

# Interpersonal entrainment in Indian instrumental music performance: Synchronization and movement coordination relate to tempo, dynamics, metrical and cadential structure

## Introduction

Group music-making showcases the remarkable capacity for precision and creativity that is possible in the coordination of rhythmic behaviour between individuals. This phenomenon entails interpersonal entrainment, a process whereby individuals interact with each other in a manner supporting both the synchronization of musical sounds and higher-level musical coordination. In this paper we explore the nature of interpersonal entrainment in music performance, in particular the relationship of both note-to-note synchronization and higher-level coordination (e.g. phrases, transition points) to musical factors, such as metrical and cadential structures, whilst contextualizing these quantifiable aspects of interpersonal entrainment in relation to previous understanding of underlying musical processes gained through ethnographic enquiry. These issues are explored in the context of a specific musical genre, the North Indian instrumental gat. ‘Gat’ refers to a short composition idiomatic to North Indian plucked lutes such as the sitar and sarod, which serves as the basis of extensive improvisations (Clayton, 1993; Dick, 2001). This genre can be regarded as essentially performed by a duo – plucked lute player and tabla accompanist – even if the duo are often joined by a third musician playing an unchanging reference part (‘drone’) on the tanpura. This focus allows us to consider a number of different factors in entrainment and to draw on a body of ethnographic and musical knowledge, while at the same time limiting ourselves to a single genre and restricting analysis to the interactions within a dyad.

## Sensorimotor synchronization and coordination

At least two main factors can be distinguished in musical entrainment (Clayton et al., submitted). The first is sensorimotor synchronization (SMS), a process by which the sound-producing actions of musicians can become tightly synchronized (with the standard deviation of signed asynchronies between event onsets often in the tens of milliseconds; Clayton et al., submitted; Rasch, 1988). This process seems to operate over relatively short timescales (typically in the hundreds of milliseconds), and its accuracy is

dependent on auditory information (Repp, 2005). The second we term ‘coordination’, referring to the management of performances over longer timescales, from a few seconds up to several minutes. This is more accessible to conscious control and reflection and can involve different modalities, including visual cues such as head nods or upper body sway. Such ‘ancillary’ (non-sound-producing) movements have been shown to play a key role in communicating expressive and structural aspects of the music between co-performers, as well as from performers to audiences (e.g. Davidson, 1993; Eerola et al., 2018; Wanderley et al., 2005; Williamon & Davidson, 2002).

One of the differences hypothesised between SMS and longer-term coordination is that while the latter occurs in relation to a body of shared, culturally-specific knowledge, the former can occur spontaneously, even in the absence of an intention to synchronize (Clayton, 2007; Lucas et al., 2011). Moreover, while synchronization dynamics may vary somewhat across musical contexts, similar phenomena are observed cross-culturally (Clayton et al., submitted): the mechanism seems based on a shared set of human capacities. Even if synchronization is in some sense universal, however, in practice it incorporates elements of culture-specific knowledge, often at very basic levels: clapping along with music, for example, often involves choices of beat level and downbeat position where the ‘correct’ answer is a matter of culturally-shared knowledge.

Coordinating music metrically requires knowledge of how different musical parts can relate to and create the abstract metrical structure – how long a metrical unit or cycle is, where the principal downbeat (if any) is located, and so on. At longer timescales transitions such as tempo changes or the switch to a new section require a shared understanding of the structures, processes, or transitions that are possible within a given musical style. The precise nature of these transitions varies between musical genres and traditions, as do methods for managing them successfully, but again some aspects of this coordination seem to be common to otherwise diverse musical cultures, such as the tendency for performers to use bodily gestures to visually indicate section boundaries (Bishop & Goebel, 2015; Huberth & Fujioka, 2018).

A dissociation between SMS and coordination can be observed in practice: tight synchronization does not guarantee the smooth management of transitions, and vice

versa. Nonetheless previous studies, and theoretical speculation, suggest that metrical position may influence synchronization: specifically, that downbeats show less variability (more precision) in asynchrony than other metrical positions (London, 2012; Polak et al., 2016). In this sense the interaction between low-level automatic processes and higher-level, culturally-mediated processes remains a matter for investigation.

### Experimental investigations of sensorimotor synchronization and coordination

In addition to specific work on Indian music, there is of course a large literature of experimental research on synchronization. Keller and Appel (2010) analysed performances by piano duos and found that synchronization accuracy was not significantly affected by visual contact, although synchronization was more accurate when the movements of the pianist playing the melody part preceded those of the accompanist. Ragert et al. (2013) studied the effect of familiarity on both keystroke synchronization and body sway coordination in piano duos, finding that while knowledge of the other musician's *part* facilitates prediction of their actions at the longer timescales associated with body movement, knowledge of the other's playing *style* is important in achieving tight synchronization. In a related study, Bishop and Goebel (2015) found that players of an accompaniment part were able to synchronize using only visual information about movement but were more accurate when audio information was also available. The present study aims to build on this work by investigating aspects of SMS and movement coordination in natural performances.

One difference between the work reviewed above and the music that is the focus of our research is that gats are not played from notation, but comprise short fixed compositions followed by extensive improvisation. Duos typically rehearse the fixed compositions before performing – especially if they involve any unusual rhythmic features – and also often final cadences, which can be very complex, but nothing else. Whereas in notated music a musician knows which notes to expect, but not necessarily the detail of timing and dynamics; in this music the musicians are familiar with a wide repertoire of rhythmic and expressive possibilities idiomatic to the genre, but are tasked with responding instantly to the choices of their co-performers, which may be more or less predictable. Not surprisingly performers tend to prefer working in stable partnerships, where they have a chance to become familiar with each other's habits: if not, accompanists often

familiarize themselves with soloists' styles in advance by watching recordings on the internet.

### **Interaction and co-performer relations in Indian instrumental music**

The present study is the first to quantitatively explore interpersonal entrainment in Indian instrumental music (see Clayton, 2007, for one additional quantitative study of entrainment in Indian vocal music). There is, however, a more substantial literature on co-performer communication and interaction, in particular in the instrumental gat genre. For instance Moran (2011) explores, via interviews with musicians, the conceptualization of musical interactions in this genre in terms of social relationships, while an observational study by Moran (2013) demonstrates a relationship between behaviours indicative of mutual attention (gesture and looking behaviour) and factors such as musical role and mutual familiarity and stresses the importance of eye contact in managing performance. Our study does not directly address gesture quality or eye contact, but quantifies the degree of coordination between upper body movements, which has been shown to be a significant predictor of expert annotations of co-performer interaction (Eerola et al., 2018). Future research could explore the relationship between such movement coordination and mutual attention via eye contact.

Exploring the nature of co-performer relations in more ethnographic depth and framed also by a Goffmanian perspective on social interaction, Clayton and Leante (2015) explored the relationship between musical role (as soloist and accompanist), social identity and status (especially seniority) and hierarchy in North Indian musical ensembles. They demonstrate that the overlap between different hierarchies (especially, those of musical role and seniority) creates possibilities for antagonistic relationships to develop; musicians performing with unfamiliar partners, in particular, frequently express anxiety about their co-performer's behaviour as well as their musical competence and sensitivity. Synchronization accuracy is not part of this discourse. Nonetheless, an instruction from soloist to accompanist to adjust or correct the tempo is one of the few examples of an instruction that can be given openly (i.e. in a way that can easily be understood by the audience), and this can be a source of antagonism in performance because it implies a dominant relationship and in some instances a public criticism.

### Metrical and cadential structures in Indian instrumental music

The principal musical challenge in terms of ensemble coordination in instrumental gat music relates to the maintenance of metrical, or tala, structure. Whereas in some kinds of music the smooth performance of structural transitions is an important issue, this is not often so in the North Indian gat. Performers certainly transition between sections of a composition, or between different kinds of improvisatory techniques, but these transitions are led by the soloist and the accompanist is expected only to respond promptly. When the lead position is handed from one to the other, for instance when the melody player invites the tabla player to take a solo, the latter is not required to respond instantly but can do so after a few seconds. Such transitions therefore cause little concern: the main issue is that both players should be confident that they share their understanding of the tala structure.

Talas are understood as cycles of fixed numbers of equal time units or matras (translated as 'beats', although this translation suggests implausible tempo ranges at both ends of the scale), which are organised into an intermediate level of beat groups or vibhags: thus the 16 matras of teental, the most common tala in this genre, are organised into 4 groups of 4 (see Appendix B). The tala is not just an abstract structure but is manifested by canonical patterns of drum strokes called theka, and by means of fixed sequences of hand gesture patterns.

In brief, whatever musicians do in terms of the development of melodic and rhythmic ideas has to be done in tala (with the exception of an unaccompanied introductory section, alap, that is performed without tala). This is made challenging by the often highly complex nature of those melodic and rhythmic ideas, which include a propensity for cross-rhythmic invention, syncopation, flexible expressive timing and so on. This means that musicians need to possess not only a virtuosic command of the musical techniques and materials, but also an ability to read each others' intentions and respond appropriately. The maintenance of the tala structure confirms that this mutual understanding has been achieved, and consequently the alignment of the metrical structure between co-performers is a key focus of this analysis.

At a higher level, improvised passages are punctuated with cadential figures that both close improvisatory ‘episodes’ and reaffirm the tala, often by emphasising beat 1 (‘sam’). Thus, our analysis also considers the relationship of synchronization and coordination to this *cadential* structure, over and above that of the tala structure itself (since many cycle boundaries are not marked in this way, particularly in fast tempo pieces).

### Aims of the Present Study

The aim of this article is to present the most extensive quantitative study of entrainment attempted to date based on natural performance data, in any musical genre – studying 10 pieces played by four duos, covering over two hours of recorded music – and in doing so to characterise the temporal interactions of instrumental duos in this genre. Specifically, we present a detailed analysis of synchronization between performers, and examine the role of shared periodic movement in managing coordination at metrical and cadential downbeats. These analyses are discussed in relation to the specific challenges presented by this musical genre, as understood both musicologically (what the musicians do rhythmically) and ethnographically (what musicians say about the challenges of interaction and coordination). By addressing these questions, we contribute not just empirical analyses of interpersonal synchronization in real-life musical performance, but also explore the potential interaction between synchronization and coordination mechanisms, bearing in mind the specific challenges and the nature of shared musical knowledge in this specific musical tradition. We also contribute to the understanding of metre in this genre. Since London (2012), metrical theory has been bound up with the notions of attentional periodicity and entrainment, thus the empirical investigation of relationship between entrainment and metrical structure is of relevance not only to entrainment theory but also to the theory of metre.

