TOWARDS CHARACTERISATION OF MUSIC VIA RHYTHMIC PATTERNS

Simon Dixon	*Fabien Gouyon*	*Gerhard Widmer*
Austrian Research Institute for AI	Universitat Pompeu Fabra	Medical University of Vienna
Vienna, Austria	Barcelona, Spain	Medical Cybernetics and AI

ABSTRACT

A central problem in music information retrieval is finding suitable representations which enable efficient and accurate computation of musical similarity and identity. Low level audio features are ideal for calculating identity, but are of limited use for similarity measures, as many aspects of music can only be captured by considering high level features. We present a new method of characterising music by typical bar-length rhythmic patterns which are automatically extracted from the audio signal, and demonstrate the usefulness of this representation by its application in a genre classification task. Recent work has shown the importance of tempo and periodicity features for genre recognition, and we extend this research by employing the extracted temporal patterns as features. Standard classification algorithms are utilised to discriminate 8 classes of Standard and Latin ballroom dance music (698 pieces). Although pattern extraction is error-prone, and patterns are not always unique to a genre, classification by rhythmic pattern alone achieves up to 50% correctness (baseline 16%), and by combining with other features, a classification rate of 96% is obtained.

1. INTRODUCTION

Most music can be described in terms of dimensions such as melody, harmony, rhythm, instrumentation and form. These high-level features characterise music and at least partially determine its genre, but they are difficult to compute automatically from audio. As a result, most audio-related music information retrieval research has focussed on extracting low-level features and then using machine learning to perform tasks such as classification. This approach has met with some success, but it is limited by two main factors: (1) the low level of representation may conceal many of the truly relevant aspects of the music; and (2) the discarding of too much information by the feature extraction process may remove information which is needed for the accurate functioning of the system.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.
© 2004 Universitat Pompeu Fabra.

In this work, we address one aspect of these limitations by a novel approach of extracting rhythmic patterns directly from audio in order to characterise musical pieces. It is hypothesised that rhythmic patterns are not randomly distributed amongst musical genres, but rather they are indicative of a genre or small set of possible genres. Therefore, if patterns can be extracted successfully, we can test this hypothesis by examining the usefulness of the patterns as features for genre classification.

As dance music is characterised by repetitive rhythmic patterns, it is expected that the extraction of prominent rhythmic patterns would be particularly useful for classification. However, rhythmic patterns are not necessarily unique to particular dance styles, and it is known that there is a certain amount of overlap between styles. In this work, rather than assuming a fixed dictionary of patterns, we use an automatic extraction technique which finds the most salient pattern for each piece. Thus the techniques used are generalisable to other musical genres.

First the bar length patterns in the amplitude envelope are found and clustered using the k-means algorithm with a Euclidean distance metric. The centre of the most significant cluster is used to represent the piece, and a feature vector consisting of this rhythmic pattern and various derived features is used for classification on a music database of the first 30 seconds of 698 pieces of Standard and Latin ballroom dance music.

Although we focus solely on the rhythmic aspects of music, we show that for genre classification of dance music, a very high level of accuracy is obtainable. The results show an improvement over previous methods which used periodicity or inter-onset interval histograms and features derived from these. Other possible applications of automatically extracted rhythmic patterns are query and retrieval of music, playlist generation, music visualisation, synchronisation with lights and multimedia performances.

In the following section we outline the background and related work, and then in the subsequent sections describe the pattern extraction algorithm and genre classification experiments, concluding with a discussion of the results and future work.

2. RELATED WORK

Audio feature extraction was first addressed in the context of speech recognition, and was later applied to classifi-

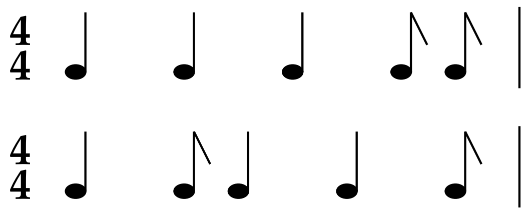

Figure 1. The importance of temporal sequence: these two different rhythm patterns have the same distribution of inter-onset intervals but are typical of two different genres, Cha Cha (above) and Rumba (below).

cation tasks in order to separate speech from non-speech signals such as music and environmental sounds [19, 26]. More recently, several authors have addressed classification tasks specific to music, such as instrument recognition [14] and detection of segments of music that contain singing [2]. Others have focussed on determining similarity judgements for content-based retrieval [8], for the organisation and navigation of large music collections [16, 17] and for computation of high level semantic descriptors [21].

