MODELING RHYTHM SIMILARITY FOR ELECTRONIC DANCE MUSIC

Maria Panteli
University of Amsterdam,
Amsterdam, Netherlands
m.x.panteli@gmail.com

Niels Bogaards
Elephantcandy,
Amsterdam, Netherlands
niels@elephantcandy.com

Aline Honingh
University of Amsterdam,
Amsterdam, Netherlands
a.k.honingh@uva.nl

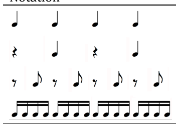
ABSTRACT

A model for rhythm similarity in electronic dance music (EDM) is presented in this paper. Rhythm in EDM is built on the concept of a 'loop', a repeating sequence typically associated with a four-measure percussive pattern. The presented model calculates rhythm similarity between segments of EDM in the following steps. 1) Each segment is split in different perceptual rhythmic streams. 2) Each stream is characterized by a number of attributes, most notably: attack phase of onsets, periodicity of rhythmic elements, and metrical distribution. 3) These attributes are combined into one feature vector for every segment, after which the similarity between segments can be calculated. The stages of stream splitting, onset detection and downbeat detection have been evaluated individually, and a listening experiment was conducted to evaluate the overall performance of the model with perceptual ratings of rhythm similarity.

1. INTRODUCTION

Music similarity has attracted research from multidisciplinary domains including tasks of music information retrieval and music perception and cognition. Especially for rhythm, studies exist on identifying and quantifying rhythm properties [16, 18], as well as establishing rhythm similarity metrics [12]. In this paper, rhythm similarity is studied with a focus on Electronic Dance Music (EDM), a genre with various and distinct rhythms [2].

EDM is an umbrella term consisting of the 'four on the floor' genres such as techno, house, trance, and the 'breakbeat-driven' genres such as jungle, drum 'n' bass, breaks etc. In general, four on the floor genres are characterized by a four-beat steady bass-drum pattern whereas breakbeat-driven exploit irregularity by emphasizing the metrically weak locations [2]. However, rhythm in EDM exhibits multiple types of subtle variations and embellishments. The goal of the present study is to develop a rhythm similarity model that captures these embellishments and allows for a fine inter-song rhythm similarity.



Rhythm in Musical Notation	Attack Positions of Rhythm	Most Common Instrumental Associations
	1/5/9/13	Bass drum
	5/13	Snare drum; handclaps
	3/7/11/15	Hi-hat (open or closed); also snare drum or synth "stabs"
	All	Hi-hat (closed)

Figure 1: Example of a common (even) EDM rhythm [2].

The model focuses on content-based analysis of audio recordings. A large and diverse literature deals with the challenges of audio rhythm similarity. These include, amongst other, approaches to onset detection [1], tempo estimation [9, 25], rhythmic representations [15, 24], and feature extraction for automatic rhythmic pattern description and genre classification [5, 12, 20]. Specific to EDM, [4] study rhythmic and timbre features for automatic genre classification, and [6] investigate temporal and structural features for music generation.

In this paper, an algorithm for rhythm similarity based on EDM characteristics and perceptual rhythm attributes is presented. The methodology for extracting rhythmic elements from an audio segment and a summary of the features extracted is provided. The steps of the algorithm are evaluated individually. Similarity predictions of the model are compared to perceptual ratings and further considerations are discussed.

2. METHODOLOGY

Structural changes in an EDM track typically consist of an evolution of timbre and rhythm as opposed to a verse-chorus division. Segmentation is firstly performed to split the signal into meaningful excerpts. The algorithm developed in [21] is used, which segments the audio signal based on timbre features (since timbre is important in EDM structure [2]) and musical heuristics.

