Novel Methods in Facilitating Audience and Performer Interaction using the Mood Conductor Framework

György Fazekas, Mathieu Barthet and Mark B. Sandler

School of Electronic Engineering and Computer Science
Center for Digital Music
Queen Mary University of London
Email: {firstname.lastname}@eecs.qmul.ac.uk

Abstract. While listeners' emotional response to music is the subject of numerous studies, less attention is payed to the dynamic emotion variations due to the interaction between artists and audiences in live improvised music performances. By opening a direct communication channel from audience members to performers, the Mood Conductor system provides an experimental framework to study this phenomenon. Mood Conductor facilitates interactive performances and thus also has an inherent entertainment value. The framework allows audience members to send emotional directions using their mobile devices in order to "conduct" improvised performances. Audience indicated emotion coordinates in the arousal-valence space are aggregated and clustered to create a video projection. This is used by the musicians as guidance, and provides visual feedback to the audience. Three different systems were developed and tested within our framework so far. These systems were trialled in several public performances with different ensembles. Qualitative and quantitative evaluations demonstrate that musicians and audiences are highly engaged with the systems, and raise new insights enabling future improvements of the framework.

Keywords: audience-performer interaction, music, emotion, mood, arousal, valence, improvisation, live music

1 Introduction

A large body of recent research in the area of music and emotions focusses on listeners' response to music, which may be studied in terms of expressed or *perceived* emotions, and in terms of induced or *felt* emotions. Thorough reviews of these works from the perspective of Psychological research as well as Music Informatics are presented in [24] and [4] respectively. However, the role of emotions felt by artists in the construction of expressive performances has been studied less frequently. Recent work [26] showed that performers experience both *music-related* and *practice-related* emotions. For some artist or musician, performing

involves feeling music-related emotions in order to "get into the mood" of the piece being played, whereas for others, performing is more a matter of deliberate conscious awareness and planned expressiveness, i.e. "knowing" music-related emotions, rather than "feeling" them [26].

Our present work deals with another aspect of music and emotion which is also scarcely studied: the communication of emotions between listeners (the audience) and performers. We describe a novel framework called Mood Conductor (MC) to create interactive music performances. The framework allows members of the audience to indicate the emotional content they find most desirable in the context of improvised music. The term "conductor" is used hereby metaphorically to refer to the act of conducting, which is the art of directing a musical performance by way of visible gestures. Through the use of the Mood Conductor App[1], audience members of a music performance are able to communicate emotional intentions to the performers using their mobile devices: smart phones, tablets or laptops. Performers are made aware of the audiences' intentions by visual feedback operating in real-time. The framework is adapted to musical improvisation situations, which may be defined as a creative activity of immediate musical composition, combining performance with communication of emotions and instrumental technique, as well as spontaneous response to other musicians [11]. In our framework, the emotional intentions communicated from spectators to performers represent exogenous directions which constrain the musical improvisation in an interactive way.

The three main components of the Mood Conductor framework are a *i)* web-based application optimised for mobile devices, *ii)* a server side engine for aggregating and clustering data from the audience, and *iii)* a visualisation client providing feedback for the musicians and the audience.

The gradual development of the framework including the tuning of the systems' parameters was based on several interactive music performances over the course of almost two years. Concerts were held at Wilton's Music Hall in London, UK, during the 'New Resonances Festival' held in conjunction with CMMR 2012, the Cathedral of Strasbourg in France, Queen Mary University of London (QMUL), UK, the Barbican Arts Centre in London, UK, and an artistic event during the International Conference on Affective Computing and Intelligent Interaction (ACII 2013) in Geneva, Switzerland [9]. Different music ensembles were involved, including the vocal quartet 'VoXP', composed of four professional singers that graduated in Jazz and Classical Music from the Conservatoire National de Strasbourg, as well as a jazz/rock trio and quintet. The different systems were evaluated using online surveys as well as a controlled comparative study conduced in the lab environment at the Media and Arts Technology Laboratory of QMUL. The results of thorough formal evaluation of the first two systems are summarised here while more details can be found in [17] and [18].

The need for the Mood Conductor framework grow out of real-world problems the VoXP quartet was facing when trying to create interactive improvised per-

---

[1] http://bit.ly/moodxp2

formances during their music studies. We invite the reader to watch a YouTube video[2] showing the system in use with the VoXP singer quartet[3].

The remainder of the paper is organised as follows. Section 2 highlights important related works and details the rationale behind selecting the underlying model of emotions for the interactive and visualisation interface of the framework. An overview of the architecture of the Mood Conductor framework is detailed in Section 3. Section 4 details the specific differences between the different MC systems, the different music performances, as well as the evaluations and the iterations of the framework. Finally, in Section 5 we summarise our findings, outline our conclusions and the future perspectives of Mood Conductor.

2   Related Works and Model Selection

The most striking examples of constrained or directed improvisation occur in the theatre [14] when an improvisational troupe invites the audience to supply a theme and some constraints and then develops a rendition of the theme that also meets the constraints [12]. This principle was adapted to the human computer interaction (HCI) domain by [12] who developed an interactive story application for children using intelligent agents.

2.1   Interactive Systems in Musical Applications

In the musical domain, directed improvisation commonly occurs in jazz and other musical performances [8]. The literature on live music improvisation systems involving audience-performer interaction is however very scarce and to the best of our knowledge, Mood Conductor is one of the first systems of its kind. The system fosters a new chain of communication between audience and performers which differs from the classic composer-performer-listener model adapted to Western written music [15]. A distributed interaction framework that supports audience and player interaction during interactive media performances was proposed in [23]. [3] describes a "cheering meter" application offering a visualisation of audience reactions which has been shown to increase the sense of participation among the audience at large-scale rap music performances. The system developed by [13] draws on facial emotion recognition technology to create art installations reacting to or mimicking audience changes in emotional expressions.

2.2   Arousal-Valence as a Model for Interaction and Visualisation

The Mood Conductor visualisation components rely on the two-dimensional arousal-valence (AV) model proposed by Thayer [25] and Russell [19] to characterise core or basic emotions (such as "happy", "calm", "sad", "anger", etc.). The experiments described in [21] demonstrated that the AV model provides a reliable means for collecting continuous emotional responses to music. In our

---

[2] http://www.youtube.com/watch?v=o9Fd7nV2IWs
[3] https://www.facebook.com/VoXPerformance

framework, the AV space is used in a reciprocal way, to collect emotional intentions in real-time.

We choose the AV model as basis for the Mood Conductor client and visualisation interface, because it overcomes some of the problems commonly associated with categorical emotion models, such as the discretisation of the problem into a set of landmarks, which prevents emotions which differ from these landmarks to be considered. Additionally, it was empirically shown to be a good representation of internal human emotion states [25], that is also relevant in the music domain, where emotion classes can be defined in terms of arousal or energy (how exciting/calming musical pieces are) and valence or stress (how positive/negative musical pieces are) [4].

Albeit higher dimensional models were also developed, such as [1], the interaction based on these models would almost certainly be less intuitive and the interface would be more cumbersome. It would also present a higher cognitive load distracting the audience members from the performance. We do not however rely solely on the AV model. The client interface also shows a mood tag closest to the selected position, allowing easier interpretation of the space, without restricting audience members to select a distinct emotion category, creating an interface that fuses dimensional and categorical emotion models.

The graphical user interface of the Mood Conductor client application shares similarities with that of 'MoodSwings' [16], a collaborative game that was developed for collecting music mood labels. However, in MC, the user interface displays mood tags, rather than emotion faces [22], to help users finding the relationships between locations in the space and their associated emotions, thus bridging the gap between the dimensional and categorical approaches [4].

