ELEMENTARY SOURCES: LATENT COMPONENT ANALYSIS FOR MUSIC COMPOSITION

Spencer S. Topel **Michael A. Casey**
Bregman Music Audio Research Studio
Dartmouth College

ABSTRACT

Complexity of music audio signals creates an access problem to specific musical objects or structures within the source samples. Instead of employing more commonly used audio analysis or production techniques to access features, we describe extraction of sub-mixtures from real-world audio using a Probabilistic Latent Component Analysis-based decomposition tool for music composition. This is highlighted with the presentation of a prior relevant compositional approach named Spectral Music along with a discussion of five compositions extending these principles using methods more commonly associated with source separation research.

1. INTRODUCTION

Music recordings of all types consist of mixtures; the recorded sources are transformed via real or virtual acoustic processes and these are summed to make a stereo or mono track. A major challenge facing machine analysis of audio is extracting information contained in mixtures for the purpose of isolating relevant content. Techniques based on independent component analysis (ICA), such as independent subspace analysis (ISA) [8] and Probabilistic Latent Component Analysis (PLCA) [20], provide ways of accessing perceptually motivated musical objects [17], which we describe here as "sub-mixtures".

For compositions based upon music audio, either for information or analysis, accessing latent features expands the creative possibilities. Specific technical innovations contribute to this perspective: fast Fourier transform (FFT) signal analysis allowed composers associated with *Spectral Music* to explore spectral profiles for the purpose of generating material for pieces [10]. PLCA likewise extends the concept of exploring audio content for music composition by providing the means necessary to extracting components with independent properties, or what we describe here as sub-mixture content, i.e. distinctive streams of sound objects grouped by their common frequency-amplitude statistics.

We present a repertoire of works that use such techniques. Pieces were included based on their relevance to the techniques presented in this paper, and their specific aesthetic insights or innovations with respect to component manipulation. Preceding this, a discussion of the historically significant *Spectral Music* composition movement is presented, where spectrogram analysis and data-mining generated material for new musical works.

2. BACKGROUND

Audio decomposition methods in computer-assisted music composition have a rich history in musical discourse over the past twenty years, notably Wishart's expansion of Pierre Schaeffer's *Music Objets* [22], and Smalley's Spectro- morphology [19]. One approach of particular relevance, formally named *Musique Spectrale*, uses extracted timbral features for the purpose of creating new works, both acoustic and electro-acoustic in nature. Spectral Music was first introduced by Henry Dufourt in 1979 [10], however by this time compositions using materials extracted from Fourier transforms had already been written by the group, most notably Gérard Grisey's "Partiels Pour 18 Musiciense" [12], where the composer proposed macro-synthesis of analyzed sources using combinations of acoustic musical instruments. In most of these early pieces, there was a transparent and straightforward process by which the composer derived materials for new works, outlined in Figure 1.

A wellspring of software development at the Institut de Recherche et Coordination Acoustique/Musique (IRCAM) occurred during this time, including the visual programming environment Max, after Max Matthews, along with many of the well-known IRCAM package modules including "Music V" brought to IRCAM by Jean-Claude Risset, who came to further timbre research, as well as *CHANT*, developed by Xavier Rodet, and the transcription tools that now belong to *Open Music*. Amongst the composers belonging to the first-wave of Spectral composition were Gérard Grisey, Tristan Murail, Hugues Dufourt, and British composer Jonathan

Harvey. Their music influenced a younger cadre of composers, including Magnus Lindberg, Marc-Andre Dalbavie and Joshua Fineberg.

The purpose of this shift in compositional thought possesses deeper connections to the European *Zeitgeist* of the 1960's and 1970's. The prevailing school of thought being "Total Serialism" was the only credible way to be a respected contemporary composer, built upon ideas first pioneered by the Second Viennese School, consisting of Arnold Schoenberg, and his disciples Anton Webern, and Alban Berg [11]. The Spectralists realized the aesthetics associated with mainstream Serialism of the time disregarded the final sounded musical experience. Instead, these techniques favored abstraction in notation and formalism blinded, to a certain extent, by the ideas that all 12-tone (half-step relationships) were equal, and that non-western tuning systems were not relevant to mainstream Western "art" music.

