EASY DOES IT:
THE ELECTRO-ACOUSTIC MUSIC ANALYSIS TOOLBOX

Tae Hong Park
Tulane University
park@tulane.edu

Zhiye Li
Tulane University
zli3@tulane.edu

Wen Wu
Tulane University
wwu1@tulane.edu

ABSTRACT

In this paper we present the EASY (Electro-Acoustic muSic analYsis) Toolbox software system for assisting electro-acoustic music analysis. The primary aims of the system are to present perceptually relevant features and audio descriptors via visual designs to gain more insight into electro-acoustic music works and provide easy-to-use "click-and-go" software interface paradigms for practical use of the system by non-experts and experts alike. The development of the EASY system exploits MIR techniques with particular emphasis on the electro-acoustic music repertoire – musical pieces that concentrate on timbral dimensions rather than traditional elements such as pitch, melody, harmony, and rhythm. The project was mainly inspired by the lack of software tools available for aiding electro-acoustic music analysis. The system's frameworks, feature analysis algorithms, along with the initial analyses of pieces are presented here.

1. INTRODUCTION

The idea for EASY Toolbox originated between 1999-2000 in the form of a master's thesis entitled "Salient Feature Extraction of Musical Instrument Signals" [11] which included a Java-based feature extraction and visualization software called Jafep (Java Feature Extraction Program). Since then, the project has somewhat been dormant, at least in direct relation to its original intention. Portions of the research evolved to automatic instrument recognition studies and further lead to the FMS software synthesis system [10] and most recently has developed into the EASY Toolbox to assist in the analysis of electro-acoustic music.

There is much interest and on-going research in MIR on various dimensions of music and a wealth of research can be found in pertinence to traditional music especially popular music. Some examples include rhythm analysis, melody analysis, tonality, traditional harmony, music recommendation, genre classification, instrument identification, and composer identification to name a few [2,16]. As far as MIR techniques and its applications in the area of music are concerned, much of the focus seems to be *outside* the realm of electro-acoustic music. One of the reasons for the scarcity in MIR-based research for electro-acoustic music may perhaps be attributed to the need for MIR researchers to be interested *and* actively be involved in composing or be deeply engaged in electro-acoustic music on a musical level. Another reason for this somewhat imbalance may be that the community seems to prioritize resources to the more standard musical repertoire that the general public accesses.

Some works related to the topic of music analysis software include Jafep (Java Feature Extraction Program), jAudio, Wavesurfer, Vivo, JRing, SoniXplorer, Sonic Visualizer [1, 3, 4, 6, 8, 11, 14], and others [17, 18]. Jafep is a Java-based feature extraction system for displaying feature vectors in a two-dimensional canvas and includes a harmonic follower designed mainly for analysis of musical instruments which is similar to jAudio. However, jAudio further concentrates on providing a feature extraction library/repository. Wavesurfer is a system for speech research and displays waveforms, pitch information, spectrograms, and formants. Vivo and JRing focus on pitch-based music, where JRing additionally deals with incorporating traditional scores for musicological studies. SoniXplorer is an interesting application which again primarily pays attention to traditional and popular music using self-organizing clustering algorithms. The Sonic Visualizer seems to be designed to address traditional music also, that is, pieces involving pitch, harmony, and rhythm. Although it has the ability to display audio features, perhaps due to the original design of the software architecture, when analyzing electro-acoustic music type signals, the visualization environment does not seem to be ideally suited for such situations. Marsyas (and to a lesser degree MIR Toolbox) is probably the most extensive environment for MIR research. It seems especially well suited for "DSP experts" and for the more experienced software developers/researchers but is perhaps not ideal for "users" who are looking for out-of-the-box software applications with simple and intuitive GUI interfaces as well as viewing capabilities – ready to use applications for specific purposes.

