TOWARDS A GENERATIVE ELECTRONICA: HUMAN-INFORMED MACHINE TRANSCRIPTION AND ANALYSIS IN MAXMSP

Arne Eigenfeldt
School for the Contemporary Arts
Simon Fraser University
Vancouver, Canada
arne_e@sfu.ca

Philippe Pasquier
School of Interactive Arts and Technology
Simon Fraser University
Surrey, Canada
pasquier@sfu.ca

ABSTRACT

We present the initial research into a generative electronica system based upon analysis of a corpus, describing the combination of expert human analysis and machine analysis that provides parameter data for generative algorithms. Algorithms in MaxMSP and Jitter for the transcription of beat patterns and section labels are presented, and compared with human analysis. Initial beat generation using a genetic algorithm utilizing a neural net trained on the machine analysis data is discussed, and compared with the use of a probabilistic model.

1. INTRODUCTION

The goal of this research is to create a generative electronica using rules derived from a corpus of representative works from within the genre of electronica, also known as electronic dance music (EDM). As the first author and research assistants are composers, we have approached the problem as a compositional one: what do we need to know about the style to accurately generate music within it?

EDM is a diverse collection of genres whose primary function is as dance music. As such, the music tends to display several key characteristics: a constant beat, repeating rhythmic motives, four beat measures grouped in eight measure phrases. Despite these restrictions, a great deal of variety can be found in other elements within the music, and can define the different genres – the specific beat pattern, the overarching formal structure, the presence and specific locations of the breakdown (the release of tension usually associated with the drop out of the beat) – and it is these variations that create the musical interest in each track.

The primary goal of this work is creative. We are looking for methods – many of which are borrowed from MIR – that can be used both for offline analysis, as well as real-time generation in performance: we are not interested in genre recognition or classification. Our initial research is concerned with the analysis of a corpus from both a bottom-up (e.g. beat patterns) as well as top-down (e.g. formal structures) perspective, as both are defining characteristics of the style. Although some generation has already been undertaken, creative use of these analyses will be the future focus.

2. RELATED WORK

Little research has been done exclusively upon EDM, with the exception of Diakopoulos et al. [1], who used MIR techniques to classify one hundred 30-second excerpts into six EDM genres for live performance using a multi-touch surface. Gouyon and Dixon [2] approached non-electronic dance music classification using a tempo-based approach.

Automatic transcription of polyphonic music is, as Hainsworth and MacLeod suggest, one of the "grand challenges" facing computational musicology [3]. Klapuri gives an excellent overview of the problem [4].

Research specifically into drum transcription has recently been undertaken [5, 6, 7], including a very thorough overview by FitzGerald [8]. The parsing of compositions into sections from audio data has been researched as well [9, 10, 11, 12, 13].

Our research is unique in that it is carried out by composers using a combination of two of the standard live performance software tools, MaxMSP and Ableton Live, and is specific to electronic dance music.

3. DATA COLLECTION

One hundred tracks were chosen from four styles of EDM: *Breaks*, *Drum and Bass*, *Dubstep*, and *House*. The selection of these styles were based upon a number of factors: they are produced for a dance-floor audience and display clear beat patterns; the styles are clearly defined, and significantly different from one another; there is common instrumentation within each of the separate styles; they are less complex than some other styles.

Individual tracks were chosen to represent diverse characteristics and time periods, ranging from 1994-2010, with only four artists being represented twice. The tracks contain many common formal and structural production traits that are typical of each style and period.

Breaks tempi range from 120-138 beats per minute (BPM), and is derived from sped-up samples of drum breaks in Soul and Funk music which are also commonly associated with hip-hop rhythms. Off-beats occur in the hi-hat, similar to *House*, with many parts being layered to add variety. The beat is moderately syncopated, empha-

sizing two and four. Notable artists in this genre are Crystal Method, Hybrid, and Stanton Warriors.

Drum and Bass (D&B) has a tempo range of 150-180 BPM, with a highly syncopated beat containing one or more sped-up sampled breakbeats. As the name suggests, the bass line is very important, most often a very low frequency (sub-bass) sampled or synthesized timbre. Notable artists in this genre are Dom & Roland, Seba, and Klute.

