

Moodplay: an interactive mood-based musical experience

Mathieu Barthet, György Fazekas, Alo Allik, Mark Sandler
Centre for Digital Music, Queen Mary University of London
{m.barthet,g.fazekas,a.allik,mark.sandler}@qmul.ac.uk

ABSTRACT

Moodplay is a system that allows users to collectively control music and lighting effects to express desired emotions. The interaction is based on the Mood Conductor participatory performance system that uses web, data visualisation and affective computing technologies. We explore how artificial intelligence, semantic web and audio synthesis can be combined to provide new personalised and immersive musical experiences. Participants can choose degrees of energy and pleasantness to shape the music played using a web interface. Semantic Web technologies have been embedded in the system to query mood coordinates from a triple store using a SPARQL endpoint and to connect to external linked data sources for metadata.

Categories and Subject Descriptors

C.2.4 [Computer Communication Networks]: Distributed Systems—*Distributed applications*; I.2.4 [Artificial Intelligence]: Knowledge Representation Formalisms and Methods—*Semantic networks*; J.5 [Arts and Humanities]: Music

Keywords

Affective computing, artificial intelligence, semantic web, audio synthesis

1. INTRODUCTION

A number of technologies have recently been developed in response to the exponential growth of music collections, both for private users and content providers. This development is fundamentally changing music search strategies for both personal and commercial use as the traditional search methods are rendered ineffective by the sheer volume of available data. New search algorithms are typically based on content-based audio feature extraction and focus on similarity measures between tracks in a collection, however, low-level features in general provide insufficient meaning in terms of musical attributes with which most people are familiar. The

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

AM15, October 07 - 09, 2015, Thessaloniki, Greece

© 2015 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-3896-7/15/10...\$15.00

DOI: <http://dx.doi.org/10.1145/2814895.2814922>

Moodplay system allows users to collaboratively explore music in a large collection through an intuitive touch interface in which moods are represented in a 2-dimensional mood space. The played music expresses user selected emotions. Through collaborative voting process, the users create a trajectory through the mood space that is visualised with individualised selection indicators and constantly evolving average mood path. Moodplay uses crowd-sourced tag statistics to determine the location of each track in the mood space, rather than content-based features to better approximate the perceived emotion the music expresses. It is a distributed client-server system with multiple modular components each performing a specific task while communicating with other modules over the network. The system embeds Semantic Web technologies to query mood coordinate values for each audio track from a triple store and music metadata from external linked data sources.

2. BACKGROUND

2.1 Related work

Mood driven music navigation systems and interactive social audio players have been explored in different contexts for various purposes. These have employed a number of different strategies to enable interactive browsing of large music collections. Content-based feature extraction systems like SmartDJ[1] are based on the concept of building feature similarity networks of songs. SmartDJ uses a set of low-level features and then applies dimension reduction techniques to place the songs into a 2-dimensional space. The users are able to visualise and transition between songs based on tempo, spectral similarity and tonality. A different way to select music is available in affective biofeedback players that use previous listening data to reinforce user experiences. For example, AffectiveDJ[3] and AMP[9] capture listeners' biosignals and derive mood information from skin conductance. Both these systems monitor and store skin conductance data during the listening experience to be able to match users' previous experiences and select music based on listening history on subsequent occasions. An alternative strategy to feature extraction is crowd-sourcing mood tags from social networks and creating track networks based on tag statistics. Mood Cloud 2.0 [10] employs this kind of approach by extracting information from Last.fm tags and aligns these to a 2-dimensional mood space. In this respect, it is most similar to the mood selection method used in Moodplay. However, it does not implement a collaborative interface for multiple users to interact simultaneously.

The collaborative control feature of Moodplay is similar to Jukola [14] which is an example of an experimental public jukebox where participants are able to democratically choose music in a collaborative way using hand-held mobile devices together with a more traditional central jukebox control.

2.2 Mood Conductor

Mood Conductor [6] is an interactive audience-performer system that facilitates audience feedback during improvised live music performances [5]. In order to “conduct” such performances, Mood Conductor allows audience members to send emotional directions using their mobile devices. Audience indicated emotion coordinates in the arousal-valence (AV) space, a model proposed by Thayer [20] and Russell [15] to characterise core or basic emotions, are aggregated and clustered to create a video projection. This is used by the musicians as guidance as well as visual feedback to the audience. Mood Conductor addresses the problem of investigating and exploring the dynamic emotion variations due to the interaction between artists and audiences in live performances. It opens a direct communication channel from audience members to performers, providing an experimental framework to study this phenomenon. The system proved highly successful in several public performances delivering both entertainment value and useful insight into the interplay of music and emotions [7, 12].

