Automated Video Analysis of Interpersonal Entrainment in Indian Music Performance

Paolo Alborno,
Gualtiero Volpe,
Antonio Camurri

DIBRIS
University of Genoa
16145, Genova, Italy
paoloalborno@gmail.com,
gualtiero.volpe@unige.it,
antonio.camurri@unige.it

Martin Clayton
Department of Music
Durham University
Durham, UK
martin.clayton@durham.ac.uk

Peter Keller
MARCS Institute
University of Western Sydney
Sydney, Australia
p.keller@uws.edu.au

***Abstract*— Interpersonal entrainment is an important and underexplored feature of musical performance. Better understanding of this phenomenon has important implications for musicology and for the understanding of joint action and coordination from a psychological perspective, and is likely also to be applicable in entertainment, educational, and therapeutic contexts. In order to achieve this, new methods of investigation are needed through which entrainment can be investigated in ecological contexts, keeping time-consuming manual intervention to a minimum. Preliminary research results are presented here on a system for the automated analysis of entrainment, and leadership from video recordings of performances of Indian music. The system is based on the EyesWeb XMI open software platform. The context and aims of the research are discussed along with an analysis of entrainment on the basis of head movements tracked on a video recording.**

Keywords: interpersonal entrainment; synchronisation; ethnomusicology;

I. Introduction

Interpersonal entrainment – that is, rhythmical coordination and synchronisation of movement between two or more individuals – is an important and underexplored feature of musical performance. Better understanding of this phenomenon has important implications for musicology, because it can shed light on the emergence and stability of musical structures and the reasons for their efficacy, and for the understanding of joint action and coordination from a psychological perspective. Knowledge of interpersonal entrainment is likely also to be applicable in a range of entertainment, educational, and therapeutic contexts [1] [21] [24].

Although entrainment in music has been demonstrated and some of its features uncovered, much remains to be discovered regarding the range of entrained behaviours experienced in music performance of various kinds. Existing methods, while effective, suffer some limitations. Studies dependent on detailed expert mark-up of either significant musical events or movements are time-consuming, and of limited effectiveness in musical genres where periodic sound-producing actions are not easily identified; lab-based studies using motion capture technologies may reduce some of the manual element, but suffer from a lack of ecological validity. In order to progress study in this area effectively, therefore, new methods of investigation are needed through which entrainment can be investigated in ecological contexts, keeping time-consuming manual intervention to a minimum.

This motivates the current study, which takes an important step towards automated analysis of interpersonal entrainment in a form of music performance in which identifying periodic musical actions can be problematic, namely north Indian classical vocal (and vocal-style instrumental) music. Manual mark-up is not completely eliminated, but we nonetheless demonstrate the potential to investigate entrainment using only audiovisual recordings of real-life performances, algorithms for extracting visual information and statistical techniques. The musicological efficacy of this approach is demonstrated by testing a hypothesis regarding the relationship between synchronisation and metrical structure in the performance of this music at slow tempi.

II. Background

Interpersonal entrainment was described by Condon in the 1960s on the basis of observational studies of filmed conversations. Condon's original investigations involved the segmentation of behaviour, both speech and movement. He posited that both intrapersonal synchrony (e.g. a speaker's speech with his or her own body movement) and interpersonal synchrony (e.g. a listener's body movements with another's speech) were ubiquitous in interaction, and their absence was a sign of pathology [8] [24].

INTETAIN 2015, June 10-12, Torino, Italy

DOI 10.4108/icst.intetain.2015.259521

More recent studies have used more advanced methods for both data collection (e.g., motion tracking) and data analysis. They have confirmed that interpersonal entrainment does occur spontaneously when two individuals attend to each other in conversation or in carrying out a joint task, and is strongly linked to visual attention [22][23][17]. An increasing body of research has also investigated the relationship between this spontaneous interpersonal entrainment and social factors such as affiliation and prosocial behaviour [12].

In ethnomusicology, the topic relates to historic theories that communal singing and dancing are crucial mechanisms by which 'fellow-feeling' is facilitated and social bonds strengthened: although such arguments for the importance of music have been vigorously promoted by important figures in the field [13], empirical support has been hard to come [5].