Automatic musical genre classification has a shorter history. Tzanetakis et al. [23, 22] used three sets of features representing the timbral texture, rhythmic content and pitch content of musical pieces, and trained statistical pattern recognition classifiers to achieve a 61% classification rate for ten musical genres. McKinney and Breebaart [15] examined the use of various low level feature sets and obtained 74% classification on 7 musical genres. Dixon et al. [7] compared two methods of periodicity detection, and developed a simple rule based system to classify 17 styles of Standard and Latin ballroom dance music based only on the distribution of periodicities, with an accuracy of 37%.

In the abovementioned work, limited rhythmic information is encoded in the beat histogram [22], modulation energy [15] or periodicity distribution [7]. Each of these representations provides information about the relative frequency of various time intervals between events, but discards the information about their temporal sequence. For example, consider the two rhythmic patterns in Figure 1. Both patterns have the same distribution of inter-onset intervals (three quarter notes and two eighth notes), but the patterns are perceptually very different. The upper pattern, which is typical of a Cha Cha rhythm, would not be described as syncopated, whereas the lower pattern, more likely to be found in a Rumba piece, is somewhat syncopated.

Rhythmic patterns were used by Chen and Chen [4] for song retrieval using symbolic queries and a database of symbolic music, in which approximate string matching provided the similarity measure. The patterns used were not general patterns which summarise a piece or a genre, but specific patterns which did not need to occur more than once in the piece.

Genre	Pieces	Metre	Tempo (nominal)	Tempo (actual)
Cha Cha	111	4	128	92–137
Jive	60	4	176	124–182
Quickstep	82	4	200–208	189–216
Rumba	98	4	104	73–145
Samba	86	4	200	138–247
Tango	86	4	128–132	112–135
Viennese Waltz	65	3	174–180	168–186
Waltz	110	3	84–90	78–106

Table 1. Statistics of the data used in the experiments. Tempo is given in BPM, where a beat corresponds to a quarter note (except for Samba and some Viennese Waltzes which have an eighth note beat). Nominal tempo values are according to the overviews at the `www.ballroomdancers.com` web site.

The only work we are aware of in which rhythmic patterns were automatically extracted from audio data is by Paulus and Klapuri [18], who extracted bar-length patterns represented as vectors of loudness, spectral centroid and MFCCs, and then used dynamic time warping to measure similarity. Their work did not include genre classification, although they did indicate that the similarity of drum patterns was higher within genre than between genres.

Other relevant research that involves the extraction of rhythmic content from a musical performance is beat tracking [10, 11, 20, 3, 5], that is, finding the times of the beats (at various metrical levels). If we assume that rhythmic patterns exist within a particular metrical unit, e.g. within bars, then finding the boundaries of these metrical units becomes a prerequisite to pattern finding. The main difficulty in beat tracking is not in finding the periodicities, but their phase. That is, the length of a pattern can be estimated much more reliably than its starting point. We use an interactive beat tracking system [6] in order to annotate the first bar of each piece.

3. PATTERN EXTRACTION

3.1. Data

Two major difficulties for developing music information retrieval systems are the lack of reliably labelled data sets, and the fuzziness of class boundaries of the attributes. Ballroom dance music has the advantage of providing a set of genres for which there is a high level of agreement among listeners concerning the genre. We collected 698 samples of Standard and Latin ballroom dance music (`http://www.ballroomdancers.com`), each consisting of approximately the first 30 seconds of a piece. The music covers the following eight classes: Cha Cha, Jive, Quickstep, Rumba, Samba, Tango, Viennese Waltz and (Slow) Waltz. The distribution of pieces over the classes, the nominal tempo of each class, and the actual tempo ranges of the excerpts are shown in Table 1.

3.2. Audio Processing

The samples were uncompressed from Real Audio format to a standard PCM format at the same sampling rate as the original file (either 44100, 16000 or 11025 Hz, always mono). The amplitude envelope was extracted from the signal using an RMS filter. The frame rate was set so that a bar would contain a fixed number b of samples at any tempo (where the tempo is already known, as described in the following subsection).

If $x(n)$ is the input signal with sampling rate r and bar length l seconds, then its amplitude envelope is calculated with a sampling rate of b samples per bar using a hop size h given by:

$$h = \frac{rl}{b} \tag{1}$$

The amplitude envelope $y(n)$ is given by:

$$y(n) = \sqrt{\frac{\sum_{i=nh}^{(n+k)h-1} x(i)^2}{kh}} \tag{2}$$

where k is the overlap factor. The bar lengths l ranged from 0.97 to 3.30 sec. Best results were obtained with $b = 72$ and $k = 2$, although values of b from 48 to 144 gave similar results.

Two alternative representations for $y(n)$ were also tried, by taking respectively the square and the absolute value of the signal $x(n)$, passing it through an eighth order Chebyshev Type I lowpass filter, and decimating to a sampling rate of b samples per bar. The choice of representation had only a small influence on results.

