RHYTHMIC SIMILARITY USING METRICAL PROFILE MATCHING

Leigh M. Smith

IRCAM
Analyse-Synthese Group, Paris, France
leigh.smith@ircam.fr

ABSTRACT

A method for computing the similarity of metrical rhythmic patterns is described as applied to the audio signal of recorded music. For each rhythm, a combined feature vector of metrical profile and syncopation, separated by spectral subbands, hypermetrical profile, and tempo are compared. The descriptive capability of this feature vector is evaluated by it's use in a machine learning rhythm classification task, identifying ballroom dance styles using a support vector machine algorithm. Results indicate that with the full feature vector a result of 67% is achieved. This improves on previous results using rhythmic patterns alone, but does not exceed the best reported results. By evaluating individual features, measures of metrical, syncopation and hypermetrical profile are found to play a greater role than tempo in aiding discrimination.

1. SIMILARITY OF MUSICAL RHYTHMS

A computational model of rhythmic similarity of music has applications in several fields of music research. In music information retrieval (MIR), measures of similarity matching listeners judgements promise retrieval that better matches users expectations. Identification of common rhythmic structures can be used in tracing musical sources in western and ethnological musicology. A concrete example application is "automatic DJing" [4] to create playlists that require a common mood, such as for dancing.

Palmer and Krumhansl [13] argue that pre-established mental frameworks ("schemas") for musical meter[1] are used during listening. These schemas enable robust interpretation despite sometimes contradicting, ambiguous or absent objective cues. In this study they tested the types of mental structures for meter ("perceptual hierarchies") evoked from simple event sequences. Listeners rated the relative acceptability of audible events at different locations in the metrical grid [13, Experiment 2]. They found a significant difference in performance between musicians and non-musicians, arguing that musicians hold more resilient representations of meter, which favours hierarchical subdivision of the measure, than the non-musicians.

[1] A periodic repetition of accentuation, notated in music as $\frac{4}{4}, \frac{3}{4}$ etc.

The complexity of a rhythm may contribute to how it is judged as similar to another rhythm [20, 9]. Several approaches to formalise rhythm complexity have been proposed [3, 17, 18], with experimental verification. Ladinig [9] had subjects rate sets of 2-4 musical rhythms according to complexity, chosen to compare different metrical positions. While an hierarchical, metrical structure was found in both musicians and non-musicians, serial primacy and recency effects were also found, particularly in non-musician subjects. While not the only factor, metrical structure can be seen to contribute to judgements of rhythmic similarity.

This paper proposes a model of rhythmic similarity with the aim to match human judgements when comparing two pieces of recorded music. This model represents musical concepts such as meter, syncopation, *hypermeter* (grouping over multiple measures) and tempo. These form a combined feature vector of metrical and syncopation profile for separate spectral subbands, hypermetrical profile and median tempo, described in Section 3. The applicability of the features to generate similarity measures are evaluated in Section 4 by using them for a machine learning classification task, using the well developed support vector machine (SVM) learning function in Section 5.

2. PREVIOUS WORK

Considerable work has been devoted to computational models for measuring rhythm pattern similarity that directly process musical recordings. Paulus and Klapuri [14] developed a system using normalised spectral centroid as an event onset feature. Feature vectors are matched using dynamic time warping across the duration of the audio signal. Foote et al. [4] used distances between a measure of continuous periodicity representation (the "beat spectrum") of musical pieces.

Tzanetakis and Cook [22] and Gouyon et al. [5] used beat and interval histograms that capture the relative occurrence of rhythmic periodicities. The issue of tempo dependency of these histograms is addressed by Gruhne et al. [6]. Gouyon et al. [5] adopted cepstrum like analysis of periodicities of onset intervals which effectively measures the degree of interval multiples and subdivisions.

