METER DETECTION IN SYMBOLIC MUSIC USING INNER METRIC ANALYSIS

W. Bas de Haas
Utrecht University
W.B.deHaas@uu.nl

Anja Volk
Utrecht University
A.Volk@uu.nl

ABSTRACT

In this paper we present PRIMA: a new model tailored to symbolic music that detects the meter and the first downbeat position of a piece. Given onset data, the metrical structure of a piece is interpreted using the Inner Metric Analysis (IMA) model. IMA identifies the strong and weak metrical positions in a piece by performing a periodicity analysis, resulting in a weight profile for the entire piece. Next, we reduce IMA to a feature vector and model the detection of the meter and its first downbeat position probabilistically. In order to solve the meter detection problem effectively, we explore various feature selection and parameter optimisation strategies, including Genetic, Maximum Likelihood, and Expectation-Maximisation algorithms. PRIMA is evaluated on two datasets of MIDI files: a corpus of ragtime pieces, and a newly assembled pop dataset. We show that PRIMA outperforms autocorrelation-based meter detection as implemented in the MIDItoolbox on these datasets.

1. INTRODUCTION

When we listen to a piece of music we organise the stream of auditory events seemingly without any effort. Not only can we detect the beat days after we are born [31], as infants we are able to develop the ability to distinguish between a triple meter and duple meter [18]. The processing of metrical structure seems to be a fundamental human skill that helps us to understand music, synchronize our body movement to the music, and eventually contributes to our musical enjoyment. We believe that a system so crucial to human auditory processing must be able to offer great merit to Music Information Retrieval (MIR) as well. But what exactly constitutes meter, and how can models of metrical organisation contribute to typical MIR problems? With the presentation of the PRIMA [1] model we aim to shed some light on these matters in this paper.

The automatic detection of meter is an interesting and challenging problem. Metrical structure has a large influence on the harmonic, melodic and rhythmic structure of a piece, and can be very helpful in many practical situations. For instance, in [30] a statistical exploration of common syncopation patterns in a large corpus of symbolic ragtime pieces is presented. For correct analysis of syncopation patterns knowledge of the meter is essential. However, many corpora lack reliable meter annotations, making automatic meter detection a prerequisite for rhythmic pattern analysis. Similarly, chord recognition algorithms have been shown to improve when metrical information is taken into account, e.g. [3]. Finally, also melodic similarity estimation benefits from (automatically derived) metrical information. Humans appear to be more tolerant to note transformations placed on weak metrical positions [11].

In this paper we present PRIMA: a new model for detecting the meter and the first downbeat in a sequence of onsets. Where most other approaches reduce the problem to a binary duple / triple meter detection, PRIMA estimates all time signatures that are available in the training set and also detects the first downbeat position. PRIMA's architecture is outlined as follows: the model employs Inner Metric Analysis [28, IMA] to determine the strong and weak metrical positions in an onset sequence. The IMA is folded into one-bar profiles, which are subsequently optimised. This metrical analysis feature serves as input to a probabilistic model which eventually determines the meter. Finally, two feature optimisation strategies are discussed and evaluated.

PRIMA is trained and tested on two datasets of MIDI files: the RAG collection [30] and the newly collected FMpop collection. The main motivation for choosing the RAG collection for evaluation is that there is a clear need for meter and first downbeat detection for facilitating corpus-based studies on this dataset. Since Ragtime is a genre that is defined by syncopated rhythms [30], information on meter and the location of the first downbeat is crucial for corpus-based rhythm analyses. In order to assess the flexibility of PRIMA, we also train and evaluate the model on a new dataset of pop music: the FMpop collection. All data has been produced by music enthusiasts and is separated into a test and a training set. Both datasets are too big to manually check all meter annotations. Therefore, we assume that in the training set the majority of the meters are correctly annotated. In the test set, the meter and first downbeat positions are manually corrected, and this confirms the intuition that the majority of the meters is correct, but annotation errors do occur.

