

Using Affective and Behavioural Sensors to Explore Aspects of Collaborative Music Making





Evan Morgan*, Hatice Gunes, Nick Bryan-Kinns

*School of Electronic Engineering and Computer Science, Queen Mary University of London,
Mile End Road, London, E1 4NS, UK*

Abstract

Our research considers the role that new technologies could play in supporting emotional and non-verbal interactions between musicians during co-present music making. To gain a better understanding of the underlying affective and communicative processes that occur during such interactions, we carried out an exploratory study where we collected self-report and *continuous* behavioural and physiological measures from pairs of improvising drummers. Our analyses revealed interesting relationships between creative decisions and changes in heart rate. Self reported measures of creativity, engagement, and energy were correlated with body motion; whilst EEG beta-band activity was correlated with self reported positivity and leadership. Regarding co-visibility, lack of visual contact between musicians had a negative influence on self reported creativity. The number of glances between musicians was positively correlated with rhythmic synchrony, and the average length of glances was correlated with self reported boredom. Our results indicate that ECG, motion, and glance measurements could be particularly suitable for the investigation of collaborative music making.

Keywords: Psychophysiology, affect, collaboration, creativity, improvisation, music

©2015, Elsevier. Licensed under the Creative Commons    
Accepted manuscript available online:
<http://dx.doi.org/10.1016/j.ijhcs.2015.05.002>

*Corresponding author: *Phone*, +44(0)7815 481 041, *Address*, Queen Mary University of London, 327 Mile End Road, Computer Science Room CS300, London, E1 4NS
Email addresses: e.l.morgan@qmul.ac.uk (Evan Morgan), h.gunes@qmul.ac.uk (Hatice Gunes), n.bryan-kinns@qmul.ac.uk (Nick Bryan-Kinns)

1. Introduction

Modern technologies and interfaces for musical expression have provided new ways for humans to collaborate creatively and make music together (Bevilacqua et al., 2013). These advances are being adopted by a wide range of people, from professional musicians, to those less able to play traditional instruments. The devices can be controlled by various means, such as touch surfaces, gesture recognition, and accelerometers (Kuhara and Kobayashi, 2011; Mitchell et al., 2012; Kuyken et al., 2008), allowing sounds to be created through novel modes of expression. Despite their heightened ability to sense human input, these devices predominantly focus on single-user operation and do very little to sense and support the important emotional, interpersonal, and communicative elements of collaborative music making (Carlile and Hartmann, 2005; Fencott and Bryan-Kinns, 2010).

Our research is motivated by a desire to develop interfaces that specifically take into account the presence, behaviours, and emotions of musical collaborators. Such an interface could be integrated into the musical instrument itself, or it could function alongside both traditional and modern instruments. The underlying aim is that it should support and enrich the affective aspects of collaborative performance, and in turn, the creative musical outcomes. In order to achieve this, we must first gain a better understanding of the behavioural and affective processes that accompany collaborative music making and creativity. There is very little existing research in this area, and our work is largely influenced by the fields of Affective Computing and Psychophysiology; both of which deal with the detection of people’s inner psychological states based upon externally observable and quantifiable measurements. In the following section we provide an overview of related research in these fields. The remainder of the paper documents a study, where we collected continuous behavioural and affective measurements from pairs of improvising drummers. We use these measures to explore various aspects of collaborative music making.

2. Related Work

Our multidisciplinary approach is influenced by existing research in a number of different fields. We can broadly organise these studies into two categories: i) studies relating to the measurement of behaviour and affect, and ii) studies dealing with the analysis of co-present, collaborative music making activities. This section discusses relevant work within these two categories.

2.1. Measuring Behaviour and Affect

A major component of our research is the use of sensor technologies to measure and quantify aspects of human behaviour and affect. The apparatus and techniques that we employ are influenced by research across a range of disciplines, but most prevalently *Psychophysiology*, *Affective Computing*, and *Social Signal Processing*. We will briefly introduce these disciplines before highlighting literature that is specifically relevant to this study.

Psychophysiology is the study of how psychological experiences (thoughts, feelings, emotions) relate to the physiological activity of the body. The typical approach to psychophysiological studies is to measure physiological variables in the lab, using equipment developed for medical diagnostics. These measurements are then compared to qualitative and quantitative measures of behaviour and experience, according to the specific focus of the study. Over recent years equipment for physiological measurement has become increasingly non-invasive, miniaturised and affordable, making it easier to conduct studies, not just in the lab but also in naturalistic settings (Morgan et al., 2013). Furthermore, these developments are leading towards the integration of physiological sensors in everyday technologies such as phones, smartwatches, and computer game consoles. An advantage of adopting physiological measurement as a means for inferring psychological states is that it often requires no conscious effort on behalf of the user. The development of a car seat that can sense the driver’s heart rate and detect tiredness (Edwards, 2013) is great example of this paradigm; involving technology development (Walter et al., 2011), research in psychophysiology (Patel et al., 2011), and an application that revolves around the need for subconscious measurement.

The field of *Affective Computing* also concerns the measurement of psychological states. However, it focuses more specifically on the development of technologies that are able to recognise, react to, and/or express emotions. The work in this field has mainly focused on categorising the discrete emotional responses of individuals who are presented with pre-recorded, static or virtual stimuli, usually in a laboratory setting (Sariyanidi et al., 2014).

Social Signal Processing (SSP) (Pentland, 2005) is a relatively new domain, which aims to provide computers with the ability to sense and understand human social signals (Vinciarelli et al., 2009). In this context a social signal is defined as a “communicative or informative signal that, either directly or indirectly, provides information about ‘social facts’, that is, about social interactions, social emotions, social attitudes, or social relations” (Pantic et al., 2011, p. 8). SSP researchers develop tools and techniques for the sensing and machine analysis of behavioural and psychological constructs. A thorough survey of current work in SSP can be found in Vinciarelli et al. (2012).

A common theme within these disciplines is that human experience and affect can be segmented into three components; cognitive (thoughts), behavioural (expressions and actions) and physiological (biochemical and electrical changes in the body). Measuring each of these components presents varying challenges, requiring distinct technologies and processing techniques. The following sections discuss relevant research relating to each component, since our study incorporates measurements of all three.

2.1.1. Cognitive Measurement

Neuroimaging techniques allow us to obtain information about cognitive processes inside the brain. The most commonly used techniques are EEG, which involves placing electrodes on the subject’s scalp; and fMRI, which involves the subject lying still in a large magnetic scanner. Both techniques have been used

to identify felt emotions by analysing the brain’s response to affective stimuli such as images (Schneider et al., 1997) and music (Schmidt and Trainor, 2001). Neuroimaging studies have also sought to uncover links between brain activity and creative behaviour. A comprehensive review of such studies is provided by Dietrich and Kanso (2010), where the authors highlight that the literature is, on the whole, fragmented and inconclusive. They broadly conclude that tasks involving creative cognition induce changes in pre-frontal activity of the brain. On top of this they suggest that creativity may not even be localisable in the brain, given limitations of current neuroimaging systems.

Regarding the practicalities of neuroimaging, more user-friendly EEG systems have been developed in recent years, however they still suffer from high susceptibility to noise, and a poor spatial imaging resolution. fMRI has far better spatial imaging resolution, but the requirement for participants to be stationary inside a large and immobile scanner make it impossible to conduct studies in naturalistic settings. For example, in a study of the neural aspects of musical improvisation, jazz pianists were asked to play whilst lying in an fMRI scanner (Limb and Braun, 2008).

2.1.2. Behavioural Measurement

In the context of our research we are interested in measuring relatively short-term behaviours (in the order of seconds and minutes), many of which can be categorised as non-verbal communicative acts. Argyle (1978) outlines seven forms of non-verbal communication (NVC): *facial expressions, gaze, gestures, bodily posture, bodily contact, spatial behaviour* (e.g. proximity), and *appearance* (e.g. clothing). He models NVC as a simple, communications theory-inspired sequence, whereby a sender encodes a ‘social signal’, which is subsequently decoded by a receiver. It is implicit that this signal transfer is not error free. One person will never perfectly interpret the non-verbal communicative act of another person. Furthermore, the process of NVC does not always involve conscious awareness. This makes it a particularly interesting parameter in the study of human interactions, as it indicates subtle features of the interaction that cannot be revealed through self-report based measures.

Because we are investigating co-present musical interactions, we will be most concerned with gaze, bodily posture, and spatial behaviour. Gaze can reveal a great deal about the dynamics and nature of co-present human interactions. Numerous studies have shown how gaze is closely synchronised with speech during conversations (Kendon, 1967; Cummins, 2012; Oertel et al., 2012). Additionally, the amount of time people spend looking at each other has been shown to relate to dominance and rapport (Argyle, 1978). Mutual gaze has also been shown to be physiologically arousing (Mazur et al., 1980). Gaze is commonly measured by manually annotating video footage, which is a time-consuming task. However, modern eye-tracking glasses are able to continuously track where someone is looking within a scene captured from a head-mounted camera.

With respect to bodily posture and spatial behaviour, accurate measurements can be obtained using a marker-based motion tracking system, which uses multiple cameras to detect small reflective markers positioned on the body.

Glowinski et al. (2013) used such a system to study the bodily movements of a string quartet. Their results suggest that head movement features can be used to distinguish between an engaging and non-engaging performance (as rated by the performers). Healey et al. (2005) 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. The trade-off with marker-based systems is that they take some time to set up and are not particularly portable. With the advent of the Microsoft Kinect, researchers have been able to measure body movements using a single depth sensitive camera (Hadjakos et al., 2013; Morgan et al., 2013). Other solutions involve using worn accelerometers (Zhou et al., 2008), and pressure sensitive floors (Bränzel et al., 2013).

It should be noted that the non-verbal behaviours defined by Argyle, and discussed above, are all forms of *visual* communication and do not encompass auditory behaviours. In the study of musical interactions it is also important to consider the effects that auditory feedback might have on NVC between musicians. For example, the timing synchrony between two musicians tapping a rhythm has been shown to improve when the taps trigger sounds, but to be invariable to the effects of visual information (Nowicki et al., 2013). Measuring the physical properties of auditory information, such as loudness and pitch, is easily achieved. However, it is significantly more challenging to measure how auditory information is processed in the brain.

2.1.3. *Physiological Measurement*

Mechanical, electrophysical and biochemical changes in the body can be measured using surface electrodes and sensors positioned at specific sites on the body. In a study of flow¹ during piano playing, de Manzano et al. (2010) measured heart rate, respiration and facial muscle movements while professional pianists gave five performances of a pre-prepared piece. For each performance the pianists were subsequently asked to rate their level of flow using a questionnaire. A significant relationship was found between flow and heart rate variability, respiratory depth, and facial muscle movements. Physiological studies have also been conducted to try and measure musicians' emotions during musical performance (Knapp et al., 2009). A potential application of live physiological measurement in music collaboration has been demonstrated by Mealla et al. (2011), who created an interactive musical tabletop, where physiological signals contributed to the generated sounds.

