

In Search of Automatic Rhythm Analysis Methods for Turkish and Indian Art Music

Ajay Srinivasamurthy, André Holzapfel and Xavier Serra

[This is an Author's Original Manuscript of an Article whose final and definitive form, the Version of Record, has been published in the Journal of New Music Research, Volume 43, Issue 1, 31 Mar 2014, available online at: <http://www.tandfonline.com/doi/full/10.1080/09298215.2013.879902>]

Abstract

The aim of this paper is to identify and discuss various methods in computational rhythm description of Carnatic and Hindustani music of India, and Makam music of Turkey. We define and describe three relevant rhythm annotation tasks for these cultures - beat tracking, meter estimation, and downbeat detection. We then evaluate several methodologies from the state of the art in Music Information Retrieval (MIR) for these tasks, using manually annotated datasets of Turkish and Indian music. This evaluation provides insights into the nature of rhythm in these cultures and the challenges to automatic rhythm analysis. Our results indicate that the performance of evaluated approaches is not adequate for the presented tasks, and that methods that are suitable to tackle the culture specific challenges in computational analysis of rhythm need to be developed. The results from the different analysis methods enable us to identify promising directions for an appropriate exploration of rhythm analysis in Turkish, Carnatic and Hindustani music.

Keywords: Musical meter, Rhythm, Beat tracking, Makam music of Turkey, Hindustani music, Carnatic music

1 Introduction

If we ask American music listeners to sing along to a Karaoke version of an American pop song, they are likely to comprehend at which moment of the song they are supposed to start singing. If,

however, we confront the same group of listeners with some Turkish songs, they are likely to have difficulties in synchronizing their performance with the pieces due to the unknown rhythms in the presented repertoire (Hannon, Soley, & Ullal, 2012). When we ask a specialist in composition at a conservatory about relations between rhythm in Turkish music and the music of European Renaissance period, a shrug of the shoulders paired with an inquiring gaze is the most likely answer. When we gather some specialists in Music Information Retrieval (MIR) and ask them to evaluate their computational methods for rhythm analysis on Indian and Turkish music, what result can we expect?

While, admittedly, this question appears a bit too suggestive, addressing it is motivated by some practical needs. There are no culture specific tools available for the discovery, understanding, and an enriched listening of music in many musical cultures even for the main target audience in that specific culture. Further, music listeners all over the world show increasing interest in expanding their music collections towards cultures not originally familiar to them, and musicians are increasingly seeking cooperation and influences from other cultures. This has motivated the quest for a World music science (Tenzer, 2006), which should approach music of all cultures in a more unified way. Therefore, providing musicians, listeners and researchers with computational tools that can help them to approach music in a more unified way are of high value. Tools for the analysis of rhythm, as one important aspect of music analysis (LaRue, 1992), are the focus of this paper.

As observed already many decades ago, discussions about rhythm tend to suffer from inconsistencies in their terminology (Sachs, 1953). Let us therefore try to locate definitions for some basic terms, in order to establish a consistent representation in this paper. György Ligeti, a European composer who showed a high interest in rhythmic structures in music of African cultures, defined rhythm as “every temporal sequence of notes, sounds and musical *Gestalten*”, while he referred to meter as “a more or less regular structuring of temporal development” (Ligeti, 2007). According to that definition, rhythm is contained in all kinds of music, while pulsating rhythm which is subordinated to a meter is not found in all music. Kolinski (1973) describes the regular structuring caused by meter as organized pulsation, which functions as a framework for rhythmic design. Pulsations in meter are organized into hierarchical levels of differing time spans, usually demanding the presence of pulsation at least on three levels; these levels are referred to as subdivisions, beats, and measures - from short- to long time-span (London, 2001). The pulsation at the beat level was referred to as primary rhythmic level by Cooper and Meyer (1960), and they define it as the lowest level on which a complete rhythmic group is realized. The same authors identify this level with a subjective notion as well, by

referring to it as the level at which the beats are felt and counted. As listeners tend to count the beats at varying levels (Moelants & McKinney, 2004; Parncutt, 1994), and what can be considered a complete rhythmic group can be argued as well, we are confronted with a significant amount of ambiguity in determining this level for a piece of music. Finding a clear terminology is further hampered by the fact that the pulsation at the beat level is commonly also referred to as “beat” or “pulse” as observed by London (2004, 2001). A shortcoming of most attempts to describe rhythm and meter is the assumption about pulsation to consist of recurring, precisely equivalent and equally-spaced stimuli (Lerdahl & Jackendoff, 1983). Such preconditions cause difficulties when analyzing music of other cultures since in many cases, mutual equidistance becomes an exception rather than the rule, taking the important role of additive meters in Greek and Turkish music as one example.

In order to obtain a consistent terminology, we consider meter as being an organization of pulsation into different levels related to the time-spans of the individual pulsations. Note that this may include an irregular pattern of unequal time-spans at some level, *e.g.* due to the presence of an additive meter. We will consider pulsations on three levels. On the (lower) subdivision level, we will refer to subdivision pulsation or subdivision pulses, depending on if we refer to the time series that constitutes the pulsation or to the individual instances of the pulsation, respectively. On the beat level, we will differentiate between the beat pulsation, and beats as its instances (instead of the inconvenient term of “beat pulses”). On the measure level, the term pulsation is not appropriate due to the often larger time-span at this level. Therefore, we use the notions of measure length to describe the time-span at this level and downbeat as the beginning of a measure. The overall structure of a meter is defined by the time-span relations between these three levels. Typically, the time-span relation between measure- and beat level is denoted as the measure length (in beats), and the relation between beat- and subdivision level as the subdivision meter (in subdivision pulses). It was observed by *e.g.* Clayton (2000) that meter in music causes a transformation of time from linear development to a repetitive and cyclic structure. This happens because its levels of pulsation constitute what we want to refer as rhythmic cycles. Throughout the text, we put emphasis on terms containing the notion of a cycle (such as the measure cycle for the cycle repeating on every downbeat), a notion suitable in the musical context of both Turkish and Indian music.

Automatic approaches that aim at annotating pieces of music with certain aspects of rhythm can provide users with an understanding of the temporal structure in music at various levels. Müller et al. (2011) discuss

the estimation of beat pulsation, *beat tracking*, and the estimation of higher-level rhythmic structures such as the measure length. Approaches such as the one presented by Klapuri, Eronen, and Astola (2006) aim at estimating structures at several metrical levels at the same time in order to track the development of the metrical grid of a piece. Other approaches concentrate on specific aspects, such as beat tracking (Ellis, 2007), estimating the type of meter (Pikrakis, Antonopoulos, & Theodoridis, 2004) in terms of a time signature, or tracking the downbeat locations in a piece of music (Davies & Plumbley, 2006; Hockman, Davies, & Fujinaga, 2012), which is commonly referred to as downbeat detection. Most presented approaches assume the presence of clear cues for a regular metrical grid in the surface structure of the music. Hence, they are reported to work reasonably well, especially for European and American popular music. When applied to music with a less clear surface structure, such as European classical music with wide timbral and dynamic diversity, performance is observed to decrease (Klapuri et al., 2006; Holzapfel, Davies, Zapata, Oliveira, & Gouyon, 2012).

A well-formed metrical grid is often assumed to be present in Eurogenetic¹ tonal music (Lerdahl & Jackendoff, 1983). Forms of music that deviate from the well-formedness in terms of meter have rarely been in the focus of computational rhythm analysis (*e.g.* the work by Antonopoulos et al. (2007)). In this paper, we concentrate our studies on Carnatic and Hindustani music of India, as well as Makam music of Turkey. All the three musics have a long standing history within the respective cultures. They are also well established traditions that exist in current social context with a large audience and significant musicological literature. While each of these music traditions includes a huge variety of styles and forms in itself, we maintain a broader point of view here, and ask in how far the state of the art in rhythm analysis can present us with meaningful insights when applied to them.

Until now, in the context of Hindustani music, approaches for tempo estimation and time signature recognition were presented by Gulati et al. (Gulati & Rao, 2010; Gulati, Rao, & Rao, 2011), and transcription of percussive instruments was approached by Chordia (2005) and Miron (2011). For Carnatic music,

¹We introduce this term because we want to avoid the misleading dichotomy of Western and non-Western music. The discussed well-formed meters are theoretical constructs that are motivated by music of the European common practice period. We use the word “genetic” rather with its connotation as “pertaining to origins”, coined in 1831 by Carlyle from Gk. *genetikos* “genitive”, from *genesis* “origin”, and not in its biological sense as first applied by Darwin in 1859 (<http://www.etymonline.com>). The term was proposed by Prof. Robert Reigle (MIAM, Istanbul) in personal communication.

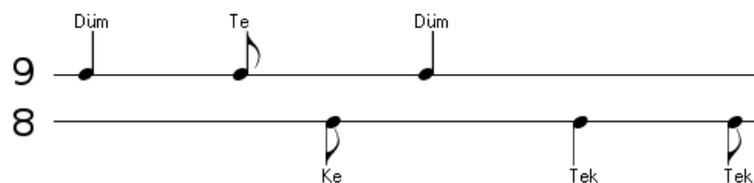
Srinivasamurthy, Subramanian, Tronel, and Chordia (2012) proposed a system to describe meter in terms of the time-span relations between pulsations at measure, beat and subdivision levels. An approach to discriminate between classes of rhythm in Turkish music was proposed by Holzapfel and Stylianou (2009), which can serve as a tool for time signature recognition as well. Recently, Holzapfel and Bozkurt (2012) investigated how the melodies of Turkish makam music are related with the underlying rhythmic mode. Both mentioned approaches for Turkish music examined repertoire consisting of symbolic data, and to the best of our knowledge no rhythmic analysis of audio signals has been attempted for Turkish music.

The available descriptions of rhythm and meter in Indian and Turkish music imply that they differ from Eurogenetic music in terms of the time-spans at various levels of the meter hierarchy and in terms of the possibly irregular pulsation at one of the levels (see e.g. Clayton (2000); Özkan (1984)). Therefore, in this article we address the problem of tracking the pulsations at different metrical levels, and of determining the relations between the time-spans between these levels. By doing so we can potentially derive meaningful descriptions of the metric structure of Indian and Turkish music pieces. In Section 2 we provide the reader with the basic concepts related to meter in Turkish Makam, Carnatic and Hindustani music, and describe some of the opportunities and challenges to rhythm analysis in these music cultures. We will then define three analysis tasks motivated by the structure of meter in the three music cultures - beat tracking, cycle length recognition, and downbeat detection. As none of these tasks have been addressed before on the presented repertoire, we had to compile and annotate the music collections that are necessary for the evaluation of the analysis algorithms. The related collections and annotations will be described in Section 3. In Section 4 we give a detailed description of all the analysis algorithms that we will evaluate for the three tasks along with description of the evaluation methods. The following Sections 5-7 give the detailed results of our experiments, separately for each of the analysis tasks. In Section 8 we sum up the results and describe what they imply for rhythm analysis in Indian and Turkish music. The final section specifies future directions for research in rhythm analysis.

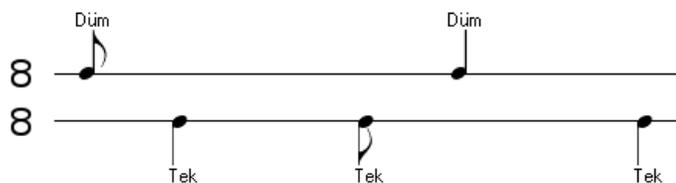
2 Problem definition

2.1 Musical background

In Turkish folk and art music, melodic progression is based on the modal concept of *makam*². While improvisations offer the opportunity to express characteristics of a makam without the presence of a continuous pulsation, all other forms of composition can be assumed to be based on a grid of organized pulsation, *i.e.* a meter. The structure of that metrical grid is partially defined by the *usul*. An usul is a rhythmic pattern of a certain length that defines a sequence of strokes with varying accent. Two examples are shown in Figure 1. The usul *Aksak* depicted in Figure (1a) has a length of nine pulses, the usul *Düyek* has a length of eight pulses. The notes on the upper line labeled *düm* denote strongly accented strokes, while the notes on the lower line denote less accented strokes. Notes on the upper line are executed with the right hand, while notes on the lower line are executed with the left hand. For some acoustic examples in the two depicted rhythmic modes, please refer to the web page of the paper (*Usul in Turkish Makam Music: Examples*, 2013).



(a) The Aksak usul, usually notated as 9/8



(b) The Düyek usul, usually notated as 8/8

Figure 1: Two usul examples, with the usual notation for right hand strokes on the upper line and left hand strokes on the lower line, respectively (Graphics from Keçecioğlu (2010)).

²The terms folk and art music are literal translations of the terms *Türk Halk* and *Türk Sanat Müziği*, which are commonly used in Turkish. Folk songs are usually assigned to a makam. If this assignment of an art music concept to folk music is acceptable in all cases is an issue that is not in the scope of this paper.