The EASY Toolbox is a modest initial step towards applying MIR theories with particular emphasis on electro-acoustic music focusing on its potential in gaining musical insights based on salient feature extraction techniques and clustering with the primary objective being that of an analytical tool wrapped with an intuitive GUI environment.

2. THE EASY TOOLBOX

2.1 Core Concept

One of the important characteristics of numerous electro-acoustic music, especially those pieces that are in the tape music genre, is that they are often concerned with aspects of timbre and sound color opposed to traditional musical elements such as pitch, harmony, and rhythm. However, although there are examples of software systems for analyzing "pitch-based music" as discussed in the introduction, there does not seem to be much of any software that is available for the analysis of music that do not adhere to those time-honored musical parameters. There is much software available for viewing raw waveforms and spectrograms but that type of information does not really offer too much insight by itself. Hence, our approach is to utilize salient feature extraction techniques as the basis for music analysis to uncover hidden information that is timbrally and perceptually relevant and perhaps even helpful in revealing additional data about a given work. We have also included segmentation/clustering algorithms using model-based and distance-based techniques. The algorithms that are implemented and used for displaying various features are hidden from the user as much as possible in order to render an easy-to-use interface. Furthermore, we have attempted to present the feature vectors in intuitive ways by plotting data in the time/frequency-domain and timbre spaces using 3D representation/navigation techniques. With a straightforward "click-and-go" environment provided by EASY, we hope that users will be encouraged to explore various timbral dimensions thereby help better understand sound objects and music.

2.2 EASY Features

2.2.1 The EASY Interface

The two main canvases in EASY are time-domain and frequency-domain displays as shown in Figure 1.

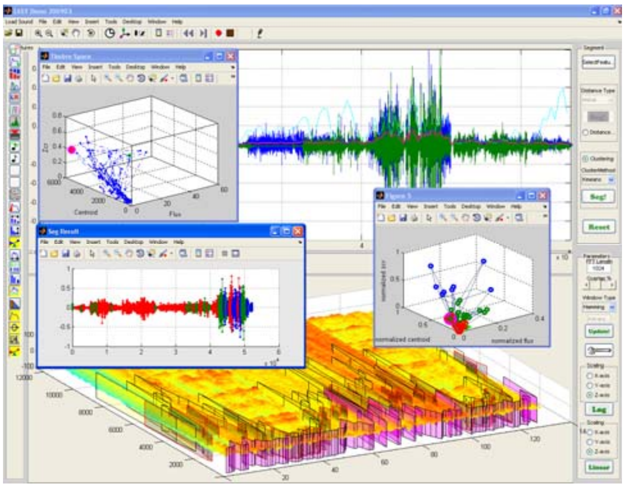

Figure 1. Screenshot of EASY

The approach of designing the EASY interface was driven by the aim of providing the user a 3D visualization environment for sonic exploration and interaction. For example, the waveforms for stereo files or multichannel files are presented in a cascading style along with the corresponding spectrogram.

The control areas of EASY include time/frequency-domain parametric control and feature selection for analysis/display. Standard functionalities such as zoom-in, zoom-out, 3D navigation/rotation, viewing options inherited from MATLAB®, the real-time input DAQ option (see Section 2.2.3), and a transport control are also included. Further controls are available for clustering and segmentation such as feature selection for clustering, number of clusters, and clustering algorithms as further discussed in Section 3.

2.2.2 EASY 3D Timbre Space Plots: the timbregram

EASY provides intuitive 3D timbre space representations adopted from [7] for sonic exploration which we call timbregrams. Figure 2 shows a timbregram example of a time-sequenced three instrument signal – bass guitar followed by clarinet and French horn with three timbre dimensions (spectral spread, spectral centroid, and spectral flux).