3.3. Bar Finding

Much research has been conducted on beat tracking, that is, finding the times of musical beats in audio files [10, 11, 20, 3, 5]. Although beat tracking is not a solved problem, the extraction of periodicities is reliable, with the remaining difficulties being the mapping of periodicities to metrical levels (e.g. estimating which periodicity corresponds to the rate of quarter notes), and choosing the correct phase for a metrical level (e.g. estimating which quarter note beats correspond to the first beat of each bar).

Since the focus of this work was not to perform beat or measure finding, we used values for the first bar generated by BeatRoot [6] and corrected manually. This also allowed us to skip irregular (i.e. tempo-less) introductions, which are difficult to detect automatically in short (30 sec) excerpts.

Once the first bar was known, the process of finding subsequent bars was performed automatically, by searching within $\pm 5\%$ of the end of the previous bar for a starting point which has maximum correlation with the sum of previous bars. That is, for each bar i, a correction factor $\delta(i)$ was calculated which determined the offset of the beginning of the following bar $m(i+1)$ from its expected position ($m(i) + b$). If $d = \lfloor \frac{b}{20} \rfloor$ and $m(i)$ is the index of the beginning of the ith bar, where $m(1)$ is given by BeatRoot, then:

$$m(i+1) = m(i) + b + \delta(i) \tag{3}$$

where

$$\delta(i) = \arg\max_{k=-d}^{d} \sum_{j=0}^{b-1} y(m(i) + b + k + j) * z(i,j) \tag{4}$$

and

$$z(i,j) = \sum_{k=1}^{i} y(m(k) + j) \tag{5}$$

3.4. Extracting Rhythmic Patterns

Once the bar positions were determined, bar length rhythmic patterns were then extracted, consisting of the amplitude envelope of the signal between the start and end points of the bar. The ith pattern v_i is a vector of length b:

$$v_i = \langle y(m(i)), y(m(i)+1), \ldots, y(m(i)+b-1) \rangle \tag{6}$$

In order to remove outliers, k-means clustering (with $k = 4$) was used to find clusters of similar bars, and the largest cluster was taken as defining the most prominent rhythmic pattern for each piece.

If C_j is the cluster containing the most bars, then the characteristic rhythmic pattern $p(n)$ of a piece is given by:

$$p(n) = \frac{1}{|C_j|} \sum_{k \in C_j} y(m(k) + n) \tag{7}$$

Furthermore, we can define the distance $D(i,j)$ between two rhythmic patterns $p_i(n)$ and $p_j(n)$ by the Euclidean metric:

$$D(i,j) = \sqrt{\sum_{k=1}^{b} (p_i(k) - p_j(k))^2} \tag{8}$$

For example, Figure 2 shows the pattern vectors of all 15 bars of one Cha Cha excerpt, where the colours indicate the clusters to which the bars belong, and the thick black curve shows the centre of the largest cluster, that is, the extracted pattern $p(n)$. The perceptual onset of a sound occurs slightly before its energy peak [24], so it is valid to interpret peaks occurring immediately after a metrical boundary as representing an onset at that metrical position. For example, the extracted pattern in Figure 2 has a peak at each eighth note, clearly implying a quadruple metre, and if the five highest peaks are taken, the resulting pattern corresponds to the upper rhythmic pattern in Figure 1.

Viewing the representative patterns for each song provides some feedback as to the success of the pattern extraction algorithm. If the measure finding algorithm fails, the chance of finding a coherent pattern is reduced, although the clustering algorithm might be able to separate the pre-error bars from the post-error bars. The remainder

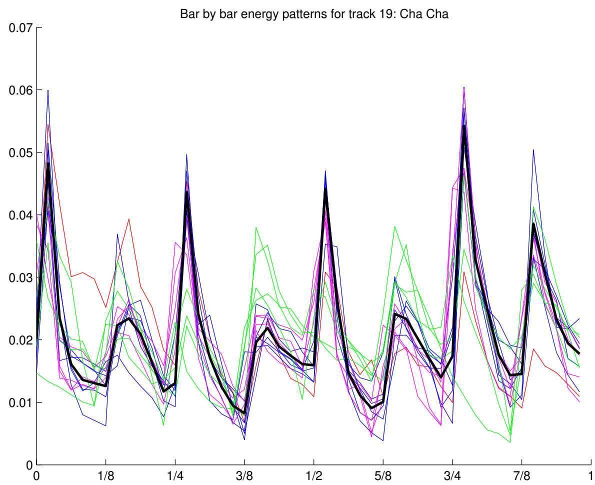

Figure 2. The amplitude envelope of the fifteen bars of excerpt 19 are shown, with the colours representing the four clusters. The thick black line is the centre of the largest cluster, that is, the rhythmic pattern which is extracted for this excerpt. This pattern is somewhat typical of the Cha Cha. The labels on the x-axis (showing musical units) were added for illustrative purposes, and were not known to the system.

of this section gives examples of extracted rhythmic patterns which have features typical of the genres they represent.

