ANALYSIS, VISUALIZATION, AND TRANSFORMATION OF AUDIO SIGNALS USING DICTIONARY-BASED METHODS

Bob L. Sturm[*], *Curtis Roads*[†], *Aaron McLeran*[†], *and John J. Shynk*[*]
Department of Electrical and Computer Engineering[*]
Media Arts and Technology Program[†]
University of California
Santa Barbara, CA 93106

ABSTRACT

This article provides an overview of dictionary-based methods (DBMs), and reviews recent work in the application of such methods to working with audio and music signals. As Fourier analysis is to additive synthesis, DBMs can be seen as the analytical counterpart to a generalized granular synthesis, where a sound is built by combining heterogeneous *atoms* selected from a user-defined *dictionary*. As such, DBMs provide novel ways for analyzing and visualizing audio signals, creating multiresolution descriptions of their contents, and designing sound transformations unique to a description of audio in terms of atoms.

1. INTRODUCTION

The development of dictionary-based methods (DBMs) — also called sparse approximation — has been motivated by the desire to represent signals in ways that are more sparse, efficient, robust to noise, meaningful, and malleable than can be obtained using standard transform methods [1, 2]. DBMs attempt to adapt a representation to a signal, and give a user the freedom to define the set of functions over which a decomposition is performed. These properties provide many benefits over other methods of time-frequency or time-scale decomposition, such as adaptability to specific signals. When the functions used are *atomic*, i.e., localized in time, the end result of the decomposition embodies a "score" to reproduce the given sound with atoms [3]. In this sense, DBMs can be seen as the analytical equivalent to granular synthesis [4, 5], but their application is much wider than this. Researchers have applied DBMs to the and coding and compression of audio [6, 7] and image [8] data; data denoising and recovery [9, 10], blind source separation [11]; musical analysis and transcription [12, 13], etc. A recent development is compressive sampling [14], where the use of DBMs for *compressible signals* allows one to sample at rates much lower than required by the Nyquist-Shannon sampling theorem. In our work, we explore the use of DBMs to provide a rich and flexible interface to the *content* in an audio signal (e.g., transients, harmonics, notes).

After providing an overview of DBMs, we review some of our recent work and results in this area, specifically in the analysis, visualization, and transformation of audio signals, significantly extending the results presented in [15]. Throughout we discuss the theoretical and practical benefits of DBMs, as well as some of their problems, such as *uniqueness* and *bias*. We show how an atomic decomposition can lead to a sparse and structured representation of a musical signal, providing new methods of visualization. Furthermore, through the molecules described in section 3.1, these representations can ultimately be used to provide an interface to the contents of the signal at many 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, ^{T} denotes transpose, and $||\cdot||$ denotes an ℓ_2-norm.

2. OVERVIEW OF D-B 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 columns in a *dictionary* $\mathbf{D}_{K\times N}$. More formally, we want to find a solution to the following problem:

$$\arg\min_{\mathbf{s}} f\left(C(\mathbf{s}), D(\mathbf{x}, \mathbf{Ds})\right) \text{ such that } \mathbf{x} = \mathbf{Ds} \qquad (1)$$

where $C(\mathbf{s})$ is a cost function, $D(\mathbf{x}, \mathbf{Ds})$ is a distortion function, and $\mathbf{s}$ is a column vector of N weights. If $N = K$ and $\mathbf{D}^H$ is the orthonormal discrete Fourier transform matrix, then $\mathbf{s}$ is just the discrete Fourier transform of $\mathbf{x}$. In DBMs, however, $N \gg K$ and $\text{rank}(\mathbf{D}) = K$, which is the meaning of the term *overcomplete*. Thus, solving (1) is more complex than using orthogonal least-squares projection. For any real $\mathbf{x}$, there could exist an infinity of solutions $\mathbf{s}$; and none will be unique unless $C(\mathbf{s})$ and $D(\mathbf{x}, \mathbf{Ds})$ are well-defined.

A constraint on sparsity is requiring that the number of atoms selected from $\mathbf{D}$ be minimized, i.e., the number of nonzero elements in $\mathbf{s}$ is minimized. This, however, makes finding the best $\mathbf{s}$ unsolvable in a reasonable amount of time [16], since it requires checking all possible linear combinations of atoms. One might instead require that the ℓ_1-norm of $\mathbf{s}$, i.e., the sum of the magnitudes in $\mathbf{s}$, be minimized. This constraint is specified in basis pursuit (BP) [2], which, while providing a solvable problem and guaranteeing a certain amount of sparsity in the solution, requires a linear program to solve. Another set of approaches use a gradient descent approach to find solutions, e.g., matching pursuit (MP) [1, 17]. In these methods, the procedure involves minimizing an intermediate distortion at each step. These methods are straight-forward and simple to implement [18], but often their results can be heavily biased. To find one atom in MP requires on the order of a fast Fourier transform of the entire signal [1, 18].