In Figure 1, both usuls are notated using a time signature. While the numerator denotes the number of pulses in the measure cycle and is constant for a given usul, the denominator (*mertebe*) may change depending on the tempo of a piece. For instance, in an Aksak notated as 9/8, the strokes are performed in a higher tempo than when a piece is notated as 9/4 (referred to as *Ağır aksak*). Makam music of Turkey is usually notated using Eurogenetic staff notation, and therefore we can interpret the length of an usul as measure length, and the regular pulsation related to the denominator of the time signature as beat pulsation. It should be pointed out that *e.g.* in 9/8 Aksak pieces, listeners would usually clap an irregular sequence at the quarter note level instead of the regular sequence at the 1/8-level. However, by considering the 1/8-level as our beat level, we will be able to evaluate existing algorithmic approaches that can only track regular sequences. As shown by Holzapfel and Bozkurt (2012), the note onsets in a piece tend to be highly correlated with the strokes of the usul. In most cases, usul patterns represent additive rhythms as they can usually not be divided into equal parts. In a performance, a percussive accompaniment may be present, playing patterns similar to the shown usul or to a more dense stroke pattern, referred to as *velveleli*. But, pieces may also be performed without any percussive accompaniment. Recently, the use of Western drum sets has become more common, with the played accompaniment being less similar to the underlying usul pattern. The described combination of additive rhythm characteristic and a wide variety of performance styles may pose severe obstacles to existing algorithms for rhythm annotation.

Hindustani (Hindustāni) and Carnatic (Karnāṭaka) music are two of the most predominant art music traditions in India. While they differ in the performance practices, they share similar melodic and rhythmic concepts. The melodic framework is based on *rāg* in Hindustani music and *rāga* in Carnatic music. The rhythmic framework is based on rhythmic cycles called the *tāl* in Hindustani music and *tāḷa* in Carnatic music. Both the words *tāl* and *tāḷa* are Sanskritic in origin and have the same literal meaning of a “hand-clap”. The difference apparent in transliteration is due to their usage in Hindustani and Carnatic music - the transliteration in this paper uses Hindi for Hindustani music and Kannada for Carnatic music. For consistency and convenience, in this article, we will use the word *tāḷa* to mean both *tāl* and *tāḷa* when we refer collectively to the two Indian musics, and use the respective terms while referring to each music culture individually. Further we will use the term Indian music to refer collectively to Carnatic and Hindustani music. Though the idea and purpose of a *tāḷa* in both music cultures is the same, there are significant fundamental differences in performance practices and terminology.

In Carnatic music, the *tāla* provides a broad structure for repetition of music phrases, motifs and improvisations. It consists of fixed length time cycles called *āvartana* which can be referred to as the *tāla* cycle. In the *āvartana* of a *tāla*, phrase refrains, melodic and rhythmic changes occur usually at the beginning of the cycle. An *āvartana* is divided into basic equidistant time units called *akṣaras*. The first *akṣaras* pulse of each *āvartana* (called the *sama*) is often accented, with notable melodic and percussive events. Each *tāla* also has a distinct, possibly non-regular division of the cycle period into parts called the *aṅga*. The *aṅgas* serve to indicate the current position in the *āvartana* and aid the musician to keep track of the movement through the *tāla* cycle. A movement through a *tāla* cycle is explicitly shown by the musician using hand gestures, based on the *aṅgas* of the *tāla*. Therefore, as a descriptive pattern of physical movements, it can be considered to take a similar role as the *usul* pattern in Turkish music.

The *akṣaras* in an *āvartana* are grouped into possibly unequal length units which are shown with hand gestures. We will refer to it as the beats of the *tāla*. The sub-division grouping structure of the *akṣaras* in a beat is called the *naḍe* (also spelt *naḍai*) or *gati*. Another important aspect of rhythm in Carnatic music is the tempo class or *kāla*. *Kāla* could be *viḷambita* (slow), *madhyama* (medium) and *dhṛta* (fast). The *kāla* is equivalent to a tempo multiplying factor and determines the number of *akṣaras* played in each beat of the *tāla*. Another important aspect of rhythm in Carnatic music is the *eḍupu*, the “phase” or offset of the lead melody, relative to the *sama* of the *tāla*. With a non-zero *eḍupu*, the composition does not start on the *sama*, but before (*aṭita*) or after (*anāgata*) the beginning of the *tāla* cycle. This offset is predominantly for the convenience of the musician for a better exposition of the *tāla* in certain compositions. However, *eḍupu* is also used for ornamentation in many cases. Sambamoorthy (1998) provides a comprehensive description of *tālas* in Carnatic music. Though there are significant differences in terms of scale and length, as an analogy, the concepts of *akṣara*, the beat, and the *āvartana* of Carnatic music bear analogy to the subdivision, beat and the measure metrical levels of Eurogenetic music. The notions of Carnatic music are visualized in Figure 2, and the web page of the paper gives acoustic examples (*Tāla in Carnatic Music: Examples*, 2013).

Figure 2 shows three popular Carnatic *tālas* - *miśra chāpu*, *rūpaka*, and *ādi tāla* in *madhyama kāla*. The figure shows the *akṣaras* with time-ticks, beats of the cycle with numbered longer time-ticks, and the *sama* in the cycle using ×. The *aṅga* boundaries are highlighted using bold and long time-ticks. *Ādi tāla* has 8 beats in a cycle, with 4 *akṣaras* in each beat leading to 32 *akṣaras* in a cycle. *Rūpaka tāla* has 12 *akṣaras* in a cycle, with 4 *akṣaras* in each of its 3 beats.

The case of *miśra chāpu* needs a special mention here. Figure (2a) shows *miśra chāpu* to consist 14 *akṣaras* in a cycle. The 14 *akṣaras* have a definite grouping structure of 6+4+4 (or 6+8 in some cases) and the boundaries of these groups are shown with visual gestures, and hence form the beats of this *tāla* (Sambamoorthy, 1998). However, in common practice, *miśra chāpu* can also be divided into seven equal beats. In this paper, we consider *miśra chāpu* to consist of seven uniform beats (as numbered in Figure (2a)), with beats \times , 4 and 6 being visually displayed. Similarly, another popular *tāla* *khaṇḍa chāpu* has 10 *akṣaras* in a cycle grouped into two groups as 4+6. In the scope of this article, *khaṇḍa chāpu* can be interpreted to consist of 5 equal length beats.

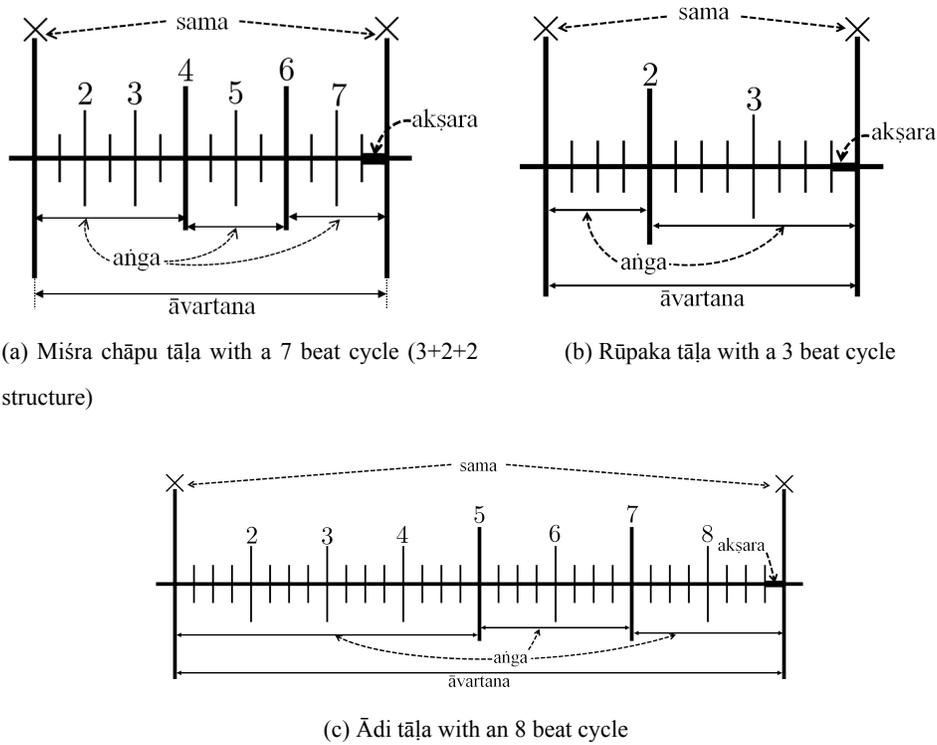


Figure 2: An *āvartana* of three popular Carnatic *tālas*, showing the *akṣaras* (all time ticks), beats (numbered time ticks), *anṅas* (long and bold time ticks) and the *sama* (\times). Each *tāla* is also illustrated using the terminology used in this article.

Most performances of Carnatic music are accompanied by the *Mridaṅgam* (*Mṛdaṅgam*), a double-sided barrel drum. There could however be other rhythmic accompaniments such as *Ghatam*, *Khanjira*, *Thevil*

and *Morsing*, which follow the Mridangam closely. All these instruments (except the Khanjira) are pitched percussion instruments and are tuned to the tonic of the lead voice. Since the progression through the tāla cycles is explicitly shown through hand gestures, the Mridangam is provided with substantial freedom of rhythmic improvisation during the performance. The solo performed with the percussion accompaniments, called a *tani āvartana*, demonstrates the wide variety of rhythms that can be played in the particular tāla.

The definition of tāl in Hindustani music is similar to the tāla in Carnatic music. A tāl has fixed-length cycles, each of which is called an āvart. An āvart is divided into basic time units called *mātrā*. The mātrās of a tāl are grouped into sections, sometimes with unequal time-spans, called *vibhāgs*. Vibhāgs are indicated through the hand gestures of a *thāli* (clap) and a *khāli* (wave). The beginning of an āvart is referred to as *sam* (Clayton, 2000). There are also tempo classes called *lay* in Hindustani music which can vary between *ati-vilāmbit* (very slow), *vilāmbit* (slow), *madhyam* (medium), *dhṛt* (fast) to *ati-dhṛt* (very fast). Depending on the *lay*, the mātrā may be further subdivided into shorter time-span pulsations. However, since these pulses are not well defined, we consider mātrā to be lowest level pulse in the scope of this paper. As with Carnatic music, even in Hindustani music, there are significant differences to the terminology describing meter in Eurogenetic music. However, the concepts of mātrā, the vibhāg, and the āvart of Hindustani music bear analogy to the subdivision, beat and the measure metrical levels of Eurogenetic music.

Figure 3 shows three Hindustani tāls - *tīntāl*, *jhaptāl*, and *rūpak tāl*. The figure also shows the sam (shown as ×), and the vibhāgs, indicated with thāli/khāli pattern using numerals. A khāli is shown with a 0, while the thālis are shown with non-zero numerals. The thāli and khāli pattern of a tāl decides the accents of the tāl. The sam has the strongest accent (with certain exceptions) followed by the thāli instants. The khāli instants have the least accent. A jhaptāl āvart has 10 mātrās with four unequal vibhāgs (Figure (3a)), while a tīntāl āvart has 16 mātrās with four equal vibhāgs (Figure (3c)). We can also note from Figure (3b) that the sam is a khāli in rūpak tāl, which has 7 mātrās with three unequal vibhāgs. Miron (2011), Clayton (2000), Dutta (1995) and Naimpalli (2005) provide a detailed discussion of tāl in Hindustani music including thēkās for commonly used tāls. The web page of the paper gives acoustic examples (*Tāl in Hindustani Music: Examples*, 2013)

Hindustani music uses the *Tabla* as the main rhythmic accompaniment. Tabla acts as the timekeeper during the performance indicating the progression through the āvarts. The lead musician improvises over these cycles, with limited rhythmic improvisation during the main piece. A percussion solo shows a variety

of improvisation possible in the framework of the tāl, with the role of timekeeping taken up by the lead musician during the solo.

