

Matching Pursuit and Atomic Signal Models Based on Recursive Filter Banks

Michael M. Goodwin, *Member, IEEE* and Martin Vetterli, *Fellow, IEEE*

Abstract—The matching pursuit algorithm can be used to derive signal decompositions in terms of the elements of a dictionary of time–frequency atoms. Using a structured overcomplete dictionary yields a signal model that is both parametric and signal adaptive. In this paper, we apply matching pursuit to the derivation of signal expansions based on damped sinusoids. It is shown that expansions in terms of complex damped sinusoids can be efficiently derived using simple recursive filter banks. We discuss a subspace extension of the pursuit algorithm that provides a framework for deriving real-valued expansions of real signals based on such complex atoms. Furthermore, we consider symmetric and asymmetric two-sided atoms constructed from underlying one-sided damped sinusoids. The primary concern is the application of this approach to the modeling of signals with transient behavior such as music; it is shown that time–frequency atoms based on damped sinusoids are more suitable for representing transients than symmetric Gabor atoms. The resulting atomic models are useful for signal coding and analysis modification synthesis.

Index Terms—Damped sinusoids, matching pursuit, overcomplete expansions, pre-echo, recursive filters, signal models, time–frequency atoms, wavelets.

I. INTRODUCTION

TIME–FREQUENCY atomic signal decompositions have been of ongoing interest since their introduction by Gabor [1], [2]. Basis expansions, especially orthogonal cases such as Fourier and wavelet bases, are the most common examples of such atomic models, but these exhibit serious drawbacks for modeling arbitrary signals; for instance, the Fourier method provides a poor representation of time-localized signals. To overcome such difficulties, signals can be modeled using overcomplete sets of atoms that exhibit a wide range of time–frequency behaviors [3]–[6].

Overcomplete expansions allow for compact representation of arbitrary signals for the sake of compression or analysis [6], [7]. Examples include best basis methods and adaptive wavelet packets. In these approaches, the overcomplete set is a collection of bases; a basis for the signal expansion is chosen from the set of bases according to a metric such as

Manuscript received May 27, 1998; revised December 16, 1998. This work was supported by the Department of Electrical Engineering and Computer Science, University of California, Berkeley. The associate editor coordinating the review of this paper and approving it for publication was Dr. Hitoshi Kiya.

M. M. Goodwin is with SGI, Mountain View, CA 94043 USA (e-mail: mgoodwin@alum.mit.edu).

M. Vetterli is with the Laboratoire de Communications Audiovisuelles, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, and the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720 USA (e-mail: vetterli@de.epfl.ch).

Publisher Item Identifier S 1053-587X(99)04674-7.

entropy or rate distortion [8]–[10]. Signal decomposition using more general overcomplete sets has also been considered; approaches include the method of frames [11], basis pursuit [5] FOCUSS [12], and matching pursuit and its variations [3], [4], [6], [13]. The matching pursuit algorithm of [3] is the focus of this paper since it is particularly amenable to the issue of interest here: modeling of arbitrary signals using parameterized time–frequency atoms in a successive refinement framework.

In the literature, the time–frequency atoms used in matching pursuit are typically symmetric Gabor atoms [3], [14]. While these are useful for many applications, such symmetric functions are not well suited for modeling asymmetric events. For instance, applying matching pursuit to a musical signal with a sharp onset introduces an objectionable artifact known as *pre-echo*; this refers to energy in the reconstruction before the original onset and is a common problem in audio coding [6]. High-resolution matching pursuit addresses this issue but fundamentally still relies on symmetric atoms for representing asymmetric features [13], [14]. In this paper, the problem is addressed by using asymmetric atoms, namely, damped sinusoids and related functions.

Several approaches in the literature have dealt with time–frequency atoms having exponential behavior. In [15], frames of damped sinusoids are used to construct a time–frequency representation that is useful for transient detection and characterization. In [16], IIR filter banks that derive orthogonal wavelet expansions are considered. This paper focuses on the more general case of overcomplete expansions based on recursive filter banks. Signal analysis based on similar recursive computation has been considered; short-time Fourier transforms using exponential windows are discussed in [17], and more arbitrary short-time transforms are examined in [18].

This paper is organized as follows. In Section II, signal decompositions are discussed; it is demonstrated that overcomplete expansions provide advantages over basis expansions and that computation of overcomplete models calls for a signal-adaptive approach. In Section III, the matching pursuit algorithm for deriving overcomplete expansions is reviewed and extended. Section IV discusses time–frequency dictionaries for signal modeling, namely, symmetric Gabor dictionaries, dictionaries of damped sinusoids, and composite dictionaries consisting of both symmetric and asymmetric atoms. Section V presents an efficient algorithm for matching pursuit with damped sinusoids; the computation is based on simple recursive filters. Section VI extends this method to two-sided atoms constructed by coupling causal and anticausal damped

sinusoids. Computational costs are examined in Section VII, and conclusions are given in Section VIII. Some of the results herein have been presented in preliminary form in [6], [19], and [20].

II. SIGNAL DECOMPOSITIONS

In signal processing applications it is often useful to decompose a signal into elementary building blocks. In such a decomposition, a signal $x[n]$ is represented as a linear combination of expansion functions $d_m[n]$

$$x[n] = \sum_{m=1}^M \alpha_m d_m[n] \quad (1)$$

which can be expressed in matrix notation as

$$x = D\alpha \quad \text{with } D = [d_1 \quad d_2 \quad \cdots \quad d_m \quad \cdots \quad d_M] \quad (2)$$

where the signal x is a column vector ($N \times 1$), α is a column vector of expansion coefficients ($M \times 1$), and D is an $N \times M$ matrix whose columns are the expansion functions $d_m[n]$.

The set of expansion coefficients and functions in (1) provide a representation or model of the signal. If the model is compact or sparse, the decomposition indicates basic signal features and is useful for signal analysis, compression, and enhancement [6]. It should be noted that compact models tend to involve expansion functions that are highly correlated with the signal.

A. Basis Expansions

When the functions $d_m[n]$ constitute a basis, the matrix D in (2) is square ($N = M$) and invertible, and the expansion coefficients α for a signal x are uniquely given by $\alpha = D^{-1}x$. In general biorthogonal cases such as wavelets, there is a dual basis \tilde{D} such that $D^{-1} = \tilde{D}^H$ and $\alpha = \tilde{D}^H x$, which indicates that the coefficients in a basis expansion can be derived independently using the formula $\alpha_m = \tilde{d}_m^H x = \langle \tilde{d}_m, x \rangle$. In orthogonal cases such as Fourier bases, $\tilde{D} = D$, and the expansion coefficients are given simply by the correlations $\langle d_m, x \rangle$.

Basis expansions have a serious drawback in that a given basis is not well-suited for modeling a wide variety of signals. For example, the Fourier basis does not provide a compact model of a time-localized signal; similar difficulties can be readily found for any basis. This shortcoming of basis expansions results from the attempt to represent arbitrary signals using a limited set of functions. Better models can be derived by using expansion functions that are signal adaptive; this can be achieved by using a parametric approach such as the sinusoidal model [6], [20] or by choosing the expansion functions in a signal-dependent fashion from an overcomplete set of time–frequency atoms as in adaptive wavelet packets or matching pursuit [3], [6], [10]. The term *overcomplete* refers to a set of vectors that spans the signal space but includes more functions than is necessary to do so, i.e. a linearly dependent set; the term *complete* refers simply to any set that spans the space. Using a highly overcomplete set of time–frequency atoms enables compact representation of a wide range of time–frequency behaviors.

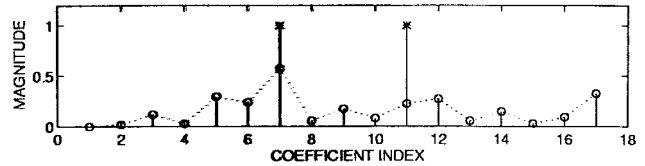


Fig. 1. Overcomplete expansions and compaction. An exact sparse expansion of a signal in an overcomplete set (*) and the dispersed expansion given by the SVD pseudo-inverse (o).

B. Overcomplete Expansions

When the functions $d_m[n]$ constitute an overcomplete or redundant set ($M > N$), the matrix D is of rank N , and the linear system in (2) is underdetermined. The null space of D then has nonzero dimension, and there are an infinite number of expansions of the form of (1). One solution is given by the pseudo-inverse D^+ , which can be derived using the singular value decomposition (SVD); the coefficient vector $\alpha = D^+x$ has the minimum two-norm of all possible solutions. The caveat is that this minimization of the two-norm tends to spread energy throughout all of the vector's elements, which undermines the goal of compaction; the SVD-based approach does not yield compact signal models.

An example of the dispersion of the SVD approach is given in Fig. 1. The signal in question is constructed as the sum of two functions from an overcomplete set; there is thus an expansion in that overcomplete set with only two nonzero coefficients. This sparse expansion is shown in the plot by the asterisks; the dispersed expansion computed using the SVD pseudo-inverse is indicated by the circles. The representations can be immediately compared with respect to two applications: First, the sparse model is clearly more appropriate for compression; second, it provides a more useful analysis of the signal in that it identifies fundamental signal structures.

