OPEN ACCESS Freely available online

From Interactive to Adaptive Mood-Based Music Listening Experiences in Social or Personal Contexts

MATHIEU BARTHET, *AES Associate Member*, **GYÖRGY FAZEKAS,** *AES Associate Member*,

(m.barthet@qmul.ac.uk)   (g.fazekas@qmul.ac.uk)

ALO ALLIK, FLORIAN THALMANN, AND MARK B. SANDLER, *AES Fellow*

(a.allik@qmul.ac.uk)   (f.thalmann@qmul.ac.uk)   (mark.sandler@qmul.ac.uk)

Centre for Digital Music, School of Electronic Engineering and Computer Science,
Queen Mary University of London, London E1 4NS, UK

Consumer cultures are increasingly shifting to cultures of participation supported by technology such as social computing. In the domain of interactive audio, listeners' roles are revisited giving more importance to desire, context, and the sense of control. We present new developments in our mood-driven music player Moodplay allowing both social (virtual jukebox) and personal usage. Moodplay relies on semantic computing technologies for musical mood using social tags and informative and aesthetic browsing visualizations. The prototype runs with a dataset of over 10,000 songs covering various genres and arousal and valence levels. Changes in the design of the system were made in response to a user evaluation including over 120 participants from 15 different sectors of work or education. The proposed client/server architecture integrates modular components powered by semantic web technologies and audio content feature extraction. This enables a shift from interactivity to adaptivity where recorded music content can be controlled in flexible and non linear ways.

1 INTRODUCTION

This work is concerned with the development of novel listening experiences of recorded music and is part of our larger project Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption[1]. The music recording industry created the concept of fixation—to place a song into a tangible medium such as the Compact Disc [34]. Although this revolutionized accessibility to music, it positioned recorded music as a finished goods meant to be consumed passively from the creative perspective, which consequently settled a clear distinction between producers and consumers [7]. Rooted in such tradition, the first generation of digital music broadcast services left little room for interactions from end-users, i.e., listeners. We are nowadays experiencing a shift from consumer cultures to cultures of participation in different domains, from end-user software development [34] to music entertainment, with the introduction of multimedia systems responding to the desires

and needs of the receivers and new interaction paradigms revisiting the roles and activities of receivers [13, 17, 39]. With regard to music listening such shift manifests itself by the development of adaptive music systems [8] defined here as systems enabling to select or alter music content according to user-related factors. Music recommender systems [33], which can be seen as a specific case of adaptive music systems, aim to present listeners with tracks matching their tastes and activities. Several methodologies for music recommendation have been proposed over time that may be categorized as follows according to the sources of information they rely on: (i) collaborative filtering (relies on inter-user similitude of interests, see, e.g., Last.fm[2]); (ii) content-based (relies on inter-tracks similitudes, e.g., derived from audio [12]); (iii) user-based (relies on user personal factors, such as felt mood [15] or perceived mood [4]); (iv) context-aware (relies on the user situation, such as activity [11], location [18], motion [24], time [3]); and (v) a multimodal combination of information (see, e.g., [15, 26, 38]).

In this paper we present advances made to our mood-driven and semantic web-based music player suited to both social, Moodplay [4], and personal, myMoodplay

---

This article is part of the Special Issue on Intelligent Audio Processing, Semantics, and Interaction. See the guest editors' note on page 464 of the 2016 July/August issue.

[1] http://www.semanticaudio.ac.uk

[2] http://www.last.fm

[2], usages. Among factors driving music recommendation, mood[3] has received a growing interest in the past decade [5]. Music psychology studies highlighted that music can actively be used for mood regulation [1] which provides a foundation to explain why music consumers express strong interest in being able to search and browse music by mood [20]. The communication of emotional meaning through music (perceived mood) differs from mood induction (felt mood), a difference that is often disregarded in music information retrieval research. A recommender system performing a one-to-one mapping between a user's felt mood and the mood conveyed by a song (e.g., play happy tracks when one feels happy) would discard the mood regulation function of music (e.g., play party tracks to shift from calm to excited states). To address this issue, our system confers users the possibility to choose music based on a *desired mood*, which can or not correspond to their felt mood, depending on cases. We follow an iterative user-centered design approach [22] where the design is driven and refined by evaluations with users. The survey conducted in [4] highlighted areas of improvements and new applications to our system. We focus here on improving aspects such as personalization, identification, and mixing and present a new software architecture informed by user needs and expectations.

2 RELATED WORK

The design of the music player Moodplay (MP) was inspired by our participatory live music performance system Mood Conductor (MC) [13] that allows audiences to conduct performers using mood directions. In both MC and MP, the user-facing mood voting interface (see Fig. 1) combines categorical and dimensional models of emotions by using a continuous two-dimensional space inspired by Russell's circumplex model of affect [27] (proved relevant for music by Thayer [36]) together with discrete emotion categories represented by mood tags (such as angry, haunting, sweet, and fun). The two dimensions are arousal (A) or energy (how exciting/calming musical pieces are) and valence (V) or stress (how positive/negative musical pieces are). A color gradient is also used in the voting space according to color-emotion associations obtained from the http://wefeelfine.org platform. We aimed to provide users with a self-explicit interface helping them to find desired moods easily which proved successful [21]. The first Moodplay prototype [4] can be seen as a virtual jukebox and targets social music listening experiences where people collaboratively vote for the music to be played (see Fig. 2). Visualizations and lighting effects provide informative and aesthetic feedback about individual mood votes and the mood selected by the system according to time and vote frequency considerations [13]. Time-varying individual mood