## Methods

### Materials

The analysis draws on a corpus of four complete raga performances by professional performers. These performances form a subset of the North Indian Raga corpus released by the Interpersonal Entrainment in Music Performance (IEMP) project: see Clayton,

Leante, and Tarsitani (2018).<sup>1</sup> The metred sections of these performances are ten in total (each performance also includes an unmetred alap in which the tabla artist does not play, which is not analysed here). The mean duration of the analysed sections was 732 sec (SD=327 sec). Summary statistics including names of performers, ragas and talas, duration and tempo can be found in Appendix A. These recordings are representative of the genre in a number of ways: the three most common plucked lutes are represented (sitar, sarod and slide guitar); the talas and tempi are typical of the genre, with a high representation of teental (16 beats) at different tempi from vilambit (slow, 40-65 bpm) to drut (fast, 200-650 bpm) and a lower incidence of other talas (dhamar tal = 14 beats, ektal = 12 beats and rupak tal = 7 beats); and the gender of the musicians, with most string players and almost all tabla accompanists male, is also representative of the genre (the one female soloist plays on performance 2). The duos are not long-established pairs, but all are experienced professional artists. Two of the performances were recorded in public performances in front of paying audiences in the UK, and one was recorded in front of an invited audience in a house concert in Kolkata, India. The last was recorded in a private studio session in the UK with a small audience: in this case the artists were asked to record a complete raga performance as if for a concert or broadcast. The ragas are all common melodic modes in the North Indian tradition. Selection of these examples took into account the availability of both audio recordings with separate tracks for the two instruments, and video recordings with clear and static views of the upper bodies of the two musicians. Performances are referred to in the text as performance 1, 2, 3 & 4; section names take the form <2a\_StrA\_Slow> where 2a refers to the first section of performance 2, StrA means ‘Sitarist [with initial] A’ and ‘Slow’ identifies the tala and tempo (‘Slow’ and ‘Fast’ always refer to vilambit and drut teental respectively; the others refer to the tala names: Teen(tal, at medium tempo), Ek(tal), Dham(ar tal) and Rup(ak tal)). The main tala structures are represented in Appendix B.

### Event onsets and audio annotation

The discussions below draw on the following analyses of the audio recordings:

---

<sup>1</sup> See Appendix A.

1. Extraction of event onsets for each instrument in each section.<sup>2</sup>
2. Manual annotation of metrical downbeats, allowing beat positions to be estimated by dividing each cycle into equal parts.
3. Alignment of selected event onsets with metrical (beat) positions. This was done by
  - a. identifying the closest tabla onset to the estimated beat position, within a specified time window (windows varied according to tempo, between +/- 30ms and +/- 160ms). This was done for each matra in all cases, as well as for half-matra divisions in the slowest pieces in each performance (1a, 2a, 3a, 4a).<sup>3</sup>
  - b. identifying the nearest melody instrument onset within a defined time window of the tabla onset or, if not present, of the estimated beat position.
  - c. manual checking of the selected onsets to remove false positives and correct mislabelling. This also allowed identification of performance errors such as dropped or added beats. This process is illustrated in Figure 1.
4. Manual annotations of structural units and major events (e.g. tabla solos, cadential figures).<sup>4</sup> The most common cadential figures are the mukhra – the opening phrase of the fixed composition used as a refrain – and various forms of tihai or triple repetition patterns. They are hardly used at all in the jhala part that concludes each drut teental section: these portions are omitted from the cadence analysis. The number of cadences annotated in each section can be found in Appendix A.

<INSERT FIG 1 HERE>

---

<sup>2</sup> Onsets were extracted based on envelope characteristics using MIR Toolbox (Lartillot, Toivainen, & Eerola, 2008). First the audio signal is band-pass filtered to eliminate frequencies above 900-1200 Hz to focus on body resonances of these plucked instruments. The envelope of the filtered signal is then extracted and subjected to low-pass filtering and half-wave rectification before applying peak-picking algorithms with three parameters. These parameters determine (1) the local contrast threshold value, (2) normalised amplitude threshold value, and (3) threshold value for the minimum difference between peak values. The optimal values of these parameters were defined for each instrument using a 1-minute sample of manually annotated onsets against which the algorithm performance was compared using F-measure and 70 ms tolerance for the correct onsets, reaching F-measures between 0.75 and 0.95. The rationale for the relying on one particular onset extraction technique that relies on envelope energy instead of other possible components such as timbral flux, for instance, was to keep the onset timings comparable across instruments in terms of the lag inherent in onset detection. This technique provides an excellent onset detection accuracy for tabla and with the above-mentioned parameters; the onsets in the plucked instruments could also be extracted reliably using an identical temporal resolution.

<sup>3</sup> Selected onsets were also checked for double selection, i.e. where the same onset fell within the specified window of two adjacent estimated beats.

<sup>4</sup> In both Elan (Wittenburg et al 2006) and Sonic Visualiser.



Onset extraction and labelling present significant challenges in this repertoire, even when using high quality recordings with separate microphones for each instrument. Challenges for onset extraction include the highly complex auditory signals produced by the stringed instruments, and the very wide dynamic range of all instruments. Specific challenges requiring manual correction include, apart from false positives due to over-sensitive detection or bleed between instruments, the tendency for tempo to fluctuate within cycles in slow-tempo pieces, and for rhythmic figures on the melody instruments to be articulated slightly before or after the nominal (equal) metrical position. The proportion of onsets requiring intervention (to add, remove or reassign the automatically extracted onsets) was typically 5-15% of the total assigned by windowing. These judgements were informed by expert knowledge of the genre, in particular in matters of rhythm and including playing experience (see Clayton 2000), but inevitably include a subjective element.

### Movement extraction

To quantify aspects of movement coordination, we performed two-dimensional (2D; vertical and horizontal) tracking of each performer from each video recording within the corpus. Specifically, we tracked performers' ancillary upper body movements (head and upper torso). Despite the relatively low sampling rate (25 frames per second) of these video recordings, it is still feasible to track these types of ancillary movements, due to the longer timescale over which movements such as head nods and body sway occur (in comparison to sound-producing movements, such as key presses on a piano; see Jakubowski et al., 2017).

Movement tracking was done using a semi-automated process in EyesWeb ([http://www.infomus.org/eyesweb\\_ita.php](http://www.infomus.org/eyesweb_ita.php)). A region of interest (ROI) was manually selected around the head and upper torso of each performer, and 2D movement was then automatically tracked within each ROI using a dense optical flow algorithm, based on that of Farnebäck (2003). This technique has been validated for tracking ancillary movements of musical performers in Jakubowski et al. (2017).

Once 2D movement coordinates had been extracted from each performer, we used cross wavelet transform (CWT) analysis as a measure of co-occurring periodic movement between the two performers. This technique allows one to detect co-occurrences of

periodic movements across multiple frequencies over time. In addition, CWT Energy across a broad frequency range (0.3 to 2.0 Hz) has been shown to be a significant predictor of annotations of ‘bouts of visual interaction’ by expert musicians (Eerola et al., 2018), indicating that this computational measure is effective in detecting co-occurrences of movement that are considered meaningful and communicative by human raters. In the present study, we first detrended the x- and y-coordinates from each performer from the optical flow data, converted these into polar coordinates, and retained the radial coordinates ( $\rho$ ) for the CWT analysis. The CWT analysis was then applied across the frequency range from 0.3 to 2.0 Hz, as in Eerola et al. (2018), and the resultant CWT Energy measure was normalized to a range of 0–1 for subsequent analysis.

## Results 1: Sensorimotor synchronization

Sensorimotor synchronization was analysed across the corpus using the annotated event onset times. Approximately 16,700 pairwise asynchrony values were calculated from over 43,000 annotated onsets (i.e. asynchronies were computed as onset time differences where both onsets were assigned to the same metrical position). Two aspects of synchronization were quantified using these pairwise asynchronies: **mean relative position** and **precision**.

In this study, the mean relative position of the two instruments was measured as the **Mean signed asynchrony**. This measure gives an indication of whether (and by how much) one instrument tends to play ahead in time with respect to the other; negative signed asynchrony values in this case indicate the melody instrument playing ahead of the tabla. Precision, or tightness of synchronization, was measured using **Asynchronization**, which is defined as the standard deviation (SD) of the signed asynchronies (lower values indicate more precise synchronization; Rasch, 1988, p. 73). In cases where the standard deviation is not an appropriate statistical measure (e.g. in cases where raw asynchrony data is more appropriate as input for analysis than a summary measure such as SD), we used the **Mean absolute (unsigned) asynchrony** as a measure of ‘precision’. Application of this measure requires clarification, since usage in the literature varies significantly. For instance, Ragert et al. (2013) refer to mean absolute

asynchrony as a measure of synchronization ‘accuracy’,<sup>5</sup> distinguishing it from ‘precision’, while Bishop and Goebel (2015) use the same measure to judge how “successful” synchronization was within duos, which could be interpreted as ‘precision’. In fact, mean absolute asynchrony is highly correlated with asynchronization in our dataset (comparing results by 2-minute Segment, see following paragraph;  $r(58) = 0.939$ ,  $p < .001$ ), and therefore may be used as an alternative measure for ‘precision’.<sup>6</sup>

To compute summary measures for some of the following analyses (e.g. mean signed asynchrony, tempo, etc., by section), the pieces in the corpus were divided into Segments of roughly equal duration from which these summary measures were calculated ( $N = 60$ ,  $M_{\text{duration}} = 120.6$  s,  $SD = 8.1$  s). Each piece was divided roughly equally in order to produce Segment lengths around 2 minutes; boundaries were adjusted to the nearest metrical boundaries. The analysis by Segment allowed us to calculate correlations between mean relative position/precision and other parameters such as mean tempo, and also ensured that slow and fast sections were represented in compatible fashion (the raw data for fast sections comprised many more points than those for slow sections).

#### Distribution of signed asynchrony values

The overall distribution of signed asynchronies between melody instrument and tabla is symmetrical, with a mean close to zero ( $M = -2.34$  ms,  $SD = 27.87$  ms; see Figure 2). The mean position indicates that the melody instrument plays slightly ahead. If this can be regarded as an example of the ‘melody lead’ phenomenon (Goebel, 2001; Palmer, 1989; Repp, 1996) then the effect is small however: previous studies reported melody lead in Western music performance of up to 30ms. ‘Melody lead’ is not evident in the slow sections: the mean signed asynchrony for all slow sections was +1.25 ms, in contrast to -2.63 ms for fast sections.

