Exploring music with a probabilistic projection interface

B. Vad, J. Williamson, D. Boland, R. Murray-Smith MAIL@BEA-VAD.DE,
School of Computing Science, University of Glasgow, Scotland. JHW, DANIEL, ROD@DCS.GLA.AC.UK

P. Steffensen PBS@SYNTONETIC.COM
Syntonetic A/S, Copenhagen, Denmark.

Abstract

We present the design and evaluation of an interactive tool for music exploration, with musical mood and genre inferred directly from tracks. It uses probabilistic representations of multivariable predictions of subjective characteristics of the music to give users subtle, nuanced visualisations of the 2D map. These explicitly represent the uncertainty and overlap among features and support music exploration and casual playlist generation. A longitudinal trial in users' homes showed that probabilistic highlighting of subjective features led to more focused exploration in mouse activity logs, and 6 of 8 users preferred the probabilistic highlighting.

1. Introduction

Our aim is to build an interactive music exploration tool which offers interaction at a range of levels of engagement and which can foster directed exploration of music spaces, casual selection (Pohl & Murray-Smith, 2013; Boland et al., 2013; 2015) and serendipitous playback. It should provide a consistent, understandable and salient layout of music, in which users can learn music locations, select music and generate playlists, promoting discovery, rediscovery and accommodating widely varying collections.

To address these goals, we built and evaluated a system (Figure 2) to interact with 2D music maps (Stober, 2011), based on dimensionally-reduced inferred subjective aspects such as mood and genre. We used a flexible pipeline of acoustic feature extraction, nonlinear dimensionality reduction and probabilistic feature mapping. The features are generated by the commercial Moodagent Profiling Service[1] for each song, computed from low-level acoustic features, based on a machine-learning system which learns feature ratings from a small training set of human subjective classifications. These inferred features are uncertain due to this subjectivity and the general limitations of music classification, despite the apparently high performance achieved in the literature (Sturm, 2014).

2. Probabilistic Music Interface

As shown in Figure 1, the interface builds on features derived from raw acoustic features and transforms these into a mood-based interface, where nearby songs will have a similar subjective "feeling". Our feature extraction service provides over thirty predicted subjective features for each song including its mood, genre, style, vocals, instrument, beat, tempo, energy and other attributes. The features associated with moods chosen for later highlighting in the visualisation are *Happy, Erotic, Angry, Fear, Sad* & *Tender*. *Tempo*, not strictly a mood, was also included. For interaction, we used the t-distributed stochastic neighbour embedding (t-SNE, (van der Maaten & Hinton, 2008)) model for nonlinear dimensionality reduction down to 2D.

To support user exploration, probabilistic models present high dimensional features in the 2D space. This probabilistic back-projection gives insight into the structure of the layout, and the uncertainties associated with the classifications. Over the pipeline (Figure 1), we built an efficient, scalable web-based UI for collections of 20000 songs. The tracks can be seen as random variables drawn from a probabilistic distribution with respect to a specific feature. The distribution parameters can be estimated and used for prediction, allowing smoothed interpolation of features as shown in Figure 1c) and 2. We used Gaussian Process (GP) priors, a powerful nonparametric Bayesian regression method. We used a squared exponential covariance function on the 2D (x, y) coordinates, predicting the

[1] http://www.moodagent.com/

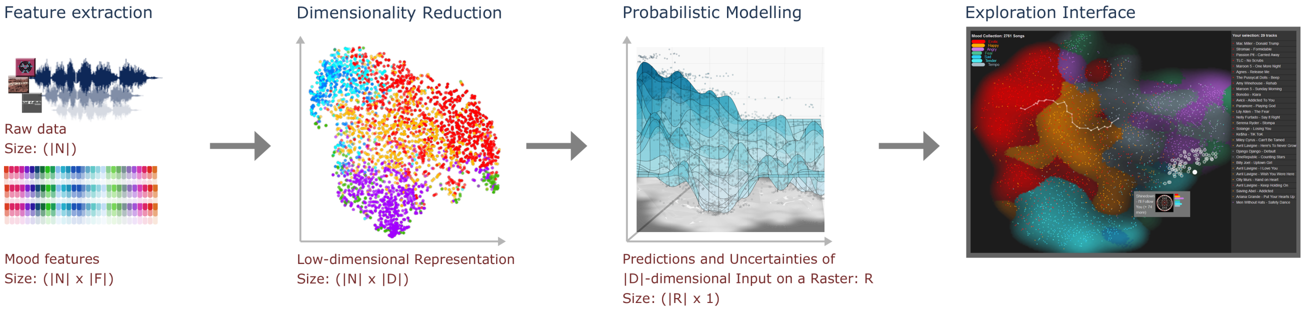

Figure 1. (a) An audio collection, described by features extracted from tracks. (b) visualisation of this high-dimensional dataset in 2D using dimensionality reduction (c) probabilistic models showing distribution of specific features in the 2D space (d) combining dimensionality reduction with these models to build an interactive exploration interface. This overview of the interactive web interface is in its 'winner takes all' colouring. A path playlist selection as well as an area selection playlist is visible in the mood space.

mood features P_f over the map. The GP infers the uncertainty σ_f^2 of the feature relevance for each point.