2.1. Matching Pursuit Algorithm

MP [1] decomposes signals using a gradient descent approach. At step $n+1$, the column in $\mathbf{D}$ that has the largest magnitude inner product with the nth residual signal is selected, as in the following rule:

$$\mathbf{g}_n = \arg\max_{\mathbf{d}\in\mathbf{D}} \big|\mathbf{d}^T\mathbf{r}(n)\big|/||\mathbf{d}|| \tag{2}$$

where $\mathbf{r}(n) = \mathbf{x} - \widetilde{\mathbf{x}}(n)$ is the nth-order residual signal ($\mathbf{r}(0) \equiv \mathbf{x}$), and $\widetilde{\mathbf{x}}(n)$ is the nth-order approximation waveform ($\widetilde{\mathbf{x}}(0) \equiv \mathbf{0}$). This selects the dictionary waveform that is most correlated with the current residual signal. (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.) The weight of $\mathbf{g}_n$ is then calculated by

$$a_n = \mathbf{g}_n^T\mathbf{r}(n)/||\mathbf{g}_n|| \tag{3}$$

and MP produces the new residual signal $\mathbf{r}(n+1) = \mathbf{r}(n) - a_n\mathbf{g}_n$. The above process is repeated until the residual energy is lower than some limit, or a specified estimation order has been reached. Orthogonal MP (OMP) [17] performs the additional step of orthogonalizing the residual for all selected atoms, which in effect recomputes every weight. After n iterations of MP, we have a nth-order approximation of $\mathbf{x}$

$$\widetilde{\mathbf{x}}(n) = [\mathbf{g}_0|\mathbf{g}_1|\cdots|\mathbf{g}_{n-1}]\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1}\end{bmatrix} = \mathbf{G}(n)\mathbf{a}(n). \tag{4}$$

It should be noted that the signal is not windowed in this process, which avoids an arbitrary segmentation.

The simple selection criterion given in (2) can be generalized to weighted MP

$$\mathbf{g}_n = \arg\max_{\mathbf{d}\in\mathbf{D}} \big|\mathbf{d}^T\mathbf{W}\mathbf{r}(n)\big| \tag{5}$$

where $\mathbf{W}$ is a weighting matrix that can be, for example, based on perceptually significant measures [19]. One can also specify a selection criterion based upon measures other than an inner product [20].

As long as the dictionary is at least complete, the solution s is convergent [1], i.e., $\lim_{n\to\infty}\widetilde{\mathbf{x}}(n) = \mathbf{x}$. Although both BP and OMP guarantee convergence in a finite number of steps $\leq K$ [2, 17], a solution found using MP usually requires an infinite number of iterations to converge. Depending on the application, however, an exact or sparse solution may be less important that acquiring a useful and meaningful representation of $\mathbf{x}$.

2.2. Overcomplete Dictionaries

A major advantage of DBMs is the flexibility to choose the contents of the dictionary. This gives a user the ability to make a decomposition adaptable to specific structures in a signal. Unlike in Fourier or wavelet analysis, the dictionary can be any collection of waveforms without restrictions. However, with this freedom comes a higher computational complexity, a lack of uniqueness, and the manifestation of artifacts from aspects of the decomposition process.

A common approach to creating a dictionary is by combining families of discretized, scaled, translated, and modulated lowpass functions $h(k;s)$. An simple example of a real dictionary waveform is

$$g(k) = Ah(k-u;s)\cos\big(k\omega + \phi\big) \tag{6}$$

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 [21], can also be created. The scalar A is set for an atom such that $\sum |g(k)|^2 = 1$.

The shape of each waveform is specified by $h(k;s)$, which can be likened to a window. For instance, a *Gabor atom* consists of a translated discrete Gaussian function:

$$h(k;s) = \begin{cases} \exp\big(-\frac{(k-s/2)^2}{2(\alpha s)^2}\big), & k = 0, 1, \ldots, s-1 \\ 0, & \text{else} \end{cases} \tag{7}$$

where α controls the variance, and s is the scale. An example Gabor atom is shown in Fig. 1. To create a *Gabor dictionary*, also called a *time-frequency dictionary* [1], each column of $\mathbf{D}$ is created by evaluating the functions in (6) and (7) at a number of different scales, translations, and modulations.

The dictionary shown in Table 1 consists of several atoms of different shapes $h(k;s)$ and scales s. The values Δ_u and Δ_ω specify increments for the translation and modulation frequency, respectively, such that for a signal of length K, an atom of scale s can be translated to $u = i\Delta_u$ for $0 \leq i < (K - \frac{s}{2})/\Delta_u$, and modulated to $\omega = j\Delta_\omega$ for $0 \leq j \leq \pi/\Delta_\omega$. This dictionary includes a family of Hann-windowed harmonic atoms [22], a real sampled waveform (e.g., a training sequence), and waveforms that are learned for particular classes of signals [13].