Taks description: We define the meter detection task

[1] Probabilistic Reduction of Inner Metric Analysis

as follows: given a series of onsets, automatically detect the time signature and the position of the first beat of the bar. This first beat position is viewed as the offset of the meter measured from the starting point of an analysed segment, and we will refer to this offset as the *rotation* of the meter. [2] After all, a metrical hierarchy recurs every bar, and if the meter is stable, the first beat of the bar can easily be modelled by rotating the metrical grid. In this paper we limit our investigation to the $\frac{2}{2}$, $\frac{2}{4}$, $\frac{4}{4}$, $\frac{3}{4}$, $\frac{6}{8}$, $\frac{12}{8}$ meters that occur at least in 40 pieces of the dataset (five different meters in the RAG, four in the FMpop Collection). Naturally, additional meters can be added easily. In the case of duple / triple classification $\frac{2}{2}$, $\frac{2}{4}$, and $\frac{4}{4}$ are considered duple meters and $\frac{3}{4}$, $\frac{6}{8}$, and $\frac{12}{8}$ are considered triple meters. Within this study we assume that the meter does not change throughout an analysed segment, and we consider only MIDI data.

Contribution: The contribution of this paper is threefold. First, we present a new probabilistic model for automatically detecting the meter and first downbeat position in a piece. PRIMA is conceptually simple, based on a solid metrical model, flexible, and easy to train on style specific data, Second, we present a new MIDI dataset containing 7585 pop songs. Furthermore, for small subsets of this new FMpop Collection and a collection of ragtime pieces, we also present new ground-truth annotations of the meter and rotation. Finally, we show that all variants of PRIMA outperform the autocorrelation-based meter detection implemented in the MIDItoolbox [5].

2. RELATED WORK

The organisation of musical rhythm and meter has been studied for decades, and it is commonly agreed upon that this organisation is best represented hierarchically [13]. Within a metrical hierarchy strong metrical positions can be distinguished from weaker positions, where strong positions positively correlate with the number of notes, the duration of the notes, the number of equally spaced notes, and the stress of the notes [16]. A few (computational) models have been proposed that formalise the induction of metrical hierarchies, most notable are the models of Steedman [20], Longuet-Higgins & Lee [14] Temperley [21], and Volk [28]. However, surprisingly little of this work has been applied to the automatic detection of the meter (as in the time signature) of a piece of music, especially in the domain of symbolic music.

Most of the work in meter detection focusses on the audio domain and not on symbolic music. Although large individual differences exists, in the audio domain the meter detection systems follow a general architecture that consists of a feature extraction front-end and a model that accounts for periodicities in the onset or feature data. In the front-end typically features are used that are associated with *onset detection* such as spectral difference, or flux, and energy spectrum are used [1]. Or, in the symbolic case, one simply assumes that onset data is available [9, 22], like we do in this paper.

After feature extraction the periodicity of the onset data is analysed, which is typically done using auto-correlation [2, 23], a (beat) similarity matrix [6, 8], or hidden Markov models [17, 12]. Next, the most likely meter has to be derived from the periodicity analysis. Sometimes statistical machine learning techniques, such as Gaussian Mixture Models, Neural Networks, or Support Vector Machines [9], are applied to this task, but this is less common in the symbolic domain. The free parameters of these models are automatically trained on data that has meter annotations. Frequently the meter detection problem is simplified to classifying whether a piece uses a *duple* or *triple* meter [9, 23], but some authors aim at detecting more fine-grained time signatures [19, 24] and can even detect odd meters in culturally diverse music [10]. Albeit focussed on the audio domain, for a relatively recent overview of the field we refer to [24].

2.1 Inner Metric Analysis

Similar to most of the meter detection systems outlined in the previous section PRIMA relies on periodicity analysis. However, an important difference is that it uses the Inner Metric Analysis [28, IMA] instead of the frequently used autocorrelation. IMA describes the *inner* metric structure of a piece of music generated by the actual onsets opposed to the *outer* metric structure which is associated with an abstract grid annotated by a time signature in a score, and which we try to detect automatically with PRIMA.

What distinguishes IMA from other metrical models, such as Temperley's Grouper [21], is that IMA is flexible with respect to the number of metric hierarchies induced. It can therefore be applied both to music with a strong sense of meter, e.g. pop music, and to music with less pronounced or ambiguous meters. IMA has been evaluated in listening experiments [25], and on diverse corpora of music, such as classical pieces [26], rags [28], latin american dances [4] and on 20th century compositions [29].

IMA is performed by assigning a *metric weight* or a *spectral weight* to each onset of the piece. The general idea is to search for all chains of equally spaced onsets within a piece and then to assign a weight to each onset. This chain of equally spaced onsets underlying IMA is called a *local meter* and is defined as follows. Let On denote the set of all onsets of notes in a given piece. We define every subset $m \subset On$ of equally spaced onsets to be a local meter if it contains at least three onsets and is not a subset of any other subset of equally spaced onsets. Each local meter can be identified by three parameters: the starting position of the first onset s, the period denoting the distance between consecutive onsets d, and the number of repetitions k of the period (which equals the size of the set minus one).

