NOT JUST PRETTIER:  
FMS TOOLBOX MARCHES ON

Tae Hong Park, Zhiye Li
Tulane University  
Music Department  
New Orleans, LA, USA  
{park, zli3}@tulane.edu

ABSTRACT
In this paper we present current development and updates in the FMS (Feature Modulation Synthesis) Toolbox. The FMS Toolbox is a fully functional standalone GUI-based synthesis-by-analysis software system developed in MATLAB® using MIR (Music Information Retrieval) techniques for modulating timbral dimensions of sound objects. An overview of newly developed feature extraction algorithms and their corresponding re-synthesis engines, cross-synthesis methodologies, feature-specific sound morphing algorithms, the timbre space analysis /display/modulation system, 3D editing/displaying, and 3D exploration environments are given.

1. INTRODUCTION
Some of the most important and representative parameters in musical sounds are pitch, dynamics, duration, and timbre. For the first three parameters, we somewhat have a decent grasp of the theories and understanding surrounding them. For timbre, however, perhaps largely due to its multidimensional structure and complexity, our knowledge of its theories and concepts are rather less well understood. However, much advancement in this area, especially accelerated by active research in MIR in recent years, we have closed the gap between the known and unknowns of musical timbre, especially in the realm of research concerning automatic timbre recognition and sound identification. Feature Modulation Synthesis (FMS) fully exploits the research and findings in timbre research. It is a software system for sound modification via analysis, modulation, and re-synthesis of sonic feature vectors [4] and concentrates on musical timbre and salient audio feature vectors. It is a system that allows users to directly engage, modulate, and experiment with perceptually relevant dimensions not only to provide tools for modulating timbral features but also provide the user an environment for sonic exploration and study. The motivation for developing such a system can be attributed to the somewhat lack of similar software applications – software that focus on synthesis by means of perceptual aspects of sound. By exploiting MIR research and utilizing audio descriptors we thus provide a tool that allows the user the possibility for flexible manipulation and alteration of timbral dimensions.

The current FMS Toolbox includes 26 time and frequency-domain feature modulation modules providing deep insights into sounds from different feature angles. FMS also offers numerous options to view, edit, and modify sound objects which can be easily used by musicians, researchers, and “non-experts” alike. With the aim of making the modulation interface more intuitive and easy-to-use, we have developed 3D displays and modulation interfaces for most of the modules using the MATLAB® GUI API. Other updates include the 3D timbre space [9] display system for modulation and analysis, a real-time timbre vector tracking feature (using the MATLAB® Data Acquisition Toolbox), and making FMS a standalone program that can be run on any machine that has the MATLAB® runtime library.

2. FEATURE VECTOR DEVELOPMENT
The newly developed feature modulation include MFCC, spectral roll-off, spectral smoothness, critical band, low energy ratio, crest factor, release time, sound field vector and dynamic tightness. Each of these features has an accompanying modulation and re-synthesis engine. From the list of 26 features, zero-crossing rate and modality/harmonicity/noisiness are only used for analysis and do not have corresponding re-synthesis module at this time.

2.1. Dynamic Tightness and Low Energy Ratio
A new feature that we call dynamic tightness is shown in Figure 1 and 2. It is a feature which provides insights into the “tightness” in dynamics in such a way that the analysis output displays a histogram relationship between vacant discrete amplitude slots vs. amplitude slots occupied in a given windowed frame. Figure 1 shows the “dynamic roll-off” amplitude boundary in the middle (90% of energy in this example) and at the left-hand side the frame-by-frame overall tightness measure is shown. This feature is modulated via dynamic compression and expansion techniques.
Figure 1. Example of Blues (Electric) Guitar Riff.

Figure 2 shows an electric bass sample run through a dynamic expander effects processor. We can clearly see the reduction in dynamic tightness where the expanded region is pushed towards the zero amplitude values leaving a “drained pool” so-to-speak between the expander threshold region and lowest amplitude region.