2.3 Music Mood Annotation

Moodplay uses a database of audio recordings exceeding 10,000 tracks (see Section 3.1) automatically annotated with mood ratings in the two dimensional arousal-valence space. Manually annotating songs on this scale is not feasible, however, several automatic and semi-automatic approaches have been proposed to solve this problem. These solutions range from games-with-a-purpose [21] and crowd-sourcing, to the estimation of mood labels using machine learning models trained on small manually labelled datasets. Typically, these models learn the association between audio features and mood labels. A prime example of the former approach is MoodSwings [19], a collaborative online game that leverages crowd-sourcing to collect AV annotations directly. Audio-based approaches include auto tagging that rely on early categorical models of emotions. For instance, [11] used an audio database with manually labelled adjectives belonging to one of 13 categories and trained Support Vector Machines (SVM) on timbre, rhythmic and pitch features. However, categorical representations of emotions have been criticised for their numerous restrictions [13, 2]. For instance, the discretisation of moods into a set of “landmarks” prevents the use of emotions which differ from these landmarks. To solve this issue, music emotion labelling may be formulated as a regression problem [22] to map high-dimensional features extracted from audio directly to the AV space.

Audio-based mood annotation is still challenging however and far from being a fully solved problem [18]. This is especially true for the valence dimension, where typical performance is low ($R^2 < 0.5$) compared to human labelling. Crowd-sourcing AV annotation for a large dataset using online games or Mechanical Turk is also problematic, due to the long time it takes and the potential cost it may involve. For these reasons, in the present work we opt for a third approach which relies on folksonomy or crowd-sourced tag data collected from an online radio station Last.fm. We then map

tagged tracks to the AV space using the approach proposed in [17] and validated in [16]. This process yields the mood annotations for the dataset described in Section 3.1.

3. MOODPLAY SYSTEM

Moodplay uses an extended version of the client-server architecture originally developed for Mood Conductor, with several new components developed particularly for this work. It also uses a large audio dataset automatically annotated with arousal and valence ratings. In this section we provide an overview of the database and the Moodplay software architecture including several new components.

3.1 Audio dataset

In order to create a new musical experience, Moodplay relies on large dataset called ILM10K consisting of samples from 10,199 commercial tracks from iLikeMusic¹ (ILM). To create this database, first social tag data was collected from Last.fm for over a million tracks as in [17]. We found 218,032 tracks in the ILM database that matched one of the Last.fm tracks using string matching between the artist names and song titles and a duration difference below 0.5s. We then applied a two stage sampling method, removing duplicates and fulfilling several potentially conflicting criteria, to arrive to the 10,199 tracks used in Moodplay.

To provide a unique musical experience, the aim was to create a diverse dataset that includes many artists, albums and genres and provides a good coverage of the mood space. To this end, we sort songs into 16 genre buckets representing broad genre categories such as rock and jazz, obtained by applying hierarchical clustering to expert annotations available from ILM. We also assign each track to one from a list of unique artists available in the collection and track the number of songs sampled from each single artist. We then assign artists to multiple genre buckets and the genre categories are ordered by the number of unique artists associated with them. To provide a good coverage of moods, we apply Multi Dimensional Scaling (MDS) to the mood tags and obtain a coordinate in a 3 dimensional space for each track. A Gaussian Mixture Model (GMM) with 5 components is fitted on this data. The likelihood under the model is then used to sort tracks in the respective genre and artist buckets before sampling. During iterative sampling, we first choose a genre category using a pseudo random process that favours genres with fewer artists associated with them. Then we randomly pick an artist from this genre bucket and choose the least likely song from the selected artist under the GMM model fitted on the MDS space. This ensures entropy in the resulting sample is maximised without having to directly estimate it. Selected songs are removed from the pool and the process is repeated until a desired number of tracks is reached or no more songs can be selected given the criteria related to the maximum number of songs chosen from the same artist, album or genre. We experimented with different values for genre and artist thresholds, as well as the number of components for the GMM and evaluated different collections by looking at maximising the number of unique artists as well as the coverage of the mood space. The selected collection includes tracks from over 5600 unique artists. Ten of 16 genre categories are equally well represented in the database (about 7% each) including pop, rock and jazz, while classi-

¹www.ilikemusic.com

cal and latin are somewhat underrepresented. The database was then automatically annotated with arousal and valence mood ratings using the Affective Circumplex Transformation (ACT) proposed in [17]. This method ensures that coordinates obtained by applying MDS to crowd-sourced tags conform to locations of mood adjectives validated in human Psychological experiments.

