SYSTEM IDENTIFICATION IN THE WAVELET DOMAIN WITH CROSSBAND FILTERS

Erez Cohen and Israel Cohen

Department of Electrical Engineering, Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel
{erezc@tx,icohen@ee}.technion.ac.il

ABSTRACT

In this paper, we present a wavelet domain system identification scheme, which takes into consideration the cross-terms between different frequency bands. Wavelet domain subband processing is advantageous whenever it produces a sparse representation of the processed signals. Perfect representation of linear time invariant (LTI) systems in the discrete-time wavelet transform (DTWT) domain requires time-invariant band-to-band filters and time-varying crossband filters between distnict subbands. We represent the response between two different subbands as a convolution with an appropriate multirate crossband filter. This reduces the model mismatch, which improves the identification in high SNR environment. The crossband filters formulation is extended to the two-dimensional wavelet domain. Experimental results demonstrate the advantages of crossband filters usage.

Index Terms— System identification, discrete-time wavelet transform, subband filtering, system modeling.

1. INTRODUCTION

Time-domain identification of LTI systems may suffer from high computational complexity and low convergence rate when the impulse response is very long. Identification in different frequency bands (or in a time-frequency domain) is attractive due to subsampling in each frequency band, by reducing the number of coefficients to be identified. However, a reliable identification requires an accurate modeling of the system, including band-to-band terms for each subband and cross-terms between different subbands [1]. Ignoring the cross-terms introduces aliased versions of the input signal, in the error signal [2].

In [1], crossband filters estimation is studied in the short-time Fourier transform (STFT) domain for acoustic echo cancelation. Increasing the cross-terms number does not necessarily improves the identification quality. Cross-terms number selection depends on two factors : as the input signal becomes stronger (with respect to the noise energy in the degradation model), more crossband filters may be considered. The same holds when the number of samples becomes larger (corresponds to slower time variations of a general linear system, restricting the time-invariance assumption to hold for a longer time period).

The wavelet transform enables a multiersolution analysis with variable length basis functions. This results in a sparse representation of signals in the wavelet domain, i.e. many coefficients are close to zero while most of the signal's energy is concentrated in a small number of coefficients [3]. This valuable property, especially in signals and images coding and estimation, is desirable in the subband system identification as well, since consideration of a small number of subbands that carry most of the signal's energy, significantly reduces its complexity without substantial model mismatch.

Recently, a subband identification scheme was explored, using the discrete-time wavelet transform (DTWT) domain [4]. The main difference from the STFT domain representation is the filter bank interpretation of the DTWT. The diverse lengths of the moultiresolution analysis and synthesis windows dictate a nonuniform filter bank [5]. Consequently, modeling the system in the wavelet domain requires time-varying cross-terms, while the band-to-band terms remain time-invariant. One approach to overcome this problem is to use band-to-band filters only [4]. However, in order to restrict the model-mismatch caused by disregarding the cross-terms, the analysis and synthesis windows must be long enough to produce a selective frequency bands separation. Apart from the computational cost and additional delay caused by increasing the windows length, feasible windows are incapable of producing an efficient representation.

The main contribution of this work, is the extension of the scheme in [4], by enabling an estimation of the crossband filters. Although these terms vary in time, it is possible to represent the cross-terms as convolution with an appropriate multirate filter. Estimating the multirate filter coefficients takes care of the aliasing effects caused by imperfect frequency localization of the analysis and synthesis windows. Thus, the model accuracy is improved and allows usage of shorter windows which in turn reduces the system identifier delay. In addition, we generalize the crossband filters to two-dimensional domain, using separable wavelet bases.

The paper is organized as follows. Section 2 presents a brief description of the DTWT and LTI system representation in this time-frequency domain. An emphasis is given on the

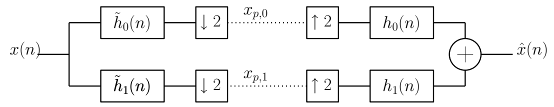


Fig. 1. 1-level PRFB with conjugate mirror filters.