The metric weight of an onset o is calculated as the weighted sum of the length k_m of all local meters m that coincide at this onset ($o \in m$), weighted by parameter p that regulates the influence of the length of the local meters on the metric weight. Let $M(\ell)$ be the set of all local

[2] We chose the new term *rotation* for the offset of the meter because the musical terms generally used to describe this phenomenon, like *anacrusis*, *upbeat figure*, or *pickup*, are sometimes interpreted differently.

meters of the piece of length at least ℓ, then the metric weight of an onset, $o \in On$, is defined as follows:

$$W_{\ell,p}(o) = \sum_{\{m \in M(\ell): o \in m\}} k_m^p. \quad (1)$$

The spectral weight is calculated in a similar fashion, but for the spectral weight each local meter is extended throughout the entire piece. The idea behind this is that the metrical structure induced by the onsets stretches beyond the region in which onsets occurs. The extension of a local meter m is defined as $ext(m_{s,d,k}) = \{s + id, \forall i\}$ where i is an integer number. For all discrete metrical positions t, regardless whether it contains an onset or not, the spectral weight is defined as follows:

$$SW_{\ell,p}(t) = \sum_{\{m \in M(\ell): t \in ext(m)\}} k_m^p. \quad (2)$$

In this paper we have used the standard parameters $p = 2$, and $\ell = 2$. Hence, we consider all local meters that exist in a piece. A more elaborate explanation of the IMA including examples can be found in [28].

3. IMA BASED METER DETECTION

In this section we will outline the PRIMA model in a bottom-up fashion. We start with the input MIDI data, and describe how we transform this into onset data, perform IMA and finally optimise a feature based on IMA. Next, we explain how this feature is used in a probabilistic model to detect the meter and rotation of sequence of onsets, and we elaborate on two different training strategies.

3.1 Quantisation and Preprocessing

Before we can perform IMA, we have to preprocess the MIDI files to obtain a quantised sequence of onsets. The following preprocessing steps are taken:

To be able to find periodicities, the onset data should be quantised properly. Within Western tonal music duple as well as triple subdivisions of the beat occur commonly. Hence, we use a metrical grid consisting of 12 equally spaced onset positions per quarter note. With this we can quantise both straight and *swung* eight notes. Here, *swing* refers to the characteristic long-short rhythmical pattern that is particularly common in Jazz, but is found throughout popular music.

In the quantisation process we use the length of a quarter note as annotated in the MIDI file. This MIDI *time division* specifies the number of MIDI ticks per quarter note and controls the resolution of the MIDI data. Because the MIDI time division is constant, strong tempo deviations in the MIDI data might distort the quantisation process and the following analyses. To estimate the quality of the alignment of the MIDI data to the metrical grid, we collect the quantisation deviation for every onset, and the average quantisation deviation divided by the MIDI time division gives a good estimate of the quantisation error. To make sure that the analysed files can be quantised reasonably well, we discard all MIDI files with an average quantisation error higher than 2 percent.

After quantising the MIDI data, we collect all onset data from all voices and remove duplicate onsets. Next, the MIDI data is segmented at all positions where a meter change is annotated in the MIDI file. Segments that are empty or shorter than 4 bars are excluded from further analysis. Also, MIDI files that do not contain meter annotations at all are ignored in the training phase.