3.2 Software architecture

Moodplay uses a client-server architecture similar to that of Mood Conductor [6] but extended with several new methods in the component client application programming interfaces (API) as well as a triple store providing a SPARQL end-point² and exposing music mood metadata. This new architecture is shown in Figure 1.

The Moodplay system is implemented using a REST-based API to collect mood directions from participants using a smart-phone friendly Web application as well as APIs for several clients which are collectively responsible for creating the Moodplay experience. These clients may communicate with each other using the API, providing a distributed environment in which Moodplay is run. The system consists of the components discussed in the next section in detail.

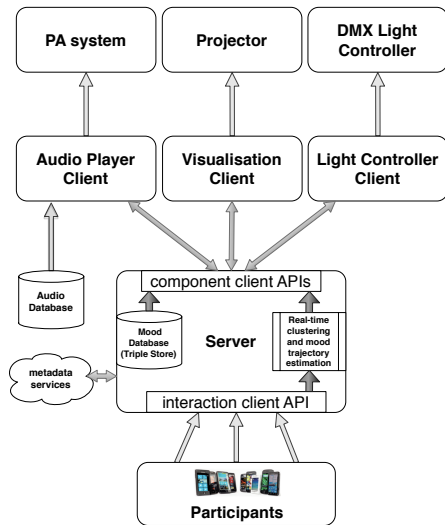


Figure 1: Overview of the technical components of the Moodplay installation. Participants interact with the system using a Web-based smartphone application that allows to select emotion cues.

3.3 Components

Specific components of Moodplay consists of a server, a participant interface running on mobile devices, a visualisation client, an audio player client, a light controller client, as well as a method for estimating mood trajectory based on data from the participants, and a semantic information management framework that builds on audio related Semantic Web technologies [8].

3.3.1 Moodplay Server

The Moodplay server component uses CherryPy³, a self-contained Web application server library. It provides APIs for communication with the participant clients as well as

the clients creating the Moodplay experience. The server accepts data from participants and registers the coordinates of the selected emotion, as well as the corresponding time. The server then aggregates the input using a real-time constrained clustering process described in Section 3.4. The purpose of this is to reduce the complexity of user inputs and facilitate the creation of a smooth path of transitions across the mood space. The Moodplay server also provides APIs for the visualisation, audio player and light controller clients 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, the time when each cluster was spawned, the estimated path across the mood space and metadata associating tracks with mood annotation. This metadata is exposed using a SPARQL end-point provided by the Fuseki server, a REST-style server in the Apache Jena⁴ framework using the SPARQL protocol over HTTP.

3.3.2 Participant interface

The participant interface is a Web-based application written in JavaScript that runs in any modern Web browser and shown in Figure 2. It is based on audience client of Mood Conductor detailed in [6] and it is automatically downloaded when a designated URL⁵ is dereferenced. This design allows Moodplay to be used in a platform independent manner. The client displays an interactive screen showing the arousal-valence space and allows the selection of a mood coordinate. The client also recognises the location of prominent mood words and displays these to facilitate the interpretation of the AV space. The selected mood coordinates are sent to the server using an XML HTTP request via the API.



Figure 2: One of the smartphone interfaces used in the Moodplay experience.

3.3.3 Audio player client

The Moodplay audio client implements a simple audio playback algorithm in the SuperCollider environment⁶. SuperCollider is a programming language for real time audio synthesis and algorithmic composition. The synthesis algorithm is based on a nearest neighbour search in a k-d tree instance that contains AV coordinates assigned to each track in the ILM10K dataset and links the coordinates to a physical audio file. User interaction data is synchronised from the visualisation client via the Mood Conductor server. The weighted average of the user data retrieved from the server determines which track to play back by matching the coordinates to the track with the nearest coordinates on the valence-arousal plane. The query rate is modifiable and determines the frequency of cross-fades between tracks. An-

⁴<http://jena.apache.org/>

⁵<http://bit.ly/moodxp2>

⁶<http://supercollider.github.io/>

²<http://www.w3.org/TR/rdf-sparql-protocol/>

³<http://www.cherrypy.org/>

other variable is density which controls the number of maximum concurrent player instances. In practice, only 2 tracks have been allowed to overlap at the same time to allow audience members interacting with the system to have a clearer perception and identification of the music, but more experimental settings could potentially be used for other purposes. For example, if the query rate is increased significantly as well as the density, the playing tracks are harder to identify, however, the mood expressed by the audio output would still approximately correspond to the average location on the mood coordinate plane. The timeout parameter defines the length of time the system keeps playing a selected track if the user data query returns no results.