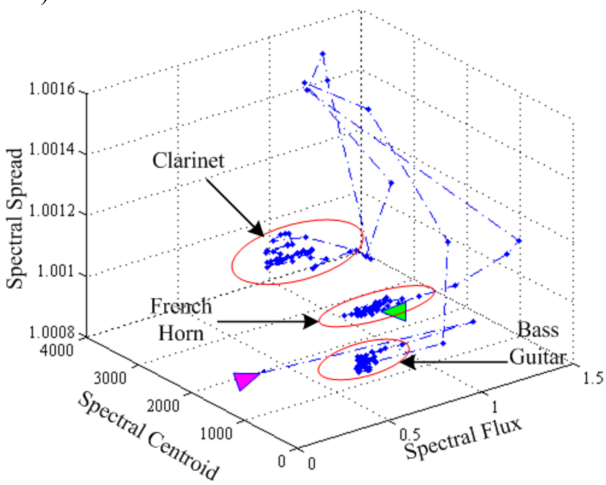

Figure 2. Timbre Space Example

The dots and dashed lines portray the 3D timbral trajectory as a function of time where the right pointing triangle refers to the beginning of the sample and the left pointing triangle the end of the sample. Each node represents a time unit equal to the frame/hop size. During audio playback, feature vector following occurs not only in the time-domain and frequency-domain canvases but also in the timbregram canvas itself (displayed in a separate window as shown in Fig. 1). This allows intuitive observation of sonic events via synchronization between the visuals and the audio that is played back.

2.2.3 "Real-Time" and MATLAB® Data Acquisition Toolbox

One of the advantages in using MATLAB® is the incredible resource of toolboxes available for data analysis and manipulation. One such example is the Data Acquisition (DAQ) Toolbox used for real-time analysis applications. The EASY system exploits the DAQ for analyzing and displaying input signals (mic/line input) in "real-time." It can display one or multiple features (selectable by the user) in the time and frequency-domain as well as the timbregram canvas.

2.3 EASY Algorithms

A total of 26 features in the time and frequency-domain are implemented in this current version of EASY – amplitude envelope, amplitude modulation, attack time, crest factor, dynamic tightness, frequency modulation, low energy ratio, noise content, pitch, release time, sound field vector, temporal centroid, zero-crossing rate, 3D spectral envelope, critical band, harmonic compression, harmonic expansion, inharmonicity, MFCC, modality/harmonicity/noisiness, spectral centroid, spectral flux, spectral jitter, spectral roll-off, spectral shimmer, and spectral smoothness. Many of the feature extraction algorithms themselves were developed in the FMS Toolbox [10, 12] and have been customized for use in EASY. Below, we present a short description of a select number of new features that we developed.

The dynamic tightness feature measures the quantized time-amplitude histogram on a frame-by-frame basis and provides insights into the "tightness" or "holiness" of the distribution of quantized sample values. This idea is shown in Figure 3 showing a highly compressed electric bass slide sample displaying a densely populated bed of samples throughout the amplitude axis bounded by the compressor threshold value.

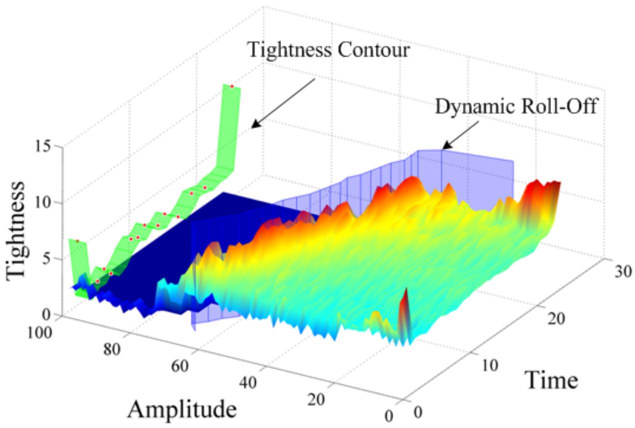

Figure 3. Electric Bass Slide: Compressed

Modality/harmonicity/noisiness is a method for analyzing a signal in terms of its harmonic, modal, and noise content. As shown in Figure 4, the harmonicity, modality, and noise floor levels of a signal are computed and displayed over time. One way of computing the harmonicity and modality is via the fundamental frequency (f_0) and the drift (e_k) in Hz. e_k is found by first determining spectral peaks, followed by computing their distances with respect to the closest ideal harmonic locations. The modality ("excessive inharmonicity") of each harmonic component can be then computed as the ratio of the drift and the fundamental frequency. As expressed in Equation (1) and (2), taking the mean of the inharmonicities of all the harmonics can be used to derive the modality of a signal.

