Instrumenting the Interaction: Affective and Psychophysiological Features of Live Collaborative Musical Improvisation

Evan Morgan
Queen Mary University of London
Mile End Rd
London E1 4NS
e.l.morgan@qmul.ac.uk

Hatice Gunes
Queen Mary University of London
Mile End Rd
London E1 4NS
h.gunes@qmul.ac.uk

Nick Bryan-Kinns
Queen Mary University of London
Mile End Rd
London E1 4NS
n.bryan-kinns@qmul.ac.uk

ABSTRACT

New technologies have led to the design of exciting interfaces for collaborative music making. However we still have very little understanding of the underlying affective and communicative processes that occur during such interactions. We carried out a study where we collected both self-report and *continuous* behavioural, and physiological measures from pairs of improvising drummers. Correlations were found between self-report scores and continuous measures. Absence of visual contact between participants was also shown to affect some of these measures. We discuss how our findings could influence the design of enhanced, collaborative interfaces for musical creativity and expression.

Keywords

Psychophysiology, affect, improvisation, music, creativity.

1. INTRODUCTION

Advances in wireless communication, touch screen sensing, and motion recognition have greatly facilitated the design and development of exciting new interfaces for collaborative music making. Researchers readily utilise such interfaces as a means evaluating designs and investigating the nature of joint music composition. This is a valuable approach, however it is inherently centred around the affordances and restrictions of the technology, as opposed to the sensitivities and needs of the users. We still lack a good understanding of the basic communicative and affective processes which accompany collaborative music making. To investigate this we carried out a study in which we asked pairs of experienced drummers to perform improvised drum beats, with and without visual contact. During the performances we collected physiological, behavioural, and MIDI data, as well as post-performance subjective reports.

2. RELATED WORK

Our work is influenced by theories, models and tools that are drawn from research in musical interactions, group creativity, affect recognition, and psychophysiology. For clarity we separate our review of existing research according to these four distinct fields, however it is common to find some overlap between research in these areas.

Musical Interactions: Mutual engagement is an important feature of multi-user musical interactions. Bryan-Kinns identifies five important design features for supporting mutual engagement: i) mutual awareness of action; ii) annotation, iii) shared and consistent representations, iv) mutual modifiability, and v) spatial organisation [2]. Motion tracking studies involving groups of string musicians have shown that head movement features can indicate levels of engagement [12] as well as complex interaction patterns and rhythmic synchronisation [15].

Rhythmic interaction has also been studied in relation to audio and visual coupling. Konvalinka et al. [22] looked at mutual prediction and adaptation during joint tapping experiments. They found that when both participants could hear each other they continuously adapted to each other's millisecond beat timings, such that no leader-follower relationship emerged. Vera et al. [33] studied the effect of line-of-sight on the precise note timings of a string duet. They found that even partial line of sight was sufficient to improve synchrony.

Group Creativity: Examples of group creativity are commonly found in everyday conversation. Conversation analysts have described how interlocutors use turn taking [27], eye gaze [17], and body position [18] to maintain successful conversations. It seems reasonable to infer that similar phenomena may exist in creative musical interactions. Healey et al. [16] examined the spatial behaviour of a group of seven improvising musicians. They observed how the use of space played a complex role in maintaining the coherence of the performance, and drew a number of parallels with conversational interactions.

An important idea that spans all forms of group creativity is that of emergence [29] - "the arising of novel and coherent structures, patterns, and properties during the process of self-organization in complex systems" [13]. Sawyer [28] adopts the term *collaborative emergence* to refer specifically to emergence in small groups. He points out that group members are often constrained as to what they can contribute to the emergent creative act. For example, improvising musicians might be constrained to play within the confines of a specific key and tempo. In the context of live, co-present group creativity group members need to work within these constraints, whilst also continuously monitoring and providing novel contributions to the interaction. Evidently this involves a combination of conscious and subconscious processing. However, Sawyer's interviews with

NIME'14, June 30 – July 03, 2014, Goldsmiths, University of London, UK.

musicians suggest a preference for the dominance of non-conscious mental states during improvised performance [29].

