FEATURE MODULATION SYNTHESIS (FMS)

Tae Hong Park, Jonathan Biguenet, Zhiye Li, Conner Richardson, Travis Scharr

Tulane University
Music Department
New Orleans, LA 70118

ABSTRACT
This paper introduces FMS (Feature Modulation Synthesis), a synthesis-by-analysis sound synthesis system. Unlike many of the traditional synthesis methods such as AM, FM, additive/subtractive synthesis, waveshaping, wavetable, and physical modeling which results in sound generation via sonic parameters that in essence were developed in mimicking or modeling existing target acoustic instruments, FMS takes the approach of extracting salient timbral features directly from a sound object where the end goal is the re-synthesis of an altered sound object reflecting the modulated timbral features. The FMS system comprises of analysis of a sound object and salient feature extraction; modulation of single or multiple features; and synthesizing the new sound object via the modulated features. The essential idea behind FMS is thus synthesizing a sound object by modulating and affecting only the desired timbral dimensions while ideally leaving all other timbral aspects intact giving very fine control over specific sonic attributes.

1. INTRODUCTION
The origin of the idea for FMS came about during 2003-2004 research in Music Information Retrieval and feature extraction for automatic timbre recognition studies [3] as outlined as a “future work” project evolving to the present project. The motivation for this research can be traced to the somewhat limited number of synthesis/re-synthesis tools for composers who wish to directly deal with perceptually relevant dimensions of sound and be able to control and manipulate such dimensions in the form of alteration of salient features. Classic synthesis algorithms commonly generate sound from scratch or by using sine tones or sculpting noisy signals via control parameters that are often abstract and seemingly have little relationship to perceptually relevant nomenclatures. FMS tries to alleviate some of this disconnect between abstract control parameters and perceptually relevant ones by facilitating timbral control via modulation of salient features. The architecture follows a synthesis-by-analysis model with salient feature extraction, feature modulation, and synthesis modules as shown in figure 1.

\[ s \rightarrow \text{feature extraction} \rightarrow \text{feature modulation} \rightarrow \text{synthesis} \rightarrow s' \]

Figure 1. Basic FMS architecture

where \( s \) is the input sound object, \( s' \) synthesized result, \( v \) and \( v' \) the extracted and modulated feature vectors respectively. The fundamental approach to sound synthesis in FMS is to analyze a sound object for salient features, modulate a select number of these feature vectors, and synthesize a new altered sound object while keeping all other timbral features untouched. Although modulating one specific timbral dimension ideally should not affect any other features, in practice, some artifacts do occur and are perhaps inevitable to a certain degree. This synthesis method is not a completely unique one in a sense that FM synthesis, granular synthesis, or subtractive synthesis is and includes general ideas that touch upon sound morphing and spectral shaping research [2, 7, 8].

2. FEATURE EXTRACTION
The current set of implemented feature sets for FMS include amplitude envelope detection using RMS (Root Mean Square) and windowing, spectral envelope, harmonic expansion/compression [3], inharmonicity, spectral spread, spectral flux, spectral centroid, and spectral jitter/shimmer. The more general feature spectral envelope is also included which is a three-dimensional envelope representation of a signal in the frequency-domain with frequency, magnitude, and time axes. The frequency-domain analysis is achieved using Short-Time Fourier Transform. Harmonic analysis is conducted via a combination of harmonic product spectrum [1], autocorrelation [4], mode analysis, and other custom algorithms.

3. FEATURE MODULATION AND SYNTHESIS
Feature modulation in FMS is divided into time-domain and frequency-domain processes where most occur in the latter using STFT. There are two main approaches used for frequency-domain manipulation that are common for each modulation module – convolution based modification of the spectrum and direct manipulation of each frequency bin using an interpolated spectral envelope as reference.

3.1. Amplitude Envelope
The amplitude envelope feature is modulated by computing the product of the time-domain signal \( s \) and the envelope shaper \( v_{shaper} \) where the RMS amplitude envelope of \( s \) is used as a reference as shown in figure 2. The envelope shaper \( v_{shaper} \) is formed by modulating a copy of the original amplitude envelope of \( s \) via addition of nodes and computing the difference between the modulated envelope and original envelope \( Env_s \) as shown in eq. 1 and 2. The synthesized output is then computed via eq. 3.
The desired centroid is reached by finding the appropriate “see-saw” in an iterative manner using the method of gradient descent [5]. The gradient descent takes the usual form shown in eq. 8 with error \( E \) (eq. 9) between the actual centroid and desired centroid, and step size \( \eta \).

