# Rhythmic Interaction for Song Filtering on a Mobile Device

Andrew Crossan[1] and Roderick Murray-Smith[1,2]

[1] Department of Computing Science,
University of Glasgow,
Scotland, UK, G12 8QQ
{ac, rod}@dcs.gla.ac.uk
[2] Hamilton Institute,
NUI Maynooth, Co. Kildare, Ireland

**Abstract.** This paper describes a mobile implementation of song filtering using rhythmic interaction. A user taps the screen or shakes the device (sensed through an accelerometer) at the tempo of a particular song in order to listen to it. We use the variability in beat frequency to display ambiguity to allow users to adjust their actions based on the given feedback. The results of a pilot study for a simple object selection task showed that although the tapping interface provided a larger range of comfortable tempos, participants could use both tapping and shaking methods to select a given song. Finally, the effects of variability in a rhythmic interaction style of interface are discussed.

## 1 Introduction

Mobile music players are an important part of everyday life for many people on the move. While the capacity of these devices and therefore the amount of information stored on them has increased dramatically in the last few years, they have got smaller and smaller. While this is important for portability, it presents interesting challenges in allowing people to interact with these devices in a fun, but efficient manner. Especially while on the move, where staring at a small screen while navigating a busy environment might prove difficult and potentially dangerous. One interesting example of a prototype system, where the music adapts to aid navigation is discussed in [1]. Directional information and ambiguity are presented to the user through volume and audio panning of the current track to direct the user to a destination or alert them when a choice of different paths is available.

This paper examines a novel concept for allowing users to browse their music library by tempo. It attempts to provide an intuitive, fun and, importantly, 'eyes free' method of interaction that allows the user to interact with the device while walking and potentially without removing it from a bag or pocket.

### 1.1 Mobile Gesturing

Mobile devices are now widely used for a variety of everyday tasks. However, due to the requirement for a small screen and keyboard, interacting with these devices often

D. McGookin and S. Brewster (Eds.): HAID 2006, LNCS 4129, pp. 45–55, 2006.

proves to be difficult. On-screen buttons are generally closely grouped together making interactions slow and error prone. The combination of reduced resources and the need for mobility may mean that a direct translation of the desktop interaction techniques may not provide the best method of interacting with these devices. The development of new interaction techniques and technologies provides the opportunity for a more continuous form of interaction. There are currently several examples of gesturing in mobile devices available in the literature. Pirhonen, Brewster and Holguin [2] demonstrate an example of stylus gesturing as an input technique for controlling a PDA based MP3 player. These interactions are designed to be intuitive for the task performed. Pirhonen, Brewster and Holguin were able to demonstrate significant usability benefits with the gesture interface over the standard interface, with users indicating that the gesture system required a lower workload to perform the task.

Recent commercially produced mobile devices have been developed with integrated low cost accelerometers. For example, both Samsung and Nokia have developed phones (the Samsung SCH-S310 and Nokia 3220) that use integrated accelerometers as an interaction mechanism. These sensors have previously been examined for use in context sensing applications, but also offer the potential for allowing a user to control a device through gestures. Strachan *et al.* [3] describe the bodyspace project where a mobile device can be controlled through movement. Acclerometers attached to the mobile device detect the movement of the device from one pre-defined position to different areas around the body. A control action can then be assigned to each area of the body allowing a user to open documents or applications through body-related gesturing.

Hinkley *et al.* [4] demonstrate how a combination of sensors can be used to interact with a mobile device. Recognition of orientation of the device was, for example, used to start a voice recording application (in combination with proximity sensors and touch sensors) or automatically change the screen orientation. Hinkley *et al.* demonstrate how these novel sensors can be used to provide a natural and intuitive interaction mechanism.

### 1.2 Rhythmic Interaction

The methods described above all use point-to-point gestures where the user starts in one position and moves through a specific trajectory to another set position to finish the gesture. Rhythmic gesture methods of interaction have largely been ignored as an input mechanism for computers. However, they have the potential to offer a natural method of interacting with a device for several types of task. This is particularly important in the case of mobile devices, where their small size and need for portability can lead to slow and frustrating interactions. The introduction of new, affordable sensor technology, allows greater possibilities for interacting with these devices.