Dubstep has a tempo range of 137-142 BPM, with a half-time feel that emphasizes the third beat (rather than two and four). It tends to be rather sparse, with a predominant synthesized bass line that exhibits a great deal of rhythmic low frequency modulation, known as a "wobble bass". Notable artists in this genre are Nero, Skream, and Benga.

House has a tempo range of 120-130 BPM, with a non-syncopated beat derived from Disco that emphasizes all four beats on the kick, two and four on the snare, and off-beats in the hi-hat. *House* music typically involves more complex arrangements, in order to offset the straight-forward repetitive beat, and often has Latin and Soul/R&B music influences, including sampled vocals. Notable artists in this genre are Cassius, Deep Dish, and Groove Armada.

Each recording was imported into Ableton Live[1], and, using the software's time-warp features, and adjusted so that each beat was properly and consistently aligned within the 1/16 subdivision grid. As such, each track's tempo was known, and analysis could focus upon the subdivisions of the measures.

4. BEAT ANALYSIS

Initial human analysis concentrated upon beat patterns, and a database was created that listed the following information for each work:

- tempo;
- number of measures;
- number of measures with beats;
- number of unique beat patterns;
- length of pattern (1 or 2 measures);
- average kicks per pattern;
- average snare hits per pattern;
- number of instrumental parts per beat pattern;
- number of fills.

From these, we derived the following features:

1. kick density (number of measures with beats / (pattern length / kicks per pattern));
2. snare density (number of measures with beats / (pattern length / snares per pattern));
3. density percentile (number of measures / number of measures with beats);
4. change percentile (number of measures / number of unique beat patterns).

In order to determine whether these were useful features in representing the genres, a C4 Decision-Tree (J48) classifier was run, using the features 1-4, above (note that tempo was not included, as it is the most obvious classifier). The Decision-Tree showed that snare density and kick density differentiated *Dubstep* and *House* from the other genres, and, together with the change percentile, separated *D&B* from *Breaks*. The confusion matrix is presented in Table 1. Note that differentiating *Breaks* from *D&B* was difficult, which is not surprising, given that the latter is often considered a sped-up version of the former.

	Breaks	Dubstep	D&B	House
Breaks	0.75	0.05	0.12	0.08
Dubstep	0.04	0.96	0.00	0.00
D&B	0.33	0.00	0.59	0.08
House	0.00	0.00	0.00	1.00

Table 1. Confusion matrix, in percent, for kick and snare density, and change and density percentile.

While this information could not be used for generative purposes, it has been used to rate generated patterns. Actual beat patterns were hand transcribed, a task that is not complex for human experts, but quite complex for machines.

4.1 Machine Analysis: Beat Pattern Detection

In order to transcribe beat patterns, a Max for Live[2] patch was created for Ableton Live that transmitted bar, beat, and subdivision information to Max 5[3], where the actual analysis occurred. Audio was analyzed in real-time using a 512 band FFT, with three specific frequency bands selected as best representing the spectrum of the kick, snare, and hi-hat onsets: 0-172 Hz (kick); 1 kHz-5kHz (snare); 6 kHz-16kHz (hi-hat). Frame data from these regions were averaged over 1/16th subdivisions of the measure.

Derivatives for the amplitude data of each subdivision were calculated in order to separate onset transients from more continuous timbres; negative values were discarded, and values below the standard deviation were considered noise, and discarded: the remaining amplitudes were considered onsets. The 16 value vectors were then combined into a 16×1 RGB matrix within Jitter, with hi-hat being stored in R, snare in G, and kick in B (see Figure 1).

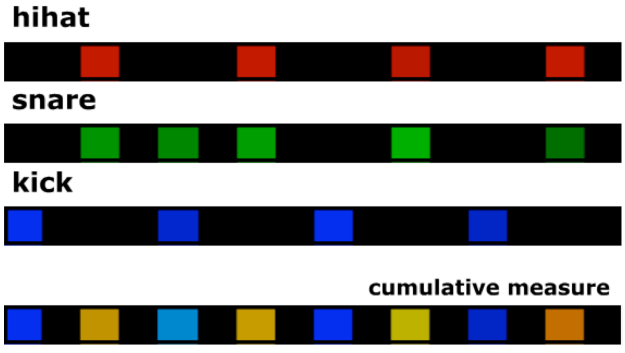

Figure 1. Example beat transcription via FFT, into 16×1 Jitter matrices. Brightness corresponds to amplitude.