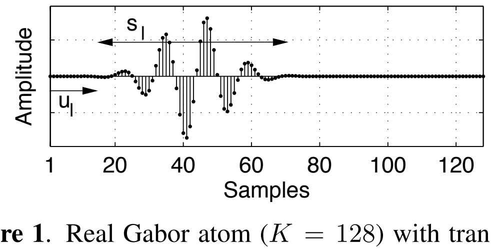

Figure 1. Real Gabor atom ($K = 128$) with translation u_l, scale s_l, modulation frequency ω_l, and phase ϕ_l.

$h(k;s)$	s	Δ_u	Δ_ω	details
Dirac	1	1	-	-
Rectangle	4	1	$\pi/2$	-
Rectangle	8	4	$\pi/4$	-
Gaussian	64	16	$\pi/64$	$\alpha = 0.1$
Gaussian	128	32	$\pi/128$	$\alpha = 0.1$
Gaussian	128	32	$\pi/128$	$\alpha = 0.2$
Hann	1024	128	$\pi/2048$	Harmonic
Hann	1024	128	$\pi/2048$	chirp = 0.1
Sampled	12,452	500	-	-
Learned	2048	1024	-	-

Table 1. Example Dictionary

2.3. Example Decomposition

Consider the simple signal shown in Fig. 2, which has three periods of a 300 Hz sinusoid. Decomposing this signal using MP with a Gabor dictionary produces a good approximation using the first three atoms as shown. Table 2 shows the *book* of this decomposition, which provides details about each of the atoms. The first atom selected $\mathbf{g}_0$ has a modulation frequency very close to the original signal, as well as a phase that is close to $-\pi/2$. It provides a very good first-order representation of the signal. The next two atoms selected have a much smaller amplitude than the first, and serve to correct the errors created at the edges of the original signal by destructively interfering with the tails of $\mathbf{g}_0$. This phenomenon is discussed further in Section 3.2.

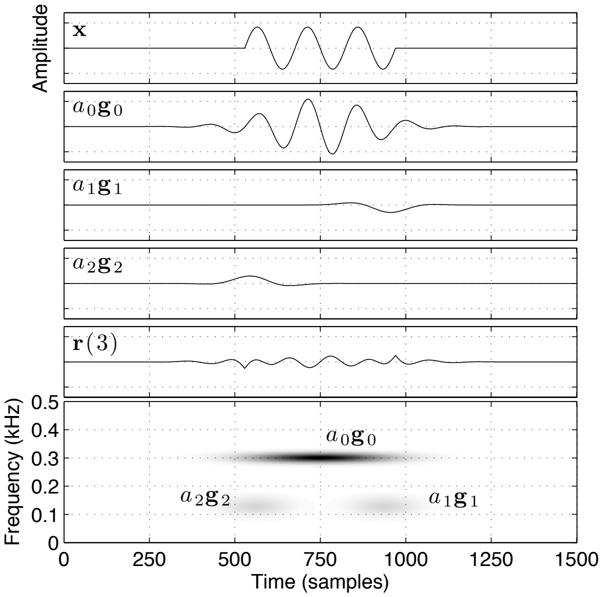

Figure 2. Signal $\mathbf{x}$, first three Gabor atoms found by MP, and residual signal $\mathbf{r}(3)$. Wivigram of atoms (bottom).

n	Type	a_n	s	u	$F_s\omega/\pi$	ϕ	
0	Gabor	18	1024	238	301.5	$-.49\pi$	$\alpha = 0.1$
1	Gabor	3.5	512	681	129.2	$-.67\pi$	$\alpha = 0.1$
2	Gabor	3.5	512	307	129.2	$.67\pi$	$\alpha = 0.1$

Table 2. Book of Decomposition from Fig. 2

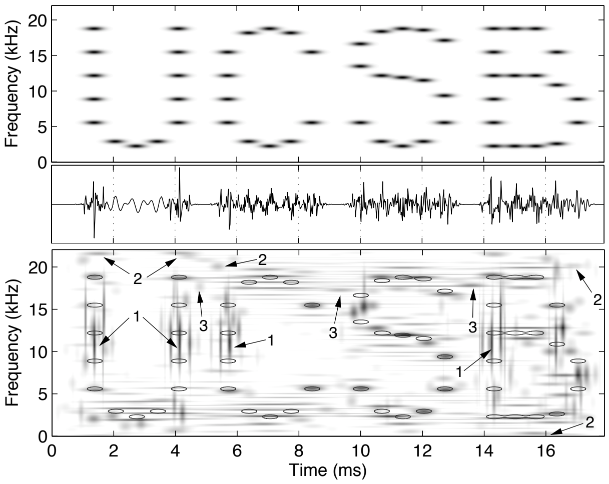

Figure 3. Signal (middle) built from 57 Gabor atoms seen in wivigram (top). Wivigram of decomposition (bottom) with outlines of atoms in top wivigram.