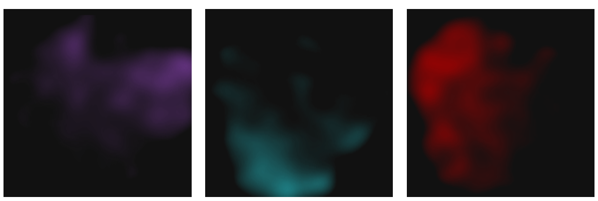

Figure 2. Background highlighting for the *angry* feature (left), *erotic* (center), and *tender* (right). Compared with the 'winner takes all view', the subtle fluctuations of features are apparent.

To present the inferred subjective results to the users, the GP mean and standard deviation is evaluated over a 200×200 grid covering the 2D music space. A continuously coloured *background highlighting* is created where areas of high feature scores stand out above areas with higher uncertainty or lower scores. To highlight areas with high prediction scores and low uncertainty, a nonlinear transform is used: $\alpha_f = P_f^2 - \sigma_f^2$, for each mood feature f, having a standard deviation σ_f and a predicted feature value P_f. The clusters in the music space can be emphasised as in the upper part of Figure 2 by colouring areas with the colour associated with the highest α_f score – a winner-takes-all view. This not only divides the space into discrete mood areas but also shows nuanced gradients of mood influences within those areas. However, once a user starts to dynamically explore a specific area, the system transitions to *implicit background Highlighting* such that the background distribution of the mood with the highest value near the cursor is blended in dynamically as in Figure 2, giving the user more subtle insights into the nature of the space.

3. Evaluation results

Can users create viable mental models of the music space?
The feedback from the 'in-the-wild' evaluation indicates that people enjoyed using these novel interfaces on their own collections, at home, and that mood-based categorisation can usefully describe personal collections, even if initially unfamiliar. *"fun to explore the galaxy", "easy generation of decent playlists"* Users also appreciated the casual nature of the interface: *"It was very easy to take a hands-off approach", "I didn't have to think about specific songs".* Log analysis revealed distinct strategies in experiencing the mood space. Some users explored diverse parts of the mood space and switched among them, while others quickly homed in on areas of interest and then concentrated on those. Their responses suggest they learned the composition of the space and used it more constructively in later sessions. Users make plausible mental models of the visualisation – they know where the favourite songs are – and can use this to discover music and formulate playlists.

Which interface features enable people to navigate and explore the music space?
Interactive background highlighting seemed to reduce the need to browse intensively with the mouse. Subjective feedback confirmed that it helped understand the music space with 6/8 users preferring it over no highlighting. Most users did not feel disturbed by the implicitly changing background highlighting. Both the neighbourhood and trajectory playlist generators were used by the participants, although neighbourhood selections were subjectively preferred and were made $3\times$ more often than trajectories.

Can a single interface enable casual, implicit and focused interaction?
Users valued the ability to vary the level of engagement with the interface. *"It was very easy to take a hands-off approach", "I didn't have to think about specific songs".* Their feedback also suggested that incorporating preview and control over the playing time of playlists would be useful, e.g. move towards "happy" over 35 minutes. A recurring theme was that playlists tended to be repetitive. One solution would be to allow the jittering of playlist trajectories and to do this jittering in high-dimensional space. The low-dimensional path then specifies a prior in the high-dimensional music space which can be perturbed to explore alternative expressions of that path.

References

Boland, Daniel, McLachlan, Ross, and Murray-Smith, Roderick. Inferring Music Selections for Casual Music Interaction. *EuroHCIR*, pp. 15–18, 2013.

Boland, Daniel, McLachlan, Ross, and Murray-Smith, Roderick. Engaging with mobile music retrieval. In *MobileHCI 2015, Copenhagen*, 2015.

Pohl, Henning and Murray-Smith, Roderick. Focused and casual interactions: allowing users to vary their level of engagement. In *Proc. ACM SIGCHI Conf. on Human Factors in Computing Systems*, CHI '13, pp. 2223–2232, New York, NY, USA, 2013. ACM.

Rasmussen, Carl Edward and Williams, Christopher K. I. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

Schedl, Markus and Flexer, Arthur. Putting the User in the Center of Music Information Retrieval. In *Proc. of the Int. Conf. on Music Information Retrieval (ISMIR)*, Porto, Portugal, 2012.

Stober, Sebastian. *Adaptive Methods for User-Centered Organization of Music Collections*. PhD thesis, Otto-von-Guericke-Universitt Magdeburg, 2011.

Sturm, Bob L. A simple method to determine if a music information retrieval system is a horse. *IEEE Transactions on Multimedia*, 16(6):1636–1644, 2014.

van der Maaten, Laurens and Hinton, Geoffrey. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

Acknowledgements

This work was supported by contributions from the University of Glasgow, Nokia, the Danish Council for Strategic Research of the Danish Agency for Science Technology and Innovation under the CoSound project, case number 11-115328, and Bang & Olufsen, This publication only reflects the authors' views.