[1] http://www.ableton.com/

[2] http://www.ableton.com/maxforlive

[3] http://cycling74.com/

4.1.1 Transcribing Monophonic Beat Patterns

15 drum loops were chosen to test the system against isolated, monophonic beat patterns. These patterns ranged in tempo from 120-130 BPM, and consisted of a variety of instruments, with one or more kick, snares, tuned toms, hi-hats, shakers, tambourines and/or cymbals. Table 2 describes the success rate.

Onsets	Transcriptions	Correct	Missed	False positives
389	373	0.84	0.12	.10

Table 2. Transcription success rates given 15 drum loops. Missed onsets tended to be of low amplitude, while false positives included those onsets transcribed early ("pushed beats") or late ("laid-back beats").

4.1.2 Transcribing Polyphonic Beat Patterns

Transcribing beat patterns within polyphonic music was less successful, mainly due to the variety of timbres that shared the same spectral regions. Furthermore, specific instruments, such as the bass in the low frequency, or synthesizer textures in the mid and high frequencies often used percussive envelopes that were difficult to discriminate from beat patterns (whose timbres themselves were not limited to noise).

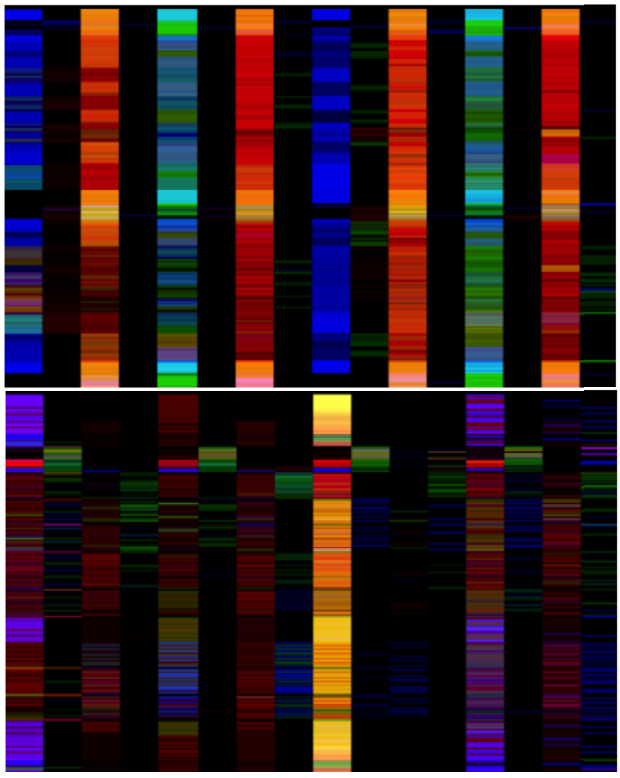

Figure 2. Two "beat fingerprints" for entire compositions: a single measure is presented as a horizontal line, with successive measures displayed top to bottom. Top, the House track "Funky Enuff": blue indicates mainly kick, red hi-hat, demonstrating the "four to the floor" with hi-hat off-beats typical of House music. Bottom, the Dubstep track "Age of Dub": yellow indicates snare and hi-hat, demonstrating the half-time feel of Dubstep.

Successive measures were accumulated into longer matrices, with the second dimension corresponding to the number of measures within the composition. This resulted in generating a track's "beat pattern fingerprint", visually displaying similarities and differences between individual compositions and genres (see Figure 2).

While the track fingerprints provided interesting visual information, a fair degree of noise remained, due to the difficulty in separating actual beats from other timbres that shared the same spectrum. For example, human analysis determined that the track in Fig. 3, top, contained only a single beat pattern, present throughout the entire duration; machine analysis calculated 31 unique kick patterns, 40 snare patterns, and 20 hi-hat patterns. As a result additional filtering was done, removing all onsets whose amplitudes were below the mean. This tended to remove false positive onsets from breakdowns.

