Exploring the effect of rhythmic style classification on automatic tempo estimation

Matthew E. P. Davies and Mark D. Plumbley
Centre for Digital Music, Queen Mary, University of London
Mile End Rd, E1 4NS, London, United Kingdom
email: matthew.davies@elec.qmul.ac.uk
web: www.elec.qmul.ac.uk/digitalmusic

March 11, 2008

Abstract

Within ballroom dance music, tempo and rhythmic style are strongly related. In this paper we explore this relationship, by using knowledge of rhythmic style to improve tempo estimation in musical audio signals. We demonstrate how the use of a simple 1-NN classification method, able to determine rhythmic style with 75% accuracy, can lead to an 8% point improvement over existing tempo estimation algorithms with further gains possible through the use of more sophisticated classification techniques.

1 Introduction

The automatic extraction of tempo from musical audio forms a key component in many aspects of rhythmic analysis and has received wide attention in the music signal processing research community [1, 2]. Perhaps the most common use for tempo is within the task of *beat tracking* where the aim is to replicate human foot-tapping in time to music. For this task, the tempo indicates the rate at which the beats occur. Therefore to maintain a consistent beat output it is imperative to have an accurate method for finding and tracking the tempo. While considerable progress has been made in this field (see [1, 2] for an overview of existing techniques) an ongoing difficulty has been in identifying the tempo in a manner consistent with a human listener. The highest performing tempo estimation algorithms are able to infer the tempo with 85% accuracy provided the evaluation method used allows for the estimated tempo to be "correct" if it can be related by a factor of two to the annotated tempo [1]. This double/half ambiguity is known as the *tempo octave* problem [3]. When these related tempo octaves aren't considered accurate, the overall performance of the best performing algorithms drops by approximately 20% points [1].

For certain applications, e.g. beat-dependent audio effects [4], octave ambiguity may not be critical, but for others finding the annotated tempo becomes far more important. One such example is the classification of ballroom dance music. Most existing work on rhythmic style classification [5, 6, 7] has made use of the same ballroom dance database. It contains 698 excerpts (each 30 seconds in length) across 8 rhythmic styles: Jive, QuickStep, Tango, Waltz, Viennese-Waltz, Samba, ChaCha and Rumba. Ballroom dances are typically characterised by a repeating rhythmic pattern at a particular tempo [6]. The restriction of ballroom dances to small ranges of tempi has meant that tempo has been identified as an important discriminating feature for dance music classification; however tempo alone is not sufficient to provide a perfect classification [8].

To avoid the issue of tempo octave ambiguity in automatic tempo estimation, rhythmic style classification algorithms (e.g. [6, 7]) use annotated tempo rather than automatically extracted values. The tempo is then combined with multiple features extracted from rhythmic pattern representations and passed to a classification algorithm to return a style label for a given input signal. To characterise the rhythmic properties Dixon et al [6] use a predominant bar length pattern, where as Peeters [7] uses autocorrelation functions and spectral rhythmic patterns.

In a more recent study, Seyerlehner et al [9] explore the relationship between tempo and rhythmic style from a different perspective. Again using the ballroom data they use rhythmic pattern matching as means for identifying tempo. Given a periodicity pattern for each musical excerpt and its ground truth tempo, they find the tempo for an unknown excerpt by taking the average of the ground truth tempi resulting from a k-NN classification (where k=5). They compare two rhythmic features: an autocorrelation function signal similar to that used in [7]; and a fluctuation pattern which has been used in previous work on music similarity [10]. For which they find the fluctuation pattern to be more successful feature.

We extend their approach by investigating a simple style-dependent method for tempo estimation, where knowledge of musical style with a known nominal tempo is used to guide the range of likely tempi within our existing tempo extraction algorithm [11]. In contrast to the approach of Seyerlehner et al [9] which requires that all 698 patterns from the ballroom set with associated tempo annotations be stored, we simply store one pattern per musical style and use a single nominal tempo value. For each unknown excerpt we then perform a 1-NN classification and pass the nominal tempo of the nearest neighbour to our existing tempo estimation algorithm.

Our results indicate that using this simple classification we can achieve rhythmic style classification of 75% which in turn improves the performance of our tempo estimation algorithm from 71% to 79%. With the use of a more sophisticated classification algorithm (the Adaboost classifier, as used for this task in [6, 7] we can identify rhythmic style with 85% accuracy which leads to a tempo accuracy of 86%.

