AUDIO ANALYSIS, VISUALIZATION, AND TRANSFORMATION WITH THE MATCHING PURSUIT ALGORITHM

Garry Kling and Curtis Roads

Media Arts and Technology
3431 South Hall
University of California,
Santa Barbara, CA 93106-6065 USA
http://www.mat.ucsb.edu/
g.kling@mat.ucsb.edu clang@create.ucsb.edu

ABSTRACT

The matching pursuit (or MP) algorithm decomposes audio data into a collection of thousands of constituent sound particles or gaborets. These particles correspond to the "quantum" or granular model of sound posited by Dennis Gabor. This robust and high-resolution analysis technique creates new possibilities for sound visualization and transformation. This paper presents an account of a first round of experiments with MP-based visualization and transformation techniques.

1. THE GRANULAR REPRESENTATION OF SOUND

In the 1940s, the Nobel Prize winning physicist Dennis Gabor proposed that any sound could be decomposed into acoustical quanta bounded by discrete units of time and frequency [1,2,3]. This quantum representation formed the famous Gabor matrix. Like a sonogram, the vertical dimension of the Gabor matrix indicated the location of the frequency energy, while the horizontal dimension indicated the time region in which this energy occurred. In a related project, Gabor built a machine to granulate sound into particles. This machine could alter the duration of a sound without shifting its pitch. In these two projects, the matrix and the granulator, Gabor accounted for both important domains of sound representation. The matrix was the original windowed frequency domain representation. The granulation machine, on the other hand, operated on a time domain representation.

Today such representations are labeled by a multiplicity of terms: "acoustic quantum," "grain," "gaboret," "Gaussian elementary signal," "short-time segment," "Gabor atom," "wavelet," etc. Roads [4] cites 32 different names. In this paper, we refer to all such techniques as *granular representations*, echoing the term used by the composer-engineer Iannis Xenakis, who first proposed a granular representation of musical sound [5,6,7].

Techniques that exploit granular representations have emerged as highly useful tools for the synthesis and transformation of musical sound. Recent advances let us probe and explore the beauties of this formerly unseen world. Granular techniques dissolve the rigid bricks of music architecture–the notes–into a more fluid and supple medium. Sounds may coalesce, evaporate, or mutate into other sounds. The sensations of point, pulse (regular series of points), line (tone), and surface (texture) appear as the density of particles increases. Sparse emissions leave rhythmic traces. When the particles line up in rapid succession, they induce the illusion of tone continuity that we call pitch. As the particles meander, they flow into streams and rivulets. Dense agglomerations of particles form swirling sound clouds whose shapes evolve over time.

The potential of granular representations has yet to be fully explored [4]. New approaches to signal analysis have demonstrated a variety of techniques that serve as analytical correlates to granular synthesis, under the broad category of wavelet or atomic decompositions [8]. For clarity and consistency, we refer to these analytical methods as "granular decompositions."

1.1. The Matching Pursuit Algorithm

Amongst the garden of granular decompositions lies the matching pursuit (MP) algorithm, pioneered by Mallat and Zhang [9]. The MP is a generalized framework for computing adaptive, granular signal representations. Many different flavors of the MP have been proposed [10]. The concept of the algorithm derives from Gabor: given an input signal, elementary particles can be combined to reconstitute that signal.

Figure 1 shows the operational flow of the MP algorithm. In step 1, the sound input, dictionary, residue, and output buffers are initialized. In step 2, the algorithm searches the particle dictionary to find the best match to the sound energy. The search procedure varies by implementation.

The core of this search algorithm is the inner product space used to calculate the correlation between the sound and the dictionary particles. In the simplest case, this is the standard dot product. Some varieties of the search procedure also contain a step that refines the current grain selection [13], or optimizes the search using previous results [10].

The dictionary construction varies according to the flavor of the MP being run. The structure and contents of the dictionary are flexible; the only requirement of the dictionary is that its vectors (grains) form a basis for the set of all input signals. In the common case, the dictionary is redundant, made up of pure sines, gaborets (a sine wave modulated by a gaussian envelope), and diracs (transient functions). This dictionary is useful for musical signals since it contains particles that can reconstitute the basic structures of a musical event: harmonic steady-state spectra and transient attack structures. In addition to this form, the MP algorithm has been adapted to use dictionaries of chirplets [11], and damped sinusoids [12].