3.3.4 Visualisation client

The Moodplay experience is complemented by a visualisation projected on a large screen that is part of the installation. The visualisation client creates a graphical representation of the aggregated and clustered user input as well as an estimated trajectory over the mood space given the participant data. The display of clustered data consists of coloured spheres representing the participant activity in each region of the mood space. The continuously evolving emotion trajectory is represented by a moving blob with a coloured tail that fades over time. The calculation of this trajectory is detailed in Section 3.4. The visualisation client accesses the server and retrieves aggregated data from the participants via its REST-based API. Its implementation uses PyGame⁷, a real-time game engine based on the Simple DirectMedia Layer (SDL), a cross-platform, free and open source multimedia library written in C.

3.3.5 Lighting client

In order to make the experience more immersive, stage lighting effects reacting to the mood votes of users were generated. We used the DMX512 (Digital Multiplex with 512 pieces of information) digital communication standard for the control of stage lighting and effects. DMX allows for controlling multiple slave devices through either XLR-5pin, XLR-3pin, or RJ-45 Ethernet connections. The lighting client was built upon the Open Lighting Architecture (OLA) which is an open source framework to communicate DMX512 data to lighting control systems via a network or hardware device. The lighting client sends RGB data matching the colour corresponding to the mood trajectory to a given set of LED PARs.

3.4 Emotion Trajectory Estimation

In order to create the Moodplay experience, a smooth trajectory over the mood space is desired, given the data received from the participants. To this end, we first cluster user data using a time-constrained real-time process described in [6] in order to simplify the interpretation. Then dynamically calculate a trajectory based on the current cluster centres using exponentially weighted moving average complemented by a set of heuristics.

In the Moodplay system, participant input is organised using N clusters B_i ($i = 1, 2, \dots, N$) associated with the 5-tuple $(\mathbf{x}_i, c_i, t_i, w_i, m_i)$, where \mathbf{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. Real-time clustering is governed by Equation 1,

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

where K and T are a spatial and temporal kernel such that $K_s(x) = 1$ if $\|x\| \leq \lambda_s$, while $T_s(t) = 1$ if $t \leq \tau_s$, otherwise the values are zero, and λ_s and τ_s are spatial and temporal tolerances. If $nb_S \geq 1$, the input is associated to cluster B' that minimises $d(x_s, x_i)$ for all B_i where d is the Euclidean distance, otherwise a new cluster is created.

The cluster weights w_i are updated using a set of heuristics discussed in [7] that makes them dependent on the quantity and novelty (age) of participant inputs, but also introduces a penalty when the trajectory already crossed a particular region of the AV space. These rules ensure that the trajectory clearly follows the majority of user inputs but avoid the dominance of any single cluster, and ensure a dynamic user experience with newer participant responses weighted higher than clusters with older input. We also consider the time elapsed since the trajectory was first observed within a spacial tolerance of cluster B_i . The cluster weight is diminished once the trajectory represented by $M(\mathbf{x})$ is observed within the tolerance of B_i , that is, $d_{Eucl}(B_i(\mathbf{x}), M(\mathbf{x})) < \lambda_p$ (empirically set to $\lambda_p = 0.13$ in normalised Euclidean space). This facilitates smooth transitions from one area to another on the AV plane, while avoiding a single emotion area to dominate the experience. The trajectory then follows the weighted average of all active clusters with using $B_i(w)$ instead of simply weighting the clusters by the number of participants indicating similar emotions.