The remainder of this paper is structured as follows. In section 2 we describe our simplified method for rhythmic style classification. In section 3 we review our existing tempo extraction algorithm and then illustrate the modifications necessary to encode knowledge of rhythmic style. We evaluate our method for rhythmic style classification and demonstrate its effect on the performance of our tempo estimation algorithm in section 4. We present discussion and conclusions in section 5.

2 Rhythmic style classification

Our method for rhythmic style classification requires two components: (i) a suitable feature derived from the musical audio which maximises intra-style rhythmic similarity and minimises inter-style similarity; and (ii) a classification method able to exploit the properties of the input feature. Our motivation is towards a simple solution for each component - ideally one that can be incorporated into our tempo extraction algorithm will minimal extra processing. To this end, we derive a feature for rhythmic style classification directly from the input to our tempo extraction algorithm and embed the style classification method into the tempo calculation.

2.1 Classification feature

The input to our tempo extraction algorithm is the complex spectral difference onset detection function [12] – a mid-level representation of the input audio signal which emphasises the locations of note onsets. Given an input signal $s(n)$ we calculate the m^{th} sample of the onset detection function $\Gamma(m)$ by measuring the sum of the Euclidean distance between an observed short term spectral frame $S_k(m)$ and a predicted frame $\hat{S}_k(m)$ for each bin, k:

$$\Gamma(m) = \sum_{k=1}^{K} |S_k(m) - \hat{S}_k(m)| \tag{1}$$

where each detection function (DF) sample has a temporal resolution t_{DF}=11.6ms. For a complete derivation see [12].

As the basis for rhythmic style classification, Dixon et al [6] extract a predominant bar length pattern derived from an onset detection function type representation. While a suitable feature for describing the rhythmic properties of the input signal, its extraction requires prior knowledge of the bar locations. Due to limitations in the automatic detection of bar boundaries, Dixon et al [6] extracted them in a semi-automatic manner. Since our interest is in performing a fully automatic style classification, we cannot make use of such information. As an alternative to a temporal rhythmic pattern, Peeters [7] and later Seyerlehner et al [9] adopted a periodicity pattern based on the autocorrelation function (ACF) of an onset detection function type representation. Because our tempo extraction method [11] extracts a salient periodicity from the autocorrelation function of the onset detection function we also follow this approach.

To emphasise the peaks in the onset detection function (prior to deriving the autocorrelation function) we calculate an adaptive moving mean threshold:

$$\bar{\Gamma}(m) = \text{mean}\{\Gamma(q)\} \quad m - \tfrac{Q}{2} \leq q \leq m + \tfrac{Q}{2} \tag{2}$$

where Q indicates the approximate width of a typical peak in $\Gamma(m)$. In earlier work we found Q=16 DF samples to be a suitable value. We then subtract the adaptive threshold from $\Gamma(m)$ to give a modified onset detection function:

$$\tilde{\Gamma}(m) = \text{HWR}(\Gamma(m) - \bar{\Gamma}(m)) \tag{3}$$

where HWR performs half-wave rectification such that $\text{HWR}(x) = (x + |x|)/2$.

The autocorrelation function $A(l)$ for lag l is calculated using

$$A(l) = \frac{\sum_{m=1}^{L} \tilde{\Gamma}(m)\tilde{\Gamma}(m-l)}{|l-L|} \quad l = 1, \ldots, L \tag{4}$$

where the denominator corrects for the bias which occurs as a function of lag.

The ACF used by Seyerlehner et al [9] includes lags up to 4 seconds. If the tempo of each excerpt is not constant, then the peaks of the ACF at longer lags will be smeared. To reduce this affect we use a smaller range of lags, by setting L=144 DF samples in equation (4) as used by Dixon et al [6] as the duration of their bar length feature.This corresponds to $L.t_{DF}$=1.67 seconds.

In our approach the location of the peaks in $A(l)$ are the important features which we use to infer the style of the input. To emphasise the peaks of $A(l)$ we employ a second thresholding processing. We create a modified autocorrelation function $\tilde{A}(l)$ by substituting $\Gamma(m)$ for $A(l)$ and applying equations (2) and (3). In comparison to Seyerlehner et al [9] our ACF feature covers a shorter range of lags and has been subject to a peak-preserving adaptive threshold.