A variety of methods has been employed to study interpersonal entrainment. For instance, the 'stroboscopic' method involves the identification of a reference point in quasi-periodic behaviour, such as the moment at which a limb movement changes direction: phase relationships between two such periodic behaviours are calculated and stabilisation of this phase relationship (in particular following a disturbance) is taken as strong evidence for entrainment [18][19]. Where such reference points cannot be determined, or where patterns of behaviour are too complex to characterise as 'quasi-periodic', other methods can be employed to investigate entrainment, including cross-recurrence quantification and event synchronisation (e.g., see [25] as an example of how such methods were applied to analysis of entrainment in music performance).

III. Entrainment and Interpersonal Synchrony in Music Performance

Empirical studies of music performance have begun to elucidate the dynamics of interpersonal entrainment. Clayton studied the emergence, in an Indian music concert, of entrainment between individual parts that were explicitly intended to be uncoordinated [7], while Lucas et al [15] demonstrated the emergence of entrainment between distinct groups of musicians and dancers in the Afro-Brazilian ritual Congado, which again was clearly not intended. Doffman's studies of jazz trios demonstrated, by relating interview data to analyses of entrainment between musicians in performance, that a degree of fluidity in interpersonal synchrony – an alternation between tighter and looser synchronisation – is valued by players [10][11], while Poole's study of Cuban dance bands illustrates a regular alternation between looser and more precise alignment within the measure [20]. Yoshida et al have referred to this phenomenon as 'soft entrainment' [27]. Each of these studies used the stroboscopic method described above, applied to data derived from recorded audio and/or visual information.

Other methods used to study interpersonal entrainment in music and dance include correlation [9] [16]. For example, the cross-correlation of co-performers' body sway motion has been used to estimate the degree of interpersonal coordination and leader-follower relations in piano duos [14]. This is the approach taken in the present study.

These studies have opened a window on the richness and complexity of interpersonal entrainment in music performance: for instance, the ways in which some genres of music involve resistance to entrainment, between individuals or between groups; the great variety in terms of how tightly synchronised a successful performance should be, and whether the degree of coordination is (and should be) varied from beat to beat or over longer time scales; and the ways in which entrainment sheds light on leadership dynamics within ensembles. Further exploration of such topics is clearly necessary before confidently generalising about interpersonal synchrony in expressive performance, let alone claim to have fully explored its relationship to such issues as the development of affiliation and social identities.

IV. Entrainment in Indian Music Performance

As noted above, Clayton's 2007 [7] study explored unintended entrainment in a Hindustani (North Indian classical) music performance. In this study, manual annotation of a video recording was used to determine the periodic nature of the plucking of long-necked lutes called tanpura, which could not be heard or extracted from the sound recordings. In a related study, Will et al [26] have explored listeners' tapping responses to theoretically unmetred sections of Hindustani music performance. They found that in the introductory alap section (in which it is often stated no pulse should be felt), listeners' responses were largely determined by their own endogenous rhythms – as elicited in the form of spontaneous tapping – which interacted weakly with the temporal structure of the auditory stimulus; in the later jor section (which is supposed to be 'pulsed'), listeners' responses align much more closely with the periodic structure of the stimulus. These studies have begun to clarify the role of entrainment in both production and perception of theoretically unmetred parts of Indian music performance.

Paradoxically, metred sections of Hindustani music performances – those set to one of the theoretically-recognised talas, or metrical patterns [6] – have proved harder to investigate. In these performances there can be no doubt that ensemble members share an understanding of the tala or temporal structure and orient themselves to that structure: they are doubtless mutually entrained. The structure and style of the music, however, makes it difficult to explore the dynamics of that entrainment. Typically, a tabla player (drummer) plays variations of a standard pattern (theka) that references the tala structure. Vocal or instrumental soloists have considerable freedom within that structure: they do not need to play on the beat to demonstrate that they are playing in time. Moreover, in classical singing, or performance on bowed or blown instruments which is closely modelled on singing, notes often have no clearly identifiable onset (indeed, blending the sound so that an attack or release point is not identifiable is valued aesthetically). Synchrony between musicians is demonstrated by the fact that a soloist will either return to a refrain just before beat one (sam) in the tala, and/or will conclude an improvisation at this point. In between these points it may be regulated by a soloist's hand gestures (used to direct the drummer to adjust the tempo), or affirmed by gestures such as synchronised head nods on the sam (beat one).

