Evaluation and Recommendation of Pulse and Tempo Annotation in Ethnic Music

Olmo Cornelis[1], Joren Six[1], Andre Holzapfel[2], and Marc Leman[3]

[1]University College Ghent, School of Arts
Hoogpoort 64, 9000 Ghent - Belgium
olmo.cornelis@hogent.be
joren.six@hogent.be

[2]Bahçeşehir University, Electrical and Electronics Engineering
Çirağan Cd. 4, Beşiktaş, 34353 Istanbul - Turkey
xyzapfel@gmail.com

[3]Ghent University, IPEM - Department of Musicology
Blandijnberg 2, 9000 Ghent - Belgium
marc.leman@ugent.be

February 4, 2013

Abstract

Large digital archives of ethnic music require automatic tools to provide musical content descriptions. While various automatic approaches are available, they are to a wide extent developed for Western popular music. This paper aims to analyze how automated tempo estimation approaches perform in the context of Central-African music. To this end we collect human beat annotations for a set of musical fragments, and compare them with automatic beat tracking sequences. We first analyze the tempo estimations derived from annotations and beat tracking results. Then we examine an approach, based on mutual agreement between automatic and human annotations, to automate such analysis, which can serve to detect musical fragments with high tempo ambiguity.

Keywords: Ethnic Music, Beat Estimation, Tempo Annotation, Tempo Perception, Ambiguity

1 Introduction

In an effort to preserve the musical heritage of various cultures, large audio archives with ethnic music have been created at several places throughout the world[1]. With the widespread availability of digital audio technology, many archiving institutions have started to digitize their audio collections to facilitate better preservation and access[2]. Meanwhile, a good number of audio collections have been fully digitized, which enables the next step to make these au-

[1]British Library (London), CREM and SDM (Paris), Ethnologisches Museum (Berlin), RMCA (Brussels), Essen Folksong Collection (Warsaw), GTF (Vienna) and many more

[2]See Appendix C for a number references to digitization projects.

dio archives more accessible for researchers and general audiences.

Computational Ethnomusicology from this perspective, aims at providing better access to ethnic audio music collections using modern approaches of content-based search and retrieval (Tzanetakis et al., 2007; Cornelis et al., 2010). This research field has its roots in Western Musicology, as well as in Ethnomusicology and Music Information Retrieval. Current computational tools for the content-based analysis of Western musical audio signals are well established and have begun to reach a fair performance level as seen in many applications, publications and the MIREX initiative[3]. However, for the field of ethnic music, it is still unclear which computational tools for content-based analysis can be applied successfully. Given the diversity and oral character of ethnic music, Computational Ethnomusicology faces many challenges. A major difficulty is concerned with the influence and dominance of Western musical concepts in content-based analysis tools. It is generally believed that the influence of Western concepts may affect the interpretation of the extracted audio features. However, there is little information about the exact nature of this possible contamination. It may be that tools based on low-level acoustical features perform reasonably well, while tools that focus on higher-level musical concepts perform less well. In this context, one could question whether existing beat tracking and tempo extraction tools, typically developed and tested on, mainly, Western music, can be readily applied to African music.

In this paper, we focus on tools for beat tracking and tempo extraction from Central-African music. The overall aim of this study is to see to what extent meaningful results can be expected from the automatic tempo analysis of Central-African music. The research in this paper relies on existing computational tools, and does not aim to introduce novel approaches in beat tracking and tempo estimation. A useful byproduct of this research could be a new way to identify ethnic music with ambiguous tempo relations and reveal information of a higher metrical hierarchy: from beats to meter.

Our goal is to explore whether a set of 17 automatic beat trackers and tempo estimators (i) can be used as a tool for extracting tempo from Central-African musical audio, (ii) can give insight into the ambiguity of tempo perception, (iii) can detect problematic cases for tempo annotation, and (iv) if it can provide information about a higher metrical level.

In order to be able to evaluate the performance of the beat trackers, we compare them with the performance of 25 professional musicians, who manually annotated the beat for 70 audio fragments. The results of both human and computational annotations are analyzed and compared with each other. The goal is to see how large the variability is in both sets of annotations (automatic and manual) and whether ambiguity in human annotations implies ambiguity in computational annotations, and how well the two match.

The paper is structured as follows; Section 2 presents aspects of tempo in music. Section 3 gives an overview of related literature. Section 4 outlines our methodology and describes the used data collection. Section 5 contains the results of these experiments. Section 6 elaborates on considerations in the field of approaching ethnic music. Section 7 concludes the paper.

2 On the concept of tempo

Willenze (1964) points out the relationship between the measurable, or objective time and the time that is experienced, the subjective time. This reflects the traditional distinction between the theoretical tempo that is implied in a score, and the tempo that comes out of performance. Although the score written by a composer is handled as a primary source, musical notation in the case of transcription is typically considered to be a subjective assessment of the transcriber. Especially in the area of ethnic music this has been mentioned several times, as for example in the work of Brandel (1961).

Subjective assessments of tempo in music are determined by studying synchronization with the pulse.

[3]The Music Information Retrieval Evaluation eXchange (MIREX) is an annual evaluation campaign for Music Information Retrieval (MIR) algorithms. More info about MIREX can be found on `http://www.music-ir.org`

However, at least in Western music, the pulse often functions within a larger structure that is called the meter. Lerdahl & Jackendoff (1983) speak about strong and weak beats (*instances of a pulse*) and they approach meter as a super structure on top of *"a relatively local phenomenon"*. The perception of pulse and meter is associated with a perceivable regularity that creates expectations in a time span. For this reason, one can tap along with any music that has a regular/repetitive basis. Therefore, meter facilitates the structuring of the beats over time.

Non-Western rhythmical phenomena are different from Western rhythmical phenomena. Ethnomusicologists tend to recognize the concept of pulse that organizes music in time, but they assess the structuring of pulses in a way that is different from the concept of meter. From all their theories and concepts, the idea of the fastest pulse as a basis for understanding aspects of timing seems to be the most fundamental, general, and useful, as it allows the widest variety of interpretations. In this context, Arom (1985) states that African music is not based on bars, which define the meter as in classical music, but on pulsations, a succession of isochronous time units.

Thus, rather than using the concept of meter, the structuring of pulses is based on the concept of sequences, forming the starting point for further analysis of rhythms. The best-known approach is the Time Unit Box System (TUBS) notation, developed by Kubik (1994); Koetting (1970) for annotating West African drums. It is a graphical annotation approach that consists of boxes of equal length put in horizontal sequence. Each box represents an instance of the fastest pulse in a particular musical piece. If an event occurs, the box is marked, if not the box is left empty. TUBS are most useful for showing relationships between layers of complex rhythms. An example of this notation can be found in 8

The approach of rhythmical organization by Kubik (1994); Koetting (1970) is based on three levels. The first level is the elementary pulsation, a framework of fast beats that define the smallest regular units of a performance as a unheard grid in the mind of the performer. The second level is formed by a subjective reference beat. There are no preconceived strong or weak parts of the meter, and the beats are often organized in a repetitive grid of 3, 4, 6, 8 or 12 units. The point of departure is so ingrained that it needs no special emphasis. For this reason, the first beat is often acoustically veiled or unsounded. For outsiders this can cause a phase shift. On top of these two levels, Kubik adds a third level, which he calls the cycle. A cycle would typically contain 16 to 48 beats. The introduction of numbered cycles (Kubik, 1960) replaced conventional Western time signatures in many transcriptions of African music. The main advantage of conceiving these large cycles is that polymeter structures resolve in it.

Agawu (2003) introduced *topoi*, which are short distinct, memorable rhythmic figures of modest duration that serve as a point of temporal reference. The presence of these repetitive topoi shows that there is an underlying pulse. He writes that "*West and Central African dances feature a prominently articulated, recurring rhythmic pattern that serves as an identifying signature*". Seifert et al. (1995) followed a similar path of the smallest pulse as basis for a theoretical and integrated research strategy for the interpretation of non-Western rhythmical phenomena, based on the TUBS of Kubik and Koetting.

Connected to the idea of the fastest pulse, Jones (1959) was the first to describe the asymmetric structure of the higher rhythmical patterns. A well-known common example of such pattern is the 12-beat pattern that contains a seven and a five stroke component, of which one is prevalent while its complementary pattern is latent, and is tapped as a syncopated pulse. The pattern appears later as an example in Section 5 and is illustrated by Figure 3.

Another prominent rhythmical phenomenon in African music are interlocking patterns. They consist of two or more (rhythmic or melodic) lines that have different starting points, running one smallest beat apart from each other. Kubik suggests that the origin of these interlocking patterns could have initiated from pestle-pounding strokes by two or three women that alternately strike in a mortar. The patterns are fundamental to much African music.

A final remark concerns a call by Agawu (1995) for rebalancing the presumed importance of rhythmical elements in African music over the other musical parameters. Agawu (2003) believes that the rhythmical elements and their organization in African music are over-conceptualized. In his writings he lists, quotes, and reviews many of the great ethnomusicologists' ideas of the 20th century. Contrary to these ideas, he suggests a more explorative bottom-up approach and he warns ethnomusicologists against the eagerness of constructing African music as essentially different from the West.

This shows that the concepts of pulse, meter, and tempo are still a topic of discussion, and that this discussion should be taken into account when trying to apply computational content-based analysis methods to Central-African music.

3 Literature on Tapping Experiments

Apart from concepts on pulse, meter, sequences, and tempo, it is also of interest to consider experiments on tapping. Experiments on synchronized finger tapping along with beat of the music (Repp, 2006; Large, 2000; Desain & Windsor, 2000; Moelants & McKinney, 2004; Wohlschlager & Koch, 2000) reveal some interesting aspects that should be taken into account when studying beat and tempo in Central-African Music.

One aspect concerns the range in which musical tempo can be perceived, namely, between 200 to 1500 milliseconds, or 40 to 300 Beats Per Minute (*bpm*) (Pöppel et al., 1978; Moelants & McKinney, 2004). In cases of slower tempi one tends to subdivide, while faster tempi physically cannot be performed. Within that space, Moelants mentions there is a preferred tempo-octave lying between 81 and 162 *bpm*.

It is perhaps superfluous to mention that the regularity of beats is never strictly rigid. In musical performances as well as in human synchronization tapping tasks, minor deviations are present in the signal and data, but these are inherent to musical and to human performance. They do not influence the global tempo, but are characteristics of the microtiming in the music. A related aspect concerns the *negative asynchrony* (Repp, 2006), the phenomenon that subjects tend to tap earlier than the stimulus (typically between 20 and 60 ms), which shows that subjects perform motor planning, and thus rely on anticipation, during the synchronization task (Dixon, 2002).

Another aspect concerns *tempo octaves*, the phenomenon that subjects tend to synchronize their taps with divisions or multiplications of the main tempo. These tempo octaves are regularly reported and they are the main argument to identify a tempo as being *ambiguous*. Indeed, the human perceivable tempo limitations (40-300 *bpm*) span a large range of tempi, namely, more or less three tempo-octaves. Consequentially, the listener has different possibilities in synchronizing (tapping) with the music. Therefore, ambiguity arises in the tempo annotations of a group of people. These choises are related to personal preference, details in the performance, and temporary mood of the listener (Moelants, 2001). This subjectivity has large consequences in approaching tempo and meter in a scientific study. However, McKinney & Moelants (2006) demonstrate that for pieces with tempi around 120 *bpm*, a large majority of listeners are very likely to perceive this very tempo, whereas faster and slower tempi induce more ambiguity, with responses spread over two tempo-octaves (Moelants & McKinney, 2004). This connects to the *2Hz resonance theory of tempo perception* (Van Noorden & Moelants, 1999), according to which tempo perception and production is closely related to natural movement, with humans functioning as a resonating system with a natural frequency. The preferred tempo is located somewhere between 110 and 130 *bpm*, and therefore creates a region in which music is tapped less ambiguously (Moelants, 2002).