2.2 Classification methods

The ballroom dance database used in this work is comprised of 8 rhythmic styles: Jive (J), QuickStep (Q),Tango (T), Waltz (W), Viennese-Waltz (V), Samba (S), ChaCha (C) and Rumba (R). We use parameter X to refer to a generic rhythmic style and give the following arbitrary ordering $X = \{$J,Q,T,W,V,S,C,R$\}$. For the z^{th} excerpt of each rhythmic style X we calculate an ACF pattern $\tilde{A}_{X,z}(l)$ as described above.

The basis for our simple approach to style classification is to define one ACF pattern, $P_X(l)$ per style. We follow the clustering approach of Dixon et al [6], who derive a predominant rhythmic pattern by clustering the bar length patterns (using k-means) for a given each excerpt and returning the temporal average of the largest cluster. Our ACF feature $\tilde{A}(l)$ already summarises each excerpt in one signal, therefore to summarise a rhythmic style, we cluster $\tilde{A}_{X,z}(l)$ for all z using k-means (with k=2), and find the predominant pattern for each style $P_X(l)$ as the temporal average of the largest cluster. The predominant patterns for each style are shown along with the nominal tempo for each rhythmic style in figure 1.

Given an incoming ACF pattern feature we employ a 1-NN (nearest neighbour) classifier by measuring the Euclidean distance $D(X)$ between $\tilde{A}(l)$ and each $P_X(l)$ where each signal has been normalised to sum to unity

$$D(X) = \sum_{l=1}^{L} \left| |P_X(l)|^2 - |\tilde{A}(l)|^2 \right|^{(1/2)} \tag{5}$$

where the classified style $\hat{X}$ is found as

$$\hat{X} = \arg\min_X (D(X)). \tag{6}$$

While this 1-NN approach is simple both conceptually and in terms of implementation, in order to gauge how accurate it is as a classifier we also explore the use of a more sophisticated classification algorithm. For this purpose, we select the Adaboost classifier as used by Dixon et al [6] and Peeters [7] from the open source data mining software WEKA [13].

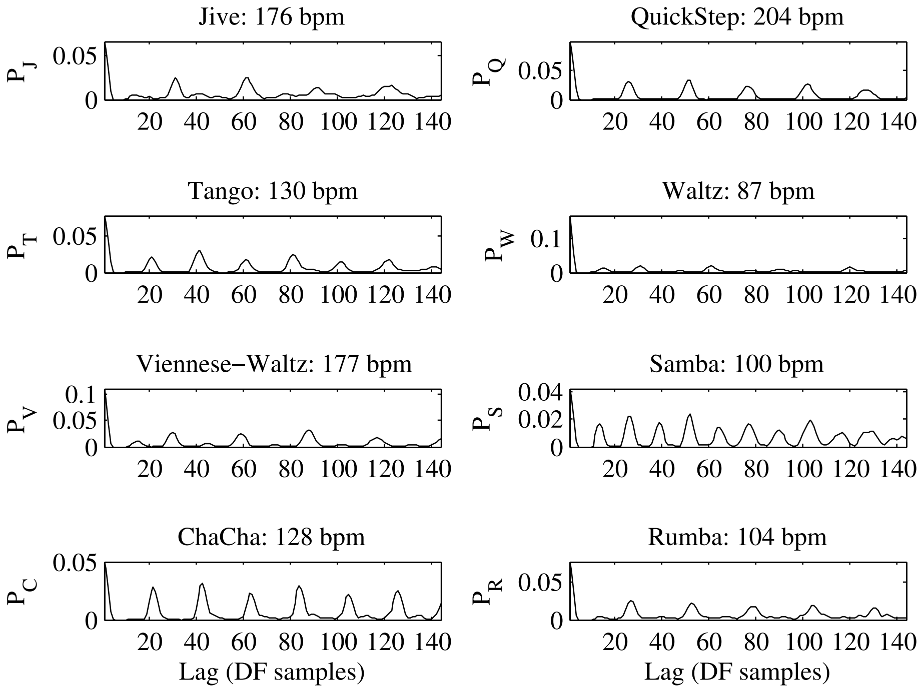

Figure 1: Predominant periodicity patterns $P_X(l)$ with ground-truth nominal tempi: Jive, QuickStep, Tango, Waltz, Viennese-Waltz, Samba, ChaCha, Rumba. Each pattern has been normalised to sum to unity.