<INSERT FIG 2 HERE>

In order to explore the variability of the data in more depth, the distribution of asynchronies was explored by 2-minute Segment. This reveals considerable variation

---

<sup>5</sup> ‘Accuracy’ is often used in the literature to refer to the aspect of synchronisation measured here as mean signed asynchrony.

<sup>6</sup> There is a correlation between mean absolute asynchrony and mean signed asynchrony, but this reflects the fact that Segments with more precise synchrony tend to have higher melody lead, i.e. in this corpus more negative mean signed asynchrony values.

within and between performances. Figure 3 plots the mean signed asynchrony of each Segment: the mean position for performance 2 shows small positive values in all Segments (melody lag,  $M = 5.53$  ms), while many Segments of performances 3 and 4 show larger negative values (melody lead,  $M = -5.11$  ms and  $-8.67$  ms respectively) and performance 1 has a mean close to zero ( $M = -0.33$  ms). The spread of values for the different 2-minute Segments ranges from a melody lead of about 16 ms to a melody lag of about 14 ms. The following sections will explore some of the factors in the music that account for this variability.

<INSERT FIG 3 HERE>

Another way to investigate the existence of ‘melody lead’ is to consider what happens in instances in which the tabla takes solos and the string player provides accompaniment (i.e. the normal roles are reversed). Across the corpus the lead instrument plays ahead of the one performing an accompanying role – the ‘melody lead’ phenomenon is in fact a ‘lead instrument’ effect whether that instrument is melodic or tabla ( $M = 4.27$  ms for tabla solos,  $-3.04$  ms for the non-tabla solo sections:  $t(1811.1) = 8.45$ ,  $p < .001$ ). This is visually evident in Figure 4, which illustrates the section with the most extensive tabla solos as well as the most pronounced difference in synchronization between tabla solos and other sections (1a\_Srd\_Rup: Tabla solo sections:  $M = 13.84$  ms; remaining sections:  $M = -8.07$  ms,  $t(1016.7) = 11.68$ ,  $p < .001$ ). This pattern is only statistically significant ( $p < .05$ ) in 4 of the 7 pieces including tabla solos, however.

<INSERT FIG 4 HERE>

The asynchronization for each Segment is represented as error bars on Figure 3, and indicates a tendency for this variability to reduce (i.e. precision to increase) over the course of each performance. This is to be expected since all performances increase in tempo, while previous studies have found variability of asynchrony to increase with the duration of the inter-onset interval (IOI; e.g. Zendel et al., 2011.). The overall asynchronization for the slow sections is 36.30 ms, compared with 24.13 ms for the fastest four sections (in drut teental). Synchrony also tends to be looser for tabla solos than for sections in which the melody leads ( $SD = 33.28$  vs.  $27.16$  ms; non tabla-solo sections are more precise than tabla solos in 6 out of 7 sections).

### Relationship of asynchrony with tempo and event density

It should be borne in mind when considering the summary results above that performances in this genre tend to increase in tempo and event density. Occasional local decelerations are encountered in a couple of these performances, but the dominant pattern is one of acceleration. There seems therefore to be a pattern of reducing variability with increased tempo, and an indication that melody lead is overall higher in faster sections. In this section we explore these relationships between synchronization and speed more systematically.

First, the mean tempo of each Segment was calculated from the metrical annotations. Significant negative correlations were found between tempo and mean signed asynchrony ( $r(58) = -.257$ ,  $p = .047$ ) and tempo and asynchronization ( $r(58) = -.722$ ,  $p < .001$ ). Thus, with increasing tempo synchrony becomes tighter and the melody instrument tends to play further ahead.

It should be noted that the nominal mean tempi calculated, as here, as the number of matras (basic time intervals) per minute, cover a very wide range, from roughly 40-600 bpm: this is not a realistic estimate of the range of perceived tactus rates. We therefore tested this relationship within specific tala and tempi, to see if the correlation holds. The two categories with the most data were slow and fast teental sections. In slow teental sections we found no significant correlation between tempo and either mean signed asynchrony or asynchronisation ( $p > .05$ ). In the fast teental sections, however, we found

a significant correlation between precision (asynchronisation) and the mean tempo of the Segments ( $r(24) = -.472, p = .015$ ).

An alternative method of checking the relationship between speed of playing and synchrony was employed, using event density calculations. For each onset, a local event density for that instrument was calculated as the number of detected (both selected and unselected) events per second, calculated over 2-second windows. These figures were averaged over each Segment to produce mean event density figures. Correlations were explored between these event densities (for each instrument separately, and for the sum of the two figures) and the asynchrony data.

Significant negative correlations were found between event density and asynchronization (Melody density:  $r(58) = -.347, p = .007$ , Tabla density:  $r(58) = -.615, p < .001$ , Summed densities:  $r(58) = -.541, p < .001$ ). In the slow teental sections we found a significant correlation between mean signed asynchrony and event density of the string instrument ( $r(14) = -.800, p < .001$ , for the tabla,  $r(14) = -.398, p = .127$ ; for the summed density,  $r(14) = -.434, p = .093$ ), and a marginally significant correlation between asynchronisation and summed density ( $r(14) = -.497, p = .05$ ). In the fast teental sections, in contrast, we found no significant correlations with density ( $p > .05$ ).

Thus, synchronization was significantly more precise at higher tempi and event densities. Tempo and event density are, not surprisingly, correlated ( $r(58) = .706, p < .001$ ), and this holds also within the slow teental and fast teental sections. However, there is scope for divergence between the two, for example in slow sections the musicians might use a very fast subdivision of a slow beat. Thus, it is worth enquiring whether the increase in precision is more closely linked to tempo or event density. In slow teental sections we found a correlation between precision and the melody instrument's event density but not tempo; in fast teental sections the opposite pattern of a correlation with tempo but not density emerged. This may simply be due to the fact that the range of values is greater this way round (e.g. density varies more than tempo in slow sections), and thus given a stable tempo, precision varies with density, while given a stable density it will vary with tempo.

There was no overall effect of event density on mean signed asynchrony. In summary, melody lead increases with increased tempo (but not increased event density); synchrony becomes tighter at higher speeds, whether measured as tempo or as event density.

### Relationship of asynchrony with metrical position

Next we examined the relationship between synchronization and metrical position. This analysis was carried out at different metrical levels, which are illustrated in Figure 5, using the vilambit (slow tempo) teental as an example. All metrical structures evidence multiple pulse levels in a hierarchical relationship, but the North Indian tala system is unusual in having three levels that are theoretically recognized – cycle, vibhag (beat group) and matra (beat). At slow tempi, more than one level of matra subdivision is also evident: in the slow pieces in our corpus we assigned onsets to the half-matra level. The illustration also includes a 2-matra level in the hierarchy: this has no place in North Indian music theory, but can nonetheless be hypothesized as potentially metrically significant in the overall binary structure.

<INSERT FIG 5 HERE>

Our analysis thus explores differences in synchronization at four metrical levels:

- (i) Between matras and half-matra subdivisions (i.e. beats and their subdivisions)
- (ii) Between odd and even matras. The rationale for this is that if the ‘naïve’ (unrepresented in music theory) 2-matra level is metrically significant, odd beats should be metrically stronger than even beats.
- (iii) Between matras that fall at the start of vibhags and all other matras. In contrast to (ii) this comparison is driven by Indian music theory: the matras beginning vibhags should be metrically stronger than the others.
- (iv) Between matra 1 (sam) – the strongest metrical position – and other matras falling at the start of vibhags.

First, then, we tested whether synchronization differed between **(i) matra and half-matra positions**, for the 4 slow pieces only (1a, 2a, 3a, 4a). We computed the mean signed asynchrony and asynchronization for matra and half-matra positions separately for each of the 2-minute Segments described in the previous section. In paired-samples t-tests, there was no significant effect of matra/half-matra position on either mean signed asynchrony ( $t(25) = -1.23, p = .229$ ) or asynchronization ( $t(25) = 0.088, p = .931$ ).

Next, we investigated synchronization differences between **(ii) odd and even matras** for all pieces in the corpus (with the exception of 1a\_Srd\_Rup, which has an odd number of matras, and 3a\_Gtr\_Dham, which is divided into halves of 7 matras). Mean signed asynchrony and asynchronization were computed across Segments for odd versus even matras. There was a significant effect of odd/even matra position on mean signed asynchrony ( $t(49) = -2.92, p = .005$ ), with the melody instrument playing more ahead on even ( $M = -2.19$  ms) than odd ( $M = 0.27$  ms) matras. No relationship was found with asynchronization ( $t(48) = -1.70, p = .095$ ).

We then investigated differences in synchronization between **(iii) matras at the start of groups (vibhags) and non-vibhag matras**. In this case, vibhag matras are defined as:

Teental (all tempi): 1, 5, 9, 13

Rupak tal: 1, 4, 6

Ektal: 1, 5, 9, 11<sup>7</sup>

Dhamar tal: 1, 6, 8, 11

Because there are many fewer instances of matras that fall at the start of vibhags than non-vibhag matras, this analysis used a different method based on random sampling of an equal number of asynchronies to the vibhag matras and computing the difference in mean asynchronies between each group. We randomly sampled (without replacement) an equal number of asynchronies falling on non-vibhag matras as the number of asynchronies falling on vibhag matras for each piece in the corpus. This analysis was repeated 1,000 times, and mean differences and 95% confidence intervals (CIs) were computed for each piece. In this analysis we used mean absolute asynchrony for the ‘precision’ measure, in order to use an analogous sampling procedure for both signed and absolute asynchronies (rather than using a summarized SD calculation). Results are displayed in Figures 6 and 7. There was a non-significant tendency across the corpus for the melody instrument to play earlier on non-vibhag than vibhag matras (mean difference in signed asynchronies =  $-1.32$  ms, 95% CI:  $[-3.99, 1.39]$ ) and a non-significant tendency toward smaller absolute

---

<sup>7</sup> In fact, there are six vibhags in ektal, starting on matras 1, 3, 5, 7, 9 and 11. Since this is identical to the odd/even matra analysis, however, we substituted a simplified four-vibhag structure often evident in performance.



asynchronies (greater precision) on non-vibhag matras ( $M_{\text{difference}} = -1.25$  ms, 95% CI: [-3.03, 0.58]).