EDM rhythm is expressed via the 'loop', a repeating pattern associated with a particular (often percussive) instrument or instruments [2]. Rhythm information can be extracted by evaluating characteristics of the loop: First, the rhythmic pattern is often presented as a combination of instrument sounds (eg. Figure 1), thus exhibiting a certain 'rhythm polyphony' [3]. To analyze this, the signal is split into the so-called rhythmic streams. Then, to describe the underlying rhythm, features are extracted for each stream based on three attributes: a) The attack phase of the onsets is considered to describe if the pattern is performed on

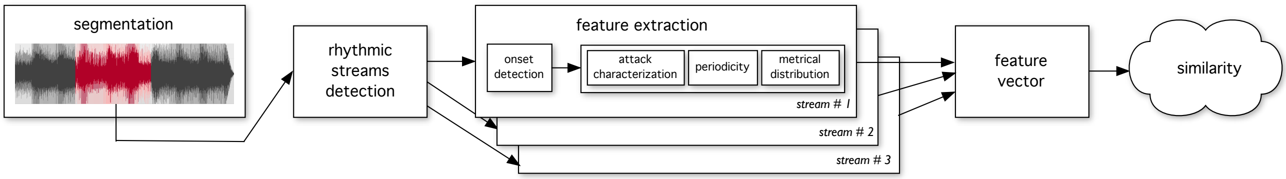


Figure 2: Overview of methodology.

percussive or non-percussive instruments. Although this is typically viewed as a timbre attribute, the percussiveness of a sound is expected to influence the perception of rhythm [16]. b) The repetition of rhythmic sequences of the pattern are described by evaluating characteristics of different levels of onsets' periodicity. c) The metrical structure of the pattern is characterized via features extracted from the metrical profile [24] of onsets. Based on the above, a feature vector is extracted for each segment and is used to measure rhythm similarity. Inter-segment similarity is evaluated with perceptual ratings collected via a specifically designed experiment. An overview of the methodology is shown in Figure 2 and details for each step are provided in the sections below. Part of the algorithm is implemented using the MIRToolbox [17].

2.1 Rhythmic Streams

Several instruments contribute to the rhythmic pattern of an EDM track. Most typical examples include combinations of bass drum, snare and hi-hat (eg. Figure 1). This is mainly a functional rather than a strictly instrumental division, and in EDM one finds various instrument sounds to take the role of bass, snare and hi-hat. In describing rhythm, it is essential to distinguish between these sources since each contributes differently to rhythm perception [11].

Following this, [15, 24] describe rhythmic patterns of latin dance music in two prefixed frequency bands (low and high frequencies), and [9] represents drum patterns as two components, the bass and snare drum pattern, calculated via non-negative matrix factorization of the spectrogram. In [20], rhythmic events are split based on their perceived loudness and brightness, where the latter is defined as a function of the spectral centroid.

In the current study, rhythmic streams are extracted with respect to the frequency domain and loudness pattern. In particular, the Short Time Fourier Transform of the signal is computed and logarithmic magnitude spectra are assigned to bark bands, resulting into a total of 24 bands for a 44.1 kHz sampling rate. Synchronous masking is modeled using the spreading function of [23], and temporal masking is modeled with a smoothing window of 50 ms. This representation is hereafter referred to as loudness envelope and denoted by L_b for bark bands $b = 1, \ldots, 24$. A self-similarity matrix is computed from this 24-band representation indicating the bands that exhibit similar loudness pattern. The novelty approach of [8] is applied to the 24×24 similarity matrix to detect adjacent bands that should be grouped to the same rhythmic stream. The peak locations P of the novelty curve define the number of the bark band that marks the beginning of a new stream, i.e., if $P = \{p_i \in \{1, \ldots, 24\} | i = 1, \ldots, I\}$ for total number of peaks I, then stream S_i consists of bark bands b given by,

$$S_i = \begin{cases} \{b | b \in [p_i, p_{i+1} - 1]\} & \text{for } i = 1, \ldots, I - 1 \\ \{b | b \in [p_I, 24]\} & \text{for } i = I. \end{cases} \tag{1}$$

An upper limit of 6 streams is considered based on the approach of [22] that uses a total of 6 bands for onset detection and [14] that suggests a total of three or four bands for meter analysis.

The notion of rhythmic stream here is similar to the notion of 'accent band' in [14] with the difference that each rhythmic stream is formed on a variable number of adjacent bark bands. Detecting a rhythmic stream does not necessarily imply separating the instruments, since if two instruments play the same rhythm they should be grouped to the same rhythmic stream. The proposed approach does not distinguish instruments that lie in the same bark band. The advantage is that the number of streams and the frequency range for each stream do not need to be predetermined but are rather estimated from the spectral representation of each song. This benefits the analysis of electronic dance music by not imposing any constraints on the possible instrument sounds that contribute to the characteristic rhythmic pattern.