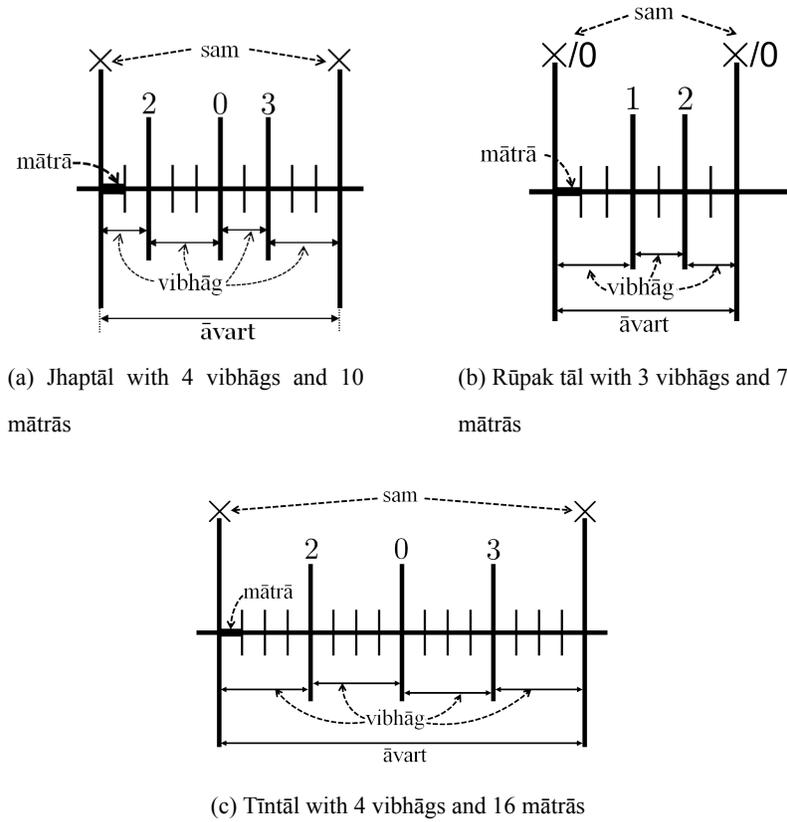


Figure 3: An āvart of three popular Hindustani tāls, showing the sam (×), vibhāg (shown with numerals) and mātrās (shown with small and large time ticks). Each tāl is also illustrated using the terminology used in this article.

The surface rhythm in all three cultures provide cues to the underlying tāḷa or usul structures. In Hindustani music, tabla is a very important cue to the underlying tāl progression. All tāls have a definite accent and Tabla stroke pattern defined by the ṭhēkā which is mostly followed except in improvisatory passages. The surface rhythm consists of these accents and specific strokes, but is also replete with other strokes, fillers and expressive additions. Filler strokes are employed in slow pieces with long cycles. In Carnatic music, as discussed earlier, the progression through the tāḷa is shown through visual gestures and hence there is no need for definitive cues in the surface rhythm. However, the ṭhēkā played on the

Mridangam, the melodic phrases and the lyrics of the composition provide cues to the underlying tāḷa. Unlike Tabla strokes, Mridangam strokes are less indicative of the current position in the cycle of a tāḷa. For Turkish music, percussive instruments have an important role as timekeeper especially in ensemble performances, and their degrees of freedom regarding improvisation are rather limited. Therefore, their strokes can give important insight into the progression of the usul. In performances without percussion, which are usually solo or duet performances, note onsets and emphasis on particular notes can be considered to be strongly related to the usul.

Unmetered forms of music exist in all three music cultures. The most important unmetered form in Hindustani music is the *ālāp* and in Carnatic music is the *ālāpanā*, both of which are melodic improvisational forms based on a rāga. Taksim is the most prominent unmetered form in Turkish makam music. An understanding of the rhythmic behavior of unmetered forms is far from trivial for musicologists and even practicing musicians (Clayton, 1996). Widdess (1994) presented an interesting discussion of the notion of pulsation in ālāps and a disagreement about it among performers. For this reason, we believe that rhythmic analysis of unmetered forms should be reserved for a study more from a musicological perspective and hence we do not consider it in this paper.

2.2 Challenges in Indian and Turkish music

We identified the āvartana (or āvart) and usul cycles as the most important and musically relevant cycles in Indian and Turkish music, respectively. For Indian music, the theoretical frameworks described previously also show cyclical structures at time-spans different from the āvartana (or āvart). There exist sub-cycles which can be perceived at the vibhāg level in certain tāḷs. A tīntāl can be seen to have four sub-cycles in an āvart, one at each vibhāg. Similarly, rūpaka (See Figure (2b)) in Carnatic music can be seen to be comprised of three sub-cycles of four akṣaras each. This implies that depending on the metrical levels we focus upon, the metric structure is determined by either duple or triple relations. While this is not a distinct feature of meter in Indian music, it is encountered quite frequently here. For Turkish music, this situation is apparently simpler, as no subdivision is explicitly defined by the usul.

Furthermore, in Carnatic music, metric structure might also vary within a piece while maintaining the same tāḷa. E.g., though rūpaka tāḷa is generally defined with four akṣaras in a beat and three beats in an

āvartana, it might change within a piece to be grouped as 4 units of 3 akṣaras. For the purpose of analysis in this paper, we consider rūpaka to have a ternary meter. This further indicates that ideally, the metrical structure of the piece needs to be estimated at all levels, taking into account possible changes in the meter. If we estimate the cycle length at only a particular time scale, we might not obtain the required cycle length at āvartana level.

An important aspect of meter in both Indian and Turkish music is the presence of pulsation at some metrical level with unequal time-spans. The vibhāgs in Hindustani music and aṅgas in Carnatic music are examples of such possibly unequal pulsations. Such forms of additive meter have so far not been widely considered for computational analysis.

In all three music cultures, the same composition can be rendered in different tempi, at the convenience of the musician. This lack of a definite tempo value and the choice of a wide variety of tempo classes further complicate the choice of a relevant timescale for cycle length tracking. In Hindustani music, the āvart cycle lengths vary from 1.5 second in ati-dhṛt (very fast) tīntāl to 65 second in ati-vilāmbit (very slow) ēktāl (Clayton, 2000, p. 87). Long time scales such as these are far too long to be perceived as single entities (Clayton, 2000). At such long time scales, the rhythm of the piece is rather characterized by the grouping structure of the piece (Lerdahl & Jackendoff, 1983). This implies that algorithmic approaches for rhythm analysis that are solely based upon estimation of pulsation might not be capable to analyze the temporal structure in presence of such long cycles. Carnatic music has a smaller range of tempi and the tālas are more concretely defined, and hence the choice of time scale is an easier problem. In Turkish music, cycle lengths can be assumed to vary very widely as well, ranging from a very fast played three pulse cycle, to a slowly played cycle of 48 or more pulses. Even within one type of usul, the *mertebe* (*i.e.* the denominator of the time signature) can specify tempo classes factors of two apart and affects the rhythmic style notably. Within these tempo classes, the full range of the tempo octave can be utilized by a composition (Holzapfel & Stylianou, 2011).

In summary, the cycle lengths in Indian music can be meaningfully tracked at multiple time levels. Distinguishing between these multiple time levels is difficult due to the wide variety of tempo classes. The same variety of tempi can be expected to cause problems in tracking the progression through an usul for Turkish music. Especially for Hindustani and Turkish music, the presence of additive meters is expected to pose challenges to existing analysis approaches. Finally, we must not forget that we attempt to track the

meter as a theoretical concept in music performance. However, in all three cultures artists can be assumed to deviate from such concepts, which results in a divergence between surface rhythm and theoretical framework that is hard to conceive in any kind of rhythm analysis using only the acoustic signal.

We can now determine a number of meaningful rhythm annotation tasks. For Carnatic music, important automatic rhythm annotation tasks are estimating the beats and the āvartana cycle length (in beats), detecting the akṣara and sama instants, and estimating the naḍe (the sub-division of the beat). The most important automatic rhythm annotation tasks for Hindustani music are estimation of the mātrā pulsation, the āvart cycle length (in mātrās), the sam instants and the vibhāg boundaries. For Turkish music, a complete description of the usul alignment to a recording can be obtained by tracking its beat pulsation, determining its cycle length in beats, and then determining the location of the downbeats.

2.3 Opportunities in rhythm analysis of Turkish and Indian music

In this paper, we only consider audio for rhythmic analysis. But the associated notations, lyrics and information regarding form also carry rhythm information which can be used for a combined approach to rhythm analysis. Tabla uses onomatopoeic syllables representing its strokes called *bōls* and Mridangam has similar syllables called *solkaṭṭu*. These syllables define a language for Indian percussion and play a very important role in defining rhythmic structures in Indian music. In Hindustani music, the ṭhēkā are defined using these bōls and hence these bōls can be used to track the movement through the āvart. The oral recitation of these syllables is a music form called *konnakōl* in Carnatic music. These syllables also form an important component of the Indian art dance forms Kathak (using Hindustani bōls) and Bharatanāṭyam (using solkaṭṭu).

For Turkish music, the availability of notations implies that the length and the type of the usul can be derived directly from the score. Furthermore, aligning an audio recording to a score using a technique as described by Şentürk, Holzapfel, and Serra (2012) can be applied to constrain both beat tracking and downbeat detection, which can be assumed to increase accuracy significantly. Using audio only, rhythm analysis can be supported by the phrasing of melodies, which is in most cases correlated with the usul (Holzapfel & Bozkurt, 2012). Also, for Indian music, tracking of the percussion strokes has a potential of increasing the accuracy of rhythm analysis.

Another important aspect of Carnatic music is that the progression through the āvartana is explicitly

shown through visual hand gestures. In a performance context, these visual cues play an important role in communication of rhythm among the performers, as well as between the performers and the audience. In fact, in many concerts, these visual cues become a part of expressiveness of the musician and the appreciation of the audience, and hence is a part of the complete experience of the music. Since these cues consist mainly of claps, they can be quite sonorous and it is not very surprising that they can be audible in some recordings. These unique rhythm aspects of Indian and Turkish music need to be explored and used in a more complete computational rhythm analysis.

A complete description of rhythm for effective discovery and experience of music involves integrating various sources of information such as audio, scores, lyrics, visual cues, information about musical form, and other culture specific aspects. Tools for rhythm analysis need to combine these data-sources in order to arrive at a more consistent analysis than by just taking into account the acoustic signal.

2.4 Automatic annotation tasks

Given the above outline of rhythmic properties of Turkish, Carnatic, and Hindustani music, we will now describe specific analysis tasks that will be discussed in detail in this paper. All three music cultures have in common that their pieces, apart from improvisations, have a meter. This meter is related to the rhythmic patterns of the *usul* for Turkish and *tāla* for Indian art music. Hence, when aiming at an understanding of the underlying meter of an unknown piece, we are able to pose similar questions as for metered music in a Eurogenetic context. We formulate and describe three analysis goals, which we investigate in this paper. We will then evaluate some state-of-the-art methods available to approach these goals, and find promising directions for solving relevant rhythm annotation problems in Indian and Turkish music.

If existing methods from MIR are capable of handling the following tasks in a satisfying way for Indian and Turkish music, we will be able to automatically analyze the content of these music signals in a well-structured way. However, as recent research results show (Holzapfel et al., 2012), these tasks are far from being solved even for the Eurogenetic forms of music, for which most methods have been presented. We evaluate several approaches for each of the three tasks and analyze which of those are promising in their results and can provide directions for future work. It should be pointed out here that there are algorithmic approaches which tackle more than one of the tasks in a combined way (e.g., Klapuri et al., 2006). We will

report the accuracy of individual tasks for such systems in our experiments as well. Further, it is also to be noted that we use only audio and its associated metadata in these tasks, because none of the available methods is capable of combining audio processing with the several other cues that were specified in Section 2.3. While it is not our primary goal to compare rhythmic aspects of Indian and Turkish music, we will point out performance differences of the applied methods between the cultural contexts. These differences will be assigned to specific properties of the analyzed music.

2.4.1 Beat tracking

In the context of MIR, beat tracking is commonly defined as determining the time instances where a human listener is likely to tap his foot to the music. This definition is likely to cause problems in our context, as for example in Turkish Aksak rhythms, listeners familiar to makam music tend to tap an irregular sequence of beats at the quarter-note level, instead of the faster regular pulsation at the eighth-note level. In Indian music, the perceived beats can be irregularly spaced as well. As none of the approaches we consider in the present context can handle an irregular sequence of beats, we pursue this task here using a more adapted definition of a beat for the purpose of evaluation, defined as a uniform pulsation. This way, we are able to obtain a notion of beat for Turkish and Carnatic music which is rather objective for evaluation with the presented algorithms, and can serve as a basis for other analysis tasks.

For Turkish music, tracking the note values specified in the mertebe (*e.g.* 8 in 9/8) is the pulsation we aim at. For Carnatic music, the beats we aim to track correspond to the beats of the tāḷa - positions in the tāḷa cycle that are visually displayed through hand gestures, as defined earlier. In Carnatic music, beat tracking is evaluated only on pieces from ādi and rūpaka since they have uniform beats. In Hindustani music, beats of a piece are not well defined, since it depends also on the *lay*. The beats could be defined at the mātrā level for slow *lay* and at vibhāg level for faster pieces and hence there exists no unambiguous definition of beats (Clayton, 2000, p. 91). Due to this ambiguity which needs a deeper exploration, we do not explore beat tracking for Hindustani music formally in this paper.

We carefully note that the beat tracking task as defined in this paper is simplified and the more appropriate task of tracking the possibly irregularly spaced beats is more relevant for Turkish and Indian music. However, the present definition of an equidistant beat pulsation can later help in obtaining the musically relevant

possibly irregular beat sequence that is a subset of the equidistant pulses. This simplification of beat to an equidistant pulsation stems mainly from the limitations of the algorithms under study.