Given the desire to derive compact representations for signal analysis, coding, denoising, and modeling in general, the SVD is not a particularly useful tool. The SVD-based expansion is by nature not sparse, and thresholding small expansion coefficients to improve the sparsity is not a useful approach [6], [19], [21]. A more appropriate paradigm for deriving an overcomplete expansion is to apply an algorithm specifically designed to arrive at sparse solutions. Because of the complexity of the search, however, it is not computationally feasible to derive an optimal sparse expansion that perfectly models a signal. It is likewise not feasible to compute approximate sparse expansions that minimize the error for a given sparsity; this is an NP-hard problem [22]. For this reason, it is necessary to narrow the considerations to methods that either derive sparse approximate solutions according to suboptimal criteria or derive exact solutions that are not optimally sparse. Methods of the latter type tend to be computationally costly and to lack an effective successive refinement framework [5], [12]; therefore, the former category is of more interest here. Specifically, the matching pursuit algorithm introduced in [3] is the method of choice here since it provides a framework for deriving sparse approximate models with successive refinements and since it can be carried out at low cost as will be seen.

III. MATCHING PURSUIT

Matching pursuit is a greedy iterative algorithm for deriving signal decompositions in terms of expansion functions chosen from a *dictionary* [3], [4]. To achieve compact representation of arbitrary signals, it is necessary that the dictionary elements or *atoms* exhibit a wide range of time–frequency behaviors and that the appropriate atoms from the dictionary be chosen to decompose a particular signal. When a well-designed overcomplete dictionary is used in matching pursuit, the nonlinear nature of the algorithm leads to compact signal-adaptive models [6], [7].

A dictionary can be likened to the matrix D in (2) by considering the atoms to be the matrix columns; then, matching pursuit can be interpreted as an approach for computing sparse approximate solutions to inverse problems [6], [19], [21]. Related approximation methods have indeed been used in linear algebra for some time [21]. Furthermore, matching pursuit is similar to some forms of vector quantization and is related to the projection pursuit method investigated in the field of statistics for the task of finding compact models of data sets [23].

A. One-Dimensional Pursuit

The greedy iteration in matching pursuit is carried out as follows. Using the two-norm as the approximation metric because of its mathematical convenience, the atom that best approximates the signal is chosen; the contribution of this atom is subtracted from the signal, and the process is iterated on the residual. Denoting the dictionary by D , the task at the i th stage of the algorithm is to find the atom $d_{m(i)}[n] \in D$ that minimizes the two-norm of the residual

$$r_{i+1}[n] = r_i[n] - \alpha_i d_{m(i)}[n] \quad (3)$$

where α_i is the expansion coefficient for the atom, and $m(i)$ is its dictionary index; note that the iteration begins with $r_1[n] = x[n]$. To simplify the notation, the atom chosen at the i th stage is hereafter referred to as $g_i[n]$, where

$$g_i[n] = d_{m(i)}[n] \quad (4)$$

from (3). The subscript i refers to the iteration when $g_i[n]$ was chosen, whereas $m(i)$ is the actual dictionary index of $g_i[n]$.

Treating the signals as column vectors, the optimal atom to choose at the i th stage is

$$g_i = \arg \min_{g_i \in D} \|r_{i+1}\|^2 = \arg \min_{g_i \in D} \|r_i - \alpha_i g_i\|^2. \quad (5)$$

The orthogonality principle gives the value of α_i as

$$\langle r_{i+1}, g_i \rangle = \langle r_i - \alpha_i g_i, g_i \rangle = (r_i - \alpha_i g_i)^H g_i = 0 \quad (6)$$

$$\Rightarrow \alpha_i = \frac{\langle g_i, r_i \rangle}{\langle g_i, g_i \rangle} = \frac{\langle g_i, r_i \rangle}{\|g_i\|^2} = \langle g_i, r_i \rangle \quad (7)$$

where the last step follows from restricting the atoms to be unit-norm. The norm of $r_{i+1}[n]$ can then be expressed as

$$\|r_{i+1}\|^2 = \|r_i\|^2 - \frac{|\langle g_i, r_i \rangle|^2}{\|g_i\|^2} = \|r_i\|^2 - |\alpha_i|^2 \quad (8)$$

which is minimized by maximizing $|\alpha_i|^2 = |\langle g_i, r_i \rangle|^2$. This is simply equivalent to choosing the atom with the largest

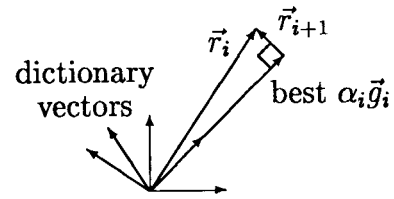


Fig. 2. Matching pursuit and the orthogonality principle. The two-norm or Euclidean length of r_{i+1} is minimized by choosing g_i to maximize $|\langle g_i, r_i \rangle|$ and α_i such that $\langle r_{i+1}, g_i \rangle = 0$.

magnitude correlation $|\alpha_i|$; therefore, (5) can be rewritten as

$$g_i = \arg \max_{g_i \in D} |\langle g_i, r_i \rangle|. \quad (9)$$

An example of this optimization is illustrated in Fig. 2. Note that (8) shows that the norm of the residual decreases as the algorithm progresses, provided that an exact model has not been reached and that the dictionary is complete; for an undercomplete dictionary, the residual may belong to a subspace that is orthogonal to all of the dictionary vectors, in which case, the model cannot be further improved by pursuit.

In deriving a signal decomposition, the matching pursuit is iterated until the residual energy is below some threshold or until some other halting criterion is met. After I iterations, the pursuit provides the sparse approximate model

$$x[n] \approx \sum_{i=1}^I \alpha_i g_i[n] = \sum_{i=1}^I \alpha_i d_{m(i)}[n]. \quad (10)$$

According to (8), the mean-squared error of such a model approaches zero as the number of iterations increases [3]. This convergence implies that I iterations will yield a reasonable I -term model; this model, however, is in general not optimal in the mean-squared sense because of the term-by-term greediness of the algorithm.

To enable representation of a wide range of signal features, a large dictionary of time–frequency atoms is used in the matching pursuit algorithm. The computation of the correlations $\langle g, r_i \rangle$ for all $g \in D$ is, thus, costly. As derived in [3], this computation can be substantially reduced using an update formula based on (3); the correlations at stage $i+1$ are given by

$$\langle g, r_{i+1} \rangle = \langle g, r_i \rangle - \alpha_i \langle g, g_i \rangle \quad (11)$$

where the only new computation required for the correlation update is the dictionary cross-correlation term $\langle g, g_i \rangle$, which can be precomputed and stored if enough memory is available.

B. Subspace Pursuit

Although searching for the optimal high-dimension subspace is not computationally reasonable, it is worthwhile to consider the related task of finding an optimal low-dimension subspace at each pursuit iteration if the subspaces under consideration have a simplifying structure. In subspace pursuit, the i th iteration consists of searching for an $N \times R$ matrix G , whose R columns are dictionary atoms, that minimizes the two-norm of the residual $r_{i+1} = r_i - G\alpha$, where α is now an $R \times 1$ vector of weights. This R -dimensional formulation is similar to the one-dimensional (1-D) case; the orthogonality

constraint $\langle r_i - G\alpha, G \rangle = 0$ yields a solution for the weights

$$\alpha = (G^H G)^{-1} G^H r_i. \quad (12)$$

The energy of the residual is then given by

$$\langle r_{i+1}, r_{i+1} \rangle = \langle r_i, r_i \rangle - r_i^H G (G^H G)^{-1} G^H r_i \quad (13)$$

which is minimized by choosing G to maximize the second term. This approach is clearly expensive unless G consists of orthogonal vectors or has some other special structure.

C. Conjugate Subspaces

One useful subspace to consider is the two-dimensional subspace spanned by an atom and its complex conjugate. Here, the two columns of G are simply an atom g and its conjugate g^* . If the signal r_i is real and if g has nonzero real and imaginary parts so that G has full column rank and $G^H G$ is invertible, the results given above can be simplified. The metric to maximize through the choice of g , i.e. the second term in (13), can be rewritten as

$$\frac{1}{1 - |\langle g, g^* \rangle|^2} (2|\langle g, r_i \rangle|^2 - \langle g, g^* \rangle \langle \langle g, r_i \rangle^* \rangle^2 - \langle g, g^* \rangle^* \langle g, r_i \rangle^2) \quad (14)$$

and the optimal weights are

$$\alpha = \begin{bmatrix} \alpha(1) \\ \alpha(2) \end{bmatrix} = \frac{1}{1 - |\langle g, g^* \rangle|^2} \begin{bmatrix} \langle g, r_i \rangle - \langle g, g^* \rangle \langle g, r_i \rangle^* \\ \langle g, r_i \rangle^* - \langle g, g^* \rangle^* \langle g, r_i \rangle \end{bmatrix}. \quad (15)$$

Note that the above metric can also be written as

$$\langle g, r_i \rangle^* \alpha(1) + \langle g, r_i \rangle \alpha(1)^* = 2 \operatorname{Re} \{ \langle g, r_i \rangle^* \alpha(1) \} \quad (16)$$

and that $\alpha(1) = \alpha(2)^*$, meaning that the algorithm simply searches for the atom g_i that minimizes the two-norm of the residual

$$\begin{aligned} r_{i+1}[n] &= r_i[n] - \alpha_i(1)g_i[n] - \alpha_i(1)^*g_i^*[n] \\ &= r_i[n] - 2 \operatorname{Re} \{ \alpha_i(1)g_i[n] \} \end{aligned} \quad (17)$$

which is real valued; the orthogonal projection of a real signal onto the subspace spanned by a conjugate pair is again real. Using such conjugate subspaces yields decompositions of the form

$$x \approx 2 \sum_{i=1}^I \operatorname{Re} \{ \alpha_i(1)g_i[n] \}. \quad (18)$$

This approach thus provides real decompositions of real signals using an underlying complex dictionary. A similar notion based on a different computational framework is discussed in [3].

