbf{A}, \boldsymbol{\pi}, \mathcal{B} \rangle$, where $\mathcal{Q}$ is the set of states taking values $q \in 1, 2, \ldots, N_Q$; $\mathbf{A}$ is the matrix of state transition probabilities; $\boldsymbol{\pi}$ is the initial state probability vector; and $\mathcal{B} = \{b_k(\mathbf{w}^t)\}$, is the set of observation (or emission) probability distributions.

Suppose now that $\mathbf{w}^t = [w_1^t, w_2^t, \ldots, w_N^t]$, with $w_n^t \in \mathbb{R}$, results from a DWT analysis of the signal with J scales and that w_0, the approximation coefficient at the coarsest scale, is not included in the vector so that $N = 2^J - 1$. The HMT in the state k of the HMM can be defined with the structure $\theta^k = \langle \mathcal{U}^k, \mathcal{R}^k, \boldsymbol{\kappa}^k, \boldsymbol{\epsilon}^k, \mathcal{F}^k \rangle$, where $\mathcal{U}^k$ is the set of nodes in the tree; $\mathcal{R}^k$ is the set of states in all the nodes of the tree, being $\mathcal{R}_u^k$ the set of states in the node u which take values $r_u \in 1, 2, \ldots, M$; $\boldsymbol{\epsilon}^k = \left[\epsilon_{u,mn}^k\right]$ is the array whose elements hold the conditional probability of node u being in state m given that the state in its parent node $\rho(u)$ is n; $\boldsymbol{\kappa}^k$ are the probabilities for the initial states in the root node; and $\mathcal{F}^k = \left\{ f_{u,m}^k(w_u) \right\}$ is the set of observation probability distributions, that is, $f_{u,m}^k(w_u^t)$ is the probability of observing the wavelet coefficient w_u^t with the state m (in the node u). In addition, we will refer to the set $\mathcal{C}_u = \{c_1(u), c_2(u), \ldots, c_N(u)\}$ of children nodes of node u.

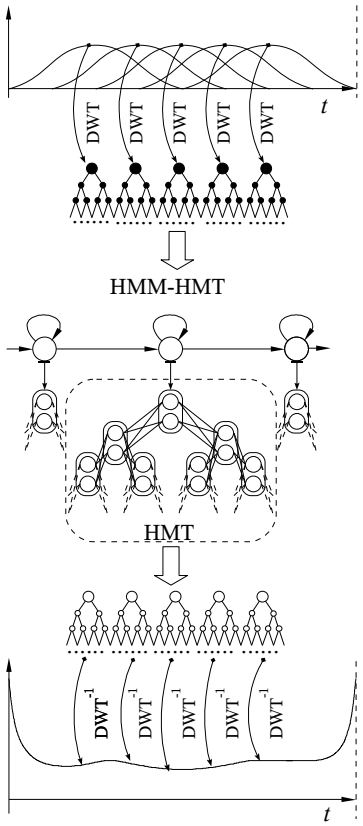

Fig. 1. Diagram of the proposed model.

2.2. Joint Likelihood

First, we restate for convenience the three basic assumptions regarding an HMT:

i) $\Pr(r_u = m | r_v / v \neq u) = \Pr(r_u = m | r_{\rho(u)}, r_{\mathcal{C}_u})$,

ii) $\Pr(\mathbf{w}|\mathbf{r}) = \prod_u \Pr(w_u|\mathbf{r})$,

iii) $\Pr(w_u|\mathbf{r}) = \Pr(w_u|r_u)$.