3 Tempo estimation with rhythmic style

In section 2.1 we introduced the onset detection function and the subsequent calculation of the autocorrelation function feature $A(l)$. In our existing tempo extraction algorithm [11, 2] we identify a salient periodicity (the *beat period*) by passing the autocorrelation function through a *shift-invariant* comb filterbank which is scaled by a perceptually motivated weighting over possible beat periods. The weighting function $W(l)$ is derived from the Rayleigh distribution function which strongly attenuates very short lags while decays more gently for longer lags

$$W(l) = \frac{l}{\beta^2} \exp\left(\frac{-l^2}{2\beta^2}\right) \qquad l = 1, \ldots, L \tag{7}$$

where the constant β is set to 43 DF samples, which is equivalent to 120 beats per minute (bpm) using the following relationship for converting ACF lag into tempo

$$\text{tempo} = \frac{60}{l \times t_{DF}}. \tag{8}$$

The beat period is then extracted as the index of the maximum value of the output of the comb filterbank, which can be converted to tempo using equation

(8). For a complete description of our tempo estimation algorithm see [11, 2].

While the Rayleigh weighting $W(l)$ is suitable when the rhythmic style is unknown, once we know the style $W(l)$ becomes too broad and can leave the tempo estimation susceptible to octave errors. We therefore restrict the likely range of observable periodicities, through the use of a style-dependent weighting, $W_{\hat{X}}(l)$ which we define in terms of a Gaussian centred on the nominal periodicity $\tau_{\hat{X}}$ for the classified style $\hat{X}$ with standard deviation set at $\tau_{\hat{X}}/2$

$$W_{\hat{X}}(l) = \exp\left(\frac{-\left(l-\tau_{\hat{X}}\right)^2}{2(\tau_{\hat{X}}/2)^2}\right) \qquad l = 1, \ldots, L \tag{9}$$

where τ_X can take values $\{29, 25, 40, 59, 29, 52, 40, 50\}$ DF samples by applying equation (8) to the nominal tempi from figure 1 given the arbitrary ordering $X = \{\text{J,Q,T,W,V,S,C,R}\}$. We can then identify the beat period (and therefore the tempo) by finding the index of the maximum value of output of the style-dependent weighted comb filterbank.

4 Results

We evaluate the performance of our style classification method and subsequent tempo estimation on the 698 excerpt ballroom dance database which has been used for both these tasks in previous work [6, 9] and is publicly available[1].

4.1 Style Classification

We calculate the accuracy of the simple 1-NN classifier and the Adaboost classifier as the ratio of the number of correct classifications to the total number of excerpts to classify. To maintain consistency with the methods of Dixon et al [6] and Peeters [7] we undertake a 10-fold cross validation, where there is 90%/10% split between training and testing data, where each excerpt can only be in the testing group once. For our 1-NN classifier we therefore generated a new set of predominant patterns $P_X(l)$ for each fold of the validation rather than use a single global pattern for each style. The raw decisions of each classification algorithm are shown figure 2. The overall performance of our two classifiers in comparison with existing algorithms on the same dataset are summarised in Table 1.

Of the fully automatic style classification methods the 1-NN classifier is the weakest at 75% but is still comparable to the other classifiers. It is important to note that our 1-NN approach makes use of a just single pattern $P_X(l)$ per cross-validation fold, where as each of the other classifiers has access to all of the training examples. The 86% accuracy of our Adaboost classifier (which is able to draw on all the training examples) actually exceeds the performance of all existing fully automatic algorithms on this dataset (e.g. 81% accuracy of Peeters [7]). This suggests that the extra processing applied to our ACF feature in section 2.1 had a positive effect on the outcome. The Adaboost classifier is still less accurate than the best performing semi-automatic approaches [6, 7] but each of these has access to ground truth tempo annotations; data which our classifiers cannot be permitted to use.