In a study of the physiological reactions of audience members to a live music performance, Egermann et al. (2013) found that unexpected musical events were generally associated with a rise in skin conductance, and decreased heart rate. A real-world application of these ideas saw the musical score and sequence of scenes in a film being dictated by the emotional responses of the audience, as

¹Csikszentmihalyi's theory of Flow (Csikszentmihalyi and Rathunde, 1993) describes the mental state of being completely absorbed in an activity.

inferred from physiological measurements (Price, 2011).

Regarding human-human interactions, a study of partner influence during conversation found ‘physiological linkage’ between the blood pressure (BP) measurements of romantic couples (Reed et al., 2013). Research into user experience with game technologies found differing physiological responses when participants were playing against a computer compared with playing against another human (Mandryk et al., 2006).

In summary, cognitive, behavioural and physiological measurements provide many ways in which we can infer and quantify affective aspects of human experience. In recent years the use of these techniques has started to make the transition from laboratory-based studies to real-world applications. However, there has been comparatively little focus on situations where the interaction is occurring *between* people.

2.2. Co-Present Collaborative Music Making

In the context of our research, we define co-present collaborative music making as any situation where two or more people are jointly involved in the process of creating live music. The following sections discuss research that is relevant to group musical interactions in general, before focusing on improvisation, which is the specific case of co-present collaborative music making that we investigate in this paper.

2.2.1. Group Interactions

Sawyer (2003) proposes two generalised approaches to the study of group interactions: the *input-output (IO) approach* and the *process approach*. The former concerns things that take place before and after the interaction, whilst the latter looks at what occurs during the interaction. Quantitative methods can provide interesting insights into group interaction processes. For example, Fencott and Bryan-Kinns (2010) used interaction logs to analyse specific aspects of software-based co-present collaborative music making, such as the amount of co-editing that occurred. Hadjakos et al. (2013) developed a quantitative method for analysing the rhythmic synchronisation of a violin duo using the Kinect motion tracking device. By tracking head movements they were able to demonstrate how complex interaction patterns could be observed.

As an extension to the Theory of Flow (see section 2.1.3), Sawyer (2003) conceives the idea of Group Flow, 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. However, Sawyer does not back up his theory of group flow with a substantial amount of evidence, and leaves it somewhat unsatisfactorily defined. A similar concept is that of mutual engagement, which has been highlighted as an important feature of group musical interactions (Bryan-Kinns, 2013). Bryan-Kinns and Hamilton (2012) have developed a Mutual Engagement Questionnaire (MEQ) for evaluating the mutual engagement qualities of different musical interfaces.

Emotion also plays an important part in group interactions. In particular, we may be interested in the processes by which the emotional representations of a person or group influence the emotions of another person or group. This is often termed *emotional contagion*. There is a lack of research on emotional contagion in performing musicians, however it has been suggested that it is one of the mechanisms by which people experience emotions while listening to music (Lundqvist et al., 2008). Studies of group interactions have shown that emotional contagion affects group processes such as task performance and cooperation (Barsade, 2002). Similar processes of influence have been observed in relation to behavioural displays. *Behavioural mimicry* is the process whereby actions or emotions represented by one person subconsciously cause congruent behaviour in another person. There is a growing body of evidence for behavioural mimicry (see Chartrand and Lakin (2012) for an extensive review), and its links to cognitive processes. In the musical domain it has been suggested that behavioural mimicry contributes to temporal assimilation and coordinated variations in intensity and intonation during ensemble performances (Keller et al., 2014).

2.2.2. Group Improvisation

Group musical improvisation is defined as a spontaneous process, whereby creative contributions are made within the restrictions of real-time performance (Kenny and Gellrich, 2002; Wilson and MacDonald, 2012). Research on group musical improvisation is limited, and existing studies have predominantly focused on jazz music. Seddon (2005) used videotapes of six jazz musicians during rehearsal and performance in order to investigate modes of communication during jazz improvisation. He defined three modes of non-verbal communication: *instruction*, *cooperation*, and *collaboration*. Instruction involves the demonstration of musical ideas through vocalisation, or use of an instrument. Cooperation occurs when the musicians are producing a cohesive performance with the inclusion of musical and visual cues. Collaboration is the state where musicians are able to stimulate the spontaneous generation of creative contributions within the group, communicating exclusively through musical interaction. During collaboration, Seddon defines the musicians as being ‘empathetically attuned’ - a state of mutual empathy that encourages the musicians to take risks and challenge each other’s creativity.

Regarding the cognitive factors involved in improvisation, Kenny and Gellrich (2002) propose eight processes: short, medium, and long-term *anticipation*; short, medium, and long-term *recall*; *flow status* (see section 2.1.3); and *feedback* (decisions based upon previous experiences). Similar cognitive processes were also identified by Biasutti and Frezza (2009), who developed an Improvisation Process Questionnaire, which they gave to 76 experienced musicians. In addition to anticipation, use of repertoire, flow, and feedback, the authors highlight *emotive communication* as an important ability required by improvisers.

Comparisons have been made between jazz improvisation and conversation (Monson, 1996; Sawyer, 2005), whereby a common language or vocabulary is used alongside rules (e.g. grammar) in order to form a coherent and emergent

interaction between two or more people. Conversation analysts have described how interlocutors use non-verbal behaviours such as eye gaze (Kendon, 1967), and body position (Kendon, 2010) to maintain successful conversations. There is already some evidence that similar processes may occur during collaborative musical creativity (Healey et al., 2005). A noteworthy difference between conversation and musical improvisation is that the former is characterised by turn taking (Sacks et al., 1974), whilst the latter involves simultaneous contributions. Musicians must, therefore, continuously monitor the contributions of their collaborators, whilst also providing their own novel contributions. This involves a combination of both conscious and sub-conscious processing (Sawyer, 2003).

Finally, it is worth considering how the moods and emotions of group members might influence the creative aspects of improvisation. A theory for the relationship between emotion and creativity is provided by the ‘dual pathway’ model (De Dreu et al., 2008) (see Fig. 1). This model suggests that emotions with high arousal (e.g. anger, elation) lead to greater originality and creative fluency (the number of ideas, insights and solutions generated) when compared to low arousal emotions (e.g. sadness, serenity). It also proposes that positive emotions contribute to this process by facilitating greater cognitive flexibility and inclusiveness; whilst negative emotions facilitate increased persistence and perseverance. It is important to note that this model was developed based upon studies of individual creativity, and it is not clear how it might apply to group creativity. Existing studies of group creativity have predominantly explored the influence of long-term mood on creative output, often in a workplace setting (Amabile et al., 2005; Jamison, 1996). A notable feature of these studies is that they tend to focus on correlations between discrete emotional states and overall creativity. There is an absence of research that addresses continuous affective interactions and their real-time influences on creative tasks.

2.3. Summary

Our research is concerned with the use of sensor *technologies* to explore aspects of collaborative music making. Each of these technologies has associated *techniques* for the collection of useful data. Following data collection, existing *concepts* can be used to guide the analysis and interpretation of the data. In this section we have reviewed a wide range of topics with associated technologies, techniques, and concepts. We have seen how motion sensing technologies can be used in tracking and mediating interactions, and measuring the subtle movements of musicians. Whilst small, wireless physiological sensors are used to sense and recognise affect and social signals. Techniques for collecting data in both controlled and naturalistic settings have been described. These include the strictly controlled fMRI scans of improvising pianists, and the collection of subjective and physiological data from audience members watching a live performance. Finally, concepts relating to the phenomena we wish to investigate have been discussed. These include the theory of group flow, the dual pathway approach to creativity, and concepts surrounding non-verbal communication and auditory behaviours. In the following sections we will see how specific sensor

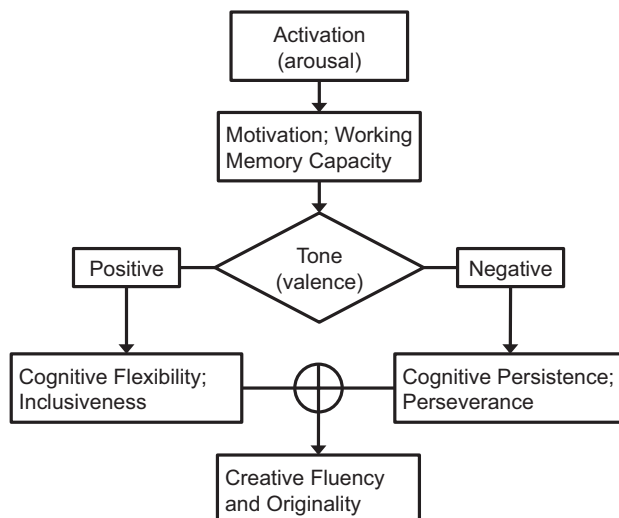


Figure 1: Schematic of the dual pathway to creativity model (reproduced from (De Dreu et al., 2008))

technologies and techniques were chosen for our study, and how certain concepts influenced the subsequent analysis of our data-set.

3. The Study

Incorporating methods and techniques from the research discussed above, we designed an exploratory study to gather subjective and continuous quantitative measures from pairs of co-present, improvising drummers. Co-present interaction was chosen due to our interest in non-verbal communication and the effects of presence. To investigate the importance of visual contact between participants we decided to use two conditions - one where the participants were visible to each other, and one where they were separated by a screen. Under both conditions the pairs of drummers performed two 5-10 minute improvisations. This was deemed long enough for them to engage with the task and improvise a range of rhythms, without becoming bored or fatigued. Improvisation was chosen due to our interest in musical creativity. Our study was not founded on specific hypotheses, instead our aims were to:

Assess challenges and issues associated with the experimental use of behavioural and affective sensors to investigate collaborative music making.

Report exploratory findings to guide and inspire future research.

Identify which measurements and features are most suitable for the investigation of collaborative music making.

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 visually, through movement. There is also an absence of melodic content, which might otherwise influence the participants’ affective responses. We simplified the experiment further by requiring that each participant only used one hand to drum on a single drum pad. This also enabled us to use the non-drumming hand to collect physiological data, using sensors attached to the fingers. The following sections provide more detailed descriptions of the study design and data collection methods, as well as the steps involved in processing this data.

3.1. Research Design and Data Collection

We used a within-subjects design, where each pair of drummers performed drumming tasks under each of the two conditions - visual (V) and non-visual (NV). This design was selected due to high individual variability in drumming ability, behaviour, and affective responses. Throughout the remainder of this paper we use the word “dyad” to refer to a specific pairing of two drummers. We then distinguish each drummer within a dyad as participant 1 or 2. We use the notation $Dx.py.C$ to indicate the dyad (D), participant (p), and condition (C).

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 dyad knew each other, and three of the dyads had previously played music together. We specifically chose to pair people who were known to each other, since we believed that this would maximise the amount of non-verbal communication that we could observe. However, it is beyond the scope of this study to include inter-participant relationship variables as part of the analyses. The participants were aged 26 to 34 ($M=29.1$, $SD=3.1$), they had been playing percussion for between 1 and 17 years ($M=7.4$, $SD=5.0$), and their self-rated 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). Participants were offered £20 as an incentive and signed a consent form before partaking in the study. The study was given ethical approval (QMREC2013/48)².

