MATCHING PURSUIT WITH DAMPED SINUSOIDS

Michael Goodwin
Department of Electrical Engineering and Computer Science &
Center for New Music and Audio Technologies (CNMAT)
University of California at Berkeley
e-mail: michaelg@eecs.berkeley.edu

ABSTRACT

The matching pursuit algorithm derives an expansion of a signal in terms of the elements of a large dictionary of time-frequency atoms. This paper considers the use of matching pursuit for computing signal expansions in terms of damped sinusoids. First, expansion based on complex damped sinusoids is explored; it is shown that the expansion can be efficiently derived using the FFT and simple recursive filterbanks. Then, the approach is extended to provide decompositions in terms of real damped sinusoids. This extension relies on generalizing the matching pursuit algorithm to derive expansions with respect to dictionary subspaces; of specific interest is the subspace spanned by a complex atom and its conjugate. Developing this particular case leads to a framework for deriving real-valued expansions of real signals using complex atoms. Applications of the damped sinusoidal decomposition include system identification, spectral estimation, and signal modeling for coding and analysis–modification–synthesis.

1. 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 $g_m[n]$; in matrix notation,

$$x = D\,\alpha\,, \qquad D = [g_1\ g_2 \cdots g_m \cdots g_M] \tag{1}$$

where the signal x is a column vector $(N \times 1)$, α is a column vector of weights $(M \times 1)$, and D is an $N \times M$ matrix whose columns are the expansion functions $g_m[n]$ as indicated. The subscript m denotes an index set that describes the features of the building block $g_m[n]$, for instance time location, modulation, and scale. A wide variety of such decompositions, ranging from Fourier and sinusoidal models to wavelet and frame expansions, have been explored in the literature. These approaches find use in coding applications, where compression is often achieved by discarding components with low-valued expansion coefficients; in such cases the expansion is intended to provide an accurate but not necessarily perfect reconstruction of the signal.

For Fourier transforms, wavelets, and other expansions where the functions $g_m[n]$ constitute a basis $(N = M)$, the matrix D is invertible and the expansion coefficients α for a given signal are unique. These basis expansions exhibit a certain rigidity, however, in that a given basis is not well-suited for a wide variety of signals. Consider the Fourier case: for a time-localized signal, the frequency domain representation does not readily indicate the time localization; the Fourier analysis does not provide information about the relevant signal features. This shortcoming results from attempting to represent arbitrary signals in terms of a very limited set of functions. Better representations can be achieved by using a larger number of expansion functions, *i.e.* by choosing the $g_m[n]$ from a highly redundant dictionary that not only spans the signal space but also includes a wide range of functions beyond the spanning set; this enables appropriate representation of a wide range of time-frequency behaviors.

When the functions $g_m[n]$ constitute a redundant set $(M > N)$, the linear system in equation 1 is underdetermined. One solution is provided by the pseudo-inverse of D, which can be derived using the singular value decomposition (SVD); the weight vector $\tilde{\alpha} = D^+ x$ has the minimum two-norm of all solutions. This minimization of the two-norm in the SVD solution is inappropriate for compression, however, in that it tends to spread energy throughout all of the elements of $\tilde{\alpha}$; compression is only achievable by discarding elements below some threshold. In comparison to this SVD approach of computing a non-sparse solution and thresholding it, the goal of compression is better served by simply searching for a sparse approximate solution to the underdetermined inverse problem. One algorithm for computing such sparse approximate solutions is known as *matching pursuit* [1]. Investigation of matching pursuit and similar methods is readily motivated by the compaction improvement that can be achieved with respect to traditional linear methods such as the SVD. An example of this is given in figure 1, which shows a plot of the thresholded SVD expansion coefficients (solid) and a sparse matching pursuit solution (circles) for the same reconstruction error. To achieve the same error as the sparse representation, which has 16 non-zero elements, the SVD approach must use a very small threshold, which leads to poor compaction; the SVD solution retains 330 non-zero elements. In this simulation, the original signal is the sum of five dictionary elements, which means there is an exact solution with only five non-zero values. For reasons to be discussed, the matching pursuit does not find this optimal sparse solution; however, it does reliably identify the fundamental signal structure.