[1] http://mtg.upf.edu/ismir2004/contest/tempoContest/node5.html

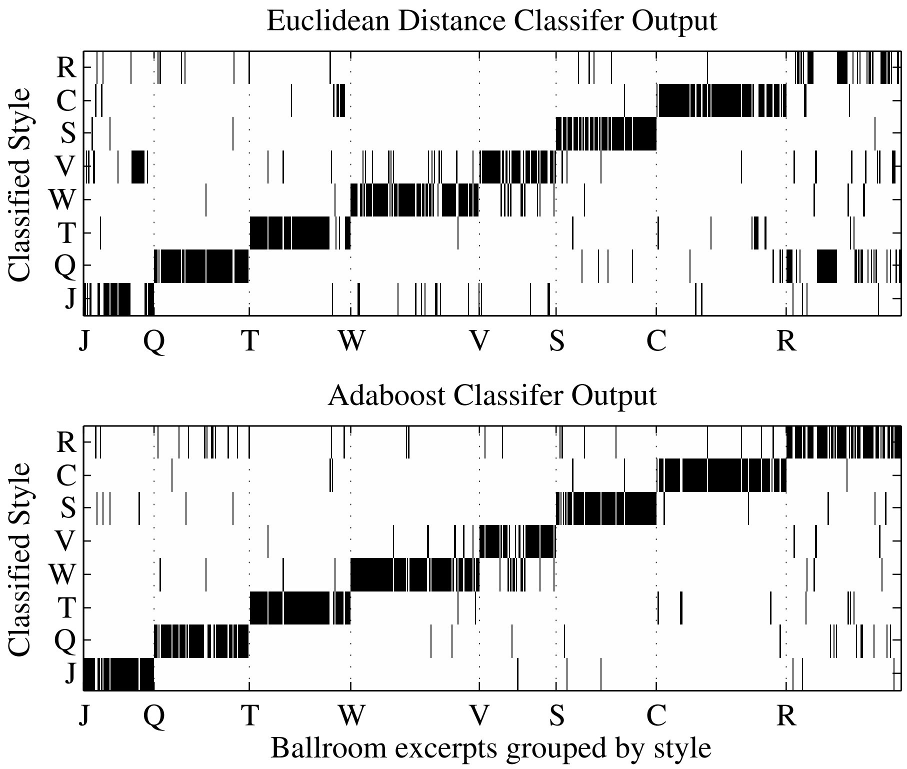

Figure 2: Raw decisions by rhythmic style classifiers. Top: Euclidean distance classifier. Bottom: Adaboost Classifier.

4.2 Tempo Estimation

We now explore the effect of style classification on tempo estimation. The performance of our tempo estimation algorithm is measured for four cases: (i) tempo estimation with no access to style information (our baseline system) [11, 2]; (ii) tempo estimation given the output of the Euclidean distance classifier; (iii) tempo estimation given the output of the Adaboost classifier; and (iv) tempo estimation given hypothetical perfect style classification. Tempo accuracy is calculated according to the two methods in [1]: T1 where the a given tempo is accurate if it within is $\pm 4\%$ of the ground truth value and T2 which allows for the tempo to be within $\pm 4\%$ of double or half the annotated tempo. The results are summarised according to rhythmic style in Table 2.

By inspection of the “Overall T1” row of Table 2 we can see that knowledge of musical style can lead to an improvement in tempo accuracy, even when the style classifier used is only 75% accurate itself. It is interesting to note that while the knowledge of rhythmic style leads to a drastic improvement for some styles (e.g. Jive, QuickStep) the tempo accuracy for the Rumba is reduced by almost 50% when using the output of the Euclidean distance based classifier. Referring back to figure 2, we can see that many of the Rumba examples were mis-classified as QuickStep. This is not an unexpected result given the predominant patterns in figure 1. The tempo of the QuickStep is approximately twice that of the

	Classification Feature(s)	Accuracy (%)
Dixon et al [6]:	Pattern Only	50.1*
	Automatic Features (62)	82.2
	Auto+Semi-auto Features(79)	96.0*
Gouyon et al [5]:	MFCC Features	79.6
Peeters [7]:	Pattern Only	80.8
	Pattern + Tempo	90.4*
DP	Pattern Only (Euclidean)	75.3
	Pattern Only (Adaboost)	85.0

Table 1: Accuracy of Rhythmic Style Classification. Accuracy values marked with * were calculated with access to ground truth annotated data.

Rhythmic Style	No Style (%)	Euc. Style (%)	Ada. Style (%)	Perfect Style (%)
Jive: 176 bpm	35.0	97.4	96.7	98.3
QuickStep: 204 bpm	13.4	83.2	76.8	95.1
Tango: 130 bpm	95.3	93.4	93.0	95.4
Waltz: 87 bpm	55.5	65.6	79.1	85.5
Viennese-W: 177 bpm	72.3	86.9	80.0	100.0
Samba: 100 bpm	93.0	85.6	89.5	93.0
ChaCha: 128 bpm	98.2	91.5	97.3	97.3
Rumba: 104 bpm	85.7	44.1	75.5	88.8
Overall T1	70.9	79.4	85.8	93.6
Overall T2	93.3	94.0	94.4	94.6

Table 2: Effect of style classification on tempo accuracy. Performance is divided between each rhythmic style under conditions of increasing style classification performance. Euc. refers to 1-NN classifier by Euclidean distance. Ada. refers to the Adaboost classifier.