In this perspective, it is possible to distinguish between beat rate and/or tapping rate on the one hand, and the perceived tempo on the other hand (Epstein, 1995). The beat rate is the periodicity which best affords some form of bodily synchronization with the rhythmic stimulus. It may or it may not directly correspond to the perceived tempo, especially when the latter is considered as a number that reflects a rather complex *Gestalt* that comes out of the sum of musical factors, combining the overall sense of a work's themes, rhythms, articulations, breathing, motion, harmonic progressions, tonal movement, and contrapuntal activity. As such, the beat could be different from the perceived tempo. Early research by Bolton (1894) reported already the *phenomenal grouping* as an aspect of synchronized tapping; when he presented perfectly isochronous and identical stimuli to subjects they spontanously subdivided, by accentuation, into units of two, three, or four. London (2011) speaks of *hierarchically-nested periodicities* that a rhythmic pattern embodies. The observation of subdivisions and periodicity brings Parncutt (1994) to the question what *phase* listeners tend to synchronise to when listening to music and what cues in the musical structure influence these decisions.

Another aspect concerns the *ambiguity of meter perception* (McKinney & Moelants, 2006). In music theory, the meter of a piece is considered as an unambiguous factor, but some music could be interpreted both with a binary or a ternary metric structure. Handel & Oshinsky (1981) presented a set of polyrhythmic pulses and asked people to synchronize along with them. The general outcome was that 80% of the subjects tapped in synchrony with one of the two pulses, whereas 12% of the subjects tapped the co-occurrence of the two pulses, and 6% tapped every second or third beat. The choice of preferred pulse however was not clear. A conclusion was that subjects tend to follow the fastest of the two pulses that make the polyrhythm when the global tempo is slow, and that subjects tend to follow the slowest pulse in a fast global tempo. When the global tempo is too high, people switch to a lower tempo octave. If the presented polyrhythm consists of different pitch content, the lower pitch element was the preferred frequency. Finally, Handel and Oshinsky concluded that if the tempo of the presented series of beats is very high, the elements are temporally so tightly packed that the pulse becomes part of the musical foreground instead of the pulsation that is part of the musical background. For polyrhythms, this transition point is about $200ms$ or $300bpm$.

The above overview shows that research on synchronized tapping tasks has to take into account several aspects that are likely to be highly relevant in the context of Central-African stimuli where we typically deal with complex polyrhythms.

4 Methodology

4.1 Experiment 1: Human

Procedure: Tap Along

Tempo annotation is the ascription of a general tempo to a musical piece, expressed in beats per minute (*bpm*). Beat synchronisation is the underlying task for the identification of a basic pulse from which the tempo is derived. Subjects were asked to *tap to the most salient beat* of the audio fragments. More information on the stimuli can be found in section 4.1.1. For each tap annotation containing taps at the time instances $t_1, ..., t_N$ (s), we obtain a set of $N-1$ inter-tap distances $D = d_1, ..., d_{N-1}$ (s). Then, a tempo in *bpm* is assigned to the piece by calculating the median of $60/D$.

The experiment was done on a laptop with the subjects listening to the audio fragments on headphones while tapping on the keyboard space bar. Since manual annotation of tempo is an intense and time consuming task, the data was recorded in two sessions with a small pause between the two. Subjects could restart any fragment if they had doubts about their annotation. The number of retries and the tapping data for each retry were recorded together with the final tapping data. All the data was organized and recorded by the software Pure Data. To ensure that the data is gathered correctly a test with a click track was done, with the interval between the clicks being constantly $500ms$. The average tapping interval was $499.36ms$, with a standard deviation of $20ms$. The low standard deviation implies that the measurement system has sufficient granularity for a tapping experiment.

4.1.1 Stimuli: Audio Fragments

The stimuli used in the experiment were 70 sound fragments, each with a length of 20 seconds, selected from the digitized sound archive of the Royal Museum for Central Africa (RCMA), Tervuren, Belgium. The archive of the Department of Ethnomusicology contains at present about 8,000 musical instrument and 50,000 sound recordings, with a total of 3,000 hours of music, most of which are field recordings made in Central Africa with the oldest recordings dating back to 1910. The archive has been digitized not only to preserve the music but also to make it more accessible (Cornelis et al., 2005). Results of the digitisation project can be found at `http://music.africamuseum.be`. The 70 fragments were chosen from the RMCA archive. It was attempted to cover a wide range of tempi and to include only fragments without tempo changes. The songs contained singing, percussion, and other musical instruments, in soloist or in group performances. This set of 70 stimuli will be referred to as *fragments* in the subsequent sections.

4.1.2 Participants: Musicians

The experiment was carried out by 25 participants. All of them were music students at the University College Ghent - School of Arts (Belgium), who were expected to play, practice, and perform music for several hours a day. The group consisted of 14 men and 11 women, ranging in age from 20 to 34 years.

4.2 Experiment 2: Software

Within the Music Information Retrieval community automated tempo estimation and beat tracking are important research topics. While the goal of the former is usually the estimation of a tempo value in *bpm*, the latter aims at estimating a sequence of time values that coincides with the beat of the music. Beat tracking and tempo estimation are applied in diverse applications, such as score alignment, structure analysis, play-list generation, and cover song identification. This paper however does not compare or evaluate such algorithmic approaches. For these matters, please refer to Gouyon et al. (2006); Zapata & Gómez (2011), and the yearly MIREX competition[4].

Automatic tempo analysis was done on the stimuli by a set of 17 beat trackers and tempo estimation algorithms (see appendix B). All parameters for each algorithm were left on the default values and no adaption to the stimuli was pursued. Some algorithms only give an ordered list of tempo suggestions (Beatcounter, Mixmeister, Auftakt), here only the primary tempo annotation was considered. For the beat tracking algorithms, a tempo estimation was derived from the beat sequences in the same way as for the human taps as described in Section 4.1. To be able to compare the results of the automatic tempo analysis with the human annotations, the same stimuli were used as in the first experiment (see Section 4.1).

[4]http://www.music-ir.org

4.3 Comparison: Measuring beat sequence/annotation agreement

Recently, a method based on mutual agreement measurements of beat sequences was proposed by Holzapfel et al. (2012). This method was applied for the automatic selection of informative examples for beat tracking evaluation. It was shown that the Mean Mutual Agreement (MMA) between beat sequences can serve as a good indicator for the difficulty of a musical fragment for either automatic or human beat annotation. A threshold on MMA could be established above which beat tracking was assumed to be feasible to a subjectively satisfying level. For the beat sequence evaluation in this paper, 5 out of the 17 algorithms were selected (Oliveira et al., 2010; Degara et al., 2011; Ellis, 2007; Dixon, 2007; Klapuri et al., 2006). This selection was made for several reasons. First, some of the 17 approaches are pure tempo estimators that give only tempo values in *bpm*, and not beat sequences. Second, in Holzapfel et al. (2012) it was shown that this selection increases diversity and accuracy of the included beat sequences, and, third, this selection guarantees comparability with results presented in Holzapfel et al. (2012).

Comparing beat sequences is not a straightforward task; two sequences should be considered to agree not only in case of a perfect fit, but also in the presence of deviations that result in perceptually equal acceptable beat annotations. Such deviations include small timing deviations, tempi related by a factor of two, and a phase inversion (off-beat) between two sequences, to name only the most important factors that should not be considered as complete disagreement. Because of the difficulty of assessing agreement between beat sequences, various measures have been proposed that differ widely regarding their characteristics (Davies et al., 2009). In this paper we restrict ourselves to two evaluation measures that are suitable for the two tasks at hand, which are spotting complete disagreement between sequences and investigating the types of deviations between sequences.

1. Information Gain (Davies et al., 2011): Local timing deviations between beat sequences are summarized in a beat error histogram. The beat error histogram is characterized by a concentration of magnitudes in one or a few bins if sequences are strongly related, and by a flatter shape if the two sequences are unrelated. The deviation of this histogram from the uniform distribution, the so-called "information gain", is measured using K-L divergence. The range of values for Information Gain is from 0 bits to 5.3 bits, with the default parameters as proposed in (Davies et al., 2011). This measure punishes completely unrelated sequences with a value of 0 bits, while all sequences with some meaningful relation tend to score higher. Such meaningful relations include a constant beat-relative phase shift, or simple integer relations between the tempi of the sequences. This means that off-beat or octave differences do not lead to a strong decrease in this measure. The maximum score can only be reached when all beats errors between the two sequences fall into the same beat error histogram bin, with the bin width being, for example $12.5ms$ at $120bpm$. MMA measured with this measure will be denoted as MMA_D.

2. F-measure: A beat in one sequence is considered to agree with the second sequence if it falls within a ±70ms tolerance window around a beat in the second sequence. Let the two sequences have $|A|$ and $|B|$ beats, respectively. We denote the number of beats in the first sequence that fall into such a window of the second sequence as $|A_{win}|$, and the number of beats in the second sequence that have a beat of the first sequence in their tolerance window as $|B_{win}|$. Note that if several beats of the first sequence fall into one tolerance window, $|A_{win}|$ is only incremented by one. Then F-measure is calculated as

$$F = \frac{2 * P * R}{P + R} \tag{1}$$

with $P = |A_{win}|/|A|$ and $R = |B_{win}|/|B|$. The F-measure has a range from 0% to 100% and drops to about 66% when two sequences are related by a factor of two, while a value of 0% is usually only observed when two sequences have the exact same period, but a phase offset. Note that two unrelated sequences do not score zero but about 25% (Davies et al., 2009). MMA measured with this measure will be denoted as MMA_F.

We will investigate, how many fragments in the RMCA subset can be successfully processed with automatic beat tracking, and to what extent the human annotations correlate with the estimated beat sequences. For this task MMA_D will be applied, as it was shown in Holzapfel et al. (2012) to reliably spot difficult musical fragments. For the fragments, which were judged to be processable by automatic beat tracking, we will apply MMA_F, as we can differentiate which types of errors occured for a given fragment. For example, values of 66% are mostly related to octave relations between the compared sequences, and an off-beat relation is in practice the only case which results into a value of 0%.

The MMA values for a fragment will be obtained by computing the mean of the $N(N-1)/2$ mutual agreements, with $N = 5$ for beat trackers, and $N = 25$ for human annotations. We will differentiate between beat sequences, which are obtained from algorithms (referred to as BT), and tapped annotations from human annotators (referred to as TAP).

5 Results

5.1 Human tempo annotations

In Appendix A we list the tempo annotations for all songs and all annotators. We assigned a general tempo value to each song by choosing the tempo that most people tapped. A tempo was considered similar if it did not deviate by more than $5bpm$ from

Type	#	%	Track ID's
Unanimous tempo	2	2.9%	5, 56
+ Tempo Octaves (no related)	23	32.9%	4, 6, 7, 8, 9, 10, 13, 14, 15, 17, 23, 25, 35, 42, 44, 50, 51, 55, 57, 58, 60, 65, 70
Tempo octaves < Related tempi	19	27.1%	28, 1, 62, 22, 20, 59, 63, 18, 41, 66, 53, 54, 37, 43, 52, 26, 39, 19, 64
Tempo octaves = Related tempi	3	4.3%	29, 34, 45
Tempo octaves > Related tempi	19	27.1%	69, 32, 38, 48, 61, 30, 33, 40, 24, 27, 47, 68, 12, 31, 67, 36, 49, 11, 3
+ Related tempi (no octaves)	2	2.9%	2, 46
No tempo	2	2.9%	16, 21
Total number of records	70		

Table 1: Overview of audio fragments organized by sorts of human assigned tempi.

the assigned tempo. The other tempi were considered in relation to this assigned tempo, and could be divided into tempo octaves (halve, double, triple tempo), related tempi (usually a mathematical relation with the assigned tempi), related octaves (halve, double, triple of the related tempo), unrelated tempi (no relation with the assigned tempo). Also some people tapped annotations of different length creating a pattern as, for example, 2 + 3 in a meter of 5 and 2 + 3 + 3 for some songs in 8, and those were specified as patterns without attempting to derive a tempo value from them.