2.4. Wivigrams

The solution to (1) is stated in a book, which contains the parameters of all the dictionary atoms found in the decomposition process. We can obtain a picture of how energy is distributed in the representation by superposing the Wigner-Ville distribution (WVD) [23] of each atom in the book [1], 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 (7). Of all possible time-frequency structures, Gabor atoms have the least amount of spread in time and frequency [3].

2.5. Greed, Bias, and Uniqueness in Matching Pursuit

Consider the signal in Fig. 3, 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 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. 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.

Because MP selects at each step the atom that maximizes the energy removed from the residual signal, MP is called *greedy*. In the example shown in Fig. 3, MP decomposes the vertical portions of the letters "U," "C," and "B," into small-scale wideband atoms without considering them the result of atoms of a large scale that are in-phase. The "ideal" solution is, in a sense, lost from the very first atom selections, which is a clear example of how MP can heavily bias the results. We can force MP to reproduce the original representation by specifying a Gabor dictionary with atoms of scale 64 samples only.

These results demonstrate three important aspects of MP in particular, and 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 bias the solution, such as placing atoms in time and frequency regions where no energy exists in the original. We discuss this phenomena further in section 3.2.

3. ANALYSIS

DBMs can produce representations that are much less redundant and more meaningful than those provided by other transform methods, such as the short-time Fourier transform (STFT). Specifying the contents of the dictionary gives one adaptability in representing specific data or signals. With a good choice of a dictionary, the dimensionality of the original signal will become much smaller, from which analysis applications can greatly benefit.

3.1. Higher-level Representations Through Molecules

To work with data 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 thus designed an algorithm that builds molecular representations from atomic ones [24, 25]. Each molecule is a group of atoms that act together to represent a high-level feature. This approach was inspired by the McAulay-Quatieri algorithm [26], where a STFT is used to build a parametric sinusoidal model of speech. Molecular MP [13, 27] takes a similar approach, except molecules are built jointly with the signal decomposition.

The wivigram at top in Fig. 4 visualizes an MP decomposition of a bird call signal. Here *time-frequency tiles* are shown to highlight 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 [24]. Examples of tonal molecules are outlined in the bottom of Fig. 4. The relationships are now more clear between particular atoms in the atomic decomposition and the harmonic contents in the original signal. With these molecules, one can work more directly with the content of a signal through its atomic decomposition.

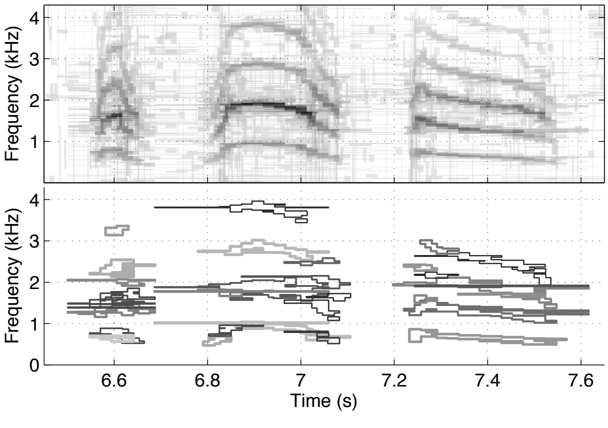

Figure 4. Wivigram (top) of a bird signal. Several outlined molecules (bottom).

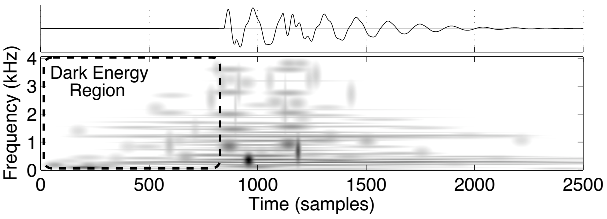

Figure 5. Wivigram (bottom) shows terms found by MP decomposition of signal (top) at times of no energy.

3.2. Dark Energy and Interference

Because of the non-orthogonal nature of the dictionaries used in DBMs, waveforms may interact and interfere in a representation [28, 29]. In the most extreme case, an atom of a representation will disappear in the resynthesis when superposed with the others. Several examples of this are seen in Fig. 5. Because of this effect we call all interference exhibited by a non-orthogonal representation *dark energy* [28, 30]. Such terms can be created by the decomposition algorithm to correct for "poor" atom choices made in earlier iterations, which obviously reduces the efficiency and meaningfulness of the representation.