Figure 2. Expanded Electric Bass Slide.

Low energy ratio of a signal is calculated as the energy ratio of the low passed version of the signal versus the original one. Low energy ratio modulation and re-synthesis is accomplished in time-domain by applying input signal $x$ to two filters, a LPF and a HPF sharing the same cutoff frequency. The resulting signals are multiplied with the energy weighting parameters $\alpha$ and $\beta$ ($\alpha = 1/\beta$) before being summed to render the new signal. The input and output energy levels are kept the same by satisfying Equation (1) where $x$ is the input signal, $x_L$ the low passed signal, and $x_H$ the high passed signal. The low energy ratio of the synthesized signal can be expressed as Equation (2).

$$\|x\| = \|\alpha x_L + \beta x_H\|$$  \(1\)

$$\text{LER'} = \frac{\|\alpha x_L\|}{\|x\|} = \frac{\|\alpha x_L + \beta x_H\|}{\|\alpha x_L + \beta x_H\|} = \frac{\alpha^2 \|x_L\| + \beta^2 \|x_H\|}{\alpha^2 \|x_L\| + \beta^2 \|x_H\|}$$  \(2\)

With the 3D modulation interface, users are able to control the cutoff frequency of the filters as well as the weighting parameters while viewing the change of the low energy ratio feature vector in a separate window. The following figure is a snapshot of the LER modulation interface.

Figure 3. Low Energy Ratio Modulation

2.2. Spectral Smoothness and Spectral Roll-off

The spectral smoothness feature refers to the smoothness or irregularity of the spectrum. To modulate the smoothness of a spectrum, that is, to make a spectrum smoother or rougher, a low pass and high pass filter is applied to the spectrum itself as if it were a time signal. The filtered signal is then shifted by the amount of filter delay to synchronize with the original signal and finally followed by normalization.

Figure 4. Spectral Roll-off Modulation

The spectral roll-off feature is modulated by directly addressing the roll-off frequency ($f_{ro}$) – moving $f_{ro}$ towards the DC or the Nyquist frequency to shift the energy distribution. In the modulation process, to keep the entire energy of the signal unchanged, the boosted/attenuated energy located on each side of the roll-off frequency is equally re-distributed over the linear frequency axis.
Cross-fading is used to smooth the transition region of the spectrum. The 3D roll-off modulation interface allows users to manipulate the roll-off point of each frame individually, as shown in the following illustration.

2.3. Additional Features

The MFCC modulation and re-synthesis algorithm provides another way of modulating perceptual audio features in the frequency-domain. Since the majority of the spectral information is contained in the first several coefficients, modification of these coefficients will have substantial impact on the re-synthesized sound. Other features that we have implemented include the crest factor, release-time, sound field vector, and critical band. For crest factor modulation we keep \( |x|_{\text{peak}} \) (x is time signal) unchanged for each frame and alter \( x_{\text{rms}} \). The RMS energy is increased and decreased via dynamic compression and expansion techniques followed by normalization similar to the approach taken in dynamic tightness modulation. The release-time is altered by first finding the release start point and end point for a given signal envelope and then applying phase vocoding techniques for time-shifting. The sound field vector feature (developed by Travis Scharr while at Tulane University) is the vector sum of the energy in each audio channel. This feature is modulated by creating and moving a node to a desired sound field location in a 3-dimensional space. Critical band modulation is used to change the energy in each frequency band/channel defined by its center frequency and bandwidth.

3. CROSS-SYNTHESIS AND MORPHING

To make the FMS system even more effective as a compositional tool, we have developed cross-synthesis and morphing modules as describe in this section.

3.1. Cross-Synthesis

In cross-synthesis, we extract a feature from one sound source \( (x_a) \) and apply the same feature to another sound object \( (x_b) \) for modulation. The target sound object is then re-synthesized to render the cross-synthesis result as shown in Figure 5.