$$Modality = \frac{e_1 / f_0 + ... + e_k / f_0}{k} \quad (1)$$

$$Harmonicity = 1 - Modality \quad (2)$$

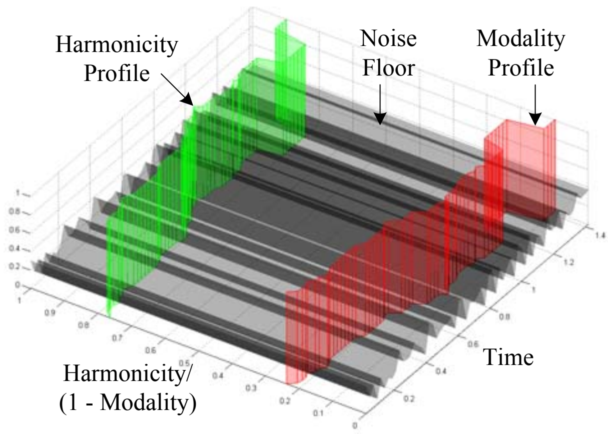

Figure 4. Modality/Harmonicity/Noisiness

The computation of the noise floor is based on sound flatness measure (SFM) – the ratio of the geometric mean and the arithmetic mean which has been used in speech research to extract voiced and unvoiced signals. When the signal is considered to be above the noise threshold (via SFM), the fundamental frequency is estimated which is then followed by modality analysis. On the other hand, if the signal's SFM value is determined to be below the noise threshold, it will be considered as noise. Another feature included in EASY is the multi-channel sound field vector developed by Travis Scharr while at Tulane University. This feature enables mutli-channel audio file display as a vector sum of the energy in each of the audio channels as a function of time.

3. SEGMENTATION ALGORITHMS

The two segmentation methods that we developed are based on clustering and distance measurement-based techniques as described in this section.

3.1 Model-based Segmentation: Clustering

The model-based approach for segmentation exploits a timbral feature vector clustering scheme. The audio input is first subjected to a silence detector followed by frame-by-frame feature extraction. The *N*-dimensional feature space is then piped to the clustering algorithm (eg. k-means). The clusters are then remapped to the time-domain in a color-coded fashion for visual clarity as shown in Figure 5.

3.2 Distance-based Segmentation

The distance-based segmentation algorithm applies statistical analysis of extracted features selectable by the user. The statistical analysis itself uses a long-term windowing scheme (main frames) to compute the average feature trajectory on a window-by-window (via sub-frames) basis – each sub-frame represents a single data point. Each main frame is then analyzed for its mean and standard deviation – the standard deviation is the distance measure used for segmentation. The distance can be computed via Euclidian distance, Kullback-Leibler distance, Bhattacharyya distance, Gish distance, Entropy loss or Mahalanobis distance.