5. BEAT GENERATION

Although generation is not the focus of our research at this time, some initial experiments have been undertaken.

5.1 Genetic Algorithm using a Neural Network

We trained a neural network (a multilayer perceptron with four nodes in the hidden layer) using patterns from the machine analysis described in Section 4.1. A fifth output was specified in which random patterns were fed in order for the neural network to be able to identify non-genre based patterns. The three individual patterns – kick, snare, hi-hat – were concatenated into a single 48 value floating point vector which was fed to the network.

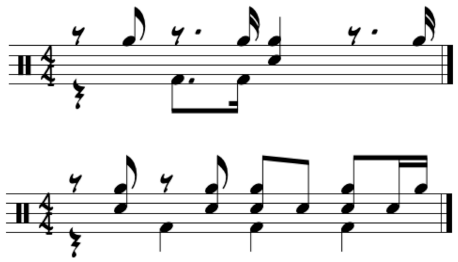

Figure 3. Example beats created by the genetic algorithm using a neural network as fitness function; top, a *Dubstep* pattern; bottom, a *House* pattern.

A genetic algorithm was created in MaxMSP in order to generate a population of beat patterns, using the trained neural network as the fitness function. Individuals, initially randomly generated, were fed to the neural network, which rated each individual as to its closeness to the patterns of a user-selected genre (similarity being determined by an algorithm that compares weighted onsets and density); individuals ranked highest within the genre were considered strong, and allowed to reproduce through crossover. Three selection methods were used, including top 50%, roulette-wheel, and tournament selection, resulting in differences in diversity in the final population. Mutation included swapping beats, and removing onsets, as randomly generated patterns tended to be much more dense than required. Using an initial population of

100 individuals, a mutation rate of 5%, and evolving 20 generations, two examples are shown in Figure 3.

5.2 Genetic Algorithm using a Probabilistic Model

A second approach was explored within the genetic algorithm – the fitness function being the Euclidean distance from prototype patterns from each genre. These prototype patterns were calculated by accumulating onsets for all measures in every analyzed track, eliminating those scores below 0.2, and generating a probabilistic model (see Figure 4).

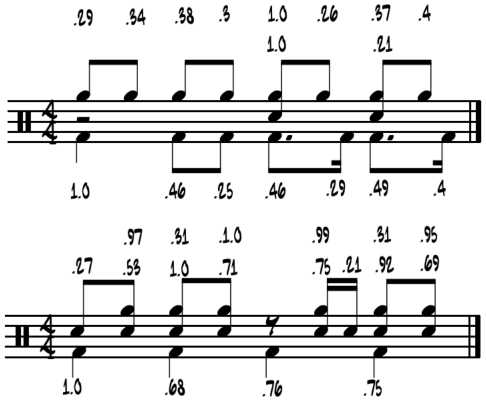

Figure 4. Proto-patterns for *Dubstep*, top, and *House*, bottom, based upon onset probabilities derived from machine analysis, with probabilities for each onset.

The machine analysis for these proto-patterns can be compared to those generated from the human analysis using the same criteria (see Figure 5). Note within *House*, only a single pattern occurs; the more active snare in the machine analysis suggests difficulty in the algorithm in separating percussive midrange timbres – such as guitar – from the snare.

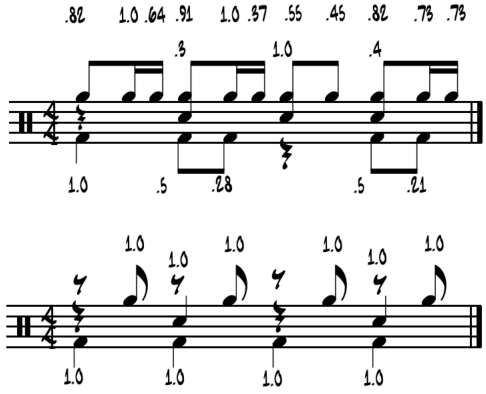

Figure 5. Proto-patterns for *Dubstep*, top, and *House*, bottom, based upon onset probabilities derived from human analysis, with probabilities for each onset.