---

[3] We will employ the words emotion and mood interchangeably as their distinctions are out of the scope of this article. We refer the reader to [23] for a discussion on the differences between emotion ("temporary and evanescent") and mood ("relatively permanent and stable").
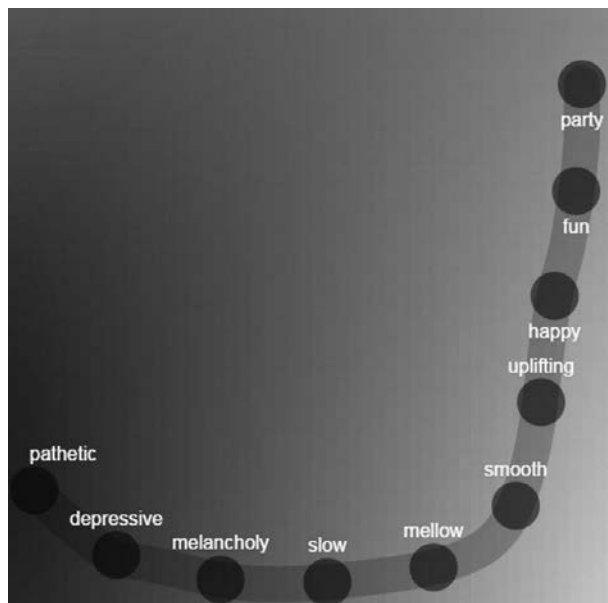

Fig. 1. Example of user-defined mood trajectory in the myMoodplay player. Vertical and horizontal dimensions are arousal and valence, respectively.
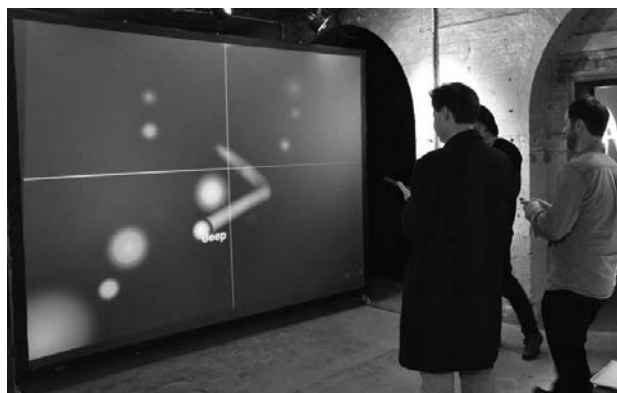

Fig. 2. Example of use of the Moodplay prototype in a social context (Digital Shoreditch 2015).

vote representations are likewise a dynamic bubble chart where areas reflect vote frequency, while a moving trace similar to Dixon's performance worm [10] is used to express current and past "elected" moods in the AV space. MC and MP are the first systems of their kind to the best of our knowledge.

The social jukebox presented in [25] presents similarities to Moodplay in terms of control (voting through a smartphone-friendly app) but music is selected in a more traditional way based on bibliographic (artist, tiltle, etc.) rather than creative (mood) metadata. Following user evaluation (Sec. 4), we developed another prototype, myMoodplay (mMP), destined to be used on a personal level in a web browser environment [2]. As described in Sec. 3, MP and mMP rely on data-driven semantic mood space inferred from crowd sourced tag-to-song associations and onto which songs from a music collection are projected [28]. A similar approach is followed in Mood Cloud 2.0 [19] but the technique to obtain the semantic mood space is

prefix mo: <http://purl.org/ontology/mo/ > .
prefix mood: <http://isophonics/content/mood-play/> .
prefix muto: <http://purl.org/muto/core#> .
prefix afo: <http://http://sovarr.c4dm.eecs.qmul.ac.uk/af/ontology/1.0#> .
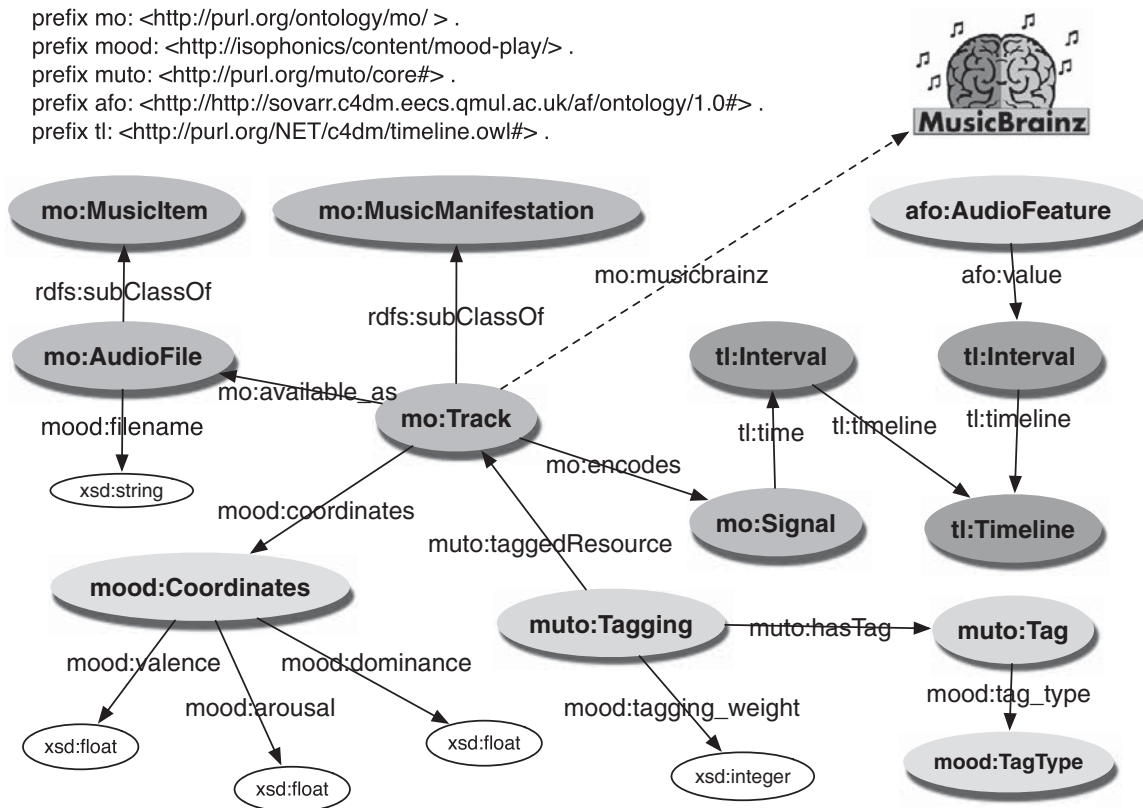prefix tl: <http://purl.org/NET/c4dm/timeline.owl#> .