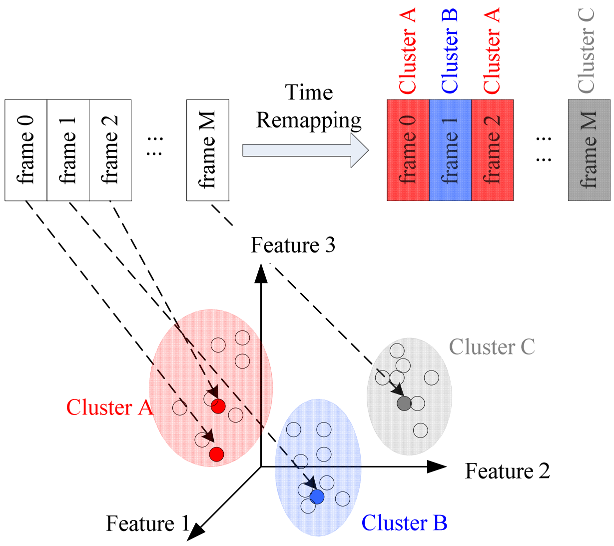

Figure 5. Time-Remapping in Clustering-based Segmentation for 3 Features

4. PREMLINARY ANALYSIS RESULTS

We used two pieces to conduct preliminary analysis of electro-acoustic works – *Machine Stops* (Tae Hong Park) and *Riverrun* (Barry Truax). We chose *Machine Stops* as we have first-hand detailed knowledge about the construction of the piece and *Riverrun* as it's not only an electro-acoustic masterpiece, but also because it is very much based on timbral compositional strategies.

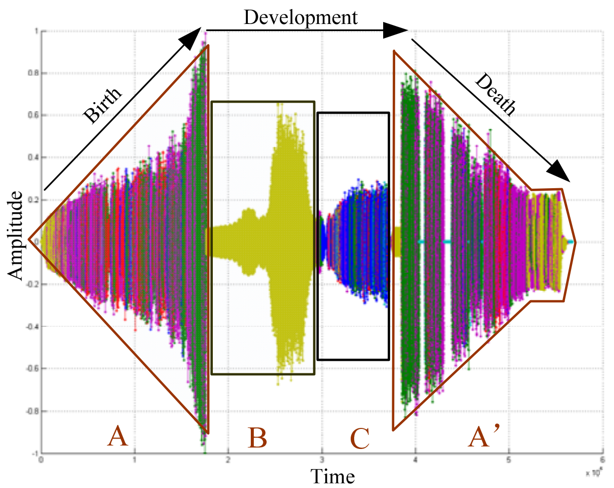

Figure 6. Segmentation Map of *Machine Stops*

A number of general observations could be made just by using single features such as modality/harmonicity/noisiness (MHN), dynamic tightness (DT), spectral centroid (SC), and the spectrogram (SG) itself. The extracted information included insights about where harmonic sections started and ended, where more modal sections occurred (via MHN), locating timbrally bright sounding parts (SC), exposing dynamically compressed areas (DT), and observing overall energy distributions and shifts (SG). However, what was most interesting in our initial analysis was discovering "segmentation maps," "timbregram trajectories," and "segmentation/cluster tracks" as shown in Figures 6, 7, and 8. Looking at the segmentation map we can generally identify four sections (A, B, C, A') via the color-coded segmentation regions and the amplitude envelope. The intro A (labeled as "birth") shows a triangular structure with a general build-up of energy. This is mirrored, slightly fragmented, in A' during the "death" phase of the piece which illustrates the overall arching shape of the piece itself. A' also includes an extended portion of the beginning part of the piece, adding a prolongation of decay towards the end (the "machine" coming to a "stop").

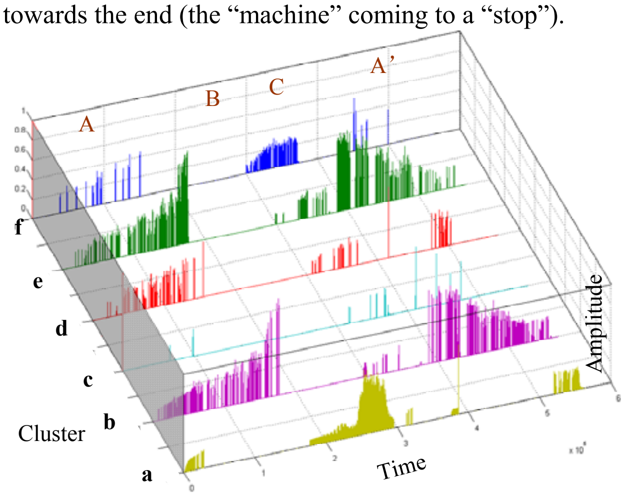

Figure 7. Segmentation/Cluster tracks (*Machine Stops*)

