Inferring Music Selections for Casual Music Interaction

Daniel Boland
University of Glasgow
United Kingdom
daniel@dcs.gla.ac.uk

Ross McLachlan
University of Glasgow
United Kingdom
r.mclachlan.1@research.gla.ac.uk

Roderick Murray-Smith
University of Glasgow
United Kingdom
rod@dcs.gla.ac.uk

ABSTRACT

We present two novel music interaction systems developed for casual exploratory search. In casual search scenarios, users have an ill-defined information need and it is not clear how to determine relevance. We apply Bayesian inference using evidence of listening intent in these cases, allowing for a belief over a music collection to be inferred. The first system using this approach allows users to retrieve music by subjectively tapping a song's rhythm. The second system enables users to browse their music collection using a radio-like interaction that spans from casual mood-setting through to explicit music selection. These systems embrace the uncertainty of the information need to infer the user's intended music selection in casual music interactions.

Categories and Subject Descriptors

H.5.2 [**Information interfaces**]: User Interfaces

General Terms

Design, Human Factors, Theory

1. INTRODUCTION

When interacting with a music system, listeners are faced with selecting songs from increasingly large music collections. With services like Spotify, these libraries can include many songs the user has never heard of. This retrieval is often a hedonic activity and may not serve a particular information need. Users do not always have a song in mind and are often just interested in setting a mood or finding something 'good enough' [9]. This type of *casual search* has recently been identified as not being well supported within IR literature [14]. In particular, the concept of relevance becomes nebulous where the information need is not well defined. By inferring a belief over a music collection using the likelihood of a user's input, we implement interactions which incorporate this uncertainty. These interactions can account for subjectivity and span from casual, serendipitous listening through to highly engaged music selection.

Music listeners are not always fully engaged with the selection of music - as evidenced by the success of the shuffle playback feature. Large libraries of music such as Spotify are available but users often just want background music, not a specific song out of millions. In these casual search scenarios, users often *satisfice* i.e. search for something which is 'good enough' [11]. As this information need is poorly defined, so too is relevance, placing these interactions outside of typical Information Retrieval approaches.

2. UNCERTAIN MUSIC SELECTION

By asking *'What would this user do?'*, we can develop a likelihood model of user input within an interaction. With Bayes theorem, this allows for an uncertain belief over a music space to be inferred. Users can provide evidence of their listening intent as part of a casual music interaction, not needing to be fully engaged in the music retrieval. This is an explicitly user-centered approach, focusing on how a user will interact with the system. Both the systems discussed here have been iteratively developed by comparing real user behaviour against that predicted by the user input models. We present two novel music retrieval systems which explore two challenges with this approach: i) how to correctly interpret evidence which may be subjective and ii) how to allow users to set their current level of engagement:

i) 'Query by Tapping' is a music retrieval technique where users tap the rhythm of a song in order to retrieve it [1]. As part of a user-centred development process, we identified that rhythmic queries are often subjective and so developed a model of rhythmic input which captures some of this subjective behaviour. This allows for the system to be trained to the user's tapping style, giving significant improvements over previous efforts at rhythmic music retrieval.

ii) FineTuner is a prototype of a radio-like music interface that enables users to retrieve music at a level of engagement suited to their current information need. Users navigate their music collection using a dial, with the system using prior knowledge of the user to inform the music selection. A pressure sensor enables users to assert varying levels of control over the system – with no pressure, users can casually tune in to sections of their music collection to hear recommended music with common characteristics. As pressure is applied, the user is able to make increasingly specific selections from the collection. The inferred music selection is conditioned upon the asserted control, allowing for the seamless transition from casual mood-setting to engaged music interaction.