3.1.2. Measures

We chose to collect a wide range of measurements so that we could evaluate different techniques and explore correlations between various types of data and extracted features. To measure heart rate and perspiration we used small

²Queen Mary Research Ethics Committee reference number.

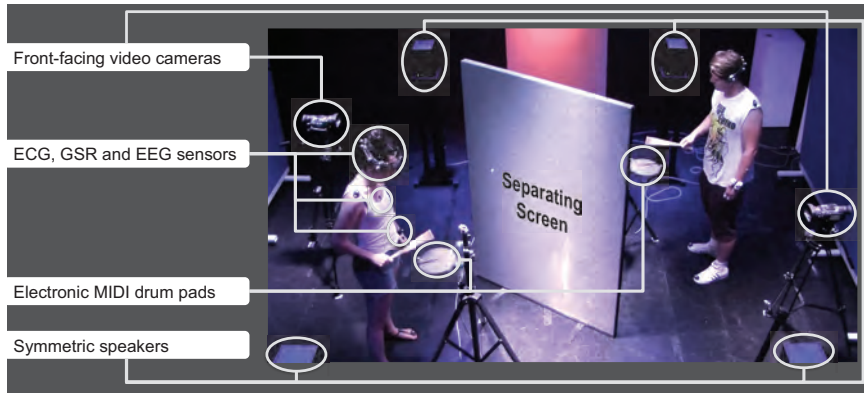


Figure 2: Image taken from the overhead camera illustrating the setup and the equipment used in the study

($53mm \times 32mm \times 19mm$) wireless ECG and GSR sensors developed 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. 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 2 shows an annotated image taken from the overhead camera.

A post-performance questionnaire (PPQ) was designed to collect self-report (SR) 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 with the other participant, 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 full PPQ is shown in Fig. 3. The first item is loosely based upon the widely used Consensual Assessment Technique (CAT) (Amabile, 1982), which proposes that the best way to measure the creativity of an artefact is to simply ask experts in that field to provide a creativity rating. In our case, we were specifically interested in the participant’s own subjective rating of their creativity. Previous studies have found moderate correlations between self and expert-rated creativity using the CAT (Hennessey et al., 2011). Item 2 is adapted from the Mutual Engagement Questionnaire (Bryan-Kinns and Hamilton, 2012), developed specifically for musical interactions. Items 3, 4 and 6 are closely related to the items activation, valence, and dominance from the Self Assessment Manikin (SAM), which is commonly used in affect research (Gunes and Schuller, 2013). Item 5 is adapted from a questionnaire used to study user experience with entertainment technologies (Mandryk et al., 2006; Mandryk and Atkins, 2007).

metronome click track at a tempo of 110 bpm. The second warm up task (~ 1 min) required them to repeat a set rhythmic phrase, which they listened to and learnt prior to the task. These initial tasks allowed the participants to get used to playing the electronic drum and to drumming with one another. The improvisation task (*Improv*, ~ 6 -10 min) required the participants to improvise with one another, where the only restriction was that they did not use verbal communication. Verbal communication was prohibited because we wanted to simulate a live performance environment, where musicians would not normally use verbal communication. Following completion of the drum tasks, the participants sat individually and watched the overhead videos of their two improvised performances. After each minute³ of video they were asked to complete all the SR 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, MIDI, and self-report (SR) data were imported into MATLAB. For each dyad 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 task, relative to t_0 . For each data source these time points were used to extract and label data-sets corresponding to measurements for each participant within each task and visibility condition.

3.2.2. Feature Extraction

Features were extracted from individual data-sets according to the type of data they contained. We initially segmented the data from the *Improv* tasks according to the 1 or 2 minute time windows used for the self report questionnaires. We then extracted features from each of these windows ($Improv_w$), for each participant within each condition (NV or V). This was done to enable us to analyse relationships between extracted features and self report measures, as detailed in the following section. We found that some of the EEG, ECG, and GSR data contained artefactual readings, due to movement and poor electrode contact. We manually labelled these noisy data using a binary coding (1 = noisy, 0 = clean), so that it could be recognised and automatically excluded from our subsequent analysis.

ECG: We used ECGtools⁴ to filter the raw ECG data (sampled at 51.2 Hz) 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, SD, maximum, minimum, the positions of maxima and minima, and the number of extrema divided by the task duration.

³For dyad 5, 2 minute segments were used because the *Improv* tasks were longer.

⁴<http://www.ecgtools.org/>

GSR: Skin conductance response (SCR) has been shown to be a useful metric in analysis of GSR data (Kim et al., 2004; Kim and André, 2008). We used Ledalab⁵ to extract the timing and amplitude of SCR events from the raw GSR data (sampled at 5 Hz) using Continuous Decomposition Analysis (CDA). Again, interpolation was performed and the mean, SD, positions of maxima and minima, and number of extrema divided by task duration, were calculated from the SCR amplitude series.

EEG: EEG signals contain frequency components that relate to the firing activity of neurones in the brain. Standard Theta, Alpha, and Beta frequency band power values are often computed in EEG studies, as they provide information on cognitive activity (Chaouachi and Frasson, 2010). Using EEGLab (Delorme and Makeig, 2004) we initially bandpass filtered the EEG signal (sampled at 128 Hz) between 3 and 30 Hz. We then performed manual artefact rejection to remove noisy segments of data caused by head and facial muscle movements. The Emotiv EEG recordings consist of 14 channels of data, relating to sensors at different positions on the scalp. 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).

Motion: We took the accelerometer readings from the ECG, GSR and EEG sensors and summed the absolute values of the axial components for each sensor. This gave us approximate mean quantity of motion (QoM) values for the head (EEG), torso (ECG), and non-drumming hand (GSR). QoM has previously been shown to be one of the most successful motion features for classifying gestural representations of emotions (Castellano et al., 2007).

MIDI: The number of beats per second, SD in time between consecutive beats, and mean velocity were computed as MIDI features. We compared the beat onset times for participants in each dyad ($tP1$ and $tP2$) and considered any beats that occurred within 70ms of each other to be perceptually synchronous rhythmic events, as suggested by Dixon (2001). For these beats we calculated the time difference ($tP1 - tP2$). We then found the mean over all the absolute difference values. This provides an indicator of the timing synchrony within the dyads. The MIDI data was also used to manually annotate the rhythmic change points (RCPs) for each participant. RCPs were defined as the time points in seconds (relative to t_0) when the participant changed their rhythm from a previously established pattern - defined as a fixed-length sequence of beats, which had been repeated at least twice. The identification of RCPs is illustrated in Fig. 4, which shows a section of MIDI data from one of the participants. We can see that the position of the MIDI notes on the score, as well as the velocity

⁵<http://www.ledalab.de/>

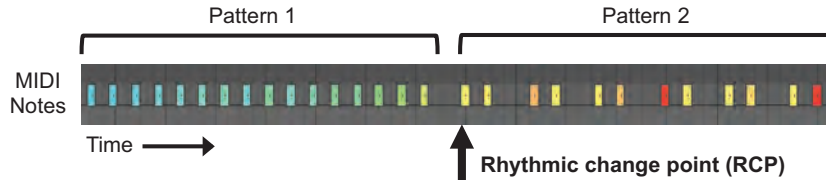


Figure 4: A short segment of MIDI data for one of the participants, illustrating how rhythmic change points were identified. The blocks represent individual beats, and their colour denotes the velocity (strength) of the beat. There are two distinct patterns. Pattern 1 consists of soft (blue-green) and evenly spaced beats, and Pattern 2 consists of paired beats, played with more velocity (yellow-red).

of the notes (denoted by colour) helps identify two distinct rhythmic patterns. In addition to visual analysis we were also able to listen to the rhythm in order to identify RCPs.

Video: We have already described how video footage was used as an elicitation tool to help the participants provide PPQ responses (see Section 3.1.2). In existing studies of conversational interactions video footage has also been used to annotate gaze behaviour (Oertel et al., 2012; Cummins, 2012). We used the footage from the front-facing video cameras to manually annotate the time points when a participant was glancing towards their collaborator during the *Improv* tasks. The start and end time of each glance was recorded in seconds (relative to t_0). We then used these data to calculate the number of glances, percentage glance time, number of mutual glances, and percentage mutual glance time, within each time window ($Improv_w$).

4. Analyses & Results

The aim of our analyses was to use our rich data-set to undertake an exploratory investigation of dyadic, collaborative music making. To facilitate this we structured our analyses over three levels - i) individual; ii) dyad; and iii) whole study. This allowed us to analyse general trends and observations over the entire study whilst also paying attention to interesting and suggestive idiosyncrasies at the level of the individual and the dyad. Sample videos of the performances are provided online.

4.1. Individual-Level Analyses

4.1.1. Data visualisation

At the level of the individual we were particularly interested in how specific events within the improvisation tasks might be linked to observable changes in the continuously captured physiological and motion-based measures. This interest was inspired by previous studies identifying patterns of physiological

Table 1: A summary of the measures, sensors, and features used in this study.

Measure	Sensor	Features
Self Report	PPQ responses (video elicited)	Creativity, engagement, energy, positivity, boredom, leadership. (see Fig. 3)
ECG	Shimmer ECG sensors	Heart rate (HR); mean and SD of HR; positions of HR extrema; number of extrema per min.
GSR	Shimmer GSR sensors	Skin conductance response (SCR); mean and SD of SCR; positions of SCR extrema; number of extrema per min.
EEG	Emotiv EEG headset	Mean power in Theta (4-7 Hz), Alpha (7.5-12.5 Hz), L-Beta (12.5-25 Hz) and H-Beta (25-30 Hz) bands.
Motion	Accelerometers in the Shimmer sensors	Mean quantity of motion (QoM - head, torso, hand).
MIDI	Roland V-Drums	Beats per second; SD in time between consecutive beats; mean velocity; mean absolute inter-beat lag between participants; rhythmic change points (RCPs).
Video	Front-facing and overhead cameras	Number of glances; % glance time; number of mutual glances; % mutual glance time.

change in response to musical events (Egermann et al., 2013), and computer game events (Ravaia et al., 2006). Within the context of our study, the two main event based measures that we obtained were the *rhythmic change points* (RCPs), and *phases of glance* (see Section 4.3) for each participant. We refer to these as continuous features since they derive from continuous measures - audio/midi and video. Figure 5 illustrates a simple time-series plot that we designed to enable us to visualise these events alongside other continuous features. By generating such a plot for each dyad within each condition, we were able to visually explore potential patterns within, and relationships between, continuous measures.