![Cross-synthesis Modulation Flow Chart](image)

Figure 5. Cross-synthesis Modulation Flow Chart

When using cross-synthesis, the feature vectors for source \( v_a(k) \) and target \( v_b(k) \) are of the same type.

3.2. Morphing

Unlike cross-synthesis, in morphing, a smooth modulation from sound object \( x_a \) to sound object \( x_b \) occurs. This concept is shown in Figure 6 for a single feature vector example. The basic morphing algorithm is straightforward – we simply cross-fade between the source feature vector \( v_a(k) \) and the target vector \( v_b(k) \) where \( v_a \) and \( v_b \) refer to the feature vector \( k \) for sound object \( x_a \) and \( x_b \). In general, morphing entails applying multiple features, however.

![FMS Morphing Interface](image)

Figure 6. FMS Morphing Interface

4. INTERFACE UPDATES

The feature modules developed in the previous version of FMS [4] have been enhanced to incorporate a 3D display and modulation environment whenever possible – this is one of the core interface updates of the FMS software system. The current FMS Toolbox software system provides an intuitive and easy-to-use colour-coded 3D display and interactive modulation interface. It also integrates a 3D timbre space or “timbregram” display and modulation interface providing an insightful method for viewing and modifying feature vector trajectories while observing correlations between different timbral dimensions. Other updates include saving and opening sessions, selective playback in 2D and 3D canvases, scrolling/scrolling in 2D and 3D, and real-time analysis using the Data Acquisition Toolbox (DAQ) in all canvases including the timbregram canvas.

5. SUMMARY AND FUTURE WORK

In this paper we presented current development and updates in the Feature Modulation Synthesis (FMS) Toolbox software application. FMS employs concepts for
analysis, modulation, and re-synthesis of salient sonic feature vectors to render a desired target sound object exploiting MIR techniques, and further provides the user a simple and intuitive 3D display and modulation interface for sound exploration and modification. 26 feature extraction modules and their corresponding modulation and re-synthesis algorithms are implemented in this version of FMS. The FMS interface was designed with the aim of creating a simple, user-friendly, and intuitive software environment inspired by “interface-free” and touch-driven HCI paradigms. Although the current computer interaction technology, largely based on pointing devices and mouse clicks is not ideally suited for the 3D interface that FMS provides, we hope that in the near future, once multi-touch screen technology and touch-driven HCI systems become widely available, the “hands-on” interaction with FMS will be something to look forward to.

**Figure 7.** Tellegen-Watson-Clark model for mood

A potentially exciting area for applying FMS is investigating modulation of emotional content. In this scenario we would ideally be able to alter a sound object’s emotional content by manipulating its timbral feature vectors. The research in music emotion recognition and musical mood recognition is still very much in its infant stages, however. It is a difficult subject for research as 1) the perception of emotion is intrinsically subjective, 2) it is difficult to use universal adjectives to describe emotion, and 3) it is still not understood how music actually evokes emotion. However, researchers have found some interesting results [1, 2, 5] and have also been able to loosely map descriptors of emotion on various emotion planes including the Hevner’s adjective circle, Thayer’s arousal-valence emotion plane, and Tellegen-Watson-Clark model of mood [1, 6, 7] as shown in Figure 7. Some of the features that have been found important in emotion detection have included flux, tonality, spectral roll-off, spectral centroid, SFM, and spectral crest factor [8]. We thus plan on developing a “mood space” interface which can be used to shape the dimensions of mood as a function of time and feature vectors.

Although MATLAB® is perhaps one of the best environments for research and development, due to its scripting nature; it is not ideally suited for flexible real-time control. We thus plan on porting the FMS Toolbox to a compiler-based platform such as Cocoa once the development phase reaches a stage where only minor updates are required. This is especially an exciting prospect as it will open up possibilities for application of the FMS Toolbox system to live and real-time performance situations as well as developing VST/RTAS FMS plug-ins.

6. REFERENCES