2. MATCHING PURSUIT

Matching pursuit is an iterative algorithm for deriving signal decompositions in terms of expansion functions chosen from a highly redundant dictionary [1]. Equivalently, it is

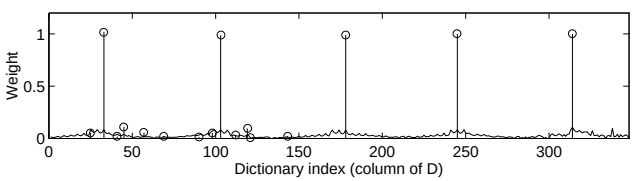

Figure 1. *A thresholded SVD solution (solid) and a sparse matching pursuit approximation (circles) to an underdetermined inverse problem; the representations have the same reconstruction error.*

an approach for computing sparse approximate solutions to underdetermined inverse problems; the dictionary elements, or *atoms*, correspond to the columns of D in the linear system. In [1], the dictionary consists of Gabor atoms, which are scaled, modulated, and translated versions of a single window function. This set of functions, which notably includes both Fourier and wavelet-like bases, is highly redundant when the scale, modulation, and translation parameters are not tightly restricted. This redundancy implies that the atoms exhibit a wide range of time-frequency behavior, and can thus provide better decompositions of a wide range of signals than a basis expansion.

Matching pursuit refers specifically to a greedy iterative algorithm for determining an expansion given a signal and a dictionary of atoms. At each stage of the iteration, the atom that best approximates a portion of the signal is chosen; then the weighted contribution of this atom to the signal is subtracted and the iteration proceeds on the residual. Mathematically, the task at the i-th stage is to find the atom $g_{m(i)}[n]$ that minimizes the two-norm of the residual signal

$$r_{i+1}[n] \;=\; r_i[n] - \alpha_i g_{m(i)}[n] \tag{2}$$

where α_i is a weight that describes the contribution of the atom to the signal, *i.e.* the expansion coefficient, and $m(i)$ is the dictionary index of the atom; the iteration begins with $r_1[n] = x[n]$. The solution for α_i and $g_{m(i)}[n]$ follows from the orthogonality principle; treating the signals as column vectors, the energy or two-norm of the residual r_{i+1} is a minimum if it is orthogonal to the atom:

$$\langle r_i - \alpha_i g_{m(i)}, g_{m(i)} \rangle \;=\; \left(r_i - \alpha_i g_{m(i)}\right)^H g_{m(i)} \;=\; 0 \tag{3}$$

$$\Longrightarrow\quad \alpha_i \;=\; \frac{\langle g_{m(i)}, r_i \rangle}{\langle g_{m(i)}, g_{m(i)} \rangle} \;=\; \langle g_{m(i)}, r_i \rangle$$

where the last step follows from restricting the atoms to be unit-norm. Then, the energy $\langle r_{i+1}, r_{i+1} \rangle$ of the error is

$$\langle r_i, r_i \rangle \;-\; \frac{|\langle g_{m(i)}, r_i \rangle|^2}{\langle g_{m(i)}, g_{m(i)} \rangle} \;=\; \langle r_i, r_i \rangle \;-\; |\alpha_i|^2. \tag{4}$$

This energy is minimized by choosing the atom $g_{m(i)}$ that has the largest magnitude correlation with the signal r_i, and the expansion coefficient for that atom is $\langle g_{m(i)}, r_i \rangle$.

In deriving a signal decomposition, the matching pursuit iteration is continued until the residual energy is below some threshold, or until some other halting criterion is met. After I iterations, the decomposition (or sparse approximate solution) corresponds to the estimate

$$x[n] \;\approx\; \sum_{i=1}^{I} \alpha_i g_{m(i)}[n]. \tag{5}$$

The mean-squared error of this approximation, namely the energy of the residual, converges to zero as the number of iterations approaches infinity [1]. This convergence property implies that I iterations will provide a reasonable I-term estimate; this I-term approximate solution, however, is in general not optimal in the mean-squared sense. Since the dictionary is not orthogonal, the term-by-term matching pursuit approach does not find the optimal I-term expansion; determining the optimal I-term expansion based on a non-orthogonal dictionary requires finding the minimum projection error over all I-dimensional dictionary subspaces, which is not computationally feasible for large I [2].