Because of the unexpected difference in energies between those in the representation $\{\mathbf{G}(n), \mathbf{a}(n), \mathbf{r}(n)\}$ and in the approximation $\widetilde{\mathbf{x}}(n) = \mathbf{G}(n)\mathbf{a}(n)$, we have defined the *dark energy* associated with a given atom as the magnitude difference between the energy of the new approximation, and the energy of the approximation that would result were the new atom orthogonal to the current approximation [28]:

$$\Xi(n+1) = \left| ||\widetilde{\mathbf{x}}(n+1)||^2 - (||\widetilde{\mathbf{x}}(n)||^2 + |a_n|^2) \right| \quad (8)$$

$$= 2\left| a_n \mathbf{g}_n^T \widetilde{\mathbf{x}}(n) \right| \; n = 0, 1, \ldots \quad (9)$$

where (9) is true for a Euclidean vector space. This expression shows that the dark energy associated with the new atom $\mathbf{g}_n$ is proportional to the extent to which the atom is already present in the current approximation. Using a short-term measure of dark energy [30], we can see how dark energy is spread throughout a representation with respect to the signal. Figure 6 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 symmetric Gabor atoms.

Other researchers have attempted to avoid this phenomenon by changing the selection criterion used in MP [20, 31], or by specifying different functions for the dictionary [32]. We have instead sought ways to productively

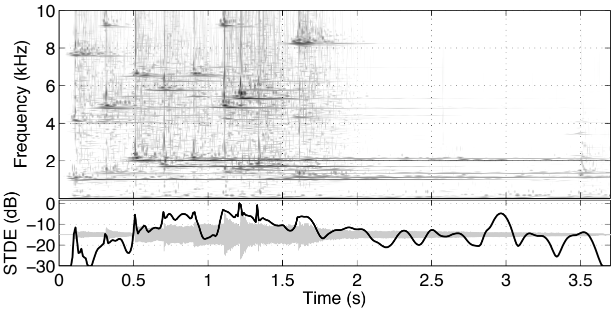

Figure 6. Wivigram (top) of Glockenspiel signal. Short-term dark energy overlaid on signal waveform (bottom).

use this behavior to learn about the signal and its relationship to the dictionary, and to measure the efficiency and meaningfulness of a decomposition [28, 30, 29].

4. 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. 7.

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. 8(a) is a short musical excerpt from *Pictor Alpha* [33]. Below it are three different representations. The spectrogram (log magnitude of STFT) is shown in Fig. 8(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 content is difficult. The scalogram in Fig. 8(c) shows the magnitudes of a dyadic wavelet transform (DWT) using the Gabor wavelet [34]. 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. 8(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.

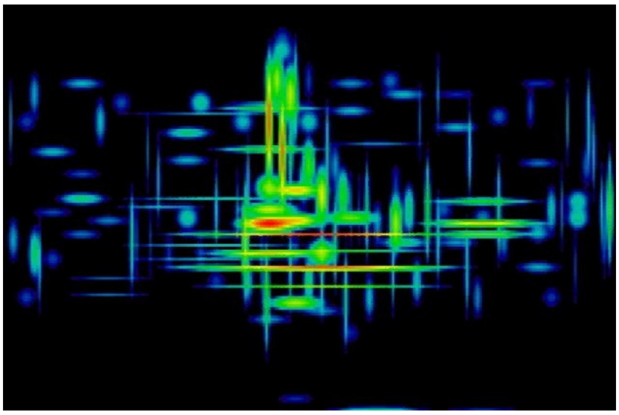

Figure 7. Wivigram visualization of a segment of electroacoustic work *Concrete PH* by Iannis Xenakis.

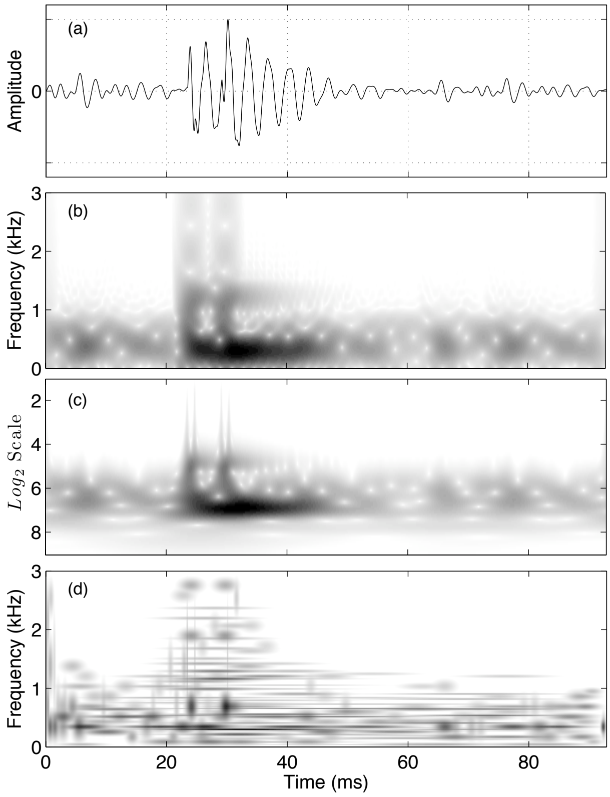

Figure 8. (a) Short musical signal. (b) Spectrogram from STFT. (c) Scalogram from DWT using Gabor wavelet. (d) Wivigram from MP using Gabor dictionary.