4.1.2. Relationships between heart rate and RCPs

One of the most interesting observations to come from our visual analysis was that RCPs for three of the participants seemed to be closely aligned with extrema (peaks and/or troughs) in the participant’s heart rate plot. To test whether this relationship was significant we extracted the heart rate extrema time points, for each participant, within each condition, and calculated how many of the corresponding RCPs were aligned. We arbitrarily defined points as

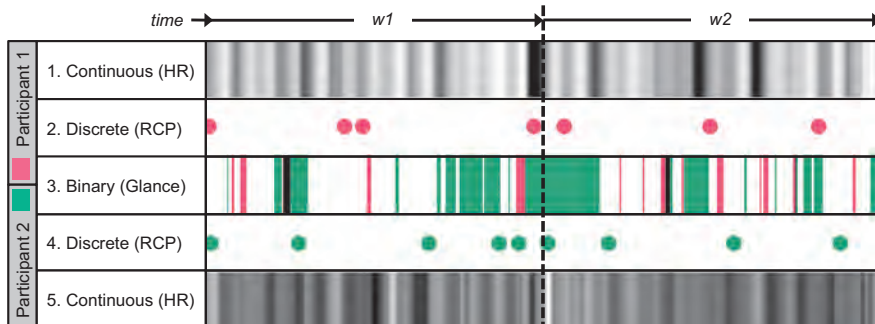


Figure 5: Strip plot of individual participant data for a two window ($w1$ and $w2$) segment of a single improvisation task (under visual-contact condition). Rows 1 and 5 show continuous heart rate data (lighter colours represent higher values). Rows 2 and 4 represent discrete rhythmic change points (RCPs). Row 3 shows glance, which is binary data (glancing or not-glancing) shaded according to the participant. Black shading represents a mutual glance.

aligned if they fell within a quarter of a second of each other. Figure 6 shows an example of the extracted extrema and RCPs, which in this case appear to be aligned to heart rate minima.

Our null hypothesis was that the number of aligned points would not be significantly greater than the chance outcome. In order to estimate the likelihood of a given number of alignments, we considered each set of RCPs as a randomly distributed set of discrete time points, each of which has a probability of being aligned with an extrema point, defined by:

$$P(\text{aligned}) = \frac{n_e(W/res)}{T_s/res} \quad (1)$$

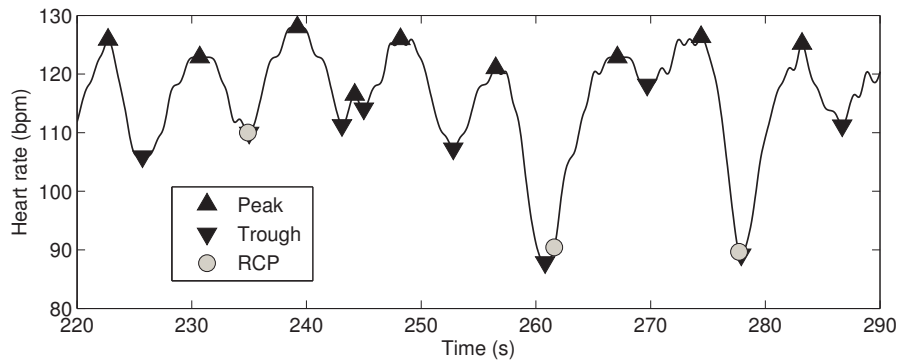


Figure 6: Plot of a segment of heart rate data showing detected extrema and corresponding timing of RCPs.

The numerator in Equation 1 represents the total number of time points on which an RCP could be considered aligned, where n_e is the number of extrema, W is the window size, and res is the time resolution (seconds between samples). The denominator represents the total number of time points, where T_s is the length of the task in seconds. Our window size and resolution were 0.5 and 0.1 seconds respectively. These values were chosen based upon the precision of our heart rate data and expected error due to RCP annotation inaccuracies. Our p-values were subsequently calculated using the complementary cumulative distribution function (CCDF):

$$p = 1 - F(n_a | n_r, P(\text{aligned})) \quad (2)$$

Where F is the binomial CDF, Equation 2 gives the one-tailed probability of obtaining *at least* n_a aligned points, given the total number of RCPs (n_r) and the probability ($P(\text{aligned})$) of a single alignment. This is a suitable statistical test for our null hypothesis, since we consider the number of aligned points to be randomly distributed, with a known probability distribution. Table 2 presents the results of this analysis. We can see that four of the participants show significant alignment percentages, where:

$$\%_a = 2(n_a)/(n_r + n_e) \quad (3)$$

For each of the participants with significant alignments in Table 2, we plotted their average heart rates over the 7 seconds preceding and following all of their

Table 2: Percentages of rhythmic change points (RCPs) and heart rate extrema that are aligned for individual participants within each condition.

Dyad	Par	% Aligned RCPs and Extrema (% _a)		
		all extrema	minima	maxima
<i>Non-visual:</i>				
1	1	13.5***	7.4	14.8***
1	2	8.6	5.8	9.3
2	1	4.9	4.4	4.4
3	1	4.5	2.1	6.2
3	2	7.2	5.3	7.9
5	1	5.0	8.9**	0.0
5	2	5.6	7.1	2.8
<i>Visual:</i>				
1	1	6.6	6.8	4.5
1	2	7.3	1.4	11.5*
2	1	4.0	2.4	4.9
2	2	7.5	6.9	7.0
3	1	4.5	2.7	5.4
3	2	11.9**	13.6**	6.9
5	1	2.4	4.4	0.0
5	2	1.9	1.2	2.3

Note: Results of one-tailed cumulative CDF: * $p < .05$, ** $p < .01$, *** $p < .001$. Data is absent for D4, and D2.p2.NV due to sensor failure.

RCPs. We performed this analysis because it allowed us to look in more detail at how individual participants' heart rates changed on average leading up to, and preceding RCPs. This method was successfully used by Egermann et al. (2013) and Ravaja et al. (2006) to reveal time-windowed patterns in event-based physiological responses. The results are shown in Fig. 7. For D1.p1.NV, heart rate rises in the pre-RCP phase and peaks around 2 seconds after the RCP. This moderately corresponds with our results in Table 2, where we showed a significant percentage of maxima alignments for this participant. The strength of the correspondence may be diluted by the fact that we have averaged the HR across all RCPs, and this participant also had a significant alignment percentage for all extrema. For D1.p2.V the changes in HR are also subtle, however the mean HR does appear to rise prior to the RCP and drop immediately after. This supports our significant results for maxima alignments in the visual condition. For D3.p2.V we see that the mean HR peaks around 1s prior to the RCP, and troughs around 3s after the RCP. This does not correspond well with the results in Table 2. Again, this could be a result of the averaging effect, as we can see that this participant also had significant alignments with all extrema. For D5.p1.NV we can see that a substantial drop in HR occurs during the Pre-RCP phase, followed by an immediate rise during the Post-RCP phase. This strongly supports the significant result for non-visual minima alignments

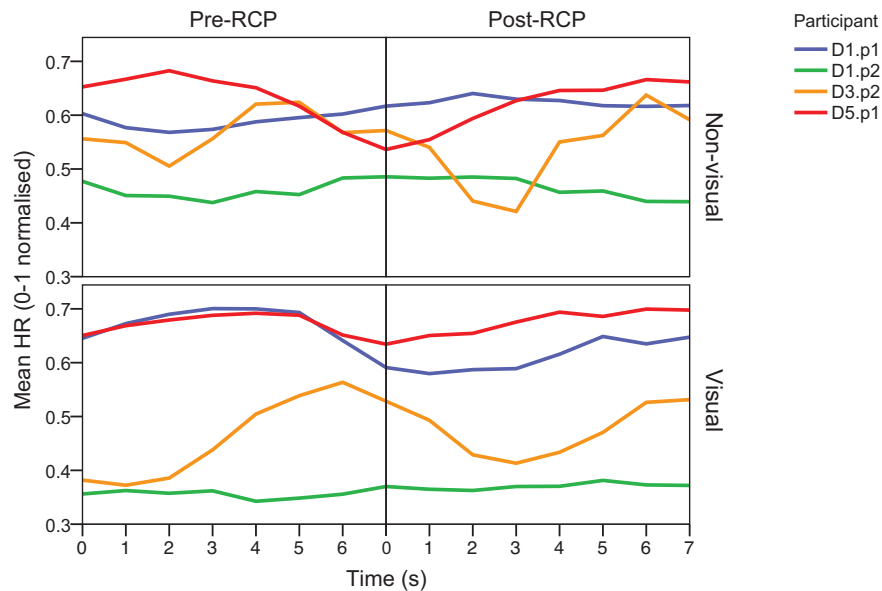


Figure 7: Plots of the mean heart rates (HR) during the 7 seconds before and after RCPs for participants with significant RCP-extrema alignments.

presented in Table 2. In this case the participant had no corresponding maxima alignments in that condition, which might be why we see such a pronounced minima alignment in the plot.

4.2. Dyad-Level Analyses

4.2.1. Correlations between self report scores

We were interested in how well participants’ self report scores correlated within dyads. The reason for this is that it provides an indication of the level of within-dyad agreement with respect to the subjective evaluation of collaborative music making experiences. We used the windowed SR data in order to perform one-tailed, pairwise Spearman correlation analysis between sets of equivalent SR scores within dyads. Spearman correlation is an appropriate test because we cannot assume that linear correlations exist, nor that the variables are normally distributed. The results are shown in Table 3, along with the mean correlation across dyads for each SR item. We can see that the majority of correlations are weak, indicating that participants did not generally agree with each other on the subjectively rated aspects of their performances. Correlations for the item *leadership* are strongest (mean $r_S = -0.46$), with significant results for dyad 3 ($r_S(8) = -0.72, p < .01$) and dyad 4 ($r_S(10) = -0.54, p < .05$).

4.2.2. Heart rate synchrony

For the HR measurements we were interested in undertaking a more focused analysis of correlations between the continuous time series data for each participant across entire *Improv* tasks. This interest was inspired by the studies on emotional contagion and physiological linkage reviewed in Section 2, which suggest that aspects of co-present experiences might be apparent in physiological measurements. As a starting point for our analysis, we were curious as to whether synchronised changes in HR occurred within dyads, as this would serve as a basic indication of potential physiological linkage. The requirement for synchrony is that a change in one participant’s HR should be accompanied by a corresponding change in the other participant’s HR. Therefore, to carry out this analysis we differentiated each participant’s continuous HR data and performed within-dyad comparisons of the slope at each time point. If two participants’ heart rates shared the same direction of change (positive or negative slope) then we labelled that time point as ‘in-sync’. We were then able to calculate the proportion of synchrony between participants. The results are shown in Table 4, where we see that the values all lie close to the chance value of 50%, since 0% and 100% correspond to perfect negative and positive synchrony respectively. This suggests that there was no synchrony between HR changes within dyads. Consequently, we decided not to explore a more detailed analysis of heart rate synchrony.

4.3. Study-Level Analyses

In order to perform our analyses at the level of the entire study, we had to take into account the nested structure of our study. Each windowed data

Table 3: Spearman correlation coefficient (r_s) values for significant within-dyad correlations between self report ratings

Data	Feature	rs within dyads (df)					Mean
		D1 (10)	D2 (10)	D3 (8)	D4 (10)	D5 (10)	
SR	Creativity	0.28	0.37	-0.24	0.29	0.02	0.15
-	Engagement	0.41	-0.55	0.42	0.33	-0.12	0.10
-	Energy	-0.07	0.08	-0.11	0.54*	0.35	0.16
-	Positivity	0.56*	0.11	0.00	0.12	0.12	0.18
-	Boredom	-0.38	0.18	0.15	0.49	-0.27	0.04
-	Leadership	-0.09	-0.49	-0.72**	-0.54*	-0.46	-0.46

Note: Results of Spearman's rho: * $p < .05$, ** $p < .01$. df = degrees of freedom.