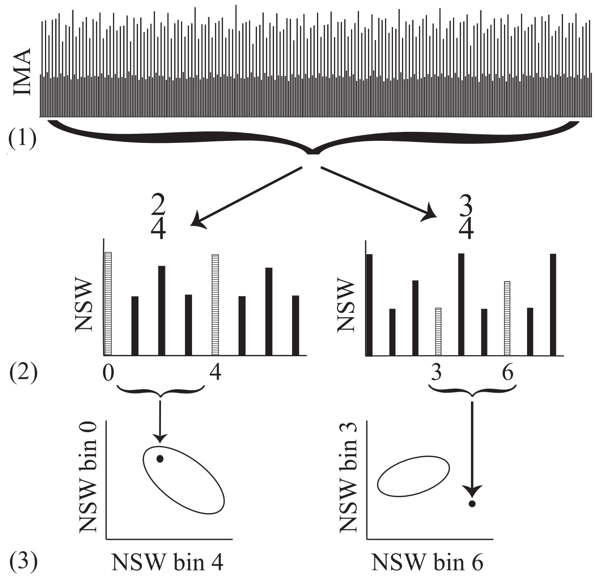

Figure 1. The construction of NSW profiles for a piece in $\frac{2}{4}$: (1) displays IMA, (2) displays the NSW profiles derived from IMA for a $\frac{2}{4}$ and a $\frac{3}{4}$ meter, and (3) shows how two bins are selected from each profile and used to estimate the probability of that meter. The ellipse represents the Gaussian distribution fitted to selected bins of the NSW profiles in the training phase. Note that the $\frac{3}{4}$ NSW profile does not resemble a typical $\frac{3}{4}$ and receives a low probability. Also, the selected bins may differ per optimisation strategy.

3.2 Normalised spectral weight profiles

We use the spectral weights of IMA to construct a feature for detecting the meter in a piece. More specifically, this feature will model the conditional probability of a certain meter given a sequence of onsets. As we will explain in more detail in the next section, the distribution of these features will be modelled with a Gaussian distribution. We call this feature a *Normalised Spectral Weight* (NSW) profile, and discern three stages in constructing them: (1) perform IMA, (2) folding the IMA in one-bar profiles and normalising the weights profiles, and (3) selecting the most relevant bins for modelling the Gaussian distribution. These three stages are displayed schematically in Figure 1, and are detailed below.

IMA marks the metrical importance of every quantised onset position in a piece. Because of the large numbers of spectral weights and the large differences per piece, IMA cannot be used to detect the meter directly. However, we can fold the analysis into one-bar profiles to get a more concise metrical representation for every candidate meter. These one-bar profiles are created by summing the spectral weights per quantised position within a bar. Consequently, the shape of these profiles is determined by the meter (beats per bar), the length of piece (longer pieces yield higher spectral weights), and the number of quantisation bins.

We normalise spectral weights in the one-bar profiles by applying Formula 3:

$$\text{normalise}(w) = \log(\frac{w}{n^{\beta}} + \alpha) \qquad (3)$$

Here, w is the summed spectral weight of a particular quantised beat position and n is the number of bars used to create the profile. We use a parameter β to control the effect of the length of the piece in the normalisation. Furthermore, because many quantised beat positions might have a summed metrical weight of 0, and this will cause problems when we fit Gaussian models to these profiles, we use Laplace smoothing [15, p. 260] and add a constant factor α to all weights. Finally, because statistical analysis of large amounts of profiles showed that differences in weights are distributed normally on a logarithmic scale, we apply the natural logarithm in Eq. 3. For the results in this report we have used $\beta = 2$ and $\alpha = 1$. We call these profiles *Normalised Spectral Weight* (NSW) Profiles.

The raw NSW profiles cannot yet be conveniently used as a feature for meter detection: the dimensionality of the NSW profiles is relatively high, and the dimensionality differs per meter. Also, not every metrical position within a bar is equally important. For instance, the first beat of the bar will have a high spectral weight, while the metrical position of the second eighth note will generally have a much lower spectral weight. Hence, we select profile bins that contain the information most relevant for meter detection.

The selection of the relevant profile bins is a special case of feature dimensionality reduction where the feature bins are highly dependent on each other. In this section we introduce two selection methods that will be experimentally verified in Section 4.3. A first intuition is to select the n profile bins that have the highest weights on average for a given dataset. A brief analysis showed that these bins roughly correspond to the first n principal components. However, a preliminary evaluation shows that NSW profiles containing only these bins perform relatively poorly. Hence, in order to learn more about what are the distinctive bins in the profiles, we use a Genetic Algorithm (GA) to explore the space of possible bin selections. [3] When we analyse these bin selections, we notice that the GA selects bins for a meter that contain weights that are maximally different to other meters.

[3] Note that these n profile bins may differ between meters, and n does not have to be the same for all meters, as long as the bin selection of the examined profile is exactly the same as the selection used in the template profile for that meter. For the implementation of the GA we use the Haskell library `https://github.com/boegel/GA`, using a population size of 100 candidates, a crossover rate of 0.7, a mutation rate of 0.2, and Eq. 5 as fitness function.