As an extension of Csikszentmihalyi's theory of flow [5], Sawyer conceives the idea of group flow [29], referring to a state of peak performance at the level of the group, rather than the individual. He points to the importance of factors such as parallel processing (simultaneous awareness of self and collaborator(s)) and visual attention in establishing a state of group flow.

Affect Recognition: Human emotion is commonly separated into three components; behavioural (expressions and actions), cognitive (thoughts and feelings), and physiological (biochemical and electrical changes in the body). In the field of Affective Computing researchers utilise these components with the goal of developing technologies that are able to recognise, react to, and/or express emotions. For example, functional MRI and EEG are able to identify felt emotions by analysing the brain's response to affective stimuli such as music [30] and images [31]. The analysis of facial expressions and posture can successfully discriminate emotional states [35], and in the study of anxiety, physiological parameters such as heart rate [1], galvanic skin response (GSR) [34], and salivary cortisol [32], have been shown to vary with levels of stress.

In recent years affective computing research has matured from controlled laboratory-based investigations to more true-to-life, spontaneous settings. Software developed by researchers at MIT Media Lab has collected global data using webcam images to continuously monitor the facially-expressed emotion of people viewing online videos [24]. In another application, the musical score and sequence of scenes in a film were guided by the emotional responses of the audience, as inferred from physiological measurements [25]. Such measures have also been used to detect musicians' emotions during musical performance [21].

Psychophysiology: Psychophysiology involves the study of how psychological experiences (thoughts, feelings, emotions) relate to the physiological activity of the body. Equipment for physiological measurement has become increasingly non-invasive, miniaturised and affordable, making it easier to conduct studies outside the laboratory. These developments are also leading towards the integration of physiological sensors in everyday technologies such as phones and computer games consoles.

In a study of flow during piano playing, Manzano et al. [7] measured heart rate, respiration and facial muscle movements while professional pianists gave five performances of a pre-prepared piece. They found a significant relationship between self-reported flow and heart rate variability, respiratory depth, and facial muscle movements. The same measures were employed, along with skin conductance, in a study of audience reactions to a live music performance [11]. The study used a computational model to determine high information content (IC) segments of the performed piece, whilst participants provided continuous subjective ratings of expectedness. Unexpected and high-IC events were generally associated with a rise in skin conductance, and decreased heart rate. Respiration rate increased only after the onset of unexpected events, and facial muscle movements showed no event-related responses.

Regarding human interactions, research into user experience with game technologies found differing physiological responses when participants were playing against a computer compared with playing against another human [23]. A study of partner influence during conversation found 'physiological linkage' between the blood pressure (BP) measurements of romantic couples [26]. High partner influence resulted in an in-phase relationship between the partners' BP measurements, and low influence resulted in an anti-phase relationship.

Numerous studies have sought to uncover links between brain activity and creativity. A comprehensive review of neuroimaging studies of creativity can be found in [9], where the authors highlight that the literature is, on the whole, fragmented and inconclusive.

3. THE STUDY

Incorporating methods and techniques from the research discussed above, we designed a study to gather both subjective, and continuous quantitative measures from pairs of co-present, improvising drummers. In each session the drummers performed two 5-10 minute improvisations, once where they were not visible to each other, then again where they were fully visible. The main aims of the study were to:

- **Assess** the practicalities of using various types of physiological and behavioural monitoring devices in a live performance setting.
- **Identify** which measurements and features are most informative/useful for our future work on the design of a collaborative interface for musical expression.
- **Report** some findings linking creativity, engagement, and emotion to quantitative features and measures such as motion and physiology.

We chose to use drumming in our study because it presents some noteworthy advantages over other forms of musical expression. In particular, beat timing and velocity can be accurately recorded using electronic pads. Large amounts of motion are involved, which increases the information conveyed through movement. There is also far less melodic content, which might otherwise influence participant emotion and constrain improvisational freedom. We simplified the experiment further by requiring that each participant only used one hand to drum on a single drum pad.