Figure 3 shows another Cha Cha piece which has a rhythmic pattern very similar to the one shown in Figure 2. By thresholding below the level of the highest 5 peaks, we again obtain the prototypical Cha Cha rhythmic pattern shown in the upper part of Figure 1.

Music for Jive and Quickstep is usually characterised by swing eighth notes. That is, each quarter note is broken into an unequal pair of "eighth" notes, where the first is longer than the second. The ratio of the lengths of the two notes is known as the *swing ratio*, which is typically around 2:1. Figure 4 shows an extracted pattern where a swing ratio around 2:1 is clearly visible.

One of the characteristics of Rumba is the use of syncopation in the percussion instruments. Accents on the 4th and 6th eighth notes are typical, and this is seen in many of the extracted patterns, such as in Figure 5. This pattern is similar (but not identical) to the rhythm shown in the lower part of Figure 1.

Finally, the two Waltz patterns in Figure 6 clearly show a triple metre, distinguishing these pieces from the other patterns which have a quadruple or duple meter. However, we also note that these two patterns are quite dissimilar, in that the upper one has peaks for each quarter note, whereas the lower pattern has peaks for each eighth note. It is also noticeable in Figure 6 that there is much greater variability between the bars of each piece. The lack of prominent percussion instruments makes the amplitude peaks less pronounced, making bar finding and pattern extraction less reliable. As a result, a number of the Waltz patterns failed

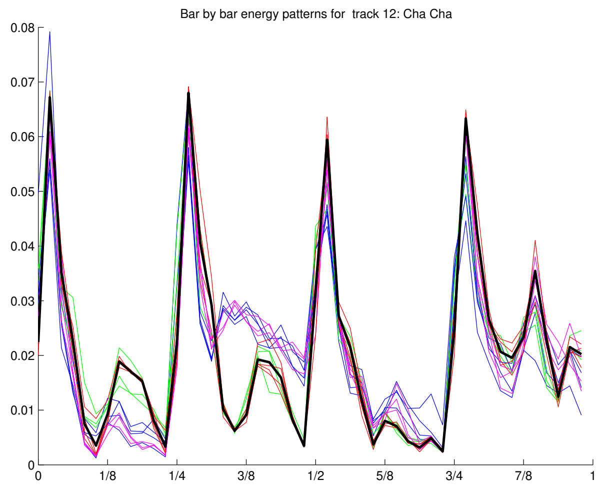

Figure 3. Another Cha Cha piece, which has a pattern very similar to the piece in Figure 2.

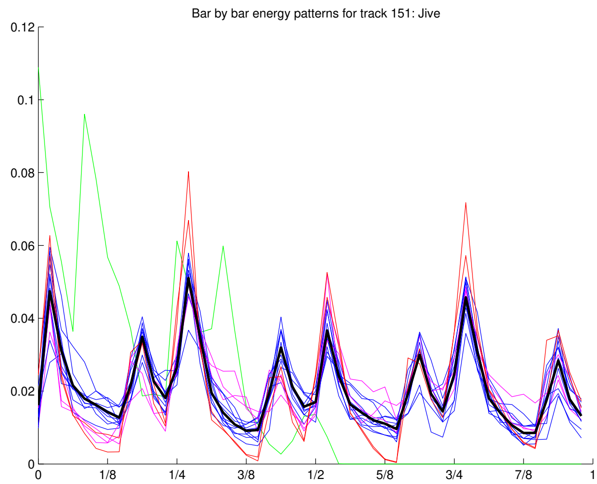

Figure 4. A Jive pattern showing a swing eighth note rhythm.

to show any regularity at all.

4. GENRE CLASSIFICATION EXPERIMENTS

The relevance of the discovered patterns was evaluated in several genre (dance style) classification experiments. Various supervised learning algorithms and data representations (see below) were compared empirically. Classification accuracy was estimated via a standard 10-fold cross-validation procedure: in each experiment, the training examples were randomly split into 10 disjoint subsets (folds), 9 of these folds were combined into a training set from which a classifier was induced, and the classifier was then tested on the remaining tenth fold; this was repeated 10 times, with each fold serving as test set exactly once.