\[
u[n] = u[n-1] - \eta \frac{\partial E}{\partial u} \quad (8)
\]

\[
E = |SC[X'] - \text{SC}_{\text{desired}}| \quad (9)
\]

### 3.2. Spectral Centroid

The approach in modulating the spectral centroid was inspired by the mechanics of a “see-saw.” That is, more energy towards the lower frequencies will result in the see-saw being tilted towards the DC and more energy in the high frequency area will result in tilting towards the Nyquist. The see-saw is thus initially balanced at Nyquist/2 frequency and tilts either towards the DC or Nyquist depending on the desired centroid location during modulation. Similar to eq. 3 a multiplication of the “see-saw” and the spectrum occurs in the frequency-domain resulting in convolution as seen in eq. 5.

\[
X' = X \cdot V_{\text{see-saw}} \quad (9, 5)
\]

The tilting of the see-saw effectively moves the centroid either towards the DC or Nyquist. Additionally, to achieve more “tilt” the see-saw is “bent” causing the line to become a polynomial shape as seen in eq. 6 and figure 3 where convolver \( V_{\text{see-saw}} \) represents the degree of bending determined via polynomial order \( Q \), scalar \( \alpha \), line \( u \), frequency bin index \( k \), and DFT length \( N \).

\[
V_{\text{see-saw}}[k] = \sum_{q=0}^{Q} \alpha \cdot u^q \quad (6)
\]

\[
u[k] = k, \quad k = 0 \ldots N/2 \quad (7)
\]
whereas for the spread region defined by $K_l$ and $K_u$ eq. 12 is used. The combination of Envflanking and Envspread results in the modulated envelope shown in eq. 13.

Equation 13 is used to create the target modulation envelope where $k$ is the frequency bin number from DC to Nyquist.

$Env_{spread}[k] = \begin{cases} 
Env_{orig}[k], & \text{spread increase} \\
Env_{orig}[k], & \text{spread decrease}
\end{cases}$ (12)

The remaining equations are used to determine $Env_{conv}$ where the convolution product results in $X'[k]$ which is then synthesized to the time-domain signal $s'[n]$ after adjustment of spectral energy to conserve the original energy level.

$\kappa[k] = Env_{nov}[k] - Env_{orig}[k]$ (14)

$Env_{conv}[k] = \left(\kappa[k] + X[k]\right) \times \frac{X[k]}{X'[k]} = X'[k] \cdot Env_{conv}[k]$ (15)

3.4. Harmonic Expansion/Compression

Harmonic expansion/compression [4] is a feature that describes the deviation of actual harmonics from their ideal locations as depicted in figure 6.

Figure 4. Spread expansion

**Figure 5.** Spread compression

for $K_l < k \leq K_u$

$Env_{nov}[k] = Env_{spread}[k] + Env_{flanking}[k]$ (13)

Equation 17 shows the modulation algorithm where $f_n'$ denotes the extracted $m$th harmonic versus its ideal counterpart $mf_0$, where $f_0$ is the fundamental frequency, and $f_n'$ the modulated harmonic.

$$f_n' = f_n + g[m]$$ (17)

$$g[m] = m^2 \cdot df_0$$ (18)

Figure 6. Harmonic expansion/compression

Harmonic expansion and compression thus consists of harmonic analysis and extraction followed by altering a frame’s expansion or compression profile with a non-linear function $g[m]$. A non-linear $g[m]$ example is shown in eq. 18 where $\alpha$ denotes the expansion or compression characteristic. Compression occurs when $\alpha$ is negative and expansion when positive. Harmonic regions are defined by upper and lower frequency bounds determined by percentage $P_m$ of each harmonic’s magnitude. Harmonic regions that are shifted to a new region are cross-faded with the surrounding region and the “hole” left behind is attenuated and cross-faded using $P_m$ as reference as seen in figure 7.

Figure 7. Harmonic shifting
3.5. Inharmonicity

The inharmonicity for each harmonic is modulated using eq. 19

\[ f'_n = mf_0 + (f_m - mf_0)e^\beta \] (19)

where \( f'_n \) is the modulated \( m^{th} \) harmonic, \( f_0 \) is the fundamental, and \( \beta \) the inharmonicity coefficient. When \( \beta = 0 \) the degree of inharmonicity is unchanged whereas positive \( \beta \) values result in increased inharmonicity and negative values in decreased inharmonicity. Inharmonicity uses the same procedure for shifting the harmonic regions as used in harmonic expansion/compression modulation.