<INSERT FIG 6 HERE>

<INSERT FIG 7 HERE>

As a final step in testing the effects of metrical position, we compared synchronization on **(iv) matra 1 (sam) and other matras falling at the start of vibhags**. This analysis followed the same random sampling procedure as reported above, due to the unequal distribution of the two sets of data being compared. Across the corpus there was only a small, non-significant difference in both signed ( $M_{\text{difference}} = -1.25$  ms, 95% CI: [-3.03, 0.581]) and absolute asynchronies ( $M_{\text{difference}} = 0.537$  ms, 95% CI: [-2.06, 3.17]), in the direction of the melody instrument playing earlier in time on matra 1 and tighter synchronization on matra 1 than on other vibhag beats.

#### Relationship of asynchrony with specific performance factors

Beyond global factors such as tempo, event density and metrical position, to what extent can specific musical practices affect synchronization? Many factors could be considered here, of which we focus on two that relate to musical energy, tension and relaxation – specifically, loudness and cadential figures. We also report observations on the occurrence of metrical errors.

#### Dynamic level

When musicians play quietly they are often playing in a relaxed mode, while loud playing is associated with high energy, tension and musical climax. Does playing more loudly lead one to play further ahead of one's partner, and does it have any effect on tightness of synchronization? Peak levels were calculated during the onset detection process using peak RMS amplitude. The peak values were then averaged over each Segment for each instrument. It might be thought that loudness is positively correlated with speed; however, in this corpus mean peak levels are not significantly correlated with mean event density levels (Melody:  $r(58) = -.210$ ,  $p = .107$ , Tabla:  $r(58) = -.042$ ,  $p = .750$ ). Mean peak levels were negatively correlated with mean signed asynchrony but this is significant only for the melody instrument (Melody:  $r(58) = -.458$ ,  $p < .001$ , Tabla:  $r(58) = -.212$ ,  $p = .104$ ). Thus, the melody instrument tends to play further ahead when it plays

louder. This may be linked to the ‘melody lead’ effect, in that the melodic part is given prominence by both playing louder and further ahead at certain points. No significant correlation was found between peak level and asynchronization.

### Cadences

To test whether synchronization patterns differed at cadences (as defined in the Methods section), we compared asynchronies at cadences, all of which occurred on matra 1, to asynchronies at non-cadential instances of matra 1. As there are many more instances of non-cadential than cadential matra 1, we used the same random sampling procedure for comparing equal numbers of asynchronies as reported above (with 1,000 iterations). All pieces were included in this analysis except 3d\_Gtr\_Fast, which contains only 5 annotated cadences. Figure 8 shows the mean signed asynchronies for cadential and non-cadential instances of matra 1. The melody instrument is more ahead of the tabla player on cadences than non-cadences for 8 out of the 9 pieces. Although there is no significant difference across the corpus as a whole (mean difference in signed asynchronies = 8.65 ms, 95% CI: [-4.66, 21.88]), this difference was statistically significant for 4 of the individual pieces (2a, 2b, 3b, 3c). Mean absolute asynchronies were not significantly different at cadences than non-cadences (mean difference in absolute asynchronies = 0.562 ms, 95% CI: [-7.43, 9.47]).

<INSERT FIG 8 HERE>

### Errors

Three metrical performance errors were identified in this corpus.<sup>8</sup> Only the first of these is associated with any noticeable uncertainty, which may be due to a momentary loss of concentration following a cadential downbeat. The second occurs during a passage of melodic improvisation that seems tightly synchronized – in fact it is difficult to work out exactly where the error occurred – while the third occurs in the accompaniment of a tabla solo and again no disruption to synchronization is evident. These instances suggest that while a metrical error may be associated with a loss of synchronization it is not

---

<sup>8</sup> 1b\_Srd\_Fast, cycle 54 (added beat), 1b\_Srd\_Fast, cycle 113 (half cycle dropped), and 2a\_StrA\_Slow, cycle 66 (added beat).

necessarily so: it is possible to play through an error without compromising synchronization.

### Summary

The global picture of sensorimotor synchronization in this corpus of North Indian instrumental gat performances shows a symmetrical distribution of signed asynchronies with a mean melody lead of about 2 ms and a standard deviation of about 28 ms. The mean relative position of the instruments conceals considerable variation, with either instrument up to c. 15 ms ahead over particular 2-minute Segments. The melody instrument tends to play further ahead in sections with faster tempi, when the melody instrument plays louder, and at cadences, while the tabla tends to lead during tabla solos.

As for variations in the tightness of synchrony, this increases with faster tempo or event density. Metrical position showed very little effect on synchronization; only one small effect was found which indicated that the melody instrument played more ahead on even than odd matras, albeit with a difference in means of less than 2 ms.

## Results 2: Movement coordination

We pursued two primary strands of analysis to examine the relationship between co-performer ancillary movement coordination and the musical structure. First, we tested whether increased movement coordination (operationalised here as CWT Energy across the two performers) was present at metrical section boundaries (operationalised here as the start of each metrical cycle, i.e. *matra 1* or *sam*), in comparison to non-section boundaries. Second, we tested whether increased movement coordination could be evidenced at cadences in comparison to non-cadential instances of *matra 1*.

### Metrical boundaries

This first analysis, comparing metrical cycle boundaries versus non-boundaries, used the four slow pieces only. We focused on these slow pieces as the metrical cycle occurs over a timescale of 12.67 s on average (range of mean cycle durations: 4.79 - 19.16 s). In the faster pieces, where the cycle duration can be 3 s or less, it is unlikely that performers will show increased visual coordination at the start of every cycle, particularly

considering the relatively slow timescale over which periodic ancillary movements such as body sway occur (see Eerola et al., 2018).

We first computed the mean CWT Energy in a window centered around each section boundary. The window size varied as a function of the metrical cycle length for each performance, and was computed as 10% of the mean metrical cycle duration for that particular performance. To select sections of the performance of equal duration for comparison that did *not* include section boundaries, we shifted each section boundary to a new randomly selected timepoint, with the conditions that a) each boundary could move either forward or backwards in time up to the 50% boundary with the next section boundary, and b) each boundary had to be moved forward/backward in time at least the length of the window size as computed above. This preserves the number and relative spacing of section boundaries, whilst ensuring that the CWT Energy values analysed around these ‘fake’ section boundaries do not contain any of the same data accounted for in the analysis of CWT Energy around the real section boundaries. We then compared the mean CWT Energy around the real versus fake section boundaries, using the same window size for each, and repeated this process 1,000 times, with a new round of random sampling of fake boundaries for each iteration. We calculated the pairwise mean difference between CWT Energy in the window around each real section boundary and its corresponding fake section boundary for each iteration, and computed an overall mean difference and 95% CI across all iterations of the analysis for each of the four pieces.

This analysis revealed that movement coordination between the two performers was significantly higher at section boundaries (i.e. around *matra 1*) than non-section boundaries in two of the four pieces (2a and 4a – the two sitar pieces in *vilambit teental*), as mean differences were greater than zero (mean differences of 0.057 and 0.042, respectively), and the 95% confidence intervals around these means did not include zero (CIs: [0.042, 0.069], [0.022, 0.057], see Figure 9).

<INSERT FIG 9 HERE>

A subsequent question raised by this analysis is whether the higher degree of movement coordination evidenced at section boundaries is related to the presence of cadences on *matra 1*. To begin to investigate this question, we re-ran the analysis above, but excluded

all cadential instances of matra 1. CWT Energy at section boundaries was not significantly higher than at non-boundaries, suggesting that cadence points contributed substantially to the CWT Energy differences seen at section boundaries in the initial analysis. This led us to explore the role of cadences in more detail, as reported in the following section.

### Cadences

Next we compared movement coordination on cadential versus non-cadential instances of matra 1 (see Figure 10 for an illustration). This analysis was performed on all pieces within the corpus, with the exception of 3d, which only included 5 cadences. The remaining pieces contained 14.78 cadences on average (range = 9-25 cadences).

<INSERT FIG 10 HERE>

For each piece of music we randomly sampled an equal number of non-cadence instances of matra 1 to the number of cadences, and calculated the mean CWT Energy within a window centred around each selected timepoint (10% of the mean cycle duration, similarly to the metrical cycle analysis above). We repeated this random sampling and comparison between cadences and non-cadences 1,000 times and computed the mean difference in CWT Energy and 95% CI across all iterations for each piece. All pieces in the corpus showed greater movement coordination at cadential than non-cadential matra 1 positions (see Figure 11, mean difference across the corpus = 0.145, 95% CI: [0.143, 0.147]), with mean differences of greater than zero for all pieces and no CI spanning zero. There are some differences within individual performers by piece, although there does not seem to be a consistent pattern related to, for instance, whether a fast or slow tempo piece is being performed.

<INSERT FIG 11 HERE>

### Discussion

This study has characterized interpersonal entrainment in the North Indian instrumental gat genre. Two separate sets of analysis were carried out, one of synchronization (the ongoing, mostly automatic, beat to beat temporal alignment between the musicians, effected to a large extent using auditory information and extending over hundreds of milliseconds; see ‘Sensorimotor synchronization and coordination’ above), and the other termed coordination (describing longer-term processes of aligning metrical and formal

structures and managing transitions, more accessible to conscious control and often manifesting in coordinated movement). Variability of synchronization with tempo, event density, metrical position, lead instrument and cadential structure, and of movement coordination with metrical and cadential structures, were explored. The aim was to understand both of these entrainment mechanisms, and subsequently combine the perspectives in order to build up a picture of temporal coordination, including the ways in which these mechanisms interact and/or complement each other, taking into account musicological and ethnographic perspectives on this genre.

The event onset analysis shows that **precision** is higher in faster pieces and increases with tempo and event density. The precision of Segments from the slow pieces (most lying in the range of 30-50 ms) is consistent with Western chamber music groups (Rasch, 1988), while the faster Segments, in the range 17-21 ms, are comparable with Afrogenic drum ensembles from Mali and Uruguay (SD=16-18 ms, Clayton et al., submitted).

The **mean relative position** shows the melody instrument plays ahead of the tabla by the small margin of about 2 ms. This ‘melody lead’ effect is found in fast tempo pieces but not slow; it increases with the loudness of the melody instrument, with tempo and at cadences, and it is reversed in tabla solo interludes. It also varies significantly between performances, with 2\_StrA showing no melody lead in any Segment.

The possibility of a relationship between SMS and metrical position was explored at four different metrical levels. Virtually no effects of theoretically ‘strong’ versus ‘weak’ beats were found (with one small significant difference between mean signed asynchrony on odd vs. even matras). A larger, though non-significant difference was found at cadential downbeats, which see the melody instrument > 8 ms earlier on average with respect to the tabla than on other examples of matra 1.

