Queen Mary
University of London

Analysis, Visualization, and Transformation of Audio Signals Using Dictionary-based Methods

Sturm, BL; Roads, C; McLeran, A; Shynk, JJ

© 2009 Routledge

For additional information about this publication click this link.
http://qmro.qmul.ac.uk/xmlui/handle/123456789/11290

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact scholarlycommunications@qmul.ac.uk

Analysis, Visualization, and Transformation of Audio Signals Using Dictionary-based Methods

Bob L. Sturm, Curtis Roads[†], John J. Shynk, and Aaron McLeran[*]

March 6, 2009

Abstract

In this article we provide an overview of dictionary-based methods (DBMs) — also called sparse approximation — and review recent work in the application of such methods to working with signals, in particular audio and music signals. As Fourier analysis is to additive synthesis, DBMs can be seen as the analytical counterpart to granular synthesis since a signal is rebuilt by a linear combination of heterogeneous *atoms* selected from a user-defined *dictionary*. We demonstrate how DBMs provide novel means for analyzing, visualizing, and transforming audio signals by creating multiresolution and parametric descriptions of their contents.

1 Introduction

The development of dictionary-based methods (DBMs) — also called sparse approximation, and atomic decomposition — has been motivated by the desire to represent signals having diverse structures in ways that are more efficient, robust to noise, meaningful, and malleable than can be obtained using standard linear transform methods with orthogonal bases [Mallat and Zhang, 1993, Chen et al., 1998, Mallat, 1999]. Efficiency refers to the

[*]Bob L. Sturm and John J. Shynk are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA, 93106-9560 USA. E-mail: {boblsturm, shynk}@ece.ucsb.edu. Phone: 805-893-3977 Fax: 805-893-3262. C. Roads ([†] contact author) and Aaron McLeran are with the Media Arts and Technology Program, University of California, Santa Barbara, CA 93106-6065 USA. E-mail: {clang@mat.ucsb.edu, amcleran@gmail.com} Phone: 805-893-5244 Fax: 805-893-2930.

number of terms used to describe a signal to some degree of precision. Noise robustness implies that the representation will not be greatly affected by the corrupting influence of noise. Malleability refers to the possibility of manipulating parameters of the signal model to bring about desired changes, for instance source separation, or pitch shifting. Finally, meaningfulness refers to a clarity in the correspondence between model parameters and signal content.

Instead of transforming a signal into a new function space, e.g., the Fourier transform takes a signal from the time-domain into the frequency-domain, a DBM adapts a representation to the signal with respect to a model through selection of the "best" subset of functions from a user-defined and usually redundant and overcomplete *dictionary*. When the functions used are *atomic*, i.e., localized in time, the end result of the decomposition embodies a "score" to reproduce the given signal with atoms [Gabor, 1947]. In this sense, DBMs can be seen as the analytical equivalent to granular synthesis [Xenakis, 1971, Roads, 2001], but their application is much wider than this. Researchers have applied DBMs to the coding and compression of audio [Lewicki, 2002, Christensen and van de Par, 2006] and image [Figueras i Ventura et al., 2006] data; denoising and data recovery [Elad and Aharon, 2006, Aharon et al., 2006], blind source separation [Lesage et al., 2006]; musical analysis and transcription [Derrien, 2006, Leveau et al., 2008], etc. Recent related works have focused on compressive sampling [Candès and Wakin, 2008], where the use of DBMs for compressible signals allows one to sample at rates much lower than that specified by the Nyquist-Shannon sampling theorem.

In this article we discuss the use of DBMs to provide a rich and flexible interface to analyzing, visualizing, and transforming the *content* in audio signals, e.g., transients, harmonics, notes, voices. After providing an overview of DBMs, we review our recent work in these areas, significantly extending the results presented in [Kling and Roads, 2004]. Throughout we discuss the theoretical and practical benefits of DBMs, as well as some of their problems, such as *uniqueness* and *dark energy*. Atomic decomposition can provide an efficient and meaningful representation of an acoustic signal, and we show that these properties can be enhanced by considering the interactions between atoms in the signal

model. Furthermore, we propose that through the molecules described in Section 3.1, a sparse approximation can ultimately be used to provide an interface to the contents of the signal at multiple levels of detail. Finally, we present several novel sound transformations via atomic representations.

The following notation is used throughout the text: column vectors are bold lower-case, matrices are bold upper-case, ^{H} denotes conjugate transpose, and ^{T} denotes transpose. The ith element of a vector $\mathbf{g}$ is notated $\mathbf{g}_i$. The inner-product of two vectors is defined $\langle \mathbf{g}, \mathbf{h} \rangle \triangleq \mathbf{h}^H \mathbf{g}$. The ℓ_2-norm is induced by the inner product of a vector with itself $||\mathbf{g}||_2^2 \triangleq \langle \mathbf{g}, \mathbf{g} \rangle$.

2 Overview of Dictionary-based Methods

Consider a real sampled signal represented by a column vector $\mathbf{x}$ of length K. We wish to find a way to describe $\mathbf{x}$ as a linear combination of N waveforms specified a priori as unit-norm columns of length-K in a *dictionary* $\mathcal{D}$, notated in matrix form as $\mathbf{D}$. More formally, we want to find a solution $\mathbf{s}$ producing a minimum squared-error, i.e.,

$$\min_{\mathbf{s}} ||\mathbf{x} - \mathbf{D}\mathbf{s}||_2^2 = \min_{\mathbf{s}} \left|\left| \mathbf{x} - \sum_{n=1}^{N} \mathbf{d}_i \mathbf{s}_i \right|\right|_2^2 \leq \epsilon \tag{1}$$

where $\epsilon \geq 0$ is some minimum squared error, and $\mathbf{s}$ is a column vector of N real weights, one for each function in the dictionary. If $N = K$ and $\mathbf{D}$ is the orthonormal discrete Fourier transform (DFT) matrix, then the best solution $\mathbf{s} = \mathbf{D}^H \mathbf{x}$ is simply the discrete Fourier transform of $\mathbf{x}$. Usually in DBMs, however, $N \gg K$ and $\text{rank}(\mathbf{D}) = K$, which is the meaning of the term *overcomplete*. In such a case, one possible solution to (1) can be found using least-squares projection [Meyer, 2001]:

$$\mathbf{s} = \left[\mathbf{D}^H \mathbf{D}\right]^{-1} \mathbf{D}^H \mathbf{x}. \tag{2}$$

However, this solution will most likely not be efficient as many elements in $\mathbf{s}$ will be non-zero. This is demonstrated by the following example.

Consider the simple length-4 vector $\mathbf{x}$ shown in Fig. 1(a). The magnitudes of its DFT are shown in (b). We can also represent $\mathbf{x}$ using elements from a dictionary of a size-16×16 DFT matrix truncated to be size 4×16. Such a dictionary is overcomplete in the signal space, which implies that there exists an infinity of solutions to (1). Figure 1(c) shows the magnitudes of the solution found by least-squares projection in (2), which we see is not efficient. The magnitudes of another possible solution are shown in Fig. 1(d). The most efficient solution in this dictionary is shown in Fig. 1(e).

Since for any real $\mathbf{x}$ and an overcomplete dictionary there will exist an infinity of solutions to (1), none will be unique unless constraints on $\mathbf{s}$ are also specified. The constraint implicit in (2) is the minimization of $||\mathbf{s}||_2$, i.e., the length of the solution; but as shown in Fig. 1, this solution is not sparsity preserving. In terms of sparsity, the best solution is the one having the fewest non-zero elements. This, however, makes finding the best $\mathbf{s}$ unsolvable in a reasonable amount of time [Davis et al., 1997] since it requires inspecting all m-tuples of atoms in the dictionary. For the simple example in Fig. 1, finding the most efficient solution is simple, but for larger scale problems it becomes combinatorially prohibitive. Toward this end, several methods, or *pursuits*, have been proposed to find solutions to (1) that are satisfactorily efficient. We now review several of these methods.

2.1 Basis Pursuit Algorithm

Basis Pursuit (BP) [Chen et al., 1998] proposes solving (1) by the following linear programming principles:

$$\min_{\mathbf{s}} ||\mathbf{s}||_1 \text{ such that } \mathbf{x} = \mathbf{D}\mathbf{s} \tag{3}$$

$$\min_{\mathbf{s}} ||\mathbf{s}||_1 \text{ such that } ||\mathbf{x} - \mathbf{D}\mathbf{s}||_2^2 \leq \epsilon \tag{4}$$

where the 1-norm of a vector is defined as the sum of the magnitudes of its elements. The first statement requires the solution to reproduce the original signal exactly, while the second statement assumes that the signal of interest is contaminated by noise in $\mathbf{x}$. The solution to (3) and (4) can be found through approaches in convex optimization

[Boyd and Vandenberghe, 2004], which, while not combinatorial, are expensive for large dictionaries and high-dimensional signals such as acoustic signals.

Our work has focused on another set of approaches that instead solve the problem in (1) using an iterative descent strategy. These methods involve minimizing an intermediate distortion at each iteration of the decomposition, and are straightforward and simple to implement [Krstulovic and Gribonval, 2006]; but often the resulting solutions can contain artifacts from poor atom selections by the algorithm, and a mismatch of the model to the signal — discussed in Section 2.9. We now review several of these iterative descent pursuit algorithms.