2.4.2 Cycle length recognition

For Indian and Turkish music, recognizing the type of the *tāla* or *usul* is a central task for the listener and musician in order to follow the piece. An important step to arrive at this understanding is the correct recognition of the length of the measure cycle in beats. In MIR, related tasks are referred to as time signature recognition (Pikrakis et al., 2004), meter tracking (Klapuri et al., 2006), and rhythmic similarity (Holzapfel & Stylianou, 2011). Meter tracking aims at tracking pulsation at several levels simultaneously, and the length of a measure cycle can be derived from the relation between beat and measure pulsation time-spans. Time signature recognition estimates this relation without necessarily tracking the positions of the pulses in a signal. Rhythmic similarity, on the other hand, can be applied to the task of cycle length recognition by assigning an unknown piece to a class of rhythm it is deemed to be most similar to, based on some low level signal features. If the classes of rhythm present in a collection have distinct measure cycle lengths, we can obtain the length of the measure cycle for an unknown piece through this classification into a rhythm class.

In Indian music, we can track well-defined metrical cycles at several levels of the meter. As described earlier, though the aim of the task is to estimate the length of *tāla* cycle (length of a *āvart/āvartana*), the algorithms might track a different time scale, which need not correspond to the *tāla* cycle. One of such other specific levels we are interested in, apart from the *tāla* cycle is the sub-division within the beat pulsation, *i.e.* the periodic structure dominating the rhythmic content between the beat instances. In Carnatic music, this corresponds to the *naḍe*, and in Hindustani music, this broadly corresponds to the number of *mātrās* in the *vibhāgs*. Further, we need to point out here that there is a clear definition of the subdivision meter in the form of *naḍe* in Carnatic music. However, there is no such explicit distinction in Hindustani and Turkish music and hence the subdivision meter is not clearly given by the theoretical framework.

Though the *tāla* cycle is an important part of rhythmic organization, it is not necessary that all phrase changes occur on the *sama*. Therefore, in recordings where we know about such phrasing properties, we will observe if cycle lengths determined are consistent with the related structure in the phrases.

2.4.3 Downbeat recognition

With the given information of the beat pulsation of a music signal and the knowledge about the length of a measure cycle in beats, we finally need to understand where a measure begins. This knowledge provides us with the ability to comprehend the melodic development of a piece, which is typically synchronized with the phase of the measure cycle (Holzapfel & Bozkurt, 2012). This corresponds to detecting the sam (or sama) instants of the tāla cycle in Indian music. The downbeat in Turkish music corresponds to the first pulse in each usul cycle. There are several approaches in MIR which pursue this goal in the context of Eurogenetic music (e.g. Davies and Plumbley (2006); Hockman et al. (2012)). Note that while the term downbeat has been mostly applied to Eurogenetic music, we apply it here as well because it generally denotes the pulse at the beginning of a measure cycle. The downbeat does not necessarily correspond to the strongest accent in the cycle. For instance, in the Düyek usul in Turkish music the second stroke of the pattern shown in Figure (1b) is often referred to as stronger accented than the first stroke on the downbeat. In this sense, downbeat in Indian, Turkish, and Eurogenetic musics are likely to be concepts with different meaning.

3 Audio Music collections

The recordings used for evaluation in this paper are a subset of the bigger collection being compiled as a part of the CompMusic project (Serra, 2011). The CompMusic collection is still being built and when complete, will form a comprehensive collection representative of the forms of music that we discuss in this paper. It will comprise not only audio recordings and metadata, but also include scores, lyrics, ontologies and other culture specific information. However, in the context of this article, we use only a subset of the audio collection and the associated rhythm metadata from commercially available releases. The audio recordings are short clips extracted from full length pieces.

In order to evaluate the algorithms that will be outlined in Section 4, we need collections of audio recordings that are annotated in various aspects. For cycle length recognition we only need high-level information of the class of rhythm (usul, tāl, tāla), which decides the length of the measure cycle. For the tasks of beat tracking and downbeat recognition however, we need low-level annotations that specify the alignment between organized pulsation and music sample. Because no such annotated music collection is available, a collection of samples had to be manually annotated. As the process of manual annotation is very

time consuming, we decided to compile a bigger set of recordings with high-level annotation and selected a smaller set of recordings for the evaluation of beat tracking and downbeat recognition.

3.1 High-level annotated selections

For Turkish music, four usuls were selected based on the number of recordings that were available in the CompMusic collection. The number of pieces in each of the selected usuls *Aksak*, *Curcuna*, *Düyek* and *Yürüksemai*, and the length of the usul cycles (in beats) are listed in Table 1. Recordings comprise a large variety of recording qualities and periods. Most of the recordings contain a percussion accompaniment. For the remainder of the paper, we refer to this dataset as the Turkish dataset.

Usul	# pieces	# Beats
Aksak	60	9
Curcuna	58	10
Düyek	64	8
Yürüksemai	48	6
Total	230	

Table 1: Turkish dataset with the number of pieces of each usul. The length of each usul cycle (in beats) is shown in the right column.

For both Hindustani and Carnatic music, recordings from four popular tālas were selected. These tālas are the most prominent tālas in performance practice and also comprehensively cover most of the metrical structures that occur in common practice. The Hindustani dataset has compositions in three *lay* classes - vilāmbit, madhyam and dhṛt. The duration of āvart cycle in the dataset varies from about 2 second (dhṛt pieces) to over 20 second (vilāmbit pieces). All the recordings have a percussion accompaniment. Table 2 shows the number of pieces in the Carnatic dataset in different tālas, along with the number of beats, number of akṣaras per beat (naḍe), and akṣaras in an āvartana, in *caturaśra* (double/quadruple) naḍe and madhyama kāla. Table 3 shows the number of pieces in Hindustani music dataset in different tāls, along with the vibhāgs and the mātrās in each āvart.

Tāḷa	# Pieces	# Beats (Measure Length)	naḍe (Subdivision Meter)	# Akṣara (Subd. Pulses per Measure)
Ādi	61	8	4	32
Rūpaka	63	3	4	12
Mīśra chāpu	60	7	2	14
Khaṇḍa chāpu	33	5	2	10

Table 2: Carnatic dataset with the number of pieces in the dataset for each tala. The table also shows the structure of the tāḷa indicated by the naḍe (akṣaras per beat), the number of beats and akṣaras in each cycle

In the datasets summarized in Tables 1, 2, and 3, the pieces are 2 minute long excerpts sampled at 44100 Hz. Though the audio recordings are stereo, they are down-mixed to mono since none of the algorithms evaluated in this study make use of additional information from stereo audio and primarily work on mono. They include instrumental as well as vocal recordings. The tāḷa/tāl/usul annotation of these pieces were directly obtained from the accompanying editorial metadata contained in the CompMusic collection.

3.2 Low-level annotated selections

The low-level annotated selections are a subset of the high-level annotated datasets. Two out of the four classes were chosen in Turkish and Carnatic music and annotated manually. The annotations include downbeat annotations and beat annotations as a regular pulsation at the level described in Section 2.4.1. Thirty two examples in Aksak usul and thirty examples in Düyek form the annotated Turkish music dataset, referred to in the remainder of the paper as Turkish low-level-annotated dataset. These audio files are one minute long, in contrast to Turkish dataset, where pieces are two minute long.

Thirty two examples in ādi tāḷa and thirty four examples in rūpaka tāḷa of Carnatic music have beat and sama instants manually annotated, and we refer to the Carnatic low-level-annotated dataset. Similar to the Carnatic dataset, Carnatic low-level-annotated dataset also consists of two minute long excerpts.

All annotations were manually done using Sonic visualizer (Cannam, Landone, & Sandler, 2010) for Carnatic music, and using the Beat-Station (Miron, Gouyon, Davies, & Holzapfel, 2013) for the Turkish

Tāl	# Pieces	# Vibhāg (Measure Length)	Division (Subdivision Meter)	# mātrās (Subd. Pulses per Measure)
Tīntāl	62	4	4,4,4,4	16
Ēktāl	61	6	2,2,2,2,2,2	12
Jhaptāl	15	4	2,3,2,3	10
Rūpak tāl	19	3	3,2,2	7

Table 3: Hindustani dataset showing the number of pieces for each tāl. For each tāl, it also shows the structure of the tāl with the number of vibhāgs in each avart and number of mātrās in each avart. The last column of the table shows the grouping of the mātrās in the avart into vibhāgs, and the length of each vibhāg, *e.g.* each avart of rūpak tāl has three vibhāgs consisting of three, two, two mātrās respectively.

music, by tapping along to a piece and then manually correcting the annotations. The reason for performing annotations using different tools lie in personal preferences of the annotators (first author for the Indian dataset and second author for the Turkish dataset).

4 Computational methods

In this section, we describe the computational methods that we evaluate in the scope of the three defined rhythm analysis tasks. We describe the used approaches, and provide references to the related works. We want to point out that the various algorithms described in this section were in their majority not developed for the music repertoire we intend to analyze. Therefore, we will use the performance differences not to judge their quality, but rather to draw conclusions about suitable approaches to rhythm analysis in Indian and Turkish music.

4.1 Beat tracking

So far, to the best of our knowledge, no beat tracking has been attempted on Indian and Turkish music. Hence, we chose two acknowledged approaches presented in the last decade (Klapuri et al., 2006; Ellis, 2007) for evaluation. Holzapfel et al. (2012) provide a detailed discussion of the available state of the art.

4.1.1 Approaches

Klapuri et al. (2006) proposed a method for analyzing the musical meter jointly at the metrical levels of measure, beat and subdivision, which are referred to as measure, tactus and tatum, respectively. This method will be referred to as KLA in this paper (When possible, we will use the acronyms for the algorithms as used by McKinney, Moelants, Davies, and Klapuri (2007)). A time frequency analysis is applied to obtain accent signals in four frequency bands, which are aimed at emphasizing changes caused by note onsets in the signal. A bank of comb filter resonators is applied for periodicity analysis to each of the four accent signals. The periodicities thus found are processed by a probabilistic model that incorporates certain assumptions regarding musical meter to perform a joint estimation of the tatum, tactus, and measure pulsations. In this paper, we employ the causal version of the algorithm.

Ellis (2007) developed a beat tracking algorithm based on dynamic programming (referred to as ELL in the remainder of this paper), which computes a global set of optimal beat candidates, given an accent signal and a tempo estimate. For the beat tracking algorithm used in this paper, the accent signal used is the detection function computed by Davies and Plumbley (2007) and the tempo estimate is provided from the tempo estimation algorithm described by Srinivasamurthy et al. (2012). ELL algorithm pursues a tradeoff between the temporal continuity of beats and the salience of the detection function using the dynamic programming approach. The main drawback of the algorithm is the assumption of a constant tempo, which causes problems for music with varying tempo. Further, errors in tempo estimation translate to an incorrect estimation of the beats.

4.1.2 Evaluation criteria

For our evaluations, we can compare the pulsations estimated by the algorithms with our ground truth annotations at all three metrical levels to determine if the large possible tempo ranges cause the beat to be tracked at different levels of the meter. Using the beat ground truth annotations from both the Turkish and Carnatic low-level annotated datasets, we can define the median inter-annotation time-span as T_a for each music piece. From the estimates obtained from KLA for downbeats, beats and subdivision pulses on a specific piece, we define the following time-spans: let T_m denote the median measure time-span, T_b the median beat time-span, and T_s the median subdivision time-span. Let $T_{b,ELL}$ denote the median beat time-span as

estimated by the ELL algorithm. The median is computed to obtain an estimate that is robust to spurious deviations.

Further, given the estimated beat instants from the algorithm, we have a wide variety of performance measures for beat tracking. Davies, Degara, and Plumbley (2009) and McKinney et al. (2007) provide a detailed description of the commonly used performance measures. For our evaluation, we will use the f-measure, Information Gain, and CMLt measures which are often used in beat tracking evaluations (MIREX)³.

The f-measure uses the information retrieval performance metrics of *precision* and *recall*. It can take a maximum value of 100%, while beats tapped on the off-beat relative to annotations will be assigned an f-measure of 0%. Estimated beats with time-spans either half or double the annotated time-span are penalized with a value of 66%. The CMLt measure (Correct Metrical Levels, total) takes a value of 100% only for sequences that coincide with the annotations. It does not penalize discontinuities in beat tracking as the CMLc (Correct Metrical Levels, continuity required) measure, but penalizes any beats tracked at half or double time-spans of the annotated metrical level. For both f-measure and CMLt, small mis-alignments in the estimated beats are allowed for by applying tolerance windows. Information gain aims at determining if there exists any kind of relation between the estimated beats and the annotations. It uses the entropy of the beat error distribution and can be interpreted as an information theoretic measure. It takes a value of 0 bits only for completely unrelated sequences and by using the default setting of 40 bins in the beat error histogram, a maximum value of 5.3 bits for highly related beat sequences. The study by Davies et al. (2009) provides more details on the measures. Measuring performance using these three different measures help us to interpret the results of beat tracking better and identify several types of errors present in the estimations.