Synchronization between the musicians is therefore affected by factors including tempo, dynamics, and leadership, with only small effects of metrical position.

Analysis of the movement data showed that coordination was greater at metrical boundaries (around matra 1) than at other points in the cycles for 2 out of the 4 slow pieces. Further exploration demonstrated that this effect was largely due to coordination around cadential downbeats. Across the whole corpus, movement coordination is significantly higher at cadential downbeats than at other instances of matra 1.

How, then, do musicians synchronize and coordinate their performance with reference to the tala structure, and how is their coordination evidenced in performance timing and movement? These results appear to suggest that SMS occurs from beat to beat with little mediation by the metrical structure. Cadences show a greater difference, but more in relation to melody lead (which is greater on cadential downbeats) than to precision. On the other hand, the metrical and particularly cadential structures leave a clear trace in terms of the coordination of movement between the musicians. This would suggest a distinction between short-term synchronization, where several factors affect melody lead but only tempo and event density have a significant effect on precision, and longer-term coordination, which is related to higher movement coordination at those metrical boundaries marked by the occurrence of cadences.

The impact of cadential structure on SMS, in the form of increased melody lead, can be understood as the result of lead instruments reaching a climax at the end of improvisation episodes, speeding up slightly and playing ahead of their partners on the downbeat: this is not linked to higher precision. Analyses of temporal causality of movements between musicians in orchestras (D'Ausilio et al., 2012) and string quartets (Glowinski et al., 2015) have provided similar evidence and suggested that being ahead is a communicational cue and related to the leadership roles in ensembles. The greater movement coordination identified in this study relates to Moran's finding that musicians engage in eye contact particularly on the first beat of the cycle (2013). Investigation of the relationships between eye contact, increased movement coordination and cadential downbeats would shed light on the social as well as musical performance dynamics. Observation of the small number of errors (beats dropped or added) suggests that these occurrences need not result in a loss of synchrony, consistent with the hypothesis that SMS and coordination are distinct mechanisms. The dissociation and complementarity of function between the two is not the whole story however. It is evident while checking the onset selection that deciding which events fall on which metrical position is frequently a non-trivial task. This is evidenced in a number of different ways:

- a. Expressive timing. Particularly in slow sections, soloists may deliberately place onsets that are associated with a metrical position significantly early or late (c. 150-250 ms).

- b. Cross-rhythmic variation. This music employs a number of techniques for rhythmic variation, known as *laykari*, involving different beat subdivisions and contrametric accenting. Maintaining coordination depends on the accompanist correctly interpreting the soloist's intention.
- c. Tempo fluctuation. Tempo sometimes increases sharply due to a deliberate decision by the soloist, and can also fluctuate within cycles.

A good accompanist, of course, has to be able to judge whether an onset that falls earlier than expected is an indication of an acceleration, of expressive timing or of a cross-rhythmic figure. This is where musicians' discourse about the need for attentive and sensitive accompanists is relevant. This is partly about social interaction, but also has a technical sense: a sensitive co-performer is also one who can distinguish expressive timing from a random variation in asynchrony, an acceleration from a cross-rhythm, and a deliberate off-beat accent from an error.

Even if SMS is a universal and innate mechanism, in this music it cannot be effective without an understanding what one is synchronizing to: which note or drum stroke in a dense and varied acoustic pattern is the one that falls on the beat? In this sense SMS *is* mediated by a shared, culturally-specific body of knowledge: however, it is not just the metrical cycle that is important, but rather the full range of idiomatic musical expressions that are possible and the ways in which each reinforces or challenges the perception of that metrical structure. This factor is not unique to this genre of instrumental *gat* or to Indian *raga* music. However, factors such as the avoidance of notation or detailed pre-planning, the value placed on virtuosity, the maintenance of strictly observed metrical structures and the diversity of rhythmic possibilities place a premium on a performer's ability to quickly grasp which of the myriad possibilities is in play.

If that knowledge is insufficient then coordination of the metrical structure will be lost, and without that coordination the musicians do not know what should be synchronized with what. This does not normally depend on the increased mutual attention at cadences, associated with greater movement coordination, whose function is in general not to achieve coordination but to reaffirm it. (In making eye contact and nodding together with a co-performer one does not say, "This is where I consider beat one to fall", but rather "We both clearly agree that this is beat one.")



Analyses of interpersonal entrainment on the basis of expert musical performance, such as this, are rare in comparison with those of performance in experimental conditions, and even more so in comparison with synchronization experiments. These studies are important in helping to understand how our capacities for mutual entrainment are expressed in the context which most creatively exploits them – musical performance – and how these expressions become socially and personally meaningful. The present contribution, by focusing on analysis of performance data in a specific genre, throws light also on more general issues such as the relationship between different entrainment mechanisms. We thus contrast a short-term synchronization process that is largely automatic and enabled by auditory information, which is affected by tempo and event density but largely blind to metrical structure, with a longer-term coordination mechanism in which coherent movement aligns with the music's metrical and cadential structure. The extent to which this pattern of complementary entrainment mechanisms at different time-scales extends to other genres of music warrants further discussion.

## Funding

This work was supported by the Arts and Humanities Research Council (grant no. AH/N00308X/1).

## Acknowledgements

The authors thank Simone Tarsitani, Technical Specialist at Durham University, for support, and Paolo Alborno, Antonio Camurri and Gualtiero Volpe from University of Genoa for providing support in computer vision extraction of the movement using EyesWeb software, and to all musicians, recordists and promoters involved in the analysed performances.

## References

Bishop, L. & Goebel, W. (2015). When they listen and when they watch: Pianists' use of nonverbal audio and visual cues during duet performance. *Musicae Scientiae*, 19 (1), 84-110. 10.1177/1029864915570355.

- Clayton, M. (1993). Two gat forms for the sitār: A case study in the rhythmic analysis of North Indian music. *British Journal of Ethnomusicology*, 2(1), 75-98.
- Clayton, M. (2000). *Time in Indian Music: rhythm, metre and form in North Indian rag performance*. Oxford: Oxford University Press.
- Clayton, M. (2007). Observing entrainment in music performance: video-based observational analysis of Indian musicians' tanpura playing and beat marking. *Musicae Scientiae*, 11(1), 27-59.
- Clayton, M, Eerola, T., Jakubowski, K., Keller, P., Camurri, A., Volpe, G. and Alborno, P. (submitted) Interpersonal entrainment in music performance: Theory and method. *Physics of Life Reviews*.
- Clayton, M. & Leante, L. (2015). Role, status and hierarchy in the performance of North Indian classical music. *Ethnomusicology Forum*, 24(3), 414-442.
- Clayton, M., Leante, L., & Tarsitani, S. (2018). *IEMP North Indian Raga*. <https://doi.org/10.17605/OSF.IO/KS325>
- D'Ausilio, A., Badino, L., Li, Y., Tokay, S., Craighero, L., Canto, R., Aloimonos, Y., & Fadiga, L. (2012). Leadership in orchestra emerges from the causal relationships of movement kinematics. *PLoS one*, 7(5), e35757.
- Davidson, J. W. (1993). Visual perception of performance manner in the movement of solo musicians. *Psychology of Music*, 21, 103-113. 0.1177/030573569302100201
- Dick, A. (2001) India, subcontinent of, §III, 6, (i) (b) Sitar and sarod. *Oxford Music Online*. doi-org.ezphost.dur.ac.uk/10.1093/gmo/9781561592630.article.43272
- Eerola, T., Jakubowski, K., Moran, N., Keller, P. & Clayton, M. (2018). Shared Periodic Performer Movements Coordinate Interactions in Duo Improvisations. *Royal Society Open Science* 5(2): 171520.
- Farnebäck G. (2003) Two-Frame Motion Estimation Based on Polynomial Expansion. In: Bigun J., Gustavsson T. (eds) *Image Analysis. SCIA 2003. Lecture Notes in Computer Science*, vol 2749. Springer, Berlin, Heidelberg. 10.1007/3-540-45103-X\_50.
- Glowinski, D., Dardard, F., Gnecco, G., Piana, S., & Camurri, A. (2015). Expressive non-verbal interaction in a string quartet: an analysis through head movements. *Journal on Multimodal User Interfaces*, 9(1), 55-68.

- Goebel, W. (2001). Melody lead in piano performance: Expressive device or artifact? *Journal of the Acoustical Society of America*, 110, 563–572.
- Huberth, M. & Fujioka, T. (2018). Performers' motions reflect the intention to express short or long melodic groupings. *Music Perception: An Interdisciplinary Journal*, 35(4), 437-453.
- Jakubowski, K., Eerola, T., Alborn, P., Volpe, G., Camurri, A. & Clayton, M. (2017). Extracting coarse body movements from video in music performance: A comparison of automated computer vision techniques with motion capture data. *Frontiers in Digital Humanities, section Digital Musicology* 4: 9.
- Keller, P. & Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Perception*, 28, 27–46. 10.1525/mp.2010.28.1.27.
- Lartillot, O., Toiviainen, P. & Eerola, T. (2008). A Matlab Toolbox for Music Information Retrieval. In *Data Analysis, Machine Learning and Applications*. Preisach, C., Burkhardt, H., Schmidt-Thieme, L. & Decker, R. Berlin, Germany: Springer. 261-268.
- London, J. (2012). *Hearing in time: Psychological aspects of musical meter*, 2<sup>nd</sup> edition. Oxford, New York: Oxford University Press.
- Lucas, G., Clayton, M. & Leante, L. (2011). Inter-group entrainment in Afro-Brazilian Congado ritual. *Empirical Musicology Review*, 6(2), 75-102.
- Moran, N. (2011). Music, bodies and relationships: An ethnographic contribution to embodied cognition studies. *Psychology of Music*. 41, 5-17. 10.1177/0305735611400174.
- Moran, N. (2013). Social Co-Regulation and Communication in North Indian Duo Performances. In Clayton, M. Dueck, B. & Leante, L. eds, *Experience and Meaning in Music Performance*. Oxford, New York; Oxford University Press, pp. 40-61. 10.1093/acprof:oso/9780199811328.003.0003.
- Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 331–346.
- Polak, R., London, J., & Jacoby, N. (2016). Both isochronous and non-isochronous metrical subdivision afford precise and stable ensemble entrainment: A corpus study

- of Malian jembe drumming. *Frontiers in Neuroscience*, 10, 285. 10.3389/fnins.2016.00285.
- Ragert, M., Schroeder, T. & Keller, P. E. (2013) Knowing too little or too much: the effects of familiarity with a co-performer's part on interpersonal coordination in musical ensembles. *Frontiers in Psychology*, 4, 368. 10.3389/fpsyg.2013.00368.
- Rasch, R. A. (1988). Timing and synchronization in ensemble performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 70-90). Oxford, UK: Clarendon Press.
- Repp, B. H. (1996). Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America*, 100, 3917–3932.
- Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*. 12. 969-92. 10.3758/BF03206433.
- Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, 34(1), 97-113. 10.1080/09298210500124208.
- Williamon A. & Davidson JW (2002), Exploring co-performer communication. *Musicae Scientiae*, 6, 53-72.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In: *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*.
- Zendel, B., Ross, B. & Fujioka, T. (2011). The effects of stimulus rate and tapping rate on tapping performance. *Music Perception*, 29, 65-78. 0.1525/mp.2011.29.1.65.