A first glance at the results, Table 1, shows that 68 songs could be assigned a general tempo, two songs had such wide range of tempi that no general tempo could be assigned. They were both a capella vocal songs, that contained rather recitation than singing. Of the remaining 68 songs, only two songs were labeled unanimously. For 64 songs people tapped tempo octaves, and for 43 songs also related tempi were present. For the songs that had both octaves and related tempi, the distribution was equal: 19 songs had more octaves than related tempi, and 19 songs had more related tempi than octaves. This last group, which formed 27%, can be seen as songs with high ambiguity in tempo perception. These songs contained several instruments that combined polymetric layers. People tended to have distributed preference in following different instruments.

Table 2 lists the distribution of all 1750 annotations: 60% correspond to the assigned tempo, 17% correspond to tempo octaves, while only 9% correspond to related tempi. Apparently, in many songs (61%) some people do hear related tempi, but mostly this is a small group of people. But, even after applying a threshold on the minimum number of relation occurrances as in Table 3, 23% of the songs were still tapped in a related tempo by 5 or more persons (from the 25). This shows that related tempi are not coincidental or individual cases, but that a quarter of the audio set had tempo ambiguity, similar to what was derived in the previous paragraph.

Type	Human (%)	BT (%)
Identical	60%	48%
Octave	17%	18%
Related	9%	19%
Related Tempo Octave	3%	3%
Unrelated	9%	6%
Pattern	2%	0%

Table 2: Distribution of all annotations (1750 human annotations, 1190 BT tempi) over available classes.

The individual differences on the median over the 70 songs was remarkable, with personal medians ranging from 77 up to 133*bpm*. In affirmation with some elements from the literature, there is indeed a large agreement on tempo annotations in the region 120-130*bpm*, namely 83% (10% tapped a tempo oc-

		At least one		More than One		More than Two		More than Five	
Human	Tempo octaves	64	91%	56	80%	44	63%	28	40%
	Related tempo	43	61%	32	46%	25	36%	16	23%
	Related octave	25	36%	13	19%	7	10%	1	1%
	Pattern	37	53%	24	34%	15	21%	10	14%
	Unrelated tempo	19	27%	11	16%	6	9%	2	3%
BT	Identical	64	94%	61	90%	58	85%		
	Octave	52	76%	41	60%	28	41%		
	Related	52	76%	38	56%	28	41%		
	Related Octave	18	26%	9	13%	5	7%		
	Unrelated tempo	31	46%	20	29%	13	19%		

Table 3: Distribution of all annotations over available classes if a threshold is set.

Meter	Identical	Octave	Related
(1)	1 25 58 60	0	0
2	5 20 27 31 34 35 43 44 47 51 53	42 64 70	3
3	2 10 12 26 33 40 45	0	19 29 36 38 59 61
4	4 6 9 13 15 17 18 23 37 50 54 55 56 69	8 14 30 57	24 67
5	22 41 46 49	52	11
6	7 28 32 39 48 66	63	62 65 68

Table 4: BT annotations organized by meter and their classification along the human tempo references.

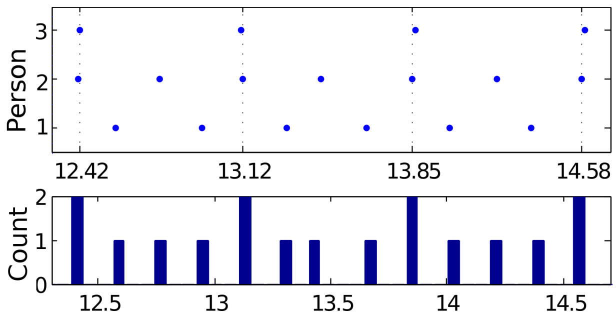


Figure 1: Small fragment (Track 25) of tapped onsets of three persons, one following the tempo octave (tempo halving), and two persons in different phase. Histogram below.

tave, and only 2% tapped a related tempo for this tempo region). 8 of the 10 songs in this tempo region were tapped with a binary meter. In the other tempi regions, ambiguity was much higher, but the set was too small to deduce tendencies. What was noticed is that songs around 90*bpm* received only few tempo octaves, but more related tempi.

When we focus on the properties of individual songs, the pieces with a meter in five deserve special attention. The annotations were very diverse, and can be divided into different groups. Some people tapped exactly on the fastest pulse, while others only tapped each fifth beat of this pulse, creating a tempo range of 5 tempo octaves. Some people tapped every second beat of the fastest pulse level, which implies going "on and off beat" per bar, creating an alternating syncopation. Several people tapped a subdivided pattern of 2 and 3 beats and some people tapped every 2.5 beats, subdividing the meter of five into two equal parts. This diversity reoccurred for each song that had a meter in five.

Agawu mentions that cultural insiders easily identify the pulse. For those who are unfamiliar with such specific culture, and especially if the dance or choreographic movements cannot be observed, it can be difficult to locate the main beat and express it in movement (Agawu, 2003). De Hen (1967) considers that rhythm is an alternation of tension and relaxation. The difference between Western music and African music, he writes, lies in the opposite way of counting, where Western music counts heavy-light and African music the other way around. The human annotations support these points. Figure 1 zooms in on a tap annotation where persons 1 and 2 tap the same tempo but in a different phase. Figure 2 visualizes a similar example where the binary annotations vary in phase. This specific fragment was very ambigious—13 persons tapped ternary, 10 binary—what is especially remarkable is that the group of ternary people synchronize in phase, while the binary annotations differ much more. It is clear that the ambiguity is not only between binary and ternary relations, but that there is a phase ambiguity as well. As an explorative case study, a small group was asked to write down the rhythmical percussive ostinato pattern from an audio fragment. The result shown in Figure 3 is striking by its variance. At first sight it seems so incomparable one would even question if they were listening to the same song. To summarize, it appears that people perceive different tempi, different meter, different starting points and assign different accents and durations to the percussive events.

As a final insight, we have transposed the idea of the TUBS notations (Time Unit Box System) to the human annotations (see Section 2). While TUBS is most useful for showing relationships between complex rhythms, it is here used for visualizing the annotation behavior where the place of the marker in the box indicates the exact timing of the tapped event. Hence, it visualizes the human listeners' synchronization to music. In Figure 4, a fragment of the tapped annotations is given. One sees clearly that there is quite some variance in trying to synchronize with the music, although the global tempo was unambiguous. This variance is mainly caused by the individual listeners tapping stable but in different phases than the others.

5.2 Tempo annotation by Beat Trackers

The tempo annotations of the 17 BT's are listed in Appendix B; each column containing the tempo estimates of each song.

The reference tempo for evaluating the tempo esti-

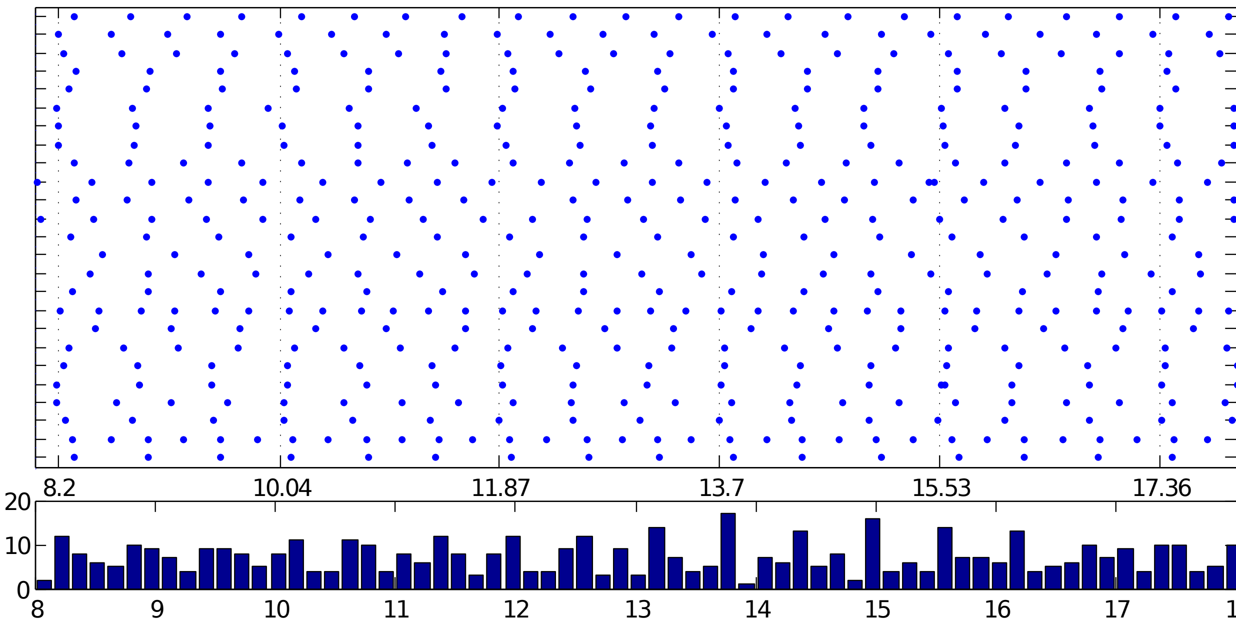


Figure 2: Fragment of track 61, where the group is divided in binary and ternary tapping. Two people follow the smallest pulse (tempo doubling). Time indications were manually added to mark the bars. The histogram shows this polymetric presence.

mations was the tempo that most people tapped (see Appendix A). As with the analysis of the human annotations, the other categories were: tempo octaves, related tempo, related tempo octaves and unrelated tempi. The category of patterns was left out, as beat tracking algorithms are designed to produce a regular pulse.

In most cases the majority of the 17 beat trackers did match the tempo assigned by humans, namely for 46 fragments (67.6%), listed in Table 4. For nine songs the tempo octave was preferred by the beat trackers, in most instances (seven), they suggested the double tempo. For the remaining 13 songs, the beat trackers preferred the related tempo above the assigned tempo, 10 times they preferred the binary pulse for the ternary pulse tapped by humans, and only two times the ternary for the binary. One instance concerned a meter of five where the tempo estimation of the BT split up the meter in 2.5. Looking at Table 3, the assigned tempo was detected by at least one BT in 64 songs (94%), and by three of the five BT still in 58 songs (85%).

Table 2 contains the distribution of the 1190 annotations which are comparable to the overall human annotations. At 48%, there is a slight decrease in identical tempo annotations, while the category of the related tempi increases up to 19%.

We can conclude that the beat trackers give a reliable result: two thirds of the tempi were analyzed identically to the human annotations. For the other songs the majority of the BT's suggested a tempo octave or a related tempo. In songs with higher ambiguity (where people assigned several tempi), it appears that the BT's tend to prefer binary meter over ternary, and higher tempi over slower. The preference for higher tempo is also reflected in the medians for each beattracker over the 70 songs, with a range of 109-141bpm, and one outlier of 191bpm, higher than the human medians mentioned in Section 5.1.