3.1 Method

3.1.1 Participants

Participants were recruited via email lists and word of mouth. We required that all participants had prior drumming experience and were confident enough to improvise rhythms 'on-the-fly'. Five pairs of participants took part in the study (2 mixed-sex pairs, 3 male pairs). Participants in each pair knew each other, and three of the pairs had previously played music together. The participants were aged 26 to 34 (M=29.1, SD=3.1), their drumming experience ranged from 1 to 17 years (M=7.4, SD=5.0), and their level of expertise ranged from 2 to 4 (M=2.7, SD=0.7) on a five point scale representing novice (1) to expert (5).

3.1.2 Measures

Given the exploratory nature of our study, we chose to collect a wide range of measurements so that we would have the flexibility to test various hypotheses in our post-study analysis. To measure heart rate and perspiration we used small, wireless ECG and GSR sensors provided by Shimmer Research. We used the Emotiv EEG headset to wirelessly record 14 channel EEG measurements from each participant. All of the physiological sensors contained accelerometers for recording motion. To provide more accurate motion measurements, we also used a Vicon marker-based motion tracking system to record continuous head, torso, arm, and feet position.

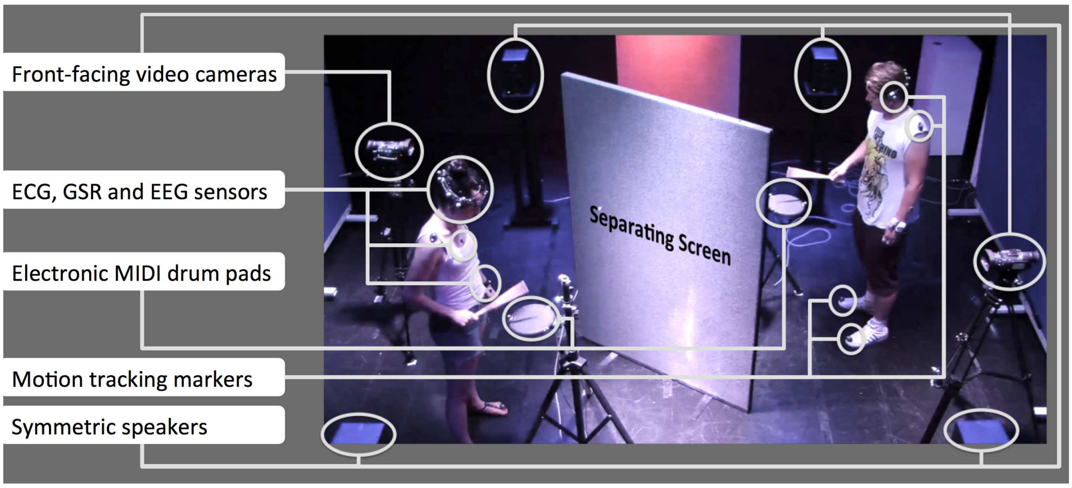

Figure 1: Image taken from the overhead camera illustrating the setup and the equipment used in the study

For the drums we used two identical Roland V-Drum electronic drum pads. By recording MIDI data from the pads we were able to log the exact timing and velocity (strength) of each drum beat. Three video cameras were set up - one facing each participant, and one overhead camera to capture the entire interaction. Figure 1 illustrates an image taken from the overhead camera. The annotations indicate the positioning of the measurement apparatus.

A post-performance questionnaire (PPQ) was designed to collect subjective report data from each participant while they reviewed video footage of their improvised performances. The PPQ asked participants to rate their individual levels of creativity, engagement, energy, positivity and boredom on a 9-point scale; as well as who they thought was leading the performance (1 = 'All me', 9 = 'All them'). The first two items assess subjective interpretations of the drumming task, while the latter four items were chosen to gauge emotional state, relating closely to the dimensions of valence, arousal, and dominance, commonly used in affect research [14].