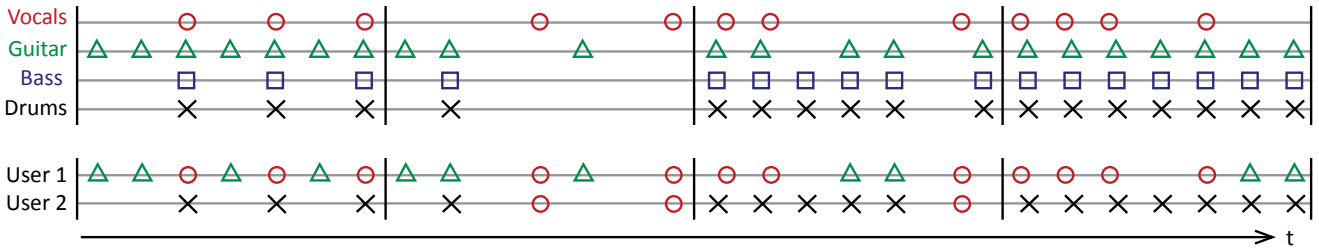

Figure 1: Users construct queries by sampling from preferred instruments. User 1 prefers Vocals and Guitar whereas User 2 prefers Drums and Bass.

3. MODELLING SUBJECTIVITY

In this section we describe our efforts to model the subjectivity of rhythmic queries, yielding a query by tapping system for casual music retrieval which can be trained to users to account for their subjective querying style. After training the system, a user can tap a rhythm to re-order their music collection by rhythmic similarity to their query. The top 20 highly ranked results are listed on-screen as a music playlist, from which the user could also then select a specific song.

Query by tapping provides an example of a casual music interaction which suffers from subjective queries. In mobile music-listening contexts, it can often be inconvenient for users to remove their mobile device from their pocket or bag and engage with it to select music. This tapping of music as a querying technique for music is depicted in figure 2. Tapping a rhythm is already a common act and rhythm is a universal aspect of music [13]. In an exploratory design session where users were asked to provide rhythmic queries, it became apparent that users differed in querying style. We describe this subjective behaviour and our approach to modelling it in previous work [1]. One of the key aspects of the model is that users have preferences for which instruments they tap to, as depicted in figure 1.

In order to assign a belief to the songs in the music collection given a rhythmic query, we compare the query to those predicted by the user input model. This comparison is done using the edit distance from string comparison methods, scaling the mismatch penalty to the time differences between the rhythmic sequences [5].

3.1 Query By Tapping

'Query by Tapping' has received some consideration in the Music Information Retrieval community. The term was introduced in [7] which demonstrated that rhythm alone can be used to retrieve musical works, with their system yielding a top 10 ranking for the desired result 51% of the time. Their work is limited however in considering only monophonic rhythms i.e. the rhythm from only one instrument, as opposed to being polyphonic and comprising of multiple instruments. Their music corpus consists of MIDI representations of tunes such as "You are my sunshine" which is hardly analogous to real world retrieval of popular music.

Rhythmic interaction has been recognised in HCI [8, 15] with [4] introducing rhythmic queries as a replacement for hot-keys. In [2] tempo is used as a rhythmic input for exploring a music collection – indicating that users enjoyed such a method of interaction. The consideration of human factors is also an emerging trend in Music Information Retrieval [12]. Our work draws upon both these themes, being the first QBT system to adapt to users. A number of key techniques for QBT are introduced in [5] which describes rhythm as a sequence of time intervals between notes – termed inter-onset intervals (IOIs). They identify the need for such intervals to be defined relative to each other to avoid the user having to exactly recreate the music's tempo.

In previous implementations of QBT, each IOI is defined relative to the preceding one [5]. This sequential dependency compounds user errors in reproducing a rhythm, as an erroneous IOI value will also distort the following one.

Figure 2: Users are able to select music by simply tapping a rhythm or tempo on the device, enabling a casual eyes-free music interaction.