Training a GA on large amounts of data takes a lot of time, even if the NSW profiles are pre-calculated. Since we have a clear intuition about how the GA selects profile bins, we might be able to mimic this behaviour without exploring the complete space of possible bin selections. Recall when we classify a single piece, we calculate multiple NSW profiles for a single IMA: one for each meter. If we select the same bins in each profile for matching, i.e. every first beat of a bar, the chances are considerable that this selection will match multiple meters well. Hence, we select the n bins of which the NSW profiles of the ground-truth meter are maximally different from the NSW profiles of other meters. In this calculation we define maximally different as having a maximal absolute difference in spectral weight, and n does not differ between meters. We call this method the *Maximally Different Bin* (MDB) Selection.

3.3 A probabilistic meter classification model

To restate our initial goal: we want to determine the meter and its rotation given a sequence of note onsets. Ignoring rotation, a good starting point is to match NSW profiles with template profiles of specific meters. However, although the spectral weights of IMA reflect human intuitions about the musical meter [27], it is rather difficult to design such templates profiles by hand. Moreover, these template profiles might be style specific. Hence, we propose a model that learns these templates from a dataset.

Another style dependent factor that influences meter detection is the distribution of meters in a dataset. For instance, in Ragtime $\frac{2}{2}$ and $\frac{2}{4}$ occur frequently, while pop music is predominantly notated in a $\frac{4}{4}$ meter. Just matching NSW profiles with meter templates will not take this into account. When we combine simple profile matching with a weighting based on a meter distribution (prior), this conceptually equals a Naive Bayes classifier [15]. Therefore, probabilistically modelling meter detection is a natural choice.

If we ignore the rotations for sake of simplicity, we can express the probability of a meter given a set of note onsets with Equation 4:

$$P(\text{meter}|\text{onsets}) \propto P(\text{onsets}|\text{meter}) \cdot P(\text{meter}) \qquad (4)$$

Here, $P(\text{onset}|\text{meter})$ reflects the probability of an onset sequence given a certain meter, and $\propto$ denotes "is proportional to". Naturally, certain meters occur more regularly in a dataset than others which is modelled by $P(\text{meter})$. The conditional probability $P(\text{onset}|\text{meter})$ can be estimated using NSW profiles. Given a piece and a specific meter we create an NSW profile that can be used as multidimensional feature. Given a large dataset that provides us with sequences of onsets and meters, we can model the distribution of the NSW profiles as Gaussian distributions. For every meter in the dataset we estimate the mean and covariance matrix of a single Gaussian distribution with the expectation-maximization algorithm [7]. The prior probability of a certain meter, $P(\text{meter})$, can be estimated with maximum likelihood estimation, which equals the number of times a certain meter occurs in a dataset divided by the total number of meters in the dataset.

Adding the estimation of the rotation makes the problem slightly more complicated. A natural way of incorporating rotation is to add it as a single random variable that is dependent on the meter. This makes sense because it is likely that the kind of rotation depends on the kind of meter: an anacrusis in a $\frac{4}{4}$ meter is likely to differ from an anacrusis in a $\frac{3}{4}$ meter. Hence, we can transform Eq. 4 into the following equation:

$$P(x, r|y) \propto P(y|x, r) \cdot P(r|x) \cdot P(x) \tag{5}$$

Similar to Eq. 4, we estimate the meter x given an onset pattern y, but now we also add the rotation r. The term $P(y|x, r)$ can again be modelled with NSW profiles, but now the profiles should also be rotated according to the rotation r. The term $P(x)$ reflects the probability of a meter and can be estimated with maximum likelihood estimation.

We do not consider all possible rotations. For a $\frac{4}{4}$ meter there are $4 \cdot 12 = 48$ possible rotations, many of which are not likely to occur in practise. The rotations are modelled as a fraction of a bar, making the rotation meter independent. Furthermore, we rotate clock-wise, e.g. $\frac{1}{4}$ represents an anacrusis of one quarter note in a $\frac{4}{4}$ meter. The space of possible rotations can be further reduced by only considering two kinds of rotations: rotations for duple and triple meters. After all, given the very similar metrical structure of $\frac{2}{4}$ and $\frac{4}{4}$, we expect that the rotations will be similar as well (but on another absolute metrical level, e.g. eighth instead of quarter notes). For duple meters we explore eight, and for triple meters we explore six different rotations.