2.2 Matching Pursuit Algorithm

Matching Pursuit (MP) [Mallat and Zhang, 1993] iteratively refines the representation of a signal by reducing its squared error with each atom selection. Consider the nth-order *representation* of a signal $\mathbf{x}$, defined as $\mathcal{X}_n \triangleq \{\mathbf{H}(n), \mathbf{a}(n), \mathbf{r}(n)\}$, where $\mathbf{H}(n) \triangleq [\mathbf{h}_0|\mathbf{h}_1|\cdots|\mathbf{h}_{n-1}]$ is a matrix of n columns selected from the dictionary, $\mathbf{a}(n)$ is a length-n column vector of weights, and $\mathbf{r}(n)$ is the nth-order residual, such that the model of the signal is expressed

$$\mathbf{x} = \mathbf{H}(n)\mathbf{a}(n) + \mathbf{r}(n) \tag{5}$$

$$= \hat{\mathbf{x}}(n) + \mathbf{r}(n) \tag{6}$$

where $\hat{\mathbf{x}}(n) \triangleq \mathbf{H}(n)\mathbf{a}(n)$ is the nth-order approximation of $\mathbf{x}$. (Note that here n specifies the order of the model, or iteration of the decomposition process, and is not a time index of the signal—for which we use k.) MP is initialized with $\mathbf{r}(0) = \mathbf{x}$. MP updates this representation by adding a new column to the representation basis $\mathbf{H}(n)$, a new row to the weights $\mathbf{a}(n)$, and updating the residual, i.e.:

$$\mathcal{X}_{n+1} = \left\{ \begin{array}{l} \mathbf{H}(n+1) = [\mathbf{H}(n)|\mathbf{h}_n], \\ \mathbf{a}(n+1) = [\mathbf{a}(n)^T, \langle \mathbf{r}(n), \mathbf{h}_n \rangle]^T, \\ \mathbf{r}(n+1) = \mathbf{r}(n) - \langle \mathbf{r}(n), \mathbf{h}_n \rangle \mathbf{h}_n \end{array} \right\}. \tag{7}$$

Each atom is selected according to the criterion

$$\mathbf{h}_n = \arg\max_{\mathbf{d}\in\mathcal{D}} \frac{|\langle \mathbf{r}(n), \mathbf{d}\rangle|^2}{||\mathbf{d}||_2^2}. \tag{8}$$

This atom selection rule comes from minimizing the squared error $||\mathbf{r}(n) - \langle \mathbf{r}(n), \mathbf{d}\rangle \mathbf{d}||_2^2$. Essentially, this selects the dictionary waveform that is most correlated with the current residual signal. This process is repeated until the residual energy $||\mathbf{r}(n)||_2^2$ is lower than some limit, or a specified estimation order n has been reached. A simple example of MP decomposition is shown in Fig. 2. It should be noted that a signal decomposed using a DBM need not be windowed — as is done using the short-term Fourier transform — which avoids an arbitrary segmentation of a signal.

2.3 Orthogonal Matching Pursuit Algorithm

Orthogonal MP (OMP) [Pati et al., 1993] performs the additional step of orthogonalizing the residual for all selected atoms, which in effect recomputes every weight of the representation. Given the nth-order representation $\mathcal{X}_n$, the update rule for OMP is given by:

$$\mathcal{X}_{n+1} = \left\{ \begin{array}{l} \mathbf{H}(n+1) = [\mathbf{H}(n)|\mathbf{h}_n], \\ \mathbf{a}(n+1) = \mathbf{H}^{\dagger}(n+1)\mathbf{x}, \\ \mathbf{r}(n+1) = \mathbf{x} - \mathbf{H}(n+1)\mathbf{a}(n+1) \end{array} \right\} \tag{9}$$

where $\mathbf{H}^{\dagger} \triangleq (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$, and each atom is selected by (8). The change here is the orthogonal projection of $\mathbf{x}$ onto the new representation basis $\mathbf{H}(n+1)$, which ensures that each intermediate residual signal is orthogonal to the space spanned by the representation basis.

2.4 Optimized Orthogonal Matching Pursuit Algorithm

A further optimization to MP is made in Optimized OMP (OOMP) [Rebollo-Neira and Lowe, 2002], where each atom is selected such that the residual energy will be reduced by the maximum amount possible. OOMP updates the nth-order representation $\mathcal{X}_n$ according to the rule in

(9), but with each atom selected by the criterion

$$\mathbf{h}_n = \arg\max_{\mathbf{d}\in\mathcal{D}} \frac{|\langle \mathbf{r}(n), \mathbf{d}\rangle|^2}{||\mathbf{d}_{\mathcal{H}_n^\perp}||_2^2} \tag{10}$$

where $\mathbf{d}_{\mathcal{H}_n^\perp} \triangleq \mathbf{d} - \mathbf{H}(n)\mathbf{H}^\dagger(n)\mathbf{d}$.

For four example signals, Fig. 3 shows how the three pursuits MP, OMP, and OOMP perform in terms of how the residual energy $||\mathbf{r}(n)||_2^2$ decays as a function of pursuit iteration n, or, equivalently, model order in (6). Each pursuit was performed with the same dictionary of functions, a union of Gabor atoms and Dirac spikes (explained further below). It can easily be seen that OOMP outperforms both OMP and MP with respect to this metric.

2.5 Other Pursuits

Other methods for solving (1) have been proposed and studied. For instance, High Resolution Pursuit [Jaggi et al., 1998, Gribonval et al., 1996] attempts to create signal models that accurately embody features of the data. Molecular Matching Pursuit (MMP) [Daudet, 2006, Leveau et al., 2008] takes advantage of knowing the structures expected in acoustic signals, i.e., transients and tonals, to create sparse and structured models of each. Psychoacoustic Adaptive MP [Heusdens et al., 2002] provides a perceptual measure for selecting the atoms of a model. And Cyclic Matching Pursuit (CMP) [Christensen and Jensen, 2007] refines the estimates of the parameters of atoms in the model to reduce its distortion.

2.6 Dictionaries

The performance of a pursuit in finding a solution to (1) depends in a highly complex way on the selected dictionary. In order for a pursuit to be capable of yielding an efficient representation of a signal, the dictionary must have elements that are sufficiently similar to the structures in the signal. For instance, a dictionary of sinusoids is not well-suited to model a signal of impulses; and a dictionary of impulses is not well-suited to model a signal of sinusoids. However, a dictionary that contains sinusoids *and* impulses is well-suited to model signals having either or both types of structures [Donoho and Huo, 2001]. This can

be seen in the results of decomposing the simple signal in Fig. 2.

The freedom of selecting the contents of a dictionary is a major advantage of DBMs over traditional transforms. This gives a user the ability to make a decomposition adaptable to specific structures in a signal. There are three ways of creating dictionaries: combining bases, modulating prototype lowpass functions, and learning dictionary functions. MMP [Daudet, 2006] combines two bases into a dictionary, one for each expected type of content. For tonal content it uses the set of monoresolution atoms from the modified discrete cosine transform (MDCT); and for transient content it uses a tree of dyadic wavelets. One can also combine several MDCT bases of different scales [Ravelli et al., 2008] to create a multiresolution dictionary for audio coding purposes.

Another way to create a dictionary is by combining families of discretized, scaled, translated, and modulated lowpass functions $h(k; s)$. One simple example of a real dictionary waveform of length K is

$$g(k) = Ah(k - u; s)\cos\big(k\omega + \phi\big) \tag{11}$$

where $0 \leq k \leq K - 1$ is a time index, $0 \leq u < K - s/2$ is a translation, $1 \leq s \leq K$ is the scale in samples, and $0 \leq \omega \leq \pi$ and $0 \leq \phi < 2\pi$ are the modulation frequency and phase, respectively. Atoms with quadratic phase, such as chirps [Gribonval, 2001], or harmonic atoms [Gribonval and Bacry, 2003], can also be created. The scalar A is set for an atom such that it has unit norm, i.e., $\sum |g(k)|^2 = 1$. The shape of each waveform is specified by $h(k; s)$, which can be likened to a window. For instance, a Gabor atom [Gabor, 1946] consists of a translated and truncated discrete Gaussian function:

$$h(k; s) = \begin{cases} \exp\big[-\frac{(k-s/2)^2}{2(\alpha s)^2}\big], & k = 0, 1, \ldots, s-1 \\ 0, & \text{else} \end{cases} \tag{12}$$

where α controls the variance, and s is the scale. An example Gabor atom is shown in Fig. 4. To create a dictionary of Gabor atoms — also called a *time-frequency dictionary* [Mallat and Zhang, 1993] — each column of $\mathbf{D}$ is created by evaluating the functions in (11) and (12) at a number of different scales, translations, and modulations. A benefit to using

such dictionaries is that each atom in the model is associated with a set of parameters that are in themselves meaningful, for instance, modulation frequency. This provides ways to analyze, visualize, and modify the content of a decomposed signal.

Dictionaries in sparse approximation can also be learned, much the way a vector codebook is learned in applications of digital communications [Gersho and Gray, 1991]. Various methods that have been studied use independent component analysis [Lewicki and Sejnowski, 2000, Abdallah and Plumbley, 2006], maximum a posteriori estimation [Kreutz-Delgado et al., 2003], a generalized Lloyd algorithm or K-means singular value decomposition [Aharon et al., 2006, Elad and Aharon, 2006], and eigenvector decomposition [Jost et al., 2006]. Though atoms learned from these methods will have limited meaningfulness with regards to describing a signal in a parametric way, the pursuit of a signal model using such a dictionary can excel in terms of efficiency [Lewicki and Sejnowski, 2000, Elad and Aharon, 2006].

These methods for creating a dictionary can also be blended together. For instance, considering the harmonic atoms above, one can learn a good set of parameters from signals so that content can be identified and separated, e.g., harmonic atoms associated with specific musical instruments [Leveau et al., 2008]. Decomposing a musical signal using such dictionaries can simultaneously provide a separation and description of sources in a signal [Daudet, 2006].