Though searching for the optimal high-dimension subspace is not reasonable, it is reasonable to consider the related problem of finding an optimal low-dimension subspace at each iteration of the matching pursuit algorithm. In this variation of the algorithm, the i-th iteration consists of searching for an $N \times J$ matrix G, whose J columns are dictionary atoms, that minimizes the two-norm of the residual $r_{i+1} = r_i - G\alpha$, where α is a $J \times 1$ vector of weights. This J-dimensional formulation is similar to the one-dimensional case; the orthogonality constraint $\langle r_i - G\alpha, G \rangle \;=\; 0$ leads to a solution for the weights:

$$\alpha \;=\; \left(G^H G\right)^{-1} G^H r_i \tag{6}$$

The energy of the residual is then given by

$$\langle r_i, r_i \rangle \;-\; r_i^H G \left(G^H G\right)^{-1} G^H r_i \tag{7}$$

which is minimized by choosing G so as to maximize the second term. Clearly, this approach is computationally expensive unless G consists of orthogonal vectors or possesses some other special structure.

One such structured case, which will prove useful in section 4, is the two-dimensional case where the two columns of G are an atom g and its complex conjugate; the general results can be significantly simplified for this case. Assume that the original signal is real and that g has non-zero real and imaginary parts so that G has full column rank and $G^H G$ is invertible. Then, letting $\Gamma = \langle g, g^* \rangle$ and $\beta = \langle g, r_i \rangle$, the metric to maximize through the choice of g is

$$\frac{1}{1 - |\Gamma|^2} \left(2|\beta|^2 - \Gamma(\beta^*)^2 - \Gamma^* \beta^2\right) \tag{8}$$

and the optimal weights are

$$\alpha \;=\; \begin{bmatrix} \alpha(1) \\ \alpha(2) \end{bmatrix} \;=\; \frac{1}{1 - |\Gamma|^2} \begin{bmatrix} \beta - \Gamma\beta^* \\ \beta^* - \Gamma^* \beta \end{bmatrix}. \tag{9}$$

Note that the above metric can also be written as

$$\beta^* \alpha(1) \;+\; \beta \alpha(1)^* \tag{10}$$

and that $\alpha(1) = \alpha(2)^*$. The new residual is then

$$r_{i+1} \;=\; r_i - 2\Re\{\alpha(1) g\}. \tag{11}$$

Note that the orthogonal projection of a real signal onto the subspace spanned by a conjugate pair is again real.

3. DAMPED SINUSOIDAL ATOMS

In many applications of Gabor functions, the function set is derived from an even-symmetric window, resulting in a dictionary of atoms that exhibit symmetric time-domain behavior. This underlying symmetry is problematic for representing asymmetric signal features such as transients, which occur frequently in natural signals such as music. Consider a typical transient, the damped sinusoid. Figure 2(a) shows a damped sinusoidal signal; the first stage of a matching pursuit based on symmetric Gabor functions chooses the atom shown in figure 2(b). This atom matches the frequency behavior of the signal, but its time-domain symmetry results in a *pre-echo* artifact in the residual as shown in figure 2(c). The residual has energy before the onset of the original signal, which the matching pursuit algorithm then attempts to remove at subsequent stages. One approach to this problem is the high-resolution matching pursuit algorithm suggested in [3], where symmetric atoms are still used in the pursuit, but the correlation metric is modified so that atoms that introduce such artifacts are not chosen for the signal decomposition. Another approach is to use asymmetric atoms such as damped sinusoids.