Figure 7 which displays the decomposition of the segmentation map into individual "cluster tracks," further exposes this build up and loss of energy of parts A and A' and also depicts the introduction of section C (cluster f) as new material (⑤ in Figure 8). Section B generally represents a sparse timbral construct exemplified by single and harmonically distorted sine-waves (in the HMN analysis plot, harmonicity is maximal in region B – not shown here).

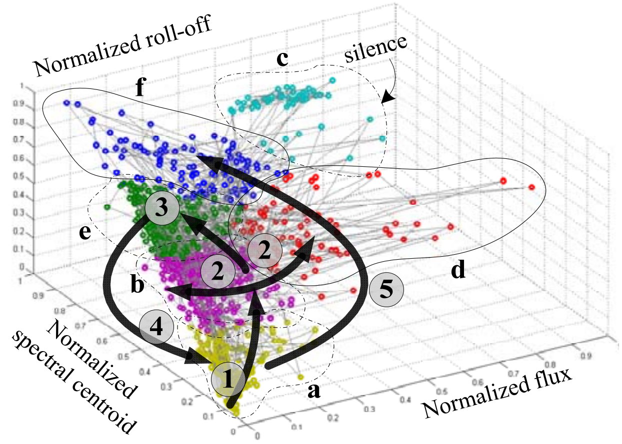

Figure 8. Timbregram Trajectory of *Machine Stops*

As shown in the timbregram plot (Figure 8) we can clearly view (when following the cursor during playback) the timbral trajectory which generally follows ① to ②, ③, ④, and ⑤ during the "birth" and "development" sections of the piece. The timbregram is also useful in displaying continuous timbral changes between cluster **a**, **b**, and **e** while also showing abrupt jumps in the timbre space between clusters **e** and **a** as well as **a** and **f**. The closing triangular portion follows the inverse trajectory ③ to ①.