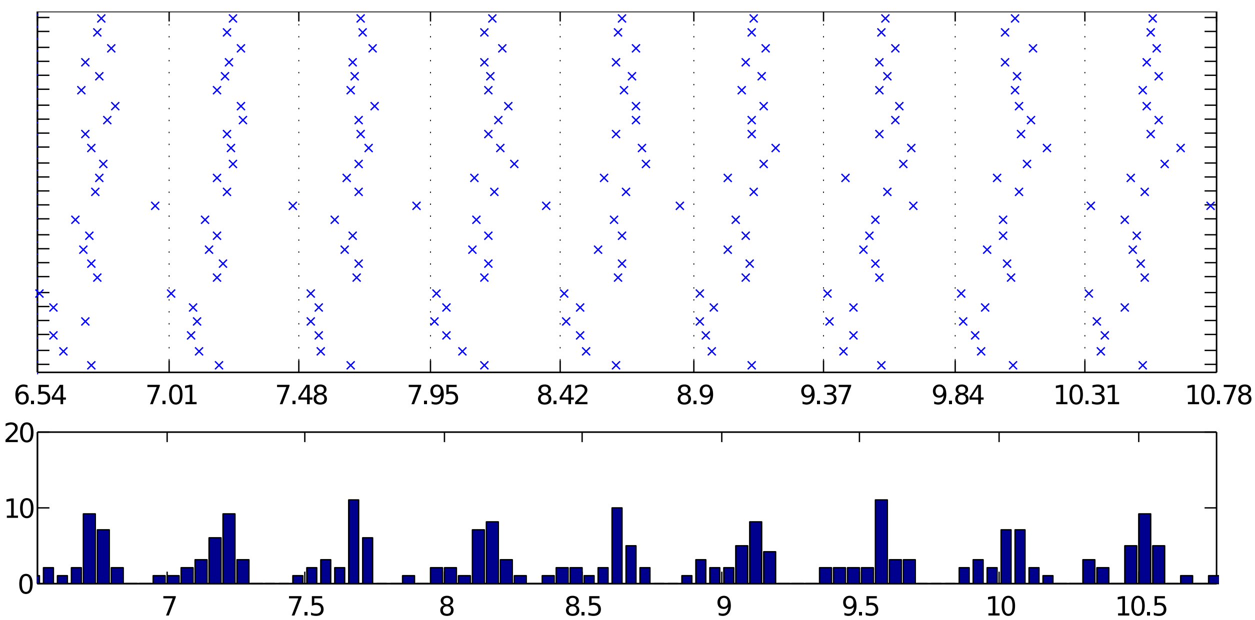
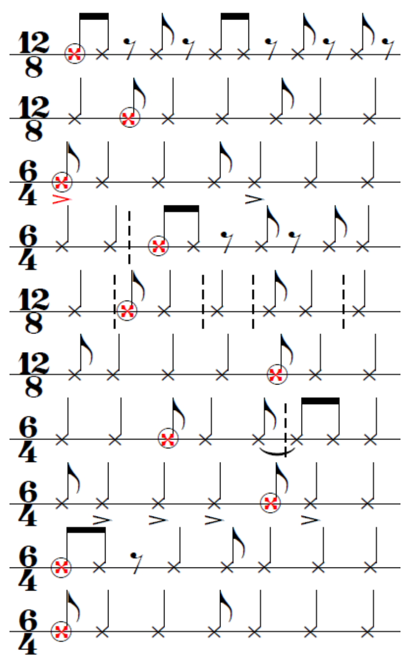


Figure 4: Fragment of track 56 where each box represents one beat, as in a TUBS representation. The unanimously assigned tempo however conceals large time differences in human onsets. The dotted lines are manually added as a reference.

5.3 Human annotations versus Beat Trackers

As a first step we determined all mutual agreements between the 5 beat trackers that are contained in our committee, using the Information Gain measure (see Section 4.3). In Figure 5 the histograms of these mutual agreements for all musical fragments in RMCA subset are shown, where the histograms are sorted by their MMA_D value. It can be observed that there is an almost linear transition from histograms with concentration at low agreement values to histograms with very high agreements on the right side of Figure 5. The vertical red line marks the threshold for perceptually satisfying beat sequences (MMA=1.5bits), which was established in listening tests (Zapata et al., 2012). Out of the 70 fragments in the dataset 57 lie on the right side of this threshold, which implies that for 81% of this data at least one of the five beat sequences can be considered as perceptually acceptable. This percentage is higher than the one reported for a dataset of Western music (73%, Zapata et al. (2012)). In the previous Section we showed that 59 songs have either correct or half/double tempo. That proportion is quite close to the 81% we measure here.

We will show the difference between songs having beat sequences with low MMA and those having a high MMA between their sequences using two examples. One example was taken from the left side of the red line in Figure 5 and the other from the right side of it. An excerpt of the beat sequences for the low-MMA_D song is shown in Figure 6. It is apparent that the beat sequences are largely unrelated, both in terms of tempo as well as in terms of phase alignment. On the other hand, in Figure 7 the song with high MMA_D has beat sequences that are more strongly related. Their phase is well aligned, however, there are octave relationships between the tempi of the beat sequences. This can also be seen from the TUBS representation, which is less randomly distributed than for the low-MMA_D song depicted in Figure 6. This



(a) Different transcriptions of the same rhythmical pattern derived from listening to a song (in case MR.1973.9.19-2A) by 10 people. The circled note indicates same place in the shifted pattern.

♪	♩	♩	♪	♩	♩	♩
4			2		2	2

(b) Number of transcriptions at different starting points in the pattern.

x	x		x		x	x		x		x	
x	x		x		x		x	x		x	
x		x		x	x		x		x		x
x		x	x		x		x	x		x	

(c) TUBS notation of the general pattern with 4 different starting points.

Figure 3: Different transcriptions the wide-spread asymetrical 12-pulses ostinato rythmical pattern / timeline.

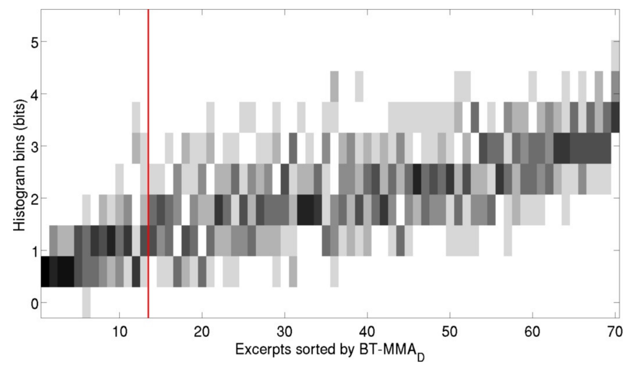


Figure 5: Each column of the image depicts a histogram obtained from $5 * 4/2$ mutual agreements of the 5 beat sequences for each song in the RMCA subset. The histograms are sorted by their mean values (BT-MMA). Dark colors indicate high histogram values. The dotted red line marks the threshold above which a perceptually satisfying beat estimation can be performed.

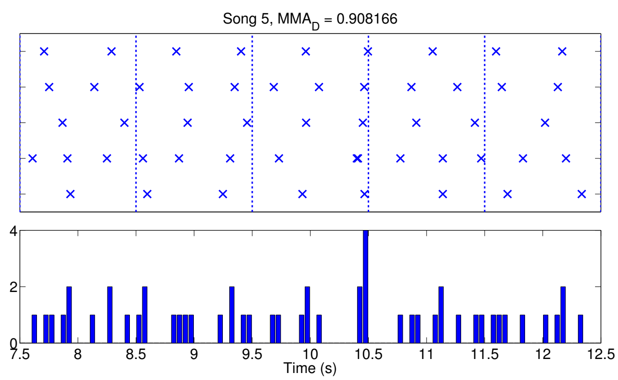


Figure 6: Beat sequences of the 5 beat trackers in the committee for a song with low MMA_D.

clarifies that by calculating the MMA$_D$ we can obtain an estimation about the agreement between beat sequences or annotations without the necessity of a time consuming manual analysis.

When directing our attention towards the human annotations, we obtain an unexpected result. In Figure 8 it can be seen that from low agreement among beat sequences follows low agreement among human annotations, which can be seen by the population of the lower-left rectangle formed by the 1.5-

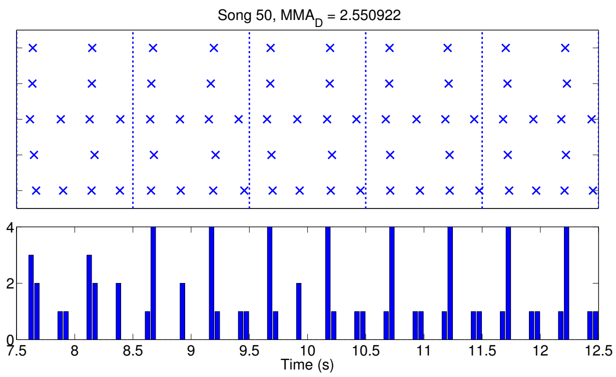


Figure 7: Beat sequences of the 5 beat trackers in the committee for a song with high MMA_D.

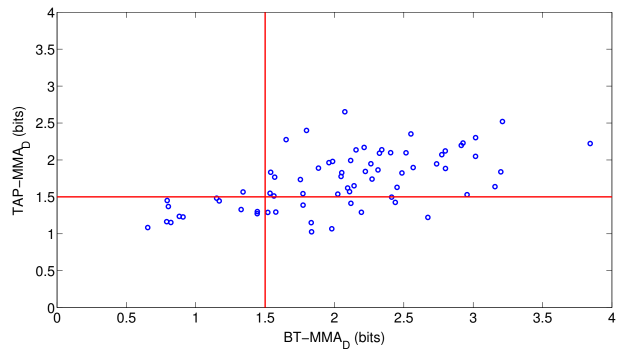


Figure 8: Scatter plot of the MMA_D values obtained from human tappings and the beat tracking algorithms. Red lines indicate the $1.5bit$-threshold.

bit threshold lines. However, high agreement among beat trackers does not imply high agreement among human tappers; a significant amount of fragments with a BT-MMA$_D$ above the threshold has quite low TAP-MMA$_D$ values (lower-right rectangle). This is quite different from the result for Western music presented in Holzapfel et al. (2012), where this quadrant was not populated at all, indicating that good beat tracker performance always implied high agreement among human tappers. Inspection of the human annotations related to the fragments in the lower-right quadrant revealed that they are indeed characterized by a large variability for each fragment. The audio for these fragments appears to have several polyrhythmic layers, almost independent polyphony, often with flute, rattle, singing, and dense percussion. Several fragments in the lower quadrants contained rattles, which have an unclear attack, resulting in poorly aligned tapped sequences.

From the 12 fragments in the lower-left quadrant only one had a binary meter while six of them were ternary. Two were in five and three were undefined. From the 11 fragments in the lower-right quadrant, the meters were equally distributed, but for this selection the average tempo stands out with 140bpm, whereas it was 102bpm for the lower-left quadrant and 109bpm for the upper quadrants. The BT tempi follow the same tendency, but less distinct. The upper quadrants had and average of 17 persons tapping the same tempo, while the lower quadrants 12. When we add the number of the retries of the human annotations, which can indicate the more difficult files since people were in doubt with their first annotation, we see a very large portion of the retries appearing in the lower-left quadrant. For the lower-right quadrant, some fragments barely had any retries while others had many. There was no relation between meter and retries, except for the meter in five which apparently needed one or more retries from most people.

As we can now determine for which fragments some meaningful relation can be found in a set of beat sequences or annotations by using MMA$_D$, we now go one step further and explore which kind of tempo relations might be encountered between these sequences, and if there are off-beat relationships. For example, in Figure 7 we saw a set of beat sequences that are well aligned in phase, but were characterized by octave relationships. To this end we will analyse the MMA$_F$, which results in characteristic values in presence of specific tempo and phase relations, as explained in Section 4.3. For the 57 fragments above the MMA$_D$ threshold in Figure 5 we depict the MA histograms obtained using the F-measure in Figure 9, sorted again by MMA$_D$. Hence, this plot represents the BT-mutual agreement histograms of the same fragments as on the right side of the red line in Figure 5, but the histograms are computed using the F-measure. The curve on the right side of the histograms depicts the sum of each bin over all 57 fragments. We can see that the largest amount of sequences agree perfectly (100%). The peak close to 66% is mainly caused by sequences that are well aligned but have tempo relations of factor two. High values in the histogram at zero help identifying sets of sequences with identical tempi, but off-beat relations. Our example shown in Figure 7 which contained octave errors finds itself in column 42 of the image in Figure 9. It has a large peak in the bin related to 66% which can be seen by the black spot in that area. Hence, by observing the shape of a histogram (*i.e.* a single column in Figure 9), we can obtain valuable insight into what relations exist between an arbitrary set of beat sequences or annotations. While tempo relations between regular sequences can easily be obtained by determining the relations between their average inter beat distances, this says nothing about the accuracy of their alignment in phase. Thus, examining the existence of peaks in the F-measure MA histograms can give a better understanding about this alignment. Furthermore, these histograms have the property that they give an even more accurate representation when the number of compared sequences is high. This is quite helpful, as for a large number of sequences manual analysis gets more and more difficult. While we showed example sequences for beat tracking algorithm outputs, such insight can also be obtained for human annotations.