One benefit of rhythmic gestures is that they can be designed to be naturally repeatable. Tapping out a rhythm on a desk with a finger or dancing to music would be two everyday examples of repeatable rhythmic gestures. If the user taps out a rhythm or performs a rhythmic gesture, it feels natural for the user to repeat this gesture until recognition occurs. This allows the user to perform the gesture to excite a state in the system (akin to finding the ‘resonant frequency’ of particular states or options).

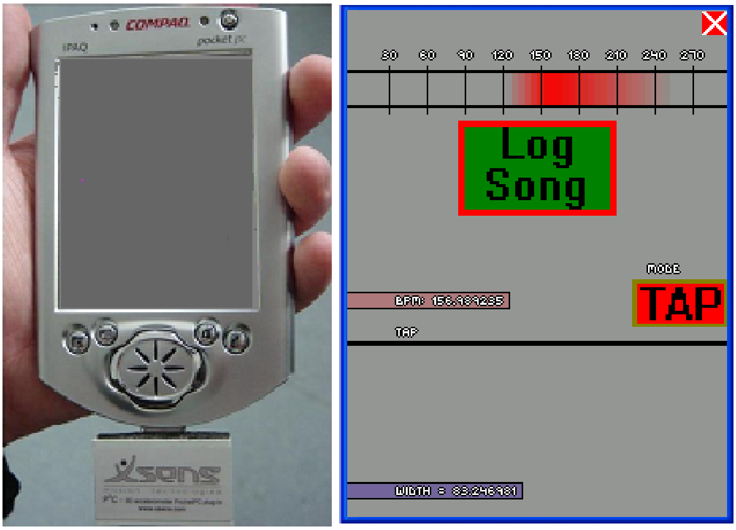


**Fig. 1.** PDA with the Xsens P3C accelerometer attached to the serial port (**left**). The song browsing environment (**right**).

The system can then respond by presenting a series of options based on a probabilistic model of how likely it is that the user is trying to select each option. More likely options can be displayed to the user more prominently, and the dynamics of the system can be altered such that it is easier to select more probable options. The user can then refine his or her movements, responding to the feedback, to select the target. Lantz and Murray-Smith [5] conducted an initial usability study using rhythmic gestures with a mobile device. They examined 10 different rhythmic gesture movements recorded through an accelerometer attached to a mobile device to look at consistency and ease of movement for each of the gestures. They demonstrate a dynamic movement primitive approach that is used to model the gestures to eventually provide recognition. Rhythmic interaction in the form of 'haptic dancing' with a PHANToM force-feedback device is described in [6]. One of the insights from this work is that future interaction with a computer might not be similar to the current command-and-control style of interaction, but more like the flowing transitions of control and the give-and-take of dancing.

Synchronisation of oscillators and phase entrainment has been studied extensively in the physics literature [7], and the theory has recently been applied to the design of rhythmic interaction methods. In [8], we examined extraction of gait phase from an accelerometer attached to a mobile device held by a user while walking applied to mobile usability. The acceleration signal, allowed extraction of each step and further allowed an estimate to be made at each point of the user's current phase within a step. Here we present a rhythmic interaction system for song filtering on a mobile device. For input, we use repeated screen tapping or shaking gestures sensed using an accelerometer. Songs are selected by synchronising the tapping or device oscillations with the tempo of the song being selected.

## 2 The System

This system was developed using an HP 5450 PDA with the Xsens P3C 3 degree of freedom linear accelerometer attached to the serial port (shown on the left of Figure 1). Its effect on the balance of the device is negligible (its weight is 10.35g). The accelerometer was used to detect movement of the device, sampling at a rate of approximately 90Hz.

The interface used for the study is displayed on the right of Figure 1. The top section of the screen shows a scale indicating the user's current beats per minute (BPM), and the uncertainty in tempo due to variability in the timing of the user's beats. To produce a beat, the user must either tap the designated area of the screen (below the dark line shown on the right of Figure 1), or shake the device. The button on the right allows the user to change between tap mode where the user taps out a tempo, and accelerometer mode where the user moves the device rhythmically at a tempo through the air. When processing the accelerometer data, a low-pass filter is applied in real time to remove muscle tremor and sensor noise from the trace, and a high pass filter is used to correct for drift over time (as the user's posture may change). Examples of processed acceleration traces for the device's vertical accelerometer axis are shown in Figure 2.