Interpersonal synchrony between soloist and drummer is most easily identified, then, at the sam. In a slow composition the tala cycle might extend for as long as a minute: what happens to the synchrony over this time span? Are they tightly synchronised throughout, via their shared understanding of the tala structure? The jazz and Cuban studies cited above have shown how in much shorter metrical cycles the tightness of the coordination can ebb and flow from moment to moment: we might hypothesise, then, that the synchrony between Indian classical singer and tabla accompanist is relatively loose over the course of a long tala cycle, becoming more precise in the approach to the sam[1]. Can this be demonstrated empirically by analysing the body movement of musicians? If so, when do musicians start to synchronise more tightly? Is there a particular point in the cycle when this starts to happen, or can it be cued at any point, for instance by the musicians making eye contact?

These are some of the issues that need to be investigated in Hindustani music performance. This work aims to do so, moreover, using audio-visual recordings made in actual public performances, in which no constraints were placed on performers and no special motion capture markers used. (The only possible limit on ecological validity in such cases is the presence of video cameras: this is unlikely to be disturbing to professional musicians however, who are used to being videoed in concert.) Is it possible to extract detailed movement information from such recordings, without laborious manual coding, and to do so robustly enough to enable empirical studies of entrainment and leadership?

V. Murad Ali and Gurdain Rayatt Performance

A considerable archive of digital audio-visual recordings of Hindustani performances is held at Durham University. Most comprise multiple view video recordings using locked-down cameras (no panning or zooming), and multitrack audio. The recording used for this study was made at a public concert promoted by Gem Arts at the Sage Gateshead on 24th September 2012, recorded by a team from Durham University using two HD video cameras (Canon XF305), with separate audio tracks fed via the mixing desk to a Tascam HS-P82 recorder.

The soloist, Murad Ali (MA), is highly-regarded sarangi (bowed lute) player from India; his tabla accompanist, Gurdain Rayatt (GR), is a young UK-based musician. The sarangi was for a long time the favoured melodic accompaniment instrument for Hindustani vocal performance. It became established as a solo concert instrument in the latter half of the twentieth century: when used as a solo instrument, as here, its repertory and style remain very closely modelled on vocal style. Reasons for the choice of this recording for a first trial were the good quality of both performance and recording, and the use of the sarangi, which means that vocal-style music can be studied in a duo format (without additional melodic accompaniment).

The item research focuses on here is a performance of the raga (melodic mode) Shyam Kalyan. The thirty minute performance has 4 sections:

- **S1**: 0:00-4:30: *Alap*. MA on unaccompanied sarangi.
- **S2**: 04:30-20:10: *Slow Ektal.* 12 beats (matras), the cycle lasts about 44s at the beginning, accelerating to c. 35s by the end.
- **S3**: 20:10-26:15: *Fast Teental.* 16 beats, cycle lasts c. 2.5s.
- **S4**: 26:15-30:50: *Fast Ektal*. 12 beats, cycle c. 2.5s.

Every slow ektal cycle has been considered for the analysis and divided into three different parts: *preparation*, *development* and *conclusion*. This is a novel division: it reflects the fact that the soloist typically fills the cycle with one coherent episode of improvisation, in the course of which a new idea is presented, then further developed, before he starts to orient himself towards the conclusion of the cycle. The division into equal thirds is arbitrary.

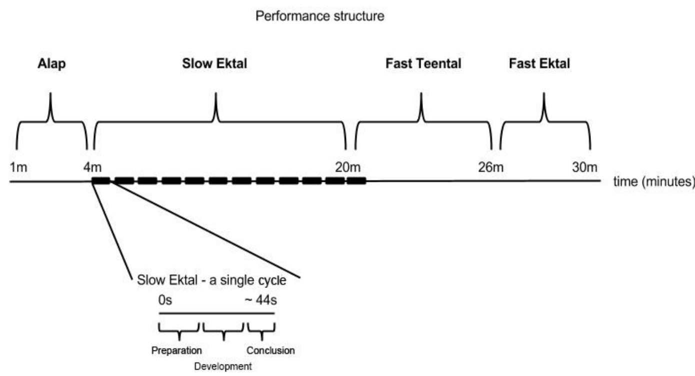

Figure 1. The structure of the musical performance

In order to refine questions for investigation here, the frontal video view was watched several times, the beats of the tala cycles labelled as a reference (tapping the beats in *Sonic Visualiser*), and initial observations noted. Since it is impossible to visually identify many periodic movements – other than the recurrence of head nods at the start of each cycle in the slow ektal section – the following observations are largely qualitative in nature:

a. In the alap (S1) there is no visible coordination between MA (who plays) and GR (who does not), although the latter does maintain a respectful and attentive demeanour.