3.6. Spectral Shimmer and Jitter

The first step in shimmer modulation requires tracking of harmonics over each STFT frame \( (X_r) \) followed by modulation of each harmonic via function \( g[l] \) (eg. Gaussian noise) to “spread” the magnitude as shown in eq. 20 where \( l \) is the index for \( g[l] \), scalar \( \lambda \) determines its magnitude, and \( k_u \) the harmonic bin index.

\[ X_r[k_u] = X_r[k_u] \cdot (1 + \lambda \cdot g[l]) \] (20)

As seen in figure 8 shimmer amplification is realized via “magnitude interpolation” for the region defined by upper \( (k_u) \) and lower frequency \( (k_l) \) bounds. The bounds are dynamically determined as a function of the fundamental as shown in eq. 21 where \( L \) is a constant defining the neighboring area of each harmonic. Similarly, for harmonic attenuation the algorithm finds boundaries \( k_u \) and \( k_l \) which is then followed by shifting, interpolation, and cross-fading as depicted in figure 9.

\[ k_{u,l} = k_u \pm \text{round} \left( \frac{k_l}{L} \right) \] (21)

Spectral jitter is modulated in a similar fashion to shimmer as shown in figure 10, the main difference being the method of shaping the altered spectral envelope around each harmonic. As we can see in figure 10 each half of the shifted envelope around the modulated harmonic is shaped either via down-sampling or up-sampling to preserve the general shape of the envelope bounded by the upper and lower frequency limits.

3.7. Spectral Flux

The current implementation of flux utilizes comparisons of successive STFT frames yielding quantifiable changes in frequency content over time needed to measure spectral flux. This is done for all successive frames, and the values produced from each pair are averaged to get an overall spectral flux measure for an input signal. The current version is only one method of analyzing flux. We have also found that measuring this feature in sub-bands can give an alternative method of flux analysis.

As seen in figure 11 spectral flux increase (left) and decrease (right) are modulated either by increasing or decreasing the Euclidian distance for each bin between successive frames. An increase represents a movement of the current frame’s frequency components away from the previous frame’s corresponding frequency bins, and vice-versa for a decrease in flux as depicted in figure 11.
Modulation of the spectral flux for each frame follows the same procedure used in spectral spread in re-shaping each STFT frame.

3.8. Spectral Envelope

After computing the envelope for each STFT frame followed by interpolation to match the STFT frame size each of the interpolated frames are combined to render a three-dimensional spectral representation. The three-dimensional spectral envelope is modulated first by selecting an existing point on the 3-D representation and changing its magnitude value. Once the magnitude value is altered, interpolation of the surrounding regions is computed in three dimensions.

The “range,” or the “decay rate,” of interpolation for each dimension is individually controlled as seen in figure 12. The “envelope shaper” initially set to a plane with 0 magnitude (no modulation) is then used to modulate the original sound object rendering a new sound object characterized by $X'_r[k]$.

4. PRELIMINARY RESULTS AND CONCLUSION

Various sounds were used in testing the algorithms which included electric basses, flutes, percussion samples, guitar riffs, and popular music samples. FMS was also used in compositions and presented at electro-acoustic music concerts by Zhiye Li composing a tape piece and a live piece for electric bass and computer altering the harmonic expansion and contraction characteristics. Most of the FMS algorithms performed as we had anticipated although there were degrees of success that depended on the musical signals and not surprisingly DFT parameters. For example, spectral spread modulation presented an issue as some input signals exhibited a large amount of low frequency energy, thus allowing the spread only to expand by a limited amount. Forcing a wider spread requires the main energy concentration to extend further towards the Nyquist frequency than it does towards DC, and this “imbalance” caused additional distortion. In one spectral centroid modulation example we used the centroid profile of a flute to impose it onto the electric bass guitar which had some interesting musical results – the overall timbre becoming brighter as expected but at the same time adding high frequency noise thus changing the resulting timbre. Also, simple changes to the amplitude envelope in synthesizing a “flattened envelope” version of the original bass drum surprisingly exposed potential musical applications as it brought to the foreground what normally would be masked out thus ultimately altering the identity of the sound object drastically. By no means is this study the end but rather the beginning of FMS where we anticipate much improvement to the algorithms by tweaking the harmonic tracker and making it more robust to various types of signals and adding a segmentation/onset analysis module. Our plan is to have a growing set of features on top of the current basic yet important ones, including LPC for noise content analysis via residual analysis. Sound examples can be found at tmt.tulane.edu.

5. REFERENCES