3.1.3 Data Synchronisation

We used two computers and four separate applications to record the continuous measurements. Consequently we were faced with the problem of how best to synchronise all of the data. Our solution was to place the physiological and EEG sensors on top of one of the drum pads and use a beater (with a motion capture marker on it) to tap the drum 10 times. This meant that we had 10 clearly identifiable, short-duration peak events in the EEG and physiological accelerometer data, accompanied by 10 MIDI note events and 10 visible motion capture/video events. When processing the data, we were able to use these events as reference points, enabling us to align all of the data sources to a high (millisecond) precision.

3.1.4 Setup

The study was held in a performance lab with stage lighting set up to make it feel more like a live music venue. The drum pads were positioned in the centre of the room, with speakers either side (see Fig. 1). The two computers were placed out of sight behind blank screens at one end of the room; this is also where the experimenter sat during the drum performances. ECG modules were strapped around each participant's waist, with the electrodes attached to their chest. GSR modules were placed around the wrist of their non-drumming hand, and the electrodes were strapped to their index and middle finger. The EEG headsets were positioned and fitted with motion capture markers, which were also placed around the participants' wrists, and on their shoulders and toes.

3.1.5 Tasks

The experiment consisted of three drumming tasks. The first task (tM, duration ~ 1 min) required the participants to play along to a metronome click track at a tempo of 110 bpm. The second (tS, duration ~ 1 min) required them to repeat a set rhythmic phrase, which they listened to and learnt prior to the task. These initial two tasks were designed to provide baseline measurements of the participants' rhythmic timing and physiological measures. For the third and final task (tI, duration ~ 6-10 min) the participants were asked to improvise with one another, where the only condition was that they did not use verbal communication. All three tasks were performed twice, once under a non-visual (NV) condition, then again under a visual (V) condition. In the NV condition participants were either facing away from each other (sessions 1-3), or blocked by a screen (sessions 4 and 5). In the V condition they faced towards each other with no obstruction, other than the drum pad. The participants performed all the tasks as a pair, except for in the V condition, where they performed the tM task individually. Following completion of the drum tasks, the participants sat individually and watched the overhead videos of their two improvised performances. After each minute[1] of video they were asked to complete all the items on the PPQ, in relation to that particular minute of their performance.

3.2 Data Processing

3.2.1 Preparation

The EEG, ECG, GSR, and MIDI data was imported into MATLAB[2]. For each session the accelerometer synchronisation peaks and MIDI note events were used to align the data to a common start point (t_0). Using the video footage we found the start and end times of each experimental task, relative to t_0. For each data source these time points were used to extract and label blocks of data corresponding to measurements for each participant and each task.

[1]For session 5, two minute segments were used because the improvisation tasks were longer in duration.
[2]Due to software issues, we have not yet been able to process the Vicon motion capture data.

3.2.2 Feature Extraction

Features were extracted from individual data blocks according to the type of data they contained. We manually labelled anomalous physiological data so that it could be excluded from further analysis.

ECG: We used ECGtools[3] to filter the raw ECG data and extract the R-peaks, which correspond to individual heart beats. The distance between consecutive peaks was then used to find the instantaneous heart rate (HR) values. These values were interpolated to give an evenly spaced time series from which we extracted the mean, variance, SD, maximum, minimum, the positions of maxima and minima, and the number of extrema divided by the task duration.

GSR: Skin conductance response (SCR) has been shown to be a useful metric in analysis of GSR data [19, 20]. We used Ledalab[4] to extract the timing and amplitude of SCR events using Continuous Decomposition Analysis (CDA). Again, interpolation was performed and the mean, variance, SD, positions of maxima and minima, and number of extrema divided by task duration, were calculated from the SCR amplitude series.