Dixon et al. [2] use a method of averaging the onset detection function over the metrical period. Peaks in the

meter averaged onset detection function are used to identify the onsets occurring in the measure. This metrical profile is used together with rhythmic measures of swing ratio, and a syncopation measure, with machine learning classifiers. Peeters [15] uses a rhythm spectrum determined from the product of spectral and autocorrelation functions applied to the onset detection function. This produces discrimination closely matching the performance of [2, 5].

Despite reporting very good results, there are areas for improvement to these approaches. One is the issue of dimensionality of the feature vectors used. Since approaches such as beat histograms produce significant data, descriptive statistics of them are used as the features. The approach taken in this paper is to instead aim for cognitively associated representations which are interpretable in musical terms. This aims to increase the inductive bias (i.e. domain knowledge) [1] applied to the machine learning task of rhythm classification.

3. METHOD

The model of similarity is based on identification and comparison of rhythmic patterns. An approach related to that of [2] has been adopted, where a recurring metric pattern in the piece is identified. The use of high level rhythmic descriptions such as syncopation and metrical profile depends on identification of the beat period, metrical period (duration of the bar) and phase (downbeat location) from the audio signal of the musical example. This is achieved using a beat-tracker, developed by Peeters [15, 16].

The feature vector for each rhythm is then used to compute a kernel distance measure between vectors. This produces a similarity matrix of comparisons between all rhythms. The actual distance measure used is dependent on the classification algorithm used, with Euclidean and Cosine distances as the most common.

3.1. Metrical Profile

The metrical profile, indicating the relative occurrence of events in each metrical position within the measure, has been demonstrated by [13] to represent metrical structure and matches closely with listeners judgements of metrical well-formedness. The metrical profile is computed from the likelihood of an onset at each *tatum* (shortest temporal interval) within a measure. The likelihood of onsets are determined from the presence of onset detection function (ODF) energy e described in [16]. The probability of an onset o_t at each tatum location t is

$$o_t = \begin{cases} \frac{\bar{e}_t}{\bar{e}+\gamma\sigma_e+\varepsilon}, & o_t < 1 \\ 1 & o_t > 1 \end{cases} \quad (1)$$

where $\bar{e}_t$ is the mean energy of the ODF over the region of the tatum t, $\bar{e}$ and σ_e are the mean and standard deviation of the entire ODF energy respectively, ε is a small value to guard against zero $\bar{e}$, and γ is a free parameter determining the maximum number of standard deviations above the mean to assure an onset has occurred. By informal testing, $\gamma = 2$. The onset likelihoods are then used to create an histogram of the relative amplitude and occurrence at each tatum, by averaging each o_t across all measures.

To normalise for varying tempo across each piece and between pieces, the duration of each measure is derived from the beat-tracker [16]. Using the beat locations identified by the beat-tracker, each beat duration is uniformly subdivided into 1/64th notes (hemi-demi-semiquavers), that is $0 < t < 64$ for a measure of a semibreve (whole note) duration. Such a high subdivision attempts to categorise swing timing occurring within the measure and to provide sufficient resolution for accurate comparisons of metrical structure. Using the tatum duration set to equal subdivisions of each beat duration does not capture expressive timing occuring within that time period. However, the error produced from this is minimal since the expressive timing which modifies each beat and measure period is respected. The effect of this error is to blur the peak of each tatum onset. To reduce dimensionality, the metrical profile is then downsampled (by local averaging of 4 tatums) to semiquavers (1/16 notes).

3.2. Spectral Matching

In the case of the Rock/Pop idiom, where drums are commonplace, the presence of bass, snare, tom-tom drums and/or cymbals at each tatum is a salient feature used by listeners in rhythmic identification. Thus the spectral character (centroid, dispersion) of each onset contributes greatly to its perceived time-keeping role and therefore similarity judgements. To match the categorization used by listeners, rhythmic patterns need to be compared when separated by their spectral character.