Songs of differing tempo are stored in the system. A song is presented to the user by a short clip of looped audio from the song and is associated in the search space with its tempo. The volume of the song played to the user is then a function of the distance between the tempo of the song and the current detected BPM. The volume is also dependent on the variability in the user's tapping. A high level of variability may be due to the user browsing the space. The system therefore presents the user with more context information, represented by a higher level of volume from nearby songs. Low BPM variability will occur when a user has decided on a target and adjusts his or her tempo to match the selected song. The system then presents a more focussed area of the space to the user thereby reducing the volume of songs whose tempo is nearby but does not match the current BPM. The variability is displayed to the user by the width of the distribution at the current position shown on the BPM scale.

This can be viewed as a form of the pointerless selection mechanism described by Williamson and Murray-Smith [9], which is based on detection of control behaviour in the user's observed actions.In this case, the metaphor is that each track has a 'resonant frequency' which can be 'excited' by the user's tapping actions, such that the initial beats narrow down the search space. The current uncertainty in the negotiation process is indicated by a volume which increases the closer the user is to the resonant frequency for that track. The display indicates the ambiguity in the tapping behaviour, by visually showing the spread of frequencies being excited, and also in sound by spreading the excitation to songs proportional to the spread in beat frequencies. Options close to the excitation frequency have a volume proportional to their likelihood, as a stimulus to the user, which allows the user to entrain more precisely with the desired song, homing in to the final selection.

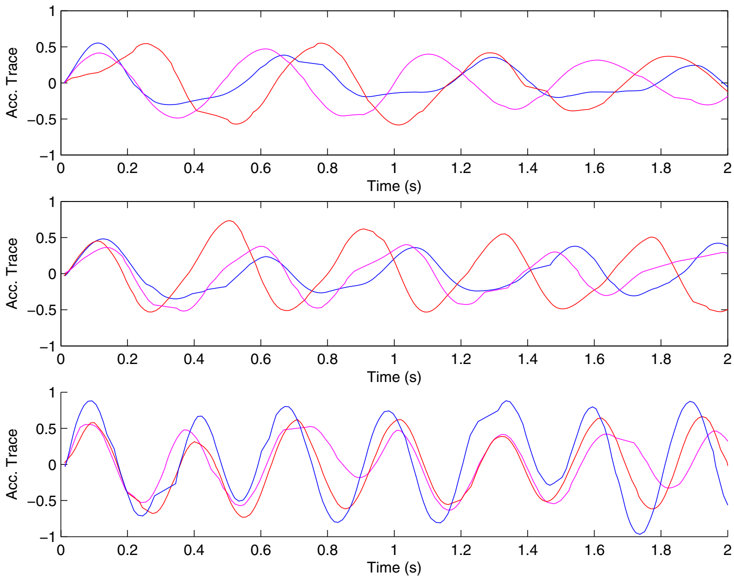


**Fig. 2.** Data from three participants when in acceleration mode and synchronising with a song of 112BPM (top), 135BPM(Middle) and 192BPM (Bottom)

## 3 Pilot Results

A pilot study was carried out with a small range of tracks to examine how the system was used to browse within the space. Seven participants took part in the pilot study using the application described with seven songs of differing tempo stored in the environment. The songs chosen had tempos of 72, 87, 112, 135, 150, 170 and 192 beats per minute. Users were asked to target each song using accelerometer and tapping modes in a counterbalanced order. Users successfully managed to target and maintain synchronisation for several beats with a mean of 6.7 of 7 songs using tapping mode and 4.9 of 7 songs using accelerometer mode. This difference is to be expected as the tapping interface allowed the users to easily look at the screen and therefore use the BPM scale to browse the space. Also, *post hoc* analysis of the data suggests that 5 unsuccessful targeting attempts in the accelerometer condition were due to poor performance of the recognition algorithm for slower tempo songs. The tapping mode allowed a larger range of motions to be incorporated due to the lower physical demand required to tap the stylus on the screen as opposed to moving the device through the air. The acceleration traces for one axis for three users targeting three songs of tempo 112, 135, and 192 beats per minute are shown in Figure 2. In this instance for the specific movement that the participants were instructed to perform, each cycle of the sinusoid corresponds to one oscillation with the device. This would not necessarily be true for different motions.