Damped oscillations occur commonly in natural signals and, by no coincidence, damped sinusoids are a fundamental part of linear system theory. In this light, damped sinusoids are a sensible candidate for use as building blocks in signal decompositions; it is physically reasonable to model a signal as a sum of damped sinusoids. Estimation of the parameters for such models has been explored in the literature in

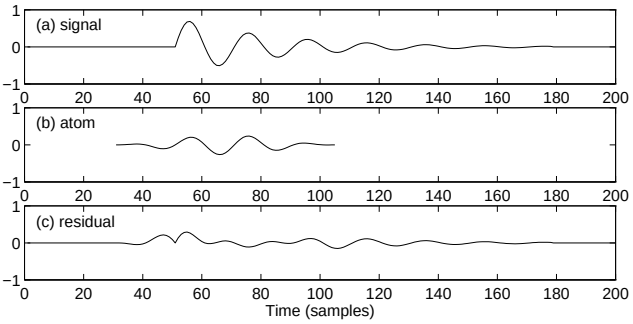

Figure 2. *(a) a damped sinusoidal signal, (b) the optimal first atom chosen from a symmetric Gabor dictionary, and (c) the residual; note the artifact near the onset time.*

the framework of system identification, spectral estimation, and signal modeling [4, 5, 6, 7]; these techniques, however, have difficulties when the time origin of the sinusoids is unknown. In [8], complex damped sinusoids are used to provide a time-frequency representation in which transients are identifiable; this method however assumes prior knowledge of the damping factor, which is inappropriate for the application of deriving decompositions of arbitrary signals. Matching pursuit using a dictionary of damped sinusoids overcomes the drawbacks of these various approaches.

Like the atoms in a Gabor dictionary – which are indexed by scale, modulation, and translation parameters – complex damped sinusoids can be indexed by characteristic parameters. The damping factor a_0, modulation frequency ω_0, and start time n_0 uniquely specify these atoms:

$$g[a_0, \omega_0, n_0] \;=\; S_0 \; a_0^{(n-n_0)} e^{j\omega_0 (n-n_0)} u[n-n_0] \qquad (12)$$

where S_0 is a scaling factor needed to satisfy the unit-norm requirement. For the sake of realizability, an atom is truncated when its amplitude falls below a threshold T; the corresponding length is $L = \lceil \log T / \log a_0 \rceil$, and the appropriate scaling factor is $S_0 = \sqrt{(1-a_0^2)/(1-a_0^{2L})}$. Note that this truncation results in sensible localization properties; heavily damped atoms are short-lived, and lightly damped atoms persist in time. Also note that the atoms are one-sided; an atom thus resembles the impulse response of a filter with a single complex pole; this is a suitable property given the intent of representing transient signals.

For the dictionary of complex atoms specified by equation 12, the correlations and hence the expansion coefficients for signal decompositions will generally be complex; the coefficient provides both a magnitude and a phase for the atom in the expansion. For real signals, decomposition in terms of such complex atoms can be misleading; for a signal that consists of one real damped sinusoid, the matching pursuit does not simply find the appropriate 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:

$$\hat{g}[a_0, \omega_0, n_0, \phi_0] \;=\; \hat{S}_0 a_0^{(n-n_0)} \cos\left[\omega_0 (n-n_0) + \phi_0\right] \qquad (13)$$

In the complex case, the matching pursuit dictionary is indexed by the three parameters $\{a_0, \omega_0, n_0\}$, and the phase of an atom in the expansion is given precisely by its correlation. In contrast, the real dictionary requires the phase parameter as an additional index. The phase is not supplied by the correlation computation; like the other parameters, it is estimated from a discretized set by the matching pursuit. This explicit appearance of the phase results in both a larger dictionary and a more complicated correlation computation. This problem can be circumvented by using the complex dictionary and considering expansions onto the subspace spanned by an atom and its conjugate as formulated in section 2. This framework provides a method for deriving real-valued expansions in terms of real damped sinusoids where the phase is provided by the correlation computation and an explicit search over a phase index is not required.

4. 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 $\langle g, r_i \rangle$ is thus intensive. As noted in [1], this computation can be reduced using an update formula derived from equation 2; the correlations at stage $i+1$ are given by

$$\langle g, r_{i+1} \rangle \;=\; \langle g, r_i \rangle \;-\; \alpha_i \langle g, g_{m(i)} \rangle \qquad (14)$$

where the only new computation required for the correlation update is the dictionary cross-correlation term $\langle g, g_{m(i)} \rangle$, which can be precomputed and stored if enough memory is available. For some dictionaries, the atomic structure can be exploited to simplify the correlation computation irrespective of this update formula. Such an approach is readily applicable to dictionaries of complex damped sinusoids.