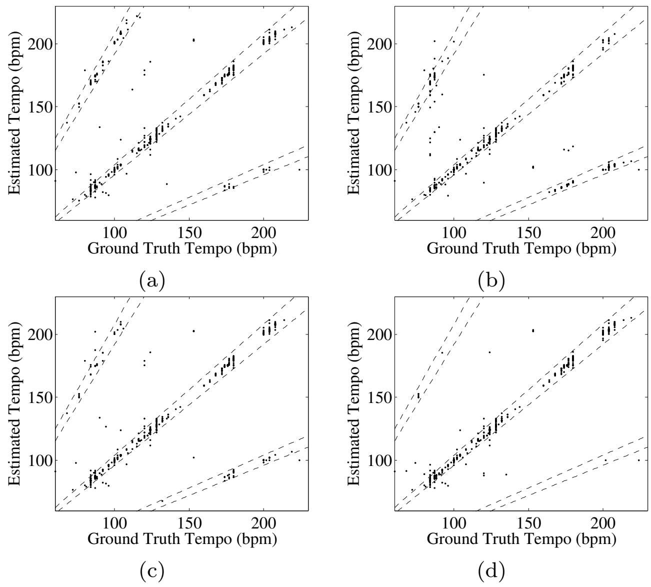

Figure 3: Effect of rhythmic style on tempo classification. Dotted lines indicate $\pm 4\%$ tolerance window for accurate tempo estimation allowing for tapping at the notated tempo, double and half. (a) Tempo estimates without style information; (b) Tempo estimates with Euclidean style classification; (c) Tempo estimates with Adaboost style classification; (d) Tempo estimates given perfect style classification.

Rumba, therefore the peaks of P_R are in very similar locations to those in P_Q, this leaves the Euclidean distance measure unable to rigorously distinguish the two.

Comparing the "Overall T1" row to the "Overall T2" row we can observe a steady convergence of T1 towards T2 as increasingly accurate knowledge of rhythmic style is included. This can be confirmed visually by inspection of the scatter plots of ground truth tempo against estimated tempo in figure 3. Looking in particular at figure 3(d) we can see that, given perfect style information, very few of the estimated values are related to the ground truth by a factor of two. Also, the vast majority of accurate tempi (along the main diagonal) are contained within the $\pm 4\%$ allowance window, suggesting it is an appropriate size for measuring tempo estimation.

4.3 Style vs. Tempo Relationship

Let us now examine the relationship in greater detail. We know the tempo accuracy given the output of the Euclidean distance based classifier (79%) and the tempo accuracy given perfect style information (94%). We now examine the tempo accuracy when style classification accuracy is controlled. We exercise control by forcing a correct classification (i.e. by setting the Euclidean distance to be zero for the known style) for each excerpt with probability p. By allowing p to increase from 0 (where the Euclidean based style classification accuracy is 75%) and 1 (where it is 100%) we can observe how improvements in the classifier would affect tempo accuracy. The relationship between probability of forced classification and the resulting tempo accuracy is shown as the dashed line in figure 4.

To discover whether the mis-classifications for the Euclidean classifier help or hinder the style-dependent tempo estimation, we repeat the controlled experiment but replace the Euclidean distances with white noise. In this scenario when p=0, the style classification will be totally random and when p=1 we will have perfect style classification. This is shown as the solid line in figure 4.

Inspection of figure 4 reveals a number of interesting properties. First, given a completely random style classification, we can still achieve a tempo accuracy of 57%. While less accurate than our baseline tempo estimation algorithm (71%) this is comparable with the "KEA" (63%) the best performing system on this dataset from [1]. The tempo accuracy which uses the ACF pattern based Euclidean distance classification is more accurate than both systems presented by Seyerlehner et al [9] which are marked "S1" and "S2" and correspond to the accuracy using fluctuation patterns and ACF patterns respectively. By comparing the tempo accuracy of S2 (74%) with that resulting from our Adaboost classifier (86%) we can see that our ACF based feature offers better discrimination than that of Seyerlehner et al [9].