For dictionaries consisting of both complex and purely real (or purely imaginary) atoms, the real atoms must be considered independently of the various conjugate subspaces since the above formulation breaks down when g and g^* are linearly dependent; in that case, $|\langle g, g^* \rangle| = 1$ and the matrix G is singular. It is thus necessary to compare metrics of the form given in (14) and (16) for conjugate subspaces with metrics of the form $|\langle g, r_i \rangle|^2$ for real atoms. These metrics quantify the amount of energy removed from the residual in either case and, thus, provide for a fair choice between conjugate subspaces and real atoms in the pursuit decomposition.

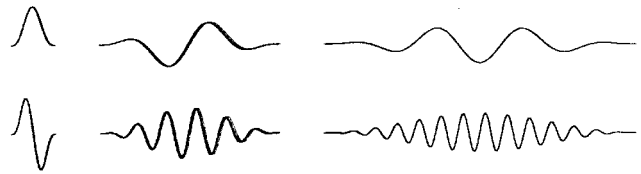


Fig. 3. Time–frequency dictionary elements: Gabor atoms derived from a symmetric window.

IV. TIME–FREQUENCY DICTIONARIES

In a compact model, the atoms in the expansion correspond to basic signal features. This is especially useful for analysis and coding if the atoms can be described by meaningful parameters such as time location, frequency modulation, and scale; then, the basic signal features can be identified and parameterized. Matching pursuit using a large dictionary of such atoms provides a compact, adaptive, parametric time–frequency representation of a signal [3], [4], [6]. Several types of dictionaries are discussed below.

A. Gabor Atoms

Localized time–frequency atoms were introduced by Gabor from a theoretical standpoint and according to psychoacoustic motivations [1], [2]. The literature on matching pursuit has focused on using dictionaries of Gabor atoms since these are generally appropriate components for time–frequency signal models [3], [4]. Such atoms are derived from a single unit-norm function $g(t)$ by scaling, modulation, and translation as

$$g_{\{s,\omega,\tau\}}(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-\tau}{s}\right) e^{j\omega(t-\tau)}. \quad (19)$$

This definition can be extended to discrete time by a sampling argument as in [3]; fundamentally, the extension simply indicates that Gabor atoms can be represented in discrete time as

$$g_{\{s,\omega,\tau\}}[n] = f_s[n - \tau] e^{j\omega(n-\tau)} \quad (20)$$

where

- $f_s[n]$ unit-norm function supported on a scale s ;
- ω atom's modulation frequency;
- τ discrete-time translation.

Note that Gabor atoms are scaled to have unit norm and that each is indexed in the dictionary by a parameter set $\{s, \omega, \tau\}$. This parametric structure allows for a simple description of a specific dictionary, which is useful for compression. When the atomic parameters are not tightly restricted, Gabor dictionaries are highly overcomplete and can include both Fourier and wavelet bases; examples of Gabor atoms are depicted in Fig. 3. One issue to note is that the modulation of an atom can be defined independently of the time shift, or *dereferenced*:

$$\bar{g}_{\{s,\omega,\tau\}}[n] = f_s[n - \tau] e^{j\omega n} = e^{j\omega\tau} g_{\{s,\omega,\tau\}}[n]. \quad (21)$$

This simple phase relationship will have a significant impact in later considerations.

In applications of Gabor functions, $g[n]$ is typically an even-symmetric window. The associated dictionaries thus consist

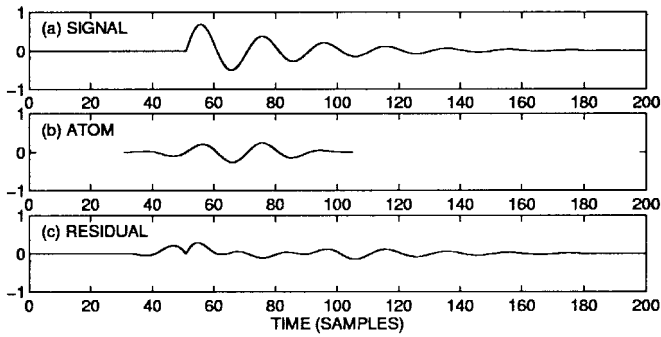


Fig. 4. Pre-echo artifact. (a) Damped sinusoidal signal. (b) First atom chosen from a symmetric Gabor dictionary by matching pursuit. (c) Residual; note the artifact at the onset time.

of atoms that exhibit symmetric time-domain behavior. This is problematic for modeling asymmetric features such as transients, which occur frequently in natural signals such as music. Fig. 4(a) shows a typical transient from linear system theory, the damped sinusoid; the first stage of a matching pursuit based on symmetric Gabor functions chooses the atom shown in Fig. 4(b). This atom matches the frequency behavior of the signal, but its time-domain symmetry results in a pre-echo as indicated. The atomic model has energy before the onset of the original signal; as a result, the residual has both a pre-echo and a discontinuity at the onset time as shown in Fig. 4(c). In later stages, then, the matching pursuit must incorporate small-scale atoms into the decomposition to remove the pre-echo and to model the discontinuity. One approach to this problem is the high-resolution matching pursuit algorithm proposed in [13] and [14], where symmetric atoms are still used but the selection metric is modified so that atoms that introduce drastic artifacts are not chosen for the decomposition. Another approach is to use a dictionary of asymmetric atoms, e.g., damped sinusoids.

B. Damped Sinusoids

The common occurrence of damped oscillations in natural signals justifies considering damped sinusoids as signal model components; in addition, damped sinusoids are physically better suited than symmetric Gabor atoms for representing transients. Like the atoms in a general Gabor dictionary, damped sinusoidal atoms can be indexed by characteristic parameters, namely, the damping factor a , modulation frequency ω , and start time τ

$$g_{\{a,\omega,\tau\}}[n] = S_a a^{(n-\tau)} e^{j\omega(n-\tau)} u[n - \tau] \quad (22)$$

or, if the modulation is dereferenced

$$\bar{g}_{\{a,\omega,\tau\}}[n] = S_a a^{(n-\tau)} e^{j\omega n} u[n - \tau] \quad (23)$$

where the factor S_a is included for unit-norm scaling. Examples are depicted in Fig. 5. It should be noted that these atoms can be interpreted as Gabor functions derived from a one-sided exponential window; their asymmetry distinguishes them from typical Gabor atoms, however. In addition, their atomic structure is more readily indicated by a damping factor than a scale parameter; therefore, the dictionary index set $\{a, \omega, \tau\}$ is used instead of the general Gabor set $\{s, \omega, \tau\}$.

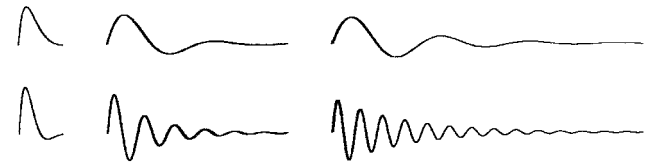


Fig. 5. Damped sinusoids: Gabor atoms derived from a one-sided exponential window.

A damped sinusoidal atom corresponds to the impulse response of a filter with a single complex pole; this is a suitable property given the intent of representing transient signals, especially if the signal's source can be well modeled by simple linear systems. For the sake of realizability, however, it is necessary to deviate somewhat from this relationship between the atoms and IIR filters. Specifically, a damped sinusoidal atom is truncated to a finite duration when its amplitude envelope falls below a threshold T ; the corresponding length is $L = \lceil \log T / \log a \rceil$, and the appropriate scaling factor is then $S_a = \sqrt{(1 - a^2) / (1 - a^{2L})}$. Note that this truncation results in sensible localization properties; heavily damped atoms are short-lived, and lightly damped atoms persist in time.

C. Composite Dictionaries

The simple example of Fig. 4 shows that symmetric atoms are inappropriate for modeling some signals. While that example is motivated by physical considerations, i.e. simple linear models of physical systems, it certainly does not encompass the wide range of complicated behaviors observed in natural signals. It is, of course, trivial to construct examples for which asymmetric atoms would prove similarly ineffective. Thus, given the task of modeling arbitrary signals, it can be argued that a wide range of both symmetric and asymmetric atoms should be present in the dictionary. Such *composite* dictionaries are considered here.

A simple way to build a composite dictionary is to merge a dictionary of symmetric atoms with a dictionary of damped sinusoids. The pursuit described in Section III can indeed be carried out using such a dictionary, but the atomic index set requires an additional parameter to specify which type of atom the set refers to. In addition, the nonuniformity of the dictionary would cause difficulties in the computation and storage of the dictionary cross-correlations needed for the correlation update of (11). Such issues will be discussed in Section VII.