Additional mutation functions were employed that used musical variations, in which musically similar rhythms could be substituted – see [14] for a description of this process. Example patterns evolved using this model are given in Figure 6, using an initial population of 100 individuals, a mutation rate of 5%, and evolving 20 generations.

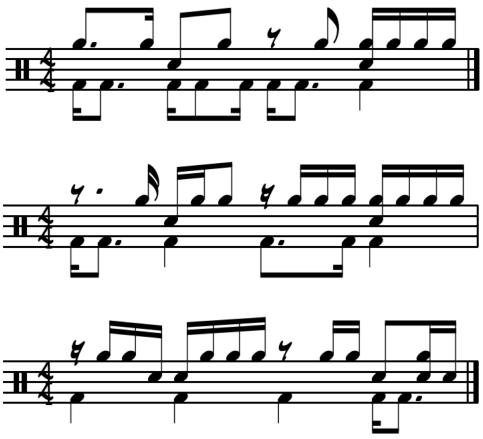

Figure 6. Three *House* patterns evolved using a genetic algorithm using machine-derived prototype patterns as fitness functions.

The use of a genetic algorithm in this second model to generate beat patterns might seem superfluous, given that a target is already extant. However, the result of the GA is a population of patterns that can be auditioned or accessed in real-time, a population that resembles the prototype target in musically interesting ways. No variation methods need to be programmed: instead, each pattern has evolved in a complex, organic way from the genre's typical patterns. Lastly, unlike generating patterns purely by the probability of specific onsets found in the proto-pattern, new onsets can appear within the population (for example, sixteenths in the *House* patterns shown in Figure 6).

6. STRUCTURAL ANALYSIS

Within Ableton Live, phrases were separated by hand into different sections by several expert listeners (however, only one listener per track):

- Lead-in – the initial section with often only a single layer present: synth; incomplete beat pattern; guitar, etc.;
- Intro – a bridge between the Lead-in and the Verse: more instruments are present than the Lead-in, but not as full as the Verse;
- Verse – the main section of the track, in which all instruments are present, which can occur several times;
- Breakdown – a contrasting section to the verse in which the beat may drop out, or a filter may remove all mid– and high–frequencies. Will tend to build tension, and lead back to the verse;
- Outro – the fade-out of the track.

The structures found within the tracks analysed were unique, with no duplication; as such, form was in no way formulaic in these examples.

Interestingly, there was no clear determining factor as to why section breaks were considered to occur at specific locations. The discriminating criteria tended to be the addition of certain instruments, the order of which was not consistent. Something as subtle as the entry of specific synthesizer timbres were heard by the experts as sectional boundaries; while determining such edges may not be a difficult task for expert human listeners, it is extremely difficult for machine analysis. Furthermore,

many of the analyses decisions were debatable, resulting from the purely subjective criteria.

6.1 Machine Analysis: Section Detection

These fuzzy decisions were emulated in the machine analysis by searching for significant changes between phrases: therefore, additional spectral analysis was done, including:

- spectral energy using a 25 band Bark auditory modeler [15], which provides the spectral energy in these perceptually significant bands;
- spectral flux, in which high values indicate significant energy difference between frames, e.g. the presence of beats;
- spectral centroid, in which high values indicate higher overall central frequency, e.g. a full timbre, rather than primarily kick and bass;
- spectral roll-off, in which high values indicate the presence of high frequencies, e.g. hi-hats.

These specific features were found to be most useful in providing contrasting information, while other analyses, such as MFCC, 24 band Mel, and spectral flatness, were not as useful. Spectral analysis was done using Malt & Jourdan's zsa externals for MaxMSP[4]

As with beat pattern analysis, these features were analyzed over 1/16 subdivisions of the measure, and stored in two separate RGB Jitter matrices, the first storing the Bark data (3.3-16 kHz in R, 450-2800 in G, 60-350 Hz in B), the second the spectral data (Flux in R, Centroid in G, Roll-off in B). See Figure 7 for examples of these spectral fingerprints.