Fig. 3. Moodplay uses structured data built on the Music Ontology, Audio Feature Ontology, and the MUTO tag ontology.

different; Moodplay uses dimensionality reduction techniques together with the affective circumplex transform (ACT) [28] to conform the semantic mood space to the AV space, while Mood Cloud 2.0 uses self-organizing maps (SOMs). [37] proposed an audio content-based approach combined with two-dimensional visualizations of probabilistic mappings of features such as mood and genre. Although probabilistic mappings offer a promising way to counteract uncertainties of audio-based predictions, the generated browsing space required a learning curve for users as its dimensions are not explicit. [15] introduced a hybrid content- and context-based recommender system that supports reasoning about users' desired emotion using a domain-specific semantic web ontology. Users' desired emotions are inferred from context and user factors such as situation and taste, however user-based training of an emotion state transition model must be conducted for the system to operate. Some commercial proprietary services[4] present similarities to our system but their methods are not disclosed. Music visualization interfaces relying on features other than mood have also been proposed. For instance, [40] applies support vector regression to map high dimensional audio feature vectors to a two-dimensional tempo/timbre space.

3 SEMANTIC COMPUTING OF MUSICAL MOOD

Creating mood-based listening experiences requires the association of audio recordings with perceived moods. Var-

ious approaches are possible, ranging from using purposefully composed "mood" music also known as production music [28] and expert annotations of back catalogues, to automatic identification of mood by machine relying on acoustic features of recordings (see [6] for a review). While recent advances in content-based audio analysis allow for recognizing the mood automatically, audio-based mood detection is challenging and far from being a solved problem [30]. We therefore developed methods that build mood models using social information.

The objective for this study was the construction of a well-balanced music dataset with good coverage of different genres and moods. To this end, we relied on the diverse music catalogue of I Like Music[5] (ILM), a library consisting of over 3 million songs at the time of analysis. In order to obtain mood related tags, we first sampled over 1m songs from Last.fm using the approach introduced in [29]. From the initial corpus, we found 218,032 tracks in the ILM database that matched one of the Last.fm tracks using string matching between artist names, song titles, and track durations with penalties for spelling mistakes and other metadata discrepancies. We applied a two stage sampling method to fulfill several potentially conflicting criteria. We aimed to create a dataset that represents all broad genre categories following expert classification available from ILM as well as crowd sourced genre tags.

We first compute normalized Term Frequency-Inverse Document Frequency (TF-IDF) scores from the Last.fm

---

[4] http://musicovery.com and http://www.moodagent.com

[5] http://www.ilikemusic.com

tags, then create a three-dimensional mood space by using non-metric Multi-Dimensional Scaling (MDS). To obtain a good coverage of this mood space, a Gaussian Mixture Model (GMM) with five components is fitted on the data. We then estimate the log-likelihood for each track under the fitted model. This likelihood is then used to sort songs in the respective genre and artist buckets before sampling. During iterative sampling, we first choose a genre category using a pseudo random process that favors genres with fewer artists associated with them. We then 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. The song is then removed from the pool. We also limit the number of tracks picked from any single artist or genre. This process maximizes the number of unique artists in the collection, while ensures a balanced coverage of genres and moods.

The database, denoted ILM10K in the following, is then labelled with AV coordinates using the Affective Circumplex Transform (ACT) as in [31]. This transform uses MDS coordinates but also conforms the mood space to a reference configuration of terms. Such configuration can be extracted from Russell's and Scherer's studies [27, 32]. This is done via Procrustes transformation [14], which performs a linear mapping from one space to another while retaining the relative distances between objects in the original space.

The track metadata together with arousal-valence coordinates for tracks in the ILM10k dataset produced in the ACT experiments are stored in a triplestore and structured according to a light-weight OWL ontology. The ontology uses concepts from the Music Ontology[6] for music production metadata and links these with tagging data structures described in the Modular Unified Tagging Ontology (MUTO)[7]. Each audio track entity in the triplestore is represented as a Track class from the Music Ontology and linked to a set of Last.FM tags by adding tagging properties to this class.