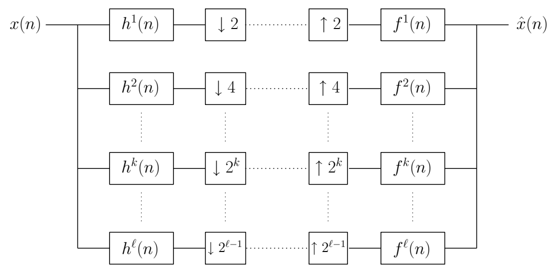
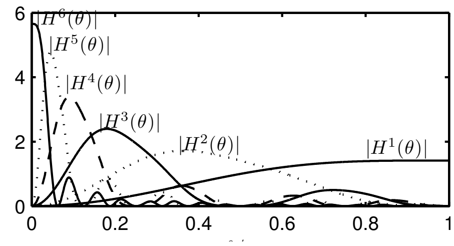
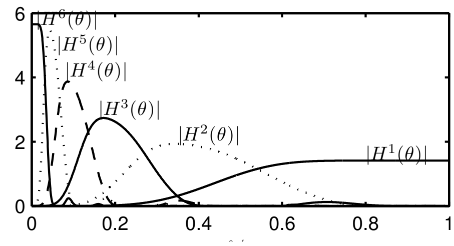


Fig. 2. DTWT nonuniform filter bank.



(a) $N = 4$



(b) $N = 12$

Fig. 3. Magnitude responses of $\{H^k\}_{k=1}^{\ell}$ for a 5 levels filter-bank using Daubechies wavelets of length N.

extraction of the crossband filters as multirate filtering. In Section 3, an identification scheme using least-squares (LS) criterion is proposed. Experimental results on real-world image data are discussed in Section 4.

2. LTI SYSTEM REPRESENTATION IN THE DTWT DOMAIN

2.1. The Discrete Time Wavelet Transform

In this section, the orthogonal discrete time wavelet transform (DTWT) and its implementation are briefly described.

Starting with a single-level decomposition, a discrete-time signal $x(n) \in \ell^2(\mathbb{Z})$ is convolved with analysis high-pass and low-pass filters, $\tilde{h}_0(n)$ and $\tilde{h}_1(n)$ respectively, then subsampled by a factor of 2 (see Fig. 1) :

$$x_{p,k} = \sum_m x(m)\tilde{h}_k(2p - m), \tag{1}$$

for $k \in \{0, 1\}$. The indices p and k represent the time index and the frequency-bin, respectively. In the reconstruction, both signals are upsampled and convolved with synthesis filters, $h_0(n) = \tilde{h}_0(-n)$ and $h_1(n) = \tilde{h}_1(-n)$:

$$\hat{x}(n) = \sum_p x_{p,0}h_0(n - 2p) + \sum_p x_{p,1}h_1(n - 2p). \tag{2}$$

A perfect reconstruction is achieved by choosing $\tilde{h}_0(n)$ and $\tilde{h}_1(n)$ as *conjugate mirror filters* [6]. A cascade of J decomposition units as described in (1), where in each level only the low-pass signal is decomposed, generates a tree-structured filter bank, having an equivalent nonuniform pattern (see Fig. 2), with $\ell = J + 1$ frequency-scale subbands and a varying multirate factor for each subband $\mu^{(k)} \triangleq 2^{\min(k,\ell-1)}$. The equivalent analysis and synthesis filters $h^k(n)$ and $f^k(n)$, acquired with the "Noble identities" [5], separate the frequency domain to a set of octave bands, as can be seen in Fig. 3. It is easy to verify that the length of each analysis or synthesis filter is $N_k = (\mu^{(k)} - 1)(N - 1) + 1$. Since the DTWT is implemented as an FIR filter bank, perfect separation is unobtainable. However, longer analysis and synthesis filters lead to better separation.

Following this structure, the decomposition and reconstruction are given by

$$x_{p,k} = \sum_m x(m)h^k(\mu^{(k)}p - m), \tag{3}$$

$$x(n) = \sum_{k=1}^{\ell}\sum_p x_{p,k}f^k(n - \mu^{(k)}p). \tag{4}$$

2.2. Perfect Representation of LTI systems in the DTWT domain

Typically, the system identification problem is associated with a model, originally formulated in time domain (or in the space domain, for two-dimensional systems). When the system is assumed to be linear time-invariant (LTI), it is fully characterized by an impulse response $a(n)$. The system's output for an input signal $x(n)$ is given by $d(n) = \sum_{i=0}^{L-1} a(i)x(n - i)$. Subband domain approach to system identification necessitates an explicit subband characterization of the LTI system. The relation between the respective wavelet coefficients of $x(n)$ and $d(n)$ is given in [4] as

$$d_{p,k} = \sum_{k'=1}^{\ell}\sum_{p'} x_{p',k'}a_{p,k,p',k'}, \tag{5}$$

where $a_{p,k,p',k'}$ is the *crossband filter* :