EEG: Frequency band power values are often computed in EEG studies, as they provide information on cognitive activity. Using EEGlab [8] we initially bandpass filtered the signal between 3 and 30 Hz. We then performed manual artefact rejection to remove noisy segments of data caused by head and facial muscle movements. Artefactual channels were removed entirely and the average power over all remaining channels was computed within the following standard frequency bands: Theta (4-7 Hz), Alpha (7.5-12.5 Hz), L-Beta (12.5-25 Hz), and H-Beta (25-30 Hz).

MIDI: The number of beats per second (BPS), SD in time between consecutive beats, and mean velocity were computed as basic MIDI features. To measure the timing synchrony between one participant (Px) and the other (Py) we compared their individual beat onsets (tPx and tPy) and considered any beats which occurred within 70ms of each other to be single rhythmic events [10]. For these beats we calculated the time difference (tPx - tPy). We then found the mean, and absolute mean over all the difference values. For tM data the same procedure was used to measure the synchrony between individual participants and the MIDI encoded metronome events.

Motion: We took the accelerometer readings from the ECG, GSR and EEG sensors and summed the absolute values of the axial components across the entire measurement block for each sensor. This gave us approximate quantity of motion (QoM) values for the head (EEG), torso (ECG), and non-drumming hand (GSR).

4. ANALYSIS & RESULTS

4.1 Subjective Reports Versus Continuous Interaction Features

The first part of our feature analysis aims to test whether participants' post-performance subjective reports were correlated with their within-performance continuous measures for the improvisation task. To do this we segmented the continuous tI data into 1 or 2 minute windows (tI_w), identical to those used for the PPQs. This was done for each participant within each condition (NV or V). Features were then extracted from each tI_w. We performed baseline scaling on the features using two separate procedures. The first method ($bAdj_I$) divided each tI_w feature by the equivalent feature extracted over the entire tI task. The second method ($bAdj_S$) scaled relative to the tS values. For the SR scores we used both adjusted and non-adjusted values. In this case adjustment was performed using the $bAdj_I$ method only, as we did not have SR scores for any of the other tasks. Treating every window as an independent row of samples, we ran pairwise Pearson correlation analysis between each column of SR scores and each column of features. Significant correlations are shown in Table 1.

We can see that all of the physiological data sources have at least one feature which correlates with at least one SR item. For ECG and GSR, the mean HR/SCR and number of HR/SCR extrema are the most informative features. For EEG, MIDI, and motion data, the informative features are the four spectral band powers, number of beats, and mean body QoM respectively. Strong ($r > 0.4$) correlations are highlighted in bold, with the majority of these falling under energy, positivity, and boredom SR items. Of particular note are the correlations with no. of SCR extrema, and with mean H-Beta power. The correlations with mean body QoM are to be expected, given that high amounts of movement are generally linked to high arousal and valence [4]. Self reported creativity is most significantly correlated with BPS, followed by average heart rate and mean body QoM. Engagement is correlated negatively with heart rate and positively with BPS and QoM. Leadership is positively correlated with Beta activity and BPS.

4.2 Effect of Visibility

To test for effects of participant visibility we performed paired-sample t-tests comparing both SR and continuous features in NV and V conditions. In this case we used features averaged over the entire improvisation session for each participant, under each condition. The continuous feature values were all baseline adjusted using $bAdj_S$, meaning that we compared tI features relative to tS features within each visual condition. The results are shown in Table 2. We see that engagement is the only SR measure which shows significant differences, whilst the p-values for creativity and boredom suggest potential significance if more trials were performed. The sign of the t-values indicates that Creativity and Engagement were given lower ratings in the NV sessions than the V sessions, and Boredom was given higher ratings. Regarding continuous features, the mean heart rate appears to be significantly higher in the NV condition, whereas the SD in SCR amplitude is lower. Again, the effects of visibility on SD in heart rate and mean SCR amplitude show potential significance given more trials. The same can be said for MIDI and motion features, where we see that the mean velocity and mean bodily QoM were higher in the NV condition.