4 USER-CENTERED DESIGN

Our Mood Conductor and Moodplay systems have evolved and continue to evolve following a circular cycle involving prototyping, user evaluation, and inferred user and technical requirements. The Mood Conductor system is the result of a participatory design process involving performers with the objective of enabling creative interactions with audiences. Moodplay inherited from lessons learned with Mood Conductor [21], however its context of application is different as it deals with recorded music content rather than live performances. In [4] we conducted a user evaluation of the social Moodplay prototype with over 120 participants from 15 different sectors of work or education across two public exhibitions held in London in May 2015. Photos and a short video showing interactions are provided at: http://bit.ly/mp_photos and http://bit.ly/mp_video. As in

[25] we found that the process of voting for music collectively created a rich source of interaction between participants as exemplified by the creation of different groups of users sharing the experience together, the generation of discussions within these groups (e.g., on musical tastes and music technology), playful competition, and fun. About two-thirds of the participants found their experience with the system very satisfying. The most striking result was the wide range of applications suggested by the participants (see Fig. 4) including over 100 different contexts of use (from, e.g., party and home usage to marketing and therapy), some of which we had not envisioned. Among the different areas of improvement identified in [4] we focus in this paper on the following ones: (i) personalization (e.g., personal usage with own local or cloud-based music collection, customization through user preference and history, etc.), (ii) identification (users, tracks), and (iii) mixing (e.g., smoother and slower transitions, feature-based mixing). The next section describes the technical changes we implemented to provide improvements in these three areas. To be able to support different range of social or personal applications we opted for a decentralized model enabling a service-oriented architecture providing increased flexibility and accessibility. We also built on semantic web technologies following a linked data approach for content identification. In order to improve mixing aspects and produce seamless transitions between tracks as described in [9], we integrated automated sequencing technologies exploiting audio content-based features.

5 ARCHITECTURE

The components from the Moodplay system architecture can be divided into a three-tiered structure by functionality: user-facing components, web application logic, and data storage. The original [4] and proposed architectures share the underlying data storage components including the ILM10K audio dataset and a separate triplestore with AV mood coordinates, tag statistics, and a set of identifiers for each track. The original software architecture includes user, audio, visualization, and lighting clients, and the server application used to communicate between the individual client components and process user interaction data (Fig. 5). As an installation, the original system requires a location-specific setup, including a PA system, projector, and lighting effects controller for creating the user experience. The new architecture (Fig. 6) adds a JSON database storing content-based audio features to the data storage layer to improve mixing between tracks (see Sec. 5.2). The application logic is implemented in several independent modular units that can be located on one or multiple servers. The Web service API receives requests from client-facing components and information processing agents and connects to the data stores. The enhanced functionality to audio playback has introduced two more application logic software components, one to process content-based audio feature data and the other to efficiently stream audio fragments on-demand according to user interaction.

Fig. 4.   Top 100 suggested applications for Moodplay [4].

5.1 Personalization and Identification

The user interface (UI) design and implementation depends on the particular context for which the architecture is deployed. In the original system the user client runs a Web-based application written in JavaScript supported by any modern Web browser. This design allows Moodplay to be used in a platform-independent manner and the interface serves solely as a control device for users to interact. This is in contrast to the decentralized architecture implemented for the web client version myMoodplay, in which different components directly involved in creating the user experi-

ence, are required to be incorporated into a common client application. In this case the user interface is presented in the same combined space as the visualization functionality. This is achieved by layering an interactive surface on top of animated and background visualizations displaying user selections and playback indicators, to better support identification (Fig. 7). In order to improve the customization of the player, we added the possibility to build playlists and store them by directly drawing trajectories in the AV space. The Moodplay social experience is organized around a visualization projected on a large screen (Fig. 2). In a

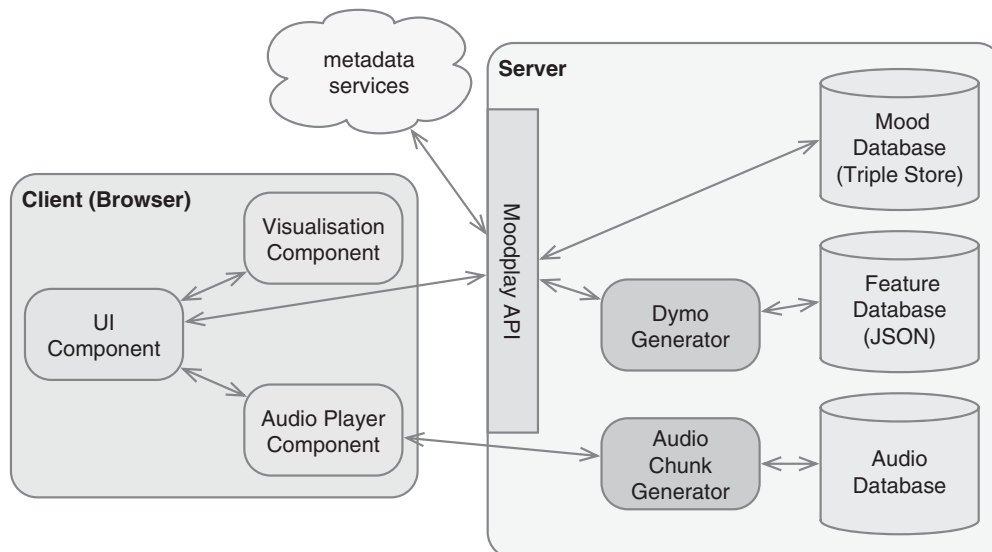Fig. 5.   The architecture of the original Moodplay system.

Fig. 6.   New architecture for the personal browser-based myMoodplay system. Dymo: dynamic music objects [35].

browser environment the principle of visually representing the mood votes remains the same regardless of whether the interaction is collaborative or personalized. The myMoodplay app combines user interface with visualization, metadata, and playback indicator as shown in Fig. 7.

