

Engaging with Mobile Music Retrieval

Daniel Boland
daniel@dcs.gla.ac.uk

Ross McLachlan
r.mclachlan.1@research.gla.ac.uk

Roderick Murray-Smith
roderick.murray-smith@glasgow.ac.uk

School of Computing Science
University of Glasgow, United Kingdom

ABSTRACT

This paper contributes novel measures of user engagement in mobile music retrieval, linking these to work in music psychology, and illustrating resulting design guidelines in a demonstrator system. The large music collections available to users today can be overwhelming in mobile settings, they offer ‘too-much-choice’ to users, who often resort to shuffle-based playback. Work in music psychology has introduced the concept of music engagement – listeners vary in their desired control over their music listening, and engagement varies with listening context. We develop a series of metrics to capture music listening behaviour from users’ interaction logs. In a survey of 94 music listeners, we show significant correlations between music engagement from questionnaires and the presented quantitative metrics. We show how music retrieval can adapt to this engagement, developing a tablet-based demonstrator system, with an exploratory evaluation.

Author Keywords

Engagement; Music; Retrieval; Adaptation

ACM Classification Keywords

H5.2 Information interfaces and presentation: User Interfaces. – Graphical user interfaces.

INTRODUCTION

Music collections have grown to an extent that common retrieval techniques, such as scanning a list or navigating menus for something to listen to, would be excessively time-consuming. Users often have a choice between this burden of control, or yielding control entirely, using a recommender system or shuffling music at random. This issue is compounded in mobile contexts such as walking or driving, where fully attending to a retrieval process would be inappropriate. There is a need for music retrieval interactions that allow users to choose how much to engage and explore the large collections available. The music streaming service Spotify has over 20 million tracks available but only accessible by recommendation or specific query. Users have little means of exploring unknown content, with some 20% of tracks (4 million) never being listened to by anyone even once.¹

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.
MobileHCI '15, August 25 - 28, 2015, Copenhagen, Denmark
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3652-9/15/08 \$15.00
DOI: <http://dx.doi.org/10.1145/2785830.2785846>

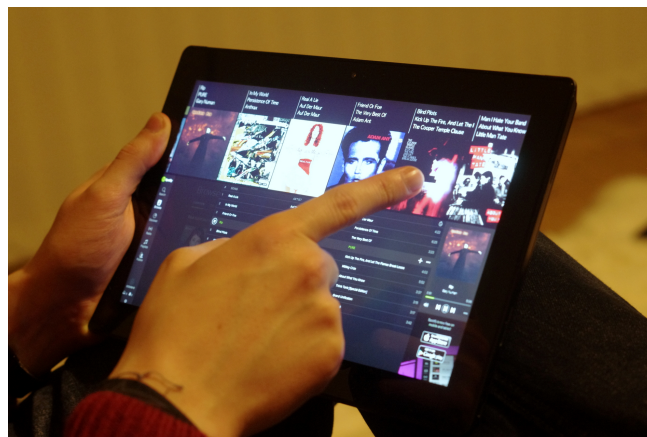


Figure 1. Our system (above) adapts to the user’s engagement, allowing users to explore music at a high-level (e.g. selecting angry music) or seamlessly engage to make specific selections (e.g. albums in that area).

There is increasing evidence that users vary significantly in terms of their engagement with music retrieval. More highly engaged users listen to more music and take more control over their music selection. Engagement with music also varies with listening context, including situational, social and multi-tasking factors [9]. Existing retrieval interfaces cater to specific levels of user engagement. Mobile interfaces often allow for casual input e.g. shuffle playback, tapping a rhythm [1] or radio-like tuning [11]. These are often paired with high engagement options, e.g. the menu hierarchy and textual retrieval seen in the Apple iPod² and Spotify.¹ Users are thus often forced to choose between extremes of engagement. The mobile user, often unable to fully engage, is disempowered in their music listening and left with a crippled experience.

We identify design guidelines for adapting music retrieval to the user’s level of engagement, allowing users to take as much control over their retrieval as is appropriate to their context. We envisage recommender systems that can be steered by users, to a degree determined by the user’s engagement. An example of such adaptation can be seen in Pandora,³ where users like or dislike tracks to steer recommendation. We show in an exemplar system how users can seamlessly take control over recommendation, from hearing any popular music, through to more specific sub-regions of music, all the way to the user explicitly selecting individual items.

1. <http://news.spotify.com/us/2013/10/07/the-spotify-story-so-far/> (11/08/14)
2. <http://www.apple.com/uk/ipodclassic/> (11/08/14)
3. <http://www.pandora.com> (11/08/14)

BACKGROUND

Too Much Choice?

While increased choice has been empirically shown to improve motivation and satisfaction [12], a growing body of literature claims that too much choice can overload users, in what is termed the ‘too-much-choice effect’ [21]. Requiring users to make an explicit selection from many choices not only incurs greater cognitive load and selection time (Hick-Hyman’s law [10]) but may also influence the happiness of the user. One cause of this unhappiness was proposed by Scheibehenne et al. to be a subsequent requirement to justify choices [21]. Given that music selection is often part of a social interaction, with a user judged by their selections [5], we can expect this effect to apply. This issue is compounded in mobile contexts such as when walking, which are often inappropriate for making specific selections or queries.

Satisficing & Selection

It has been widely observed that music listeners often have no particular song in mind when engaging with a music system and instead browse for a suitable piece of music [16]. This activity of finding the first ‘good enough’ item is termed *satisficing* whereas trying to optimise to find the best option is termed *maximising*. Schwartz et al. explored maximising versus satisficing as a personality construct, claiming that individuals vary with respect to how much they optimise selections [22]. Whether a user wishes to satisfice or make more exact selections will also depend upon their present context and how much they wish to engage in the music retrieval. Where users do not wish to expend any effort, they turn to recommendation or random shuffling, seeking a serendipitous music selection [17].