5. DISCUSSION

Some of the most significant correlations in our analyses in 4.1 came from EEG measurements. This is somewhat surprising, as we had expected that the susceptibility to movement artefacts might have distorted any trends in the data. In comparison with SR measures of energy and positivity, Beta activity was positively correlated, whilst Theta and Alpha activity were negatively correlated. These results concur with previous studies which associate Beta activity with engagement and cognitive challenge; and Theta and Alpha activity with drowsy states, and reflective states of relaxation, respectively [3].

[3]http://www.ecgtools.org/

[4]http://www.ledalab.de/

Table 1: Pearson correlation coefficient (r) values for significant correlations between extracted features and self report ratings (using windowed epochs of tI data from both V and NV conditions)

Data	**Feature**	**Self report items**					
		Creativity	Engagement	Energy	Positivity	Boredom	Leadership
ECG	Mean HR	$.32^{*\bullet\dagger}$	$-.39^{**\circ}$	$.39^{*\bullet\dagger}$	$.34^{*\bullet\dagger}$	$.32^{**\circ}$	
-	No. of HR extrema				$.27^{*\circ}$	$-.26^{*\circ}$	
GSR	Mean SCR amp.					$\mathbf{.42}^{***\circ}$	
-	No. of SCR extrema			$\mathbf{.42}^{***\circ}$	$\mathbf{.48}^{***\circ}$		
-	-			$.34^{*\bullet\dagger}$	$.35^{*\bullet\dagger}$		
EEG	Mean Theta power			$-.31^{*\circ}$	$-.35^{*\circ}$		
-	Mean Alpha power			$-.30^{*\circ}$	$-.39^{**\circ}$		
-	Mean L-Beta power			$.34^{*\bullet}$	$\mathbf{.48}^{***\bullet}$	$-.32^{*\bullet}$	$.31^{*\circ}$
-	Mean H-Beta power			$\mathbf{.57}^{***\bullet}$	$\mathbf{.78}^{***\bullet}$	$\mathbf{-.40}^{**\bullet}$	$.38^{**\circ}$
-	-				$.38^{**\circ}$		
MIDI	No. of beats per sec.	$.29^{**\bullet}$	$.30^{*\dagger}$	$.33^{*\dagger}$	$.38^{**\dagger}$		$.30^{*\dagger}$
-	-	$\mathbf{.41}^{**\dagger}$					
Motion	Mean body QoM	$.27^{*\bullet}$	$.22^{*\bullet}$	$.22^{*\bullet}$	$\mathbf{.52}^{***\bullet\dagger}$	$\mathbf{-.44}^{**\bullet\dagger}$	
-	-	$.31^{**\circ}$	$.38^{**\bullet\dagger}$	$\mathbf{.57}^{***\bullet\dagger}$			

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$, features adjusted using $^{\circ}bAdj_S$ or $^{\bullet}bAdj_I$, †SR scores adjusted using $bAdj_I$, $r > 0.4$ highlighted in **bold**.

Table 2: Paired-sample t-test results for effect of visual condition on SR items and continuous features

Data		**t**	**p**	**df**
SR	Creativity	-1.80	.146	4
-	Engagement	-2.85^{*}	.047	4
-	Boredom	1.98	.120	4
ECG	Mean HR	4.17^{**}	.009	5
-	SD HR	1.97	.106	5
GSR	Mean SCR amp.	-2.24	.075	5
-	SD SCR amp.	-3.95^{*}	.011	5
MIDI	Mean velocity	1.96	.091	7
Motion	Mean body QoM	1.76	.122	7

Note: t = t-score, p = p-value, df = degrees of freedom, $^{*}p < .05$, $^{**}p < .01$.

Extracting the number of extrema as a feature of continuous HR and SCR data is not a method commonly used in other studies. However, we found this to be one of the features which showed the strongest correlations to our SR data. In particular, the number of SCR extrema showed strong correlations with reported energy and positivity. Correlations were weaker for the number of heart rate extrema.