2.7 Example Decomposition

Consider again the simple signal shown in Fig. 2(a), which has three periods of a sinusoid and a single impulse. Decomposing this signal using MP with a dictionary that is a union of Gabor atoms and Dirac impulses produces a good approximation using the first five atoms as shown. The first four atoms embody aspects of the sinusoid, and the fifth atom embodies the impulse. Even when this signal is embedded in additive white Gaussian noise (AWGN) at a level of 5 dB signal-to-noise energy ratio (SNR), the signal model built by MP in Fig. 2(b) is quite similar to the previous one. This shows the robustness possible with pursuit methods.

2.8 Wivigrams

We can obtain a picture of how energy is distributed in a pursuit representation by superposing the Wigner-Ville distribution (WVD) [Preis and Georgopoulos, 1999] of each scaled atom [Mallat and Zhang, 1993] — which we call a *wivigram*. The WVD of a Gabor atom is a two-dimensional Gaussian, centered on its modulation frequency and time translation. Its spread in time is proportional, and its spread in frequency is inversely proportional, to the variance of the Gaussian function — given by $(\alpha s)^2$ in (12). Of all possible time-frequency structures, Gabor atoms have the smallest spread of energy in time and frequency [Gabor, 1946].

2.9 Artifacts of Greedy Iterative Descent Methods

With the freedom in selecting the contents of a dictionary, and the process of decomposition, also comes artifacts from the mismatch between the signal and a model, and poor atom selections by a pursuit. Consider the signal in Fig. 5, built using 57 Gabor atoms of scale 64 samples. If we decompose this time-domain signal using MP and a Gabor dictionary—which includes the same atoms used to build the signal— the representation found is very different from the most sparse and efficient solution. The arrows labeled "1" show the small-scale atoms that coincide with the spikes in the time-domain signal, which are among the first five atoms selected by MP. Similarly, MP decomposes each vertical portion of the letters "U," "C," and "B," into small-scale wideband atoms without considering them the result of atoms of a larger scale that are in-phase. The arrows labeled "2" point to atoms at frequencies that do not exist in the original signal; and the arrows labeled "3" point to atoms at times where the original signal has no energy. These "spurious" terms reveal less about the original signal, and more about the way it is represented [Jaggi et al., 1998, Gribonval et al., 1996, Goodwin and Vetterli, 1999, Sturm, 2009, Sturm et al., 2008b]. They can be seen as self-correction by the pursuit for atoms that are mismatched to the signal, or atoms that have been poorly selected. These are discussed further in Section 3.2.

The "ideal" decomposition is, in a sense, lost from the very first atom selections, which is a clear example of how iterative descent pursuits can choose poorly from the first iteration.

We can make MP recover the original model (57 Gabor atoms) by specifying a Gabor dictionary having only atoms of scale 64 samples; but without this knowledge there is little we can do. These results demonstrate three important aspects of DBMs in general. First and foremost, the content of a dictionary has significant impact on the performance of the decomposition algorithm, and also the usefulness of the resulting representation. Second, since an overcomplete dictionary by definition provides several possible ways to approximate a given signal, any solution will lack uniqueness, and could be quite different from the "ideal" or expected representation. Third, characteristics of a decomposition algorithm, for instance, the greediness inherent in the selection of dictionary waveforms in MP, can manifest in unexpected ways and fill the model with erroneous atoms, such as atoms in time and frequency regions where no energy exists in the original. We discuss these phenomena further in Section 3.2.

3 Application to Musical Signal Analysis

DBMs can produce representations that are smaller in dimension than those provided by other transform methods, such as the short-time Fourier transform (STFT). Specifying the contents of the dictionary gives a pursuit the ability to adapt a representation to particular data or signals. With a good choice of a dictionary, the dimensionality of the original signal becomes much smaller, which can be taken advantage of by applications of signal and model analysis.

3.1 Higher-level Representations Through Molecules

To work with signal content that is represented by multiple dictionary waveforms, such as an attack, harmonic, or complete note, one must first find and delimit the atoms that are related. We have designed an algorithm that builds molecular representations from atomic ones [Sturm et al., 2008c, Sturm et al., 2008d, Sturm, 2009]. Each molecule is a group of atoms that act together to represent a high-level feature. This approach is inspired by the McAulay-Quatieri algorithm [McAulay and Quatieri, 1986], where a STFT is used in-

stead of a sparse atomic model to build a parametric sinusoidal model of speech. MMP [Leveau et al., 2008, Daudet, 2006] takes a similar approach, except molecules are built simultaneously with the signal decomposition according to two separable models for transient and tonal content.

The wivigram at top in Fig. 6 visualizes an MP decomposition of a bird call signal. Here we show *time-frequency tiles* [Mallat, 1999] to emphasize the overlap between terms. Using an agglomerative clustering approach, our algorithm combines atoms into molecules based on simple measures of similarity in time and frequency [Sturm et al., 2008c]. Examples of the resulting tonal molecules are outlined in the bottom of Fig. 6. The relationships between atoms and specific signal content are now more clear, and thus one can work more directly with the content of a signal through its atomic decomposition. Post-processing can refine each structure to remove artifacts of greedy pursuits, and model mismatches.

3.2 Dark Energy and Interference

Because of the non-orthogonal nature of the dictionaries typically used in DBMs, atoms may interact with each other in the model (6) of a signal [Sturm et al., 2008b, Sturm et al., 2008a, Sturm, 2009]. In the most extreme case, an atom of a representation will completely disappear in the resynthesis through destructive interference with other atoms. Several examples of this are seen in Fig. 7. Because of this effect we call all interaction exhibited by a non-orthogonal representation *dark energy* [Sturm et al., 2008b, Sturm et al., 2007]. Such terms are created by the decomposition algorithm to correct for "poor" atom choices made in earlier iterations, and mismatches between the signal and its model. These obviously reduce the efficiency and meaningfulness of any sparse signal representation.

We have proposed and studied several measures of model efficiency and meaningfulness in DBMs based on the interference between atoms. We define the *interference* associated with a real atom $\mathbf{h}_m$ in the real representation $\{\mathbf{H}(n), \mathbf{a}(n), \mathbf{r}(n)\}$ for $m = 0, 1, \ldots n-1$ over

the entire signal space [Sturm, 2009, Sturm and Shynk, 2008]:

$$\Delta(m) = \frac{1}{2}\left[||\hat{\mathbf{x}}(n)||_2^2 - (||\hat{\mathbf{x}}(n) - a_m\mathbf{h}_m||_2^2 + |a_m|^2)\right] \tag{13}$$

$$= \langle \hat{\mathbf{x}}(n) - a_m\mathbf{h}_m, a_m\mathbf{h}_m \rangle. \tag{14}$$

This expression defines the interference associated with the atom $\mathbf{h}_m$ as proportional to the extent to which it already exists in the current approximation from the other atoms. If $\Delta(m) < 0$, then $\mathbf{h}_m$ is negatively correlated with the signal model, and thus destructively interferes with it. And if $\Delta(m) > 0$, then $\mathbf{h}_m$ is positively correlated with the signal model, and thus constructively interferes with it — i.e., it contributes to representing the signal, and not to correcting the model by removing parts of other atoms.

Figure 8 shows how we can classify the elements of a signal model based on the sign of the interferences. Notice that all atoms appearing before the attack are placed in the set of destructively interfering atoms. These atoms do not productively contribute to the description of the signal, and rather serve to correct the model. For purposes of signal analysis, it is important to be able to distinguish between such features of a sparse approximation to make sure the analysis is of the signal and not of the model or pursuit.

We may also define a short-term measure of interference, or dark energy [Sturm et al., 2007], to see how interference is spread throughout a representation with respect to the signal. This helps answer questions about where a sparse approximation can be trusted as an efficient and accurate representation of the signal. Figure 9 shows how dark energy in a MP decomposition of a musical signal (using a Gabor dictionary) is often concentrated around times of transients. In these regions MP is attempting to represent the asymmetric onsets of energy using greedily selected large-scale symmetric Gabor atoms.

Other researchers have attempted to avoid these artifacts by changing the selection criterion used in MP [Gribonval et al., 1996, Jaggi et al., 1998], or by specifying different functions for the dictionary [Goodwin and Vetterli, 1999]. We have instead sought ways to productively use this behavior to learn about the signal and its relationship to the dictionary, and to measure the efficiency and meaningfulness of a decomposition, and the effective-

ness of a pursuit [Sturm et al., 2008b, Sturm et al., 2007, Sturm et al., 2008a, Sturm, 2009, Sturm and Shynk, 2008]. Instead of selecting atoms based on (8) for MP and OMP, or (10) for OOMP, atoms are selected with considerations of interference, for example in MP and OMP [Sturm, 2009, Sturm and Shynk, 2008]:

$$\mathbf{h}_n = \arg\max_{\mathbf{d}\in\mathcal{D}} |\langle \mathbf{r}(n), \mathbf{d}\rangle| + \lambda(n)\langle \mathbf{r}(n), \mathbf{d}\rangle\langle \hat{\mathbf{x}}(n), \mathbf{d}\rangle \tag{15}$$

where the last term is just the interference (14) associated with the atom $\mathbf{d}$, and $\lambda(n)$ weights its importance at the nth pursuit iteration. In such case, the pursuit adapts not only to the signal and its error, but also how the signal has been modeled so far.