2.1.1 Onset Detection

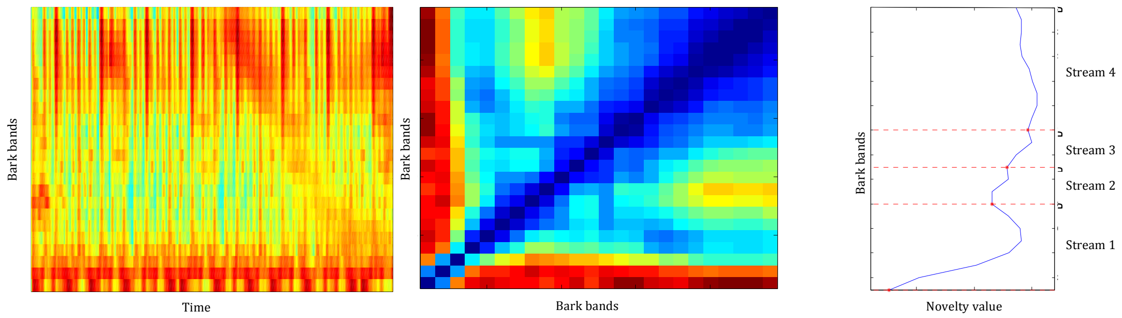
To extract onset candidates, the loudness envelope per bark band and its derivative are normalized and summed with more weight on loudness than its derivative, i.e.,

$$O_b(n) = (1 - \lambda) N_b(n) + \lambda N'_b(n) \tag{2}$$

where N_b is the normalized loudness envelope L_b, N'_b the normalized derivative of L_b, $n = 1, \ldots, N$ the frame number for a total of N frames, and $\lambda < 0.5$ the weighting factor. This is similar to the approach described by Equation 3 in [14] with reduced λ, and is computed prior summation to the different streams as suggested in [14,22]. Onsets are detected via peak extraction within each stream, where the (rhythmic) content of stream i is defined as

$$R_i = \Sigma_{b \in S_i} O_b \tag{3}$$

with S_i as in Equation 1 and O_b as in Equation 2. This onset detection approach incorporates similar methodological concepts with the positively evaluated algorithms for the task of audio onset detection [1] in MIREX 2012, and tempo estimation [14] in the review of [25].



(a) Bark-band spectrogram. (b) Self-similarity matrix. (c) Novelty curve.

Figure 3: Detection of rhyhmic streams using the novelty approach; first a bark-band spectrogram is computed, then its self-similarity matrix, and then the novelty [7] is applied where the novelty peaks define the stream boundaries.

2.2 Feature Extraction

The onsets in each stream represent the rhythmic elements of the signal. To model the underlying rhythm, features are extracted from each stream, based on three attributes, namely, characterization of attack, periodicity, and metrical distribution of onsets. These are combined to a feature vector that serves for measuring inter-segment similarity. The sections below describe the feature extraction process in detail.

2.2.1 Attack Characterization

To distinguish between percussive and non-percussive patterns, features are extracted that characterize the attack phase of the onsets. In particular, the attack time and attack slope are considered, among other, essential in modeling the perceived attack time [10]. The attack slope was also used in modeling pulse clarity [16]. In general, onsets from percussive sounds have a short attack time and steep attack slope, whereas non-percussive sounds have longer attack time and gradually increasing attack slope.

For all onsets in all streams, the attack time and attack slope is extracted and split in two clusters; the 'slow' (non-percussive) and 'fast' (percussive) attack phase onsets. Here, it is assumed that both percussive and non-percussive onsets can be present in a given segment, hence splitting in two clusters is superior to, e.g., computing the average. The mean and standard deviation of the two clusters of the attack time and attack slope (a total of 8 features) is output to the feature vector.

2.2.2 Periodicity

One of the most characteristic style elements in the musical structure of EDM is repetition; the loop, and consequently the rhythmic sequence(s), are repeating patterns. To analyze this, the periodicity of the onset detection function per stream is computed via autocorrelation and summed across all streams. The maximum delay taken into account is proportional to the bar duration. This is calculated assuming a steady tempo and $\frac{4}{4}$ meter throughout the EDM track [2]. The tempo estimation algorithm of [21] is used.