Once a grain is chosen, it is subtracted from the signal in step 3. The remainder is called the residue, and is used for subsequent

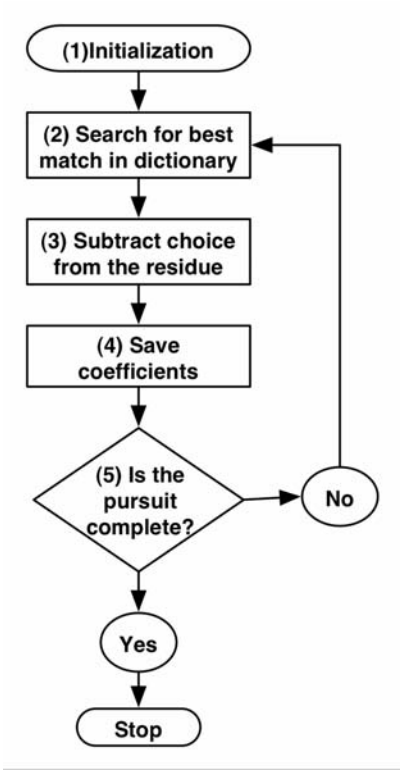

Figure 1: *The Basic Flow of the Matching Pursuit. See text for explanation.*

search. The coefficients (or dictionary indices) are saved (step 4), and the exit condition is tested in step 5. The pursuit is terminated when either a specified number of particles are found, or a percentage of the signal's energy has been dissolved. If this condition is not met, the process returns to step 2.

The result of the analysis is a granular representation. Grains are listed parametrically by the grain's center time, frequency, phase, amplitude, and duration.

1.2. Matching Pursuit analysis versus Fourier analysis

The Short-Time Fourier Transform (STFT) forms the core of most analysis and transformation methods employed by musicians. The toolbox created around the STFT is extensive, with new applications being added every year. However, certain properties of the Fourier transform make it less than ideal for certain applications. Initial experimentation with MP analysis has yielded provocative results that suggest that it may perform certain tasks at a higher level of quality than the STFT.

Perhaps the most notable defect of the STFT is poor localization of sound structures in both time and frequency. The STFT is not well suited to describe events that are smaller or larger than the hop size of the analysis [9]. The frequency resolution is directly tied to the window size, which further complicates event onset localization. The time-frequency tradeoff of the STFT distorts the reality of sound, fogging the acoustic lens of computer music with spectral clutter. Since MP analysis uses a multiscale dictionary of atoms with arbitrary frequency resolution, time-frequency resolution is not tied to the size of the analysis sample. Visual comparisons of this phenomenon are given below.

Another important difference between the STFT and MP is translation invariance. This property of the MP representation makes it useful for pattern search and recognition, since a feature's representation is not dependent on its time-frequency location. The STFT is translation invariant as well, but this property is destroyed by sampling the translation parameter uniformly [8].

The translation invariance of the MP representation lets us alter and extract information while conserving the energy content of the original signal. It is well known that changing the parameters of a single frequency bin in an STFT alters the entire spectrum, and thus character of a sound. Initial experiments have shown that deleting, transposing, or otherwise altering a selected group of grains does not destroy the timbral character of a sound.

2. SOUND VISUALIZATION WITH THE MATCHING PURSUIT ALGORITHM

The granular representation created by MP analysis affords a unique and powerful visualization technique. Utilizing the Wigner-Ville distribution in a simple but novel way, these visualizations, which we would like to term "microsonograms", let one look inside the life of a sound. The following is a brief overview of the techniques involved with some pictorial examples. We will present video examples during our talk.

2.1. Spectral Visualizations

Visualizations of spectral analysis are essential tools for computer music. For a history of spectrum analysis and visualization see [14]. Spectral analysis entered the modern era in the 1960s when Cooley and Tukey pioneered the fast Fourier transform and digital computing made its calculation practical [14]. Researchers were finally able to capitalize on the local Fourier analysis proposed by Gabor [1].

The Fourier spectrogram depicts sound as continuous strata, zones of intensity blurred across time and frequency. Pointed sonic gestures become dull smears. The granular representation and its visualization, as presented by Mallat and Zhang [9], give us a more detailed visual account of sound's inner structure.