3   Overview of the Mood Conductor Framework

The Mood Conductor framework consists of a shared system architecture and some common organising principles that facilitate a novel form of interaction between performers and the audience. Different systems have been built and evaluated so far within this framework. This section describes the common system components and principles, while the implementation details of the respective system variants are given in Section 4.

3.1   Common System Architecture

The Mood Conductor framework consists of three technical components that form a client-server architecture. The first component is a smartphone friendly Web-based client program. This client allows for any member of the audience with an internet and Web browser enabled smartphone (or other mobile device) to indicate emotions during a music performance, thus collaboratively "conduct" the performance.

The second component is a server-side application that provides two conceptually separate application programming interfaces (API) implemented within

the same process. The smartphone client API allows for collecting data from the audience, and also exposes a JavaScript program that is run by the mobile client's web browser. The visual client API allows for retrieving aggregated and clustered user input from the server, as well as configuring some of the server-side parameters. Concerning client-server communication, Mood Conductor follows a representational state transfer (REST) style design [10]. More details about the server side implementation and the clustering process are provided in the the respective descriptions of each Mood Conductor system variant.

The final component of Mood Conductor is a visualisation client that may be run on a conventional PC or laptop with video output. This client allows for projecting or otherwise visualising cumulative user input. The video output is displayed to the performers who may use this as guidance during improvisation. The visualisation is also projected on a (typically large) screen behind the performers for visual feedback to the audience. Figure 1 shows the components of the framework. In the rest of this section, we describe how these components work and interact.

3.2 Mobile Client Application

Mood Conductor opens a real-time communication channel in the context of improvised music performances. The mobile client allows the audience to indicate a target mood or emotion they would consider desirable, interesting, appropriate or otherwise aesthetically pleasing, and they would like to be expressed by the performers. This is achieved using a Web-based application suitable for smartphones that provides a easy to use interface to indicate emotion on the two dimensional arousal-valence (AV) space [19, 25], ranging form roughly positive to negative on the x axis, and calm to energetic on the y axis. The emotion tags closest to the selected point on the AV plane, such as "joyful" or "relaxed", are also shown on screen.

The client is a Web-based application written in JavaScript and HTML5 that runs in any modern Web browser. It is automatically downloaded when a designated URL[4] is dereferenced. This design allows Mood Conductor to be used in a platform independent manner on any mobile phone that is capable of accessing the internet and running a conventional Web browser. The client displays an interactive screen showing the arousal-valence space (see Figure 3a) and allows for selecting an emotion coordinate. The client recognises the coordinates of prominent mood words using different databases of mood word to AV coordinate mappings. These databases are described in Section 4. The selected mood coordinates are communicated to server by a designated API call using an XML HTTP request.

3.3 Mood Conductor Server

Individual user inputs are clustered by a server-side application that is also accessed by a visualisation client which produces a projection of the cumulative
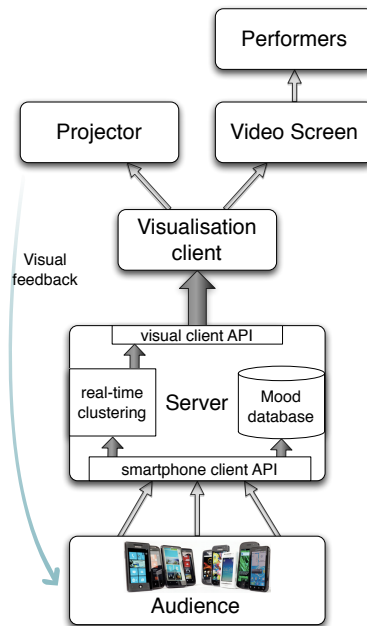
---

[4] http://bit.ly/moodxp2

Fig. 1: Overview of the Mood Conductor framework. The audience interacts with the system using a Web-based smartphone application that allows for indicating mood. Visual feedback is received via a projection of clustered responses situated behind the performers.

responses. The Mood Conductor server is a Python program written using CherryPy[5], a self-contained Web application server library. The server provides two simple APIs for communication with the audience (mobile) and visualisation clients. Most importantly, the server accepts data from the audience and registers the coordinates of the selected emotion, as well as the corresponding time. Note that for simplicity potential network delays are ignored in the current implementation. The server aggregates the user input using a real-time constrained clustering process described in Section 4.1. The purpose of the clustering is to reduce the complexity of user inputs, and aid visualisation that facilitates the interpretation of audience interaction by the performers. Essentially, the server provides an API function to be accessed by the visualisation client to retrieve data from the server. This provides information such as the cluster centres, the number of input samples or observations assigned to each cluster, and the time when each cluster was spawned.

3.4 Visualisation Client

The Mood Conductor visualisation client dynamically creates a graphical representation of the aggregated and clustered user input. The program is written

---

[5] http://www.cherrypy.org/

using PyGame[6], a Python-based real-time game engine, which in turn is based on the Simple DirectMedia Layer (SDL), a cross-platform, free and open source multimedia library written in C. Thus, the Mood Conductor visualisation client can be executed on any modern personal computer or laptop. The visualisation client accesses the server and retrieves aggregated data from the audience via its REST-based API. In later versions of the system (from MC System 2), it also serves as a client for an operator who may modify the server side clustering parameters during a performance. The need for this capability became apparent when trialling the system with a large audience.

Different visualisation modes have been implemented, all relying on projecting aggregated user input into the the arousal-valence model of emotions [19,25]. These modes range from the display of simple clustered data represented by coloured spheres, to showing a continuously evolving emotion trajectory in the AV space. The details of these visualisation models are presented in Section 4 in the context of each respective system variant.

3.5 User Data Collection

The Mood Conductor server maintains a performance log in which we collect information about the audience indicated emotions registered by the server. This process logs the IP address of each device for each received input, together with the coordinates of the indicated position on the AV plane, and the time of registration for each data point. The data collection is implemented simply by using the logging facilities of the web application server.

Table 1 summarises the data collected in four Mood Conductor performances in different venues and with significantly different audience sizes. Albeit the system logs IP addresses, individual user tracking is not yet implemented as we preferred an anonymous model for the interaction with the mobile client requiring no registration or login. However, we plan for assigning individual random identifier for connecting clients in the future, to make better use of individual emotion trajectories. As can be seen from the data, IP addresses alone cannot server this purpose, since mobile operators may assign IP addresses in a way that does not resolve individual clients, and also clients may be connected to a central WiFi network of the venue. This was the case during the last concert held in conjunction with the ACII2013 conference, hence the low number of registered IP addresses despite almost all conference participants were present during the conference and interacted with the application.

3.6 Performance Simulation

The Mood Conductor framework has the capability to simulate interaction based on data collected from previous performances as described in Section 3.5, or based simply on random data. This data allows us to replay a performance, i.e. recreate the visualisation by providing the clustering and visualisation process with input data identical to what was sent by the audience during a concert

---

[6] www.pygame.org

Table 1: Summary of data collected during the most well attended Mood Conductor performances.

concert and venue	approx. audience size	unique IPs	duration (min.)	number of emotion cues	mean response rate per sec.
Cathedral of Strasbourg (Concert #2)	150	456	15	5392	6.22
Harold Pinter Drama Studio (Concert #3)	45	68	29	5429	3.72
Hack the Barbican (Concert #5)	70	49	60	5931	2.32
ACII Artistic Event (Concert #7)	200	11	45	11723	6.08

performance. The purpose of simulation is twofold. On one hand, it enables fine tuning the visualisation parameters in the absence of a real audience. On the other hand, it serves the means of practicing the interpretation of the kind of visuals generated by the system during rehearsals. Albeit the initial concept of Mood Conductor was developed in collaboration with the VoXP quartet, several performances were held with a diverse range of artists who have not experienced the system before. The simulation capability proved very valuable in preparation for these performances. The simulator itself is implemented as a separate application that parses server logs and sends timestamped messages to the Mood Conductor server. Therefore it can also be used to test the capabilities of the network available at different performance venues.