Figure 10(a) shows the increase in model efficiency when interference ($\lambda(n) \neq 0$) is considered, as in (15) for OMP. We see that to model the signal at a SRR of 60 dB, we can reduce the model order from 67 to 47 atom by considering interference. The time-domain distribution of atoms in Fig. 10(b) also shows an increase in the correspondence between atoms and the signal. When $\lambda(n) = \lambda = 0$, we see a large number of atoms placed at a time when no energy exists in the signal. By considering interference these non-informative atoms are eliminated from the model, and we are left with a description of the signal that is much more clear and useful for applications of analysis.

4 Applications to Musical Signal Visualization

The decomposition of data into meaningful heterogeneous units provides novel ways to see, find, and work with a variety of content at many different resolutions. We have explored the use of DBMs to provide low- and high-level structured representations of audio signals and their morphological features, as well visualizing the results of decompositions with wivigrams. For instance, we compiled the wivigrams of a decomposition of the electroacoustic composition *Concrete PH* by Iannis Xenakis to produce a scrolling animation of it, a still of which is seen in Fig. 11.

Comparing the visualizations created using different time-frequency decomposition methods provides insight into how DBMs provide a novel alternative to picturing and working

with sound. The time-domain signal shown in Fig. 12(a) is a short musical excerpt from *Pictor Alpha* [Roads, 2004]; and below it are three different representations. The spectrogram, or log magnitude short-term Fourier transform (STFT), is shown in Fig. 12(b), and was created using a Hann window of length 5.8 ms and a constant overlap of 99%. It is possible to determine when and where energy exists in both time and frequency, but finding and delimiting particular structures is difficult. The scalogram in Fig. 12(c) shows the magnitudes of a dyadic wavelet transform (DWT) using the Gabor wavelet [Mallat, 1999]. Precise times of sharp discontinuities in the original signal (e.g., ≈ 22 ms) can be found, in addition to a concentration of energy at wavelets with larger scales. The wivigram in Fig. 12(d) is significantly less redundant than both the scalogram and spectrogram, and is able to simultaneously resolve various aspects of the signal at high and low frequencies and large and small scales — such as transient and tonal structures. Furthermore, since these atoms are parameterized, we automatically have a means to make adjustments to the model itself.

Using DBMs with a multiresolution dictionary (e.g., even a union of a wavelet and Fourier basis [Daudet, 2006]), one can separate the stationary and transient content of an audio signal. Figure 13 shows two wivigrams of atoms from an MP decomposition of a glockenspiel signal separated based on scale. This clearly separates the signal structures associated with the attacks from those associated with the ringing tones. Further, by advancing the work of agglomerating terms in atomic decompositions, presented in Section 3.1, many other options for visualization become possible.

5 Applications to Musical Signal Transformation

Describing a sound in terms of heterogeneous waveforms provides several unique ways in which to perform transformations [Kling and Roads, 2004]. Atoms selected from the dictionary by a DBM can be modified by varying their parameters, independently or in groups. In the latter case, groups can be determined either by molecular clustering algorithms as presented in Section 3.1, or through variety of parameteric-based selection or filtering processes, for instance, all atoms longer than 100 ms.

Due to the non-uniqueness inherent to (1) when using overcomplete dictionaries and minimally defined constraints, some decompositions may provide more malleability than others, or may be well-suited to a particular class of transformations at the expense of others. Because of this, it is often beneficial to experiment with different dictionaries for a particular signal.

Though the number of atoms in a decomposition can be in the hundred-thousands, the limiting factor in real-time synthesis is waveform density. However, since synthesis is only a process of table-lookup, most transformations can be performed and synthesized in realtime. This is important for providing immediate feedback to a user and for allowing the possibility using decompositions for musical performance.

Below, we outline and describe four classes of sound transformation using representations built by DBMs: filtering, parametric manipulations, substitution, and granular transformations. Furthermore, we have created an application, *Scatter*, that provides an interface to analyzing, visualizing, and working with atomic decompositions produced by pursuits. An example screen-shot is shown in Fig. 14. Audio examples and a video demonstration of the interface can be found on-line at `http://www.mat.ucsb.edu/~b.sturm/ICMC2008/`.

5.1 Filtering

Each atom in a decomposition is described by a set of parameters, each dependent on the atom type. All atoms have at least a defined scale s, translation u, and amplitude. Other waveforms have parameters specific to their atom type, for example, those of Gabor atoms in (11) and (12) include modulation frequency ω and phase ϕ. All of these parameters can be used as guides for filtering atoms from a decomposition.

5.1.1 Frequency Filtering

Selection and attenuation in a decomposition based on modulation frequency allow transformations that are analogous to traditional frequency-based filtering, such as low-pass, high-pass, bandpass and notch filtering. This transformation only works when the atom itself has a frequency parameter, i.e., not an impulse. Though not unique, this class of

filtering is useful in conjunction with other filter classes.

5.1.2 Amplitude Filtering

Filtering based on the weights $\mathbf{a}(n)$ in (6) selects atoms having energies within specified ranges. Since the energy of each atom selected by MP grows smaller as the approximation order increases [Mallat and Zhang, 1993], keeping those atoms that have energies above a given threshold will result in a decomposition that is equivalent to a low-order approximation of the signal. This results in a decomposition which retains the loudest energy components of the signal, which for a harmonic sound, correspond to the lower partials. The opposite, more exotic, effect is achieved if only the low-energy atoms are retained or amplified.

5.1.3 Scale Filtering

Multiresolution atomic decompositions provide the ability to filter based on the scale, or duration, of atoms. One can also achieve this using a wavelet decomposition, except that scale and frequency are inversely related: specifying large-scale wavelets also selecting those of low frequency, and vice versa. DBMs have the capacity to make scale and frequency independent parameters in the model. For example, one may synthesize the transients or tonals of a signal by using only the shortest or longest atoms, respectively. Such an approach using a time-frequency dictionary works best on signals that have very distinct separations between such components, as seen in Fig. 13.

5.1.4 Morphological Filtering

Using larger structures, such as the molecules discussed in Section 3.1, a decomposition can be filtered based on the morphologies to which atoms contribute. For instance, we can target distinct groups of atoms that model the harmonics, such as those seen in Fig. 6. Atoms may even be found that are specific to instrumental morphologies, such as features particular to a piano [Leveau et al., 2008]. An attractive element of using a heterogeneous atomic representation is the ease with which the transient and tonal portion of a signal may be separated and modified independently, as demonstrated in Fig. 13.

5.2 Parametric Manipulations

Another class of transformations also uses the ability to select atoms based on parameters, but instead of removal, attenuation, or amplification, the parameters are modified.

5.2.1 Pitch-shifting

Transposing an audio signal in frequency without affecting its temporal characteristics can be done using dictionary atoms that can be associated with pitch, such as Gabor atoms. A naïve approach alters the modulation frequency of each atom [Sturm et al., 2006]. For instance, a doubling of the modulation frequencies can change the pitch of the resynthesis by an octave. Modifying the frequencies without accounting for the phase of each atom, however, results in pre-echo and artifacts that sound like tremolo [Sturm et al., 2006]. Since the time-frequency relationships between the atoms are in a delicate balance, as described in Section 3.2, one must pay careful attention to the phase relationships between atoms and adjust accordingly to preserve the envelope of the decomposed waveform. Still, the naïve approach works remarkably well for signals with soft transients, such as a flute. One can combine this approach with morphological filtering such that only the tonal content of a musical signal is transposed while the transient content is preserved.

5.2.2 Time-scaling

One can alter the duration of a signal without affecting its frequency by changing the scale of every atom and adjusting the translations appropriately. This approach still suffers from not accounting for the interactions between atoms. The results do not sound as natural as a simple phase vocoder method, but they are unique. In the case of a drum sample stretched by a factor of four, a cymbal crash maintains its cymbal qualities, but transients begin to sound like "damped chimes." A less naïve approach shifts the atoms in time and fills in the resulting gaps with additional atoms having parameters interpolated between their neighbors. As in pitch shifting, there also exists the possibility of modifying only those atoms with morphologies that make sense to scale in time, such as tonals as opposed to transients. In this case, time scaling would not be done to atoms in transient morphologies.

5.2.3 Stochastic Transformations

Unique transformations can be created by modifying parameters according to stochastic models or probability distribution functions. Stochastically offsetting atom translation or frequency creates a time-frequency jitter while changing atom duration, and preserving the atom center-times, creates a temporal bleed or pointalistic transformation. A stochastically scrambled transformation in the time-frequency plane is made by combining these other stochastic transformations.

5.3 Substitution

Given a signal decomposition that uses one dictionary, we may replace any or all of those atoms with new ones. This technique has been explored using wavelets, but with varying degrees of success [Roads, 2001]. Through DBMs the results can sound smooth and lack sharp discontinuities, i.e., missing the distortions that often appear when substituting one wavelet type for another. Replacing the entire dictionary used for the analysis with a different one for synthesis can produce dramatic effects. For example, replacing the Gabor atoms in a dictionary used in a decomposition of speech with one containing only damped sinusoidal atoms produces something that sounds like "speaking chimes." Replacing damped sinusoidal atoms with Gabor atoms creates smoothed transients, and a transformation similar to a "reverse echo."

5.4 Granular Transformations

The rich history of granular methods [Gabor, 1946, Xenakis, 1971, Roads, 2001] gives a plentiful number of sound manipulations. After decomposing a signal into atoms, one can manipulate atom density to create unique transformations such as evaporation (sonic disintegration), coalescence (sonic formation), and mutation (sonic metamorphosis). Sonic disintegration, or evaporation, is created by stochastically reducing atom density as a function of time. The reverse process, e.g., stochastically increasing atom density as a function of time, produces sonic coalescence. A sonic mutation is realized by simultaneously disintegrating one signal and coalescing another signal.