Unfortunately, estimating the prior probability of the rotation given a certain meter, i.e. $P(r|x)$, is not trivial because we rely on MIDI data in which the rotation is not annotated. Hence, we need another way of estimating this prior probability of the rotation. We estimate the rotation by calculating all rotations of the NSW profiles and pick the rotation that maximises probability of the annotated ground-truth meter. Having an estimation of the best fitting rotation per piece, we can perform maximum likelihood estimation by counting different rotations for each meter in order to obtain the rotation probabilities.

3.4 Training

We train and evaluate PRIMA on two datasets (see Sec. 4.1 and 4.2). These datasets consist of MIDI files created by music enthusiasts that might have all sorts of musical backgrounds. Hence, it is safe to assume that the meter annotations in these MIDI files might sometimes be incorrect. A likely scenario is, for instance, that MIDI creation software adds a $\frac{4}{4}$ meter starting at the first note onset by default, while the piece in question starts with an upbeat and is best notated in $\frac{3}{4}$. Nevertheless, we assume that the *majority* of the meters is annotated correctly, and that incorrect meters will only marginally effect the training of PRIMA.

In this paper we evaluate two different ways of training PRIMA. We use Maximally Different Bin (MDB) selection in the feature training phase, or alternatively, we use a GA to select the most salient NSW profile bins. After the bin selection, we use Maximum Likelihood estimation to learn the priors and rotation, as described in the previous section, and Expectation-Maximisation for fitting the Gaussian distributions.

4. EVALUATION

To assess the quality of the meter and rotations calculated by PRIMA, we randomly separate our datasets into testing and training sets. The test sets are manually corrected and assured to have a correct meter and rotation. The next two sections will detail the data used to train and evaluate PRIMA. The manual inspection of the meters and rotations confirms the intuition that most of the meters are correct, but the data does contain meter and rotation errors.

4.1 RAG collection

The RAG collection that has been introduced in [30] currently consists of 11545 MIDI files of ragtime pieces that are collected from the Internet by a community of Ragtime enthusiasts. The collection is accompanied by an elaborate compendium [4] that stores additional information about individual ragtime compositions, like year, title, composer, publisher, etc. The MIDI files in the RAG collection describe many pieces from the ragtime era (approx. 1890 $\sim$ 1920), but also modern ragtime compositions. The dataset is separated randomly in a test set of 200 pieces and a training set of 11345 pieces. After the preprocessing detailed in Sec. 3.1, 74 and 4600 pieces are considered suitable for respectively testing and training. For one piece we had to correct the meter and for another piece the rotation.

4.2 FMpop collection

The RAG corpus only contains pieces in the ragtime style. In order to study how well PRIMA predicts the meter and rotation of regular pop music, we collected 7585 MIDI files from the website Free-Midi.org. [5] This collection comprises MIDI files describing pop music from the 1950 onwards, including various recent hit songs, and we call this collection the FMpop collection. For evaluation we randomly select a test set of 200 pieces and we use the remainder for training. In the training and test sets, 3122 and 89 pieces successfully pass the preprocessing stage, respectively. Most of the pieces that drop out have a quantisation error greater than 2 percent. For three pieces we had to correct the meter, and for four pieces the rotation.

4.3 Experiments

We perform experiments on both the RAG and the FMpop collections in which we evaluate the detection performance by comparing the proportion of correctly classified meters, rotations, and the combinations of the two. In these experiments we probe three different training variants of PRIMA: (1) a variant where we use Maximally Different Bin (MDB)

[4] see http://ragtimecompendium.tripod.com/ for more information

[5] http://www.free-midi.org/

RAG Collection

(Training) model	Meter	Rotation	Both
Duple / Triple meters			
MIDItoolbox	.76	–	–
MDB selection (2 bins)	.97	.88	.86
MDB selection (3 bins)	.97	.97	.95
GA optimized	.97	.99	**.96**
Meters: $\frac{2}{2}$, $\frac{2}{4}$, $\frac{3}{4}$, $\frac{4}{4}$, $\frac{6}{8}$			
MDB selection (2 bins)	.85	.92	.80
MDB selection (3 bins)	.80	.92	.76
GA optimised	.84	.93	**.82**

Table 1. The proportion of correctly detected meter and rotation in the RAG collection. The first section shows the duple / triple meter classification, the second section shows the proportions for the five most used time signatures.