Table 4: Synchrony in continuous heart rate data between participants, within dyads. Expressed as the percentage of time where both participants’ heart rates were changing in the same direction (positive or negative slope).

Condition	% Synchrony				Mean
	D1	D2	D3	D5	
Non-visual	46	44	51	53	49
Visual	52	52	49	55	52
Mean	49	48	50	54	50

point comes from an individual, nested within one of the two conditions. These individuals are, in turn, nested within dyads, which are nested under the entire study group. A popular approach to the analysis of data that is structured over multiple levels is the use of linear mixed models (LMMs). Numerous studies have adopted LMMs for the analysis of continuous and subjective human measurements (Glowinski et al., 2013; Egermann et al., 2013; Baayen et al., 2008; Ravaja et al., 2006). Linear mixed models estimate the relationship between a dependent variable and associated covariates by taking into account both fixed and random effects. They also allow for missing data points for subjects (West et al., 2007), which is the case with our data-set.

We used the MIXED procedure in SPSS (Version 22) statistical analysis software to model the relationships between pairs of continuous features. For the random effects we specified random slopes and intercepts for participants. Random effects were not included at the level of the dyad, since dyad association did not generally have systematic effects on the individual participant responses. The analysis was carried out with data across both conditions, as well as within conditions. For the former, we specified in the random effects model that the covariates were nested within the condition. The Restricted Maximum Likelihood (REML) method was used for parameter estimation. To select the best covariance structure for each model (either *variance components*, or *unstructured*) we used likelihood ratio tests to compare the -2 Restricted Log Likelihood values for the models obtained using each of the two covariance structures. If the two models were not significantly different ($p < .05$) then we selected the simpler covariance structure (variance components). The covariance structure used for each analysis is specified in the notes below the corresponding results table.

4.3.1. Correlations between continuous features

Much like our analysis at the level of the individual in Section 4.1, we were interested in how glance and RCP events related to changes in other continuous measures. Table 5 presents the results obtained in our LMM analysis (see Appendix A.1 for more details). Since there are many potential combinations of our 28 continuous features, we only present the significant results. All of the results come from the analysis within the visual condition. We see that the mean absolute inter-beat lag is significantly negatively correlated with the number of individual glances ($t(29) = -2.37, p < .05$) and mutual glances ($t(27) = -2.79,$

Table 5: Linear mixed effects modelling (LMM) estimates of fixed effects between continuous features.

Parameter	Dependent Variable	Estimates of Fixed Effects		
		Estimate	Std. Error	df
No. of glances	Mean abs. inter-beat lag	-0.0006	0.0002	28.8290
No. of mutual glances	Mean abs. inter-beat lag	-0.0011	0.0004	27.1060
No. of RCPs	Mean HR	-0.9204	0.4202	26.2340

Note: t = t-value, df = degrees of freedom. Significance estimates: * $p < .05$, ** $p < .01$. All estimates calculated using *variance components* (VC) covariance type.

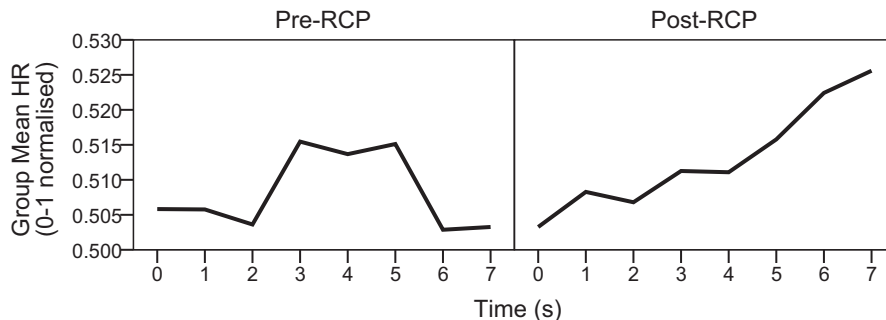


Figure 8: Plot of the mean heart rate (HR) for all participants during the 7 seconds before (Pre-RCP), and after (Post-RCP) all of the RCPs from both conditions.

$p < .01$). Mean HR is significantly negatively correlated with the number of RCPs ($t(26) = -2.19$, $p < .05$). The first two results are closely related, since there is a collinearity between the number of mutual glances and the number of individual glances. This is due to the fact that mutual glances occur when the individual glances of two participants within a dyad are concurrent.

In a similar manner to our individual level analysis in Fig. 7, we decided to plot the study-level mean heart rate during the 7 seconds preceding, and following RCPs. This was calculated as the mean of the set of all the 15 second normalised HR recordings centred upon every corresponding RCP. The results can be seen in Fig. 8, which shows that in the 7 seconds following rhythmic change (post-RCP) there is a fairly linear average increase in HR. For the pre-RCP phase we see a slight rise and fall in HR occurring in the 5 seconds leading up to a RCP.

4.3.2. Correlations between continuous features and self report scores

Using the same LMM procedure as above (see Appendix A.2 and Appendix A.3), we performed an analysis of relationships between continuous features and self report scores. Again, the analysis was performed across both conditions, and within conditions. Our aim was to explore which continuous measures might be suitable as predictors of subjective experience. Significant ($p < 0.05$) correlations are presented in Table 6. We see that across the entire analysis all of the self report items have at least one associated continuous measure that gives a significant correlation. The mean body quantity of motion is positively correlated with creativity ($t(37) = 3.59$, $p < .001$), engagement ($t(31) = 3.01$, $p < .01$), and energy ($t(33) = 3.16$, $p < .01$). There are significant correlations between the beta-band EEG features and leadership. Again, these correlations are strongest in the NV condition, with leadership correlating with mean L-Beta power ($t(21) = 3.33$, $p < .01$), and mean H-Beta power ($t(34) = 3.25$, $p < .01$). In the NV condition mean H-Beta power is also positively correlated with positivity ($t(22) = 2.70$, $p < .05$), and negatively correlated with boredom ($t(21) = -2.30$, $p < .05$). Leadership is also strongly correlated with the number

Table 6: Linear mixed effects modelling (LMM) estimates of Fixed Effects between self report measures and continuous features.

Parameter	Dependent Variable	Estimates of Fixed Effects			t
		Estimate	Std. Error	df	
<i>Both conditions:</i>					
Creativity	Mean body QoM	0.10	0.04	68.28	2.44*
Engagement	Mean body QoM	0.10	0.04	57.67	2.41*
Energy	Mean body QoM	0.12	0.05	54.33	2.56*
Leadership	Mean L-Beta power	0.35	0.17	24.03	2.12*
-	Mean H-Beta power	0.49	0.16	31.53	3.15**
-	No. of beats per sec.	0.17	0.06	83.45	2.75**
<i>Non-visual:</i>					
Creativity	Mean body QoM [†]	0.21	0.06	37.01	3.59***
Engagement	Mean body QoM	0.17	0.06	30.52	3.01**
Energy	Mean body QoM	0.23	0.07	33.34	3.16**
Positivity	Mean H-Beta power	0.66	0.25	21.64	2.70*
Boredom	Mean H-Beta power	-0.49	0.21	21.34	-2.30*
Leadership	Mean L-Beta power	0.63	0.19	20.86	3.33**
-	Mean H-Beta power	0.81	0.25	34.05	3.25**
-	No. of beats per sec.	0.18	0.08	20.05	2.37*
<i>Visual:</i>					
Energy	Mean body QoM	0.15	0.06	15.85	2.44*
Engagement	Mean velocity	-4.14	1.56	35.99	-2.66*
Boredom	No. of SCR extrema	-0.02	0.01	27.10	-2.77**
-	% Glance time	6.60	3.14	26.16	2.10*
Leadership	No. of beats per sec.	0.21	0.10	41.27	2.06**

Note: t = t-value, df = degrees of freedom. Significance estimates: * $p < .05$, ** $p < .01$, *** $p < .001$. Estimates calculated using *variance components (VC)* covariance type or [†]*unstructured* covariance type.

of beats played, with the strongest correlation in the cross-condition analysis ($t(83) = 2.75, p < .01$).

The visual condition analyses result in the fewest correlations, with three results that are exclusive to this condition. Firstly, engagement is negatively correlated with the mean MIDI velocity ($t(36) = 2.66, p < .05$). Secondly, there is a significant correlation between percentage glance time and boredom ($t(26) = 2.10, p < .05$). Finally, the only GSR-related correlation arises here, with a significant negative relationship between the number of SCR extrema and boredom ($t(27) = -2.77, p < .01$).

4.3.3. Effects of visibility condition

Our final study-level analyses were carried out to evaluate the potential effects of the visibility condition on both subjectively reported and continuously measured aspects of the improvised performances. We used a similar LMM approach to that adopted above, with visibility condition set as a factor, and random slopes and intercepts for participants (see Appendix A.4). Significant ($p < .05$) and noteworthy ($p < .15$) results are shown in Table 7, where we see that the only significant effects of visibility condition are on the creativity SR item, $t(9) = -2.26, p < .05$. The negative t-value indicates that participants felt more creative in the visual condition, than the non-visual condition.

5. Discussion

We believe that this study provides three distinct contributions to research on collaborative music making and the use of affective and behavioural sensors in studies of human interactions. Firstly, the techniques, challenges and issues associated with our novel study *design* serve as a valuable reference for researchers planning to conduct similar studies. In addition to this, our *findings* pose interesting research questions, which relate to existing concepts, and serve as inspiration for future studies. Finally, our assessment of various *measures* will assist researchers in deciding which measures and associated sensor technologies are best suited to their particular studies. The following sections discuss these three contributions of our study in more detail.

5.1. Design

We designed a study to assess the challenges and issues associated with the experimental use of behavioural and affective sensors to investigate collaborative music making. The conflict between statistical and ecological validity is a noteworthy challenge for researchers attempting to undertake such studies. In our case we are unable to account for the ways in which the experimental environment and measuring devices may have influenced the participants. In Section 2 we saw how researchers investigating musical improvisation and ensemble interactions have chosen to tolerate unnatural and highly controlled experimental settings, such as the confines of an MRI scanner; and abstract tasks, such as finger tapping. We chose a one-handed improvised drumming task because we

Table 7: Linear mixed effects modelling (LMM) estimates of Fixed Effects of visibility condition on all self report measures and noteworthy ($p < .15$) continuous measures.

Data	Feature	Estimates of Fixed Effects of Visibility Condition			
		Estimate	Std. Error	df t	
SR	Creativity	-0.66	0.29	9.27	-2.26*
-	Engagement	-0.89	0.55	17.79	-1.64
-	Boredom	0.44	0.26	9.35	1.69
ECG	No of HR extrema	-0.02	0.01	60.39	-1.63
GSR	No of SCR extrema	-0.02	0.01	55.00	-1.60
MIDI	No of RCPS	0.41	0.26	95.08	1.54
Motion	Mean body QoM	-0.38	0.21	7.07	-1.76
-	Mean arm QoM	0.49	0.25	4.88	1.94

Note: t = t-value, df = degrees of freedom. Significance estimates: * $p < .05$. All estimates calculated using *variance components* (VC) covariance type.

felt it was representative of a basic on-the-fly musical collaboration, whilst also satisfying practical requirements. However, the specificity of this task means that we must take caution in generalising our findings to other forms of musical interaction.