Observation during the experiment suggested that some of the participants were able to use the variability in tempo to browse the space. A common technique was to start with slow beats, and increase rapidly to move the current BPM position up through the scale. This presented the user with a range of tempos from the slowest to the fastest BPM while maintaining a high level of variability. This ensured that a wide distribution of the space was always presented to the user at different positions on the BPM scale, which is useful for browsing the space.

### 3.1 Considering Variability in Tapping

When considering this, or any gesture-interaction mechanism, it is important to take into account the effect of variability of the user's actions in the recognition process. When analysing beats per minute, it can be seen that:

$$BPM = \frac{60}{IBG}, \tag{1}$$

where BPM is beats per minute, and IBG is the inter beat gap in seconds. If we introduce a bias error in the time between taps into this equation, a new value for the predicted beats per minute is given by:

$$BPM_{varLim} = \frac{60}{(IBG + v)}, \tag{2}$$

where BPM$_{varLim}$ is the value of the BPM measured given where variability in inter beat gap is $v$. If the target BPM was 120, and therefore the target inter beat gap 0.5seconds, a variability of 0.1 would correspond to the user tapping with an inter beat gap of 0.6 seconds. The offset can then be obtained by looking at the the difference between these values,

$$BPM - BPM_{varLim} = x, \tag{3}$$

where $x$ is the difference in target BPM and recorded BPM for a given variability. Substituting for BPM and BPM$_{varLim}$ using equations 1 and 2, we can derive an equation for $v$ in terms of IBG such that

$$v = \frac{xIBG^2}{1 - xIBG}, \tag{4}$$

again, substituting BPM for IBG using equation (1), we can see that

$$v = \frac{60x}{BPM^2 - 60xBPM}, \tag{5}$$

This equation can be used to plot the effect of tap variability on BPM. Figure 3 shows the inter tap variability required to target at different BPMs for accuracy thresholds of +3 to +21BPM. Figure 4 shows the corresponding plot for an accuracy threshold below the target BPM for -3BPM to -21BPM. It can immediately be seen that positive inter tap variability have a far larger effect on the BPM level than negative inter tap variabilities. Also, it can be seen that as the beats per minute increase, a smaller variability is required to keep the BPM within a given threshold.

These show the effect of variability in one tap or a group of taps with a constant variability. However, what it does not take account of is a users ability to synchronise to a particular beat. For example, if a user is given a song to tap along with, he or she may display variability with individual beats, but can use his or her knowledge of the

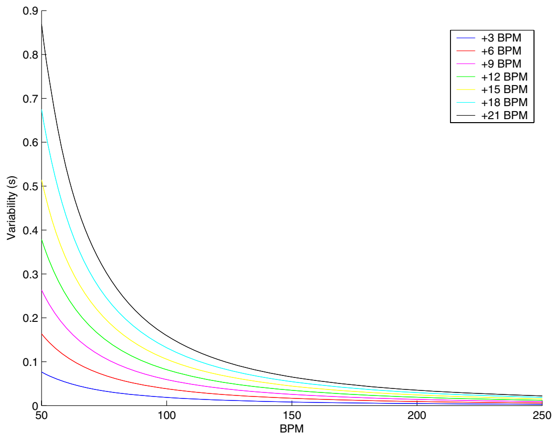


**Fig. 3.** Inter tap variability required to target within the given accuracy threshold above the target BPM for different BPM levels