For each of the nine vectors (three each, for Bark, Spectral, and Pattern), derivatives of amplitude differences between subdivisions of successive measures were calculated; these values were then also summed and compared to successive measures in order to discover if section changes occurred at locations other than eight bar multiples[5]. Having grouped the measures into phrases, phrase amplitudes were summed, and derivatives between phrases calculated; as with pattern recognition, negative values and values below the mean where dropped. This same mean value served as a threshold in scoring potential section breaks, as each phrase in each of the nine vectors were assigned positive scores if the difference between successive values was greater than this threshold (a new section) or below this value for subsequent differences (reinforcing the previous section change). Summing the scores and eliminating those below the mean identified virtually all section changes.

Sections were then assigned labels. Overlaying the human analysis section changes with the mean values for the nine features, it was found that breakdowns had the lowest energy in the low and high Bark regions, while verses had the highest energy in all three Bark regions (when compared to the entire track's data). See Figure 8 for an example.

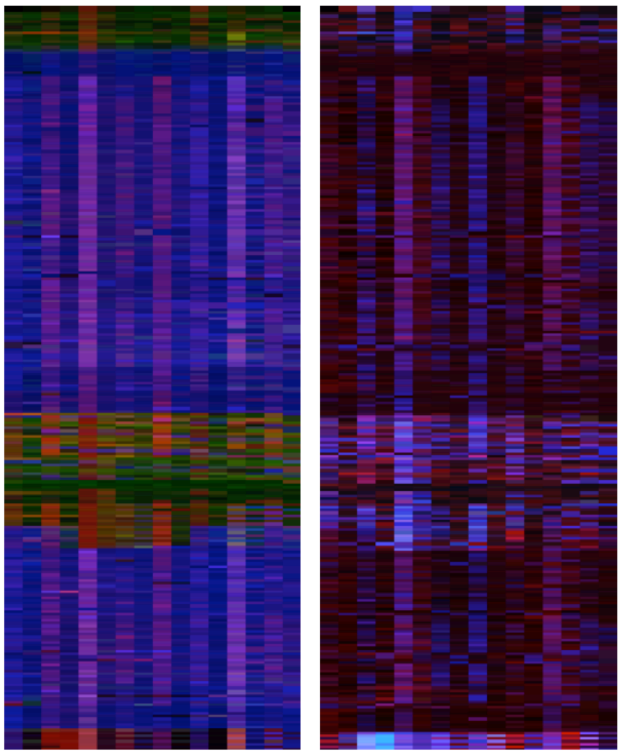

Figure 7. Spectral fingerprints for the *Breaks* track "Blowout", with Bark analysis, left, and Flux/Centroid/Roll-off, right. The section changes are clearly displayed: in this track, both low and high frequencies are removed during the breakdown, leaving primarily the midrange, shown green in the Bark analysis.

Thus, those sections whose mean values for low and high Bark regions were below the mean of all sections, were tentatively scored as breakdowns, and those sections whose mean values for all three Bark regions were above the mean of all sections, were tentatively scored as verses.

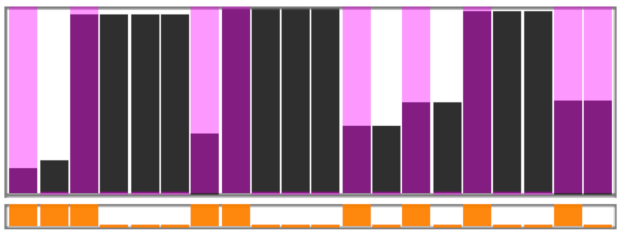

Figure 8. Mean amplitudes per section for twenty phrases for the *Breaks* track "Burma". Gray represents the normalized amplitudes over the sections, pink represents the human-analyzed section divisions, orange the machine-analyzed section divisions, including a false positive in the lead-in.

A Markov transition table was generated from the human analysis of all sections, and the machine labels were then tested against the transition table, and the scores adjusted. Thus, a low energy section near the beginning of a track (following the lead-in) may have been initially labeled a breakdown, but the transition table suggested a higher probability for a continued lead-in. After all possi-

[4] http://www.e--j.com

[5] The most formal variation occurred in *House* music, ironically considered the most static genre.

ble transitions (forwards and backwards) were taken into account, the label with the highest probability was selected.

Each phrase within 32 tracks was machine labeled for its section: Table 3 presents the results. 5 tracks that displayed unusual forms (e.g. low energy verses) in the first three genres brought the scores down significantly.