It is shown in Section V that correlations with damped sinusoidal atoms can be computed with low cost without using the update formula of (11). The approach applies both to causal and anticausal damped sinusoids, which motivates considering two-sided atoms constructed by coupling causal and anticausal components. This construction can be used to generate symmetric and asymmetric atoms; furthermore, these atoms can be smoothed by simple convolution operations. Such atoms take the form

$$g_{\{a,b,J,\omega,\tau\}}[n] = f_{\{a,b,J\}}[n - \tau] e^{j\omega(n-\tau)} \quad (24)$$

or, if the modulation is dereferenced

$$\bar{g}_{\{a,b,J,\omega,\tau\}}[n] = f_{\{a,b,J\}}[n - \tau] e^{j\omega n} \quad (25)$$

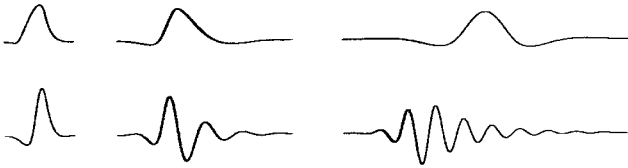


Fig. 6. Composite atoms. Symmetric and asymmetric atoms can be constructed by coupling causal and anticausal damped sinusoids and using low-order smoothing.

where the amplitude envelope is a unit-norm function constructed using a causal and an anticausal exponential according to the formula

$$f_{\{a,b,J\}}[n] = S_{\{a,b,J\}}(a^n u[n] + b^{-n} u[-n] - \delta[n]) * h_J[n] \quad (26)$$

where $\delta[n]$ is subtracted because the causal and anticausal components, as written, overlap at $n = 0$. The function $h_J[n]$ is a smoothing window of length J ; later considerations will be limited to the case of a rectangular window. A variety of composite atoms are depicted in Fig. 6.

The unit-norm scaling factor for a composite atom is

$$S_{\{a,b,J\}} = \frac{1}{\sqrt{\Upsilon(a,b,J)}} \quad (27)$$

where $\Upsilon(a,b,J)$ denotes the squared-norm of the atom prior to scaling

$$\Upsilon(a,b,J) = \sum_n |(a^n u[n] + b^{-n} u[-n] - \delta[n]) * h_J[n]|^2 \quad (28)$$

which can be simplified to

$$\Upsilon(a,b,J) = \sum_{l=0}^{J-1} \sum_{k=0}^{J-1} \frac{a^{|l-k|}}{1-a^2} + \frac{b^{|l-k|}}{1-b^2} + \frac{a^{|l-k|}b - ab^{|l-k|}}{a-b} \quad (29)$$

which does not take truncation of the atoms into account. This approximation does not introduce significant error if a small truncation threshold is used; however, if some error is introduced, the analysis-by-synthesis iterations in the pursuit work to remove the error at later stages. This scale factor affects the complexity of the algorithm, but primarily with respect to precomputation.

The composite atoms described above can be written in terms of unit-norm constituent atoms

$$\bar{g}_{\{a,b,J,\omega,\tau\}}[n] = S_{\{a,b,J\}} \left(\frac{\bar{g}_{\{a,\omega,\tau\}}^+[n]}{S_a} + \frac{\bar{g}_{\{b,\omega,\tau\}}^-[n]}{S_b} - \delta[n] \right) * h_J[n] \quad (30)$$

$$= S_{\{a,b,J\}} \sum_{\Delta=0}^{J-1} \frac{\bar{g}_{\{a,\omega,\tau+\Delta\}}^+[n]}{S_a} + \frac{\bar{g}_{\{b,\omega,\tau+\Delta\}}^-[n]}{S_b} - \delta[n + \Delta] \quad (31)$$

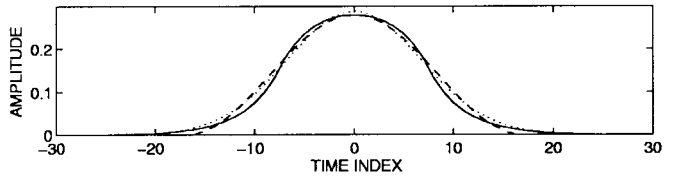


Fig. 7. Symmetric composite atoms. An example of a smoothed composite atom (solid) that roughly matches a Hanning window (dashed) and a Gaussian window (dotted).

where $\bar{g}_{\{a,\omega,\tau\}}^+[n]$ is a causal atom, and $\bar{g}_{\{b,\omega,\tau\}}^-[n]$ is an anticausal atom defined as

$$\bar{g}_{\{b,\omega,\tau\}}^-[n] = S_b b^{-(n-\tau)} e^{j\omega n} u[-(n-\tau)], \quad (32)$$

Note that atoms with dereferenced modulation are used in (31) so that the modulations add coherently in the sum over the time lags Δ ; otherwise, the constituent atoms would require phase shifts $e^{j\omega\Delta}$ to achieve coherent modulation of the composite atom. As will be seen, this construction leads to a simple relationship between the correlations of the signal with the composite atom and with the underlying damped sinusoids, especially in the dereferenced case.

The special case of symmetric atoms ($a = b$), one example of which is shown in Fig. 6, suggests using this approach to construct atoms similar to symmetric Gabor atoms based on common windows. Given a unit-norm window $w[n]$, the issue is to find a damping factor a and a smoothing order J such that the resultant $f_{\{a,a,J\}}[n]$ accurately mimics $w[n]$. Using the two-norm as a metric, this amounts to choosing a and J to minimize $\|f_{\{a,a,J\}}[n] - w[n]\|^2$, which is equivalent to maximizing the correlation of $f_{\{a,a,J\}}[n]$ and $w[n]$. Note that this process itself corresponds to a matching pursuit. In an implementation, this would not be an on-line operation but rather a precomputation indicating values of a and J for the parameter set of the composite dictionary; these values determine the scales of symmetric behavior in the dictionary. Fig. 7 shows a composite atom that roughly matches a Hanning window and a Gaussian window.

It has been shown that a composite dictionary containing a wide range of symmetric and asymmetric atoms can be constructed by coupling causal and anticausal damped sinusoids. Atoms resembling common symmetric Gabor atoms can readily be generated; therefore, standard symmetric atoms can be included as a dictionary subset; there is no generality lost by constructing atoms in this way. Furthermore, the construction is useful in that the pursuit computations can be carried out efficiently; the computational framework is developed in Sections V and VI.

D. Signal Modeling

In atomic modeling by matching pursuit, the characteristics of the signal estimate fundamentally depend on the structure of the time–frequency dictionary used in the pursuit. Consider the successively refined model in Fig. 8, which is derived by matching pursuit with a dictionary of symmetric Gabor atoms. In the early stages of the pursuit, the algorithm arrives at smooth estimates of the global signal behavior because the

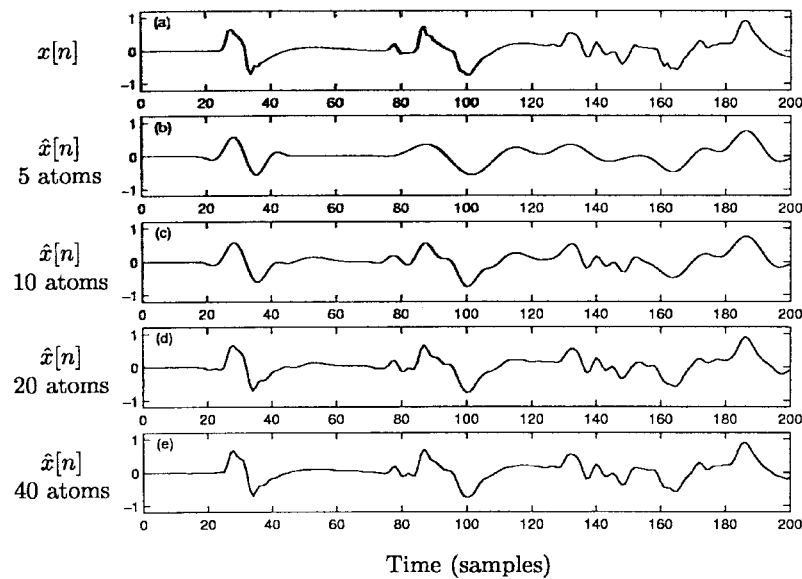


Fig. 8. Signal modeling with symmetric Gabor atoms. The original signal in (a), which is the onset of a gong strike, is modeled by matching pursuit with a dictionary of symmetric Gabor atoms derived from a Hanning prototype. Approximate reconstructions at various pursuit stages are given. (b) Five atoms. (c) Ten atoms. (d) 20 atoms. (e) 40 atoms.

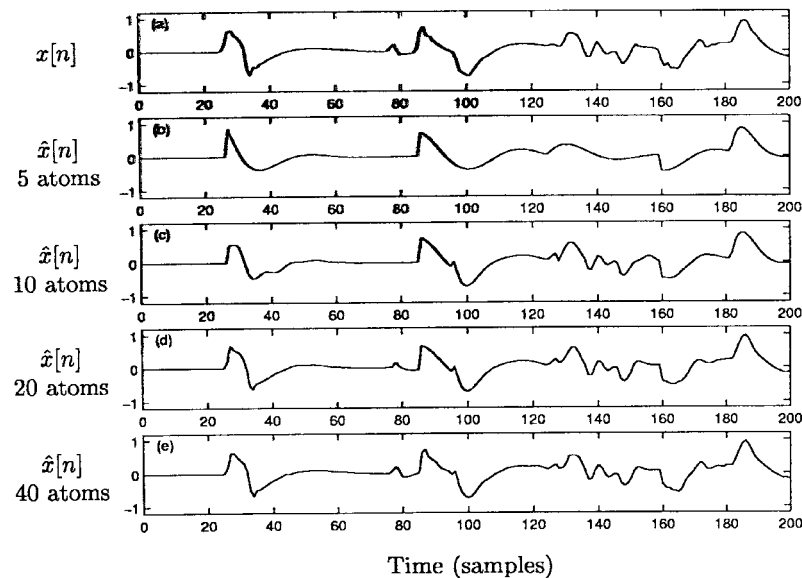


Fig. 9. Signal modeling with damped sinusoidal atoms. The signal in (a) is modeled by matching pursuit with a dictionary of damped sinusoids. Approximate reconstructions at various pursuit stages are given. (b) Five atoms. (c) Ten atoms. (d) Twenty atoms. (e) Forty atoms.

large-scale dictionary elements to choose from are themselves smooth functions. At later stages, the algorithm chooses atoms of smaller scale to refine the estimate; for instance, small-scale atoms are incorporated to remove pre-echo artifacts.