4   Evaluations and Iterations of the Framework

The Mood Conductor framework follows a common system architecture and some basic principles with regards to the interaction and visualisation models. Several variants with different features have been built and evaluated so far. These variants can be grouped into three systems based on their user clients and visualisation modes, representing three landmarks in the development of the framework. Figure 2 illustrates the development process, the distinctive features of these systems, and their relation to concert performances and evaluations conducted in conjunction with these performances. In this section, we provide the technical details of each system, describe the respective performances where each system was used and trialled, detail on the evaluation methods and observations during these performances, and outline the results and conclusions which enabled further developments of the framework.

4.1   MC System 1

The initial version of Mood Conductor (MC System 1) was developed for the 'New Resonances Festival' held during the CMMR 2012 conference. This version featured the simplest mobile client interface and cluster-based visualisation as

MC System 1

Cluster-based visualisation

Smartphone interface with 16 mood tags and simple interpolated colors taken from wefeelfine.org

MC System 2

Cluster-based visualisation combined with moving emotion trajectory

Smartphone interface with improved colors and interaction

MC System 3

Improved calculation of moving emotion trajectory

Improved mood tag selection and simulation capability

2012		2013			
June	Sept.	April	July	August	Sept.

<<Concert #1>>
CMMR 2012
(Wilton's Music Hall London, UK)

Performers: VoXP vocal quartet

Audience size: 60

Feedback:
- qualitative obs.

<<Concert #2>>
exhibitronic#2
(Strasbourg Cathedral, France)

Performers: VoXP vocal quartet

Audience size: 150

Feedback:
- data collection
- online survey

<<Concert #3>>
New Musical Interfaces
(Harold Pinter Drama Studio, London, UK)

Performers: Jazz/rock trio

Audience Size: 60-80

Feedback:
- data collection
- online survey

<<Concert #4>>
Evaluation Experiment
(QMUL, London, UK)

Performers: Impro. quintet

Audience size: 27

Feedback:
- data collection
- comparative survey

<<Concerts #5-6>>
Hack The Barbican
(Barbican Arts Centre, London, UK)

Performers: Impro. quintet & duo

Audience size: 70

Feedback:
- data collection
- qualitative obs.

<<Concert #7>>
ACII 2013
(Geneva, Switzerland)

Performers: VoXP vocal quartet

Audience Size: 200

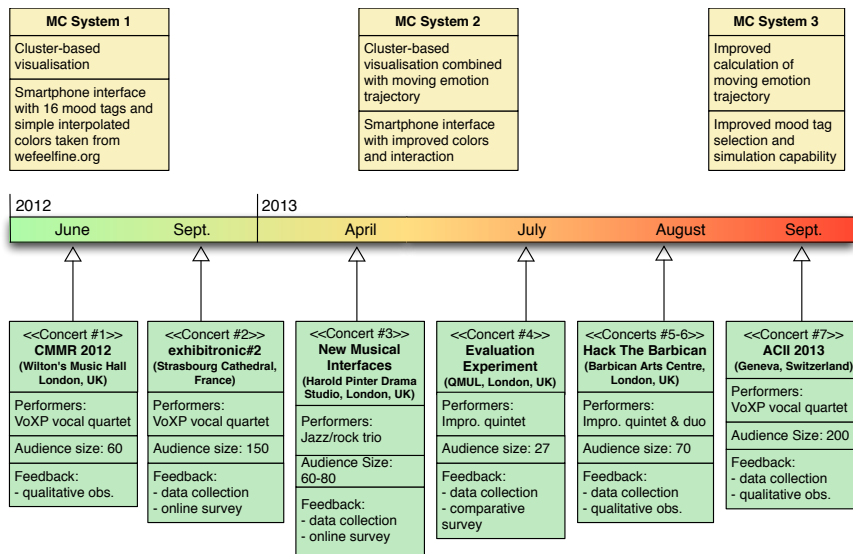Feedback:
- data collection
- qualitative obs.

Fig. 2: Development timeline of the Mood Conductor Framework.

described in the following sections. System 1 was used during the first three Mood Conductor performances. Figure 3 shows its interface and a photo taken during concert #3 with the visualisation projected in the background.
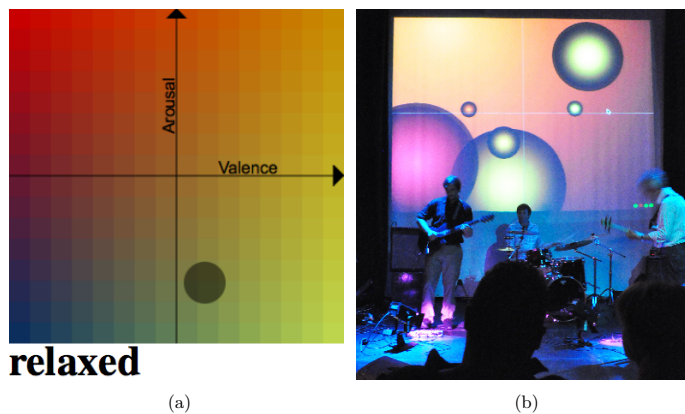

(a)   (b)

Fig. 3: (3a) Graphical user interface (GUI) of the Mood Conductor application for audience's suggestions. (3b) Visualisation of the audience's emotional directions with coloured spherical blobs. The higher the number of users indicating an emotion cue, the larger the blobs.

Mobile client interface The interface of the first system (see Figure 3a) allows the user to indicate emotion on the two dimensional arousal-valence (AV) space, which was shown to be a good representation of internal human emotion states [25], that is also relevant in the music domain. While this provides an effective way to gather the intent of each individual user in a continuous space, we do not consider the space to be readily interpretable by any member of the audience. We therefore combine the continuous representation with a discrete emotion model provided by a small number of mood tags displayed in the AV space. To find a mapping between mood words and AV coordinates, system one uses the Affective norm for English words (ANEW) database [5] developed to provide a set of normative emotional ratings for a large number of words in English. To avoid confusion resulting from the potential proximity of several words on the AV plane, especially when a small screen is considered, the system uses only some selected words from ANEW such as "serious", "relaxed", "calm", "fun" or "sad", that were deemed dominant in the authors opinion.

To prevent any single user from overwhelming the system by repeatedly selecting the same or different emotions, the client employs a blocking strategy that does not allow transmitting a different emotion within a predefined time limit. This time limit is experimentally defined, and currently set to 1000 ms. Albeit, this mitigates problems resulting form quickly repeated clicks, it still enables users to emphasise a certain area in the emotion space, potentially competing with other users.

Visualisation The visualisation created by MC System 1 is also based on the arousal-valence model. Clustered user inputs are projected on the AV plane. Each cluster is represented by a spherical coloured blob drawn over dark background, whose size is proportional to the number of input samples associated with the cluster. The colours are selected using a gradient colour map that associate areas of the AV space with conventionally accepted colours[7] (e.g. anger with red) of prominent moods. While this selection is fairly subjective, it serves only an aesthetic purpose in our application. Each blob is then drawn using a colour gradient generated between the selected emotion colour and the background. This produces a 3D effect which may create the impression that the clusters are represented by "planets" or "stars" in the "universe of emotions"[8]. The size of each sphere may be changed dynamically as new user inputs are assigned to clusters using a slow animation. To better reflect the continuously changing nature of emotion expressed by improvised music, the visualisation of each cluster is time limited. Unless new input is registered for a given cluster, its corresponding blob is faded towards the background colour at a constant rate by decreasing its opacity (alpha channel value). In the current implementation, each blob is faded away in a fixed period of time (several seconds) after the

---

[7] Colour codes for some prominent emotion were taken from the following resource: http://www.wefeelfine.org/data/files/feelings.txt

[8] See photos at: `http://bit.ly/moodcphotos`

last input is registered for that cluster. The next section provides more specific details about the clustering and visualisation method.