$$a_{n,k,k'} \triangleq a(n) * h^k(n) * f^{k'}(n), \tag{6}$$

$$a_{p,k,p',k'} = a_{n,k,k'} \mid_{n=\mu^{(k)}p-\mu^{(k')}p'} . \tag{7}$$

The filter $a_{n,k,k'}$ in (6) is termed the *undecimated crossband filter*. The response of $d_{p,k}$ to an impulse excitation on $x_{p',k'}$ does not depend on the wavelet domain time difference $p - p'$. It follows that the crossband filters are not time-invariant. This is a major difficulty in the system identification context, since (6) implies that for each time index p, a new filter should be estimated, leading to infeasible model order. Fortunately, it is possible to exploit the fact that the relation between $x_{p',k'}$ and $d_{p,k}$ is always derived from the same undecimated crossband filter. Furthermore, an explicit relation exists in the form of multirate filtering with a downsampled version of $a_{n,k,k'}$. By substituting (7) into (5) we get

$$d_{p,k} = \sum_{k'=1}^{\ell} \sum_{p'} x_{p',k'} a_{\mu^{(k)}p-\mu^{(k')}p',k,k'}. \tag{8}$$

Assume that $k, k' < \ell$ (This assumption is made only for keeping the equations below readable). In this case, $\mu^{(k)} = 2^k$ and (8) reduces to :

$$d_{p,k} = \sum_{k'=1}^{\ell} \sum_{p'} x_{p',k'} a_{2^k p-2^{k'}p',k,k'} = \sum_{k'=1}^{\ell} d_{p,k,k'}, \tag{9}$$

where $d_{p,k,k'}$ is the contribution of frequency band k' to frequency band k. When $k = k'$, both $x_{p,k'}$ and $d_{p,k}$ are defined at the same resolution :

$$d_{p,k,k} = x_{p,k} * a_{2^k p,k,k}. \tag{10}$$

When $k > k'$, the relation is described as a convolution with $2^{k-k'}$-fold decimation filter :

$$d_{p,k,k'} = \sum_{p'} x_{p',k'} a_{2^{k'}(2^{k-k'}p-p'),k,k'}. \tag{11}$$

Finally, when $k < k'$, we have :

$$d_{p,k,k'} = \sum_{p'} x_{p',k'} a_{2^k(p-2^{k'-k}p'),k,k'}. \tag{12}$$

Considering the case where $k = \ell$ or $k' = \ell$ is possible by substituting 2^k and $2^{k'}$ with $\mu^{(k)}$ and $\mu^{(k')}$, respectively, in equations (10)-(12). Generalizing the three cases discussed above, we define the *decimated crossband filter*

$$\bar{a}_{p,k,k'} = a_{n,k,k'} \mid_{p=\min(\mu^{(k)},\mu^{(k')})n}, \tag{13}$$

to get :

$$d_{p,k,k'} = \begin{cases} x_{p,k'} * \bar{a}_{p,k,k'} & \mu^{(k)} = \mu^{(k')} \\ x_{q,k'} * \bar{a}_{q,k,k'} \mid_{q=\frac{\mu^{(k)}}{\mu^{(k')}}p} & \mu^{(k)} > \mu^{(k')} \\ \breve{x}_{p,k'} * \bar{a}_{p,k,k'} & \mu^{(k)} < \mu^{(k')}, \end{cases} \tag{14}$$

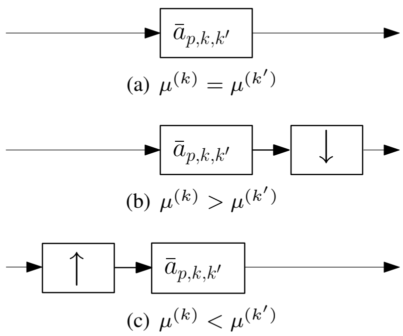


Fig. 4. *Different multirate operations depend on the ratio between the scales $\mu^{(k)}$ and $\mu^{(k')}$.*

where $\breve{x}_{p,k}$ is the upsampled version of $x_{p,k}$ by factor $\frac{\mu^{(k')}}{\mu^{(k)}}$.