5.2 Adaptive Audio Playback

The audio playback engine used in the original Moodplay system, implemented in SuperCollider[8], simply performs crossfades between two successive tracks [4]. To offer advanced audio mixing capabilities we developed a new engine based on semantic web technologies and audio features. The engine uses the *Dynamic Music Objects (dymos)* linked data music format defined by the Mobile Audio Ontology [35]. The format consists of an abstract representation of the musical structure and its semantics[9] that can be annotated with any type of audio features. *Dymos* can be queried and navigated in various ways and can include playback instructions that define how various controls and features map to predefined and custom musical parameters. The domain of such mappings can cover algorithmic controls, mobile sensor data, user interface controls, contextual information, web-sourced musical metadata, and analytical information extracted from the music itself. *Dymos* can be loaded and played back in any web environment simply by embedding the dedicated JavaScript library using the Bower package management system[10]. This library ensures that any necessary sensors or control units are allocated and it includes an audio scheduler that works for a great variety of use cases.

The audio client in the original Moodplay system connects to a SPARQL endpoint during system initialization and retrieves mood coordinate data for all tracks in the

dataset. However, this solution is not feasible for the on-demand data access model implemented in the browser-based version of the system. Instead, the nearest neighbor search is performed by sending a query to the SPARQL endpoint (example shown in Listing 1) that calculates the Manhattan distance between user selections and track coordinates in the mood space. The query determines a fixed number of nearest tracks along with their features from which the server generates *dymos*. These *dymos* are sent back to the playback engine that queries them locally to find which track will best follow the current track according to configurable transition rules (e.g., similar key, harmony, meter, tempo, etc.). On the server, the audio feature data (extracted using the Vamp plugins framework with Sonic Annotator[11]) are structured according to the Audio Feature Ontology[12] and capture acoustic information that relate to perceptual attributes (e.g., rhythm, tonality, timbre, structural segmentation), often at a high temporal resolution. In *dymos* the features are summarized depending on the custom needs of the playback engine based on statistical methods. The use of semantic web technologies for both the audio feature store and the *dymos* ensures that musical concepts and their relationships are properly defined throughout the process. At all times the playback engine is aware of the musical objects it deals with and it can rely on advanced querying methods of the semantic web to find optimal transition candidates and locations.

Transitions between successive tracks can be achieved by interpolating in various musical dimensions. For the myMoodplay system we define a main `mix` dymo as a conjunction of two tracks that can be played back simultaneously (Listing 2). The *dymo* tracks are modeled as three-level hierarchical structures consisting of the main audio track, a bar-level, and a beat-level. To create the transitions, we add custom higher-level parameters and define how they map to

---
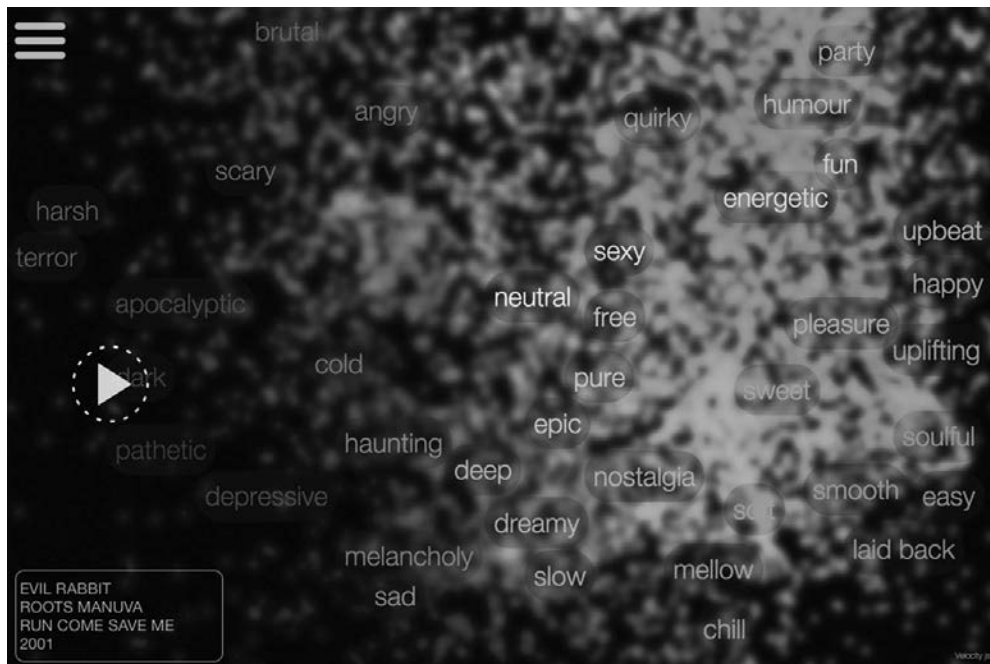
Fig. 7.   myMoodplay browser-based UI with mood space visualization.

```
PREFIX mo: <http://purl.org/ontology/mo/>
PREFIX mood: <http://isophonics.net/content/mood/>

SELECT ?valence ?arousal ?mbid ?path
WHERE
{
        SELECT ?valence ?arousal ?mbid ?path
        ((ABS(@x-?valence)+ABS(@y-?arousal)) as ?diff)
        WHERE {
                ?coords mood:valence ?valence ;
                        mood:arousal ?arousal ;
                        mood:configuration mood:actfold4 .
                ?lfmid mood:coordinates ?coords ;
                        mo:available_as ?fileid ;
                        mo:musicbrainz_guid ?mbid .
                ?fileid mood:filename ?path .
        } ORDER BY ?diff
} LIMIT 5
```