Classification was performed with the software Weka (`www.cs.waikato.ac.nz/ml/weka`) [25], using the following classification algorithms. The simplest method used was the k-Nearest Neighbours (k-NN) algorithm. For

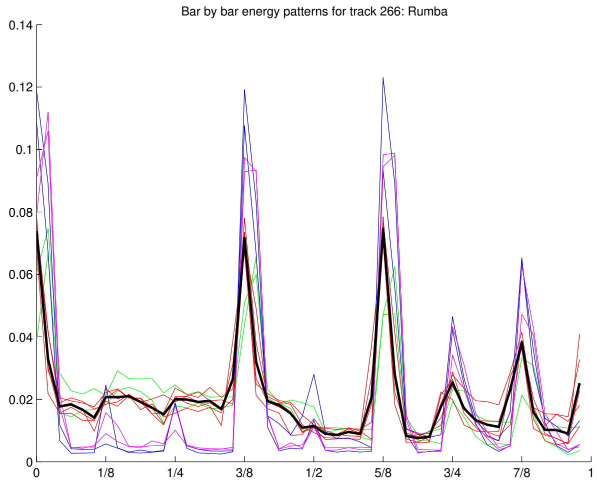

Figure 5. A Rumba pattern showing a strong emphasis on the 4th and 6th eighth notes. (Note that the first eighth note is at 0.)

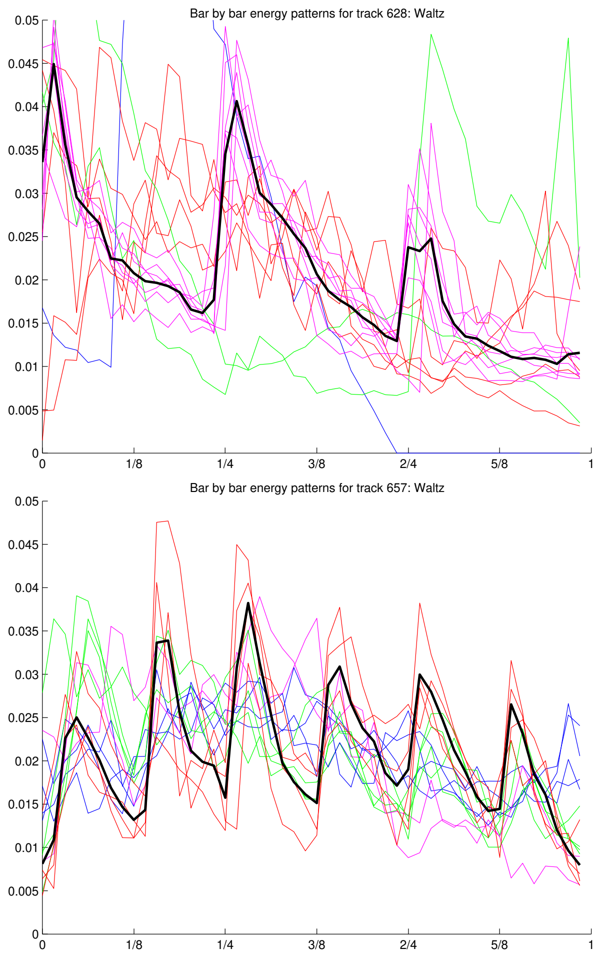

Figure 6. Two Waltz patterns: one in quarter notes (above), and one in eighth notes (below).

Representation	Resolution			
	72	96	120	144
RMS ($k = 1$)	46.4%	45.7%	48.1%	45.1%
RMS ($k = 2$)	47.4%	46.0%	47.1%	46.1%
ABS	43.8%	46.1%	45.8%	46.8%
SQR	44.7%	44.7%	50.1%	45.1%

Table 2. Genre classification results using the rhythmic patterns alone as feature vectors. The rows are different pattern representations, and the columns are the number of points used to represent the patterns.

$k = 1$ this amounts to assigning each test set instance to the class of the nearest element in the training set. For $k > 1$, the k nearest instances in the training set are found, and the greatest number of these neighbours which belong to the same class determines the class of the test instance. Various values of k were used in the experiments. The standard decision tree learning algorithm, J48, was also used, as well as two meta-learning algorithms, AdaBoost and Classification via Regression. AdaBoost [9] runs a given weak learner (in this case J48) several times on slightly altered (reweighted) training data and combines their predictions when classifying new cases. Classification via regression (using M5P and linear regression as base classifiers) builds a regression model for each class and combines the models via voting.

4.1. Classification by Rhythmic Pattern Alone

The first set of classification experiments was performed with $p(n)$ as the feature vector for each excerpt, that is, using the rhythmic pattern alone for classification. Note that this representation is totally independent of tempo. The classification rates for various pattern representations described in subsection 3.2 are shown in Table 2. The best classification rate, 50%, was achieved using the AdaBoost classifier, with the decimated squared signal representation with $b = 120$. This is well above the baseline for classification of this data, which is 16%.