The contribution pattern from the k''th subband to the k'th subband is a multirate operation, obeys the scales ratio between them. In Fig. 4, the different multirate operations appear, with respect to each case. From (7), the length of $a_{n,k,k'}$ is $N_k + N_{k'} + L - 2$ with $N_{k'} - 1$ noncausal coefficients. Therefore, the length of $\bar{a}_{p,k,k'}$ is

$$N_{k,k'} = \left\lfloor \frac{N_{k'} - 1}{\min(\mu^{(k)}, \mu^{(k')})} \right\rfloor + \left\lfloor \frac{N_k + L - 2}{\min(\mu^{(k)}, \mu^{(k')})} \right\rfloor + 1. \tag{15}$$

A reasonable assumption when working in subband processing, is that the LTI system's impulse response length is very long comparing to any analysis or synthesis window, i.e. $L >> N_k, \quad 1 \leq k \leq l$. Then, (15) reduces to

$$N_{k,k'} \approx \frac{L}{\min(\mu^{(k)}, \mu^{(k')})}. \tag{16}$$

For achieving a full representation of the LTI system, there is no need to fully estimate the crossband filters. The decimated crossband filter coefficients, defined in (13), are sufficient. The coefficients number in (16) for all the band-to-band filters and crossband filters dictate the model order and its complexity.

2.3. Extension to two-dimensional systems

Extension to wavelet representation of two dimensional systems, given by

$$d(n_1, n_2) = \sum_{i_1,i_2=0}^{L-1} a(i_1, i_2) x(n_1 - i_1, n_2 - i_2), \tag{17}$$

is immediate when using separable wavelet bases [3]. A single level decomposition of a two-dimensional signal $x(n_1, n_2) \in \ell^2(\mathbb{Z}^2)$ is described by a quadratic filter-bank, in which the signal is decomposed to 4 subbands in the two-dimensional frequency domain (see Fig. 5). The analysis filters in (1) construct the two-dimensional filters:

$$\tilde{h}_{i_1,i_2}(n_1, n_2) \triangleq \tilde{h}_{i_1}(n_1)\tilde{h}_{i_2}(n_2) \tag{18}$$

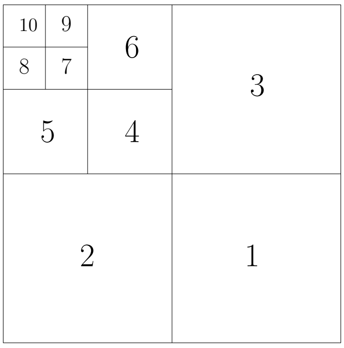


Fig. 5. Two-dimensional frequency domain, schematically separated by the different subbands. The number in each square corresponds to subband index k.

for $i_1, i_2 \in \{0,1\}$. The same holds for the synthesis filters. For a cascade of J decompositions levels, the equivalent structure is a nonuniform filter-bank with $\ell = 3J+1$ frequency subbands. In the wavelet domain, the system in (17) is modeled using two-dimensional crossband filters:

$$d_{p_1,p_2,k} = \sum_{k'=1}^{\ell} \sum_{p'_1,p'_2} x_{p'_1,p'_2,k'} a_{p_1,p_2,k,p'_1,p'_2,k'}, \quad (19)$$

$$a_{p_1,p_2,k,p'_1,p'_2,k'} \triangleq a_{n_1,n_2,k,k'} \mid_{n_i=\mu^{(k)}p_i-\mu^{(k')}p'_i \quad i=1,2}, \quad (20)$$

$$a_{n_1,n_2,k,k'} = a(n_1,n_2) * h^k(n_1,n_2) * f^{k'}(n_1,n_2). \quad (21)$$

The multirate filtering is similar to the one-dimensional case, where the upsampling or downsampling factor is equal in both dimensions.

3. ESTIMATION OF THE CROSSBAND FILTERS

In this section, we describe a system identification scheme, relying on the estimation of the band-to-band filters and the crossband filters. Consider the LTI system presented in (17), under the presence of an additive noise $\xi(n_1,n_2)$:

$$y(n_1,n_2) = x(n_1,n_2) * a(n_1,n_2) + \xi(n_1,n_2). \quad (22)$$

On the DTWT domain, (22) is represented by

$$y_{p_1,p_2,k} = \sum_{k'=1}^{\ell} d_{p_1,p_2,k,k'} + \xi_{p_1,p_2,k}. \quad (23)$$