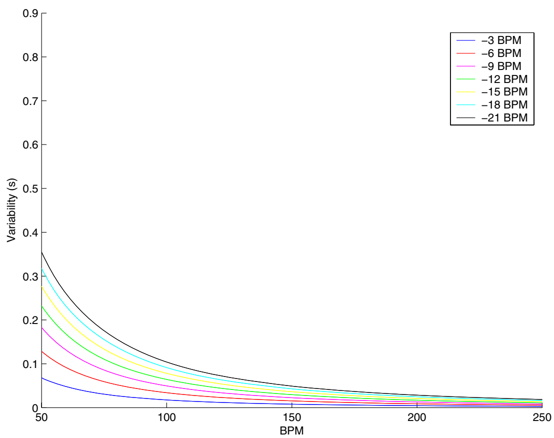


**Fig. 4.** Inter tap variability required to target within the given accuracy threshold below the target BPM for different BPM levels

song so far to predict the next beat, and therefore compensate for the initial variability. Measurement of actual user variability is important to determine how the user uses the music to synchronise.

In Figure 6 we show the effect of variability in tapping accuracy as a test user tries to locate songs at a range of BPM. The standard deviation of the Inter Beat Gap for songs at BPM of 72,112, 135, 170 and 192, was respectively 0.0383, 0.0435, 0.0200, 0.0210, 0.0171s, as shown in Figure 5, but although the slower songs have a lot of variability, in absolute terms, it does not translate into greater uncertainty, as shown in Figure 6. Note that the means are also extremely close to the target BPM, showing the precision users can achieve with the system.

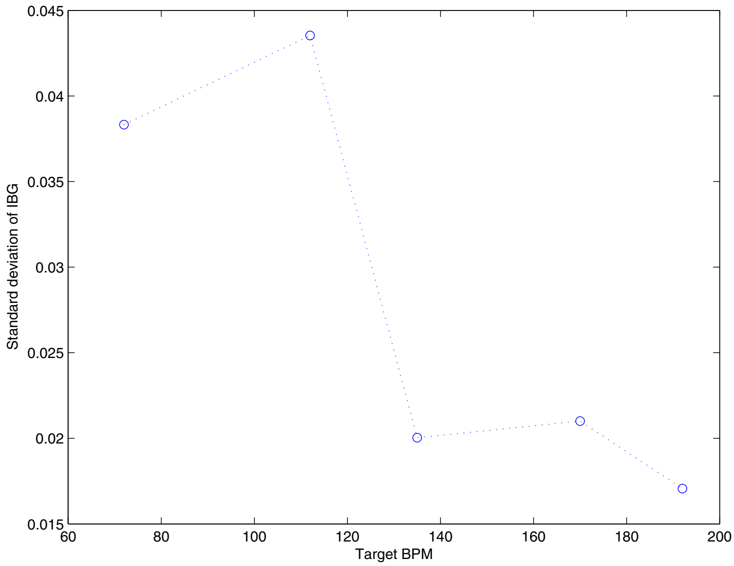


**Fig. 5.** Inter Beat Gap standard deviation at different BPM levels

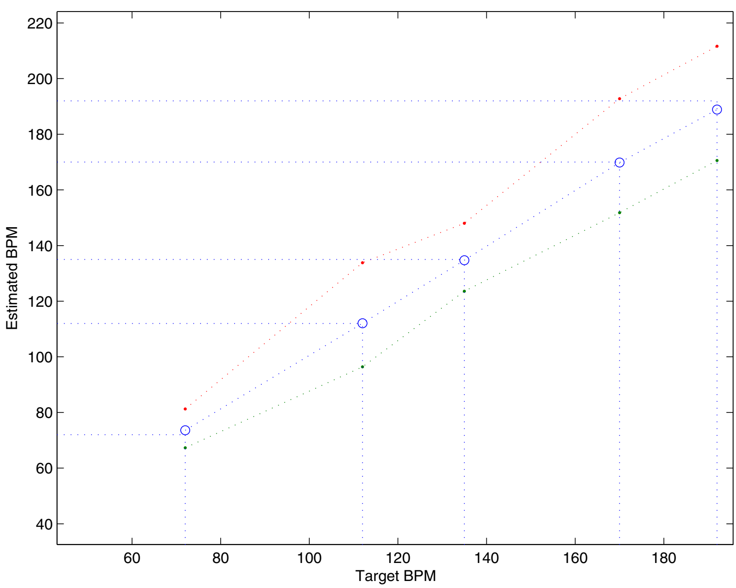


**Fig. 6.** Uncertainty of BPM estimate at different reference BPM levels (showing the spread of BPM associated with $\pm 2\sigma$ – two standard deviations in the inter-beat gap)