b. In the slow ektal portion (S2), most occurrences of sam (beat one) are clearly marked by shared head movements by the two musicians. Other than this there is little obvious gestural communication between the two.

c. Looking at the joint head nods in the slow ektal (S2) in more detail, sometimes the musicians look towards each other, making eye contact; at others both musicians look ahead at the audience. Sometimes there seem to be clear preparatory movements: either

[1] To be precise, the basic durational unit of the tala cycle should be referred as a *matra*: the term is loosely translated as 'beat' here for ease of reading.

a nodding of MA's head in the last beat or so, or a joint head-nod on beat 9; sometimes MA looks to GR and makes eye contact just before the nod. It was noticed that after periods in which MA has been giving a lot of visual attention to a section of the audience, he deliberately reconnects with GR in this way (see e.g. cycle 15, s 769-799).

d. In cycle 20 (s 968-1001) the musicians make a mistake, adding an extra beat to the cycle. Their movements suggest a lack of mutual attention here, although it is not clear what caused the error. (As is usually the case after such mistakes, the musicians make no obvious reference to the fact either to each other or to the audience).

e. In contrast to the slow ektal, in the two fast sections (S3, S4) the joint head nods on sam are only occasional. However, much more miscellaneous gestural communication occurs: in particular, instructions from MA to GR, mostly to either adjust the tempo (he asks GR to speed up at least 5 times) or to take a solo (4 times).

f. Transitions between sections are obviously significant in terms of coordination. Observation of these points (c. 20-21 mins and 26-26:30) does not reveal much movement coordination, however: what seems to happen is that MA stops one section, then after a few sections begins the next, and GR starts up when he has picked up the new tala pattern (during the transition to S3 MA also briefly retunes his instrument).

g. There is nothing in the video to suggest that MA's leadership is being contested by GR. The soloist role and the fact that MA is a touring artist and GR a local accompanist mean that the musical 'seniority' clearly lies with the former. When GR takes solos, for instance, this is clearly at the invitation of MA.

h. It was also noticed in a few occasions that, although there is no eye contact or shared gesture, MA is clearly working on a rhythmic figure while concentrating hard on the tabla.

There are various aspects of interpersonal entrainment and leadership that could be investigated. For the purposes of this work in progress paper, research concentrated on analysing the relationship between the two musicians in S2 (the slow ektal portion) between the three phases: the preparation (beats 1-4), development (5-8) and conclusion (9-12); based on the observations above, the hypothesis here is that the relationship will be closest in the last third of the cycle.

VI. Conceptual Framework and System Architecture

In order to perform the analysis a system was developed for the automated extraction of synchronisation and entrainment cues from the movements of the two musicians. In particular the analysis focused on the head movements of both musicians.

The system is derived from the multimodal framework developed by Camurri and colleagues [3] [4].

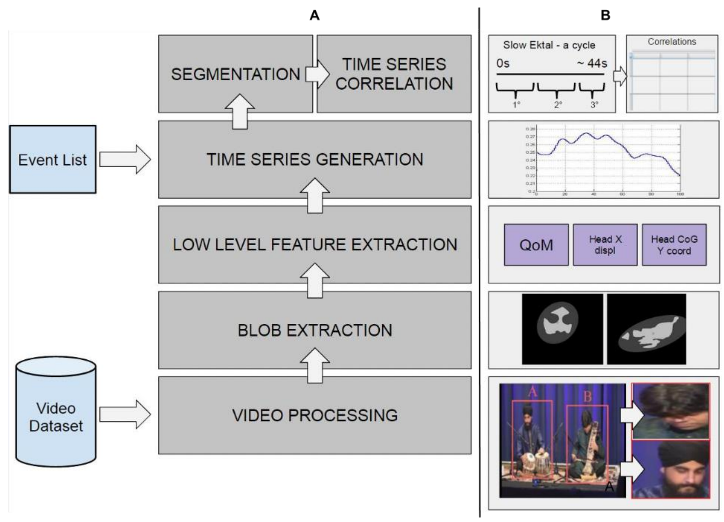

Figure 2. The system architecture for automatic feature extraction from music performance recordings.