3.5 Semantic data modelling and retrieval

The ACT coordinate data from different experimental configurations and tagging information from Last.fm is linked to internal ILM identifiers and physical audio file paths using an embedded light-weight OWL ontology. The structure, as illustrated in Figure 3, relies on concepts defined in the Music Ontology and the Modular Unified Tagging Ontology (MUTO)⁸. The Music Ontology⁹ uses the RDF/OWL framework to describe properties and concepts related to music, particularly metadata for published musical works. MUTO is a tag ontology which unifies core concepts of various ontologies on tagging and folksonomies in one consistent schema. The mood player namespace extends the Track class in the Music Ontology by adding an object property for coordinates, which is necessary to associate a set of alternative mood reference configurations of the ACT coordinate data to each track. AudioFile class is extended as well with a property for storing a file name for each track in the dataset. Last.fm tagging data is linked in the triple store by extending the MUTO ontology to facilitate expression of tag statistics for each track by individual tags as well as by tag types, including mood and genre. An example of a track in the dataset expressed in Turtle syntax can be seen in Listing 1. The audio client connects to a SPARQL endpoint during system initialisation and retrieves data to train the k-d tree instance. The local audio file path is used to label the valence-arousal points in the 2-d tree space. The query is shown in Listing 2.

⁸<http://muto.socialtagging.org/core/v1.html>

⁹<http://musicontology.com/>

⁷www.pygame.org

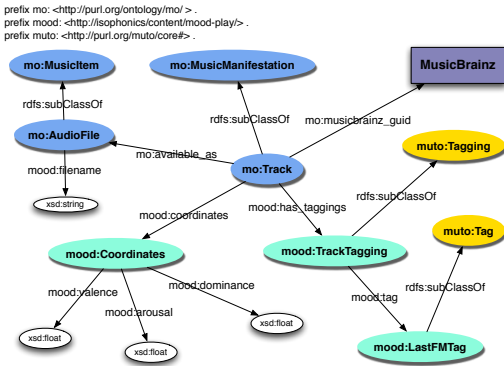


Figure 3: Moodplay uses structured data built on the Music Ontology and the MUTO tag ontology.

```

@prefix mood: <http://isophonics.net/content/mood-play/>
@prefix mo: <http://purl.org/ontology/mo/> .
@prefix muto: <http://purl.org/muto/core#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

mood:lfm13628 a mo:Track ;
mo:available_as mood:af13628 ;
mo:musicbrainz_guid
  "55743cb9-46c0-4d45-96f4-f80ed9131450" ;
mood:coordinates mood:actfold4-13628 ;
mood:ilm_id "62400:14" ;
mood:has_taggings mood:tagging_lfm13628_tag_darkgroove,
  mood:tagging_lfm13628_tag_spooky ;
mood:lastfm_id 13628 .

mood:af13628 a mo:AudioFile ;
mood:filename "62400-14.01.wav" ;

mood:actfold4-13628 a mood:Coordinates ;
mood:arousal -0.744659578346668 ;
mood:configuration mood:actfold4 ;
mood:dominance 0.148601591725518 ;
mood:valence -0.896364633037682 .

mood:actfold4 a mood:Configuration ;
mood:reference "actfold4" .

mood:tagging_lfm13628_tag_darkgroove a mood:TrackTagging ;
mood:tag mood:tag_alternative_dark_groove ;
mood:tagging_weight 18 .

mood:tagging_lfm13628_tag_spooky a mood:TrackTagging ;
mood:tag mood:tag_spooky ;
mood:tagging_weight 18 .

mood:tag_spooky a mood:LastFMTag ;
rdfs:label "spooky" ;
mood:observations 55 ;
mood:total_count 775 .

mood:tag_darkgroove a mood:LastFMTag ;
rdfs:label "dark groove" ;
mood:observations 3 ;
mood:total_count 56 .

```

Listing 1: Track and corresponding audio file representation in Turtle syntax.

The audio client sends an internal ILM dataset track identifier retrieved from the SPARQL query to the Mood Conductor server that the visual client can use to retrieve a MusicBrainz identifier from the SPARQL endpoint. MusicBrainz¹⁰ is an open and public encyclopedia for music

¹⁰<http://musicbrainz.org/>

```

prefix mood: <http://isophonics.net/content/mood-play/>

SELECT ?id ?valence ?arousal ?filename
WHERE {
  ?coords mood:configuration mood:actfold4 ;
  mood:valence ?valence ;
  mood:arousal ?arousal .
  ?id mood:coordinates ?coords ;
  mood:filename ?filename .
}

```

Listing 2: SPARQL query to retrieve ACT values for a specified configuration.

metadata that allows anyone to access or contribute data in a structured, reliable format. The MusicBrainz search server exposes an Apache Lucene service that handles metadata queries and returns results in XML and JSON formats. The visual client displays track metadata accessed from the Recording search interface at http://musicbrainz.org/ws/2/recording/specifying_the_rid_parameter_as_the_MusicBrainz_identifier_in_the_query_URI: `?query=rid:55743cb9-46c0-4d45-96f4-f80ed9131450&fmt=json`

4. EVALUATION

Two public exhibitions of Moodplay were organised as part of the Digital Shoreditch (DS) Festival¹¹ and the BBC R&D “Sound: Now and Next” (BBCSNN) conference¹² held in London in May 2015. We conducted a user survey during these events in order to assess the system in real scenarios with a large number of users from different backgrounds.