This is produced by computing spectral sub-bands of the half wave rectified spectral energy. The sub-bands are computed by summing over non-overlapping frequencies:

$$F_{c,t} = \sum_{b=b_c}^{b'_c} e_{HWR}(\omega_b, t), \quad (2)$$

where $F_{c,t}$ is the spectral flux for the sub-band channel c at time t, over the spectral bands $b = [\omega_c, \omega'_c]$ of the half-wave rectified spectral energy $e_{HWR}(\omega_b, t)$ at frequency band ω_b computed as described by [16]. The sub-band channels used are $[\omega_1 = 60, \omega'_1 = 100]$, $[\omega_2 = 3500, \omega'_2 = 4000]$. These have been chosen to select representative frequency ranges (in Hz) that capture the bass and high frequency components for channels 1 and 2 respectively.

3.3. Syncopation

An early formal representation of syncopation is by Longuet-Higgins and Lee [12], that assumes the listener will attempt to interpret a rhythm according to a given meter so as to minimize syncopations. Syncopation is consequently defined by them as a beat stronger than the previous sounded note falling on a rest or tied note [12, 19]. A syncopation occurs if and only if a (sounded) note outlasts the highest-level *metrical unit* it initiates. Metrical units are defined in an hierarchy identical to Lerdahl and Jackendoff's [10] metrical hierarchy, although inverse in polarity.

This model has been shown to be well correlated to listener judgements of rhythmic complexity [19, 21]. The original model of Longuet-Higgins and Lee relies on the identification of an onset at each tatum within the measure. Onset detection in polyphonic audio is a currently unsolved research problem in MIR. Two approaches are possible, reformulating the model of Longuet-Higgins and Lee in probablistic terms, or relying on outputs of the model which may be prone to error equal to the error rate of the onset detection. Initially, the second approach has been adopted.

The same measure of onset likelihood (Equation 1) is used to compute the binary decision whether there is an onset present,

$$onset_t = \begin{cases} 1 & o_t > 0.5, \\ 0 & o_t \leq 0.5. \end{cases} \tag{3}$$

While this simplistic approach to onset detection is error prone, the syncopation is computed for each measure and averaged to compute a single syncopation metric for the song. With an average of 80 measures per Rock/Pop song, this error is not expected to bias results to a particular metrical form. Since the tatum locations that match the beat locations will always produce a syncopation measure of zero at that location, these positions are discarded to avoid feature dimensions which do not contribute to the similarity discrimination.

3.4. Hypermetrical Profile

In several genres, patterns can occur which consistently span multiple measures. The averaging of the onset detection function to compute the metrical profile will not capture such patterns. In order to compare rhythmic patterns beyond the period of the measure, the feature vector includes a profile of *hyper-meter* [11]. An hypermetrical profile is formed from an histogram of beat occurrences over a phrase spanning multiple measures.

$$H = \bar{t}_0, \bar{t}_1, \ldots, \bar{t}_i \tag{4}$$

where $\bar{t}_i$ is the tatum at location $0 < i < |H|$ averaged across the number of hypermeters of each piece.

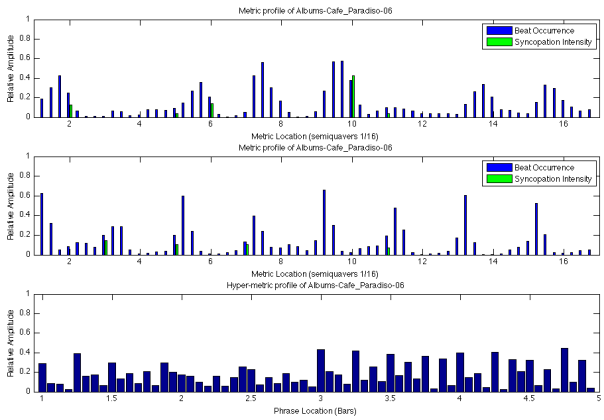

Figure 1. Metrical, syncopation and hypermetrical profiles of an example from the ballroom dancers dataset. The top and center plots indicate metrical and syncopation profiles for bass ($c = 1$) and treble ($c = 2$) subbands respectively represented in 1/64's of a measure. The lower plot indicates the hypermetrical profile for four measures, with a resolution of 1/16 of a measure.