In the example of Fig. 9, the model is derived by matching pursuit with a dictionary of damped sinusoids. Here, the early estimates have sharp edges since the dictionary elements are one-sided functions. In later stages, edges that require smoothing are refined by inclusion of overlapping atoms in the model; in addition, as in the symmetric atom case, atoms of small scale are chosen in late stages to counteract any inaccuracies brought about by the early atoms.

In the examples of Figs. 8 and 9, the dictionaries are designed for a fair comparison in that they consist of atoms

with comparable scales and are structured so that the mean-squared errors of the respective atomic models have similar convergence properties. The convergence behaviors are compared in Fig. 10(a); the plot in Fig. 10(b) shows the energy of the pre-echo in the symmetric Gabor model and indicates that the pursuit devotes atoms at later stages to remove the pre-echo artifact. The model based on damped sinusoids does not introduce a pre-echo.

Modeling with a composite dictionary is depicted in Fig. 11. The dictionary contains the same causal damped sinusoids as in the example of Fig. 9 plus an equal number of anticausal damped sinusoids and a few smoothing orders. As will be seen, computing the correlations with the underlying damped sinusoids is the main factor in the cost of the composite

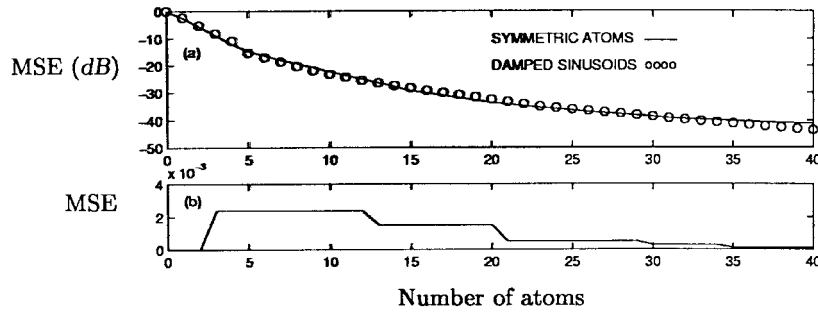


Fig. 10. Mean-squared convergence of atomic models. Plot (a) shows the mean-squared error of the atomic models depicted in Figs. 8 and 9. The dictionaries of symmetric Gabor atoms (solid) and damped sinusoids (circles) are designed to have similar mean-squared convergence for the signal in question. Plot (b) shows the mean-squared energy in the pre-echo of the symmetric Gabor model; the pursuit devotes atoms at later stages to reduce the pre-echo energy. The damped sinusoidal decomposition does not introduce a pre-echo.

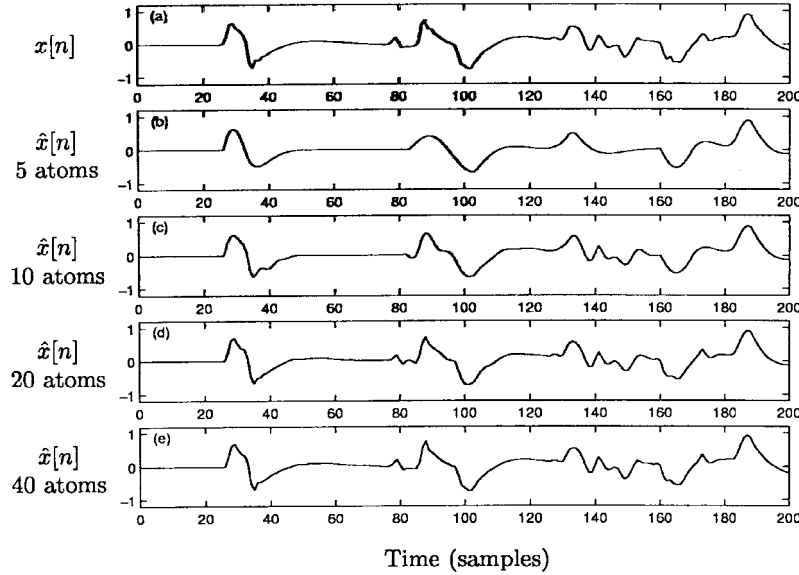


Fig. 11. Signal modeling with composite atoms. The signal in (a) is modeled by matching pursuit with composite atoms. Approximate models at various pursuit stages are given. (b) Five atoms. (c) Ten atoms. (d) 20 atoms. (e) 40 atoms. The composite dictionary contains the same causal damped sinusoids used in the example of Fig. 9, plus an equal number of anticausal damped sinusoids and a few smoothing orders.

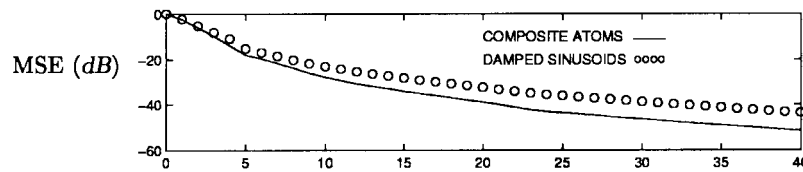


Fig. 12. Mean-squared error of an atomic model using composite atoms (solid) and the mean-squared error of a model based on only the underlying causal damped sinusoids (circles). This plot corresponds to the composite atomic models given in Fig. 11 and the damped sinusoidal decompositions of Fig. 9.

pursuit; therefore, deriving the composite atomic model in Fig. 11 requires roughly twice as much computation as the pursuit based on damped sinusoids alone. As shown in Fig. 12, this additional computation leads to a lower mean-squared error for the model. Noting that the parameter set for composite atoms is larger than that for damped sinusoids or Gabor atoms, it is clear that a full comparison of the various models requires consideration of the interplay of computation, rate, and distortion [6].

V. PURSUIT OF DAMPED SINUSOIDAL ATOMS

In matching pursuit using a dictionary of complex damped sinusoids, correlations must be computed for every combination of damping factor, modulation frequency, and time shift.

The correlation of a signal $x[n]$ with a truncated causal atom $g_{\{a,\omega,\tau\}}^+[n]$ is given by

$$\eta_+(a, \omega, \tau) = S_a \sum_{n=\tau}^{\tau+L-1} x[n] a^{(n-\tau)} e^{-j\omega(n-\tau)}. \quad (33)$$

In the following, correlations with unnormalized atoms will be used:

$$\rho_+(a, \omega, \tau) = \sum_{n=\tau}^{\tau+L-1} x[n] a^{(n-\tau)} e^{-j\omega(n-\tau)} = \frac{\eta_+(a, \omega, \tau)}{S_a}. \quad (34)$$

Using unnormalized atoms will serve to simplify the notation and to reduce the cost of the pursuit algorithm developed in Section VI.

For dictionaries of complex damped sinusoids, the correlation computation can be simplified, irrespective of the update formula in (11); it is shown in the following that the correlation in (34) can be readily computed with recursive filter banks and FFT's. These simplifications allow for a substantial reduction of the computation requirements with respect to the time shift and modulation parameters. Note that the correlation uses the atoms defined in (22), for which the modulation is phase-referenced to τ ; results for dereferenced atoms are given later.

A. Time-Domain Simplification

The exponential structure of the atoms can be used to simplify the pursuit computation over the time index; correlations at neighboring times are related by the recursion

$$\begin{aligned} \rho_+(a, \omega, \tau - 1) &= ae^{-j\omega} \rho_+(a, \omega, \tau) + x[\tau - 1] \\ &\quad - a^L e^{-j\omega L} x[\tau + L - 1]. \end{aligned} \quad (35)$$

This is just a one-pole filter with a correction to account for truncation. If truncation effects are ignored, which is reasonable for small truncation thresholds, the formula becomes

$$\rho_+(a, \omega, \tau - 1) = ae^{-j\omega} \rho_+(a, \omega, \tau) + x[\tau - 1]. \quad (36)$$

Note that the equation is operated in reversed time to make the recursion stable for causal damped sinusoids; the similar forward recursion is unstable for $a < 1$. For anticausal atoms, the correlations are given by the recursion

$$\begin{aligned} \rho_-(b, \omega, \tau + 1) &= be^{j\omega} \rho_-(b, \omega, \tau) + x[\tau + 1] \\ &\quad - b^L e^{j\omega L} x[\tau - L + 1] \end{aligned} \quad (37)$$

or, if truncation is neglected

$$\rho_-(b, \omega, \tau + 1) = be^{j\omega} \rho_-(b, \omega, \tau) + x[\tau + 1]. \quad (38)$$

These recursions are operated in forward time for the sake of stability.

The equivalence of (36) and (38) to filtering operations suggests interpreting the correlation computation over all possible parameters $\{a_i, \omega_i, \tau_i\}$ as an application of the signal to a dense grid of one-pole filters in the z -plane; these are the *matched filters* for the dictionary atoms. The filter outputs are the correlations needed for the matching pursuit; the maximally correlated atom is indicated by the maximum magnitude output of the filter bank. Of course, pursuit based on arbitrary atoms can be interpreted in terms of matched filters, but in the general case, this insight is not particularly useful; here, it provides a framework for reducing the computation. Note that the dictionary atoms themselves correspond to the impulse responses of a grid of one-pole filters; as in the wavelet filter bank case, then, the atomic synthesis can be interpreted as an application of the expansion coefficients to a synthesis filter bank. Fig. 13 depicts z -plane dictionary structures that provide for various tradeoffs in time-frequency resolution.