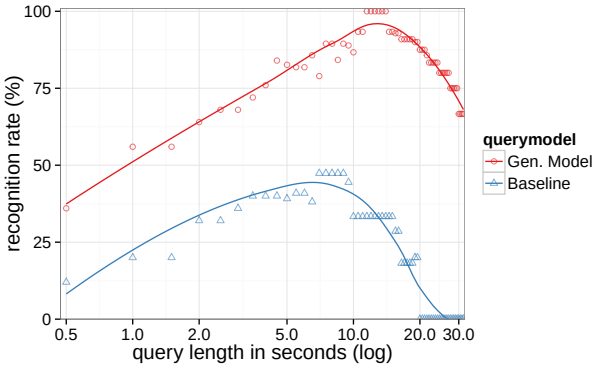

Figure 3: Percentage of queries yielding a highly ranked result (in the top 20 i.e. 6.7%) plotted against query length in seconds.

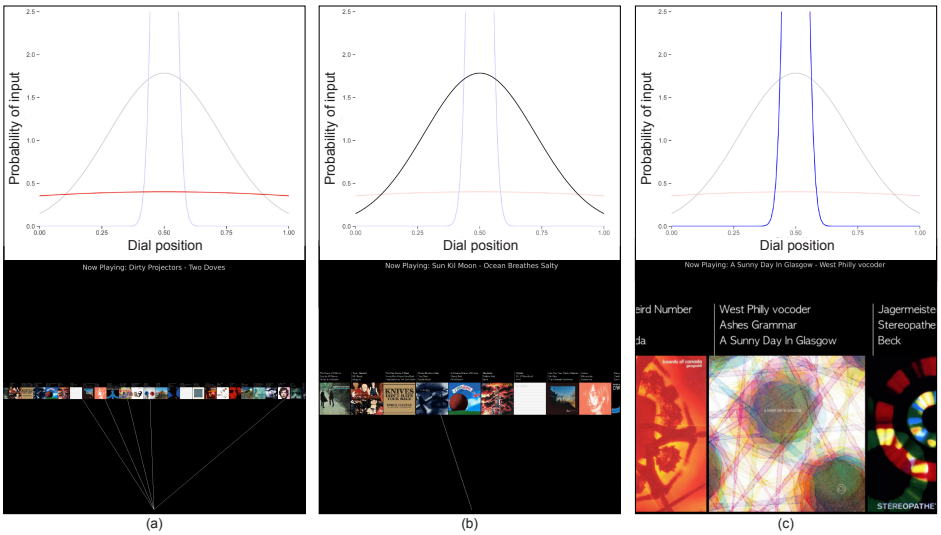

Figure 4: As the user asserts control, the distribution of predicted input for a given song becomes narrower. This adds weight to the input, meaning a belief is inferred over fewer songs and the view zooms in.

The approach to rhythmic interaction in [4] however used k-means clustering to classify taps and IOIs into three classes based on duration. The clustering based approach avoids the sequential error however loses a great deal of detail in the rhythmic query and so we explore a hybrid approace.

3.2 Evaluation

The most important metric for the system to be usable was whether a rhythmic input produced an on-screen (top 20) result. We asked eight participants to provide queries for songs selected from a corpus of 300 songs which we had complete note onset data for. Participants listened to the songs first to ensure familiarity and were asked to provide training queries for each song. These training queries were used to train the generative model using leave-one-out cross-validation. We use a state-of-the-art onset detection algorithm (based on measuring spectral flux [10]) as a baseline which does not account for subjectivity. Performance typically improves with query length as seen in figure 3. Higher rankings are achieved for all query lengths when using the generative model. Interestingly, queries over 10 seconds lead to a rapid fall-off in performance - possibly due to errors accumulating beyond the initial query the user had in mind or due to users becoming bored.

4. MODELLING ENGAGEMENT

We consider casual search interactions as spanning a range of levels of engagement. How much a user is willing to engage with a system and provide evidence of their listening intent will undoubtedly vary with listening context. An interaction which is fixedly casual would be as problematic as one which requires a user's full attention, with users unable to take control when they wish to. An example of this would be old analogue radios – whilst they offer a simple music interaction, users have limited control over what they hear. Previous work by Hopmann et al. sought to bring the benefits of interaction with vintage analog radio to modern digital music collections [6], however their work also required explicit selection (a fixed level of engagement).