Interaction Engagement

Previous work has explored interactions which bridge from casual to engaged styles of use. Pohl and Murray-Smith outline a ‘focused-casual’ continuum, pointing out that users are empowered by selecting a level of engagement suitable to their social, physical or mental context [19]. If a music retrieval system enabled users to set their level of engagement, it could cover the maximising-satisficing scale – catering to individual users and their music-listening contexts. This dynamic has important implications for the user’s sense of ownership of the music interaction. It is likely that users would no longer feel the pride or embarrassment related to a track selection (as described in [5]) if control was shared with an intelligent music system.

In casual interactions, where the user exerts less control, a system can act more autonomously – making inferences from prior evidence about what the user intended. This way of handing over control was termed the ‘H-metaphor’ by Flemisch et al. where it was likened to riding a horse – as the rider loosens the reins, and exerts less control, the horse behaves more autonomously [7]. Music retrieval would benefit from spanning this ‘focused-casual’ continuum, allowing users to set a level of engagement appropriate to their context. Users would have a corresponding degree of control over the recommendations, ranging from biasing them to a style of music, through to the explicit selection of a particular track.

Music Engagement

The field of Music Psychology has explored the concept of *music engagement*, given by Greasley [8] as:

The extent to which music plays an integral role in a person’s everyday life, including the importance of music; the amount of music owned; uses of music; involvement in music-related activities; and various other factors ... people’s levels of engagement with music represents a spectrum, with clearly identifiable extremes (e.g. those who are either very highly engaged or who rarely engage with music in daily life), and overlaps in the middle (e.g. those showing characteristics of both high and low engagement).

It is clear that this concept of engagement is of importance to mobile music retrieval, however this definition refers to the relationship between listener and music, rather than the user’s engagement in a specific interaction or act of mobile retrieval. Greasley goes on to link music engagement to a user’s desire to control the music they listen to, with lower levels of engagement linked to radio usage rather than controlled song selection [8]. This aspect of music engagement overlaps significantly with the concepts of satisficing vs. maximising and interaction engagement. While the notion of engagement is multi-faceted and hard to define [15], there is work across disciplines identifying a common theme of user effort and desired control. Work within the field of Music Information Retrieval (MIR) has also touched on the idea of users having differing degrees of involvement with music, illustrating the range from *musical indifferents* to *musical savants* [3].

Measurement

As a relatively new area of research, there has been little work exploring how to measure a user’s music engagement. Greasley used 5 point Likert scales for a series of largely qualitative questions, aimed at different factors of music engagement elicited in user interviews. These factors were the user’s preferred level of *Control* over music selection, their *Collection* size, the user’s *Motivation* for music acquisition, *Memory* of first music purchase and the importance of *Lyrics* [8]. We adopt this questionnaire in this work, updating the wording to suit digital collections (see Appendix) and explore its relationship to other measures of engagement.

The use of experience sampling with wider Likert scales was introduced by Greasley and Lamont, demonstrating that more engaged music listeners tended to have greater control over the music they heard, often selecting it to evoke specific moods [9]. Related work by Krause et al. further explored how users control their music selection, also using experience sampling methodology (ESM) to show that music selection behaviour changes with music engagement as well as age [13]. The use of ESM allows for an in-depth analysis of a user’s music-listening behaviour, capturing music interactions in mobile contexts as well as in the home. It is, however, time-consuming to run and difficult to scale to many participants. Notably, ESM illuminates the variance in behaviour between sessions of music-listening within one participant – music engagement varies with listening context, including situational, social and multi-tasking factors [9].

Music Retrieval Interfaces

Music retrieval interfaces typically afford very low or very high user engagement and control, as discussed in the introduction. These are often complementary, for example the textual search and ‘radio’ recommenders in Spotify, or the hierarchical menu and shuffle features of the iPod. Adaptivity is sometimes employed, for example the relevance feedback in Pandora. Stober and Nürnberger give a thorough treatment of approaches for adapting music retrieval to user context [23].

The radio metaphor for casual music interaction is compelling for mobile use, building upon a familiar and successful music retrieval interface, especially in driving contexts where users can spare limited attention. Previous work by Hopmann et al. sought to bring the benefits of interaction with vintage analog radio to an engaged and controlling style of retrieval in modern digital music collections [11]. They point out that the status quo of using a hierarchical organisation of music does not support free navigation. Their system enabled navigation of a 1D music space, arranged by genre/artist/album, with the look-and-feel of a vintage radio. Earlier work on radio-like interaction by Crossan and Murray-Smith, using rhythmic input, incorporated the uncertainty of the user’s input into the retrieval [4]. This approach allowed users to choose how well defined they wished their rhythmic queries to be. This query specificity and effort required to produce a query mirrors the concepts of music engagement discussed. Later work on rhythmic retrieval made linked query specificity and user engagement, allowing users to casually query by tempo, or reproduce a specific song [1].

MEASURING USER INTERACTION

By combining detailed music listening histories (acquired from Last.fm), with song durations acquired from Spotify, we propose a number of measures of users’ interaction with their music retrieval systems. Notably, many mobile music players support Last.fm play logging. Adding a song’s duration to the time its playback started gives an indication of when the song would end without further intervention. Where a new song begins immediately after the expected end time of a previous song, this is indicative of automatic playback such as playlists or albums – evidence of a lack of user intervention. Songs which are interrupted early by a subsequent song, or which are followed by another song after a brief pause are evidence of users intervening to find new music to listen to.

Album-Based Playback

Despite the shift to digital music consumption, listening to music in the form of albums remains a common music-listening behaviour. Detecting this behaviour from a user’s music listening logs is trivial, as consecutive tracks will have the same album metadata A_s . We assume that where consecutive tracks are played from the same album, this was due to their being from the same album. The ordering of tracks on the album is not considered, as this would introduce a confound with shuffle versus sequential playback. The detection of selection being an *Album*-based behaviour is thus:⁴

$$Album = \frac{1}{n-1} \sum_{s=2}^n [A_s \equiv A_{s-1}].$$

Interventions

The user interventions described can be stated formally as *Skipping* and *Switching*. For a song s with start time t_s , song duration d_s and subsequent song start time t_{s+1} :

$$\Delta t_s = t_{s+1} - (t_s + d_s)$$

$$Skipping = \frac{1}{n-1} \sum_{s=1}^{n-1} [\Delta t_s < 0].$$

Assuming that automatic song playback occurs within a threshold of $t_a = 100ms$ of a previous song ending and that user intervention to switch or select a new song will occur within a threshold of $t_t = 30s$, yields:

$$Switching = \frac{1}{n-1} \sum_{s=1}^{n-1} [\Delta t_s > t_a \wedge \Delta t_s < t_t].$$

The assumed value for t_t is conservative to ensure that the measured pauses are part of a continuous music retrieval interaction. The aim is to detect whether a song was selected automatically or by manual intervention; delays greater than 30s become increasingly likely to be due to external events.