4.2 Cycle length recognition

The algorithms which will be evaluated for the recognition of the cycle length can be divided into two substantially different categories. On one hand, we have approaches that examine characteristics in the surface rhythm of a piece of music, and try to derive an estimate of cycle length solely based on the pulsations found in that specific piece. We will refer to such approaches as self-contained through the remainder of this

³Music Information Retrieval Evaluation eXchange (MIREX) 2012, http://www.music-ir.org/mirex/wiki/2012:Audio_Beat_Tracking

paper. On the other hand, there are rhythmic similarity approaches that can give an insight into the rhythmic properties of a piece by comparing with other pieces of known rhythmic content. To this end, we will use the music collections described in Section 3.1, which contain pieces with known cycle lengths. There, we can determine the rhythmic similarity of an unknown piece to all the pieces in our collection. We can then assign a cycle length to the piece according to the observation of the cycle lengths of other similar pieces. We will give a summary of the approaches used in this paper for cycle length recognition, dividing the approaches into the specified two categories.

4.2.1 Self-contained approaches

Gulati et al. (2011) proposed an algorithm (referred to as GUL in the remainder of the paper) to estimate and classify the subdivision meter of the piece to be duple (2/4/8), triple (3/6), or septuple(7). A mel-scale frequency decomposition of the signal is used to drive a two stage comb filterbank. The filterbank output is used to estimate the subdivision time-spans and the subdivision meter of the song. This is one of the first proposed approaches to rhythm modeling applied specifically to Indian music. However, the algorithm was tested on a different repertoire than Hindustani and Carnatic music. Though proposed for Indian music, the algorithm is general in approach and does not consider any specific characteristics of rhythm in Indian music.

In the method proposed by Pikrakis et al. (2004), referred to as PIK henceforth, a time signature is estimated from a self-distance matrix computed from Mel-frequency cepstral coefficients (MFCC) of the audio signal. To this end, minima in the distance matrix are assumed to be caused by repetitions related to the metrical structure of the piece. Hence, this algorithm does not track pulsations in a piece, but relies on existence of patterns caused by general repetitions in the MFCC features. Because MFCC features capture timbral characteristics, it can be stated that similarities in local timbre are used by the algorithm. The algorithm was tested on East-European music styles, including Greek traditional dance music. The paper presents two different diagonal processing techniques and we report the performance for both methods (Method-A and Method-B). As suggested by Pikrakis et al. (2004), we also report the performance using a combination of the two methods.

We note that KLA approach can be used for cycle length recognition task by using the measure, beat, and subdivision time-spans. Ideally, dividing the measure time-span by the beat time-span should present us

with the measure length in beats. However, we explore the use of the measure-beat, beat-subdivision, and measure-subdivision time-span relations to estimate cycle length and evaluate how well they coincide with the known cycle lengths of a piece.

Srinivasamurthy et al. (2012) proposed a culture-specific approach (referred to as SRI in the remainder of the paper) to rhythm description of Indian music. The algorithm uses a beat similarity matrix and inter-onset interval histogram to automatically extract the subdivision and measure cycle time-span. From this information, a rank ordered set of candidates for the tāla cycle period and the naḍe can be obtained. For the present evaluation, the tempo estimation in the algorithm, which is adapted from Davies and Plumbley (2007), is modified to peak at 90 bpm. Though there is no notated tempo in Carnatic music, it was observed from the Carnatic dataset that the histogram of the beat timespans peaks at 90 bpm. Further, the tempo analysis was modified to include a wide range of tempi (from 20 bpm to 180 bpm). The algorithm applies the ELL beat tracker (see Section 4.1) with the estimated tempo as input. Though formulated using the knowledge of the tāla, the algorithm does not make an effort to resolve metrical level ambiguities, which can severely affect the accuracy since the performance of the algorithm depends mainly on reliable beat tracking at the correct metrical level.

4.2.2 Comparative approaches

One approach to compare rhythmic content of music is by using onset patterns (*OP*), as initially presented by Pohle, Schnitzer, Schedl, and Knees (2009). Starting from a magnitude spectrum obtained from the Short-Time Fourier Transform (STFT) of a monophonic piece of music, a set of energy coefficients are computed in 32 logarithmically spaced frequency bands. A band-wise accent signal is then derived by applying a moving average filter and half wave rectification to each of the 32 bands. A second STFT operating on longer time scale (8s window with 1s hop) is applied to each band-wise accent signal. This way, a description of periodicities referred to as *OP* features (Holzapfel, Flexer, & Widmer, 2011) is obtained for 5 bands per octave, and 5 periodicity octaves from 30 bpm to 960 bpm. The rhythm of a whole sample is described by the mean of the *OP* obtained from the various segments of this sample.

A different descriptor for rhythmic content based on Scale transform was proposed by Holzapfel and Stylianou (2011). The first step is a computation of an accent signal. To this end, the sum of the 32

band-wise accent signals used for the *OP* features are applied to obtain a single vector describing the note onset characteristics. Then, within the moving windows of eight seconds length, autocorrelation coefficients are computed from this accent signal and then transformed into the scale domain by applying a discrete Scale Transform. For one piece, the mean of the Scale Transform Magnitudes (*STM*) obtained from all the analysis windows are the *STM* descriptors of the rhythmic content of the piece.

Both the mapping onto a logarithmic axis of the magnitudes in the second STFT in the *OP* features, and the application of a Scale transform in the *STM* features provide varying degrees of robustness to tempo changes. Holzapfel and Stylianou (2011) provide more details and the exact computation of parameters of the two descriptors.

4.2.3 Evaluation criteria

For comparative approaches, we apply a 1-nearest-neighbor classification in a *leave-one-out* scheme, and report the accuracy for a dataset. For self-contained approaches, we examine the accuracy of the outputs obtained from various algorithms.

We note that the algorithms PIK and GUL consider short time scales for cycle lengths and may track cycles of shorter length than the measure cycle. Hence, as explained in Section 2.2, the algorithms may track meter at a subdivision level. As the algorithms are not specifically designed to perform the task of cycle length recognition as defined in Section 2.4, the evaluation has to be adapted to the algorithms. For example, GUL classifies the audio piece into three classes - duple, triple, and septuple meter. For this reason, samples in the the dataset are labeled as being duple, triple or septuple based on the *tāla/usul* for evaluating GUL. Rhythm classes in the datasets that do not belong to any of these categories are excluded from evaluation.

For KLA and SRI algorithms we report the accuracy of estimating the annotated cycle length at the correct Metrical Level (CML). We also report the Allowed Metrical Levels (AML) accuracy considering cycle length estimates by the algorithms to be correct that are related to the annotated cycle length by a factor of 2 or 0.5, which is referred to as doubling or halving, respectively. For cycle lengths which are odd (e.g. Aksak usul, with cycle length 9), we only consider doubling of cycle length estimates in AML. Halving and doubling of cycle lengths can be interpreted as estimating sub-cycles and supra-cycles related to the annotated cycle length by a multiple, and can provide insights on tempo estimation errors committed by the algorithms.

4.3 Downbeat recognition

The methods described in this section were developed for the identification of downbeats within sequences of beats. So far mainly music with a 4/4 time signature was focused upon in evaluations, usually in the form of collections of Eurogenetic popular and/or classical music. Hence, we will address the questions if such approaches can cope with the lengths of measure cycles present in our data (*i.e.* 3, 5, 7, 8 and 9 beats, and ranging from 2 to 15 seconds), and if Indian or Turkish music pose challenges of unequal difficulty.

4.3.1 Approaches

The approach presented by Davies and Plumbley (2006), which will be referred to as DAV, is based on the assumption that percussive events and harmonic changes tend to be correlated with the downbeat position. Therefore, they partition an audio signal into beat segments and compute an STFT of each segment, neglecting frequencies above 1.4 kHz. Then the magnitude differences between all neighboring blocks are computed. Subsequently, for a given measure length l_m in beats, the sequence of l_m -distant segments that is related to the maximum spectral change is chosen as downbeats.

Recently, Hockman et al. (2012) presented an algorithm for detecting downbeats in music signals of electronic dance music genres. Their approach, which will be referred to as HOC, is an extension of a downbeat detection system proposed by Jehan (2005) that applies support vector regression. The features of the regression consist of Mel-frequency spectral coefficients, loudness descriptors, and chroma features, all computed for the separate beat segments. The extension proposed by Hockman et al. comprises a post-processing of the regression, a combination with a low-frequency onset detection, and a beat-time weighting. While the post-processing compensates for spurious downbeat detections, the combination of the regression with a low-frequency onset feature is motivated by the fact that strong bass drums tend to be located at the downbeat for the form of music they considered.

It is apparent that both systems are conceptualized for styles of music with notable differences to the music considered in this paper. The system by Davies and Plumbley (2006) is mainly sensitive to harmonic changes, whereas none of the musical cultures incorporates a notion of harmony similar to the Eurogenetic concept of functional harmony. On the other hand, the system by Hockman et al. (2012) is customized to detect the bass kick on a downbeat, which will not occur in the music we investigate here. As the latter system

contains this low-frequency feature as a separate module, we will examine the influence of the low-frequency onsets and the regression separately in our later experiments.

4.3.2 Evaluation criterion

The evaluation metrics we use are the same as the continuity-based approach applied by Hockman et al. (2012). This measure applies a tolerance window of 6.25% of the inter-annotation-interval to the annotations. Then it accepts a detected downbeat as correct, if

1. The detection falls into a tolerance window.
2. The precedent detection falls into the tolerance window of the precedent annotation.
3. The inter-beat-interval is equal to the inter-annotation-interval (accepting a deviation of the size of the tolerance window).

5 Beat tracking experiment

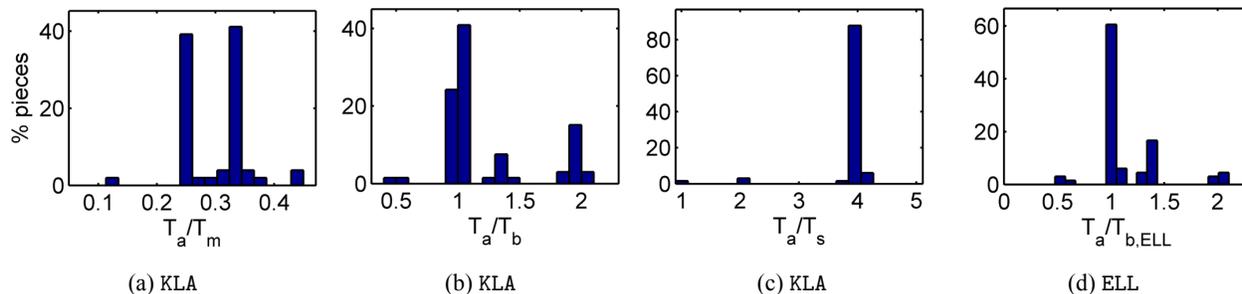


Figure 4: Relative time-span histograms on Carnatic low-level-annotated dataset. The panels (a), (b), (c) correspond to KLA algorithm. Panel (d) corresponds to the ELL algorithm. The ordinate is the fraction of pieces (in percentage) and is normalized to sum to 100%.

We evaluate beat tracking on the Turkish low-level-annotated dataset and the Carnatic low-level-annotated dataset introduced in Section 3.2, using KLA and ELL beat tracking algorithms. Since KLA algorithm estimates meter at three levels, we evaluate beat tracking performance of the algorithm with the annotated pieces at the subdivision, beat and measure level to ascertain whether the algorithm

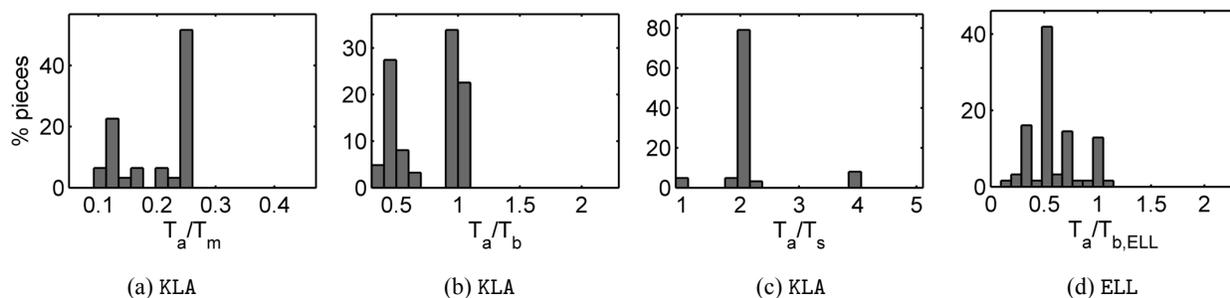


Figure 5: Relative time-span histograms on Turkish low-level-annotated dataset. The panels (a), (b), (c) correspond to KLA algorithm. Panel (d) corresponds to the ELL algorithm. The ordinate is the fraction of pieces (in percentage) and is normalized to sum to 100%.

tracks the annotated beats on any of those levels. We evaluate in how far the estimated beats are accurate in terms of the estimated time-spans, and if they are correctly aligned to the phase of the annotations. We further discuss various kinds of errors observed in the beat tracking results.