Time-constrained Real-time Clustering The main algorithm driving the visualisation system uses a time-constrained real-time process, conceptually related to nearest neighbour classification [7], mean-shift clustering [6], and ensemble tracking [2].

User input is organised using a maximum of N clusters that correspond to blobs B_i $(i = 1, 2, ..., N)$ visualised on screen. Each cluster is associated with the 3-tuple $(\boldsymbol{x}_i, c_i, t_i)$, where $\boldsymbol{x}_i$ is the spatial centre of the cluster on the AV plane, c_i is the number of observations or user inputs associated with cluster i, and t_i represents the time of the cluster object construction. Each input sample S received via the smartphone API is associated with the tuple $(\boldsymbol{x}_s, t_s)$, where $\boldsymbol{x}_s$ is the spatial coordinate of the user-indicated emotion on the AV plane, and t_s is the time when the input is registered. Let K_s be a spatial kernel such that

$$K_s(x) = \begin{cases} 1 & \text{if} \quad ||x|| \leq \lambda_s \\ 0 & \text{if} \quad ||x|| > \lambda_s, \end{cases} \tag{1}$$

and T_s a temporal kernel such that

$$T_s(t) = \begin{cases} 1 & \text{if} \quad t \leq \tau_s \\ 0 & \text{if} \quad t > \tau_s, \end{cases} \tag{2}$$

where λ_s and τ_s are server-side parameters representing spatial and temporal tolerances. For every user input S received via the smartphone API, a new cluster is constructed if nb_S in equation (3) evaluates to zero, that is, if we can not find an existing cluster within the spatial and temporal constraints designated by λ_s and τ_s to which the new input might be assigned to.

$$nb_S = \sum_{i=1}^{N} K_s(x_s - x_i) T_s(t_s - t_i) \tag{3}$$

In case $nb_s \geq 1$, the input is associated to the cluster denoted B' that minimises $d(x_s, x_i)$ for all B_i, where d is the Euclidean distance. In essence, this makes the process adhere to a nearest neighbour classification rule which minimises the probability of classification error [7]. The parameters of $B'(\boldsymbol{x}', c', t')$ are updated according to the following: $c' \leftarrow c + 1$ while $t' \leftarrow t_s$; however, x', the spacial centre of the cluster remains unchanged in the current implementation. In an alternative implementation to be tested in the future, we may update x' to shift towards the cluster centre defined by the new sample mean of registered audience inputs, similarly to a k-means or mean-shift clustering approach [6]. However, this may create abrupt changes in the appearance of already displayed clusters. It remains future work to assess the usefulness of this alternative technique from our application's point of view.

In the current implementation, we set N, the number of clusters displayed at any one time, to an experimentally defined fixed value (see next section).

Clusters and associated colour blobs are removed if this value is exceeded. The mechanism is implemented using a fixed length first-in, first-out (FIFO) queue. Moreover, regardless of their position in the queue, colour blobs are removed by the visualisation client when their age exceeds the predefined time threshold τ_c. This parameter is linked to the server side parameter τ_s, albeit they may be adjusted separately.

Performances with MC System 1 Three interactive music performances were organised using MC system 1 (see Fig. 2). The firs two were held in collaboration with vocal VoXP, while the third performance was held in collaboration with a jazz/rock trio (drums, bass, guitar). These performances provided means for fine tuning some of the parameters of the system during rehearsals, as well as for evaluating the system using informal feedback from the audience as well as the musicians, and by conducting online surveys.

Concert #1 was held during the CMMR 2012 "New Resonances Festival: Music, Machines and Emotions"[9] with an audience of about 60 people. The three primary parameters defined for the visualisation and clustering algorithm are N the total number of clusters or blobs allowed, λ_s the spatial proximity tolerance that governs how inputs are assigned to existing clusters, and τ_s the temporal constraint on clustering. During rehearsals these values were set to $N = 15$, $\lambda_s = 0.08$ (in normalised Euclidean space) and $\tau_s = 17s$.

Concert #2 was held as part of the Electroacoustic Music festival "exhibitronic#2". More than 150 members of the audience used the system[10]. During this concert, we used the same parameters as described above. While this worked well overall, we found that the visualisation became cluttered or chaotic during parts of the performance, especially in the first couples of minutes. This suggested that the clustering parameters need to be tuned differently for larger audiences, and also that there is a learning curve while the audience tries to familiarise with the application, and discover the meaning of locations and associated mood words on the AV plane. Musicians may also need time to learn how to adapt to more diverse responses of a large audience. Mutual adaptation however gradually allowed the performance to converge towards common directions and the audience suggested emotions to become more uniform.

The third performance was held in the Harold Pinter Drama Studio as part of the 'New Musical Interfaces' concert, with a drums, bass and guitar trio, playing improvised electroacoustic music influenced by jazz and rock. During concert #3, the audience was asked to start and stop the performance using the Mood Conductor application to emphasise the idea of "conducting" the ensemble. To this end, the musicians started the performance as the audience started to interact, and stopped a piece as the audience stopped the interaction. This idea however was only partly successful as the audience preferred to interact continuously, therefore the performance consisted of a single improvised piece. Based on our experience from the Strasbourg performance, two modifications

---

[9] http://cmmr2012.eecs.qmul.ac.uk/music-programme
[10] Photos can be found online at http://bit.ly/moodcphotos

to the system were introduced before this concert, in order to better accommodate for a larger audience. We configured the visualisation parameters slightly differently ($N = 18$, $\lambda_s = 0.15$ and $\tau_s = 12s$), which provided smoother visual feedback with a larger audience. This is probably due to the fact that more similar moods were clustered together as a result of the increased λ_s parameter, and higher number of colour blobs were allowed at any one time, resulting in fewer deletions as blobs are dropping out of a fixed length queue, i.e., more blobs were allowed to fade out gradually. Finally, to accommodate potentially swifter responses, the τ_s parameter was decreased.

Collected data analytics We collected data from the audience during the performances in the Cathedral of Strasbourg (concert #2) and the Harold Pinter Drama Studio (concert #3, see summary in Table 1).

During concert #2, the system received between 3 and 17 indicated emotion responses each second with 6.22 responses per second on average. With respect to the AV plane [19, 25], the highest number of responses occurred along the diagonal corresponding to "tiredness" vs. "energy" in Thayer's model, with a high number of responses in the negative-low ("melancholy", "dark", "atmospheric") and positive-high ("humour", "silly", "fun") quadrants. A richer cluster of responses was observed in the middle of the plane and at the positive end of the valence axis corresponding to mood words such as "happy", "pleased", or "glad".

During the third concert, between 1 and 15 indicated emotion responses were received each second with an average of 3.72. Since the performance was almost twice as long, but the size of the audience was smaller (about 40-50% of that in concert #2), these data seem to indicate a very similar level of engagement during these two concerts. A similar pattern of audience responses can be observed with regards to the AV plane with one notable difference. A more emphasised cluster of mood indications can be observed in the quadrant corresponding to negative valence and high arousal ("aggressive", "energetic", "brutal"), which seems to suggest a different genre bias in case of the rock/jazz influenced performance.