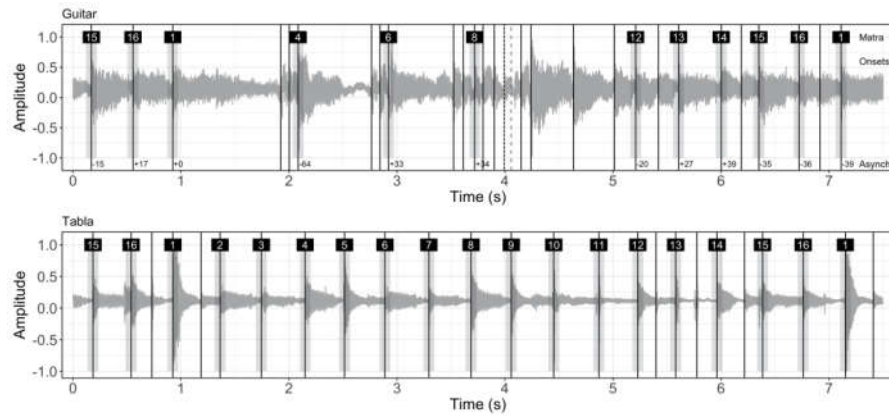


Fig. 1. Example of onset extraction and labelling. In this extract all detected onsets for the guitar (top) and tabla (bottom) are indicated by solid lines. Grey shading indicates the selection window, in this case  $\pm 100$  ms from each estimated beat position. White numbers in black boxes indicate matra numbers: in this case the onset selected at matra 9 was manually removed (first dotted line), as it was incorrectly identified; the correct onset for this beat position was not extracted (second dotted line). Asynchronies represented are given in ms. Example from piece 3b\_Gtr\_Teen, 562.8 to 570.5 secs.

1111x476mm (96 x 96 DPI)

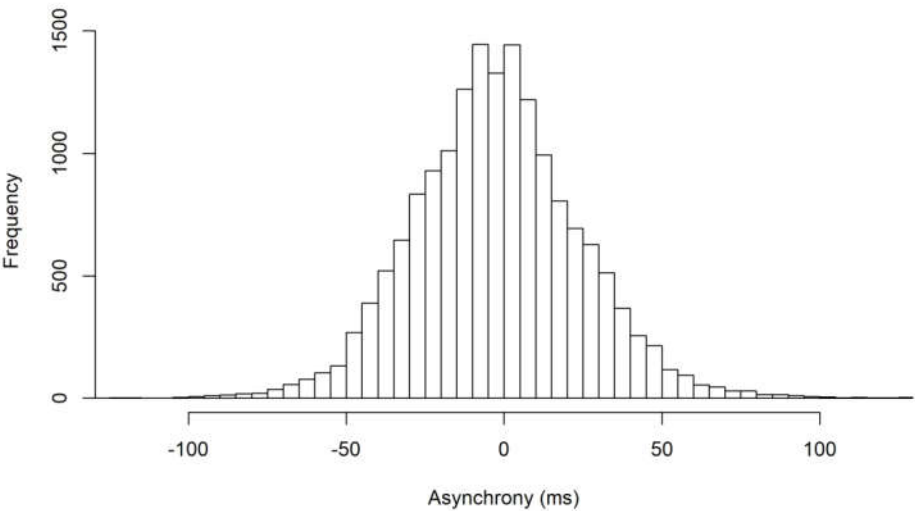


Fig. 2. Pairwise asynchronies between plucked strings and tabla (negative values indicate melody lead). A small number of values (0.25%) with a magnitude > 120 ms are excluded for ease of reading.

203x127mm (300 x 300 DPI)

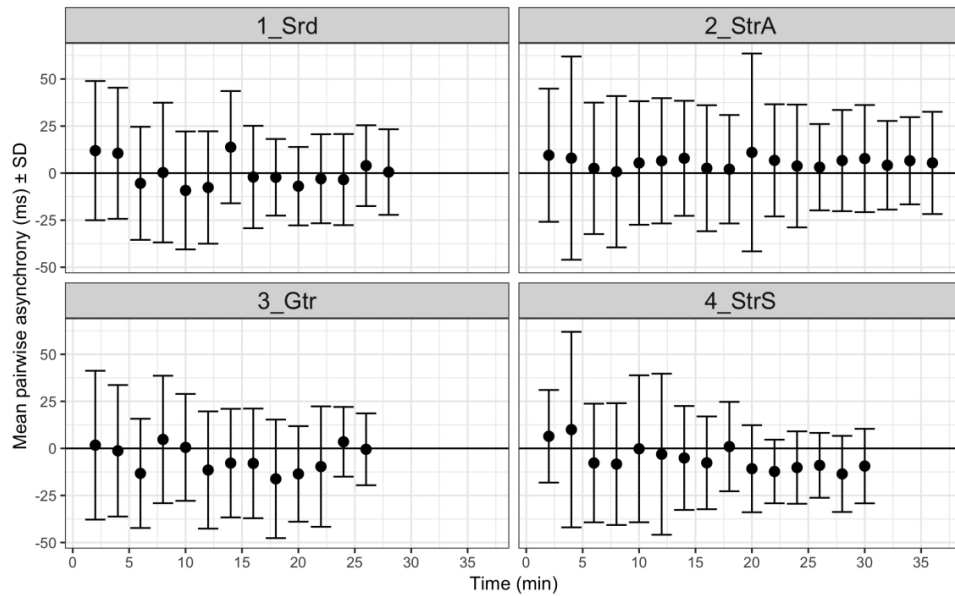


Fig. 3. Mean signed asynchronies for 2-minute Segments of the four North Indian gat performances in the corpus. Error bars show SD. For Segment summary data see Appendix C.

846x529mm (72 x 72 DPI)

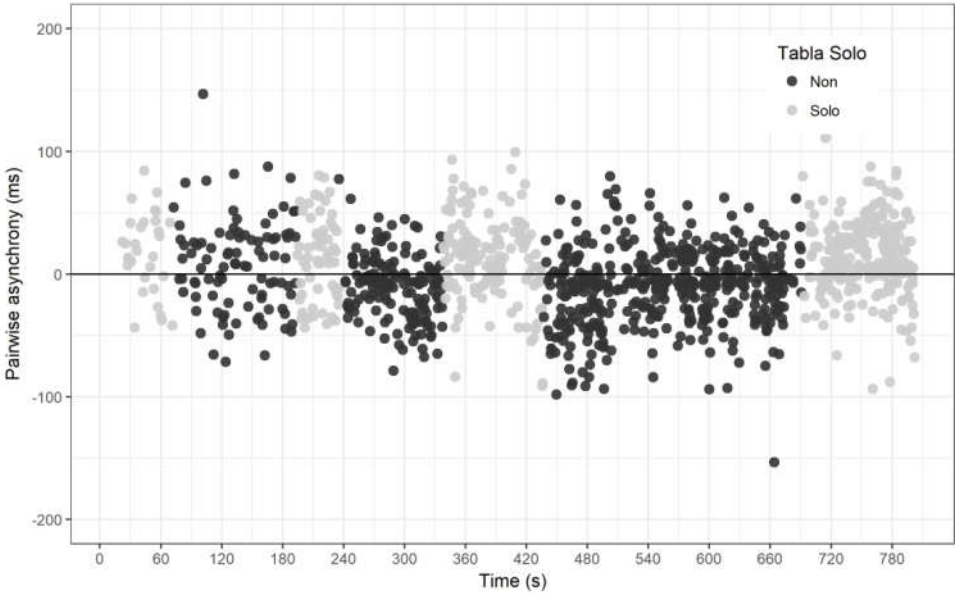


Fig. 4. Plot of raw signed asynchrony data against time for performance 1a\_Srd\_Rup.  
203x127mm (300 x 300 DPI)



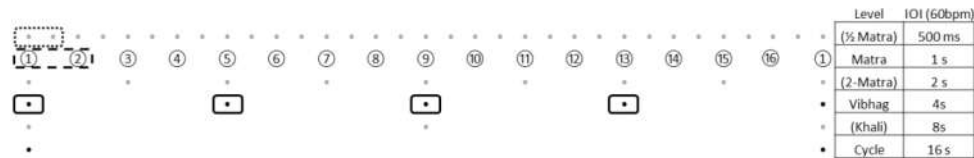


Figure 5. Vilambit (slow) teental, showing comparisons of metrical positions. The grid is based on the metrical grid diagrams popularised by Lerdahl and Jackendoff (1983): five metrical levels are shown in total, four as dots and the fifth as circled numbers (the matras). The top layer represents the half-matra subdivision, which forms part of our analysis (in practice, other subdivision layers are also used, but the half-matra is the most common). The second row represents the matras; the third, 2-matra groupings (which are not recognized as significant in the Indian theory). The fourth row represents the vibhags, which are an important part of the theory. The fifth row represents the half-cycle subdivision, which is recognized in as much as the half-way point in the cycle in this tala is marked by a distinctive hand gesture (a wave, known as khali), and the final row is the cycle (avartan). Labels and example IOIs (assuming a tempo of 60 bpm, which is within the normal range of this tala) are given on the right-hand side. The dotted box highlights the contrast between matras and half-matras (i.e. beats and offbeats): the diagram suggests that matra positions are 'stronger' than half-matra positions. The dashed box highlights the contrast between odd and even matras: the diagram suggests that odd beat positions should be stronger than even beat positions, although only if the 2-matra level is metricaly significant. The solid boxes highlight the vibhags, which are marked for emphasis in North Indian music theory, and indicated in practice by means of (optional) hand gestures: this would suggest that each of the solid-boxed matras – 1, 5, 9 and 13 – is metricaly stronger than other metrical positions.