The Spectralists thus considered observation of sounded acoustic sources to be the starting point of their work. They further rejected both ideas that 12-tones were equal. Their early repertoire emphasized these points, with pieces such as "Godwana" [15], that evokes a sense of non-western tuning by exploring the relationship between a synthetic bell source and a spectrally analyzed trombone sample and later with "Désintégrations" [16] that utilizes the careful blending of timbres between electronic sounds and acoustic sources.

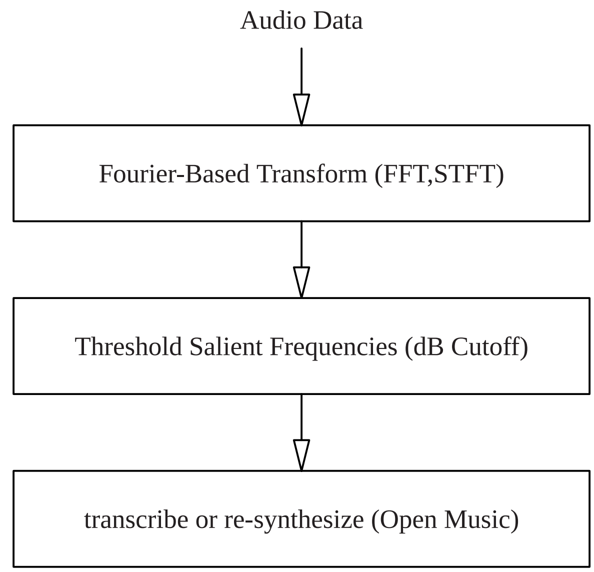

Figure 1. Spectralist composers used these techniques to generate novel harmonic material. The first step would be to perform a Fourier Transform, that either extracted only frequency information (FFT), or preserved temporal resolution of the frequencies (STFT). Then by using the resulting frequency information, the composers built chords with specific dynamics and orchestration to re-create the spectral profiles of the source materials.

3. REVEALING LATENT STRUCTURE IN AUDIO

Aside from socio-political currents in their work, the Spectralists' perspective emphasized uncovering hidden information in sound to generate new musical ideas. In the same way, *Latent* structure refers to identifying distinctive or salient parts of recorded audio that otherwise remain hidden. For *Spectralists* this meant identifying structural partials that distinguished one instrument from another playing the same perceived note (e.g. an "A" or "Bb"), as these partials relate to some organization of harmonic material.

Independent Subspace Analysis extraction techniques offer another way to access structure in audio, since re-synthesized latent components retain correlated behaviors between frequency and amplitude information in each component. When PLCA is used on magnitude-only STFT representations, the extracted components have similar characteristics to the output of phase vocoder methods with an important distinction: in addition to spectrum and envelope decompositions, components are further segmented by statistical independence of information content or recurrence of embedded acoustic patterns in the sound.

3.1 Probabilistic Latent Component Analysis

PLCA is a generalization of Non-negative Matrix Factorization (NMF) and a multi-variate generalization of Hoffman's bi-variate Probabilistic Latent Semantic Analysis (PLSA) [13] [18] [20]. For the purposes of modeling a time-frequency distribution such as the magnitude STFT, the PLCA model has the following form:

$$P(\mathbf{x}) = \sum_z P(z)P(w|z)P(h|z) \tag{1}$$

where $P(\mathbf{x})$ is the 2-dimensional distribution of the random variable $\mathbf{x} = wh$. z is a latent variable which we interpret to be an additive spectrogram component of $\mathbf{x}$. These marginals, w and h, are frequency components and amplitude components, respectively, of independent latent magnitude spectrograms. The marginal distributions themselves are dependent on a latent variable z. The objective of this analysis is to find out the underlying structure of a probability distribution. This is done by estimating $P(w|z), P(h|z)$ and $P(z)$ from an observed $P(\mathbf{x})$ using a version of the Expectation Maximization (EM) algorithm [9]. Following [18], the expectation step estimates the contribution of the latent variable z:

$$R(\mathbf{x}, z) = \frac{P(z)P(w|z)P(h|z)}{\sum_{z'} P(z')P(w|z')P(h|z')} \tag{2}$$

and in a maximization step we re-estimate the marginals using the above weighting to obtain a new and more accurate estimate:

$$P(z) = \int P(\mathbf{x})R(\mathbf{x}, z)d\mathbf{x} \tag{3}$$

with

$$P(w|z) = \frac{\int P(\mathbf{x})R(\mathbf{x}, z)dh}{P(z)} \tag{4}$$

and

$$P(h|z) = \frac{\int P(\mathbf{x})R(\mathbf{x}, z)dw}{P(z)}. \tag{5}$$

Repeating these steps converges to a solution for the marginals and the latent variable priors. Figure 3 illustrates the decomposition of magnitude STFT time-frequency distributions into probabilistic latent components using this 2-dimensional marginal decomposition algorithm. By considering the distributions as signal matrices, the final form of the matrix factorization using PLCA is:

$$\mathbf{X} = \mathbf{WZH}^{\mathrm{T}}. \tag{6}$$

Where W is the spectral distribution and H is the temporal distribution, the product of which produces the spectrogram reconstruction matrix $\mathbf{X}$. A prior, Z is introduced that weights the relative contribution of each component within the spectrogram reconstruction matrix. For basic procedures aimed at generating composition material, the Z prior can be omitted since we do not need to preserve the relative contribution (or loudness) of each component to the original mixture when re-composing with them. With the basic mechanics of Independent Component Analysis methods described, we now examine applications of these algorithms for music composition.

4. SOUNDSPLITTER: COMPONENT-WISE RE-SYNTHESIS

The compositions described below used a Matlab tool called SoundSplitter. Sounds were loaded from 44.1kHz samplerate 16-bit WAV format and analyzed using the short-time Fourier transform (STFT), yielding a sequence of vectors, $\mathbf{X}$, with 4096 samples per frame and an overlap of 2048 samples. For each frame, only the first 2049 magnitude Fourier coefficients were retained to eliminate redundancy due to the symmetry of the Fourier transform for real-valued signals. Optionally, the sequence of analysis frames was divided into segments using fixed-length blocks of STFT frames, with block-length typically between 1s and 10s in duration. Each block was analyzed using the PLCA2D algorithm [20] to yield three matrices per block corresponding to the frequency marginals, amplitude (probability) coefficients, and time marginals respectively: $\mathbf{W}, \mathbf{Z}$ and $\mathbf{H}$. The number of columns in these matrices corresponded to the number of components, n, requested in the analysis.

Component-wise re-synthesis produced a magnitude spectrogram for each marginal component, k, using the re-synthesis equation $\mathbf{X}_k = \mathbf{W}_k\mathbf{Z}_{k,k}\mathbf{H}_k^T$ for component spectrogram $\mathbf{X}_k$, a column vector for the kth frequency marginal, $\mathbf{W}_k$, an amplitude scalar from the kth diagonal entry in $\mathbf{Z}$, $\mathbf{Z}_{k,k}$, and the transposed kth column vector from the time marginals, $\mathbf{H}_k^T$. The compnent spectra, $\hat{\mathbf{X}}_k$, were re-synthesized by symmetrically expanding the magnitude spectrum around the Nyquist frequency and multiplying by the complex exponenentiated phase argument from the STFT of the source signal $\mathbf{X}^\star$, such that $\hat{\mathbf{X}}_k = \mathbf{X}_k e^{j\arg(\mathbf{X}^\star)}$ for $k \in \{1 \ldots n\}$ where the sum of the marginals forms the identity: $X^\star = (\sum_{i=1}^{n} \mathbf{X}_i)\, e^{j\arg(\mathbf{X}^\star)}$. The k-th component signal was computed using overlap-add re-synthesis, via the inverse short-time Fourier transform of $\hat{\mathbf{X}}_k$, with overlap corresponding to the hop size used in the analysis step and each window multiplied by a raised cosine window to smooth the transition between adjacent frames. The component-wise audio signals were controlled independently by the composers using digital audio workstation software to yield the compositions described below.