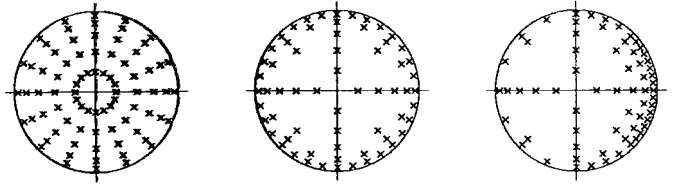


Fig. 13. Filter bank interpretation and dictionary structures. The atoms in a dictionary of damped sinusoids correspond to the impulse responses of a bank of one-pole filters; for decaying causal atoms, the poles are inside the unit circle. These dictionaries can be structured for various tradeoffs in time-frequency resolution. The correlations in the pursuit are computed by the corresponding matched filters, which are time-reversed and, thus, have poles outside the unit circle.

B. Frequency-Domain Simplification

A simplification of the correlation computations across the frequency parameter can be achieved if the z -plane filter bank, or, equivalently, the matching pursuit dictionary, is structured such that the modulation frequencies are equi-spaced for each damping factor. If the filters (atoms) are equi-spaced angularly on circles in the z -plane, the discrete Fourier transform can be used for the computation over ω . For $\omega = 2\pi k/K$, the correlation is given by

$$\begin{aligned} \rho_+(a, 2\pi k/K, \tau) &= \sum_{n=0}^{L-1} x[n + \tau] a^n e^{-j2\pi kn/K} \\ &= \text{DFT}_K \{x[n + \tau] a^n\}_k \end{aligned} \quad (39)$$

where $n \in [0, L - 1]$, and $K \geq L$. Thus, an FFT can be used to compute correlations over the frequency index. This formulation applies to any dictionary of harmonically modulated atoms.

At a fixed scale, correlations must be computed at every time-frequency pair in the index set. There are two ways to cover this time-frequency index plane; these correspond to the dual interpretations of the short-time Fourier transform [6]. The first approach is to use a running DFT with an exponential window; windowing and the DFT require L and $K \log K$ multiplies per time point, respectively; therefore, this method requires roughly $N(L + K \log K)$ multiplies for a signal of length N . The second approach is to use a DFT to initialize the K matched filters across frequency and then compute the outputs of the filters to evaluate the correlations across time; indeed, the signal can be zero padded such that the filters are initialized with zero values, and no DFT is required. Recalling the recursion of (35), this latter method requires one complex multiply and one real-complex multiply per filter for each time point; therefore, it requires $5KN$ real multiplies, $2KN$ of which account for truncation effects and are not imperative. For large values of K , this is significantly less than the multiply count for the running DFT approach; therefore, the matched filter approach is the method of choice.

C. Results for Dereferenced Modulation

The results given in the previous sections hold for an atom whose modulation is referenced to the time origin of the atom, as in (19), (22), and (24). This local time reference has been adhered to since it allows for an immediate filter bank

interpretation of the matching pursuit analysis; in addition, synthesis based on such atoms can be directly carried out using recursive filters. For the construction and pursuit of composite atoms, however, the dereferenced atoms defined in (21), (23), and (25) are of importance. The correlation formulae for these atoms can be derived by combining the relation in (21) with the expression in (34) to arrive at

$$\bar{\rho}_+(a, \omega, \tau) = e^{-j\omega\tau} \rho_+(a, \omega, \tau). \quad (40)$$

Therefore, (36) and (38) can be reformulated as

$$\bar{\rho}_+(a, \omega, \tau - 1) = a\bar{\rho}_+(a, \omega, \tau) + e^{-j\omega(\tau-1)} x[\tau - 1] \quad (41)$$

$$\bar{\rho}_-(b, \omega, \tau + 1) = b\bar{\rho}_-(b, \omega, \tau) + e^{-j\omega(\tau+1)} x[\tau + 1]. \quad (42)$$

When the modulation depends on the atomic time origin, the pursuit can be interpreted in terms of a modulated filter bank; for dereferenced modulation, however, the filter bank has a heterodyne structure. As will be seen in Section VI, this dereferencing simplifies the relationship between the signal correlations with composite atoms and the correlations with underlying damped sinusoids.

D. Real Decompositions of Real Signals

If dictionaries of complex atoms are used in matching pursuit, the correlations and, hence, the expansion coefficients for signal decompositions will generally be complex; a given coefficient thus provides both a magnitude and a phase for the atom in the expansion. For real signals, decomposition in terms of complex atoms can be misleading. For instance, for a signal that consists of one real damped sinusoid, the pursuit does not simply find the constituent conjugate pair of atoms as might be expected; this occurs because an atom and its conjugate are not orthogonal. For real signals, then, it is preferable to consider expansions in terms of real atoms as in

$$g_{\{a, \omega, \tau, \phi\}}^+ = S_{\{a, \omega, \phi\}} a^{(n-\tau)} \cos[\omega(n-\tau) + \phi] u[n-\tau] \quad (43)$$

or, in the case of dereferenced modulation

$$\bar{g}_{\{a, \omega, \tau, \phi\}}^+ = \bar{S}_{\{a, \omega, \phi\}} a^{(n-\tau)} \cos[\omega n + \phi] u[n-\tau] \quad (44)$$

where the two cases differ by a phase offset, which affects the scaling as well as the modulation.

In the case of a complex dictionary, the atoms are indexed by the three parameters $\{a, \omega, \tau\}$, and the phase of an atom in the expansion is given by its correlation. In contrast, a real dictionary requires the phase parameter as an additional index because of the explicit presence of the phase in the argument of the cosine in the atom definition. The phase is not supplied by the correlation computation, as in the complex case; like the other parameters, it must be discretized and incorporated as a dictionary parameter in the pursuit, which results in a larger dictionary and, thus, a more complicated search. Furthermore, the correlation computations are more difficult than in the complex case because the recursion formulae derived earlier do not apply for these real atoms. These problems can be circumvented by using a complex dictionary and considering conjugate subspaces according to the formulation of Section III.

Conjugate subspace pursuit can be used to search for conjugate pairs of complex damped sinusoids; the derivation leading to (17) verifies that this approach will arrive at a decomposition in terms of real damped sinusoids if the original signal is real. The advantage of this method is indicated by (15) and (16), which show that the expansion coefficients and the maximization metric in the conjugate pursuit are both functions of the correlation of the residual with the underlying complex atoms. The computational simplifications for a dictionary of complex damped sinusoids can thus be readily applied to calculation of a real expansion of the form

$$x[n] \approx 2 \sum_{i=1}^I S_{a_i} A_i a_i^{(n-\tau_i)} \cos[\omega_i n + \phi_i] \quad (45)$$

where $A_i e^{j\phi_i} = \alpha_i(1)$, and the modulation is dereferenced. As in the complex case, the phases of the atoms in this real decomposition are provided directly by the computation of the expansion coefficients; the phase is not required as a dictionary index, i.e., an explicit search over a phase index is not required in the pursuit. By considering signal expansions in terms of conjugate pairs, the advantages of the complex dictionary are fully maintained; furthermore, note that the dictionary for the conjugate search is effectively half the size of the full complex dictionary since atoms are considered in conjugate pairs.

It is important to note that (45) neglects the inclusion of unmodulated exponentials in the signal expansion. Such atoms are indeed present in the complex dictionary, and all of the recursion speedups apply trivially; furthermore, the correlation of an unmodulated atom with a real signal is always real; therefore, there are no phase issues to be concerned with. An important caveat, however, is that the conjugate pursuit algorithm breaks down if the atom is purely real; the pursuit requires that the atom and its conjugate be linearly independent, meaning that the atom must have nonzero real and imaginary parts. Thus, a fix is required if real unmodulated exponentials are to be admitted into the signal model. The i th stage of the fixed algorithm is as follows: First, the correlations $\langle g, r_i \rangle$ for the entire dictionary of complex atoms are computed using the simplifications described. Then, energy minimization metrics for both types of atoms are computed and stored; for real atoms, the metric is simply $|\langle g, r_i \rangle|^2$, and for complex atoms, the metric is $\langle g, r_i \rangle^* \alpha(1) + \langle g, r_i \rangle \alpha(1)^*$ as given in (16), where $\alpha(1)$ is as defined in (15), and

$$\langle g_{\{a, \omega, \tau\}}, g_{\{a, \omega, \tau\}}^* \rangle = S_a^2 \left(\frac{1 - a^{2L} e^{-j2\omega L}}{1 - a^2 e^{-j2\omega}} \right). \quad (46)$$

These metrics quantify the amount of energy removed from the residual in the two cases; maximization over these metrics indicates which real component should be added to the signal expansion at the i th stage to minimize the energy of the new residual $r_{i+1}[n]$.

The description of a signal in terms of conjugate pairs does not require more data than a model using complex atoms. Either case simply requires the indices $\{a, \omega, \tau\}$ and the complex number $\alpha(1)$ for each atom in the model. There is, however, additional computation in the conjugate-pair analysis and synthesis, but this computation improves the model's

ability to represent real signals. In a sense, the improvement arises because the added computation enables the model data to encompass twice as many atoms in the conjugate pair case as in the complex case.