We explore how the inference of listening intent can be conditioned upon the user's level of engagement, with the music interaction spanning from casual mood-setting through to specific song selection. While it would be desirable to bring the simplicity of radio-like interaction to modern music collections, mapping a modern music collection to a dial such as in figure 5 would require prolonged scrolling. An alternative would be to instead support scrolling through an overview of the music space however this removes granularity of control from the user, leaving them unable to select specific items. We developed a radio-like system called FineTuner that allows users to navigate their music, which is arranged along a mood axis. Users can 'tune in' to a mood to hear recommended songs based on their listening history. FineTuner allows the user to assert control over the music recommendation by applying pressure to a sensor. This enables users to seamlessly transition from a casual style of interaction akin to a radio to controlling styles such as specifying a particular sub-area of interest in a music space, or even selecting individual songs. FineTuner provides a single interaction which supports casual search through to fully engaged retrieval.

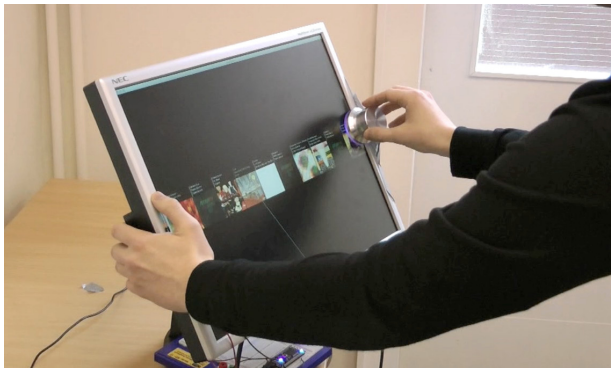

Figure 5: Users share control over an intelligent radio system, using a knob and pressure sensor.

4.1 Varying Engagement

Our system enables both casual and engaged forms of interaction, giving users varying degrees of control over the selection of music. In casual interactions where users apply less pressure, the system can become more autonomous – making inferences from prior evidence about what the user intended. This handover of control was termed the 'H-metaphor' by Flemisch et al. where it was likened to riding a horse – as the rider asserts less control the horse behaves more autonomously [3]. By allowing users to make selections from the *general* to the *specific*, the system supports both specific selections and satisficing. Users can make broad and uncertain *general* selections to casually describe what they want to listen to. However, they can also assert more control over the system and force it to play a *specific* song. Control is asserted by applying force to a pressure sensor.

As the user begins an interaction, they have not applied pressure and therefore are not asserting control over the system. The inferred selection is thus broad, covering an entire region of their collection and is biased towards popular tracks (fig. 4a). The music in the inferred selection is visualised by randomly sampling tracks from it and drawing beams from the dial position to the album art. The user may press in the knob to accept the selection and the sampled track is played. At low levels of assertion it is likely that most tracks played would be highly popular tracks. This behaviour is a design assumption, users may want the system to use other prior evidence. When the user applies pressure, the system interprets this as an assertion of control. The inferred selection is smaller and the spread of beams becomes narrower, the album art visualisation zooms in to show the smaller selection (fig. 4b). This selection is a combination of evidence from the dial position with prior evidence i.e. their last.fm music history. When users fully assert control (max. pressure), they navigate the collection album by album (fig. 4c) and can make exact selections. By varying the pressure, users seamlessly move through this continuous range of control.

The smooth change in engagement is achieved using a simple model of user input. We assume that in an engaged interaction, users will point precisely at the song of interest (as in fig. 4c). For more casual selection, we assume that users will point in the general area (mood) of the music they want, modelled using a normal distribution as in (fig. 4b). As less pressure is applied the distribution is widened, leading to less precise selection and a greater role for a prior belief over the music collection such as listening history.