User intervention in music listening can be considered as a Bernoulli process i.e. a binary choice between whether a user selected a song or it was automatic. The examples in this work consider users across all their listening sessions, assuming interventions are a homogeneous process (with a fixed probability). In practice, user behaviour will inevitably change with context, however these assumptions still allow for a user’s overall behaviour to be modelled. Users can thus be characterised in terms of their *Intervention Rate* – the rate at which they intervene with a *Skipping* or *Switching* event.

Speed

While the occurrence of an intervention may provide some information about music-listening behaviour, we hypothesise that the speed at which users intervene may add further detail. We propose measures of *Switching speed* – how quickly a user switches to a new track after a previous track ends, and *Skip speed* – how far, proportionally, into a song does a user decide to skip track. In the case of *Skipping*, we normalise time against track length, given the large variation in track durations. To mitigate the effect of large outliers where a track nearly completed, the harmonic mean of time is used. As we are interested in speed, we then take the reciprocal, yielding a simple equation. Considering only the n songs where a *Skipping* event has been detected:

$$Skip\ speed = \frac{1}{n-1} \sum_{s=1}^{n-1} \frac{d_s}{\Delta t_s}.$$

We calculate *Switch speed* in a similar manner. Considering only the n songs where a *Switching* event has been detected:

$$Switch\ speed = \frac{1}{n-1} \sum_{s=1}^{n-1} \frac{1}{\Delta t_s}.$$

4. Iverson bracket notation used for clarity.

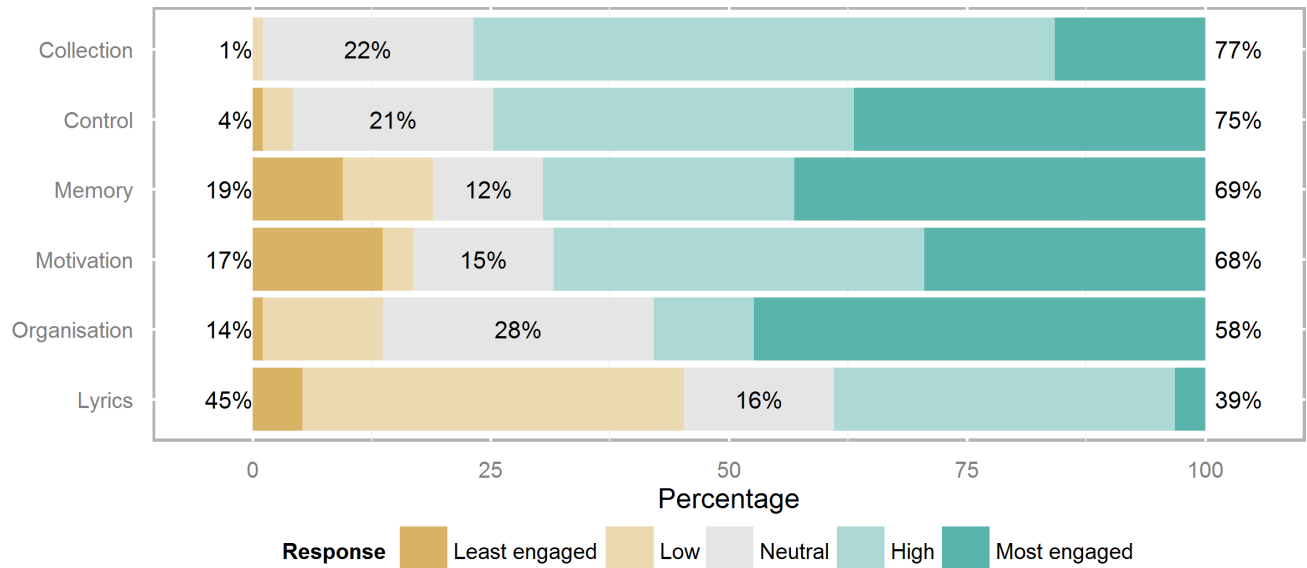


Figure 2. Responses to our adapted music engagement questionnaire, $n = 95$. Full questions and responses can be viewed in the Appendix. Respondents were Last.fm users and thus responses were more likely to show higher engagement with music.

EXPLORATORY STUDY

We conducted an exploratory study to link the concept of music engagement to the way users interact with their music retrieval system. The music engagement and music selection questionnaires from music psychology were combined and updated for digital collections (see Appendix) and posted along with a request for Last.fm usernames to a messaging board for Last.fm users. These usernames were used to data mine respondents' listening history in order to calculate our intervention measures, giving an insight into how users interacted with music retrieval systems over many years. We received 95 responses, shown in Figure 2, of which we were able to mine 89 usable listening histories. 88% reported use of mobile music selection, 93% use of a computer, 60% use of streamed music and only 18% use of radio.

The questionnaire responses were mostly skewed toward high engagement with music. The importance of *Lyrics* has a distinctly different distribution. Cronbach's α for all factors was 0.43, indicating that the questionnaire is covering a number of distinct facets of users' music engagement. How much *Control* a user likes to have over music selection is the most relevant of the factors to the measures of intervention, with the others addressing more qualitative aspects. We calculated the measures of *Album* based listening, *Intervention Rate*, *Skip speed*, *Select speed* from the listening histories of the respondents, and their correlations with *Control*, given in table 1.

	Control	Album	Interv.	Skip Speed
Control				
Album	0.26*			
Interventions	-0.29*	-0.51**		
Skip Speed	0.48**	0.29**	-0.74**	
Select Speed	0.25*	0.51**	-0.84**	0.47**

Table 1. Polyserial correlations of users' self-reported *Control* over music selection and the measures of music-listening behaviour derived from their listening histories. * $p < 0.05$, ** $p < 0.01$.