2.2. The Wigner-Ville Distribution and the Microsonogram

Mallat and Zhang [9] introduced a visualization method for the MP analysis data that uses Wigner-Ville distribution to plot the time-frequency energy of an analyzed signal. The Wigner-Ville Distribution (WVD) is a development that parallels the theories of Gabor and representative of a second (and fundamentally different) direction in time-frequency analysis [15]. While brought into the signal processing world by Ville in 1948, Wigner originally developed it in 1932 in the context of quantum thermodynamics [8]. In contrast to the granular model of signal energy, the WVD is a quadratic time-frequency energy distribution that is computed by correlating a signal with a time and frequency translation of itself. The result is a time-frequency distribution with a continuous character, as opposed to the granular model above.

The utility of the standard WVD is limited due to interference terms that make interpretation problematic [15]. However, this property is overcome using the granular representation produced by the MP. The standard MP uses only static frequency gaborets.

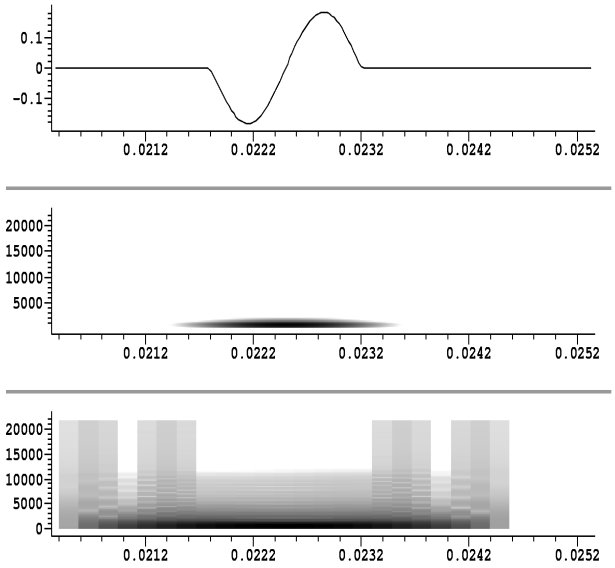

Figure 2: *A single cycle of a sine wave at 440 hertz. At top is the signal plot, in the middle the single particle found with MP analysis, and at the bottom is spectrogram of a 256 point STFT with a Hamming window.*

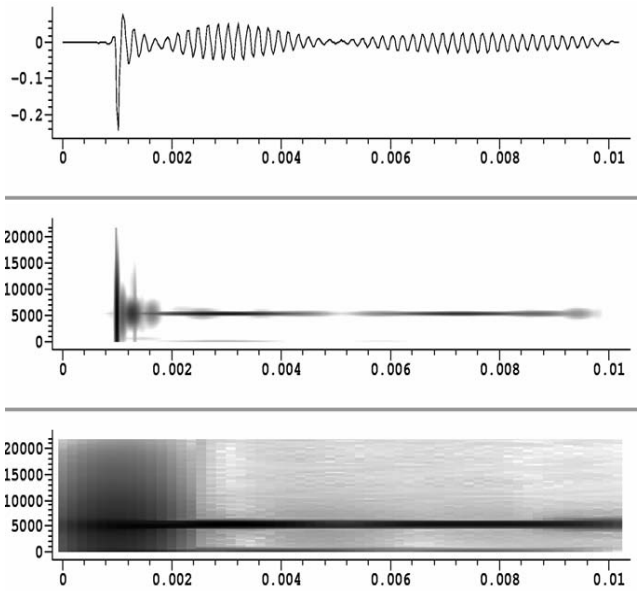

Figure 3: *A 100ms excerpt from the composition Pictor alpha by Curtis Roads. At top is a plot of the excerpt. In the middle is a microsonogram of the first 100 particles of a MP analysis of the signal. At the bottom is a spectrogram of a 256-point STFT with a Hamming window.*

The WVD of a single gaboret does not contain interference terms since it is well localized in time and frequency. The WVD of each grain found in the MP is summed [9], and the result is a representation that is distinct from the sonogram. It is a collection of grains whose parameters are independent, rather than a continuum of sine waves that are locked together in time.