VI. PURSUIT OF COMPOSITE ATOMS

Using matching pursuit to derive a signal model based on composite atoms requires computation of the correlations of the signal with these atoms. Recalling the form of the composite atoms given in (30) and (31), these correlations have, by construction, a simple relationship to the correlations with the underlying one-sided atoms as in

$$\bar{\rho}(a, b, J, \omega, \tau) = S_{\{a,b,J\}} \sum_{\Delta=0}^{J-1} \left[\frac{\bar{\eta}_+(a, \omega, \tau + \Delta)}{S_a} + \frac{\bar{\eta}_-(b, \omega, \tau + \Delta)}{S_b} - x[\tau + \Delta] \right] \quad (47)$$

$$= S_{\{a,b,J\}} \sum_{\Delta=0}^{J-1} [\bar{\rho}_+(a, \omega, \tau + \Delta) + \bar{\rho}_-(b, \omega, \tau + \Delta) - x[\tau + \Delta]]. \quad (48)$$

The correlation with any composite atom can thus be computed based on the correlations derived by the recursive filter banks discussed earlier; this computation is most straightforward if dereferenced modulation is used in the constituent atoms and if these underlying atoms are unnormalized. Essentially, any atom constructed according to (31), which includes simple damped sinusoids, can be added to the modeling dictionary at the cost of one multiply per atom for scaling. Note that for composite atoms, real decompositions of real signals take the form

$$x[n] \approx 2 \sum_{i=1}^I A_i f_{\{a_i, b_i, J_i\}}[n - \tau_i] \cos[\omega_i n + \phi_i] \quad (49)$$

where $f_{\{a_i, b_i, J_i\}}$ is as defined in (26), and $A_i e^{j\phi_i} = \alpha_i(1)$ from (15).

VII. COMPUTATION CONSIDERATIONS

This section compares the computational cost of two matching pursuit implementations: pursuit based on correlation updates [3] and pursuit based on recursive filter banks. The cost is measured in terms of memory requirements and multiplicative operations. Simple search operations, table lookups, and conditionals are neglected in the cost measure. Precomputation is allowed without a penalty, but storage of precomputed data is included in the memory cost. Startup cost for the first pursuit iteration is considered separately; in cases where only a few atoms are to be derived, the startup arithmetic in the update algorithm may constitute an appreciable percentage of the overall computation.

The following treatment involves modeling a real signal of length N using a composite dictionary based on damped sinusoids. The dictionary parameters consist of

- A different causal damping factors;
- B anticausal damping factors;
- H smoothing orders;

- K modulations;
- N time shifts.

The dictionary thus has $M = ABHKN$ atoms; using S to denote the number of scales, namely, $S = ABH$, the dictionary size is given by $M = SKN$. The average scale or atom length will be denoted by L ; the correlation $\langle g, x \rangle$ thus requires L real-complex multiplies on average. The following comparison focuses on pursuit of complex atoms since the evaluation of a real model based on a complex pursuit has equal cost in both matching pursuit implementations; in addition, deriving the correlation magnitudes requires the same amount of computation in both approaches. The relevant computation to compare is that required to calculate $\langle g, r_i \rangle$ for all of the complex atoms $g \in D$ at some stage i of the algorithm.

The update approach computes the correlations needed for the pursuit using (11), which relates the correlations at stage $i + 1$ to those computed at stage i . This method relies on precomputation and storage of the dictionary cross-correlations $\langle g, g_i \rangle$ to reduce the cost of the pursuit. If this storage is done without taking the sparsity or redundancy of the data into account, M^2 cross-correlations must be stored. A simple example shows that such a brute force approach is prohibitive. Consider analysis of a 10-ms frame of high-quality audio consisting of $N = 400$ samples. In a rather small dictionary with $K = 32$, $A = 10$, $B = 1$, and $H = 1$, there are roughly $M = 10^5$ atoms. Storage of the complex-valued cross-correlations then requires $2M^2 = 2 \times 10^{10}$ memory locations. This is altogether unreasonable; therefore, it is necessary to investigate the possibility of memory-computation tradeoffs.

The memory requirement can be reduced by considering the sparsity and redundancy of the cross-correlation data. First, many of the atom pairs have no time overlap and, thus, zero correlation; these cases can be handled with conditionals. For atoms that do overlap, the correlation storage can be reduced using the following formulation. Introducing the notation

$$g(s_0, \omega_0, \tau_0) = g_{\{a_0, b_0, J_0, \omega_0, \tau_0\}}[n] = f_{\{s_0\}}[n - \tau_0] e^{j\omega_0 n} \quad (50)$$

$$g(s_1, \omega_1, \tau_1) = g_{\{a_1, b_1, J_1, \omega_1, \tau_1\}}[n] = f_{\{s_1\}}[n - \tau_1] e^{j\omega_1 n} \quad (51)$$

where s_0 and s_1 serve as shorthand for the effective scales of the atoms and $f[n]$ is a unit-norm envelope constructed as in (26), the cross-correlation of two composite atoms is given by

$$\langle g(s_0, \omega_0, \tau_0), g(s_1, \omega_1, \tau_1) \rangle = \sum_n f_{\{s_0\}}[n - \tau_0] f_{\{s_1\}}[n - \tau_1] e^{j(\omega_1 - \omega_0)n} \quad (52)$$

$$\{\text{letting } m = n - \tau_0\} = \sum_m f_{\{s_0\}}[m] f_{\{s_1\}}[m - (\tau_1 - \tau_0)] e^{j(\omega_1 - \omega_0)(m + \tau_0)} \quad (53)$$

$$= e^{j(\omega_1 - \omega_0)\tau_0} \langle g(s_0, 0, 0), g(s_1, \omega_1 - \omega_0, \tau_1 - \tau_0) \rangle. \quad (54)$$

Thus, with the exception of a phase shift, the cross-correlation depends only on the relative time locations and modulation

frequencies of the atoms. Furthermore, it only depends on the absolute frequency difference since negative values of $\omega_1 - \omega_0$ can be accounted for by conjugation as

$$\begin{aligned} & \langle g(s_0, 0, 0), g(s_1, \omega_1 - \omega_0, \tau_1 - \tau_0) \rangle \\ &= \langle g(s_0, 0, 0), g(s_1, \omega_0 - \omega_1, \tau_1 - \tau_0) \rangle^*. \end{aligned} \quad (55)$$

Conjugation can also be used to handle redundancy in the cross-correlations for scale pairs

$$\begin{aligned} & \langle g(s_1, \omega_1, \tau_1), g(s_0, \omega_0, \tau_0) \rangle \\ &= \langle g(s_0, \omega_0, \tau_0), g(s_1, \omega_1, \tau_1) \rangle^* \end{aligned} \quad (56)$$

$$= e^{j(\omega_0 - \omega_1)\tau_0} \langle g(s_0, 0, 0), g(s_1, \omega_1 - \omega_0, \tau_1 - \tau_0) \rangle^*. \quad (57)$$

This scale property serves to reduce the memory requirements by roughly a factor of two.

The formulations given above drastically reduce the amount of memory required to store the dictionary cross-correlations. For the modulation frequencies, there are K distinct possibilities for $|\omega_1 - \omega_0|$. For the time shifts, the S different scales can be considered in pairs using L to approximate the number of lags that lead to overlap and nonzero correlation; there are roughly S^2L different configurations. In total, then, $2S^2KL$ memory locations are required to store the distinct cross-correlation values; the scale-pair redundancy reduces this count to S^2KL . For the simple example discussed above, this amounts to about 6×10^4 locations for $L = 20$. Noting the phase shift in (54), this reduction in the memory requirements is achieved at the cost of a complex multiply, or three real multiplies, for each correlation update.

In the filter bank approach, the pursuit computation is based on correlations with unnormalized atoms as formalized in (48), which holds for any dictionary of composite atoms or simple damped sinusoids (where $H = 1$ and $B = 1$). This correlation computation requires scaling by $S_{\{a,b,J\}}$; therefore, these scaling factors are precomputed and stored. When the causal and anticausal damping factors do not have any particular symmetry, storing the scaling factors requires $S = ABH$ memory locations.

In the first stage of the update algorithm, all of the signal correlations with the dictionary atoms must be computed, which requires $ML = SKL$ real-complex multiplies or $2ML$ real multiplies; storing the results requires $2M$ locations so that the total memory needed in the update algorithm is $S^2KL + 2M$. Note that the computation could be carried out with recursive filter banks at a lower cost, but such a merged approach will not be treated here.

Once the dictionary cross-correlations have been precomputed and the correlations for the first stage of the pursuit have been calculated and stored, the cost of the update algorithm depends only on the update formula. Each stage of the algorithm involves M complex-complex multiplies ($3M$ real) to multiply the M cross-correlations by α_i , plus another M complex-complex multiplies to carry out the phase shift given in (54), for a total of $6M$ real multiplies per iteration. Note that in the update algorithm, it is not necessary to keep the signal in memory after the first iteration or to ever actually compute the residual signal.

Method	MEMORY (real numbers)		COMPUTATION (real multiplies)	
	Precomp.	Algorithm	First iteration	Later iterations
Update	S^2KL $= A^2B^2H^2KL$	$2M$ $= 2ABHKN$	$2ML$ $= 2ABHKNL$	$6M$ $= 6ABHKN$
Filter bank	$S = ABH$	$N + 2(A+B)KN$	$5L + 2ABH + 6(A+B)KN$	

Fig. 14. Tabulation of computation considerations. Memory and computation requirements for matching pursuit using the update algorithm and the recursive filter bank method. N is the length of the signal; the dictionary index set contains A causal damping factors, B anticausal damping factors, H smoothing orders, $S = ABH$ scales, K modulations, and N time shifts, meaning that the dictionary contains $M = SKN = ABHKN$ distinct atoms. L is the average time support of a dictionary atom.