5.4.1 Granular Spatialization

It is also possible spatialize atoms individually [McLeran et al., 2008]. The spatially transformed sound retains its coherence (identity) with respect to the original, but with different atoms projected from different spatial locations. We also can spatially transform the signal in consort with the other manipulations above, for instance based on morphological filtering (transients, harmonics, loud atoms, short atoms, etc.).

6 Conclusion

We have presented an overview of DBMs as well our research, which contributes to exploring the application of DBMs to analyzing, visualizing, and transforming audio and music signals in novel ways. We have shown how a sparse approximation method, such as OOMP, can produce a parsimonious parametric signal model that can be useful in analysis. By superposing the WVD of each atom in a model, we can obtain a multiresolution time-frequency domain image of the signal model. Finally, using a dictionary of parametric atoms provides ways of altering the content embodied in the model.

One of the most attractive features of DBMs is the flexibility of specifying how a signal is decomposed, and the set of functions over which it is decomposed. However, this freedom imposes tradeoffs, including (1) increased computation, (2) non-unique solutions, and (3) non-trivial interactions between atoms that can produce artifacts, which diminish the efficiency and meaningfulness the results. While (1) and (2) are important for applications like real-time communications, they are not especially critical to musical applications. The problem of artifacts (3) is more critical to musical signal processing, as it directly correlates with audio quality. We have addressed this problem by proposing and studying an alternative atom selection criterion that takes into consideration how the signal has been modeled by the pursuit and dictionary. We show that the resulting models can provide a more efficient and meaningful representation than those obtained using the selection criterion of MP, OMP, or OOMP. Going further, we have demonstrated how the atoms of a decomposition can be combined into molecules that represent larger morphological structures, such as

harmonics. These combinations clarify the correspondence between individual atoms in the model to content in the signal, and can further aid the application of sparse approximation to sound and music analysis, visualization, and transformation.

7 Acknowledgments

Thank you to Garry Kling for providing Fig. 11. This work was supported in part by the National Science Foundation under Grant CCF-0729229, as well as a Bourse Chateaubriand Fellowship administered by le Ministere Affaires Étrangeres de France, N. 634146B.

References

[Abdallah and Plumbley, 2006] Abdallah, S. A. and Plumbley, M. D. (2006). Unsupervised analysis of polyphonic music by sparse coding. *IEEE Trans. Neural Networks*, 17(1):179–196.

[Aharon et al., 2006] Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.*, 54(11):4311–4322.

[Boyd and Vandenberghe, 2004] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge, UK.

[Candès and Wakin, 2008] Candès, E. and Wakin, M. B. (2008). An introduction to compressive sampling. *IEEE Signal Process. Mag.*, 25(2):21–30.

[Chen et al., 1998] Chen, S. S., Donoho, D. L., and Saunders, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61.

[Christensen and Jensen, 2007] Christensen, M. G. and Jensen, S. H. (2007). The cyclic matching pursuit and its application to audio modeling and coding. In *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA.

[Christensen and van de Par, 2006] Christensen, M. G. and van de Par, S. (2006). Efficient parametric coding of transients. *IEEE Trans. Audio, Speech, Lang. Process.*, 14(4):1340–1351.

[Daudet, 2006] Daudet, L. (2006). Sparse and structured decompositions of signals with the molecular matching pursuit. *IEEE Trans. Audio, Speech, Lang. Process.*, 14(5):1808–1816.

[Davis et al., 1997] Davis, G., Mallat, S., and Avellaneda, M. (1997). Adaptive greedy approximations. *J. Constr. Approx.*, 13(1):57–98.

[Derrien, 2006] Derrien, O. (2006). Multi-scale frame-based analysis of audio signals for musical transcription using a dictionary of chromatic waveforms. In *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, volume 5, pages 57–60, Toulouse, France.

[Donoho and Huo, 2001] Donoho, D. L. and Huo, X. (2001). Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory*, 47(7):2845–2862.

[Elad and Aharon, 2006] Elad, M. and Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.*, 15(12):3736–3745.

[Figueras i Ventura et al., 2006] Figueras i Ventura, R. M., Vandergheynst, P., and Frossard, P. (2006). Low-rate and flexible image coding with redundant representations. *IEEE Trans. Image Process.*, 15(3):726–739.

[Gabor, 1946] Gabor, D. (1946). Theory of communication. *J. IEE*, 93(3):429–457.

[Gabor, 1947] Gabor, D. (1947). Acoustical quanta and the theory of hearing. *Nature*, 159(4044):591–594.

[Gersho and Gray, 1991] Gersho, A. and Gray, R. M. (1991). *Vector Quantization and Signal Compression*. Kluwer Academic, Norwell, MA.

[Goodwin and Vetterli, 1999] Goodwin, M. M. and Vetterli, M. (1999). Matching pursuit and atomic signal models based on recursive filter banks. *IEEE Trans. Signal Process.*, 47(7):1890–1902.

[Gribonval, 2001] Gribonval, R. (2001). Fast matching pursuit with a multiscale dictionary of gaussian chirps. *IEEE Trans. Signal Process.*, 49(5):994–1001.

[Gribonval and Bacry, 2003] Gribonval, R. and Bacry, E. (2003). Harmonic decompositions of audio signals with matching pursuit. *IEEE Trans. Signal Process.*, 51(1):101–111.

[Gribonval et al., 1996] Gribonval, R., Bacry, E., Mallat, S., Depalle, P., and Rodet, X. (1996). Analysis of sound signals with high resolution matching pursuit. In *Proc. IEEE-SP Int. Symp. Time-Freq. Time-Scale Anal.*, pages 125–128, Paris, France.

[Heusdens et al., 2002] Heusdens, R., Vafin, R., and Kleijn, W. B. (2002). Sinusoidal modeling using psychoacoustic-adaptive matching pursuits. *IEEE Signal Process. Lett.*, 9(8):262–265.

[Jaggi et al., 1998] Jaggi, S., Karl, W. C., Mallat, S., and Willsky, A. S. (1998). High resolution pursuit for feature extraction. *Applied and Computational Harmonic Analysis*, 5(4):428–449.

[Jost et al., 2006] Jost, P., Vandergheynst, P., Lesage, S., and Gribonval, R. (2006). MoTIF: an efficient algorithm for learning translation invariant dictionaries. In *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, volume 5, pages 857–860, Toulouse, France.

[Kling and Roads, 2004] Kling, G. and Roads, C. (2004). Audio analysis, visualization, and transformation with the matching pursuit algorithm. In *Proc. Int. Conf. Digital Audio Effects*, pages 33–37, Naples, Italy.

[Kreutz-Delgado et al., 2003] Kreutz-Delgado, K., Murray, J. F., Rao, B. D., Engan, K., Lee, T., and Sejnowski, T. J. (2003). Dictionary learning algorithms for sparse representation. *Neural Computation*, 15(2):349–396.

[Krstulovic and Gribonval, 2006] Krstulovic, S. and Gribonval, R. (2006). MPTK: Matching pursuit made tractable. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, volume 3, pages 496–499, Toulouse, France.

[Lesage et al., 2006] Lesage, S., Krstulovic, S., and Gribonval, R. (2006). Underdetermined source separation: Comparison of two approaches based on sparse decompositions. In *Proc. Int. Conf. Independent Component Analysis Blind Source Separation*, pages 633–640, Charleston, South Carolina.

[Leveau et al., 2008] Leveau, P., Vincent, E., Richard, G., and Daudet, L. (2008). Instrument-specific harmonic atoms for mid-level music representation. *IEEE Trans. Audio, Speech, Lang. Process.*, 16(1):116–128.

[Lewicki, 2002] Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature Neuroscience*, 5(4):356–363.

[Lewicki and Sejnowski, 2000] Lewicki, M. S. and Sejnowski, T. J. (2000). Learning overcomplete representations. *Neural Computation*, 12:337–365.

[Mallat, 1999] Mallat, S. (1999). *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, CA, 2nd edition.

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

[McAulay and Quatieri, 1986] McAulay, J. and Quatieri, T. (1986). Speech analysis/synthesis based on a sinusoidal representation. *IEEE Trans. Acoustics, Speech, Signal Process.*, 34(4):744–754.

[McLeran et al., 2008] McLeran, A., Roads, C., Sturm, B. L., and Shynk, J. J. (2008). Granular methods of sound spatialization using overcomplete representations. In *Sound and Music Computing Conference*, Berlin, Germany.

[Meyer, 2001] Meyer, C. (2001). *Matrix Analysis and Applied Linear Algebra*. SIAM: Society for Industrial and Applied Mathematics, Philadelphia, PA.

[Pati et al., 1993] Pati, Y., Rezaiifar, R., and Krishnaprasad, P. (1993). Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proc. Asilomar Conf. Signals, Syst., Comput.*, volume 1, pages 40–44, Pacific Grove, CA.

[Preis and Georgopoulos, 1999] Preis, D. and Georgopoulos, V. C. (1999). Wigner distribution representation and analysis of audio signals: An illustrated tutorial review. *J. Audio Eng. Soc.*, 47(12):1043–1053.

[Ravelli et al., 2008] Ravelli, E., Richard, G., and Daudet, L. (2008). Union of MDCT bases for audio coding. *IEEE Trans. Audio, Speech, Lang. Proc.*, 16(8):1361–1372.

[Rebollo-Neira and Lowe, 2002] Rebollo-Neira, L. and Lowe, D. (2002). Optimized orthogonal matching pursuit approach. *IEEE Signal Process. Lett.*, 9(4):137–140.

[Roads, 2001] Roads, C. (2001). *Microsound*. MIT Press, Cambridge, MA.