An additional issue, specific to our design, was the ordering of the visual and non-visual conditions. Our decision to use a fixed order was made partly because randomised ordering would not have been effective with our small study size. In addition to this we thought that holding the V condition first would have allowed participants to use non-verbal communication to appraise aspects of their performance. These appraisals could then have had a noteworthy influence on how the participants performed in the second improvisation. As a consequence of this decision, our results may have been influenced by ordering effects. We highlight the potential presence of these effects when discussing specific findings below.

5.2. Findings

The level at which various analyses should be performed is an important consideration for researchers investigating dyadic interactions. It depends both on the nature of the data, and of the questions posed by the researcher. Our broad analyses were performed at the level of the individual, dyad, and entire study. At the level of the individual we looked at how specific events in the performances related to continuous measures. This allowed us to reveal correlations that may not have emerged during study-level analyses, due to cross-participant random effects such as personality factors and musical ability. Conversely, study-level correlations between subjective and continuous features may not have achieved significance at the level of the individual, due to small sample sizes. Our findings are discussed in more detail below, according to their level of analysis.

5.2.1. Individual-level

At the level of the individual, our visual analysis of the continuous data resulted in us pursuing a more detailed analysis of potential relationships between HR extrema and RCPs. The results in Table 2 and Fig. 7 indicate that the extent of these relationships varies a great deal between participants. For the majority of participants the relationship is weak or non-existent. However, for some participants it is significant. A plausible explanation for these alignments is psychophysiological linkage. The reasoning behind this is that heart rate is linked to arousal and stress, and participants are most likely to change rhythm when their arousal is too high (e.g. the rhythm is too challenging) or too low (e.g. the rhythm has become boring). Another interesting consideration is that HR and RCP alignments are a representation of the participant attempting to maintain a state of flow, whereby they never become overly challenged, or overly bored by the task. The alignments for D5.p1 are especially interesting, as they appear to correspond solely with HR minima. During an informal post-performance interview this participant commented that she had been particularly nervous about playing in the presence of the other participant,

who she knew well. She also had 11 years less drumming experience than her collaborator. It is possible that this anxiety could have contributed to a pronounced rise in HR following new rhythmic contributions. This lends support to our suggestion that HR and RCP alignments were a result of psychophysiological linkage. Furthermore, it highlights the possibility that other factors, such as the participants' personalities, expertise, and relationships, may have influenced these findings. A more focused study is required in order to verify whether a causal relationship exists between HR extrema and on-the-fly creative musical decisions. In particular, it would be worth considering whether there are other performance-related events that also coincide with HR extrema (e.g. changes in the musical dynamics or specific non-verbal exchanges between musicians).

5.2.2. Dyad-level

Our analysis of correlations between SR scores at the level of the dyad indicated that participants did not generally agree upon aspects of their experiences, and the creativity of the performances. These results are similar to those of Schober and Spiro (2014), who found that two jazz musicians did not have a high level of agreement on self reported evaluations of their performances, despite the performances being judged to be successful improvisations. The most significant and consistent correlations were for the *leadership* item. This is a potentially interesting result as it suggests that dyads were more attuned to the functional leader-follower aspects of their interaction than they were to the affective aspects, such as energy and positivity. It has been suggested that the establishment and communication of leader-follower relationships is an important tool in the creation of dynamic musical collaborations (Reidsma et al., 2014).

Our analysis of HR synchrony within dyads suggested that synchrony did not generally occur for participants in this study. The existence of physiological linkage between co-present individuals is understudied and previous evidence has been found only in resting participants (Reed et al., 2013). In our case, the active nature of the drumming task may have been the dominant factor in determining changes in HR over time.

5.2.3. Study-level

Our study-level analyses of relationships between paired sets of continuous features revealed significant correlations between glance-based features and the mean absolute participant inter-beat lag. These results suggest that glancing at the other participant increases the amount of timing synchrony (decreased lag) between participants. This agrees with existing research, which has shown that visual contact can have a positive influence on timing synchrony between musicians (Vera et al., 2013). We also found a significant negative correlation between the number of RCPs and the mean HR ($t(26) = -2.19, p < .05$). This supports our suggestion that RCPs occur at points of low physiological arousal (see Section 5.2.1). If this is the case, then we would expect the arousal to increase following rhythmic change. Evidence for this can be seen in the plot

in Fig. 8, which shows an average increase in HR in the 7 seconds following rhythmic change.

Our study-level analyses of correlations between continuous features and self report scores revealed that body motion is positively correlated with *creativity* ($t(37) = 3.59, p < .001$), *engagement* ($t(31) = 3.01, p < .01$), and *energy* ($t(33) = 3.16, p < .01$). This aligns with previous research, which found that quantity of movement features can discriminate between low and high arousal emotions (Castellano et al., 2007). The fact that body movement is also correlated with *creativity* lends support to the *dual pathway model* discussed in Section 2.2.2, which highlights the importance of arousal in the generation of creative ideas. H-Beta EEG power is significantly correlated with *leadership* in both the cross-condition ($t(32) = 3.15, p < .01$) and NV analyses ($t(34) = 3.25, p < .01$). We also found H-beta correlations with *positivity* ($t(22) = 2.70, p < .05$) and negative H-beta correlations with *boredom* ($t(21) = -2.30, p < .05$). Previous studies have associated beta activity with *engagement* and *cognitive challenge* (Budzynski et al., 2008). This concurs with our findings, as we would expect people to be more engaged and cognitively aroused when leading the performance. Within the visual condition, the negative correlation between *engagement* and mean velocity ($t(36) = -2.66, p < .05$) is interesting, since it suggests that participants hit the drum more softly during periods of high engagement. This might be reflective of participants playing more quietly in order to direct their attention towards the other musician’s playing. This result is consistent with descriptions of jazz musicians’ multiple levels of attention, incorporating an awareness of both the self and the other (Sawyer, 2003). Also of interest is the correlation between percentage glance time and *boredom*, which suggests that prolonged glances might be used within the performance to communicate boredom; or that they might be interpreted as indicators of boredom when reviewing the video recordings of the performance. The same can be said for the negative correlation between *boredom* and the number of SCR extrema, since SCR events have been shown to be concomitant with emotional arousal (Benedek and Kaernbach, 2010).

The lack of significant correlations between *creativity* and EEG features is interesting because it may be indicative of the use of contrasting thought processes during creative action. This concurs with previous EEG research, which has struggled to show conclusive links between creativity and localised brain activity (Dietrich and Kanso, 2010). Indeed, the literature on creativity often refers to the roles and interplay of different styles of thinking in the development and generation of ideas. For example, the terms *associative* and *analytic* have also been used to define two contrasting modes of thought involved in the creative process (Gabora, 2002; Simonton, 1975).

During our study-level analyses of correlations between continuous features and self report scores we drew attention to the fact that the results for the NV condition yielded more, and stronger correlations than those for the V condition. A potential explanation for this is that participants’ non-verbal behaviours and physiological responses are more closely tied to their personal subjective experiences when they are unable to see the other musician. This idea is influenced

by the theories of emotional contagion and mimicry (see Section 2.2.1), which suggest that being able to see the other participant might influence changes in behaviour and physiology. More work would need to be done to establish whether or not this is an influencing factor when correlating subjective scores with measures of behaviour and affect.

Our results for study-level effects of the visibility condition indicated that the only significant effects of co-visibility were on self reported creativity ($t(9) = -2.66, p < .05$). This indicates that creativity was rated more highly in the V condition. However, we should be cautious of the potential influence of ordering effects on this result. Given the small size of our study, we also reported noteworthy results ($p < .15$). These indicate that self reported engagement was generally lower in the NV condition ($t(18) = -1.64$), whilst boredom was higher ($t(9) = 1.69$). With regard to ordering effects, in this case we would have expected boredom to be rated more highly in the second condition (V). These results support the idea that co-visibility is influential to the subjective experience of collaborating musicians. The HR and SCR extrema features are the only physiological features that show noteworthy visibility condition effects. These features represent the number of phasic shifts in a participant's physiological arousal over time. Our results suggest that there are more shifts during the V condition. This may be reflective of the fact that the participants have more to attend to when they are able to see their fellow musician. This could lead to them shifting their attention more frequently, resulting in corresponding shifts in physiological arousal due to psychophysiological linkage.

As can be noted from the analyses discussed above, we predominantly investigated relationships between pairs of features and conditions. This approach was chosen due to the exploratory nature of our study and the broad array of measures and features collected. A more comprehensive analysis of specific correlations and effects would require the inclusion of other factors, based upon specific hypotheses. Such a multi-factorial approach could yield results that contrast with those presented in this paper. In addition to this, the small sample size used in this study means that our findings should be viewed speculatively in relation to existing studies, and as inspiration for future research.

5.3. Measures

We assess the suitability of our measures based on three factors; *practicality*, *reliability*, and *informativeness*. Practicality concerns how difficult it is to collect the measurements, given the constraints of live musical performance. Reliability addresses the issue of whether the raw measurements are likely to contain reliable information, as opposed to noisy, or artifactual data. Finally, informativeness concerns the amount of useful information that we are able to obtain by extracting features and analysing the measurements. These factors were chosen in light of study aims (see Section 3).

Of the sensor-based measures, ECG and accelerometer data were easy to acquire using wireless sensors, and were not particularly prone to noise. On the contrary, the EEG sensor required a lot of preparation and adjustment

prior to use, and the data required manual processing (artefact removal), and were prone to noise. Regarding the informativeness of sensor-derived features, we found that the extraction of heart rate extrema time points was useful for the analysis of time-based performance events. The bodily quantity of motion feature appeared to be a good indicator of the arousal dimension of subjective experience. Glance related features were acquired by manually annotating video footage, however it is feasible that modern eye-tracking sensors could perform this task automatically. Glance count and average duration appear to be promising features for the analysis of collaborative interactions where the musicians are visible to one another.

Performance related features, acquired through MIDI data, assisted in the identification of rhythmic changes, and provided velocity and timing synchrony features. Timing synchrony was extracted from beats that were perceptually synchronous, so would not have accounted for more pronounced and intentional changes in tempo. Reliable measures of tempo changes are difficult to extract automatically, which is why we chose not to perform tempo analysis. However, this is an important aspect of improvisation and would be worth considering in future studies. One issue is that MIDI data can only be obtained from certain instruments. Alternative methods, such as audio feature-extraction, should be considered for more generalised studies of collaborative music making.