The confusion matrix for this classifier is shown in Table 3. Viennese Waltzes were the most poorly classified, while classification of Cha Cha pieces was the most accurate. The greatest mutual confusion was between the Waltz and Viennese Waltz, which is to be expected, since they have the same metre and often use the same rhythmic patterns and instruments, and the clearly distinguishing feature, the tempo, is not encoded in the rhythmic pattern.

4.2. Calculation of Derived Features

The rhythmic patterns themselves do not contain information about their time span, that is, they are independent of the tempo. Since the tempo is one the most important features in determining dance genre [7, 12], we tested classification with a combination of the rhythmic patterns,

	C	J	Q	R	S	T	V	W	Rec%
C	74	6	0	14	7	7	0	3	67
J	10	23	11	1	1	10	1	3	38
Q	0	11	35	2	6	9	5	14	43
R	20	0	3	53	1	3	2	16	54
S	8	0	9	11	43	5	3	7	50
T	15	8	7	6	5	35	0	10	41
V	0	2	6	3	0	1	23	30	35
W	1	0	9	6	4	7	19	64	58
Prec%	58	46	44	55	64	46	43	44	

Table 3. Confusion matrix for classification based on rhythmic patterns alone. The rows refer to the actual style, and the columns the predicted style. The rightmost column shows the percentage recall for each class and the bottom row shows the percentage precision. The abbreviations for the dance styles are: Cha Cha (C), Jive (J), Quickstep (Q), Rumba (R), Samba (S), Tango (T), Viennese Waltz (V), Waltz (W).

features derived from the rhythmic patterns, features derived from the audio data directly, and the tempo.

The features derived from the rhythmic patterns were: the mean amplitude of the pattern, the maximum amplitude of the pattern, the relative maximum amplitude of the pattern (maximum divided by mean), the standard deviation of the pattern amplitudes, an estimate of the metre, a syncopation factor, and the swing factor. The metre was estimated by calculating two weighted sums of the pattern, the first with higher weights around the positions of a division of the bar into 4 quarter notes (8 eighth notes), the second with the weights set for a division of the bar into 3 quarter notes (6 eighth notes). The greater of the two sums determined the metre as a binary attribute, indicating either a quadruple or ternary metre. The syncopation factor was calculated as the relative weights of the offbeat eighth notes (i.e. the 2nd, 4th, etc.) to the on-beat eighth notes. The swing factor was calculated using a pulse train of Gaussian curves spaced at quarter note intervals, correlating with the signal and finding the highest 2 peaks, which usually correspond to the positions of the quarter note and eighth note respectively. If the duration of the quarter note is q and the interval between the two peaks is r, then the swing factor s is given by:

$$s = \max\left(\frac{r}{q-r}, \frac{q-r}{r}\right) \tag{9}$$

If only one peak in the correlation was found, the swing factor was set to 0.

An additional feature set, containing three groups of descriptors (as described by Gouyon et al. [12]) was also used. The first group of descriptors was tempo-related features, including the measured tempo calculated from the bar length. The second group consisted of features derived from the periodicity histogram representation, and the third group of features were derived from inter-onset interval histograms. Apart from the measured tempo, all

	C	J	Q	R	S	T	V	W	Rec%
C	102	1	0	5	0	3	0	0	92
J	0	58	0	0	0	2	0	0	97
Q	0	0	80	0	0	0	2	0	98
R	1	1	0	92	0	4	0	0	94
S	0	0	2	0	82	0	1	1	95
T	1	0	0	2	0	83	0	0	97
V	0	0	0	0	0	0	65	0	100
W	0	0	0	1	1	0	0	108	98
Prec%	98	97	98	92	99	90	96	99	

Table 4. Confusion matrix for classification using rhythmic patterns and other features. Compare with Table 3.

of these values were calculated automatically (see Table 5 and [12] for more details).

4.3. Classification Using All Features

In the following classification experiments using all features, a classification rate of 96% was achieved with the AdaBoost classifier, using the RMS signal with $k = 2$ and $b = 72$. This is remarkable, considering that spectral features are not represented at all in the data, and there is certainly some ambiguity in the relationship between music pieces and dance styles. The confusion matrix is shown in Table 4. More than half (16 out of 28) of the errors are caused by confusion of Cha Cha, Tango and Rumba. From Table 1, we see that these styles have strongly overlapping tempo ranges and the same metre, so other features must be used to distinguish these classes.