Listing 1. SPARQL query to retrieve the nearest 5 track identifiers to user specified x-y coordinates (query parameters @x and @y are substituted by floating point values when query is executed).

```
{
    "@context":"http://tiny.cc/dymo-context",
    "@id":"mix",
    "@type":"Dymo",
    "ct":"conjunction",
    "parts":["track1.json","track2.json"],
    "mappings":[{
        "domainDims":[
            {"name":"Tempo","@type":"Parameter"},
            {"name":"duration","@type":"Feature"}],
        "function":{"args":["a", "b"],
            "body":"return a/60*b;"},
        "dymos":{"args":["d"],
            "body":"return d.getLevel() == 3;"},
        "range":"TimeStretchRatio"
    },{
        "domainDims":[
            {"name":"Fade","@type":"Parameter"},
            {"name":"afv:Tempo","@type":"Feature"}],
        "function":{"args":["a", "b"],
            "body":"return b[0]+a*(b[1]-b[0]);"},
        "dymos":["mix"],
        "range":"Tempo"
    },...
    ]
}
```

Listing 2.   The main mix dymo used in myMoodplay.

lower-level ones. For instance, the first mapping in Listing 2 defines the relationship between our custom parameter `Tempo` and the standard parameter `TimeStretchRatio` of all beat-level objects, also considering their individual `duration` feature. This parameter not only adjusts the overall tempo of the songs but also evens out the duration of the beats in order to ensure that the beats can be matched during playback. The second mapping defines how the `Fade` parameter controls an interpolation between the tempos of the two tracks (via the `Tempo` parameter just defined). In a similar fashion, we then define an interpolation between the tracks' amplitudes also using low-pass filters. The interpolation mechanism can be controlled by mapping a `Ramp` control with a given duration to the `Fade` parameter. Whenever a new track is chosen by the system

we replace the inaudible track of the `mix` dymo with the new one and trigger the `Ramp`.

The new adaptive playback engine also makes use of a specialized audio server that minimizes network traffic between components and responds more flexibly to user interactions. A dedicated service sends audio chunks rather than entire files upon request, based on the segmentation structure represented in the currently playing *dymos*. This is optimal for Moodplay due to the unpredictability and potential high frequency of user choices. The audio chunk server also paves the way for future developments by enabling unpredictable behavior of the playback engine such as nonlinear playback and real-time mashups.

5.3 Discussion

The modular flexible architecture of Moodplay has enabled experimenting with different solutions to user interaction. The two-dimensional interactive control interface can be supplemented or replaced by other means of mood estimation, including extracting valence and arousal features from biosignals using, for example, an electroencephalogram (EEG) to measure brain activity or electromyogram (EMG) for detecting muscle activity. An experimental interface for Moodplay using biosignals was designed at the Sonar Music Hack Day 2015, where brain, muscle, and heart activity were mapped to valence and arousal coordinates with the Neuroelectrics Enobio[13] and BITalino[14] biosignals systems and sent to the Moodplay server. Through mapping, the participants' votes were registered and the mood trajectory calculated by the existing mechanism despite the nature of the sources and data input rates being significantly different from what the original touch surface interface affords. The new system can be used in various ways fulfilling different use cases. For instance, we can build a personalized browser-based version (myMoodplay) where a single user selects a mood or defines a mood trajectory and the music is played back locally in the client application (Fig. 6). We can also build a web-based Moodplay that works in the same way as the original system, with multiple users collaboratively selecting moods and a centralized server generating the music that the clients then stream (in this case, the audio component shown as part of the client in Fig. 6 is externalized to its own server).

6 CONCLUSIONS AND FUTURE WORK

We presented a novel client/server architecture for our mood-based music player Moodplay, which integrates modular components powered by semantic web technologies and audio content feature extraction. The proposed framework supports social (virtual jukebox) as well as personal applications and enables an adaptive control of music tracks responding to user-based (e.g., mood trajectory) and/or contextual factors (e.g., based on mobile phone sensor data). We plan to develop dynamic user personalization techniques (e.g., using like/dislike) and explore other types of browsing interface using multimodal features (e.g., tempo and mood). With regard to audio playback, we plan to define different mixing instructions where, for instance, harmony or timbre are taken into account for even smoother transitions. Dynamic music objects could be used to create mashups on the fly of two or more simultaneous songs to be able to follow votes for multiple moods. We are also considering non-linear audio techniques that could transform the player into a creative tool, for instance by reorganizing, compressing or expanding temporally pre-recorded content.

---

[13] http://www.neuroelectrics.com/products/enobio
[14] http://www.bitalino.com

8 REFERENCES

[1] J. J. H. Adrian, C. North, and David J. Hargreaves "Uses of Music in Everyday Life," *Music Perception: An Interdisciplinary Journal*, vol. 22, no. 1, pp. 41–77 (2004).

[2] A. Allik, G. Fazekas, M. Barthet, and M. Sandler "myMoodplay: An Interactive Mood-Based Music Discovery App," *Proc. of the 2nd Web Audio Conference (WAC)*, Atlanta, GA, USA (2016).

[3] L. Baltrunas and X. Amatriain "Towards Time-Dependant Recommendation Based on Implicit Feedback," *Proc. of Int. Conference on Context-Aware Recommender Systems (CARS)*, NY, USA (2009).

[4] M. Barthet, G. Fazekas, A. Allik, and M. B. Sandler "Moodplay: An Interactive Mood-Based Musical Experience," *Proceedings of the Audio Mostly 2015 on Interaction With Sound, AM '15, Thessaloniki, Greece, October 7–9, 2015*, pp. 3:1–3:8 (2015).