Finally, our self report measures were obtained using a post-performance questionnaire (PPQ). We designed the PPQ to be as concise as possible, due to the fact that each participant had to complete it from 6 up to 10 times. Instead of including multiple questions addressing similar constructs, specific questions were selected and adapted from existing questionnaires used in related studies. Verifying the validity and reliability of questionnaires is a challenge for any researcher attempting to collect subjective measures of collaborative music making. In particular, reliability is difficult to ascertain, given that the unique nature of each collaboration prohibits the collection of repeated measures. Furthermore, the value and general reliability of introspective self report data has been called into question (Pronin, 2009). These issues provide weight to the argument that alternative measures, such as physiology, might be more suitable tools for the measurement of behaviour and affect. The predicament facing researchers is that, in order to interpret and understand how such measurements relate to affective and behavioural phenomena, we must have some existing knowledge about the phenomena we are measuring. This is especially difficult when it comes to investigating collaborative music making - an activity that involves a complex tapestry of contextual and subjective information that changes continuously as the interaction unfolds.

In summary, researchers should take caution in drawing links between self report and continuous physiological and behavioural data. In particular, it is necessary to carefully consider the causal mechanisms and pathways that might explain such links.

6. Conclusions

Our exploratory study provides an insight into the use of affective and behavioural sensors as tools for investigating collaborative music making. We have highlighted some of the challenges of working with these sensors and have addressed experimental design considerations that could be relevant to researchers planning to undertake similar studies. Our detailed analyses led to us reporting a range of interesting findings, which we hope will inspire future research in this field. In particular, there is a need to build upon this work by employing appropriate sensors to test specific hypotheses surrounding the behavioural and affective aspects of collaborative music making.

Acknowledgements

This work was funded by the EPSRC through the Media and Arts Technology Programme (EP/G03723X/1), an RCUK Centre for Doctoral Training. We would like to thank Marcus Pierce⁶ for his advice on the processing of EEG data, and Hugo Maruri-Aguilar⁶, Heiko Grossmann⁶, and Miriam Koschate-Reis⁷ for their advice on statistical methods.

Amabile, T. M., 1982. Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology* 43 (5), 997–1013.

URL <http://content.apa.org/journals/psp/43/5/997>

Amabile, T. M., Barsade, S. G., Mueller, J. S., Staw, B. M., Sep. 2005. Affect and Creativity at Work. *Administrative Science Quarterly* 50 (3), 367–403.

URL <http://asq.sagepub.com/content/50/3/367.short>

Argyle, M., 1978. *The psychology of interpersonal behaviour*. Penguin Books.

Baayen, R., Davidson, D., Bates, D., Nov. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59 (4), 390–412.

URL <http://linkinghub.elsevier.com/retrieve/pii/S0749596X07001398>

Barsade, S., 2002. The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly* (December), 644–675.

Benedek, M., Kaernbach, C., Jun. 2010. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods* 190 (1), 80–91.

URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2892750&tool=pmcentrez&rendertype=abstract>

⁶Queen Mary University of London

⁷University of Exeter

- Bevilacqua, F., Fels, S., Jensenius, A. R., Lyons, M. J., Schnell, N., Tanaka, A., 2013. SIG NIME: Music, Technology, and Human-Computer Interaction. In: CHI '13 Extended Abstracts on Human Factors in Computing Systems - CHI EA '13. ACM Press, New York, New York, USA, p. 2529.
URL <http://dl.acm.org/citation.cfm?doid=2468356.2468822>
- Biasutti, M., Frezza, L., May 2009. Dimensions of Music Improvisation. *Creativity Research Journal* 21 (2-3), 232–242.
- Bränzel, A., Holz, C., Hoffmann, D., Schmidt, D., Knaust, M., Lühne, P., Meusel, R., Richter, S., Baudisch, P., 2013. GravitySpace. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13. ACM Press, New York, New York, USA, p. 725.
URL <http://dl.acm.org/citation.cfm?doid=2470654.2470757>
- Bryan-Kinns, N., Jan. 2013. Mutual engagement and collocation with shared representations. *International Journal of Human-Computer Studies* 71 (1), 76–90.
URL <http://linkinghub.elsevier.com/retrieve/pii/S1071581912000328>
- Bryan-Kinns, N., Hamilton, F., Feb. 2012. Identifying mutual engagement. *Behaviour & Information Technology* 31 (2), 101–125.
- Budzynski, T., Budzynski, H., Evans, J., Abarbanel, A., 2008. Introduction to Quantitative EEG and Neurofeedback: Advanced Theory and Applications, 2nd Edition. Academic Press.
- Carlile, J., Hartmann, B., 2005. OROBORO : A Collaborative Controller with Interpersonal Haptic Feedback. In: *New Interfaces for Musical Expression (NIME05)*. pp. 250–251.
- Castellano, G., Villalba, S. D., Camurri, A., 2007. Recognising Human Emotions from Body Movement and Gesture Dynamics. In: *Affective Computing and Intelligent Interaction (ACII07)*. Springer-Verlag, Lisbon, Portugal, pp. 71–82.
URL <http://www.springerlink.com/index/aj2461vw0635g238.pdf>
- Chaouachi, M., Frasson, C., 2010. Exploring the relationship between learner EEG mental engagement and affect. In: Alevin, V., Kay, J., Mostow, J. (Eds.), *Intelligent Tutoring Systems*. Springer Berlin Heidelberg, pp. 291–293.
- Chartrand, T. L., Lakin, J. L., Sep. 2012. The Antecedents and Consequences of Human Behavioral Mimicry. *Annual Review of Psychology* (September 2012).
URL <http://www.ncbi.nlm.nih.gov/pubmed/23020640>
- Csikszentmihalyi, M., Rathunde, K., 1993. The measurement of flow in everyday life: Toward a theory of emergent motivation. Vol. 40. University of Nebraska Press, Lincoln, NE, US, pp. 57–97.

- Cummins, F., Dec. 2012. Gaze and blinking in dyadic conversation: A study in coordinated behaviour among individuals. *Language and Cognitive Processes* 27 (10), 1525–1549.
- De Dreu, C. K. W., Baas, M., Nijstad, B. a., May 2008. 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.
URL <http://www.ncbi.nlm.nih.gov/pubmed/18444736>
- de Manzano, O., Theorell, T., Harmat, L., Ullén, F., Jun. 2010. The psychophysiology of flow during piano playing. *Emotion (Washington, D.C.)* 10 (3), 301–11.
URL <http://www.ncbi.nlm.nih.gov/pubmed/20515220>
- Delorme, A., Makeig, S., Mar. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* 134 (1), 9–21.
URL <http://www.ncbi.nlm.nih.gov/pubmed/15102499>
- Dietrich, A., Kanso, R., Oct. 2010. A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological Bulletin* 136 (5), 822–48.
URL <http://www.ncbi.nlm.nih.gov/pubmed/20804237>
- Dixon, S., Mar. 2001. Automatic Extraction of Tempo and Beat From Expressive Performances. *Journal of New Music Research* 30 (1), 39–58.
- Edwards, L., Sep. 2013. Ford S-MAX Concept will measure your heart rate from the driver’s seat. Online, (Accessed 25th September 2013).
URL <http://www.pocket-lint.com/news/123510-ford-s-max-concept-will-measure-your-heart-rate-from-the-driver-s-seat>
- Egermann, H., Pearce, M. T., Wiggins, G. a., McAdams, S., Sep. 2013. Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective & Behavioral Neuroscience* 13 (3), 533–53.
- Fencott, R., Bryan-Kinns, N., 2010. Hey Man , you re invading my Personal Space! Privacy and Awareness in Collaborative Music. In: *New Interfaces for Musical Expression (NIME10)*. pp. 198–203.
- Gabora, L., 2002. Cognitive mechanisms underlying the creative process. In: *Proceedings of the fourth conference on Creativity & cognition - C&C '02*. ACM Press, New York, New York, USA, pp. 126–133.
URL <http://portal.acm.org/citation.cfm?doid=581710.581730>
- Glowinski, D., Gnecco, G., Piana, S., Camurri, A., 2013. Expressive Non-Verbal Interaction in String Quartet. In: *Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII'13)*. IEEE Computer Society, Geneva, Switzerland, pp. 233–238.

- Gunes, H., Schuller, B., Jul. 2013. Categorical and Dimensional Affect Analysis in Continuous Input: Current Trends and Future Directions. *Image and Vision Computing* 31 (2), 120–136.
 URL <http://linkinghub.elsevier.com/retrieve/pii/S0262885612001084>
- Hadjakos, A., Groß hauser, T., Goebel, W., 2013. Motion Analysis of Music Ensembles with the Kinect. In: *International Conference on New Interfaces for Musical Expression (NIME)*. KAIST, Daejeon, Korea, pp. 106–110.
- Healey, P. G. T., Leach, J., Bryan-kinns, N., 2005. Inter-Play : Understanding Group Music Improvisation as a Form of Everyday Interaction. In: *Proceedings of Less is More: Simple Computing in an Age of Complexity*.
- Hennessey, B. A., Amabile, T. M., Mueller, J. S., 2011. Consensual Assessment. In: Runco, M., Pritzker, S. (Eds.), *Encyclopedia of Creativity*, 2nd Edition. Vol. 1. Elsevier, pp. 253–260.
- Jamison, K. R., Dec. 1996. *Touched with Fire: Manic-Depressive Illness and the Artistic Temperament*. Simon & Schuster Ltd.
- Keller, P. E., Fabian, D., Timmers, R., Schubert, E., 2014. Ensemble performance: interpersonal alignment of musical expression. Expressiveness in music performance: Empirical approaches across styles and cultures, 260–282.
- Kendon, A., Jan. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26, 22–63.
 URL <http://linkinghub.elsevier.com/retrieve/pii/0001691867900054>
- Kendon, A., 2010. Spacing and Orientation in Co-present Interaction. *Lecture Notes in Computer Science* 5967, 1–15.
- Kenny, B. J., Gellrich, M., Apr. 2002. Improvisation. In: Parncutt, R., McPherson, G. (Eds.), *The Science and Psychology of Music Performance: Creative Strategies for Teaching and Learning*. Oxford University Press, Ch. 8, pp. 117–134.
- Kim, J., André, E., Dec. 2008. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (12), 2067–83.
 URL <http://www.ncbi.nlm.nih.gov/pubmed/18988943>
- Kim, K. H., Bang, S. W., Kim, S. R., May 2004. Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing* 42 (3), 419–427.
- Knapp, R. B., Jaimovich, J., Coghlan, N., Sep. 2009. Measurement of motion and emotion during musical performance. In: *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*.