4.1 User survey

A self-completion questionnaire was designed to assess aspects related to user interaction (UI), user experience (UX), and to gain insights on possible the applications of the system and on how to improve it. We kept the questionnaire simple so that it could be quickly completed after trialling the system in the context of an exhibition. We mixed closed-ended questions (Likert scale, Yes/No), usually easy to answer, and open-ended questions, usually providing richer information, in the following sets of 7 questions:

1. *How would you rate your overall experience? (Very satisfactory/Somewhat satisfactory/Neutral/Not really satisfactory/Not at all satisfactory)*
2. *Do you think the system was easy to control using the mobile app?(Yes/No. If no, why?)*
3. *Was the mood representation in the visualisation clear and intuitive?(Yes/No. If no, why?)*
4. *Did you feel the music expressed the mood indicated on the screen?(Yes/No. If no, why?)*
5. *Did the ambient lighting complement the experience well?(Yes/No. If no, why?)*
6. *What would you use the system for?*
7. *How would you improve the system?*

Basic demographic information were also collected (gender, age range and, optionally, job title and affiliation) in the questionnaires which were anonymous.

4.2 Apparatus

¹¹<http://digitalshoreditch.com/>

¹²<http://www.bbc.co.uk/rd/events/sound2015>

The Moodplay audio and visual clients were set up on an Apple iMac computer connected to an external hard drive which stored the music database. Participants tried the system using either their own or a lent smartphone or tablet (iOS and Android) connected to the web either via 3G, 4G or WiFi. The DS2015 installation was set up in a 4 m x 7m space allowing the use of a large projection screen and stage lighting. Visual projections were made using a short throw rear projector and a portable fast-folding 4 m x 3 m projector screen. Music was played back on a stereo PA system including a subwoofer. Six LED PAR lights (Chauvet Slim Par) were hung on stage bars near ceiling level to provide immersive lighting effects. In the BBCSNN installation music was played back on headphones and the visual projections were made on the screen of the iMac as the space was limited and shared with other exhibitors. Only 2 LED PAR lights located at floor level were employed. Photos of the set up used in both events are provided at this URL¹³.

4.3 Participants

Visitors from the DS and BBCSNN events who came to the Moodplay installation were invited to interact with the system. Consenting visitors were given brief instructions on how to use the system which they could then explore and experience without timing or task-based constraints. After their trial of the system visitors were invited to take part in our survey. In total 128 participants completed the survey (113 from DS and 13 from BBCSNN, 47.6% female, 52.4% male). Their age range spreads as follows: 20-29 yo (46.8%), 30-39 yo (34.9%), 40-49 yo (14.3%), 50-59 yo (4%). The participants had a very diverse range of backgrounds which is one of the advantages of conducting surveys at wide audience events such as DS and BBCSNN: Academia (28%), Creative industry (20%), Designer (9%), Software/Web development (7%), Broadcast industry (6%), Advertising/Marketing (6%), Environment (6%), Music industry (4%), Business/Finance (4%), Healthcare (3%), Art (3%), Writing industry (2%), Social (1%), Fashion (1%), IT (1%) and Food and beverage industry (1%).

4.4 Results

4.4.1 Quantitative results

In terms of user satisfaction, 61.7% of participants found the overall experience very satisfactory, 35.2% somewhat satisfactory while 3.1% neutral. No participant found it dissatisfactory. Statistics on categorical answers to questions 2

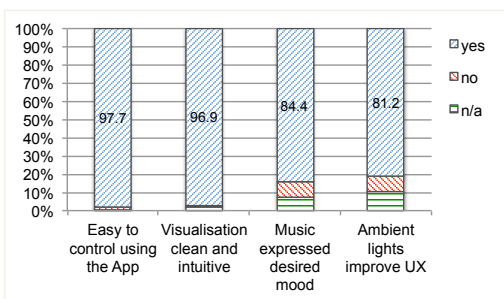


Figure 4: Participant ratings to UI and UX questions (survey questions 2 to 5).

to 5 (UX and UI) can be found in Figure 4. The system received very positive feedback from the wide majority of the participants. Indeed, 61.7% of the participants found the overall experience very satisfactory. Over 95% of the participants found the app easy to control and the visualisation clear and intuitive, and over 80% found that the music expressed well the desired moods and that the lighting effects complemented well the experience. However the fact that 35.2% of the participants only found the experience somewhat satisfactory shows that the system could benefit from improvements. This was further investigated by analysing the answers to open-ended questions as described in the next section.