We can also get further insight into the working of the algorithm by looking at the time evolution of the mean and standard deviation of the BPM estimate, as shown in Figure 7, where 3 different runs are shown where a user acquires the desired BPM

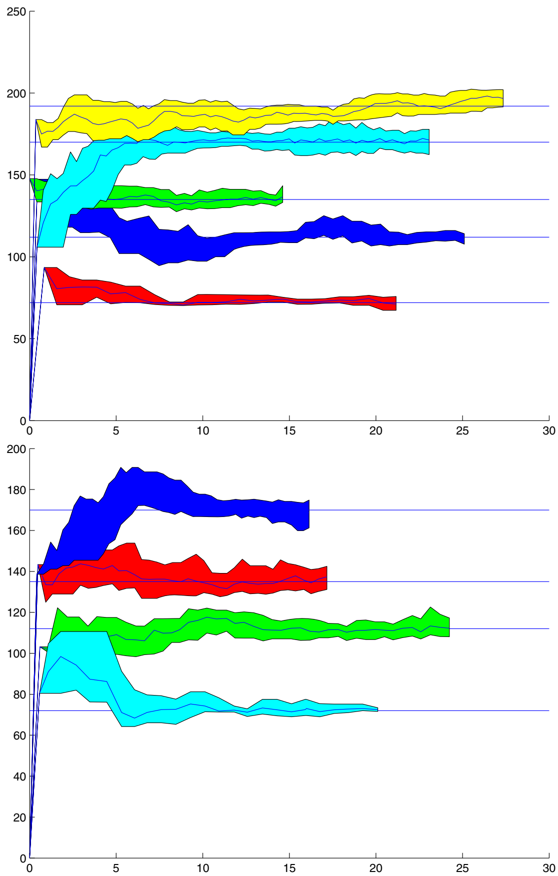


**Fig. 7.** Illustration of mean and standard deviation evolution as the user attempts to select each of 5 tracks with different BPM

level, while receiving feedback from the system. It can be seen that initially the user estimates the BPM for the appropriate song. There is a high level of variability as the user searches the space. The system supports this by increasing the volume of nearby songs due to this high variability of tapping. As the user hears and recognises the desired song, the mean tapping BPM moves towards the song's BPM and the level of variability of tapping tempo decreases presenting the user with a focussed area of the space centred around the song.

## 4 Discussion and Extensions

This system was successful in allowing users to browse an audio space of 7 songs. However, obvious scalability issues need to be examined before this interaction mechanism is shown to provide some benefit in this area. We envisage a system based on tempo to be used in conjunction with other complementary song selection mechanisms.

This should not be seen as a song selection technique, but as a system of filtering a selection of songs, as using tempo alone for selection would require the system itself to have more in-built intelligence. An extension that has been developed to the system described above combines different filtering mechanisms to provide more differentiability. The user rhythmically taps to screen to provide the tempo, while tilting the device to different orientations to browse different genres simultaneously.

In a rhythmic interaction system, it is important to consider how variability in the user's actions and timing will affect the performance of the system. In the initial implementations of the above interface, the user's BPM was determined by averaging the time between a fixed number of beats. This method was found to penalise convergence with slower songs, as with this system, synchronising with a slower song will always require more time to generate the same number of beats. An alternative method is to only consider beats within a certain time period. For example, if we only consider beats with 5 seconds of the current time, tapping with a tempo of 60BPM will use 5 beats when determining the mean. tapping at a rate of 120BPM will lead to 10 beats being considered when calculating the mean. This method provides a good compromise to robustness from the effects variability and rate of convergence for slower songs.

In this work we have used meta-data as a reference for the response frequency for each song. A more sophisticated approach is to derive the timing of beats for an individual track using signal processing techniques, and then require the user to tap in time with the automatically extracted beat-information. There are interesting challenges in this, including the subjective nature of music perception (different people might want to entrain with different aspects of the track, and this should be supported). For examples of modern automated analysis of music tracks see the Intelligent Sound project `www.intelligentsound.org`.