Using DBMs with a multiresolution dictionary (e.g., even a union of a wavelet and Fourier basis [27]), one can separate the stationary and transient content of an audio signal. Figure 9 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.

5. TRANSFORMATION

Describing a sound in terms of heterogeneous waveforms provides several unique ways in which to perform transformations [15]. Individual waveforms selected from the dictionary by DBMs can be modified independently or as groups, such as the molecules presented in Section 3.1. And due to the non-uniqueness inherent to (1) when using overcomplete dictionaries and minimally defined constraints, some solutions may provide more malleability than others, or suggest additional ways of modifying the content in a signal. Furthermore, since a resynthesis is

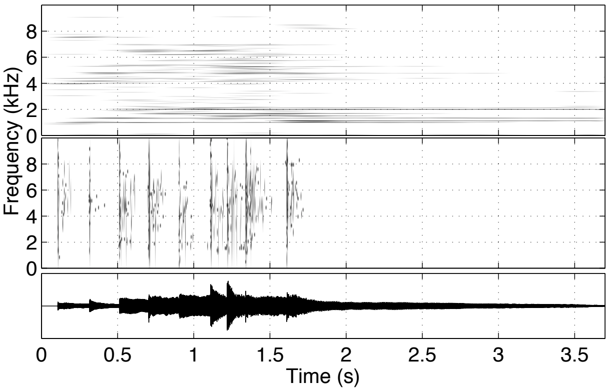

Figure 9. Wivigrams showing long (top) and short atoms (middle) in a musical signal (bottom).

essentially only a process of table-lookup and summation, many of these transformations can be done in real-time. Below, we outline and describe four classes of sound transformation using representations built by DBMs. A screenshot of an application that will provide an interface to working with these representations is shown in Fig. 10. Audio examples are on-line at `http://www.mat.ucsb.edu/~b.sturm/ICMC2008/`.

5.1. Filtering

Each waveform in a decomposition is described by a set of parameters, some of which are shown in Table 2 for Gabor atoms. These can include scale s, modulation frequency ω, translation u, and the weighting a_n.

5.1.1. Frequency Filtering

When dictionary waveforms can be associated with frequencies, transformations that are analogous to bandpass and notch filtering are realized by selecting only waveforms having a particular range of frequencies.

5.1.2. Amplitude Filtering

Filtering based on the weights $\mathbf{a}(n)$ in (4) basically selects components having energies within specified ranges. Since the energy of each waveform selected by MP grows smaller as the order increases [1], keeping those atoms that have energies above a given threshold will only generate low-order approximations of the signal. However, one can obtain exotic effects by amplifying atoms with low-energies, for instance, amplifying the ones found 20 dB below the fundamental.

5.1.3. Scale Filtering

Multiresolution atomic decompositions provide the ability to filter based on the scale, or duration, of waveforms. 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 works best on signals that have very distinct separations between such components, as seen in Fig. 9.

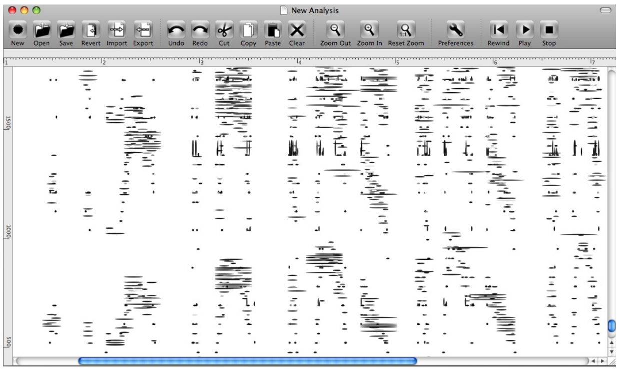

Figure 10. Screenshot of an interface for working with OM decompositions.

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 waveforms contribute. For instance, we can target distinct groups of waveforms that model the harmonics, such as those seen in Fig. 4. Atoms may even be found that are specific to instrumental morphologies, such as features particular to a piano [13]. An attractive element of using a heterogeneous atomic representation is the ease with which the transient and tonal portions of a signal may be separated and modified independently, as demonstrated in Fig. 9.