With this in mind, the observation density for each HMM state reads

$$b_{q^t}(\mathbf{w}^t) = \sum_{\forall \mathbf{r}} \prod_{\forall u} \epsilon_{u,r_u r_{\rho(u)}}^{q^t} f_{u,r_u}^{q^t}(w_u^t), \qquad (1)$$

where $\mathbf{r} = [r_1, r_2, \ldots, r_N]$ is a combination of hidden states in the HMT nodes. Thus, the complete joint likelihood for the HMM-HMT can be obtained as

$$\begin{aligned}\mathcal{L}_\Theta(\mathbf{W}) &= \sum_{\forall \mathbf{q}} \prod_t a_{q^{t-1}q^t} b_{q^t}(\mathbf{w}^t) \\ &= \sum_{\forall \mathbf{q}} \prod_t a_{q^{t-1}q^t} \sum_{\forall \mathbf{r}} \prod_{\forall u} \epsilon^{q^t}_{u,r^t_u r^t_{\rho(u)}} f^{q^t}_{u,r^t_u}(w^t_u) \\ &= \sum_{\forall \mathbf{q}} \sum_{\forall \mathbf{R}} \prod_t a_{q^{t-1}q^t} \prod_{\forall u} \epsilon^{q^t}_{u,r^t_u r^t_{\rho(u)}} f^{q^t}_{u,r^t_u}(w^t_u) \\ &\triangleq \sum_{\forall \mathbf{q}} \sum_{\forall \mathbf{R}} \mathcal{L}_\Theta(\mathbf{W}, \mathbf{q}, \mathbf{R}), \end{aligned} \tag{2}$$

where $a_{01} = \pi_1 = 1$. $\forall \mathbf{q}$ says that the sum is over all possible state sequences $\mathbf{q} = q^1, q^2, \ldots, q^T$ and $\forall \mathbf{R}$ accounts for all possible sequences of all possible combinations of hidden states $\mathbf{r}^1, \mathbf{r}^2, \ldots, \mathbf{r}^T$ in the nodes of each tree.

2.3. EM Formulation

In this section we will obtain a maximum likelihood estimation of the model parameters. For the optimization, the auxiliary function can be defined as

$$\mathcal{D}(\Theta, \bar{\Theta}) \triangleq \sum_{\forall \mathbf{q}} \sum_{\forall \mathbf{R}} \mathcal{L}_\Theta(\mathbf{W}, \mathbf{q}, \mathbf{R}) \log\left(\mathcal{L}_{\bar{\Theta}}(\mathbf{W}, \mathbf{q}, \mathbf{R})\right). \tag{3}$$

Using (2), this function can be separated in independent functions for the estimation of a_{ij}, $\epsilon^k_{u,mn}$, and the parameters of $f^k_{u,m}(w_u)$. For the estimation of the transition probabilities in the external HMM, a_{ij}, it is then easy to see that no changes from the standard formulas are needed.

Let be $q^t = k$, $r^t_u = m$ and $r^t_{\rho(u)} = n$. To obtain the reestimation formula for $\epsilon^k_{u,mn}$, the restriction $\sum_m \epsilon^k_{u,mn} \stackrel{\circ}{=} 1$ should be satisfied. Using Lagrange multipliers, we can thus optimize

$$\hat{\mathcal{D}}(\Theta, \bar{\Theta}) \triangleq \mathcal{D}(\Theta, \bar{\Theta}) + \sum_n \lambda_n \left(\sum_m \epsilon^k_{u,mn} - 1 \right), \tag{4}$$

and the reestimation formula results

$$\epsilon^k_{u,mn} = \frac{\sum_t \gamma^t(k) \xi^{tk}_u(m,n)}{\sum_t \gamma^t(k) \gamma^{tk}_{\rho(u)}(n)}, \tag{5}$$

where $\gamma^t(k)$ is the probability of being in state k (of the external HMM) at time t (computed as usual for HMM); $\gamma^{tk}_{\rho(u)}(n)$ is the probability of being in state m of the node u, in the HMT corresponding to the state k in the HMM and at time t; and $\xi^{tk}_u(m,n)$ is the probability of being in state m at node u, and in state n at its parent node $\rho(u)$, at time t and in the HMT of the state k in the HMM. The last two expected variables, $\gamma^{tk}_{\rho(u)}(n)$ and $\xi^{tk}_u(m,n)$, can be estimated with the upward-downward algorithm [4].