4.4.2 Qualitative results

Qualitative answers were analysed using an approach borrowed from ethnography which consists in identifying and naming specific analytic dimensions and categories (coding) and analysing by themes which reflect recurrent or underlying patterns [4]. Ten main themes corresponding to potential areas of improvement of the system emerged from the thematic analysis of open-ended questions 2 to 5 and 7.

Theme 1: Personalisation: music library, preference and recommendation.

17 participants wished to be able to use the system with their own music collection (“Can you upload your own music?”, “tag my own collection of music”), metadata (“it would be really cool if you could pick your own songs for the mood - tag & label yourself”), and outside the exhibition context (“be able to use outside exhibition”, “have the app on a headset”). Some expressed interests to link the system to existing music streaming services (“link to Spotify or other music service (Last.fm) playlists of the user and vote cast are based on their library”) or to live streams (“allow for importing live streams”). Another interesting proposed use case scenario puts the focus on a single piece of music rather than a music collection (“would be interesting to control the mood of a single composition, so you could hear the adaptations of a single theme.”). Usage with larger music collection was reported several times (“add more music bands, genres, moods”, “a bigger music library would be a plus”, “not enough varieties of mood in some more extreme moods”). 13 participants suggested to add functionalities for user preference and history monitoring (“upvoting”, “ability to save playlist i created”), user customisation (“have an individual base level song to calibrate system”, “tailoring the music to suit the user?”, “ability to configure?”, “select genres”) and recommendation (“if it told you the song and you could also say if you liked the choice so the app could learn about your preferences”, “personalisation based on history, tree-like searching / search under categories”).

Theme 2: Visualisation.

A second area which was deemed to be important to improve system concerns the visualisation. 14 participants reported that the visuals could be “enhanced” and made “stronger” to make the system more “attractive”, “appealing” and “interesting”. Some participants suggested to use “image backgrounds instead of dots on the screen” and to design a “music visualisation to better connect the lighting and the on-screen visuals”.

Theme 3: Identification: user, track.

4 participants suggested to improve the user identifica-

¹³http://bit.ly/mp_photos

applications in different settings were considered by participants such as parties (38% of the participants), home (14%), club and disco (9%), bars and pubs (7%), radio (7%), festival/concerts/performance (5%), office (4%), hotel and hospitality (3%), streaming service such as Spotify (3%), therapy and fitness (2%), retail shop such as supermarket (1%) and marketing (1%). One notable and humorous comment was: “having wild parties & starting fights with my family at christmas!”

5. CONCLUSIONS AND FUTURE WORK

Moodplay has been developed in the form of an interactive installation to serve as a prototype for new types of applications that use affective computing and data visualisation technologies for music selection. The public survey results imply that most of the time the music played approximates the represented mood quite accurately, even though moods and musical preferences are rather individual and subjective. This relative success could perhaps be attributed to the crowd-sourced mood tags that enable more accurate reflection of the general consensus on the relationship between a musical work and the mood it represents. Building the system on Semantic Web technologies enables connections to external linked data sources that enables enriched metadata queries beyond the basic information about the particular audio track. There were a number of insightful ideas, suggestions for improvements and potential new applications of Moodplay in the user survey. The themes in this regard included music player for public and private parties, mood jukebox in public places, feedback device for DJs, mood-based playlist generator, accompanying or background music for public events like fashion shows, music education and even therapy. The current data visualisation could also be enhanced by introducing more dynamic animation or exploring visual textures that correspond to the mood space. The audio playback would perhaps benefit from more sophisticated search algorithms that use content-based features to match by tempo, spectrum and tonality in addition to mood coordinates.

6. ACKNOWLEDGMENTS

This work is supported by the “Fusing Semantic and Audio Technologies for Intelligent Music Production and Consumption” (FAST-IMPACT) project (EP/L019981/1). The authors thank QMUL student James Woodbridge for his work on the mood-based stage lighting effects.