4.1. Complex Damped Sinusoids

For the dictionary of complex damped sinusoids, correlations must be computed for every combination of damping factor, modulation frequency, and time shift. The correlation $\langle g, x \rangle$ of a signal $x[n]$ with an atom $g[a_0, \omega_0, n_0]$ is

$$\rho(a_0, \omega_0, n_0) \;=\; S_0 \sum_{n=n_0}^{n_0+L-1} x[n] \; a_0^{(n-n_0)} e^{-j\omega_0 (n-n_0)} \qquad (15)$$

where the atoms are truncated to a length L that is a function of the damping factor a_0. The structure of this correlation allows for substantial reductions in the computation requirements with respect to the time shift and modulation parameters. These are examined in turn.

A simplification of the correlation computation over the time index is provided by the exponential structure of the atoms, which results in a recursion relationship between correlations at neighboring times:

$$\begin{aligned} \rho(a_0, \omega_0, n_0 - 1) \;&=\; a_0 e^{-j\omega_0} \rho(a_0, \omega_0, n_0) \qquad (16) \\ &+ S_0 \;\left(x[n_0 - 1] \;-\; a_0^L e^{-j\omega_0 L} x[n_0 + L - 1] \right). \end{aligned}$$

This is simply a one-pole filter with a correction to account for the atom truncation. It is operated in reversed time to make the recursion stable for damped sinusoids; the similar forward recursion is unstable for $a_0 < 1$. This filter structure suggests interpreting the correlation computation over all possible indices $\{a_i, \omega_i, n_i\}$ as an application of the signal to a dense grid of one-pole filters in the z-plane, which are the *matched filters* for the dictionary atoms. The filter outputs are the correlations needed for the matching pursuit; the maximally correlated atom is directly indicated by the maximum magnitude output of the filter bank.

A further simplification can be achieved if the z-plane filterbank, *i.e.* the matching pursuit dictionary, is structured such that the modulation frequencies are equispaced for each damping factor. If the filters (atoms) are equispaced angularly on circles in the z-plane, the discrete Fourier transform can be used for the computation over ω_0. Specifically, for $\omega_0 = 2\pi k_0 / K$, the correlation is given by

$$\begin{aligned} \rho(a_0, k_0, n_0) \;&=\; S_0 \sum_{n=0}^{L-1} x[n+n_0] a_0^n e^{-j2\pi k_0 n / K} \qquad (17) \\ &=\; S_0 \; \mathrm{DFT}_K \left\{ x[n+n_0] \; a_0^n \right\} |_{k_0} \end{aligned}$$

which shows that FFT algorithms can be used to reduce the cost of the correlation computation over the frequency index. Note that such an FFT-based simplification can be applied to any dictionary of harmonically modulated atoms.

As mentioned earlier, correlations with complex atoms are generally complex; the computation of $\rho(a_0, \omega_0, n_0)$ thus provides both a magnitude and a phase for the signal expansion. For real atoms, however, the phase becomes part of the index set and appears explicitly in each atom. As a result, the structure of the correlation with real atoms does not allow for the simplifications discussed above. In the real case, then, not only is the dictionary larger because of the phase index, but the computation is adversely affected as well. These problems can both be avoided while maintaining the advantages of the complex approach by considering signal expansions in terms of conjugate pairs.

4.2. Real Damped Sinusoids

A decomposition in terms of real damped sinusoids can be arrived at by using the subspace matching pursuit algorithm discussed in section 2 to search for optimal conjugate pairs of complex damped sinusoids to use in the expansion. At the i-th iteration, this algorithm searches for the atom $g_{m(i)}$ that minimizes the two-norm of the residual

$$r_{i+1}[n] \;=\; r_i[n] - \alpha_i(1) g_{m(i)}[n] - \alpha_i(2) g^*_{m(i)}[n]. \qquad (18)$$