[5] M. Barthet, G. Fazekas, and M. Sandler "Multidisciplinary Perspectives on Music Emotion Recognition: Implications for Content and Context-Based Models," *9th International Symposium on Computer Music Modeling and Retrieval (CMMR'12)*, London, UK, June 19–22 (2012).

[6] M. Barthet, G. Fazekas, and M. Sandler *Music Emotion Recognition: From Content- to Context-Based Models*, volume 7900 of *Lecture Notes in Computer Science Volume* (Springer-Verlag, Heidelberg, Germany), From Sounds to Music and Emotions edition, 2013.

[7] Y. Bencher *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (Yale University Press, New Haven, 2006).

[8] A. R. Brown and T. Kerr "Adaptive Music Techniques," *Improvise: The Australasian Computer Music Conference*, pp. 26–31, Brisbane, Australia (2009).

[9] D. Cliff "Hang the DJ: Automatic Sequencing and Seamless Mixing of Dance-Music Tracks," Technical report, Hewlet Packard (2000).

[10] S. Dixon, W. Goebl, and G. Widmer "The Performance Worm: Real Time Visualisation of Expression Based on Langner's Tempo-Loudness Animation," *Proc. of Int. Computer Music Conference (ICMC)* (2005).

[11] S. Dornbush, J. English, T. Oates, Z. Segall, and A. Joshi "XPod: A Human Activity Aware Learning Mobile Music Player," *Proc. of the Workshop on Ambient Intelligence, 20th Int. Joint Conference on Artificial Intelligence (IJCAI)* (Jan. 2007).

[12] G. Fazekas, M. Barthet, and M. Sandler "The BBC Desktop Jukebox Music Recommendation System: A Large-Scale Trial with Professional Users," *Proc. of the*

IEEE International Conference on Multimedia and Expo (ICME), San Jose, CA, USA, July 15–19 (2013).

[13] G. Fazekas, M. Barthet, and M. Sandler "The Mood Conductor System: Audience and Performer Interaction Using Mobile Technology and Emotion Cues," *10th International Symposium on Computer Music Multidisciplinary Research (CMMR'13)*, Marseille, France, Oct. 15–18 (2013).

[14] J. C. Gower and G. B. Dijksterhuis *Procrustes Problems*, vol. 3 (Oxford University Press, 2004).

[15] B.-J. Han, S. Rho, S. Jun, and E. Hwang "Music Emotion Classification and Context-Based Music Recommendation," *Multimedia Tools and Applications*, vol. 47, no. 3, pp. 433–460 (2009).

[16] M. Harris, A. Smaill, and G. Wiggins "Representing Music Symbolically," *Proceedings of the IX Colloquio di Informatica Musicale*, Venice (1991).

[17] K. Hayes, M. Barthet, Y. Wu, L. Zhang, and N. Bryan-Kinns "A Participatory Live Music Performance with the Open Symphony System," *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pp. 313–316, New York, NY, USA (2016. ACM).

[18] A. Hazzard, S. Benford, and G. Burnett "You'll Never Walk Alone: Designing a Location-Based Soundtrack," *Proc. of the Int. Conference on New Interfaces for Musical Expression (NIME)*, pp. 411–414 (2014).

[19] C. Laurier, M. Sordo, and P. Herrera "Mood Cloud 2.0: Music Mood Browsing Based on Social Networks," *10th International Society for Music Information Conference (ISMIR), Kobe, Japan* (2009).

[20] J. A. Lee and J. S. Downie "Survey of Music Information Needs, Uses, and Seeking Behaviors: Preliminary Findings," *Proc. of the 5th International Society for Music Information Retrieval (ISMIR) Conference* (2004).

[21] T. Lou, M. Barthet, G. Fazekas, and M. Sandler "Evaluation and Improvement of the Mood Conductor Interactive System," Proc. of the *53rd AES International Conference on Semantic Audio*, conference paper P1-10 (Jan. 2014).

[22] T. Lowdermilk *User-Centered Design* (O'Reilly, 2013).

[23] L. B. Meyer *Emotion and Meaning in Music* (The University of Chicago Press, 1956).

[24] B. Moens, C. Muller, L. van Noorden, M. Frank, B. Celie, J. Boone, J. Bourgois, and M. Leman "Encouraging Spontaneous Synchronisation with D-Jogger, an Adaptive Music Player that Aligns Movement and Music," *PLoS ONE*, vol. 9, no. 12, pp. 1–40 (2014).

[25] K. O'Hara, M. Lipson, M. Jansen, A. Unger, H. Jeffries, and P. Macer "Jukola: Democratic Music Choice in a Public Space," *5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (ACM, 2004).

[26] H.-S. Park, J.-O. Yoo, and S.-B. Cho *Fuzzy Systems and Knowledge Discovery*, chapter "A Context-Aware Music Recommendation System Using Fuzzy Bayesian Networks with Utility Theory," pp. 970–979 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2006).

[27] J. A. Russell "A Circumplex Model of Affect," *J. Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178 (1980).

[28] P. Saari, M. Barthet, G. Fazekas, T. Eerola, and M. Sandler "Semantic Models of Musical Mood: Comparison between Crowd-Sourced and Curated Editorial Tags," *International Conference on Multimedia & Expo*, San Jose, CA, USA, July 15–19 (2013).

[29] P. Saari and T. Eerola "Semantic Computing of Moods Based on Tags in Social Media of Music," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2548–2560 (2014).

[30] P. Saari, T. Eerola, G. Fazekas, and M. Sandler "Using Semantic Layer Projection for Enhancing Music Mood Prediction with Audio Features," *Sound and Music Computing Conference, Stockholm, Sweden* (2013).