The interpretation of the plots of forced classification probability with tempo accuracy using random data (the solid line) and using Euclidean distance from ACF patterns (the dashed line) is less intuitive. The dependent variable is the probability of forced correct classification not the style classification accuracy directly. The ACF pattern plot covers the range of style classification from 75% to 100% where as the random classification plot covers approximately 12.5% (the baseline rate for 8-way classification) to 100%. Incrementing p by 0.01 for the ACF patterns leads to an increase in style classification of $0.01(100\% - 75\%) = 0.25\%$; but for the random classification the increase is $0.01(100\% - 12.5\%) = 0.875\%$. Using this relationship we can find the equivalent point on the solid line to the starting point of the ACF plot; this occurs when $p = (75\% - 12.5\%)/0.875 = 0.715$. For this value of p, the corresponding tempo accuracy is approximately 84%, and is higher than the 79% from the ACF pattern classification. Examined in this way, all points on the ACF plot are lower than the equivalent points on the random classification plot.

In the context of our style-dependent tempo estimation, this demonstrates that the mis-classifications for the Euclidean classifier are more harmful for tempo accuracy than mis-classifying the rhythmic style in a random fashion. We have already observed this limitation of our classifier where many Rumba excerpts were classified QuickStep (see figure 2). This particular mis-classification will almost guarantee an incorrect tempo assignment (or octave error), as the (true) periodicity for a Rumba, which should be close to the nominal value τ_R, will be outside of the range of $W_Q(l)$ from equation (9). We discover that small Euclidean distances in our classifier do not necessarily correspond to small differences in tempo; they can be the result of octave related tempi. The more sophisticated Adaboost classifier however is not so susceptible to this problem.

5 Discussion and Conclusions

Through the results presented we have shown that improvements in tempo estimation for ballroom dance music can be made through a fully automatic clas-

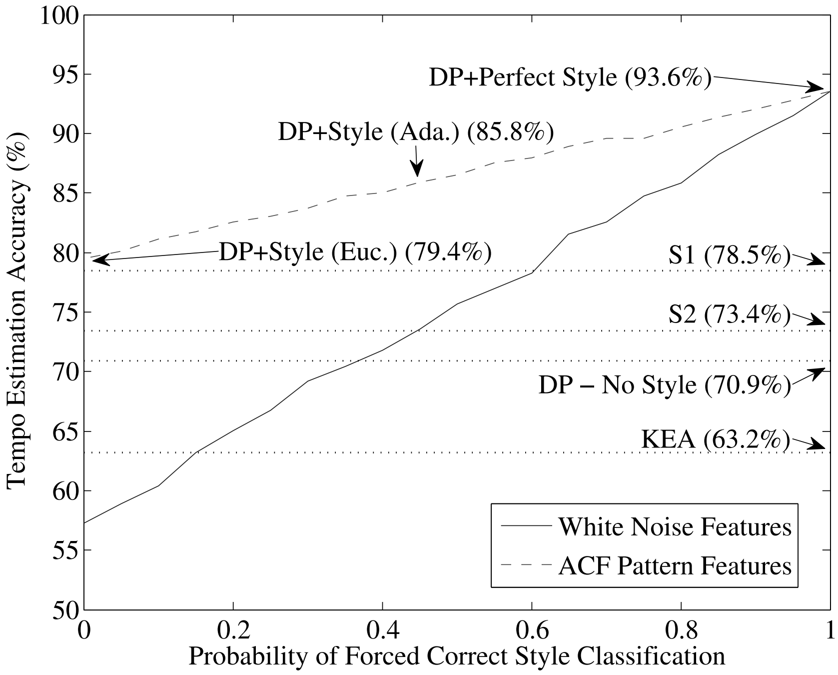

Figure 4: The effect of rhythmic style classification on tempo estimation accuracy. The solid line represents the relationship between style and tempo using random features. The dashed line shows the relationship given our ACF pattern features. DP+Style (Ada.) show the tempo accuracy resulting from the Adaboost classifier. The horizontal dotted lines sown the performance of existing systems: KEA [1], S1 and S2 are the fluctuation pattern approach and ACF pattern approach respectively from [9] and DP – No Style is our baseline tempo estimation algorithm.

sification of rhythmic style. Within the evaluation our main focus has been on the Euclidean distance based classifier rather than the Adaboost classifier despite this being the more successful for this task. We justify this emphasis in the wider context of style-dependent rhythmic analysis. While it is reasonable to perform a cross fold validation in terms of a proof of concept, given a larger real-world collection (perhaps in the order of 10,000 tracks) we would not want to undertake the computational burden of a large scale classification of this nature. We consider being able to summarise particular rhythmic styles by a single ACF pattern with only a small reduction in overall tempo accuracy to be an important result.