Finally, an application has been developed to demonstrate how rhythmic interaction methods can be used to adapt the user interface. The system analyzes the oscillatory acceleration history from the user walking or jogging, and will select the song of the closest tempo to the user's step rate in order to synchronise with the user. With an online Hilbert transform for gait phase estimates and techniques to adapt the playing of the music, such as granular synthesis, a self-tuning approach could adjust the music to synchronise with the jogging, creating a phase-locked loop between user and music.

## 5 Conclusions and Future Work

This paper describes an interaction technique for allowing users to filter songs by interacting with a mobile device through rhythmic gestures, and synchronising the tempo of the interactions with the tempo of a song stored in the environment. Different classes of object (not just songs) could be accessed in such a manner. Using tempo to browse songs seemed to be an intuitive and enjoyable concept to all users tested. With minimal or no training, all users were able to to pick up the device and synchronise with songs.

When selecting a song, the user is asked to produce a constant tempo. There is potential for a very rich and expressive method of interaction with the system by allowing differing rhythms to be entered. Particularly for the accelerometer case where the continuous interaction will allow users to adjust the form of their movements to provide more information to the system. This will become important for the scalability of the system where, for example, users could be more expressive, and modify their action between beats to select genre.

A key feature of this interface is the fun people had in relatively carefree exploration of the music space. Efficiency of selection might not always be the only metric for applications that focus on entertainment, such as audio or photo browsing. One user commented that it brought a bit of the more ‘grainy’ feel of old-fashioned radios with tuning dials to the relatively ‘antiseptic’ world of digital audio.

Finally the accelerometer offers the potential to produce a system to allow context sensitive selection of objects. For example, the system could detect a user walking or jogging, and select a random song from the user’s music library of the appropriate tempo to synchronise with the user’s step rate. Alternatively, the system could monitor the user’s heart rate and coax them to increase or decrease their step rate by playing songs of a slightly different tempo to the step rate and by exploiting phase entrainment. Future work will require a formal evaluation of these techniques, and fine-tuning of phase relationships during selection.

## Acknowledgements

The authors are grateful for the support of SFI grant 00/PI.1/C067. RM-S is grateful for support of EPSRC grant GR/R98105/01, and BRG project *Continuous Gestural Interaction with Mobile Devices*. This work was supported in part by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the authors’ views.

## References

1. Warren, N., Jones, M., Jones, S., Bainbridge, D.: Navigation via continuous adapted music. In: ACM SIG CHI 2005, Portland (2005)
2. Pirhonen, A., Brewster, S.A., Holguin, C.: Gestural and audio metaphors as a means of control for mobile devices. In: Proceedings of ACM CHI 2002 (Minneapolis, MN), ACM Press Addison Wesley (2002) 291–298
3. Strachan, S., Murray-Smith, R., Oakley, I., Ängeslevä, J.: Dynamic primitives for gestural interaction. In Stephen Brewster, M.D., ed.: LNCS 3160: Proceedings of Mobile Human-Computer Interaction MobileHCI 2004: 6th International Symposium, Glasgow, UK, Springer-Verlag (2004) 325–330
4. Hinckley, K., Pierce, J., Sinclair, M., Horvitz, E.: Sensing techniques for mobile interaction. In: Proceedings User Interface Software and Technology (UIST 2000), ACM (2000) 91–100
5. Lantz, V., Murray-Smith, R.: Rhythmic interaction with a mobile device. In: NordiCHI ’04, Tampere, Finland, ACM (2004) 97–100
6. Gentry, S., Murray-Smith, R.: Haptic dancing: human performance at haptic decoding with a vocabulary. In: IEEE International conference on Systems Man and Cybernetics, IEEE (2003)
7. Pikovsky, A., Rosenblum, M., Kurths, J.: Synchronization: A universal concept in nonlinear sciences. Cambridge University Press (2001)
8. Crossan, A., Murray-Smith, R., Brewster, S., Kelly, J., Musizza, B.: Gait phase effects in mobile interaction. In: ACM SIG CHI 2005, Portland. (2005)
9. Williamson, J., Murray-Smith, R.: Pointing without a pointer. In: ACM SIG CHI, ACM (2004) 1407–1410