5. LATENT COMPONENT ANALYSIS FOR MUSIC COMPOSITION

The PLCA2D algorithm used in the current version of SoundSplitter has the attribute of extracting *fixed* re-occurring patterns in frequency and amplitude. These components retain a qualitatively higher level of structure of the original audio than non-pattern extraction methods, (e.g. bandpass filtering, spectral frequency decomposition, or phase-vocoder decomposition). The following section discusses five works employing these techniques in the order of their first performances. Each work manipulates components differently, but they share a common trait of using components to articulate specific and recognizable characteristics of the original audio samples at specific moments.

5.1 *Strange-Charmed* (1999), by Michael Casey and Simon Atkinson

Strange-Charmed [7] used independent subspace analysis (ISA) of spectrogram data [8] to generate an expanded set of sound materials from a set of textural and granular source sounds consisting of Geiger counters, insects, band-pass filtered water sounds, and scraped metallic objects. In contrast to PLCA, the JADE algorithm [2] for independent component analysis was used which yields components having both positive and negative values. For time-frequency distributions with negative values, such as non-rectified filterbank outputs, the ISA method is well formed, but it is inefficient due to the vast quantity of data generated by the filterbank. For real-time use the magnitude Fourier transform was employed, and any negative values in the magnitude Fourier transform reconstruction had to be truncated to zero for re-synthesis. A custom real-time synthesizer software

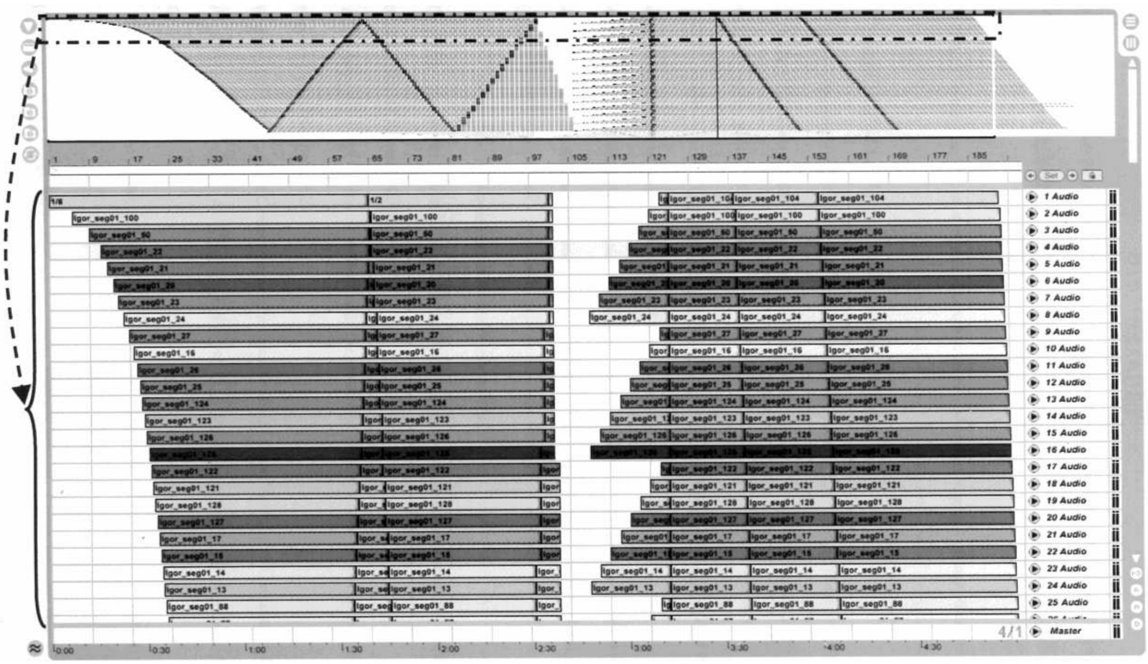

Figure 2. Components were introduced one at a time, with the misaligned components gradually aligned to bring the heterophony of asynchronous components into a state of order, clearly revealing the source as Stravinsky's iconic Rite of Spring chord.