5. SUMMARY

The scenarios explored here involve casual music retrieval, where users have an ill-defined information need and browse for hedonic purposes or to satisfice a music selection. In these cases, considering what input a user would provide for target songs and inferring selections is an intuitive approach which avoids the issue of defining relevance. We show two music interactions which support the uncertain selection of music, inferred from casual user input such as tapping a rhythm or turning a radio dial.

We have shown that modelling user input for inferring music selection can address issues of subjectivity by taking a user-centered approach to model development. The model can be iterated by comparing its predictions against actual user behaviour. Accounting for this subjectivity can yield significant improvements in retrieval performance as well as creating a more personalised search experience. A key feature of the second system, FineTuner, is its ability to span seamlessly from casual search scenarios, such as satisficing, through to more explicit selections of music. By conditioning the inference upon the user's level of engagement, we are able to interpret the same input space (in this case the dial) according to the current context.

Our approach to casual music interaction empowers the user to enjoy their music while expending as much or as little effort in the retrieval as they wish, providing queries in their own subjective style. Instead of focusing solely on optimising the retrieval process, we consider it equally important to design retrieval systems which suit how the user currently wants to interact. By considering how users might provide casual evidence for their listening intent, we achieve music interactions as simple as tapping a beat or tuning a radio.

6. ACKNOWLEDGMENTS

We are grateful for support from Bang & Olufsen and the Danish Council for Strategic Research.

7. REFERENCES

[1] Boland, D., and Murray-Smith, R. Finding My Beat: Personalised Rhythmic Filtering for Mobile Music Interaction. In *MobileHCI 2013* (2013).

[2] Crossan, A., and Murray-Smith, R. Rhythmic Interaction for Song Filtering on a Mobile Device. *Haptics and Audio Interface Design* (2006), 45–55.

[3] Flemisch, O., Adams, A., Conway, S. R., Goodrich, K. H., Palmer, M. T., and Schutte, P. C. NASA/TMâĂŤ2003-212672 The H-Metaphor as a Guideline for Vehicle Automation and Interaction, 2003.

[4] Ghomi, E., Faure, G., Huot, S., and Chapuis, O. Using rhythmic patterns as an input method. *Proc. CHI* (2012), 1253–1262.

[5] Hanna, P. Query by tapping system based on alignment algorithm. In *Proc. ICASSP* (2009), 1881–1884.

[6] Hopmann, M., Vexo, F., Gutierrez, M., and Thalmann, D. Vintage Radio Interface: Analog Control for Digital Collections. In *CHI 2012: Case Study* (2012).

[7] Jang, J., Lee, H., and Yeh, C.-H. Query by Tapping: A New Paradigm for Content-based Music Retrieval from Acoustic Input. *Proc. PCM* (2001).

[8] Lantz, V., and Murray-Smith, R. Rhythmic interaction with a mobile device. In *Proc. NordiCHI*, ACM (2004), 97–100.

[9] Laplante, A., and Downie, J. S. Everyday life music information-seeking behaviour of young adults, 2006.

[10] Masri, P. *Computer modelling of sound for transformation and synthesis of musical signals.* PhD thesis, University of Bristol, 1996.

[11] Scheibehenne, B., Greifeneder, R., and Todd, P. M. What Moderates the Too-Much-Choice Effect? *Journal of Psychology & Marketing 26(3)* (2009), 229–253.

[12] Stober, S., and Nürnberger, A. Towards user-adaptive structuring and organization of music collections. *Adaptive Multimedia Retrieval. Identifying, Summarizing, and Recommending Image and Music* (2010), 53–65.

[13] Trehub, S. E. Human processing predispositions and musical universals. In *The Origins of Music*, N. L. Wallin, B. Merker, and S. Brown, Eds. MIT Press, 2000, ch. 23, 427–448.

[14] Wilson, M. L., and Elsweiler, D. Casual-leisure Searching: the Exploratory Search scenarios that break our current models. In *HCIR 2010* (2010).

[15] Wobbrock, J. O. Tapsongs: tapping rhythm-based passwords on a single binary sensor. In *Proc. UIST* (2009), 93–96.