7. REFERENCES

- [1] M. S. Aw, C. S. Lim, and A. W. Khong. Smartdj: An interactive music player for music discovery by similarity comparison. In *Signal and Information Processing Association (APSIPA'13)*. IEEE, 2013.
- [2] R. Cowie, G. McKeown, and E. Douglas-Cowie. Tracing emotion: an overview. *International Journal of Synthetic Emotions, Special Issue on Benefits and Limitations of Continuous Representations of Emotions*.
- [3] F. Dabek, J. Healey, and R. Picard. A new affect-perceiving interface and its application to personalized music selection. In *Workshop on Perceptual User Interfaces*, 1998.
- [4] R. M. Emerson, R. I. Fretz, and L. L. Shaw. *Writing Ethnographic Fieldnotes*. University of Chicago Press, 1995.
- [5] G. Fazekas, M. Barthet, and M. Sandler. Mood conductor: Emotion-driven interactive music performance. In *Affective Computing and Intelligent Interaction (ACII'13)*, 2-5 September, Geneva, Switzerland, 2013.
- [6] G. Fazekas, M. Barthet, and M. Sandler. The Mood Conductor System: Audience and Performer Interaction using Mobile Technology and Emotion Cues. In *10th International Symposium on Computer Music Multidisciplinary Research (CMMR'13)*, 15-18 October, Marseille, France., 2013.
- [7] G. Fazekas, M. Barthet, and M. Sandler. *Novel Methods in Facilitating Audience and Performer Interaction using the Mood Conductor Framework*, volume 8905 of *Lecture Notes In Computer Science (LNCS)*. Springer-Verlag, Heidelberg, Germany., Sound Music and Motion edition, 2014.
- [8] G. Fazekas, Y. Raimond, K. Jakobson, and M. Sandler. An overview of Semantic Web activities in the OMRAS2 Project. *Journal of New Music Research special issue on Mus. Inf. and the OMRAS2 Project*, 39(4):295-311, 2010.
- [9] J. H. Janssen, E. L. van den Broek, and J. H. Westerink. Tune in to your emotions: a robust personalized affective music player. *User Modeling and User-Adapted Interaction*, 22(3):255-279, 2012.
- [10] C. Laurier, M. Sordo, and P. Herrera. Mood cloud 2.0: Music mood browsing based on social networks. In *10th International Society for Music Information Conference (ISMIR)*, Kobe, Japan, 2009.
- [11] T. Li and M. Ogiwara. Detecting emotion in music. *4th International Society of Music Information Retrieval Conference*, pages 239-240, 2003.
- [12] T. Lou, M. Barthet, G. Fazekas, and M. Sandler. Evaluation and Improvement of the Mood Conductor Interactive System. *53rd AES International Conference on Semantic Audio*, 26-29 January, London, UK, 2014.
- [13] M. Mortillaro, B. Meuleman, and R. Scherer. Advocating a componential appraisal model to guide emotion recognition. *Journal of Synthetic Emotions, Spec. Issue on Benefits and Limitations of Continuous Repr. of Emotions*, 2012.
- [14] K. O'Hara, M. Lipson, M. Jansen, A. Unger, H. Jeffries, and P. Macer. Jukola: democratic music choice in a public space. In *5th conference on Designing interactive systems: processes, practices, methods, and techniques*. ACM, 2004.
- [15] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161-1178, 1980.
- [16] 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 *International Conference on Multimedia & Expo*, 15-19 July 2013, San Jose, CA, USA, 2013.
- [17] P. Saari and T. Eerola. Semantic computing of moods based on tags in social media of music. *IEEE Transactions on Knowledge and Data Engineering*, 26(10):2548-2560, 2014.
- [18] P. Saari, T. Eerola, G. Fazekas, and M. Sandler. Using Semantic Layer Projection for Enhancing Music Mood Prediction With Audio Features. In *Sound and Music Computing Conference*, Stockholm, Sweden, 2013.
- [19] E. M. Schmidt and Y. E. Kim. Modeling musical emotion dynamics with conditional random fields. In *12th International Society for Music Information Retrieval Conference*, Miami, Florida, USA, 2011.
- [20] R. E. Thayer. *The Biopsychology of Mood and Arousal*. Oxford University Press, New York, USA, 1989.
- [21] L. von Ahn and L. Dabbish. Designing games with a purpose. *Communications of the ACM*, 51(57), 2008.
- [22] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H.-H. Chen. A regression approach to music emotion recognition. *IEEE Trans. on Audio, Speech, and Language Proc.*, 16(2):448-457, 2008.