Since the downbeat, beat and duration of the measure is known, for many genres, it is possible to make strong assumptions about higher order periodicities, for example, hypermetrical structures tend to occur over four measures. While hypermeter can not be guaranteed to be phase aligned with the meter, the use of hypermetrical profiles is designed to capture common grouping periods. When computing a distance measure, each hypermeter can be aligned by finding the maximum cross-correlation to each hypermeter it is compared to. The alignment a of the metrical profiles U and V is computed by:

$$a = |\operatorname{argmax}(U * V) - |V||, \tag{5}$$

where $|V|$ is the length of vector V. The alignment provides the rotation around the hypermetrical period of U to best match the two metrical profiles. The match between the aligned profiles p_i and p_j can then be computed by the Euclidean distance between vectors. For the evaluation using rhythm classification however, the phase of the hypermeter is currently assumed to match the downbeat of the meter, while it is not clear that the two periodicities will always synchronise, even in Rock/Pop music.

An example of a rhythm represented as metrical, syncopation and hypermetrical profiles is shown in Figure 1.

3.5. Tempo matching

It is often reported [2, 5, 15] that tempo is a significant discriminator in rhythm classification studies. An initial implementation incorporated the median computed tempo of all beats as an element in the feature vector. Future work may need to represent separate regions for periods of radically differing tempi. The best method to represent tempo vari-

	C	J	Q	R	S	T	Rec %
C	94	1	2	8	4	0	86
J	2	39	2	7	4	2	69
Q	3	1	34	5	10	1	63
R	14	6	5	40	8	8	49
S	8	5	6	7	51	3	63
T	7	6	1	8	6	50	64
Pre %	73	67	68	53	61	78	

Table 1. Confusion matrix for ballroom dancer $\frac{4}{4}$ rhythms. Columns indicate how the ground truth (rows) were classified as: *C*haChaCha, *J*ive, *Q*uickstep, *R*umba, *S*amba and *T*ango. *Pre*cision and *Rec*all are also indicated.

ation between rhythms and the impact on similarity judgements remains unresolved. For example, many cover versions of songs with the same rhythm as the original are often played at radically different tempi.

4. EVALUATION

Evaluation of rhythmic similarity remains a problem due to it's informal definition. Paulus and Klapuri [14] compared similarity measures of two patterns of the same rhythm to patterns of different rhythms (in-song vs. inter-song). While this establishes an upper bound on similarity, this approach does not measure perceived similarity between different pieces. Guastavino et al [7] compared algorithms to listener judgements of flamenco rhythms, however the rhythms were synthesized and lacking typical accompaniment (i.e melodic instruments). Several researchers [2, 5, 15] have used classification between labelled ballroom dance rhythms as an evaluation method. While this has good ecological validity, the dataset may be biased towards music which have overtly differentiated rhythms for the purpose of dancing. For example, the high discriminability by tempo may be specific to the ballroom dataset. Comparisons of pieces of music which are not so obviously rhythmically different (i.e. differing in meter or significantly in tempo) are more common in MIR or musicological tasks.

Mindful of these issues, a reduced set of the ballroom dancer dataset was used for evaluation. Since the model relies on a beat tracker which can identify the metrical period, the metrical, syncopation and hypermetrical profiles vary in length depending on the meter. Meters like $\frac{3}{4}$ and $\frac{4}{4}$ are therefore firstly classified using the meter assigned by the beat-tracker since it is such a strong differentiation. To test the capability of the similarity model alone, a reduced dataset consisting of only those pieces in $\frac{4}{4}$ and classified as such by the beat-tracker were tested. The number of rhythms in each style were then ChaChaCha (109), Jive (56), Quickstep (54), Rumba (81), Samba (80), and Tango (78), 458 pieces total of 30 seconds duration each.