Listener Profiles

The results in table 1 paint a somewhat surprising picture of how music engagement influences user interaction. *Album* based listening showed significant moderate correlation with users' self-reported preferred *Control*. Naturally, this also meant a significant inverse correlation with user intervention in music playback. While it might appear counter-intuitive that desire for *Control* leads to fewer interventions, engaged listeners are more likely to invest in selecting a specific album. Casual music listening, by shuffle or recommendation, requires less initial investment, but then more corrective actions are required. We confirmed our album detection with an independent-samples t-test between users who had or had not listed album-based selection, with significant differences in *Album*, *Skip speed* ($p < .01$), and *Interventions* ($p < .05$).

Control also had significant correlations with *Skip speed* and *Select speed*. More musically engaged listeners made quicker interventions in their music listening. This adds to the profile of musically engaged listeners who know what they want and who need to make few interventions, which are decisive and quick. Combining the qualitative work of music psychology with our empirical measures of how these users interacted with their music retrieval interfaces, we can identify the following listener profiles as targets:

Engaged These users have a high initial engagement in the interaction, with more specific retrieval queries e.g. selecting a particular album. They then make very few further interventions, which are quick and decisive.

Casual These users wish to satisfice, investing little effort in the retrieval at any given point. Their lack of initial control means that these users need to be able to easily make corrective interventions.

Mixed Most users vary between levels of *Casual* to *Engaged* music listening, depending on their listening context.

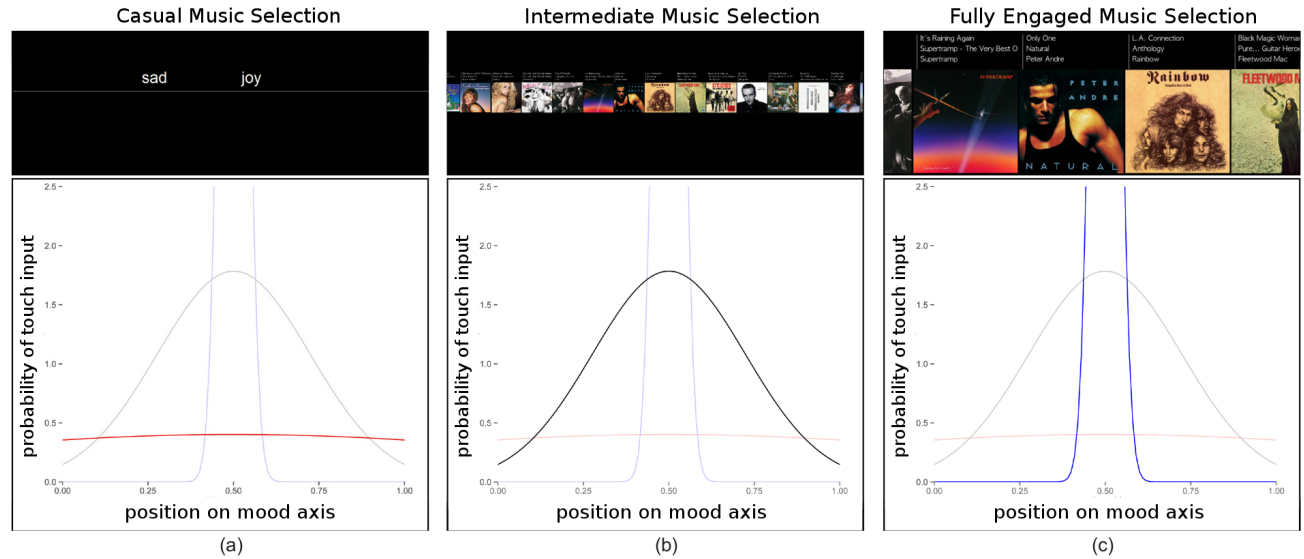


Figure 3. As the user exerts control, the distribution of predicted input for a given song becomes narrower. This means the system can give more weight to the user’s input and infer a belief about their listening intent over fewer songs. (a) shows a casual music selection, with no pressure applied, the user sees the entire axis with mood labels to give a sense of the layout. (b) shows a more engaged music selection, with the user now applying pressure, the view zooms semantically to show album artwork for a range of music of a particular mood, allowing users to get more specific recommendations of popular music. (c) shows a fully engaged music selection, when maximum pressure is applied, users are able to zoom in to view a specific album artwork and make an exact selection, with the system deferring entirely to their input.

EXAMPLE SYSTEM

To demonstrate how mobile music retrieval can scale with a user’s engagement, we augmented a popular music retrieval interface (Spotify) on a tablet as in Figure 1. We add a semantic zooming view of a simple linear music space, enabling both casual and engaged forms of interaction – giving users varying degrees of control over the selection of music. The music was arranged on one axis as this is the simplest possible, and requires the least engagement. By allowing users to make selections from the *general* to the *specific*, the new interface supports both maximising and satisficing. Users can make broad and uncertain *general* selections to casually describe what they want to listen to. However, they can also exert more control over the system and force it to play a *specific* song. A recommender system infers relevant music given user input, engagement and prior evidence of listening intent.

Inferring Listening Intent

As the user yields control over the retrieval, the system is able to act in an increasingly autonomous fashion. The input from the user can be considered coarser as they become less engaged, having less weight on the behaviour of the system. The intended selection can then be inferred from prior evidence of music listening intent, for example song popularity, previous listening history or social recommendations. While the user’s input itself may be coarser when the user is less engaged,⁵ in other cases similar input must be evaluated in the context of different engagement levels. An example would be a touchscreen event when a user is casually browsing through an overview of a music library versus an explicit selection of a piece of music. In order to interpret this input, the system must be aware of the engagement context, for example by sensing it with the mobile device’s accelerometers [14].