It should be noted that although MP analysis is amenable to dictionaries composed of a wide variety of grain waveforms, certain waveforms cannot be visualized with the WVD without interference. Gaborets, chirplets, and other stationary or linearly modulated, single component particles can be visualized with good results. Grains with multiple frequency components and non-linear frequency modulations require more complicated procedures to visualize clearly, with a limited degree of success [15].

2.3. Visualization Examples

Figure 2 is an example using a single cycle of a sine wave at 440 cps. At top is a signal plot, in the middle is the WVD of the single grain found with an MP, and at bottom is a 256-point STFT with a Hamming window. It can easily be seen how the MP representation is better localized in time and frequency. Figure 3 is a brief excerpt from the composition *Pictor alpha* by Curtis Roads [18]. At top, the signal plot of the excerpt, composed of grains and a sharp transient, in the middle, a microsonogram of the first 100 particles found in an MP analysis, and at bottom, a 256 point STFT with a Hamming window. Note how the STFT (bottom) while recovering the main pitched component in rough form, dissolves the transient across the time-frequency plane. The microsonogram (middle) represents the transient well, as well as better showing the amplitude variation of the pitched component. All components of the sound are localized.

3. SOUND TRANSFORMATIONS USING MATCHING PURSUIT ANALYSIS

Granular representations of sound present us with new possibilities for sound transformation that are impossible or impractical using STFT-based methods. Granular representations have advantages that free us of many of the conceptual and technical encumbrances of the Fourier representation. But in order to forge new sound effects with this data, we must, to some degree, leave our Fourier sensibilities behind us. This will free us to take full advantage of the granular nature of sound energy.

Additive synthesis techniques draw from STFT-based methods due to their direct conceptual relationship. Similarly, we can find a wealth of information to fertilize our experiments with granular representations by looking at granular synthesis and transformation techniques.

For our experiments, we wrote software that analyses sound data using the MP functions [19] in the LastWave software package [20], which contains the most complete implementation of MP analysis. The transformations are performed on the resultant data, which is resynthesized with the CSound score language. We will present an overview of the results of our experiments thus far. Further results will be presented during our demonstration.

3.1. Pitch-Time Effects

Experiments have shown the MP representation of sound to be well suited for creating transposition and time stretching effects. As with the standard time-domain granulation versions of these effects [21], there are myriad ways that we can use to achieve these results.

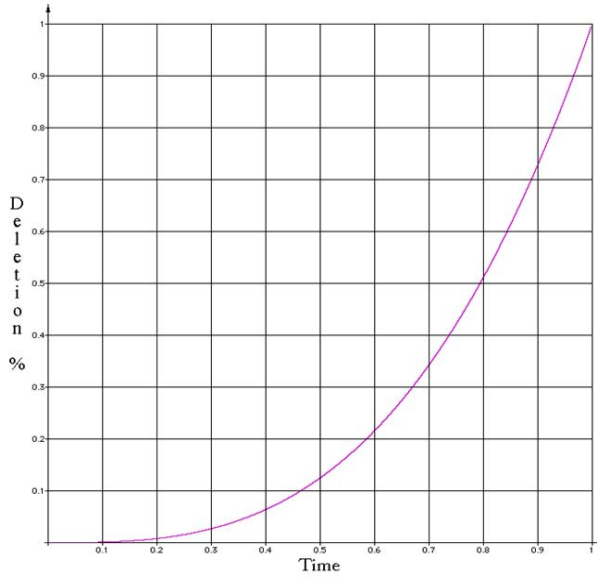

Figure 4: *Schematic view of the Gabor matrix and an overlaid disintegration function.*