- IEEE, pp. 1–5.
 URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5349469
- Kuhara, Y., Kobayashi, D., 2011. Kinetic Particles Synthesizer Using Multi-Touch Screen Interface of Mobile Devices. In: *New Interfaces for Musical Expression (NIME11)*. No. June. pp. 136–137.
- Kuyken, B., Verstichel, W., Bossuyt, F., Vanfleteren, J., Demey, M., Leman, M., 2008. The HOP sensor : Wireless Motion Sensor. In: *New Interfaces for Musical Expression (NIME08)*. pp. 229–232.
- Limb, C. J., Braun, A. R., Jan. 2008. Neural substrates of spontaneous musical performance: an FMRI study of jazz improvisation. *PloS one* 3 (2).
 URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2244806&tool=pmcentrez&rendertype=abstract>
- Lundqvist, L.-O., Carlsson, F., Hilmersson, P., Juslin, P. N., Aug. 2008. Emotional responses to music: experience, expression, and physiology. *Psychology of Music* 37 (1), 61–90.
 URL <http://pom.sagepub.com/cgi/doi/10.1177/0305735607086048>
- Mandryk, R. L., Atkins, M. S., Apr. 2007. A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies* 65 (4), 329–347.
 URL <http://linkinghub.elsevier.com/retrieve/pii/S1071581906001881>
- Mandryk, R. L., Inkpen, K. M., Calvert, T. W., Mar. 2006. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & Information Technology* 25 (2), 141–158.
- Mazur, A., Rosa, E., Faupel, M., Heller, J., Leen, R., Thurman, B., Jul. 1980. Physiological aspects of communication via mutual gaze. *American Journal of Sociology* 86 (1), 50–74.
 URL <http://www.ncbi.nlm.nih.gov/pubmed/7435770>
- Mealla, S., Våljamäe, A., Bosi, M., Jordà, S., 2011. Listening to Your Brain: Implicit Interaction in Collaborative Music Performances. In: *The International Conference on New Interfaces for Musical Expression (NIME)*. No. June. Oslo, Norway, pp. 149–154.
 URL http://mtg.upf.es/system/files/publications/listening_to_your_brain_camera_ready.pdf
- Mitchell, T., Madgwick, S., Heap, I., 2012. Musical Interaction with Hand Posture and Orientation : A Toolbox of Gestural Control Mechanisms. In: *New Interfaces for Musical Expression (NIME12)*.

- Monson, I., 1996. *Saying Something: Jazz Improvisation and Interaction*. Chicago Studies in Ethnomusicology. University of Chicago Press, Chicago, IL.
URL <http://books.google.co.uk/books?id=Pus8-ccYW1MC>
- Morgan, E. L., Gunes, H., Harris, D., 2013. Gesturing at Architecture: Experiences and Issues with New Forms of Interaction. In: Proc. of CHI'13 Workshop on Experiencing Interactivity in Public Spaces (EIPS). Paris, France, pp. 117–121.
URL http://www.cs.tut.fi/ihte/EIPS_workshop_CHI13/papers.shtml
- Nowicki, L., Prinz, W., Grosjean, M., Repp, B. H., Keller, P. E., 2013. Mutual adaptive timing in interpersonal action coordination. *Psychomusicology: Music, Mind, and Brain* 23 (1), 6–20.
- Oertel, C., Wlodarczak, M., Edlund, J., Wagner, P., Gustafson, J., Feb. 2012. Gaze Patterns in Turn-Taking. In: *Proceedings of Interspeech*. Vol. 29.
- Pantic, M., Cowie, R., D'Errico, F., Heylen, D., Mehu, M., Pelachaud, C., Poggi, I., Schroeder, M., Vinciarelli, A., 2011. Social signal processing: The research agenda. *Visual Analysis of Humans*, 511–538.
- Patel, M., Lal, S., Kavanagh, D., Rossiter, P., Jun. 2011. Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert Systems with Applications* 38 (6), 7235–7242.
URL <http://linkinghub.elsevier.com/retrieve/pii/S0957417410013916>
- Pentland, A., Mar. 2005. Socially aware, computation and communication. *Computer* 38 (3), 33–40.
URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1413116>
- Price, J., Mar. 2011. Plug in to movies that know how you're feeling. Online, (Accessed 10th December 2012).
URL <http://www.newscientist.com/blogs/culturelab/2011/03/plugin-for-emotional-response-cinema.html>
- Pronin, E., 2009. The Introspection Illusion. In: *Advances in Experimental Social Psychology*. Vol. 41. Elsevier Inc., Ch. 1, pp. 1–67.
- Ravaja, N., Saari, T., Salminen, M., Laarni, J., Kallinen, K., Nov. 2006. Phasic Emotional Reactions to Video Game Events: A Psychophysiological Investigation. *Media Psychology* 8 (4), 343–367.
- Reed, R. G., Randall, A. K., Post, J. H., Butler, E. a., Jun. 2013. Partner influence and in-phase versus anti-phase physiological linkage in romantic couples. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology* 88 (3), 309–16.
URL <http://www.ncbi.nlm.nih.gov/pubmed/22922526>

- Reidsma, D., Radha, M., Nijholt, A., 2014. Mediated Interactions and Musical Expression - A Survey. In: Lee, N. (Ed.), *Digital Da Vinci: Computers in Music*. Springer, New York, NY, USA, pp. 79–98.
- Sacks, H., Schegloff, E., Jefferson, G., 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50 (4), 696–735.
- Sariyanidi, E., Gunes, H., Cavallaro, A., 2014. Automatic analysis of facial affect: A survey of registration, representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8828 (AUGUST 2013), 1–1.
 URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6940284>
- Sawyer, R. K., 2003. *Group Creativity: Music, Theater, Collaboration*. Psychology Press.
- Sawyer, R. K., Jul. 2005. Music and Conversation. In: Miell, D., MacDonald, R., Hargreaves, D. J. (Eds.), *Musical Communication*. Oxford University Press, Ch. 3.
- Schmidt, L. A., Trainor, L. J., Jul. 2001. Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cognition & Emotion* 15 (4), 487–500.
- Schneider, F., Grodd, W., Weiss, U., Klose, U., Mayer, K. R., Nägele, T., Gur, R. C., Dec. 1997. Functional MRI reveals left amygdala activation during emotion. *Psychiatry Research: Neuroimaging* 76 (2-3), 75–82.
- Schober, M. F., Spiro, N., Jan. 2014. Jazz improvisers’ shared understanding: a case study. *Frontiers in psychology* 5 (August), 808.
 URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4126153&tool=pmcentrez&rendertype=abstract>
- Seddon, F. a., Mar. 2005. Modes of communication during jazz improvisation. *British Journal of Music Education* 22 (1), 47–61.
 URL http://www.journals.cambridge.org/abstract_S0265051704005984
- Simonton, D. K., Oct. 1975. Creativity, task complexity, and intuitive versus analytical problem solving. *Psychological reports* 37 (2), 351–4.
 URL <http://www.ncbi.nlm.nih.gov/pubmed/1187912>
- Vera, B., Chew, E., Healey, P., 2013. A Study of Ensemble Performance Under Restricted Line of Sight. In: *Proceedings of the International Conference on Music Information Retrieval*. Curitiba, Brazil.
 URL http://www.ppgia.pucpr.br/ismir2013/wp-content/uploads/2013/09/171_Paper.pdf

- Vinciarelli, A., Pantic, M., Bourlard, H., Nov. 2009. Social signal processing: Survey of an emerging domain. *Image and Vision Computing* 27 (12), 1743–1759.
URL <http://linkinghub.elsevier.com/retrieve/pii/S0262885608002485>
- Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D’Errico, F., Schroeder, M., Jan. 2012. Bridging the Gap between Social Animal and Unsocial Machine: A Survey of Social Signal Processing. *IEEE Transactions on Affective Computing* 3 (1), 69–87.
URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5989788>
- Walter, M., Eilebrecht, B., Wartzek, T., Leonhardt, S., Jan. 2011. The smart car seat: personalized monitoring of vital signs in automotive applications. *Personal and Ubiquitous Computing* 15 (7), 707–715.
- West, B. T., Welch, K. B., Galecki, A. T., 2007. *Linear Mixed Models: A Practical Guide Using Statistical Software*. Chapman and Hall/CRC.
- Wilson, G. B., MacDonald, R. a. R., Aug. 2012. The sign of silence: Negotiating musical identities in an improvising ensemble. *Psychology of Music* 40 (5), 558–573.
- Zhou, H., Stone, T., Hu, H., Harris, N., Jan. 2008. Use of multiple wearable inertial sensors in upper limb motion tracking. *Medical engineering & physics* 30 (1), 123–33.
URL <http://www.ncbi.nlm.nih.gov/pubmed/17251049>

Author Biographies

Evan Morgan is a Ph.D. Student on the Media and Arts Technology programme at Queen Mary, University of London. Prior to this he completed an integrated master's degree (M.Eng) in Biomedical Engineering at Imperial College London. He is interested in using physiological and behavioural measurement technologies as tools for investigating and enhancing co-present human interactions. His research is specifically focused on the area of musical creativity, and incorporates ideas from the fields of affective computing, cognitive science, psychophysiology, and social signal processing.

Dr. Hatice Gunes is a Senior Lecturer at Queen Mary University of London, leading the Affective and Human Computing Lab. Her research is in the areas of automatic analysis of emotional and social behaviour and human aesthetic canons, multimodal information processing and human-robot interactions. She published over 70 technical papers in these areas (Google scholar citations > 1500, H-index = 20) and was a recipient of awards for Outstanding Paper (IEEE FG'11), Quality Reviewer (IEEE ICME'11) and Best Demo (IEEE ACII'09). She serves on the Management Board of Association for the Advancement of Affective Computing and the Steering Committee of IEEE Transactions on Affective Computing.

Dr. Nick Bryan-Kinns is Reader in Interaction Design, and the Deputy Dean for Science and Engineering at Queen Mary, University of London. He holds a Royal Academy of Engineering Industrial Secondment, leads Interactional Sound and Music in the Centre for Digital Music at QM, and has published award winning international journal and conference papers on his funded research on mutual engagement, cross-modal interaction, and tangible interfaces. He was a panel member for the National Science Foundation's CreativeIT funding panel, and provided expert consultation for the European Commission's funding of Creativity and ICT.

Appendix A. SPSS Linear Mixed Model Command Syntaxes

Appendix A.1. Command syntax for the analysis of correlations between continuous features.

```
MIXED MeanInterBeatLagWITH NumMutualGlancemin  
/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001)  
HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)  
/FIXED=NumMutualGlancemin | SSTYPE(3)  
/METHOD=REML  
/PRINT=SOLUTION  
/RANDOM=INTERCEPT NumMutualGlancemin | SUBJECT(Par) COVTYPE(VC).
```

Appendix A.2. Command syntax for the analysis of correlations between continuous features and self-report scores across both conditions.

```
MIXED MeanBody BY Vis WITH Creativity  
/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001)  
HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)  
/FIXED=Creativity | SSTYPE(3)  
/METHOD=REML  
/PRINT=SOLUTION  
/RANDOM=INTERCEPT Creativity(Vis) | SUBJECT(Par) COVTYPE(VC).
```

Appendix A.3. Command syntax for the analysis of correlations between continuous features and self-report scores within a single condition.

```
MIXED Lbeta WITH Positivity  
/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001)  
HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)  
/FIXED=Positivity | SSTYPE(3)  
/METHOD=REML  
/PRINT=SOLUTION  
/RANDOM=INTERCEPT Positivity | SUBJECT(Par) COVTYPE(VC).
```

Appendix A.4. Command syntax for the analysis of effects of visibility condition.

```
MIXED Creativity BY Vis  
/CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001)  
HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)  
/FIXED=Vis | SSTYPE(3)  
/METHOD=REML  
/PRINT=SOLUTION  
/RANDOM=INTERCEPT Vis | SUBJECT(Par) COVTYPE(VC).
```