Qualitative observations Informal observation on the evolution of these performance suggests that the behaviour of the audience members includes *exploratory*, genuinely *musical*, as well as possibly *game-like* interaction. Exploratory interaction commonly happens during the first phase of a performance, and may be explained by two factors: *i)* the need for exploring the emotion space (e.g. the assignment of mood words to positions on the AV plane), and *ii)* observing the initial response of the performers to the clustered and visualised audience input. This phase is occasionally followed by a period of time where audience members focus on different quadrants of the AV plane, or different contrasting emotions. This behaviour was observed for short time durations of up to 20-30 seconds. It may be interpreted as game-like, i.e., audience members converge, but try to steer the performance towards different contending directions.

As performances evolve and the audiences' understanding of the system deepens, a slower and more balanced interplay between performers and listeners de-

velops. During this phase, the majority of audience members appear to focus on a prominent emotion area, typically indicating related emotions that are close to each other on the AV plane. The audience either moves slowly from a given emotion to other musically relevant emotion, or start to follow a new emotion selected by a minority in the audience. The observed convergence of listeners often leads to a slowly evolving trajectory spanning the AV plane, and allows the performers to express each emotion area more deeply, as well as to explore how musical improvisation techniques may be used to convey particular emotions. These informal observations allow us to form a number of hypotheses about the commonly occurring interaction types with Mood Conductor and the overall effectiveness of the system. Testing these hypotheses using the collected data (see Section 3.5) constitutes future work.

Evaluation We conducted a user survey [17] to assess the various components of MC System 1 (web application, data visualisation technique), collect feedback on the experience and level of engagement during live performances, and collect suggestions for improvement. Two sets of self-completion questionnaires were designed for audience members and performers using the surveymonkey.com platform. Participants from Concerts #2 and #3 (see Figure 2) were invited to take part in the online survey by email. In total, 35 participants comprising 29 audience members and 6 performers (3 singers, 1 guitarist, 1 bass player, 1 drummer) took part in the experiment. The answers from audience and performer participants are summarised in Figures 4a and 4b.

As can be seen in Figure 4a, about two third of the audience participants found the web application easy to use. This result highlighted the relevance and intuitiveness of the arousal/valence (AV) space as a means to convey emotional cues (note that audience members were given explanations on the AV space and how to use the app prior to the performances). Although a large proportion of participants found the mood tags and AV space colour mapping helpful, the fact that it didn't reach a consensus showed scope for improvement in the design of the user interface. The strongest flaw of the system was shown to be the cluster-based visualisation as more participants found it unclear and confusing rather than the opposite. Relatedly, most audience participants were not sure as to whether the emotional cues conveyed using the app had been followed by the performers well. The feedback from performers on the visualisation (see Figure 4b) corroborates that of audience participants, as two third of the performers found the visualisation unclear and none of them found the emotional cues easy to follow. Although MC System 1 was generally judged to be distracting when creating musical improvisation by performers, all of them expressed a strong interest in performing again with the system and its future iterations.

The audience participants highlighted several issues with the web application. A better adaptation of the user interface to the dimensions of mobile devices' screen and a more attractive design was sought for. Some participants mentioned that there were some latency after selection on the screen. It is worth noting that a delay was introduced intentionally to throttle the bandwidth, and also to
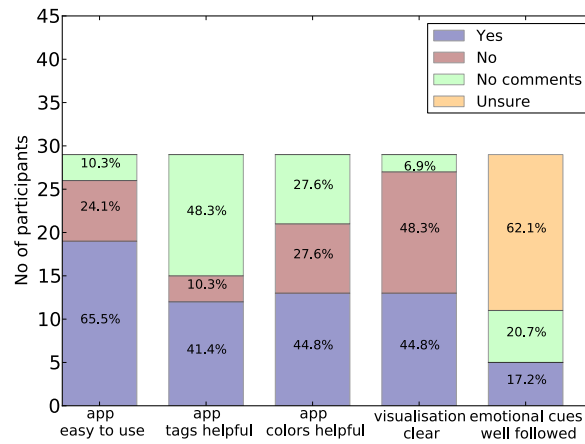
(a) Evaluation of the MC System 1 by audience participants.
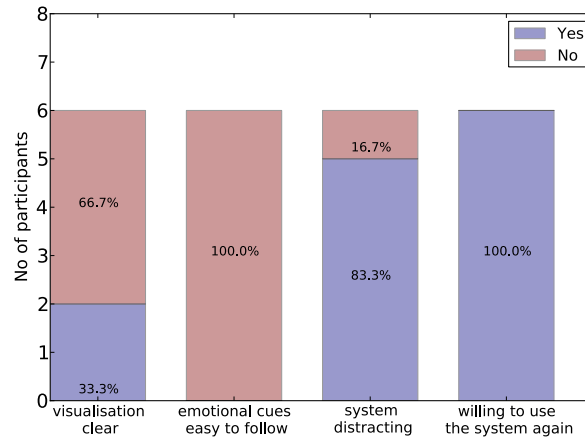
(b) Evaluation of the the MC System 1 by performers.

Fig. 4: Evaluation of the MC System 1 by audience participants (a) and performers (b).

prevent a single user from overwhelming the system. Some participants didn't know whether they had to tap a point continuously or only once on their touch screen for the data to be sent and wished to be able to see the mood tags before selecting an emotion point in the space. Refining the discretisation of the AV space to cover more mood categories was another suggested improvement. Some participants suggested to add an "introduction" or "clear instructions" in the application to guide users.

Table 2: Issues in the MC System 1 and proposed solutions.

MC System 1 web application issues	Proposed solutions in MC System 2
The interface does not consistently fit mobile devices' screen sizes.	Platform-dependent adjustment of the UI size.
More attractive user interface	Partial solutions: refinement of the color transitions in the AV space canvas. Mood tags displayed in the AV space rather than at the bottom.
No clear instructions on how to communicate with performers.	Adding of help instructions in the web app.
Data are sometimes sent unintentionally.	Adding of a button to choose when to send the data.
Occurrences of delay between selection and visualisation.	Adding of warnings when emotional cues are sent too rapidly.
MC System 1 visualisation issues	**Proposed solutions in MC System 2**
The clustered bubbles shown on screen are confusing.	New visualisation with based on the moving vote average.
Audience and performers have difficulties to interpret the data.	New visualisation with based on the moving vote average.
Changes of emotional intentions are too fast to follow for performers.	Adding constraints on moving average trajectory.

Some suggestions made by both audience participants and performers indicated that the visualisation would become clearer if it was possible to follow a single emotional cue, for instance the average of the audience's votes. Some participants indicated that, as in the app, mood tags should be displayed to help uncovering the emotions related to the clusters. The suggestions of the performers reached a consensus on several points: (i) when split votes occurred in the audience, it was very hard to figure out which emotional direction to follow, (ii) even in the case of a single emotional direction, time was required to be able to satisfactorily interpret the associated emotion, (iii) the changes of emotional directions were too fast to follow. Table 2 summarises the issues arising from the evaluation of MC System 1 and the proposed solutions implemented in MC System 2.

4.2 MC System 2

Mood Conductor System 2 was used during the evaluation experiment held at the Media and Arts Technology (MAT) Laboratory of QMUL. The system features an improved mobile device interface and a new visualisation model that uses a time-varying emotion trajectory projected in the AV space. The purpose of this system was to evaluate the changes proposed in the previous section in a controlled laboratory environment. Figure 5 illustrates the new mobile client interface and the changes to the visualisation model. The following sections describe the implementation details between system 1 and system 2.