Creativity appears to share correlation features with self reported energy and positivity. This lends support to previous research which highlights the importance of arousal and positive valence in the generation of creative ideas [6]. The lack of relations between creativity and EEG features is interesting, because it may be indicative of the contrasting use of both non-conscious and conscious thought during creative action. This holds true with previous EEG research, which has struggled to show conclusive links between creativity and localised brain activity [9]. The correlations between leadership and Beta activity make sense, since we would expect leadership to induce higher cognitive engagement.

Our results in 4.2 suggest that participant visibility has effects, not just upon self reported aspects of interaction, but also on physiology and performance features. Further trials need to be performed in order to verify the statistical significance of these effects. Our experimental design also means that these results may be subject to an ordering bias, due to the NV tasks always being held prior to the V ones. However, if validated by further experiments, these findings could have a large impact upon the design of collaborative musical interfaces.

Throughout our analysis we found that the choice of baseline adjustment method had a large effect on the results. Understanding the nature of these effects will be important, especially if such sensors are to be incorporated into interfaces for public use, where baseline data collection is challenging.

In summary, our findings indicate that continuous physiological, motion and performance measures can be used to infer subjective aspects of participant engagement, creativity and affect during live collaborative music making. Such measures could be adopted as a means of gathering continuous evaluation metrics during the testing of new interfaces. This would allow designers to manipulate the layout of their interface on-the-fly, whilst obtaining quantitative indicators of how each layout influenced the user experience. Regarding the design of interfaces, we envisage that such measures could be used in a similar way, enabling the interface to adapt to the participant in real-time. For example, an interface might detect boredom and respond by providing new options, whilst also conveying this emotional state to the other participants, so that they may choose to adjust their contributions. It is foreseeable that as physiological measurement and motion capture technology becomes increasingly non-invasive and user friendly, such devices could be readily incorporated into interface designs.

6. CONCLUSIONS

The scale of this study means that our experimental results are more suggestive than conclusive. However, our findings support the hypothesis that continuous measures of affect, psychophysiology, and performance are potentially valuable in the evaluation and design of interfaces for collaborative music making. Our future work will explore how these measures can be used to provide live emotional and behavioural feedback to interacting musicians. Further experiments will then allow us to evaluate how such interventions influence the participant experience, and the outcomes of the interaction.

7. ACKNOWLEDGEMENTS

This work was funded by EPSRC through the Media and Arts Technology Programme, an RCUK Centre for Doctoral Training.

8. REFERENCES

[1] Appelhans et al. Heart rate variability as an index of regulated emotional responding. *Review of General Psychology*, 10(3):229–240, 2006.

[2] N. Bryan-Kinns. Mutual engagement and collocation with shared representations. *Int'l Journal of Human-Computer Studies*, 71(1):76–90, Jan. 2013.

[3] T. Budzynski et al. *Introduction to Quantitative EEG and Neurofeedback: Advanced Theory and Applications*. Academic Press, 2nd editio edition, 2008.

[4] A. Castellano et al. Recognising Human Emotions from Body Movement and Gesture Dynamics. In *Cont. on Affective Computing and Intelligent Interaction*, pages 71–82, Lisbon, Portugal, 2007. Springer-Verlag.

[5] M. Csikszentmihalyi and K. Rathunde. The measurement of flow in everyday life: Toward a theory of emergent motivation. *Developmental perspectives on motivation. Current theory and research in motivation*, 40:57–97, 1992.

[6] C. K. W. De Dreu et al. Hedonic tone and activation level in the mood-creativity link: toward a dual pathway to creativity model. *Journal of personality and social psychology*, 94(5):739–56, May 2008.

[7] O. de Manzano et al. The psychophysiology of flow during piano playing. *Emotion (Washington, D.C.)*, 10(3):301–11, June 2010.

[8] A. Delorme and S. Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21, Mar. 2004.

[9] A. Dietrich and R. Kanso. A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological bulletin*, 136(5):822–48, Oct. 2010.