We can proceed in a similar way to estimate de parameters of $f^k_{u,m}(w_u)$. For observations given by $f^k_{u,r^t_u}(w^t_u) = \mathcal{N}(w^t_u, \mu^k_{u,m}, \sigma^k_{u,m})$, we find:

$$\mu^k_{u,m} = \frac{\sum_{t=1}^T \gamma^t(k) \gamma^{tk}_u(m) w^t_u}{\sum_{t=1}^T \gamma^t(k) \gamma^{tk}_u(m)}, \tag{6}$$

and

$$(\sigma^k_{u,m})^2 = \frac{\sum_{t=1}^T \gamma^t(k) \gamma^{tk}_u(m) \left(w^t_u - \mu^k_{u,m}\right)^2}{\sum_{t=1}^T \gamma^t(k) \gamma^{tk}_u(m)}. \tag{7}$$

2.4. Multiple Observations

In several applications, we have a large number of observed signals $\mathcal{W} = \left\{\mathbf{W^1}, \mathbf{W^2}, \ldots, \mathbf{W}^P\right\}$, where each observation consists of a sequence of evidences $\mathbf{W}^p = \mathbf{w}^{p,1}, \mathbf{w}^{p,2}, \ldots, \mathbf{w}^{p,T_p}$, with $\mathbf{w}^{p,t} \in \mathbb{R}^N$. Assuming that each sequence is independent of the others, we define the auxiliary function

$$\begin{aligned}\mathcal{D}(\Theta, \bar{\Theta}) \triangleq \sum_{p=1}^P \frac{1}{\Pr\left(\mathbf{W}^p|\theta\right)} \sum_{\forall \mathbf{q}} \sum_{\forall \mathbf{R}} \mathcal{L}_\Theta(\mathbf{W}^p, \mathbf{q}, \mathbf{R}) \times \\ \times \log\left(\mathcal{L}_{\bar{\Theta}}(\mathbf{W}^p, \mathbf{q}, \mathbf{R})\right). \end{aligned} \tag{8}$$

The derivation of the reestimation formulas is similar to the single-sequence case. The reestimation formula for transition probabilities is

$$\epsilon^k_{u,mn} = \frac{\sum_{p=1}^P \sum_{t=1}^{T_p} \gamma^{p,t}(k) \xi^{p,tk}_u(m,n)}{\sum_{p=1}^P \sum_{t=1}^{T_p} \gamma^{p,t}(k) \gamma^{p,tk}_{\rho(u)}(n)}. \tag{9}$$

Analogous extensions are found for the parameters of the observation densities.

3. MODEL-BASED DENOISING

The wavelet transform allows to succesfully estimate a signal corrupted by additive white Gaussian noise by means of simple scalar transformations of individual wavelet coefficients, that is, thresholding or shrinkage. To further exploit

the structure of actual signals, we propose a model-based signal estimation approach based on the composite HMM-HMT model described in the previous section.

The method is an extension of the wavelet-domain empirical Wiener filter used in [11] and [1], which provides a conditional mean estimate for the signal coefficients given the noisy ones. Unlike them, however, the proposed method is applied in a short-term basis. Frame by frame, each local feature is extracted using a Hamming window of width N_w, shifted in steps of N_s samples. The first window begins N_o samples out (with zero padding) to avoid the information loss at the beginning of the signal. The same procedure is used to avoid information loss at the end of the signal.

In the next step, the DWT is applied to each windowed frame. In our case, we do not restrict the signal to have zero-mean wavelet coefficients. Thus, this mean is subtracted before filtering and added back before reconstruction. The structure of the filter is

$$\bar{w}_u^t = \sum_k \gamma^t(k) \sum_m \gamma_u^{tk}(m) \left(\frac{(\sigma_{u,m}^k)^2}{(h_u \tilde{\sigma}_w)^2 + (\sigma_{u,m}^k)^2} \cdot \cdot (w_u^t - \mu_{u,m}^k) + \mu_{u,m}^k \right), \quad (10)$$

where w_u^t is the noisy wavelet coefficient and $\bar{w}_u^t$ the denoised one. Note that the estimated noise deviation, $\tilde{\sigma}_w$, is multiplied by h_u, the corresponding attenuation introduced by the window in the frame analysis, subsampled as the wavelet coefficient in the node u.