Mobile Client Interface The client interface of this system (see Figure 5a) features a colour space with finer colour gradient on the canvas, and most importantly, a different interaction model that requires the user to press an additional button to send information to the server, rather than sending information at every touch of the screen (assuming a touchscreen device). Additionally, mood
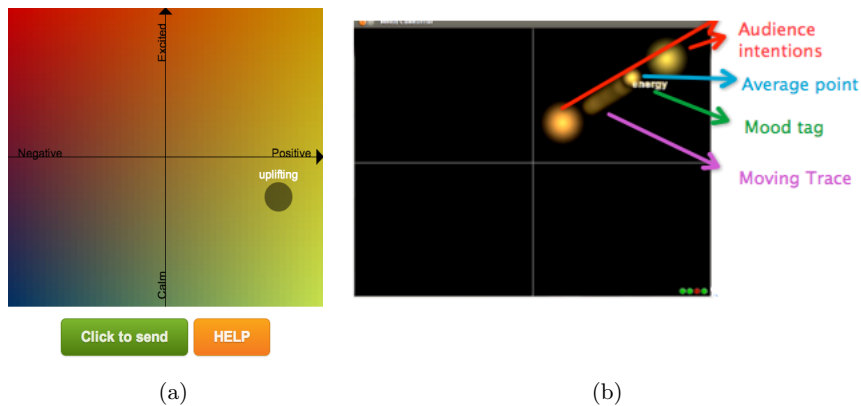
(a)   (b)

Fig. 5: (5a) Updated (GUI) of the Mood Conductor application for audience's suggestions. (5b) Visualisation of the audience's emotional directions including a moving blob and trace indicating an emotion trajectory.

words are displayed next to a selected area in the AV space, and a help feature was added providing some information about the usage of the client interface.

The rationale behind changing the interaction model arose from two requirements. First, it became apparent during the evaluation of the previous system that audience members often need more time to discover the relationships between mood words and locations on the AV plane. This was supported both by survey data and qualitative observations as well as the tendency of the collected data following a uniform distribution during the first few minutes of the performance while audience members were experimenting with the interface. Second, although System 1 throttled the communication form client to server, it was still possible to overwhelm MC by continuously clicking on a single location, encouraging occasional game-like interaction between audience members. Having to press a button to send a new emotion direction is assumed to encourage more gradual changes in intent during interaction.

Visualisation in System 2 In MC System 1, clusters of emotional intentions are represented as blobs or spheres in the AV space whose size depends on the number of people who have selected an emotional intention in the corresponding area. Using this visualisation model, musicians typically relied on the dominant emotional intention to improvise. However, as pointed out by performers and audience members, it was not always easy to determine which blobs corresponds to the emotional intention of the majority, since blobs located in different parts of the AV space happened to have similar sizes, especially in the case of large audiences [18]. To overcome this issue, and thus ease the cognitive load on musicians interpreting the visualisation, a continuously moving sphere was added to the visualisation that intends to represent the average emotional direction indicated by the audience. This visualisation follows the normalised weighted average of all active clusters represented by the blob $M(\boldsymbol{x})$, where $\boldsymbol{x}$ is the spa-

tial coordinate of the moving sphere. A gradually fading trace of $M(\boldsymbol{x})$ was also added to the visualisation. The coordinates of M are updated upon each change in the cluster configuration according to equation (4):

$$\boldsymbol{x} = \frac{\sum_{i=1}^{N} B_i(\boldsymbol{x}) B_i(c)}{\sum_{i=1}^{N} B_i(c)}, \tag{4}$$

where $B_i(\boldsymbol{x})$ is the spatial centre of cluster B_i, and $B_i(c)$ is the number of users inputs associated with that cluster.

4.3 Comparative evaluation of MC Systems 1 and 2

We conducted an experiment to compare MC Systems 1 and 2 in the context of a live performance and to determine which of the two systems was judged best by the audience and performers [18]. An improvisation ensemble gathering 5 performers (harp and singer, flutist and singer, guitarist, drummer, keyboard player) was created for the purpose of the experiment which was held in the Performance space of the Media, Arts and Technology program at QMUL. A rehearsal was organised with the performers to let them become familiar with the two systems before the experiment. Two projector screens were used for visual feedback, one located behind the performers for the audience, and the other one located behind the audience, for performers.

27 audience participants were recruited amongst students and staff at QMUL (10 males and 17 females, aged between 21 and 43 years old). Each audience participant and performer was paid for the experiment. The audience participants were divided into two groups (a group of 13 and a group of 14). To account for possible order effects in judgements about the systems, the order of use of the two systems was alternated for the two groups. Each group was first introduced to the MC framework and was given explanations about the arousal/valence space. For each system, two short performances were given during which the audience participants had to use the MC app to convey emotional cues to the performers from their mobile devices or laptops. Right after the end of the interactive performances, the audience participants had to complete an online questionnaire[11] on computers from the MAT laboratory. The performers also completed an online questionnaire[12] after their performances.
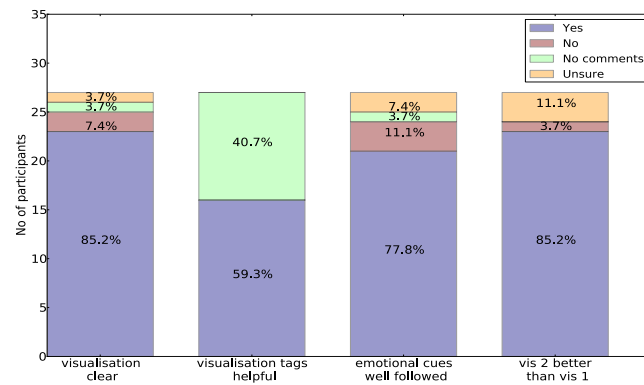
Figure 6 presents a summary of the audience participants' assessment of the web application and visualisation technique used in MC System 2. 82% of the participants preferred the app in MC System 2 which validated the benefit of the modifications made after the first evaluation, which were listed in Table 2 (e.g. send button, refinement of color transitions, display of mood tags in AV space). However, only 22% of the participants found the help button useful which showed that the instructions have to be presented in a better manner.

---

[11] Audience participants' questionnaires: session 1: `http://bit.ly/mcs1q1` and `http://bit.ly/mcs1q2` ; session 2: `http://bit.ly/mcs2q1` and `http://bit.ly/mcs2q2`

[12] Performers' questionnaire: `http://bit.ly/mcperformer`
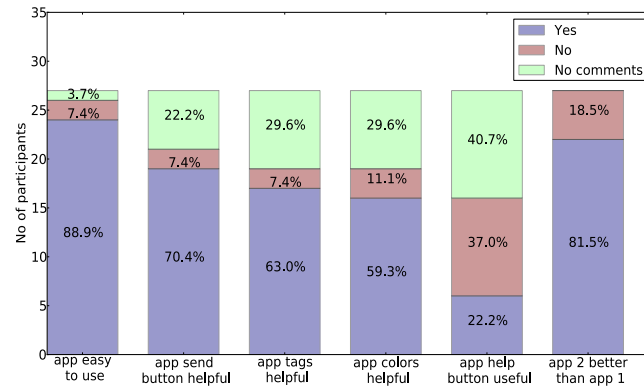
(a) Evaluation of the MC System 2 app by audience participants.

(b) Evaluation of the MC System 2 visualisation by audience participants.

Fig. 6: Evaluation of the MC System 2 app (a) and visualisation (b) by audience participants.