From the autocorrelation curve (cf. Figure 4), a total of 5 features are extracted:

Lag duration of maximum autocorrelation: The location (in time) of the second highest peak (the first being at lag 0) of the autocorrelation curve normalized by the bar duration. It measures whether the strongest periodicity occurs in every bar (i.e. feature value = 1), or every half bar (i.e. feature value = 0.5) etc.

Amplitude of maximum autocorrelation: The amplitude of the second highest peak of the autocorrelation curve normalized by the amplitude of the peak at lag 0. It measures whether the pattern is repeated in exactly the same way (i.e. feature value = 1) or somewhat in a similar way (i.e. feature value < 1) etc.

Harmonicity of peaks: This is the harmonicity as defined in [16] with adaptation to the reference lag l_0 corresponding to the beat duration and additional weighting of the harmonicity value by the total number of peaks of the autocorrelation curve. This feature measures whether rhythmic periodicities occur in harmonic relation to the beat (i.e. feature value = 1) or inharmonic (i.e. feature value = 0).

Flatness: Measures whether the autocorrelation curve is smooth or spiky and is suitable for distinguishing between periodic patterns (i.e. feature value = 0), and non-periodic (i.e. feature value = 1).

Entropy: Another measure of the 'peakiness' of autocorrelation [16], suitable for distinguishing between 'clear' repetitions (i.e. distribution with narrow peaks and hence feature value close to 0) and unclear repetitions (i.e. wide peaks and hence feature value increased).

2.2.3 Metrical Distribution

To model the metrical aspects of the rhythmic pattern, the metrical profile [24] is extracted. For this, the downbeat is detected as described in Section 2.2.4, onsets per stream are quantized assuming a $\frac{4}{4}$ meter and 16-th note resolution [2], and the pattern is collapsed to a total of 4 bars. The latter is in agreement with the length of a musical phrase in EDM being usually in multiples of 4, i.e., 4-bar, 8-bar, or 16-bar phrase [2]. The metrical profile of a given stream is thus presented as a vector of 64 bins (4 bars $\times$ 4 beats $\times$ 4 sixteenth notes per beat) with real values ranging between 0 (no onset) to 1 (maximum onset strength) as shown in Figure 5. For each rhythmic stream, a metrical pro-

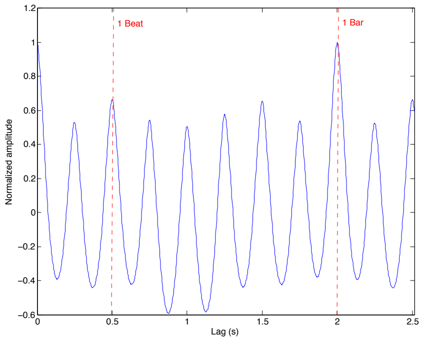


Figure 4: Autocorrelation of onsets indicating high periodicities of 1 bar and 1 beat duration.

	Bar 1	Bar 2
	♩ ♩ ♩ ♩	♩ ♩ ♩ ♩
Stream 1	1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0	1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0
Stream 2	0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0	0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0
Stream 3	0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0	0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0
Stream 4	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

Figure 5: Metrical profile of the rhythm in Figure 1 assuming for simplicity a 2-bar length and constant amplitude.

file is computed and the following features are extracted. Features are computed per stream and averaged across all streams.

Syncopation: Measures the strength of the events lying on the weak locations of the meter. The syncopation model of [18] is used with adaptation to account for the amplitude (onset strength) of the syncopated note. Three measures of syncopation are considered that apply hierarchical weights with, respectively, sixteenth note, eighth note, and quarter note resolution.

Symmetry: Denotes the ratio of the number of onsets in the second half of the pattern that appear in exactly the same position in the first half of the pattern [6].

Density: Is the ratio of the number of onsets over the possible total number of onsets of the pattern (in this case 64).

Fullness: Measures the onsets' strength of the pattern. It describes the ratio of the sum of onsets' strength over the maximum strength multiplied by the possible total number of onsets (in this case 64).

Centre of Gravity: Denotes the position in the pattern where the most and strongest onsets occur (i.e., indicates whether most onsets appear at the beginning or at the end of the pattern etc.).