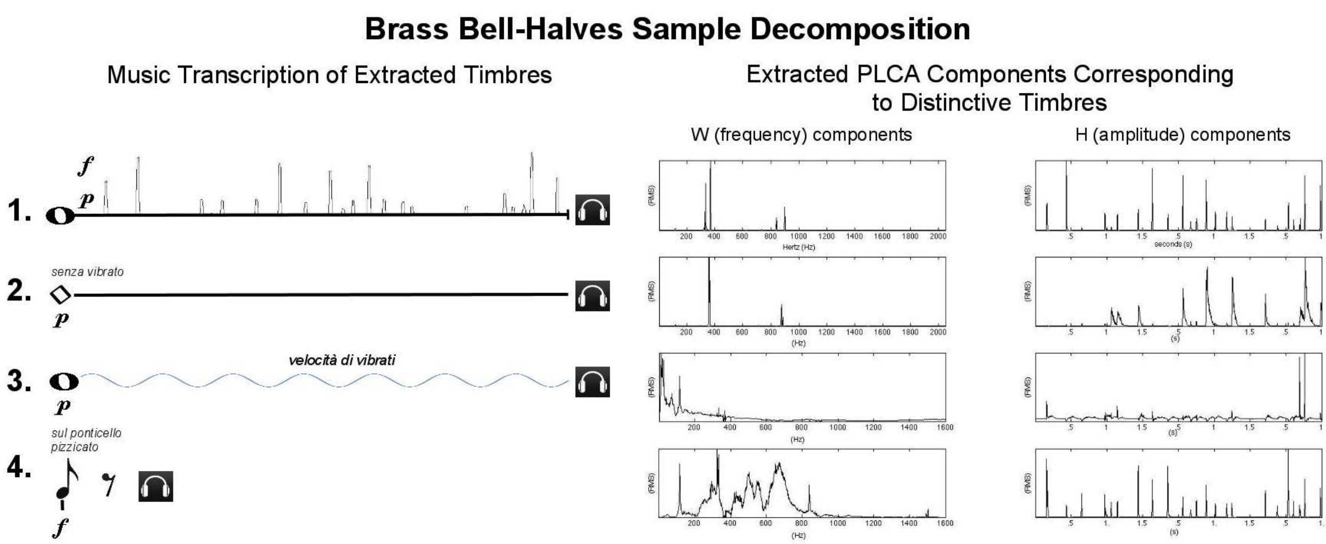

Figure 3. Each horizontal row corresponds to an extracted component, where the left-hand side of the figure shows the transcriptions to music notation, and the right-side shows the plots of the **W** and **H** decompositions per-component. From top to bottom: 1)A transient-laden component over the relative duration of a whole-note, where the peaks represent loud articulations and the troughs are equivalent to pianissimo. 2) A clean bell-partial, 3)"wobbling" of the bell-halve settling on the concrete. 4) Cement "click", articulated by a near-pitchless pizzicato near the bridge of the instrument.

package was designed and implemented that used MIDI to control the balance of amplitudes, and the relative delay, of each component of each sound independently. The real-time system, described in detail in [6], was used to shape the sounds and the output recorded for off-line editing and layered into the final composition using digital audio workstation software.

5.2 *Stratovinksy* (2010), by Paul Osentinsky

The concept behind *Stratovinksy* is the gradual revealing of an important musical instance. Namely, the iconic, jaggedly repeating, chord from the "Omens of Spring: Dances of the Youths and Maidens" movement of Igor Stravinsky's "Rite of Spring" [21] [14]. Using SoundSplitter, 128 components were extracted from 2 seconds of audio, re-synthesized to audio and then normalized. The resulting characteristics of these components were micro-tonal with low-energy levels distributed across the many extracted components.