.

Figure 2 shows an overview of the system architecture and its major layers. The main building blocks, or modules, are described below. The architecture includes low-level (highly synchronised) layers (e.g. Video Processing) up to higher-level layers (e.g. Time Series Correlation), where asynchronous events are typically detected.

Layer description:

Video processing: the system accepts as input a video of a music performance (fixed camera view, high quality colour video). This low-level video analysis layer consists of the automated tracking of the regions of interests (ROIs), i.e. the minimum rectangles containing the blobs corresponding to the locations of musician's heads. A blob is defined as a cluster of pixels extracted from the ROI characterised by invariant properties (i.e. pixels that are part of the same cluster vary within a statistically coherent range of values). In this case, two ROIs are identified, corresponding to the two musicians' heads. Once identified, a video binarization is applied.

Blob extracting: the output of this layer is a binarized image containing blobs rendered as white textures. The extracted blobs correspond to the heads of the two performers.

Low-level feature extraction: low-level features are extracted from blobs. Computer vision techniques are applied to compute features such as velocities, accelerations, energy, contraction, directness, impulsivity and so on. In this case, three features of interest were identified:

- **Quantity of Motion**: the amount of movement detected by the video camera and computed as the pixel-based difference between two consecutive frames containing head blobs. A mathematical description is available in [3] [4].
- **Y coordinate of the head Centre of Gravity (CoG)**: the blob's barycentre coordinate on the vertical axis. This feature was used to analyse joint head movements (nods) that typically characterise the end of musical (tala) cycles.
- **Head X displacement**: the overall translation and rotation components of the head movements, computed from optical flow techniques applied to the head of each musician. This measure can be useful for approximating the gaze direction, to identify whether a musician is looking for eye contact during the performance.

Time Series Generation: the values of the features are logged in a set of files as time series. Time series are synchronised at a fine-grained level, which is enabled by the EyesWeb platform for multimodal synchronisation and analysis, developed in the SIEMPRE EU ICT FET Project (www.siempre.infomus.org). This layer includes the definition of the important events, and the creation of records of events occurrence and their timings.

Segmentation: performances are manually segmented into different sections, corresponding to sections in the musical structure. The slow ektal portion is segmented into 22 cycles (average duration is between 35 and 44 s).

Time Series Correlation: the Pearson's correlation coefficient between the extracted time series of each measured signal is computed. Statistical analysis is performed on the correlation values.

VII. System Implementation

The system was implemented using the EyesWeb platform (www.infomus.org/eyesweb). Each software module (corresponding to each layer) was modelled as an EyesWeb patch, using the algorithms included in the EyesWeb Video Analysis and Social Signals software libraries.

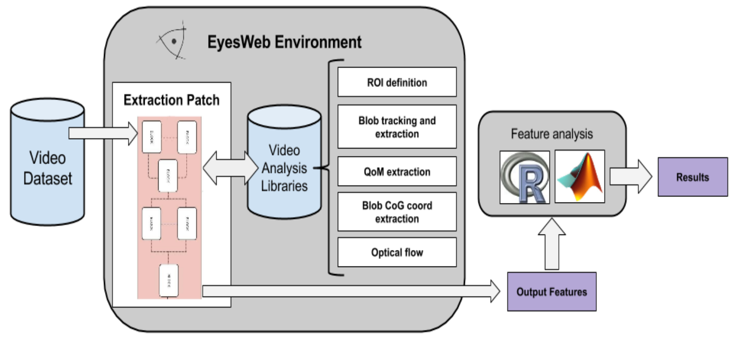

Figure 3. The EyesWeb XMI platform was adopted to perform tracking and analysis. The study of correlation between time series and statistical analysis was performed in Matlab and R.

VIII. Results

As a first step to assess synchronisation between the movements of the two musicians, the Pearson's linear correlation coefficient ρ was computed on the time series of each extracted movement feature for each of the three parts of each slow ektal cycle (preparation, development, and conclusion).