Genre	Phrases	Correct	Percentile
Breaks	174	122	0.70
D&B	264	189	0.72
Dubstep	184	124	0.67
House	152	122	0.80

Table 3. Success rate for machine labeling of sections.

7. CONCLUSIONS AND FUTURE WORK

Accurately creating music within an EDM genre requires a thorough knowledge of the model; while this knowledge may be implicit within composers, this research is the first step in making every decision based upon explicit analysis.

7.1 Improvements

Several improvements in the system are currently being made, including:

• Better beat detection involving comparing FFT matrix data between different regions of the tracks to determine similarities and differences within a phrase (i.e. comparing measure *n* and *n* + 4) and between phrases (*n* and *n* + 8).

• Incorporating fill detection to determine sectional change. Fills occur in the last 1, 2, 4, or even 8 measures of a section, and display significantly different features than the phrase, and lead to a significantly different section following.

7.2 Future Directions

Signal processing is an integral element of EDM, and we are currently involved in human analysis of typical DSP processes within the corpus, in determining representative processes, and their use in influencing structure. Similarly, pitch elements – bass lines, harmonies – are also being hand-transcribed.

Acknowledgments

Thanks to Christopher Anderson, Alan Ranta, and Tristan Bayfield for their analyses, and David Mesiha for his work with Weka. This research was funded in part by a New Media grant from the Canada Council for the Arts.

8. REFERENCES

[1] D. Diakopoulos, O. Vallis, J. Hochenbaum, J. Murphy, and A. Kapur, "21st Century Electronica: MIR Techniques for Classificaiton and Performance," Int. Soc. for Music Info. Retrieval Conf. (ISMIR), Kobe, 2009, pp. 465-469.

[2] F. Gouyon and S. Dixon, "Dance Music Classification: a tempo-based approach," ISMIR, Barcelona, 2004, pp.501-504.

[3] S. Hainsworth, M. Macleod, M. D. "The Automated Music Transcription Problem," Cambridge University Engineering Department, 2004, pp.1-23.

[4] A. Klapuri, "Introduction to Music Transcription," in A. Klapuri & M. Davy (Eds.), Signal Processing Methods for Music Transcription, 2006, pp. 3-20.

[5] J. Paulus, "Signal Processing Methods for Drum Transcription and Music Structure Analysis," PhD the- sis, Tampere University of Technology, Tampere, Fin- land, 2009.

[6] O. Gillet, G. Richard, "Automatic transcription of drum loops," Evaluation, 4, 2004, pp. 2-5.

[7] O. Gillet, G. Richard, "Transcription and Separation of Drum Signals From Polyphonic Music," IEEE Transactions On Audio Speech And Language Processing, 16(3), 2008, pp.529-540.

[8] D. Fitzgerald, "Automatic drum transcription and source separation," PhD thesis, Dublin Institute of Technology, 2004.

[9] M. Goto, "A Chorus Section Detection Method for Musical Audio Signals and Its Application to a Music Listening Station," IEEE Trans. Audio, Speech, and Lang. Proc. 14(5), 2006, pp. 1783-1794.

[10] N. Maddage, "Automatic Structure Detection for Popular Music," IEEE Multimedia, 13(1), 2006, pp.65–77.

[11] R. Dannenberg, "Listening to Naima: An Automated Structural Analysis of Music from Recorded Audio," Proc. Int. Computer Music Conf. 2002, pp.28-34.

[12] R. Dannenberg, M. Goto. "Music structure analysis from acoustic signals," in D. Havelock, S. Kuwano, and M. Vorländer, eds, Handbook of Signal Processing in Acoustics, v.1, 2008, pp. 305–331.

[13] J. Paulus, "Improving Markov Model-Based Music Piece Structure Labelling with Acoustic Information," ISMIR, 2010, pp.303-308.

[14] A. Eigenfeldt, "The Evolution of Evolutionary Software: Intelligent Rhythm Generation in Kinetic Engine," in Applications of Evolutionary Computing, Berlin, 2009, pp.498-507.

[15] E. Zwicker, E. Terhardt, "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," J. Acoustical Society of America 68(5) 1980: pp.1523-1525.