Description	Features	# Correct	% Success
Full Features	122	308	67%
No Tempo	121	290	63%
No Syncopation	98	285	62%
Bass Only	94	274	60%
No Hypermeter	58	253	55%
No Meter Prof.	90	238	52%

Table 2. Results of classification by feature inclusion. The Features column indicates the number of feature attributes used in each run, percent success is the number of rhythms correctly classified compared to the total 458 rhythms.

The machine learning system Weka was used with the SVM function using sequential minimal optimization [8] for evaluation. The rhythms were classified as belonging to one of the six dance styles. A standard 10-fold cross-validation was used, where 10 randomly chosen subsets of the dataset are used with 90% of the examples used for training, remaining 10% for testing, and the final result is the average over the folds. Baseline evaluation (derived using the **ZeroR** function) is 24%.

5. RESULTS

Running the SVM on the reduced ballroom dataset produced the confusion matrix in Table 1. This corresponds to a performance of 67% correctly classified instances. In order to evaluate the contributions of each feature, the difference in performance when removing features is summarised in Table 2. It can be seen that removing the 32 metrical profile features, for both bass and treble subbands, had the greatest impact on the classification. In similar fashion, the long temporal structure of the hypermeter had a significant effect on the classification performance. The "Bass Only" features refers to removing the treble features for metrical and syncopation profiles, to determine the contribution that the subband separation makes. Interestingly, these features and the syncopation profiles had a greater impact than the tempo feature in classification.

This runs counter to the finding of [2, 5] who reported tempo being a significant classification feature. Such a finding would seem to be in part due to the AdaBoost classifier used by [2] being less able to handle the higher dimensionality produced by the full-feature vector than the SVM. AdaBoost produced significantly lower results (37%), mainly because boosting was not possible, but removal of tempo dropped performance (28%) while removal of metrical or syncopation profile produced no change in performance.

Principle components analysis (PCA) was used to reduce the dimensionality of the features, from 122 to 78 features, but this was found to not improve performance with

SVM (56%) or AdaBoost (27%). An approach of reducing dimensionality by using higher level musical representations that introduces greater inductive bias [1] seems a more advantageous approach.

Compared to the results presented by [2], when incorporating their full complement of rhythmic patterns and other features (their Table 4), the model presented here performs significantly worse. However, when compared against the rhythmic patterns used by them alone (their Table 3), the model here performs better except for Rumba patterns (49% vs. 54%). In principle, the extra features they use could also be combined with the model to improve its results, although increasing dimensionality may not guarantee this. In comparison to [15], however, the model here performs significantly worse, with the additional cost of higher dimensionality.

6. CONCLUSIONS

A model of rhythm using features drawn from music perception has been proposed as a means of computing rhythmic similarity of music recordings. It's performance has been analysed by using it in a rhythmic style classification task. Such features, when combined with a sufficiently capable machine learning algorithm (the SVM), has been demonstrated to improve on previously reported results drawn from metric patterns [2], but remains below best performance [15]. The use of features which correspond to musical concepts aids the improvement by closer matching listeners cognitive processes (inasfar as musicological concepts as syncopation and metrical profile match listeners perception).

There are many improvements that can be made. The dimensionality is very large and therefore computationally high. One feasible method of reduction of dimensionality is to use the lower Fourier coefficients [4] of the metrical profiles. The features that are compared are insufficient to identify all rhythms which are perceived as similar. The features used summarise behaviour over long spans of time. Averaging across measures to produce metrical profiles is unlikley to sufficiently capture rhythmic features when the corpus consists of highly similar rhythms. For example, an homogenous corpus such as Rock/Pop rhythms are not able to be cleanly separated into rhythm patterns such as dance styles. For such rhythms, matching on short term rhythmic figures or "riffs" independent of the meter may be required.

The presence of syncopation on each measure and it's occurrence across the piece are currently mixed. A low syncopation measure can indicate there weren't many occurances across the track, or that the onsets were corrupted by noise. A possible method to address this is by modelling the observation of onsets as a hidden Markov process. This is a current research topic.

7. ACKNOWLEDGEMENTS

This research was supported by the French project Oseo "Quaero". Thanks are due to Geoffroy Peeters for provision of the beat-tracker.