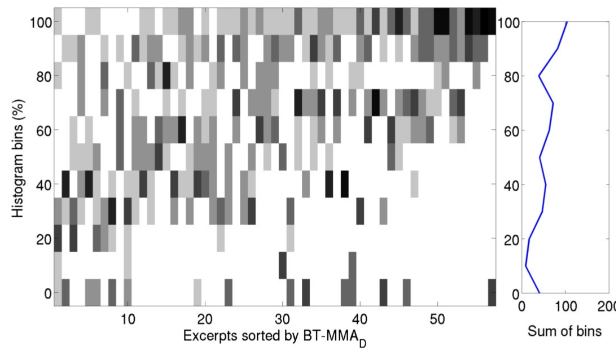


Figure 9: Each column of the image depicts a histogram obtained from $5 * 4/2$ mutual agreements of the 5 beat sequences for each song in the RMCA audio subset, measured with the F-measure. The histograms are sorted by the BT-MMA$_D$. Dark colors indicate high histogram values. On the right we see the result of summing up each bin ovber all histograms.

6 MIR and ethnic music discussion

6.1 Awareness on possible biased approaches

Most music software applications, interfaces, and underlying databases are optimised for descriptions related to Western popular music. A common practise of such music information retrieval software is to take the musical characteristics and semantic descriptions of Western music as a standard, and to develop tools that are based upon a series of Western cultural concepts and assumptions. These assumptions apply to structural aspects (e.g. tonality, assumption of octave equivalence, instrumentation), social organization of the music (e.g. composers, performers, audience) and technical aspects (e.g. record company, release date). For non-Western music however, there is no guarantee that these concepts can be easily applied (Tzanetakis et al., 2007). On the contrary, imposing Western concepts onto non-Western music can lead to incorrect or incomplete information. The predominant focus on the composer and performer illustrates this typically Western approach, whereas in non-Western music this information is often unknown or even irrelevant. In turn, non-Western music often has a very specific function, such as working song, rowing, hunting, which is a unfamiliar concept for Western music. There is, however, a need for reorienting methodologies since over the last decade several national and European projects were launched which aim at digitization of musical libraries, with at least a part of ethnic music. See Appendix C for a limited list of such as projects.

On the other hand, as can be seen in the results of this research, the beat tracking software does perform well, even without any specific fine-tuning towards the set of Central-African music. The paradigm of focusing on the smallest pulse, as some ethnomusicologists suggest, is an effective starting point which the beat trackers are capable of.

6.2 Transcription

A general concern is the indirect relation between the sounding music, its written representation, and the musical intentions of the composer/performer as described by Leman (2007). This relationship is even weaker in the context of ethnic music. Any musical performance is an intense and individual interpretation of its performers' knowledge and history. The ethnomusicologist who listens to this musicalized language faces an immense challenge if he wants to (re)produce scores starting from the audio as it is heard.

In such tasks, transcription has since long been the first step before studying an oral culture. Often a transcription relies on Western notation, sometimes specially invented symbols are added, and some others prefer to use graphical visualization of the audio. More about the complexities of transcriptions can be found in Nettl (1983), two chapters by Ter Ellingson in Myers (1993) and the chapter *Notation and Oral Tradition* by Shelemay (2008).

As a final note, we identify some polarizing issues: Namely the descriptive notation, a meticulously detailed notation that tries to capture every aspect of the audio but makes it hard to read or even understand, versus the prescriptive transcription that merely consists of the information needed by the insider (Nettl, 1983). And secondly, in the context of African music which is of very repetitive nature, one can ask if a full transcription is needed, or that it is allowed to summarize the song to its essential components by filtering out small variations (Wade, 2009).

With the aim of developing automated tools for transcription, one must be aware of all these elements. They set out rules that should not be seen as a additional difficulties, rather they should be seen as guidelines which a multidisciplinary approach of musicology, ethnomusicology and computer engineering should follow.

7 Conclusions & Future work

This paper presents the preliminary research on the development of a computational approach for analyzing temporal elements in ethnic music. For a good understanding of tempo in ethnic music, a case study with Central-African music was conducted. Both human annotations, and the output of a set of beat trackers were compared to discover insights in the tempo estimations results, in the computational potential, and in some perceptual phenomena themselves. Tempo is based on the regular and repetitive pulse of music, and will form a basis for any further analysis, annotation and transcription. The experiment showed the ambiguity in perception of tempo and meter, both for humans and for beat trackers. The beat trackers obtained comparable results with the human annotations, with a slight tendency to prefer binary pulsation in ambiguous situations and to prefer a higher tempi octave. We also found a notable ambiguity in phase indication.

Gathering multiple beat trackers entails some advantages: if their results are combined, they appear to detect temporal ambiguity in songs where humans showed a similar perception. Detecting such information is important for the user, as it is, after all, our intention to create a realistic analysis platform where the user makes the final decision on any annotation or transcription. The software only makes suggestions that can be followed, adapted or ignored. Another interesting advantage is that the combination of the several tempo estimations does tell us something about the temporal organisation behind the pulsation: combining the group of tempo estimations can give suggestions about the metrical organization of the piece.

The given hypotheses can be affirmed by this research: i) a set of BT can be used as a reliable method for tempo extraction in Central-African music with results comparable with human annotations, ii) the set of BT gives similar insights into the ambiguity of tempo perception as in human tempo perception, and iii) the set of BT does mostly detect problematic cases for tempo annotation. The fourth hypothesis seems promising namely that the combined results of the set of BT can provide information of a higher metrical level, but this has not been investigated further in a computational way.

It is the intention to add the proposed approach into the existing software package Tarsos (Six & Cornelis, 2011), which currently is focused on analysis of pitch organization in ethnic music.

8 Acknowledgments

This research was supported by the University College Ghent and by the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

We are grateful to the RMCA (Royal Museum for Central Africa) in Belgium for providing access to its unique archive of Central African music.

Finally we would like to thank José R. Zapata for his support in running beat tracking algorithms.

9 Bibliography

Agawu, K. (1995). The Invention of African Rhythm. *Journal of the American Musicological Society*, *48*.

Agawu, K. (2003). *Representing African Music.* Routledge, New York.

Arom, S. (1985). *African Polyphony and Polyrhythm: Musical Structure and Methodology.* Cambridge University Press.

Bolton, T. (1894). Rhythm. *American Journal of Psychology*, 145–238.

Brandel, R. (1961). *The Music of Central Africa.* Martinus Nijhof.

Cornelis, O., De Caluwe, R., Detré, G., Hallez, A., Leman, M., Matthé, T., et al. (2005). Digitisation of the ethnomusicological sound archive of the rmca. *IASA Journal*, *26*, 35-44.

Cornelis, O., Lesaffre, M., Moelants, D., & Leman, M. (2010, April). Access to ethnic music: Advances and perspectives in content-based music information retrieval. *Signal Processing*, *90*(4), 1008–1031.

Davies, M. E. P., Degara, N., & Plumbley, M. D. (2009). *Evaluation Methods for Musical Audio Beat Tracking Algorithms* (Tech. Rep. No. C4DM-TR-09-06). Queen Mary University of London, Centre for Digital Music.

Davies, M. E. P., Degara, N., & Plumbley, M. D. (2011). Measuring the Performance of Beat Tracking Algorithms Using a Beat Error Histogram. *IEEE Signal Processing Letters*, *18*(3), 157–160.

Degara, N., Argones, E., Pena, A., Torres, M., Davies, M. E. P., & Plumbley, M. D. (2011). Reliability-informed Beat Tracking of Musical Signals. *IEEE Transactions on Audio, Speech and Language Processing*(1), 290–301.

De Hen, F. (1967). De muziek uit afrika. *Bulletin d'information de la coopération au développement*.

Desain, P., & Windsor, L. (2000). *Rhythm, Perception and Production.* Swets & Zeitlinger.

Dixon, S. (2002). Pinpointing the Beat: Tapping to Expressive Performances. In *International Conference on Music Perception and Cognition (ICMPC 2002).*

Dixon, S. (2007, 1). Evaluation of the Audio Beat Tracking System BeatRoot. *Journal of New Music Research (JNMR)*, *36*(1), 39–50.

Ellis, D. P. W. (2007). Beat Tracking by Dynamic Programming. *Journal of New Music Research (JNMR)*, *36*, 51–60.

Epstein, D. (1995). *Shaping Time: Music, the Brain, and Performance.* New York, Schirmer.

Gouyon, F., Klapuri, A., Dixon, S., Alonso, M., Tzanetakis, G., Uhle, C., et al. (2006). An Experimental Comparison of Audio Tempo Induction Algorithms. *IEEE Transactions on Speech and Audio Processing*, *14*.

Handel, S., & Oshinsky, J. (1981). The Meter of Syncopated Auditory Polyrhythms. *Percept Psychophys*(1), 1-9.

Holzapfel, A., Davies, M. E. P., Zapata, J. R., Lobato Oliveira, J. ao, & Gouyon, F. (2012). Selective sampling for beat tracking evaluation. *IEEE Transactions on Audio, Speech and Language Processing*, *20*(9).

Jones, A. (1959). *Studies in African Music.* London, Oxford University Press.

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.

Koetting, J. (1970). Analysis and Notation of West African Drum Ensemble. *Selected Reports in Ethnomusicology*, *1*.

Kubik, G. (1960). The structure of kiganda xylophone music. *African Music*, *2*(3), 6–30.

Kubik, G. (1994). *Theory of African music.* The University of Chicago Press.

Large, E. W. (2000). On Synchronizing Movements to Music. *Human Movement Science*, *19*, 527–566.

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology.* The MIT Press. Hardcover.

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music.* Cambridge, MA.: The MIT Press.

London, J. (2011). Tactus and Tempo: Some Dissociations Between Attentional Focus, Motor Behavior, and Tempo Judgment. *Empirical Musicology Review*, *43*(1).

McKinney, M. F., & Moelants, D. (2006). Ambiguity in Tempo Perception: What Draws Listeners to Different Metrical Levels? *Music Perception*, *24*(2), 155–166.

Moelants, D. (2001). *Een model voor ritmeperceptie toegepast op de muziek van de 20ste eeuw.* Unpublished doctoral dissertation, University Ghent.

Moelants, D. (2002). Preferred tempo reconsidered. In *Proceedings of the 7th International Conference on Music Perception and Cognition (ICMPC 2002)* (p. 580-583).

Moelants, D., & McKinney, M. F. (2004). Tempo Perception and Musical Content: What Makes a Piece Fast, Slow or Temporally Ambiguous? In *Proceedings of the International Conference on Music Perception and Cognition (ICMPC 2004).*

Myers, H. (1993). *Ethnomusicology, Historical and Regional Studies.* Routledge.

Nettl, B. (1983). *The Study of Ethnomusicology, 31 Issues and Concepts.* University of Illinois Press.

Oliveira, J., Gouyon, F., Martin, L., & Reis, L. (2010). IBT: A Realtime Tempo and Beat Tracking System. In *Proceedings of the 11th International Symposium on Music Information Retrieval (ISMIR 2010)* (pp. 291–296).

Parncutt, R. (1994). A Perceptual Model of Pulse Salience and Metrical Accent in Musical Rhythms. *Music Perception*, *11*(4), pp 409–464.

Pöppel, E., Held (Eds), R., Teuber, H., & Leibowits, H. (1978). *Handbook of sensory physiology, vol. 8: Perception* [Book]. Berlin Springer Verlag.

Repp, B. (2006). Music, Motor Control and the Brain. In E. Altenmuller, M. Wiesendanger, & J. Kesselring (Eds.), (pp. 55–76). Oxford: Oxford University Press.

Seifert, U., Schneider, A., & Olk, F. (1995). On Rhythm Perception: Theoretical Issues, Empirical Findings. *Journal of New Music Research (JNMR)*, *24*(2), 164-195.

Shelemay, K. (2008). The Garland Handbook of African Music. In R. M. Stone (Ed.), (pp. 24–44). Routledge New York.

Six, J., & Cornelis, O. (2011). Tarsos - a Platform to Explore Pitch Scales in Non-Western and Western Music. In *Proceedings of the 12th International Symposium on Music Information Retrieval (ISMIR 2011).*

Tzanetakis, G., Kapur, A., Schloss, W. A., & Wright, M. (2007). Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*, *1*(2).