While engagement could be inferred from user behaviour, we allow users to express it explicitly by pressing on a pressure sensor with their left hand as they navigate with their right – employing a design metaphor of physical exertion as engagement. The levels of engagement (and handover of control) may be continuous or occur on a number of discrete levels. Discretisation allows distinct operating modes that are clearly delineated, but burdens users with understanding the implications of each mode, and with managing transitions between multiple modes. As an alternative, we explore the use of engagement as a continuous variable, acting as a parameter in the inferential system and changing the dynamics of a single and consistent interaction.

As the user begins an interaction, they have not applied pressure and therefore are not highly engaged in the retrieval. The view of the music collection is zoomed out and shows mood labels describing the layout of the music space. The inferred selection is thus broad, covering an entire mood region and is biased towards popular tracks (Figure 3a). At low levels of engagement, users receive recommendations of highly popular tracks. When the user applies pressure, the system interprets this as the user taking control to make a more exact selection. The inferred selection is smaller, the view zooms semantically to show the album art of the selection (Figure 3b). This selection is a combination of evidence from the user’s navigated position with prior evidence i.e. song popularity. Users retain the ability to make explicit album selections by fully applying pressure (Figure 3c). By varying the pressure, users can seamlessly set their level of engagement and control over music selection.

5. for example in Pohl and Murray-Smith’s work on casual interaction [19]

Inference

The handover from casual to engaged interaction relies upon an inferential model of user listening intent. We assume that when making casual selections, users will navigate to a broad area of musical interest, e.g. happy music. We also assume that as users engage and wish to make more specific selections, they will use the richer album art and metadata feedback to navigate to a few (or one) albums of interest. We can encode these assumptions about the user’s behaviour as a generative model i.e. one that predicts the user’s inputs for a given level of engagement, as in Figure 3. Put simply, it predicts casual users will point roughly near the type of music they want and engaged users will navigate to exactly what they wish to hear.

We implement this generative model of user input i_x along the x axis conditioned upon a target song and level of engagement. The input is modelled using a Gaussian Mixture Model, with a component corresponding to each song s_i within the music collection. Each component is a Gaussian distribution, with the centroid positioned according to the song’s position x_s in the one-dimensional music projection, with distribution width controlled by the precision τ :

$$p(i_x | s_i) = \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau(i_x - x_s)^2}{2}}.$$

The precision τ parameter is inversely proportional to the unit variance σ_d^2 of the distribution. As users apply pressure, the precision is scaled by the denoted engagement $E \in (0, 1]$:

$$\tau = \frac{E}{\sigma_d^2}.$$

This generative model can be used to infer a belief that a given song s_i is of interest to the user, conditioned upon an input position i_x and the level of user engagement E :

$$p(s_i | i_x, E) = \frac{p(i_x | s_i, E) p(s_i | E)}{p(i_x | E)}.$$

The prior belief $p(s_i | E)$ over the music space allows existing evidence of listening intent to be incorporated, conditioned on the current level of engagement. In our work we use the music’s current popularity on Spotify so that the system can recommend popular music to the user. We marginalise out engagement from this prior, making the simplifying assumption that only expected input varies with engagement and not music preference. Our approach does, however, allow for supporting differing listening preferences based on engagement, if such data were available.

Mapping Interaction to Inference

The level of detail presented to the user should be appropriate for their current level of engagement. This is a natural application for semantic zooming – users making broad selections would not benefit from seeing the artwork of each album and so we zoom out to showing only general labels such as a mood. As users engage and the interface zooms in, the mood labels give way to popular artists and then artwork.

The greatest disadvantage to generative modelling is that where the model does not conform well to the user, the user will be unable to predict the behaviour of the system. As the interaction is increasingly conditioned upon the inference, the exposure to this problem is greater. This can be mitigated by adhering to the principle of ‘direct manipulation’ such that the mapping between input and output is clear to the user. In this work we take a conservative approach, where engagement is taken as being linear with exerted pressure and the semantic zooming is done linearly with engagement. We sample tracks from the inferred posterior distribution of songs in real-time and highlight their album art with a faint glow. Tracks are also shown in the playlist frame of the Spotify UI in order of inferred listening intent.

Projecting the music space

The simplest representation of a music space is a one dimensional projection. An assumption of our system is that similar songs are placed near to each other, such that when users are disengaged, they can select a whole coherent region of the music space. We use information about the mood of the music to arrange the collection however genre or other features and metadata could be used. We take a high dimensional feature space produced by MoodAgent,⁶ a commercial music signal processing system focused on mood-related music features. While we could use a single one of these features (e.g. tempo) to arrange the music collection, we attempt to maintain more of the information by using non-linear dimensionality reduction. Venna et al. introduce an information retrieval perspective on such techniques, presenting their state-of-the-art NeRV algorithm for balancing recall and precision in the projection [24]. This trade-off between keeping similar items together and dissimilar items apart can be modified using a λ parameter. Using NeRV we project the high dimensional music feature space into a one dimensional space – we show how the original features influence this projection in Figure 4. In this example, anger is inversely correlated with tenderness, with joyful music peaking in the centre of the projection.

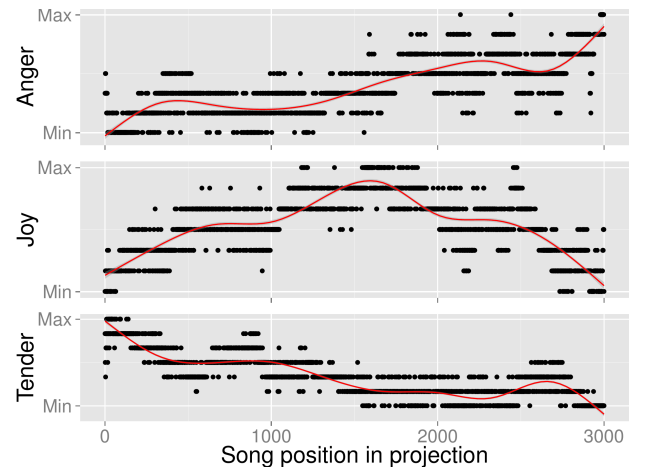


Figure 4. The music is arranged by taking the mood features such as Anger, Joy and Tender and creating a one-dimensional projection.