In the final stage, the synthesis consists of inverting each DWT for the processed trees and add each one with the corresponding shift. Then, the inverse of the sum of all used windows is applied.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Practical issues

The reestimation formulas were implemented in logarithmic scale in order to make a more efficient computation of products and to avoid underflow errors in the probability accumulators [4]. HMTs with 2 states per node were used in all the experiments. The external models are left-to-right HMM with transitions $a_{ij} = 0 \; \forall j > i+1$. The only limitation of this architecture is that it is not possible to model sequences with less frames than states in the model, but there is not a constraint on the maximum number of frames in the sequence.

The DWT was implemented by the fast pyramidal algorithm [12], using periodic convolutions and the Daubechies-8 wavelet. Preliminary tests were carried out with other wavelets of the Daubechies and Splines families but no important differences in results were found. To avoid the border effects due to the periodic convolutions in the DWT, the first and last 8 samples in the inverted frame were not considered. Noise deviation was estimated as in [10] but taking the median of the medians in all frames:

$$\tilde{\sigma}_w = \frac{1}{0.67} \operatorname*{med}_t \left\{ \frac{1}{0.54} \operatorname*{med}_{2^{J-1} < u \leq 2^J} \left\{ |w_u^t| \right\} \right\}, \quad (11)$$

where 0.54 is the median of the Hamming window and 0.67 is a constant empirically determined from the data (see [10]). The experiments were conducted with the same test signals used in [1, 10] and many other works about wavelet denoising. In all tests, noisy signals were synthesized adding white Gaussian noise of unity variance. Performance of the method was evaluated computing the mean-squared error (MSE) between the clean and denoised signals like in [1].

Table 1. Denoising results for Doppler signals corrupted with additive white noise of variance 1.0.

N_x	N_w	N_s	N_Q	min. MSE	ave. MSE
1024	256	128	7	0.05349	0.07158
2048	512	128	10	0.04756	0.05814
4096	512	256	11	0.03248	0.04084

4.2. Tests with fixed-length signals

In a first stage, the performance of the proposed method was assessed using fixed-length sequences both for training and testing. Experiments were conducted for signals of 1024, 2048, and 4096 samples. A different model was trained for each signal length. For each case, several trials were carried out to test the impact of the most important parameters regarding signal analysis and the HMM-HMT architecture. The analysis stage was tested for $N_w \in \{128, 256, 512\}$ and $N_s \in \{64, 128, 256\}$ [1].

Table 1 and Table 2 show a summary of the best results for Doppler and Heavisine signals, respectively. Presented results are MSE averages over 10 test signals. It should be noted that these results are clearly better than all of those reported in [1] and [10]. For example, for Doppler signals of length 1024, the best MSE reported in [10] is 0.24 and for the HMT used in [1] it is 0.13. As it can be seen in Table 1, as N_x increases it is convenient to increase the window size N_w and the window step N_s in the analysis. In this experiments the number of states in the external HMM, N_Q, also grows to fit the frames in the sequence, thus avoiding modeling two frames with the same HMT. We also verified that the number of observed signals is not so important

[1] Note that not all the combinations are possible, for example, the reconstruction would be impossible if $N_w = 128$ and $N_s = 128$.

Table 2. Denoising results for Heavisine signals corrupted with additive white noise of variance 1.0.

N_x	N_w	N_s	N_Q	min. MSE	ave. MSE
1024	512	256	3	0.03595	0.05188
2048	512	256	8	0.01361	0.01889
4096	512	256	15	0.01708	0.02042

because the reestimation algorithms can extract the relevant characteristics with only three of them.