Van Noorden, L., & Moelants, D. (1999). Resonance in the Perception of Musical Pulse. *Journal of New Music Research*, *28*(1), 43–66.

Wade, B. C. (2009). *Thinking Musically.* Oxford University Press.

Willenze, T. (1964). *Algemene muziekleer* [Book]. AULA.

Wohlschlager, A., & Koch, R. (2000). Synchronisation Error: an Error in Time Perception. In P. Desain & W. L. Windsor (Eds.), *Rhythm Perception and Production.* (p. 115-127). Swets & Zeitlinger.

Zapata, J. R., & Gómez, E. (2011, July). Comparative Evaluation and Combination of Audio Tempo Estimation Approaches. In A. E. Society (Ed.), *42nd AES Conference on Semantic Audio.* Ilmenau, Germany: Audio Engineering Society.

Zapata, J. R., Holzapfel, A., Davies, M. E. P., Lobato Oliveira, J. ao, & Gouyon, F. (2012). Assigning a Confidence Threshold on Automatic Beat Annotation in Large Datasets. In *Proceedings of the 13th International Symposium on Music Information Retrieval (ISMIR 2012).*

A Human Annotations

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18
1	**112**	**83**	**105**	59	**119**	**129**	75	**83**	**94**	66	38	**114**	72	**81**	**120**	103	**101**	58
2	54	*27*	52	**117**	**122**	167	**150**	**81**	47	**133**	83	*38*	**144**	**80**	60	49	195	66
3	**119**	**81**	*79*	**118**	**122**	**129**	**150**	161	**95**	**133**	81	**117**	**146**	**79**	**120**	88	**101**	103
4	**112**	66	52	**117**	**120**	65	**150**	**80**	**94**	66	40	**115**	72	**80**	**120**	54	**100**	108
5	**114**	**81**	52	59	**122**	**128**	**152**	**81**	**95**	66	103	58	**146**	40	**120**	57	**100**	*49*
6	**109**	*111*	52	**117**	**121**	**129**	**150**	**83**	**94**	**133**	**79**	**115**	**145**	**80**	**120**	120	204	**150**
7	126	145	52	58	**130**	**129**	**150**	167	**96**	**133**	134	**115**	**145**	**80**	**120**	110	**101**	**148**
8	**108**	**81**	**108**	58	**123**	**128**	**150**	165	**94**	**133**	**77**	**115**	**146**	**80**	**120**	92	**101**	123
9	**109**	136	**105**	**117**	**122**	**131**	75	**83**	**95**	66	**77**	*77*	**146**	**80**	**120**	114	127	**150**
10	62	123	53	**117**	**126**	**129**	**150**	**82**	**94**	67	*99*	**114**	**146**	**80**	61	108	110	**144**
11	*74*	*56*	52	**117**	**125**	**128**	**150**	**83**	**95**	**134**	80	**115**	**146**	40	**120**	59	**102**	75
12	*77*	*109*	*157*	58	**123**	**129**	**150**	**83**	**94**	**133**	112	*76*	**146**	**80**	**120**	74	**102**	75
13	**115**	*111*	**105**	59	**120**	**129**	75	**83**	**95**	**133**	**80**	**115**	**146**	**80**	**120**	126	**102**	**150**
14	103	**82**	**103**	59	**125**	**128**	77	**83**	**95**	**133**	*97*	**114**	**144**	**80**	**120**	96	**101**	110
15	**114**	*111*	*157*	**117**	**125**	**129**	**152**	**83**	**94**	**133**	*98*	**117**	**146**	**80**	**120**	108	207	**148**
16	**111**	**82**	**103**	58	**117**	**131**	**150**	**82**	**95**	**132**	**80**	**115**	72	**80**	**120**	88	**101**	*99*
17	**117**	*113*	**103**	**115**	**125**	**128**	**152**	167	**95**	**133**	**78**	**116**	**146**	161	**120**	154	**102**	108
18	**111**	56	117	59	**122**	**131**	75	167	**95**	66	94	**116**	**144**	**80**	**120**	111	215	74
19	56	**83**	*157*	58	**122**	**132**	**154**	**83**	**94**	**131**	**79**	**114**	**146**	159	**120**	99	**99**	114
20	124	133	*157*	**119**	**124**	**128**	**150**	164	**94**	**133**	154	**116**	**147**	159	**120**	130	**101**	**146**
21	95	**83**	*157*	59	**123**	**129**	75	**83**	**94**	66	*99*	**116**	**146**	**80**	**122**	122	**104**	**144**
22	220	134	*153*	**116**	**126**	**129**	**150**	**83**	**94**	**133**	92	**114**	**146**	**80**	**120**	59	**101**	207
23	**118**	**83**	**105**	**123**	**123**	**128**	**154**	**82**	**95**	**134**	*98*	**116**	**146**	**81**	**122**	68	**100**	129
24	**117**	**81**	*157*	59	**123**	**128**	**150**	164	**95**	**134**	154	**115**	**146**	**81**	**120**	97	**103**	107
25	**120**	*111*	*157*	**119**	**122**	**128**	**150**	162	**94**	**133**	**80**	**116**	**143**	162	**125**	92	**101**	92
tempo	**114**	**82**	**105**	**117**	**123**	**129**	**150**	**83**	**95**	**133**	**79**	**115**	**146**	**80**	**120**	**?**	**101**	**148**
identical	**15**	**10**	**8**	**13**	**25**	**23**	**19**	**17**	**24**	**18**	**8**	**21**	**22**	**19**	**23**		**19**	**8**
octave	**3**		**7**	**12**		**1**	**6**	**8**	**1**	**7**	**4**	**1**	**3**	**6**	**2**		**2**	**3**
related	**2**	**6**	**8**								**5**	**2**						**1**
rel oct		**3**	**1**									**1**						**1**
unrelated	**5**	**6**									**2**					**25**	**2**	**12**
pattern			**1**			**1**					**6**						**2**	
meter	1	3	2	4	2	4	6	4	4	3	5	3	4	4	4	u	4	4
$2^{nd}meter$		4	3									2						3

	T19	T20	T21	T22	T23	T24	T25	T26	T27	T28	T29	T30	T31	T32	T33	T34	T35
1	**65**	**111**	104	75	69	77	83	61	65	93	*38*	*95*	72	*72*	67	**98**	**128**
2	85	**109**	49	117	71	*51*	83	0	258	55	**114**	**63**	105	252	*92*	102	63
3	133	**109**	112	*117*	**140**	*103*	158	60	**128**	94	105	*185*	**144**	*140*	*90*	**98**	**126**
4	**67**	**110**	105	74	70	**152**	**162**	**122**	199	92	**115**	**62**	**145**	*140*	*45*	50	**126**
5	**65**	**109**	54	**150**	71	*51*	**167**	**120**	**129**	57	57	**63**	**144**	**96**	*45*	50	63
6	128	220	117	**150**	**140**	**153**	**167**	**120**	**128**	67	**115**	**63**	**144**	0	*46*	*125*	**126**
7	133	**109**	128	**150**	34	*100*	333	60	**129**	92	**117**	175	*94*	**93**	*46*	*125*	252
8	**66**	**109**	88	76	70	**150**	90	75	0	*77*	**112**	**63**	*95*	**94**	67	**99**	**128**
9	**65**	73	110	*120*	70	*51*	83	**121**	**129**	**112**	**114**	**63**	*47*	**94**	*45*	**100**	63
10	**65**	157	135	*60*	**142**	**153**	83	**120**	**128**	55	105	142	**145**	*141*	*45*	**102**	**126**
11	43	**109**	63	*59*	79	*51*	**165**	64	**131**	55	*38*	**63**	75	46	*45*	109	**125**
12	130	**111**	57	**150**	71	*101*	83	**121**	*80*	**114**	225	**64**	*96*	**94**	**134**	150	**128**
13	**65**	**109**	91	**150**	**144**	**154**	**167**	60	*87*	**109**	**114**	*95*	*96*	**94**	**136**	**99**	**128**
14	**66**	**109**	93	*60*	**144**	**150**	**167**	**120**	**128**	105	**111**	**62**	**144**	*70*	*46*	**97**	**126**
15	133	**111**	122	**150**	**143**	**152**	**167**	**122**	*85*	**111**	**116**	128	**144**	*142*	**136**	*125*	**128**
16	**66**	**110**	43	74	**140**	77	83	60	**131**	*77*	**112**	**63**	72	120	98	**100**	63
17	130	**109**	123	**150**	**140**	*102*	**167**	**120**	**129**	74	**115**	**64**	**144**	**95**	**136**	**100**	**126**
18	84	59	91	140	71	*52*	**167**	**121**	*87*	69	**115**	**63**	*98*	**94**	86	102	**126**
19	**65**	**109**	61	122	120	**146**	**167**	62	**131**	**112**	**110**	*190*	**144**	96	95	**98**	**127**
20	0	**109**	73	**150**	**144**	**157**	**166**	**120**	**130**	**114**	**116**	**64**	**144**	*142*	**134**	**101**	**126**
21	**65**	**109**	135	74	71	**148**	85	**120**	**128**	128	**114**	**63**	72	*71*	**136**	**99**	**126**
22	133	**109**	99	**149**	70	*102*	**169**	63	*86*	84	*77*	**63**	**144**	**94**	**136**	191	**125**
23	131	**109**	112	**146**	**140**	*101*	148	**125**	**131**	120	**114**	*96*	**144**	*140*	**136**	*126*	**126**
24	134	**109**	101	*283*	**144**	139	**164**	**121**	**129**	*81*	**117**	*97*	**142**	**94**	140	**98**	**128**
25	131	**108**	102	**152**	**140**	**150**	**165**	61	**129**	87	**112**	*195*	**144**	*70*	**136**	106	287
tempo	**65**	**109**	**?**	**150**	**142**	**150**	**167**	**121**	**129**	**112**	**114**	**63**	**144**	**94**	**136**	**99**	**126**
identical	**11**	**21**		**11**	**12**	**11**	**14**	**14**	**16**	**6**	**18**	**15**	**14**	**10**	**9**	**13**	**19**
octave	**10**	**2**		**5**	**11**	**2**	**8**	**9**	**2**	**4**	**2**	**1**	**3**	**1**	**2**	**4**	**5**
related	**1**	**1**		**3**		**6**		**1**	**5**	**3**	**2**	**4**	**5**	**6**	**8**	**4**	
rel oct		**1**		**3**		**5**					**1**	**3**	**1**	**4**	**2**		
unrelated			**25**		**1**	**1**	**3**		**1**	**12**	**2**	**2**			**1**	**2**	**1**
pattern	**2**			**3**	**1**								**2**	**3**	**3**	**2**	
meter	3	2	u	5	4	3	u	3	2	6	3	4	2	6	3	2	2
$2^{nd}meter$		3				2			3		2	3	3	4			

Table 5: Human tempo annotations on a set of seventy sound fragments.