The improvements in the visualisation of MC System 2 (moving average) were also clearly demonstrated by the survey as 85% of the audience participants found it better than that of the initial system (cluster). While only 17% of the audience participants had found that the emotional cues were well followed by the performers in the previous evaluation of MC System 1 (see Figure 4a), most participants (79%) found that the performers well followed their emotional cues with MC System 2 (see Figure 6b). The visualisation improvements provided by MC System 2 were also acknowledged by performers who found (80%) the second visualisation system was better than the previous one and all found

that the emotional cues were easy to follow (see Figure 7), albeit by qualitative judgement, the moving trace was not responsive enough and often stuck in the middle of the space.
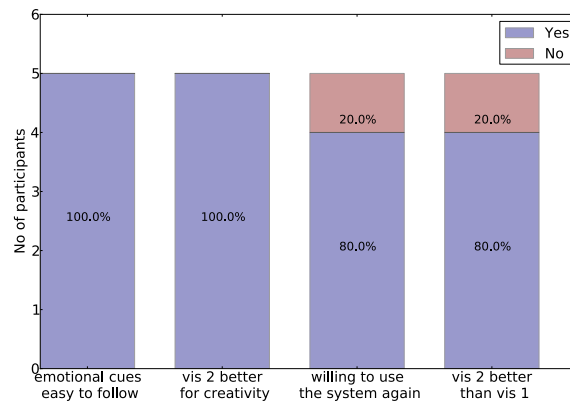

Fig. 7: Evaluation of the MC System 2 visualisation by performers.

4.4 MC System 3

The final system in our current framework builds upon the improvements introduces in MC System 2, with additional modifications in the visualisation model. Both mobile client interfaces remain usable in this system, but in the last performances (concerts #5-7) the user interface of System 2 was used. However, instead of using tags from the ANEW database, we used a music specific tag database obtained by mapping production music tags to locations in the AV space as described in [20]. The main difference lies in the way the desired emotion trajectory is computed.

Visualisation in System 3 To overcome some of the limitations in the way System 2 estimates the emotion trajectory suggested by the audience, we implemented a number of refinements in updating the location of the moving blob and associated trace. Intuitively, we would like new user input that is associated with larger clusters to be weighted higher than other user input, to avoid the emotion directions to be simply 'averaged out' over time. At the same time, to help the musicians' interpretation and artistic rendering of the trajectory, we would like relatively smooth transitions from one area of the emotion space to another, which still depends on choices made by the majority of the audience. However, previous experiments showed that audience members' emotional cues occasionally cluster in dissimilar emotion areas. Albeit we would like the system to be unbiased, artists following the emotion trajectory should be able to explore

different areas of the space. To realise these divergent requirements, System 3 adopts the following model in estimating the emotion trajectory.

Adaptive Emotion Trajectory Estimation Similarly to previous variants of Mood Conductor, in System 3 each user input is organised into N clusters associated with blobs B_i $(i = 1, 2, ..., N)$ visualised on screen. Clusters in this system are associated with the 5-tuple $(\boldsymbol{x}_i, c_i, t_i, w_i, m_i)$, where $\boldsymbol{x}_i$ is the spatial centre of the cluster, c_i is the number of observations, t_i is the timestamp of the last user input associated with cluster i, w_i is an adaptively updated weight, and m_i is the time the emotion trajectory is first observed within a spacial tolerance λ_p of the cluster.

The update rules for $\boldsymbol{x}_i$, c_i, and t_i are identical to those described in Section 4.1. The cluster weights are updated using $w' \leftarrow (w_c c + w_a a)^7 p$ at every new user input, where c, a and p are calculated according to Equations 5, 6, and 7 respectively, while w_c and w_a are experimentally defined weights using the simulation capability of the framework. The values are currently set to $w_c = 16$ and $w_a = 0.5$.

The parameter c relates to the requirement that we want the trajectory to clearly follow the majority of user inputs. Therefore the weights of clusters with a higher number of observations is higher. However, the dominance of any single cluster should be avoided, therefore we added an adaptively calculated upper limit on the number of user inputs considered. This parameter is calculated as follows:

$$c = \begin{cases} B_i(c) & \text{if} \quad B_i(c) < \lambda_c \\ \lambda_c & \text{otherwise,} \end{cases} \tag{5}$$

where λ_c represents the maximum number of user inputs considered. This value is updated once per second and depends on the current input rate, i.e., the number of audience directions sent to the server per second, reflecting the size and activity of the audience. The typical values of λ_c fluctuate between 15-30 with audience sizes of concerts #5-7 (see Table 1).

The parameter a allows clusters with newer audience responses to be weighted higher than clusters with older input. It is calculated according to Equation 6:

$$a = \begin{cases} \tau_f - (t - B_i(t)) & \text{if} \quad t - B_i(t) < \tau_f \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where t is the current time and $B_i(t)$ is the time the last input was assigned to cluster B_i. Therefore $t - B_i(t)$ represents the relative age of the cluster. The time constant τ_f is experimentally defined and is set to 10s.

The parameter p depends on the time elapsed since the trajectory was first observed within a spacial tolerance ($\lambda_p = 0.13$ in the normalised Euclidean space) of cluster B_i. This parameter relates to the requirement that we would like to facilitate smooth transitions from one area to another on the AV plane, while avoiding a single emotion area to dominate the performance. The value of p is calculated as follows:

$$p = \begin{cases} \frac{1}{10(B_i(m) - \tau_p + 1)} & \text{if} \quad B_i(m) \geq \tau_p \\ 1 & \text{otherwise,} \end{cases} \qquad (7)$$

where $B_i(m)$ is the time the emotion trajectory is first observed within a spacial tolerance of the cluster, and τ_p is a time constant currently set to 35s. The value of $B_i(m)$ is zero until $M(x)$ is observed within the tolerance of B_i, that is, $d_{Eucl}(B_i(\boldsymbol{x}), M(\boldsymbol{x})) > \lambda_p$. The effect of this rule is that the weight of cluster B_i is gradually diminished if the area has been active and the moving blob $M(\boldsymbol{x})$ has been close to the cluster for more than the time designated by τ_p. The actual visualisation of the trajectory then follows the normalised weighted average of all active clusters, similarly to Equation 4, but with the adaptive cluster weighting $B_i(w)$ used in place of $B_i(c)$ which depends only on input counts.

Performances Observations MC System 3 was used in two performances (Concerts #5 & 6) held at the end of the Mood Conductor residency during the Hack The Barbican festival (Barbican Arts Centre, London, August 2013)[13]. Concert #5 was given by an ensemble of five musicians including a singer/flutist, a singer/harpist, a keyboard player, a guitarist and a cahon/percussion player while Concert #6 was given in a duo configuration (vocals/flute and guitar). The video provided at the link below[14] shows some excerpts of Concert #5 corresponding to interactions in the four main quadrants of the arousal/valence space (funny, calm, depressive, scary). The concert was well-received with approximately 70 audience members including 49 active participants (see Table 1). Qualitative comparisons of video recordings from Concert #3 and #5 shows that the emotional cues from the audience spanned a wide range of emotions across the arousal/valence space in Concert #5 whilst cues in the "sad" and "angry" quadrants were predominant in Concert #3. We hypothesise that this finding comes from two factors, first the improved visualisation system (moving average) encourages the audience to build emotional trajectories in the AV space (in case of completely split votes, the average would be stuck in the middle resulting in neutral/boring performances, so audience members may choose their cues creating movements in the space), second, the emotional cues selected by the audience are influenced by the instrumentation due to a feedback loop with the music being played (the jazz/rock trio configuration of Concert #3 was favourable to playing sad and angry music, therefore audience members tended to be biased toward these emotions, whereas the wide diversity of instruments in Concert #5 and the presence of a harp producing particularly soft timbres invited audience members to try the various quadrants of the AV space).