We use the different time-span definitions from Section 4.1.2 to obtain histograms of estimated relative tempo for both low-level-annotated datasets, as shown in Figures 4 and 5. The figure shows the histogram of the ratios T_a/T_m , T_a/T_b , T_a/T_s , obtained from the estimates by the KLA algorithm, and $T_a/T_{b,ELL}$ for the estimation by ELL algorithm. A ratio of one would imply that the annotation time-span and the estimated pulse time-span matched exactly, and the pulsation is thus tracked at the annotated level. A ratio of, *e.g.* 2, would indicate that the estimated tempo is two times faster than the annotated one.

From the panels (a)-(c) of Figures 4 and 5, we see that in a majority of pieces in both the low-level-annotated datasets, the KLA algorithm tracks the annotated beats at the beat level, as indicated by Figures (4b) and (5b). It is also clear from the figure that the measure time-spans estimated by the algorithm are at least twice as long as the annotated beat time-spans and hence we conclude that the estimated downbeats never correspond to the annotated beats in both datasets. Further, we see that the estimated subdivision time-spans are much shorter than the annotated beat time-spans. Based on these observations, we present the beat tracking results using only the estimated beats for KLA algorithm. In general, we can also see that there is a tendency towards tempo under-estimation for Turkish music (Figure (5b)), and towards tempo over-estimation for Carnatic music (Figure (4b)).

Dataset	Turkish low-level annot. dataset		Carnatic low-level annot. dataset	
	KLA _b	ELL	KLA _b	ELL
f-measure(%)	67.79	33.17	46.73	23.27
Information Gain (bits)	1.31	0.4065	1.09	0.4254
CMLt(%)	44.01	10.18	34.85	24.78

Table 4: Beat Tracking Results, KLA_b indicates that the results are reported using the beat pulsation estimates of KLA algorithm

The beat tracking performance is shown in Table 4, averaged over the songs of the individual collections. In the table, KLA_b refers to evaluating using the beat estimates by the KLA algorithm. ELL refers to the performance using Ellis beat tracking algorithm. KLA_b provides us with the best accuracy while beat tracking on Turkish music seems to work slightly better than for Carnatic music. However, accuracies are low in general, with *e.g.* CMLt accuracies of 44.01% and 34.85% using KLA_b on Turkish and Carnatic low-level-annotated datasets, respectively. This indicates that we are able to obtain an absolutely identical sequence in less than half of the cases on both datasets. Indeed, only 33% of the samples in the Carnatic dataset, and 29% of the samples in Turkish dataset have a CMLt higher than 50%, which means that only a third of the pieces are absolutely correctly tracked at least in more than half of their duration. Information gain values are very low especially for Carnatic, which indicates a lot of sequences occur that are not simply related to the annotation (*e.g.* by tempo halving/doubling, or off-beat). In fact, only 21% of Carnatic, and 34% of Turkish low-level-annotated datasets have an Information Gain value larger than 1.5 bits. This value was established as a threshold for perceptually acceptable beat tracking (Zapata, Holzapfel, Davies, Oliveira, & Gouyon, 2012). Even though this value was established on Eurogenetic popular music, the low values for Information Gain in Table 4 imply that a large part of the beat tracking results is not perceptually acceptable. This seems to stand in conflict with the observation made in Figures 4 and 5 that in a majority of cases KLA seems to get a median tempo that is related to the annotation. We observed however that many beat estimations suffer from unstable phase and tempo, which causes low values especially for Information Gain and CMLt.

ELL algorithm uses the tempo estimated by SRI approach (Srinivasamurthy et al., 2012), and uses a tempo weighting function which peaks at 90 bpm. Since the median inter-annotation time-span in Turkish low-level-annotated dataset corresponds to 147 bpm, the tempo estimate is not accurate and hence the CMLt performance is poor. The median inter-annotation time-span in Carnatic dataset corresponds to 85 bpm, which is closer to the peak of the weighting function than for Turkish music. We can also see this from Figures (4d) and (5d), where we observe that ELL tracks the tempo of the annotated beats in a majority of cases in Carnatic, while a majority of pieces in the Turkish dataset are tracked with slower tempo compared to the annotations. In many cases, the estimated beats tended to drift away from the annotations. This is initially caused by the resolution of the tempo estimate (about 11.6 ms, as used by Davies and Plumbley (2007)), which can result in small deviations from the true tempo. While such errors should be compensated for in the ELL algorithm by the matching of the beat estimations to the accent signal, this compensation seems to fail in many cases especially for Carnatic music.

In order to check if beat tracking tends to happen on the off-beat for some songs in the collections, we took advantage of the properties of the evaluation measures (see Section 4.1). Since estimations with a constant phase misalignment are practically the only reason for the f-measure to take values close to zero, we located the appearance of those songs. Interestingly, for none of the Turkish music samples such a case appeared which shows that at least in the small available dataset there is no ambiguity regarding phase alignment. However, for Carnatic music, phase alignment caused problems for both ELL and KLA, revealed by a large number of samples with very low f-measure (<15%). This effect is partly related to a non-zero *eđupu* (a phase shift of the start of the composition relative to the *sama* - see Section 2.1). In such a case, the composition starts not on the *sama* but 2, or 4, or 6 *akşaras* after *sama*. If the *eđupu* is 2 or 6 *akşaras*, the accents of the composition in the piece are more likely to be on the off-beat and hence the beat tracker tracks the off-beat, with the tracked beat instant lying half way in between the two annotated beat instants. In general, we find poorer performance with pieces with a non-zero *eđupu*. In Carnatic dataset we have a total of 11 pieces with a non-zero *eđupu*, and the average CMLt accuracy values using KLA_b for those pieces was observed to be clearly lower than the average (CMLt: 20.19% instead of 34.85%).

6 Cycle length recognition experiment

We are primarily interested in estimating the cycle length at the *āvart/āvartana* level for Indian music and at the *usul* level in Turkish music, a problem related to estimating the measure length in Eurogenetic music. However, as explained in Section 2.2, cycles may exist at several metrical levels, with especially Carnatic *tālas* having equal subdivisions at lower metrical levels in many cases. In connection with the fact that the measure cycles might extend over a long period of time, these shorter cycles contribute an important aspect to forming what can be perceived as beats. For the evaluations in this section, we will refer to the subdivision meter and the measure length as given in Tables 1, 2, and 3.

We differentiated between self-contained and comparative approaches, and the self-contained approaches are divided into two types of methods (see Section 4.2). The first type attempts to estimate the meter or the time signature based on repetitions observed in the signals, while the second type aims at tracking the pulsations related to those repetitions. We will start our evaluations with methods that belong to the first type (GUL, PIK), and evaluate then the tracking methods (KLA, SRI). The comparative approaches are evaluated in Section 6.2.

6.1 Self-contained approaches

Dataset	Accuracy (%)
Turkish (Düyek+Yürüksemai)	83.04
Carnatic (without khaṇḍa chāpu)	75.27
Hindustani(without jhaptāl)	49.30

Table 5: Performance of meter estimation using GUL algorithm

Gulati’s algorithm (GUL) is focused on music with a divisive, *i.e.* a regular, meter. Further, the algorithm only considers a classification into double, triple, or a septuple meter. Therefore, we had to restrict the evaluation to those classes that are based on such a meter. For Turkish music, we evaluated if the algorithm correctly estimates the type of the cycle length for the *usuls* under consideration, and *düyek* was considered as duple meter and *yürüksemai* as triple meter. For Carnatic music, *ādi*, *rūpaka*, and *miśra chāpu tālas* have a

double, triple and septuple meter respectively. In Hindustani music, *tīntāl*, *ēktāl*, and *rūpak tāl* were annotated to belong to double, triple and septuple meter classes.

Table 5 shows the accuracies for the three datasets, using the types of rhythms which can be processed by the algorithm. The performance on Turkish and Carnatic music is better than the performance on Hindustani music. A detailed analysis revealed that the performance on *rūpaka* (ternary) is only 65.08%, which leads to considerable decrease in the performance on Carnatic music. This poorer performance can be attributed to the ambiguity between duple and triple meter that is an intrinsic property of this *tāla* (see Section 2.2). Furthermore, the performance on Hindustani music was found to be poor on *rūpak tāl* and *ēktāl* while the performance on just *tīntāl* is 80.64%. This can be attributed to the fact that there are very long cycles in Hindustani music in *vilāmbit lay*, where the long subdivision time-spans restrains the algorithm from a correct estimation. In most of such cases in *ēktāl* and *rūpak tāl*, the estimated meter is a duple meter, which might be related to the further subdivision of the *mātrās*.

Dataset	Method-A (%)	Method-B (%)	Combined (%)
Turkish	29.56	33.48	44.78
Carnatic	52.53	49.30	64.06
Hindustani	35.67	53.50	57.96

Table 6: Performance of cycle length recognition using PIK algorithm. The Method-A and Method-B refer to the two methods suggested by Pikrakis et al. (2004)

Pikrakis algorithm (PIK) looks for measure lengths between 2 and 12. We report the accuracy accepting an answer if it is correct at one of the metrical levels. For example, for *ādi tāla* and *tīntāl*, 4/4, 8/4, 4/8, 8/8 are all evaluated to be correct estimates, because 4 is the subdivision meter, and 8 is the length of the *āvartana* cycle (measure length). Further, the algorithm outputs an estimation for every 5 second frame of audio, and therefore time signature of a song is obtained by using a majority vote for a whole song. The performance is reported as the accuracy of estimation (% correctly estimated) for both the diagonal processing methods (Method-A and Method-B) in Table 6. As suggested by Pikrakis et al. (2004), we also use both methods to combine the decision and it improves the performance, as can be seen from the table. The performance on

Dataset	CML (%)			AML (%)		
	L_{mb}	L_{ms}	L_{bs}	L_{mb}	L_{ms}	L_{bs}
Turkish	1.30	27.82	0	31.30	27.83	10.00
Carnatic	11.06	8.76	4.15	34.10	45.16	25.81
Hindustani	0	25.4	0	45.22	46.50	32.48

Table 7: Accuracy of cycle length recognition using KLA algorithm

Dataset	CML (%)			AML (%)		
	L_{mb}	L_{ms}	L_{bs}	L_{mb}	L_{ms}	L_{bs}
Turkish	5.65	1.74	7.39	25.22	40.43	13.91
Carnatic	3.69	0.46	6.45	40.55	50.69	14.28
Hindustani	14.64	9.55	0.64	43.95	55.41	14.65

Table 8: Accuracy of cycle length recognition using SRI algorithm

Carnatic music is better than that on Turkish and Hindustani music. Though the performance on Hindustani dataset is poor, further analysis shows that for *tīntāl*, the accuracy is 74.19%. PIK algorithm performs better in the cases where the meter is a simple duple or triple, while the performance is worse with other meters. For example, *mīśra chāpu* (length 7) has an additive meter and the cycle can be visualized to be a combination of 3/4 and 4/4. On that class the PIK algorithm estimates most of *mīśra chāpu* pieces to have either a 3/4 meter or a 4/4 meter. The performance on *Aksak* is also very poor (additive 9-cycle), with most pieces being estimated to have a meter of 3/4 or 4/4.

To obtain the cycle length estimates with the KLA algorithm, we first compute the median values of measure-, beat- and subdivision time-spans, defined in Section 4.1.2 as T_m , T_b and T_s , respectively. We then compute the cycle length estimates as,

$$L_{mb} = \left\lfloor \frac{T_m}{T_b} \right\rfloor \quad L_{ms} = \left\lfloor \frac{T_m}{T_s} \right\rfloor \quad L_{bs} = \left\lfloor \frac{T_b}{T_s} \right\rfloor$$

where $\lfloor \cdot \rfloor$ indicates rounding to the nearest integer. We examine which of the three estimates more closely represents the cycle length. As described in Section 4.2, we report both the CML and AML accuracy of cycle

length recognition. Table 7 shows the recognition accuracy (in percentage) of KLA algorithm separately for L_{mb} , L_{ms} , or L_{bs} as the cycle length estimates.

We see in Table 7 that there is a large difference between CML and AML performance, which indicates that in many cases tracked level is related to the annotated level by a factor 2 or 1/2. We also see that for Hindustani music, the cycle length is best estimated using L_{ms} , with the CML accuracy being very low or zero when we use the other cycle length estimates instead. As discussed earlier, in Hindustani music, the cycle length is defined as the number of mātrās in the cycle, *i.e.* the number of subdivision pulses per measure. However, in the case of vilāmbit pieces, the mātrās are longer than the range of the tatum pulse time-span estimated by the algorithm and hence the performance is poor. Interestingly, we see a good performance when evaluated with L_{mb} only with tīntāl, which resembles the Eurogenetic 4/4 meter, with an AML accuracy of 88.71% in spite of the CML accuracy being zero. In fact, it is seen that L_{mb} is always four in the case of a correct estimation (AML), which is the estimate of the number of vibhāgs in the tāl. Further, it follows from Klapuri et al. (2006, Figure 8) that relation between neighboring levels in KLA cannot be larger than 9, which implies longer cycle length estimates (as needed by e.g. ēktāl or tīntāl) could possibly appear only in the L_{ms} length. Hence, for Hindustani music, in which the cycle length is described with mātrās (which are analogous to subdivisions), L_{ms} is the best estimate of cycle length.