Comparisons of classification rates with various subsets of features were performed to determine the relative contribution of each subset (see Table 5). The left hand column shows the results from Gouyon et al. [12]; classification using tempo alone achieved up to 82%, classification using other features not including tempo also reached 82%, and by combining these features, a classification rate of 93% was obtained. The right hand column shows the results of adding rhythmic patterns and their derived features to the feature vectors: in each case an improvement was made, with overall classification rates improving to 84% (compared with 82%) without the tempo and 96% (compared with 93%) including the tempo. For all of these results, the rhythmic patterns were generated with the RMS filter with $k = 2$ and $b = 72$, and the AdaBoost learning algorithm was used (hence the difference from published results in [12]).

5. DISCUSSION AND FURTHER WORK

It is not to be expected that a single rhythmic pattern could uniquely determine the genre of a piece of dance music. Many other features which are not represented in this work are also relevant to genre, such as the choice of musical instruments, which could perhaps be represented with standard timbral features such as MFCCs. Examination

Feature sets from [12]	Without RP	With RP
None (0)	15.9%	50.1%
Periodicity histograms (11)	59.9%	68.1%
IOI histograms (64)	80.8%	83.4%
Periodicity & IOI hist. (75)	82.2%	85.7%
Tempo attributes (3)	84.4%	87.1%
All (plus bar length) (79)	95.1%	96.0%

Table 5. Comparison of classification rates using various sets of features. The columns show rates without and with the rhythmic patterns (RP) and their derived features; the rows show the different feature subsets from Gouyon et al. [12], with the number of features shown in parentheses.

of the extracted patterns shows that some of the patterns are quite trivial, such as those which show sharp peaks on each of the quarter note beats, thus only serving to distinguish triple from quadruple metre. Nevertheless, even with these limitations, the results demonstrate that rhythmic patterns are a useful feature for classification.

The fact that only 30 seconds of each song was used may have adversely influenced the results, as many songs have an introduction which does not match the style of the rest of the piece. Because of the shortness of the tracks, it was considered better to extract only one rhythmic pattern. With longer tracks it would be worthwhile to investigate classification using multiple patterns per song. It is also expected that the statistical reliability of pattern extraction would increase with the length of the excerpt.

One restriction of the current work is that it relies on an accurate estimate of the first bar. Automatic methods of finding metrical boundaries have made great progress in recent years, but they are still far from perfect, and manual correction for very large music databases is not feasible. However the errors of such systems are not random; they belong to a very small class of possibilities: tempo errors of a factor of 2 or 3, and phase errors of half (or occasionally a third or quarter) of the metrical unit. If we allow these cases, no longer considering them as errors, the classification algorithm could possibly succeed in implicitly recognising these different cases.

Another limitation is that although we do not explicitly detect percussive onsets, the methodology assumes peaks in energy (e.g. for correlation) for extracting the patterns. This limitation is seen in the patterns extracted from Waltz and Viennese Waltz excerpts. An explicit onset detection step which includes the detection of soft (i.e. non-percussive) onsets [1] could be used to alleviate this problem. Another approach would be to use features other than amplitude or energy. Paulus and Klapuri [18] found that the spectral centroid, normalised by the energy, provided the best feature vector for describing patterns.

The high dimensionality of the pattern vectors reduces the ability of learning algorithms to build suitable classifiers. Dimensionality reduction either by PCA or by a more explicit symbolic encoding (i.e. in musical symbols) would be a step in the direction of solving this problem. If the patterns were quantised and encoded into musical units, they could be matched to explicit patterns such as those found in instructional books. Even without such an encoding, matching in the other direction, i.e. from explicit patterns to the audio data could be performed as a method of generating further features.

A related issue that is yet to be explored is the choice of distance metrics between patterns. The Euclidean distance is not necessarily ideal, as it treats all time points independently, so that, for example, peaks which almost match are penalised as heavily as peaks which are far from being aligned.

Another avenue of further research would be to extract patterns in various frequency bands in order to detect between-instrument patterns (e.g. bass drum, snare drum, hi-hat). Alternatively, recent work on drum sound recognition [13] could be used to determine multi-dimensional rhythmic patterns. These ideas would necessitate the development of more complex methods of encoding and comparing patterns.

There are numerous other directions of possible further development. The current experiments are limited in the genres of music on which they have been performed. As other labelled data sets become available, it will be possible to test the generality of this method of pattern extraction and comparison for classification of other genres. The algorithms are general purpose; no domain specific knowledge is encoded in them. The unknown issue is the extent to which other genres are characterised by rhythmic patterns.

6. CONCLUSION

We described a novel method of characterising musical pieces by extracting prominent bar-length rhythmic patterns. This representation is a step towards building higher level, more musically relevant, parameters which can be used for genre classification and music retrieval tasks. We demonstrated the strength of the representation on a genre recognition task, obtaining a classification rate of 50% using the patterns alone, 84% when used in conjunction with various automatically calculated features, and 96% when the correct tempo was included in the feature set. These classification rates represent a significant improvement over previous work using the same data set [12], and higher rates than have been published on other data sets [22, 15, 7]. However, we acknowledge the preliminary nature of these investigations in the quest to extract semantic information from audio recordings of music.