338x60mm (96 x 96 DPI)

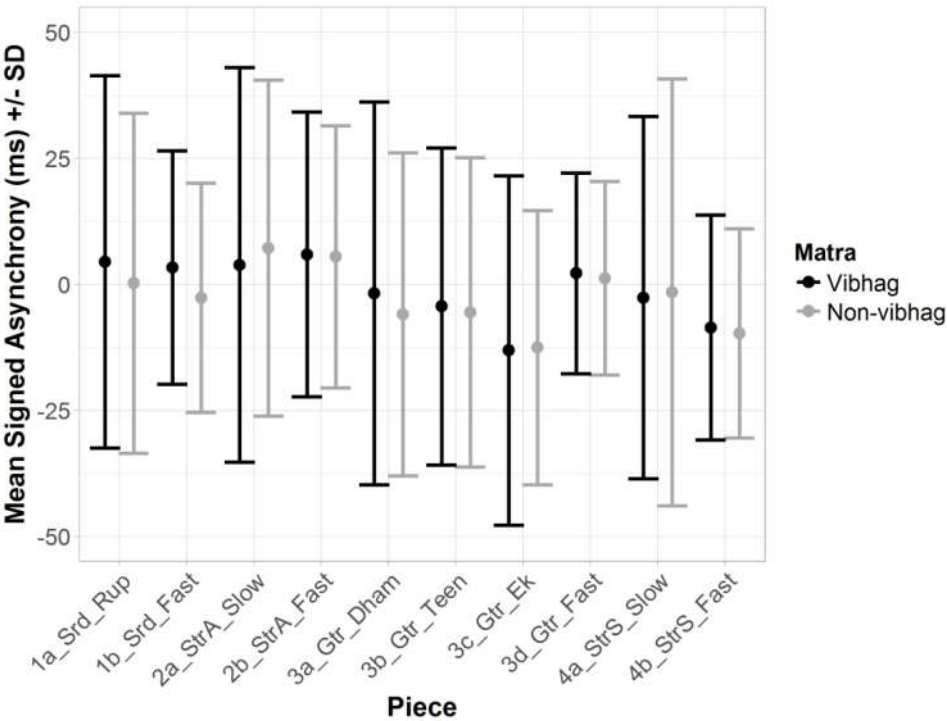


Fig. 6. Mean signed asynchrony (+/- standard deviation) for matras at the start of vibhags vs. sampled, non-vibhag matras.

203x152mm (300 x 300 DPI)

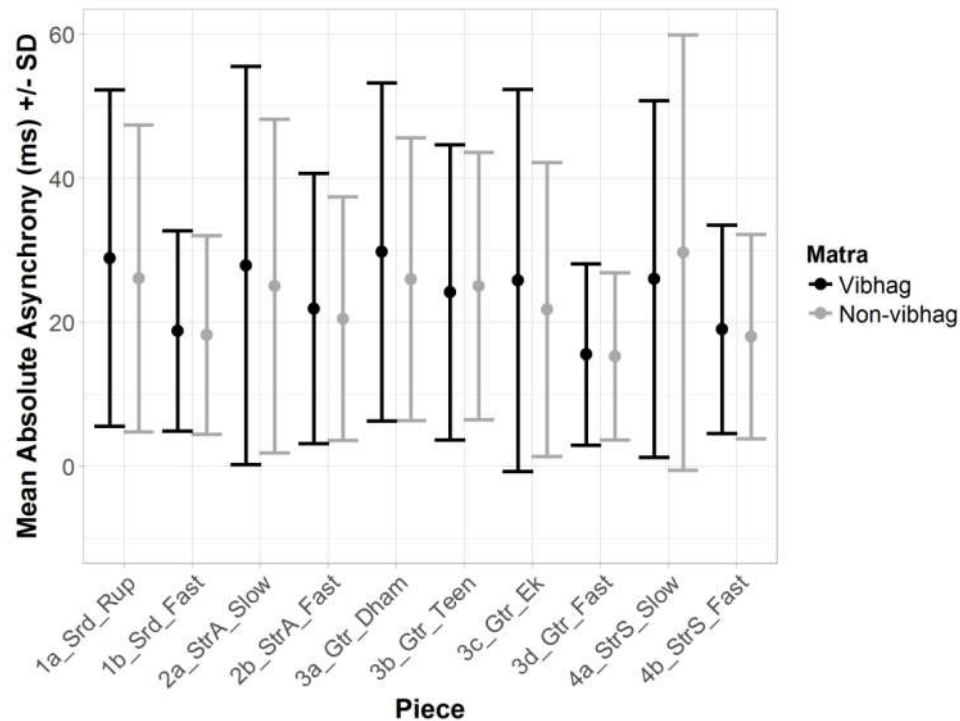


Fig. 7. Mean absolute asynchrony (+/- standard deviation) for matras at the start of vibhags vs. sampled, non-vibhag matras.

203x152mm (300 x 300 DPI)

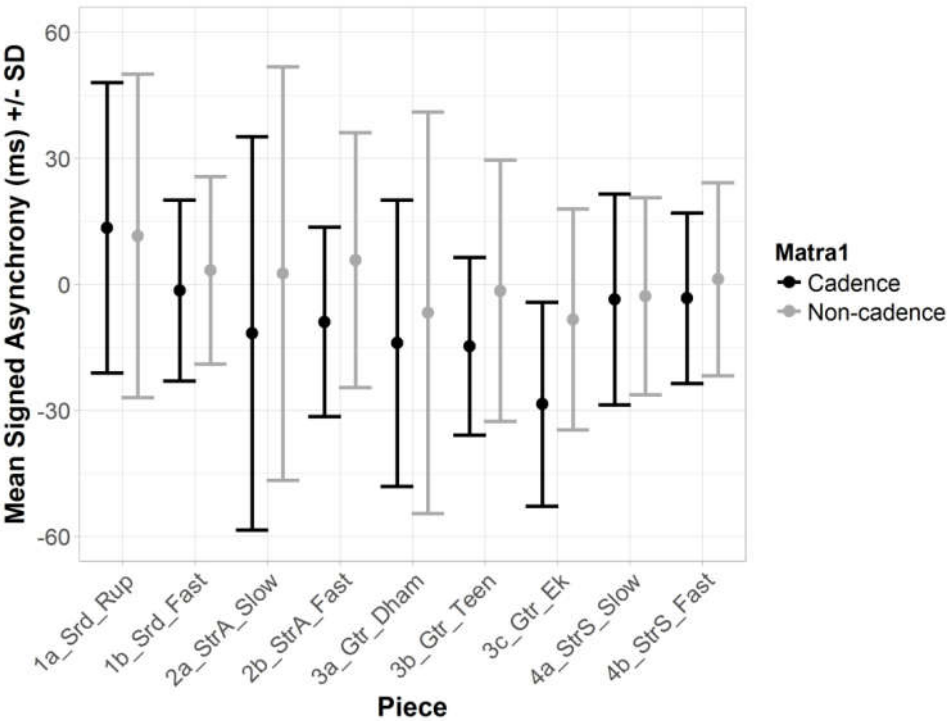


Fig. 8. Mean signed asynchrony (+/- standard deviation) for cadences on matra 1 vs. sampled, non-cadence instances of matra 1.

203x152mm (300 x 300 DPI)

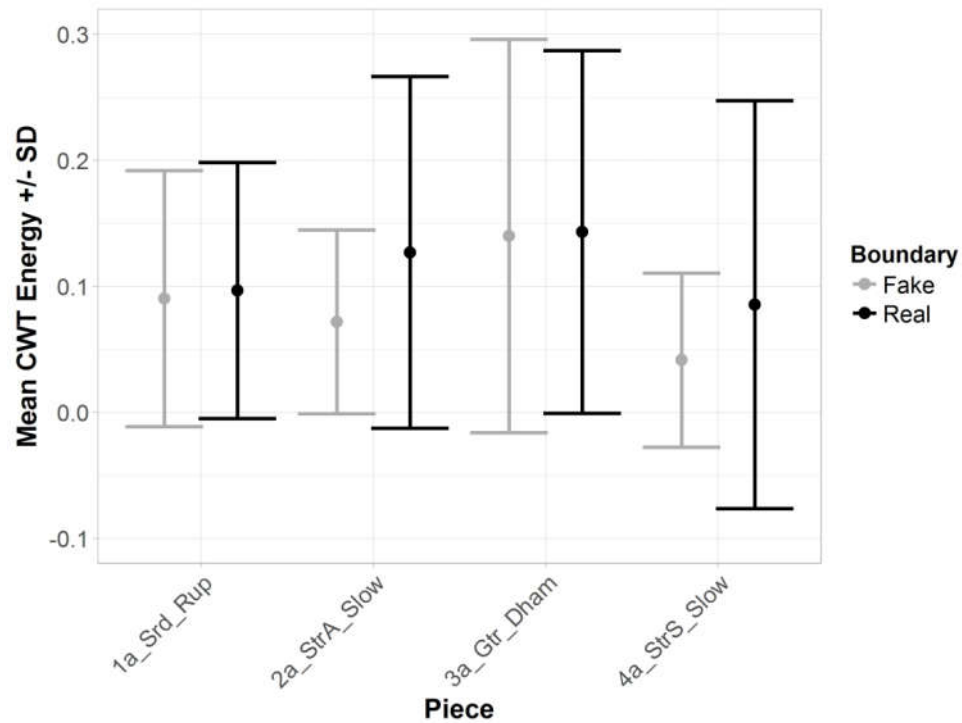


Fig 9. Mean CWT Energy ( $\pm$  standard deviation) for windows around real vs. 'fake' section boundaries by piece.