	T36	T37	T38	T39	T40	T41	T42	T43	T44	T45	T46	T47	T48	T49	T50	T51	T52
1	70	75	**105**	72	**93**	70	**82**	**91**	63	78	**122**	*73*	**94**	*44*	59	57	**81**
2	*92*	35	*69*	71	*60*	115	**81**	46	63	80	103	*74*	**94**	88	**117**	**114**	41
3	**136**	**139**	96	72	**101**	**142**	162	**92**	**128**	**157**	**122**	**108**	**94**	131	**117**	**115**	164
4	**136**	*52*	**104**	72	49	69	**81**	**90**	**128**	70	**122**	54	**94**	37	59	57	**81**
5	**140**	**140**	*70*	**144**	**94**	70	**81**	45	**128**	**159**	*81*	**107**	**95**	28	**117**	58	**83**
6	**138**	**140**	*140*	**146**	*32*	**140**	163	183	**127**	**161**	**120**	*150*	*142*	102	**118**	**115**	**81**
7	**138**	**145**	215	**144**	115	**140**	164	181	149	154	**119**	167	**97**	**111**	**117**	**117**	*99*
8	*94*	**140**	*71*	73	*62*	78	**81**	**88**	**126**	79	**120**	**108**	**92**	*44*	**117**	**114**	**81**
9	**140**	**140**	*70*	48	*62*	70	**82**	**91**	**129**	80	**120**	*162*	*142*	*44*	**117**	57	55
10	**136**	**140**	*123*	73	*65*	150	164	**91**	**128**	100	**120**	*152*	**94**	127	**117**	**115**	**82**
11	69	34	**104**	72	47	*46*	**81**	**89**	63	**162**	**121**	**108**	47	88	**117**	58	**80**
12	*92*	**138**	*70*	**145**	**95**	70	**81**	**91**	**126**	150	**122**	**109**	**96**	**117**	**117**	**114**	**81**
13	**138**	**140**	**104**	**146**	*63*	109	162	**92**	**125**	**162**	**122**	**109**	**96**	*88*	**119**	**115**	**82**
14	**138**	69	**104**	73	91	*49*	**82**	**90**	**126**	*109*	**120**	**109**	116	**107**	**119**	**114**	*64*
15	**139**	**138**	**104**	**146**	**94**	**141**	**81**	**90**	**128**	148	**120**	**107**	*144*	**111**	**117**	**115**	167
16	**140**	69	*69*	72	*63*	73	**82**	**91**	61	*104*	**120**	**109**	**92**	111	59	108	**82**
17	**136**	**136**	*140*	**146**	*128*	**136**	164	**91**	**126**	**159**	**120**	**108**	**94**	**109**	**117**	**115**	**81**
18	**136**	**142**	**103**	**146**	*62*	112	**82**	*60*	**125**	**161**	**120**	215	**94**	91	**117**	57	**82**
19	**138**	**139**	*139*	**146**	*124*	**141**	162	181	**128**	149	**120**	**108**	*140*	215	**120**	**115**	**82**
20	**137**	**140**	**105**	**146**	**100**	**140**	164	181	**128**	*211*	**120**	*162*	**95**	178	**117**	**117**	164
21	**136**	69	**104**	**144**	54	133	**81**	**91**	65	**162**	**120**	**109**	**95**	123	**119**	**115**	**82**
22	**136**	*100*	**104**	**144**	*61*	117	162	**91**	**125**	**162**	*80*	*162*	**92**	97	**117**	**116**	**81**
23	**138**	**140**	**103**	**145**	**101**	**139**	**82**	**91**	**126**	*107*	**120**	**108**	**94**	**110**	**119**	**117**	**82**
24	**136**	**140**	**103**	**145**	**99**	153	162	**90**	**128**	*106*	**120**	**109**	**93**	**118**	**118**	172	167
25	**139**	**140**	**104**	**142**	**101**	*98*	**81**	**90**	140	95	138	**109**	*140*	**117**	**120**	**116**	162
tempo	**138**	**140**	**104**	**145**	**99**	**140**	**81**	**91**	**127**	**161**	**120**	**108**	**94**	**111**	**117**	**114**	**82**
identical	**20**	**17**	**13**	**15**	**9**	**8**	**15**	**18**	**18**	**9**	**21**	**15**	**18**	**8**	**22**	**17**	**16**
octave	**2**	**6**	**1**	**9**	**2**	**6**	**10**	**6**	**5**	**4**		**2**	**1**	**2**	**3**	**6**	**6**
related	**3**	**1**	**6**	**1**	**8**	**2**		**1**		**4**	**2**	**5**	**5**	**3**			**1**
rel oct		**1**	**4**		**3**	**1**				**1**		**2**		**1**		**1**	**1**
unrelated			**1**		**3**	**5**			**2**	**7**	**1**	**1**		**5**		**1**	**1**
pattern						**3**					**1**		**1**	**6**			
meter	3	4	3	6	3	5	2	2	2	3	5	2	6	5	4	2	5
$2^{nd}meter$	4		4		2					4	4	3	2				

	T53	T54	T55	T56	T57	T58	T59	T60	T61	T62	T63	T64	T65	T66	T67	T68	T69	T70
1	**91**	59	65	**125**	**58**	80	**57**	62	**98**	69	**59**	90	**93**	**109**	86	39	*70*	**41**
2	**92**	59	65	**125**	**58**	81	**59**	**122**	**96**	118	**59**	91	**93**	54	**154**	**79**	**92**	**41**
3	**92**	**117**	**128**	**128**	**58**	**162**	**57**	**121**	*133*	**131**	*88*	90	**92**	**108**	**159**	**79**	**92**	**41**
4	**91**	**119**	**128**	**128**	**59**	**159**	*74*	**122**	**98**	67	**59**	**180**	**92**	**109**	**150**	*60*	*140*	**41**
5	181	59	**131**	**125**	**58**	**163**	**58**	**123**	*131*	**133**	**61**	**180**	**92**	**109**	*103*	**79**	*140*	**41**
6	181	**119**	**128**	**128**	**58**	**164**	*22*	**122**	199	**136**	**60**	**178**	186	220	**154**	*117*	*140*	**41**
7	185	128	**129**	**131**	116	**162**	*78*	**125**	*129*	**133**	**59**	**181**	136	**111**	*211*	**79**	**94**	**42**
8	**92**	*91*	65	**128**	**58**	**162**	**62**	**122**	**98**	**128**	102	**181**	**92**	**109**	**154**	**79**	120	**42**
9	**92**	60	**128**	**126**	**57**	**164**	*19*	**122**	*131*	**136**	**59**	90	**92**	**109**	**154**	39	*138*	**42**
10	**95**	**117**	**131**	**126**	114	**162**	**61**	61	**99**	**136**	**59**	**181**	**94**	**109**	34	240	**92**	**43**
11	181	59	**128**	**128**	**58**	**162**	**61**	**122**	*132*	40	**59**	136	115	**110**	77	*117*	*140*	**43**
12	188	**117**	**128**	**125**	**58**	82	**62**	**122**	**98**	**138**	177	91	188	**108**	*207*	*117*	**92**	82
13	185	**118**	**128**	**125**	**58**	**164**	**58**	**123**	195	**136**	118	**181**	**92**	**108**	*103*	**79**	**93**	82
14	162	**117**	**130**	**125**	**57**	154	45	**122**	**98**	*94*	**59**	**178**	**91**	**107**	**154**	**78**	*140*	82
15	181	**117**	**129**	**128**	115	**164**	114	**123**	*131*	**134**	117	**181**	125	**108**	**152**	**79**	**92**	83
16	**91**	**122**	**128**	**125**	**59**	150	**57**	**122**	**98**	**136**	*89*	**178**	80	**110**	76	*59*	**92**	86
17	**93**	**117**	**128**	**126**	115	**162**	114	**122**	*133*	*95*	175	91	188	**109**	*104*	**79**	**92**	124
18	**93**	**119**	**128**	**126**	114	**163**	123	**125**	*129*	69	**59**	91	185	70	203	**78**	*138*	125
19	*136*	**117**	**128**	**125**	112	**164**	128	**125**	**100**	**134**	180	**181**	186	**109**	116	*117*	*140*	125
20	185	**122**	**131**	**125**	116	**162**	133	**123**	**98**	**136**	177	**181**	185	215	**157**	**77**	185	125
21	**92**	**120**	**128**	**127**	**57**	**159**	117	61	**98**	**129**	*88*	91	**93**	**108**	*103*	*117*	*71*	126
22	**92**	125	**131**	**125**	**58**	**167**	*78*	**122**	**99**	97	178	**181**	188	**109**	191	**77**	**93**	126
23	*136*	**112**	**131**	**125**	116	151	*81*	**125**	*131*	**140**	*88*	91	**92**	**107**	138	*119*	*140*	126
24	175	**117**	**129**	**125**	**59**	**162**	129	**124**	*131*	117	178	**181**	**91**	**109**	*101*	*117*	*140*	136
25	193	**114**	**131**	**129**	**56**	101	128	**125**	**102**	**136**	**59**	89	**91**	**105**	92	**80**	**94**	141
tempo	**92**	**117**	**129**	**126**	**58**	**162**	**59**	**123**	**98**	**136**	**59**	**180**	**92**	**109**	**154**	**79**	**92**	**42**
identical	**12**	**17**	**22**	**25**	**17**	**18**	**10**	**22**	**13**	**16**	**12**	**14**	**14**	**21**	**9**	**13**	**11**	**11**
octave	**8**	**5**	**3**		**8**	**3**	**8**	**3**	**2**	**3**	**8**	**10**	**6**	**3**	**3**	**3**	**1**	**11**
related	**2**	**1**					**4**		**10**	**2**	**4**	**1**		**1**	**5**	**7**	**10**	
rel oct							**2**								**2**	**2**	**2**	
unrelated	**3**	**2**				**4**	**1**			**2**	**1**		**4**		**5**		**1**	**2**
pattern										**2**			**1**		**1**			**1**
meter	2	4	4	4	4	1	3	1	3	6	6	2	6	6	4	6	4	2
$2^{nd}meter$	3	3									2				3	3	6	

Table 6: Human tempo annotations on a set of seventy sound fragments. Continued.