From (14), it is possible to formulate a vector-matrix representation to $d_{p_1,p_2,k,k'}$. Let $\bar{\mathbf{a}}_{k,k'}$ and $\mathbf{d}_{k,k'}$ be the lexicographically orderings of the undecimated crossband filter from k' to k and of $d_{p_1,p_2,k,k'}$, respectively. Since, $d_{p_1,p_2,k,k'}$ is a result of a linear operation, for each $1 \leq k, k' \leq \ell$, the vector $\mathbf{d}_{k,k'}$ can be can be expressed as :

$$\mathbf{d}_{k,k'} = \mathbf{X}_{k,k'} \bar{\mathbf{a}}_{k,k'}. \quad (24)$$

When $\mu^{(k)} = \mu^{(k')}$, $\mathbf{X}_{k,k'}$ is a Block-Toeplitz matrix of size $((P_k+L-1)^2 \times N^2_{k,k'})$ with $x_{p'_1,p'_2,k'}$ on its columns [1],[4]. When $\mu^{(k)} > \mu^{(k')}$, the convolved signal is downsampled, equivalent to downsampling the columns of $\mathbf{X}_{k,k'}$. Note that the downsampling is performed on a noncausal signal, and therefore its zero coefficient should be kept. When $\mu^{(k)} < \mu^{(k')}$, the input signal $x_{p'_1,p'_2,k'}$ is upsampled before convolution, and therefore $\mathbf{X}_{k,k'}$ is built as a Block-Toeplitz matrix with the upsampled input on its columns.

Similarly to [1], concatenation of the vectors $\{\bar{\mathbf{a}}_{k,k'}\}_{k'=1}^{\ell}$ and the matrices $\{\mathbf{X}_{k,k'}\}_{k'=1}^{\ell}$:

$$\bar{\mathbf{a}}_k = \begin{pmatrix} \bar{\mathbf{a}}^T_{k,1} & \dots & \bar{\mathbf{a}}^T_{k,\ell} \end{pmatrix}^T$$

$$\mathbf{X}_k = \begin{pmatrix} \mathbf{X}_{k,1} & \dots & \mathbf{X}_{k,\ell} \end{pmatrix},$$

leads to the following vector-matrix representation of $\mathbf{d}_k$:

$$\mathbf{d}_k = \mathbf{X}_k \bar{\mathbf{a}}_k. \quad (25)$$

It is important to note that not only that the decimated crossband filters are noncasual, but also the number of noncausal coefficients differs on each frequency band. In order to compensate it, an appropriate artificial delay is set to each matrix $\{\mathbf{X}_{k,k'}\}_{k'=1}^{\ell}$. The delay is related to the finest scales subbands (determining the total delay of the system identifier). In addition, zero padding may be added for compenstating the unequal matrices size. Estimation of the decimated crossband filters is performed using least-squares (LS) criterion :

$$\hat{\bar{\mathbf{a}}}_k = \arg\min_{\bar{\mathbf{a}}_k} \|\mathbf{y}_k - \mathbf{X}_k \bar{\mathbf{a}}_k\|^2 = \left(\mathbf{X}^H_k \mathbf{X}_k\right)^{-1} \mathbf{X}^H_k \mathbf{y}_k. \quad (26)$$

Estimating $\bar{\mathbf{a}}_k$ as in (26) may result in severe over-fitting, when the data amount is limited, or in low SNR. In the STFT domain, it is proposed to perform the estimation for each subband, with a fixed number of crossband filters from adjacent subbands [1]. This is motivated by the fact that the crossband filters energy decreases as $\mid k - k' \mid$ increases. The same holds in the two-dimensional DTWT domain [4] with numerous disparities. First, in the two-dimensional domain, $\mid k - k' \mid$ is no longer a measure for adjacency between subbands (see Fig. 5). Second, in this nonuniform structure there is no justification for a fixed number of crossbands, since the overlap between subbands is not fixed as well. A generalized reduced model, allowing much more flexibility is :

$$\tilde{d}_{p_1,p_2,k} = \sum_{k' \in \Omega_k} \sum_{p'_1,p'_2} x_{p'_1,p'_2,k'} a_{p_1,p_2,k,p'_1,p'_2,k'}, \quad (27)$$

for $k = 1, \dots, \ell$, where $\Omega_k \subset \{1, \dots, \ell\}$ is the frequency bands subset that contribute to frequency band k. For example, the selection $\Omega_k = \{k\}$ is associated with the band-to-band model of [4]. The LS estimation is based only on crossband filters $\bar{\mathbf{a}}_{\mathbf{k},\mathbf{k'}}$ such that $k' \in \Omega_k$. If for certain subband k, $\Omega_k = \emptyset$, then none of the crossbands that contribute to k is estimated. On some setup, this may be a legitimate choice.