[31] P. Saari, G. Fazekas, T. Eerola, M. Barthet, O. Lartillot, and M. Sandler "Genre-Adaptive Semantic Computing and Audio-Based Modelling for Music Mood Annotation," *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 122–135 (2016 Apr.).

[32] K. R. Scherer "Emotion as a Multicomponent Process: A Model and Some Cross-Cultural Data," *Review of Personality and Social Psychology*, vol. 5, pp. 37–63 (CA: Sage, Beverly Hills, 1984).

[33] Y. Song, S. Dixon, and M. Pearce "A Survey of Music Recommendation Systems and Future Perspectives," *Proc. of the International Symposium on Computer Music Modeling and Retrieval (CMMR)* (2012).

[34] R. M. Stone, ed., *The Garland Handbook of African Music* (Routledge, 2nd ed., 2008).

[35] F. Thalmann, A. Perez Carillo, G. Fazekas, G. A. Wiggins, and M. Sandler "The Mobile Audio Ontology: Experiencing Dynamic Music Objects on Mobile Devices," *Tenth IEEE International Conference on Semantic Computing*, Laguna Hills, CA (2016).

[36] R. E. Thayer *The Biopsychology of Mood and Arousal* (Oxford University Press, New York, USA, 1989).

[37] B. Vad, D. Boland, J. Williamson, R. Murray-Smith, and P. B. Steffensen "Design and Evaluation of a Probabilistic Music Projection Interface," *Proc. of Int. Symposium on Music Information Retrieval (ISMIR)*, Malaga, Spain (2015).

[38] X. Wang, D. Rosenblum, and Y. Wang "Context-Aware Mobile Music Recommendation for Daily Activities," *Proc. ACM MM'* (2012).

[39] L. Zhang, Y. Wu, and M. Barthet "A Web Application for Audience Participation in Live Music Performance: The Open Symphony Use Case," *Proc. of the Int. Conf. on New Interfaces for Musical Expression (NIME)* (2016).

[40] J. Zhu and L. Lu "Perceptual Visualization of a Music Collection," *IEEE International Conference on Multimedia and Expo (ICME'05)*, pp. 1058–1061 (2005 July).

THE AUTHORS

Mathieu Barthet   György Fazekas   Alo Allik   Florian Thalmann   Mark B. Sandler

Mathieu Barthet Ph.D. is a Lecturer in Digital Media at Queen Mary University of London (QMUL). He received MSc degrees in electronics and computer science from University of Paris VI in 2003 and in acoustics from Ecole Centrale Marseille in 2004. He was awarded a Ph.D. from Aix-Marseille University and CNRS-Laboratory of Mechanics and Acoustics in 2008 ("From Performer to Listener: An Acoustical and Perceptual Analysis of Musical Timbre"). From 2009 to 2014 he was a Postdoctoral Researcher at QMUL's Centre for Digital Music on EP-SRC and AHRC projects in collaboration with the BBC, the British Library, and I Like Music. He has published over 50 academic papers in the fields of music informatics, music perception, affective computing, human computer interaction, computational musicology, big data, and semantic web. He is co-investigator on the Audio Commons H2020 EU Project (2016–2019). He was general chair of the CMMR 2012 conference "Music and Emotions." He is a member of the AES and ACM. He is a guitarist and regularly plays in rock/jazz ensembles.

●

György Fazekas Ph.D. is a Lecturer in Digital Media at the Centre for Digital Music. His research investigates how semantic audio technologies streamline music production workflows. He has published over 50 academic papers in the fields of music information retrieval, ontologies, and semantic audio analysis. He participated in research and knowledge transfer projects as researcher, developer, and at management level. He is QMUL's Principal Investigator of the € 2.9M EU/H2020 funded AudioCommons project. He was co-investigator of several research projects funded by JISC, EPSRC. He worked with BBC R&D to create mood-based music recommendation systems in the TSB funded Making Musical Mood Metadata (M4) project. He was papers co-chair of the AES 53rd International Conference on Semantic Audio, London, UK in 2014. He is a member of the IEEE, ACM, BCS, and AES and received the Citation Award of the AES for his work on the Semantic Audio Analysis Technical Committee.

●

Alo Allik Ph.D. is a Postdoctoral Researcher at the Centre for Digital Music, Queen Mary University of London. He has completed a B.A. degree in computer music at the University of Washington, M.A. degree in sonology at the Royal Conservatoire in the Hague, and Ph.D. in music composition at the University of Hull. His most recent research has focused on algorithmic music performance systems, machine learning and artificial intelligence, interactive composition, live coding, and musical knowledge representation.

●

Florian Thalmann Ph.D. is a Postdoctoral Researcher in semantic audio, computer music, and mathematical music theory at the Centre for Digital Music, Queen Mary University of London. He received an MSc in computer science from the University of Bern, Switzerland, and after working as a freelance software developer and musician for several years, he received an MA in art history and a Ph.D. in music theory from the University of Minnesota. His research interests lie in ontological and mathematical representations of music, creative processes, gestural control and interaction design, musical spaces and transformations, dynamic music, and musical applications of machine learning.

●

Prof. Mark Sandler Ph.D. FAES FIET FIEEE CEng is Founding Director of the Centre for Digital Music, a world-leading research group in audio and music technology with over 80 members. The Centre is in Queen Mary University of London's School of Electronic Engineering & Computer Science, where he holds the chair in Signal Processing. He is Principal Investigator of the EPSRC-funded grant, Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption (www.semanticaudio.ac.uk). He is a recipient of the Royal Society Wolfson Research Merit Award (2015–19) and has published over 400 papers in conferences and journals.