[10] S. Dixon. Automatic Extraction of Tempo and Beat From Expressive Performances. *Journal of New Music Research*, 30(1):39–58, Mar. 2001.

[11] H. Egermann et al. Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, affective & behavioral neuroscience*, 13(3):533–53, Sept. 2013.

[12] D. Glowinski et al. Expressive Non-Verbal Interaction in String Quartet. In *Conf. on Affective Computing and Intelligent Interaction*, pages 233–238, Geneva, Switzerland, 2013. IEEE Computer Society.

[13] J. Goldstein. Emergence as a Construct: History and Issues. *Emergence*, 1(1):49–72, Mar. 1999.

[14] H. Gunes and B. Schuller. Categorical and Dimensional Affect Analysis in Continuous Input: Current Trends and Future Directions. *Image and Vision Computing*, 31(2):120–136, July 2013.

[15] A. Hadjakos, T. Groß hauser, and W. Goebl. Motion Analysis of Music Ensembles with the Kinect. In *International Conference on New Interfaces for Musical Expression (NIME)*, pages 106–110, KAIST, Daejeon, Korea, 2013.

[16] P. G. T. Healey et al. Inter-Play : Understanding Group Music Improvisation as a Form of Everyday Interaction. In *Proc. of Less is More - Simple Computing in an Age of Complexity*, 2005.

[17] A. Kendon. Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26:22–63, Jan. 1967.

[18] A. Kendon. Spacing and Orientation in Co-present Interaction. *Lecture Notes in Computer Science*, 5967:1–15, 2010.

[19] J. Kim and E. André. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, 30(12):2067–83, Dec. 2008.

[20] K. H. Kim et al. Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing*, 42(3):419–427, May 2004.

[21] R. B. Knapp et al. Measurement of motion and emotion during musical performance. In *3rd Int'l Conference on Affective Computing and Intelligent Interaction*, pages 1–5. IEEE, Sept. 2009.

[22] I. Konvalinka et al. Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Quarterly journal of experimental psychology (2006)*, 63(11):2220–30, Nov. 2010.

[23] R. L. Mandryk et al. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & Information Technology*, 25(2):141–158, Mar. 2006.

[24] D. McDuff, R. el Kaliouby, and R. Picard. Crowdsourced data collection of facial responses. In *Proceedings of the 13th international conference on multimodal interfaces - ICMI '11*, page 11, New York, New York, USA, Nov. 2011. ACM Press.

[25] J. Price. Plug in to movies that know how you're feeling. *New Scientist Online*, 2011.

[26] R. G. Reed et al. Partner influence and in-phase versus anti-phase physiological linkage in romantic couples. *Int'l Journal of Psychophysiology*, 88(3):309–16, June 2013.

[27] H. Sacks et al. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735, 1974.

[28] R. K. Sawyer. The emergence of creativity. *Philosophical Psychology*, 12(4):447–469, Dec. 1999.

[29] R. K. Sawyer. *Group Creativity: Music, Theater, Collaboration*. Psychology Press, 2003.

[30] L. A. Schmidt and L. J. Trainor. Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cognition & Emotion*, 15(4):487–500, July 2001.

[31] F. Schneider et al. Functional MRI reveals left amygdala activation during emotion. *Psychiatry Research: Neuroimaging*, 76(2-3):75–82, Dec. 1997.

[32] N. Takai et al. Effect of psychological stress on the salivary cortisol and amylase levels in healthy young adults. *Archives of oral biology*, 49(12):963–8, Dec. 2004.

[33] B. Vera et al. A Study of Ensemble Performance Under Restricted Line of Sight. In *Proceedings of the International Conference on Music Information Retrieval*, Curitiba, Brazil, 2013.

[34] T. Wu et al. The impact of anxiety on social decision-making: Behavioral and electrodermal findings. *Social Neuroscience*, 8(1):11–21, Jan. 2013.

[35] Z. Zeng et al. A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence*, 31(1):39–58, Jan. 2009.