[Roads, 2004] Roads, C. (2004). Pictor Alpha. In *Point, Line, Cloud*, compact disc and digital video disc. Asphodel Records.

[Sturm, 2009] Sturm, B. L. (2009). *Sparse Approximation and Atomic Decomposition: Considering Atom Interactions in Evaluating and Building Signal Representations*. PhD thesis, University of California, Santa Barbara, CA.

[Sturm et al., 2006] Sturm, B. L., Daudet, L., and Roads, C. (2006). Pitch-shifting audio signals using sparse atomic approximations. In *Proc. ACM Workshop Audio Music Comput. Multimedia*, pages 45–52, Santa Barbara, CA.

[Sturm and Shynk, 2008] Sturm, B. L. and Shynk, J. J. (2008). Sparse approximation and the pursuit of meaningful signal models. *IEEE Trans. Acoustics, Speech, Signal Process. (submited)*.

[Sturm et al., 2007] Sturm, B. L., Shynk, J. J., and Daudet, L. (2007). A short-term measure of dark energy in sparse atomic estimations. In *Proc. Asilomar Conf. Signals, Syst., Comput.*, pages 1126–1129, Pacific Grove, CA.

[Sturm et al., 2008a] Sturm, B. L., Shynk, J. J., and Daudet, L. (2008a). Measuring interference in sparse atomic estimations. In *Proc. Conf. Info. Sciences Syst.*, pages 961–966, Princeton, NJ.

[Sturm et al., 2008b] Sturm, B. L., Shynk, J. J., Daudet, L., and Roads, C. (2008b). Dark energy in sparse atomic estimations. *IEEE Trans. Audio, Speech, Lang. Process.*, 16(3):671–676.

[Sturm et al., 2008c] Sturm, B. L., Shynk, J. J., and Gauglitz, S. (2008c). Agglomerative clustering in sparse atomic decompositions of audio signals. In *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, pages 97–100, Las Vegas, NV.

[Sturm et al., 2008d] Sturm, B. L., Shynk, J. J., McLeran, A., Roads, C., and Daudet, L. (2008d). A comparison of molecular approaches for generating sparse and structured multiresolution representations of audio and music signals. In *Proc. Acoustics*, pages 5775–5780, Paris, France.

[Xenakis, 1971] Xenakis, I. (1971). *Formalized Music*. Indiana University Press, Bloomington, Indiana.

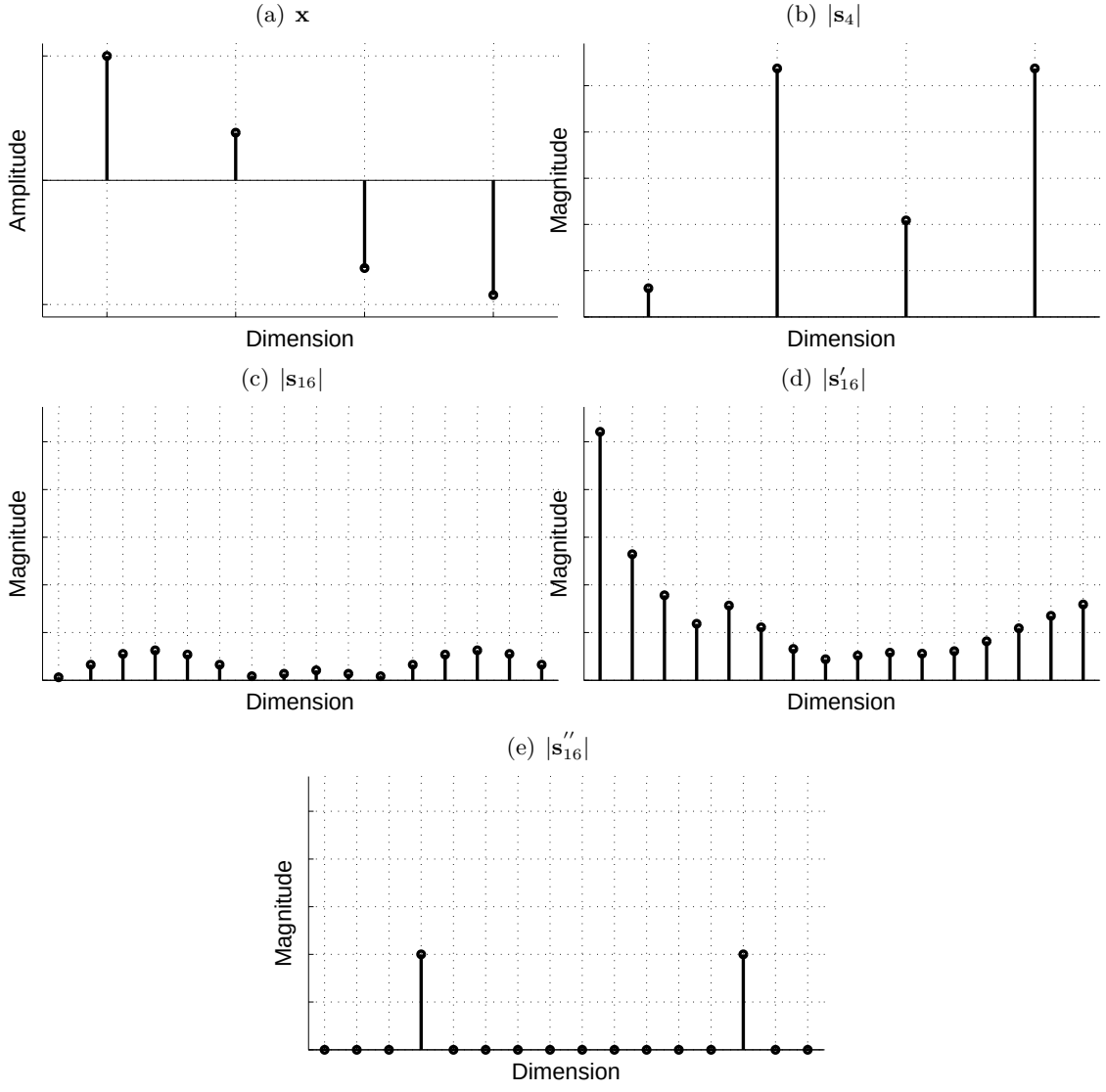

Figure 1: Vector $\mathbf{x}$ (a) represented by several solutions in different dictionaries. (b) Length-4 discrete Fourier transform. (c) 4-times zero-padded Fourier dictionary. (d) Another valid solution. (e) The most efficient solution.

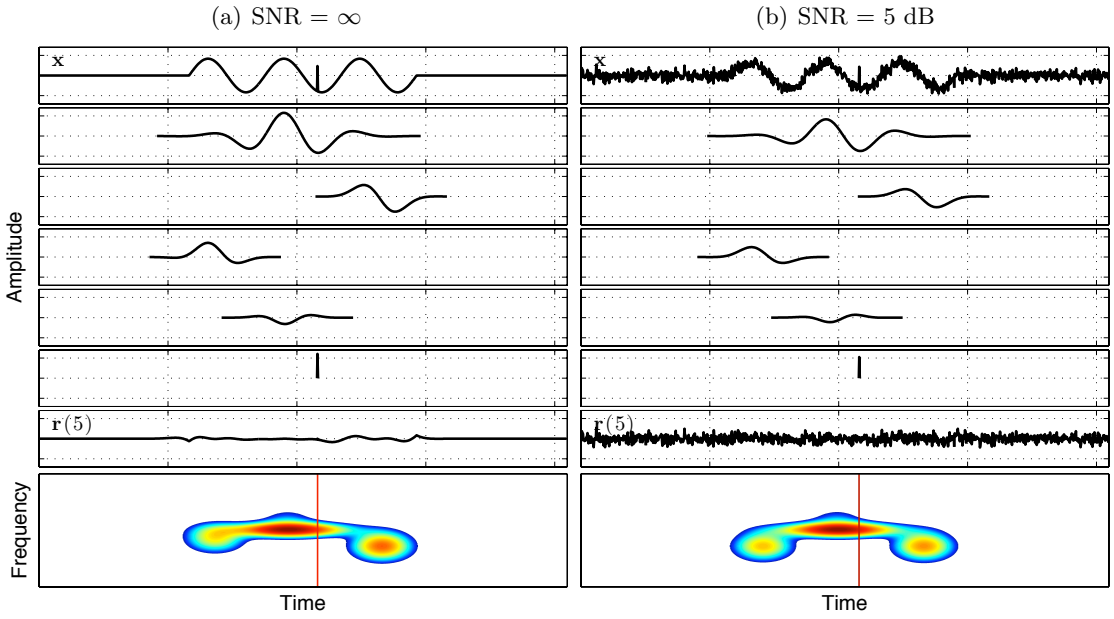

Figure 2: MP decomposition of a simple signal with and without noise over a Gabor and Dirac dictionary. Signal is at top; from top to bottom are the first five atoms found by MP in order. The fifth-order residual is labeled $\mathbf{r}(5)$. At bottom is a superposition of the Wigner-Ville distribution of each atom.