7. ACKNOWLEDGEMENTS

This work was funded by the EU-FP6-IST-507142 project SIMAC (Semantic Interaction with Music Audio Contents). The Austrian Research Institute for Artificial Intelligence also acknowledges the financial support of the Austrian Federal Ministries of Education, Science and Culture and of Transport, Innovation and Technology.

References

[1] Bello, J. and Sandler, M. (2003). Phase-based note onset detection for musical signals. In *International Conference on Acoustics, Speech and Signal Processing*.

[2] Berenzweig, A. and Ellis, D. (2001). Locating singing voice segments within musical signals. In *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 119–123, Mohonk, NY.

[3] Cemgil, A., Kappen, B., Desain, P., and Honing, H. (2000). On tempo tracking: Tempogram representation and Kalman filtering. In *Proceedings of the 2000 International Computer Music Conference*, pages 352–355, San Francisco CA. International Computer Music Association.

[4] Chen, J. and Chen, A. (1998). Query by rhythm: An approach for song retrieval in music databases. In *Proceedings of the 8th IEEE International Workshop on Research Issues in Data Engineering*, pages 139–146.

[5] Dixon, S. (2001a). Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30(1):39–58.

[6] Dixon, S. (2001b). An interactive beat tracking and visualisation system. In *Proceedings of the International Computer Music Conference*, pages 215–218, San Francisco CA. International Computer Music Association.

[7] Dixon, S., Pampalk, E., and Widmer, G. (2003). Classification of dance music by periodicity patterns. In *4th International Conference on Music Information Retrieval (ISMIR 2003)*, pages 159–165.

[8] Foote, J. (1997). Content-based retrieval of music and audio. In *Multimedia Storage and Archiving Systems II*, pages 138–147.

[9] Freund, Y. and Schapire, R. (1996). Experiments with a new boosting algorithm. In *Proceedings of the Thirteenth International Conference on Machine Learning*, pages 148–156.

[10] Goto, M. and Muraoka, Y. (1995). A real-time beat tracking system for audio signals. In *Proceedings of the International Computer Music Conference*, pages 171–174, San Francisco CA. International Computer Music Association.

[11] Goto, M. and Muraoka, Y. (1999). Real-time beat tracking for drumless audio signals. *Speech Communication*, 27(3–4):311–335.

[12] Gouyon, F., Dixon, S., Pampalk, E., and Widmer, G. (2004). Evaluating rhythmic descriptors for musical genre classification. In *Proceedings of the AES 25th International Conference*, pages 196–204.

[13] Herrera, P., Dehamel, A., and Gouyon, F. (2003a). Automatic labeling of unpitched percussion sounds. In *Presented at the 114th Convention of the Audio Engineering Society*, Amsterdam, Netherlands.

[14] Herrera, P., Peeters, G., and Dubnov, S. (2003b). Automatic classification of musical instrument sounds. *Journal of New Music Research*, 32(1):3–22.

[15] McKinney, M. and Breebaart, J. (2003). Features for audio and music classification. In *4th International Conference on Music Information Retrieval (ISMIR 2003)*, pages 151–158.

[16] Pampalk, E. (2001). Islands of music: Analysis, organization, and visualization of music archives. Master's thesis, Vienna University of Technology, Department of Software Technology and Interactive Systems.

[17] Pampalk, E., Dixon, S., and Widmer, G. (2004). Exploring music collections by browsing different views. *Computer Music Journal*, 28(2):49–62.

[18] Paulus, J. and Klapuri, A. (2002). Measuring the similarity of rhythmic patterns. In *Proceedings of the 3rd International Conference on Musical Information Retrieval*, pages 150–156. IRCAM Centre Pompidou.

[19] Saunders, J. (1996). Real time discrimination of broadcast speech/music. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pages 993–996.

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

[21] Scheirer, E. (2000). *Music-Listening Systems*. PhD thesis, Massachusetts Institute of Technology, School of Architecture and Planning.

[22] Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302.

[23] Tzanetakis, G., Essl, G., and Cook, P. (2001). Automatic musical genre classification of audio signals. In *International Symposium on Music Information Retrieval*.

[24] Vos, J. and Rasch, R. (1981). The perceptual onset of musical tones. *Perception and Psychophysics*, 29(4):323–335.

[25] Witten, I. and Frank, E. (1999). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Francisco, CA.

[26] Wold, E., Blum, T., Keislar, D., and Wheaton, J. (1996). Content-based classification, search, and retrieval of audio. *IEEE Multimedia*, 3(2):7–36.