Similarly, L_{ms} is the best estimate of cycle length at CML for Turkish music. However, the CML performance of L_{ms} is high mainly on Düyek usul (65.62%). But, in a significant number of cases, the annotated level is often in between the beat and the subdivision levels tracked by the algorithm. This severely affects the cycle length recognition, especially of Aksak that has a cycle length of 9, which is almost never estimated accurately.

We saw in Section 5 that KLA tracked correct tempo in a majority of cases in Carnatic music. The CML accuracy with L_{mb} is hence better than the other cycle length estimates. However, the performance is poor because the algorithm often under-estimates the measure length. Further, in tālas of Carnatic music that have two akṣaras in a beat (khaṇḍa chāpu and miśra chāpu), L_{ms} is a better indicator of the cycle length than L_{mb} , since akṣaras are closer to the estimated subdivision time-span. In general, L_{bs} performs poorly compared to L_{ms} or L_{mb} , which is not astonishing since the measure lengths we are looking for are longer than the estimated subdivision meter. Summing up, none of the estimated meter relations can serve as a robust estimate for the measure cycle length.

SRI algorithm estimates the cycle length at two metrical levels using the beat estimation by ELL algorithm, one being at the measure level (measure length in beats), and the second at the beat level (subdivision meter). The algorithm computes a list of possible candidates for the beat cycle length and measure cycle length, ordered by a score. We consider the top candidate in the list and compute the cycle length estimates L_{mb} , L_{bs} , and the L_{ms} , assuming that the beats tracked by the ELL algorithm correspond the beat time-span T_b . Similar to KLA algorithm, we present the CML and AML accuracy of performance in Table 8.

We see that there is large disparity between the CML and AML accuracy, which indicates that the beat tracker and the correct beat are related by a factor of 2 or 1/2, a tendency that was observed at least for Turkish music in Section 5. In general, the algorithm performs poorly, which can be mainly attributed to errors in tempo and beat tracking. The tempo estimation uses a weighting curve that peaks at 90 beats per minute, which is suitable for Carnatic music, but leads to an incorrect estimation of cycle length for Turkish and Hindustani music. A beat tracking based approach as the SRI algorithm might in general not be well suited for Hindustani music which often includes long cycles that might be more reflected in the structure of melodic phrases than in pulsation and rhythmic aspects.

The poor performance on Carnatic music can in part be also attributed to changes in *kāla* (tempo multiplication) throughout a song and to variation in percussion accompaniment, which is completely free to improvise within the framework of the *tāla*. Further, the algorithm is based on the implicit assumption that beats at the same position in a measure cycle are similar between various recurrences of the cycle. For certain music pieces where there are no inherent rhythmic patterns or the patterns vary unpredictably, the algorithm gives a poorer performance. For Carnatic music, the algorithm specifically estimates the subdivision-meter (*naḍe*), as the number of *aḡaras* per beat. Using L_{bs} as an estimate of the *naḍe*, we obtain a reasonably good performance comparable to GUL with an accuracy of 39.63% and 79.72% at CML and AML (of the subdivision meter), respectively. We see that a reasonable performance when demanding an exact numerical result for the meter (CML) is only reached for the *naḍe* estimation in Carnatic music.

We observe that the duration of cycles in seconds is often estimated correctly, but the presence or absence of extra beats causes the estimated length in beats to be wrong. ELL algorithm is sensitive to tempo value and cannot handle small tempo changes effectively. This leads to addition of beats into the cycle and the cycle length in many cases were estimated to be one-off from the actual value, though the actual duration of

the cycle (in seconds) was estimated correctly.

6.2 Comparative approaches

Dataset	OP (%)	STM (%)
Turkish	50.4	56.5
Carnatic	41.0	42.2
Hindustani	47.8	51.6

Table 9: Accuracy of cycle length recognition using comparative approaches

The comparative approaches as introduced in Section 4.2 are based on a description of periodicities that can be derived from the signal without the need to perform meter tracking. Performances of the two evaluated methods, OP and STM, is the average accuracy in a 1-nearest neighbor classification. It tells us how often a piece found to be most similar to a test piece belongs actually to the same class of rhythm as the test piece. The results of this classification experiment are depicted in Table 9. It is apparent that the comparative approaches lead to a performance significantly better than random, which would be 25% for our compiled four-class datasets. In fact, accuracies are in the same range as the results of the PIK algorithm, with the STM approach outperforming the PIK on Turkish music (56.5% instead of 44.8%), but the PIK performing better on Carnatic music (64.1% instead of 42.2%). This might indicate the potential of combining self-contained and comparative approaches, because none of the approaches evaluated for cycle length recognition provides us with a sufficient performance for a practical application.

In a second experiment we evaluated in how far these approaches can serve to discriminate between the musical cultures. For that we applied the same 1-nearest-neighbor classification to the task of assigning a song to either Carnatic, Hindustani or Turkish music. The obtained accuracy values are 84.9% for the OP features and 71.8% for the STM features. These results indicate that the descriptors are able to capture important aspects of rhythm in Indian and Turkish music, but that those low-level features cannot serve as a sufficient solution on its own to estimate the length of rhythmic cycles in none of the three cultural contexts.

7 Downbeat recognition experiment

Dataset	Turkish		Carnatic	
	low-level annot. dataset		low-level annot. dataset	
Method	Aksak (9)	Düyek (8)	Ādi (8)	rūpaka (3)
DAV	37.2	36.3	21.7	41.2
HOC-SVM	60.7	59.8	22.9	42.1
HOC	86.5	85.5	49.9	64.4

Table 10: Accuracy of downbeat recognition. The measure lengths are indicated in parentheses next to the usul/tāla.

In Table 10, we depict the downbeat recognition accuracies (in %) for the two systems outlined in Section 4.3. The results are given separately for each of the two rhythm classes in the Carnatic and Turkish low-level-annotated datasets. The HOC algorithm was applied with and without emphasizing the low-frequency onsets, denoted as HOC and HOC-SVM, respectively. The DAV algorithm has the lowest accuracies for all presented rhythm classes. This is caused by the focus of the method on changes in harmony that is related to chord changes - concepts not present in Turkish or Indian music. However, the results obtained from HOC are more accurate and allow for two interesting conclusions. First, taking onsets in the low-frequency region into account improves recognition for all contained rhythms, and, second, downbeat recognition in Turkish music is much more accurate than in Carnatic music. The extent of accuracy for Turkish music is in the same range as the accuracy reported for electronic dance music (Hockman et al., 2012). However, Carnatic music with its wide rhythmic variations and its flexible rhythmical style seems to represent a more difficult challenge for downbeat recognition.

We want to remind the fact that the samples for Turkish music were chosen to contain percussion accompaniment. Pieces without such phenomenal cues are very likely to present both automatic systems and human listeners with a more difficult challenge when looking for the downbeat. Furthermore, the accuracies depicted in Table 10 can only be achieved with known cycle length, and correctly annotated beats. As we showed in the previous sections, it is currently not possible to automatically estimate these with a high

accuracy.

8 Discussion

We summarize and discuss the key results of the experiments in the previous three sections. Estimating the relevant metrical level at which beats need to be tracked is an important task. In Turkish music, we saw that KLA estimated the right metrical level in only about half the pieces and ELL tracked the relevant level in less than 20% of the pieces. Estimating the right tempo class is thus an imminent problem in Turkish music. A robust method for tempo estimation and beat tracking is therefore needed.

Though we described the beat pulsation as being mostly equally spaced for the purpose of evaluation in this paper, as we mentioned earlier, the beats in Turkish and Indian music can be non-uniform. Therefore, estimating the actual non-uniform beat instants in these additive rhythms is the musically meaningful task. This task would help further in estimating the right and left hand stroke pattern in Turkish music, the vibhāg structure in Hindustani music and the aṅga structure in Carnatic music. A non-zero eḍupu (phase offset from sama) poses a severe problem to estimating the correct phase of the tracked beat in Carnatic music. Since eḍupu is more of a musical concept without a direct manifestation in the signal, it is difficult to estimate. However, the beat tracking performance strongly depends on it and an estimation of eḍupu is critical.

Long āvart cycles, ambiguous definition of a beat, unequal vibhāgs are the most important challenges to beat tracking in Hindustani music. For an effective estimation of the beats, an estimation of the *lay* is necessary. Once we have the right *lay*, we can estimate the right tempo of this beat pulsation easily. For tāls with a very long cycle duration, methods that aim at tracking short time-span pulsation will be not adequate. For such cases, similarity-based methods based on concepts e.g. applied in the PIK algorithm might be more suitable, as they are capable of detecting repetitions caused by the grouping structure of a piece.

Cycle length recognition is challenging in Turkish and Indian music since cycles of different lengths exist at different time-scales. Although we defined the most important cycle to be at the āvart and usul level, the other cycles, mainly at the beat and subdivision level, also provide useful rhythm related information. The evaluated approaches PIK and GUL estimate the subdivision meter and time signature. This is possible to an acceptable level of accuracy, when restricting to a subset of rhythm classes with relatively simple subdivision meters. Though they do not provide a complete picture of the meter, they estimate the underlying metrical

structures at short time scales and can be used as pre-processing steps for estimating longer and more complex cycles.

Both SRI and KLA aimed to estimate the longer cycle lengths but show a performance that is inadequate for any practical application involving cycle length recognition. Since KLA has a better beat tracking performance than ELL, combining SRI and KLA to use the KLA beat tracker in SRI might lead to a better performance and needs to be explored further. Tempo estimation and beat tracking have a significant effect on cycle length recognition, especially in the self-contained approaches and also needs to be explored further.

The comparative approaches show that the applied signal features capture important aspects of rhythm but are not sufficient to be used standalone for cycle recognition. A combination of self-contained and comparative approaches might provide useful insights in rhythm description of Turkish and Indian music through mutual reinforcement. In summary, cycle length recognition in Turkish and Indian music has been addressed inadequately so far with the evaluated methods presenting results of generally insufficient accuracy. Developing methods that take the culture-specific properties of rhythm into account is therefore necessary to proceed towards a reliable computational rhythm analysis.

Downbeat recognition, *i.e.* the estimation of the instant of beginning of the measure cycle was explored using HOC and DAV algorithms. HOC provided reasonable results with Turkish music, but fell short of such good performance on Carnatic music. With a wide variety of rhythms, coupled with the perceptual edupu, Carnatic music poses a difficult challenge in sama recognition. Since there is no time adherence to an absolute clock, tempo drifts are common and lead to small shifts in the sama instants. The sam in Hindustani music is easily identified by listeners through the *ṭhēkā* of the Tabla and phrasing of the composition. Using the *ṭhēkā* information for sam estimation might improve the performance.

The downbeat detectors presented in this paper needed an estimation of beats and the cycle length of the piece, which are difficult to estimate presently. Since downbeat information can help in estimating the cycle length and also beat tracking, an independent approach to downbeat recognition will be very useful. A joint estimation of the beat, cycle length, and downbeat might be a potential solution since each of these parameters are mutually useful for estimating the others. A combination of bottom up and top down knowledge based approach which performs a joint estimation of these parameters is a promising direction to pursue future work.

We emphasize that complete automatic annotation of a *tāla* or *usul* cycle would require all three tasks

to be tractable with sufficient accuracy. In conclusion, beat tracking and cycle length recognition seem to represent bottlenecks for a complete annotation. Moreover, there are other challenges in the context of rhythmic analysis which need further exploration. They include distinguishing tālas and usuls with same cycle length, eḍupu recognition, tāla cycle tracking during highly complex rhythmic improvisation sections, and handling tempo drifts - all of which help in other music analysis tasks such as structural segmentation and motivic analysis.

9 Conclusions

Automatic rhythm annotation can provide important information for a better discovery and enhanced experience with music collections. We discussed three important automatic rhythm annotation tasks for Turkish, Carnatic and Hindustani music in this paper. We also provided an evaluation of the state of the art for these tasks. The results provide us with useful insights to indicate promising directions for future work.

Though we considered beat tracking, cycle length estimation and downbeat recognition as separate independent tasks, it might be better to consider a holistic approach and build a framework of methods in which the performance of each element can be influenced by estimations in another method. Ironically, we see from the HOC algorithm, a downbeat detector for electronic dance music, that sometimes the most rigid specialization leads to good performance on apparently completely different music. Thus, it still remains an open question if we need more specialist approaches, or more general approaches that are able to react to a large variety of music. Generally, it appears desirable to have generic approaches that can be adapted to a target music using machine learning methods. It is illusionary to believe that specialist systems can be developed for each of the musics of the world. However, mechanisms of music distribution, marketing and copyright management will increasingly depend on automatic analysis tools. Therefore, the development of generic and adaptive approaches has the potential to prevent various music cultures from being excluded from commercial distribution.