Figure 2 displays the denoising results for a realization of Doppler signal with $N_x = 1024$. For this qualitative analysis we selected the sample used to compute the average MSE in Table 1 whose MSE was closer to the average. The denoised Doppler shows that the larger errors are in the high frequencies, at the beginning of the signal. This residual noise can be explained noting that in these frames the variances of the signal and the noise are similar at the finest scale. Therefore, the estimated variances in the corresponding HMT are relatively large and application of (10) has a minor effect. Additionally, in this signal the first part of the first frame is not overlapped with other frames and thus the Hamming window has an stronger impact in the estimation of the model parameters and the reconstruction of the denoised signal. Moreover, the signal to noise ratio in this part is lower than in other regions of the signal.

The main advantage of the HMM-HMT is that it provides a set of small models that fit each region of the signal. However, it can be seen that at the end of the Doppler signal (around the sample 800) the mean in the denoised signal follows that of the noisy one. Similar errors can be seen in some other parts of the signal, for example, in the peak around the sample 350. For the Heavisine signal, a similar behavior can be observed at the beginning, the middle and the end of the signal (not shown). In this point, recall that the HMT does not model the approximation coefficient. Thus, the approximation coefficient used in the re-synthesis is the noisy one, which is never modified. Therefore, if the noise has significant low-frequency components, they will appear as remaining noise in the denoised signal.

4.3. Tests with variable-length signals

Although the forementioned experiments allow for direct comparison with other wavelet-based methods for signal estimation, they are not suitable to exploit the main advantage of the proposed model, that is, its ability to deal with variable-length sequences. In order to test for this flexibility, another set of experiments was carried out in which the length of the training signals was allow to vary over different ranges. Only Doppler signals were used for these

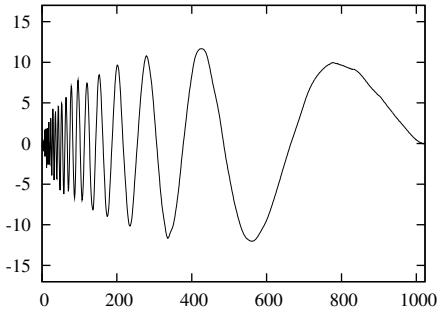

Fig. 2. Average case for denoised Doppler signal ($N_x = 1024$, MSE=0.0678).

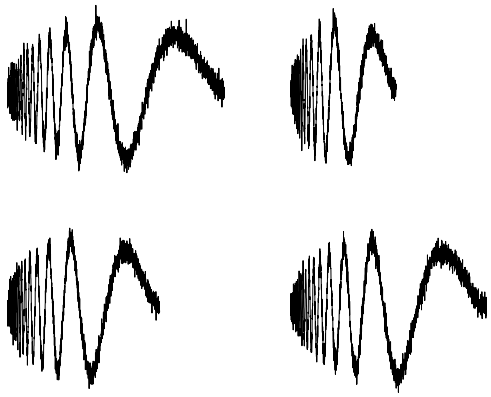

Fig. 3. Typical variability of the length of the signals used to train the model. The examples are from the set of signals with random length in $[1280, 2816]$ samples (ΔM_6).

tests. In each trial, the length of training signals was set to $2048 + \Delta M_k$ samples, where ΔM_k was randomly generated from a uniform distribution in $[-128k, +128k]$, with $k = 1, 2, \ldots, 8$. Figure 4.3 shows typical examples of signals in the training set for one range of length variation. The length of the test signals remained fixed at 2048 samples. In all these experiments, an external HMM with $N_Q = 7$ was used, and the signal analysis was done with $N_w = 256$ and $N_s = 128$. Results are shown in figure 4, averaged over 30 test signals for each range of length variation of the training signals. It can be seen that though the estimation is not as good as for the experiments where the model is trained always with fixed-length sequences, both average MSE and their related standard deviations remain fairly the same over a broad range of variability in the length of the training signals. Even more, in all cases results are found to be better than all of those wavelet-based signal estimation results reported in [10] using various threshold methods.