FMpop Collection

(Training) model	Meter	Rotation	Both
Duple / Triple meters			
MIDItoolbox	.74	–	–
MDB selection (2 bins)	.94	.90	**.85**
MDB selection (3 bins)	.90	.93	.84
GA optimized	.94	.88	.83
Meters: $\frac{3}{4}$, $\frac{4}{4}$, $\frac{6}{8}$, $\frac{12}{8}$			
MDB selection (2 bins)	.94	.81	.79
MDB selection (3 bins)	.94	.81	.78
GA optimised	.94	.91	**.87**

Table 2. The correctly detected proportion for on the FMpop collection for duple / triple meter classification and for the four most used time signatures.

selection in which we select the two most salient bins and (2) a variant in which we select the three most salient bins. Finally, (3) we also use a Genetic algorithm to select the bins and estimate the rotation priors.

To place the performance of PRIMA into context, we compare the results to the meter detection model implemented in the MIDItoolbox [5]. This model only predicts whether a meter is duple or triple and does not predict the time signature. Therefore, we can compare the MIDItoolbox meter finding to PRIMA only in the duple / triple case. To ensure we use the exact same input data, we have written our own NMAT export script that transforms the MIDI as preprocessed by PRIMA into a matrix that can be parsed by the MIDItoolbox. All source code and data reported in this study is available on request.

4.4 Results

We evaluate the performance PRIMA and its different training strategies on duple / triple meter detection and the detection of five different time signatures. In Table 1 the proportions of correctly detected meters in the RAG collection are displayed. In the duple / triple meter detection experiments all variants of PRIMA outperform the MIDItoolbox meter detection. We tested the statistical significance of all individual differences between MIDItoolbox meter detection and PRIMA using McNemar's χ^2 test, and all differences are significant ($p < 0.001$). In the classification of five different time signatures the performance drops considerably. However, rags are mostly notated in $\frac{2}{4}$, $\frac{4}{4}$, and $\frac{2}{2}$ meters, and even experienced musicians have difficulty determining what is the correct meter. Still PRIMA achieves a 96 percent correct estimation for meter and rotation in the duple / triple experiment and 82 percent correct estimation on the full time signature detection.

In Table 2 the proportions of correctly classified meters in the FMpop Collection are displayed. Also on onsets extracted from popular music, PRIMA outperforms the MIDItoolbox meter finding. Again, we tested the statistical significance of the differences between all PRIMA variants using McNemar's χ^2 test, and all differences are statistically significant ($p < 0.002$ for GA and MDB selection (2 bins), and $p < 0.017$ for MDB selection (3 bins)). Overall, PRIMA's performance on the FMpop Collection is lower than on the RAG Collection for the duple / triple detection, but higher for time signature detection. Respectively, 85 and 87 percent correct classification is achieved for both meter and rotation. Generally, the GA seems to yield the best results.

5. DISCUSSION AND CONCLUSION

We presented a new model for detecting the meter and first downbeat position of a piece of music. We showed that IMA is valuable in the context of meter and first downbeat detection. PRIMA is flexible, can be easily trained on new data, and is conceptually simple. We have shown that PRIMA performs well on the FMpop and RAG Collections and outperforms the MIDItoolbox meter finding model. However, while PRIMA can be trained on data of specific styles, the parameters of the MIDItoolbox meter detection model are fixed. Hence, the performance of the MIDItoolbox should be seen as a baseline system that places PRIMA's results into context.

In this study we applied PRIMA to MIDI data only because we believe that corpus based analyses on collections like the RAG collection can really benefit from meter finding. Nevertheless, PRIMA's IMA based feature and probabilistic model are generic and can be easily applied to onset sequences extracted from audio data. Hence, it would be interesting to investigate how PRIMA model performs on audio data, and compare it to the state-of-the-art in audio meter detection. We strongly believe that also in the audio domain meter detection can benefit from IMA. We are confident that IMA has the potential to aid in solving many MIR tasks in both the audio and the symbolic domain.

6. ACKNOWLEDGMENTS

A. Volk and W.B. de Haas are supported by the Netherlands Organization for Scientific Research, through the NWO-VIDI-grant 276-35-001 to Anja Volk.

7. REFERENCES

[1] J.P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M.B. Sandler. A tutorial on onset detection in music signals. *IEEE Trans. ASLP*, 13(5):1035–1047, 2005.

[2] J.C. Brown. Determination of the meter of musical scores by autocorrelation. *JASA*, 94(4):1953–1957, 1993.

[3] R. Chen, W. Shen, A. Srinivasamurthy, and P. Chordia. Chord recognition using duration-explicit hidden markov models. In *Proc. ISMIR*, pages 445–450, 2012.