8. REFERENCES

[1] E. Alpaydin, *Introduction to Machine Learning*. Cambridge, Mass: MIT Press, 2004, 415p.

[2] S. Dixon, F. Gouyon, and G. Widmer, "Towards characterisation of music via rhythmic patterns," in *Proceedings of the International Symposium on Music Information Retrieval*, 2004.

[3] P. Essens, "Structuring temporal sequences: Comparison of models and factors of complexity," *Perception and Psychophysics*, vol. 57, no. 4, pp. 519–32, 1995.

[4] J. Foote, M. Cooper, and U. Nam, "Audio retrieval by rhythmic similarity," in *Proceedings of the International Symposium on Music Information Retrieval*, 2002, pp. 265–266.

[5] F. Gouyon, S. Dixon, E. Pampalk, and G. Widmer, "Evaluating Rhythmic Descriptors For Musical Genre Classification," in *Proceedings of the AES 25th International Conference*, 2004, pp. 196–204.

[6] M. Gruhne, C. Dittmar, and D. Gaertner, "Improving rhythmic similarity computation by beat histogram transformations," in *Proceedings of the International Symposium on Music Information Retrieval*, 2009, pp. 177–182.

[7] C. Guastavino, F. Gómez, G. Toussaint, F. Marandola, and E. Gómez, "Measuring similarity between flamenco rhythmic patterns," *Journal of New Music Research*, vol. 38, no. 2, pp. 129–38, 2009.

[8] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, 2009.

[9] O. Ladinig, "Temporal expectations and their violations," Ph.D. dissertation, Institute for Logic, Language and Computation, Universiteit van Amsterdam, The Netherlands, October 2009.

[10] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*. Cambridge, Mass: MIT Press, 1983, 368p.

[11] J. London, *Hearing in Time: Psychological Aspects of Musical Meter*. Oxford University Press, 2004.

[12] H. C. Longuet-Higgins and C. S. Lee, "The rhythmic interpretation of monophonic music," *Music Perception*, vol. 1, no. 4, pp. 424–41, 1984.

[13] C. Palmer and C. L. Krumhansl, "Mental representations for musical meter," *Journal of Experimental Psychology - Human Perception and Performance*, vol. 16, no. 4, pp. 728–41, 1990.

[14] J. Paulus and A. Klapuri, "Measuring the similarity of rhythmic patterns," in *Proceedings of the International Symposium on Music Information Retrieval*, 2002.

[15] G. Peeters, "Rhythm classification using spectral rhythm patterns," in *Proceedings of the International Symposium on Music Information Retrieval*, 2005.

[16] ——, "Template-based estimation of time-varying tempo," *EURASIP Journal on Advances in Signal Processing*, no. 67215, p. 14 pages, 2007, doi:10.1155/2007/67215.

[17] J. Pressing, "Cognitive complexity and the structure of musical patterns," in *Proceedings of the 4th Conference of the Australasian Cognitive Science Society*, 1999. [Online]. Available: http://psy.uq.edu.au/CogPsych/Noetica/OpenForumIssue8/Pressing.html

[18] I. Shmulevich and D.-J. Povel, "Measures of temporal pattern complexity," *Journal of New Music Research*, vol. 29, no. 1, pp. 61–9, 2000.

[19] L. M. Smith and H. Honing, "Evaluating and extending computational models of rhythmic syncopation in music," in *Proceedings of the International Computer Music Conference*. International Computer Music Association, 2006, pp. 688–91.

[20] S. Streich, "Music complexity: A multifaceted description of audio content," Ph.D. dissertation, Music Technology Group, The Universitat Pompeu Fabra, 2006.

[21] E. Thul and G. Toussaint, "Rhythm complexity measures: A comparison of mathematical models of human perception and performance," in *Proceedings of the International Symposium on Music Information Retrieval*, 2008, pp. 663–8.

[22] G. Tzanetakis and P. R. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.