Aside from these features, the metrical profile (cf. Figure 5) is also added to the final feature vector. This was found to improve results in [24]. In the current approach, the metrical profile is provided per stream, restricted to a total of 4 streams, and output in the final feature vector in order of low to high frequency content streams.

2.2.4 Downbeat Detection

The downbeat detection algorithm uses information from the metrical structure and musical heuristics. Two assumptions are made:

Assumption 1: Strong beats of the meter are more likely to be emphasized across all rhythmic streams.

Assumption 2: The downbeat is often introduced by an instrument in the low frequencies, i.e. a bass or a kick drum [2, 13].

Considering the above, the onsets per stream are quantized assuming a $\frac{4}{4}$ meter, 16-th note resolution, and a set of downbeat candidates (in this case the onsets that lie within one bar length counting from the beginning of the segment). For each downbeat candidate, hierarchical weights [18] that emphasize the strong beats of the meter as indicated by Assumption 1, are applied to the quantized patterns. Note, there is one pattern for each rhythmic stream. The patterns are then summed by applying more weight to the pattern of the low-frequency stream as indicated by Assumption 2. Finally, the candidate whose quantized pattern was weighted most, is chosen as the downbeat.

3. EVALUATION

One of the greatest challenges of music similarity evaluation is the definition of a ground truth. In some cases, objective evaluation is possible, where a ground truth is defined on a quantifiable criterion, i.e., rhythms from a particular genre are similar [5]. In other cases, music similarity is considered to be influenced by the perception of the listener and hence subjective evaluation is more suitable [19]. Objective evaluation in the current study is not preferable since different rhythms do not necessarily conform to different genres or subgenres [1]. Therefore a subjective evaluation is used where predictions of rhythm similarity are compared to perceptual ratings collected via a listening experiment (cf. Section 3.4). Details of the evaluation of rhythmic stream, onset, and downbeat detection are provided in Sections 3.1 - 3.3. A subset of the annotations used in the evaluation of the latter is available online [2].

3.1 Rhythmic Streams Evaluation

The number of streams is evaluated with perceptual annotations. For this, a subset of 120 songs from a total of 60 artists (2 songs per artist) from a variety of EDM genres and subgenres was selected. For each song, segmentation was applied using the algorithm of [21] and a characteristic segment was selected. Four subjects were asked to evaluate the number of rhythmic streams they perceive in each segment, choosing between 1 to 6, where rhythmic stream was defined as a stream of unique rhythm.

For 106 of the 120 segments, the subjects' responses' standard deviation was significantly small. The estimated number of rhythmic streams matched the mean of the subject's response distribution with an accuracy of 93%.

[1] Although some rhythmic patterns are characteristic to an EDM genre or subgenre, it is not generally true that these are unique and invariant.

[2] `https://staff.fnwi.uva.nl/a.k.honingh/rhythm_similarity.html`

3.2 Onset Detection Evaluation

Onset detection is evaluated with a set of 25 MIDI and corresponding audio excerpts, specifically created for this purpose. In this approach, onsets are detected per stream, therefore onset annotations should also be provided per stream. For a number of different EDM rhythms, MIDI files were created with the constraint that each MIDI instrument performs a unique rhythmic pattern therefore represents a unique stream, and were converted to audio.

The onsets estimated from the audio were compared to the annotations of the MIDI file using the evaluation measures of the MIREX Onset Detection task [3]. For this, no stream alignment is performed but rather onsets from all streams are grouped to a single set. For 25 excerpts, an F-measure of 85%, presicion of 85%, and recall of 86% are obtained with a tolerance window of 50 ms. Inaccuracies in onset detection are due (on average) to doubled than merged onsets, because usually more streams (and hence more onsets) are detected.

3.3 Downbeat Detection Evaluation

To evaluate the downbeat the subset of 120 segments described in Section 3.1 was used. For each segment the annotated downbeat was compared to the estimated one with a tolerance window of 50 ms. An accuracy of 51% was achieved. Downbeat detection was also evaluated at the beat-level, i.e., estimating whether the downbeat corresponds to one of the four beats of the meter (instead of off-beat positions). This gave an accuracy of 59%, meaning that in the other cases the downbeat was detected on the off-beat positions. For some EDM tracks it was observed that high degree of periodicity compensates for a wrongly estimated downbeat. The overall results of the similarity predictions of the model (Section 3.4) indicate only a minor increase when the correct (annotated) downbeats are taken into account. It is hence concluded that the downbeat detection algorithm does not have great influence on the current results of the model.