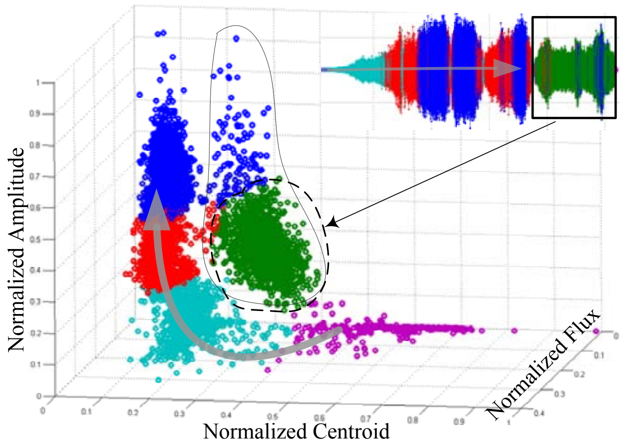

Figure 9. Timbregram and Segmentation Map of *Riverrun*

A similar analysis was conducted for *Riverrun* where we concentrated in particular on the segmentation map, cluster tracks, and timbgregram. It was quite straightforward to identify sectional divisions in the spectrogram as expected, but what was particularly interesting in the timbrgram was the finding that, unlike in *Machine Stops*, the colonization of the timbre space portrayed a distinct separation of one particular cluster from the rest – the timbral cluster pertaining to the closing section of the piece with high spectral centroid as shown in Figure 9. At the same time, the continuous development as described in [13] of *Riverrun* can also be clearly seen in Figure 91 beginning from a sparse quiet group of droplets, developing to rivulets, streams, and massive oceans towards the main part of the piece. Various feature sets have been employed in generating clusters, segmentation maps, and timbregrams. Interestingly enough, for the majority of the cases, the ensuing results have been quite similar when interpreting the various plots. The shapes, however, at times looked quite different in the timbregrams for example, but the overall timbral trajectories usually gravitated to the same conclusions. The same was also true when changing the number of clusters. In general, more clusters gave finer detail in grouping subtleties in the timbre space, whereas smaller number of clusters merged closely spaced clusters into a "supercluster." This is evident in Figure 9, where the ultimate section of the piece becomes one large cluster extending vertically (amplitude) when employing 5 clusters.

5. SUMMARY AND FUTURE WORK

5.1 Summary

In this paper we presented a new software system for assisting analysis of electro-acoustic music with particular emphasis on timbre. We described the functionalities of the toolbox, some of the feature extraction algorithms, the timbre space display interface, real-time possibilities using EASY, conducted preliminary analysis of two musical examples, and discussed pattern recognition modules to help reveal structural elements of an audio signal. The system has been designed with ease of use in mind by providing a "click-and-go" interface while at the same time offering advanced options for more detailed parametric control.

5.2 Future Work

The current version of the EASY Toolbox already includes 26 features but we foresee that more features, especially those that are specific to electro-acoustic music will be encountered in the future as we further develop this system. To facilitate adding new features we plan on providing a template for third party development. We plan to further extensively test and use the EASY Toolbox for analyzing a number of classic electro-acoustic works and expect to report our findings in the near future.

One very interesting and potentially exciting area that could provide promising application for EASY is exploiting more pattern recognition techniques on feature vectors to analyze for "horizontal" and "vertical" relationships and correlations in a given audio signal. That is, analyzing and displaying feature trajectories and patterns not only by comparing frames as one unit but also analyzing the vertical relationships vs. time as shown in Figure 9.
This could be very useful in displaying detailed relationships between frames, sections, motifs, formal structures, referential cues, and many other patterns that can provide insights into the music under scrutiny. One way of implementing such a feature would be using labels to display various icons in the time/frequency-domain canvases and timbregram, which will further allow for annotation possibilities.

Another area that we are interested in exploring is the literature concerning cognitive studies especially those that are related to mood and sound [5, 15]. We are not explicitly interested in measuring mood per se but we would also like to examine other angles to help extract perceptual and cognitive dimensions from the music that is being analyzed.

On top of providing analysis results from feature vectors, we also plan on offering supplementary cultural information acquired from the Internet via search engines and online digital libraries. One approach is using search strings as implemented in jWebMiner [9], which is a software package for extracting cultural features from the web using hit counts. Current MIR technologies such as fingerprinting, artist identification, and genre classification are used for automatically recommending similar musical styles, composers, and artists. Although these technologies have not been specifically applied to "music analysis" software systems that we know of, we foresee great potential in incorporating and exploiting such technologies not just for electro-acoustic music alone, but also for musical research, musicological studies, pedagogy, and composition in general. It is not difficult to imagine being able to have easy access to supplementary information such as scores, program notes, composer/performer/"machine" biographical information, graphics/pictures/videos, or any other related materials/media at one's fingertips and at the click of one button.

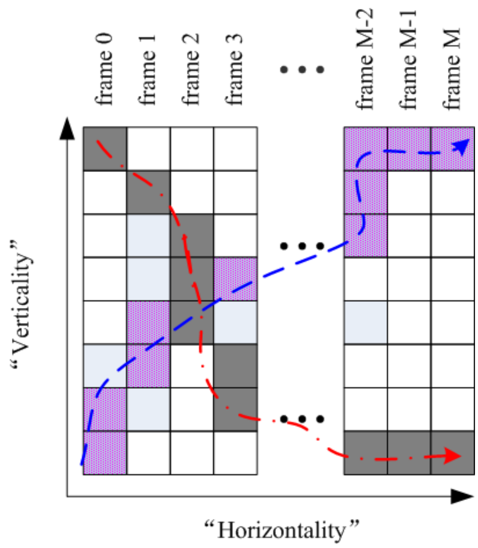

Figure 10. Verticality AND horizontality

Although the current software version is already a stand-alone MATLAB® application and can run on any machine that has the MATALB® run-time library, we plan on porting it to faster and more efficient compiler-based platforms like Cocoa.