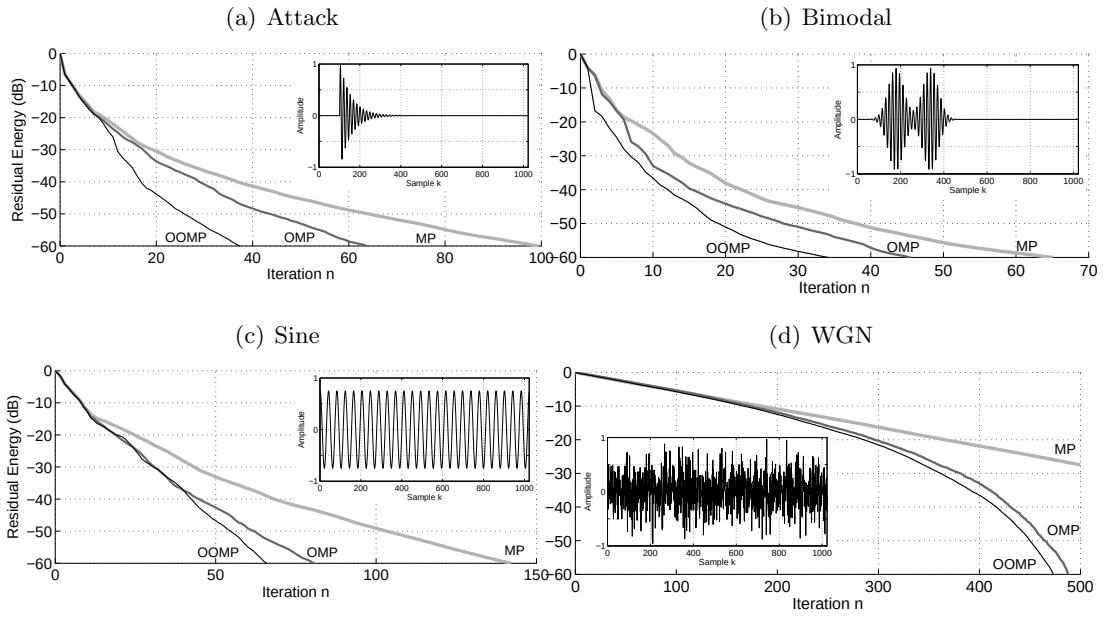

Figure 3: Residual energy decay as a function of iteration for four example signals (inset). Dictionary is a union of Gabor atoms and Dirac spikes. (Note change in abscissa in (d).)

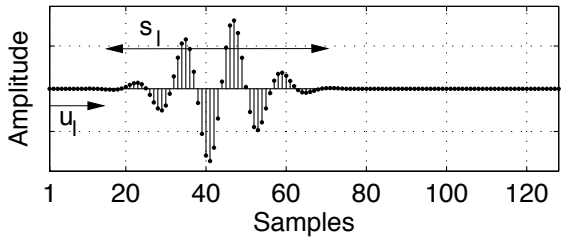

Figure 4: Real Gabor atom ($K = 128$) with translation $u_l = 15$, scale $s_l = 64$, modulation frequency ω_l, and phase ϕ_l.

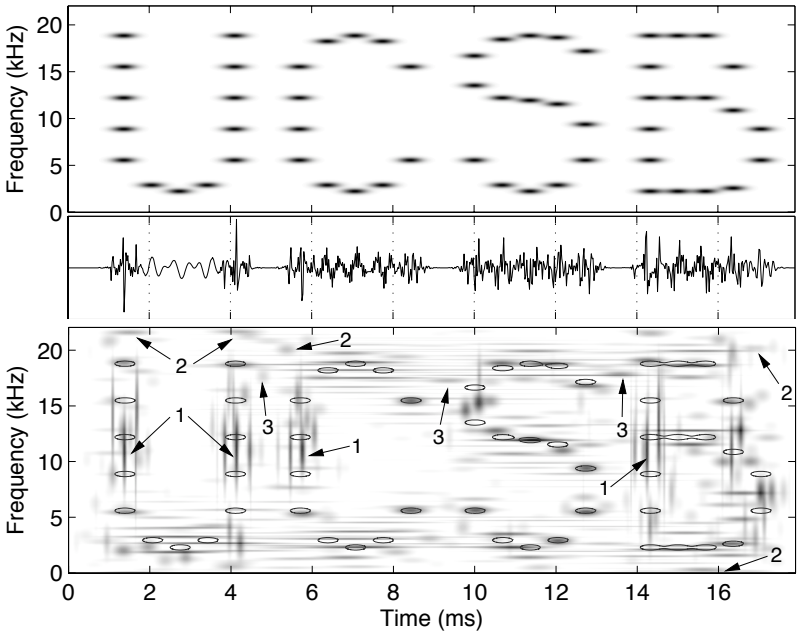

Figure 5: Signal (middle) built from 57 Gabor atoms seen in wivigram (top). Wivigram of decomposition (bottom) with outlines of atoms in top wivigram.

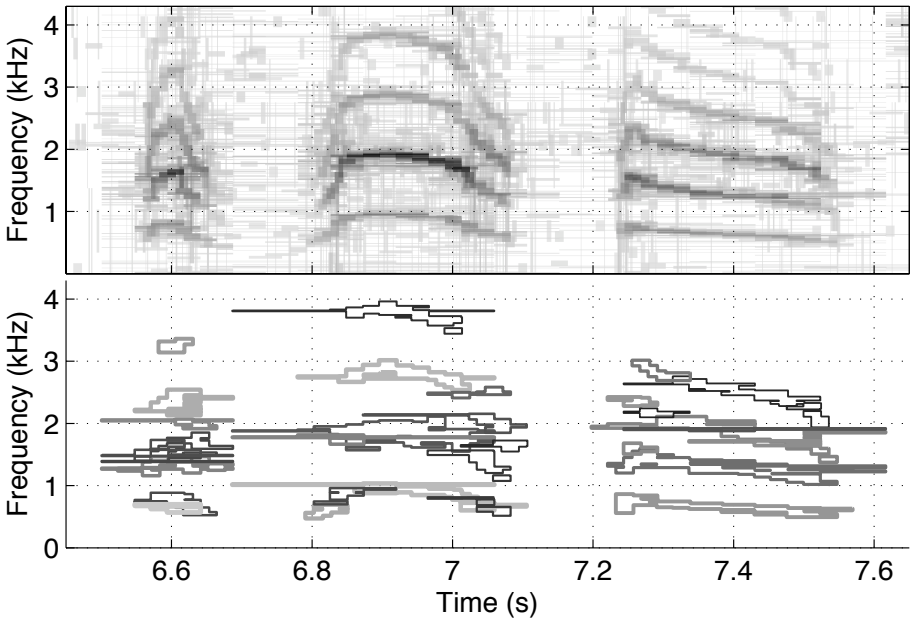

Figure 6: Wivigram (top) of a bird signal. Several outlined molecules (bottom).

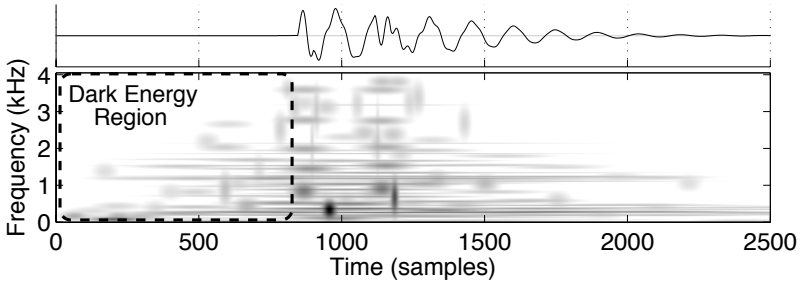

Figure 7: Wivigram (bottom) shows terms found by MP decomposition of signal (top) at times of no energy. These atoms cancel in synthesis and thus disappear through destructive interference.

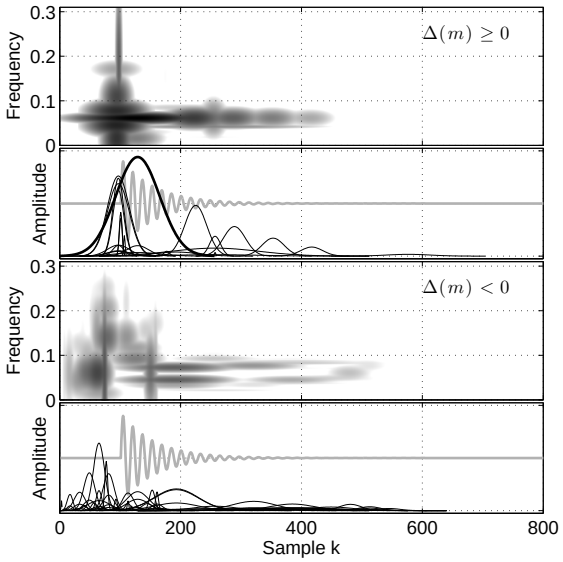

Figure 8: A model of Attack (seen in the inset to Fig. 3(a)) split into two sets by sign of interference $\Delta(m)$, shown with wivigrams and atom envelopes.

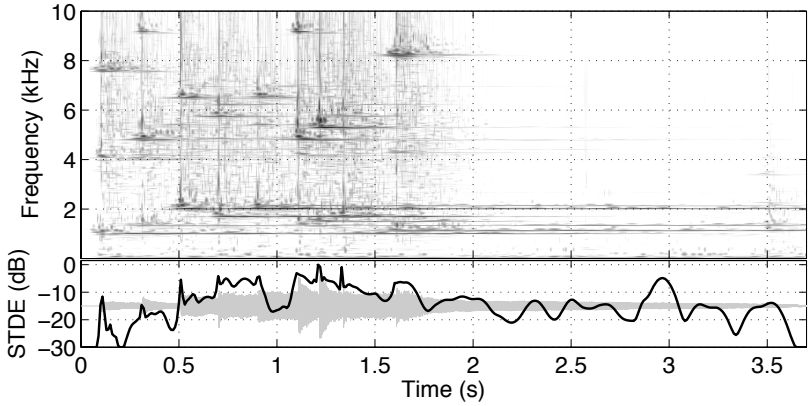

Figure 9: Wivigram (top) of Glockenspiel signal. Short-term dark energy overlaid on signal waveform (bottom).