The final part of each cycle (conclusion) proved to be the one displaying the highest synchronisation between the two musicians. Table I reports mean ($\bar{\rho}$) and standard deviation (σ_ρ) of the correlation coefficient computed for each movement feature.

TABLE I. Mean

Motion Feature	$\bar{\rho}$		
	Preparation	***Development***	***Conclusion***
Quantity of Motion	.006	.036	.391
Y coordinate of CoG	.018	.058	.325
Head X displacement	.007	.046	.028

TABLE II. Standard Deviations

Motion Feature	σ_ρ		
	Preparation	***Development***	***Conclusion***
Quantity of Motion	.125	.204.	.289
Y coordinate of CoG	.103	.186	.192
Head X displacement	.218	.268	.264

A one-way repeated-measures ANOVA was conducted to compare the effect of every cycle part (*preparation*, *development*, and *conclusion*) on synchronisation for each feature. ANOVA hypotheses were checked with the commonly used Shapiro-Wilk and Mauchly's tests.

Quantity of Motion: data did not deviate from a normal distribution (Shapiro-Wilk test). Mauchly's Test of Sphericity indicated that the assumption of sphericity was not violated ($p = .302$). The effect of cycle part on synchronisation was significant ($F(2,42)= 27.73$, $p < 10^{-7}$). Three paired samples t-tests were used to make post hoc comparisons between conditions. A Bonferroni correction was applied to account for multiple comparisons. The first paired samples t-test indicated that there was no significant difference in synchronisation for the preparation part (M=-.006, SD=.125) and the development part (M=.036, SD=.204), p=.999. The second t-test indicated that there was a significant difference in synchronisation for the preparation part (M=-.006, SD=.125) and the conclusion (M=.391, SD =.289), $p < 10^{-4}$. The third t-test indicated that there was a significant difference in synchronisation for the development part (M=.036, SD =.204) and the conclusion (M=.391, SD=.289), $p < 10^{-4}$.

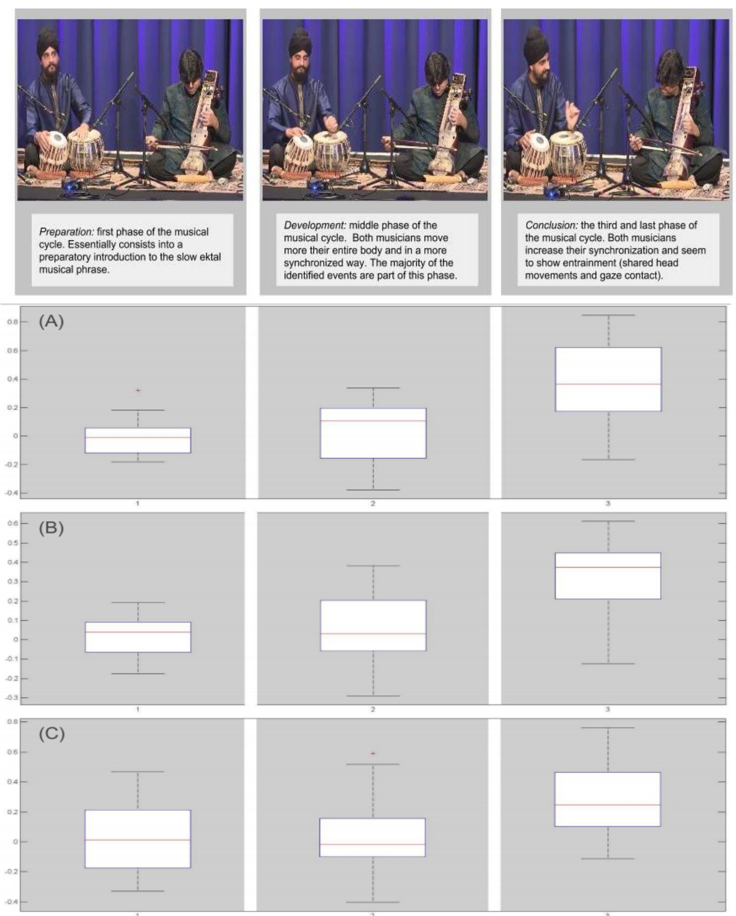

Figure 4: Boxplots relative to the three parts of a cycle. (A) Quantity of Motion. (B) Y coordinate of CoG. (C) Head X displacement.