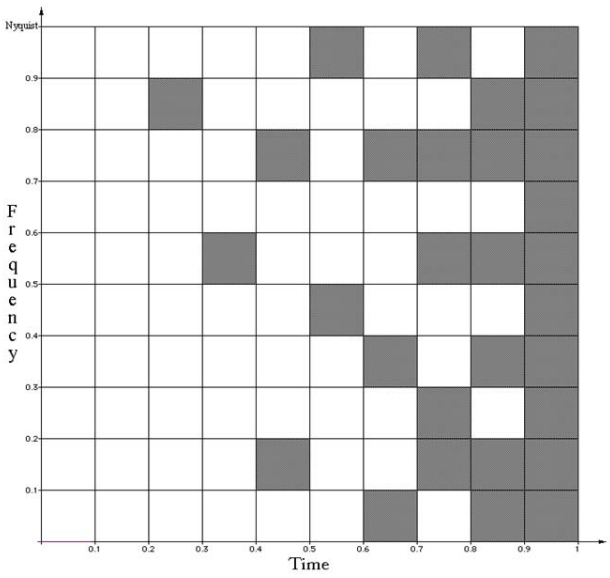

Figure 5: *Schematic view of a shattering gradient.*

Transposition of the granular representation is particularly effective for percussive, noisy sounds, in contrast to phase vocoder techniques. Since transients are well localized, the attack portion is retained, which makes the effect convincing. Sounds that are harmonic are also transposed well, owing to better frequency localization than the STFT. Policies for transposing harmonics while keeping noise in place will be needed to improve this effect.

Time compression and expansion are effective even using a simple multiplicative method. Time compression is quite excellent using the granular representation, again owing to preservation of transient structures. An intelligent model, such as the deterministic plus stochastic model in SMS [23], will produce superior results. Grains cannot be lengthened or shortened limitlessly without losing the character of the original sound.

3.2. Thresholding

Amplitude thresholding is an effect that can sift noise or harmonicity from a sound. This transformation is carried out by resynthesizing grains whose amplitude is lesser or greater than some coefficient. Resynthesis of grains that fall below some level tends to recover the noise components of a sound. Resynthesis of grains greater than some level extracts harmonic content and strong transients from a sound. The warbling effect created by an incomplete analysis or a threshold that recovers only a small portion of a sound is sure to become a cliché of MP based techniques.

3.3. Octave Splitting

In the parlance MP analysis, the octave of a grain denotes its duration in samples, expressed as a power of two. For efficiency reasons, the dictionary used in the LastWave implementation of the algorithm contains only grains of some power of two in duration.

Given that we have a representation that sorts grains by duration, an obvious experiment is to synthesize grains of each size separately. It was found that shorter octaves retain the prosody of the original signal while the longer octaves take on the harmonic characteristics of the input. Shorter octaves have a characteristic noisy and transient timbre, while the longer ones are “warbled” or “watery” and are more pitched. This effect was found to have a very consistent character no matter what sound sample was used.

3.4. Coalescence and Disintegration

Past implementations of these effects were carried out with the tracking phase vocoder (TPV) in the QuickMQ program [22]. In this program, the sinusoidal tracks can be altered with a variety of algorithms. The Granny algorithm realizes disintegration and coalescence [22] through a process that cuts holes in the time-frequency tracks of a sound. This transformation, however, is marred by transient artifacts that compromise the overall effect.

For the present realization of coalescence and disintegration, a time frequency grid is laid across the granular representation, as shown in Figure 4. The cells of this grid roughly correspond to the cells of the Gabor matrix, and grains whose centers fall within a given cell are said to occupy them. The duration in seconds and size in frequency of the grid are adjustable.

The amount of the effect is governed by an arbitrary function (Figure 4). When the function governing the effect deletes more grains as the sound progresses, disintegration results. Less and less grains are kept to produce coalescence. The function can, however, take any form. A random number is drawn for each cell in the matrix. If the value of this number is greater than the value of the function during that cell, the grains contained by that cell are deleted. The resulting grid is called a shattering gradient, which can be stored for later editing and reuse (Figure 5).

These effects are delicate, and require a great deal of attention for clarity. Successful coalescence results in a sound being reconstituted in all areas of the spectrum at once, but in a chaotic fashion. Well-constructed disintegrations leave a sound falling to pieces from within, as if by some internal force. In the future, the

simplest versions of these two cases will be automated for quick preview use, suitable for subsequent customization.

4. CONCLUSIONS AND FURTHER WORK

Granular representations of sound offer us powerful new ways to look at and interact with audio material. Early experiments have shown that these tools give us promising new directions in musical signal processing. The properties of the MP give us a robust, high resolution representation that affords new methods for understanding and transforming sound.