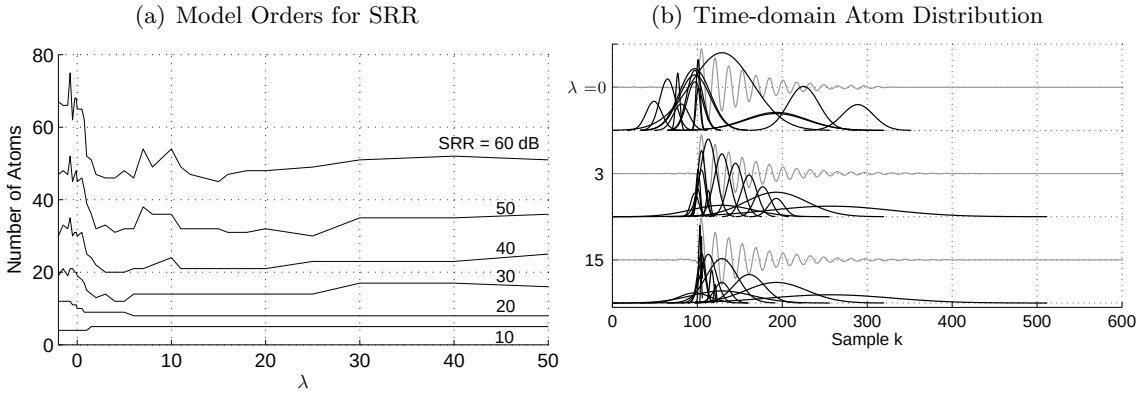

Figure 10: For Attack (seen in the inset to Fig. 3(a)): (a) Model order at specific SRR as a function of $\lambda(n) = \lambda$ modeled by OMP with interference adaptation. (b) Time-domain distribution of atoms in models created with specific λ.

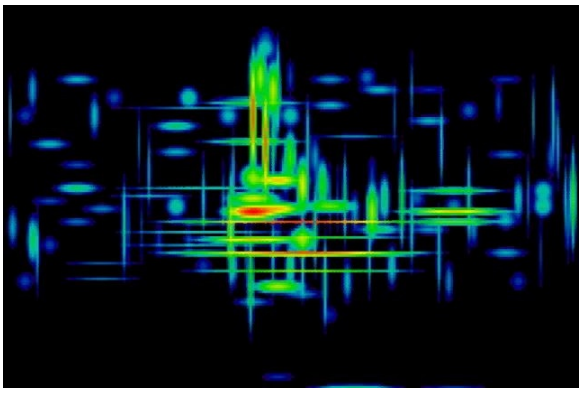

Figure 11: Wivigram visualization of a segment of electroacoustic work *Concrete PH* by Iannis Xenakis.

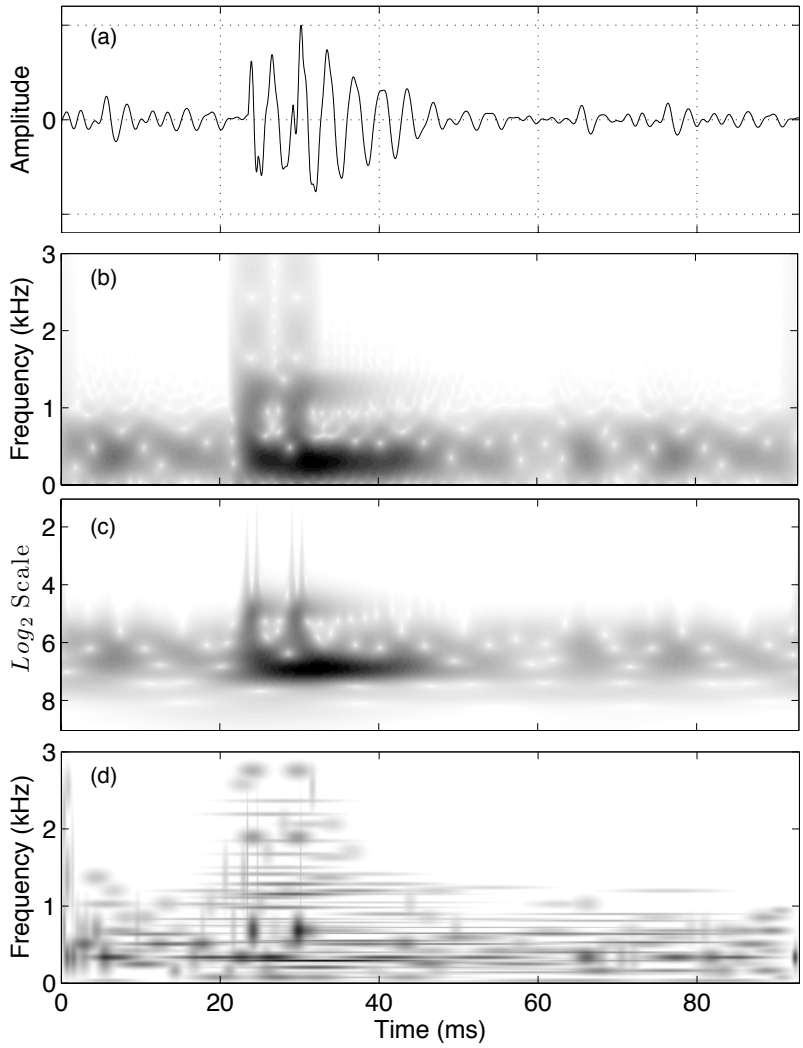

Figure 12: (a) Short musical signal. (b) Spectrogram from STFT. (c) Scalogram from DWT using Gabor wavelet. (d) Wivigram from MP using Gabor dictionary.

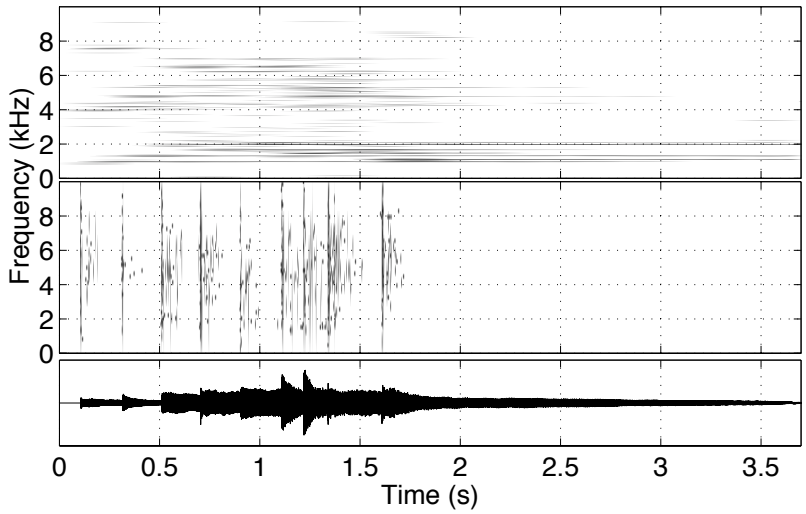

Figure 13: Wivigrams showing long (top) and short atoms (middle) in a musical signal (bottom).

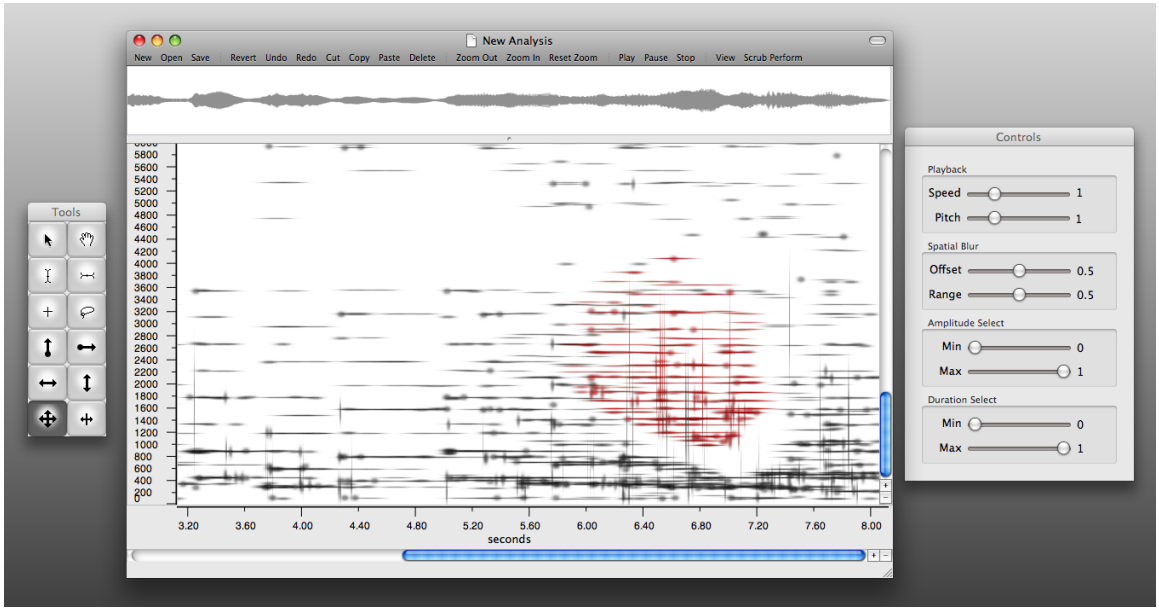

Figure 14: Screenshot of an interface for working with OM decompositions.