For estimating the components of meter from audio, we need signal descriptors that can be used to reliably infer the underlying meter from the surface rhythm in audio. The availability of such descriptors will greatly enhance the performance of automatic annotation algorithms. At present, we have suitable audio descriptors for low level rhythmic events such as note onsets and percussion strokes, but we aim to develop

descriptors for higher level rhythmic events in both Indian and Turkish music.

The inadequate performance of the presented approaches leads us to explore the specific problems more comprehensively. It also motivates us to explore varied and unconventional approaches to rhythm analysis. Further, building a larger annotated collection is necessary to generalize the algorithms better. As described in Section 2.3, extending the types of data that can be used for rhythm analysis, and using multiple datasources for a better rhythm description need to be explored. The automatic rhythmic description achieved through these tasks forms a part of the comprehensive system that helps to discover and explore these music traditions.

10 Acknowledgments

We are thankful to Dr. R. Reigle for inspiring discussion and help regarding terminology. We would also like to thank Dr. A. Klapuri, Dr. A. Pikrakis, Dr. M. E. P. Davies, J. Hockman, and S. Gulati providing their code and implementation of the algorithms for evaluation in this paper. This work is partly supported by the European Research Council under the European Union’s Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

11 Glossary

Carnatic Music

akṣara: The lowest metrical level pulse defined in theory

aṅga: The sections of a tāḷa

āvartana: One complete cycle of a tāḷa

eḍupu: The phase/offset of the composition relative to the sama

kāla: The tempo multiplying factor

mṛdaṅgam: The primary percussion accompaniment

naḍe: The subdivision meter within a beat (also called gati)

rāga: The melodic framework of Carnatic music

sama: The first akṣara of an āvartana (Downbeat)

tāla: The rhythmic framework of Carnatic music

ādi: A tāla with 32 akṣaras in a cycle

khaṇḍa chāpu: A tāla with 10 akṣaras in a cycle

miśra chāpu: A tāla with 14 akṣaras in a cycle

rūpaka: A tāla with 12 akṣaras in a cycle

Hindustani Music

āvart: One complete cycle of a tāl

lay: The tempo class

dhṛt: Fast tempo class

madhyam: Medium tempo class

vilāmbit: Slow tempo class

mātrā: The lowest metrical pulse level defined in theory

rāg: The melodic framework of Hindustani music

sam: The first mātrā of an āvart (Downbeat)

tāl: The rhythmic framework of Hindustani music

ektāl: A tāl with 12 mātrās in a cycle

jhaptāl: A tāl with 10 mātrās in a cycle

rūpak tāl: A tāl with 7 mātrās in a cycle

tīntāl: A tāl with 16 mātrās in a cycle

vibhāg: The sections of a tāl

Turkish Makam Music

makam: The melodic framework of Turkish art and folk music

meretebe: The denominator of the time signature of an usul. Defines tempo class.

usul: The rhythmic framework

aksak: An usul with 9 beats in a cycle

curcuna: An usul with 10 beats in a cycle
düyek: An usul with 8 beats in a cycle
yürüksemai: An usul with 6 beats in a cycle

12 References

- Antonopoulos, I., Pikrakis, A., Theodoridis, S., Cornelis, O., Moelants, D., & Leman, M. (2007). Music Retrieval by Rhythmic Similarity applied on Greek and African Traditional Music. In *Proc. of 8th International Conference on Music Information Retrieval (ISMIR 2007)* (pp. 297–300). Vienna, Austria.
- Cannam, C., Landone, C., & Sandler, M. (2010, October). Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files. In *Proc. of the ACM Multimedia 2010 International Conference* (pp. 1467–1468). Florence, Italy.
- Chordia, P. (2005). *Automatic Transcription of Solo Tabla Music* (PhD Thesis). Stanford University.
- Clayton, M. (1996). Free Rhythm: Ethnomusicology and the Study of Music Without Metre. *Bulletin of the School of Oriental and African Studies*, 59, 323–332.
- Clayton, M. (2000). *Time in Indian Music : Rhythm, Metre and Form in North Indian Rag Performance*. Oxford University Press.
- Cooper, G., & Meyer, L. B. (1960). *The Rhythmic Structure of Music*. University of Chicago Press.
- Şentürk, S., Holzapfel, A., & Serra, X. (2012, July). An Approach for Linking Score and Audio Recordings in Makam Music of Turkey. In *Proc. of the 2nd CompMusic Workshop* (pp. 95–106). Istanbul, Turkey.
- Davies, M. E. P., Degara, N., & Plumbley, M. D. (2009, October). Evaluation Methods for Musical Audio Beat Tracking Algorithms. *Technical Report C4DM-TR-09-06, Queen Mary University of London*.
- Davies, M. E. P., & Plumbley, M. D. (2006, September). A spectral difference approach to downbeat extraction in musical audio. In *Proc. of the 14th European Signal Processing Conference (EUSIPCO 2006)*. Florence, Italy.
- Davies, M. E. P., & Plumbley, M. D. (2007, March). Context-Dependent Beat Tracking of Musical Audio. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3), 1009–1020.

- Dutta, A. E. (1995). *Tabla: Lessons and Practice*. Ali Akbar College.
- Ellis, D. P. W. (2007). Beat Tracking by Dynamic Programming. *Journal of New Music Research*, 36(1), 51–60.
- Gulati, S., & Rao, P. (2010, December). Rhythm Pattern Representations for Tempo Detection in Music. In *Proc. of the First International Conference on Intelligent Interactive Technologies and Multimedia (IITM-2010)* (pp. 241–244). Allahabad, India.
- Gulati, S., Rao, V., & Rao, P. (2011, March). Meter detection from audio for Indian music. In *Proc. of 8th International Symposium on Computer Music Modeling and Retrieval (CMMR)* (pp. 34–43). Bhubaneswar, India.
- Hannon, E. E., Soley, G., & Ullal, S. (2012). Familiarity overrides complexity in rhythm perception: A cross-cultural comparison of American and Turkish listeners. *Journal of Experimental Psychology: Human Perception and Performance*, 38(3), 543–548.
- Hockman, J. A., Davies, M. E. P., & Fujinaga, I. (2012, October). One in the Jungle: Downbeat Detection in Hardcore, Jungle, and Drum and Bass. In *Proc. of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)* (pp. 169–174). Porto, Portugal.
- Holzapfel, A., & Bozkurt, B. (2012, July). Metrical Strength and Contradiction in Turkish Makam Music. In *Proc. of 2nd CompMusic Workshop* (pp. 79–84). Istanbul, Turkey.
- Holzapfel, A., Davies, M., Zapata, J., Oliveira, J., & Gouyon, F. (2012, November). Selective Sampling for Beat Tracking Evaluation. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(9), 2539–2548.
- Holzapfel, A., Flexer, A., & Widmer, G. (2011, July). Improving Tempo-sensitive and Tempo-robust Descriptors for Rhythmic Similarity. In *Proc. of Conference on Sound and Music Computing (SMC 2011)* (pp. 247–252). Padova, Italy.
- Holzapfel, A., & Stylianou, Y. (2009, October). Rhythmic Similarity in Traditional Turkish Music. In *Proc. of 10th International Conference on Music Information Retrieval (ISMIR 2009)* (pp. 99–104). Kobe, Japan.
- Holzapfel, A., & Stylianou, Y. (2011). Scale transform in rhythmic similarity of music. *IEEE Transactions on Speech and Audio Processing*, 19(1), 176–185.
- Jehan, T. (2005). *Creating music by listening* (PhD Thesis). Massachusetts Institute of Technology.

- Keçecioglu, M. U. (2010). *Usul-Velvele Editor*. Retrieved October 27, 2013, from http://notist.org/usul_velvele%20editor_ing.html
- Klapuri, A. P., Eronen, A. J., & Astola, J. T. (2006). Analysis of the Meter of Acoustic Musical Signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1), 342–355.
- Kolinski, M. (1973). A Cross-Cultural Approach to Metro-Rhythmic Patterns. *Ethnomusicology*, 17(3), 494–506.
- LaRue, J. (1992). *Guidelines for Style Analysis, 2nd Ed, Detroit Monographs in Musicology*. Harmonie Park Press, Michigan.
- Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press Cambridge.
- Ligeti, G. (2007). Brief an Kai Jakobs. In M. Lichtenfeld (Ed.), *Gesammelte Schriften*. Paul Sacher Stiftung.
- London, J. (2001). Metre. In L. Macy (Ed.), *Grove Music Online. Oxford Music Online*. Oxford University Press. Retrieved 27 October, 2013, from <http://www.grovemusic.com>
- London, J. (2004). *Hearing in Time: Psychological Aspects of Musical Meter*. Oxford: Oxford University Press.
- McKinney, M. F., Moelants, D., Davies, M. E. P., & Klapuri, A. (2007). Evaluation of Audio Beat Tracking and Music Tempo Extraction Algorithms. *Journal of New Music Research*, 36(1), 1–16.
- Miron, M. (2011). *Automatic Detection of Hindustani Talas* (Master's Thesis). Universitat Pompeu Fabra, Barcelona, Spain.
- Miron, M., Gouyon, F., Davies, M. E., & Holzapfel, A. (2013). Beat-Station: A real-time rhythm annotation software. In *Proc. of Sound and Music Computing Conference (SMC 2013)* (pp. 729–734). Stockholm, Sweden.
- Moelants, D., & McKinney, M. F. (2004). Tempo Perception and Musical Content: What makes a piece fast, slow or temporally ambiguous ? In *Proc. of the 8th International Conference on Music Perception and Cognition (ICMPC)* (pp. 558–562). Evanston, IL, USA.
- Müller, M., Ellis, D. P. W., Klapuri, A., Richard, G., & Sagayama, S. (2011). Introduction to the Special Issue on Music Signal Processing. *IEEE Journal of Selected Topics in Signal Processing*, 5(6), 1085–1087.
- Naimpalli, S. (2005). *Theory and practice of Tabla*. Popular Prakashan.
- Özkan, I. H. (1984). *Türk Müsikisi Nazariyati ve Usulleri*. Otuken Nesriyat.
- Parncutt, R. (1994). A Perceptual Model of Pulse Saliency and Metrical Accent in Musical Rhythms. *Music*

- Perception*, 11(4), 409–464.
- Pikrakis, A., Antonopoulos, I., & Theodoridis, S. (2004, October). Music meter and tempo tracking from raw polyphonic audio. In *Proc. of 5th International Conference on Music Information Retrieval (ISMIR 2004)*. Barcelona, Spain.
- Pohle, T., Schnitzer, D., Schedl, M., & Knees, P. (2009, October). On Rhythm and General Music Similarity. In *Proc. of 10th International Conference on Music Information Retrieval (ISMIR 2009)* (pp. 525–530). Kobe, Japan.
- Sachs, C. (1953). *Rhythm and Tempo*. W. W. Norton & Co.
- Sambamoorthy, P. (1998). *South Indian Music Vol. I-VI*. The Indian Music Publishing House.
- Serra, X. (2011, October). A Multicultural Approach in Music Information Research. In *Proc. of 12th International Society for Music Information Retrieval Conference (ISMIR 2011)* (pp. 151–156). Miami, USA.
- Srinivasamurthy, A., Subramanian, S., Tronel, G., & Chordia, P. (2012, July). A Beat Tracking Approach to Complete Description of Rhythm in Indian Classical Music. In *Proc. of the 2nd CompMusic Workshop* (pp. 73–78). Istanbul, Turkey.
- Tāla in Carnatic Music: Examples*. (2013). Retrieved 27/10/2013, from <http://compmusic.upf.edu/node/185>
- Tāl in Hindustani Music: Examples*. (2013). Retrieved 27/10/2013, from <http://compmusic.upf.edu/node/186>
- Tenzer, M. (2006). Introduction: Analysis, Categorization and Theory of Musics of the World. In M. Tenzer (Ed.), *Analytical Studies in World Music* (pp. 3–38). Oxford University Press.
- Usul in Turkish Makam Music: Examples*. (2013). Retrieved 27/10/2013, from <http://compmusic.upf.edu/node/187>
- Widdess, R. (1994). Involving the Performers in Transcription and Analysis: A Collaborative Approach to Dhrupad. *Ethnomusicology*, 38(1), 59–79.
- Zapata, J. R., Holzapfel, A., Davies, M. E. P., Oliveira, J. L., & Gouyon, F. (2012, October). Assigning a Confidence Threshold on Automatic Beat Annotation in Large Datasets. In *Proc. of 13th International Society for Music Information Retrieval Conference (ISMIR 2012)* (pp. 157–162). Porto, Portugal.