6. <http://www.moodagent.com/> (11/08/14)

EVALUATION

It is inherently difficult to find an objective means of evaluating such a personalised music system. Schedl and Flexer identify this challenge as a key reason that little work has been done in personalised music systems [20]. They note the need to consider the user from an early point in the development process. Our generative modelling based approach allows us to simulate user input to characterise the behaviour of the system. We also evaluated our approach and design assumptions in a series of design sessions with music listeners. In our evaluations, the music space used was taken from the UK Top 40 singles charts over the past 50 years. Audio was streamed from Spotify, with our engagement-dependent interface integrated within the Spotify web player UI.

Engagement-Dependent Retrieval

As the system is based upon a generative model of user input, we are able to easily characterise the input required from users in various scenarios. Using the design assumption of users casually choosing popular music, we show how input required from users varies with the popularity of the retrieved track. Figure 5 shows how accurately the user must navigate to a song for it to be played, at different levels of engagement and for songs of different popularity. We can thus link the behaviour of our system to the music listening profiles we identified as targets from our exploratory study.

Casual This profile called for allowing users to satisfice and make corrective actions with little effort. The example system supports this with recommendations of popular music from the broad area navigated to, requiring only swipe gestures to select or change music. The recommender system largely controls the music selection, with coarse high-level influence from user input. Notably, it is impossible to select the least popular songs at low engagement.

Engaged This profile called for allowing users to invest effort in making an exact selection of an album. The system supports this by letting users apply pressure to zoom into a specific mood region, or even an album of interest. At high levels of engagement, the user input dwarfs the recommender’s influence, with popularity having little effect.

Mixed The intermediate levels of engagement are supported by letting users quickly set their engagement, with the continuous handover of control using the pressure sensor. As users engage, their input has increasing influence. Semantic zooming gives users feedback about how specific a query at a given level of engagement would be.

Design sessions

Six participants, in groups of 2, took part in informal design sessions which focused on discussing the music interaction described in this work, and in particular the benefits of being able to engage with a mood-based semantically zooming view. Each was given time to freely use the tablet to explore the collection and experiment with making a variety of music selections at different levels of engagement. We used this experience to generate discussion with each of them about how they currently find and listen to music and how our system might support or enhance those habits.

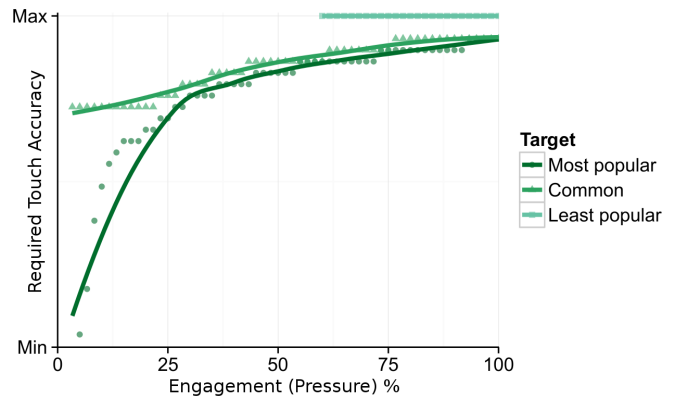


Figure 5. The required accuracy for song selection against user engagement. Popular songs can be selected by casually navigating near them. Unpopular songs are only selected when the user is highly engaged and navigates exactly to them. As engagement increases, the user takes control from the recommender, and songs become equally targetable.

Shuffle and Casual Music Selection

Five of the participants reported using both shuffle and explicit selection to play music, while only one exclusively used explicit selection. Participants stated that they typically use shuffle for background music or alleviating the social pressures of selecting music, as discussed in the literature review. This supports our prediction that the ability to vary the degree of engagement is desirable, with four participants often creating playlists of songs specifically to shuffle over. These playlists were based on personal heuristics and were often kept to a small number of songs to guarantee similarity between items. We highlight that users are seeking to engage with their music system at a level between shuffle and explicit selection, which is often unsupported in existing interfaces.

Curation and Taking Control

Participants all expressed a desire to curate the collection. One suggested that there could be more room for subtle differences between tracks in a personal collection, which would allow finer grained control over even uncertain selections. One pair felt that having more control over the music in the collection would lead to a greater understanding of the organisation, commenting that “I’d like to be able to apply more of my own decision making.” This sentiment was echoed in comments about the labels. One participant commented that self-defined labels would be better, even if not fully representative of the underlying mood – “people might be annoyed if all their favourite albums are classed as ‘sad’... I would prefer ‘melancholic’, it has nicer connotations.” Clearly, a desire for customisation is present – a form of even further engagement in the music retrieval interaction.

Sources of Evidence

Our implementation uses music popularity as a source of evidence for weighting songs as more likely to be played. This comes from a design assumption that satisficing is supported by playing such songs. The design session revealed that users felt that in doing this recent additions to their music collection were underplayed. By including when the album was added to the collection as prior evidence, we can bias the selection towards recent additions as well and support this use case.

DISCUSSION

The exploratory study confirmed the arguments made in the literature review, showing that music engagement and desired control over music retrieval varies with user and listening context. This result is certainly not controversial, however we are also able to link new measures of listening behaviour to self-reported desire for control, yielding listening profiles at different levels of engagement. Of particular note is that desire for greater control over music actually correlated with less interventions and the use of album-based playback. This shows that high music engagement manifests as an initial burst of engaged music retrieval, and it is possible that over an entire listening session, the casual listener may even have interacted more with their music system.

The measures of music listening we developed allow us to make a useful distinction in users' listening behaviour. While users wishing to have control actually intervened less, where they did intervene, it was done quickly. This raises an interesting question regarding how engagement in interaction is presently considered – whether more interaction events means greater engagement versus a few decisive and controlling interaction events. We must also consider that musically engaged listeners may be more likely to consider albums as a single work of art, rejecting track by track selection. Album-based listening inherently means fewer interventions (hence the strongest correlation), and engaged music listeners may not wish to interfere with an album. Music retrieval is shaped by such cultural factors. For example, Last.fm users logging their listens are biased to high engagement, and further work is needed to test if our results generalise.