635x476mm (96 x 96 DPI)

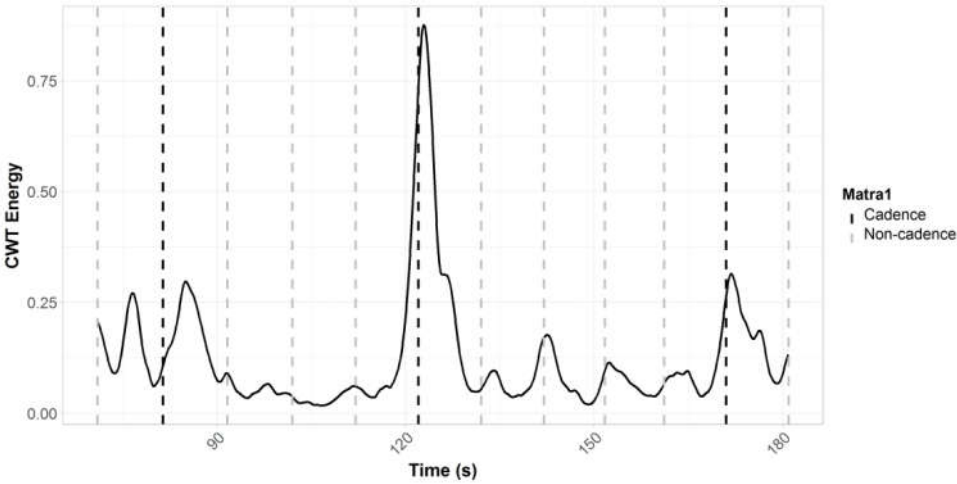


Fig. 10. Extract from 3a\_Gtr\_Dham, showing CWT Energy of both performers over time, with cadences and non-cadences (both occurring on matra 1) indicated with vertical lines.

952x476mm (96 x 96 DPI)

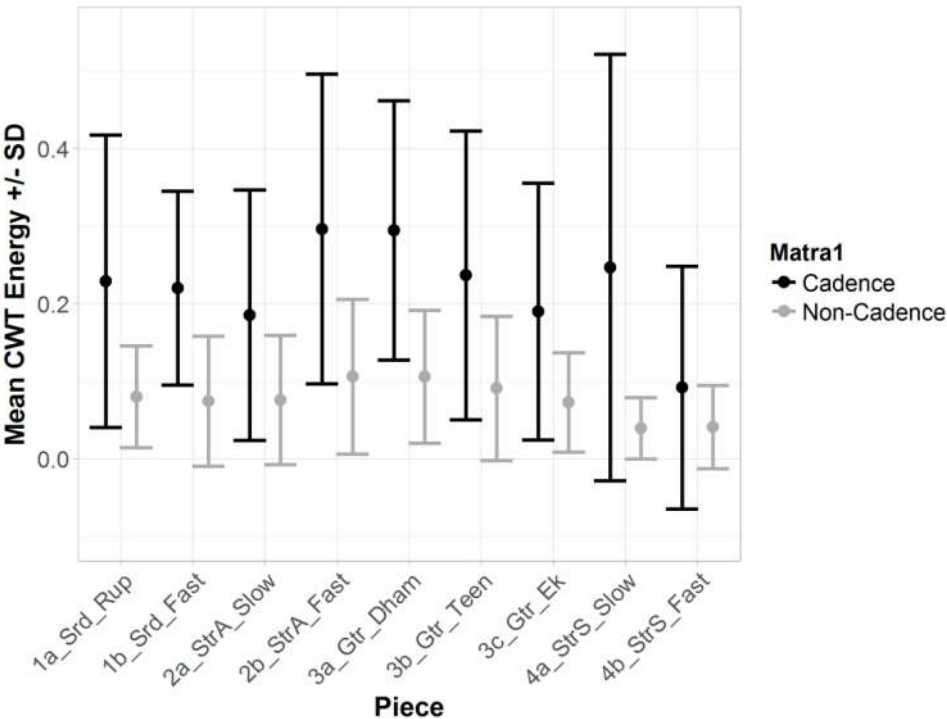


Fig. 11. Mean CWT Energy (+/- standard deviation) for cadences on matra 1 vs. sampled, non-cadence instances of matra 1.

635x476mm (96 x 96 DPI)

Code	Corpus ID	Artists	Piece	Duration (s)	Mean tempo (bpm)	Tempo range (bpm)	Onsets	Asynchronies	Mean signed asynchrony (ms)	Asynchronisation (ms)	No. cadences/No. cycles
1a Srd Rup	NIR_PrB_Jhinjhoti	Pratyush Banerjee (sard), Gauri Shankar Karmaakar (tabla)	Rag Jhinjhoti, rupak tal	792	91	70-100	3220	1189	0.72	33.43	18/165
1b Srd_Fast			Rag Jhinjhoti, drut teental	808	367	200-600	5490	1913	-0.99	23.03	13/139 before jhala
1c StrA_Slow	NIR_ABh_Puriya	Anupama Bhagwat (sitar), Gurdain Rayatt (tabla)	Rag Puriya, vilambit teental	1243	56	45-65	3364	1369	5.38	37.23	12/71
1d StrA_Fast			Rag Puriya, drut teental	977	374	250-625	8872	3593	5.59	26.71	17/190 before jhala
1e 3a Gtr_Dham	NIR_DbH_Malhar	Debashish Bhattacharya (guitar), Gurdain Rayatt (tabla)	Rag Miyan Malhar, dhamar tal	378	96	80-120	1177	385	-5.30	34.57	9/41
1f 3b Gtr_Teen			Rag Miyan Malhar, teental	563	166	140-180	1971	679	-5.25	30.89	19/95
1g 3c Gtr_Ek			Rag Miyan Malhar, drut ektal	397	277	240-320	2326	738	-12.74	29.71	11/149
1h 3d Gtr_Fast			Rag Miyan Malhar, drut teental	268	533	450-600	3008	867	1.59	19.47	5/19 before jhala
1i 4a StrS_Slow	NIR_ShKh_Jhinjhoti	Shujaat Khan (sitar), Shahbaz Hussein (tabla)	Rag Jhinjhoti, vilambit teental	727	66	40-65	1708	665	-2.51	39.20	9/38
1j 4b StrS_Fast			Rag Jhinjhoti, drut teental	1107	417	200-650	12069	5338	-9.43	21.21	25/114 before jhala

Appendix A. Corpus. Mean length of section = 775 secs, SD = 316 secs.



Teental  
16  
matras

Dha	Dhin	Dha	Dha	Dha	Dhin	Dha	Dhin	Tin	Ta	Tete	Dhin	Dha			
X				2		0				3					
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

Dhamar tal  
14  
matras

matrix													
Ke	Dhet	Ta	Dhet	Ta	Dha	Ga	Din	Na	Din	Na	Ta	–	
X					2	0			3				
1	2	3	4	5	6	7	8	9	10	11	12	13	14

Ektal  
12  
matras

Dhin	Dhin	Dhage	Tirakita	Tun	Na	Kat	Ta	Dhage	Tirakita	Dhin	Na
X		0		2		0		3		4	
1	2	3	4	5	6	7	8	9	10	11	12

Rupak  
tal  
7  
matras

Tin	Tin	Na	Dhin	Na	Dhin	Na
0			1		2	
1	2	3	4	5	6	7

Appendix B. Tala structures. In each case the first row comprises an archetypal drum pattern called a theka referred to above: written syllables iconically represent distinct drum sounds. The second row marks out the theoretical ‘vibhag’, or beat-grouping structure, and the third gives a count of the matras – the basic theoretical time units (translated as ‘beats’). Shaded parts are the ‘khali’ portions, which are vibhags marked gesturally with a silent wave and sonically by a predominance of drum strokes without bass resonance (Tin, Ta, Na, Kat etc).

<http://mc.manuscriptcentral.com/MSX>

1	2R	651	113.0	561.0	5.413	27.166	21.833	8.579	8.616	0.195	0.099
2	3a_Gtr_Dham	100	130.9	84.1	1.713	39.529	31.389	2.741	1.913	0.258	0.180
3	3B	131	123.7	91.0	-1.269	34.936	28.118	2.670	2.061	0.285	0.198
4	3C	154	122.0	105.6	-13.271	29.014	24.615	2.933	2.689	0.264	0.221
5	3b_Gtr_Teen	147	118.7	146.3	4.752	33.865	27.422	3.044	4.132	0.245	0.143
6	3E	40	102.8	156.3	0.547	28.386	21.579	2.434	2.574	0.194	0.081
7	3F	115	108.9	166.7	-11.475	31.143	26.697	3.520	3.363	0.247	0.113
8	3G	212	117.6	172.0	-7.805	28.840	23.919	4.397	4.101	0.200	0.157
9	3H	165	108.7	178.7	-7.962	29.122	23.126	4.456	5.238	0.201	0.205
10	3I	189	132.5	247.3	-16.179	31.485	24.819	3.676	5.399	0.216	0.143
11	3c_Gtr_Ek	267	132.6	271.3	-13.554	25.382	22.056	4.995	6.297	0.245	0.141
12	3J	282	131.1	309.0	-9.657	31.988	22.598	6.543	6.371	0.221	0.145
13	3K	470	133.2	468.4	3.516	18.520	15.141	5.576	7.549	0.206	0.136
14	3d_Gtr_Fast	395	134.9	582.2	-0.471	19.085	15.371	5.207	8.628	0.180	0.156
15	3M	29	108.2	44.5	6.440	24.594	17.304	4.440	5.102	0.097	0.162
16	4a_StrS_Slow	70	110.5	44.4	9.986	51.938	34.573	3.199	1.845	0.105	0.110
17	4C	115	123.6	47.4	-7.754	31.518	23.554	8.312	2.398	0.153	0.142
18	4D	125	113.0	50.8	-8.334	32.356	26.587	8.884	2.421	0.161	0.132
19	4E	133	121.8	55.5	-0.201	39.033	24.862	6.774	4.789	0.111	0.147
20	4F	193	143.6	61.1	-3.092	42.782	32.406	8.181	2.759	0.195	0.130
21	4b_StrS_Fast	379	122.1	231.7	-5.070	27.602	21.545	7.601	4.354	0.136	0.086
22	4H	388	123.7	274.0	-7.664	24.648	19.927	8.151	4.933	0.155	0.118
23	4I	352	123.1	301.1	0.983	23.738	16.441	6.860	6.248	0.097	0.144
24	4J	544	121.9	367.0	-10.784	23.132	20.434	8.936	5.640	0.162	0.112
25	4K	606	124.1	390.2	-12.263	16.905	17.072	8.327	5.530	0.140	0.073
26	4L	576	122.8	409.7	-10.177	19.238	17.485	8.630	5.506	0.132	0.068
27	4M	722	121.8	429.8	-8.992	17.236	15.138	9.574	6.612	0.177	0.092
28	4N	832	122.4	476.4	-13.551	20.220	20.234	9.968	7.567	0.216	0.132
29	4O	939	123.9	599.6	-9.391	19.799	17.675	10.587	8.885	0.262	0.169

Appendix C. Segment data.