The entirety of the data was then imported into Ableton Live as WAV files, temporally mis-aligned, looped, and distributed across sixteen virtual channels; see Figure 2. The layered components gradually came into alignment over the course of several minutes, with the effect of this process being equivalent to seeing a blurred object slowly come into focus.

5.3 *Decomposing Autumn* (2010), by David Plans Casal

Decomposing Autumn [3] was a live performance utilizing component-wise decomposition and live improvised reconstruction of *Autumn in Warsaw*, the sixth in György Ligeti's 2nd book of piano etudes. The latent component analysis method is coupled with real-time component-wise audio retrieval using *SoundSpotter* [4] as a foundation for a structured improvisation. Here, the PLCA2D algorithm was employed as a decomposition technique to obtain a corpus of separated sound fragments, which were then queried by a live improviser performing on a custom-built acoustic guitar. The approach used in this composition effectively bridges the fields of latent component analysis, music information retrieval by audio matching, and composition.

5.4 *Violine* (2011), by Spencer Topel

A work with a similar aim as *Stratovinksy*, at highlighting and revealing musical structure is *Violine*, for solo violin and laptop. The source material for each movement consisted of a short $12 - 90$ second audio clip of J.S. Bach's Chaconne in d minor from Partita No. 2 for solo violin. Doing so preserved not only the composed structure explicit in the J.S. Bach's notation, but also the timings and articulations supplied by the performer, (e.g. rolling of chords, chord voicing, and rubato).

The approach in this composition was two-fold: first to use SoundSplitter to perform a decomposition to isolated individual notes or pitch-classes, and then employ SoundSpotter to match a live violin signal directly on audio features analyzed on the extracted component database [5]. The basic function of SoundSpotter allows for matching between pitch and timbre characteristics, both of which were utilized in *Violine*. Combined together, SoundSplitter and SoundSpotter, the compositional material becomes a combination of the composers intentions and the performers interpretation, similar to the pieces discussed in [4].

The component decomposition parameter was again critical in the pre-composition phase of this piece. An eight component-wise decomposition proved to be effective at extracting clear, well-formed sounding components. A musical score was then written using the analysis provided by the SoundSplitter decomposition. A recent version of Real-time SoundSpotter as a VST plug-in, provided an immediate and interactive way to match sounds of the live violin, resulting in a near-seamless counterpoint between the extracted components and the composition.

5.5 *Elementary Sources* (2011), by Spencer Topel

Elementary Sources examines SoundSplitting as a means of decomposing audio into different timbral objects that contribute the identity of the original source. The movement discussed here, of five movements for string quartet and laptop, was written using a single 101 second audio recording of brass bell-halves Case dropped on concrete and recorded by artist Case Hathaway-Zepeda. SoundSplitter was used to extract components relating to different events segregated by timbre, which included cement "clicks", resulting from the moment the metal hit the ground, in-harmonic partials from the metals as the bell-halves rang, and the oscillation, or wobbling, of the halves as they came to rest on the cement. Different component extraction parameters were explored and components were re-synthesized and auditioned. Through trial and error, eight components were identified as the yielding the best sounding results. Additional experimentation with component re-synthesis included creating components that had cross-spectral characteristics with the three categories described above. Specifically, spectral signatures and time-trajectories from SoundSplitter were recombined in different ways to extend the timbral palette without extracting new audio sources.

A prevailing idea in *Elementary Sources* was to acoustically synthesize a specific audio sample, like the bell-halves, with an entirely different set of sources, such as a string quartet. This is not unlike Gérard Grisey's ideas for his landmark work Partiels, where he describes the concept of using the orchestra for the purpose of Macro-synthesis, where each instrument of the orchestra contributes specific time and frequency behaviors that culminate in an overall synthesis of a spectral profile, rather than distinctive sections and instrumental motives [12].