In matching pursuit based on recursive filter banks, the scaling factors $S_{\{a,b,J\}}$ are precomputed and available via lookup. In addition to the scaling factors, the residual signal must be stored, which requires N memory locations. The final memory requirement is the storage of the correlations with the constituent unnormalized damped sinusoids, which are needed to compute the correlations with the composite atoms. For some smoothing order J , correlations with J causal and J anticausal damped sinusoids are required. Storing these underlying correlations in a local manner requires $2(A+B)KJ$ locations, where the factor of two arises because the correlations are complex; global storage of these correlations requires $2(A+B)KN$ locations; therefore, the worst-case memory requirement in the filter bank case is $S + N + 2(A+B)KN$.

The filter bank algorithm uses $(A+B)K$ recursive filters to derive the correlations. In the dereferenced case of (41) and (42), each recursion requires four real-real multiplies for each of the N time points if atom truncation is neglected or six if truncation is included. As given by (48), correlations with composite atoms are computed by adding the correlations with constituent unnormalized damped sinusoids and then scaling with the appropriate factor; this process introduces $S = ABH$ real-complex multiplies or $2S$ real multiplies. Thus, $6(A+B)KN + 2ABH$ real multiplies are needed to compute the pursuit correlations. Once an atom is chosen based on these correlations, the residual must be updated; this requires roughly $5L$ multiplies to generate the unit-norm atomic envelope, modulation to the proper frequency, and weighting with its expansion coefficient prior to subtraction from the signal. The total computational cost per iteration for the filter bank algorithm is thus $5L + 6(A+B)KN + 2S$.

The results of this section are summarized in Fig. 14. To quantify the comparison, consider modeling a signal of length $N = 400$ with a dictionary having $A = 20$, $B = 10$, $H = 10$, $L = 30$, and $K = 32$. The update method requires storage of 3.8×10^9 precomputed values and 5.1×10^7 values for a given iteration, whereas the filter bank method requires 2×10^3 precomputation locations and 7.7×10^5 locations for a given iteration; the filter bank approach requires less memory. The update method carries out 1.5×10^9 multiplies for the first iteration and 1.5×10^8 multiplies for each iteration thereafter. The filter bank framework requires 2.3×10^6 multiplies for each iteration; therefore, it provides a considerable reduction in the cost of the pursuit computation.

VIII. CONCLUSION

This paper has focused on signal modeling using matching pursuit with a dictionary of damped sinusoids. It was shown that the required computation can be carried out efficiently using first-order recursive filter banks. Composite atoms constructed from damped sinusoids were also considered; these proved useful since they can represent a wide range of time-frequency behaviors. It was also demonstrated that the filter bank pursuit method is more efficient with respect to both memory and computation than pursuit based on the update framework of [3].

Estimation of the parameters of damped sinusoids in a signal has been widely considered for the applications of system identification and spectral estimation. These applications, however, usually involve an underlying source-filter model of the signal and, thus, differ from the task of generalized signal modeling. Of course, the estimation of damped sinusoidal parameters by matching pursuit could be tailored for spectral estimation and system identification, but this would require constraining the pursuit in various ways. The relationship between overcomplete expansions based on parametric dictionaries and parameter estimation methods such as ESPRIT has yet to be fully considered. Early work suggests that greedy algorithms can be effective for spectral estimation of nonstationary signals [24]; greedy parameter estimation seems to be well-suited for scenarios where little or no *a priori* information about the signal is available.

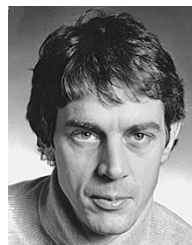
REFERENCES

- [1] D. Gabor, "Theory of communication," *J. Inst. Elect. Eng.*, vol. 93, pp. 429–457, 1946.
- [2] ———, "Acoustical quanta and the theory of hearing," *Nature*, vol. 159, no. 4044, pp. 591–594, May 1947.
- [3] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, Dec. 1993.
- [4] S. Qian and D. Chen, "Signal representation using adaptive normalized Gaussian functions," *Signal Process.*, vol. 36, pp. 1–11, Mar. 1994.
- [5] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," Feb. 1996. [Online.] Available: ftp.playfair.stanford.edu.
- [6] M. Goodwin, *Adaptive Signal Models: Theory, Algorithms, and Audio Applications*. Boston, MA: Kluwer, 1998.
- [7] V. Goyal, M. Vetterli, and N. Thao, "Quantized overcomplete expansions in \mathcal{R}^n : Analysis, synthesis and algorithms," *IEEE Trans. Inform. Theory*, vol. 44, pp. 16–31, Jan. 1998.
- [8] R. Coifman and M. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory*, vol. 38, pp. 713–718, Mar. 1992.
- [9] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. Image Processing*, vol. 2, pp. 160–175, Apr. 1993.
- [10] C. Herley, J. Kovacevic, K. Ramchandran, and M. Vetterli, "Tilings of the time-frequency plane: Construction of arbitrary orthogonal bases and fast tiling algorithms," *IEEE Trans. Signal Processing*, vol. 41, pp. 3341–3359, Dec. 1993.
- [11] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
- [12] B. Rao, "Analysis and extensions of the FOCUSS algorithm," in *Conf. Rec. Twenty-Ninth Asilomar Conf. Signals, Syst., Comput.*, Nov. 1996, vol. 2, pp. 1218–1223.
- [13] S. Jaggi *et al.*, "High resolution pursuit for feature extraction," Tech. Rep. LIDS-P-2371, Mass. Inst. Technol., Cambridge, Nov. 1996.
- [14] R. Gribonval *et al.*, "Analysis of sound signals with high resolution matching pursuit," in *Proc. IEEE-SP Int. Symp. Time-Freq. Time-Scale Anal.*, June 1996, pp. 125–128.
- [15] B. Friedlander and A. Zeira, "Oversampled Gabor representation for transient signals," *IEEE Trans. Signal Processing*, vol. 43, pp. 2088–2094, Sept. 1995.
- [16] C. Herley and M. Vetterli, "Wavelets and recursive filter banks," *IEEE Trans. Signal Processing*, vol. 41, pp. 2536–2556, Aug. 1993.
- [17] S. Tomazic, "On short-time Fourier transform with single-sided exponential window," *Signal Process.*, vol. 31, no. 2, pp. 141–148, Dec. 1996.
- [18] M. Unser, "Recursion in short-time signal analysis," *Signal Process.*, vol. 18, no. 3, pp. 229–240, May 1983.
- [19] M. Goodwin, "Matching pursuit with damped sinusoids," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1997, vol. 3, pp. 2037–2040.
- [20] M. Goodwin and M. Vetterli, "Atomic decompositions of audio signals," in *Proc. IEEE ASSP Workshop Appl. Signal Process. Audio Acoust.*, Oct. 1997.
- [21] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, Apr. 1995.
- [22] G. Davis, "Adaptive nonlinear approximations," Ph.D. dissertation, New York Univ., New York, NY, Sept. 1994.
- [23] P. Huber, "Projection pursuit," *Ann. Stat.*, vol. 13, no. 2, pp. 435–475, 1985.
- [24] P. Prandoni and M. Vetterli, "An FIR cascade structure for adaptive linear prediction," *IEEE Trans. Signal Processing*, vol. 46, pp. 2566–2571, Sept. 1998.



Michael M. Goodwin (M'97) was born in Florida in 1970. He received the B.S. and M.S. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1992 and the Ph.D. degree in electrical engineering from the University of California, Berkeley in 1997.

From 1989 to 1992, he was an intern at AT&T Bell Laboratories, Murray Hill, NJ, where he started out as a Circuit Designer and ended up working on the theory and implementation of electroacoustic beamformers. While a graduate student at Berkeley, he did research on signal processing for computer music at the Center for New Music and Audio Technologies and furthered some of this work as a consultant for the Gibson Guitar Corporation. He spent the summer of 1996 developing audio analysis and synthesis algorithms at E-Mu Systems, Scotts Valley, CA. Since 1997, he has worked in a digital media group at Silicon Graphics, Inc., Mountain View, CA. His research interests include signal modeling, wavelets, multimedia signal processing, and computer music, and he has consulted on topics ranging from audio signal processing to medical imaging. He is the author of several research articles and the book *Adaptive Signal Models: Theory, Algorithms, and Audio Applications* (Boston, MA: Kluwer, 1998).



Martin Vetterli (F'95) received the B.E. degree from the Swiss Federal Institute of Technology, Zurich, in 1981, the M.S. degree from Stanford University, Stanford, CA, in 1982, and the Ph.D. from the Swiss Federal Institute of Technology, Lausanne, in 1986.

In 1986, he joined the Center for Telecommunications Research and the Department of Electrical Engineering, Columbia University, New York, NY, where he was an Associate Professor of electrical engineering and co-director of the Image and Advanced Television Laboratory. In 1993, he joined the Faculty of Electrical Engineering and Computer Science, University of California, Berkeley, where he was a Full Professor until 1997 and is now an Adjunct Professor. In 1995, he joined the Communication Systems Division of the Swiss Federal Institute of Technology, Lausanne, which he headed from 1996 to 1997. He works on signal processing and communications, in particular, wavelet theory and applications, image and video compression, digital video communications, and computational complexity.

Dr. Vetterli's work won him several prizes (best paper awards from EURASIP in 1984 and from the IEEE Signal Processing Society in 1991 and 1996). He received the Swiss National Latsis Prize in 1996 and is the co-author, with J. Kovacevic, of the book *Wavelets and Subband Coding* (Englewood Cliffs, NJ: Prentice-Hall, 1994) and of numerous research papers. He is a Member of the Technical Board of C-Cube Microsystems.