B Beat Tracker Annotations

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	T21	T22	T23	T24	T25	T26	T27	T28	T29	T30	T31	T32	T33	T34	T35
IBT*	215	**82**	52	**116**	*98*	258	75	167	**96**	66	*99*	**115**	73	161	**120**	235	207	75	**65**	215	152	**152**	52	102	**167**	*81*	258	*167*	92	126	**144**	**94**	**136**	**99**	63
Audiosculpt	**118**	164	*156*	**117**	102	**129**	**149**	164	**95**	**132**	**79**	**115**	**145**	161	**120**	80	**101**	76	**65**	**109**	139	*120*	**141**	102	83	**121**	**129**	*167*	**114**	126	**143**	**93**	**136**	**100**	**126**
beatCounter106	68	202	*78*	89	*185*	64	75	165	**96**	*171*	126	**115**	*108*	160	240	51	**101**	99	54	218	89	*90*	*196*	*51*	83	**121**	199	56	230	*95*	**143**	**93**	68	**99**	**126**
Böck	230	176	*157*	**115**	**125**	240	**150**	166	**95**	**133**	*98*	230	72	162	**120**	146	200	222	*86*	**109**	230	**153**	*92*	101	176	**122**	**130**	133	230	127	*95*	**93**	**136**	**100**	**125**
Stark-Davies	**117**	**81**	*156*	**119**	104	**129**	**150**	**83**	**95**	**132**	*106*	*76*	**144**	*102*	**120**	101	**101**	**141**	*89*	**110**	94	*60*	132	101	**166**	*78*	**129**	**113**	136	125	*95*	**93**	*90*	**97**	**126**
Klapuri*	**112**	**81**	*157*	**117**	**126**	**129**	**150**	**83**	**95**	**133**	*99*	**115**	**146**	**80**	**120**	109	**101**	**148**	*87*	**109**	115	*120*	*94*	102	83	**122**	**129**	**112**	**115**	126	*95*	**93**	*91*	**100**	**126**
ILSP	*75*	**80**	*78*	**117**	**124**	**129**	75	165	**95**	**132**	*99*	**115**	**145**	160	**120**	143	203	213	129	217	197	*60*	**140**	101	176	**120**	**129**	144	230	127	**143**	**94**	**136**	**99**	**126**
aufTAKT	**112**	**80**	*78*	**118**	**124**	**130**	75	**83**	**94**	66	*98*	**116**	*110*	160	90	70	88	136	*88*	**109**	165	*90*	**141**	**153**	83	**121**	**128**	**112**	**115**	128	**142**	**94**	58	**102**	**127**
Ellis	**116**	*111*	*155*	**118**	**125**	**128**	**148**	*111*	**96**	**131**	*100*	**115**	**144**	*108*	**120**	93	135	**140**	128	**109**	129	**153**	**147**	100	**161**	**124**	**128**	136	*153*	128	*96*	**94**	**137**	**100**	**126**
MixMeister	**112**	134	*156*	93	*95*	**128**	**149**	**82**	**94**	**135**	*100*	123	**144**	160	**120**	**128**	97	117	130	**108**	145	**150**	**141**	99	95	**120**	**128**	104	96	125	*95*	**94**	*91*	*152*	**126**
Aubio	129	*108*	*161*	128	**126**	**129**	**150**	**83**	**95**	**127**	131	**115**	**145**	115	**120**	**128**	**101**	**141**	*88*	**109**	88	*60*	127	102	**169**	**121**	**128**	128	*154*	127	**143**	**94**	**135**	130	**125**
BtRule	86	167	*157*	**116**	**121**	*86*	**150**	165	**95**	**133**	*197*	*153*	**144**	160	178	**122**	**101**	128	130	**109**	111	**150**	105	**152**	66	**120**	*86*	*165*	**114**	*94*	**143**	*140*	**136**	**100**	**126**
Degara*	133	**81**	*157*	**115**	103	**129**	**152**	**82**	**94**	**133**	**78**	**115**	**144**	*108*	**120**	**129**	133	**148**	*86*	**108**	101	*60*	**140**	101	**167**	**120**	**129**	**112**	*152*	126	*96*	**94**	*91*	**99**	**126**
Ellis*	231	188	*156*	**117**	**123**	259	**150**	326	**95**	**133**	*99*	231	**146**	319	**119**	92	203	**142**	259	217	300	**153**	*91*	**152**	**167**	242	259	366	*153*	127	283	**94**	**136**	246	250
Beatroot*	126	183	*154*	185	171	**129**	**150**	164	**94**	**133**	*194*	**113**	**146**	160	**120**	150	80	**150**	130	**109**	182	**150**	**141**	103	**167**	**121**	**130**	**113**	*153*	128	**146**	**94**	**136**	**100**	**125**
MIRtoolbox	**115**	161	*156*	186	63	**128**	**149**	*110*	*124*	**128**	*95*	**116**	**144**	161	**120**	55	58	89	*88*	**109**	148	*60*	56	100	**168**	**122**	*86*	95	*153*	126	*96*	**94**	**137**	**100**	**126**
Sonic Annotator	66	**83**	*52*	**117**	107	**132**	**152**	**83**	**96**	**136**	*105*	**117**	**148**	**81**	**123**	**129**	58	76	*88*	**110**	103	*60*	**144**	103	**167**	**123**	**132**	89	*157*	129	*97*	**96**	**136**	**99**	**129**
Tempo (human)	**114**	**82**	**105**	**117**	**123**	**129**	**150**	**83**	**95**	**133**	**79**	**115**	**146**	**80**	**120**	**?**	**101**	**148**	**65**	**109**	**?**	**150**	**142**	**150**	**167**	**121**	**129**	**112**	**114**	**63**	**144**	**94**	**136**	**99**	**126**
BT = Human	1	1	0	1	1	1	1	0	1	1	0	1	1	0	1	0	0	1	0	1	0	1	1	0	1	1	1	0	0	1	1	1	1	1	1
identical	7	7	0	12	8	12	13	7	16	14	2	13	13	2	14		6	7	2	13		7	8	3	9	14	12	5	4	0	8	16	11	14	15
octave	2	3	2		1	3	4	8		2		2	2	11	1		4	3	6	4				1	4	1	2	2	3	15	1		1		2
related	1	2	12	2	7	1		2	1	1	11	1	2	3	1			2	8			6	3	13		2	2	3	7	2	8	1	4	1	
rel oct			3	2	1						2	1			1							4	1												
unrelated	7	5		1		1					2			1			7	5	1				5		4		1	7	3				1	2	

Table 7: Tempo from a set of beat tracker and tempo estimation tools, * by Zapata.

	T36	T37	T38	T39	T40	T41	T42	T43	T44	T45	T46	T47	T48	T49	T50	T51	T52
IBT*	*91*	*103*	*69*	72	207	**140**	**81**	178	258	**161**	246	**108**	93	**108**	235	57	*162*
Audiosculpt	**138**	*207*	*138*	**144**	**92**	**141**	164	**91**	102	**159**	**121**	**108**	94	**109**	**119**	**114**	**82**
beatCounter106	*92*	**138**	*69*	72	109	*58*	**81**	**91**	**127**	**160**	**122**	55	93	170	**119**	*86*	**82**
Böck	*92*	68	*69*	**146**	240	92	162	181	**127**	**162**	**122**	**109**	93	73	240	*157*	*162*
Stark-Davies	*92*	**138**	*139*	**144**	**99**	*114*	**81**	**90**	116	*105*	**117**	**107**	93	95	**119**	**113**	*163*
Klapuri*	*91*	**138**	*69*	**144**	**93**	**140**	**81**	**91**	**126**	*107*	**120**	**109**	94	**109**	**117**	**114**	**81**
ILSP	*91*	*207*	*69*	**143**	**92**	**142**	**81**	**91**	*84*	**160**	**120**	**108**	94	**110**	**118**	**114**	**81**
aufTAKT	*91*	69	**103**	*94*	*138*	*83*	**82**	**90**	**121**	**159**	**118**	**106**	95	**109**	**117**	**114**	*163*
Ellis	**132**	**138**	**101**	**144**	**94**	**139**	158	*123*	**127**	*107*	**122**	**109**	94	**110**	**119**	**114**	*163*
MixMeister	*93*	*103*	**103**	**143**	**94**	*112*	162	133	**125**	**159**	**121**	**104**	93	**108**	**118**	**114**	*154*
Aubio	*92*	69	*35*	**144**	**96**	**144**	41	181	**128**	**160**	*85*	**108**	93	128	**119**	57	*162*
BtRule	**138**	*207*	**104**	72	**96**	**141**	**81**	*120*	*167*	**160**	**120**	**107**	94	**110**	**118**	**114**	**82**
Degara*	*92*	**138**	*69*	**144**	103	*112*	161	185	*167*	*108*	**120**	**108**	94	*86*	**117**	**115**	*161*
Ellis*	*183*	**138**	238	**144**	186	**142**	163	*366*	254	319	242	**109**	94	**109**	234	*155*	*163*
Beatroot*	*91*	**136**	*141*	**146**	188	**143**	164	182	**128**	**160**	**120**	**109**	94	154	**118**	**114**	*162*
MIRtoolbox	*189*	**139**	*69*	**147**	183	94	161	*123*	**126**	**160**	*86*	**107**	84	119	**118**	*156*	*163*
Sonic Annotator	*93*	**140**	*70*	**148**	**96**	*56*	167	**92**	**129**	**162**	*85*	**108**	94	**112**	**120**	**115**	*167*
Tempo (human)	**138**	**140**	**104**	**145**	**99**	**140**	**81**	**91**	**127**	**161**	**120**	**108**	**94**	**111**	**117**	**114**	**82**
BT = Human	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0
identical	3	9	4	13	9	9	7	7	10	12	12	16	17	11	14	11	5
octave		3		3	3		10	5	2	1	2	1			3	2	12
related	12	3	8	1	1	3		3	2	4	3			1		4	
rel oct	2	2	4			3		1	1								
unrelated			1		4	2		1	2					5			

	T53	T54	T55	T56	T57	T58	T59	T60	T61	T62	T63	T64	T65	T66	T67	T68	T69	T70
IBT*	**92**	**120**	**129**	**126**	**57**	**167**	*161*	246	*133*	258	117	**185**	184	**108**	162	*117*	184	82
Audiosculpt	184	**118**	**129**	**126**	114	**164**	*77*	**122**	*131*	**130**	175	90	*124*	**108**	**154**	**79**	**93**	83
beatCounter106	**92**	104	**127**	**126**	*85*	**163**	*77*	**122**	196	*92*	*88*	89	*119*	**109**	**155**	**78**	184	83
Böck	181	162	**130**	166	*187*	**162**	*150*	**122**	**98**	*90*	171	**181**	187	222	*101*	**77**	**92**	84
Stark-Davies	**91**	**115**	**128**	**125**	115	82	*77*	**124**	**99**	*93*	117	90	*125*	**107**	*104*	**79**	**94**	84
Klapuri*	**92**	**117**	**129**	**128**	115	107	*80*	**122**	**102**	*91*	117	91	*123*	**108**	*103*	*117*	**92**	85
ILSP	**92**	**117**	**130**	**127**	114	**164**	*161*	**122**	**98**	*91*	118	90	*124*	217	*103*	**78**	**93**	85
aufTAKT	**93**	144	**135**	154	115	82	*77*	**122**	**98**	*94*	145	91	**92**	**108**	**155**	*117*	**93**	85
Ellis	111	**115**	**128**	**125**	118	**163**	*155*	**124**	*130*	**129**	120	*134*	*125*	*142*	*102*	*120*	**91**	125
MixMeister	**91**	**117**	**127**	**125**	95	**162**	97	**123**	**100**	*95*	131	90	*124*	91	123	*127*	**94**	127
Aubio	**91**	162	**127**	**123**	117	*126*	129	**122**	*131*	*96*	**59**	96	*124*	*144*	*103*	**80**	**93**	129
BtRule	138	**119**	**127**	**125**	115	**163**	*78*	**122**	196	**133**	176	91	186	*144*	**154**	*117*	**93**	162
Degara*	**92**	**117**	**129**	**126**	115	**161**	144	**123**	*129*	*90*	117	**178**	*123*	**108**	*103*	*117*	**92**	167
Ellis*	273	221	**129**	169	224	**165**	*155*	246	*133*	**144**	341	**181**	*124*	217	*207*	234	185	167
Beatroot*	182	**119**	**130**	**126**	114	**162**	180	**122**	*130*	*91*	176	**182**	187	*141*	**154**	*118*	187	169
MIRtoolbox	183	**120**	**128**	167	*84*	**164**	*152*	**123**	**99**	*94*	72	**179**	*124*	**107**	*103*	**78**	*133*	170
Sonic Annotator	**92**	**120**	**129**	**126**	129	**167**	147	**123**	*133*	*96*	117	91	*126*	**110**	*103*	*120*	**94**	172
Tempo (human)	**92**	**117**	**129**	**126**	**58**	**162**	**59**	**123**	**98**	**136**	**59**	**180**	**92**	**109**	**154**	**79**	**92**	**42**
BT = Human	1	1	1	1	0	1	0	1	0	0	0	0	0	1	0	0	1	0
identical	10	12	17	13	1	12	0	15	7	4	1	6	1	9	5	7	12	0
octave	5	1			11	3		2	2	1	12	10	4	3		1	4	17
related				3	2		6		8	12	1	1	12	4	9	9	1	
rel oct					1	1	6								1			
unrelated	2	4		1	2	1	5				3			1	2			

Table 8: Tempo from a set of beat tracker and tempo estimation tools, continued, * by Zapata

C List of Digitization Efforts

Acronym	Name	Coordinator	URL
DEKKMMA	Digitization of the Ethnomusicological Sound Archive of the Royal Museum for Central Africa	IPEM, Gent University	http://music.africamuseum.be
DELOS	Network of Excellence on Digital Libraries	Association	http://www.delos.info
DISMARC	Discovering Music Archives Across Europe	Consortium	http://www.dismarc.org
EASAIER	Enabling access to sound archives integration enrichment retrieval	Queen Mary University, London	http://www.elec.qmul.ac.uk/easaier
EthnoArc	Linked European Ethnomusicological Archives	Wissenschaftskolleg zu Berlin	http://www.ethnoarc.org
Kopal	Co-operative development of a long-term digital information archive	Niedersaechsische Staats- und Universitaetsbibliothek Goettingen	http://kopal.langzeitarchivierung.de
Nestor	Network of Expertise in Long-Term Storage of Digital Resources	Deutsche Nationalbibliothek, Frankfurt am Main	http://www.langzeitarchivierung.de
MICHAEL	Multilingual Inventory of Cultural Heritage in Europe	EU countries	http://www.michael-culture.org
POFADEAM	Preservation and On-line Fruition of the Audio Documents from the European Archives of etnic Music	University of Udine	http://www.ipem.ugent.be/2005POFADEAM
PrestoSpace	Preservation towards storage and access. Standardized Practices for Audiovisual Contents in Europe.	Institut National de l'audiovisuel, Paris	http://www.prestospace.org
TAPE	Training for Audiovisual Preservation in Europe	European Commission on Preservation and Access (ECPA)	http://www.tape-online.net

Table 9: Digitization efforts of music collections with, at least some, ethnic music.