6. REFERENCES

[1] Cannam, C., Landone C., Sandler M., Bello J., "The Sonic Visualizer: A Visualization Platform for Semantic Descriptors", *Proceedings of the International Conference on Music Information Retrieval 2006*, Victoria Canada

[2] Cao, C., Li M., Liu J., Yan, Y., "Singing Melody Extraction in Polyphonic Music by Harmonic Tracking", *Proceedings of the International Conference on Music Information Retrieval 2007*, Vienna, Austria.

[3] Kornstädt, A., "The JRing System for Computer-Assisted Musicological Analysis", *Proceedings of the International Conference on Music Information Retrieval 2001*, Indiana, USA

[4] Kuuskankare, M., Laurson, M., "Vivo - Visualizing Harmonic Progressions and Voice-Leading in PWGL", *Proceedings of the International Conference on Music Information Retrieval 2007*, Vienna, Austria.

[5] Li, T., Ogihara, M., "Detecting emotion in music", *Proceedings of the International Conference on Music Information Retrieval 2003*, Washington D.C., USA

[6] Lubbers, D., "SoniXplorer: Combining Visualization and Auralization for Content-Based Exploration of Music Collections", *Proceedings of the International Conference on Music Information Retrieval 2005*, London, U.K.

[7] McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. 1995. *Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions, Specificities, and Latent Subject Classes*, Psychological Research 58, 177 - 192.

[8] McEnnis D., McKay C., Fujinaga I., Depalle P., "jAudio: An Feature Extraction Library", *Proceedings of the International Conference on Music Information Retrieval 2005*, London, U.K.

[9] McKay, C., Fujinaga, I., "jWebMiner: A Web-based Feature Extractor", *Proceedings of the International Conference on Music Information Retrieval 2007*, Vienna, Austria.

[10] Park T. H., Biguenet J., Li Z., Richardson C., Scharr T., "Feature Modulation Synthesis", *Proceedings of the International Computer Music Conference 2007*, August, 2007, Copenhagen, Denmark.

[11] Park, T. H. *"Salient Feature Extraction of Musical Instrument Signals"*, Dartmouth College, M.A. Dissertation, 2000.

[12] Park, T. H., Z. Li, Biguenet J., "Not Just More FMS: Taking It To The Next Level", Proceedings of the 2008 ICMC, Belfast, Ireland.

[13] Simoni, M., "Analytical Methods of Electroacoustic Music", Routledge, 2006, Ch. 8, pp. 187 – 238.

[14] Sjölander, K., Beskow, J., "Wavesurfer – An Open Source Speech Tool", *Proceedings of ICSLP 2000*, Beijing, China

[15] Thayer, R.E. 1989. *The Biopsychology of Mood and Arousal*. New York: Oxford University Press

[16] Tzanetakis, G., Essel, G. Cook, P., "Musical genre classification of audio signals", *Proceedings of ISMIR 2001*, Indiana, USA

[17] Tzanetakis, G., Cook, P., "MARSYAS: a framework for audio analysis", *Organised Sound*, Vol. 4 , Issue 3, 1999, Cambridge University Press

[18] Lartillot, O., Toiviainen, P. "A MATLAB TOOLBOX FOR MUSICAL FEATURE EXTRACTION FROM AUDIO", *Proceedings of the 10th Int. Conference on Digital Audio Effects (DAFx-07)*, Bordeaux, France.