Extensive exploration is needed to realize the full potential of the MP and the resultant granular representation. Computation of the MP is very time-consuming, making the algorithm useful to only those with patience or powerful equipment. Optimizations, dictionary design, and search strategies are topics that are being actively researched.

The visualizations produced by MP analysis and the WVD can reveal structures that are blurred in the traditional spectrogram, particularly transients and finely spaced frequencies. This is something that shows promise for examining the acoustics of musical instruments and other detailed sonic phenomena.

Creation of granular synthesis-inspired transformations extends the artistic toolbox of composers and sound designers. The effects presented here are simple examples of the many possibilities that granular representations have to offer.

In the near future, we will be able to use the granular representation of sound to create a broad menu of audio analysis, visualization, and transformation techniques unheard of in the Fourier domain. The already versatile toolbox that the STFT affords is being augmented with ones using a particulate theory of sound. This is a universe in which sound has a time pattern and a frequency pattern, an elementary principle [2] that was only partially acknowledged in Gabor's time, but one that we fully appreciate in our own. It is only in the present time, with analytical tools such as MP analysis and the granular representation that we are beginning to exploit this property of sound directly.

5. REFERENCES

[1] D. Gabor, "Theory of communication," *J. Inst. Elect. Eng.*, vol. 93, no. III: 429–457, 1946.

[2] D. Gabor, "Acoustical Quanta and the Theory of Hearing," *Nature*, 159(4044): 591–594, May 1947.

[3] D. Gabor. "Lectures on communication theory," Technical Report 238, Research Laboratory of Electronics. Cambridge, Massachusetts: Massachusetts Institute of Technology, 1952.

[4] C. Roads, *Microsound.* Cambridge, Massachusetts: MIT Press, 2002.

[5] I. Xenakis, "Elements of stochastic music," Gravensaner Blätter 18: 84–105, 1960.

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

[7] I. Xenakis, *Formalized Music.* Revised edition. New York: Pendragon Press, 1992.

[8] S. Mallat, *A Wavelet Tour of Signal Processing.* San Diego: Academic Press, 1998.

[9] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Proc.*, vol. 41, no. 12, pp. 3397–3414, 1993.

[10] G. Davis, S. Mallat and Z. Zhang, "Adaptive Time-Frequency Approximations with Matching Pursuits," Technical Report 657, Computer Science Department, NYU, March 1994.

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

[12] M. Goodwin, "Matching Pursuit With Damped Sinusoids," *Proc. IEEE ICASSP*, 1997.

[13] R. Gribonval, P. Depalle, X. Rodet, E Bacry, S. Mallat. "Sound signal decomposition using a high resolution matching pursuit," in *Proceedings of International Computer Music Conference (ICMC'96)*, Clear Water Bay, Hong-Kong. pp. 293–296, August 1996.

[14] C. Roads, *The Computer Music Tutorial.* Cambridge, Massachusetts: MIT Press, 1996.

[15] P. Flandrin, *Time-Frequency/Time Scale Analysis.* San Diego: Academic Press.

[16] E. Wigner, "On the quantum correction for thermodynamic equilibrium," *Physical Review*, 40: 749–759, 1932.

[17] J. Ville, "Théorie et applications de la notion de signal analytique," *Cables et Transmission, 2éme.* A(1): 61–74, 1948.

[18] C. Roads, *Pictor alpha.* From *Point, Line, Cloud.* Compact disc and digital video disc, Asphodel. Forthcoming.

[19] R. Gribonval, E. Bacry, and J. Abadia. MP Software and Documentation. Internet: `http://www.cmap.polytechnique.fr/~bacry/LastWave/packages/mp/mp.html`, 2003.

[20] E. Bacry. LastWave Software and Documentation. Internet: `http://www.cmap.polytechnique.fr/~bacry/LastWave/`, 2003.

[21] P. Dutilleux, G. De Poli, and U. Zölzer, "Time-Segment Processing," in *Digital Audio Effects*, ed. U. Zölzer, pp. 201–236. Chichester: Wiley, 2002.

[22] S. Berkely, "QuickMQ: a software tool for the modification of time-varying spectrum analysis file," M.S. thesis. Hanover: Department of Music, Dartmouth College, 1994.

[23] X. Serra and J. Smith III, "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition," *Computer Music Journal*, vol. 14, no. 4, Winter 1990.