Measuring Music Engagement

It is clear from the music engagement questionnaire, and the measured internal reliability, that the questions cover a number of loosely connected factors. This is to be expected, given the broad definition of music engagement, and there is thus a need to consider factors of engagement individually. In the context of music retrieval, we have highlighted the *Control* factor as being particularly relevant. While it is important to consider the affective factors in order to fully capture user engagement [15], there is an opportunity to develop a quantitative, inter-disciplinary theory of engagement as control, building on work using control theory and measuring interaction in bits [19]. Such an approach can be also linked to the specificity of information need in Information Retrieval, as has been done here with music.

The example system we present is useful in showing how retrieval can adapt to a user's engagement. It also serves to highlight the challenges facing engagement-dependent retrieval. Most notably, we asked users to explicitly denote their engagement, having them squeeze a pressure sensor as a proxy. While this is a reasonable metaphor to employ, it would be desirable to infer a user's engagement from enhanced sensing and previous behaviour. Also, our approach to adapting to engagement involved developing models of user behaviour across levels of engagement, requiring design insight or model fitting. The evaluative design sessions highlighted that users would prefer to customise such a system, as part of a highly engaged music interaction.

Music Spaces

The behaviour of this system is closely linked to the music space used. In this work we used a music space likely to be familiar to all participants, however this does introduce some issues. In early testing we used the most popular 5000 tracks from the UK charts over the last five decades and found that the resulting music space was too homogeneous to navigate meaningfully. In an effort to increase the diversity of the music space, we randomly sampled from the music collection instead, yielding the distinct mood regions seen. The 1D space broadly reflected the culturally universal arousal aspect of mood [6]. The use of mood itself is supported by evidence that mood features describe music as well as genre [2] and that users listen to music by mood [18].

Users disliked when dissimilar songs were together. A benefit of using the NeRV algorithm to project the music arrangement is being able to adjust the λ parameter to favour precision (keeping dissimilar items apart) over recall (keeping similar items together). Using feedback from users while iterating the design of our system, we biased λ strongly toward precision to keep dissimilar items apart and improve perceived system quality. The need to avoid stark outliers has been noted within MIR; Paul Lamere of Echonest introduced the 'WTF test' for automatic playlists, with systems scored negatively for each outlier.⁷ The true test for an interface such as the one presented would be to use the user's own music collection for the projection, as part of a longitudinal study of music listening behaviour.

Market-Based Evaluation

We have designed an engagement-dependent retrieval system and developed measures of music listening behaviour and engagement. While we have shown the behaviour of the system using technical measures and qualitative feedback, it would be desirable to study the longitudinal impact that using such a system would have on listening behaviour. Though we could conduct further small-scale studies, it is preferable to widely deploy a full music system, which users can adopt for their day to day listening. The concepts we present in this work have been applied in a commercial, tablet format, engagement-dependent music retrieval system – the BeoSound Moment, announced at CES 2015. This product will be used for further evaluation, measuring the impact of these approaches on listener behaviour.

Wider Application

The engagement-dependent approach has applications beyond music retrieval. Users have access to large volumes of media such as movies and photos, e.g. with Netflix and Flickr. We argue that users are better served by letting them engage in a retrieval that is appropriate to their context. While recommender systems have done much in the way of alleviating users of the issue of too-much-choice, there will always be cases where users wish to have more input into the selection. Incorporating user engagement into recommender systems allows users to benefit from recommendation, while retaining a degree of control that suits their current context.

7. <http://musicmachinery.com/2011/05/14/how-good-is-googles-instant-mix/> (11/08/14)

CONCLUSION

This work introduces users' music engagement as important for the development of music retrieval interfaces, especially in mobile contexts. We combined the music engagement questionnaires from music psychology to survey music listeners, and developed novel metrics to capture their retrieval behaviour from their listening histories. We then identified relationships between these measures and presented listener profiles to guide the design of music retrieval interfaces.

A tablet-based system was developed that can adapt to users' engagement, supporting the listener profiles we identified. A flexible recommender system allows users to broadly select a style of music, or to take control and make more specific music selections. A generative model of user input at different levels of engagement enables the system to infer relevant music, with music popularity providing prior evidence. In an agent based evaluation, we show that the system supports the range of interaction styles given in the listener profiles, to allow mobile music retrieval at a user's chosen level of engagement. Design sessions were also held, showing that users liked to vary their engagement but that some users wished to have even greater control. This work contributes a set of novel metrics of music listening using insights from psychology, MIR and HCI, and also an engagement-dependent approach to music retrieval adopted in a commercial product.