3.4 Mapping Model Predictions to Perceptual Ratings of Similarity

The model's predictions were evaluated with perceptual ratings of rhythm similarity collected via a listening experiment. Pairwise comparisons of a small set of segments representing various rhythmic patterns of EDM were presented. Subjects were asked to rate the perceived rhythm similarity, choosing from a four point scale, and report also the confidence of their rating. From a preliminary collection of experiment data, 28 pairs (representing a total of 18 unique music segments) were selected for further analysis. These were rated from a total of 28 participants, with mean age 27 years old and standard deviation 7.3. The 50% of the participants received formal musical training, 64% was familiar with EDM and 46% had experience as EDM musician/producer. The selected pairs were rated between 3 to 5 times, with all participants reporting confidence in their rating, and all ratings being consistent, i.e., rated similarity was not deviating more than 1 point scale. The mean of the ratings was utilized as the ground truth rating per pair.

[3] www.MIREX.org

r	p	features
-0.17	0.22	attack characterization
0.48	0.00	periodicity
0.33	0.01	metrical distribution excl. metrical profile
0.69	0.00	metrical distribution incl. metrical profile
0.70	0.00	all

Table 1: Pearson's correlation r and p-values between the model's predictions and perceptual ratings of rhythm similarity for different sets of features.

For each pair, similarity can be calculated via applying a distance metric to the feature vectors of the underlying segments. In this preliminary analysis, the cosine distance was considered. Pearson's correlation was used to compare the annotated and predicted ratings of similarity. This was applied for different sets of features as indicated in Table 1.

A maximum correlation of 0.7 was achieved when all features were presented. The non-zero correlation hypothesis was not rejected ($p > 0.05$) for the attack characterization features indicating non-significant correlation with the (current set of) perceptual ratings. The periodicity features are correlated with $r = 0.48$, showing a strong link with perceptual rhythm similarity. The metrical distribution features indicate a correlation increase of 0.36 when the metrical profile is included in the feature vector. This is in agreement with the finding of [24].

As an alternative evaluation measure, the model's predictions and perceptual ratings were transformed to a binary scale (i.e., 0 being dissimilar and 1 being similar) and their output was compared. The model's predictions matched the perceptual ratings with an accuracy of 64%. Hence the model matches the perceptual similarity ratings at not only relative (i.e., Pearson's correlation) but also absolute way, when a binary scale similarity is considered.

4. DISCUSSION AND FUTURE WORK

In the evaluation of the model, the following considerations are made. High correlation of 0.69 was achieved when the metrical profile, output per stream, was added to the feature vector. An alternative experiment tested the correlation when considering the metrical profile as a whole, i.e., as a sum across all streams. This gave a correlation of only 0.59 indicating the importance of stream separation and hence the advantage of the model to account for this.

A maximum correlation of 0.7 was reported, taking into account the downbeat detection being 51% of the cases correct. Although regularity in EDM sometimes compensates for this, model's predictions can be improved with a more robust downbeat detection.

Features of periodicity (Section 2.2.2) and metrical distribution (Section 2.2.3) were extracted assuming a $\frac{4}{4}$ meter, and 16-th note resolution throughout the segment. This is generally true for EDM, but exceptions do exist [2]. The assumptions could be relaxed to analyze EDM with ternary divisions or no $\frac{4}{4}$ meter, or expanded to other music styles with similar structure.

The correlation reported in Section 3.4 is computed from a preliminary set of experiment data. More ratings are currently collected and a regression analysis and tuning of the model is considered in future work.

5. CONCLUSION

A model of rhythm similarity for Electronic Dance Music has been presented. The model extracts rhythmic features from audio segments and computes similarity by comparing their feature vectors. A method for rhythmic stream detection is proposed that estimates the number and range of frequency bands from the spectral representation of each segment rather than a fixed division. Features are extracted from each stream, an approach shown to benefit the analysis. Similarity predictions of the model match perceptual ratings with a correlation of 0.7. Future work will fine-tune predictions based on a perceptual rhythm similarity model.