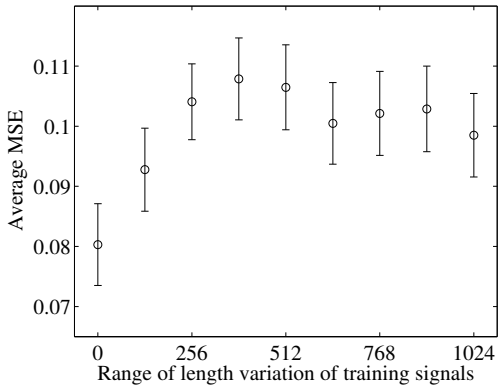

Fig. 4. Average MSE obtained training the model with variable-length Doppler signals and testing with sequences of 2048 samples. Error bars show the standard deviation of the results from the average MSE.

Table 3. Denoising results for variable-length Doppler signals corrupted with additive white noise of variance 1.0.

Range of ΔM	$N_Q = 3$	$N_Q = 5$	$N_Q = 7$
± 128	0.14593	0.11275	0.10332
± 256	0.15494	0.12266	0.10930
± 512	0.15752	0.12773	0.11963
± 1024	0.16250	0.12630	0.11350

The flexibility of the proposed method was also assessed using variable-length Doppler sequences both for training and testing. Table 3 shows average MSE values for several ranges of signal lengths and for architectures with different number of hidden states in the external HMM model. Averages are over 30 testing signals for each range of length variation. For the architecture used in the previous experiments, it can be seen that the length variability of the testing signals do not give rise to a significant degradation in performance. Results also show that models with more hidden states in the external model consistently reach a better estimation in all the tested conditions.

5. CONCLUSIONS

The proposed architecture allows to model variable-length signals in the wavelet domain. The algorithms for parameter estimation were derived using the EM framework, resulting in a set of reestimation formulas with a simple structure. Model-based denoising with the proposed method showed important qualitative and quantitative improvements over previous methods. Performance remained fairly the same over important variations in the length of both training and test signals, showing the robustness of the method. Future work will be oriented to test the proposed model for estimation of signals observed in non-stationary noise as well as for joint classification and estimation tasks.

6. REFERENCES

[1] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. on Signal Proc.*, vol. 46, no. 4, pp. 886–902, 1998.

[2] M. Borran and R. Nowak, "Wavelet-based denoising using hidden Markov models," in *Proc. of the ICASSP '2001*, Salt Lake City, UT, 2001, pp. 3925–3928.

[3] G. Fan and X.-G. Xia, "Improved hidden Markov models in the wavelet-domain," *IEEE Trans. on Signal Proc.*, vol. 49, no. 1, pp. 115–120, Jan. 2001.

[4] J.-B. Durand, P. Gonçalvès, and Y. Guédon, "Computational methods for hidden Markov trees," *IEEE Trans. on Signal Proc.*, vol. 52, no. 9, pp. 2551–2560, 2004.

[5] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, New Jersey, 1993.

[6] Y. Bengio, "Markovian Models for Sequential Data," *Neural Computing Surveys*, vol. 2, pp. 129–162, 1999.

[7] N. Dasgupta, P. Runkle, L. Couchman, and L. Carin, "Dual hidden Markov model for characterizing wavelet coefficients from multi-aspect scattering data," *Signal Proc.*, vol. 81, no. 6, pp. 1303–1316, June 2001.

[8] Jiuliu Lu and Lawrence Carin, "HMM-based multiresolution image segmentation," in *Proc. of the ICASSP '2002*, Orlando, FL, 2002, vol. 4, pp. 3357–3360.

[9] K. Weber, S. Ikbal, S. Bengio, and H. Bourlard, "Robust speech recognition and feature extraction using HMM2," *Computer Speech & Language*, vol. 17, no. 2-3, pp. 195–211, 2003.

[10] D. Donoho and I. Johnstone, "Adapting to unknown smoothness by wavelet shrinkage," *J. of the Amer. Stat. Assoc.*, vol. 90, no. 432, pp. 1200–1224, 1995.

[11] S. Ghael, A. Sayeed, and R. Baraniuk, "Improved wavelet denoising via empirical Wiener filtering," in *Proc. of SPIE*, San Diego, CA, Oct. 1997, vol. 3169, pp. 389–399.

[12] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 2nd edition, 1999.