Another concept, relevant to the DTWT domain, is the sparse representation of many applicative signals [3]. For example, it is known that for real-world images, most of the energy is concentrated in the low frequencies. Therefore, it is possible to exploit such a priori knowledge, and define a nonuniform model:

$$\Omega_k = \begin{cases} \{k' \mid \mu^{(k)} = 2^J\} & \mu^{(k)} = 2^J \\ \{k+3\} & \mu^{(k)} < 2^J. \end{cases} \tag{28}$$

The nonuniform model distinguishes the low-frequency subbands by involving all the cross-terms related to the coarsest scales. In the high-frequency subbands, only a single band contributes, from a coarser resolution and the same orientation (see Fig. 5). The discrimination above is justified mainly from two reasonings. First, the SNR is very low in the high-frequency subbands. Therefore, including them in the estimation will eminently increase the variance. Second, due to the small decimation factor, the high frequency bands contain a considerable amount of data (see (16)), increasing the complexity in the pseudo-inverse calculation of (26).

4. EXPERIMENTAL RESULTS

This section compares the nonuniform model with the band-to-band and the full-band models.

We assume an LSI system whose impulse response is uniform of size 25×25. The additive white Gaussian noise has a variance σ_n^2, set to establish a desired blurred image SNR $\left(BSNR = \frac{\|\mathbf{d}-\bar{\mathbf{d}}\|^2}{\sigma_n^2}\right)$, independent of the original image. The DTWT-based system identifier used here consists of 3 decomposition levels. We define the identification error on a mean-square-error (MSE) term :

$$MSE = \frac{\|\mathbf{d} - \mathbf{X}\hat{\mathbf{a}}\|^2}{\|\mathbf{d}\|^2}. \tag{29}$$

It is well known that the MSE consist of two elements : the bias error, stems from inaccurate modeling of the system, and the variance error, stems from consideration of noisy data in the identification process.

Fig. 6 shows the MSE curves as a function of the input SNR for different identification configurations. First, the decomposition is performed using *Daubechies* wavelets of length $N = 4$ and $N = 12$. The identification is based on the band-to-band filters only and ignores the crossband filters. Next, the nonuniform model is applied for the same BSNR range and wavelet length $N = 4$. Obviously, the bias error is dominant when the input SNR is high enough [1]. It can be seen, that the use of the crossband filters reduces the bias error by $11dB$, while increasing the wavelet length decreases the bias error by only $9dB$. On the other-hand, estimating the crossband filters in low SNR environment is not effective and results in increased variance error. This result settles with the analysis presented in [1].

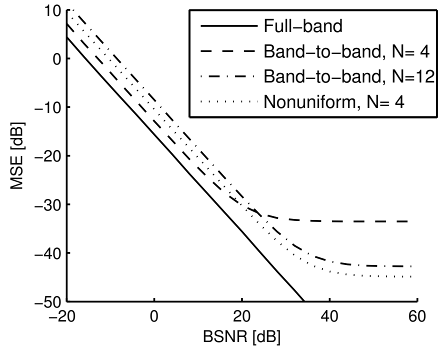


Fig. 6. MSE curves of different system identification configurations.

5. SUMMARY AND CONCLUSIONS

We have considered cross band filters estimation in the DTWT domain. We showed that modeling with crossband filters significantly reduces the identification bias error. In addition, a priori knowledge of the signal's properties, is used for neglecting cross-terms related to low SNR subbands, thus reducing the variance error. Identification in the wavelet domain has further challenges, like specification of adaptive wavelet bases for signals and systems with known spectral properties, and identification of time-varying systems.

6. REFERENCES

[1] Y. Avargel and I. Cohen, "System identification in the short-time Fourier transform domain with crossband filtering," *IEEE Trans. Audio, Speech, Lang. Process*, vol. 15, no. 4, pp. 1305–1319, May 2007.

[2] A. Gilloire and M. Vetterli, "Adaptive filtering in subbands with critical sampling: Analysis, experiments, and application to acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 40, no. 8, pp. 1862–1875, Aug. 1992.

[3] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[4] Y. Avargel and I. Cohen, "Representation and identification of systems in the wavelet transform domain," in *Proc. IASTED Int. Conf. Applied Simulation and Modelling (ASM)*, Palma De Mallorca, Spain, Aug. 2007, pp. 463–468.

[5] P. Vaidyanathan, *Multirate systems and filter banks*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1993.

[6] M. Vetterli and C. Herley, "Wavelets and filter banks: theory and design," *IEEE Trans. Signal Process.*, vol. 40, no. 9, pp. 2207–2232, Sep. 1992.