5.2. Parametric Manipulations

One can create exotic transformations of a decomposed signal by altering the various parameters used to describe the waveforms in the dictionary.

5.2.1. Pitch Shifting

Transposing an audio signal in frequency without affecting its temporal characteristics can be done using dictionary waveforms that can be associated with pitch, such as Gabor atoms. A naïve approach alters the modulation frequency of each atom [35]. 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 [35]. Since the relationships between the waveforms 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 original 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 changing its frequency by changing the scale of every waveform and adjusting its translation appropriately. This approach still suffers from not accounting for the interactions between waveforms. 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 waveforms in time and fills in the resulting gaps with additional waveforms having parameters interpolated between their neighbors. As in pitch shifting, there also exists the possibility of modifying only those atoms in 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. Granular Spatialization

Through atomic representations we can spatialize grains individually, thus inducing the decomposed sounds to take on novel perceptual qualities. In particular, the reconstructed sound retains its coherence (identity) with respect to the original, but different time-frequency components of the sound can be projected from different locations in a large-scale facility like the UCSB Allosphere [36]. Based on the representation, we also can parse the signal in many different ways based on its audio content (transients, harmonics, loud atoms, short atoms, etc.), and each parsing provides a basis for a novel spatialization.

5.2.4. Jitter, Bleed, and Scramble

A jittering effect can be created by offsetting waveform parameters, such as translation or frequency, according to a stochastic model. Retaining the center times of the waveform, but adjusting their durations — in a sense, "bleeding" them in time — creates a unique effect. One may also rearrange the waveforms in time and frequency, in essence scrambling their positions in the time-frequency plane.

5.3. Substitution

Given a signal decomposition that uses one dictionary, we may replace any or all of those waveforms with new ones. This technique has been explored using wavelets, but with varying degrees of success [5]. 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 dictionary used in a decomposition of speech with one containing only damped sinusoidal atoms produces "speaking chimes." Replacing damped sinusoidal atoms with Gabor atoms creates smoothed transients, and an effect similar to "reverse echo."

5.4. Physical Analogs

Thinking of sound as a combination of elementary units, and the metaphor of decomposing sound into atoms, motivates the conception of physically inspired transformations. Specifically, through manipulations of particle density, we can realize transformations such as evaporation (sonic disintegration), coalescence (sonic formation), and mutation (sonic metamorphosis). We can cause a sound to disintegrate by reducing its density by removing more and more atoms over time. In effect, we insert gaps in the representation until the sound evaporates. Imagine some dense pitch cluster that has been riddled with gaps by a process of atomic cavitation. It is transformed into a sparse sound cloud and becomes sonically "diaphanous." It is now possible to mix in another sound and hear it through the gaps in the original cavitated signal. The opposite of disintegration—coalescence—can be realized by simply permitting more atoms to be included in the resynthesis over time.

6. CONCLUSION

We have presented an overview of DBMs, as well new research exploring their application to analyzing, visualizing, and transforming audio and music signals in novel ways. 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. When time-localized waveforms are used, such as Gabor atoms, DBMs can be seen as providing the analytical counterpart to granular synthesis. However, a price is paid for the freedom to specify the dictionary, among which are increased computation, non-unique solutions, and complex interactions between atoms that can bias the results. While the first two are important for some applications, e.g., realtime communications, they are not critical to off-line audio and music signal processing. The third problem, clear accessibility to and meaningful representation of signal content through an atomic decomposition, is more important. To deal with these problem we have shown how the atoms of a decomposition can be combined into larger morphological structures, such as harmonics, which make more clear the significance of individual atoms to signal content, and which can be used for analysis, visualization, and transformation. Further work incorporating dark energy and interference into the decomposition procedure will reduce the negative effects of bias in the results.

7. ACKNOWLEDGMENTS

Garry Kling provided Fig. 7. This work was supported in part by the National Science Foundation under Grant CCF-0729229.

8. REFERENCES

[1] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[2] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Aug. 1998.

[3] D. Gabor, "Acoustical quanta and the theory of hearing," *Nature*, vol. 159, no. 4044, pp. 591–594, May 1947.

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

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

[6] M. S. Lewicki, "Efficient coding of natural sounds," *Nature Neuroscience*, vol. 5, no. 4, pp. 356–363, Mar. 2002.

[7] M. G. Christensen and S. van de Par, "Efficient parametric coding of transients," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1340–1351, July 2006.

[8] R. M. Figueras i Ventura, P. Vandergheynst, and P. Frossard, "Low-rate and flexible image coding with redundant representations," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 726–739, Mar. 2006.

[9] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[10] M. Aharon, M. Elad, and A.M. Bruckstein, "K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.