Two methods were explored to determine how best to have the four acoustic instruments perform the different extracted components. Firstly, a music notation relating to each distinctive component was devised to best articulate the time-frequency behaviors, shown in Figure 3. Secondly, audio samples were provided for the performers to audition the sounds for themselves to best determine the execution of these components. The two methods proved to work better in combination than in isolation, since the notation provided a starting-point, and the playback of components were of high enough quality as to provide additional information for each performer that could not be captured in the music notation.

With basic elements of the bell-halves translated to the quartet, it was now possible to use a combination of live-electronics (e.g. pitch-shifting, reverberation, compression), and re-synthesized component sample playback to achieve a fairly close relationship between the sampled sources and the acoustic instruments. The result was the creation of an interstitial space between bell samples and their transcriptions, where relationships in the compositional materials were a subsequent outgrowth of the timbres from the original bell-halve sources.

6. SUMMARY

A new approach to music composition using Latent Component Analysis techniques is described, along with five compositions and accompanying examples that demonstrate the usages of these techniques, which are accessible online [1]. We also show here that Spectralism overlaps with these concepts, since a shared theme in both repertoires is the application of computer analysis to uncover latent features in music audio. Future work will examine how component estimation more deeply influences compositional material and how recent innovations on PLCA and related algorithms can be used for better extraction of compositionally relevant information.

7. REFERENCES

[1] http://digitalmusics.dartmouth.edu/ismir2011, September 2011.

[2] Jean-François Cardoso. Source separation using higher order moments. In *Proc. ICASSP-89*, pages 2109–2112, 1989.

[3] D.P. Casal and M. Casey. Decomposing autumn: A component-wise recomposition. In *Proc. ICMC*, 2010.

[4] M. Casey. Soundspotting: A new kind of process. *The Oxford Handbook of Computer Music*, 2009.

[5] M. A. Casey. http://soundspotter.org/.

[6] M.A. Casey. *Auditory Group Theory with Applications to Statistical Basis Methods for Structured Audio, Ph.D. Thesis*. MIT Media Laboratory, 1998.

[7] M.A. Casey and S. Atkinson. *Strange-Charmed: MIT EMS@25 CD, Track 9*. MIT Media Laboratory, 1999.

[8] M.A. Casey and A. Westner. Separation of mixed audio sources by independent subspace analysis. In *Proceedings of the International Computer Music Conference*, pages 154–161, 2000.

[9] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc. B*, 39(1):1–38, 1977.

[10] H Dufourt. Musique spectrale. *Musique, pouvoir, écriture*, pages 289–290, 1979.

[11] M.J. Grant. *Serial music, serial aesthetics: compositional theory in post-war Europe*, volume 16. Cambridge Univ Press, 2005.

[12] Gérard Grisey. *Partiels Pour 18 Musiciense*. Ricordi, 1975.

[13] T. Hoffman. Probabilistic latent semantic indexing. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 50–57. ACM Press, New York, 1999.

[14] L.A. Philharmonic I. Salonen. Stravinsky: Le sacre du printemps, 2006.

[15] T. Murial. *Godwana*. Transatlantiques, 1980.

[16] T. Murial. *Désintégrations*. Éditions Salabert, 1989.

[17] P. Schaeffer. Solfege de l'objet sonore. *INA/GRM*, 1967.

[18] Madhusudana Shashanka, Bhiksha Raj, and Paris Smaragdis. Probabilistic latent variable models as non-negative factorizations. Technical report, Mitsubishi Electric Research Laboratories, December 2007.

[19] D. Smalley. Spectromorphology: explaining sound-shapes. *Organised Sound*, 2(2):107–126, 1997.

[20] Paris Smaragdis, Bhiksha Raj, and Madhusudana V. S. Shashanka. Sparse and shift-invariant feature extraction from non-negative data. In *ICASSP*, pages 2069–2072. IEEE, 2008.

[21] I. Stravinsky. *Rite of Spring*. Dover Publications, 1989.

[22] T. Wishart and S. Emmerson. On sonic art. *12*, Routledge, 1996.