Y coordinates of head's center of gravity: data did not deviate from a normal distribution (Shapiro-Wilk test). Since Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated ($p = .004$), a Greenhouse-Geisser correction was used ($\varepsilon = .707$). The effect of cycle part on synchronisation was significant ($F(1.41,29.69)=25.17$, $p<10^{-5}$). Three paired samples t-tests were used to make post hoc comparisons between conditions. A Bonferroni correction was applied to account for multiple comparisons. The first paired samples t-test indicated that there was no significant difference in synchronisation for the preparation part (M=.018, SD=.103) and the development part (M=.058, SD=.186), p=.908. The second t-test indicated that there was a significant difference in synchronisation for the preparation part (M=.018, SD=.103) and the conclusion (M=.325, SD =.192), $p < 10^{-6}$. The third t-test indicated that there was a significant difference in synchronisation for the development part (M=.058, SD =.186) and the conclusion (M=.325, SD=.192), p = 0.0007.

Head X displacement: data did not deviate from a normal distribution (Shapiro-Wilk test). Mauchly's Test of Sphericity indicated that the assumption of sphericity was not violated ($p = .5673$). The effect of beat part on synchronisation was significant ($F(2,42) = 9,753$, $p < 10^{-4}$). Three paired samples t-tests were used to make post hoc comparisons between conditions. A Bonferroni correction was applied to account for multiple comparisons. The first paired samples t-test indicated that there was no significant difference in synchronisation for the preparation part (M=.007, SD=.218) and the development part (M=.046, SD=.268), p=.870. The second t-test indicated that there was a significant difference in synchronisation for the preparation part (M=.007, SD=.218) and the conclusion (M=.028, SD =.264), p = 0.0011. The third t-test indicated that there was a significant difference in synchronisation for the development part (M=.046, SD=.268) and the conclusion (M=.028, SD =.264), p = 0.0138.

IX. Discussion and Future Work

Table I and II and Figure 4 summarise some preliminary results on the analysis of entrainment.

A key result is the confirmation of the main hypothesis: data analysis reveals how the synchrony between Indian sarangsoloist MA and tabla accompanist GR is relatively loose over the course of a long tala cycle, becoming more precise in the approach to the sam. This is consistent with the findings in [20][27] and related findings in other recent literature.

The system for extraction of features from video recordings of music performances, derived from the EyesWeb SIEMPRE Platform for multimodal synchronised recordings and analysis, was demonstrated to be flexible and suitable for use with a dataset recorded in an ecological settings typical of the ethnomusicology field studies here considered. The system therefore provides a promising tool to support the analysis of interpersonal entrainment between musicians on the basis of observational studies of recordings.

Future work will include further analysis of large ethnomusicology video repositories, such as the Durham University collection, and more sophisticated analyses based on the EyesWeb Social Signals Processing library [25], using techniques such as Event Synchronisation and Recurrence

Quantification Analysis. The system will also be extended to fully support analysis of leadership.

Analysis of entrainment and leadership can be exploited in many application fields, including interactive entertainment, serious games, and therapy and rehabilitation.

ACKNOWLEDGMENTS

This work was partially supported by the EU H2020 ICT DANCE Project no.645553.

REFERENCES

[1] A. Behrends, S.Müller & I. Dziobek, (2012). Moving in and out of synchrony: A concept for a new intervention fostering empathy through interactional movement and dance. The Arts in Psychotherapy, *39*(2), 107-116.

[2] J. Blacking, (1977). Towards an anthropology of the body. In J. Blacking, ed. The Anthropology of the Body. London: Academic Press, pp.1-28.

[3] A. Camurri, B. Mazzarino, M. Ricchetti, R.Timmers, and G. Volpe (2004). Multimodal analysis of expressive gesture in music and dance performances. Lecture Notes in Artificial Intelligence, 2915:20–39.

[4] A. Camurri, G. Volpe, G. De Poli, and M. Leman. (2005). Communicating expressiveness and affect in multimodal interactive systems. Multimedia, IEEE, 12(1):43 – 53, Jan.- 2005.

[5] M. Clayton, R Sager and U.Will, (2005). In time with the music: The concept of entrainment and its significance for ethnomusicology. EME, 11, ESEM-Counterpoint Series.

[6] M. Clayton, (2000) Time in Indian Music: rhythm, metre and form in north Indian rag performance. Oxford: Oxford University Press.