[11] S. Lesage, S. Krstulovic, and R. Gribonval, "Underdetermined source separation: Comparison of two approaches based on sparse decompositions," in *Proc. Int. Conf. Independent Component Analysis Blind Source Separation*, Charleston, South Carolina, Mar. 2006, pp. 633–640.

[12] O. Derrien, "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.*, Toulouse, France, Apr. 2006, vol. 5, pp. 57–60.

[13] P. Leveau, E. Vincent, G. Richard, and L. Daudet, "Instrument-specific harmonic atoms for mid-level music representation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 1, pp. 116–128, Jan. 2008.

[14] E. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Sig. Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[15] G. Kling and C. Roads, "Audio analysis, visualization, and transformation with the matching pursuit algorithm," in *Proc. Int. Conf. Digital Audio Effects*, Naples, Italy, Oct. 2004, pp. 33–37.

[16] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *J. Constr. Approx.*, vol. 13, no. 1, pp. 57–98, Jan. 1997.

[17] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1993, vol. 1, pp. 40–44.

[18] S. Krstulovic and R. Gribonval, "MPTK: Matching pursuit made tractable," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Toulouse, France, Apr. 2006, vol. 3, pp. 496–499.

[19] R. Heusdens, R. Vafin, and W. B. Kleijn, "Sinusoidal modeling using psychoacoustic-adaptive matching pursuits," *IEEE Signal Process. Lett.*, vol. 9, no. 8, pp. 262–265, 2002.

[20] R. Gribonval, E. Bacry, S. Mallat, Ph. Depalle, and X. Rodet, "Analysis of sound signals with high resolution matching pursuit," in *Proc. IEEE-SP Int. Symp. Time-Freq. Time-Scale Anal.*, Paris, France, June 1996, pp. 125–128.

[21] R. Gribonval, "Fast matching pursuit with a multiscale dictionary of gaussian chirps," *IEEE Trans. Signal Process.*, vol. 49, no. 5, pp. 994–1001, May 2001.

[22] R. Gribonval and E. Bacry, "Harmonic decompositions of audio signals with matching pursuit," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 101–111, Jan. 2003.

[23] D. Preis and V. C. Georgopoulos, "Wigner distribution representation and analysis of audio signals: An illustrated tutorial review," *J. Audio Eng. Soc.*, vol. 47, no. 12, pp. 1043–1053, Dec. 1999.

[24] B. L. Sturm, J. J. Shynk, and S. Gauglitz, "Agglomerative clustering in sparse atomic decompositions of audio signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, Las Vegas, NV, Apr. 2008, pp. 97–100.

[25] B. L. Sturm, J. J. Shynk, A. McLeran, C. Roads, and L. Daudet, "A comparison of molecular approaches for generating sparse and structured multiresolution representations of audio and music signals," in *Proc. Acoustics*, Paris, France, June 2008.

[26] J. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 34, no. 4, pp. 744–754, Aug. 1986.

[27] L. Daudet, "Sparse and structured decompositions of signals with the molecular matching pursuit," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 5, pp. 1808–1816, Sept. 2006.

[28] B. L. Sturm, J. J. Shynk, L. Daudet, and C. Roads, "Dark energy in sparse atomic estimations," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 3, pp. 671–676, Mar. 2008.

[29] B. L. Sturm, J. J. Shynk, and L. Daudet, "Measuring interference in sparse atomic estimations," in *Proc. Conf. Info. Sciences Syst.*, Princeton, NJ, Mar. 2008.

[30] B. L. Sturm, J. J. Shynk, and L. Daudet, "A short-term measure of dark energy in sparse atomic estimations," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2007, pp. 1126–1129.

[31] S. Jaggi, W. C. Karl, S. Mallat, and A. S. Willsky, "High resolution pursuit for feature extraction," *Applied and Computational Harmonic Analysis*, vol. 5, no. 4, pp. 428–449, Oct. 1998.

[32] M. Goodwin and M. Vetterli, "Matching pursuit and atomic signal models based on recursive filter banks," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 1890–1902, July 1999.

[33] C. Roads, "Pictor Alpha," in *Point, Line, Cloud*, compact disc and digital video disc. Asphodel Records, 2004.

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

[35] B. L. Sturm, L. Daudet, and C. Roads, "Pitch-shifting audio signals using sparse atomic approximations," in *Proc. ACM Workshop Audio Music Comput. Multimedia*, Santa Barbara, CA, Oct. 2006, pp. 45–52.

[36] X. Amatriain, J. Castellanos, T. Höllerer, J. Kuchera-Morin, S. T. Pope, G. Wakefield, and W. Wolcott, "Experiencing audio and music in a fully immersive environment," in *Lecture Notes in Computer Science: Sense to Sound*, K. K. Jensen, R. Kronland-Martinet, and S. Ystad, Eds. Springer Verlag, Berlin, Germany, in press 2008.