[4] E. Chew, A. Volk, and C.-Y. Lee. Dance music classification using inner metric analysis. In *The next wave in computing, optimization, and decision technologies*, pages 355–370. Springer, 2005.

[5] T. Eerola and P. Toiviainen. *MIDI Toolbox: MATLAB Tools for Music Research*. University of Jyväskylä, Jyväskylä, Finland, 2004.

[6] J. Foote and S. Uchihashi. The beat spectrum: a new approach to rhythm analysis. *Proc. IEEE ICME*, 0:224–228, 2001.

[7] C. Fraley and A. E. Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458):611–631, 2002.

[8] M. Gainza. Automatic musical meter detection. In *Proc. ICASSP*, pages 329–332, April 2009.

[9] F. Gouyon and P. Herrera. Determination of the meter of musical audio signals: Seeking recurrences in beat segment descriptors. In *Proc. AES*. Audio Engineering Society, 2003.

[10] A. Holzapfel, F. Krebs, and A. Srinivasamurthy. Tracking the "odd": Meter inference in a culturally diverse music corpus. In *Proc. ISMIR*, pages 425–430, 2014.

[11] P. van Kranenburg, A. Volk, F. Wiering, and R. C. Veltkamp. Musical models for folk-song alignment. In *Proc. ISMIR*, pages 507–512, 2009.

[12] F. Krebs, S. Böck, and G. Widmer. Rhythmic pattern modeling for beat and downbeat tracking in musical audio. In *Proc. ISMIR*, pages 227–232, 2013.

[13] F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, 1996.

[14] H. C. Longuet-Higgins and C. S. Lee. The rhythmic interpretation of monophonic music. *Music Perception*, 1(4):424–441, 1984.

[15] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.

[16] C. Palmer and C. L. Krumhansl. Mental representations for musical meter. *JEPHPP*, 16(4):728, 1990.

[17] G. Peeters and H. Papadopoulos. Simultaneous beat and downbeat-tracking using a probabilistic framework: Theory and large-scale evaluation. *IEEE Trans. ASLP*, 19(6):1754–1769, Aug 2011.

[18] J. Phillips-Silver and L. J. Trainor. Feeling the beat: Movement influences infant rhythm perception. *Science*, 308(5727):1430, 2005.

[19] A. Pikrakis, I. Antonopoulos, and S. Theodoridis. Music meter and tempo tracking from raw polyphonic audio. In *Proc. ISMIR*, pages 192–197, 2004.

[20] M. J. Steedman. The perception of musical rhythm and metre. *Perception*, 6(5):555–569, 1977.

[21] D. Temperley. *The Cognition of Basic Musical Structures*. Cambridge, MA, MIT Press, 2001.

[22] P. Toiviainen and T. Eerola. The role of accent periodicities in meter induction: A classification study. In *Proc. ICMPC*, pages 422–425, 2004.

[23] P. Toiviainen and T. Eerola. Autocorrelation in meter induction: The role of accent structure. *JASA*, 119(2):1164–1170, 2006.

[24] M. Varewyck, J.-P. Martens, and M. Leman. Musical meter classification with beat synchronous acoustic features, DFT-based metrical features and support vector machines. *JNMR*, 42(3):267–282, 2013.

[25] A. Volk. The empirical evaluation of a mathematical model for inner metric analysis. In *Proc. ESCOM*, pages 467–470, 2003.

[26] A. Volk. Metric investigations in Brahms symphonies. In G. Mazzola, T. Noll, and E. Lluis-Puebla, editors, *Perspectives in Mathematical and Computational Music Theory*, pages 300–329. epOs Music, Osnabrück, 2004.

[27] A. Volk. Persistence and change: Local and global components of metre induction using inner metric analysis. *J. Math. Music*, 2(2):99–115, 2008.

[28] A. Volk. The study of syncopation using inner metric analysis: Linking theoretical and experimental analysis of metre in music. *J. New Music Res.*, 37(4):259–273, 2008.

[29] A. Volk. Applying inner metric analysis to 20th century compositions. In *MCM 2007*, CCIS 37, pages 204–210. Springer, 2009.

[30] A. Volk and W. B. de Haas. A corpus-based study on ragtime syncopation. In *Proc. ISMIR*, pages 163–168, 2013.

[31] I. Winkler, G. P. Háden, O. Ladinig, I. Sziller, and H. Honing. Newborn infants detect the beat in music. *PNAS*, 106(7):2468–2471, 2009.