6. REFERENCES

[1] S. Böck, A. Arzt, K. Florian, and S. Markus. Online real-time onset detection with recurrent neural networks. In *International Conference on Digital Audio Effects*, 2012.

[2] M. J. Butler. *Unlocking the Groove*. Indiana University Press, Bloomington and Indianapolis, 2006.

[3] E. Cambouropoulos. Voice and Stream: Perceptual and Computational Modeling of Voice Separation. *Music Perception*, 26(1):75–94, 2008.

[4] D. Diakopoulos, O. Vallis, J. Hochenbaum, J. Murphy, and A. Kapur. 21st Century Electronica: MIR Techniques for Classification and Performance. In *ISMIR*, 2009.

[5] S. Dixon, F. Gouyon, and G. Widmer. Towards Characterisation of Music via Rhythmic Patterns. In *ISMIR*, 2004.

[6] A. Eigenfeldt and P. Pasquier. Evolving Structures for Electronic Dance Music. In *Genetic and Evolutionary Computation Conference*, 2013.

[7] J. Foote and S. Uchihashi. The beat spectrum: a new approach to rhythm analysis. In *ICME*, 2001.

[8] J. T. Foote. Media segmentation using self-similarity decomposition. In *Electronic Imaging*. International Society for Optics and Photonics, 2003.

[9] D. Gärtner. Tempo estimation of urban music using tatum grid non-negative matrix factorization. In *ISMIR*, 2013.

[10] J. W. Gordon. The perceptual attack time of musical tones. *The Journal of the Acoustical Society of America*, 82(1):88–105, 1987.

[11] T. D. Griffiths and J. D. Warren. What is an auditory object? *Nature Reviews Neuroscience*, 5(11):887–892, 2004.

[12] C. Guastavino, F. Gómez, G. Toussaint, F. Marandola, and E. Gómez. Measuring Similarity between Flamenco Rhythmic Patterns. *Journal of New Music Research*, 38(2):129–138, June 2009.

[13] J. A. Hockman, M. E. P. Davies, and I. Fujinaga. One in the Jungle: Downbeat Detection in Hardcore, Jungle, and Drum and Bass. In *ISMIR*, 2012.

[14] A. Klapuri, A. J. Eronen, and J. T. Astola. Analysis of the meter of acoustic musical signals. *IEEE Transactions on Audio, Speech and Language Processing*, 14(1):342–355, January 2006.

[15] F. Krebs, S. Böck, and G. Widmer. Rhythmic pattern modeling for beat and downbeat tracking in musical audio. In *ISMIR*, 2013.

[16] O. Lartillot, T. Eerola, P. Toiviainen, and J. Fornari. Multi-feature Modeling of Pulse Clarity: Design, Validation and Optimization. In *ISMIR*, 2008.

[17] O. Lartillot and P. Toiviainen. A Matlab Toolbox for Musical Feature Extraction From Audio. In *International Conference on Digital Audio Effects*, 2007.

[18] H. C. Longuet-Higgins and C. S. Lee. The Rhythmic Interpretation of Monophonic Music. *Music Perception: An Interdisciplinary Journal*, 1(4):424–441, 1984.

[19] A. Novello, M. M. F. McKinney, and A. Kohlrausch. Perceptual Evaluation of Inter-song Similarity in Western Popular Music. *Journal of New Music Research*, 40(1):1–26, March 2011.

[20] J. Paulus and A. Klapuri. Measuring the Similarity of Rhythmic Patterns. In *ISMIR*, 2002.

[21] B. Rocha, N. Bogaards, and A. Honingh. Segmentation and Timbre Similarity in Electronic Dance Music. In *Sound and Music Computing Conference*, 2013.

[22] E. D. Scheirer. Tempo and beat analysis of acoustic musical signals. *The Journal of the Acoustical Society of America*, 103(1):588–601, January 1998.

[23] M. R. Schroeder, B. S. Atal, and J. L. Hall. Optimizing digital speech coders by exploiting masking properties of the human ear. *The Journal of the Acoustical Society of America*, pages 1647–1652, 1979.

[24] L. M. Smith. Rhythmic similarity using metrical profile matching. In *International Computer Music Conference*, 2010.

[25] J. R. Zapata and E. Gómez. Comparative Evaluation and Combination of Audio Tempo Estimation Approaches. In *Audio Engineering Society Conference*, 2011.