As shown in section 2, if $r_i[n]$ is real, the expansion coefficients are conjugates and the new residual r_{i+1} is also real. Equations 9 and 10 in section 2 show that the expansion coefficients and the maximization metric in this pursuit are both functions of the correlation of the residual with the underlying complex dictionary atoms; this means that the computational simplifications for the complex dictionary can be readily applied to calculation of a real expansion. The decomposition found by this approach is

$$2 \sum_i \Re \left\{ \alpha_i(1) g_{m(i)}[n] \right\} \qquad (19)$$

which can be written explicitly as

$$2 \sum_i S_i A_i a_i^{(n-n_i)} \cos\left[\omega_i (n - n_i) + \phi_i\right] \qquad (20)$$

where $A_i e^{j\phi_i} = \alpha_i(1)$. 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 real decomposition is derived without requiring the phase as a dictionary index. Furthermore, the dictionary for this conjugate search is effectively half the size of the full complex dictionary since atoms are considered in conjugate pairs.

One caveat to note is that the conjugate pursuit algorithm breaks down if the atom g is purely real; this occurs because the derivation in section 2 requires that the atom and its conjugate be linearly independent, *i.e.* the atom must have non-zero real and imaginary parts. Thus, a fix is required if real unmodulated exponentials, which are elements of the complex dictionary, are to be admitted in the real signal expansions. The i-th stage of the fixed overall algorithm is as follows: first, the correlations $\beta = \langle g, r_i \rangle$ for the entire dictionary of complex atoms are computed using the simplifications of section 4. Then, energy-minimizing metrics for both types of atoms are computed and stored: for real g, the metric is $|\beta|^2$ as indicated in equation 4; for complex g, the metric is $\beta^* \alpha(1) + \beta \alpha^*(1)$, where $\alpha(1) = (\beta - \Gamma\beta^*)/(1 - |\Gamma|^2)$ and $\Gamma = \langle g, g^* \rangle$ is given by

$$\Gamma(a_0, \omega_0) \;=\; S_0^2 \left(\frac{1 - a_0^{2L} e^{-j2\omega_0 L}}{1 - a_0^2 e^{-j2\omega_0}} \right). \qquad (21)$$

These metrics provide the proper comparison of the energy reduction of the residual for 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}.

5. CONCLUSIONS

Generalizing the matching pursuit algorithm to finding optimal dictionary subspaces provides a framework for computing real expansions of real signals using complex atoms. Fundamentally, this is based on projecting onto the subspace spanned by an atom and its conjugate. This method is advantageous over matching pursuit with real atoms in that it simplifies the treatment of the phase of the expansion components. For decompositions in terms of real damped sinusoids, this method allows for the use of fast correlation algorithms that apply to an underlying dictionary of complex damped sinusoids. The use of damped sinusoids for signal decompositions is motivated by the commonality of damped oscillations in natural signals and the shortcomings of symmetric atoms for representing transient signal behavior. Future work includes exploration of this approach and other similarly aggressive methods for audio coding and analysis–modification–synthesis as well as comparisons with the high-resolution matching pursuit discussed in [3].

6. ACKNOWLEDGEMENTS

The author thanks Martin Vetterli for various contributions to this work and Vivek Goyal for a number of valuable insights, as well as Edward Lee for his continuous support.

REFERENCES

[1] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, December 1993.

[2] G. Davis. *Adaptive Nonlinear Approximations.* PhD thesis, New York University, 1994.

[3] R. Gribonval et al. Analysis of sound signals with high resolution matching pursuit. *Proceedings of Time-Frequency Time-Scale Symposium*, June 1996.

[4] S. Kay and S. Marple. Spectrum analysis – a modern perspective. *Proceedings of the IEEE*, 69(11):1380–1419, November 1981.

[5] R. Kumaresan and D. Tufts. Estimating the parameters od exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 30(6):833–840, December 1982.

[6] R. Roy and T. Kailath. ESPRIT - Estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(7), July 1989.

[7] J. Laroche. The use of the matrix pencil method for the spectrum analysis of musical signals. *Journal of the Acoustical Society of America*, 94(4):1958–1965, October 1993.

[8] B. Friedlander and A. Zeira. Oversampled Gabor representation for transient signals. *IEEE Transactions on Signal Processing*, 43(9):2088–2094, September 1995.