The Mood Conductor project was selected as the artistic event of the 5th International Conference on Affective Computing (ACII 2013). This concert was the one with the largest number of participants so far with an audience of about 200 people. The concert was performed by the Strasbourg-based vocal quartet

---

[13] http://lanyrd.com/scmwgx and http://lanyrd.com/scmxby
[14] http://bit.ly/mcbarbican

VoXP. A video recording of the performance can be found at the link provided below[15]. As for Concert #5, a wide range of emotions spanning the AV space were performed by the vocal quartet over the course of the performance. Qualitative observations of the video recording show that the large amount of incoming data induce frequent changes of directions of the average cluster. Although this encourages musical diversity, this may hinder the performers to establish clear emotions which requires a sufficiently large time window. Finding additional mechanisms preventing too rapid shifts of the moving average in the case of "big" incoming data will be the focus of further research.

5 Conclusions and Future Work

We described Mood Conductor, a novel framework for audience-directed interactive performances that is well suited for facilitating improvisation in music, using a new form of audience feedback. The framework opens a new communication channel between the audience and musicians that proved to be valuable in seven public improvised music performances.

Several variants of the system were built and evaluated in substantially different venues with different audience sizes. We found that although some parameters of Mood Conductor are sensitive to the number of participants using its Web-based mobile application, the communication works even with suboptimal parameter settings after some adaptation by both the audience and the performers, supported by the analyses of a survey we obtained both from performers and audience members. Several solutions to the parameter optimisation problem might exist. An easily implemented extension allows the use of the new configuration API provided by the server, in order to manually tune the parameters during a performance. Further criteria might be introduced in the algorithm described in Section 4.1 to allow automatic adaptation to the rate and distribution of the incoming data. Lastly, data recorded during concerts may be replayed during rehearsals for further evaluation by the artists. This can be used to create a parameter database that allows for estimating the correct settings for each type of performance and for different audience sizes.

From a research perspective, one of the main merits and novelty of the Mood Conductor system is that it allows for examining the interaction between artists and audience using technology. The recorded data can be used in music emotion studies, and analysed in the context of recorded audio. More work is required however to evaluate further aspects of the system. It may be possible to further improve the visualisation by employing different clustering strategies and additional visualisation models to the cluster and trajectory based models tested so far. For instance, mean-shift clustering [6] may represent the user input more accurately, on the expense of potentially more complex visualisations. The AV space may be replaced by other dimensional representations of emotions, and the blobs may be replaced by several continuous trajectories.

---

[15] http://bit.ly/mc_acii2013_video

In the first parts of Concerts #5-7, the performers played several improvisations based on predetermined emotion trajectories (e.g. from calm to neutral to angry). Audience members were asked to use the MC app in a reverse manner, i.e. to rate the emotions perceived in the musical improvisations played by the performers, rather than to send emotional cues. Future research will look at finding the correlations between the emotional intentions of the performers and the listeners' data collected with the MC app.

An intriguing research question is presented by the need to define a reliable and objective measure of coherency that reflects the overall quality of communication between musicians and the audience. To this end, we may develop two alternative client interfaces that allow to split the audience into two groups: one conducting the performance as initially devised in our system, the other indicating perceived emotion. A possible way to define a coherency measure is then to select suitable correlation statistics to be calculated between the data sets collected from the two groups, appropriately corrected for a possible time lag.

References

1. E. P. Asmus. Nine affective dimensions (Test manual). Technical report, University of Miami, 1986.
2. Shai Avidan. Ensemble tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, Calufornia, USA*, 2005.
3. Louise Barkhuus and Tobias Jørgensen. Engaging the crowd: studies of audience-performer interaction. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 2925–2930, New York, NY, USA, 2008.
4. M. Barthet, G. Fazekas, and M. Sandler. Music emotion recognition: From content- to context-based models. In M. Aramaki, M. Barthet, R. Kronland-Martinet, and S. Ystad, editors, *From Sounds to Music and Emotions*, Lecture Notes in Computer Science (LNCS) 7900. Springer, 2013.
5. M.M. Bradley and P.J Lang. Affective norms for english words: Instruction manual and affective ratings. *Technical Report C-2. University of Florida*, 2010.
6. Yizong Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995.
7. Thomas M. Cover and Peter E. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.
8. R. Dean. *Creative Improvisation: Jazz, Contemporary Music and Beyond*. Open University Press,, Philadelphia, 1989.
9. G. Fazekas, M. Barthet, and M. Sandler. Mood Conductor: Emotion-Driven Interactive Music Performance. *International Conference on Affective Computing and Intelligent Interaction (ACII'13), 2-5, September, Geneva, Switzerland*, 2013.

10. Roy T. Fielding and Richard N. Taylor. Principled design of the modern Web architecture. *ACM Transactions on Internet Technology*, 2(2):115–150, May 2002.

11. R. Gorow. *Hearing and Writing Music: Professional Training for Today's Musician.* September Publishing, Gardena, CA, 2nd edition, 2002.

12. Barbara Hayes-Roth, Erik Sincoff, Lee Brownston, Ruth Huard, and Brian Lent. Directed improvisation. Technical report, Stanford University, 1994.

13. M. Iacobini, T. Gonsalves, N.B. Berthouze, and C. Frith. Creating emotional communication with interactive artwork. In *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–6, 2009.

14. K. Johnstone. *Improvisation and the Theatre.* Impro (Performance Books). Methuen Drama.

15. R. A. Kendall and E. C. Carterette. The communication of musical expression. *Music Perception*, 8(2):129–164, 1990.

16. Y. Kim, E. M. Schmidt, and L. Emelle. Moodswings: A collaborative game for music mood label collection. In *Proc. of the International Society for Music Information Retrieval (ISMIR) Conference*, 2008.

17. T. Lou, M. Barthet, G. Fazekas, and M. Sandler. Evaluation of the Mood Conductor interactive system based on audience and performers' perspectives. In *Proc. of the 10th International Symposium on Computer Music Multidisciplinary Research (CMMR)*, pages 594–609, 2013.

18. T. Lou, M. Barthet, G. Fazekas, and M. Sandler. Evaluation and Improvement of the Mood Conductor Interactive System. *53rd AES Internationa Conference on Semantic Audio, 26-29 January, London, UK*, 2014.

19. J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

20. P. Saari, M. Barthet, G. Fazekas, T. Eerola, and M. Sandler. Semantic models of musical mood: Comparison between crowd-sourced and curated editorial tags. *in proc. IEEE International Conference on Multimedia Expo (ICME2013) International Workshop on Affective Analysis in Multimedia (AAM), 15-19 July, San Jose, CA, USA.*, 2013.

21. E. Schubert. Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology*, 51(3):154–165, 1999.

22. E. Schubert, S. Ferguson, N. Farrar, D. Taylor, and G. E. McPherson. Continuous response to music using discrete emotion faces. In M. Barthet and S. Dixon, editors, *Proc. of the 9th Int. Symposium on Computer Music Modeling and Retrieval (CMMR'12)*, pages 3–19, 2012.

23. N.-M. Sgouros. Supporting audience and player interaction during interactive media performances. In *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*, volume 3, pages 1367–1370 vol.3, 2000.

24. J. A. Sloboda and P. N. Juslin. Psychological perspectives on music and emotion. In Patrik N. Juslin and John A. Sloboda, editors, *Music and Emotion Theory and Research*, Series in Affective Science, pages 71–104. Oxford University Press, 2001.

25. R. E. Thayer. *The Biopsychology of Mood and Arousal.* Oxford University Press, New York, USA, 1989.

26. A. G. W. Van Zijl and J. Sloboda. Performers' experienced emotions in the construction of expressive musical performance: An exploratory investigation. *Psychology of Music*, 39(2):196–219, 2010.