It is important to note that this ballroom dataset has certain properties which allow this summarisation to be particularly successful, for example the disjoint distribution of tempi between styles and the constraint of approximately constant tempo for each excerpt. Nevertheless we believe there is scope to extend this approach to a wider variety of signals. The properties of the ballroom dataset allowed us to present this task as one of using style to inform tempo, but in fact we are performing a tempo classification – where the spacing of the peaks of the ACF feature implicitly encode the tempo. Therefore on a wider range of data, where the styles cannot be grouped by tempo (e.g. Jazz or Rock songs cover a wide range of tempi), we would use a several periodicity patterns to cover a small tempo range. In this scenario the style label itself would not be important, rather getting a match to a periodicity pattern close to the correct tempo would be sufficient to improve tempo accuracy. We plan to explore this one aspect of our future work.

Looking beyond tempo extraction we intend to investigate style-dependent rhythmic analysis in a wider context. Collins [14] raises the issue that universal solutions to rhythmic analysis problems do not exist, and that next-generation systems should make greater use of style-specific information. Within our current approach, there is scope to use style related information to aid in the extraction of time-signature (given that the two Waltzes are in 3/4 time, but the remaining styles are in 4/4 time), bar boundaries by using temporal bar patterns (e.g. from Dixon et al [6]) and given both of these pieces of information, recovering style dependent beat locations.

References

[1] F. Gouyon, A. Klapuri, S. Dixon, M. Alonso, G. Tzanetakis, C. Uhle, and P. Cano, "An experimental comparison of audio tempo induction algorithms," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 5, pp. 1832–1844, 2006.

[2] M. F. McKinney, D. Moelants, M. E. P. Davies, and A. Klapuri, "Evaluation of audio beat tracking and music tempo extraction algorithms," *Journal of New Music Research*, vol. 36, no. 1, pp. 1–16, 2007.

[3] J. Laroche, "Efficient tempo and beat tracking in audio recordings," *Journal of the Audio Engineering Society*, vol. 51, no. 4, pp. 226–233, April 2003.

[4] A. M. Stark, M. D. Plumbley, and M. E. P. Davies, "Audio effects for real-time performance using beat tracking," in *Proceedings of the 122nd AES Convention*, Vienna, Austria, May, 5–8 2007, Pre-print 7156.

[5] F. Gouyon, S. Dixon, E. Pampalk, and G. Widmer, "Evaluating rhythmic descriptors for musical genre classification," in *Proceedings of 25th International AES Conference on Semantic Audio*, London, UK, 2004, pp. 196–204.

[6] S. Dixon, F. Gouyon, and G. Widmer, "Towards characterisation of music via rhythmic patterns," in *Proceedings of 5th International Conference on Music Information Retrieval*, Barcelona, Spain, 2004, pp. 509–517.

[7] G. Peeters, "Rhyhm classification using spectral rhythm patterns," in *Proceedings of 6th International Conference on Music Information Retrieval*, London, UK, September 2005, pp. 644–647.

[8] F. Gouyon and S. Dixon, "Dance music classification: A tempo based approach," in *Proceedings of 5th International Conference on Music Information Retrieval*, Barcelona, Spain, 2004, pp. 501–504.

[9] K. Seyerlehner, G. Widmer, and D. Schnitzer, "From rhythm patterns to perceived tempo," in *Proceedings of the 8th International Conference on Music Information Retrieval*, Vienna, Austria, 2007, pp. 519–524.

[10] E. Pampalk, *Computational Models of Music Similarity and their Application to Music Information Retrieval*, Ph.D. thesis, Vienna University of Technology, 2006.

[11] M. E. P. Davies and M. D. Plumbley, "Context-dependent beat tracking of musical audio," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 1009–1020, 2007.

[12] J. P. Bello, C. Duxbury, M. E. Davies, and M. B. Sandler, "On the use of phase and energy for musical onset detection in the complex domain," *IEEE Signal Processing Letters*, vol. 11, no. 6, pp. 553–556, 2004.

[13] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann, San Francisco, 2nd edition, 2005.

[14] N. Collins, "Towards a style-specific basis for computational beat tracking," in *Proceedings of the 9th International Conference on Music Perception and Cognition*, Bologna, Italy, August, 22–26 2006, pp. 461–467.