[7] M. Clayton (2007). Observing entrainment in music performance: Video-based observational analysis of Indian musicians' tanpura playing and beat marking. Musicae Scientiae 11 (1), 27-59.

[8] W. S. Condon and W.D. Ogston, (1966). Sound film analysis of normal and pathological behaviour patterns. Journal of Nervous and Mental Disease, 143, 338–347.

[9] M. Demey, C. Muller, and M. Leman (2013). DanSync: A Platform to Study Entrainment and Joint-Action during Spontaneous Dance in the Context of a Social Music Game. In Proceedings 5th International ICST Conference on Intelligent Technologies for Interactive Entertainment, INTETAIN 2013, Mons, Belgium.

[10] M. Doffman, (2009). Making it groove! Entrainment, participation and discrepancy in the 'conversation' of a jazz trio, Language and History 52 (1), 130-147.

[11] M. Doffman, (2013). Groove: temporality, awareness and the feeling of entrainment in jazz performance, in M. Clayton, B. Dueck and L. Leante (Eds.) Experience and Meaning in Music Performance. Oxford: Oxford University Press, pp.62-85.

[12] E. Goodman, (2002). Ensemble performance. In J. Rink (Ed.), Musical performance: A guide to understanding (pp. 153-167). Cambridge, UK: Cambridge University Press.

[13] C. Keil, (1994 [1987]). Participatory discrepancies and the power of music. In Charles Keil and Steven Feld , *Music Grooves*, pp. 9 6–108. Chicago: University of Chicago Press.

[14] P. Keller, & M. Appel, (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. Music Perception, 28, 27-46.

[15] G. Lucas, M. Clayton, and L. Leante, (2011). Inter-group entrainment in Afro-Brazilian Congado ritual. Empirical Musicology Review 6 (2), 75-102.

[16] G. Luck. and P. Toivianen, (2006). Ensamble Musicians' Synchronization with Conductors' Gestures: An Automated Feature-Extraction Analysis. Music Perception, 4(2), 189-200.

[17] K. L. Marsh, M. J.Richardson, R. C. Schmidt, (2009). Social connection through joint action and interpersonal coordination. Topics in Cognitive Science, 1 (2), 320-339.

[18] E. Pereda, R.Q. Quiroga, and J. Bhattacharya, (2005). Nonlinear multivariate analysis of neurophysiological signals. Progress in Neurobiology 77(1-2), 1-37.

[19] A. Pikovsky, M. Rosenblum, and J. Kurths (2001). Synchronization. A Universal Concept in Nonlinear Science. Cambridge: Cambridge University Press.

[20] A. Poole, (2012). The Aesthetics of Groove in Cuban Dance Music: An Analysis of Son and Salsa. PhD Dissertation, Open University.

[21] T. C. Rabinowitch, I. Cross, & P. Burnard, (2013). Long-term musical group interaction has a positive influence on empathy in children. Psychology of Music, 41, 484-498.

[22] M. J. Richardson, K. L. Marsh. and R. C. Schimdt, (2005). Effects of visual and verbal interaction on unintentional interpersonal coordination. Journal of Experimental Psychology: Human Perception and Performance, 31 (1): 62-79.

[23] Schmidt, R. C., and O'Brien, B. (1997). Evaluating the dynamics of unintended interpersonal coordination. Ecological Psychology, 9 (3): 189-206.

[24] M. Thaut, G. C. McIntosh, & V. Hoemberg, (2015). Neurobiological foundations of neurologic music therapy: Rhythmic entrainment and the motor system. Frontiers in Psychology, 5.

[25] G. Varni, G. Volpe, and A. Camurri (2010). A System for Real-Time Multimodal Analysis of Nonverbal Affective Social Interaction in User-Centric Media. IEEE Transactions on Multimedia, vol. 12, n. 6, October 2010.

[26] U. Will, M. Clayton, I. Wertheim, L. Leante, and E. Berg, (2015). Pulse and entrainment to non-isochronous auditory stimuli: The case of North Indian alap. Plos ONE. 10(4): e0123247. doi:10.1371/journal.pone.0123247

[27] T. Yoshida, S. Takeda, and S. Yamamoto, (2002). The application of entrainment to musical ensembles. ICMAI 2002.