ACKNOWLEDGMENTS

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

REFERENCES

1. Boland, D., and Murray-Smith, R. Finding my beat: personalised rhythmic filtering for mobile music interaction. In *Proc. MobileHCI*, ACM (2013).
2. Boland, D., and Murray-Smith, R. Information-theoretic measures of music listening behaviour. In *Proc. ISMIR* (2014).
3. Celma, O. *Music Recommendation and Discovery: The Long Tail, Long Tail, and Long Play in the Digital Music Space*. Springer, 2010.
4. Crossan, A., and Murray-Smith, R. Rhythmic interaction for song filtering on a mobile device. In *Proc. HAID*, Springer (2006).
5. Cunningham, S. J., Jones, S., and Jones, M. Organizing digital music for use: an examination of personal music collections. In *Proc. ISMIR* (2004).
6. Egermann, H., Fernando, N., Chuen, L., and McAdams, S. Music induces universal emotion-related psychophysiological responses: Comparing canadian listeners to congolese pygmies. *Frontiers in Psychology* 5, 1341 (2015).
7. Flemisch, O., Adams, A., Conway, S. R., Goodrich, K. H., Palmer, M. T., and Schutte, P. C. The H-Metaphor as a Guideline for Vehicle Automation and Interaction. Tech. rep., NASA, 2003.
8. Greasley, A. E. *Engagement with music in everyday life: and in-depth study of adults' musical preferences and listening behaviours*. PhD thesis, University of Keele, 2008.
9. Greasley, A. E., and Lamont, A. Exploring engagement with music in everyday life using experience sampling methodology. *Musicae Scientiae* 15, 1 (2011), 45–71.
10. Hick, W. E. On the rate of gain of information. *Quarterly Journal of Experimental Psychology* 4 (1952), 11–26.
11. Hopmann, M., Vexo, F., Gutierrez, M., and Thalmann, D. Vintage radio interface: Analog control for digital collections. In *Proc. CHI* (2012).
12. Iyengar, S. S., and Lepper, M. R. When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology* 79, 6 (2000), 995–1006.
13. Krause, A., North, A., and Hewitt, L. Music selection behaviors in everyday listening. *Journal of Broadcasting & Electronic Media* 58, 2 (2014), 306–323.
14. Kuhn, M., Wattenhofer, R., Wirz, M., Fluckiger, M., and Troster, G. Sensing Dance Engagement for Collaborative Music Control. In *Proc. ISWC*, IEEE (2011).
15. Lalmas, M., O'Brien, H., and Yom-Tov, E. *Measuring User Engagement*. Information Concepts, Retrieval, and Services. Morgan & Claypool, 2014.
16. Laplante, A. *Everyday Life Music Information-Seeking Behaviour of Young Adults: An Exploratory Study*. Master's thesis, McGill University, Quebec, 2008.
17. Leong, T., Vetere, F., and Howard, S. The serendipity shuffle. In *Proc. OzCHI* (2005).
18. Lonsdale, A. J., and North, A. C. Why do we listen to music? a uses and gratifications analysis. *British Journal of Psychology* 102, 1 (2011), 108–134.
19. Pohl, H., and Murray-Smith, R. Focused and casual interactions: Allowing users to vary their level of engagement. In *Proc. CHI* (2013).
20. Schedl, M., and Flexer, A. Putting the User in the Center of Music Information Retrieval. In *Proc. ISMIR* (2012).
21. Scheibehenne, B., Greifeneder, R., and Todd, P. M. What moderates the too-much-choice effect? *Journal of Psychology & Marketing* 26(3) (2009), 229–253.
22. Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., and Lehman, D. R. Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology* 83 (2002), 1178–1197.
23. Stober, S., and Nürnberger, A. Adaptive music retrieval-a state of the art. *Multimedia Tools and Applications* 65, 3 (2013), 467–494.
24. Venna, J., Peltonen, J., Nybo, K., Aidos, H., and Kaski, S. Information retrieval perspective to nonlinear dimensionality reduction for data visualization. *Journal of Machine Learning Research* 11 (2010), 451–490.

APPENDIX

The following music engagement and use questionnaire is based on similar questionnaires in [9, 13]. They covered the factors of music engagement and the selection methods used in music listening. We combined these questionnaires and also updated the wording of some questions to reflect digital music listening. This following questions were presented to users in an interactive Google Forms webpage, accessible via a link posted to forums of last.fm users. The first two questions are multiple choice, while the remaining are forced choice Likert scale (1–5) to reflect music engagement.

Questionnaire

The aim of this questionnaire is to gather data on people's music listening behaviour. You will be asked to provide your Last.fm username so that your publicly shared Scrobbles can be analysed. Completing this questionnaire is taken as consent for this analysis.

Please enter your Last.fm username:

We will only analyse your public Scrobbles to identify your music-listening behaviour.

Think about the last few times you listened to music and tick the selection methods you used.

- ☐ I did not have any control
- ☐ It was performed live at the time
- ☐ Watched TV
- ☐ Personal premade playlist
- ☐ Specific artist
- ☐ Premade playlist - by someone else
- ☐ Specific song
- ☐ Listened to the radio
- ☐ Specific album
- ☐ Someone I was with chose
- ☐ Random/Shuffle
- ☐ Computer Recommendation
- ☐ Website streaming
- ☐ Other:

Please tick the devices you often use to listen to music.

- ☐ Mobile MP3
- ☐ Mobile Phone
- ☐ Mobile CD
- ☐ Computer - own music
- ☐ Computer - streamed music
- ☐ Computer - music store in the cloud
- ☐ Stereo - MP3
- ☐ Stereo - CD
- ☐ Radio
- ☐ TV
- ☐ In public - live
- ☐ In public - recorded

How much control do you like to have over the music you listen to?

- ☐ I generally listen to whatever is played
- ☐ I listen to a particular radio channel or recommendations in a style of music I like
- ☐ Sometimes I like to choose, sometimes I'll listen to radio or recommendations
- ☐ I generally prefer listening to music I've chosen e.g. in a playlist I made
- ☐ I like to have full control over which album or song I'm listening to

How much music do you have in your collection (in iTunes, MP3 player, Spotify etc)?

- ☐ Up to 5 albums or about 50 MP3s
- ☐ Up to 25 albums or about 250 MP3s
- ☐ Up to 125 albums or about 1250 MP3s
- ☐ Up to 5000 albums or about 50,000 MP3s
- ☐ More than 5000 albums or 50,000 MP3s

What best describes your music collection?

- ☐ All my collection is organised by hand in a certain way
- ☐ My collection is broadly categorised by hand (e.g. genre or mood playlists/folders)
- ☐ My music system organises my collection for me
- ☐ There is no specific organisation to my music collection
- ☐ I can never find anything when I want it!

Can you remember the first album you bought?

- ☐ Yes, I can remember what it was and exactly when and where I bought it
- ☐ I can remember what it was, but not exactly when I bought it
- ☐ Not off hand, I probably could if I thought about it
- ☐ I doubt it
- ☐ I have no idea

Why did you make your last music purchase?

- ☐ I had to have it, I heard it and I just had to go and buy it
- ☐ I knew I would like it, because I've built up a knowledge of what I like
- ☐ I'd heard a couple of the tracks I liked so bought it to see if the whole thing was good
- ☐ It was recommended to me, so I thought I'd give it a go
- ☐ I can't remember

Lyrics in music, which best describes you?

- ☐ I never really listen to the lyrics in songs
- ☐ With some of my favourite music I know the lyrics, but otherwise I don't really pay attention to the lyrics in songs
- ☐ I like to know the lyrics so I can sing along
- ☐ I have to know the lyrics because it's central to understanding what the artist is trying to convey
- ☐ I have to know the lyrics because I don't want to be singing along to something that might be at odds with my beliefs